“But Where Are You Really From?” The Cost of Movement:

A Longitudinal Analysis of Migrant Life Satisfaction and Health Outcomes in Germany

Sophie Gabreldar

An honors thesis submitted to the Program in Symbolic Systems in partial fulfilment of the requirements for the degree of Bachelor of Science with Honors

Primary Advisor: Nilam Ram

Second Reader: Harriett Jernigan

Stanford University

June 2026

To the Directors of the Program in Symbolic Systems:

I certify that I have read the thesis of Sophie Gabreldar in its final form for submission and have found it to be satisfactory for the degree of Bachelors of Science with Honors.

05/30/2026

Nilam Ram

Principal Advisor

Departments of Psychology and Communication

To the Directors of the Program in Symbolic Systems:

I certify that I have read the thesis of Sophie Gabreldar in its final form for submission and have found it to be satisfactory for the degree of Bachelors of Science with Honors.

Harriett Jernigan

Second Reader

Program in Writing and Rhetoric

Table of Contents

List of Figures and Tables

Acknowledgments

I am deeply grateful to everyone who made this honors thesis possible. I especially want to thank my advisor, Professor Nilam Ram, for continually encouraging me to pursue my ideas and for supporting my growth as a researcher. This project would not have been possible without your support, guidance, and mentorship. I also want to thank A Garron Torres, the lab manager of The Change Lab at Stanford, for patiently guiding me through this process with kindness and endless support.

I would next like to thank my mom, dad, and brother for their unwavering love and support in both me and my pursuits. To my mom, thank you for teaching me true patience and for always reminding me to be kind to myself and others. To my dad, thank you for teaching me the power of resilience in the face of strife and for always encouraging me to persevere. Finally, to my brother, thank you for being my lighthouse in the dark and my sounding board. I will never be able to fully express how grateful I am to have you all in my life. Thank you for believing in possibilities I had yet to see.

To my friends at Stanford—Kelly, Amelia, Armita, and Alyssa—I cannot imagine completing this project without your support and encouragement. Thank you for the endless laughs and memories. To my friends back home—Lian, Jacob, Chris, Evelyn, Megan, Sarah, Vanessa, and Sonya—thank you for always being a place I can run to, no matter how long it has been.

Abstract

A growing body of research has documented the "healthy immigrant effect" (HIE), whereby immigrants initially exhibit better health outcomes than native-born populations but experience more rapid health decline over time. However, much of this research relies on cross-sectional data and focuses primarily on first-generation migrants, leaving questions about how health trajectories evolve across generations. Applying linear growth models to longitudinal data from the German Socio-Economic Panel (SOEP), I examine changes in life satisfaction, health satisfaction, and self-reported health limitations for N = 99,433 participants (M_{age} = 39 , SD = 17; 50% female) over a 37-year period. The sample includes individuals with no migration background (n = 64,954), direct migration background (n = 29,016), and indirect migration background (n = 5,463).

I found that life satisfaction ($p < .001$) and health satisfaction ($p < .001$) declined significantly over time across the sample, while physical ($p < .001$) and emotional health limitations increased ($p < .001$). Individuals with migration backgrounds reported higher baseline life satisfaction and health satisfaction, but experienced steeper declines and greater decreases over time compared to non-migrants. Patterns of health limitations varied by migration background: direct migrants showed faster deterioration in physical health, while indirect migrants exhibited more moderate, and in some cases slower, decline. Emotional health trajectories also differed, with direct migrants reporting worse baseline emotional health but slower decline over time, and indirect migrants showing patterns more similar to non-migrants. These findings showing that disparities in migrant health emerge over time suggest cumulative impact of post-migration experiences on long-term health outcomes.

Introduction

Motivations

"But where are you *really* from?" is a question I have been asked many times, often leaving me feeling out of place in spaces I consider home. In these moments, I am reminded that movement across borders does not necessarily translate into acceptance or a secure sense of belonging. While personal, the experience of feeling out of place is not unique; across history, people have moved in search of safety, opportunity, and improved living conditions, shaping societies and identities along the way. A search that, in many ways, connects us all.

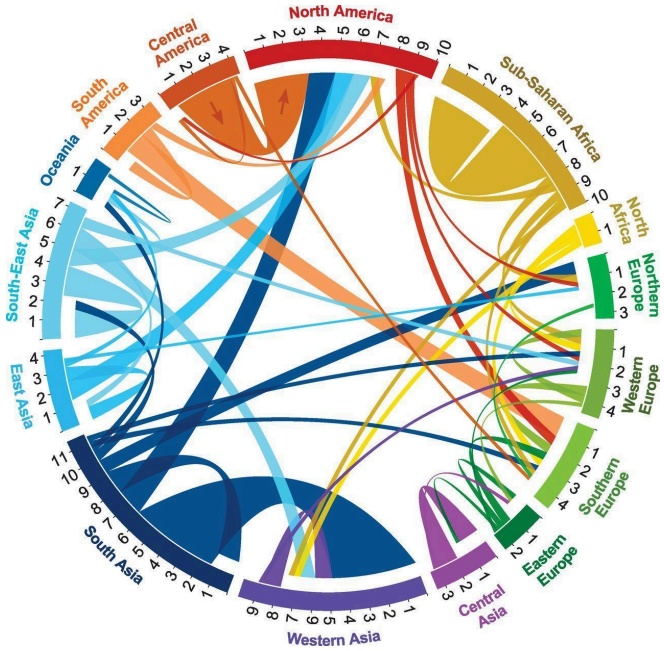
Today, migration is a significant and growing global phenomenon, with an estimated 304 million international migrants worldwide in 2024 (United Nations, 2024). While industrialization and globalization have accelerated the scale and speed of migration, the motivations and experiences of migrants remain highly varied and are shaped by shifting social, political, and economic contexts. Over time, these migration patterns have shifted, with migrants increasingly moving into a relatively small number of key destination countries while originating from a growing range of countries of birth (Wallace, 2024).

Migration refers to the movement of individuals across national borders for a variety of reasons, ranging from the pursuit of economic opportunity to the need for safety from conflict or persecution. The term "migrant" is used as a broad term to capture these varied experiences, while an immigrant refers more specifically to someone who relocates with the intention of settling in a new country, known as a host country. Movement to a new country is not always

voluntary. Refugees are individuals who have been forced to leave their home country and have received legal protection, whereas asylum seekers are still awaiting recognition of that status, and are often living in uncertain conditions (United Nations, 1951; UNHCR, 2024).

Figure 1

Plot of migration flows between and within world regions during 2005 to 2010



Note. Tick marks show the number of migrants inflows and outflows in millions. Only flows containing at least 170,000 migrants are shown. Reprinted from Abel, G. J., & Sander, N. (2014). *Quantifying global international migration flows. Science, 343*(6178), 1520–1522. https://doi.org/10.1126/science.1248676

Migration occurs both within and across world regions with especially large migration flows concentrated between Asia, Europe, and North America (Figure 1; Abel & Sander, 2014). Although migrants make up a relatively small proportion of the global population, they have a

substantial presence in the United Kingdom, the European Union, the United States, and Australia (Wallace, 2024). Additionally, the composition of migrant flows has changed. Earlier waves of migration, particularly in the early to mid-20th century, were largely characterized by the movement of predominantly European labor migrants (Wallace, 2024). In contrast, contemporary migration is far more diverse, with individuals arriving from a wide range of low- and middle-income countries for varied reasons, including employment, education, family reunification, and asylum.

While migration continues to play a central role in shaping societies, its benefits and challenges are experienced unevenly across migrants and the countries in which they settle. Outcomes related to health and well-being are deeply influenced by structural factors, including socioeconomic status, access to resources, and broader systems of inequality (Abdi et al., 2025). These differences highlight that migration is more than the act of simply moving across borders. Rather, migration includes navigating the social, economic, and institutional conditions that shape life after arrival in the host country. In this way, migration extends beyond relocation to impact the lived experiences of migrants that will ultimately influence their long-term health and well-being.

Trends in Migrant Health

Although existing research provides important insight into patterns of health among migrant populations, there are still some key gaps in our understanding. Studies show that, compared to non-immigrants, immigrants tend to live longer but also often spend more years in

relatively poor health (Wallace et al., 2022). These outcomes are not uniform, as health varies substantially by country of origin, destination, and duration of residence. For example, cardiovascular disease patterns differ across migrant groups, with migrants from sub-Saharan Africa experiencing higher rates of stroke, likely linked to higher prevalence of hypertension and diabetes while coronary heart disease remains less common compared to native European populations (Agyemang & van den Born, 2019).

Building on these differences, migrants also face elevated risks of several chronic conditions. There is consistent evidence that migrant populations, compared to non-migrant populations, have higher incidence, prevalence, and mortality rates of diabetes, often with earlier onset compared to host populations, reflecting a combination of genetic factors, environmental changes, and challenges in disease management (Agyemang & van den Born, 2019). In contrast, patterns for cancer differ, as migrants in Europe generally experience lower cancer incidence and mortality compared to non-migrant populations(Agyemang & van den Born, 2019). However, this pattern is not consistent across all groups, as certain migrant populations face higher rates of cancers associated with infectious diseases, including stomach, liver, and cervical cancers (Arnold et al., 2010). In addition to physical health, migrants and non-migrants may also differ in mental health. While findings are mixed overall, refugees, asylum seekers, and undocumented migrants face greater exposure to risk factors, including violence in their countries of origin and stress during migration and after arrival (Lindert et al., 2009).

While the existing evidence shows that migrant health varies across conditions and population, it remains unclear how these outcomes might change over time. Much of the literature relies on cross-sectional comparisons, which limits the ability to capture how health

evolves with duration of residence. As a result, researchers still lack a clear understanding of how structural, behavioral, and environmental factors interact to shape long-term health trajectories among migrant populations. This gap highlights the need for longitudinal approaches that can better capture the dynamic nature of migrant health.

Generational Patterns in Migrant Health

While research has examined migrant health, much of it has focused on first-generation migrants (direct migration), with more limited attention to second-generation populations (indirect migration). Existing evidence suggests that health outcomes differ across generations in important ways, though the direction and magnitude of these differences are not always consistent. For example, second-generation migrants in Europe have been found to face elevated risks of early mortality compared to individuals with non-migrant backgrounds (Wallace et al., 2023). This risk appears to vary by maternal country of origin, with higher mortality observed among children of migrant mothers from regions such as Turkey, sub-Saharan Africa, and Pakistan in Germany (Wallace et al., 2023). These differences may result from specific medical causes, such as congenital conditions, but also from broader structural factors, including differential access to care, socioeconomic disadvantage, and exposure to racism and discrimination.

Differences across generations are also evident in mental health outcomes. Second-generation migrants show slightly higher rates of mental health diagnoses (Salas-Wright et al., 2014). This generational shift has been linked to early exposure to childhood adversity, as well as the intergenerational transmission of migration-related stress and trauma (Salas-Wright et al., 2014). The differences in diagnosis may also reflect variation in health-seeking behavior,

cultural stigma surrounding mental health, and changing interactions with healthcare systems across generations (Bauldry & Szaflarski, 2017). As a result, it remains difficult to disentangle whether these patterns reflect differences in underlying health or differences in access to and use of care.

Other research highlights evidence of an initial health advantage among migrant populations. Using population-wide administrative data from Norway, Evensen and colleagues (2024) find that immigrant-background adolescents have lower rates of both somatic and psychiatric diagnoses compared to non-migrant peers, with a clear gradient across generations. Health advantages are strongest among first-generation migrants and diminish with greater exposure to the host-country environment, with second-generation individuals showing outcomes more similar to native populations (Evensen et al., 2024). Similarly, longitudinal evidence from Germany shows that immigrants exhibit better health than native-born individuals upon arrival, with declines over time more pronounced for subjective and mental health than for objective health measures (Ferrara et al., 2024). These findings suggest that the processes shaping migrant health may differ across domains, with mental health potentially more sensitive to post-migration environments and social experiences.

These patterns reflect the interaction of multiple mechanisms. Selection processes may contribute to initial health advantages among first-generation migrants, while acculturation and behavioral change can reshape health over time. At the same time, exposure to structural and social stressors, such as discrimination, socioeconomic inequality, and barriers to healthcare, may accumulate and contribute to declining health outcomes. These dynamics can persist across

generations, as the effects of disadvantage are carried forward and continue to shape health among second- and later-generation migrants

Evidence from intergenerational studies further supports the idea that these processes may lead not only to convergence with native populations, but in some cases to widening disparities. For example, Aradhya et al. (2022) show that birthweight disparities increase across generations among descendants of non-Western immigrants in Sweden, despite overall improvements in birth outcomes. Similarly, Loi et al. (2021) find evidence of a negative health assimilation process in Finland, in which later generations experience worse physical and mental health outcomes compared to both first-generation migrants and native populations. Together, these findings suggest that exposure to host-country environments does not necessarily lead to improved health outcomes over time but may instead contribute to the accumulation of disadvantage across generations.

The Healthy Immigrant Effect

The "healthy immigrant effect" (HIE), also referred to as the "immigrant paradox," describes the pattern in which immigrants initially exhibit better health than native-born individuals, but experience a more rapid decline in health over time (Elshahat et al., 2021). One way researchers assess these patterns, particularly in relation to mortality, is through the all-cause standardized mortality ratio (SMR), which compares the number of observed deaths in a population to the number expected based on a reference population (Johnson et al., 1996). In migration research, SMR is used to evaluate whether migrant populations experience higher or lower mortality relative to the general population in host countries (Aldridge et al., 2018).

A systematic review of 96 studies found that the summary estimate of all-cause SMR for international migrants was lower than one when compared with the general population in destination countries (Aldridge et al., 2018). That is, migrants have lower mortality in comparison to the general population. This decreased mortality was also shown to persist among both male and female migrants (Aldridge et al., 2018). Interestingly, while there was a mortality advantage for refugees, this was not the case for asylum seekers, who instead have a higher mortality rate in comparison to the general populations of their host countries (Aldridge et al., 2018). While SMRs for all causes of death were lower in migrants compared with general populations, migrants were found to have increased mortality due to infectious diseases and external causes of death (Aldridge et al., 2018).

Beyond mortality, the healthy immigrant effect also manifests in patterns of physical health among migrant populations. Migrants to Europe tend to have lower initial rates of cancer incidence and mortality compared to non-migrant populations, suggesting a health advantage at the time of arrival (Rechel et al., 2013). However, this advantage is not uniform across all conditions. Cardiovascular outcomes vary across migrant groups; for example, migrants from Africa tend to have higher rates of stroke, which may be related to higher prevalence of hypertension and diabetes, while rates of coronary heart disease are often lower compared to native populations (Rechel et al., 2013).

However, relatively little attention has been given to how these patterns persist over time, particularly into midlife and older age. Some evidence suggests that older immigrants experience decreased physical functioning and lower self-reported health compared to native-born individuals (Reus-Pons et al., 2018). Additionally, there is evidence that this initial health

advantage tends to diminish relatively quickly with length of stay; approximately 10 years after arrival, most health differences between immigrants and non-immigrants disappear (Loi and Hale, 2019).

To explain these patterns, the observed health advantage has often been attributed to the positive selection of healthier individuals into migration, at both the individual and national levels (Elshahat et al., 2021). At the national level, this selection process frequently occurs through strict migration policies, such as points-based systems, which favor healthy and well-educated individuals (Ramraj & Siddiqi, 2015). In points-based systems, governments award points to potential migrants based on a set of characteristics deemed valuable, including language ability, education, age, and work experience (Papademetriou & Sumption, 2010). Many of the countries that implement points-based systems are also major destination countries for migrants, such as the United Kingdom, Australia, Canada, and Germany (Papademetriou & Sumption, 2010; Nintcheu, 2025).

While the attribution of the healthy immigrant effect to positive selection is compelling, it does not fully account for the experiences of all migrant groups. In particular, refugees and undocumented immigrants often migrate under conditions that do not reflect positive health selection at either the individual or population level (Elshahat et al., 2021). At the individual level, the observed health advantage has also been attributed to resilience, which enables immigrants to navigate and overcome the challenges associated with adapting to a new country (Estrada-Moreno et al., 2025). In this context, resilience supports the development of strategies that facilitate adaptation, including fostering a sense of community, maintaining life satisfaction,

engaging socially, and navigating experiences of social rejection (Bobowik et al., 2015; González-Rábago et al., 2017).

The magnitude of the healthy immigrant effect varies depending on the specific health outcome assessed and the comparison group used (Holtz, 2021). Research in Germany suggests that these advantages are often concentrated at younger ages and diminish over time, as immigrants report a steeper decline in health satisfaction regardless of socioeconomic gains (Ronellenfitsch & Razum, 2004). This pattern is supported by findings from Holtz and colleagues, who show that health trajectories differ across migrant groups (2021). Non-European migrants tend to exhibit a physical health advantage upon arrival, but a mental health disadvantage compared to native-born individuals, while European migrants experience a more pronounced decline in physical functioning over time. Despite these initial differences, the physical health advantages observed among both groups decrease over time when compared to native-born populations (Holtz, 2021). The healthy immigrant effect therefore suggests that migrant populations often arrive with certain health advantages relative to native-born populations, particularly in measures of physical health and mortality. However, the literature also indicates that these advantages are dynamic rather than permanent, as immigrants may experience steeper declines in health and well-being over time due to a combination of social, structural, and behavioral factors.

The Weathering Hypothesis

Complementary to the healthy immigrant effect, two additional processes may help explain how health differences between migrant and non-migrant populations change over time: weathering and acculturation. While the healthy immigrant effect primarily describes

immigrants' relatively favorable health at the time of arrival, the weathering and acculturation provide possible explanations for why these advantages may diminish across the lifespan. In particular, they explain how migrants may experience steeper declines in health and well-being over time as a result of cumulative social, economic, and environmental stressors.

The weathering hypothesis proposes that the combined effects of social, economic, and environmental adversity can accelerate health deterioration among marginalized populations (Geronimus, 1992). Specifically, weathering refers to the process by which chronic exposure to social and economic disadvantage accelerates aging and leads to an earlier onset of adverse physical health conditions among disadvantaged individuals compared to their more advantaged peers of the same age.

The weathering hypothesis was originally motivated by observations of the earlier onset of chronic conditions, such as hypertension, and their impact on birth outcomes among Black women in comparison to their white counterparts (Geronimus et al., 1991). The weathering hypothesis explains elevated rates of illness, disability, and mortality among African Americans as a physiological response to chronic exposure to structural barriers, material hardship, discrimination, and identity-based stressors (Geronimus, 1992). Although originally formulated in the context of racial health disparities, the weathering hypothesis provides a useful framework for understanding how migrants may experience accelerated declines in health over time due to cumulative social and structural stressors.

The weathering hypothesis suggests that health disparities persist even after accounting for socioeconomic status, as the cumulative effects of lifelong stress gradually weaken physiological resilience (Loi et al., 2025). Experiences of marginalization, which are common

among immigrant populations, can contribute to chronic stress, which has been linked to poorer health outcomes (Schneiderman & Siegel, 2005). The stress associated with social exclusion and systemic barriers may, in turn, lead to accelerated aging and declining health over time. While there is a substantial body of literature on differential aging by race and ethnicity, relatively few studies have examined patterns of weathering across immigration backgrounds. This suggests that migrant populations may experience health trajectories shaped not only by initial conditions, but also by the cumulative impact of structural and social stressors over time.

Unhealthful Acculturation

Another process that may contribute to changes in migrant health over time is acculturation. While the weathering hypothesis emphasizes the physiological effects of chronic stress and structural disadvantage, acculturation focuses on how prolonged exposure to a host society can influence migrants' behaviors, lifestyles, and social practices.

As the world becomes more connected, cultural contact can lead to the sharing, and eventual adaptation, of practices between formerly distinct groups (Fox et al., 2017). For example, a migrant who moves from a country where home-cooked meals and active daily routines are the norm may, over time, begin to adopt more sedentary habits and rely on processed or convenience foods that are more common in the host country. In the case of individuals, exposure to another culture may lead to differences in their beliefs, values, and practices. At both the individual and group level, this process is known as acculturation. Acculturation can be defined as "changes in beliefs, values, identity, or behaviors such as language, customs, diet, or social relationships that occur in minority-culture individuals (immigrant or indigenous) as a result of prolonged contact with the majority culture" (Fox et al., 2021).

This process often unfolds gradually across multiple domains of daily life. For instance, the same individual may initially maintain dietary traditions from their country of origin, but over time adopt host-country eating patterns, shift social networks toward majority-group peers, and experience changes in health behaviors such as decreased physical activity or increased alcohol consumption. At the same time, they may face new stressors, such as navigating unfamiliar institutions, language barriers, or discrimination, which can further shape both behavior and well-being.

One theory explaining the decline in immigrants' health over time has been linked to "unhealthful acculturation" into new lifestyles. Unhealthful acculturation describes how individuals adopt host-country behaviors that are detrimental to overall well-being while losing health practices associated with their ethnic identities. This includes the adoption of less healthy lifestyles and dietary practices, stress associated with adapting to new socioeconomic conditions, and exposure to environmental factors such as racial discrimination (Lu et al., 2017). These changes do not occur all at once, but accumulate over time, contributing to the observed decline in health with increased length of residency (Tuggle et al., 2018).

Using longitudinal data across European contexts, Jang and colleagues (2023) found that while immigrants often arrive with relatively favorable health profiles, changes in health over time are not uniform across outcomes or populations. Instead, health trajectories diverge depending on the type of health measure considered and the degree of exposure to host-country environments. In particular, their findings suggest that declines are more pronounced in domains, such as self-reported health, while more objective indicators show weaker or less consistent patterns of deterioration (Jang et al., 2023). This distinction points to the role of social and

environmental factors, including stress, discrimination, and changes in daily life, in shaping how individuals perceive and experience their health over time (Jang et al., 2023).

The existing evidence also suggests that assimilation does not uniformly lead to improved well-being and may instead produce unintended trade-offs. Research on minority populations in Central Europe finds that greater assimilation is associated with weaker social integration, lower levels of generalized trust, and less clarity in life goals, even in contexts where economic conditions are more favorable (Pusztai & Márkus, 2019). This "paradox of assimilation" highlights that integration into dominant social systems may coincide with the erosion of community ties and identity-based support structures that are protective for well-being (Pusztai & Márkus, 2019).

While adaptation to host-country norms may provide certain opportunities, it can also introduce new risks, as migrants navigate shifting identities, changing lifestyles, and exposure to structural stressors over time. Unhealthful acculturation suggests that declines in immigrant health over time may be shaped not only by aging itself, but also by prolonged exposure to host-country environments and lifestyles. As migrants adapt to new social norms, behaviors, and institutions, the cumulative effects of changing lifestyle practices may contribute to gradual declines in health and well-being across the lifespan.

Migration in Germany

To better understand these patterns in migrant health, Germany acts as a useful case study due to its migration history and the availability of longitudinal data that allow for the examination of changing health trajectories over time. The following section outlines migration in Germany and provides context for the analysis that follows.

Migration has played a central role in shaping Germany's demographic, economic, and social landscape over time. Between the late 17th and 19th centuries, large numbers of individuals left German-speaking regions for Eastern and Southeastern Europe and later for the United States (Hanewinkel & Oltmer, 2018). This trend began to shift in the 20th century, particularly during and after the World Wars, when Germany experienced substantial inflows of refugees and displaced populations (Hanewinkel & Oltmer, 2018). In particular, The Second World War (WWII), driven by Nazi Germany, devastated Europe, most notably through the Holocaust, which resulted in the murder of six million Jewish people and the destruction of entire communities. These events reshaped Europe's population through widespread loss of life and displacement. In the final years of the war, an estimated 12–14 million Germans were forcibly displaced from Eastern regions, with around 600,000 deaths occurring amid widespread violence and trauma (Kuwert, 2009). The aftermath of World War II was marked by the arrival of over 14 million ethnic Germans and displaced persons from Eastern Europe, fundamentally reshaping the country's population structure (Hanewinkel & Oltmer, 2018).

For a long time after World War II, West Germany recruited foreign labor through bilateral agreements with countries such as Italy, Turkey, and Yugoslavia (Hanewinkel & Oltmer, 2018). These "guest workers" were initially expected to stay temporarily and were expected to leave the country after 3–5 years (Razum & Wenner, 2016). A total number of around 14 million guest workers came to Germany starting in 1955 until recruitment stopped in 1973 (Razum & Wenner, 2016). While many of these guest workers returned to their home countries, those who

stayed usually had their families follow under a regulation allowing family reunification during the late 1970s and early 80s after recruitment ended (Razum & Wenner, 2016). In recent years, the majority of migrants to Germany have originated from European countries, particularly other European Union member states (Hanewinkel & Oltmer, 2018). However, 2015 marked an exception, as a significant proportion of migrants arrived from outside Europe, particularly from Syria, which became the leading country of origin (Hanewinkel & Oltmer, 2018).

However, it is important to note that while Germany has been receiving immigrants for a long time, German integration policy is considered to be exclusionist (Malmusi, 2015). Through this system, integration or migrant naturalization was intended, with the exception of ethnic German resettlers (Razum & Wenner, 2016). Migrants were considered to be temporary guests who would eventually return to their countries of origin (Razum & Wenner, 2016). In 2001, the report of the commission on migration was a turning point in German immigration policy (Razum & Wenner, 2016). This report led to changes in citizenship and naturalization, effectively allowing the acquisition of German nationality for children born to foreign parents with regular status and at least 8 years of residence in Germany (Razum & Wenner, 2016). Following this report, a new immigration act was introduced in 2005 with the aim of improving integration through the use of integration and language courses (Razum & Wenner, 2016).

Immigrants in Germany are required to learn German, and this is promoted by the promise that this language acquisition will allow them to have greater economic opportunities (Heinemann, 2017). These courses operate with the assumption that those with non-European

citizenship do not share the same values as German citizens. The state-subsidized German courses, ‘Integrationskurse’ (integration courses) in Germany, are always supplemented by an ‘Orientation Course’ (Orientierungskurs), which is meant to convey the mainstream values and rules of German society (Heinemann, 2017). This transition from an exclusionist model of immigration to one that promotes integration may influence how migrants experience social and institutional environments, which can, in turn, shape health outcomes over time.

These newfound institutional expectations may also create conditions that reinforce processes associated with both weathering and unhealthful acculturation. Pressure to integrate into dominant cultural norms may encourage migrants to adopt host-country behaviors, lifestyles, and social practices relatively quickly, particularly in areas such as language use, diet, work culture, and social relationships. As a result, exposure to host-country environments and behavioral norms may begin to shape health trajectories not only among first-generation migrants, but also among second-generation individuals who are raised within these systems from birth. At the same time, expectations of integration may coexist with experiences of discrimination, social exclusion, or perceived cultural difference, creating chronic social and psychological stress. The combination of adapting to dominant social norms while simultaneously navigating structural barriers may therefore contribute to the gradual erosion of the initial health advantages often observed among migrant populations.

The Socio-Economic Panel (SOEP)

The Socio-Economic Panel (SOEP) is a large, nationally representative longitudinal household study in Germany that has been conducted annually since 1984. It was originally established in 1983 under the leadership of Hans-Jürgen Krupp as part of a broader research initiative on the microanalytical foundations of social policy (DIW Berlin: The Institutional History of the SOEP, 2018). It was originally developed within the Collaborative Research Center "Microanalytic Foundations of Social Policy" and has since become part of Germany's research infrastructure under the Leibniz Association (Goebel et al., 2019). The study is currently based at the German Institute for Economic Research (DIW Berlin).

Over time, the SOEP has expanded in both scope and population coverage. Notably, it incorporated respondents from East Germany following the fall of the Berlin Wall and later included samples of migrants and ethnic German repatriates (DIW Berlin: The Institutional History of the SOEP, 2018). This enables researchers to examine long-term societal changes, life course dynamics, and the relationships between early life conditions and later outcomes. In addition, the inclusion of migrant samples and respondents from both East and West Germany enhances its suitability for analyzing convergence and inequality across population groups.

The survey includes approximately 15,000 households and 30,000 individuals and collects a broad range of information on both objective and subjective aspects of well-being (Goebel et al., 2019). Objective measures include variables such as income, education, employment status, and demographic characteristics, while subjective measures capture dimensions such as life satisfaction, perceptions of fairness, and psychological traits. Although

the SOEP does not include detailed clinical or diagnostic health data, aside from measures such as grip strength, it relies on self-reported and proxy-reported indicators of health (Schupp, 2012).

The SOEP interview methodology relies on a set of pre-tested questionnaires administered at both the household and individual levels. Interviewers aim to conduct face-to-face interviews with all household members aged 16 and older, and proxy interviews are not used for adult respondents. In addition, a designated household member, typically the head of the household, completes a separate household questionnaire that collects information on housing conditions, housing costs, and sources of income, including social transfers such as assistance or housing allowances. This questionnaire also gathers information on children under the age of 16, particularly regarding childcare, kindergarten, and school attendance.

This structure allows researchers to examine within-household dynamics, such as income distribution and decision-making processes, as well as individual-level outcomes (Goebel et al., 2019). The SOEP uses a combination of standardized and specialized survey instruments. Core questionnaires are administered annually, ensuring consistency over time, while additional rotating and topical modules expand the range of topics covered (Goebel et al., 2019). These include areas such as wealth, social networks, neighborhood conditions, and family dynamics. Over time, the survey has also expanded to include age-specific questionnaires, such as youth surveys and instruments for children at different developmental stages, as well as parental reports on younger children.

The Present Study

Building on prior research highlighting disparities in both physical and mental health among migrant and non-migrant populations, as well as evidence from the healthy immigrant effect and weathering hypothesis, the present study examines how migration background shapes multiple dimensions of health and well-being over time. While existing literature has often focused on specific outcomes or first-generation migrants, less attention has been given to how subjective and functional health measures vary across migration backgrounds using longitudinal data. Using German Socio-Economic Panel (SOEP) data, this study investigates whether migration background is associated with differences in how life satisfaction, health satisfaction, physical limitations, and emotional health change over time.

Specifically, the following hypotheses are examined:

1. Migration background is associated with differences in life satisfaction, such that individuals with a migration background will report lower baseline life satisfaction and experience steeper declines over time compared to native-born Germans.
2. Migration background is associated with differences in health satisfaction, with individuals with a migration background reporting lower baseline health satisfaction and in declines over time relative to native-born Germans.
3. Migration background is associated with higher levels of health-related limitations, including both general and physical limitations, as well as steeper increases in these limitations over time compared to native-born Germans.

4. Migration background is associated with higher levels of emotional health limitations and greater worsening of emotional health outcomes over time.

Table 1

1a) Participant Demographics and 1b) Migration Background

Participant Demographics	
	N = 99433[1]
Age	39 (17)
Unknown	4
Sex	
Male	49,438 (50%)
Female	49,987 (50%)
Unknown	8
Birth Country	
Germany	70,417 (71%)
Other	17,665 (18%)
Syria	5,000 (5.0%)
Turkey	2,665 (2.7%)
Poland	1,985 (2.0%)
Russian Federation	1,701 (1.7%)
[1] Mean (SD); n (%)	

Participant Demographics	
	N = 99433[1]
Migration Background	
No Migration Background	64,954 (65%)
Direct Migration Background	29,016 (29%)
Indirect Migration Background	5,463 (5.5%)
[1] n (%)	

Note. Data on participants' race and ethnicity was not available in the dataset used for this analysis.

Methods

Participants

This analysis uses data collected from 99,433 participants between 1984 and 2021. These participants were born between 1882 and 2004. At the time of their first survey participation (i.e., first observed wave), respondents ranged in age from 14 to 102 years, with a mean age of

39.11 (SD = 17; Table 1a). 50% of participants were female and 50% of participants were male (Table 1a). The majority of the participants were born in Germany (71%), while 18% were born in other countries, a notable number of participants were born in Syria (5.0%), Turkey (2.7%), Poland (2.0%), and Russia (1.7%; Table 1a). These participants completed an average of 7.29 (SD = 7.72) annual surveys. Over the course of the study, a total of 6,641 participants died, with a mean age at death of 73.51 (SD = 14.1; range: 18 – 104).

Measures

Migration Background

Migration background was categorized using the SOEP variable *migback*, a composite variable designed to capture both first-generation and subsequent-generation migration status. The construction of this variable draws on multiple sources of information within the SOEP, including respondents' country of birth, parental country of birth, citizenship, and additional biographical data (SOEP Group, 2025). Respondents were categorized into three groups: no migration background, direct migration background, and indirect migration background (SOEP Group, 2025).

Individuals born outside of Germany were classified as having a direct migration background. Individuals born in Germany were further distinguished based on parental information: those whose parents were also born in Germany were classified as having no migration background, while those with at least one parent who had migrated were classified as having an indirect migration background. In cases where parental information was missing or

incomplete, the SOEP used additional indicators, such as grandparental background or citizenship, to infer migration status where possible (SOEP Group, 2025). In the sample analyzed here, 65% of participants were classified as no migration background, 29% as direct migration background, and 5.5% as indirect migration background (Table 1b).

Since this variable is constructed using information collected across multiple waves, it is subject to retrospective updates. That is, respondents' classification may change over time as new or more complete information becomes available (e.g., if citizenship or parental background is reported in later waves; SOEP Group, 2025). While this approach enhances the accuracy and completeness of the measure, it may also introduce minor inconsistencies across waves. Additionally, the variable may underestimate the number of individuals with an indirect migration background in cases where parental or grandparental information is missing and no alternative indicators are available (SOEP Group, 2025).

Life Satisfaction

Life satisfaction was included as an outcome measure because it captures individuals' overall evaluation of their well-being and provides a broad indicator of subjective quality of life. It is particularly relevant to the hypotheses, as it allows for the examination of how migration background is associated with long-term trajectories of well-being beyond specific physical or mental health outcomes.

General life satisfaction has been measured as part of the individual questionnaire in the SOEP on an annual basis since 1984. Participants were asked, "Wie zufrieden sind Sie

gegenwärtig, alles in allem, mit Ihrem Leben?", which translates to English as "How satisfied are you with your life, all things considered?" Responses were recorded on a scale ranging from 0 (completely dissatisfied / ganz und gar unzufrieden) to 10 (completely satisfied/ ganz und gar zufrieden). Since participants are asked to reflect on their lives as a whole, these answers serve as an indicator of life satisfaction. In the sample analyzed here, the average life satisfaction score was 7.16 (SD = 1.68; Table A1).

Satisfaction with Health

Health satisfaction was included as an outcome measure because it captures individuals' subjective perceptions of their physical well-being and is another satisfaction domain assessed annually in the SOEP, alongside other aspects of satisfaction. Health satisfaction has been measured as part of the individual questionnaire in the SOEP on an annual basis since 1984.

Participants were asked, "Wie zufrieden sind Sie… mit Ihrer Gesundheit?" ("How satisfied are you with... your health?"). Responses were recorded on a scale from 0 (completely dissatisfied/ ganz und gar unzufrieden) to 10 (completely satisfied/ ganz und gar zufrieden). In the sample analyzed here the average satisfaction with health score was 6.77 (SD = 2.27; Table A1).

Limitations Due to Health

Limitations due to health were included as a measure because they capture the extent to which health interferes with daily functioning, providing a more concrete indicator of health

status beyond subjective perceptions. This variable was assessed from 2011 to 2013, and 2015 to 2021 as part of the individual questionnaires.

General health limitations were assessed using the item: “Sind Sie durch ein gesundheitliches Problem bei Tätigkeiten des normalen Alltagslebens eingeschränkt?” (Do you have a health problem that restricts you from performing normal, everyday activities?) Responses were recorded on a three-point scale indicating the extent to which their health limits their daily activities, 1 (“yes, seriously restricts me” / ja, stark eingeschränkt), 2 (“yes, restricts me to some extent”/ ja, etwas eingeschränkt), and 3 (“no, does not restrict me at all”/ nein, nicht eingeschränkt). Notably, for this measure, lower scores indicate higher activity limitations due to health problems with a mean value of 2.6 (SD = 0.64; Table A1).

Limitations Due to Physical Problems

To better understand how migration background relates to health, multiple dimensions of health were examined rather than relying on a single measure. Physical limitations due to health were included to capture functional health, allowing for an assessment of how migration background may influence individuals’ ability to carry out daily activities. This measure was recorded biennial during the individual questionnaire from 2002 to 2014 annually from 2016 to 2018, and in 2020.

Physical health limitations assessed as using the item: “Bitte denken Sie einmal an die letzten vier Wochen." Wie oft kam es in dieser Zeit vor, ... dass Sie wegen gesundheitlicher Probleme körperlicher Art in Ihrer Arbeit oder Ihren alltäglichen Beschäftigungen in der Art

Ihrer Tätigkeiten eingeschränkt waren?" (During the last four weeks… how often did you feel limited in some way at work or in everyday activities because of physical health problems?) Participants responded using a five-point scale with options including 1 ("always" / immer), 2 ("often" / oft), 3 ("sometimes"/ manchmal), 4 ("hardly ever"/ fast nie), and 5 ("never" / nie). Across the study period, the mean score for this measure was 3.99 (SD = 1.13; Table A1).

Limitations Due to Emotional Problems

Emotional limitations were included to capture how mental health affects daily functioning. This measure reflects the extent to which emotional or psychological factors interfere with everyday activities. It is particularly relevant in this context, as migration-related stressors and social experiences may influence mental health differently across groups. This measure was recorded during the individual questionnaire biennially from 2002 to 2014, and annually from 2016 to 2021.

Emotional health limitations were measured using the item: "Bitte denken Sie einmal an die letzten vier Wochen. Wie oft kam es in dieser Zeit vor... dass Sie wegen seelischer oder emotionaler Probleme in Ihrer Arbeit oder Ihren alltäglichen Beschäftigungen weniger geschafft haben als Sie eigentlich wollten?" (During the last four weeks, how often did you feel that you achieved less than you wanted to at work or in everyday activities due to mental health or emotional problems?) Participants responded using a five-point scale: 1 ("always" / immer), 2 ("often" / oft), 3 ("sometimes"/ manchmal), 4 ("hardly ever"/ fast nie), and 5 ("never" / nie). Across the study period, the mean score for this measure was 4.24 (SD = 0.99; Table A1).

Time-in-Study

For each participant per survey wave, time-in-study was calculated as the number of years since a participant first entered the study. On average, participants spent 7.36 (SD = 7.43; range: <1-37; Table A1) years in the SOEP.

Data Analysis

Linear growth models (Grimm et al., 2016) were estimated to examine trajectories of life satisfaction, health satisfaction, and self-reported health limitations across the 37-year study period. Migration background was included as a key predictor to assess whether baseline levels of well-being and health outcomes, as well as rates of change over time, differed among individuals with no migration background, a direct migration background, or an indirect migration background. This modeling approach allows for the estimation of both within-person changes over time and between-person differences in baseline levels and rates of change, accommodating the nested structure of the data, with repeated yearly observations nested within individuals in a multilevel framework.

The Level 1 model of within-person change over time is specified as:

$$y_{it} = b_{1i} + b_{2i} TimeInStudy_{it} + u_{it}$$

where y_{it} is the outcome for individual *i* at time *t*, $TimeInStudy_{it}$ is modeled as a function of person-specific intercepts, b_{1i}, that indicate expected level of the outcome at baseline,

person-specific slopes, b_{2i}, that indicate rate of change across time, and u_{it} are occasion-specific residuals not captured by the linear trajectory.

At Level 2, person-level baseline levels and rates of change over time are modeled as a function of migration background. The Level 2 model is specified as:

$$b_{1i} = \beta_{01} + \beta_{11} MigrationDirect_i + \beta_{21} MigrationIndirect_i + d_{1i}$$

and

$$b_{2i} = \beta_{02} + \beta_{12} MigrationDirect_i + \beta_{22} MigrationIndirect_i + d_{2i}$$

In this specification, $MigrationDirect_i$ represents whether an individual has a direct migration background (compared to no migration background), and $MigrationIndirect_i$ represents whether an individual has an indirect migration background (compared to no migration background). The terms d_{1i} and d_{2i} are residual individual-level differences in baseline levels and rates of change that are not explained by differences in migration background.The same model specification was fit separately for each outcome variable of interest: life satisfaction, health satisfaction, general health limitations, physical health limitations, and emotional health limitations.

All data management and analysis were conducted using R in RStudio. In particular, the following packages were used: tidyverse (Wickham et al., 2019), psych (Revelle, 2024), lmerTest (Kuznetsova et al., 2017), lme4 (Bates et al., 2015), nlme (Pinheiro et al., 2000), and ggplot2 (Wickham, 2016). Statistical significance was evaluated at $p < 0.05$.

Results

Life Satisfaction

To examine trajectories of life satisfaction over time, life satisfaction was modeled as a function of migration background and time-in-study. Parameter estimates and fit statistics for the model of Life Satisfaction are presented in Table 2. Baseline life satisfaction for non-migrants at the start of the survey was 7.33 ($p < 0.001$). In line with hypotheses, for those with a direct migration background, this value was 0.092 points higher ($p < 0.001$) in comparison to participants without a migration background. Furthermore, those with an indirect migration background were 0.262 points higher ($p < 0.001$) than those without a migration background.

In line with hypotheses, all migration backgrounds experience a decline in life satisfaction over time. For individuals with no migration background, life satisfaction decreased by -0.026 ($p < 0.001$) for each year of participation in the survey. Individuals with a direct migration background experienced a steeper decline in life satisfaction by an additional -0.0081 ($p < 0.001$) points per year. Similarly, individuals with an indirect migration background experience a -0.01 ($p = 0.0014$) steeper decline in life satisfaction .

Table 2
Linear Growth Model of Life Satisfaction Over Survey Duration by Migration Background

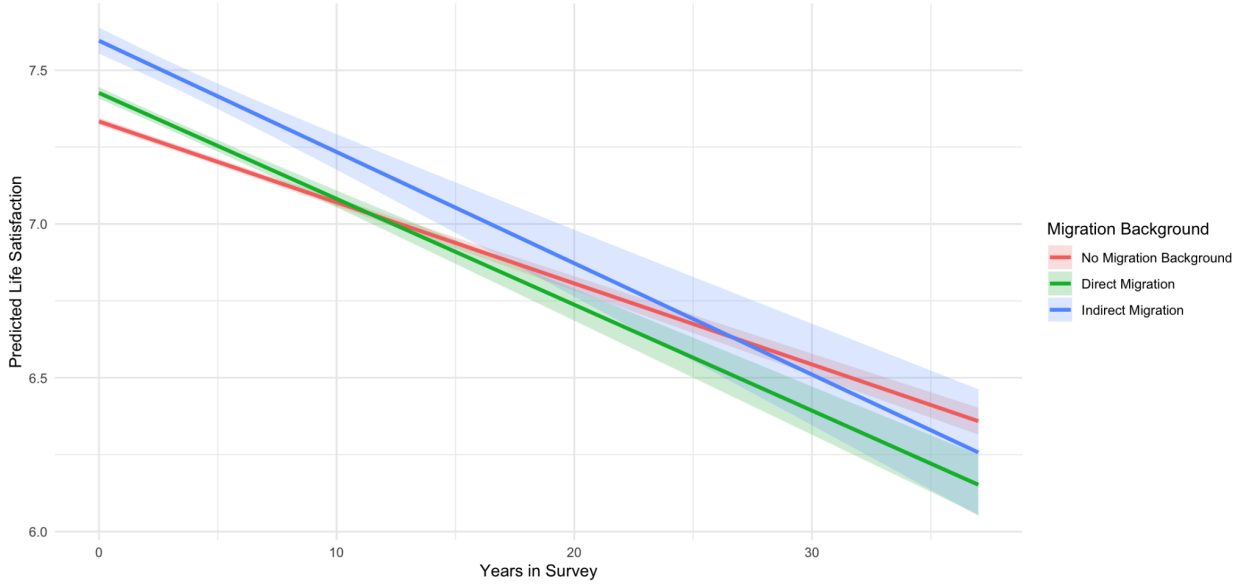
Parameter	Estimate	SE
Fixed effects estimates		
Intercept	**7.33** ***	0.0060
Direct Migration Background	**0.092** ***	0.011
Indirect Migration Background	**0.26** ***	0.022
Time -in-Study	**-0.026** ***	0.00064
Direct Migration Background x Time-in-Study	**-0.0081** ***	0.0016
Indirect Migration Background x Time-in-Study	**-0.01** **	0.0031
Random Effects		
Variance Intercept	1.74	
Variance: Time-in-Study	0.0066	
Corr (Intercept, Time-in-Study)	-0.39	
Residual	1.64	

Note. Unstandardized estimates and standard errors are presented. ** $p < .01$, *** $p < .001$.

These results, as shown in Figure 3, indicate that although individuals with migration backgrounds reported higher life satisfaction at baseline, their life satisfaction is predicted to decline more rapidly over time compared to individuals with no migration background.

Figure 2

Predicted Life Satisfaction over Survey Duration by Migration Background



Note: Life satisfaction predictions were generated from the fitted linear mixed-effects model and applied to a constructed dataset spanning 0 to 37 years in the survey. Shaded areas represent 95% confidence intervals around the predicted values.

Health Satisfaction

To examine trajectories of health satisfaction over time, health satisfaction was modeled as a function of migration background and time-in-study. Parameter estimates and fit statistics for this model are presented in Table 3. Contrary to the hypotheses, baseline health satisfaction was higher among individuals with migration backgrounds compared to non-migrants. Baseline health satisfaction for non-migrants at the start of the survey was 7.10 ($p < 0.001$). For those with a direct migration background, this value was 0.44 points higher ($p < 0.001$) compared to participants without a migration background. Furthermore, individuals with an indirect migration

background had baseline health satisfaction that was 0.83 ($p < 0.001$) points higher than those without a migration background.

Consistent with the hypotheses, all migration backgrounds experience a decline in health satisfaction over time. For individuals with no migration background, health satisfaction is expected to decrease by 0.066 ($p < 0.001$) for each year of participation in the survey. Individuals with a direct migration background are expected to experience a steeper decline in health satisfaction by an additional 0.020 ($p < 0.001$) points per year. Similarly, individuals with an indirect migration background are expected to experience a steeper decline in health satisfaction by an additional 0.01 ($p = 0.015$).

Table 3
Linear Growth Model of Health Satisfaction Over Survey Duration by Migration Background

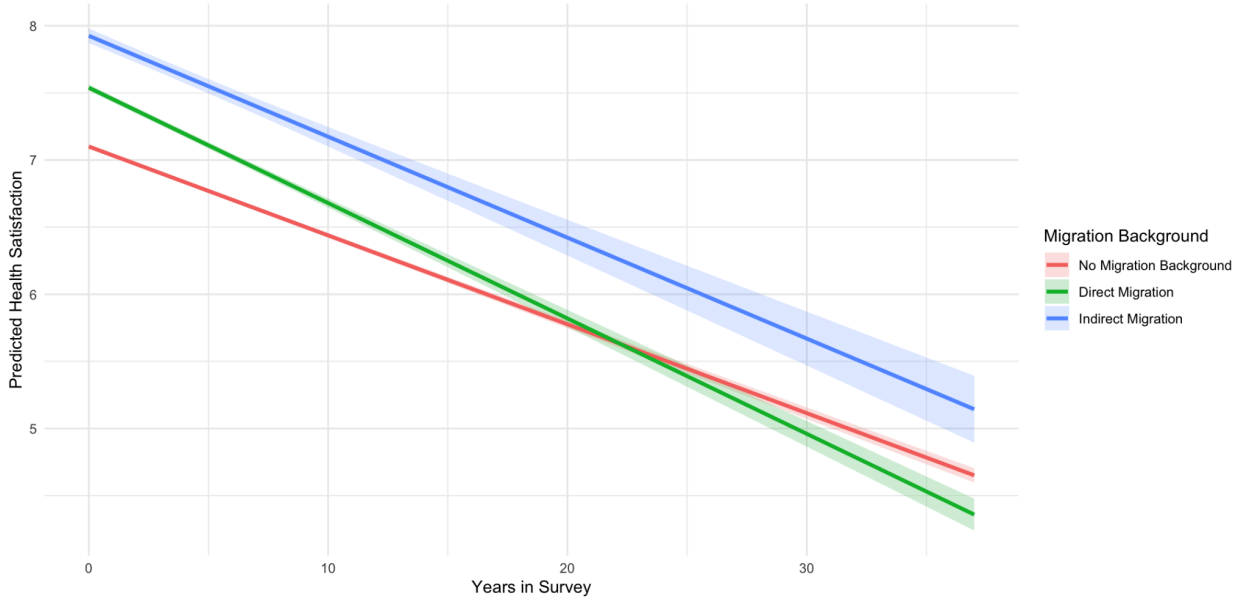
Parameter	Estimate	SE
Fixed effects estimates		
Intercept	**7.10** ***	0.0077
Direct Migration Background	**0.44****	0.014
Indirect Migration Background	**0.83** ***	0.028
Time-in-Study	**-0.066** ***	0.00078
Direct Migration Background x Time-in-Study	**-0.020** ***	0.0019
Indirect Migration Background x Time-in-Study	**-0.01** *	0.0037
Random Effects		
Variance Intercept	2.96	
Variance: Time-in-Study	0.0097	
Corr (Intercept, Time-in-Study)	-0.34	
Residual	2.24	

Note. Unstandardized estimates and standard errors are presented. * $p < .05$, *** $p < .001$.

These results, as shown in Figure 4, indicate that although individuals with migration backgrounds reported higher health satisfaction at baseline, their health satisfaction is predicted to decline more rapidly over time compared to individuals with no migration background.

Figure 3

Predicted Health Satisfaction over Survey Duration by Migration Background



Note: Health satisfaction predictions were generated from the fitted linear mixed-effects model and applied to a constructed dataset spanning 0 to 37 years in the survey. Shaded areas represent 95% confidence intervals around the predicted values.

Limitations Due to Health

To examine trajectories of limitations due to health over time, daily health limitations were modeled as a function of migration background and time-in-study. Parameter estimates and fit statistics for this model are presented in Table 4. Since the scale is coded such that 3 indicates "not at all limited" and 1 indicates "extremely limited," decreases in the outcome reflect worsening health.

Table 4
Linear Growth Model of Limitations in Daily Life due to Health Over Survey Duration by Migration Background

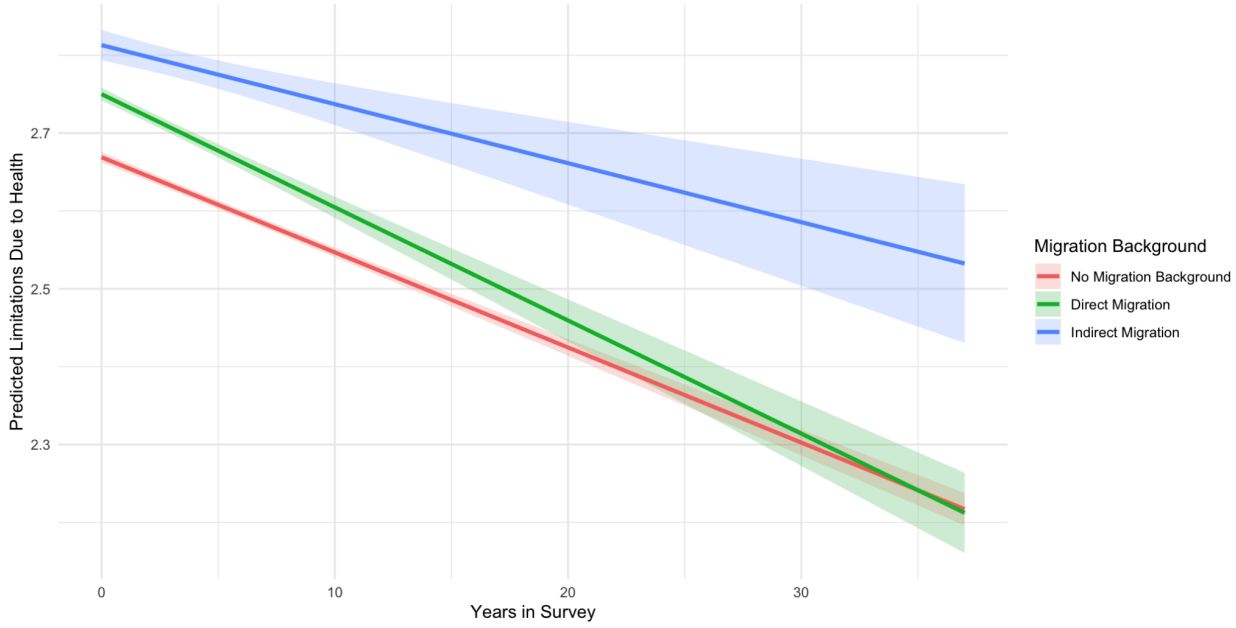
Parameter	Estimate	SE
Fixed effects estimates		
Intercept	**2.67** ***	0.0034
Direct Migration Background	**0.081** ***	0.0053
Indirect Migration Background	**0.14** ***	0.010
Time-in-Study	**-0.012** ***	0.00033
Direct Migration Background x Time-in-Study	**-0.0023** **	0.00082
Indirect Migration Background x Time-in-Study	**0.0046** **	0.0015
Random Effects		
Variance Intercept	0.18	
Variance: Time-in-Study	0.00043	
Corr (Intercept, Time-in-Study)	-0.04	
Residual	0.17	

Note. Unstandardized estimates and standard errors are presented. ** $p < .01$, *** $p < .001$.

Partially consistent with the hypotheses, individuals across migration backgrounds experienced increasing health limitations over time, although the rate of change differed by migration background. For individuals with no migration background, health limitations increased gradually over time, with scores declining by 0.012 points per year ($p < .001$). Individuals with a direct migration background experienced a steeper decline in health, with an additional decrease of 0.0023 points per year ($p = 0.0045$). However, contrary to expectations, individuals with an indirect migration background showed a slower decline in health limitations compared to individuals with no migration background by 0.0046 points ($p = 0.0027$).

Figure 4

Predicted Daily Health Limitations over Survey Duration by Migration Background



Note: Predicted health limitations were generated from the fitted linear mixed-effects model and applied to a constructed dataset spanning 0 to 37 years in the survey. Shaded areas represent 95% confidence intervals around the predicted values.

As shown in Figure 5, although individuals with migration backgrounds begin with fewer health limitations, their trajectories diverge over time, with direct migrants experiencing faster declines and indirect migrants showing a slower rate of deterioration compared to non-migrants.

Table 5
Linear Growth Model of Limitations in Daily Life due to Physical Problems Over Survey Duration by Migration Background

Parameter	Estimate	SE
Fixed effects estimates		
Intercept	**4.17** ***	0.0052
Direct Migration Background	**0.043** ***	0.0091
Indirect Migration Background	**0.25****	0.019
Time-in-Study	**-0.026** ***	0.00049
Direct Migration Background x Time-in-Study	-0.0017	0.0012
Indirect Migration Background x Time-in-Study	**-0.0053** *	0.0025
Random Effects		
Variance Intercept	0.59	
Variance: Time-in-Study	0.0014	
Corr (Intercept, Time-in-Study)	-0.023	
Residual	0.63	

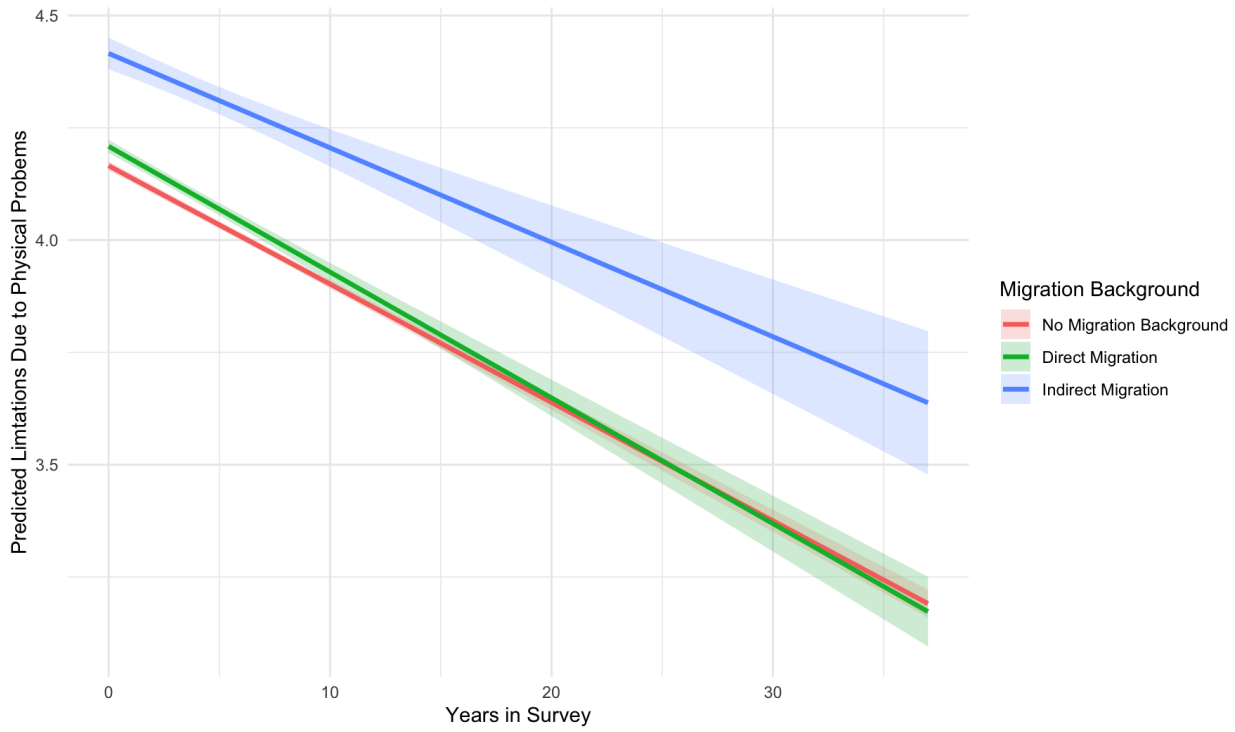
Note. Unstandardized estimates and standard errors are presented. * $p < .05$, *** $p < .001$.

Limitations Due to Physical Problems

To examine trajectories of limitations due to physical problems over time, physical limitations were modeled as a function of migration background and time-in-study. Parameter estimates and fit statistics for this model are presented in Table 5. Since the scale is coded such that 5 indicates “never” and 1 indicates “always,” decreases in the outcome reflect worsening health.

Figure 5

Predicted Daily Physical Limitations over Survey Duration by Migration Background



Note: Predicted physical limitations were generated from the fitted linear mixed-effects model and applied to a constructed dataset spanning 0 to 37 years in the survey. Shaded areas represent 95% confidence intervals around the predicted values.

Consistent with the hypotheses, baseline physical limitations for non-migrants at the start of the survey was 4.17 ($p < 0.001$). For those with a direct migration background, this value was 0.043 points higher ($p < 0.001$) in comparison to participants without a migration background. Furthermore, those with an indirect migration background had baseline physical limitations that were 0.25 ($p < 0.001$) points higher than those without a migration background.

In line with the hypothesized expectations, physical limitations increased over time across all migration backgrounds. For individuals with no migration background, physical limitations increase gradually over time, with scores declining by 0.026 points per year ($p < 0.001$). Individuals with a direct migration background experience a slightly steeper decline in physical health, with an additional decrease of 0.0017 points per year, but this finding was not statistically significant. However, individuals with an indirect migration background also show a steeper decline in physical health in comparison to individuals with no migration background by 0.0053 points ($p = 0.0031$).

As shown in Figure 6, although individuals with migration backgrounds begin with fewer physical limitations, their physical health declines more rapidly over time, particularly for those with an indirect migration background.

Limitations Due to Emotional Problems

Parameter estimates and fit statistics for this model are presented in Table 6. Since the scale is coded such that 5 indicates "never" and 1 indicates "always," decreases in the outcome reflect worsening health. To examine trajectories of limitations due to emotional problems over time, emotional limitations were modeled as a function of migration background and years in the survey.

Partially consistent with the hypotheses, baseline emotional limitations for non-migrants at the start of the survey was 4.35 ($p < 0.001$). For those with a direct migration background, this value was 0.12 points lower ($p < 0.001$) in comparison to participants without a migration background. In contrast, those with an indirect migration background had baseline emotional limitations that were 0.077 ($p < 0.001$) points higher than those without a migration background.

Table 6
Linear Growth Model of Limitations in Daily Life due to Emotional Problems Over Survey Duration by Migration Background

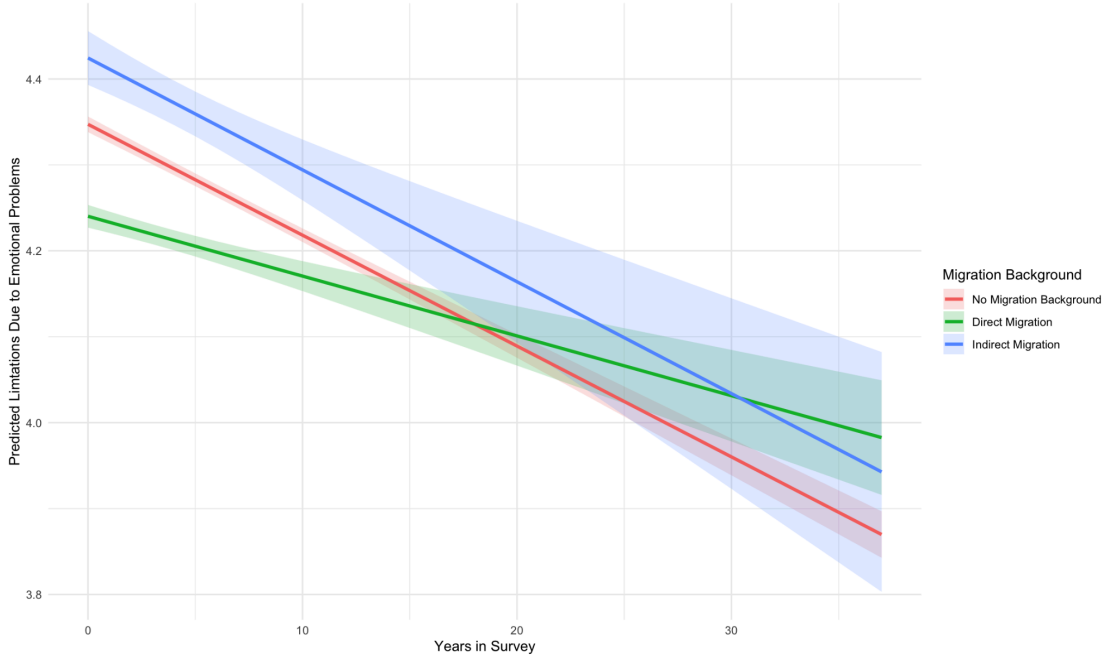
Parameter	Estimate	SE
Fixed effects estimates		
Intercept	**4.35** ***	0.0046
Direct Migration Background	**-0.12** ***	0.0081
Indirect Migration Background	**0.077** ***	0.017
Time-in-Study	**-0.013** ***	0.00043
Direct Migration Background x Time-in-Study	**-0.0059** ***	0.0011
Indirect Migration Background x Time-in-Study	-0.00012	0.0022
Random Effects		
Variance Intercept	0.43	
Variance: Time-in-Study	0.0011	
Corr (Intercept, Time-in-Study)	-0.37	
Residual	0.56	

Note. Unstandardized estimates and standard errors are presented. *** $p < .001$.

Mixed support was found for the hypotheses regarding emotional health limitations over time. For individuals with no migration background, emotional health limitations increase gradually over time, with scores declining by 0.013 points per year ($p < 0.001$). Individuals with a direct migration background experience a slower decline in emotional health, with a smaller decrease in scores over time (an additional +0.0059 points per year; $p < 0.001$), indicating a less steep worsening of emotional health compared to non-migrants.

Figure 6

Predicted Daily Emotional Limitations over Survey Duration by Migration Background



Note: Predicted emotional limitations were generated from the fitted linear mixed-effects model and applied to a constructed dataset spanning 0 to 37 years in the survey. Shaded areas represent 95% confidence intervals around the predicted values.

In contrast, individuals with an indirect migration background show a similar trajectory to those with no migration background, as the difference in their rate of decline is very small (−0.00012 points per year) and not statistically significant. Although the differences between indirect migrants and non-migrants remain relatively small and not statistically significant, as shown in Figure 5, individuals with a direct migration background begin with slightly worse emotional health but experience a slower decline over time compared to non-migrants.

Discussion

This study begins to parse the complex and often competing dynamics that shape migrant health and well-being over time. Migrant populations and their health trajectories are influenced by a combination of initial conditions at arrival, long-term exposure to social and economic environments, and processes of acculturation within the host society (Tuggle et al., 2018). In some cases, migrants may arrive with relative health advantages, yet these advantages do not necessarily remain stable (Elshahat et al., 2021). Over time, cumulative exposure to structural disadvantage, stress, and unequal access to resources can contribute to declining health outcomes (Geronimus, 1992). Simultaneously, the process of adapting to a new cultural and social context can introduce additional risks, as changes in behavior, environment, and social networks reshape everyday life (Lu et al., 2017; Tuggle et al., 2018).

These dynamics are particularly relevant in the German context, where migration has been shaped by a long history of labor recruitment, displacement, and more recent inflows from both within and outside Europe (Razum and Wenner, 2016; Hanewinkel & Oltmer, 2018). The institutional framing of migrants as temporary "guests," combined with limited early efforts toward integration, has contributed to structural conditions that may continue to influence outcomes across generations (Razum and Wenner, 2016). Although more recent policy changes have expanded access to citizenship and introduced integration and language programs, these efforts often emphasize individual adaptation rather than addressing broader structural inequalities (Razum and Wenner, 2016). As a result, differences in health and well-being may

reflect both enduring constraints and gradual shifts associated with generational change. The present study examined whether migration background is associated with differences in how life satisfaction, health satisfaction, and both physical and emotional health limitations changed over time.

Contrary to the first and second hypotheses, individuals with a migration background reported higher baseline levels of both life satisfaction and health satisfaction compared to non-migrants. These findings do not support the expectation that migrants would begin with lower subjective well-being. Instead, they align more closely with the initial advantages described in the healthy immigrant effect, which suggests that migrants may be positively selected on health and well-being at the time of arrival. However, consistent with theoretical expectations, these initial advantages were not sustained over time. Both direct and indirect migrants experienced significantly steeper declines in life satisfaction and health satisfaction compared to non-migrants. This pattern suggests that while migrants may begin with a relative advantage, their well-being deteriorates more rapidly, potentially reflecting processes of cumulative disadvantage, social stress, or barriers to integration.

The third hypothesis, which predicted higher levels of health-related limitations among migrants, was only partially supported. At baseline, migrants reported fewer health-related limitations, indicating better initial functional health. Individuals with a direct migration background experienced a more rapid increase in health limitations over time, suggesting a faster deterioration in physical functioning. This finding is consistent with the weathering hypothesis,

which posits that prolonged exposure to social and structural stressors accelerates health decline among marginalized populations (Geronimus, 1992). In contrast, individuals with an indirect migration background showed a more moderate or, in some cases, slower rate of decline, highlighting important heterogeneity between first- and second-generation migrants.

The fourth hypothesis regarding emotional health limitations yielded similarly partially supported results. Individuals with a direct migration background reported worse emotional health at baseline, indicating greater initial mental health challenges relative to non-migrants. However, their rate of decline over time was slower, suggesting that while they begin at a disadvantage, their trajectories do not worsen as rapidly. In contrast, individuals with an indirect migration background exhibited patterns more similar to non-migrants, with no significant differences in the rate of change over time. These findings suggest that mental health trajectories may be shaped by distinct mechanisms compared to physical health, and that generational status plays a role in shaping emotional well-being.

In summary, the findings provide support for both the healthy immigrant effect and the weathering hypothesis, suggesting that migrant health is shaped by a combination of initial selection and subsequent exposure to structural and social stressors.

Limitations

The study findings should be interpreted in the context of some limitations. The reliance on self-reported measures of health and well-being introduces the possibility of reporting bias. Perceptions of health and well-being are shaped by cultural norms and expectations, and migrant and non-migrant populations may interpret survey questions differently. As a result, differences observed in outcomes such as life satisfaction or health satisfaction may reflect variation in

reporting styles as much as underlying differences in health. While the SOEP does include the option for participants to self - disclosure measures related to health such as illness/disease incidence, it does not include detailed clinical or diagnostic health measures. Aside from limited indicators, such as grip strength, the study mainly relies on subjective and functional measures of health. This restricts the ability to assess objective health conditions and may contribute to discrepancies between perceived and actual health outcomes.

The classification of migration background also presents important limitations. Although the migration background variable incorporates multiple sources of information, it simplifies a highly heterogeneous population. Differences in country of origin, migration pathway, socioeconomic position, and exposure to discrimination are not fully captured, yet are likely to play a significant role in shaping health trajectories. Additionally, this measure does not capture the distinction between the different types of migrants such as refugees and asylum seekers.

Although survey weights are available in the SOEP, they were not incorporated into the current analysis. While the SOEP is designed to be nationally representative, failing to apply these weights may result in certain groups being over- or underrepresented in the sample. As a result, the generalizability of these findings to the broader German population may be limited. Finally, although the growth models capture changes over time, they cannot fully account for unobserved factors that may influence health trajectories, such as access/utilization to healthcare, social integration, or cumulative experiences of marginalization. These factors are central to understanding migrant health and may help explain the observed patterns of decline over time.

Future Directions

The findings of the present study highlight several important avenues for future research. One key direction is the need to more closely examine how health trajectories vary by time since migration. While this study captures changes over time using years in the survey, it does not explicitly distinguish between early and later stages of the migration experience. Future work could incorporate more precise measures of duration since migration to better understand whether health declines occur gradually, accelerate after a certain threshold, or differ across periods of adjustment.

Future studies would also benefit from incorporating more objective health measures alongside self-reported indicators. While subjective measures capture important dimensions of well-being, the inclusion of clinically verified data, or more detailed health diagnostics would allow for a more comprehensive assessment of health and help address potential reporting biases. Combining subjective and objective measures could also provide insight into the extent to which perceived and actual health diverge across migrant and non-migrant populations.

In addition, there is a need for a more detailed examination of heterogeneity within migrant populations. Migration background, as operationalized in this study, captures broad generational differences but does not account for variation by country of origin, migration pathway, or cultural context. Future research should disaggregate migrant groups to examine how health outcomes differ across specific origin countries or regions, reasons for migration, such as labor migration, family reunification, or asylum, as well as demographic and social factors including gender and family or caregiving roles. These distinctions are likely to play a significant role in shaping both initial health advantages and subsequent trajectories.

Another important direction involves the inclusion of structural and institutional factors, particularly access to and utilization of social services. Understanding how migrants engage with healthcare systems, social welfare programs, and other public resources may provide critical insight into the mechanisms driving changes in health over time. Differences in access, eligibility, and utilization of these services may help explain why initial health advantages are not sustained and why certain groups experience more rapid declines.

Future research should also consider how the relationship between migration background and health is moderated by socioeconomic factors, such as income, education, and employment status. These variables may either buffer against or exacerbate declines in health over time and incorporating them into analyses would provide a more nuanced understanding of how inequalities develop across groups. Examining these interactions would also help distinguish the extent to which observed differences are driven by migration background itself versus broader structural inequalities. There should be the continued exploration of the role of broader social and structural processes, including discrimination, social integration, and socioeconomic mobility, in shaping migrant health trajectories.

Conclusion

The findings of this study suggest that disparities in migrant health may develop through a complex and interconnected set of processes that unfold over time. Individuals with migration backgrounds often begin with higher levels of life satisfaction and health satisfaction, but experience steeper declines across the life course, alongside worsening physical and emotional health outcomes. These patterns suggest that health disadvantages may not be fixed at arrival, but

instead shaped by ongoing exposure to social, structural, and environmental conditions over time. While the present study cannot directly identify the causal mechanisms underlying these trajectories, the findings are consistent with theories emphasizing the cumulative effects of acculturative stress, changes in health behaviors, and prolonged exposure to structural challenges such as discrimination and barriers to navigating new social and institutional environments.

These findings also suggest that disparities in migrant health are not necessarily inevitable outcomes of migration itself but may instead reflect broader social and structural conditions that shape health over time. As migration continues to influence increasingly interconnected societies, there is a need not only to document these patterns, but also to better understand the mechanisms through which they emerge and persist. This may require greater attention to equitable healthcare access, policies that reduce structural inequalities, and interventions that address the long-term effects of stress, discrimination, and social exclusion. Without such efforts, early health advantages observed among some migrant groups may continue to diminish over time, potentially contributing to persistent or widening disparities across generations. Ultimately, addressing migrant health disparities is not only important for improving individual outcomes, but also for building more equitable and responsive social and healthcare systems.

References

Abdi, Y. H., Abdi, M. S., Bashir, S. G., Ahmed, N. I., & Abdullahi, Y. B. (2025). Understanding Global Health Inequality and Inequity: Causes, Consequences, and the Path Toward Justice in Healthcare. Public health challenges, 4(4), e70156. https://doi.org/10.1002/puh2.70156

Agyemang, C., & van den Born, B. J. (2019). Non-communicable diseases in migrants: an expert review. Journal of travel medicine, 26(2), tay107. https://doi.org/10.1093/jtm/tay107

Aldridge, R. W., Nellums, L. B., Bartlett, S., Barr, A. L., Patel, P., Burns, R., Hargreaves, S., Miranda, J. J., Tollman, S., Friedland, J. S., & Abubakar, I. (2018). Global patterns of mortality in international migrants: a systematic review and meta-analysis. *Lancet (London, England)*, *392*(10164), 2553–2566.

Aradhya S, Katikireddi SV, Juárez SP. Immigrant ancestry and birthweight across two generations born in Sweden: an intergenerational cohort study. BMJ Global Health. 2022;7:e007341. https://doi.org/10.1136/bmjgh-2021-007341

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauldry, S., & Szaflarski, M. (2017). Immigrant-based Disparities in Mental Health Care Utilization. *Socius: Sociological Research for a Dynamic World*, *3*, 237802311668571. https://doi.org/10.1177/2378023116685718

Bobowik, M., Basabe, N., & Páez, D. (2015). The bright side of migration: hedonic, psychological, and social well-being in immigrants in Spain. Social science research, 51, 189–204. https://doi.org/10.1016/j.ssresearch.2014.09.0119861.

Close, C., Kouvonen, A., Bosqui, T., Patel, K., O'Reilly, D., & Donnelly, M. (2016). The mental health and wellbeing of first generation migrants: a systematic-narrative review of reviews. Globalization and health, 12(1), 47. https://doi.org/10.1186/s12992-016-0187-3

Czaika, M., Reinprecht, C. (2022). Migration Drivers: Why Do People Migrate?. In: Scholten, P. (eds) Introduction to Migration Studies. IMISCOE Research Series. Springer, Cham. https://doi.org/10.1007/978-3-030-92377-8_3

Davide Malmusi, Immigrants' health and health inequality by type of integration policies in European countries, European Journal of Public Health, Volume 25, Issue 2, April 2015, Pages 293–299, https://doi.org/10.1093/eurpub/cku156

David Richter, Julia Rohrer, Maria Metzing, Wiebke Nestler, Michael Weinhardt, Jürgen Schupp. 2017. SOEP Scales Manual (updated for SOEP-Core v32.1). SOEP Survey Papers 423: Series C. Berlin: DIW/SOEP

DIW Berlin: The Institutional History of the SOEP. (2018). DIW Berlin. https://www.diw.de/en/diw_01.c.603717.en/the_institutional_history_of_the_soep.html

Dolores Acevedo-Garcia, Emma V. Sanchez-Vaznaugh, Edna A. Viruell-Fuentes, Joanna Almeida, Integrating social epidemiology into immigrant health research: A cross-national framework, Social Science & Medicine, Volume 75, Issue 12, 2012, Pages 2060-2068, ISSN 0277-9536, https://doi.org/10.1016/j.socscimed.2012.04.040.

Elshahat, S., Moffat, T., & Newbold, K. B. (2022). Understanding the Healthy Immigrant Effect in the Context of Mental Health Challenges: A Systematic Critical Review. Journal of immigrant and minority

Estrada-Moreno, I. S., Palma-Garcia, M. de las O., Gomez Jacinto, L., & Hombrados-Mendieta, M. I. (2025). Resilience in immigrants: a facilitating resource for their social integration. Journal of Ethnic & Cultural Diversity in Social Work, 1–16. https://doi.org/10.1080/15313204.2024.2447274

Evensen, M., Cools, S., & Hermansen, A. S. (2024). Adolescent Health Inequality Across Immigrant Generations. *Journal of Adolescent Health*, *75*(5), 792–800. https://doi.org/10.1016/j.jadohealth.2024.05.017

Ferrara, A., Grindel, C., & Brunori, C. (2024). A longitudinal perspective to migrant health: Unpacking the immigrant health paradox in Germany. Social science & medicine (1982), 351, 116976. https://doi.org/10.1016/j.socscimed.2024.116976

Fox, M., Thayer, Z., & Wadhwa, P. D. (2017). Acculturation and health: the moderating role of socio-cultural context. *American anthropologist*, *119*(3), 405–421. https://doi.org/10.1111/aman.12867

Geronimus, A. T., Andersen, H. F., & Bound, J. (1991). Differences in hypertension prevalence among U.S. black and white women of childbearing age. *Public health reports (Washington, D.C. : 1974)*, *106*(4), 393–399.

Geronimus A. T. (1992). The weathering hypothesis and the health of African-American women and infants: evidence and speculations. Ethnicity & disease, 2(3), 207-221.

Goebel, J., Grabka, M., Liebig, S., Kroh, M., Richter, D., Schröder, C. & Schupp, J. (2019). The German Socio-Economic Panel (SOEP) . Jahrbücher für Nationalökonomie und Statistik, 239(2), 345-360. https://doi.org/10.1515/jbnst-2018-0022

Grimm, K. J., Nilam Ram, & Ryne Estabrook. (2017). *Growth modeling : structural equation and multilevel modeling approaches*. The Guilford Press.

Guy J. Abel, Nikola Sander, ,Quantifying Global International Migration Flows. Science 343, 1520-1522(2014). DOI:10.1126/science.1248676

Ha, J. T., Dewaard, J., Abel, G., Tsuchiya, K., Pinchoff, J., Levesque, C., & Price, K. (2025). The Globalization of International Migration? A Conceptual and Data-Driven Synthesis. *Population and development review*, *51*(1), 421–448. https://doi.org/10.1111/padr.12686

Hanewinkel, V., & Oltmer, J. (2018, January 11). *Historical and Current Development of Migration to and from Germany*. Bpb.de. https://www.bpb.de/themen/migration-integration/regionalprofile/english-version-country-profiles/262758/historical-and-current-development-of-migration-to-and-from-germany/

Heinemann, A. M. B. (2017). The making of 'good citizens': German courses for migrants and refugees. Studies in the Education of Adults, 49(2), 177–195. https://doi.org/10.1080/02660830.2018.1453115

Holz, M. (2021). Health inequalities in Germany: differences in the 'Healthy migrant effect' of European, non-European and internal migrants. Journal of Ethnic and Migration Studies, 48(11), 2620-2641. https://doi.org/10.1080/1369183X.2021.1901675

International Organization for Migration. (n.d.). Key Migration Terms. International Organization for Migration; IOM. https://www.iom.int/key-migration-terms

Jang, S. Y., Oksuzyan, A., Myrskylä, M., van Lenthe, F. J., & Loi, S. (2023). Healthy immigrants, unhealthy ageing? Analysis of health decline among older migrants and natives across European countries. *SSM - population health*, *23*, 101478. https://doi.org/10.1016/j.ssmph.2023.101478

Johnson, J. C., Thaul, S., Page, W. F., Crawford, H., & Institute of Medicine (US) Committee on the CROSSROADS Nuclear Test. (1996). *Mortality of Veteran Participants in the Crossroads Nuclear Test*. National Academies Press (US). https://doi.org/10.17226/5428

Kelsey, J. L. (2008). Observational Epidemiology. Elsevier EBooks, 609–620. https://doi.org/10.1016/b978-012373960-5.00193-3

Kuwert, P., Brähler, E., Glaesmer, H., Freyberger, H. J., & Decker, O. (2009). Impact of forced displacement during World War II on the present-day mental health of the elderly: a population-based study. *International Psychogeriatrics*, *21*(4), 748–753. https://doi.org/10.1017/s1041610209009107

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Loi, S., & Hale, J. M. (2019). Migrant health convergence and the role of material deprivation. Demographic Research, 40, 933–962. https://doi.org/10.4054/demres.2019.40.32

Loi, S., Li, P., & Myrskylä, M. (2025). Unequal weathering: How immigrants' health advantage vanishes over the life-course. *Journal of migration and health*, *11*, 100303. https://doi.org/10.1016/j.jmh.2025.100303

López, S., van Dorp, L., & Hellenthal, G. (2016). Human Dispersal Out of Africa: A Lasting Debate. *Evolutionary bioinformatics online*, *11*(Suppl 2), 57–68. https://doi.org/10.4137/EBO.S33489

Lu, Y., Denier, N., Wang, J. S.-H., & Kaushal, N. (2017). Unhealthy assimilation or persistent health advantage? A longitudinal analysis of immigrant health in the United States. Social Science & Medicine, 195, 105–114. https://doi.org/10.1016/j.socscimed.2017.11.019

Nintcheu, J. M. (2025). Skilled immigration via the Opportunity Card. Nintcheu IW-Kurzbericht, 96. https://www.iwkoeln.de/en/studies/jeannette-michaelle-nintcheu-skilled-immigration-via-the-opportunity-card.html

Nolte, H., Weischer, C., Wilkesmann, U., Maetzel, J., & Tegethoff, H. G. (1997). Kontrolleinstellungen zum Leben und zur Zukunft. Auswertung eines neuen, sozialpsychologischen Itemblocks im Soziooekonomischen Panel. Ruhr-Universität Bochum: Diskussionspapiere aus der Fakultaet für Sozialwissenschaft.

Papademetriou, D., & Sumption, M. (2011). Rethinking Points Systems and Employer- Selected Immigration. https://www.migrationpolicy.org/sites/default/files/publications/rethinkingpointssystem.pdf

Pinheiro, José & Bates, Douglas. (2000). Mixed-Effects Models in S and S-PLUS. 10.1007/b98882.

Pusztai, G., & Márkus, Z. (2019). Paradox of assimilation among indigenous higher education students in four central European countries. *Diaspora, Indigenous, and Minority Education*, *13*(4), 201–216. https://doi.org/10.1080/15595692.2019.1623193

Ramraj, C., Pulver, A., & Siddiqi, A. (2015). Intergenerational transmission of the healthy immigrant effect (HIE) through birth weight: A systematic review and meta-analysis. Social Science & Medicine, 146, 29–40. https://doi.org/10.1016/j.socscimed.2015.10.023

Razum, O., & Wenner, J. (2016). Social and health epidemiology of immigrants in Germany: past, present and future. *Public health reviews*, *37*, 4. https://doi.org/10.1186/s40985-016-0019-2

Rechel, B., Mladovsky, P., Ingleby, D., Mackenbach, J. P., & McKee, M. (2013). Migration and health in an increasingly diverse Europe. The Lancet, 381(9873), 1235–1245. https://doi.org/10.1016/s0140-6736(12)62086-8

Reus-Pons, M., Mulder, C. H., Kibele, E. U. B., & Janssen, F. (2018). Differences in the health transition patterns of migrants and non-migrants aged 50 and older in southern and western Europe (2004-2015). BMC medicine, 16(1), 57. https://doi.org/10.1186/s12916-018-1044-4

Revelle , W. (2024.). psych: Procedures for Psychological, Psychometric, and Personality Research . Northwestern University; https://CRAN.R-project.org/package=psych[Google Scholar]

Rodríguez-Álvarez, E., González-Rábago, Y., Borrell, L. N., & Lanborena, N. (2017). Perceived discrimination and self-rated health in the immigrant population of the Basque Country, Spain. *Gaceta sanitaria*, *31*(5), 390–395. https://doi.org/10.1016/j.gaceta.2016.12.014

Ronellenfitsch, U., & Razum, O. (2004). Deteriorating health satisfaction among immigrants from Eastern Europe to Germany. International Journal for Equity in Health, 3, 4. doi: 10.1186/1475-9276-3-4

Salas-Wright, C. P., Kagotho, N., & Vaughn, M. G. (2014). Mood, anxiety, and personality disorders among first and second-generation immigrants to the United States. Psychiatry research, 220(3), 1028–1036. https://doi.org/10.1016/j.psychres.2014.08.045

Schneiderman, N., Ironson, G., & Siegel, S. D. (2005). Stress and health: psychological, behavioral, and biological determinants. *Annual review of clinical psychology*, *1*, 607–628. https://doi.org/10.1146/annurev.clinpsy.1.102803.144141

Schupp J. Das Sozio-oekonomische Panel (SOEP) [The Socioeconomic Panel (SOEP)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2012 Jun;55(6-7):767-74. German. doi: 10.1007/s00103-012-1496-1. PMID: 22736154.

SOEP Group, 2025. SOEP-Core v40.1 – PPATHL: Person-Related Meta-Dataset . SOEP Survey Papers 1546: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP

Tuggle, A. C., Cohen, J. H., & Crews, D. E. (2018). Stress, migration, and allostatic load: a model based on Mexican migrants in Columbus, Ohio. Journal of physiological anthropology, 37(1), 28. https://doi.org/10.1186/s40101-018-0188-4

UNHCR. (2024). Asylum-seekers. UNHCR. https://www.unhcr.org/about-unhcr/who-we-protect/asylum-seekers

United Nations. (1951, July 28). Convention relating to the Status of Refugees. OHCHR; United Nations. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-relating-status-refugees

United Nations (2024). *International Migrant Stock* 2024: *Key facts and figures*. UN DESA/POP/2024/DC/NO. 13.

Uretsky, M.C., Mathiesen, S.G. The Effects of Years Lived in the United States on the General Health Status of California's Foreign-Born Populations. *J Immigrant Health* 9, 125–136 (2007). https://doi.org/10.1007/s10903-006-9017-7

Wallace, M., Hiam, L., & Aldridge, R. (2023). Elevated mortality among the second-generation (children of migrants) in Europe: what is going wrong? A review. British medical bulletin, 148(1), 5–21. https://doi.org/10.1093/bmb/ldad027

Wallace, Matthew (2024). Living longer in poor health? A narrative review of the evidence among international migrants. Stockholm Research Reports in Demography. Preprint. https://doi.org/10.17045/sthlmuni.27193233.v1

Ware, J., Jr, Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Medical care, 34(3), 220–233. https://doi.org/10.1097/00005650-199603000-00003

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019).

Welcome to the Tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686

Appendix A: Descriptives Statistics

Table A1
Descriptive Statistics for Data Used in the Models

	N	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis
Life Satisfaction	718666	7.16	1.81	8	0	10	10	-0.94	1.18
Health Satisfaction	717459	6.77	2.27	7	0	10	10	-9.74	0.13
Limitations Due to Health	184799	2.6	0.64	3	1	3	2	-1.37	0.63
Limitations Due to Physical problems	233312	3.99	1.13	4	1	5	4	-0.86	-0.24
Limitations Due to Emotional Problems	233826	4.24	0.99	5	1	5	4	-1.15	0.45

Note. This table reflects the data included in the final models; observations with values below 0 were excluded from the analysis.

Table A2
Descriptive Statistics Prior to Data Cleaning

	N	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis
Life Satisfaction	724542	7.07	2.04	8	-5	10	15	-1.77	6.27
Health Satisfaction	724542	6.66	2.5	7	-5	10	15	-1.28	2.83
Limitations Due to Health	724542	-4.92	4.53	-8	-8	3	11	1	-0.84
Limitations Due to Physical problems	724542	-3.74	5.46	-8	-8	5	13	0.71	-1.35
Limitations Due to Emotional Problems	724542	-3.66	5.58	-8	-8	5	13	0.7	-1.38

Note. Descriptive statistics are based on the raw data prior to any data cleaning and may include values below 0, which indicate missing or incomplete responses according to SOEP coding conventions.

Invalid Responses

The SOEP dataset uses standardized negative value codes to denote invalid or non-substantive responses. These include -1 (no answer), -2 (does not apply), -3 (implausible value), -4 (inadmissible multiple response), -5 (not included in this version of the questionnaire), -6 (version of questionnaire with modified filtering), -7 (only available in less restricted edition), -8 (question not part of the survey in a given year), and -9 (missing due to a terminated interview).

Appendix B: Correlations Among Measures and Data Cleaning

Table B1

Correlations between Measures

Variable	1	2	3	4	5
1. Life Satisfaction					
2. Health Satisfaction	0.53				
3. Limitations Due to Health	0.027	-0.011			
4. Limitations Due to Physical problems	0.044	0.050	-0.011		
5. Limitations Due to Emotional Problems	0.045	0.036	-0.015	0.99	

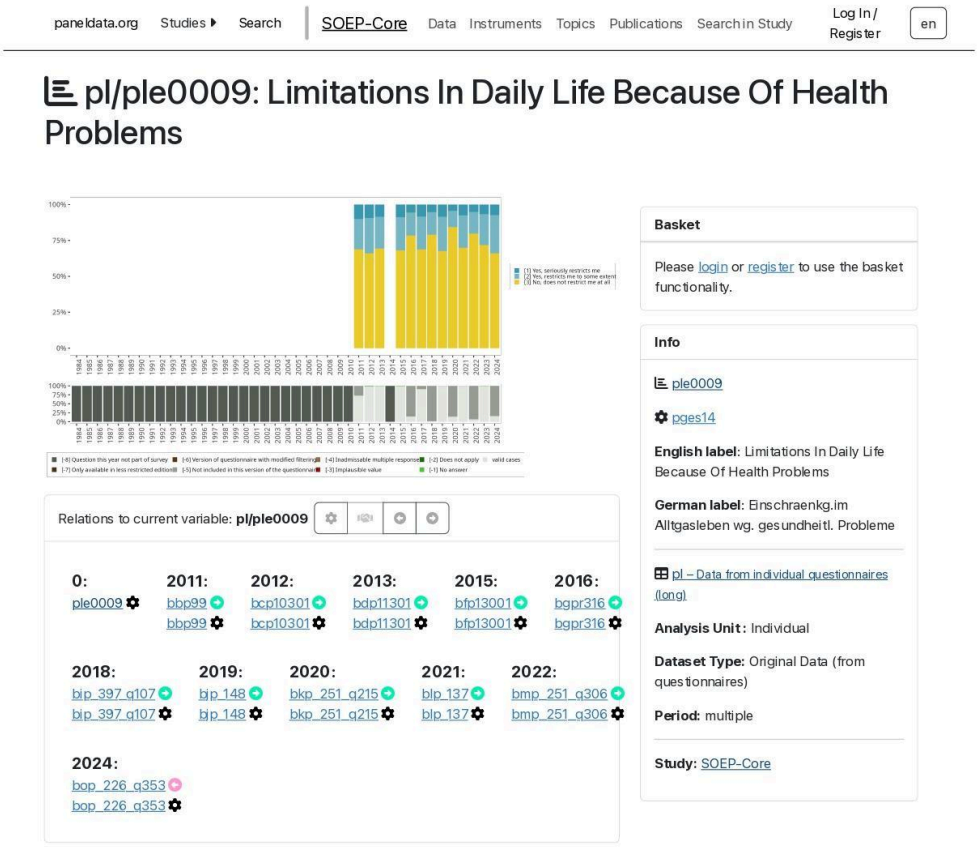
Data Cleaning

Outcome variables were derived from the SOEP individual questionnaire dataset (PL), which contains annual, individual-level responses collected through the core survey instrument (SOEP CORE). This dataset is structured in long format and keyed by a unique person identifier and survey year. Demographic and migration background variables were obtained from the SOEP individual tracking file (PPATHL), which provides information on stable individual characteristics, household identifiers, and response status.

Prior to merging, both datasets were subsetted to include only the variables relevant to the present analysis. The datasets were then merged using participants' unique person identifier and

survey year to create a dataset containing repeated yearly observations for each participant. The resulting dataset was structured in long format. In line with SOEP data conventions, standardized negative value codes indicating invalid or non-substantive responses, were re - coded as missing values for the purposes of analysis.

Appendix C: Survey Measures



Value Labels

	en	de
1	[1] Yes, seriously restricts me	[1] Ja, stark eingeschraenkt
2	[2] Yes, restricts me to some extent	[2] Ja, etwas eingeschraenkt
3	[3] No, does not restrict me at all	[3] Nein, nicht eingeschraenkt
-1	[-1] No answer	[-1] keine Angabe
-2	[-2] Does not apply	[-2] trifft nicht zu
-3	[-3] Implausible value	[-3] unplausibler Wert
-4	[-4] Inadmissable multiple response	[-4] unzulässige Mehrfachantwort
-5	[-5] Not included in this version of the questionnaire	[-5] in Fragebogenversion nicht enthalten
-6	[-6] Version of questionnaire with modified filtering	[-6] Fragebogenversion mit geänderter Filterführung
-7	[-7] Only available in less restricted edition	[-7] nur in weniger eingeschränkter Edition verfügbar
-8	[-8] Question this year not part of survey	[-8] Frage in diesem Jahr nicht Teil des Frageprogramms
-9	[-9] Missing due to a terminated interview	[-9] fehlend aufgrund eines abgebrochenen Interviews

Questions

- Do you have a health problem that restricts you from performing normal, everyday activities? -
- Do you have a health problem that restricts you from performing normal, everyday activities? -
- Do you have a health problem that restricts you from performing normal, everyday activities? -
- Do you have a health problem that limits you in performing normal, everyday activities? -
- Does a health problem hinder you in performing normal, everyday tasks? -

Show all questions

Statistics

Measure	Value
valid	229966
invalid	625100

Codes of related variables with the same concept

This table provides you with an overview of label definitions across related variables to identify changes over time in longitudinal variables. The first number indicates the value code, the second number (in brackets) represents the frequency in the data. Please note that labels are simplified and values with frequency = 0 are hidden.

Search:
Search all columns

Label	Long	2011	2012	2013	2015	2016	2017	2018	2019	2020	2021
	pl	bbp	bcp	bdp	bfp	bgp	bhp	bip	bjp	bkp	blp
	ple0009	bbp99	bcp10301	bdp11301	bfp13001	bgpr316	bhp_161	bip_397_q107	bjp_148	bkp_251_q215	blp_1

Label	Long	2011	2012	2013	2015	2016	2017	2018	2019	2020	2021
	pl	bbp	bcp	bdp	bfp	bgp	bhp	bip	bjp	bkp	blp
	ple0009	bbp99	bcp10301	bdp11301	bfp13001	bgpr316	bhp_161	bip_397_q107	bjp_148	bkp_251_q215	blp_1

Yes, seriously restricts me	1										
Yes, restricts me to some extent	2										
No, does not restrict me at all	3										
Yes, Strongly		1	1	1	1	1	1				

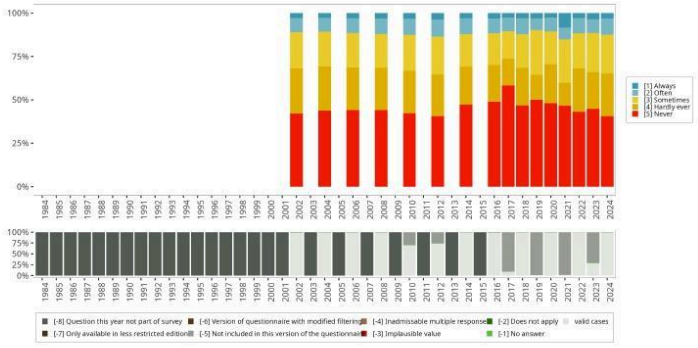
Yes, Strongly Limitated		1	1	1	1	1	1				
Yes, Somewhat Limitated		2	2	2	2	2	2				
No, Not Limitated At All		3	3	3	3	3	3				
Yes, severely											1
Yes, somewhat											2

Contact / feedback

German Socio-economic Panel (SOEP) Imprint

paneldata.org Studies Search | SOEP-Core Data Instruments Topics Publications Search in Study Log In / Register en

pl/ple0032: Limitations Due To Physical Problems



Relations to current variable: **pl/ple0032**

0:	2002:	2004:	2006:	2008:	2010:
ple0032	sp8907	up8607	wp9007	yp10207	bap9007
	sp8907	up8607	wp9007	yp10207	bap9007

2014:	2016:	2017:	2018:	2019:
bep9207	bgp10807	bhp_428_q56	bip_137_07	bjp_427_q155
bep9207	bgp10807	bhp_428_q56	bip_137_07	bjp_427_q155

2021:	2022:	2023:	2024:
blp_427_q289	bmp_141_07	bnp_251_07_q344	bop_157_07
blp_427_q289	bmp_141_07	bnp_251_07_q344	bop_157_07

Basket

Please login or register to use the basket functionality.

Info

ple0032

pges10

English label: Limitations Due To Physical Problems

German label: Inhaltliche Einschränkung wegen körperlicher Probleme

pl – Data from individual questionnaires (long)

Analysis Unit: Individual

Dataset Type: Original Data (from questionnaires)

Period: multiple

Study: SOEP-Core

Questions

- Please think about the last four weeks. How often was it the case in this time ... -
- During the last four weeks, how often did you: -
- During the last four weeks, how often did you: -
- If you think about the last four weeks, how often have you felt... -
- During the last four weeks, how often did you: -

Show all questions

Value Labels

Statistics

	en	de
1	[1] Always	[1] Immer
2	[2] Often	[2] Oft
3	[3] Sometimes	[3] Manchmal
4	[4] Hardly ever	[4] Fast nie
5	[5] Never	[5] Nie
-1	[-1] No answer	[-1] keine Angabe
-2	[-2] Does not apply	[-2] trifft nicht zu
-3	[-3] Implausible value	[-3] unplausibler Wert
-4	[-4] Inadmissable multiple response	[-4] unzulässige Mehrfachantwort
-5	[-5] Not included in this version of the questionnaire	[-5] in Fragebogenversion nicht enthalten
-6	[-6] Version of questionnaire with modified filtering	[-6] Fragebogenversion mit geänderter Filterführung
-7	[-7] Only available in less restricted edition	[-7] nur in weniger eingeschränkter Edition verfügbar
-8	[-8] Question this year not part of survey	[-8] Frage in diesem Jahr nicht Teil des Frageprogramms
-9	[-9] Missing due to a terminated interview	[-9] fehlend aufgrund eines abgebrochenen Interviews

Measure	Value
valid	316691
invalid	538375

Codes of related variables with the same concept

This table provides you with an overview of label definitions across related variables to identify changes over time in longitudinal variables. The first number indicates the value code, the second number (in brackets) represents the frequency in the data. Please note that labels are simplified and values with frequency = 0 are hidden.

Search:

Search all columns

Label	Long	2002	2004	2006	2008	2010	2012	2014	2016	2017	2018	2
	pl	sp	up	wp	yp	bap	bcp	bep	bgp	bhp	bip	b
	ple0032	sp8907	up8607	wp9007	yp10207	bap9007	bcp9407	bep9207	bgp10807	bhp_428_q56	bip_137_07	b

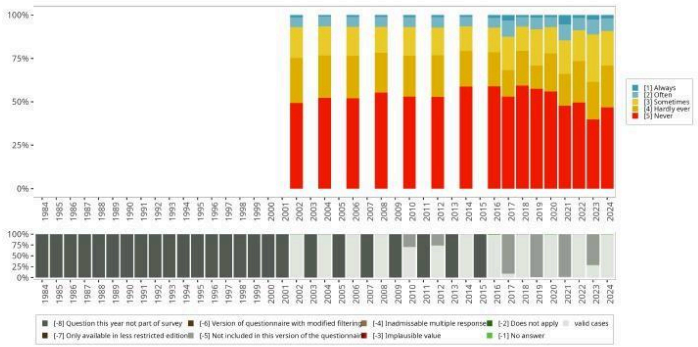
Label	Long	2002	2004	2006	2008	2010	2012	2014	2016	2017	2018	2
	pl	sp	up	wp	yp	bap	bcp	bep	bgp	bhp	bip	b
	ple0032	sp8907	up8607	wp9007	yp10207	bap9007	bcp9407	bep9207	bgp10807	bhp_428_q56	bip_137_07	b
Always	1	1	1	1	1	1	1	1	1	1	1	
Often	2	2	2	2	2	2	2	2	2	2	2	
Sometimes	3	3	3	3	3	3	3	3	3	3	3	
Hardly	4											

ever												
Never	5	5	5	5	5	5	5	5	5	5	5	
Almost Never		4	4	4	4	4	4	4	4	4		
[de] Immer												1
[de] Oft												2
[de] Manchmal												3
[de] Fast nie												4

Contact / feedback

German Socio-economic Panel (SOEP) Imprint

pl/ple0033: Accomplished Less Due To Emotional Problems



Relations to current variable: **pl/ple0033**

0:	2002:	2004:	2006:	2008:	2010:
ple0033	sp8908	up8608	wp9008	yp10208	bap9008
	sp8908	up8608	wp9008	yp10208	bap9008

2014:	2016:	2017:	2018:	2019:
bep9208	bgp10808	bhp_429_q56	bip_137_08	bjp_428_q155
bep9208	bgp10808	bhp_429_q56	bip_137_08	bjp_428_q155

2021:	2022:	2023:	2024:
blp_428_q289	bmp_141_08	bnp_251_08_q344	bop_157_08
blp_428_q289	bmp_141_08	bnp_251_08_q344	bop_157_08

Basket

Please login or register to use the basket functionality.

Info

ple0033

pages11

English label: Accomplished Less Due To Emotional Problems

German label: Weniger geschafft wegen seelischer Probleme

pl – Data from individual questionnaires (long)

Analysis Unit: Individual

Dataset Type: Original Data (from questionnaires)

Period: multiple

Study: SOEP-Core

Questions

- During the last four weeks, how often did you: -
- Please think about the last four weeks. How often was it the case in this time ... -
- During the last four weeks, how often did you: -
- If you think about the last four weeks, how often have you felt... -
- Please think about the last four weeks. How often was it the case in this time ... -

Show all questions

Value Labels

	en	de
1	[1] Always	[1] Immer
2	[2] Often	[2] Oft
3	[3] Sometimes	[3] Manchmal
4	[4] Hardly ever	[4] Fast nie
5	[5] Never	[5] Nie
-1	[-1] No answer	[-1] keine Angabe
-2	[-2] Does not apply	[-2] trifft nicht zu
-3	[-3] Implausible value	[-3] unplausibler Wert
-4	[-4] Inadmissable multiple response	[-4] unzulässige Mehrfachantwort
-5	[-5] Not included in this version of the questionnaire	[-5] in Fragebogenversion nicht enthalten
-6	[-6] Version of questionnaire with modified filtering	[-6] Fragebogenversion mit geänderter Filterführung
-7	[-7] Only available in less restricted edition	[-7] nur in weniger eingeschränkter Edition verfügbar
-8	[-8] Question this year not part of survey	[-8] Frage in diesem Jahr nicht Teil des Frageprogramms
-9	[-9] Missing due to a terminated interview	[-9] fehlend aufgrund eines abgebrochenen Interviews

Statistics

Measure	Value
valid	317201
invalid	537865

Codes of related variables with the same concept

This table provides you with an overview of label definitions across related variables to identify changes over time in longitudinal variables. The first number indicates the value code, the second number (in brackets) represents the frequency in the data. Please note that labels are simplified and values with frequency = 0 are hidden.

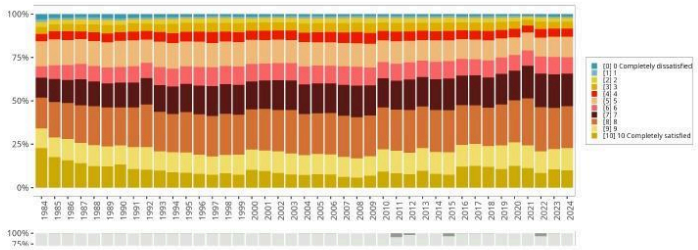
Search:

Search all columns

Label	Long	2002	2004	2006	2008	2010	2012	2014	2016	2017	2018
	pl	sp	up	wp	yp	bap	bcp	bep	bgp	bhp	bip
	ple0033	sp8908	up8608	wp9008	yp10208	bap9008	bcp9408	bep9208	bgp10808	bhp_429_q56	bip_137_08

Label	Long	2002	2004	2006	2008	2010	2012	2014	2016	2017	2018
	pl	sp	up	wp	yp	bap	bcp	bep	bgp	bhp	bip
	ple0033	sp8908	up8608	wp9008	yp10208	bap9008	bcp9408	bep9208	bgp10808	bhp_429_q56	bip_137_08
Always	1	1	1	1	1	1	1	1	1	1	1
Often	2	2	2	2	2	2	2	2	2	2	2
Sometimes	3	3	3	3	3	3	3	3	3	3	3

pl/plh0171: Satisfaction With Health



paneldata.org Studies Search | SOEP-Core Data Instruments Topics Publications Search in Study Log In / Register en

Basket

Please login or register to use the basket functionality.

Info

plh0171

German label: Zufriedenheit Gesundheit

pl – Data from individual questionnaires (long)

Analysis Unit: Individual

Dataset Type: Original Data (from questionnaires)

Period: multiple

Study: SOEP-Core

Relations to current variable: **pl/plh0171**

0:	1984:	1985:	1986:	1987:	1988:	1989
plh0171	ap0301	bp0101	cp0101	dp0101	ep0101	fp0101
	ap0301	bp0101	cp0101	dp0101	ep0101	fp0101

1991:	1992:	1993:	1994:	1995:	1996:	1997:
hp1001	ip9801	jp0101	kp0101	lp0101	mp0101	np0101
hp1001	ip9801	jp0101	kp0101	lp0101	mp0101	np0101

1999:	2000:	2001:	2002:	2003:	2004:	2005:
pp0101	qp0101	rp0101	sp0101	tp0101	up0101	vp0101
pp0101	qp0101	rp0101	sp0101	tp0101	up0101	vp0101

2007:	2008:	2009:	2010:	2011:	2012:	20
xp0101	yp0101	zp0101	bap0101	bbp0101	bcp0101	bdp
xp0101	yp0101	zp0101	bap0101	bbp0101	bcp0101	bdp

2014:	2015:	2016:	2017:	2018:	2019:
bep0101	bfp0101	bgp0101	bhp_01_01	bip_01_01	bjp_01_01
bep0101	bfp0101	bgp0101	bhp_01_01	bip_01_01	bjp_01_01

2021:	2022:	2023:	2024:
blp_04_01	bmp_01_01	bnp_01_01	bop_01_01
blp_04_01	bmp_01_01	bnp_01_01	bop_01_01
		short0023	

Questions

- How satisfied are you currently with the following areas of your life? -
- How satisfied are you currently with the following areas of your life? -
- How satisfied are you with the following areas of your life at the moment: How satisfied are you with your current health? -
- How satisfied are you with the following areas of your life at the moment? How satisfied are you with your current health? -
- How satisfied are you today with the following areas of your life? -

Show all questions

Value Labels

Statistics

	en	de
0	[0] 0 Completely dissatisfied	[0] 0 Zufrieden: Skala 0-Niedrig bis 10-Hoch
1	[1] 1	[1] 1 Zufrieden: Skala 0-Niedrig bis 10-Hoch
2	[2] 2	[2] 2 Zufrieden: Skala 0-Niedrig bis 10-Hoch
3	[3] 3	[3] 3 Zufrieden: Skala 0-Niedrig bis 10-Hoch
4	[4] 4	[4] 4 Zufrieden: Skala 0-Niedrig bis 10-Hoch
5	[5] 5	[5] 5 Zufrieden: Skala 0-Niedrig bis 10-Hoch
6	[6] 6	[6] 6 Zufrieden: Skala 0-Niedrig bis 10-Hoch
7	[7] 7	[7] 7 Zufrieden: Skala 0-Niedrig bis 10-Hoch
8	[8] 8	[8] 8 Zufrieden: Skala 0-Niedrig bis 10-Hoch
9	[9] 9	[9] 9 Zufrieden: Skala 0-Niedrig bis 10-Hoch
10	[10] 10 Completely satisfied	[10] 10 Zufrieden: Skala 0-Niedrig bis 10-Hoch
-1	[-1] No answer	[-1] keine Angabe
-2	[-2] Does not apply	[-2] trifft nicht zu
-3	[-3] Implausible value	[-3] unplausibler Wert
-4	[-4] Inadmissable multiple response	[-4] unzulässige Mehrfachantwort
-5	[-5] Not included in this version of the questionnaire	[-5] in Fragebogenversion nicht enthalten
-6	[-6] Version of questionnaire with modified filtering	[-6] Fragebogenversion mit geänderter Filterführung
-7	[-7] Only available in less restricted edition	[-7] nur in weniger eingeschränkter Edition verfügbar
-8	[-8] Question this year not part of survey	[-8] Frage in diesem Jahr nicht Teil des Frageprogramms
-9	[-9] Missing due to a terminated interview	[-9] fehlend aufgrund eines abgebrochenen Interviews

Measure	Value
valid	845512
invalid	9554

Codes of related variables with the same concept

This table provides you with an overview of label definitions across related variables to identify changes over time in longitudinal variables. The first number indicates the value code, the second number (in brackets) represents the frequency in the data. Please note that labels are simplified and values with frequency = 0 are hidden.

Search:

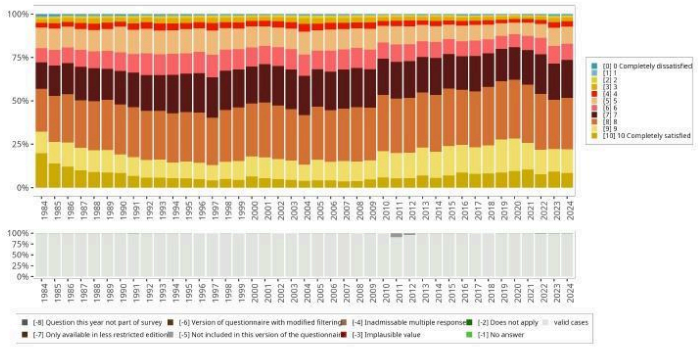
Search all columns

Label	Long	1984	1985	1986	1987	1988	1989	1990	1990	1991	1992	1993	1994	199

	pl	ap	bp	cp	dp	ep	fp	gpost	gp	hp	ip	jp	kp	lp
	plh0171	ap0301	bp0101	cp0101	dp0101	ep0101	fp0101	gp5501e	gp0101	hp1001	ip9801	jp0101	kp0101	lp01
Label	**Long**	**1984**	**1985**	**1986**	**1987**	**1988**	**1989**	**1990**	**1990**	**1991**	**1992**	**1993**	**1994**	**199**
	pl	ap	bp	cp	dp	ep	fp	gpost	gp	hp	ip	jp	kp	lp
	plh0171	ap0301	bp0101	cp0101	dp0101	ep0101	fp0101	gp5501e	gp0101	hp1001	ip9801	jp0101	kp0101	lp01
1	1													
2	2													
3	3													
4	4													
5	5													
6	6													
7	7													
8	8													
9	9													
10 Completely satisfied	10													
1 Satisfied: On Scale		1	1	1	1	1	1	1	1	1	1	1	1	1

Sometimes	3	3	3	3	3	3	3	3	3	3	3
Hardly ever	4										
Never	5	5	5	5	5	5	5	5	5	5	5
Almost Never		4	4	4	4	4	4	4	4	4	
[de] Immer											
[de] Oft											
[de] Manchmal											
[de] Fast nie											

pl/plh0182: Current Life Satisfaction



Basket

Please login or register to use the basket functionality.

Info

plh0182

pzule1

English label: Current Life Satisfaction

German label: Lebenszufriedenheit gegenwaertig

pl – Data from individual questionnaires (long)

Analysis Unit: Individual

Dataset Type: Original Data (from questionnaires)

Period: multiple

Study: SOEP-Core

Relations to current variable: **pl/plh0182**

0:	1984:	1985:	1986:	1987:	1988:	1989:
plh0182	ap6801	bp9301	cp9601	dp9801	ep89	fp108
	ap6801	bp9301	cp9601	dp9801	ep89	fp108

1991:	1992:	1993:	1994:	1995:	1996:	1
hp10901	ip10901	jp10901	kp10401	lp10401	mp11001	n
hp10901	ip10901	jp10901	kp10401	lp10401	mp11001	n

1998:	1999:	2000:	2001:	2002:	2003:	2004:
op12301	pp13501	qp14301	rp13501	sp13501	tp14201	up14501
op12301	pp13501	qp14301	rp13501	sp13501	tp14201	up14501

2005:	2006:	2007:	2008:	2009:	2010:	2011:	2012:
vp154	wp142	xp149	yp15501	zp15701	bap160	bbp15201	bcp151
vp154	wp142	xp149	yp15501	zp15701	bap160	bbp15201	bcp151

2013:	2014:	2015:	2016:	2017:	2018:	2019:
bdp15801	bep151	bfp174	bgp175	bhp_205	bip_202_01	bjp_210
bdp15801	bep151	bfp174	bgp175	bhp_205	bip_202_01	bjp_210

2020:	2021:	2022:	2023:	2024:
bkp_204	blp_205	bmp_234	bnp_214	bop_246
bkp_204	blp_205	bmp_234	bnp_214	bop_246
			short014	

Value Labels

Questions

	en	de
0	[0] 0 Completely dissatisfied	[0] 0 Zufrieden: Skala 0-Niedrig bis 10-Hoch
1	[1] 1	[1] 1 Zufrieden: Skala 0-Niedrig bis 10-Hoch
2	[2] 2	[2] 2 Zufrieden: Skala 0-Niedrig bis 10-Hoch
3	[3] 3	[3] 3 Zufrieden: Skala 0-Niedrig bis 10-Hoch
4	[4] 4	[4] 4 Zufrieden: Skala 0-Niedrig bis 10-Hoch
5	[5] 5	[5] 5 Zufrieden: Skala 0-Niedrig bis 10-Hoch
6	[6] 6	[6] 6 Zufrieden: Skala 0-Niedrig bis 10-Hoch
7	[7] 7	[7] 7 Zufrieden: Skala 0-Niedrig bis 10-Hoch
8	[8] 8	[8] 8 Zufrieden: Skala 0-Niedrig bis 10-Hoch
9	[9] 9	[9] 9 Zufrieden: Skala 0-Niedrig bis 10-Hoch
10	[10] 10 Completely satisfied	[10] 10 Zufrieden: Skala 0-Niedrig bis 10-Hoch
-1	[-1] No answer	[-1] keine Angabe
-2	[-2] Does not apply	[-2] trifft nicht zu
-3	[-3] Implausible value	[-3] unplausibler Wert
-4	[-4] Inadmissable multiple response	[-4] unzulässige Mehrfachantwort
-5	[-5] Not included in this version of the questionnaire	[-5] in Fragebogenversion nicht enthalten
-6	[-6] Version of questionnaire with modified filtering	[-6] Fragebogenversion mit geänderter Filterführung
-7	[-7] Only available in less restricted edition	[-7] nur in weniger eingeschränkter Edition verfügbar
-8	[-8] Question this year not part of survey	[-8] Frage in diesem Jahr nicht Teil des Frageprogramms
-9	[-9] Missing due to a terminated interview	[-9] fehlend aufgrund eines abgebrochenen Interviews

- In conclusion, we would like to ask you about your satisfaction with your life in general. How satisfied are you with your life, all things considered? -
- How satisfied are you with your life at present, all things considered? -
- In conclusion, we would like to ask you about your satisfaction with your life in general. How satisfied are you with your life, all things considered? -
- In conclusion, we would like to ask you about your satisfaction with your life in general. How satisfied are you with your life, all things considered? -
- In conclusion, we would like to ask you about your satisfaction with your life in general. How satisfied are you with your life, all things considered? -

Show all questions

Statistics

Measure	Value
valid	847994
invalid	7072

Codes of related variables with the same concept

This table provides you with an overview of label definitions across related variables to identify changes over time in longitudinal variables. The first number indicates the value code, the second number (in brackets) represents the frequency in the data. Please note that labels are simplified and values with frequency = 0 are hidden.

Search:

Search all columns

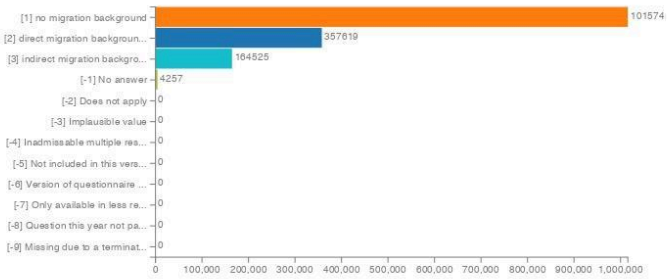
Label	Long	1984	1985	1986	1987	1988	1989	1990	1990	1991	1992	1993	1994	19

	pl	ap	bp	cp	dp	ep	fp	gpost	gp	hp	ip	jp	kp	lp
	plh0182	ap6801	bp9301	cp9601	dp9801	ep89	fp108	gp6401e	gp109	hp10901	ip10901	jp10901	kp10401	lp
Label	**Long**	**1984**	**1985**	**1986**	**1987**	**1988**	**1989**	**1990**	**1990**	**1991**	**1992**	**1993**	**1994**	**19**
	pl	ap	bp	cp	dp	ep	fp	gpost	gp	hp	ip	jp	kp	lp
	plh0182	ap6801	bp9301	cp9601	dp9801	ep89	fp108	gp6401e	gp109	hp10901	ip10901	jp10901	kp10401	lp
1	1													
2	2													
3	3													
4	4													
5	5													
6	6													
7	7													
8	8													
9	9													
10 Completely satisfied	10													
1 Satisfied: On Scale		1	1	1	1	1	1	1	1	1	1	1	1	1

paneldata.org Studies Search | SOEP-Core Data Instruments Topics Publications Search in Study Lo Re

ppathl/migback: Migration background

Percent Hide Missings



Relations to current variable: **ppathl/migback**

0:
migback
migback
migback

Basket

Please login or register to use the basket functionality.

Info

migback

Harmonized from:
migback

migback

English label: Migration background

German label: Migrationshintergrund

ppathl – Individual Tracking File

Analysis Unit : Individual

Dataset Type: Tracking Data

Period: multiple

Attachments:

- Codebook (PDF)

Study: SOEP-Core

Value Labels

Description

	en	de
1	[1] no migration background	[1] kein Migrationshintergrund
2	[2] direct migration background	[2] direkter Migrationshintergrund
3	[3] indirect migration background	[3] indirekter Migrationshintergrund
-1	[-1] No answer	[-1] keine Angabe
-2	[-2] Does not apply	[-2] trifft nicht zu
-3	[-3] Implausible value	[-3] unplausibler Wert
-4	[-4] Inadmissable multiple response	[-4] unzulaessige Mehrfachantwort
-5	[-5] Not included in this version of the questionnaire	[-5] in Fragebogenversion nicht enthalten
-6	[-6] Version of questionnaire with modified filtering	[-6] Fragebogenversion mit geaenderter Filterfuehrung
-7	[-7] Only available in less restricted edition	[-7] nur in weniger eingeschraenkter Edition verfuegbar
-8	[-8] Question this year not part of survey	[-8] Frage in diesem Jahr nicht Teil des Frageprogramms
-9	[-9] Missing due to a terminated interview	[-9] fehlend aufgrund eines abgebrochenen Interviews

MIGBACK contains information on respondents' migration background for all persons who have ever been a part of a SOEP household (i.e., the population from PPFAD, PPATH or PPATHL). In comparison to GERMBORN, the variable MIGBACK is useful to identify immigrants' descendants by combining information on respondents' country of birth (see GERMBORN) and (grand-

Show full description

Statistics

Measure	Value
valid	1537889
invalid	4257

Codes of related variables with the same concept

This table provides you with an overview of label definitions across related variables to identify changes over time in longitudinal variables. The first number indicates the value code, the second number (in brackets) represents the frequency in the data. Please note that labels are simplified and values with frequency = 0 are hidden.

Search:

Search all columns

Label	Long
	ppathl
	migback

Label	Long
	ppathl

	migback
no migration background	1
direct migration background	2
indirect migration background	3
migration background, not differentiable	

Contact / feedback

German Socio-economic Panel (SOEP) Imprint

Appendix D: R Code

Note: The variable years_in_survey was recoded as Time-in-Survey as described in the study

Demographics and Variable Exploration

Set-Up

```
        #Change this to your working directory
setwd("/Users/sophiegabreldar/Library/CloudStorage/Box-Box/Sophie\ Gabreldar\'s\
Files/Gabreldar_HonorsThesis2026/R_analysis")


#Load Necessary Libraries
library(psych)          #for general use functions
library(gt)             # table formatting
library(gtable)         # table formatting
library(gtsummary)      # table summaries
library(kableExtra)     # table formatting
library(knitr)          # html table formatting
library(data.table)
library(tidyverse)   #ggplot2, stringr, dplyr, readr, forcats, tidyr, purr
```

Data Import

```
        people_long
<-fread("/Users/sophiegabreldar/Library/CloudStorage/Box-Box/SOEP/SOEP-CORE.v38.1_inte
rnational_CSV/soepdata/pl.csv")
```

```
people_values <-
read_csv("/Users/sophiegabreldar/Library/CloudStorage/Box-Box/SOEP/SOEP-CORE.v38.1_international_CSV/soepdata/pl_values.csv")

        ## Rows: 102947 Columns: 7
## ── Column specification ──────────────────────────────────────────────────────
─

## Delimiter: ","
## chr (6): study, dataset, version, variable, label, label_de
## dbl (1): value
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.

        demo <-
read_csv("/Users/sophiegabreldar/Library/CloudStorage/Box-Box/SOEP/SOEP-CORE.v38.1_international_CSV/soepdata/ppathl.csv")

        ## Rows: 1219858 Columns: 45
## ── Column specification ──────────────────────────────────────────────────────
─

## Delimiter: ","
```

```
## dbl (45): cid, hid, pid, syear, parid, partner, sex, gebjahr, eintritt, erst...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.

        demo_values <-
read_csv("/Users/sophiegabreldar/Library/CloudStorage/Box-Box/SOEP/SOEP-CORE.v38.1_in
ternational_CSV/soepdata/ppathl_values.csv")

        ## Rows: 678 Columns: 7
## ── Column specification
────────────────────────────────────────────────────────────────────────────────
─
## Delimiter: ","
## chr (6): study, dataset, version, variable, label, label_de
## dbl (1): value
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Subsetting Data

```
        people_sub <- people_long %>%
 select(pid, syear, plj0019,ple0008, ple0009, ple0032, ple0033, plh0171, plh0182, plb0471_h,
```

```
plb0471_v2, plj0019, plj0047)

demo_sub <- demo %>%

 select(pid, syear, corigin, migback, gebjahr,sex, phrf, pbleib, todjahr )
```

Data Cleaning

```
        merged_data <- people_sub %>%

 left_join(demo_sub, by = c("pid", "syear"))

#creating a categorical migrant variable

merged_data <- merged_data %>%

 mutate(

  migrant_binary = case_when(

   migback == 1 ~ 0,

   migback %in% c(2, 3) ~ 1,

   TRUE ~ NA_real_

  )

 )
```

```
        merged_data <- merged_data %>%
 mutate(
  income = case_when(
   syear <= 2001 ~ plb0471_h,
   syear >= 2002 ~ plb0471_v2
  )
 ) #re - label birth country variable
country <- demo_values %>%
 filter(variable == "corigin") %>%
 select(value, label)
merged_data <- merged_data %>%
 left_join(country, by = c("corigin" = "value")) %>%
 rename(birthcountry = label)

        merged_data <- merged_data %>%
 mutate(
  bcountry_clean = case_when(
   birthcountry %in% c("No Answer", "Don't Know", "Inadmissible", "Q.Not.Survey") ~
NA_character_,
   TRUE ~ birthcountry
  ),
  bcountry_clean = factor(bcountry_clean)
```

```
 )
country_unique <- merged_data %>%
 group_by(pid) %>%
 summarise(
  bcountry_clean = first(bcountry_clean),
  .groups = "drop"
 )
```

Demographics

```
        pid_static <- merged_data %>%
 distinct(pid, migback, sex, bcountry_clean)


# Age at first survey
age_first <- merged_data %>%
 filter(gebjahr >= 0) %>%
 group_by(pid) %>%
 summarise(
  first_year = min(syear, na.rm = TRUE),
  birthyear  = first(gebjahr),
  age_at_first = first_year - birthyear,
  .groups = "drop"
```

```
 )
psych::describe(age_first$age_at_first)

        ##    vars     n  mean sd median trimmed   mad min max range skew kurtosis   se
## X1    1 99429 39.11 17     37   37.67 19.27  14 102    88 0.59    -0.42 0.05

        # income
income_pid <- merged_data %>%
 group_by(pid) %>%
 filter(income >= 0) %>%
 summarise(
  income = first(income[!is.na(income)]),
  .groups = "drop"
 )


# combine into one demographics table + recode labels
demo_pid <- pid_static %>%
 left_join(age_first, by = "pid") %>%
 left_join(income_pid, by = "pid") %>%
 mutate(
  migback = factor(
   migback,
   levels = c(1, 2, 3),
```

```
    labels = c("No Migration Background", "Direct Migration Background", "Indirect Migration
Background")
   ),
   migback = fct_explicit_na(migback, "Missing"),
   bcountry_clean = fct_explicit_na(as.factor(bcountry_clean), "Missing")
 )

        ## Warning: There was 1 warning in `mutate()`.
## ℹ In argument: `migback = fct_explicit_na(migback, "Missing")`.
## Caused by warning:
## ! `fct_explicit_na()` was deprecated in forcats 1.0.0.
## ℹ Please use `fct_na_value_to_level()` instead.

        demo_pid <- demo_pid %>%
 mutate(
  sex= factor(
    sex,
    levels = c(1, 2),
    labels = c("Male", "Female")
   )
 )


# 5) Summary table by migration background
```

```r
demo_pid %>%
 select(migback, sex, age_at_first, income, bcountry_clean) %>%
 tbl_summary(
  by = migback,
  label = list(
   sex ~ "Sex",
   age_at_first ~ "Median Age at First Survey",
   income ~ "Median Annual Income",
   bcountry_clean ~ "Birth Country"
  ),
  statistic = list(
   all_continuous() ~ "({median})",
   all_categorical() ~ "{n} ({p}%)"
  ),
  missing = "ifany"
 ) %>%
 modify_header(label ~ "") %>%
 bold_labels() %>%
 as_gt() %>%
 tab_header(title = "Demographics by Migration Background (pid-level)")
```

Table 1: Demographics by Migration Background (pid-level)

	No Migration Background N = 64,954[1]	Direct Migration Background N = 29,016[1]	Indirect Migration Background N = 5,463[1]
Sex			
Male	31,522 (49%)	15,259 (53%)	2,657 (49%)
Female	33,424 (51%)	13,757 (47%)	2,806 (51%)
Unknown	8	0	0
Median Age at First Survey	(40)	(34)	(20)
Unknown	3	1	0
Median Annual Income	(1,796)	(1,534)	(1,200)

Unknown	24,849	14,639	1,689
Birth Country			
Afghanistan	0 (0%)	1,195 (4.1%)	0 (0%)
Albania	0 (0%)	244 (0.8%)	0 (0%)
Algeria	0 (0%)	43 (0.1%)	0 (0%)
American Samoa	0 (0%)	1 (<0.1%)	0 (0%)
Angola	0 (0%)	15 (<0.1%)	0 (0%)
Argentina	0 (0%)	34 (0.1%)	0 (0%)
Armenia	0 (0%)	68 (0.2%)	0 (0%)

Australia	0 (0%)	23 (<0.1%)	0 (0%)
Austria	0 (0%)	225 (0.8%)	0 (0%)
Azerbaijan	0 (0%)	70 (0.2%)	0 (0%)
Bangladesh	0 (0%)	34 (0.1%)	0 (0%)
Belarus	0 (0%)	66 (0.2%)	0 (0%)
Belgium	0 (0%)	35 (0.1%)	0 (0%)
Benin	0 (0%)	3 (<0.1%)	0 (0%)
Bolivia	0 (0%)	7 (<0.1%)	0 (0%)
Bosnia and Herzegovina	0 (0%)	443 (1.5%)	0 (0%)

Botswana	0 (0%)	3 (<0.1%)	0 (0%)
Brazil	0 (0%)	101 (0.3%)	0 (0%)
British Indian Ocean Territory	0 (0%)	1 (<0.1%)	0 (0%)
Bulgaria	0 (0%)	553 (1.9%)	0 (0%)
Burkina Faso	0 (0%)	4 (<0.1%)	0 (0%)
Cambodia	0 (0%)	1 (<0.1%)	0 (0%)
Cameroon	0 (0%)	47 (0.2%)	0 (0%)
Canada	0 (0%)	23 (<0.1%)	0 (0%)
Cape Verde	0 (0%)	1 (<0.1%)	0 (0%)

Chad	0 (0%)	6 (<0.1%)	0 (0%)
Chile	0 (0%)	26 (<0.1%)	0 (0%)
China	0 (0%)	70 (0.2%)	0 (0%)
Colombia	0 (0%)	47 (0.2%)	0 (0%)
Congo	0 (0%)	1 (<0.1%)	0 (0%)
Congo, Democratic Republic	0 (0%)	26 (<0.1%)	0 (0%)
Costa Rica	0 (0%)	31 (0.1%)	0 (0%)
Croatia	0 (0%)	284 (1.0%)	0 (0%)
Cuba	0 (0%)	27 (<0.1%)	0 (0%)

Czech Republic	0 (0%)	212 (0.7%)	0 (0%)
Denmark	0 (0%)	24 (<0.1%)	0 (0%)
Djibouti	0 (0%)	2 (<0.1%)	0 (0%)
Dschibutisch	0 (0%)	12 (<0.1%)	0 (0%)
Ecuador	0 (0%)	14 (<0.1%)	0 (0%)
Egypt	0 (0%)	58 (0.2%)	0 (0%)
El Salvador	0 (0%)	2 (<0.1%)	0 (0%)
Eritrea	0 (0%)	457 (1.6%)	0 (0%)
Estonia	0 (0%)	14 (<0.1%)	0 (0%)

Ethiopia	0 (0%)	52 (0.2%)	0 (0%)
Finland	0 (0%)	18 (<0.1%)	0 (0%)
France	0 (0%)	146 (0.5%)	0 (0%)
Georgia	0 (0%)	63 (0.2%)	0 (0%)
Germany	64,954 (100%)	0 (0%)	5,463 (100%)
Ghana	0 (0%)	65 (0.2%)	0 (0%)
Greece	0 (0%)	758 (2.6%)	0 (0%)
Guinea	0 (0%)	76 (0.3%)	0 (0%)
Haiti	0 (0%)	1 (<0.1%)	0 (0%)
Honduras	0 (0%)	2 (<0.1%)	0 (0%)

Hong Kong	0 (0%)	3 (<0.1%)	0 (0%)
Hungary	0 (0%)	166 (0.6%)	0 (0%)
India	0 (0%)	212 (0.7%)	0 (0%)
Indonesia	0 (0%)	26 (<0.1%)	0 (0%)
Iran	0 (0%)	424 (1.5%)	0 (0%)
Iraq	0 (0%)	1,500 (5.2%)	0 (0%)
Ireland, Republic of	0 (0%)	13 (<0.1%)	0 (0%)
Israel	0 (0%)	16 (<0.1%)	0 (0%)
Italy	0 (0%)	1,123 (3.9%)	0 (0%)

Ivory Coast	0 (0%)	9 (<0.1%)	0 (0%)
Jamaica	0 (0%)	2 (<0.1%)	0 (0%)
Japan	0 (0%)	29 (<0.1%)	0 (0%)
Jordan	0 (0%)	31 (0.1%)	0 (0%)
Kazakhstan	0 (0%)	1,285 (4.4%)	0 (0%)
Kenya	0 (0%)	25 (<0.1%)	0 (0%)
Korea, Republic of	0 (0%)	7 (<0.1%)	0 (0%)
Kosovo	0 (0%)	512 (1.8%)	0 (0%)
Kuwait	0 (0%)	21 (<0.1%)	0 (0%)

Kyrgyzstan	0 (0%)	134 (0.5%)	0 (0%)
Laos	0 (0%)	3 (<0.1%)	0 (0%)
Latvia	0 (0%)	30 (0.1%)	0 (0%)
Lebanon	0 (0%)	168 (0.6%)	0 (0%)
Lesotho	0 (0%)	1 (<0.1%)	0 (0%)
Liberia	0 (0%)	1 (<0.1%)	0 (0%)
Libya	0 (0%)	36 (0.1%)	0 (0%)
Lithuania	0 (0%)	45 (0.2%)	0 (0%)
Luxembourg	0 (0%)	8 (<0.1%)	0 (0%)
Madagascar	0 (0%)	10 (<0.1%)	0 (0%)

Malaysia	0 (0%)	4 (<0.1%)	0 (0%)
Mali	0 (0%)	4 (<0.1%)	0 (0%)
Mauritius	0 (0%)	3 (<0.1%)	0 (0%)
Mexico	0 (0%)	42 (0.1%)	0 (0%)
Moldova	0 (0%)	45 (0.2%)	0 (0%)
Mongolia	0 (0%)	9 (<0.1%)	0 (0%)
Montenegro	0 (0%)	23 (<0.1%)	0 (0%)
Morocco	0 (0%)	174 (0.6%)	0 (0%)
Mozambique	0 (0%)	6 (<0.1%)	0 (0%)
Myanma	0 (0%)	3 (<0.1%)	0 (0%)

Namibia	0 (0%)	9 (<0.1%)	0 (0%)
Nepal	0 (0%)	11 (<0.1%)	0 (0%)
New Zealand	0 (0%)	8 (<0.1%)	0 (0%)
Nicaragua	0 (0%)	2 (<0.1%)	0 (0%)
Niger	0 (0%)	5 (<0.1%)	0 (0%)
Nigeria	0 (0%)	226 (0.8%)	0 (0%)
North Macedonia	0 (0%)	200 (0.7%)	0 (0%)
Norway	0 (0%)	9 (<0.1%)	0 (0%)
Pakistan	0 (0%)	222 (0.8%)	0 (0%)

Palestina	0 (0%)	45 (0.2%)	0 (0%)
Paraguay	0 (0%)	8 (<0.1%)	0 (0%)
Peru	0 (0%)	22 (<0.1%)	0 (0%)
Philippines	0 (0%)	68 (0.2%)	0 (0%)
Poland	0 (0%)	1,985 (6.8%)	0 (0%)
Portugal	0 (0%)	94 (0.3%)	0 (0%)
Qatar	0 (0%)	2 (<0.1%)	0 (0%)
Romania	0 (0%)	1,145 (3.9%)	0 (0%)
Russian Federation	0 (0%)	1,701 (5.9%)	0 (0%)

Rwanda	0 (0%)	3 (<0.1%)	0 (0%)
Samoa	0 (0%)	3 (<0.1%)	0 (0%)
Saudi-Arabia	0 (0%)	26 (<0.1%)	0 (0%)
Senegal	0 (0%)	10 (<0.1%)	0 (0%)
Serbia	0 (0%)	358 (1.2%)	0 (0%)
Sierra Leone	0 (0%)	7 (<0.1%)	0 (0%)
Singapore	0 (0%)	2 (<0.1%)	0 (0%)
Slovakia	0 (0%)	53 (0.2%)	0 (0%)
Slovenia	0 (0%)	54 (0.2%)	0 (0%)
Somalia	0 (0%)	218 (0.8%)	0 (0%)

South Africa	0 (0%)	38 (0.1%)	0 (0%)
Spain	0 (0%)	564 (1.9%)	0 (0%)
Sri Lanka	0 (0%)	58 (0.2%)	0 (0%)
St Kitts and Nevis	0 (0%)	1 (<0.1%)	0 (0%)
Sudan	0 (0%)	33 (0.1%)	0 (0%)
Suriname	0 (0%)	2 (<0.1%)	0 (0%)
Sweden	0 (0%)	28 (<0.1%)	0 (0%)
Switzerland	0 (0%)	79 (0.3%)	0 (0%)
Syria	0 (0%)	5,000 (17%)	0 (0%)

Taiwan	0 (0%)	16 (<0.1%)	0 (0%)
Tajikistan	0 (0%)	36 (0.1%)	0 (0%)
Tanzania	0 (0%)	4 (<0.1%)	0 (0%)
Thailand	0 (0%)	51 (0.2%)	0 (0%)
The Gambia	0 (0%)	55 (0.2%)	0 (0%)
The Netherlands	0 (0%)	115 (0.4%)	0 (0%)
Togo	0 (0%)	20 (<0.1%)	0 (0%)
Trinidad and Tobago	0 (0%)	1 (<0.1%)	0 (0%)
Tunisia	0 (0%)	86 (0.3%)	0 (0%)

Turkey	0 (0%)	2,665 (9.2%)	0 (0%)
Turkmenistan	0 (0%)	6 (<0.1%)	0 (0%)
Uganda	0 (0%)	7 (<0.1%)	0 (0%)
Ukraine	0 (0%)	390 (1.3%)	0 (0%)
United Arab Emirates	0 (0%)	14 (<0.1%)	0 (0%)
United Kingdom	0 (0%)	92 (0.3%)	0 (0%)
United States of America	0 (0%)	158 (0.5%)	0 (0%)
Uruguay	0 (0%)	5 (<0.1%)	0 (0%)

Uzbekistan	0 (0%)	52 (0.2%)	0 (0%)
Venezuela	0 (0%)	20 (<0.1%)	0 (0%)
Vietnam	0 (0%)	74 (0.3%)	0 (0%)
Yemen	0 (0%)	11 (<0.1%)	0 (0%)
Zambia	0 (0%)	1 (<0.1%)	0 (0%)
Zimbabwe	0 (0%)	8 (<0.1%)	0 (0%)
Missing	0 (0%)	943 (3.2%)	0 (0%)

[1] n (%); (Median)

Deaths During Study

```
pid_death <- merged_data %>%
distinct(pid, todjahr, gebjahr)
```

```
#counting number of deaths that occured
sum(pid_death$todjahr > -2, na.rm = TRUE)

        ## [1] 6641

        #participants who died at an unknown date
sum(pid_death$todjahr == -1, na.rm = TRUE)

        ## [1] 28

        #participants whose status is unknown
sum(pid_death$todjahr == -2, na.rm = TRUE)

        ## [1] 92792

        #ages of death
pid_death <- pid_death %>%
  mutate(death_age = ifelse(todjahr > -1 & gebjahr > -1,
                            todjahr - gebjahr,
                            NA))


psych::describe(pid_death$death_age)

        ##    vars    n  mean   sd median trimmed   mad min max range skew kurtosis   se
## X1    1 6613 73.51 14.1     76   74.73 13.34  18 104    86 -0.9     0.96 0.17
```

Variable Exploration

Making Factor Variables

```
        # current health factor
ple0008_factor <- people_values %>%
 filter(variable=="ple0008")


merged_data <- merged_data %>%
 mutate(ple0008_f = factor(ple0008,
                           levels=c(-8, -4, -1, 1, 2, 3, 4, 5),
                           labels=c("Q.Not.Survey",
                                    "InAdmiss",
                                    "No Answer", "Very Good", "Good", "Satisfactory", "Poor",
                                    "Bad")))

        # limitations in daily life because of health
ple0009_factor <- people_values %>%
 filter(variable=="ple0009")


merged_data <- merged_data %>%
 mutate(ple0009_f = factor(ple0009,
                           levels=c(-8, -4, -1, 1, 2, 3),
                           labels=c("Q.Not.Survey", "InAdmiss", "No Answer",
                                    "Severly Restricted", "Somewhat Restricted", "Not Restricted")))
```

```
        # physical limitations
ple0032_factor <- people_values %>%
 filter(variable=="ple0032")

merged_data <- merged_data %>%
 mutate(ple0032_f = factor(ple0032,
                           levels=c(-8, -5, -1, 1, 2, 3, 4, 5),
                           labels=c("Q.Not.Survey", "Not Included", "No Answer",
                                    "Always", "Often", "Sometimes",
                                    "Almost Never", "Never")))
# limitations due to emotions
ple0033_factor <- people_values %>%
 filter(variable=="ple0033")

merged_data <- merged_data %>%
 mutate(ple0033_f = factor(ple0033,
                           levels=c(-8, -5, -1, 1, 2, 3, 4, 5),
                           labels=c("Q.Not.Survey", "Not Included", "No Answer",
                                    "Always", "Often", "Sometimes",
```

```
                              "Almost Never", "Never")))
#satisfaction with health
plh0171_factor <- people_values %>%
 filter(variable=="plh0171")


merged_data <- merged_data %>%
 mutate(plh0171_f = factor(plh0171,
                           levels=c(-5, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10),
                           labels=c("Not Included", "Does Not Apply", "No Answer",
                                    "0", "1", "2", " 3", "4", "5",
                                    "6", "7", "8",
                                    "9", "10"
                                    )))
#current life satisfaction
plh0182_factor <- people_values %>%
 filter(variable=="plh0182")


merged_data <- merged_data %>%
 mutate(plh0182_f = factor(plh0182,
                           levels=c(-5, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10),
                           labels=c("Not Included", "Does Not Apply", "No Answer",
                                    "0", "1", "2", " 3", "4", "5",
```

```
                              "6", "7", "8",
                              "9", "10"
                              )))



#hostility
plj0047_factor <- people_values %>%
 filter(variable=="plj0047")


merged_data <- merged_data %>%
 mutate(plj0047_f = factor(plj0047,
                           levels=c(-8, -5, -2, -1, 1, 2, 3),
                           labels=c("Q.Not.Survey", "Not Included", "Does Not Apply", "No Answer",
                                    "I Worry A Lot", "I Worry Somewhat", "I Don't Worry At All"
                                   )))
```

Preliminary Correlations

```
        cor(
merged_data[, c("ple0008", "ple0009",
                "ple0032", "ple0033",
```

```
            "plh0171", "plh0182")]
)
```

```
        ##           ple0008     ple0009     ple0032     ple0033     plh0171     plh0182
## ple0008  1.00000000  0.25207840  0.29005219  0.29452340 -0.16954717 -0.08084343
## ple0009  0.25207840  1.00000000 -0.01125661 -0.01486484 -0.01114523  0.02691110
## ple0032  0.29005219 -0.01125661  1.00000000  0.99343367  0.05021281  0.04397971
## ple0033  0.29452340 -0.01486484  0.99343367  1.00000000  0.03641680  0.04498570
## plh0171 -0.16954717 -0.01114523  0.05021281  0.03641680  1.00000000  0.51577001
## plh0182 -0.08084343  0.02691110  0.04397971  0.04498570  0.51577001  1.00000000
```

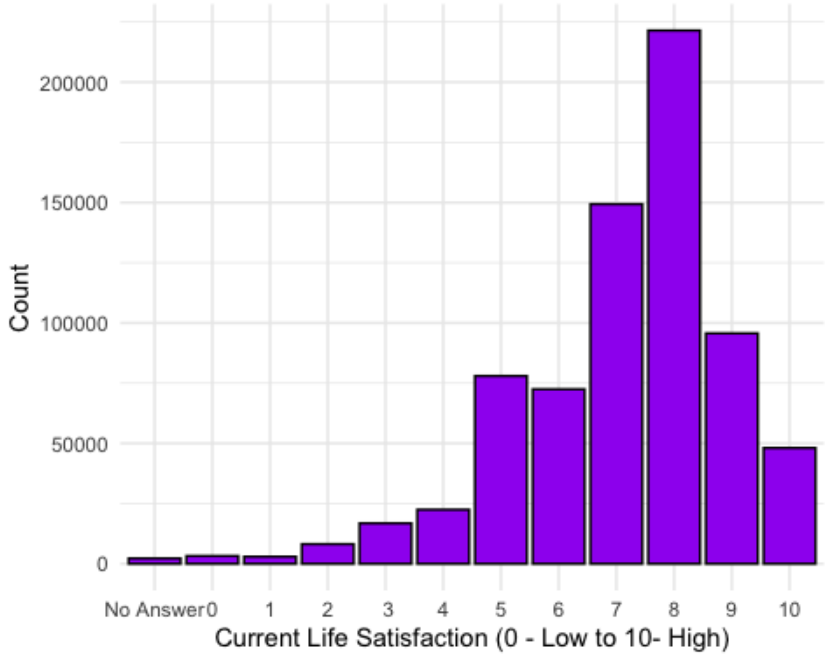
Life Satisfaction

```
        # Current Life Satisfaction
psych::describe(merged_data$plh0182_f)
```

```
        ##    vars      n  mean   sd median trimmed  mad min max range  skew kurtosis se
## X1*    1 724542 11.08 1.99     12   11.28 1.48   1  14    13 -1.46     3.88  0
```

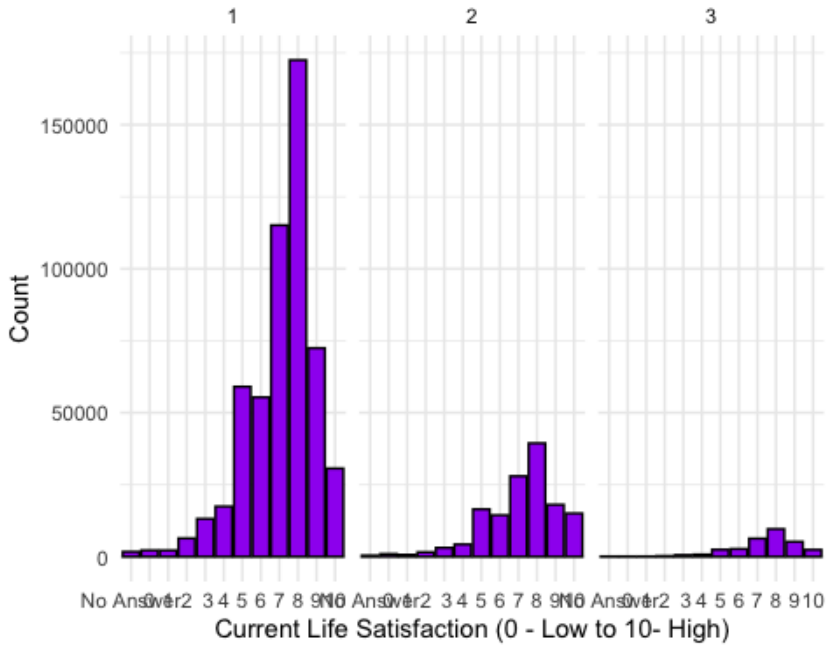
```
        merged_data %>%
 filter( plh0182 >= -1) %>%
 ggplot(aes(x=plh0182_f)) +
 geom_histogram(stat="count", fill="purple", color="black") +
 xlab("Current Life Satisfaction (0 - Low to 10- High)") + ylab("Count") +
 theme_minimal()
```

```
## Warning in geom_histogram(stat = "count", fill = "purple", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```



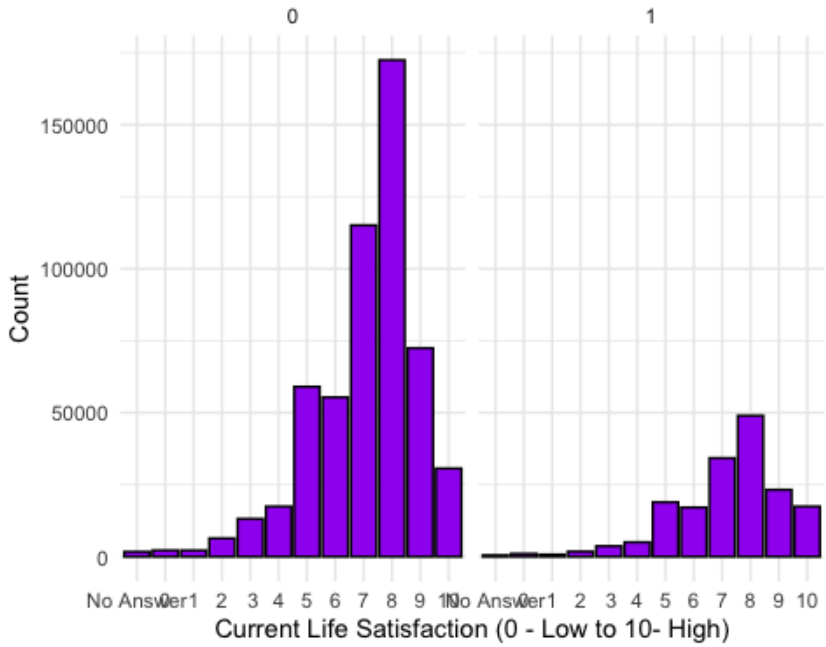

```
merged_data %>%
filter( plh0182 >= -1) %>%
ggplot(aes(x=plh0182_f)) +
geom_histogram(stat="count", fill="purple", color="black") +
xlab("Current Life Satisfaction (0 - Low to 10- High)") + ylab("Count") +
theme_minimal() +
facet_wrap(~migback)
```

```
## Warning in geom_histogram(stat = "count", fill = "purple", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```




```
        merged_data %>%
filter( plh0182 >= -1) %>%
ggplot(aes(x=plh0182_f)) +
geom_histogram(stat="count", fill="purple", color="black") +
xlab("Current Life Satisfaction (0 - Low to 10- High)") + ylab("Count") +
theme_minimal() +
facet_wrap(~migrant_binary)
```

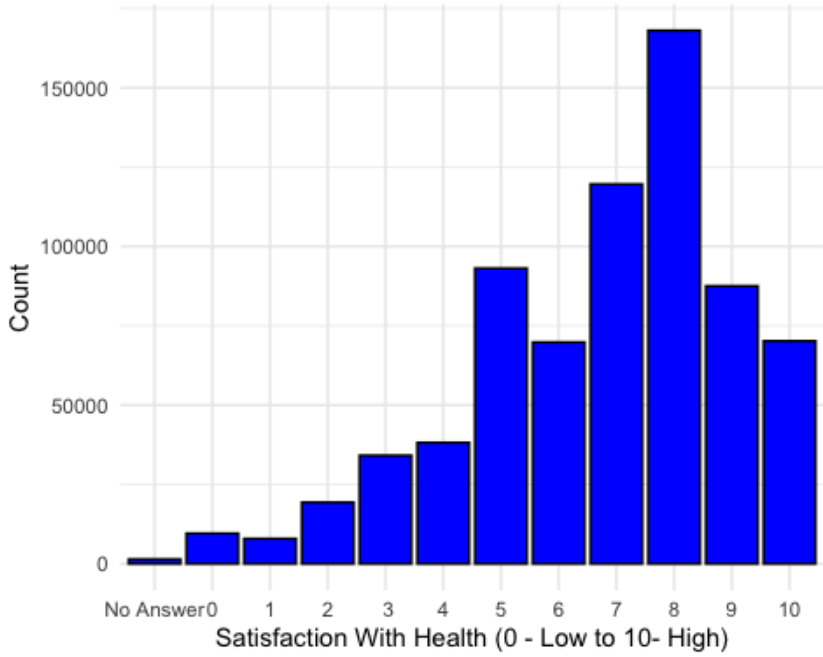
```
        ## Warning in geom_histogram(stat = "count", fill = "purple", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```



Satisfaction With Health

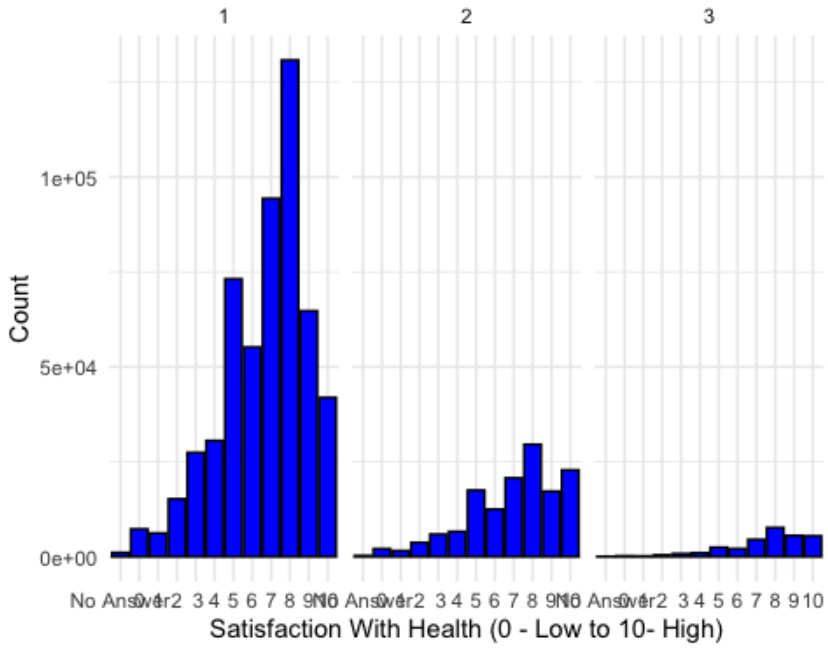
```
        #Satisfaction with Health
psych::describe(merged_data$plh0171_f)

        ##     vars      n  mean   sd median trimmed  mad min max range  skew kurtosis se
## X1*    1 724540 10.68 2.44     11    10.91 2.97   1  14    13 -1.04     1.37  0

        merged_data %>%
 filter(plh0171 >= -1) %>%
 ggplot(aes(x=plh0171_f)) +
 geom_histogram(stat="count", fill="blue", color="black") +
 xlab("Satisfaction With Health (0 - Low to 10- High)") + ylab("Count") +
 theme_minimal()
```

```
## Warning in geom_histogram(stat = "count", fill = "blue", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```

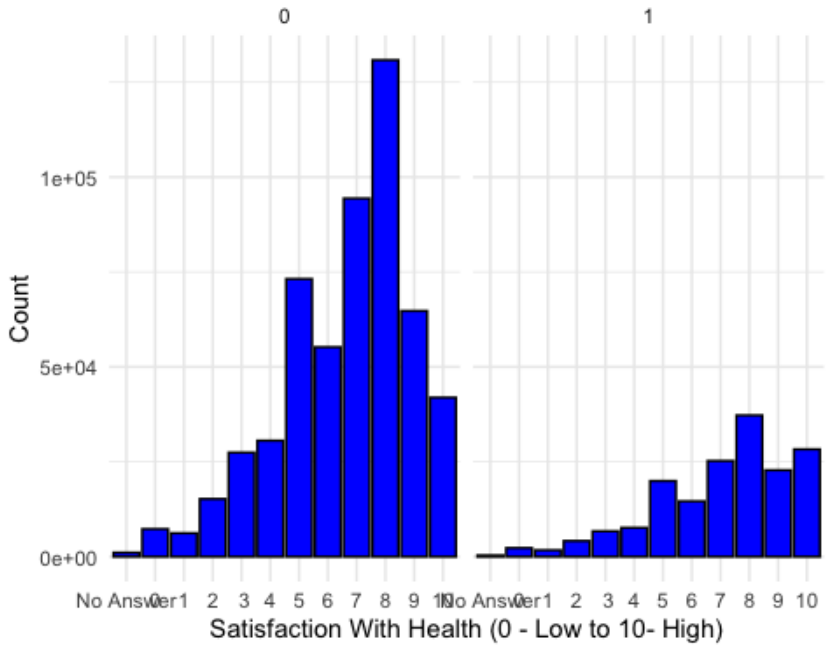



```
merged_data %>%
filter(plh0171 >= -1) %>%
ggplot(aes(x=plh0171_f)) +
geom_histogram(stat="count", fill="blue", color="black") +
xlab("Satisfaction With Health (0 - Low to 10- High)") + ylab("Count") +
theme_minimal() +
facet_wrap(~migback)
```

```
## Warning in geom_histogram(stat = "count", fill = "blue", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```



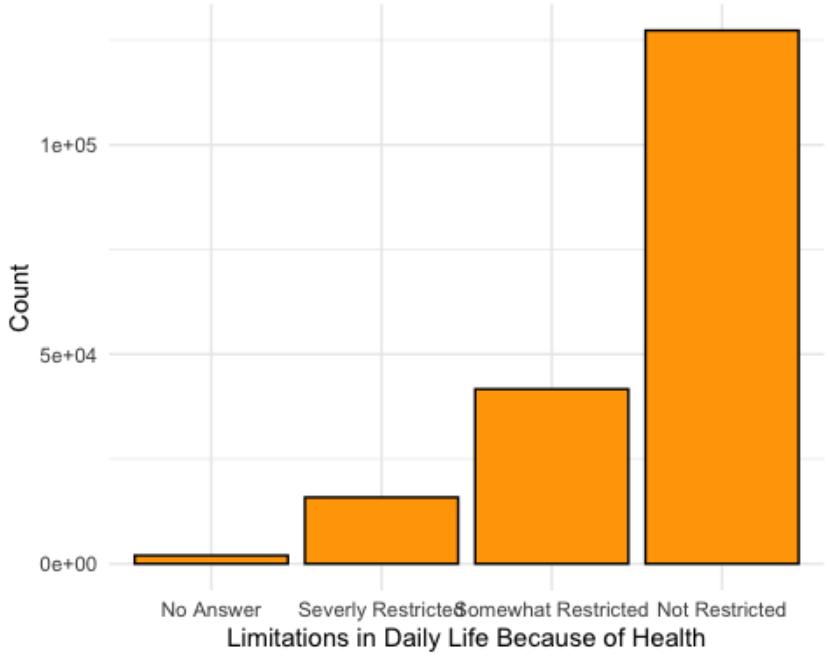

```
        merged_data %>%
filter(plh0171 >= -1) %>%
ggplot(aes(x=plh0171_f)) +
geom_histogram(stat="count", fill="blue", color="black") +
xlab("Satisfaction With Health (0 - Low to 10- High)") + ylab("Count") +
theme_minimal() +
facet_wrap(~migrant_binary)

        ## Warning in geom_histogram(stat = "count", fill = "blue", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```



Limitations in Daily Life Because of Health

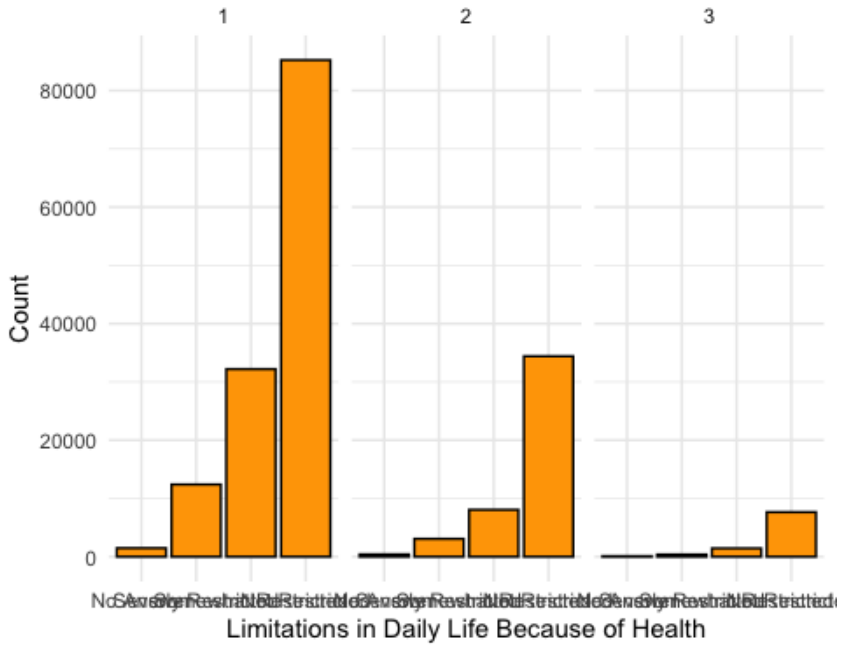
```
        psych::describe(merged_data$ple0009_f)

        ##     vars      n mean   sd median trimmed mad min max range skew kurtosis se
## X1*    1 638004 2.34 2.12      1    2.05   0   1   6     5    1    -0.92  0

        merged_data %>%
 filter(ple0009 >= -1) %>%
 ggplot(aes(x=ple0009_f)) +
 geom_histogram(stat="count", fill="orange", color="black") +
 xlab("Limitations in Daily Life Because of Health") + ylab("Count") +
 theme_minimal()
```

```
## Warning in geom_histogram(stat = "count", fill = "orange", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```

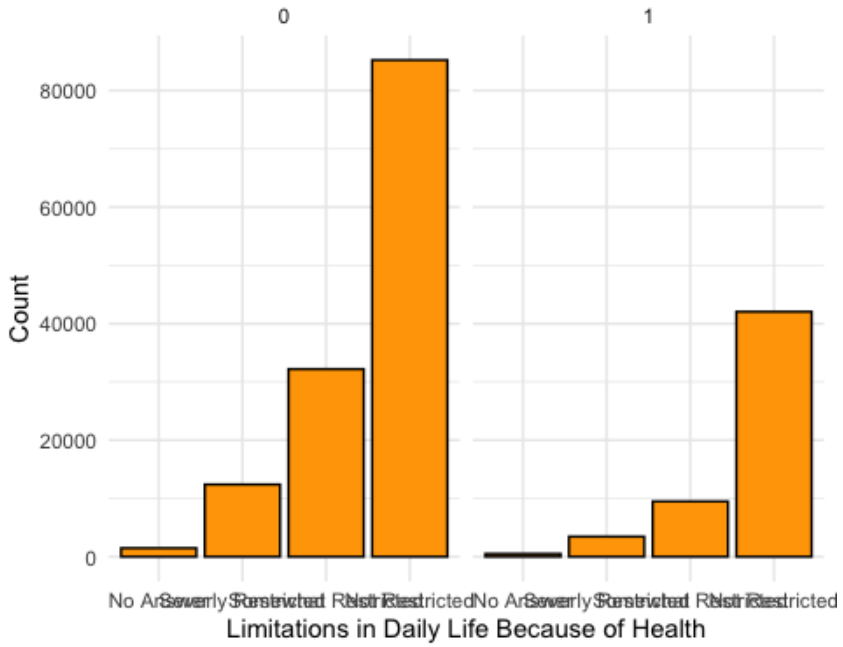



```
merged_data %>%
filter(ple0009 >= -1) %>%
ggplot(aes(x=ple0009_f)) +
geom_histogram(stat="count", fill="orange", color="black") +
xlab("Limitations in Daily Life Because of Health") + ylab("Count") +
theme_minimal() +
facet_wrap(~migback)
```

```
## Warning in geom_histogram(stat = "count", fill = "orange", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```



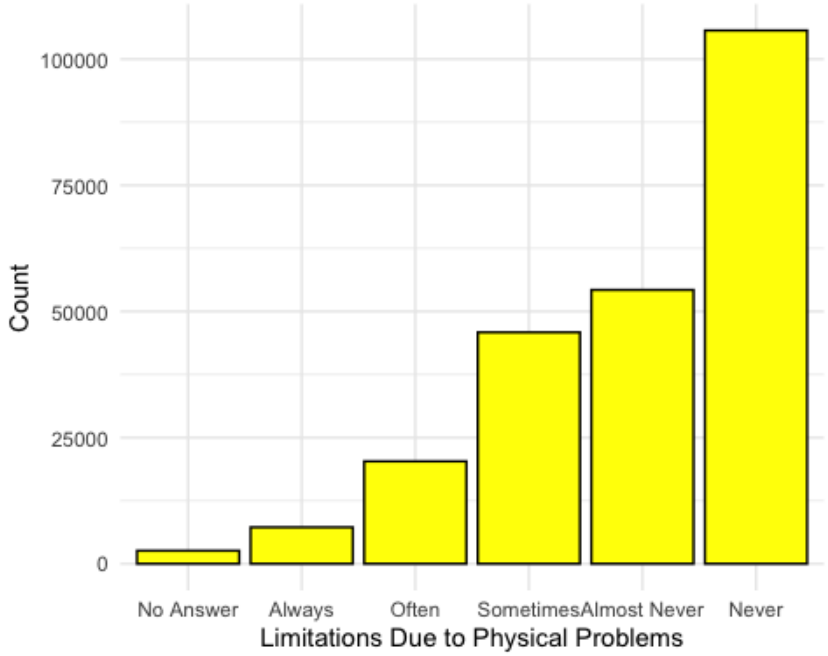

```
        merged_data %>%
filter(ple0009 >= -1) %>%
ggplot(aes(x=ple0009_f)) +
geom_histogram(stat="count", fill="orange", color="black") +
xlab("Limitations in Daily Life Because of Health") + ylab("Count") +
theme_minimal() +
facet_wrap(~migrant_binary)

        ## Warning in geom_histogram(stat = "count", fill = "orange", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```



Limitations Due to Physical Problems

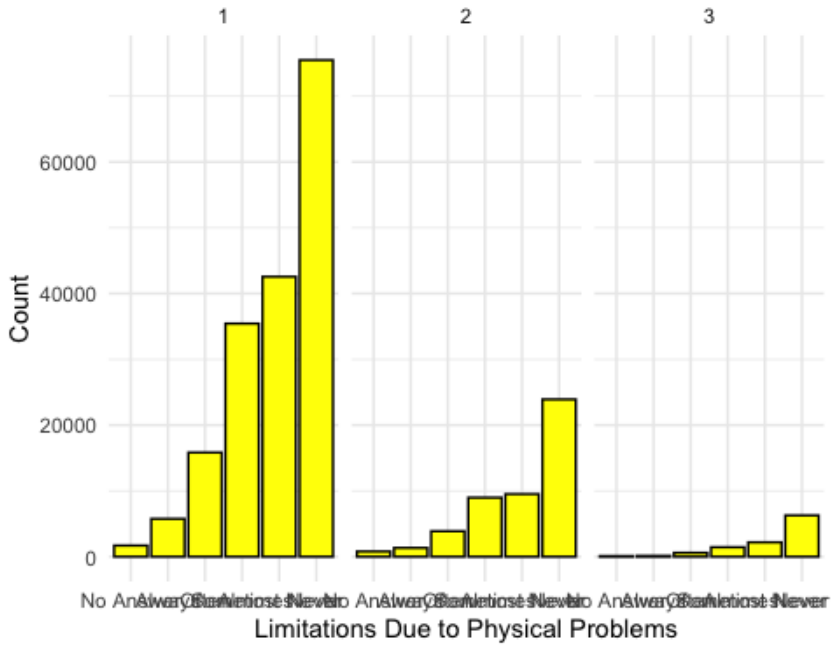
```
psych::describe(merged_data$ple0032_f)

##     vars      n mean  sd median trimmed mad min max range skew kurtosis se
## X1*    1 724542 3.06 2.8      1     2.7   0   1   8     7 0.86    -1.05  0

merged_data %>%
 filter(ple0032 >= -1) %>%
 ggplot(aes(x=ple0032_f)) +
 geom_histogram(stat="count", fill="yellow", color="black") +
 xlab("Limitations Due to Physical Problems") + ylab("Count") +
 theme_minimal()
```

```
## Warning in geom_histogram(stat = "count", fill = "yellow", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```

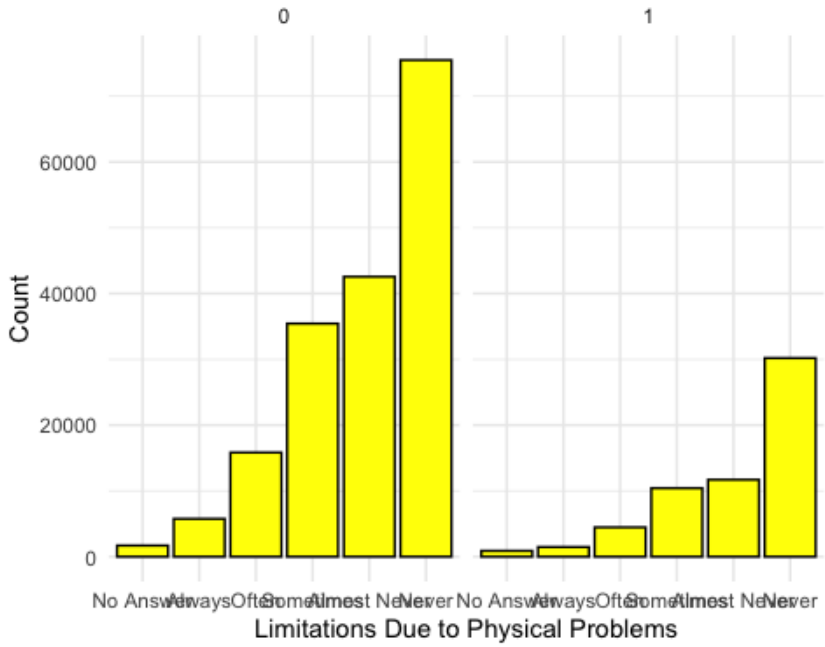



```
merged_data %>%
filter(ple0032 >= -1) %>%
ggplot(aes(x=ple0032_f)) +
geom_histogram(stat="count", fill="yellow", color="black") +
xlab("Limitations Due to Physical Problems") + ylab("Count") +
theme_minimal() +
facet_wrap(~migback)
```

```
## Warning in geom_histogram(stat = "count", fill = "yellow", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```



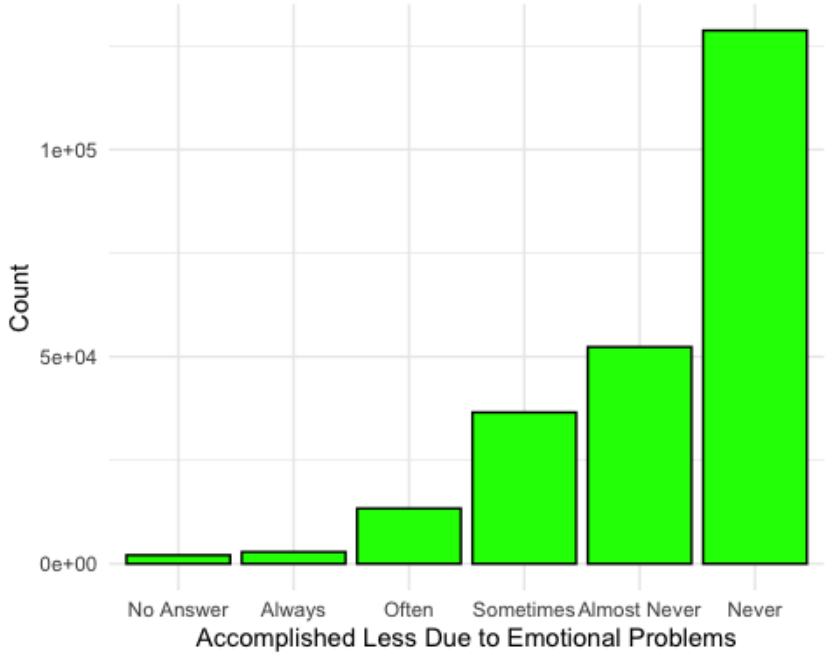

```
        merged_data %>%
filter(ple0032 >= -1) %>%
ggplot(aes(x=ple0032_f)) +
geom_histogram(stat="count", fill="yellow", color="black") +
xlab("Limitations Due to Physical Problems") + ylab("Count") +
theme_minimal() +
facet_wrap(~migrant_binary)

        ## Warning in geom_histogram(stat = "count", fill = "yellow", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```



Accomplished Less Due to Emotional Problems

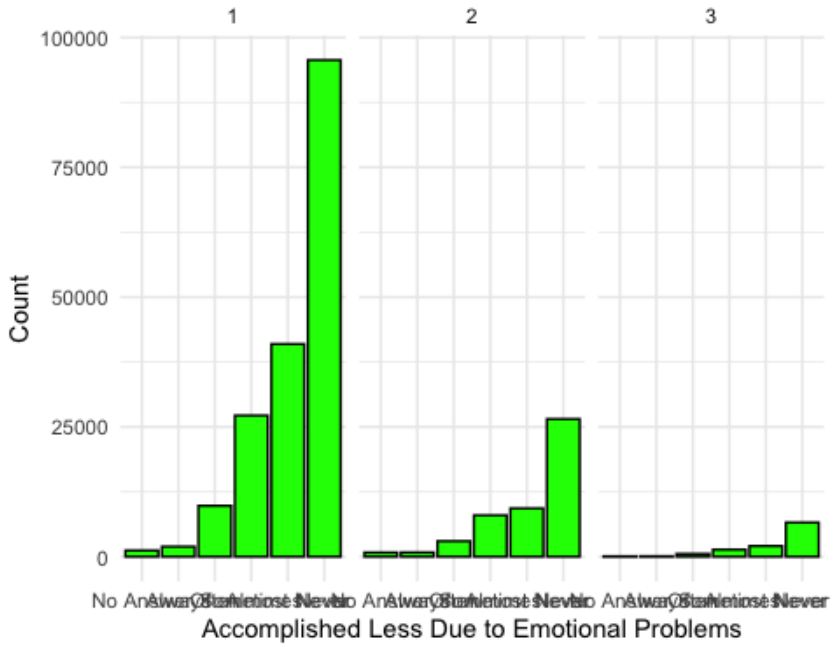
```
psych::describe(merged_data$ple0033_f)

##     vars      n mean  sd median trimmed mad min max range skew kurtosis se
## X1*    1 724542 3.14 2.9      1    2.81   0   1   8     7 0.82    -1.17  0

merged_data %>%
 filter(ple0033 >= -1) %>%
 ggplot(aes(x=ple0033_f)) +
 geom_histogram(stat="count", fill="green", color="black") +
 xlab("Accomplished Less Due to Emotional Problems") + ylab("Count") +
 theme_minimal()
```

```
## Warning in geom_histogram(stat = "count", fill = "green", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```

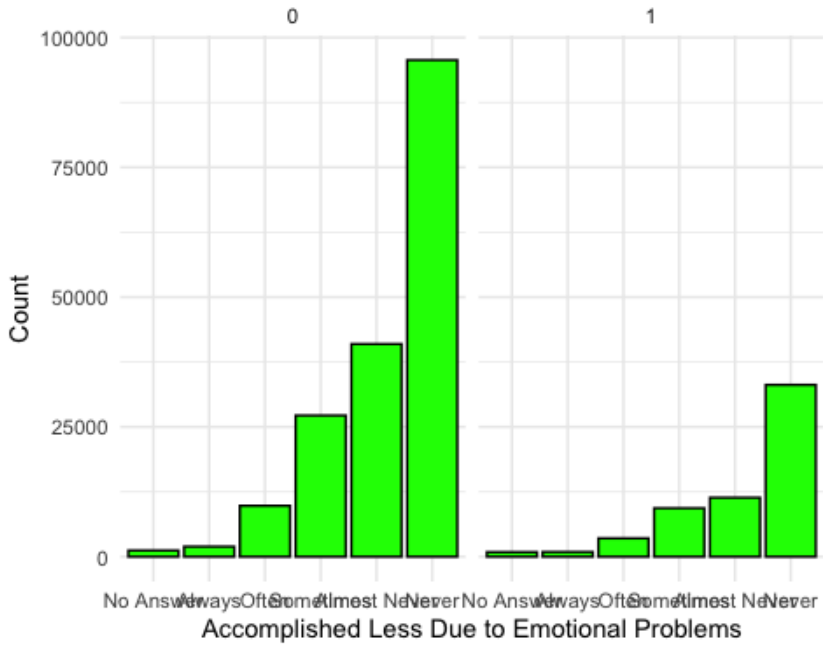



```
merged_data %>%
filter(ple0033 >= -1) %>%
ggplot(aes(x=ple0033_f)) +
geom_histogram(stat="count", fill="green", color="black") +
xlab("Accomplished Less Due to Emotional Problems") + ylab("Count") +
theme_minimal() +
facet_wrap(~migback)
```

```
## Warning in geom_histogram(stat = "count", fill = "green", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```




```
        merged_data %>%
filter(ple0033 >= -1) %>%
ggplot(aes(x=ple0033_f)) +
geom_histogram(stat="count", fill="green", color="black") +
xlab("Accomplished Less Due to Emotional Problems") + ylab("Count") +
theme_minimal() +
facet_wrap(~migrant_binary)

        ## Warning in geom_histogram(stat = "count", fill = "green", color = "black"):
## Ignoring unknown parameters: `binwidth` and `bins`
```



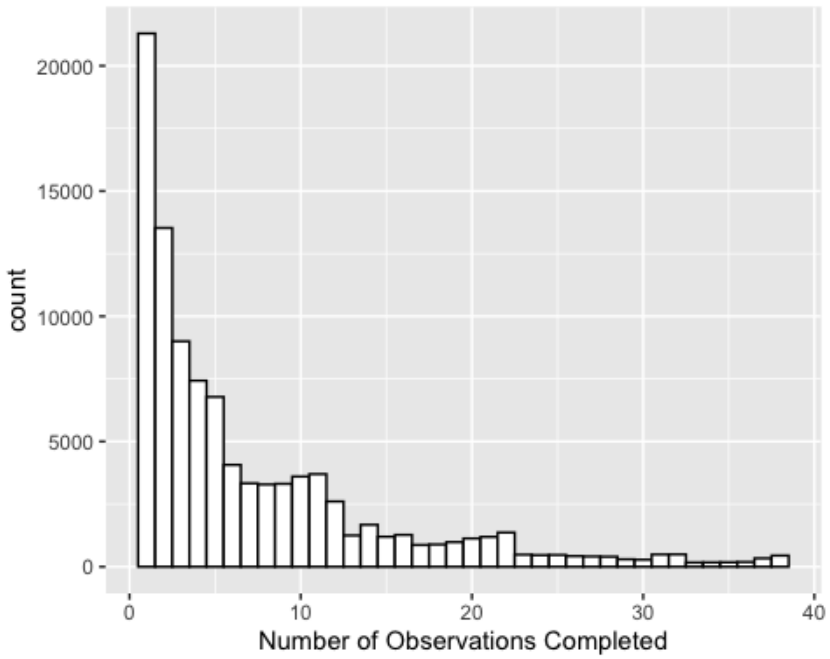
Overall Compliance

```
#counts of surveys completed
table(merged_data$syear)

##
##  1984  1985  1986  1987  1988  1989  1990  1991  1992  1993  1994  1995  1996
## 11654 10503 10101  9970  9509  9195 13245 12941 12676 12475 12697 13031 12779
##  1997  1998  1999  2000  2001  2002  2003  2004  2005  2006  2007  2008  2009
## 12560 13882 13367 23332 21206 22670 21455 20920 20049 21208 19837 18703 19758
##  2010  2011  2012  2013  2014  2015  2016  2017  2018  2019  2020  2021
## 25390 27328 26588 29429 26127 25861 27690 30883 28776 28398 28940 19409
```

```
        #counting number of survey and how many observations completed for each person
compliance_stats <- merged_data %>%
 group_by(pid) %>%
 summarise(
  waves_completed = n_distinct(syear)
 )
# looking at counts
describe(compliance_stats)

        ##                vars     n       mean         sd  median     trimmed
## pid               1 99433 16045135.41 15103769.29 8659301 14588074.24
## waves_completed    2 99433       7.29       7.72       4       5.75
##                       mad min      max    range skew kurtosis       se
## pid            12725600.58 101 55572102 55572001 0.60    -0.76 47898.30
## waves_completed       4.45   1       38       37 1.74     2.84     0.02

        table(compliance_stats$waves_completed, useNA = "always")

        ##
##     1     2     3     4     5     6     7     8     9    10    11    12    13
## 21291 13525  9000  7430  6780  4066  3329  3286  3310  3600  3698  2607  1249
##    14    15    16    17    18    19    20    21    22    23    24    25    26
##  1676  1198  1271   869   888   983  1130  1191  1365   482   463   471   424
```

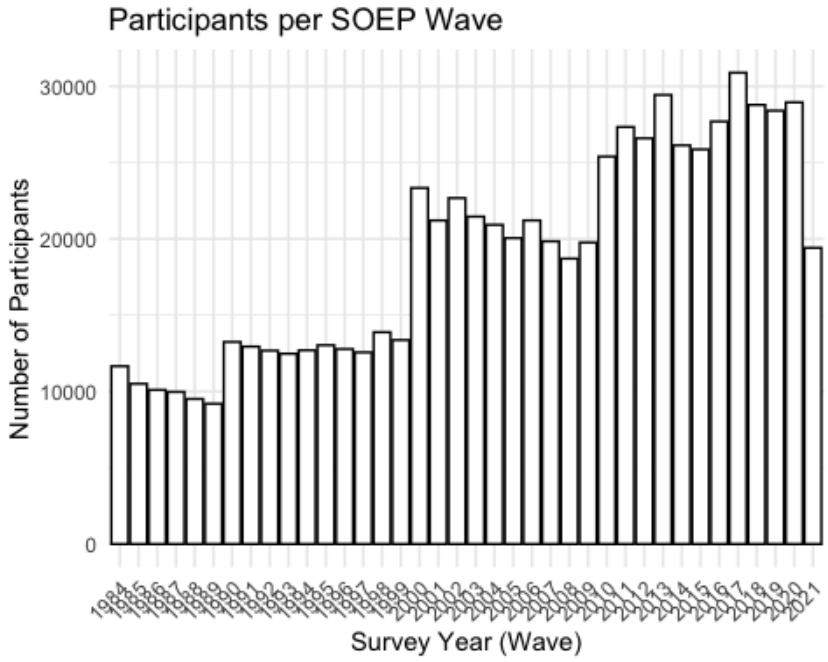
```
##   27   28   29   30   31   32   33   34   35   36   37   38 <NA>
##  410  401  297  281  490  493  166  170  175  188  335  445    0
```

```
        obs_per_id <- merged_data %>%
 group_by(pid) %>%
 summarise(num_obs = n())

        #histogram
obs_per_id %>%
 ggplot(aes(x = num_obs)) +
 geom_histogram(fill = "white", color = "black", binwidth = 1) +
 labs(x = "Number of Observations Completed")
```



Participants Per Wave

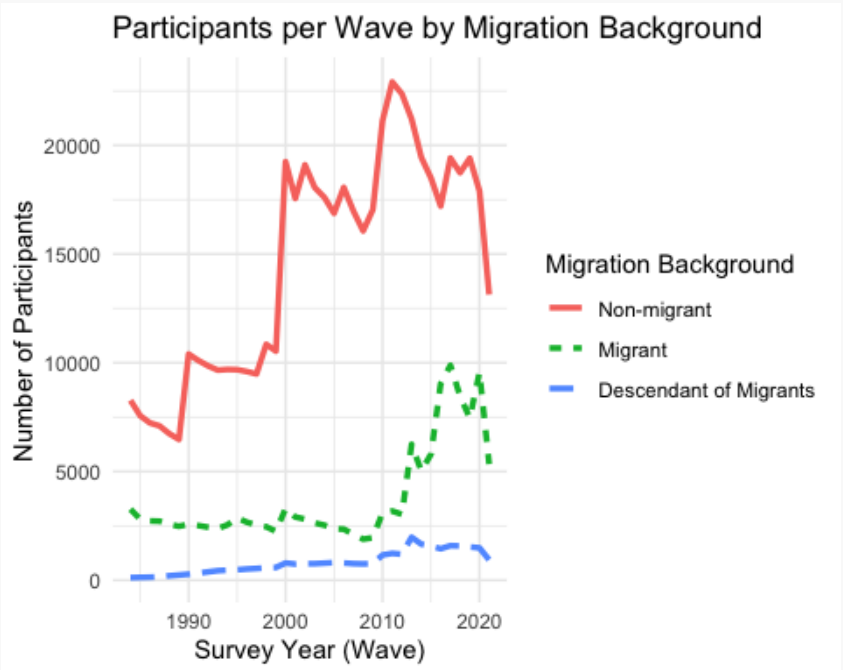
```
        merged_data %>%
distinct(pid, syear) %>%
count(syear, name = "n_participants") %>%
ggplot(aes(x = factor(syear), y = n_participants)) +
geom_col(fill = "white", color = "black") +
labs(
  x = "Survey Year (Wave)",
  y = "Number of Participants",
  title = "Participants per SOEP Wave"
) +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Participants Over Time

```
        merged_data %>%
distinct(pid, syear, migback) %>%
mutate(
  migback = factor(
    migback,
    levels = c(1, 2, 3),
    labels = c("Non-migrant",
               "Migrant",
               "Descendant of Migrants")
  )%>%
count(syear, migback, name = "Participants") %>%
ggplot(aes(x = syear,
           y = Participants,
           color = migback,
           linetype = migback)) +
geom_line(linewidth = 1.2) +
```

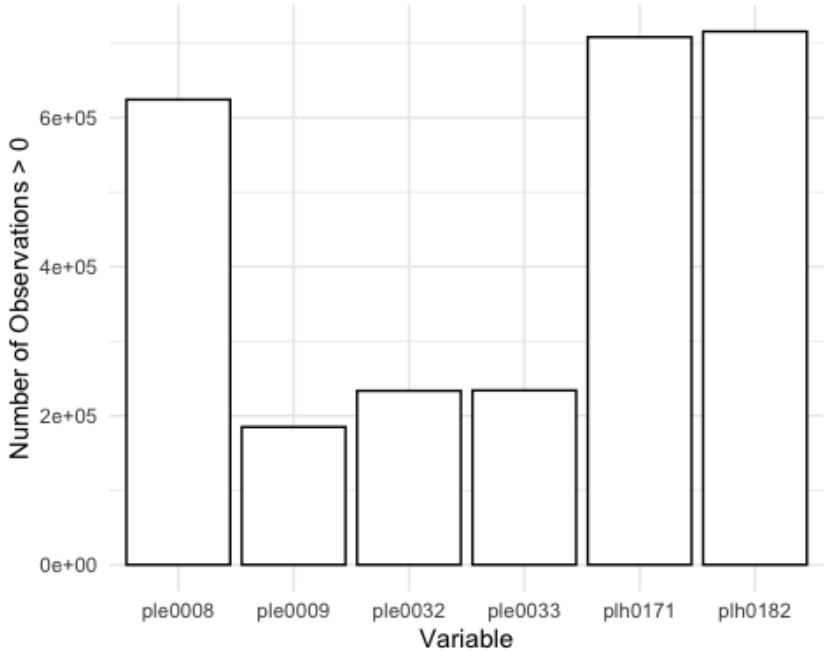
```
        labs(
  x = "Survey Year (Wave)",
  y = "Number of Participants",
  color = "Migration Background",
  linetype = "Migration Background",
        title = "Participants per Wave by Migration Background"
) +
theme_minimal()
```



Variable Compliance

```
          var_list <- c("ple0008", "ple0009",
            "ple0032", "ple0033",
            "plh0171", "plh0182")


# Count observations where value > 0
valid_counts <- merged_data %>%
  summarise(
    across(all_of(var_list),
           ~ sum(. > 0, na.rm = TRUE))


          ) %>%
  pivot_longer(
    everything(),
    names_to = "variable",
    values_to = "n_gt_zero"
  )


ggplot(valid_counts, aes(x = variable, y = n_gt_zero)) +
  geom_col(fill = "white", color = "black") +
  labs(
    x = "Variable",
```

```
  y = "Number of Observations > 0"
) +
theme_minimal()
```

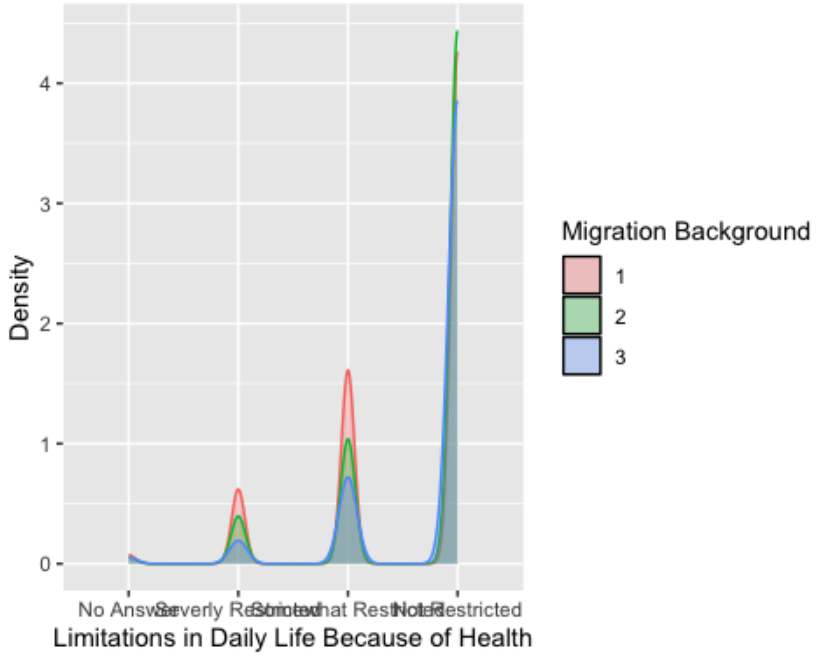


Preliminary Plots

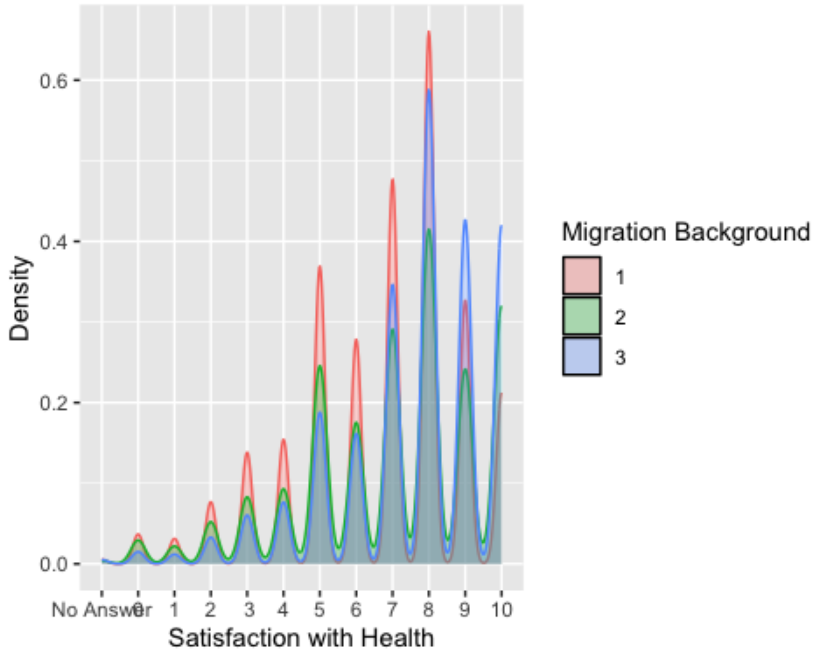
Density Distributions

```
        merged_data %>%
filter(ple0009 >= -1) %>%
ggplot(aes(x=ple0009_f)) +
geom_density(aes(group=factor(migback),
                 colour=factor(migback), fill=factor(migback)), alpha=0.3) +
guides(colour="none",
```

```
      fill=guide_legend(title="Migration Background")) +
labs(x="Limitations in Daily Life Because of Health", y="Density")
```



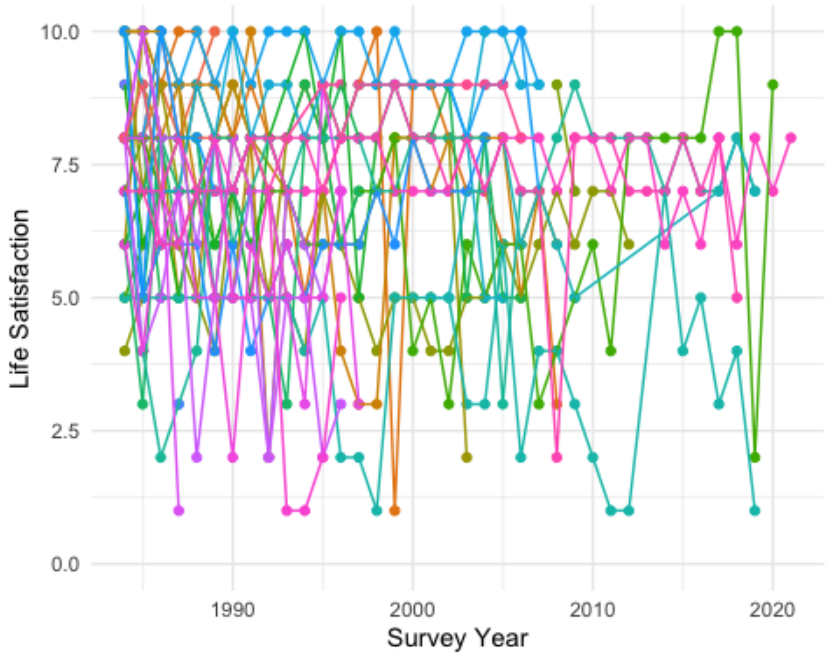

```
        merged_data %>%
filter(plh0171 >= -1) %>%
ggplot(aes(x=plh0171_f)) +
geom_density(aes(group=factor(migback),
            colour=factor(migback), fill=factor(migback)), alpha=0.3) +
guides(colour="none",
      fill=guide_legend(title="Migration Background")) +
labs(x="Satisfaction with Health", y="Density")
```



Variables Over Time

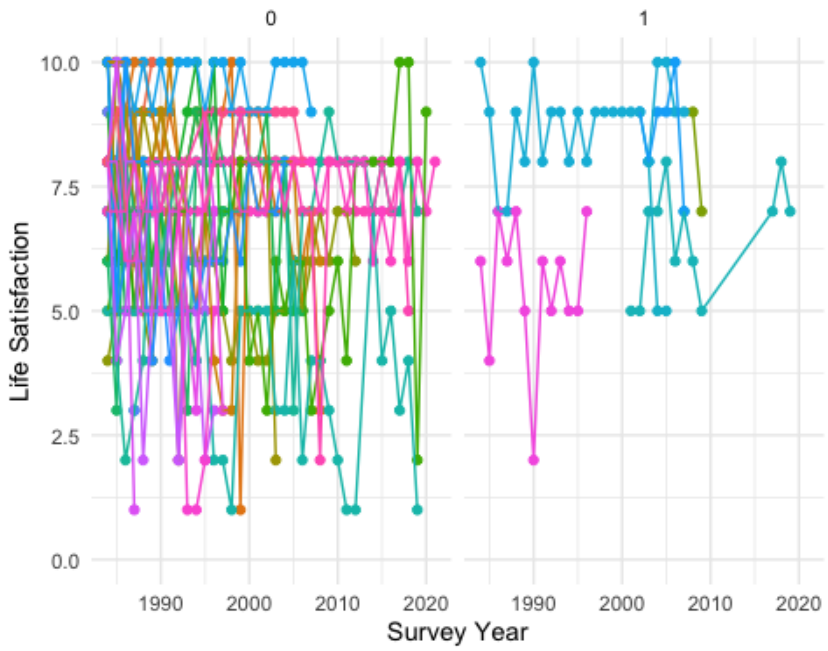
```
        #create new id numbers for plotting
merged_data <- merged_data %>%
 mutate(pid_seq = dense_rank(pid))
#create time variable for plotting
merged_data <- merged_data %>%
 group_by(pid) %>%
 mutate(
  time = syear - min(syear, na.rm = TRUE)
 ) %>%
 ungroup()
```

```
        merged_data %>%
filter(plh0182 > 0) %>%
filter(pid_seq < 50) %>%
ggplot(aes(x = syear, y = plh0182, color=factor(pid_seq))) +
geom_point() +
geom_line() +
labs(
  x = "Survey Year",
  y = "Life Satisfaction"
) +
ylim(0, 10) +
theme_minimal() +
theme(legend.position = "none")
```



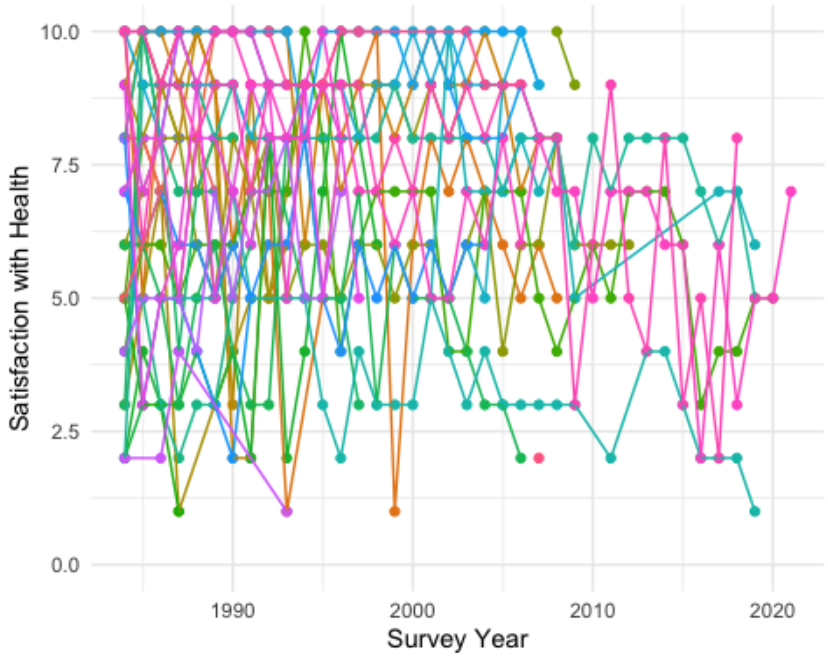

```
        merged_data %>%
filter(plh0182 > 0) %>%
filter(pid_seq < 50) %>%
ggplot(aes(x = syear, y = plh0182, color=factor(pid_seq))) +
geom_point() +
geom_line() +
labs(
  x = "Survey Year",
  y = "Life Satisfaction"
) +
ylim(0, 10) +
theme_minimal() +
```

```
theme(legend.position = "none") +

facet_wrap(~migrant_binary)
```



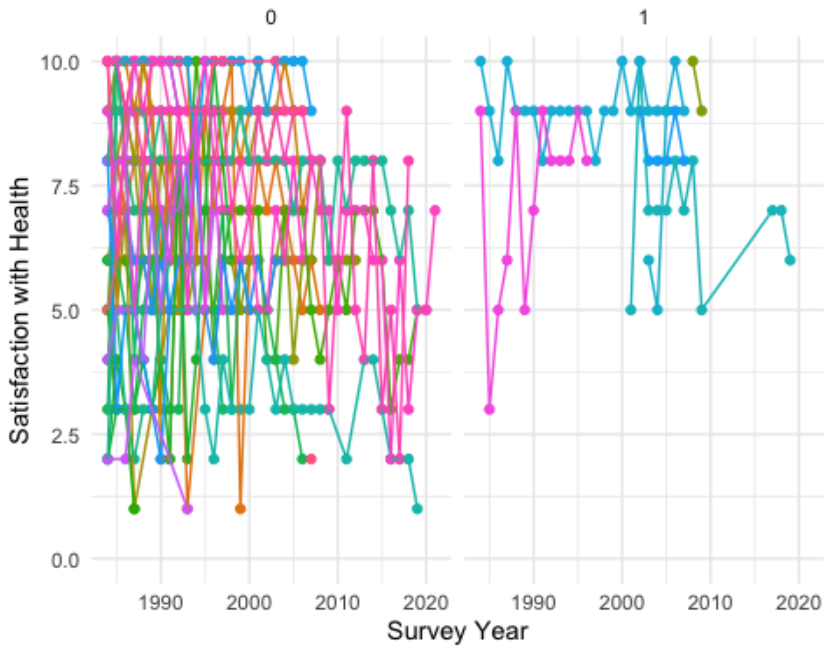

```
        merged_data %>%

filter(plh0171 > 0) %>%

filter(pid_seq < 50) %>%

ggplot(aes(x = syear, y = plh0171, color=factor(pid_seq))) +

geom_point() +

geom_line() +

labs(

  x = "Survey Year",

  y = "Satisfaction with Health"
```

```
) +
ylim(0, 10) +
theme_minimal() +
theme(legend.position = "none")
```




```
        merged_data %>%
filter(plh0171 > 0) %>%
filter(pid_seq < 50) %>%
ggplot(aes(x = syear, y = plh0171, color=factor(pid_seq))) +
geom_point() +
geom_line() +
labs(
```

```
  x = "Survey Year",
  y = "Satisfaction with Health"
) +
ylim(0, 10) +
theme_minimal() +
theme(legend.position = "none")  +
facet_wrap(~migrant_binary)
```

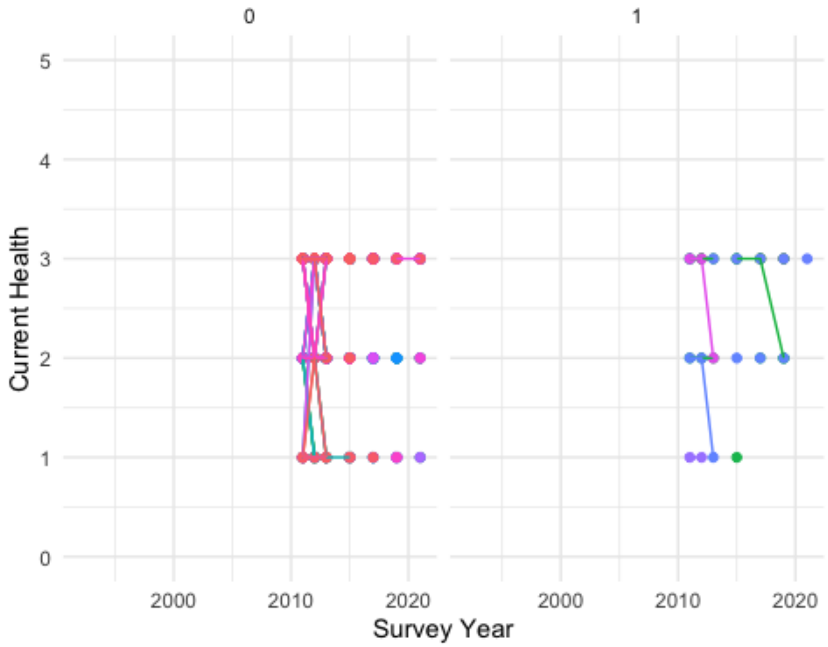



```
        merged_data %>%
filter(ple0008 > 0) %>%
filter(pid_seq < 500) %>%
ggplot(aes(x = syear, y = ple0009, color=factor(pid_seq))) +
```

```
  geom_point() +
  geom_line() +
  labs(
    x = "Survey Year",
    y = "Current Health"
  ) +
  ylim(0, 5) +
  theme_minimal() +
  theme(legend.position = "none") +
  facet_wrap(~migrant_binary)
```

```
## Warning: Removed 3642 rows containing missing values or values outside the scale
range
## (`geom_point()`).
```

```
## Warning: Removed 3419 rows containing missing values or values outside the scale
range
## (`geom_line()`).
```

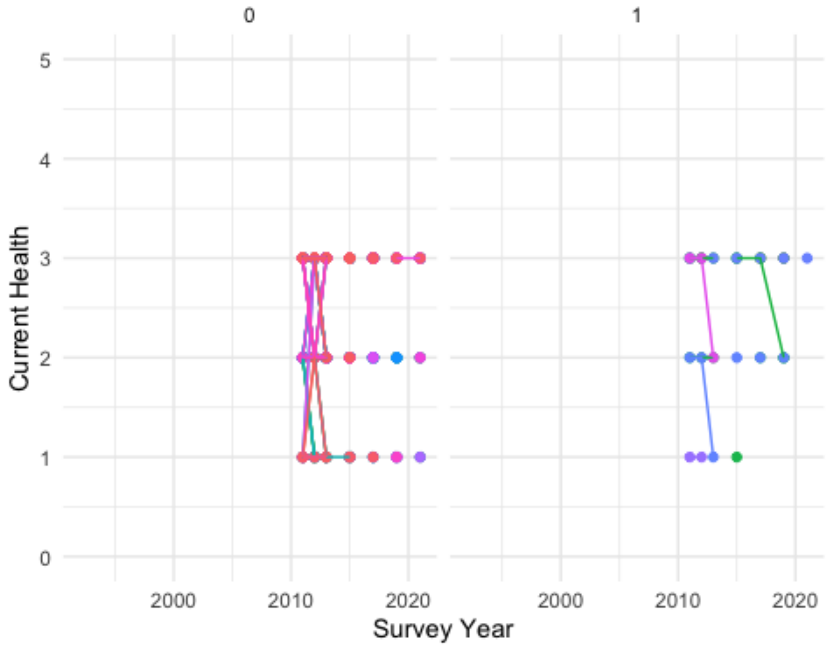



```
        merged_data %>%
filter(ple0008 > 0) %>%
filter(pid_seq < 500) %>%
ggplot(aes(x = syear, y = ple0009, color=factor(pid_seq))) +
geom_point() +
geom_line() +
labs(
  x = "Survey Year",
  y = "Current Health"
) +
ylim(0, 5) +
theme_minimal() +
```

```
 theme(legend.position = "none") +
 facet_wrap(~migrant_binary)
```

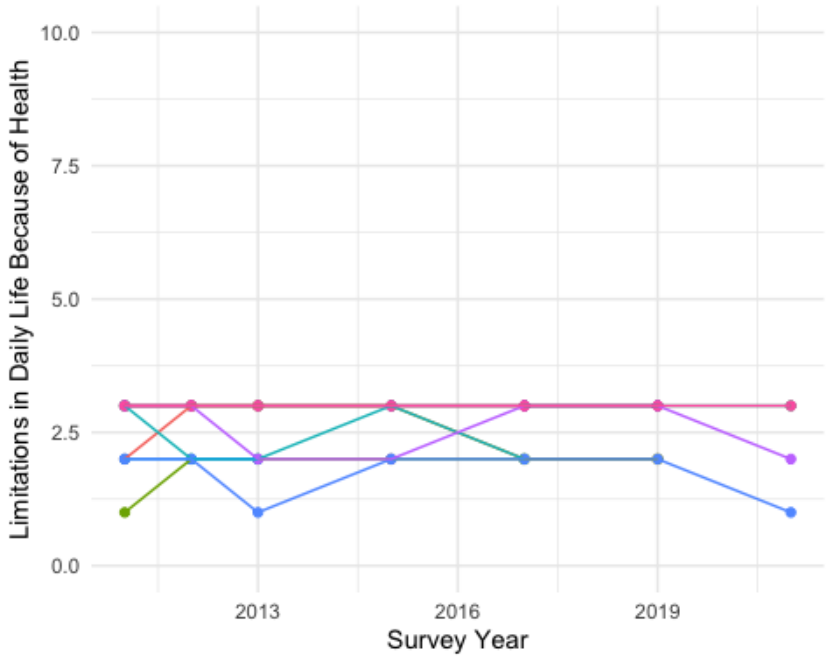
```
## Warning: Removed 3642 rows containing missing values or values outside the scale
range
## (`geom_point()`).
```

```
## Warning: Removed 3419 rows containing missing values or values outside the scale
range
## (`geom_line()`).
```

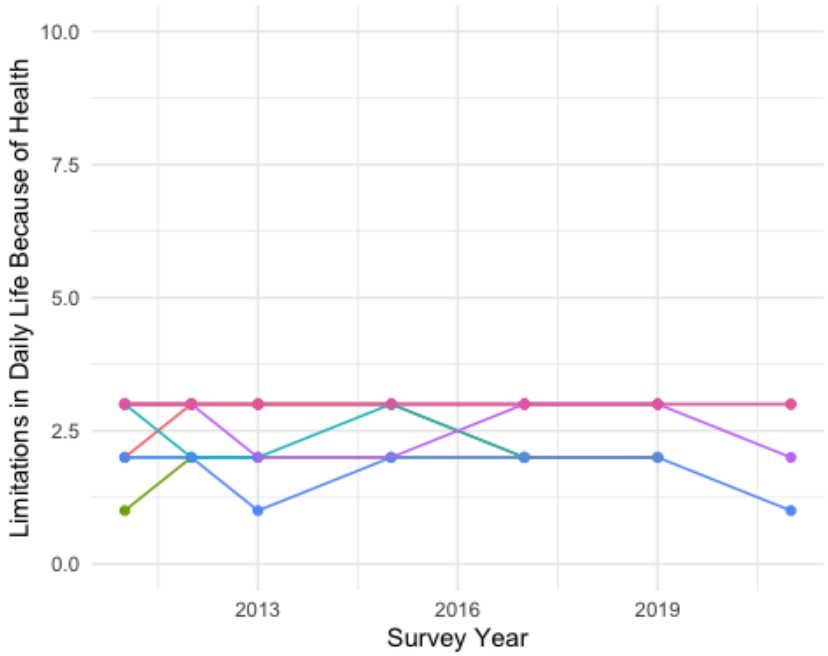



```
merged_data %>%
 filter(ple0009 > 0) %>%
 filter(pid_seq < 100) %>%
```

```
ggplot(aes(x = syear, y = ple0009, color=factor(pid_seq))) +
geom_point() +
geom_line() +
labs(
  x = "Survey Year",
  y = "Limitations in Daily Life Because of Health"
) +
ylim(0, 10) +
theme_minimal() +
theme(legend.position = "none")
```

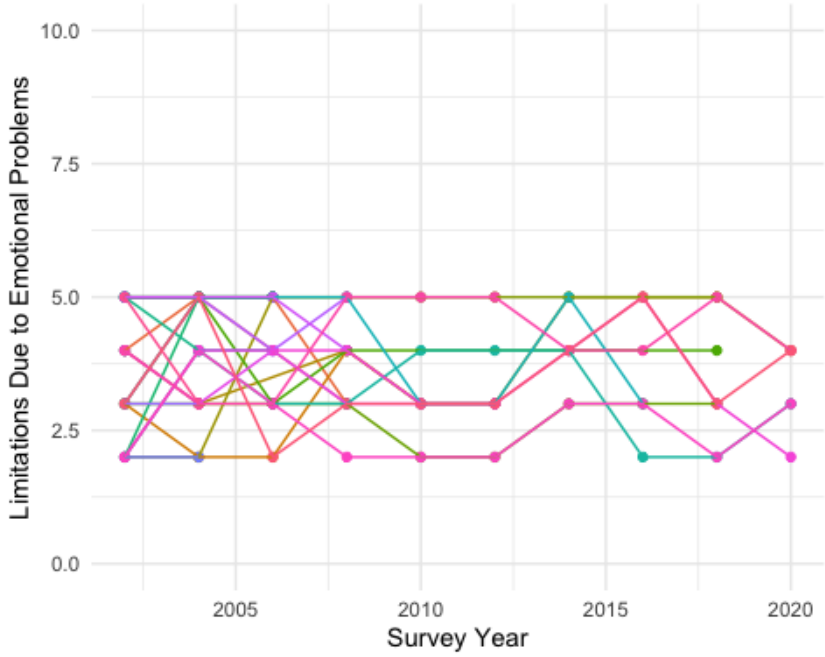


```
        merged_data %>%
filter(ple0009 > 0) %>%
filter(pid_seq < 100) %>%
ggplot(aes(x = syear, y = ple0009, color=factor(pid_seq))) +
geom_point() +
geom_line() +
labs(
  x = "Survey Year",
  y = "Limitations in Daily Life Because of Health"
) +
ylim(0, 10) +
theme_minimal() +
theme(legend.position = "none")
```



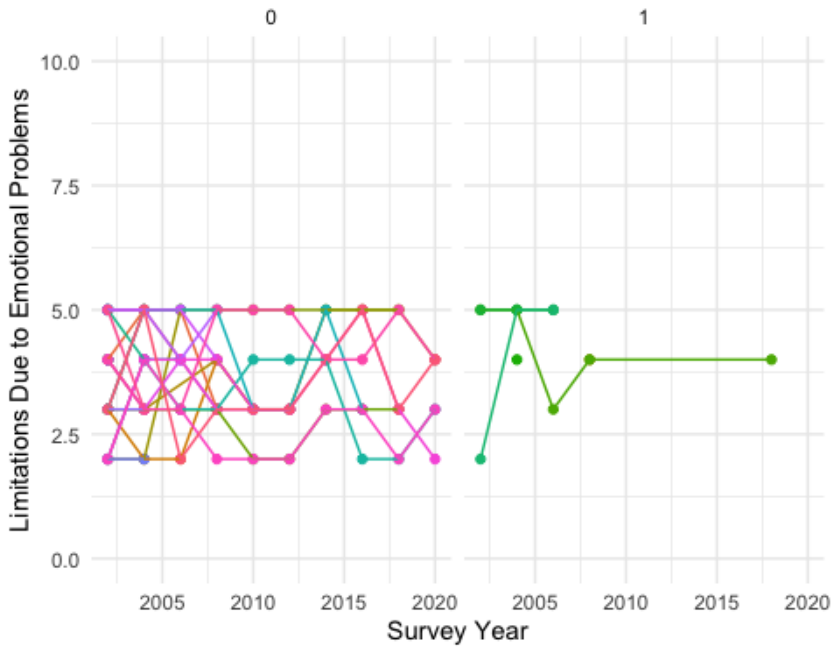

```
        merged_data %>%
filter(ple0033 > 0) %>%
filter(pid_seq < 100) %>%
ggplot(aes(x = syear, y = ple0033, color=factor(pid_seq))) +
geom_point() +
geom_line() +
labs(
  x = "Survey Year",
  y = "Limitations Due to Emotional Problems"
) +
ylim(0, 10) +
```

```
theme_minimal() +
theme(legend.position = "none")
```




```
        merged_data %>%
filter(ple0033 > 0) %>%
filter(pid_seq < 100) %>%
ggplot(aes(x = syear, y = ple0033, color=factor(pid_seq))) +
geom_point() +
geom_line() +
labs(
  x = "Survey Year",
  y = "Limitations Due to Emotional Problems"
```

```
) +
ylim(0, 10) +
theme_minimal() +
theme(legend.position = "none") +
facet_wrap(~migrant_binary)
```



Saving Data Set

```
write.csv(merged_data, file = "merged_data.csv")
```

SOEP Growth Modeling

Set Up

```
#Change this to your working directory
setwd("/Users/sophiegabreldar/Library/CloudStorage/Box-Box/Sophie\ Gabreldar\'s\
Files/Gabreldar_HonorsThesis2026/R_analysis")


#Load Necessary Libraries
library(psych)      # for general use functions
library(nlme)       # for mixed effects models
library(lme4)       # for mixed effects models

## Loading required package: Matrix

##
## Attaching package: 'lme4'

## The following object is masked from 'package:nlme':
##
##     lmList

library(lmerTest)   # p - values

##
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':
##
##     lmer
```

```
## The following object is masked from 'package:stats':
##
##     step

library(effects)     # calculating and plotting predictions

## Warning: package 'effects' was built under R version 4.5.2

## Loading required package: carData

## lattice theme set by effectsTheme()
## See ?effectsTheme for details.

library(ggeffects)   # plotting predictions

## Warning: package 'ggeffects' was built under R version 4.5.2

library(tidyverse)   # ggplot2, stringr, dplyr, readr, forcats, tidyr, purr

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   4.0.0     ✔ tibble    3.3.0
## ✔ lubridate 1.9.4     ✔ tidyr     1.3.1
## ✔ purrr     1.0.4

##         ──         Conflicts         ──────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ ggplot2::%+%()    masks psych::%+%()
```

```
## ✖ ggplot2::alpha()  masks psych::alpha()
## ✖ dplyr::collapse() masks nlme::collapse()
## ✖ tidyr::expand()   masks Matrix::expand()
## ✖ dplyr::filter()   masks stats::filter()
## ✖ dplyr::lag()      masks stats::lag()
## ✖ tidyr::pack()     masks Matrix::pack()
## ✖ tidyr::unpack()   masks Matrix::unpack()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Data Import

```
merged_data <- read_csv("/Users/sophiegabreldar/Library/CloudStorage/Box-Box/Sophie\ Gabreldar\'s\
Files/Gabreldar_HonorsThesis2026/R_analysis/merged_data.csv")

## New names:
## Rows: 724542 Columns: 53
## ── Column specification
##       ────────────────────────────────────────────────────────────────────────────────
Delimiter: "," chr
## (9): birthcountry, bcountry_clean, ple0008_f, ple0009_f, ple0032_f, ple... dbl
## (44): ...1, pid, syear, pla0045, plj0175, plj0019, plj0662, ple0008, ple...
## ℹ Use `spec()` to retrieve the full column specification for this data. ℹ
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## • `` -> `...1`
```

Data Cleaning

```r
merged_data <- merged_data %>%
 group_by(pid) %>%
 mutate(years_in_survey = syear - min(syear, na.rm = TRUE)) %>%
 ungroup()



#convert negative codes into NAs for modeling
merged_data <- merged_data %>%
 mutate(across(
  where(is.numeric),
  ~ ifelse(. < 0, NA, .)
 ))


#making factored variable for  migration background
merged_data <- merged_data %>%
 mutate(
  migback_f = factor(
   migback,
   levels = c(1, 2, 3),
   labels = c(
    "No Migration Background",
    "Direct Migration",
    "Indirect Migration"
   )
```

```
  )
 )
```

Descriptions of Recoded Variables

```
#life satisfaction
psych::describe(merged_data$plh0182)

##    vars      n mean   sd median trimmed  mad min max range  skew kurtosis se
## X1    1 718666 7.16 1.81      8     7.3 1.48   0  10    10 -0.94     1.18  0

#health satisfaction
psych::describe(merged_data$plh0171)

##    vars      n mean   sd median trimmed  mad min max range  skew kurtosis se
## X1    1 717459 6.77 2.27      7    6.95 2.97   0  10    10 -0.74     0.13  0

#limitations due to health
psych::describe(merged_data$ple0009)

##    vars      n mean   sd median trimmed mad min max range  skew kurtosis se
## X1    1 184799  2.6 0.64      3    2.74   0   1   3     2 -1.37     0.63  0

# physical limitations
psych::describe(merged_data$ple0032)

##    vars      n mean   sd median trimmed  mad min max range  skew kurtosis se
## X1    1 233312 3.99 1.13      4    4.15 1.48   1   5     4 -0.86    -0.24  0
```

```
# emotional limitations

psych::describe(merged_data$ple0033)

##    vars      n mean   sd median trimmed mad min max range  skew kurtosis se

## X1    1 233826 4.24 0.99      5    4.41   0   1   5     4 -1.15     0.45  0
```

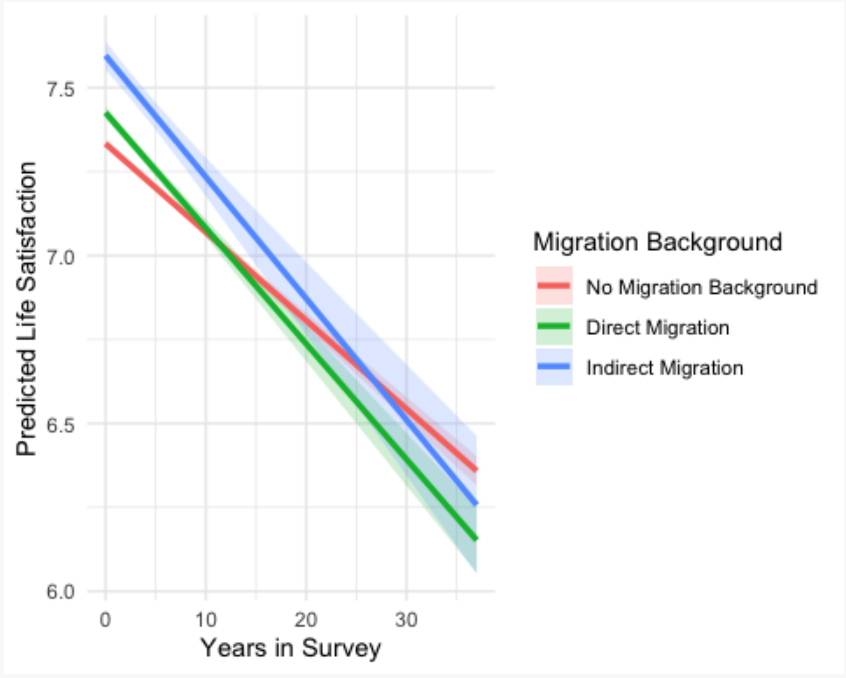
Growth Models for Life Satisfaction

```
#fitting linear growth model and assigning it to an object

lg_lifesat_tic_lmer <- lmer( plh0182 ~ years_in_survey * migback_f + (1 + years_in_survey | pid),

                        data = merged_data,

                        REML = TRUE,

                        na.action = na.exclude)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :

## Model failed to converge with max|grad| = 0.013664 (tol = 0.002, component 1)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable: very large eigenvalue

##  - Rescale variables?

summary(lg_lifesat_tic_lmer)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [

## lmerModLmerTest]

## Formula: plh0182 ~ years_in_survey * migback_f + (1 + years_in_survey |

##     pid)

##    Data: merged_data
```

```
##
## REML criterion at convergence: 2596376
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.0273 -0.4439  0.0958  0.5551  5.6007
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  pid      (Intercept)     1.736693 1.31784
##           years_in_survey 0.006616 0.08134  -0.39
##  Residual                 1.639577 1.28046
## Number of obs: 718666, groups:  pid, 98632
##
## Fixed effects:
##                                             Estimate Std. Error         df
## (Intercept)                                 7.334e+00  6.073e-03  8.897e+04
## years_in_survey                            -2.634e-02  6.469e-04  2.032e+04
## migback_fDirect Migration                   9.241e-02  1.128e-02  1.005e+05
## migback_fIndirect Migration                 2.624e-01  2.236e-02  9.917e+04
## years_in_survey:migback_fDirect Migration  -8.082e-03  1.554e-03  2.648e+04
## years_in_survey:migback_fIndirect Migration -9.843e-03  3.076e-03  3.163e+04
##                                             t value Pr(>|t|)
## (Intercept)                                1207.487  < 2e-16 ***
```

```
## years_in_survey                          -40.722  < 2e-16 ***
## migback_fDirect Migration                  8.196 2.52e-16 ***
## migback_fIndirect Migration               11.734  < 2e-16 ***
## years_in_survey:migback_fDirect Migration  -5.200 2.01e-07 ***
## years_in_survey:migback_fIndirect Migration  -3.200  0.00137 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) yrs_n_ mgb_DM mgb_IM y__:_DM
## yers_n_srvy -0.405
## mgbck_fDrcM -0.539  0.218
## mgbck_fIndM -0.272  0.110  0.146
## yrs_n_s:_DM  0.168 -0.416 -0.369 -0.046
## yrs_n_s:_IM  0.085 -0.210 -0.046 -0.391  0.088
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.013664 (tol = 0.002, component 1)
## Model is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?

# get fitted effects
effects_lifesat <- Effect(
 c("years_in_survey", "migback_f"),
 lg_lifesat_tic_lmer,
```

```
  xlevels = list(years_in_survey = 0:37)
)


# turn into data frame
effectsdata_lifesat <- as.data.frame(effects_lifesat)


# plot with confidence intervals
ggplot(
  data = effectsdata_lifesat,
  aes(x = years_in_survey, y = fit,
      group = migback_f, color = migback_f, fill = migback_f)
) +
  geom_ribbon(aes(ymin = lower, ymax = upper),
              alpha = 0.2, color = NA) +
  geom_line(linewidth = 1.2) +
  theme_minimal() +
  labs(
    x = "Years in Survey",
    y = "Predicted Life Satisfaction",
    color = "Migration Background",
    fill = "Migration Background"
  )
```



Growth Model for Satisfaction With Health with Migration Background

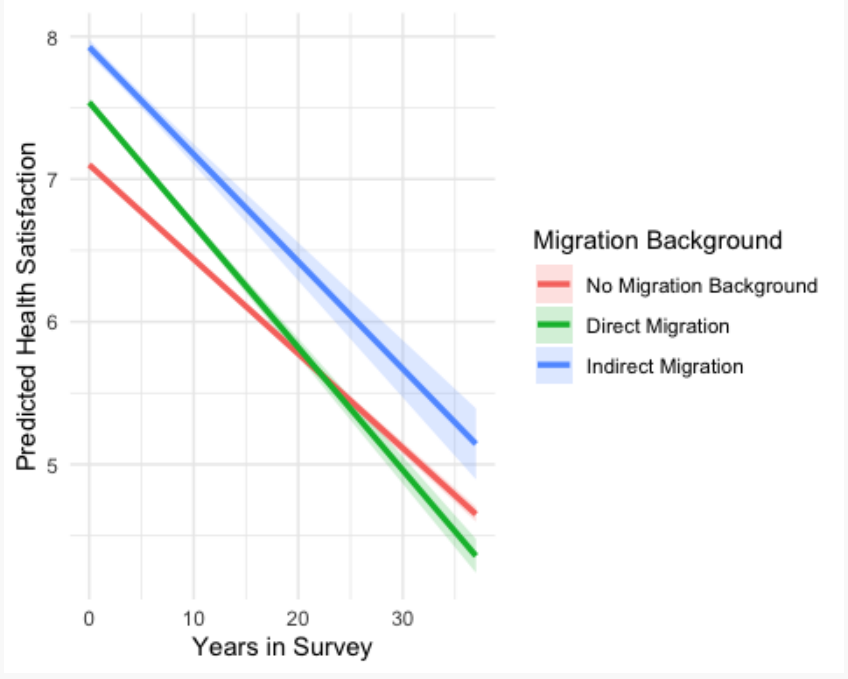
```
#fitting linear growth model and assigning it to an object
lg_healthsat_tic_lmer <- lmer( plh0171 ~ years_in_survey * migback_f + (1 + years_in_survey | pid),
                    data = merged_data,
                    REML = TRUE,
                    na.action = na.exclude)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.607883 (tol = 0.002, component 1)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?
```

```
summary(lg_healthsat_tic_lmer)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: plh0171 ~ years_in_survey * migback_f + (1 + years_in_survey |
##     pid)
##    Data: merged_data
##
## REML criterion at convergence: 2834250
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.0065 -0.4725  0.0958  0.5652  6.5586
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  pid      (Intercept)     2.9638   1.72156
##           years_in_survey 0.0097   0.09849  -0.34
##  Residual                 2.2358   1.49524
## Number of obs: 717459, groups:  pid, 98035
##
## Fixed effects:
##                                      Estimate Std. Error         df
## (Intercept)                          7.101e+00  7.731e-03  8.967e+04
```

```
## years_in_survey                          -6.621e-02  7.793e-04  2.008e+04
## migback_fDirect Migration                 4.383e-01  1.440e-02  9.987e+04
## migback_fIndirect Migration               8.250e-01  2.843e-02  9.829e+04
## years_in_survey:migback_fDirect Migration   -1.975e-02  1.881e-03  2.624e+04
## years_in_survey:migback_fIndirect Migration -8.986e-03  3.702e-03  3.134e+04
##                                          t value Pr(>|t|)
## (Intercept)                              918.393   <2e-16 ***
## years_in_survey                          -84.964   <2e-16 ***
## migback_fDirect Migration                 30.446   <2e-16 ***
## migback_fIndirect Migration               29.014   <2e-16 ***
## years_in_survey:migback_fDirect Migration   -10.497   <2e-16 ***
## years_in_survey:migback_fIndirect Migration  -2.427   0.0152 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) yrs_n_ mgb_DM mgb_IM y__:_DM
## yers_n_srvy -0.365
## mgbck_fDrcM -0.537  0.196
## mgbck_fIndM -0.272  0.099  0.146
## yrs_n_s:_DM  0.151 -0.414 -0.338 -0.041
## yrs_n_s:_IM  0.077 -0.210 -0.041 -0.356  0.087
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.607883 (tol = 0.002, component 1)
```

```
## Model is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?

# get fitted effects
effects_healthsat <- Effect(
  c("years_in_survey", "migback_f"),
  lg_healthsat_tic_lmer,
  xlevels = list(years_in_survey = 0:37)
)


# turn into data frame
effectsdata_healthsat <- as.data.frame(effects_healthsat)


# plot with confidence ribbons
ggplot(
  data = effectsdata_healthsat,
  aes(x = years_in_survey, y = fit,
      group = migback_f, color = migback_f, fill = migback_f)
) +
  geom_ribbon(aes(ymin = lower, ymax = upper),
              alpha = 0.2, color = NA) +
  geom_line(linewidth = 1.2) +
  theme_minimal() +
  labs(
```

```
  x = "Years in Survey",
  y = "Predicted Health Satisfaction",
  color = "Migration Background",
  fill = "Migration Background"
)
```



Growth Models For Limitations in Daily Life Because of Health with Migration Background

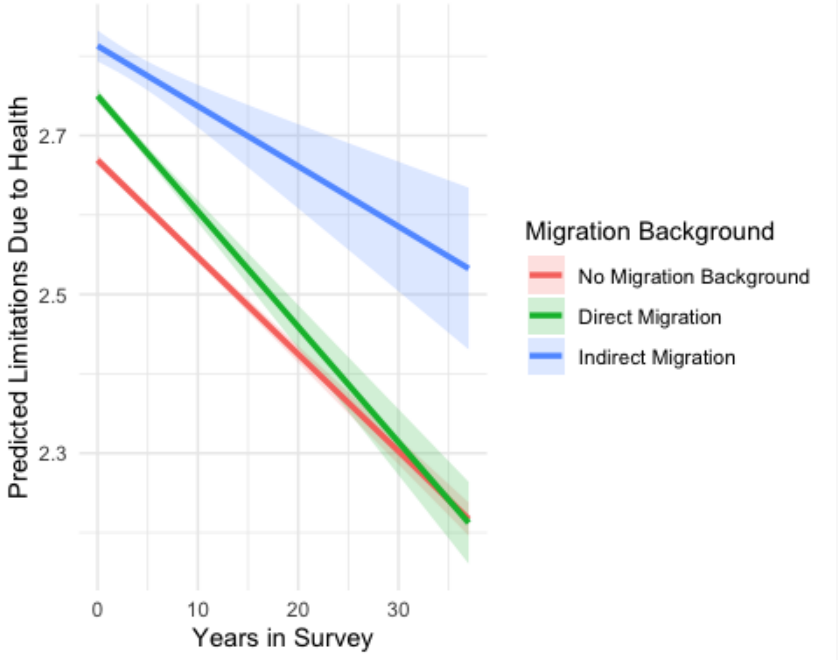
```
#fitting linear growth model and assigning it to an object
lg_health_limits_tic_lmer <- lmer( ple0009 ~ years_in_survey * migback_f + (1 + years_in_survey | pid),
                                   data = merged_data,
```

```
                    REML = TRUE,
                    na.action = na.exclude)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0416796 (tol = 0.002, component 1)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?

summary(lg_health_limits_tic_lmer)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: ple0009 ~ years_in_survey * migback_f + (1 + years_in_survey |
##     pid)
##    Data: merged_data
##
## REML criterion at convergence: 288422.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.1233 -0.2202  0.1643  0.3855  4.1653
##
## Random effects:
##  Groups   Name            Variance  Std.Dev. Corr
```

```
##  pid      (Intercept)     0.1841446 0.42912
##           years_in_survey 0.0004278 0.02068  -0.04
##  Residual                 0.1733214 0.41632
## Number of obs: 184799, groups:  pid, 60038
##
## Fixed effects:
##                                          Estimate Std. Error         df
## (Intercept)                              2.669e+00  3.356e-03  5.415e+04
## years_in_survey                          -1.222e-02  3.266e-04  1.479e+04
## migback_fDirect Migration                8.098e-02  5.346e-03  5.305e+04
## migback_fIndirect Migration              1.439e-01  1.047e-02  5.402e+04
## years_in_survey:migback_fDirect Migration   -2.316e-03  8.152e-04  2.026e+04
## years_in_survey:migback_fIndirect Migration  4.637e-03  1.543e-03  2.456e+04
##                                          t value Pr(>|t|)
## (Intercept)                              795.222  < 2e-16 ***
## years_in_survey                          -37.405  < 2e-16 ***
## migback_fDirect Migration                15.149  < 2e-16 ***
## migback_fIndirect Migration              13.749  < 2e-16 ***
## years_in_survey:migback_fDirect Migration    -2.841  0.00451 **
## years_in_survey:migback_fIndirect Migration   3.005  0.00266 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##             (Intr) yrs_n_ mgb_DM mgb_IM y__:_DM
## yers_n_srvy -0.553
## mgbck_fDrcM -0.628  0.347
## mgbck_fIndM -0.321  0.177  0.201
## yrs_n_s:_DM  0.222 -0.401 -0.421 -0.071
## yrs_n_s:_IM  0.117 -0.212 -0.073 -0.467  0.085
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.0416796 (tol = 0.002, component 1)
## Model is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?

# get fitted effects
effects_health_limits <- Effect(
 c("years_in_survey", "migback_f"),
 lg_health_limits_tic_lmer,
 xlevels = list(years_in_survey = 0:37)
)


# turn into data frame
effectsdata_health_limits <- as.data.frame(effects_health_limits)


# plot with confidence ribbons
ggplot(
 data = effectsdata_health_limits,
```

```
  aes(x = years_in_survey, y = fit,
      group = migback_f, color = migback_f, fill = migback_f)
) +
  geom_ribbon(aes(ymin = lower, ymax = upper),
              alpha = 0.2, color = NA) +
  geom_line(linewidth = 1.2) +
  theme_minimal() +
  labs(
    x = "Years in Survey",
    y = "Predicted Limitations Due to Health",
    color = "Migration Background",
    fill = "Migration Background"
  )
```



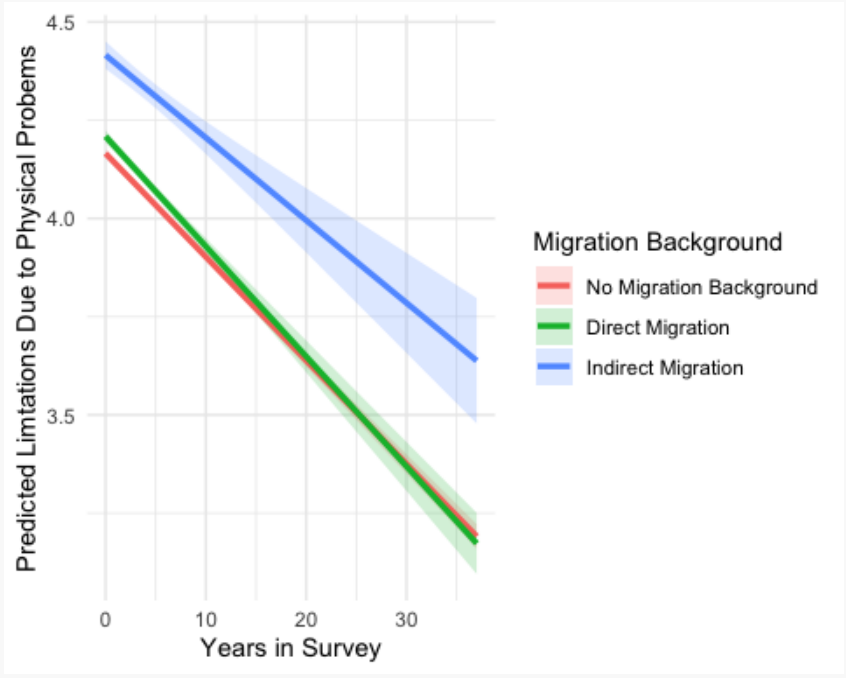
Growth Models Limitations Due to Physical Problems with Migration Background

```
#fitting linear growth model and assigning it to an object
lg_physical_limits_tic_lmer <- lmer( ple0032 ~ years_in_survey * migback_f + (1 + years_in_survey |
pid),
                                     data = merged_data,
                                     REML = TRUE,
                                     na.action = na.exclude)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0219472 (tol = 0.002, component 1)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly
unidentifiable: very large eigenvalue
##  - Rescale variables?

summary(lg_physical_limits_tic_lmer)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: ple0032 ~ years_in_survey * migback_f + (1 + years_in_survey |
##     pid)
##    Data: merged_data
##
## REML criterion at convergence: 647704.4
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -4.4662 -0.5339  0.2063  0.5443  4.2444
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  pid      (Intercept)     0.587192 0.76628
##           years_in_survey 0.001363 0.03692  -0.23
##  Residual                 0.625842 0.79110
## Number of obs: 233312, groups:  pid, 70792
##
## Fixed effects:
##                                          Estimate Std. Error         df
## (Intercept)                              4.165e+00  5.163e-03  6.203e+04
## years_in_survey                         -2.634e-02  4.888e-04  1.664e+04
## migback_fDirect Migration                4.328e-02  9.121e-03  6.525e+04
## migback_fIndirect Migration              2.501e-01  1.855e-02  6.866e+04
## years_in_survey:migback_fDirect Migration   -1.653e-03  1.246e-03  2.181e+04
## years_in_survey:migback_fIndirect Migration  5.322e-03  2.466e-03  2.695e+04
##                                         t value Pr(>|t|)
## (Intercept)                             806.838  < 2e-16 ***
## years_in_survey                         -53.893  < 2e-16 ***
## migback_fDirect Migration                 4.745 2.09e-06 ***
## migback_fIndirect Migration              13.485  < 2e-16 ***
## years_in_survey:migback_fDirect Migration    -1.327   0.1845
```

```
## years_in_survey:migback_fIndirect Migration   2.158   0.0309 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) yrs_n_ mgb_DM mgb_IM y__:_DM
## yers_n_srvy -0.548
## mgbck_fDrcM -0.566  0.310
## mgbck_fIndM -0.278  0.152  0.158
## yrs_n_s:_DM  0.215 -0.392 -0.467 -0.060
## yrs_n_s:_IM  0.109 -0.198 -0.061 -0.528  0.078
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.0219472 (tol = 0.002, component 1)
## Model is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?

# get fitted effects
effects_physical_limits <- Effect(
 c("years_in_survey", "migback_f"),
 lg_physical_limits_tic_lmer,
 xlevels = list(years_in_survey = 0:37)
)
```

```r
# turn into data frame
effectsdata_physical_limits <- as.data.frame(effects_physical_limits)


# plot with confidence ribbons
ggplot(
  data = effectsdata_physical_limits,
  aes(x = years_in_survey, y = fit,
      group = migback_f, color = migback_f, fill = migback_f)
) +
  geom_ribbon(aes(ymin = lower, ymax = upper),
              alpha = 0.2, color = NA) +
  geom_line(linewidth = 1.2) +
  theme_minimal() +
  labs(
    x = "Years in Survey",
    y = "Predicted Limtations Due to Physical Probems",
    color = "Migration Background",
    fill = "Migration Background"
  )
```



Growth Models: Limitations Due to Emotional Problems with Migration Background

```
#fitting linear growth model and assigning it to an object
lg_emotional_limits_tic_lmer <- lmer(ple0033 ~ years_in_survey * migback_f + (1 + years_in_survey |
pid),
                                     data = merged_data,
                                     REML = TRUE,
                                     na.action = na.exclude)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0173464 (tol = 0.002, component 1)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly
unidentifiable: very large eigenvalue
##  - Rescale variables?
```

```
summary(lg_emotional_limits_tic_lmer)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: ple0033 ~ years_in_survey * migback_f + (1 + years_in_survey |
##     pid)
##    Data: merged_data
##
## REML criterion at convergence: 611215.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6803 -0.5080  0.2510  0.5282  3.5575
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  pid      (Intercept)     0.428615 0.65469
##           years_in_survey 0.001098 0.03314  -0.37
##  Residual                 0.562759 0.75017
## Number of obs: 233826, groups:  pid, 70773
##
## Fixed effects:
##                                      Estimate Std. Error         df
## (Intercept)                          4.347e+00  4.605e-03  6.083e+04
```

```
## years_in_survey                          -1.290e-02  4.331e-04  1.604e+04
## migback_fDirect Migration                -1.070e-01  8.145e-03  6.421e+04
## migback_fIndirect Migration               7.719e-02  1.658e-02  6.759e+04
## years_in_survey:migback_fDirect Migration    5.941e-03  1.083e-03  2.000e+04
## years_in_survey:migback_fIndirect Migration -1.170e-04  2.177e-03  2.547e+04
##                                         t value Pr(>|t|)
## (Intercept)                             944.044  < 2e-16 ***
## years_in_survey                         -29.787  < 2e-16 ***
## migback_fDirect Migration               -13.138  < 2e-16 ***
## migback_fIndirect Migration               4.656 3.22e-06 ***
## years_in_survey:migback_fDirect Migration     5.484 4.21e-08 ***
## years_in_survey:migback_fIndirect Migration  -0.054    0.957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) yrs_n_ mgb_DM mgb_IM y__:_DM
## yers_n_srvy -0.596
## mgbck_fDrcM -0.565  0.337
## mgbck_fIndM -0.278  0.166  0.157
## yrs_n_s:_DM  0.238 -0.400 -0.502 -0.066
## yrs_n_s:_IM  0.119 -0.199 -0.067 -0.565  0.080
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.0173464 (tol = 0.002, component 1)
```

```
## Model is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?

# get fitted effects
effects_emotional_limits <- Effect(
  c("years_in_survey", "migback_f"),
  lg_emotional_limits_tic_lmer,
  xlevels = list(years_in_survey = 0:37)
)


# turn into data frame
effectsdata_emotional_limits <- as.data.frame(effects_emotional_limits)


# plot with confidence ribbons
ggplot(
  data = effectsdata_emotional_limits,
  aes(x = years_in_survey, y = fit,
      group = migback_f, color = migback_f, fill = migback_f)
) +
  geom_ribbon(aes(ymin = lower, ymax = upper),
              alpha = 0.2, color = NA) +
  geom_line(linewidth = 1.2) +
  theme_minimal() +
  labs(
```

```
  x = "Years in Survey",
  y = "Predicted Limtations Due to Emotional Problems",
  color = "Migration Background",
  fill = "Migration Background"
)
```