

Happiness and Commuting (SOEP)

Anastasiia Apatenko (matriculation number: 143732)

2024-08-30

Contents

Introduction	2
SOEP data	2
Definition of Key Terms	2
Literature Review	2
Data sets and Variables	3
Hypotheses	4
Demographics Analysis	4
Average Age by Year	4
Gender Distribution	6
Marital Status Distribution	7
Household and individual income	8
Average monthly household income	8
Average Income Per Person in household	9
Education and Employment	10
Average Years of Education by Year	10
Average Years of Education by Year and Education Level	11
Education and employment status	13
Life satisfaction and employment status	14
Commuting	15
Mean Commuting Distance and Mean Commuting Time	15
Commuting Distance vs. Time by Frequency of Commuting and Year	16
Interactive Scatter Plot of Commuting Time vs. Life Satisfaction by Year	17

Income and life satisfaction	18
Correlation between household income and life satisfaction	18
Correlation between Overall Life Satisfaction and Satisfaction With Personal Income by Year . . .	18
Conclusion	20
Appendix	21
References	22

Introduction

Understanding the relationship between happiness and commuting is a significant area of interest in social science research. Commuting, defined as the daily travel time to and from work, is an integral part of many individuals' lives, particularly in urban settings. Previous studies have highlighted that commuting time can impact various dimensions of well-being, including physical health, mental health, and overall life satisfaction. As cities grow and commuting times lengthen, investigating how commuting affects happiness becomes increasingly relevant.

SOEP data

The Socio-Economic Panel (*SOEP*) is a longitudinal survey that has been conducted annually since 1984 by the German Institute for Economic Research (DIW Berlin). It provides a rich, multi-dimensional dataset that includes detailed information on households and individuals in Germany, covering aspects such as income, employment, education, health, and life satisfaction. The SOEP is one of the largest and most comprehensive panel datasets available, allowing researchers to analyze the dynamic aspects of socio-economic conditions and behaviors over time. It is a valuable resource for studying trends and patterns within the German population, providing insights into how economic, social, and personal factors interact to shape individuals' lives and societal outcomes.

Definition of Key Terms

In this study, ***Happiness*** is defined as the self-reported life satisfaction score, as measured in the Socio-Economic Panel (SOEP) dataset. ***Commuting*** refers to the total time an individual spends traveling from home to work and back, typically measured in minutes per day. ***Life satisfaction*** is a subjective assessment of one's overall well-being and contentment with life. Life satisfaction was chosen as the measure of happiness because no other more suitable variables were available in the data sets accessible to us to describe the level of happiness.

Literature Review

The existing literature indicates that commuting has significant implications for an individual's subjective well-being, with various studies revealing both direct and indirect effects on happiness. Stutzer and Frey (Stutzer & Frey, 2008) describe the "commuting paradox," where longer commuting times, despite potential financial or employment benefits, are consistently linked to lower levels of life satisfaction. This paradox underscores the psychological stress and reduced time for personal and social activities that longer commutes often entail. Kahneman et al. (Kahneman et al., 2004) further emphasize the negative impact of commuting on daily well-being by using the Day Reconstruction Method (DRM) to categorize commuting as one of the

least enjoyable daily activities. Their research shows that commuting significantly contributes to daily stress and overall dissatisfaction with life, underscoring its role in shaping daily experiences.

Meier and Stutzer (Meier & Stutzer, 2008) contribute to this discussion by exploring the broader context of time use and its influence on happiness. Although their study primarily investigates the benefits of volunteering, they note that time spent commuting can reduce opportunities for engaging in fulfilling activities, such as socializing or volunteering, which are crucial for maintaining happiness. This aligns with findings from Petrunoff et al. (Petrunoff et al., 2017), who highlight the benefits of active commuting modes, such as walking or cycling, particularly in work settings. Their systematic review suggests that interventions encouraging active commuting can mitigate some of the negative impacts associated with longer or more stressful commutes, thus enhancing overall life satisfaction and reducing stress levels.

Clark et al. (Clark et al., 2019) delve deeper into the nuances of how different commuting modes affect subjective well-being. They found that individuals commuting by car tend to report lower happiness levels compared to those who walk or cycle, potentially due to the associated physical activity and exposure to more pleasant environments. Moreover, the study underscores the importance of perceived control over one's commuting method in influencing its impact on life satisfaction. This suggests that the ability to choose one's mode of transport, and thereby potentially reduce commute-related stress, plays a significant role in determining the overall well-being of commuters.

Despite the extensive body of research on commuting and happiness, there remains a gap in understanding how various socio-economic and demographic factors, such as income, employment status, and geographic region, interact with commuting to affect happiness levels.

Data sets and Variables

We had access to six datasets: `gripstr.dta`, `hgen.dta`, `hl.dta`, `pequiv.dta`, `pgen.dta`, and `pl.dta`. After thoroughly examining each dataset, I selected four that contained the necessary information for my analysis: **`hgen.dta`**, **`pequiv.dta`**, **`pgen.dta`**, and **`pl.dta`**. The datasets `gripstr.dta` and `hl.dta` were not included in the final analysis, as they did not contain the relevant variables for my research.

The chosen datasets provided comprehensive data on employment status, household characteristics, life satisfaction, and educational background, which were crucial for understanding the relationships between life satisfaction and commuting behavior. However, it's important to note that some variables were not available due to restricted access. For instance, the residence information variable (`l11101`) in the `pequiv.dta` was inaccessible, which limited the analysis of certain aspects related to geographic and residential factors.

Variables included:

1. Age
2. Gender
3. Marital Status
4. Number of persons in household
5. Number of years of education
6. Education with respect to High School
7. Household income satisfaction
8. Commuting distance
9. Commuting Frequency
10. Commuting Time (Minutes)

11. Monthly household income

12. Employment Status

Each variable in the datasets is available for different time periods. Some cover the whole duration of the SOEP study, from 1984 to 2020, while others are only available for specific years. As a result, I used the primary variables for the entire study period, but some detailed indicators are only available for shorter times.

To handle the large amount of data, I also shortened the research period for some analyses. This made the data more manageable and helped focus on specific aspects of the study.

Overall, this approach allowed me to analyze life satisfaction and commuting behavior in a clear and organized way, despite the challenges with data availability and size.

Hypotheses

Based on the literature review, I hypothesize that longer commuting times are associated with lower levels of happiness. Additionally, I claim that this relationship may vary depending on socio-economic factors such as income, employment status, and education.

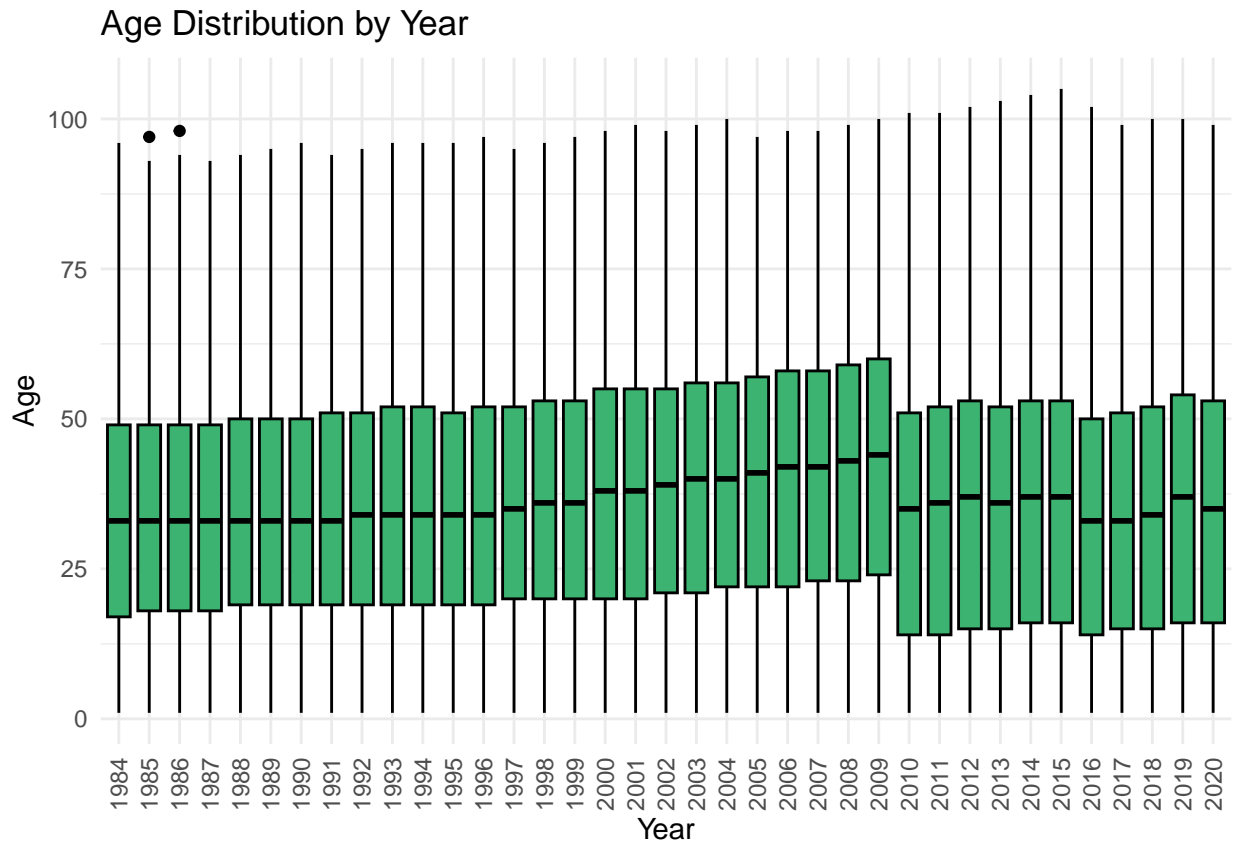
Demographics Analysis

To begin, we will examine the key demographic variables of SOEP data: *age*, *gender*, and *marital status*. All of these variables are contained in the *pequiv.dta* database. These demographic factors are crucial as they provide foundational context for understanding the characteristics of the population under study. Analyzing these variables will help us identify patterns and correlations that may influence other aspects of the study, such as happiness and commuting time. Understanding the distribution and characteristics of these demographics allows us to better interpret the broader social and economic dynamics at play.

Average Age by Year

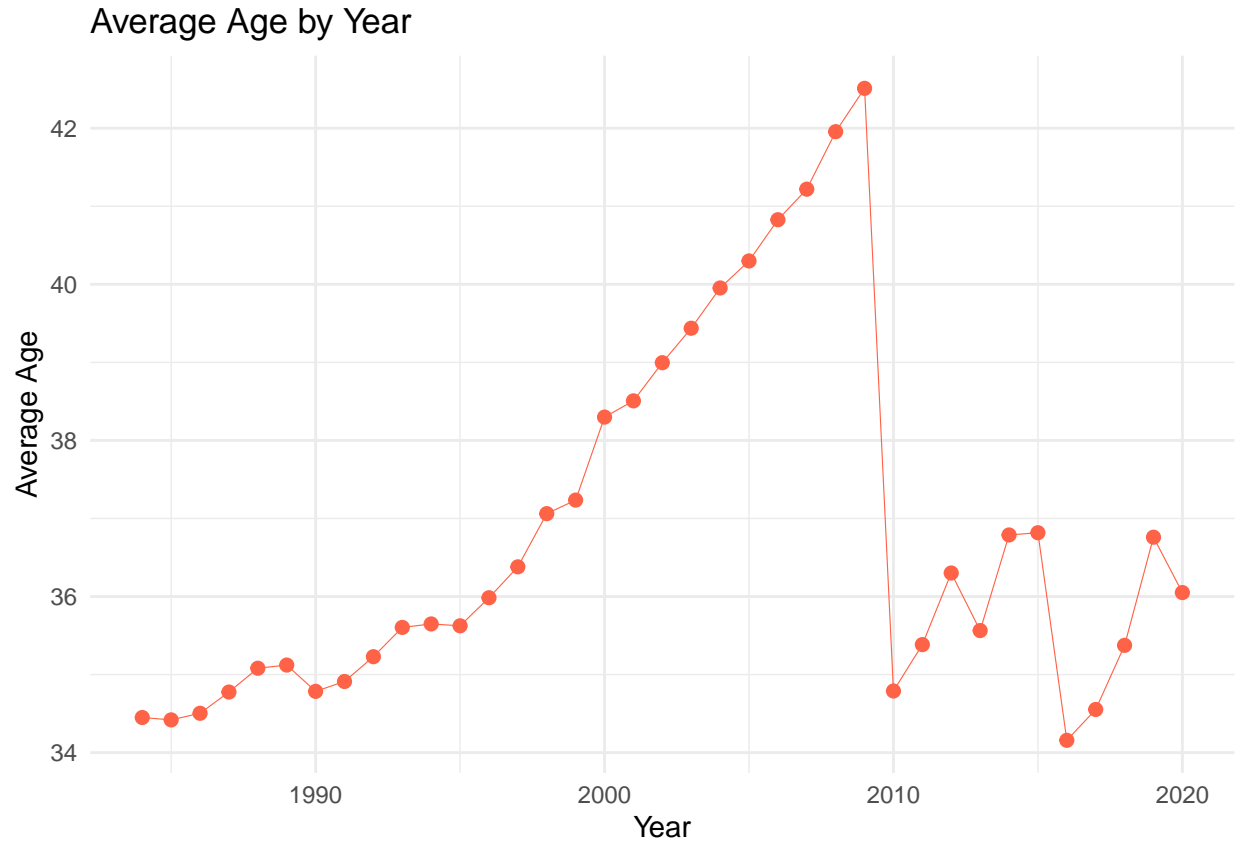
Age Distribution by Year

This box plot effectively illustrates the distribution of the ages of respondents from 1984 to 2020. It provides a clear view of the *minimum* and *maximum* ages for each year, allowing for an easy identification of the age range over time. Additionally, the plot highlights the *median* age and showcases the variability in ages across different years.



Average Age by Year

The graph shows fluctuations in the *average* age over the years. The line illustrates the general direction or trend of the average age, whether it is increasing, decreasing, or remaining stable.

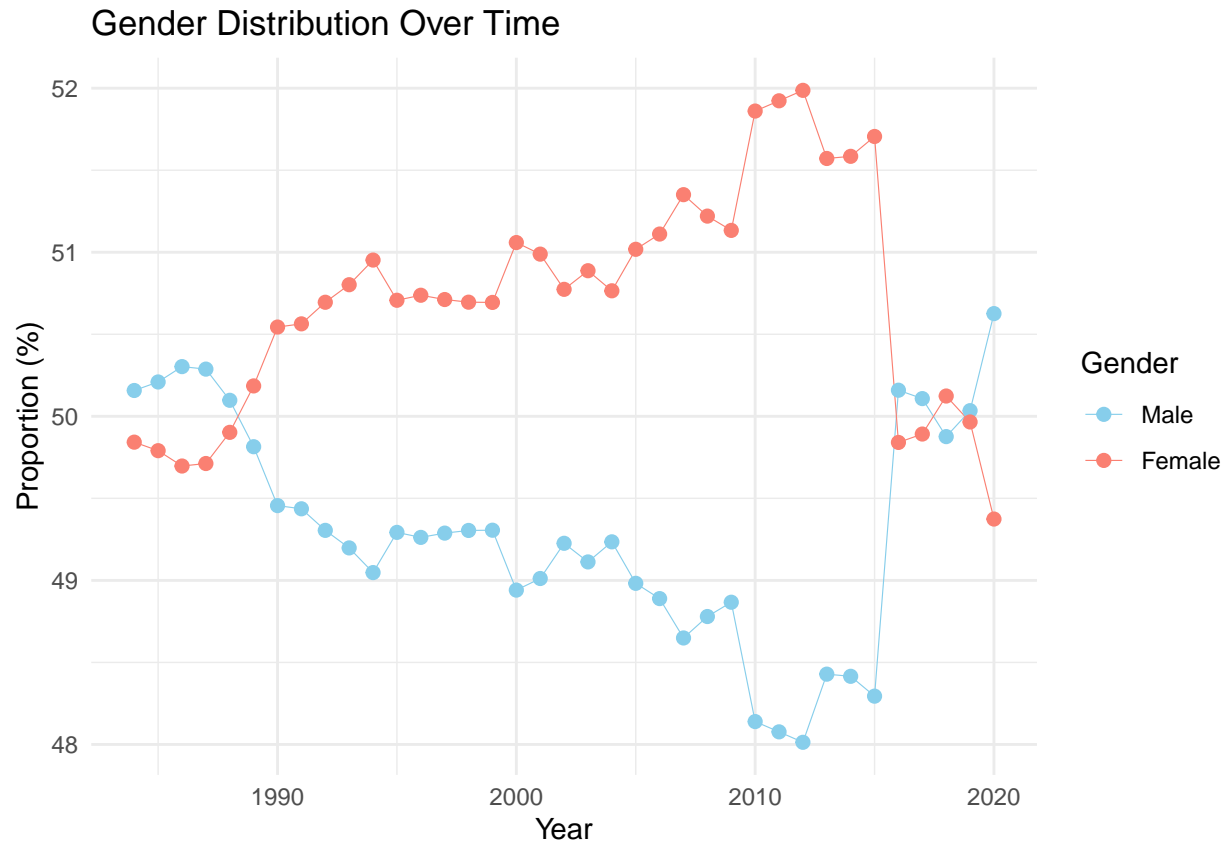


Starting from 1990, the plot reveals a gradual increase in the mean age of respondents up until nearly the end of the 2010s. Notably, there is a significant drop in median age in 2010 compared to 2009, with the median age falling from approximately 42 to 35. Following this sharp decline, the median age has not risen above 36 in subsequent years, indicating a sustained trend of younger respondents in the more recent data. This shift may reflect broader demographic changes or shifts in the survey population over time.

Gender Distribution

The histogram illustrates the gender distribution over time from 1984 to 2020.

- The gender proportions exhibit fluctuations over the years, with noticeable variations in the ratio of males to females across different periods.
- Significant shifts can be seen at certain points in time, such as the early 1990s and around 2000, reflecting changes in gender representation within the dataset.
- Recent years show a more balanced gender distribution, with proportions converging towards equal representation.

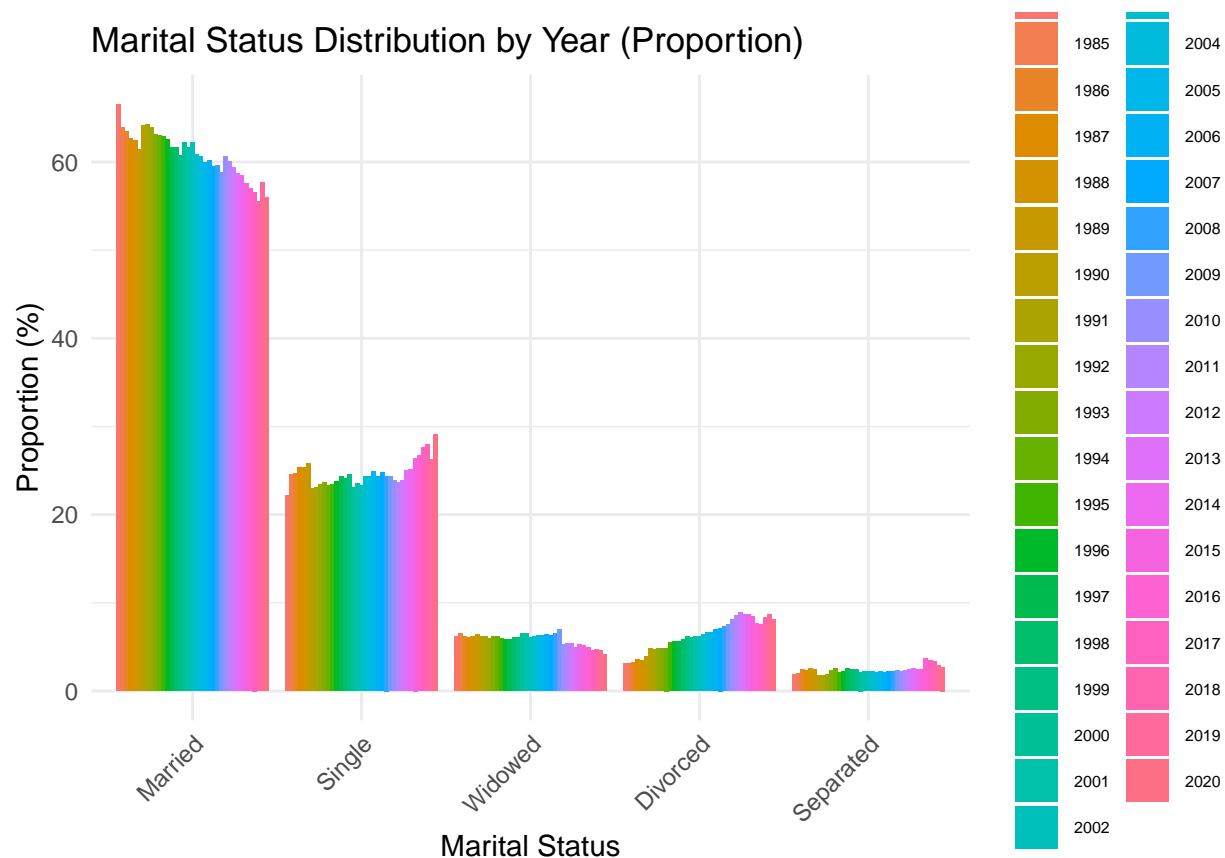


Marital Status Distribution

The bar plot shows the distribution of marital statuses across different years. Each color represents a different year, and the height of the bars represents the proportion of respondents in each marital status category.

- The majority of respondents are married, with a consistent proportion of over 60% across the years.
- The proportion of single individuals is the second highest, followed by widowed, divorced, and separated individuals.
- We can also observe that since the mid-2000s, the popularity of divorces has increased, while the number of marriages has significantly declined.
- Similarly, the number of single individuals began to rise sharply around 2010.

These trends may point to evolving cultural, social, and economic factors influencing people's choices regarding marriage and relationships.



Household and individual income

I analyzed household income data to explore how average monthly household income varies with the number of household members and how income per person in a household changes over the years. These analyses provide insights into income distribution patterns and their potential impact on commuting behaviors and overall happiness.

Average monthly household income

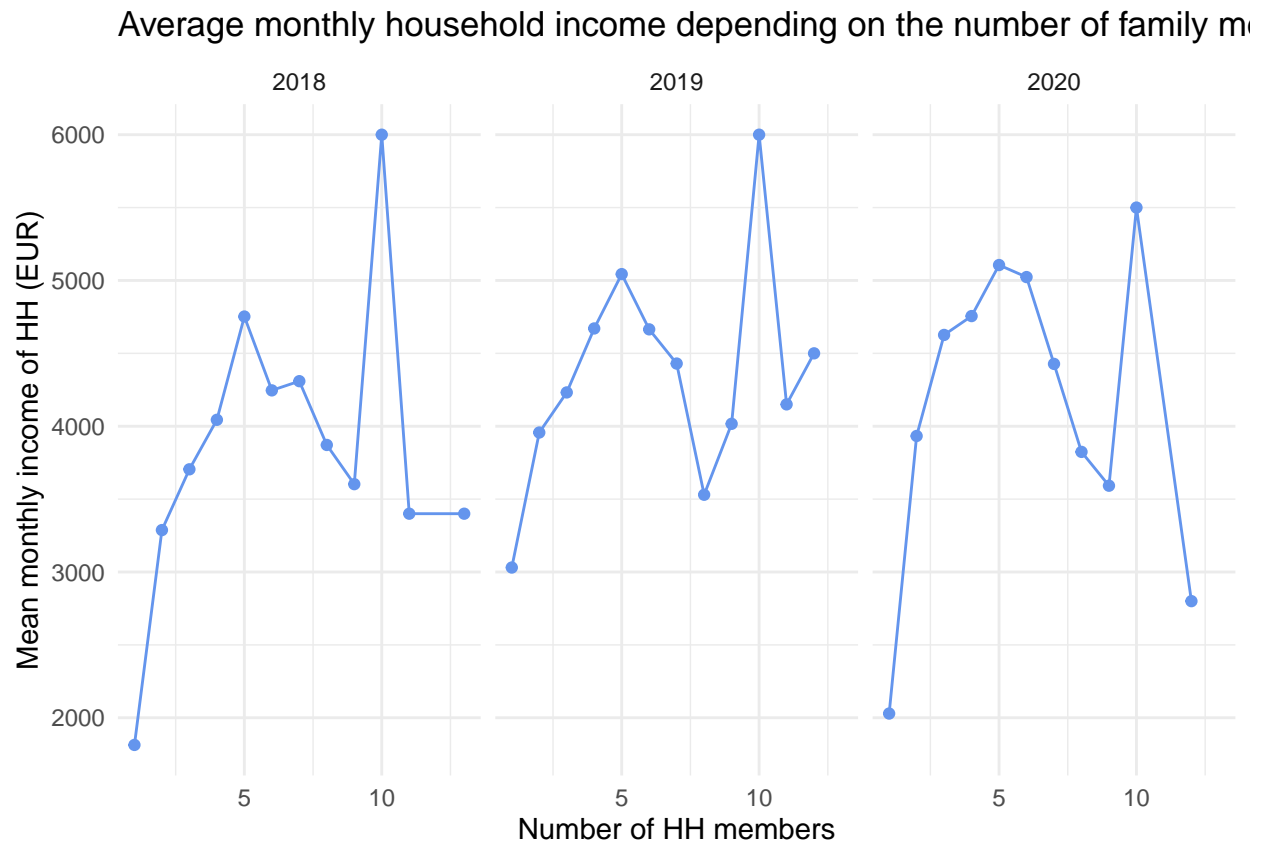
Here I wanted to merge and clean household income and size data, then visualize how average monthly household income varies with the number of household members across the years. But the data was so vast and there were more than 1 million variables. So I decided to choose only years from 2018 to 2020. The plot helps to identify trends and patterns in income distribution relative to household size during these years. The facet wrap allows for a comparative view of income trends across the specified years, helping to identify any significant changes or consistent patterns.

Main observations:

The analysis of the average monthly household income in relation to the number of household members from 2018 to 2020 reveals several interesting trends. The data shows that household income increases as the number of family members grows, but this trend only holds up to a certain point. Specifically, the average income rises steadily with household size until it reaches five members. Beyond this threshold, the income begins to decline gradually, suggesting that additional family members may lead to diminishing financial returns, possibly due to increased living costs or resource allocation challenges.

However, the trend takes a surprising turn when the number of household members exceeds nine. At this point, there is a sharp increase in average household income, which could indicate the presence of additional income earners in larger households or economies of scale that benefit very large families. Following this unexpected spike, the income drops sharply again, highlighting potential volatility and financial strain in extremely large households.

These findings suggest that while larger households may benefit from multiple income sources, they also face increased financial pressures that can impact their overall economic stability.



Average Income Per Person in household

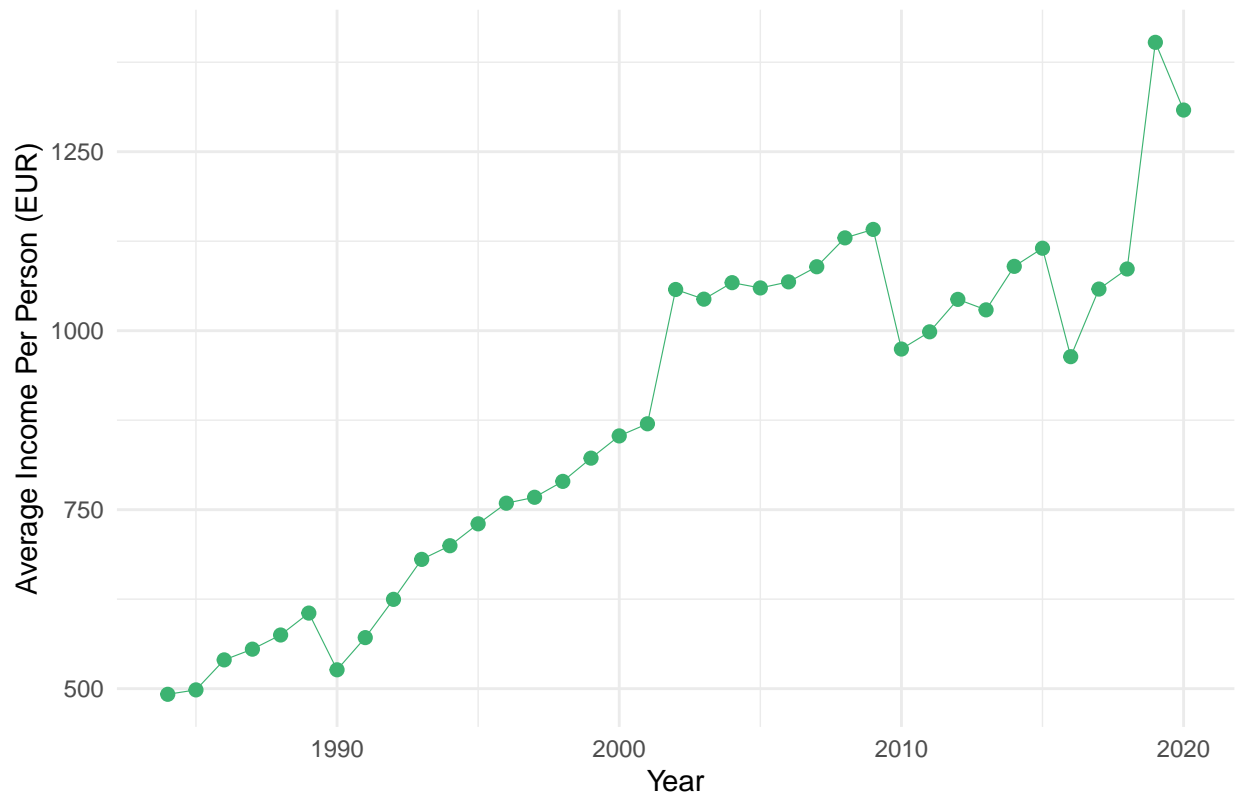
To further analyze household income dynamics, I calculated the income per household member to understand the average income per individual in each household. This calculation was necessary because the available dataset did not provide a variable for monthly individual income. By deriving this metric, I could more accurately assess the financial situation on a per-person basis within households.

Main observations:

The resulting plot displays the trend of average income per person from 2018 to 2020. The graph shows a gradual increase in the average income per person over the years, indicating an overall improvement in individual economic conditions within households. This upward trend suggests that household incomes are rising in a way that benefits each member more significantly over time.

A particularly notable observation from the graph is the marked increase in per capita income in 2001. This significant jump may reflect broader economic changes, policy impacts, or shifts in the labor market that particularly benefitted households during that year.

Average Income Per Person in HH by Year



Education and Employment

Education generally contributes to higher happiness through improved financial stability, better health, and enhanced social and psychological benefits. It often leads to greater job satisfaction and a sense of achievement.

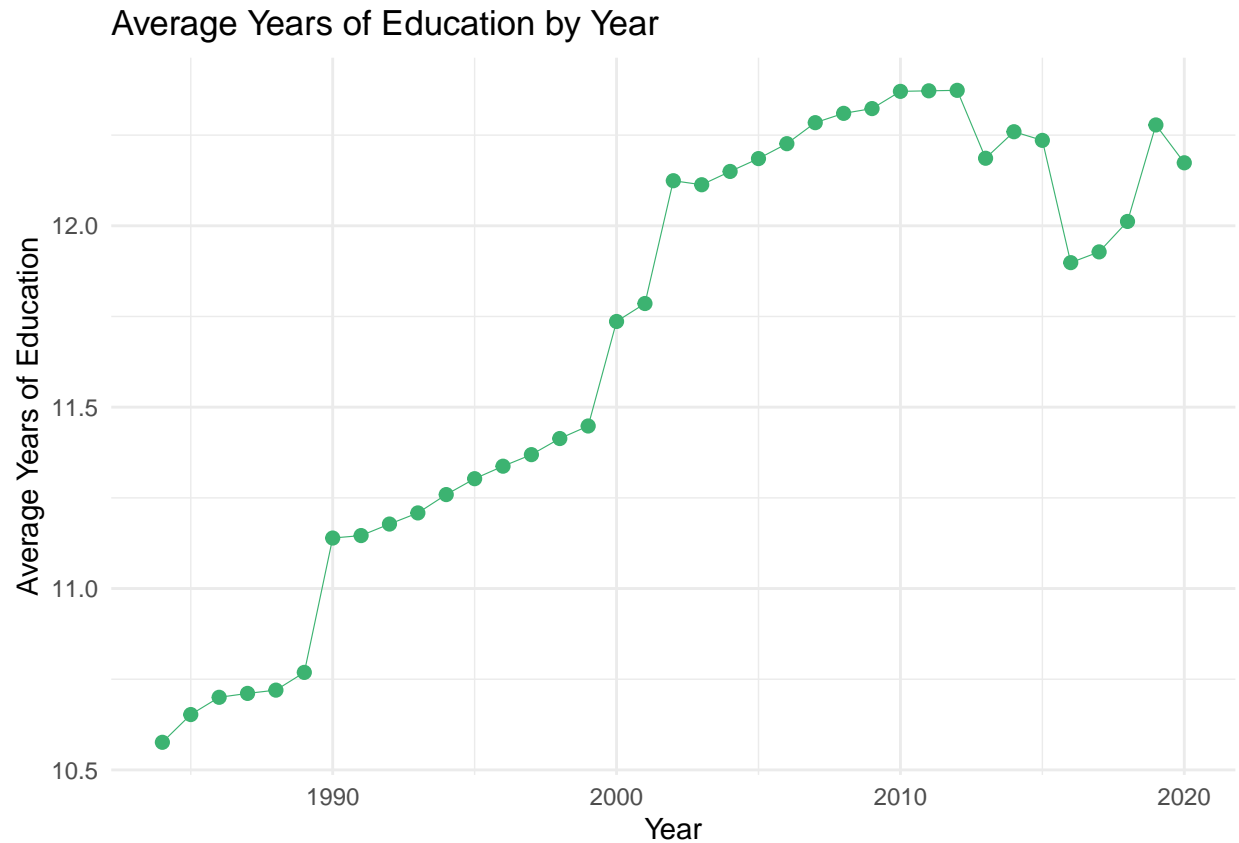
Education plays a crucial role in shaping an individual's life opportunities and overall well-being. This part explores the relationship between educational attainment and employment status, as well as their combined impact on life satisfaction.

Average Years of Education by Year

The graph illustrates the trend in average years of education from 1984 to 2020.

Main observations:

The graph shows a clear upward trend in the average years of education from 1980 to around 2010, followed by fluctuations in the subsequent years. This indicates a general increase in the number of years individuals spend in education over the 40-year period.



Average Years of Education by Year and Education Level

The faceted plot highlights distinct trends for each educational level, emphasizing the different trajectories and growth patterns.

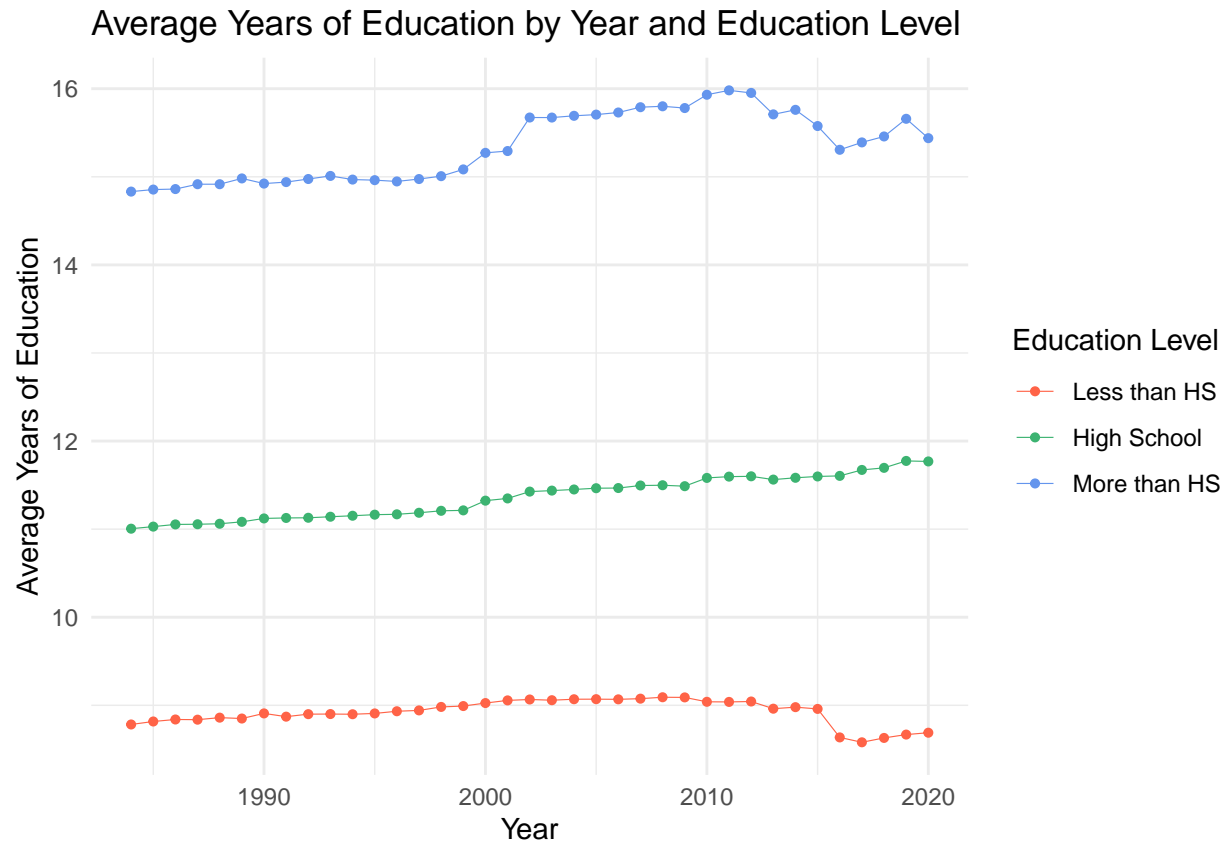
Main observations:

For “*Less than HS*,” there is a noticeable increase followed by stabilization, indicating efforts to enhance basic education access.

The “*High School*” category shows consistent growth, reflecting increased retention and completion rates.

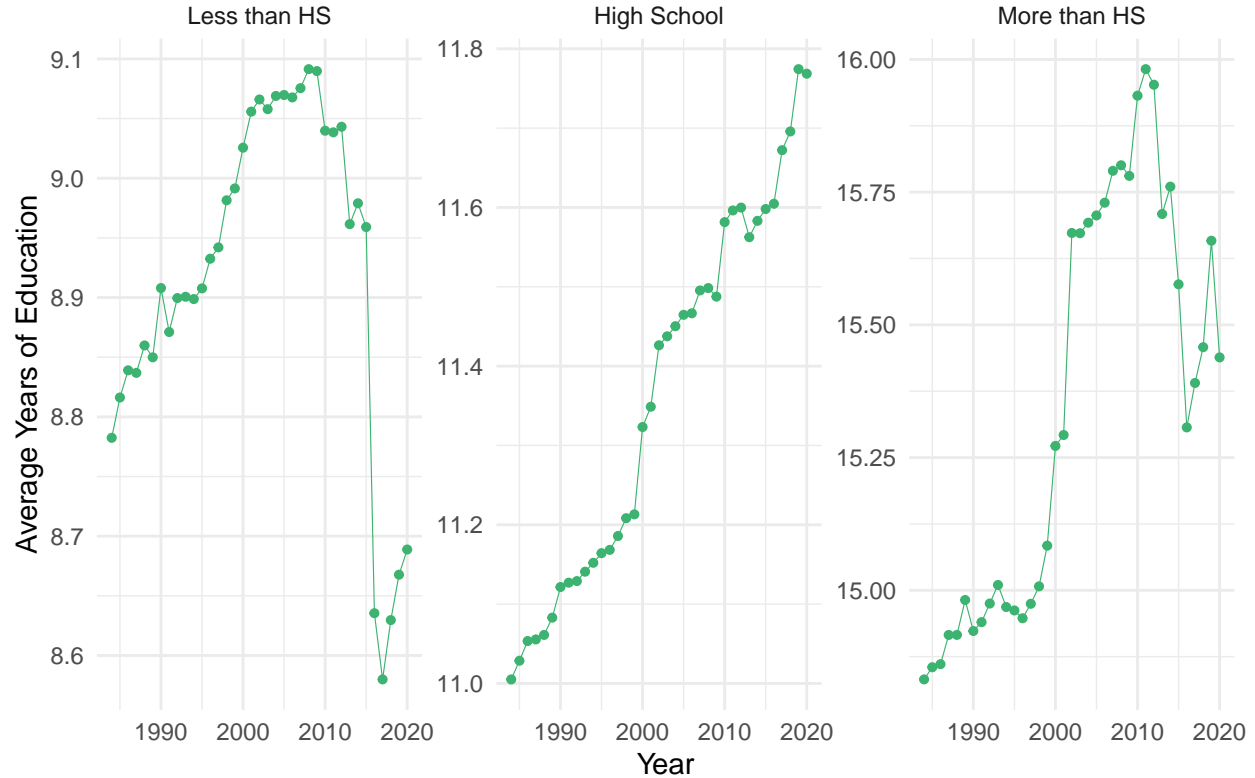
The “*More than HS*” category shows a pronounced upward trend, emphasizing the growing demand for advanced education and professional qualifications.

The gap between the average years of education for those with “More than HS” and those with only “High School” education has widened over time. This suggests increasing stratification based on educational attainment, which could have implications for income inequality, employment opportunities, and social mobility.



This is a *faceted plot*, which divides the data into three panels, each corresponding to one of the educational levels, allowing for a detailed examination of trends within each category over the specified years.

Average Years of Education by Year and Education Level

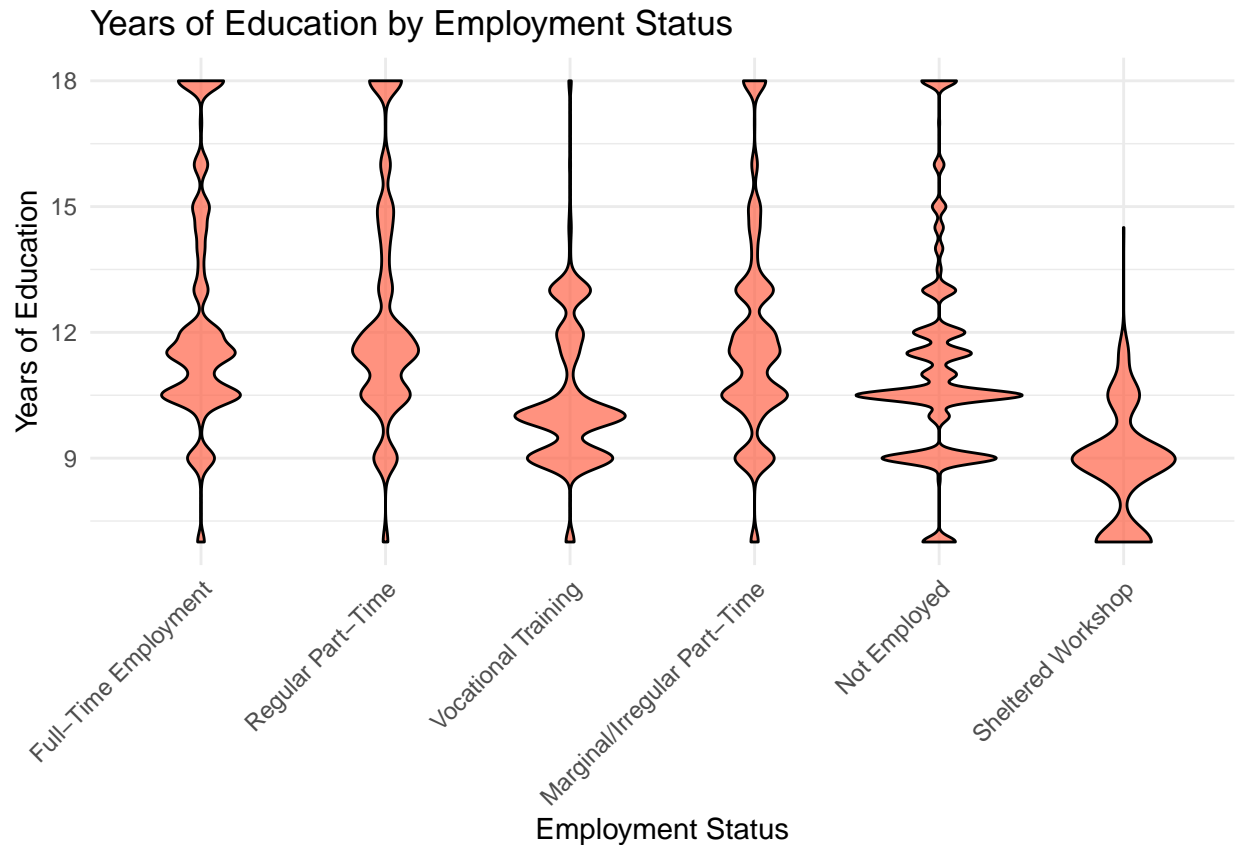


Education and employment status

This is a violin plot showing the distribution of years of education across different employment statuses, considered in *pgen.dta*.

Main observations:

1. *Full-Time Employment:* Individuals with full-time employment generally have a higher level of education. The plot is more concentrated around higher years of education, with a notable peak at around 15 years, indicating that many individuals in this category have some college education or higher.
2. *Regular Part-Time and Vocational Training:* These groups show similar patterns, with a slight peak around 12-15 years of education. This suggests that many individuals in these categories have completed high school and some have attended college or vocational training programs.
3. *Marginal/Irregular Part-Time and Not Employed:* These groups have a broader distribution of years of education, with many individuals having less than a high school education and others having more. This spread indicates a wide range of educational backgrounds among these individuals.
4. *Sheltered Workshop:* This group shows a unique distribution, with most individuals having fewer years of education, typically less than high school.



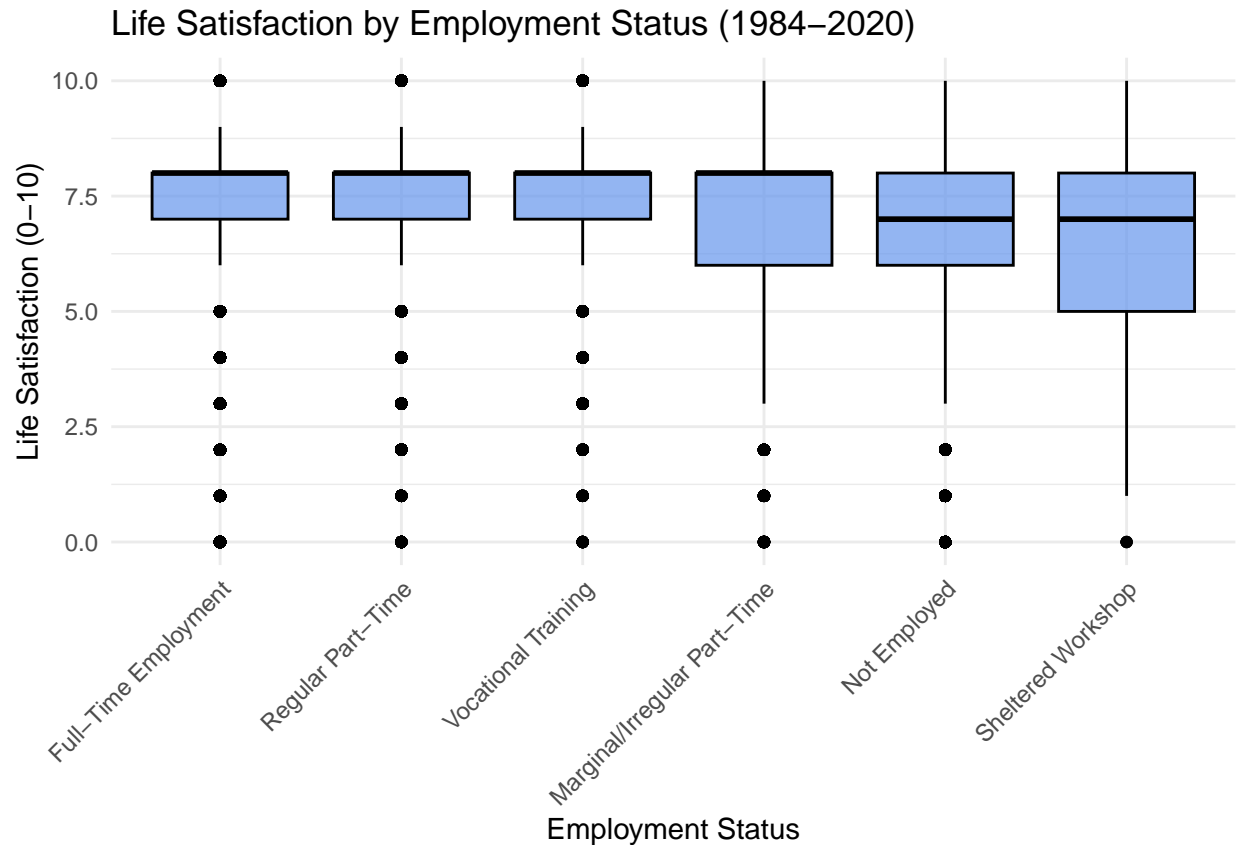
Life satisfaction and employment status

This box plot showing the distribution of life satisfaction scores (on a scale from 0 to 10) across the same employment statuses.

Main observations:

Not Employed and Sheltered Workshop groups generally report lower life satisfaction, with median scores slightly lower than the employed groups. The range of scores is wider, suggesting greater variability in life satisfaction among individuals who are not employed or in sheltered workshops. Some individuals report high life satisfaction, while others report very low satisfaction.

The gap in education levels between different employment categories (e.g., full-time vs. not employed) suggests a stratification in employment opportunities based on educational attainment. Those with higher education levels are more likely to secure stable employment, which is associated with higher life satisfaction.



Commuting

This section explores the relationship between commuting patterns and overall life satisfaction. By examining various metrics such as commuting distance, time, and frequency over different years, we can gain a deeper understanding of how changes in commuting behavior might influence happiness.

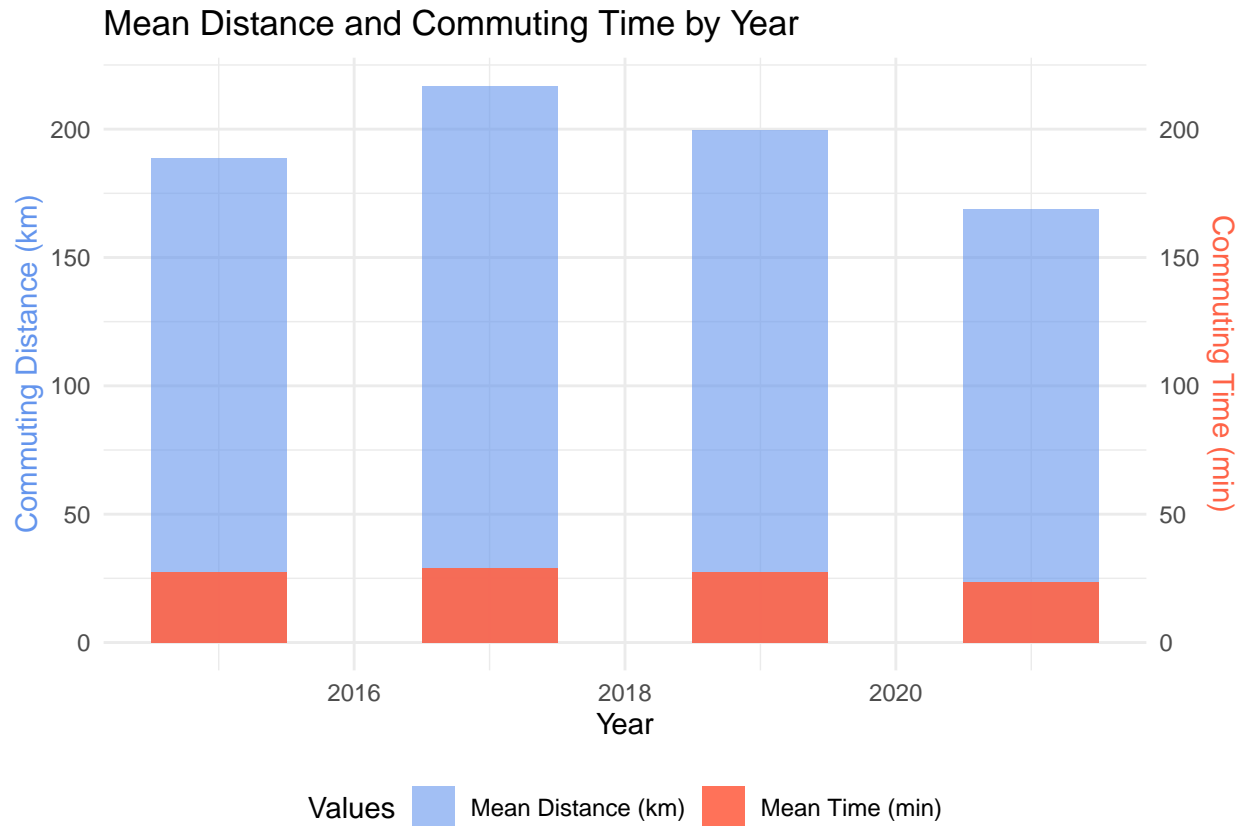
Mean Commuting Distance and Mean Commuting Time

The bar plot compares two different metrics over several years: Mean Commuting Distance (in kilometers) and Mean Commuting Time (in minutes). Each bar represents a different year, showing both metrics on the same graph using two different scales.

Main observations:

Over the years shown (2015, 2017, 2019 and 2021), we see variations in both commuting distance and time. For instance, the mean commuting distance increased from 2015 to 2017 and started slightly decreasing after, while the mean commuting time remained almost constant across these years.

Since 2017, commuting has decreased, likely due to the increase in remote work and online job opportunities. However, for those who continued to commute, the time spent traveling remained significant, possibly due to heavy traffic in cities.



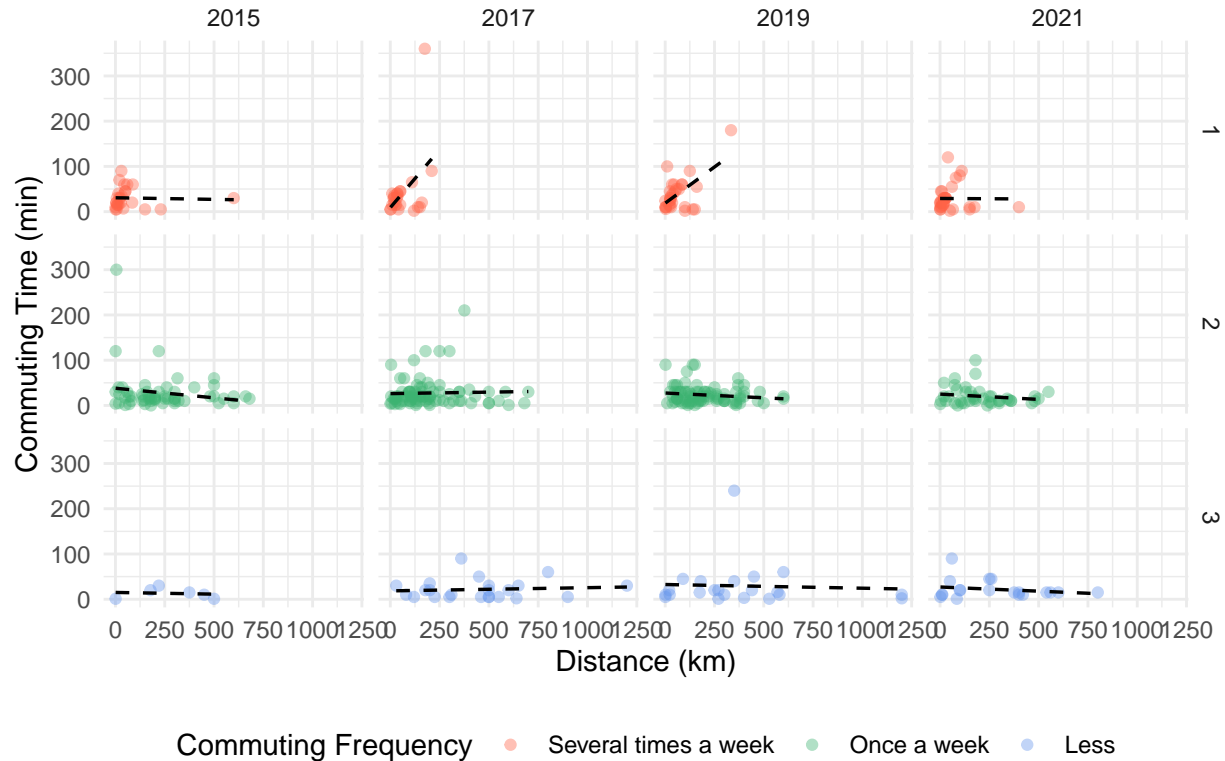
Commuting Distance vs. Time by Frequency of Commuting and Year

This scatter plot matrix visualizes the relationship between commuting distance (in km) and commuting time (in minutes) for different commuting frequencies and years. The different panels represent different years (2015, 2017, 2019), and the points are colored based on commuting frequency.

Main observations:

- As expected, there is a general trend where longer commuting distances correspond to longer commuting times.
- The black dashed line represents the mean commuting time for each distance category, and it appears that several people commute long distances relatively quickly (likely due to faster transportation means).
- High-frequency commuters (red dots) tend to have more consistent commuting times, even at longer distances, especially in 2019, compared to lower-frequency commuters (blue and green dots).

Commuting Distance vs. Time by Frequency of Commuting and Year



Interactive Scatter Plot of Commuting Time vs. Life Satisfaction by Year

The scatter plot illustrates the relationship between commuting time (in minutes) and life satisfaction across three different years: 2015, 2017, and 2019. During the data preparation process, the initial dataset of circa 400,000 respondents was significantly reduced, leaving only 354 respondents for analysis.

Main observations:

- Across all three years, life satisfaction scores range from 0 to 10. The distribution is somewhat consistent, with a majority of scores falling between 4 and 10.
- Over the years, there is a consistent observation that moderate commuting times (under 100 minutes) are associated with higher life satisfaction scores, which generally range from 6 to 8.
- A noticeable trend is the increase in lower life satisfaction scores among those with longer commuting times, especially in 2019.

This correlation proves the plot's results and shows the absence of almost any *direct* connection between commuting time and life satisfaction.

```
## [1] "Correlation coefficient of life satisfaction and commuting time: 0.0140971713975234"
```

The analysis of commuting patterns over the years reveals several important trends.

- * First, the decrease in commuting distances post-2017 likely reflects the growing adoption of remote work.
- * Second, frequent commuters tend to have more predictable and consistent commuting times, suggesting that commuting behavior is a significant factor in life satisfaction.
- * Finally, while longer commutes can negatively impact life satisfaction, the relationship is complex and multifaceted.

These insights emphasize the importance of understanding the broader context of commuting behaviors and their impact on well-being.

Income and life satisfaction

In a previous section, I examined household income and individual income per family member, observing a steady increase in both over nearly 40 years. Building on these findings, this part delves deeper into the relationship between individual income and life satisfaction. Understanding how personal financial stability influences well-being can provide valuable insights into the broader socioeconomic factors that contribute to happiness and overall quality of life. This analysis aims to explore whether rising incomes are consistently linked to higher levels of life satisfaction.

Correlation between household income and life satisfaction

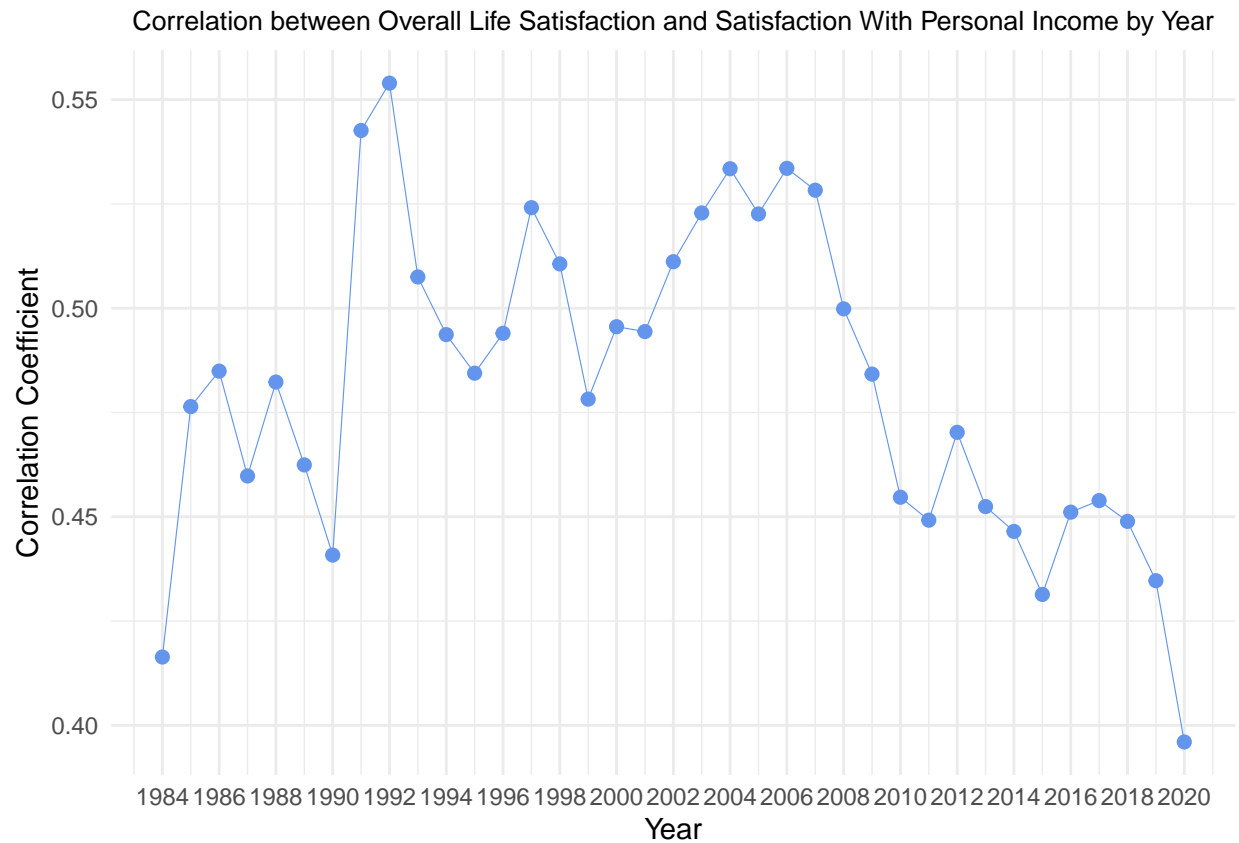
This correlation suggests that increasing household income can contribute to improved life satisfaction, but it is not the sole factor.

```
## [1] "Correlation coefficient between satisfaction with household income and overall life satisfaction"
```

Correlation between Overall Life Satisfaction and Satisfaction With Personal Income by Year

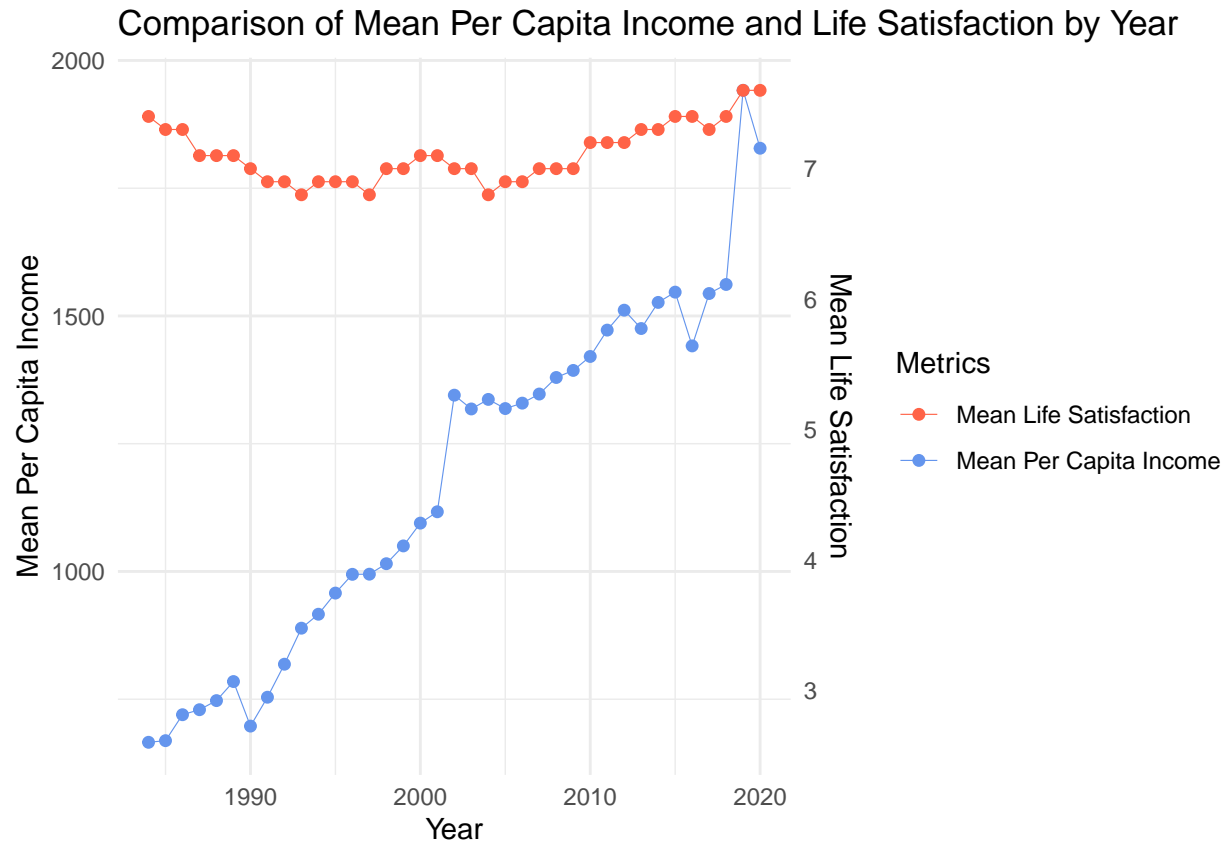
The provided graph illustrates the correlation between overall life satisfaction and satisfaction with personal income from 1984 to 2020.

Post-1990, there is a decline in the correlation, followed by several fluctuations. The decline in the correlation from the mid-2000s onward suggests that the association between life satisfaction and income satisfaction became weaker over time. This could imply that factors other than income satisfaction may have started to play a more significant role in determining overall life satisfaction.



The “Comparison of Mean Per Capita Income and Life Satisfaction by Year” graph compares the trends in mean per capita income and mean life satisfaction over time, illustrating their separate trajectories and relative scales.

The parallel increase in both income and life satisfaction suggests a positive correlation between financial well-being and overall happiness. However, the less pronounced rise in life satisfaction compared to income growth indicates diminishing returns. ***After a certain point, additional income may have a reduced impact on life satisfaction.***



A correlation coefficient of 0.465 indicates a moderate positive relationship between the two variables in question. It suggests that there is a moderate tendency for higher satisfaction with personal income to be associated with higher overall life satisfaction, but it also indicates that other factors may be contributing to overall life satisfaction.

```
## [1] "Correlation coefficient of life satisfaction and commuting time: 0.465"
```

The analysis of income and life satisfaction data over several decades reveals a nuanced relationship between financial well-being and happiness. While income growth is positively correlated with increased life satisfaction, the relationship is not absolute. The correlation between income satisfaction and overall life satisfaction has declined over time, suggesting the growing importance of non-financial factors in shaping well-being. These findings highlight the complexity of happiness, where financial security is necessary but not sufficient.

Conclusion

The analysis of the relationship between happiness (measured as life satisfaction) and commuting, as well as other socio-economic factors, reveals a complex interplay of variables that influence individual well-being.

Main findings:

- **Commuting and Happiness:** Consistent with the literature, longer commuting times are generally associated with lower levels of life satisfaction. The analysis reveals that individuals with longer commuting distances tend to report lower happiness levels. This supports the “commuting paradox”

identified in previous studies, where the negative impact of longer commutes on well-being often outweighs the potential benefits of better employment opportunities or higher income associated with longer distances.

- **Income and Happiness:** The analysis of income data highlights a moderate positive correlation between household income and life satisfaction, suggesting that financial stability contributes to happiness. However, the diminishing returns observed indicate that beyond a certain point, additional income has a reduced impact on life satisfaction. This finding aligns with the notion that while income is crucial for well-being, it is not the sole determinant of happiness. Interestingly, when commuting is factored in, individuals with higher incomes and longer commutes do not necessarily report higher happiness levels, emphasizing the complex trade-offs between financial benefits and the time costs of commuting.
- **Education and Employment:** Higher educational attainment is generally associated with higher life satisfaction, as it often leads to better employment opportunities and financial stability. This analysis shows that individuals with more years of education are more likely to be employed full-time and report higher happiness levels. However, those in lower-paying jobs or with irregular employment patterns report lower life satisfaction, even if their commuting times are shorter. This suggests that employment quality, not just the presence of a job, is a critical factor in the commuting-happiness equation.
- **Commuting Patterns:** The analysis of commuting patterns over time reveals a trend towards shorter commuting distances and times, likely influenced by the rise of remote work and flexible working arrangements. While this shift has the potential to enhance life satisfaction by reducing daily stress and time lost to commuting, the overall impact on happiness also depends on other socio-economic factors, such as income, employment status, and household composition.
- **Income Satisfaction and Life Satisfaction:** The correlation between overall life satisfaction and satisfaction with personal income has weakened over time, suggesting that other factors beyond income, such as work-life balance, commuting time, and personal relationships, are increasingly important in determining overall happiness.
- **Demographics and Happiness:** In my analysis, I did not delve deeply into the detailed correlation between demographic factors and commuting behaviors. However, based on the conducted research and the comparison of the generated graphs, it is possible to draw some insights. The findings suggest that demographic factors such as age, gender, and marital status do play a role in shaping commuting behaviors and their impact on happiness. Specifically, younger respondents, particularly those from more recent data, reported higher life satisfaction despite variations in commuting times, which may reflect greater adaptability or resilience. Additionally, married individuals, who generally have shorter commutes compared to single individuals, tend to report higher levels of happiness. This underscores the influence of social support systems on overall well-being.

All factors examined in this study—demographics, income, education, employment, and commuting—are interconnected in their influence on happiness. Demographics shape commuting behaviors and preferences, which, in turn, affect life satisfaction. Income and education not only determine the quality of employment but also influence the choice of commuting methods and distances. Employment status directly affects commuting patterns, as full-time and stable jobs may require longer commutes, while flexible or remote jobs may reduce the need for commuting altogether.

Appendix

Variables' names

pequiv.dta

d11101: Age

d11102ll: Gender
d11104: Marital Status
d11106: Number of persons in HH
d11109: Number of Years of Education
d11108: Education With Respect to High School

pgen.dta

pgemplst: Employment Status

pl.dta

plh0175: HH income satisfaction
plb0590: Distance 1. Dwelling - 2. Dwelling
plb0591: Commuting Frequency
plb0592: Commuting Time (Minutes)

hgen.dta

hghinc: Monthly HH income

References

- Clark, B., Chatterjee, K., Martin, A., & Davis, A. (2019). How commuting affects subjective wellbeing. *Transportation*, 47, 2777–2805. <https://doi.org/10.1007/s11116-019-09983-9>
- Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702), 1776–1780. <https://doi.org/10.1126/science.1103572>
- Meier, S., & Stutzer, A. (2008). Is volunteering rewarding in itself? *Economica*, 75(297), 39–59. <https://doi.org/10.1111/j.1468-0335.2007.00597.x>
- Petrunoff, N., Rissel, C., & Wen, L. M. (2017). The effect of active travel interventions conducted in work settings on driving to work: A systematic review. *International Journal of Environmental Research and Public Health*, 14(10), 1176. <https://doi.org/10.3390/ijerph14101176>
- Stutzer, A., & Frey, B. S. (2008). Stress that doesn't pay: The commuting paradox. *Scandinavian Journal of Economics*, 110(2), 339–366. <https://doi.org/10.1111/j.1467-9442.2008.00542.x>