

Final Project

Group Members:

Name / Cnetid / Github_username / section

1. Pei-Chin Lu / Peichin / Pei0504 / section 4

2. Yuan Qi / yuanqi / freyaqi / section 4

3. Huiting(Aurora) Zhang / zhanght / aurorazhang688 / section 4

Data File

Due to file size limitations, the necessary data file for this project is hosted on Google Drive. You can download it using the following link:

[Download Original Data File]

(<https://drive.google.com/file/d/1TGlCQULvQDOUOzWxdpDD60O2IJmcQer0/view?usp=sharing>)

Make sure to download this file and place it in the appropriate directory before running the project.

Research Question and the approach we took

Research in various fields has shown that **residential mobility** influences key aspects of how individuals think about themselves, interact with others, and perceive public rules. Based on this, our project primarily investigates the relationship between social mobility (e.g., economic opportunities, migration patterns, and educational access) and a range of personal characteristics. Specifically, we will pay attention to residents' **Happiness (HAPPY), Trust (TRUST), and Fairness (FAIR)**. These variables were selected because they represent essential aspects of individual well-being, interpersonal dynamics, and societal norms. Establishing reliable connections between these factors is crucial for designing effective public policies that enhance social creativity and public well-being.

To achieve this, we conduct a series of regression analyses. Additionally, to ensure that our variable selection is free from selection bias or "**cherry-picking**" (hand-picking variables to show favorable insights), we have implemented additional measures. We developed a method to objectively verify the validity of our variable selection, including the use of Exploratory Graph Analysis (EGA) and statistical tests, to classify variables as "Interested," "Proximate," or "Distal." These measures and visualizations help enhance the credibility of our research and ensure that our conclusions are derived from objective data analysis. All these charts and visualizations will ultimately be presented in our Shiny app.

Literature Base

A growing body of literature highlights the significant impact of residential mobility on both individual and cultural dynamics. At the individual level, residential mobility has been linked to increased individualism, a heightened sense of freedom, optimism, and well-being. At the group level, it fosters broader but shallower social networks, higher relational mobility, and greater trust in strangers.

A study **Shifts in Residential Mobility Predict Shifts in Culture** by Buttrick, Cha, and Oishi used quantitative methods to examine the positive impact of residential mobility on cultural variables such as trust, fairness, and happiness.

Research Objective

Building on these studies, we aim to revisit this topic using the methodology of Buttrick, Cha, and Oishi to verify the impact of residential mobility on key cultural dimensions. The results will offer valuable insights for shaping mobility-related policies and understanding their broader cultural implications.

Setup

```
C:\Users\freya\Python\Lib\site-packages\dask\dataframe\__init__.py:49: FutureWarning:
```

```
Dask dataframe query planning is disabled because dask-expr is not installed.
```

```
You can install it with `pip install dask[dataframe]` or `conda install dask`.  
This will raise in a future version.
```

Load Dataset

We utilize publicly available datasets such as the General Social Survey (GSS) from 1978 to 2018, which provides comprehensive data on social trends, public happiness, and socio-economic factors in the United States. In addition, we will incorporate data on U.S. Immigration and GDP, sourced from government and financial databases like the Federal Reserve Economic Data (FRED). These datasets will be merged and preprocessed within our Shiny app to ensure consistency and compatibility.

Immigration Data (DHS)

The yearly immigration population in the United States, as one covariate of our regression model. Data source: <https://www.dhs.gov/immigration-statistics>

Gross Domestic Product (GDP) Data (FRED)

Yearly gross domestic product data in the United States as one covariate of our regression model. Data source: <https://fred.stlouisfed.org/>

General Social Survey (GSS)

All Likert scale variables related to culture comes from this dataset, including measurements about Social Issues, Political Views, Health and Well-being, Religious Beliefs in the United States, as dependent variables of our regression model. Data source: <https://gss.norc.umd.edu/>

Residential Mobility Data (ACS)

Yearly mobility level in the United States, as independent variables of our regression model Data source: <https://www.census.gov/programs-surveys/acs>

In labels_data, we categorized each personal characteristic variable into one of the following types for user selection in the Shiny app interface: Likert Scale Variables, Binary Variables, Continuous Variables, Multichoice Variables, Administration Variables.

Clean Data

We cleaned the GSS dataset by recoding specific variables to ensure consistency and handle missing values. Columns with specific values (e.g., 4 or 5) were recoded to NaN, while others had their values swapped to align with our analysis needs. After preprocessing, we calculated the mean of each column grouped by year and saved the results to a CSV file for further analysis.

We calculated the number of missing values for each column in data_by_year and mapped these counts to the corresponding variables in labels_data.

We standardized column names, cleaned data by removing unnecessary columns, and calculated yearly averages. External datasets like US Immigration, GDP, and mobility data were processed and merged with GSS data using the "year" column. GDP data was filtered for October observations, ensuring all datasets were aligned and ready for analysis.

We prepared the data by interpolating missing values using KNN imputation, standardizing the variables for consistency, and merging external datasets like immigration and GDP data to add context. Lagged variables were created to capture temporal relationships, ensuring the dataset was ready for regression and statistical analysis.

One key difference between our analysis and the authors' methodology is how we handled missing data. While the authors used interpolation to address missing values—preserving more of the time-series structure in the data—we used KNN imputation in our analysis. This difference in data handling might explain some of the variations in the results, particularly in terms of the short-term impacts of mobility on happiness and trust.

Shiny app

To further help other researchers who are interested in studying effect of mobility on culture factors in GSS dataset, we build up a shiny app.

By using this Shiny App, users can freely choose any variables they are interested in, run the analysis, and visualize the results, ensuring the research process remains unbiased and transparent. This approach will allow future studies to validate their findings and ensure their analysis is not influenced by subjective variable selection.

UI Design

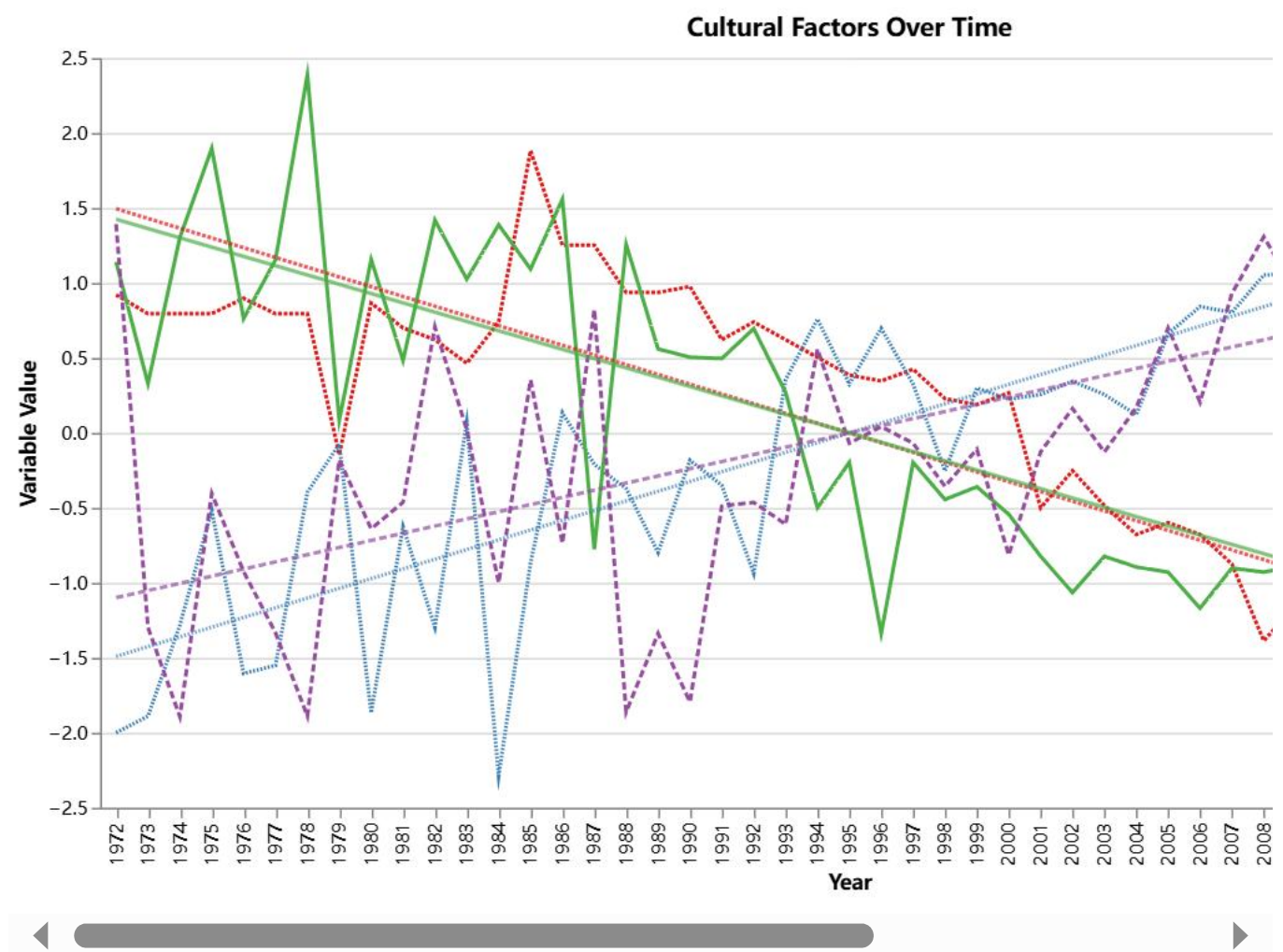
In the Shiny app interface, users select variables based on two main criteria: Variable Types and Missing Value Threshold for the period from 1972 to 2018. Users can filter variables by selecting specific types and adjusting the threshold slider. Based on the selected variables, the system generates three visualizations and a regression results table.

Server Logic

Altair-Static Plot

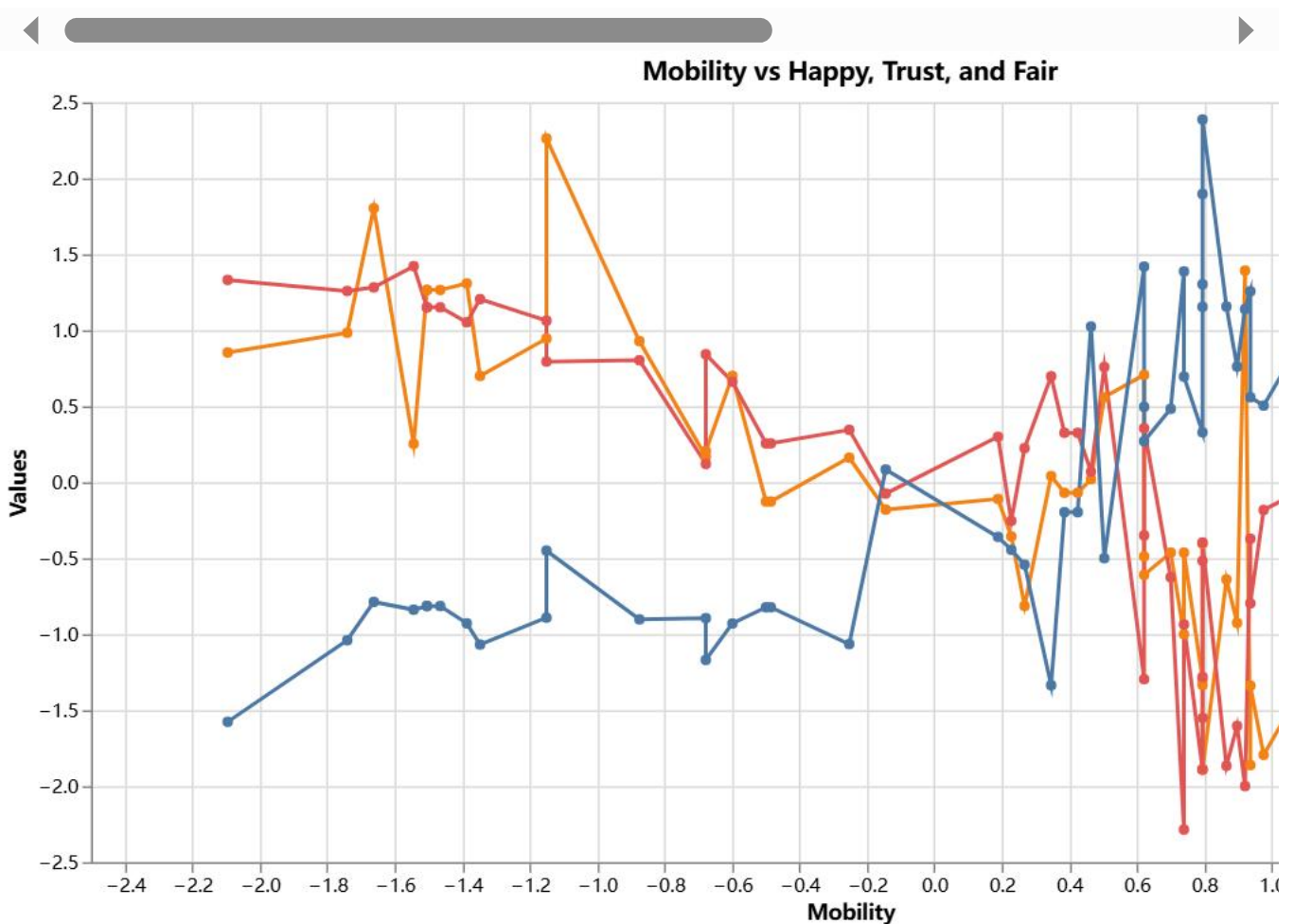
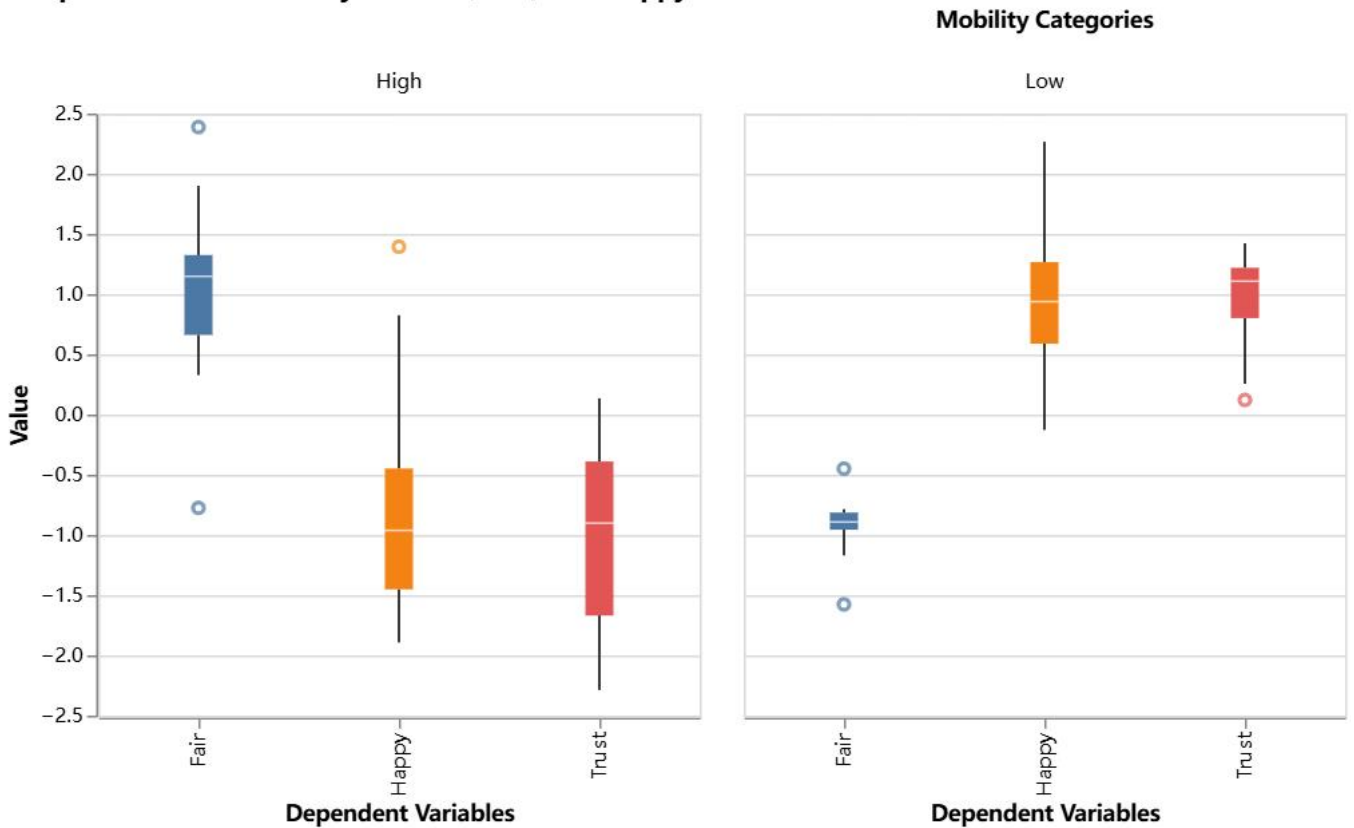
Before doing regression and cherry-picking test, let's draw some pictures using altair to observe the patterns and trends between mobility and Happy, Trust and Fair first.

Focused variables over Time



Effects of mobility on Trust, Fair, Happy

Boxplot: Effect of Mobility on Trust, Fair, and Happy



From these plots, we can generally conclude that mobility has a positive relationship with Fair as opposed to a negative relationship with Happy and Trust.

Regression and Dynamic Plots

Now, let's do the regressions and draw some dynamic plots.

The first visualization graph involves Exploratory Graph Analysis (EGA), where variables are clustered using a structured process. Selected variables are standardized and subjected to Principal Component Analysis (PCA) for dimensionality reduction, mapping the data onto two principal components for visualization. These components are then clustered using the K-Means algorithm, dividing the variables into four distinct groups. The results are displayed in a PCA plot, showing the distribution of variables across clusters. From this plot, we know clearly about which cluster our selected variables belong or don't belong to and what are the other variables in the same cluster as our selected variables so that we are able to do some further analysis.

The regression results table complements the visualizations by presenting coefficients, t-values, p-values, and R-squared values for the relationship between the selected variables and mobility. This dual approach enhances the analytical depth, offering users both statistical and visual insights into the data.

Then, we categorized variables into three distinct types for our analysis:

- **Interested/focal Variables:** These are user-selected variables that represent specific areas of focus. They are explicitly identified by the user and belong to a particular cluster of interest.
- **Proximate Variables:** These variables share the same cluster as the Interested Variables but are not directly chosen by the user. They are closely related but remain unselected.
- **Distal Variables:** These variables belong to clusters other than the one containing the Interested Variables. They represent less direct or more distant relationships to the primary focus.

Based on these three categories, we created the histogram chart and the raincloud chart.

The first chart is a histogram plot that visualizes the absolute t-values for regression coefficients, categorizing variables into the three types: Interested, Proximate-other, and Distal-other. Each bar represents a variable, with the t-value magnitude indicating the strength of the relationship between that variable and the independent variable—mobility. The color coding helps distinguish the variable types based on their clustering.

The second chart is a raincloud plot comparing the distribution of absolute t-values across the same three clusters: Interested, Proximate-other, and Distal-other, along with an additional cluster, All-other. The plot shows the spread, central tendency, and individual data points of t-values for each cluster. The annotated Z and p-values compare the statistical significance of differences between clusters, highlighting potential variations in regression performance among the clusters. This detailed visualization further underscores the robustness of our approach, demonstrating that our variable selection methodology maintains objectivity and credibility.

Cherry-picking test–Dynamic Plots

Selected Variables: ['INCOM16', 'INCOME', 'RINCOME', 'PARTYID', 'POLVIEWS', 'NATSPAC', 'NATENVIR', 'NATHEAL', 'NATCITY', 'NATCRIME', 'NATDRUG', 'NATEDUC', 'NATRACE', 'NATARMS', 'NATAID', 'NATFARE', 'COURTS', 'FUND', 'ATTEND', 'RELITEN', 'FUND16', 'SPFUND', 'HAPPY', 'HAPMAR', 'HEALTH', 'LIFE', 'HELPFUL', 'FAIR', 'TRUST', 'CONFINAN', 'CONBUS', 'CONCLERG', 'CONEDUC', 'CONFED', 'CONLABOR', 'CONPRESS', 'CONMEDIC', 'CONTV', 'CONJUDGE', 'CONSCI',

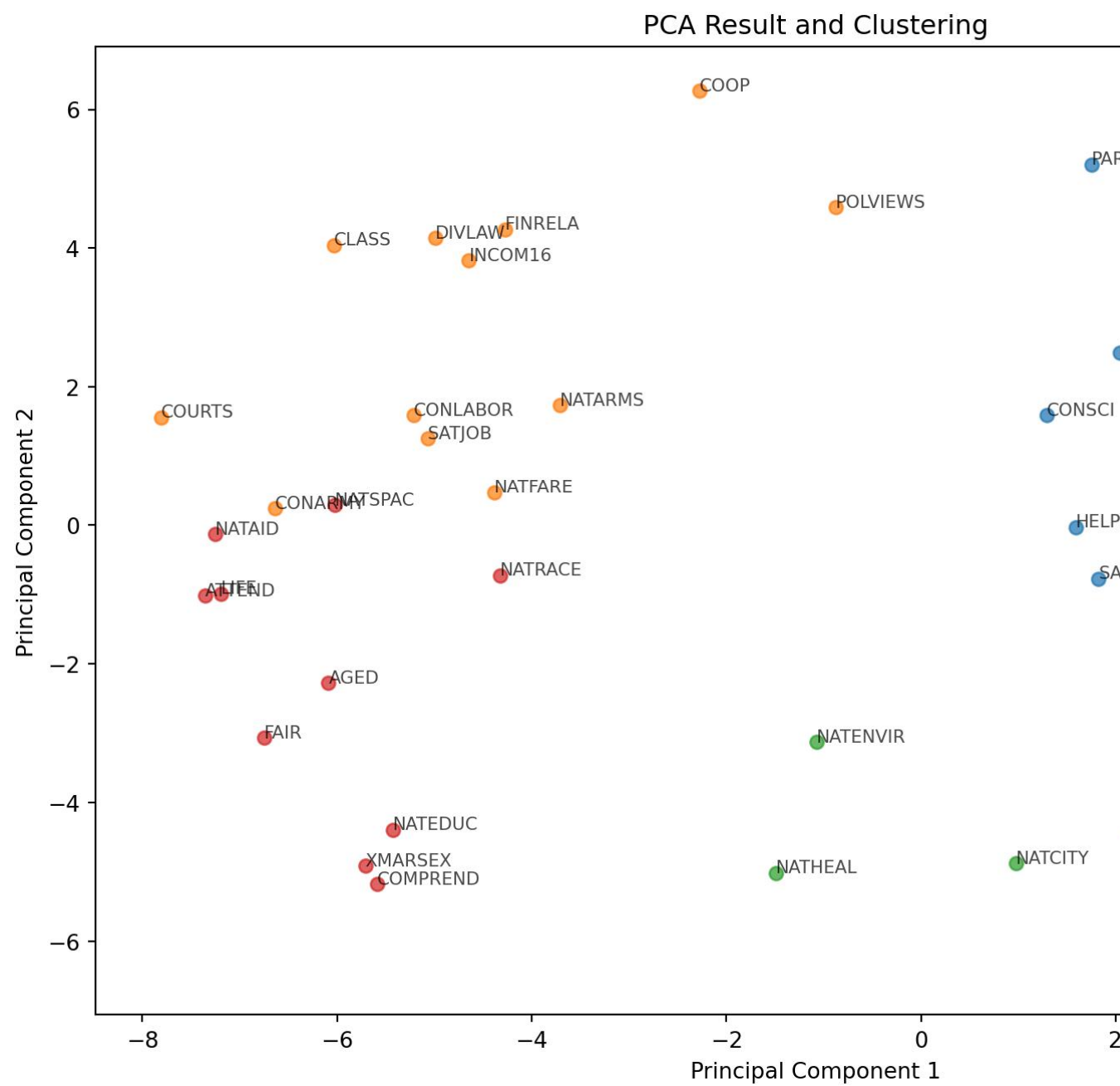
'CONLEGIS', 'CONARMY', 'AGED', 'SATJOB', 'CLASS', 'SATFIN', 'FINALTER', 'FINRELA', 'DIVLAW', 'PREMARSEX', 'XMARSEX', 'HOMOSEX', 'PORNLA', 'NEWS', 'COOP', 'COMPREND']

Cluster 1: ['INCOME', 'RINCOME', 'PARTYID', 'NATCRIME', 'NATDRUG', 'FUND', 'RELITEN', 'FUND16', 'SPFUND', 'HAPMAR', 'HELPFUL', 'TRUST', 'CONFINAN', 'CONCLERG', 'CONEDUC', 'CONFED', 'CONPRESS', 'CONMEDIC', 'CONTV', 'CONSCI', 'CONLEGIS', 'SATFIN', 'PREMARSEX', 'HOMOSEX', 'NEWS']

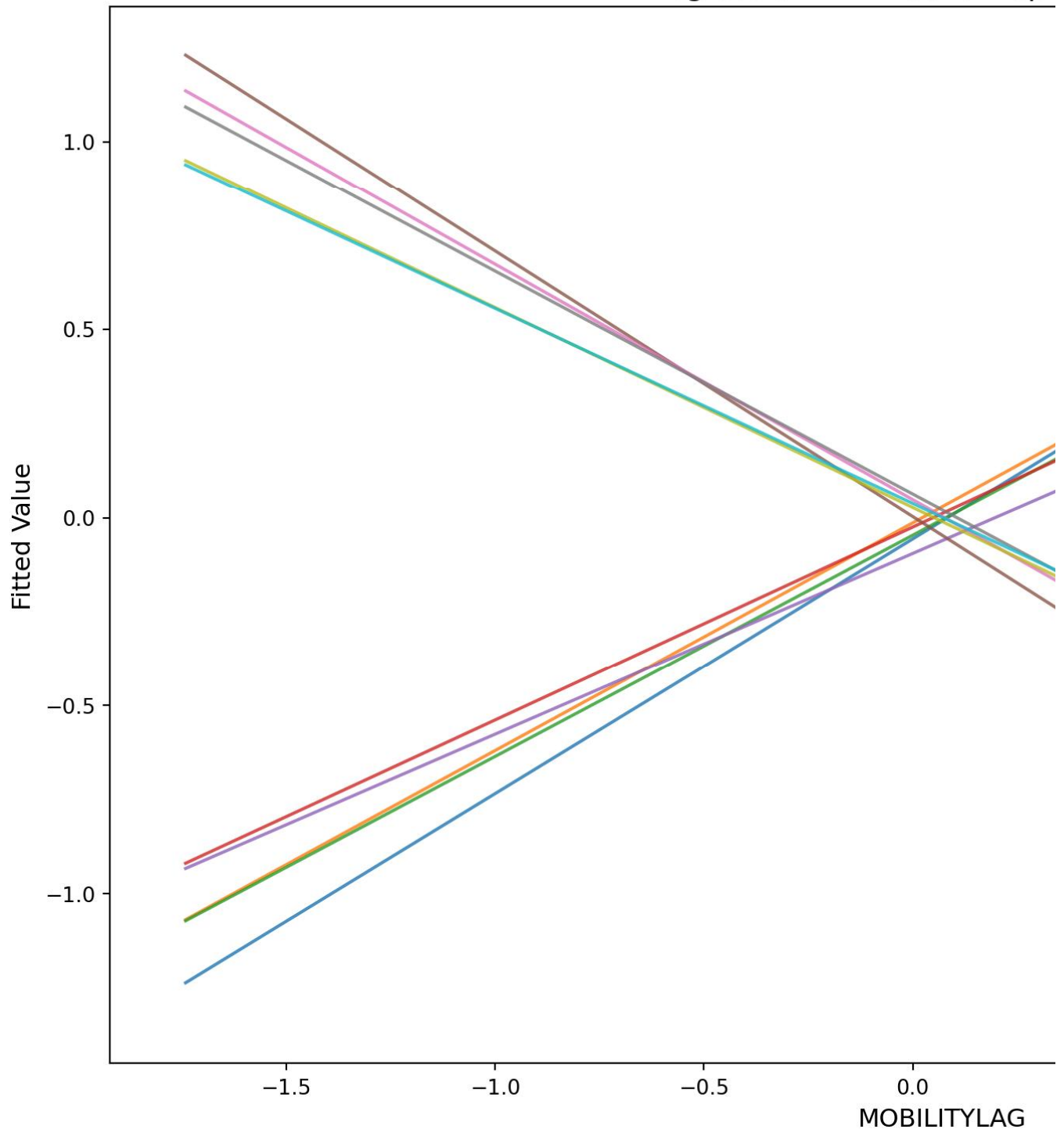
Cluster 2: ['INCOM16', 'POLVIEWS', 'NATARMS', 'NATFARE', 'COURTS', 'CONLABOR', 'CONARMY', 'SATJOB', 'CLASS', 'FINRELA', 'DIVLAW', 'COOP']

Cluster 3: ['NATENVIR', 'NATHEAL', 'NATCITY', 'HAPPY', 'HEALTH', 'CONBUS', 'CONJUDGE', 'FINALTER', 'PORNLA']

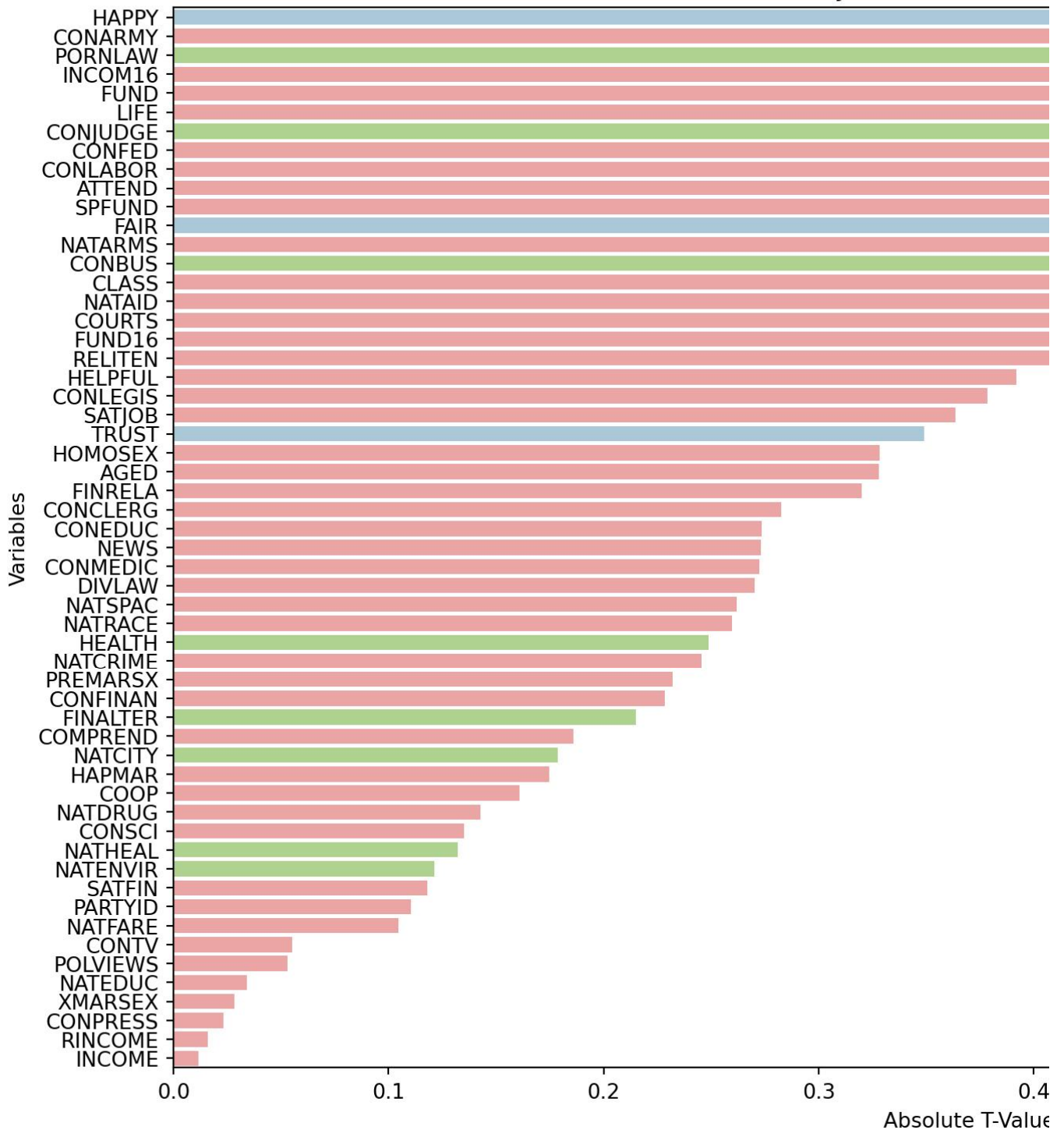
Cluster 4: ['NATSPAC', 'NATEDUC', 'NATRACE', 'NATAID', 'ATTEND', 'LIFE', 'FAIR', 'AGED', 'XMARSEX', 'COMPREND']



Regression Fit Lines for Top



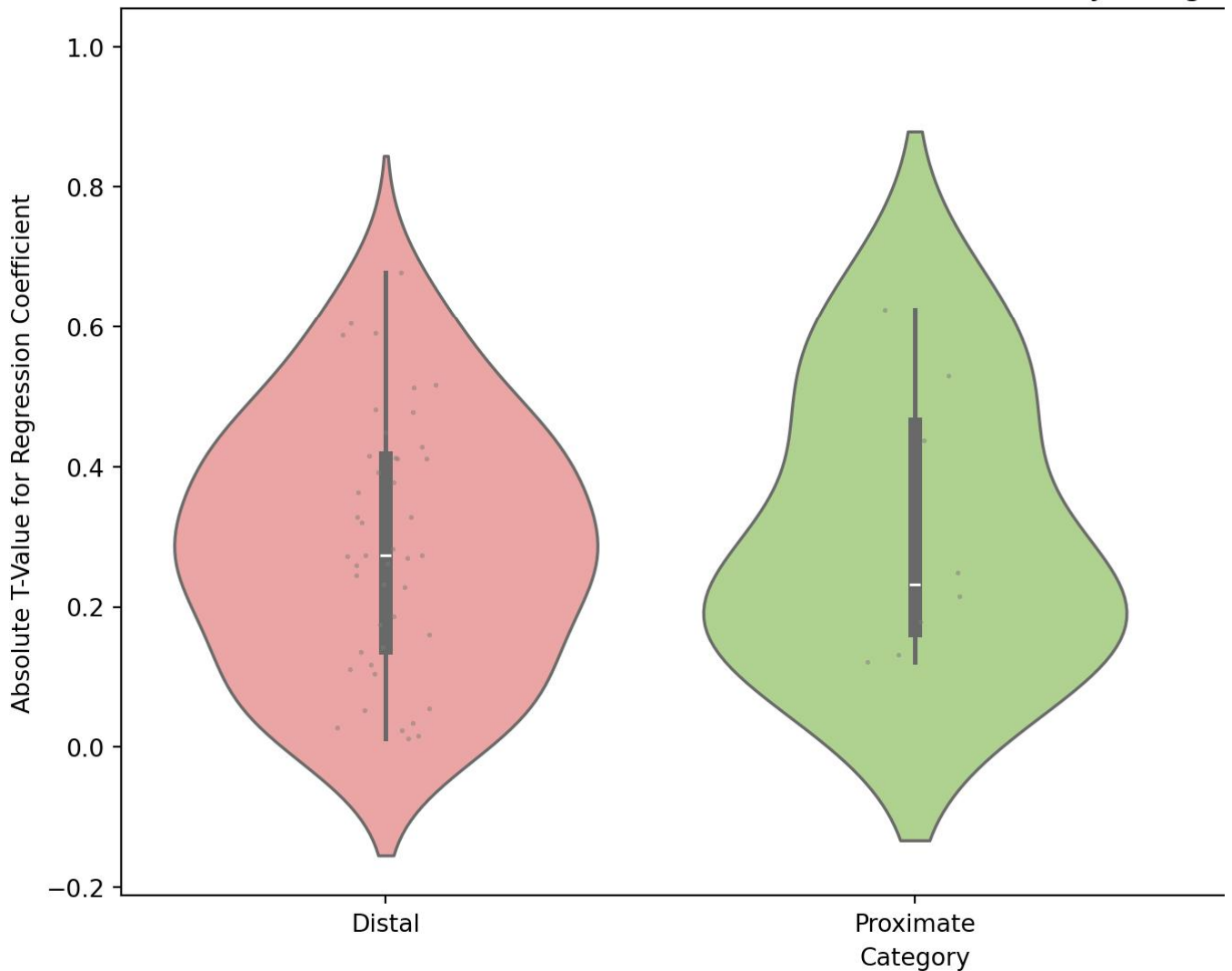
Variables by Absolute T-Value



C:\Users\freya\AppData\Local\Temp\ipykernel_12076\3407294881.py:68: UserWarning:

No artists with labels found to put in legend. Note that artists whose label start with an underscore are ignored when legend() is called with no argument.

Raincloud Plot: Absolute T-Values by Category



```
# Create and run the app
import nest_asyncio
nest_asyncio.apply()
app = App(app_ui, server)
app.run(host="127.0.0.1", port=8273)
```

Output analysis

Based on the results, we observe that the p-values for happy, fair, and trust are all below or around 0.05, indicating that mobility has a statistically significant impact on these variables. Notably, the coefficient for fair is positive(0.25), suggesting that as mobility increases, individuals perceive a greater sense of fairness in societal interactions. This could indicate that greater residential mobility exposes individuals to more diverse environments, fostering a sense of equitable opportunity and fairness in society.

On the other hand, while the literature highlights a positive link between mobility and happiness, our results show that the coefficients for happy and trust are negative(-0.58 and -0.26), implying that increased residential mobility correlates with a decrease in reported

happiness and trust. This finding reveals a potential social insight: while mobility may broaden individuals' exposure and opportunities, it can also disrupt social ties and create instability, leading to feelings of alienation or a loss of trust in others. These results suggest a dual-edged nature of mobility—it can promote perceptions of fairness while simultaneously eroding the emotional and social foundations of happiness and trust. This highlights the need for policymakers to balance efforts to increase mobility with initiatives that strengthen community cohesion and social support systems.

From the results of the histogram and raincloud plots, we uncover insights that diverge from those reported in the existing literature. While there is no cherry-picking issue in the literature, it does exist in our study.

Bar Chart

Blue bars represent our focal variables, showing the strongest effects, consistently ranking from 1st to 19th. In contrast, proximal-other variables (green bars), which are more strongly correlated with the focal variables, display a wider range of effect sizes, ranking from 2nd to near the bottom. Similarly, distal-other variables (red bars), which are less correlated with the focal variables, exhibit a similar wide range, ranking from 2nd to last.

Notably, the focal variables have a smaller range of effect sizes and higher t-values overall, suggesting a stronger and more consistent influence. However, this does not entirely rule out the possibility of cherry-picking.

Raincloud Plot

The raincloud plot further tests whether the mean effect-size distribution associated with our focal variables differs from that of other dependent variables.

A permutation-based comparison of effect sizes revealed a significant difference between the focal variables and all other groups. This finding suggests that cherry-picking might be present, as the observed differences indicate a non-random selection of variables.

Policy Implications

The findings suggest that residential mobility has a complex and dual-edged impact on individual and societal well-being. The positive correlation between mobility and perceived fairness indicates that mobility can create an environment where individuals experience greater equity and societal justice. Policymakers can leverage this by encouraging mobility through programs such as job relocation assistance, education grants, or housing subsidies that promote equitable access to opportunities.

However, the negative impact of mobility on happiness and trust raises concerns about the social and psychological costs of increased mobility. As individuals move more frequently, the erosion of trust and community ties can lead to feelings of alienation and lower overall life satisfaction. Policymakers must address these unintended consequences by fostering stronger social networks and support systems. For example:

- **Community-Building Programs:** Investing in initiatives that strengthen community bonds, such as local engagement programs, shared public spaces, or neighborhood events, can help mitigate the social disconnection caused by mobility.
- **Mental Health Support:** Introducing mental health resources and support systems for individuals experiencing mobility-related stress or loneliness could alleviate negative emotional outcomes.
- **Trust-Building Campaigns:** Promoting social trust through targeted campaigns, inclusive policies, and fostering shared cultural or civic values may counteract the erosion of interpersonal trust.

These findings highlight the importance of designing mobility-enhancing policies that are socially conscious and proactive in addressing the trade-offs between opportunity creation and emotional well-being.

Directions for Future Work

While this study sheds light on the impact of residential mobility on fairness, happiness, and trust, there are several areas for future exploration:

- **Longitudinal Analysis:** Since we only study patterns across 1978-2018, investigating a longer-term effects of mobility on well-being continuously could provide deeper insights into whether the observed impacts persist, diminish, or intensify over time.
- **Geographical Variations:** Analyzing mobility's effects in urban vs. rural settings or across different socioeconomic strata could uncover regional or demographic differences in outcomes.
- **Mechanisms Behind Trust Decline:** Future research could focus on understanding the mechanisms that drive the decline in trust due to mobility. For example, does it stem from weaker interpersonal relationships, cultural dissonance, or perceived competition?
- **Experimental Approaches:** Designing controlled experiments or natural experiments, for example, randomly assigning residents to move from one place to another to avoid pre-existing different baseline personal characteristics and validate causal relationships between mobility and personal characteristics could strengthen the robustness of these findings.

By addressing these areas, future work can provide more granular and actionable insights for policymakers, ensuring that residential mobility fosters not only equitable opportunities but also emotional and social well-being.