MEMORANDUM

To:	Carole Voulgaris, Course Instructor, SES 5215
From:	Jane Doe, Student, SES 5215
Date:	October 28, 2022
Subject:	Assignment 2, Describing and visualizing data

The purpose of this memo is to present and describe the variables in the dataset I am using the answer the question:

What is the effect of a person's income on their typical travel time to work, after accounting for the effects of transportation mode and population density?

Over the course of this semester, I will be addressing this question using census microdata from the 2015-2019 American Community Survey, which I accessed through IPUMS (Ruggles et al. 2022). My dataset includes survey all Survey respondents from California who indicated a usual commute time of more than zero minutes (804,800 people). The variables in my data set are:

- Commute time: The self-reported typical time (in minutes) it took to get from home to work in the week prior to the survey.
- Commute mode: The most common mode of transportation used to reach work in the week prior to the survey.
- Income: The person's pre-tax income in the year prior to the survey.
- Population density: The number of people per square mile of land area in the public-use microdata area (PUMA) in which the person lives.

Table 1 present basic descriptive statistics for each continuous variable in the dataset.

Table 1: Descriptive statistics for continuous variables

	Commute time (minutes)	Income	Population density (people per km^2)
Full range	1 – 142	-$11,654 - $1,453,158	3 – 15,303
Interquartile range	15 – 40	$21,902 - $83,435	444 – 2,974
Standard deviation	24	$85,687	2,172
Mean	30	$68,186	2,102
Median	25	$43,779	1,559

Commute times in the sample have a minimum of one minute, since those with zero-minute commutes were excluded from the sample, and a maximum of 142 minutes (two hours and 22 minutes. Half of the individuals in the sample had commute times between fifteen and 40 minutes, representing an interquartile range of 25 minutes, which is close to the standard deviation of 24 minutes. The median value of 25 minutes is slightly less than the average value of 30 minutes, which suggests some left skew in the distribution, as illustrated in Figure 1.

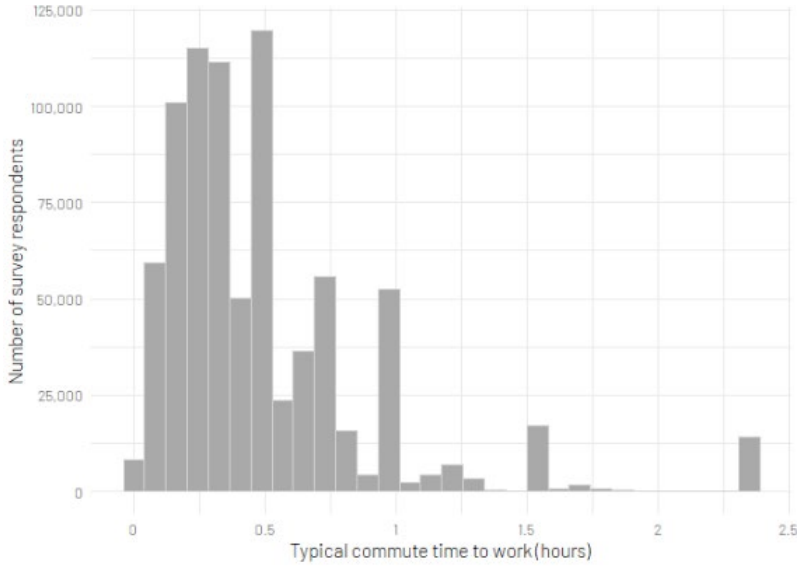

Figure 1: Distribution of commute times

Incomes in the sample have a minimum of -$11,654 (representing a new loss over the course of a year) and a maximum of $1.45 million. Half of the individuals in the sample had earned between $21,902 and $83,435, representing an interquartile range of $61,533, which narrower than the standard deviation of $85,687. The median value of $43,779 is substantially lower than the average value of $43,779. As Figure 2 shows, the distribution is approximately log-normal (the x-axis is on a log scale).

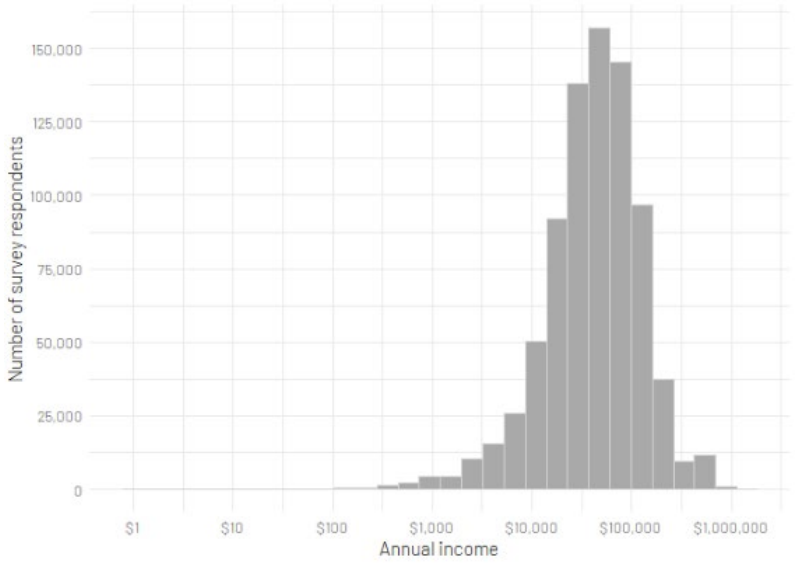

Figure 2: Distribution of incomes

The population density of the PUMA in which a person in the sample lives ranges from 3 to 15,303 people per square kilometer, with half of the sample living at densities between 444 and 2,974 people per square kilometer, representing an interquartile range of 2,530, which is a little narrower

than the standard deviation of 2,172 people per square kilometer. The average value of 2,102 is higher than the median of 1,559, suggesting some left skew in the distribution. Figure 3 illustrates the distribution in a histogram with a log-transformed x-axis. Note that the distribution of the log-transformed variable is right-skewed.

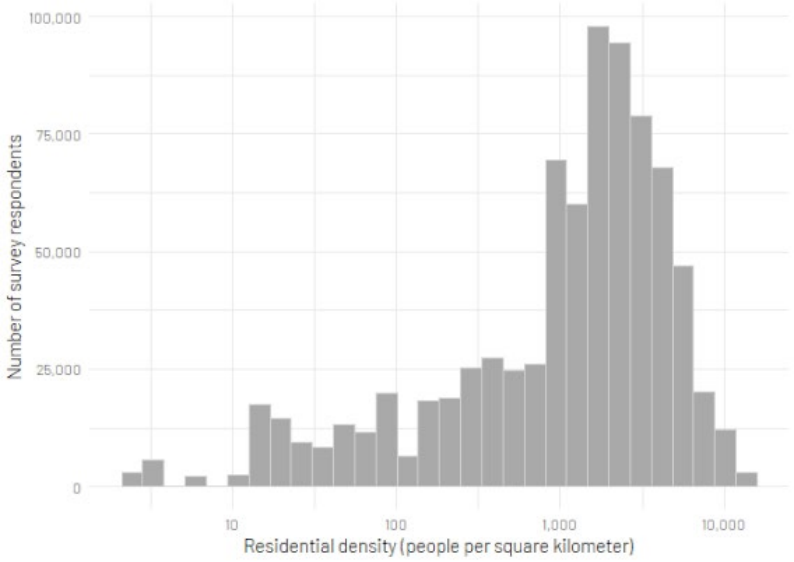

Figure 3: Distribution of residential density

Figure 4: Proportions of sample commuting by each mode

Figure 4 illustrate the proportions of the sample who commute by car, transit, walking, cycling, and a single category for all other modes. As show, a majority of the sample (89 percent) commute by car. The remaining individuals are almost evenly split between transit (5 percent) and other modes (3 percent walking, 2 percent other, and 1 percent cycling).

References

Steven Ruggles, Sarah Flood, Ronald Goeken, Megan Schouweiler and Matthew Sobek. *IPUMS USA: Version 12.0* [dataset]. Minneapolis, MN: IPUMS, 2022. https://doi.org/10.18128/D010.V12.0