MEMORANDUM

To:	Carole Voulgaris, Course Instructor, SES 5215
From:	Ramona Quimby, 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 Boston census tract's median age on the number of crashes per square kilometer that occur there, after accounting for the effects of poverty and population density?

Over the course of this semester, I will be addressing this question using crash data from Vision Zero Boston (https://data.boston.gov/dataset/vision-zero-crash-records) and demographic data from the 2016-2020 American Community Survey. My dataset includes 187 census tracts in Boston (excluding tracts with zero households), shown in Figure 1 below.

Figure 1: Census tracts in the study area

The variables in my data set are:

1. Crash density: The number of crashes recorded in each census tract between January 1, 2015 and September 30, 2022, divided by the area (square kilometers) of land within the census tract.
2. Median age: The median age of census tract residents.
3. Population density: The number of residents per square kilometer of land area in each census tract.
4. Majority poverty: A categorical variable indicating whether the majority of census tract residents have income below the poverty level.

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

Table 1: Descriptive statistics for continuous variables

	Crash density (crashes per km^2)	Median age (years)	Population density (people per km^2)
Full range	35 – 1,906	20 – 62	46 – 36,893
Interquartile range	155 – 519	31 – 37	5,436 – 13,532
Standard deviation	274	6	6,746
Mean	368	34	10,455
Median	324	34	8,810

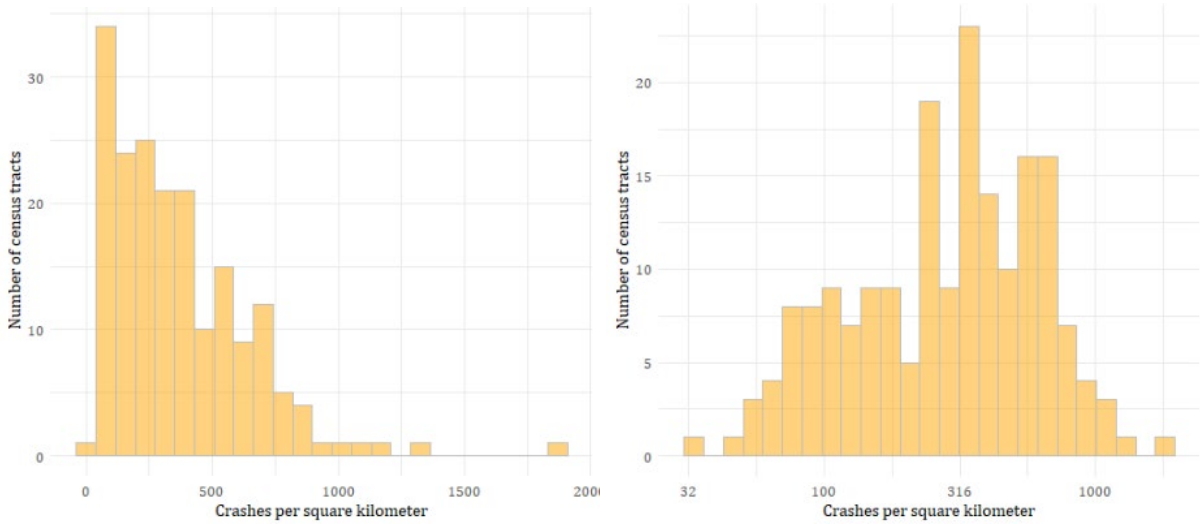

Figure 2: Distribution of crash density on a linear scale (left) and on a log scale (right)

The density of crashes within a study-area census tract ranges from 35 to 1,906 crashes per square kilometer, with the values for half of all tracts falling between 155 and 519 crashes per square kilometer. The standard deviation for the distribution is 274 crashes per square kilometer, which is less than the average value of 368 crashes per square kilometer or the median value of 324 crashes per square kilometer. As Figure 2 shows, the distribution is left-skewed, and more closely approximates a normal distribution when shown on a logarithmic scale.

The median age of a study-area census tract ranges from 20 to 62 years, with the values for half of all tracts falling between 31 and 37 years. This represents an interquartile range of six years, which is the same as the standard deviation. The mean and median values for the median age are both 34 years. The distribution is approximately normal, as shown in Figure 3.

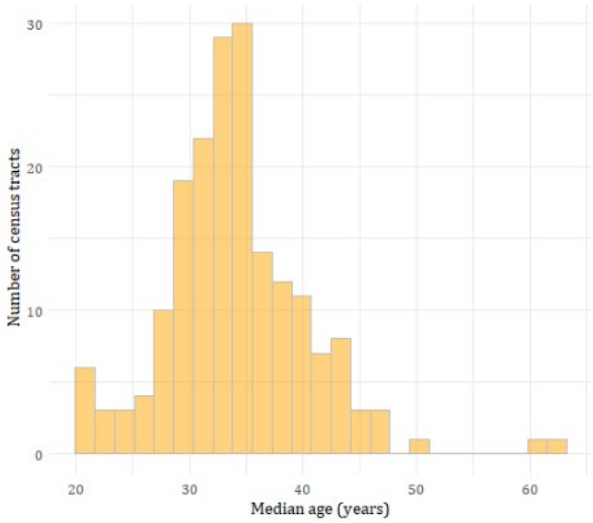

Figure 3: Distribution of median age

The population density within a study-area census tract ranges from 46 to 36,893 people per square kilometer, with the values for half of all tracts falling between 5,436 and 13,532 people per square kilometer. The standard deviation for the distribution is 13,532 people per square kilometer, which is less than the average value of 10,455 people per square kilometer or the median value of 8,810 people per square kilometer. As Figure 4 shows, the distribution is somewhat left-skewed, and more closely approximates a normal distribution (with one outlier) when shown on a logarithmic scale.

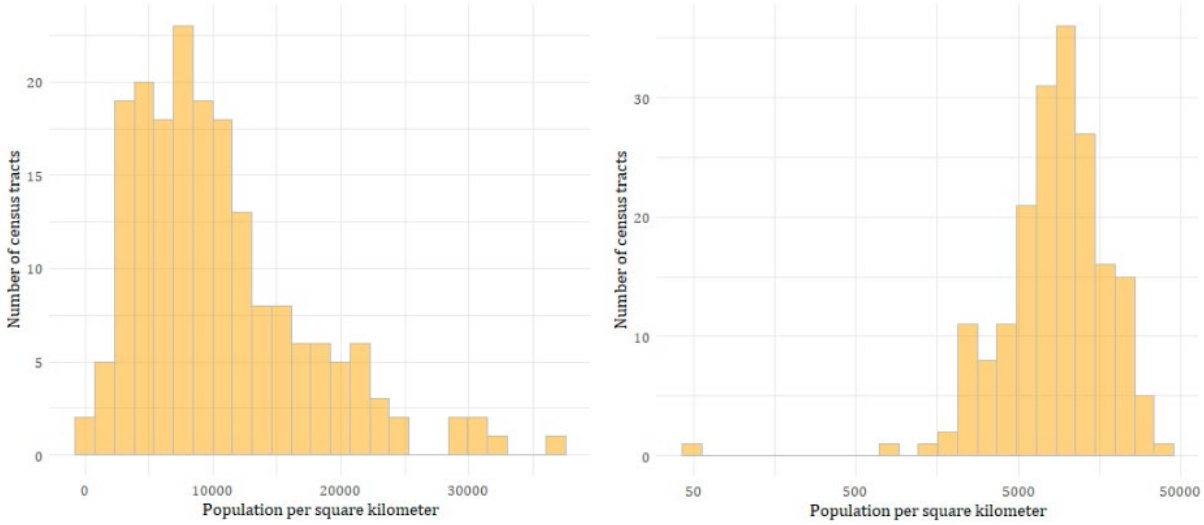

Figure 4: Distribution of population density on a linear scale (left) and on a log scale (right)

Nineteen out of the 187 tracts in the sample (just over ten percent) have a majority of households with incomes at or below the poverty level. These are shown in Figure 5.

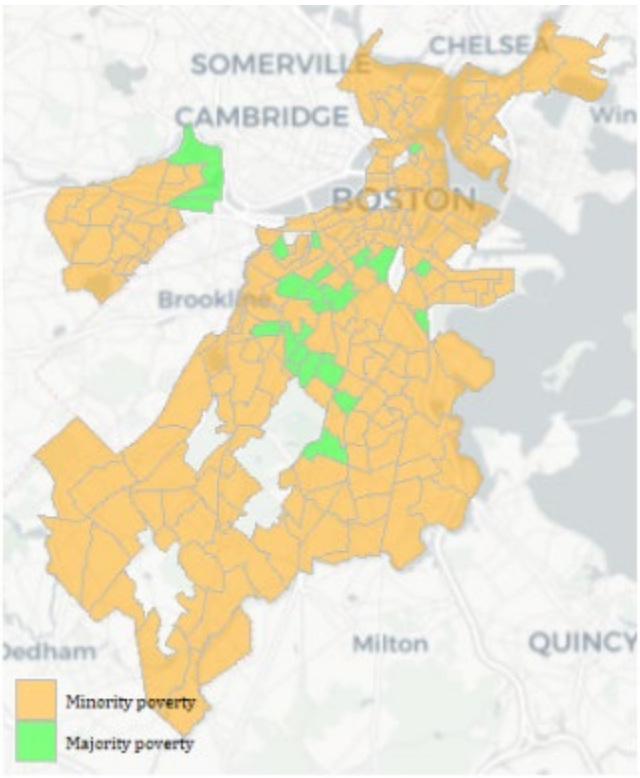

Figure 5: Locations of majority poverty census tracts