# Homework #1

POSC 3410 - Quantitative Methods in Political Science

Steven V. Miller

1. Get the mean of the percentage of the vote Trump got in the primary. Get the median. What does this suggest in terms of skew of the variable?

The mean is 36.13 and The median is 35.55. The difference between the two is 0.58, which implies there is probably not a huge problem of skew in this variable.

2. Get a density plot of the Trump share variable. What does this suggest about the distribution of this variable?
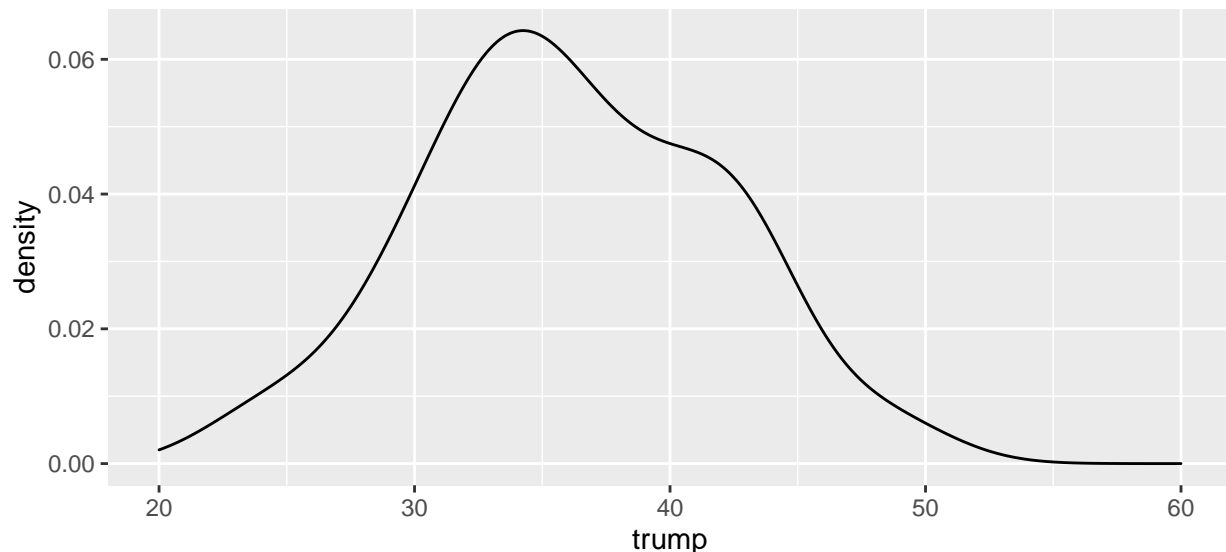
A density plot will describe the shape and distribution of a variable. Here, a density plot suggests the data look fairly normal even as there appears to be a small right skew that would account for why the mean is slightly larger than the median.

3. Look at the mean of the population variable and then look at the median. Get a density plot too. What does this suggest in terms of skew of the variable?

The mean county population in South Carolina is 105,054 while the median county population is 58,398.5. This implies a major issue of right skew in the data, which the density plot below also shows.
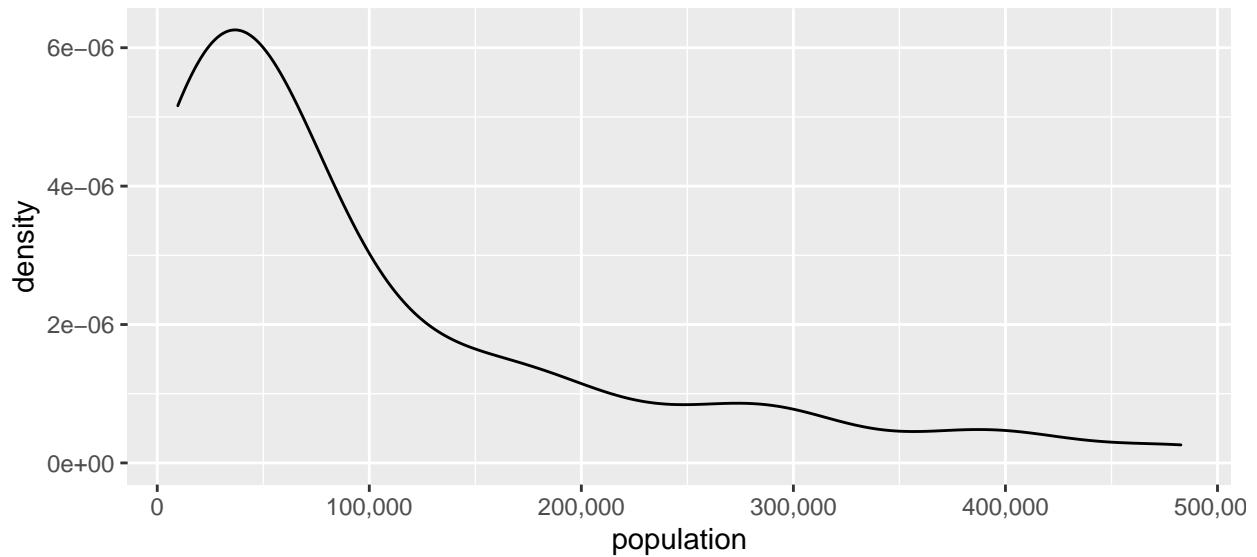
4. Why do you think the population variable has this distribution?

South Carolina has 46 counties, of which a handful (e.g. Greenville, Richland, Charleston, Horry, Spartanburg) include relatively large cities (e.g. Greenville, Columbia, Charleston, Rock Hill, and Spartanburg) while the majority of counties are quite small by comparison without major metropolitan areas. The effect of these larger counties is to pull the mean right of the median because the mean is a statistic of central tendency that is sensitive to large outliers.

5. Take a natural logarithm of the population variable and then get the mean and median again. Get a density plot too. What do you see?

The mean of this log-transformed variable is 11.02 and the median is 10.97. A log transformation effectively coerces a more normal distribution from positive, right-skewed variables like this county population variable. The density plot below will show this.
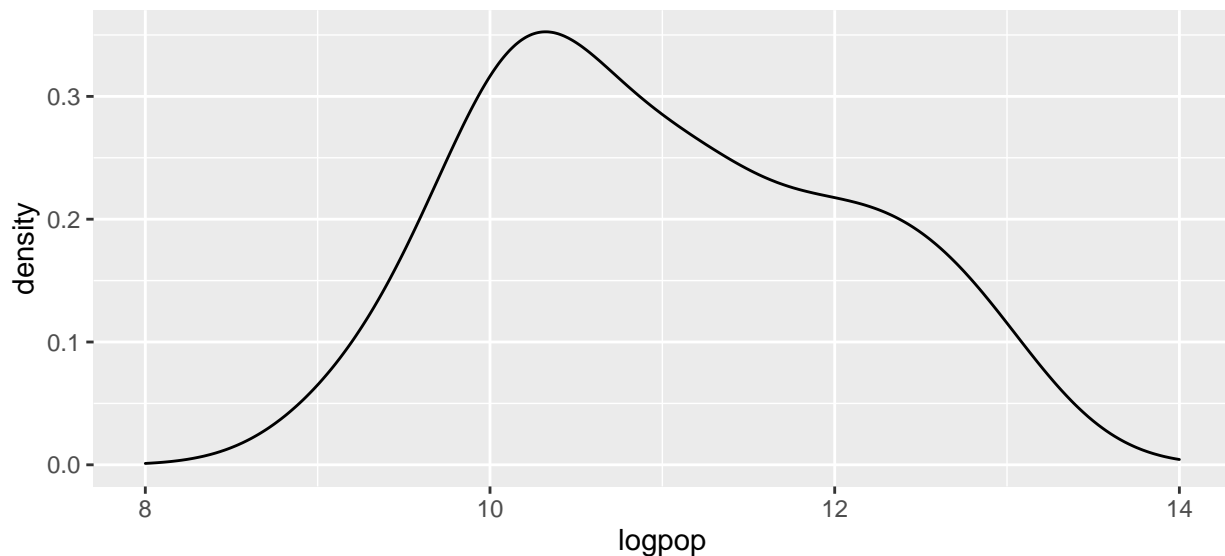
Extra Credit

6. Run the regression that seeks a minimal model to explain the county-level vote share for Trump in the Republican primary. What do you think these results suggest? Be mindful of ecological fallacies.

Increasing county-level unemployment and county-level illiteracy rates coincided with an increased county-level vote share for Trump in the 2016 South Carolina Republican primary. Further, Trump appeared to do worse in counties that had a higher percentage of African Americans and had a higher county-level vote share in counties with fewer black people as a percentage of the county population. All three factors are statistically significant, suggesting the effect of these factors is not zero.