

Practice Assignment: Interpretation of Mean and Median

- 1) In which of the following situations is it likely that the mean will be **lower** than the median?
 - a) The dataset has many low and high outliers, which are symmetrically distributed.
 - b) The dataset has 500 data values.
 - c) The dataset has right skew and several high outliers.
 - d) The dataset is mostly symmetric but contains several extremely low outliers and no high outliers.

- 2) When is it most appropriate to use the mean as a measure of center?
 - a) When the distribution has no outliers
 - b) When the distribution is symmetric
 - c) When the distribution is skewed
 - d) When the distribution has outliers

- 3) Determine whether this statement is true or false: It doesn't make sense to calculate a mean or median when describing categorical data.

Questions 4–8: These questions reference the *FiveThirtyEight* `hate_crimes` dataset,¹ which describes hate crime rates within different states.

The following table displays selected variables from the full dataset for the first 10 states (including DC).

state_abbrev	median_house_inc	hate_crimes_per_100k_spl c	avg_hatecrimes_per_100k_fb i
AL	42278	0.125838926	1.806410489
AK	67629	0.143740118	1.656700109
AZ	49254	0.225319954	3.413927994
AR	44922	0.069060773	0.869208872
CA	60487	0.255805361	2.397985899
CO	60940	0.390523301	2.804688765
CT	70161	0.335392269	3.772701469
DE	57522	0.322754169	1.469979563
DC	68277	1.52230172	10.95347971

¹ *Hate crimes*. (2018, February 9). FiveThirtyEight.
<https://github.com/fivethirtyeight/data/tree/master/hate-crimes>

The following are descriptions of each variable in the table:

state_abbrev: State abbreviation

median_house_inc: Median household income, 2016

hate_crimes_per_100k_splc: Hate crimes per 100,000 population, Southern Poverty Law Center, Nov. 9–18, 2016

avg_hatecrimes_per_100k_fbi: Average annual hate crimes per 100,000 population, FBI, 2010–2015

4) Why do you believe the dataset authors chose to report the median household income within each state rather than the mean?

5) Go to the *Describing and Exploring Quantitative Variables* tool at https://lumen-learning.shinyapps.io/eda_quantitative/. Select the **Single Group** tab, locate the drop-down menu under “Dataset,” and select “Hate Crimes (SPLC).” This dataset includes the data from the variable *hate_crimes_per_100k_splc*.

Part A: Find the **mean** hate crimes committed per 100,000 people across the states and DC. Note that four states with missing values were excluded.

Part B: Find the **median** hate crimes committed per 100,000 people across the states and DC.

Part C: Use the histogram displayed on the online tool to visualize the distribution of hate crimes committed per 100,000 people across the states and DC. Comment on the shape of the distribution and how it relates to the relationship between the mean and median you calculated previously.

6) There is one data value that has a particularly powerful influence on the previous mean calculation. Use the previous abbreviated table to identify which U.S. state (including DC) is associated with that data value and comment on whether it should be removed before calculating the mean hate crimes per 100,000 people.