

Practice Assignment: Boxplot Data and Displays

Questions 1–3: These questions reference the *FiveThirtyEight* `hate_crimes` dataset,¹ which describes hate crime rates within different states. This dataset was also used in Practice Assignment 4.C.

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

<code>state_abbrev</code>	<code>median_house_inc</code>	<code>hate_crimes_per_100k_splc</code>	<code>avg_hatecrimes_per_100k_fbi</code>
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

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

- 1) Go to the *Describing and Exploring Quantitative Variables* tool at https://lumen-learning.shinyapps.io/eda_quantitative/. Select the **Several Groups** tab, locate the drop-down menu under “Dataset,” and select “Hate Crimes (Source).”

Construct parallel boxplots (boxplots displayed side-by-side) to visualize the hate crimes per 100,000 people as reported by the FBI and the hate crimes per 100,000 people as reported by the Southern Poverty Law Center (SPLC). Draw a

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

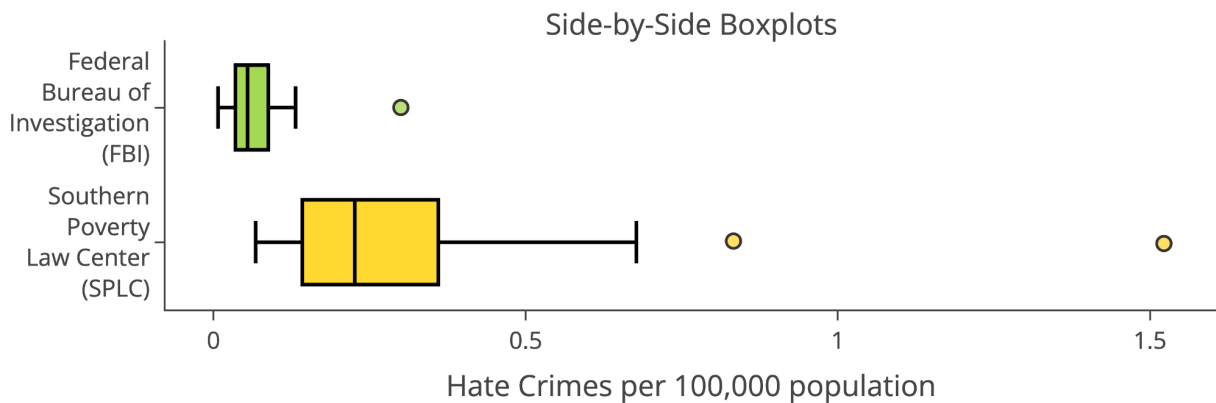
well-labeled sketch of your visualization. Note: You can ignore states with missing data.

- 2) The FBI data describes the average annual hate crime rates from 2010–2015. The SPLC data describes the hate crime rate in 2016 between November 9–18.

Part A: Compare the two distributions, using your visualization from Question 1.

Part B: Why do you think these two distributions are so different? Explain.

- 3) To put these datasets on the same time scale, we divide the FBI hate crime rates by 36.5—this yields the average 10-day hate crime rates for the years 2010–2015. We visualize this transformed FBI data alongside the SPLC data as follows:



Part A: Comment on what effect our transformation of the FBI dataset had on the parallel boxplots of these distributions.

Part B: The 10-day period covered by the SPLC data (Nov. 9–18, 2016) were the 10 days following Donald Trump’s presidential election in 2016. Some claim that the election caused a spike in hate crimes. Does the above visualization provide enough evidence to prove this claim? Explain why or why not. If not, describe further information you would need to evaluate this claim.

- 4) In what situation would a dotplot tend to be more effective than a boxplot?
a) When we’re working with a very small dataset

- b) When the distribution is symmetric
 - c) When we're especially interested in the median
 - d) When we're working with a very large dataset
- 5) Determine whether this statement is true or false: Boxplots tell you the number of data points between the 3rd quartile and the maximum.