

Data Visualizations and Transformations

In this lab, you will be asked to find patterns and draw conclusions using data collected from the CIA World Fact Book, <https://www.cia.gov/library/publications/resources/the-world-factbook/index.html>.

Our dataset consists of four variables collected in 2014:

- **country**: the name of each country within our dataset
 - **totalmilitary**: Total annual military expenditures of each country
 - **GDP**: (purchasing power parity) compares the gross domestic product (GDP) or value of all final goods
 - **population**: estimated population of that country
- 1) Go to the website, <http://rstudio.grinnell.edu/> and select the **Military Spending (Basic)** link. If not already showing, select the **Histogram** tab and select **Military Spending** from the dropdown menu. In the histogram, you should see one value that is significantly larger than the others. In the tables below the histogram, the data is sorted from smallest amount spend on military spending to the largest. Which country has the largest military spending? How much do they spend each year?
 - 2) Use the arrow keys on the **Number of Observations** to view **5** observations per group instead of 3. List the 5 countries that spend the most on military spending.
 - 3) Now select the **Spatial Map** tab and select **Military Spending** from the dropdown menu. Compare the histogram to the map.¹ Explain how both the histogram and spatial map demonstrate that the data is highly skewed. Explain how the spatial map provides more useful information than the histogram.

¹ Missing countries are due to missing data for that country.

- 4) Notice that one or two outliers can keep us from seeing trends in the rest of the data. Too better see overall trends in the data, we transformed the data by creating a new variable called **Log of Military Spending**. View the histogram and spatial map of this transformed data. Does either map indicate that the log of military spending is still highly skewed? Explain how using the log of military spending is better at demonstrating trends in the data than military spending?

- 5) Clearly larger countries with more people are likely to spend more money on military expenditures. Instead of using military spending, we will create a new variable called **Military Spending per Capita**, which is Military Spending divided by Population. Which country spends the most on military spending per capita? How much money does this country spend per person?

- 6) We can also look at military spending as a percentage of the countries economy (we will use **Military Spending/GDP**). Which country spends the most based upon GDP? Where does the United States fit into this ranking?

- 7) All four military spending variables (Military Spending, Log of Military Spending, Military Spending per Capita, and Military Spending/GDP) used in this lab are accurate, but they can lead to very different interpretations. If you wanted to use data to make an argument to reduce U. S. Military spending, which graph do you think would be most effective in supporting your views?

- 8) If you wanted to use data to demonstrate a need for increased U. S. Military spending which graph do you think would most effectively support your views? Of the variables in this file, which variable would most likely be used to indicate a need for increased military spending? Why did you choose this variable? On the variable you chose, where does the U.S. Military spending fall?

- 9) In your opinion, which graph is the best in helping us truly understand military spending in each country? Give a brief explanation for your choice.