

R Bridge Course Final Project

This is a final project to show off what you have learned. Select your data set from the list below:

<http://vincentarelbundock.github.io/Rdatasets/> (click on the csv index for a list). Another good source is found here:

<https://archive.ics.uci.edu/ml/datasets.html>

The presentation approach is up to you but it should contain the following:

1. Data Exploration: This should include summary statistics, means, medians, quartiles, or any other relevant information about the data set. Please include some conclusions in the R Markdown text.
2. Data wrangling: Please perform some basic transformations. They will need to make sense but could include column renaming, creating a subset of the data, replacing values, or creating new columns with derived data (for example – if it makes sense you could sum two columns together)
3. Graphics: Please make sure to display at least one scatter plot, box plot and histogram. Don't be limited to this. Please explore the many other options in R packages such as ggplot2.
4. Meaningful question for analysis: Please state at the beginning a meaningful question for analysis. Use the first three steps and anything else that would be helpful to answer the question you are posing from the data set you chose. Please write a brief conclusion paragraph in R markdown at the end.
5. BONUS – place the original .csv in a github file and have R read from the link. This will be a very useful skill as you progress in your data science education and career.

Please submit your .rmd file and the .csv file as well as a link to your RPub.