**[HW 4](https://blackboard.syr.edu/webapps/assignment/uploadAssignment?content_id=_4391970_1&course_id=_363064_1&assign_group_id=&mode=view)**

**Visually describe a dataset**

The purpose of this exercise is to use what you have learned so far to visually describe a dataset with R plos. This means, showing a few single dimension plots as well as at least one multi-dimension plot. One of the hardest parts of this assignment is going to be finding the right data. You will turn in a pdf with the plots and your R script. Make sure your plots are vector graphics (saved as PDF). Each plot should be on it's own page. Note that work done in this assignment can be used in both the Work in Progress Report (coming up) and the final poster.

You have 2 weeks for this assignment.

**Data**

You need to find a substantial dataset for this project. Your dataset must score at least 100 in the formula below:

**(NumberOfColumns \* 4) \* (NumberOfRows/100) >= 100**

Examples:

A dataset with 5 columns and 10,000 rows would score: (5 \* 4) \* (10,000/100) = 2000

A dataset with 17 columns and 3,000 rows would score: (17 \* 4) \* (350/100) = 238

A dataset with 5 columns and 500 rows would score: (5 \* 4) \* (500/100) = 100

The point of this is to make sure you have a large enough dataset to work with. You need enough columns to look at how how different variables might related. You need enough rows to show that you can work with a decent sized dataset.

**Descriptive plots**

Your data may have any types. For example, you might have a mix of continuous, text, and categorical data columns, or you might have several columns of just continuous data. The type of the data should dictate what kinds of plots you make.

To visually describe the data, first, make some plots in R that show how the data is distributed. For continuous data this might be histograms, boxplots or density plots. For categorical, a good choice is a barchart that shows the frequencies of different categories. Yes, you can use a pie chart too, but they are kind of boring. Include at least 3 of these single dimension plots. If you tell me your data only has 3 columns, I will probably say find a different dataset.

Next, make a multidimension plot using at least two columns, though as you have seen in labs, some 2-dimension plots can take three different columns to make.

The plots don't need to be "pretty" unless you want, but they do need context items that help me understand your data:

* All plots should have a title and possibly context text under the title so I know what the data is about.
* Include titles and axis labels in the plots.
* All plots should list the source in the lower right corner in small (but readable) text.

**Format of PDF**

The file should be a single pdf. The first page should include a title, your name, the source of the data, and a table of contents of the plots. Since I am asking for a title page and 3 single and one multi dimension plots, you should have a minimum of 5 pages, but you can make more plots if you wish, just make one page per plot.

Note that in this assignment the plots and their context text should be all I need to get a sense of what your data is about.

Upload this pdf with the file name: DataReport-<your name>.pdf

Also upload your R script.