## **Assignment 3 - Visualize data using R and Tableau**

**Due:** October 10, 2017

### **R**

**Data set**

We will use the [WDV survey data](https://bcourses.berkeley.edu/courses/1465668/files/71551328/download?verifier=cy9OGmzagagJZzj3cN1Zx7nvLyhKmZi6Ovc3Fqtn&wrap=1) from Assignment 0.

* Download [R (Links to an external site.)Links to an external site.](https://www.r-project.org/) and optionally [RStudio (Links to an external site.)Links to an external site.](https://www.rstudio.com/products/rstudio/#Desktop)
  + Use the instructions here
* Check out these R tutorials that will walk you through multiple charts and analyzing data in R:
  + [http://swcarpentry.github.io/r-novice-gapminder/04-data-structures-part1/ (Links to an external site.)Links to an external site.](http://swcarpentry.github.io/r-novice-gapminder/04-data-structures-part1/)
  + [http://www.datacarpentry.org/R-ecology-lesson/index.html (Links to an external site.)Links to an external site.](http://www.datacarpentry.org/R-ecology-lesson/index.html)
  + [http://flowingdata.com/2012/12/17/getting-started-with-charts-in-r/ (Links to an external site.)Links to an external site.](http://flowingdata.com/2012/12/17/getting-started-with-charts-in-r/)

1. Walk through the line charts section in[this (Links to an external site.)Links to an external site.](http://www.harding.edu/fmccown/r/) R tutorial
   1. Save a pdf or png for the final line chart and add it to the homework document
      * The last part line chart needs a file named auto.dat, which contains the data. You can download the file from [here](https://bcourses.berkeley.edu/courses/1465668/files/71811852/download?verifier=vqjzqlTj8aYGjQjqRuIXdGIymtr4fi6KwwhwJBBs).
   2. Include the R code used to generate the chart
2. Read the WDV survey data into R:
   1. Plot any **ordinal** data. Add a title for your graph and name axises
   2. Save a pdf or png for the plot and add it to the homework document
   3. Include the code you generated

Note that you can use the console or the R notebook from RStudio

File -> New File -> R Notebook

### **Tableau**

**Data set**

We will use the data from the [National Center for Education Statistics (Links to an external site.)Links to an external site.](http://nces.ed.gov/). I filtered and clean the data so you can work on it for the assignment. Here is the data: [SAT\_score\_states.xlsxreview the documentiew in a new window](https://bcourses.berkeley.edu/courses/1465668/files/71749556/download?verifier=eIqQBVkTogFvWqNe6Isw0QRV1HtdjfoWxu9lqi2M&wrap=1)

* Download Tableau Desktop (use [http://www.tableau.com/tft/activation (Links to an external site.)Links to an external site.](http://www.tableau.com/tft/activation)) and activate it with the code I posted here
* Walk through the [Get Started with Tableau Desktop tutorial. (Links to an external site.)Links to an external site.](http://onlinehelp.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-home.html) Here is also a video that shows how to visualize data using Tableau:

[https://www.youtube.com/watch?v=Z8NIsJckumI&feature=youtu.be (Links to an external site.)Links to an external site.](https://www.youtube.com/watch?v=Z8NIsJckumI&feature=youtu.be)

* We would like to determine if states with high student/teacher ratios had low SAT scores

1. Generate a graph (more than graph if possible) to explore the data and answer the question above
2. You need to look at the data for additional factors that might affect SAT scores and/or education quality. Be creative!
3. Take a screenshot of your final chart/graph and add it to the assignment file
4. Provide a short write-up about the visualization and what insights you discovered from exploring the data. What is the message that you are trying to present to the user?
5. How did you find Tableau or what is your feedback on Tableau?

**Grading:** I will evaluate if you did what the assignment asked and the quality of the graphs you will generate. 

**Submission:** Upload a pdf of your answers on bCourses.