# Homework (HW06) Data Viz: Using GGPLOT

### General Instructions

For this homework you will upload 1 R file into blackboard.

Reminder:

* All HW must start with an Identification Block like this sample…

################################################

# IST 387/687, Standard Homework Heading

#

# Student name:

# Homework number:

# Date due:

#

# Attribution statement: (choose the statements that are true)

# 1. I did this work by myself, with help from the book and the professor

# 2. I did this work with help from the book and the professor and these Internet sources: <provide the urls>

# 3. I did this work with coaching from <Name of another student> but did not cut and paste any code

# Run these three functions to get a clean test of homework code

dev.off() # Clear the graph window

cat('\014') # Clear the console

rm(list=ls()) # Clear all user objects from the environment!!!

# Set working directory

# Change to the folder containing your homework data files

setwd("~/MyDesktop/ISTX87/Homework")

This homework builds on our efforts from the Prep Exercise and depends on a careful read of Chapter 12 of *An Introduction to Data Science*. This homework uses one of the most powerful tools for data visualization in R - ggplot. Technically speaking it is ggplot2. Written by computer/data scientist Hadley Wickham, this “graphics grammar” tool builds visualizations in layers. This method provides immense flexibility, but *takes a bit of practice to master*. The person who wrote that last sentence may wish to revisit their definition of “a bit”!

### HW06

**Step 1: Explore the Merged Data – Understanding distributions**

1. Use the merged dataset created during this week’s prep work (mergedDF)
2. Create separate histograms using ggplot2() for the Pop2011, murder rate, assault and rape columns. Make sure each line of code is explained (comments) in terms of what it is doing. What parameter will you have to adjust to make the other histograms look right?
3. Create a boxplot for the Pop2011, and a different boxplot for the murder rate.
4. Create a block comment explaining which visualization (boxplot or histogram) you thought was more helpful (explain why)

**Step 2: Which State had the Most Murders – bar charts**

1. Calculate the number of murders per state
   1. *Hint: use the population and murder rate percentage from your new dataframe*
2. Generate a bar chart, with the number of murders per state  
   *Hint: use the geom\_col() function*
3. Generate a bar chart, with the number of murders per state. Rotate text (on the X axis), so we can see x labels, also add a title named “Total Murders”.
   1. *Hint: use theme(axis.text.x = element\_text(angle = 90, hjust = 1))*
4. Generate a new bar chart, the same as in the previous step, but also sort the x-axis by the murder rate from low to high
5. Generate a third bar chart, the same as the previous step, but also showing UrbanPop as the color of the bar

**Step 3: Explore Murders – scatter chart**

1. Generate a scatter plot – have Pop2011 on the X axis, the UrbanPop on the y axis, and the size & color represent the number of murders.

Be sure you’ve included the most useful resources that helped you learn how to use ggplot2. Please note, your instructor should not be your first or only resource. Remember, this course is about learning how to learn!

***You must submit all Homework to blackboard prior to the deadline specified for each assignment.***

Late HW assignments will not be accepted for credit.

*Materials provided by the instructors of this course are copyrighted 2018 by Jeffrey Saltz and Jeffrey Stanton, or by the respective instructor that produced them. These materials are provided to enrolled students for their exclusive use during the semester of instruction. Uploading instructor-produced materials to Internet sites (e.g., Coursehero) without permission is considered a breach of professional ethics and will be reported as a possible academic integrity violation.*