Choose a topic and use at least one data set to analyze. Use R to analyze the data. Create an R Markdown file to share your findings, which you will also have the opportunity to talk through with your peers. Your analysis should include the following:

1. A data-backed claim.
2. Calculation of at least 5 key statistics. Examples include: mean, median, mode, correlation, variation, and standard deviation.
3. An interpretation of the statistics calculated.
4. At least 4 graphs.

We’ll be looking for the following:

1. Your use of R. This portion of the grade will be based on:
   1. The quality of your comments.
   2. The demonstrated ability to write your own function.
   3. The demonstrated ability to work with data frames.
2. The quality of the conclusions of your analysis.
3. The quality of your explanations to make the problem and your conclusions understandable.
4. Your use of graphs to visualize data.

You’ll have 15 minutes to present your project to your colleagues – this is *not* a formal presentation; you’re just showing us what you’ve learned and what cool tricks you used to get there.

Things to keep in mind:

1. Data sets are often incomplete. Make sure to look for that and handle it when it comes up.
2. R will often convert data into strings if it is not sure how to handle it. Use functions like as.integer() and as.numeric() to address this problem and do not forget the argument na.rm = TRUE.
3. You should not create your own dataset. You can find many data sets online. Some websites you may consider for data include Kaggle, wprdc.org, and data.gov.
4. Pick a topic that interests you, and try to learn something about the topic. Do not just create graphs – use the tools in R to uncover insights about the data. Ask interesting questions then use analytics to answer those questions!
5. I recommend looking back at your notes and the materials on blackboard to review the methods we covered in class – you have lots of tools at your disposal!