**Section 1 – Week 9 – Getting Started**

* Provide an introduction that explains the problem statement you are addressing. Why would someone be interested in this?
  + For dog people such as myself, a significant factor in deciding where to live is how dog-friendly a particular location is.
  + This is a significant problem when trying to find a place in a big city. Many landlords will not allow dogs, or not allow them over a certain weight, or not allow certain breeds.
  + New York City has plenty of good data sets on dogs, so this study will focus on that area.
* Draft 5-10 Research questions that focus on the problem statement.
  + How do restricted breeds correlate with dog bites in NYC?
  + How does a dog size correlate with dog bites in NYC?
  + How does dog size correlate with commonly restricted dogs in NYC?
  + What are popular dog breeds in NYC?
  + Do people prefer large or small dogs in NYC?
  + How easy is it to find a place to walk your dog in NYC?
* Provide a concise explanation of how you plan to address this problem statement.
  + I plan on finding data sets on dogs in NYC, and finding correlations between and within the different data sets. Given the correlation or lackthereof, analysis will be performed that will attempt to explain the results observed.
* Discuss how your proposed approach will address (fully or partially) this problem.
  + My approach will help to determine if NYC is a dog-friendly city
* Do some digging on a dataset that you can use to address the issue.
  + <https://catalog.data.gov/dataset/nyc-parks-dog-runs>
    - Locations and details of off leash dog runs in NYC
    - KML shows polygons of the boundary of the dog runs
    - Includes zip codes as well
  + <https://catalog.data.gov/dataset/dohmh-dog-bite-data>
    - Each entry is a single bite
    - Records breed, date, age, gender, spay/neuter, Borough, and zip code.
    - The zip code may be relatable to the dog bite issue
  + <https://catalog.data.gov/dataset/nyc-dog-licensing-dataset>
    - This will help to serve as a control for the other data sets.
    - (for example, say 99% of dogs are a certain breed, which would affect the probability that it would show up in the list of dog bites.)
    - Helpful information for distribution
* Identify the Packages that are needed for your project.
  + ggmap
  + ggplot
  + Rgooglemaps
* What types of plots and tables will help you to illustrate the ﬁndings to your research questions?
  + Scatter plots
  + Histogram
  + HeatMap
* What do you not know how to do right now that you need to learn to answer your research questions?
  + I need to find a way to visualize geo-spatial data. There will likely be areas of NYC that are better for being a dog owner, and it would be good to know what those areas are. Maybe it would be broken down by zip code or something.