**Week 3**

-will need to have one univariate numerical plot for our final

-in R, go to search in the bottom left folder area to find what you cloned from github

* **Counting** 
  + Expresses numerical values
  + May be tempted to use bar plots; discrete values
  + With too many values, bar plot is not a good idea because it is difficult to read and highlight what you want to show
    - Schools with reduced lunch is 172
  + Sometimes, aggregate
    - Use data, use reduced lunch count but aggregate by county (count by number of schools by each county, sum those values)
    - Makes it into 39 observations
    - Can then count in order by decreasing value
    - ADDING NEW COLUMNS; what is the percentage that counted variable represents for the whole? Can add percentage of what the count variable (county) has for the value you care about (reduced lunch)
    - Then can do this cumulative count for a bar plot
      * Using cumulative column; see counties in descending order—this is ordering by alphabet of the counties (not of number of reduced lunch schools)
      * Still not the most informative, change the parameters (scale\_x\_discrete) scale so the order is reduced.lunch.cumulative
      * Change colors—if county is below 80%, have one color, if not have another color *(that is why creating this column was important)*
        + From all counties, you can see how many represent 80% of reduced lunch schools in those counties
      * FOR PROJECT—stop at the line of code at 130 (the other thing he did after is just extra to show what you can do)
    - Installing packages
      * Right side, install and search for what you want
  + Now trying to look at actual values, not just the cases per county
    - Organize into intervals—look at summary statistics (of raw data)
    - Median—50% of the schools at most, offer at most 25.50 students with reduced lunch
      * Standard deviation bigger than the mean (36.55)
    - 75% of schools do not benefit more than 47 students
    - Max is 301—that shows some pretty high inequality
    - Histogram shows the skewed data; then run a code to add 4th value (the mean from the statistics)
* **Measurement** 
  + Tells you the times a particular unit is present in the unit of analysis: allows for the presence of decimal places
  + This is not about counting, it is about measuring something (DO NOT USE BAR PLOTS)
    - Ex. Student/teacher ratio
  + Common to have outliers, may mean the existence of inequality
* ASSINGMENT
  + One plot prepared for our dashboard (can be any that we have discussed so far), with data that we plan to use

**Call with Brain; Loaded Library**

* Tidyverse loads a bunch of packages
* Read\_csv—is reading that file in (but you needed to have tidy verse in)
  + Best to save excel files as csv (save as csv)
  + Save in folder to visual analytics—created class folder with R Project—everything he does in project dumps into that folder
* Df: data that you want and you can change this; and then the whole thing entire below auto runs

Total Number of Projects in each phase

-check in with Amara about specifics of it all

-what do the data mean?

-where are the performance reference?

-get information from the Brain and see what we can get

**BIG PICTURE QUESTIONS**

-SCHEDULE: Where are we getting hung up? What phases are most responsible for delays

-FINANCE: what projects are approaching gates

-What projects are within 3 months of approaching gate 1/2/3/etc?

-SCHEDULE/FINANCE: How staff members are charging hours (power bi dashboard)

-staff hours, financials,

-How contractors are doing submittals

If we are trying to offload to county to make this a valuable resource; may have to replicate this in Looker BI

-automated report that we are pulling from Looker BI is the ultimate goal

**Create a New Notebook**

Bring your data in

Do your plotting

Output is one plot; save as assignment 1 in the folder of the repo in your computer (rmd file)

Preview, knit to HTML, pulls up a webpage with code and plot—that is what professor will check in our repo (from github)

**Feb 14 Class**

One univariate, one bivariate, one map (spatial)—this is what is due in the final dashboard

Dashboard

* Flex dashboard, easiest to develop
* Packages- flexdashboard; install
* File, new file, R Markdown, from template, flex dashboard, ok
  + New data screen pops up
  + You copy and paste what you did to create your plots and they will then come up in the dashboard
  + Name of Chart—have to put the name in
* saveRDS (dataCSV, file= ‘dataCSV.rds’); he recommends saving as RDS
* source code, embed—allows you to have the code you used to make the plots up at the top of your html file
* Dashboard: good titles, good colors, good names of axis, subtitles, showing the right things
* No class next week, working to create dashboard and then can submit any time