NCQA Data Goals:

1. Set up a folder structure in the project folder, with folders for
   1. raw\_data
   2. processed\_data
   3. scripts
   4. figures
   5. tables
   6. Rmd\_documents
2. And a single R script at the top (folder) level, titled run\_all.R, which consists of source commands, like
   1. Source(‘01\_script.R’)
   2. Source(‘02.script.R’)
   3. Source(‘03.script.R’)
   4. Etc.
3. Save old script as old\_script.R
4. Break into chunks that are really well commented, so that future you will understand them
   1. 01\_setup.R **PDRH**
      1. Includes libraries, functions, constants
      2. Statenames vector, state abbreviations
      3. Link to website
      4. Plantypelist, instypes
   2. 02\_read\_ncqa\_data.R **PDRH**
      1. Loops to read in 2015, 2016, 2017, 2018
      2. Save each as a separate csv file in processed\_data
      3. Outputs 02p\_ncqa2015.csv, etc.
   3. 03\_read\_kff\_data.R **JB**
      1. Read in, clean KFF data variables
      2. Remove PR, DC
      3. Save as 50 rows for 50 states
      4. Name each variable as varname\_year, so that we know what year derived from
      5. Save as csv in processed\_data
      6. Output 03p\_kff\_data.csv, etc.
   4. 10\_clean\_ncqa.R **MM**
      1. Read in 2015, clean
      2. Read in 2016, clean
      3. Read in 2017, clean
      4. Read in 2018, clean
      5. Assign plantypes
      6. Remove junk lines, label families, Blue as needed
      7. Save all 4 as distinct csvs in processed\_data, 10p\_ncqa2015data.csv, etc.
   5. 20\_viz\_distributions.R **MM**
      1. Visualize distribution of insurance satisfaction
      2. Violin plots plus geom\_jitter
      3. Facet by instype and plantype (facet\_grid)
      4. Save numbered figures as jpeg with ggsave, descriptive name in figures folder, 20p\_F1\_name.jpg
   6. 30\_setup\_choropleths.R **JB**
      1. Read in data for 2015, 2016, 2017
      2. Left\_join with region, pop, abbreviation
      3. Loop to get mean commercial rating by state
      4. Left join to states
      5. Save in 30p\_processed\_data.csv
   7. 31\_map\_choropleths.R **JB**
      1. Read in processed ncqa data
      2. Make choropleths for 2015, 2016, 2017, 2018 with geom\_sf
      3. Can do with geom\_sf and facet
      4. Use patchwork package to make a 4 panel plot, with tag by year
      5. Save numbered figure as jpeg with ggsave, descriptive name in figures folder, here(‘figures/1p\_F2\_year\_maps.jpeg’)
   8. 40\_join\_kff.R  **PDRH**
      1. Read in 2016 data from ncqa
      2. Read in kff data
      3. Left join
      4. Check, clean
      5. Test to confirm no missing
      6. Save as csv with descriptive name in here(‘processed\_data/40p\_kff\_ncqa\_joined\_data.csv’)
   9. 50\_correlations.R  **PDRH**
      1. Set up correlations of predictor vars (remove outcome var [ins sat])
      2. Make correlation matrix or corrplot
      3. Look for correlations >0.4 before modeling
      4. Save correlation grid or plot in figures folder
   10. 51\_modeling.R  **JB**
       1. Try multiple models, find best model to predict insurance satisfaction
          1. Include state level predictors, instype, plantype (not state)
       2. Watch out for interactions
       3. Use leaps to help
       4. Make a nice table of coefficients and p values with broom::tidy
       5. Make the table pretty with flextable OR gt, export to word
       6. Save to tables folder
   11. 52\_predictions.R  **PDRH**
       1. Predict insurance sat for each plan
       2. Create residuals (Actual – predicted) for each plan
       3. Arrange by residuals – high to low
       4. Make a nice looking table (use flextable package)
          1. Save to tables folder
       5. Make a version with only the top 10, bottom 10
          1. Save to tables folder
   12. 60\_word\_output.Rmd **MM**
       1. Summarize basic points, bullet points
       2. Include all figures, tables for paper
       3. Make Rmd document output to word
   13. 61\_ppt\_output.Rmd **MM**
       1. Summarize basic points, bullet points
       2. Include all figures, tables for paper
       3. Make Rmd document output to word
   14. 70\_run\_all.R **MM**
       1. Update top level script run\_all.R