Steps:

President Example

* Download datasets. ‘Data/Presidents’ folder. president2008.csv, president2012.csv, president2016.csv, president2020.csv are data files.
* Import all csv files as datasets.
* Merge all 4 datasets together by fips code to make master file.
* Subset dataset. Keep rows where variable ‘type’ is ‘County’.
* Create new variables in dataset that are proportion of 2008 democrats’ total county votes, proportion of 2008 republicans’ total county votes and proportion of 2020 republicans’ total county votes respectively.
* Compare democrats and republican proportions through t.test, specifically for the year 2008. Use variables created in previous step.
* Compare proportion of republican voting percentage from 2008 to 2020 using t.test
* FTest comparing standard deviations of democratic votes and republican votes specifically for 2008.
* Data is in wide format. Convert to long format. 1 column for democrats, 1 column for republicans, 1 column for year instead of variables. Use code online.
* Compare average proportion of democrats to republicans for all years. T.test comparing average statistic by group.
* Compare democrats and republican proportions for 2008 again through t.test now that data is in long format.
* Create overlapping line graph showing change in democrats and republican voting percentage from year to year.
* Save Graph
* Create dataset/table that shows total sum of democratic and republican county votes by year. Along with average proportion of votes to democrats and republicans. Tidyverse, group\_by
* Create log and script files to track code and output. Make sure code runs in full with no errors.

SAT Example

* Will perform conditions for regression equation. Look up assumptions. Predict csat from expense, percent income high college variables
* Download data. Data/csat
* Import dataset.
* Examine contents of dataset. Variable names, variable types. Descriptive statistics.
* Run preliminary regression equation. Is there anything wrong with the output.
* Create histograms to look at distribution of independent and dependent variables.
* Any missing values? Any outliers? Any other errors with data? Fix them.
* Perform correlation matrix of all important variables. csat expense perc income high college
* Perform scatterplots of independent variables with csat. How do the relationships look.
* Apply variable transformation for better regression results. Perform square and log transformation for Percent variable. Which gives the best results?
* Perform final regression and export results to neat table.

Weather Exercise

* Download data. Keep same folder structure.
* Set directory to data folder.
* Import and combine each dataset in folder. Combine in vertical fashion. Append data.
* Create 3 new variables: convert degrees for tMin, tMax from Celsius to farenheit. Calculate average daily farenheit temperature using newly created tMin and tMax farenheit variables. How to differentiate between 1 dataset to the next.
* Reduce dataset to show average tMin, tMax, TAvg for each day each fips code. 1 row per day.
* Perform one way anova to test if average temperature is equal for all fips.