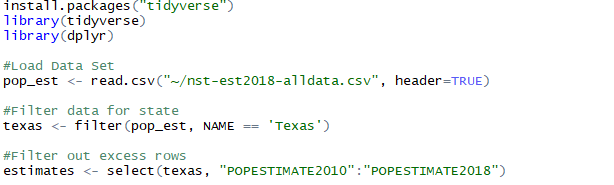
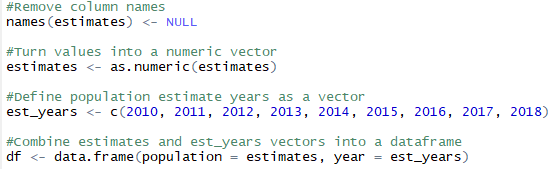
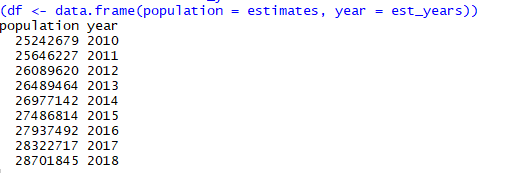
Project C996

1. A linear regression analysis is created with R for use in predicting the size of the population for the state selected, based on the Current Estimates Data dataset.
   1. First, I uploaded the packages I needed to complete the project: tidyverse and dplyr
   2. I then used read.csv to upload the csv file
   3. I selected Texas in the filtering criteria and then reduced the columns with the select function
2. Data was prepared and imported into R and includes a screenshot of the results

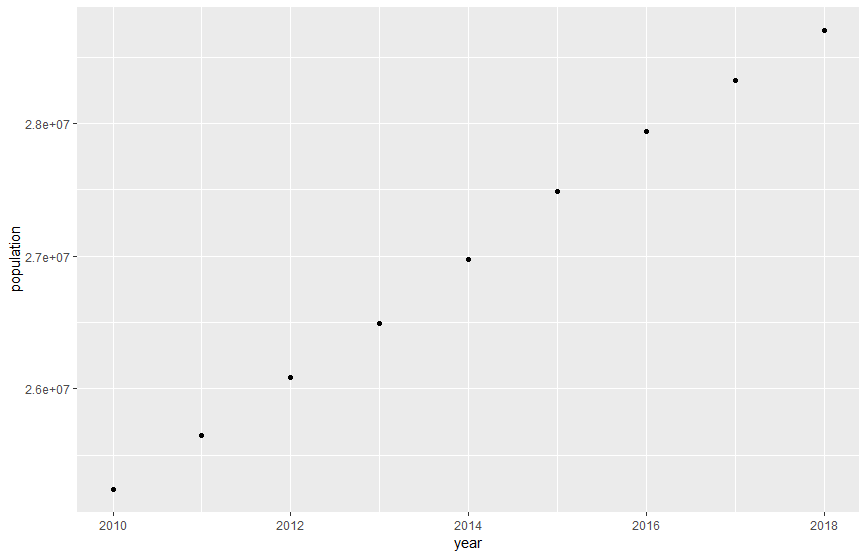


* 1. I removed the column names in the Texas dataframe with the names as NULL function and then built a vector with the as.numeric function
  2. I then created an est\_years vector. After this, I merged the two vectors into a dataframe so I could start visualizing the data in ggplot
  3. Once I was complete, the dataframe was neatly organized into two columns called population and year

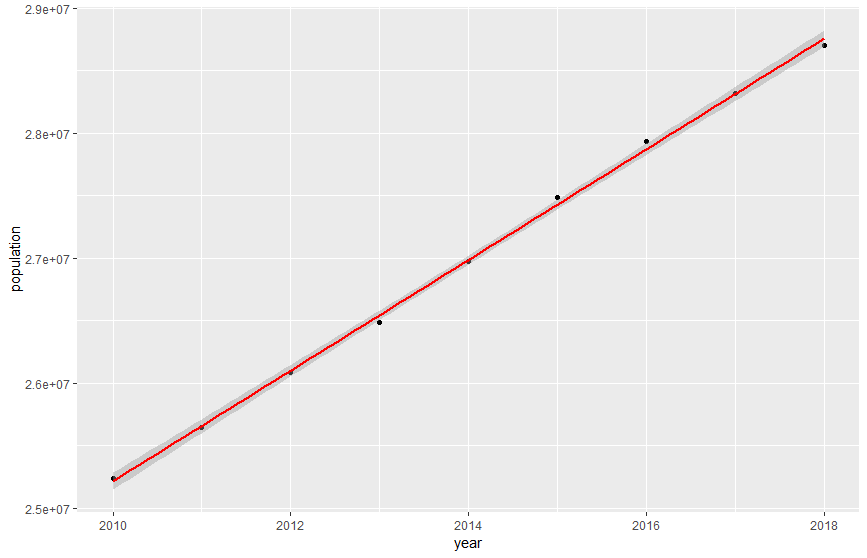
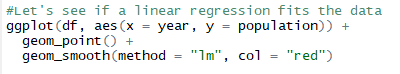


1. A Linear Regression Model is created to accurately portray the dataframe. The summary method is utilized for a statistical description of the model.
   1. I first use a scatterplot with geom\_point to show the data as is and try to determine the best model to represent the information with

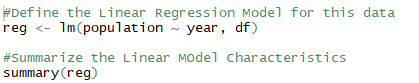


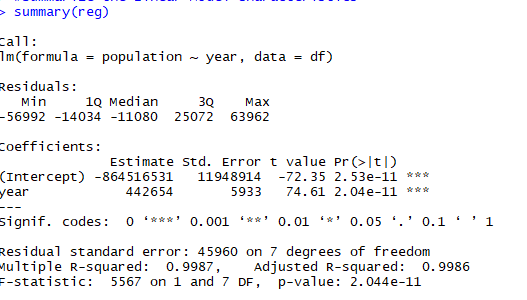


* 1. From the scatterplot, it is evident that a linear plot would fit this data best. I utilized geom\_smooth with lm to create the linear model in a red line across the data points.



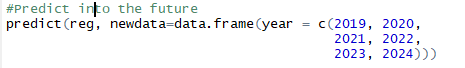
* 1. I then defined the linear model in order to utilize the summary statistics function to analyze the data



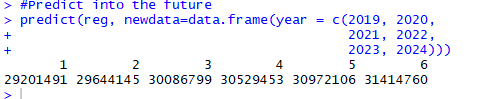


* 1. This statistical summary shows how well the model represents the data. A key indicator is the p value. It is far smaller than the standard .05 allowance. The r-squared value is quite high as well, showing the model accounts for 99% of the variance in the data.

1. Predict the population size of Texas in 5 years, based on the data points for each year.



* 1. I used the predict function with the linear model and prediction years to show the progression of the estimated population for the next five years



* 1. The values show very reasonable estimates for the following years and seem relatively reliable. The main factor to consider before trusting these estimates is that the data we are using is a set of estimates in it’s own right.