STAT 184 Final Project

Prajwal Bhandari, Sam Johnson, Jack Mariani

Introduction

For our final project our group decided to explore data relating to previous presidential elections and explore questions relating to it. We explored a couple of topics before landing on our key topic. The reason we decided presidential election data is because it is easily accessible; it is public information. Also we reasoned that we could explore a plethora of different questions to explore.

Before going into the questions we explore we must first find the data we wish to use and tidy it so that it is accessible in "R". We found our data on presidency.ucsb.edu, a website whose purpose is giving presidential data to students across the country to use and interpret. We came across an issue after choosing this website. We wanted to use the results from multiple elections and we had a hard time finding an efficient way to import this data. We ended up using the paste0 function to import each year into R.

After importing the data into R we noticed slight differences within the data from year to year, the tables were not all tided properly, some had extra rows above the data. Also we aimed to look at data based on party and the data frames did not have each party's candidate in the same position each time, the winner was listed first. For example, in 1988 the data frame listed George Bush first, a Republican. The next Election Bill Clinton, a democrat won the election and was put first in the data frame. To get around this we went through each to make sure each data frame had each party in the same order every time. We had to create several functions to wrangle the data in order to start making assumptions with the data. This process took a majority of our time, but when we finally got them all the same we took the data from each year and combined them into one major data frame so we can look at the data overall the entire time interval.