1. The first thing required for this project is to gather the data. The list of Presidents is available on Wikipedia at https://en.wikipedia.org/wiki/List_of_Presidents_of_the_United_States#Presidents. There are two ways of getting this data.

- The Ideal Way: The best way of fetching this data would be to use BeautifulSoup4 or a similar library to parse the HTML and make a tabular structure and load it in a DataFrame using read_html() method. This way, in case if the Table updates, we can be sure that our data will update along with it and would not require manual intervention, thereby making it almost a 'no touch' way of gathering data
- The Easy Way: This is manually interpreting the Data on the Wiki page, and creating a CSV by hand. This saves us hours of coding and HTML parsing, and we get straight to business. In a research environment, where the Dataset may not be under constant use, and a small sample set should suffice, this method is ideal. However, we are at the risk of our data being outdated, and therefore, our findings being skewed to one side.

We use method #2 because by the time we end this Mini Project, the dataset wouldn't change drastically (if at all), and given the fact that we don't need to constantly analyze this dataset, it need not be updated all the time by using some esoteric HTML parsing and read_html() methods.

2. Pandas was used to load the CSV into python's memory space, keeping a tabular structure. Since the CSV was manually made, it wasn't a problem to load the CSV in desired shape/dimensions, since we intuitively sanitized that data's shape before we bought it to python. We can take the case here that there are, at times, small tweaks that can be made out of the python environment to give us a head start in our research endeavours.

3. Downloading the Data proved slightly challenging, since I could not use my regular method of fetching data from Yahoo Finance. Hence, Quandl was used, but Quandl's DJIA index didn't have 2017's data. Therefore, there is no analysis with respect to the recent Republican President. (2017-Present). Quandl was easier to get Data from, though that 1 year's missing dataset in their free tier made me slightly uneasy. What was most interesting to see was the 'collapse' option that gave me the ability to have a year-wise dataset, which saved me a bunch of daily returns to monthly to yearly returns calculations (though, with a daily dataset, this could have also been done in one line, just by changing the pandas 'pct_change' functions parameter to the number of trading days in a year)

4. Calculating the yearly returns from '1920' onwards was taken in quite the literal sense. That is, the start year was 1920, and since I don't have 1919's data, the annual return on this returned NaN. Therefore, this row was dropped. This later also proved to be an accidental convenience, since Harding's term began right in 1921.

5. The segregation of presidency on the basis of party was a slight challenge. I had to convert the indices in the president's dataframe to datetime objects, then check in an O(n^2) for loop, for each index, whether it lied between every presidents tenure. Since we have annual returns in the form of, say, 2003-12-31, a persons presidency could start in 2000 and end in say, August of 2003, and another president starts his tenure in August of 2003. The annual returns for the year 2003 will be counted for the later president and not the former. This was something I observed could skew the data slightly. Luckily for us, such instances are very rare in our dataset, and doesn't lend itself heavily to skewing our data. Though this caveat was noted, and such caveats must be presented while approving research.

6. Pandas' describe() method gives all the necessary data for calculating Central Tendency. It gives mean, 50% Quantile (which is Median), Standard Deviation (whose square is variance). The other

aspect was we used describe() on both the democratic filtered DataFrame and the Republican filtered DataFrame. A combination of this DataFrame was necessary for the next part.

7. The aforementioned describe() dataframes for both democrat and republican were combined so that we could plot it better. A barplot was chosen because at one glimpse it conveys the overall picture, being that **stock markets were more bullish (performed better) during democratic years over republican years**. This is seen easily in the graph, since The mean and median of democratic years is higher than republican years. Also, markets tend to be more volatile during republican years as indicated by the variance during republican years (We have a republican in power, maybe market turbulence could be used as an advantage now)