US 2020 vs 2024 Election Results Analysis

2248-ASDS-5301-002-STAT THEORY AND APPLICATIONS

GROUP-4 PROJECT REPORT

BY

SHALINI DRONADULA, AUSTIN STEEL, KARUNAKAR KAVALA

Contents

1.INTRODUCTION

1.1 PROBLEM STATEMENT

The U.S. presidential elections of 2020 and 2024 provide a unique opportunity to analyze shifts in voting patterns and party performance across states. While both elections featured common voter concerns like immigration, abortion, and the economy, they also differed in their contextual challenges—COVID-19 in 2020 and a late Democratic ticket change in 2024. This analysis aims to determine whether there were statistically significant differences in vote counts and percentages for the Republican, Democrat, and Other parties between the two elections. By comparing state-level data using statistical techniques, the study seeks to uncover trends that highlight changes in voter preferences and the evolving political landscape of the United States.

1.2 OBJECTIVES

This project's goal is to find statistically significant differences in the ways that the Republican, Democratic, and other parties voted in the 2020 and 2024 U.S. presidential elections. The analysis uses paired t-tests to determine statistical significance and focuses on changes in vote counts and percentages using state-level data. The objective is to identify significant patterns in voter behavior, draw attention to changes in party support, and offer insights into how American politics are changing.

1.3 SCOPE

The primary goal of this project is to examine state-level voting trends in the 2020 and 2024 U.S. presidential elections. Examining the Republican, Democratic, and Other party vote totals and percentages in each of the 50 states and the District of Columbia is part of the scope. The study finds statistically significant trends and changes in voter behavior over the course of the two elections using paired t-tests. Because the analysis is restricted to NBC News data, the results are guaranteed to be accurate and consistent. The project's insights are intended to improve comprehension of changes in political preferences and add to conversations about the changing electoral landscape in the US.

2.DESCRIPTIVE ANALYSIS

This project analyzes voting trends in the 2020 and 2024 U.S. presidential elections using a custom dataset created by our team. The dataset includes vote counts and percentages for the Republican, Democrat, and Other parties across all 50 states and Washington D.C., totaling 52 rows and 15 columns. Initially, the dataset contained raw data that included both state-level details and summary rows for national totals. This dataset was specifically created by our team to examine voting behavior and identify trends in party performance over the two election years, with a focus on whether shifts in votes and percentages occurred across different states.

Before conducting any analysis, the dataset underwent a cleaning process to remove irrelevant rows and columns unintentionally imported, focusing solely on state-level data. New columns were added to capture the differences in vote counts and percentages for each party between 2020 and 2024. The dataset was also checked for missing values, and no missing data was found, ensuring the integrity of the analysis. Initial exploratory data analysis (EDA) was performed using visualizations such as histograms and boxplots to assess the distribution of votes and to check for any anomalies or outliers. This provided a clearer understanding of the data's structure, ensuring that it was ready for further statistical analysis.

The CONTENTS Procedure

Data Set Name	WORK.ELECTION_DATA	Observations	53
Member Type	DATA	Variables	15
Engine	V9	Indexes	0
Created	12/04/2024 20:43:12	Observation Length	136
Last Modified	12/04/2024 20:43:12	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
3	2020 D %	Num	8	PERCENT12.2		2020 D %
2	2020 D Votes	Num	8	COMMA15.		2020 D Votes
7	2020 Other %	Num	8	PERCENT12.2		2020 Other %
6	2020 Other Votes	Num	8	BEST.		2020 Other Votes
5	2020 R %	Num	8	PERCENT12.2		2020 R %
4	2020 R Votes	Num	8	COMMA15.		2020 R Votes
8	2020 Total Votes	Num	8	COMMA15.		2020 Total Votes
10	2024 D %	Num	8	PERCENT12.2		2024 D %
9	2024 D Votes	Num	8	BEST.		2024 D Votes
14	2024 Other %	Num	8	PERCENT12.2		2024 Other %
13	2024 Other Votes	Num	8	BEST.		2024 Other Votes
12	2024 R %	Num	8	PERCENT12.2		2024 R %
11	2024 R Votes	Num	8	BEST.		2024 R Votes
15	2024 Total Votes	Num	8	BEST.		2024 Total Votes
1	State	Char	20	$20.	$20.	State

3.DATA PREPROCESSING

Before any statistical analysis could be performed, the dataset underwent several preprocessing steps to ensure it was clean, consistent, and ready for analysis. The goal of this preprocessing phase was to transform the raw data into a format suitable for statistical tests and to ensure that all the necessary information was accurately captured.

3.1 DATA CLEANING

The first step in the preprocessing process was cleaning the dataset. The raw data initially contained several issues that could potentially affect the analysis. These included irrelevant rows such as summary totals and blank entries, which did not contribute to state-level comparisons. To address this, all rows with aggregated totals (e.g., "Total") and any empty rows were removed, ensuring the dataset focused solely on state-level information.

3.2 ADDING NEW COLUMNS

To facilitate the comparison of vote changes between the two election years (2020 and 2024), new columns were added to the dataset. These columns calculated the differences in votes and percentages for each political party across the two years. The new columns included:

- **Change in Republican Votes**: The difference in the number of Republican votes between 2024 and 2020.
- **Change in Democrat Votes**: The difference in the number of Democrat votes between 2024 and 2020.
- **Change in Other Votes**: The difference in the number of votes for third-party candidates between 2024 and 2020.
- **Change in Total Votes**: The overall difference in total votes between the two elections.
- **Change in Republican Percentages**: This column calculates the change in the percentage of votes received by the Republican party from 2020 to 2024.
- **Change in Democrat Percentages**: Like the Republican percentage change, this column calculates the change in the percentage of votes received by the Democrat party from 2020 to 2024.
- **Change in Other Percentages**: This column calculates the change in the percentage of votes received by third-party candidates between 2020 and 2024.

3.3 CHECKING ROWS AND DATA INTIGRITY

After importing the dataset, it was essential to ensure that all rows were correctly imported, and no data was omitted or misformated. This step involved a thorough review of the dataset to verify that every state and its corresponding data were correctly represented. The dataset was checked for any anomalies in formatting or missing values.

3.4 HADLING MISSING VALUES

Fortunately, the dataset did not contain any missing values, which meant no imputation or additional data handling was required. Having a complete dataset

ensured that the analysis would be based on full, consistent data without any gaps that could distort the results.

The MEANS Procedure

Variable	Label	N	N Miss
2020 D Votes	2020 D Votes	51	0
2020 D %	2020 D %	51	0
2020 R Votes	2020 R Votes	51	0
2020 R %	2020 R %	51	0
2020 Other Votes	2020 Other Votes	51	0
2020 Other %	2020 Other %	51	0
2020 Total Votes	2020 Total Votes	51	0
2024 D Votes	2024 D Votes	51	0
2024 D %	2024 D %	51	0
2024 R Votes	2024 R Votes	51	0
2024 R %	2024 R %	51	0
2024 Other Votes	2024 Other Votes	51	0
2024 Other %	2024 Other %	51	0
2024 Total Votes	2024 Total Votes	51	0
Change_D_Votes		51	0
Change_R_Votes		51	0
Change_Other_Votes		51	0
Change_Total_Votes		51	0
Change_D_Percent		51	0
Change_R_Percent		51	0
Change_Other_Percent		51	0

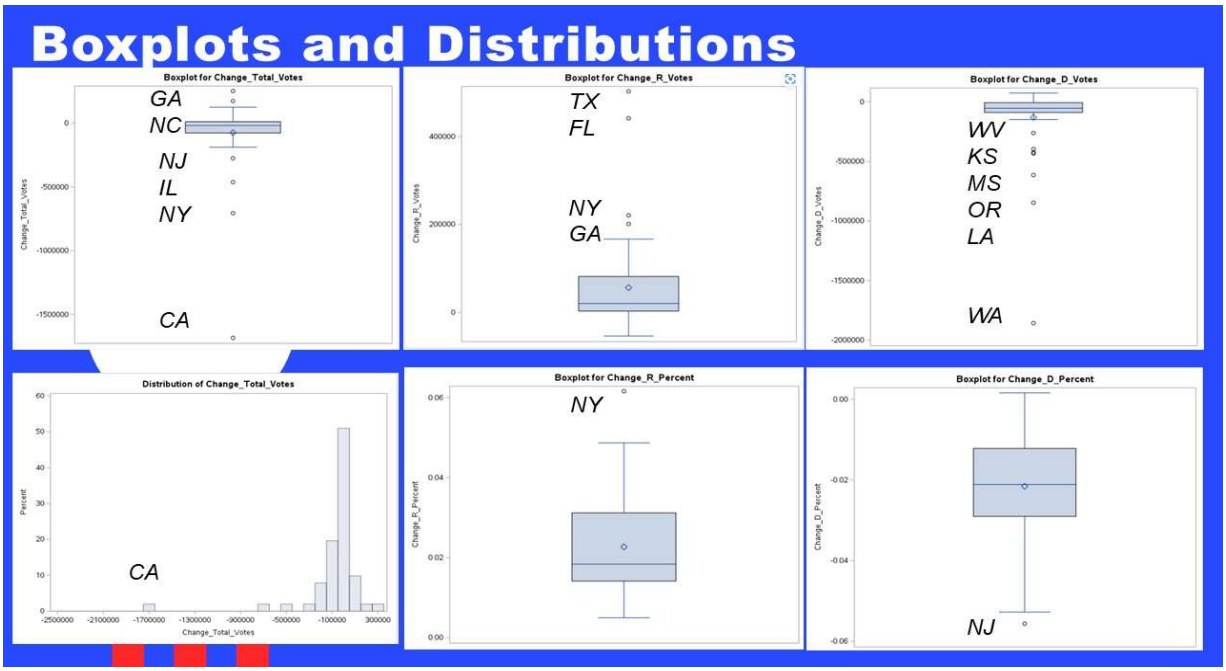
3.5 EXPLORATORY DATA ANALYSIS

Once the data was cleaned and verified, initial exploratory data analysis (EDA) was performed. The purpose of EDA was to understand the distributions of the variables, check for outliers, and create several initial visualizations.

The MEANS Procedure

Variable	Label	Mean	Std Dev	Minimum	Maximum	Median
2020 D Votes	2020 D Votes	1593794.14	1916171.84	73491.00	11110639.00	856034.00
2020 D %	2020 D %	0.4863000	0.1204811	0.2655000	0.9215000	0.4936000
2020 R Votes	2020 R Votes	1455372.06	1413252.57	18586.00	6006518.00	1020280.00
2020 R %	2020 R %	0.4920353	0.1198747	0.0540000	0.6994000	0.4924000
2020 Other Votes	2020 Other Votes	57297.16	59618.78	7475.00	384223.00	36685.00
2020 Other %	2020 Other %	0.0216588	0.0075623	0.0092000	0.0439000	0.0200000
2020 Total Votes	2020 Total Votes	3106463.35	3306713.21	276765.00	17501380.00	2148062.00
2024 D Votes	2024 D Votes	1449497.84	1620942.76	69527.00	8937402.00	769391.00
2024 D %	2024 D %	0.4658824	0.1180899	0.2610000	0.9250000	0.4750000
2024 R Votes	2024 R Votes	1500784.12	1474654.68	20888.00	6375376.00	1036213.00
2024 R %	2024 R %	0.5159216	0.1185767	0.0670000	0.7230000	0.5100000
2024 Other Votes	2024 Other Votes	49385.66	68767.77	0	471270.89	31022.54
2024 Other %	2024 Other %	0.0181961	0.0078001	0	0.0410000	0.0170000
2024 Total Votes	2024 Total Votes	2999667.62	3081053.33	266422.76	15202286.89	2068169.54
Change_D_Votes		-144296.29	336594.01	-2173237.00	74381.00	-53746.00
Change_R_Votes		45412.06	111017.86	-212904.00	485029.00	14235.00
Change_Other_Votes		-7911.49	23632.01	-119978.00	87047.89	-6625.46
Change_Total_Votes		-106795.73	350186.26	-2299093.11	253189.19	-31512.25
Change_D_Percent		-0.0204176	0.0137984	-0.0533000	0.0035000	-0.0181000
Change_R_Percent		0.0238863	0.0121109	0.0043000	0.0636000	0.0213000
Change_Other_Percent		-0.0034627	0.0055567	-0.0191000	0.0091000	-0.0029000

- **Boxplots:** To identify any outliers in the change in vote data.



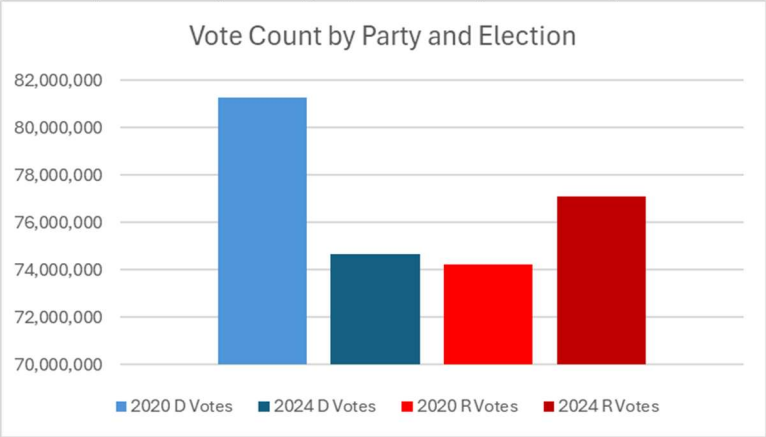
From these boxplots we can see Georgia and North Carolina were outliers with an increase in Total votes, where New Jersey, Illinois, New York, and California were outliers with a decrease in Total Votes. California is notably the largest change in Total Votes. However, California is also the most populous state, and has had a net decrease in population for the first time in the last few years. Therefore, this would make sense.

For the difference in Republican Votes we only see outliers with increases like Texas, Florida, New York, and Georgia. For the difference in Democrat Votes we only see outliers with decreases like West Virgina, Kansas, Mississippi, Oregon, Louisiana, and Washington.

Because the populations vary dramatically from state to state, it is important to consider the percentage changes of party votes as well. Looking at these boxplots, we see only New York as an outlier for increase in percentage Republican Votes, and New Jersey as and outlier for decrease in percentage Democrat Votes. These two outliers are particularly interesting because both New York and New Jersey have been considered solid Democrat states for the last several elections. What we see here is that these two states changed the largest away from that trend in terms of percentage vote.

- **Total Vote Tables and Graphs**

	D Votes	R Votes	Other Votes	Total
2020	81,283,501	74,223,975	2,922,155	158,429,631
2024	74,666,439	77,100,099	2,812,712	154,579,250
Change	(6,617,062)	2,876,124	(109,443)	(3,850,381)



This table shows us the total number of votes for each party in each election and the change at the bottom. Overall, we see that Democrats received 6.6 million less votes in 2024 than 2020, and Republicans received 2.8 million more votes. The Total vote count across all parties and states went down by 3.85 million votes. While many interpretations can be drawn as to why this may be, most simply it appears that enthusiasm for Democrats was much lower this election than in 2020, and enthusiasm for Republicans was much higher in 2024 than 2020.

From these boxplots, tables and graphs we can see there is some difference in voting patterns between the two elections. EDA helped identify the overall patterns in the data and highlighted any potential issues, such as skewed distributions or outliers, that needed to be addressed before conducting statistical tests.

4. PAIRED T-TESTS

To assess whether there were significant changes in voting patterns, we will perform a paired t-tests for each variable we can match between the two elections. This means we will do seven paired t-tests. We are looking to see if the mean is the same in each variable in the respective elections. For example, for Democrat Votes, the null and alternative hypothesis are:

H_0 = There is no significant change in the mean number of Democrat Votes between 2020 and 2024 ($\mu_1 = \mu_2$).

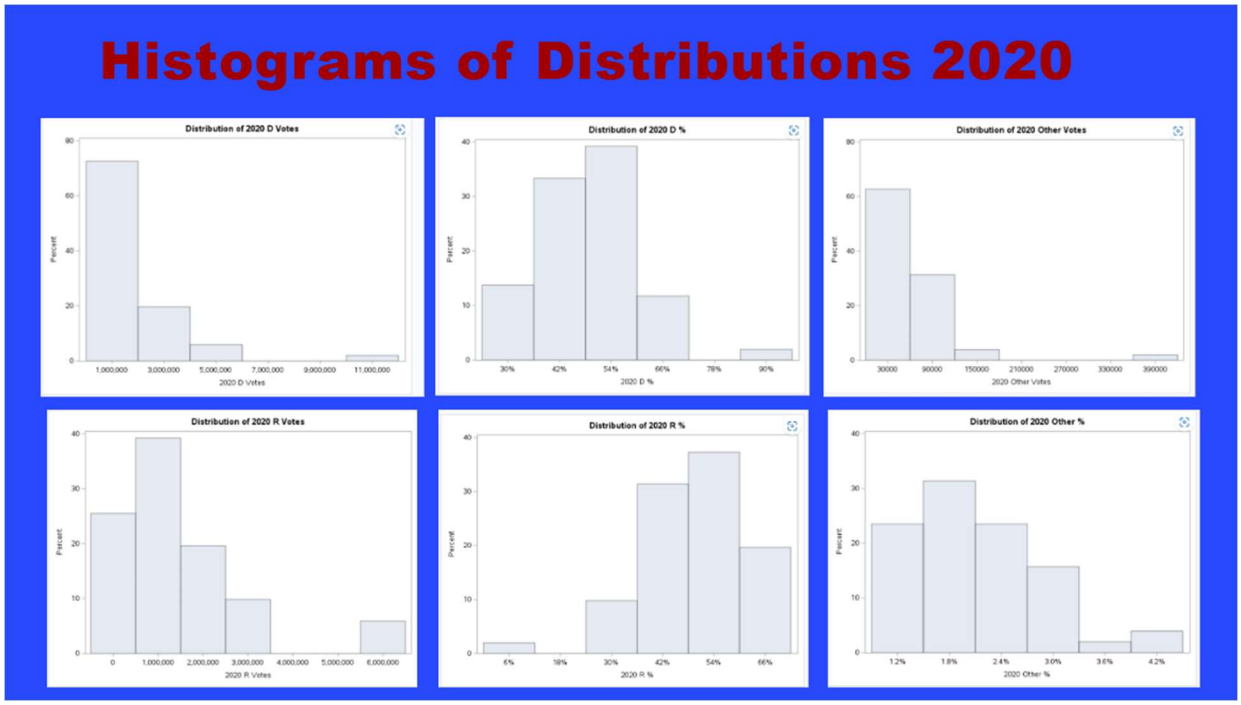
H_a = There is a significant change in the mean number of Democrat Votes between 2020 and 2024 ($\mu_1 \neq \mu_2$).

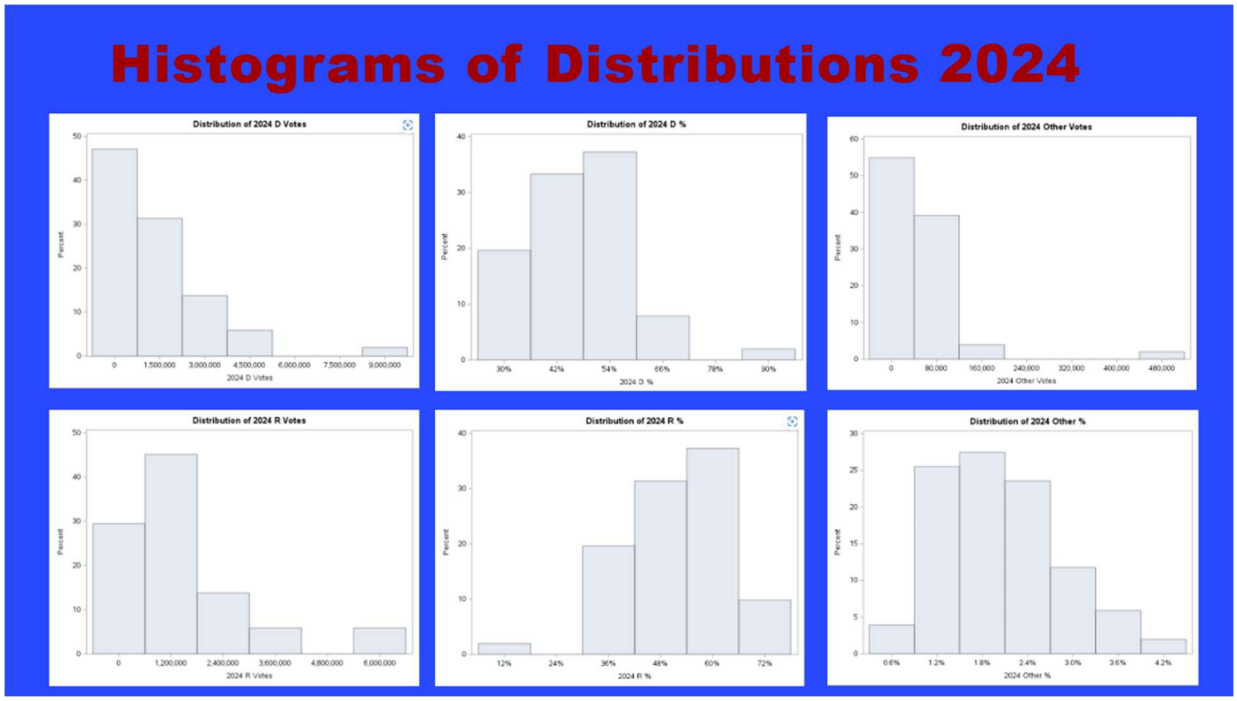
4.1 CHECK ASSUMPTIONS

Before we can perform our seven t-tests, we need to check to ensure that the data meets the assumptions required for statistical analysis. The three assumptions which should be met before continuing are Randomness, Independence, and Normality. The sample can be classified as Random since we are using all states in both elections. Each row entry is a state and therefore is independent of all the other entries. To check for normality, we used *proc univariate* looking for p-values of $\alpha < 0.05$, and considered the distributions by looking at the histograms of each variable.

Normality Test	
Variable	*P-Test*
2020 D Votes	<0.0001
2020 D %	0.0457
2020 R Votes	<0.0001
2020 R %	0.0288
2020 Other Votes	<0.0001
2020 Other %	0.0049
2020 Total Votes	<0.0001
2024 D Votes	<0.0001
2024 D %	0.0152
2024 R Votes	<0.0001
2024 R %	0.0119
2024 Other Votes	<0.0001
2024 Other %	0.1458
2024 Total Votes	<0.0001

All of our variables have a p-value of less than 0.05 except for the variable *2024 Other %*. From the histograms we see mostly normal distributions as well.





Since only one variable is not normal, *2024 Other %,* and we are mostly concerned with changes in Republican and Democrat votes, not Other, we can assume normality. All of our assumptions are met, thus we can proceed with our t-tests!

4.2 T-Tests

There are seven pairs of means we are comparing. R Votes, D Votes, Other Votes, Total Votes, R %, D %, and Other %. For each pair, we are testing the mean of all states to see if they are the same or not. Using *proc ttest* with $\alpha = 0.05$, we obtained the following results:

	Diff D	Diff D %	Diff R	Diff R %	Diff Other	Diff Other %	Diff Total
Mean	-129,746	-2.16%	56,394	2.27%	-2,145	-0.11%	-75,498
95% Min	-213,548	-2.54%	26,907	1.93%	-8,784	-0.26%	-151,996
95% Max	-45,944	-1.78%	85,882	2.60%	4,492	0.04%	1,001
P-Value	0.0031	0.0001	0.0003	0.0001	0.5191	0.1542	0.053
Null Hyp	Reject	Reject	Reject	Reject	Fail To Reject	Fail To Reject	Fail to Reject, close to cutoff
Alt Hyp	Evidence For	Evidence For	Evidence For	Evidence For			

For the pairs Democrat votes, Republican votes, Democrat %, and Republican % we observe p-values less than $\alpha = 0.05$, and therefore we reject the null hypothesis that the means of these pairs is the same. There is evidence for the alternative hypothesis for these pairs, that the means are not the same.

However, for the pairs Other votes, Other %, and Total votes, we observe p-values greater than $\alpha = 0.05$. For these pairs we will fail to reject the null hypothesis that the means of these pairs is the same. In the case of Total votes, it is noteworthy to mention that the p-value is just slightly above our cutoff of $\alpha = 0.05$. While we still fail to reject the null hypothesis here that the means are the same, it is close for this variable.

5. CONCLUSIONS

In summary, we observed from multiple visualizations and tables there appear to be significant differences in voting patterns between the 2020 and 2024 elections. After preprocessing our data, conducting an exploratory analysis, and checking the necessary assumptions, we performed seven paired t-tests to measure if the differences in voting pattern were statistically significant. For the Democrat votes, Republican votes, Democrat %, and Republican % we were able to reject the null hypothesis that the means were the same, and we did observe evidence for the alternative hypothesis that the means were significantly different. For Other votes, Other %, and Total votes, we failed to reject the null hypothesis that the means were the same. In terms of raw numbers:

- The means for Republican votes and Republican % by state went up an average of 56,395 and 2.27%.
- 95% of states saw in increase of between 26,907 to 85,882 in Republican votes.
- 95% of states saw in increase in % Republican between 1.93% to 2.60%.
- The means for Democrat votes and Democrat % by state went down an average of 129,746 and 2.16%.
- 95% of states saw a decrease in Democrat votes of between 45,944 to 213,548.
- 95% of states saw a decrease in % Democrat between 1.78% to 2.54%.
- The mean number of Total votes by state decreased by an average of 75,498.
- 95% of states saw a change in Total vote counts between –151,996 and 1,001.

Aside from these, some interesting observations were:

- The mean difference between Republican % (+2.27%) and the mean difference between Democrat % (-2.16%) are similar opposites.
- All states had an increase in % Republican vote.
- All states except one (Utah) had a decrease in % Democrat vote.
- New York and New Jersey both saw some of the largest increases in % Republican vote and decrease in % Democrat vote.
- California received significantly less Total votes, and Democrat Votes, but a relatively small increase in Republican votes.

References:

https://www.nbcnews.com/politics/2024-elections/president-results

https://www.nbcnews.com/politics/2020-elections/president-results/