# The Election of 2000 in Palm Beach County Was a Sham

**Case Study 1**

Radiah Khan Catherine Weeks Mercer Mercer

2025-10-06

## Introduction

We will introduce our research questions, background of the problem and the motive behind analyzing these questions.

### Background

During the US presidential election of 2000 Palm Beach County used a new confusing ballot that could have potentially influenced the election results for this county. Amid the tight race between Gore and Bush, Palm Beach County had an unexpectedly high number of votes for Buchanan, from the Reform party. This may be due to the new ballot used in Palm Beach, where if not read closely, someone attempting to cast a vote for Gore could have mistakenly filled in the circle for Buchanan. In this case study, we plan to look at the relationship between Bush and Buchanan votes within Palm Beach County and compare that relationship to those in other Florida counties. We also plan to make a prediction of how many miscast votes there were.

### Research Questions

In this case study, we aim to answer three questions:

- Is there a significant relationship between Bush and Buchanan votes within Palm Beach County?
- How does the relationship between Bush and Buchanan votes differ across other Florida counties?
- How many miscast votes were there for Buchanan in Palm Beach County?

### Motivation behind the research questions

We want to determine if the ballot in the Palm Beach county lead to people mistakenly casting votes for Buchanan when they meant to cast the vote for Gore.

## Data Description

In this section, we will introduce the variables we are going to use for this case study. Later, we are going to perform an exploratory data analysis to see the distribution of votes for both Buchanan and Bush.

### Variables Used

- `County` : Names of the county in Florida during the 2000s election.
- `Buchanan2000` : Number of votes for Buchanan in Palm Beach County during the 2000 election.
- `Bush2000` : Number of votes for Bush in Palm Beach County during the 2000 election.

### Exploratory Data Analysis

```
#creating a new variable that only contains Buchanan's election data
election_summary_buchanan <- election|>
  select(Buchanan2000)

#printing out the first 5 rows of the Buchanan table
head(election_summary_buchanan, 5)
```

```
  Buchanan2000
1          262
2           73
3          248
4           65
5          570
```

Above we can see the first few rows of Buchanan's vote totals from counties in Florida.

```
#creating a new variable that only contains Bush's election data
election_summary_bush <- election|>
  select(Bush2000)

#printing out the first 5 rows of the Bush table
head(election_summary_bush, 5)
```

```
  Bush2000
1    34062
2     5610
3    38637
4     5413
5   115185
```

Above is the first few rows of Bush's vote totals from counties in Florida.

```
#Saving summary table of Buchanan election data
election_count_summary_buchanan <-summary(election_summary_buchanan)

#printing the summary table created above with captions.
kable(election_count_summary_buchanan,
      caption = "Summary of Vote Counts (Buchanan) in the 2000s Election")
```

Table 1: Summary of Vote Counts (Buchanan) in the 2000s Election

| Buchanan2000 |
|---|
| Min. : 9.0 |
| 1st Qu.: 46.5 |
| Median : 114.0 |
| Mean : 258.5 |
| 3rd Qu.: 285.5 |
| Max. :3407.0 |

These are the summary statistics from Buchanan's vote counts. We can see overall Buchanan did not have a lot of voters across the different Florida counties.

```
#Saving summary table of Bush election data
election_count_summary_bush <-summary(election_summary_bush)

#printing the summary table created above with captions.
kable(election_count_summary_bush,
      caption = "Summary of Vote Counts (Bush) in the 2000s Election")
```

Table 2: Summary of Vote Counts (Bush) in the 2000s Election

| Bush2000 |
|---|
| Min. : 1316 |
| 1st Qu.: 4746 |
| Median : 20196 |
| Mean : 43356 |
| 3rd Qu.: 56542 |
| Max. :289456 |

These are the summary statistics from Bush's vote counts. We can see according to these summary statistics Bush had a fairly large voter base across all of the Florida Counties and just his mean vote count surpasses Buchanan's maximum.

```r
#Creating a new data set that is better for EDA
election_new <- election |>

  #pivoting longer to make separate box plots later
  pivot_longer(cols = c("Buchanan2000", "Bush2000"),
               names_to = "Candidate",
               values_to = "Count") |>
  mutate(Candidate = recode(Candidate,
                            "Bush2000" = "Bush",
                            "Buchanan2000" = "Buchanan"))

#printing out first 5 lines of new data frame
head(election_new, 5)
```

```
# A tibble: 5 x 3
  County  Candidate Count
  <chr>   <chr>     <dbl>
1 Alachua Buchanan    262
2 Alachua Bush      34062
3 Baker   Buchanan     73
4 Baker   Bush       5610
5 Bay     Buchanan    248
```

```r
#creating a box plot with the pivoted data
ggplot(data = election_new,
       aes(x = Count, fill = Candidate))+
  geom_boxplot() +

  #adding a title and x label
  labs(title ="Vote Counts for Bush and Buchanan in Florida",
       x = "Vote Counts (In log scale)") +

  #creating multiple plots based on the candidate
  facet_wrap(~Candidate)+

  #scaling the x axis to log10 to make plot more legible
  scale_x_log10()+

  #making the plot pretty :)
  theme_minimal()+
  theme(plot.margin = margin(10,10,10,10))
```

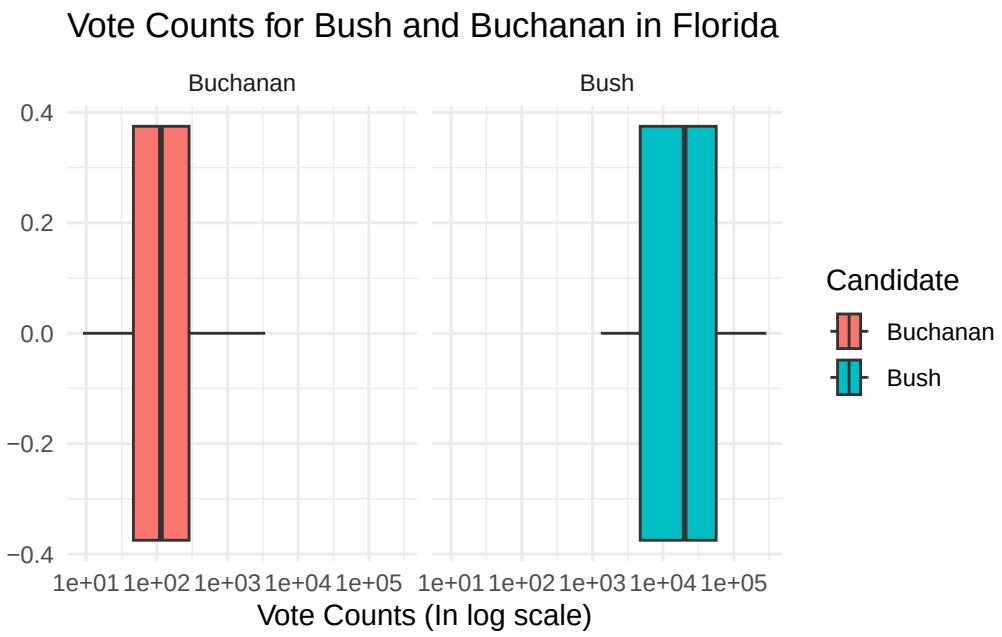



Figure 1

From the Figure 1, these box plots show the distribution of Bush and Buchanan votes in the different Florida counties. It is worth noting that the x axis has been changed to be log10, so the differences between the two plots are more significant. From this we can gather that Bush had an overall greater mean count of votes by a significant margin as well as a more consistent voter base across all counties. Meanwhile, Buchanan had less consistent voters, with some counties having much higher counts of votes than others, all while still maintaining lower vote counts than Bush overall.

## Analysis Model

We will use a simple linear regression model to determine the relationship between the votes for Bush and the votes for Buchanan. To begin we will assume conditions are met for linear regression model. Later we will asses their validity.

```
#creating a plot of Bush and Buchanan's votes in each county
ggplot(election, aes(x = Bush2000, y = Buchanan2000)) +
  geom_point() +

  #adding labels
  labs(title = "Buchanan vs Bush in the 2000 Elections",
       x = "Votes for Bush",
       y = "Votes for Buchanan")+
  theme_minimal()
```

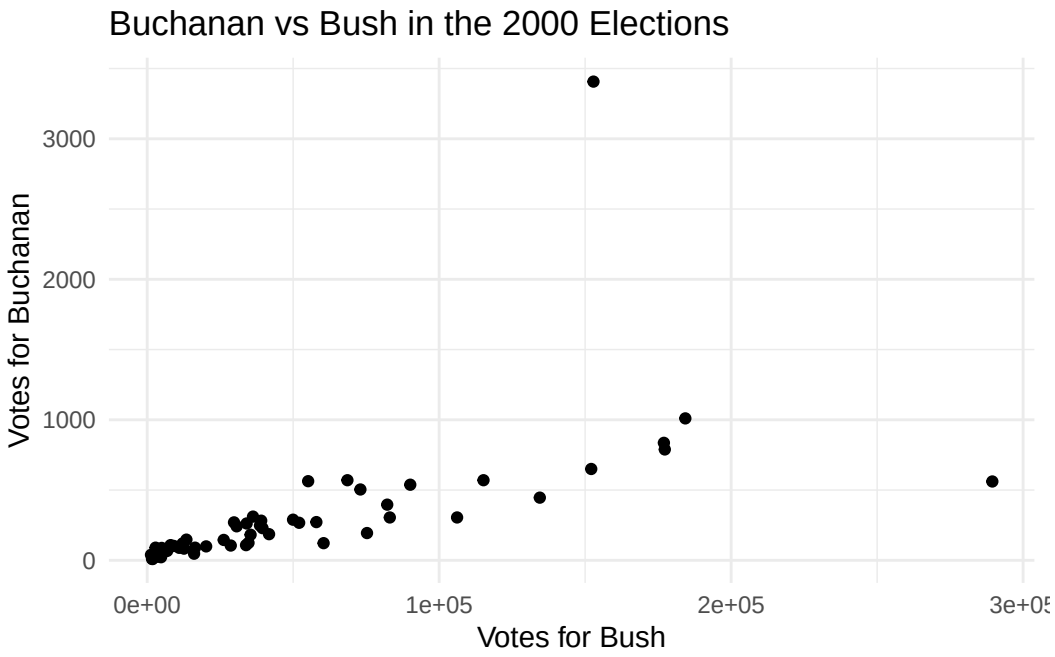



Figure 2

```
#creating a linear model of the Buchanan and Bush Votes
election_lm_pre <- lm((Buchanan2000) ~(Bush2000), data = election)
election_lm_pre
```

```
Call:
lm(formula = (Buchanan2000) ~ (Bush2000), data = election)

Coefficients:
(Intercept)     Bush2000
  45.289861     0.004917
```

```
#printing creating a summary table
election_summary <- tidy(election_lm_pre)

#printing the summary table
kable(election_summary, caption = "Regression summary Before Transformation")
```

Table 3: Regression summary Before Transformation

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | 45.2898613 | 54.4794230 | 0.8313205 | 0.4088361 |
| Bush2000 | 0.0049168 | 0.0007644 | 6.4319610 | 0.0000000 |

```r
#Plotting residuals of the pre-transformation model
plot(fitted(election_lm_pre), residuals(election_lm_pre))
abline(0,0)
```

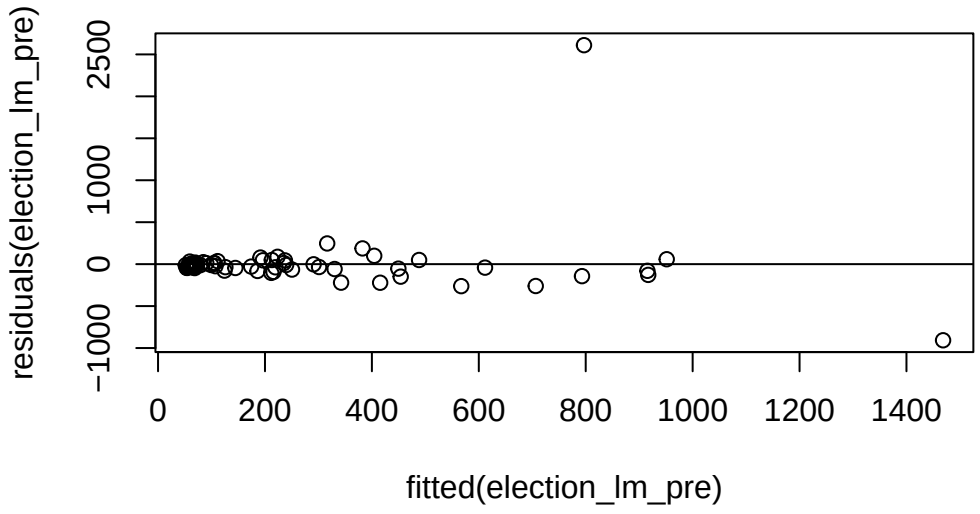



Figure 3

```r
#saving the residuals of the pre transformation model as a variable
residuals <- resid(election_lm_pre)

#creating a qq plot of the residuals
qqnorm(residuals)

#adding a line to help with legibility
qqline(residuals, col="red")
```

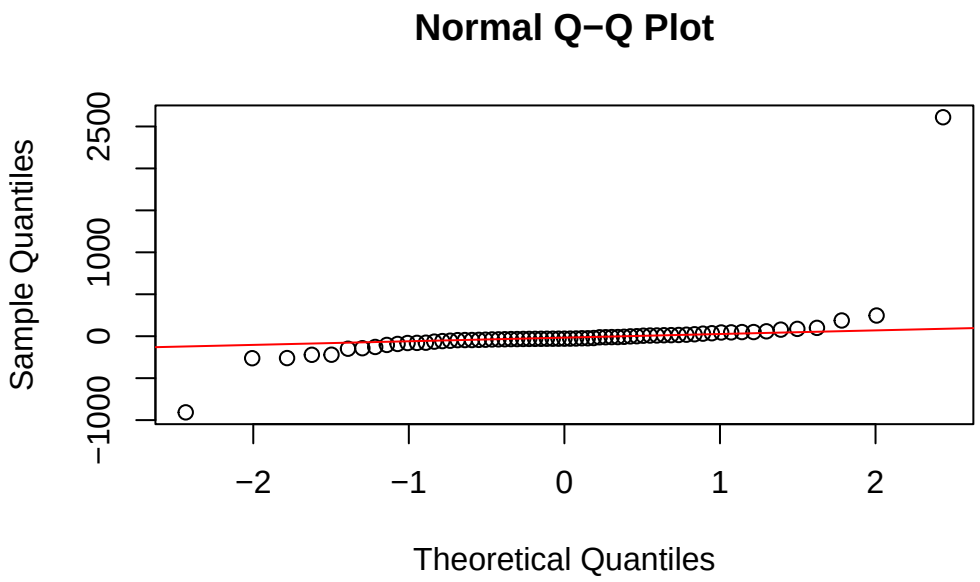



Figure 4

The first plot Figure 2 shows comparing Buchanan and Bush's vote totals by county tells us that there is at least a somewhat linear relationship between the two, but that we will likely need to preform a transformation to properly evaluate that relationship.

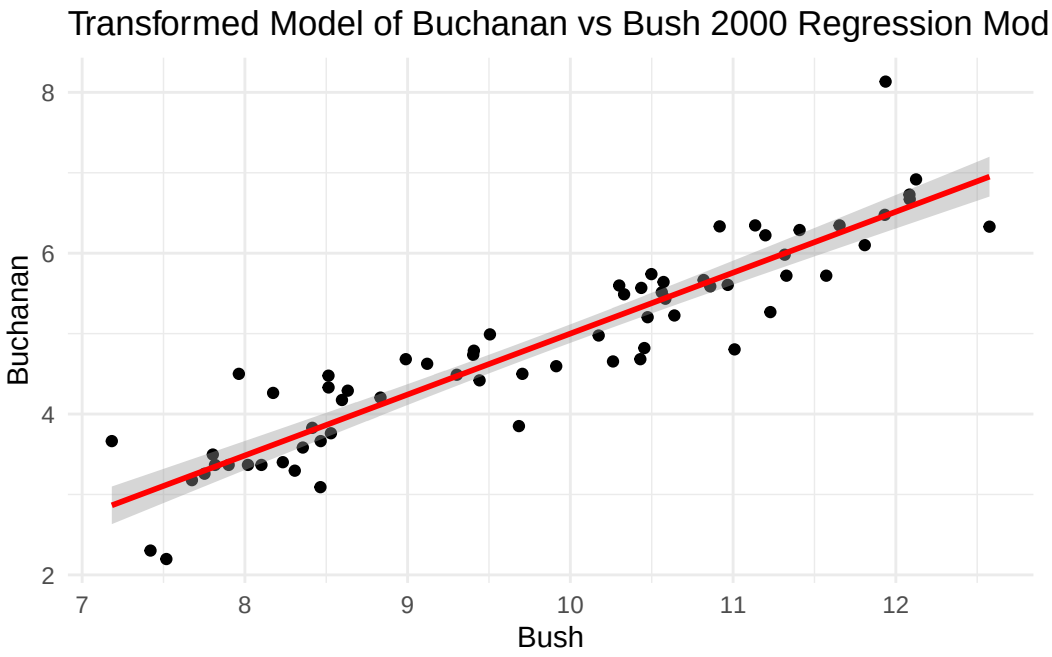
Together Figure 3 and Figure 4 confirm this suspicion. It shows us that these residuals are definitely not evenly distributed about the x axis and they are also clearly not evenly distanced vertically. This tells us that there are violations of linearity and of equal variance. And that the errors are not normally distributed, because of all of this we are going to re express the explanatory and response variables to hold the conditions.

Here is our post-transformation scatter plot. You can see that the data is already significantly better fit jsut visually, but we will do further analysis to confirm this.

```
#plotting the data with a log scale on the x and y axis.
ggplot(election, aes(x = log(Bush2000), y = log(Buchanan2000))) +
  geom_point() +

  #adding a line for legibility
  geom_smooth(method = "lm", col = "red") +

  #adding labels and a title
  labs(title = "Transformed Model of Buchanan vs Bush 2000 Regression Model",
       x = "Bush",
       y = "Buchanan")+
  theme_minimal()
```




**Linear Model**

```
#Creating a linear model of Buchanan votes and bush votes but they have both been log scaled.
election_lm_post <- lm(log(Buchanan2000) ~ log(Bush2000), data = election)
```

**Residual vs Fitted**

```
#creating a residualplot of the transformed model
plot(fitted(election_lm_post), residuals(election_lm_post))
abline(0,0)
```

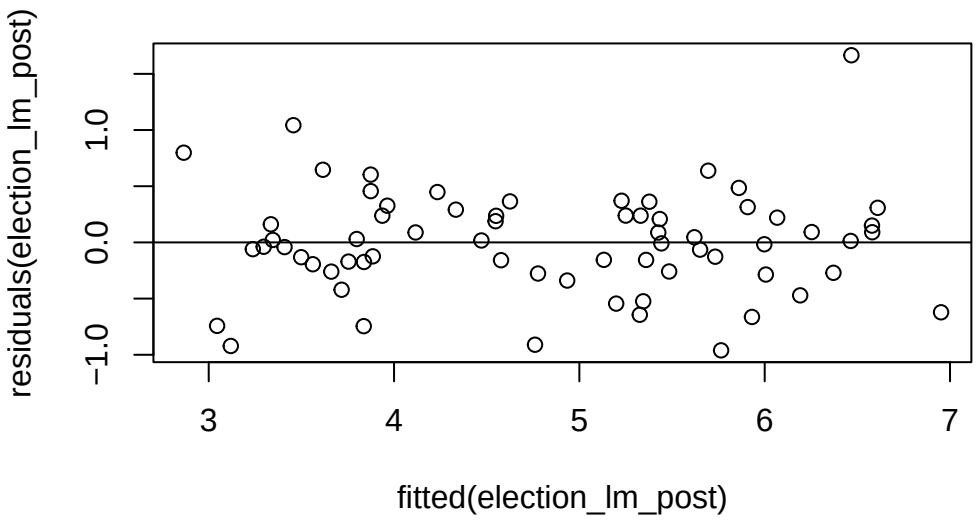



Figure 5

**Normality**

```
#saving the residuals as a variable
residuals_post <- resid(election_lm_post)

#creating a qq plot
qqnorm(residuals_post)

#adding a line for legibility
qqline(residuals_post, col="red")
```

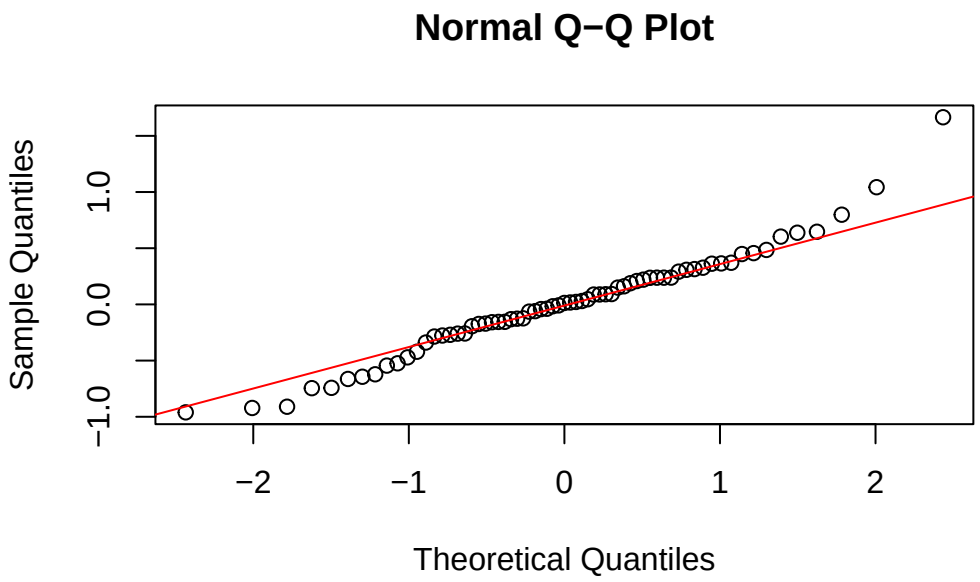



Figure 6

After it is transformed, $\log(y_i) = \beta\_0 + \beta_1 log(x_i)$

- **Null hypothesis** ($H_0$): $\beta = 0$
  There is no linear relationship between the votes for Bush and the votes for Buchanan in the 2000 election across Florida counties.
- **Alternative hypothesis** ($H_A$): $\beta \neq 0$
  There is a linear relationship between the votes for Bush and the votes for Buchanan in the 2000 election across Florida counties.

The significance level for this test is $p = 0.05$, and the confidence interval is 95.

```
#AN EQUIVELENT OF THIS ALREADY EXISTS
#Creating a lm comparing Buchanan's and Bush's 2000 election data
election_lm <- lm(Buchanan2000 ~ Bush2000, data = election)

#saving the summary table of the election lm
election_summary <- tidy(election_lm)

#printing the summary table
kable(election_summary, caption = "Regression summary")
```

Table 4: Regression summary

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | 45.2898613 | 54.4794230 | 0.8313205 | 0.4088361 |
| Bush2000 | 0.0049168 | 0.0007644 | 6.4319610 | 0.0000000 |

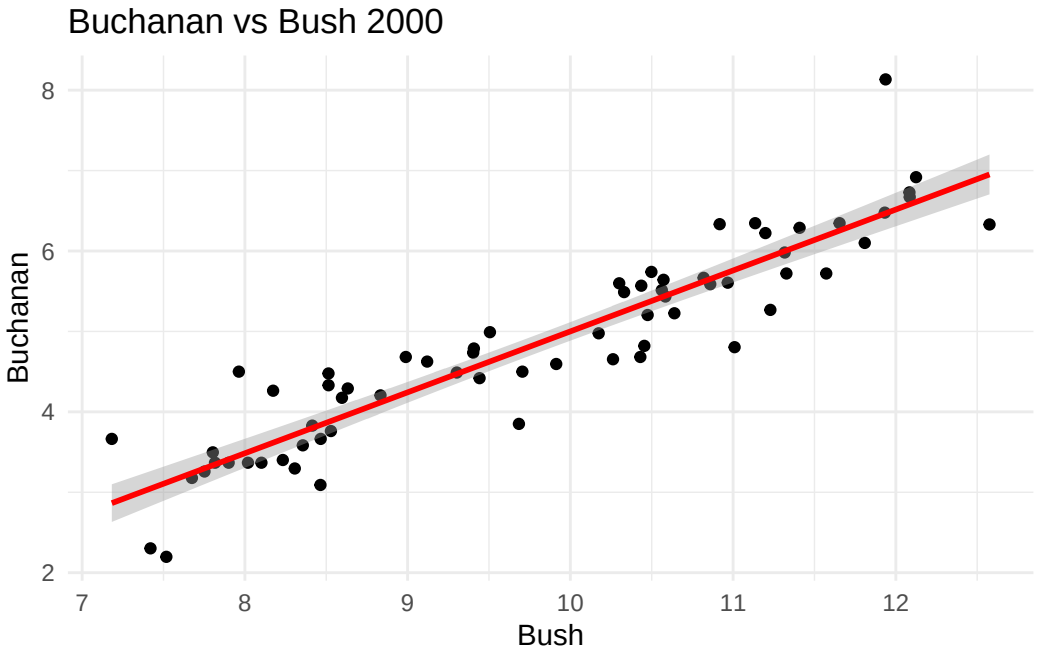
The p-value is near 0; suggesting significant relationship.

```
#Plotting the Bush and Buchanan data with a log transformation.
ggplot(election, aes(x = log(Bush2000), y = log(Buchanan2000))) +
  geom_point() +

  #adding a line to the plot
  geom_smooth(method = "lm", col = "red") +

  #adding labels
  labs(title = "Buchanan vs Bush 2000",
       x = "Bush",
       y = "Buchanan")+

  #making it pretty :)
  theme_minimal()
```




```
#saving the summary table of the election lm (post transformation)
election_summary_post <- tidy(election_lm_post )
```

```
#printing the summary table (post transformation)

kable(election_summary_post, caption = "Regression summary Post Transformation")
```

Table 5: Regression summary Post Transformation

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -2.5771236 | 0.3891909 | -6.621747 | 0 |
| log(Bush2000) | 0.7577224 | 0.0393594 | 19.251368 | 0 |

As seen in Table @ref(tab:post-summary), we set up the regression equation:

$$E[log(Buchanan2000)|log(Bush2000)] = \beta_0 + \beta_1 * Bush2000 + \epsilon$$
$$= -2.5771236 + 0.7577224 * Bush2000$$

p-value is 0; so the results are statistically significant which rejects the null hypothesis.

```
#saving our prediction number as the number of votes Bush got in Palm County in 2000
pred <- data.frame(Bush2000 = 152846)

#creating the prediction interval with a 95% conf level
prediction_interval <- election_lm_post |>
  augment(newdata = pred,
          interval = "prediction",
          conf.level = 0.95)
#|>
  #select(c(".fitted", ".lower", ".upper"))|>
  #exp()

prediction_interval
```

```
# A tibble: 1 x 4
  Bush2000 .fitted .lower .upper
     <dbl>   <dbl>  <dbl>  <dbl>
1   152846    6.47   5.51   7.42
```

```
#creating the prediction interval with a 95% conf level without specifying Palm Beach
prediction_interval1 <- election_lm_post |> augment(election, interval = "prediction", conf.lev

prediction_interval1
```

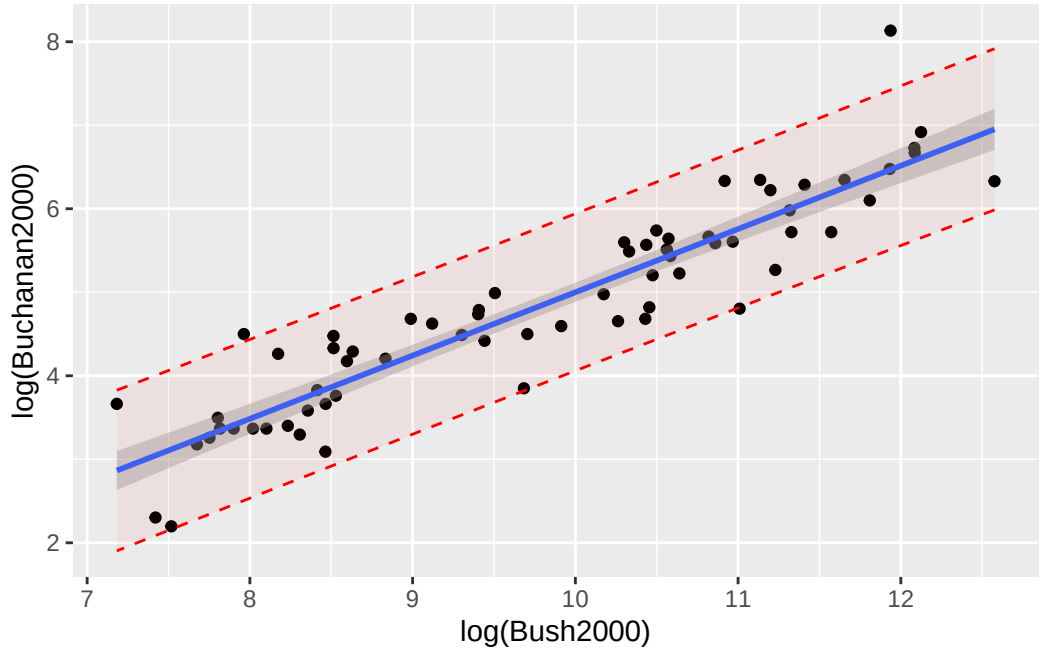
```
# A tibble: 67 x 11
   County    Buchanan2000 Bush2000 .fitted .lower .upper  .resid   .hat .sigma
   <chr>            <dbl>    <dbl>   <dbl>  <dbl>  <dbl>   <dbl>  <dbl>  <dbl>
```

```
 1 Alachua             262    34062    5.33   4.39   6.27  0.238  0.0180  0.470
 2 Baker                73     5610    3.96   3.02   4.91  0.327  0.0243  0.469
 3 Bay                 248    38637    5.43   4.48   6.37  0.0875 0.0193  0.471
 4 Bradford             65     5413    3.94   2.99   4.88  0.238  0.0249  0.470
 5 Brevard             570   115185    6.25   5.30   7.21  0.0920 0.0398  0.471
 6 Broward             789   177279    6.58   5.62   7.54  0.0905 0.0526  0.471
 7 Calhoun              90     2873    3.46   2.51   4.41  1.04   0.0384  0.452
 8 Charlotte           182    35419    5.36   4.42   6.30 -0.156  0.0183  0.471
 9 Citrus              270    29744    5.23   4.29   6.17  0.371  0.0168  0.469
10 Clay                186    41745    5.48   4.54   6.43 -0.259  0.0202  0.470
# i 57 more rows
# i 2 more variables: .cooksd <dbl>, .std.resid <dbl>
```

```r
#plotting the prediction interval
ggplot(data = election,
       mapping = aes(x = log(Bush2000), y = log(Buchanan2000))
       ) +
  geom_point() +

  #adding the lm
  geom_smooth(method = lm, formula = y~x) +

  #adding the prediction interval
  geom_ribbon(data = prediction_interval1,
              mapping = aes(ymax = .upper, ymin = .lower),
              color = "red", fill = "red",
              alpha = .05, linetype = 2
              )
```



We can say with 95% confidence that the number of Buchanan votes based on the number of Bush notes should be between 248 and 1,674 in Palm Beach County, FL. This is significantly different than the actual number of Buchanan votes in Palm Beach County, which is 3,407. Assuming some of the votes cast for Buchanan were meant to be cast for Gore, our prediction interval suggests there were likely 2,763 miscast votes.

## Summary

In this case study, we set out to determine if the ballot in the Palm Beach County lead to people mistakenly casting votes for Buchanan when they meant to cast the vote for Gore. Based on our models and prediction intervals, the Buchanan votes from Palm Beach County, FL are significantly different than what we would expect to see. This brings us to the conclusion that the confusing ballot for this county likely affected the votes, and led many people to mistakenly vote for Buchanan when they meant to vote for Gore.

### Limitations

Our major limitation is that we are limited to election data within Florida. We also don't know if Buchanan had an abnormally large voter base located within Palm Beach Florida. We also cannot prove that the reason Buchanan had an abnormality large vote count was because of voter fraud. If that was the case the independence condition would have been violated. However all of these limitations do not seem either likely or significant enough to affect the overall outcome to the extent we saw in the election of 2000.