# A Statistical Analysis of Palm Beach County Election Votes in 2000

**Case Study #1**

Sydney Weisberg and Molly Daniel

2025-03-05

## Introduction

During the U.S. presidential election of November 7, 2000, Reform Party candidate Pat Buchanan received an unusually high amount of votes in the Palm Beach county of Florida. This sparked debate about the layout of ballots in Palm Beach. Democratic voters in Palm Beach County complained that the confusing butterfly ballot format caused them to accidentally vote for Buchanan instead of the Democratic candidate Gore. Additionally, 19,000 ballots were discarded because voters had marked two circles, possibly indicating that they got confused by the ballot and attempted to rectify their vote. Could the influx of votes for Buchanan in Palm Beach have occurred due to chance, or is there significant evidence that some external factor, such as the butterfly ballot, changed the outcome of the election? This analysis will examine Buchanan votes as a function of Bush votes in Florida counties in order to draw a conclusion about whether or not the Palm Beach county vote count could have happened by chance given the trend/distribution of Florida county votes.

## Data Description

Our data is sourced from "The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)". The dataset consists of the numbers of votes for Buchanan and Bush in 67 counties in Florida during the 2000 presidential election. We exclude Palm Beach so that after we fit our model, we can see how the potentially abnormal Buchanan votes in that county compare to the rest of the data. The relationship between Bush and Buchanan votes is visualized in the scatter plot below

## Association between votes for Bush and votes for Buchanan.
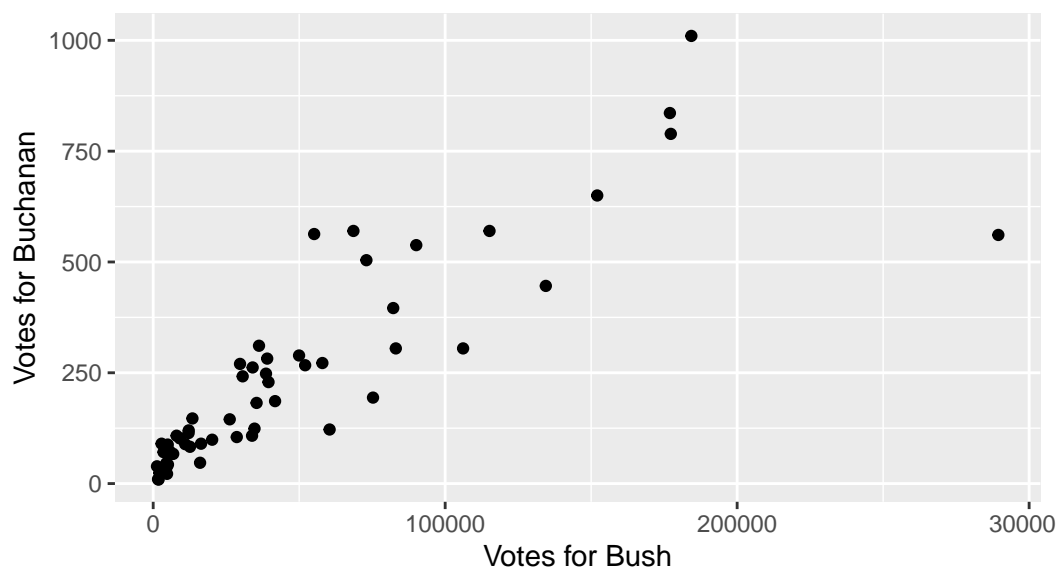


Relationship between Bush and Buchanan votes without Palm Beach county data

This plot shows that there is a positive correlation between votes for Bush and votes for Buchanan, with points clustering in the bottom left of the plot. We see that the scale on the x axis is much higher than the scale on the y axis, which makes sense given that Bush tends to receive more votes than Buchanan. The average amount of Buchanan votes versus Bush votes is shown in the table below.

| Average Buchanan Votes | Average Bush Votes |
|:---:|:---:|
| 210.76 | 41696.82 |

## Modeling Process

In order to create a model, we first observed the relationship in the scatter plot showing the votes for Buchanan and Bush in each of the Florida counties. From this, we noticed data clustering occurring in the bottom left corner of the scatter plot. We created diagnostic plots and checked to see if the model violated conditions for statistical inference. The diagnostic plots revealed violations of conditions of linearity, equal variance, and normality. In order to expand the scale and find a model that does not violate those conditions, we explored different transformations and their associated diagnostic plots. We concluded that a logarithmic transformation on the explanatory and response variable led to the regression conditions being best met. We applied a log transformation to both the Buchanan and Bush votes, revealing a multiplicative relationship between the two. Below is a plot of both variables transformed by log:
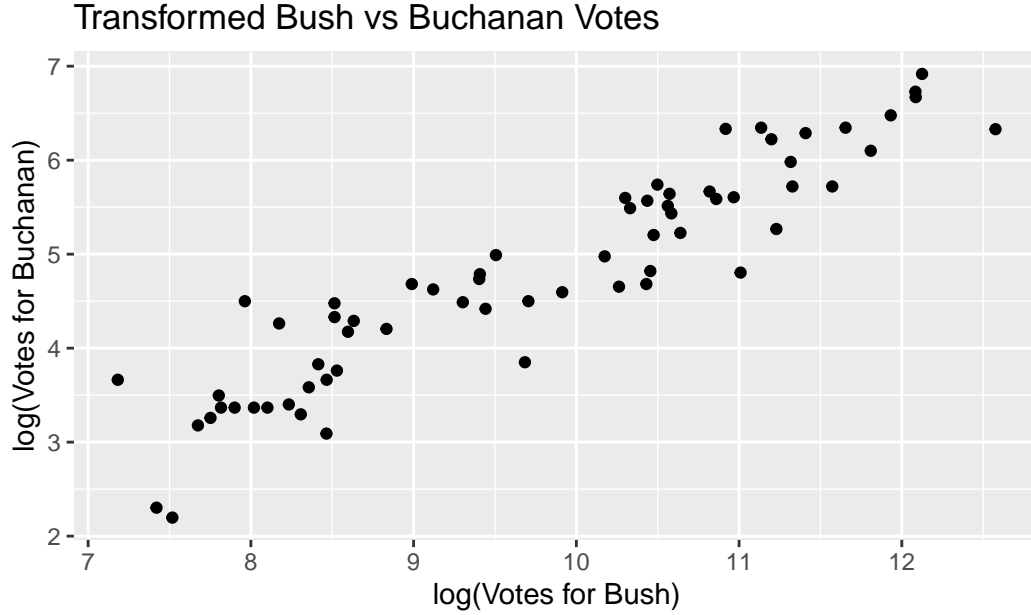
## Transformed Bush vs Buchanan Votes

Transformed relationship between Bush and Buchanan votes without Palm Beach county data

Let $buchanan_i$ denote the number of votes cast for Buchanan in county $i$ and $bush_i$ denote the number of votes cast for Bush in Florida county $i$ during the U.S. presidential election of November 7, 2000. Our final linear regression model for the mean is

$$E[log(buchanan_i)|log(bush_i)] = \beta_0 + \beta_1 \log(bush_i)$$

We fit our sample data to this model, and found estimates for the coefficients.

|              | Estimate | Std. Error | t value | Pr(>\|t\|) |
|--------------|----------|------------|---------|-----------|
| (Intercept)  | -2.34    | 0.35       | -6.61   | 0         |
| log(Bush2000)| 0.73     | 0.04       | 20.32   | 0         |

Our sample intercept, $\hat{\beta}_0$, is -2.34149 with a standard error of 0.35442. Our sample slope, $\hat{\beta}_1$, is 0.73096 with a standard error of 0.03597. Both coefficients have a p-value less than 0.05, making them statistically significant. Thus, our fitted model is

$$\log(\widehat{buchanan_i}) = -2.34149 + 0.73096 \log(bush_i).$$

Following this, we created a 95% prediction interval for the number of predicted Buchanan votes in Palm Beach County during the 2000 election. Our prediction interval means that we are 95% confident that the number of votes cast for Buchanan in the U.S. presidential election of November 7, 2000 when the number of votes cast for Bush is 152,846 (the reported number of votes for Bush in Palm Beach county) is between 250 and 1399 votes. By computing the difference in votes between the prediction interval and the actual number of Buchanan votes, we determine that there are likely between 2008 to 3157 miscast ballots.

## Conclusion

Since the 3407 votes for Buchanan in Palm Beach County in the 2000 election falls outside of our prediction interval of 250 to 1399 where Bush receives 152,846 votes, we can conclude that it is extremely unlikely for this unusually high number of votes for Buchanan to occur by chance, and that there likely is some external factor impacting the votes of the residents in Palm Beach. One limitation of this analysis is the lack of information regarding the sociopolitical climate of Palm Beach County in 2020. We do not know the demographic breakdown of the county. If there were a higher percentage of centrists who would vote for a member of the Reform Party in 2020 then that could explain the abnormal amount of Buchanan votes. Additionally, we don't know what campaigning was like in Palm Beach County. There could have been additional campaigns for Buchanan there that caused a vote increase particularly in that county. Based on the assumption that some of the votes cast for Buchanan in Palm Beach were actually intended to be votes for Gore, we make a generalization that there are more Democratic voters in Palm Beach County than represented in our data. Without the cultural knowledge of the environment where the data comes from, we can't be fully confident that our model is a good fit for the data.

## R Appendix

```r
# Importing relevant packages
library(tidyverse)
library(Sleuth2)
library(broom)
library(kableExtra)

# Importing election data from the textbook
election <- Sleuth2::ex0825

# Removing the extreme Palm Beach observation
election_wo_pb <- election |> filter(County != "Palm Beach")

# Displaying an untransformed scatter plot between Bush and Buchanan votes
election_wo_pb |> ggplot(aes(x = Bush2000, y = Buchanan2000)) + geom_point()
  + ggtitle("Association between votes for Bush and votes for Buchanan.")+
  labs(caption = "Relationship between Bush and Buchanan votes without
  Palm Beach county data") + xlab("Votes for Bush") + ylab("Votes for
  Buchanan")
```

Association between votes for Bush and votes for Buchanan.
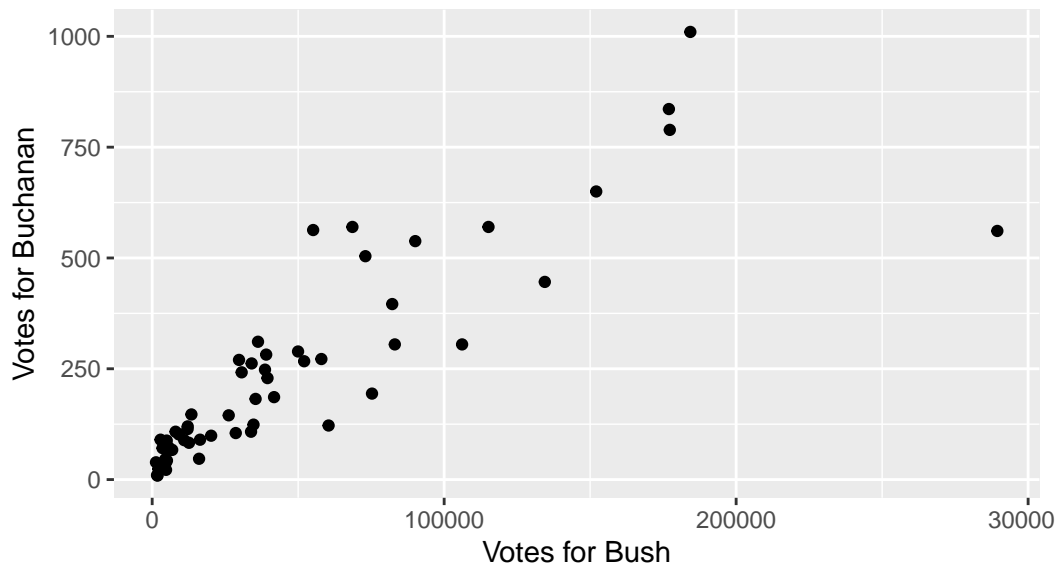
Relationship between Bush and Buchanan votes without Palm Beach county data

```
# Getting the average number of votes for each of the candidates excluding
↪  Palm Beach
buchanan <- mean(election_wo_pb$Buchanan2000)
bush <- mean(election_wo_pb$Bush2000)

mean_election_wo_pb_table <- data.frame(Buchanan2000 = buchanan, Bush2000 =
↪  bush)

mean_election_wo_pb_table |> kbl(col.names = c("Mean Buchanan Votes", "Mean
↪  Bush Votes"), align = "c", booktabs = T, linesep="", digits = c(2, 2, 2,
↪  4)) |> kable_classic(full_width = F, latex_options = c("HOLD_position"))
```
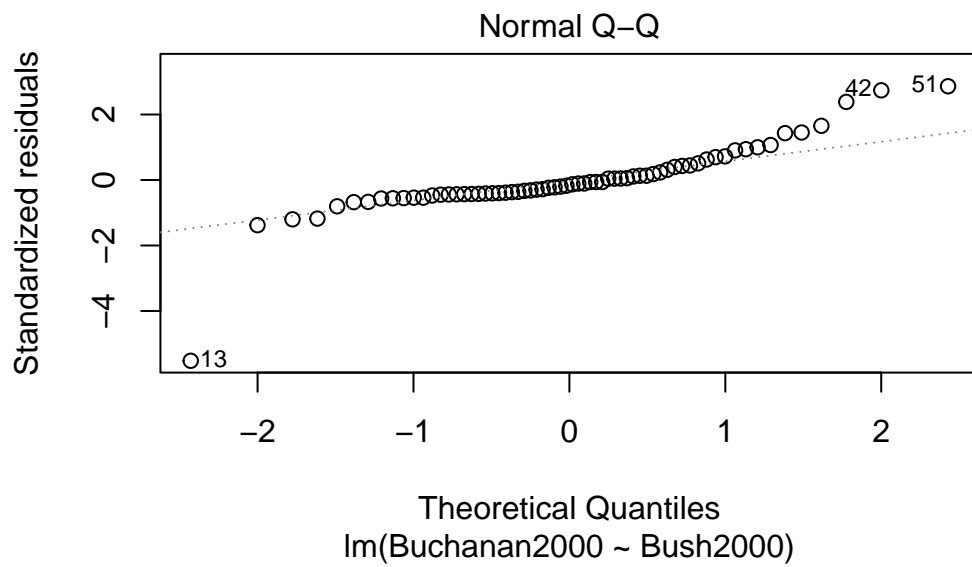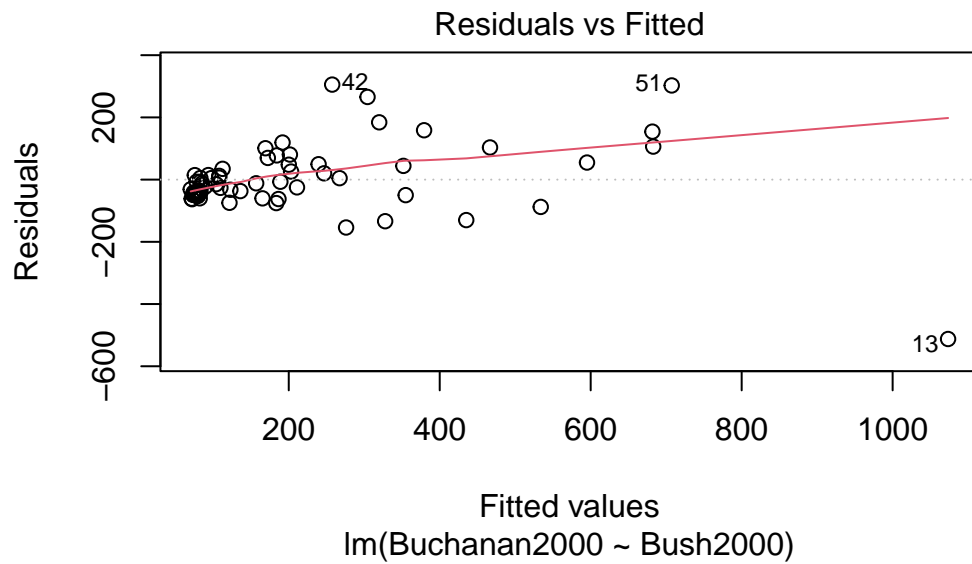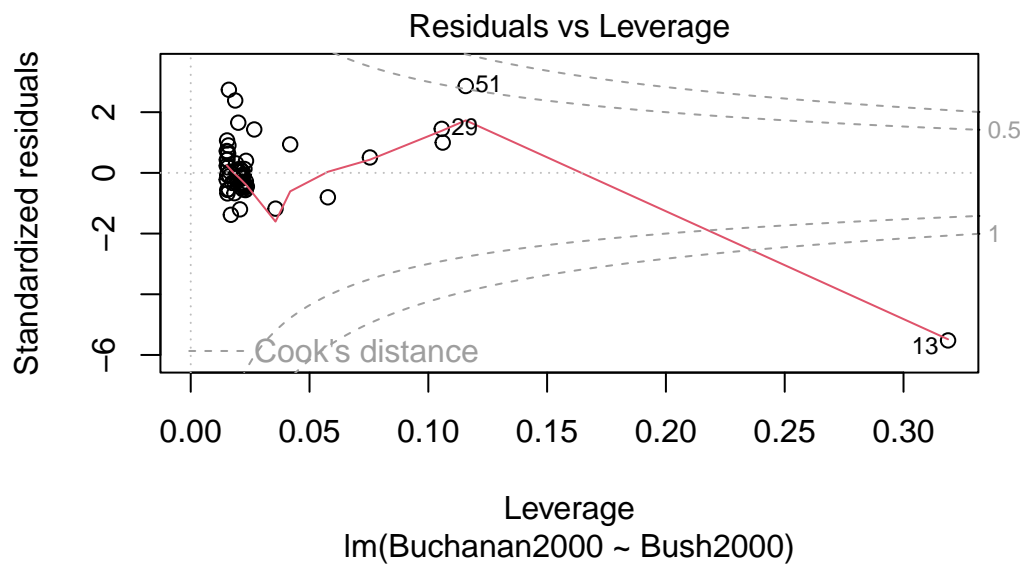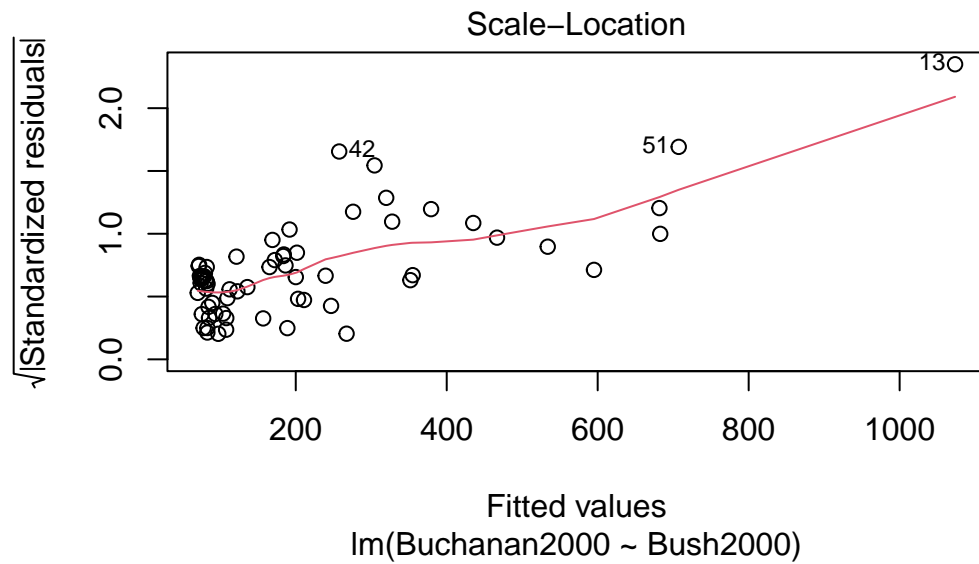
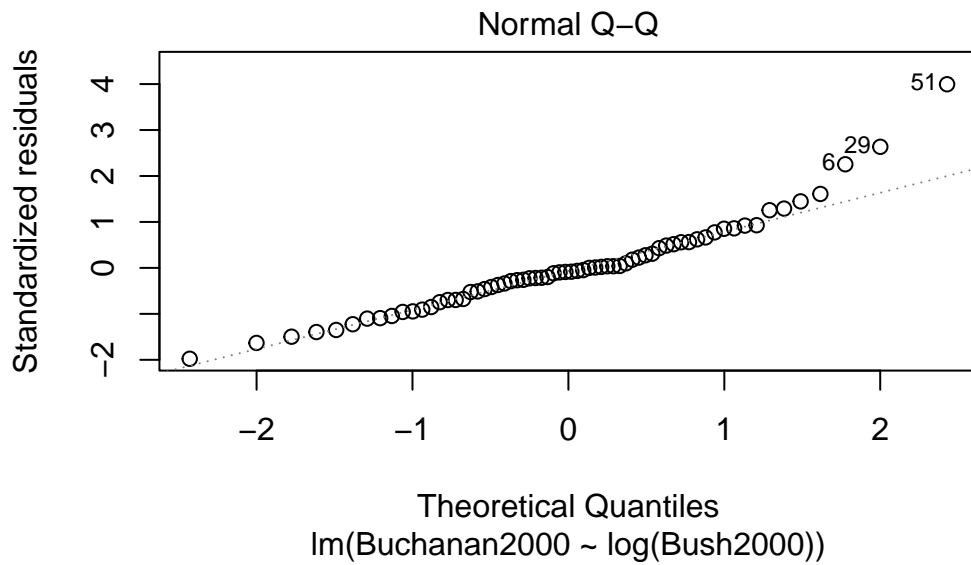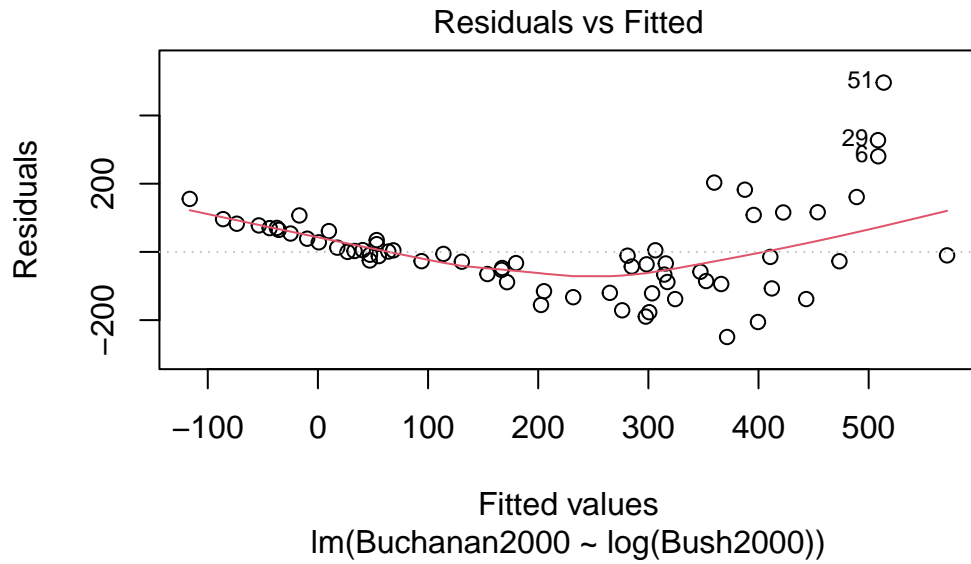| Mean Buchanan Votes | Mean Bush Votes |
|:---:|:---:|
| 210.76 | 41696.82 |

```
# Testing transformations to determine the best fit for the model

# Untransformed model
untransformed <- lm(Buchanan2000 ~ Bush2000, data = election_wo_pb)
plot(untransformed)
```
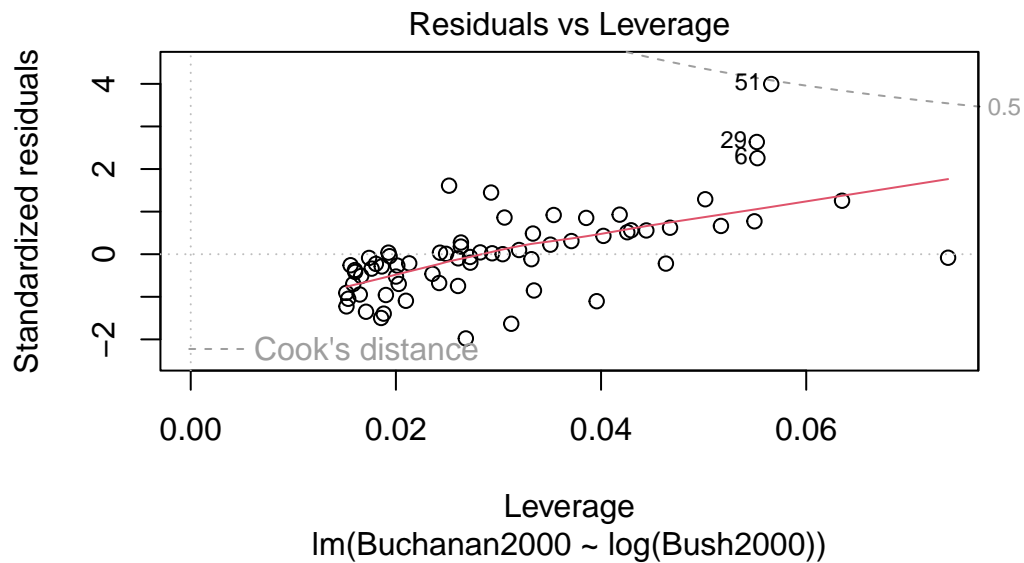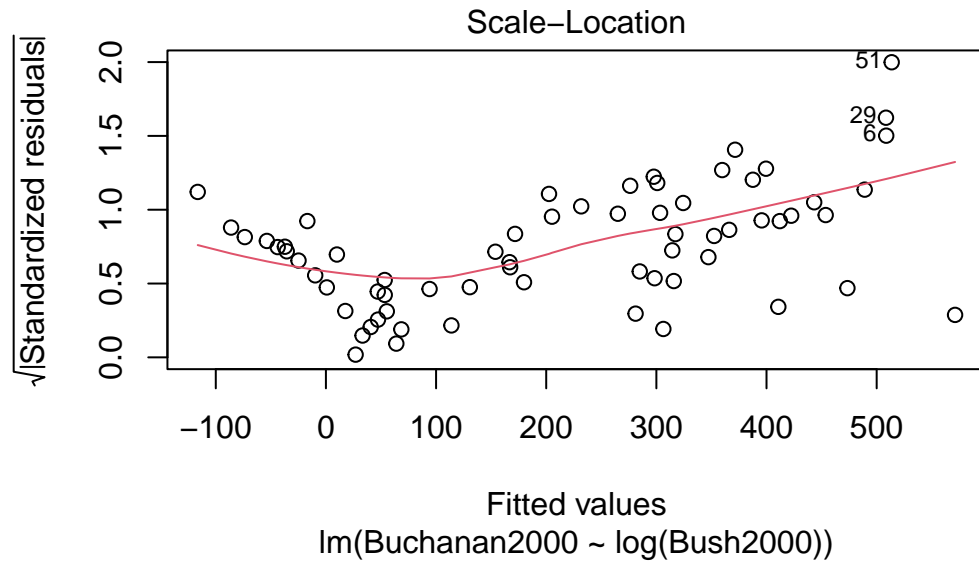
5

# Residuals vs Fitted



Residuals

42

51

13

200    400    600    800    1000

Fitted values
lm(Buchanan2000 ~ Bush2000)

# Normal Q–Q



Standardized residuals

42    51

13

−2    −1    0    1    2

Theoretical Quantiles
lm(Buchanan2000 ~ Bush2000)

## Scale–Location

$\sqrt{|\text{Standardized residuals}|}$ vs Fitted values

lm(Buchanan2000 ~ Bush2000)

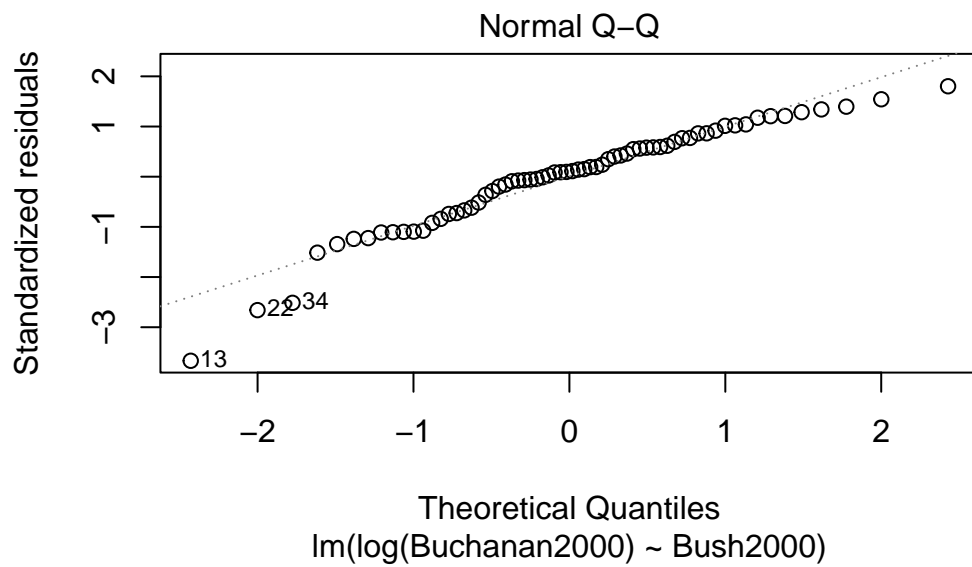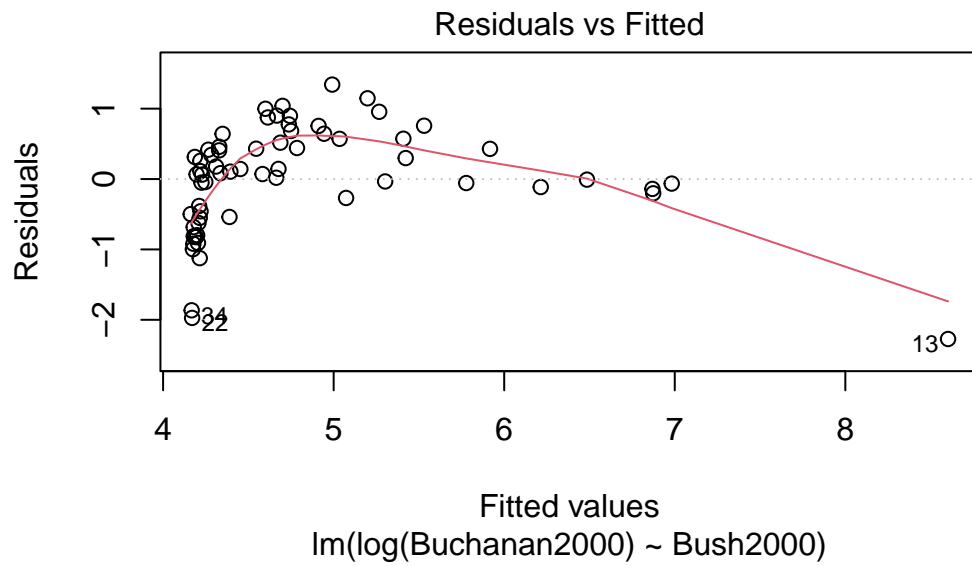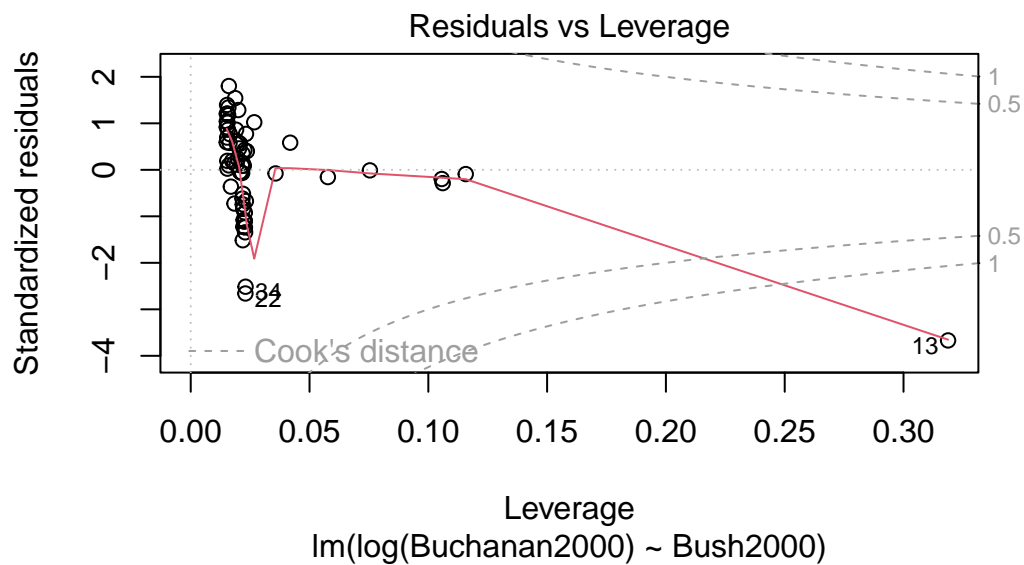## Residuals vs Leverage

lm(Buchanan2000 ~ Bush2000)

```r
# Model with a logarithmic explanatory variable transformation
xtransformed <- lm(Buchanan2000 ~ log(Bush2000), data = election_wo_pb)
plot(xtransformed)
```

## Residuals vs Fitted



Fitted values
lm(Buchanan2000 ~ log(Bush2000))

## Normal Q–Q



Theoretical Quantiles
lm(Buchanan2000 ~ log(Bush2000))

Scale-Location

√|Standardized residuals|

Fitted values
lm(Buchanan2000 ~ log(Bush2000))



Residuals vs Leverage

Standardized residuals

Cook's distance

Leverage
lm(Buchanan2000 ~ log(Bush2000))

```r
# Model with a logarithmic response variable transformation
ytransformed <- lm(log(Buchanan2000) ~ Bush2000, data = election_wo_pb)
plot(ytransformed)
```
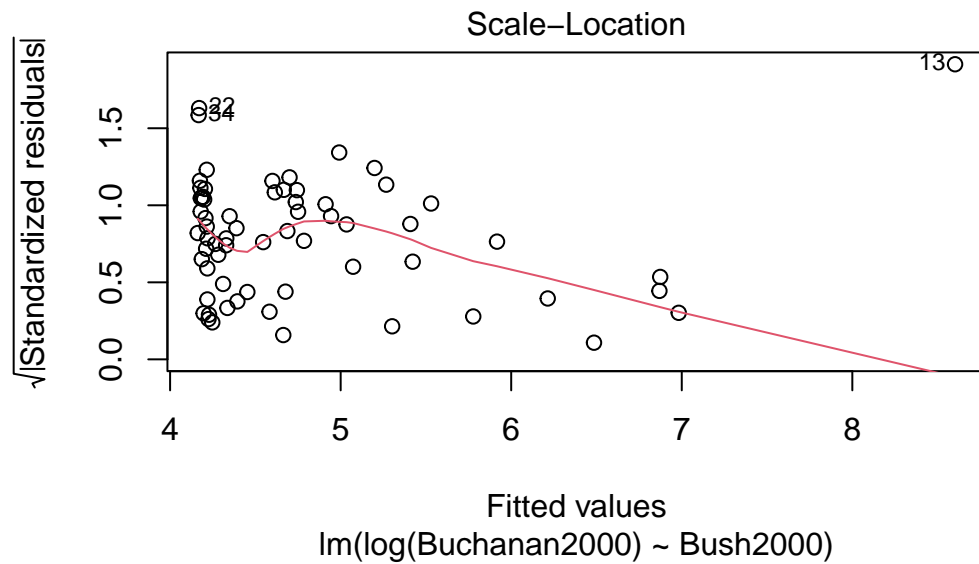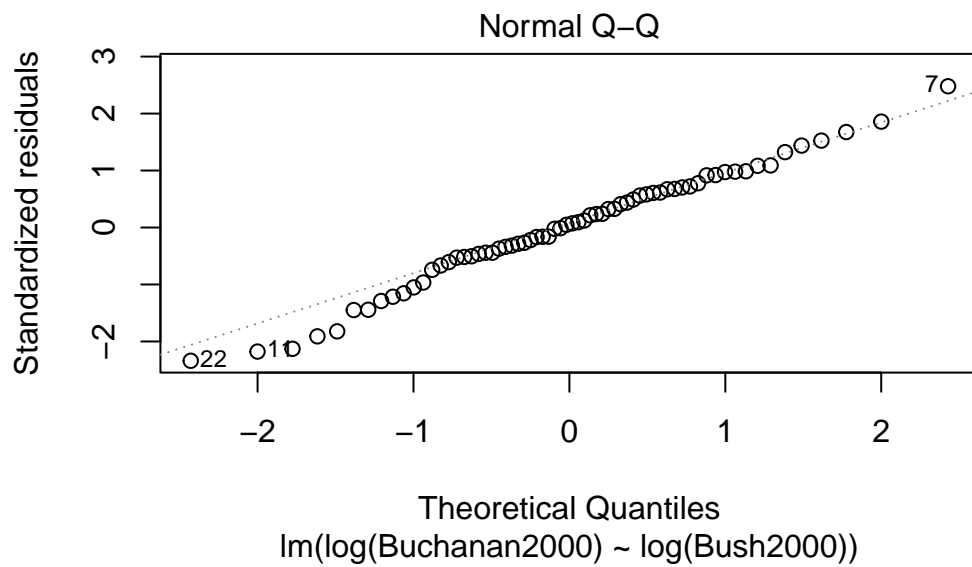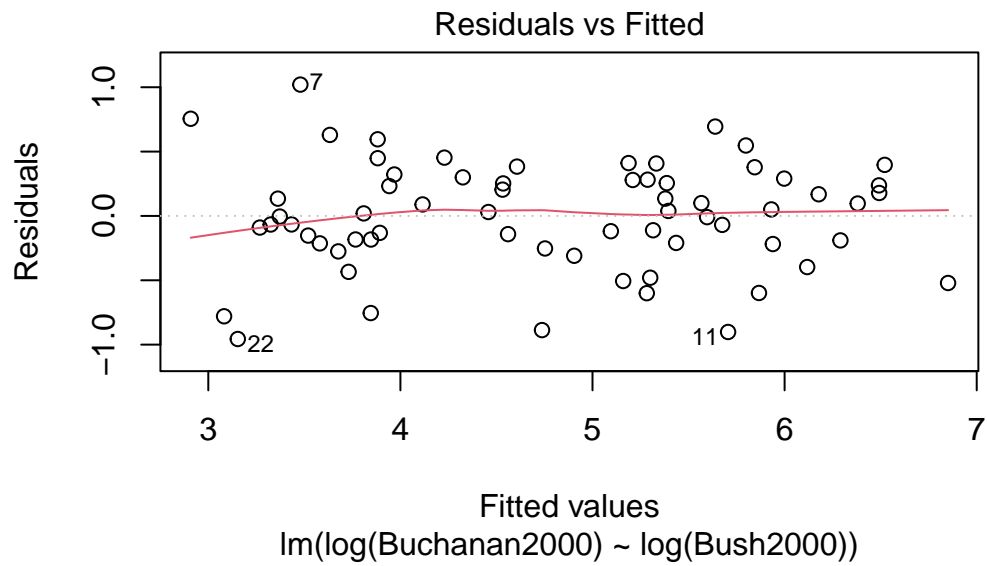
## Residuals vs Fitted



Residuals

Fitted values
lm(log(Buchanan2000) ~ Bush2000)

## Normal Q–Q



Standardized residuals

Theoretical Quantiles
lm(log(Buchanan2000) ~ Bush2000)

## Scale−Location



√|Standardized residuals|

Fitted values
lm(log(Buchanan2000) ~ Bush2000)

## Residuals vs Leverage



Standardized residuals

Leverage
lm(log(Buchanan2000) ~ Bush2000)

```r
# Model with a logarithmic explanatory and response variable transformation
both_transformed <- lm(log(Buchanan2000) ~ log(Bush2000), data =
 ↪  election_wo_pb)
plot(both_transformed)
```
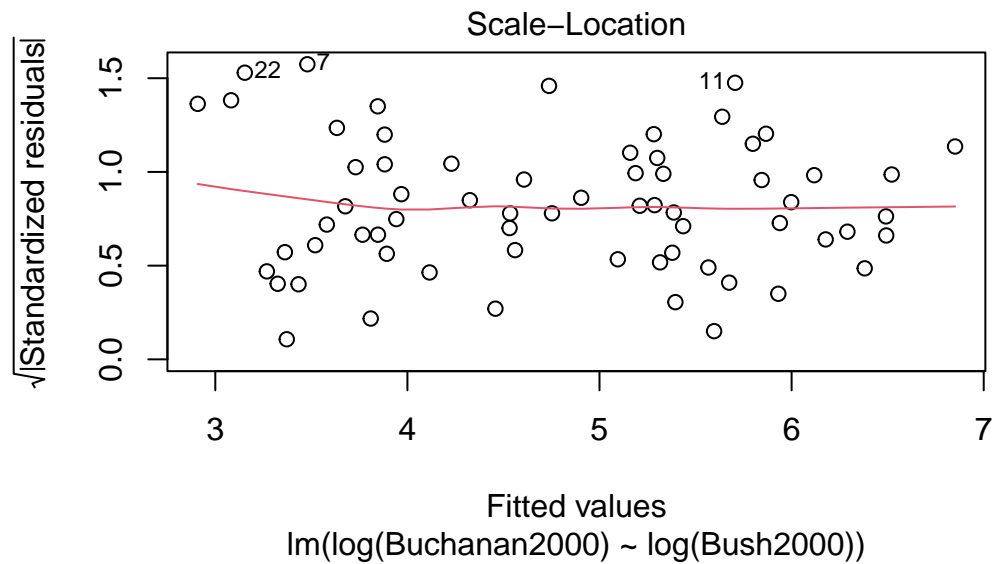
## Residuals vs Fitted



Fitted values
lm(log(Buchanan2000) ~ log(Bush2000))

## Normal Q–Q



Theoretical Quantiles
lm(log(Buchanan2000) ~ log(Bush2000))

## Scale–Location



√|Standardized residuals|

22 7

11

Fitted values
lm(log(Buchanan2000) ~ log(Bush2000))

## Residuals vs Leverage



7

39

Cook's distance

22

Leverage
lm(log(Buchanan2000) ~ log(Bush2000))
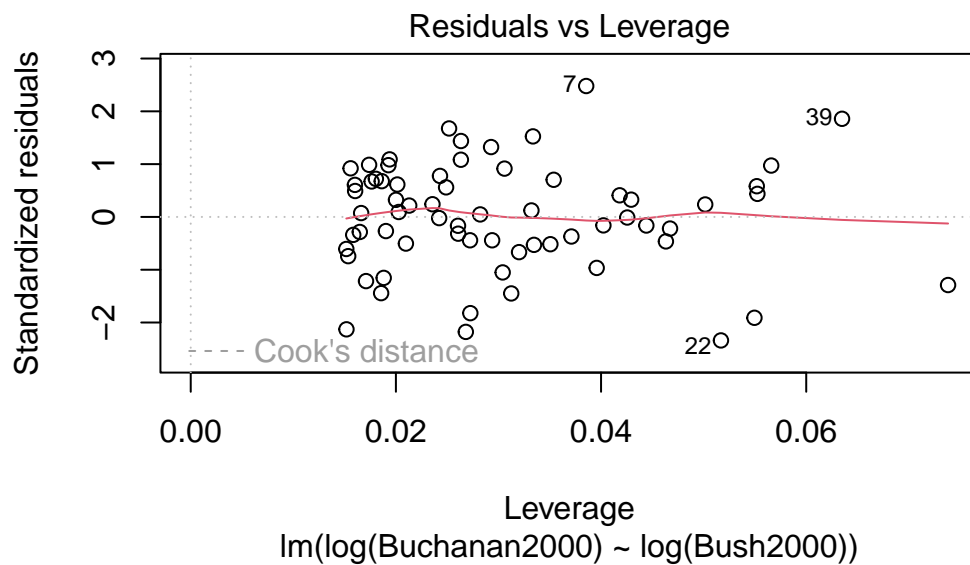
```
# Displaying a doubly transformed scatter plot between Bush and Buchanan
↪  votes
```

```
election_wo_pb |> ggplot(aes(x = log(Bush2000), y = log(Buchanan2000))) +
  ↳  geom_point() + ggtitle("Transformed Bush vs Buchanan Votes") +
  ↳  labs(caption = "Transformed relationship between Bush and Buchanan votes
  ↳  without Palm Beach county data") + xlab("log(Votes for Bush)") +
  ↳  ylab("log(Votes for Buchanan)")
```
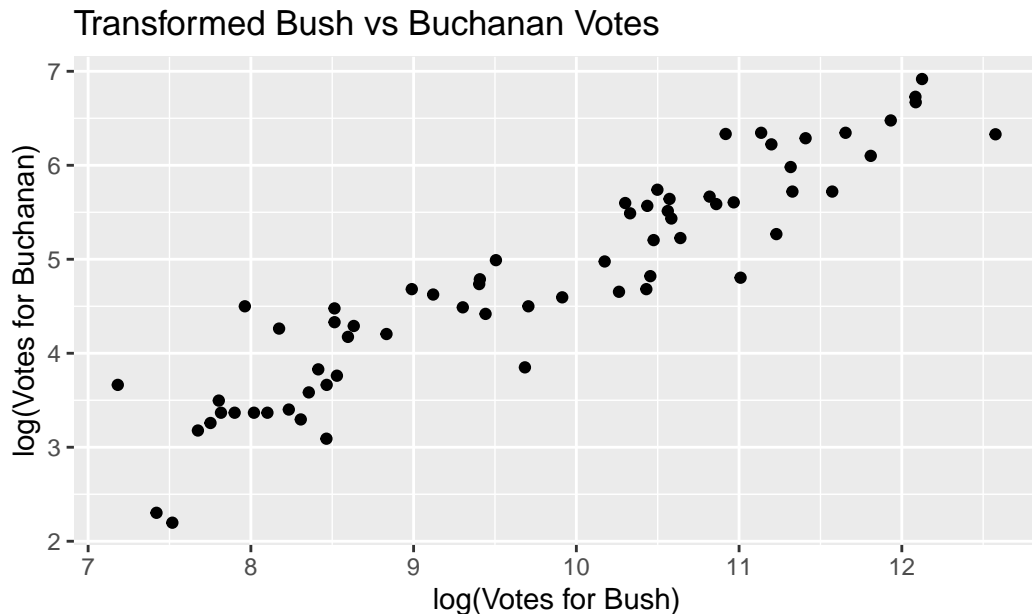
### Transformed Bush vs Buchanan Votes



Transformed relationship between Bush and Buchanan votes without Palm Beach county data

```
# Getting the coefficients for the doubly transformed model
both_transformed_table <- summary(both_transformed)$coefficients

# Creating a visible, clean table to display the coefficients from the
  ↳  doubly transformed model
both_transformed_table |> kbl(col.names = c("Estimate", "Std. Error", "t
  ↳  value", "Pr(>|t|)"), align = "c", booktabs = T, linesep="", digits =
  ↳  c(2, 2, 2, 4)) |> kable_classic(full_width = F, latex_options =
  ↳  c("HOLD_position"))
```

|              | Estimate | Std. Error | t value | Pr(>\|t\|) |
|--------------|----------|------------|---------|------------|
| (Intercept)  | -2.34    | 0.35       | -6.61   | 0          |
| log(Bush2000) | 0.73    | 0.04       | 20.32   | 0          |

```
summary(both_transformed)
```

```
Call:
```

```
lm(formula = log(Buchanan2000) ~ log(Bush2000), data = election_wo_pb)


Residuals:
     Min       1Q   Median       3Q      Max
-0.95631 -0.21236  0.02503  0.28102  1.02056


Coefficients:
              Estimate Std. Error t value            Pr(>|t|)
(Intercept)   -2.34149    0.35442  -6.607        0.00000000907 ***
log(Bush2000)  0.73096    0.03597  20.323 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.4198 on 64 degrees of freedom
Multiple R-squared:  0.8658,    Adjusted R-squared:  0.8637
F-statistic:   413 on 1 and 64 DF,  p-value: < 0.00000000000000022
```

```r
# Creating a prediction interval
predicted_palm_beach = data.frame(Bush2000 = 152846)
both_transformed |> augment(newdata = predicted_palm_beach, interval =
  ↪  "prediction", conf.level = 0.95)
```

```
# A tibble: 1 x 4
  Bush2000 .fitted .lower .upper
     <dbl>   <dbl>  <dbl>  <dbl>
1   152846    6.38   5.52   7.24
```