A Statistical Analysis of Palm Beach County Election Votes in 2000

Sydney Weisberg and Molly Daniel

2005-03-01

Introduction

During the US presidential election of November 7, 2000, Democratic voters in Palm Beach County complained that a confusing ballot format caused them to accidentally vote for the Reform Party candidate Pat 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 changed the outcome of the election? This analysis will examine the relationship between Bush votes (explanatory variable) and Buchanan votes (response variable) 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 county votes.

Data Description

The election data consists of the numbers of votes for Buchanan and Bush in all 67 counties in Florida. We construct plots and fit our model using every county except Palm Beach. We do this so after we fit our model, we can see how Palm Beach data compares to the rest of the data. Here is a scatter plot showcasing the relationship between Bush votes and Buchanan votes.

Association between votes for Bush and votes for Buchanan.
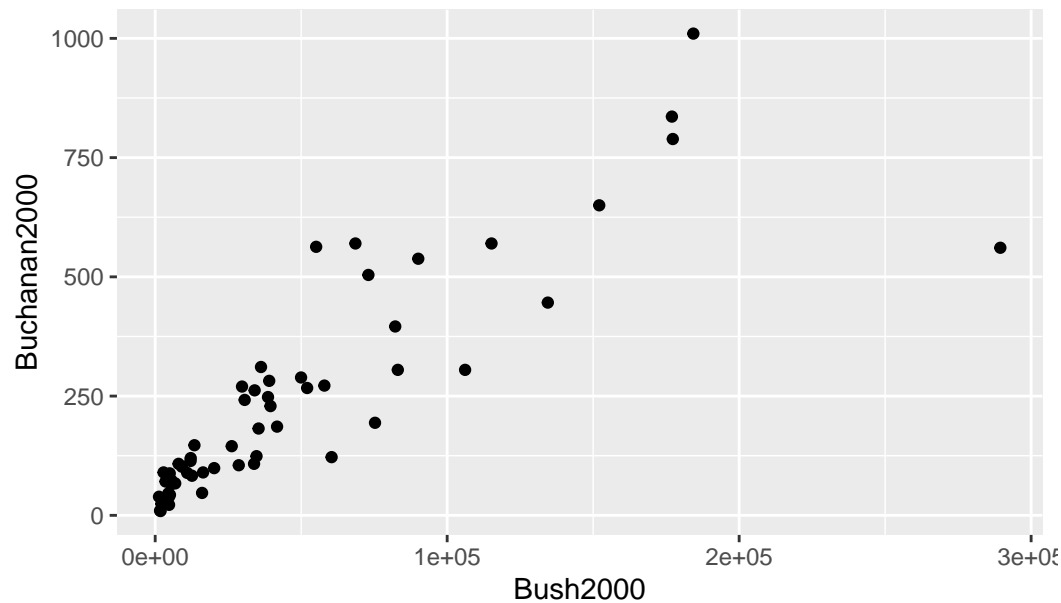
Modeling Process

In order to create a model, we first observed a 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. When creating diagnostic plots of the data, it was clear that regression conditions were not met. In order to expand the scale, we applied a log transformation to both the Buchanan and Bush votes, revealing a multiplicative relationship between the two. After transformation, the diagnostic plots reveal that regression conditions are met. Below is a plot of both variables transformed by log:

```
election_wo_pb |> ggplot(aes(x = log(Bush2000), y = log(Buchanan2000))) +
 ↪   geom_point() + ggtitle("Transformed Association between votes for Bush
 ↪   and votes for Buchanan.")
```

Transformed Association between votes for Bush and votes for E
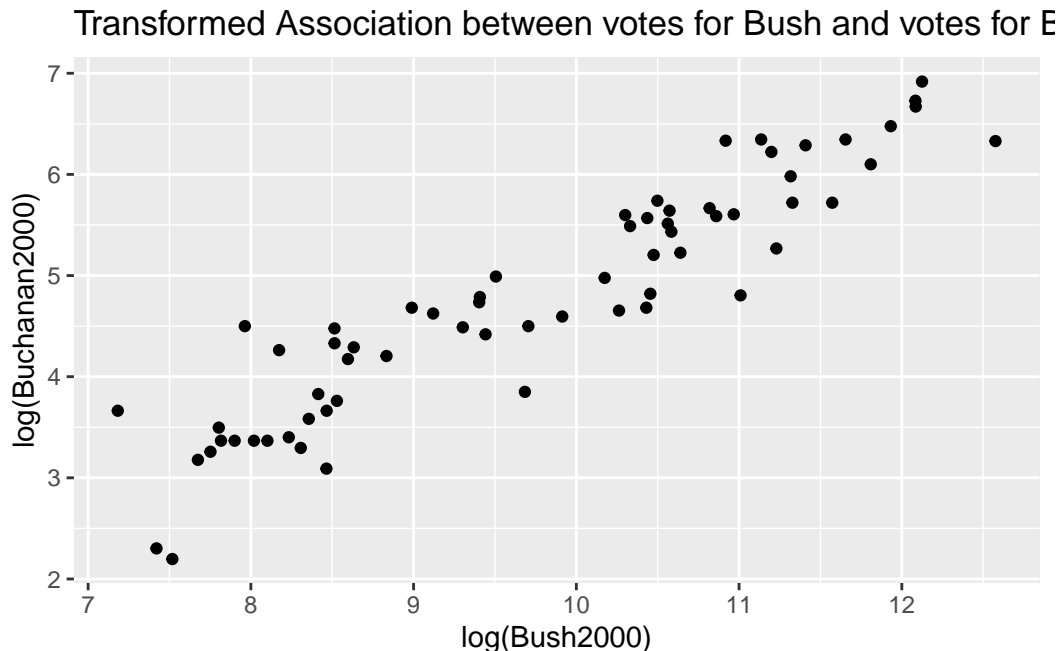
Let $buchanan_i$ denote the number of votes cast for Buchanan in county i and $bush_i$ denote the number of votes cast for Bush in Florida county i during the U.S. presidential election of November 7, 2000. Our final linear regression model for the mean is $E[log(buchanan_i)|log(bush_i)] = \beta_0 + \beta_1 \log(bush_i)$. 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. So our fitted model is $log(\widehat{buchanan_i}) = -2.34149 + 0.73096 \log(bush_i)$.

Point Estimate $= e\hat{}6.384143 = 592.376848042$

Prediction Interval $= (e\hat{}5.524656, e\hat{}7.24363) = (250.8, 1399.164)$

This 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 152846 (the reported number of votes for Bush in Palm Beach county) is between 250.8 and 1399.164 votes.

Conclusion

Since the 3407 votes for Buchanan in Palm Beach County in the 2000 election falls outside of our prediction interval of 250.8 to 1399.164 where Bush receives 15,2846 votes, we can conclude that it

is likely that there 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.")
```
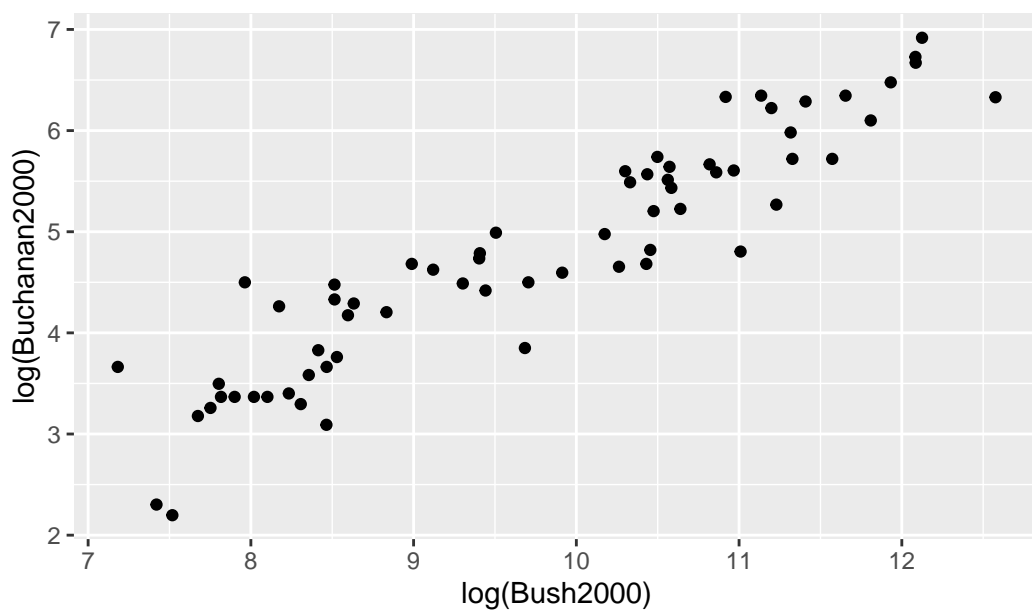
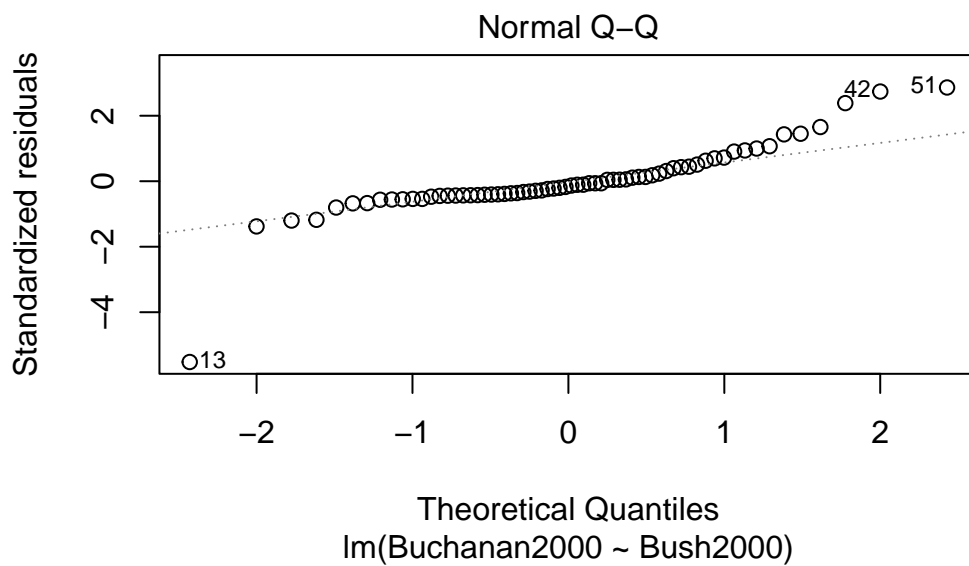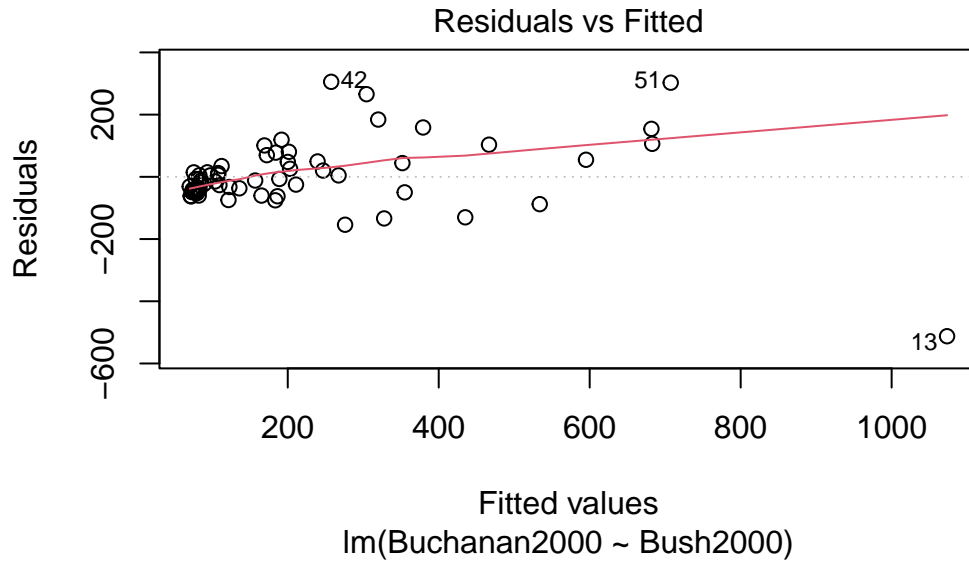Association between votes for Bush and votes for Buchanan.
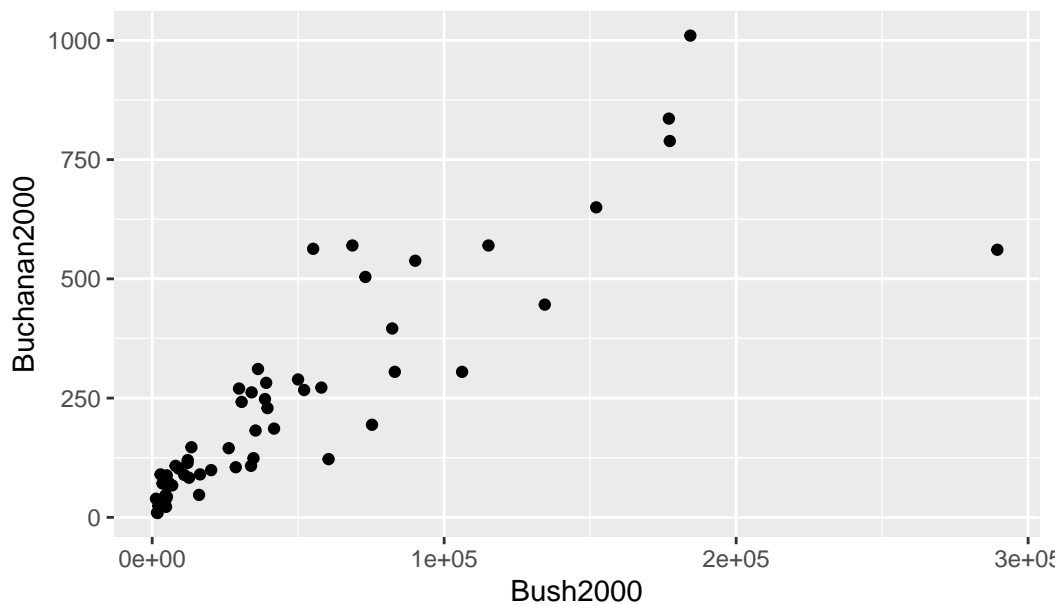
```
# Displaying a doubly transformed scatter plot between Bush and Buchanan
↪  votes
election_wo_pb |> ggplot(aes(x = log(Bush2000), y = log(Buchanan2000))) +
↪  geom_point() + ggtitle("Transformed Association between votes for Bush
↪  and votes for Buchanan.")
```
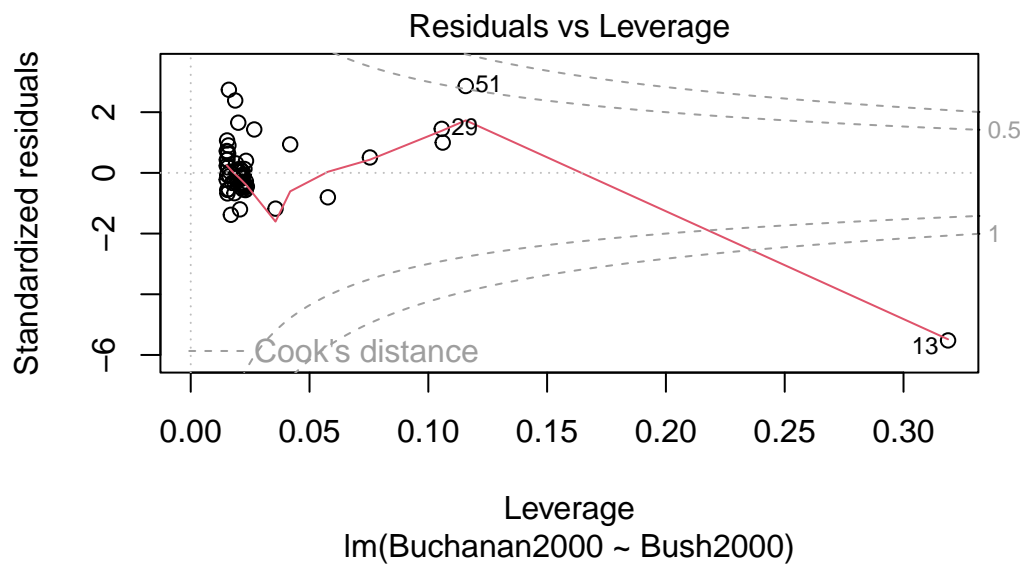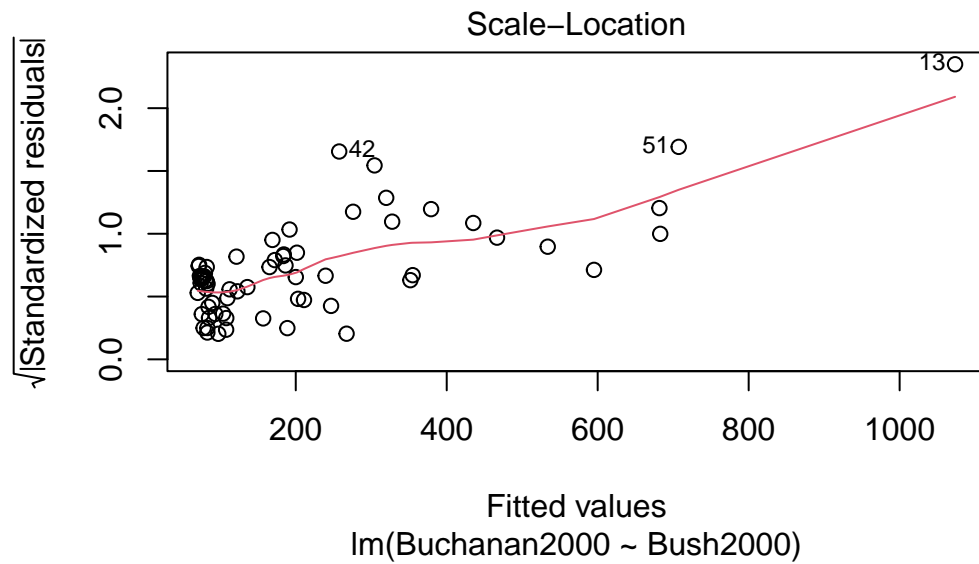

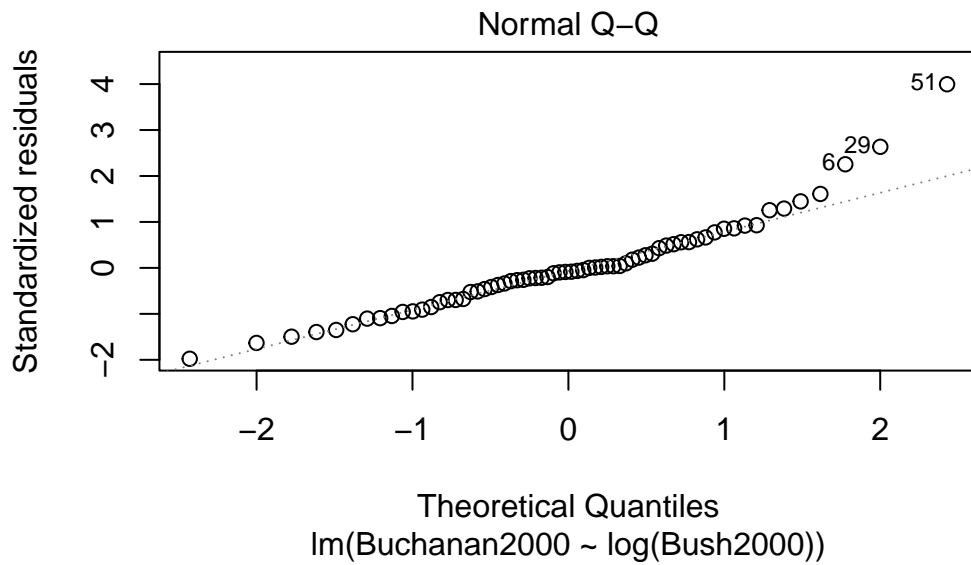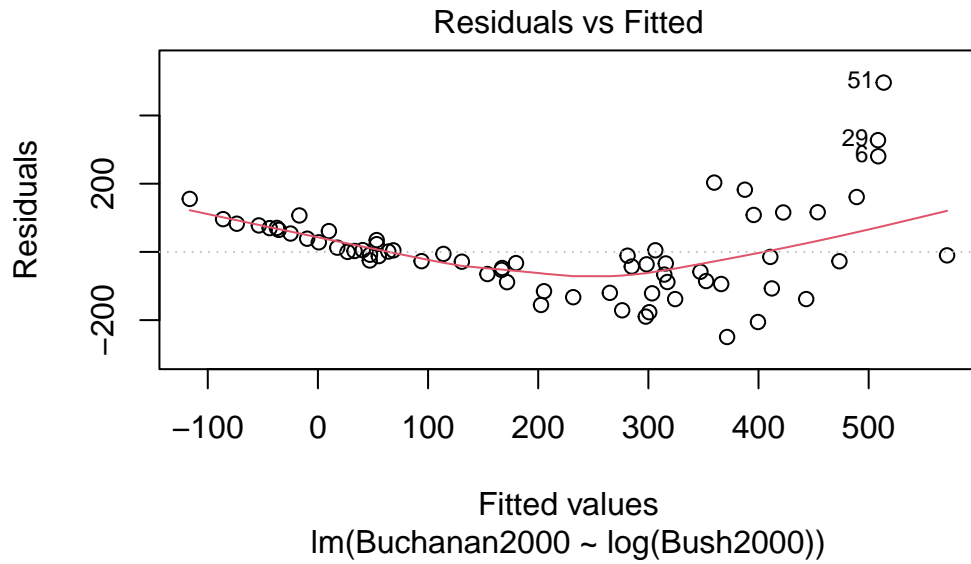Transformed Association between votes for Bush and votes for B

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

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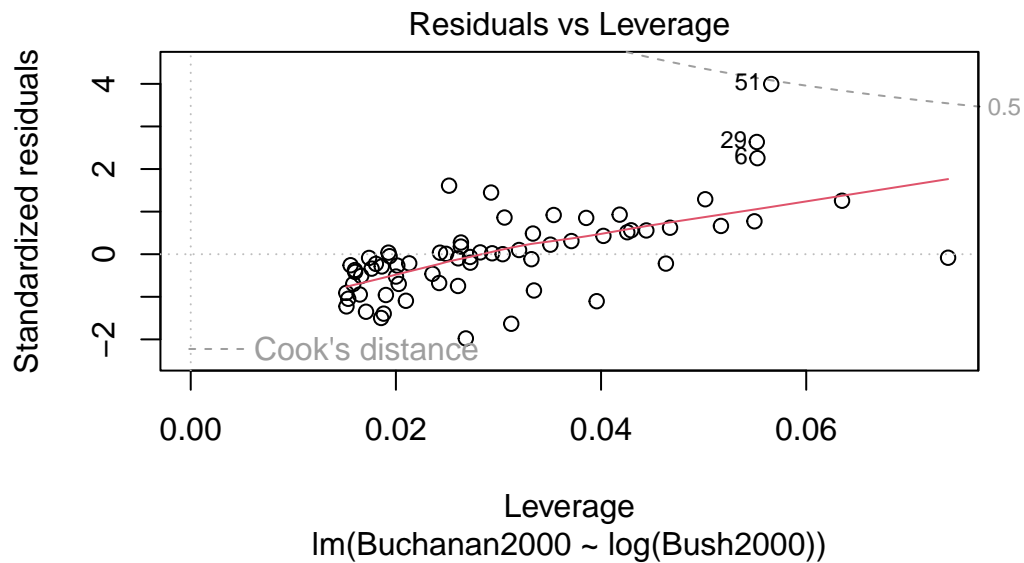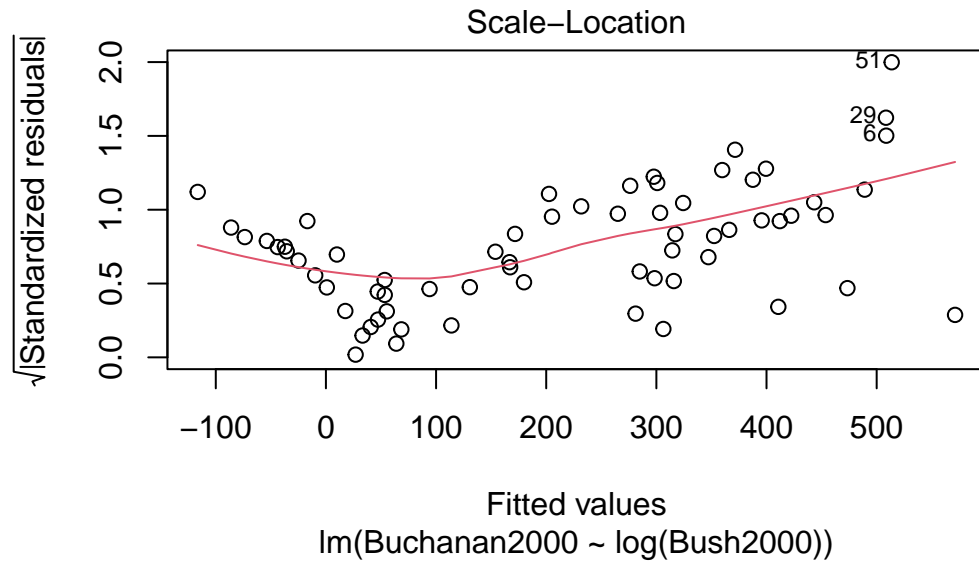
Residuals vs Fitted

Fitted values
lm(Buchanan2000 ~ Bush2000)

Normal Q–Q

Theoretical Quantiles
lm(Buchanan2000 ~ Bush2000)

Scale−Location

lm(Buchanan2000 ~ Bush2000)

Residuals vs Leverage
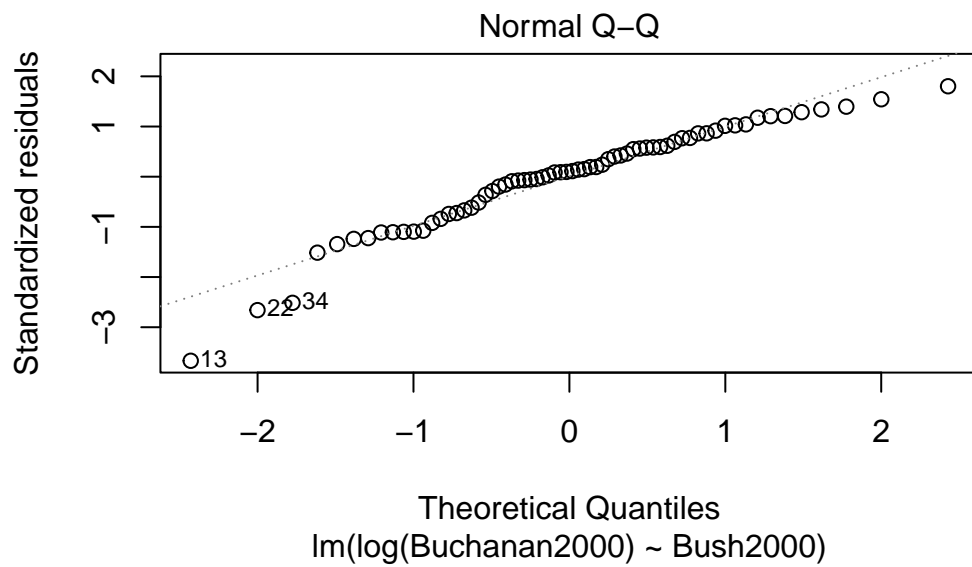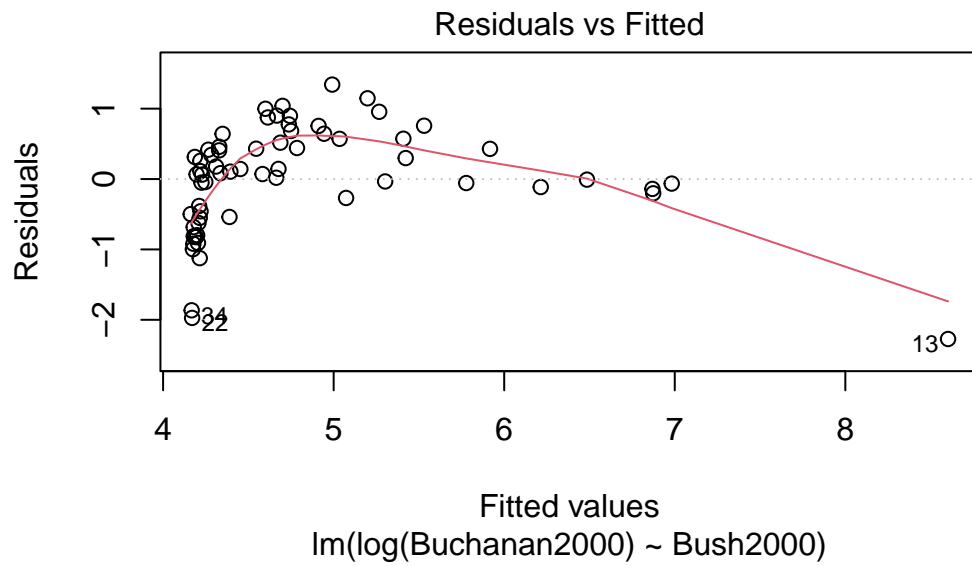

lm(Buchanan2000 ~ Bush2000)

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

Residuals vs Fitted

Residuals

51

29
60

Fitted values
lm(Buchanan2000 ~ log(Bush2000))

Normal Q−Q

Standardized residuals

51

29
6

Theoretical Quantiles
lm(Buchanan2000 ~ log(Bush2000))

Scale–Location

lm(Buchanan2000 ~ log(Bush2000))

Residuals vs Leverage

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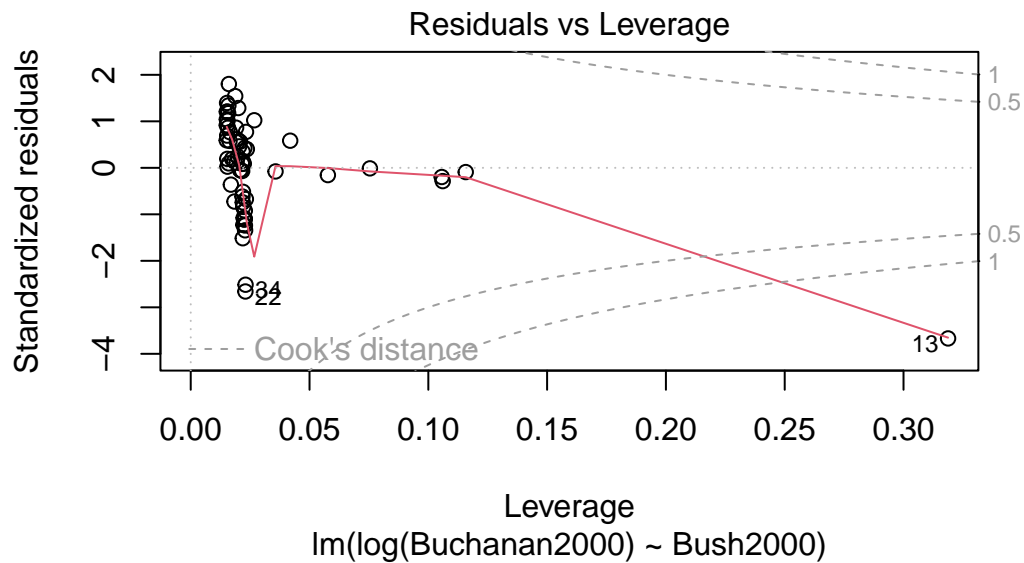
```r
# Model with a logarithmic response variable transformation
ytransformed <- lm(log(Buchanan2000) ~ Bush2000, data = election_wo_pb)
plot(ytransformed)
```
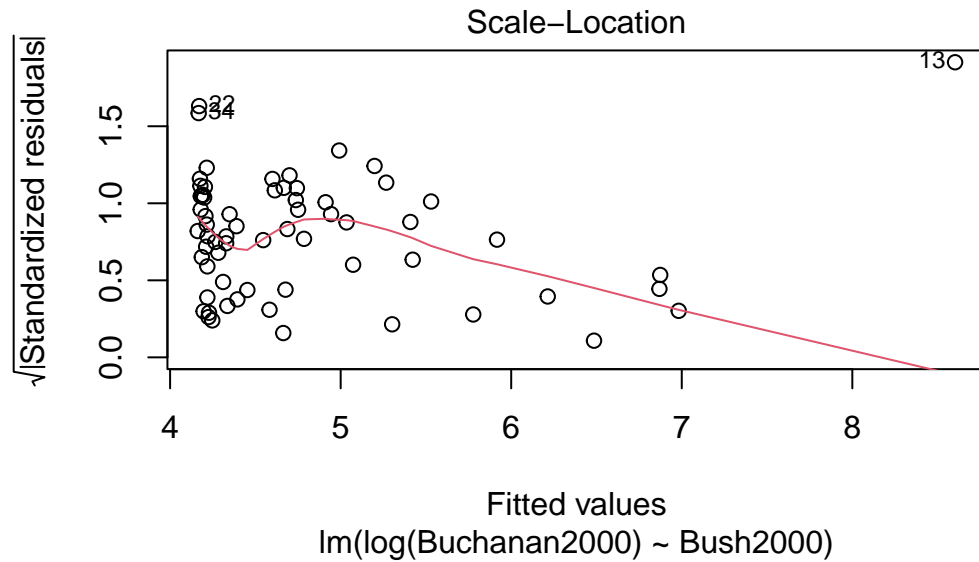
Residuals vs Fitted

lm(log(Buchanan2000) ~ Bush2000)

Normal Q−Q

lm(log(Buchanan2000) ~ Bush2000)

Scale–Location

lm(log(Buchanan2000) ~ Bush2000)

Residuals vs 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)

# 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 9.07e-09 ***
log(Bush2000)  0.73096    0.03597  20.323  < 2e-16 ***
---
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: < 2.2e-16
```

```
# 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
```