

Example Project: Hanging Chads

Jason Bryer, Ph.D.

2000 Election

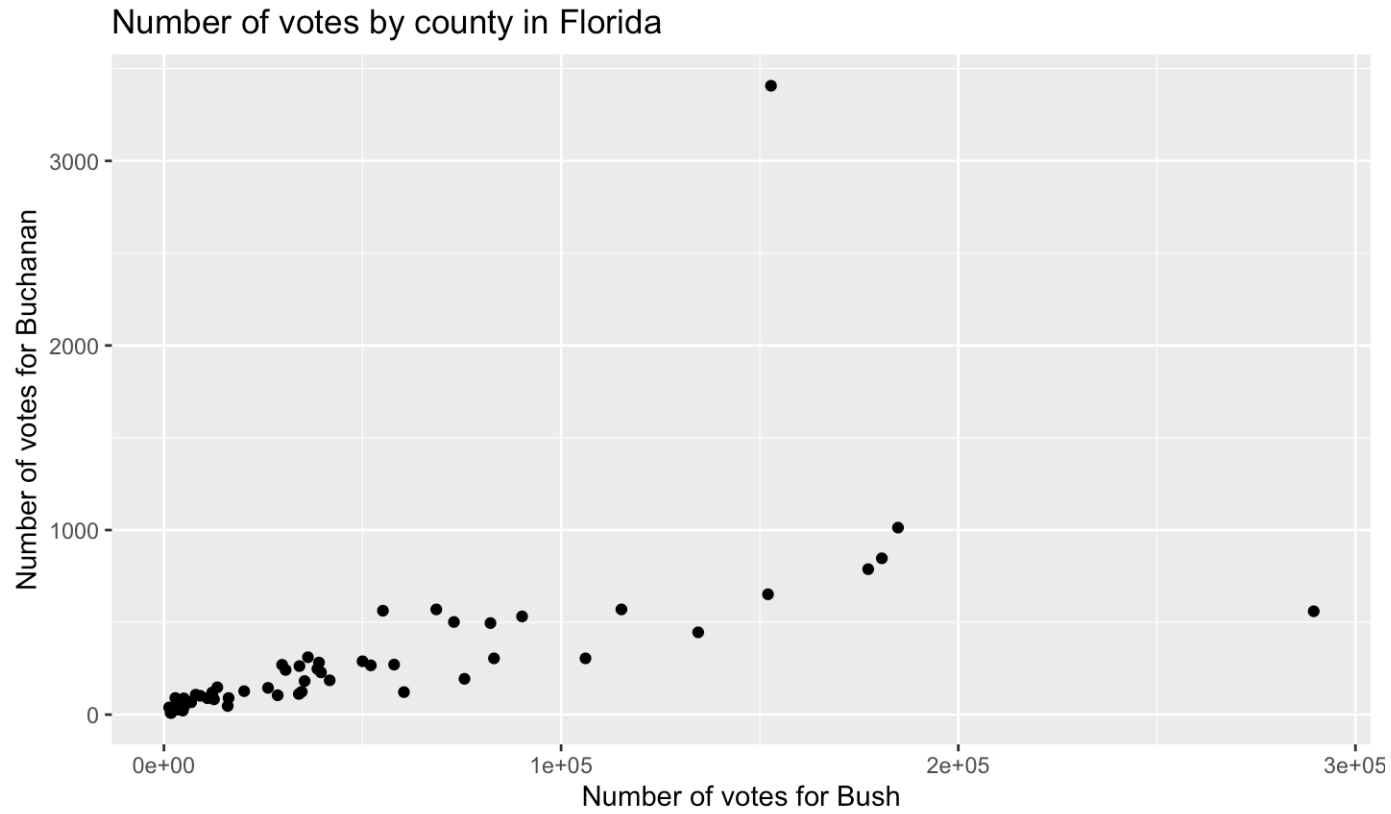
The 2000 election between George Bush and Al Gore was ultimately decided in Florida. However, there was a third candidate on the ballot, Pat Buchanan, and one county with an unpredictable outcome. Is there evidence that a large number of votes were cast for a mistaken candidate?

The `elections` data frame contains the breakdown of votes by each of the 67 counties in Florida.

There are 67 counties in Florida that cast a total of 2,910,078 votes for George Bush and 2,909,117 resulting in Bush winning by 961 votes.

However, in the days following the election there was much controversy surrounding so called “hanging chads.” That is, there were a number of ballots where it was not clear who the vote was for. This was a particular issue in Palm Beach.

Number of votes by county in Florida



Correlation

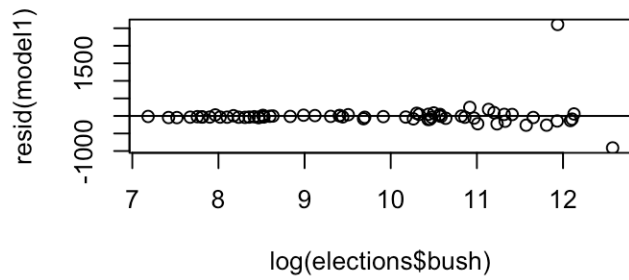
```
##  
## Pearson's product-moment correlation  
##  
## data: elections$buch and elections$bush  
## t = 6.455, df = 65, p-value = 1.574e-08  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.4527668 0.7522709  
## sample estimates:  
##      cor  
## 0.6250012
```

Linear Regression Model

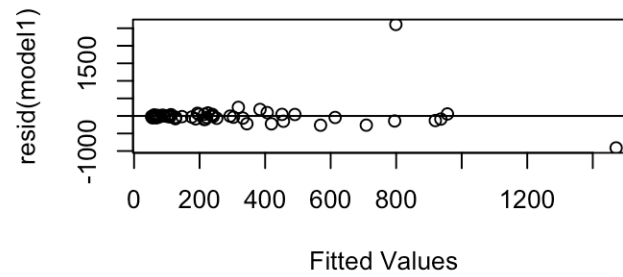
```
##  
## Call:  
## lm(formula = buch ~ bush, data = elections)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -911.30  -46.11  -26.05   12.01 2608.01   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) 4.697e+01  5.446e+01   0.863   0.392      
## bush         4.920e-03  7.622e-04   6.455 1.57e-08 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 353.9 on 65 degrees of freedom  
## Multiple R-squared:  0.3906, Adjusted R-squared:  0.3813   
## F-statistic: 41.67 on 1 and 65 DF,  p-value: 1.574e-08
```

Residual Analysis

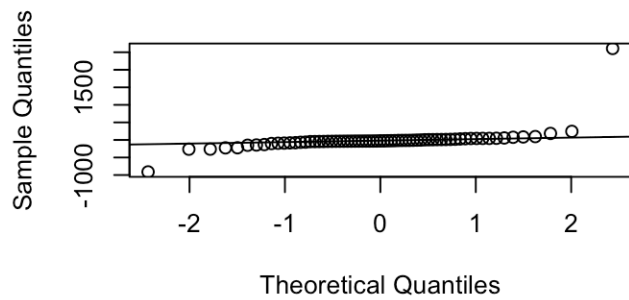
Predictors vs Residuals



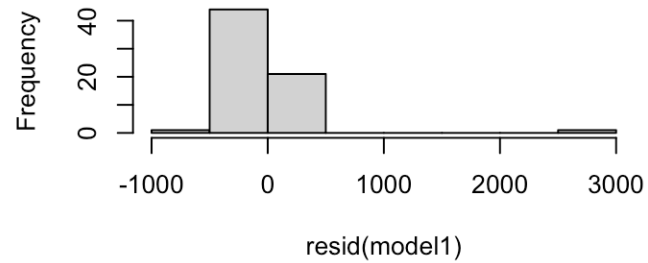
Fitted vs Residuals



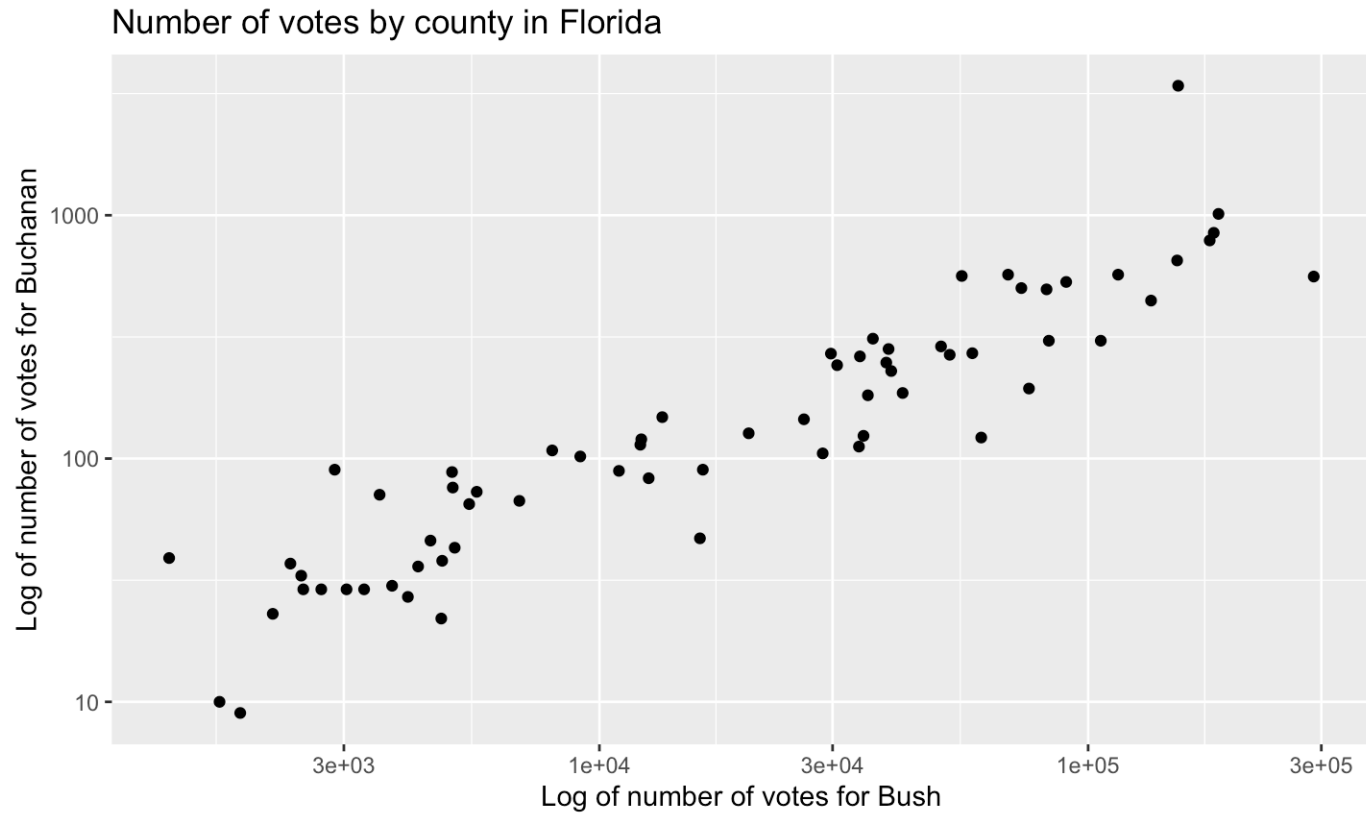
QQ-Plot of Residuals



Histogram of Residuals



Log Transform



Correlation with log transformations

```
##  
## Pearson's product-moment correlation  
##  
## data: log(elections$buch) and log(elections$bush)  
## t = 19.222, df = 65, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.8760098 0.9515894  
## sample estimates:  
##      cor  
## 0.9221706
```

Linear Regression Model (log transform)

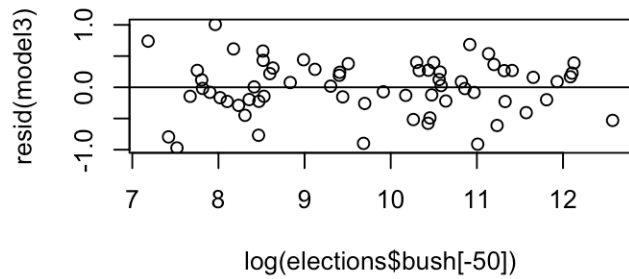
```
##  
## Call:  
## lm(formula = log(buch) ~ log(bush), data = elections)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.97038 -0.24247  0.00825  0.25452  1.65752   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -2.55079    0.38903  -6.557 1.04e-08 ***   
## log(bush)    0.75620    0.03934  19.222 < 2e-16 ***   
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.4672 on 65 degrees of freedom  
## Multiple R-squared:  0.8504, Adjusted R-squared:  0.8481   
## F-statistic: 369.5 on 1 and 65 DF,  p-value: < 2.2e-16
```

Regression model without Palm Beach

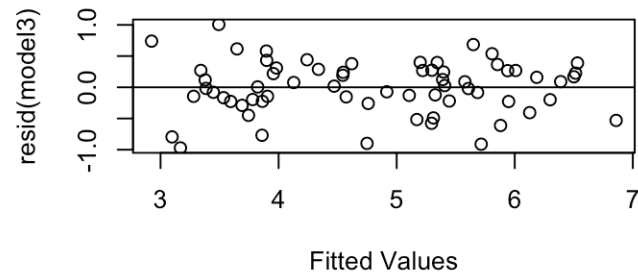
```
##  
## Call:  
## lm(formula = log(buch) ~ log(bush), data = elections[-50, ])  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.97136 -0.22384  0.02279  0.26959  1.00652  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -2.31657    0.35470  -6.531 1.23e-08 ***  
## log(bush)    0.72960    0.03599  20.271 < 2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.4203 on 64 degrees of freedom  
## Multiple R-squared:  0.8652, Adjusted R-squared:  0.8631  
## F-statistic: 410.9 on 1 and 64 DF,  p-value: < 2.2e-16
```

Residual Analysis (log)

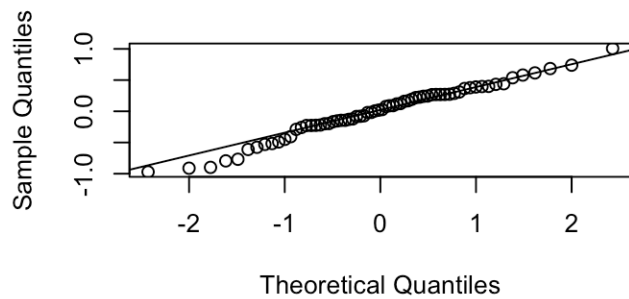
Predictors vs Residuals



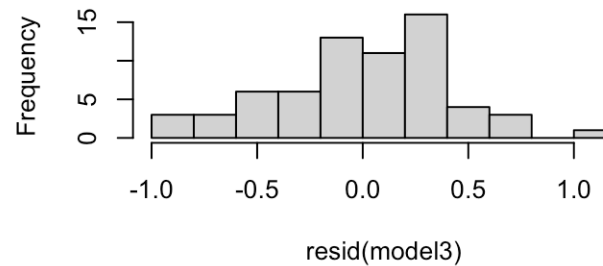
Fitted vs Residuals



QQ-Plot of Residuals



Histogram of Residuals



Predict Palm Beach from the model

Obtain the predicted vote count for Palm Beach given the fitted model without

The difference between predicted on the original scale and the observed vote count

```
##          1  
## 2809.498
```

Predict Palm Beach from the model (cont.)

Prediction Confidence Interval for log(vote count)

```
##           fit      lwr      upr
## 1 6.392757 5.532353 7.253162
```

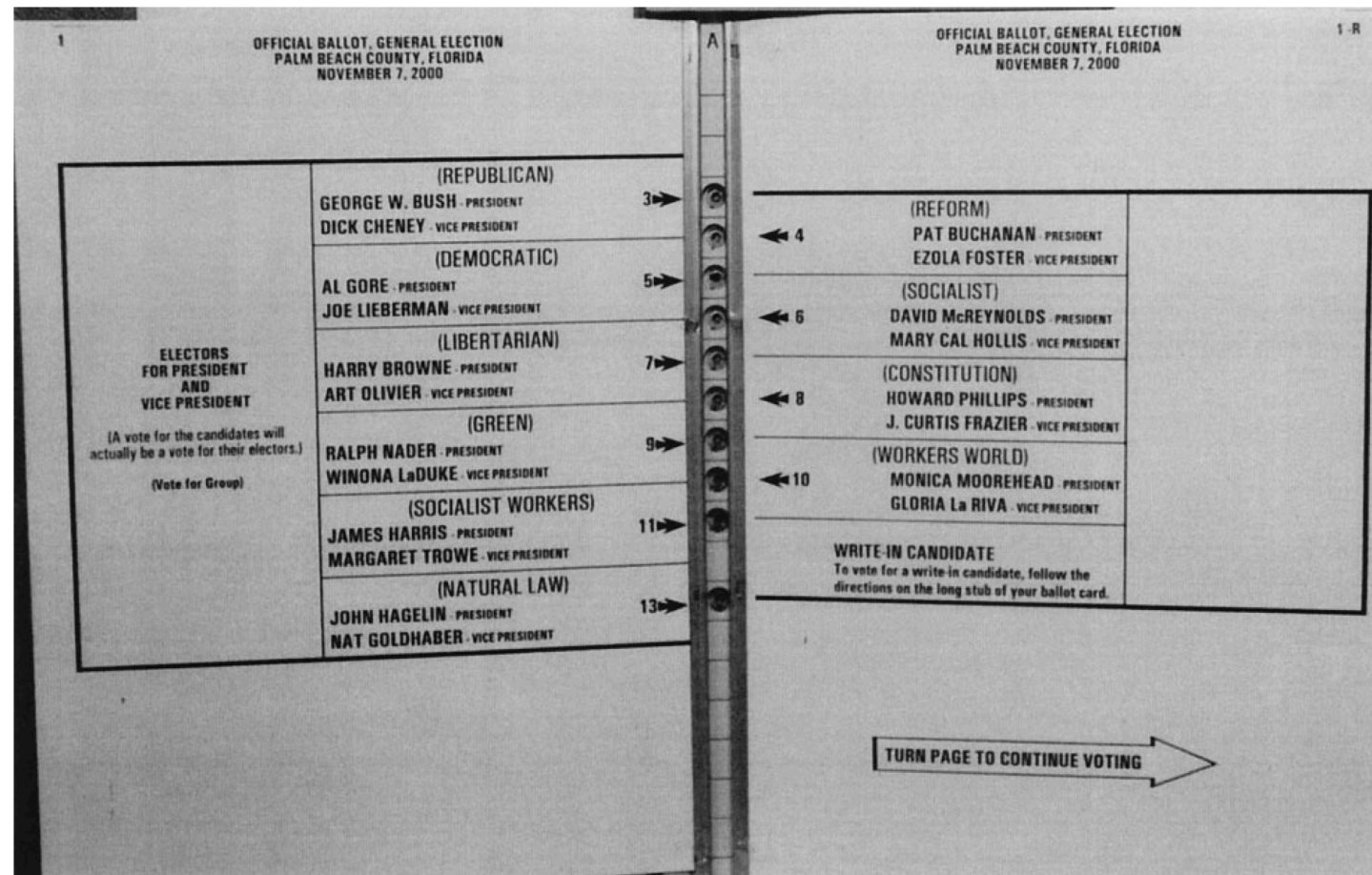
Prediction Confidence Interval on the original scale

```
##           fit      lwr      upr
## 1 597.5019 252.738 1412.564
```

```
## [1] 3407
```

Therefore, what we can say is that it is likely that Palm Beach is a different community.

FIGURE 1. The Palm Beach County Butterfly Ballot



Source: AP Worldwide Photos, Gary I. Rothstein. Reprinted with permission.

References

Wand, J.N., Shotts, K.W., Sekhon, J.S., Mebane, W.R., Herron, M.C., & Brady, H.E. (2001). [The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida](#). *American Political Science Review*, 95(4).

Smith, R.L. (2002). [A Statistical Assessment of Buchanan's Vote in Palm Beach County](#). *Statistical Science*, 17(4).

[Regression Analysis - 1.3.3 - 2000 Elections in Florida \(Youtube\)](#)