

ballots

Chris Parrish

January 9, 2016

ballots

references:

- Cannon, et al., Stat2, chapter 04, example 4.3

Import the data.

```
data <- read.csv("PalmBeach.csv", header=TRUE)
head(data)
```

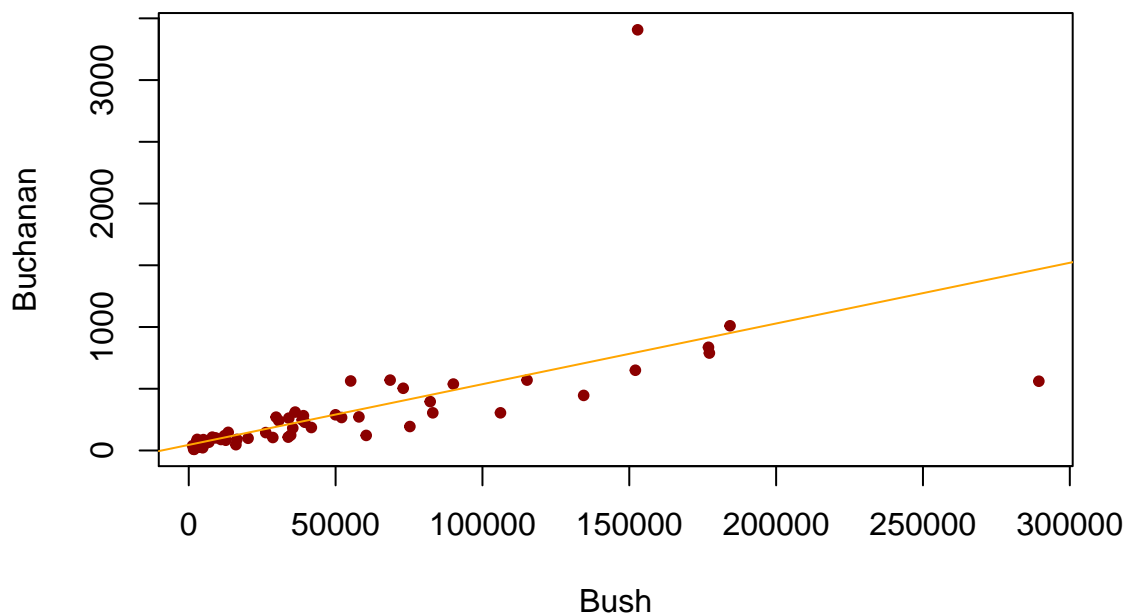
```
##      County Buchanan  Bush
## 1 ALACHUA      262 34062
## 2  BAKER       73  5610
## 3   BAY       248 38637
## 4 BRADFORD     65  5413
## 5 BREVARD     570 115185
## 6 BROWARD     789 177279
```

```
dim(data)
```

```
## [1] 67 3
```

View the data.

```
plot(Buchanan ~ Bush, data=data,
     pch=20, col="darkred")
ballots.lm <- lm(Buchanan ~ Bush, data=data)
abline(ballots.lm, col="orange")
```



Unusual points.

```
ballots.diag <- ls.diag(ballots.lm)
summary(ballots.diag)
```

```
##           Length Class  Mode
## std.dev      1  -none- numeric
## hat          67  -none- numeric
## std.res      67  -none- numeric
## stud.res     67  -none- numeric
## cooks       67  -none- numeric
## dfits        67  -none- numeric
## correlation   4  -none- numeric
## std.err       2  -none- numeric
## cov.scaled    4  -none- numeric
## cov.unscaled  4  -none- numeric
```

Leverage of point (x_i, y_i) .

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}$$

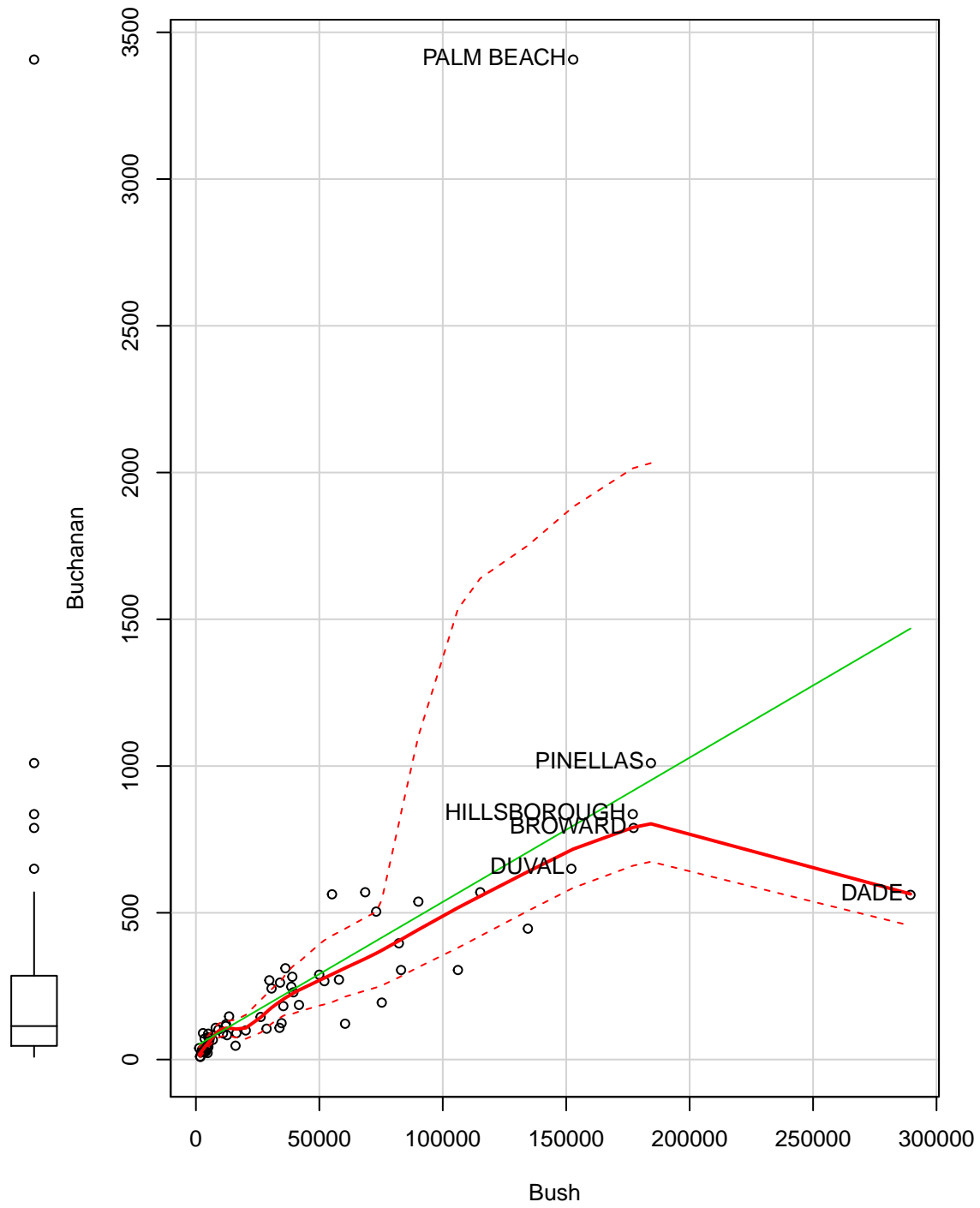
Points with high leverage.

```
n <- nrow(data)
typical.leverage <- 2 / n
hi.threshold <- 2 * typical.leverage
data[ballots.diag$hat > hi.threshold, ]
```

```
##           County Buchanan  Bush
## 6           BROWARD      789 177279
## 13            DADE       561 289456
## 16            DUVAL       650 152082
## 29 HILLSBOROUGH       836 176967
## 50     PALM BEACH     3407 152846
## 52     PINELLAS     1010 184312
```

Identify unusual points.

```
library(car)
scatterplot(Buchanan ~ Bush, data=data,
            id.n=6, labels=data$County)
```



County	Count
BROWARD	6
DADE	13
DUVAL	16
HILLSBOROUGH	29
PALM BEACH	50
PINELLAS	52

Standardized residuals.

```
hi.threshold <- 3 * typical.leverage
std.res.threshold <- 2
idx <- (1:67)[ballots.diag$hat > hi.threshold |
            abs(ballots.diag$std.res) > std.res.threshold]
results <- cbind(data[idx, ],
                ballots.lm$fit[idx],
                ballots.lm$resid[idx],
                ballots.diag$std.res[idx])
names(results) <- c("County", "Buchanan", "Bush", "Fits", "Resids", "Std. Resids")
results
```

##	County	Buchanan	Bush	Fits	Resids	Std. Resids
## 6	BROWARD	789	177279	916.9402	-127.94023	-0.3807500
## 13	DADE	561	289456	1468.4953	-907.49526	-3.0591800
## 29	HILLSBOROUGH	836	176967	915.4062	-79.40618	-0.2362617
## 50	PALM BEACH	3407	152846	796.8074	2610.19263	7.6510719
## 52	PINELLAS	1010	184312	951.5203	58.47972	0.1749128