Assignment - 2

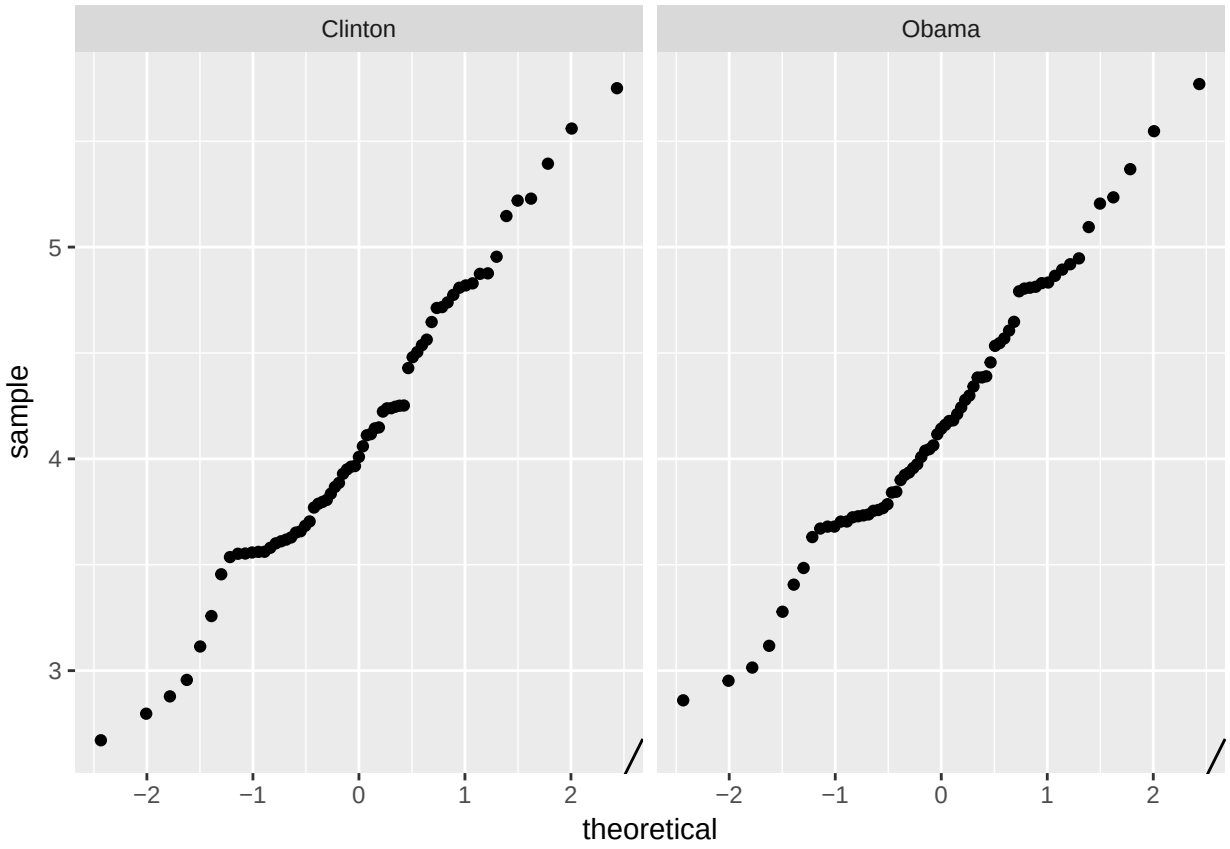
Dwipam

1/25/2017

Question 1:

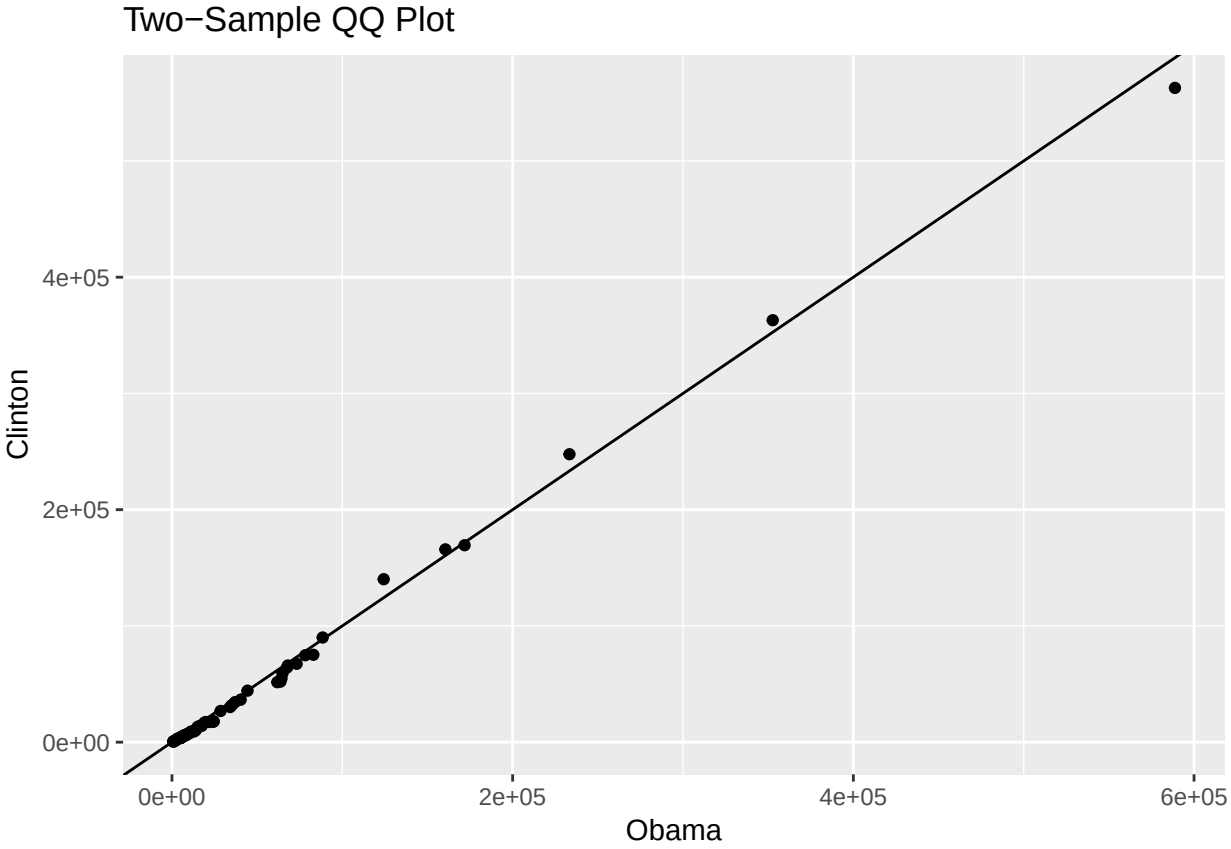
```
library(tidyr)
votes = read.table("pennsylvania.txt",sep="",header = TRUE)
votes.long = votes %>% gather(President,votes,Obama:Clinton)
```

Question 2:

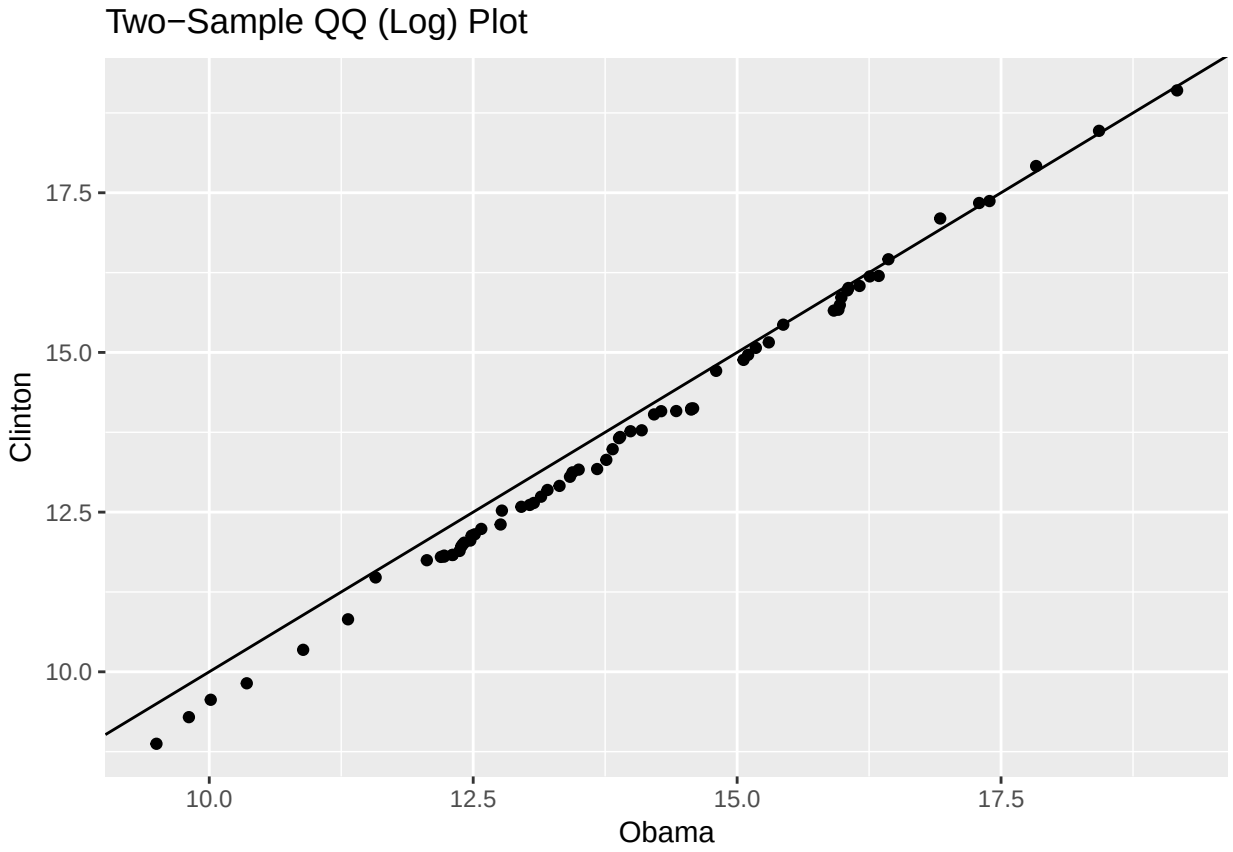


Looking visually at the log 10 qq plot of votes by we can say that, after the transformation, data is approximately normal, as it does have fairly straight line. By looking at the spread it sems that Obama, has slightly higher votes than clinton.

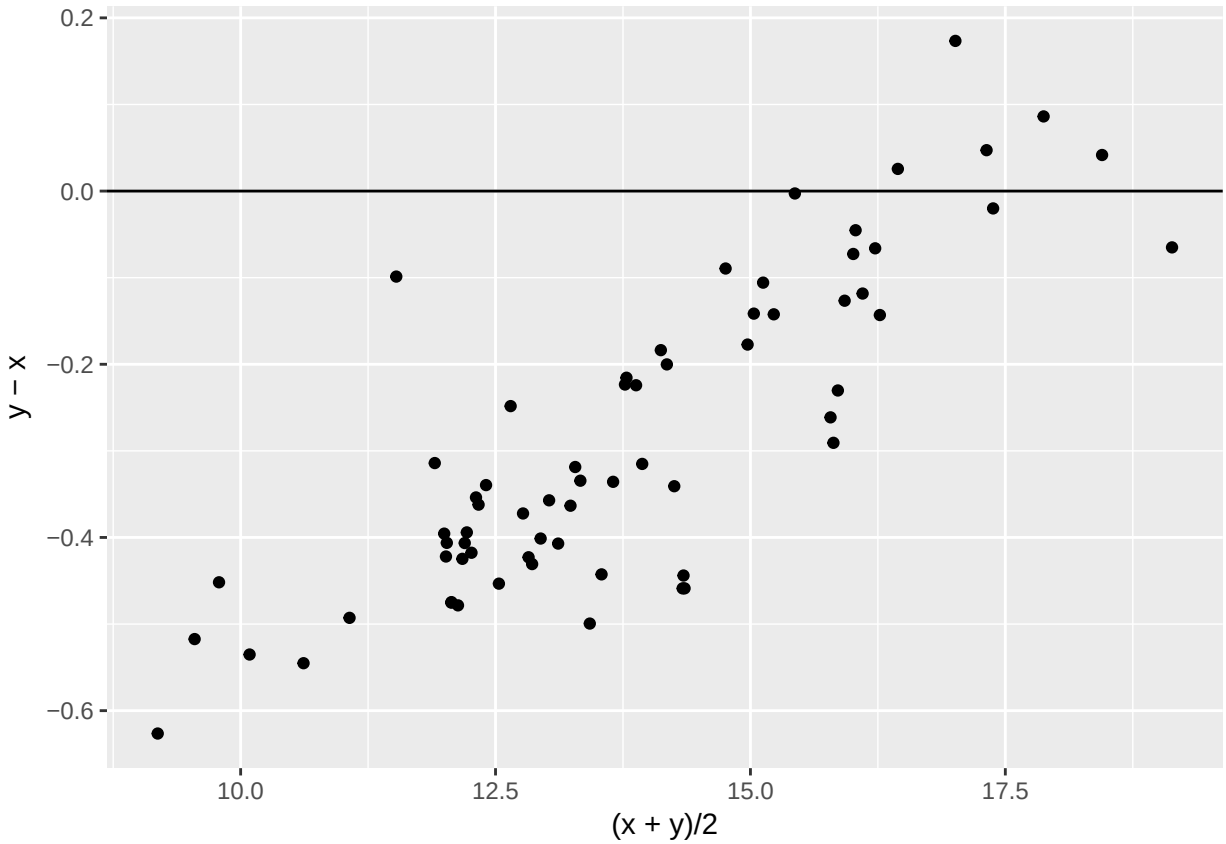
Question 3:



It seems that y=x does fit the plot. Line with this slope seems to fit the data well. However to assess this futher we can check the log transformation.

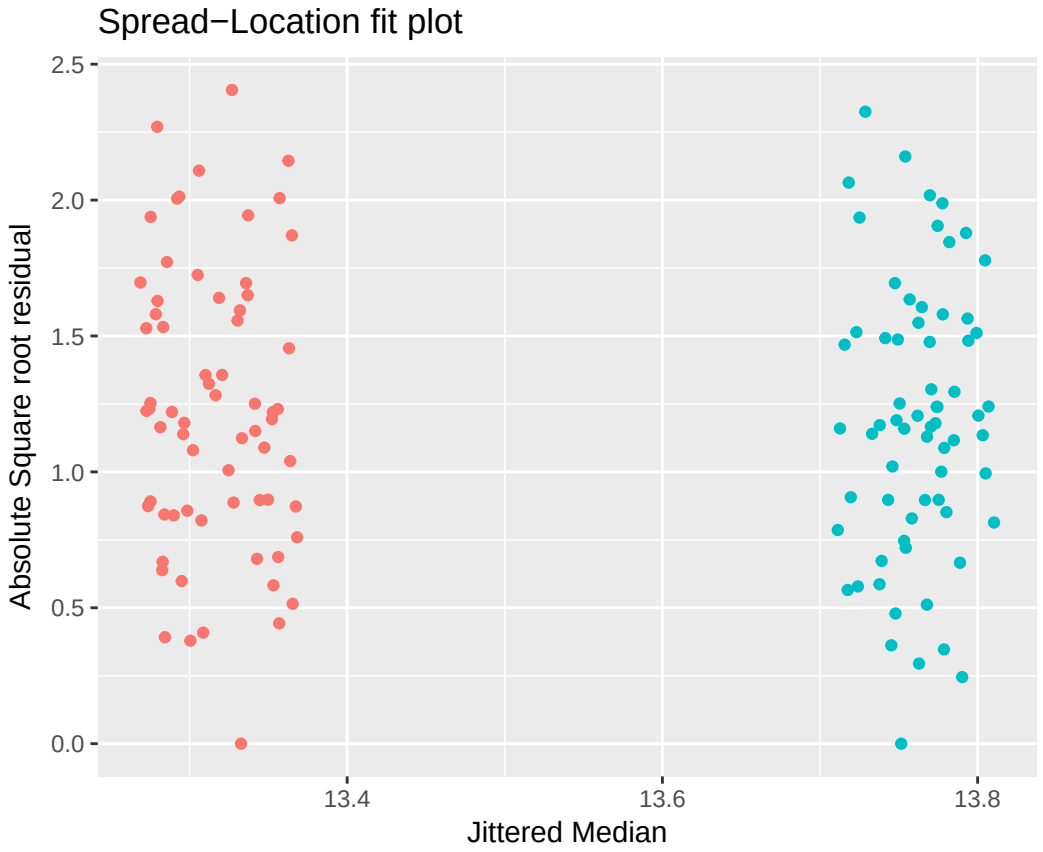


After doing log transformation, it seems that slope is changing slightly, as the points lie above and below the line. Let's assess with Tuke-difference plot.



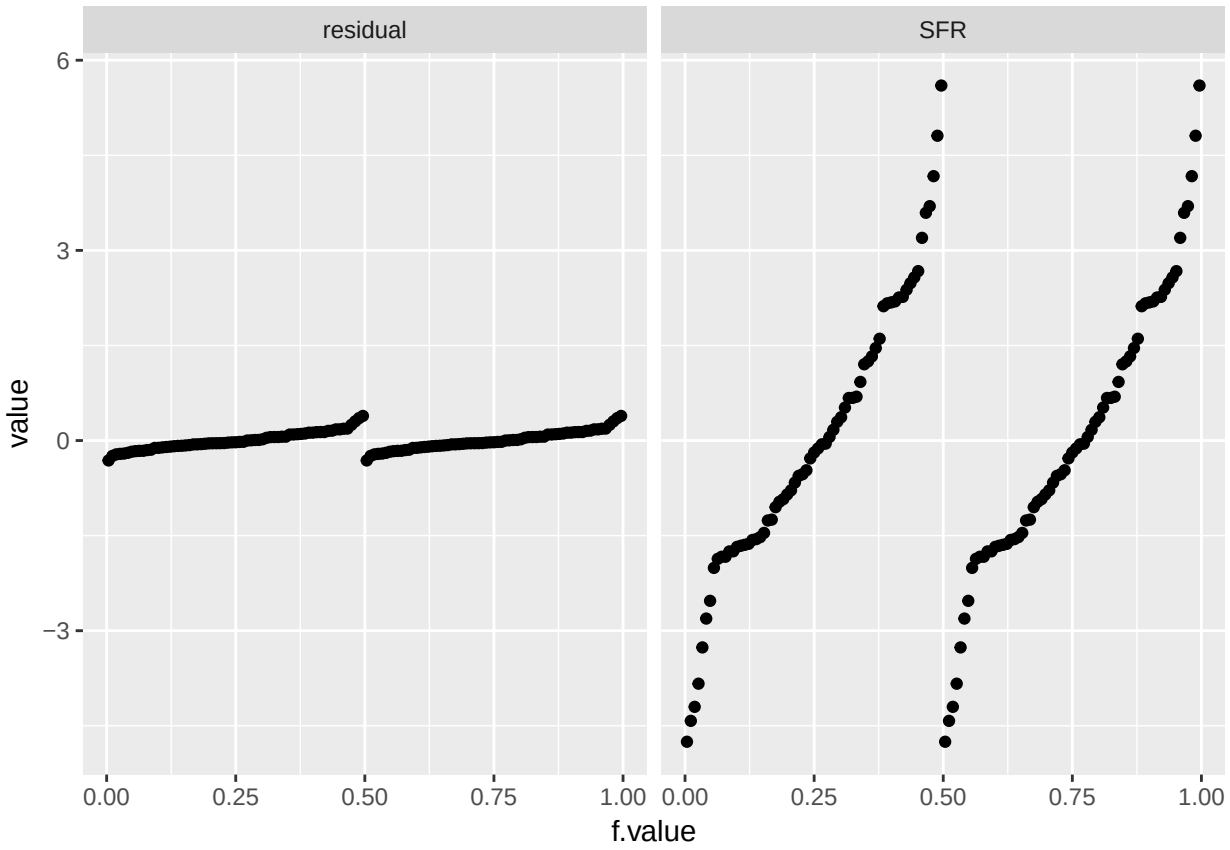
It seems that, As major points are below 0, comparing to the average Clinton did have few votes than Obama in general. We can also say that, as the average increases Clinton starts receiving more votes. We can say that model is complicated. Let's assess with spread-residual plot. Here, let's consider median as it is robust to outlier.

This seems to have a pretty good spread of residual. Lets check with Log2 transform. Also difference between



the median is quite significant.

This looks better than previous plot in terms of spread. Let's fit a line and assess the residual-fit plot to Log2 transformation.



We can see that, good amount of variance is being captured, still there is some variance left in residuals.