EDA - Assignment2

Jivitesh Poojary

January 23, 2017

Answer 1

Q1. After reading the data into R, use gather() to convert it to "long form". (You don't need to write anything up for the question, just give the code.)

Code Provided.

```
##
## Attaching package: 'dplyr'
```

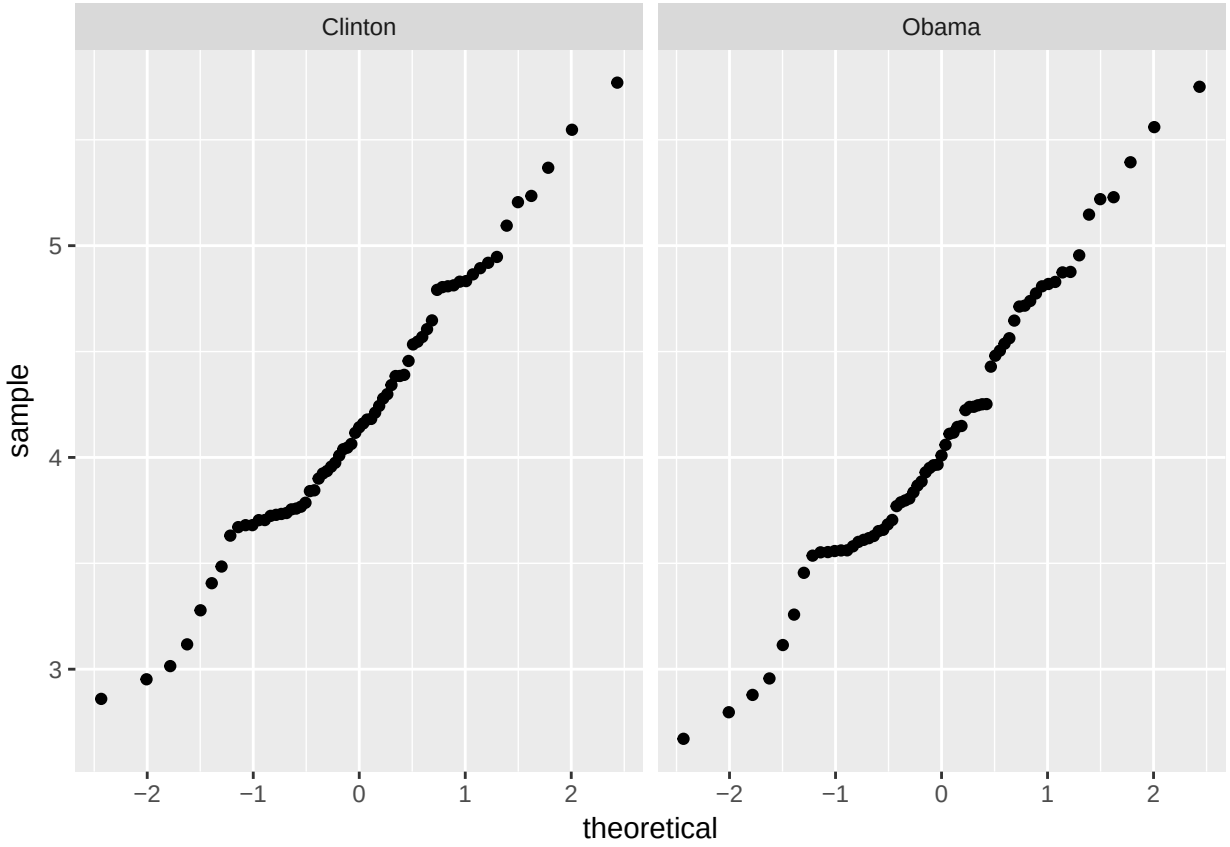
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

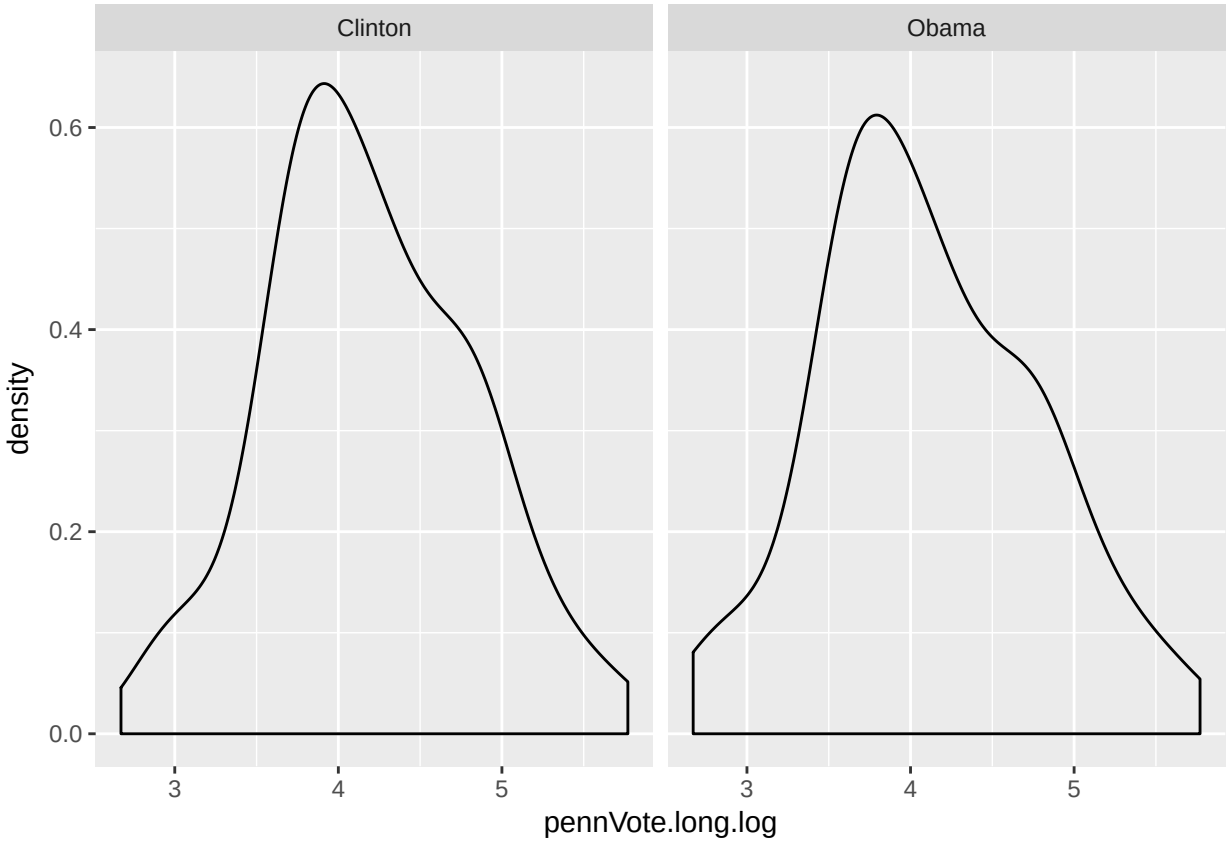
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

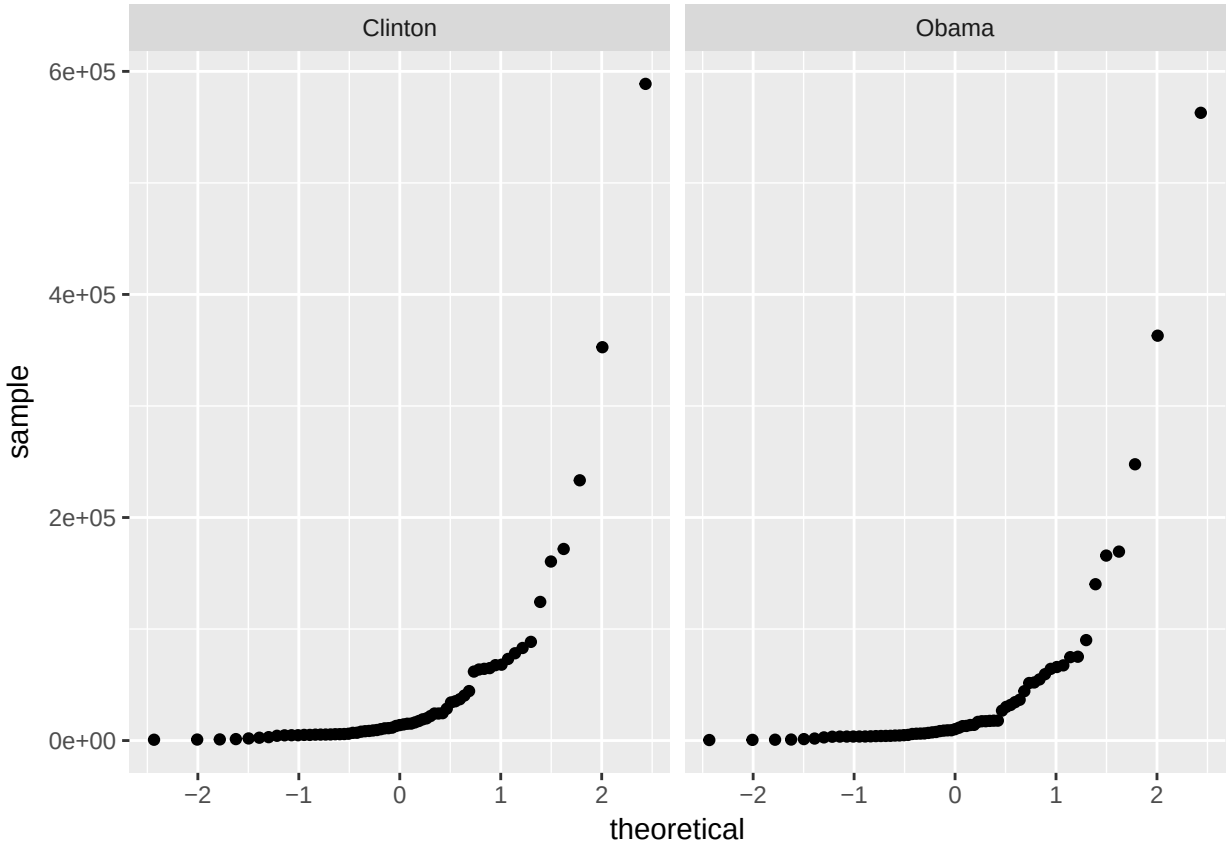
Answer 2

Q2. Use ggplot() to reproduce this normal QQ plot of the log_10 transformed data, and explain what it tells you:

The log transformation changes the skewness to a more symmetric distribution rendering a simplified structure. The two plots are similar in distribution. The plots obtained do not follw a normal distribution. We can say that they may be a case of over tranformation as we are getting a slightly 'S' shaped curve.



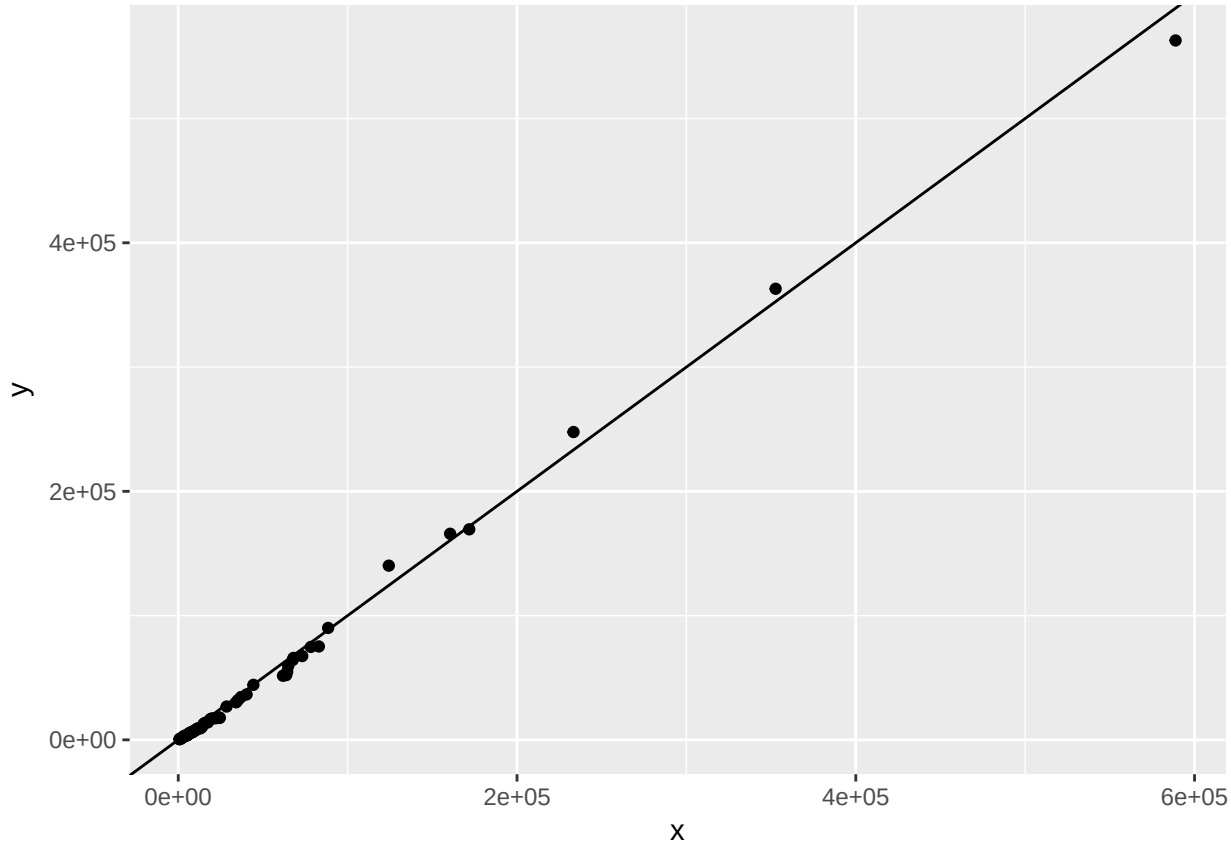

Clinton
Obama
density
0.6
0.4
0.2
0.0
3
4
5
pennVote.long.log

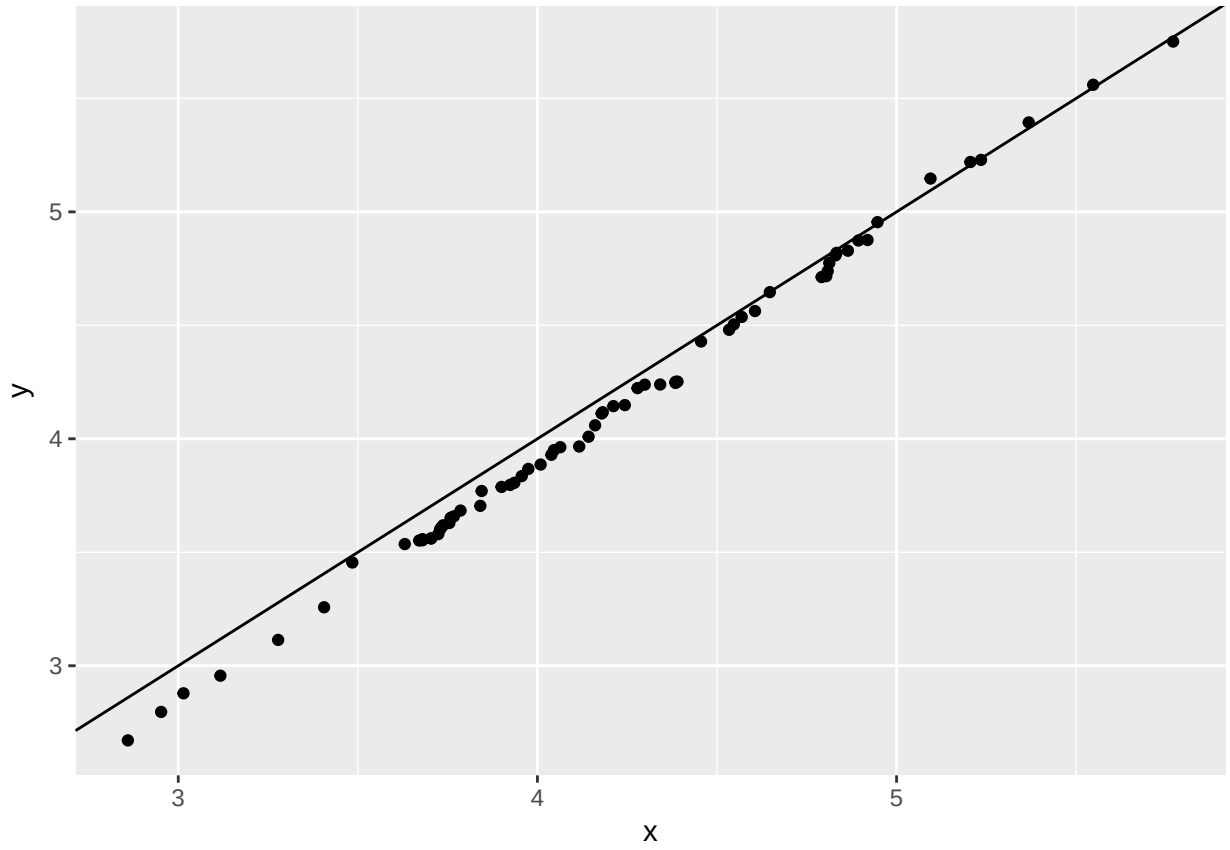

Clinton
Obama
6e+05
4e+05
2e+05
0e+00
sample
−2
−1
0
1
2
theoretical

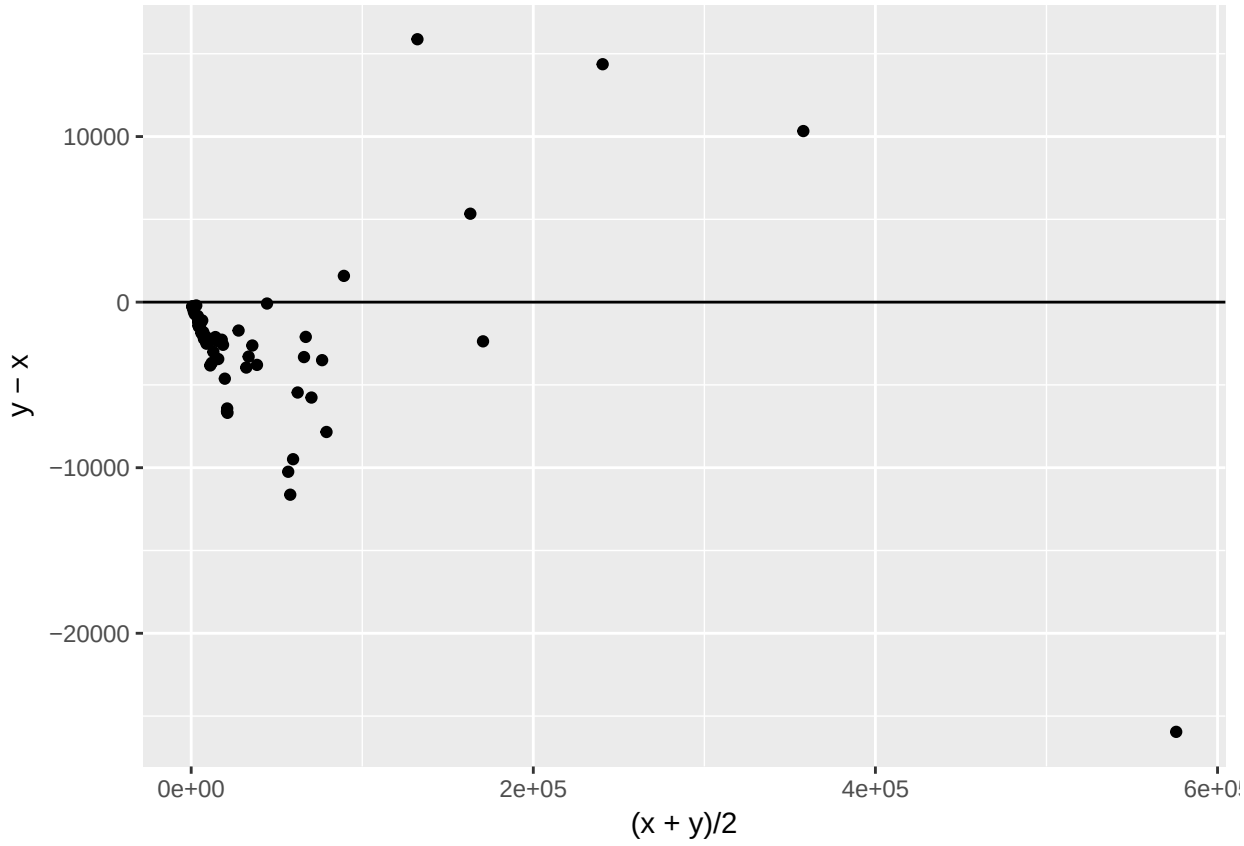
Answer 3

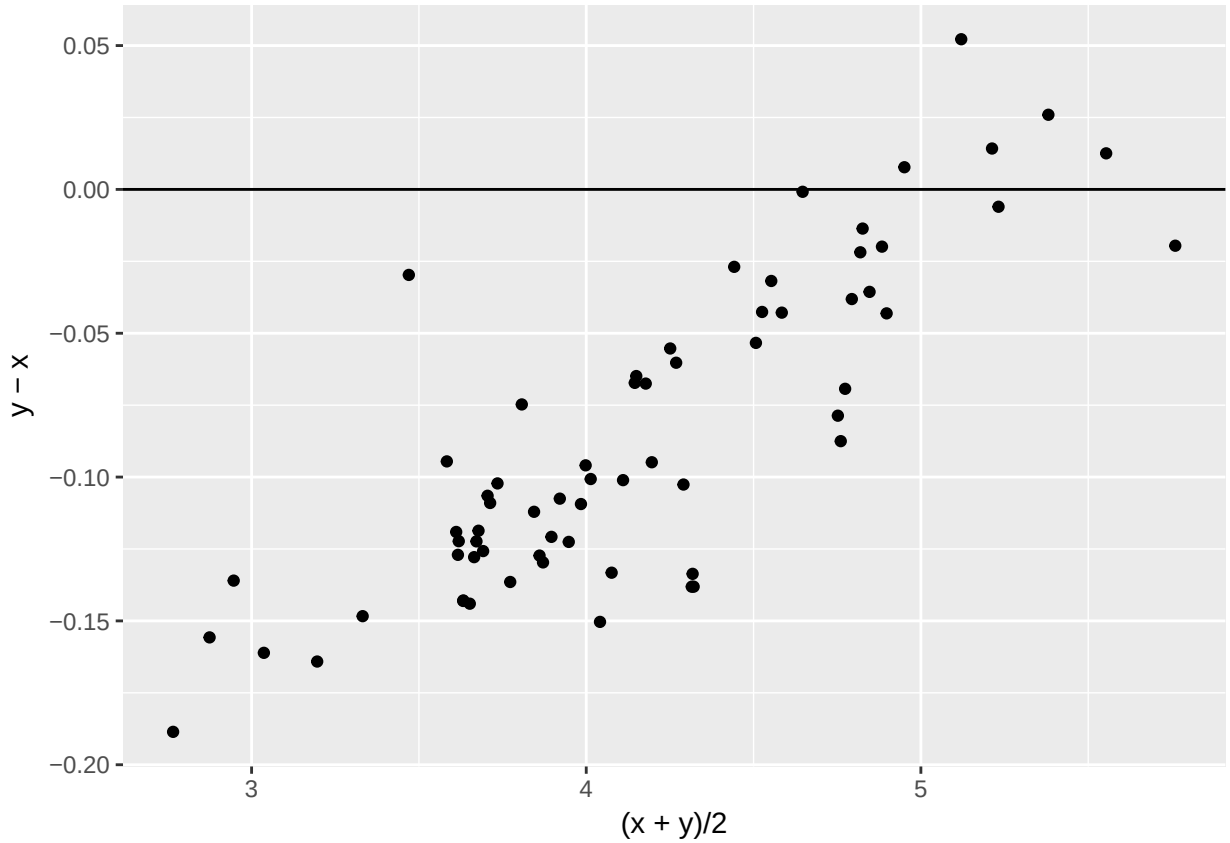
Q3. Does it look like the relationship between Clinton's vote and Obama's vote is (i) additive, (ii) multiplicative, or (iii) more complicated? Justify your answer. (You'll need to draw further graphs to answer this question; include one or two of them in your answer.)

The plot obtained without transformation is not normal.




5
4
3
y
3
4
5
x


10000
0
−10000
−20000
y − x
0e+00
2e+05
4e+05
(x + y)/2


0.05
0.00
−0.05
−0.10
−0.15
−0.20
y − x
3
4
5
(x + y)/2