

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

##           County Candidate  Votes
## 1           Adams      Obama  15091
## 2      Allegheny      Obama 352687
## 3      Armstrong      Obama   9045
## 4           Beaver      Obama  37055
## 5       Bedford      Obama   4788
## 6           Berks      Obama  83011
## 7           Blair      Obama 16276
## 8      Bradford      Obama   8624
## 9           Bucks      Obama 160521
## 10          Butler      Obama  28550
## 11         Cambria      Obama  24249
## 12         Cameron      Obama    724
## 13          Carbon      Obama  11580
## 14          Centre      Obama  34176
## 15         Chester      Obama 124311
## 16         Clarion      Obama   5056
## 17    Clearfield      Obama  11121
## 18         Clinton      Obama   5734
## 19         Columbia      Obama  10937
## 20         Crawford      Obama  13883
## 21    Cumberland      Obama  44367
## 22         Dauphin      Obama  64965
## 23        Delaware      Obama 171792
## 24             Elk      Obama   5463
## 25            Erie      Obama  68036
## 26         Fayette      Obama  21971
## 27          Forest      Obama    896
## 28        Franklin      Obama  18995
## 29          Fulton      Obama   1310
## 30         Greene      Obama   5852
```

## 31	Huntingdon	Obama	5409
## 32	Indiana	Obama	14473
## 33	Jefferson	Obama	4787
## 34	Juniata	Obama	2547
## 35	Lackawanna	Obama	61838
## 36	Lancaster	Obama	88481
## 37	Lawrence	Obama	17513
## 38	Lebanon	Obama	19900
## 39	Lehigh	Obama	78283
## 40	Luzerne	Obama	64307
## 41	Lycoming	Obama	15203
## 42	McKean	Obama	5297
## 43	Mercer	Obama	24232
## 44	Mifflin	Obama	4273
## 45	Monroe	Obama	35221
## 46	Montgomery	Obama	233356
## 47	Montour	Obama	3053
## 48	Northampton	Obama	67606
## 49	Northumberland	Obama	13072
## 50	Perry	Obama	5685
## 51	Philadelphia	Obama	588806
## 52	Pike	Obama	10210
## 53	Potter	Obama	1897
## 54	Schuylkill	Obama	24546
## 55	Snyder	Obama	4687
## 56	Somerset	Obama	9436
## 57	Sullivan	Obama	1034
## 58	Susquehanna	Obama	6935
## 59	Tioga	Obama	5357
## 60	Union	Obama	6109
## 61	Venango	Obama	7945
## 62	Warren	Obama	6995
## 63	Washington	Obama	40345
## 64	Wayne	Obama	8396
## 65	Westmoreland	Obama	63722
## 66	Wyoming	Obama	5061
## 67	York	Obama	73191
## 68	Adams	Clinton	14077
## 69	Allegheny	Clinton	363017
## 70	Armstrong	Clinton	6849
## 71	Beaver	Clinton	30225
## 72	Bedford	Clinton	3613
## 73	Berks	Clinton	75169
## 74	Blair	Clinton	13093
## 75	Bradford	Clinton	6263
## 76	Bucks	Clinton	165861
## 77	Butler	Clinton	26834
## 78	Cambria	Clinton	17348
## 79	Cameron	Clinton	469
## 80	Carbon	Clinton	8917
## 81	Centre	Clinton	36555
## 82	Chester	Clinton	140188
## 83	Clarion	Clinton	4256
## 84	Clearfield	Clinton	7700

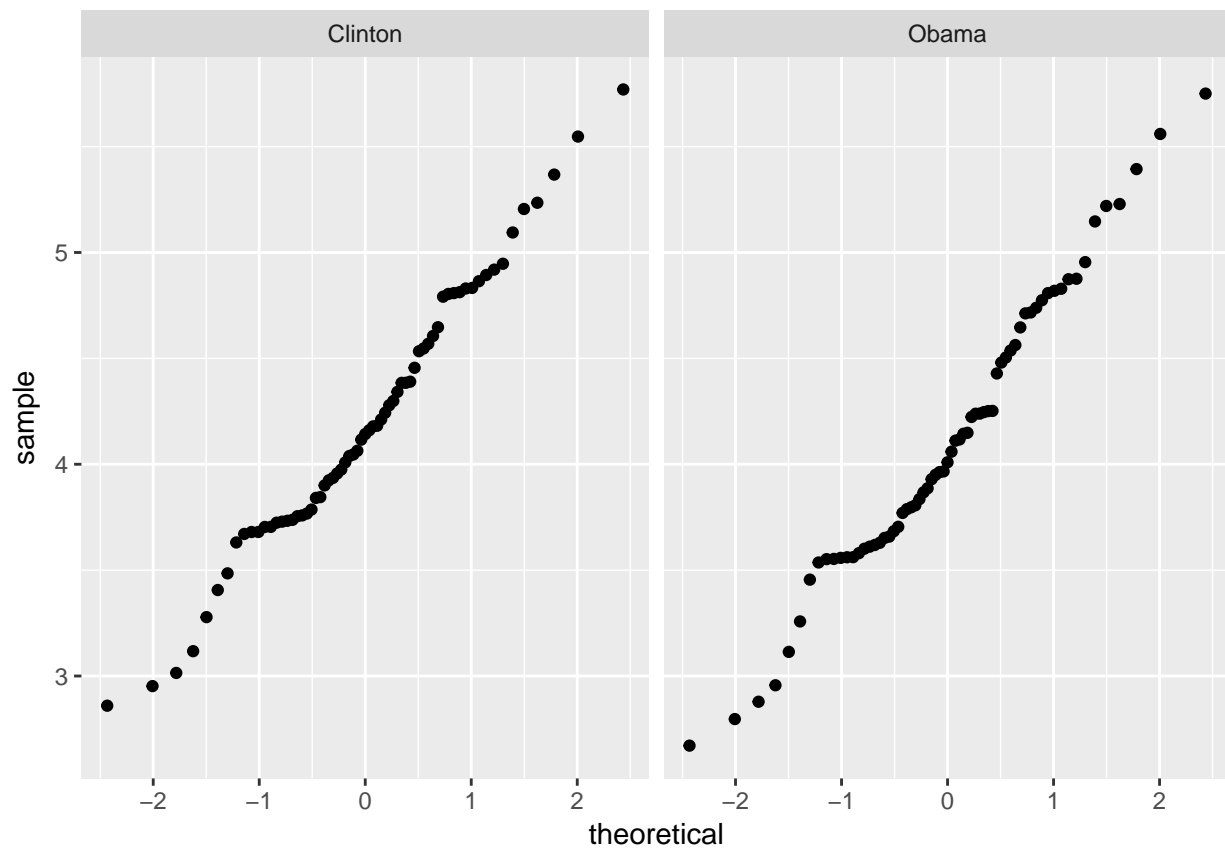
## 85	Clinton	Clinton	4553
## 86	Columbia	Clinton	8502
## 87	Crawford	Clinton	10215
## 88	Cumberland	Clinton	44282
## 89	Dauphin	Clinton	64287
## 90	Delaware	Clinton	169423
## 91	Elk	Clinton	3637
## 92	Erie	Clinton	54820
## 93	Fayette	Clinton	17826
## 94	Forest	Clinton	626
## 95	Franklin	Clinton	17322
## 96	Fulton	Clinton	904
## 97	Greene	Clinton	4157
## 98	Huntingdon	Clinton	4487
## 99	Indiana	Clinton	11468
## 100	Jefferson	Clinton	3437
## 101	Juniata	Clinton	1810
## 102	Lackawanna	Clinton	51593
## 103	Lancaster	Clinton	90066
## 104	Lawrence	Clinton	13933
## 105	Lebanon	Clinton	17860
## 106	Lehigh	Clinton	74777
## 107	Luzerne	Clinton	52092
## 108	Lycoming	Clinton	12926
## 109	McKean	Clinton	3802
## 110	Mercer	Clinton	17631
## 111	Mifflin	Clinton	3563
## 112	Monroe	Clinton	31930
## 113	Montgomery	Clinton	247724
## 114	Montour	Clinton	2851
## 115	Northampton	Clinton	65936
## 116	Northumberland	Clinton	9184
## 117	Perry	Clinton	4081
## 118	Philadelphia	Clinton	562853
## 119	Pike	Clinton	9247
## 120	Potter	Clinton	1300
## 121	Schuylkill	Clinton	16724
## 122	Snyder	Clinton	3991
## 123	Somerset	Clinton	7367
## 124	Sullivan	Clinton	756
## 125	Susquehanna	Clinton	5065
## 126	Tioga	Clinton	3642
## 127	Union	Clinton	6138
## 128	Venango	Clinton	5889
## 129	Warren	Clinton	4828
## 130	Washington	Clinton	34436
## 131	Wayne	Clinton	6398
## 132	Westmoreland	Clinton	59506
## 133	Wyoming	Clinton	3573
## 134	York	Clinton	67428

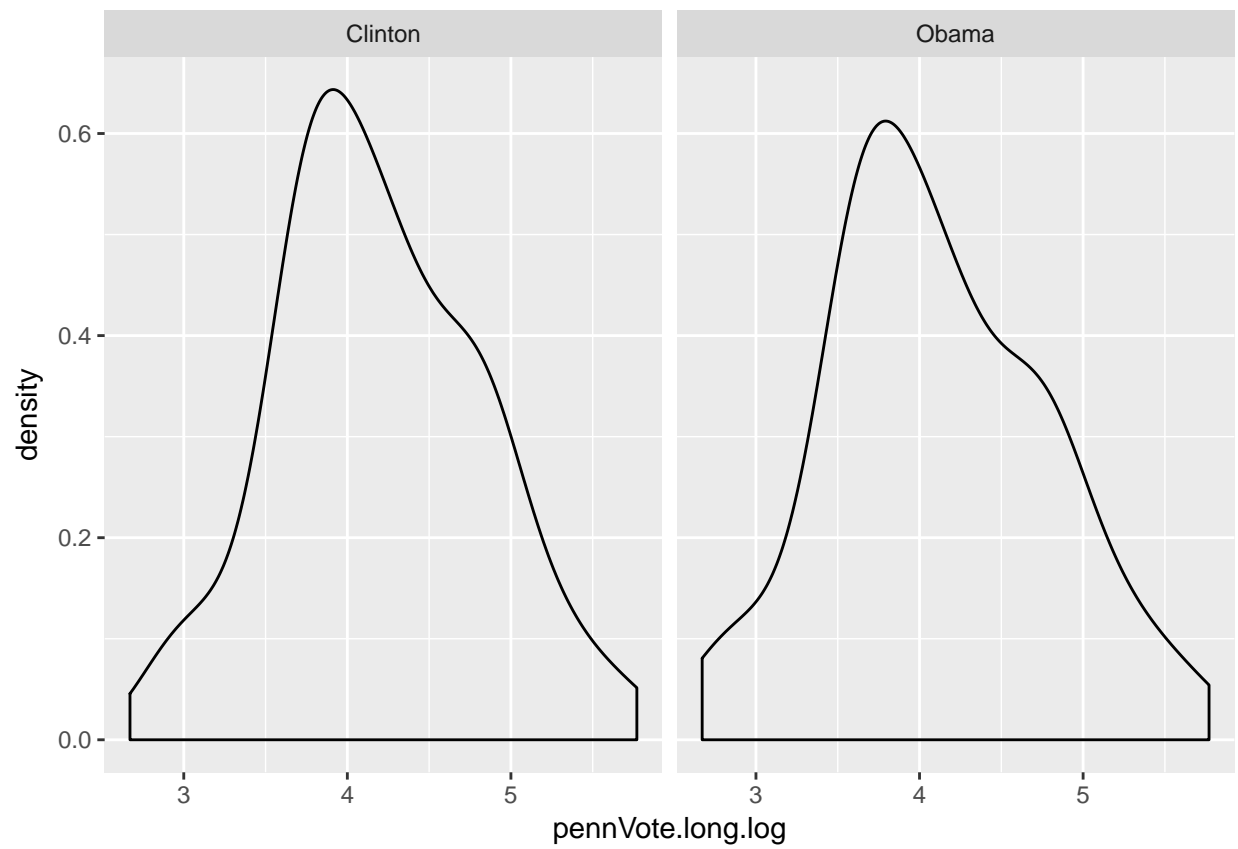
Answer 2

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 follow a normal distribution. This is because there is slight bump in the distribution around the -1 standard deviation value on the plot.

From the density plots of the two distributions we can confirm that the plots are skewed to the right.





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.)

Since the data is highly skewed, log transformation (with base 10) was performed on the data. The base 10 was chosen in this case as the maximum of votes per county per candidate is in the hundreds of thousands. The Tukey Mean difference was plotted for the original data as well as for the log transformed data. The first plot is of the original data while the second plot is for the log transformed data. The points on the plots are not spread randomly. If there were randomly distributed we would get some evidence of additive relationship in the first plot and multiplicative relationship in the second.

Since the points have a pattern associated with their positions. We have found no evidence of additive or multiplicative relationship. Hence we can safely say that something more complicated is being happening in this case.

