STA2201 Lab #2

Timothy Jordan Regis

2023-01-22

Q1

Highest Delay Stations by Line
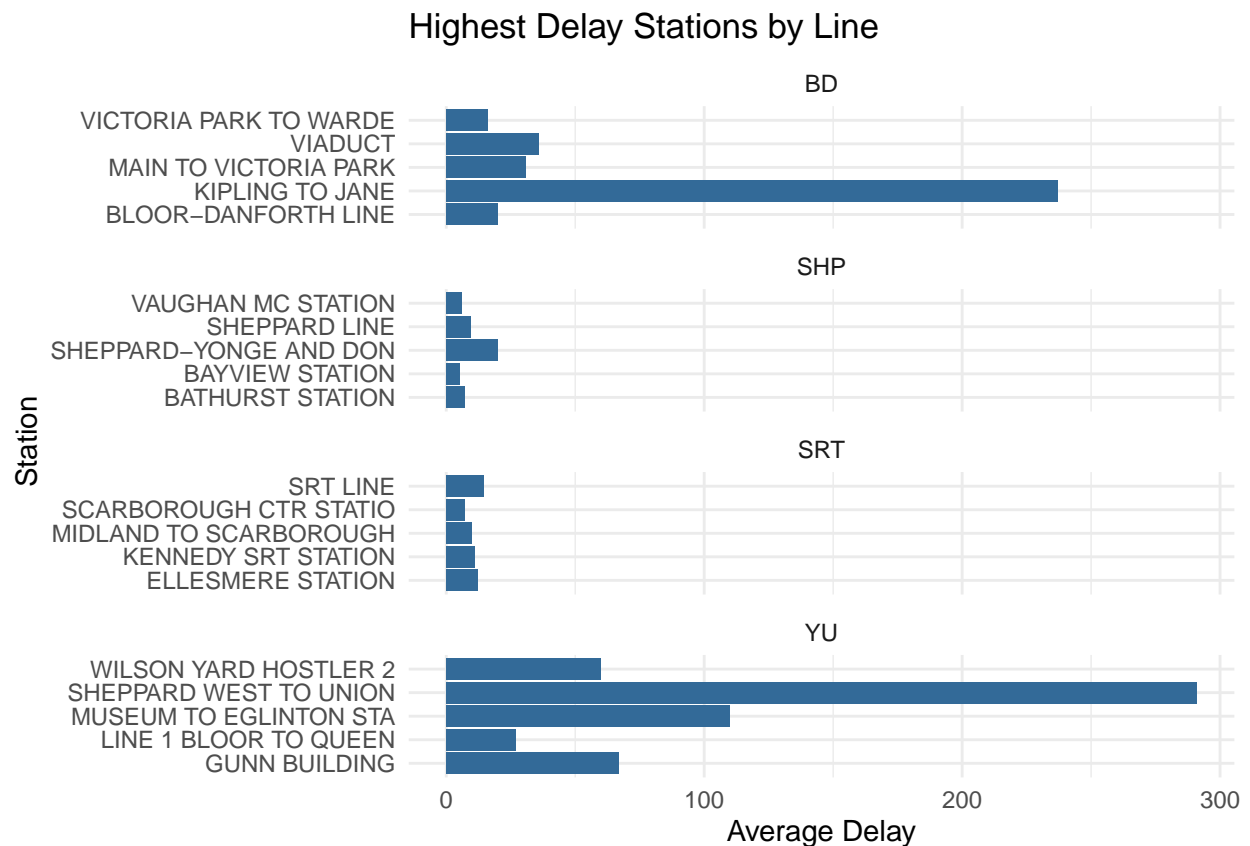

Q2 & Q3

```
## # A tibble: 6 x 13
##   contributors~1 contr~2 contr~3 contr~4 contr~5 goods~6 contr~7 relat~8 presi~9
##   <chr>          <chr>   <chr>   <chr>   <chr>   <chr>   <chr>   <chr>   <chr>
## 1 A D'Angelo, T~ <NA>    M6A 1P5 300     Moneta~ <NA>    Indivi~ <NA>    <NA>
## 2 A Strazar, Ma~ <NA>    M2M 3B8 300     Moneta~ <NA>    Indivi~ <NA>    <NA>
## 3 A'Court, K Su~ <NA>    M4M 2J8 36      Moneta~ <NA>    Indivi~ <NA>    <NA>
## 4 A'Court, K Su~ <NA>    M4M 2J8 100     Moneta~ <NA>    Indivi~ <NA>    <NA>
## 5 A'Court, K Su~ <NA>    M4M 2J8 100     Moneta~ <NA>    Indivi~ <NA>    <NA>
## 6 Aaron, Robert~ <NA>    M6B 1H7 250     Moneta~ <NA>    Indivi~ <NA>    <NA>
```

```
## # ... with 4 more variables: authorized_representative <chr>, candidate <chr>,
## #   office <chr>, ward <chr>, and abbreviated variable names
## #   1: contributors_name, 2: contributors_address, 3: contributors_postal_code,
## #   4: contribution_amount, 5: contribution_type_desc,
## #   6: goods_or_service_desc, 7: contributor_type_desc,
## #   8: relationship_to_candidate, 9: president_business_manager
```
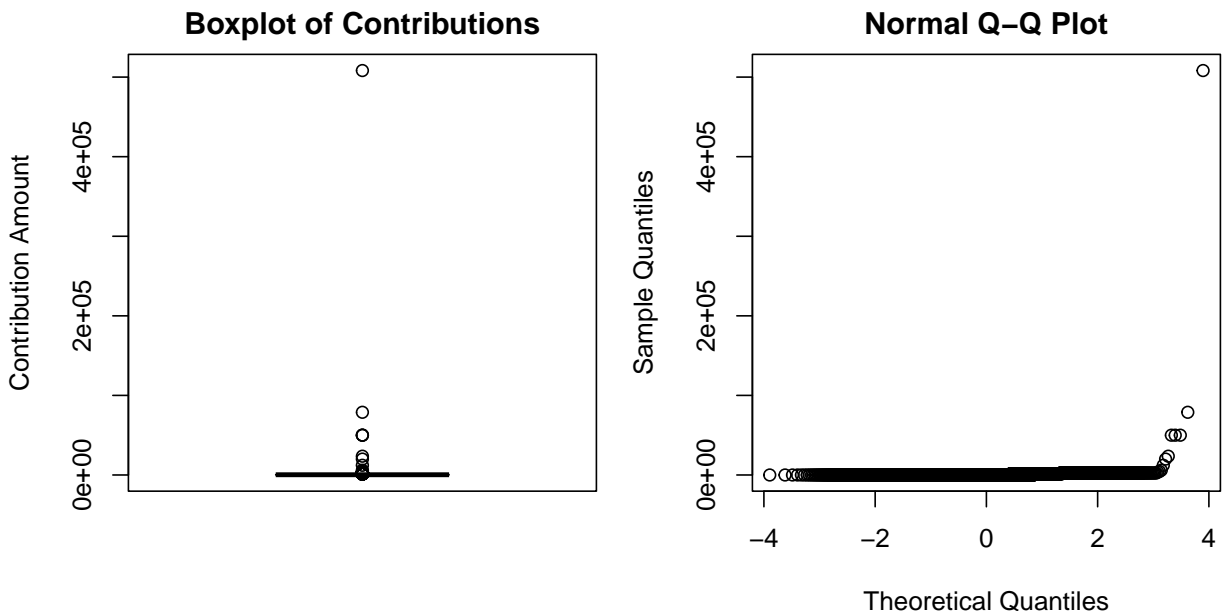
Q4

Summaries

```
##  contributors_name  contributors_address contributors_postal_code
##  Length:10199        Length:10199          Length:10199
##  Class :character    Class :character      Class :character
##  Mode  :character    Mode  :character      Mode  :character
##  contribution_amount contribution_type_desc goods_or_service_desc
##  Length:10199        Length:10199          Length:10199
##  Class :character    Class :character      Class :character
##  Mode  :character    Mode  :character      Mode  :character
##  contributor_type_desc relationship_to_candidate president_business_manager
##  Length:10199          Length:10199              Length:10199
##  Class :character      Class :character          Class :character
##  Mode  :character      Mode  :character          Mode  :character
##  authorized_representative  candidate          office
##  Length:10199              Length:10199      Length:10199
##  Class :character          Class :character  Class :character
##  Mode  :character          Mode  :character  Mode  :character
##      ward
##  Length:10199
##  Class :character
##  Mode  :character
```

NAs

```
##          contributors_name        contributors_address
##                          0                       10197
##   contributors_postal_code          contribution_amount
##                          0                           0
##     contribution_type_desc        goods_or_service_desc
##                          0                       10188
##     contributor_type_desc  relationship_to_candidate
##                          0                       10166
## president_business_manager  authorized_representative
##                      10197                       10197
##                  candidate                      office
##                          0                           0
##                       ward
##                      10199
```
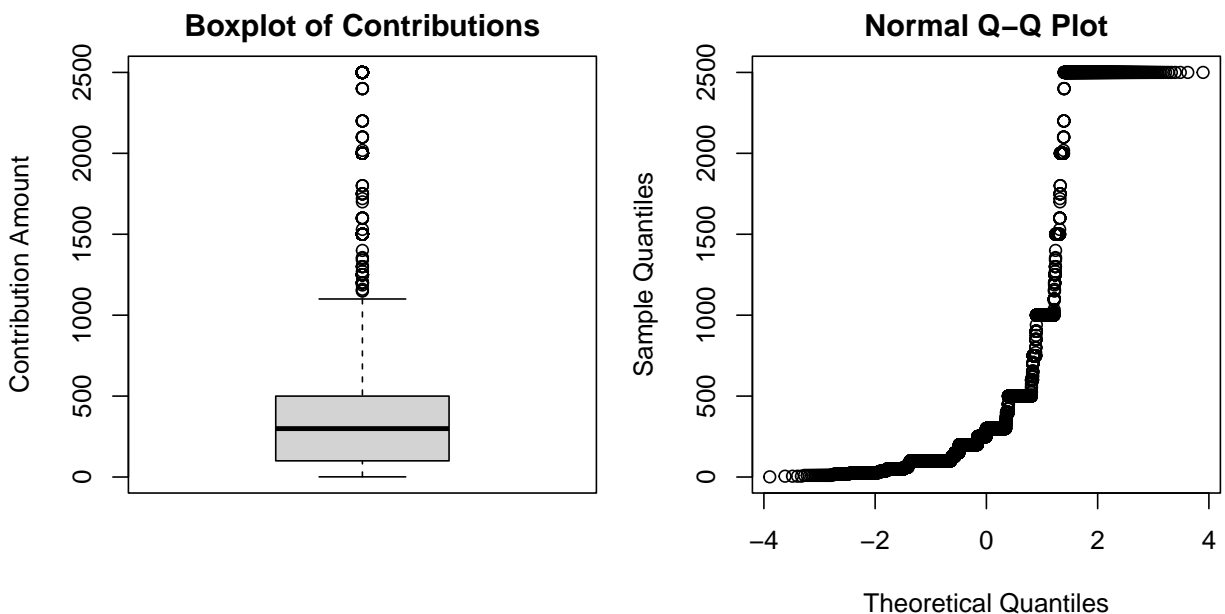
In this instance, we first find that the column tracking the amount of the contribution made is currently represented as a character value, which we will change to a numeric. Second, we notice that multiple columns have many NAs, and some are filled entirely with missing values, however, given the questions we will be looking at, this will not cause any major problems, thus we can move on with our inspection.

Q5

Boxplot of Contributions

Normal Q–Q Plot

We notice that a large majority of the contributions made in this dataset are below $3000. Examining the data more closely, we can see a relatively evenly spaced distribution of contributions up until $2500, which is where many contributors reach their maximums. However, after $2500, we can see that there are 11 more contributions that have been made. Among these, 10 of the contributions are made from the candidate to themselves, while the last remaing contribution has the lowest valuation and was in the form of a photography service. Thus, it clearly seems that this group of self-funding contributions are outliers in the data, and should be removed if we wish to inspect the distributions any further.

Boxplot of Contributions

Normal Q–Q Plot

Finally, we can see a much clearer picture of the distribution of contribution amounts after removing the previous outliers. As we mentioned, a majority of these contributions are below $2500, and the flat section on the qq-plot displays the significant number of contributors that contribute exactly $2500.

Q6

candidate	total_amount
Tory, John	2767868.7
Chow, Olivia	1638265.9
Ford, Doug	889897.3
Ford, Rob	387648.2
Stintz, Karen	242805.0

candidate	average_amount
Sniedzins, Erwin	2025.000
Syed, Hïmy	2018.000
Ritch, Carlie	1886.667
Ford, Doug	1456.460
Clarke, Kevin	1200.000

candidate	contributions_received
Chow, Olivia	5708
Tory, John	2602
Ford, Doug	611
Ford, Rob	538
Soknacki, David	314

Q7

candidate	total_amount
Tory, John	2765368.7
Chow, Olivia	1634765.9
Ford, Doug	331172.6
Stintz, Karen	242805.0
Ford, Rob	174509.5

candidate	average_amount
Ritch, Carlie	1886.667
Sniedzins, Erwin	1866.667
Tory, John	1063.194
Gardner, Norman	1000.000
Tiwari, Ramnarine	1000.000

candidate	contributions_received
Chow, Olivia	5706
Tory, John	2601
Ford, Doug	608
Ford, Rob	531
Soknacki, David	314

Q8

For all contributions, the total number of multi-contributors is:

[1] 184

OR, without self-funding

[1] 181