Assignment 9gw : Exercise 16: Clustering

Ragunath Gunasekaran

2020-11-01

Footnote

This is a Footnote test.

Citations

- R for Everyone
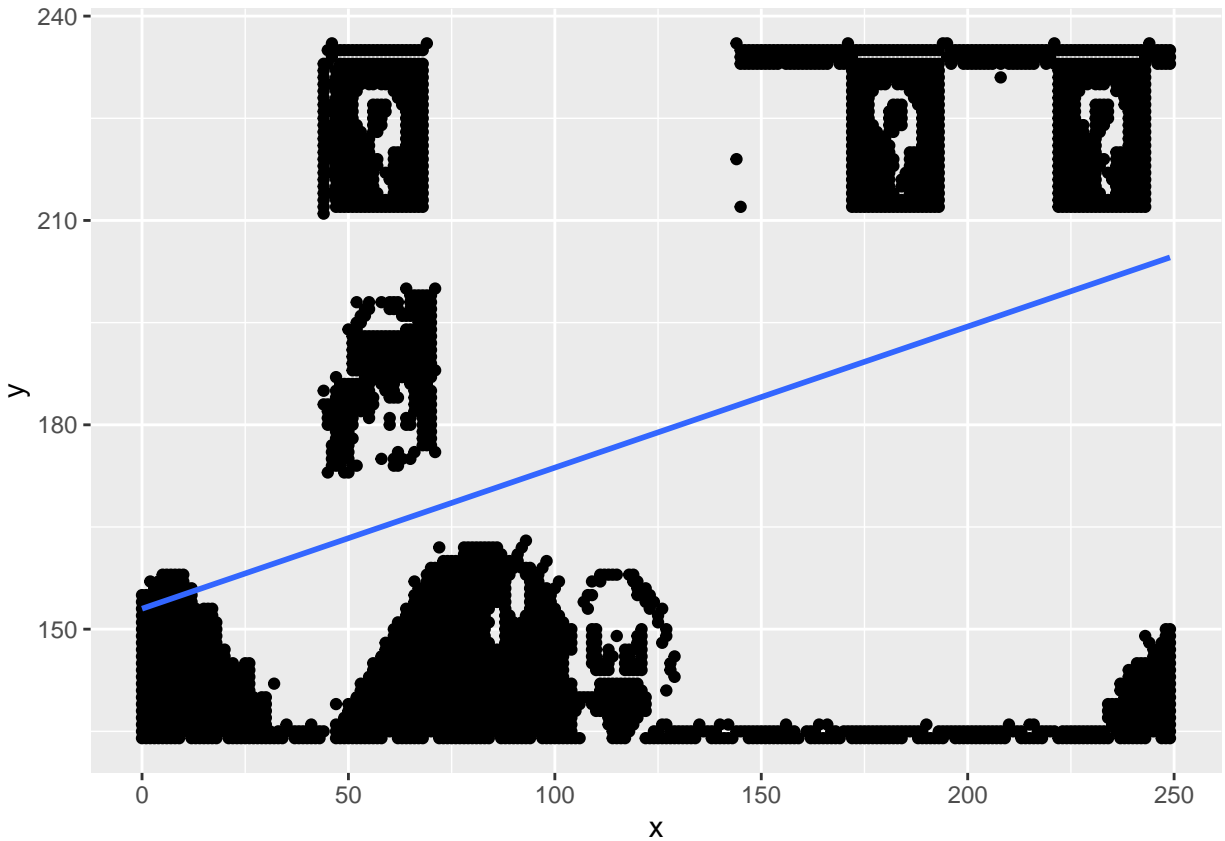- Discovering Statistics Using R

Data Load - Data from CSV file load into Dataframe

Structure Of An Arbitrary R Object

```
## 'data.frame':    4022 obs. of  2 variables:
##  $ x: int  46 69 144 171 194 195 221 244 45 47 ...
##  $ y: int  236 236 236 236 236 236 236 236 235 235 ...
```

Scatter plot of clustering dataset by using ggplot2

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

Fiting the dataset using the k-means algorithm from k=2 to k=12. Create a scatter plot of the resultant clusters for each value of k.

Cluster Data for each K Value

#Ref - Source 5,4,6

```
## Warning: package 'cluster' was built under R version 4.0.2
```

```
##          x          y
## 1  62.2815 162.4090
## 2 207.8005 203.2898
```

```
##           x          y
## 1 209.16575 205.1791
## 2  87.29886 144.3702
## 3  38.73348 180.4610
```

```
##           x          y
## 1  57.42711 207.3043
## 2  88.12209 144.5305
## 3 209.19733 205.2351
## 4  14.47973 142.9662
```

3

```
##           x        y
## 1 221.61981 137.6677
## 2  14.10580 143.0324
## 3  86.55457 144.6858
## 4  57.42711 207.3043
## 5 203.98579 224.8406
```

```
##           x        y
## 1 231.08994 224.3683
## 2 221.61981 137.6677
## 3  14.10580 143.0324
## 4  57.42711 207.3043
## 5 179.55019 225.2664
## 6  86.55457 144.6858
```

```
##           x        y
## 1 179.55019 225.2664
## 2 112.54167 142.3063
## 3  13.07206 143.2460
## 4 227.08362 137.9477
## 5 231.08994 224.3683
## 6  74.28817 145.8355
## 7  57.31518 207.7575
```

```
##           x        y
## 1 231.08994 224.3683
## 2 179.55019 225.2664
## 3  13.07206 143.2460
## 4  58.82698 186.9824
## 5  56.34467 223.0181
## 6  74.49892 145.4643
## 7 227.08362 137.9477
## 8 112.74316 142.3032
```

```
##           x        y
## 1  12.91166 143.2915
## 2 231.08994 224.3683
## 3  58.82698 186.9824
## 4  71.78875 145.4550
## 5 103.68796 143.7500
## 6  56.34467 223.0181
## 7 179.55019 225.2664
## 8 237.85841 138.8628
## 9 171.17391 134.5739
```

```
##           x        y
## 1  65.81143 143.4457
## 2 179.55019 225.2664
## 3 179.35514 134.5514
## 4 231.08994 224.3683
## 5  89.41301 146.7431
## 6  12.61071 143.3839
## 7 118.62931 141.4440
```

```
## 8   239.28704 139.0648
## 9    56.34467 223.0181
## 10   58.82698 186.9824


##             x        y
## 1    89.27110 146.7792
## 2    12.61071 143.3839
## 3   178.21495 134.5514
## 4    56.34467 223.0181
## 5    65.72169 143.4299
## 6   118.17167 141.4549
## 7   239.00917 139.0229
## 8   231.35714 224.2619
## 9   157.85714 233.6310
## 10  184.00683 223.7677
## 11   58.82698 186.9824


##             x        y
## 1   179.93333 134.5524
## 2    94.84158 143.7327
## 3   166.78378 228.9946
## 4    77.33189 149.5922
## 5    58.82698 186.9824
## 6    60.12052 139.9218
## 7    56.34467 223.0181
## 8   120.57895 141.6077
## 9   187.30814 223.5378
## 10   12.29295 143.4955
## 11  231.66009 224.1382
## 12  239.28704 139.0648
```
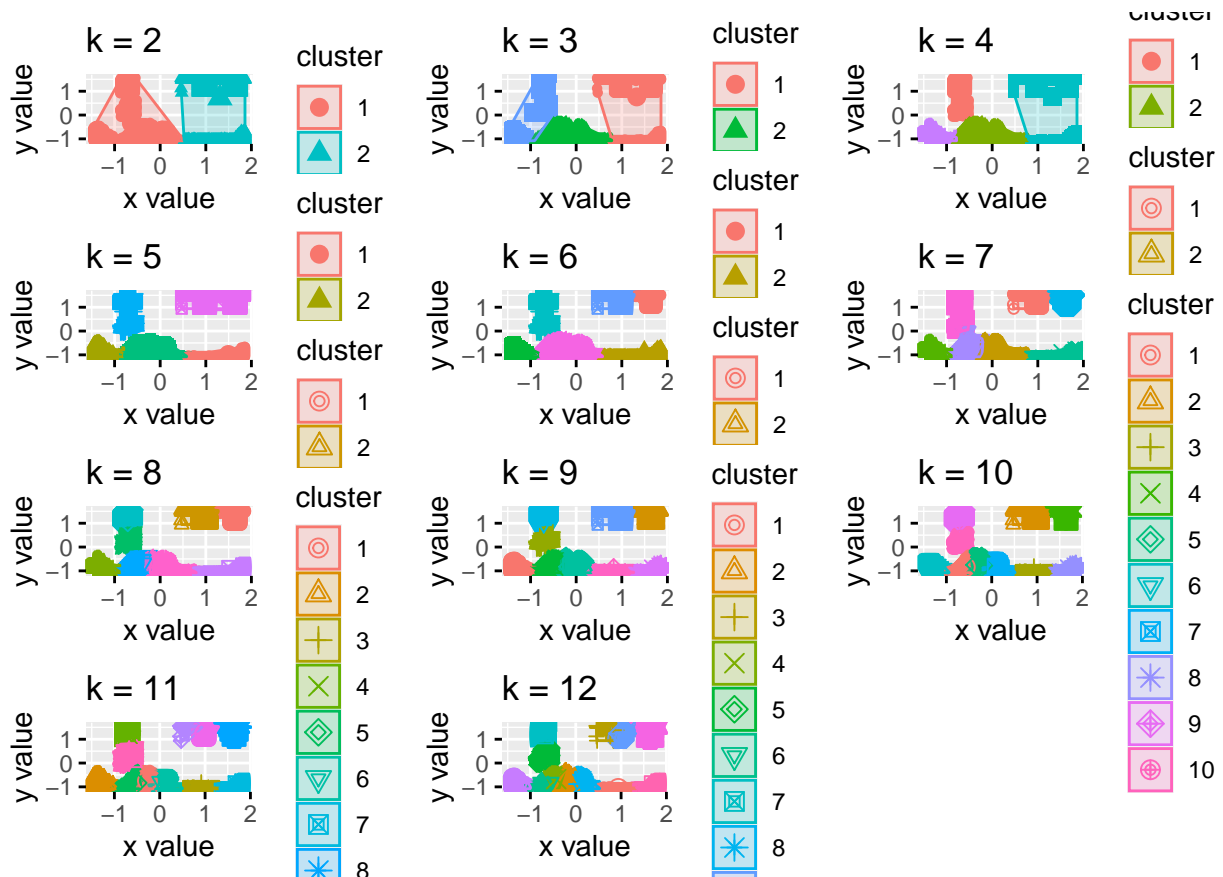
#install.package("factoextra")

Plots Comparision

#Ref Source -6,3,7

```
## Warning: package 'factoextra' was built under R version 4.0.2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

## Warning: package 'gridExtra' was built under R version 4.0.2
```
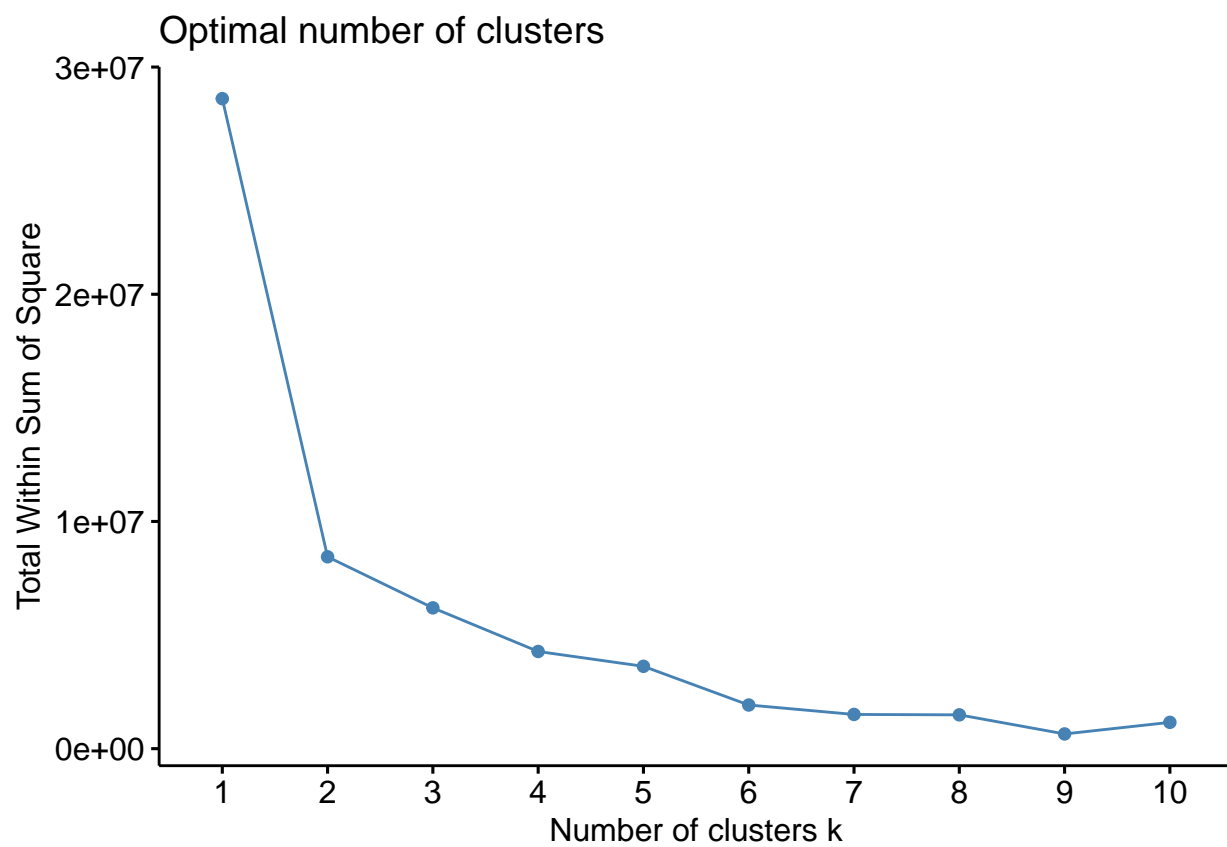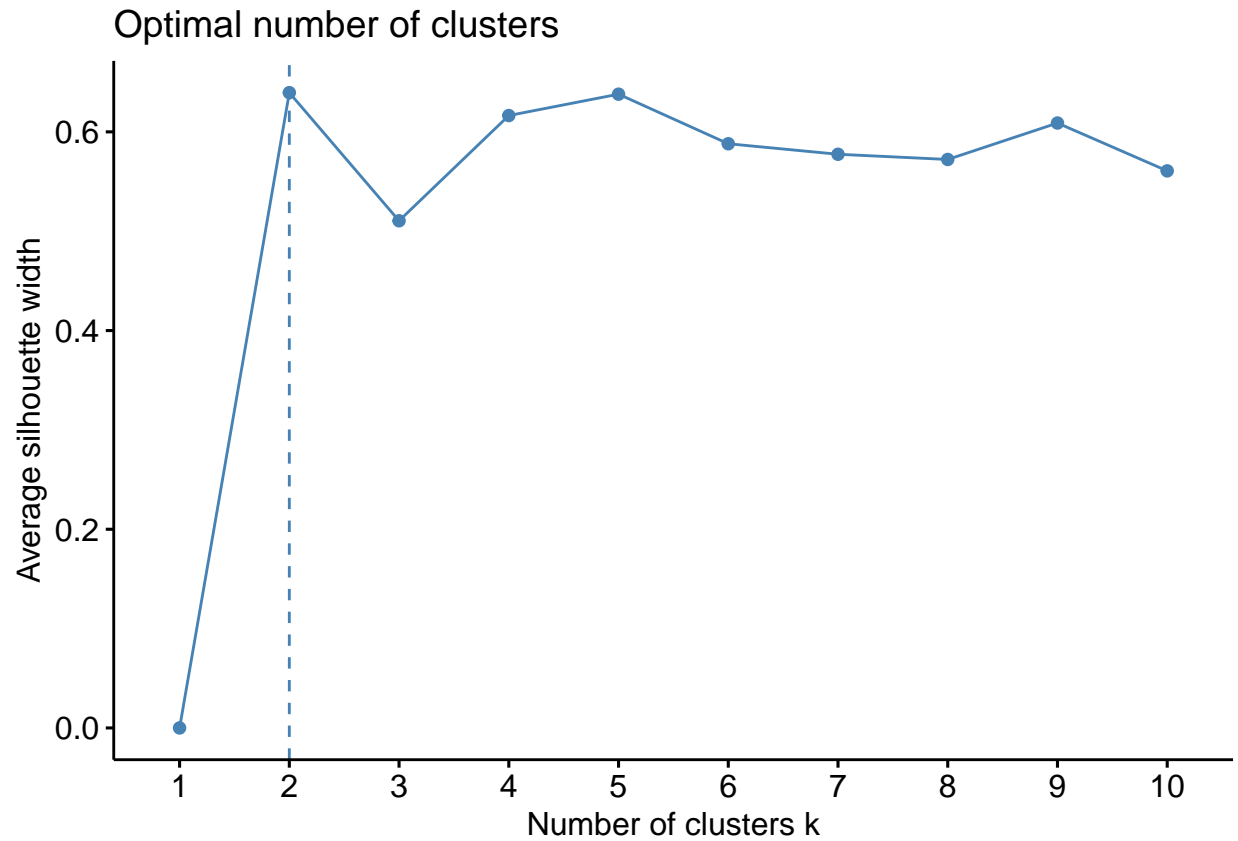
k = 2 · **k = 3** · **k = 4** · **k = 5** · **k = 6** · **k = 7** · **k = 8** · **k = 9** · **k = 10** · **k = 11** · **k = 12**

Calculate this average distance from the center of each cluster for each value of k and plot it as a line chart where k is the x-axis and the average distance is the y-axis.

#Ref Source -6,3,7

Optimal number of clusters

Optimal number of clusters

References

1. Sejal Jaiswal. (2018). K-Means Clustering in R Tutorial
2. https://rpubs.com/dnchari/ggplot2
3. https://towardsdatascience.com/clustering-with-k-means-1e07a8bfb7ca
4. https://rpubs.com/dnchari/kmeans
5. https://www.youtube.com/watch?v=3GorGZgTTEk
6. https://rpubs.com/abdul_yunus/Kmeans_Clustering
7. https://cran.r-project.org/web/packages/factoextra/factoextra.pdf