

JenksBreaks

Avishek

2023-06-01

Introduction

The Jenks Natural Breaks Classification (or Optimization) system is a data classification method designed to optimize the arrangement of a set of values into “natural” classes.

- A Natural class is the most optimal class range found “naturally” in a data set.
- Natural breaks are determined with a frequency histogram. Class boundaries are identified as troughs in the data.
- Many dataset will not have obvious natural breaks which means that this method would tend to show breaks where none really exists.

Jenks Natural Breaks Classification is done by seeking to minimize each class’s average deviation from the class mean, while maximizing each class’s deviation from the means of the other classes. In other words, the method seeks to reduce the variance within classes and maximize the variance between classes.

Algorithm Steps

Step 1: The user selects the attribute, x, to be classified and specifies the number of classes required, k

Step 2: A set of k-1 random or uniform values are generated in the range $[\min\{x\}, \max\{x\}]$. These are used as initial class boundaries

Step 3: The mean values for each initial class are computed and the sum of squared deviations of class members from the mean values is computed. The total sum of squared deviations (TSSD) is recorded

Step 4: Individual values in each class are then systematically assigned to adjacent classes by adjusting the class boundaries to see if the TSSD can be reduced. This is an iterative process, which ends when improvement in TSSD falls below a threshold level, i.e. when the within class variance is as small as possible and between class variance is as large as possible. True optimization is not assured. The entire process can be optionally repeated from Step 1 or 2 and TSSD values compared

```
##
## 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
```

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

```
## Package 'clustertend' is deprecated. Use package 'hopkins' instead.
```

```
## Loading required package: MASS
```

```
##
```

```
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##     select
```

```
## Loading required package: ape
```

```
##
```

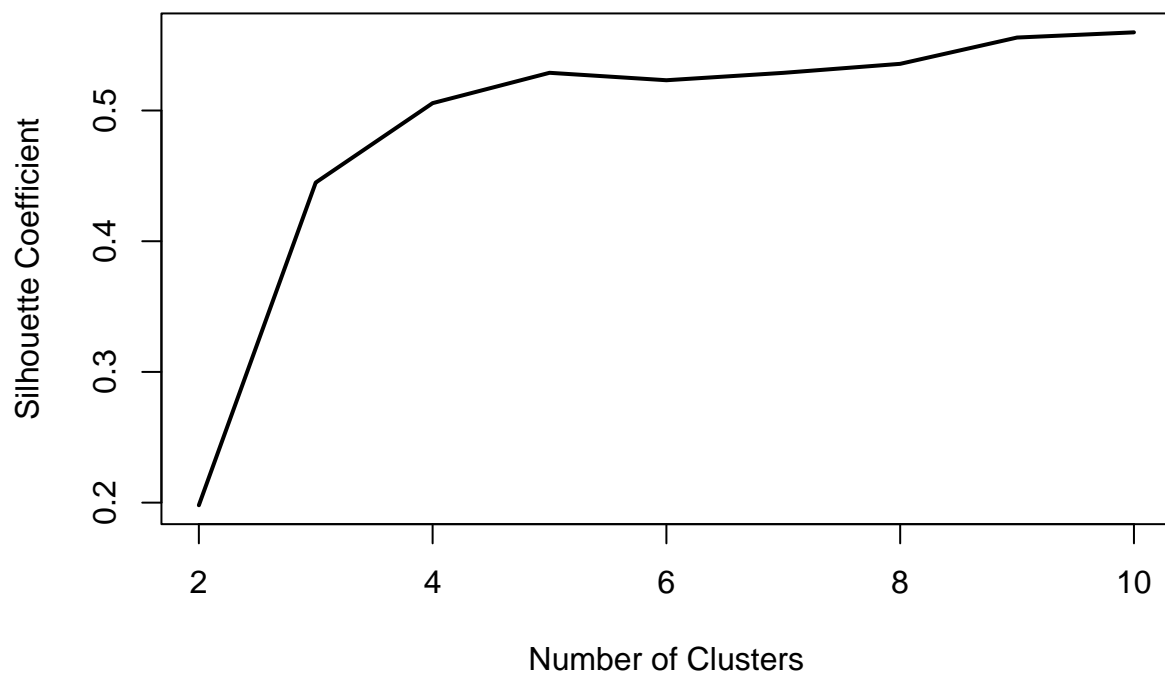
```
## Attaching package: 'ape'
```

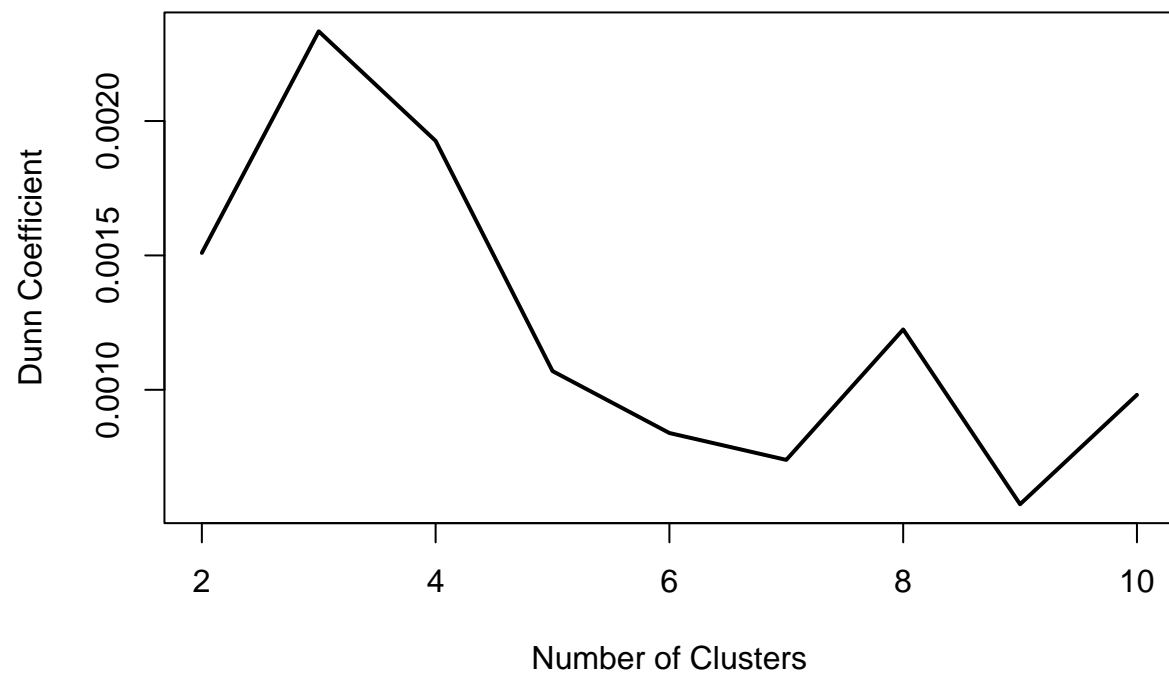
```
## The following object is masked from 'package:dplyr':
```

```
##
```

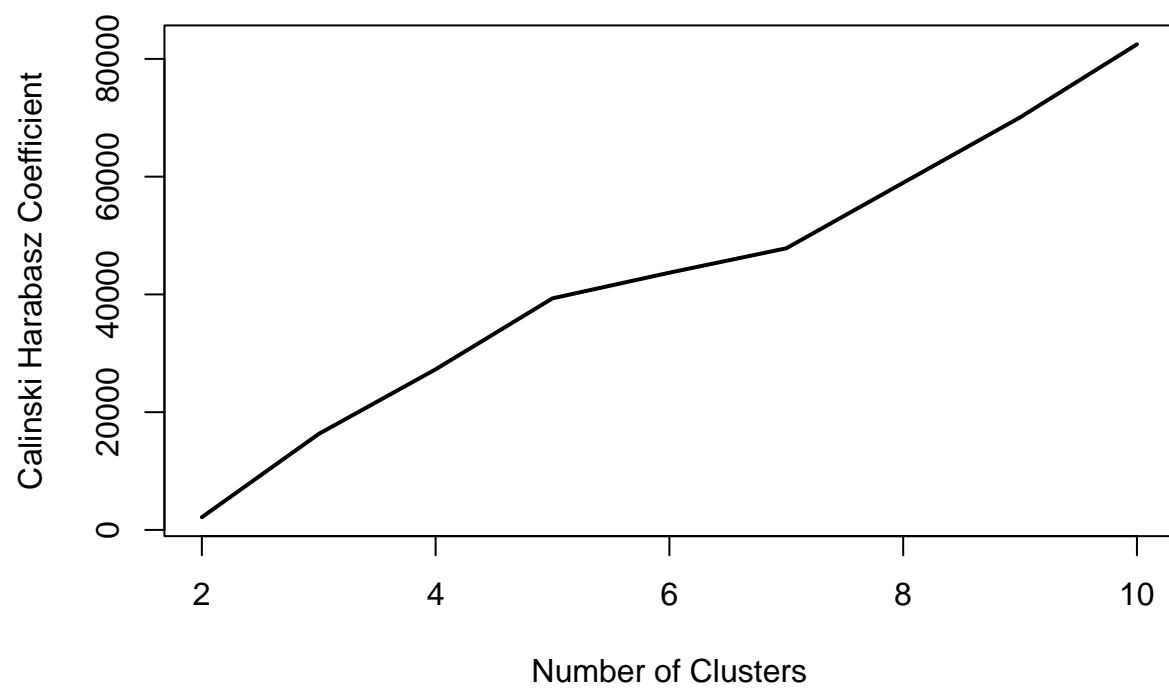
```
##     where
```

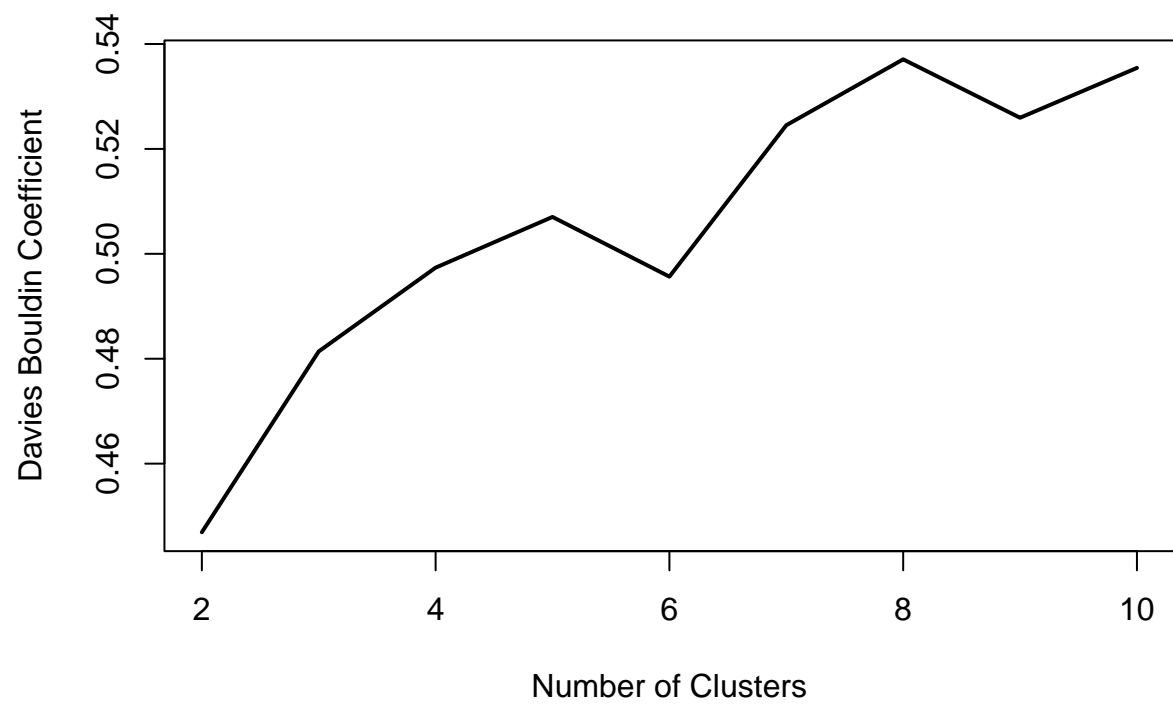
Employment



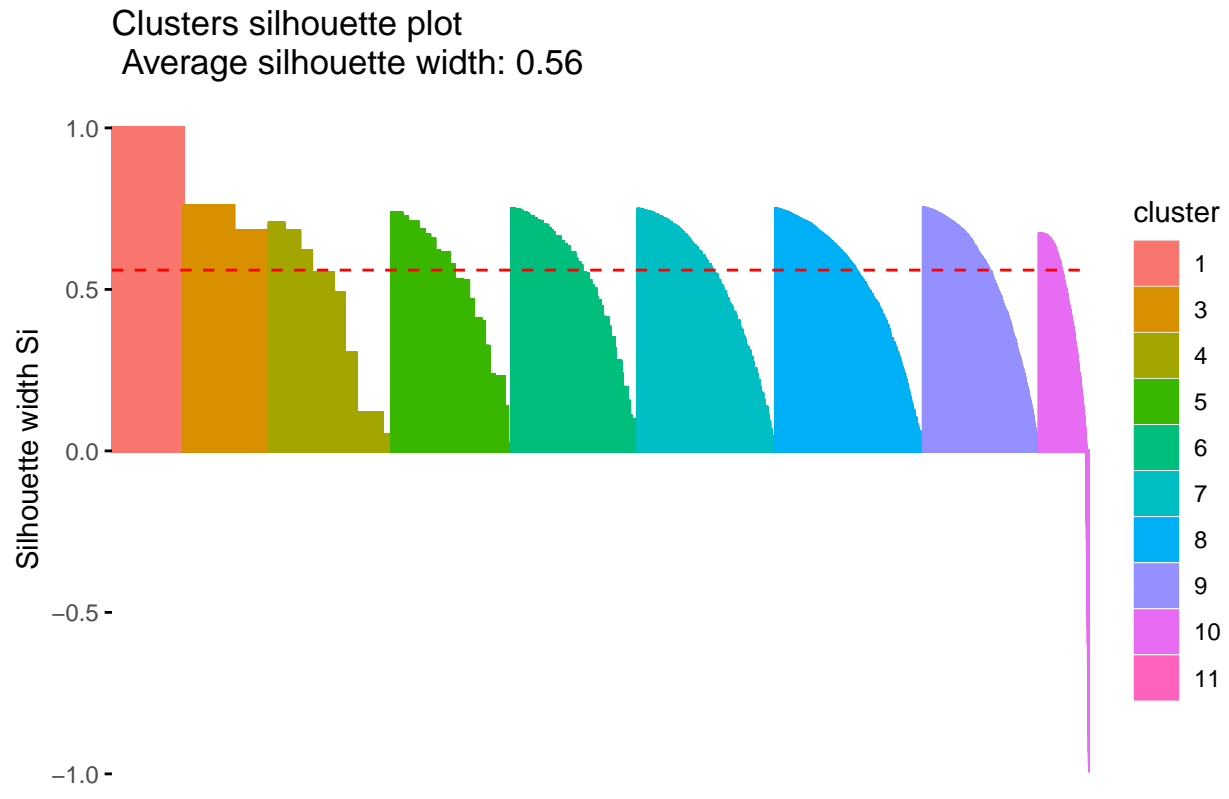






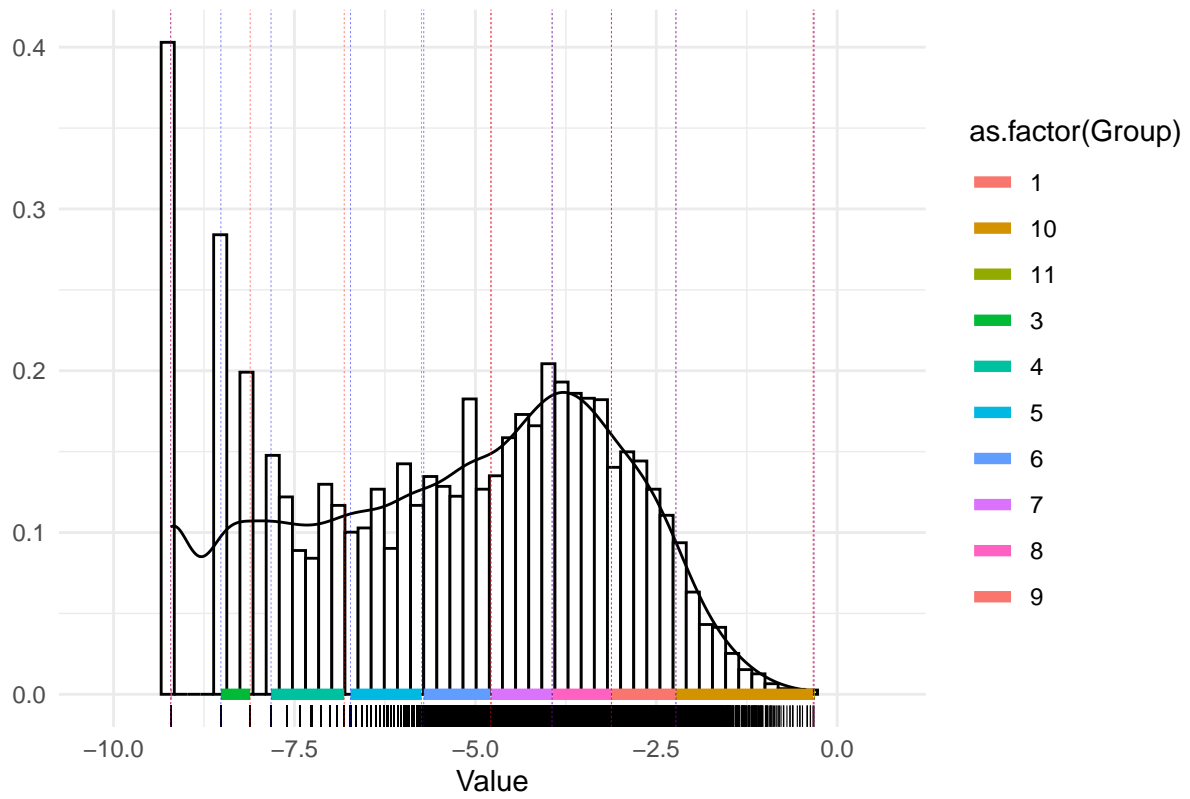


##	cluster	size	ave.sil.width
## 1	1	925	1.00
## 3	3	1109	0.73
## 4	4	1582	0.43
## 5	5	1559	0.52
## 6	6	1633	0.53
## 7	7	1780	0.52
## 8	8	1913	0.52
## 9	9	1500	0.53
## 10	10	651	0.42
## 11	11	1	0.00

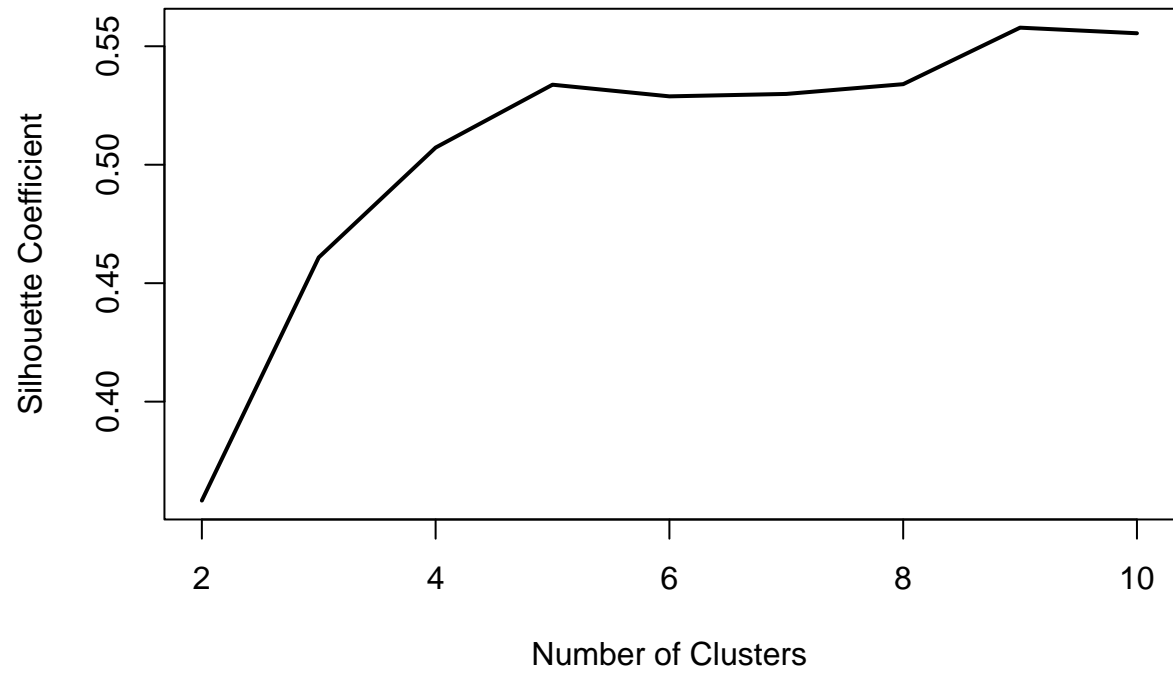


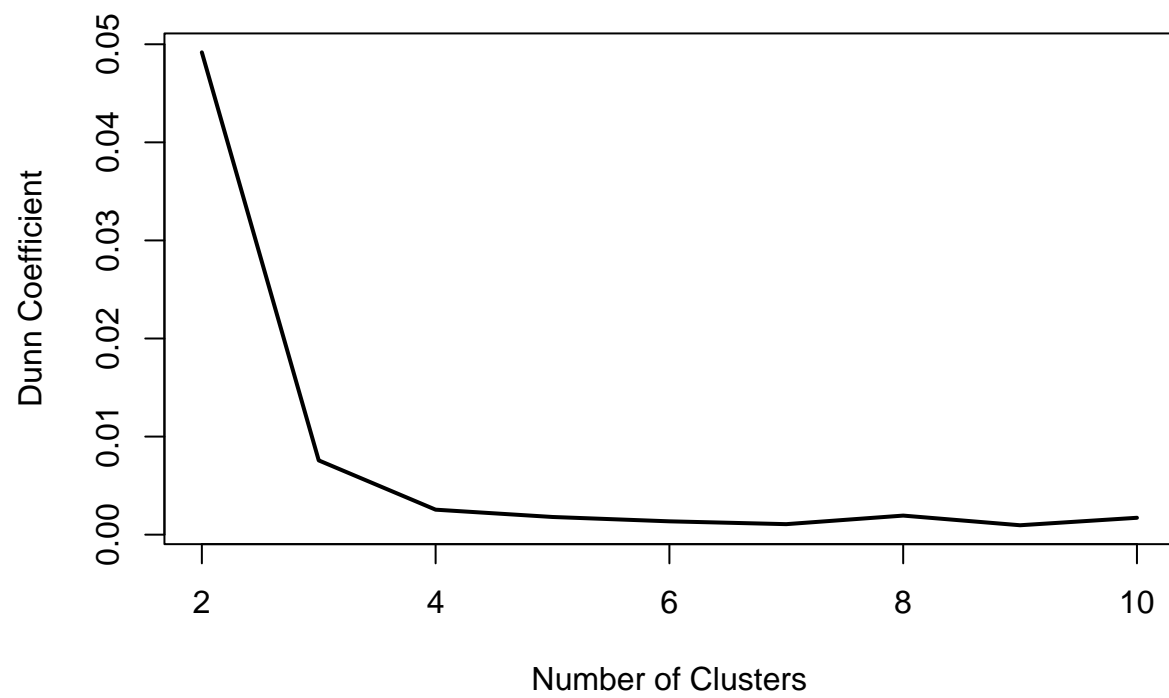
In the cutoff line figures below Blue vertical line indicates a cluster starting point and red vertical line indicates cluster's ending point. If you can't see the blue line before the red line that means either the cluster contains only one point or the starting and ending point of that cluster is so close.

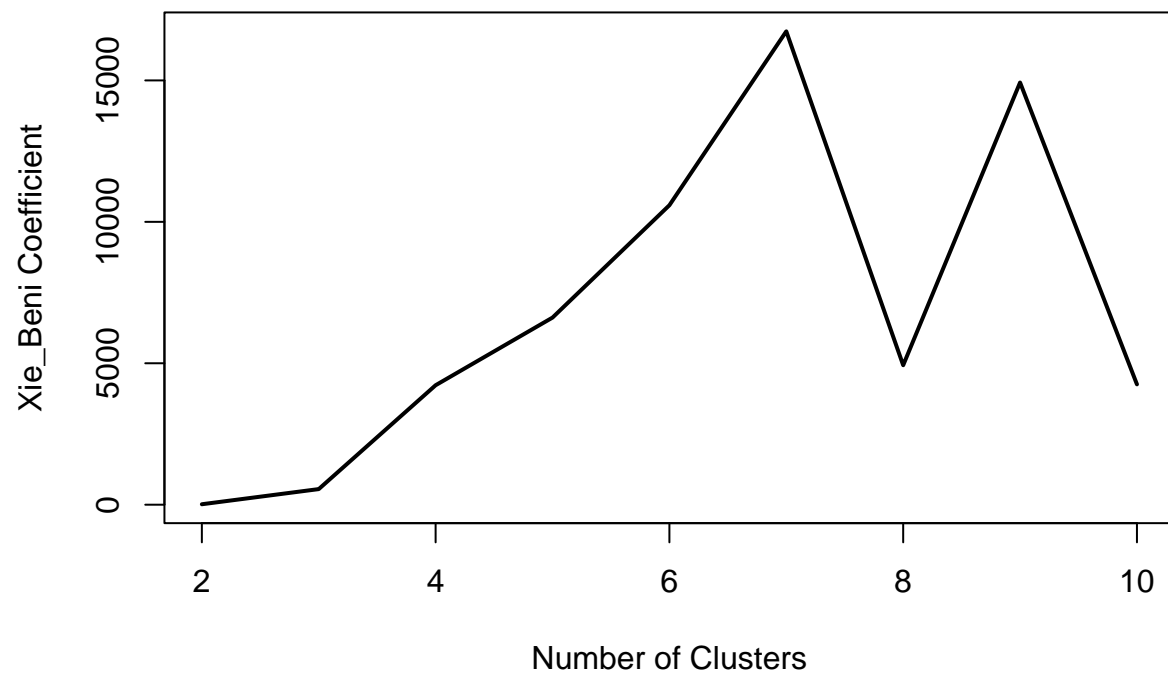
Cluster Rugs with Histogram & Cutoff Line

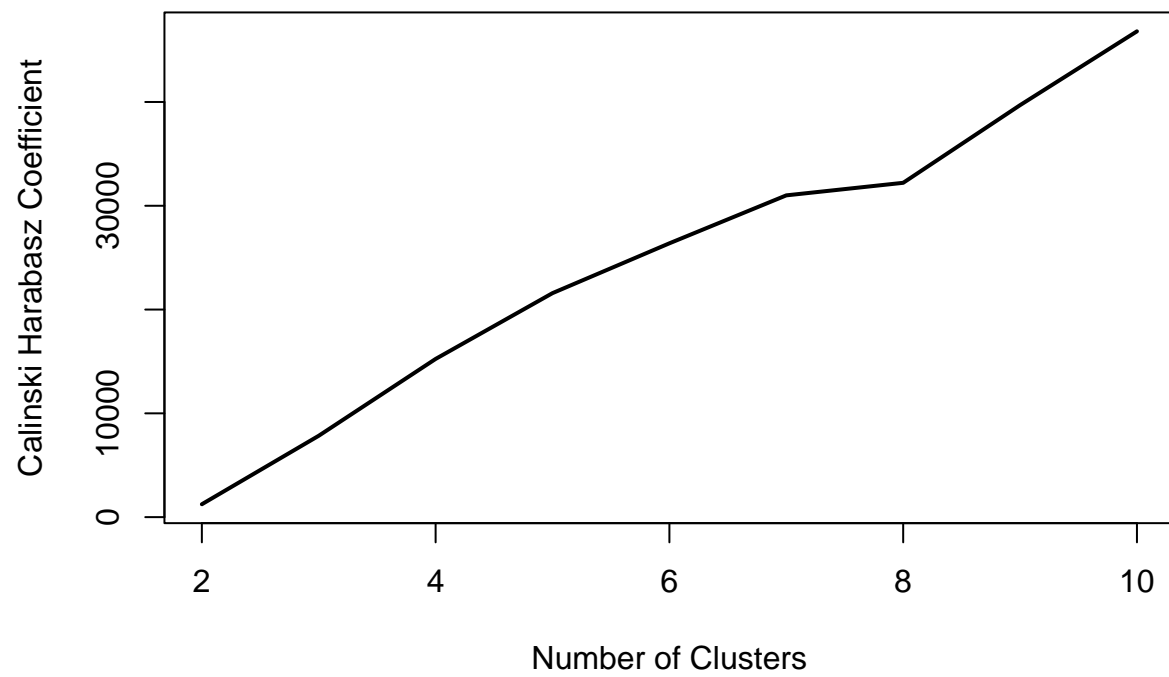


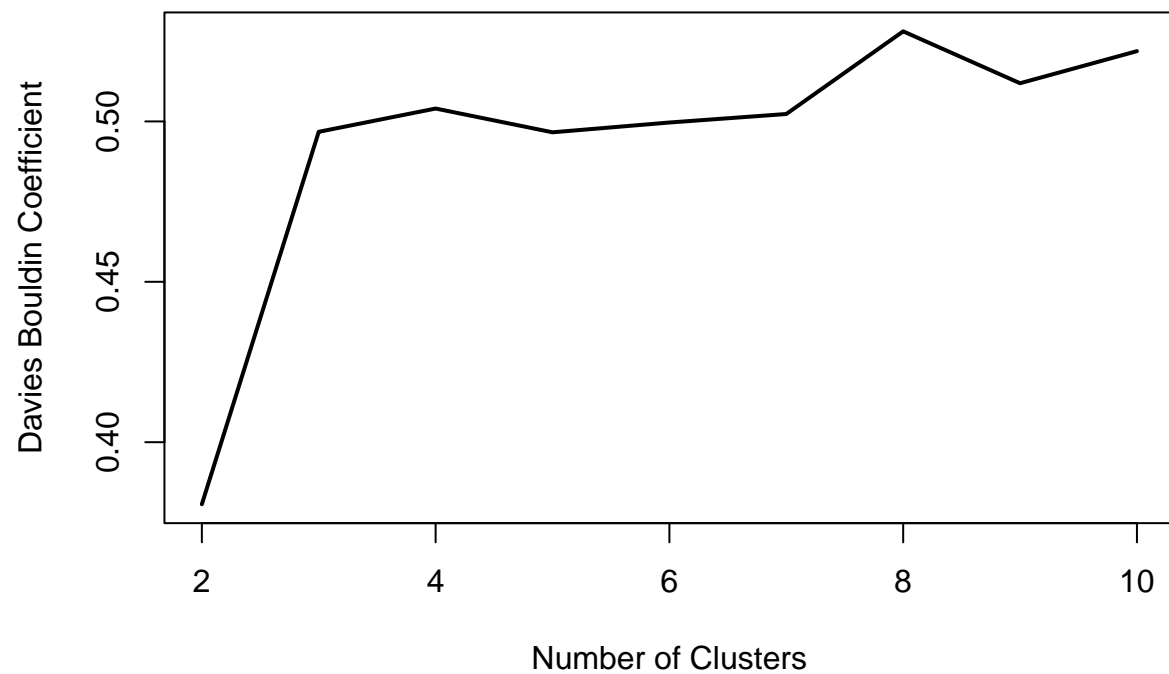
Health



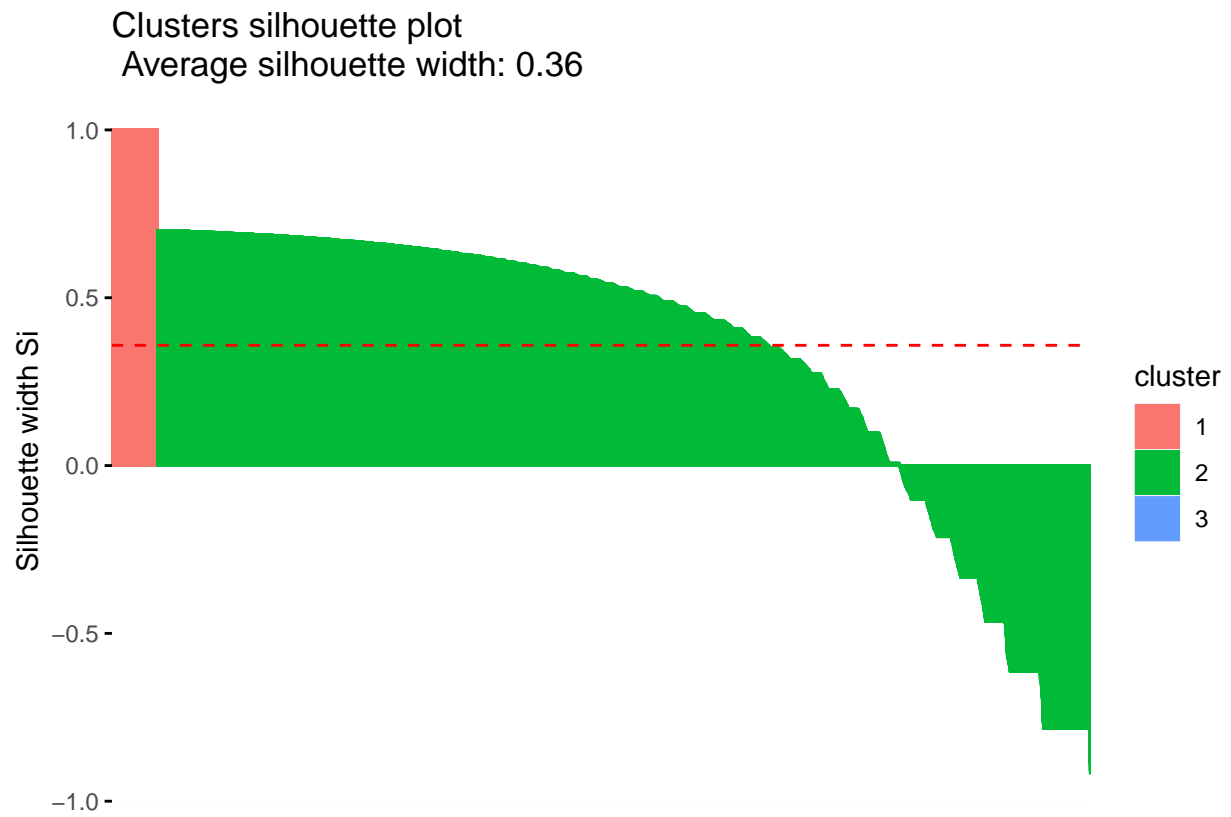




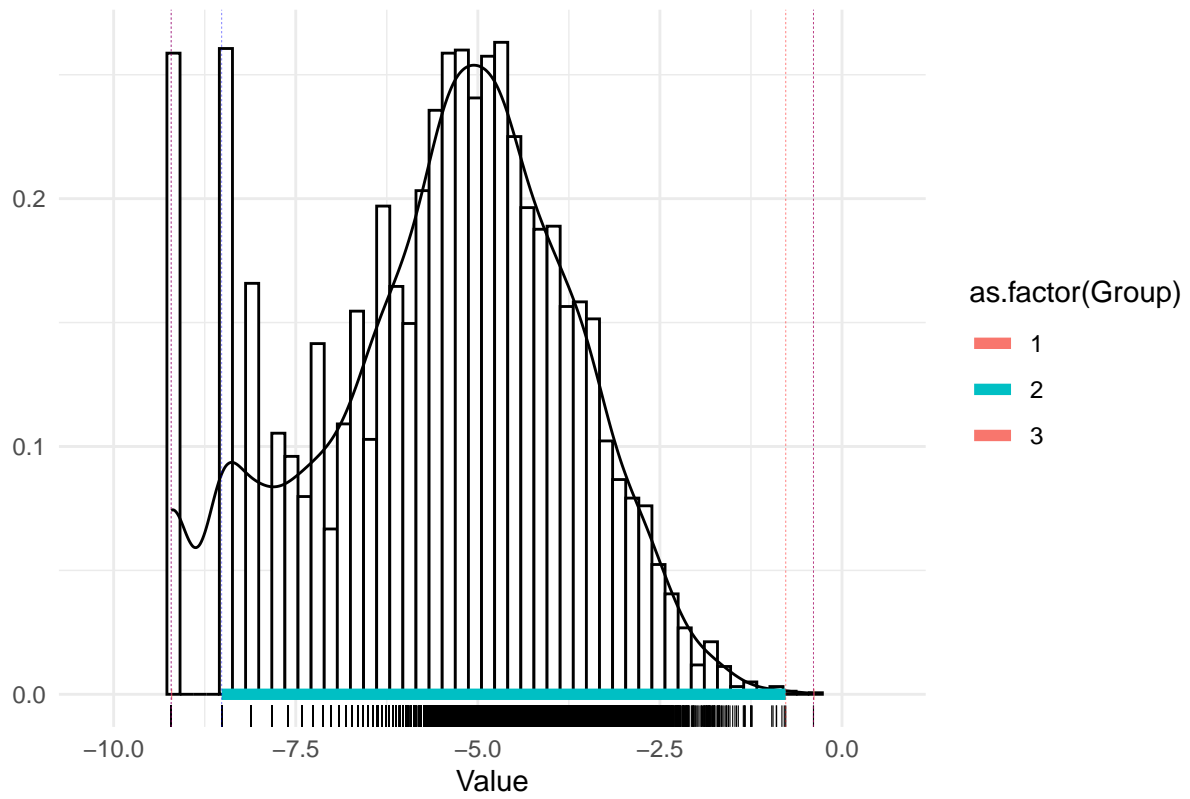




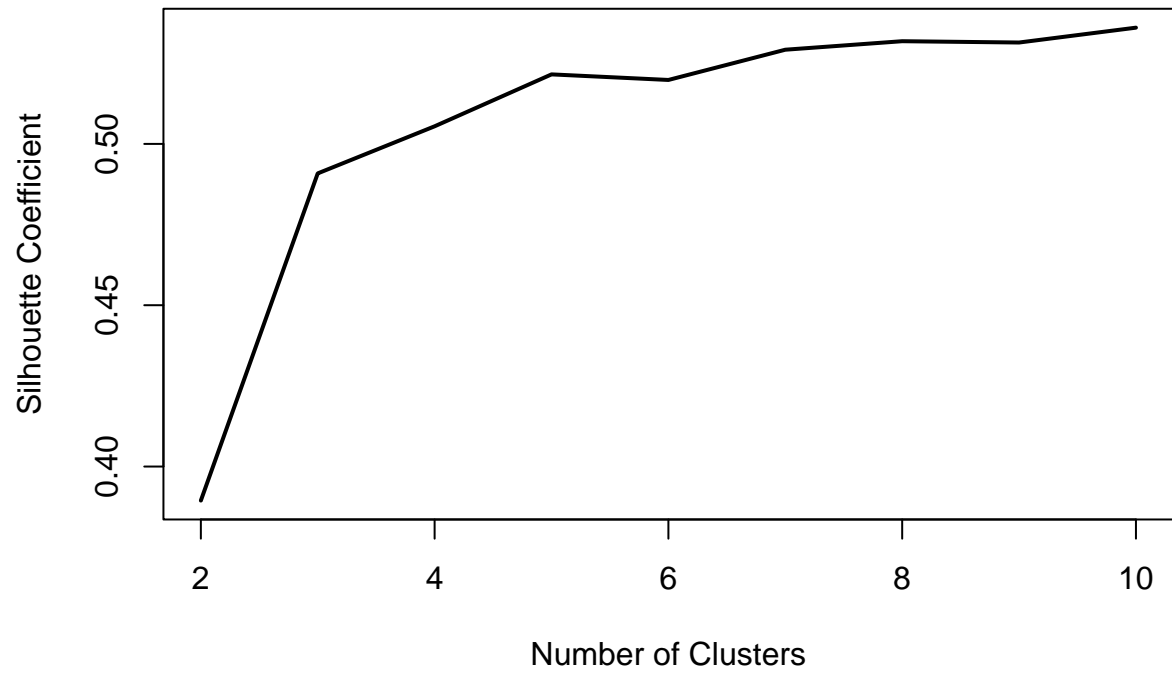
```
## cluster size ave.sil.width
## 1      1  415          1.00
## 2      2 8497          0.33
## 3      3   1          0.00
```

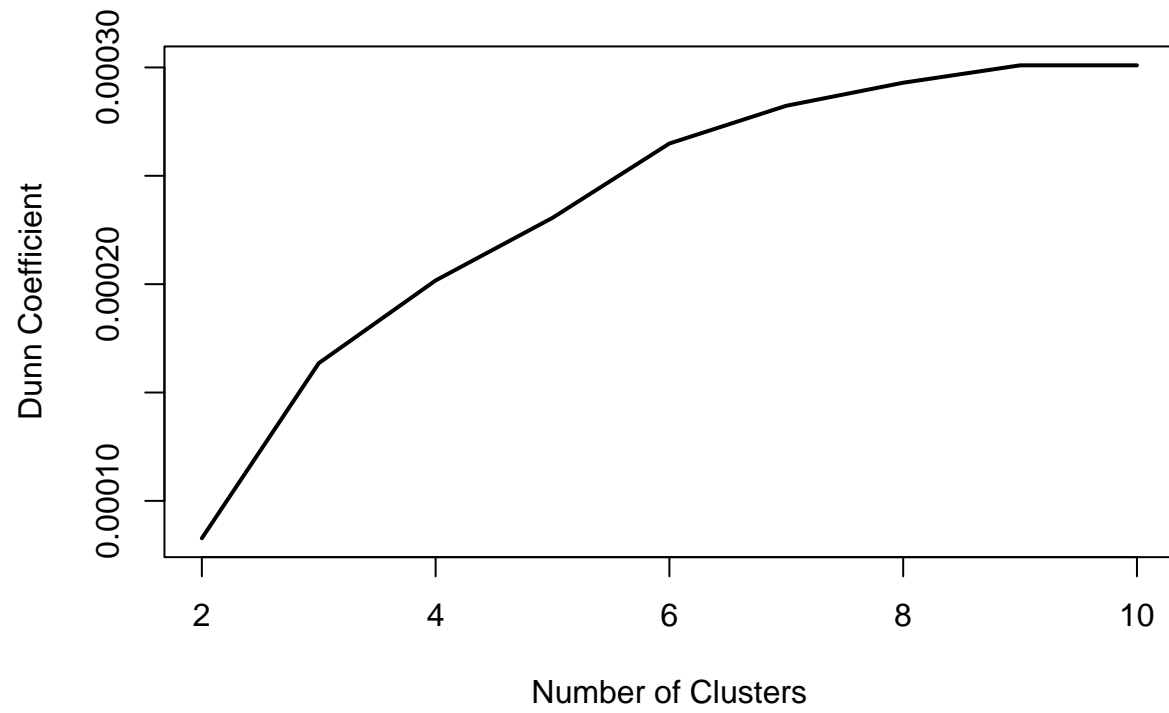


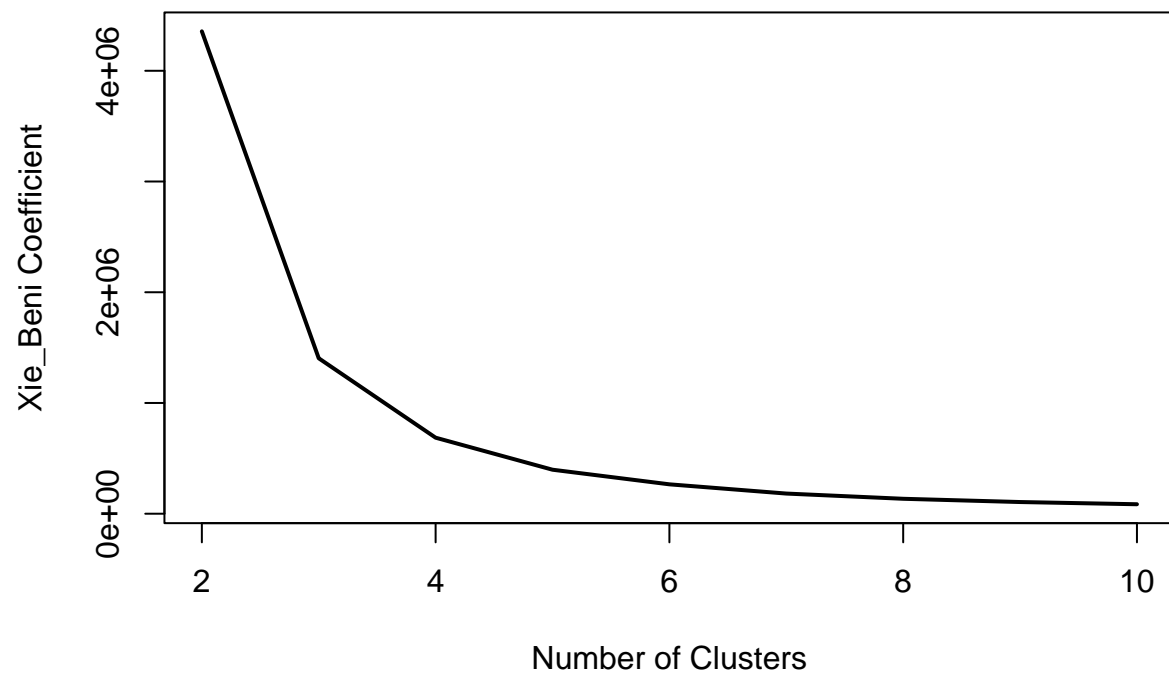
Cluster Rugs with Histogram & Cutoff Line

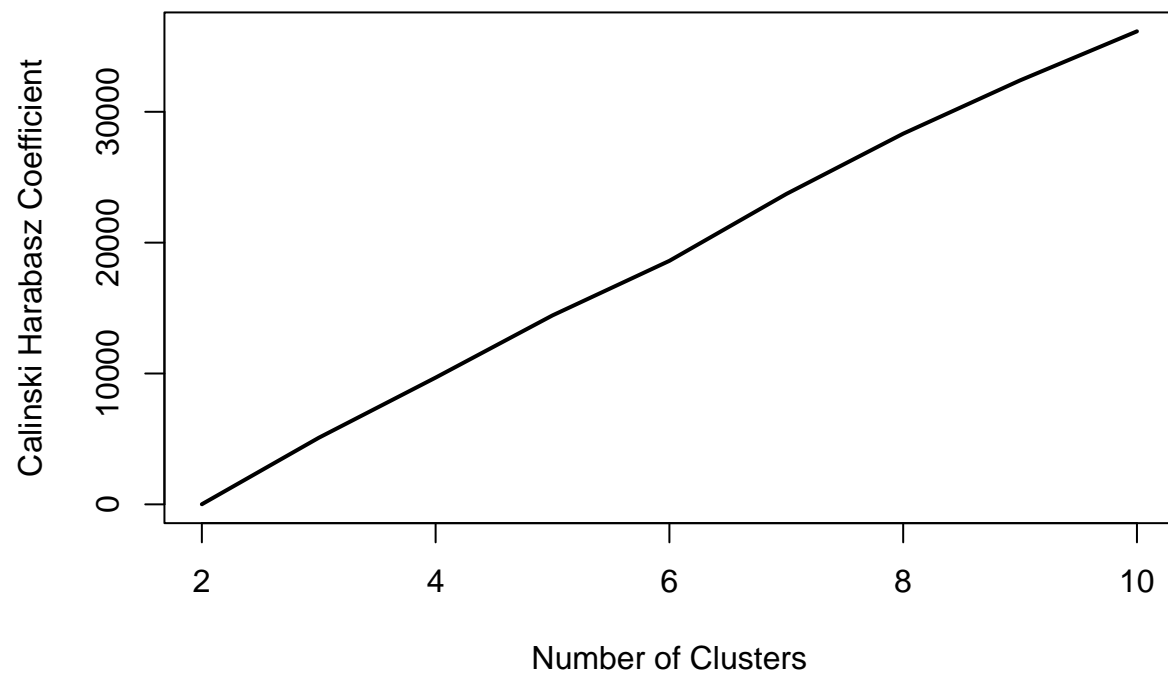


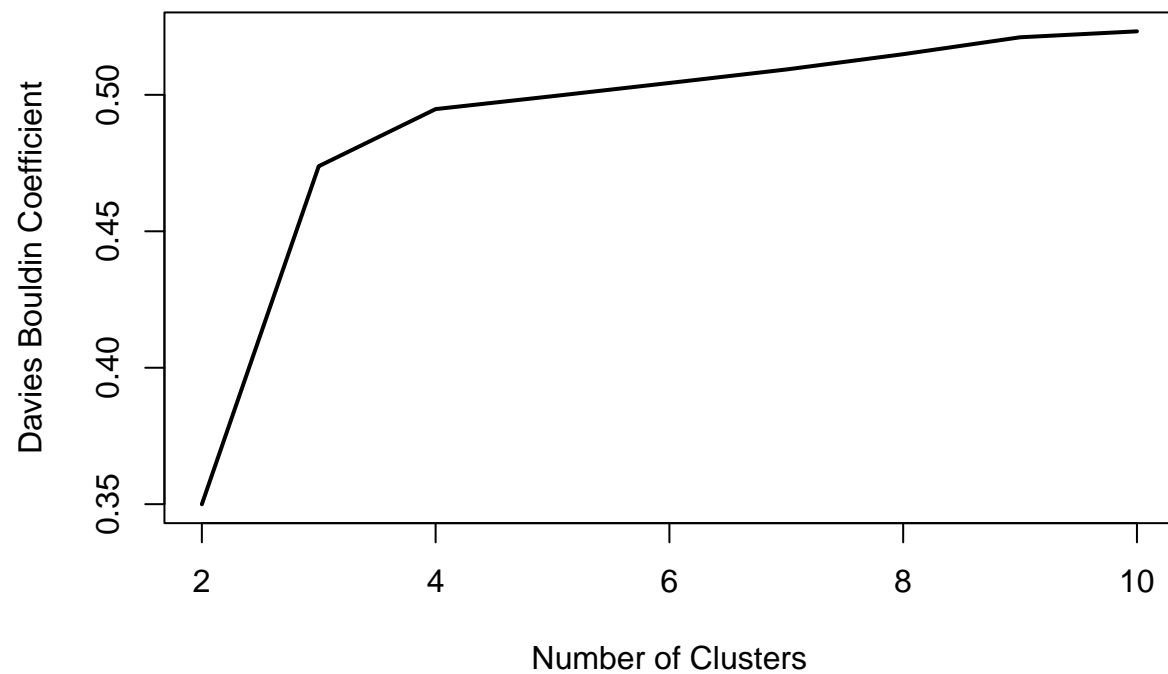
Childcare





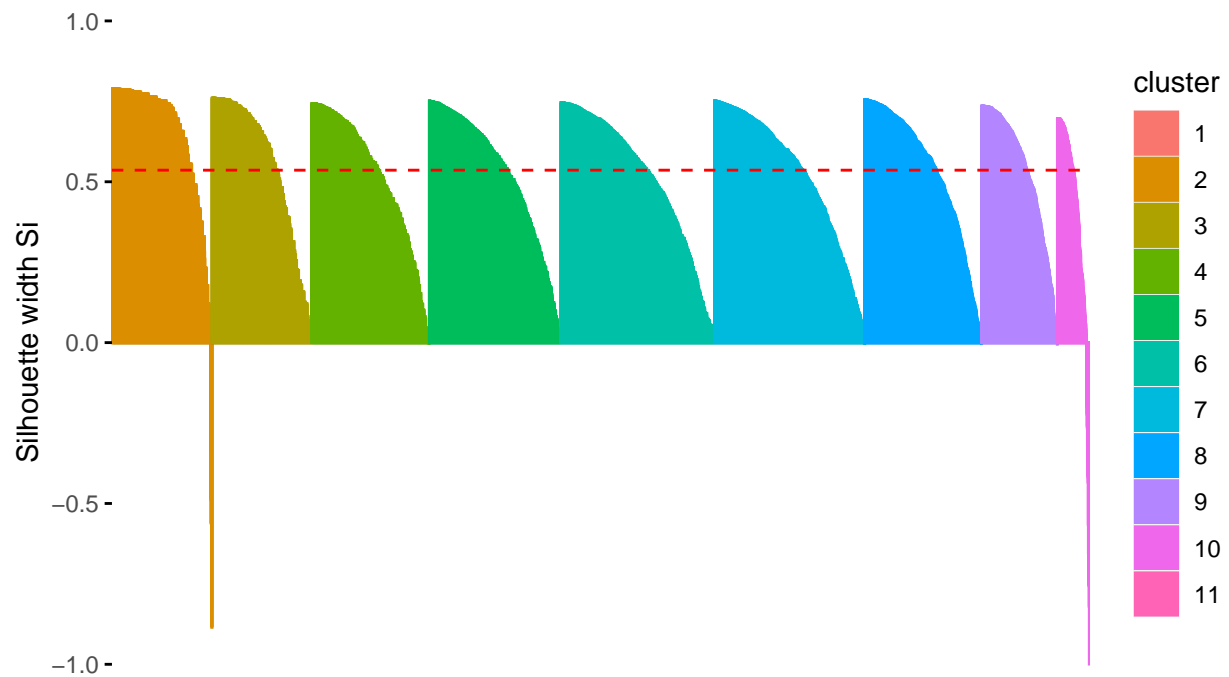




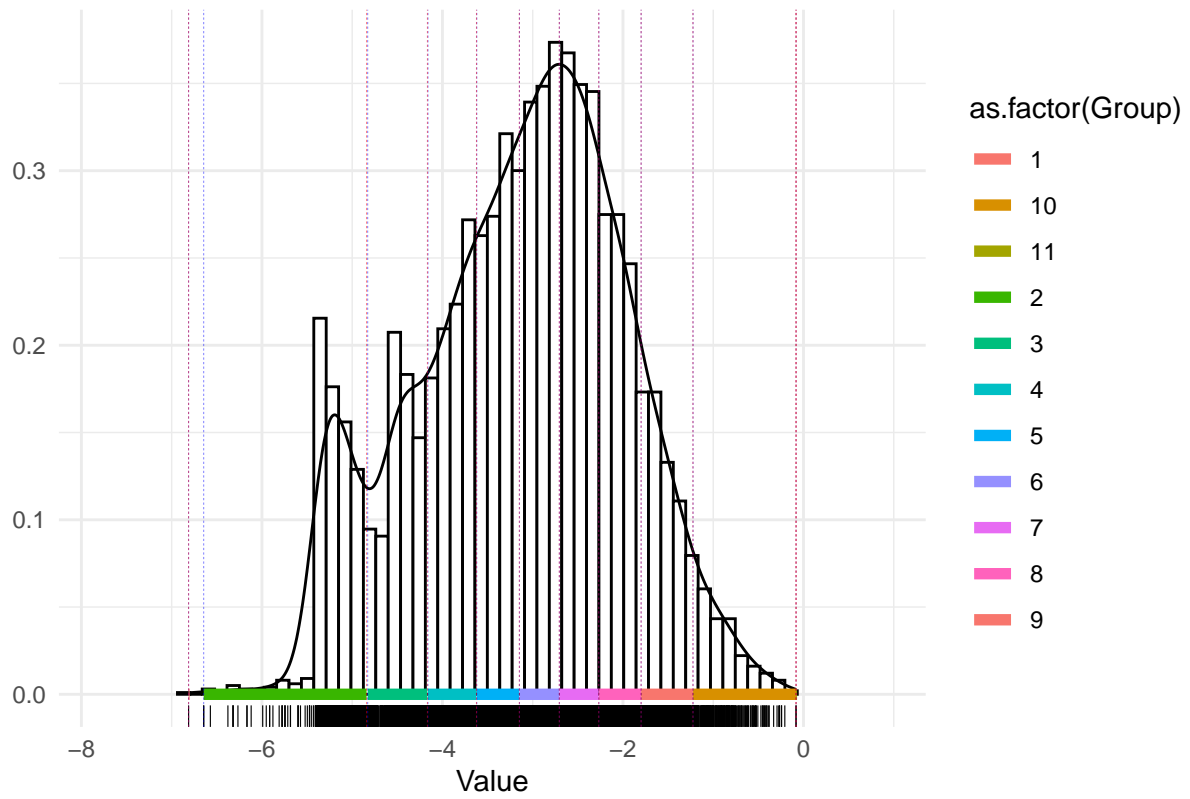


##	cluster	size	ave.sil.width
## 1	1	1	0.00
## 2	2	739	0.65
## 3	3	738	0.55
## 4	4	869	0.52
## 5	5	974	0.53
## 6	6	1136	0.51
## 7	7	1111	0.52
## 8	8	864	0.54
## 9	9	561	0.53
## 10	10	237	0.40
## 11	11	1	0.00

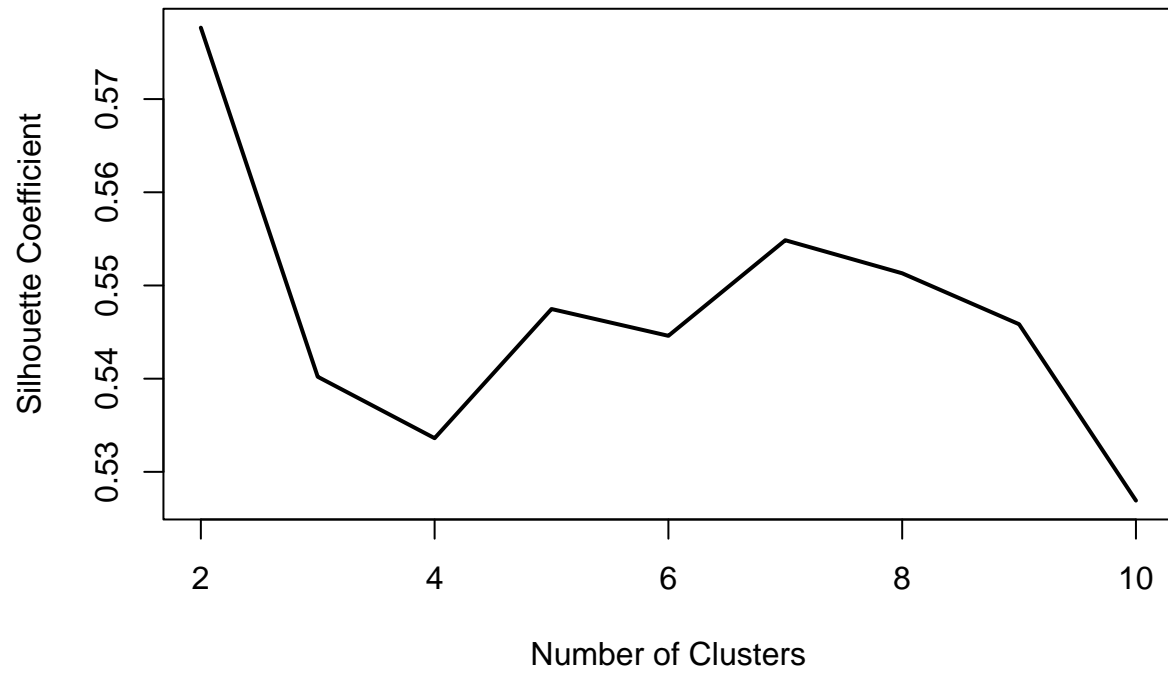
Clusters silhouette plot
Average silhouette width: 0.54

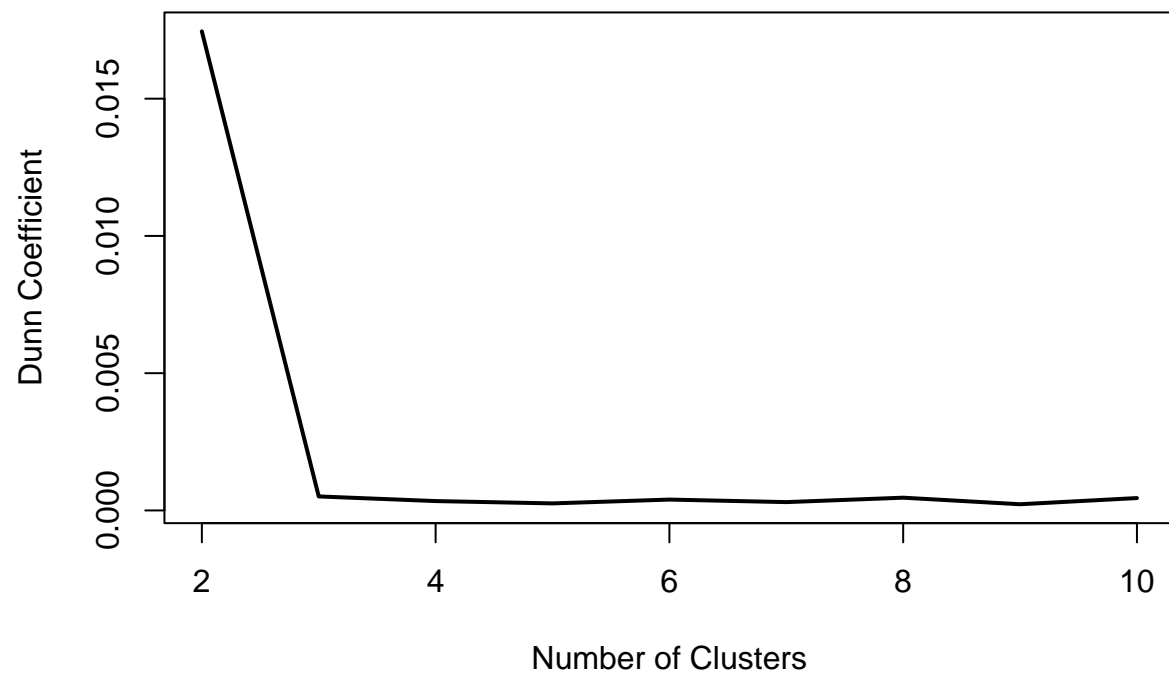


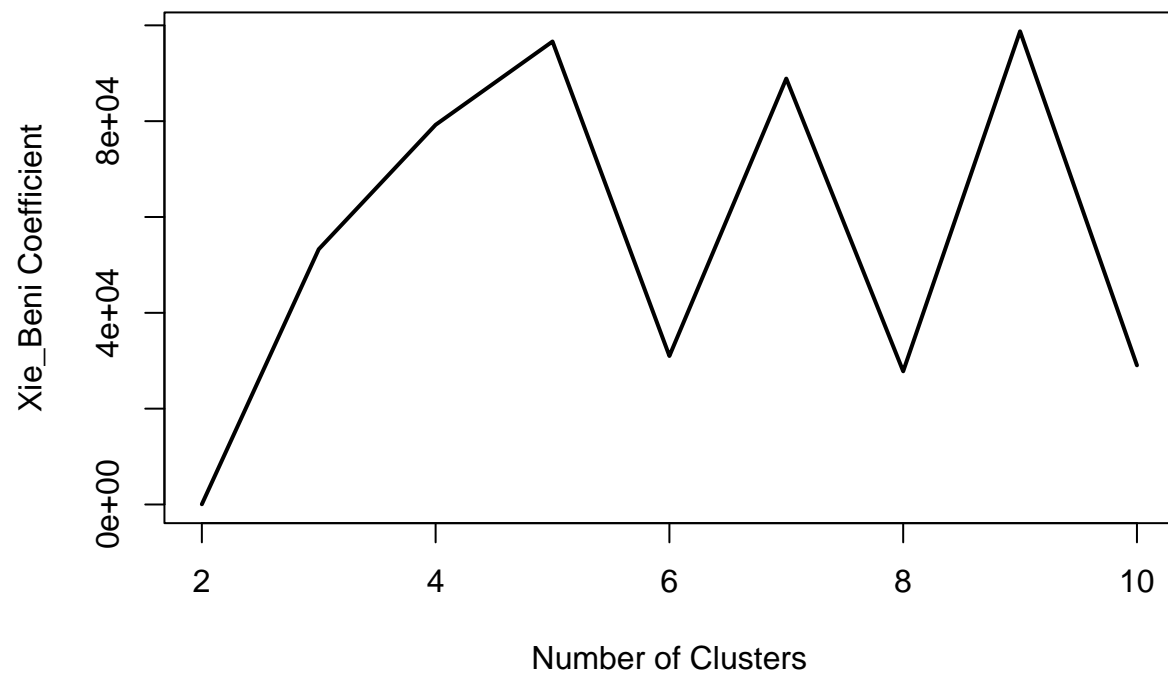
Cluster Rugs with Histogram & Cutoff Line

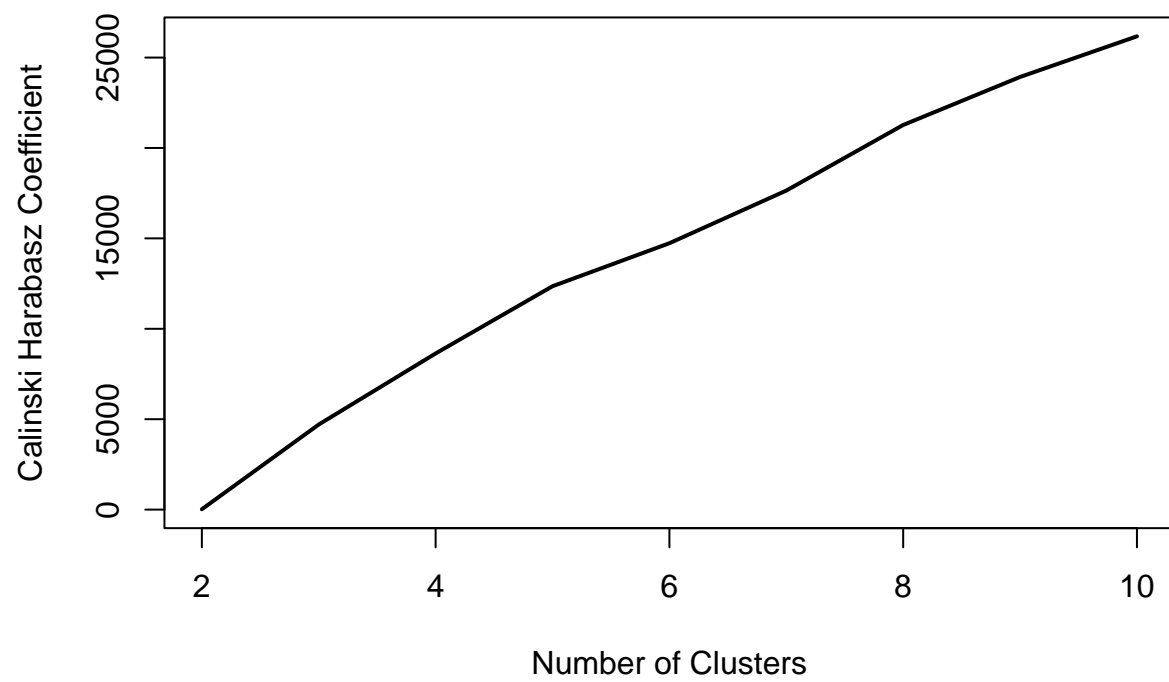


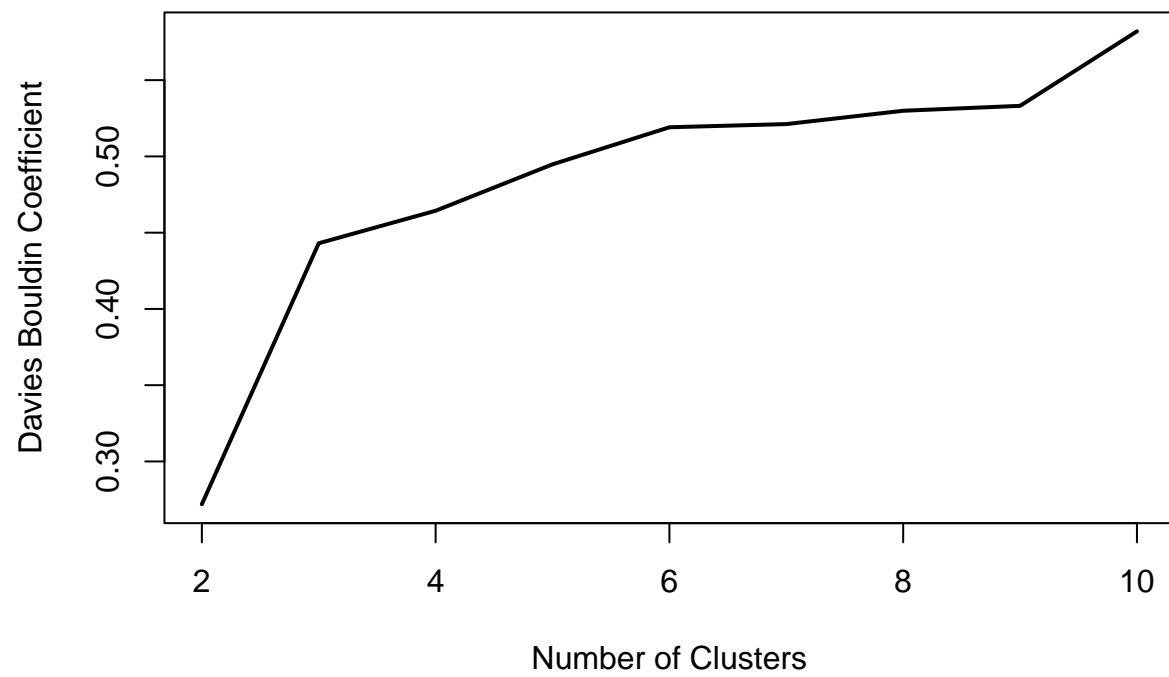
Parks





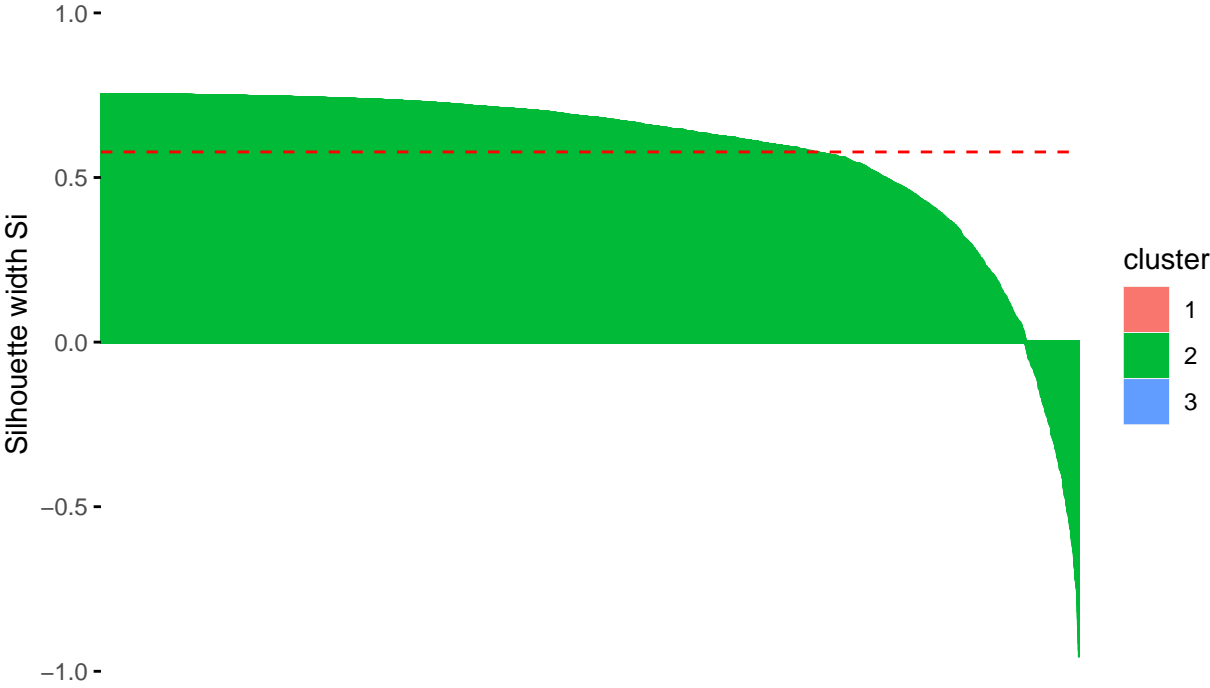




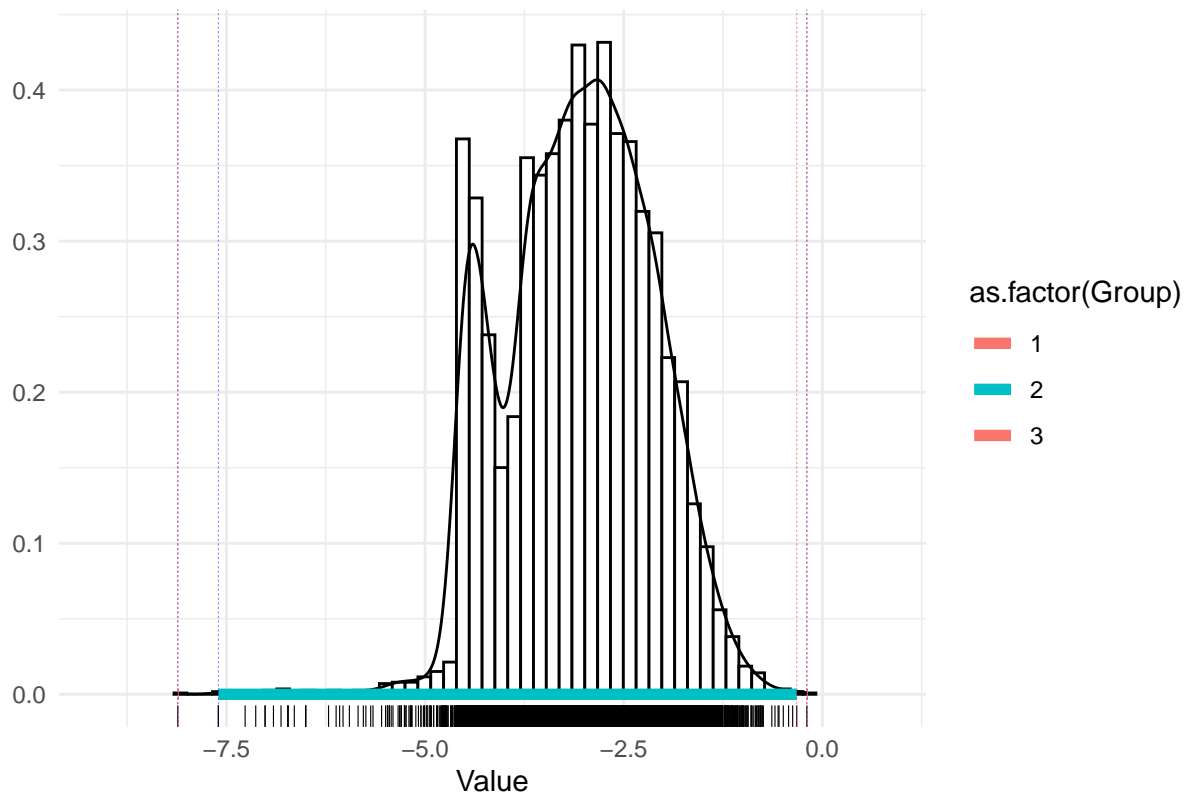


```
## cluster size ave.sil.width
## 1      1      1          0.00
## 2      2 6967          0.58
## 3      3      1          0.00
```

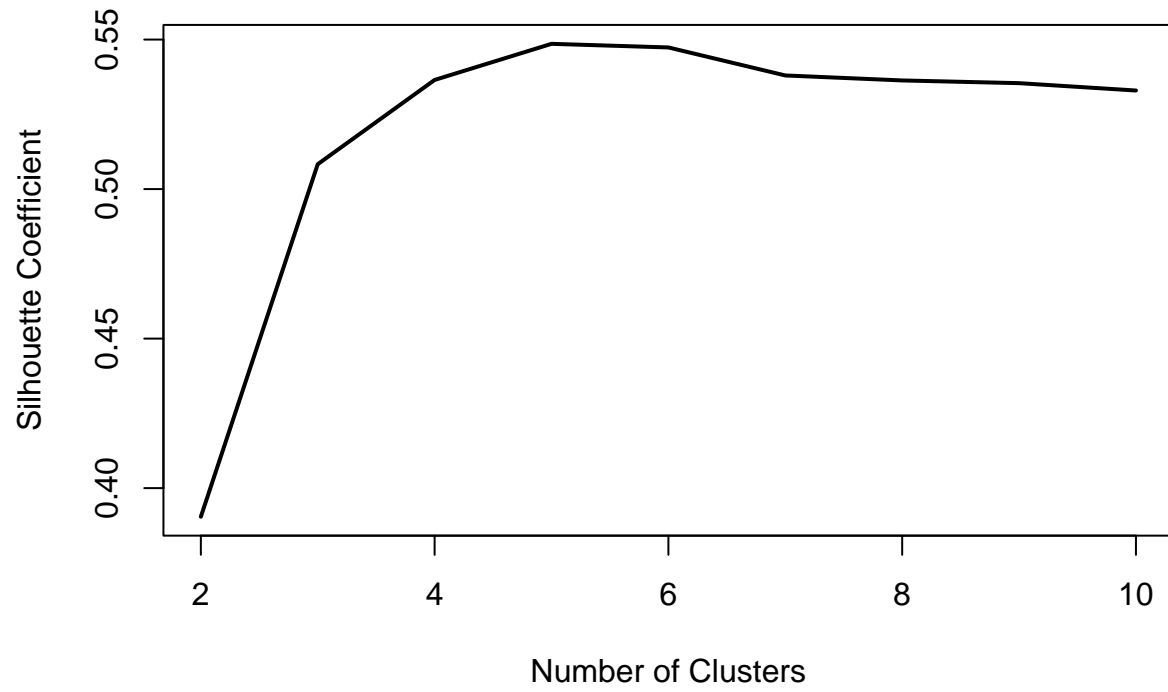
Clusters silhouette plot
Average silhouette width: 0.58

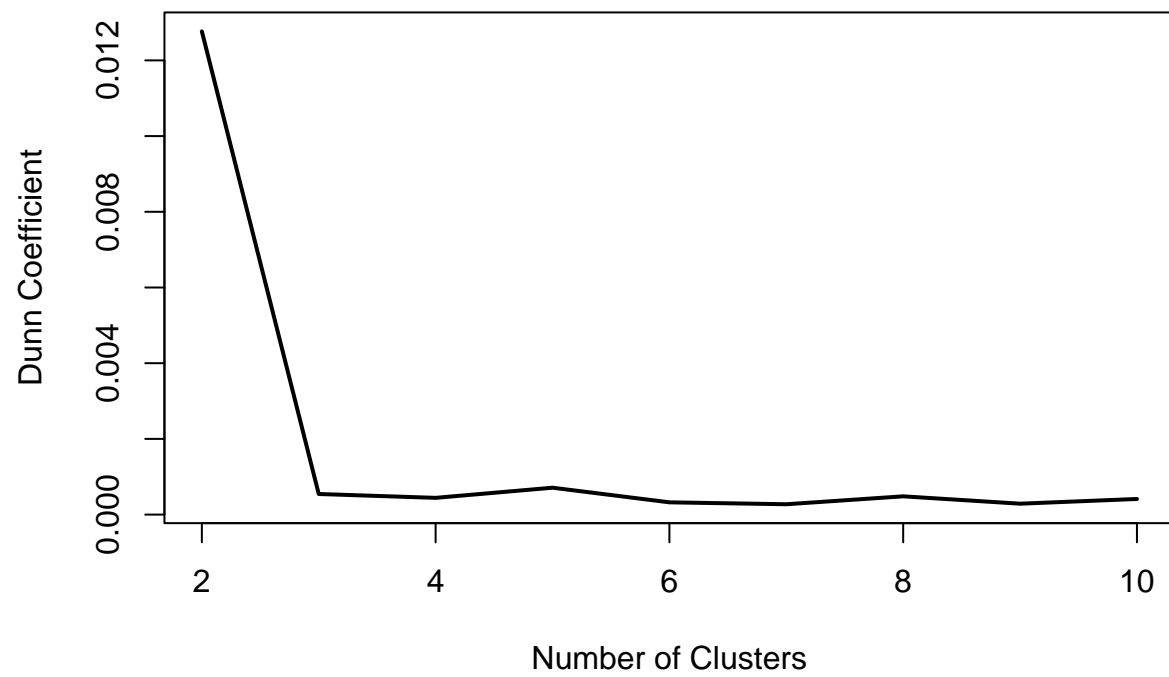


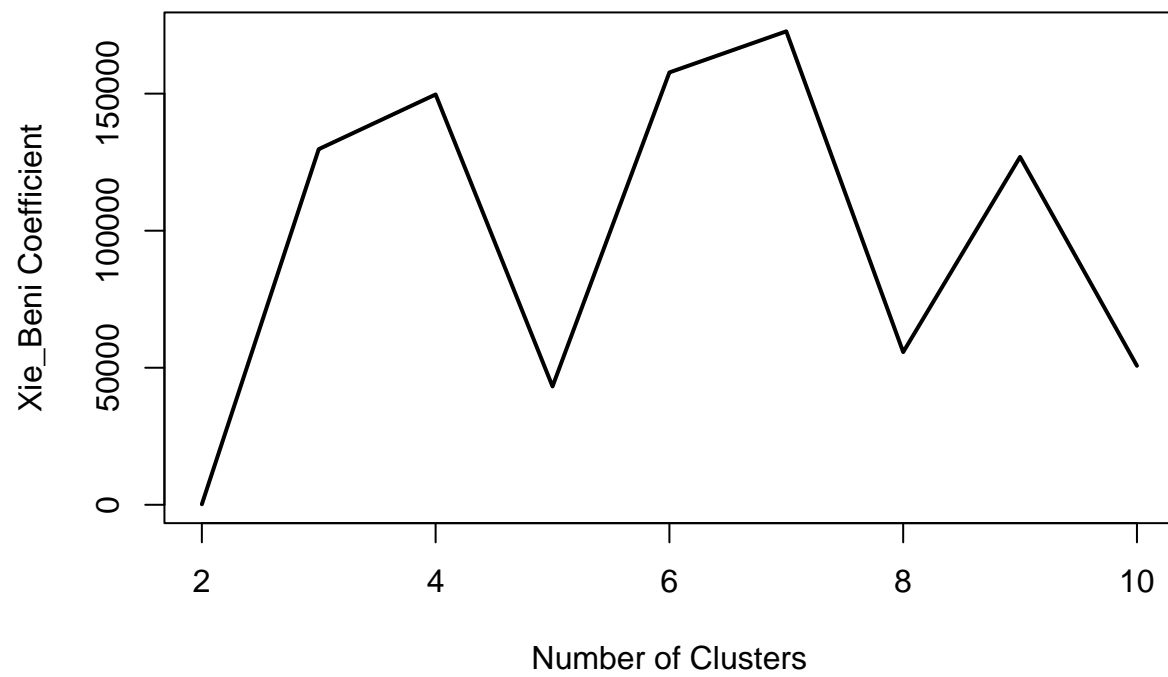
Cluster Rugs with Histogram & Cutoff Line

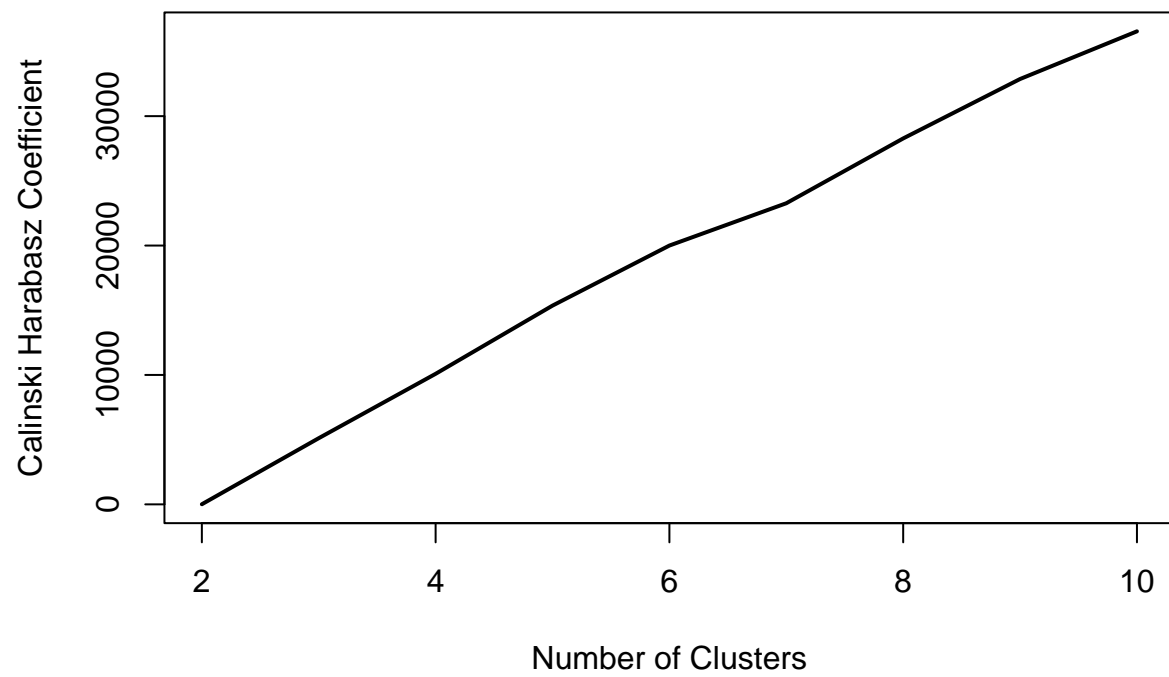


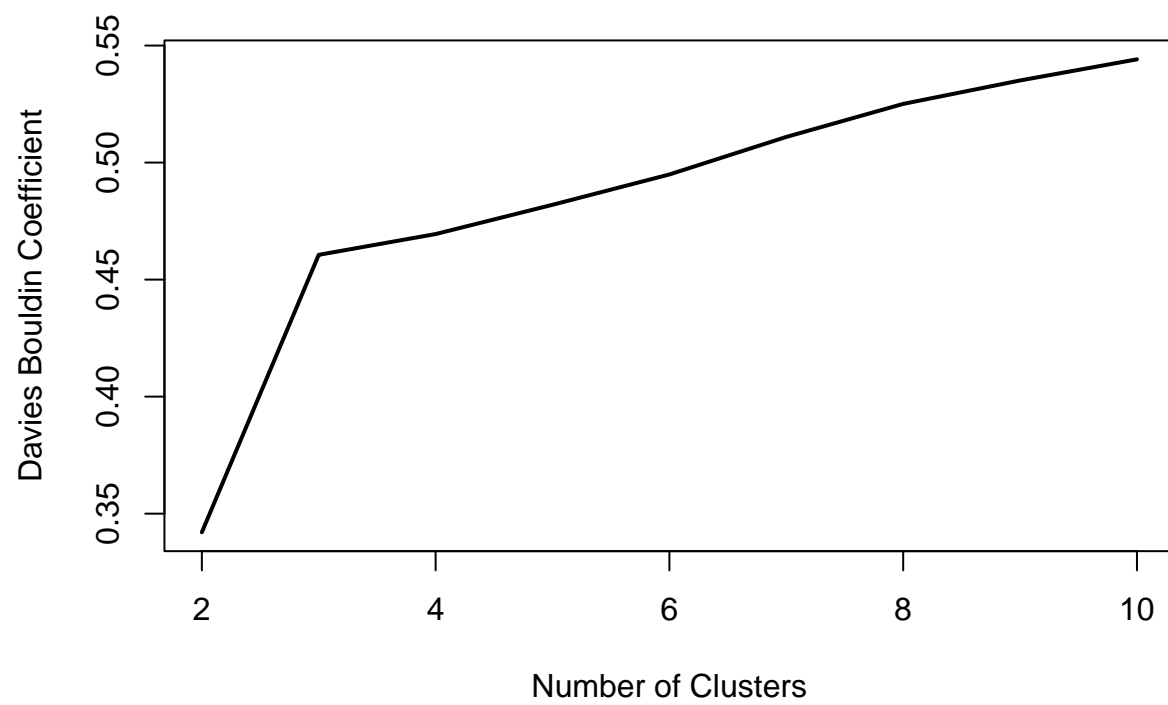
Education Primary



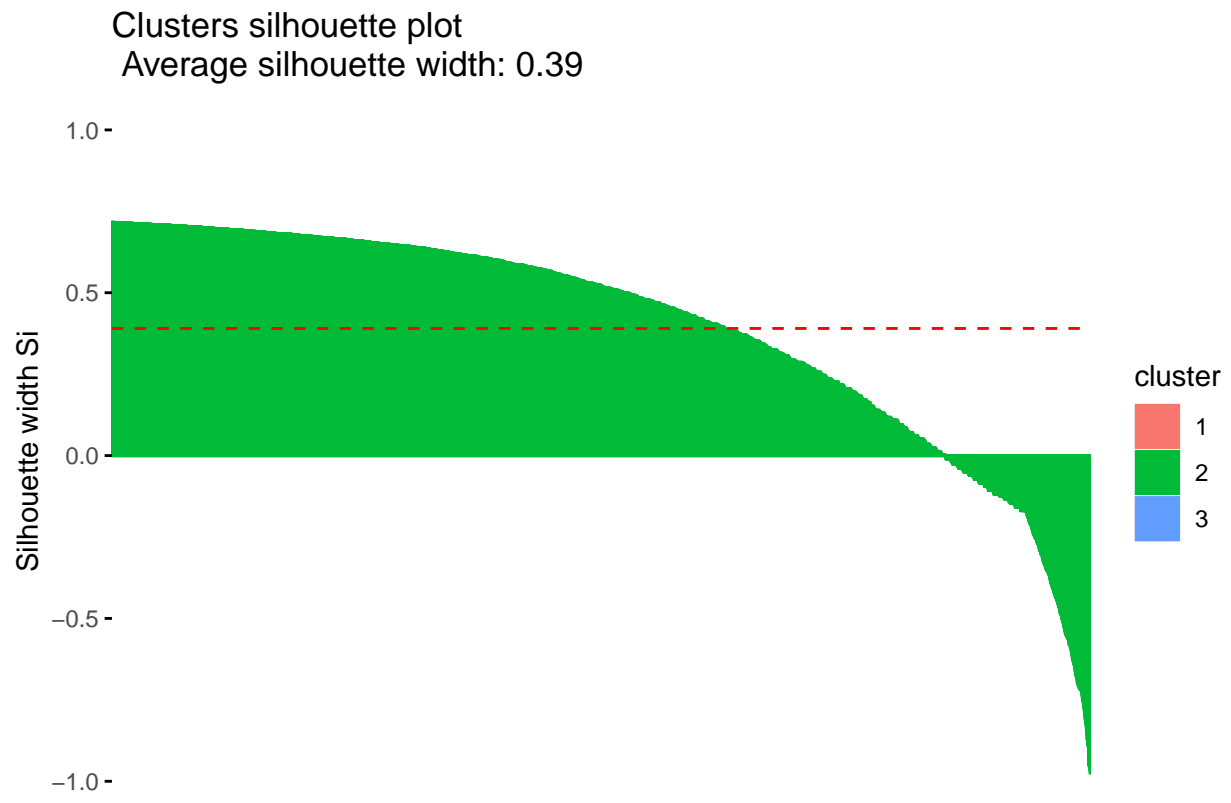




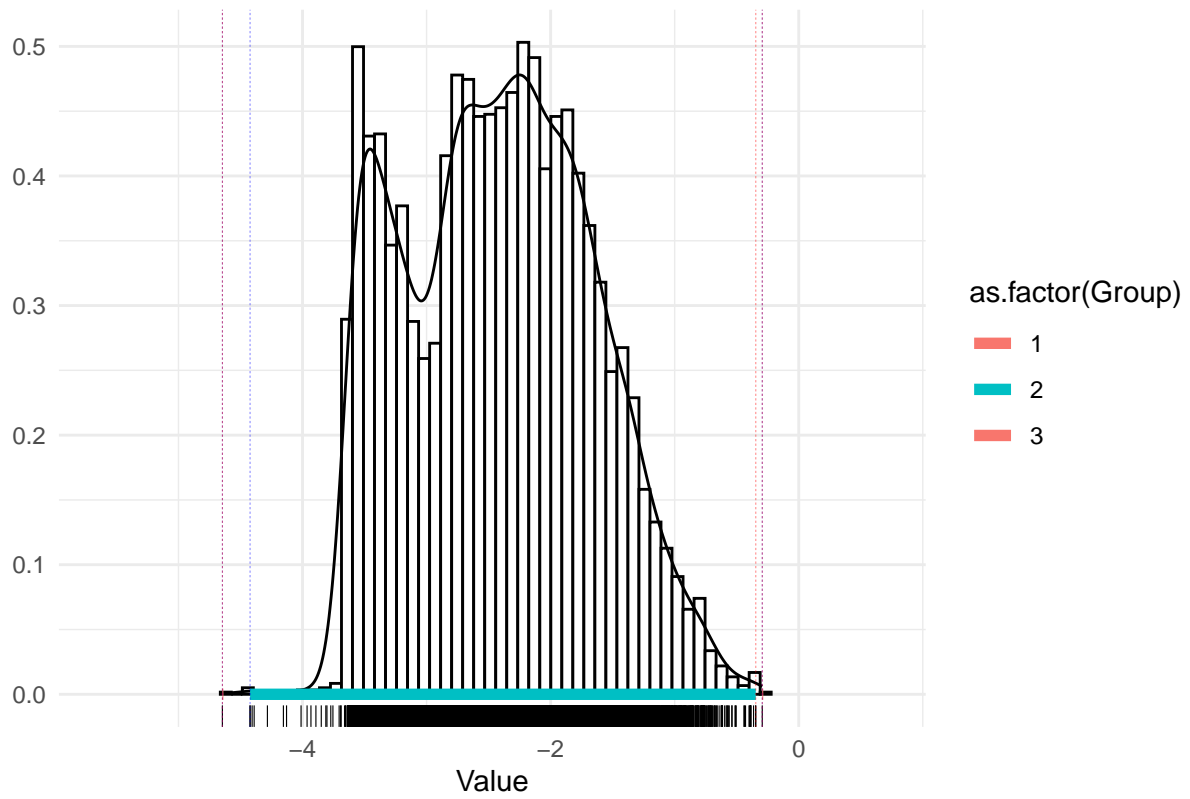




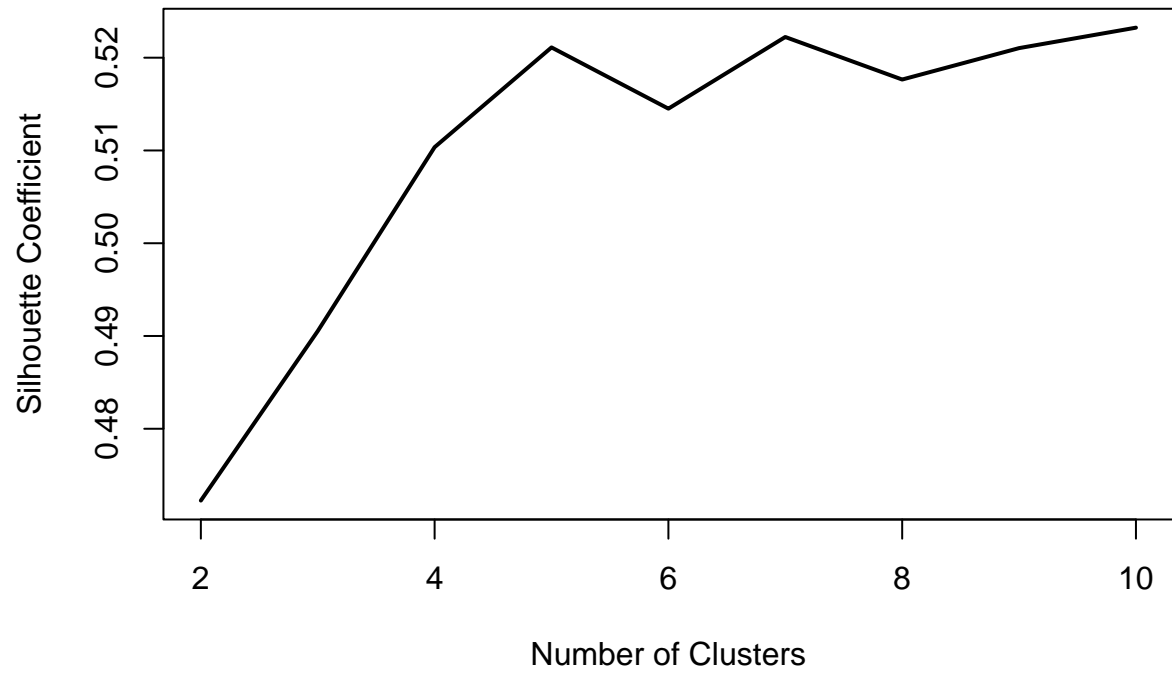
```
## cluster size ave.sil.width
## 1      1      1           0.00
## 2      2 6689           0.39
## 3      3      1           0.00
```

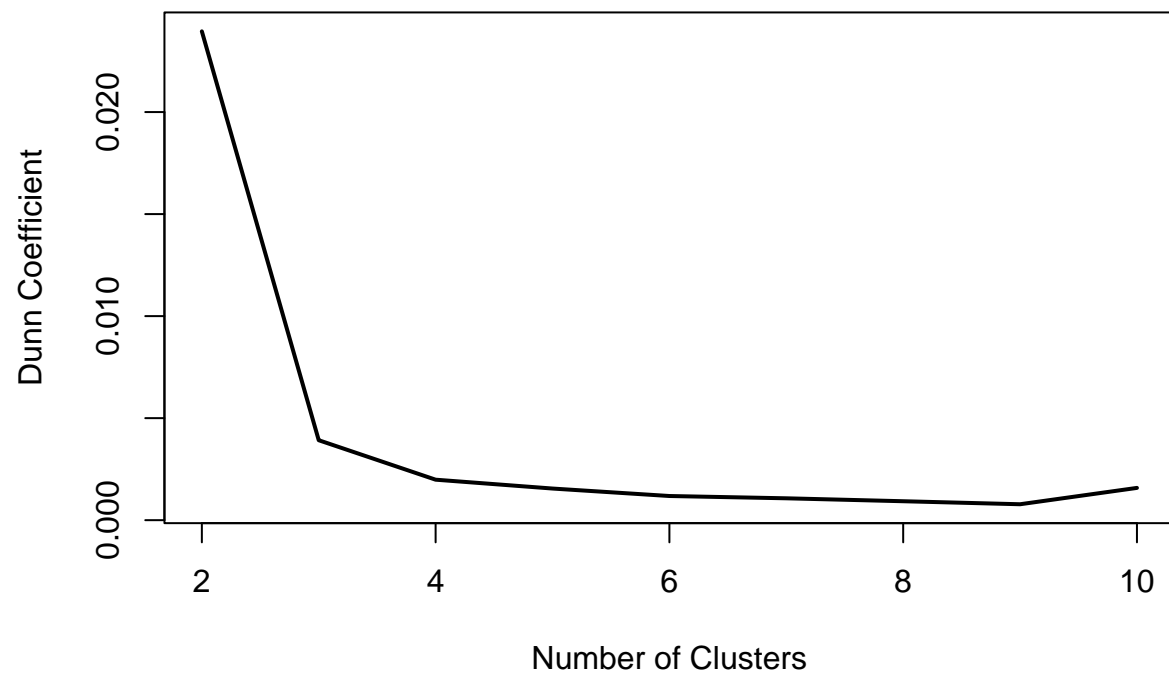


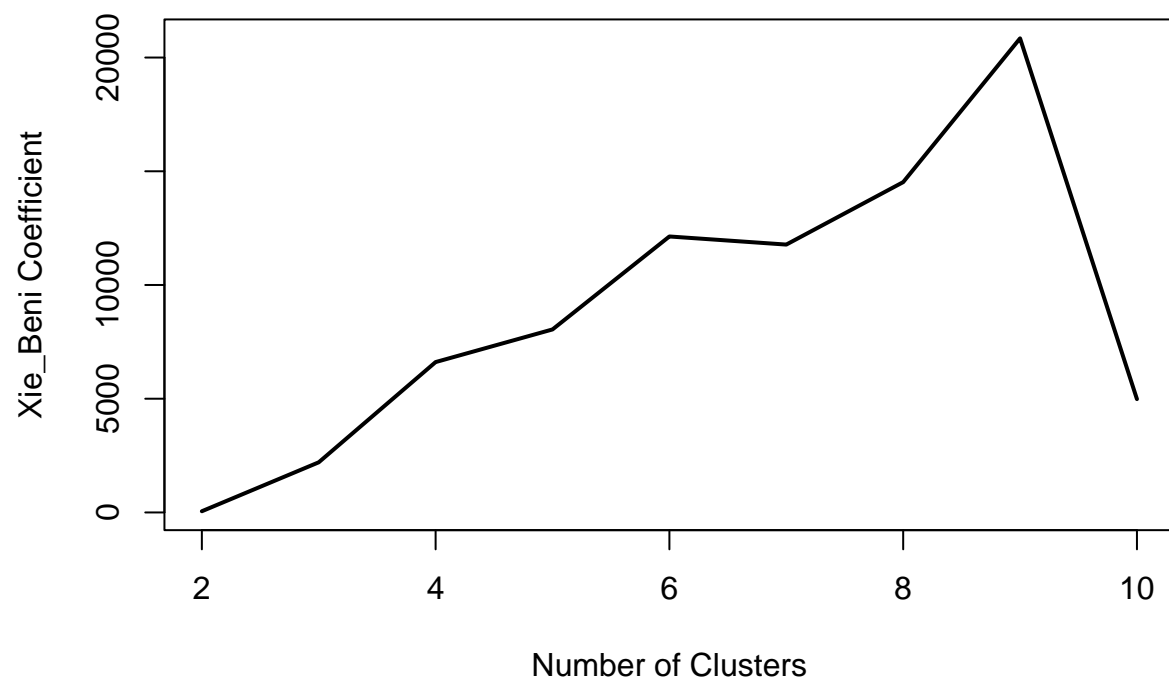
Cluster Rugs with Histogram & Cutoff Line

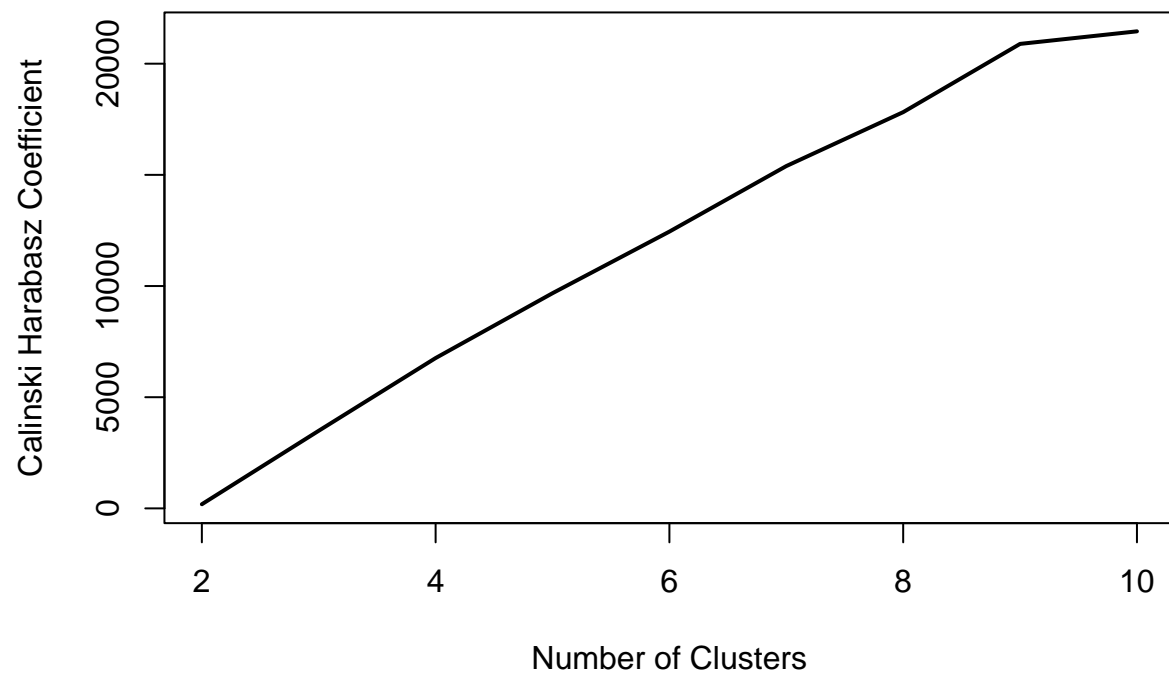


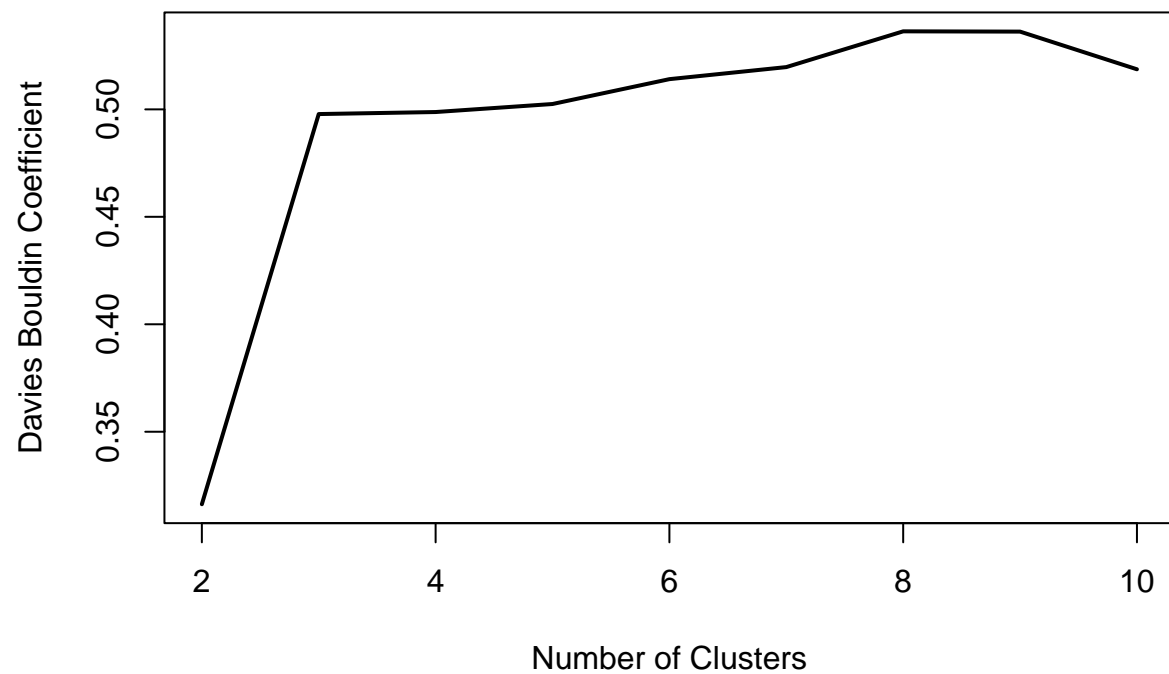
Transit



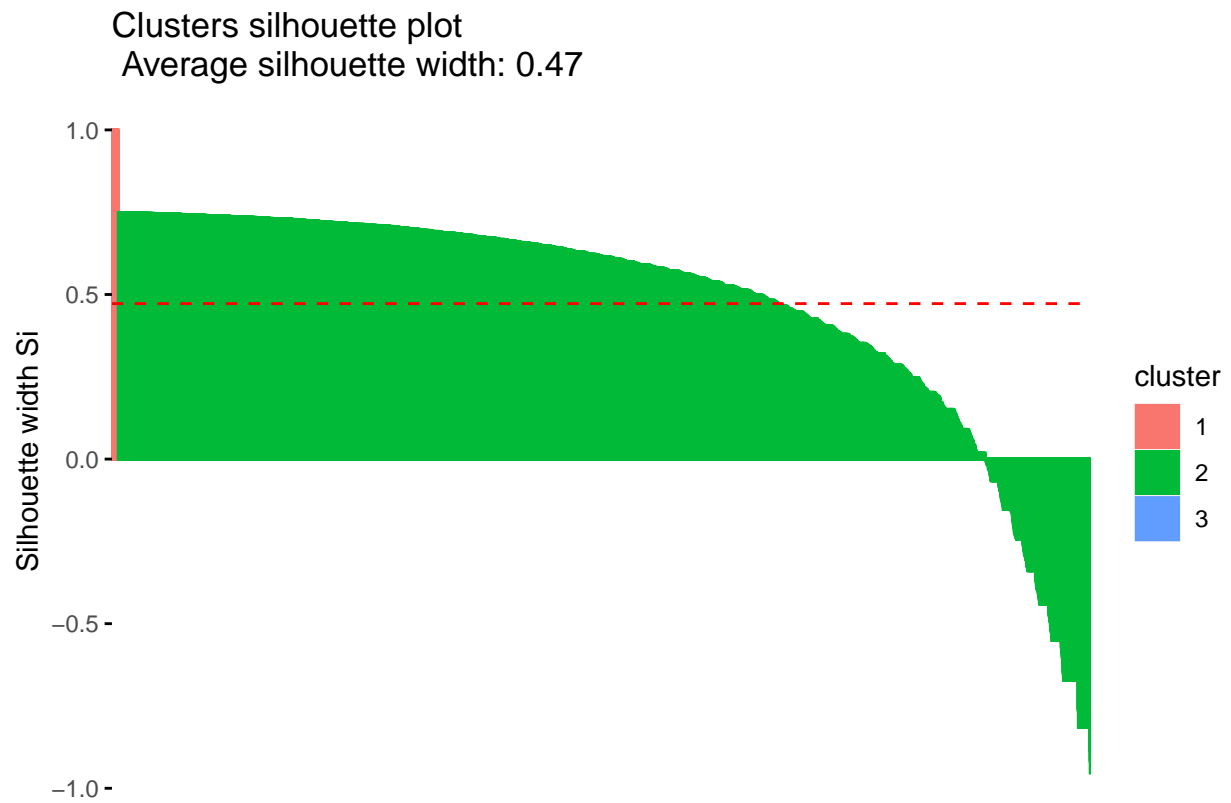




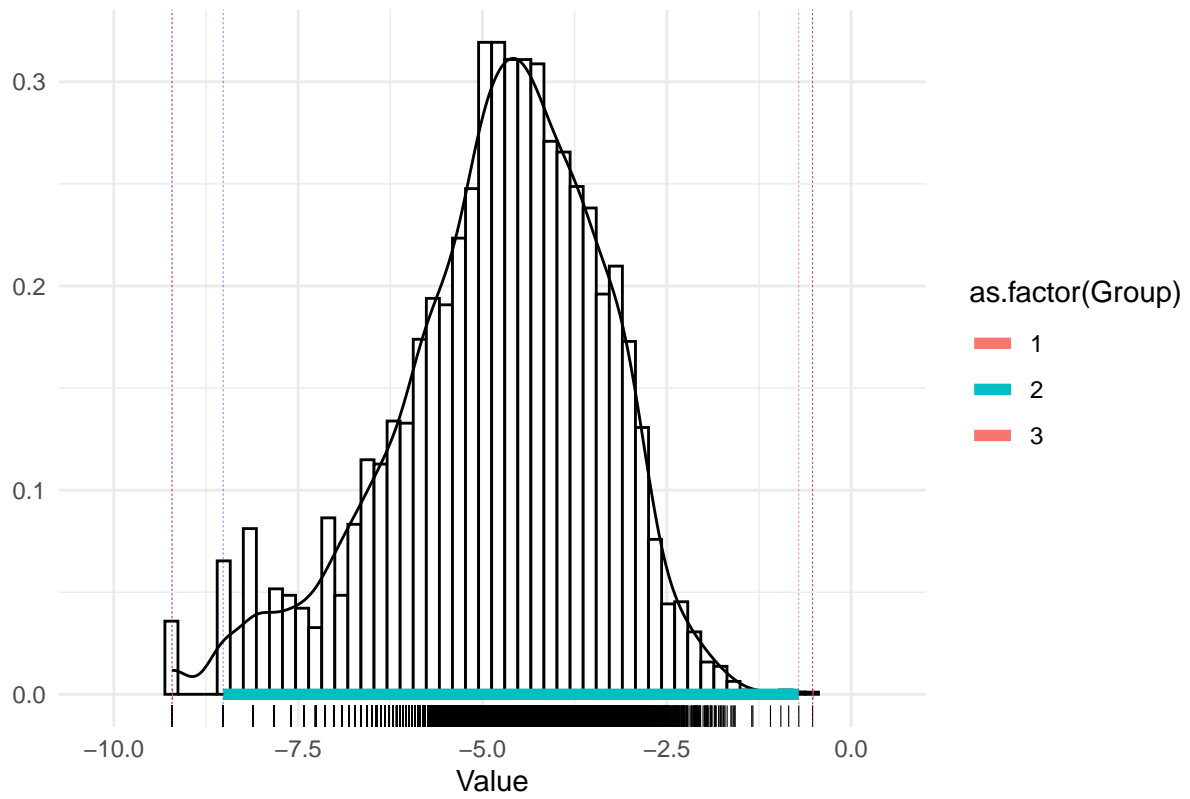




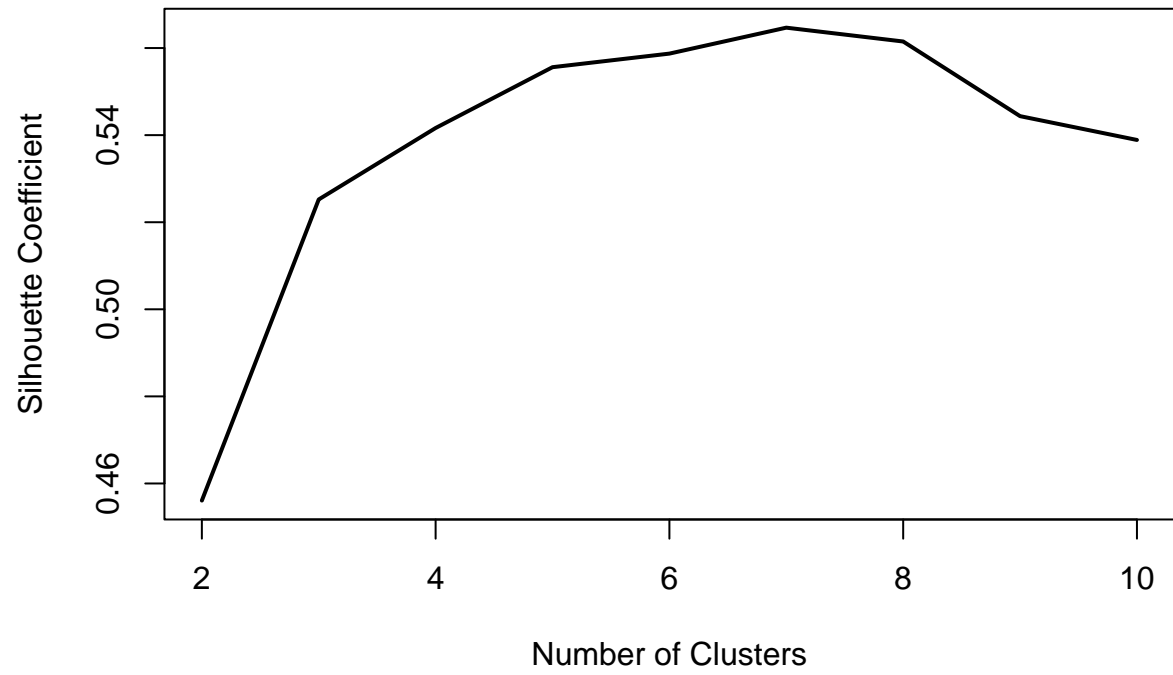
```
## cluster size ave.sil.width
## 1      1  34          1.00
## 2      2 5317          0.47
## 3      3   1          0.00
```

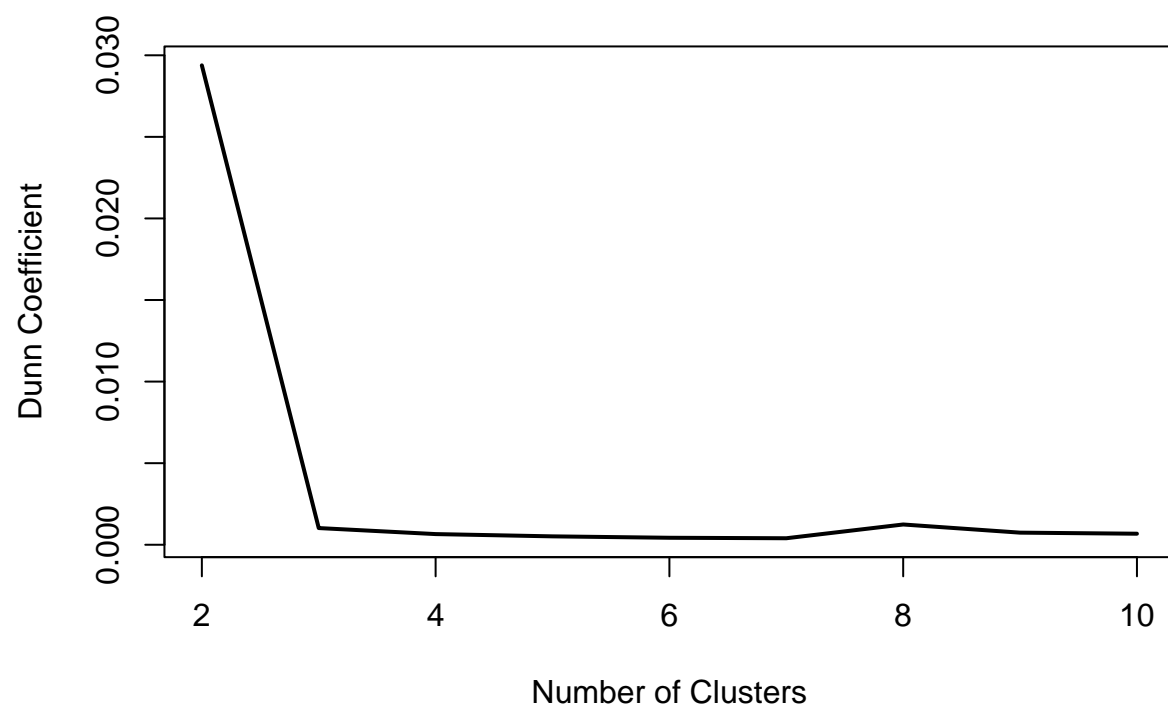


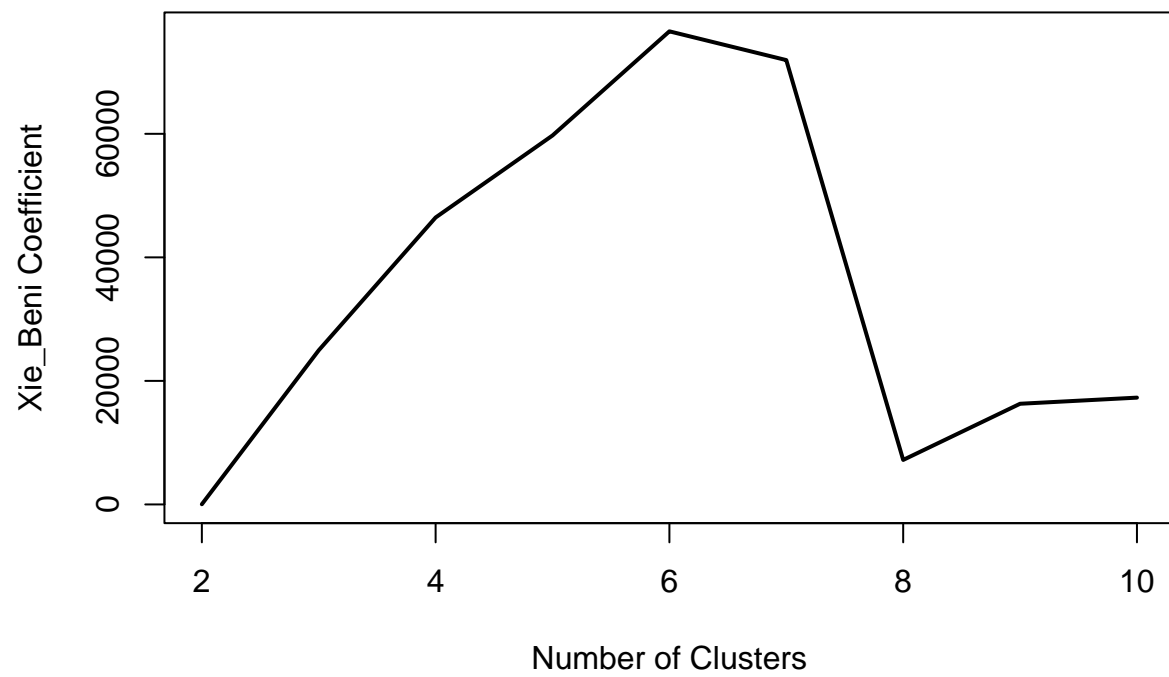
Cluster Rugs with Histogram & Cutoff Line

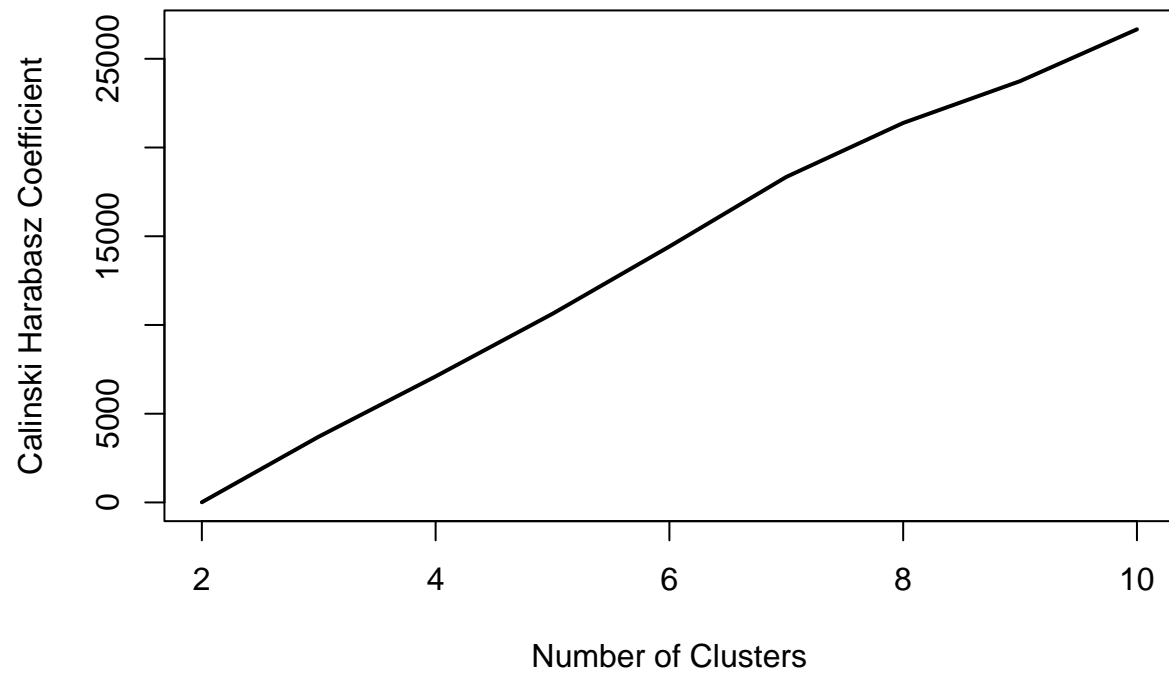


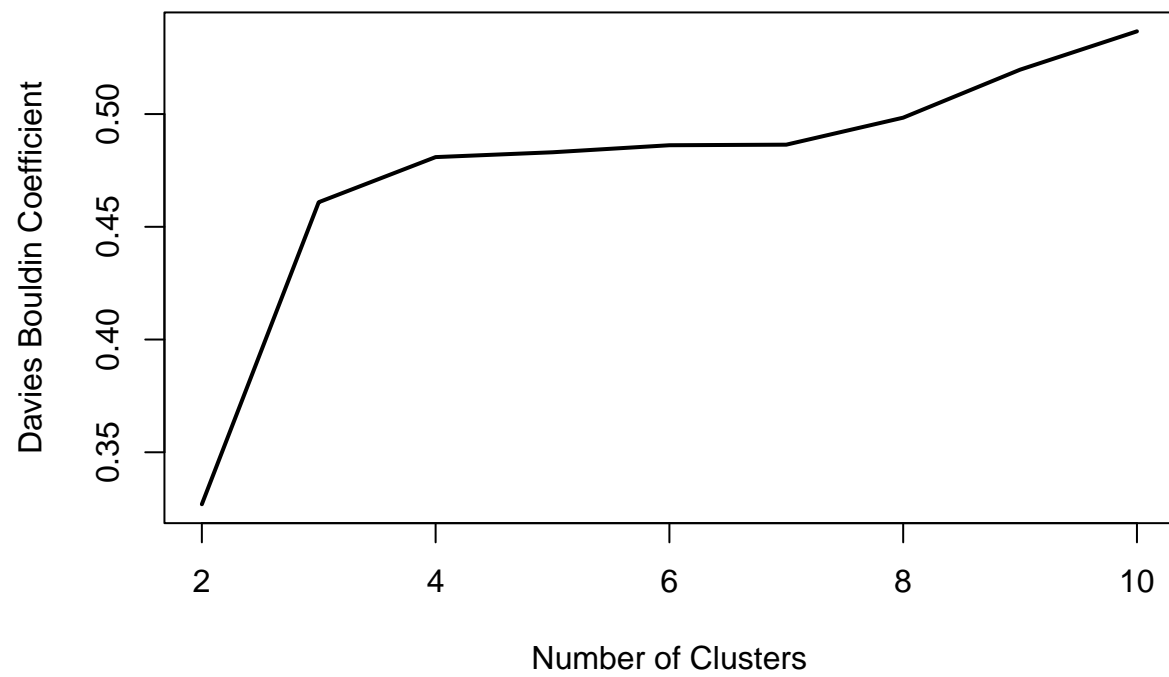
Pharma



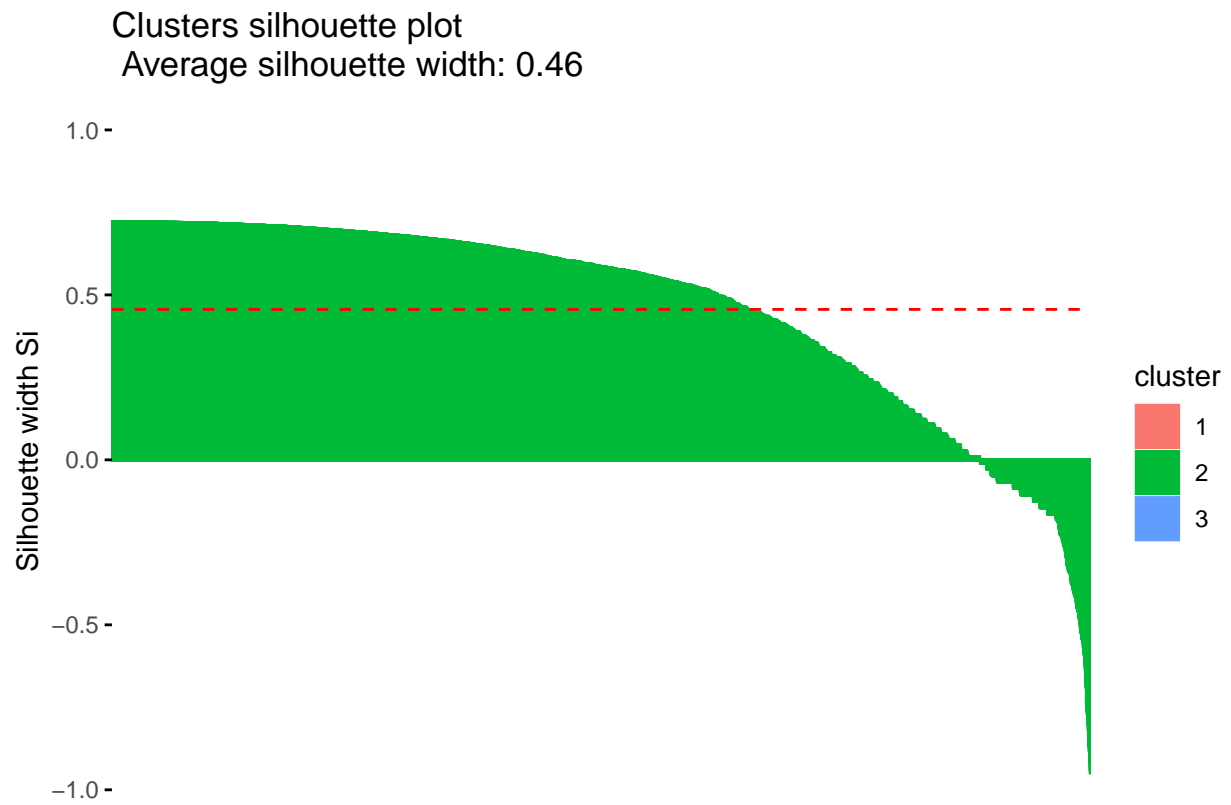


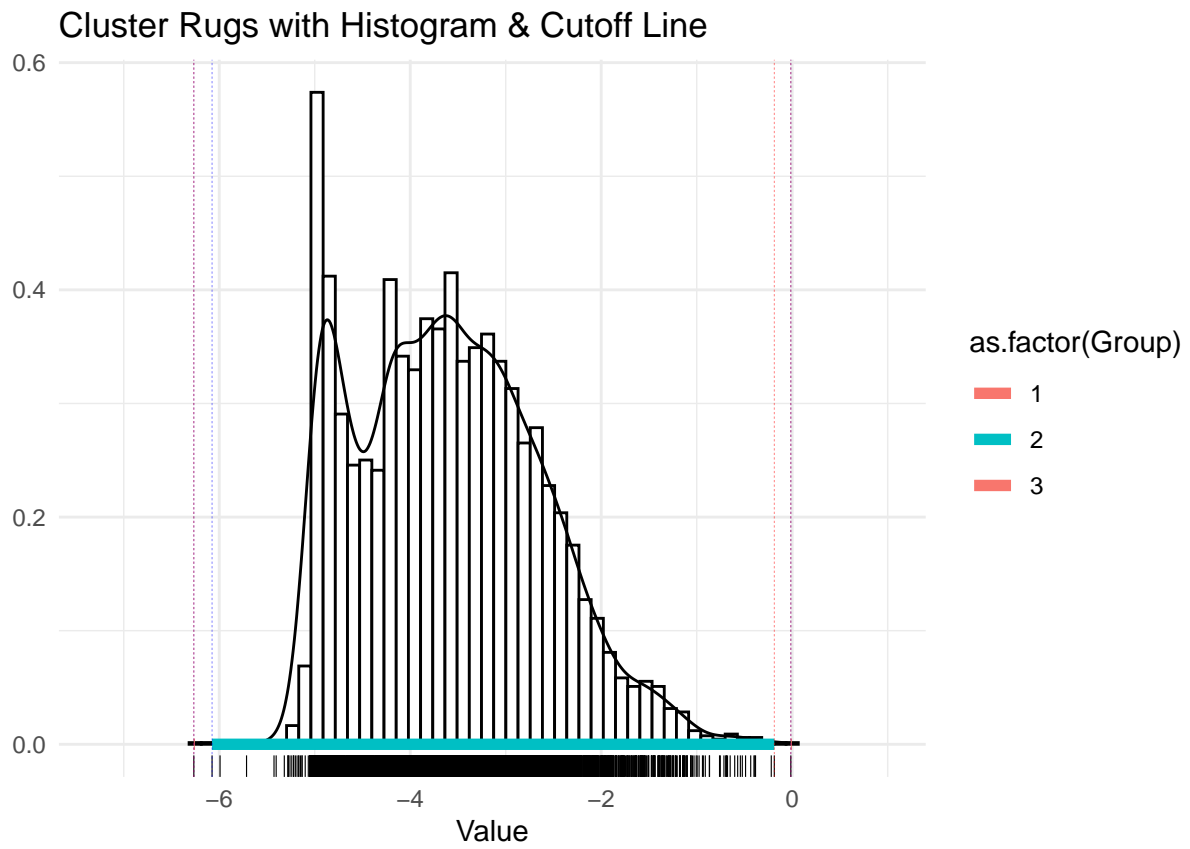




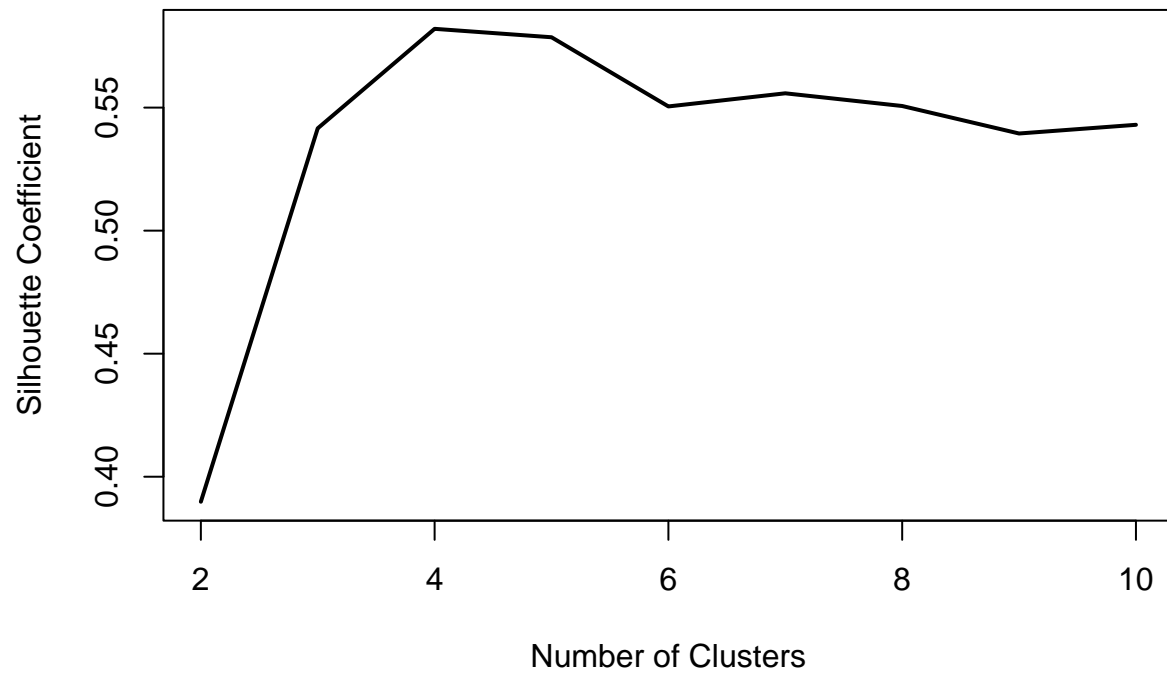


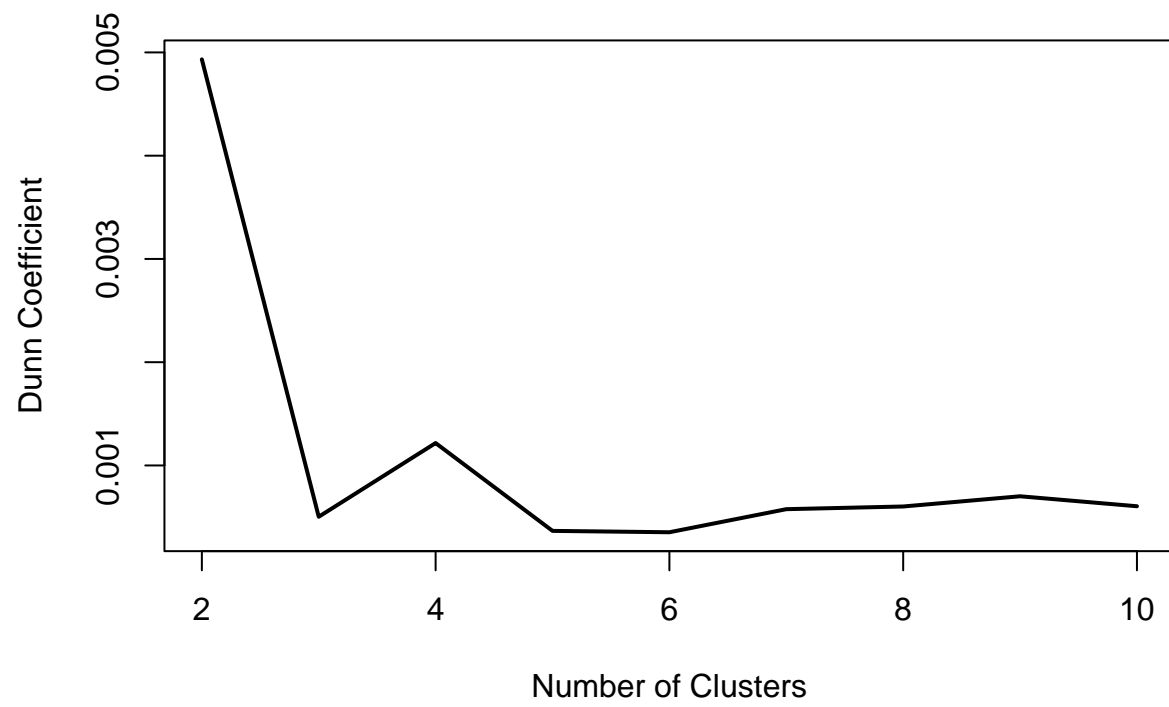
```
## cluster size ave.sil.width
## 1      1      1          0.00
## 2      2 5229          0.46
## 3      3      1          0.00
```

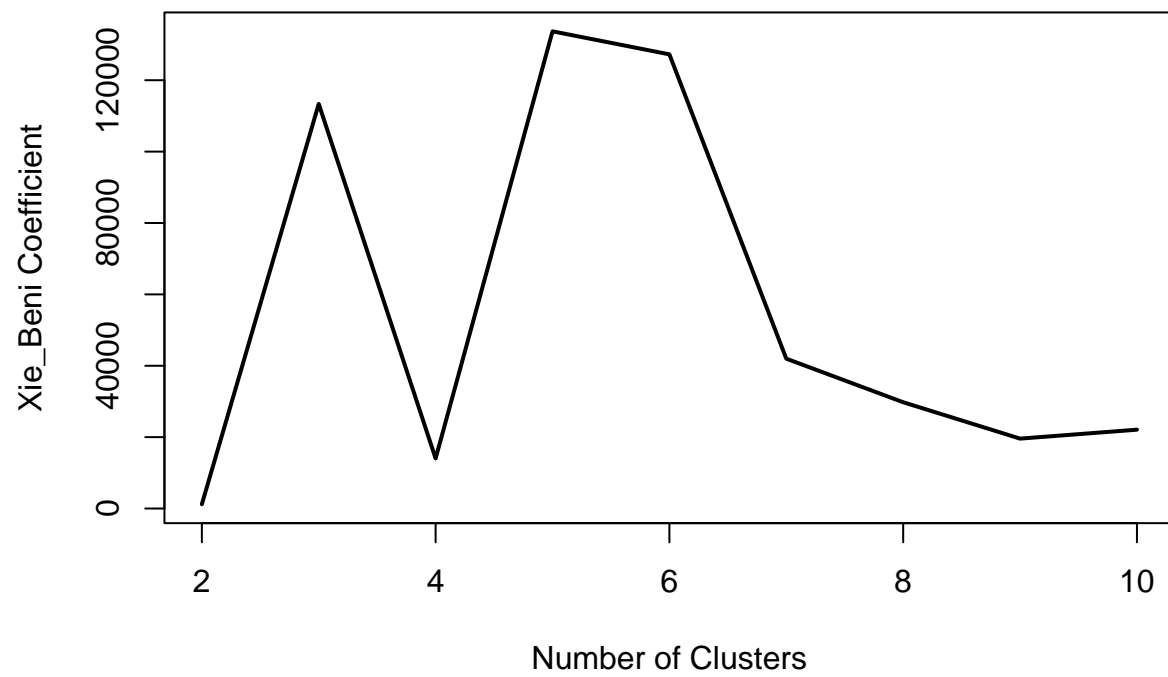



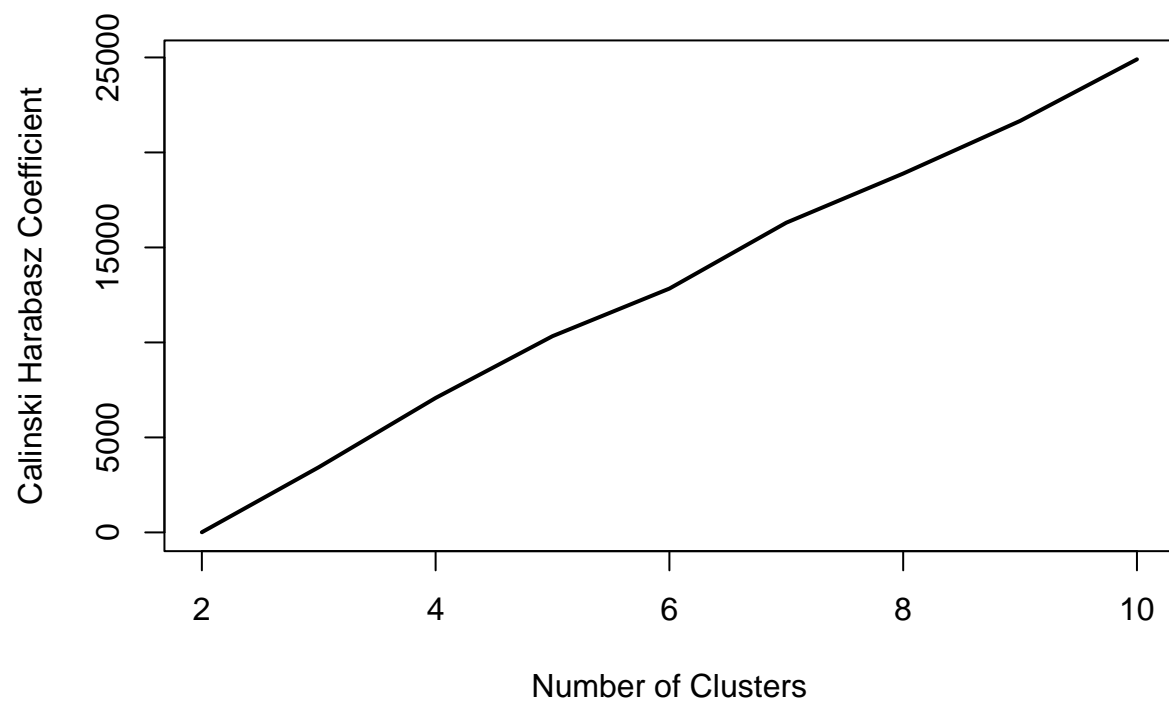


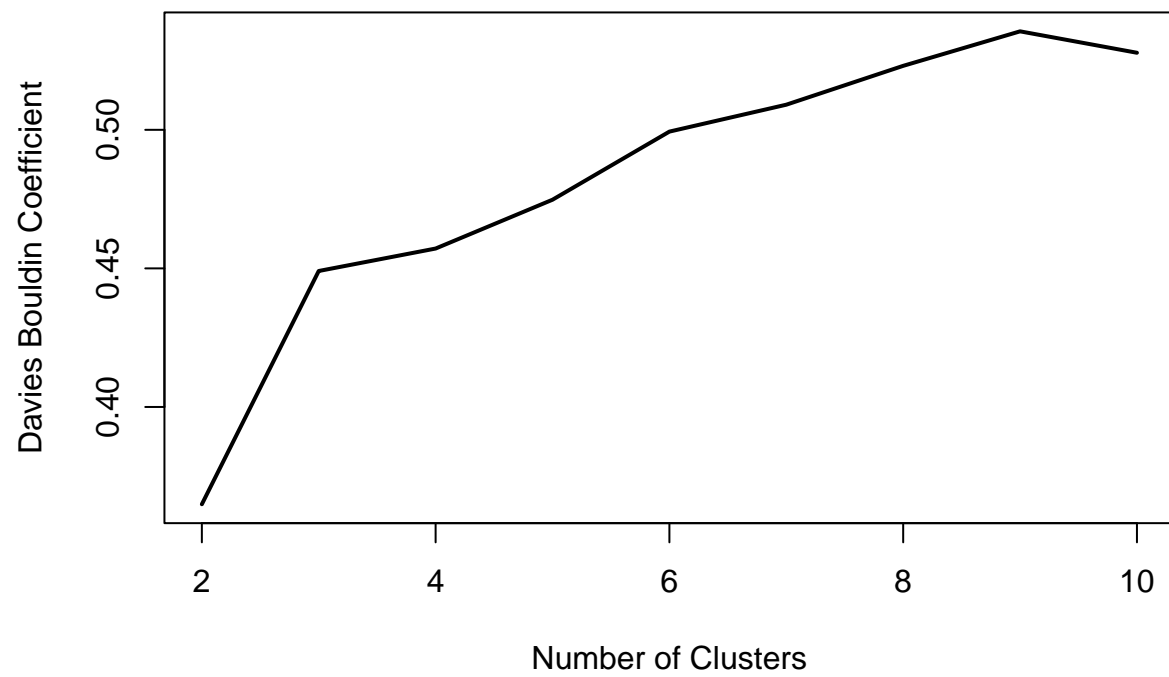
Education Secondary



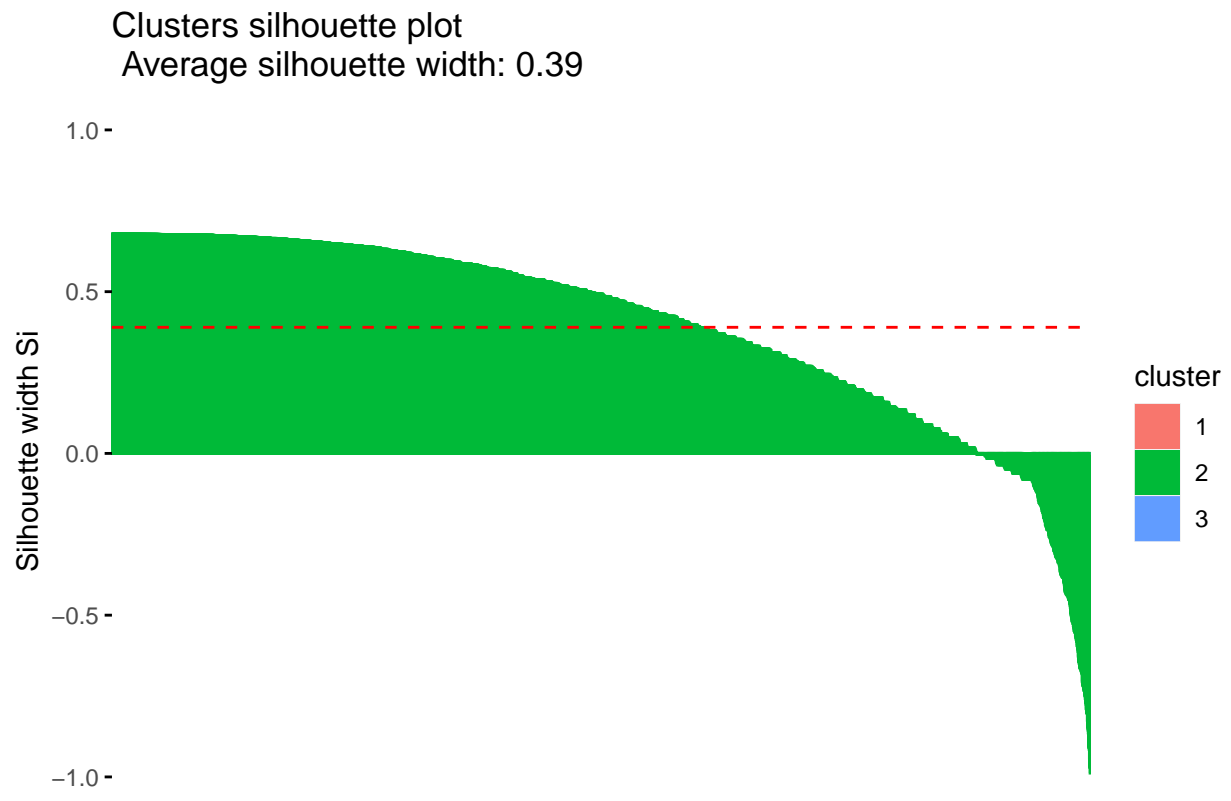




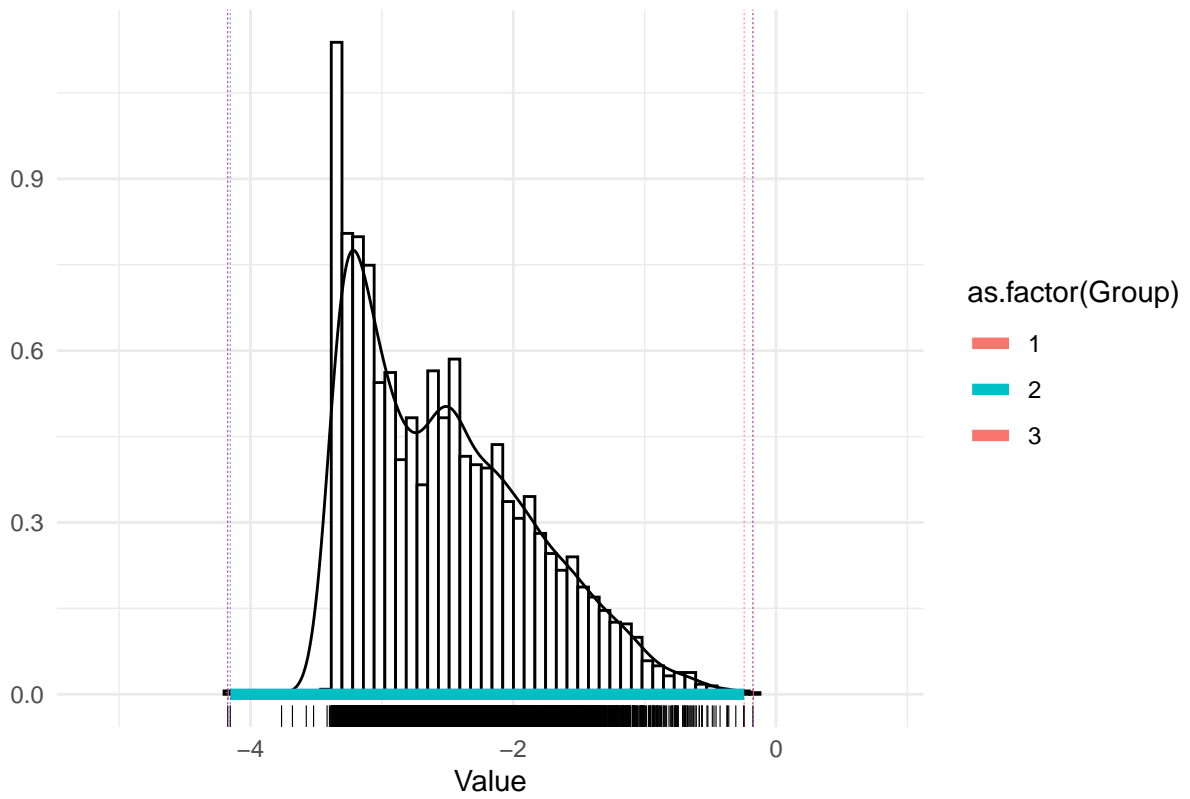




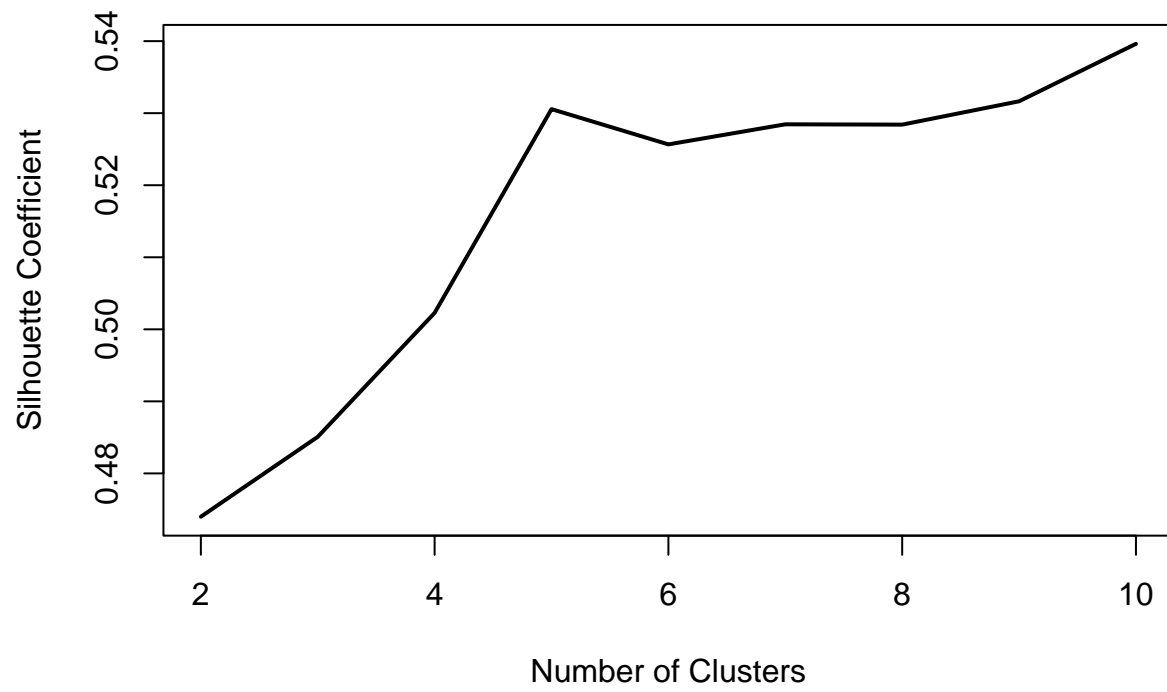
```
## cluster size ave.sil.width
## 1      1      1           0.00
## 2      2 4187           0.39
## 3      3      1           0.00
```

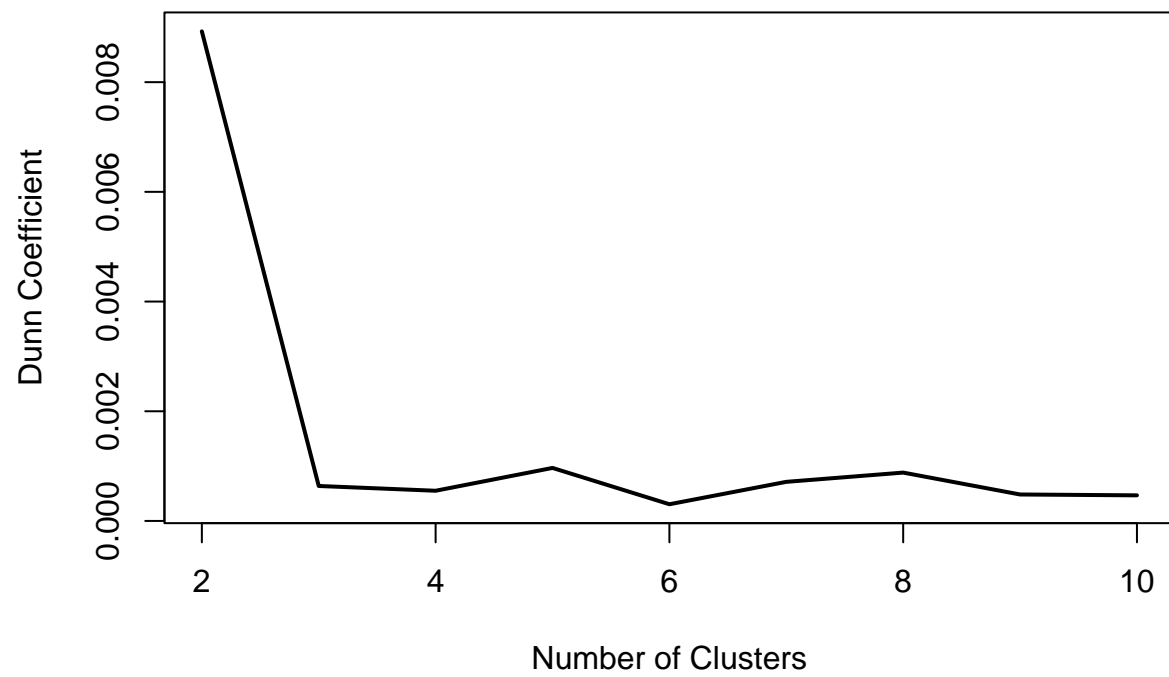


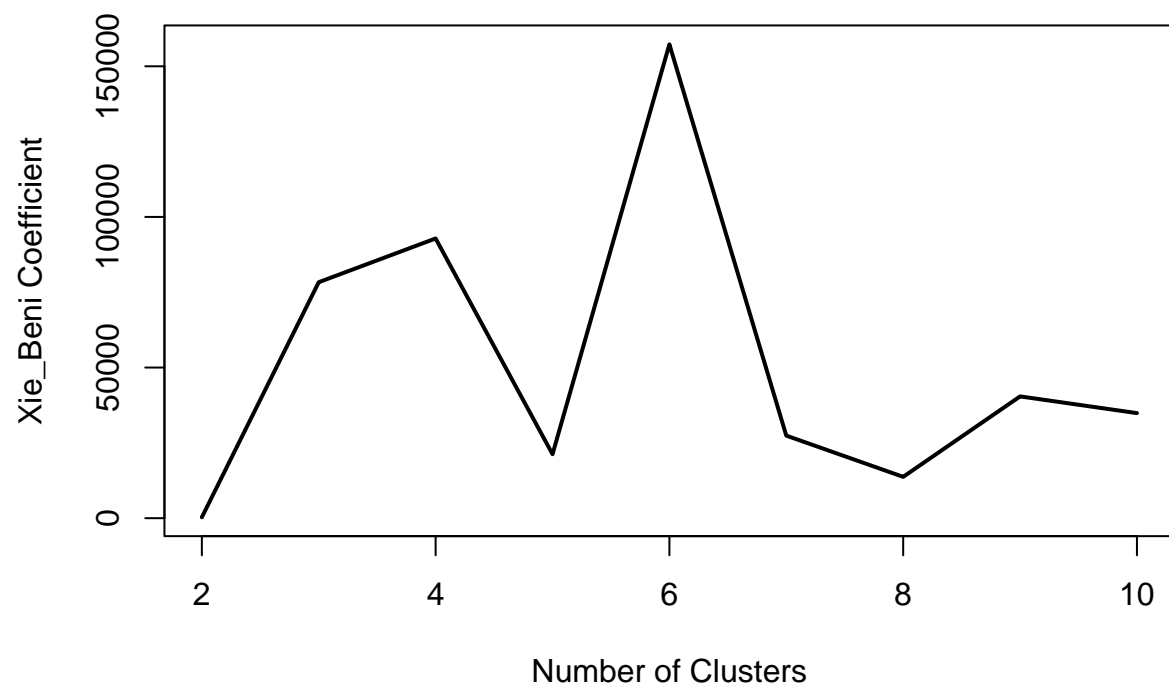
Cluster Rugs with Histogram & Cutoff Line

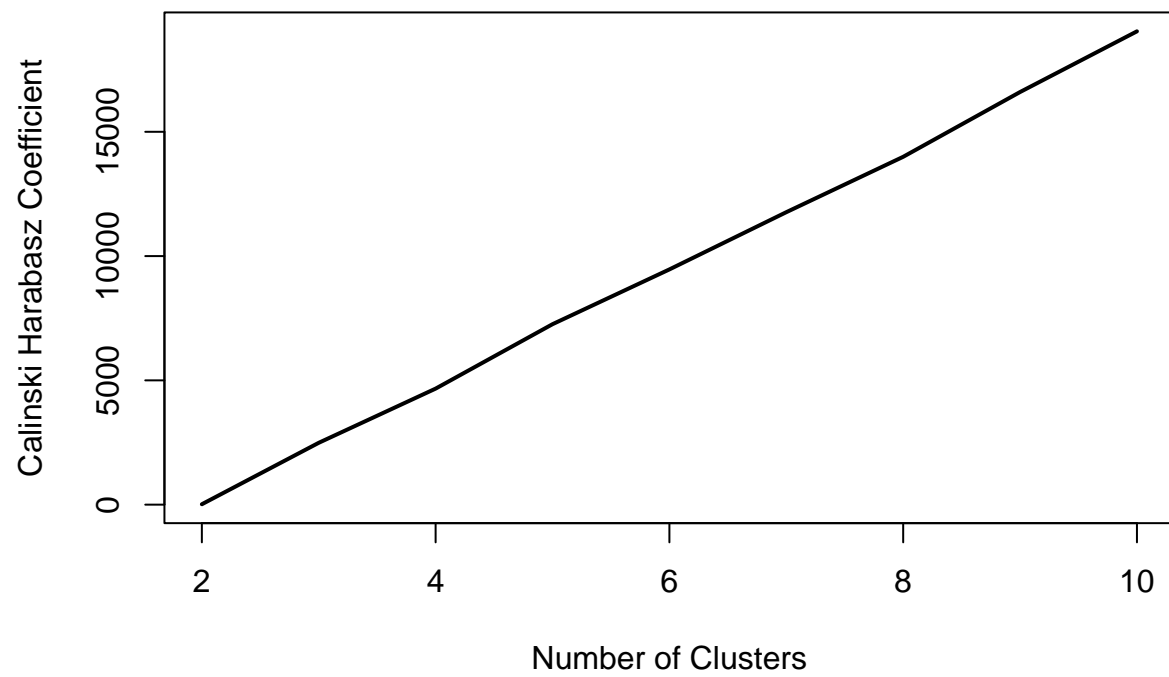


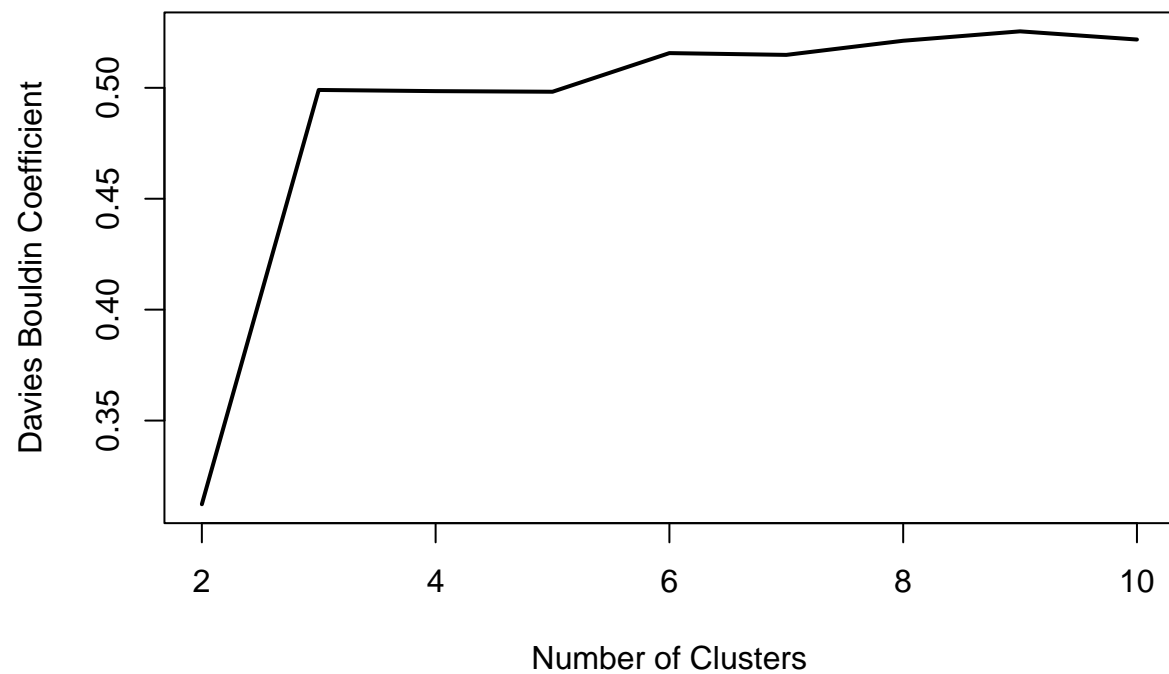
Grocery



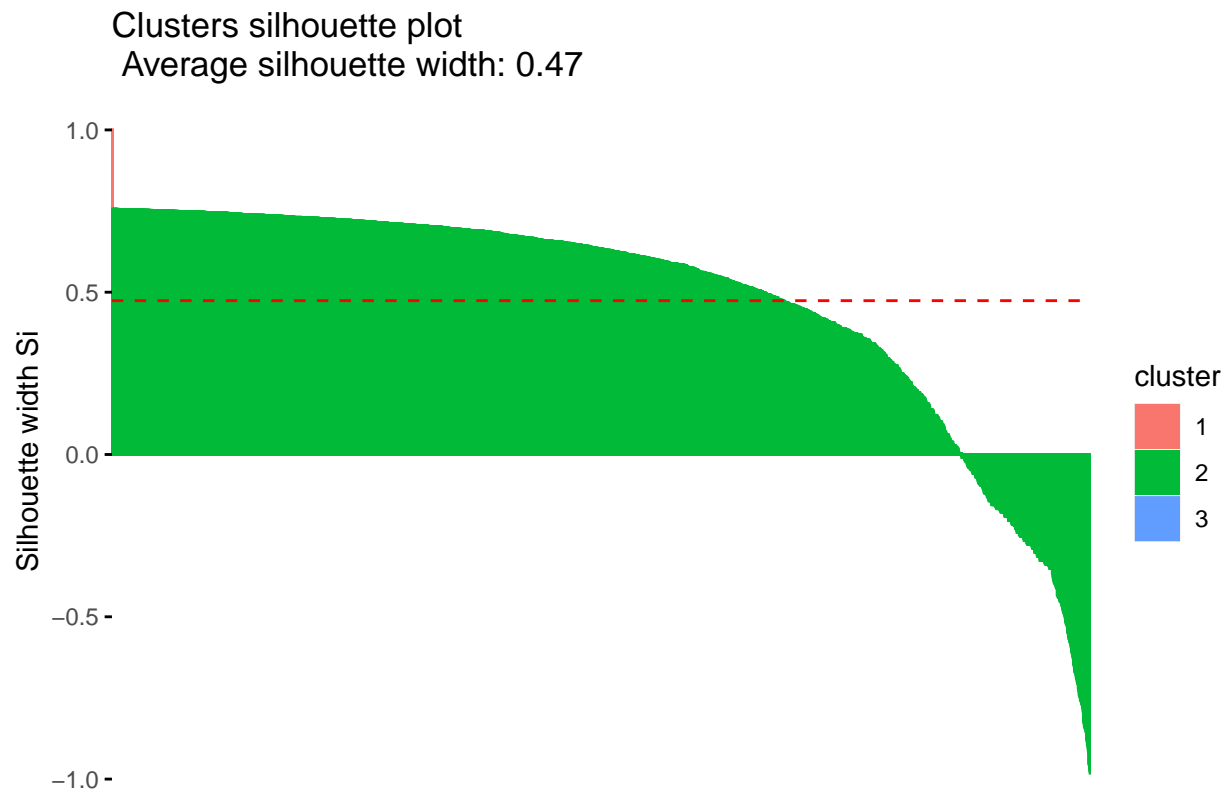




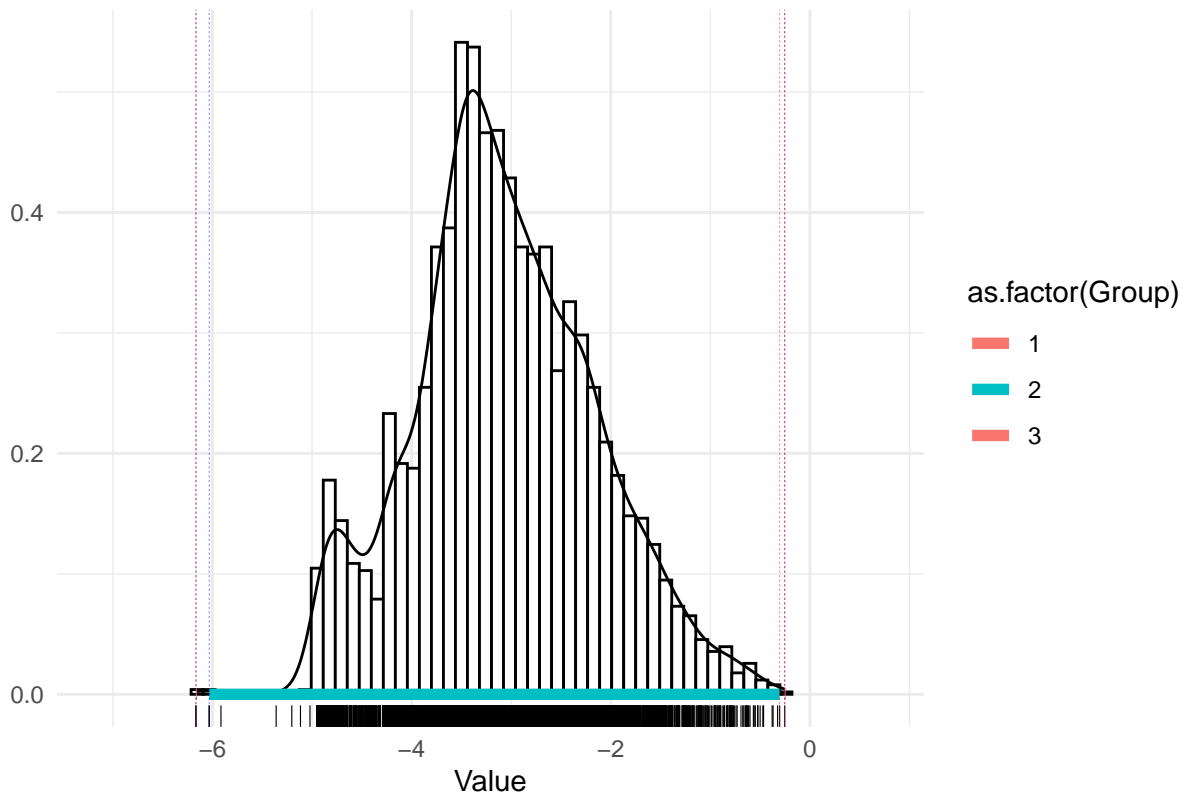




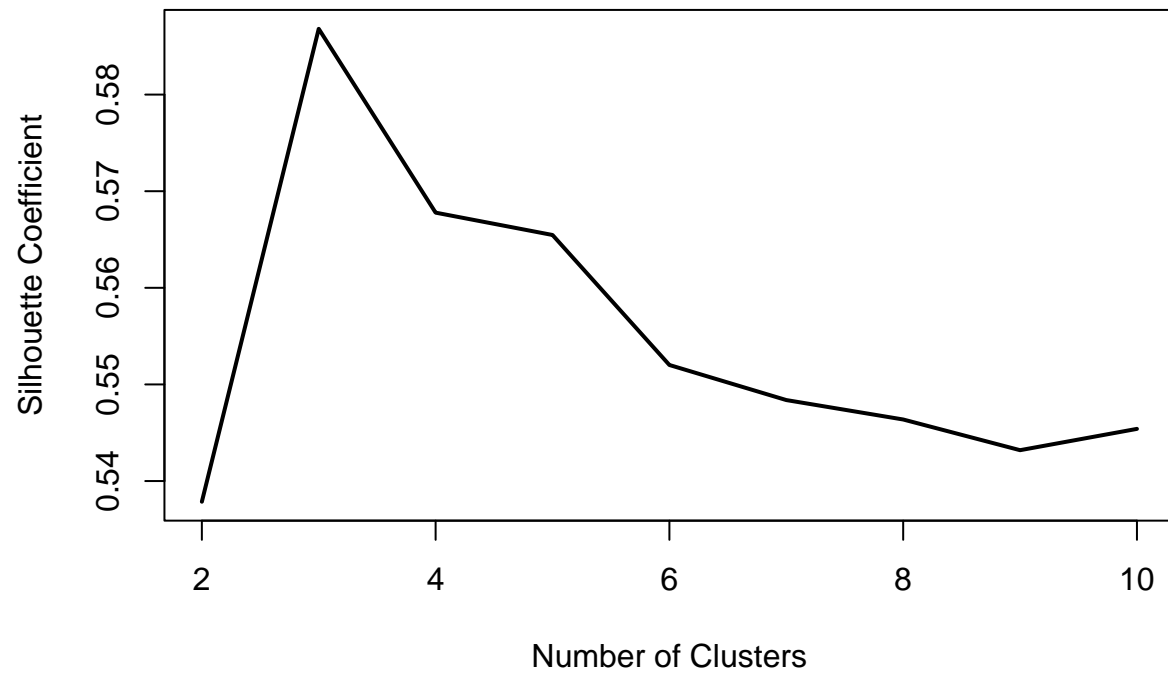
```
## cluster size ave.sil.width
## 1      1      2          1.00
## 2      2 4192          0.47
## 3      3      1          0.00
```

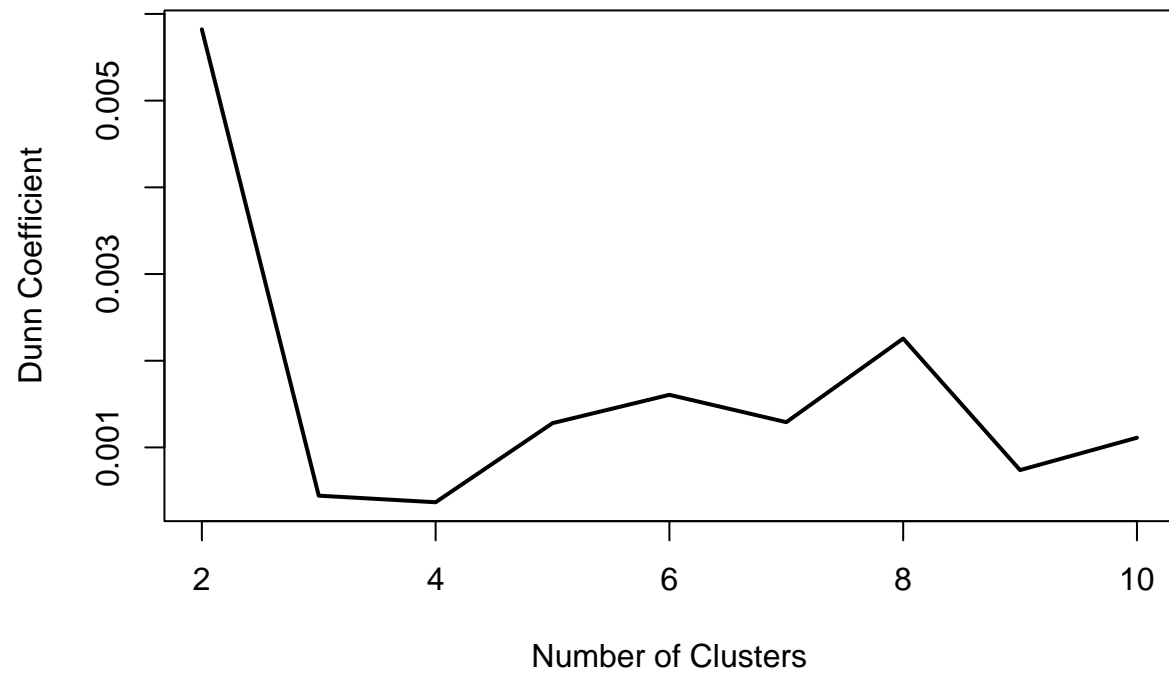


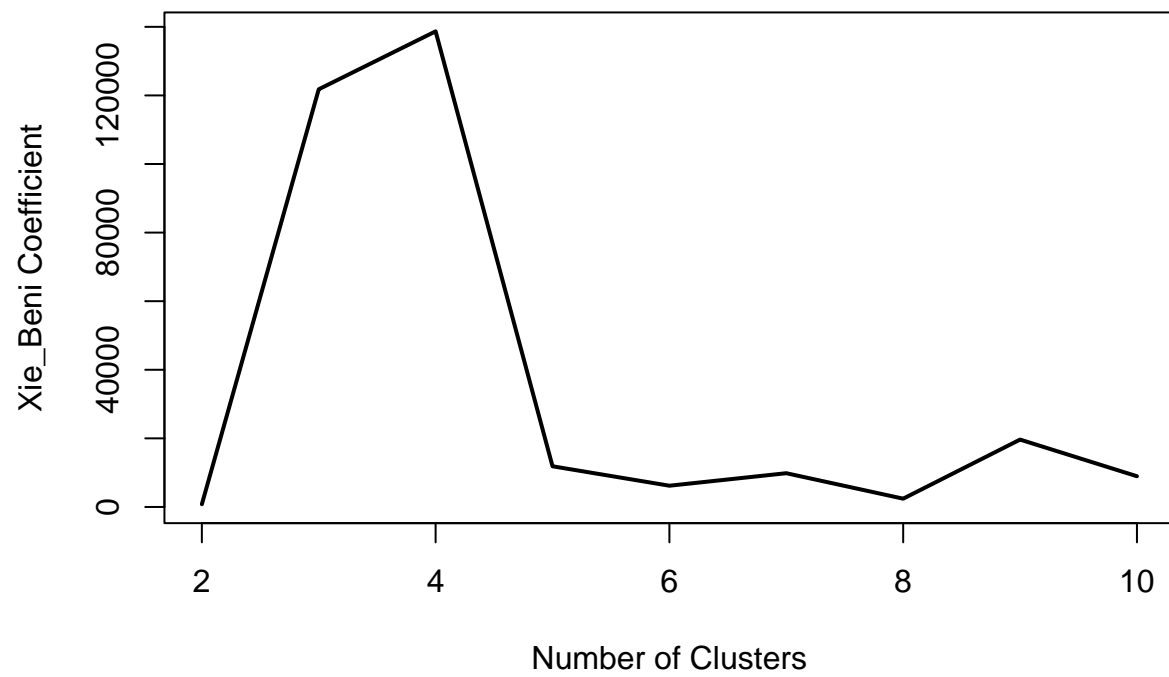
Cluster Rugs with Histogram & Cutoff Line

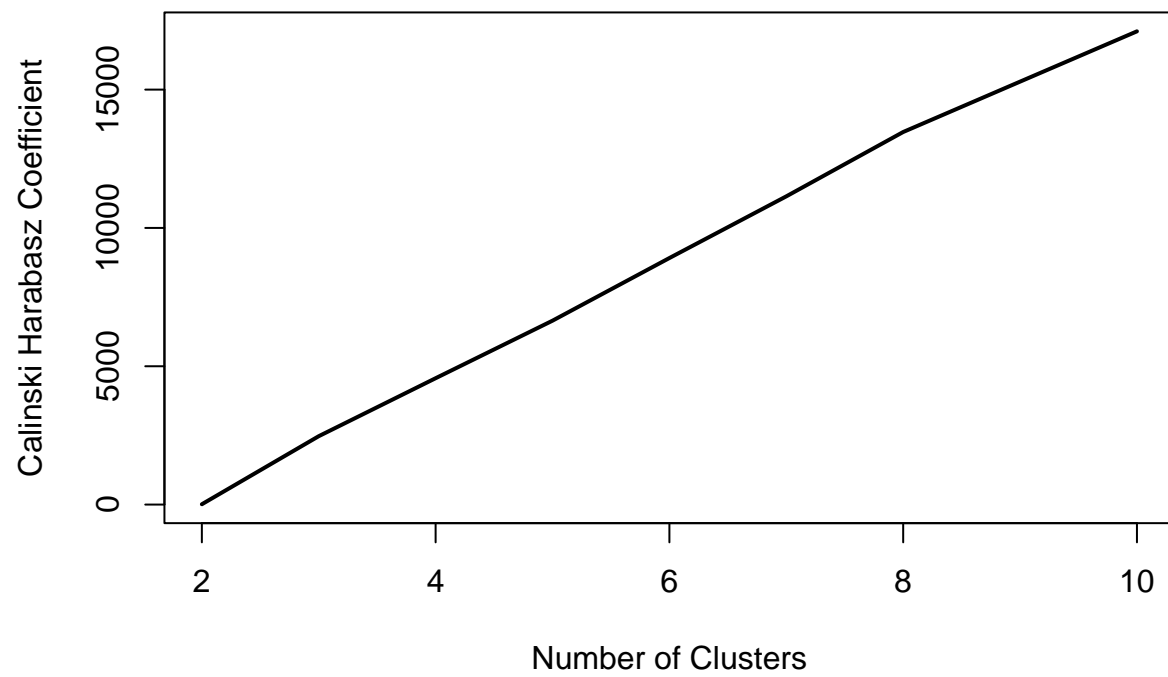


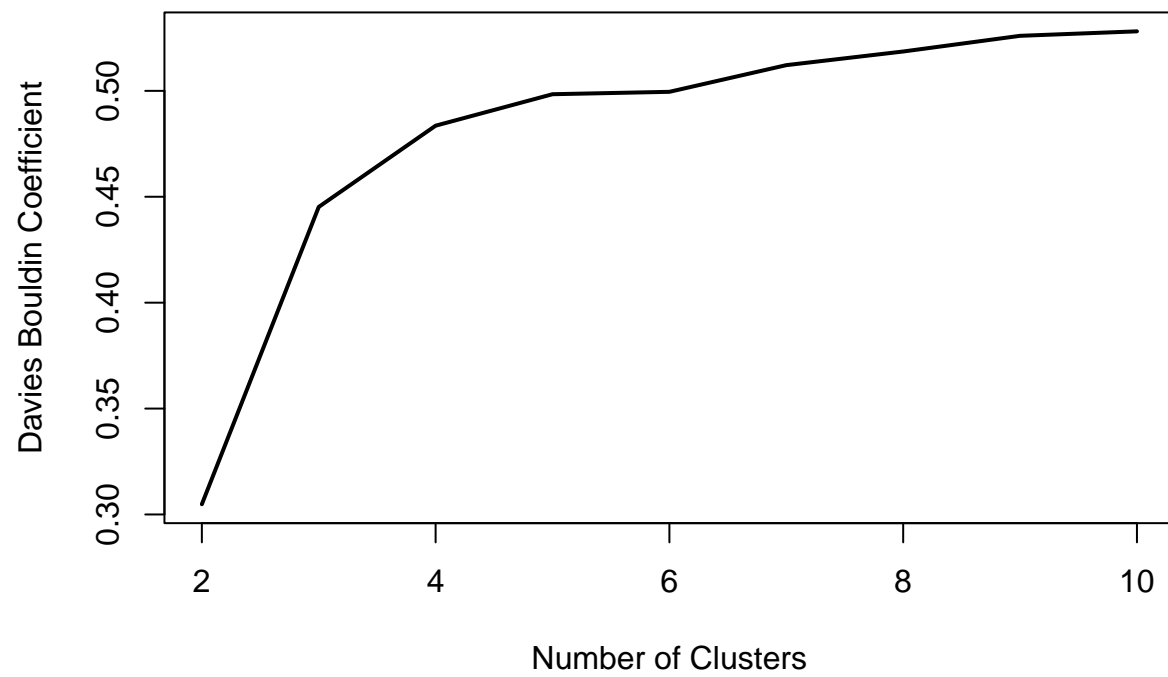
Library



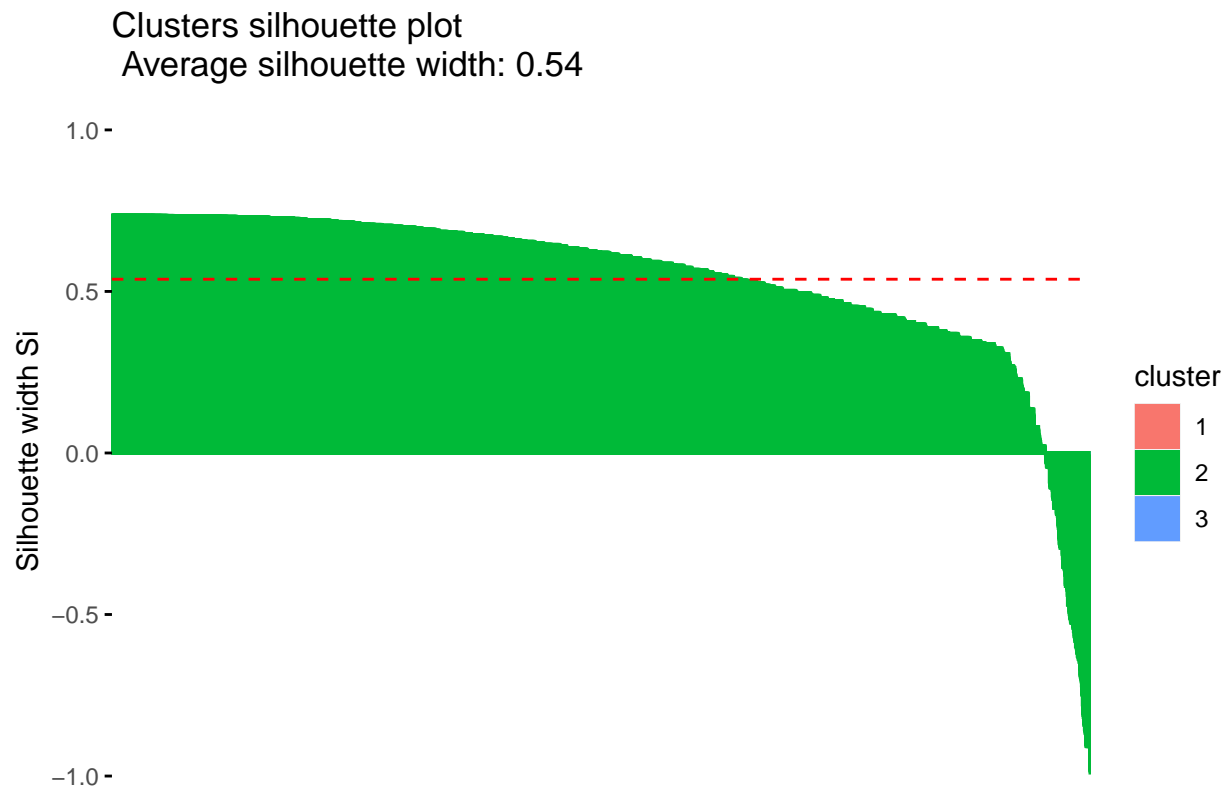








```
## cluster size ave.sil.width
## 1      1      1           0.00
## 2      2 3288           0.54
## 3      3      1           0.00
```



Cluster Rugs with Histogram & Cutoff Line

