EV project - Feynn

Aryan

2024-02-14

EV Market Segmenatation - R

```
data1 <- read.csv('EV_fil.csv')
```

```
data2 <- subset(data1, select = -rating)
```

PCA

You can also embed plots, for example:

```
EV.pca <- prcomp(data2)
summary(EV.pca)
```

```
## Importance of components:
##                           PC1     PC2     PC3     PC4     PC5     PC6
## Standard deviation     3.2829 0.71506 0.57404 0.54038 0.51412 0.48781
## Proportion of Variance 0.8683 0.04119 0.02655 0.02353 0.02129 0.01917
## Cumulative Proportion  0.8683 0.90946 0.93601 0.95953 0.98083 1.00000
```

```
print(EV.pca, digits = 1)
```

```
## Standard deviations (1, .., p=6):
## [1] 3.3 0.7 0.6 0.5 0.5 0.5
##
## Rotation (n x k) = (6 x 6):
##                     PC1   PC2   PC3   PC4   PC5   PC6
## Visual.Appeal       0.4  0.86  0.29 -0.05  0.17 -0.12
## Reliability         0.4 -0.13 -0.05 -0.76 -0.44 -0.06
## Performance         0.4 -0.05 -0.29 -0.05  0.43  0.76
## Service.Experience  0.5 -0.44  0.73  0.24  0.08 -0.02
## Comfort             0.4  0.12 -0.28  0.59 -0.64  0.10
## Value.for.Money     0.4 -0.18 -0.46  0.11  0.42 -0.63
```

```
library("flexclust")
```

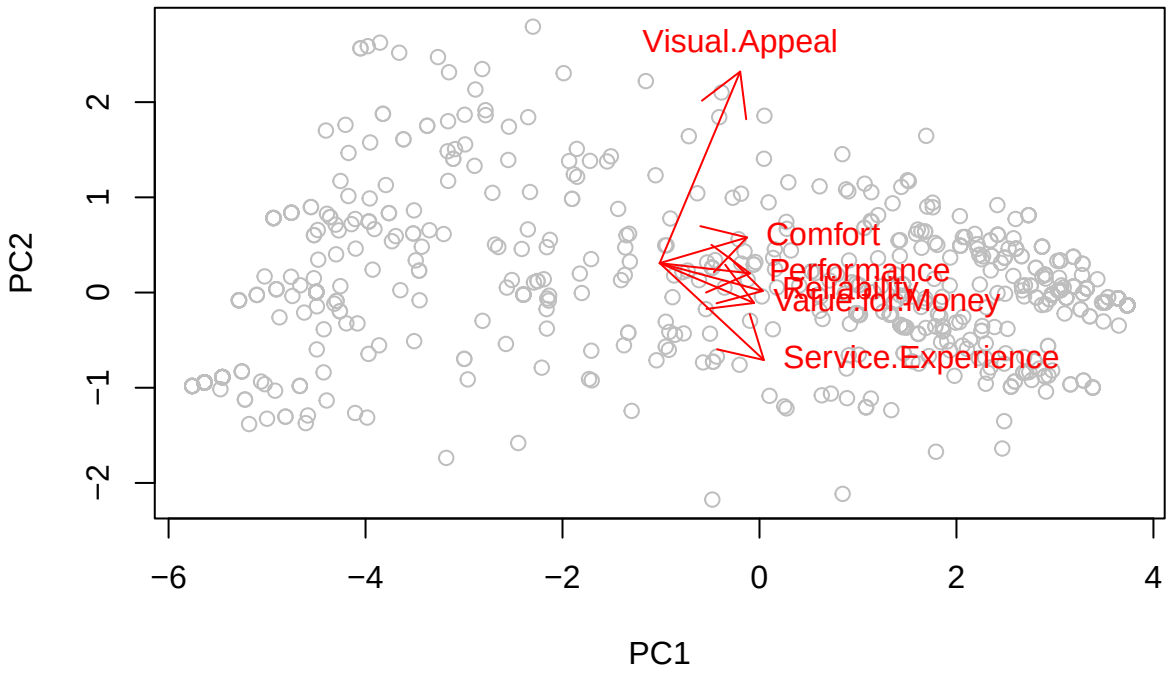
```
## Loading required package: grid
```

```
## Loading required package: lattice
```

```
## Loading required package: modeltools
```

```
## Loading required package: stats4
```

```
plot(predict(EV.pca), col = "grey")
projAxes(EV.pca)
```



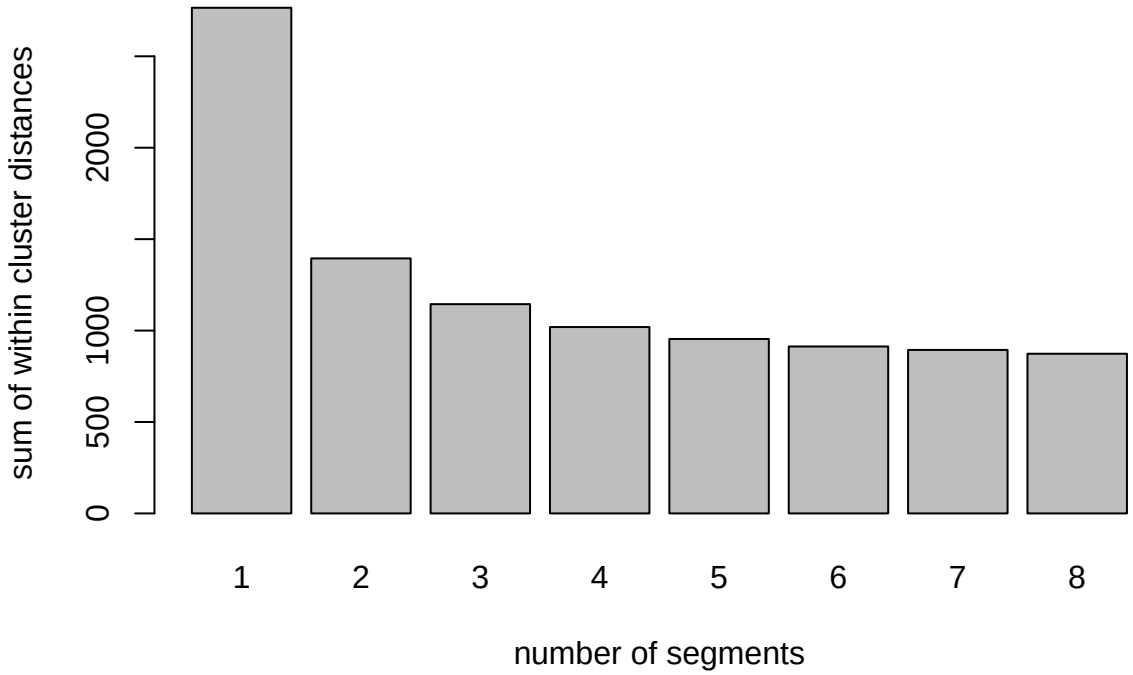
The attribute Visual Appeal plays a key role in the evaluation of two-wheelers in India. The remaining attributes align with what can be interpreted as positive perceptions: as all attricbutes point in the same direction in the perceptual chart.

These initial exploratory insights represent valuable information for segment extraction. Results indicate that all attributes are strongly related to one another, and that the Visual appeal dimension may be critical in differentiating between groups of consumers.

Extracting Segments

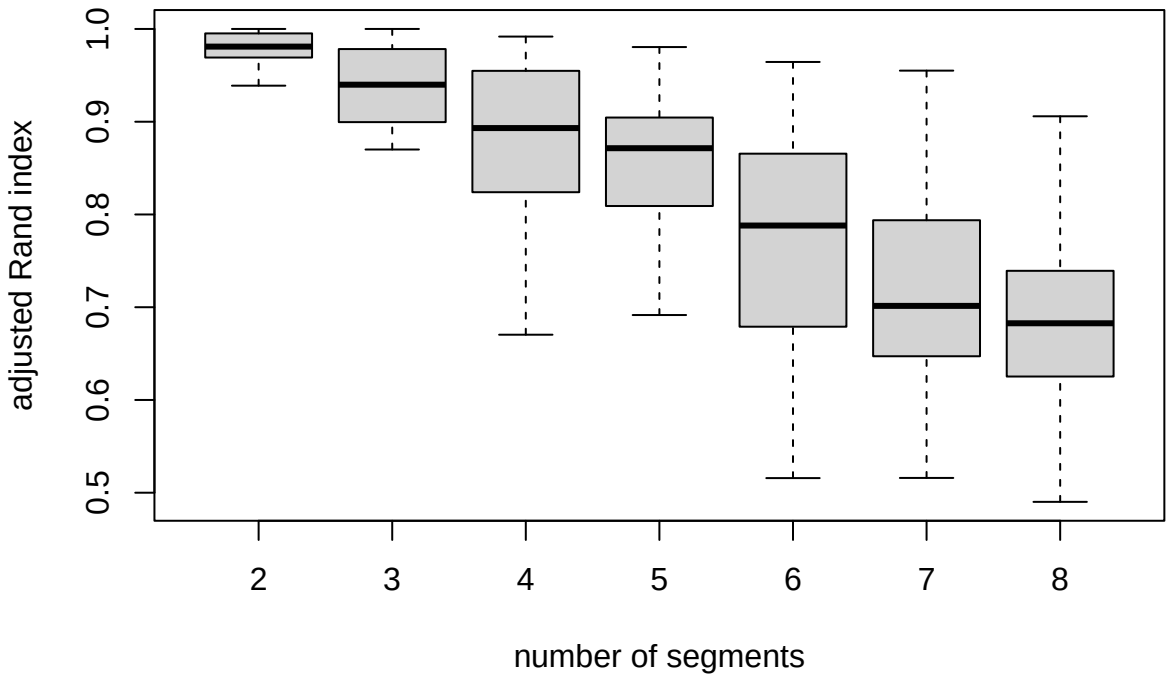
```
set.seed(1234)

EV.km28 <- stepFlexclust(data2, 2:8, nrep = 10, verbose = FALSE)
EV.km28 <- relabel(EV.km28)
```

```
plot(EV.km28, xlab = "number of segments")
```



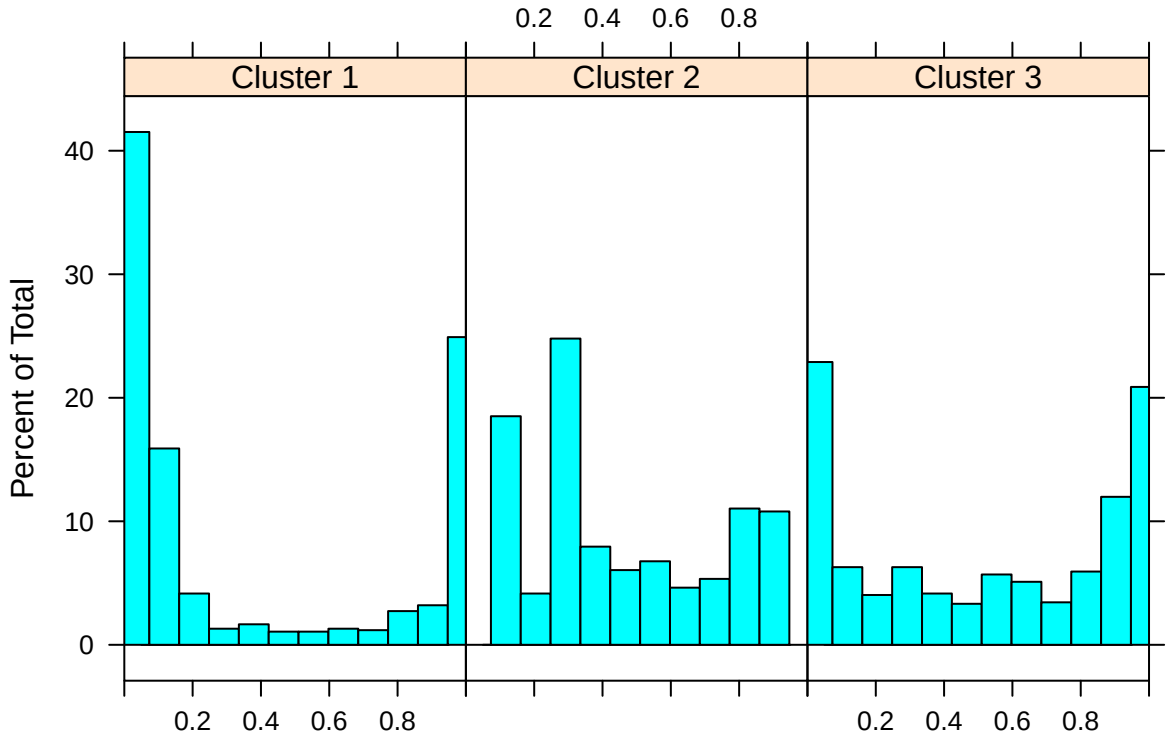

```
set.seed(1234)
EV.b28 <- bootFlexclust(data2, 2:8, nrep = 10, nboot = 100)
```

```
plot(EV.b28, xlab = "number of segments", ylab = "adjusted Rand index")
```



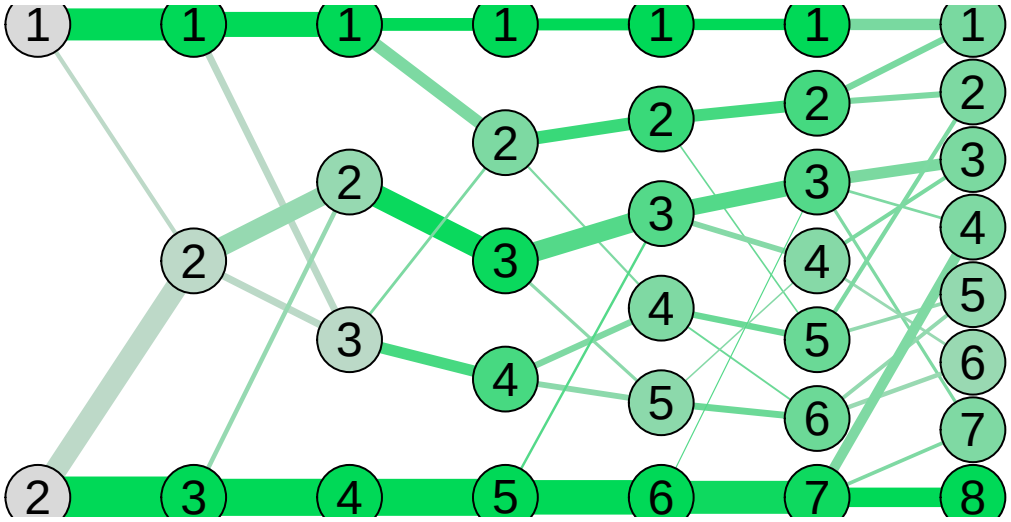
Inspecting the box plot points to the 2, 3, 4 and 5 segment solutions as being quite stable. Despite low number of segments not lacking interesting insights considering the same directional feature PCA plot (Fig 3.1.2) and the elbow method from the scree plot (Fig 3.1.3) the 3-segment solution still marks to be the best solution – the solution with the most mark segments with a high degree of replication. This is further backed by a noticeable drop in replicability in the 4-segment solution.

```
histogram(EV.km28[["3"]], data = data2, xlim = 0:1)
```



None of the segments shown in the gorge plot is well separated from the other segments, and proximity to at least one other segment is present as indicated by the similarity values all being between 0 and 1.

```
slsaplot(EV.km28)
```

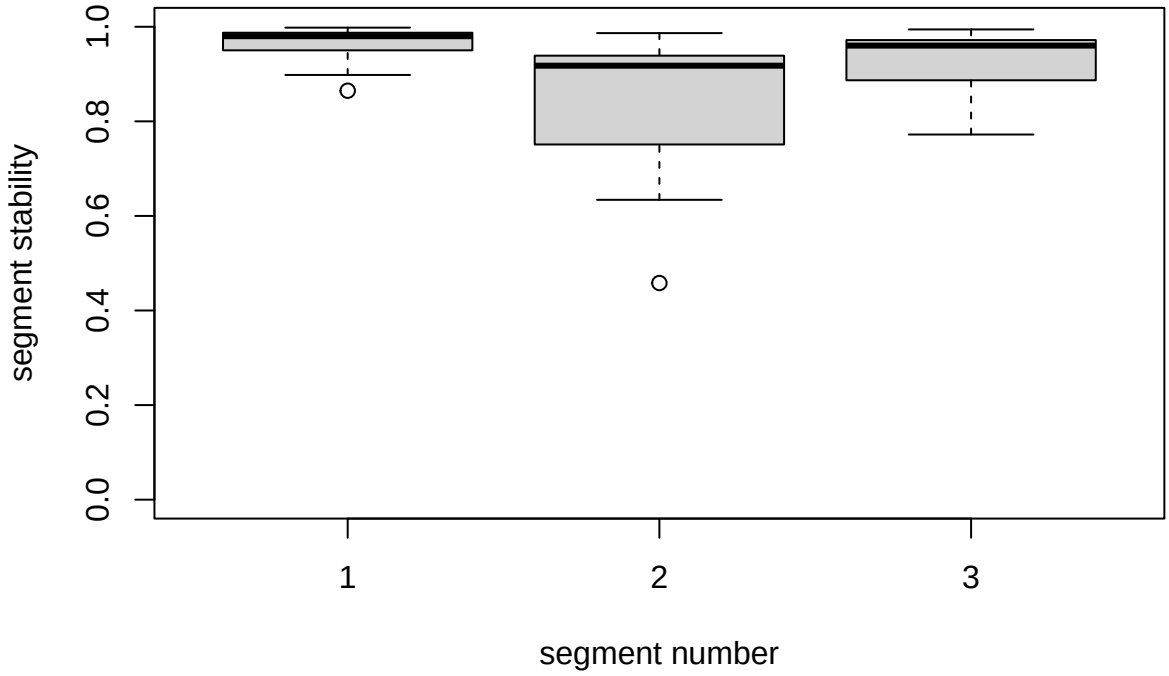


Thick green lines indicate that many members of the segment to the left of the line move across to the segment on the right side of the line. Segment 2 in the two-segment solution (in the far left column of the plot) remains almost unchanged until the seven-segment solution, then it starts losing members. This seems in line with the global stability box plot. However keeping in mind earlier analysis, the SLSA plot in view of the earlier determination that the three-segment solution looks fine and but it seems segments 2 and 3 are nearly identical.

```
EV.k3 <- EV.km28[["3"]]
```

```
EV.r3 <- slswFlexclust(data2, EV.k3)
```

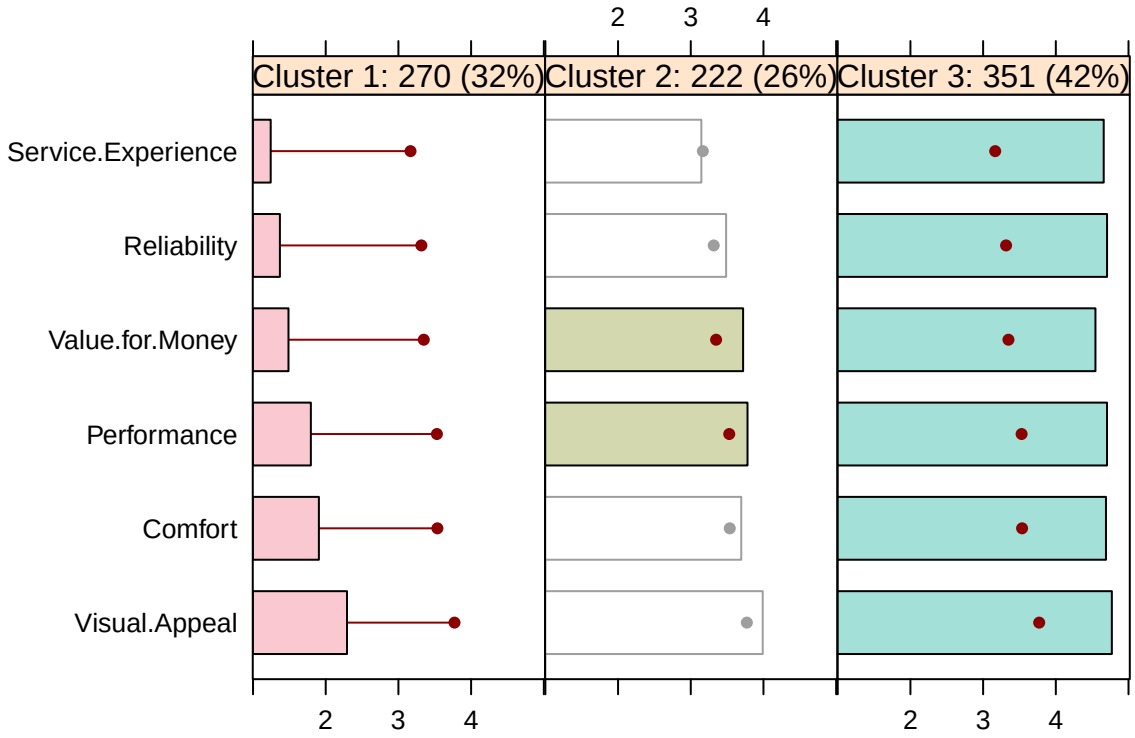
```
plot(EV.r3, ylim = 0:1, xlab = "segment number", ylab = "segment stability")
```



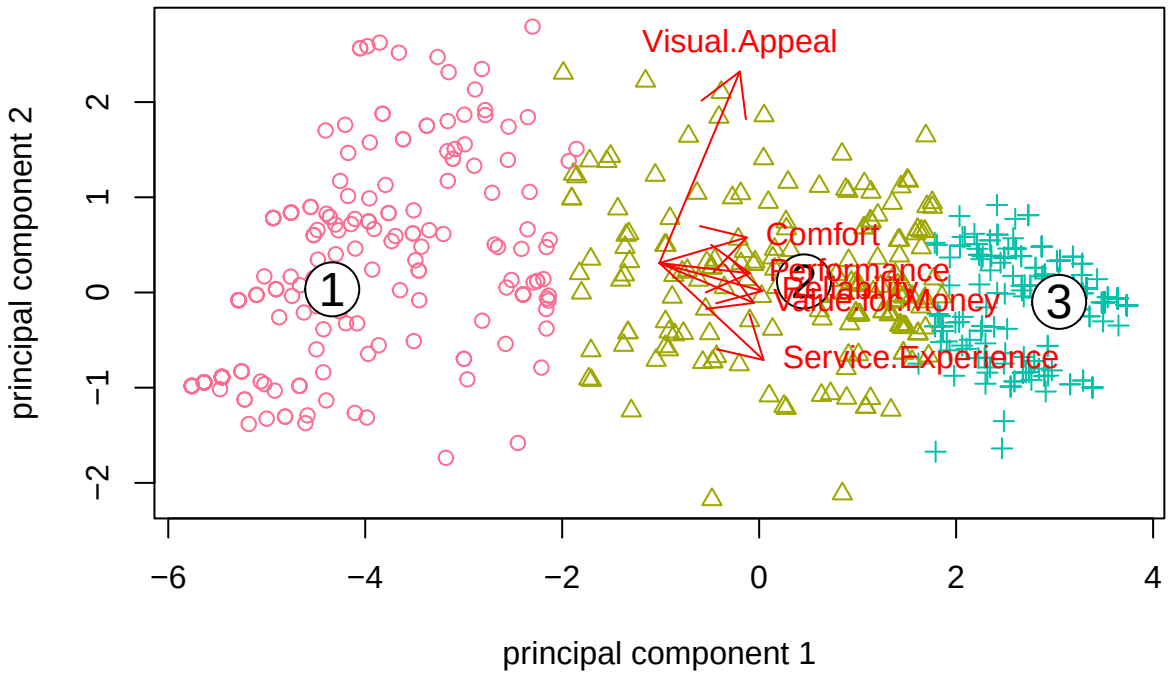
Profiling Segments

```
EV.vclust <- hclust(dist(t(data2)))
```

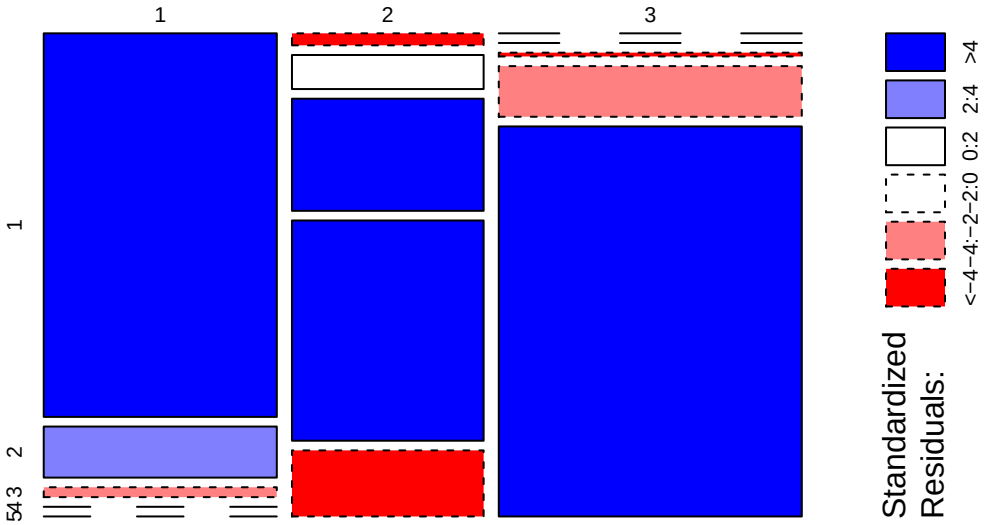
```
barchart(EV.k3, shade = TRUE, which = rev(EV.vclust$order))
```




```
plot(EV.k3, project = EV.pca, data = data2, hull = FALSE, simlines = FALSE, xlab = "principal component

projAxes(EV.pca)
```



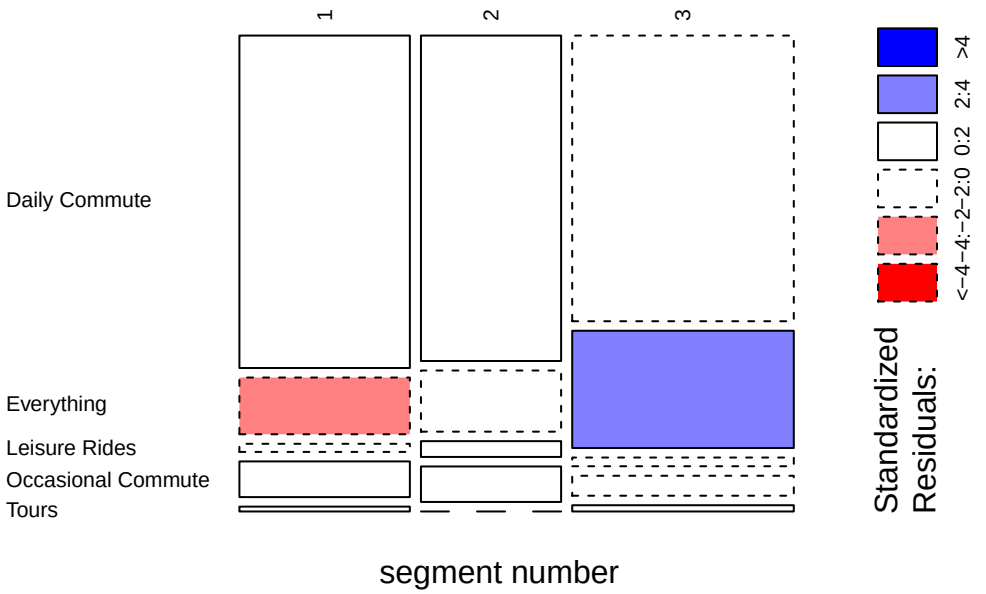

```
k3 <- clusters(EV.k3)
mosaicplot(table(k3, data1$rating), shade = TRUE, main = "", xlab = "segment number")
```



The mosaic plot plots segment number along the x-axis, and the overall user rating their EV two-wheeler along the y-axis. The mosaic plot reveals a strong and significant association between these two variables. Members of segment 1 (depicted in the first column) do not highly rate their EV two-wheelers whatsoever shown by the completely empty boxes bottom left and seldom find them average (rating = 3), as indicated by the bottom left box being coloured in red. In stark contrast, members of segment 3 are significantly more likely to rate an EV two-wheeler highly, as indicated by the dark blue box in the bottom right of the plot. At the same time, these consumers are less likely dislike or find EV bikes average, as indicated by the completely empty boxes and very small red box at the top right of the plot. Members of segment 2 appear to have the moderate feelings towards EV bikes (rating largely 3-4); their generally have an above average liking with a small likelihood to fall into extreme liking or disliking for the product.

```
filtered_data <- read.csv('filtered_data.csv')
```

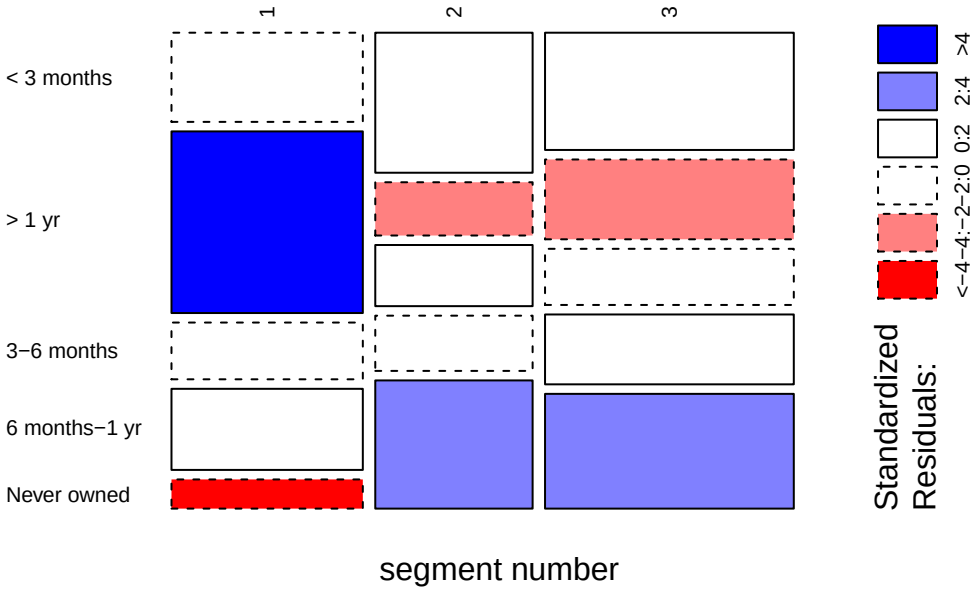
```
mosaicplot(table(k3, filtered_data$Used.it.for), shade = TRUE, main = "", xlab = "segment number", las =
```



The descriptor variable (Used it for) is plotted along the y-axis. This mosaic plot offers the following additional insights about our market segments: segment 1 and segment 2 have a similar use case distribution with an

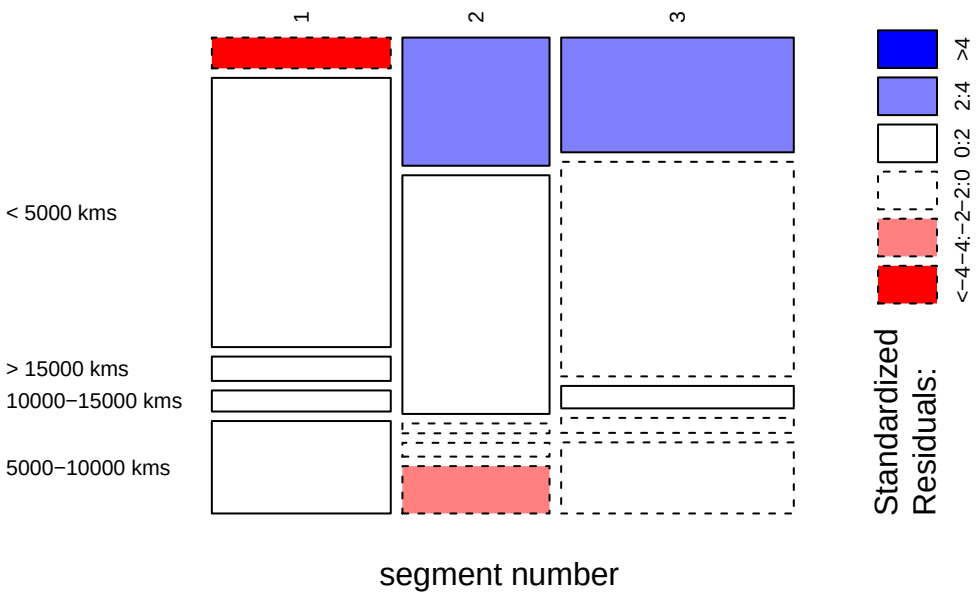
emphasis on daily commute. Segment 1 has a very low likelihood to use EV bikes for everything. Segment 3 contains more individuals who use EV bikes for everything and have a higher tendency to use their EV bikes for everything rather than a specific use case.

```
mosaicplot(table(k3, filtered_data$Owned.for), shade = TRUE, main = "", xlab = "segment number", las = 1
```



In Fig 3.1.13, the descriptor variable (Owned for) is plotted along the y-axis. Segment 2 and segment 3 have a similar distribution with regards to their length of ownership. In both these segments, users tend to have a greater likelihood to never own the vehicle. Segment 1, the haters, contains significantly more individuals owning the product for a length greater than 1 year. (as depicted by the larger blue box for the >1 yr category).

```
mosaicplot(table(k3, filtered_data$Ridden.for), shade = TRUE, main = "", xlab = "segment number", las =
```



Never Owned - high for 1 & 2, low for 3

Market segments are plotted along the x-axis. The descriptor variable (Ridden for) is plotted along the y-axis.

```r
library("partykit")
tree <- ctree( + factor(k3 == 3) ~ rating + Used.it.for + Owned.for + Ridden.for, data = filtered_data)

plot(tree)
```