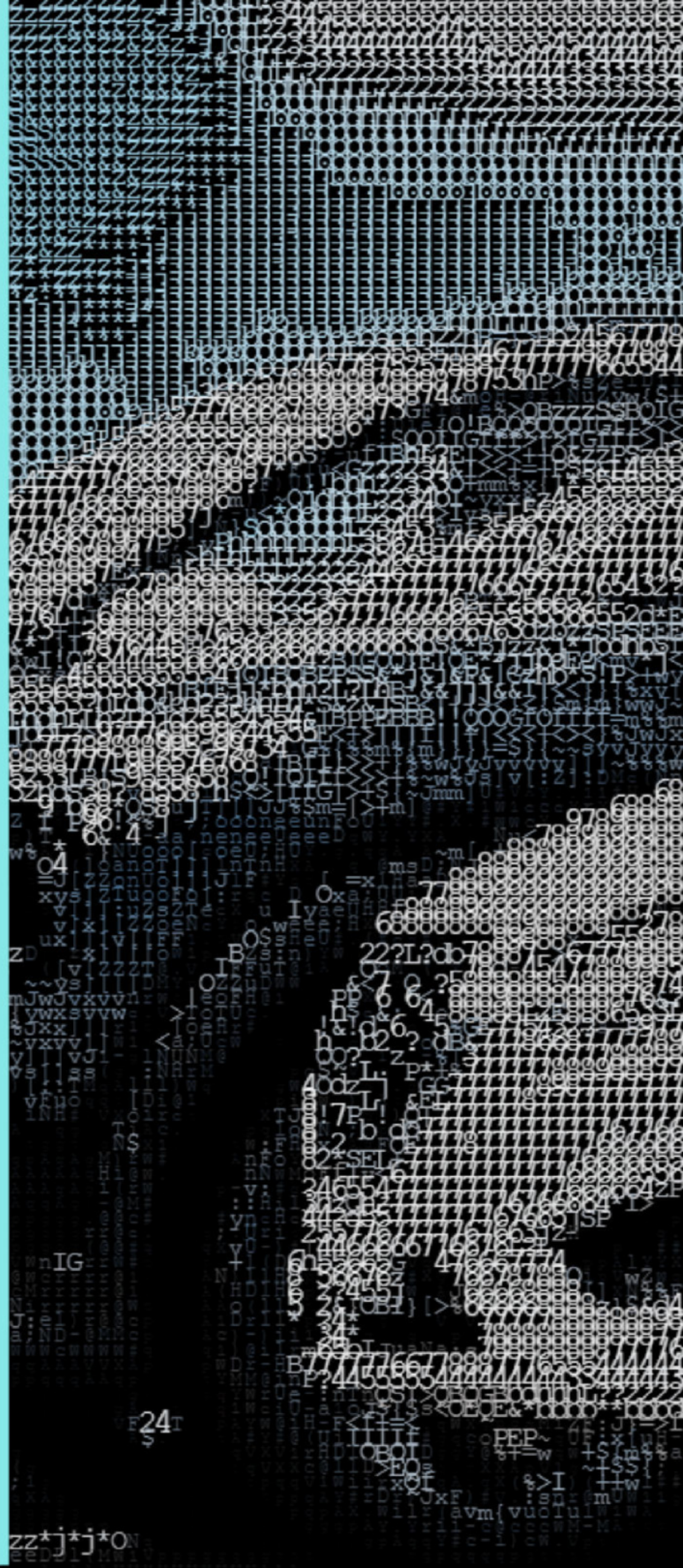
Consumer Automobile Preferences

GEORGE DREEMER



Research Process

The iterative process taking us from raw data to insights.

1 Our Dataset

2 Hierarchical Clustering Process

3 Number of Clusters Selection

4 Preliminary to Final Hierarchical Solution

5 Segment preferences and the Naming Convention

6 Who are the people behind the segment?

7 Non-Hierarchical Clustering Comparison 1/2

8 Non-Hierarchical Clustering Comparison 2/2

9 Market Potential of Environmentally-friendly cars

Our Dataset

/descriptive statistics

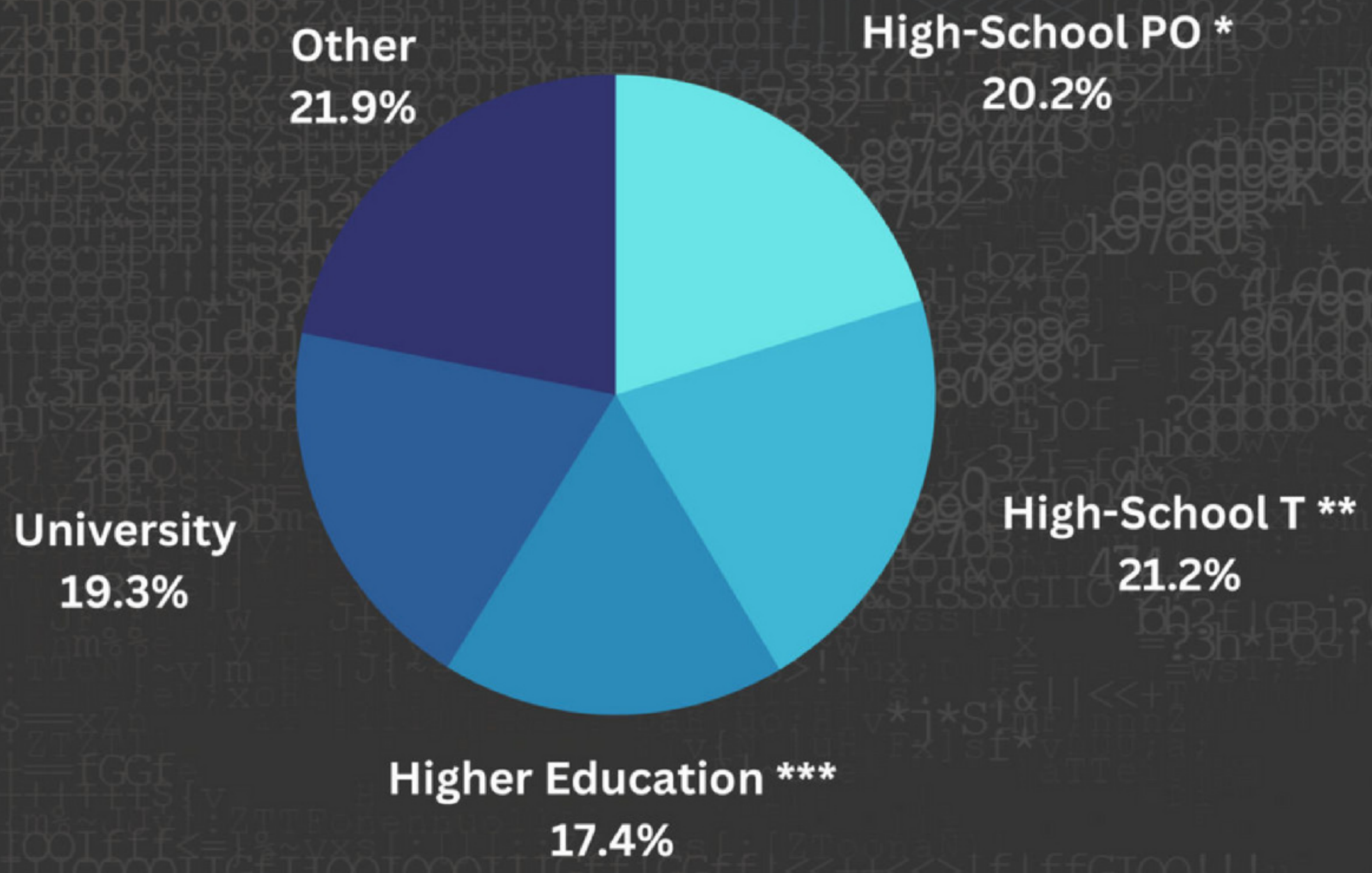
420 participants

Male Female

48% 52%

Median Age: 48

Education



* Profession-oriented ** Theory-based *** Non-university

6 Automobile Preferences

- Mileage/Range
- Comfort
- Design
- Power
- In-car Entertainment
- Environment-friendly

No profound correlations found between variables.

See Appendix A-1

Hierarchical Clustering

/process

We used *CPCC* to choose which *linkage method* has the best fit

We used *Average & Ward's linkage* for the steps below to find the optimal number of clusters

Linkage Methods → **Dendrograms** → **Agglomeration Schedules** → **Scree Plots** → **Robustness Check** → **Preliminary Cluster Solutions** → **Do clusters differ on their preferences?** → **Final Cluster Solution**

We compared our Avg. & Ward cluster solutions to Random k-means to see which solution, at what number of clusters is robust

We used ANOVA and Tukey to check if clusters' preferences differ significantly from each other and chose our final hierarchical cluster solution.

Selecting the number of clusters

/preliminary decision process

Linkage Methods

We assessed the CPCC values for each linkage method and found the highest two were for Average (0.54) & Ward's linkage (0.50).

Dendrograms

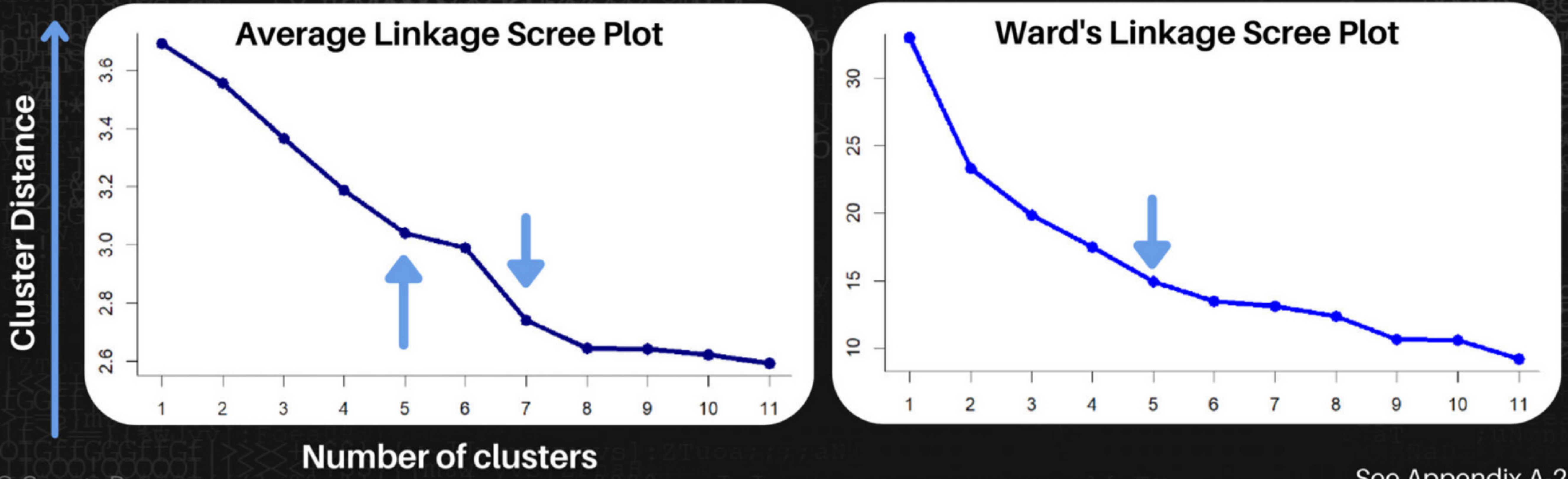
We visually inspected dendrograms from the two aforementioned methods and *perliminarily estimated 5-7 clusters could be optimal.*

Agglomeration Schedules

The values in the agglomeration schedules of the two methods both pointed towards 5 clusters, but Ward's was clearer in that regard.

Scree Plots

We look for the point after which there is a jump in the y-axis value.



See Appendix A-2:5

Hierarchical Clustering

/preliminary to final solution

Comparing Avg. & Ward to Random K-means

Comparing our Ward and Avg. solutions to Random K-means, we found robustness in *Ward with 6 clusters* & *Average with 5 clusters.*

Do clusters differ on their preferences?

ANOVAs & Tukey tests

Next, we examined if the preferences between clusters differ significantly from each other. We found the Ward's clusters to have too much overlap between preferences. On the other hand Average with 5-clusters had better defined and diverse clusters. This can be seen bellow.

Preference	Ward 6-Clusters Preference Overlap*	Average 5-Clusters Preference Overlap*
Mileage/Range	5-1, 4-2, 6-3, 5-4	5-2, 5-4
Power	3-1, 5-2	2-1, 3-2
Design	5-3, 4-1	4-1, 5-2
Comfort	6-3, 3-2	4-2
Environment-friendly	5-3, 4-1	4-1, 3-2, 5-2
In-car Entertainment	1-5, 2-3, 4-2, 4-3, 6-3	4-3, 5-1

The non-significant differences between clusters are reflected in the naming in the next slide.

Final HC Solution

Average linkage with 5 clusters

Legend: Mileage/Range, Power, Design, Comfort, In-car Entertainment, Environment-friendly

* e.g: 5-1 means the 5th and 1st clusters do not significantly differ on that particular preference

See Appendix A-6:7

What are the preferences of the segments?

/Naming Convention + Preferences per Segment

1-2 Adjectives + **1 Noun**

Adj. 1 = most important preference
Adj. 2 = second most (if there are two)
Could reflect all preferences but mileage.

+

Noun = level of preference for mileage
Relative Levels: Pilgrims (highest) > Riders (central score) > Cruisers (lowest)

Note: The only exception to the rule are the "Daily Riders" who have central or low preferences for most attributes.

Daily Riders
- #1 Comfort
- #2 Power
- Central Mileage

Doesn't care about anything except **Comfort** and a little **Power**.

Entertained Eco-Pilgrims
- #1 Entertainment
- #2 Environment
- High Mileage

Cares most of all other groups about **In-car Entertainment.**

Designer Eco-Cruisers
- #1 Design
- #2 Environment
- Low Mileage

Cares most of all other groups about a fashionable car **Design.**

Power Pilgrims
- #1 Power
- High Mileage

Cares most of all other groups about **Power** and beside mileage has no other positive preference.

Comfort Eco-Pilgrims
- #1 Comfort
- #2 Environment
- High Mileage

Cares most of all other groups about **Comfort.**

See Appendix A-8

Who are the people behind the segments?

/5 Consumer Segments for Auto-preferences

Daily Riders

33% of sample

33% | 67%

59-68 | 45%
68 & older | 26%

Theory HS | 33%
Other | 18%

Urban | 33%
Suburban | 28%

Entertained Eco-Pilgrims

25% of sample

55% | 45%

18-28 | 56%
29-38 | 26%

University | 27%
Other | 22%

Countryside | 45%
Suburban | 35%

Designer Eco-Cruisers

27% of sample

79% | 21%

39-48 | 36%
29-38 | 22%

Higher Edu | 25%
Other | 25%

Metropolitan | 65%
Urban | 29%

Power Pilgrims

10% of sample

100%

29-38 | 28%
39-48 | 28%

Prof. HS | 73%
Theory HS | 23%

Countryside | 48%
Urban | 25%

Comfort Eco-Pilgrims

5% of sample

100%

49-58 | 33%
68 & older | 33%

Other | 63%
University | 37%

Countryside | 71%
Suburban | 29%

Mainstream in Sample — Niche in Sample

See Appendix A-9

Can we reproduce this result?

/Non-hierarchical Clustering /Combined K-means

Short answer: generally yes! *(preference-wise)*

Daily Riders	Entertained Eco-Pilgrims	Designer Eco-Cruisers	Power Pilgrims	Comfort Eco-Pilgrims
#1 Comfort	#1 Entertainment	#1 Design	#1 Power	#1 Comfort
#2 Power	#2 Environment	#2 Environment		#2 Environment
Central Mileage	High Mileage	Low Mileage	High Mileage	High Mileage
Slightly higher score for **Comfort.**	Lower score for **Mileage.**	Slightly higher score for **Design** & **Entertainment.**	Slightly lower score for **Mileage** & **Power.**	Lower score for **Mileage** & **Environment.**

There are no score differences that are low/high enough to affect the top 2 preferences of segments and therefore the overall characteristics of segments remain.

In the next slide we can see that the segments do begin to differ in demographics.

See Appendix A-10

Who are the people behind the segments?

/5 Consumer Segments for Auto-preferences

Daily Riders

25% of sample

25% | 75%

59-68 | 50%
68 & older | 30%

Theory HS | 33%
All others similar %

Urban | 36%
Suburban | 26%
Metropolitan | 25%

Entertained Eco-Pilgrims

21% of sample

43% | 57%

18-28 | 66%
29-38 | 26%

University | 27%
All others similar %

Countryside | 42%
Suburban | 26%

Designer Eco-Cruisers

22% of sample

83% | 17%

39-48 | 45%
29-38 | 27%

Other | 30%
All others similar %

Metropolitan | 71%
Urban | 17%

Power Pilgrims

16% of sample

82% | 18%

39-48 | 27%
49-58 | 22%

Prof. HS | 64%
Theory HS | 27%

Countryside | 54%
Suburban | 22%

Comfort Eco-Pilgrims

16% of sample

96% | 4%

49-58 | 34%
59-68 | 30%

Other | 34%
University | 31%
Higher edu | 27%

Countryside | 37%
Suburban | 32%

We can see the respondents have been distributed differently than with hierarchical clustering.

See Appendix A-11

Market potential of Eco-friendly cars

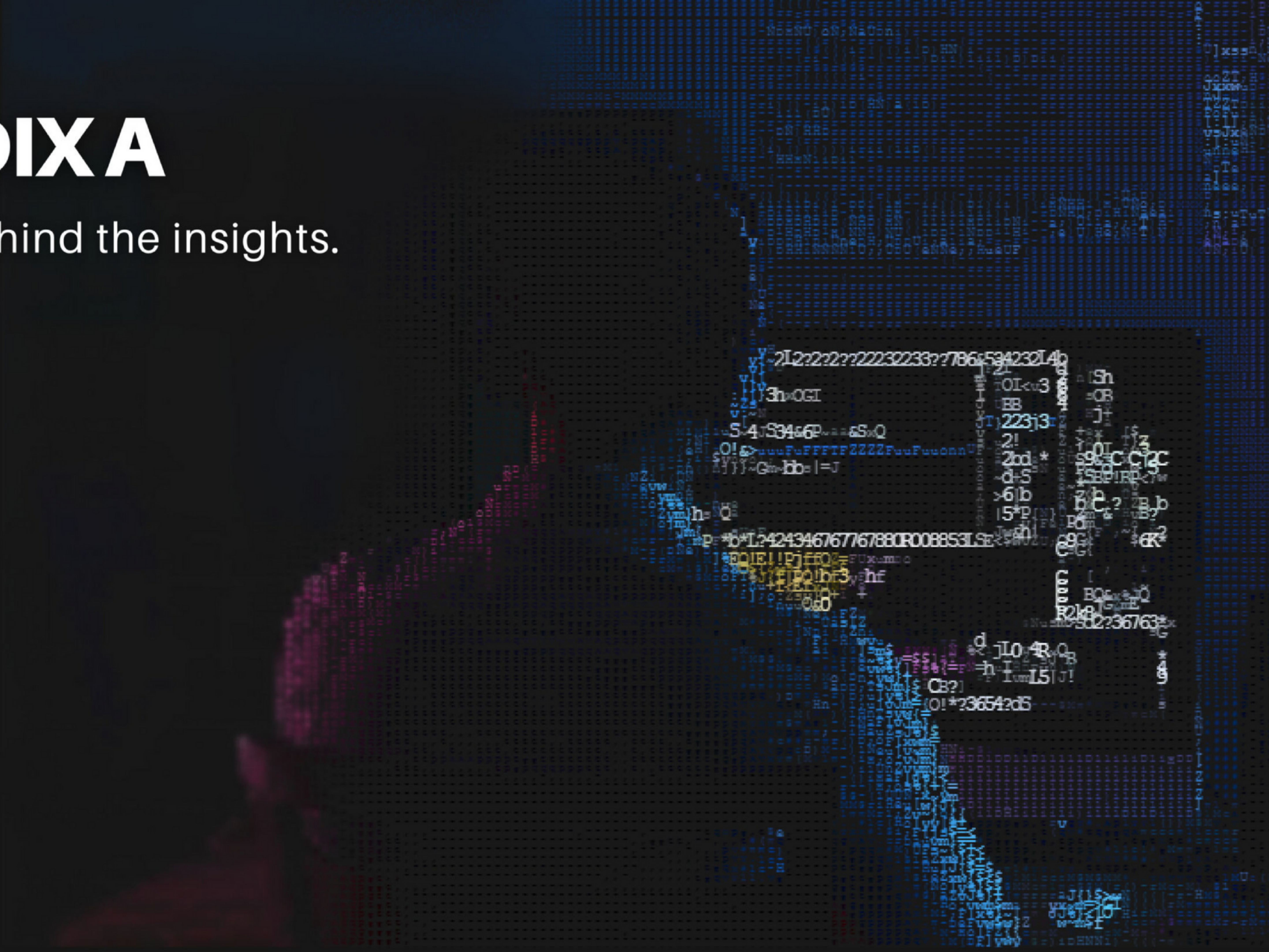
/based on our two clustering solutions

HC Clustering Solution	**Entertained Eco-Pilgrims** 25% of sample	**Designer Eco-Cruisers** 27% of sample	**Comfort Eco-Pilgrims** 5% of sample	= 57%
NHC Clustering Solution	**Entertained Eco-Pilgrims** 21% of sample	**Designer Eco-Cruisers** 22% of sample	**Comfort Eco-Pilgrims** 16% of sample	= 59%

Considering nearly 60% of all respondents according to both clustering solution have shown a preference for environmentally-friendly vehicles it is plausible to think eco-friendly cars have a great market potential. Furthermore, it is important to note the age groups of these segments vary, so not only younger generations (EEP) but also middle and older generations (DEC, CEP) show a preference. Finally, the groups are diverse in terms of area of living as well as their education level.

APPENDIX A

The raw data behind the insights.



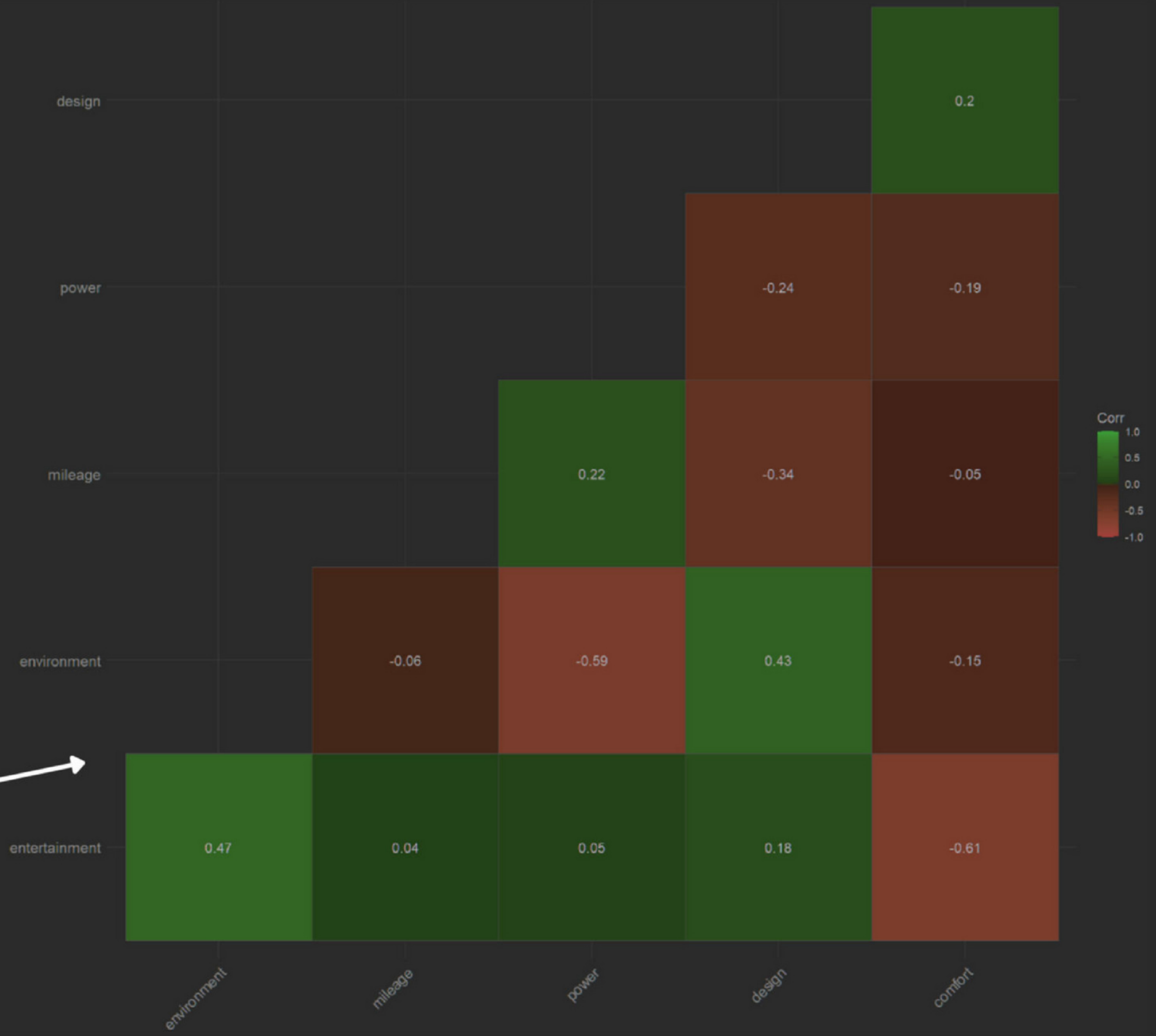
APPENDIX A-1

```r
## Descriptives ----
### Means, Medians, etc. ----
summary(herald.df)
str(herald.df)
head(herald.df)
#median age 48
```

```
      age
Min.   :18.00
1st Qu.:33.00
Median :48.00
Mean   :47.53
3rd Qu.:62.00
Max.   :75.00
```

```r
### Counting ----
ldply(herald.df, function(c) sum(c =="1"))
ldply(herald.df, function(c) sum(c =="2"))
ldply(herald.df, function(c) sum(c =="3"))
ldply(herald.df, function(c) sum(c =="4"))
#217 females, 203 males
#edu: 85,89,73,81,92
```

```r
### Check Correlation ----
h.cor <- cor(herald.df_s[,7:12])
ggcorrplot(h.cor,
           hc.order = TRUE,
           type = "lower",
           lab = TRUE,
           colors = c("red", "white", "green"))
```



APPENDIX A-2.1

```r
#### Single Linkage / Nearest Neighbor ----
herald.hc.single <- hclust(herald.daisy, method="single")
plot(herald.hc.single)
herald.agglo.single <- cbind(as.data.frame(herald.hc.single[1]),as.data.frame(herald.hc.single[2]))
herald.agglo.single
#goodness-of-fit metric for a hierarchical cluster solution - cophenetic correlation coefficient (CPCC)
#assesses how well a dendrogram matches the true distance metric (daisy)
#CPCC > 0.7 indicates a relatively strong fit (meaning hierarchical tree represents the distances between
cor(cophenetic(herald.hc.single),herald.daisy)
#CPCC = 0.3123886 => not a good fit

#### Complete Linkage ----
herald.hc.complete <- hclust(herald.daisy, method="complete")
plot(herald.hc.complete)
herald.agglo.complete <- cbind(as.data.frame(herald.hc.complete[1]),as.data.frame(herald.hc.complete[2]))
herald.agglo.complete
cor(cophenetic(herald.hc.complete),herald.daisy)
#CPCC = 0.478819 => better than Single Linkage

#### Average Linkage ----
herald.hc.avg <- hclust(herald.daisy, method="average")
plot(herald.hc.avg)
herald.agglo.avg <- cbind(as.data.frame(herald.hc.avg[1]),as.data.frame(herald.hc.avg[2]))
herald.agglo.avg
cor(cophenetic(herald.hc.avg),herald.daisy)
#CPCC = 0.5411245 => best so far
```

Average linkage highest CPCC

```r
#### Centroid Linkage ----
herald.hc.centroid <- hclust(herald.daisy, method="centroid")
plot(herald.hc.centroid)
herald.agglo.centroid <- cbind(as.data.frame(herald.hc.centroid[1]),as.data.frame(herald.hc.centroid[2]))
herald.agglo.centroid
cor(cophenetic(herald.hc.centroid),herald.daisy)
#CPCC = 0.3368034 => not a good fit
```

APPENDIX A-2.2

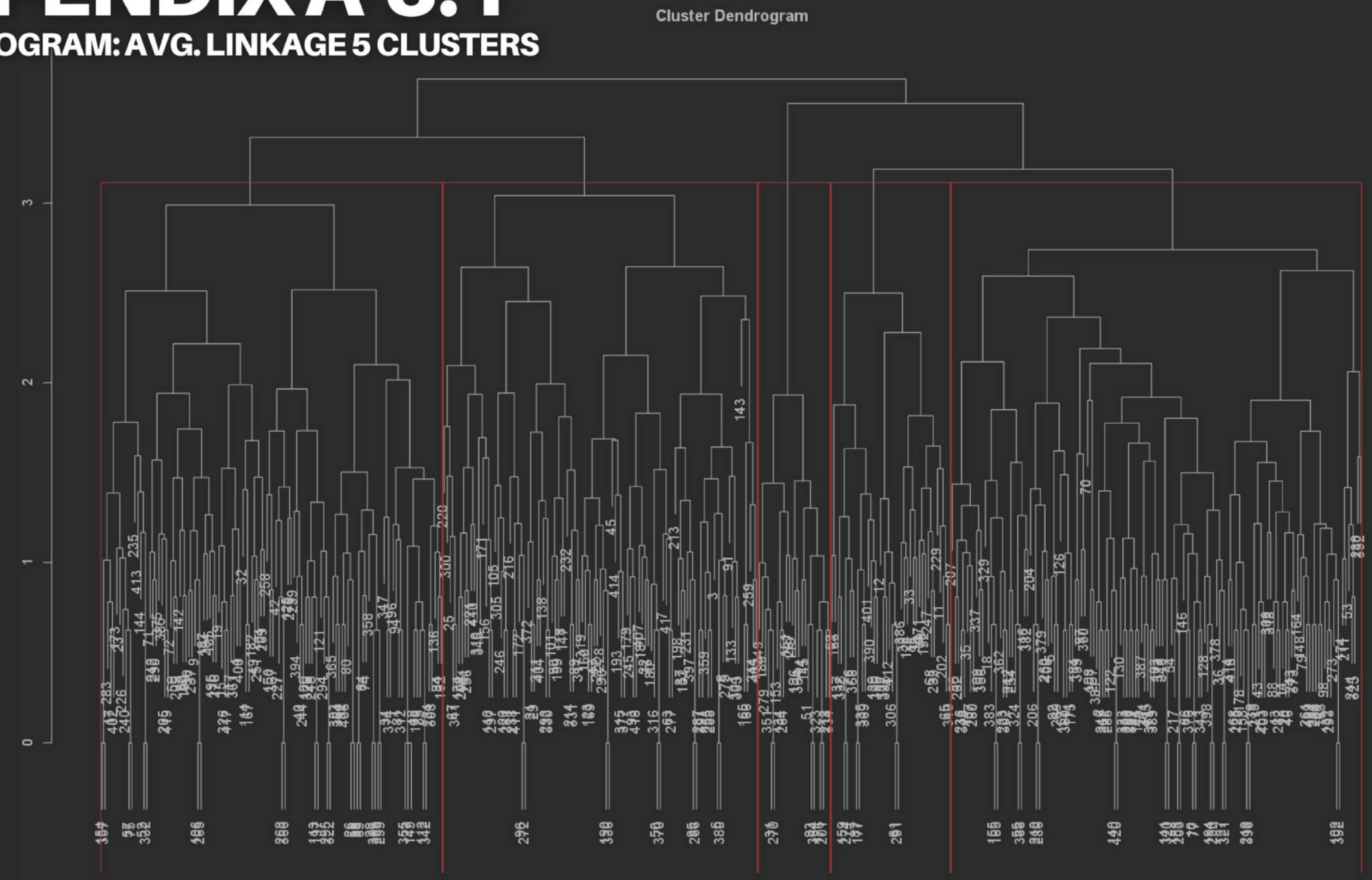
```r
#### Centroid Linkage ----
herald.hc.centroid <- hclust(herald.daisy, method="centroid")
plot(herald.hc.centroid)
herald.agglo.centroid <- cbind(as.data.frame(herald.hc.centroid[1]),as.data.frame(herald.hc.centroid[2]))
herald.agglo.centroid
cor(cophenetic(herald.hc.centroid),herald.daisy)
#CPCC = 0.3368034 => not a good fit


#### Ward's Linkage ----
herald.hc.ward <- hclust(herald.daisy, method="ward.D2")
plot(herald.hc.ward)
herald.agglo.ward <- cbind(as.data.frame(herald.hc.ward[1]),as.data.frame(herald.hc.ward[2]))
herald.agglo.ward
cor(cophenetic(herald.hc.ward),herald.daisy)
#CPCC = 0.4991169
```

Ward linkage second highest CPCC

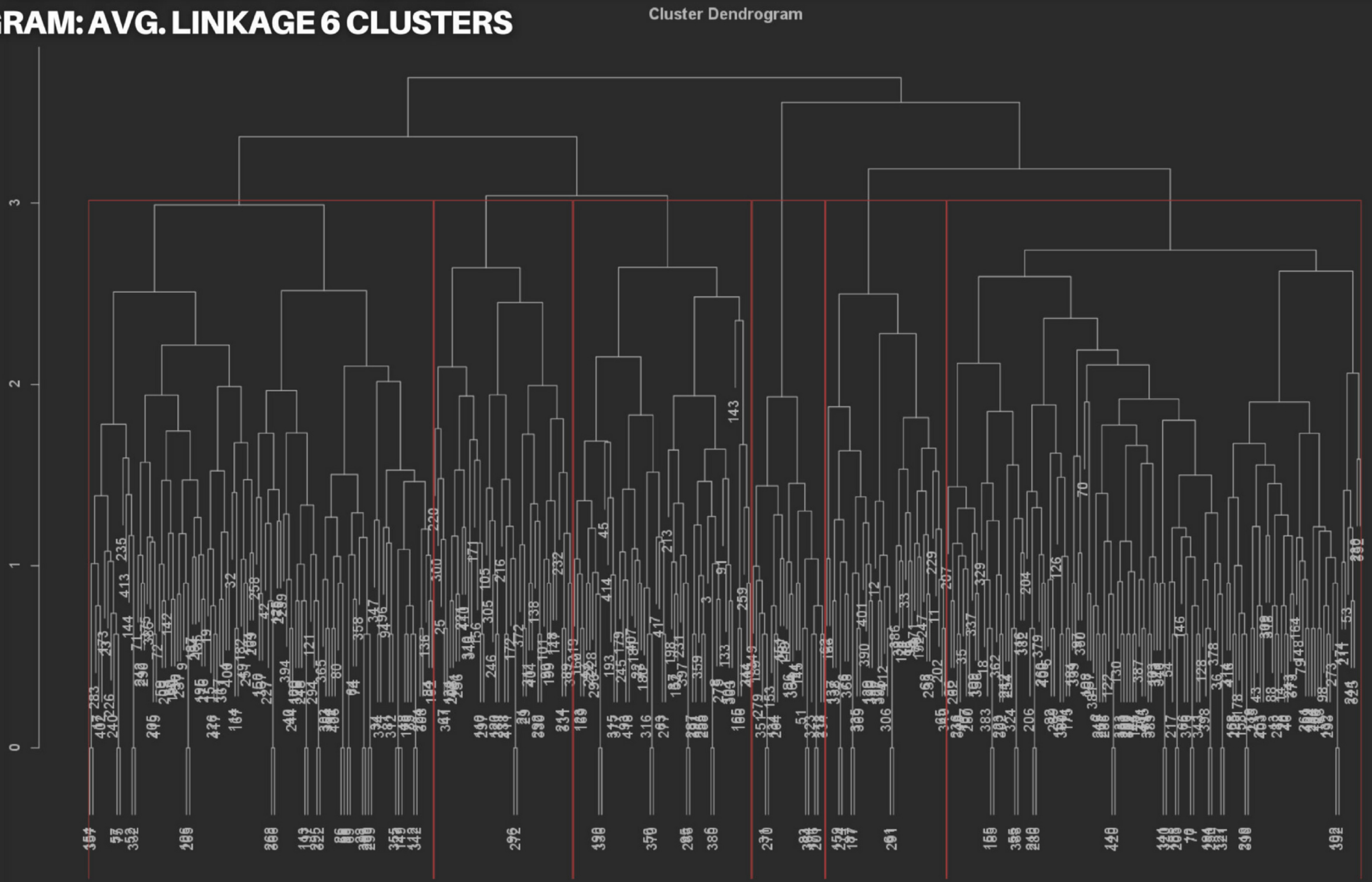
APPENDIX A-3.1

DENDROGRAM: AVG. LINKAGE 5 CLUSTERS



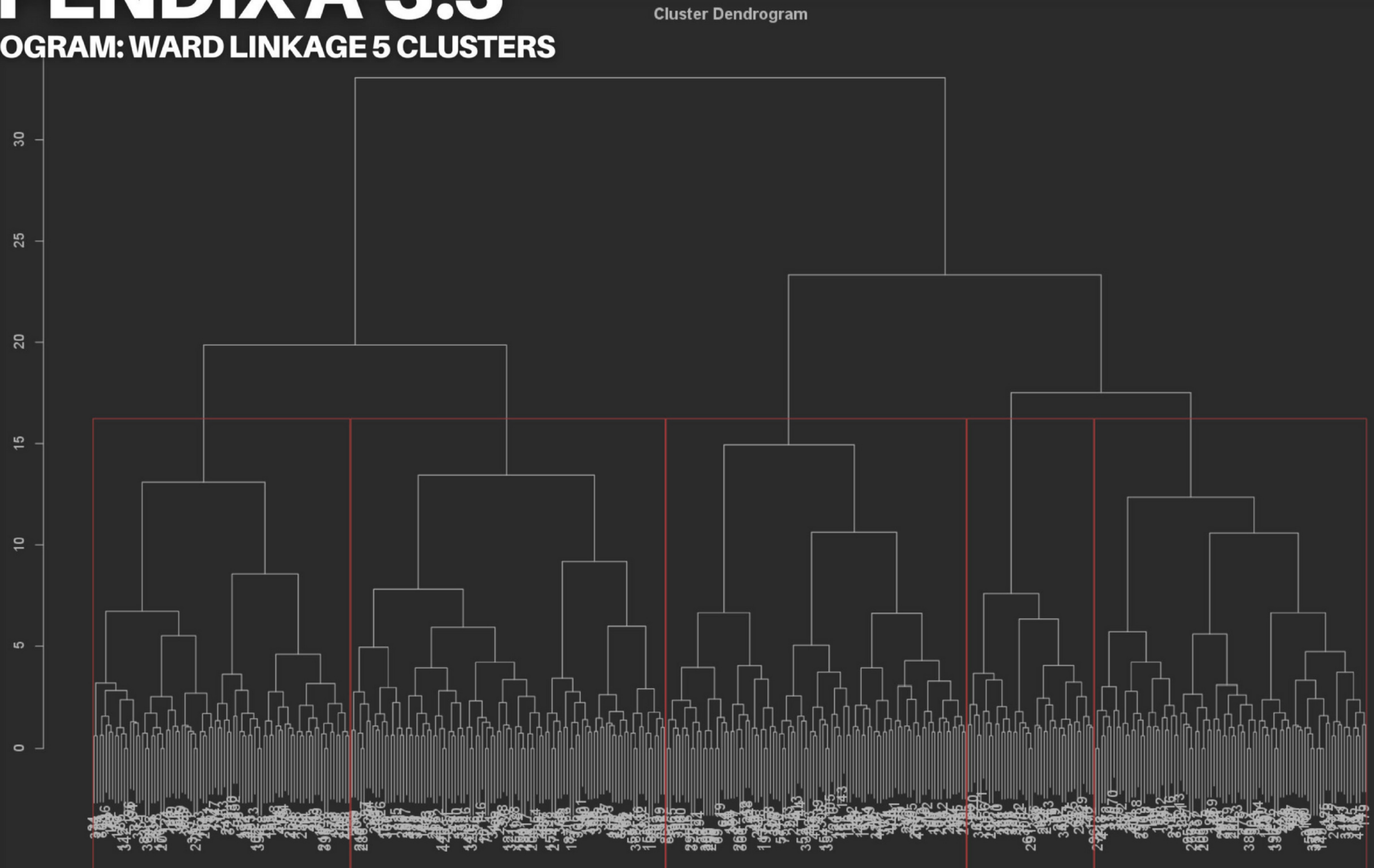
APPENDIX A-3.2

DENDROGRAM: AVG. LINKAGE 6 CLUSTERS



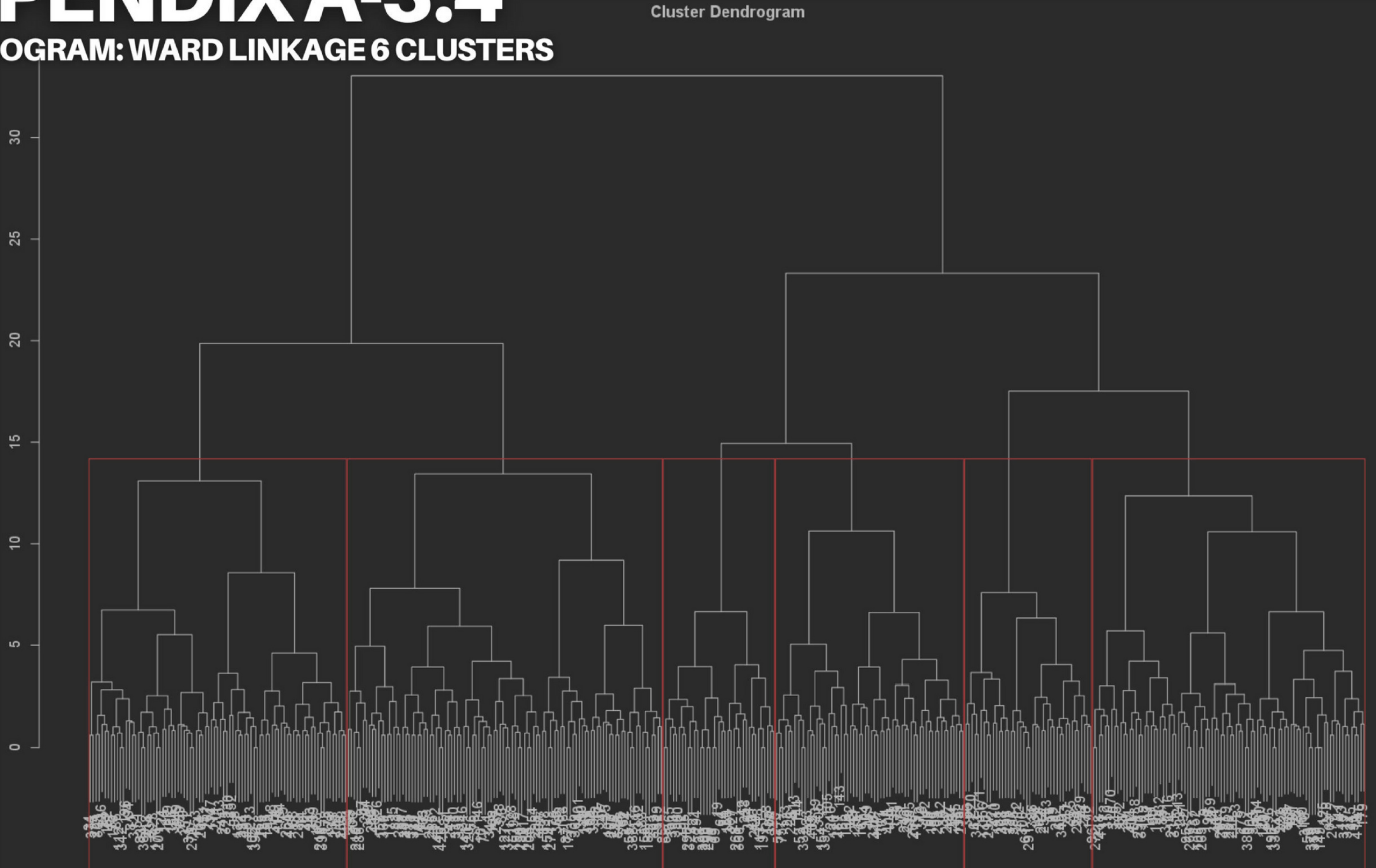
APPENDIX A-3.3

DENDROGRAM: WARD LINKAGE 5 CLUSTERS



APPENDIX A-3.4

DENDROGRAM: WARD LINKAGE 6 CLUSTERS



APPENDIX A-4

AGGLOMERATION SCHEDULES:

AVERAGE LINKAGE

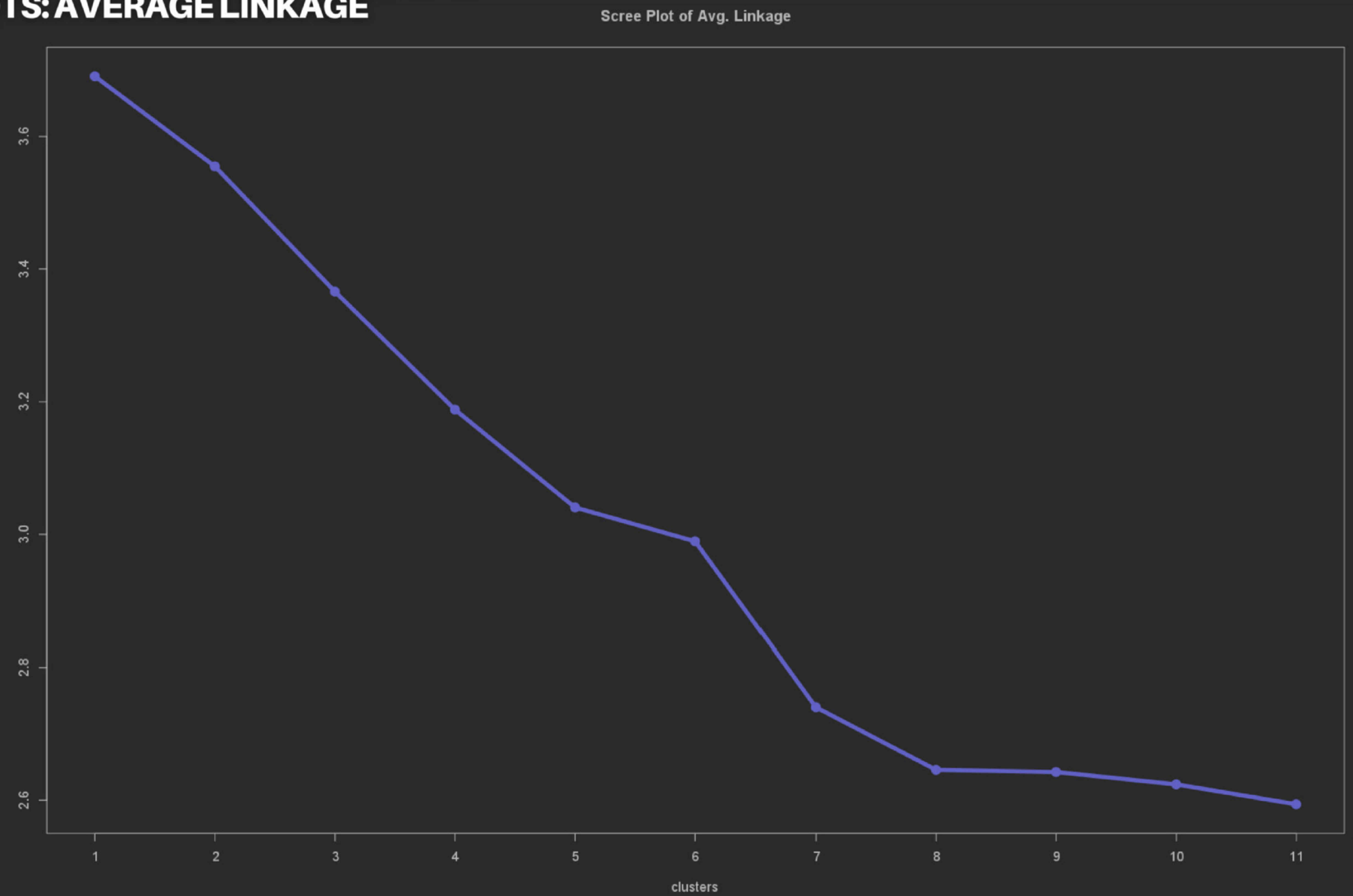
```
herald.agglo.avg[409:419,]
#merge.1 merge.2   height
#409      397     403 2.593962
#410      381     393 2.624121
#411      394     404 2.642673
#412      398     405 2.645860
#413      409     410 2.740376 <-- 7 clusters
#414      407     408 2.989426 <-- 6 clusters
#415      411     412 3.040448 <-- 5 clusters
#416      406     413 3.188129
#417      414     415 3.366040
#418      384     416 3.555047
#419      417     418 3.690821
```

WARD LINKAGE

```
herald.agglo.ward[409:419,]
#    merge.1 merge.2    height
#409     379     400  9.193091
#410     397     403 10.598295
#411     395     402 10.636004
#412     398     410 12.362224
#413     405     408 13.102808 <-- 7 clusters
#414     407     409 13.447985 <-- 6 clusters
#415     404     411 14.936958 <-- 5 clusters
#416     406     412 17.516693
#417     413     414 19.869114
#418     415     416 23.326205
#419     417     418 33.054571
```

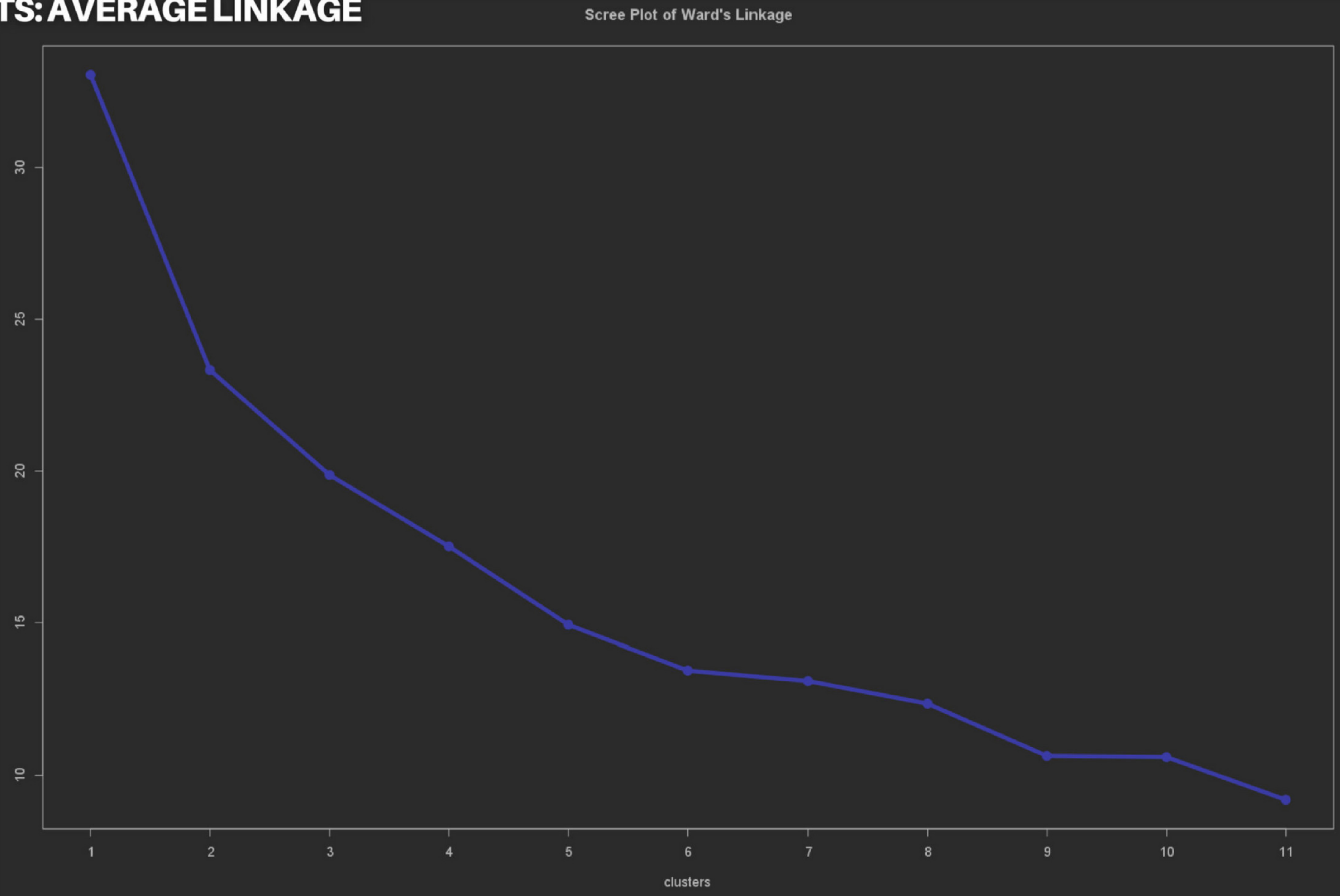
APPENDIX A-5.1

SCREE PLOTS: AVERAGE LINKAGE



APPENDIX A-5.2

SCREE PLOTS: AVERAGE LINKAGE



APPENDIX A-6.1

ROBUSTNESS CHECK: AVERAGE LINKAGE MEAN VALUES (5 & 6 CLUSTERS)

```
#Average Linkage - 5 CLUSTERS
herald.hc.avg.seg5 <- cutree(herald.hc.avg, k=5)
herald.hc.avg.mean5 <- seg.summ(herald.df_s[,c(7:12)], herald.hc.avg.seg5)
herald.hc.avg.mean5
#  Group.1     mileage       power      design    comfort entertainment environment
# 1       1 -0.07319727  0.12969328 -0.53596735  0.5551568   -0.87371925  -0.7362582
# 2       2  0.52990369 -0.05785682 -0.07180322 -0.7642550    1.09630099   0.6726285
# 3       3 -0.97106752 -0.34050420  0.94385999 -0.0160259    0.24713351   0.4488910
# 4       4  1.06181458  1.51614550 -0.89754027 -0.5676459   -0.05536086  -0.9753113
# 5       5  0.94238553 -1.39672308  0.38618490  1.1967951   -0.89045215   0.7533439

#Average Linkage - 6 CLUSTERS
herald.hc.avg.seg6 <- cutree(herald.hc.avg, k=6)
herald.hc.avg.mean6 <- seg.summ(herald.df_s[,c(7:12)], herald.hc.avg.seg6)
herald.hc.avg.mean6
#    Group.1     mileage        power     design    comfort entertainment environment
# 1       1 -0.07319727  0.129693278 -0.5359674  0.5551568   -0.87371925  -0.7362582
# 2       2  0.73103303 -0.004048218  0.6624776 -0.3397638    1.05781667   0.8240269
# 3       3 -0.97106752 -0.340504199  0.9438600 -0.0160259    0.24713351   0.4488910
# 4       4  1.06181458  1.516145501 -0.8975403 -0.5676459   -0.05536086  -0.9753113
# 5       5  0.27193345 -0.126872211 -1.0135982 -1.3087110    1.14566131   0.4784436
# 6       6  0.94238553 -1.396723082  0.3861849  1.1967951   -0.89045215   0.7533439
```

APPENDIX A-6.2

ROBUSTNESS CHECK: WARD LINKAGE MEAN VALUES (5 & 6 CLUSTERS)

```
#Ward Linkage - 5 CLUSTERS
herald.hc.ward.seg5 <- cutree(herald.hc.ward, k=5)
herald.hc.ward.mean5 <- seg.summ(herald.df_s[,c(7:12)], herald.hc.ward.seg5)
herald.hc.ward.mean5
#    Group.1     mileage      power     design    comfort entertainment environment
# 1        1  0.01289247  0.2268627 -1.0636879  0.2711422    -0.9282391 -0.91011808
# 2        2  0.61688283 -0.2388448  0.1326096 -0.5182415     0.8852540  0.80715418
# 3        3 -1.09386914 -0.1954185  0.9789584 -0.2778498     0.5352731  0.48836905
# 4        4  0.64632192  1.2179603 -0.9140356 -1.0465655     0.7937656 -0.55347715
# 5        5  0.28573238 -0.3988892  0.4724857  1.0577156    -0.8172492 -0.03640107

#Ward Linkage - 6 CLUSTERS
herald.hc.ward.seg6 <- cutree(herald.hc.ward, k=6)
herald.hc.ward.mean6 <- seg.summ(herald.df_s[,c(7:12)], herald.hc.ward.seg6)
herald.hc.ward.mean6
#    Group.1     mileage      power     design     comfort entertainment environment
# 1        1  0.01289247  0.2268627 -1.0636879  0.27114222    -0.9282391 -0.91011808
# 2        2  0.61688283 -0.2388448  0.1326096 -0.51824151     0.8852540  0.80715418
# 3        3 -1.00088602  0.3996380  0.6885472 -0.46177940     0.6202127  0.05901773
# 4        4  0.64632192  1.2179603 -0.9140356 -1.04656554     0.7937656 -0.55347715
# 5        5  0.28573238 -0.3988892  0.4724857  1.05771562    -0.8172492 -0.03640107
# 6        6 -1.24967868 -1.1925402  1.4655934  0.03035656     0.3929420  1.20782261
```

APPENDIX A-6.3

ROBUSTNESS CHECK: RANDOM SEED K-MEANS (5 & 6 CLUSTERS)

```
#Random Seed K-means - 5 CLUSTERS
set.seed( seed: 32902321)
herald.nhc.rsk5 <- kmeans(herald.df_s[,c(7:12)], centers=5, nstart = 30)
seg.summ(herald.df_s[,c(7:12)], herald.nhc.rsk5$cluster)
#   Group.1     mileage      power      design     comfort entertainment environment
# 1       1  0.05168568 -0.5566865 -0.06501103  0.96741776    -0.9489319  -0.1378888
# 2       2 -0.46694580 -1.0452944  0.94281369 -0.15192636     0.5552822   1.2094188
# 3       3 -0.57682224  0.7061500  0.97201186 -0.03831357     0.3887931  -0.0874417
# 4       4  0.39313963  0.9044006 -0.98460388 -0.06718624    -0.5852940  -1.1422455
# 5       5  0.47055534  0.3146240 -0.61612304 -1.18281580     1.1806805   0.3025010

#Random Seed K-Means - 6 CLUSTERS
set.seed( seed: 32902321)
herald.nhc.rsk6 <- kmeans(herald.df_s[,c(7:12)], centers=6, nstart = 30)
seg.summ(herald.df_s[,c(7:12)], herald.nhc.rsk6$cluster)
#   Group.1    mileage      power     design    comfort entertainment environment
# 1       1 -0.2896331 -0.1303839 -1.0024420  0.4322040    -0.9672700  -0.6965442
# 2       2 -0.5945293  0.7495608  1.0141847  0.1838324     0.1779064  -0.2809602
# 3       3  0.2232230  0.1978532 -0.5468484 -1.1304096     1.1951216   0.3851683
# 4       4 -0.5634789 -1.0234859  0.9895437 -0.2142994     0.5683526   1.2268045
# 5       5  0.4415864 -0.6423591  0.4676913  1.1110937    -0.7407439   0.1691180
# 6       6  1.0043354  1.3377199 -0.8734588 -0.3924850    -0.2771587  -1.1275699
```

APPENDIX A-7.1.1

DIFFERENCES CHECK: ANOVA & TUKEY (AVERAGE LINKAGE /W 5 CLUSTERS)

```
# MILEAGE
h.aov.avg5.mileage <- aov(mileage ~ clusmembers.avg5, data = herald.aovbase.avg5)
summary(h.aov.avg5.mileage) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.avg5.mileage) # note: we include output only of insignificant differences
# 5-2 p=0.08, the Comfort Eco-Pilgrims (5) marginally significant on Mileage from Entertained Eco-Travelers (2)
# 5-4 p=0.97, the Comfort Eco-Pilgrims (5) not sig. diff. on Mileage from Power Pilgrims (4)

# POWER
h.aov.avg5.power <- aov(power ~ clusmembers.avg5, data = herald.aovbase.avg5)
summary(h.aov.avg5.power) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.avg5.power) # note: we include output only of insignificant differences
# 2-1 p=0.37, the Entertained Eco-Travelers (2) not sig. diff. on Power from Comfort Cruisers (1)
# 3-2 p=0.069, the Fabulous Eco-Cruisers (3) marginally significant difference on Power from Entertained Eco-Travelers (2)

# DESIGN
h.aov.avg5.design <- aov(design ~ clusmembers.avg5, data = herald.aovbase.avg5)
summary(h.aov.avg5.design) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.avg5.design) # note: we include output only of insignificant differences
# 4-1 p=0.065, the Power Pilgrims (4) marginally significant difference on Design from Comfort Cruisers (1)
# 5-2 p=0.063, the Comfort Eco-Pilgrims (5) marginally significant difference on Design from Entertained Eco-Travelers (2)

# COMFORT
h.aov.avg5.comfort <- aov(comfort ~ clusmembers.avg5, data = herald.aovbase.avg5)
summary(h.aov.avg5.comfort) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.avg5.comfort) # note: we include output only of insignificant differences
# 4-2 p=0.68, the Power Pilgrims (4) not sig. difference on Comfort from Entertained Eco-Travelers (2)
```

APPENDIX A-7.1.2

DIFFERENCES CHECK: ANOVA & TUKEY (AVERAGE LINKAGE /W 5 CLUSTERS)

```r
# ENTERTAINMENT
h.aov.avg5.entertainment <- aov(entertainment ~ clusmembers.avg5, data = herald.aovbase.avg5)
summary(h.aov.avg5.entertainment) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.avg5.entertainment) # note: we include output only of insignificant differences
# 4-3 p=0.68, the Power Pilgrims (4) marginally significant difference on Ent. from Fabulous Eco-Cruisers (3)
# 5-1 p=0.99, the Comfort Eco-Pilgrims (5) not sig. diff. on Ent. from Comfort Cruisers (1)

# ENVIRONMENT
h.aov.avg5.environment <- aov(environment ~ clusmembers.avg5, data = herald.aovbase.avg5)
summary(h.aov.avg5.environment) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.avg5.environment) # note: we include output only of insignificant differences
# 4-1 p=0.37, the Power Pilgrims (4) not sig. diff.on Enviro. from Comfort Cruisers (1)
# 3-2 p=0.16, the Fabulous Eco-Cruisers (3) not sig. diff. on Enviro. from Entertained Eco-Travelers (2)
# 5-2 p=0.99, the Comfort Eco-Pilgrims (5) not sig. diff. on Enviro. from Entertained Eco-Travelers (2)
```

APPENDIX A-7.2.1

DIFFERENCES CHECK: ANOVA & TUKEY (WARD LINKAGE /W 6 CLUSTERS)

```
# MILEAGE
h.aov.ward6.mileage <- aov(mileage ~ clusmembers.ward6, data = herald.aovbase.ward6)
summary(h.aov.ward6.mileage) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.ward6.mileage) # note: we include output only of insignificant differences
# 5-1 p=0.14, Comfort Flashy-Riders (5) do not differ on Mileage from Daily Riders (1)
# 4-2 p=0.99, Entertained Power-Travelers (4) do not differ on Mileaege from Entertained Eco-Travelers (2)
# 6-3 p=0.62, Fabulous Eco-Cruisers (6) do no differ on Mileage from Entertained Flashy-Cruisers (3)
# 5-4 p=0.12, Comfort Flashy-Riders (5) do not differ on Mileage from Entertained Power-Travelers (4)

# POWER
h.aov.ward6.power <- aov(power ~ clusmembers.ward6, data = herald.aovbase.ward6)
summary(h.aov.ward6.power) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.ward6.power) # note: we include output only of insignificant differences
# 3-1 p=0.77, Entertained Flashy-Cruisers (3) do not differ on Power from Daily Riders (1)
# 5-2 p=0.78, Comfort Flashy-Riders (5) do not differ on Power from Entertained Eco-Travelers (2)

# DESIGN
h.aov.ward6.design <- aov(design ~ clusmembers.ward6, data = herald.aovbase.ward6)
summary(h.aov.ward6.design) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.ward6.design) # note: we include output only of insignificant differences
# 4-1 p=0.71,  Entertained Power-Travelers (4) do not differ on Design from Daily Riders (1)
# 5-3 p=0.22, Comfort Flashy-Riders (5) do not differ on Design from Entertained Flashy-Cruisers (3)

# COMFORT
h.aov.ward6.comfort <- aov(comfort ~ clusmembers.ward6, data = herald.aovbase.ward6)
summary(h.aov.ward6.comfort) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.ward6.comfort) # note: we include output only of insignificant differences
# 6-1 p=0.55,  Fabulous Eco-Cruisers (6) do not differ on Comfort from Daily Riders (1)
# 3-2 p=0.99, Entertained Flashy-Cruisers (3) do not differ on Comfort from Entertained Eco-Travelers (2)
```

APPENDIX A-7.2.2

DIFFERENCES CHECK: ANOVA & TUKEY (WARD LINKAGE /W 6 CLUSTERS)

```r
# ENTERTAINMENT
h.aov.ward6.entertainment <- aov(entertainment ~ clusmembers.ward6, data = herald.aovbase.ward6)
summary(h.aov.ward6.entertainment) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.ward6.entertainment) # note: we include output only of insignificant differences
# 5-1 p=0.80, Comfort Flashy-Riders (5) do not differ on Entertainment from Daily Riders (1)
# 3-2 p=0.08, Entertained Flashy-Cruisers (3) do not differ on Entertainment from Entertained Eco-Travelers (2)
# 4-2 p=0.96, Entertained Power-Travelers (4) do not differ on Entertainment from Entertained Eco-Travelers (2)
# 4-3 p=0.67, Entertained Power-Travelers (4) do not differ on Entertainment from Entertained Flashy-Cruisers (3)
# 6-3 p=0.44, Fabulous Eco-Cruisers (6) do no differ on Entertainment from Entertained Flashy-Cruisers (3)

# ENVIRONMENT
h.aov.ward6.environment <- aov(environment ~ clusmembers.ward6, data = herald.aovbase.ward6)
summary(h.aov.ward6.environment) # SIGNIFICANT (p<0.001) averages are different between clusters, but which? => TukeyHSD
TukeyHSD(h.aov.ward6.environment) # note: we include output only of insignificant differences
# 4-1 p=0.07,  Entertained Power-Travelers (4) do not differ on Environment from Daily Riders (1)
# 5-3 p=0.97, Comfort Flashy-Riders (5) do not differ on Environment from Entertained Flashy-Cruisers (3)
```

APPENDIX A-8

CLUSTER PREFERENCES OF FINAL HC SOLUTION + NAMING

```
# PREFERENCES:
#   Group.1     mileage       power      design    comfort entertainment environment
# 1       1 -0.07319727  0.12969328 -0.53596735  0.5551568   -0.87371925  -0.7362582
# 2       2  0.52990369 -0.05785682 -0.07180322 -0.7642550    1.09630099   0.6726285
# 3       3 -0.97106752 -0.34050420  0.94385999 -0.0160259    0.24713351   0.4488910
# 4       4  1.06181458  1.51614550 -0.89754027 -0.5676459   -0.05536086  -0.9753113
# 5       5  0.94238553 -1.39672308  0.38618490  1.1967951   -0.89045215   0.7533439
```

```
#### Final Cluster Names, Preferences & Differences ----
# Daily Riders (1), Entertained Eco-Travelers (2), Fabulous Eco-Cruisers (3), Power Pilgrims (4), Comfort Eco-Pilgrims (5)
# New naming: Daily Riders (1), Entertained Eco-Pilgrims (2), Designer Eco-Cruisers (3), Power Pilgrims (4), Comfort Eco-Pilgrims (5)
# Note on Mileage: Who has highest preference for mileage most? Pilgrims > Riders > Cruisers
```

APPENDIX A-9.1

CLUSTER DEMOGRAPHICS OF FINAL HC SOLUTION: RESPONDENTS PER CLUSTER

```r
#RELATIVE (%) NR. OF RESPONDENTS per cluster
hcsolution.size <- herald.hcsolution.df %>%
  group_by(cluster_name) %>%
  summarise(n.respondents = n()) %>%
  mutate(respondents.percent = formattable::percent(n.respondents / sum(n.respondents)))

print(as.data.frame(hcsolution.size))
#                  cluster_name n.respondents respondents.percent
# 1             (1) Daily Riders           137              32.62%
# 2 (2) Entertained Eco-Pilgrims           105              25.00%
# 3    (3) Designer Eco-Cruisers           114              27.14%
# 4           (4) Power Pilgrims            40               9.52%
# 5     (5) Comfort Eco-Pilgrims            24               5.71%
```

APPENDIX A-9.2

CLUSTER DEMOGRAPHICS OF FINAL HC SOLUTION: GENDER PER CLUSTER

```
#GENDER distribution per cluster
hcsolution.gender <- herald.hcsolution.df %>%
  group_by(cluster_name,gender_string) %>%
  summarise(n.respondents = n()) %>%
  mutate(respondents.percent = formattable::percent(n.respondents / sum(n.respondents)))

print(as.data.frame(hcsolution.gender))
#                   cluster_name gender_string n.respondents respondents.percent
# 1             (1) Daily Riders        female            45              32.85%
# 2             (1) Daily Riders          male            92              67.15%
# 3 (2) Entertained Eco-Pilgrims        female            58              55.24%
# 4 (2) Entertained Eco-Pilgrims          male            47              44.76%
# 5    (3) Designer Eco-Cruisers        female            90              78.95%
# 6    (3) Designer Eco-Cruisers          male            24              21.05%
# 7           (4) Power Pilgrims          male            40             100.00%
# 8     (5) Comfort Eco-Pilgrims        female            24             100.00%
```

APPENDIX A-9.3

CLUSTER DEMOGRAPHICS OF FINAL HC SOLUTION: AGE PER CLUSTER

```
#AGE distribution per cluster
hcsolution.agegroup <- herald.hcsolution.df %>%
  group_by(cluster_name,age_group) %>%
  summarise(n.respondents = n()) %>%
  mutate(respondents.percent = formattable::percent(n.respondents / sum(n.respondents)))

print(as.data.frame(hcsolution.agegroup))
#                  cluster_name    age_group n.respondents respondents.percent
# 1            (1) Daily Riders        39-48            10               7.30%
# 2            (1) Daily Riders        49-58            29              21.17%
# 3            (1) Daily Riders        59-68            63              45.99%
# 4            (1) Daily Riders 68 and older            35              25.55%
# 5  (2) Entertained Eco-Pilgrims        18-28            59              56.19%
# 6  (2) Entertained Eco-Pilgrims        29-38            27              25.71%
# 7  (2) Entertained Eco-Pilgrims        39-48            12              11.43%
# 8  (2) Entertained Eco-Pilgrims        49-58             6               5.71%
# 9  (2) Entertained Eco-Pilgrims        59-68             1               0.95%
# 10    (3) Designer Eco-Cruisers        18-28            13              11.40%
# 11    (3) Designer Eco-Cruisers        29-38            25              21.93%
# 12    (3) Designer Eco-Cruisers        39-48            41              35.96%
# 13    (3) Designer Eco-Cruisers        49-58            20              17.54%
# 14    (3) Designer Eco-Cruisers        59-68            12              10.53%
# 15    (3) Designer Eco-Cruisers 68 and older             3               2.63%
# 16          (4) Power Pilgrims        18-28             2               5.00%
# 17          (4) Power Pilgrims        29-38            11              27.50%
# 18          (4) Power Pilgrims        39-48            11              27.50%
# 19          (4) Power Pilgrims        49-58             6              15.00%
# 20          (4) Power Pilgrims        59-68             5              12.50%
# 21          (4) Power Pilgrims 68 and older             5              12.50%
# 22     (5) Comfort Eco-Pilgrims        39-48             2               8.33%
# 23     (5) Comfort Eco-Pilgrims        49-58             8              33.33%
# 24     (5) Comfort Eco-Pilgrims        59-68             6              25.00%
# 25     (5) Comfort Eco-Pilgrims 68 and older             8              33.33%
```

APPENDIX A-9.4

CLUSTER DEMOGRAPHICS OF FINAL HC SOLUTION: AREA PER CLUSTER

```
#AREA distribution per cluster
hcsolution.area <- herald.hcsolution.df %>%
  group_by(cluster_name,area_name) %>%
  summarise(n.respondents = n()) %>%
  mutate(respondents.percent = formattable::percent(n.respondents / sum(n.respondents)))

print(as.data.frame(hcsolution.area))
#                   cluster_name       area_name n.respondents respondents.percent
# 1              (1) Daily Riders (1) Metropolitan            29              21.17%
# 2              (1) Daily Riders        (2) Urban            42              30.66%
# 3              (1) Daily Riders     (3) Suburban            39              28.47%
# 4              (1) Daily Riders  (4) Countryside            27              19.71%
# 5  (2) Entertained Eco-Pilgrims (1) Metropolitan             5               4.76%
# 6  (2) Entertained Eco-Pilgrims        (2) Urban            17              16.19%
# 7  (2) Entertained Eco-Pilgrims     (3) Suburban            36              34.29%
# 8  (2) Entertained Eco-Pilgrims  (4) Countryside            47              44.76%
# 9     (3) Designer Eco-Cruisers (1) Metropolitan            74              64.91%
# 10    (3) Designer Eco-Cruisers        (2) Urban            33              28.95%
# 11    (3) Designer Eco-Cruisers     (3) Suburban             7               6.14%
# 12            (4) Power Pilgrims (1) Metropolitan             2               5.00%
# 13            (4) Power Pilgrims        (2) Urban            10              25.00%
# 14            (4) Power Pilgrims     (3) Suburban             9              22.50%
# 15            (4) Power Pilgrims  (4) Countryside            19              47.50%
# 16     (5) Comfort Eco-Pilgrims     (3) Suburban             7              29.17%
# 17     (5) Comfort Eco-Pilgrims  (4) Countryside            17              70.83%
```

APPENDIX A-9.4

CLUSTER DEMOGRAPHICS OF FINAL HC SOLUTION: EDUCATION PER CLUSTER

```r
#EDUCATION distribution per cluster
hcsolution.edu <- herald.hcsolution.df %>%
  group_by(cluster_name,education_name) %>%
  summarise(n.respondents = n()) %>%
  mutate(respondents.percent = formattable::percent(n.respondents / sum(n.respondents)))

print(as.data.frame(hcsolution.edu))
#                  cluster_name    education_name n.respondents respondents.percent
# 1             (1) Daily Riders (1) HS Profession            23              16.79%
# 2             (1) Daily Riders     (2) HS Theory            45              32.85%
# 3             (1) Daily Riders    (3) HE Non-Uni            24              17.52%
# 4             (1) Daily Riders    (4) University            20              14.60%
# 5             (1) Daily Riders         (5) Other            25              18.25%
# 6  (2) Entertained Eco-Pilgrims (1) HS Profession            17              16.19%
# 7  (2) Entertained Eco-Pilgrims     (2) HS Theory            19              18.10%
# 8  (2) Entertained Eco-Pilgrims    (3) HE Non-Uni            18              17.14%
# 9  (2) Entertained Eco-Pilgrims    (4) University            28              26.67%
# 10 (2) Entertained Eco-Pilgrims         (5) Other            23              21.90%
# 11    (3) Designer Eco-Cruisers (1) HS Profession            16              14.04%
# 12    (3) Designer Eco-Cruisers     (2) HS Theory            16              14.04%
# 13    (3) Designer Eco-Cruisers    (3) HE Non-Uni            29              25.44%
# 14    (3) Designer Eco-Cruisers    (4) University            24              21.05%
# 15    (3) Designer Eco-Cruisers         (5) Other            29              25.44%
# 16           (4) Power Pilgrims (1) HS Profession            29              72.50%
# 17           (4) Power Pilgrims     (2) HS Theory             9              22.50%
# 18           (4) Power Pilgrims    (3) HE Non-Uni             2               5.00%
# 19     (5) Comfort Eco-Pilgrims    (4) University             9              37.50%
# 20     (5) Comfort Eco-Pilgrims         (5) Other            15              62.50%
```

APPENDIX A-10

CLUSTER PREFERENCES OF FINAL NHC SOLUTION (VS. OF HC MEANS)

```
# Combined K-Means - 5 CLUSTERS
seg.summ(herald.df_s[,c(7:12)], herald.nhc.k5$cluster)
#  Group.1    mileage      power     design    comfort entertainment environment
# 1       1 -0.2591953  0.1624302 -0.6211107  0.6248499  -0.952242113  -0.8508116
# 2       2  0.3739799 -0.1293350 -0.3177337 -0.9141179   1.215883283   0.7472291
# 3       3 -1.0109362 -0.3300484  1.1746703 -0.1948118   0.426614251   0.5749511
# 4       4  0.9617836  1.2618013 -0.6235208 -0.4807406  -0.005749057  -0.8629723
# 5       5  0.3452403 -0.8957069  0.4105152  0.9603073  -0.670498345   0.4381120
```

```
# HC-Means (AVG5) <- The results of the HC-Means (AVG5) are quite similar to the Combined K-Means!
#  Group.1     mileage       power      design    comfort entertainment environment
# 1       1 -0.07319727  0.12969328 -0.53596735  0.5551568   -0.87371925  -0.7362582
# 2       2  0.52990369 -0.05785682 -0.07180322 -0.7642550    1.09630099   0.6726285
# 3       3 -0.97106752 -0.34050420  0.94385999 -0.0160259    0.24713351   0.4488910
# 4       4  1.06181458  1.51614550 -0.89754027 -0.5676459   -0.05536086  -0.9753113
# 5       5  0.94238553 -1.39672308  0.38618490  1.1967951   -0.89045215   0.7533439
```

APPENDIX A-11.1

CLUSTER DEMOGRAPHICS OF FINAL NHC SOLUTION: RESPONDENTS PER CLUSTER

```r
#RELATIVE (%) NR. OF RESPONDENTS per cluster
nhcsolution.size <- herald.nhcsolution.df %>%
  group_by(cluster_name) %>%
  summarise(n.respondents = n()) %>%
  mutate(respondents.percent = formattable::percent(n.respondents / sum(n.respondents)))

print(as.data.frame(nhcsolution.size))
#                   cluster_name n.respondents respondents.percent
# 1              (1) Daily Riders           106              25.24%
# 2 (2) Entertained Eco-Pilgrims            88              20.95%
# 3    (3) Designer Eco-Cruisers            92              21.90%
# 4             (4) Power Pilgrims            67              15.95%
# 5     (5) Comfort Eco-Pilgrims            67              15.95%
# sizes seem more equally distributed than hc solution
```

APPENDIX A-11.2

CLUSTER DEMOGRAPHICS OF FINAL NHC SOLUTION: GENDER PER CLUSTER

```
#GENDER distribution per cluster
nhcsolution.gender <- herald.nhcsolution.df %>%
  group_by(cluster_name,gender_string) %>%
  summarise(n.respondents = n()) %>%
  mutate(respondents.percent = formattable::percent(n.respondents / sum(n.respondents)))

print(as.data.frame(nhcsolution.gender))
#                   cluster_name gender_string n.respondents respondents.percent
# 1             (1) Daily Riders        female            27              25.47%
# 2             (1) Daily Riders          male            79              74.53%
# 3  (2) Entertained Eco-Pilgrims        female            38              43.18%
# 4  (2) Entertained Eco-Pilgrims          male            50              56.82%
# 5     (3) Designer Eco-Cruisers        female            76              82.61%
# 6     (3) Designer Eco-Cruisers          male            16              17.39%
# 7           (4) Power Pilgrims        female            12              17.91%
# 8           (4) Power Pilgrims          male            55              82.09%
# 9      (5) Comfort Eco-Pilgrims        female            64              95.52%
# 10     (5) Comfort Eco-Pilgrims          male             3               4.48%
```

APPENDIX A-11.3

CLUSTER DEMOGRAPHICS OF FINAL NHC SOLUTION: AGE PER CLUSTER

```
#AGE distribution per cluster
nhcsolution.agegroup <- herald.nhcsolution.df %>%
  group_by(cluster_name,age_group) %>%
  summarise(n.respondents = n()) %>%
  mutate(respondents.percent = formattable::percent(n.respondents / sum(n.respondents)))

print(as.data.frame(nhcsolution.agegroup))
#                  cluster_name    age_group n.respondents respondents.percent
# 1              (1) Daily Riders        39-48             5               4.72%
# 2              (1) Daily Riders        49-58            17              16.04%
# 3              (1) Daily Riders        59-68            53              50.00%
# 4              (1) Daily Riders 68 and older            31              29.25%
# 5  (2) Entertained Eco-Pilgrims        18-28            58              65.91%
# 6  (2) Entertained Eco-Pilgrims        29-38            23              26.14%
# 7  (2) Entertained Eco-Pilgrims        39-48             4               4.55%
# 8  (2) Entertained Eco-Pilgrims        49-58             2               2.27%
# 9  (2) Entertained Eco-Pilgrims        59-68             1               1.14%
# 10    (3) Designer Eco-Cruisers        18-28            11              11.96%
# 11    (3) Designer Eco-Cruisers        29-38            25              27.17%
# 12    (3) Designer Eco-Cruisers        39-48            41              44.57%
# 13    (3) Designer Eco-Cruisers        49-58            12              13.04%
# 14    (3) Designer Eco-Cruisers        59-68             3               3.26%
# 15           (4) Power Pilgrims        18-28             5               7.46%
# 16           (4) Power Pilgrims        29-38            14              20.90%
# 17           (4) Power Pilgrims        39-48            18              26.87%
# 18           (4) Power Pilgrims        49-58            15              22.39%
# 19           (4) Power Pilgrims        59-68            10              14.93%
# 20           (4) Power Pilgrims 68 and older             5               7.46%
# 21     (5) Comfort Eco-Pilgrims        29-38             1               1.49%
# 22     (5) Comfort Eco-Pilgrims        39-48             8              11.94%
# 23     (5) Comfort Eco-Pilgrims        49-58            23              34.33%
# 24     (5) Comfort Eco-Pilgrims        59-68            20              29.85%
# 25     (5) Comfort Eco-Pilgrims 68 and older            15              22.39%
```

APPENDIX A-11.4

CLUSTER DEMOGRAPHICS OF FINAL NHC SOLUTION: AREA PER CLUSTER

```
#AREA distribution per cluster
nhcsolution.area <- herald.nhcsolution.df %>%
  group_by(cluster_name,area_name) %>%
  summarise(n.respondents = n()) %>%
  mutate(respondents.percent = formattable::percent(n.respondents / sum(n.respondents)))

print(as.data.frame(nhcsolution.area))
#                  cluster_name       area_name n.respondents respondents.percent
# 1              (1) Daily Riders (1) Metropolitan            27              25.47%
# 2              (1) Daily Riders        (2) Urban            39              36.79%
# 3              (1) Daily Riders     (3) Suburban            28              26.42%
# 4              (1) Daily Riders  (4) Countryside            12              11.32%
# 5  (2) Entertained Eco-Pilgrims (1) Metropolitan             7               7.95%
# 6  (2) Entertained Eco-Pilgrims        (2) Urban            21              23.86%
# 7  (2) Entertained Eco-Pilgrims     (3) Suburban            23              26.14%
# 8  (2) Entertained Eco-Pilgrims  (4) Countryside            37              42.05%
# 9     (3) Designer Eco-Cruisers (1) Metropolitan            66              71.74%
# 10    (3) Designer Eco-Cruisers        (2) Urban            16              17.39%
# 11    (3) Designer Eco-Cruisers     (3) Suburban            10              10.87%
# 12           (4) Power Pilgrims (1) Metropolitan             5               7.46%
# 13           (4) Power Pilgrims        (2) Urban            11              16.42%
# 14           (4) Power Pilgrims     (3) Suburban            15              22.39%
# 15           (4) Power Pilgrims  (4) Countryside            36              53.73%
# 16     (5) Comfort Eco-Pilgrims (1) Metropolitan             5               7.46%
# 17     (5) Comfort Eco-Pilgrims        (2) Urban            15              22.39%
# 18     (5) Comfort Eco-Pilgrims     (3) Suburban            22              32.84%
# 19     (5) Comfort Eco-Pilgrims  (4) Countryside            25              37.31%
```

APPENDIX A-11.4

CLUSTER DEMOGRAPHICS OF FINAL NHC SOLUTION: EDUCATION PER CLUSTER

```
#EDUCATION distribution per cluster
nhcsolution.edu <- herald.nhcsolution.df %>%
  group_by(cluster_name,education_name) %>%
  summarise(n.respondents = n()) %>%
  mutate(respondents.percent = formattable::percent(n.respondents / sum(n.respondents)))

print(as.data.frame(nhcsolution.edu))
#                  cluster_name    education_name n.respondents respondents.percent
# 1              (1) Daily Riders (1) HS Profession            18              16.98%
# 2              (1) Daily Riders     (2) HS Theory            35              33.02%
# 3              (1) Daily Riders    (3) HE Non-Uni            15              14.15%
# 4              (1) Daily Riders    (4) University            19              17.92%
# 5              (1) Daily Riders         (5) Other            19              17.92%
# 6  (2) Entertained Eco-Pilgrims (1) HS Profession             8               9.09%
# 7  (2) Entertained Eco-Pilgrims     (2) HS Theory            19              21.59%
# 8  (2) Entertained Eco-Pilgrims    (3) HE Non-Uni            16              18.18%
# 9  (2) Entertained Eco-Pilgrims    (4) University            24              27.27%
# 10 (2) Entertained Eco-Pilgrims         (5) Other            21              23.86%
# 11    (3) Designer Eco-Cruisers (1) HS Profession            15              16.30%
# 12    (3) Designer Eco-Cruisers     (2) HS Theory            13              14.13%
# 13    (3) Designer Eco-Cruisers    (3) HE Non-Uni            19              20.65%
# 14    (3) Designer Eco-Cruisers    (4) University            17              18.48%
# 15    (3) Designer Eco-Cruisers         (5) Other            28              30.43%
# 16          (4) Power Pilgrims (1) HS Profession            43              64.18%
# 17          (4) Power Pilgrims     (2) HS Theory            18              26.87%
# 18          (4) Power Pilgrims    (3) HE Non-Uni             5               7.46%
# 19          (4) Power Pilgrims         (5) Other             1               1.49%
# 20     (5) Comfort Eco-Pilgrims (1) HS Profession             1               1.49%
# 21     (5) Comfort Eco-Pilgrims     (2) HS Theory             4               5.97%
# 22     (5) Comfort Eco-Pilgrims    (3) HE Non-Uni            18              26.87%
# 23     (5) Comfort Eco-Pilgrims    (4) University            21              31.34%
# 24     (5) Comfort Eco-Pilgrims         (5) Other            23              34.33%
```

APPENDIX B

The R script behind the raw data.

R scripts available upon formal request.