HOMEWORK4_Q4 – Sadiya Amreen

```
library(ggplot2)
library(lattice)
library(dplyr)
```

```
##
## 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
```

```
library(caret)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --

## v tibble  3.1.8     v purrr   0.3.4
## v tidyr   1.2.1     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x purrr::lift()   masks caret::lift()
```

```
library(stats)
library(factoextra)
```

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

```
library(cluster)
```

4. Back to the Starwars data from a previous assignment! Remember that the variable that lists the actual names and the variables that are actually lists will be a problem, so remove them (name, films, vehicles, starships). Make sure to double check the types of the variables, i.e., that they are numerical or factors as you expect.

```
#Importing the starwars data into the playground.
head(starwars)
```

```
## # A tibble: 6 x 14
##   name           height  mass hair_~1 skin_~2 eye_c~3 birth~4 sex   gender homew~5
##   <chr>           <int> <dbl> <chr>   <chr>   <chr>     <dbl> <chr> <chr>  <chr>
```

```
## 1 Luke Skywal~    172     77 blond    fair     blue       19   male  mascu~ Tatooi~
## 2 C-3PO           167     75 <NA>     gold     yellow    112   none  mascu~ Tatooi~
## 3 R2-D2            96     32 <NA>     white,~ red        33   none  mascu~ Naboo
## 4 Darth Vader     202    136 None     white    yellow     41.9 male  mascu~ Tatooi~
## 5 Leia Organa     150     49 brown    light    brown      19   fema~ femin~ Aldera~
## 6 Owen Lars       178    120 brown,~ light    blue       52   male  mascu~ Tatooi~
## # ... with 4 more variables: species <chr>, films <list>, vehicles <list>,
## #   starships <list>, and abbreviated variable names 1: hair_color,
## #   2: skin_color, 3: eye_color, 4: birth_year, 5: homeworld
```

```
starwars <- na.omit(starwars)
data_star <- starwars
data_star <- data_star[,-c(1,12,13,14)]#removing the names, height
str(data_star)
```

```
## tibble [29 x 10] (S3: tbl_df/tbl/data.frame)
##  $ height    : int [1:29] 172 202 150 178 165 183 182 188 228 180 ...
##  $ mass      : num [1:29] 77 136 49 120 75 84 77 84 112 80 ...
##  $ hair_color: chr [1:29] "blond" "none" "brown" "brown, grey" ...
##  $ skin_color: chr [1:29] "fair" "white" "light" "light" ...
##  $ eye_color : chr [1:29] "blue" "yellow" "brown" "blue" ...
##  $ birth_year: num [1:29] 19 41.9 19 52 47 24 57 41.9 200 29 ...
##  $ sex       : chr [1:29] "male" "male" "female" "male" ...
##  $ gender    : chr [1:29] "masculine" "masculine" "feminine" "masculine" ...
##  $ homeworld : chr [1:29] "Tatooine" "Tatooine" "Alderaan" "Tatooine" ...
##  $ species   : chr [1:29] "Human" "Human" "Human" "Human" ...
##  - attr(*, "na.action")= 'omit' Named int [1:58] 2 3 8 12 15 16 18 19 22 27 ...
##   ..- attr(*, "names")= chr [1:58] "2" "3" "8" "12" ...
```

```
#creating dummies for the data processing of the data.
dummydata <- dummyVars(gender ~ ., data=data_star)
dummies_data<- as.data.frame(predict(dummydata, newdata = data_star))
dum_dat_g <- dummies_data
dum_dat_g$gender <- starwars$gender
str(dum_dat_g)
```

```
## 'data.frame':    29 obs. of  67 variables:
##  $ height                  : num  172 202 150 178 165 183 182 188 228 180 ...
##  $ mass                    : num  77 136 49 120 75 84 77 84 112 80 ...
##  $ hair_colorauburn, white: Num  0 0 0 0 0 0 1 0 0 0 ...
##  $ hair_colorblack         : Num  0 0 0 0 0 1 0 0 0 0 ...
##  $ hair_colorblond         : Num  1 0 0 0 0 0 0 1 0 0 ...
##  $ hair_colorbrown         : Num  0 0 1 0 1 0 0 0 1 1 ...
##  $ hair_colorbrown, grey   : Num  0 0 0 1 0 0 0 0 0 0 ...
##  $ hair_colorgrey          : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ hair_colornone          : Num  0 1 0 0 0 0 0 0 0 0 ...
##  $ hair_colorwhite         : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ skin_colorblue          : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ skin_colorbrown         : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ skin_colorbrown mottle : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ skin_colordark          : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ skin_colorfair          : Num  1 0 0 0 0 0 1 1 0 1 ...
##  $ skin_colorgreen         : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ skin_colorlight         : Num  0 0 1 1 1 1 0 0 0 0 ...
##  $ skin_colororange        : Num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ skin_colorpale          : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ skin_colorred           : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ skin_colortan           : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ skin_colorunknown       : Num  0 0 0 0 0 0 0 0 1 0 ...
##  $ skin_colorwhite         : Num  0 1 0 0 0 0 0 0 0 0 ...
##  $ skin_coloryellow        : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ eye_colorblack          : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ eye_colorblue           : Num  1 0 0 1 1 0 0 1 1 0 ...
##  $ eye_colorblue-gray      : Num  0 0 0 0 0 0 1 0 0 0 ...
##  $ eye_colorbrown          : Num  0 0 1 0 0 1 0 0 0 1 ...
##  $ eye_colorhazel          : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ eye_colororange         : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ eye_colorred            : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ eye_coloryellow         : Num  0 1 0 0 0 0 0 0 0 0 ...
##  $ birth_year              : num  19 41.9 19 52 47 24 57 41.9 200 29 ...
##  $ sexfemale               : Num  0 0 1 0 1 0 0 0 0 0 ...
##  $ sexmale                 : Num  1 1 0 1 0 1 1 1 1 1 ...
##  $ homeworldAlderaan       : Num  0 0 1 0 0 0 0 0 0 0 ...
##  $ homeworldBespin         : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldCerea          : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldConcord Dawn   : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldCorellia       : Num  0 0 0 0 0 0 0 0 0 1 ...
##  $ homeworldDathomir       : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldDorin          : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldEndor          : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldHaruun Kal     : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldKamino         : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldKashyyyk       : Num  0 0 0 0 0 0 0 0 1 0 ...
##  $ homeworldMirial         : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldMon Cala       : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldNaboo          : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldRyloth         : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldSerenno        : Num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldSocorro        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ homeworldStewjon        : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ homeworldTatooine       : num  1 1 0 1 1 1 0 1 0 0 ...
##  $ homeworldTrandosha      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ speciesCerean           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ speciesEwok             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ speciesGungan           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ speciesHuman            : num  1 1 1 1 1 1 1 1 0 1 ...
##  $ speciesKel Dor          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ speciesMirialan         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ speciesMon Calamari     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ speciesTrandoshan       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ speciesTwi'lek          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ speciesWookiee          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ speciesZabrak           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gender                  : chr  "masculine" "masculine" "feminine" "masculine" ...
```

```r
dum_dat_g <- na.omit(dum_dat_g)
dist_mat <- daisy(dummies_data, metric = "gower")#Using the daisy from cluster to find the dist_mat.
```

```
## Warning in daisy(dummies_data, metric = "gower"): binary variable(s) 3, 4, 5, 6,
## 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,
```

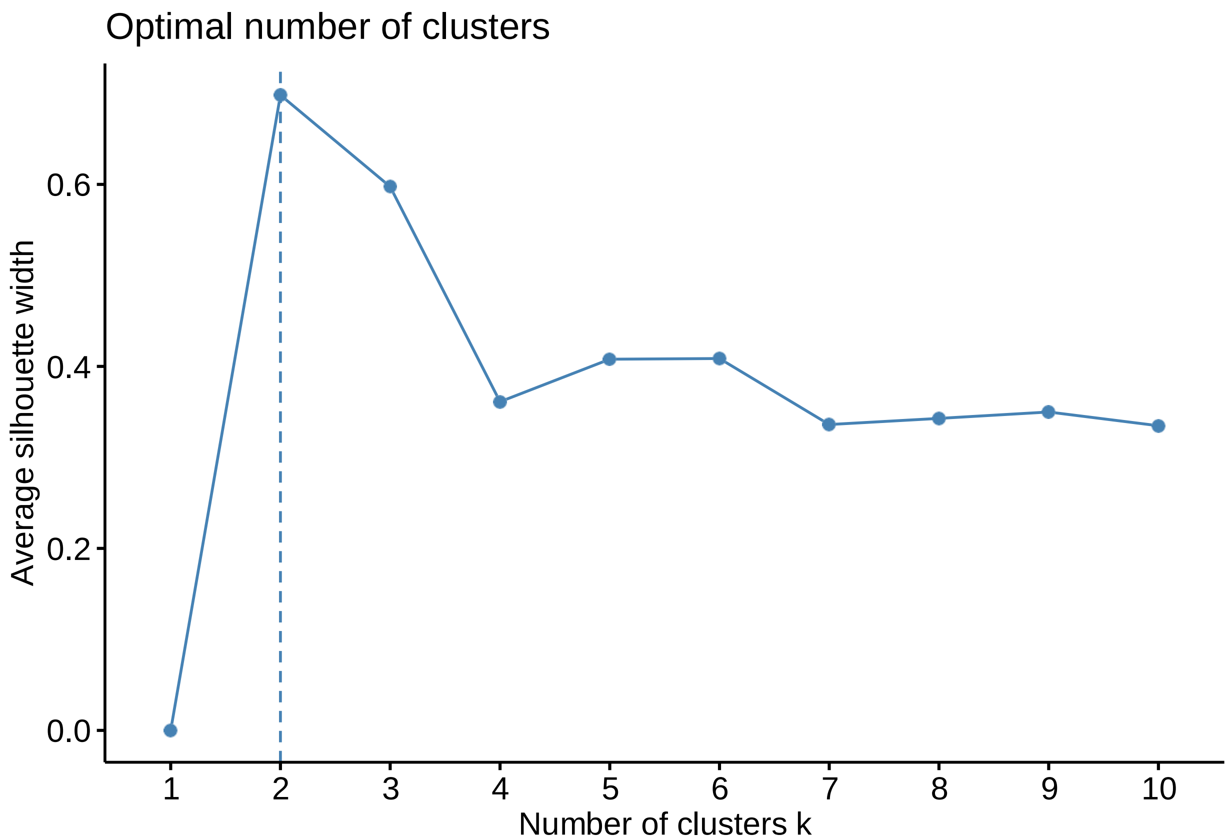
```
## 28, 29, 30, 31, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48,
## 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66 treated
## as interval scaled
```

4.A. Use hierarchical agglomeration clustering to cluster the Star wars data. This time we can leave the categorical variables in place, because we will use the gower metric from daisy in the cluster library to get the distances. Use average linkage. Determine the best number of clusters.

Now we have to perform the preprocess and predict for the k-means plotting by using the silhouette.

```
preproc <- preProcess(data_star, method=c("center", "scale"))
predictors <- predict(preproc, data_star)
fviz_nbclust(data_star, FUN = hcut, method = "silhouette")
```

```
## Warning in stats::dist(x): NAs introduced by coercion

## Warning in stats::dist(x, method = method, ...): NAs introduced by coercion

## Warning in stats::dist(x, method = method, ...): NAs introduced by coercion

## Warning in stats::dist(x, method = method, ...): NAs introduced by coercion

## Warning in stats::dist(x, method = method, ...): NAs introduced by coercion

## Warning in stats::dist(x, method = method, ...): NAs introduced by coercion

## Warning in stats::dist(x, method = method, ...): NAs introduced by coercion

## Warning in stats::dist(x, method = method, ...): NAs introduced by coercion

## Warning in stats::dist(x, method = method, ...): NAs introduced by coercion

## Warning in stats::dist(x, method = method, ...): NAs introduced by coercion
```



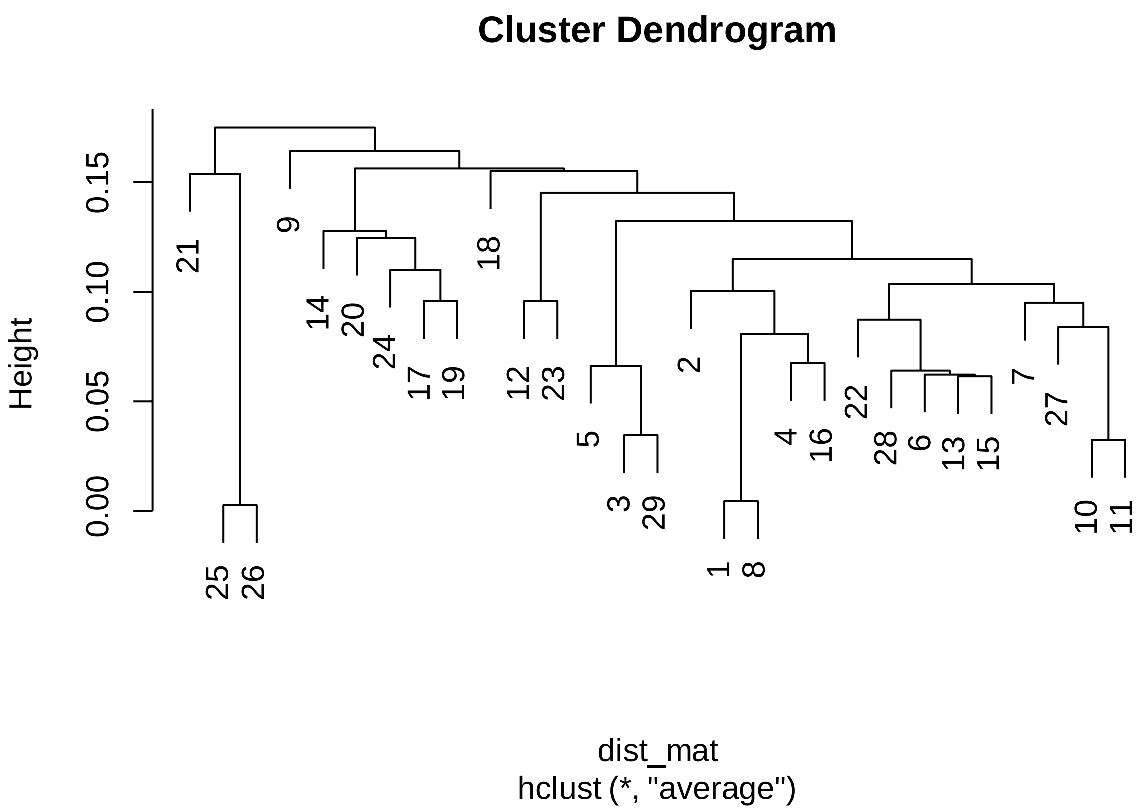
To find the best number of clusters

```
best_number_cluster <- hclust(dist_mat, method = "average")
cluster_star <- cutree(best_number_cluster, k=2)
summary(cluster_star)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.103   1.000   2.000
```

4.B. Produce the dendogram for (a). How might an anomaly show up in a dendogram? Do you see a Starwars character who does not seem to fit in easily? What is the advantage of considering anomalies this way as opposed to looking for unusual values relative to the mean and standard deviations, as we considered earlier in the course? Disadvantages?

```
dendogram <- hclust(dist_mat, method = "average")
plot(dendogram)
```



The anamolities can be easily found by the dendogram by the height of the nodes, however we can only predict the anamolities by the looks of the dendogram but they can be have similar characterstics so we have to be careful with the similarities too.

4.C.Use dummy variables to make this data fully numeric and then use k-means to cluster. Choose the best number of clusters.

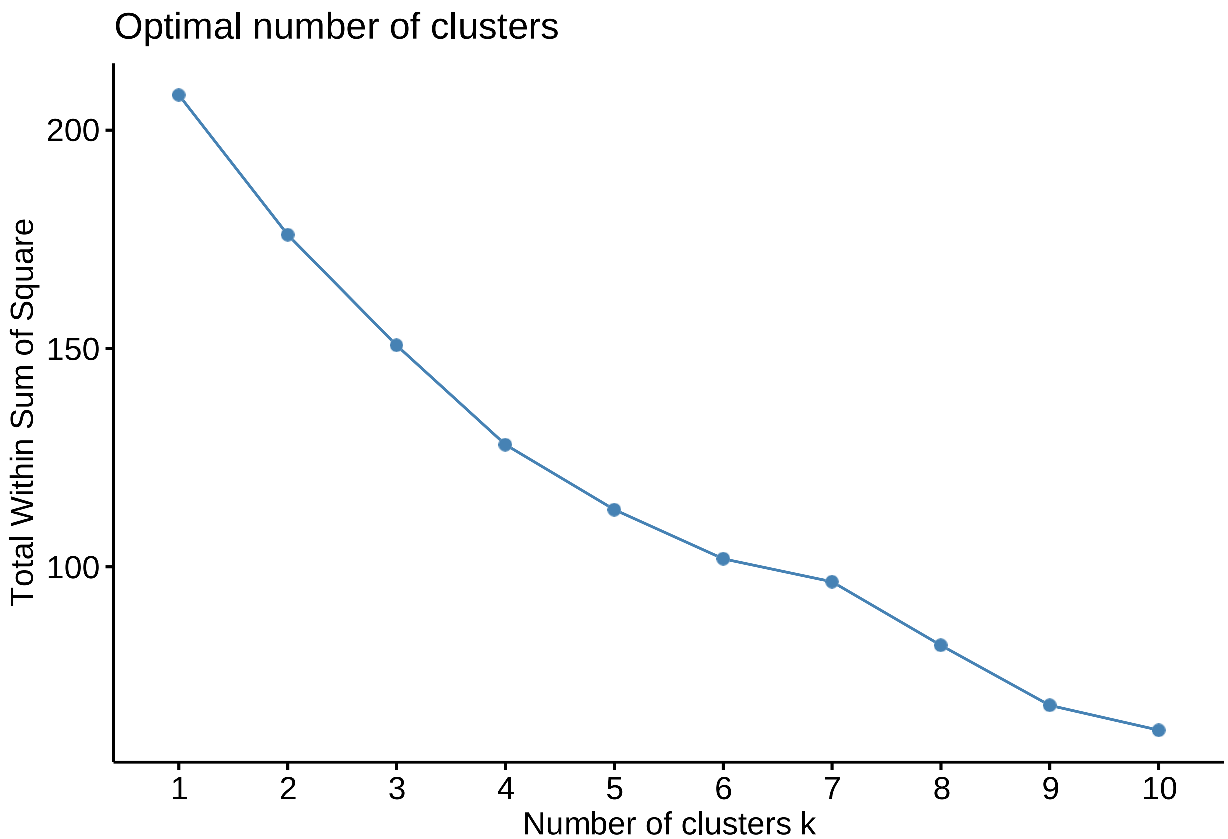
```
head(dum_dat_g)
```

```
##   height mass hair_colorauburn, white hair_colorblack hair_colorblond
## 1    172   77                       0               0               1
## 2    202  136                       0               0               0
## 3    150   49                       0               0               0
## 4    178  120                       0               0               0
## 5    165   75                       0               0               0
## 6    183   84                       0               1               0
##   hair_colorbrown hair_colorbrown, grey hair_colorgrey hair_colornone
## 1               0                     0              0              0
## 2               0                     0              0              1
## 3               1                     0              0              0
## 4               0                     1              0              0
## 5               1                     0              0              0
## 6               0                     0              0              0
##   hair_colorwhite skin_colorblue skin_colorbrown skin_colorbrown mottle
## 1               0              0               0                      0
## 2               0              0               0                      0
## 3               0              0               0                      0
## 4               0              0               0                      0
## 5               0              0               0                      0
```

```
## 6               0              0               0                      0
##   skin_colordark skin_colorfair skin_colorgreen skin_colorlight
## 1              0              1               0               0
## 2              0              0               0               0
## 3              0              0               0               1
## 4              0              0               0               1
## 5              0              0               0               1
## 6              0              0               0               1
##   skin_colororange skin_colorpale skin_colorred skin_colortan skin_colorunknown
## 1                0              0             0             0                 0
## 2                0              0             0             0                 0
## 3                0              0             0             0                 0
## 4                0              0             0             0                 0
## 5                0              0             0             0                 0
## 6                0              0             0             0                 0
##   skin_colorwhite skin_coloryellow eye_colorblack eye_colorblue
## 1               0                0              0             1
## 2               1                0              0             0
## 3               0                0              0             0
## 4               0                0              0             1
## 5               0                0              0             1
## 6               0                0              0             0
##   eye_colorblue-gray eye_colorbrown eye_colorhazel eye_colororange eye_colorred
## 1                  0              0              0               0            0
## 2                  0              0              0               0            0
## 3                  0              1              0               0            0
## 4                  0              0              0               0            0
## 5                  0              0              0               0            0
## 6                  0              1              0               0            0
##   eye_coloryellow birth_year sexfemale sexmale homeworldAldera n
## 1               0       19.0         0       1                 0
## 2               1       41.9         0       1                 0
## 3               0       19.0         1       0                 1
## 4               0       52.0         0       1                 0
## 5               0       47.0         1       0                 0
## 6               0       24.0         0       1                 0
##   homeworldBespin homeworldCerea homeworldConcord Dawn homeworldCorellia
## 1               0              0                     0                 0
## 2               0              0                     0                 0
## 3               0              0                     0                 0
## 4               0              0                     0                 0
## 5               0              0                     0                 0
## 6               0              0                     0                 0
##   homeworldDathomir homeworldDorin homeworldEndor homeworldHaruun Kal
## 1                 0              0              0                   0
## 2                 0              0              0                   0
## 3                 0              0              0                   0
## 4                 0              0              0                   0
## 5                 0              0              0                   0
## 6                 0              0              0                   0
##   homeworldKamino homeworldKashyyyk homeworldMirial homeworldMon Cala
## 1               0                 0               0                 0
## 2               0                 0               0                 0
## 3               0                 0               0                 0
```

```
## 4                0                 0               0                 0
## 5                0                 0               0                 0
## 6                0                 0               0                 0
##   homeworldNaboo homeworldRyloth homeworldSerenno homeworldSocorro
## 1              0               0                0                0
## 2              0               0                0                0
## 3              0               0                0                0
## 4              0               0                0                0
## 5              0               0                0                0
## 6              0               0                0                0
##   homeworldStewjon homeworldTatooine homeworldTrandosha speciesCerean
## 1                0                 1                  0             0
## 2                0                 1                  0             0
## 3                0                 0                  0             0
## 4                0                 1                  0             0
## 5                0                 1                  0             0
## 6                0                 1                  0             0
##   speciesEwok speciesGungan speciesHuman speciesKel Dor speciesMirialan
## 1           0             0            1              0               0
## 2           0             0            1              0               0
## 3           0             0            1              0               0
## 4           0             0            1              0               0
## 5           0             0            1              0               0
## 6           0             0            1              0               0
##   speciesMon Calamari speciesTrandoshan speciesTwi'lek speciesWookiee
## 1                   0                 0              0              0
## 2                   0                 0              0              0
## 3                   0                 0              0              0
## 4                   0                 0              0              0
## 5                   0                 0              0              0
## 6                   0                 0              0              0
##   speciesZabrak    gender
## 1             0 masculine
## 2             0 masculine
## 3             0  feminine
## 4             0 masculine
## 5             0  feminine
## 6             0 masculine
```

```
predictors <- dummies_data
set.seed(125)
prproc <- preProcess(predictors, method=c("center", "scale"))
predictors <- predict(preproc, predictors)
fviz_nbclust(predictors, kmeans, method = "wss")
```



Fit of the Kmeans is 4.

```r
w_fit_d <- kmeans(predictors, centers = 4, nstart = 25)#fit the data
w_fit_d
```

```
## K-means clustering with 4 clusters of sizes 1, 1, 21, 6
##
## Cluster means:
##       height       mass hair_colorauburn, white hair_colorblack hair_colorblond
## 1  2.2232870  1.4826220              0.00000000       0.0000000       0.0000000
## 2 -3.9976263 -2.5025035              0.00000000       0.0000000       0.0000000
## 3  0.2406354  0.3299884              0.04761905       0.1904762       0.0952381
## 4 -0.5465006 -0.9849792              0.00000000       0.3333333       0.0000000
##   hair_colorbrown hair_colorbrown, grey hair_colorgrey hair_colornone
## 1       1.0000000            0.00000000     0.00000000      0.0000000
## 2       1.0000000            0.00000000     0.00000000      0.0000000
## 3       0.0952381            0.04761905     0.04761905      0.3809524
## 4       0.5000000            0.00000000     0.00000000      0.1666667
##   hair_colorwhite skin_colorblue skin_colorbrown skin_colorbrown mottle
## 1       0.0000000      0.0000000               0             0.00000000
## 2       0.0000000      0.0000000               1             0.00000000
## 3       0.0952381      0.0000000               0             0.04761905
## 4       0.0000000      0.1666667               0             0.00000000
##   skin_colordark skin_colorfair skin_colorgreen skin_colorlight
## 1      0.0000000      0.0000000      0.00000000       0.0000000
## 2      0.0000000      0.0000000      0.00000000       0.0000000
## 3      0.0952381      0.3333333      0.04761905       0.1428571
## 4      0.0000000      0.0000000      0.00000000       0.5000000
```

```
##   skin_colororange skin_colorpale skin_colorred skin_colortan skin_colorunknown
## 1        0.0000000      0.0000000    0.00000000    0.00000000                 1
## 2        0.0000000      0.0000000    0.00000000    0.00000000                 0
## 3        0.0952381      0.0952381    0.04761905    0.04761905                 0
## 4        0.0000000      0.0000000    0.00000000    0.00000000                 0
##   skin_colorwhite skin_coloryellow eye_colorblack eye_colorblue
## 1      0.00000000        0.0000000     0.00000000     1.0000000
## 2      0.00000000        0.0000000     0.00000000     0.0000000
## 3      0.04761905        0.0000000     0.04761905     0.1904762
## 4      0.00000000        0.3333333     0.00000000     0.5000000
##   eye_colorblue-gray eye_colorbrown eye_colorhazel eye_colororange eye_colorred
## 1         0.00000000      0.0000000     0.00000000       0.0000000   0.00000000
## 2         0.00000000      1.0000000     0.00000000       0.0000000   0.00000000
## 3         0.04761905      0.3333333     0.04761905       0.0952381   0.04761905
## 4         0.00000000      0.3333333     0.16666667       0.0000000   0.00000000
##   eye_coloryellow  birth_year sexfemale sexmale homeworldAlderaan
## 1       0.0000000  4.12053386         0       1         0.0000000
## 2       0.0000000 -1.19936609         0       1         0.0000000
## 3       0.1904762 -0.07350532         0       1         0.0000000
## 4       0.0000000 -0.22959266         1       0         0.1666667
##   homeworldBespin homeworldCerea homeworldConcord Dawn homeworldCorellia
## 1      0.00000000     0.00000000           0.00000000         0.0000000
## 2      0.00000000     0.00000000           0.00000000         0.0000000
## 3      0.04761905     0.04761905           0.04761905         0.0952381
## 4      0.00000000     0.00000000           0.00000000         0.0000000
##   homeworldDathomir homeworldDorin homeworldEndor homeworldHaruun Kal
## 1        0.00000000     0.00000000              0          0.00000000
## 2        0.00000000     0.00000000              1          0.00000000
## 3        0.04761905     0.04761905              0          0.04761905
## 4        0.00000000     0.00000000              0          0.00000000
##   homeworldKamino homeworldKashyyyk homeworldMirial homeworldMon Cala
## 1      0.00000000                 1       0.0000000        0.00000000
## 2      0.00000000                 0       0.0000000        0.00000000
## 3      0.04761905                 0       0.0000000        0.04761905
## 4      0.00000000                 0       0.3333333        0.00000000
##   homeworldNaboo homeworldRyloth homeworldSerenno homeworldSocorro
## 1      0.0000000       0.0000000       0.00000000       0.00000000
## 2      0.0000000       0.0000000       0.00000000       0.00000000
## 3      0.0952381       0.0000000       0.04761905       0.04761905
## 4      0.1666667       0.1666667       0.00000000       0.00000000
##   homeworldStewjon homeworldTatooine homeworldTrandosha speciesCerean
## 1       0.00000000         0.0000000         0.00000000    0.00000000
## 2       0.00000000         0.0000000         0.00000000    0.00000000
## 3       0.04761905         0.2380952         0.04761905    0.04761905
## 4       0.00000000         0.1666667         0.00000000    0.00000000
##   speciesEwok speciesGungan speciesHuman speciesKel Dor speciesMirialan
## 1           0    0.00000000    0.0000000     0.00000000       0.0000000
## 2           1    0.00000000    0.0000000     0.00000000       0.0000000
## 3           0    0.04761905    0.7142857     0.04761905       0.0000000
## 4           0    0.00000000    0.5000000     0.00000000       0.3333333
##   speciesMon Calamari speciesTrandoshan speciesTwi'lek speciesWookiee
## 1          0.00000000        0.00000000      0.0000000              1
## 2          0.00000000        0.00000000      0.0000000              0
## 3          0.04761905        0.04761905      0.0000000              0
```

```
## 4           0.00000000          0.00000000       0.1666667                 0
##   speciesZabrak
## 1     0.00000000
## 2     0.00000000
## 3     0.04761905
## 4     0.00000000
##
## Clustering vector:
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
##  3  3  4  3  4  3  3  3  1  3  3  3  3  3  3  3  3  2  3  3  4  3  3  3  4  4
## 27 28 29
##  3  3  4
##
## Within cluster sum of squares by cluster:
## [1]   0.0000   0.0000 101.6367  21.8729
##  (between_SS / total_SS =  40.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

4.D. Compare the HAC and k-means clusterings with a crosstabulation.

```
#Creating new fit for the Kmenas by using the dummies data.
set.seed(123)
preproc <- preProcess(dummies_data, method=c("center", "scale"))
predictors <- predict(preproc, dummies_data)
dist_mat_4 <- dist(predictors, method = "euclidean")
hfit_4 <- hclust(dist_mat_4, method = "average")
h1 <- cutree(hfit_4, k=2)
fit <- kmeans(dummies_data, centers = 4, nstart = 25)
summary(cluster_star)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.103   1.000   2.000
```

```
results <- data.frame(Gender = dum_dat_g$gender, HAC_Star_Wars = h1, Kmeans = fit$cluster)
best_number_cluster
```

```
##
## Call:
## hclust(d = dist_mat, method = "average")
##
## Cluster method   : average
## Number of objects: 29
```

```
results %>% group_by(HAC_Star_Wars) %>% select(HAC_Star_Wars, Gender) %>% table()
```

```
##              Gender
## HAC_Star_Wars feminine masculine
##             1        6        22
##             2        0         1
```

```
results %>% group_by(Kmeans) %>% select(Kmeans, Gender) %>% table()
```

```
##       Gender
```

```
## Kmeans feminine masculine
##      1        6        13
##      2        0         1
##      3        0         1
##      4        0         8
```