

d212_task2.R

sigsp

2022-09-24

```
## Author: Stephen E. Porter
## Title: D212 Task 1 Clustering Analysis
## Course: WGU D212: Data Mining II
## Instructor: Dr.Keiona Middleton

# Libraries
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.5      v dplyr  1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.0.2      v forcats 0.5.1
## Warning: package 'ggplot2' was built under R version 4.1.1
## Warning: package 'tibble' was built under R version 4.1.1
## Warning: package 'tidyr' was built under R version 4.1.1
## Warning: package 'readr' was built under R version 4.1.1
## Warning: package 'dplyr' was built under R version 4.1.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
library(corrplot)

## Warning: package 'corrplot' was built under R version 4.1.2
## corrplot 0.92 loaded
library(factoextra)

## Warning: package 'factoextra' was built under R version 4.1.3
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

# Set seed
set.seed(22)

# Import CSV as data frame
df <- read.csv(file = 'C:/WGU/D212 Data Mining II/churn_clean.csv')

# Checking for nulls
sapply(df, function(x) sum(is.na(x)))
```

```
##          CaseOrder      Customer_id      Interaction
##              0              0              0
##          UID              City              State
##              0              0              0
##          County          Zip              Lat
##              0              0              0
##          Lng              Population          Area
##              0              0              0
##          TimeZone          Job              Children
##              0              0              0
##          Age              Income          Marital
##              0              0              0
##          Gender          Churn      Outage_sec_perweek
##              0              0              0
##          Email          Contacts      Yearly_equip_failure
##              0              0              0
##          Techie          Contract          Port_modem
##              0              0              0
##          Tablet      InternetService          Phone
##              0              0              0
##          Multiple      OnlineSecurity      OnlineBackup
##              0              0              0
##          DeviceProtection      TechSupport      StreamingTV
##              0              0              0
##          StreamingMovies      PaperlessBilling      PaymentMethod
##              0              0              0
##          Tenure          MonthlyCharge      Bandwidth_GB_Year
##              0              0              0
##          Item1              Item2              Item3
##              0              0              0
##          Item4              Item5              Item6
##              0              0              0
##          Item7              Item8
##              0              0
```

```
str(df)
```

```
## 'data.frame':    10000 obs. of  50 variables:
## $ CaseOrder      : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Customer_id    : chr   "K409198" "S120509" "K191035" "D90850" ...
## $ Interaction     : chr   "aa90260b-4141-4a24-8e36-b04ce1f4f77b" "fb76459f-c047-4a9d-8af9-e0f7d4..."
## $ UID            : chr   "e885b299883d4f9fb18e39c75155d990" "f2de8bef964785f41a2959829830fb8a" ...
## $ City           : chr   "Point Baker" "West Branch" "Yamhill" "Del Mar" ...
## $ State          : chr   "AK" "MI" "OR" "CA" ...
## $ County         : chr   "Prince of Wales-Hyder" "Ogemaw" "Yamhill" "San Diego" ...
## $ Zip            : int   99927 48661 97148 92014 77461 31030 37847 73109 34771 45237 ...
## $ Lat            : num   56.3 44.3 45.4 33 29.4 ...
## $ Lng            : num   -133.4 -84.2 -123.2 -117.2 -95.8 ...
## $ Population     : int   38 10446 3735 13863 11352 17701 2535 23144 17351 20193 ...
## $ Area           : chr   "Urban" "Urban" "Urban" "Suburban" ...
## $ TimeZone       : chr   "America/Sitka" "America/Detroit" "America/Los_Angeles" "America/Los_An..."
## $ Job            : chr   "Environmental health practitioner" "Programmer, multimedia" "Chief Fi..."
## $ Children       : int   0 1 4 1 0 3 0 2 2 1 ...
## $ Age            : int   68 27 50 48 83 83 79 30 49 86 ...
## $ Income         : num   28562 21705 9610 18925 40074 ...
```

```

## $ Marital      : chr "Widowed" "Married" "Widowed" "Married" ...
## $ Gender       : chr "Male" "Female" "Female" "Male" ...
## $ Churn        : chr "No" "Yes" "No" "No" ...
## $ Outage_sec_perweek : num 7.98 11.7 10.75 14.91 8.15 ...
## $ Email        : int 10 12 9 15 16 15 10 16 20 18 ...
## $ Contacts     : int 0 0 0 2 2 3 0 0 2 1 ...
## $ Yearly equip_failure: int 1 1 1 0 1 1 1 0 3 0 ...
## $ Techie       : chr "No" "Yes" "Yes" "Yes" ...
## $ Contract     : chr "One year" "Month-to-month" "Two Year" "Two Year" ...
## $ Port_modem   : chr "Yes" "No" "Yes" "No" ...
## $ Tablet       : chr "Yes" "Yes" "No" "No" ...
## $ InternetService : chr "Fiber Optic" "Fiber Optic" "DSL" "DSL" ...
## $ Phone        : chr "Yes" "Yes" "Yes" "Yes" ...
## $ Multiple     : chr "No" "Yes" "Yes" "No" ...
## $ OnlineSecurity : chr "Yes" "Yes" "No" "Yes" ...
## $ OnlineBackup  : chr "Yes" "No" "No" "No" ...
## $ DeviceProtection : chr "No" "No" "No" "No" ...
## $ TechSupport  : chr "No" "No" "No" "No" ...
## $ StreamingTV   : chr "No" "Yes" "No" "Yes" ...
## $ StreamingMovies : chr "Yes" "Yes" "Yes" "No" ...
## $ PaperlessBilling : chr "Yes" "Yes" "Yes" "Yes" ...
## $ PaymentMethod : chr "Credit Card (automatic)" "Bank Transfer(automatic)" "Credit Card (aut
## $ Tenure       : num 6.8 1.16 15.75 17.09 1.67 ...
## $ MonthlyCharge : num 172 243 160 120 150 ...
## $ Bandwidth_GB_Year : num 905 801 2055 2165 271 ...
## $ Item1        : int 5 3 4 4 4 3 6 2 5 2 ...
## $ Item2        : int 5 4 4 4 4 3 5 2 4 2 ...
## $ Item3        : int 5 3 2 4 4 3 6 2 4 2 ...
## $ Item4        : int 3 3 4 2 3 2 4 5 3 2 ...
## $ Item5        : int 4 4 4 5 4 4 1 2 4 5 ...
## $ Item6        : int 4 3 3 4 4 3 5 3 3 2 ...
## $ Item7        : int 3 4 3 3 4 3 5 4 4 3 ...
## $ Item8        : int 4 4 3 3 5 3 5 5 4 3 ...

```

```
summary(df)
```

```

##      CaseOrder      Customer_id      Interaction      UID
## Min.      :    1      Length:10000      Length:10000      Length:10000
## 1st Qu.: 2501      Class :character      Class :character      Class :character
## Median : 5000      Mode  :character      Mode  :character      Mode  :character
## Mean      : 5000
## 3rd Qu.: 7500
## Max.      :10000
##      City      State      County      Zip
## Length:10000      Length:10000      Length:10000      Min.      : 601
## Class :character      Class :character      Class :character      1st Qu.:26293
## Mode  :character      Mode  :character      Mode  :character      Median :48870
##                                     Mean  :49153
##                                     3rd Qu.:71867
##                                     Max.  :99929
##      Lat      Lng      Population      Area
## Min.      :17.97      Min.      : -171.69      Min.      :    0      Length:10000
## 1st Qu.:35.34      1st Qu.: -97.08      1st Qu.:   738      Class :character
## Median :39.40      Median : -87.92      Median :  2910      Mode  :character
## Mean      :38.76      Mean      : -90.78      Mean      :  9757

```

```

## 3rd Qu.:42.11 3rd Qu.: -80.09 3rd Qu.: 13168
## Max. :70.64 Max. : -65.67 Max. :111850
## TimeZone Job Children Age
## Length:10000 Length:10000 Min. : 0.000 Min. :18.00
## Class :character Class :character 1st Qu.: 0.000 1st Qu.:35.00
## Mode :character Mode :character Median : 1.000 Median :53.00
## Mean : 2.088 Mean :53.08
## 3rd Qu.: 3.000 3rd Qu.:71.00
## Max. :10.000 Max. :89.00
## Income Marital Gender Churn
## Min. : 348.7 Length:10000 Length:10000 Length:10000
## 1st Qu.: 19224.7 Class :character Class :character Class :character
## Median : 33170.6 Mode :character Mode :character Mode :character
## Mean : 39806.9
## 3rd Qu.: 53246.2
## Max. :258900.7
## Outage_sec_perweek Email Contacts Yearly_equip_failure
## Min. : 0.09975 Min. : 1.00 Min. :0.0000 Min. :0.000
## 1st Qu.: 8.01821 1st Qu.:10.00 1st Qu.:0.0000 1st Qu.:0.000
## Median :10.01856 Median :12.00 Median :1.0000 Median :0.000
## Mean :10.00185 Mean :12.02 Mean :0.9942 Mean :0.398
## 3rd Qu.:11.96949 3rd Qu.:14.00 3rd Qu.:2.0000 3rd Qu.:1.000
## Max. :21.20723 Max. :23.00 Max. :7.0000 Max. :6.000
## Techie Contract Port_modem Tablet
## Length:10000 Length:10000 Length:10000 Length:10000
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## InternetService Phone Multiple OnlineSecurity
## Length:10000 Length:10000 Length:10000 Length:10000
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## OnlineBackup DeviceProtection TechSupport StreamingTV
## Length:10000 Length:10000 Length:10000 Length:10000
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## StreamingMovies PaperlessBilling PaymentMethod Tenure
## Length:10000 Length:10000 Length:10000 Min. : 1.000
## Class :character Class :character Class :character 1st Qu.: 7.918
## Mode :character Mode :character Mode :character Median :35.431
## Mean :34.526
## 3rd Qu.:61.480
## Max. :71.999
## MonthlyCharge Bandwidth_GB_Year Item1 Item2
## Min. : 79.98 Min. : 155.5 Min. :1.000 Min. :1.000
## 1st Qu.:139.98 1st Qu.:1236.5 1st Qu.:3.000 1st Qu.:3.000

```

```
## Median :167.48    Median :3279.5    Median :3.000    Median :4.000
## Mean    :172.62    Mean    :3392.3    Mean    :3.491    Mean    :3.505
## 3rd Qu. :200.73    3rd Qu. :5586.1    3rd Qu. :4.000    3rd Qu. :4.000
## Max.    :290.16    Max.    :7159.0    Max.    :7.000    Max.    :7.000
##      Item3      Item4      Item5      Item6      Item7
## Min.    :1.000    Min.    :1.000    Min.    :1.000    Min.    :1.000    Min.    :1.00
## 1st Qu. :3.000    1st Qu. :3.000    1st Qu. :3.000    1st Qu. :3.000    1st Qu. :3.00
## Median  :3.000    Median  :3.000    Median  :3.000    Median  :3.000    Median  :4.00
## Mean    :3.487    Mean    :3.498    Mean    :3.493    Mean    :3.497    Mean    :3.51
## 3rd Qu. :4.000    3rd Qu. :4.000    3rd Qu. :4.000    3rd Qu. :4.000    3rd Qu. :4.00
## Max.    :8.000    Max.    :7.000    Max.    :7.000    Max.    :8.000    Max.    :7.00
##      Item8
## Min.    :1.000
## 1st Qu. :3.000
## Median  :3.000
## Mean    :3.496
## 3rd Qu. :4.000
## Max.    :8.000
```

```
# Renaming unclear columns named Item1 through Item8 for improved readability &
# confirming they have been renamed correctly
```

```
df <- df %>%
  rename(
    Response = Item1,
    Fix = Item2,
    Replacement = Item3,
    Reliability = Item4,
    Options = Item5,
    Respectful = Item6,
    Courteous = Item7,
    Listening = Item8
  )
```

```
colnames(df)
```

```
## [1] "CaseOrder"      "Customer_id"    "Interaction"
## [4] "UID"            "City"           "State"
## [7] "County"         "Zip"            "Lat"
## [10] "Lng"           "Population"     "Area"
## [13] "TimeZone"      "Job"            "Children"
## [16] "Age"           "Income"         "Marital"
## [19] "Gender"        "Churn"          "Outage_sec_perweek"
## [22] "Email"         "Contacts"       "Yearly_equip_failure"
## [25] "Techie"        "Contract"       "Port_modem"
## [28] "Tablet"        "InternetService" "Phone"
## [31] "Multiple"      "OnlineSecurity" "OnlineBackup"
## [34] "DeviceProtection" "TechSupport"   "StreamingTV"
## [37] "StreamingMovies" "PaperlessBilling" "PaymentMethod"
## [40] "Tenure"        "MonthlyCharge"  "Bandwidth_GB_Year"
## [43] "Response"      "Fix"            "Replacement"
## [46] "Reliability"   "Options"        "Respectful"
## [49] "Courteous"     "Listening"
```

```
# Getting numeric columns only
```

```
dfNumeric <- select_if(df, is.numeric)
```

```
str(dfNumeric)
```

```
## 'data.frame': 10000 obs. of 23 variables:
## $ CaseOrder : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Zip : int 99927 48661 97148 92014 77461 31030 37847 73109 34771 45237 ...
## $ Lat : num 56.3 44.3 45.4 33 29.4 ...
## $ Lng : num -133.4 -84.2 -123.2 -117.2 -95.8 ...
## $ Population : int 38 10446 3735 13863 11352 17701 2535 23144 17351 20193 ...
## $ Children : int 0 1 4 1 0 3 0 2 2 1 ...
## $ Age : int 68 27 50 48 83 83 79 30 49 86 ...
## $ Income : num 28562 21705 9610 18925 40074 ...
## $ Outage_sec_perweek : num 7.98 11.7 10.75 14.91 8.15 ...
## $ Email : int 10 12 9 15 16 15 10 16 20 18 ...
## $ Contacts : int 0 0 0 2 2 3 0 0 2 1 ...
## $ Yearly equip_failure: int 1 1 1 0 1 1 1 0 3 0 ...
## $ Tenure : num 6.8 1.16 15.75 17.09 1.67 ...
## $ MonthlyCharge : num 172 243 160 120 150 ...
## $ Bandwidth_GB_Year : num 905 801 2055 2165 271 ...
## $ Response : int 5 3 4 4 4 3 6 2 5 2 ...
## $ Fix : int 5 4 4 4 4 3 5 2 4 2 ...
## $ Replacement : int 5 3 2 4 4 3 6 2 4 2 ...
## $ Reliability : int 3 3 4 2 3 2 4 5 3 2 ...
## $ Options : int 4 4 4 5 4 4 1 2 4 5 ...
## $ Respectful : int 4 3 3 4 4 3 5 3 3 2 ...
## $ Courteous : int 3 4 3 3 4 3 5 4 4 3 ...
## $ Listening : int 4 4 3 3 5 3 5 5 4 3 ...
```

```
# Dropping unuseful numeric columns
```

```
to_drop <- c('CaseOrder', 'Zip', 'Lat', 'Lng')
```

```
dfDropped = dfNumeric[,!(names(dfNumeric) %in% to_drop)]
```

```
str(dfDropped)
```

```
## 'data.frame': 10000 obs. of 19 variables:
## $ Population : int 38 10446 3735 13863 11352 17701 2535 23144 17351 20193 ...
## $ Children : int 0 1 4 1 0 3 0 2 2 1 ...
## $ Age : int 68 27 50 48 83 83 79 30 49 86 ...
## $ Income : num 28562 21705 9610 18925 40074 ...
## $ Outage_sec_perweek : num 7.98 11.7 10.75 14.91 8.15 ...
## $ Email : int 10 12 9 15 16 15 10 16 20 18 ...
## $ Contacts : int 0 0 0 2 2 3 0 0 2 1 ...
## $ Yearly equip_failure: int 1 1 1 0 1 1 1 0 3 0 ...
## $ Tenure : num 6.8 1.16 15.75 17.09 1.67 ...
## $ MonthlyCharge : num 172 243 160 120 150 ...
## $ Bandwidth_GB_Year : num 905 801 2055 2165 271 ...
## $ Response : int 5 3 4 4 4 3 6 2 5 2 ...
## $ Fix : int 5 4 4 4 4 3 5 2 4 2 ...
## $ Replacement : int 5 3 2 4 4 3 6 2 4 2 ...
## $ Reliability : int 3 3 4 2 3 2 4 5 3 2 ...
## $ Options : int 4 4 4 5 4 4 1 2 4 5 ...
## $ Respectful : int 4 3 3 4 4 3 5 3 3 2 ...
## $ Courteous : int 3 4 3 3 4 3 5 4 4 3 ...
## $ Listening : int 4 4 3 3 5 3 5 5 4 3 ...
```

```
# Scaling data & exporting to CSV for submission
```

```
dfScaled <- as.data.frame(scale(dfDropped, center=TRUE, scale=TRUE))
summary(dfScaled)
```

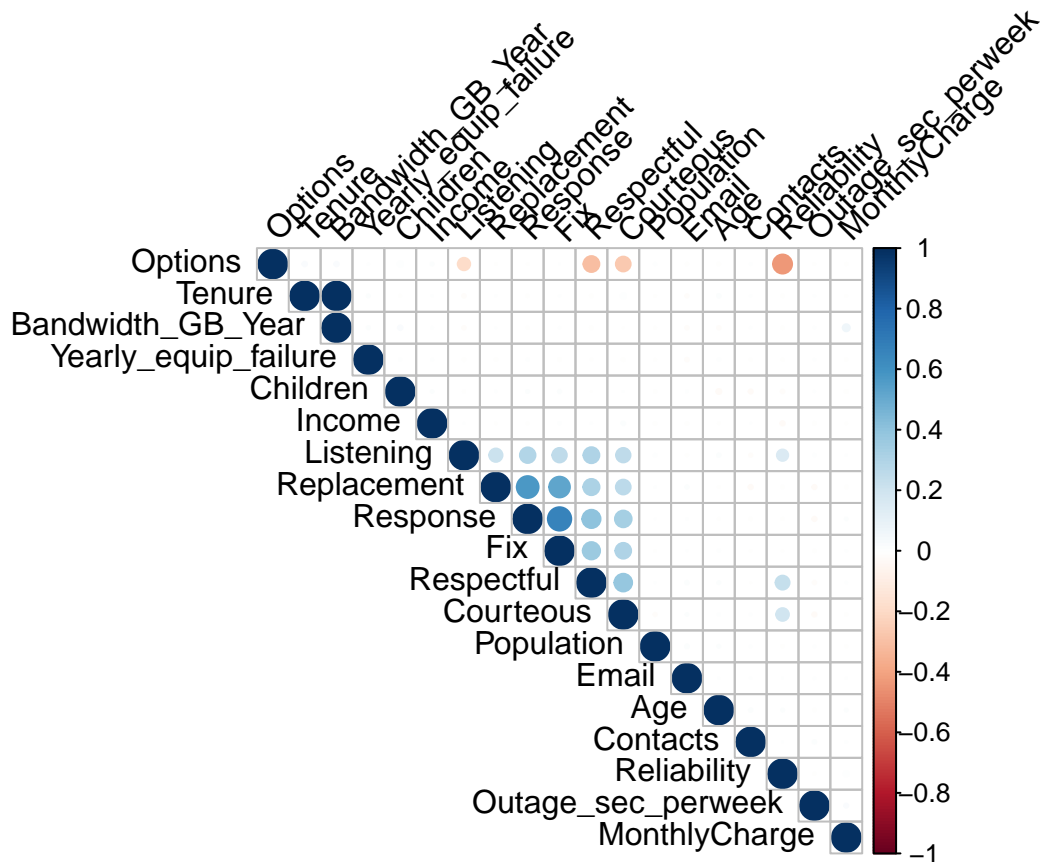
```
##      Population      Children      Age      Income
## Min.   :-0.6760  Min.   :-0.9723  Min.   :-1.694700  Min.   :-1.3992
## 1st Qu.: -0.6249  1st Qu.: -0.9723  1st Qu.: -0.873400  1st Qu.: -0.7299
## Median : -0.4743  Median : -0.5066  Median : -0.003788  Median : -0.2353
## Mean   :  0.0000  Mean   :  0.0000  Mean   :  0.000000  Mean   :  0.0000
## 3rd Qu.:  0.2364  3rd Qu.:  0.4249  3rd Qu.:  0.865825  3rd Qu.:  0.4766
## Max.   :  7.0738  Max.   :  3.6849  Max.   :  1.735437  Max.   :  7.7693
## Outage_sec_perweek      Email      Contacts
## Min.   :-3.327298  Min.   :-3.640572  Min.   :-1.005801
## 1st Qu.: -0.666539  1st Qu.: -0.666248  1st Qu.: -1.005801
## Median :  0.005616  Median : -0.005288  Median :  0.005868
## Mean   :  0.000000  Mean   :  0.000000  Mean   :  0.000000
## 3rd Qu.:  0.661164  3rd Qu.:  0.655673  3rd Qu.:  1.017537
## Max.   :  3.765225  Max.   :  3.629997  Max.   :  6.075882
## Yearly equip_failure      Tenure      MonthlyCharge      Bandwidth_GB_Year
## Min.   :-0.6258  Min.   :-1.2679  Min.   :-2.1574  Min.   :-1.48119
## 1st Qu.: -0.6258  1st Qu.: -1.0063  1st Qu.: -0.7602  1st Qu.: -0.98654
## Median : -0.6258  Median :  0.0342  Median : -0.1197  Median : -0.05162
## Mean   :  0.0000  Mean   :  0.0000  Mean   :  0.0000  Mean   :  0.00000
## 3rd Qu.:  0.9466  3rd Qu.:  1.0193  3rd Qu.:  0.6546  3rd Qu.:  1.00389
## Max.   :  8.8088  Max.   :  1.4171  Max.   :  2.7370  Max.   :  1.72363
##      Response      Fix      Replacement      Reliability
## Min.   :-2.4001  Min.   :-2.4212  Min.   :-2.4193  Min.   :-2.4346
## 1st Qu.: -0.4729  1st Qu.: -0.4882  1st Qu.: -0.4737  1st Qu.: -0.4850
## Median : -0.4729  Median :  0.4783  Median : -0.4737  Median : -0.4850
## Mean   :  0.0000  Mean   :  0.0000  Mean   :  0.0000  Mean   :  0.0000
## 3rd Qu.:  0.4907  3rd Qu.:  0.4783  3rd Qu.:  0.4990  3rd Qu.:  0.4899
## Max.   :  3.3814  Max.   :  3.3779  Max.   :  4.3902  Max.   :  3.4144
##      Options      Respectful      Courteous      Listening
## Min.   :-2.4325  Min.   :-2.4162  Min.   :-2.4400  Min.   :-2.4261
## 1st Qu.: -0.4810  1st Qu.: -0.4811  1st Qu.: -0.4954  1st Qu.: -0.4818
## Median : -0.4810  Median : -0.4811  Median :  0.4769  Median : -0.4818
## Mean   :  0.0000  Mean   :  0.0000  Mean   :  0.0000  Mean   :  0.0000
## 3rd Qu.:  0.4948  3rd Qu.:  0.4864  3rd Qu.:  0.4769  3rd Qu.:  0.4904
## Max.   :  3.4222  Max.   :  4.3564  Max.   :  3.3938  Max.   :  4.3790
```

```
write.csv(dfScaled, 'C:/WGU/D212 Data Mining II/churn_pca.csv', row.names = FALSE)
```

```
# Correlation matrix
```

```
corr_df <- round(cor(dfScaled),3)
```

```
corrplot(corr_df, method="circle", type="upper", order="hclust", tl.col="black",
          tl.srt=45)
```



```
pca_results <- prcomp(dfDropped, center=TRUE, scale.=TRUE)
summary(pca_results)
```

```
## Importance of components:
##          PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation  1.7171 1.4131 1.27928 1.02806 1.0148 1.00957 1.00234
## Proportion of Variance 0.1552 0.1051 0.08613 0.05563 0.0542 0.05364 0.05288
## Cumulative Proportion 0.1552 0.2603 0.34642 0.40205 0.4562 0.50990 0.56278
##          PC8      PC9      PC10     PC11     PC12     PC13     PC14
## Standard deviation  0.99960 0.99576 0.99027 0.98204 0.98024 0.88242 0.83074
## Proportion of Variance 0.05259 0.05219 0.05161 0.05076 0.05057 0.04098 0.03632
## Cumulative Proportion 0.61537 0.66755 0.71917 0.76992 0.82050 0.86148 0.89780
##          PC15     PC16     PC17     PC18     PC19
## Standard deviation  0.76960 0.7333 0.69407 0.56972 0.07389
## Proportion of Variance 0.03117 0.0283 0.02535 0.01708 0.00029
## Cumulative Proportion 0.92897 0.9573 0.98263 0.99971 1.00000
```

```
pca_results$rotation <- -1*pca_results$rotation
pca_results$rotation
```

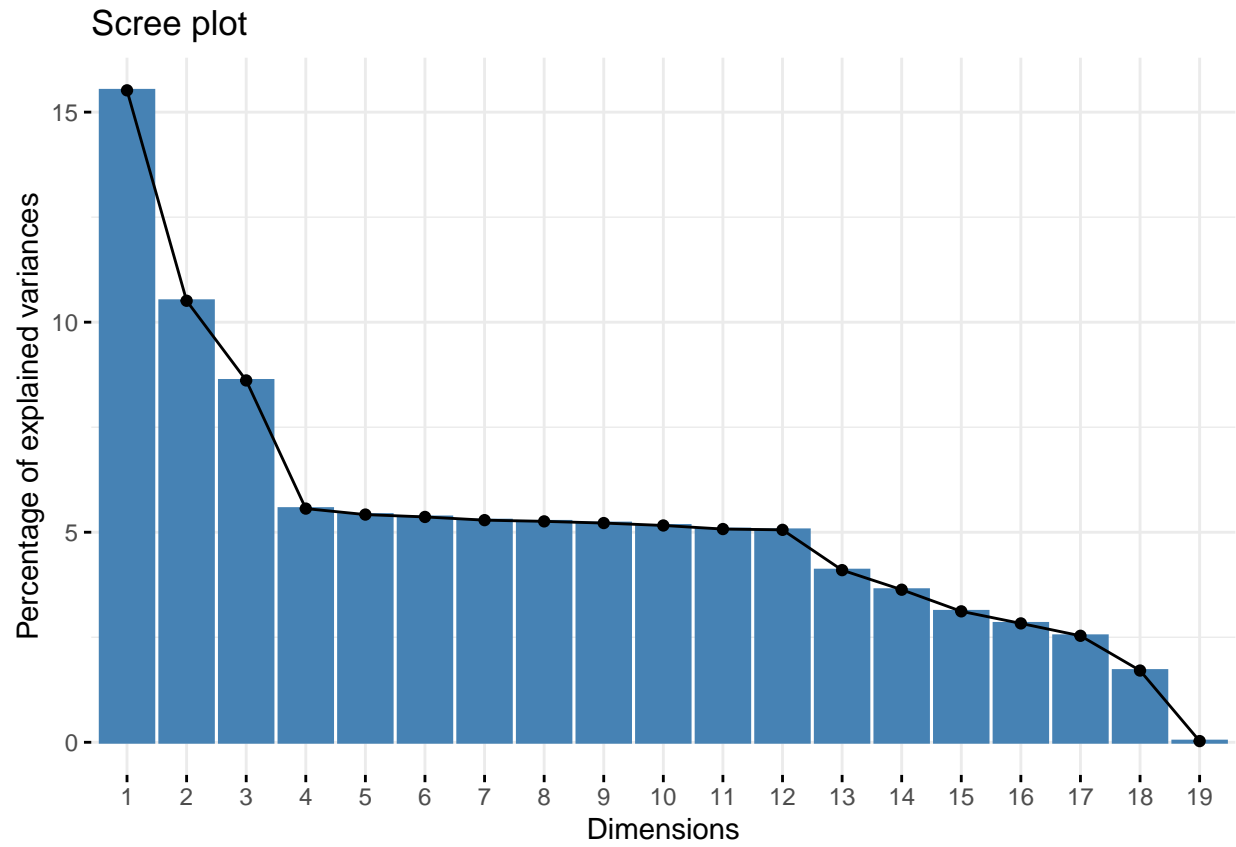
```
##          PC1      PC2      PC3      PC4
## Population -0.0021086454  0.005463171 -0.014732189 -0.2921508283
## Children   0.0040718922 -0.015861897 -0.028393430  0.5105694425
## Age        0.0064594747 -0.000293987  0.029318595 -0.4552966907
## Income     0.0010381550 -0.006034733 -0.025865385  0.2520653129
## Outage_sec_perweek -0.0175162123 -0.003927019  0.014362795 -0.2201147135
## Email      0.0087441339  0.020609151  0.003458796 -0.1904497730
```


## Contacts	-0.0087605031	-0.003317823	0.011853356	-0.4207313665
## Yearly_equip_failure	-0.0076883195	-0.017603848	-0.008198934	0.1675155298
## Tenure	-0.0163204446	-0.702322845	0.063085418	-0.0053546302
## MonthlyCharge	0.0009298264	-0.039857997	0.009499134	-0.2986903703
## Bandwidth_GB_Year	-0.0168453471	-0.703830962	0.062131817	0.0050678440
## Response	0.4586695657	-0.031339869	-0.280974313	-0.0109515532
## Fix	0.4338473173	-0.038755098	-0.282381287	-0.0201218208
## Replacement	0.4004876402	-0.035504042	-0.280527315	-0.0041077791
## Reliability	0.1458016038	0.039379777	0.568452431	0.0151424325
## Options	-0.1756983261	-0.056277863	-0.587090012	-0.0388985402
## Respectful	0.4050803054	0.006647810	0.183525296	-0.0009102063
## Courteous	0.3582462918	-0.002051209	0.181336918	0.0311939124
## Listening	0.3087328379	0.013633728	0.131655016	-0.0284346959
##	PC5	PC6	PC7	PC8
## Population	0.264958477	-0.402354910	0.3558635194	0.329127870
## Children	0.345310100	0.089376147	0.1190693067	0.226847336
## Age	-0.417933316	-0.183901702	0.1527519300	-0.024112986
## Income	-0.285029607	0.084983253	-0.4296106738	0.581477278
## Outage_sec_perweek	0.339481825	0.591283796	0.2735274188	0.262606602
## Email	0.519450086	-0.319498424	-0.1031174950	0.170129350
## Contacts	-0.124576748	0.146366053	-0.2752017218	0.508824267
## Yearly_equip_failure	-0.373154802	0.147091956	0.6864651477	0.241920647
## Tenure	-0.007568246	-0.048576026	0.0000158926	0.007553682
## MonthlyCharge	0.113921365	0.537630534	-0.1125586326	-0.284655211
## Bandwidth_GB_Year	0.022808129	-0.005062637	-0.0091323712	-0.001688322
## Response	-0.002667456	0.016377246	0.0186840354	-0.013618432
## Fix	-0.002283842	0.022965836	0.0005498379	-0.001653219
## Replacement	0.013294223	-0.022177239	0.0080475193	-0.032816999
## Reliability	0.001544489	-0.013908905	0.0108083748	-0.024874776
## Options	-0.023274666	-0.008693152	-0.0107006227	-0.011609931
## Respectful	0.002245107	-0.002666089	0.0009183955	0.025174950
## Courteous	-0.022613935	-0.010414335	-0.0593106051	0.048237629
## Listening	-0.001572577	0.034952633	0.0441293485	0.010773728
##	PC9	PC10	PC11	PC12
## Population	-0.161654293	0.580377704	-0.167295458	0.2290215524
## Children	-0.155911568	-0.175953066	-0.655598509	-0.2419736191
## Age	-0.346065788	-0.180480843	-0.234747982	-0.5908293135
## Income	-0.449648955	0.219832910	0.252659489	-0.0576742724
## Outage_sec_perweek	0.149556963	0.125520643	0.319263184	-0.4395357519
## Email	-0.290785056	-0.592268314	0.328651594	0.0611451660
## Contacts	0.434373260	-0.248703385	-0.371468450	0.2415479723
## Yearly_equip_failure	-0.114546678	-0.334113203	0.146135720	0.3653939567
## Tenure	0.028779847	-0.001590283	0.028600306	-0.0271470396
## MonthlyCharge	-0.562547351	0.029519379	-0.228176121	0.3751869589
## Bandwidth_GB_Year	-0.001653786	0.001271037	0.000742929	0.0088735824
## Response	0.012424039	-0.012898774	-0.007946334	0.0222754461
## Fix	0.014344850	-0.006536492	-0.009541507	-0.0138649714
## Replacement	0.017964951	-0.011037518	0.020962160	0.0009524185
## Reliability	0.020862649	-0.000102103	-0.008042486	0.0237632945
## Options	0.013279644	0.001989049	0.008549853	-0.0148290855
## Respectful	-0.009561595	-0.007607818	-0.001311049	0.0171871209
## Courteous	0.001604295	-0.022372911	0.005069962	0.0113863304
## Listening	-0.014506921	0.097526100	0.018232117	-0.0657554846
##	PC13	PC14	PC15	PC16

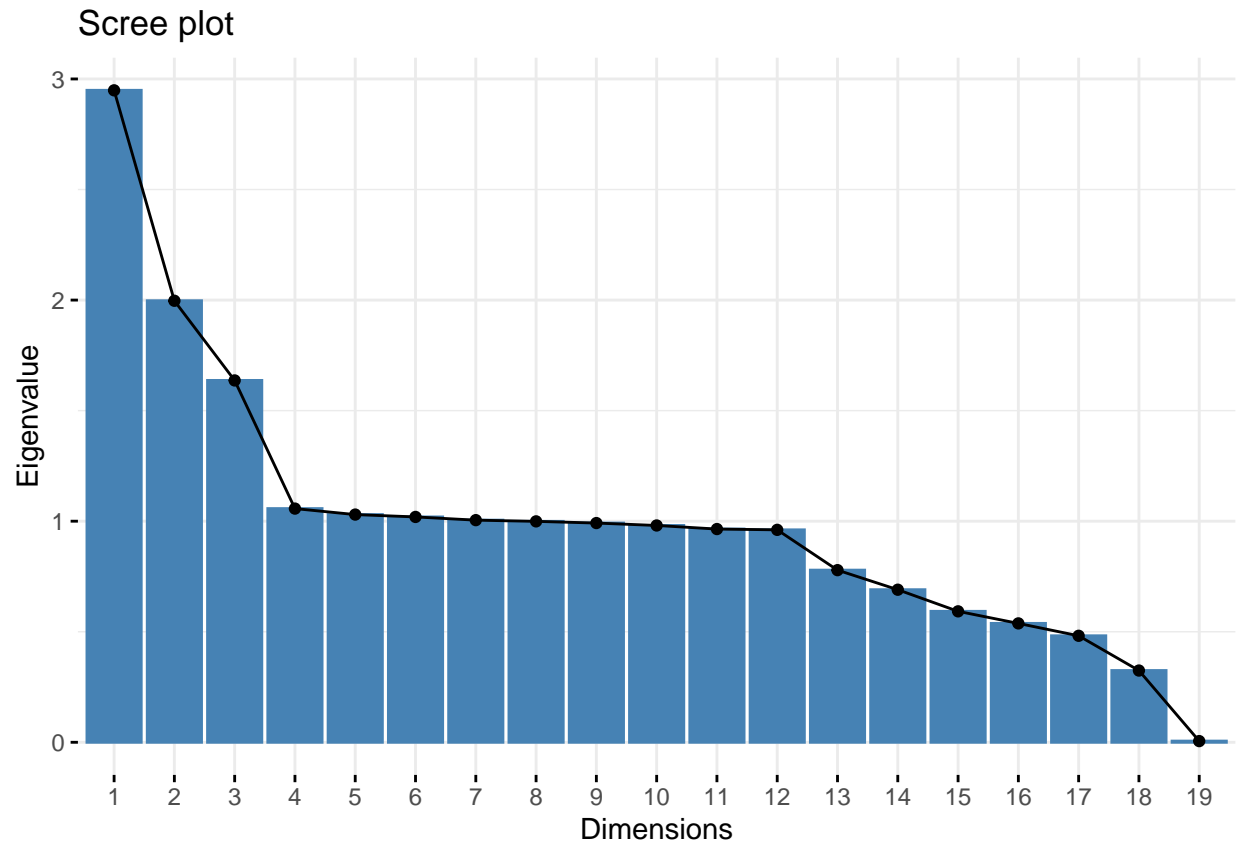
## Population	-0.057129940	0.019141937	-0.016175605	-1.209891e-03
## Children	0.017019942	-0.012250417	-0.014011113	-1.449032e-02
## Age	-0.045335881	0.002513369	-0.002495280	9.404548e-03
## Income	-0.020483920	-0.079017629	-0.007573495	2.560819e-03
## Outage_sec_perweek	-0.089843591	0.016926153	-0.008716492	-1.352929e-02
## Email	0.061158111	-0.017174988	-0.016341738	-6.448880e-03
## Contacts	0.044032025	-0.035285184	-0.003279351	2.649822e-02
## Yearly_equip_failure	0.020739444	0.006446116	-0.015852702	1.308066e-03
## Tenure	0.005939840	-0.003507182	0.006547800	7.772792e-03
## MonthlyCharge	-0.005920075	0.014551470	-0.016507844	6.813987e-05
## Bandwidth_GB_Year	0.010519731	-0.003326133	0.005611735	6.118572e-03
## Response	-0.069228332	-0.116809968	-0.046422155	-2.505747e-02
## Fix	-0.111545233	-0.170009566	-0.066138621	-7.391672e-02
## Replacement	-0.176045339	-0.249291017	-0.147591443	3.958747e-01
## Reliability	-0.173904798	-0.480655380	-0.442505121	-4.319329e-01
## Options	0.137294250	0.057896362	-0.206301634	-6.940889e-01
## Respectful	-0.060350283	0.062041440	0.759347027	-4.004522e-01
## Courteous	-0.170668511	0.804890093	-0.377414559	-7.132317e-02
## Listening	0.921694139	-0.018910668	-0.113106661	4.565779e-02
##	PC17	PC18	PC19	
## Population	-0.005661352	-0.0023563664	3.216561e-04	
## Children	0.020915032	-0.0009482978	2.161475e-02	
## Age	0.005784076	0.0136963594	-2.242136e-02	
## Income	0.005300682	0.0134658988	9.098215e-04	
## Outage_sec_perweek	0.018261706	0.0135163066	-3.612503e-04	
## Email	-0.017252809	0.0009614875	-2.258113e-04	
## Contacts	0.020255423	-0.0008129168	9.484014e-04	
## Yearly_equip_failure	0.007488293	-0.0214479090	1.445828e-04	
## Tenure	-0.004624881	0.0075193652	7.052512e-01	
## MonthlyCharge	0.021493879	-0.0120069486	4.577805e-02	
## Bandwidth_GB_Year	-0.002187745	0.0018153234	-7.067798e-01	
## Response	-0.240544956	0.7929645672	-2.979455e-03	
## Fix	-0.590695793	-0.5735473737	1.144489e-03	
## Replacement	0.673812348	-0.1768627897	-7.687323e-05	
## Reliability	0.088482694	0.0186859533	-1.050817e-04	
## Options	0.264885581	-0.0418186207	8.111189e-04	
## Respectful	0.229253460	-0.0638088587	5.934503e-04	
## Courteous	0.066812067	-0.0413240284	-4.859041e-04	
## Listening	0.045412038	-0.0432352528	1.993519e-03	

Scree plot

```
fviz_eig(pca_results, choice = "variance", ncp=19)
```

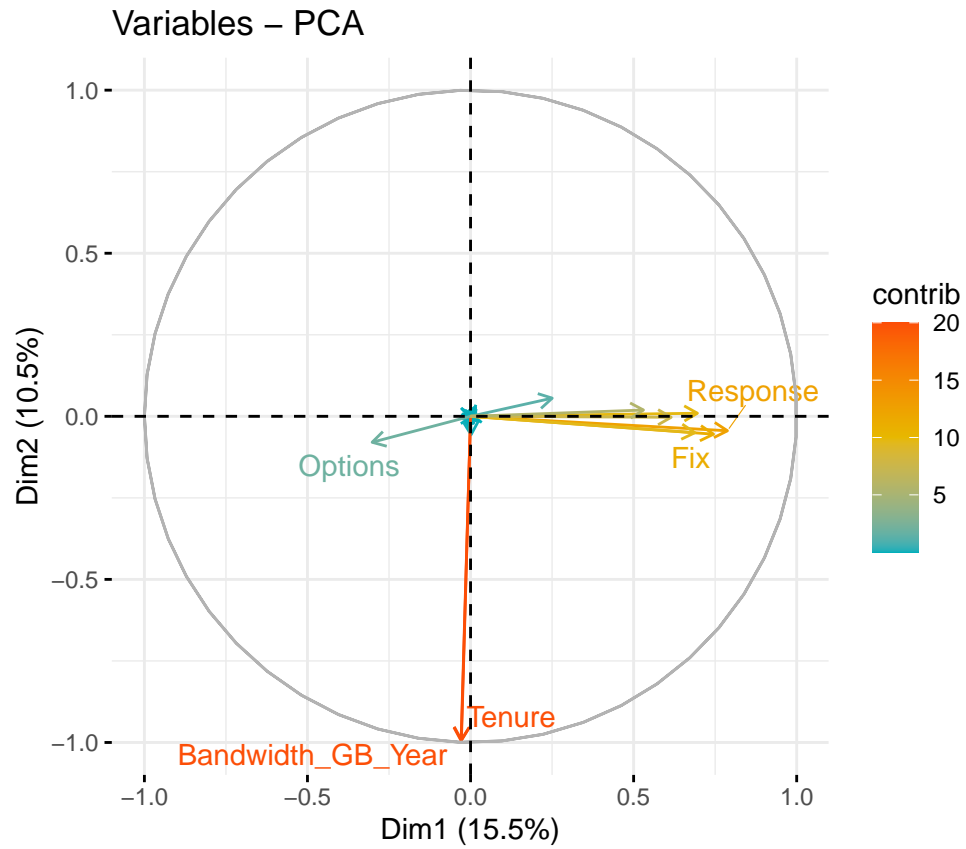


```
fviz_eig(pca_results, choice = "eigenvalue", ncp=19)
```



```
fviz_pca_var(pca_results,  
  col.var = "contrib",  
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
  repel=TRUE)
```

```
## Warning: ggrepel: 14 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```



```
fviz_pca_biplot(pca_results, col.var = "#2E9FDF", col.ind = "#696969")
```

A PCA plot showing the relationship between Dim1 (15.5%) on the x-axis and Dim2 (10.5%) on the y-axis. The plot is divided into four quadrants by dashed lines. Blue arrows indicate the direction of increasing values for 'Reliability', 'Bandwidth', and 'CPU_Yr'.

##	eigenvalue	variance.percent	cumulative.variance.percent
## Dim.1	2.948567317	15.51877535	15.51878
## Dim.2	1.996911994	10.51006313	26.02884
## Dim.3	1.636553603	8.61344002	34.64228
## Dim.4	1.056914874	5.56270986	40.20499
## Dim.5	1.029893736	5.42049335	45.62548
## Dim.6	1.019236768	5.36440404	50.98989
## Dim.7	1.004692280	5.28785410	56.27774
## Dim.8	0.999197279	5.25893305	61.53667
## Dim.9	0.991529871	5.21857827	66.75525
## Dim.10	0.980640393	5.16126523	71.91652
## Dim.11	0.964401126	5.07579540	76.99231
## Dim.12	0.960877775	5.05725145	82.04956
## Dim.13	0.778658275	4.09820145	86.14776
## Dim.14	0.690135022	3.63228959	89.78005
## Dim.15	0.592279454	3.11726029	92.89731
## Dim.16	0.537746193	2.83024312	95.72756
## Dim.17	0.481727316	2.53540693	98.26296
## Dim.18	0.324577335	1.70830176	99.97127
## Dim.19	0.005459387	0.02873361	100.00000