4th Problem Set

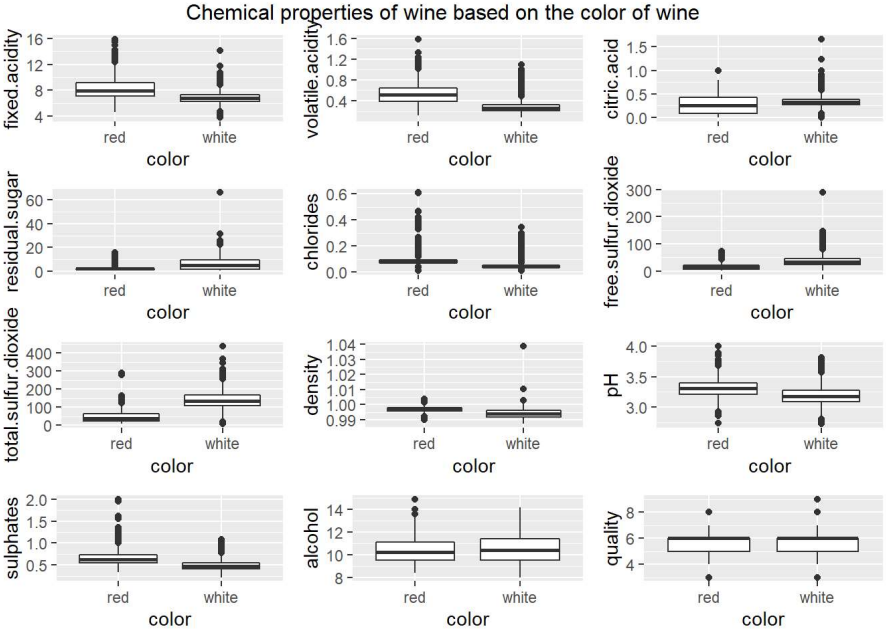
Fjolle Gjonbalaj

07/05/2021

Question 1: Clustering and PCA

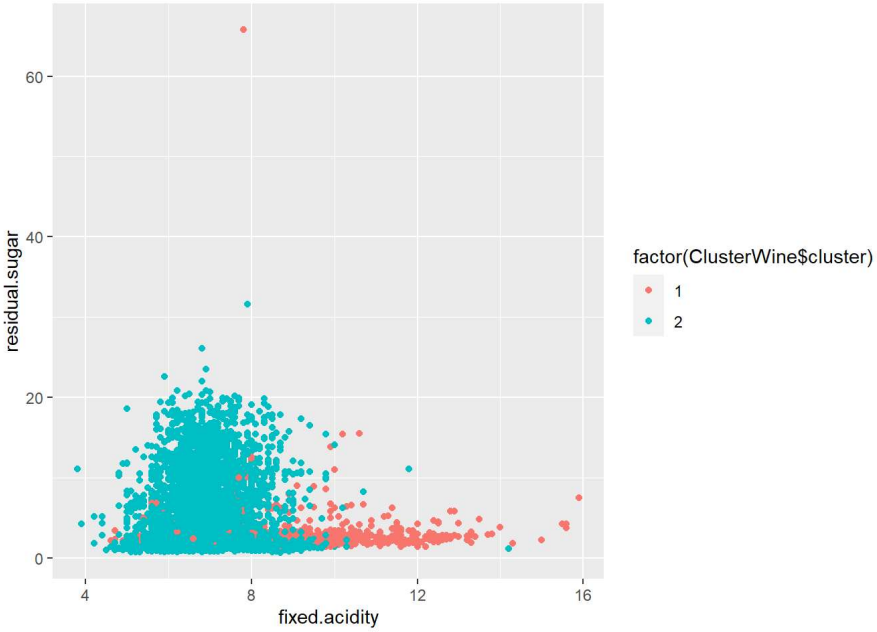
This problem compares unsupervised learning methods like Principal Component Analysis and K-means clustering to find out whether or not they can successfully distinguish the red wine from the white wine, as well as sorting the higher from the lower quality of wine based on the dataset with 11 chemical properties (or suitable transformations thereof) of 6500 different bottles of vinho verde wine from northern Portugal.

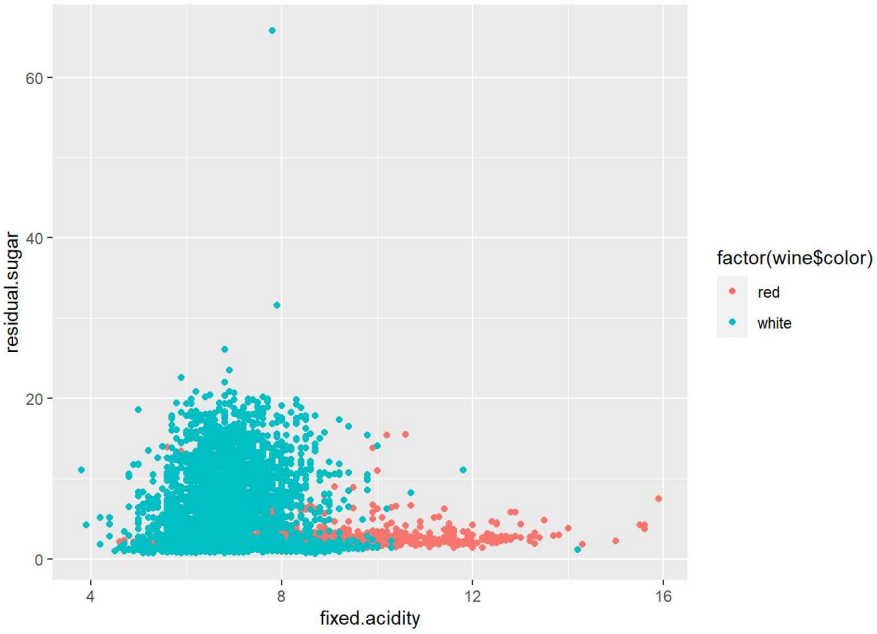
Unsupervised learning methods to distinguish red wine from white wine

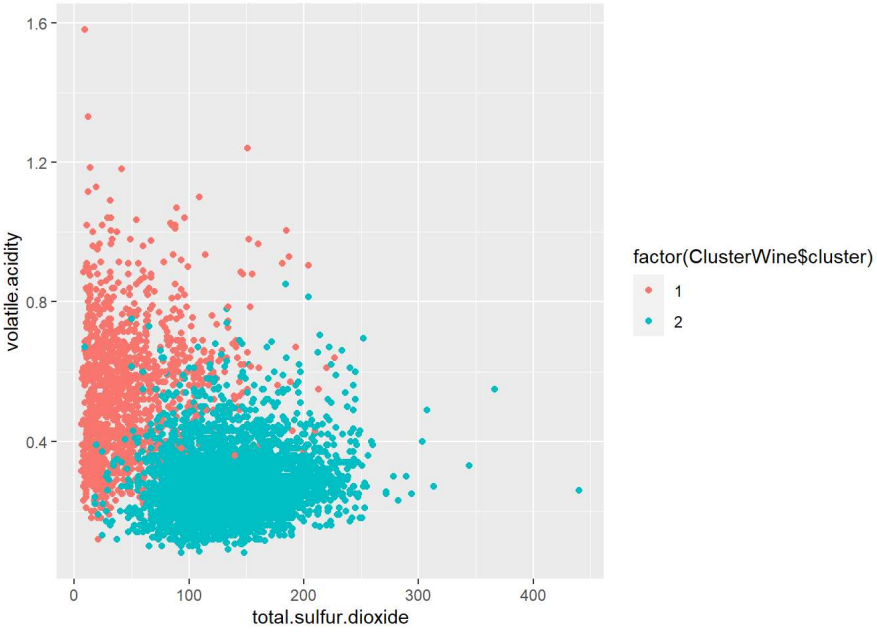


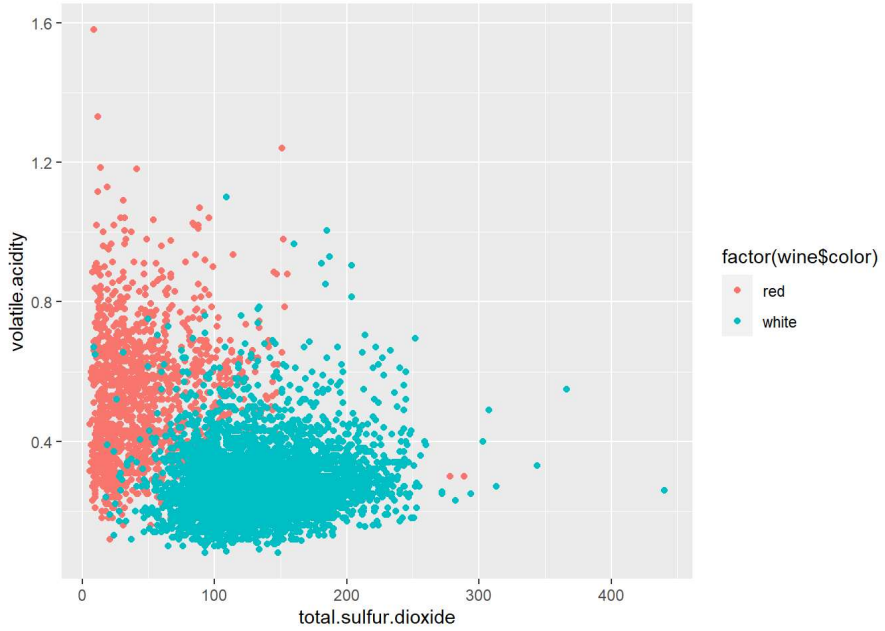
The boxplots show that red and white wine are distinct from one-another when it comes to the chemical components such as total sulfur dioxide, residual sugar, volatile acidity and fixed acidity. I use these features in order to analyze the performance of the selected methods.In the figure with boxplots the wine colors are not very different in the other chemical features. As such, a visual graphic is chosen to show the performance of K-means clustering on the selected chemical features.

K-means clustering


residual.sugar
fixed.acidity
factor(ClusterWine$cluster)
1
2


residual.sugar
fixed.acidity
factor(wine$color)
red
white


volatile.acidity
total.sulfur.dioxide
factor(ClusterWine$cluster)
1
2



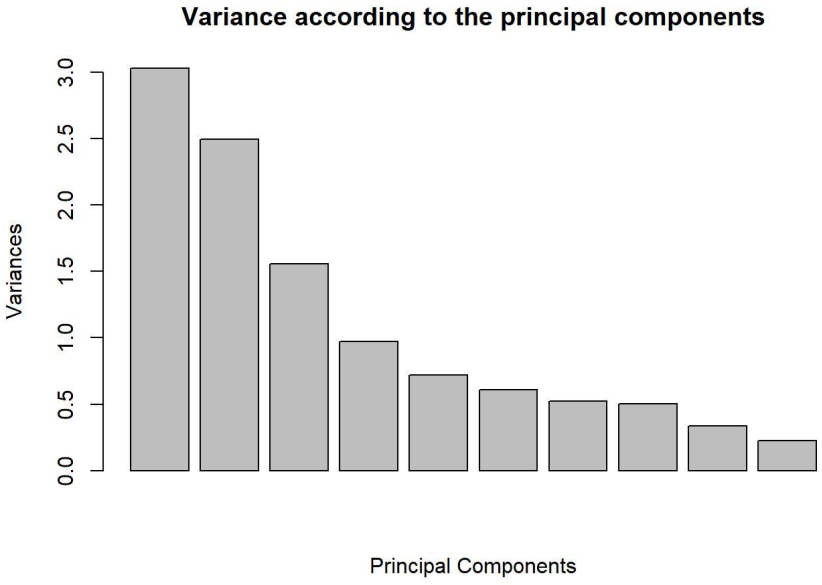
The above figures are a visual representation of the K-means clustering, and how the latter is able to successfully differentiate the colors of wine with respect to the chemical properties.

Prinicipal Components Analysis (PCA) by wine colors

Principal Components Analysis performs in a similar manner to distinguish wine by the color. The table below displays that the first five principal components explain the cumulative variation in the data.

PCA by wine colors

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
Standard deviation	1.740652	1.579185	1.247536	0.985166	0.8484544	0.7793021	0.7232971	0.7081739	0.5805377	0.4771748	0.1811927
Proportion of Variance	0.275440	0.226710	0.141490	0.088230	0.0654400	0.0552100	0.0475600	0.0455900	0.0306400	0.0207000	0.0029800
Cumulative Proportion	0.275440	0.502150	0.643640	0.731870	0.7973200	0.8525300	0.9000900	0.9456800	0.9763200	0.9970200	1.0000000

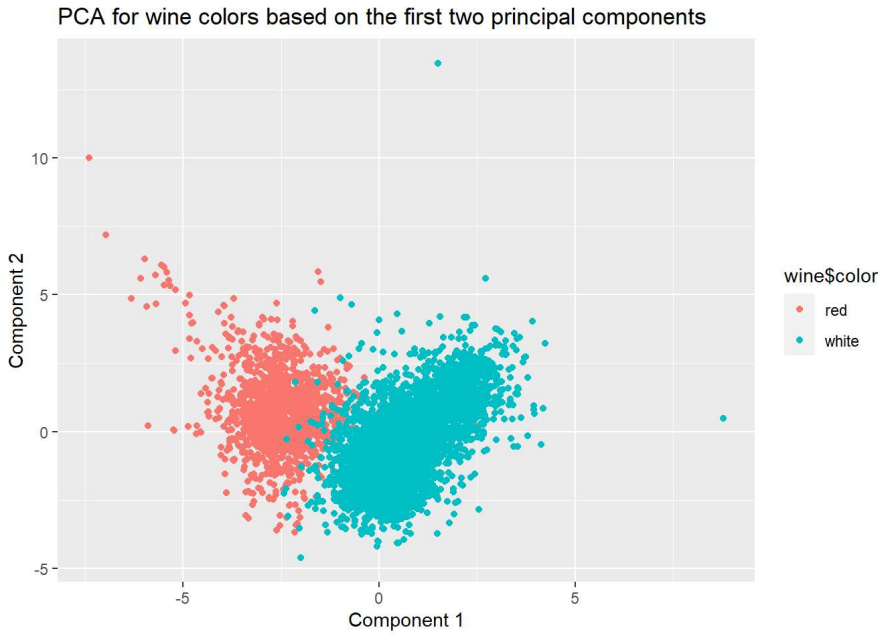


Loadings of the first five Principal Components

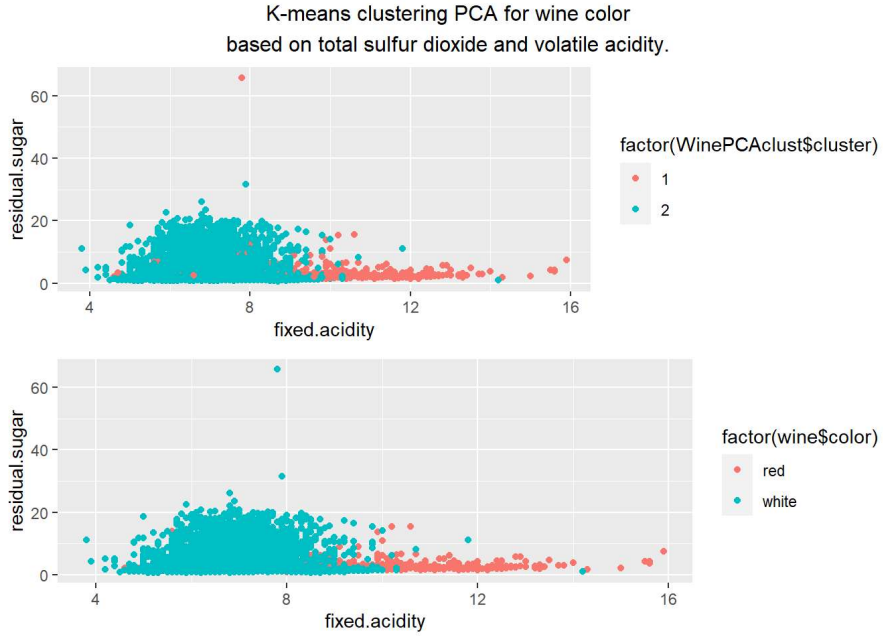
	PC1	PC2	PC3	PC4	PC5
fixed.acidity	-0.24	0.34	-0.43	0.16	-0.15
volatile.acidity	-0.38	0.12	0.31	0.21	0.15

	PC1	PC2	PC3	PC4	PC5
citric.acid	0.15	0.18	-0.59	-0.26	-0.16
residual.sugar	0.35	0.33	0.16	0.17	-0.35
chlorides	-0.29	0.32	0.02	-0.24	0.61
free.sulfur.dioxide	0.43	0.07	0.13	-0.36	0.22
total.sulfur.dioxide	0.49	0.09	0.11	-0.21	0.16
density	-0.04	0.58	0.18	0.07	-0.31
pH	-0.22	-0.16	0.46	-0.41	-0.45
sulphates	-0.29	0.19	-0.07	-0.64	-0.14
alcohol	-0.11	-0.47	-0.26	-0.11	-0.19

The loadings for the first five principal components are the linear combinations of the original chemical properties. For instance, free sulfur dioxide and total sulfur dioxide are strongly positively correlated with the first principal component, whereas volatile acidity and sulfate are strongly and negatively correlated with the first principal component.



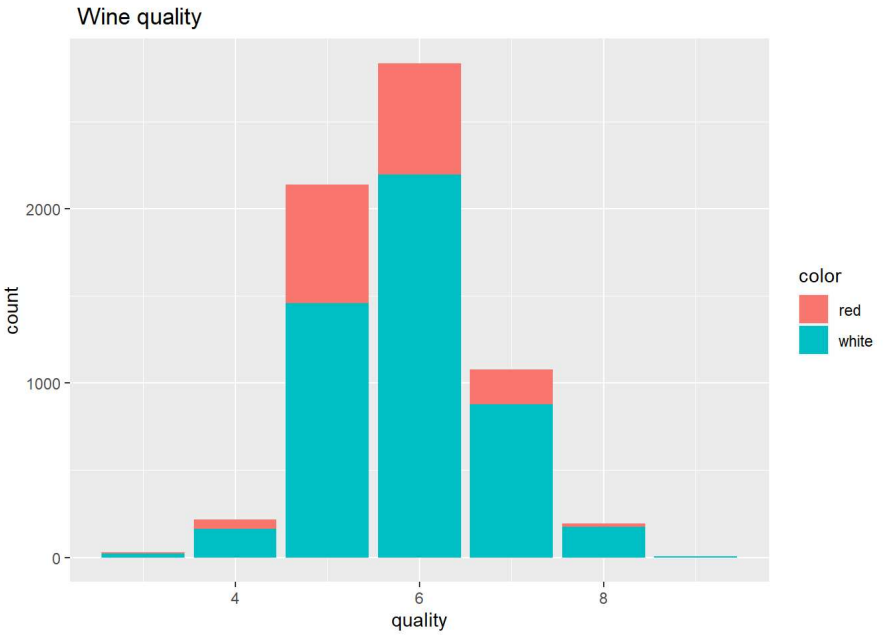
The plot above shows that PCA can also perform the clustering based on the type of wine. It can be seen that white wine generally clusters around positive values in the dimension of the first principal component(although not always), whereas red wine generally clusters around the negative values. The K-means clustering on the first five principal components can be used to check whether or not we can improve the partitioning of wine color based on K-means. This can be successfully performed by augmenting the K-means clustering.



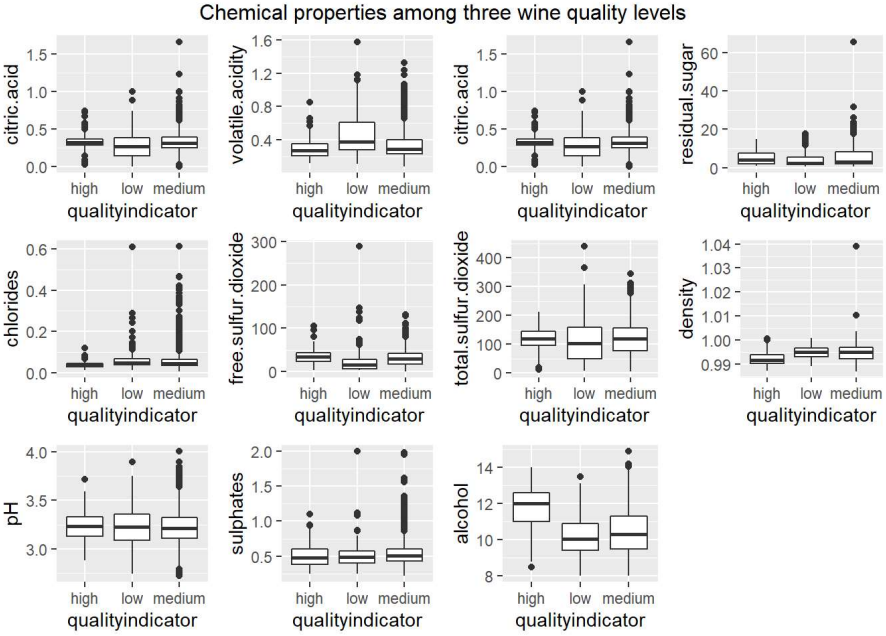
In-sample fit of two clusters

Between-cluster SS of Basic clustering 15320.72
Between-cluster SS of Clustering based PCA 15212.77
It can be shown that both PCA and K-means can successfully make the difference when it comes to the color of the wine using unsupervised learning on chemical properties. However, by first reducing the dimensions of the features by PCA and then implementing K-means would result in the colors being even better differentiated. This results in reduced within-cluster sum of squares.

Unsupervised learning on the quality of wine

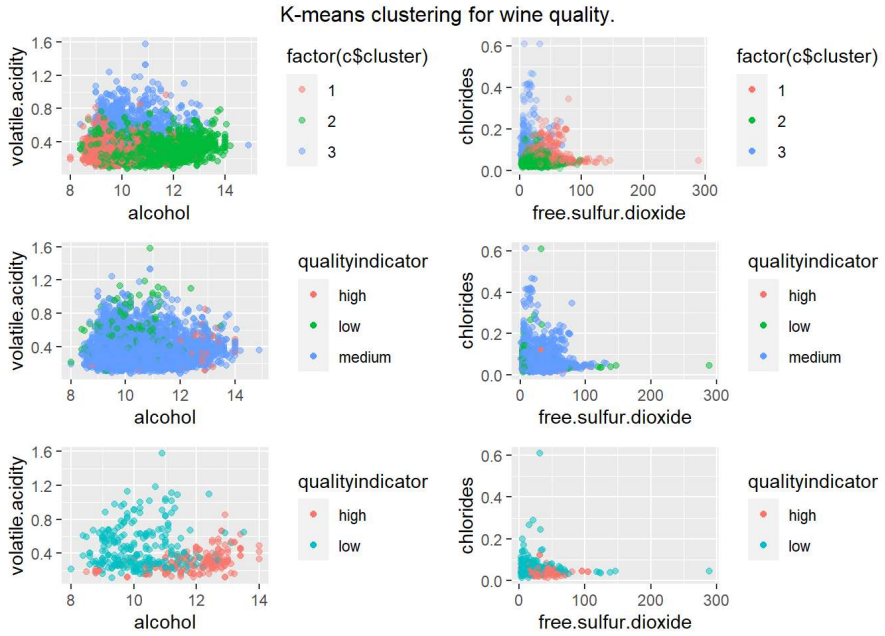


The above figure shows that there are 7 different qualities of wines that the data contains. We can further define the higher-quality wine as wine that takes a score higher than 7;The medium-quality wine takes a value somewhere between 4 and 7; whereas the lower quality wine takes a score of lower than 4. We can then further explore which chemical components of wine quality are the most different among each other.

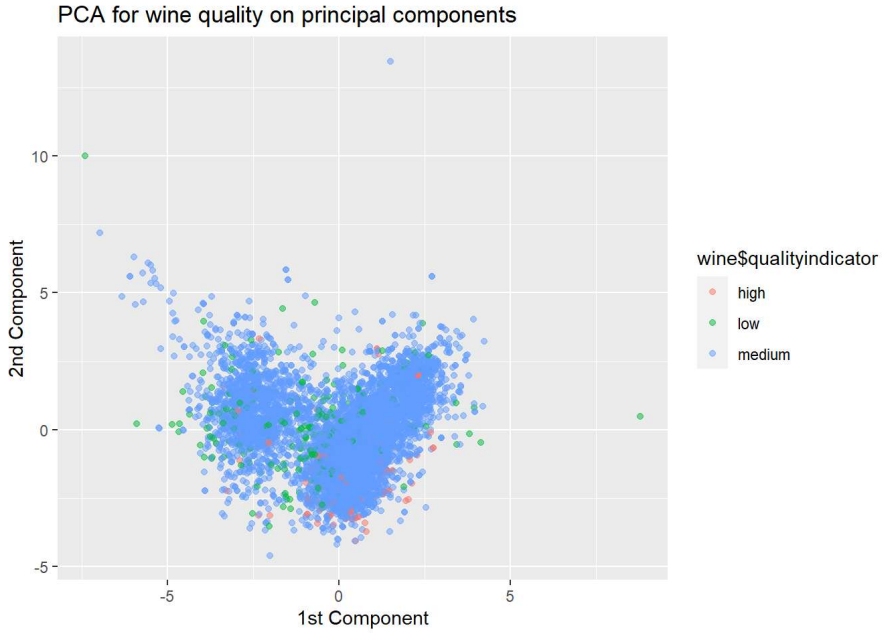


It is seen in the boxplots above that higher and lower quality wine are the ones that vary the most based on the chemical component.

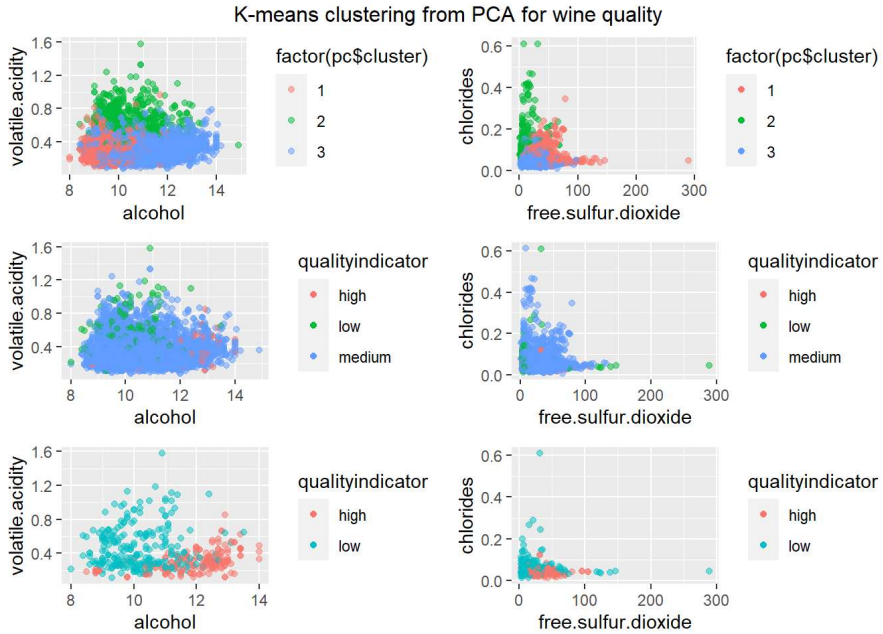
K-means clustering on the quality levels of wine



Prinicipal Components Analysis (PCA) on higher and lower quality wine



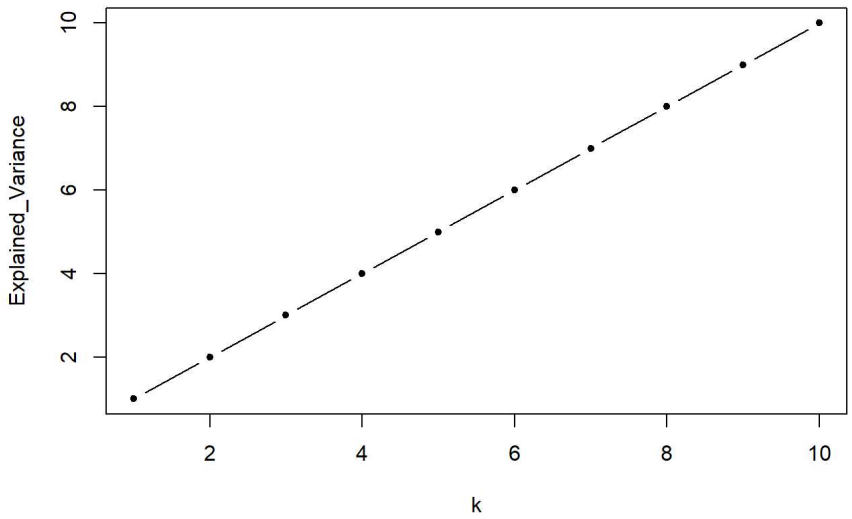
It is observed from the plot above that higher quality wine generally cluster to the right of the principal components, while the lower quality wine generally cluster to the left of them.



K-means clustering on the first five principal components is used to check whether or not partitioning of wine quality according to K-means can be improved upon. The figure above shows how k-means clustering partitions the data into three clusters in the dimensions of volatile acidity and alcohol. The middle left plot shows how the three wine quality levels are different in these dimensions. It can be difficult to observe the difference between the lower and higher quality wine due to a large number of middle quality wine that overlap. To conclude, neither K-means clustering nor PCA successfully differentiate the higher from the lower quality wines using only unsupervised learning contained on chemical properties of wine. Nevertheless, reducing the dimensions of the features by PCA and then applying K-means clustering could make the successful distinction on the quality levels of wine which could be observed from the reduced within-cluster sum of squares. Hence, K-means clustering together with PCA on the chemical components of wine work more successfully to differentiate between wine color than when we compare them based on the wine quality.

Question 2: Market segmentation

K-means



The figure above shows the elbow plot. It is not clear from the graph what the appropriate elbow is.

Result from K-means

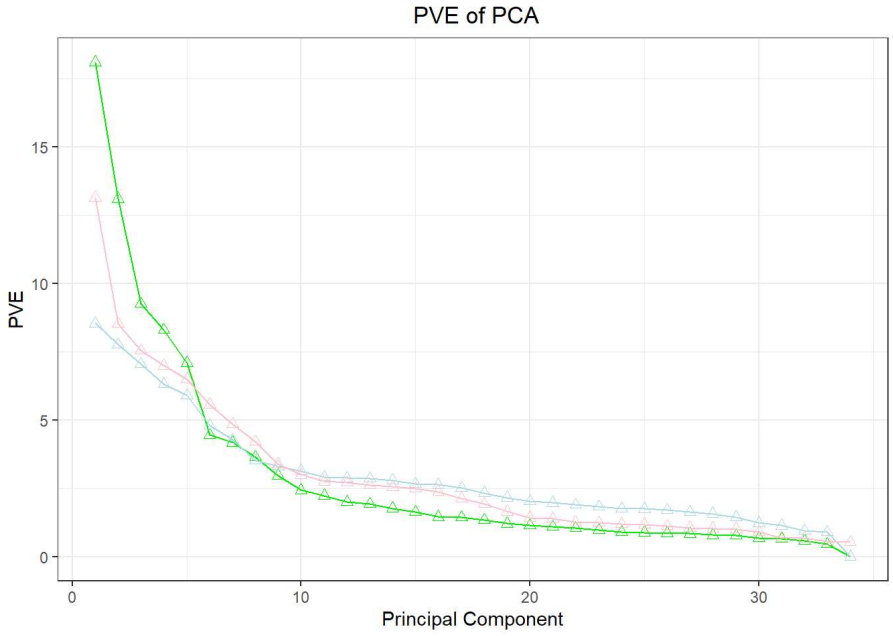
Clusters of K-means

	Category a	Category b	Category c	Category d	Category e	Consumer_no.	NA	NA	NA
Cluster a	sports_fandom	religion	food	chatter	parenting	family	school	photo_sharing	curre
Cluster b	cooking	photo_sharing	fashion	chatter	beauty	current_events	health_nutrition	travel	shop

	Category a	Category b	Category c	Category d	Category e	Consumer_no.	NA	NA	NA
Cluster c	chatter	tv_film	college_uni	current_events	music	photo_sharing	uncategorized	travel	sport
Cluster d	news	politics	chatter	automotive	sports_fandom	current_events	photo_sharing	travel	family
Cluster e	politics	travel	chatter	computers	news	current_events	photo_sharing	food	healtl
Cluster f	college_uni	online_gaming	chatter	photo_sharing	sports_playing	current_events	travel	health_nutrition	sport
Cluster g	art	tv_film	chatter	current_events	travel	photo_sharing	politics	food	uncat
Cluster h	dating	chatter	photo_sharing	fashion	school	current_events	travel	uncategorized	healtl
Cluster i	chatter	photo_sharing	shopping	current_events	travel	politics	sports_fandom	uncategorized	healtl
Cluster j	health_nutrition	personal_fitness	chatter	cooking	outdoors	photo_sharing	food	current_events	trave

K-means gave stable clusters of consumers. For each cluster, the sum of frequencies for each category over all consumers in that cluster is calculated. After ordering the categories by sums of frequencies, the first five categories with higher sums of frequencies are picked out to represent that specific cluster.

Principal Component Analysis and Hierarchical Clustering



A segmentation by PCA is performed by first scaling the data on PCA using three different scales. The proportion of variance of each principal component is calculated after performing PCA on each different scale. The figure above shows how PVE of each scale varies with each principal component, where the green line corresponds to the first scale, the light blue line corresponds to the second scale, and the pink one corresponds to the third scale. Based on the location of elbow in the figure the first ten principal components were chosen,which is a representative number of the data.

Variations in Principal Components

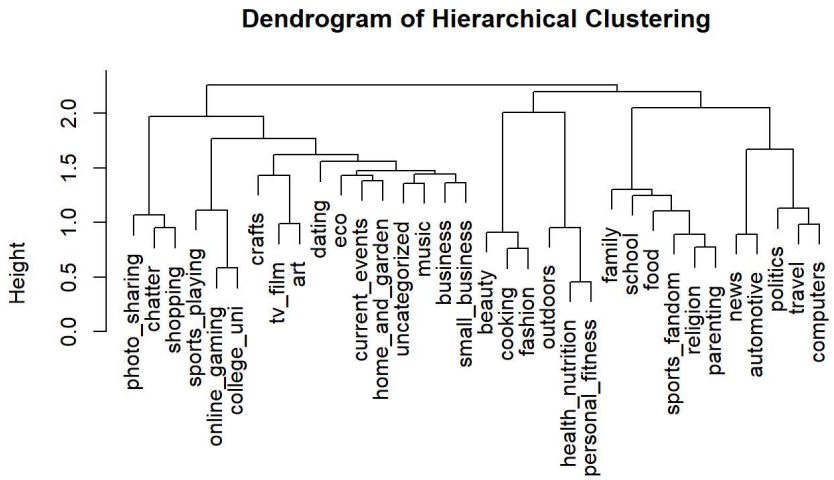
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
Standard deviation	0.1095786	0.0932282	0.0783741	0.0741907	0.0685512	0.054322	0.0526156	0.0491107	0.0443279	0.0401458	0.038444
Proportion of Variance	0.1808800	0.1309300	0.0925300	0.0829200	0.0707900	0.044450	0.0417000	0.0363300	0.0296000	0.0242800	0.022260
Cumulative Proportion	0.1808800	0.3118100	0.4043300	0.4872500	0.5580400	0.602490	0.6441900	0.6805300	0.7101200	0.7344000	0.756670

Each component is used to differentiate the consumers.

Principal Components

	Pri_Com 1	Pri_Com 2	Pri_Com 3	Pri_Com 4	Pri_Com 5	Pri_Com 6	Pri_Com 7	Pri_Com 8	Pri_Com 9	Pri_Com 10
Category 1	tv_film	0.2753544	tv_film	0.4550498	travel	0.5724341	current_events	0.8569087	cooking	0.264473
Category 2	chatter	0.2486782	photo_sharing	0.3517716	religion	0.1911620	online_gaming	0.1337311	sports_fandom	0.196339
Category 3	cooking	0.2393276	current_events	0.2955250	computers	0.1841131	family	0.0421517	tv_film	0.169524
Category 4	current_events	0.2194821	art	0.2835490	food	0.1676774	sports_fandom	0.0342870	shopping	0.154123
Category 5	travel	0.1950609	travel	0.2200802	parenting	0.1148480	sports_playing	0.0330258	college_uni	0.141513

Problem with the above result is that we are not certain who the consumers are and which ones have an interest in which cluster.To solve this problem, the scores of PCA can be used to construct a distance matrix, which measures the distance between each pair of consumers in the data. By Hierarchical Clustering and cutting we are able to obtain 10 clusters of consumers, in accordance with the value of K in K-means. The Hierarchical Clustering with K-means finds out a hierarchical structure of categories based on the proximity matrix of correlations between each pair of categories. Below you find the Dendogram of Hierarchical Clustering.



To conclude, there exists one to one mapping between ten clusters of categories and ten clusters of consumers. Moreover, the distribution of the numbers of consumers over ten clusters is extremely close. By comparing the number of consumers we can link the clusters in any of the three tables. The stable clustering structures across different approaches proves that our ten clusters are very robust based on our data. The largest cluster of the customers of NurientH20 is the one focusing on health and workout. The second largest is the online socializing and shopping.

QUESTION 3: Association rules for grocery purchases

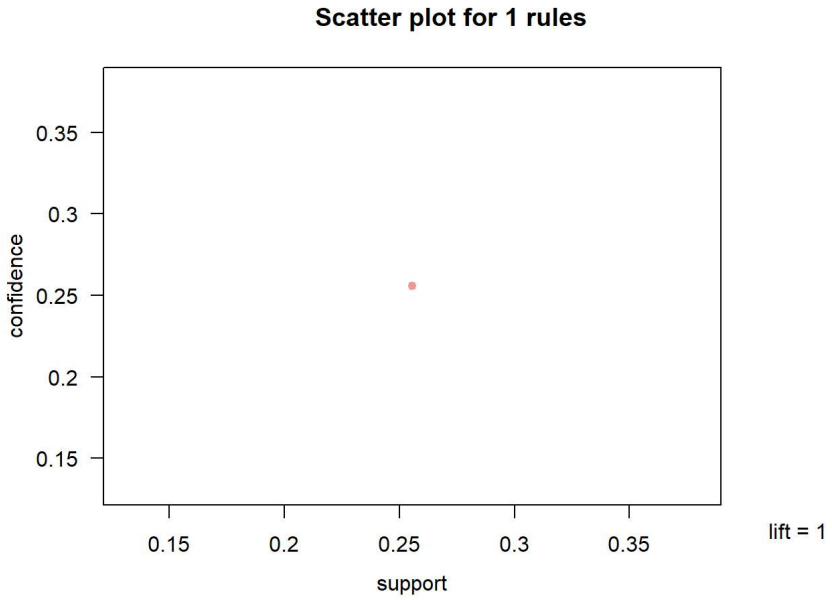
`

Among the associated rules computed, I narrow down the rules into subsets according to confidence, lift and support.

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.2    0.1    1 none FALSE            TRUE       5   0.001      1
##  maxlen target  ext
##       1  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 9
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [157 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 done [0.00s].
## writing ... [1 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
##     lhs    rhs          support  confidence coverage lift count
## [1] {}  => {whole milk} 0.255516 0.255516   1        1    2513
```

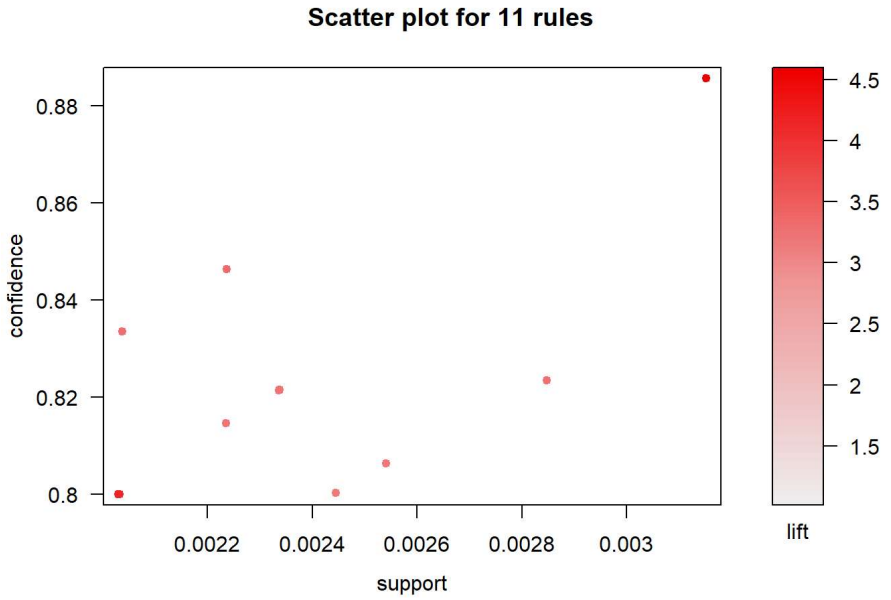
```
##     lhs    rhs          support  confidence coverage lift count
## [1] {}  => {whole milk} 0.255516 0.255516   1        1    2513
```

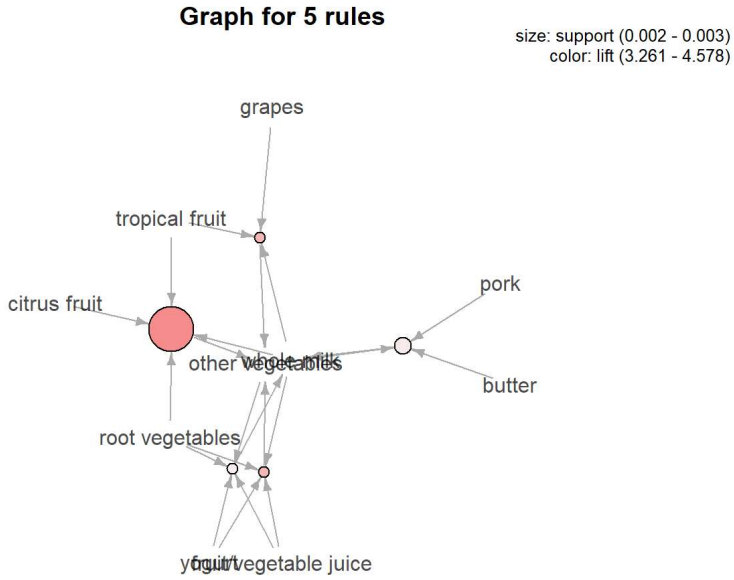



```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.8    0.1    1 none FALSE            TRUE       5   0.002      1
##  maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 19
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [147 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 done [0.01s].
## writing ... [11 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
##      lhs                       rhs                   support confidence    coverage     lift count
## [1]  {herbs,
##       tropical fruit}       => {whole milk}       0.002338587  0.8214286 0.002846975 3.214783    23
## [2]  {herbs,
##       rolls/buns}           => {whole milk}       0.002440264  0.8000000 0.003050330 3.130919    24
## [3]  {curd,
##       hamburger meat}       => {whole milk}       0.002541942  0.8064516 0.003152008 3.156169    25
## [4]  {grapes,
##       tropical fruit,
##       whole milk}           => {other vegetables} 0.002033554  0.8000000 0.002541942 4.134524    20
## [5]  {curd,
##       domestic eggs,
##       other vegetables}     => {whole milk}       0.002846975  0.8235294 0.003457041 3.223005    28
## [6]  {butter,
##       other vegetables,
##       pork}                 => {whole milk}       0.002236909  0.8461538 0.002643620 3.311549    22
## [7]  {fruit/vegetable juice,
##       other vegetables,
##       root vegetables,
##       yogurt}               => {whole milk}       0.002033554  0.8333333 0.002440264 3.261374    20
## [8]  {fruit/vegetable juice,
##       root vegetables,
##       whole milk,
##       yogurt}               => {other vegetables} 0.002033554  0.8000000 0.002541942 4.134524    20
## [9]  {citrus fruit,
##       root vegetables,
##       tropical fruit,
##       whole milk}           => {other vegetables} 0.003152008  0.8857143 0.003558719 4.577509    31
## [10] {citrus fruit,
##       other vegetables,
##       root vegetables,
##       yogurt}               => {whole milk}       0.002338587  0.8214286 0.002846975 3.214783    23
## [11] {rolls/buns,
##       root vegetables,
##       tropical fruit,
##       yogurt}               => {whole milk}       0.002236909  0.8148148 0.002745297 3.188899    22
```

```
## To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.
```





The association can be vizualized through a network graph. The larger the label, the more frequent the transaction. “Whole milk,” and “other vegetables” seem to tie together most transactions.

A strong association between “whole milk” and “other vegetables,” is observed. This is perhaps due to them generally being considered staple goods.

A low level of support chosen was driven by the fact that there weren’t too many different grocery items within the dataset.

In general the rules generated were based on the intuition that, say, that if a customer buys root vegetables he/she is also more likely to buy other vegetables. When it comes to the intuition, lift might be the most informative measure. This is because it measures the conditional probability of purchasing item set A given that you already purchased item set B. Therefore, while lift takes into consideration statistical dependence, confidence and support do not.

The chosen confidence is 80%, support rate is 2% and the minimum lift is 1. This enables for the realization of any association between different grocery baskets.

QUESTION 4: Author attribution

```
TermMatrix_train = DocumentTermMatrix(corpus_train)
TermMatrix_train
```

```
## <<DocumentTermMatrix (documents: 2500, terms: 31423)>>
## Non-/sparse entries: 425955/78131545
## Sparsity           : 99%
## Maximal term length: 36
## Weighting          : term frequency (tf)
```

```
TermMatrix_train2 = removeSparseTerms(TermMatrix_train, 0.95)
TermMatrix_train2
```

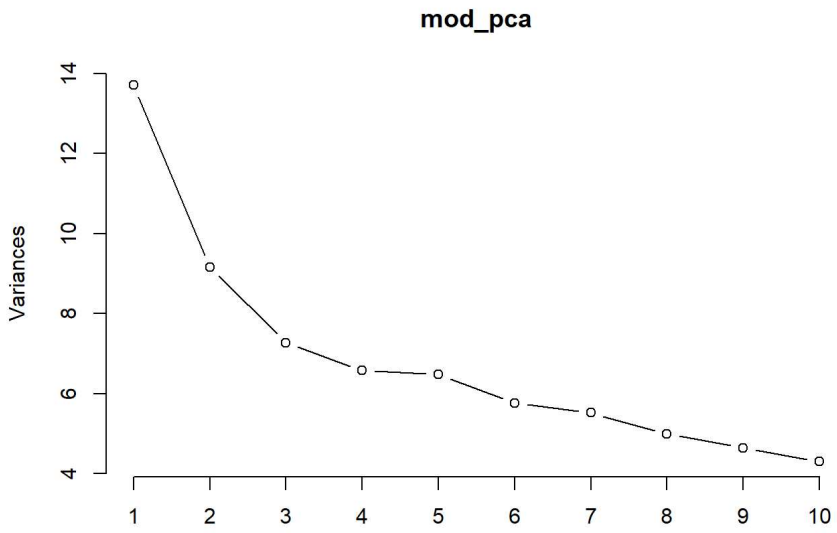
```
## <<DocumentTermMatrix (documents: 2500, terms: 641)>>
## Non-/sparse entries: 180911/1421589
## Sparsity           : 89%
## Maximal term length: 18
## Weighting          : term frequency (tf)
```

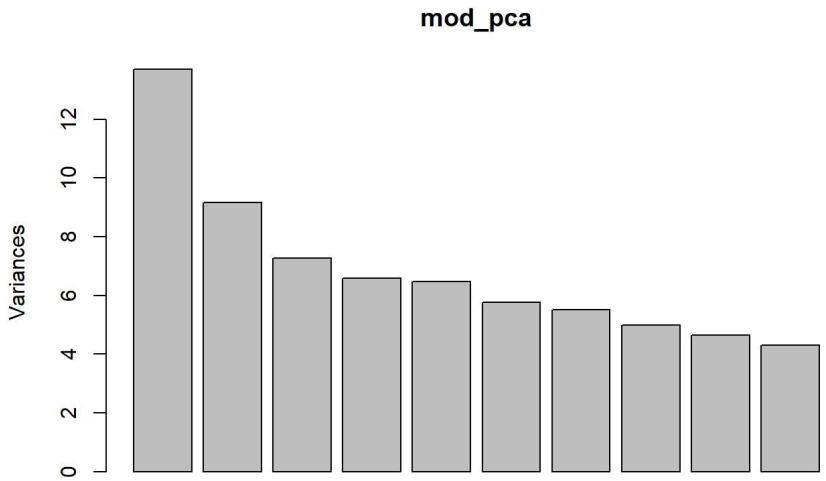
```
DF_train <- data.frame(as.matrix(TermMatrix_train2), stringsAsFactors=FALSE)
labels_train = append(labels_train, rep(author_name, length(files_to_add)))
#Cleaning label names
Author_name = labels_train %>%
    { strsplit(., '/', fixed=TRUE) } %>%
    { lapply(., tail, n=2) } %>%
    { lapply(., paste0, collapse = '') } %>%
    unlist
Author_name = as.data.frame(Author_name)
Author_name = gsub("C([0-9]+)train", "\\1", Author_name$Author_name)
Author_name = gsub("([0-9]+)", "", Author_name)
Author_name = as.data.frame(Author_name)
```

```
##  [1] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/AaronPressman"
##  [2] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/AlanCrosby"
##  [3] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/AlexanderSmith"
##  [4] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/BenjaminKangLim"
##  [5] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/BernardHickey"
##  [6] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/BradDorfman"
##  [7] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/DarrenSchuettler"
##  [8] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/DavidLawder"
##  [9] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/EdnaFernandes"
## [10] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/EricAuchard"
## [11] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/FumikoFujisaki"
## [12] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/GrahamEarnshaw"
## [13] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/HeatherScoffield"
## [14] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/JanLopatka"
## [15] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/JaneMacartney"
## [16] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/JimGilchrist"
## [17] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/JoWinterbottom"
## [18] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/JoeOrtiz"
## [19] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/JohnMastrini"
## [20] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/JonathanBirt"
## [21] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/KarlPenhaul"
## [22] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/KeithWeir"
## [23] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/KevinDrawbaugh"
## [24] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/KevinMorrison"
## [25] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/KirstinRidley"
## [26] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/KouroshKarimkhany"
## [27] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/LydiaZajc"
## [28] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/LynneO'Donnell"
## [29] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/LynnleyBrowning"
## [30] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/MarcelMichelson"
## [31] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/MarkBendeich"
## [32] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/MartinWolk"
## [33] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/MatthewBunce"
## [34] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/MichaelConnor"
## [35] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/MureDickie"
## [36] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/NickLouth"
## [37] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/PatriciaCommins"
## [38] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/PeterHumphrey"
## [39] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/PierreTran"
## [40] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/RobinSidel"
## [41] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/RogerFillion"
## [42] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/SamuelPerry"
## [43] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/SarahDavison"
## [44] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/ScottHillis"
## [45] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/SimonCowell"
## [46] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/TanEeLyn"
## [47] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/TheresePoletti"
## [48] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/TimFarrand"
## [49] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/ToddNissen"
## [50] "C:/Users/Fjolla/Documents/GitHub/ECO395M/data/ReutersC50/C50test/WilliamKazer"
```

```
## <<DocumentTermMatrix (documents: 2500, terms: 658)>>
## Non-/sparse entries: 185022/1459978
## Sparsity           : 89%
## Maximal term length: 18
## Weighting          : term frequency (tf)
```

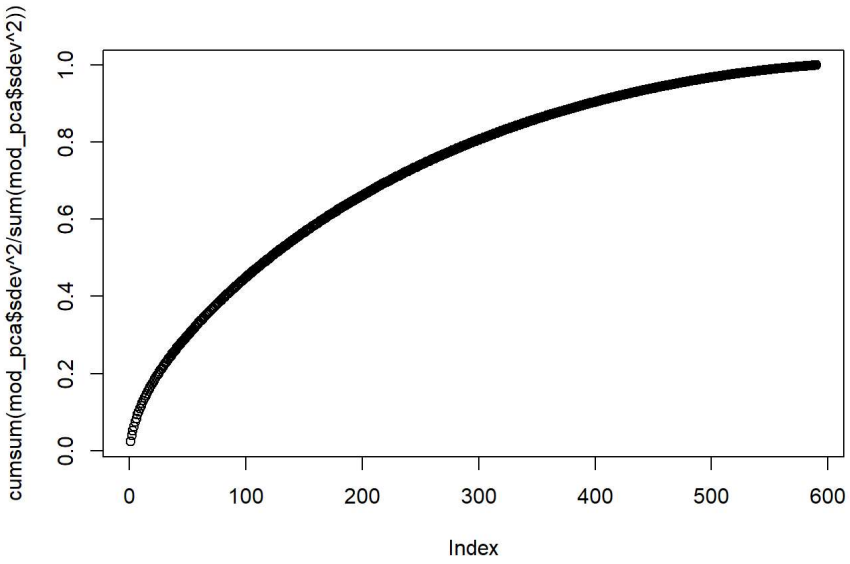
The PCA dimension reduction technique is used in order to reduce the data from a large number of features into smaller principal properties.


mod_pca
Variances
4
6
8
10
12
14
1
2
3
4
5
6
7
8
9
10


mod_pca
Variances
0
2
4
6
8
10
12

```
##        PC1        PC2        PC3        PC4        PC5        PC6        PC7
## 0.02323216 0.03876532 0.05109178 0.06223812 0.07321755 0.08298916 0.09234581
##        PC8        PC9       PC10       PC11       PC12       PC13       PC14
## 0.10082010 0.10869947 0.11599385 0.12283617 0.12946616 0.13574254 0.14174797
##       PC15       PC16       PC17       PC18       PC19       PC20       PC21
## 0.14759468 0.15334416 0.15899132 0.16444095 0.16981195 0.17504926 0.18021202
##       PC22       PC23       PC24       PC25       PC26       PC27       PC28
## 0.18521768 0.19009932 0.19492612 0.19963603 0.20428940 0.20882769 0.21330997
##       PC29       PC30       PC31       PC32       PC33       PC34       PC35
## 0.21773713 0.22211227 0.22644510 0.23065951 0.23481376 0.23888660 0.24290709
##       PC36       PC37       PC38       PC39       PC40       PC41       PC42
## 0.24691867 0.25088639 0.25484044 0.25871089 0.26256701 0.26635790 0.27012181
##       PC43       PC44       PC45       PC46       PC47       PC48       PC49
## 0.27384187 0.27753097 0.28120241 0.28482957 0.28840826 0.29196861 0.29548073
##       PC50       PC51       PC52       PC53       PC54       PC55       PC56
## 0.29896878 0.30244238 0.30588020 0.30927950 0.31267292 0.31604098 0.31940235
##       PC57       PC58       PC59       PC60       PC61       PC62       PC63
## 0.32272548 0.32602571 0.32931909 0.33256107 0.33578605 0.33899784 0.34219355
##       PC64       PC65       PC66       PC67       PC68       PC69       PC70
## 0.34538381 0.34853110 0.35166663 0.35479215 0.35790086 0.36098947 0.36407385
##       PC71       PC72       PC73       PC74       PC75       PC76       PC77
## 0.36714336 0.37019688 0.37322873 0.37625770 0.37928399 0.38228639 0.38523636
##       PC78       PC79       PC80       PC81       PC82       PC83       PC84
## 0.38817948 0.39110406 0.39400323 0.39688870 0.39976307 0.40261374 0.40544265
##       PC85       PC86       PC87       PC88       PC89       PC90       PC91
## 0.40825610 0.41105562 0.41385246 0.41663635 0.41940590 0.42216530 0.42490203
##       PC92       PC93       PC94       PC95       PC96       PC97       PC98
## 0.42763380 0.43035471 0.43305483 0.43574379 0.43843124 0.44110332 0.44376349
##       PC99      PC100      PC101      PC102      PC103      PC104      PC105
## 0.44639932 0.44902272 0.45162533 0.45422450 0.45681778 0.45940350 0.46195949
##      PC106      PC107      PC108      PC109      PC110      PC111      PC112
## 0.46451175 0.46705643 0.46958834 0.47210214 0.47460083 0.47708764 0.47957044
##      PC113      PC114      PC115      PC116      PC117      PC118      PC119
## 0.48203538 0.48449454 0.48694754 0.48938965 0.49182512 0.49424258 0.49665321
##      PC120      PC121      PC122      PC123      PC124      PC125      PC126
## 0.49905346 0.50144088 0.50380721 0.50616778 0.50852345 0.51087657 0.51321377
##      PC127      PC128      PC129      PC130      PC131      PC132      PC133
## 0.51553920 0.51785861 0.52016877 0.52246784 0.52475263 0.52703364 0.52930507
##      PC134      PC135      PC136      PC137      PC138      PC139      PC140
## 0.53156191 0.53381601 0.53605919 0.53828605 0.54050567 0.54271865 0.54492304
##      PC141      PC142      PC143      PC144      PC145      PC146      PC147
## 0.54711789 0.54930628 0.55148212 0.55365448 0.55581602 0.55797040 0.56011258
##      PC148      PC149      PC150      PC151      PC152      PC153      PC154
## 0.56224194 0.56436862 0.56648484 0.56859224 0.57069138 0.57278357 0.57486613
##      PC155      PC156      PC157      PC158      PC159      PC160      PC161
## 0.57694107 0.57900893 0.58106953 0.58312250 0.58516259 0.58718472 0.58920417
##      PC162      PC163      PC164      PC165      PC166      PC167      PC168
## 0.59121722 0.59322194 0.59522015 0.59721448 0.59919815 0.60118092 0.60315243
##      PC169      PC170      PC171      PC172      PC173      PC174      PC175
## 0.60510975 0.60706443 0.60901286 0.61094979 0.61288052 0.61480586 0.61672447
##      PC176      PC177      PC178      PC179      PC180      PC181      PC182
## 0.61862954 0.62052610 0.62241847 0.62429760 0.62617218 0.62804195 0.62990495
##      PC183      PC184      PC185      PC186      PC187      PC188      PC189
## 0.63176627 0.63361351 0.63545457 0.63729201 0.63911154 0.64092322 0.64272885
##      PC190      PC191      PC192      PC193      PC194      PC195      PC196
## 0.64452344 0.64631736 0.64810999 0.64988591 0.65165416 0.65341657 0.65517641
##      PC197      PC198      PC199      PC200      PC201      PC202      PC203
## 0.65693160 0.65868457 0.66043094 0.66216546 0.66389723 0.66562397 0.66734177
##      PC204      PC205      PC206      PC207      PC208      PC209      PC210
## 0.66905580 0.67076234 0.67246235 0.67415672 0.67584660 0.67752221 0.67919374
##      PC211      PC212      PC213      PC214      PC215      PC216      PC217
## 0.68086409 0.68252720 0.68418481 0.68583599 0.68747795 0.68911951 0.69074701
##      PC218      PC219      PC220      PC221      PC222      PC223      PC224
## 0.69236973 0.69398528 0.69559372 0.69719893 0.69880186 0.70039543 0.70198468
##      PC225      PC226      PC227      PC228      PC229      PC230      PC231
## 0.70357183 0.70515058 0.70671945 0.70828283 0.70983714 0.71138603 0.71292750
##      PC232      PC233      PC234      PC235      PC236      PC237      PC238
## 0.71446642 0.71600307 0.71753189 0.71905154 0.72056387 0.72207499 0.72358140
##      PC239      PC240      PC241      PC242      PC243      PC244      PC245
## 0.72508058 0.72657792 0.72806851 0.72955518 0.73103699 0.73251554 0.73398866
##      PC246      PC247      PC248      PC249      PC250      PC251      PC252
## 0.73545824 0.73691966 0.73837785 0.73981971 0.74125534 0.74268799 0.74411355
##      PC253      PC254      PC255      PC256      PC257      PC258      PC259
## 0.74553711 0.74695483 0.74836116 0.74976420 0.75116426 0.75255655 0.75394599
##      PC260      PC261      PC262      PC263      PC264      PC265      PC266
## 0.75533097 0.75671299 0.75808790 0.75945810 0.76082299 0.76218339 0.76354138
##      PC267      PC268      PC269      PC270      PC271      PC272      PC273
## 0.76488912 0.76623192 0.76757297 0.76890419 0.77023182 0.77155407 0.77287255
##      PC274      PC275      PC276      PC277      PC278      PC279      PC280
## 0.77418544 0.77549567 0.77679855 0.77809957 0.77939388 0.78068103 0.78196587
##      PC281      PC282      PC283      PC284      PC285      PC286      PC287
## 0.78324959 0.78452465 0.78579509 0.78705922 0.78832127 0.78957823 0.79083177
##      PC288      PC289      PC290      PC291      PC292      PC293      PC294
## 0.79207408 0.79331540 0.79455404 0.79578711 0.79701489 0.79824081 0.79946311
##      PC295      PC296      PC297      PC298      PC299      PC300      PC301
## 0.80067777 0.80188838 0.80309781 0.80430334 0.80550190 0.80669309 0.80788362
```

```
##      PC302      PC303      PC304      PC305      PC306      PC307      PC308
## 0.80906929 0.81025106 0.81142548 0.81259730 0.81376369 0.81492262 0.81607815
##      PC309      PC310      PC311      PC312      PC313      PC314      PC315
## 0.81723149 0.81837995 0.81952692 0.82067219 0.82181256 0.82294754 0.82407914
##      PC316      PC317      PC318      PC319      PC320      PC321      PC322
## 0.82520152 0.82631826 0.82743116 0.82853879 0.82964328 0.83074437 0.83184204
##      PC323      PC324      PC325      PC326      PC327      PC328      PC329
## 0.83293335 0.83401899 0.83510272 0.83618403 0.83725962 0.83833037 0.83939996
##      PC330      PC331      PC332      PC333      PC334      PC335      PC336
## 0.84046311 0.84152158 0.84257350 0.84361938 0.84466101 0.84569914 0.84673280
##      PC337      PC338      PC339      PC340      PC341      PC342      PC343
## 0.84776519 0.84879541 0.84982282 0.85084552 0.85186030 0.85287304 0.85388168
##      PC344      PC345      PC346      PC347      PC348      PC349      PC350
## 0.85488494 0.85588413 0.85687597 0.85786408 0.85884891 0.85983186 0.86081088
##      PC351      PC352      PC353      PC354      PC355      PC356      PC357
## 0.86178509 0.86275566 0.86372328 0.86468741 0.86565078 0.86660846 0.86756167
##      PC358      PC359      PC360      PC361      PC362      PC363      PC364
## 0.86851150 0.86946013 0.87040441 0.87134149 0.87227395 0.87320392 0.87412983
##      PC365      PC366      PC367      PC368      PC369      PC370      PC371
## 0.87505449 0.87597426 0.87689061 0.87780213 0.87870723 0.87961077 0.88050723
##      PC372      PC373      PC374      PC375      PC376      PC377      PC378
## 0.88139652 0.88228552 0.88316992 0.88405305 0.88493172 0.88580508 0.88667482
##      PC379      PC380      PC381      PC382      PC383      PC384      PC385
## 0.88754246 0.88840560 0.88926692 0.89012544 0.89098034 0.89183200 0.89268044
##      PC386      PC387      PC388      PC389      PC390      PC391      PC392
## 0.89352444 0.89436540 0.89520356 0.89603829 0.89686883 0.89769602 0.89851941
##      PC393      PC394      PC395      PC396      PC397      PC398      PC399
## 0.89933860 0.90015550 0.90096955 0.90178103 0.90258967 0.90339527 0.90419739
##      PC400      PC401      PC402      PC403      PC404      PC405      PC406
## 0.90499821 0.90579377 0.90658258 0.90736868 0.90815297 0.90893512 0.90971563
##      PC407      PC408      PC409      PC410      PC411      PC412      PC413
## 0.91048959 0.91126166 0.91203242 0.91280025 0.91356319 0.91432266 0.91508043
##      PC414      PC415      PC416      PC417      PC418      PC419      PC420
## 0.91583658 0.91658737 0.91733410 0.91807911 0.91881872 0.91955066 0.92028108
##      PC421      PC422      PC423      PC424      PC425      PC426      PC427
## 0.92100793 0.92173098 0.92245252 0.92317044 0.92388187 0.92459199 0.92529746
##      PC428      PC429      PC430      PC431      PC432      PC433      PC434
## 0.92600254 0.92670472 0.92739945 0.92809166 0.92878251 0.92947003 0.93015227
##      PC435      PC436      PC437      PC438      PC439      PC440      PC441
## 0.93083378 0.93151370 0.93219069 0.93286081 0.93352800 0.93419345 0.93485644
##      PC442      PC443      PC444      PC445      PC446      PC447      PC448
## 0.93551748 0.93617258 0.93682557 0.93747786 0.93812446 0.93876907 0.93941089
##      PC449      PC450      PC451      PC452      PC453      PC454      PC455
## 0.94004692 0.94068140 0.94131123 0.94193856 0.94256436 0.94318765 0.94380766
##      PC456      PC457      PC458      PC459      PC460      PC461      PC462
## 0.94442406 0.94503950 0.94565162 0.94626156 0.94686670 0.94746660 0.94806155
##      PC463      PC464      PC465      PC466      PC467      PC468      PC469
## 0.94865507 0.94924715 0.94983617 0.95042221 0.95100341 0.95157964 0.95215388
##      PC470      PC471      PC472      PC473      PC474      PC475      PC476
## 0.95272731 0.95329823 0.95386867 0.95443494 0.95499654 0.95555586 0.95611425
##      PC477      PC478      PC479      PC480      PC481      PC482      PC483
## 0.95667066 0.95722434 0.95777244 0.95831895 0.95886091 0.95939914 0.95993665
##      PC484      PC485      PC486      PC487      PC488      PC489      PC490
## 0.96046959 0.96099904 0.96152657 0.96205085 0.96257103 0.96309056 0.96360809
##      PC491      PC492      PC493      PC494      PC495      PC496      PC497
## 0.96412322 0.96463722 0.96514778 0.96565695 0.96616249 0.96666594 0.96716279
##      PC498      PC499      PC500      PC501      PC502      PC503      PC504
## 0.96765755 0.96814802 0.96863664 0.96912025 0.96960116 0.97007870 0.97055217
##      PC505      PC506      PC507      PC508      PC509      PC510      PC511
## 0.97102192 0.97148926 0.97195481 0.97241545 0.97287384 0.97332656 0.97377797
##      PC512      PC513      PC514      PC515      PC516      PC517      PC518
## 0.97422736 0.97467390 0.97511955 0.97556181 0.97600025 0.97643661 0.97687055
##      PC519      PC520      PC521      PC522      PC523      PC524      PC525
## 0.97730242 0.97772895 0.97815191 0.97857349 0.97899206 0.97940852 0.97982232
##      PC526      PC527      PC528      PC529      PC530      PC531      PC532
## 0.98023268 0.98063894 0.98104172 0.98144250 0.98184255 0.98223979 0.98263398
##      PC533      PC534      PC535      PC536      PC537      PC538      PC539
## 0.98302467 0.98341328 0.98379718 0.98417655 0.98455512 0.98493150 0.98530580
##      PC540      PC541      PC542      PC543      PC544      PC545      PC546
## 0.98567638 0.98604298 0.98640894 0.98677080 0.98713043 0.98748771 0.98784324
##      PC547      PC548      PC549      PC550      PC551      PC552      PC553
## 0.98819693 0.98854851 0.98889538 0.98924024 0.98958347 0.98991880 0.99025083
##      PC554      PC555      PC556      PC557      PC558      PC559      PC560
## 0.99057945 0.99090548 0.99122920 0.99155031 0.99186900 0.99218266 0.99249100
##      PC561      PC562      PC563      PC564      PC565      PC566      PC567
## 0.99279653 0.99310057 0.99340103 0.99370063 0.99399595 0.99428968 0.99457791
##      PC568      PC569      PC570      PC571      PC572      PC573      PC574
## 0.99486581 0.99515268 0.99543727 0.99571416 0.99598800 0.99625972 0.99652890
##      PC575      PC576      PC577      PC578      PC579      PC580      PC581
## 0.99679467 0.99705865 0.99731887 0.99757496 0.99782732 0.99807261 0.99831500
##      PC582      PC583      PC584      PC585      PC586      PC587      PC588
## 0.99855251 0.99878741 0.99901939 0.99924274 0.99945971 0.99967206 0.99986583
##      PC589      PC590
## 0.99999690 1.00000000
```


cumsum(mod_pca$sdev^2/sum(mod_pca$sdev^2))
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500
600
Index

```
## Importance of components:
##                           PC1     PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation     3.70229 3.02730 2.69678 2.56444 2.54516 2.40109 2.34956
## Proportion of Variance 0.02323 0.01553 0.01233 0.01115 0.01098 0.00977 0.00936
## Cumulative Proportion  0.02323 0.03877 0.05109 0.06224 0.07322 0.08299 0.09235
##                           PC8     PC9    PC10    PC11    PC12    PC13    PC14
## Standard deviation     2.23603 2.15611 2.07453 2.00922 1.97780 1.92433 1.88234
## Proportion of Variance 0.00847 0.00788 0.00729 0.00684 0.00663 0.00628 0.00601
## Cumulative Proportion  0.10082 0.10870 0.11599 0.12284 0.12947 0.13574 0.14175
##                           PC15    PC16    PC17    PC18    PC19    PC20    PC21
## Standard deviation     1.85730 1.84179 1.82533 1.79312 1.78014 1.75784 1.74529
## Proportion of Variance 0.00585 0.00575 0.00565 0.00545 0.00537 0.00524 0.00516
## Cumulative Proportion  0.14759 0.15334 0.15899 0.16444 0.16981 0.17505 0.18021
##                           PC22    PC23    PC24    PC25    PC26    PC27    PC28
## Standard deviation     1.71853 1.69711 1.68755 1.66699 1.65695 1.63634 1.62621
## Proportion of Variance 0.00501 0.00488 0.00483 0.00471 0.00465 0.00454 0.00448
## Cumulative Proportion  0.18522 0.19010 0.19493 0.19964 0.20429 0.20883 0.21331
##                           PC29    PC30    PC31    PC32    PC33    PC34    PC35
## Standard deviation     1.61618 1.60665 1.59887 1.57687 1.56557 1.55015 1.54016
## Proportion of Variance 0.00443 0.00438 0.00433 0.00421 0.00415 0.00407 0.00402
## Cumulative Proportion  0.21774 0.22211 0.22645 0.23066 0.23481 0.23889 0.24291
##                           PC36    PC37    PC38    PC39    PC40    PC41    PC42
## Standard deviation     1.53845 1.53002 1.52738 1.51115 1.50835 1.49554 1.49020
## Proportion of Variance 0.00401 0.00397 0.00395 0.00387 0.00386 0.00379 0.00376
## Cumulative Proportion  0.24692 0.25089 0.25484 0.25871 0.26257 0.26636 0.27012
##                           PC43    PC44    PC45    PC46    PC47    PC48    PC49
## Standard deviation     1.48150 1.47532 1.47179 1.46288 1.45308 1.44935 1.43950
## Proportion of Variance 0.00372 0.00369 0.00367 0.00363 0.00358 0.00356 0.00351
## Cumulative Proportion  0.27384 0.27753 0.28120 0.28483 0.28841 0.29197 0.29548
##                           PC50    PC51    PC52    PC53    PC54    PC55    PC56
## Standard deviation     1.43456 1.43158 1.42419 1.4162 1.41496 1.40967 1.40826
## Proportion of Variance 0.00349 0.00347 0.00344 0.0034 0.00339 0.00337 0.00336
## Cumulative Proportion  0.29897 0.30244 0.30588 0.3093 0.31267 0.31604 0.31940
##                           PC57    PC58    PC59    PC60    PC61    PC62    PC63
## Standard deviation     1.40023 1.3954 1.39395 1.38303 1.37940 1.37658 1.3731
## Proportion of Variance 0.00332 0.0033 0.00329 0.00324 0.00322 0.00321 0.0032
## Cumulative Proportion  0.32273 0.3260 0.32932 0.33256 0.33579 0.33900 0.3422
##                           PC64    PC65    PC66    PC67    PC68    PC69    PC70
## Standard deviation     1.37195 1.36268 1.36013 1.35796 1.35431 1.34992 1.34900
## Proportion of Variance 0.00319 0.00315 0.00314 0.00313 0.00311 0.00309 0.00308
## Cumulative Proportion  0.34538 0.34853 0.35167 0.35479 0.35790 0.36099 0.36407
##                           PC71    PC72    PC73    PC74    PC75    PC76    PC77
## Standard deviation     1.34574 1.34223 1.33746 1.33682 1.33623 1.3309 1.31927
## Proportion of Variance 0.00307 0.00305 0.00303 0.00303 0.00303 0.0030 0.00295
## Cumulative Proportion  0.36714 0.37020 0.37323 0.37626 0.37928 0.3823 0.38524
##                           PC78    PC79    PC80    PC81    PC82    PC83    PC84
## Standard deviation     1.31774 1.31358 1.3079 1.30477 1.30226 1.29688 1.29192
## Proportion of Variance 0.00294 0.00292 0.0029 0.00289 0.00287 0.00285 0.00283
## Cumulative Proportion  0.38818 0.39110 0.3940 0.39689 0.39976 0.40261 0.40544
##                           PC85    PC86    PC87    PC88    PC89    PC90    PC91
## Standard deviation     1.28839 1.2852 1.2846 1.28160 1.27829 1.27595 1.27070
## Proportion of Variance 0.00281 0.0028 0.0028 0.00278 0.00277 0.00276 0.00274
## Cumulative Proportion  0.40826 0.4111 0.4138 0.41664 0.41941 0.42217 0.42490
##                           PC92    PC93    PC94    PC95    PC96    PC97    PC98
## Standard deviation     1.26954 1.26702 1.2622 1.25956 1.25920 1.25560 1.25280
## Proportion of Variance 0.00273 0.00272 0.0027 0.00269 0.00269 0.00267 0.00266
## Cumulative Proportion  0.42763 0.43035 0.4330 0.43574 0.43843 0.44110 0.44376
##                           PC99   PC100   PC101   PC102   PC103   PC104   PC105
## Standard deviation     1.24705 1.24411 1.2392 1.2384 1.23695 1.23514 1.22802
## Proportion of Variance 0.00264 0.00262 0.0026 0.0026 0.00259 0.00259 0.00256
## Cumulative Proportion  0.44640 0.44902 0.4516 0.4542 0.45682 0.45940 0.46196
##                          PC106   PC107   PC108   PC109   PC110   PC111   PC112
## Standard deviation     1.22713 1.22530 1.22222 1.21784 1.2142 1.21129 1.21031
## Proportion of Variance 0.00255 0.00254 0.00253 0.00251 0.0025 0.00249 0.00248
## Cumulative Proportion  0.46451 0.46706 0.46959 0.47210 0.4746 0.47709 0.47957
##                          PC113   PC114   PC115   PC116   PC117   PC118   PC119
## Standard deviation     1.20595 1.20453 1.20303 1.20035 1.19872 1.19428 1.19259
## Proportion of Variance 0.00246 0.00246 0.00245 0.00244 0.00244 0.00242 0.00241
## Cumulative Proportion  0.48204 0.48449 0.48695 0.48939 0.49183 0.49424 0.49665
##                          PC120   PC121   PC122   PC123   PC124   PC125   PC126
## Standard deviation     1.1900 1.18683 1.18158 1.18014 1.17892 1.17828 1.17429
## Proportion of Variance 0.0024 0.00239 0.00237 0.00236 0.00236 0.00235 0.00234
## Cumulative Proportion  0.4990 0.50144 0.50381 0.50617 0.50852 0.51088 0.51321
##                          PC127   PC128   PC129   PC130   PC131   PC132   PC133
## Standard deviation     1.17132 1.16981 1.16747 1.1647 1.16105 1.16008 1.15765
## Proportion of Variance 0.00233 0.00232 0.00231 0.0023 0.00228 0.00228 0.00227
## Cumulative Proportion  0.51554 0.51786 0.52017 0.5225 0.52475 0.52703 0.52931
##                          PC134   PC135   PC136   PC137   PC138   PC139   PC140
## Standard deviation     1.15392 1.15322 1.15042 1.14623 1.14437 1.14266 1.1404
## Proportion of Variance 0.00226 0.00225 0.00224 0.00223 0.00222 0.00221 0.0022
## Cumulative Proportion  0.53156 0.53382 0.53606 0.53829 0.54051 0.54272 0.5449
##                          PC141   PC142   PC143   PC144   PC145   PC146   PC147
## Standard deviation     1.13796 1.13629 1.13302 1.13212 1.12930 1.12742 1.12423
## Proportion of Variance 0.00219 0.00219 0.00218 0.00217 0.00216 0.00215 0.00214
## Cumulative Proportion  0.54712 0.54931 0.55148 0.55365 0.55582 0.55797 0.56011
##                          PC148   PC149   PC150   PC151   PC152   PC153   PC154
```

```
## Standard deviation     1.12086 1.12015 1.11739 1.11506 1.1129 1.11103 1.10847
## Proportion of Variance 0.00213 0.00213 0.00212 0.00211 0.0021 0.00209 0.00208
## Cumulative Proportion  0.56224 0.56437 0.56648 0.56859 0.5707 0.57278 0.57487
##                           PC155   PC156   PC157   PC158   PC159   PC160   PC161
## Standard deviation     1.10644 1.10455 1.10261 1.10057 1.09711 1.09227 1.09155
## Proportion of Variance 0.00207 0.00207 0.00206 0.00205 0.00204 0.00202 0.00202
## Cumulative Proportion  0.57694 0.57901 0.58107 0.58312 0.58516 0.58718 0.58920
##                           PC162  PC163  PC164   PC165   PC166   PC167   PC168
## Standard deviation     1.08982 1.0876 1.0858 1.08474 1.08183 1.08159 1.07852
## Proportion of Variance 0.00201 0.0020 0.0020 0.00199 0.00198 0.00198 0.00197
## Cumulative Proportion  0.59122 0.5932 0.5952 0.59721 0.59920 0.60118 0.60315
##                           PC169   PC170   PC171   PC172   PC173   PC174   PC175
## Standard deviation     1.07463 1.07390 1.07218 1.06901 1.06730 1.06581 1.06395
## Proportion of Variance 0.00196 0.00195 0.00195 0.00194 0.00193 0.00193 0.00192
## Cumulative Proportion  0.60511 0.60706 0.60901 0.61095 0.61288 0.61481 0.61672
##                           PC176  PC177   PC178   PC179   PC180   PC181   PC182
## Standard deviation     1.06018 1.0578 1.05664 1.05294 1.05167 1.05032 1.04841
## Proportion of Variance 0.00191 0.0019 0.00189 0.00188 0.00187 0.00187 0.00186
## Cumulative Proportion  0.61863 0.6205 0.62242 0.62430 0.62617 0.62804 0.62990
##                           PC183   PC184   PC185   PC186   PC187   PC188   PC189
## Standard deviation     1.04794 1.04397 1.04222 1.04119 1.03611 1.03387 1.03215
## Proportion of Variance 0.00186 0.00185 0.00184 0.00184 0.00182 0.00181 0.00181
## Cumulative Proportion  0.63177 0.63361 0.63545 0.63729 0.63911 0.64092 0.64273
##                           PC190   PC191   PC192   PC193   PC194   PC195   PC196
## Standard deviation     1.02898 1.02879 1.02842 1.02362 1.02140 1.01972 1.01897
## Proportion of Variance 0.00179 0.00179 0.00179 0.00178 0.00177 0.00176 0.00176
## Cumulative Proportion  0.64452 0.64632 0.64811 0.64989 0.65165 0.65342 0.65518
##                           PC197   PC198   PC199   PC200   PC201   PC202   PC203
## Standard deviation     1.01762 1.01698 1.01507 1.01162 1.01081 1.00934 1.00673
## Proportion of Variance 0.00176 0.00175 0.00175 0.00173 0.00173 0.00173 0.00172
## Cumulative Proportion  0.65693 0.65868 0.66043 0.66217 0.66390 0.66562 0.66734
##                           PC204   PC205  PC206   PC207   PC208   PC209   PC210
## Standard deviation     1.00562 1.00342 1.0015 0.99984 0.99852 0.99429 0.99308
## Proportion of Variance 0.00171 0.00171 0.0017 0.00169 0.00169 0.00168 0.00167
## Cumulative Proportion  0.66906 0.67076 0.6725 0.67416 0.67585 0.67752 0.67919
##                           PC211   PC212   PC213   PC214   PC215   PC216   PC217
## Standard deviation     0.99273 0.99057 0.98893 0.98701 0.98425 0.98413 0.97991
## Proportion of Variance 0.00167 0.00166 0.00166 0.00165 0.00164 0.00164 0.00163
## Cumulative Proportion  0.68086 0.68253 0.68418 0.68584 0.68748 0.68912 0.69075
##                           PC218   PC219   PC220   PC221  PC222   PC223   PC224
## Standard deviation     0.97847 0.97631 0.97416 0.97318 0.9725 0.96964 0.96833
## Proportion of Variance 0.00162 0.00162 0.00161 0.00161 0.0016 0.00159 0.00159
## Cumulative Proportion  0.69237 0.69399 0.69559 0.69720 0.6988 0.70040 0.70198
##                           PC225   PC226   PC227   PC228   PC229   PC230   PC231
## Standard deviation     0.96769 0.96512 0.96210 0.96041 0.95762 0.95595 0.95366
## Proportion of Variance 0.00159 0.00158 0.00157 0.00156 0.00155 0.00155 0.00154
## Cumulative Proportion  0.70357 0.70515 0.70672 0.70828 0.70984 0.71139 0.71293
##                           PC232   PC233   PC234   PC235   PC236   PC237   PC238
## Standard deviation     0.95287 0.95217 0.94974 0.94689 0.94460 0.94422 0.94275
## Proportion of Variance 0.00154 0.00154 0.00153 0.00152 0.00151 0.00151 0.00151
## Cumulative Proportion  0.71447 0.71600 0.71753 0.71905 0.72056 0.72207 0.72358
##                           PC239  PC240   PC241   PC242   PC243   PC244   PC245
## Standard deviation     0.9405 0.9399 0.93779 0.93656 0.93502 0.93399 0.93228
## Proportion of Variance 0.0015 0.0015 0.00149 0.00149 0.00148 0.00148 0.00147
## Cumulative Proportion  0.7251 0.7266 0.72807 0.72956 0.73104 0.73252 0.73399
##                           PC246   PC247   PC248   PC249   PC250   PC251   PC252
## Standard deviation     0.93115 0.92857 0.92754 0.92233 0.92034 0.91938 0.91711
## Proportion of Variance 0.00147 0.00146 0.00146 0.00144 0.00144 0.00143 0.00143
## Cumulative Proportion  0.73546 0.73692 0.73838 0.73982 0.74126 0.74269 0.74411
##                           PC253   PC254   PC255  PC256  PC257   PC258   PC259
## Standard deviation     0.91646 0.91458 0.91090 0.9098 0.9089 0.90634 0.90541
## Proportion of Variance 0.00142 0.00142 0.00141 0.0014 0.0014 0.00139 0.00139
## Cumulative Proportion  0.74554 0.74695 0.74836 0.7498 0.7512 0.75256 0.75395
##                           PC260   PC261   PC262   PC263   PC264   PC265   PC266
## Standard deviation     0.90396 0.90299 0.90066 0.89912 0.89738 0.89590 0.89511
## Proportion of Variance 0.00138 0.00138 0.00137 0.00137 0.00136 0.00136 0.00136
## Cumulative Proportion  0.75533 0.75671 0.75809 0.75946 0.76082 0.76218 0.76354
##                           PC267   PC268   PC269   PC270   PC271   PC272   PC273
## Standard deviation     0.89172 0.89009 0.88950 0.88624 0.88505 0.88325 0.88199
## Proportion of Variance 0.00135 0.00134 0.00134 0.00133 0.00133 0.00132 0.00132
## Cumulative Proportion  0.76489 0.76623 0.76757 0.76890 0.77023 0.77155 0.77287
##                           PC274   PC275  PC276  PC277   PC278   PC279   PC280
## Standard deviation     0.88012 0.87922 0.8768 0.8761 0.87387 0.87145 0.87066
## Proportion of Variance 0.00131 0.00131 0.0013 0.0013 0.00129 0.00129 0.00128
## Cumulative Proportion  0.77419 0.77550 0.7768 0.7781 0.77939 0.78068 0.78197
##                           PC281   PC282   PC283   PC284   PC285   PC286   PC287
## Standard deviation     0.87028 0.86734 0.86577 0.86362 0.86291 0.86117 0.85999
## Proportion of Variance 0.00128 0.00128 0.00127 0.00126 0.00126 0.00126 0.00125
## Cumulative Proportion  0.78325 0.78452 0.78580 0.78706 0.78832 0.78958 0.79083
##                           PC288   PC289   PC290   PC291   PC292   PC293   PC294
## Standard deviation     0.85613 0.85579 0.85487 0.85294 0.85111 0.85047 0.84921
## Proportion of Variance 0.00124 0.00124 0.00124 0.00123 0.00123 0.00123 0.00122
## Cumulative Proportion  0.79207 0.79332 0.79455 0.79579 0.79701 0.79824 0.79946
##                           PC295   PC296   PC297   PC298  PC299   PC300   PC301
## Standard deviation     0.84655 0.84514 0.84473 0.84336 0.8409 0.83833 0.83810
## Proportion of Variance 0.00121 0.00121 0.00121 0.00121 0.0012 0.00119 0.00119
## Cumulative Proportion  0.80068 0.80189 0.80310 0.80430 0.8055 0.80669 0.80788
```

```
##                           PC302   PC303   PC304   PC305   PC306   PC307   PC308
## Standard deviation     0.83639 0.83501 0.83241 0.83149 0.82956 0.82690 0.82569
## Proportion of Variance 0.00119 0.00118 0.00117 0.00117 0.00117 0.00116 0.00116
## Cumulative Proportion  0.80907 0.81025 0.81143 0.81260 0.81376 0.81492 0.81608
##                           PC309   PC310   PC311   PC312   PC313   PC314   PC315
## Standard deviation     0.82491 0.82316 0.82262 0.82202 0.82025 0.81832 0.81709
## Proportion of Variance 0.00115 0.00115 0.00115 0.00115 0.00114 0.00113 0.00113
## Cumulative Proportion  0.81723 0.81838 0.81953 0.82067 0.82181 0.82295 0.82408
##                           PC316   PC317   PC318   PC319  PC320  PC321  PC322
## Standard deviation     0.81376 0.81171 0.81032 0.80839 0.8073 0.8060 0.8048
## Proportion of Variance 0.00112 0.00112 0.00111 0.00111 0.0011 0.0011 0.0011
## Cumulative Proportion  0.82520 0.82632 0.82743 0.82854 0.8296 0.8307 0.8318
##                           PC323   PC324   PC325   PC326   PC327   PC328   PC329
## Standard deviation     0.80241 0.80033 0.79963 0.79873 0.79662 0.79482 0.79439
## Proportion of Variance 0.00109 0.00109 0.00108 0.00108 0.00108 0.00107 0.00107
## Cumulative Proportion  0.83293 0.83402 0.83510 0.83618 0.83726 0.83833 0.83940
##                           PC330   PC331   PC332   PC333   PC334   PC335   PC336
## Standard deviation     0.79200 0.79025 0.78780 0.78554 0.78394 0.78262 0.78094
## Proportion of Variance 0.00106 0.00106 0.00105 0.00105 0.00104 0.00104 0.00103
## Cumulative Proportion  0.84046 0.84152 0.84257 0.84362 0.84466 0.84570 0.84673
##                           PC337   PC338   PC339   PC340   PC341   PC342   PC343
## Standard deviation     0.78045 0.77963 0.77857 0.77679 0.77377 0.77299 0.77142
## Proportion of Variance 0.00103 0.00103 0.00103 0.00102 0.00101 0.00101 0.00101
## Cumulative Proportion  0.84777 0.84880 0.84982 0.85085 0.85186 0.85287 0.85388
##                          PC344  PC345   PC346   PC347   PC348   PC349   PC350
## Standard deviation     0.7694 0.7678 0.76497 0.76353 0.76227 0.76154 0.76001
## Proportion of Variance 0.0010 0.0010 0.00099 0.00099 0.00098 0.00098 0.00098
## Cumulative Proportion  0.8549 0.8559 0.85688 0.85786 0.85885 0.85983 0.86081
##                           PC351   PC352   PC353   PC354   PC355   PC356   PC357
## Standard deviation     0.75814 0.75673 0.75557 0.75421 0.75392 0.75168 0.74993
## Proportion of Variance 0.00097 0.00097 0.00097 0.00096 0.00096 0.00096 0.00095
## Cumulative Proportion  0.86179 0.86276 0.86372 0.86469 0.86565 0.86661 0.86756
##                           PC358   PC359   PC360   PC361   PC362   PC363   PC364
## Standard deviation     0.74860 0.74813 0.74641 0.74356 0.74172 0.74073 0.73911
## Proportion of Variance 0.00095 0.00095 0.00094 0.00094 0.00093 0.00093 0.00093
## Cumulative Proportion  0.86851 0.86946 0.87040 0.87134 0.87227 0.87320 0.87413
##                           PC365   PC366   PC367   PC368   PC369  PC370  PC371
## Standard deviation     0.73861 0.73666 0.73529 0.73335 0.73076 0.7301 0.7273
## Proportion of Variance 0.00092 0.00092 0.00092 0.00091 0.00091 0.0009 0.0009
## Cumulative Proportion  0.87505 0.87597 0.87689 0.87780 0.87871 0.8796 0.8805
##                           PC372   PC373   PC374   PC375   PC376   PC377   PC378
## Standard deviation     0.72435 0.72423 0.72235 0.72183 0.72001 0.71783 0.71634
## Proportion of Variance 0.00089 0.00089 0.00088 0.00088 0.00088 0.00087 0.00087
## Cumulative Proportion  0.88140 0.88229 0.88317 0.88405 0.88493 0.88581 0.88667
##                           PC379   PC380   PC381   PC382   PC383   PC384   PC385
## Standard deviation     0.71548 0.71362 0.71287 0.71171 0.71020 0.70886 0.70752
## Proportion of Variance 0.00087 0.00086 0.00086 0.00086 0.00085 0.00085 0.00085
## Cumulative Proportion  0.88754 0.88841 0.88927 0.89013 0.89098 0.89183 0.89268
##                           PC386   PC387   PC388   PC389   PC390   PC391   PC392
## Standard deviation     0.70566 0.70439 0.70322 0.70178 0.70001 0.69860 0.69699
## Proportion of Variance 0.00084 0.00084 0.00084 0.00083 0.00083 0.00083 0.00082
## Cumulative Proportion  0.89352 0.89437 0.89520 0.89604 0.89687 0.89770 0.89852
##                           PC393   PC394   PC395   PC396   PC397   PC398  PC399
## Standard deviation     0.69521 0.69424 0.69303 0.69193 0.69072 0.68942 0.6879
## Proportion of Variance 0.00082 0.00082 0.00081 0.00081 0.00081 0.00081 0.0008
## Cumulative Proportion  0.89934 0.90016 0.90097 0.90178 0.90259 0.90340 0.9042
##                          PC400  PC401   PC402   PC403   PC404   PC405   PC406
## Standard deviation     0.6874 0.6851 0.68220 0.68103 0.68024 0.67932 0.67860
## Proportion of Variance 0.0008 0.0008 0.00079 0.00079 0.00078 0.00078 0.00078
## Cumulative Proportion  0.9050 0.9058 0.90658 0.90737 0.90815 0.90894 0.90972
##                           PC407   PC408   PC409   PC410   PC411   PC412   PC413
## Standard deviation     0.67575 0.67492 0.67435 0.67307 0.67092 0.66939 0.66864
## Proportion of Variance 0.00077 0.00077 0.00077 0.00077 0.00076 0.00076 0.00076
## Cumulative Proportion  0.91049 0.91126 0.91203 0.91280 0.91356 0.91432 0.91508
##                           PC414   PC415   PC416   PC417   PC418   PC419   PC420
## Standard deviation     0.66793 0.66555 0.66376 0.66299 0.66058 0.65715 0.65646
## Proportion of Variance 0.00076 0.00075 0.00075 0.00075 0.00074 0.00073 0.00073
## Cumulative Proportion  0.91584 0.91659 0.91733 0.91808 0.91882 0.91955 0.92028
##                           PC421   PC422   PC423   PC424   PC425   PC426   PC427
## Standard deviation     0.65486 0.65314 0.65246 0.65082 0.64788 0.64728 0.64516
## Proportion of Variance 0.00073 0.00072 0.00072 0.00072 0.00071 0.00071 0.00071
## Cumulative Proportion  0.92101 0.92173 0.92245 0.92317 0.92388 0.92459 0.92530
##                           PC428  PC429   PC430   PC431   PC432   PC433   PC434
## Standard deviation     0.64498 0.6436 0.64023 0.63907 0.63844 0.63690 0.63445
## Proportion of Variance 0.00071 0.0007 0.00069 0.00069 0.00069 0.00069 0.00068
## Cumulative Proportion  0.92600 0.9267 0.92740 0.92809 0.92878 0.92947 0.93015
##                           PC435   PC436   PC437   PC438   PC439   PC440   PC441
## Standard deviation     0.63410 0.63337 0.63200 0.62878 0.62741 0.62659 0.62543
## Proportion of Variance 0.00068 0.00068 0.00068 0.00067 0.00067 0.00067 0.00066
## Cumulative Proportion  0.93083 0.93151 0.93219 0.93286 0.93353 0.93419 0.93486
##                           PC442   PC443   PC444   PC445   PC446   PC447   PC448
## Standard deviation     0.62451 0.62170 0.62070 0.62036 0.61765 0.61670 0.61536
## Proportion of Variance 0.00066 0.00066 0.00065 0.00065 0.00065 0.00064 0.00064
## Cumulative Proportion  0.93552 0.93617 0.93683 0.93748 0.93812 0.93877 0.93941
##                           PC449   PC450   PC451   PC452   PC453   PC454   PC455
## Standard deviation     0.61258 0.61184 0.60959 0.60838 0.60764 0.60642 0.60482
## Proportion of Variance 0.00064 0.00063 0.00063 0.00063 0.00063 0.00062 0.00062
```

```
## Cumulative Proportion  0.94005 0.94068 0.94131 0.94194 0.94256 0.94319 0.94381
##                          PC456   PC457   PC458   PC459   PC460  PC461   PC462
## Standard deviation     0.60306 0.60259 0.60095 0.59989 0.59752 0.5949 0.59247
## Proportion of Variance 0.00062 0.00062 0.00061 0.00061 0.00061 0.0006 0.00059
## Cumulative Proportion  0.94442 0.94504 0.94565 0.94626 0.94687 0.9475 0.94806
##                          PC463   PC464   PC465   PC466   PC467   PC468   PC469
## Standard deviation     0.59176 0.59104 0.58951 0.58802 0.58558 0.58307 0.58207
## Proportion of Variance 0.00059 0.00059 0.00059 0.00059 0.00058 0.00058 0.00057
## Cumulative Proportion  0.94866 0.94925 0.94984 0.95042 0.95100 0.95158 0.95215
##                          PC470   PC471   PC472   PC473   PC474   PC475   PC476
## Standard deviation     0.58166 0.58038 0.58014 0.57801 0.57562 0.57446 0.57398
## Proportion of Variance 0.00057 0.00057 0.00057 0.00057 0.00056 0.00056 0.00056
## Cumulative Proportion  0.95273 0.95330 0.95387 0.95443 0.95500 0.95556 0.95611
##                          PC477   PC478   PC479   PC480   PC481   PC482   PC483
## Standard deviation     0.57296 0.57155 0.56866 0.56784 0.56547 0.56352 0.56314
## Proportion of Variance 0.00056 0.00055 0.00055 0.00055 0.00054 0.00054 0.00054
## Cumulative Proportion  0.95667 0.95722 0.95777 0.95832 0.95886 0.95940 0.95994
##                          PC484   PC485   PC486   PC487   PC488   PC489   PC490
## Standard deviation     0.56074 0.55891 0.55789 0.55617 0.55399 0.55364 0.55258
## Proportion of Variance 0.00053 0.00053 0.00053 0.00052 0.00052 0.00052 0.00052
## Cumulative Proportion  0.96047 0.96100 0.96153 0.96205 0.96257 0.96309 0.96361
##                          PC491   PC492   PC493   PC494   PC495  PC496  PC497
## Standard deviation     0.55129 0.55069 0.54885 0.54809 0.54614 0.5450 0.5414
## Proportion of Variance 0.00052 0.00051 0.00051 0.00051 0.00051 0.0005 0.0005
## Cumulative Proportion  0.96412 0.96464 0.96515 0.96566 0.96616 0.9667 0.9672
##                          PC498   PC499   PC500   PC501   PC502   PC503   PC504
## Standard deviation     0.54029 0.53794 0.53692 0.53417 0.53267 0.53080 0.52853
## Proportion of Variance 0.00049 0.00049 0.00049 0.00048 0.00048 0.00048 0.00047
## Cumulative Proportion  0.96766 0.96815 0.96864 0.96912 0.96960 0.97008 0.97055
##                          PC505   PC506   PC507   PC508   PC509   PC510   PC511
## Standard deviation     0.52645 0.52510 0.52410 0.52132 0.52005 0.51683 0.51607
## Proportion of Variance 0.00047 0.00047 0.00047 0.00046 0.00046 0.00045 0.00045
## Cumulative Proportion  0.97102 0.97149 0.97195 0.97242 0.97287 0.97333 0.97378
##                          PC512   PC513   PC514   PC515   PC516   PC517   PC518
## Standard deviation     0.51492 0.51328 0.51277 0.51081 0.50861 0.50740 0.50599
## Proportion of Variance 0.00045 0.00045 0.00045 0.00044 0.00044 0.00044 0.00043
## Cumulative Proportion  0.97423 0.97467 0.97512 0.97556 0.97600 0.97644 0.97687
##                          PC519   PC520   PC521   PC522   PC523   PC524   PC525
## Standard deviation     0.50478 0.50165 0.49954 0.49874 0.49694 0.49569 0.49411
## Proportion of Variance 0.00043 0.00043 0.00042 0.00042 0.00042 0.00042 0.00041
## Cumulative Proportion  0.97730 0.97773 0.97815 0.97857 0.97899 0.97941 0.97982
##                          PC526   PC527  PC528  PC529  PC530  PC531   PC532
## Standard deviation     0.49205 0.48959 0.4875 0.4863 0.4858 0.4841 0.48226
## Proportion of Variance 0.00041 0.00041 0.0004 0.0004 0.0004 0.0004 0.00039
## Cumulative Proportion  0.98023 0.98064 0.9810 0.9814 0.9818 0.9822 0.98263
##                          PC533   PC534   PC535   PC536   PC537   PC538   PC539
## Standard deviation     0.48011 0.47883 0.47592 0.47311 0.47260 0.47124 0.46993
## Proportion of Variance 0.00039 0.00039 0.00038 0.00038 0.00038 0.00038 0.00037
## Cumulative Proportion  0.98302 0.98341 0.98380 0.98418 0.98456 0.98493 0.98531
##                          PC540   PC541   PC542   PC543   PC544   PC545   PC546
## Standard deviation     0.46760 0.46507 0.46467 0.46206 0.46063 0.45912 0.45800
## Proportion of Variance 0.00037 0.00037 0.00037 0.00036 0.00036 0.00036 0.00036
## Cumulative Proportion  0.98568 0.98604 0.98641 0.98677 0.98713 0.98749 0.98784
##                          PC547   PC548   PC549   PC550   PC551   PC552   PC553
## Standard deviation     0.45681 0.45544 0.45239 0.45107 0.45000 0.44480 0.44260
## Proportion of Variance 0.00035 0.00035 0.00035 0.00034 0.00034 0.00034 0.00033
## Cumulative Proportion  0.98820 0.98855 0.98890 0.98924 0.98958 0.98992 0.99025
##                          PC554   PC555   PC556   PC557   PC558   PC559   PC560
## Standard deviation     0.44033 0.43858 0.43703 0.43526 0.43363 0.43018 0.42652
## Proportion of Variance 0.00033 0.00033 0.00032 0.00032 0.00032 0.00031 0.00031
## Cumulative Proportion  0.99058 0.99091 0.99123 0.99155 0.99187 0.99218 0.99249
##                          PC561  PC562  PC563  PC564  PC565   PC566   PC567
## Standard deviation     0.42458 0.4235 0.4210 0.4204 0.4174 0.41629 0.41238
## Proportion of Variance 0.00031 0.0003 0.0003 0.0003 0.0003 0.00029 0.00029
## Cumulative Proportion  0.99280 0.9931 0.9934 0.9937 0.9940 0.99429 0.99458
##                          PC568   PC569   PC570   PC571   PC572   PC573   PC574
## Standard deviation     0.41214 0.41140 0.40976 0.40419 0.40195 0.40039 0.39852
## Proportion of Variance 0.00029 0.00029 0.00028 0.00028 0.00027 0.00027 0.00027
## Cumulative Proportion  0.99487 0.99515 0.99544 0.99571 0.99599 0.99626 0.99653
##                          PC575   PC576   PC577   PC578   PC579   PC580   PC581
## Standard deviation     0.39599 0.39465 0.39182 0.38871 0.38587 0.38042 0.37816
## Proportion of Variance 0.00027 0.00026 0.00026 0.00026 0.00025 0.00025 0.00024
## Cumulative Proportion  0.99679 0.99706 0.99732 0.99757 0.99783 0.99807 0.99831
##                          PC582   PC583   PC584   PC585   PC586   PC587   PC588
## Standard deviation     0.37434 0.37227 0.36996 0.36301 0.35779 0.35396 0.33813
## Proportion of Variance 0.00024 0.00023 0.00023 0.00022 0.00022 0.00021 0.00019
## Cumulative Proportion  0.99855 0.99879 0.99902 0.99924 0.99946 0.99967 0.99987
##                          PC589   PC590
## Standard deviation     0.27808 0.04278
## Proportion of Variance 0.00013 0.00000
## Cumulative Proportion  1.00000 1.00000
```

The same procedure is repeated for the test data.

Using supervised learning techniques such as predictive models using Random Forest and KNN, the below results are obtainted.

```
##Random Forest
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:gridExtra':
##
##     combine
```

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

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
train_class_forest = train_class
train_class_forest$author = factor(train_class_forest$author)
randomFor = randomForest(author ~ . ,
                         data=train_class_forest, importance = TRUE)
predict_random<-predict(randomFor,data=test_class)
random_table<-as.data.frame(table(predict_random,as.factor(test_class$author)))
predicted<-predict_random
actual<-as.factor(test_class$author)
temp<-as.data.frame(cbind(actual,predicted))
temp$flag<-ifelse(temp$actual==temp$predicted,1,0)
sum(temp$flag)
```

```
## [1] 1791
```

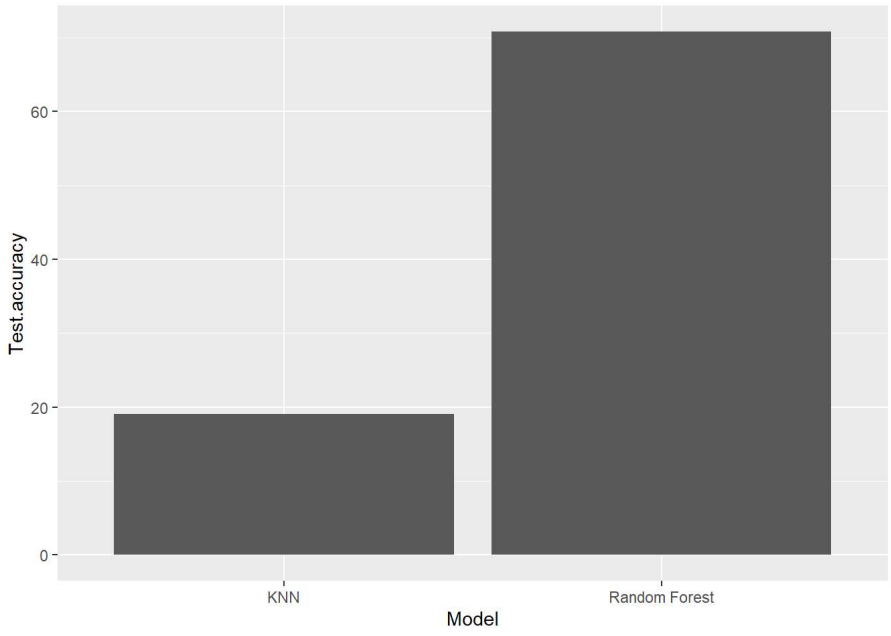
```
sum(temp$flag)*100/2500
```

```
## [1] 71.64
```

```
comp<-data.frame("Model"=c("Random Forest","KNN"), "Test.accuracy"=c(70.8,19.1))
comp
```

```
##           Model Test.accuracy
## 1 Random Forest          70.8
## 2           KNN          19.1
```

```
ggplot(comp,aes(x=Model,y=Test.accuracy))+geom_col()
```



It is clear from the results that the Random Forest technique performs significantly better than the KNN method by around 39%.