INF0613 – Aprendizado de Máquina Não Supervisionado

Trabalho 2 - Redução de Dimensionalidade

Evandro Santos Rocha Laíssa Pacheco de Oliveira Rafael Dantas de Moura

O objetivo deste trabalho é exercitar o conhecimento de técnicas de redução de dimensionalidade. Essas técnicas serão usadas tanto para obtenção de características quanto para visualização dos conjuntos de dados. Usaremos a base de dados `speech.csv`, que está disponível na página da disciplina no Moodle. A base contém amostras da pronúncia em inglês das letras do alfabeto.

Atividade 0 – Configurando o ambiente

Antes de começar a implementação do seu trabalho configure o *workspace* e importe todos os pacotes e execute o preprocessamento da base:

```
# Adicione os demais pacotes usados neste trabalho:
library(umap)
library(Rtsne)
library(datasets)
library(stats)

# Configure ambiente de trabalho na mesma pasta
# onde colocou a base de dados:
# setwd("")

# Pré-processamento da base de dados
# Lendo a base de dados
speech <- read.csv("speech.csv", header = TRUE)

# Convertendo a coluna 618 em characteres
speech$LETRA <- LETTERS[speech$LETRA]
```

Atividade 1 – Análise de Componentes Principais (*3,5 pts*)

Durante a redução de dimensionalidade, espera-se que o poder de representação do conjunto de dados seja mantido, para isso é preciso realizar uma análise da variância mantida em cada componente principal obtido. Use função `prcomp`, que foi vista em aula, para criar os autovetores e autovalores da base de dados. Não use a normalização dos atributos, isto é, defina `scale.=FALSE`. Em seguida, use o comando `summary`, analise o resultado e os itens a seguir:

```
# Executando a redução de dimensionalidade com o prcomp
speech.pca1 <- prcomp ( speech [ ,1:617])

# Analisando as componentes com o comando summary
```

```
summary(speech.pca1, options=options(max.print=2000))
```

```
## Importance of components:
##                           PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8
## Standard deviation     5.349 3.2000 2.6908 2.2953 2.1593 1.9522 1.7910 1.6023
## Proportion of Variance 0.248 0.0889 0.0629 0.0457 0.0405 0.0331 0.0278 0.0223
## Cumulative Proportion  0.248 0.3374 0.4003 0.4460 0.4865 0.5196 0.5474 0.5697
##                           PC9   PC10   PC11   PC12   PC13   PC14   PC15   PC16
## Standard deviation     1.5045 1.4245 1.2926 1.2461 1.2336 1.2017 1.1303 1.0804
## Proportion of Variance 0.0197 0.0176 0.0145 0.0135 0.0132 0.0125 0.0111 0.0101
## Cumulative Proportion  0.5894 0.6070 0.6215 0.6350 0.6482 0.6608 0.6719 0.6820
##                           PC17   PC18    PC19    PC20    PC21    PC22    PC23
## Standard deviation     1.05784 1.0295 0.98863 0.91550 0.89588 0.87390 0.84188
## Proportion of Variance 0.00972 0.0092 0.00849 0.00728 0.00697 0.00663 0.00615
## Cumulative Proportion  0.69171 0.7009 0.70940 0.71668 0.72365 0.73028 0.73643
##                           PC24   PC25    PC26    PC27    PC28    PC29    PC30
## Standard deviation     0.83022 0.8170 0.79311 0.77123 0.76799 0.71928 0.70613
## Proportion of Variance 0.00599 0.0058 0.00546 0.00516 0.00512 0.00449 0.00433
## Cumulative Proportion  0.74242 0.7482 0.75368 0.75884 0.76396 0.76846 0.77279
##                           PC31    PC32    PC33    PC34    PC35   PC36    PC37
## Standard deviation     0.70241 0.68645 0.66699 0.66407 0.65927 0.6437 0.63143
## Proportion of Variance 0.00428 0.00409 0.00386 0.00383 0.00377 0.0036 0.00346
## Cumulative Proportion  0.77707 0.78116 0.78502 0.78885 0.79263 0.7962 0.79969
##                           PC38    PC39    PC40    PC41    PC42    PC43    PC44
## Standard deviation     0.62465 0.61813 0.60593 0.59228 0.58506 0.57545 0.56941
## Proportion of Variance 0.00339 0.00332 0.00319 0.00305 0.00297 0.00288 0.00282
## Cumulative Proportion  0.80308 0.80639 0.80958 0.81263 0.81560 0.81848 0.82129
##                           PC45    PC46    PC47    PC48    PC49    PC50    PC51
## Standard deviation     0.56337 0.54403 0.54097 0.53450 0.52676 0.51723 0.51190
## Proportion of Variance 0.00276 0.00257 0.00254 0.00248 0.00241 0.00232 0.00228
## Cumulative Proportion  0.82405 0.82662 0.82916 0.83164 0.83405 0.83637 0.83865
##                           PC52    PC53    PC54    PC55    PC56    PC57    PC58
## Standard deviation     0.51159 0.50076 0.49899 0.48862 0.48508 0.48256 0.47571
## Proportion of Variance 0.00227 0.00218 0.00216 0.00207 0.00204 0.00202 0.00197
## Cumulative Proportion  0.84092 0.84310 0.84526 0.84733 0.84938 0.85140 0.85336
##                           PC59   PC60    PC61   PC62    PC63    PC64   PC65
## Standard deviation     0.46961 0.4682 0.46133 0.4550 0.45228 0.44888 0.4428
## Proportion of Variance 0.00191 0.0019 0.00185 0.0018 0.00178 0.00175 0.0017
## Cumulative Proportion  0.85528 0.8572 0.85903 0.8608 0.86260 0.86435 0.8660
##                           PC66    PC67    PC68    PC69    PC70    PC71    PC72
## Standard deviation     0.43976 0.43609 0.42528 0.42399 0.41742 0.41356 0.40654
## Proportion of Variance 0.00168 0.00165 0.00157 0.00156 0.00151 0.00149 0.00144
## Cumulative Proportion  0.86773 0.86939 0.87096 0.87252 0.87403 0.87551 0.87695
##                           PC73   PC74    PC75    PC76    PC77    PC78    PC79
## Standard deviation     0.40362 0.4011 0.39800 0.39417 0.39041 0.38466 0.38160
## Proportion of Variance 0.00141 0.0014 0.00138 0.00135 0.00132 0.00128 0.00126
## Cumulative Proportion  0.87836 0.8798 0.88114 0.88249 0.88381 0.88509 0.88636
##                           PC80    PC81    PC82    PC83    PC84    PC85    PC86
## Standard deviation     0.37905 0.37721 0.37400 0.37267 0.36786 0.36389 0.36027
## Proportion of Variance 0.00125 0.00124 0.00121 0.00121 0.00118 0.00115 0.00113
## Cumulative Proportion  0.88761 0.88884 0.89006 0.89126 0.89244 0.89359 0.89471
##                           PC87    PC88    PC89    PC90    PC91    PC92  PC93
## Standard deviation     0.35867 0.35200 0.35112 0.35003 0.34729 0.34573 0.340
```

```
## Proportion of Variance 0.00112 0.00108 0.00107 0.00106 0.00105 0.00104 0.001
## Cumulative Proportion  0.89583 0.89691 0.89798 0.89904 0.90009 0.90113 0.902
##                            PC94    PC95    PC96    PC97    PC98    PC99   PC100
## Standard deviation     0.33821 0.33678 0.33376 0.33201 0.33077 0.32809 0.32299
## Proportion of Variance 0.00099 0.00098 0.00097 0.00096 0.00095 0.00093 0.00091
## Cumulative Proportion  0.90313 0.90411 0.90508 0.90603 0.90698 0.90792 0.90883
##                           PC101   PC102   PC103   PC104   PC105   PC106  PC107
## Standard deviation     0.3211 0.31732 0.31520 0.31160 0.31076 0.30925 0.3035
## Proportion of Variance 0.0009 0.00087 0.00086 0.00084 0.00084 0.00083 0.0008
## Cumulative Proportion  0.9097 0.91059 0.91146 0.91230 0.91314 0.91397 0.9148
##                            PC108   PC109   PC110   PC111   PC112   PC113   PC114
## Standard deviation     0.30226 0.30002 0.29792 0.29655 0.29600 0.29196 0.29119
## Proportion of Variance 0.00079 0.00078 0.00077 0.00076 0.00076 0.00074 0.00074
## Cumulative Proportion  0.91556 0.91634 0.91712 0.91788 0.91864 0.91938 0.92012
##                            PC115   PC116  PC117  PC118   PC119   PC120   PC121
## Standard deviation     0.29042 0.28768 0.2846 0.2841 0.28208 0.28058 0.27765
## Proportion of Variance 0.00073 0.00072 0.0007 0.0007 0.00069 0.00068 0.00067
## Cumulative Proportion  0.92085 0.92157 0.9223 0.9230 0.92366 0.92435 0.92501
##                            PC122   PC123   PC124   PC125   PC126   PC127   PC128
## Standard deviation     0.27738 0.27558 0.27383 0.27228 0.27140 0.26928 0.26829
## Proportion of Variance 0.00067 0.00066 0.00065 0.00064 0.00064 0.00063 0.00063
## Cumulative Proportion  0.92568 0.92634 0.92699 0.92764 0.92828 0.92891 0.92953
##                            PC129   PC130  PC131   PC132   PC133   PC134   PC135
## Standard deviation     0.26515 0.26464 0.2630 0.26040 0.25715 0.25608 0.25508
## Proportion of Variance 0.00061 0.00061 0.0006 0.00059 0.00057 0.00057 0.00056
## Cumulative Proportion  0.93014 0.93075 0.9314 0.93194 0.93251 0.93308 0.93365
##                            PC136   PC137   PC138   PC139   PC140   PC141   PC142
## Standard deviation     0.25459 0.25214 0.25100 0.24977 0.24743 0.24595 0.24406
## Proportion of Variance 0.00056 0.00055 0.00055 0.00054 0.00053 0.00053 0.00052
## Cumulative Proportion  0.93421 0.93476 0.93531 0.93585 0.93638 0.93691 0.93743
##                            PC143   PC144   PC145  PC146   PC147   PC148   PC149
## Standard deviation     0.24358 0.24322 0.24144 0.2396 0.23815 0.23715 0.23668
## Proportion of Variance 0.00052 0.00051 0.00051 0.0005 0.00049 0.00049 0.00049
## Cumulative Proportion  0.93794 0.93846 0.93896 0.9395 0.93995 0.94044 0.94093
##                            PC150   PC151   PC152   PC153   PC154   PC155   PC156
## Standard deviation     0.23504 0.23430 0.23244 0.23115 0.23025 0.23008 0.22889
## Proportion of Variance 0.00048 0.00048 0.00047 0.00046 0.00046 0.00046 0.00045
## Cumulative Proportion  0.94141 0.94188 0.94235 0.94282 0.94328 0.94374 0.94419
##                            PC157   PC158   PC159   PC160   PC161   PC162   PC163
## Standard deviation     0.22793 0.22634 0.22500 0.22365 0.22322 0.22149 0.22005
## Proportion of Variance 0.00045 0.00044 0.00044 0.00043 0.00043 0.00043 0.00042
## Cumulative Proportion  0.94464 0.94509 0.94553 0.94596 0.94639 0.94682 0.94724
##                            PC164   PC165   PC166   PC167   PC168  PC169   PC170
## Standard deviation     0.21936 0.21797 0.21752 0.21705 0.21615 0.2147 0.21314
## Proportion of Variance 0.00042 0.00041 0.00041 0.00041 0.00041 0.0004 0.00039
## Cumulative Proportion  0.94766 0.94807 0.94848 0.94889 0.94930 0.9497 0.95009
##                            PC171   PC172   PC173   PC174   PC175   PC176   PC177
## Standard deviation     0.21263 0.21151 0.20922 0.20886 0.20832 0.20679 0.20586
## Proportion of Variance 0.00039 0.00039 0.00038 0.00038 0.00038 0.00037 0.00037
## Cumulative Proportion  0.95048 0.95087 0.95125 0.95163 0.95201 0.95238 0.95275
##                            PC178   PC179   PC180   PC181   PC182   PC183   PC184
## Standard deviation     0.20388 0.20310 0.20248 0.20189 0.20106 0.19938 0.19882
## Proportion of Variance 0.00036 0.00036 0.00036 0.00035 0.00035 0.00035 0.00034
## Cumulative Proportion  0.95311 0.95347 0.95382 0.95418 0.95453 0.95487 0.95522
```

```
##                           PC185   PC186   PC187   PC188   PC189   PC190   PC191
## Standard deviation     0.19766 0.19676 0.19602 0.19553 0.19416 0.19335 0.19211
## Proportion of Variance 0.00034 0.00034 0.00033 0.00033 0.00033 0.00032 0.00032
## Cumulative Proportion  0.95556 0.95589 0.95623 0.95656 0.95688 0.95721 0.95753
##                           PC192   PC193   PC194   PC195   PC196  PC197  PC198
## Standard deviation     0.19161 0.19037 0.18988 0.18932 0.18871 0.1870 0.1862
## Proportion of Variance 0.00032 0.00031 0.00031 0.00031 0.00031 0.0003 0.0003
## Cumulative Proportion  0.95785 0.95816 0.95848 0.95879 0.95910 0.9594 0.9597
##                           PC199  PC200   PC201   PC202   PC203   PC204   PC205
## Standard deviation     0.1856 0.1850 0.18359 0.18300 0.18226 0.18067 0.17996
## Proportion of Variance 0.0003 0.0003 0.00029 0.00029 0.00029 0.00028 0.00028
## Cumulative Proportion  0.9600 0.9603 0.96059 0.96088 0.96117 0.96145 0.96173
##                           PC206   PC207   PC208   PC209   PC210   PC211   PC212
## Standard deviation     0.17964 0.17882 0.17770 0.17674 0.17630 0.17587 0.17504
## Proportion of Variance 0.00028 0.00028 0.00027 0.00027 0.00027 0.00027 0.00027
## Cumulative Proportion  0.96201 0.96229 0.96257 0.96284 0.96311 0.96338 0.96364
##                           PC213   PC214   PC215   PC216   PC217   PC218   PC219
## Standard deviation     0.17442 0.17413 0.17332 0.17285 0.17251 0.17139 0.17049
## Proportion of Variance 0.00026 0.00026 0.00026 0.00026 0.00026 0.00026 0.00025
## Cumulative Proportion  0.96391 0.96417 0.96443 0.96469 0.96495 0.96520 0.96546
##                           PC220   PC221   PC222   PC223   PC224   PC225   PC226
## Standard deviation     0.17021 0.16958 0.16794 0.16744 0.16714 0.16585 0.16495
## Proportion of Variance 0.00025 0.00025 0.00024 0.00024 0.00024 0.00024 0.00024
## Cumulative Proportion  0.96571 0.96596 0.96620 0.96644 0.96669 0.96693 0.96716
##                           PC227   PC228   PC229   PC230   PC231   PC232   PC233
## Standard deviation     0.16447 0.16372 0.16306 0.16297 0.16205 0.16175 0.16148
## Proportion of Variance 0.00023 0.00023 0.00023 0.00023 0.00023 0.00023 0.00023
## Cumulative Proportion  0.96740 0.96763 0.96786 0.96809 0.96832 0.96855 0.96877
##                           PC234   PC235   PC236   PC237   PC238   PC239   PC240
## Standard deviation     0.16029 0.15991 0.15859 0.15823 0.15748 0.15660 0.15623
## Proportion of Variance 0.00022 0.00022 0.00022 0.00022 0.00022 0.00021 0.00021
## Cumulative Proportion  0.96900 0.96922 0.96944 0.96965 0.96987 0.97008 0.97029
##                           PC241   PC242   PC243   PC244   PC245  PC246  PC247
## Standard deviation     0.15551 0.15536 0.15493 0.15457 0.15417 0.1531 0.1521
## Proportion of Variance 0.00021 0.00021 0.00021 0.00021 0.00021 0.0002 0.0002
## Cumulative Proportion  0.97050 0.97071 0.97092 0.97113 0.97134 0.9715 0.9717
##                           PC248  PC249  PC250  PC251  PC252   PC253   PC254
## Standard deviation     0.1517 0.1516 0.1508 0.1506 0.1499 0.14923 0.14875
## Proportion of Variance 0.0002 0.0002 0.0002 0.0002 0.0002 0.00019 0.00019
## Cumulative Proportion  0.9719 0.9721 0.9723 0.9725 0.9727 0.97292 0.97311
##                           PC255   PC256   PC257   PC258   PC259   PC260   PC261
## Standard deviation     0.14844 0.14752 0.14729 0.14689 0.14618 0.14588 0.14569
## Proportion of Variance 0.00019 0.00019 0.00019 0.00019 0.00019 0.00018 0.00018
## Cumulative Proportion  0.97331 0.97350 0.97368 0.97387 0.97406 0.97424 0.97443
##                           PC262   PC263   PC264   PC265   PC266   PC267   PC268
## Standard deviation     0.14532 0.14497 0.14379 0.14362 0.14339 0.14220 0.14198
## Proportion of Variance 0.00018 0.00018 0.00018 0.00018 0.00018 0.00018 0.00018
## Cumulative Proportion  0.97461 0.97479 0.97497 0.97515 0.97533 0.97550 0.97568
##                           PC269   PC270   PC271   PC272   PC273   PC274   PC275
## Standard deviation     0.14178 0.14138 0.14096 0.14011 0.13927 0.13916 0.13897
## Proportion of Variance 0.00017 0.00017 0.00017 0.00017 0.00017 0.00017 0.00017
## Cumulative Proportion  0.97585 0.97603 0.97620 0.97637 0.97654 0.97671 0.97687
##                           PC276   PC277   PC278   PC279   PC280   PC281   PC282
## Standard deviation     0.13890 0.13822 0.13769 0.13717 0.13675 0.13627 0.13542
```

```
## Proportion of Variance 0.00017 0.00017 0.00016 0.00016 0.00016 0.00016 0.00016
## Cumulative Proportion  0.97704 0.97721 0.97737 0.97754 0.97770 0.97786 0.97802
##                           PC283   PC284   PC285   PC286   PC287   PC288   PC289
## Standard deviation     0.13511 0.13462 0.13437 0.13367 0.13349 0.13294 0.13274
## Proportion of Variance 0.00016 0.00016 0.00016 0.00016 0.00015 0.00015 0.00015
## Cumulative Proportion  0.97818 0.97833 0.97849 0.97865 0.97880 0.97895 0.97911
##                           PC290   PC291   PC292   PC293   PC294   PC295   PC296
## Standard deviation     0.13254 0.13188 0.13161 0.13115 0.13028 0.13009 0.12974
## Proportion of Variance 0.00015 0.00015 0.00015 0.00015 0.00015 0.00015 0.00015
## Cumulative Proportion  0.97926 0.97941 0.97956 0.97971 0.97986 0.98001 0.98015
##                           PC297   PC298   PC299   PC300   PC301   PC302   PC303
## Standard deviation     0.12965 0.12894 0.12866 0.12824 0.12793 0.12733 0.12705
## Proportion of Variance 0.00015 0.00014 0.00014 0.00014 0.00014 0.00014 0.00014
## Cumulative Proportion  0.98030 0.98044 0.98059 0.98073 0.98087 0.98101 0.98115
##                           PC304   PC305   PC306   PC307   PC308   PC309   PC310
## Standard deviation     0.12674 0.12647 0.12604 0.12567 0.12548 0.12487 0.12451
## Proportion of Variance 0.00014 0.00014 0.00014 0.00014 0.00014 0.00014 0.00013
## Cumulative Proportion  0.98129 0.98143 0.98157 0.98171 0.98184 0.98198 0.98211
##                           PC311   PC312   PC313   PC314   PC315   PC316   PC317
## Standard deviation     0.12412 0.12356 0.12313 0.12289 0.12281 0.12256 0.12218
## Proportion of Variance 0.00013 0.00013 0.00013 0.00013 0.00013 0.00013 0.00013
## Cumulative Proportion  0.98225 0.98238 0.98251 0.98264 0.98277 0.98290 0.98303
##                           PC318   PC319   PC320   PC321   PC322   PC323   PC324
## Standard deviation     0.12180 0.12127 0.12067 0.12027 0.12011 0.12004 0.11953
## Proportion of Variance 0.00013 0.00013 0.00013 0.00013 0.00013 0.00013 0.00012
## Cumulative Proportion  0.98316 0.98329 0.98341 0.98354 0.98367 0.98379 0.98391
##                           PC325   PC326   PC327   PC328   PC329   PC330   PC331
## Standard deviation     0.11895 0.11885 0.11862 0.11827 0.11792 0.11778 0.11763
## Proportion of Variance 0.00012 0.00012 0.00012 0.00012 0.00012 0.00012 0.00012
## Cumulative Proportion  0.98404 0.98416 0.98428 0.98440 0.98452 0.98465 0.98477
##                           PC332   PC333   PC334   PC335   PC336   PC337   PC338
## Standard deviation     0.11723 0.11657 0.11631 0.11581 0.11538 0.11531 0.11444
## Proportion of Variance 0.00012 0.00012 0.00012 0.00012 0.00012 0.00012 0.00011
## Cumulative Proportion  0.98488 0.98500 0.98512 0.98524 0.98535 0.98547 0.98558
##                           PC339   PC340   PC341   PC342   PC343   PC344   PC345
## Standard deviation     0.11419 0.11379 0.11367 0.11342 0.11286 0.11267 0.11233
## Proportion of Variance 0.00011 0.00011 0.00011 0.00011 0.00011 0.00011 0.00011
## Cumulative Proportion  0.98569 0.98581 0.98592 0.98603 0.98614 0.98625 0.98636
##                           PC346   PC347   PC348   PC349   PC350   PC351   PC352
## Standard deviation     0.11190 0.11141 0.11120 0.11096 0.11080 0.11065 0.11036
## Proportion of Variance 0.00011 0.00011 0.00011 0.00011 0.00011 0.00011 0.00011
## Cumulative Proportion  0.98647 0.98658 0.98669 0.98679 0.98690 0.98701 0.98711
##                           PC353  PC354  PC355  PC356  PC357  PC358  PC359  PC360
## Standard deviation     0.10999 0.1096 0.1094 0.1090 0.1086 0.1084 0.1082 0.1081
## Proportion of Variance 0.00011 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001
## Cumulative Proportion  0.98722 0.9873 0.9874 0.9875 0.9876 0.9877 0.9878 0.9879
##                           PC361  PC362  PC363  PC364  PC365  PC366  PC367  PC368
## Standard deviation     0.1079 0.1075 0.1071 0.1066 0.1064 0.1063 0.1059 0.1057
## Proportion of Variance 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001
## Cumulative Proportion  0.9880 0.9881 0.9882 0.9883 0.9884 0.9885 0.9886 0.9887
##                           PC369  PC370  PC371   PC372   PC373   PC374   PC375
## Standard deviation     0.1051 0.1051 0.1047 0.10434 0.10408 0.10398 0.10365
## Proportion of Variance 0.0001 0.0001 0.0001 0.00009 0.00009 0.00009 0.00009
## Cumulative Proportion  0.9888 0.9889 0.9890 0.98911 0.98920 0.98929 0.98939
```

```
##                           PC376   PC377   PC378   PC379   PC380   PC381   PC382
## Standard deviation     0.10361 0.10304 0.10265 0.10248 0.10231 0.10208 0.10176
## Proportion of Variance 0.00009 0.00009 0.00009 0.00009 0.00009 0.00009 0.00009
## Cumulative Proportion  0.98948 0.98957 0.98966 0.98976 0.98985 0.98994 0.99003
##                           PC383   PC384   PC385   PC386   PC387   PC388   PC389
## Standard deviation     0.10145 0.10107 0.10082 0.10050 0.10019 0.09989 0.09930
## Proportion of Variance 0.00009 0.00009 0.00009 0.00009 0.00009 0.00009 0.00009
## Cumulative Proportion  0.99012 0.99021 0.99029 0.99038 0.99047 0.99056 0.99064
##                           PC390   PC391   PC392   PC393   PC394   PC395   PC396
## Standard deviation     0.09910 0.09904 0.09886 0.09852 0.09817 0.09812 0.09749
## Proportion of Variance 0.00009 0.00009 0.00008 0.00008 0.00008 0.00008 0.00008
## Cumulative Proportion  0.99073 0.99081 0.99090 0.99098 0.99106 0.99115 0.99123
##                           PC397   PC398   PC399   PC400   PC401   PC402   PC403
## Standard deviation     0.09737 0.09714 0.09680 0.09650 0.09627 0.09582 0.09553
## Proportion of Variance 0.00008 0.00008 0.00008 0.00008 0.00008 0.00008 0.00008
## Cumulative Proportion  0.99131 0.99139 0.99148 0.99156 0.99164 0.99172 0.99180
##                           PC404   PC405   PC406   PC407   PC408   PC409   PC410
## Standard deviation     0.09485 0.09469 0.09455 0.09431 0.09393 0.09383 0.09348
## Proportion of Variance 0.00008 0.00008 0.00008 0.00008 0.00008 0.00008 0.00008
## Cumulative Proportion  0.99187 0.99195 0.99203 0.99211 0.99218 0.99226 0.99234
##                           PC411   PC412   PC413   PC414   PC415   PC416   PC417
## Standard deviation     0.09320 0.09267 0.09263 0.09250 0.09169 0.09157 0.09143
## Proportion of Variance 0.00008 0.00007 0.00007 0.00007 0.00007 0.00007 0.00007
## Cumulative Proportion  0.99241 0.99249 0.99256 0.99263 0.99271 0.99278 0.99285
##                           PC418   PC419   PC420   PC421   PC422   PC423   PC424
## Standard deviation     0.09128 0.09086 0.09062 0.09042 0.08976 0.08951 0.08908
## Proportion of Variance 0.00007 0.00007 0.00007 0.00007 0.00007 0.00007 0.00007
## Cumulative Proportion  0.99293 0.99300 0.99307 0.99314 0.99321 0.99328 0.99335
##                           PC425   PC426   PC427   PC428   PC429   PC430   PC431
## Standard deviation     0.08871 0.08853 0.08839 0.08816 0.08795 0.08735 0.08695
## Proportion of Variance 0.00007 0.00007 0.00007 0.00007 0.00007 0.00007 0.00007
## Cumulative Proportion  0.99342 0.99348 0.99355 0.99362 0.99369 0.99375 0.99382
##                           PC432   PC433   PC434   PC435   PC436   PC437   PC438
## Standard deviation     0.08653 0.08637 0.08614 0.08580 0.08559 0.08530 0.08511
## Proportion of Variance 0.00007 0.00006 0.00006 0.00006 0.00006 0.00006 0.00006
## Cumulative Proportion  0.99388 0.99395 0.99401 0.99408 0.99414 0.99420 0.99427
##                           PC439   PC440   PC441   PC442   PC443   PC444   PC445
## Standard deviation     0.08457 0.08424 0.08407 0.08402 0.08366 0.08344 0.08325
## Proportion of Variance 0.00006 0.00006 0.00006 0.00006 0.00006 0.00006 0.00006
## Cumulative Proportion  0.99433 0.99439 0.99445 0.99451 0.99457 0.99463 0.99469
##                           PC446   PC447   PC448   PC449   PC450   PC451   PC452
## Standard deviation     0.08288 0.08282 0.08234 0.08232 0.08189 0.08183 0.08162
## Proportion of Variance 0.00006 0.00006 0.00006 0.00006 0.00006 0.00006 0.00006
## Cumulative Proportion  0.99475 0.99481 0.99487 0.99493 0.99499 0.99505 0.99511
##                           PC453   PC454   PC455   PC456   PC457   PC458   PC459
## Standard deviation     0.08117 0.08089 0.08067 0.08054 0.08039 0.07972 0.07961
## Proportion of Variance 0.00006 0.00006 0.00006 0.00006 0.00006 0.00006 0.00006
## Cumulative Proportion  0.99516 0.99522 0.99528 0.99533 0.99539 0.99544 0.99550
##                           PC460   PC461   PC462   PC463   PC464   PC465   PC466
## Standard deviation     0.07915 0.07906 0.07896 0.07870 0.07824 0.07806 0.07796
## Proportion of Variance 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005
## Cumulative Proportion  0.99555 0.99561 0.99566 0.99572 0.99577 0.99582 0.99587
##                           PC467   PC468   PC469   PC470   PC471   PC472   PC473
## Standard deviation     0.07778 0.07733 0.07710 0.07685 0.07655 0.07626 0.07598
```

```
## Proportion of Variance 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005
## Cumulative Proportion  0.99593 0.99598 0.99603 0.99608 0.99613 0.99618 0.99623
##                           PC474   PC475   PC476   PC477   PC478   PC479   PC480
## Standard deviation     0.07578 0.07556 0.07545 0.07504 0.07474 0.07453 0.07426
## Proportion of Variance 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005
## Cumulative Proportion  0.99628 0.99633 0.99638 0.99643 0.99648 0.99653 0.99658
##                           PC481   PC482   PC483   PC484   PC485   PC486   PC487
## Standard deviation     0.07397 0.07382 0.07358 0.07308 0.07277 0.07259 0.07244
## Proportion of Variance 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005
## Cumulative Proportion  0.99662 0.99667 0.99672 0.99676 0.99681 0.99686 0.99690
##                           PC488   PC489   PC490   PC491   PC492   PC493   PC494
## Standard deviation     0.07221 0.07199 0.07189 0.07152 0.07121 0.07104 0.07071
## Proportion of Variance 0.00005 0.00005 0.00004 0.00004 0.00004 0.00004 0.00004
## Cumulative Proportion  0.99695 0.99699 0.99704 0.99708 0.99712 0.99717 0.99721
##                           PC495   PC496   PC497   PC498   PC499   PC500   PC501
## Standard deviation     0.07048 0.06990 0.06959 0.06922 0.06868 0.06847 0.06825
## Proportion of Variance 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004
## Cumulative Proportion  0.99725 0.99730 0.99734 0.99738 0.99742 0.99746 0.99750
##                           PC502   PC503   PC504   PC505   PC506   PC507   PC508
## Standard deviation     0.06800 0.06792 0.06773 0.06732 0.06703 0.06694 0.06645
## Proportion of Variance 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004
## Cumulative Proportion  0.99754 0.99758 0.99762 0.99766 0.99770 0.99774 0.99778
##                           PC509   PC510   PC511   PC512   PC513   PC514   PC515
## Standard deviation     0.06636 0.06612 0.06590 0.06543 0.06533 0.06507 0.06484
## Proportion of Variance 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004
## Cumulative Proportion  0.99782 0.99785 0.99789 0.99793 0.99797 0.99800 0.99804
##                           PC516   PC517   PC518   PC519   PC520   PC521   PC522
## Standard deviation     0.06444 0.06439 0.06418 0.06401 0.06348 0.06315 0.06285
## Proportion of Variance 0.00004 0.00004 0.00004 0.00004 0.00003 0.00003 0.00003
## Cumulative Proportion  0.99808 0.99811 0.99815 0.99818 0.99822 0.99825 0.99829
##                           PC523   PC524   PC525   PC526   PC527   PC528   PC529
## Standard deviation     0.06248 0.06244 0.06205 0.06166 0.06154 0.06140 0.06082
## Proportion of Variance 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003
## Cumulative Proportion  0.99832 0.99836 0.99839 0.99842 0.99845 0.99849 0.99852
##                           PC530   PC531   PC532   PC533   PC534   PC535   PC536
## Standard deviation     0.06050 0.06032 0.06029 0.05998 0.05960 0.05920 0.05916
## Proportion of Variance 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003
## Cumulative Proportion  0.99855 0.99858 0.99861 0.99865 0.99868 0.99871 0.99874
##                           PC537   PC538   PC539   PC540   PC541   PC542   PC543
## Standard deviation     0.05878 0.05866 0.05847 0.05835 0.05798 0.05764 0.05755
## Proportion of Variance 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003
## Cumulative Proportion  0.99877 0.99880 0.99883 0.99886 0.99889 0.99891 0.99894
##                           PC544   PC545   PC546   PC547   PC548   PC549   PC550
## Standard deviation     0.05700 0.05677 0.05659 0.05625 0.05576 0.05560 0.05537
## Proportion of Variance 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003
## Cumulative Proportion  0.99897 0.99900 0.99903 0.99905 0.99908 0.99911 0.99914
##                           PC551   PC552   PC553   PC554   PC555   PC556   PC557
## Standard deviation     0.05527 0.05459 0.05408 0.05402 0.05372 0.05338 0.05290
## Proportion of Variance 0.00003 0.00003 0.00003 0.00003 0.00003 0.00002 0.00002
## Cumulative Proportion  0.99916 0.99919 0.99921 0.99924 0.99926 0.99929 0.99931
##                           PC558   PC559   PC560   PC561   PC562   PC563   PC564
## Standard deviation     0.05265 0.05243 0.05212 0.05170 0.05127 0.05102 0.05067
## Proportion of Variance 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002
## Cumulative Proportion  0.99934 0.99936 0.99938 0.99941 0.99943 0.99945 0.99947
```

```
##                           PC565   PC566   PC567   PC568   PC569   PC570   PC571
## Standard deviation     0.05051 0.04988 0.04958 0.04902 0.04855 0.04811 0.04783
## Proportion of Variance 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002
## Cumulative Proportion  0.99950 0.99952 0.99954 0.99956 0.99958 0.99960 0.99962
##                           PC572   PC573   PC574   PC575   PC576   PC577   PC578
## Standard deviation     0.04700 0.04634 0.04503 0.04488 0.04432 0.04385 0.04363
## Proportion of Variance 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002
## Cumulative Proportion  0.99964 0.99966 0.99968 0.99969 0.99971 0.99973 0.99974
##                           PC579   PC580   PC581   PC582   PC583   PC584   PC585
## Standard deviation     0.04288 0.04266 0.04153 0.03982 0.03865 0.03798 0.03724
## Proportion of Variance 0.00002 0.00002 0.00001 0.00001 0.00001 0.00001 0.00001
## Cumulative Proportion  0.99976 0.99978 0.99979 0.99980 0.99982 0.99983 0.99984
##                           PC586   PC587   PC588   PC589   PC590   PC591   PC592
## Standard deviation     0.03661 0.03617 0.03540 0.03421 0.03342 0.03273 0.03203
## Proportion of Variance 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001
## Cumulative Proportion  0.99985 0.99987 0.99988 0.99989 0.99990 0.99991 0.99991
##                           PC593 PC594   PC595   PC596   PC597   PC598   PC599
## Standard deviation     0.03146 3e-02 0.02884 0.02781 0.02699 0.02503 0.02410
## Proportion of Variance 0.00001 1e-05 0.00001 0.00001 0.00001 0.00001 0.00001
## Cumulative Proportion  0.99992 1e+00 0.99994 0.99994 0.99995 0.99996 0.99996
##                          PC600  PC601  PC602 PC603  PC604  PC605  PC606  PC607
## Standard deviation     0.0229 0.0225 0.0204  0.02 0.0189 0.0176 0.0168 0.0166
## Proportion of Variance 0.0000 0.0000 0.0000  0.00 0.0000 0.0000 0.0000 0.0000
## Cumulative Proportion  1.0000 1.0000 1.0000  1.00 1.0000 1.0000 1.0000 1.0000
##                          PC608  PC609 PC610  PC611  PC612  PC613  PC614   PC615
## Standard deviation     0.0149 0.0139 0.013 0.0124 0.0116 0.0114 0.0112 0.00971
## Proportion of Variance 0.0000 0.0000 0.000 0.0000 0.0000 0.0000 0.0000 0.00000
## Cumulative Proportion  1.0000 1.0000 1.000 1.0000 1.0000 1.0000 1.0000 1.00000
##                           PC616  PC617
## Standard deviation     0.00857 0.0079
## Proportion of Variance 0.00000 0.0000
## Cumulative Proportion  1.00000 1.0000
```

Análise

a) Qual o menor número de componentes, tal que a variância acumulada seja pelo menos `80%` do total?

Resposta: Para que a variância acumulada seja de pelo menos 80%, são necessários 38 componentes.

b) Qual o menor número de componentes, tal que a variância acumulada seja pelo menos `90%` do total?

Resposta: Para que a variância acumulada seja de pelo menos 90%, são necessários 91 componentes.

c) Qual o menor número de componentes, tal que a variância acumulada seja pelo menos `95%` do total?

Resposta: Para que a variância acumulada seja de pelo menos 95%, são necessários 170 componentes.

d) Qual o menor número de componentes, tal que a variância acumulada seja pelo menos `99%` do total?

Resposta: Para que a variância acumulada seja de pelo menos 99%, são necessários 382 componentes.

e) Faça um breve resumo dos resultados dos itens *a)-d)* destacando o impacto da redução de dimensionalidade.

Resposta: As respostas acima foram obtidas a partir da análise do valor *Cumulative Proportion*, retornado pela função `summary`. Esse valor representa a soma das variâncias dos componentes à medida que estes são adicionados ao modelo. A variância, por sua vez, representa o quanto cada componente é responsável pela variabilidade nos dados.

Nota-se que apenas 91 componentes já representam 90% da variância acumulada dos dados. A primeira impressão que temos, então, desse conjunto de dados, é que é possível utilizar um conjunto muito menor de atributos e ainda assim obter um resultado muito próximo dos dados originais. Caso se quisesse ser mais conservador e utilizar uma variância de 99% do total, seria necessário utilizar apenas 382 componentes, o que representa 63% do conjuno de componentes principais. Considerando que após o componente 382 a variância acumulada aumentará até alcançar 100% - que significa que aquele componente não adicionou variabilidade nos dados, se comparados ao componentes já considerados anteriormente -, é possível manter apenas os primeiros 382 componentes. Em outras palavras, a análise do conjunto de dados e dos componentes principais indica que é possível realizar uma significante redução da dimensionalidade (aproximadamente 37%, no caso de variância acumulada de 99%), e ainda assim manter o conjunto de dados extremamente próximo ao original.

Atividade 2 – Análise de Componentes Principais e Normalização (*3,5 pts*)

A normalização de dados em alguns casos, pode trazer benefícios. Nesta questão, iremos analisar o impacto dessa prática na redução da dimensionalidade da base de dados `speech.csv`. Use função `prcomp` para criar os autovetores e autovalores da base de dados usando a normalização dos atributos, isto é, defina `scale.=TRUE`. Em seguida, use o comando `summary`, analise o resultado e os itens a seguir:

```
# Executando a redução de dimensionalidade com o prcomp
 # com normalização dos dados
speech.pca2 <- prcomp ( speech [ ,1:617], scale.=TRUE)

# Analisando as componentes com o comando summary
summary(speech.pca2, options=options(max.print=2000))
```

```
## Importance of components:
##                           PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8
## Standard deviation     10.914 7.4033 5.7923 5.2805 4.9284 4.4530 4.1197 3.8592
## Proportion of Variance  0.193 0.0888 0.0544 0.0452 0.0394 0.0321 0.0275 0.0241
## Cumulative Proportion   0.193 0.2819 0.3363 0.3815 0.4208 0.4530 0.4805 0.5046
##                           PC9   PC10   PC11   PC12  PC13   PC14   PC15  PC16
## Standard deviation     3.6160 3.5057 3.3011 3.0897 3.041 2.8763 2.7654 2.724
## Proportion of Variance 0.0212 0.0199 0.0177 0.0155 0.015 0.0134 0.0124 0.012
## Cumulative Proportion  0.5258 0.5457 0.5634 0.5789 0.594 0.6073 0.6197 0.632
##                          PC17   PC18    PC19    PC20    PC21   PC22    PC23
## Standard deviation     2.5758 2.5367 2.34228 2.30146 2.18822 2.1649 2.11362
## Proportion of Variance 0.0107 0.0104 0.00889 0.00858 0.00776 0.0076 0.00724
## Cumulative Proportion  0.6424 0.6529 0.66176 0.67034 0.67810 0.6857 0.69294
##                           PC24    PC25   PC26    PC27    PC28    PC29    PC30
## Standard deviation     2.02524 1.99286 1.9399 1.89350 1.85700 1.81195 1.76473
## Proportion of Variance 0.00665 0.00644 0.0061 0.00581 0.00559 0.00532 0.00505
## Cumulative Proportion  0.69959 0.70602 0.7121 0.71793 0.72352 0.72884 0.73389
```

```
##                            PC31    PC32    PC33    PC34    PC35    PC36    PC37
## Standard deviation     1.72980 1.70809 1.65417 1.63383 1.61280 1.58932 1.55833
## Proportion of Variance 0.00485 0.00473 0.00443 0.00433 0.00422 0.00409 0.00394
## Cumulative Proportion  0.73874 0.74347 0.74791 0.75223 0.75645 0.76054 0.76448
##                            PC38    PC39    PC40    PC41    PC42    PC43    PC44
## Standard deviation     1.50819 1.48283 1.45885 1.44102 1.41826 1.40915 1.38990
## Proportion of Variance 0.00369 0.00356 0.00345 0.00337 0.00326 0.00322 0.00313
## Cumulative Proportion  0.76816 0.77173 0.77518 0.77854 0.78180 0.78502 0.78815
##                            PC45    PC46  PC47    PC48   PC49    PC50    PC51
## Standard deviation     1.37981 1.37340 1.361 1.34277 1.3369 1.31970 1.30170
## Proportion of Variance 0.00309 0.00306 0.003 0.00292 0.0029 0.00282 0.00275
## Cumulative Proportion  0.79124 0.79429 0.797 0.80022 0.8031 0.80594 0.80868
##                           PC52   PC53    PC54    PC55    PC56    PC57    PC58
## Standard deviation     1.2672 1.2657 1.24958 1.23123 1.22014 1.20912 1.18784
## Proportion of Variance 0.0026 0.0026 0.00253 0.00246 0.00241 0.00237 0.00229
## Cumulative Proportion  0.8113 0.8139 0.81641 0.81887 0.82128 0.82365 0.82594
##                            PC59    PC60    PC61    PC62    PC63    PC64    PC65
## Standard deviation     1.16939 1.15843 1.15154 1.14573 1.12687 1.09262 1.07733
## Proportion of Variance 0.00222 0.00217 0.00215 0.00213 0.00206 0.00193 0.00188
## Cumulative Proportion  0.82816 0.83033 0.83248 0.83461 0.83667 0.83860 0.84048
##                            PC66    PC67    PC68    PC69    PC70    PC71   PC72
## Standard deviation     1.07034 1.05989 1.05122 1.04395 1.04185 1.03088 1.0248
## Proportion of Variance 0.00186 0.00182 0.00179 0.00177 0.00176 0.00172 0.0017
## Cumulative Proportion  0.84234 0.84416 0.84595 0.84772 0.84948 0.85120 0.8529
##                           PC73    PC74    PC75    PC76    PC77    PC78    PC79
## Standard deviation     0.9920 0.98325 0.97580 0.96650 0.95037 0.94873 0.94792
## Proportion of Variance 0.0016 0.00157 0.00154 0.00151 0.00146 0.00146 0.00146
## Cumulative Proportion  0.8545 0.85606 0.85761 0.85912 0.86058 0.86204 0.86350
##                           PC80    PC81    PC82    PC83    PC84    PC85    PC86
## Standard deviation     0.9293 0.92345 0.91685 0.91191 0.90634 0.89921 0.89062
## Proportion of Variance 0.0014 0.00138 0.00136 0.00135 0.00133 0.00131 0.00129
## Cumulative Proportion  0.8649 0.86628 0.86764 0.86899 0.87032 0.87163 0.87292
##                           PC87    PC88    PC89   PC90    PC91    PC92    PC93
## Standard deviation     0.88840 0.87796 0.86982 0.8607 0.85504 0.84934 0.84766
## Proportion of Variance 0.00128 0.00125 0.00123 0.0012 0.00118 0.00117 0.00116
## Cumulative Proportion  0.87420 0.87545 0.87667 0.8779 0.87906 0.88023 0.88139
##                            PC94    PC95    PC96    PC97    PC98    PC99   PC100
## Standard deviation     0.83382 0.82780 0.82699 0.81962 0.81062 0.80896 0.80160
## Proportion of Variance 0.00113 0.00111 0.00111 0.00109 0.00106 0.00106 0.00104
## Cumulative Proportion  0.88252 0.88363 0.88474 0.88583 0.88689 0.88795 0.88899
##                           PC101   PC102   PC103   PC104   PC105   PC106   PC107
## Standard deviation     0.79527 0.78247 0.78011 0.77870 0.77208 0.77014 0.76260
## Proportion of Variance 0.00103 0.00099 0.00099 0.00098 0.00097 0.00096 0.00094
## Cumulative Proportion  0.89002 0.89101 0.89200 0.89298 0.89395 0.89491 0.89585
##                           PC108   PC109   PC110  PC111   PC112   PC113   PC114
## Standard deviation     0.76021 0.75564 0.75217 0.7453 0.74159 0.73642 0.73103
## Proportion of Variance 0.00094 0.00093 0.00092 0.0009 0.00089 0.00088 0.00087
## Cumulative Proportion  0.89679 0.89771 0.89863 0.8995 0.90042 0.90130 0.90217
##                           PC115   PC116   PC117   PC118   PC119  PC120   PC121
## Standard deviation     0.72192 0.71820 0.71747 0.71403 0.71115 0.7016 0.69865
## Proportion of Variance 0.00084 0.00084 0.00083 0.00083 0.00082 0.0008 0.00079
## Cumulative Proportion  0.90301 0.90385 0.90468 0.90551 0.90633 0.9071 0.90792
##                           PC122   PC123   PC124   PC125   PC126   PC127   PC128
## Standard deviation     0.69384 0.68875 0.68769 0.67673 0.67481 0.67038 0.66661
```

```
## Proportion of Variance 0.00078 0.00077 0.00077 0.00074 0.00074 0.00073 0.00072
## Cumulative Proportion  0.90870 0.90946 0.91023 0.91097 0.91171 0.91244 0.91316
##                           PC129   PC130  PC131   PC132   PC133   PC134   PC135
## Standard deviation     0.66338 0.66054 0.6557 0.65458 0.64985 0.64773 0.64397
## Proportion of Variance 0.00071 0.00071 0.0007 0.00069 0.00068 0.00068 0.00067
## Cumulative Proportion  0.91387 0.91458 0.9153 0.91597 0.91666 0.91734 0.91801
##                           PC136   PC137   PC138   PC139   PC140   PC141   PC142
## Standard deviation     0.63905 0.63648 0.63355 0.62969 0.62619 0.61930 0.61644
## Proportion of Variance 0.00066 0.00066 0.00065 0.00064 0.00064 0.00062 0.00062
## Cumulative Proportion  0.91867 0.91933 0.91998 0.92062 0.92126 0.92188 0.92249
##                           PC143   PC144  PC145  PC146   PC147   PC148   PC149
## Standard deviation     0.61407 0.61380 0.6109 0.6078 0.60307 0.60152 0.59572
## Proportion of Variance 0.00061 0.00061 0.0006 0.0006 0.00059 0.00059 0.00058
## Cumulative Proportion  0.92310 0.92371 0.9243 0.9249 0.92551 0.92609 0.92667
##                           PC150   PC151   PC152   PC153   PC154   PC155   PC156
## Standard deviation     0.59224 0.58999 0.58872 0.58685 0.58482 0.58171 0.58132
## Proportion of Variance 0.00057 0.00056 0.00056 0.00056 0.00055 0.00055 0.00055
## Cumulative Proportion  0.92724 0.92780 0.92836 0.92892 0.92948 0.93002 0.93057
##                           PC157   PC158   PC159   PC160   PC161   PC162   PC163
## Standard deviation     0.57484 0.57472 0.56901 0.56585 0.56283 0.56205 0.56055
## Proportion of Variance 0.00054 0.00054 0.00052 0.00052 0.00051 0.00051 0.00051
## Cumulative Proportion  0.93111 0.93164 0.93217 0.93269 0.93320 0.93371 0.93422
##                           PC164  PC165   PC166   PC167   PC168   PC169   PC170
## Standard deviation     0.55912 0.5549 0.55094 0.54898 0.54566 0.54214 0.53917
## Proportion of Variance 0.00051 0.0005 0.00049 0.00049 0.00048 0.00048 0.00047
## Cumulative Proportion  0.93473 0.9352 0.93572 0.93621 0.93669 0.93717 0.93764
##                           PC171   PC172   PC173   PC174   PC175   PC176   PC177
## Standard deviation     0.53785 0.53756 0.53375 0.52910 0.52838 0.52542 0.52352
## Proportion of Variance 0.00047 0.00047 0.00046 0.00045 0.00045 0.00045 0.00044
## Cumulative Proportion  0.93811 0.93858 0.93904 0.93949 0.93994 0.94039 0.94083
##                           PC178   PC179   PC180   PC181   PC182   PC183   PC184
## Standard deviation     0.52263 0.52102 0.51724 0.51418 0.51289 0.50975 0.50713
## Proportion of Variance 0.00044 0.00044 0.00043 0.00043 0.00043 0.00042 0.00042
## Cumulative Proportion  0.94128 0.94172 0.94215 0.94258 0.94301 0.94343 0.94384
##                           PC185   PC186   PC187  PC188  PC189  PC190   PC191
## Standard deviation     0.50435 0.50197 0.50077 0.4988 0.4964 0.4953 0.49063
## Proportion of Variance 0.00041 0.00041 0.00041 0.0004 0.0004 0.0004 0.00039
## Cumulative Proportion  0.94426 0.94466 0.94507 0.9455 0.9459 0.9463 0.94666
##                           PC192   PC193   PC194   PC195   PC196   PC197   PC198
## Standard deviation     0.49023 0.48949 0.48683 0.48552 0.48340 0.48163 0.47951
## Proportion of Variance 0.00039 0.00039 0.00038 0.00038 0.00038 0.00038 0.00037
## Cumulative Proportion  0.94705 0.94744 0.94782 0.94821 0.94858 0.94896 0.94933
##                           PC199   PC200   PC201   PC202   PC203   PC204   PC205
## Standard deviation     0.47709 0.47563 0.47486 0.47336 0.46946 0.46853 0.46480
## Proportion of Variance 0.00037 0.00037 0.00037 0.00036 0.00036 0.00036 0.00035
## Cumulative Proportion  0.94970 0.95007 0.95043 0.95080 0.95115 0.95151 0.95186
##                           PC206   PC207   PC208   PC209   PC210   PC211   PC212
## Standard deviation     0.46410 0.46285 0.45944 0.45800 0.45761 0.45607 0.45439
## Proportion of Variance 0.00035 0.00035 0.00034 0.00034 0.00034 0.00034 0.00033
## Cumulative Proportion  0.95221 0.95256 0.95290 0.95324 0.95358 0.95391 0.95425
##                           PC213   PC214   PC215   PC216   PC217   PC218   PC219
## Standard deviation     0.45135 0.45057 0.45021 0.44759 0.44617 0.44432 0.44099
## Proportion of Variance 0.00033 0.00033 0.00033 0.00032 0.00032 0.00032 0.00032
## Cumulative Proportion  0.95458 0.95491 0.95524 0.95556 0.95588 0.95620 0.95652
```

```
##                           PC220   PC221   PC222   PC223  PC224  PC225  PC226
## Standard deviation     0.44008 0.43817 0.43714 0.43627 0.4337 0.4308 0.4301
## Proportion of Variance 0.00031 0.00031 0.00031 0.00031 0.0003 0.0003 0.0003
## Cumulative Proportion  0.95683 0.95714 0.95745 0.95776 0.9581 0.9584 0.9587
##                           PC227  PC228  PC229   PC230   PC231   PC232   PC233
## Standard deviation     0.4288 0.4278 0.4268 0.42488 0.42313 0.42176 0.41970
## Proportion of Variance 0.0003 0.0003 0.0003 0.00029 0.00029 0.00029 0.00029
## Cumulative Proportion  0.9590 0.9593 0.9596 0.95985 0.96014 0.96043 0.96071
##                           PC234   PC235   PC236   PC237   PC238   PC239   PC240
## Standard deviation     0.41756 0.41702 0.41547 0.41475 0.41300 0.41279 0.40929
## Proportion of Variance 0.00028 0.00028 0.00028 0.00028 0.00028 0.00028 0.00027
## Cumulative Proportion  0.96100 0.96128 0.96156 0.96184 0.96211 0.96239 0.96266
##                           PC241   PC242   PC243   PC244   PC245   PC246   PC247
## Standard deviation     0.40847 0.40617 0.40319 0.40166 0.40093 0.39981 0.39805
## Proportion of Variance 0.00027 0.00027 0.00026 0.00026 0.00026 0.00026 0.00026
## Cumulative Proportion  0.96293 0.96320 0.96346 0.96372 0.96399 0.96424 0.96450
##                           PC248   PC249   PC250   PC251   PC252   PC253   PC254
## Standard deviation     0.39751 0.39646 0.39465 0.39381 0.39281 0.39218 0.39062
## Proportion of Variance 0.00026 0.00025 0.00025 0.00025 0.00025 0.00025 0.00025
## Cumulative Proportion  0.96476 0.96501 0.96526 0.96552 0.96577 0.96601 0.96626
##                           PC255   PC256   PC257   PC258   PC259   PC260   PC261
## Standard deviation     0.38815 0.38732 0.38687 0.38548 0.38451 0.38378 0.38249
## Proportion of Variance 0.00024 0.00024 0.00024 0.00024 0.00024 0.00024 0.00024
## Cumulative Proportion  0.96651 0.96675 0.96699 0.96723 0.96747 0.96771 0.96795
##                           PC262   PC263   PC264   PC265   PC266   PC267   PC268
## Standard deviation     0.38143 0.38108 0.37933 0.37640 0.37635 0.37478 0.37292
## Proportion of Variance 0.00024 0.00024 0.00023 0.00023 0.00023 0.00023 0.00023
## Cumulative Proportion  0.96818 0.96842 0.96865 0.96888 0.96911 0.96934 0.96957
##                           PC269   PC270   PC271   PC272   PC273   PC274   PC275
## Standard deviation     0.37081 0.36985 0.36826 0.36737 0.36574 0.36524 0.36471
## Proportion of Variance 0.00022 0.00022 0.00022 0.00022 0.00022 0.00022 0.00022
## Cumulative Proportion  0.96979 0.97001 0.97023 0.97045 0.97066 0.97088 0.97110
##                           PC276   PC277   PC278   PC279   PC280   PC281   PC282
## Standard deviation     0.36177 0.36140 0.36131 0.35915 0.35882 0.35726 0.35617
## Proportion of Variance 0.00021 0.00021 0.00021 0.00021 0.00021 0.00021 0.00021
## Cumulative Proportion  0.97131 0.97152 0.97173 0.97194 0.97215 0.97236 0.97256
##                           PC283  PC284  PC285  PC286  PC287  PC288  PC289  PC290
## Standard deviation     0.35599 0.3546 0.3541 0.3533 0.3527 0.3520 0.3503 0.3494
## Proportion of Variance 0.00021 0.0002 0.0002 0.0002 0.0002 0.0002 0.0002 0.0002
## Cumulative Proportion  0.97277 0.9730 0.9732 0.9734 0.9736 0.9738 0.9740 0.9742
##                           PC291  PC292   PC293   PC294   PC295   PC296   PC297
## Standard deviation     0.3489 0.3470 0.34569 0.34504 0.34380 0.34311 0.34151
## Proportion of Variance 0.0002 0.0002 0.00019 0.00019 0.00019 0.00019 0.00019
## Cumulative Proportion  0.9744 0.9746 0.97476 0.97496 0.97515 0.97534 0.97553
##                           PC298   PC299   PC300   PC301   PC302   PC303   PC304
## Standard deviation     0.34017 0.33863 0.33814 0.33790 0.33688 0.33511 0.33490
## Proportion of Variance 0.00019 0.00019 0.00019 0.00019 0.00018 0.00018 0.00018
## Cumulative Proportion  0.97571 0.97590 0.97609 0.97627 0.97645 0.97664 0.97682
##                           PC305   PC306   PC307   PC308   PC309   PC310   PC311
## Standard deviation     0.33368 0.33214 0.33183 0.32985 0.32914 0.32803 0.32721
## Proportion of Variance 0.00018 0.00018 0.00018 0.00018 0.00018 0.00017 0.00017
## Cumulative Proportion  0.97700 0.97718 0.97736 0.97753 0.97771 0.97788 0.97806
##                           PC312   PC313   PC314   PC315   PC316   PC317   PC318
## Standard deviation     0.32562 0.32560 0.32511 0.32359 0.32235 0.32169 0.32099
```

```
## Proportion of Variance 0.00017 0.00017 0.00017 0.00017 0.00017 0.00017 0.00017
## Cumulative Proportion  0.97823 0.97840 0.97857 0.97874 0.97891 0.97908 0.97924
##                           PC319   PC320   PC321   PC322   PC323   PC324   PC325
## Standard deviation     0.32084 0.31946 0.31869 0.31780 0.31689 0.31526 0.31457
## Proportion of Variance 0.00017 0.00017 0.00016 0.00016 0.00016 0.00016 0.00016
## Cumulative Proportion  0.97941 0.97958 0.97974 0.97990 0.98007 0.98023 0.98039
##                           PC326   PC327   PC328   PC329   PC330   PC331   PC332
## Standard deviation     0.31445 0.31266 0.31203 0.31004 0.30947 0.30800 0.30764
## Proportion of Variance 0.00016 0.00016 0.00016 0.00016 0.00016 0.00015 0.00015
## Cumulative Proportion  0.98055 0.98071 0.98086 0.98102 0.98118 0.98133 0.98148
##                           PC333   PC334   PC335   PC336   PC337   PC338   PC339
## Standard deviation     0.30741 0.30630 0.30552 0.30513 0.30438 0.30229 0.30196
## Proportion of Variance 0.00015 0.00015 0.00015 0.00015 0.00015 0.00015 0.00015
## Cumulative Proportion  0.98164 0.98179 0.98194 0.98209 0.98224 0.98239 0.98254
##                           PC340   PC341   PC342   PC343   PC344   PC345   PC346
## Standard deviation     0.30091 0.29887 0.29851 0.29685 0.29648 0.29589 0.29546
## Proportion of Variance 0.00015 0.00014 0.00014 0.00014 0.00014 0.00014 0.00014
## Cumulative Proportion  0.98268 0.98283 0.98297 0.98312 0.98326 0.98340 0.98354
##                           PC347   PC348   PC349   PC350   PC351   PC352   PC353
## Standard deviation     0.29493 0.29335 0.29301 0.29235 0.29089 0.29046 0.28942
## Proportion of Variance 0.00014 0.00014 0.00014 0.00014 0.00014 0.00014 0.00014
## Cumulative Proportion  0.98368 0.98382 0.98396 0.98410 0.98424 0.98437 0.98451
##                           PC354   PC355   PC356   PC357   PC358   PC359   PC360
## Standard deviation     0.28813 0.28771 0.28613 0.28467 0.28392 0.28328 0.28260
## Proportion of Variance 0.00013 0.00013 0.00013 0.00013 0.00013 0.00013 0.00013
## Cumulative Proportion  0.98464 0.98478 0.98491 0.98504 0.98517 0.98530 0.98543
##                           PC361   PC362   PC363   PC364   PC365   PC366   PC367
## Standard deviation     0.28216 0.28169 0.28059 0.27965 0.27848 0.27760 0.27536
## Proportion of Variance 0.00013 0.00013 0.00013 0.00013 0.00013 0.00012 0.00012
## Cumulative Proportion  0.98556 0.98569 0.98582 0.98594 0.98607 0.98619 0.98632
##                           PC368   PC369   PC370   PC371   PC372   PC373   PC374
## Standard deviation     0.27517 0.27460 0.27347 0.27241 0.27084 0.27050 0.27004
## Proportion of Variance 0.00012 0.00012 0.00012 0.00012 0.00012 0.00012 0.00012
## Cumulative Proportion  0.98644 0.98656 0.98668 0.98680 0.98692 0.98704 0.98716
##                           PC375   PC376   PC377   PC378   PC379   PC380   PC381
## Standard deviation     0.26908 0.26771 0.26616 0.26562 0.26450 0.26416 0.26363
## Proportion of Variance 0.00012 0.00012 0.00011 0.00011 0.00011 0.00011 0.00011
## Cumulative Proportion  0.98728 0.98739 0.98751 0.98762 0.98774 0.98785 0.98796
##                           PC382   PC383   PC384   PC385   PC386   PC387   PC388
## Standard deviation     0.26289 0.26269 0.26183 0.26076 0.26013 0.25854 0.25810
## Proportion of Variance 0.00011 0.00011 0.00011 0.00011 0.00011 0.00011 0.00011
## Cumulative Proportion  0.98807 0.98818 0.98830 0.98841 0.98852 0.98862 0.98873
##                           PC389   PC390   PC391   PC392  PC393  PC394  PC395
## Standard deviation     0.25636 0.25548 0.25509 0.25464 0.2535 0.2526 0.2519
## Proportion of Variance 0.00011 0.00011 0.00011 0.00011 0.0001 0.0001 0.0001
## Cumulative Proportion  0.98884 0.98894 0.98905 0.98915 0.9893 0.9894 0.9895
##                           PC396  PC397  PC398  PC399  PC400  PC401  PC402  PC403
## Standard deviation     0.2510 0.2500 0.2492 0.2488 0.2480 0.2467 0.2464 0.2447
## Proportion of Variance 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001
## Cumulative Proportion  0.9896 0.9897 0.9898 0.9899 0.9900 0.9901 0.9902 0.9903
##                           PC404  PC405  PC406   PC407   PC408   PC409   PC410
## Standard deviation     0.2444 0.2431 0.2421 0.24138 0.24116 0.23997 0.23926
## Proportion of Variance 0.0001 0.0001 0.0001 0.00009 0.00009 0.00009 0.00009
## Cumulative Proportion  0.9904 0.9905 0.9906 0.99065 0.99074 0.99083 0.99093
```

```
##                           PC411   PC412   PC413   PC414   PC415   PC416   PC417
## Standard deviation     0.23861 0.23778 0.23755 0.23670 0.23580 0.23470 0.23400
## Proportion of Variance 0.00009 0.00009 0.00009 0.00009 0.00009 0.00009 0.00009
## Cumulative Proportion  0.99102 0.99111 0.99120 0.99129 0.99138 0.99147 0.99156
##                           PC418   PC419   PC420   PC421   PC422   PC423   PC424
## Standard deviation     0.23282 0.23134 0.23028 0.22985 0.22867 0.22826 0.22796
## Proportion of Variance 0.00009 0.00009 0.00009 0.00009 0.00008 0.00008 0.00008
## Cumulative Proportion  0.99165 0.99173 0.99182 0.99191 0.99199 0.99208 0.99216
##                           PC425   PC426   PC427   PC428   PC429   PC430   PC431
## Standard deviation     0.22727 0.22672 0.22584 0.22346 0.22326 0.22318 0.22164
## Proportion of Variance 0.00008 0.00008 0.00008 0.00008 0.00008 0.00008 0.00008
## Cumulative Proportion  0.99224 0.99233 0.99241 0.99249 0.99257 0.99265 0.99273
##                           PC432   PC433   PC434   PC435   PC436   PC437   PC438
## Standard deviation     0.22083 0.21980 0.21896 0.21833 0.21698 0.21650 0.21608
## Proportion of Variance 0.00008 0.00008 0.00008 0.00008 0.00008 0.00008 0.00008
## Cumulative Proportion  0.99281 0.99289 0.99297 0.99304 0.99312 0.99320 0.99327
##                           PC439   PC440   PC441   PC442   PC443   PC444   PC445
## Standard deviation     0.21546 0.21441 0.21404 0.21290 0.21266 0.21186 0.21097
## Proportion of Variance 0.00008 0.00007 0.00007 0.00007 0.00007 0.00007 0.00007
## Cumulative Proportion  0.99335 0.99342 0.99350 0.99357 0.99364 0.99372 0.99379
##                           PC446   PC447   PC448   PC449   PC450   PC451   PC452
## Standard deviation     0.21025 0.20940 0.20880 0.20767 0.20724 0.20628 0.20522
## Proportion of Variance 0.00007 0.00007 0.00007 0.00007 0.00007 0.00007 0.00007
## Cumulative Proportion  0.99386 0.99393 0.99400 0.99407 0.99414 0.99421 0.99428
##                           PC453   PC454   PC455   PC456   PC457   PC458   PC459
## Standard deviation     0.20468 0.20414 0.20354 0.20256 0.20173 0.20058 0.19938
## Proportion of Variance 0.00007 0.00007 0.00007 0.00007 0.00007 0.00007 0.00006
## Cumulative Proportion  0.99435 0.99441 0.99448 0.99455 0.99461 0.99468 0.99474
##                           PC460   PC461   PC462   PC463   PC464   PC465   PC466
## Standard deviation     0.19913 0.19831 0.19676 0.19585 0.19511 0.19439 0.19376
## Proportion of Variance 0.00006 0.00006 0.00006 0.00006 0.00006 0.00006 0.00006
## Cumulative Proportion  0.99481 0.99487 0.99493 0.99500 0.99506 0.99512 0.99518
##                           PC467   PC468   PC469   PC470   PC471   PC472   PC473
## Standard deviation     0.19237 0.19162 0.19136 0.19095 0.18983 0.18927 0.18824
## Proportion of Variance 0.00006 0.00006 0.00006 0.00006 0.00006 0.00006 0.00006
## Cumulative Proportion  0.99524 0.99530 0.99536 0.99542 0.99548 0.99553 0.99559
##                           PC474   PC475   PC476   PC477   PC478   PC479   PC480
## Standard deviation     0.18679 0.18646 0.18597 0.18519 0.18435 0.18333 0.18258
## Proportion of Variance 0.00006 0.00006 0.00006 0.00006 0.00006 0.00005 0.00005
## Cumulative Proportion  0.99565 0.99570 0.99576 0.99582 0.99587 0.99592 0.99598
##                           PC481   PC482   PC483   PC484   PC485   PC486   PC487
## Standard deviation     0.18186 0.18152 0.18028 0.17965 0.17900 0.17871 0.17773
## Proportion of Variance 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005
## Cumulative Proportion  0.99603 0.99609 0.99614 0.99619 0.99624 0.99629 0.99635
##                           PC488   PC489   PC490   PC491   PC492   PC493   PC494
## Standard deviation     0.17698 0.17596 0.17498 0.17426 0.17317 0.17255 0.17224
## Proportion of Variance 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005
## Cumulative Proportion  0.99640 0.99645 0.99650 0.99655 0.99659 0.99664 0.99669
##                           PC495   PC496   PC497   PC498   PC499   PC500   PC501
## Standard deviation     0.17146 0.17117 0.17042 0.17022 0.16984 0.16775 0.16720
## Proportion of Variance 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005 0.00005
## Cumulative Proportion  0.99674 0.99679 0.99683 0.99688 0.99693 0.99697 0.99702
##                           PC502   PC503   PC504   PC505   PC506   PC507   PC508
## Standard deviation     0.16642 0.16590 0.16539 0.16479 0.16365 0.16289 0.16259
```

```
## Proportion of Variance 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004
## Cumulative Proportion  0.99706 0.99711 0.99715 0.99720 0.99724 0.99728 0.99732
##                           PC509   PC510   PC511   PC512   PC513   PC514   PC515
## Standard deviation     0.16115 0.16025 0.15991 0.15910 0.15897 0.15830 0.15753
## Proportion of Variance 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004
## Cumulative Proportion  0.99737 0.99741 0.99745 0.99749 0.99753 0.99757 0.99761
##                           PC516   PC517   PC518   PC519   PC520   PC521   PC522
## Standard deviation     0.15681 0.15630 0.15613 0.15501 0.15490 0.15475 0.15420
## Proportion of Variance 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004
## Cumulative Proportion  0.99765 0.99769 0.99773 0.99777 0.99781 0.99785 0.99789
##                           PC523   PC524   PC525   PC526   PC527   PC528   PC529
## Standard deviation     0.15297 0.15230 0.15214 0.15137 0.15105 0.15045 0.14943
## Proportion of Variance 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004
## Cumulative Proportion  0.99792 0.99796 0.99800 0.99804 0.99807 0.99811 0.99815
##                           PC530   PC531   PC532   PC533   PC534   PC535   PC536
## Standard deviation     0.14891 0.14792 0.14697 0.14629 0.14542 0.14480 0.14457
## Proportion of Variance 0.00004 0.00004 0.00004 0.00003 0.00003 0.00003 0.00003
## Cumulative Proportion  0.99818 0.99822 0.99825 0.99829 0.99832 0.99836 0.99839
##                           PC537   PC538   PC539   PC540   PC541   PC542   PC543
## Standard deviation     0.14326 0.14306 0.14267 0.14209 0.14153 0.14066 0.13972
## Proportion of Variance 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003
## Cumulative Proportion  0.99842 0.99846 0.99849 0.99852 0.99855 0.99859 0.99862
##                           PC544   PC545   PC546   PC547   PC548   PC549   PC550
## Standard deviation     0.13948 0.13860 0.13707 0.13684 0.13664 0.13595 0.13554
## Proportion of Variance 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003
## Cumulative Proportion  0.99865 0.99868 0.99871 0.99874 0.99877 0.99880 0.99883
##                           PC551   PC552   PC553   PC554   PC555   PC556   PC557
## Standard deviation     0.13486 0.13419 0.13315 0.13277 0.13223 0.13078 0.13069
## Proportion of Variance 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003
## Cumulative Proportion  0.99886 0.99889 0.99892 0.99895 0.99898 0.99900 0.99903
##                           PC558   PC559   PC560   PC561   PC562   PC563   PC564
## Standard deviation     0.12992 0.12913 0.12850 0.12834 0.12820 0.12727 0.12644
## Proportion of Variance 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003 0.00003
## Cumulative Proportion  0.99906 0.99909 0.99911 0.99914 0.99917 0.99919 0.99922
##                           PC565   PC566   PC567   PC568   PC569   PC570   PC571
## Standard deviation     0.12631 0.12504 0.12484 0.12434 0.12388 0.12263 0.12166
## Proportion of Variance 0.00003 0.00003 0.00003 0.00003 0.00002 0.00002 0.00002
## Cumulative Proportion  0.99924 0.99927 0.99929 0.99932 0.99934 0.99937 0.99939
##                           PC572   PC573   PC574   PC575   PC576   PC577   PC578
## Standard deviation     0.12101 0.12052 0.12010 0.11923 0.11855 0.11778 0.11643
## Proportion of Variance 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002
## Cumulative Proportion  0.99942 0.99944 0.99946 0.99949 0.99951 0.99953 0.99955
##                           PC579   PC580   PC581   PC582   PC583   PC584   PC585
## Standard deviation     0.11609 0.11583 0.11523 0.11288 0.11269 0.11238 0.11166
## Proportion of Variance 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002
## Cumulative Proportion  0.99958 0.99960 0.99962 0.99964 0.99966 0.99968 0.99970
##                           PC586   PC587   PC588   PC589   PC590   PC591   PC592
## Standard deviation     0.11021 0.10968 0.10893 0.10845 0.10656 0.10454 0.10403
## Proportion of Variance 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002 0.00002
## Cumulative Proportion  0.99972 0.99974 0.99976 0.99978 0.99980 0.99981 0.99983
##                           PC593   PC594   PC595   PC596   PC597   PC598   PC599
## Standard deviation     0.10174 0.09932 0.09744 0.09704 0.09455 0.09427 0.09212
## Proportion of Variance 0.00002 0.00002 0.00002 0.00002 0.00001 0.00001 0.00001
## Cumulative Proportion  0.99985 0.99986 0.99988 0.99990 0.99991 0.99992 0.99994
```

```
##                           PC600   PC601   PC602   PC603   PC604  PC605  PC606
## Standard deviation     9e-02 0.08606 0.07748 0.06529 0.05671 0.0486 0.0391
## Proportion of Variance 1e-05 0.00001 0.00001 0.00001 0.00001 0.0000 0.0000
## Cumulative Proportion  1e+00 0.99996 0.99997 0.99998 0.99998 1.0000 1.0000
##                            PC607  PC608  PC609  PC610  PC611  PC612 PC613  PC614
## Standard deviation     0.0333 0.0298 0.0249 0.0244 0.0219 0.0216  0.02 0.0189
## Proportion of Variance 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000  0.00 0.0000
## Cumulative Proportion  1.0000 1.0000 1.0000 1.0000 1.0000 1.0000  1.00 1.0000
##                            PC615  PC616 PC617
## Standard deviation     0.0159 0.0141 0.013
## Proportion of Variance 0.0000 0.0000 0.000
## Cumulative Proportion  1.0000 1.0000 1.000
```

Análise

a) Qual o menor número de componentes, tal que a variância acumulada seja pelo menos `80%` do total?

Resposta: Para que a variância acumulada seja de pelo menos 80%, são necessários 48 componentes.

b) Qual o menor número de componentes, tal que a variância acumulada seja pelo menos `90%` do total?

Resposta: Para que a variância acumulada seja de pelo menos 90%, são necessários 112 componentes.

c) Qual o menor número de componentes, tal que a variância acumulada seja pelo menos `95%` do total?

Resposta: Para que a variância acumulada seja de pelo menos 95%, são necessários 200 componentes.

d) Qual o menor número de componentes, tal que a variância acumulada seja pelo menos `99%` do total?

Resposta: Para que a variância acumulada seja de pelo menos 99%, são necessários 400 componentes.

e) Quais as principais diferenças entre a aplicação do PCA nesse conjunto de dados com e sem normalização?

Resposta: A principal diferença é a quantidade de componentes principais necessários para se obter o mesmo nível de variância dos dados. No conjunto de dados sem a normalização, o menor número de componentes necessários para que a variância fosse de 99% do total era de 382 componentes, 18 a menos que o conjunto normalizado (que precisa de 400 componentes).

Em ambas análises, o valor de variância acumulada que o usuário reter decidirá quantos componentes ele usará. Independente do valor escolhido, porém, vimos que o PCA sem normalização sempre "demanda" menos componentes principais.

f) Qual opção parece ser mais adequada para esse conjunto de dados? Justifique sua resposta.

Resposta: O primeiro conjunto, sem a normalização, é o mais adequado, pois com ele economizamos o número de componentes retidas, ou seja, utilizamos menos features sem compometer a representação do conjunto de dados original.

Como o modelo sem normalização utiliza menos componentes, ele também é capaz de reduzir o custo de armazenamento e facilitar a implementação, se comparado com o modelo normalizado.

Atividade 3 – Visualização a partir da Redução (*3,0 pts*)

Nesta atividade, vamos aplicar diferentes métodos de redução de dimensionalidade e comparar as visualizações dos dados obtidos considerando apenas duas dimensões. Lembre de fixar uma semente antes de executar o T-SNE.

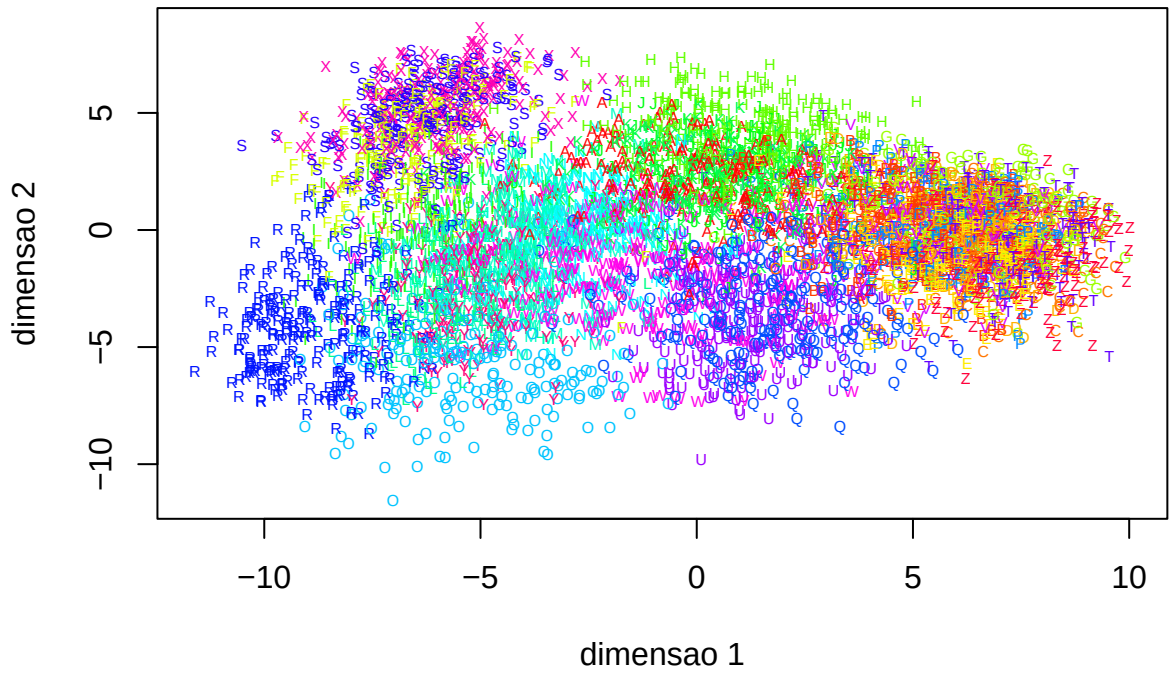
a) Aplique a redução de dimensionalidade com a técnica PCA e gere um gráfico de dispersão dos dados. Use a coluna `618` para classificar as amostras e definir uma coloração.

```
# Aplicando redução de dimensionalidade com a técnica PCA
speech.pca1 <- prcomp (speech [ ,1:617])

# Gerando o gráfico de dispersão
colors <- rainbow(length(unique(speech$LETRA)))
names(colors) <- unique(speech$LETRA)

# plot(speech.pca1$layout , t='n', main="", xlab="dimensao 1", ylab="dimensao 2")
# text(speech.pca1$layout, labels=speech$LETRA, col=colors[speech$LETRA], cex=0.5)

plot(speech.pca1$x[,1:2], t='n', main="", xlab="dimensao 1", ylab="dimensao 2")
text(speech.pca1$x[,1:2], labels=speech$LETRA, col=colors[speech$LETRA], cex=0.5)
```

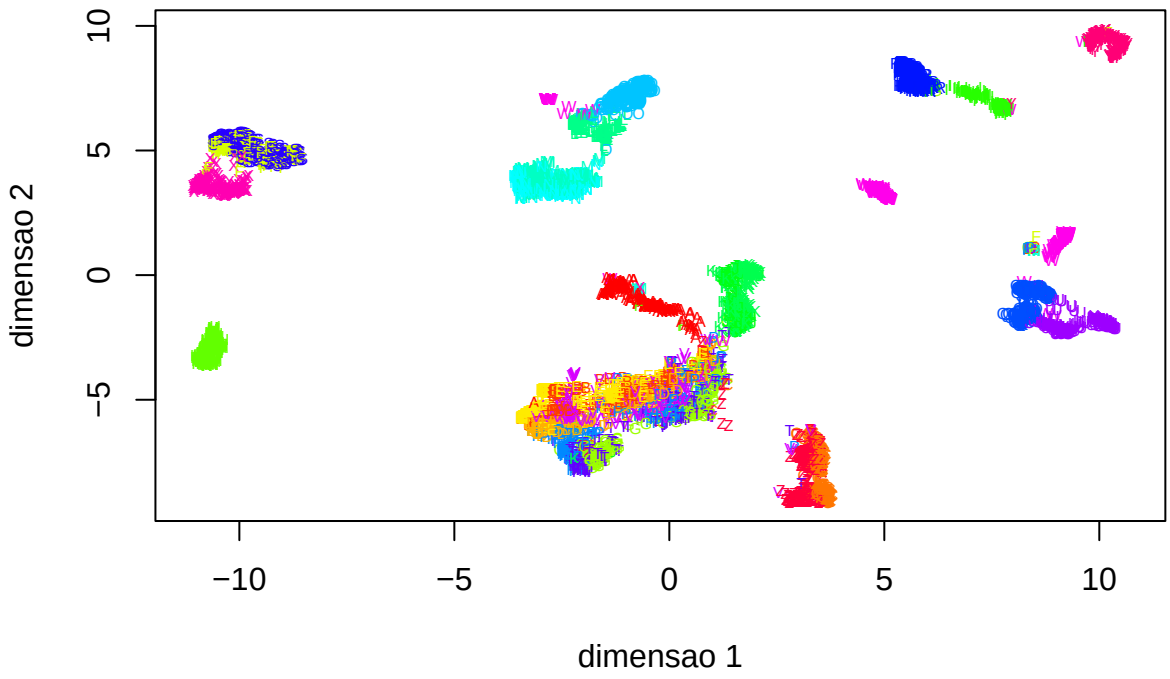


b) Aplique a redução de dimensionalidade com a técnica UMAP e gere um gráfico de dispersão dos dados. Use a coluna `618` para classificar as amostras e definir uma coloração.

```r
# Aplicando redução de dimensionalidade com a técnica UMAP
set.seed(1)
speech.umap <- umap(as.matrix(speech[,1:617]), verbose=TRUE)

# Gerando o gráfico de dispersão
colors <- rainbow(length(unique(speech$LETRA)))
names(colors) <- unique(speech$LETRA)

plot(speech.umap$layout , t='n', main="", xlab="dimensao 1", ylab="dimensao 2")
text(speech.umap$layout, labels=speech$LETRA, col=colors[speech$LETRA], cex=0.5)
```



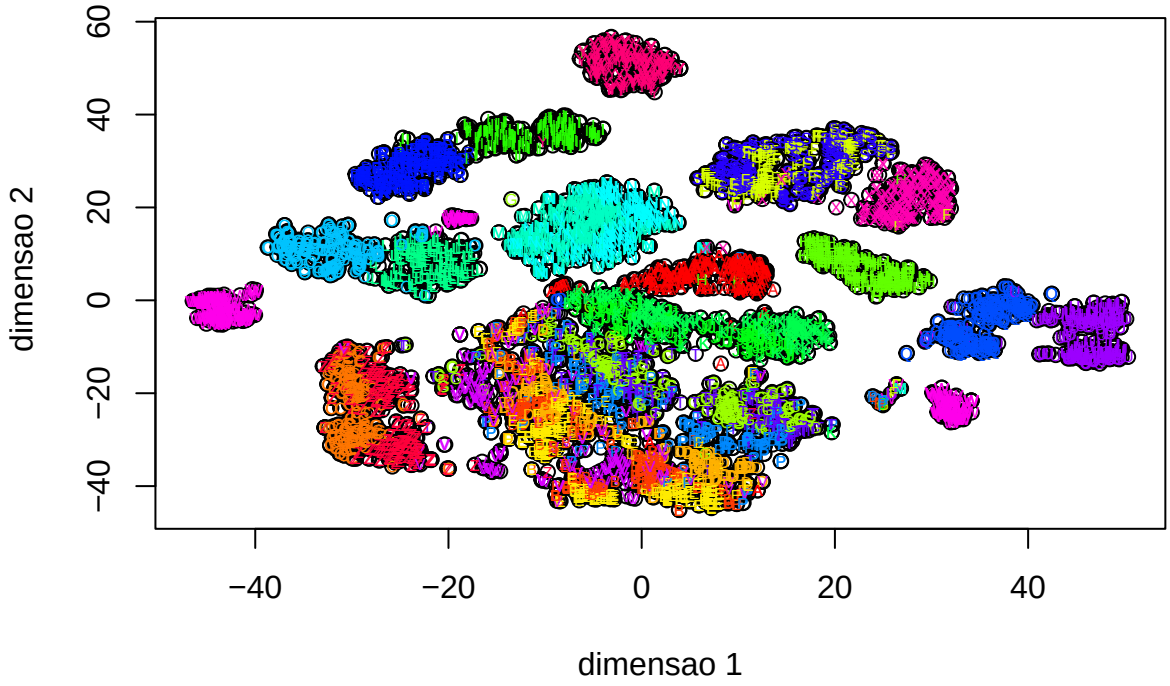
c) Aplique a redução de dimensionalidade com a técnica T-SNE e gere um gráfico de dispersão dos dados. Use a coluna `618` para classificar as amostras e definir uma coloração.

```r
# Aplicando redução de dimensionalidade com a técnica T-SNE
speech_unico <- unique(speech) # Remove dados duplicados
set.seed(1) # semente fixa para reprodutibilidade
tsne <- Rtsne(as.matrix(speech_unico[,1:617]), perplexity = 30, dims=2)

# perplexidade: 5 a 50
# Gerando o gráfico de dispersão

colors <- rainbow(length(unique(speech$LETRA))) # atribui uma cor para cada letra
names(colors) <- unique(speech$LETRA) # associa o nome da com a letra
```

```
plot(tsne$Y , xlab = " dimensao 1 " , ylab = " dimensao 2")
text(tsne$Y , labels = speech$LETRA, col=colors[speech$LETRA], cex =0.5)
```



Análise

d) Qual técnica você acredita que apresentou a melhor projeção? Justifique.

Resposta: O que apresentou o pior resultado foi o PCA, o que pode ser explicado por uma alta complexidade da relação entre os atributos desses conjunto de dados. Como o PCA realiza apenas uma comparação linear entre as features, poderíamos supor que a relação entre os atributos não é apenas linear.

Analisando os demais gráficos de dispersão, verificamos que tanto o UMAP quanto o t-SNE foram capazes de agrupar os dados que pertencem às mesmas classes, com alguma precisão. Porém, o UMAP se saiu melhor no quesito separação entre as classes, pois como pode ser visto no gráfico, a distância entre classes diferentes é maior. Além disso, por ser mais rápido, o UMAP também apresenta uma vantagem adicional, em relação ao t-SNE.

Sendo assim, consideremos que o método de redução de dimensionalidade que apresentou os melhores resultados foi o UMAP.