

# Principal Component Analysis (PCA) Coding Assignment

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library(mice)
```

```
## Warning: package 'mice' was built under R version 4.3.3
```

```
## Warning in check_dep_version(): ABI version mismatch:
```

```
## lme4 was built with Matrix ABI version 1
```

```
## Current Matrix ABI version is 0
```

```
## Please re-install lme4 from source or restore original 'Matrix' package
```

```
##
```

```
## Attaching package: 'mice'
```

```
## The following object is masked from 'package:stats':
```

```
##
```

```
## filter
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## cbind, rbind
```

```
## Load ICU dataset
```

```
file = 'Data/ICU_filtered.csv'
```

```
dfICU = read.csv(file, header = TRUE, stringsAsFactors = TRUE)
```

```
str(dfICU)
```

```
## 'data.frame': 7886 obs. of 36 variables:
```

```
## $ recordid : int 132539 132540 132541 132543 132545 132547 132548 132551 132554 132555 ...
```

```
## $ Length_of_stay : int 5 8 19 9 4 6 9 6 17 8 ...
```

```
## $ In.hospital_death: int 0 0 0 0 0 0 0 1 0 0 ...
```

```
## $ Age : int 54 76 44 68 88 64 68 78 64 74 ...
```

```
## $ Gender : int 0 1 0 1 0 1 0 0 0 1 ...
```

```
## $ Height : num NA 175 NA 180 NA ...
```

```
## $ CCU : int 0 0 0 0 0 1 0 0 0 0 ...
```

```
## $ CSRU : int 0 1 0 0 0 0 0 0 0 1 ...
```

```
## $ SICU : int 1 0 0 0 0 0 0 0 0 0 ...
```

```
## $ DiasABP_first : int NA 67 81 NA NA 79 98 50 NA 51 ...
## $ GCS_first : int 15 3 7 15 15 7 15 15 10 ...
## $ Glucose_first : int 205 105 141 129 113 264 94 132 113 106 ...
## $ HR_first : num 73 88 100 79 93 78 73 111 127 67 ...
## $ MAP_first : num NA 79 107 NA NA 101 136 83 NA 67 ...
## $ NIDiasABP_first : int 65 38 84 63 41 89 88 51 71 57 ...
## $ NIMAP_first : num 92.3 49.3 100.3 86.7 75.3 ...
## $ NISysABP_first : int 147 72 133 134 144 129 187 100 138 134 ...
## $ Temp_first : num 35.1 35.2 37.8 36.3 37.8 35.8 36.3 38 37.3 34.8 ...
## $ BUN_first : int 13 16 8 23 45 15 32 81 21 19 ...
## $ Creatinine_first : num 0.8 0.8 0.4 0.9 1 1.4 3.4 0.9 0.7 1.1 ...
## $ FiO2_first : num NA 1 1 NA NA 0.5 NA 1 NA 1 ...
## $ HCO3_first : int 26 21 24 28 18 19 25 18 21 23 ...
## $ HCT_first : num 33.7 24.7 28.5 41.3 22.6 41.6 31.9 32.6 28.3 31.5 ...
## $ K_first : num 4.4 4.3 3.3 4 6 5.1 3.7 4.2 3.9 4.6 ...
## $ Lactate_first : num NA NA 1.3 NA NA NA NA 1.4 NA NA ...
## $ Mg_first : num 1.5 3.1 1.9 2.1 1.5 1.7 1.9 2.2 1.6 1.8 ...
## $ Na_first : int 137 139 137 140 140 141 140 141 139 141 ...
## $ PaCO2_first : int NA 34 37 NA NA 45 NA 30 NA 42 ...
## $ PaO2_first : int NA 344 65 NA NA 78 NA 128 NA 486 ...
## $ Platelets_first : int 221 164 72 391 109 276 325 91 696 141 ...
## $ SysABP_first : int NA 105 148 NA NA 150 205 103 NA 98 ...
## $ WBC_first : num 11.2 7.4 4.2 11.5 3.8 24 6.2 16.1 15.2 9 ...
## $ Weight : num NA 80.6 56.7 84.6 NA NA 87 48.4 60.7 66.1 ...
## $ pH_first : num NA 7.45 7.51 NA NA 7.29 NA 7.4 NA 7.39 ...
## $ MechVent : int 0 1 1 0 0 1 0 1 0 1 ...
## $ UrineOutputSum : int NA 5 14 NA NA 12 NA 12 NA 4 ...
```

```
## Print first 5 samples of data frame
```

```
head(dfICU, n = 5)
```

```
## recordid Length_of_stay In.hospital_death Age Gender Height CCU CSRU SICU
## 1 132539 5 0 54 0 NA 0 0 1
## 2 132540 8 0 76 1 175.3 0 1 0
## 3 132541 19 0 44 0 NA 0 0 0
## 4 132543 9 0 68 1 180.3 0 0 0
## 5 132545 4 0 88 0 NA 0 0 0
## DiasABP_first GCS_first Glucose_first HR_first MAP_first NIDiasABP_first
## 1 NA 15 205 73 NA 65
## 2 67 3 105 88 79 38
## 3 81 7 141 100 107 84
## 4 NA 15 129 79 NA 63
## 5 NA 15 113 93 NA 41
## NIMAP_first NISysABP_first Temp_first BUN_first Creatinine_first FiO2_first
## 1 92.33 147 35.1 13 0.8 NA
## 2 49.33 72 35.2 16 0.8 1
## 3 100.30 133 37.8 8 0.4 1
## 4 86.67 134 36.3 23 0.9 NA
## 5 75.33 144 37.8 45 1.0 NA
## HCO3_first HCT_first K_first Lactate_first Mg_first Na_first PaCO2_first
## 1 26 33.7 4.4 NA 1.5 137 NA
## 2 21 24.7 4.3 NA 3.1 139 34
## 3 24 28.5 3.3 1.3 1.9 137 37
## 4 28 41.3 4.0 NA 2.1 140 NA
## 5 18 22.6 6.0 NA 1.5 140 NA
```

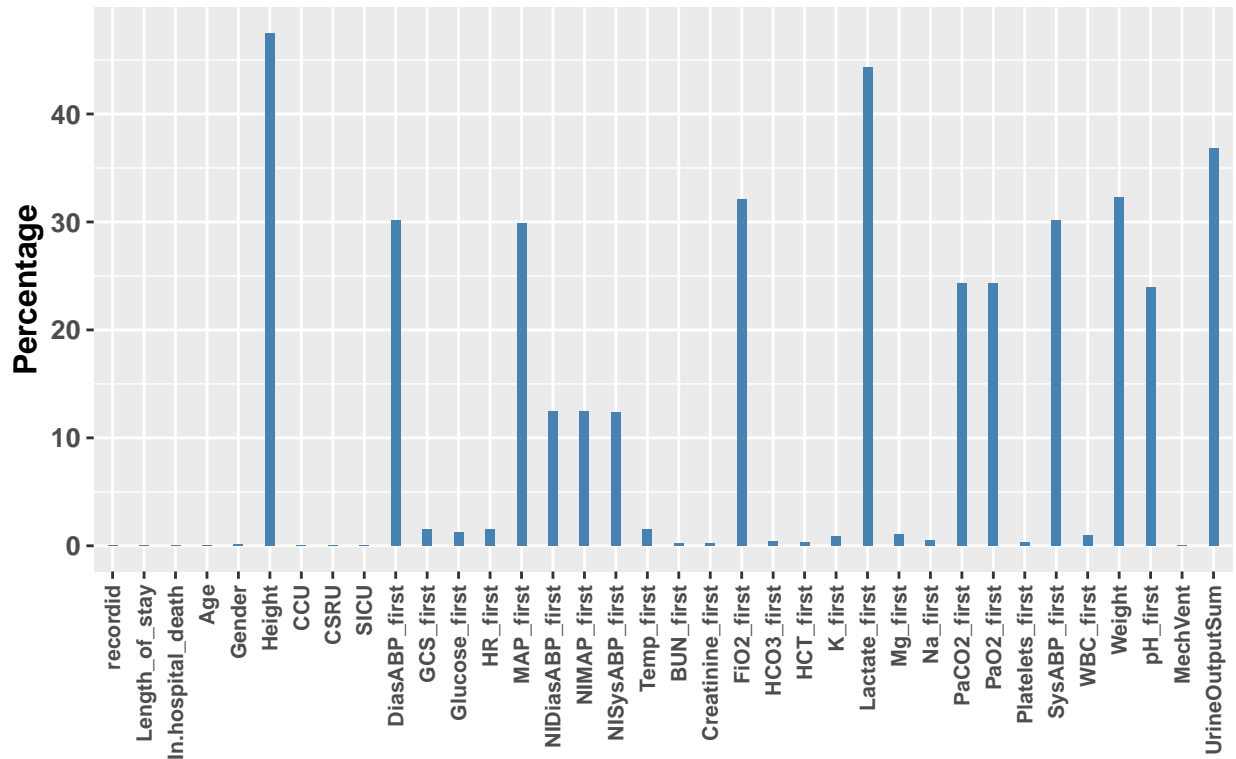
```
## PaO2_first Platelets_first SysABP_first WBC_first Weight pH_first MechVent
## 1      NA      221      NA      11.2      NA      NA      0
## 2      344      164      105      7.4      80.6      7.45      1
## 3      65      72      148      4.2      56.7      7.51      1
## 4      NA      391      NA      11.5      84.6      NA      0
## 5      NA      109      NA      3.8      NA      NA      0
## UrineOutputSum
## 1      NA
## 2      5
## 3      14
## 4      NA
## 5      NA
```

```
## Plot fraction of missing values (NAs) in each column of the data frame
```

```
pMissDF = setNames(stack(sapply(dfICU, function(x){(sum(is.na(x))/length(x))*100}))[2:1], c('Feature', 'Value'))
p = ggplot(data = pMissDF, aes(x = Feature, y = Value)) +
  geom_bar(stat = 'identity', fill = 'steelblue', width = 0.3) +
  theme(text = element_text(size = 14, face = 'bold'),
        axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5)) +
  xlab('') + ylab('Percentage') +
  ggtitle('Percentage of NAs across all features') +
  theme(plot.title = element_text(size = 12, hjust = 0.5),
        axis.text = element_text(size = 8),
        axis.text.x = element_text(size = 8),
        axis.text.y = element_text(size = 10),
        axis.title = element_text(size = 12, face = "bold"))
```

p

Percentage of NAs across all features



```
## Drop columns with more than 20% missing values
```

```
dfICU = dfICU %>% select(-c(pMissDF[pMissDF['Value'] > 20, 'Feature']))
```

```
## Collate 4 different ICU types (CCU, CSRU, SICU, CCU) into one column
```

```
## called 'ICU', remove separate ICU columns and the following columns:
```

```
## (1) 'recordid' (2) In.hospital_death (3) Length_of_stay
```

```
dfICU[dfICU['CCU'] == 1, 'ICU'] = 1
```

```
dfICU[dfICU['CSRU'] == 1, 'ICU'] = 2
```

```
dfICU[dfICU['SICU'] == 1, 'ICU'] = 3
```

```
dfICU[(dfICU['CCU'] == 0) & (dfICU['CSRU'] == 0) & (dfICU['SICU'] == 0), 'ICU'] = 4
```

```
dfICU = dfICU %>% select(-c(CCU, CSRU, SICU, recordid, In.hospital_death, Length_of_stay))
```

```
## Create list of continuous and categorical features
```

```
features = colnames(dfICU)
```

```
categorical_features = c('Gender', 'GCS_first', 'MechVent', 'ICU')
```

```
continuous_features = features[c(!(colnames(dfICU) %in% categorical_features))]
```

```
# Impute missing values using the MICE package
```

```
dfICU = complete(mice(dfICU, m = 40, maxit = 10, pred = quickpred(dfICU, minpuc = 0.25), seed = 500))
```

```
##
```

```
## iter imp variable
```

```
## 1 1 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first T
```

```
## 1 2 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first T
```

```
## 1 3 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first T
```

```
## 1 4 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first T
```

```
## 1 5 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first T
```

[illegible]

[illegible]



[illegible]



[illegible]

[illegible]

[illegible]

```
## 10 24 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 25 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 26 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 27 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 28 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 29 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 30 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 31 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 32 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 33 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 34 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 35 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 36 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 37 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 38 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 39 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 10 40 Gender GCS_first Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
```

```
## Select only continuous features
```

```
dfICU_continuous = dfICU %>% select(continuous_features)
```

```
## Warning: Using an external vector in selections was deprecated in tidysselect 1.1.0.
```

```
## i Please use `all_of()` or `any_of()` instead.
```

```
## # Was:
```

```
## data %>% select(continuous_features)
```

```
##
```

```
## # Now:
```

```
## data %>% select(all_of(continuous_features))
```

```
##
```

```
## See <https://tidysselect.r-lib.org/reference/faq-external-vector.html>.
```

```
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
```

```
## generated.
```

```
head(dfICU_continuous)
```

```
## Age Glucose_first HR_first NIDiasABP_first NIMAP_first NISysABP_first
## 1 54 205 73 65 92.33 147
## 2 76 105 88 38 49.33 72
## 3 44 141 100 84 100.30 133
## 4 68 129 79 63 86.67 134
## 5 88 113 93 41 75.33 144
## 6 64 264 78 89 102.30 129
## Temp_first BUN_first Creatinine_first HCO3_first HCT_first K_first Mg_first
## 1 35.1 13 0.8 26 33.7 4.4 1.5
## 2 35.2 16 0.8 21 24.7 4.3 3.1
## 3 37.8 8 0.4 24 28.5 3.3 1.9
## 4 36.3 23 0.9 28 41.3 4.0 2.1
## 5 37.8 45 1.0 18 22.6 6.0 1.5
## 6 35.8 15 1.4 19 41.6 5.1 1.7
## Na_first Platelets_first WBC_first
## 1 137 221 11.2
## 2 139 164 7.4
## 3 137 72 4.2
## 4 140 391 11.5
## 5 140 109 3.8
```

```
## 6      141      276      24.0

## Scale the data containing the continuous features and perform PCA
# Scale the data containing the continuous features
X = scale(dfICU_continuous) # note that the output of the scale function is a matrix

# Covariance of scaled data matrix
S = cov(X)

# Calculate eigenvectors and eigenvalues of covariance matrix
e = eigen(S)
V = e$vectors

lambda = e$values
colnames(dfICU_continuous)

## [1] "Age"          "Glucose_first"  "HR_first"      "NIDiasABP_first"
## [5] "NIMAP_first"    "NISysABP_first" "Temp_first"    "BUN_first"
## [9] "Creatinine_first" "HCO3_first"    "HCT_first"     "K_first"
## [13] "Mg_first"       "Na_first"      "Platelets_first" "WBC_first"

print(V)

##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] -0.157731074 -0.114530569 -0.388013220 -0.20059326  0.005915649
## [2,]  0.005635612 -0.172437854  0.159764114  0.00723908  0.473930391
## [3,]  0.084657475 -0.001300709  0.540994038  0.22311398 -0.233916420
## [4,]  0.494945453 -0.202688811 -0.025706623  0.10622443  0.033839267
## [5,]  0.528992203 -0.230211214 -0.118111402  0.08685209  0.042332360
## [6,]  0.449431830 -0.234527148 -0.201498710  0.02985685  0.029476245
## [7,]  0.162897380  0.047851054  0.317785388  0.02704535 -0.486713293
## [8,] -0.213571929 -0.542042198 -0.007167586  0.06753904 -0.229836010
## [9,] -0.154799799 -0.548047159  0.061673637  0.12320867 -0.191547259
## [10,]  0.157524382  0.203367765 -0.296760302 -0.33991378 -0.361045821
## [11,]  0.251551548 -0.031364913  0.080536744 -0.35760611  0.125964901
## [12,] -0.164371612 -0.317541376  0.027314011 -0.13978209  0.095937990
## [13,] -0.123474974 -0.215189739 -0.223004906 -0.29063844 -0.251198610
## [14,]  0.056542301  0.107109994 -0.096687245  0.06388189 -0.413015376
## [15,]  0.105424742 -0.077101745  0.283895858 -0.57406133  0.001832752
## [16,] -0.018719481 -0.068626251  0.369074279 -0.43437064  0.036279605
##           [,6]      [,7]      [,8]      [,9]      [,10]
## [1,]  0.377402392  0.410246656  0.01761530 -0.188522420  0.248295572
## [2,]  0.289316085  0.323140750  0.60151015  0.174142588 -0.055318183
## [3,] -0.007860540  0.009145545  0.34862544 -0.251823707 -0.025167970
## [4,] -0.062992708 -0.131462094  0.03098183 -0.114210745  0.002672089
## [5,] -0.009254811 -0.017287385 -0.03180196 -0.144260166  0.009327204
## [6,]  0.069242177  0.118243963 -0.07179352 -0.161765928 -0.007308783
## [7,] -0.105951164  0.484795171  0.03350668  0.196329234  0.286278586
## [8,]  0.172658613  0.071582144 -0.08756448  0.110829640 -0.029966589
## [9,]  0.009113297 -0.010802378 -0.23540846  0.242792450 -0.043261167
## [10,] -0.188664517  0.223446884  0.12590612  0.104476321 -0.004746577
## [11,]  0.038804711 -0.264961922  0.06683977  0.654623272  0.332417113
## [12,] -0.474222443 -0.181943298  0.14587459  0.005404975  0.089831214
## [13,] -0.216371178 -0.205871146  0.53924956 -0.255184263 -0.017027829
```

```
## [14,] 0.586538151 -0.448106907 0.20327227 0.125846863 -0.170267220
## [15,] 0.057492755 0.148928124 -0.17397606 -0.013245146 -0.690507459
## [16,] 0.260188397 -0.191742703 -0.20254861 -0.424629958 0.474456040
##      [,11]      [,12]      [,13]      [,14]      [,15]
## [1,] -0.429396734 -0.350138409 0.07256257 -0.12806968 -0.1851863129
## [2,] 0.150603663 0.326802497 -0.02950747 -0.06201052 -0.0501574721
## [3,] -0.288790517 -0.236242768 0.46445381 0.22784212 -0.0532992625
## [4,] -0.001714367 -0.073589575 0.10643445 -0.63155485 -0.0295162938
## [5,] -0.026229643 0.016036703 -0.02326936 -0.08592667 0.0233507660
## [6,] -0.023176398 0.097127449 -0.16892331 0.68809292 0.0337060161
## [7,] 0.031332168 -0.012602522 -0.49865482 -0.12160924 0.0061066931
## [8,] 0.076861992 0.007557207 0.18018624 -0.04749545 0.7117487624
## [9,] 0.159458216 0.104443485 0.16145709 0.04022027 -0.6627619853
## [10,] 0.018898119 0.473560911 0.51258509 -0.01363686 -0.0006351825
## [11,] -0.054529513 -0.349439687 0.14522622 0.14786735 0.0472908981
## [12,] -0.641042951 0.305357375 -0.22316178 -0.02026051 0.0286051143
## [13,] 0.405733293 -0.301567663 -0.18107450 0.04554407 -0.0696160600
## [14,] -0.242469445 0.227870385 -0.23305189 -0.04026719 -0.0607377346
## [15,] -0.101306038 -0.132837121 -0.09645321 -0.05139781 -0.0130771782
## [16,] 0.167487388 0.297383642 0.02258348 -0.02575046 -0.0125806713
##      [,16]
## [1,] 0.0007563845
## [2,] -0.0030633727
## [3,] -0.0156716613
## [4,] 0.4991150399
## [5,] -0.7824163101
## [6,] 0.3706481369
## [7,] 0.0082054614
## [8,] 0.0090541665
## [9,] -0.0171902008
## [10,] 0.0023612386
## [11,] -0.0042733809
## [12,] 0.0090888615
## [13,] -0.0217949905
## [14,] -0.0021386655
## [15,] -0.0064094167
## [16,] 0.0011681853
```

```
print(lambda)
```

```
## [1] 2.7106781 1.9763327 1.4941549 1.2145392 1.1661721 1.0726946 0.9719390
## [8] 0.9047355 0.8791409 0.7515047 0.7241747 0.6945458 0.6101434 0.4049811
## [15] 0.2922539 0.1320093
```

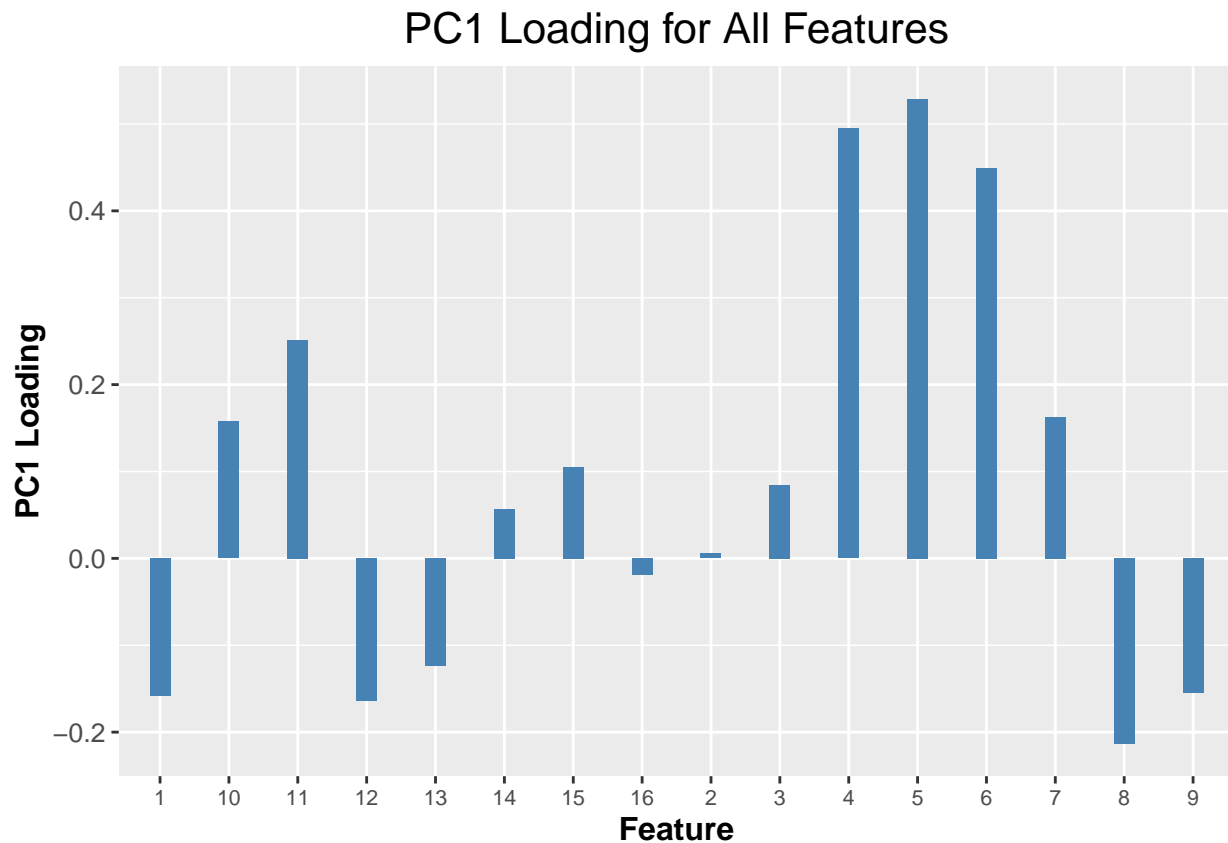
```
## Note that PCA can also be performed using R in-built function prcomp()
prcomp(X)$rotation # extracting the eigenvectors matrix
```

```
##      PC1      PC2      PC3      PC4
## Age      0.157731074 0.114530569 0.388013220 0.20059326
## Glucose_first -0.005635612 0.172437854 -0.159764114 -0.00723908
## HR_first -0.084657475 0.001300709 -0.540994038 -0.22311398
## NIDiasABP_first -0.494945453 0.202688811 0.025706623 -0.10622443
## NIMAP_first -0.528992203 0.230211214 0.118111402 -0.08685209
## NISysABP_first -0.449431830 0.234527148 0.201498710 -0.02985685
## Temp_first -0.162897380 -0.047851054 -0.317785388 -0.02704535
```

## BUN_first	0.213571929	0.542042198	0.007167586	-0.06753904
## Creatinine_first	0.154799799	0.548047159	-0.061673637	-0.12320867
## HCO3_first	-0.157524382	-0.203367765	0.296760302	0.33991378
## HCT_first	-0.251551548	0.031364913	-0.080536744	0.35760611
## K_first	0.164371612	0.317541376	-0.027314011	0.13978209
## Mg_first	0.123474974	0.215189739	0.223004906	0.29063844
## Na_first	-0.056542301	-0.107109994	0.096687245	-0.06388189
## Platelets_first	-0.105424742	0.077101745	-0.283895858	0.57406133
## WBC_first	0.018719481	0.068626251	-0.369074279	0.43437064
##	PC5	PC6	PC7	PC8
## Age	0.005915649	-0.377402392	-0.410246656	0.01761530
## Glucose_first	0.473930391	-0.289316085	-0.323140750	0.60151015
## HR_first	-0.233916420	0.007860540	-0.009145545	0.34862544
## NIDiasABP_first	0.033839267	0.062992708	0.131462094	0.03098183
## NIMAP_first	0.042332360	0.009254811	0.017287385	-0.03180196
## NISysABP_first	0.029476245	-0.069242177	-0.118243963	-0.07179352
## Temp_first	-0.486713293	0.105951164	-0.484795171	0.03350668
## BUN_first	-0.229836010	-0.172658613	-0.071582144	-0.08756448
## Creatinine_first	-0.191547259	-0.009113297	0.010802378	-0.23540846
## HCO3_first	-0.361045821	0.188664517	-0.223446884	0.12590612
## HCT_first	0.125964901	-0.038804711	0.264961922	0.06683977
## K_first	0.095937990	0.474222443	0.181943298	0.14587459
## Mg_first	-0.251198610	0.216371178	0.205871146	0.53924956
## Na_first	-0.413015376	-0.586538151	0.448106907	0.20327227
## Platelets_first	0.001832752	-0.057492755	-0.148928124	-0.17397606
## WBC_first	0.036279605	-0.260188397	0.191742703	-0.20254861
##	PC9	PC10	PC11	PC12
## Age	-0.188522420	0.248295572	0.429396734	0.350138409
## Glucose_first	0.174142588	-0.055318183	-0.150603663	-0.326802497
## HR_first	-0.251823707	-0.025167970	0.288790517	0.236242768
## NIDiasABP_first	-0.114210745	0.002672089	0.001714367	0.073589575
## NIMAP_first	-0.144260166	0.009327204	0.026229643	-0.016036703
## NISysABP_first	-0.161765928	-0.007308783	0.023176398	-0.097127449
## Temp_first	0.196329234	0.286278586	-0.031332168	0.012602522
## BUN_first	0.110829640	-0.029966589	-0.076861992	-0.007557207
## Creatinine_first	0.242792450	-0.043261167	-0.159458216	-0.104443485
## HCO3_first	0.104476321	-0.004746577	-0.018898119	-0.473560911
## HCT_first	0.654623272	0.332417113	0.054529513	0.349439687
## K_first	0.005404975	0.089831214	0.641042951	-0.305357375
## Mg_first	-0.255184263	-0.017027829	-0.405733293	0.301567663
## Na_first	0.125846863	-0.170267220	0.242469445	-0.227870385
## Platelets_first	-0.013245146	-0.690507459	0.101306038	0.132837121
## WBC_first	-0.424629958	0.474456040	-0.167487388	-0.297383642
##	PC13	PC14	PC15	PC16
## Age	0.07256257	-0.12806968	-0.1851863129	-0.0007563845
## Glucose_first	-0.02950747	-0.06201052	-0.0501574721	0.0030633727
## HR_first	0.46445381	0.22784212	-0.0532992625	0.0156716613
## NIDiasABP_first	0.10643445	-0.63155485	-0.0295162938	-0.4991150399
## NIMAP_first	-0.02326936	-0.08592667	0.0233507660	0.7824163101
## NISysABP_first	-0.16892331	0.68809292	0.0337060161	-0.3706481369
## Temp_first	-0.49865482	-0.12160924	0.0061066931	-0.0082054614
## BUN_first	0.18018624	-0.04749545	0.7117487624	-0.0090541665
## Creatinine_first	0.16145709	0.04022027	-0.6627619853	0.0171902008
## HCO3_first	0.51258509	-0.01363686	-0.0006351825	-0.0023612386

```
## HCT_first      0.14522622  0.14786735  0.0472908981  0.0042733809
## K_first       -0.22316178 -0.02026051  0.0286051143 -0.0090888615
## Mg_first      -0.18107450  0.04554407 -0.0696160600  0.0217949905
## Na_first      -0.23305189 -0.04026719 -0.0607377346  0.0021386655
## Platelets_first -0.09645321 -0.05139781 -0.0130771782  0.0064094167
## WBC_first      0.02258348 -0.02575046 -0.0125806713 -0.0011681853
```

```
## Plot PC1 loadings
dfPCA = data.frame(V)
colnames(dfPCA) = paste0('PC', c(1:ncol(V)), 'L')
ggplot(data = dfPCA) +
  geom_bar(aes(x = rownames(dfPCA), y = PC1L), stat = 'identity', width = 0.3, fill = 'steelblue') +
  labs(x = 'Feature', y = 'PC1 Loading') +
  ggtitle("PC1 Loading for All Features") +
  theme(plot.title = element_text(size = 16, hjust = 0.5),
        axis.text = element_text(size = 12),
        axis.text.x = element_text(size = 8),
        axis.text.y = element_text(size = 10),
        axis.title = element_text(size = 12, face = "bold"))
```



```
## Calculate PC scores of all samples
PC_scores = lambda
print(lambda)
```

```
## [1] 2.7106781 1.9763327 1.4941549 1.2145392 1.1661721 1.0726946 0.9719390
## [8] 0.9047355 0.8791409 0.7515047 0.7241747 0.6945458 0.6101434 0.4049811
## [15] 0.2922539 0.1320093
```



```

## How much variance is explained by each principal direction
## or each principal component?
print(cumsum(lambda))

## [1] 2.710678 4.687011 6.181166 7.395705 8.561877 9.634572 10.606511
## [8] 11.511246 12.390387 13.141892 13.866066 14.560612 15.170756 15.575737
## [15] 15.867991 16.000000

## Percentage of total variance explained by each principal component
percentage_explained_var = (cumsum(lambda) / sum(lambda)) * 100
print(percentage_explained_var)

## [1] 16.94174 29.29382 38.63229 46.22316 53.51173 60.21607 66.29069
## [8] 71.94529 77.43992 82.13682 86.66292 91.00383 94.81722 97.34836
## [15] 99.17494 100.00000

## How many minimum principal components are needed to explain more than
## 80% of the variance in the data?
n_PC = which(cumsum(percentage_explained_var) >= 80)[1]
print(n_PC)

## [1] 3

```