

# Principal Component Analysis - 19BCE1460

```
library(readr)

## Warning: package 'readr' was built under R version 4.1.3

library(psych)
library(factoextra)

## Loading required package: ggplot2

##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
## 
##     %+%, alpha

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

require("datasets")
penguinsdata <- read.csv("S:/WIN SEM 21-22/Data Visualization/Lab/penguins_lter.csv")
penguinsdata <- na.omit(penguinsdata)
str(penguinsdata)

## 'data.frame': 330 obs. of 17 variables:
##   $ studyName      : chr  "PAL0708" "PAL0708" "PAL0708" "PAL0708" ...
##   $ Sample.Number  : int  2 3 5 6 7 8 10 11 15 17 ...
##   $ Species        : chr  "Adelie Penguin (Pygoscelis adeliae)" "Adelie Penguin (Pygoscelis adeliae"
##   $ Region         : chr  "Anvers" "Anvers" "Anvers" "Anvers" ...
##   $ Island          : chr  "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...
##   $ Stage           : chr  "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" ...
##   $ Individual.ID  : chr  "N1A2" "N2A1" "N3A1" "N3A2" ...
##   $ Clutch.Completion: chr  "Yes" "Yes" "Yes" "Yes" ...
##   $ Date.Egg        : chr  "11-11-2007" "11/16/07" "11/16/07" "11/16/07" ...
##   $ CulmenLengthmm : num  39.5 40.3 36.7 39.3 38.9 39.2 42 37.8 34.6 38.7 ...
##   $ CulmenDepthmm  : num  17.4 18 19.3 20.6 17.8 19.6 20.2 17.1 21.1 19 ...
##   $ FlipperLengthmm: int  186 195 193 190 181 195 190 186 198 195 ...
##   $ BodyMassg       : int  3800 3250 3450 3650 3625 4675 4250 3300 4400 3450 ...
##   $ Sex             : chr  "FEMALE" "FEMALE" "FEMALE" "MALE" ...
##   $ Delta.15.N...o.oo.: num  8.95 8.37 8.77 8.66 9.19 ...
##   $ Delta.13.C...o.oo.: num  -24.7 -25.3 -25.3 -25.3 -25.2 ...
##   $ Comments         : chr  "" "" "" "" ...
## - attr(*, "na.action")= 'omit' Named int [1:14] 1 4 9 12 13 14 16 40 42 47 ...
## ..- attr(*, "names")= chr [1:14] "1" "4" "9" "12" ...
```

```
any(is.na.data.frame(penguinsdata))

## [1] FALSE

shapiro.test(penguinsdata$CulmenLengthmm)

##
##  Shapiro-Wilk normality test
##
## data: penguinsdata$CulmenLengthmm
## W = 0.97563, p-value = 2.219e-05

shapiro.test(penguinsdata$CulmenDepthmm)

##
##  Shapiro-Wilk normality test
##
## data: penguinsdata$CulmenDepthmm
## W = 0.97279, p-value = 7.011e-06

shapiro.test(penguinsdata$FlipperLengthmm)

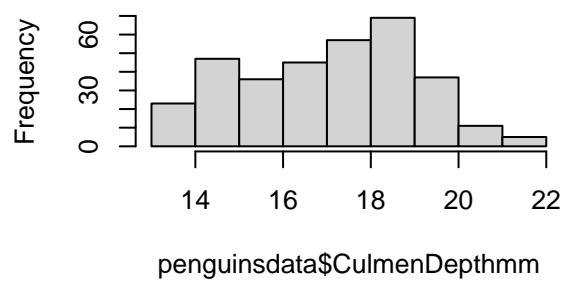
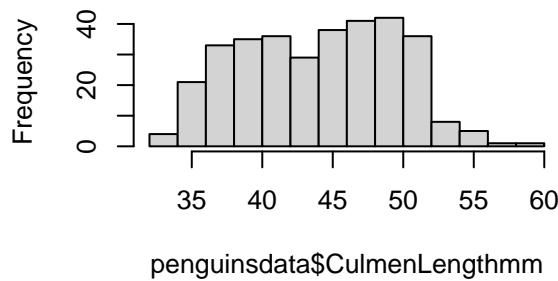
##
##  Shapiro-Wilk normality test
##
## data: penguinsdata$FlipperLengthmm
## W = 0.95228, p-value = 7.221e-09

shapiro.test(penguinsdata$BodyMassg)

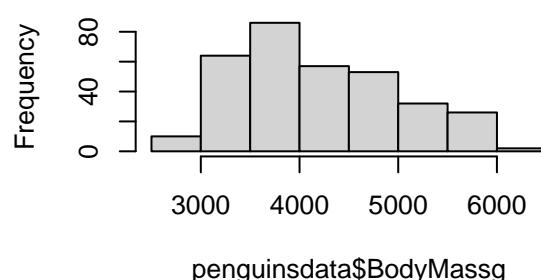
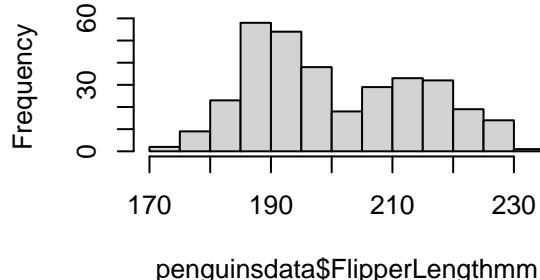
##
##  Shapiro-Wilk normality test
##
## data: penguinsdata$BodyMassg
## W = 0.9609, p-value = 1.004e-07

par(mfrow=(c(2,2)))
hist(penguinsdata$CulmenLengthmm)
hist(penguinsdata$CulmenDepthmm)
hist(penguinsdata$FlipperLengthmm)
hist(penguinsdata$BodyMassg)
```

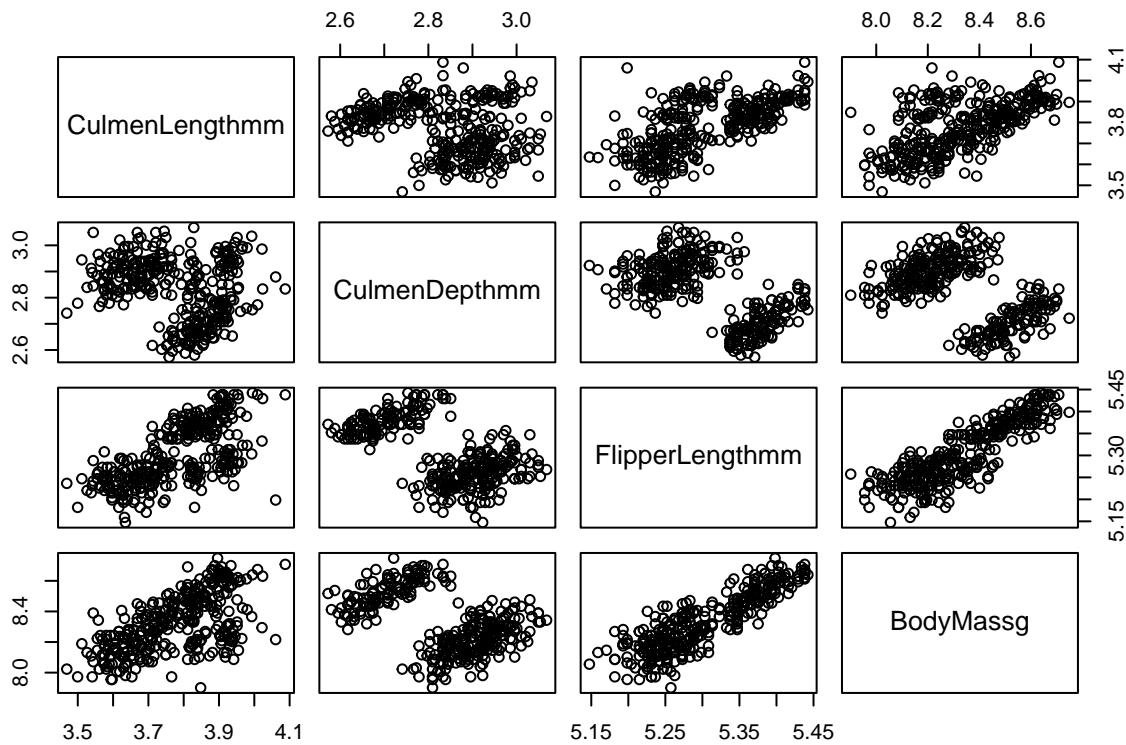
## Histogram of penguinsdata\$CulmenLengthHistogram of penguinsdata\$CulmenDepth



## Histogram of penguinsdata\$FlipperLength Histogram of penguinsdata\$BodyMass



```
penguinsdata_pca<-log(penguinsdata[, 10:13])
penguinsdata_class<-(penguinsdata[, 5])
pairs(penguinsdata_pca)
```



```
print(cortest.bartlett(cor(penguinsdata_pca), nrow(penguinsdata_pca)))
```

```
## $chisq
## [1] 789.8785
##
## $p.value
## [1] 2.36755e-167
##
## $df
## [1] 6
```

```
penguinsdata_comp<-prcomp(penguinsdata_pca,scale = T)
print(penguinsdata_comp)
```

```
## Standard deviations (1, ..., p=4):
## [1] 1.6589422 0.8743559 0.6068615 0.3393108
##
## Rotation (n x k) = (4 x 4):
##          PC1        PC2        PC3        PC4
## CulmenLengthmm  0.4584221  0.585128108 -0.6543224 -0.1390555
## CulmenDepthmm -0.4012054  0.804189688  0.3997313  0.1803554
## FlipperLengthmm  0.5748007 -0.004501677  0.2319137  0.7847292
## BodyMassg       0.5463402  0.104325342  0.5985759 -0.5764855
```

```

biplot(penguinsdata_comp, scale = 0)
#PCA:Variance
plot(penguinsdata_comp,type="lines")
summary(penguinsdata_comp)

## Importance of components:
##          PC1      PC2      PC3      PC4
## Standard deviation   1.659  0.8744  0.60686 0.33931
## Proportion of Variance 0.688  0.1911  0.09207 0.02878
## Cumulative Proportion  0.688  0.8791  0.97122 1.00000

fviz_eig(penguinsdata_comp)

fviz_pca_ind(penguinsdata_comp,
              col.ind = "cos2", # Color by the quality of representation
              gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
              repel = TRUE      # Avoid text overlapping
)

## Warning: ggrepel: 274 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

fviz_pca_var(penguinsdata_comp,
             col.var = "contrib", # Color by contributions to the PC
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE      # Avoid text overlapping
)
fviz_pca_biplot(penguinsdata_comp, repel = TRUE,
                 col.var = "#2E9FDF", # Variables color
                 col.ind = "#696969" # Individuals color
)

## Warning: ggrepel: 272 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

#Predict the PCs for new data
#Last two row of transformed penguins data
tail(penguinsdata_pca,2)

##      CulmenLengthmm CulmenDepthmm FlipperLengthmm BodyMassg
## 343       3.811097      2.694627      5.356586  8.556414
## 344       3.910021      2.778819      5.361292  8.594154

#Loading values of the PCs
print(penguinsdata_comp)

## Standard deviations (1, .., p=4):
## [1] 1.6589422 0.8743559 0.6068615 0.3393108
##
## Rotation (n x k) = (4 x 4):

```

```

##                  PC1          PC2          PC3          PC4
## CulmenLengthmm  0.4584221  0.585128108 -0.6543224 -0.1390555
## CulmenDepthmm -0.4012054  0.804189688  0.3997313  0.1803554
## FlipperLengthmm  0.5748007 -0.004501677  0.2319137  0.7847292
## BodyMassg       0.5463402  0.104325342  0.5985759 -0.5764855

#PCs of last two rows of penguins data
PC1<-predict(penguinsdata_comp, newdata=tail(penguinsdata_pca, 2))
PC1

##          PC1          PC2          PC3          PC4
## 343 1.704466 -0.6626986  0.2611904 -0.3277890
## 344 1.928683  0.3954538  0.1653766 -0.3705103

cor_penguinsdata<-cor(penguinsdata_pca)
cor_penguinsdata

##          CulmenLengthmm CulmenDepthmm FlipperLengthmm BodyMassg
## CulmenLengthmm      1.0000000   -0.2456429     0.6547169  0.6009286
## CulmenDepthmm     -0.2456429      1.0000000    -0.5870001 -0.4629559
## FlipperLengthmm     0.6547169     -0.5870001     1.0000000  0.8629381
## BodyMassg           0.6009286     -0.4629559     0.8629381  1.0000000

eigen_penguinsdata<-eigen(cor_penguinsdata)
eigen_penguinsdata$values

## [1] 2.7520891 0.7644982 0.3682808 0.1151318

penguinsdata.var.prop<-eigen_penguinsdata$values/sum(eigen_penguinsdata$values)
penguinsdata.var.prop

## [1] 0.68802228 0.19112456 0.09207021 0.02878296

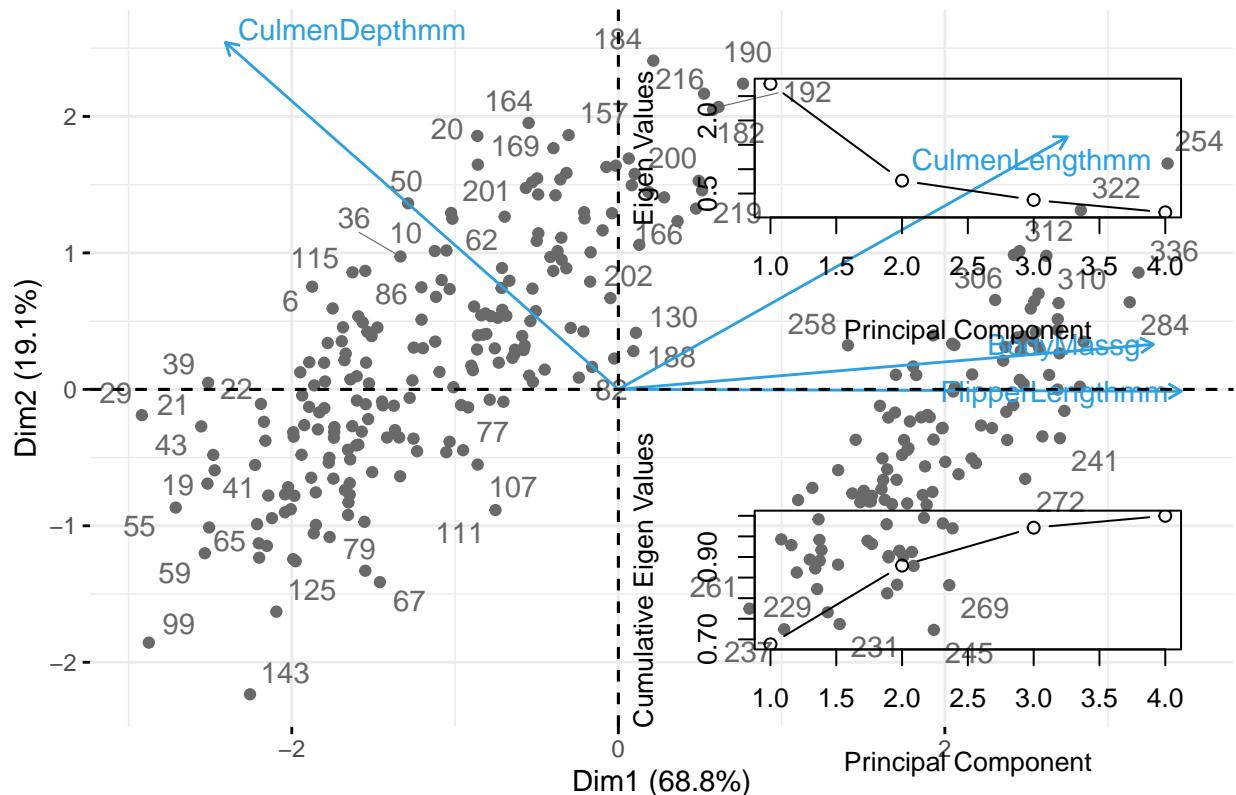
eigen_penguinsdata$vectors

##          [,1]          [,2]          [,3]          [,4]
## [1,]  0.4584221  0.585128108  0.6543224  0.1390555
## [2,] -0.4012054  0.804189688 -0.3997313 -0.1803554
## [3,]  0.5748007 -0.004501677 -0.2319137 -0.7847292
## [4,]  0.5463402  0.104325342 -0.5985759  0.5764855

plot(eigen_penguinsdata$values, xlab = "Principal Component", ylab = "Eigen Values", type = "b")
plot(cumsum(penguinsdata.var.prop), xlab = "Principal Component", ylab = "Cumulative Eigen Values", typ

```

## PCA – Biplot



```
#PCA using princomp()
penguinsdata_comp2<-princomp(penguinsdata[,10:13], scores=T)
penguinsdata_comp2$loadings
```

```
##
## Loadings:
##          Comp.1 Comp.2 Comp.3 Comp.4
## CulmenLengthmm      0.308  0.945  0.110
## CulmenDepthmm        0.146 -0.985
## FlipperLengthmm      0.947 -0.293 -0.133
## BodyMassg            1.000
##
##          Comp.1 Comp.2 Comp.3 Comp.4
## SS loadings     1.00   1.00   1.00   1.00
## Proportion Var  0.25   0.25   0.25   0.25
## Cumulative Var 0.25   0.50   0.75   1.00
```

```
penguinsdata_comp2_scores<-penguinsdata_comp2$scores
head(penguinsdata_comp2_scores)
```

```
##          Comp.1      Comp.2      Comp.3      Comp.4
## 2 -416.0327 -9.392485 -0.1062454  1.059688
## 3 -965.8256  7.956255 -2.3537433 -0.881863
## 5 -765.8966  1.691386 -4.8154781 -2.205863
```

```
## 6 -565.9580 -3.609898 -1.1255697 -2.712124
## 7 -591.0898 -11.599283  0.7071640  1.186911
## 8  458.9918 -14.913068 -1.9891751 -1.950820
```