**LAPORAN UJIAN AKHIR SEMESTER**

**STATISTIKA DESKRIPTIF**

**PRINCIPAL COMPONENT ANALYSIS (PCA)**



**NAMA : MUKHAMAD IKHSANUDIN**

**NIM : 082011633086**

**DOSEN PENGAMPU : Drs. ETO WURYANTO, DEA.**

**196609281991021001**

**PROGRAM STUDI S1 SISTEM INFORMASI**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS AIRLANGGA**

**2021**

```{R}

# PCA

library(factoextra)

library(rgl)

library(flexclust)

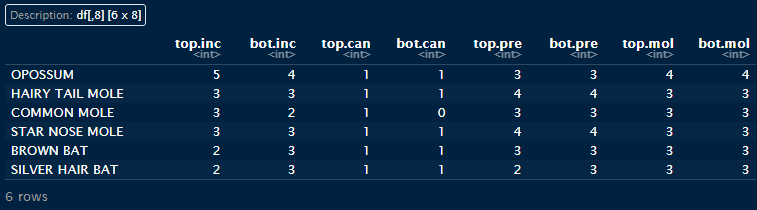
# Pembuatan PCA :

data("dentitio")

View(dentitio)

DataPCA <- dentitio

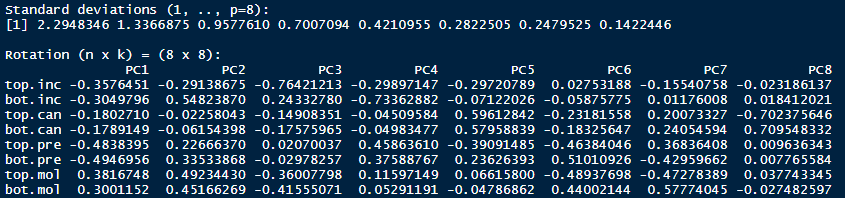
head(DataPCA)



# 1. Covariance

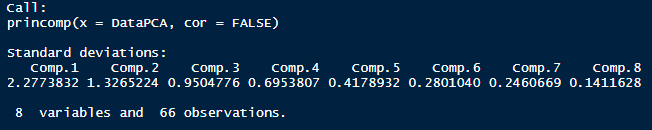
Cov\_1 <- prcomp(DataPCA, scale = FALSE)

Cov\_1



Cov\_2 <- princomp(DataPCA, cor = FALSE)

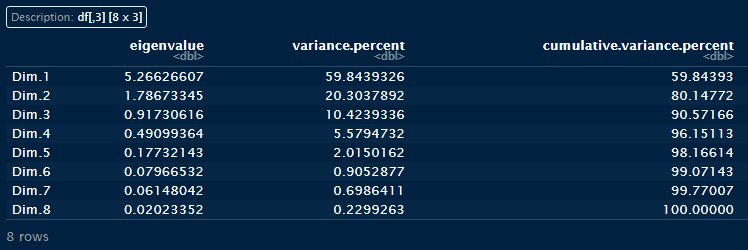
Cov\_2



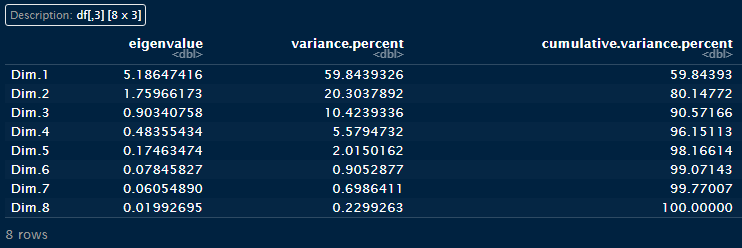
# Tampilkan eigen value

# Tabel

get\_eig(Cov\_1) # Prcomp



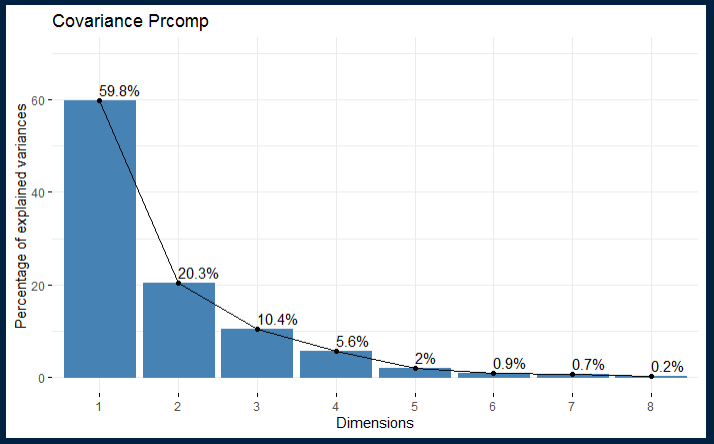
get\_eig(Cov\_2) # Princomp



# Grafik

fviz\_eig(Cov\_1, addlabels = TRUE, ylim = c(0, 70),

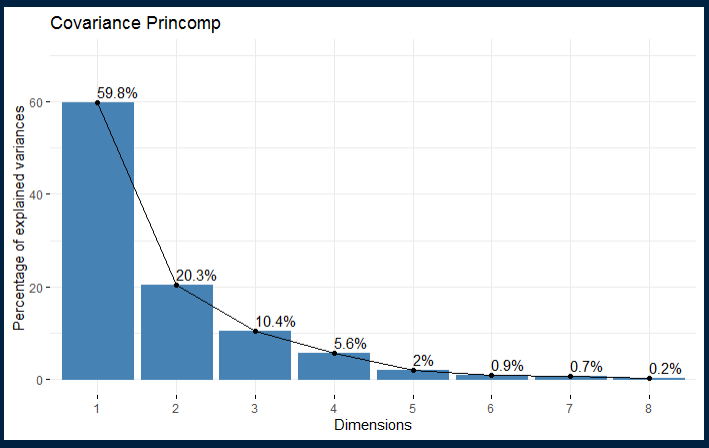
main = "Covariance Prcomp")



Berdasarkan plot prcomp di atas, didapatkan informasi persebaran data pada dimensi 1 sebesar 59,8 % dan dimensi 2 sebesar 20,3%. Berarti persebaran dari kedua dimensi tersebut adalah 80,1%. Dan data yang sudah terwakili juga sebesar 80,1% dari totalnya yang berjumlah 66.

fviz\_eig(Cov\_2, addlabels = TRUE, ylim = c(0, 70),

main = "Covariance Princomp")

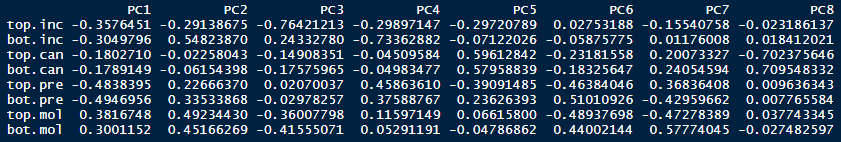


Berdasarkan plot princomp di atas, didapatkan informasi persebaran data pada dimensi 1 juga sebesar 59,8 % dan dimensi 2 juga sebesar 20,3%. Berarti persebaran dari kedua dimensi tersebut adalah 80,1%. Dan data yang sudah terwakili juga sebesar 70,1% dari totalnya yang berjumlah 66.

# Tampilkan Matriks PC

Mat\_Cov\_1 <- Cov\_1$rotation # Prcomp

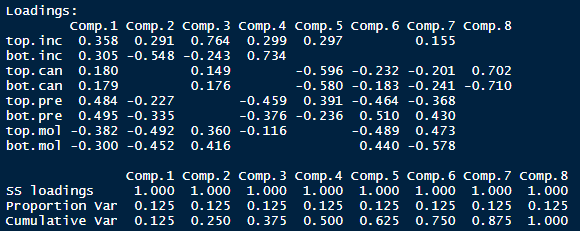
Mat\_Cov\_1



Dari 66 data, didapatkan kolom dari PC1 sampai PC8 yang mewakili matriks PC dari seluruh data setiap variabelnya.

Mat\_Cov\_2 <- Cov\_2$loadings # Princomp

Mat\_Cov\_2



Dari 66 data, didapatkan kolom dari Comp.1 sampai Comp.8 yang mewakili matriks PC dari seluruh data setiap variabelnya. Dan baris Cumulative Var membuktikan bahwa total data keseluruhan sampai Comp.8 adalah 100%.

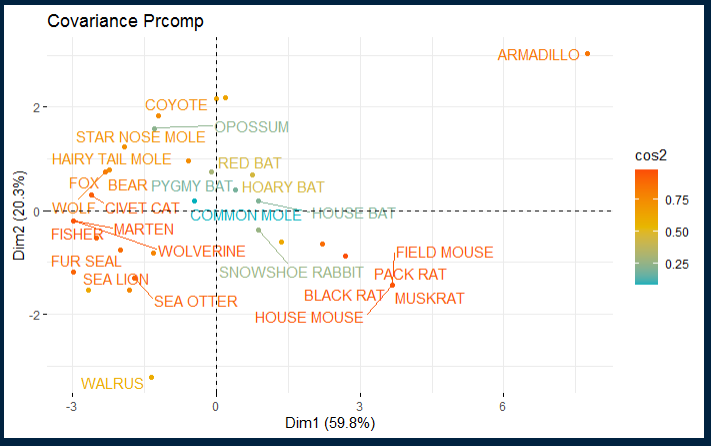
# Plot 2 dimensi untuk individu

fviz\_pca\_ind(Cov\_1,

col.ind = "cos2",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, title = "Covariance Prcomp")



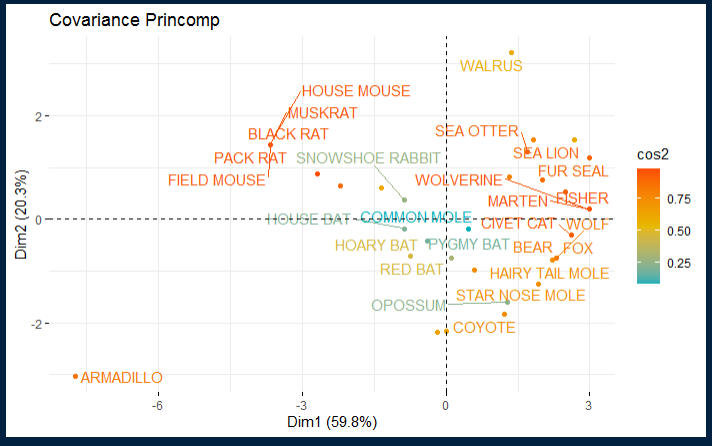
Berdasarkan plot pca mengenai persebaran jumlah gigi beberapa mamalia di atas, didapatkan 80,1% data yang sudah tersebar pada dimensi 1 dan 2.

fviz\_pca\_ind(Cov\_2,

col.ind = "cos2",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, title = "Covariance Princomp")



# "cos2" digunakan untuk melihat pengaruh individunya terhadap dimensi

Berdasarkan plot pca mengenai persebaran jumlah gigi beberapa mamalia di atas, didapatkan 80,1% data yang sudah tersebar pada dimensi 1 dan 2. Perbedaan dari kedua plot di atas adalah pada visualisasi dimensinya.

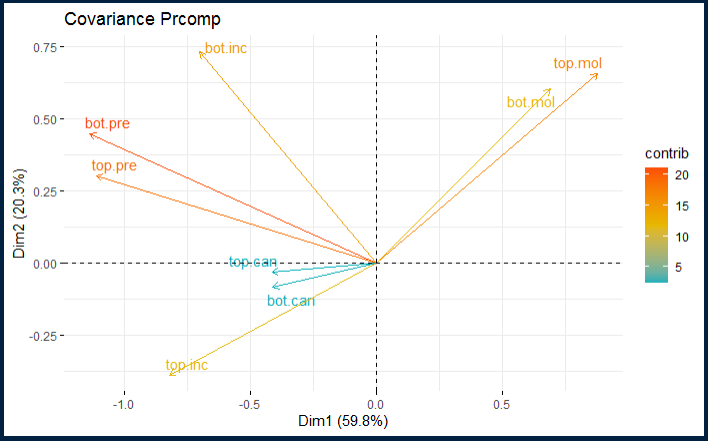
# Plot 2 dimensi untuk variabel

fviz\_pca\_var(Cov\_1,

col.var = "contrib",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, title = "Covariance Prcomp")



Berdasarkan plot variable di atas, didapatkan informasi mengenai persebaran variabel yang digunakan. Dari 8 variabel tersebut, persebaran data yang diperoleh sebesar 80,1% (pada dimensi 1 dan 2). Plot tersebut menunjukkan bahwa :

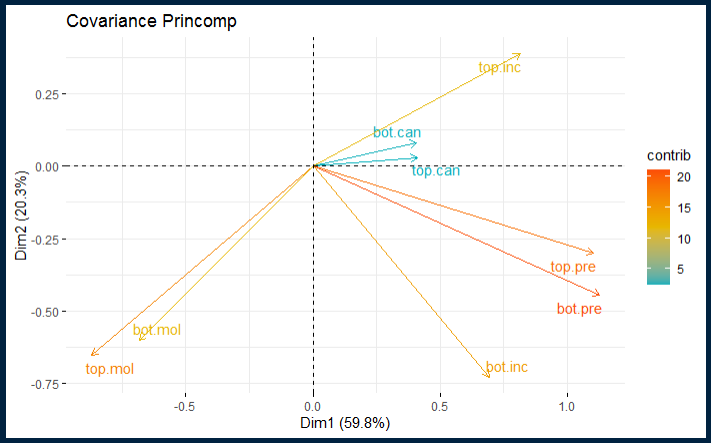
* bot.mol dan top.mol berisi data dengan rata-rata jumlah gigi terbanyak (pada dimensi 1 positif dan 2 positif)
* top.inc berisi data dengan rata-rata jumlah gigi terkecil (pada dimensi 1 negatif dan 2 negatif)
* bot.can dan top.can berisi data dengan rata-rata jumlah gigi mendekati 0 (pada dimensi 1 negatif dan 2 negatif)
* bot.inc, bot.pre, dan top.pre berisi data dengan rata-rata jumlah gigi yang lebih dominan pada dimensi 2 positif

fviz\_pca\_var(Cov\_2,

col.var = "contrib",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, title = "Covariance Princomp")



# "contrib" digunakan untuk melihat pengaruh variabel terhadap PC 1 atau 2

Berdasarkan plot variable di atas, didapatkan informasi mengenai persebaran variabel yang digunakan. Dari 8 variabel tersebut, persebaran data yang diperoleh sebesar 80,1% (pada dimensi 1 dan 2). Plot tersebut menunjukkan bahwa :

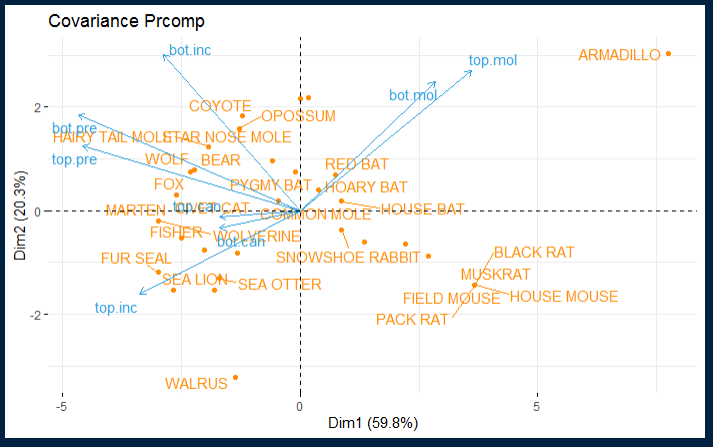
* bot.mol dan top.mol berisi data dengan rata-rata jumlah gigi terkecil (pada dimensi 1 negatif dan 2 negatif)
* top.inc berisi data dengan rata-rata jumlah gigi terbesar (pada dimensi 1 positif dan 2 positif)
* bot.can dan top.can berisi data dengan rata-rata jumlah gigi mendekati 0 (pada dimensi 1 positif dan 2 positif)
* bot.inc, bot.pre, dan top.pre berisi data dengan rata-rata jumlah gigi yang lebih dominan pada dimensi 1 positif

# Plot 2 dimensi untuk variabel + individu

fviz\_pca\_biplot(Cov\_1, repel = TRUE,

col.var = "#2E9FDF", col.ind = "#FF8C00",

title = "Covariance Prcomp")



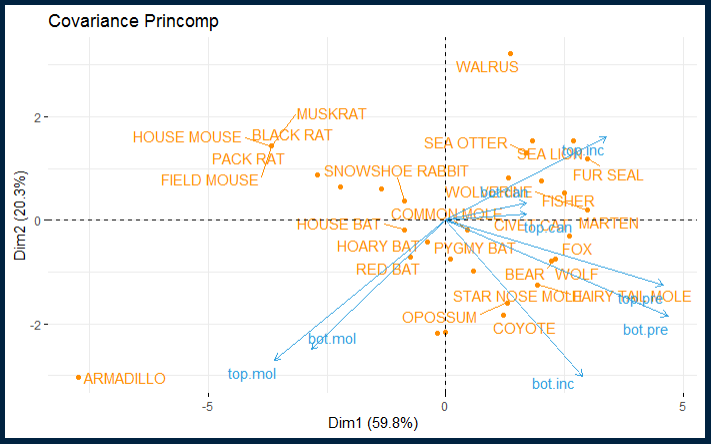
Berdasarkan biplot di atas, didapatkan informasi mengenai persebaran variabel yang digunakan. Dari 8 variabel tersebut, persebaran data yang diperoleh sebesar 80,1% (pada dimensi 1 dan 2). Plot tersebut menunjukkan bahwa :

* bot.mol dan top.mol berisi data dengan rata-rata jumlah gigi terbanyak (pada dimensi 1 positif dan 2 positif). Data-data tersebut sebagian besar terdapat pada armadillo, red bat, hoary bat, dan house bat
* bot.can, top.can, dan top.inc berisi data dengan rata-rata jumlah gigi yang kecil (pada dimensi 1 negatif dan 2 negatif). Data-data tersebut sebagian besar terdapat pada sea lion, fur seal, fisher, dan lain-lain
* bot.inc, bot.pre, dan top.pre berisi data dengan rata-rata jumlah gigi yang lebih dominan pada dimensi 2 positif. Data-data tersebut sebagian besar terdapat pada coyote, wolf, bear, nose mole, dan lain-lain
* Sedangkan data seperti rabbit, black rat, field mouse, dan lainnya tidak dominan pada salah satu bagian variabel

fviz\_pca\_biplot(Cov\_2, repel = TRUE,

col.var = "#2E9FDF", col.ind = "#FF8C00",

title = "Covariance Princomp")



Berdasarkan biplot di atas, didapatkan informasi mengenai persebaran variabel yang digunakan. Dari 8 variabel tersebut, persebaran data yang diperoleh sebesar 80,1% (pada dimensi 1 dan 2). Plot tersebut menunjukkan bahwa :

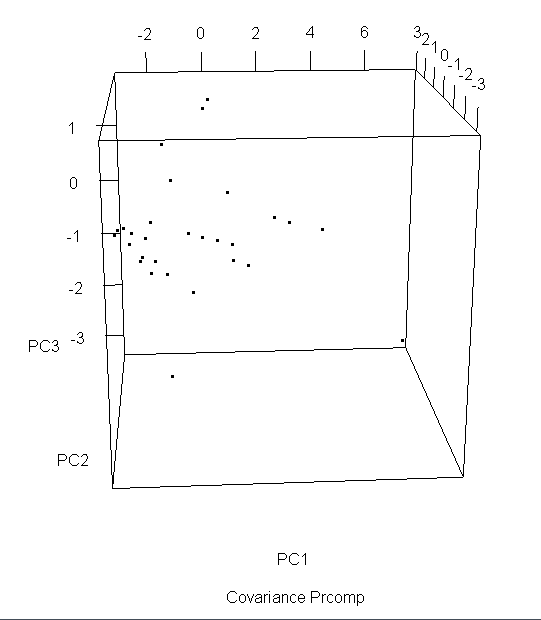
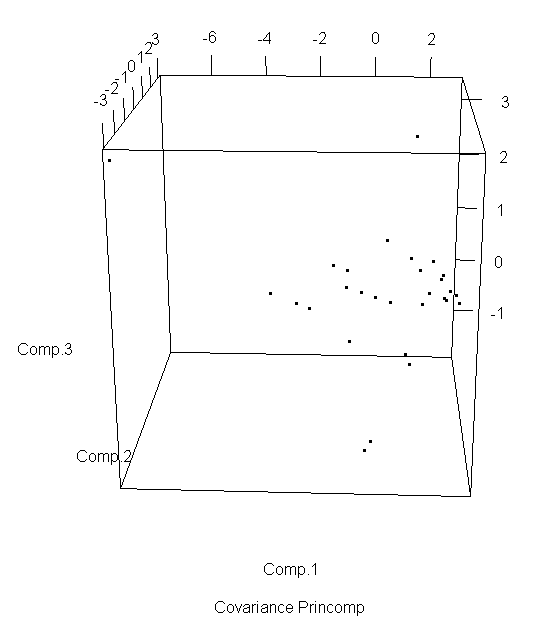
* bot.mol dan top.mol berisi data dengan rata-rata jumlah gigi terkecil (pada dimensi 1 negatif dan 2 negatif). Data-data tersebut sebagian besar terdapat pada armadillo, red bat, hoary bat, dan house bat
* bot.can, top.can, dan top.inc berisi data dengan rata-rata jumlah gigi yang kecil (pada dimensi 1 positif dan 2 positif). Data-data tersebut sebagian besar terdapat pada sea lion, fur seal, fisher, dan lain-lain
* bot.inc, bot.pre, dan top.pre berisi data dengan rata-rata jumlah gigi yang lebih dominan pada dimensi 1 positif. Data-data tersebut sebagian besar terdapat pada coyote, wolf, bear, nose mole, dan lain-lain
* Sedangkan data seperti rabbit, black rat, field mouse, dan lainnya tidak dominan pada salah satu bagian variabel

# Plot 3 Dimensi

plot3d(Cov\_1$x, col = 1, sub = "Covariance Prcomp")

plot3d(Cov\_2$scores, col = 1, sub = "Covariance Princomp")

```

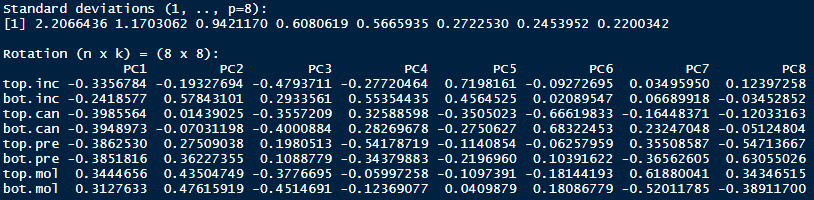
Berdasarkan plot 3d prcomp di atas, data tersebut dominan tersebar pada PC3 antara 1 dan -3. Sedangkan pada plot 3d princomp, data dominan tersebar pada Comp.3 antara 1 dan -1

```{R}

# 2. Correlation

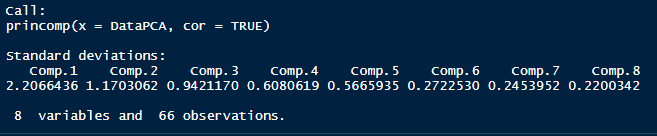
Cor\_1 <- prcomp(DataPCA, scale = TRUE)

Cor\_1



Cor\_2 <- princomp(DataPCA, cor = TRUE)

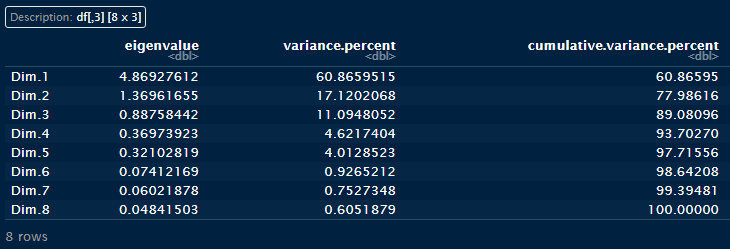
Cor\_2



# Tampilkan eigen value

# Tabel

get\_eig(Cor\_1)



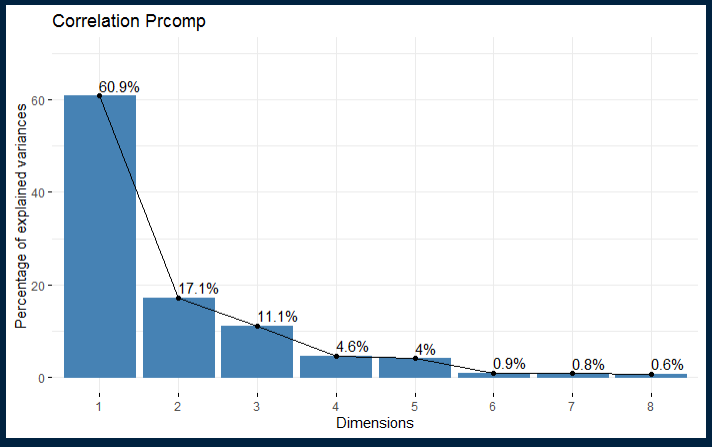
get\_eig(Cor\_2)



# Grafik

fviz\_eig(Cor\_1, addlabels = TRUE, ylim = c(0, 70),

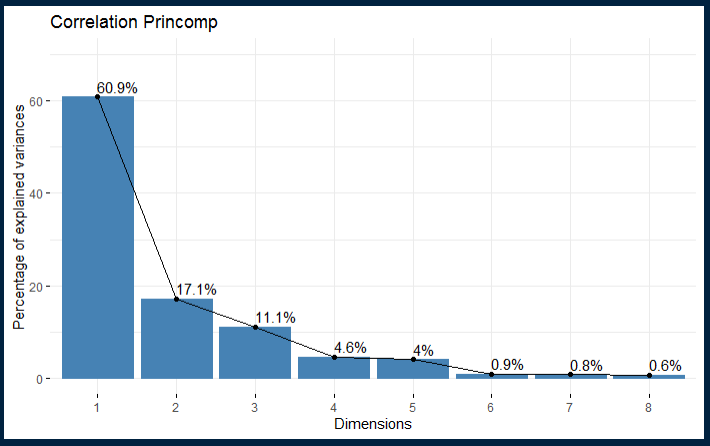
main = "Correlation Prcomp")



Berdasarkan plot prcomp di atas, didapatkan informasi persebaran data pada dimensi 1 sebesar 60,9 % dan dimensi 2 sebesar 17,1%. Berarti persebaran dari kedua dimensi tersebut adalah 78%. Dan data yang sudah terwakili juga sebesar 78% dari totalnya yang berjumlah 66.

fviz\_eig(Cor\_2, addlabels = TRUE, ylim = c(0, 70),

main = "Correlation Princomp")

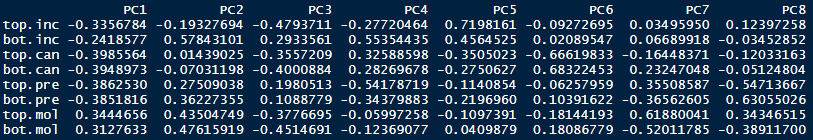


Berdasarkan plot princomp di atas, didapatkan informasi persebaran data pada dimensi 1 juga sebesar 60,9 % dan dimensi 2 juga sebesar 17,1%. Berarti persebaran dari kedua dimensi tersebut adalah 78%. Dan data yang sudah terwakili juga sebesar 78% dari totalnya yang berjumlah 66.

# Tampilkan Matriks PC

Mat\_Cor\_1 <- Cor\_1$rotation # Prcomp

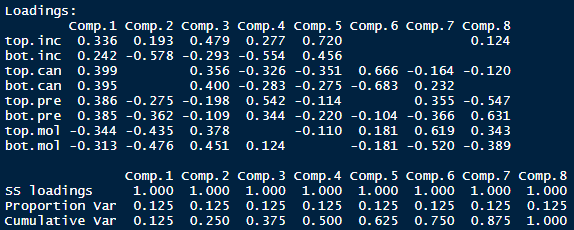
Mat\_Cor\_1



Dari 66 data, didapatkan kolom dari PC1 sampai PC8 yang mewakili matriks PC dari seluruh data setiap variabelnya.

Mat\_Cor\_2 <- Cor\_2$loadings # Princomp

Mat\_Cor\_2

Dari 66 data, didapatkan kolom dari Comp.1 sampai Comp.8 yang mewakili matriks PC dari seluruh data setiap variabelnya. Dan baris Cumulative Var membuktikan bahwa total data keseluruhan sampai Comp.8 adalah 100%

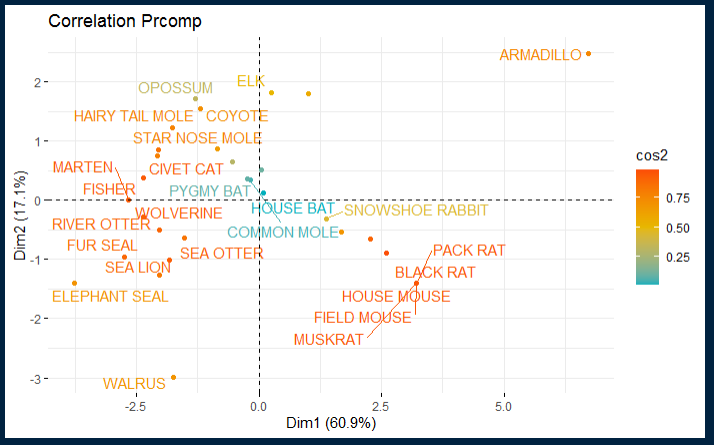
# Plot 2 dimensi untuk individu

fviz\_pca\_ind(Cor\_1,

col.ind = "cos2",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, title = "Correlation Prcomp")



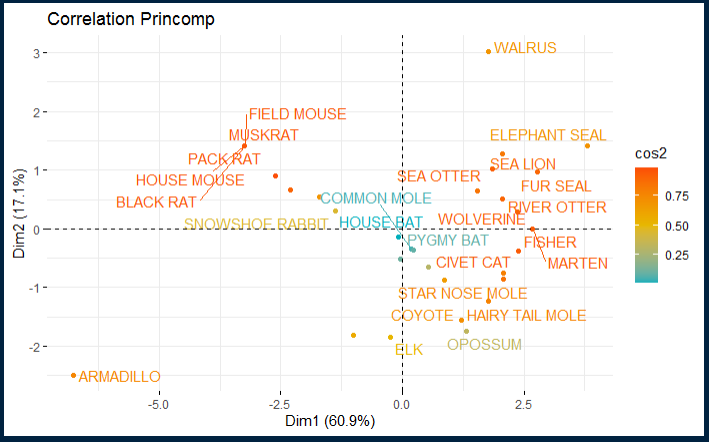
Berdasarkan plot pca mengenai persebaran jumlah gigi beberapa mamalia di atas, didapatkan 78% data yang sudah tersebar pada dimensi 1 dan 2.

fviz\_pca\_ind(Cor\_2,

col.ind = "cos2",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, title = "Correlation Princomp")



# "cos2" digunakan untuk melihat pengaruh individunya terhadap dimensi

Berdasarkan plot pca mengenai persebaran jumlah gigi beberapa mamalia di atas, didapatkan 78% data yang sudah tersebar pada dimensi 1 dan 2. Perbedaan dari kedua plot di atas adalah pada visualisasi data dengan dimensinya.

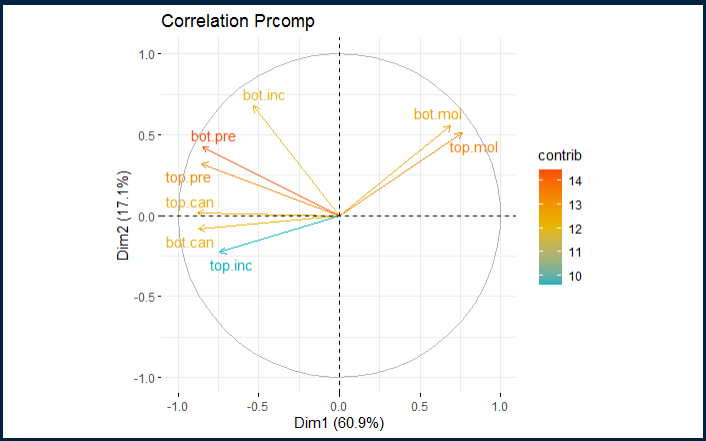
# Plot 2 dimensi untuk variabel

fviz\_pca\_var(Cor\_1,

col.var = "contrib",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, title = "Correlation Prcomp")



Berdasarkan plot variable di atas, didapatkan informasi mengenai persebaran variabel yang digunakan. Dari 8 variabel tersebut, persebaran data yang diperoleh sebesar 78% (pada dimensi 1 dan 2). Plot tersebut menunjukkan bahwa :

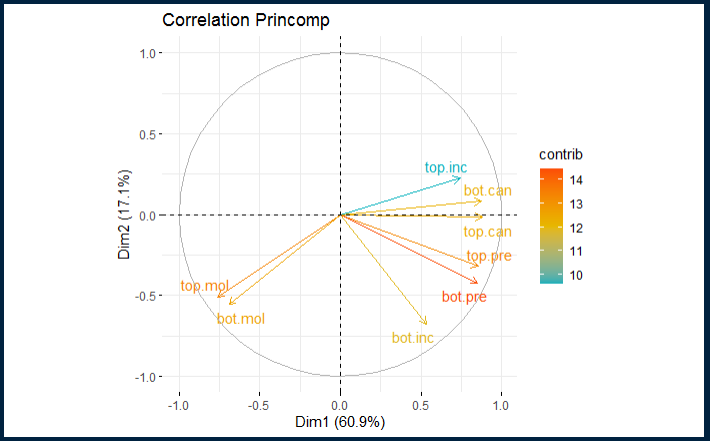
* bot.mol dan top.mol berisi data dengan rata-rata jumlah gigi terbanyak (pada dimensi 1 positif dan 2 positif)
* top.inc dan bot.can berisi data dengan rata-rata jumlah gigi dominan kecil (pada dimensi 1 negatif dan 2 negatif)
* bot.inc, bot.pre, top.pre, dan top.can berisi data dengan rata-rata jumlah gigi yang lebih dominan pada dimensi 2 positif

fviz\_pca\_var(Cor\_2,

col.var = "contrib",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, title = "Correlation Princomp")



# "contrib" digunakan untuk melihat pengaruh variabel terhadap PC 1 atau 2

Berdasarkan plot variable di atas, didapatkan informasi mengenai persebaran variabel yang digunakan. Dari 8 variabel tersebut, persebaran data yang diperoleh sebesar 78% (pada dimensi 1 dan 2). Plot tersebut menunjukkan bahwa :

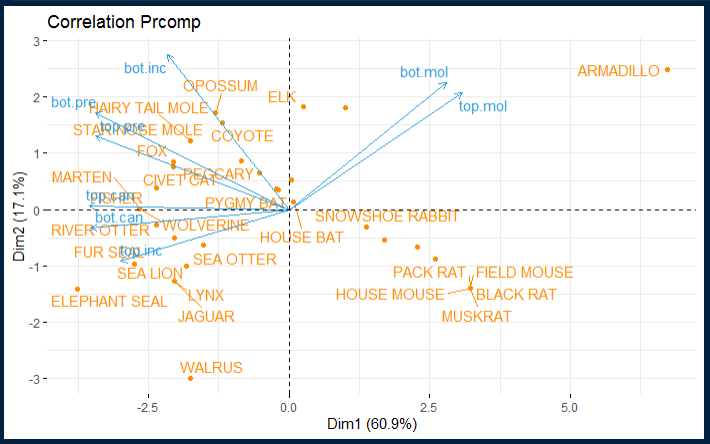
* bot.mol dan top.mol berisi data dengan rata-rata jumlah gigi terkecil (pada dimensi 1 negatif dan 2 negatif)
* top.inc dan bot.can berisi data dengan rata-rata jumlah gigi dominan besar (pada dimensi 1 positif dan 2 positif)
* bot.inc, bot.pre, top.pre, dan top.can berisi data dengan rata-rata jumlah gigi yang lebih dominan pada dimensi 1 positif

# Plot 2 dimensi untuk variabel + individu

fviz\_pca\_biplot(Cor\_1, repel = TRUE,

col.var = "#2E9FDF", col.ind = "#FF8C00",

title = "Correlation Prcomp")



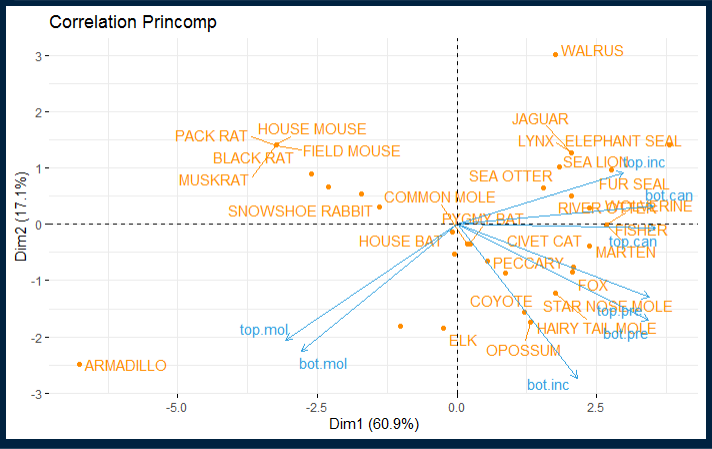
Berdasarkan biplot di atas, didapatkan informasi mengenai persebaran variabel yang digunakan. Dari 8 variabel tersebut, persebaran data yang diperoleh sebesar 78% (pada dimensi 1 dan 2). Plot tersebut menunjukkan bahwa :

* bot.mol dan top.mol berisi data dengan rata-rata jumlah gigi terbanyak (pada dimensi 1 positif dan 2 positif). Data-data tersebut sebagian besar terdapat pada armadillo dan elk
* top.inc dan bot.can berisi data dengan rata-rata jumlah gigi dominan kecil (pada dimensi 1 negatif dan 2 negatif). Data-data tersebut sebagian besar terdapat pada walrus, jaguar, elephant, seal, dan lain-lain
* bot.inc, bot.pre, top.pre, dan top.can berisi data dengan rata-rata jumlah gigi yang lebih dominan pada dimensi 2 positif. Data-data tersebut sebagian besar terdapat pada fox, coyote, fisher, dan lain-lain

fviz\_pca\_biplot(Cor\_2, repel = TRUE,

col.var = "#2E9FDF", col.ind = "#FF8C00",

title = "Correlation Princomp")



Berdasarkan biplot di atas, didapatkan informasi mengenai persebaran variabel yang digunakan. Dari 8 variabel tersebut, persebaran data yang diperoleh sebesar 78% (pada dimensi 1 dan 2). Plot tersebut menunjukkan bahwa :

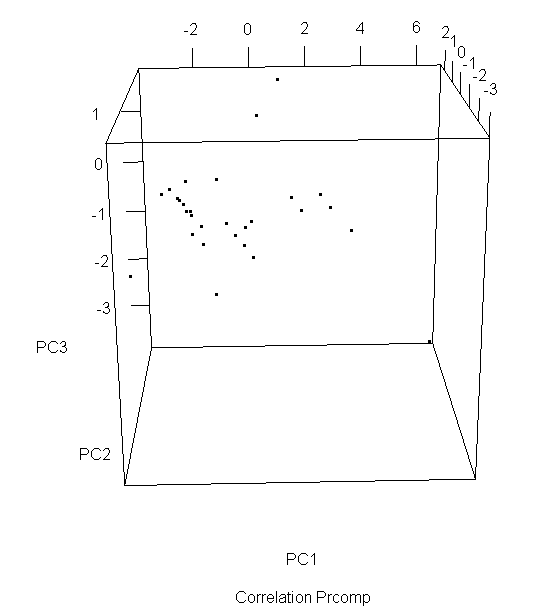
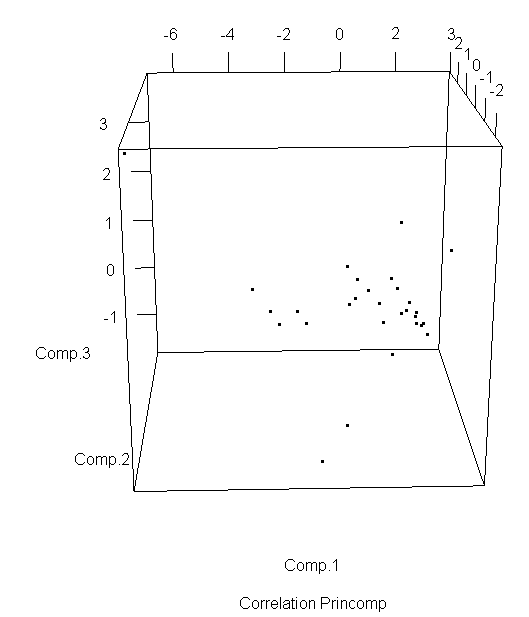
* bot.mol dan top.mol berisi data dengan rata-rata jumlah gigi terkecil (pada dimensi 1 negatif dan 2 negatif). Data-data tersebut sebagian besar terdapat pada armadillo dan elk
* top.inc dan bot.can berisi data dengan rata-rata jumlah gigi dominan besar (pada dimensi 1 positif dan 2 positif). Data-data tersebut sebagian besar terdapat pada walrus, jaguar, lynx, elephant, dan lain-lain
* bot.inc, bot.pre, top.pre, dan top.can berisi data dengan rata-rata jumlah gigi yang lebih dominan pada dimensi 1 positif. Data-data tersebut sebagian besar terdapat pada coyote, fox, fisher, dan lain-lain

# Plot 3 dimensi

plot3d(Cor\_1$x, col = 1, sub = "Correlation Prcomp")

plot3d(Cor\_2$scores, col = 1, sub = "Correlation Princomp")

```

Berdasarkan plot 3d prcomp di atas, data tersebut dominan tersebar pada PC3 antara 1 dan -3. Sedangkan pada plot 3d princomp, data dominan tersebar pada Comp.3 antara 1 dan -1

**Kesimpulan**

Data *dentition* dapat diolah menggunakan metode PCA yang akan membuatnya menjadi lebih ringkas karena dibagi menjadi beberapa bagian (PC dan Comp). Penggunaan prcomp dan princom juga mempengaruhi plot yang akan dibuat. Meski begitu, nilai eigen dari keduanya cukup mirip dan juga berlawanan antara negatif (-) dan positifnya.