**LAPORAN UJIAN AKHIR SEMESTER**

**STATISTIKA DESKRIPTIF**

**CORRESPONDENCE ANALYSIS AND MULTIPLE CORRESPONDENCE ANALYSIS (CA & MCA)**



**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}

# CA & MCA

library(flexclust)

library(cluster)

library(factoextra)

library(mclust)

library(FactoMineR)

library(corrplot)

library(ade4)

library(MASS)

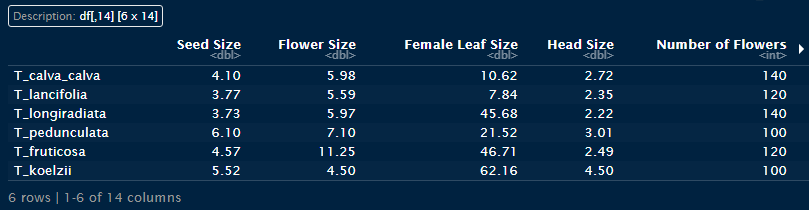
# CA

# Dataset Preparation

load("Tithonia.Rda")

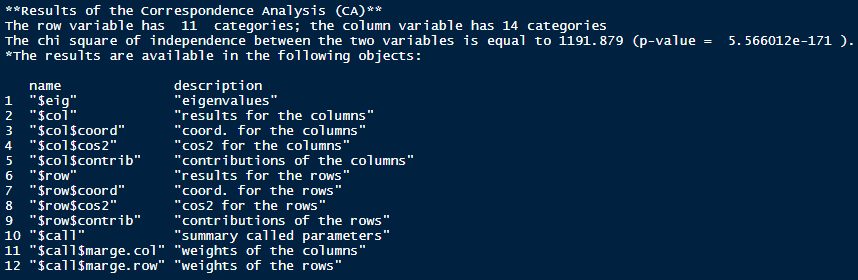
Data\_CA <- Tithonia

head(Data\_CA)



my\_CA <- CA(Data\_CA, graph = FALSE)

my\_CA



CA merupakan fungsi untuk menghitung analisis korespodensi dan mendapatkan hasil sebagai berikut, variabel baris memiliki 27 kategori dan variabel kolom memiliki 3 kategori

Hasilnya tersedia di 12 objek.

# Statisfical Significance

# Chi-Square Statistics

Chi <- 42.652

# Degree of Freedom

DF1 <- (1/(nrow(Data\_CA)-1))\*100

DF1

DF2 <- (1/(ncol(Data\_CA)-1))\*100

DF2

# P-Value

PVal <- pchisq(Chi, DF1 == DF2, lower.tail = FALSE)

PVal



Chi square bernilai 42,652

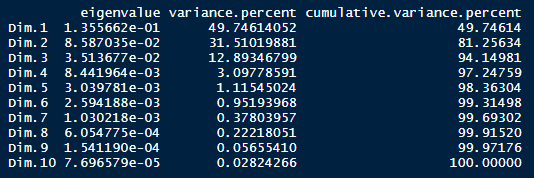
Derajat kebebasan untuk baris bernilai 10%

Derajat kebebasan untuk kolom bernilai 7,69%

Nilai P adalah 0

# Eigenvalues / Variances

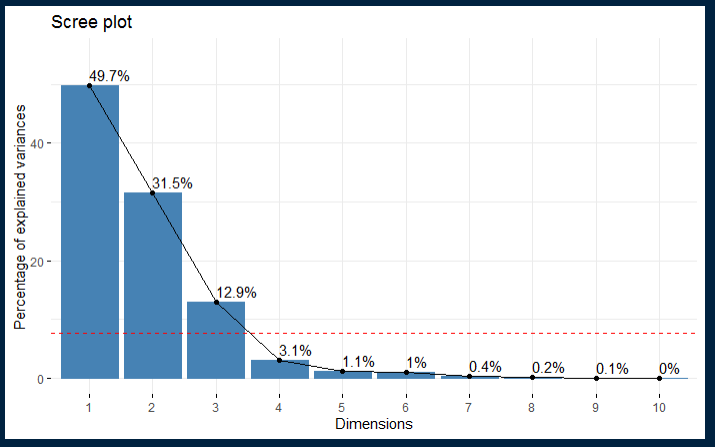
get\_eig(my\_CA)



Dari perhitungan tersebut didapat 10 nilai eigen yang diurutkan dari varians terbesar, sehingga pada dimensi 1 nilai eigennya adalah 0.0135 dan dimensi berikutnya sampai dimensi 10.

fviz\_screeplot(my\_CA, addlabels = TRUE, ylim = c(0, 55)) +

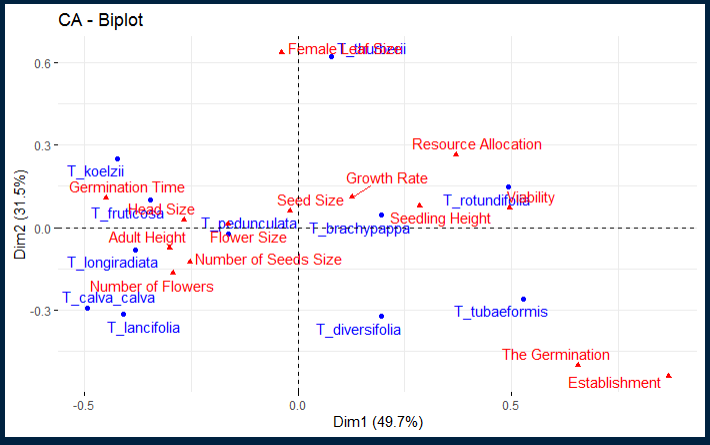
geom\_hline(yintercept = DF2, linetype = 2, color = "red")



Berdasarkan scree plot pada dimensi 1, distribusi data yang dapat dilakukan analisis berada pada persentase variance 49,7 % dimana pada dimensi 1 ini distribusi data terletak pada kondisi distribusi data yang terbaik. Hal ini karena semakin besar persentase variancenya menunjukkan bahwa data tersebut telah banyak terdistribusi dan masuk untuk dilakukan analisis. Lalu distribusi data pada dimensi 2 dengan variance 31,5% dan seterusnya.Dari plot di atas, terdapat 3 dimensi yang memenuhi derajat kebebasan.

# Biplot

fviz\_ca\_biplot(my\_CA, rep = TRUE)



Berdasarkan biplot di atas, terdapat 90,2% persebaran data yang masuk. Dari plot di atas, kita tahu berbagai spesies (biru) dengan karakteristiknya (merah). Terdapat beberapa karakteristik yang sangat dominan bagi suatu spesies. T\_thuberii sangat dominan dengan Female Leaf Size nya (memiliki nilai Female Leaf Size terbesar) dan seterusnya.

# Graph or Row variables

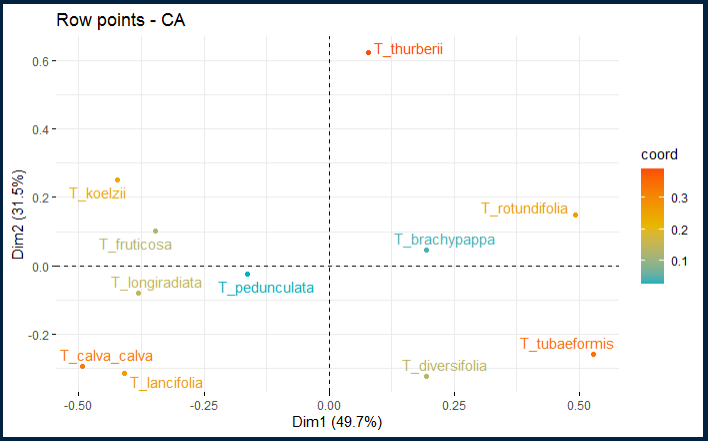
my\_CA\_Row <- my\_CA$row

# Coord

fviz\_ca\_row(my\_CA, col.row = "coord",

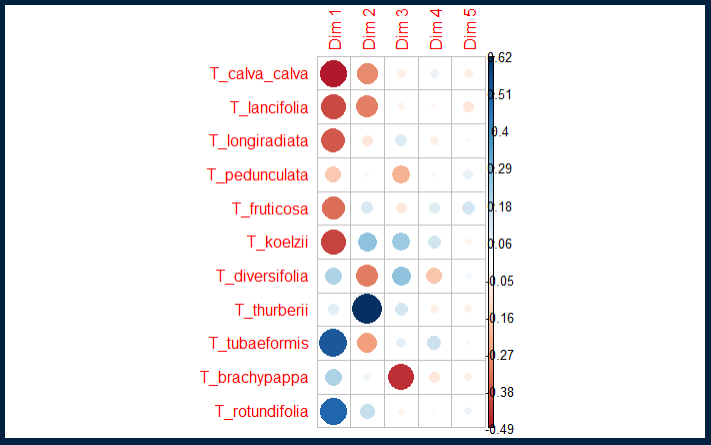
gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE)



Dari plot CA di atas, data yang telah tersebar adalah 90,2% dan didapati bahwa T\_thuberii memiliki nilai positif pada kedua dimensi dan dominan dengan dimensi 2. Sedangkan T\_rotundifolia dan T\_brachypappa memiliki nilai positif pada keduanya dan dominan dengan dimensi 1. Lalu terdapat T\_koelzii dan T\_frticosa yang memiliki nilai dimensi 2 positif dan dimensi 1 negatif. T\_diversifolia dan T\_tubaeformis memiliki nilai dimensi 1 positif namun nilai dimensi 2 nya negatif. Serta T\_calva\_calva, T\_lancifolia, T\_longiradiata dan T\_pedundulata memiliki nilai negatif di kedua dimensinya

corrplot(my\_CA\_Row$coord, is.corr = FALSE)



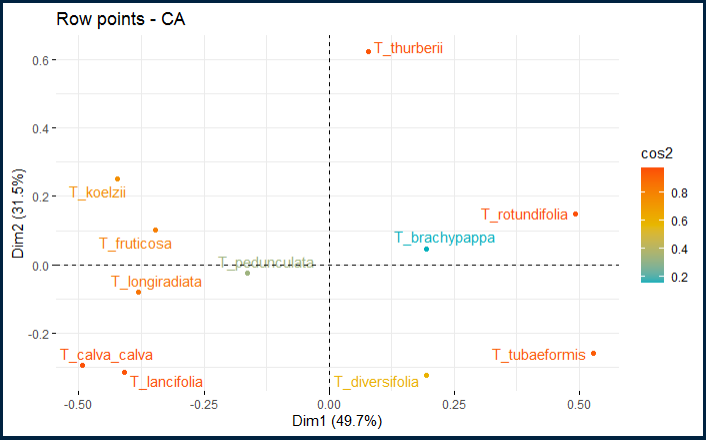
1. Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai koordinasi baris tertinggi hingga terendah. Pada dimensi 1, yang terendah adalah T\_calva\_calva dan yang tertinggi adalah T\_rotundifolia.
2. Sedangkan untuk kontribusinya, T\_calva\_calva merupakan yang tertingga dan T\_penduculata merupakan yang terendah

# Cos2

fviz\_ca\_row(my\_CA, col.row = "cos2",

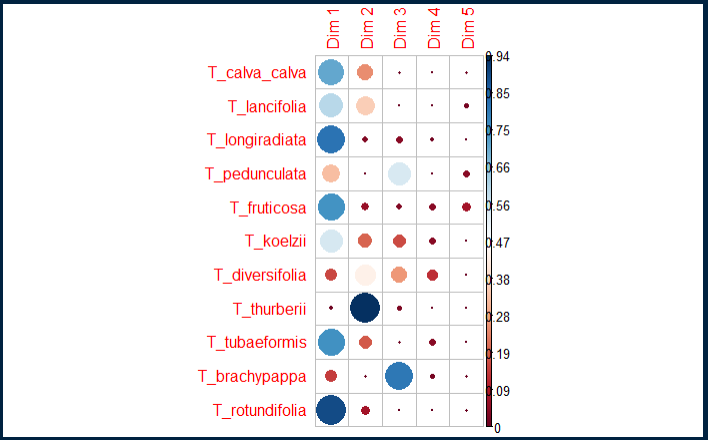
gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE)



Dari plot CA di atas, data yang telah tersebar adalah 90,2% dan didapati bahwa T\_thuberii memiliki nilai positif pada kedua dimensi dan dominan dengan dimensi 2. Sedangkan T\_rotundifolia dan T\_brachypappa memiliki nilai positif pada keduanya dan dominan dengan dimensi 1. Lalu terdapat T\_koelzii dan T\_frticosa yang memiliki nilai dimensi 2 positif dan dimensi 1 negatif. T\_diversifolia dan T\_tubaeformis memiliki nilai dimensi 1 positif namun nilai dimensi 2 nya negatif. Serta T\_calva\_calva, T\_lancifolia, T\_longiradiata dan T\_pedundulata memiliki nilai negatif di kedua dimensinya. Yang berbeda dari plot CA sebelumnya adalah skala visualisasi warnanya. Pada coord, nilai tertinggi adalah 0.5, sedangkan pada cos2 nilai tertinggi adalah 1.

corrplot(my\_CA\_Row$cos2, is.corr = FALSE)



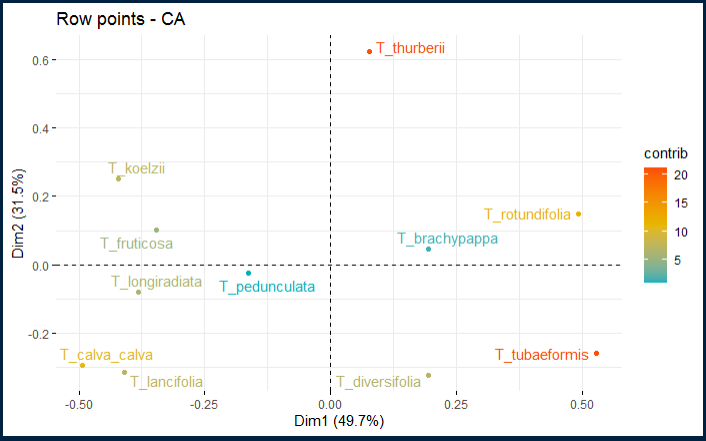
1. Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai koordinasi baris tertinggi hingga terendah. Pada dimensi 1, yang terendah adalah T\_thurberii dan yang tertinggi adalah T\_rotundifolia.
2. Sedangkan untuk kontribusinya, T\_lancifolia merupakan yang tertingga dan T\_diversifolia merupakan yang terendah

# Contrib

fviz\_ca\_row(my\_CA, col.row = "contrib",

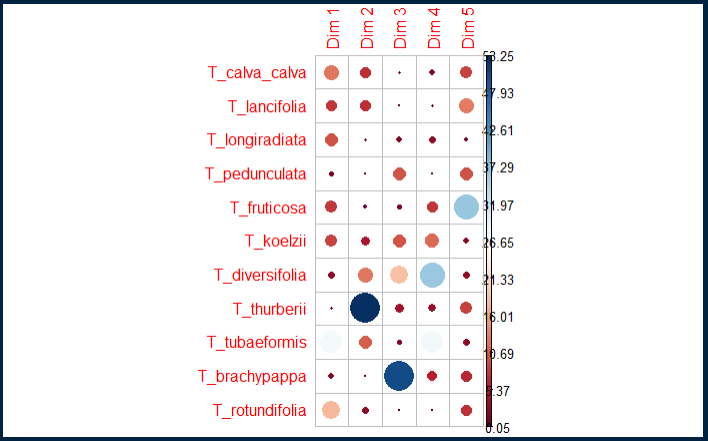
gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE)



Dari plot CA di atas, data yang telah tersebar adalah 90,2% dan didapati bahwa T\_thuberii memiliki nilai positif pada kedua dimensi dan dominan dengan dimensi 2. Sedangkan T\_rotundifolia dan T\_brachypappa memiliki nilai positif pada keduanya dan dominan dengan dimensi 1. Lalu terdapat T\_koelzii dan T\_frticosa yang memiliki nilai dimensi 2 positif dan dimensi 1 negatif. T\_diversifolia dan T\_tubaeformis memiliki nilai dimensi 1 positif namun nilai dimensi 2 nya negatif. Serta T\_calva\_calva, T\_lancifolia, T\_longiradiata dan T\_pedundulata memiliki nilai negatif di kedua dimensinya. Yang berbeda dari plot CA sebelumnya adalah skala visualisasi warnanya. Pada coord, nilai tertinggi adalah 0.5, pada cos2 nilai tertinggi adalah 1, dan pada contrib, nilai tertingginya adalah 20.

corrplot(my\_CA\_Row$contrib, is.corr = FALSE)



1. Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai koordinasi baris tertinggi hingga terendah. Pada dimensi 1, yang terendah adalah T\_disversifolia dan yang tertinggi adalah T\_tubaeformis.
2. Sedangkan untuk kontribusinya, T\_thuberii merupakan yang tertingga dan T\_longiradiata merupakan yang terendah

# Graph of Column variables

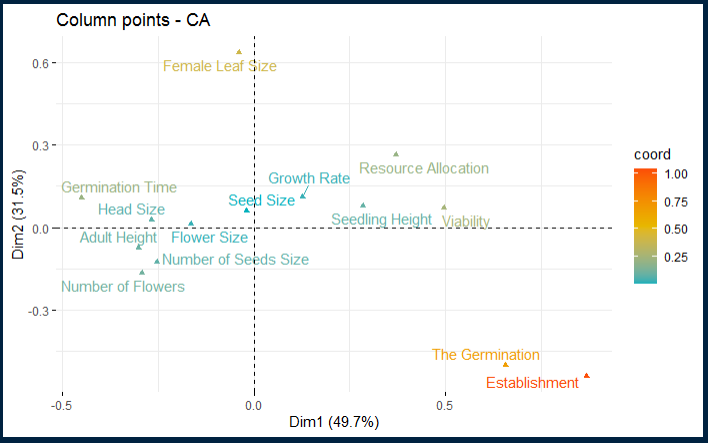
my\_CA\_Col <- my\_CA$col

# Coord

fviz\_ca\_col(my\_CA, col.col = "coord",

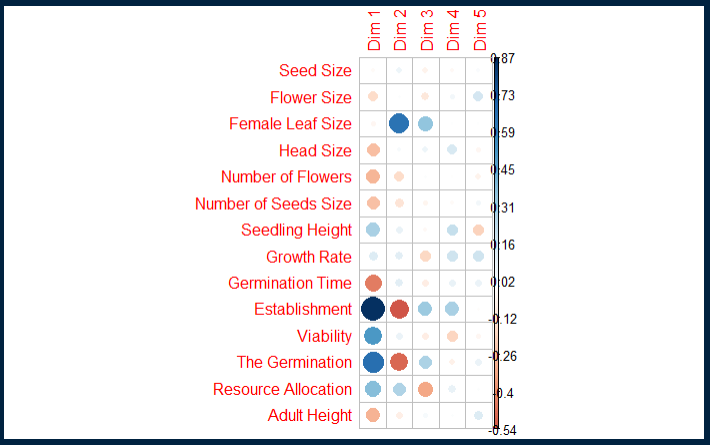
gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE)



Dari plot CA di atas, data yang telah tersebar adalah 90,2% dan didapati bahwa terdapat 2 variabel dengan nilai dimensi 1 positif dan dimensi 2 negatif. Lalu ada 4 variabel dengan nilai kedua dimensi positif.terdapat 5 variabel dengan nilai dimensi 1 negatif dan dimensi 2 positif. Serta ada 3 variabel dengan nilai kedua dimensi negatif.

corrplot(my\_CA\_Col$coord, is.corr = FALSE)



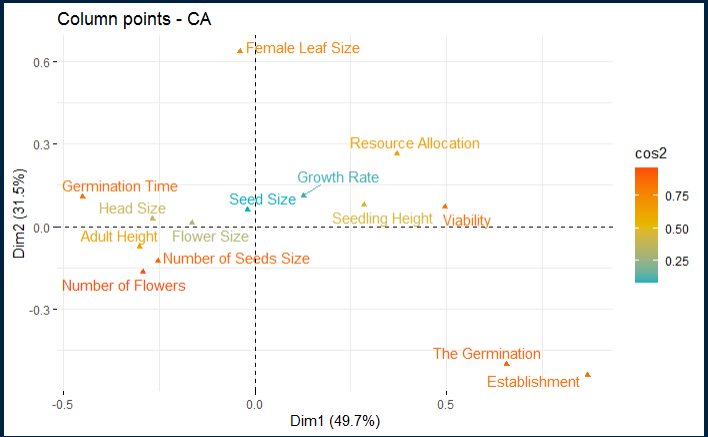
1. Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai koordinasi baris tertinggi hingga terendah. Pada dimensi 1, yang terendah adalah Germination Time dan yang tertinggi adalah Estabilshment.
2. Sedangkan untuk kontribusinya, Estabilshment merupakan yang tertingga dan Seed Size merupakan yang terendah

# Cos2

fviz\_ca\_col(my\_CA, col.col = "cos2",

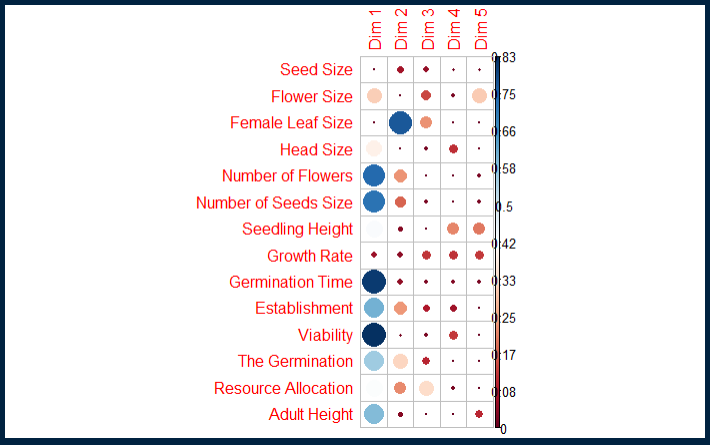
gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE)



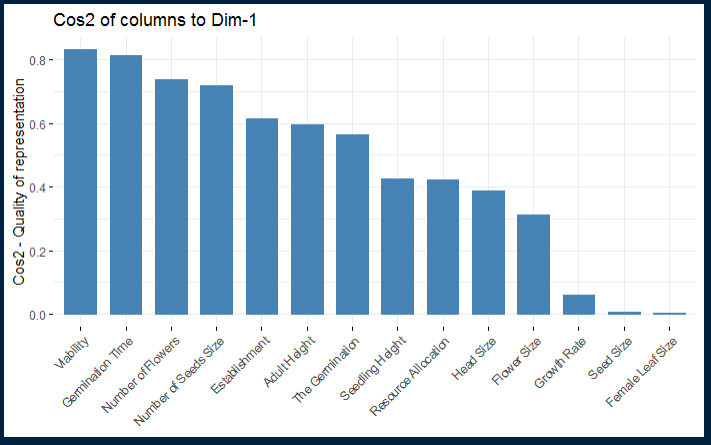
Dari plot CA di atas, data yang telah tersebar adalah 90,2% dan didapati bahwa terdapat 2 variabel dengan nilai dimensi 1 positif dan dimensi 2 negatif. Lalu ada 4 variabel dengan nilai kedua dimensi positif.terdapat 5 variabel dengan nilai dimensi 1 negatif dan dimensi 2 positif. Serta ada 3 variabel dengan nilai kedua dimensi negatif. Yang berbeda dari plot CA sebelumnya adalah skala visualisasi warnanya.

corrplot(my\_CA\_Col$cos2, is.corr = FALSE)



1. Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai koordinasi baris tertinggi hingga terendah. Pada dimensi 1, yang terendah adalah Growth Ratedan yang tertinggi adalah Viability.
2. Sedangkan untuk kontribusinya, Estabilshment merupakan yang tertingga dan Seed Size merupakan yang terendah

fviz\_cos2(my\_CA, choice = "col")



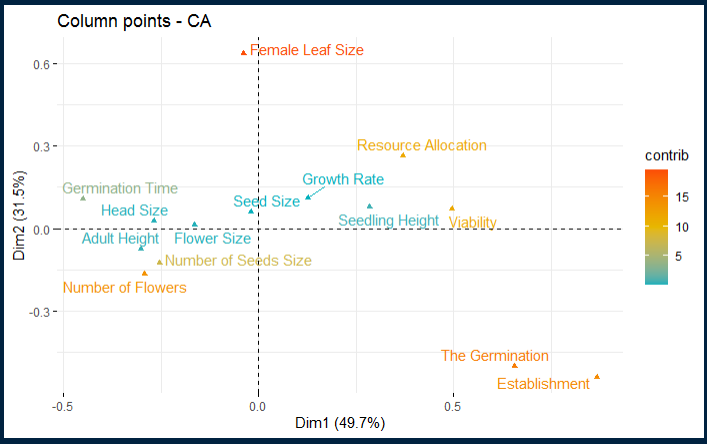
Dari grafik penyebaran dapat kita lihat bahwa tingkat nilai koordinasi kolom tersebar seperti anak tangga. Dan dari plot tersebut terlihat nilai cos2 memiliki penyebaran seperti anak tangga juga.

# Contrib

fviz\_ca\_col(my\_CA, col.col = "contrib",

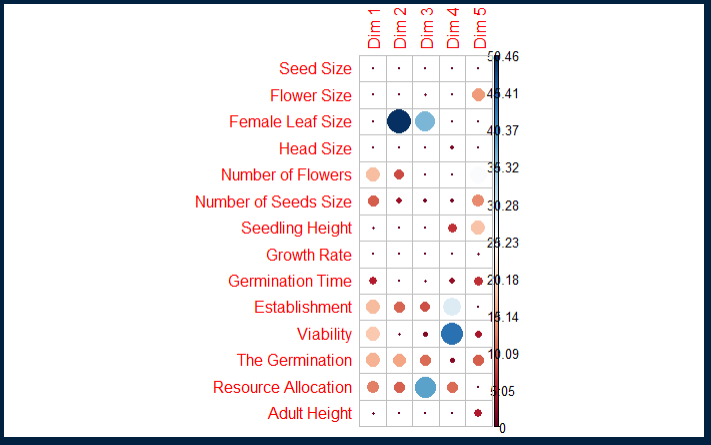
gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE)



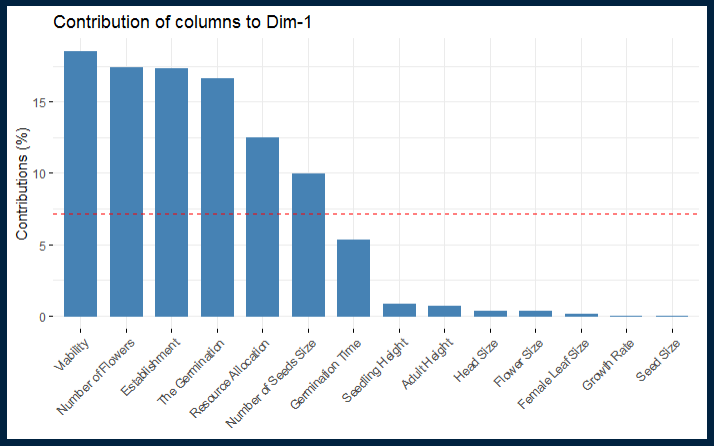
Dari plot CA di atas, data yang telah tersebar adalah 90,2% dan didapati bahwa terdapat 2 variabel dengan nilai dimensi 1 positif dan dimensi 2 negatif. Lalu ada 4 variabel dengan nilai kedua dimensi positif.terdapat 5 variabel dengan nilai dimensi 1 negatif dan dimensi 2 positif. Serta ada 3 variabel dengan nilai kedua dimensi negatif. Yang berbeda dari plot CA sebelumnya adalah skala visualisasi warnanya. Pada coord dan cos2, nilai tertinggi adalah 1 sedangkan pada contrib nilai tertingginya adalah 15-20.

corrplot(my\_CA\_Col$contrib, is.corr = FALSE)



1. Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai koordinasi baris tertinggi hingga terendah. Pada dimensi 1, yang terendah adalah Germination Time dan yang tertinggi adalah Viability.
2. Sedangkan untuk kontribusinya, The Germination merupakan yang tertingga danSeed Size dan Growth Rate merupakan yang terendah

fviz\_contrib(my\_CA, choice = "col")



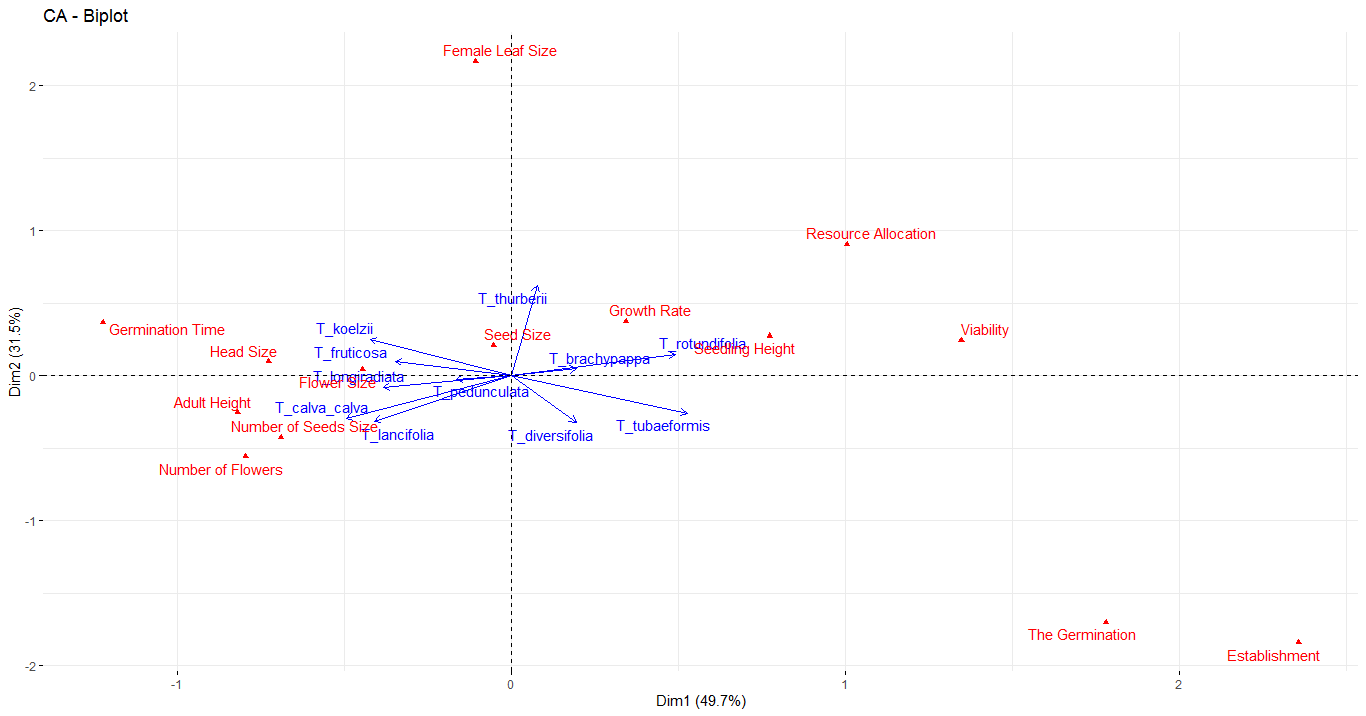
Dari grafik penyebaran dapat kita lihat bahwa tingkat nilai koordinasi kolom tersebar seperti anak tangga. Dan dari plot tersebut terlihat nilai contrib memiliki penyebaran seperti anak tangga juga. 6 variabel pertama memenuhi nilai dalam persebaran (sebagian besar data memiliki nilai yang cukup tinggi).

# Biplot Options

# Asymmetric Biplot

fviz\_ca\_biplot(my\_CA, map ="rowprincipal",

arrow = c(TRUE, FALSE), repel = TRUE)

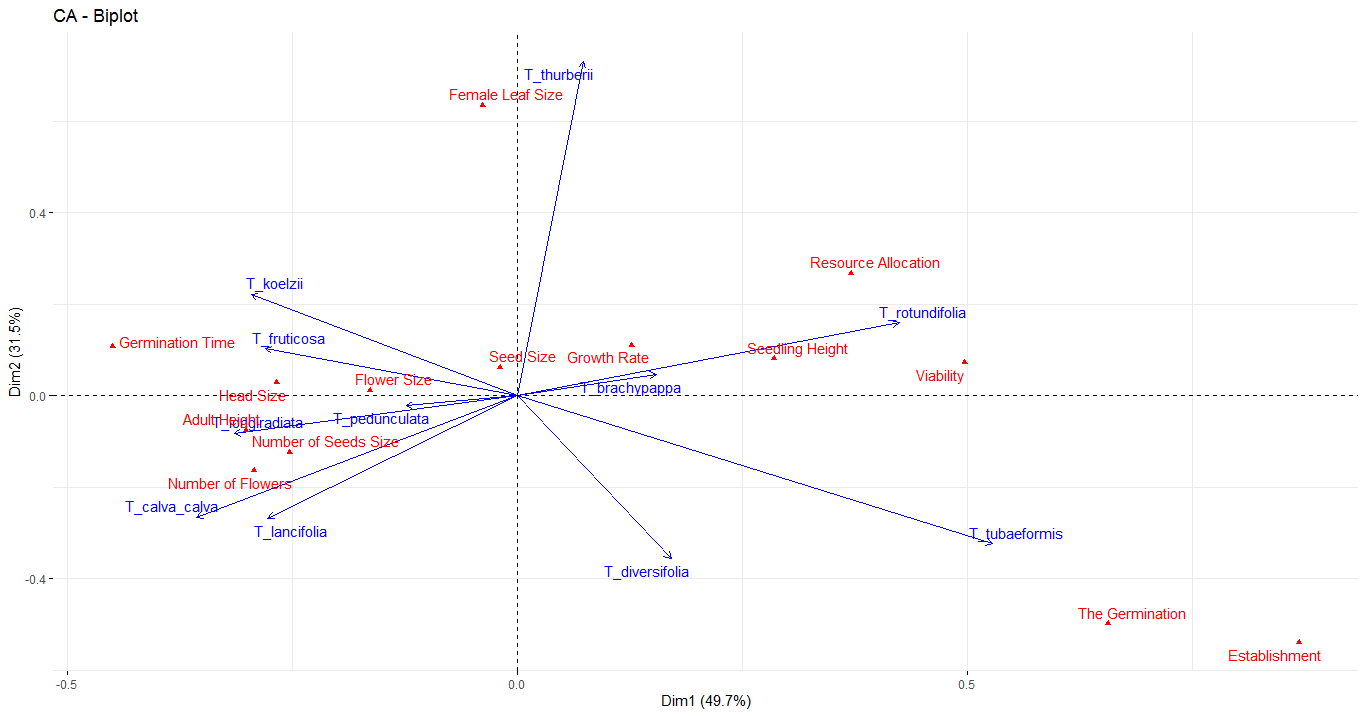


Biplot di atas menunjukkan kesesuaian/dominansi antara variabel dengan spesies. Anak panah yang ditunjukkan oleh garis biru menunjukkan spesies yang menuju pada variabel (merah).

# Contribution Biplot

fviz\_ca\_biplot(my\_CA, map ="colgreen",

arrow = c(TRUE, FALSE), repel = TRUE)



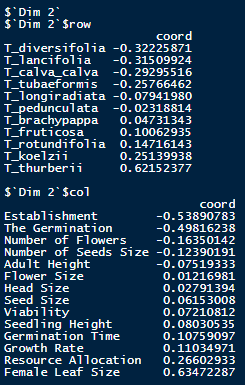
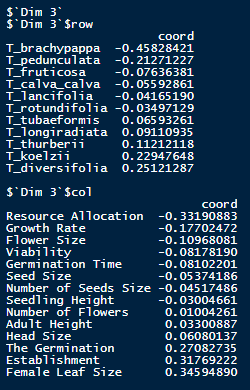
Biplot di atas menunjukkan kesesuaian/dominansi antara variabel dengan spesies. Anak panah yang ditunjukkan oleh garis biru menunjukkan spesies yang paling banyak berkontribusi (bernilai besar) pada variabel (merah).

# Dimension Description

CA\_Desc <- dimdesc(my\_CA)

CA\_Desc

```

**Kesimpulan**

Data *Tihtonia* dapat diolah menggunakan metode CA yang didasarkan pada variabel dan individual dari data. Data tersebut diolah sehingga dapat menghasilkan individual yang memiliki variabel (karakter) mencolok. Dan dari beberapa metode (coord, cos2, contrib) juga terdapat berbagai hasil.

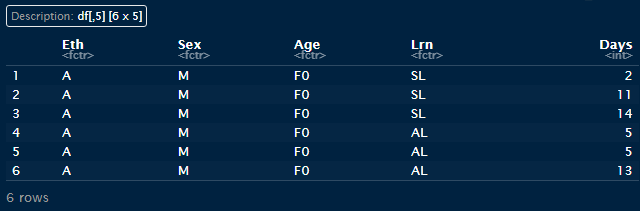
```{R}

# MCA

# Dataset Preparation

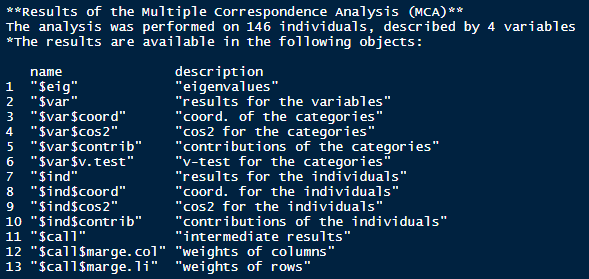
data(“quine”)

head(quine)



Data\_MCA <- MCA(quine[, 1:4], graph = FALSE)

Data\_MCA

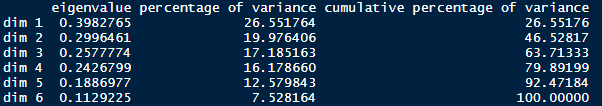


MCA digunakan untuk menganalisis kumpulan data dari survei. Tujuannya adalah untuk mengidentifikasi Sekelompok individu dengan profil yang sama. Dari output tersebut, bahwasanya Analisis dilakukan pada 146 individu, dijelaskan oleh 4 variabel.

Hasilnya tersedia di 13 objek

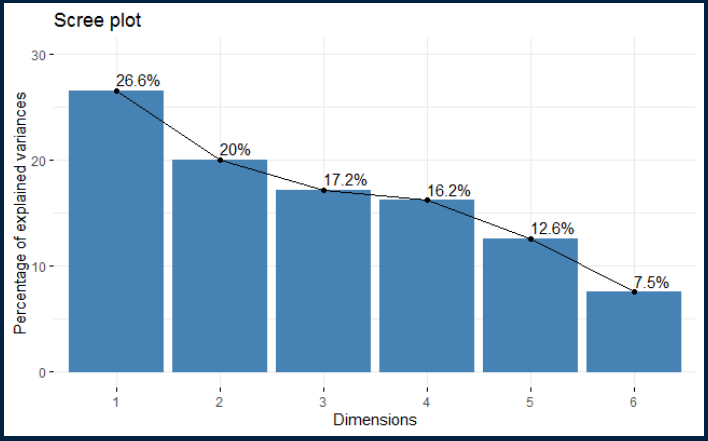
# Eigenvalues / Variances

Data\_MCA$eig



Didapatkan dari perhitungan tersebut didapat 6 nilai eigen yang diurutkan dari varians terbesar yaitu 26.551 sehingga pada dimensi 1 nilai eigennya adalah 0.392765 dan seterusnya.

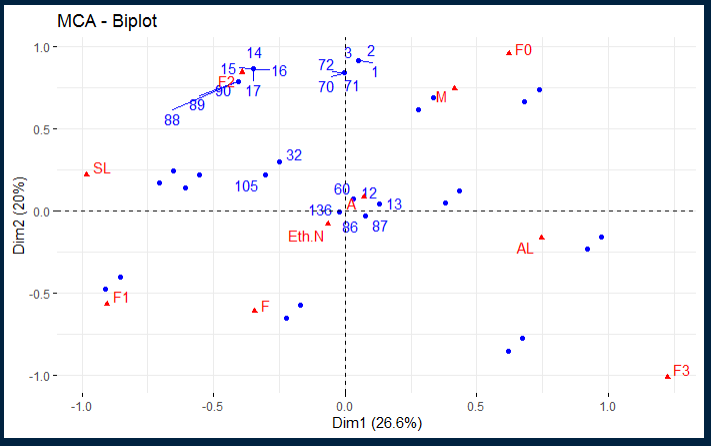
fviz\_screeplot(Data\_MCA, addlabels = TRUE, ylim = c(0, 30))



Screeplot di atas menunjukkan presentase nilai variance pada tiap dimensi. Dimensi 1 nilainya adalah 26,6% dan seterusnya.

# Biplot

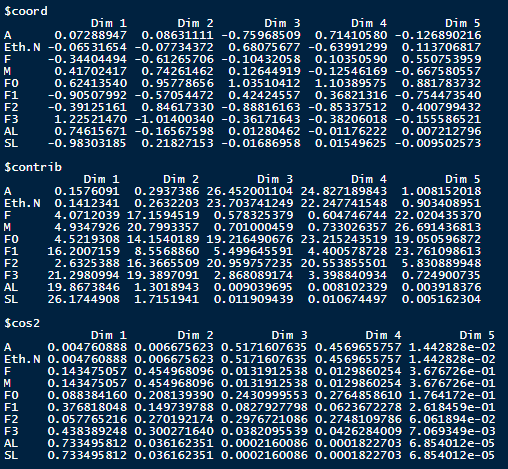
fviz\_mca\_biplot(Data\_MCA, repel = TRUE)

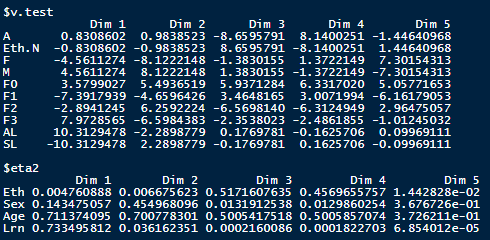


Pada biplot di atas, hanya ada 46,6% persebaran data. Faktor pada setiap variabel dijelaskan, sehingga variabelnya menjadi lebih banyak. AL (Lrn/status learner) dan F3 (age) memiliki nilai dimensi 1 positif dan dimensi 2 negatif dengan kurang lebih 4 titik dominan. A (Eth/ethnic), M (sex) dan F0 (age) memiliki nilai dimensi 1 dan 2 yang positif dengan kurang lebih 8 titik dominan. SL (Lrn) dan F2 (age) memiliki nilai dimensi 2 positif dan dimensi 1 negatif dengan kurang lebih 9 titik dominan. N (Eth), F (sex), dan F1 (age) memiliki nilai dimensi 1 dan dimensi 2 yang negative dengan kurang lebih 4 titik dominan. Setiap titik biru pada biplot di atas memungkinkan berisi banyak data.

# Graph of Variables

Data\_MCA$var





$coord merupakan nilai dari suatu koordinat untuk kategori

$cos2 memiliki nilai untuk kualitas representasi kategori

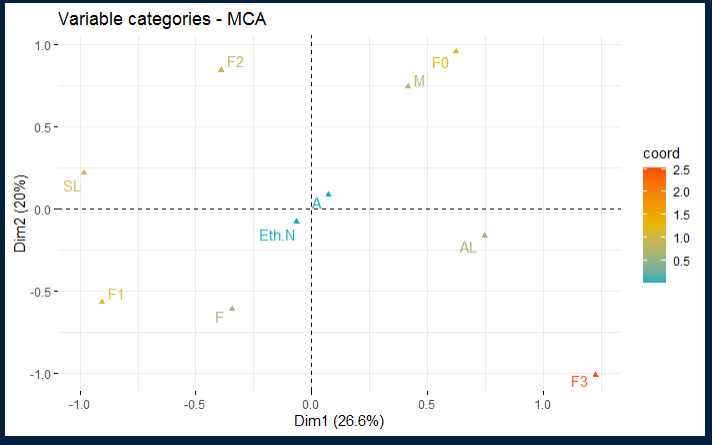
$contrib adalah nilai kontribusi kategori dalam persen untuk definisi dimensi

# Coord

fviz\_mca\_var(Data\_MCA, col.var = "coord",

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

repel = TRUE)



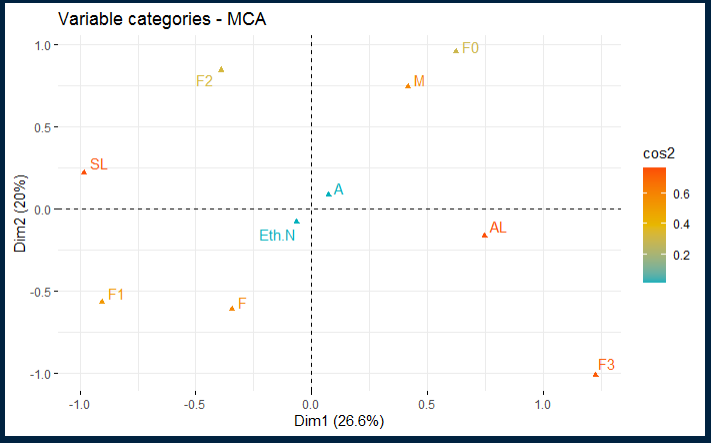
Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai koordinasi kategori terendah hingga tertinggi. Yang tertinggi adalah kategori umur F3 dan yang terendah adalah katergori etnis/suku bangsa.

# Cos2

fviz\_mca\_var(Data\_MCA, col.var = "cos2",

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

repel = TRUE)



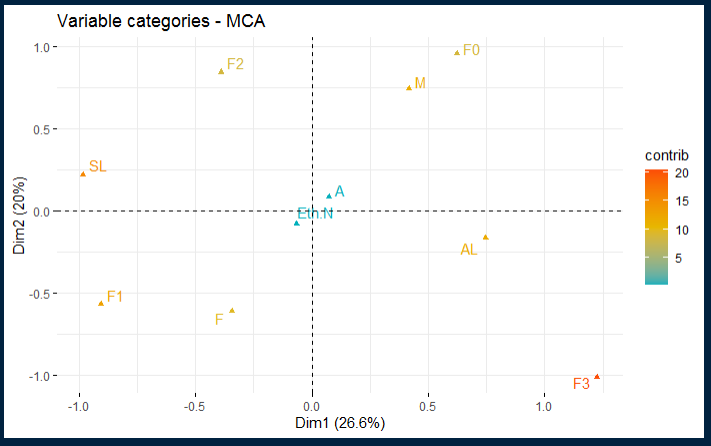
Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai koordinasi kategori terendah hingga tertinggi. Yang tertinggi adalah kategori status learner dan yang terendah adalah katergori etnis/suku bangsa.

# Contrib

fviz\_mca\_var(Data\_MCA, col.var = "contrib",

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

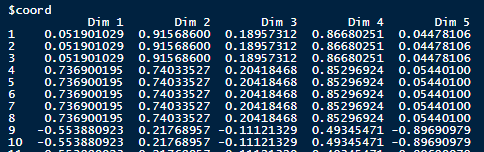
repel = TRUE)

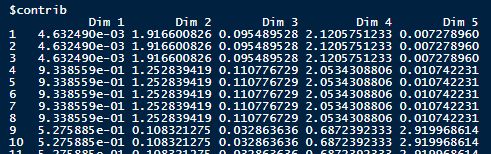


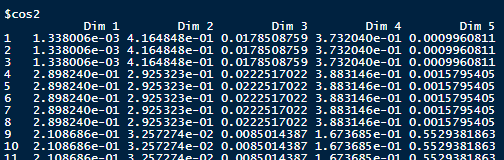
Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai koordinasi kategori terendah hingga tertinggi. Yang tertinggi adalah kategori umur F3 dan yang terendah adalah katergori etnis/suku bangsa.

# Graph of Individuals

Data\_MCA$ind





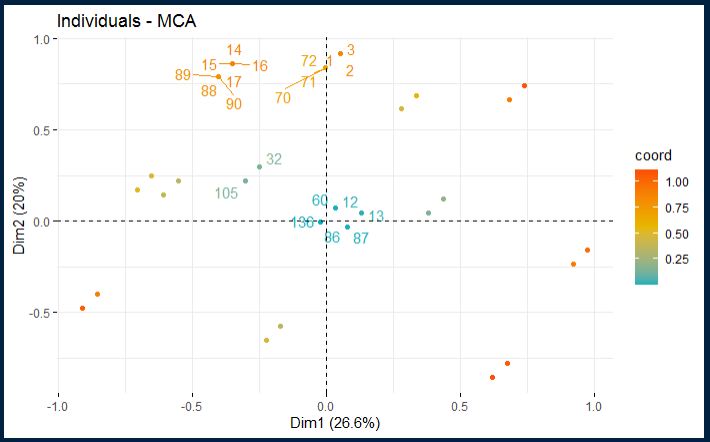


# Coord

fviz\_mca\_ind(Data\_MCA, col.ind = "coord",

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

repel = TRUE)



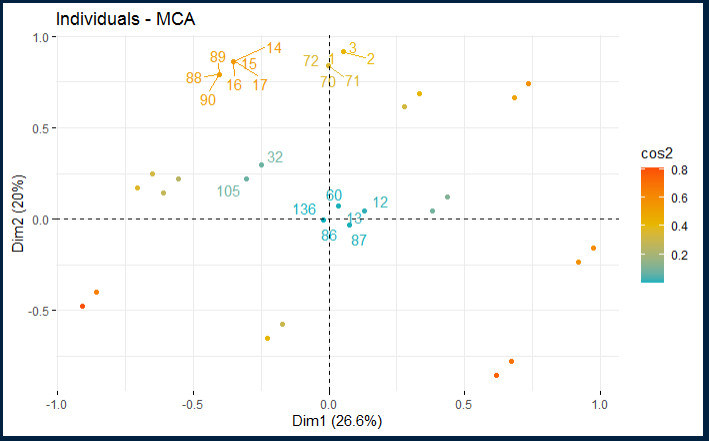
Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai koordinasi kategori terendah hingga tertinggi. Yang tertinggi berada jauh dari pusat dimensi, dan yang terendah berada di sekitar pusat dimensi seperti urutan ke-60, 12, 13, 87, dan lainnya.

# Cos2

fviz\_mca\_ind(Data\_MCA, col.ind = "cos2",

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

repel = TRUE)



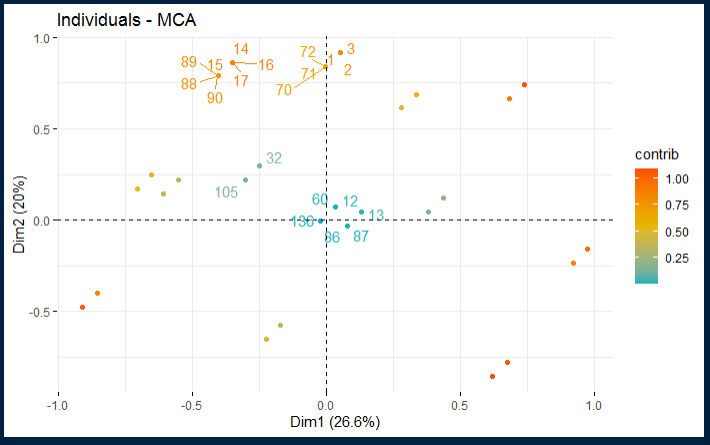
Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai koordinasi kategori terendah hingga tertinggi. Yang tertinggi berada jauh dari pusat dimensi, dan yang terendah berada di sekitar pusat dimensi seperti urutan ke-60, 12, 13, 87, dan lainnya.

# Contrib

fviz\_mca\_ind(Data\_MCA, col.ind = "contrib",

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

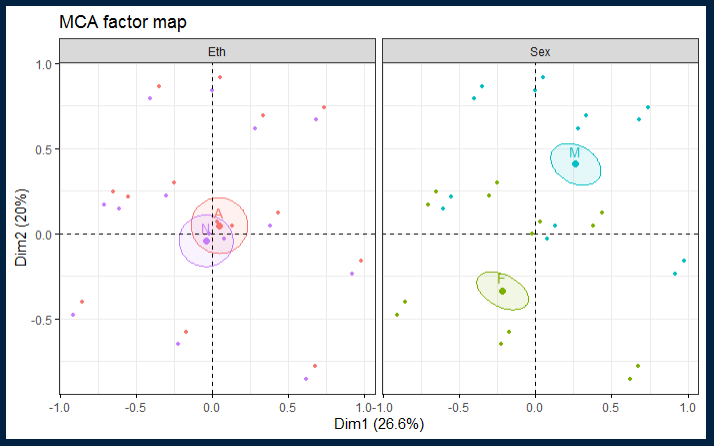
repel = TRUE)



Dari grafik penyebaran dapat kita lihat bahwa gradasi warna yang dimulai dari biru hingga kemerahan menandakan tingkat nilai kontribusi kategori terendah hingga tertinggi. Yang tertinggi berada jauh dari pusat dimensi, dan yang terendah berada di sekitar pusat dimensi seperti urutan ke-60, 12, 13, 87, dan lainnya.

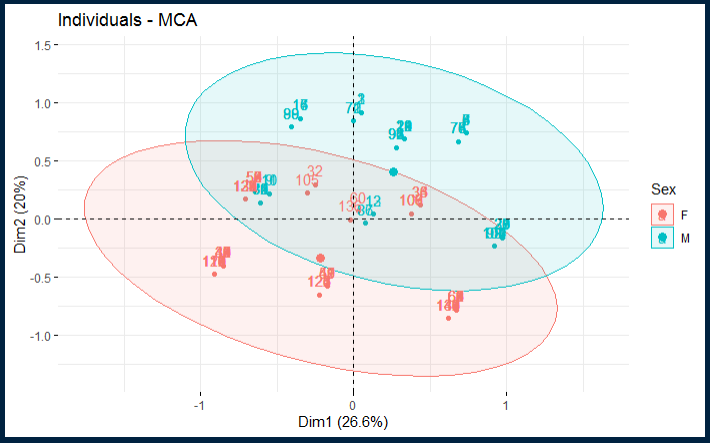
# Color Individuals by Groups

fviz\_ellipses(Data\_MCA, c("Eth", "Sex"), geom = "point")



Dari plot di atas, dapat diperoleh informasi mengenai pewarnaan data sesuai dengan faktornya. Pada variabel Eth terdapat 2 faktor yaitu A (Aborigin) dan N (Non-Aborigin). Lalu pada variabel sex terdapat 2 faktor juga yatu M (Male) dan F (Female)

fviz\_mca\_ind(Data\_MCA, habillage = 2, addEllipses = TRUE)

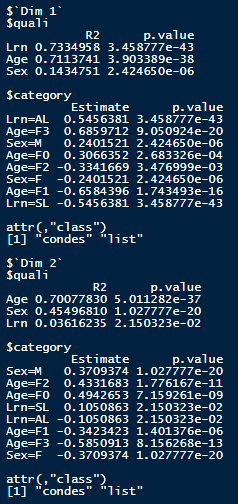
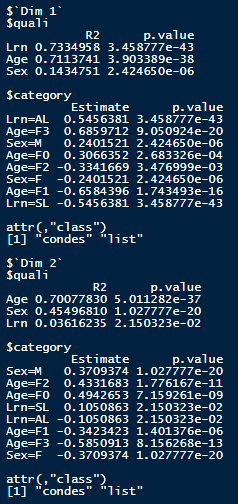


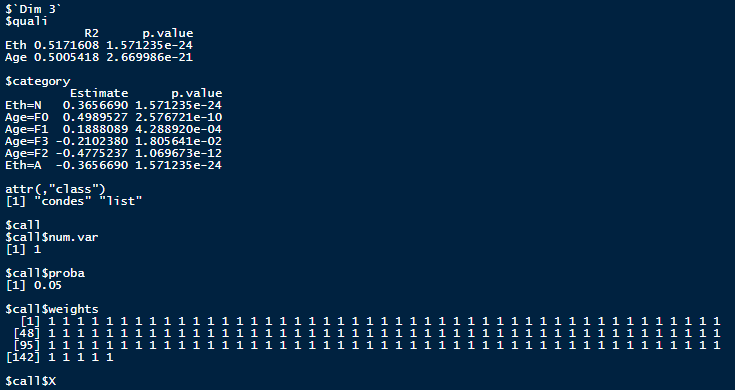
Berdasarkan plot diatas mengenai jenis kelaminnya, terbagi menjadi 2 kelompok yaitu Male dan Female. Dan dapat dilihat dari plot tersebut keduanya memiliki kelompok penyebaran yang hamper sama banyaknya.

# Habillage = Index of The Column to be Used as Grouping variable

# dimension Description

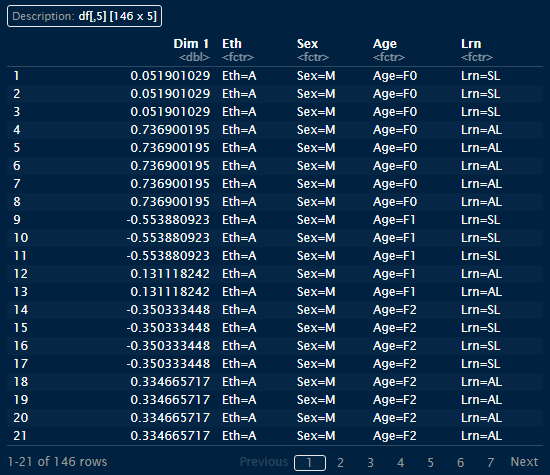
MCA\_Desc <- dimdesc(Data\_MCA)



MCA\_Desc

```



**Kesimpulan**

Data *quine* yang diolah menggunakan MCA menghasilkan plot yang membaginya berdasarkan faktor dari variabel. Data tersebut juga dapat diketahui memiliki kesamaan antar isinya. Selain itu, penggunaan MCA juga dapat membuat data lebih ringkas untuk divisualisasikan.