

## PCA\_mtcars

Ana Laguna Pradas

```
library(datasets)
library(ggplot2)
library(FactoMineR)
library(scales)
library(rgl)
library(knitr)
library(scatterplot3d)

#Principal Component Analysis and Hierarchical Clustering

# cor = TRUE indicates that PCA is performed on
# standardized data (mean = 0, variance = 1)
pcaCars <- princomp(mtcars, cor = TRUE)

# view objects stored in pcaCars
names(pcaCars)

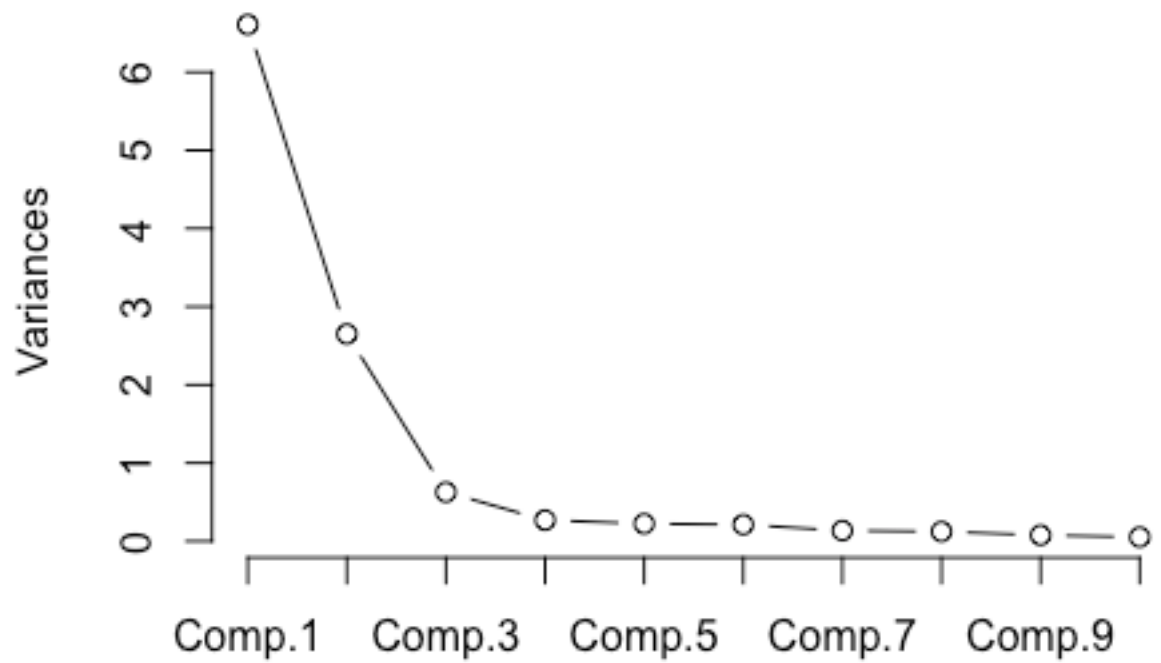
## [1] "sdev"      "loadings" "center"    "scale"     "n.obs"     "scores"
## [7] "call"

# proportion of variance explained
summary(pcaCars)

## Importance of components:
##               Comp.1   Comp.2   Comp.3   Comp.4
## Standard deviation  2.5706809 1.6280258 0.79195787 0.51922773
## Proportion of Variance 0.6007637 0.2409516 0.05701793 0.02450886
## Cumulative Proportion 0.6007637 0.8417153 0.89873322 0.92324208
##               Comp.5   Comp.6   Comp.7   Comp.8
## Standard deviation  0.47270615 0.45999578 0.36777981 0.35057301
## Proportion of Variance 0.02031374 0.01923601 0.01229654 0.01117286
## Cumulative Proportion 0.94355581 0.96279183 0.97508837 0.98626123
##               Comp.9   Comp.10   Comp.11
## Standard deviation  0.277572792 0.228112781 0.148473587
## Proportion of Variance 0.007004241 0.004730495 0.002004037
## Cumulative Proportion 0.993265468 0.997995963 1.000000000

# scree plot
plot(pcaCars, type = "l")
```

## pcaCars



```
# cluster cars
carsHC <- hclust(dist(pcaCars$scores), method = "ward.D2")

# dendrogram
plot(carsHC)
```

## Cluster Dendrogram



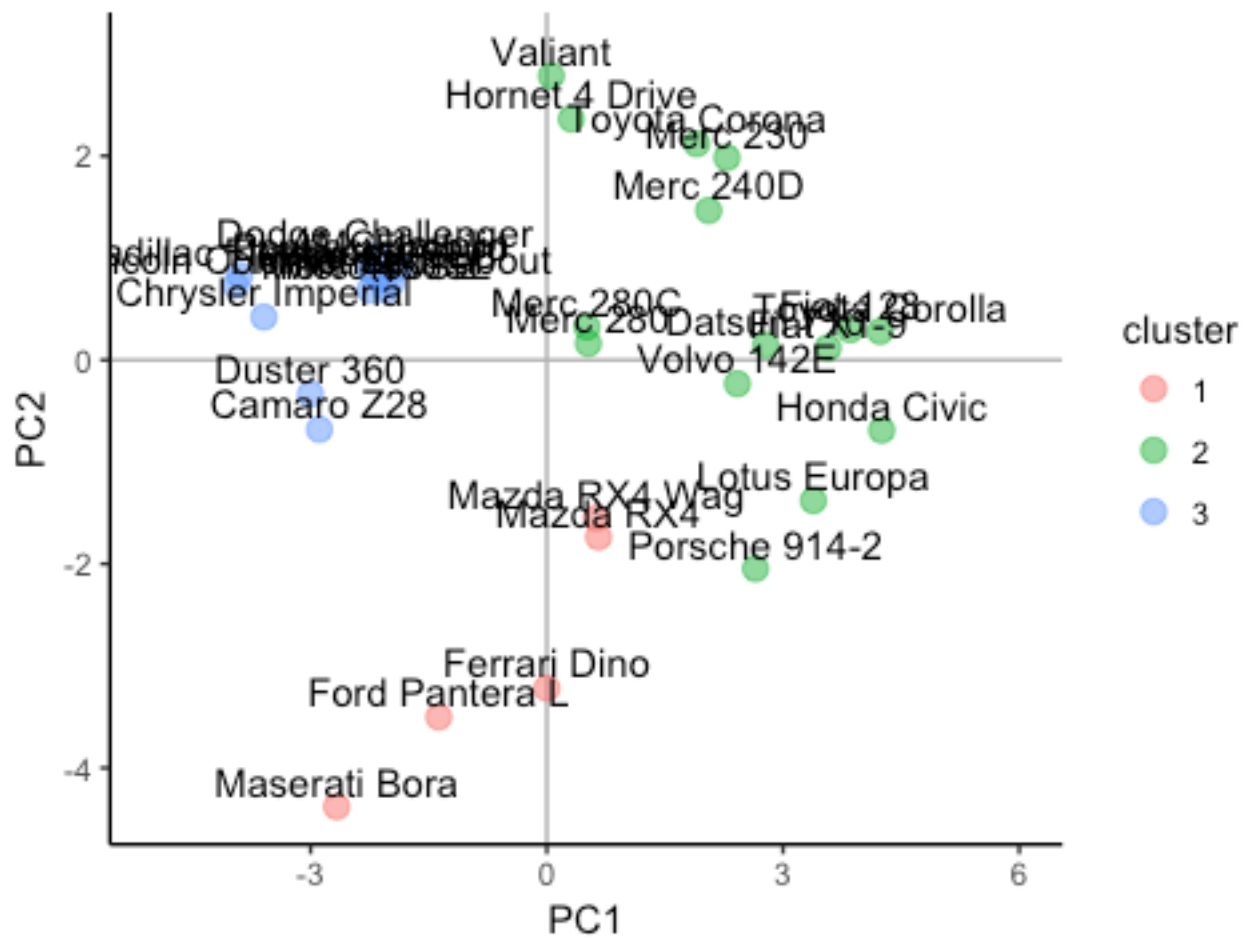
```
dist(pcaCars$scores)
hclust(*, "ward.D2")
```

```
# cut the dendrogram into 3 clusters
carsClusters <- cutree(carsHC, k = 3)

# add cluster to data frame of scores
carsDf <- data.frame(pcaCars$scores, "cluster" = factor(carsClusters))
carsDf <- transform(carsDf, cluster_name =
  paste("Cluster", carsClusters))
#First figure using ggplot2
library(ggplot2)
p1 <- ggplot(carsDf, aes(x=Comp.1, y=Comp.2)) +
  theme_classic() +
  geom_hline(yintercept = 0, color = "gray70") +
  geom_vline(xintercept = 0, color = "gray70") +
  geom_point(aes(color = cluster), alpha = 0.55, size = 3) +
  xlab("PC1") +
  ylab("PC2") +
  xlim(-5, 6) +
  ggtitle("PCA Clusters from Hierarchical Clustering of Cars
Data")

p1 + geom_text(aes(y = Comp.2 + 0.25, label = rownames(carsDf)))
```

## PCA Clusters from Hierarchical Clustering of Cars Data



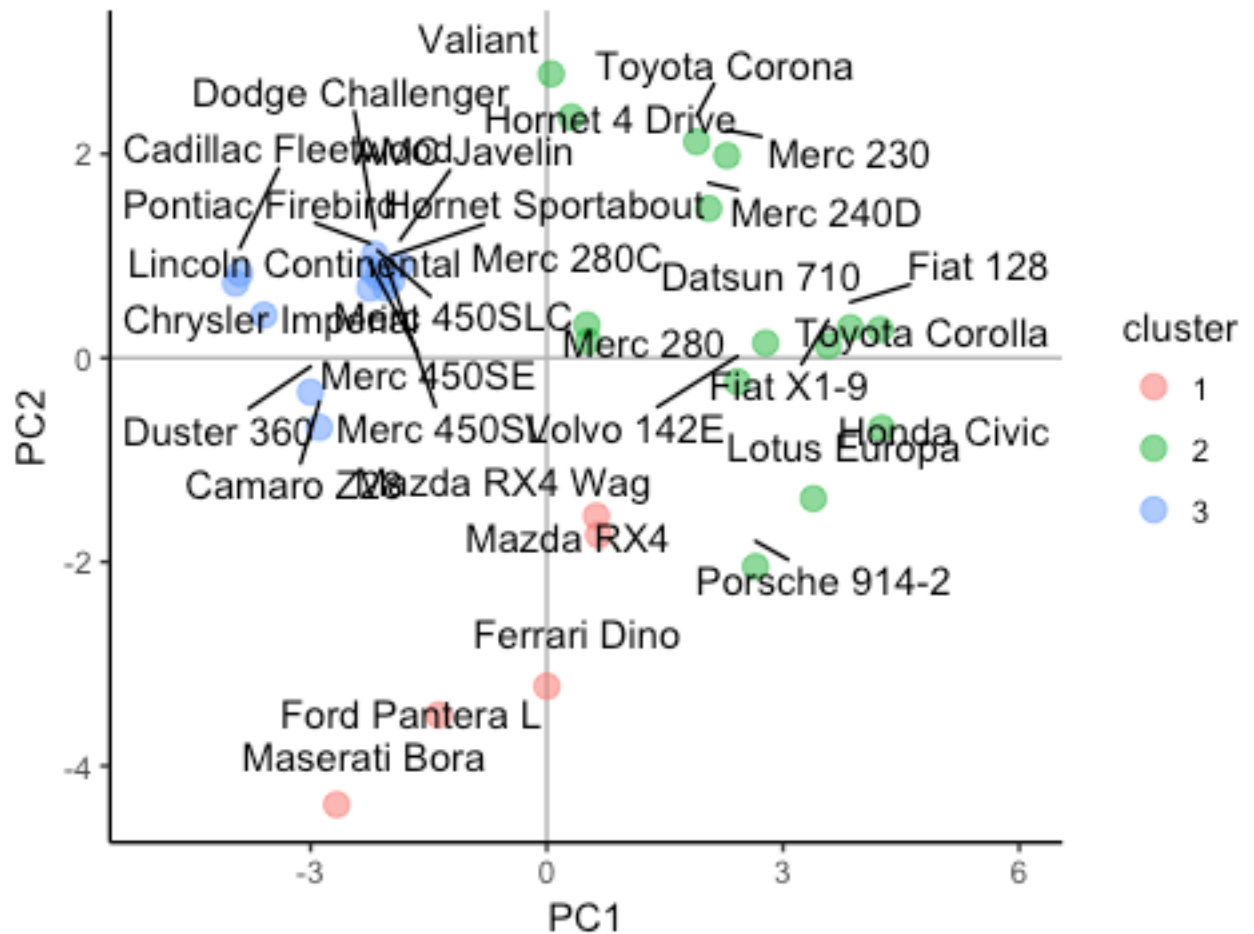
*#Second figure using ggplot2 with ggrepel*

```
library(ggplot2)
```

```
library(ggrepel)
```

```
p1 + geom_text_repel(aes(y = Comp.2 + 0.25, label = rownames(carsDf)))
```

## PCA Clusters from Hierarchical Clustering of Cars Data



```
#Interactive plot using plotly
#library(plotly)
#p <- plot_ly(carsDf, x = Comp.1 , y = Comp.2, text =
rownames(carsDf),
#           mode = "markers", color = cluster_name, marker =
List(size = 11))

#p <- layout(p, title = "PCA Clusters from Hierarchical Clustering of
Cars Data",
#           xaxis = list(title = "PC 1"),
#           yaxis = list(title = "PC 2"))

#p

# Run Principal Component on the data
PC_res = PCA(mtcars, scale.unit=TRUE, ncp = dim(mtcars)[2],
graph=FALSE)
summary(PC_res)

##
## Call:
## PCA(X = mtcars, scale.unit = TRUE, ncp = dim(mtcars)[2], graph =
```

```

FALSE)
##
##
## Eigenvalues
##          Dim.1   Dim.2   Dim.3   Dim.4   Dim.5
Dim.6
## Variance          6.608   2.650   0.627   0.270   0.223
0.212
## % of var.         60.076  24.095   5.702   2.451   2.031
1.924
## Cumulative % of var. 60.076  84.172  89.873  92.324  94.356
96.279
##          Dim.7   Dim.8   Dim.9   Dim.10   Dim.11
## Variance          0.135   0.123   0.077   0.052   0.022
## % of var.          1.230   1.117   0.700   0.473   0.200
## Cumulative % of var. 97.509  98.626  99.327  99.800 100.000
##
## Individuals (the 10 first)
##          Dist    Dim.1    ctr    cos2    Dim.2    ctr
cos2
## Mazda RX4          |  2.234 | -0.657  0.204  0.087 |  1.735  3.551
0.604 |
## Mazda RX4 Wag      |  2.081 | -0.629  0.187  0.091 |  1.550  2.833
0.555 |
## Datsun 710          |  2.987 | -2.779  3.653  0.866 | -0.146  0.025
0.002 |
## Hornet 4 Drive      |  2.521 | -0.312  0.046  0.015 | -2.363  6.584
0.879 |
## Hornet Sportabout   |  2.456 |  1.974  1.844  0.646 | -0.754  0.671
0.094 |
## Valiant             |  3.014 | -0.056  0.001  0.000 | -2.786  9.151
0.855 |
## Duster 360          |  3.187 |  3.003  4.264  0.888 |  0.335  0.132
0.011 |
## Merc 240D           |  2.841 | -2.055  1.998  0.523 | -1.465  2.531
0.266 |
## Merc 230            |  3.733 | -2.287  2.474  0.375 | -1.984  4.639
0.282 |
## Merc 280            |  1.907 | -0.526  0.131  0.076 | -0.162  0.031
0.007 |
##          Dim.3    ctr    cos2
## Mazda RX4        -0.601  1.801  0.072 |
## Mazda RX4 Wag     -0.382  0.728  0.034 |
## Datsun 710         -0.241  0.290  0.007 |
## Hornet 4 Drive     -0.136  0.092  0.003 |
## Hornet Sportabout -1.134  6.412  0.213 |
## Valiant            0.164  0.134  0.003 |
## Duster 360         -0.363  0.656  0.013 |
## Merc 240D          0.944  4.439  0.110 |
## Merc 230           1.797 16.094  0.232 |
## Merc 280           1.493 11.103  0.613 |
##
## Variables (the 10 first)
##          Dim.1    ctr    cos2    Dim.2    ctr    cos2

```

```

Dim.3
## mpg          | -0.932 13.143  0.869 |  0.026  0.026  0.001 | -
0.179
## cyl          |  0.961 13.981  0.924 |  0.071  0.191  0.005 | -
0.139
## disp         |  0.946 13.556  0.896 | -0.080  0.243  0.006 | -
0.049
## hp           |  0.848 10.894  0.720 |  0.405  6.189  0.164 |
0.111
## drat         | -0.756  8.653  0.572 |  0.447  7.546  0.200 |
0.128
## wt           |  0.890 11.979  0.792 | -0.233  2.046  0.054 |
0.271
## qsec         | -0.515  4.018  0.266 | -0.754 21.472  0.569 |
0.319
## vs           | -0.788  9.395  0.621 | -0.377  5.366  0.142 |
0.340
## am           | -0.604  5.520  0.365 |  0.699 18.440  0.489 | -
0.163
## gear         | -0.532  4.281  0.283 |  0.753 21.377  0.567 |
0.229
##              ctr    cos2
## mpg          5.096  0.032 |
## cyl          3.073  0.019 |
## disp         0.378  0.002 |
## hp           1.960  0.012 |
## drat         2.598  0.016 |
## wt          11.684  0.073 |
## qsec        16.255  0.102 |
## vs          18.388  0.115 |
## am           4.234  0.027 |
## gear         8.397  0.053 |

```

*# Plot the results of the two first Principal Components*

```

biplot(PC_res$ind$coord, PC_res$var$coord, scale=0, cex=0.7,
main="Biplot for the first two Principal Components", xlab = "First
Principal Component (explains ~60%)", ylab="Second Principal Component
(explains ~24%)")

```

```
## Warning in plot.window(...): "scale" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "scale" is not a graphical
parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...):
"scale" is not
## a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...):
"scale" is not
## a graphical parameter
```

```
## Warning in box(...): "scale" is not a graphical parameter
```

```
## Warning in title(...): "scale" is not a graphical parameter
```

```
## Warning in text.default(x, xlabs, cex = cex[1L], col = col[1L],
...):
## "scale" is not a graphical parameter
## Warning in plot.window(...): "scale" is not a graphical parameter
## Warning in plot.xy(xy, type, ...): "scale" is not a graphical
parameter
## Warning in title(...): "scale" is not a graphical parameter
## Warning in axis(3, col = col[2L], ...): "scale" is not a graphical
## parameter
## Warning in axis(4, col = col[2L], ...): "scale" is not a graphical
## parameter
## Warning in text.default(y, labels = ylabs, cex = cex[2L], col =
col[2L], :
## "scale" is not a graphical parameter
```

## Biplot for the first two Principal Components

