

PCA-Iris

Max Pereira

14/06/2020

Carregando os pacotes

```
library(FactoMineR)
```

```
## Warning: package 'FactoMineR' was built under R version 3.6.3
```

```
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 3.6.3
```

```
## Loading required package: ggplot2
```

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

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

Usando do dataset Iris

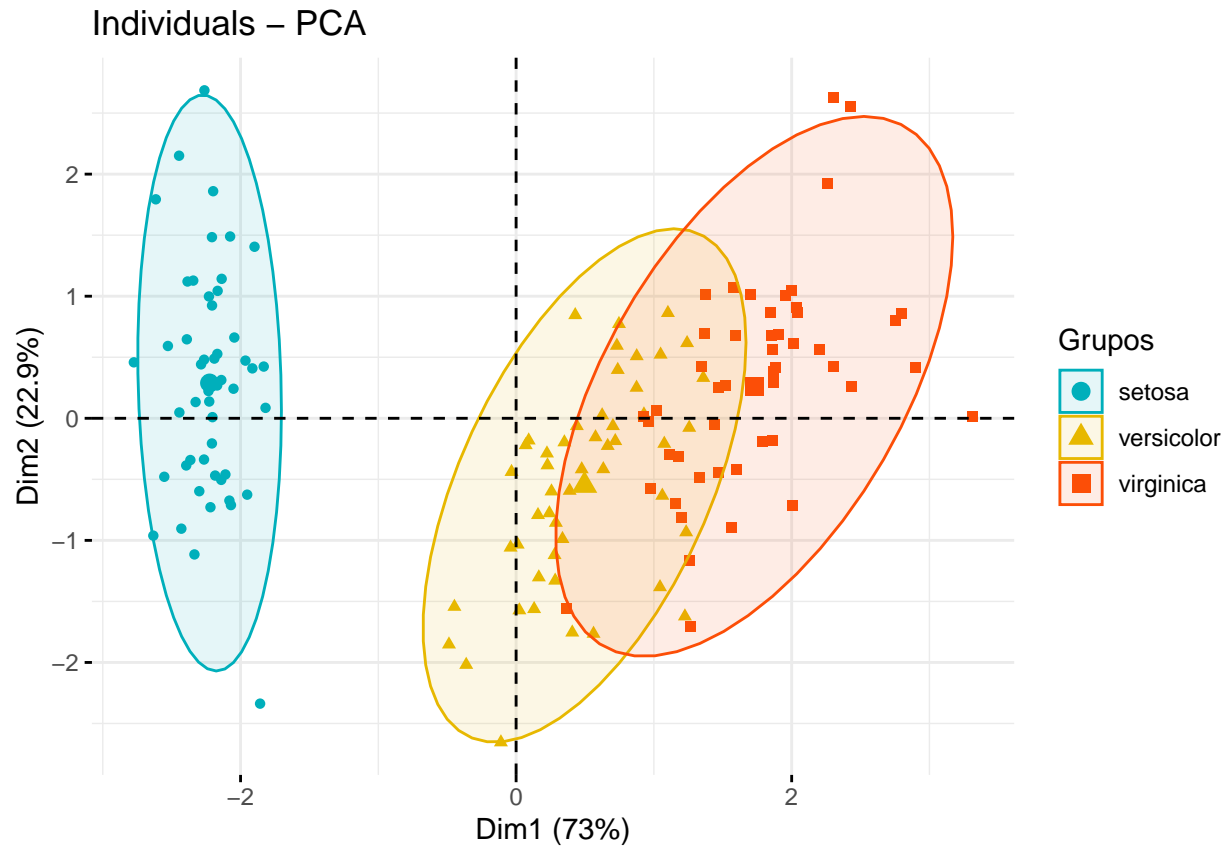
Colorindo os indivíduos por grupo. A coluna “Species” será usada para fazer o agrupamento. A variável Species (index = 5) é removida antes de executar a análise PCA

```
iris.pca <- PCA(iris[,-5], graph=FALSE)
```

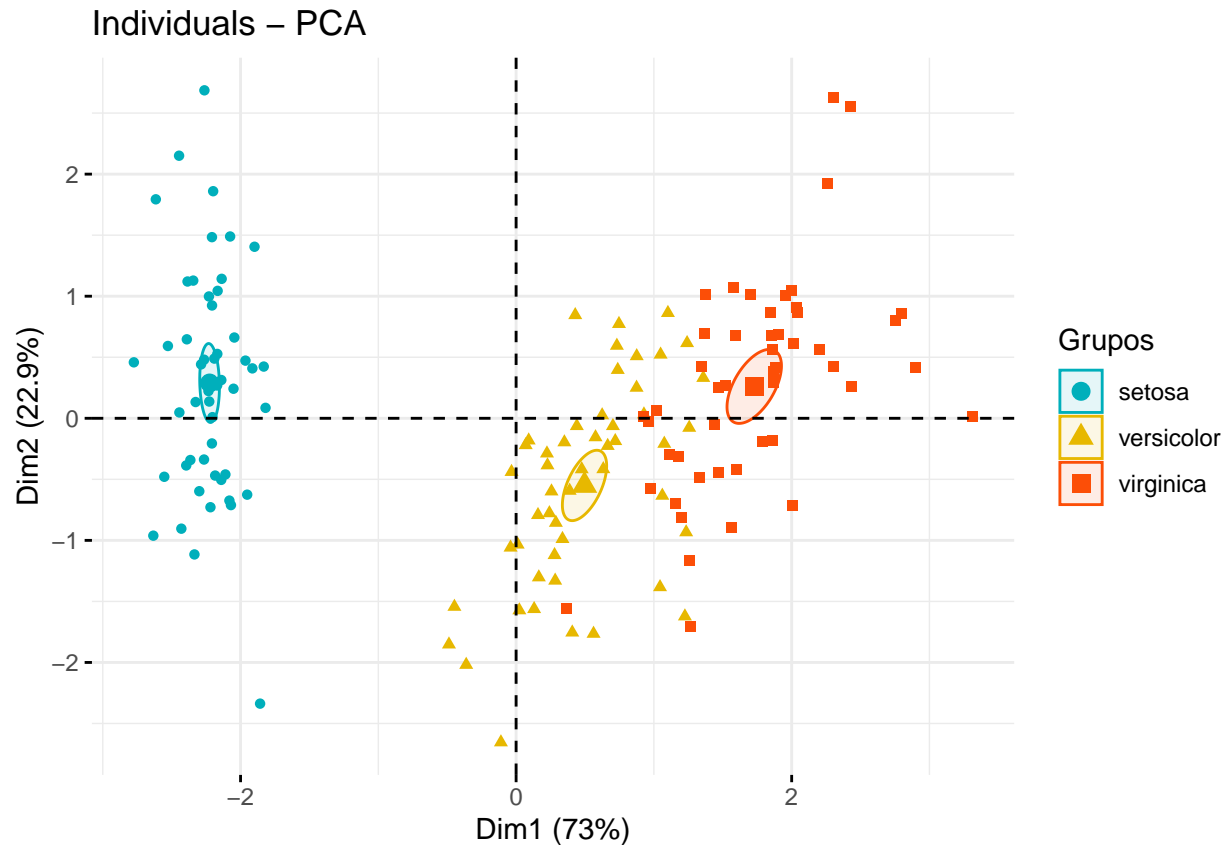
Gráficos - Agrupamento

O parâmetro habillage ou col.ind pode ser usado para especificar a variável fator usada para colorir os indivíduos por grupos. Para adicionar elipses ao redor de cada grupo, especificar o parâmetro addEllipses = TRUE.

```
fviz_pca_ind(iris.pca,
  geom.ind = "point",
  col.ind = iris$Species,
  palette = c("#00AFBB", "#E7B800", "#FC4E07"),
  addEllipses = TRUE,
  legend.title = "Grupos")
```



```
fviz_pca_ind(iris.pca, geom.ind = "point", col.ind = iris$Species,  
  palette = c("#00AFBB", "#E7B800", "#FC4E07"),  
  addEllipses = TRUE, ellipse.type = "confidence",  
  legend.title = "Grupos")
```

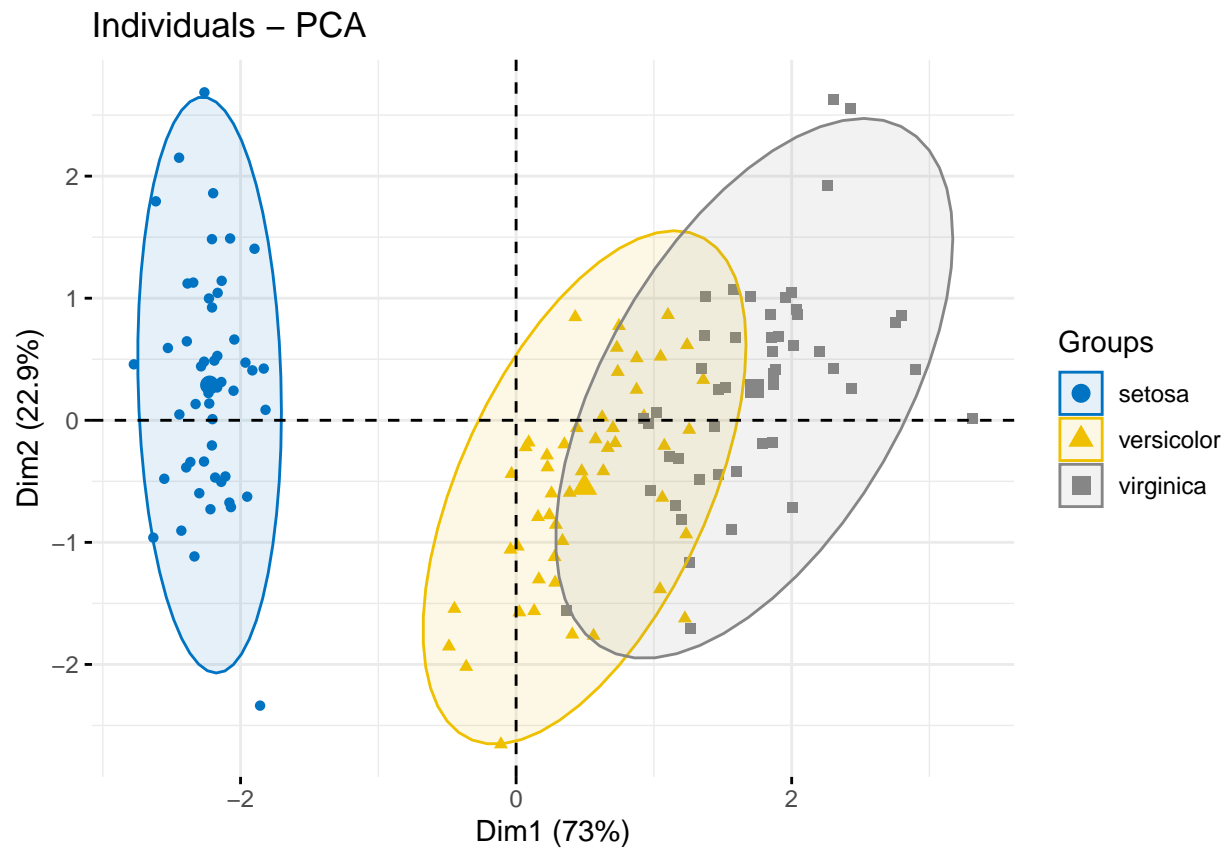


Os valores permitidos para o parâmetro `palette` são:

“grey” para paleta de cores cinzas; paletas de cores “brewer”: “RdBu”, “Blues”. Para ver todas: `RColorBrewer::display.brewer.all()` paletas customizadas: `c(“blue”, “red”)` paletas de jornais científicos do pacote “ggsci”: “npg”, “aaas”, “lancet”, “jco”, “ucscgb”, “uchicago”, “simpsons” and “rickandmorty”.

Por exemplo, para usar o jco (journal of clinical oncology)

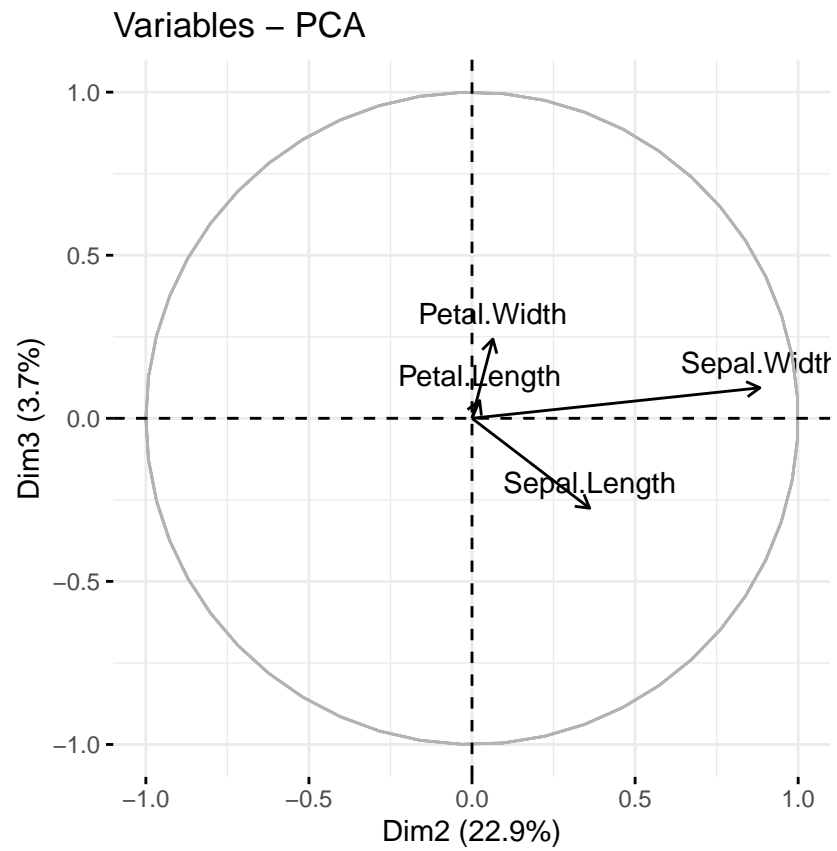
```
library(ggsci)
fviz_pca_ind(iris.pca,
  label = "none",
  habillage = iris$Species,
  addEllipses = TRUE,
  palette = "jco")
```



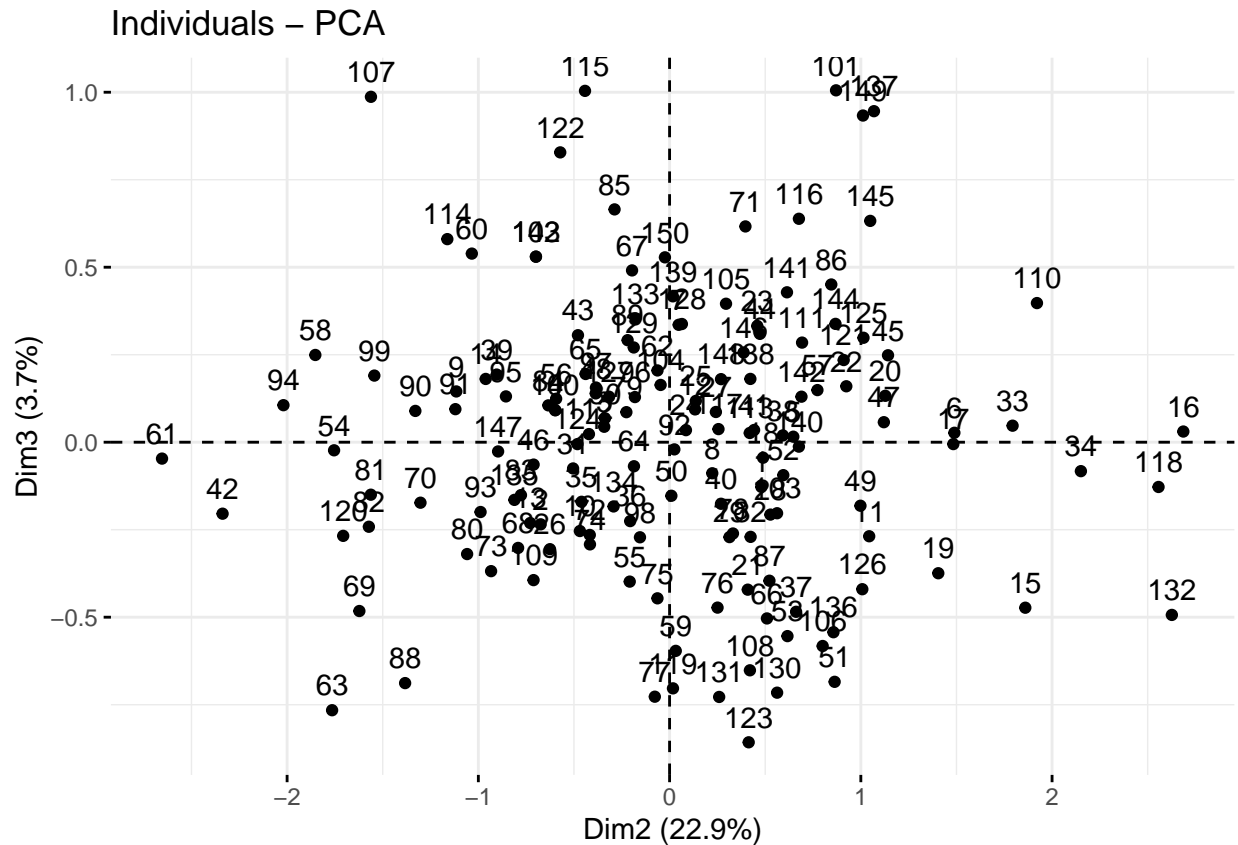
Dimensões

Por default, as variáveis/indivíduos são representados nas dimensões 1 e 2. Para visualizar as dimensões 2 e 3, por exemplo, deve ser especificado o parâmetro `axes = c(2,3)`

```
fviz_pca_var(iris.pca, axes = c(2,3))
```



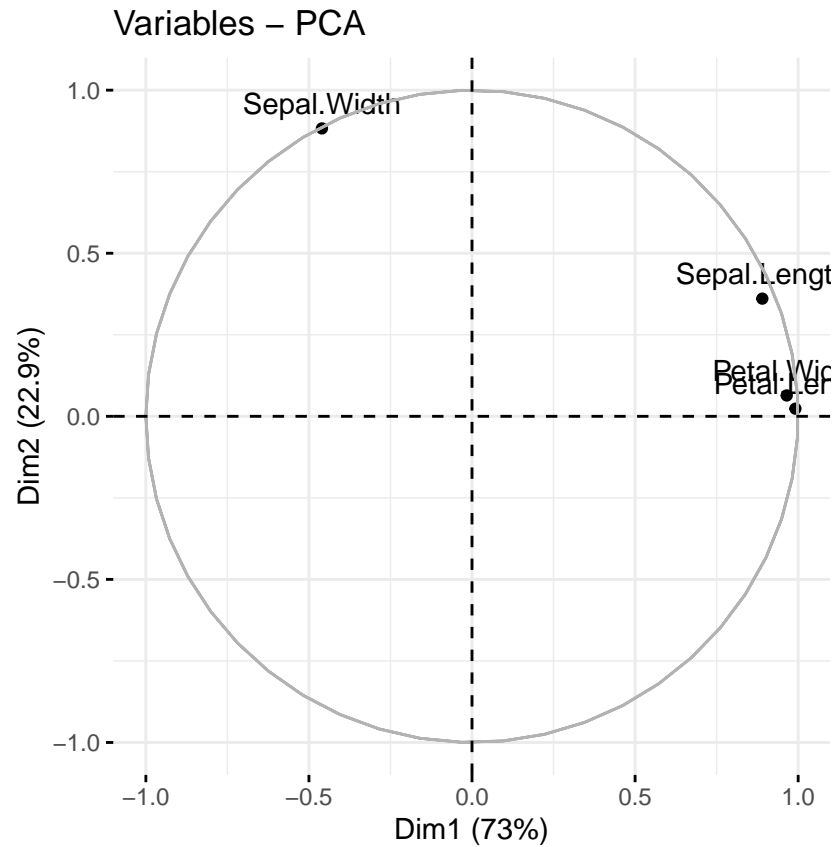
```
fviz_pca_ind(iris.pca, axes = c(2,3))
```



Elementos gráficos: point, text, arrow

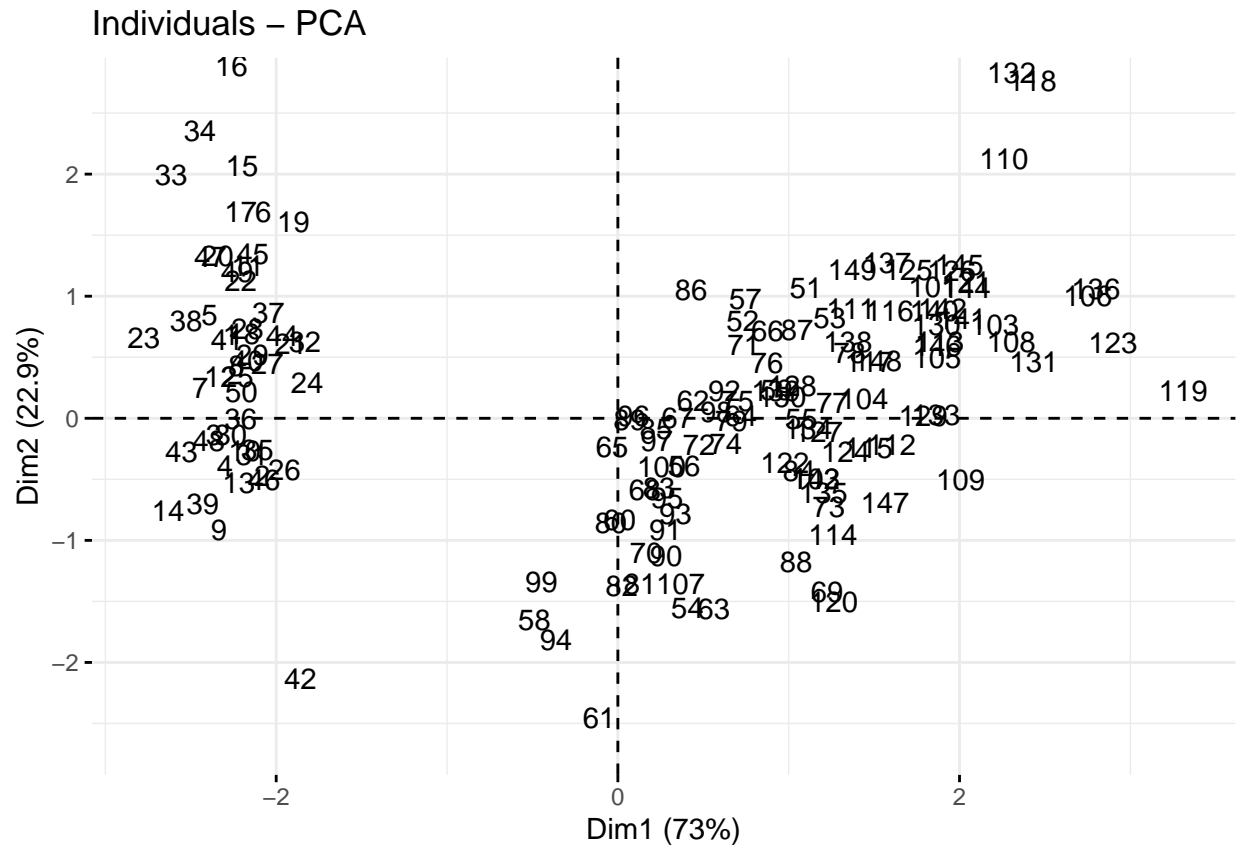
O parâmetro “geom” é usado para especificar os elementos geométricos ou elementos gráficos. geom.var: um texto para especificar a geometria a ser usada nas variáveis. geom.var = “point”: para mostrar apenas pontos geom.var = “text”: para mostrar apenas os textos das etiquetas (labels) geom.var = c(“point”, “text”): para mostrar pontos e textos dos labels geom.var = c(“arrow”, “text”): para mostrar setas e textos (default)

```
fviz_pca_var(iris.pca, geom.var = c("point", "text"))
```



geom.ind : um texto para especificar a geometria a ser usada nos individuos. geom.ind = "point" : para mostrar apenas pontos geom.ind = "text" : para mostrar apenas os textos das etiquetas (labels) geom.ind = c("point", "text") : para mostrar pontos e textos dos labels

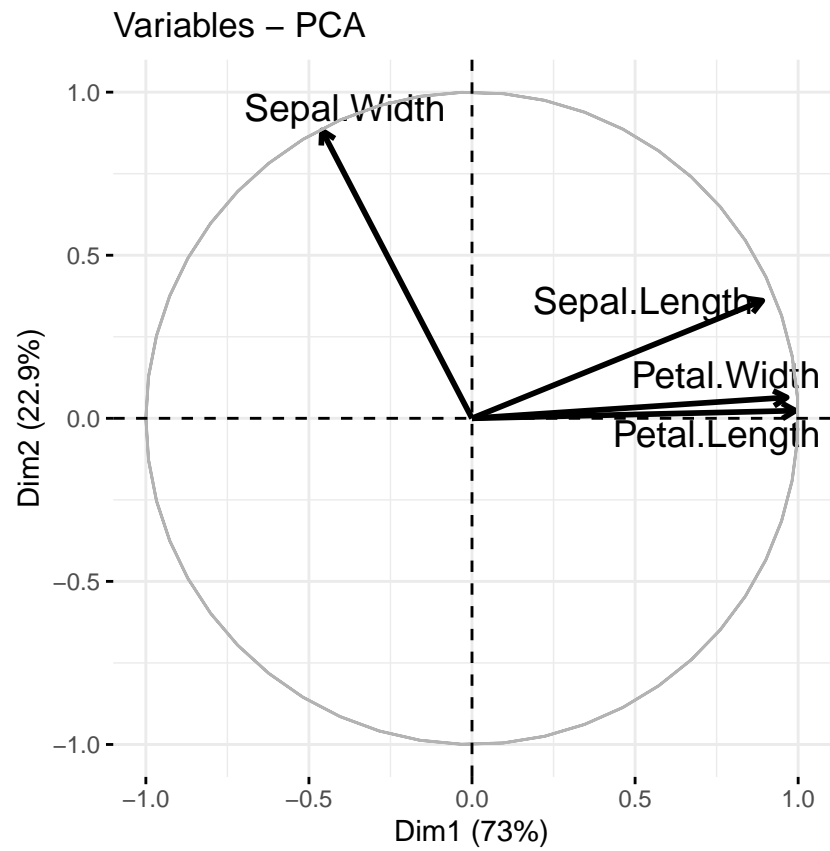
```
fviz_pca_ind(iris.pca, geom.ind = "text")
```



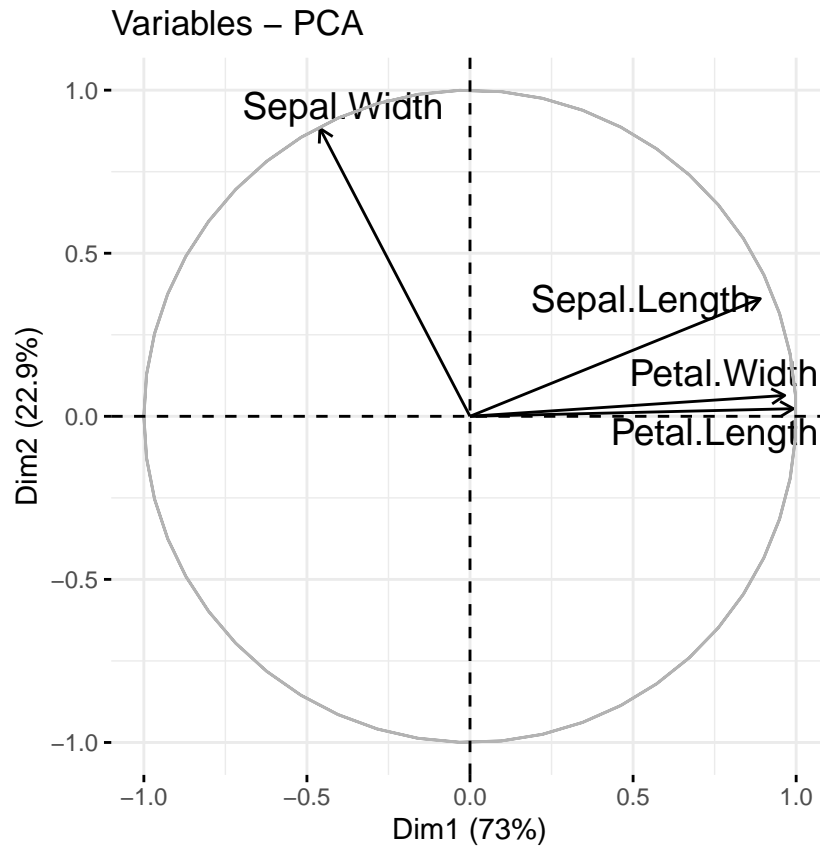
Tamanho e forma dos elementos gráficos

labelsize : tamanho da fonte dos textos, exemplo: labelsize = 4. pointsize : tamanho dos pontos, exemplo: pointsize = 1.5 arrowsize : tamanho das setas (espessura), exemplo: arrowsize = 0.5 ointshape : o formato dos pontos, exemplo : pointshape = 21. ggpubr::show_point_shapes() para ver os formatos disponíveis.

```
fviz_pca_var(iris.pca, arrowsize = 1, labelsize = 5, repel = TRUE)
```

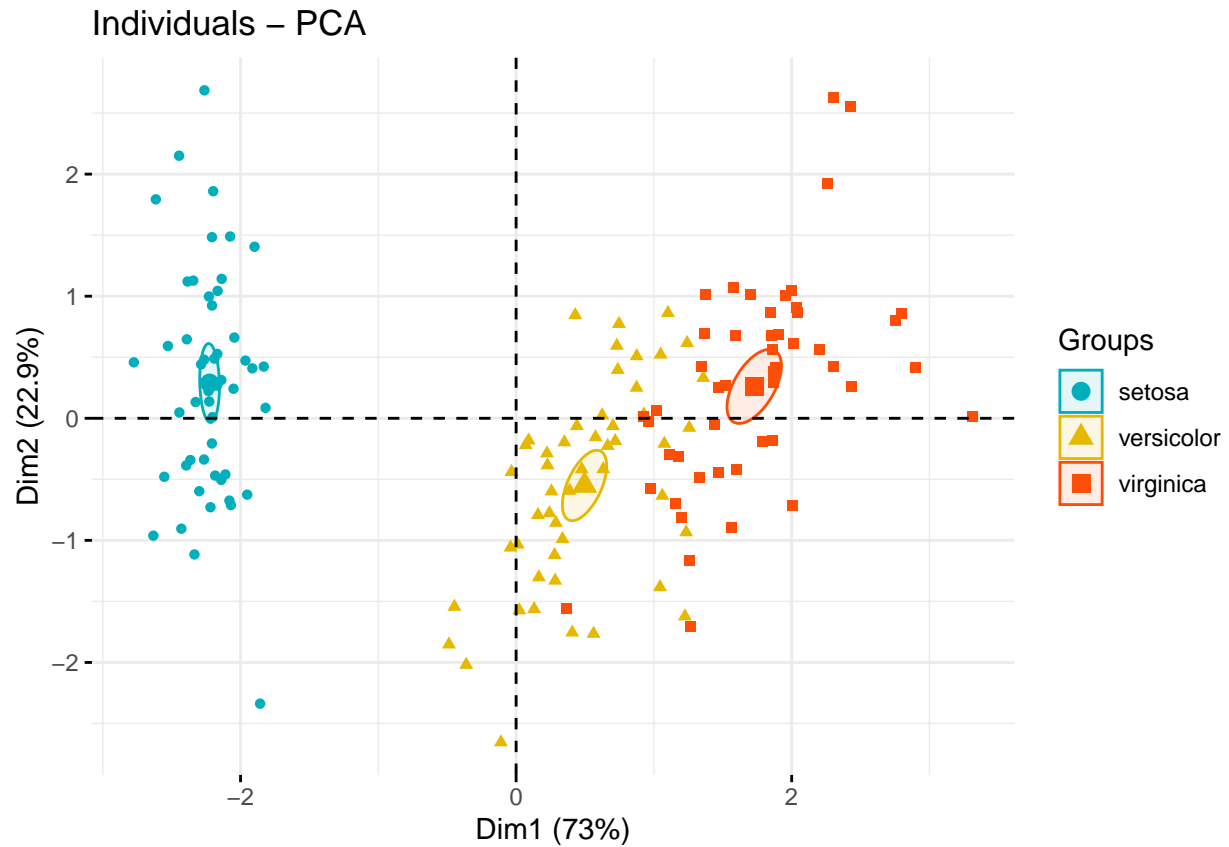
```
fviz_pca_var(iris.pca,  
             pointsize = 3, pointshape = 21, fill = "lightblue",  
             labelsiz = 5, repel = TRUE)
```



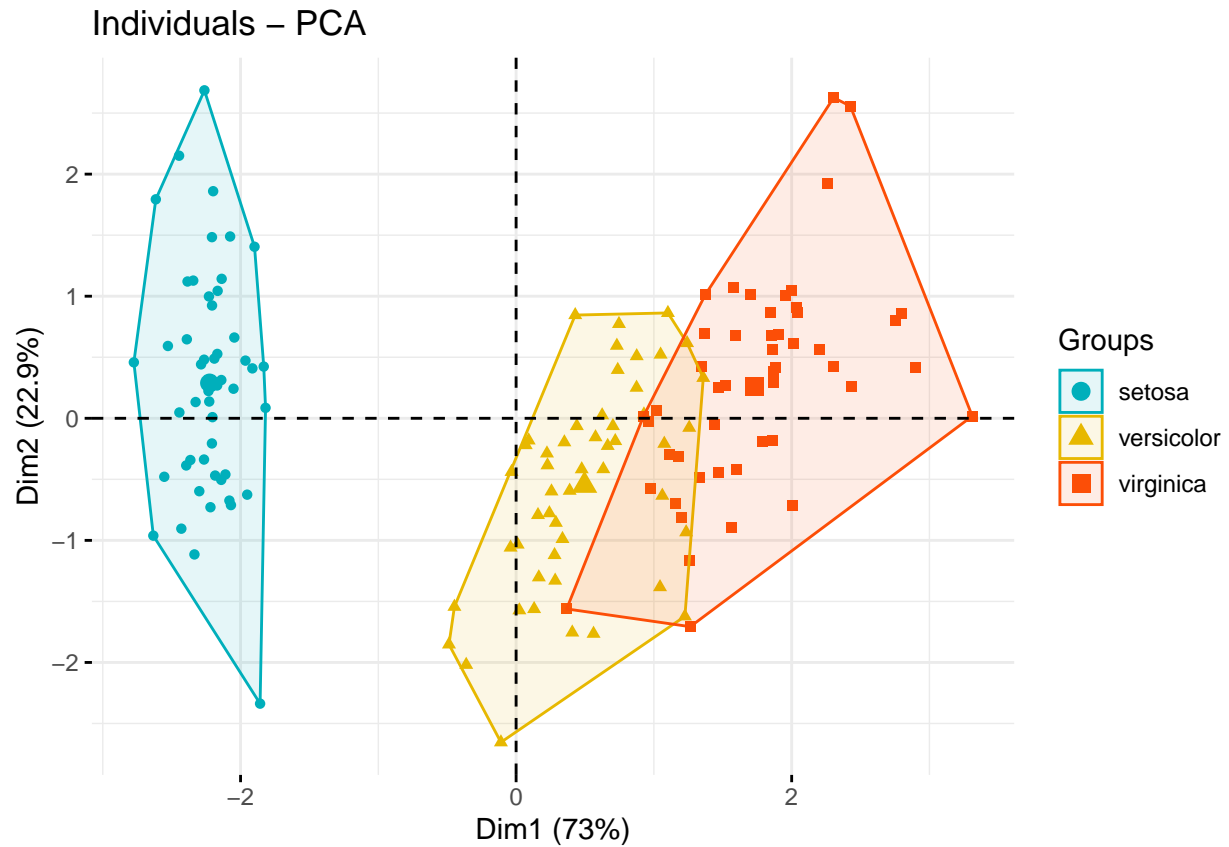
Elipses

O parâmetro `ellipse.type` pode ser usado para mudar o tipo das elipses. “convex”, “confidence”, “t”, “norm”, “euclid”

```
fviz_pca_ind(iris.pca, geom.ind = "point",
  col.ind = iris$Species,
  palette = c("#00AFBB", "#E7B800", "#FC4E07"),
  addEllipses = TRUE, ellipse.type = "confidence",
  legend.title = "Groups")
```



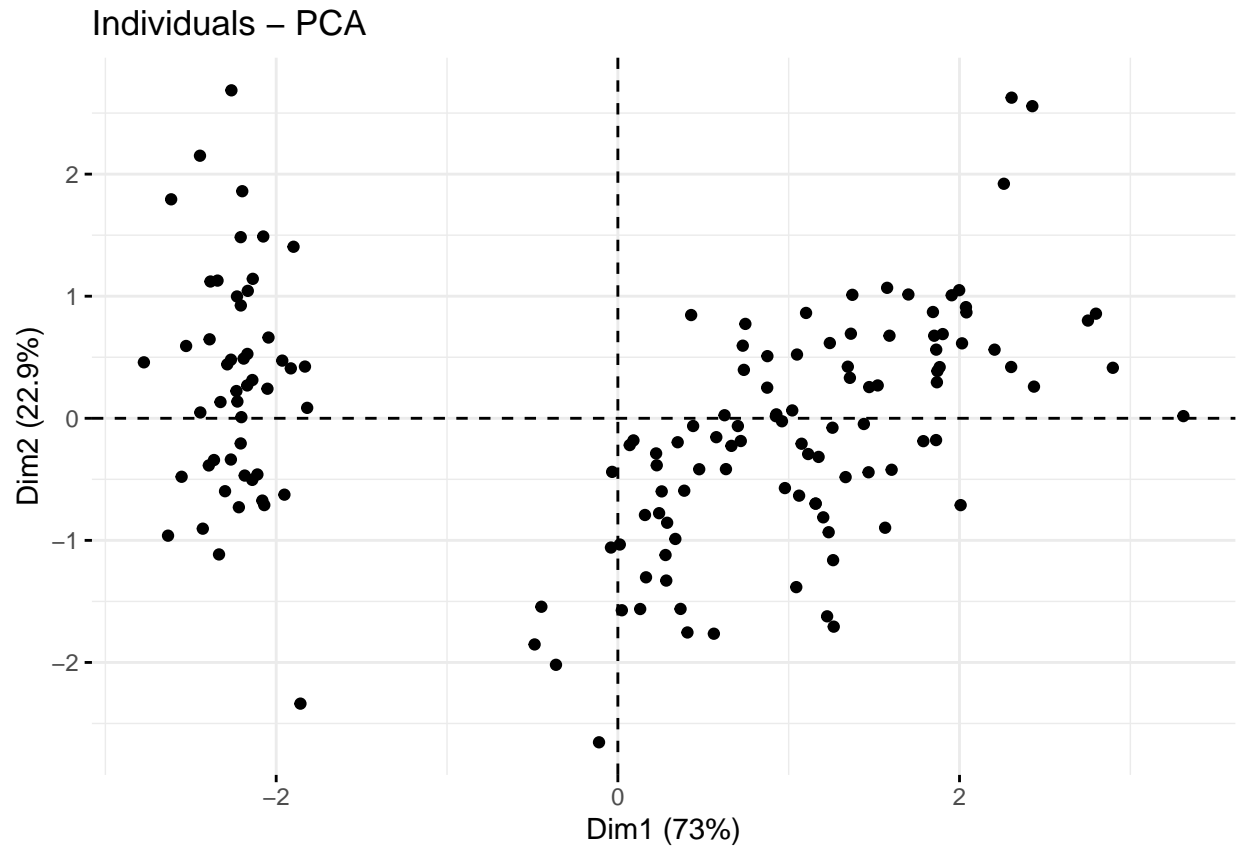
```
fviz_pca_ind(iris.pca, geom.ind = "point",  
  col.ind = iris$Species,  
  palette = c("#00AFBB", "#E7B800", "#FC4E07"),  
  addEllipses = TRUE, ellipse.type = "convex",  
  legend.title = "Groups")
```



Pontos de média dos grupos

Os pontos média dos grupos são mostrados por default Para remover as médias, usar o parâmetro `mean.point = FALSE`

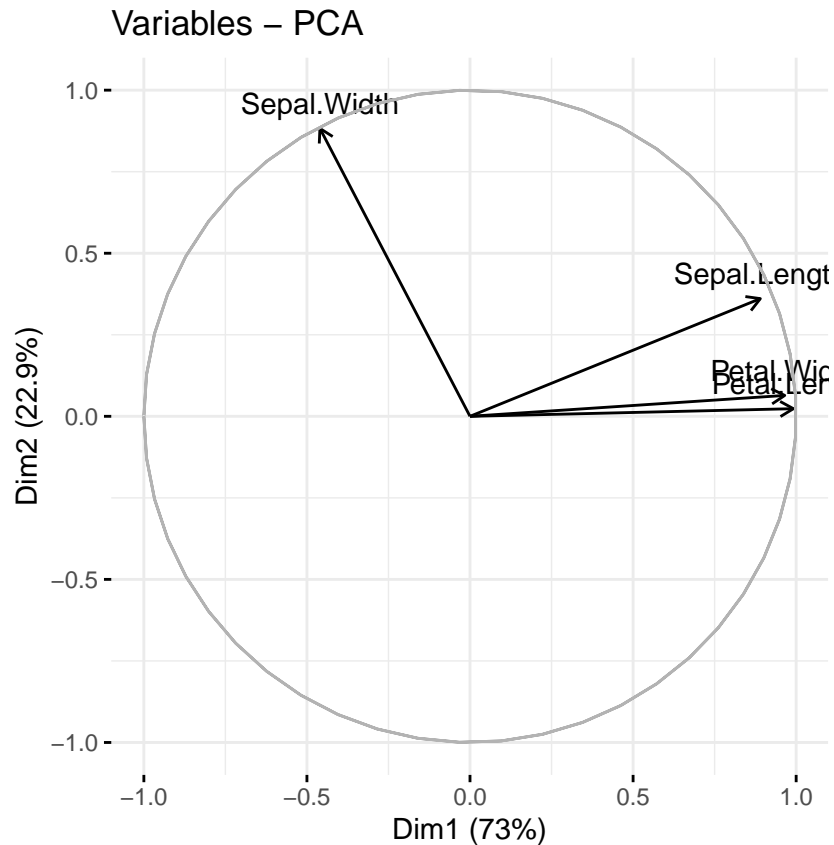
```
fviz_pca_ind(iris.pca,  
  geom.ind = "point",  
  group.ind = iris$Species,  
  legend.title = "Groups",  
  mean.point = FALSE)
```



Linhas dos eixos

O parâmetro `axes.linetype` pode ser usado para especificar o tipo de linhas dos eixos. O default é “dashed”. Os valores permitidos: “blank”, “solid”, “dotted”, etc. Para ver todos os valores possíveis `ggpubr::show_line_types()`

```
fviz_pca_var(iris.pca, axes.linetype = "blank")
```



Parâmetros gráficos

Para alterar os gráficos de qualquer ggplot, pode-se usar a função `ggpar()` [ggpubr]. Os parâmetros que podem ser alterados com `ggpar()`: títulos, labels dos eixos, legenda posição da legenda. Valores possíveis: "top", "bottom", "left", "right", "none" paleta de cores Temas: `theme_gray()`, `theme_bw()`, `theme_minimal()`, `theme_classic()`, `theme_void()`

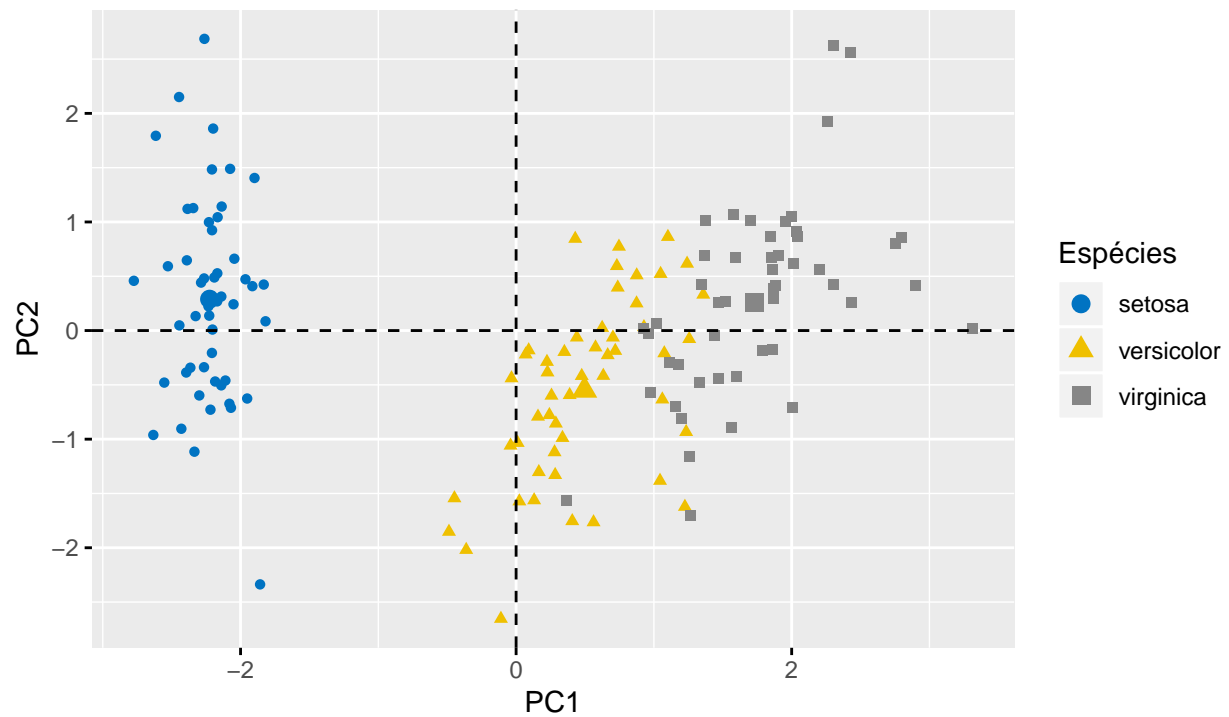
```
library(ggpubr)
```

```
## Loading required package: magrittr
```

```
ind.p <- fviz_pca_ind(iris.pca, geom = "point", col.ind = iris$Species)
ggpar(ind.p,
  title = "Análise de Componentes Principais",
  subtitle = "dataset Iris",
  caption = "Source: factoextra",
  xlab = "PC1", ylab = "PC2",
  legend.title = "Espécies", legend.position = "top",
  ggtheme = theme_gray(), palette = "jco")
```

Análise de Componentes Principais

dataset Iris



Biplot

```
fviz_pca_biplot(iris.pca,  
  col.ind = iris$Species, palette = "jco",  
  addEllipses = TRUE, label = "var",  
  col.var = "black", repel = TRUE,  
  legend.title = "Espécies")
```

