

Examen Parcial 1

ByteMiners

23/02/2020

Análisis exploratorio de datos

```
library(ggplot2)
library(cowplot)
library(dplyr)
library(grid)
```

1. Exploración y visualización de variables

1.1 Explorar la base de datos, realizar los cambios necesarios y convenientes para poder trabajar con ella.

```
df <- read.csv("Ejercicio 1.csv", sep = ",")

# Trasponer las columnas del df para que las características de cada
# observación estén ordenadas por columnas
df_transposed <- as.data.frame(t(as.matrix(df)))

# Eliminar observaciones con valores NA
df_transposed_omitted <- na.omit(df_transposed)

# Colocar nombre a las columnas y eliminar la primera fila
names(df_transposed_omitted) <- c("Peso", "Estatura", "Edad", "Complejión",
  "Papás_separados", "Cantidad_hermanos", "Trabaja", "Horas_trab_dia", "Horas_trab_mes",
  "Salario_mes")

updated_data <- df_transposed_omitted[-c(1), ]
```

2. Metodo de reducción de dimensión (PCA)

2.1 PCA base de datos PCA1

2.2 PCA base de datos PCA2

```
url <- "http://web.stanford.edu/~hastie/ElemStatLearn/datasets/zip.digits/train.2"
data <- read.csv(url)

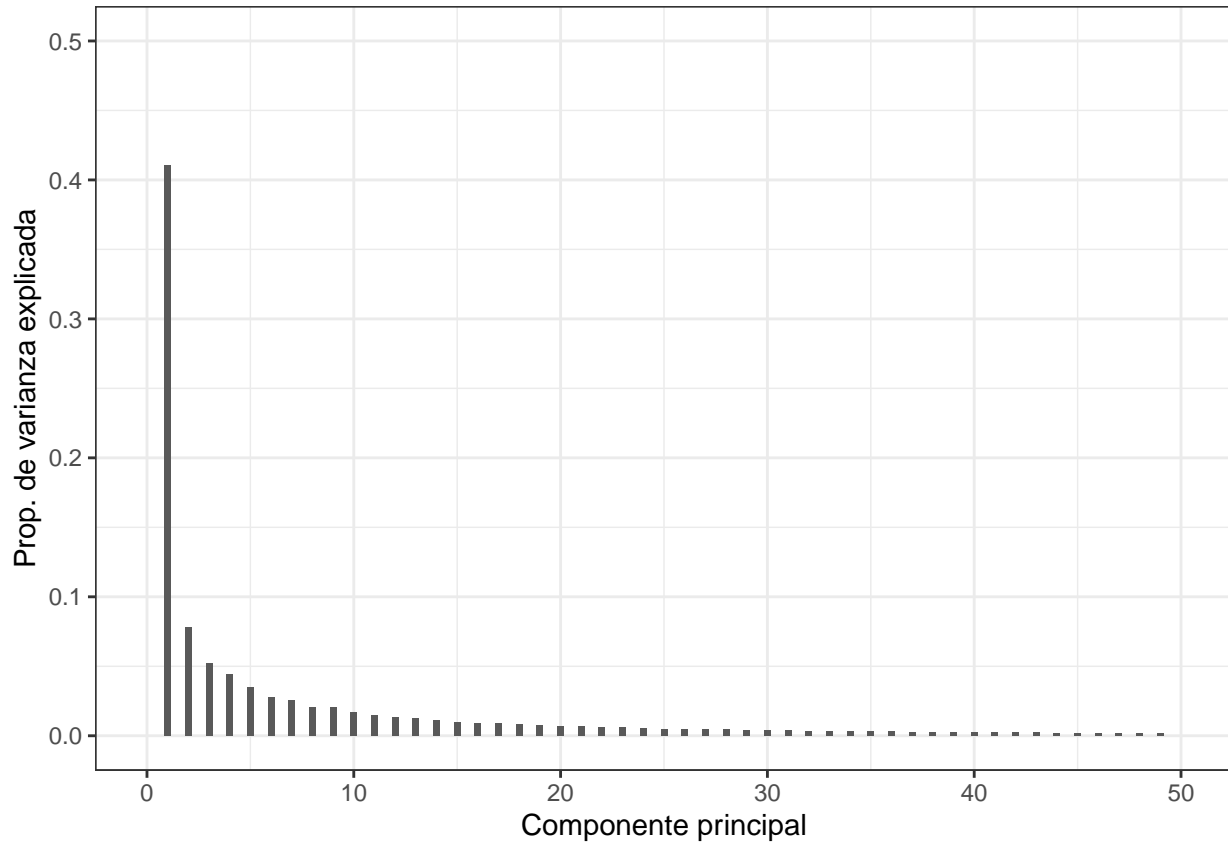
data.pca = scale(data)
data.pca <- prcomp(data, center = F, scale = T)
prop_varianza <- data.pca$sdev^2/sum(data.pca$sdev^2)
prop_varianza_acum <- cumsum(prop_varianza)

ggplot(data = data.frame(prop_varianza, pc = 1:256), aes(x = pc, y = prop_varianza)) +
```

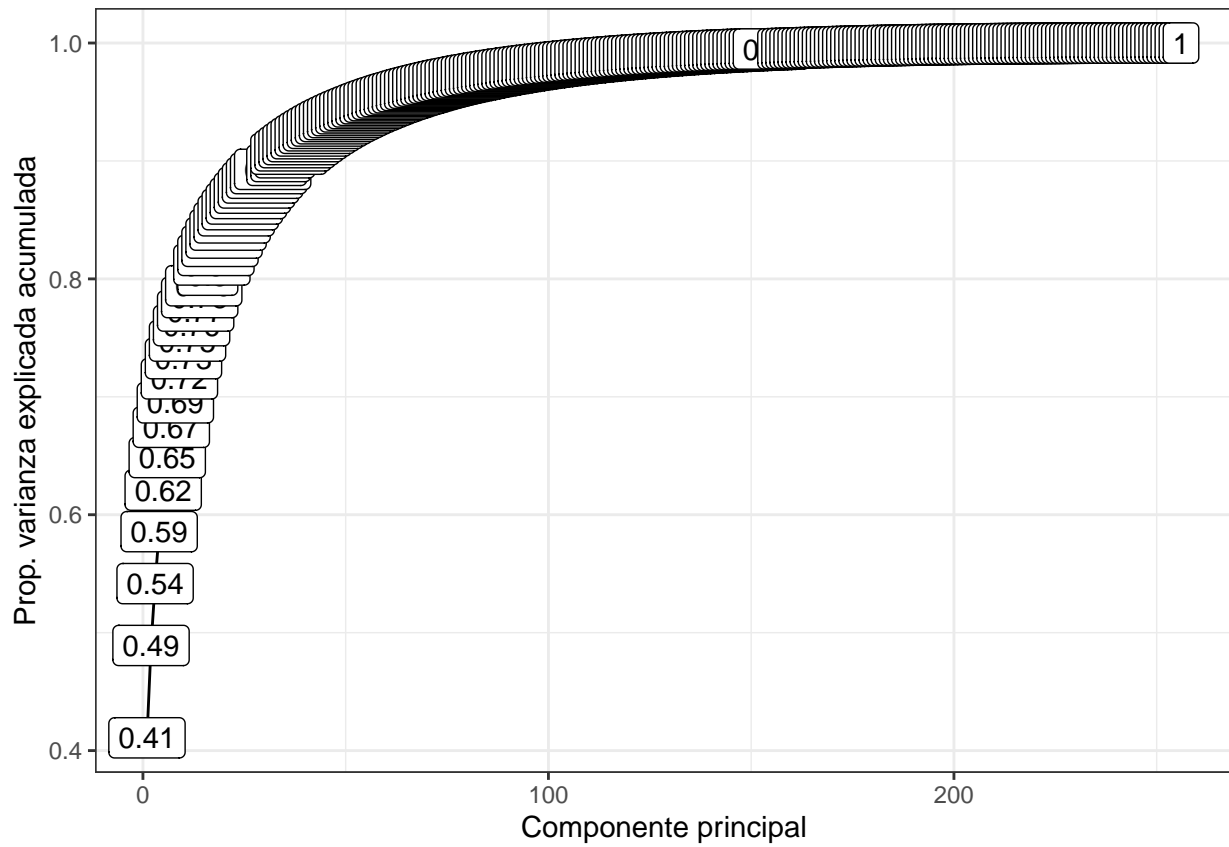
```
xlim(0, 50) + geom_col(width = 0.3) + scale_y_continuous(limits = c(0, 0.5)) +  
theme_bw() + labs(x = "Componente principal", y = "Prop. de varianza explicada")
```

Warning: Removed 206 rows containing missing values (position_stack).

Warning: Removed 1 rows containing missing values (geom_col).



```
ggplot(data = data.frame(prop_varianza_acum, pc = 1:256), aes(x = pc, y = prop_varianza_acum,  
group = 1)) + geom_point() + geom_line() + geom_label(aes(label = round(prop_varianza_acum,  
2))) + theme_bw() + labs(x = "Componente principal", y = "Prop. varianza explicada acumulada")
```



4. Metodo de reducción de dimensión (CCA)