

Plotting PCA

Plotando PCA en R

Formatando los datos

Plink vas generar dos archivos de la analisis del PCA: `plink.eigenval` y `plink.eigenvec`. Con estos dos archivos vamos plotar usando R con el paquete *tidyverse*.

Primero vamos load los paquetes y setar el directorio

```
library(tidyverse)
library(ggplot2)
setwd("~/Dropbox/Postdoc_Rosario/BiodiversityGenomics_Winter2021/dato_denovo/demographic/PCA/")
```

Despues vamos ler los archivos generados por *plink* en R

```
pca <- read_table2("./plink.eigenvec", col_names = FALSE)
eigenval <- scan("./plink.eigenval")
```

Vamos añadir nombres para las columnas de `pca`

```
names(pca)[1] <- "Species"
names(pca)[2] <- "Ind"
names(pca)[3:ncol(pca)] <- paste0("PC", 1:(ncol(pca)-2))
```

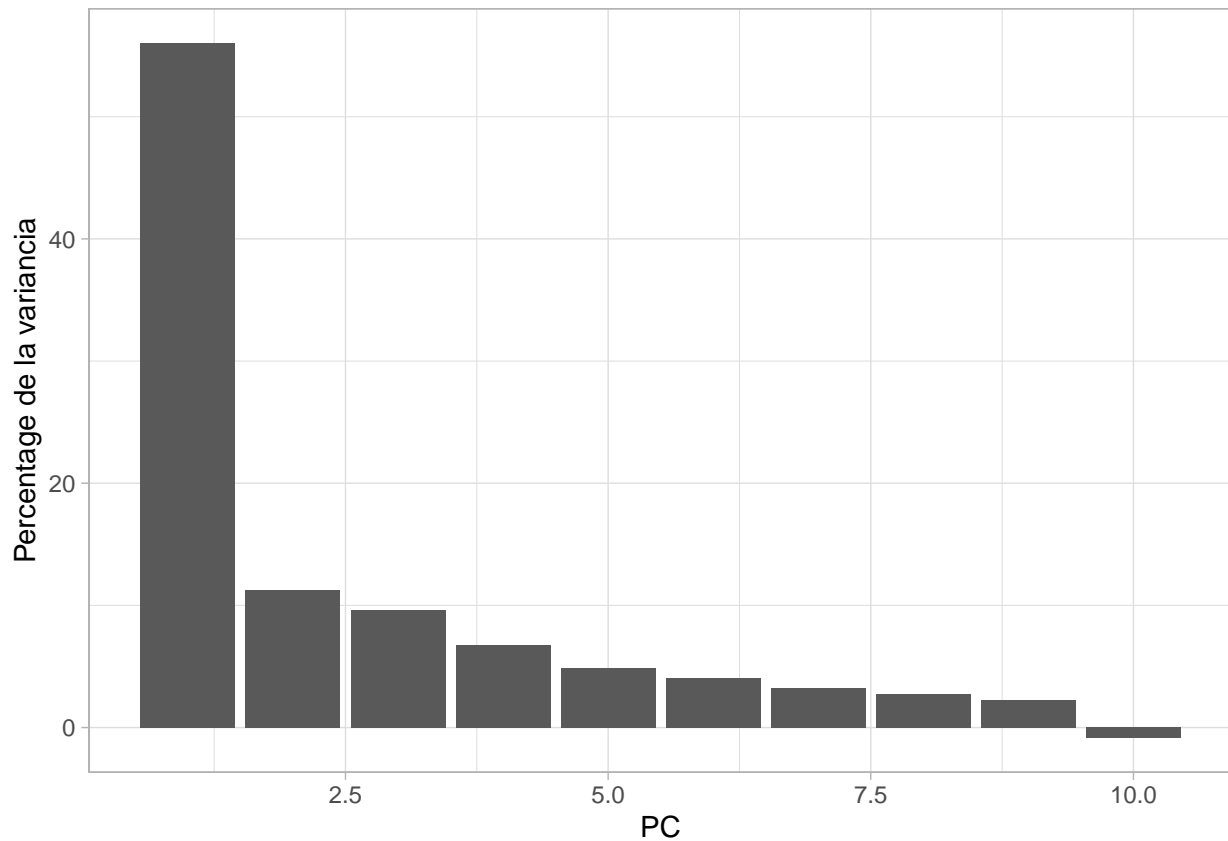
Necesitamos cambiar los valores de eigen para %

```
pve <- data.frame(PC = 1:length(eigenval), pve = eigenval/sum(eigenval)*100)
```

Ahora si, plotando

Primero vamos plotar la variancia explicado por cada componente principal

```
a <- ggplot(pve, aes(PC, pve)) +
  geom_bar(stat = "identity") +
  ylab("Percentage de la variancia") +
  theme_light(); a
```



Ahora vamos plotar el PCA

```
b <- ggplot(pca, aes(PC1, PC2, col = Species)) +
  geom_point(size = 3) +
  scale_colour_manual(values = c("red", "blue")) +
  coord_equal() +
  theme_light() +
  xlab(paste0("PC1 (", signif(pve$pve[1], 3), "%)")) +
  ylab(paste0("PC2 (", signif(pve$pve[2], 3), "%)")); b
```

