Untitled

asd

2024-11-07

rm(list=ls())  
  
# Cargar los datos de los pinguinos  
ruta = "pinguinos.csv"  
datos = read.csv(ruta)  
names(datos)

## [1] "species" "bill\_length\_mm" "bill\_depth\_mm"   
## [4] "flipper\_length\_mm" "body\_mass\_g"

# Mostrar los primeros 6 datos para ver el contenido de los datos  
head(datos)

## species bill\_length\_mm bill\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 1 Adelie 39.1 18.7 181 3750  
## 2 Adelie 39.5 17.4 186 3800  
## 3 Adelie 40.3 18.0 195 3250  
## 4 Adelie 36.7 19.3 193 3450  
## 5 Adelie 39.3 20.6 190 3650  
## 6 Adelie 38.9 17.8 181 3625

## Versión mejorada  
library(flextable) # Liberia para hacer tablas bonitas

## Warning: package 'flextable' was built under R version 4.3.3

flextable(head(datos))

| species | bill\_length\_mm | bill\_depth\_mm | flipper\_length\_mm | body\_mass\_g |
| --- | --- | --- | --- | --- |
| Adelie | 39.1 | 18.7 | 181 | 3,750 |
| Adelie | 39.5 | 17.4 | 186 | 3,800 |
| Adelie | 40.3 | 18.0 | 195 | 3,250 |
| Adelie | 36.7 | 19.3 | 193 | 3,450 |
| Adelie | 39.3 | 20.6 | 190 | 3,650 |
| Adelie | 38.9 | 17.8 | 181 | 3,625 |

## Versión aun mejor  
autofit(theme\_box(flextable(head(datos))))

| **species** | **bill\_length\_mm** | **bill\_depth\_mm** | **flipper\_length\_mm** | **body\_mass\_g** |
| --- | --- | --- | --- | --- |
| Adelie | 39.1 | 18.7 | 181 | 3,750 |
| Adelie | 39.5 | 17.4 | 186 | 3,800 |
| Adelie | 40.3 | 18.0 | 195 | 3,250 |
| Adelie | 36.7 | 19.3 | 193 | 3,450 |
| Adelie | 39.3 | 20.6 | 190 | 3,650 |
| Adelie | 38.9 | 17.8 | 181 | 3,625 |

## Revisar el tipo de variable que tiene  
# Extrae los nombres de las variables  
nombres = names(datos)  
# Extrae los tipos de las variables  
tipos = sapply(datos, class)  
# Crea un nuevo DataFrame con la información  
tiposvariable = data.frame(  
 Nombre = nombres,  
 Tipo = tipos,  
 stringsAsFactors = FALSE)  
  
# Reportar bonito  
theme\_box(flextable(tiposvariable))

| **Nombre** | **Tipo** |
| --- | --- |
| species | character |
| bill\_length\_mm | numeric |
| bill\_depth\_mm | numeric |
| flipper\_length\_mm | integer |
| body\_mass\_g | integer |

# Obtener el conteo de cuantos pinguinos hay por cada especie  
conteo = table(Especies = datos$species)  
conteo

## Especies  
## Adelie Chinstrap Gentoo   
## 151 68 123

# Convertir la tabla de frecuencias en un data frame  
conteo1 = as.data.frame(conteo)  
conteo1

## Especies Freq  
## 1 Adelie 151  
## 2 Chinstrap 68  
## 3 Gentoo 123

# Reportar bonito en una tabla  
theme\_box(flextable(conteo1))

| **Especies** | **Freq** |
| --- | --- |
| Adelie | 151 |
| Chinstrap | 68 |
| Gentoo | 123 |

# Obtener la media, la mediana, la varianza y la desviación estandar  
medias1 = apply(datos[,-1],2,mean)  
medianas1 = apply(datos[,-1],2,median)  
varianza1 = apply(datos[,-1],2,var)  
desvi1 = apply(datos[,-1],2,sd)  
# Pegar los datos juntos y hacerlos una tabla  
resumen1 = (rbind(medias1,medianas1,varianza1,desvi1))  
resumen1

## bill\_length\_mm bill\_depth\_mm flipper\_length\_mm body\_mass\_g  
## medias1 43.921930 17.151170 200.91520 4201.7544  
## medianas1 44.450000 17.300000 197.00000 4050.0000  
## varianza1 29.807054 3.899808 197.73179 643131.0773  
## desvi1 5.459584 1.974793 14.06171 801.9545

# Agregar el nombre del estadístico a la fila  
resumen2 = cbind.data.frame(Estadístico = c("Media","Mediana","Varianza","Desviación"),resumen1)  
resumen2

## Estadístico bill\_length\_mm bill\_depth\_mm flipper\_length\_mm  
## medias1 Media 43.921930 17.151170 200.91520  
## medianas1 Mediana 44.450000 17.300000 197.00000  
## varianza1 Varianza 29.807054 3.899808 197.73179  
## desvi1 Desviación 5.459584 1.974793 14.06171  
## body\_mass\_g  
## medias1 4201.7544  
## medianas1 4050.0000  
## varianza1 643131.0773  
## desvi1 801.9545

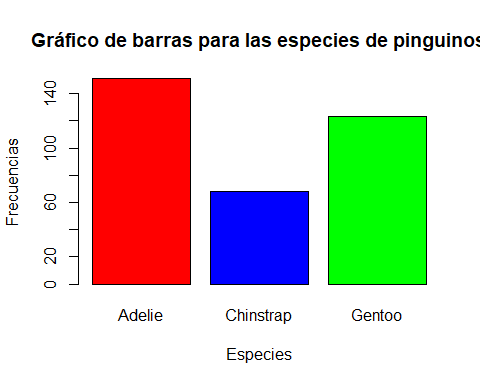
# Reportar bonito  
autofit(theme\_box(flextable(resumen2)))

| **Estadístico** | **bill\_length\_mm** | **bill\_depth\_mm** | **flipper\_length\_mm** | **body\_mass\_g** |
| --- | --- | --- | --- | --- |
| Media | 43.921930 | 17.151170 | 200.91520 | 4,201.7544 |
| Mediana | 44.450000 | 17.300000 | 197.00000 | 4,050.0000 |
| Varianza | 29.807054 | 3.899808 | 197.73179 | 643,131.0773 |
| Desviación | 5.459584 | 1.974793 | 14.06171 | 801.9545 |

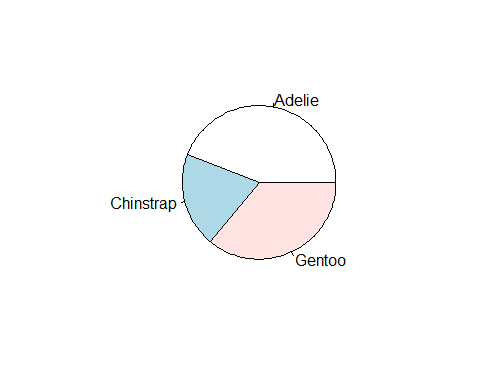
### Automatizando para extraer las estadísticas por especie ###  
# Lista de especies únicas en el dataset  
especies <- unique(datos$species)  
# Inicializar un data.frame vacío para guardar los resultados  
resumen\_total <- data.frame()  
# Ciclo for para cada especie  
for (especie in especies) {  
 # Filtrar los datos para la especie actual  
 datos\_especie <- subset(datos, species == especie)[,-1] # Quitamos la columna 'species'  
 # Calcular los estadísticos para la especie actual  
 medias <- apply(datos\_especie, 2, mean)  
 medianas <- apply(datos\_especie, 2, median)  
 varianza <- apply(datos\_especie, 2, var)  
 desviacion <- apply(datos\_especie, 2, sd)  
   
 # Crear el resumen de la especie actual en un data.frame  
 resumen\_especie <- rbind(medias, medianas, varianza, desviacion)  
 resumen\_especie <- cbind.data.frame(Estadístico = c("Media", "Mediana", "Varianza", "Desviación"),  
 resumen\_especie)  
   
 # Agregar una columna para la especie  
 resumen\_especie$Especie <- especie  
   
 # Añadir los resultados al resumen total  
 resumen\_total <- rbind(resumen\_total, resumen\_especie)  
}  
  
# Mostrar bonito el resumen  
resumen\_total = resumen\_total[,c(1,6,2:5)]  
autofit(theme\_box(flextable(resumen\_total)))

| **Estadístico** | **Especie** | **bill\_length\_mm** | **bill\_depth\_mm** | **flipper\_length\_mm** | **body\_mass\_g** |
| --- | --- | --- | --- | --- | --- |
| Media | Adelie | 38.791391 | 18.3463576 | 189.953642 | 3,700.6623 |
| Mediana | Adelie | 38.800000 | 18.4000000 | 190.000000 | 3,700.0000 |
| Varianza | Adelie | 7.093725 | 1.4802366 | 42.764503 | 210,282.8918 |
| Desviación | Adelie | 2.663405 | 1.2166498 | 6.539457 | 458.5661 |
| Media | Gentoo | 47.504878 | 14.9821138 | 217.186992 | 5,076.0163 |
| Mediana | Gentoo | 47.300000 | 15.0000000 | 216.000000 | 5,000.0000 |
| Varianza | Gentoo | 9.497845 | 0.9627922 | 42.054911 | 254,133.1801 |
| Desviación | Gentoo | 3.081857 | 0.9812198 | 6.484976 | 504.1162 |
| Media | Chinstrap | 48.833824 | 18.4205882 | 195.823529 | 3,733.0882 |
| Mediana | Chinstrap | 49.550000 | 18.4500000 | 196.000000 | 3,700.0000 |
| Varianza | Chinstrap | 11.150630 | 1.2891220 | 50.863916 | 147,713.4548 |
| Desviación | Chinstrap | 3.339256 | 1.1353951 | 7.131894 | 384.3351 |

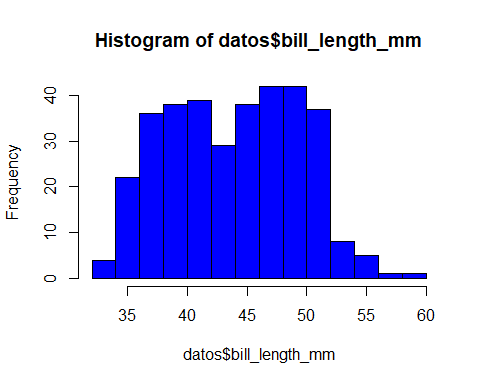
#### Exploración gráfica ####  
  
# Gráfico de barras por especie  
barplot(conteo1$Freq~conteo1$Especies,  
 col = c("red","blue","green"),  
 xlab = "Especies",  
 ylab = "Frecuencias",  
 main = "Gráfico de barras para las especies de pinguinos")



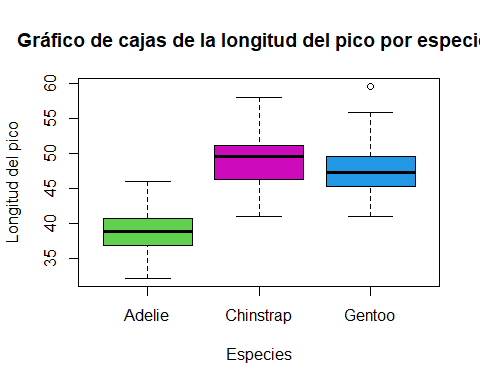
# Gráfico de pastel  
pie(conteo1$Freq,labels = conteo1$Especies)



# Histograma de la distribución de la longitud del pico  
  
hist(datos$bill\_length\_mm, col = "blue")



# Gráfico de cajas para la distribución del pico  
  
boxplot(datos$bill\_length\_mm~datos$species,  
 col = as.factor(datos$bill\_length\_mm),  
 xlab = "Especies",  
 ylab = "Longitud del pico",  
 main = "Gráfico de cajas de la longitud del pico por especies")



# Gráfico de dispersión entre la altura y profundidad del pico  
plot(datos$bill\_length\_mm,datos$bill\_depth\_mm,  
 col= as.factor(datos$species),pch = 16,  
 xlab = "Longitud del pico",  
 ylab = "Profundidad del pico",  
 main = "Gráfico de dispersión entre la longitud del pico y\n la profundidad del pico por especie")

