

# Informe en R de gráficos utilizados en el Proyecto de Google Data Analytics Course

- [Informe en R de gráficos utilizados en el Proyecto de Google Data Analytics Course](#)
  - [Autora: Carmen Witsman García](#)
- [1. Tratamiento de datos](#)
- [2. Gráficos](#)
  - [Diagrama de dispersión: Calories~WeightKg.\(Figura 4\)](#)
  - [Histograma: ActivityHour \(Figura 5\).](#)
  - [Diagrama de caja: ~ ActivityHour \(Figura 6\).](#)
  - [Histograma: ActivityHour con filtrado  \$\geq 200\$  de la variable Calories \(Figura 7\).](#)
  - [Histograma: ActivityHour con filtrado  \$\geq 200\$  de la variable Calories y personas con sobrepeso \(Figura 10\).](#)
  - [Histograma: ActivityHour con filtrado  \$\geq 200\$  de la variable Calories y personas sin sobrepeso \(Figura 11\).](#)
  - [Histograma: ActivityHour con filtrado  \$\geq 450\$  de la variable Calories y personas con sobrepeso \(Figura 12\).](#)
  - [Histograma: ActivityHour con filtrado  \$\geq 450\$  de la variable Calories y personas sin sobrepeso \(Figura 13\).](#)
  - [Gráfica lineal: Calories by Dia de usuario con Id 4558609924 \(Figura 14\).](#)

# Informe en R de gráficos utilizados en el Proyecto de Google Data Analytics Course

**Autora: Carmen Witsman García**

---

09/10/2024

## 1. Tratamiento de datos

```
# Abrimos R Commander
library(Rcmdr)
#Cargamos la librería Readr
library(readr)
# Establecemos directorio de trabajo
setwd("C://Users/witsm/Desktop/G_Analyrtics/Proyecto Final/CSV")
```

```
# Cargamos nuestros datasets
```

```
datos_bellabeat <- read_csv("datos_bellabeat.csv")
```

```
## Rows: 24618718 Columns: 7
## — Column specification —————
## Delimiter: ","
## dbl (7): Id, Calories, SedentaryMinutes, TotalDistance, WeightKg, ActivityHour...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
dailyActivity_merged <- read_csv("dailyActivity_merged.csv")
```

```
## Rows: 940 Columns: 15
## — Column specification —————
## Delimiter: ","
## chr (1): ActivityDate
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDistance...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
hourlyCalories_merged <- read_csv("hourlyCalories_merged.csv")
```

```
## Rows: 22099 Columns: 3
## — Column specification —————
## Delimiter: ","
## chr (1): ActivityHour
## dbl (2): Id, Calories
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Cargamos librería Dplyr
library(dplyr)

# Generamos un dataset donde ver la media de variables para cada uno de los
Ids
df_mean <- datos_bellabeat %>%
  group_by(Id) %>%
  summarise(across(everything(), mean, na.rm = TRUE))

# Vemos con qué usuarios tratamos en el dataset
unique(datos_bellabeat$Id)
```

```
## [1] 1503960366 1927972279 4319703577 4558609924 5577150313 6962181067
```

```
#Creamos un dataframe para cada usuario con sus datos
df_id1 <- dailyActivity_merged %>% filter(Id==1503960366)
df_id2 <- dailyActivity_merged %>% filter(Id==1927972279) # más peso y sede
ntario
df_id3 <- dailyActivity_merged %>% filter(Id==4319703577) # más tiempo cama
df_id4 <- dailyActivity_merged %>% filter(Id==4558609924) # menos tiempo ca
ma
df_id5 <- dailyActivity_merged %>% filter(Id==5577150313) # más calorías
df_id6 <- dailyActivity_merged %>% filter(Id==6962181067) # menos sedentari
o
```

```
# Dataset calorías/hora
# Cambiamos el formato a hora y numérico
hourlyCalories_merged$ActivityHour <- strptime(hourlyCalories_merged$Activi
tyHour, format = "%m/%d/%Y %I:%M:%S %p")
hourlyCalories_merged$ActivityHour <- as.numeric(format(hourlyCalories_merg
ed$ActivityHour, "%H"))
```

```
# Hora preferida para salir a andar
horas_andar <- hourlyCalories_merged %>% filter(Calories>=200)
```

```
# Analizamos personas con sobrepeso
```

```
# Filtramos por Id (previamente los IMCs fueron calculados en Python)
usuarios_sobrepeso <- hourlyCalories_merged %>% filter(Id==c("1927972279",
"4319703577", "4558609924", "5577150313"))
```

```
## warning: There was 1 warning in `filter()`.
## i In argument: `Id == c("1927972279", "4319703577", "4558609924",
##   "5577150313")`.
```

```
## Caused by warning in `Id == c("1927972279", "4319703577", "4558609924",  
"5577150313")`:  
## ! longitud de objeto mayor no es múltiplo de la longitud de uno menor
```

```
# Filtramos para las horas de actividad moderada  
horas_andar_sobrepeso <- usuarios_sobrepeso %>% filter(Calories>=200)  
# Filtramos para las horas de actividad intensa  
horas_correr_sobrepeso <- usuarios_sobrepeso %>% filter(Calories>=450)  
  
# Hacemos lo mismo para usuarios sin sobrepeso  
  
usuarios_NO_sobrepeso <- hourlyCalories_merged %>% filter(Id==c("150396036  
6", "6962181067"))
```

```
## Warning: There was 1 warning in `filter()`.  
## i In argument: `Id == c("1503960366", "6962181067")`.  
## Caused by warning in `Id == c("1503960366", "6962181067")`:  
## ! longitud de objeto mayor no es múltiplo de la longitud de uno menor
```

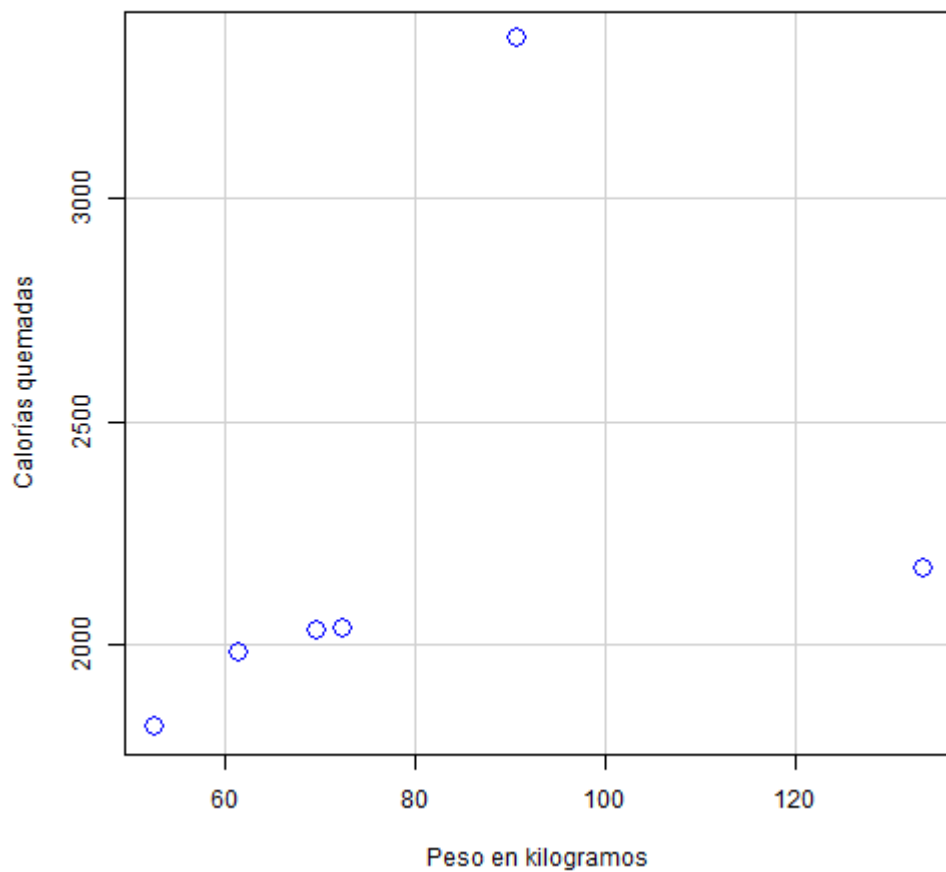
```
horas_andar_NOsobrepeso <- usuarios_NO_sobrepeso %>% filter(Calories>=100)  
horas_correr_NOsobrepeso <- usuarios_NO_sobrepeso %>% filter(Calories>=250)  
  
# Cambiamos el formato de Fecha-hora a Fecha en días  
df_id4$ActivityDate <- as.Date(df_id4$ActivityDate, format = "%m/%d/%Y")  
df_id4 <- df_id4[order(df_id4$ActivityDate), ]  
# Creamos la nueva variable "Día"  
df_id4$Dia <- as.numeric(as.factor(df_id4$ActivityDate))
```

## 2. Gráficos

### Diagrama de dispersión: Calories~WeightKg (Figura 4)

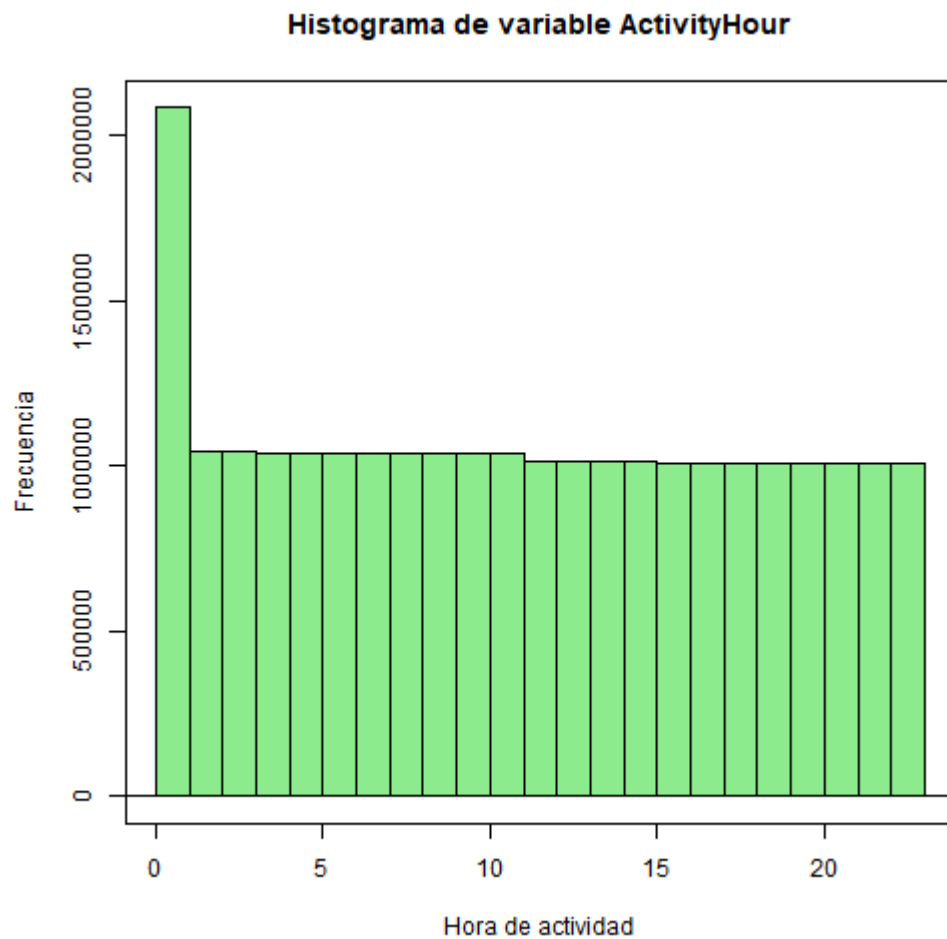
```
scatterplot(Calories~WeightKg, regLine=FALSE, smooth=FALSE, boxplots=FALSE,  
xlab="Peso en kilogramos", ylab="Calorías quemadas",  
main="Diagrama de dispersión de las variables weightkg y Calories", cex=  
1.9, data=df_mean)
```

**Diagrama de dispersión de las variables WeightKg y Calories**



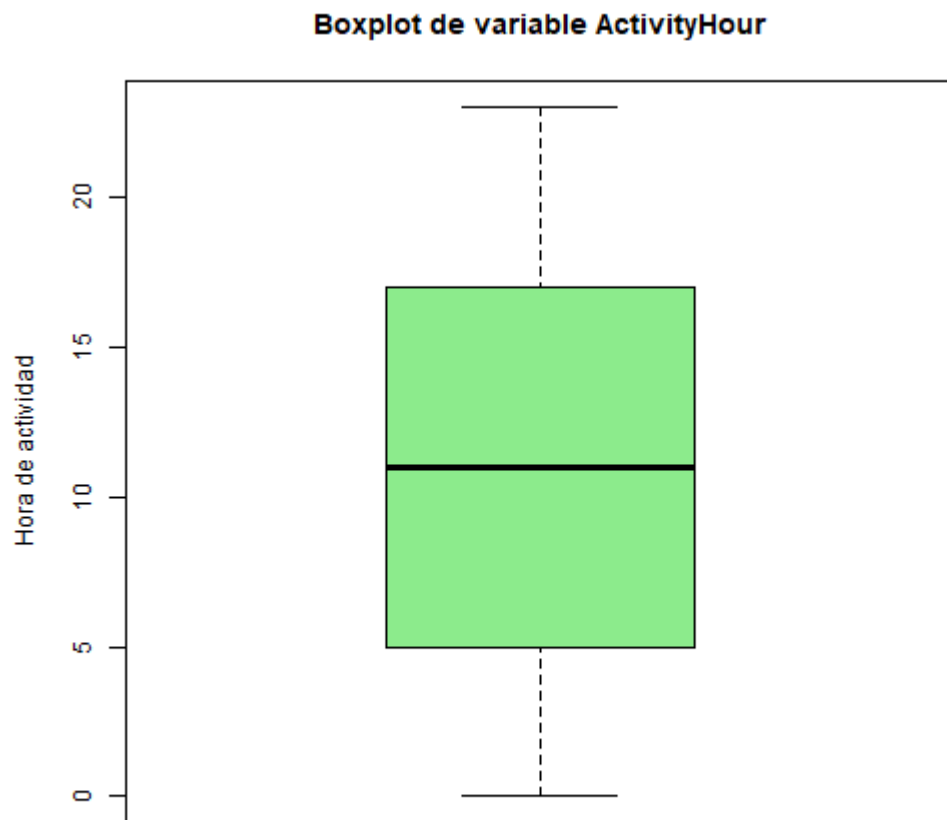
### Histograma: ActivityHour (Figura 5)

```
with(datos_bellabeat, Hist(ActivityHour, scale="frequency", breaks="Sturges", col="lightgreen", xlab="Hora de actividad", ylab="Frecuencia", main="Histograma de variable ActivityHour"))
```



### Diagrama de caja: ~ ActivityHour (Figura 6)

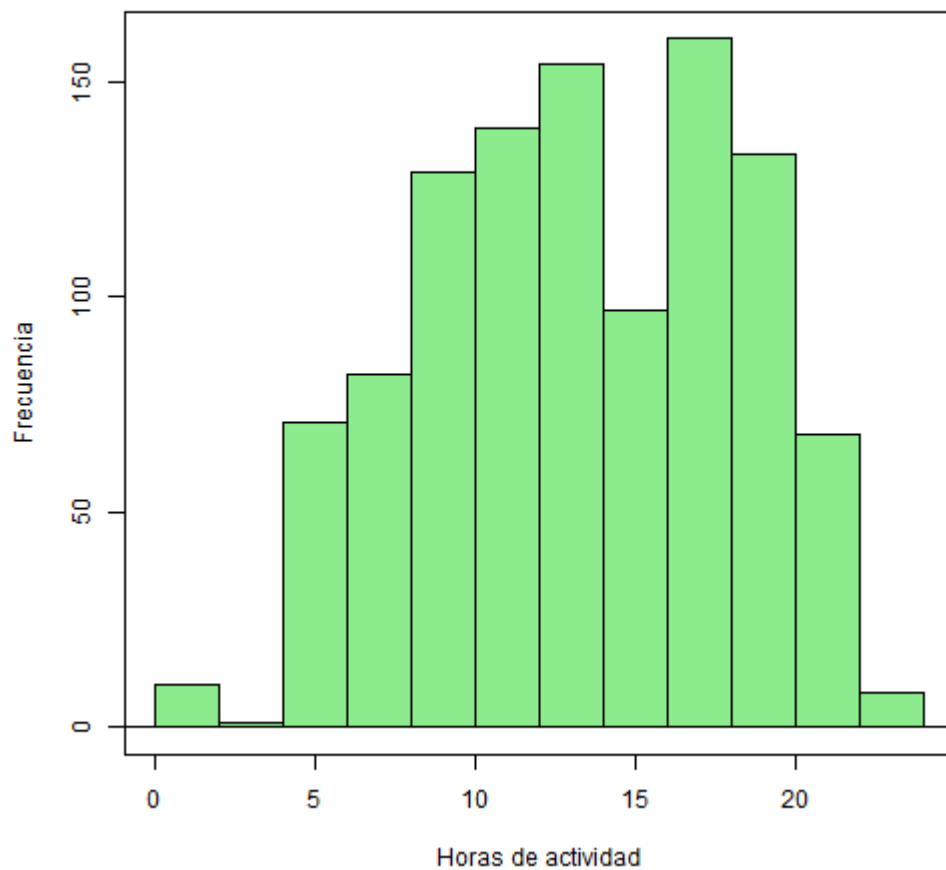
```
Boxplot( ~ ActivityHour, data=datos_bellabeat, id=list(method="y"), ylab="H  
ora de actividad", col="lightgreen",  
main="Boxplot de variable ActivityHour")
```



### Histograma: ActivityHour con filtrado $\geq 200$ de la variable Calories (Figura 7)

```
with(horas_andar, Hist(ActivityHour, scale="frequency", breaks="Sturges", col="lightgreen", xlab="Horas de actividad", ylab="Frecuencia", main="Histograma de variable ActivityHour con filtrado  $\geq 200$  de la variable calories"))
```

**Histograma de variable ActivityHour con filtrado  $\geq 200$  de la variable Calo**



**Histograma: ActivityHour con filtrado  $\geq 200$  de la variable Calories y personas con sobrepeso (Figura 10)**

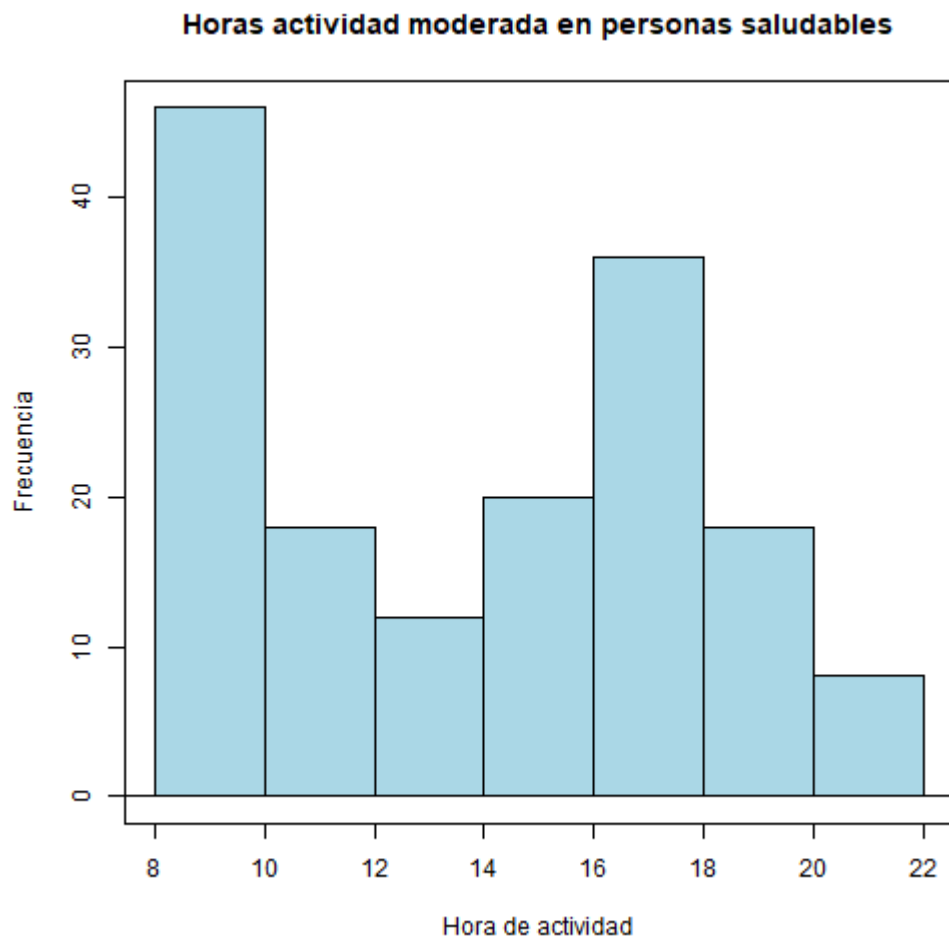
```
with(horas_andar_sobrepeso, Hist(ActivityHour, scale="frequency",  
breaks="Sturges", col="lightpink", xlab="Horas de actividad",  
ylab="Frecuencia",  
main="Horas actividad moderada en personas con sobrepeso"))
```





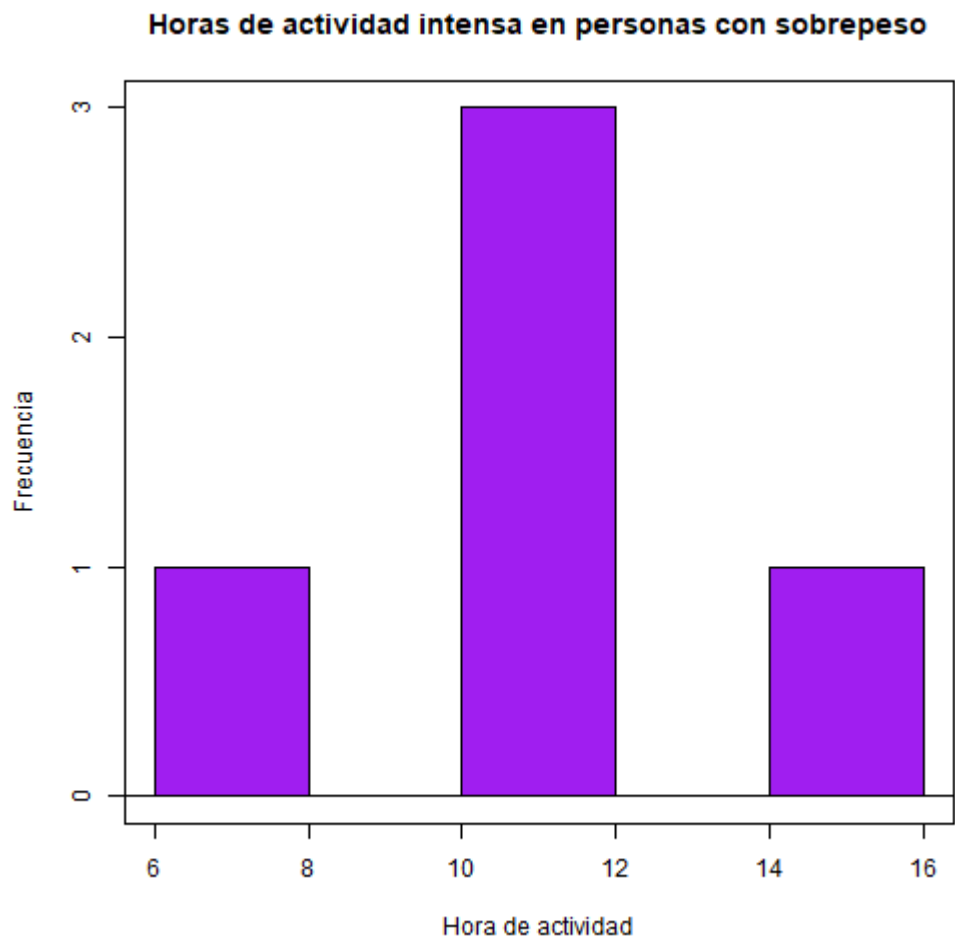
**Histograma: ActivityHour con filtrado  $\geq 200$  de la variable Calories y personas sin sobrepeso (Figura 11)**

```
with(horas_andar_NOsobrepeso, Hist(ActivityHour, scale="frequency",  
breaks="Sturges", col="lightblue", xlab="Hora de actividad",  
ylab="Frecuencia", main="Horas actividad moderada en personas saludables"))
```



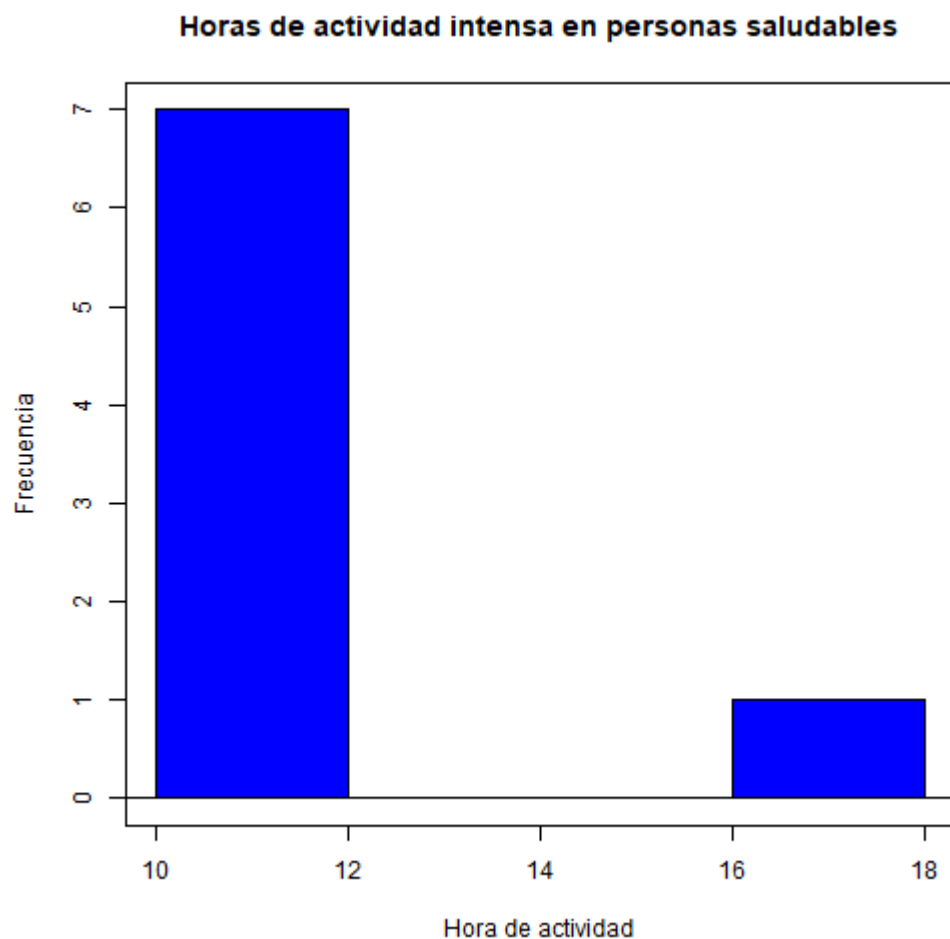
**Histograma: ActivityHour con filtrado  $\geq 450$  de la variable Calories y personas con sobrepeso (Figura 12)**

```
with(horas_correr_sobrepeso, Hist(ActivityHour, scale="frequency",  
breaks="Sturges", col="purple", xlab="Hora de actividad",  
ylab="Frecuencia",  
main="Horas de actividad intensa en personas con sobrepeso"))
```



**Histograma: ActivityHour con filtrado  $\geq 450$  de la variable Calories y personas sin sobrepeso (Figura 13)**

```
with(horas_correr_NOsobrepeso, Hist(ActivityHour, scale="frequency",  
breaks="Sturges", col="blue", xlab="Hora de actividad",  
ylab="Frecuencia",  
main="Horas de actividad intensa en personas saludables"))
```



**Gráfica lineal: Calories by Dia de usuario con Id 4558609924 (Figura 14)**

```
with(df_id4, lineplot(Dia, calories))
```

