

```
getwd()
```

Tarea Número 1 - REPRODUCIBLE RESEARCH

Leo la data, realizo estadísticos descriptivos y veo como esta estructurado

```
actividad <- read.csv("C:/Users/juana/Documents/activity.csv")  
  
summary(actividad)
```

```
##      steps      date      interval  
## Min.   : 0.00  Length:17568  Min.   : 0.0  
## 1st Qu.: 0.00  Class :character 1st Qu.: 588.8  
## Median : 0.00  Mode  :character Median :1177.5  
## Mean   : 37.38                Mean   :1177.5  
## 3rd Qu.: 12.00                3rd Qu.:1766.2  
## Max.   :806.00                Max.   :2355.0  
## NA's   :2304
```

```
head(actividad)
```

```
##  steps      date interval  
## 1    NA 2012-10-01      0  
## 2    NA 2012-10-01      5  
## 3    NA 2012-10-01     10  
## 4    NA 2012-10-01     15  
## 5    NA 2012-10-01     20  
## 6    NA 2012-10-01     25
```

¿Cuál es el número total medio de pasos dados por día?

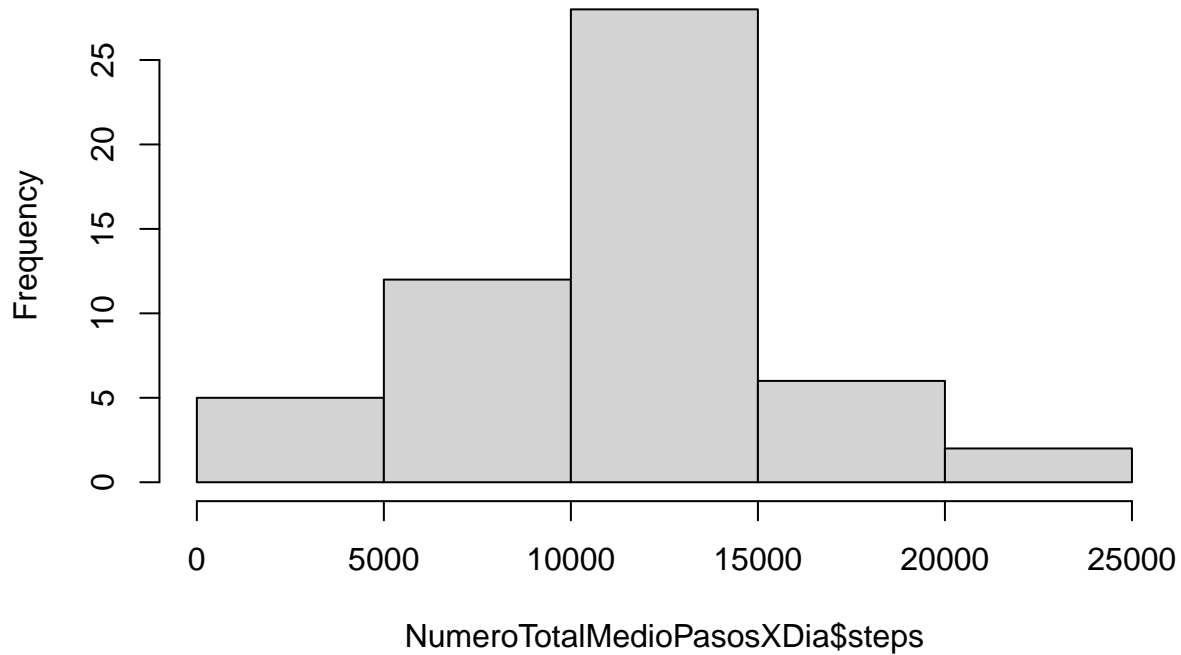
Calcule el número total de pasos dados por día.

```
NumeroTotalMedioPasosXDia <- aggregate(steps ~ date, actividad, sum, na.rm=TRUE)
```

Haga un histograma del número total de pasos dados cada día.

```
hist(NumeroTotalMedioPasosXDia$steps)
```

Histogram of NumeroTotalMedioPasosXDia\$steps



Calcule e informe la media y la mediana del número total de pasos dados por día.

```
mean(NumeroTotalMedioPasosXDia$steps)
```

```
## [1] 10766.19
```

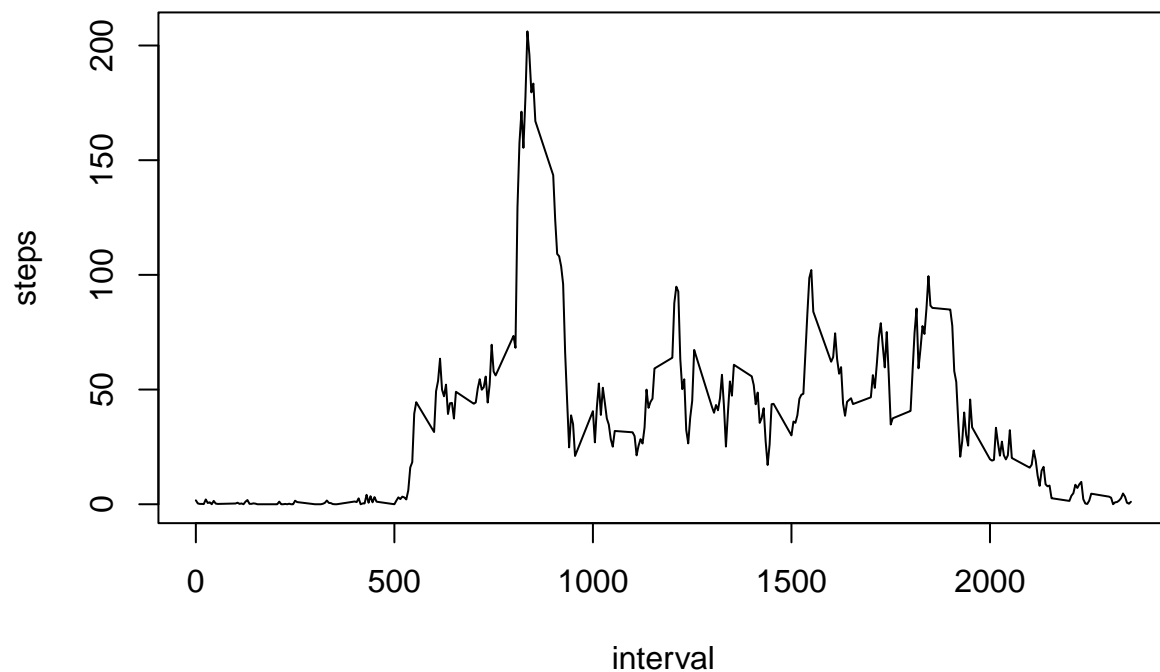
```
median(NumeroTotalMedioPasosXDia$steps)
```

```
## [1] 10765
```

¿Cuál es el patrón de actividad diaria promedio?

Haga una gráfica de series de tiempo (es decir tipo = "l") del intervalo de 5 minutos (eje x) y el número medio de pasos dados, promediado en todos los días (eje y)

```
IntervaloPasos<-aggregate(steps~interval, data=actividad, mean, na.rm=TRUE)  
plot(steps~interval, data=IntervaloPasos, type="l")
```



¿Qué intervalo de 5 minutos, en promedio en todos los días del conjunto de datos, contiene el número máximo de pasos?

```
IntervaloPasos[which.max(IntervaloPasos$steps),]$interval
```

```
## [1] 835
```

Imputación de valores perdidos

Calcule e informe el número total de valores faltantes en el conjunto de datos (es decir, el número total de filas con `color {rojo} { verbo | NA |}N / As`)

```
sum(is.na(actividad$steps))
```

```
## [1] 2304
```

Diseñe una estrategia para completar todos los valores faltantes en el conjunto de datos. La estrategia no necesita ser sofisticada. Por ejemplo, puede usar la media / mediana de ese día, o la media de ese intervalo de 5 minutos, etc.

```
MediaPasosPorIntervalo<-function(interval){
  IntervaloPasos[IntervaloPasos$interval==interval,]$steps
}
```

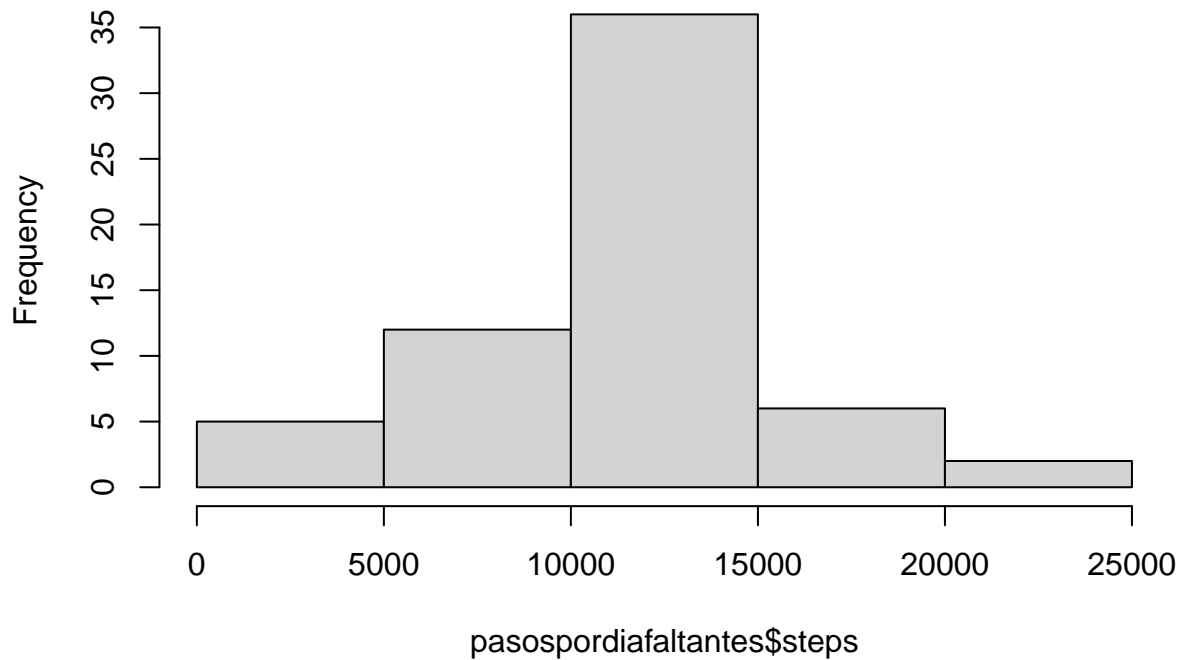
Cree un nuevo conjunto de datos que sea igual al conjunto de datos original pero con los datos faltantes completados.

```
actividadaddatosfaltantes<-actividad
for(i in 1:nrow(actividadaddatosfaltantes)){
  if(is.na(actividadaddatosfaltantes[i,]$steps)){
    actividadaddatosfaltantes[i,]$steps <- MediaPasosPorIntervalo(actividadaddatosfaltantes[i,]$interval)
  }
}
```

Haga un histograma del número total de pasos dados cada día y Calcule e informe la media y la mediana del número total de pasos dados por día. ¿Estos valores difieren de las estimaciones de la primera parte de la tarea? ¿Cuál es el impacto de imputar los datos faltantes a las estimaciones del número total diario de pasos?

```
pasospordiafaltantes <- aggregate(steps ~ date, data=actividadaddatosfaltantes, sum)
hist(pasospordiafaltantes$steps)
```

Histogram of pasospordiafaltantes\$steps



```
mediapasospordiafaltantes <- mean(pasospordiafaltantes$steps)
mediapasospordiafaltantes
```

```
## [1] 10766.19
```

```
medianapasospordiafaltantes <- median(pasospordiafaltantes$steps)
medianapasospordiafaltantes
```

```
## [1] 10766.19
```

¿Existen diferencias en los patrones de actividad entre los días de semana y los fines de semana?

Cree una nueva variable de factor en el conjunto de datos con dos niveles: “día de la semana” y “fin de semana”, que indican si una fecha determinada es un día de la semana o un día de fin de semana.

```
actividad$date <- as.Date(strptime(actividad$date, format="%Y-%m-%d"))
actividad$day <- weekdays(actividad$date)
for (i in 1:nrow(actividad)) {
  if (actividad[i,]$day %in% c("Saturday", "Sunday")) {
    actividad[i,]$day <- "weekend"
  }
}
```

```

    }
    else{
      actividad[i,]$day<-"weekday"
    }
  }
}
pasosdia <- aggregate(actividad$steps ~ actividad$interval + actividad$day, actividad, mean)

```

Haga un diagrama de panel que contenga un diagrama de serie de tiempo (es decir, color {rojo} { verb | type = "l" |}tipo = "l") del intervalo de 5 minutos (eje x) y el número medio de pasos dados, promediado entre todos los días de la semana o los días de fin de semana (eje y). Consulte el archivo README en el repositorio de GitHub para ver un ejemplo de cómo debería verse este gráfico con datos simulados.

```

names(pasosdia) <- c("interval", "day", "steps")
library(lattice)
xyplot(steps ~ interval | day, pasosdia, type = "l", layout = c(1, 2),
       xlab = "Interval", ylab = "Number of steps")

```

