

R Notebook

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.0.5
```

```
usuarios = read.table("usuarios5.csv", header = TRUE, sep = ",")
```

```
recorridos = read.table("recorridos5.csv", header = TRUE, sep = ",")
```

```
usuarios$genero_usuario = as.factor(usuarios$genero_usuario)
```

```
usuarios_recorridos = merge(usuarios, recorridos)
```

```
attach(recorridos)
```

```
dias = factor(recorridos$dia, levels = c("Domingo", "Lunes", "Martes", "Miércoles", "Jueves", "Viernes"))
```

```
usuarios_recorridos$dia = factor(dias)
```

```
direcciones = as.factor(c(direccion_estacion_origen, direccion_estacion_destino))
```

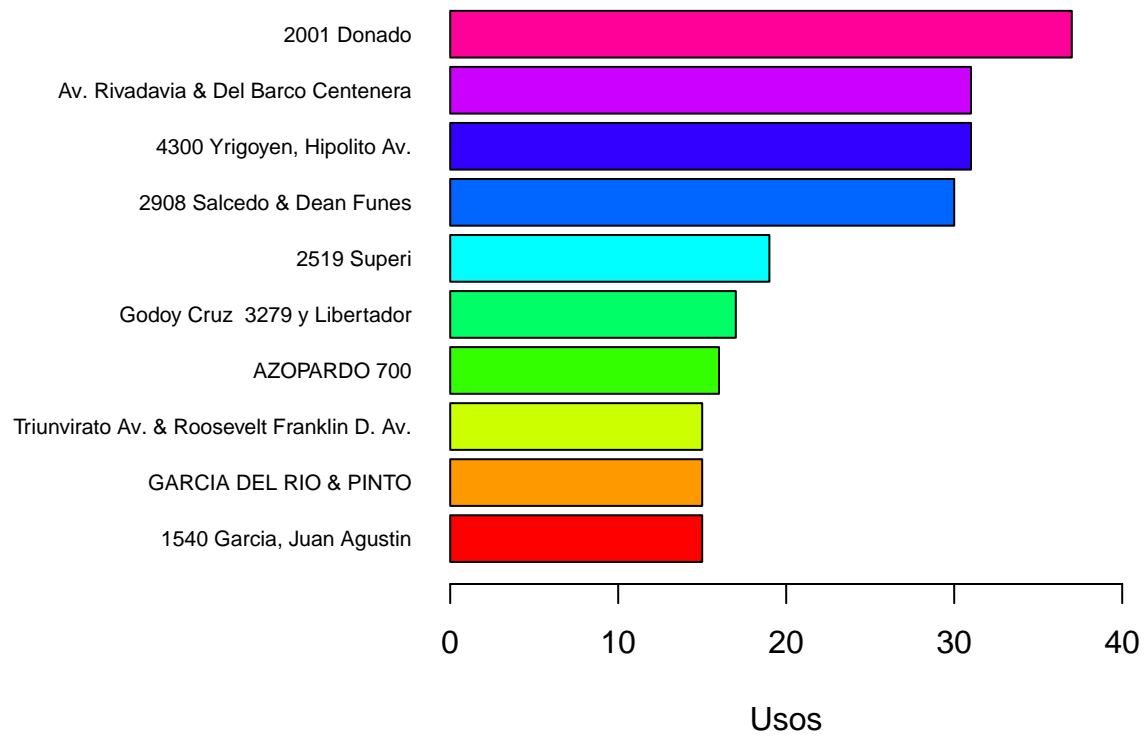
```
vn = sort(table(direcciones), decreasing = TRUE)
```

```
vn2 = sort(vn, decreasing = TRUE)[1:10]
```

Top 10 estaciones

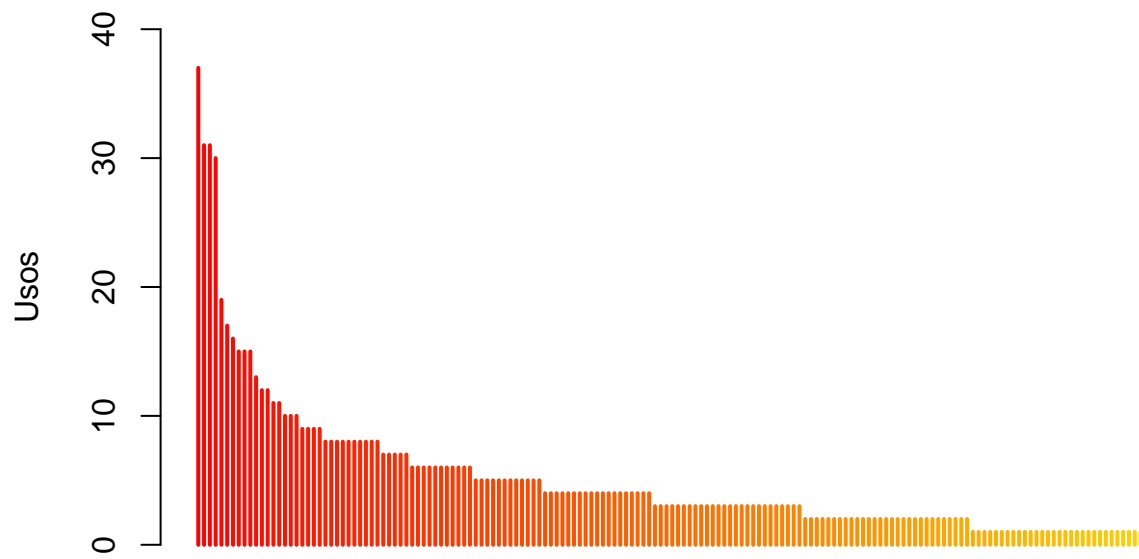
```
par(mar=c(5,13,2,2))
```

```
barplot(sort(vn2, decreasing = FALSE), las=1, xlim = c(0, 40), xlab = "Usos", col = rainbow(10), horiz = TRUE)
```



Uso de Estaciones

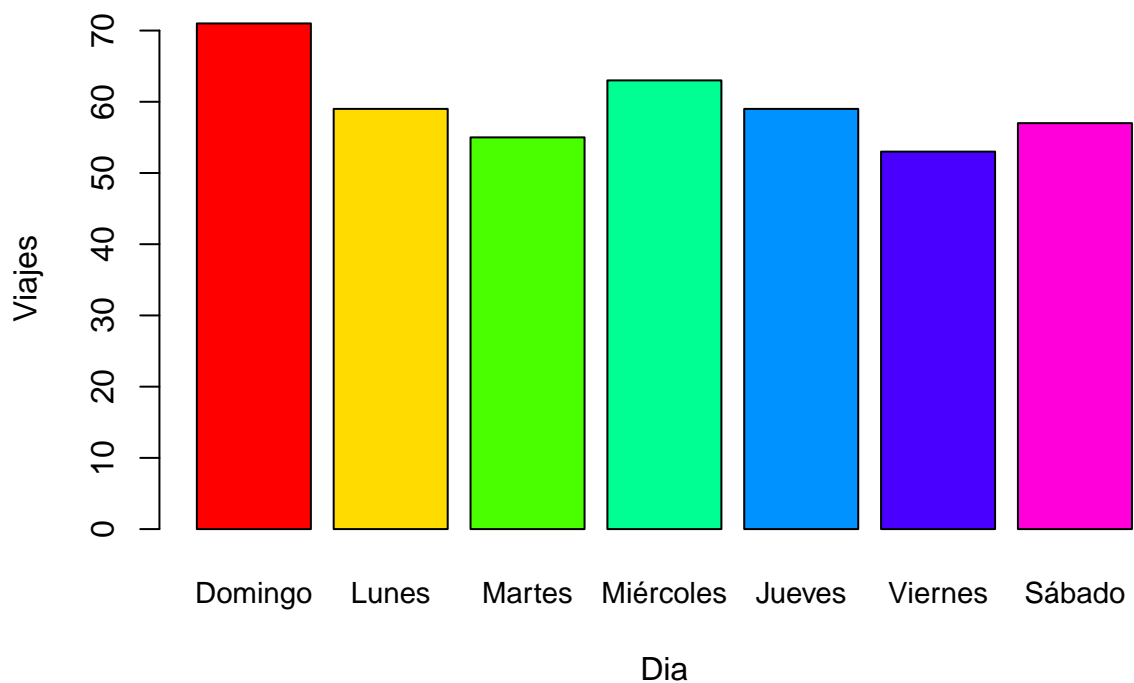
```
par(mar=c(4,4,4,2))
plot(vn, ylim = c(0, 40), xaxt='n', xlab = "Paradas", ylab = "Usos", col = rainbow(1200))
```



Paradas

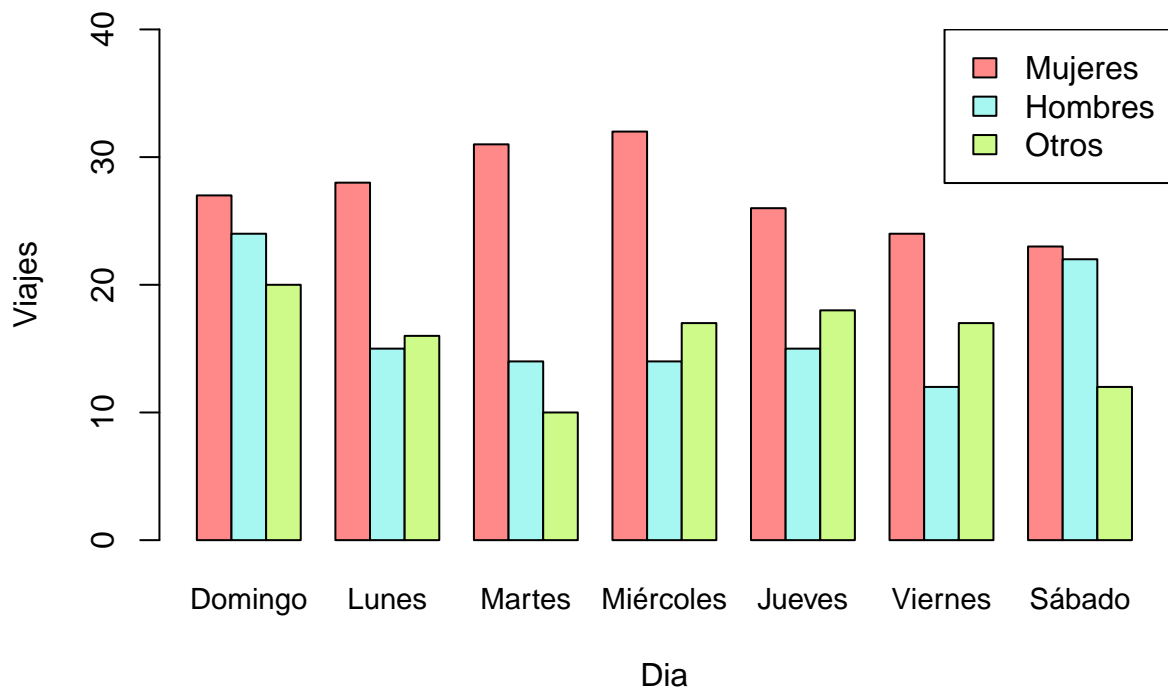
Viajes cada Dia

```
viajes_semana = table(usuarios_recorridos$dias)
barplot(viajes_semana, xlab = "Dia", ylab = "Viajes", names = levels(dias), col = rainbow(7), beside = 't')
```



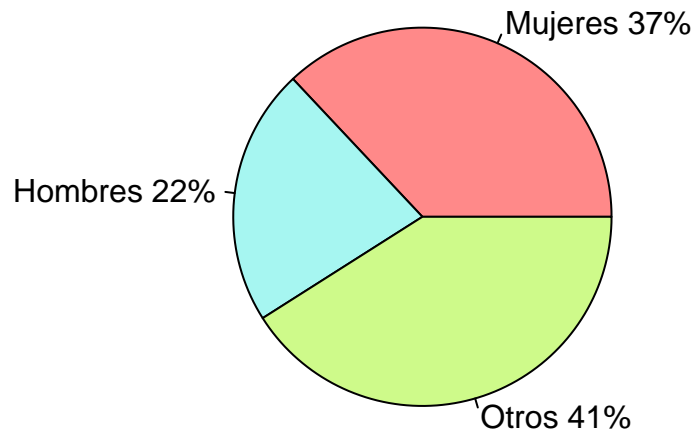
Viajes por Sexo cada Dia

```
viajes_sexo_semana = table(usuarios_recorridos$genero_usuario, usuarios_recorridos$dia)
barplot(viajes_sexo_semana, xlab = "Dia", ylab = "Viajes", names = levels(dias), col = c("#FF8989", "#A6F6F1", "#CEFA8A"),
legend("topright", legend = c("Mujeres", "Hombres", "Otros"), fill = c("#FF8989", "#A6F6F1", "#CEFA8A"))
```



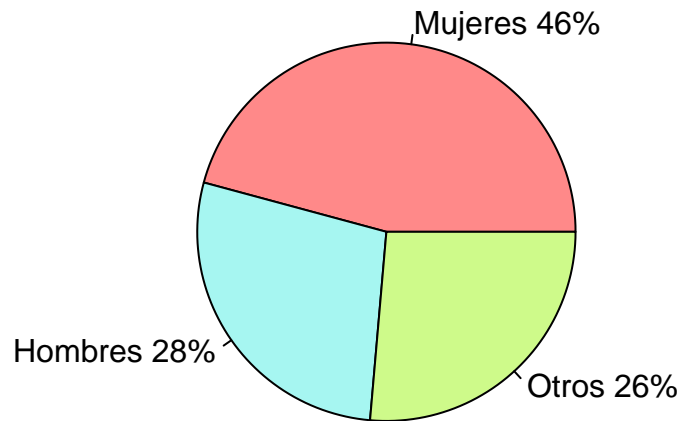
Usuarios por Sexo

```
usuarios_generos = table(usuarios$genero_usuario)
porcentajes_generos = usuarios_generos
porcentajes_generos = paste(c("Mujeres", "Hombres", "Otros"), porcentajes_generos)
porcentajes_generos = paste(porcentajes_generos, "%", sep = "")
pie(usuarios_generos, labels = porcentajes_generos, col = c("#FF8989", "#A6F6F1", "#CEFA8A"))
```



Viajes Sexos

```
viajes_sexo = table(usuarios_recorridos$genero_usuario)
porcentajes_viajes_sexo = round((viajes_sexo * 100)/417)
porcentajes_viajes_sexo = paste(c("Mujeres", "Hombres", "Otros"), porcentajes_viajes_sexo)
porcentajes_viajes_sexo = paste(porcentajes_viajes_sexo, "%", sep = "")
pie(viajes_sexo, labels = porcentajes_viajes_sexo, col = c("#FF8989", "#A6F6F1", "#CEFA8A"))
```

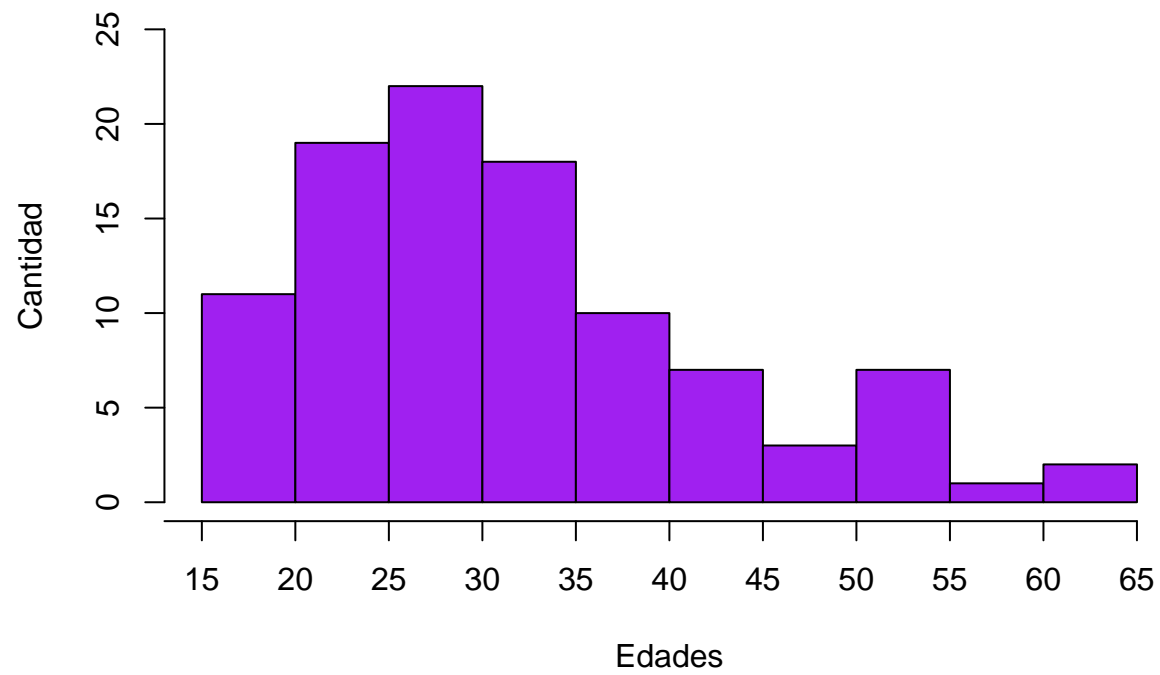


Usuarios por Edad (tallo y hoja, e histograma)

```
stem(usuarios$edad_usuario, atom = 10)
```

```
##
## The decimal point is 1 digit(s) to the right of the |
##
## 1 | 888999
## 2 | 0000011122222333333355556677777778899999
## 3 | 0000011111112222334455566777788
## 4 | 00113445568
## 5 | 01111124
## 6 | 035
```

```
hist(usuarios$edad_usuario, main = "", xlab = "Edades", ylab = "Cantidad", col = "purple", ylim = c(0, 100))
```



Viajes Edades

```
hist(usuario_recorridos$edad_usuario, main = "", xlab = "Edades", ylab = "Cantidad", col = "purple", y
```

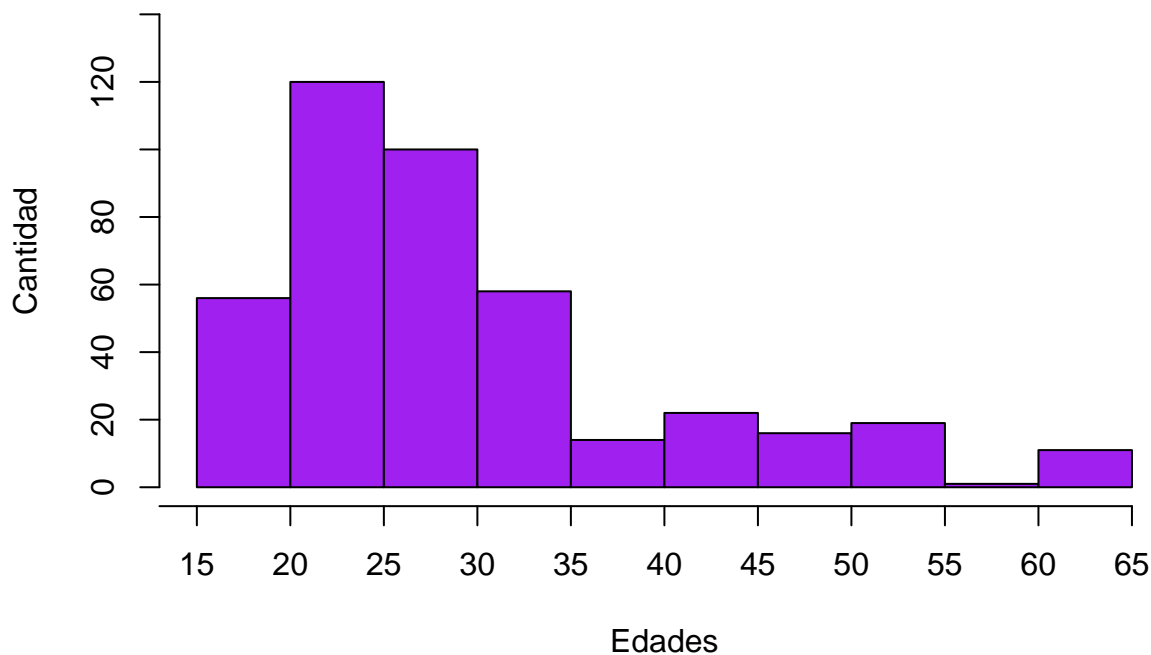



Tabla viajes Edades

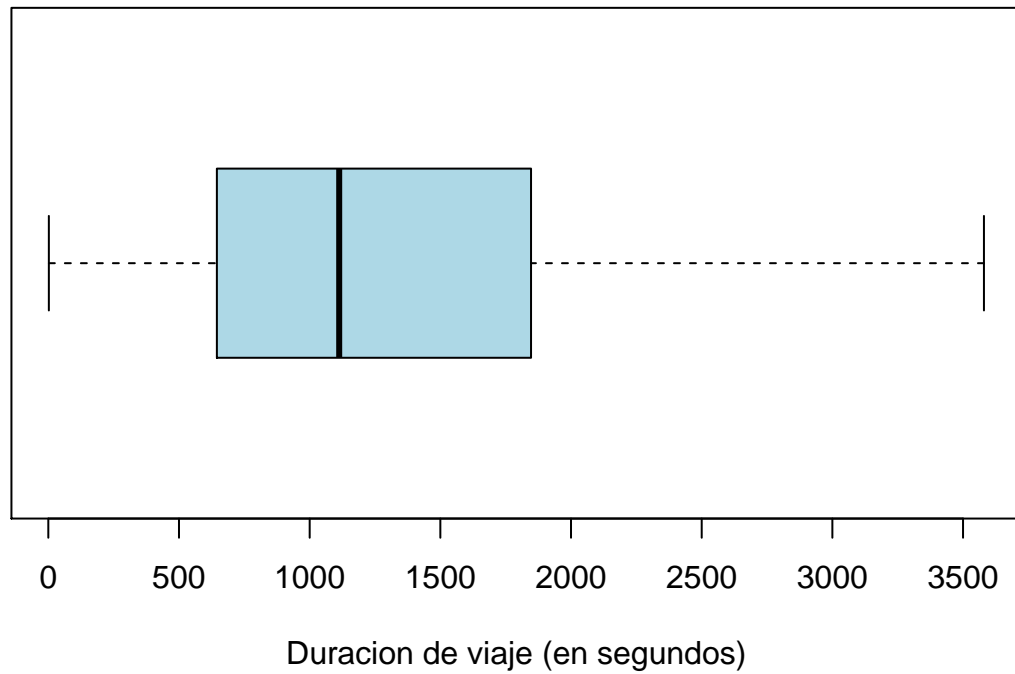
```
intervalos_edades = cut(usuarios_recorridos$edad_usuario, breaks = c(15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65))
frecuencia_absoluta = table(intervalos_edades)
frecuencia_relativa = prop.table(frecuencia_absoluta)
porcentaje = frecuencia_relativa * 100
frecuencia_absoluta_acumulada = cumsum(frecuencia_absoluta)
frecuencia_relativa_acumulada = cumsum(frecuencia_relativa)
porcentaje_acumulado = cumsum(porcentaje)
tabla_edad = cbind(frecuencia_absoluta, frecuencia_relativa, porcentaje, frecuencia_absoluta_acumulada, frecuencia_relativa_acumulada, porcentaje_acumulado)
kable(tabla_edad, caption = "Viajes por Franjas Etarias", col.names = c("Frecuencia Absoluta", "Frecuencia Relativa", "Porcentaje", "Frec abs acumulada", "Frec rel acumulada", "Porc acumulado"))
```

Table 1: Viajes por Franjas Etarias

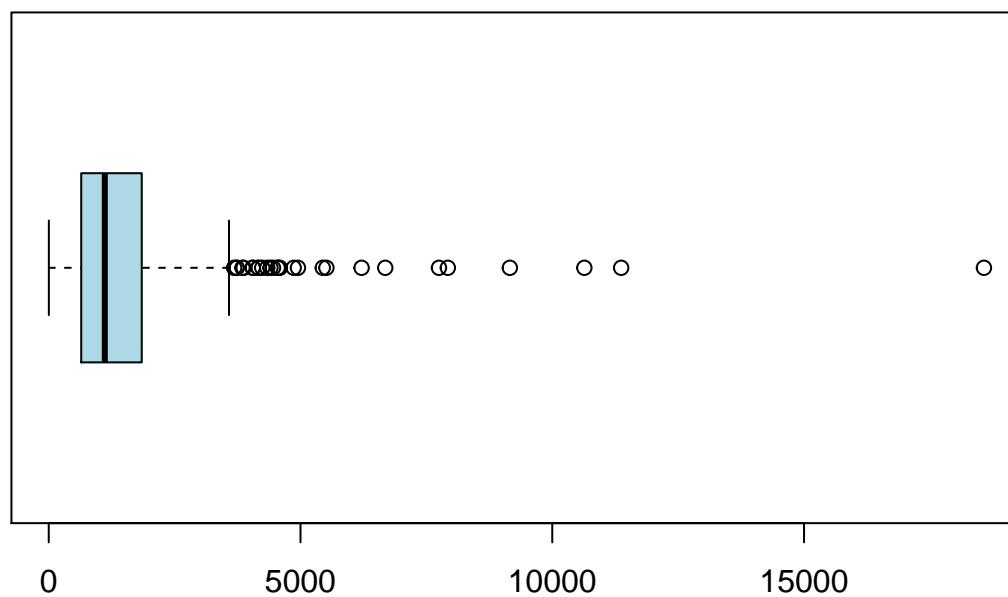
	Frecuencia Absoluta	Frecuencia Relativa	Porcentaje	Frec abs acumulada	Frec rel acumulada	Porc acumulado
(15,20]	56	0.13	13.43	56	0.13	13.43
(20,25]	120	0.29	28.78	176	0.42	42.21
(25,30]	100	0.24	23.98	276	0.66	66.19
(30,35]	58	0.14	13.91	334	0.80	80.10
(35,40]	14	0.03	3.36	348	0.83	83.45
(40,45]	22	0.05	5.28	370	0.89	88.73
(45,50]	16	0.04	3.84	386	0.93	92.57
(50,55]	19	0.05	4.56	405	0.97	97.12
(55,60]	1	0.00	0.24	406	0.97	97.36
(60,65]	11	0.03	2.64	417	1.00	100.00

Boxplot Tiempo de uso

```
boxplot(recorridos$duracion_recorrido, horizontal = TRUE, xlab = "Duracion de viaje (en segundos)", box
```



```
boxplot(recorridos$duracion_recorrido, horizontal = TRUE, xlab = "Duracion de viaje (en segundos)", box
```



Duracion de viaje (en segundos)

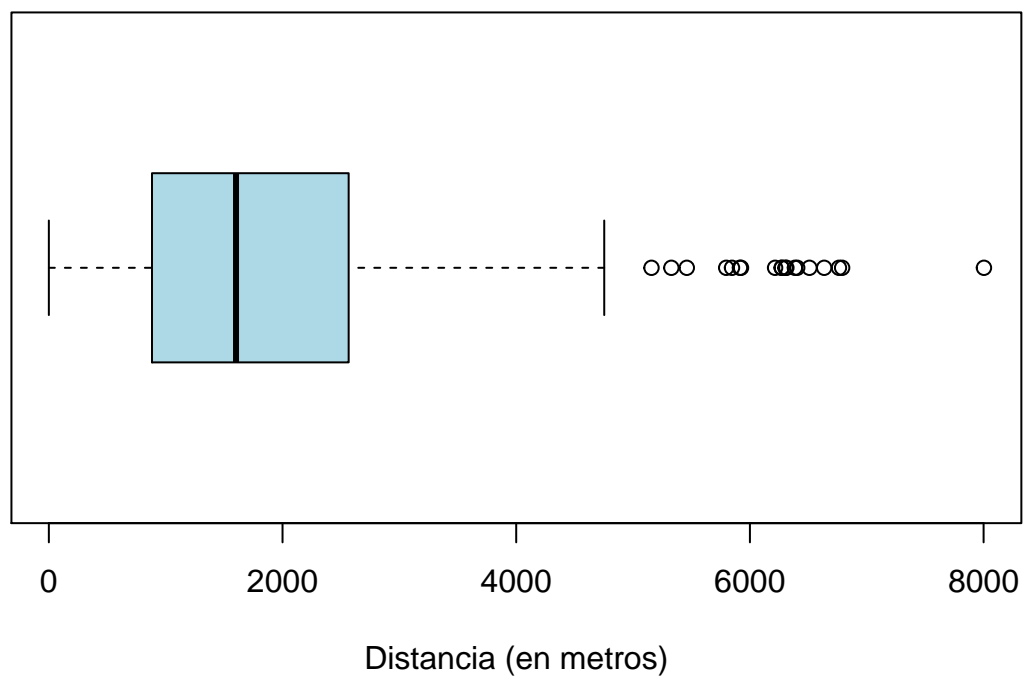
```
intervalos_horas = cut(recorridos$duracion_recorrido, breaks = c(0, 1800, 3600, 5400, 19800), dig.lab =
frecuencia_absoluta = table(intervalos_horas)
frecuencia_relativa = prop.table(frecuencia_absoluta)
porcentaje = frecuencia_relativa * 100
frecuencia_absoluta_acumulada = cumsum(frecuencia_absoluta)
frecuencia_relativa_acumulada = cumsum(frecuencia_relativa)
porcentaje_acumulado = cumsum(porcentaje)
tabla_duracion_recorridos = cbind(frecuencia_absoluta, frecuencia_relativa, porcentaje, frecuencia_absoluta_acumulada,
frecuencia_relativa_acumulada, porcentaje_acumulado)
kable(tabla_duracion_recorridos, caption = "Duraciones de recorridos", col.names = c("Frecuencia Absoluta", "Frecuencia Relativa", "Porcentaje", "Frec abs acumulada", "Frec rel acumulada", "Porc acumulado"))
```

Table 2: Duraciones de recorridos

	Frecuencia Absoluta	Frecuencia Relativa	Porcentaje	Frec abs acumulada	Frec rel acumulada	Porc acumulado
(0,1800]	308	0.74	73.86	308	0.74	73.86
(1800,3600]	81	0.19	19.42	389	0.93	93.29
(3600,5400]	18	0.04	4.32	407	0.98	97.60
(5400,19800]	10	0.02	2.40	417	1.00	100.00

Boxplot Tiempo de uso

```
boxplot(recorridos$distancia, horizontal = TRUE, xlab = "Distancia (en metros)", boxfill="lightblue")
```



```
plot(recorridos$duracion_recorrido~recorridos$distancia,ylab = "Duración",xlab = "Distancia")
```

