

# TP1

## Contents

<b>1</b>	<b>Alumnos</b>	<b>1</b>
<b>2</b>	<b>Introducción</b>	<b>1</b>
<b>3</b>	<b>Objetivos</b>	<b>1</b>
<b>4</b>	<b>Análisis</b>	<b>2</b>
4.1	Análisis univariado . . . . .	2
4.1.1	Edad . . . . .	2
4.1.2	Género . . . . .	5
4.1.3	Duración . . . . .	7
4.1.4	Distancia . . . . .	9
4.1.5	Dirección de origen y dirección de destino . . . . .	10
4.2	Análisis Bivariado . . . . .	13

## 1 Alumnos

Lucas Bachur, Sebastián Mestre, Santiago Coronel.

## 2 Introducción

EcoBici es un sistema de bicicletas compartidas que funciona en la Ciudad de Buenos Aires, cuenta con más de 200 estaciones y 1.200 rodados. Actualmente el sistema está presente en 30 de los 48 barrios de la Ciudad Autónoma de Buenos Aires, funcionando los 365 días del año 24 horas. Existen 2 formas de usar el sistema: mediante una app o mediante una tarjeta. Mediante la app, el usuario debe registrarse y, mediante su nombre de usuario y password, ingresa al sistema, elige la estación en la que se encuentra y luego elige la bici que quiere. Mediante tarjeta, el usuario debe apoyarla en un lector y el sistema le asigna una bicicleta. El usuario se moviliza por la ciudad durante un tiempo limitado (30 minutos) y luego debe devolver la bicicleta en alguna de las estaciones habilitadas. El sistema le asigna al usuario un id y registra datos tales como la edad, el género, estación de origen, estación de destino... etc.

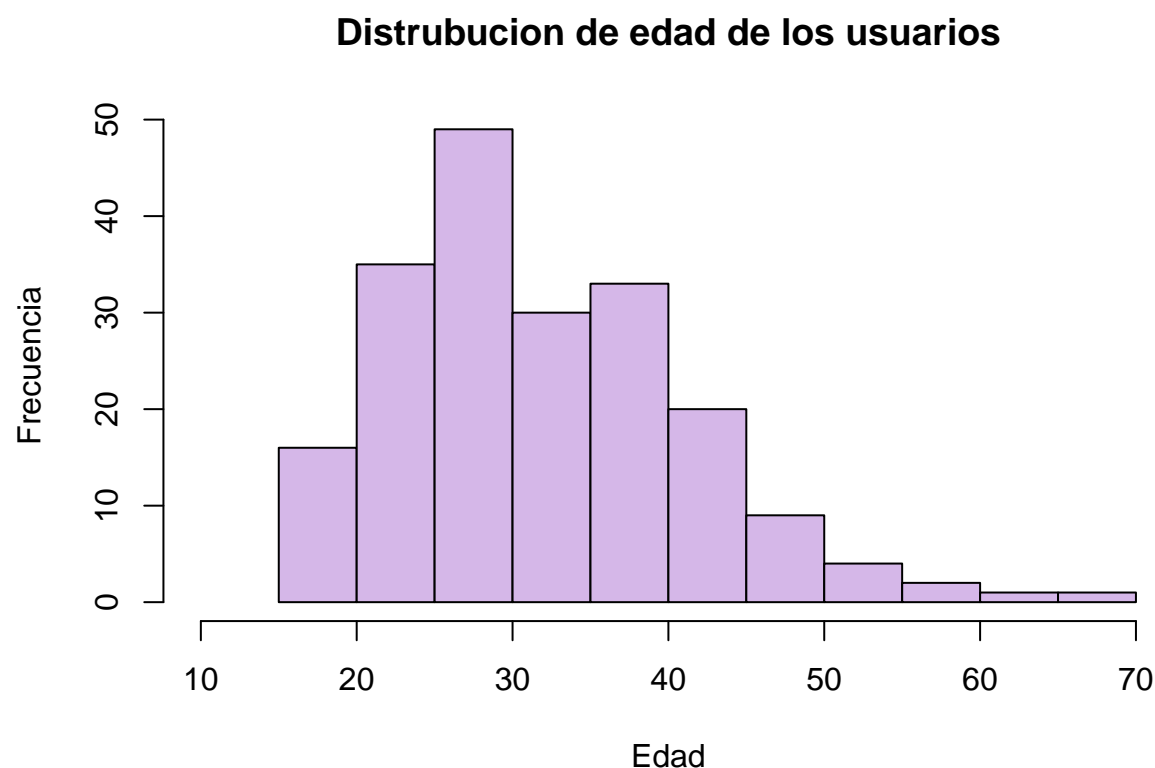
## 3 Objetivos

El objetivo de este informe es reflejar las características más notables acerca de los datos brindados. Para llevar esto a cabo se va a utilizar un software estadístico que permita efectuar un estudio descriptivo que incluya: tablas de distribución de frecuencias, gráficos y medidas descriptivas.

## 4 Análisis

### 4.1 Análisis univariado

#### 4.1.1 Edad

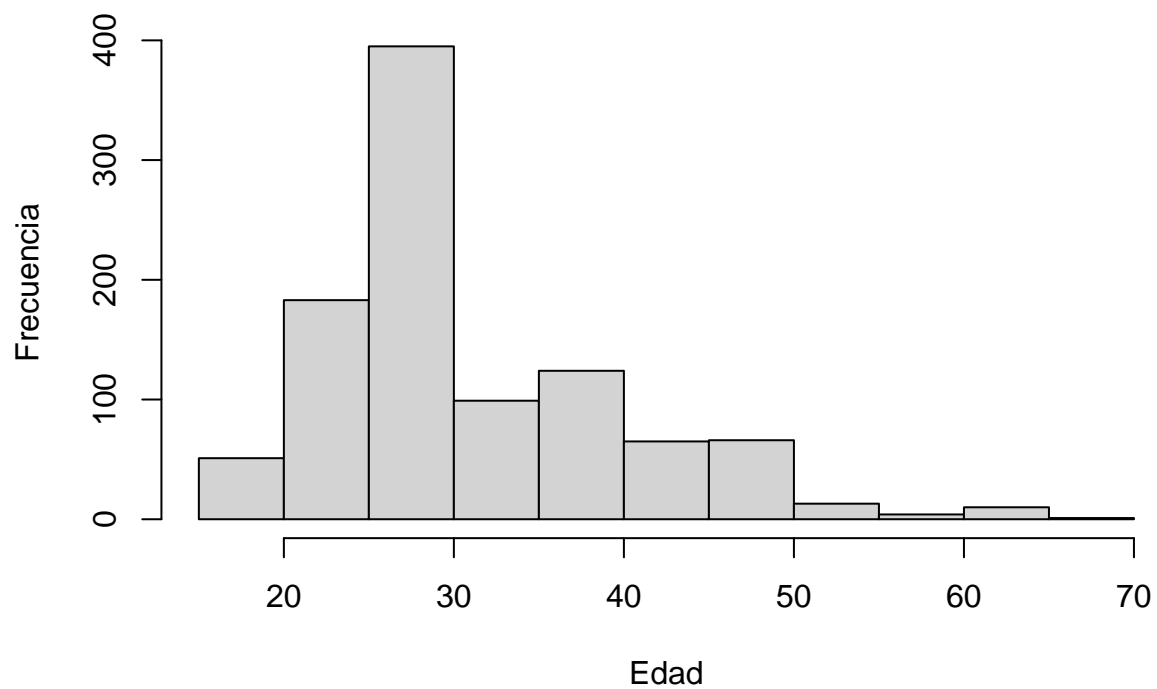


Se puede observar en este histograma que la mayoría de los usuarios están en las franjas etarias más bajas. Más específicamente, a partir de los 40 años la cantidad de usuarios baja considerablemente. Una aclaración importante es que el número de personas que figuran entre 15 y 20 años es bajo, ya que sólo estamos teniendo en cuenta los mayores a 18.

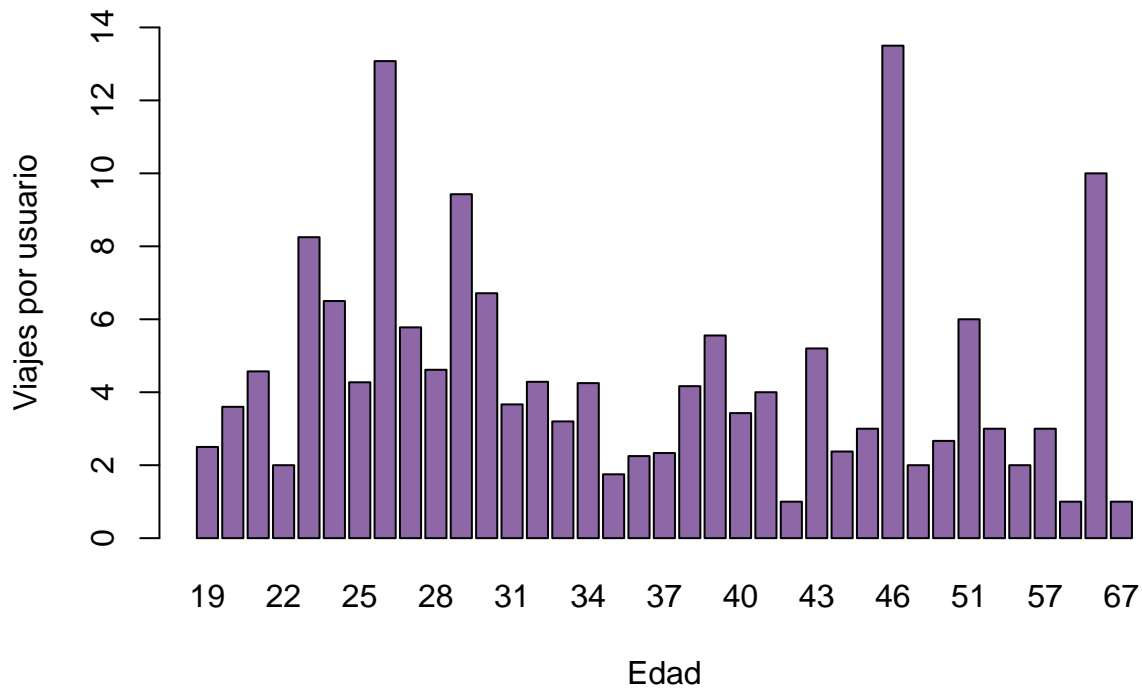
	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada
[19,24)	30	30	0.150	0.150
[24,29)	56	86	0.280	0.430
[29,33)	32	118	0.160	0.590
[33,38)	29	147	0.145	0.735
[38,43)	21	168	0.105	0.840
[43,48)	19	187	0.095	0.935
[48,53)	7	194	0.035	0.970
[53,57)	3	197	0.015	0.985
[57,62)	1	198	0.005	0.990
[62,67)	2	200	0.010	1.000

En esta tabla podemos ver incluso mejor como llegado a los 33 años, la frecuencia acumulada es casi del 60%.

### Recorridos por Edad



## Viajes realizados por usuario, repartidos por edad

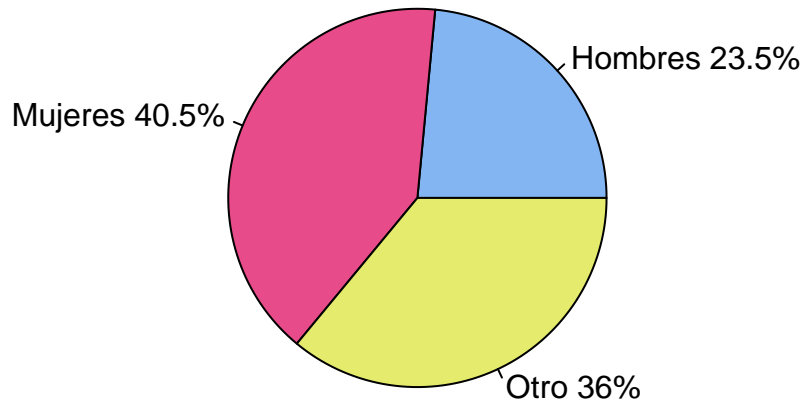


El gráfico de distribución de recorridos por edad nos muestra resultados similares a lo visto en el análisis de usuarios por edad. Sin embargo, una cosa interesante sucede cuando se analiza los viajes realizados por usuario, repartiendolos en cada edad. Se puede ver que si bien hay una leve tendencia a que baje este número, no es nada determinante. O sea, lo que podemos decir es que el hecho de que haya menos viajes realizados por gente mayor a 40 años, no tiene que ver tan directamente con que cada usuario haga menos viajes. Sino que se puede asociar también con el hecho de que son menos los usuarios en esas edades. Esto es lo que termina causando el tamaño de la disparidad.

#### 4.1.2 Género

Se analizará ahora la distrución del género. Algo interesante para acotar es que si un usuario no proporciona esta información, por defecto su género asignado será “Otro”. Esto significa que los datos no representan exactamente la distribución de género real.

**Usuarios por Genero**

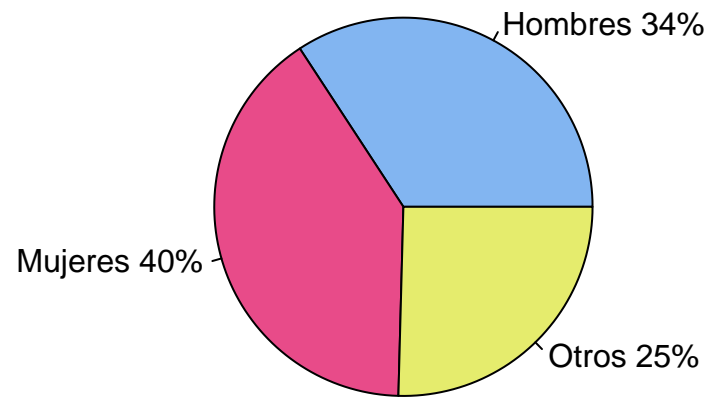


Por ambos gráficos se puede observar que:

- la mayoría de usuarios son mujeres
- la mayoría de viajes lo realizan mujeres

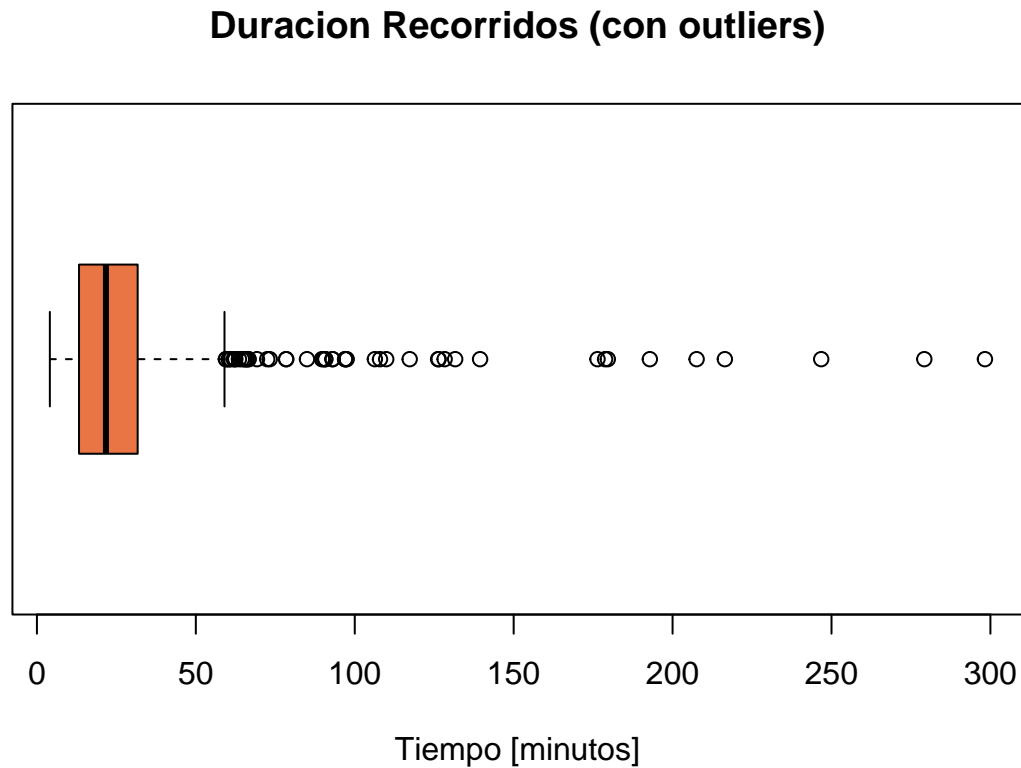
Estos datos tienen una relación directa y se puede decir que la razón por la que más viajes los realizan personas de género femenino es que estos son la mayoría.

## Viajes por Genero



### 4.1.3 Duración

Se hace primero el análisis del gráfico de la distribución de duración con “outliers”:

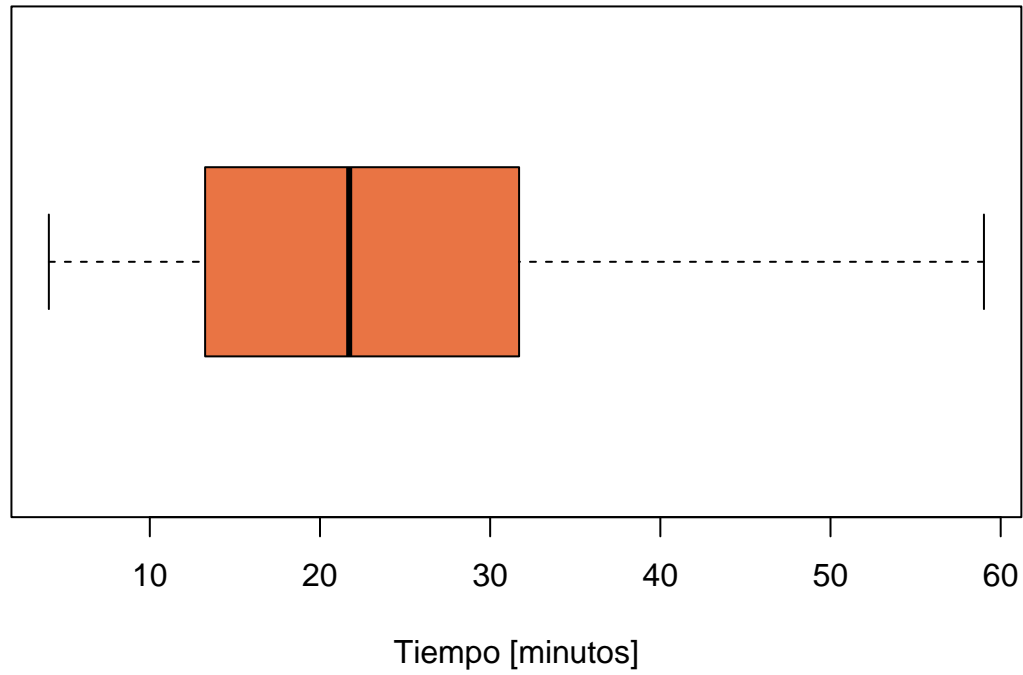


Al observarlo, se nota que hay muchos valores alejados de la media. Estas son algunas de las medidas descriptivas de la variable (en minutos):

- Mínimo: 4
- Primer Cuartil: 13
- Mediana: 22
- Promedio: 27
- Tercer Cuartil: 32
- Máximo: 298

Ahora se analizará el gráfico sin “outliers”:

### Duracion Recorridos (sin outliers)



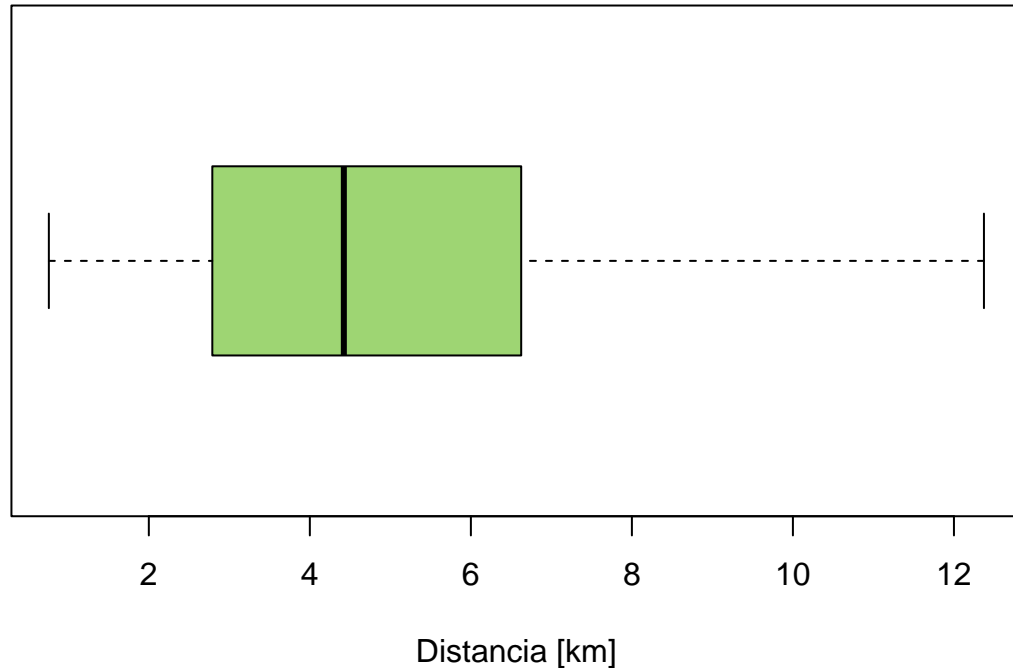
En esta gráfica se puede ver mucho mejor los datos que nos interesan.



#### 4.1.4 Distancia

En este análisis, sucede lo mismo que en el anterior: es conveniente mostrar la gráfica sin outliers. El análisis va a ser directamente de esta.

#### Distancia Recorridos (sin outliers)



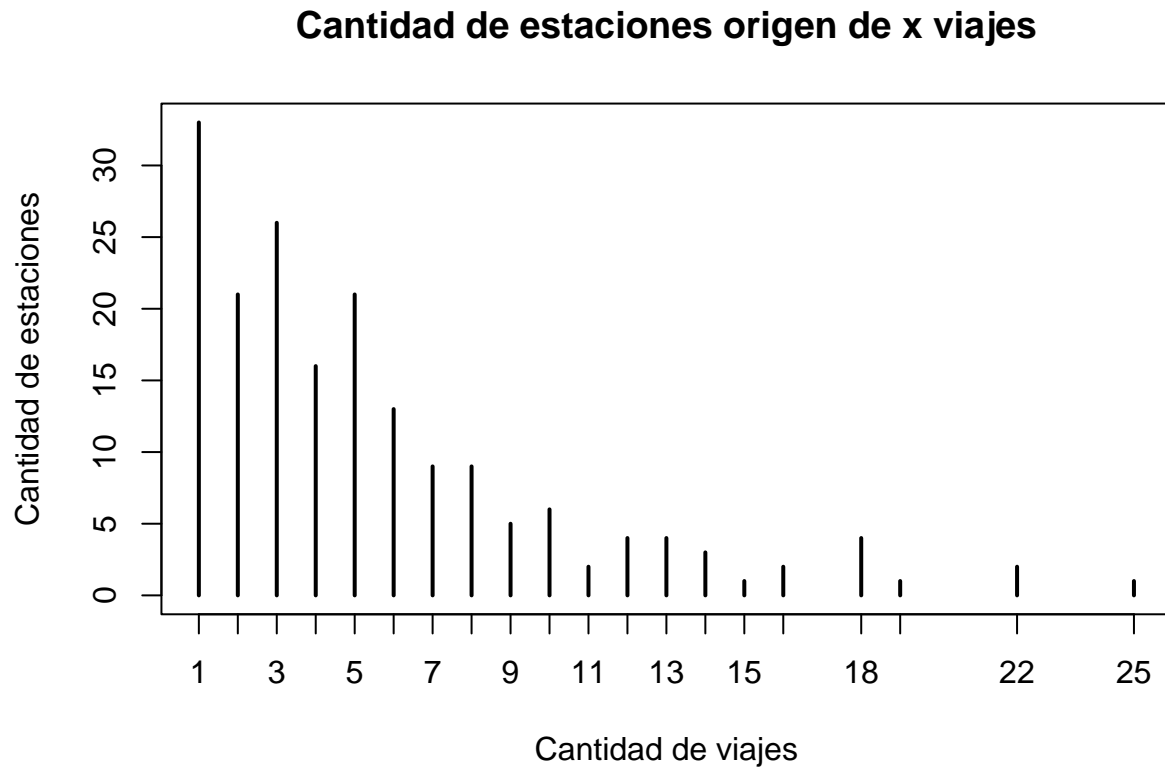
Medidas descriptivas de la variable (en metros):

- Mínimo: 758
- Primer Cuartil: 2788
- Mediana: 4422
- Promedio: 5567
- Tercer Cuartil: 6625
- Máximo: 70528

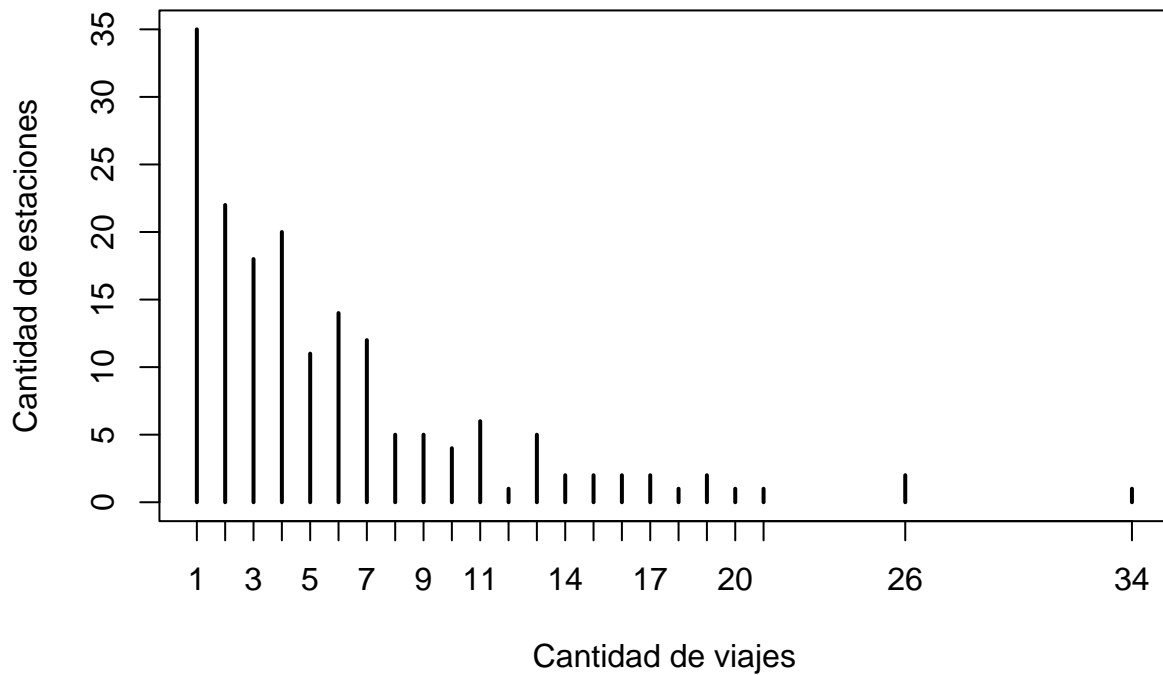
La conclusión que se puede sacar es que la mayoría de los viajes son de media distancia, variando principalmente entre los 3 y 6 kilómetros. No hay gran cantidad de viajes muy cortos ni muy largos, aunque efectivamente existen.

#### 4.1.5 Dirección de origen y dirección de destino

Para esta variable, hacer un análisis alrededor de qué estaciones en específico tienen mayor o menor tráfico no va a ser de demasiada utilidad. Así que el análisis va a ser alrededor de **cuántas** fueron salida o llegada de cada cantidad de viajes:



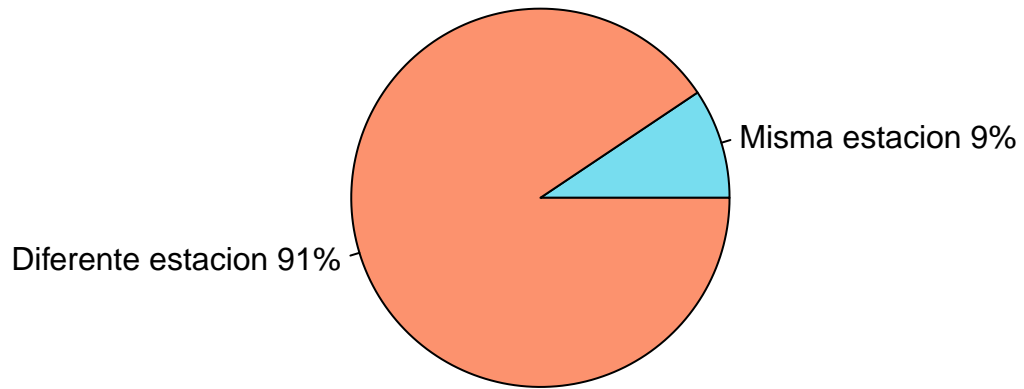
### Cantidad de estaciones destino de x viajes



Como se puede ver, hay unos pocos “outliers” en estos gráficos también, pero la información se puede ver bien igualmente. Lo que se puede notar de estos gráficos es que no hay muchas estaciones que concentren gran cantidad de tráfico. Si hay lugar para mejorar el sistema va a ser en las estaciones que tienen estos valores extremos.

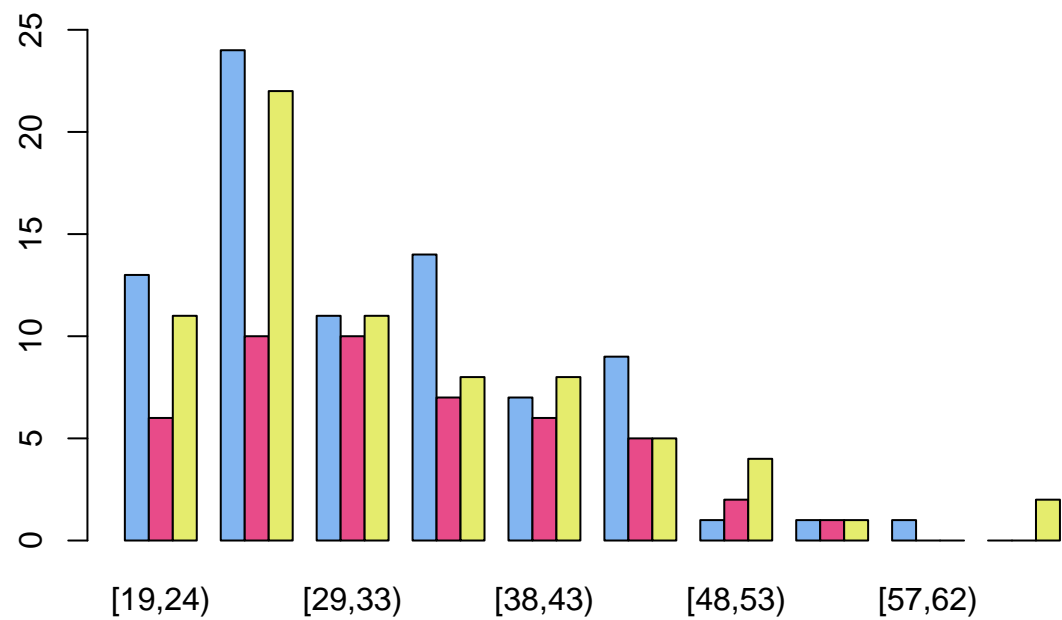
El otro foco de análisis con estas variables es cuántos viajes empiezan y terminan en la misma estación. En el gráfico de abajo se ve que la mayoría de los recorridos comienzan y acaban en diferentes lugares. Si un viaje cumple con esta condición, lo más probable es que sea del tipo recreativo. Así que la conclusión es que la mayoría de la gente no hace este tipo de viajes.

## Distribución de viajes que empiezan y terminan en el mismo lugar



## 4.2 Análisis Bivariado

El estudio se realizó relacionando la edad de los usuarios con su género.



Al observar la gráfica se puede notar que el rango de 20 a 30 años está mayormente representado por hombres. El resto de las edades tienen más repartidos los géneros.

[19,24)	[24,29)	[29,33)	[33,38)	[38,43)	[43,48)	[48,53)	[53,57)	[57,62)	[62,67)
13	37	48	62	69	78	79	80	81	81
6	16	26	33	39	44	46	47	47	47
11	33	44	52	60	65	69	70	70	72

Y con esta tabla se pueden ver bien las distribuciones acumuladas.