



TP Estadística Descriptiva

Tomás Santucci y Fernando Silva

18 de junio de 2022

Contents

Introducción	2
Análisis univariado	3
Variable género (Gráfico de Sectores)	3
Variable edad (Histograma)	3
Variable días (Tabla de Frecuencias más Gráfico de Barras)	5
Variable estación de origen (Tabla de Frecuencias más Gráfico de Barras)	6
Variable estación de Destino (Tabla de Frecuencias más Gráfico de Barras)	8
Variable distancia	10
Variable duración	11
Variable cantidad de viajes por usuario	12
Análisis bivariado	14
Conclusiones	15

Introducción

Objetivo

El objetivo del siguiente informe es mostrar estadísticamente el uso de las ecobicis de la Ciudad Autónoma de Buenos Aires. Además de brindar datos y gráficos (univariados y bivariados) para tener una apreciación de la utilidad de las bicicletas y del comportamiento de los usuarios. Se muestran falencias y posibles mejoras al sistema.

Sobre las muestras

Se utilizaron dos datasets. El primero es sobre los usuarios que utilizaron el sistema. Es de tamaño 200 y brinda el género y la edad de cada uno. También identifica a cada uno con un id.

El segundo es sobre los recorridos. Este es de tamaño 811. Dice la estación donde se originó y concluyó, la duración, la distancia y el día en el que se efectuó. Además de decir qué id de usuario realizó cada viaje.

Análisis univariado

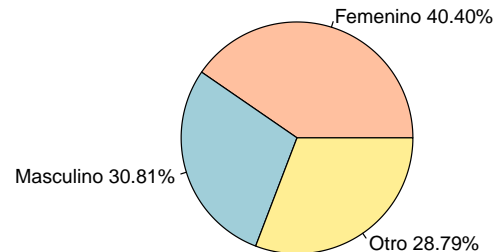
Variable género (Gráfico de Sectores)

Es una variable cualitativa cuyos valores posibles son: Masculino, Femenino y Otro. En este caso, la unidad de análisis son los usuarios. El gráfico pasa a representar el género de los usuarios que han utilizado las ecobici.

Table 1: Tabla de frecuencias

Genero	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
F	80	0.404	80	0.404
M	57	0.2879	137	0.6919
OTRO	61	0.3081	198	1

Gráfico 1: GENERO DE LOS USUARIOS DE ECOBICI CABA 2020



Fuente: Datos sustraídos de las estaciones de las bicicletas públicas de CABA

En el Gráfico 1 se puede apreciar que un gran porcentaje de usuarios de las Ecobicis son de género “Femenino”, más que aquellos que pertenecen al “Masculino”. Sobre el género “Otro” no se puede tener demasiada apreciación debido a que el término es muy abarcativo. Es interesante notar que si éste último fuera la categoría por defecto, y quienes dieron sus datos no la cambiaron siendo de géneros “Masculino” o “Femenino”, al obtener un porcentaje tan alto es posible que en caso de registrarse correctamente cambien las conclusiones.

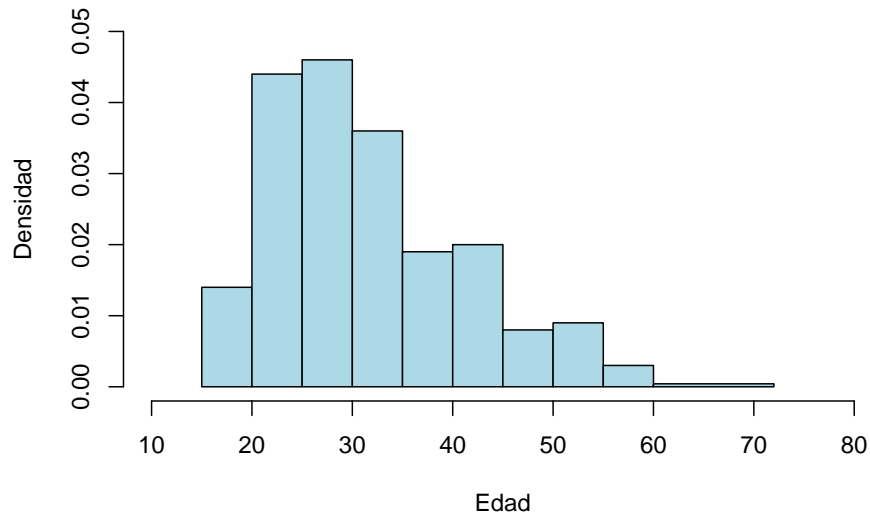
Medida de interés:

- Moda : Femenino

Variable edad (Histograma)

Es una variable cuantitativa continua cuyos valores están entre 16 y 72. Representa la edad de los usuarios de ecobici. En este caso, la unidad de análisis son los usuarios.

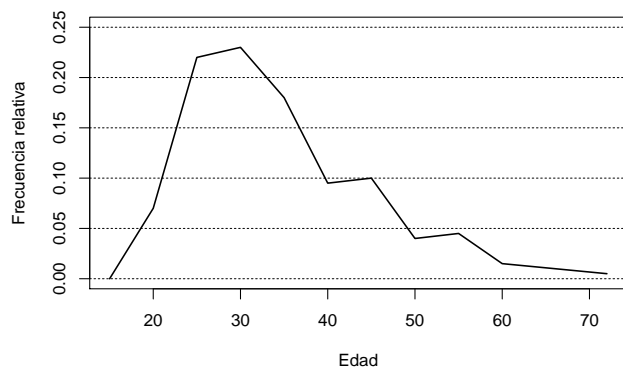
**Grafico 2: EDAD DE LOS USUARIOS DE ECOBICI
CABA 2020**



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

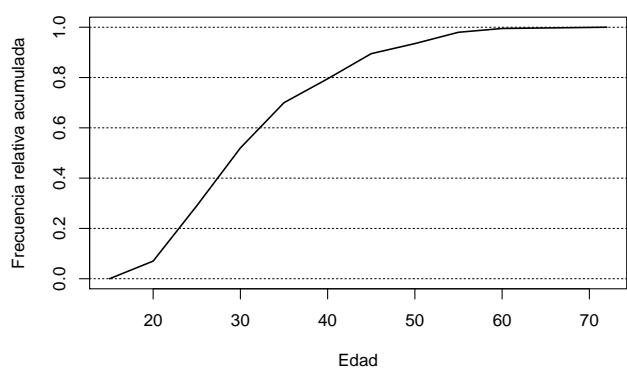
Los datos resultan en un histograma unimodal asimétrico a derecha.

**Grafico 2.1: EDAD DE LOS USUARIOS DE ECOBICI
CABA 2020**



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

**Grafico 2.2: EDAD DE LOS USUARIOS DE ECOBICI
CABA 2020**



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

En el Gráfico 2 se muestra la variación del uso del servicio respecto de la edad de los usuarios. En el Gráfico 2.1 y 2.2, se muestran los polígonos de Frecuencias Relativas y Absolutas de los usuarios respecto de la edad de los mismos.

En el Gráfico 2.1, se puede ver claramente el aumento del uso de las ecobici, de manera creciente entre los usuarios cuyas edades van desde 20 a 30 años y luego va decreciendo a medida que aumenta la edad de los mismos. En otras palabras los usuarios con edad en el rango [20,35] son los que mas utilizan el servicio, que aquellos usuarios mayores.

Complementariamente, definimos una variable “Grupo etario” cualitativa cuyos valores posibles son: Adulto menor, de 16 a 25 años. Adulto medio, de 26 a 45 años. Adulto mayor, de +45 años

Table 2: Tabla de frecuencias

Grupo etario	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Adulto mayor	84	0.1036	84	0.1036

Grupo etario	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Adulto medio	474	0.5845	558	0.688
Adulto menor	253	0.312	811	1

En la tabla se observa que la mayoría de usuarios son “Adultos medios” (alrededor del 58% de los 811 totales).

Medidas de interés:

- Mínimo: 16 años
- Mediana: 30 años
- Media: 31.97 años
- Desvío estandar: 9.89 años (desviación “típica” de la media)
- Moda: 22 años
- Máximo: 72 años

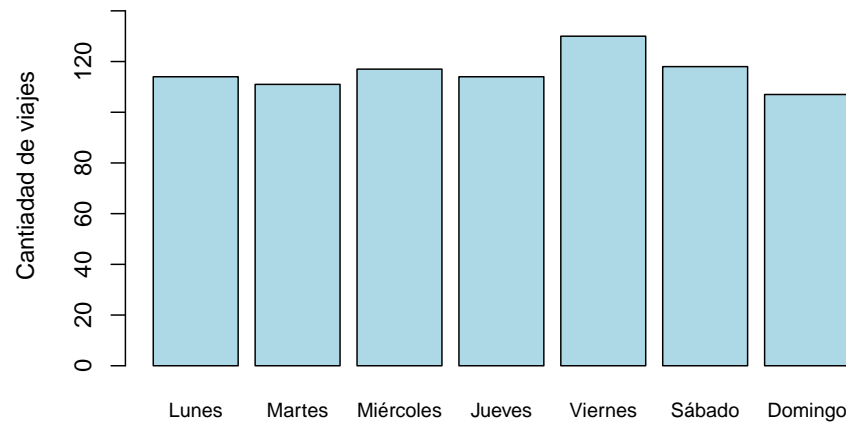
Variable días (Tabla de Frecuencias más Gráfico de Barras)

Es una variable cualitativa que representa los días en que se efectuaron los viajes. En este caso, la unidad de análisis son los recorridos.

Table 3: Tabla de frecuencias

Estación	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Lunes	114	0.1406	114	0.1406
Martes	111	0.1369	225	0.2774
Miércoles	117	0.1443	342	0.4217
Jueves	114	0.1406	456	0.5623
Viernes	130	0.1603	586	0.7226
Sábado	118	0.1455	704	0.8681
Domingo	107	0.1319	811	1

**Grafico 3: DIA DE LOS RECORRIDOS DE LAS ECOBICI
CABA 2020**



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

En el Gráfico 3 se puede observar que durante todo el año 2020 se efectuaron mas viajes los días Viernes que en el resto de la semana. Se puede suponer que ese día se efectúan dicha cantidad de viajes, debido a que es el último día laboral de la semana y la mayoría de los usuarios opta por este servicio para poder regresar a sus casas a tiempo.

Medida de interés:

- Moda: Viernes

Variable estación de origen (Tabla de Frecuencias más Gráfico de Barras)

Es una variable cualitativa. Representa a las estaciones donde se originaron los viajes.

Se representan las estaciones más utilizadas, debido a la gran cantidad de estaciones de origen. Hay 169 estaciones donde se iniciaron viajes, en total.

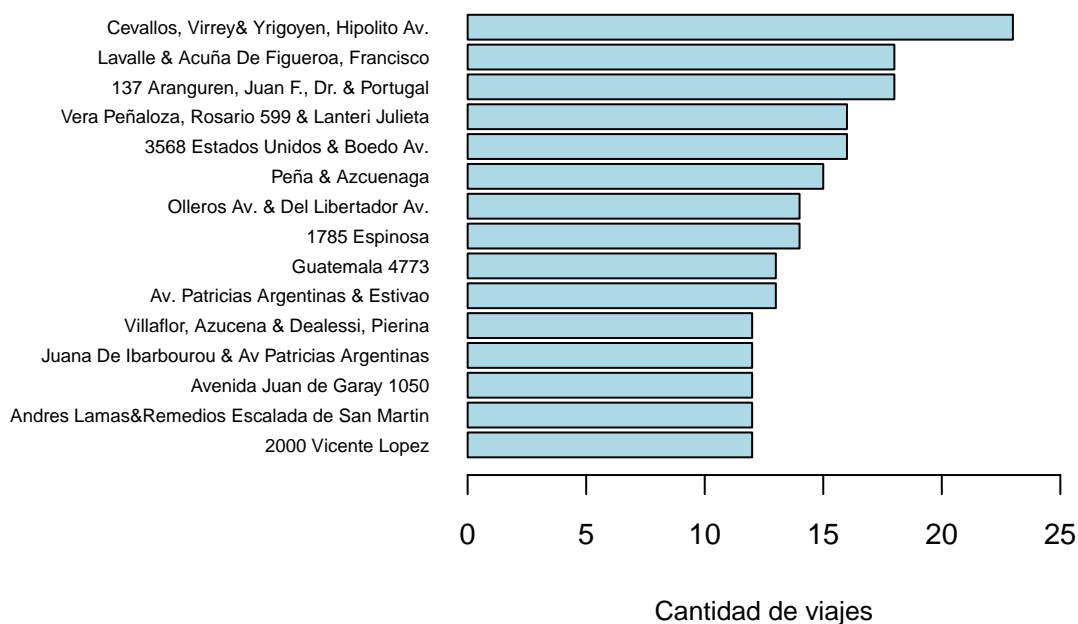
En este caso, la unidad de análisis son los recorridos.

Table 4: Tabla de frecuencias

Estación	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Cevallos, Virrey& Yrigoyen, Hipolito Av.	23	0.0284	23	0.0284
137 Aranguren, Juan F., Dr. & Portugal	18	0.0222	41	0.0506
Lavalle & Acuña De Figueroa, Francisco	18	0.0222	59	0.0727
3568 Estados Unidos & Boedo Av.	16	0.0197	75	0.0925
Vera Peñaloza, Rosario 599 & Lanteri Julieta	16	0.0197	91	0.1122
Peña & Azcuenaga	15	0.0185	106	0.1307
1785 Espinosa	14	0.0173	120	0.148
Olleros Av. & Del Libertador Av.	14	0.0173	134	0.1652

Estación	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Av. Patricias Argentinas & Estivao	13	0.016	147	0.1813
Guatemala 4773	13	0.016	160	0.1973
2000 Vicente Lopez	12	0.0148	172	0.2121
Andres Lamas&Remedios	12	0.0148	184	0.2269
Escalada de San Martin				
Avenida Juan de Garay 1050	12	0.0148	196	0.2417
Juana De Ibarbourou & Av	12	0.0148	208	0.2565
Patricias Argentinas				
Villaflor, Azucena & Dealessi, Pierina	12	0.0148	220	0.2713

Grafico 4: ESTACIONES DE ORIGEN MAS USADAS DE LAS ECOBICI, CABA 2020



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

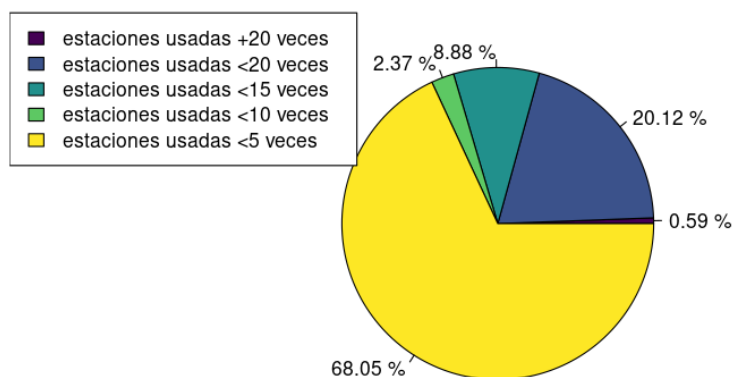
Se puede ver a partir de la tabla que alrededor del 27% de los viajes se originaron en sólo 15 de las 169 estaciones.

Medida de interés:

- Moda: Cevallos, Virrey & Yrigoyen, Hipolito Av.

Complementariamente, se pueden definir categorías para la variable anterior, de manera de poder visualizar cuántas estaciones no se utilizaron tanto.

**GRÁFICO 4.1: CANTIDAD DE VECES QUE SE UTILIZÓ UNA ESTACIÓN COMO ORIGEN
CABA 2020**



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

Del gráfico 4.1 se puede ver que alrededor del 68% de las estaciones fueron utilizadas menos de 5 veces. En otras palabras, la mayoría de las estaciones de origen fueron utilizadas relativamente poco, mientras que alrededor de un tercio de ellas fueron las que se usaron repetidas veces.

Variable estación de Destino (Tabla de Frecuencias más Gráfico de Barras)

Es una variable cualitativa. Representa a las estaciones de destino de cada viaje.

Se representan las estaciones más utilizadas, debido a la gran cantidad de estaciones de destino. Hay 171 estaciones donde se concluyeron viajes, en total.

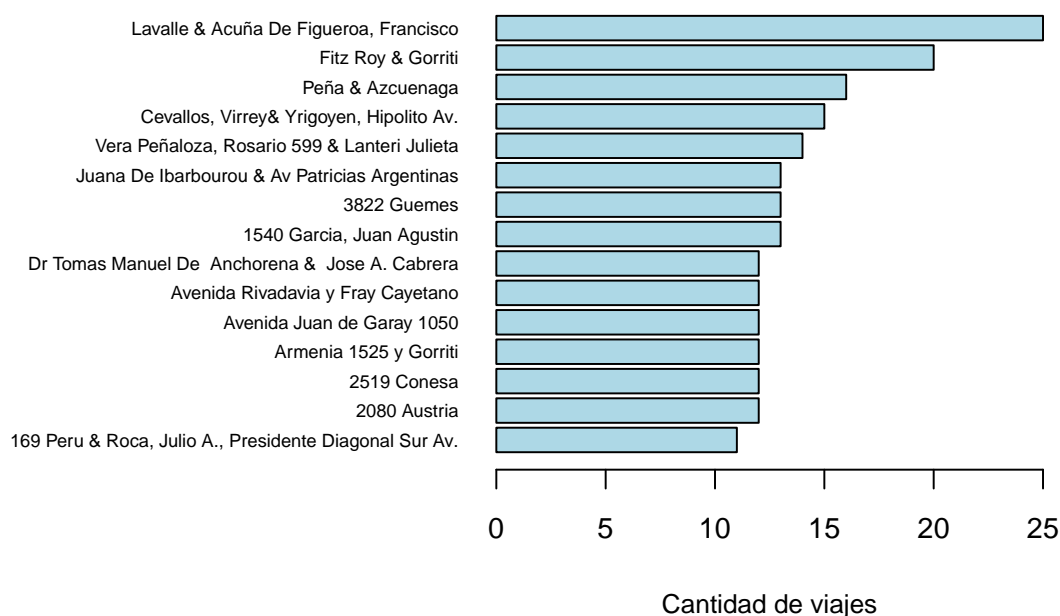
En este caso, la unidad de análisis son los recorridos.

Table 5: Tabla de frecuencias

Estación	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Lavalle & Acuña De Figueroa, Francisco	25	0.0308	25	0.0308
Fitz Roy & Gorriti	20	0.0247	45	0.0555
Peña & Azcuenaga	16	0.0197	61	0.0752
Cevallos, Virrey& Yrigoyen, Hipolito Av.	15	0.0185	76	0.0937
Vera Peñaloza, Rosario 599 & Lanteri Julieta	14	0.0173	90	0.111
1540 Garcia, Juan Agustin	13	0.016	103	0.127
3822 Guemes	13	0.016	116	0.143

Estación	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Juana De Ibarbourou & Av Patricias Argentinas	13	0.016	129	0.1591
2080 Austria	12	0.0148	141	0.1739
2519 Conesa	12	0.0148	153	0.1887
Armenia 1525 y Gorriti	12	0.0148	165	0.2035
Avenida Juan de Garay 1050	12	0.0148	177	0.2182
Avenida Rivadavia y Fray Cayetano	12	0.0148	189	0.233
Dr Tomas Manuel De Anchorena & Jose A. Cabrera	12	0.0148	201	0.2478
169 Peru & Roca, Julio A., Presidente Diagonal Sur Av.	11	0.0136	212	0.2614

Grafico 5: ESTACIONES DESTINO MAS USADAS DE LAS ECOBICI, CABA 2020



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

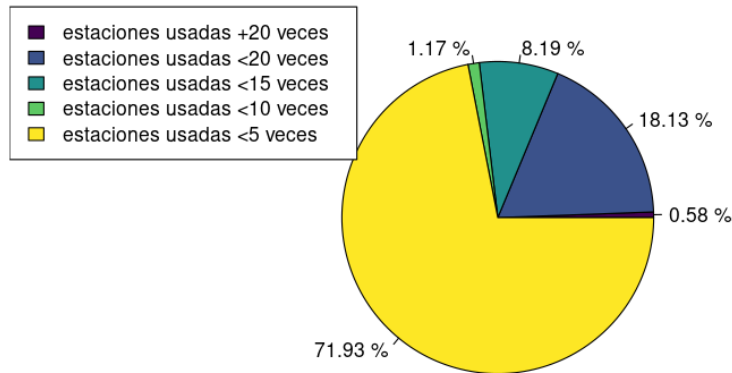
Se puede ver a partir de la tabla que, alrededor de 26% de los viajes tuvieron como destino sólo 15 de las 171 estaciones.

Medida de interés:

- Moda: Lavalle & Acuña De Figueroa, Francisco

Complementariamente, se pueden definir categorías para la variable anterior, de manera de poder visualizar cuántas estaciones no se utilizaron tanto.

**GRÁFICO 5.1: CANTIDAD DE VECES QUE SE UTILIZÓ UNA ESTACIÓN COMO DESTINO
CABA 2020**



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

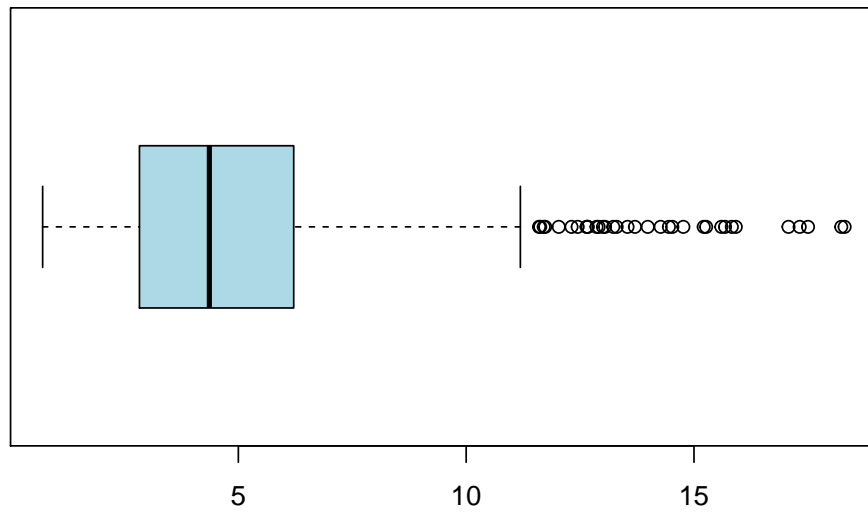
Del gráfico 5.1 se puede ver que alrededor del 72% de las estaciones fueron utilizadas menos de 5 veces. En otras palabras, la mayoría de las estaciones de origen fueron utilizadas relativamente poco, mientras que alrededor de un tercio de ellas fueron las que se usaron repetidas veces.

Variable distancia

Es una variable cuantitativa continua. Sus valores están entre 0 y 20 km. Cualquier apreciación mayor a 20 km es considerada un error de sistema o humano y por lo tanto no es tomada en cuenta en el análisis.

En este caso, la unidad de análisis son los recorridos.

**Grafico 6: DISTANCIA DE LOS RECORRIDOS
DE ECOBICI, CABA 2020**



Distancia en kilometros

Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

En el Gráfico 6 se puede ver que:

El 25% recorrió menos de 2.83 km.

El 50% recorrió menos de 4.36 km.

El 75% recorrió menos de 6.21 km.

Medidas de interés:

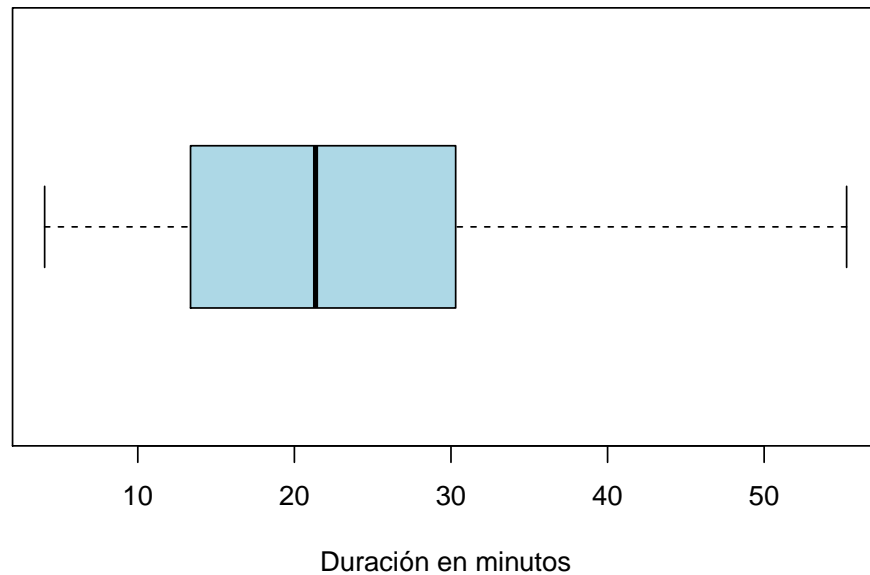
- Mínimo: 0.70 km
- Mediana: 4.36 km
- Media: 4.96 km
- Desvío estandar: 3.06 km (desviación “típica” de la media de 3.06 km)
- Máximo: 18.30 km (outlier)

Variable duración

Es una variable cuantitativa continua. Sus valores están entre 4 min y 90 min. Cualquier apreciación mayor a 90 minutos es considerada un error de sistema o humano y por lo tanto no es tomada en cuenta en el análisis.

En este caso, la unidad de análisis son los recorridos.

**Gráfico 7: DURACION DE LOS RECORRIDOS
DE ECOBICI, CABA 2020**



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

A partir del Gráfico 7 se puede ver que:

El 25% de los usuarios utilizó el servicio menos de 13 min

El 50% de los usuarios utilizó el servicio menos de 21 min

El 75% de los usuarios utilizó el servicio menos de 30 min

Es interesante notar que un 25% de los viajes se excedieron del tiempo límite de 30 minutos.

Medidas de interés:

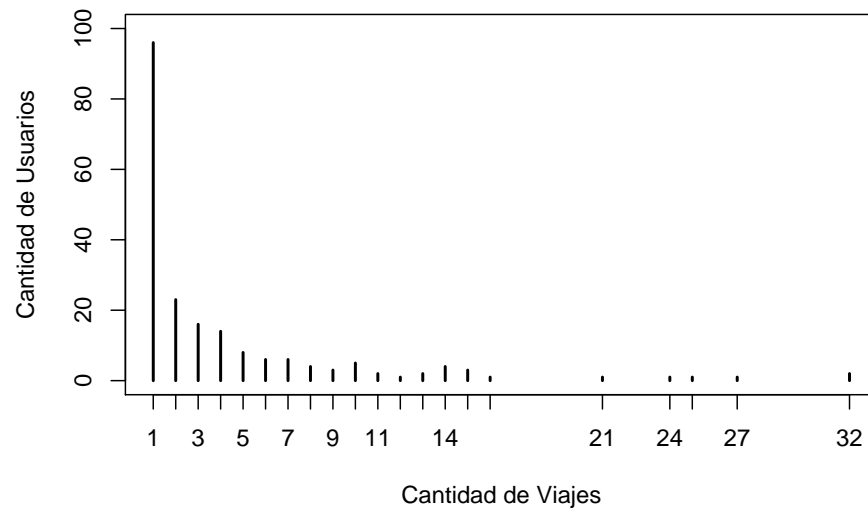
- Mínimo: 4.05 min
- Mediana: 21.35 min
- Media: 23.84 min
- Desvío estandar: 14.18 min (desviación “típica” de la media de 14.18 minutos)
- Máximo: 87.37 min (outlier)

Variable cantidad de viajes por usuario

Es una variable cuantitativa discreta, cuyos valores posibles estan entre 1 y 32 viajes.

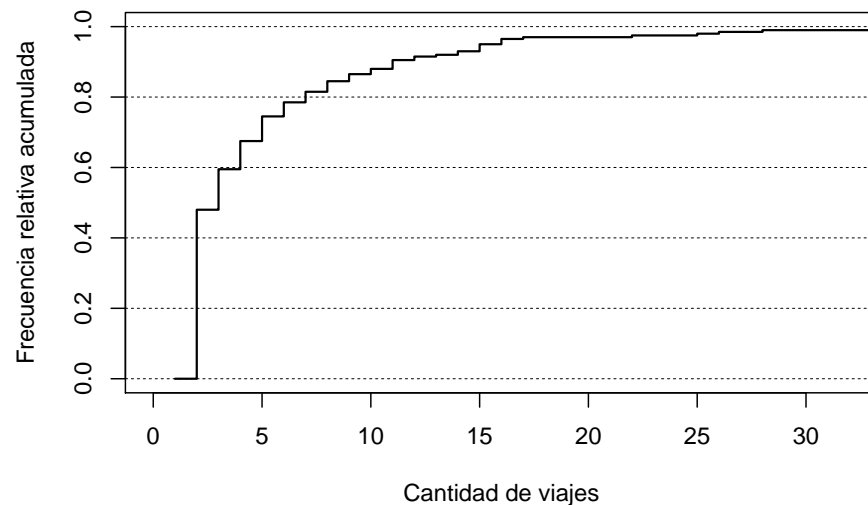
En este caso, la unidad de análisis son los usuarios.

**Grafico 8: CANTIDAD DE VIAJES POR USUARIO
CABA 2020**



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

**Grafico 9: FRECUENCIA RELATIVA ACUMULADA DE LA
CANTIDAD DE VIAJES POR USUARIO
CABA 2020**



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

A partir del gráfico 8 de bastones se puede ver que:

96 Usuarios realizaron un solo viaje.

Solamente 2 usuarios realizaron 32 viajes.

Del gráfico 9 se concluye que alrededor del 50% de usuarios realizaron sólo 1 viaje.

Medida de interes:

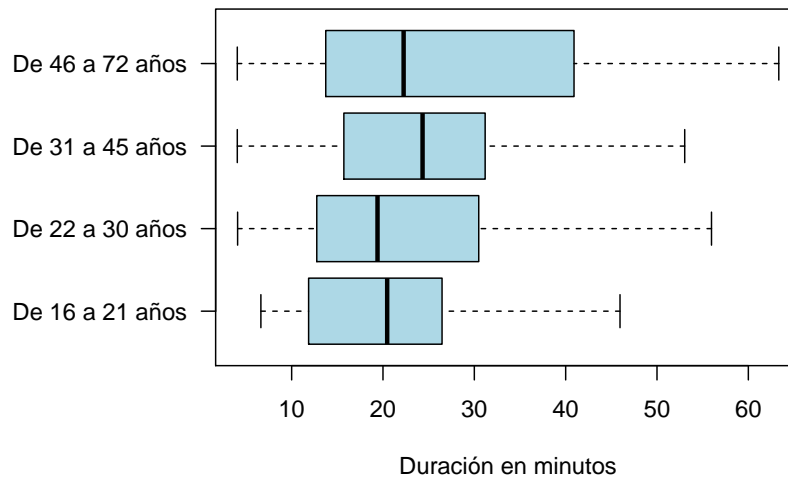
- Moda: 1 viaje

Análisis bivariado

Se realiza un análisis de las variables (cuantitativas continuas) edad y distancia recorrida. Se divide la edad a partir de 4 rangos etarios por lo que se transforma en una variable cualitativa.

En este caso, la unidad de análisis son los recorridos (por la distancia) y los usuarios (por la edad).

Grafico 10: DURACION DEL RECORRIDO SEGUN LA EDAD DEL USUARIO, CABA 2020



Fuente: Datos sustraídos de las estaciones de las bicicletas publicas de CABA

El rango de la variable en el grupo de 46 a 72 años es el más extenso y tiene el mayor máximo.

Se podría creer que una posible explicación sería que la gente mayor no acredita correctamente el fin de sus viajes.

A medida que avanzan las edades se puede ver que cada vez más gente se acerca e incluso sobrepasa el límite de 30 minutos. Siendo el grupo etario de 16 a 21 años es el que más respeta el tiempo límite. Y el grupo de 46 a 72 años es el que más lo excede.

Conclusiones

A partir de los datos analizados se pueden sacar las siguientes conclusiones:

Quiénes usan las bicicletas:

Los usuarios del servicio de bicicletas suelen ser gente de entre 20 y 40 años, de ambos géneros aunque hay una pequeña mayoría de mujeres.

Cuándo y dónde se usan:

El uso de las mismas no fluctua demasiado durante la semana pero si hay una diferencia grande en el uso de las estaciones. En unas pocas se realizaron una gran cantidad de los viajes, lo que indicaría que allí sería conveniente poner más bicicletas que en el resto de ellas.

Cómo se utilizan:

Por último es interesante notar que hay muchos viajes que son largos tanto en distancia como en duración. Esto podría indicar que dentro de los usuarios no se conoce bien el funcionamiento del servicio y por ello se pasan del tiempo permitido de 30 minutos o no acreditan bien los viajes al finalizarlos. Además, se pudo ver que los usuarios de mayor edad son los que más se excedieron en este aspecto.