



PRESENTACIÓN ACÁMICA: Dataset EcoBici “Recorridos Realizados 2018”

por Francisco Paterson

ECOBICI: EL SISTEMA PÚBLICO DE BICICLETAS PORTEÑO

Según Wikipedia:

EcoBici (...) cuenta con 400 estaciones y 4000 rodados y en 2018 más de 200.000 usuarios realizaron más de 8 millones de viajes durante el año. Una característica no usual en este tipo de sistema es que su uso, que requiere un registro previo, es totalmente gratuito tanto para residentes locales como para turistas.*

A partir de 2019, el sistema es operado por la empresa brasileña Tembici (...) Actualmente el sistema está presente en 38 de los 48 barrios de la Ciudad Autónoma de Buenos Aires, funcionando los 365 días del año 24 horas. (Wikipedia, consultado el 24/10/2019)



Funcionamiento horario durante las 24 horas. Edad mínima para inscripción: 16 años.

DATASET

Página de referencia: <https://data.buenosaires.gob.ar/dataset/bicicletas-publicas>

Fuente primaria: Ministerio de Desarrollo Urbano y Transporte. SS de Movilidad Sustentable y Segura.

Mantenedor: Secretaría General y Relaciones Internacionales. Jefatura de Gobierno GCBA.

Licencia Atribución 2.5 Argentina (CC BY 2.5 AR)

ECOBICI: EL SISTEMA PÚBLICO DE BICICLETAS PORTEÑO

1) ¿Cómo empezar? Primera aproximación al dataset

Primero me propuse ver información general sobre el dataset, de manera de tener una idea general sobre el mismo (cómo se ve, cantidad de columnas y sus valores, cantidad de filas, tipos de valores, número de NaNs y alguna información sobre su contenido específico).

```
In [6]: 1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2619968 entries, 0 to 2619967
Data columns (total 9 columns):
bici_id_usuario          int64
bici_fecha_hora_retiro    datetime64[ns]
bici_tiempo_uso           object
bici_nombre_estacion_origen  object
bici_estacion_origen      int64
bici_nombre_estacion_destino object
bici_estacion_destino      int64
bici_sexo                 object
bici_edad                  int64
dtypes: datetime64[ns](1), int64(4), object(4)
```

**9 columnas, 3 tipos de datos
y 2.619.968 filas**

```
In [9]: 1 df.isna().sum()
```

```
Out[9]: bici_id_usuario          0
bici_fecha_hora_retiro          0
bici_tiempo_uso                 43723
bici_nombre_estacion_origen     0
bici_estacion_origen            0
bici_nombre_estacion_destino    0
bici_estacion_destino           0
bici_sexo                       0
bici_edad                       0
dtype: int64
```

43.723 NaNs en una sola columna



ECOBICI: EL SISTEMA PÚBLICO DE BICICLETAS PORTEÑO



Así se ven las diez primeras instancias:

In [4]: 1 df.head(10)

Out[4]:

	bici_id_usuario	bici_fecha_hora_retiro	bici_tiempo_uso	bici_nombre_estacion_origen	bici_estacion_origen	bici_nombre_estacion_destino	bici_estacion_destino	bicisexo	biciedad
0	5453	2018-01-01 00:08:05	0 days 00:19:53.000000000	Uruguay	45	Virrey Cevallos	183	M	45
1	673	2018-01-01 00:18:05	0 days 00:26:19.000000000	Posadas	189	Guardia Vieja	110	M	61
2	179119	2018-01-01 00:20:14	0 days 00:27:39.000000000	Hospital Rivadavia	50	Padilla	31	F	52
3	400147	2018-01-01 00:20:22	0 days 00:48:51.000000000	Macacha Güemes	111	Acuña de Figueroa	54	M	27
4	400156	2018-01-01 00:20:31	0 days 00:49:27.000000000	Macacha Güemes	111	Acuña de Figueroa	54	F	27
5	476733	2018-01-01 00:21:01	0 days 00:36:10.000000000	Yatay	121	Billinghurst y Valentin Gomez	143	F	31
6	326932	2018-01-01 00:21:30	0 days 00:35:31.000000000	Yatay	121	Billinghurst y Valentin Gomez	143	F	28
7	347609	2018-01-01 00:22:48	0 days 00:31:07.000000000	Pque. Jose Evaristo Uriburu	155	Cerrito	71	M	27
8	392077	2018-01-01 00:23:46	0 days 00:34:28.000000000	Obelisco	7	Azucena Villaflor	79	F	29
9	457895	2018-01-01 00:24:19	0 days 00:17:08.000000000	Obelisco	7	Colegio Nacional Buenos Aires	47	F	26

* —
¿Por qué dice 2.619.968 viajes y 121.015 usuarios en el dataset, y Wikipedia dice que Paula Bisiau -Subsecretaria de Movilidad Sustentable- declara que fueron más de ocho millones de viajes para 200.000 usuarios en la afirmación de más arriba? Es importante verificar el discurso público cuando sea posible, y para eso ayuda la ciencia de datos.

Ambos vínculos a la supuesta afirmación de Bisiau en Wikipedia no existen más, uno está roto y el otro cambió de contenido. La explicación parece radicar en que eran proyecciones y en una mala interpretación de texto, como se ve en otro artículo que habla sobre un informe de la Secretaría de Transporte porteña, en donde dice:

"Cada día, en la ciudad de Buenos Aires se realizan 8 millones de viajes, el 84% de los cuales se hacen en transporte público. El medio preferido es el colectivo, que acapara al 57,5% de los traslados." ([Clarín 08/10/2016](#), consultado el 03/11/2019)

ECOBICI: EL SISTEMA PÚBLICO DE BICICLETAS PORTEÑO

2) Ya con algunas características del DS en la cabeza, comenzaron a surgir preguntas e hipótesis:

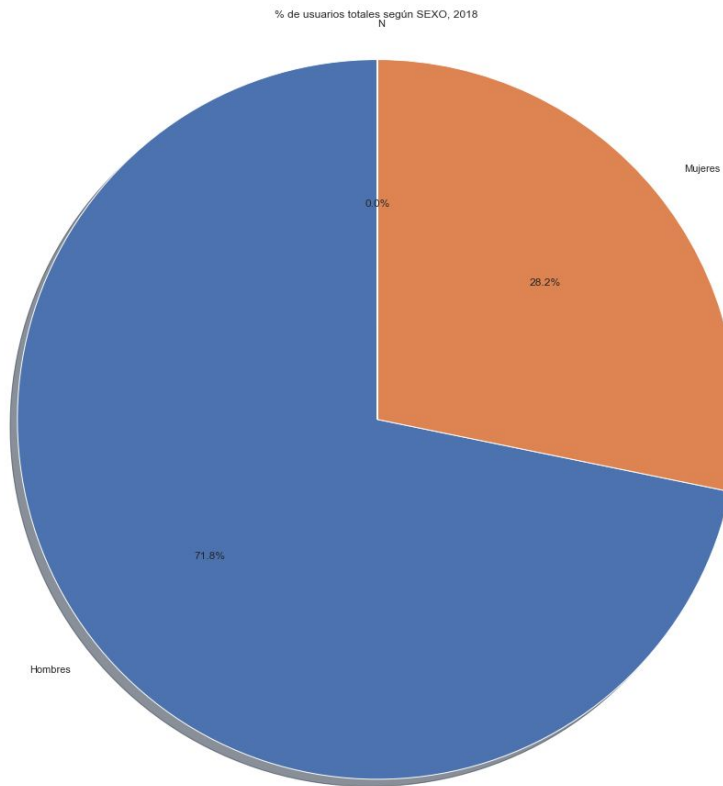
- a. ¿Usaron más el sistema los hombres o las mujeres?
- b. ¿Cómo varía la demanda total durante el año? (estacionalidad)
Hipótesis: en épocas de calor se usan menos.
- c.1. ¿Cuál estación se usó más, y cuál menos? c.2. ¿Qué porcentaje de M o F iniciaron los viajes en esas estaciones?
- d. ¿Qué segmento etario utilizó más el sistema?
- e. ¿Hay registrados viajes iniciados y no finalizados?



ECOBICI: EL SISTEMA PÚBLICO DE BICICLETAS PORTEÑO



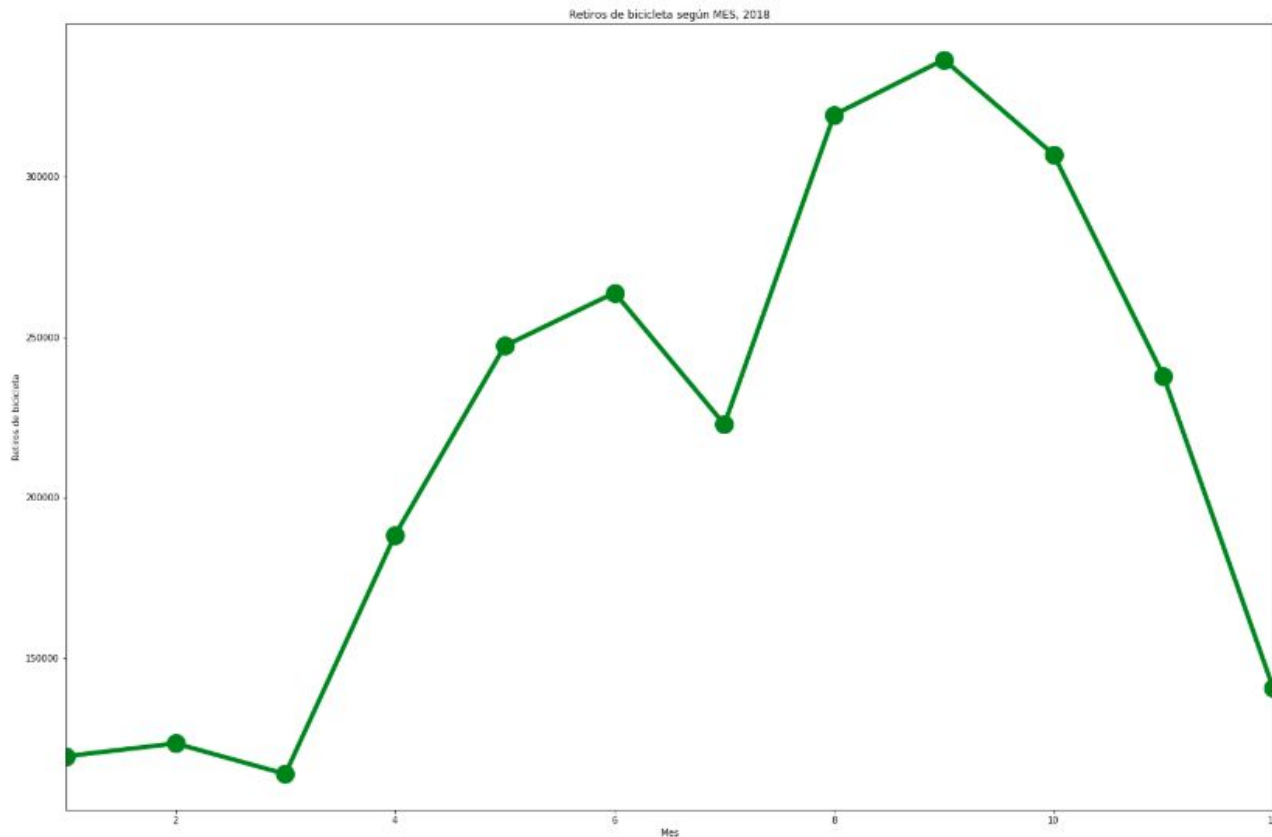
- a. Usaron más el sistema los hombres: 71,8% contra 28,2% de mujeres.



ECOBICI: EL SISTEMA PÚBLICO DE BICICLETAS PORTEÑO



b. En “épocas de calor” se usan menos, pero hubo un repunte con la primavera (y en el agosto previo).



ECOBICI: EL SISTEMA PÚBLICO DE BICICLETAS PORTEÑO



c.1. La estación de origen más usada fue Facultad de Medicina, la menos usada la de Palacio Lezama (5 retiros). Utilizaré la que la antecede para la siguiente respuesta.

c.2. Fac. Medicina: M 69,89%, F 30,10%. Pza. Bouchard: M 78,86%, F 21,13%.

```
In [32]: 1 df["bici_nombre_estacion_origen"].value_counts()
```

```
Out[32]: Facultad de Medicina      38858
         Pacífico                 36834
         Parque Las Heras        36500
         Plaza Italia             32722
         Retiro III               30765
         ...
         Santos Dumont y Otero    2419
         Plaza Houssay           2213
         Hospital Sarda          2090
         Plaza Hipolito Bouchard 1287
         PALACIO LEZAMA           5
         Name: bici_nombre_estacion_origen, Length: 199, dtype: int64
```

c. 2 ¿Usaron más los hombres o las mujeres las estaciones de origen Facultad de Medicina y Plaza Hipólito Bouchard?

```
In [44]: 1 df[df.bici_nombre_estacion_origen == "Facultad de Medicina"].bici_sexo.value_counts()#números enteros
```

```
Out[44]: M    27158
         F    11700
         Name: bici_sexo, dtype: int64
```

```
In [45]: 1 df[df.bici_nombre_estacion_origen == "Facultad de Medicina"].bici_sexo.value_counts(normalize=True)#porcentaje
```

```
Out[45]: M    0.698904
         F    0.301096
         Name: bici_sexo, dtype: float64
```

```
In [46]: 1 df[df.bici_nombre_estacion_origen == "Plaza Hipolito Bouchard"].bici_sexo.value_counts()#números enteros
```

```
Out[46]: M     1015
         F       272
         Name: bici_sexo, dtype: int64
```

```
In [47]: 1 df[df.bici_nombre_estacion_origen == "Plaza Hipolito Bouchard"].bici_sexo.value_counts(normalize=True)#porcentaje
```

```
Out[47]: M    0.788656
         F    0.211344
         Name: bici_sexo, dtype: float64
```

ECOBICI: EL SISTEMA PÚBLICO DE BICICLETAS PORTEÑO



c.1. La estación de destino más usada fue Facultad de Medicina, la menos usada la de Palacio Lezama (12 devoluciones). Utilizaré la que la antecede para la siguiente respuesta.

c.2. Fac. Medicina: M 69,91%, F 30,08%. Pza. Bouchard: M 79,21%, F 20,78%.

```
In [34]: df.bici_nombre_estacion_destino.value_counts()
```

```
Out[34]: Facultad de Medicina      39027
         Pacífico                  36402
         Parque Las Heras         34573
         Parque Centenario        29611
         Plaza Italia              28842
         ...
         Santos Dumont y Otero     2239
         Plaza Houssay             2229
         Hospital Sarda            2152
         Plaza Hipolito Bouchard   1246
         PALACIO LEZAMA            12
         Name: bici_nombre_estacion_destino, Length: 199, dtype: int64
```

¿Usaron más los hombres o las mujeres las estaciones de destino Facultad de Medicina y Plaza Hipólito Bouchard?

```
In [13]: 1 df[df.bici_nombre_estacion_destino == "Facultad de Medicina"].bici_sexo.value_counts()#números enteros
```

```
Out[13]: M    27286
         F    11741
         Name: bici_sexo, dtype: int64
```

```
In [10]: 1 df[df.bici_nombre_estacion_destino == "Facultad de Medicina"].bici_sexo.value_counts(normalize=True)#porcentaje
```

```
Out[10]: M    0.699157
         F    0.300843
         Name: bici_sexo, dtype: float64
```

```
In [14]: 1 df[df.bici_nombre_estacion_destino == "Plaza Hipolito Bouchard"].bici_sexo.value_counts()#números enteros
```

```
Out[14]: M     987
         F     259
         Name: bici_sexo, dtype: int64
```

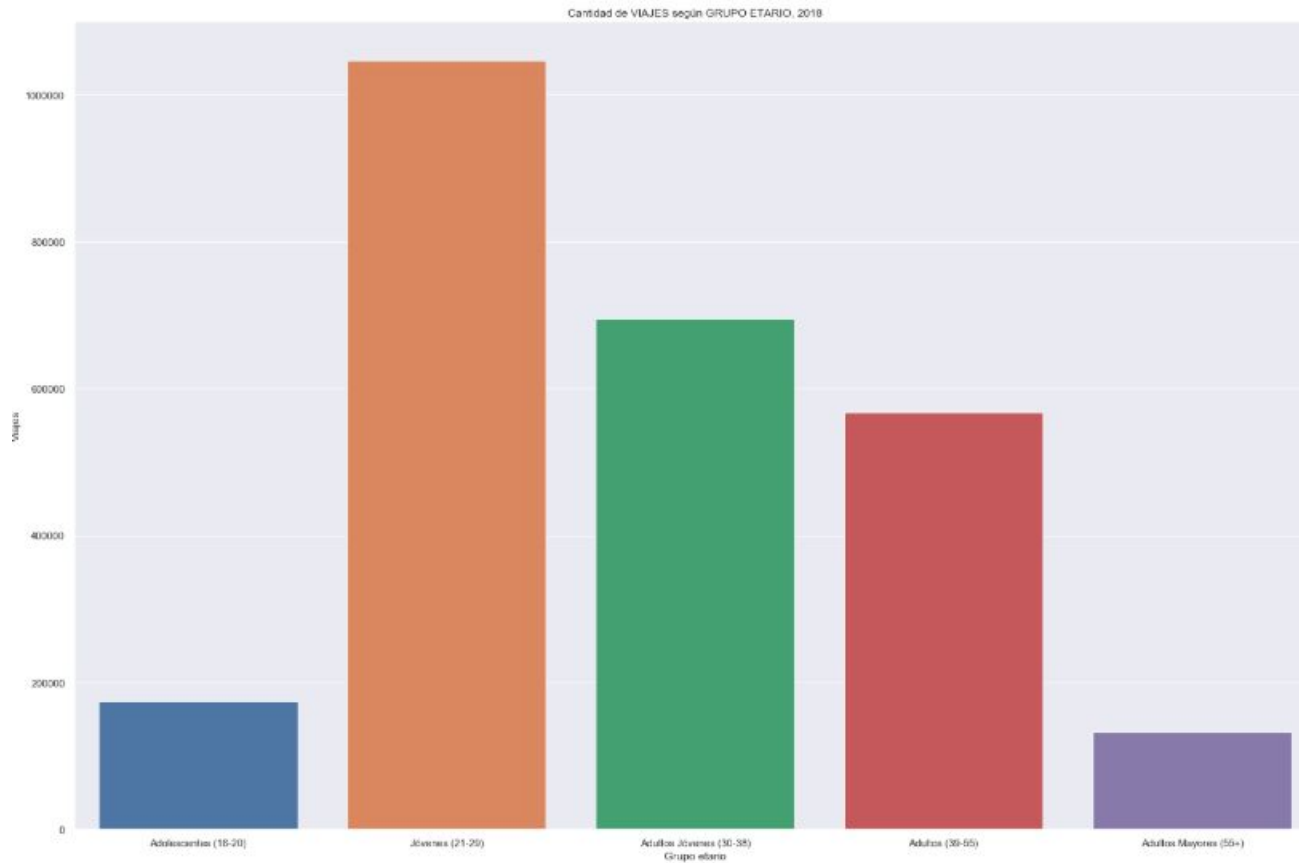
```
In [12]: 1 df[df.bici_nombre_estacion_destino == "Plaza Hipolito Bouchard"].bici_sexo.value_counts(normalize=True)#porcentaje
```

```
Out[12]: M    0.792135
         F    0.207865
         Name: bici_sexo, dtype: float64
```

ECOBICI: EL SISTEMA PÚBLICO DE BICICLETAS PORTEÑO



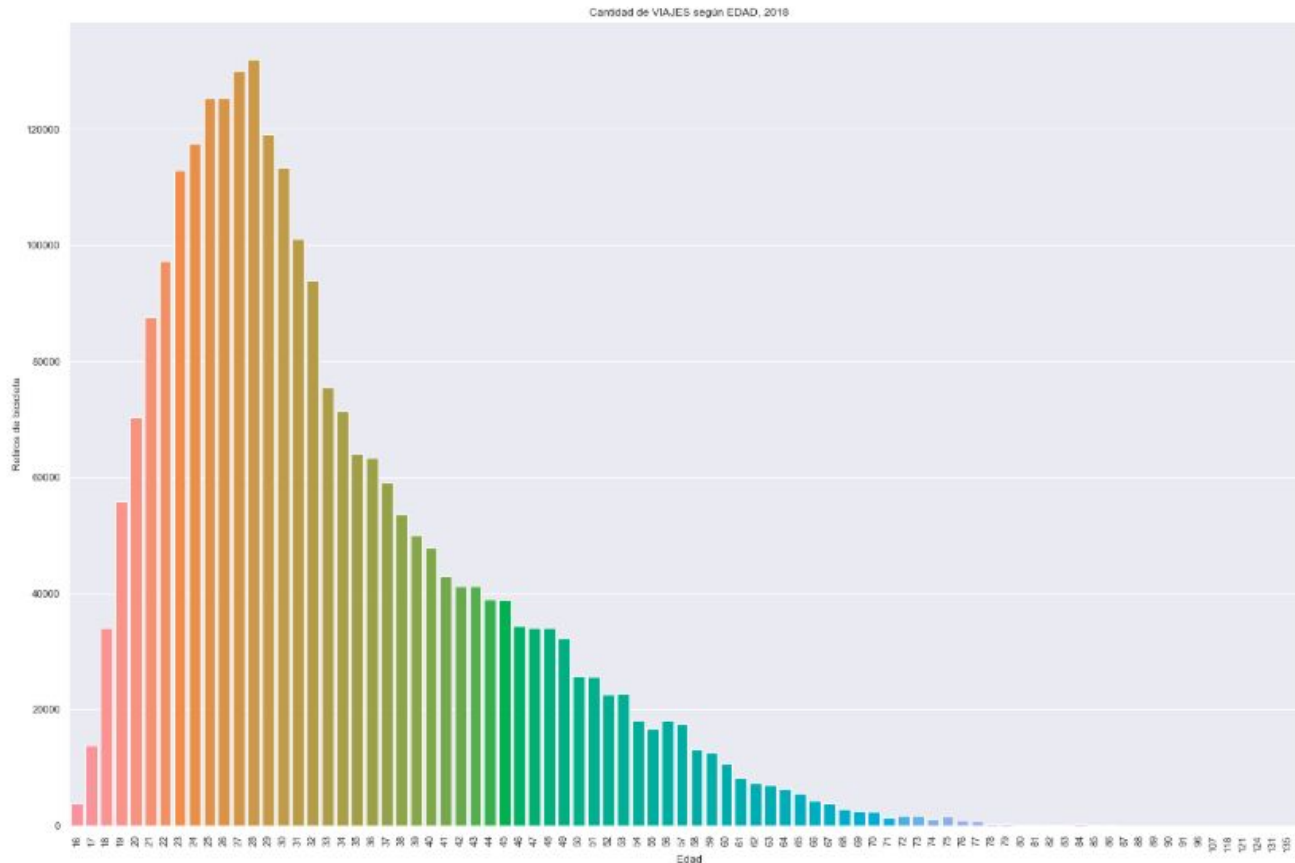
d. El grupo etario - creado por mí- que más utilizó EcoBici fue “jóvenes” con 1.046.796 retiros.



ECOBICI: EL SISTEMA PÚBLICO DE BICICLETAS PORTEÑO



EST. DESCRIPTIVA, EDAD: Moda: 28 años, mediana: 30, promedio 33.22 años.



ECOBICI: EL SISTEMA PÚBLICO DE BICICLETAS PORTEÑO



e. NO HAY VIAJES SIN FINALIZAR EN EL DATASET, dado que coinciden el **número de viajes marcados por su origen** con el **número marcado mediante su destino** -2.619.968- y que el número de estaciones únicas coincide en ambas columnas -199- sin haber ninguna etiqueta para NaNs ni ninguna etiqueta "rara".

```
Name: bici_nombre_estacion_destino, Length: 199, dtype: int64
```

```
In [57]: 1 len(df.bici_nombre_estacion_origen.value_counts()) == len(df.bici_nombre_estacion_destino.value_counts())
```

```
Out[57]: True
```

```
In [48]: 1 len(df.bici_nombre_estacion_origen)#cantidad de viajes desde origen
```

```
Out[48]: 2619968
```

```
In [49]: 1 len(df.bici_nombre_estacion_destino)#cantidad de viajes desde destino
```

```
Out[49]: 2619968
```

DESAFÍO:

¿Qué otras preguntas/consideraciones se les ocurren para hacerse con este Dataset?

Podríamos... bajar más datasets y hacer comparaciones interanuales de las variables que estudiamos y/o otras. Hay que tener en cuenta que la red de estaciones de bici era más pequeña en años anteriores.

...comparar los porcentajes de uso por parte de M o F entre BA y otras ciudades, como por ejemplo Amsterdam, ciudad paradigma del uso de la bicicleta y de la igualdad.

...reflexionar sobre las implicancias para las administraciones públicas: si con esta pequeña muestra ya encontramos un error en Wikipedia, ¿hasta qué punto las administraciones se van a plegar a abrir los datos que muestran los resultados e impactos de sus políticas públicas?

iGracias!