

Exploración de Variables Categorías y Análisis Multivariable (multivariado)

**Análisis de Datos con
Python**

Eduardo Selim Martínez Mayorga

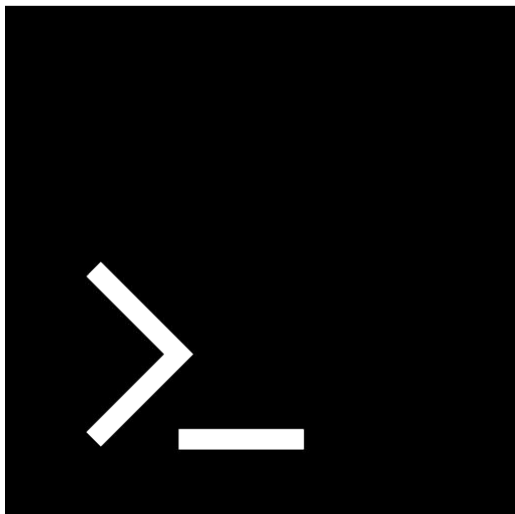


- Identificar distintas técnicas para explorar y visualizar variables categóricas.
- Añadir anotaciones a nuestras gráficas para hacerlas más comprensibles.
- Utilizar gráficas de barras.
- Graficar un conjunto de datos agrupado de acuerdo a dos variables categóricas
- Producir más de una gráfica al mismo tiempo para compararlas.
- Graficar un conjunto de datos numéricos agrupados de acuerdo a una variable categórica



- Seaborn
- Distribuciones
- Diagramas de caja
- Tablas de frecuencias
- Histogramas
- Gráficas de densidad

Antes de empezar



¡No olvides hacer pull del repo!

El material de la sesión se encuentra ahí.

```
git pull origin master
```



[Ve al Ejemplo 1](#)

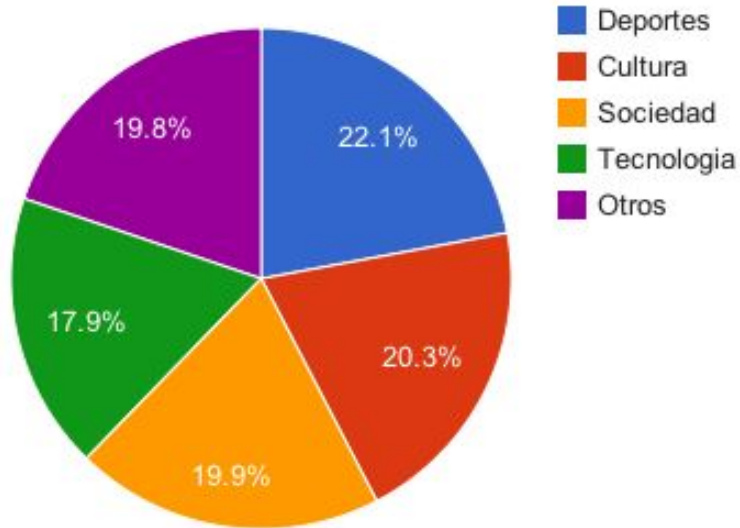
[Ve al Reto 1](#)

Hacen que podamos entender el contexto para poder interpretarlas y que otros puedan entenderlas.

- **Títulos.** Hacen claro el tema del análisis
- **Nombres de los ejes.** Hace más claro qué cosa estamos analizando.
- **Leyendas.** Cuando tenemos más de una categoría.
- **Tamaño.** Hacen más legible el contenido.
- **Dirección de los *tics*.** También mejoran la legibilidad.

Histogramas	Gráficas de barras
Las usamos para graficar variables numéricas discretas y continuas	Las usamos para graficar variables categóricas
Para poder generarlas necesitamos segmentar nuestros datos numéricos en rangos	La segmentación ya está dada de antemano ya que nuestra variable es categórica y sólo puede tomar un número predefinido de valores
Los segmentos se ordenan de menor a mayor en el eje x	No hay un orden específico a menos que nuestra variable categórica sea ordinal
Ya que todos los segmentos son rangos de una misma categoría, no hay separación entre cada barra	Se agregan espacios entre cada barra para hacer explícito que pertenecen a categorías distintas
El eje y está conformado por los conteos (la frecuencia) de cada segmento	El eje y puede indicar frecuencias, o también porcentajes

Visitas a contenidos



¡Save the pie for dessert!



- Valor o valores que más se repiten en el conjunto de datos.
- Se puede obtener desde una gráfica de barras o haciendo un conteo de valores.

Moda -> Categóricos

Media -> Numéricos

Mediana

[Ve al Ejemplo 2](#)

[Ve al Reto 2](#)

Tabla de contingencia

rating	Average	Excellent	Good	Not rated	Poor	Very Good	Total
price range							
1	12	9	33	1	0	19	74
2	47	37	116	3	2	132	337
3	41	84	141	3	1	218	488
4	19	64	56	19	2	121	281
Total	119	194	346	26	5	490	1180

[Ve al Ejemplo 3](#)

[Ve al Reto 3](#)

- Toma una de las variables categóricas como índice y el resto para crear columnas.
- Es una tabla de conteos dividida por categorías.
- Para graficarlas se usan varias gráficas en una.

Tabla de contingencia

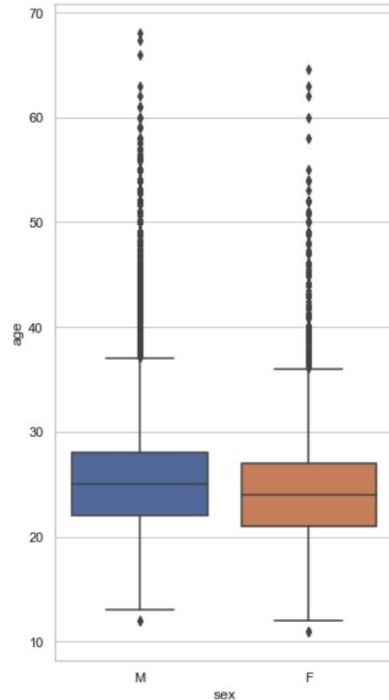
rating	Average	Excellent	Good	Not rated	Poor	Very Good	Total
price range							
1	12	9	33	1	0	19	74
2	47	37	116	3	2	132	337
3	41	84	141	3	1	218	488
4	19	64	56	19	2	121	281
Total	119	194	346	26	5	490	1180

- Toma una de las variables categóricas como índice y el resto para crear columnas.
- Es una tabla de conteos dividida por categorías.
- Para graficarlas se usan varias gráficas en una.

[Ve al Ejemplo 4](#)

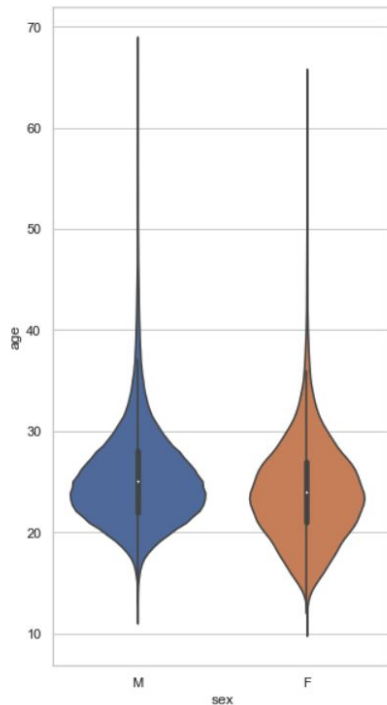
[Ve al Reto 4](#)

Edades de atletas Olímpicos separadas por género



- Se pueden usar de la misma forma que las tablas de contingencia: Separando por categorías.

Edades de atletas Olímpicos separadas por género



- A diferencia de las de caja, muestran la densidad.
- Son similares a las gráficas de densidad pero combinadas con las de caja.

[Ve al Ejemplo 5](#)

[Ve al Reto 5](#)



NO OLVIDES REVISAR TU
POSTWORK Y TU PREWORK



Preguntas

