

# Estadística descriptiva

## Medidas de tendencia y dispersión

# Repaso clase anterior

1. Exploración y Explicación
2. Distintas maneras de visualizar y explorar datos
3. Ejemplos (Barras, Puntos, Líneas, Torta, etc.)
4. Distribución de Datos (Histogramas de variables categóricas y continuas)

# Análisis estadístico

¿Cómo caracterizarían/resumirían el precio de venta de estas propiedades en un único valor? (2 min.)

PrecioDeVenta
\$108.000
\$138.000
\$138.000
\$142.000
\$186.000
\$199.500
\$208.000
\$254.000
\$254.000
\$257.500
\$298.000
\$456.400

**¿Cómo lo  
obtuvieron?**

# Medidas de tendencia

# Medidas de tendencia

Una medida de tendencia (central) es un valor único asociado a una variable para caracterizar de alguna manera el conjunto completo de valores

- Existen distintas medidas
- Cada una posee ventajas y desventajas relativas respecto a las otras

# Medidas de tendencia

**Media** (o valor promedio) es la sumatoria de todos los datos dividida la cantidad total de datos

PrecioDeVenta
\$108.000
\$138.000
\$138.000
\$142.000
\$186.000
\$199.500
\$208.000
\$254.000
\$254.000
\$257.500
\$298.000
\$456.400

$$\text{Media} = \frac{\$108.000 + \$138.000 + \dots + \$456.400}{12} = \$219.950$$

En general:

$$\text{media}(x) = \frac{\sum_{i=1}^N X_i}{N}$$

# Medidas de tendencia

**Mediana** es el número del medio; se encuentra al ordenar todos los valores y elegir el que está en el medio (o si hay dos números en el medio, tomar el promedio de esos dos números)

Están  
ordenado  
s



PrecioDeVenta
\$108.000
\$138.000
\$138.000
\$142.000
\$186.000
\$199.500
\$208.000
\$254.000
\$254.000
\$257.500
\$298.000
\$456.400

$$\text{Mediana} = \frac{\$199.500 + \$208.000}{2} = \$203.750$$

# Medidas de tendencia

**Media (promedio) → Es influenciada por valores atípicos (valores extremadamente chicos/grandes)**

**Mediana → No influenciada por valores atípicos (su cálculo es robusto)**

PrecioDeVenta
\$108.000
\$138.000
\$138.000
\$142.000
\$186.000
\$199.500
\$208.000
\$254.000
\$254.000
\$257.500
\$298.000
\$456.400

$\$219.950 > \$203.750$   
Media      Mediana

Aumenta la media  
pero no la mediana

Valor  
extremo

Si reemplazáramos los \$456.400 por \$1,5 millones  
(media = \$306.916,67; mediana = \$203.750)  
la mediana permanecería sin cambios



La mediana da una  
mejor idea del precio de  
venta

Siempre que un conjunto de datos contiene **valores extremos**, la **mediana** es la medida preferida de tendencia central (en particular para conjuntos de datos con pocas observaciones)



# Medidas de tendencia

**Moda** es el número más frecuente, es decir, el número que se repite el mayor número de veces (en caso de empate, puede existir más de una moda; en caso de no existir repeticiones los datos no tienen moda)

		PrecioDeVenta
se repite 2 veces	{	\$108.000
		\$138.000
		\$138.000
		\$142.000
		\$186.000
se repite 2 veces	{	\$199.500
		\$208.000
		\$254.000
		\$254.000
		\$257.500
		\$298.000
		\$456.400

**Moda = { \$138.000; \$254.000 }**

- Útil para variables que tienen pocos valores distintos
- Variables con muchos valores distintos (Ej. tiempos de maratonistas en una carrera)
  - Es posible que la moda no exista ¿por qué?
  - Alternativa: construir histograma y aplicar la noción de moda para referirse al bin con mayor cantidad de observaciones.

# Medidas de Dispersión

# Medidas de dispersión

Consigna. Sean:

Notas Estudiante A: 4; 5; 7; 7; 7 ; 9; 10

Notas Estudiante B: 7; 7; 7; 7; 7; 7; 7

Calcular la Media, Mediana y Moda para cada uno de los Estudiantes

# Medidas de dispersión

Consigna. Sean:

Notas Estudiante A: 4; 5; 7; 7; 7 ; 9; 10

Notas Estudiante B: 7; 7; 7; 7; 7; 7; 7

Calcular la Media, Mediana y Moda para cada uno de los Estudiantes

Respuestas.

	Estudiante A	Estudiante B
Media	7	7
Mediana	7	7
Moda	7	7

- Medidas de tendencia central **no describen de qué manera varían los valores**
- Es necesario un valor que permita caracterizar a la “**dispersión**” de los valores ¿Cuál?

# Medidas de dispersión

**Rango** es la diferencia numérica entre el valor **máximo** y el valor **mínimo**

PrecioDeVenta
\$108.000
\$138.000
\$138.000
\$142.000
\$186.000
\$199.500
\$208.000
\$254.000
\$254.000
\$257.500
\$298.000
\$456.400

Mínimo

$$\text{Rango} = \$456.400 - \$108.000 = \$348.400$$

Máximo

¿Problemas?

- Se basa sólo en 2 valores (máximo y mínimo)
- Influenciable por valores extremos

# Medidas de dispersión

**Desviación Estándar** representa cuánto se apartan los valores del valor promedio

PrecioDeVenta
\$108.000
\$138.000
\$138.000
\$142.000
\$186.000
\$199.500
\$208.000
\$254.000
\$254.000
\$257.500
\$298.000
\$456.400

**Media**

$$SD = \sqrt{\frac{(\$108.000 - \$219.950)^2 + \dots + (\$456.400 - \$219.950)^2}{12 - 1}} = \$95.100$$

# Medidas de dispersión

**Desviación Estándar** representa cuánto se apartan los valores del valor promedio

PrecioDeVenta
\$108.000
\$138.000
\$138.000
\$142.000
\$186.000
\$199.500
\$208.000
\$254.000
\$254.000
\$257.500
\$298.000
\$456.400

$$SD = \sqrt{\frac{(\$108.000 - \$219.950)^2 + \dots + (\$456.400 - \$219.950)^2}{12 - 1}} = \$95.100$$

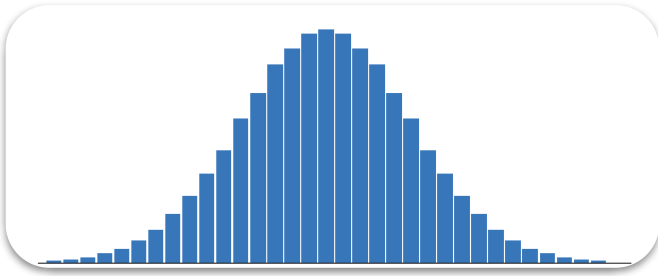
En general:

	Desviación Estándar
Población	$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$
Muestra	$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$

# Medidas de dispersión

**Desviación Estándar** representa cuánto se apartan los valores del valor promedio

- **Medida confiable cuando el histograma tiene forma de campana simétrica**

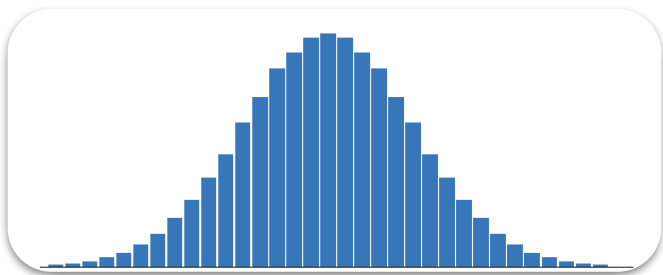




# Medidas de dispersión

**Desviación Estándar** representa cuánto se apartan los valores del valor promedio

- Medida confiable cuando el histograma tiene forma de campana simétrica



**En estos casos la variabilidad nos permite describir los datos usando intervalos :**

- el 68% de los valores están en [media - 1 SD; media +1 SD]
- el 95% de los valores están en [media - 2 SD; media +2 SD]
- > 99% de los valores están en [media - 3 SD; media +3 SD]

**¿Problemas?**

- No es confiable para distribuciones asimétricas
- Influyente por valores extremos

# Medidas de dispersión

**Percentil** es el valor que divide a una lista ordenada de datos de forma que un porcentaje de los datos sea inferior a dicho valor

# Medidas de dispersión

**Percentil** es el valor que divide a una lista ordenada de datos de forma que un porcentaje de los datos sea inferior a dicho valor

## ¿Cómo se calcula el valor del p-ésimo percentil?

1. Calcular su posición entre el conjunto de valores ordenados

*Ejemplo. ¿percentil 25? ->*

# Medidas de dispersión

$$Lp = (n+1) \frac{P}{100}$$

*Lp: Localización en la lista de datos ordenados*

*n: Tamaño muestral*

*P: Percentil a calcular*

**Percentil** es el valor que divide a una lista ordenada de datos de forma que un porcentaje de los datos sea inferior a dicho valor

## ¿Cómo se calcula el valor del p-ésimo percentil?

1. Calcular su posición entre el conjunto de valores ordenados

*Ejemplo. ¿percentil 25? ->*

¿Cuál es el Percentil 25 de altura?



# Medidas de dispersión

$$Lp = (n + 1) \frac{P}{100}$$

*Lp: Localización en la lista de datos ordenados*

*n: Tamaño muestral*

*P: Percentil a calcular*

**Percentil** es el valor que divide a una lista ordenada de datos de forma que un porcentaje de los datos sea inferior a dicho valor

## ¿Cómo se calcula el valor del p-ésimo percentil?

1. Calcular su posición entre el conjunto de valores ordenados

Ejemplo. ¿percentil 25?  $\rightarrow \frac{25 \times (12 + 1)}{100} = 3,25$

entre posiciones 3 y 4

2. Realizar la interpolación necesaria

$$\$138.000 + (3,25 - 3) \times (\$142.000 - \$138.000) = \$139.000$$

PrecioDeVenta
\$108.000
\$138.000
\$138.000
\$142.000
\$186.000
\$199.500
\$208.000
\$254.000
\$254.000
\$257.500
\$298.000
\$456.400

25% de los datos

# Medidas de dispersión

$$Lp = (n+1) \frac{P}{100}$$

*Lp: Localización en la lista de datos ordenados*

*n: Tamaño muestral*

*P: Percentil a calcular*

**Percentil** es el valor que divide a una lista ordenada de datos de forma que un porcentaje de los datos sea inferior a dicho valor

¿Cómo calcular el valor del p-ésimo percentil?

1. Calcular su posición entre el conjunto de valores ordenados

Ejemplo. ¿percentil 50?  $\rightarrow \frac{50 \times (12 + 1)}{100} = 6,5$

entre posiciones 6 y 7

2. Realizar la interpolación necesaria

$$\$199.500 + (6,5 - 6) \times (\$208.000 - \$199.500) = \$203.750$$

¡ Coincide con la mediana !

PrecioDeVenta
\$108.000
\$138.000
\$138.000
\$142.000
\$186.000
\$199.500
\$208.000
\$254.000
\$254.000
\$257.500
\$298.000
\$456.400

50% de los datos

# Medidas de dispersión

$$Lp = (n+1) \frac{P}{100}$$

*Lp: Localización en la lista de datos ordenados*

*n: Tamaño muestral*

*P: Percentil a calcular*

**Percentil** es el valor que divide a una lista ordenada de datos de forma que un porcentaje de los datos sea inferior a dicho valor

## ¿Cómo calcular el valor del p-ésimo percentil?

1. Calcular su posición entre el conjunto de valores ordenados

Ejemplo. ¿percentil 75?  $\rightarrow \frac{75 \times (12 + 1)}{100} = 9,75$

entre  
posiciones  
9 y 10

2. Realizar la interpolación necesaria

$$\$254.000 + (9,75 - 9) \times (\$257.500 - \$254.000) = \$256.625$$

PrecioDeVenta
\$108.000
\$138.000
\$138.000
\$142.000
\$186.000
\$199.500
\$208.000
\$254.000
\$254.000
\$257.500
\$298.000
\$456.400

75% de los datos

# Medidas de dispersión

**Percentil** es el valor que divide a una lista ordenada de datos de forma que un porcentaje de los datos sea inferior a dicho valor

En resumen, para calcular percentiles debemos:

- Ordena los datos
- Usa la fórmula para determinar la posición del percentil
  - Si es un número entero, toma el valor en esa posición
  - Si es decimal, interpola entre los valores adyacentes



# Medidas de dispersión

**Percentil** es el valor que divide a una lista ordenada de datos de forma que un porcentaje de los datos sea inferior a dicho valor

- Se puede calcular un percentil para cualquier valor entre 0% y 100%
- Percentiles más comunes: 25, 50 y 75 (primer cuartil, segundo cuartil y tercer cuartil)
- **Rango intercuartil (IQR)**. La diferencia entre el **tercer** y el **primer** cuartil (los percentiles 75 y 25)
- IQR **abarca el 50% medio de la distribución** de los valores y se utiliza como medida de variación
- Ventajas de percentiles y el rango intercuartil sobre el rango y la desviación estándar
  - **percentiles no requieren** que la distribución de la variable tenga **forma de campana**
  - **valores extremos no distorsionan** el valor de los percentiles

# Medidas de dispersión

**Consigna.** Sean ...

**Notas Estudiante A:** 4; 5; 7; 7; 7 ; 9; 10

**Notas Estudiante B:** 7; 7; 7; 7; 7; 7; 7

**Calcular la Media, Mediana y Moda para cada uno de los Estudiantes**  
**Calcular el rango, desvío estándar, cuartiles e IQR**

**Respuestas**

	Estudiante A	Estudiante B
<b>Media</b>	7	7
<b>Mediana</b>	7	7
<b>Moda</b>	7	7

# Medidas de dispersión

Consigna. Sean ...

Notas Estudiante A: 4; 5; 7; 7; 7 ; 9; 10

Notas Estudiante B: 7; 7; 7; 7; 7; 7; 7

Calcular la Media, Mediana y Moda para cada uno de los Estudiantes  
Calcular el rango, desvío estándar, cuartiles e IQR

Respuestas

	Estudiante A	Estudiante B
Media	7	7
Mediana	7	7
Moda	7	7

	Estudiante A	Estudiante B
Rango	6	0
STD	2,08	0,00
1Q	5	7
2Q	7	7
3Q	9	7
IQR	4	0

# Estadística descriptiva con python

- Vamos a trabajar con el dataset '*tips*'

Índice	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.5	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2

- Calculamos la **media** de la columna '*tip*'

```
In [13]:  
...: tips['tip'].mean()  
Out[13]: 2.99827868852459
```

# Estadística descriptiva con python

- Vamos a trabajar con el dataset '*tips*'

Índice	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.5	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2

- Calculamos la **mediana** de la columna '*tip*'

```
In [15]:  
...: tips['tip'].median()  
Out[15]: 2.9
```

# Estadística descriptiva con python

- Vamos a trabajar con el dataset '*tips*'

Índice	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.5	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2

- Calculamos la **moda** de la columna '*tip*'

```
In [16]:  
...: tips['tip'].mode()  
Out[16]:  
0    2.0  
Name: tip, dtype: float64
```

# Estadística descriptiva con python

- Calculamos el **rango** de la columna '*tip*'

```
In [18]:  
...: rango_tips = max(tips['tip']) - min(tips['tip'])  
...: print(rango_tips)  
9.0
```

- Calculamos la **desviación estándar** de la columna '*tip*'

```
In [17]: tips['tip'].std()  
Out[17]: 1.3836381890011822
```

# Estadística descriptiva con python

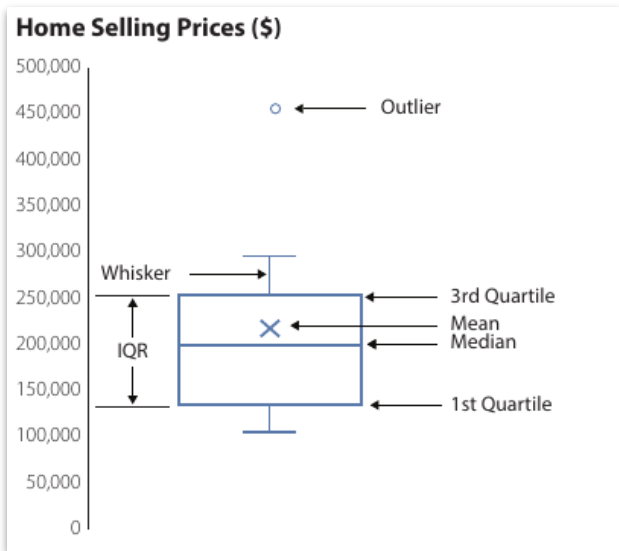
- Un comando **muy útil** es el método `.describe()`, el cual proporciona estadísticas descriptivas como la media, el percentil 25, el percentil 50 (mediana), el percentil 75, entre otros.

```
In [22]: tips['tip'].describe()
Out[22]:
count      244.000000
mean       2.998279
std        1.383638
min         1.000000
25%        2.000000
50%        2.900000
75%        3.562500
max        10.000000
Name: tip, dtype: float64
```



# Boxplot

**Boxplot** es un resumen gráfico de la distribución de los datos. Se basa en los cuartiles.



Caja (Box)

3er. cuartil = Percentil 75 = \$256.625

2do. cuartil = Percentil 50 = \$203.750  
(Mediana)

1er. cuartil = Percentil 25 = \$139.000

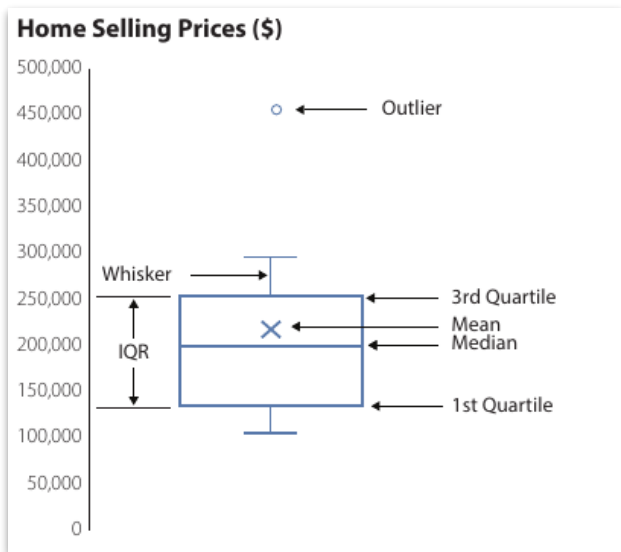
$IQR = 3er. - 1er\ cuartil = \$117.625$

Media = \$219.950

PrecioDeVenta
\$456.400
\$298.000
\$257.500
\$254.000
\$254.000
\$208.000
\$199.500
\$186.000
\$142.000
\$138.000
\$138.000
\$108.000

# Boxplot

**Boxplot** es un resumen gráfico de la distribución de los datos. Se basa en los cuartiles.



## Bigotes (Whisker)

Límite Sup. = 3er. cuartil +  $1,5 * IQR$  = \$433,062.5

Límite Inf. = 1er. cuartil -  $1,5 * IQR$  = \$-37.437

Los bigotes se extienden hasta el último dato que se encuentra dentro del límite

PrecioDeVenta
\$456.400
\$298.000
\$257.500
\$254.000
\$254.000
\$208.000
\$199.500
\$186.000
\$142.000
\$138.000
\$138.000
\$108.000

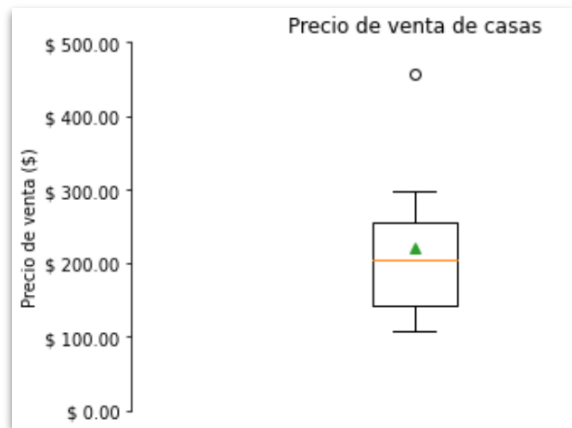
← Outlier

← Bigote Sup.

← Bigote inf.

# Boxplot

**Boxplot** es un resumen gráfico de la distribución de los datos. Se basa en los cuartiles.



## Bigotes (Whisker)

Límite Sup. = 3er. cuartil + 1,5 \* IQR = \$433,062.5

Límite Inf. = 1er. cuartil - 1,5 \* IQR = \$-37.437

Los bigotes se extienden hasta el último dato que se encuentra dentro del límite

PrecioDeVenta
\$456.400
\$298.000
\$257.500
\$254.000
\$254.000
\$208.000
\$199.500
\$186.000
\$142.000
\$138.000
\$138.000
\$108.000

← Outlier

← Bigote Sup.

← Bigote inf.

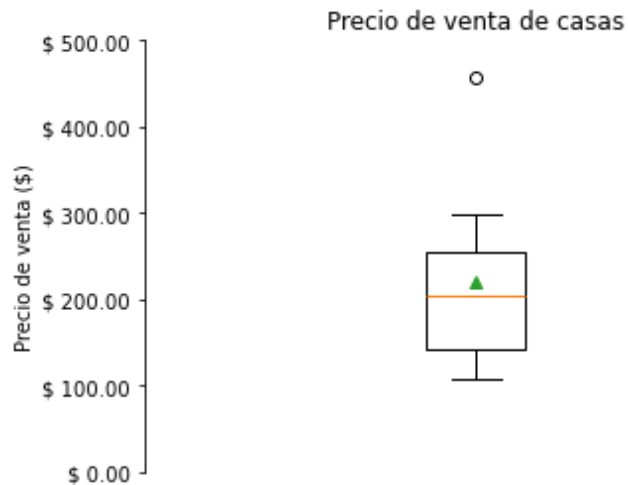
# Boxplot

```
fig, ax = plt.subplots()

ax.boxplot(ventaCasas['PrecioDeVenta'], showmeans=True)

# Agrega titulo, etiquetas a los ejes
#y limita el rango de valores de los ejes
ax.set_title('Precio de venta de casas')
ax.set_xticks([])
ax.set_ylabel('Precio de venta ($)')
# Agrega separador de decimales y signo $
ax.yaxis.set_major_formatter(ticker.StrMethodFormatter("$ {:.2f}"));

ax.set_ylim(0,500)
```

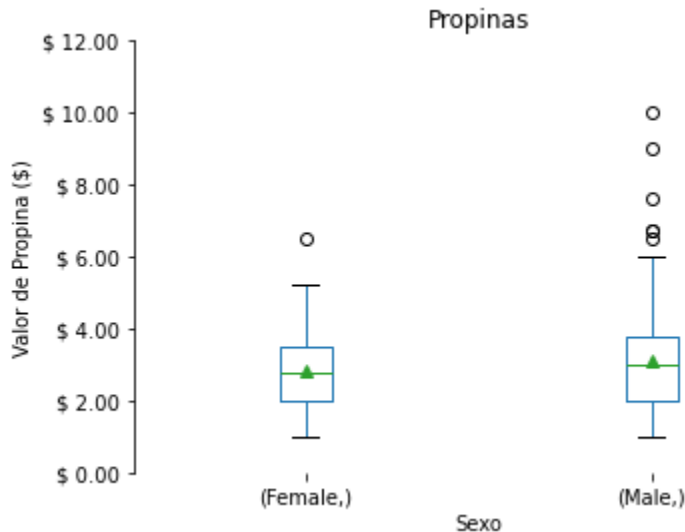


# Boxplot: incorporando variables categóricas al análisis

```
fig, ax = plt.subplots()

tips.boxplot(by=['sex'], column=['tip'],
             ax=ax, grid=False, showmeans=True)

# Agrega titulo, etiquetas a los ejes
fig.suptitle('')
ax.set_title('Propinas')
ax.set_xlabel('Sexo')
ax.set_ylabel('Valor de Propina ($)')
```

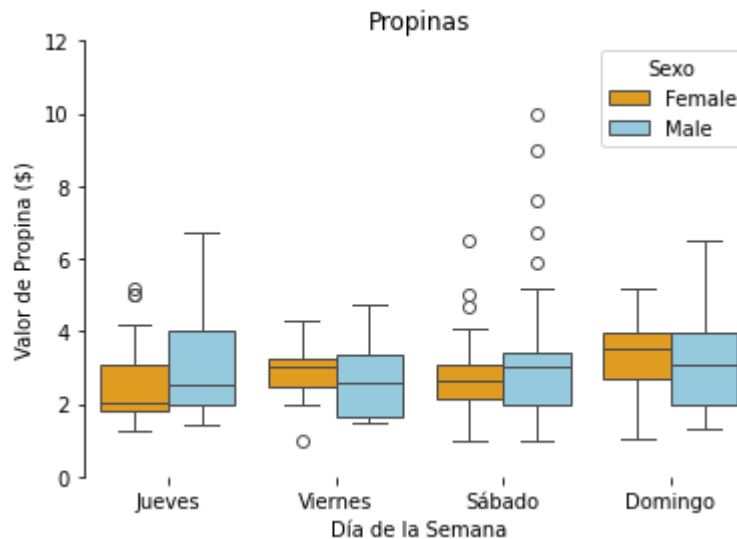


# Boxplot: incorporando variables categóricas al análisis

```
ax = sns.boxplot(x="day", ←  
                 y="tip",  
                 hue="sex", ←  
                 data=tips,  
                 order=['Thur', 'Fri', 'Sat', 'Sun'],  
                 palette={"Female": "orange", "Male": "skyblue" })
```

```
ax.set_title('Propinas')  
ax.set_xlabel('Día de la Semana')  
ax.set_ylabel('Valor de Propina ($)')  
ax.set_ylim(0,12)  
ax.legend(title="Sexo")
```

```
dias = ['Thur', 'Fri', 'Sat', 'Sun']  
labels = ['Jueves', 'Viernes', 'Sábado', 'Domingo']  
ax.set_xticks(range(len(dias)))  
ax.set_xticklabels(labels)
```

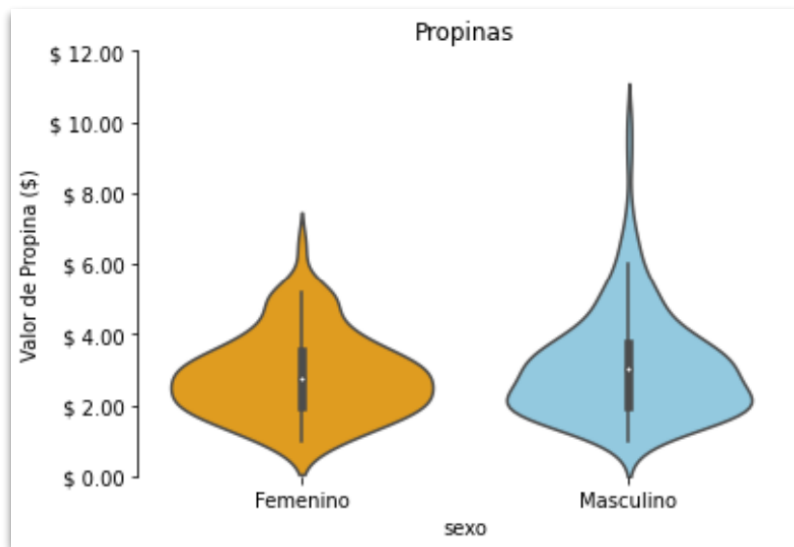


# Violinplot

**Violinplot** es similar al boxplot, salvo que muestra también la densidad de probabilidad de los datos, generalmente suavizada por un estimador de densidad kernel.

```
ax = sns.violinplot(x="sex", y="tip", data = tips,
                    palette={"Female": "orange", "Male": "skyblue" })

ax.set_title('Propinas')
ax.set_xlabel('sexo')
ax.set_ylabel('Valor de Propina ($)')
ax.yaxis.set_major_formatter(ticker.StrMethodFormatter("$ {x:,.2f}"));
ax.set_ylim(0,12)
ax.set_xticklabels(['Femenino', 'Masculino'])
```



# Ejercicio

Consigna.

Utilizando el dataset de propinas (*tips*)

- Generar un gráfico de boxplot de la proporción de propina recibida ( $\text{tip}/\text{total\_bill}$ )
- Analizar los resultados obtenidos
- Estudiar si hay diferencias en función del sexo y el sexo+día
- Discutir con el resto de la clase los resultados obtenidos



# Cómo mejorar la visualización



# Cómo mejorar la visualización de datos

	Memoria sensorial
<i>Definición</i>	Procesos cerebrales que interpretan estímulos por períodos mucho más breves que la memoria de corto plazo
<i>Sistemas derivados</i>	<ul style="list-style-type: none"><li>- Memoria icónica</li><li>- Memoria ecoica</li><li>- Memoria olfativa</li><li>- Memoria gustativa</li><li>- Memoria háptica</li></ul>
<i>Tiempo de permanencia de los datos</i>	Breve (milésimas de seg)
<i>Ejemplos</i>	Percibir un sonido en medio de una multitud

Fuente: <https://www.diferenciador.com/tipos-de-memoria/>

# Cómo mejorar la visualización de datos

	Memoria sensorial
Definición	Procesos cerebrales que interpretan estímulos por períodos mucho más breves que la memoria de corto plazo
Sistemas derivados	<div>- Memoria icónica ← <b>Visión</b> - Memoria ecoica <b>Audición</b> - Memoria olfativa <b>Olfato</b> - Memoria gustativa <b>Gusto</b> - Memoria háptica <b>Tacto</b></div>
Tiempo de permanencia de los datos	Breve (milésimas de seg)
Ejemplos	Percibir un sonido en medio de una multitud

Fuente: <https://www.diferenciador.com/tipos-de-memoria/>

# Cómo mejorar la visualización de datos

	Memoria sensorial	Memoria a corto plazo
<i>Definición</i>	Procesos cerebrales que interpretan estímulos por períodos mucho más breves que la memoria de corto plazo	Procesos cerebrales encargados de interpretar los estímulos y conservar esa información durante un tiempo breve
<i>Sistemas derivados</i>	<ul style="list-style-type: none"><li>- Memoria icónica</li><li>- Memoria ecoica</li><li>- Memoria olfativa</li><li>- Memoria gustativa</li><li>- Memoria háptica</li></ul>	<ul style="list-style-type: none"><li>- Sistema ejecutivo</li><li>- Almacén episódico</li><li>- Bucle fonológico</li><li>- Agenda visoespacial</li></ul>
<i>Tiempo de permanencia de los datos</i>	Breve (milésimas de seg)	Breve (7 a 40 seg)
<i>Ejemplos</i>	Percibir un sonido en medio de una multitud	Recordar la matrícula de un auto que acaba de pasar

Fuente: <https://www.diferenciador.com/tipos-de-memoria/>



# Cómo mejorar la visualización de datos

	Memoria sensorial	Memoria a corto plazo	Memoria a largo plazo
<i>Definición</i>	Procesos cerebrales que interpretan estímulos por períodos mucho más breves que la memoria de corto plazo	Procesos cerebrales encargados de interpretar los estímulos y conservar esa información durante un tiempo breve	Procesos cerebrales encargados de conservar información durante períodos prolongados
<i>Sistemas derivados</i>	<ul style="list-style-type: none"><li>- Memoria icónica</li><li>- Memoria ecoica</li><li>- Memoria olfativa</li><li>- Memoria gustativa</li><li>- Memoria háptica</li></ul>	<ul style="list-style-type: none"><li>- Sistema ejecutivo</li><li>- Almacén episódico</li><li>- Bucle fonológico</li><li>- Agenda visoespacial</li></ul>	<ul style="list-style-type: none"><li>- Memoria implícita</li><li>- Memoria explícita</li></ul>
<i>Tiempo de permanencia de los datos</i>	Breve (milésimas de seg)	Breve (7 a 40 seg)	Prolongado (minutos a décadas)
<i>Ejemplos</i>	Percibir un sonido en medio de una multitud	Recordar la matrícula de un auto que acaba de pasar	Recordar cómo manejar bicicleta

Fuente: <https://www.diferenciador.com/tipos-de-memoria/>

# Experimento - la memoria

Hagamos un experimento para ver los límites de la memoria a corto plazo:

**Consigna:**

- 1. Van a ver 5 series de letras**
- 2. Tienen que recordarlas en el orden correcto**

C X W





¿Las recuerdan?



C X W



M N K T Y



¿?





M N K T Y



R P J H B Z S



¿?



R P J H B Z S





G B M P V Q F J D



¿?



G B M P V Q F J D



E G Q W J P B R H K A





¿?



E G Q W J P B R H K A

- En general (a la mayoría, después de sólo 1 lectura) las siguientes series les resultaron:
  - 1 -> extremadamente fácil
  - 2 -> fácil
  - 3 -> un poco + difícil
  - 4 -> sumamente difícil
  - 5 -> casi imposible
- Muestra la limitada capacidad de la memoria de corto plazo
- Se estima que en memoria a corto plazo se puede retener alrededor de 4 fragmentos de información visual
- Consecuencia. Resulta difícil recordar qué color representa cada categoría si se utilizan más de 4 colores/categorías diferentes en un gráfico de barras o columns.

Atributos pre-atentivos

**Atributos pre-atentivos** representan aquellas características que pueden ser procesadas por la memoria icónica (memoria sensorial visual)

- Ver un gráfico o una tabla -> Ojos reciben estímulo y se transmite al cerebro
- Cerebro. Debe diferenciar y procesar

**Percepción visual.** Proceso mediante el cual nuestro cerebro interpreta esos reflejos producidos por la luz que entra por nuestros ojos. Relacionada con el funcionamiento de la memoria

- Memoria Icónica + Corto Plazo: las más importantes para el procesamiento visual
- Comprender qué aspectos se pueden procesar en estos tipos de memoria puede resultar útil para diseñar visualizaciones efectivas
- Ejemplo para ilustrar el poder de los atributos pre-atentivos

# Experimento

## Consigna ...

- Observar la siguiente figura
- Contar lo más rápido posible cuántas veces aparece el número de 7 en la figura

7	3	4	1	3	4	5	6	4	0
3	0	6	9	0	4	5	8	6	3
2	7	2	2	9	9	4	5	2	1
2	2	4	5	2	0	9	2	0	4
2	4	0	7	6	9	3	0	0	4
7	7	8	9	2	6	7	2	4	7
6	1	3	3	2	1	4	4	9	0
3	6	6	2	7	5	5	2	5	4
1	1	4	0	6	3	4	0	5	1
3	7	5	2	7	5	7	7	3	9
3	3	8	6	9	5	5	3	6	4
7	6	0	3	0	9	9	0	2	9
4	6	9	4	8	2	6	5	8	3
9	3	9	2	2	8	4	3	9	8
5	8	8	2	9	1	2	4	8	5
1	7	4	0	1	1	9	9	5	8



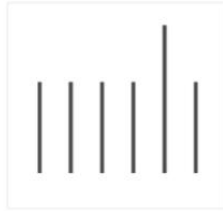
7	3	4	1	3	4	5	6	4	0
3	0	6	9	0	4	5	8	6	3
2	7	2	2	9	9	4	5	2	1
2	2	4	5	2	0	9	2	0	4
2	4	0	7	6	9	3	0	0	4
7	7	8	9	2	6	7	2	4	7
6	1	3	3	2	1	4	4	9	0
3	6	6	2	7	5	5	2	5	4
1	1	4	0	6	3	4	0	5	1
3	7	5	2	7	5	7	7	3	9
3	3	8	6	9	5	5	3	6	4
7	6	0	3	0	9	9	0	2	9
4	6	9	4	8	2	6	5	8	3
9	3	9	2	2	8	4	3	9	8
5	8	8	2	9	1	2	4	8	5
1	7	4	0	1	1	9	9	5	8

7	3	4	1	3	4	5	6	4	0
3	0	6	9	0	4	5	8	6	3
2	7	2	2	9	9	4	5	2	1
2	2	4	5	2	0	9	2	0	4
2	4	0	7	6	9	3	0	0	4
7	7	8	9	2	6	7	2	4	7
6	1	3	3	2	1	4	4	9	0
3	6	6	2	7	5	5	2	5	4
1	1	4	0	6	3	4	0	5	1
3	7	5	2	7	5	7	7	3	9
3	3	8	6	9	5	5	3	6	4
7	6	0	3	0	9	9	0	2	9
4	6	9	4	8	2	6	5	8	3
9	3	9	2	2	8	4	3	9	8
5	8	8	2	9	1	2	4	8	5
1	7	4	0	1	1	9	9	5	8

## Reflexiones ...

- Respuesta correcta. "Hay 14 sietes"
- Figura 1. Lleva tiempo y es probable haber cometido un error
- Figuras 2 y 3. Uso de atributos pre-atentivos hace el trabajo más fácil
- Figura 2. Hace uso del **tamaño** (grande vs pequeño)
- Figura 3. Hace uso del **color y forma**
- El color y el tamaño se procesan en la memoria icónica y por eso podemos diferenciarlos tan rápidamente
- El uso adecuado de atributos pre-atentivos reduce la cantidad de esfuerzo necesario para procesar de manera precisa y eficiente la información que se comunica mediante una visualización de datos

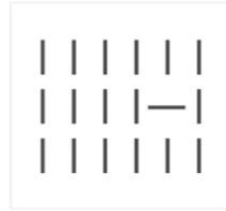
# Atributos pre-atentivos



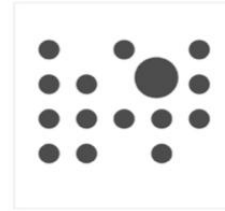
Length



Width



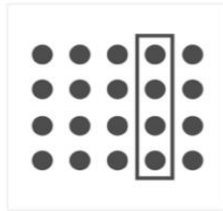
Orientation



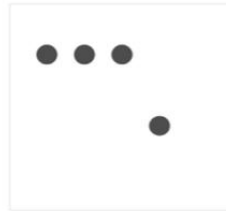
Size



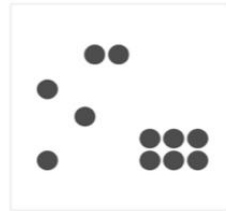
Shape



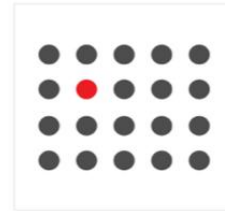
Enclosure



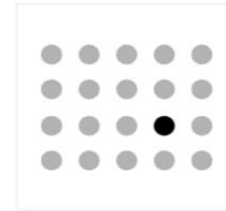
Position



Grouping



Color Hue

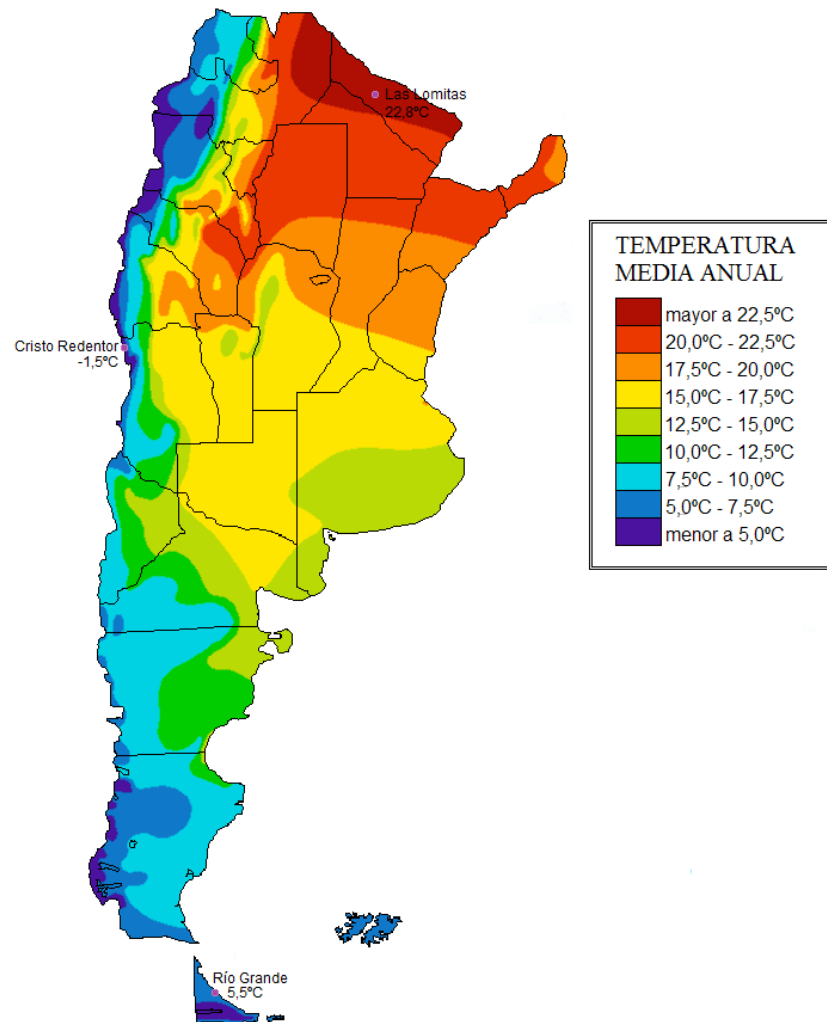
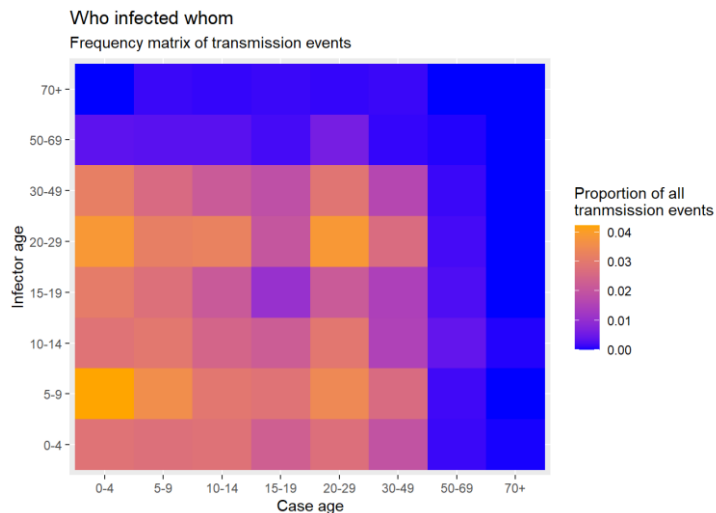


Color Intensity

- Más info en el libro

# Color

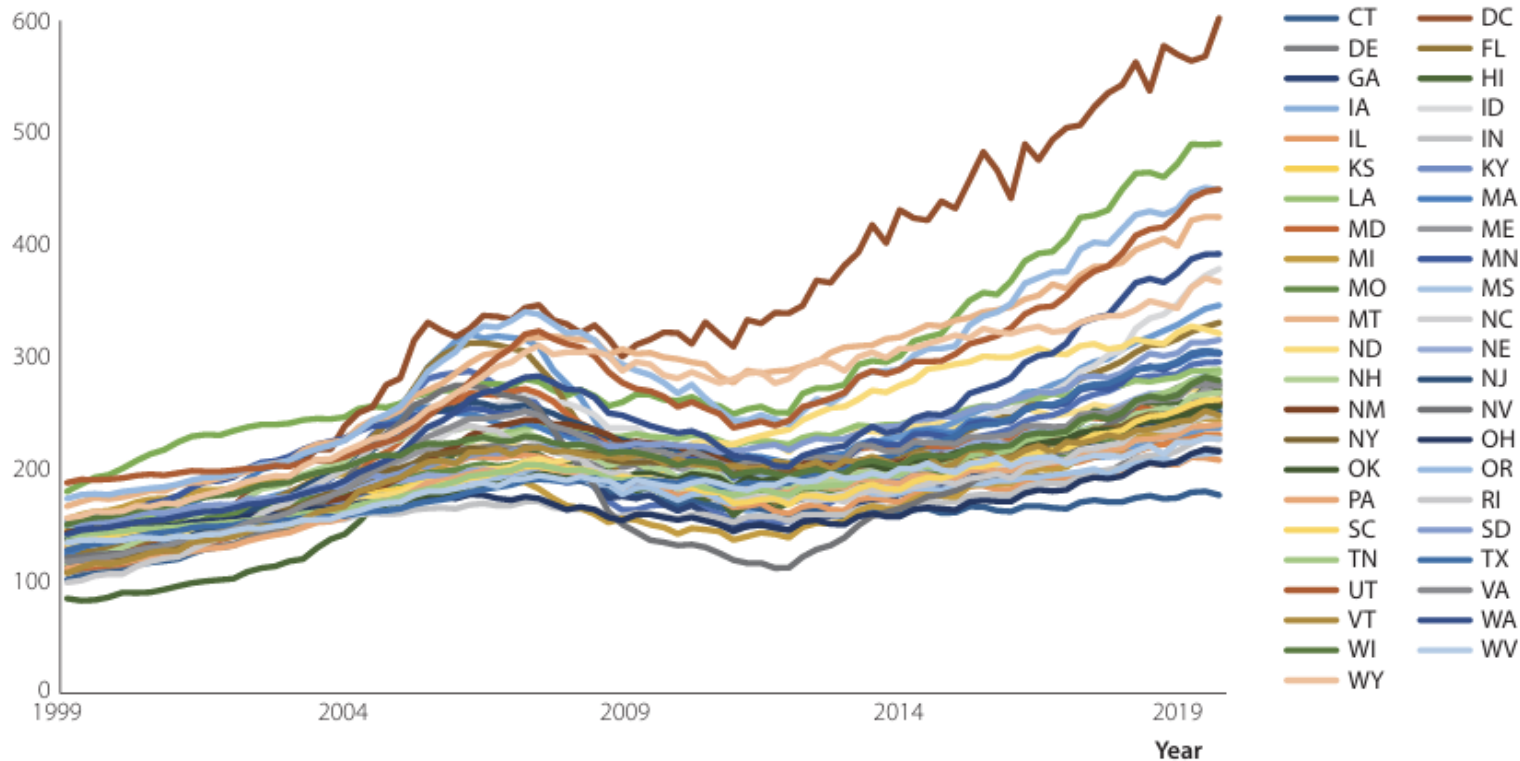
- Llama la atención
- Aumenta la comprensión
- Acelera la interpretabilidad



# Color

## Quarterly House Price Index by State (Base = 1991)

Home Price Index

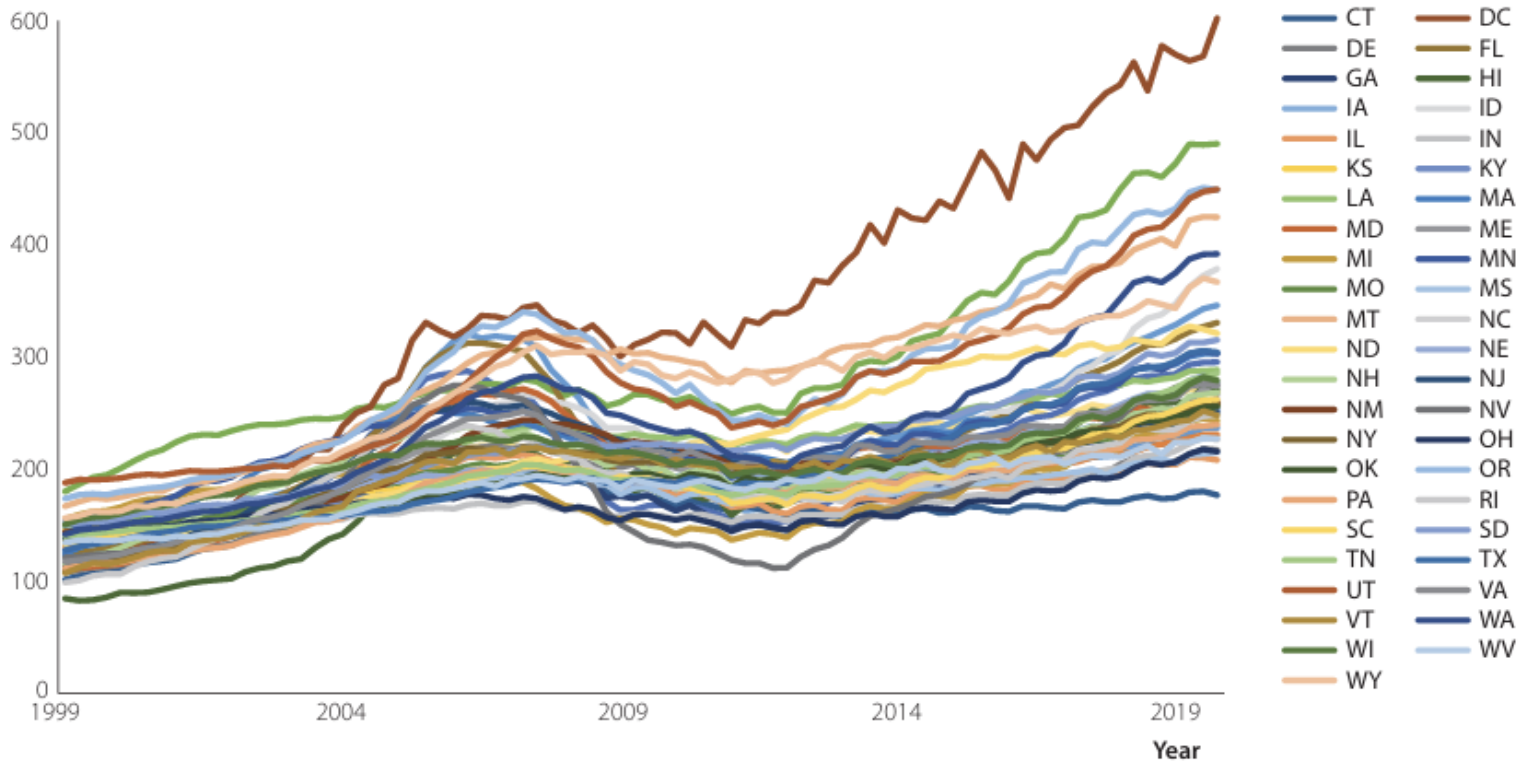


Color

Pero todo  
en su **justa**  
**medida**

Quarterly House Price Index by State (Base = 1991)

Home Price Index

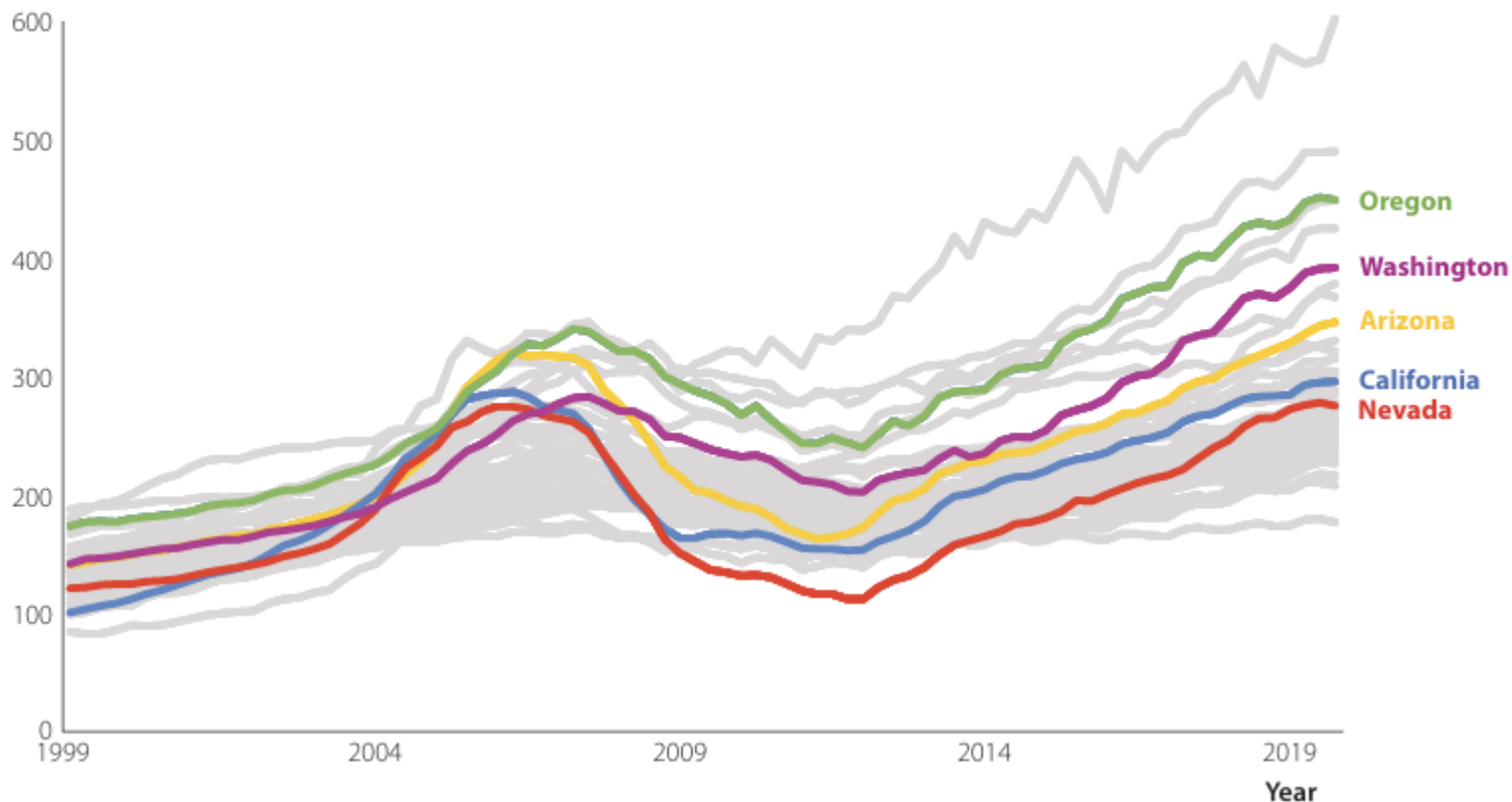


Color

Pero todo  
en su **justa**  
**medida**

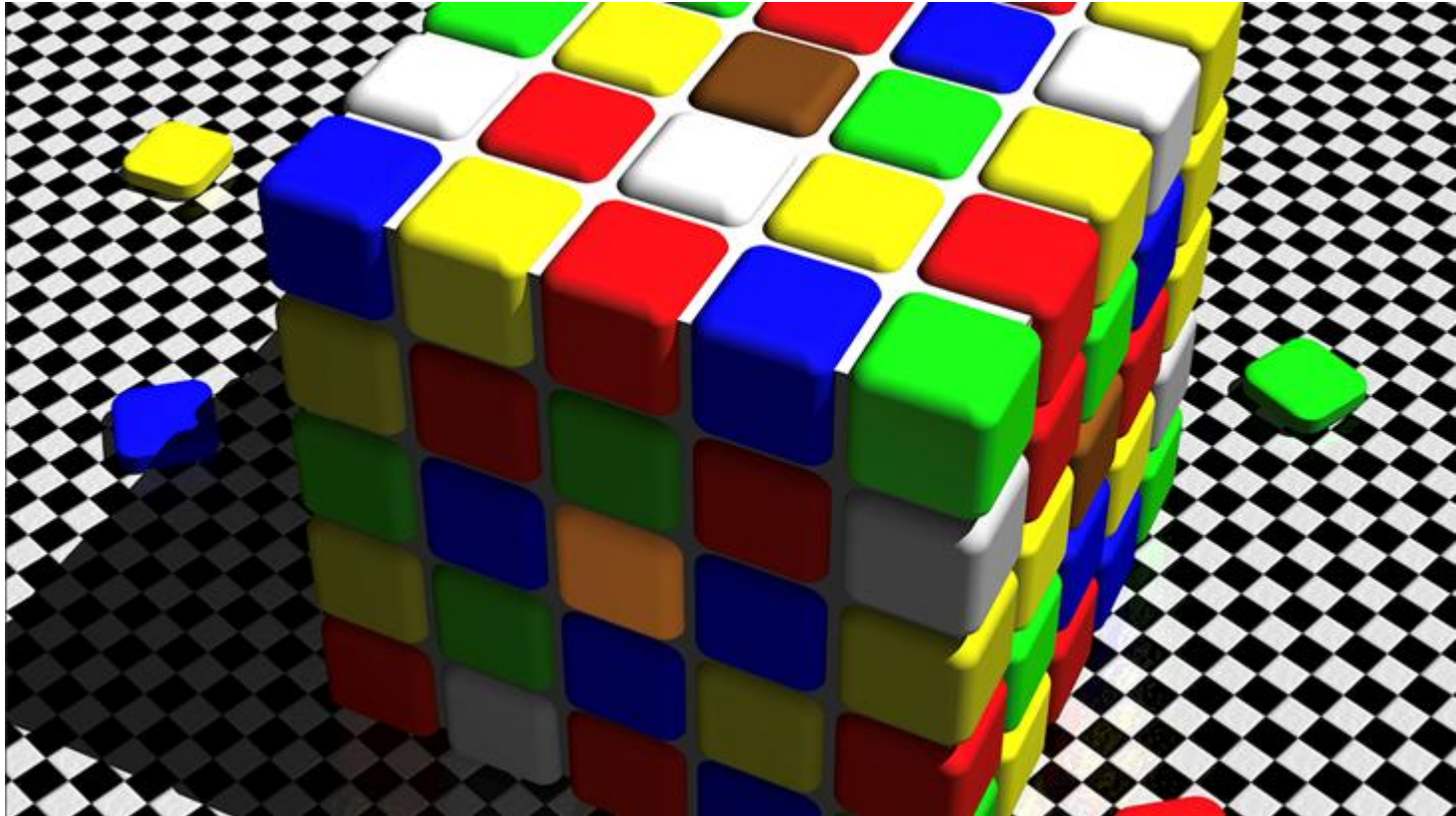
## Quarterly House Price Index (Base = 1991) for Western States Demonstrates More Volatility

Home Price Index

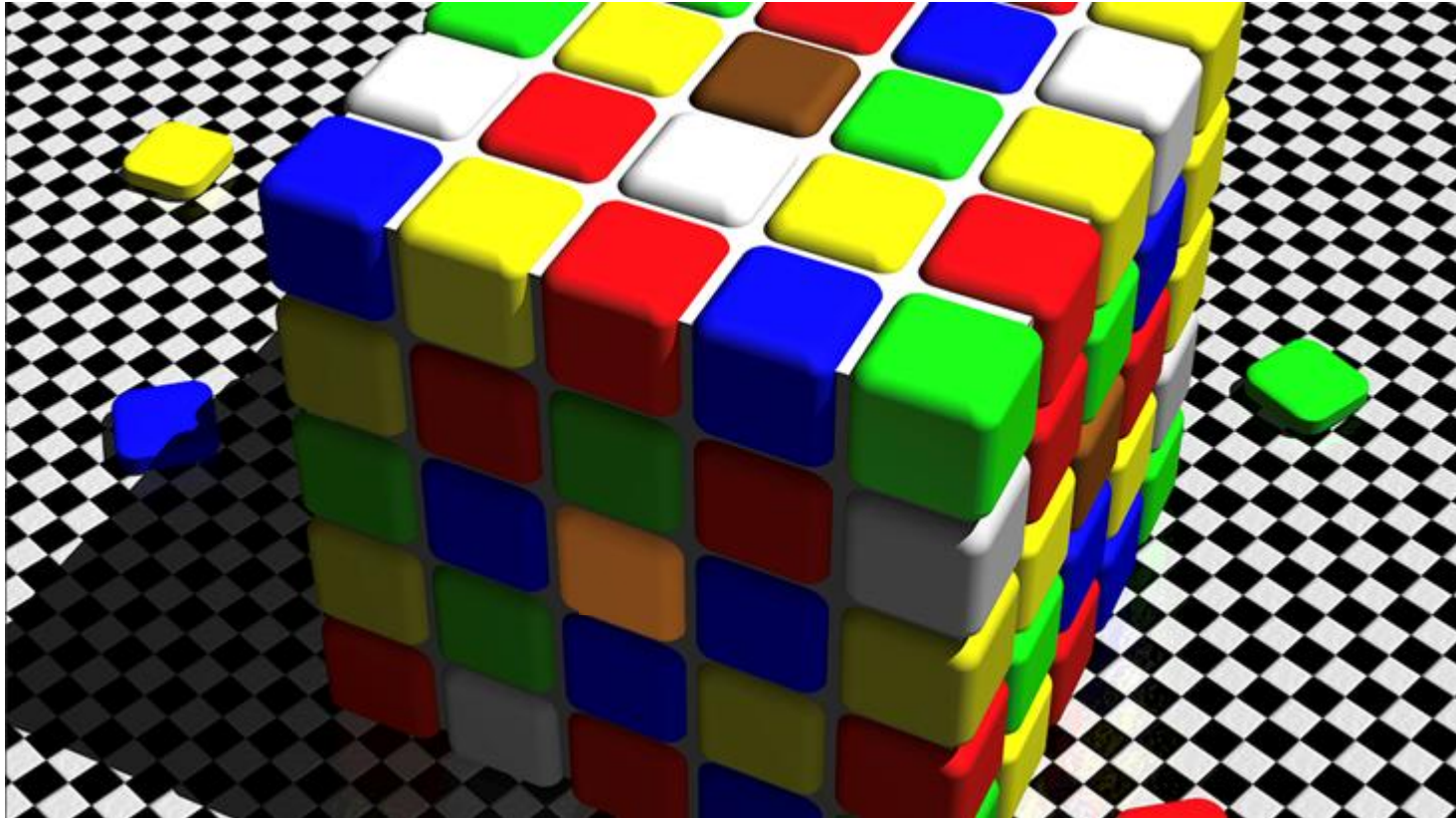




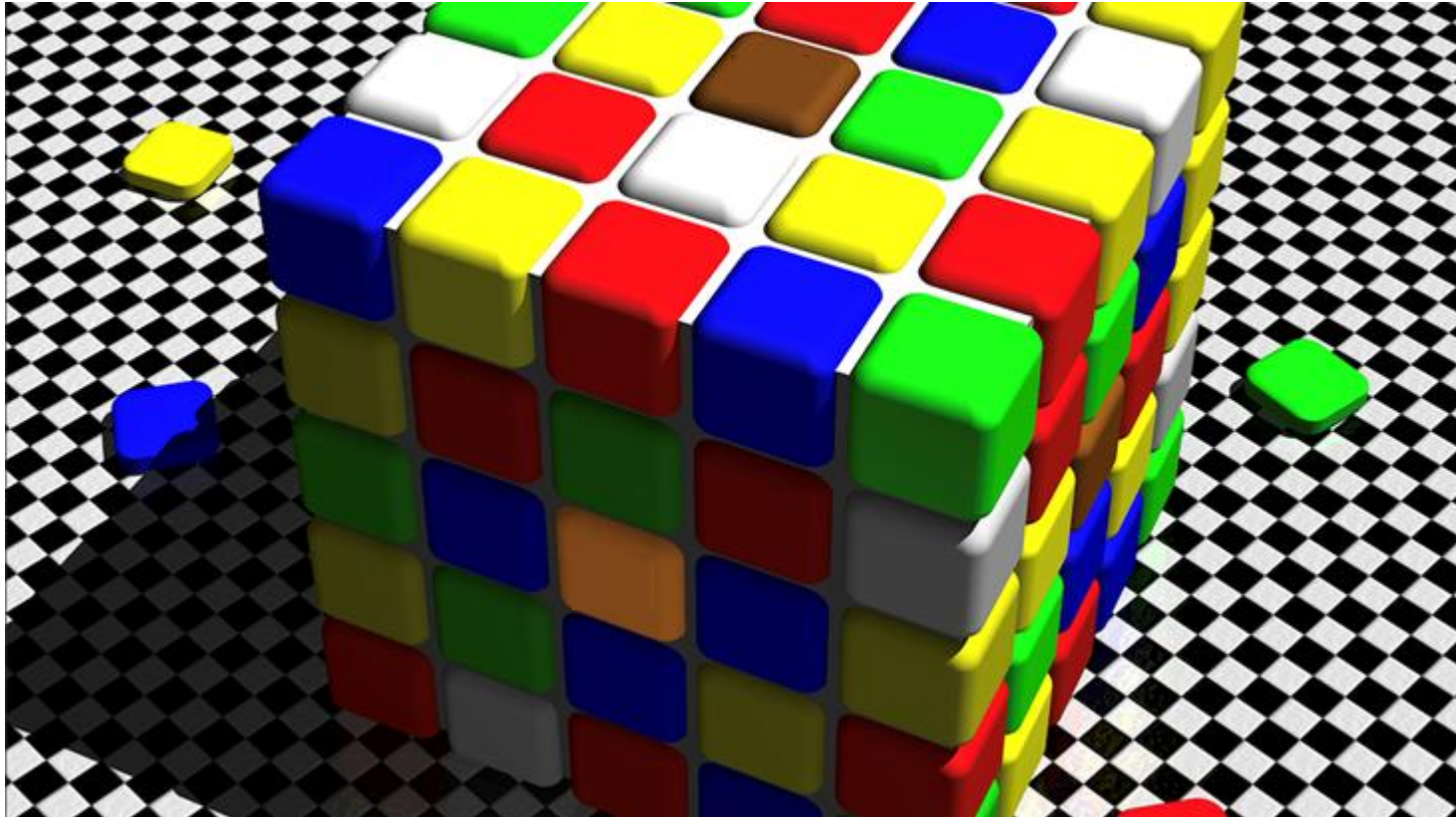
## Color: Las apariencias engañan



## Color: Las apariencias engañan



## Color: Las apariencias engañan



# Conociendo al público

# Color: Conociendo al público



Alrededor de 350.000.000 personas presentan algún tipo de ceguera al color en el mundo.

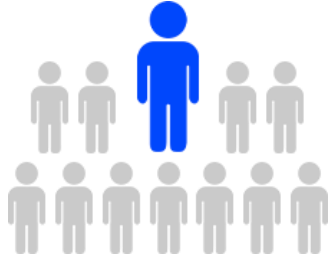
*Aproximadamente la misma cantidad que toda la población de EEUU.*



# Color: Conociendo al público



Alrededor de 350.000.000 personas presentan algún tipo de ceguera al color en el mundo.  
*Aproximadamente la misma cantidad que toda la población de EEUU.*

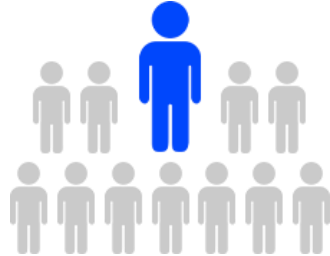


1 de cada 12 hombres presentan algún tipo de ceguera al color

# Color: Conociendo al público



Alrededor de 350.000.000 personas presentan algún tipo de ceguera al color en el mundo.  
*Aproximadamente la misma cantidad que toda la población de EEUU.*



1 de cada 12 hombres presentan algún tipo de ceguera al color



1 de cada 200 mujeres presentan algún tipo de ceguera al color

# Color: Conociendo al público



Visión normal



# Color: Conociendo al público



Visión normal



Deuteranomalía  
(verde débil)



Protanomalía  
(rojo débil)



Trianomalía  
(azul débil)



Acromatopsia  
(monocromacia)



Deuteranopia  
(verde ciego)



Protanopia  
(rojo ciego)

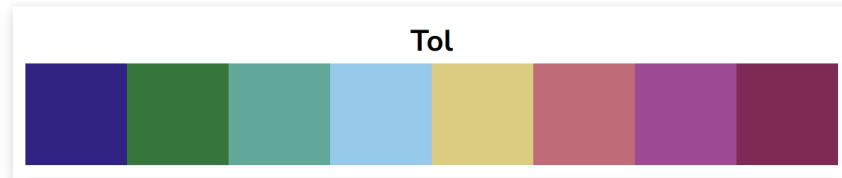
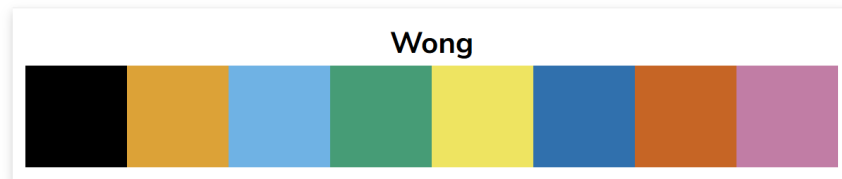
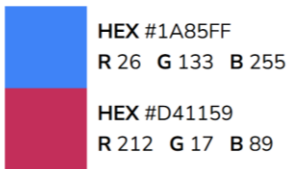
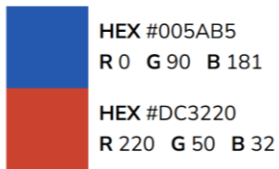
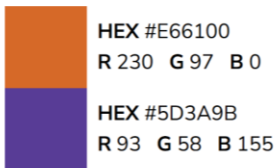
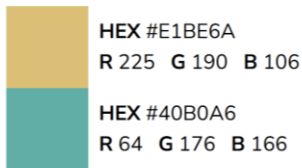
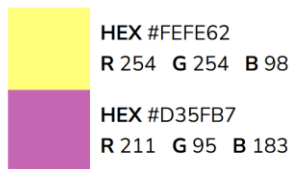
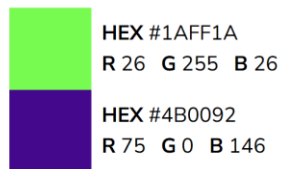
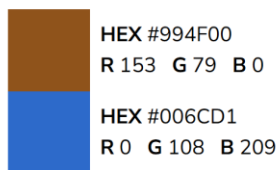
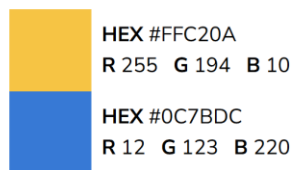


Tritanopia  
(azul ciego)



Monocromacia  
del cono azul

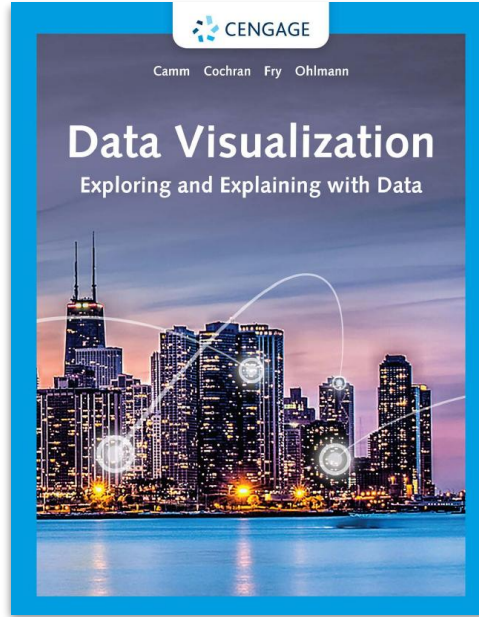
# Paletas *colorblind* friendly



# Tarea

- Resolver la guía de ejercicios de visualización

# Bibliografía



Camm/Cochran/Fry/Ohlmann, Data Visualization: Exploring and Explaining with Data, 1st. Edition, Cengage Learning, 2022