

DATA SCIENCE



Federico Baiocco
baioccofede@gmail.com
3512075440



Clase 8 - Agenda

VISUALIZACIÓN

- ¿ Para qué ?
- Tipos de datos
- Gráficos
- Matplotlib
- Seaborn

¿ Dudas de la clase
pasada ?

¿Terminaron todo lo de Pandas?





Visualización

Visualizar los datos es una parte fundamental del análisis en ciencia de datos.

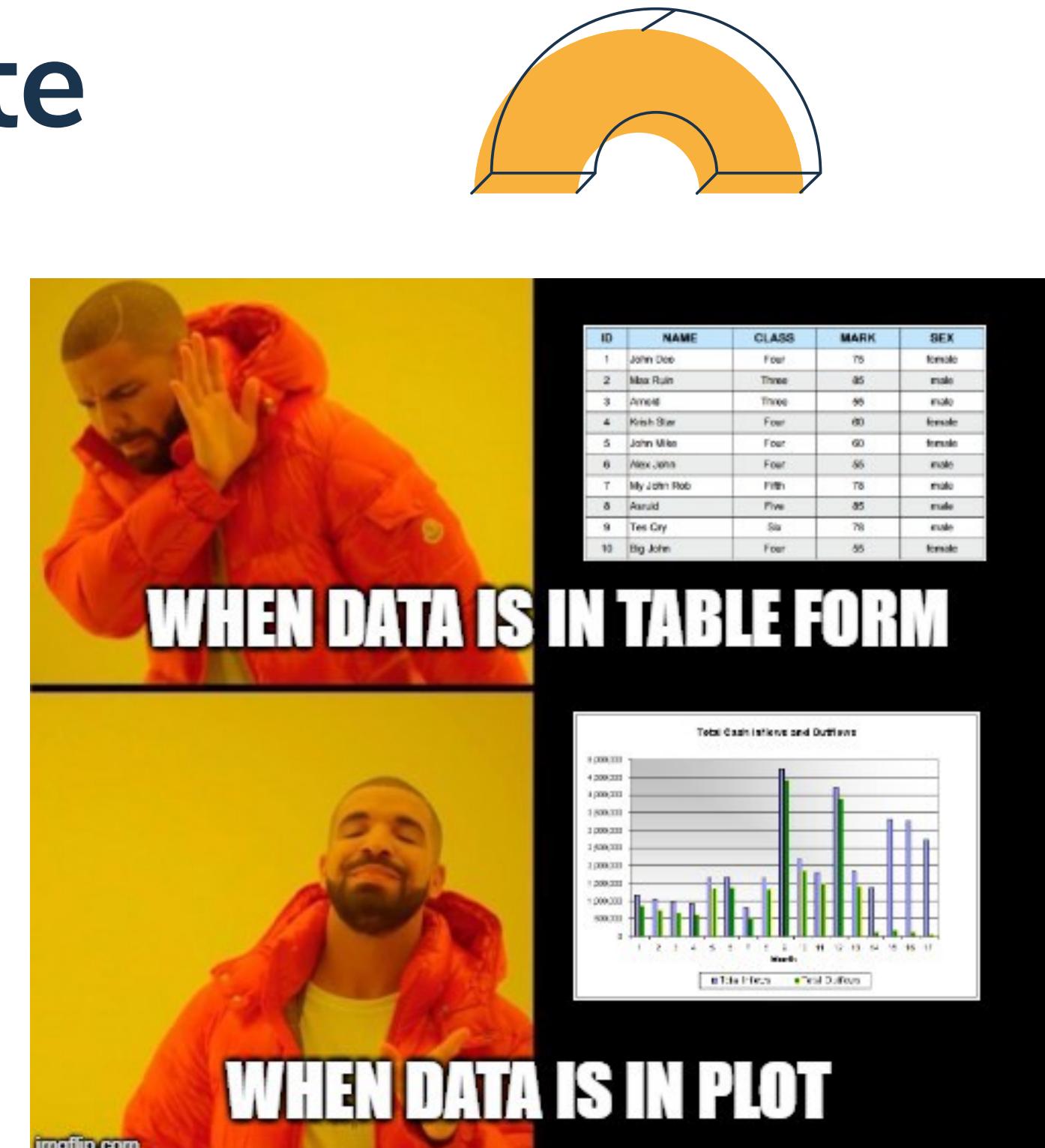
¿ Por qué es tan importante la visualización de datos ?

Por la forma en que el cerebro humano procesa la información, el uso de diagramas o gráficas para visualizar grandes cantidades de datos complejos es más fácil que vaciarlos en hojas de cálculo o reportes.

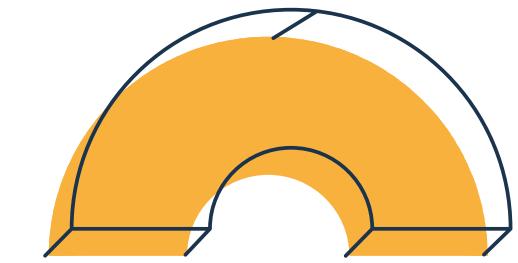
La visualización de datos es una forma fácil y rápida de transmitir conceptos de manera universal.

ADEMÁS ...

No solo sirve para **comunicar** (que es una parte fundamental del trabajo) sino que también es una herramienta esencial para **comprender el dataset** con el que estamos trabajando.

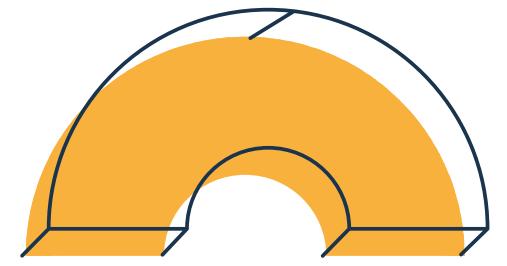


¿Qué debemos tener en cuenta a la hora de graficar?



A la hora de crear visualizaciones, hay que tener en cuenta:

- Tipo de datos:
 - cuantitativos (altura de una persona)
 - categóricos (género)
 - discretos (cantidad de personas)
- Tipo de gráfico:
 - scatter
 - barras
 - torta
- Aspectos estéticos:
 - colores
 - formas
 - ejes



Tipos de datos

En general, nos vamos a encontrar con datos de estos tipos:

- Numéricos continuos: Por ejemplo distancias
- Numéricos discretos: Por ejemplo cantidad de contagios de covid
- Categóricos: Por ejemplo la raza de un perro
- Categóricos ordenables: Por ejemplo calificaciones (bueno, regular, malo)
- Fechas
- Texto

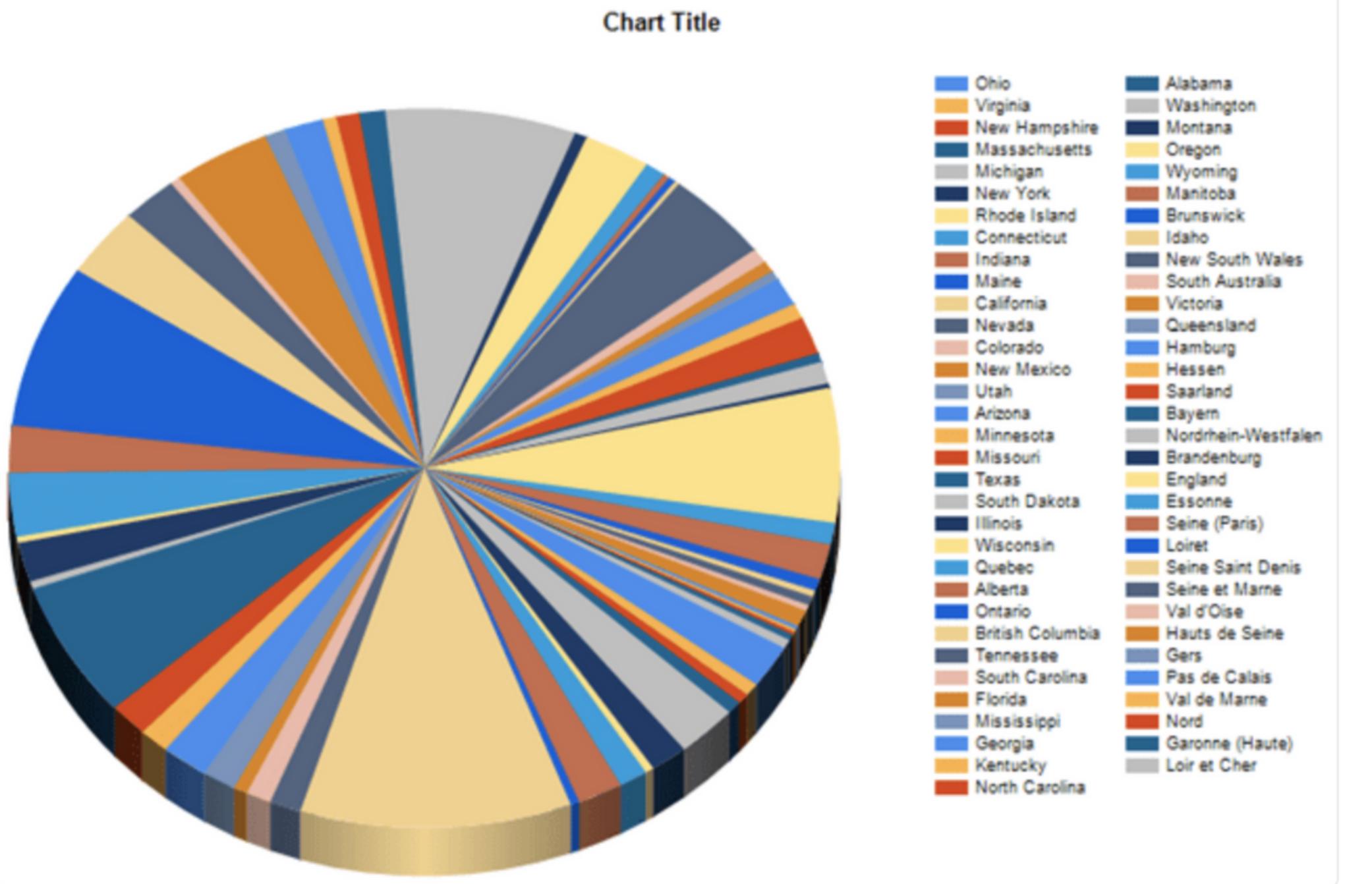
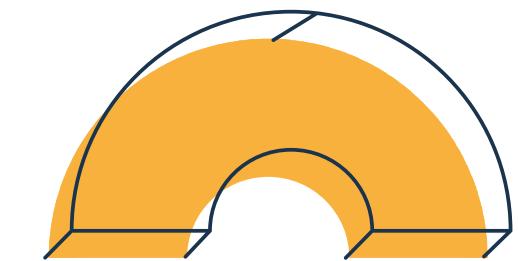
Tipos de gráficos

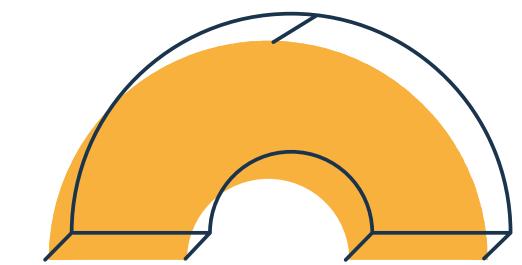


Los más comunes y que más vamos a utilizar son:

- Scatter plot
- Line plot
- Histogram
- Bar chart
- Pie chart

La importancia de elegir bien...

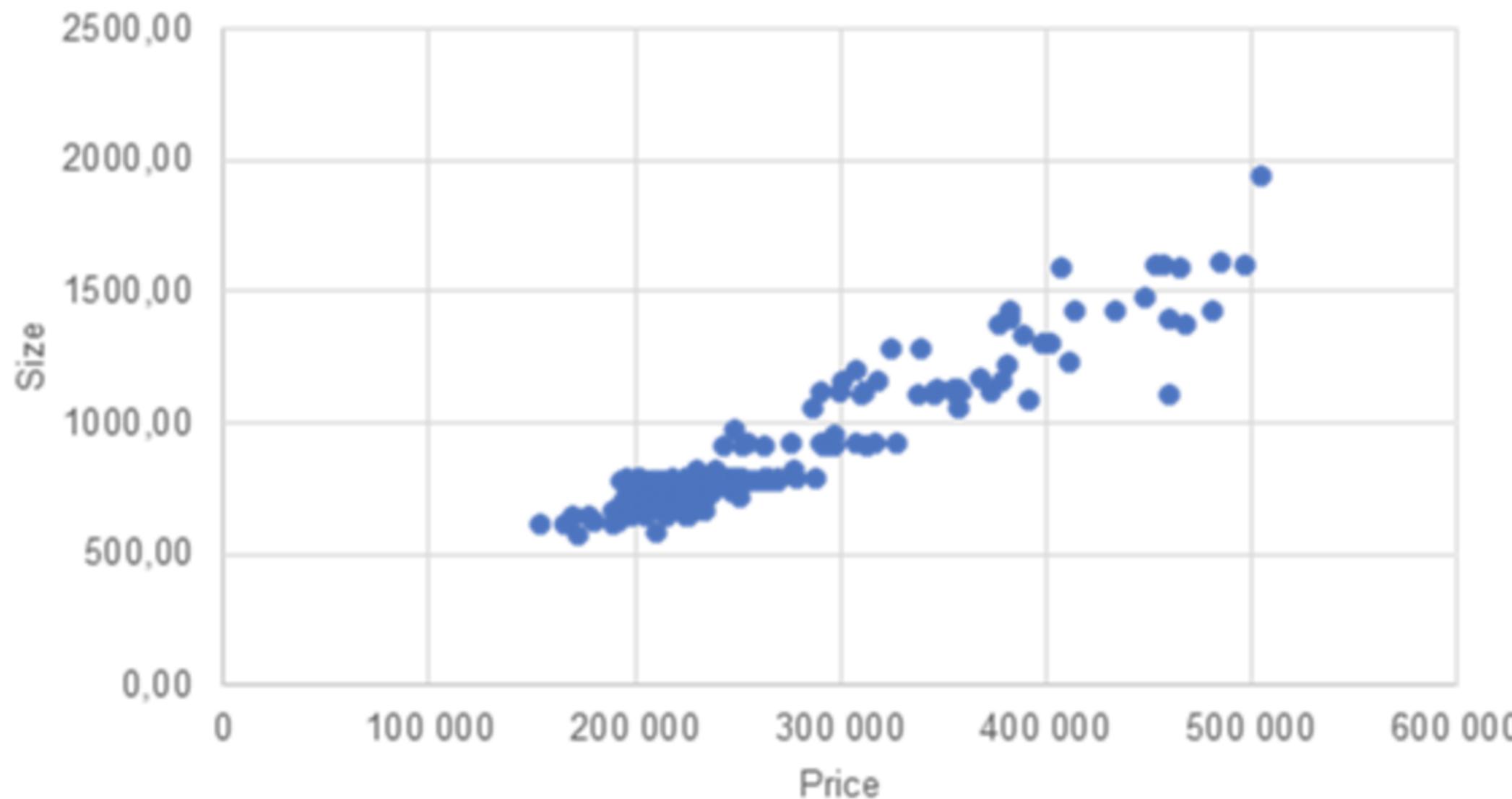


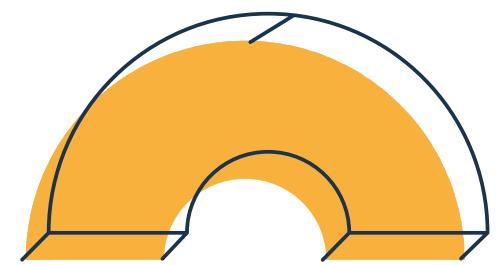


Scatter plot

Un Scatter Plot consiste en graficar una serie de puntos, con coordenadas (x,y) , sobre un plano.

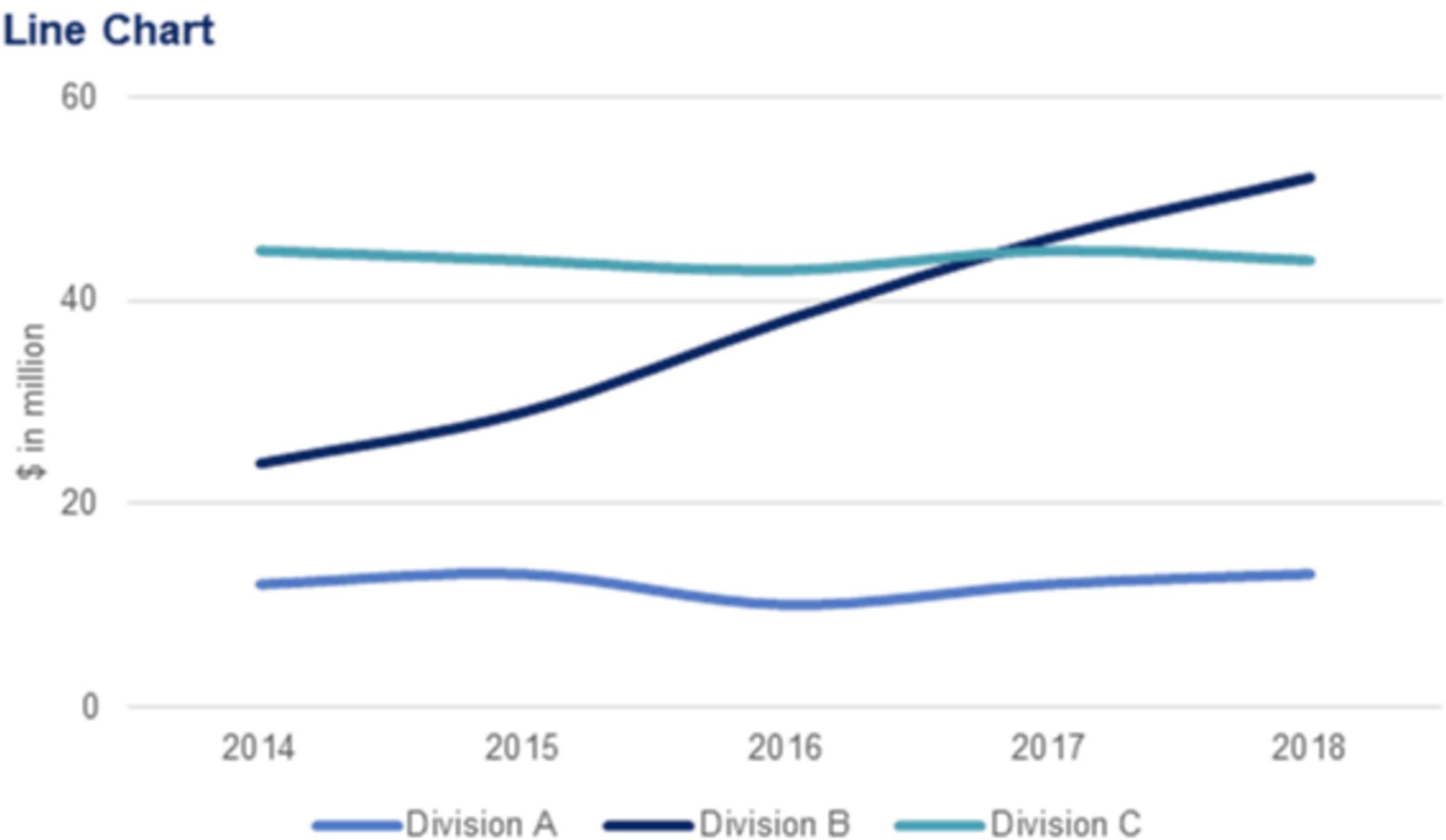
Scatter Plot - Positive Relationship



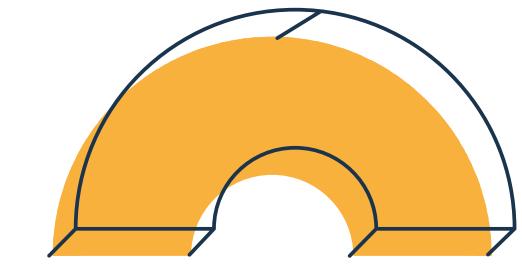


Line chart

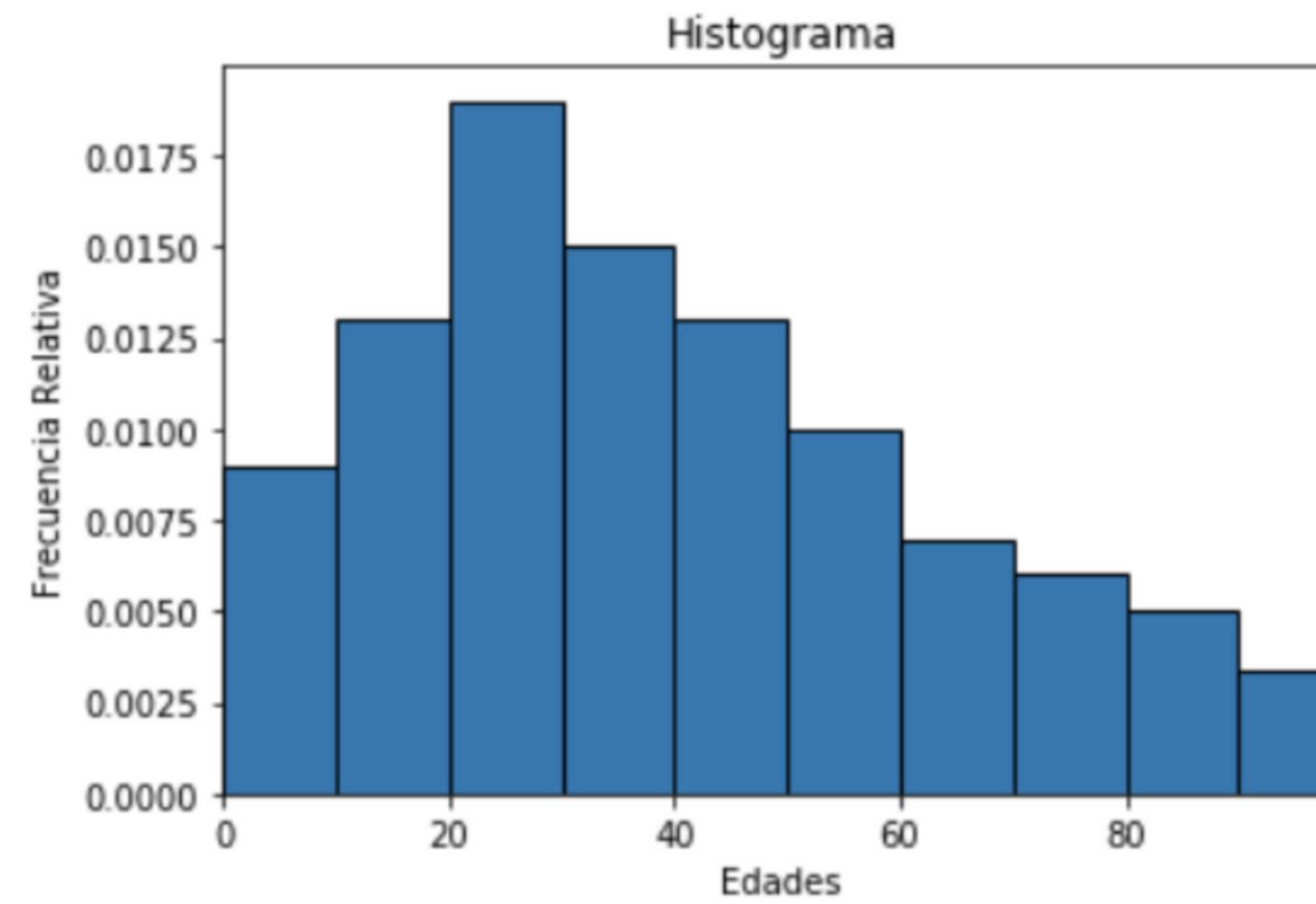
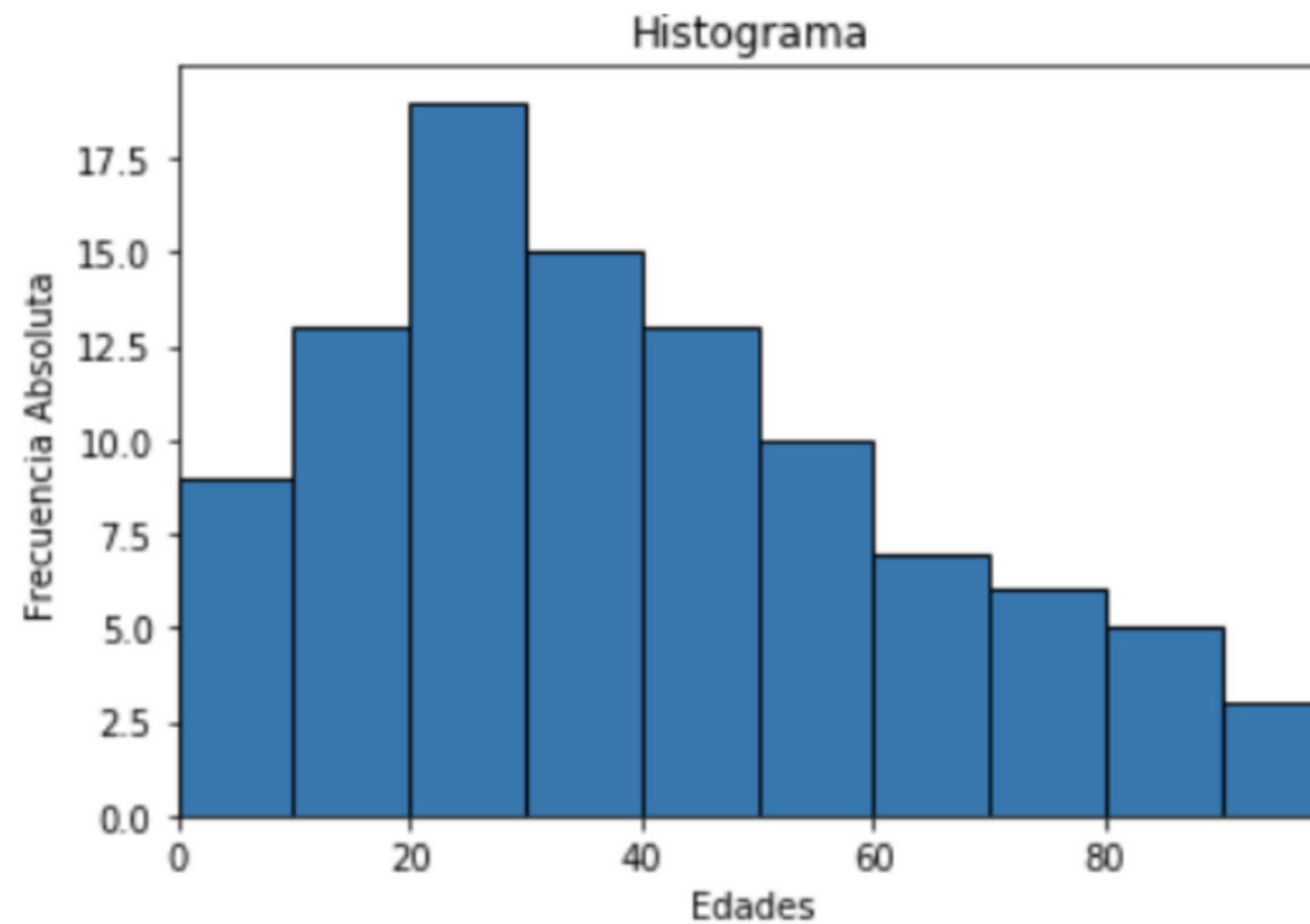
Muestra información como una serie de puntos de datos llamados "marcadores" conectados por segmentos de línea recta.



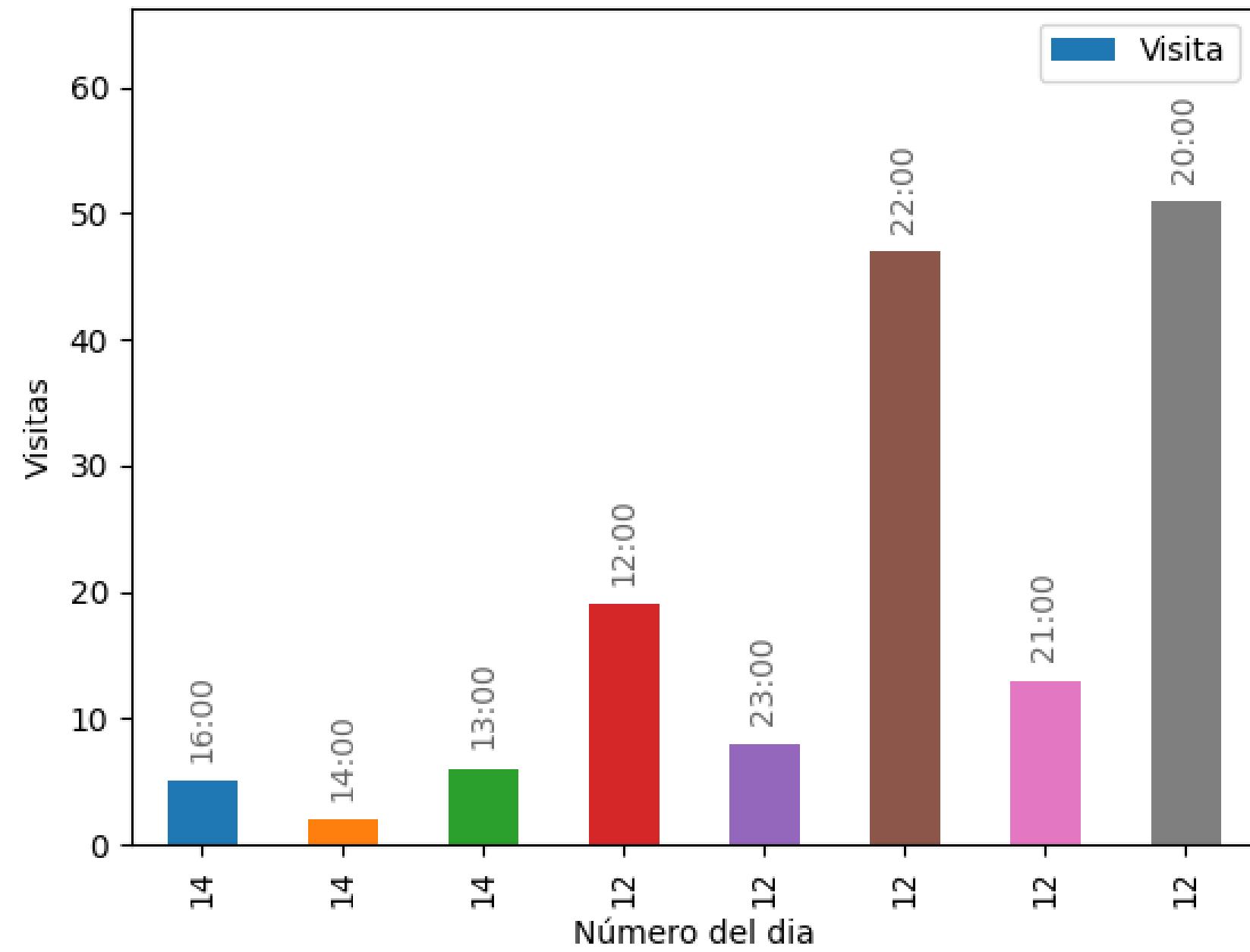
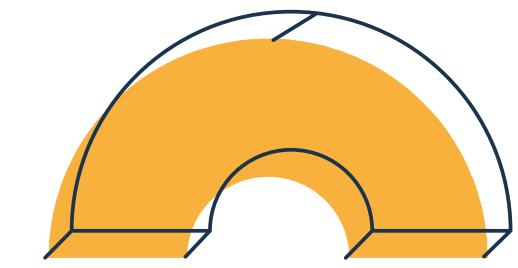
Histogram



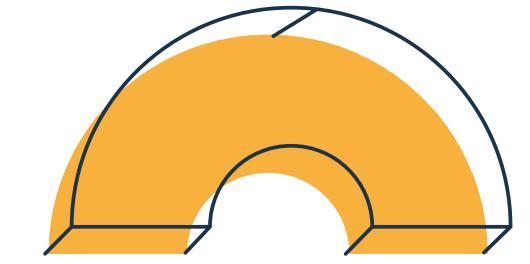
Es la representación gráfica en forma de barras, que simboliza la distribución de un conjunto de datos.



Bar chart

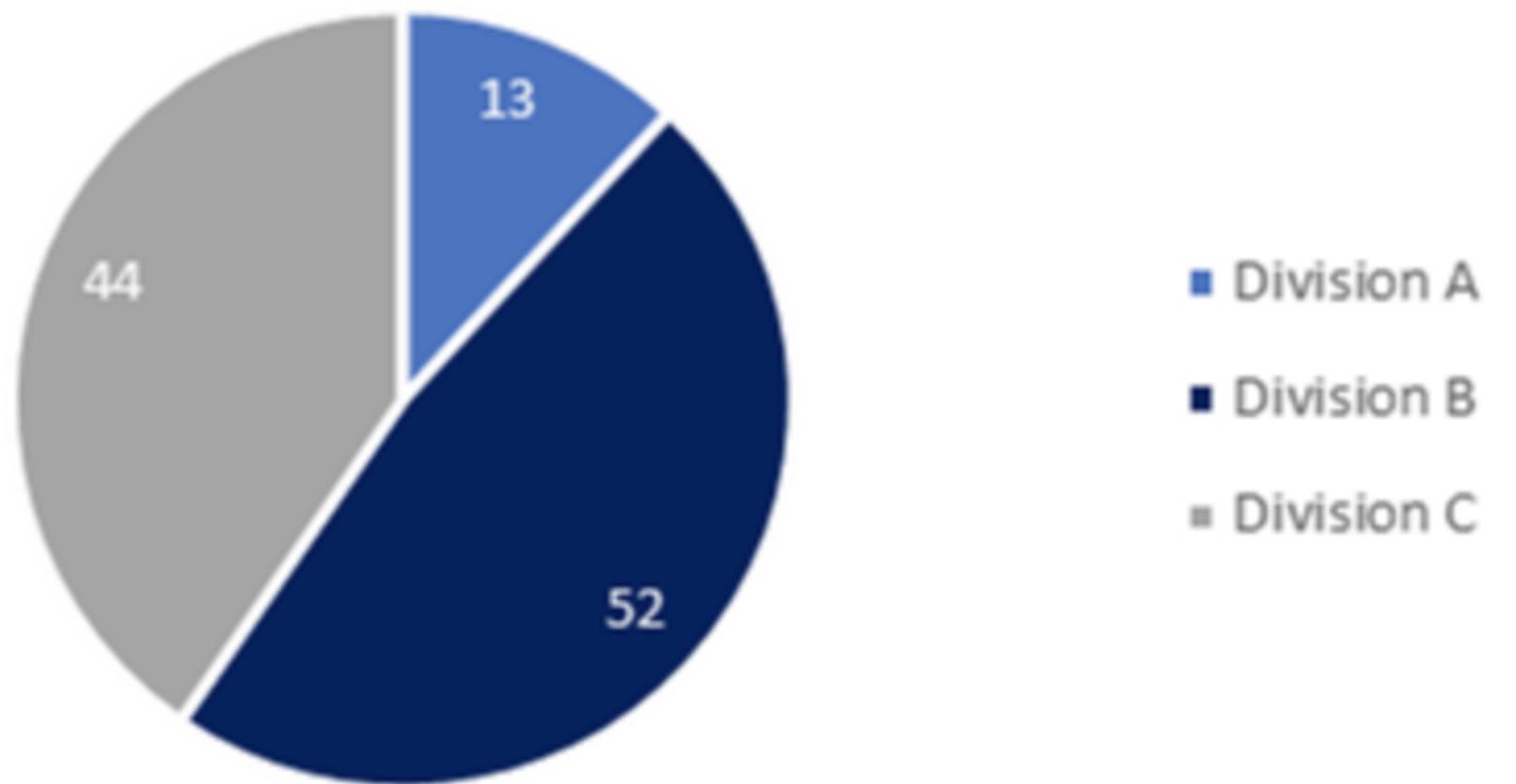


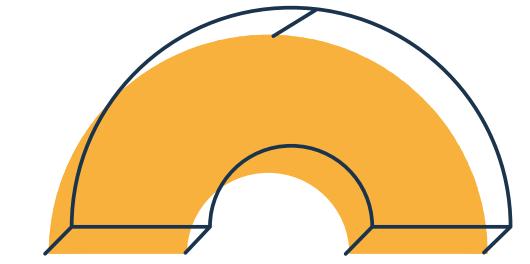
Pie chart



Es un gráfico circular dividido en porciones que representan el porcentaje de la cantidad total.

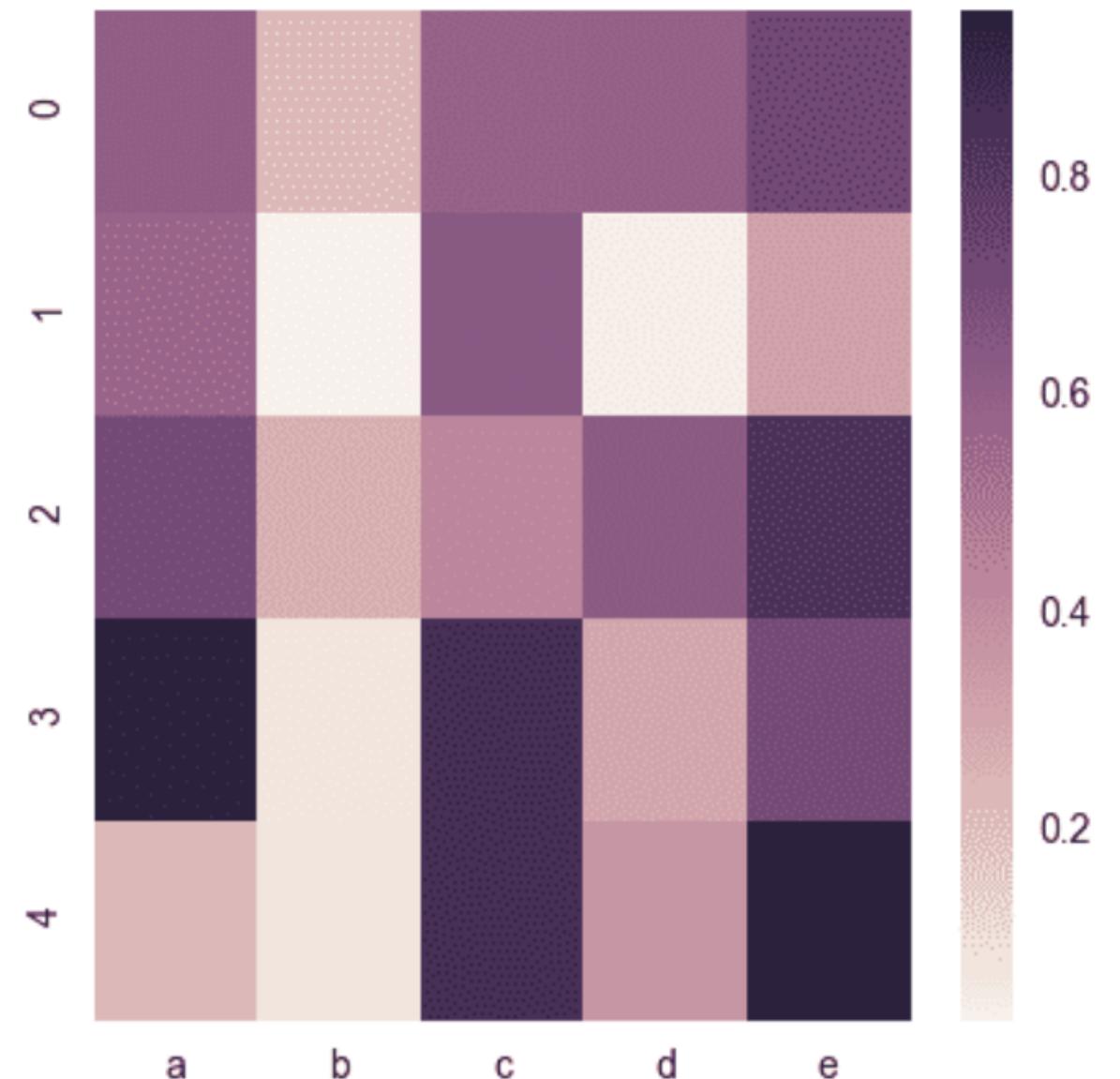
Pie Chart



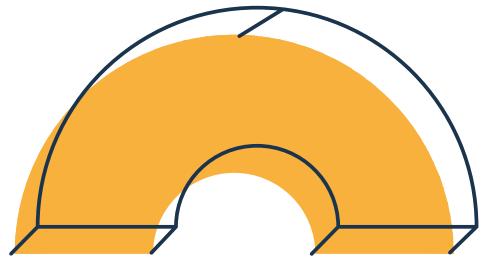


Heatmap

Un mapa de calor es una representación gráfica de datos donde cada valor de una matriz se representa como un color.



¿Qué herramientas vamos a utilizar para visualizar ?

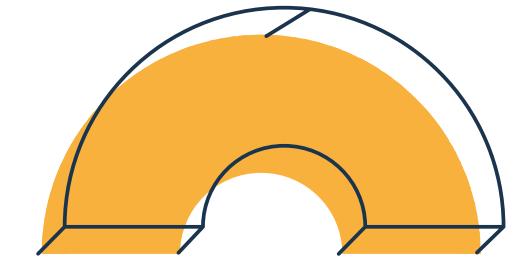


matplotlib



seaborn

Algunos links útiles



A la hora de elegir un gráfico, podemos utilizar estas herramientas para guiarnos. Existen muchos tipos de gráficos y es importante elegir bien

<https://datavizcatalogue.com/search.html>

<https://www.data-to-viz.com/>

<https://www.python-graph-gallery.com/>

matplotlib



Abrimos notebook matplotlib



Preguntas?

Para la próxima clase:

- Vamos a comenzar con análisis exploratorio de datasets y necesitamos tener bien claro lo que vimos hasta ahora