

Business Intelligence & data visualization

Département d'informatique

École supérieure d'ingénieurs Léonard-de-Vinci (ESILV)

25 novembre, 2024

Ruiwen HE

Assessment

- Assessment for this course will be conducted together with the Computer Vision module.
- Objective: Showcase and analyze your project outcomes.

1. Illustrate your problematic

- Presenting your problem

2. Dataset Visualizations

- Clear, detailed views of key data points and structures

3. Illustrate your methodology

- Flow chart or other tools to show your main idea

4. Computational Results

- Graphs, heatmaps, or charts to illustrate results

5. Analytical Insights

- Summary of findings and interpretation of results

Defense

- Mercredi 15 janvier 2025 08h15
 - Group: ESILV-5-A5-CORE-DIA3
- Jeudi 16 janvier 2025 08h15
 - Group: ESILV-5-A5-CORE-DIA1
- Vendredi 17 janvier 2025 08h15
 - Group: ESILV-5-A5-CORE-DIA1
- Presentation Guidelines
- Group Size: 4-5 students per group
- Presentation Time: 30 minutes, followed by 10 minutes for questions
- Please have each group leader send me the list of group members by December 1, 2024.

Introduction to data visualization

1. What is Business Intelligence (BI)?
2. What is data visualization?
3. Why visualization is important?
4. Data types and levels of measurement
5. Graphical Basics
6. Choosing the right type of visualization



What is Business Intelligence (BI)?

- Business Intelligence (BI) refers to the processes, technologies, tools, and systems that organizations use to collect, analyze, and present business data. The goal of BI is to support better business decision-making by transforming raw data into meaningful insights.

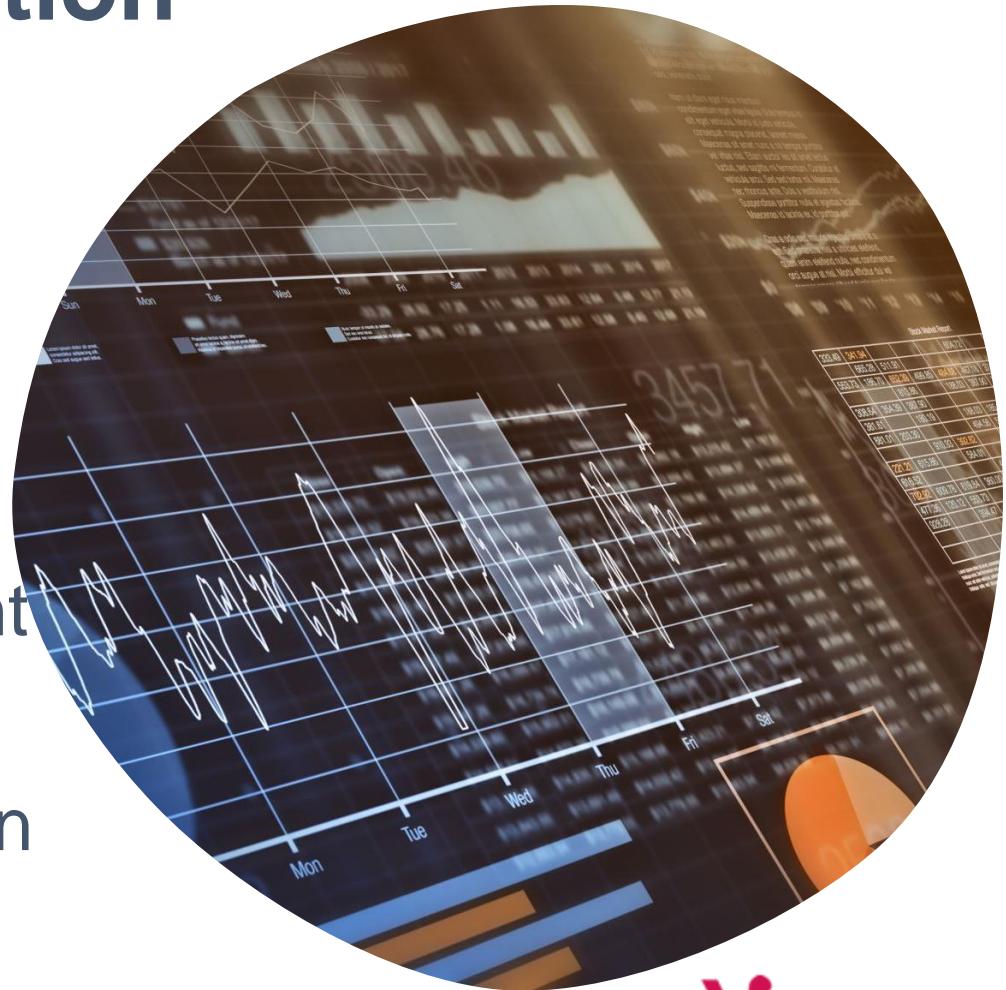


Key Components of BI

- **Data Collection:** Gathering data from internal and external sources (e.g., sales, social media, market trends).
- **Data Storage:** Storing data in structured formats (e.g., databases, data warehouses, cloud storage).
- **Data Integration & Cleaning:** Combining data from different sources and ensuring accuracy by removing errors and inconsistencies.
- **Data Analysis & Reporting:** Analyzing data to identify trends and insights. Generating reports and dashboards.
- **Data Visualization:** Creating graphs, charts, and dashboards to present data in an understandable format.
- **Decision-Making Support:** Using insights to make informed decisions and optimize business strategies.

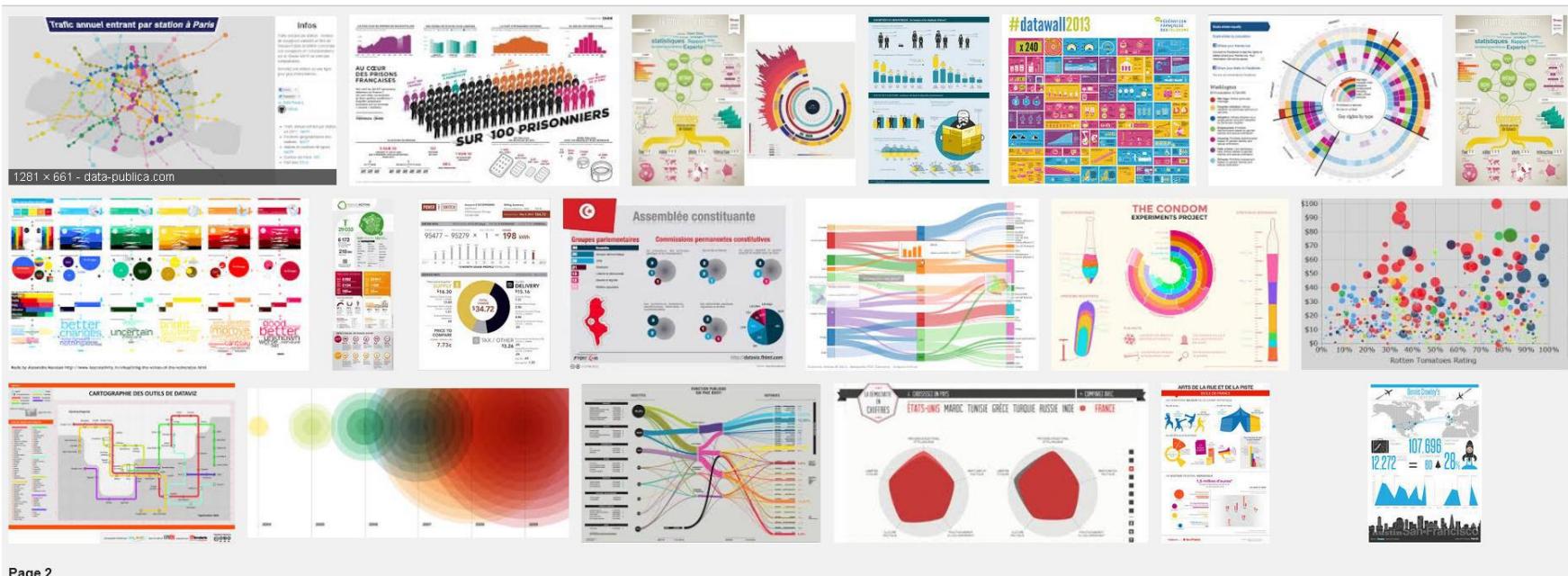
Introduction to data visualization

1. What is Business Intelligence (BI)?
2. **What is data visualization?**
3. Why visualization is important?
4. Data types and levels of measurement
5. Graphical Basics
6. Choosing the right type of visualization



What is data visualization?

- Data visualization is the technique to present data in a format of picture or graph, such as charts, plots, infographics, and even animations.



Page 2

What is data visualization?

- Data visualization enables stakeholders and decision makers to :



What is data visualization?

- Data visualization enables stakeholders and decision makers to :

 analyze data visually



What is data visualization?

- Data visualization enables stakeholders and decision makers to :



analyze data visually



identify new trends and patterns easily

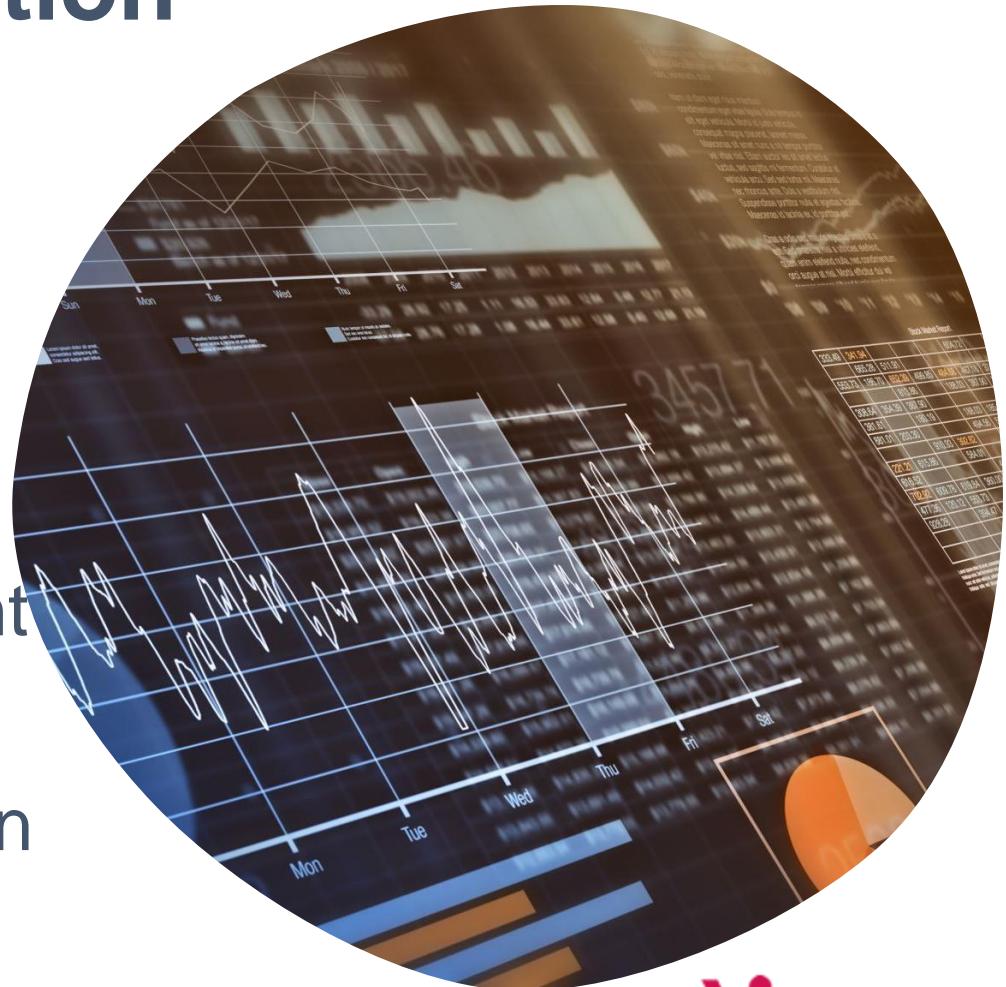


Why visualization is important?



Introduction to data visualization

1. What is Business Intelligence (BI)?
2. What is data visualization?
- 3. Why visualization is important?**
4. Data types and levels of measurement
5. Graphical Basics
6. Choosing the right type of visualization



Imagine: If you're a data engineer



<https://www.vecteezy.com/vector-art/23752994-data-scientist-icon-in-vector-illustration>

If you're a data engineer



<https://www.vecteezy.com/vector-art/23752994-data-scientist-icon-in-vector-illustration>



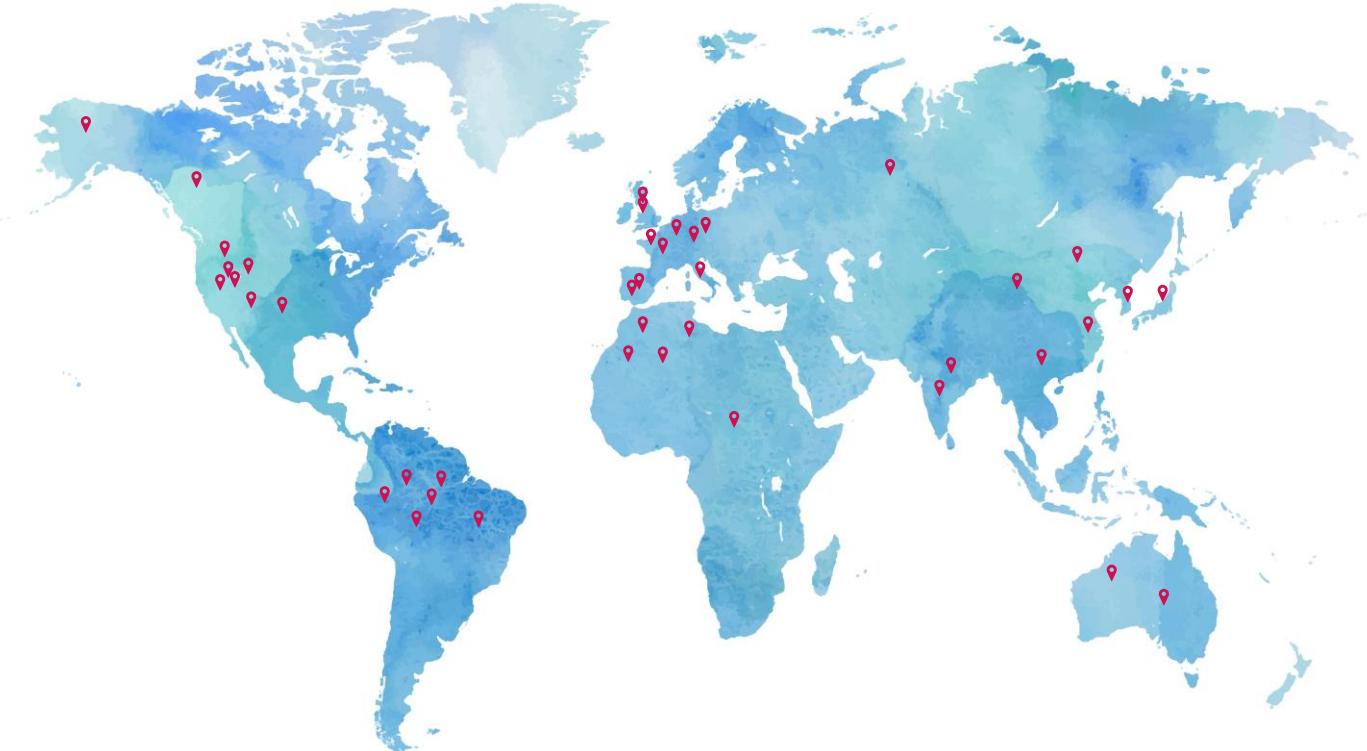
If you're a data engineer

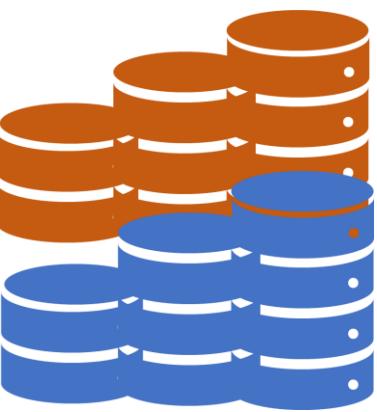


<https://www.vecteezy.com/vector-art/23752994-data-scientist-icon-in-vector-illustration>



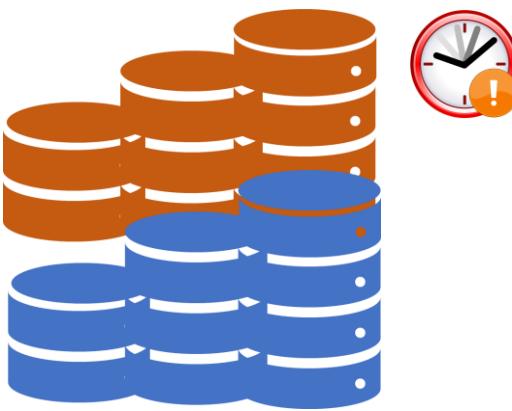
Your manager plan to study the sale details of one product





huge and complex

You don not how to visualize data



huge and complex

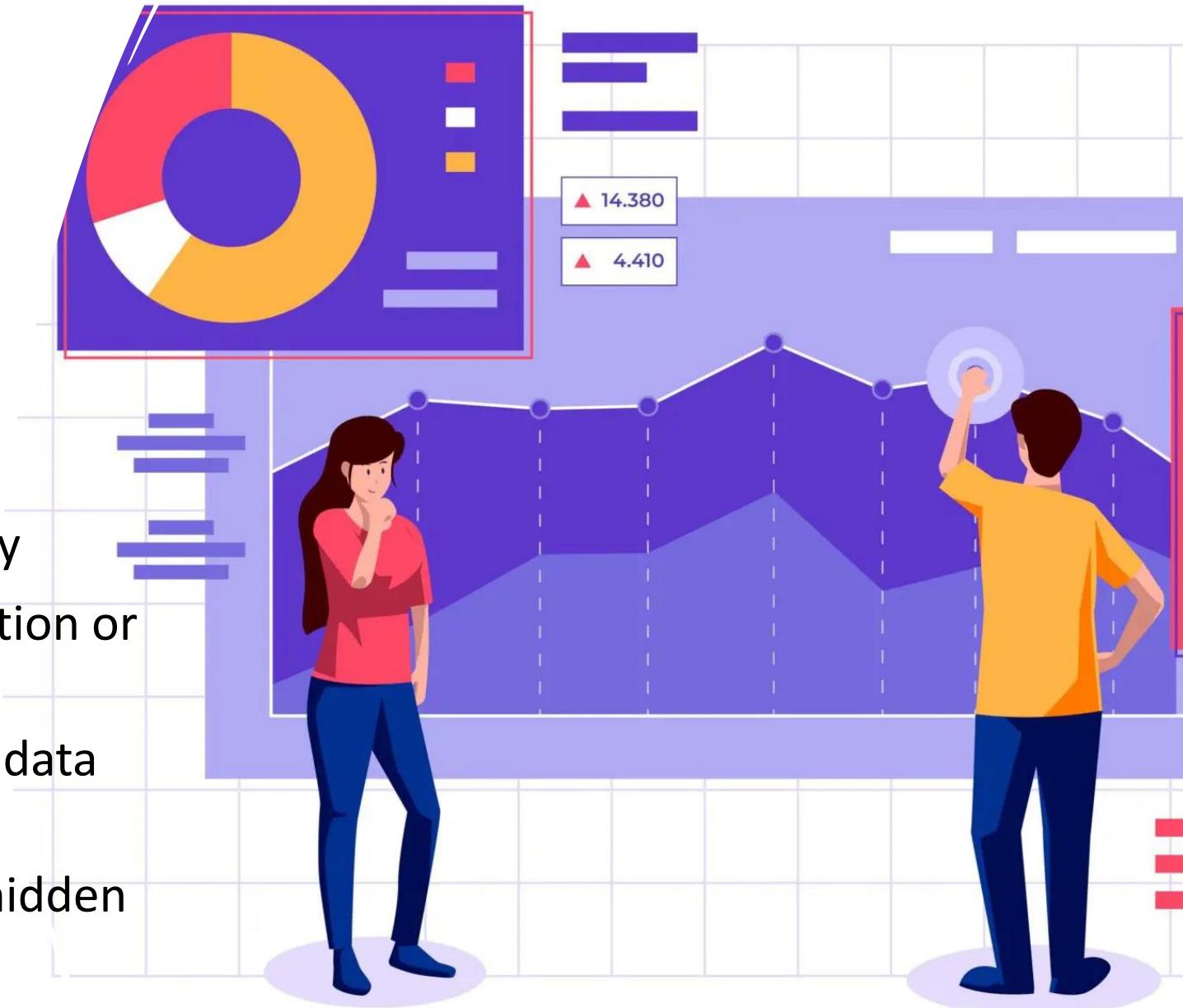
If you are a good visual designer



<https://www.livemint.com/news/business-of-life/the-art-of-decision-making-1541211945989.html>

The main benefits of data visualization

- Simplify the complex quantitative information
- Analyze and explore big data easily
- Identify the areas that need attention or improvement
- Identify the relationship between data points and variables
- Explore new patterns and reveal hidden patterns

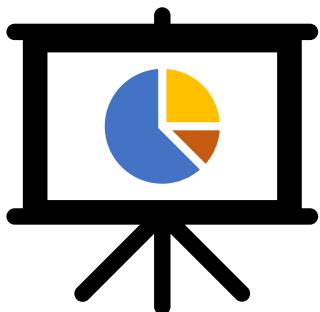


Data Visualization Considerations

There are three major considerations for data visualization they are

Data Visualization Considerations

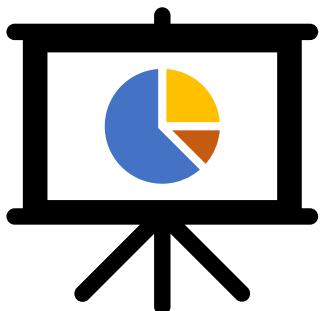
There are three major considerations for data visualization they are



Clarity

Data Visualization Considerations

There are three major considerations for data visualization they are



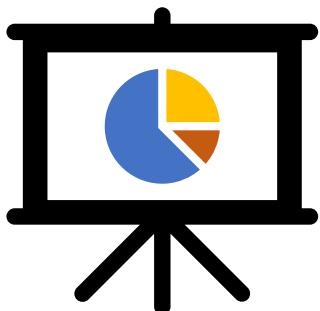
Clarity



Accuracy

Data Visualization Considerations

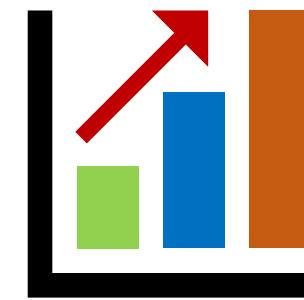
There are three major considerations for data visualization they are



Clarity



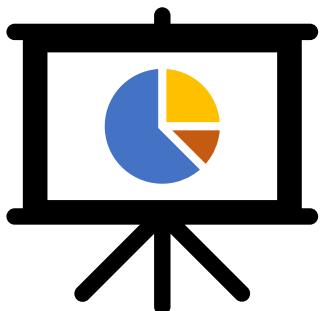
Accuracy



Efficiency

Data Visualization Considerations

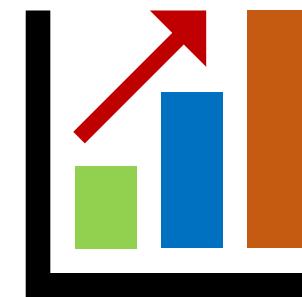
There are three major considerations for data visualization they are



Clarity



Accuracy



Efficiency

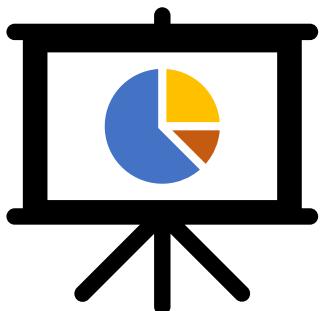
Ensure the dataset is complete and relevant.

Ensure the dataset is complete and relevant

- Relevancy: the data should meet the requirements for the intended use.
- Completeness: the data should not have missing values or miss data records

Data Visualization Considerations

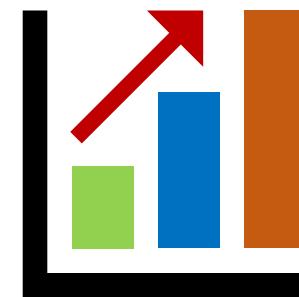
There are three major considerations for data visualization they are



Clarity



Accuracy



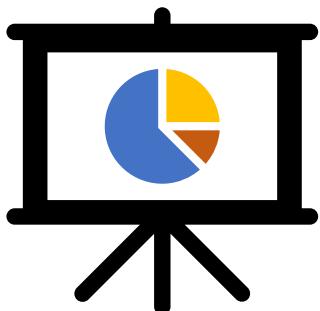
Efficiency

Ensure using appropriate graphical representation
to convey the right message

ruiwen.he@devinci.fr

Data Visualization Considerations

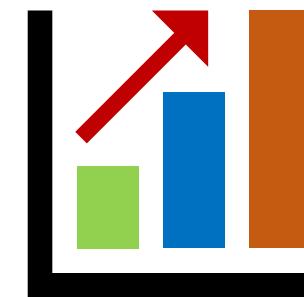
There are three major considerations for data visualization they are



Clarity



Accuracy

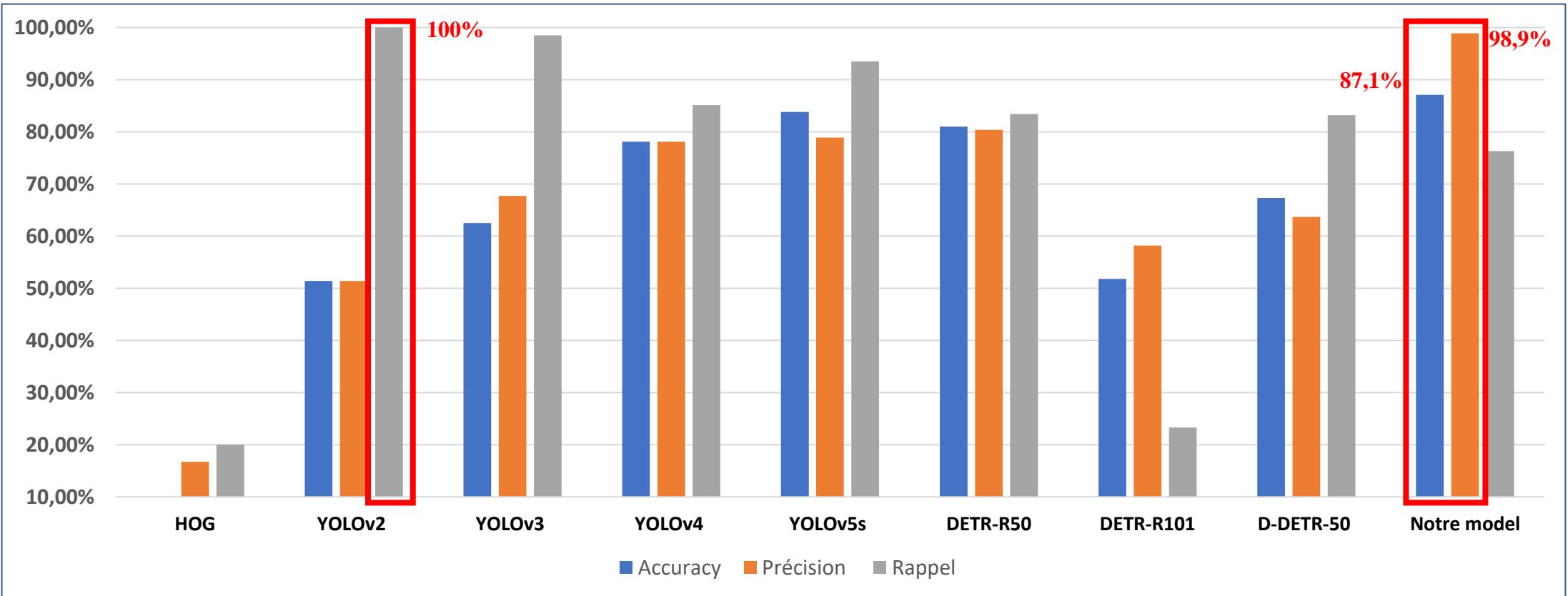


Efficiency

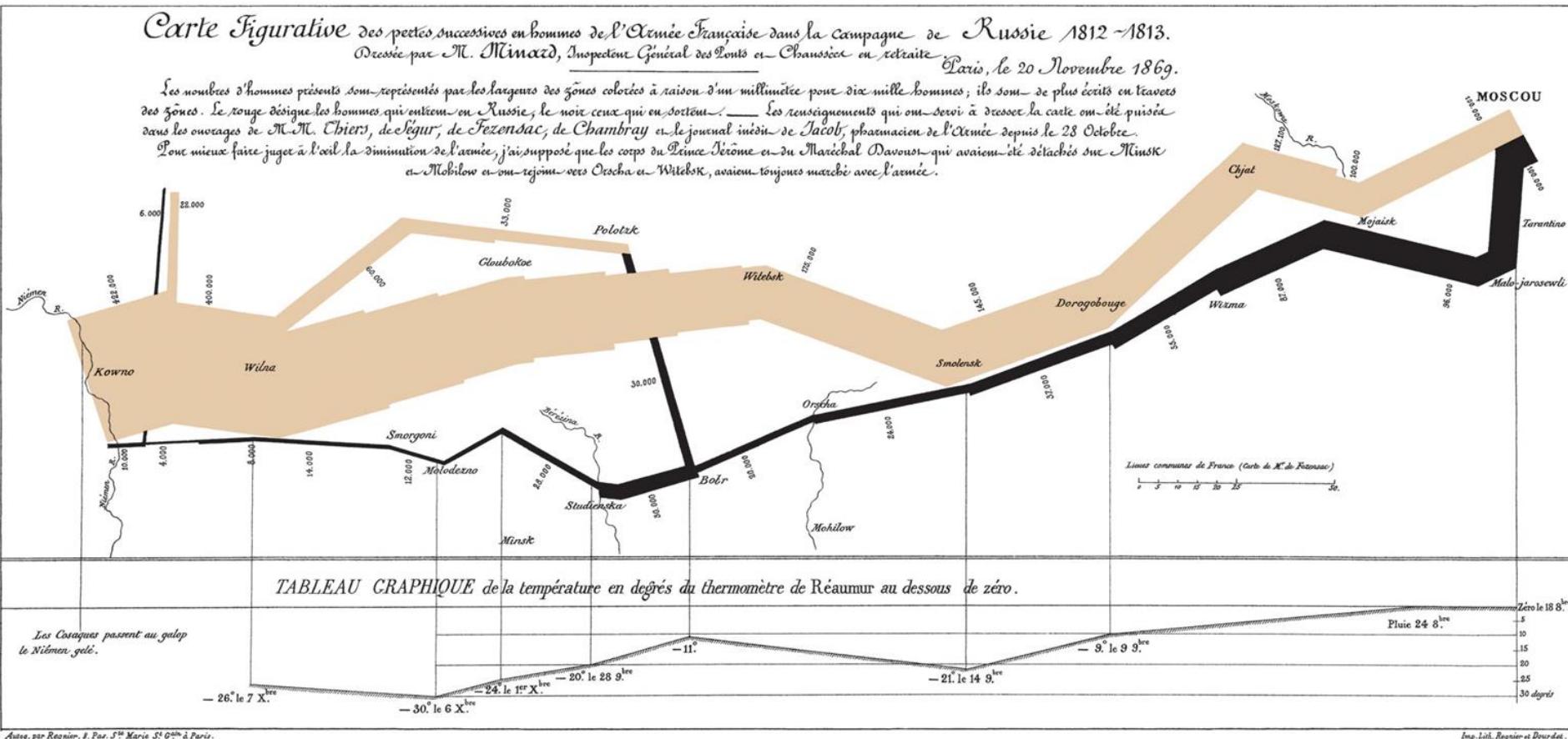
Use efficient visualization technique which
highlights all the data points

ruiwen.he@devinci.fr

Highlights important information

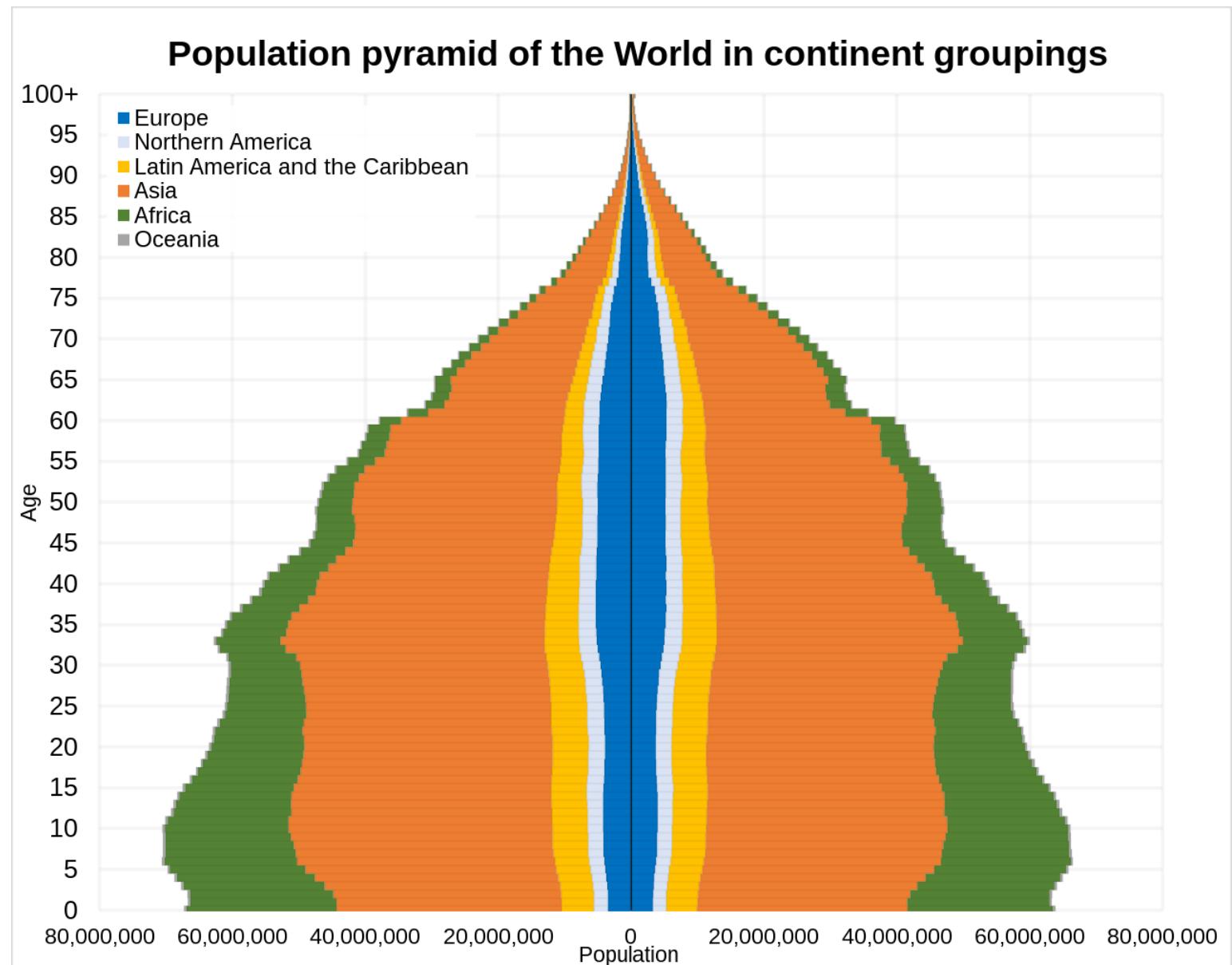


Example of visualization



Joseph Minard 1869 : Perte Napoléonienne de la campagne de Russie (diagramme de Sankey) 

Example of visualization



https://fr.wikipedia.org/wiki/Population_mondiale#:~:text=L'ONU%20l'estime%20%C3%A0%20millions%20d'habitants%20pour%201700.
ruwen.mc@devmail.fr

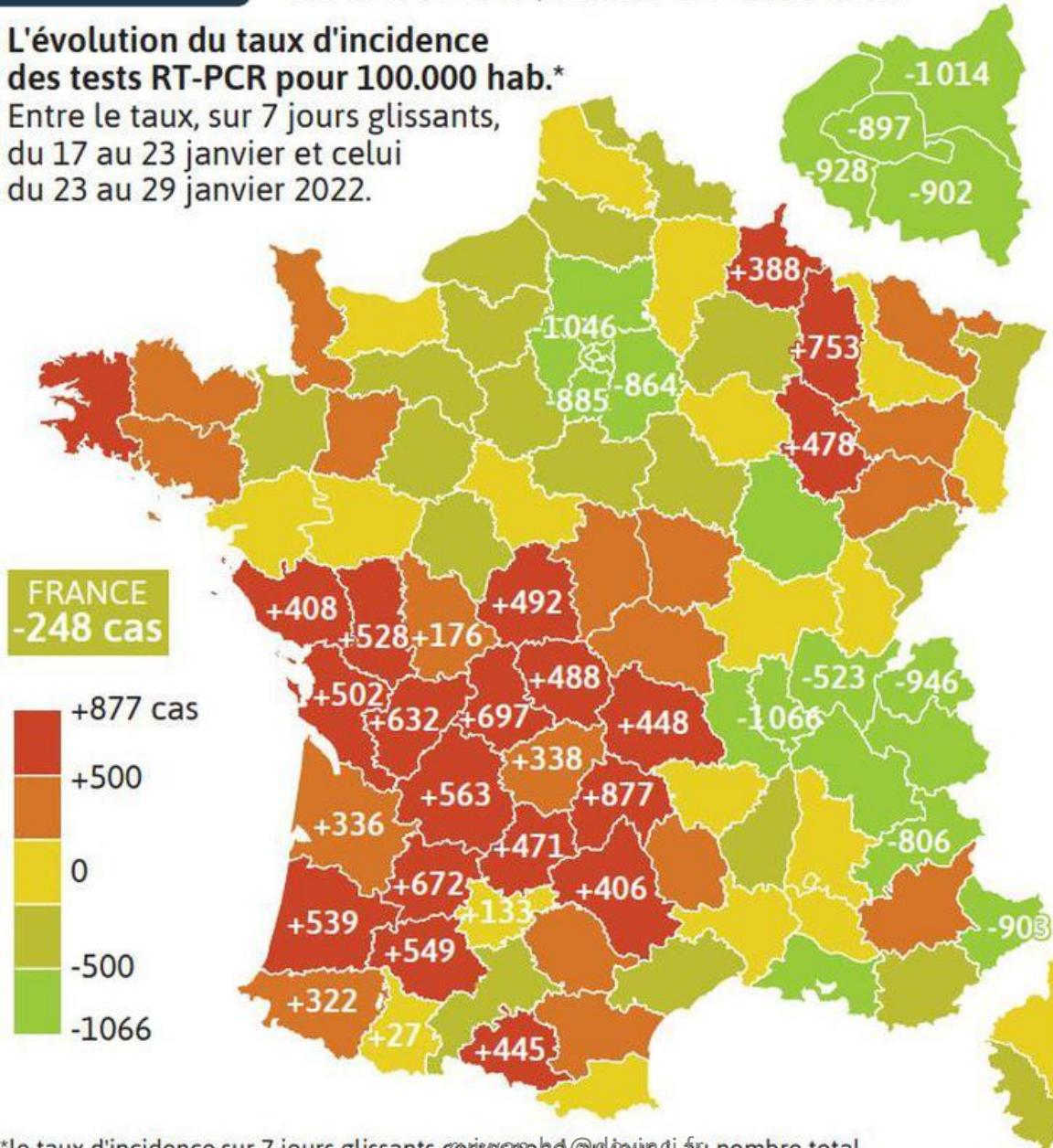
Example of visualization

COVID-19

TAUX D'INCIDENCE: LES DÉPARTEMENTS
OÙ IL MONTE, CEUX OÙ IL BAISSE

L'évolution du taux d'incidence
des tests RT-PCR pour 100.000 hab.*

Entre le taux, sur 7 jours glissants,
du 17 au 23 janvier et celui
du 23 au 29 janvier 2022.



2
fév.

Levée de certaines
restrictions sanitaires



Suppression des jauge
dans les établissements où le public
est assis et porte un masque.



Fin du port du masque
en extérieur.



Allègement
du télétravail.

16
fév.

Levée de certaines
restrictions sanitaires



Consommations autorisées
dans les stades, les cinémas
et les transports.



Consommations autorisées
debout dans les bars.



Reprise des
concerts debout.

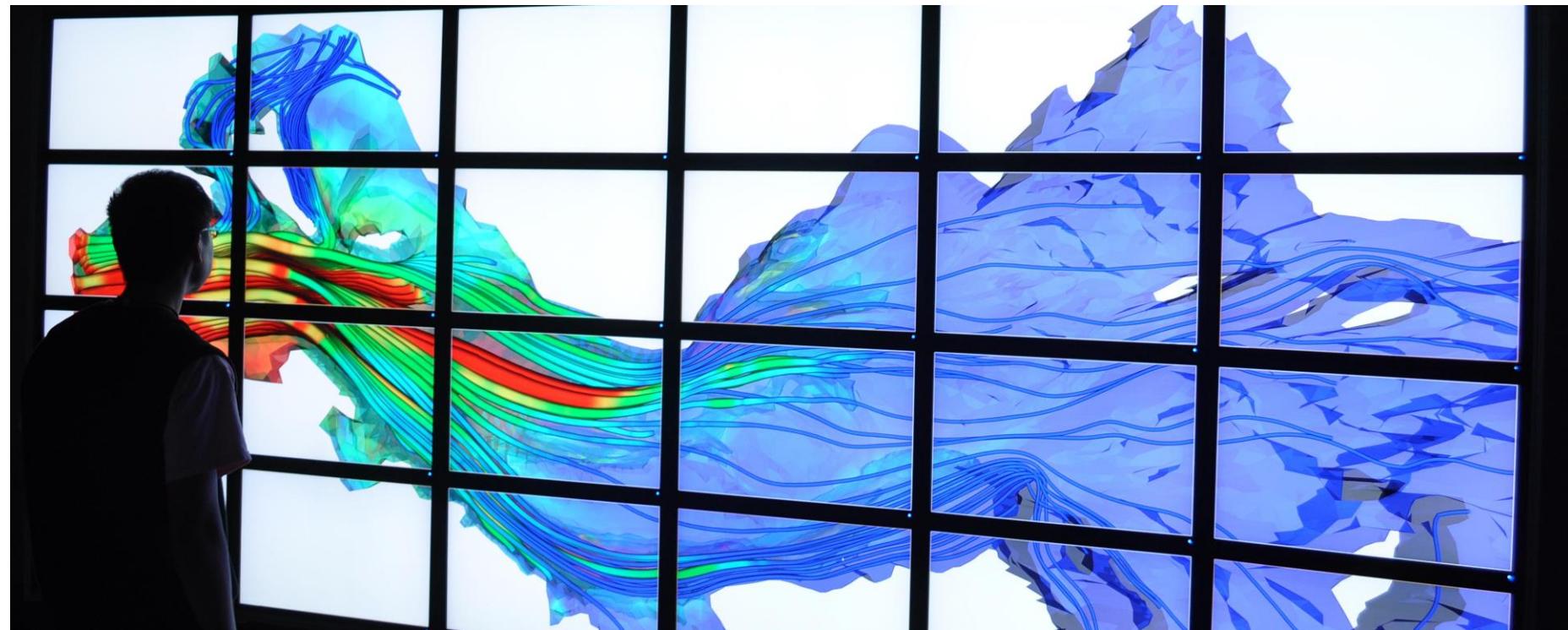


Réouverture
des discothèques.

Example of visualization

34

A river is simulated based on sensor data. The various colors show water speed in different parts of the river.

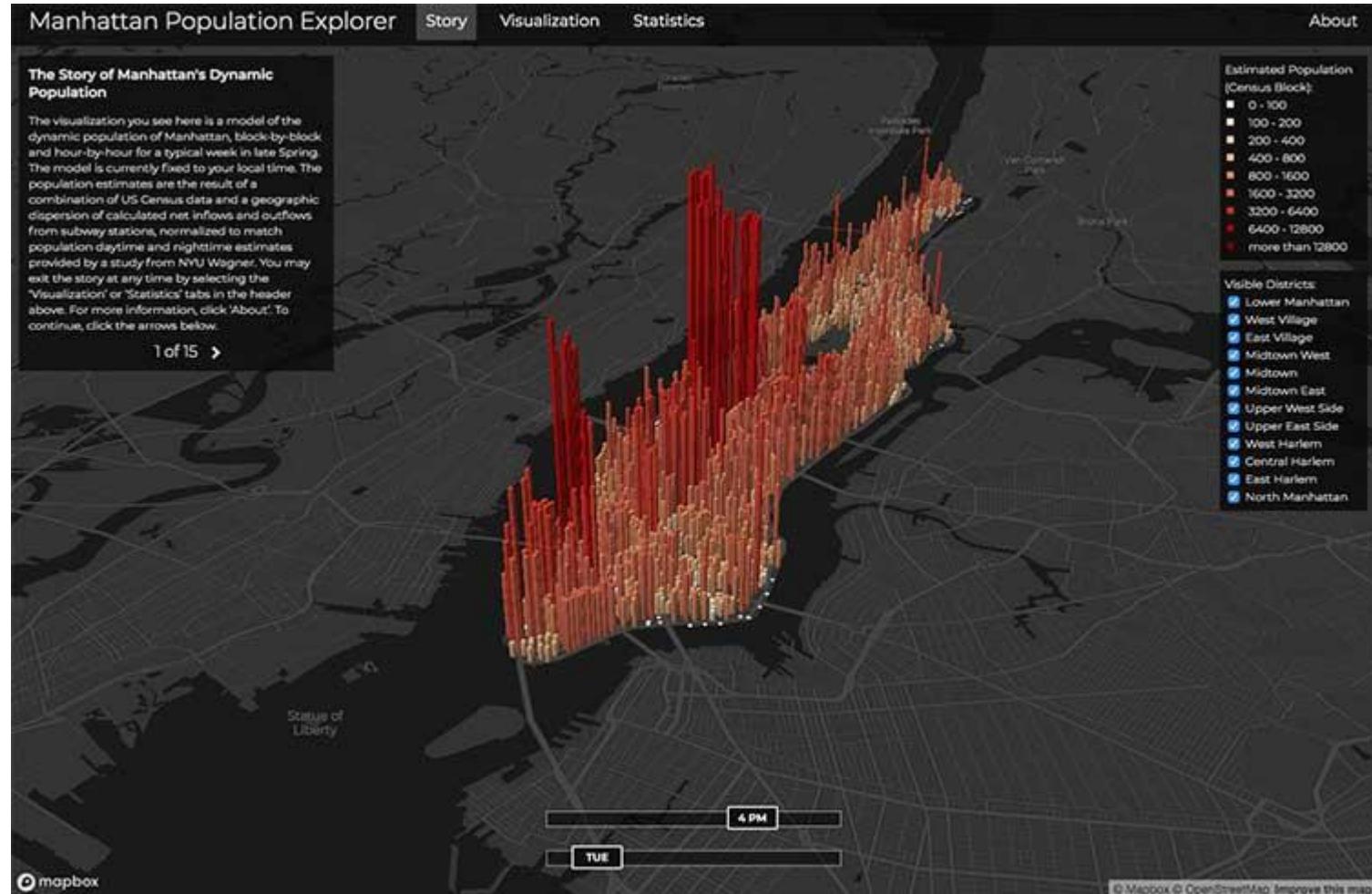


https://www.nsf.gov/discoveries/disc_images.jsp?cntn_id=114322&org=NSF

ruiwen.he@devinci.fr

Example of visualization

The invisible heartbeat of New York City



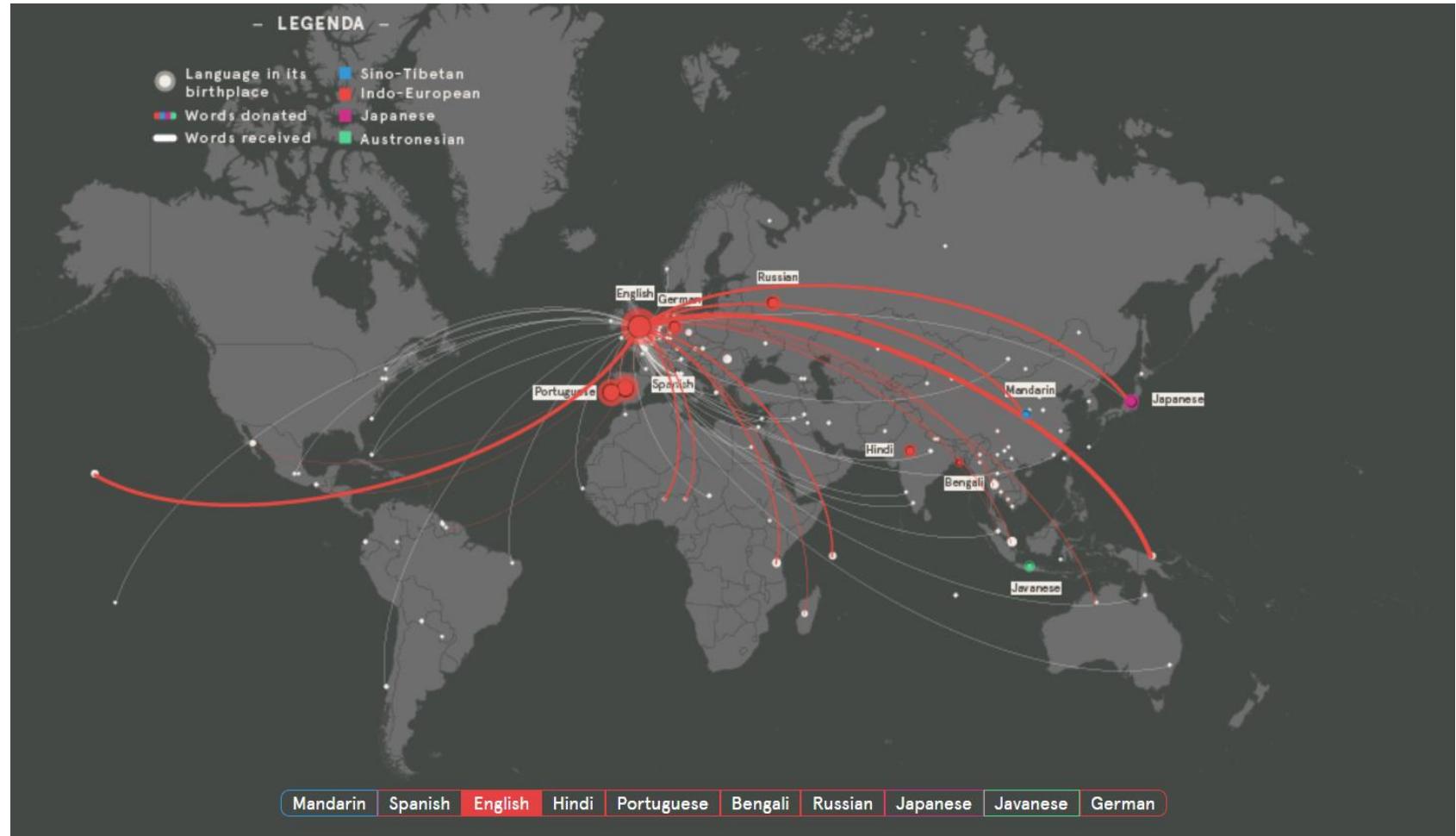
<http://manpopex.us/>

dynamic population

ruiwen.he@devinci.fr

Languages in the world

Example of visualization



<http://www.puffpuffproject.com/languages.html>

Example of visualization

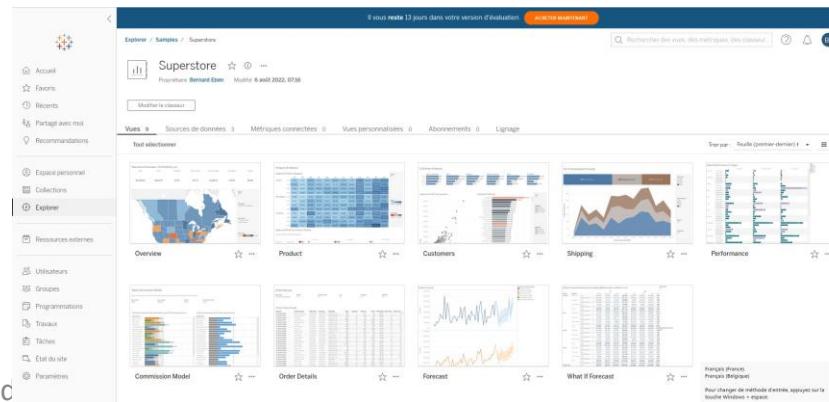
- 200 years on 4 minutes with la BBC
 - <https://youtu.be/Z8t4k0Q8e8Y>

Data Visualization Tools Overview

- **Tableau**
- **Features:** Interactive dashboards, drag-and-drop interface, supports various data sources.
- **Use Cases:** Business reporting, sales performance tracking, financial analysis.
- **Strengths:** User-friendly, powerful visualizations, large community support.



ruiwen.he@

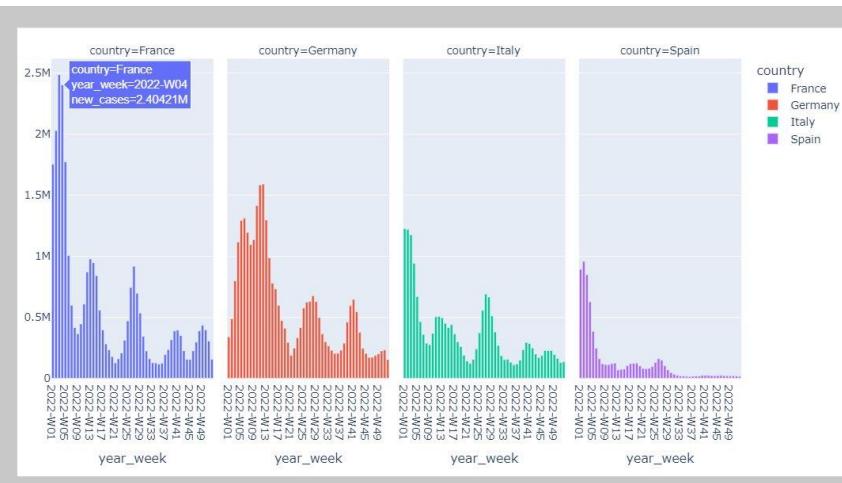


Data Visualization Tools Overview

- Plotly
- Features: Python-based, interactive visualizations, web-based integration.
- Use Cases: Data analysis, scientific charts, academic research.
- Strengths: Advanced interactivity, integrates with Python/R, customizable.



plotly



Data Visualization Tools Overview

- **Power BI**
- **Features:** Integrates seamlessly with Microsoft products, real-time dashboards, cloud sharing.
- **Use Cases:** Business analytics, KPI tracking, ad-hoc reporting.
- **Strengths:** Affordable, robust, excellent for Microsoft-based ecosystems.



Power BI

40



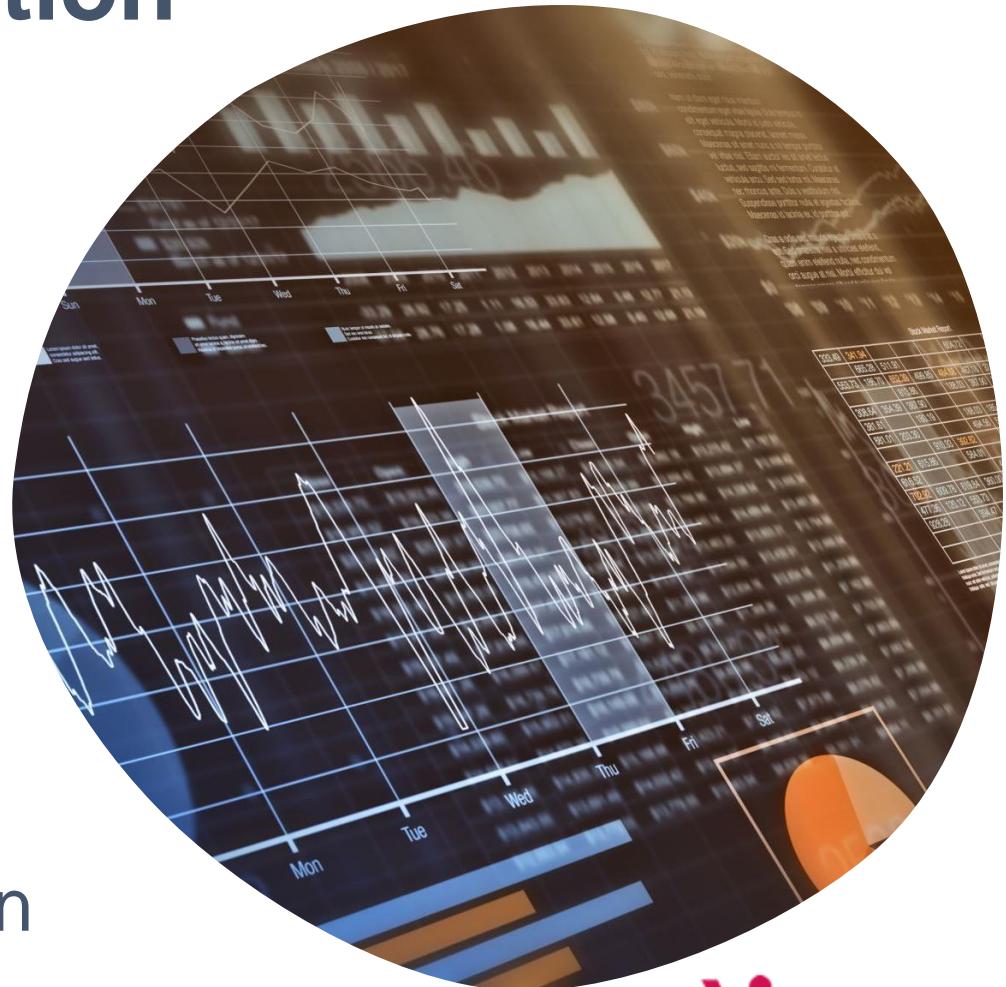
ruiv

Jun 25, 2019 09:49

V
ESILV

Introduction to data visualization

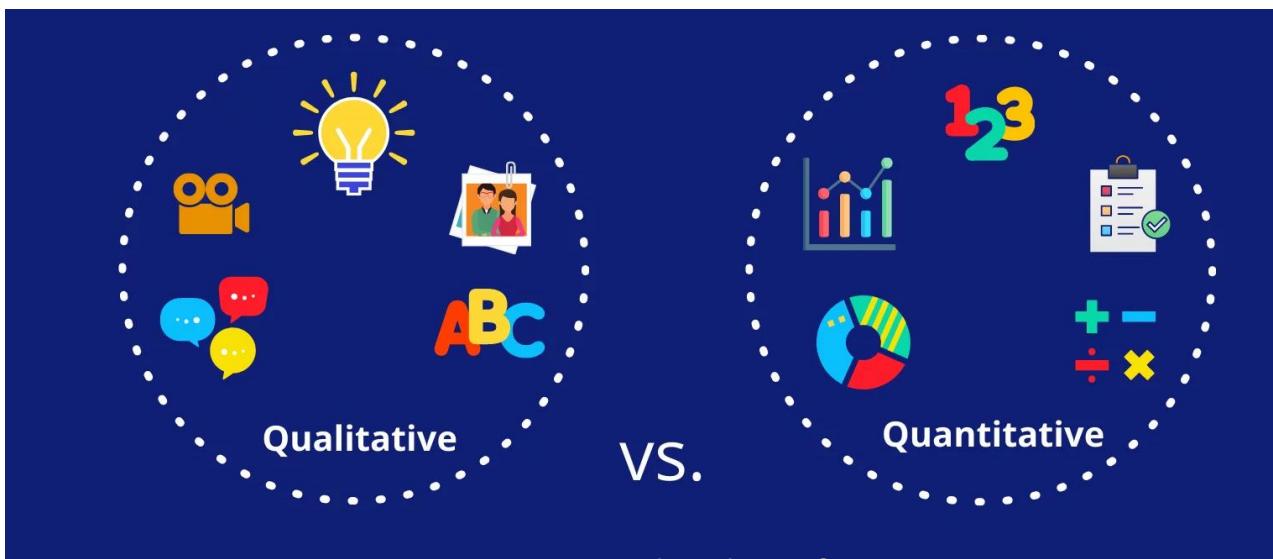
1. What is Business Intelligence (BI)?
2. What is data visualization?
3. Why visualization is important?
4. **Data types and levels of measurement**
5. Graphical Basics
6. Choosing the right type of visualization



Data Types

Qualitative (or categorical) data consist of values that can be placed into nonnumerical categories.

Quantitative data consist of values representing counts or measurements.



Exercise 1 Data Types

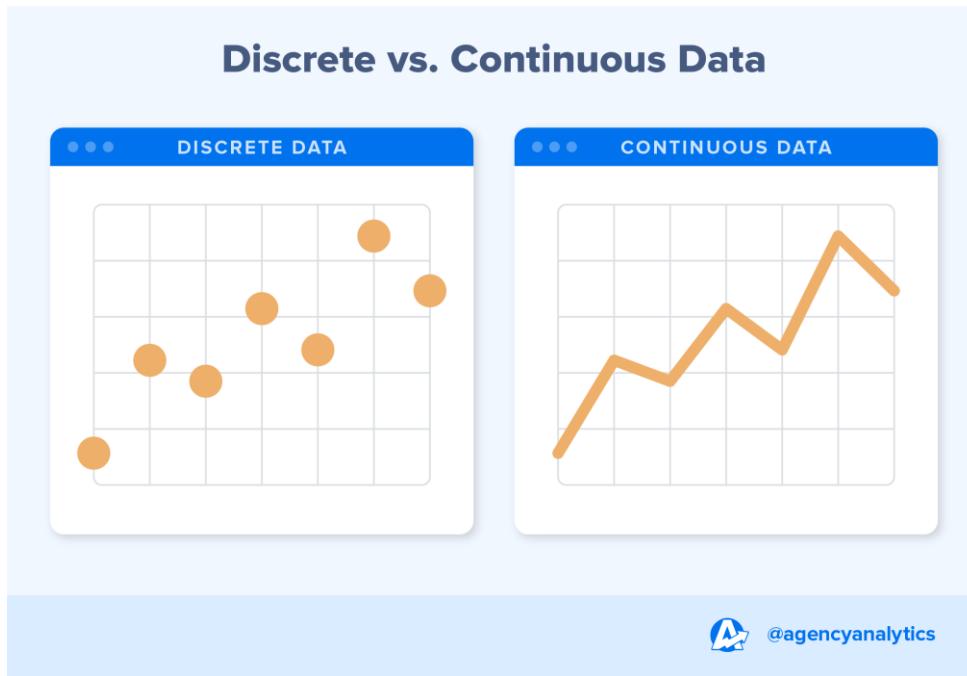
Classify each of the following sets of data as qualitative or quantitative.

- The breeds of dog: Husky Golden labrador
- The standard weight of Husky

Discrete versus Continuous Data

Continuous data can take on *any* value in a given interval.

Discrete data can take on only particular, distinct values and not other values in between.



Exercise 2 Discrete or Continuous?

For each data set, indicate whether they data are discrete or continuous.

- The speed of train from Paris to London
- The score of examens (such as 11, 17, 18)

Qualitative data

- Descriptive and non-numerical data that explains the ‘why’ behind a phenomenon
- Aim: To explore a phenomenon

Qualitative data

- Descriptive and non-numerical data that explains the ‘why’ behind a phenomenon
- Qualitative data can be of two levels of measurement:
 - Nominal
 - Ordinal

Nominal

- Nominal data is “labeled” or “named” data which can be divided into various groups that do not overlap.
- Data is not measured or evaluated in this case; it is just assigned to multiple groups.
- These groups are unique and have no common elements

Nominal data examples



Ordinal

- is classified into categories within a variable that have a natural rank order.
- However, the distances between the categories are uneven or unknown.

Ordinal data examples



Quantitative data

- Interval
 - is defined as a data type which is measured along a scale, in which each point is placed at equal distance from one another.
 - Zero has no sense
- Ratio

Interval

Please state your annual income

- Below \$40,000
- \$40,000- \$60,000
- \$60,000- \$80,000
- \$80,000- \$100,000
- Above \$100,000

Quantitative data

- Interval
 - is defined as a data type which is measured along a scale, in which each point is placed at equal distance from one another.
 - Zero has no sense
- Ratio
 - Zero has sense(nothing)
 - Physical measurement: weight, length, etc.

Rapport



<https://careerfoundry.com/en/blog/data-analytics/what-is-ratio-data/>

ruiwen.he@devinci.fr

Data Type

Qualitative

Quantitative

Discrete

Continuous

Level of Measurement

Nominal

Ordinal

Interval

Ratio

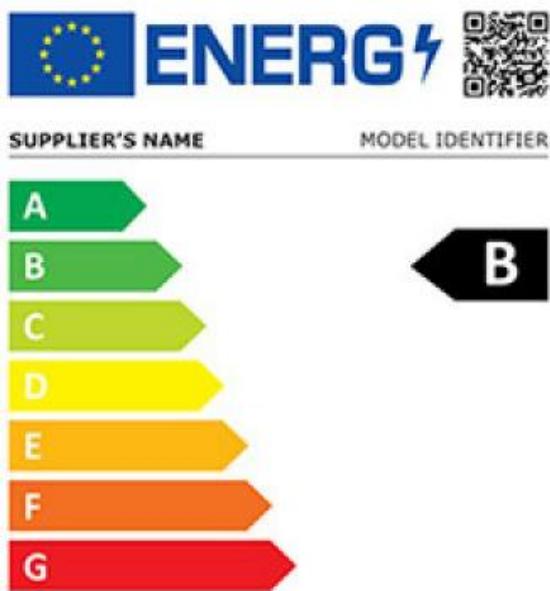
Interval

Ratio

Exercise 3 Levels of measurement

Identify the level of measurement (nominal, ordinal, interval, ratio) for each of the following sets of data.

- Energy Class of fridge



ordinal

Exercise 3 Levels of measurement

Identify the level of measurement (nominal, ordinal, interval, ratio) for each of the following sets of data.

- The standard weight of each dog breed

 **Dog size chart**

Small	Medium	Large	Giant
< than 24 lb	25 - 59 lb	60 - 99 lb	> 100 lb
< than 10 kg	11 - 26 kg	27 - 45 kg	> 45 kg
Chihuahua, Yorkshire Terrier, French Bulldog, Toy Poodle	Bulldog, Australian Shepherd, Border Collie	Golden Retriever, Rottweiler	Great Dane, Mastiff
			

interval

Exercise 3 Levels of measurement

Identify the level of measurement (nominal, ordinal, interval, ratio) for each of the following sets of data.

- score of examen and level (niveau)



Ratio and ordinal

Exercise 3 Levels of measurement

Identify the level of measurement (nominal, ordinal, interval, ratio) for each of the following sets of data.

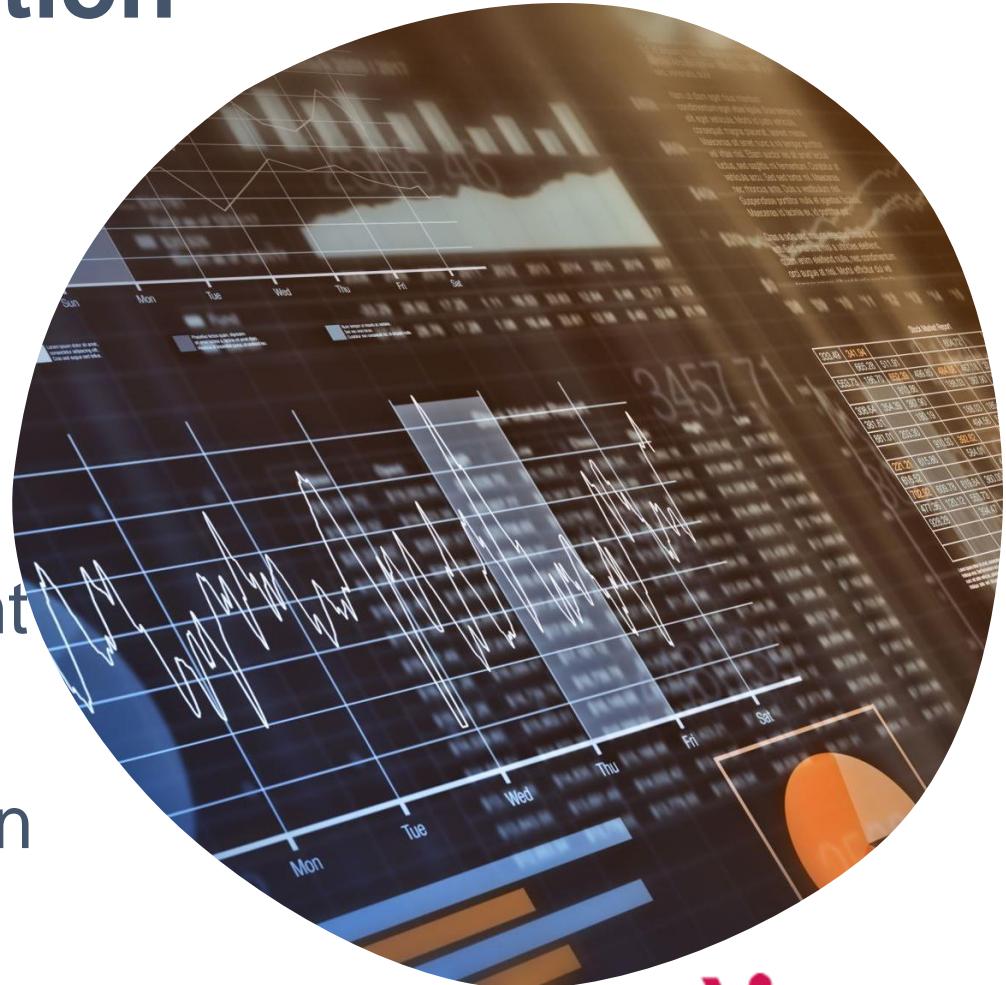
- the name of the football team



nominal

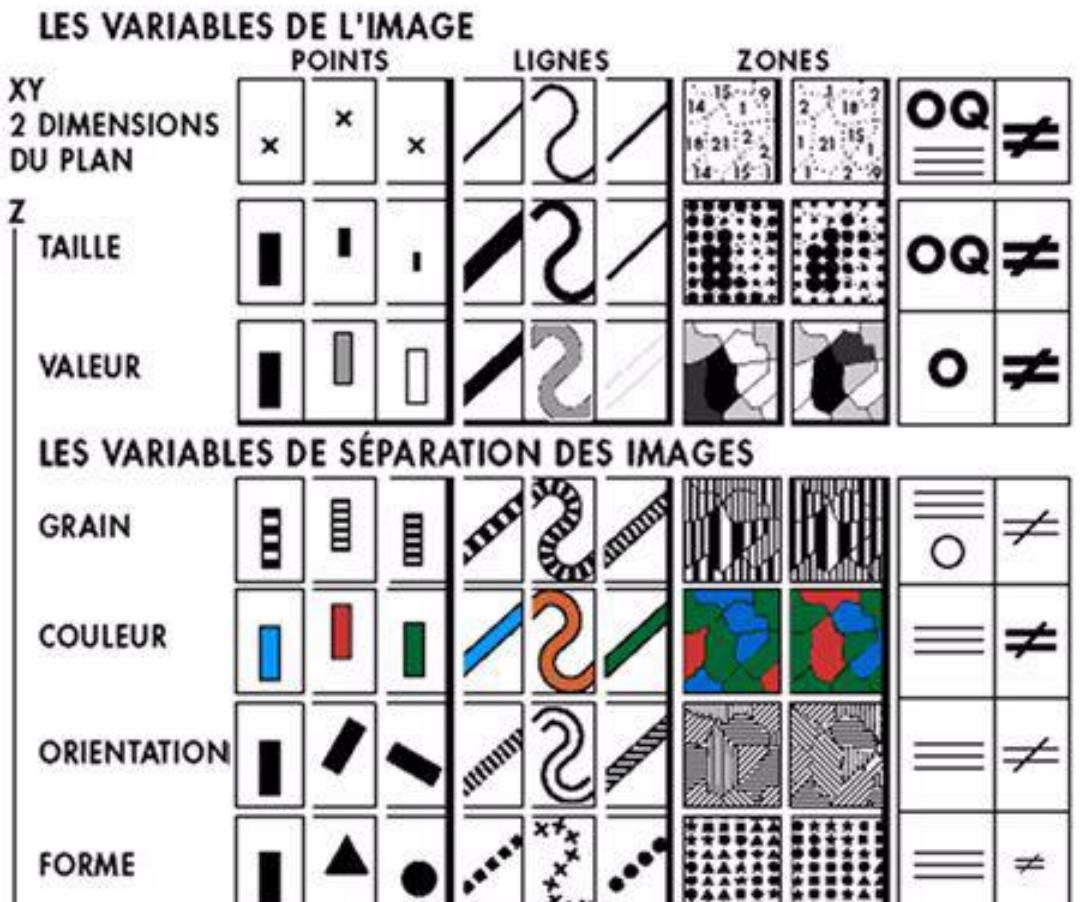
Introduction to data visualization

1. What is Business Intelligence (BI)?
2. What is data visualization?
3. Why visualization is important?
4. Data types and levels of measurement
5. **Graphical Basics**
6. Choosing the right type of visualization



Graphical Basics

- Jacques Bertin's variables
- French cartographer, author of graphic semiology



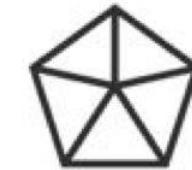
→ Points



→ Lines



→ Areas



Single brands

- Munzner, 2014,
- Visualization Analysis and Design.

Visual channels

Munzner, 2014,
Visualization Analysis and Design.

④ Position

→ Horizontal



→ Vertical



→ Both



④ Color



④ Shape



④ Tilt



④ Size

→ Length



→ Area

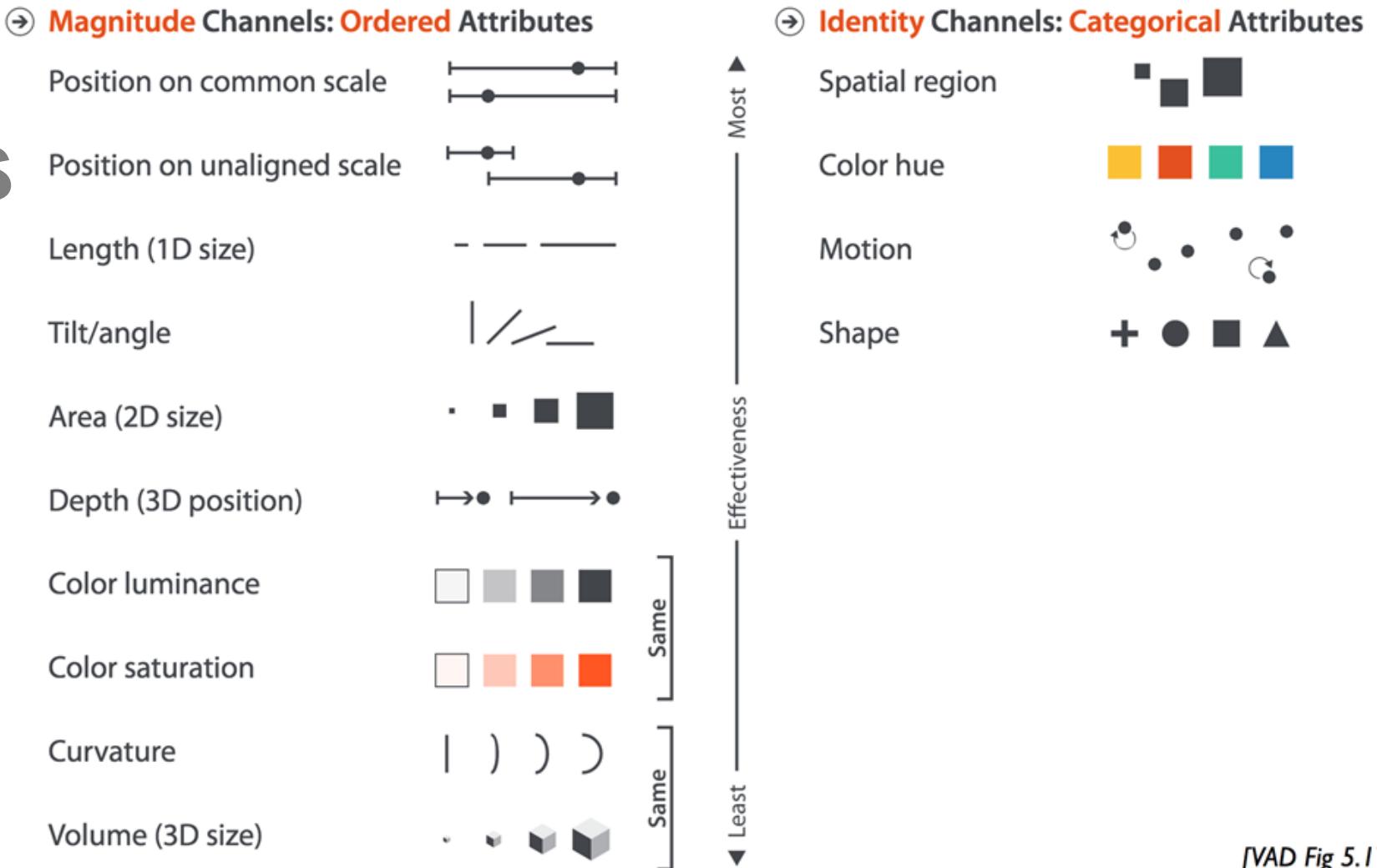


→ Volume



Channel effectiveness

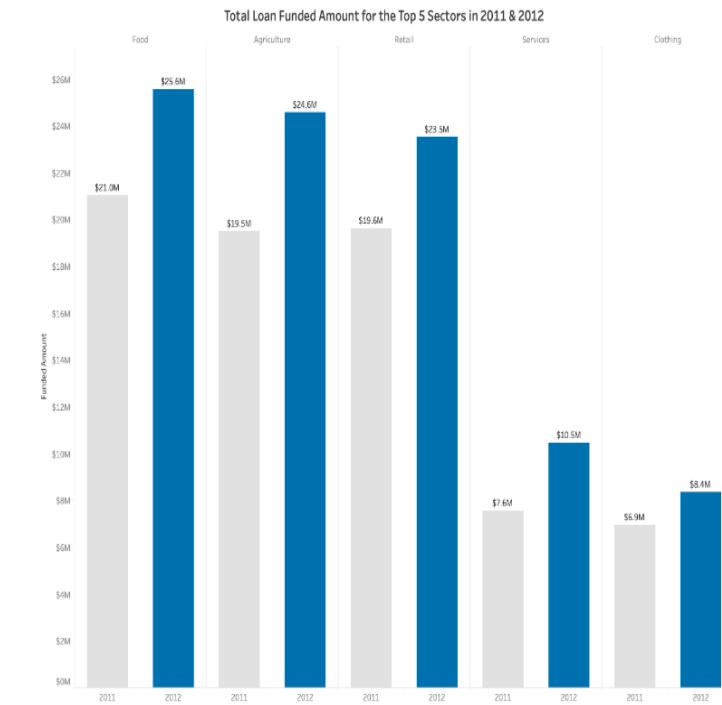
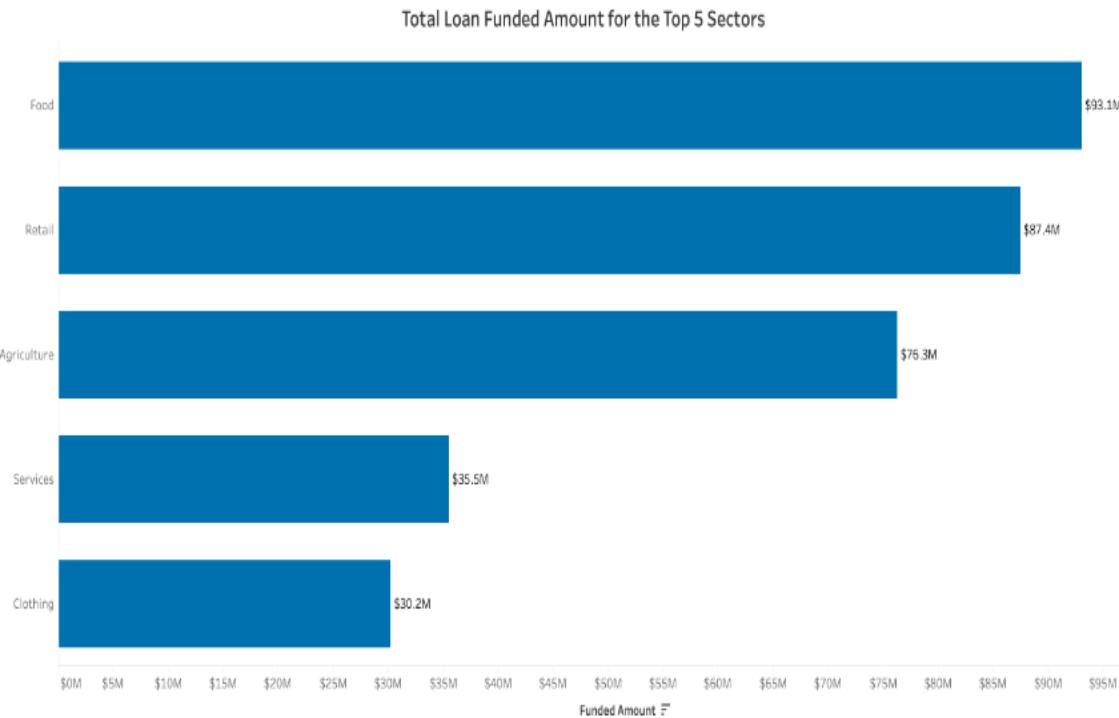
- Munzner, 2014,
Visualization Analysis
and Design.



[VAD Fig 5.1]

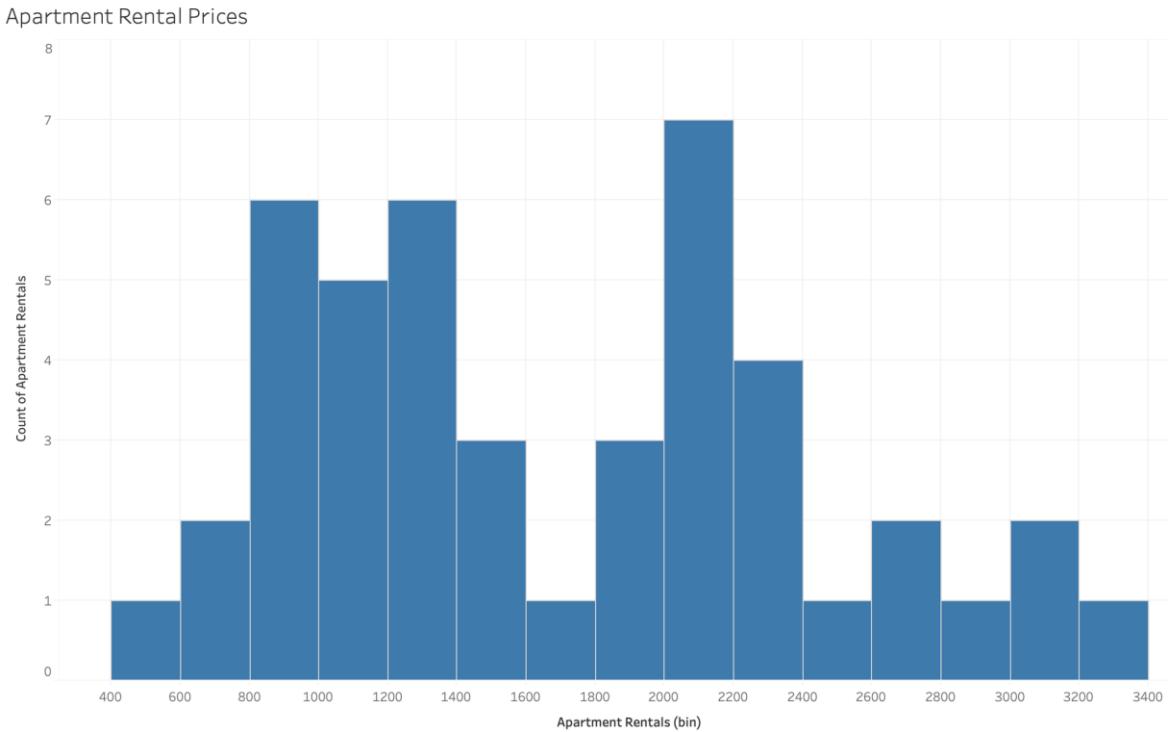
Bar Charts

- Show the data changes of multiple categories and the comparison between variables of the same category .



Histogram

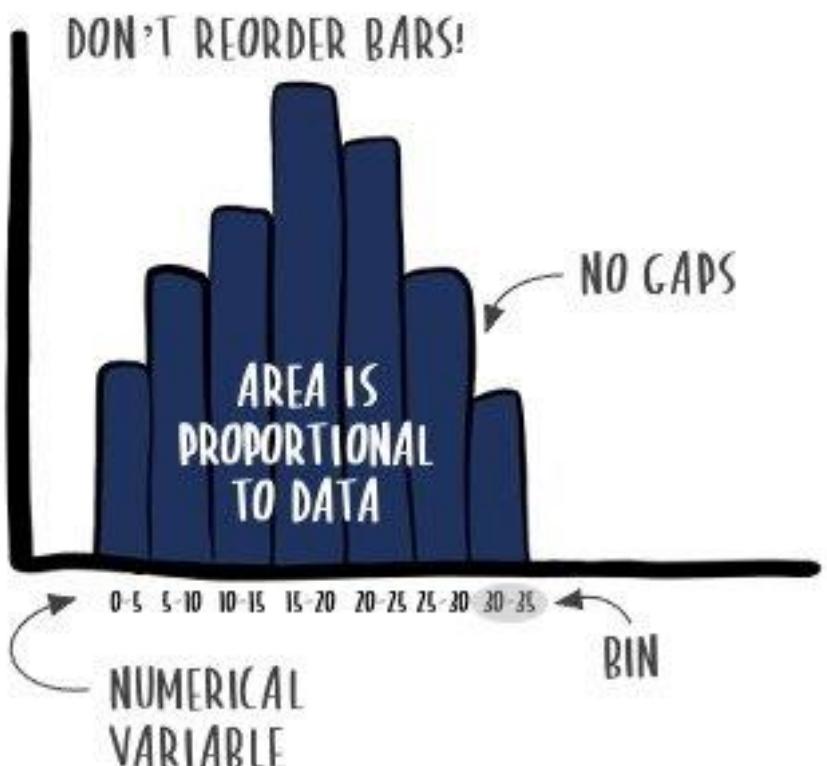
- A histogram is a graph used to represent the frequency distribution of a few data points of one variable.



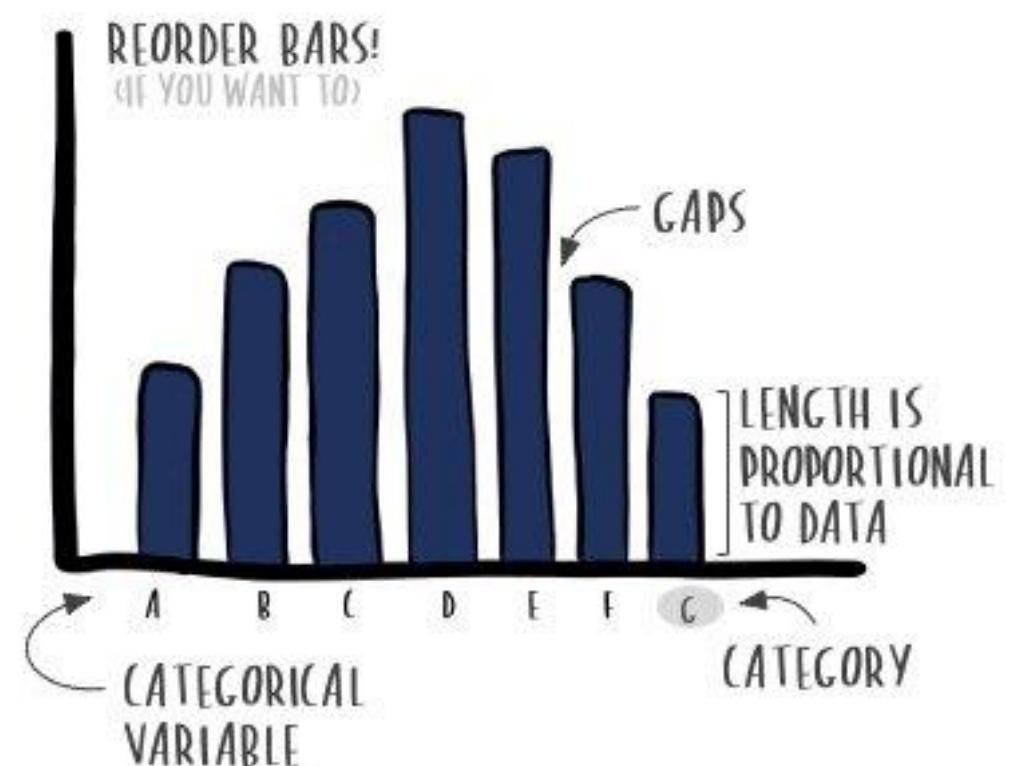
<https://www.tableau.com/data-insights/reference-library/visual-analytics/charts>

What is the difference between histogram and bar graph

This is a **histogram**...



This is a **bar chart**...

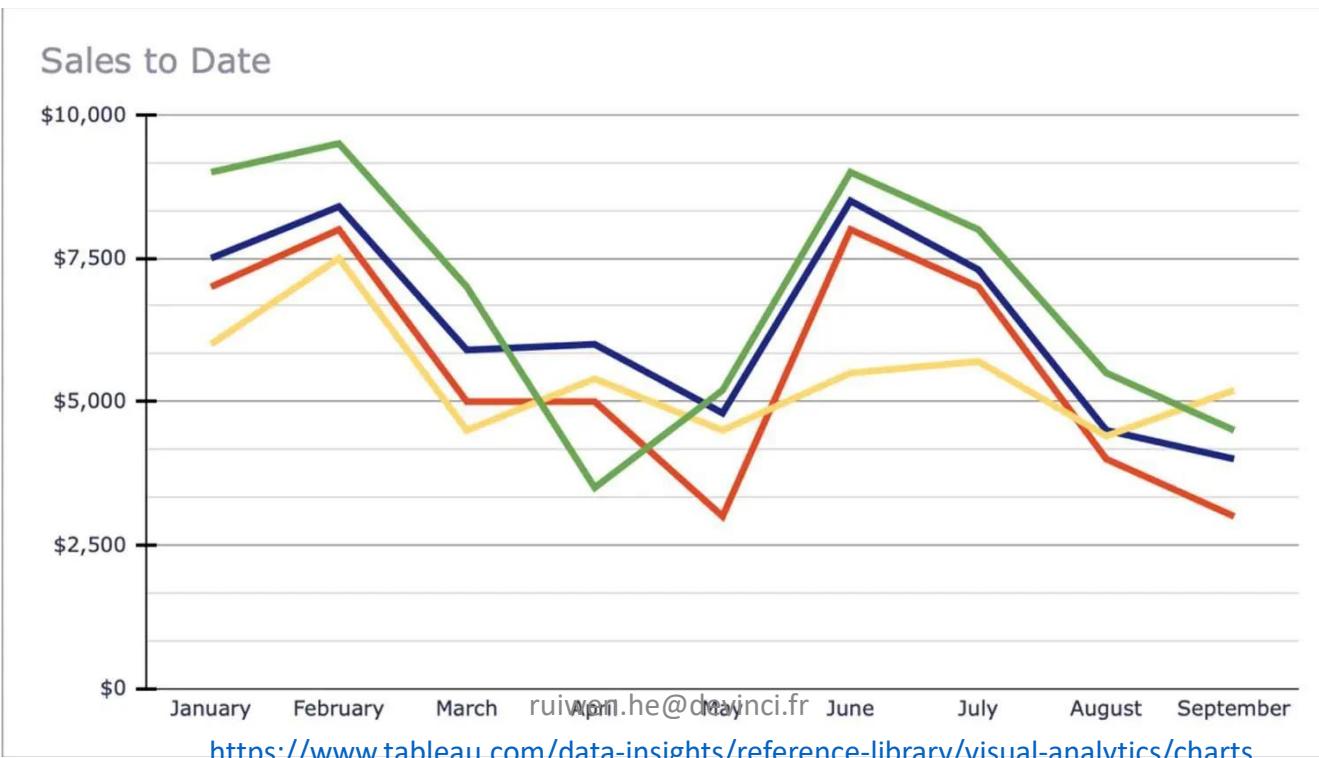


<https://www.storytellingwithdata.com/blog/2021/1/28/histograms-and-bar-charts>

ruiwen.he@devinci.fr

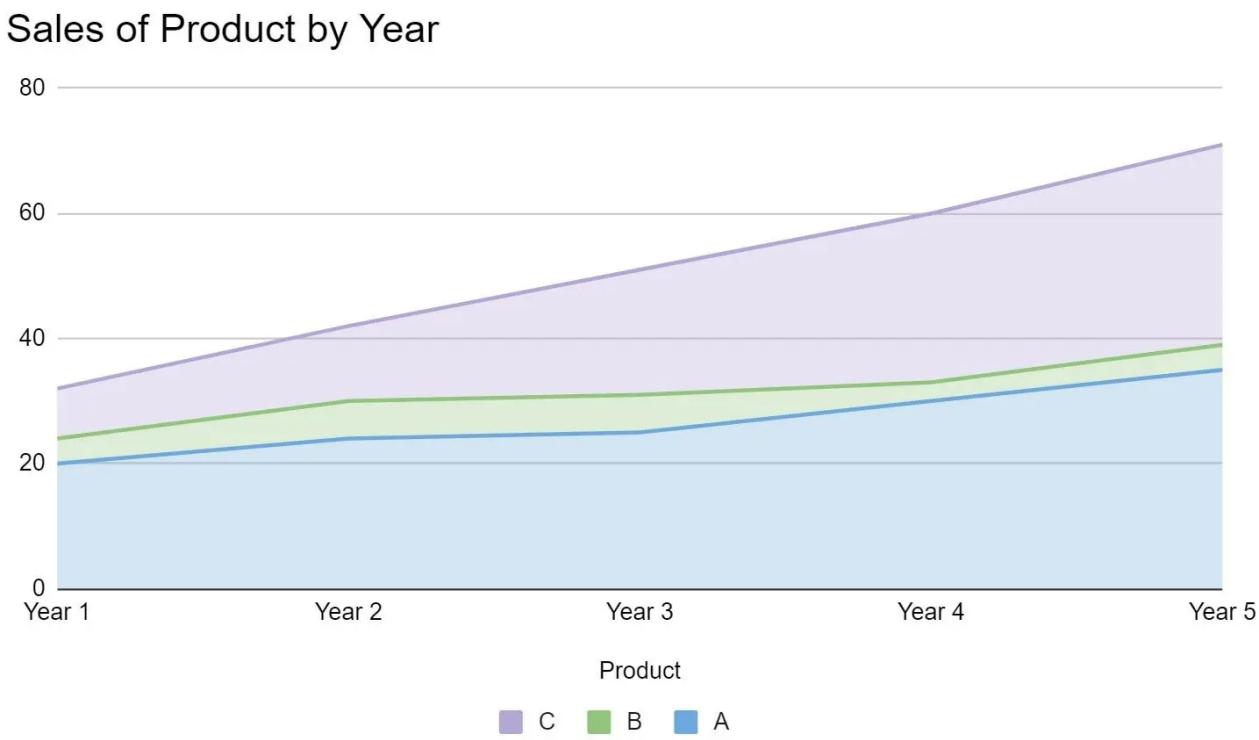
Line Graph

- It can display the trend of data changes and reflect the changes in things.
- Applicable: Ordered categories, such as time.



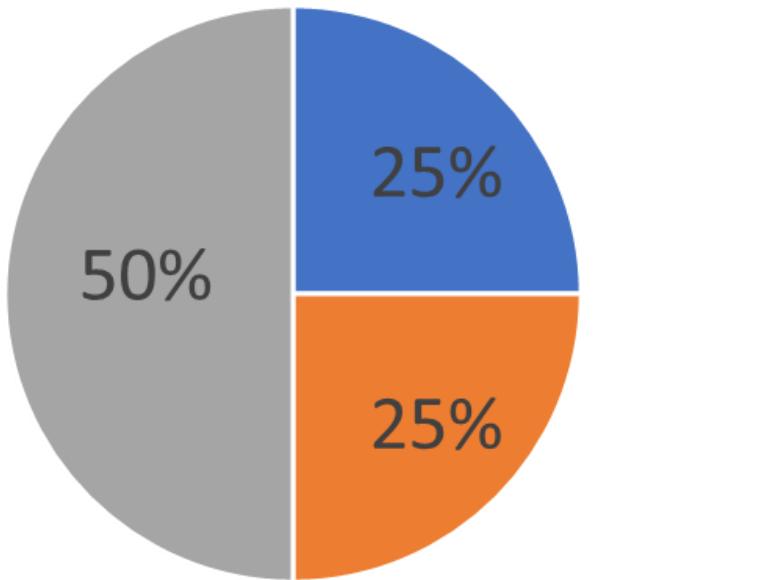
Area Chart

- Not only can it clearly reflect the trend changes of the data, but also can emphasize the gap comparison between different types of data



Pie Chart

- Very useful for highlighting scale.
- Show the relationship between the part and the whole.

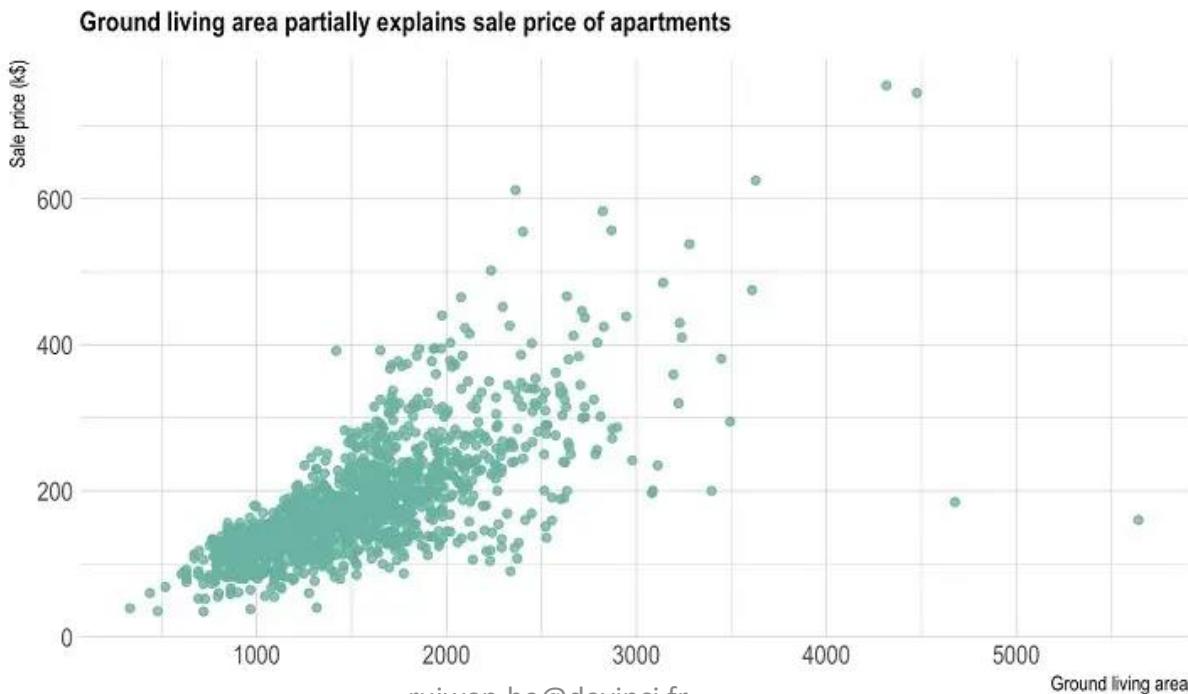


■ Attendance ■ TDs ■ Examen

ruiwen.he@devinci.fr

Scatter Chart

- Scatter charts use data values as x, and y coordinates to draw points. It can reveal relationships between values plotted on the grid and also show trends in the data.



Dashboard Chart

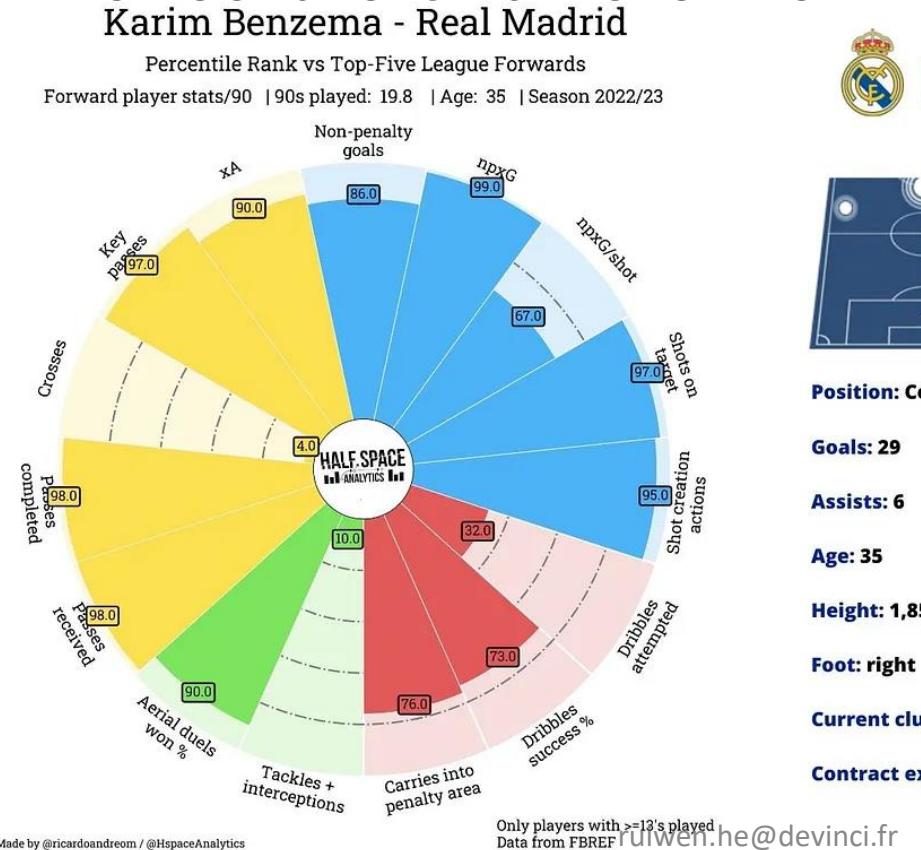
- Intuitively show the progress and completion of a certain indicator, mainly used for the display of progress or proportion.



<https://www.tableau.com/data-insights/reference-library/visual-analytics/charts>

Radar Graph

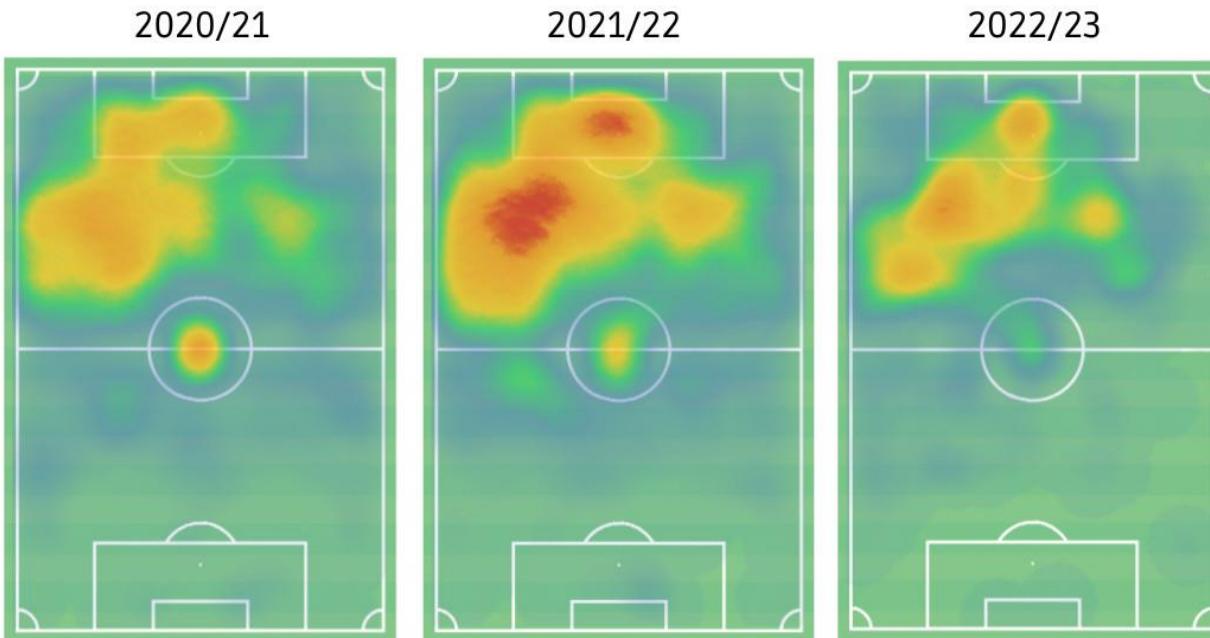
- Map the data volume of multiple classes to the coordinate axis and compare the feature of different attributes of an item.



Heatmap

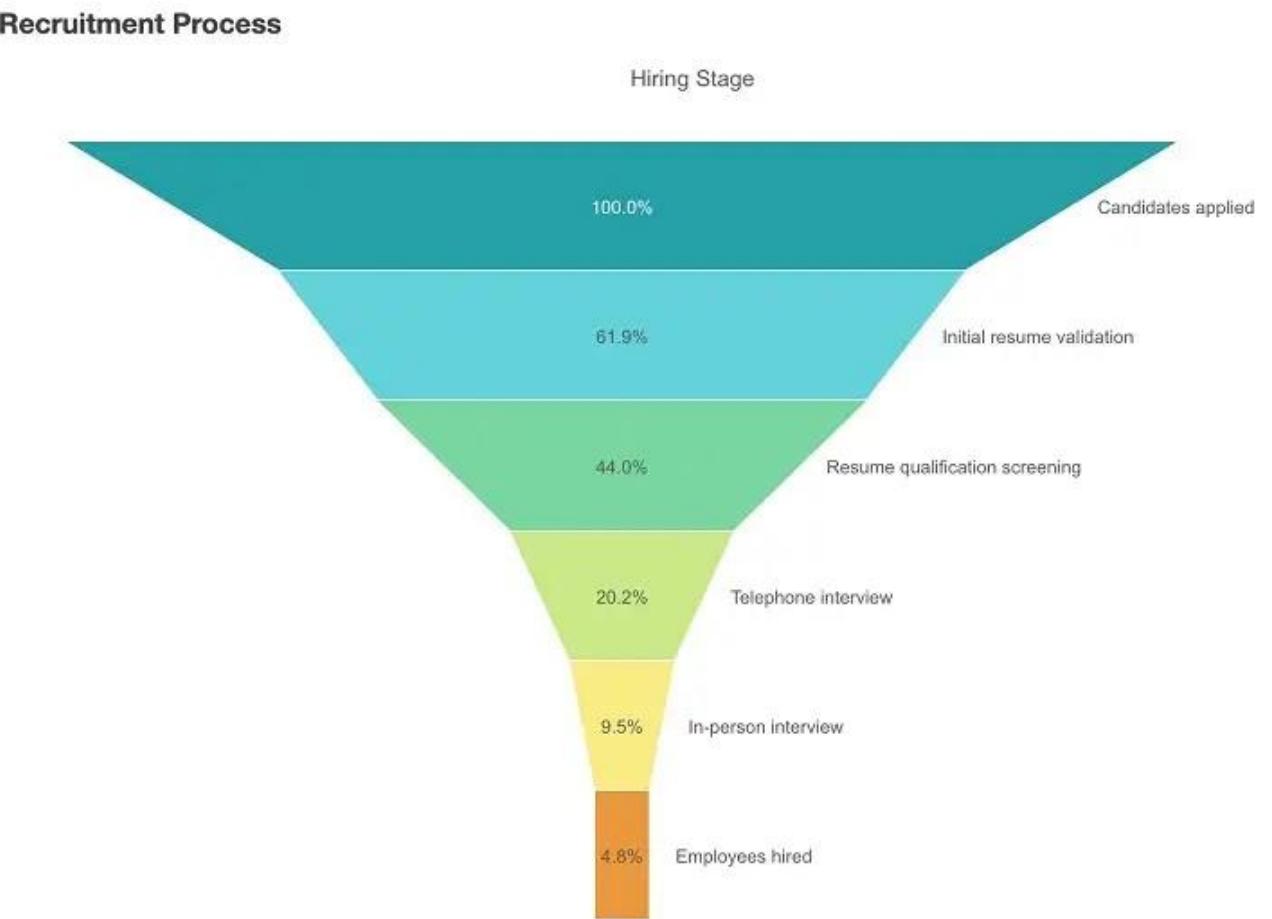
- The heatmap mainly expresses the magnitude of the value (two dimensions) through color and is generally used to reflect the degree of activity.

Karim Benzema's heat map comparison



Funnel Graph

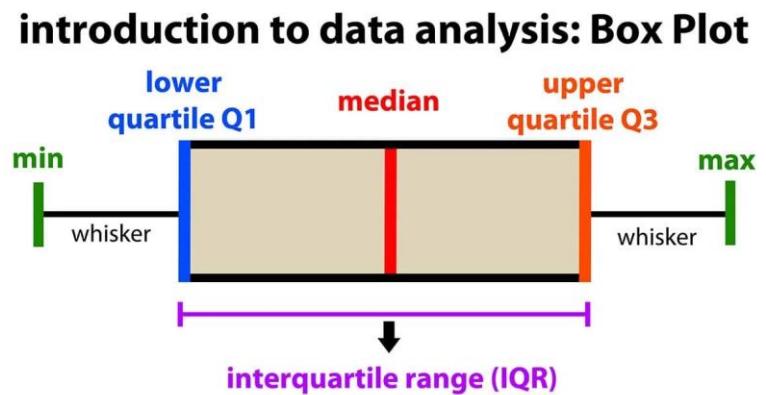
- A funnel graph is used for single-pass analysis. It consists of N process links between the start and the end.



<https://www.tableau.com/data-insights/reference-library/visual-analytics/charts>

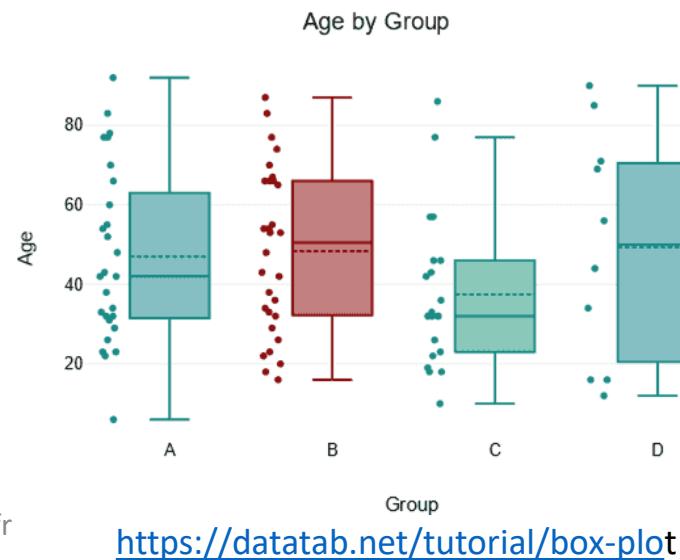
Box plot

- It is a method of describing data using five statistics in the data: minimum value, first quartile, median, third quartile and maximum value.
- Applicable: Used to show the dispersion of a set of data, especially for comparison of several samples.



77

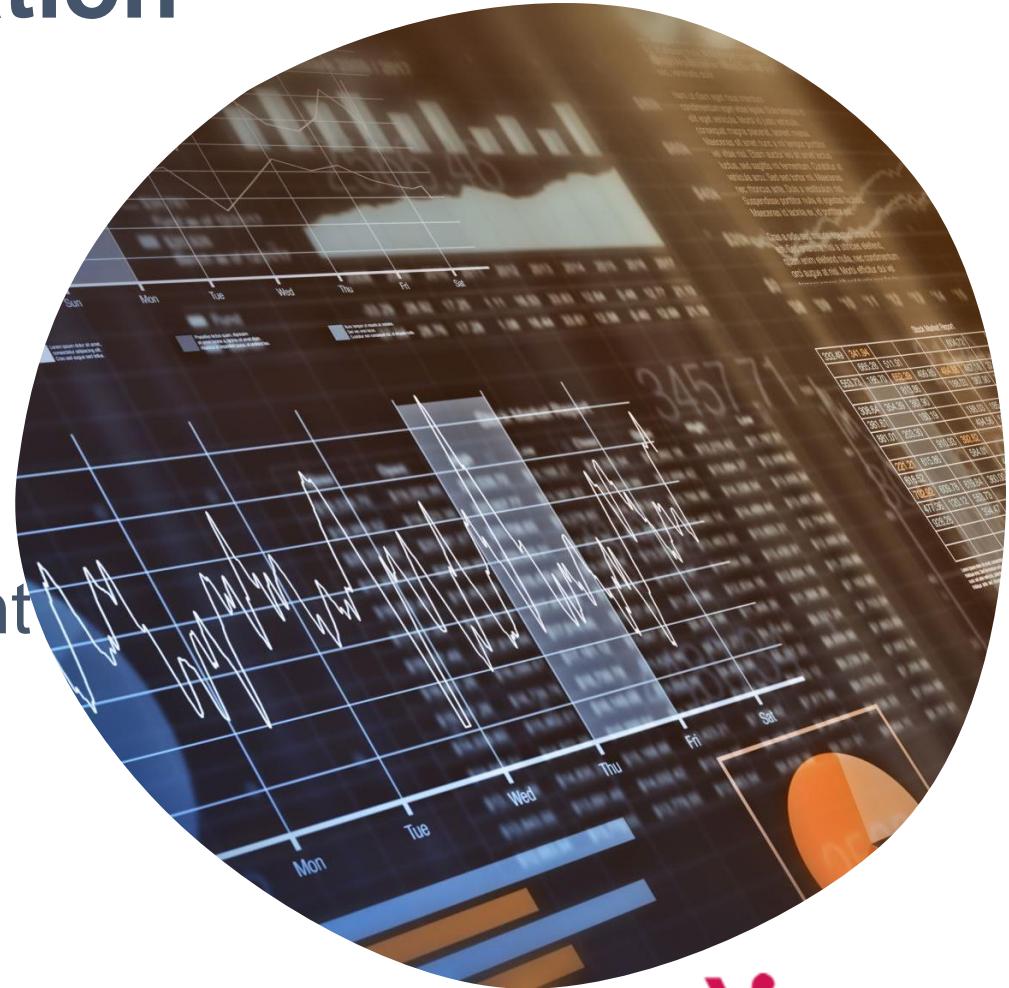
<https://www.simplypsychology.org/boxplots.html> cuiwen.he@devinci.fr



<https://databab.net/tutorial/box-plot>

Introduction to data visualization

1. What is Business Intelligence (BI)?
2. What is data visualization?
3. Why visualization is important?
4. Data types and levels of measurement
5. Graphical Basics
6. **Choosing the right type of visualization**



When good data goes bad

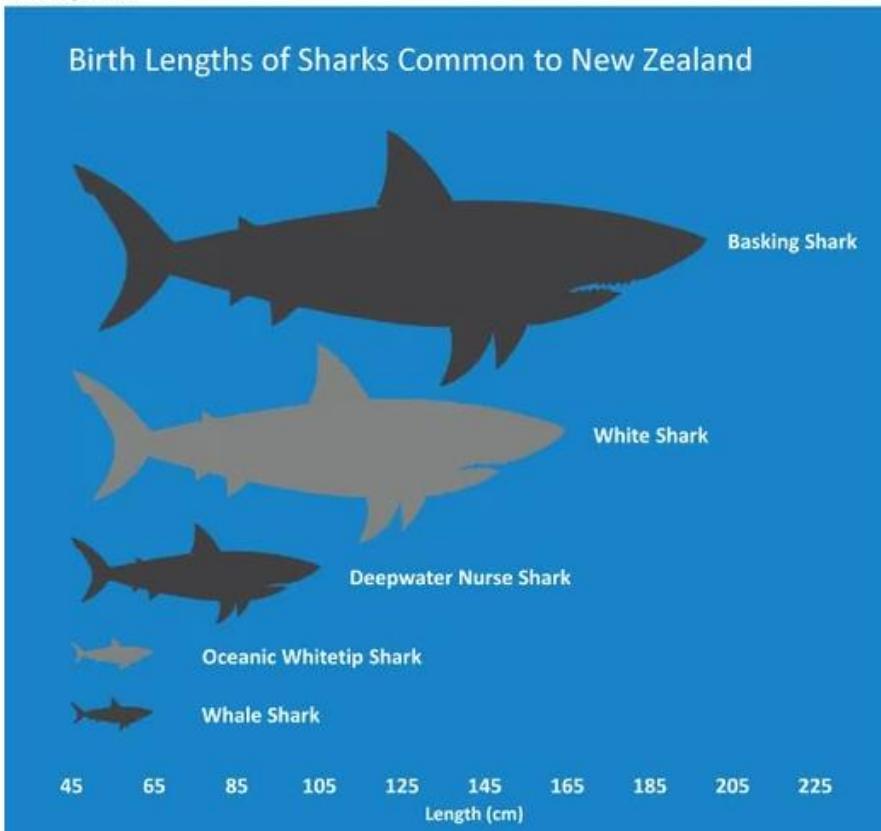
Just because data is visualized doesn't mean it's good data visualization.



Baby sharks

Are basking sharks 7 to 8 times larger than whale sharks?

Example A

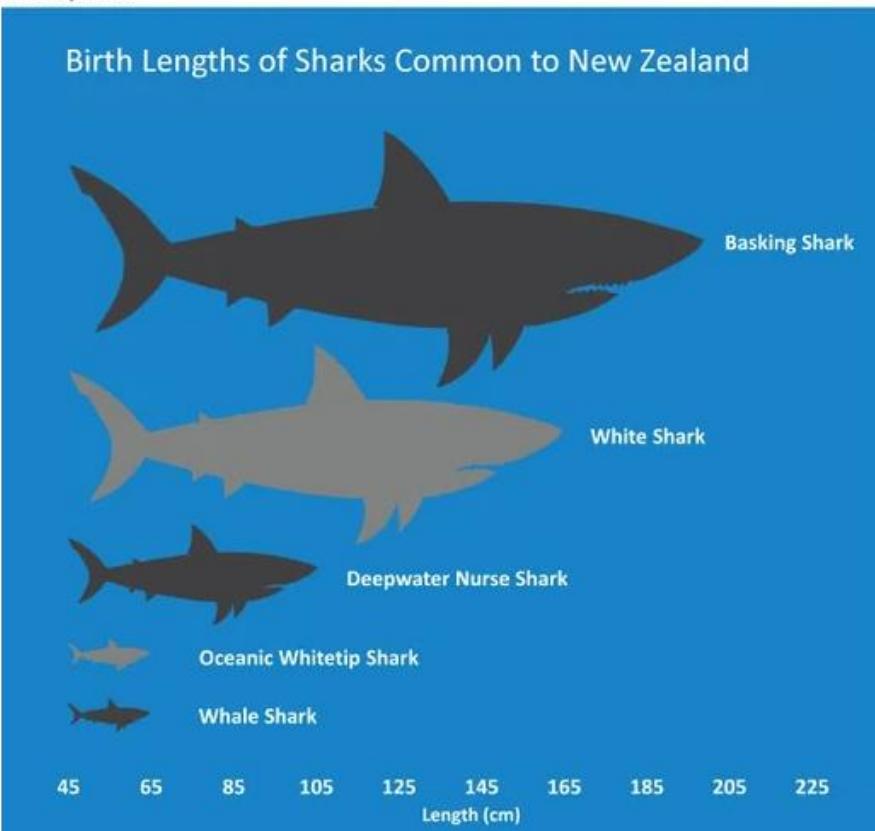


<https://www.datamine.com/datafix/index.php/2020-0>

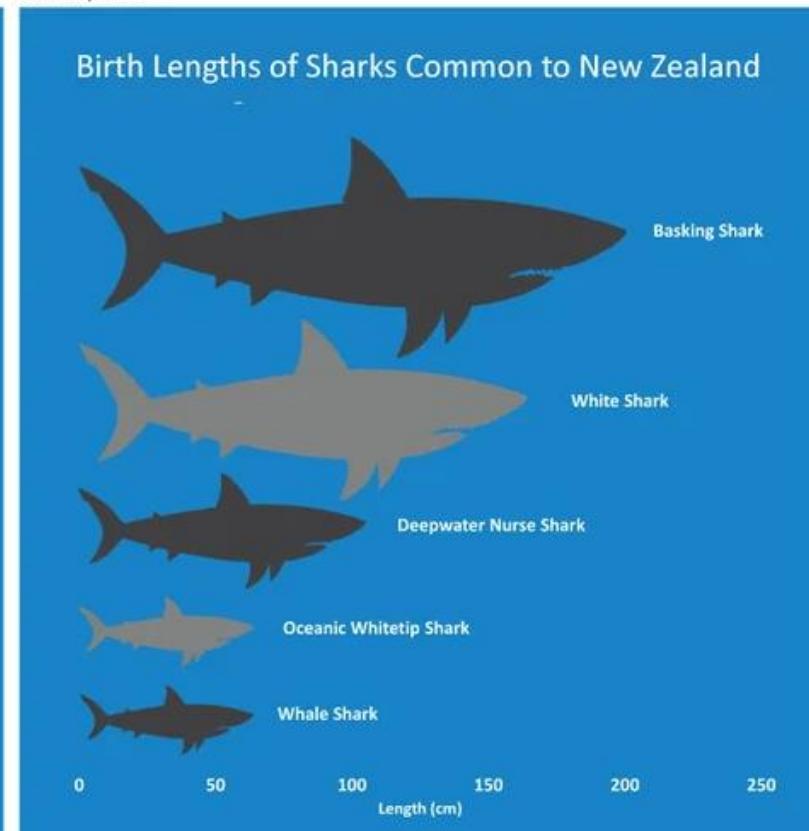
Baby sharks

Are basking sharks 7 to 8 times larger than whale sharks?

Example A



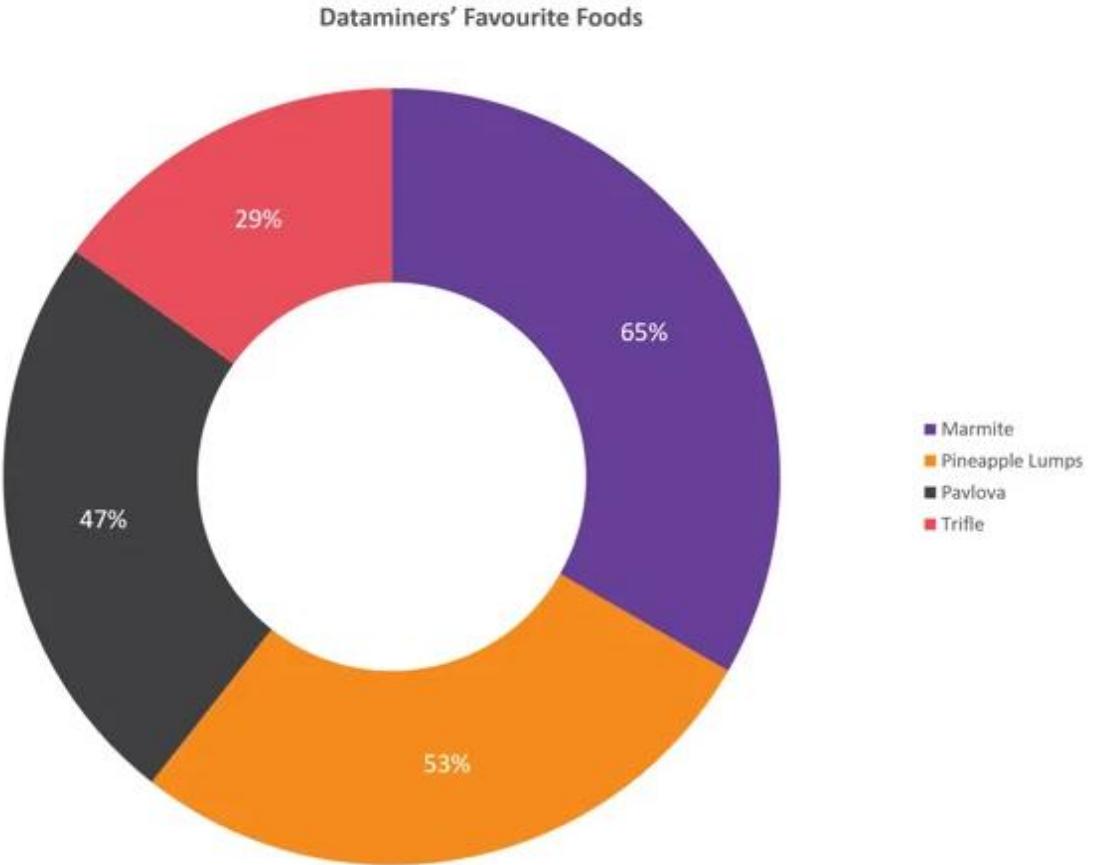
Example B



<https://www.datamine.com/datafix/index.php/2020-0>

Good taste

A keen eye would see that the sum of the parts adds to much greater than the whole, 194% in fact.



<https://www.datamine.com/datafix/index.php/2020-0>

The Lines are Lying

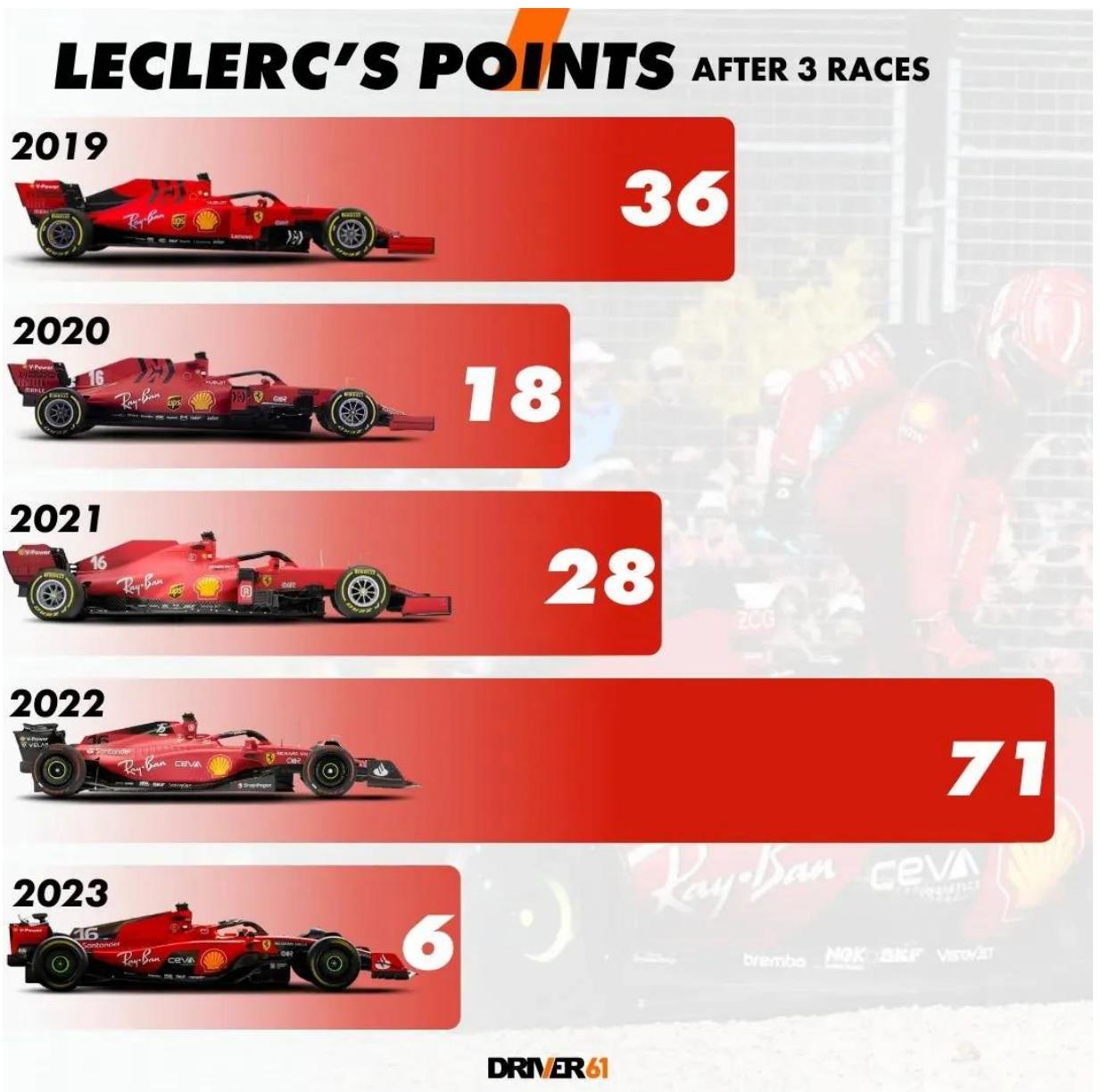
We might mistake that Fico Gutiérrez has higher popularity than Gustavo Petro.



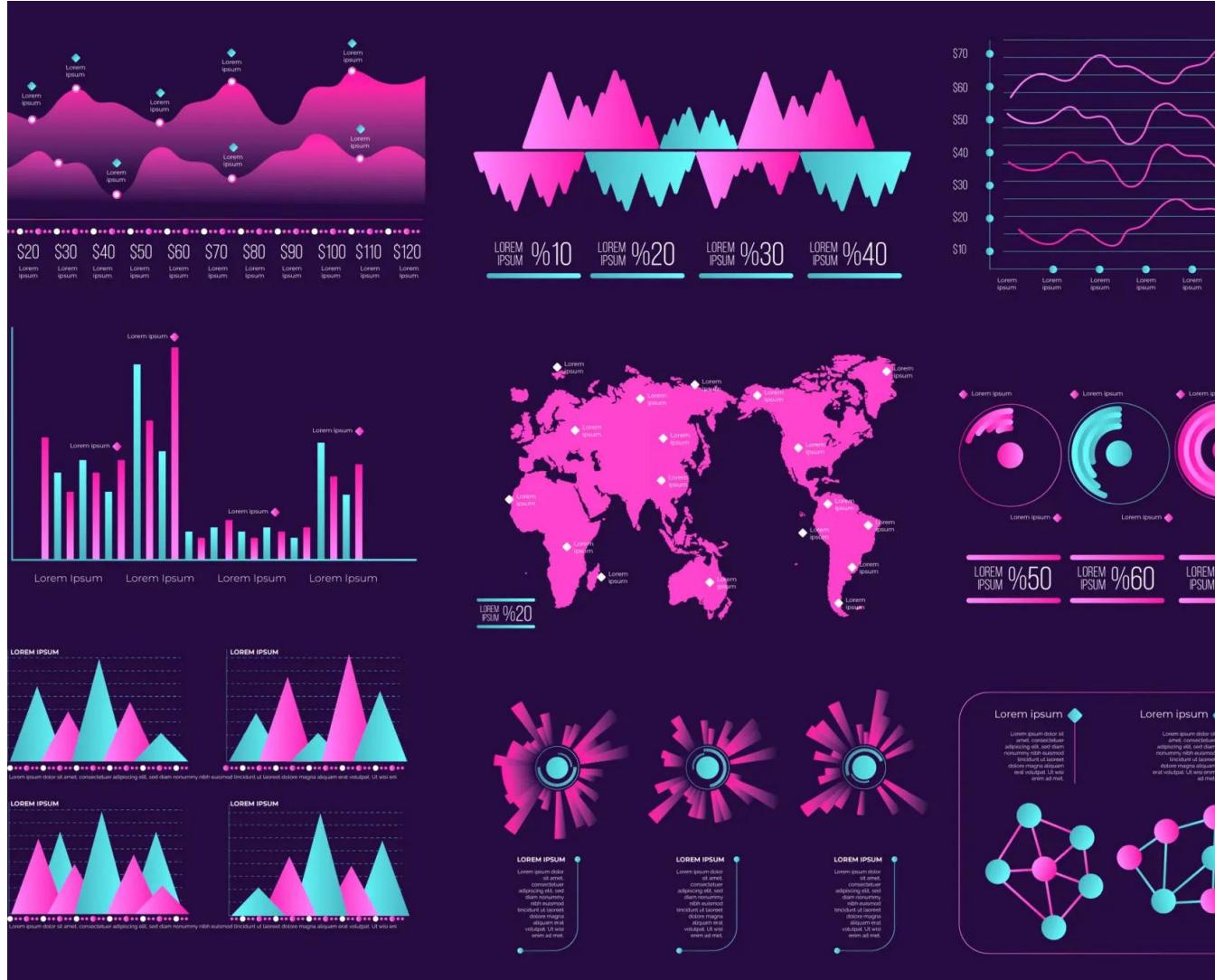
<https://www.codeconquest.com/blog/12-bad-data-visualization-examples-explained/#htoc-12-bad-data-visualization-examples>

It's all cars' fault

The only mistake here is that the car length is distorting the bars in a manner that changes all the aspects of the bars.



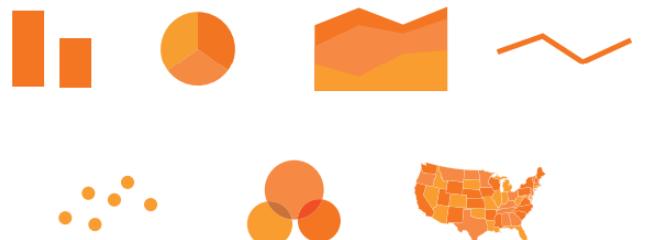
<https://www.codeconquest.com/blog/12-bad-data-visualization-examples-explained/#htoc-12-bad-data-visualization-examples>



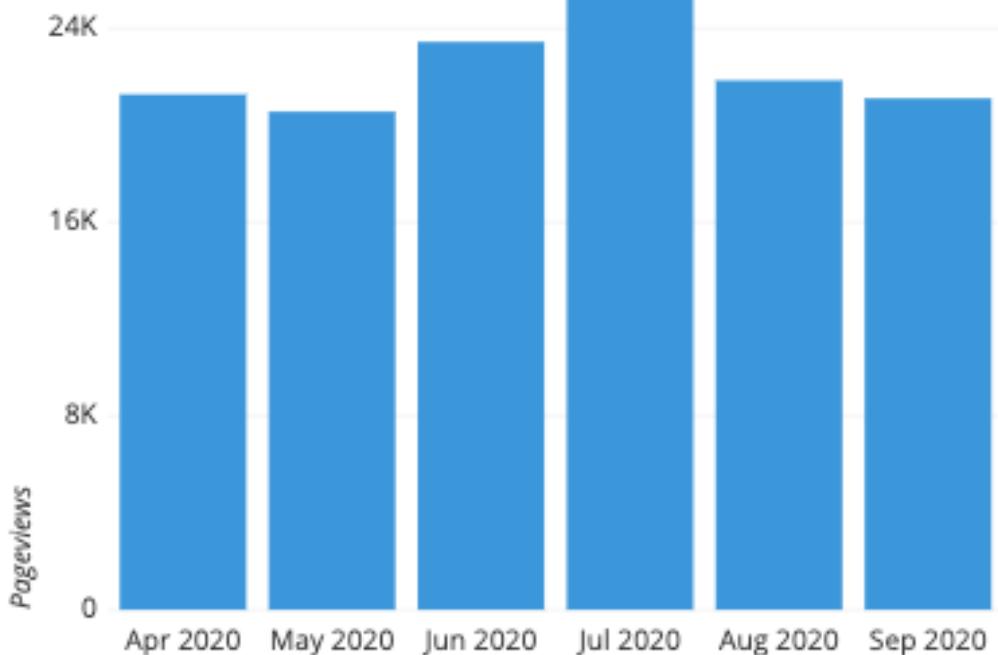
<https://datasciencedojo.com/blog/data-visualization-guide/#>

Choosing the right type of visualization

- Identify purpose and select the correct visualization for your data



Showing change over time or trends

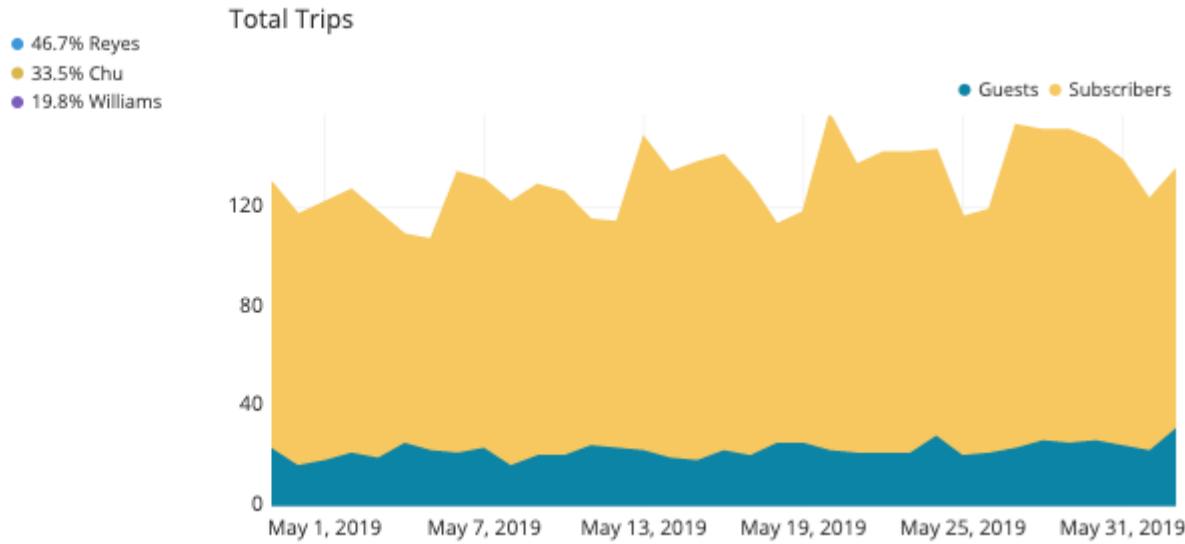
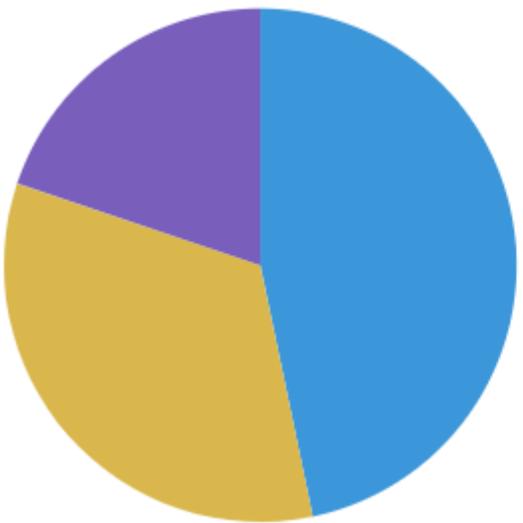


ZZD to QQY Exchange Rates



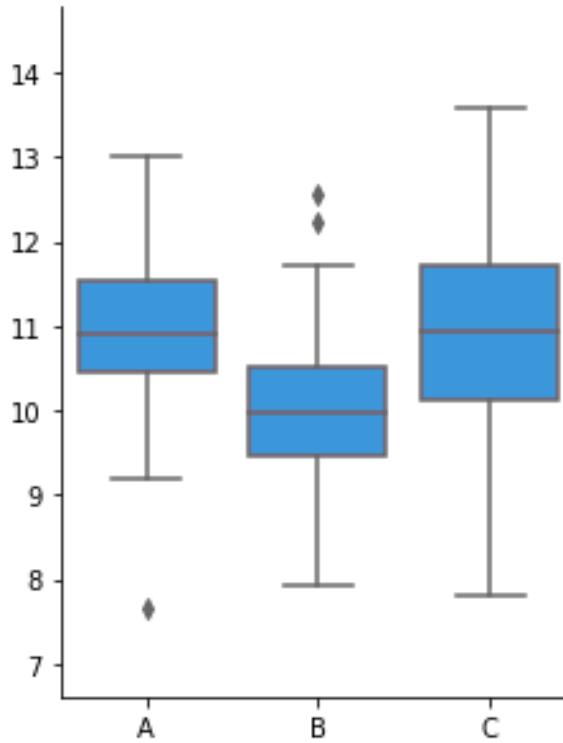
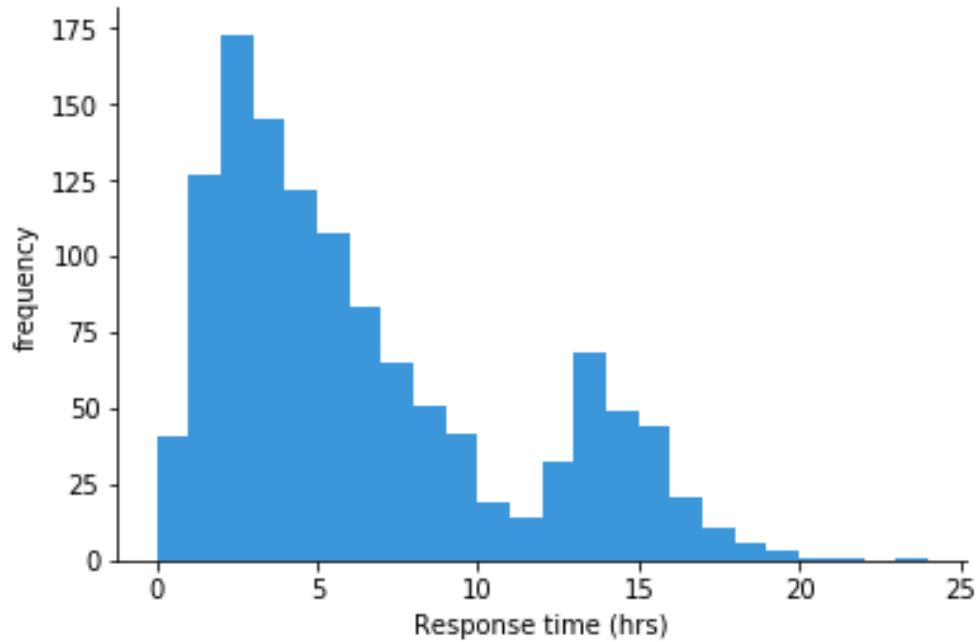
<https://chartio.com/learn/charts/how-to-choose-data-visualization/>

Showing part-to-whole composition



<https://chartio.com/learn/charts/how-to-choose-data-visualization/>

Looking at how data is distributed

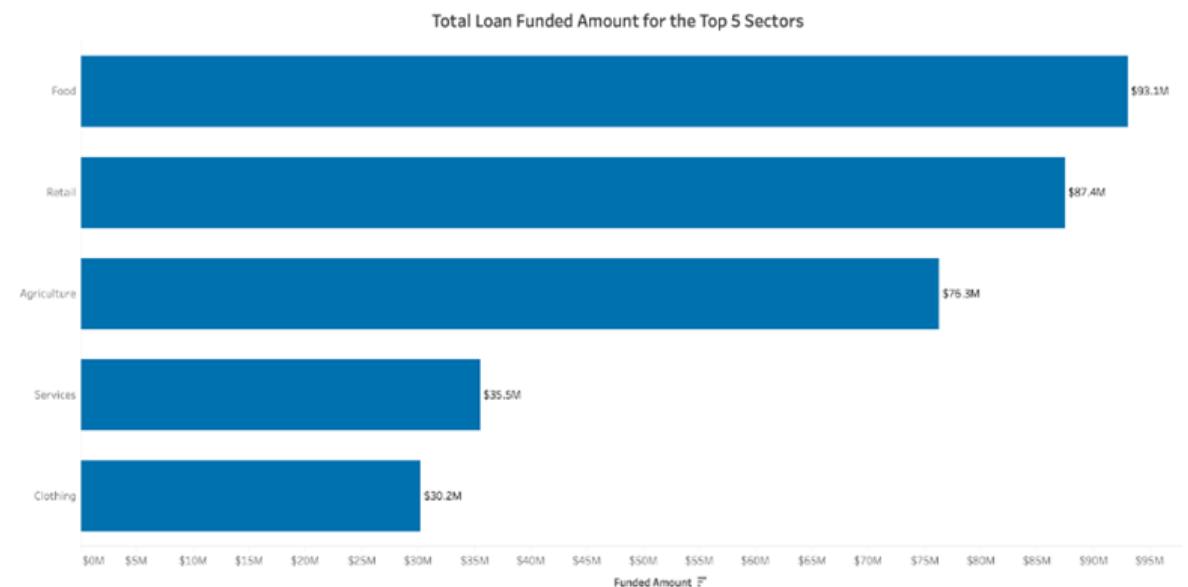


<https://chartio.com/learn/charts/how-to-choose-data-visualization/>

Comparing values between groups

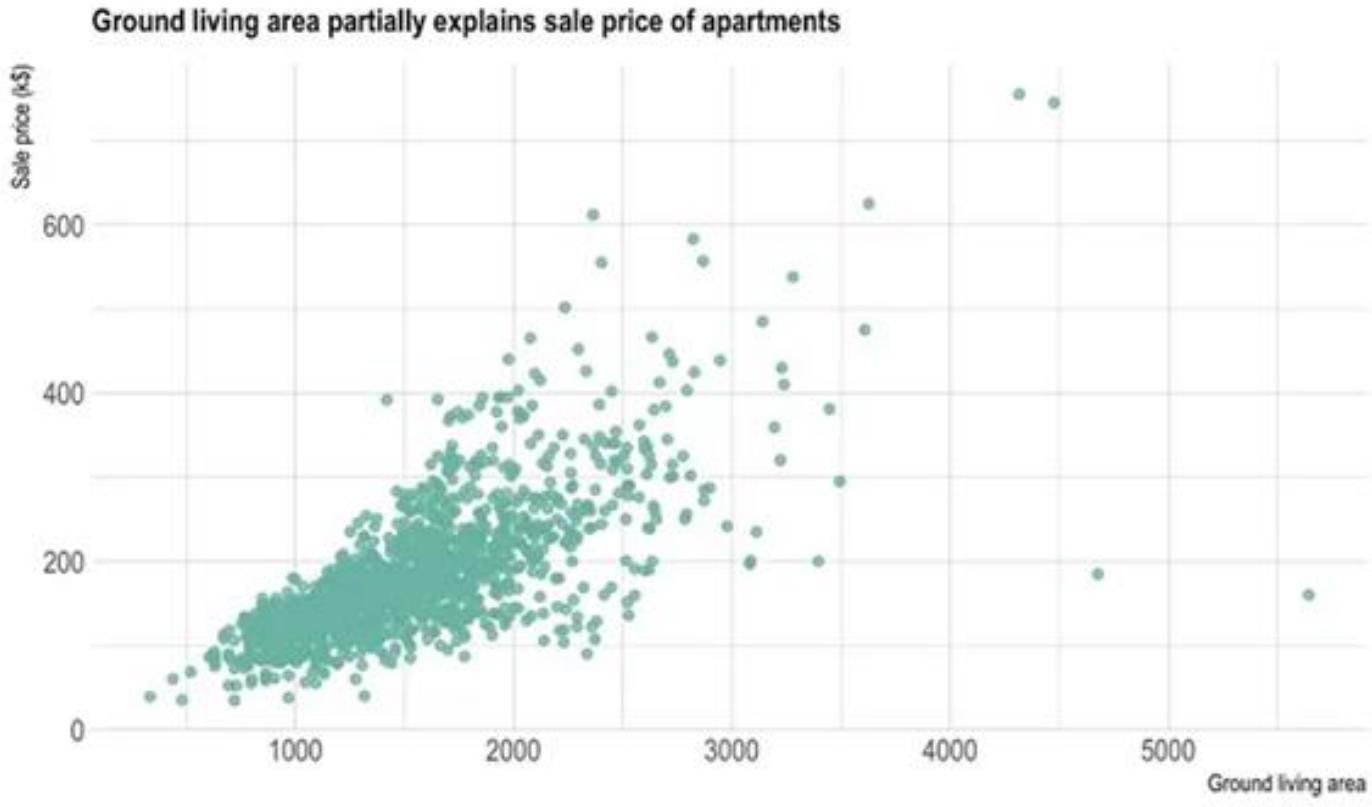


<https://chartio.com/learn/charts/how-to-choose-data-visualization/>



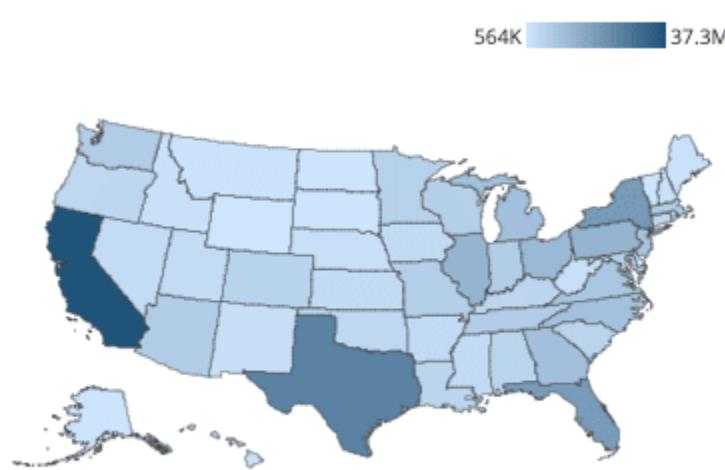
<https://www.tableau.com/data-insights/reference-library/visual-analytics/charts>

Observing relationships between variables

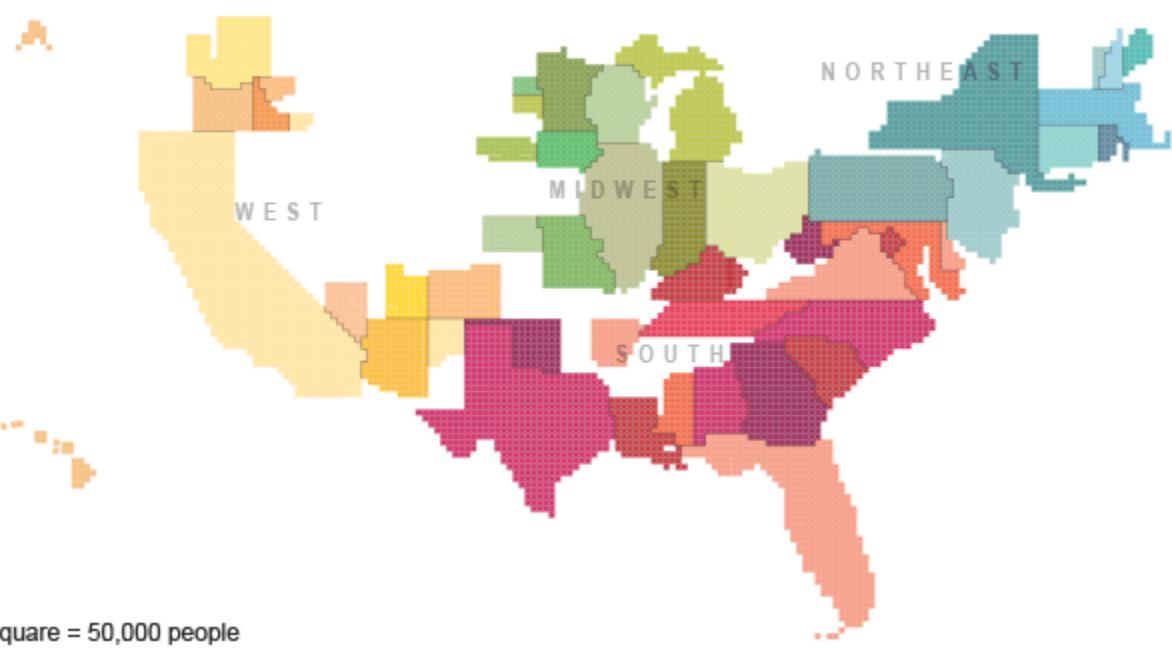


Looking at geographical data

2010 US Population

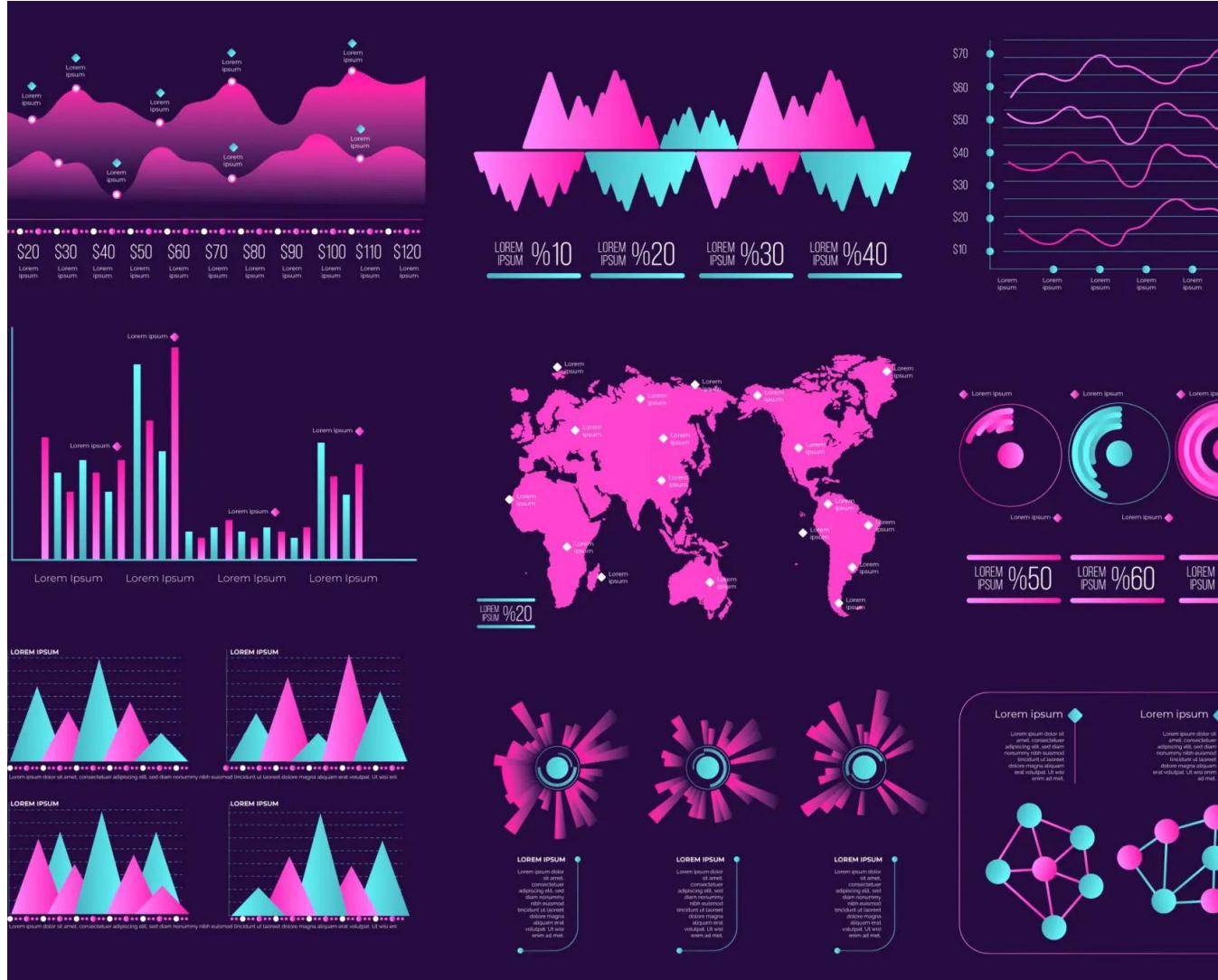


2010
Total population: 308,745,538



1 square = 50,000 people

<https://chartio.com/learn/charts/how-to-choose-data-visualization/>



<https://datasciencedojo.com/blog/data-visualization-guide/#>

Choosing the right type of visualization

- Understanding audience



Understanding audience

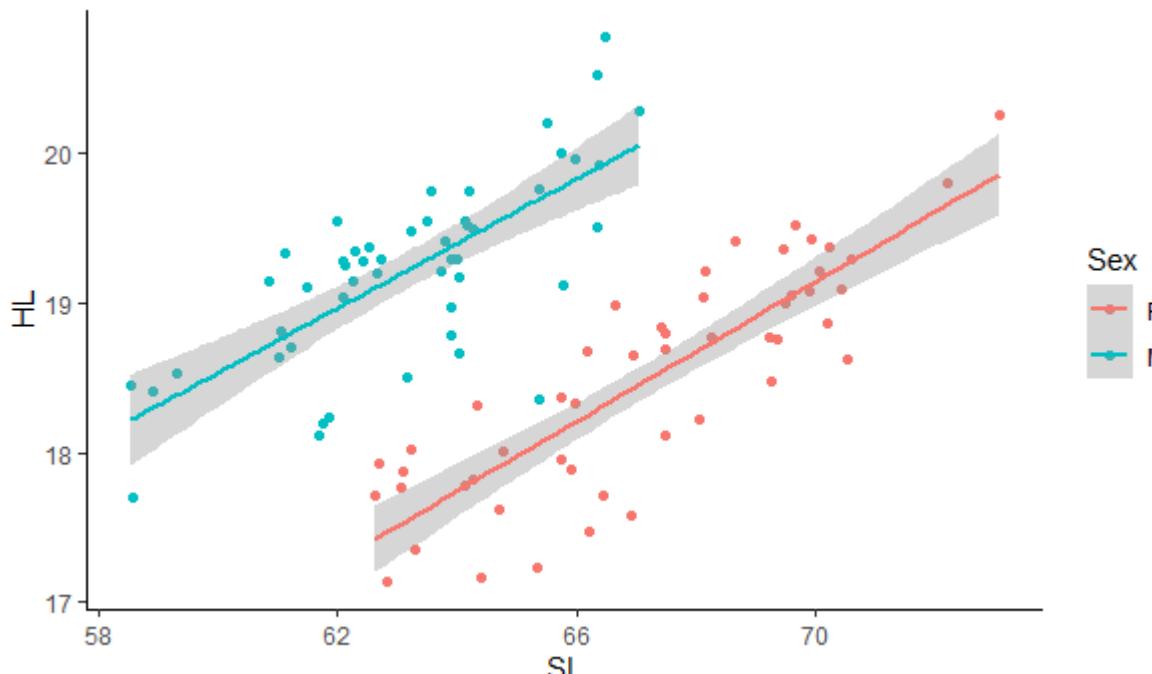
- Who your audience is
- What they care about
- What they need to know

Understanding audience

- Technical audience : more detailed and complex charts
- General audience: simpler and more colorful visuals.

Technical audience

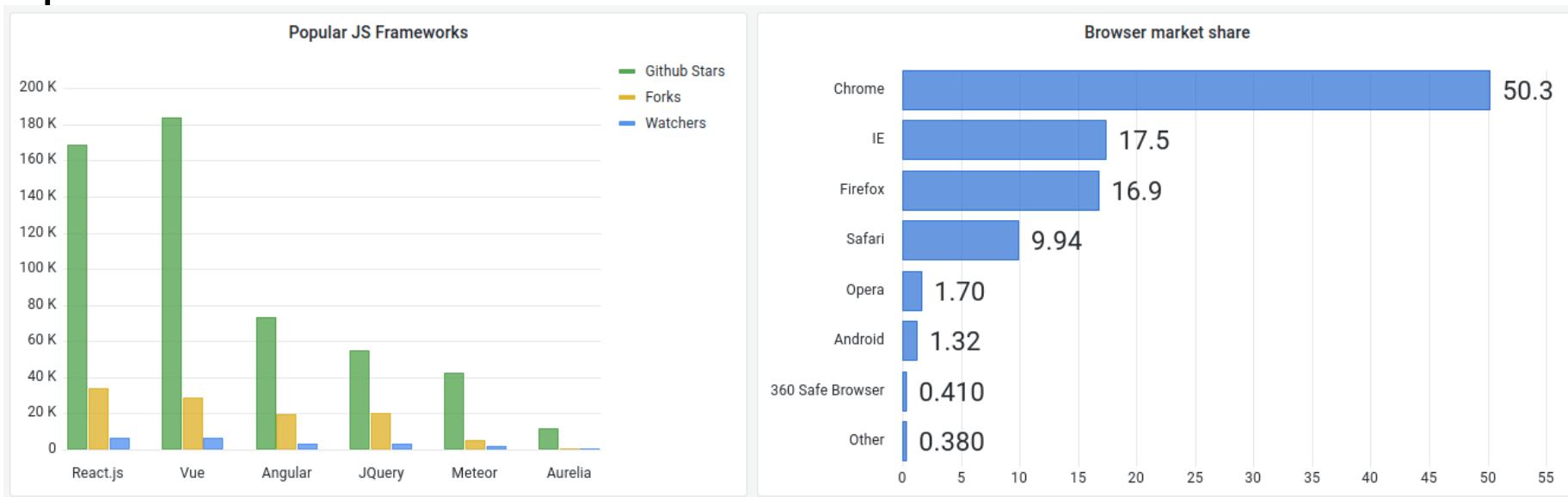
- Data scientists
 - a scatter plot with a regression line and confidence intervals to show the relationship between two variables



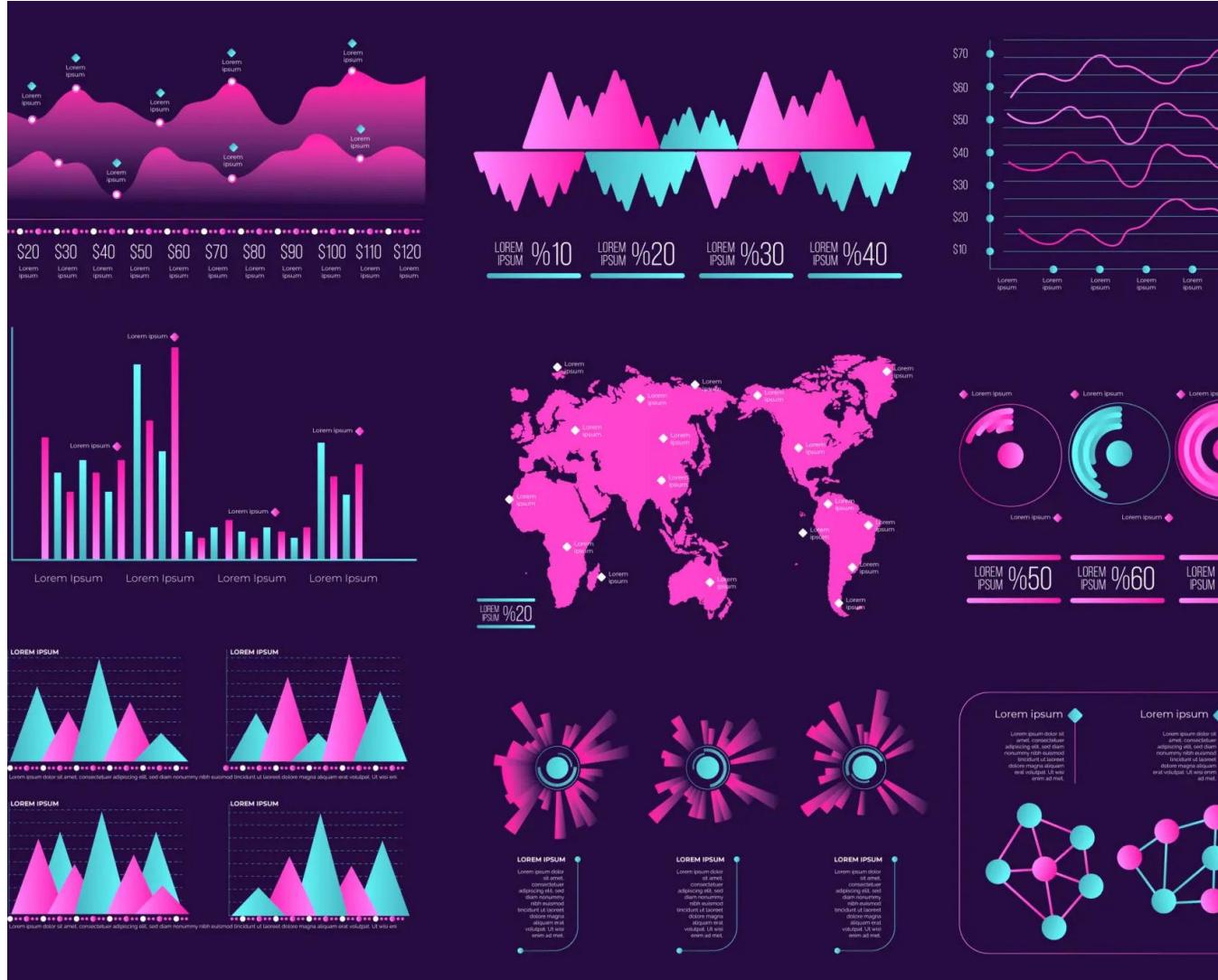
https://condor.depaul.edu/waguirre/r_scatter_plot_regression_line_ci.html
ruiwen.he@devinci.fr

Understanding audience

- Business managers
 - a bar chart with labels and annotations to highlight the key findings and implications.



<https://grafana.com/docs/grafana/latest/panels-visualizations/visualizations/bar-chart/>



Five tips to improve your data visualization

<https://datasciencedojo.com/blog/data-visualization-guide/#>

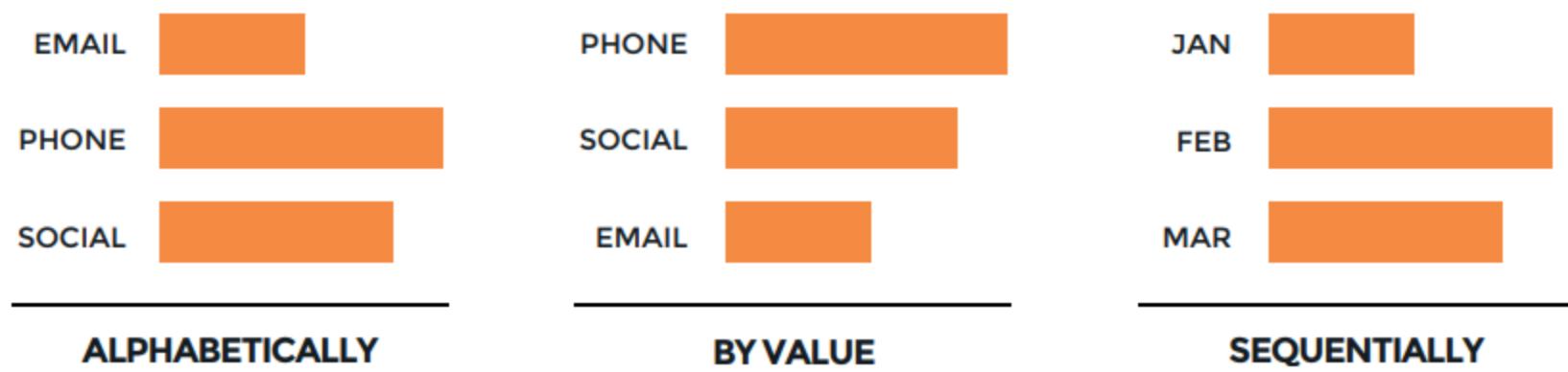
1. Accurately present the data

- Ensure that all visualizations are accurate and don't skew data

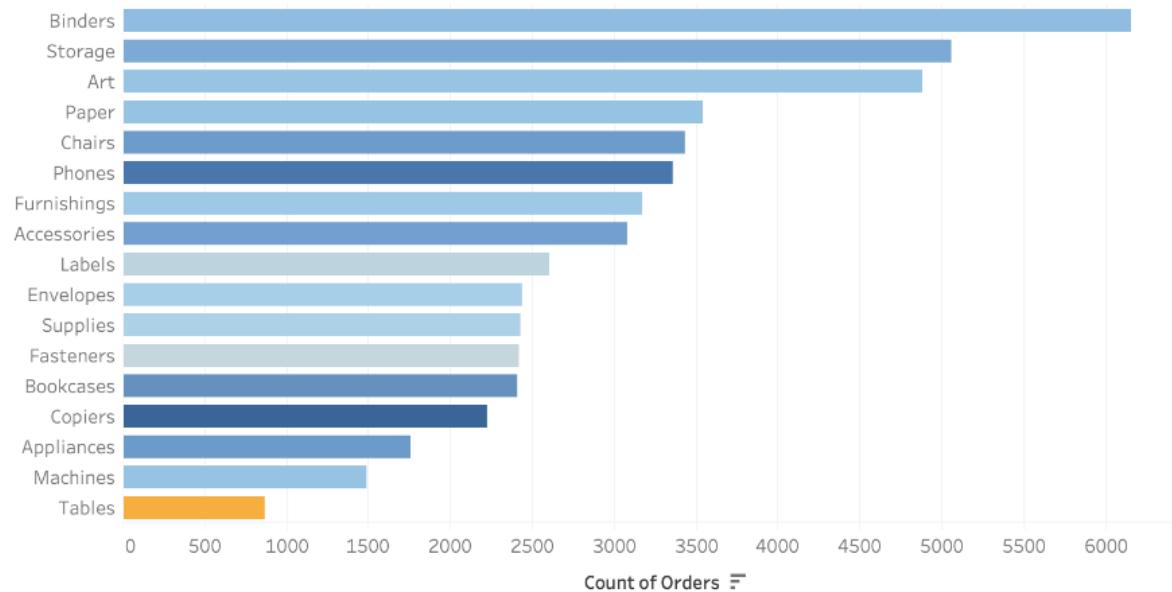
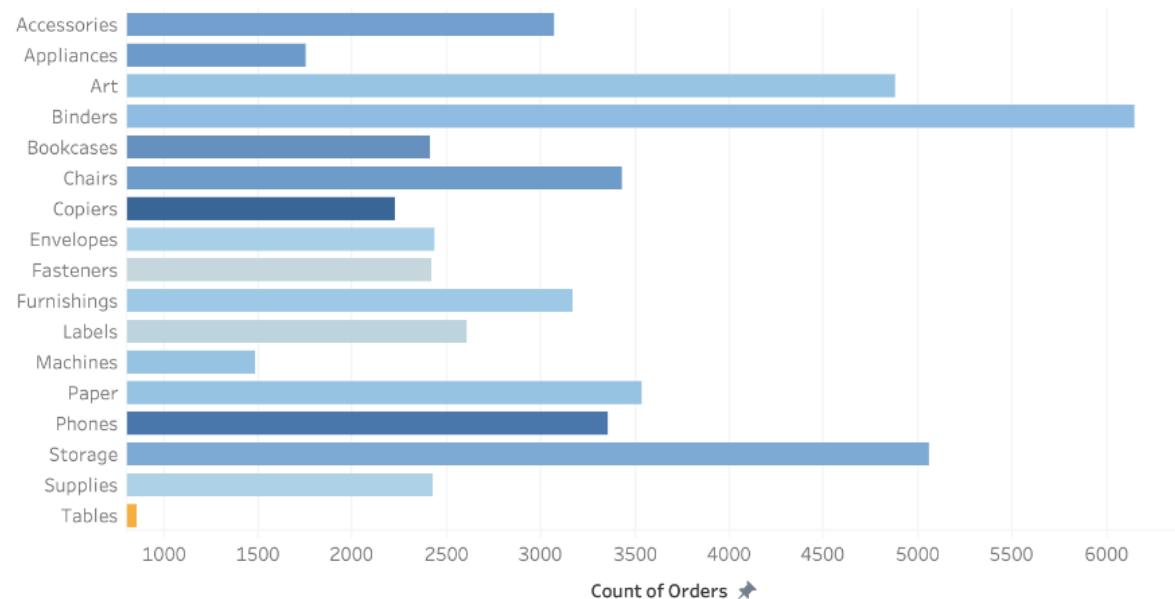


2. Order categories logically

- This allows viewers to navigate data more intuitively.



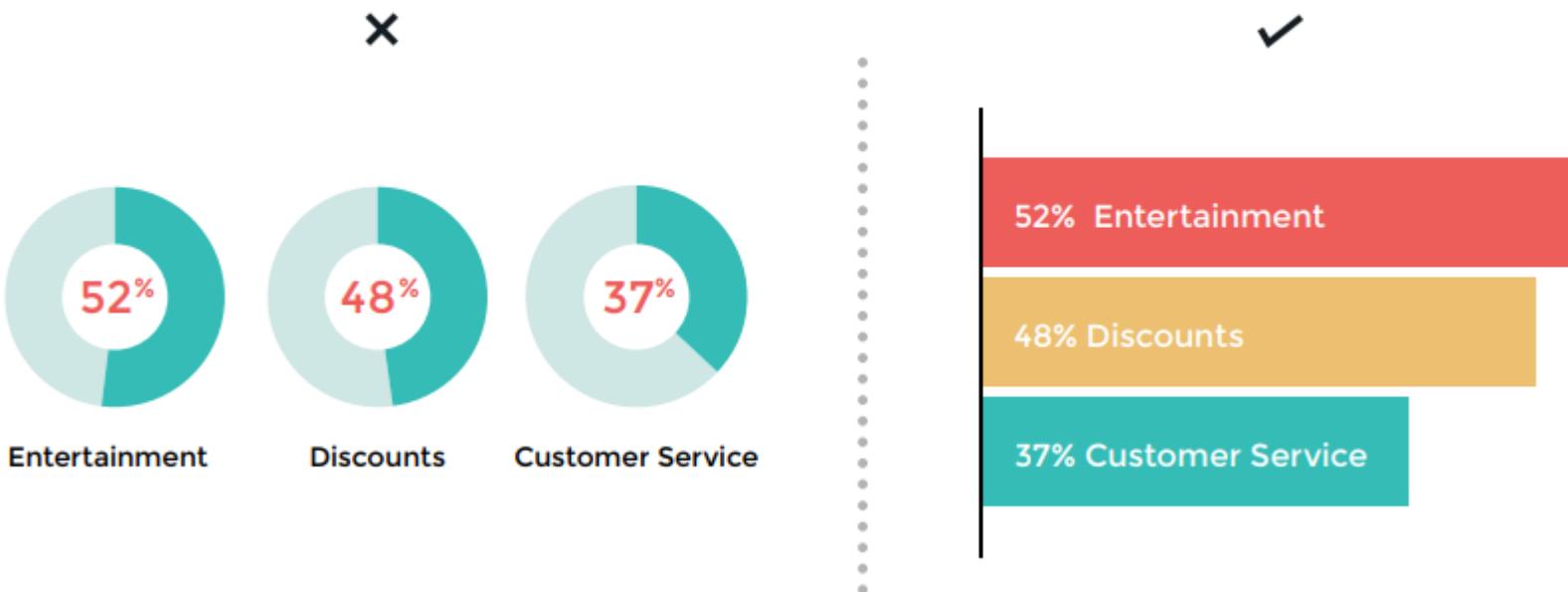
Poor example vs. better alternative



<https://www.tableau.com/data-insights/reference-library/visual-analytics/charts>

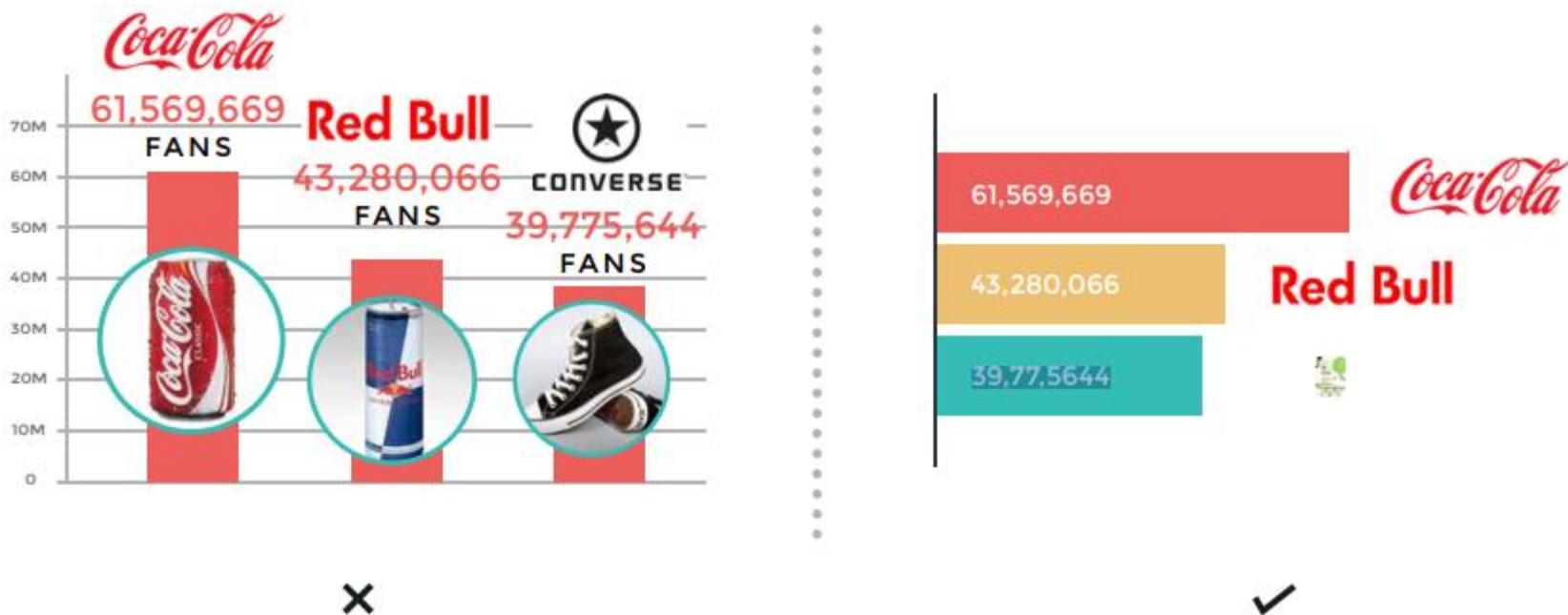
3. Visualize data in a way that is easy for readers to compare values

- Data visualization should always increase comprehension, not confuse viewers or make them work harder.



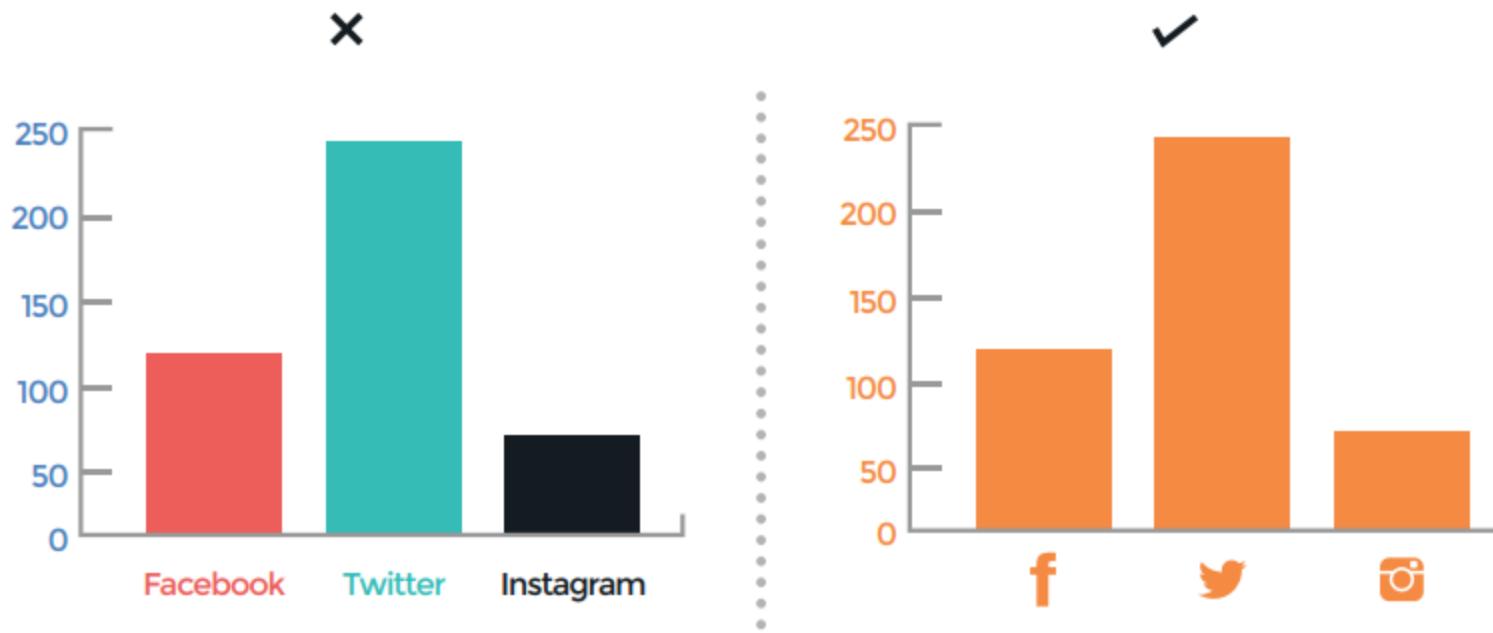
4. Don't distract from the data

- Unnecessary illustrations, dimensionality, or ornamentations (aka chart junk) can clutter the visualization



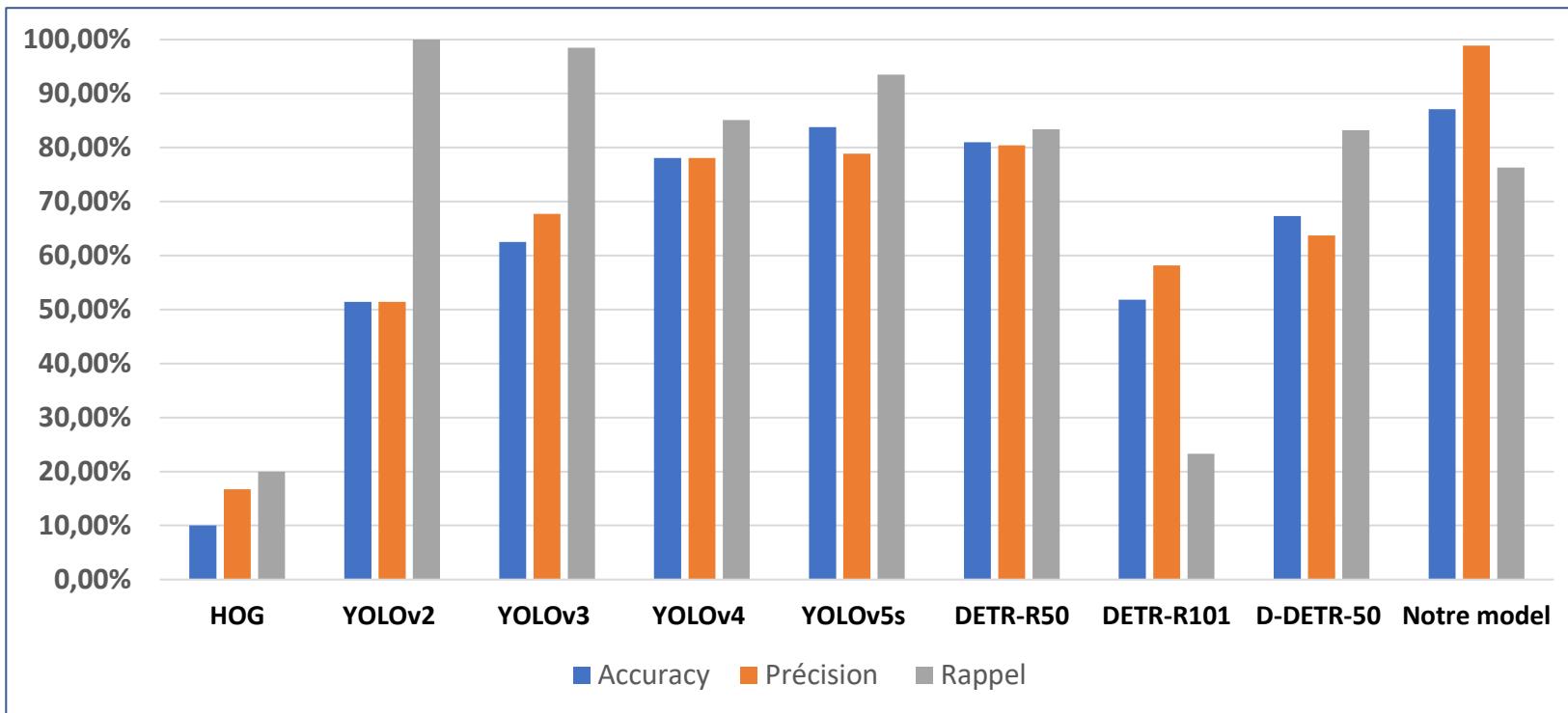
5. Use color and iconography appropriately

- This helps expedite understanding and reduce unnecessary labeling.



5. Use color and iconography appropriately

- This helps expedite understanding and reduce unnecessary labeling.



EXERCICE TABLEAU

<https://www.tableau.com/fr-fr/products/public>