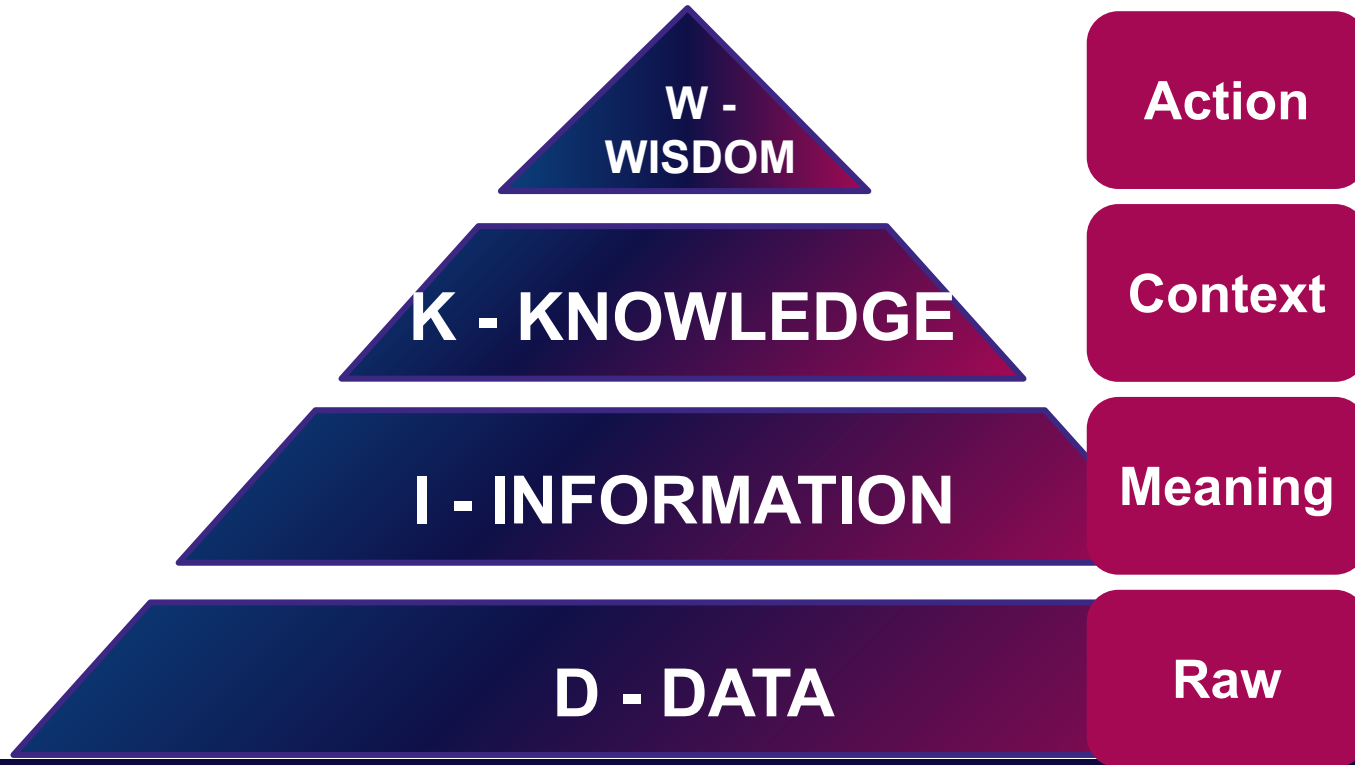


From Data to Information

Angelica Lo Duca
angelica.loduca@iit.cnr.it

The **DIKW** Pyramid



* **Introduction to Data Visualization & Storytelling: A Guide For The Data Scientist** by Jose Berengueres and Marybeth Sandell. Independently published, 2019



Numbers
Texts
...

Meaning

THE TWO-STEP PROCESS FOR DEVELOPING A DATA STORY



THE TWO-STEP PROCESS FOR DEVELOPING A DATA STORY



Tableau

THE TWO-STEP PROCESS FOR DEVELOPING A DATA STORY



THE TWO-STEP PROCESS FOR DEVELOPING A DATA STORY



INFORMATION



Numbers
Texts
...

Exploratory

Meaning

Explanatory

	Exploratory	Explanatory
Goal	Understand	Communicate

	Exploratory	Explanatory
Goal	Understand	Communicate
Audience	You	Other people

	Exploratory	Explanatory
Goal	Understand	Communicate
Audience	You	Other people
Data familiarity	Very familiar (you)	Less familiar (others)

	Exploratory	Explanatory
Goal	Understand	Communicate
Audience	You	Other people
Data familiarity	Very familiar (you)	Less familiar (others)
Visualization focus	Flexibility and speed	Simplicity, clarity, and cohesion

How to extract information from data

- **Select only relevant data:** Identify and focus on the data most important to answering our research question or solving our problem. This may involve filtering or excluding certain data points or variables unnecessary for our analysis.
- **Aggregate less important data:** Simplify the analysis and make identifying patterns or trends in the data easier. For example, group data by period, geography, or demographic characteristics to highlight key insights.
- **Focus only on what is important:** Highlight the key insights and avoid clutter or unnecessary information. Use annotations or callouts to draw attention to specific data points or trends.
- **Choose the right chart:** Choose an appropriate chart for the type of data we are presenting and the message we want to communicate.
- **Calibrate the chart to the audience and the message to communicate:** Consider the audience when designing the chart. For example, avoid complex charts or technical jargon if your audience is unfamiliar with statistical concepts. Also, consider the message to communicate and choose a chart that best supports that message.
- **Remove unnecessary noise from the chart:** This includes gridlines, borders, or other visual elements that do not add value to our analysis. A clean and simple chart emphasizes the key insights and makes our data easier to understand.

DATA PERSPECTIVE

Select
only relevant data

Aggregate
less important data

Focus
only on what is important

VISUAL PERSPECTIVE

Choose
the right chart

Calibrate
the chart to the audience

Remove
unnecessary noise

DATA PERSPECTIVE

Select
only relevant data

Aggregate
less important data

Focus
only on what is important

VISUAL PERSPECTIVE

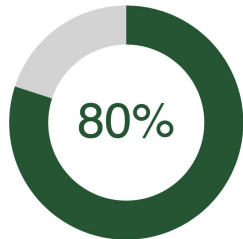
Choose
the right chart

Calibrate
the chart to the audience

Remove
unnecessary noise

Are we emphasizing only a single crucial piece of information?

80%



**Big Ass Number
(BAN)**

Donut Chart

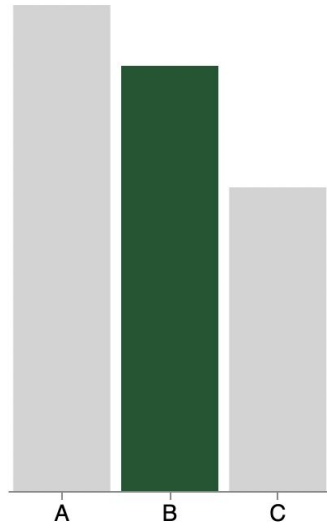
**100% Stacked Bar
Chart**

Icon Array

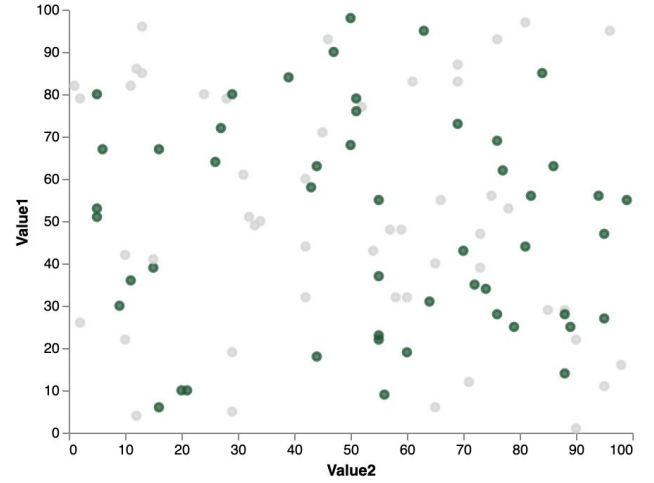
Are we drawing a comparison among two or more entities?



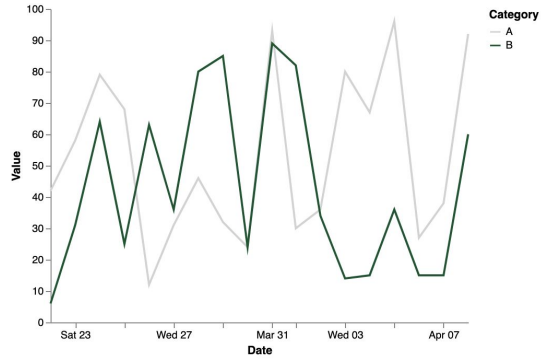
Bar Chart



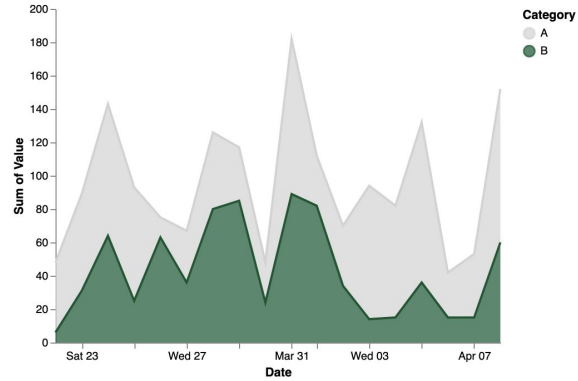
Column Chart



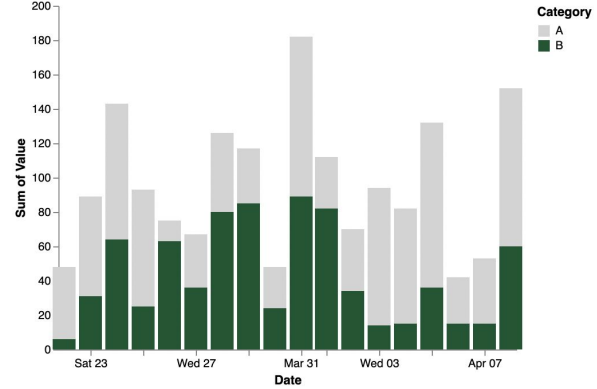
Scatter plot



Line Chart

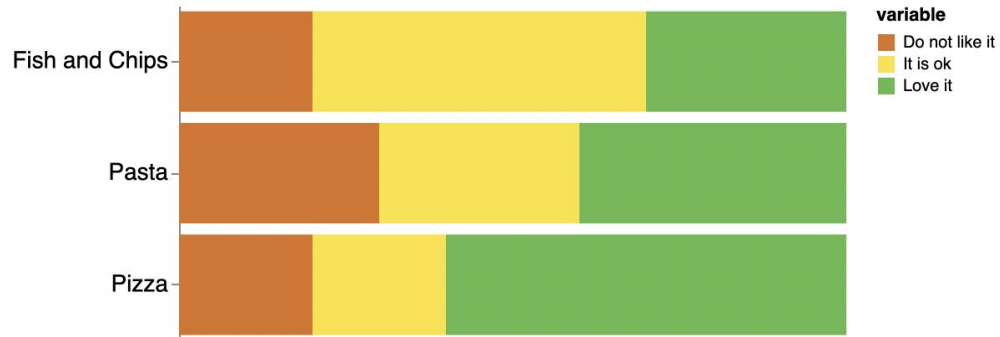


Stacked Area Chart

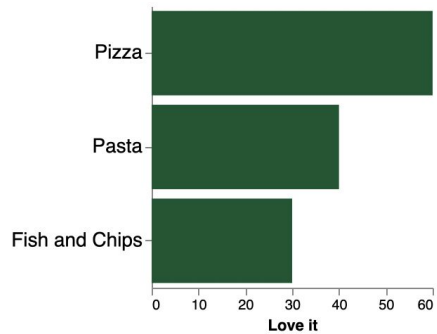


Stacked Column Chart

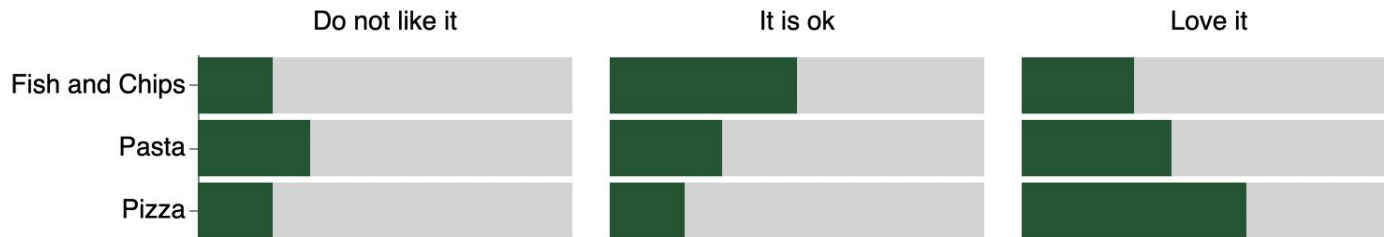
Are we displaying the outcomes of a survey
or a questionnaire?



Stacked Bar Chart

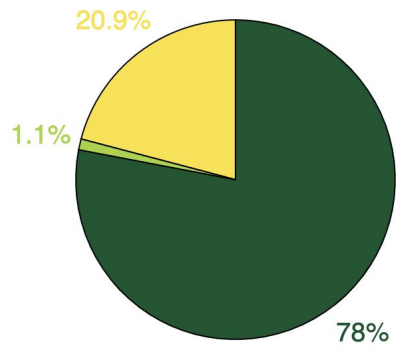


Column Chart

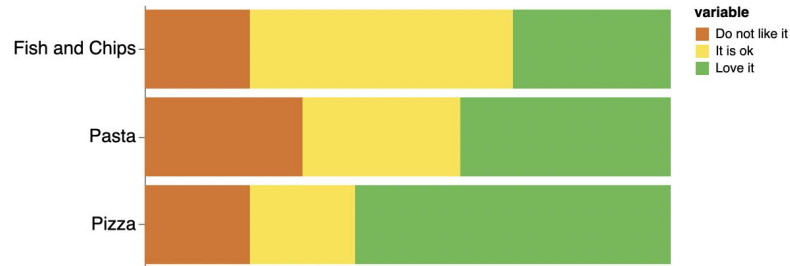


Small Bar Charts

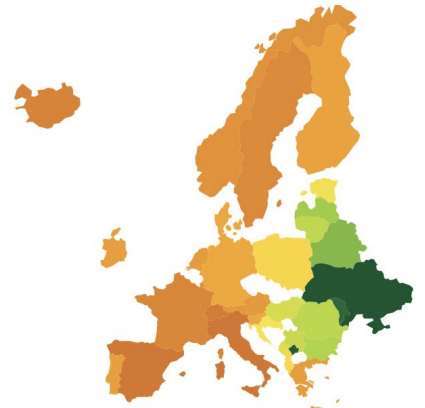
Are we explaining the relationship between individual components and the entirety?



Pie Chart



Stacked Bar Chart



Geographical Map

DATA PERSPECTIVE

Select
only relevant data

Aggregate
less important data

Focus
only on what is important

VISUAL PERSPECTIVE

Choose
the right chart

Calibrate
the chart to the audience

Remove
unnecessary noise

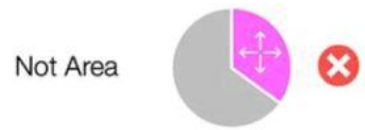
Data Visualization: Challenging Pie Chart Misconceptions

How do we interpret pie/donut charts?

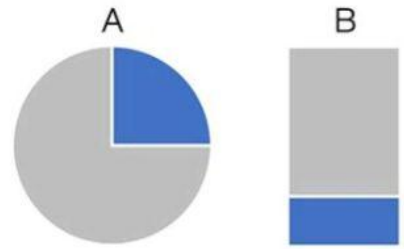


Data Visualization: Challenging Pie Chart Misconceptions

How do we interpret pie/donut charts?



Which is easier to verify the result is 25%?

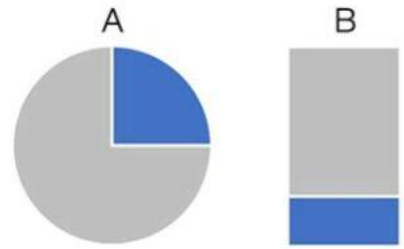


Data Visualization: Challenging Pie Chart Misconceptions

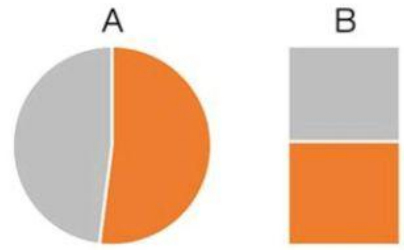
How do we interpret pie/donut charts?



Which is easier to verify the result is 25%?



Which is easier to verify the result is higher than 50%?



Effectivedatastorytelling.com

Data Storytelling: Which chart should you use?

It depends on what you're trying to communicate.

Scenario 1

Which is bigger? A or B? C or D?

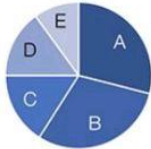


Data Storytelling: Which chart should you use?

It depends on what you're trying to communicate.

Scenario 1

Which is bigger? A or B? C or D?

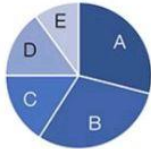


Data Storytelling: Which chart should you use?

It depends on what you're trying to communicate.

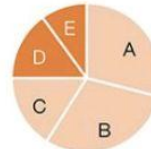
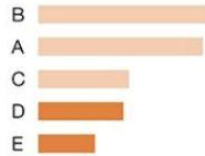
Scenario 1

Which is bigger? A or B? C or D?



Scenario 2

How much do D and E represent?

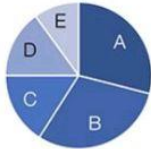


Data Storytelling: Which chart should you use?

It depends on what you're trying to communicate.

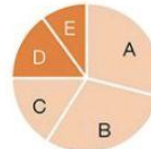
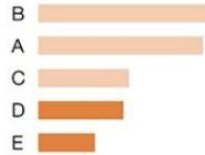
Scenario 1

Which is bigger? A or B? C or D?



Scenario 2

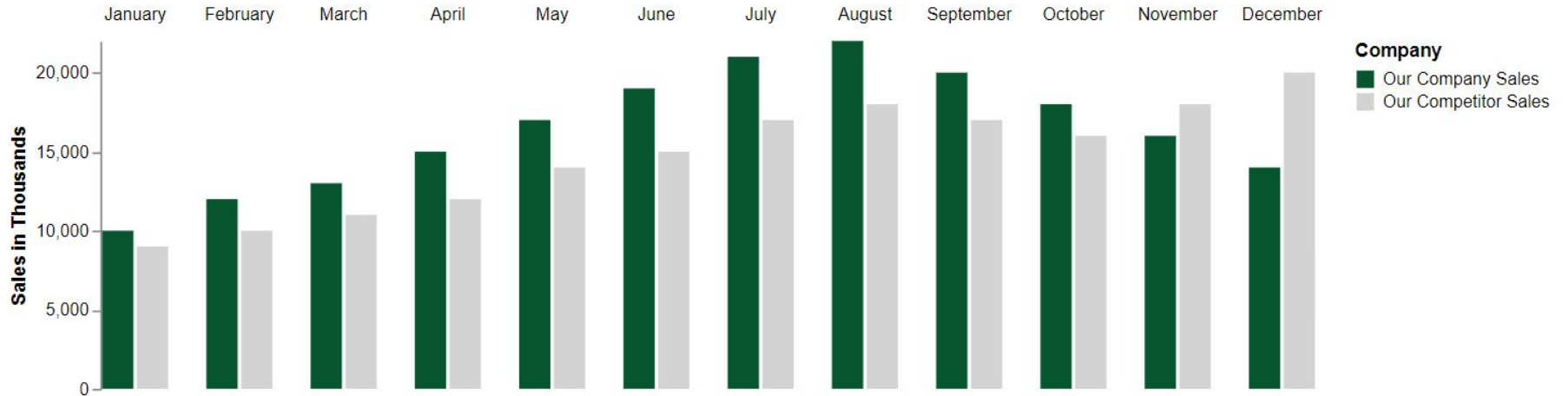
How much do D and E represent?



Effectivedatastorytelling.com

Totals

Totals refer to the raw values in a dataset. For example, consider a dataset showing our company's sales compared to our competitor's sales. We can use a bar chart showing all the trendlines to show totals.



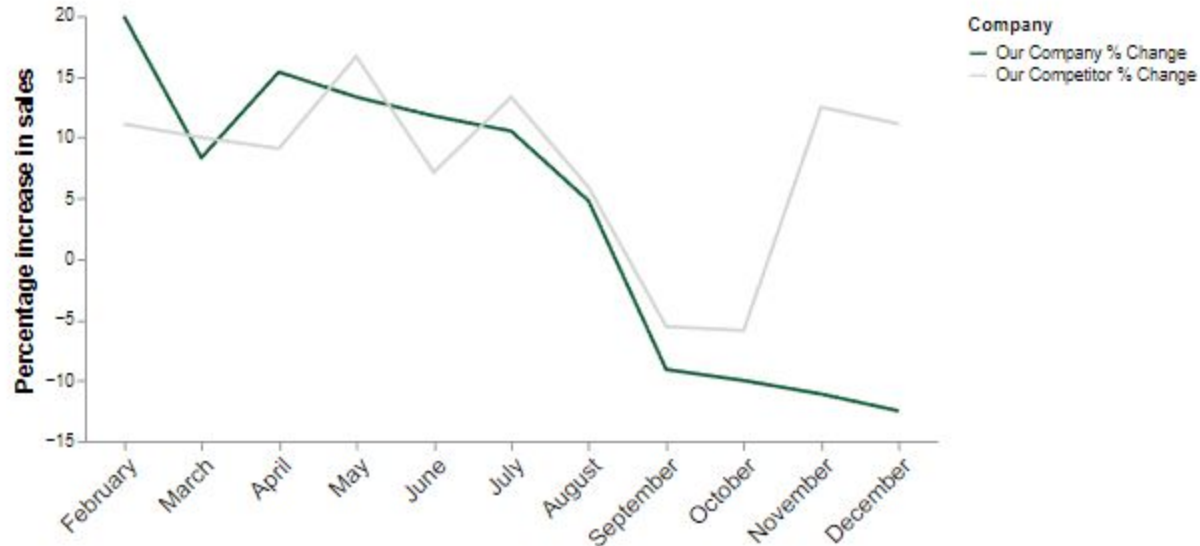
Percentage Change

Percentage changes refer to the difference between two values expressed as a percentage of the original value. Use them to measure the change in a value over time.

$$\%change = \frac{NewValue - OldValue}{OldValue} \times 100$$

Percentage Change (cont.)

Consider again the dataset containing our company's sales compared to our competitor's. Use a line chart to show the percentage change of the current month compared to the previous one, for each represented line.



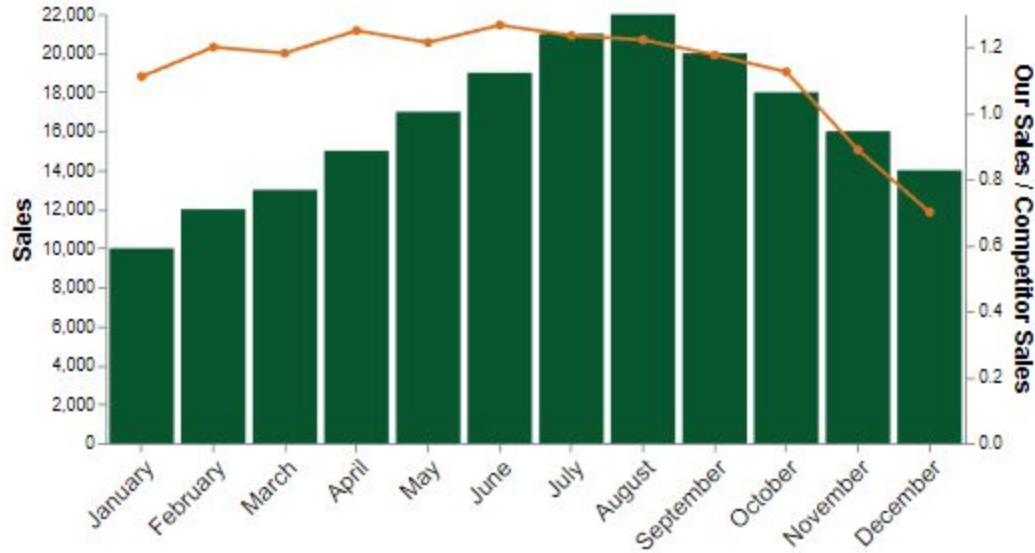
Ratio

The ratio is a comparison of two values or quantities. Use it to express the relationship between two values. The formula for calculating the ratio is:

$$\textit{Ratio} = \frac{\textit{Value1}}{\textit{Value1}}$$

Ratio (cont.)

Usually, we plot the ratio as a line on a secondary axis, as shown in the following figure:



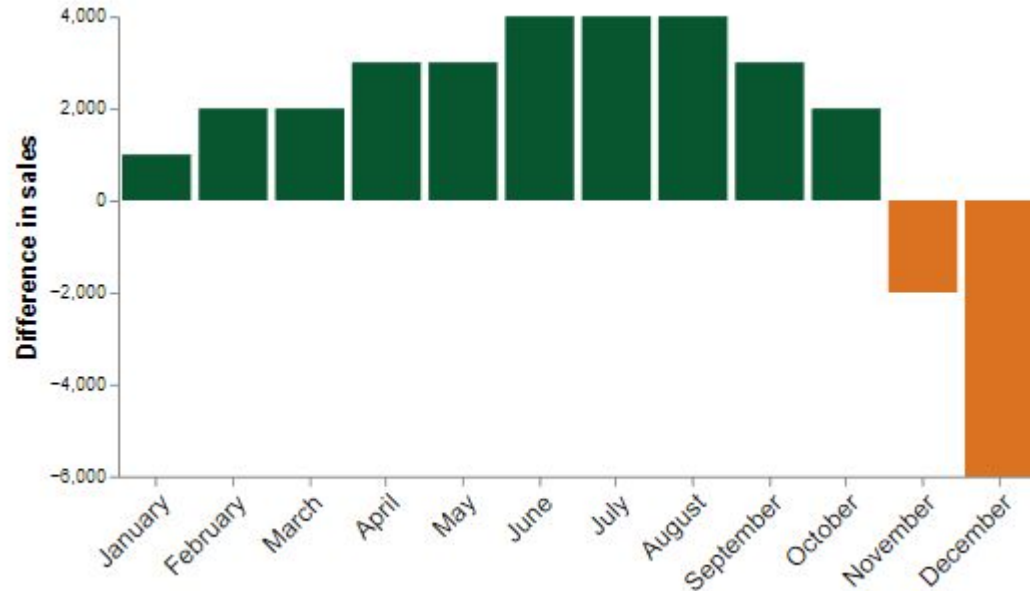
Difference

The difference compares two quantities or values directly. A positive value means that the first quantity outperforms the other, and a negative value means the contrary. The formula for calculating the difference is:

$$\textit{Difference} = \textit{Value1} - \textit{Value2}$$

Difference (cont.)

Consider the sales dataset again. The following chart describes a bar chart showing the difference between our company's and competitor's sales.



DATA PERSPECTIVE

Select
only relevant data

Aggregate
less important data

Focus
only on what is important

VISUAL PERSPECTIVE

Choose
the right chart

Calibrate
the chart to the audience

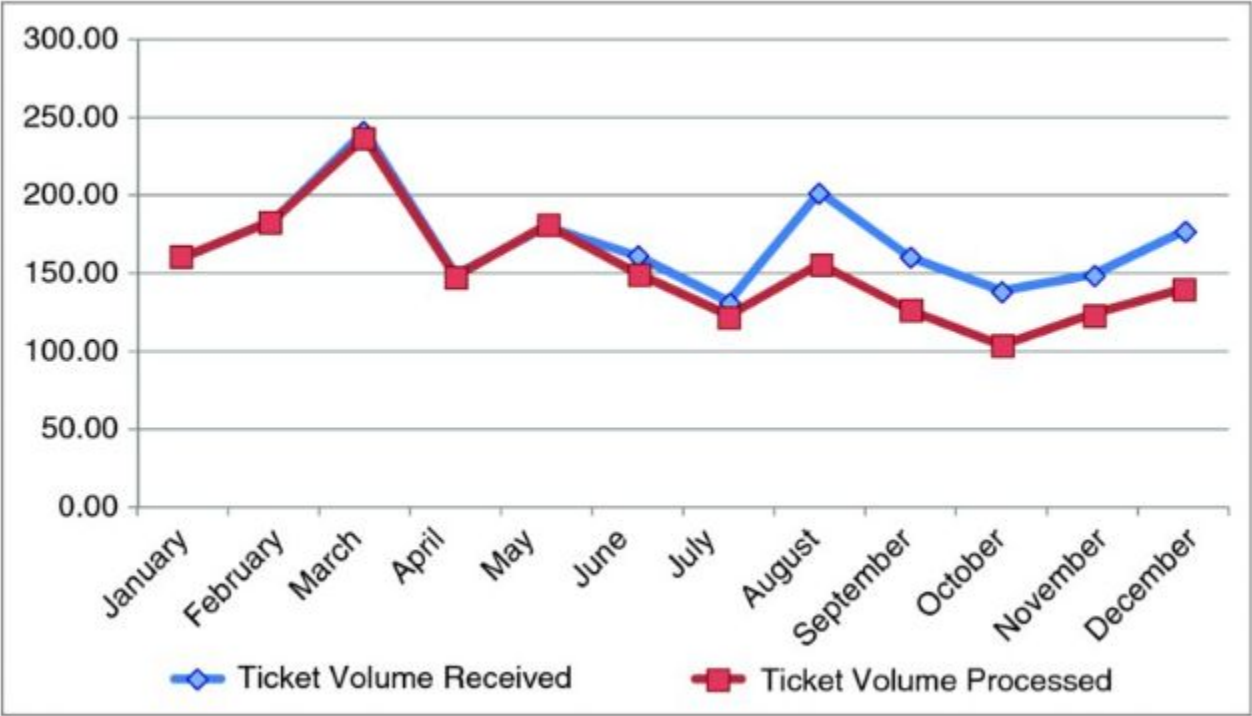
Remove
unnecessary noise

Clutters

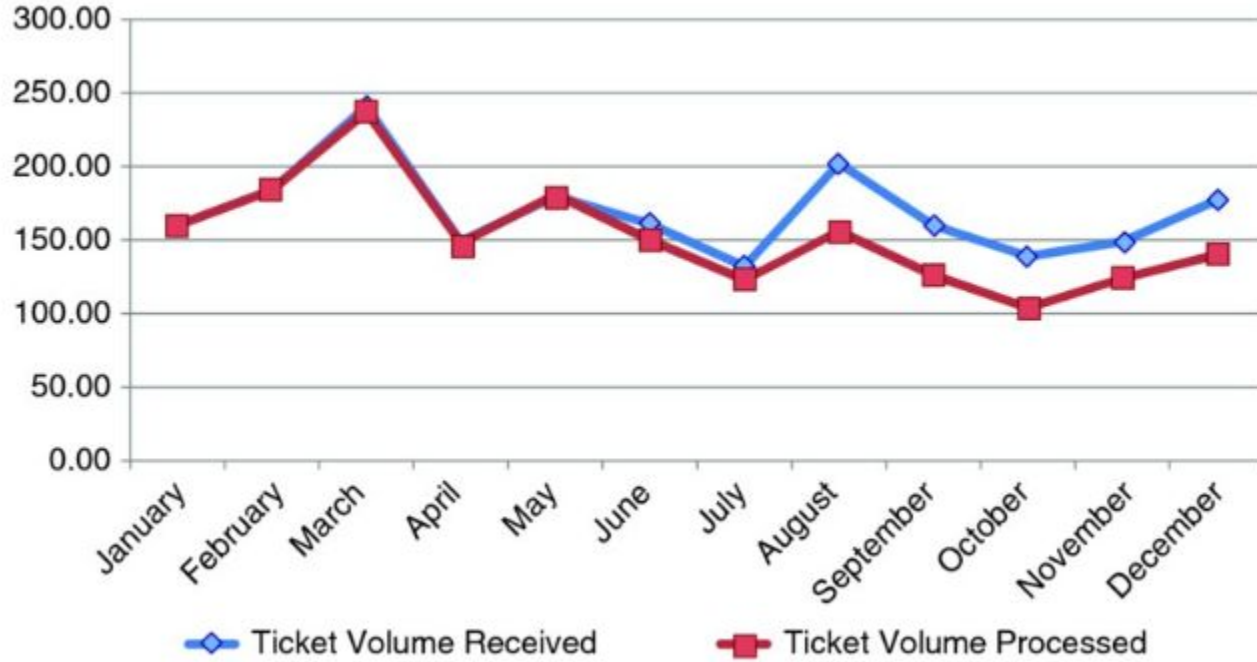
Clutters are visual elements that take up space but don't increase understanding.

We want to reduce clutter because it makes our visuals appear more complicated than necessary.

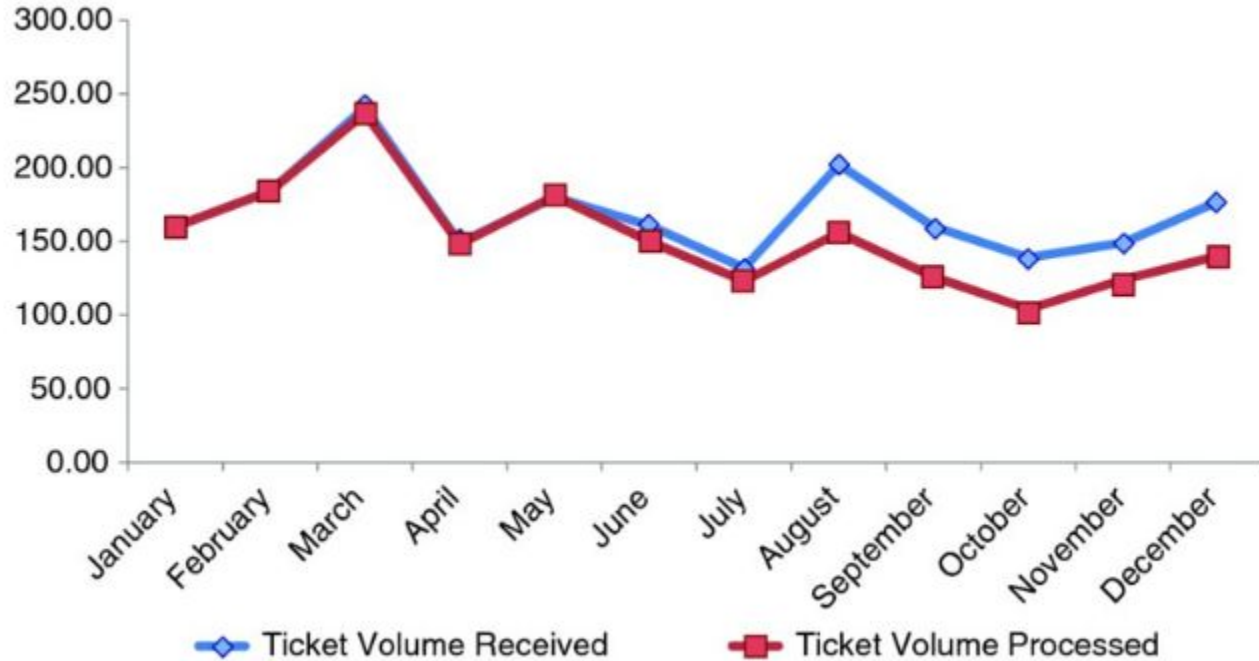
Example



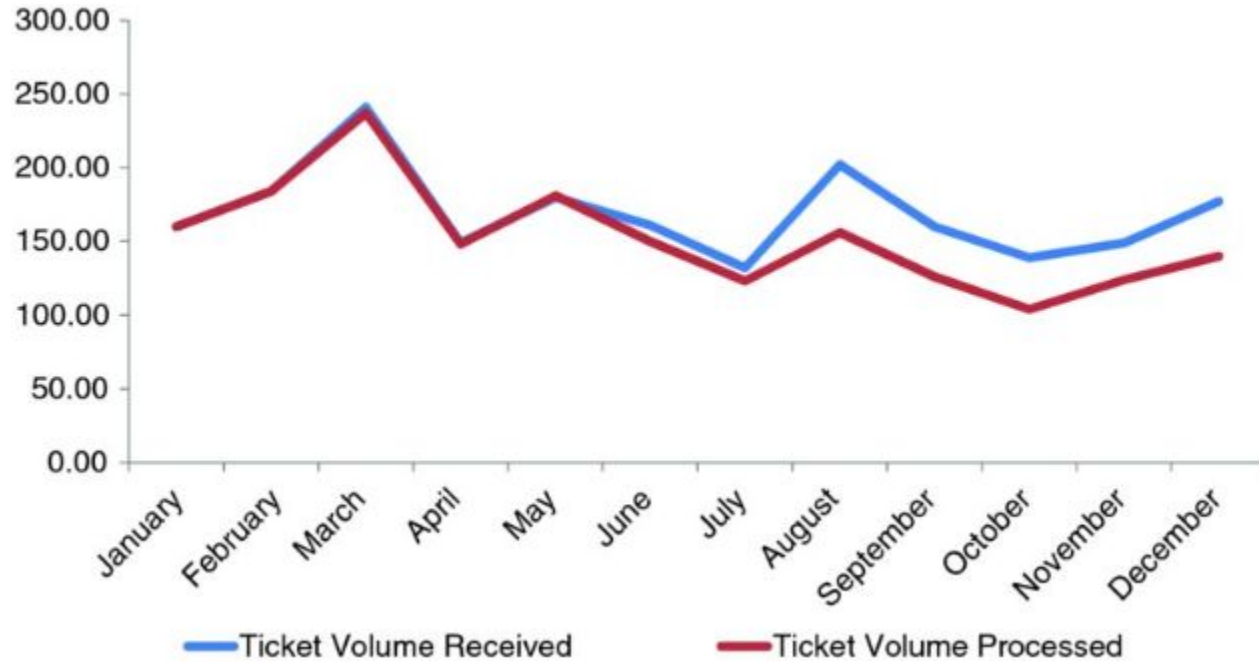
Remove chart border



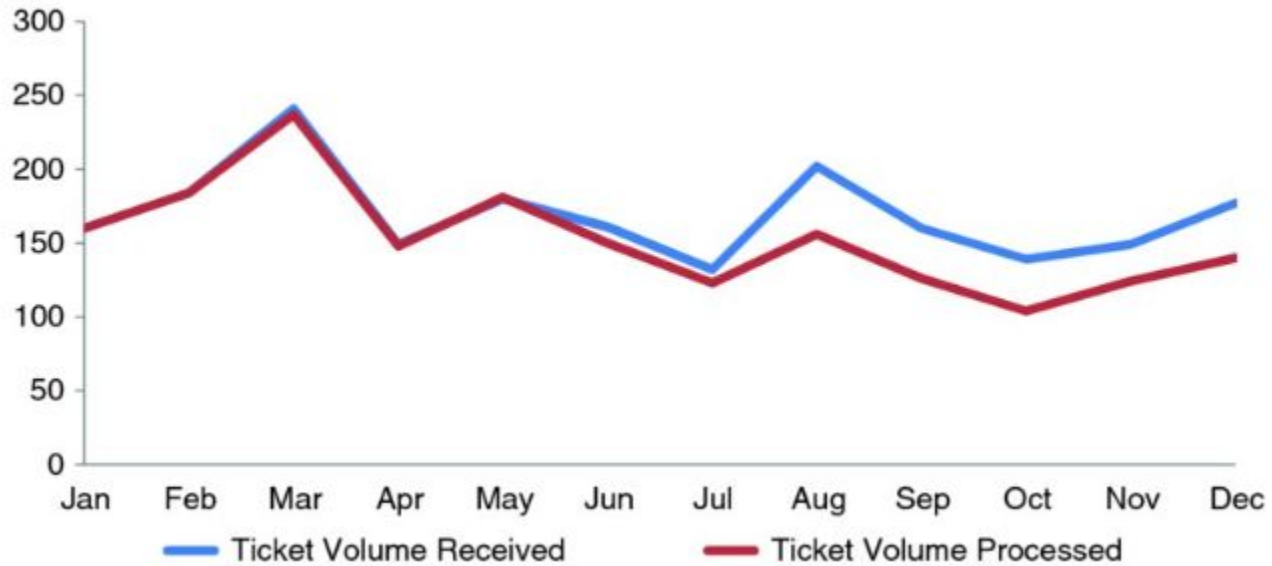
Remove gridlines



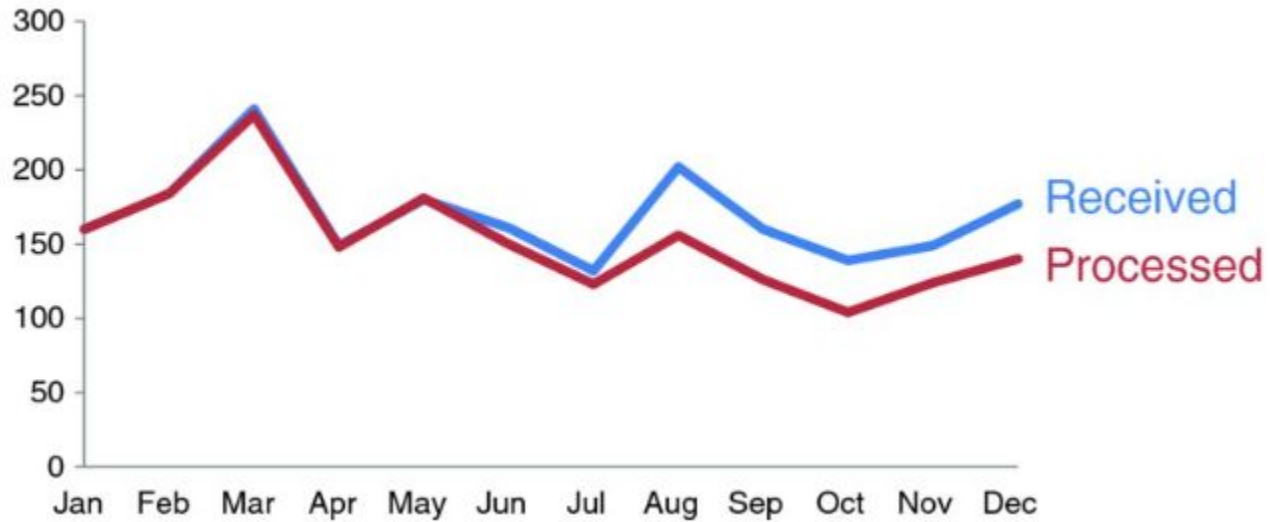
Remove data markers

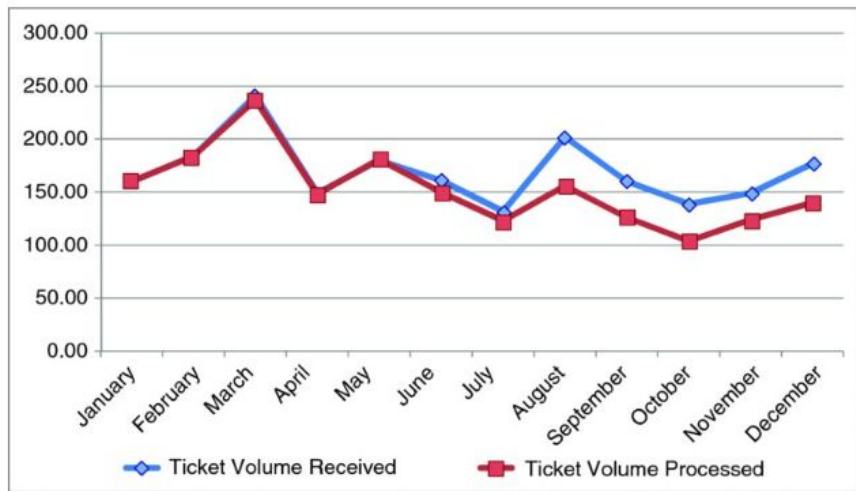


Clean up axis labels

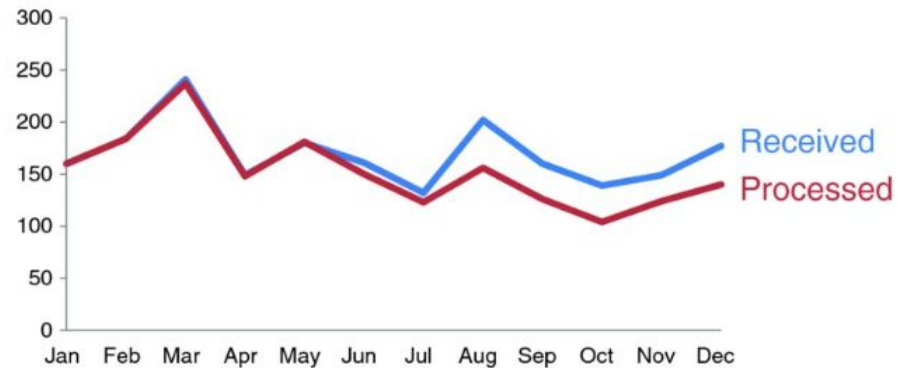


Label data directly





BEFORE



AFTER

Example

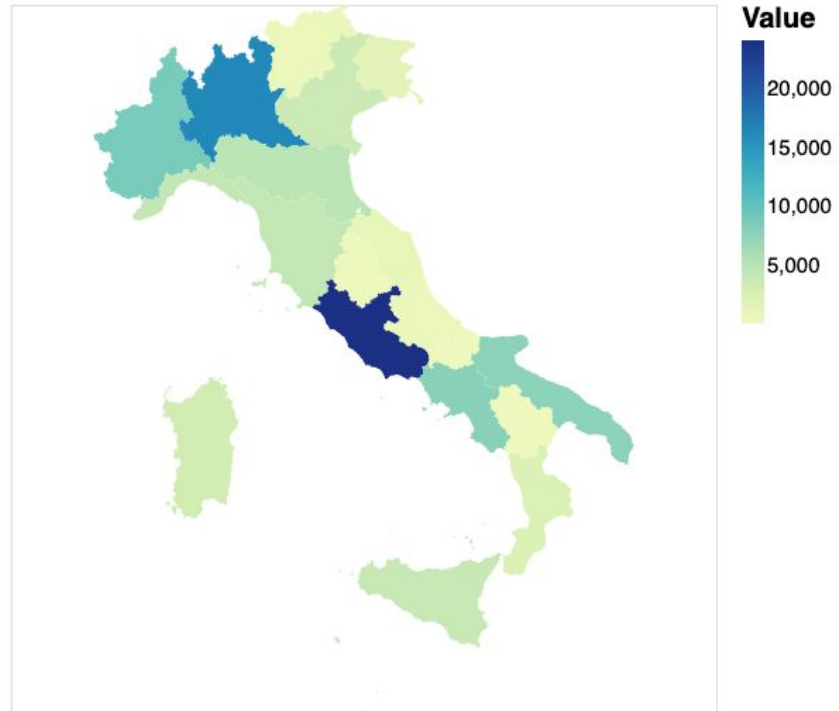
Let us suppose you work for a humanitarian organization that wants to apply for funding from the Italian government to help the homeless in Italy. The organization would like to apply for funding for the four regions with greater need. Your boss asks you to prepare a summary graph explaining the choice of the four regions.

Question to answer

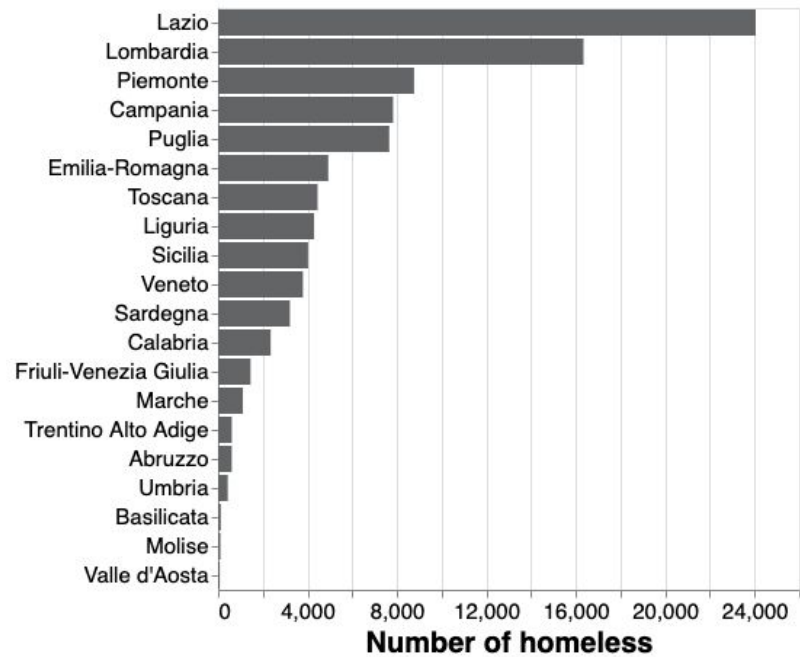
Which are the four regions for which we must ask for funding?

Territory	Sex	Age	Citizenship	Value
Piemonte	M	TOTAL	ITL	4218
Piemonte	F	TOTAL	ITL	1496
Piemonte	T	TOTAL	ITL	5714
Valle d'Aosta	M	TOTAL	ITL	41
Valle d'Aosta	F	TOTAL	ITL	17
...				

Homeless in Italy in 2021



Homeless in Italy in 2021



Number of homeless in a population of 1,000 in 2021

