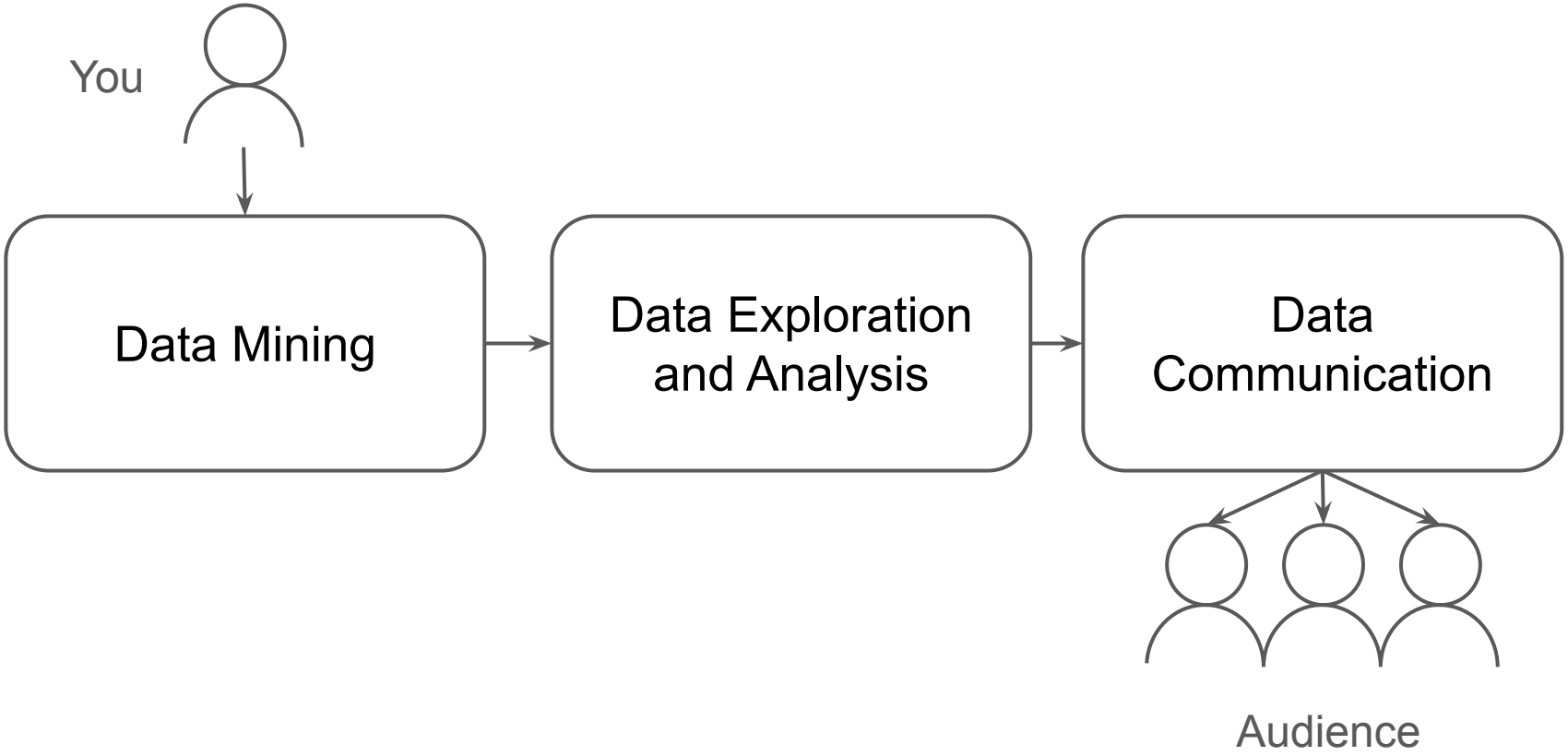
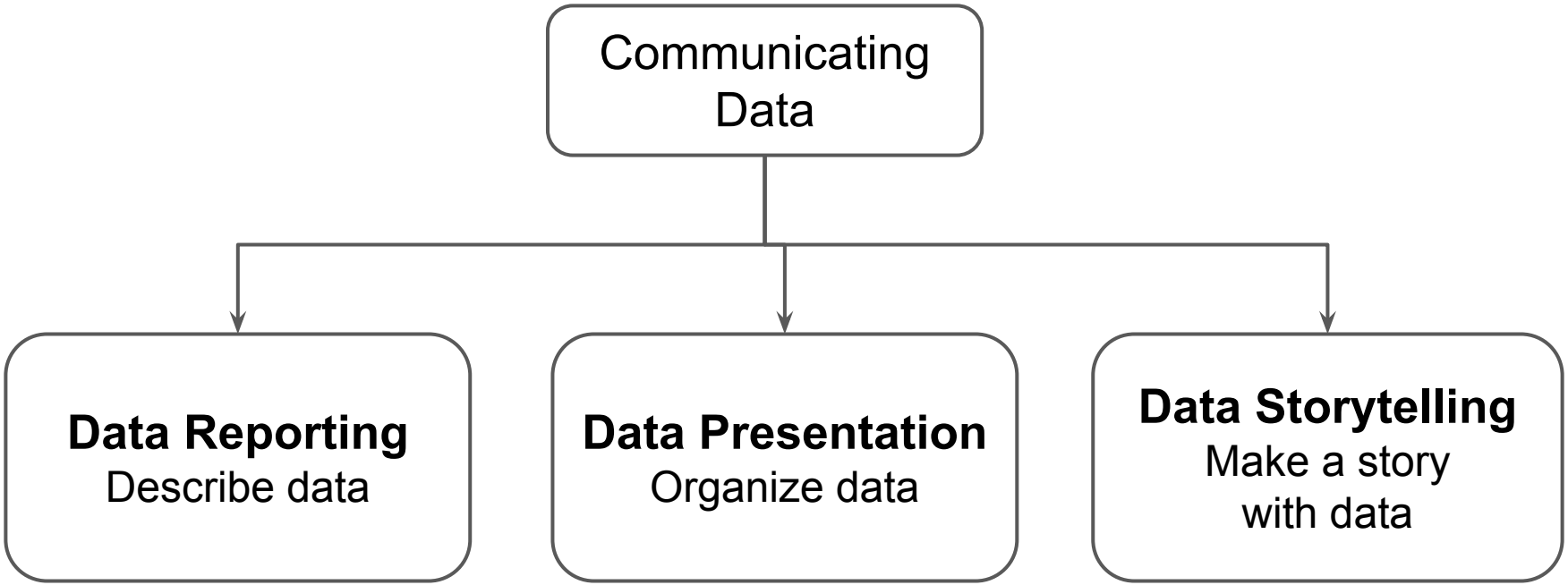
Data Storytelling in Python Altair

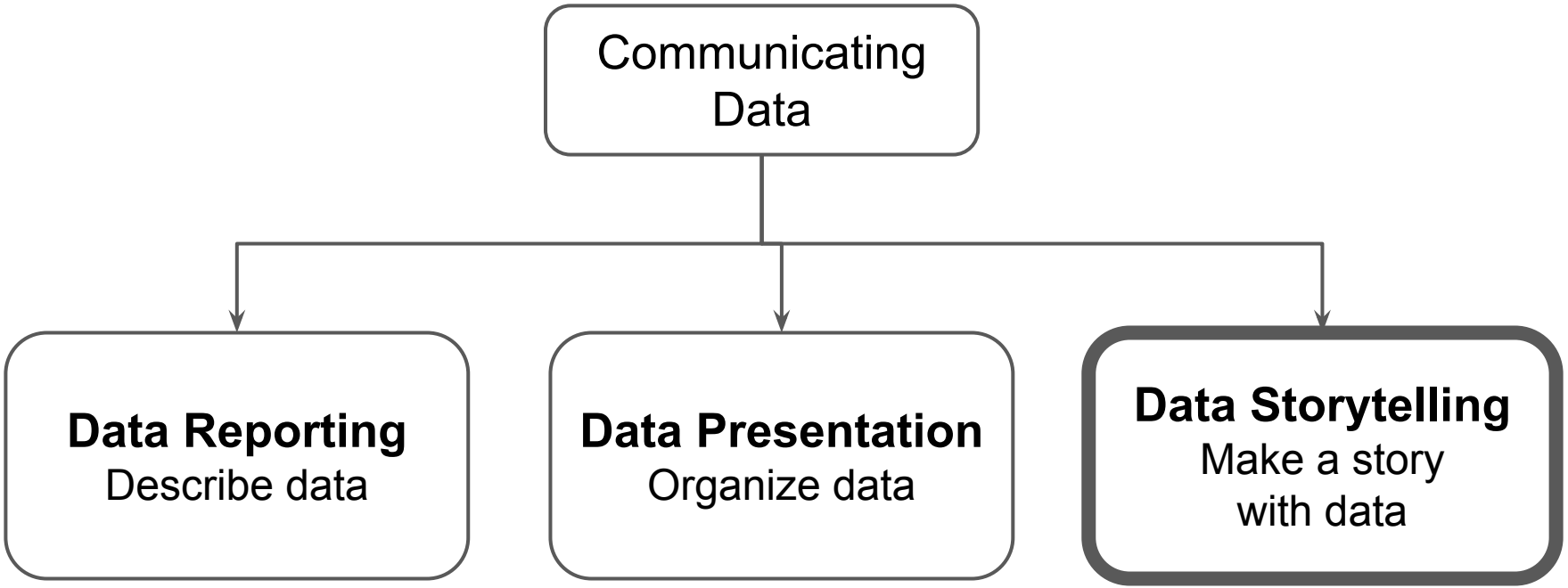
Angelica Lo Duca, PhD
Researcher @ IIT-CNR

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

- **Researcher** at the Institute of Informatics and Telematics of the National Research Council in Italy
  - Data Science, with a focus on Data Storytelling
  - Web Applications
  - Data Engineering
- **Adjunct Professor** of Data Journalism at the University of Pisa, Italy

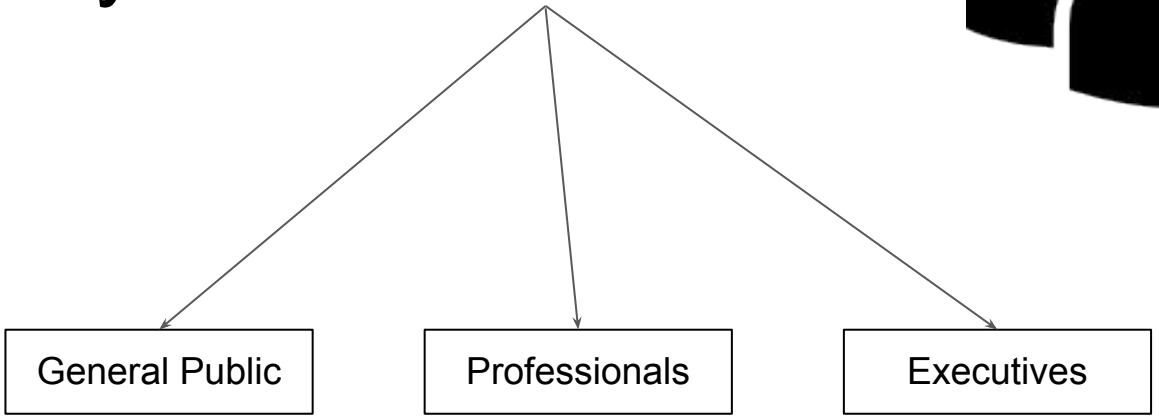

You
Data Mining
Data Exploration and Analysis
Data Communication
Audience


Communicating Data
Data Reporting
Describe data
Data Presentation
Organize data
Data Storytelling
Make a story with data

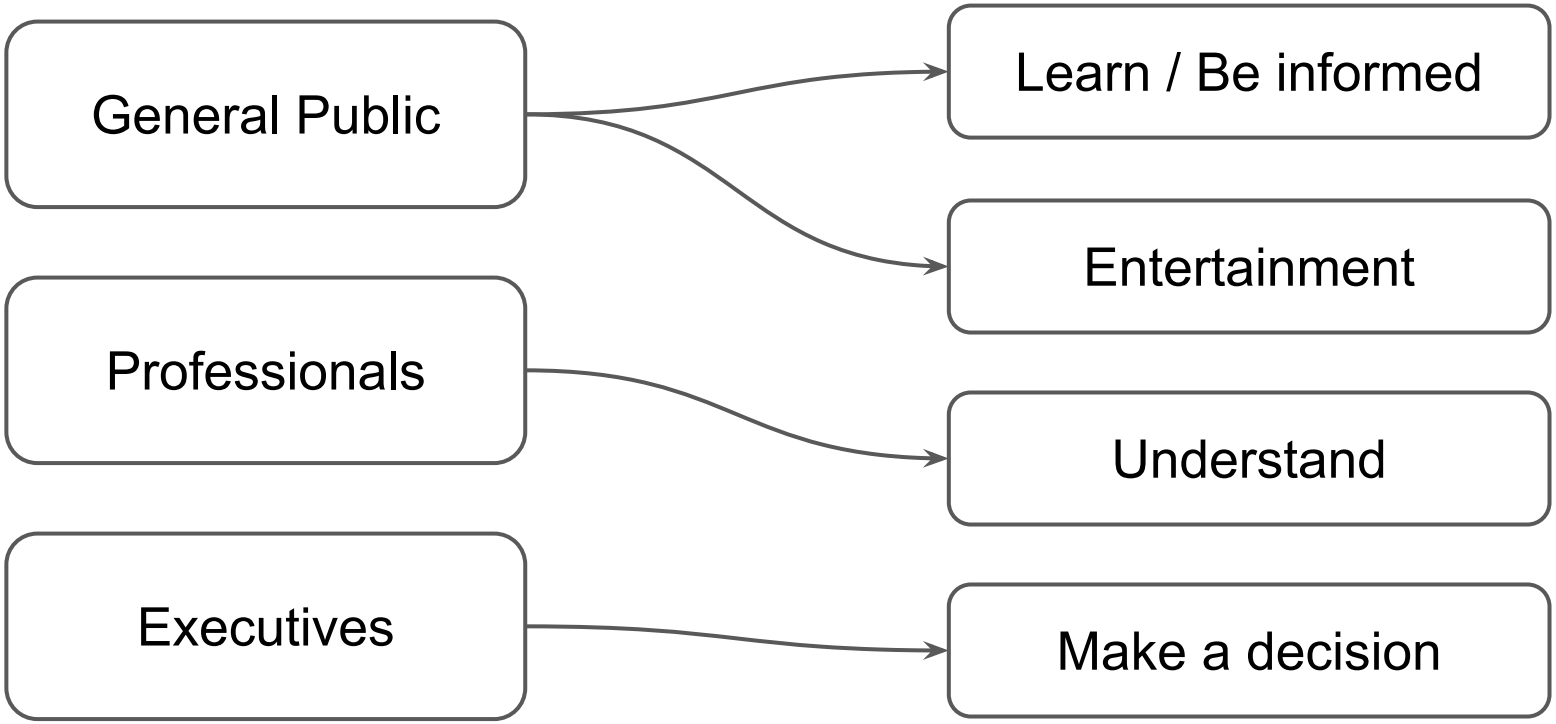

Communicating Data
Data Reporting
Describe data
Data Presentation
Organize data
Data Storytelling
Make a story with data

Data Storytelling is communicating the results of a data analysis process to an audience through a story.

You always tell a story to an **audience**



Each audience has a different **goal**



Based on your audience, you will choose an appropriate

Language and Tone

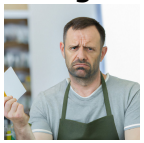
The set of words (language) and the emotional expression conveyed through them (tone)

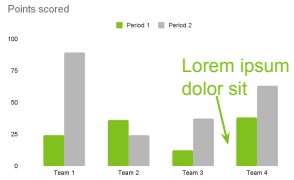
Context

The level of details to add to your story, based on the cultural sensitivity of the audience

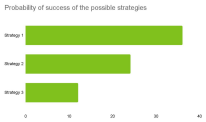
The Data-Information-Knowledge-Wisdom (DIKW) Pyramid

Background

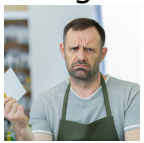


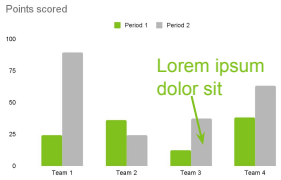


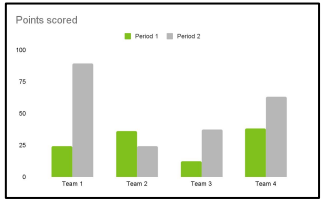
Next Steps



Background







Python Altair

Altair

The Vega-Altair library (Altair, for short) is a declarative Python library for statistical visualization based on the Vega-Lite visualization grammars.


Altair
Python Interface
Vega-lite
JSON

Altair parameters

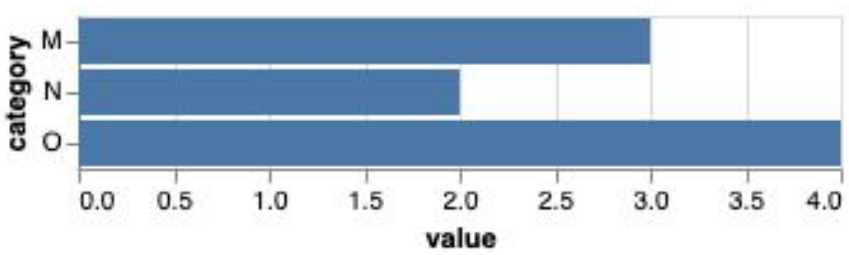
- **Marks**: define the type of chart we want to build (e.g. bar chart, line chart, …)
- **Channels**: visual properties to represent, such as axes, colors, size, …
- **Encodings**: mapping of channels to data columns in the DataFrame

```
pip install altair
```

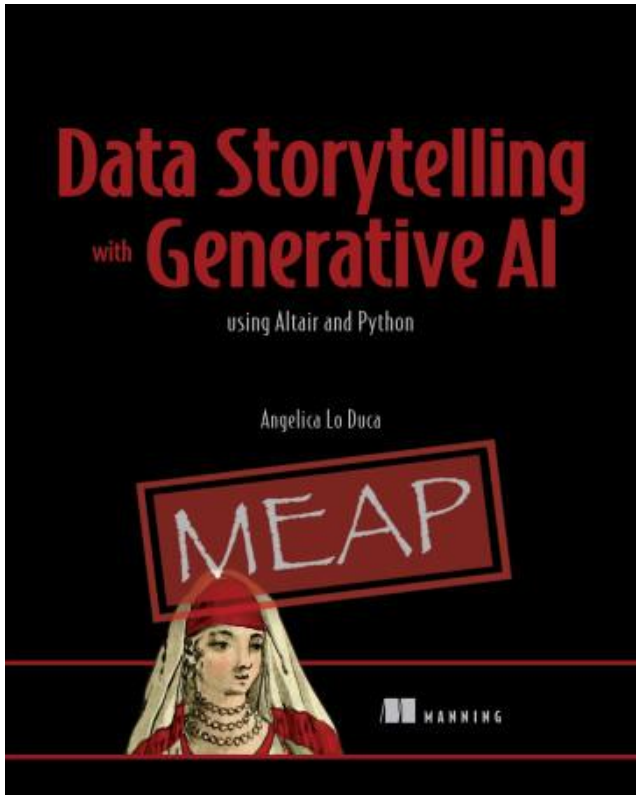
A first example

```
import pandas as pd
import altair as alt

df = pd.DataFrame({
    'value' : [3,2,4],
    'category' : ['M','N','O']
})
chart = alt.Chart(df
).mark_bar(
).encode(
     x = 'value:Q',
     y = 'category:N'
)
chart
```



References



Discount code:

datald45 (45% off all Manning products, valid through August 20, 2024)

Case Study

https://github.com/alod83/data-storytelling-in-python-altair-workshop