Selecting the right data

DATA COMMUNICATION CONCEPTS
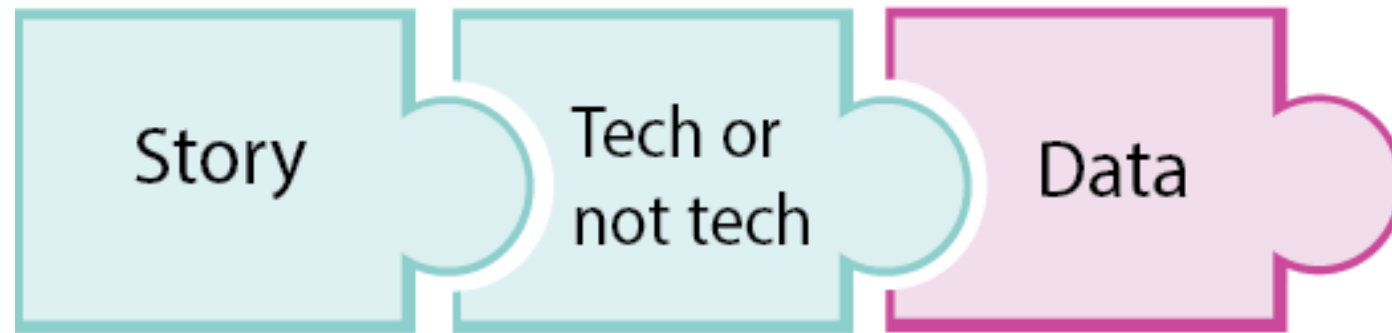
Hadrien Lacroix
Curriculum Manager

Chapter 2

How to prepare for communicating data?

- Identify **personas**

- Selecting **findings** and **statistics**

- Selecting a **visualization**

- Choose **format**
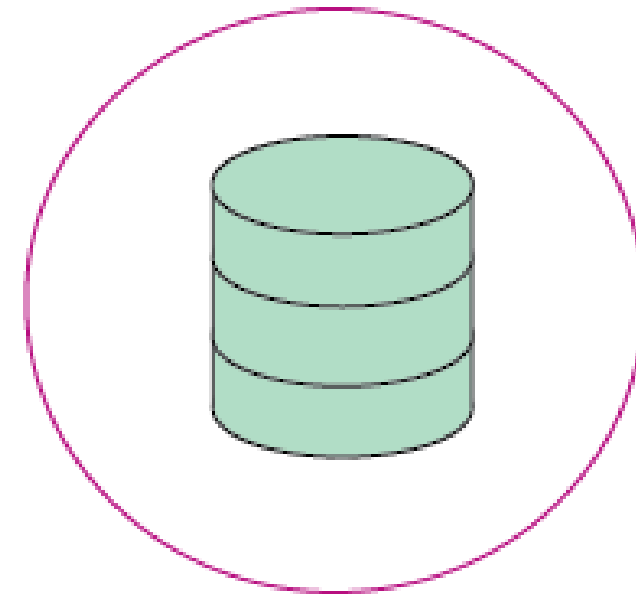
Data storytelling road
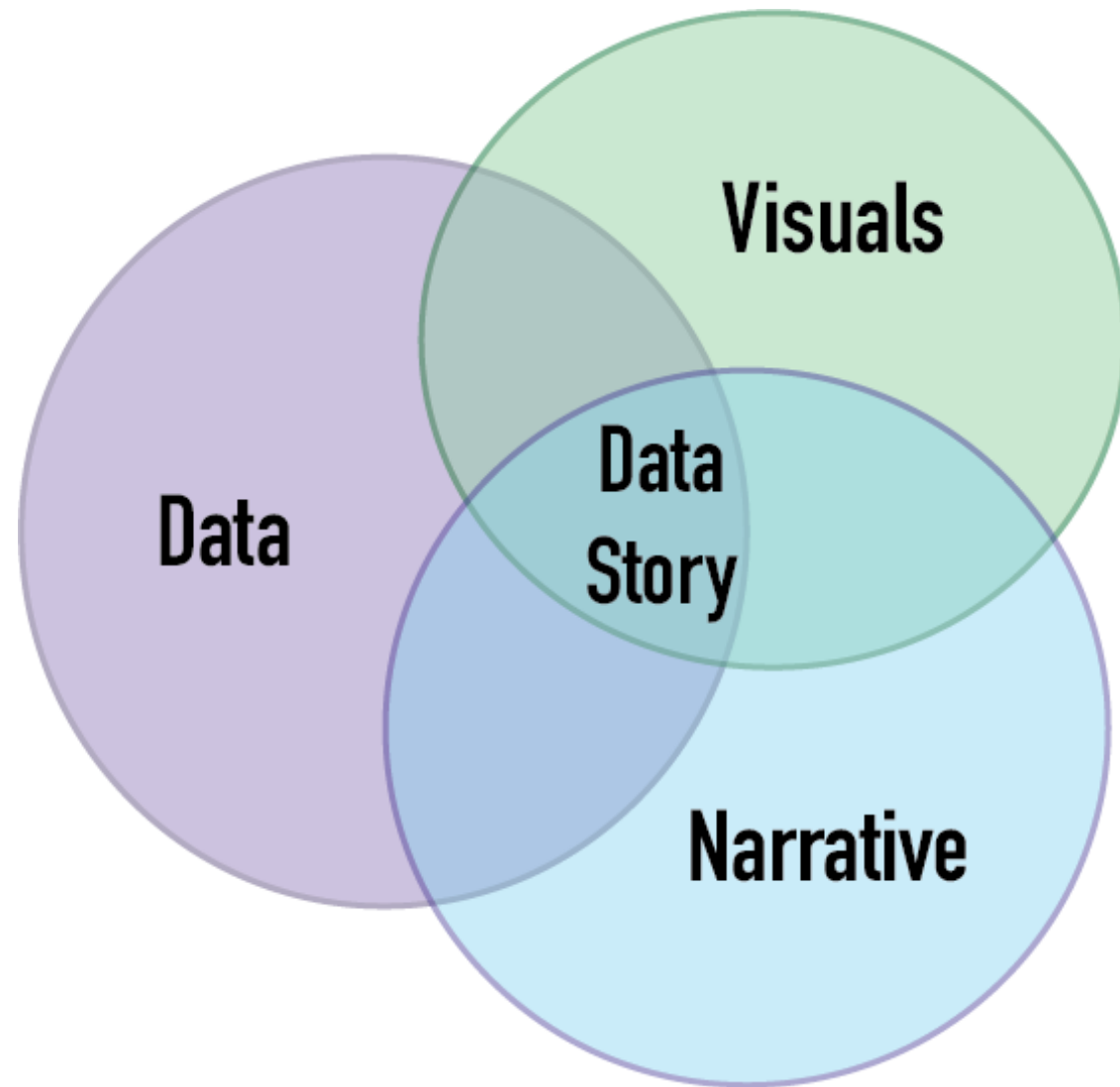
Data storytelling road

The right data

> *Selecting the right data implies including enough contextual insights in a story to better support the main point without overloading with information.*

==> Minimal amount of information to support our story

[1] Dykes, Brent. Effective Data Storytelling. Wiley.

Data storytelling

1. **Data:**
   - Garbage in, garbage out
   - Adapt to the audience
2. Narrative
3. Visuals

Stakeholders

Any person interested in the project outcome or a decision or activity derived from it.

- Technical

- Non-technical

Identifying personas

- Description
  - Interests
  - Knowledge
- Select tailored findings

Identifying personas

Food project:

- Identify personas

- Select right findings

Executive team

- **Role:** Executive level (CEO, investor, director, founder)

- **Knowledge:** Fundamentals (technical aspects)

- **Interest:** Inform their decisions based on findings

Project manager

- **Role**: Project manager

- **Interest**: Project aligns with company objectives

- **Right data**:
  - Summary data: $2M cost of marketing campaign
  - Metrics:
    - 10% monthly increase in number of customers
    - 2% risk of declining profits

Tech team

- **Role:**
  - Project collaborator
  - Technical supervisor

- **Knowledge:** Expert (Technical aspects)

- **Interest:**
  - Replicate project
  - Continue project

General audience

- **Role:**
  - Customer (external)

  - Other department staff (internal)

- **Knowledge:** Novice or generalist

- **Interests:**
  - To understand the general results and impact of the project

General audience

- **Role:**
  - Other department staff (internal)

- **Interests:**
  - To understand the general results and impact of the project

- **Right data:**
  - Historical data: Decline in profits

  - Correlation/impact:
    - Chocolate needs rebranding

    - Impact next year earnings

Audience skepticism

- Different levels of skepticism

- Different levels of argumentation
    - Convince yourself

    - Convince a friend

    - Convince a skeptic
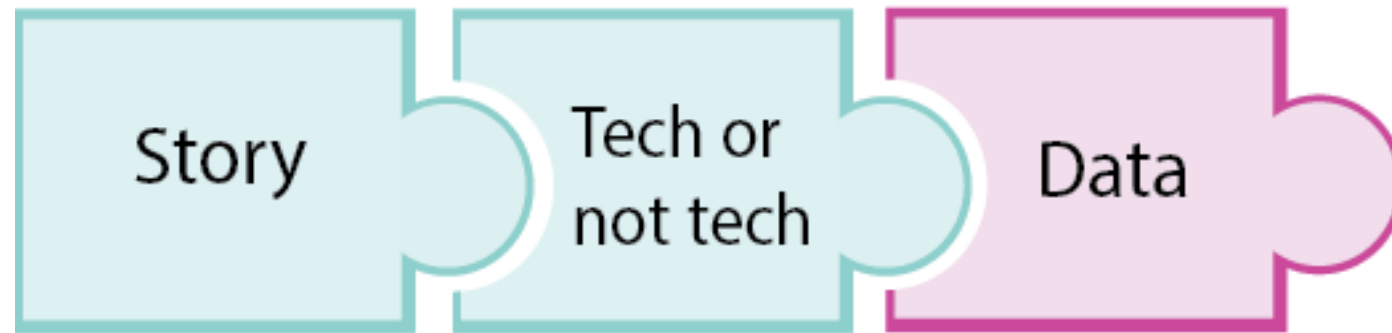
Let's practice!

DATA COMMUNICATION CONCEPTS

Showing relevant statistics

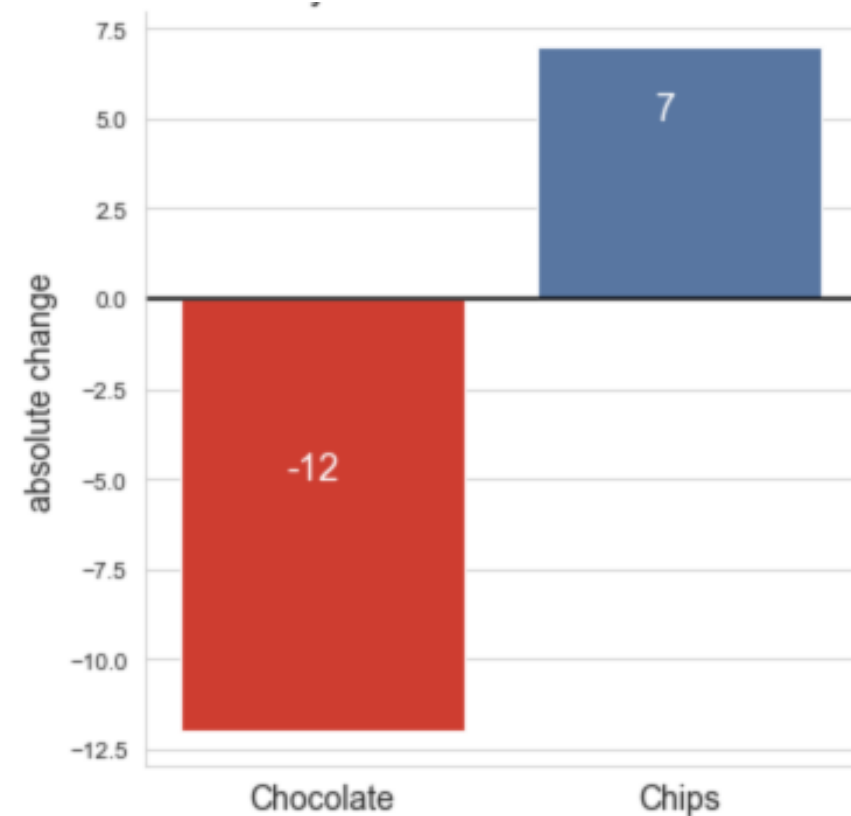## DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

datacamp

Data storytelling road
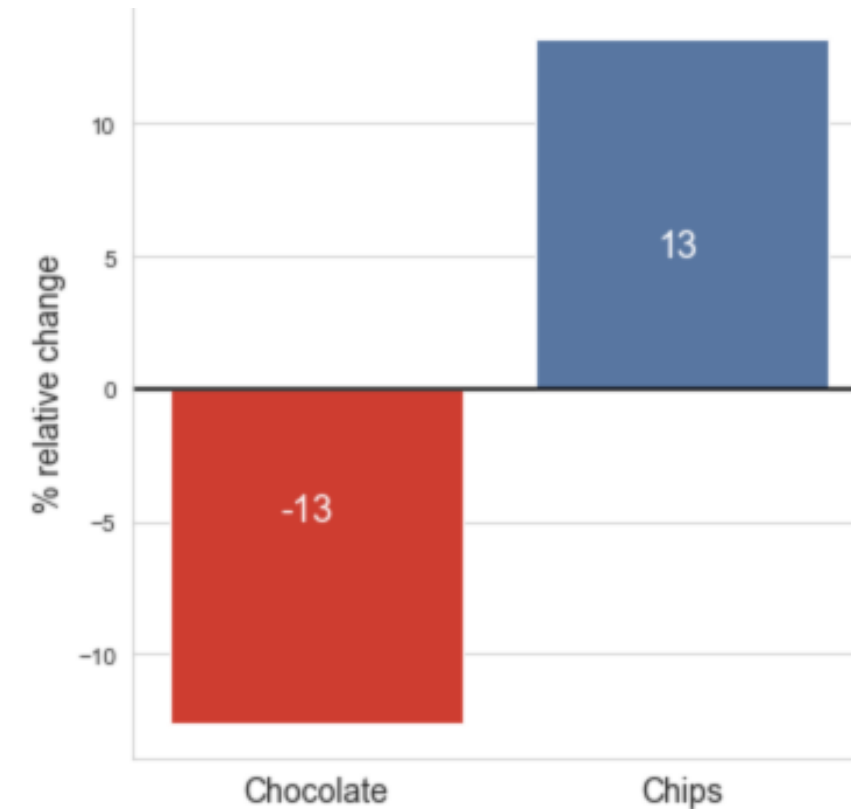
Variations of data

- **Absolute**
  - Difference between 2018 and 2017 sales
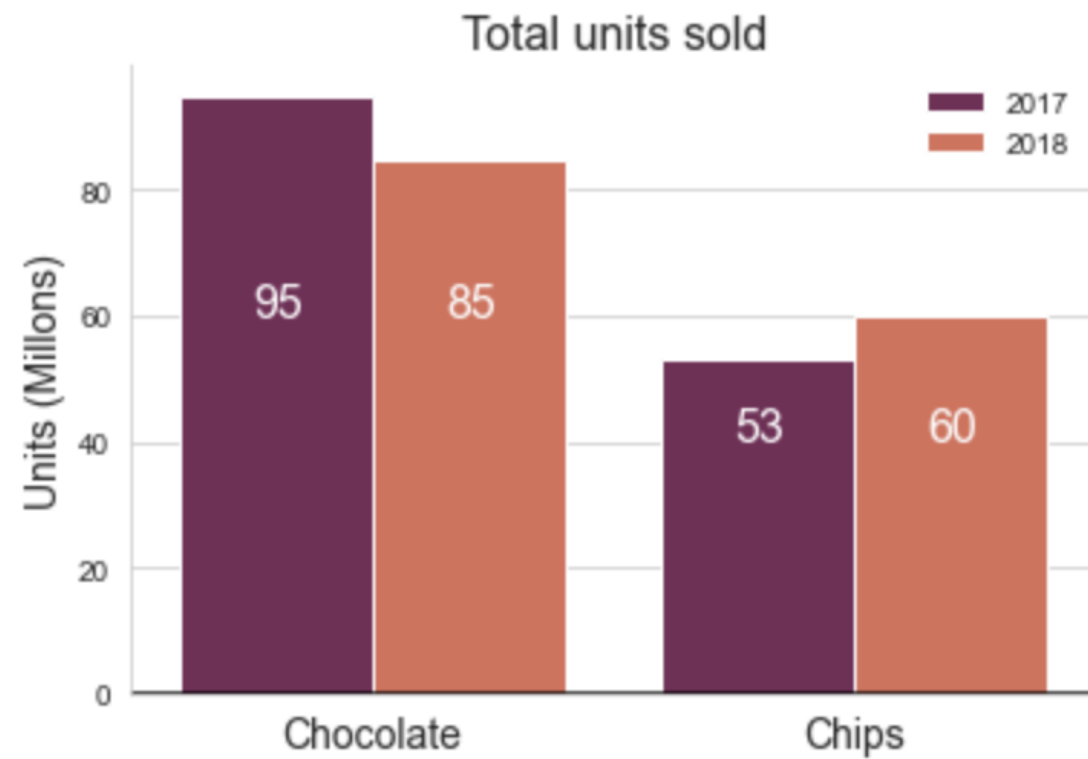 - Absolute change and relative change depend on the quantity

- **Relative**
  - Percentage variation 2018 from 2017
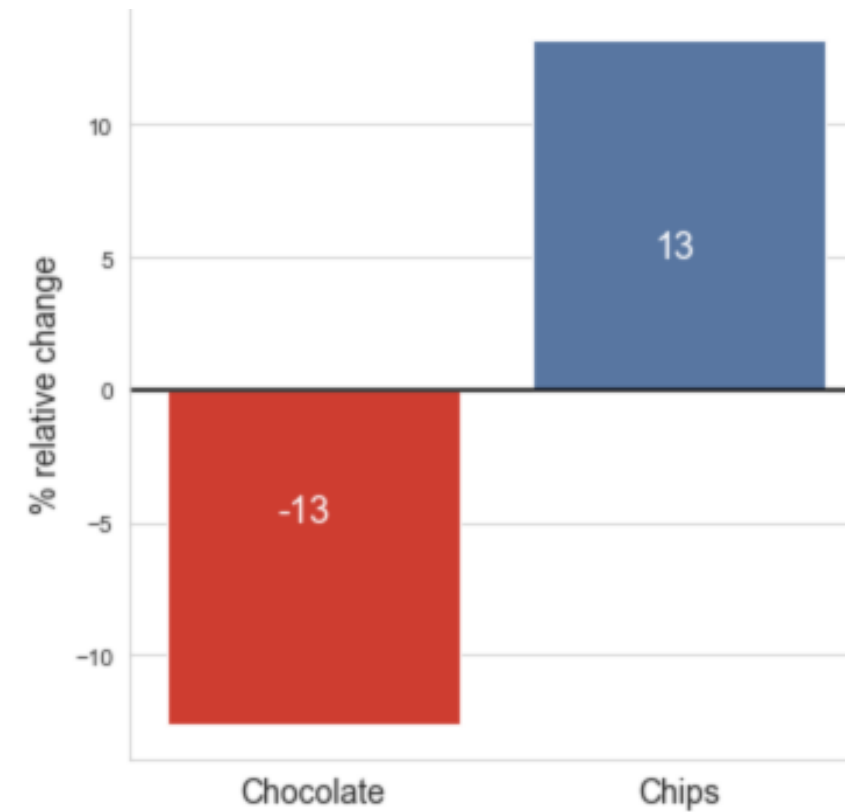 - Small numbers more significant than reality
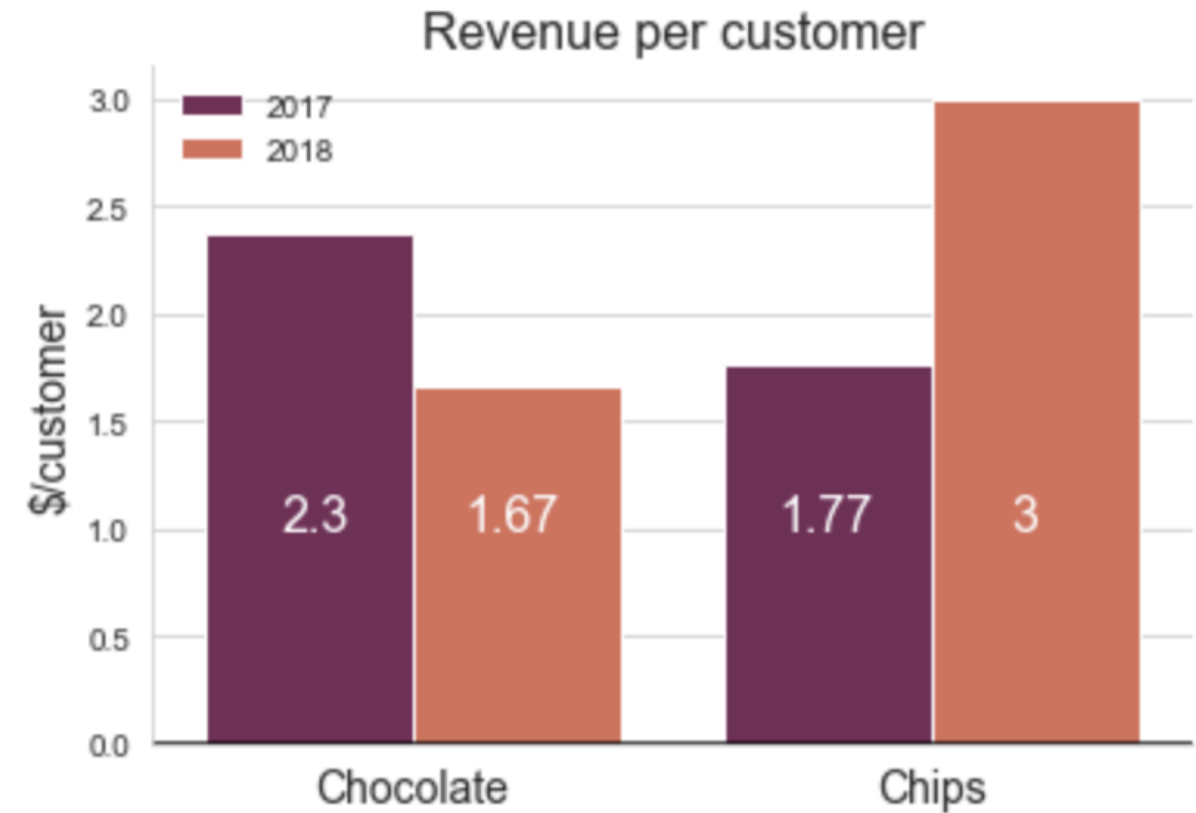
Variations of data

Absolute

Total units sold

Legend: 2017, 2018

Chocolate: 95 (2017), 85 (2018)
Chips: 53 (2017), 60 (2018)

Y-axis: Units (Millons)

Relative

% relative change

Chocolate: -13
Chips: 13

Ratio

- Quotient of two variables
  - Revenue per customer (**total product revenue/number customers**)

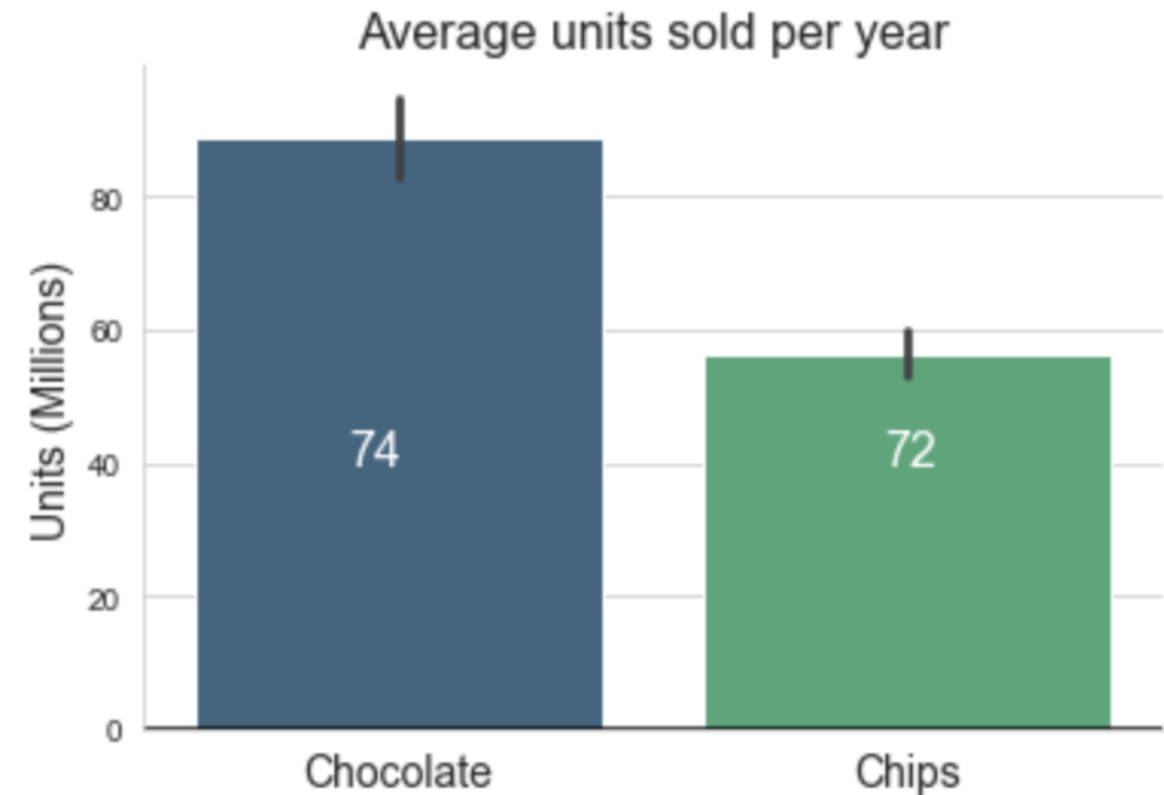- Normalize values = **better comparisons**

Aggregates

- Representative value:
  - Totals / counts

Aggregates

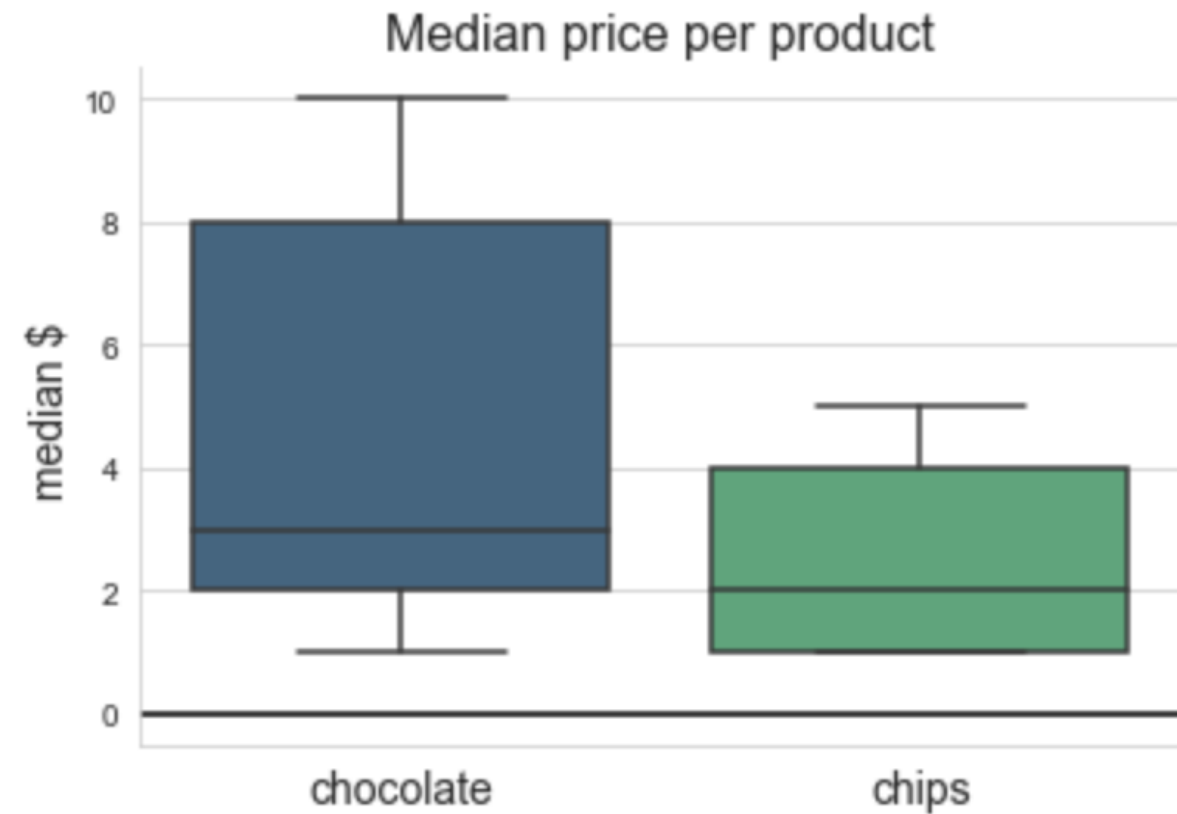- Representative value:
  - Totals / counts

  - Mean

- Average units sold per year

Aggregates

- Representative value:
  - Totals / counts
  - Mean
  - Median

- Median price

Median price per product

Aggregates

- Representative value:
  - Totals / counts

  - Mean

  - Median

- Mean can be misleading (outlier)

- Distribution of the data

- Example:
  - 2019 US **average** salary: **$51**,916.27

  - 2019 US **median** salary: **$34**,248.45

p-value

What is p-value?

- Convention:
  - Value less than 0.05: statistical significance

  - Values close to 0.05: weak indicator

What is it not?

- Not proof of evidence:
  - Reject our hypothesis, but not that is false

- Consider alternatives or complementary metrics

[1] P-Values: Misunderstood and Misused. Front. Phys. 2016.

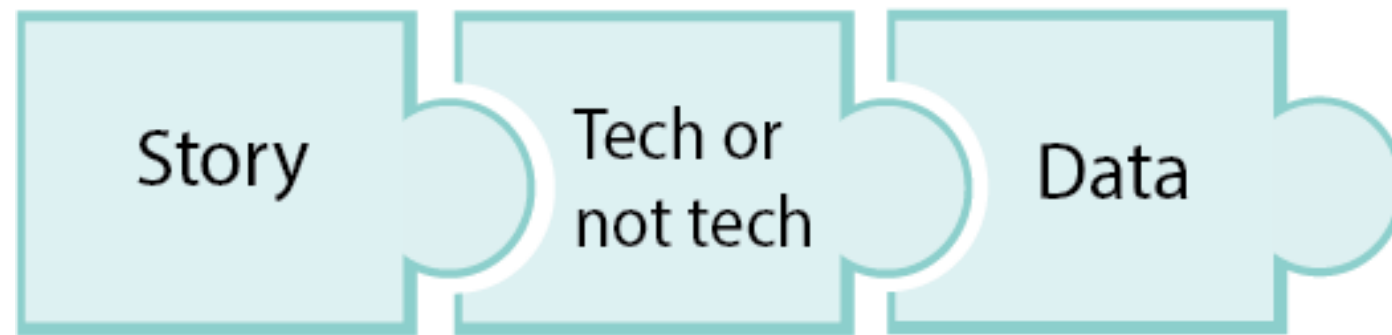# Let's practice!

DATA COMMUNICATION CONCEPTS

Visualizations for different audiences
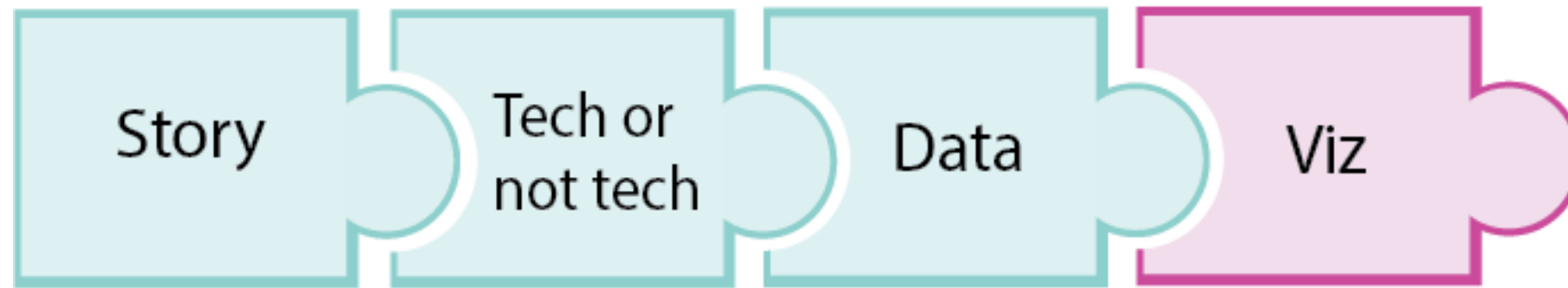
DATA COMMUNICATION CONCEPTS

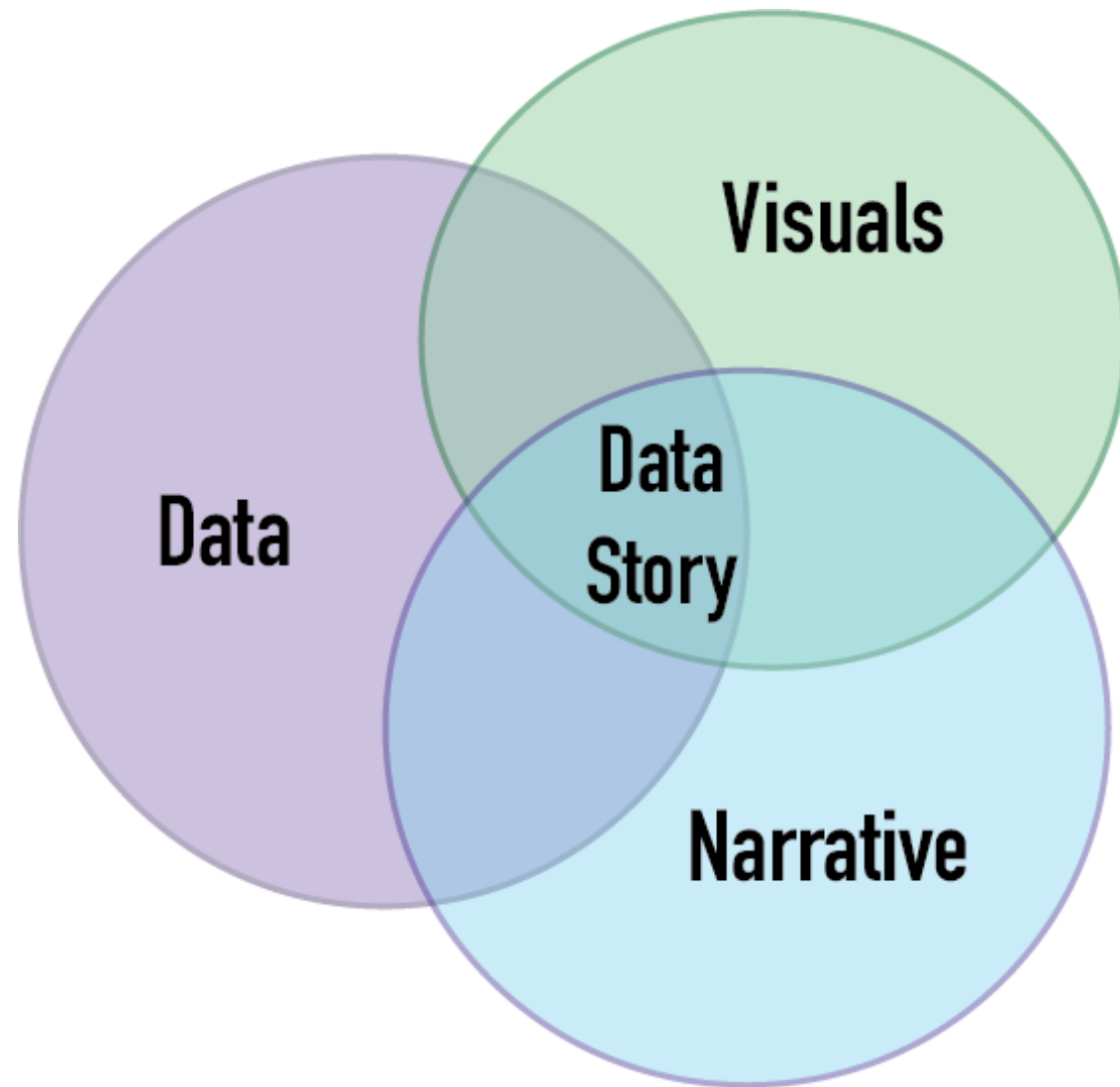**Hadrien Lacroix**
Curriculum Manager

Communication strategy

Communication strategy

Data storytelling

1. Data

2. Narrative

3. **Visuals**
   - Expertise
   - Familiarity

Tailored message

- **Investor**

Marketing campaign avoid 20% revenue drop

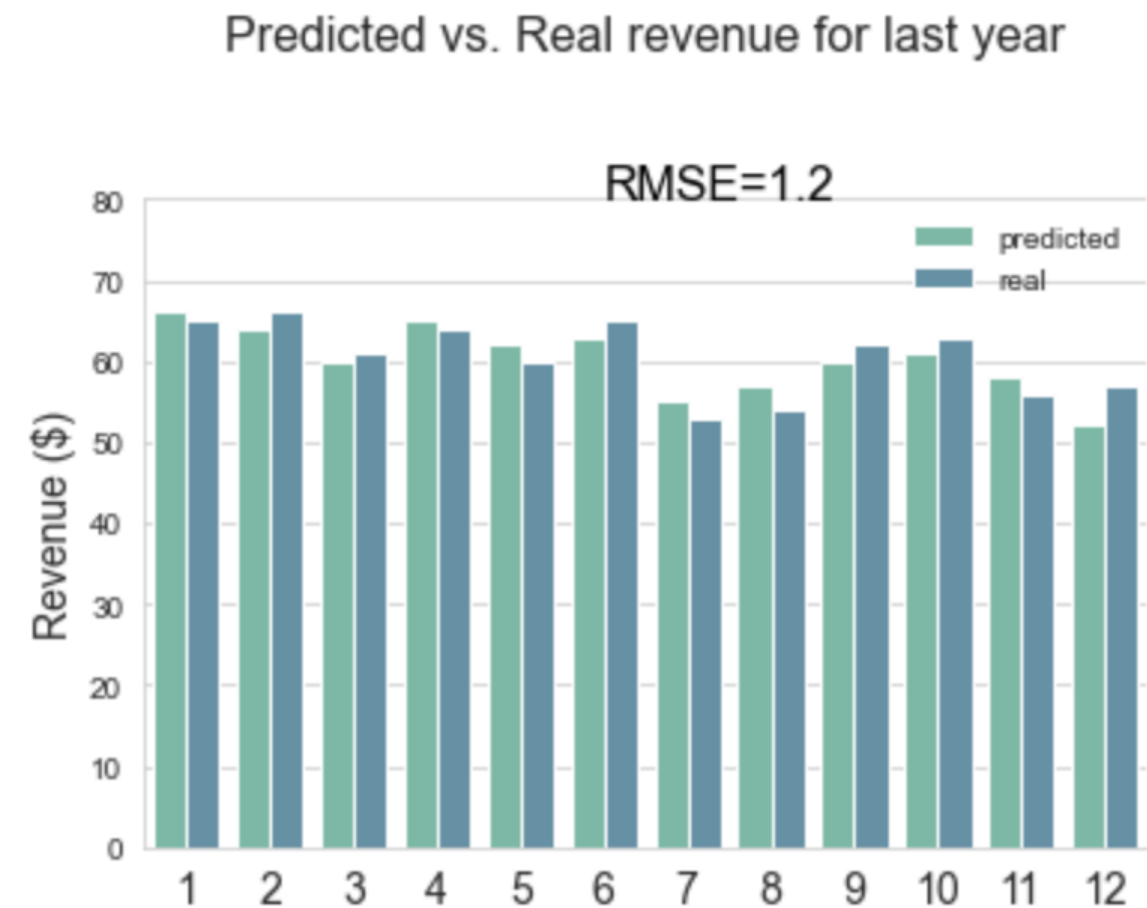- **Technical lead**

Model predicted revenue decline (RMSE: 1.2)
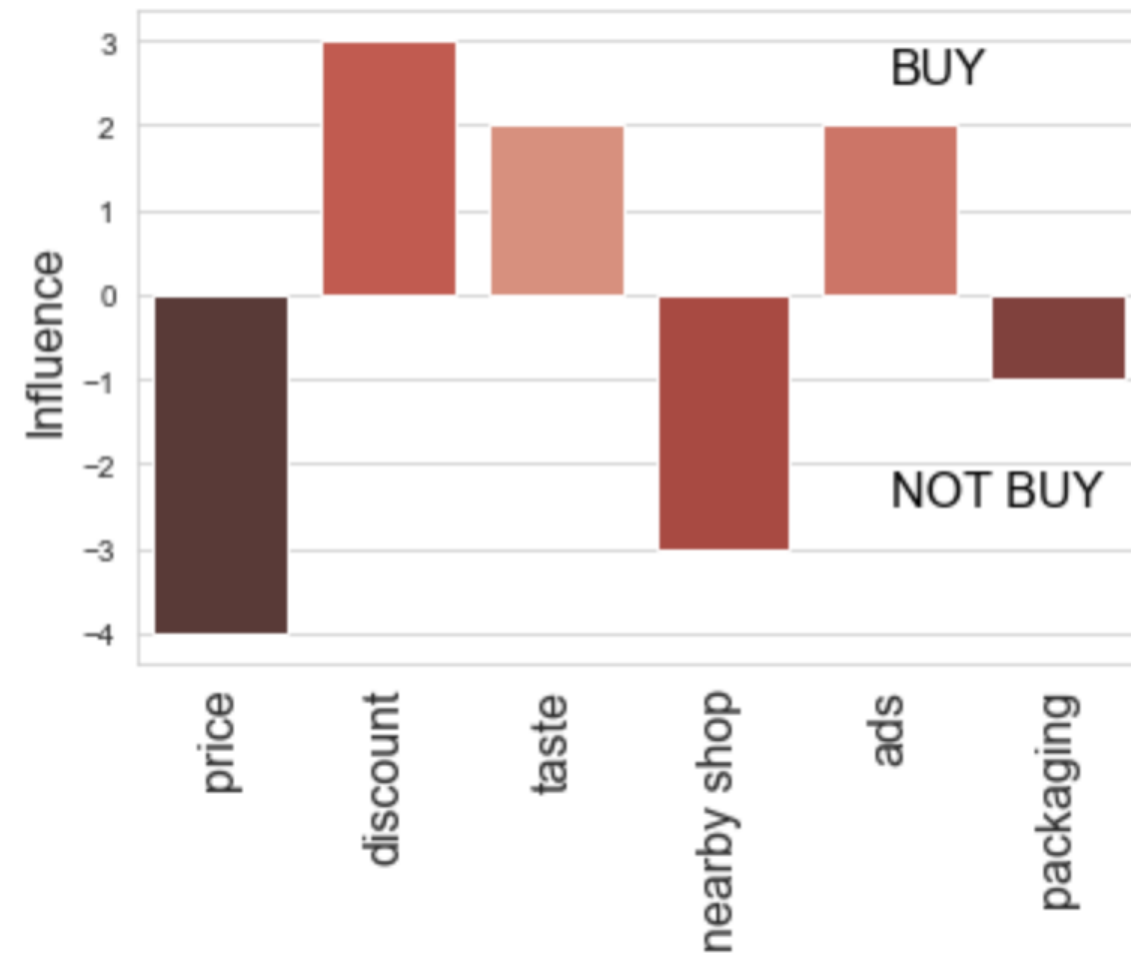
Directly linked to message
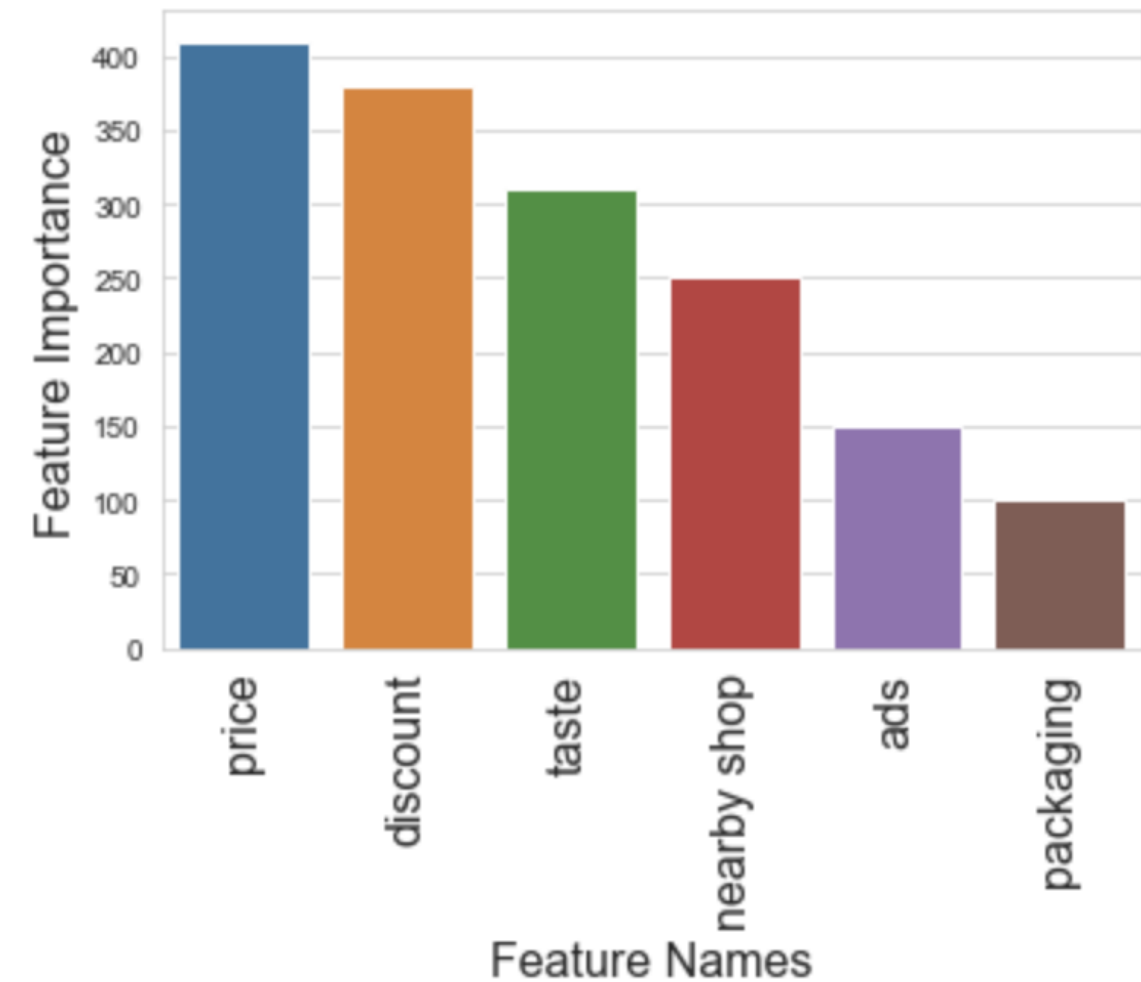
- Investor

- Technical lead

Provide context

- Investor
- Technical lead

Influence of different factors on customer behavior
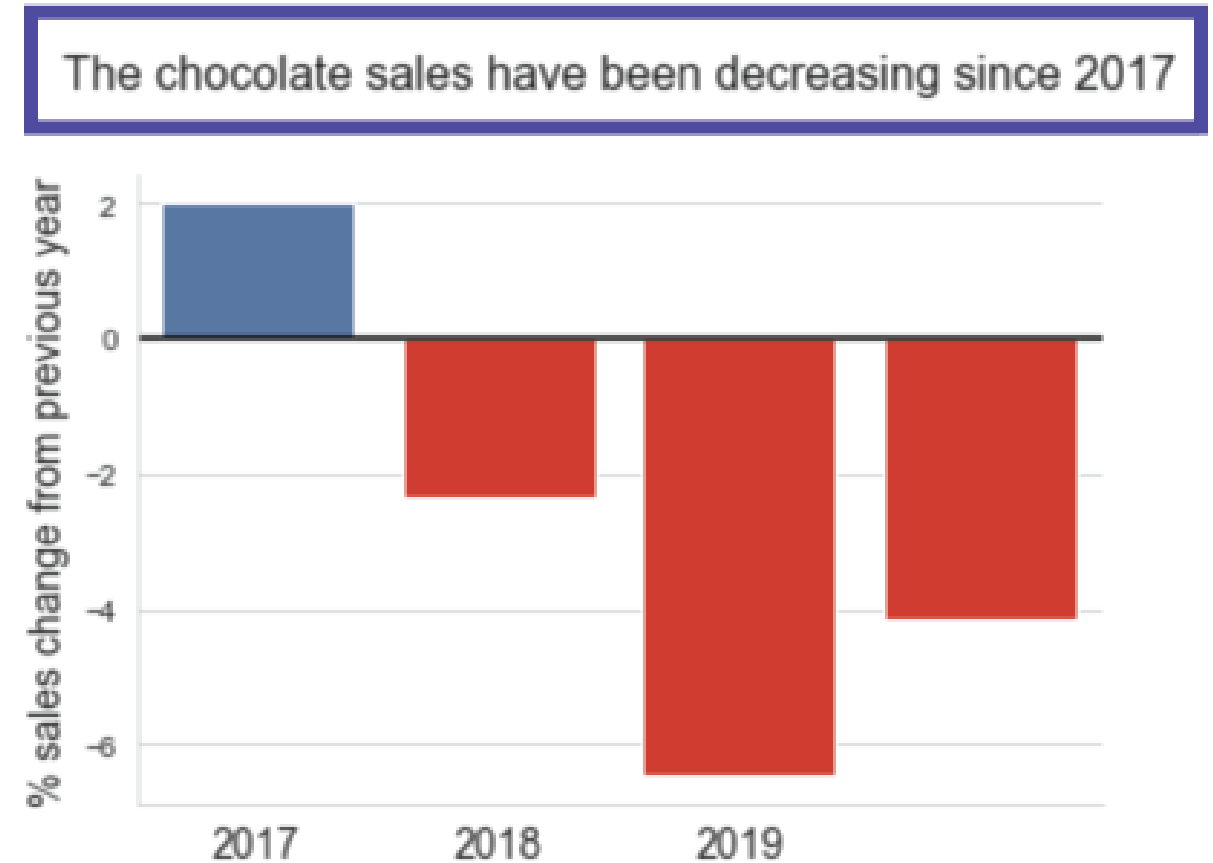

Feature Importance

More best practices

- Pareto principle:
  - Aggregate less relevant data

  - Include chocolate, chips and other products (aggregated)

- Approachable and engaging visuals

- How many / how quickly

- Less is more

McCandless method

1. **Introduce visualization by name**
   - Graph headline
   - Clear and obvious
   - y vs x technique

The chocolate sales have been decreasing since 2017

1 https://artscience.blog/home/the-mccandless-method-of-data-presentation
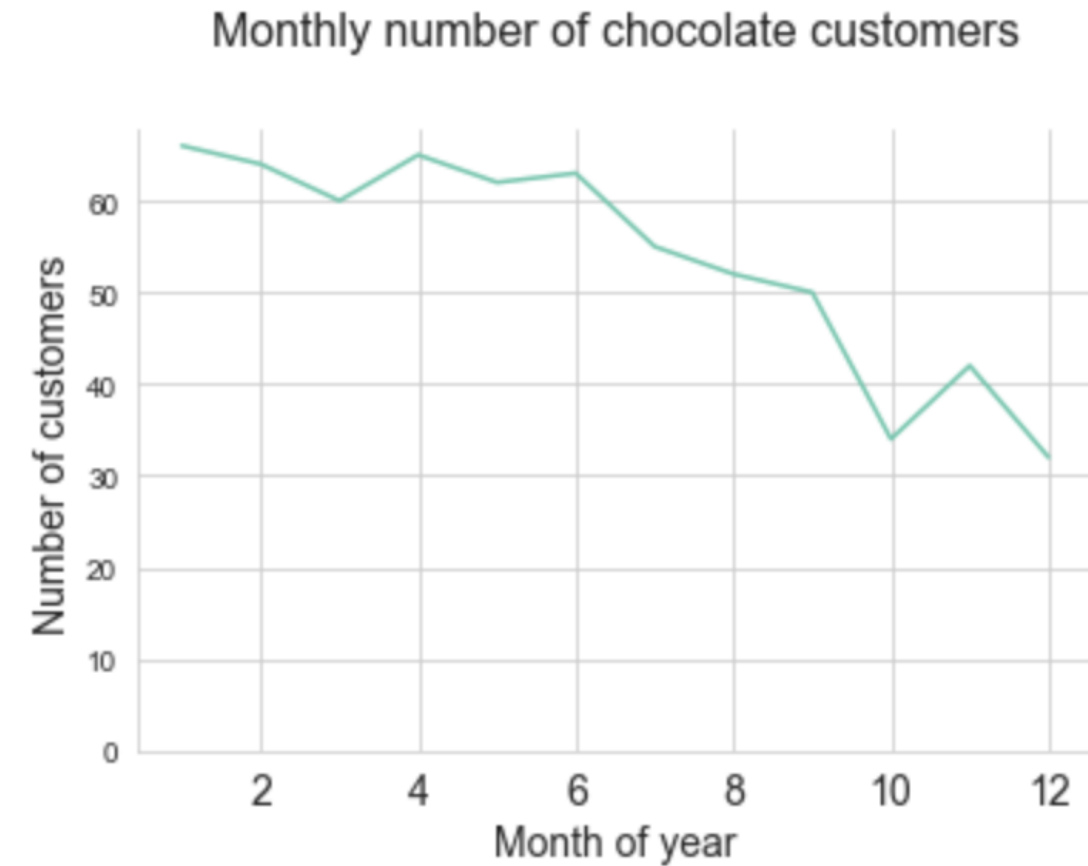
DATA COMMUNICATION CONCEPTS

McCandless method

1. Introduce visualization by name

2. **Anticipate audience's questions**
   - Focus on story not on decoding graph

McCandless method

1. Introduce visualization by name

2. Answer audience's questions

3. **State insights**

Monthly number of chocolate customers

McCandless method

1. Introduce visualization by name

2. Answer audience's questions

3. State insights

4. **Help the audience relate**

   ○ Importance

   ○ Action items

Let's practice!

DATA COMMUNICATION CONCEPTS

Choosing the appropriate format

DATA COMMUNICATION CONCEPTS
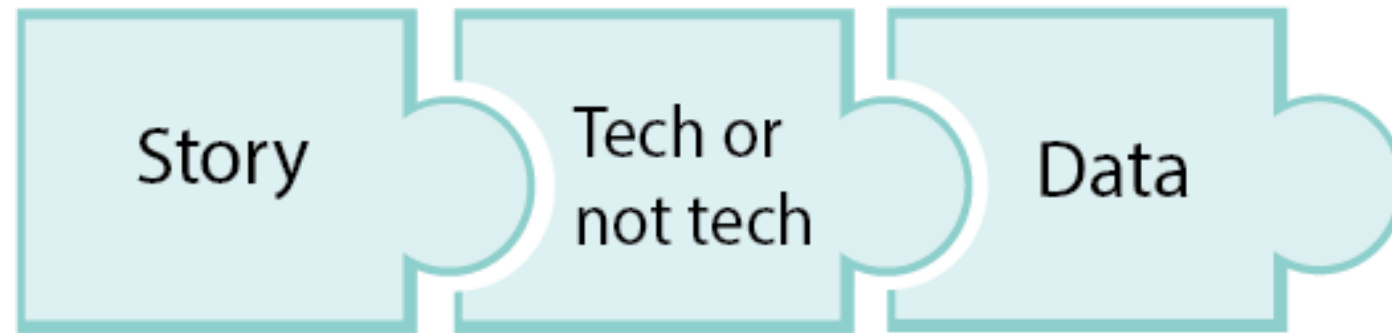
Hadrien Lacroix
Curriculum Manager

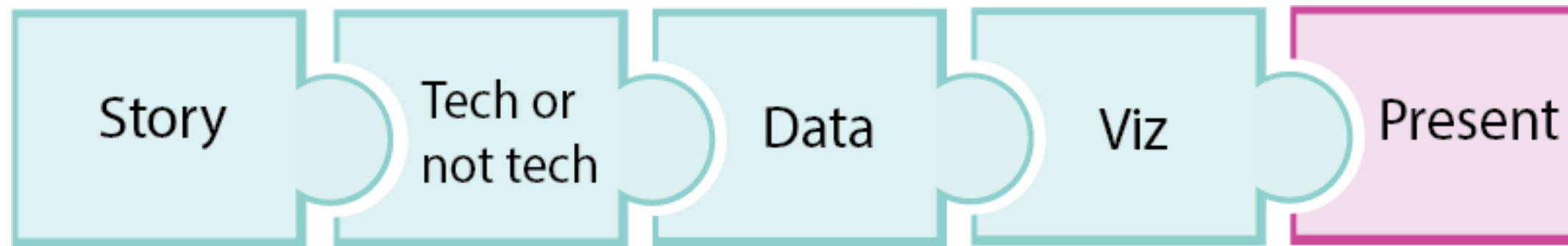# Data storytelling road

Story

Data storytelling road

Data storytelling road

Data storytelling road

Story | Tech or not tech | Data | Viz

Data storytelling road

Story — Tech or not tech — Data — Viz — Present

Which format is more effective?

Good communication format

- Key information

- Engaging

- Easy to understand

Which format is more effective?

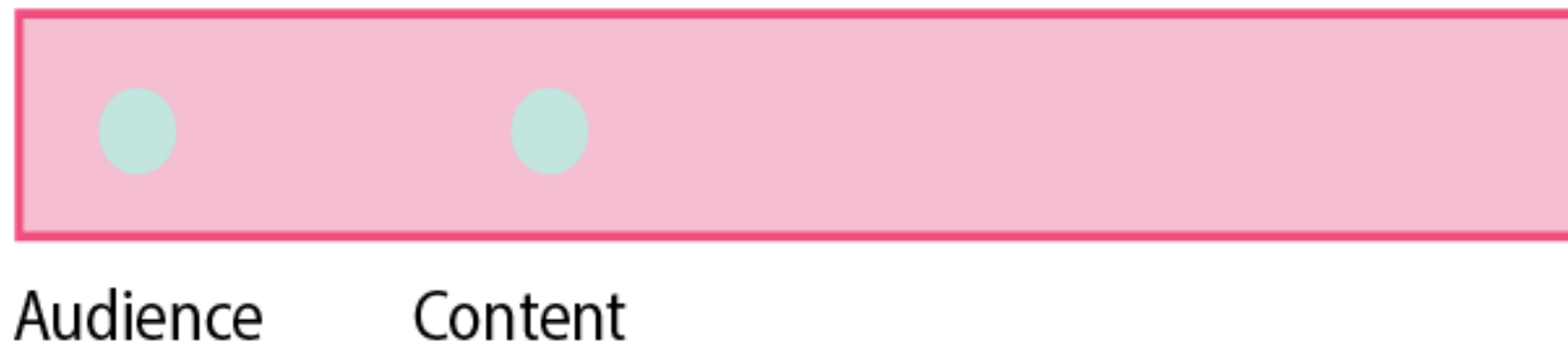1. Written reports

2. Oral presentations

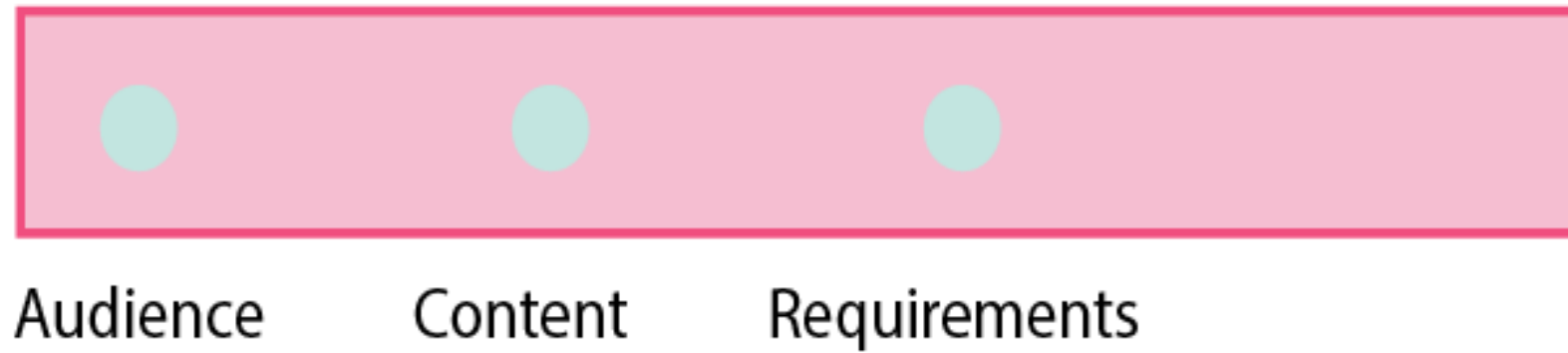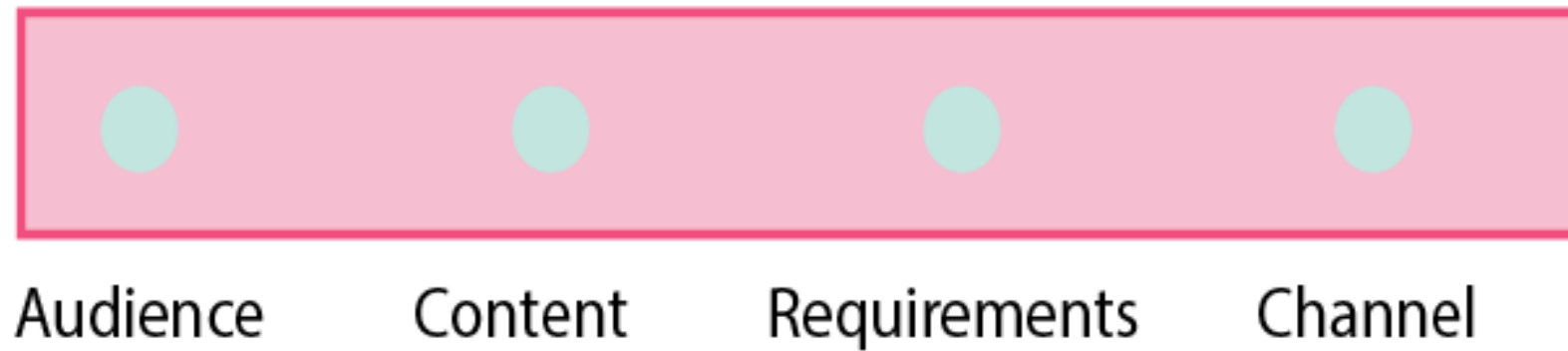# Presentation strategy

Audience

Presentation strategy

Audience   Content

Presentation strategy

Audience   Content   Requirements

Presentation strategy

Audience   Content   Requirements   Channel

Presentation strategy
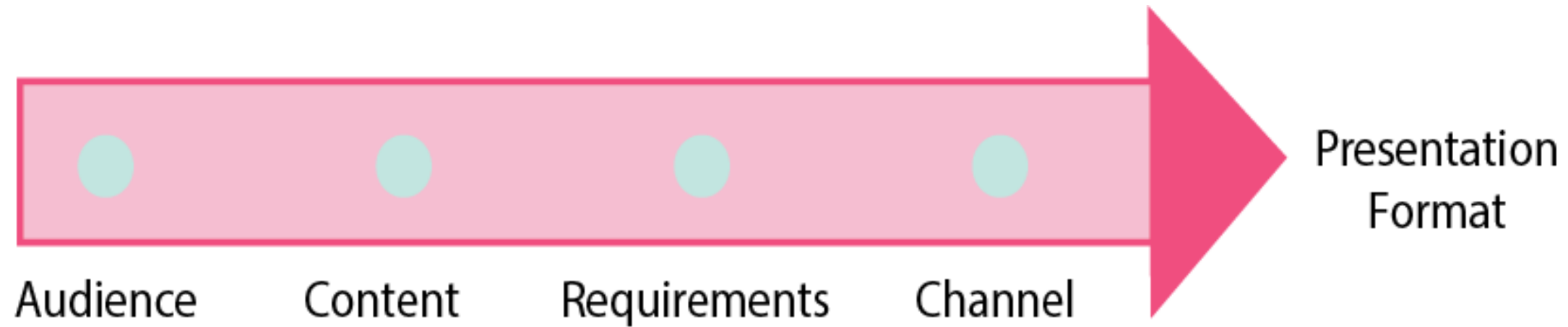
Stakeholders

- **Who?**

- **Why?**
  - Accountability
  - Methodology

- **How?**
  - Make decisions
  - Start new project

- **What?**
  - Results
  - Impact

Content

- Results?

- Conclusion?

- Recommendations?

- Methods?

[1] Photo from Unsplash

Requirements

- Time

- Authority

- Time zone

Consumption

- **Format**
  - Document

  - Slide deck

- **Delivery** mechanism
  - Live

  - Asynchronous

- **Audience**
  - Conference room

  - Ballroom

[1] Photo from Unsplash

Oral communication

Advantages

- Relationship with the audience

- Immediate feedback

- Non-verbal cues

Disadvantages

- No permanent record of communication

- Not suitable for long messages

Written communication

Advantages

- Permanent record of communication

- Shared easily with a large audience

- Less emotional reaction to message

- Suitable to share code with colleagues

Disadvantages

- Hard to see if the message was understood

- No immediate feedback

Appropriate format

Oral presentation

- Who? **CEO**

- Why? **Monthly update**

- What? **Accountability**

- Content: **Conclusions**

- Requirement: **Time**

- Channel: **Meeting**

[1] Photo from Unsplash

Appropriate format

Report

- Who? **CEO**

- Why? **Important decisions ahead**

- How? **Report to investors**

- Content: **Recommendations**

- Channel: **Email**

[1] Photo from Unsplash

Let's practice!

DATA COMMUNICATION CONCEPTS