Fundamentals of storytelling

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

datacamp

What is data storytelling?

> *Data storytelling is the practice of building a narrative around a set of data and its accompanying visualizations to help convey the meaning of that data in a powerful and compelling fashion*

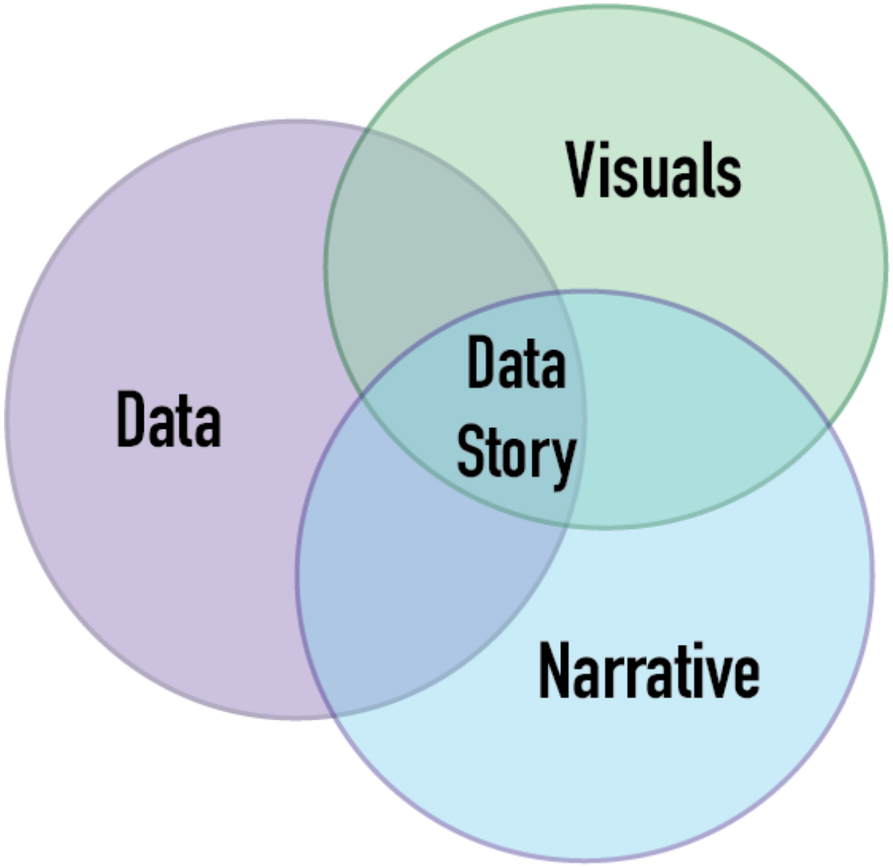
- Anecdotes = **imagination**
- Stories = **memorable**
- Add value (provide **context**)
- Capture audience's **attention**
- Facilitate **decision**-making
- Drive **change**

Benefits:

- Helps focus attention
- Meaning and context
- Helps retain insights
- Better-informed decision-making
- Persuade change-resistant stakeholders

[1] https://tdwi.org/portals/what-is-data-storytelling-definition.aspx

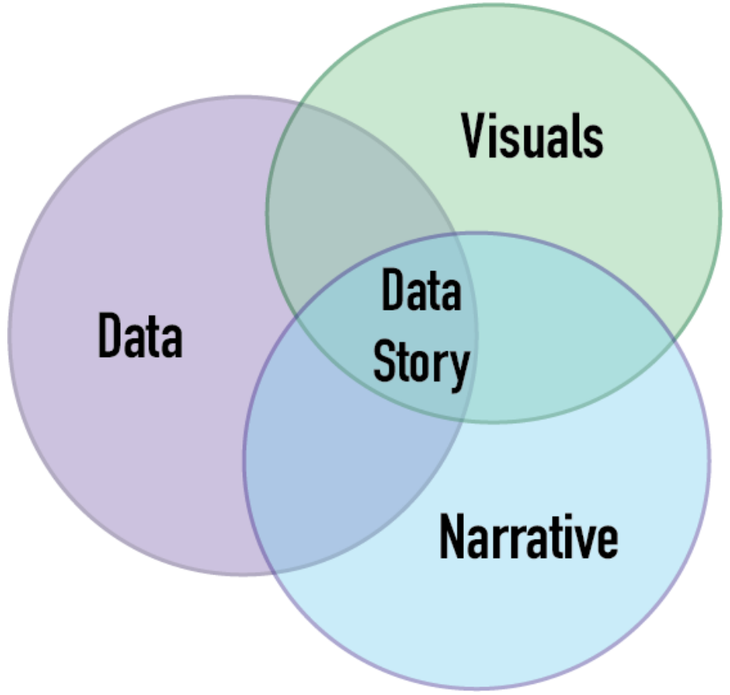
Data storytelling



Data storytelling

- **3-minutes story:**
  - What would you say in 3 minutes?
- **Big idea:**
  - Unique point of view
  - One sentence

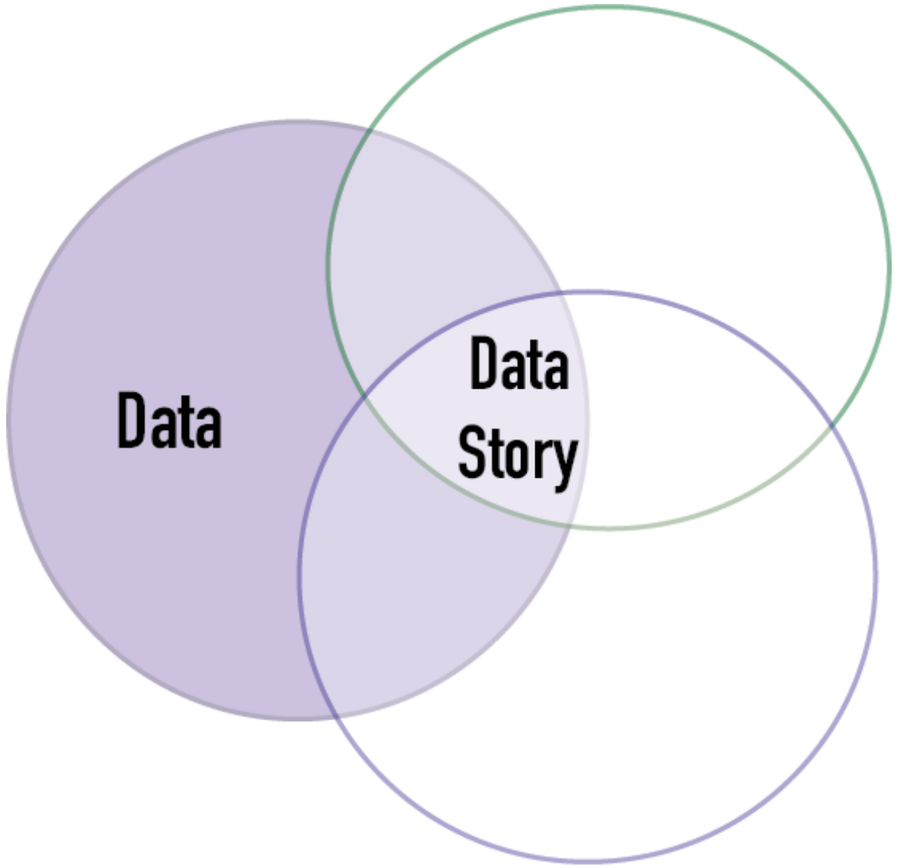
==> Clear and concise



1. **Insightful**
2. **Explanatory**
3. **Concise**

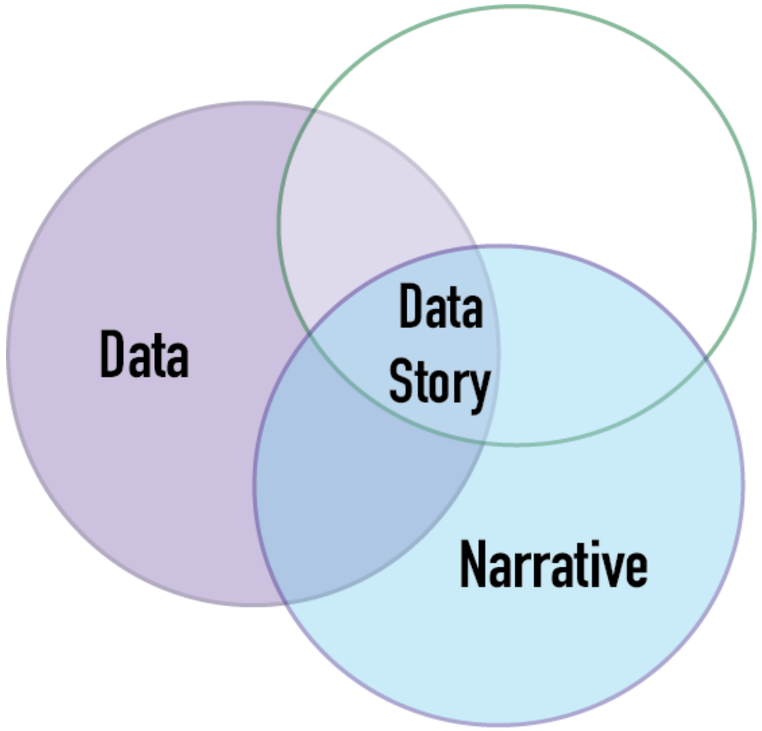
[1] Knaflic, Cole Nussbaumer. Storytelling with Data. Wiley Editorial.

Data



- **Results** (e.g predictions) and **findings** (e.g. data analysis)
- **Relevant** (methods or results and implications)
- **Accurate** and reliable
- **Actionable** insights

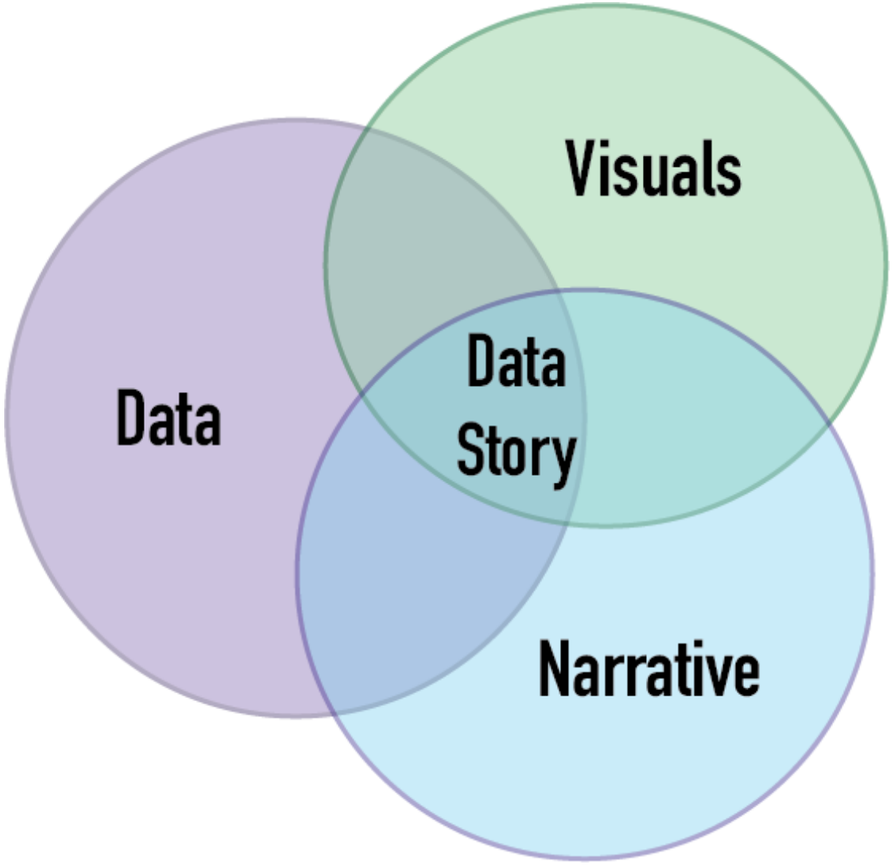
Narrative



- **Compelling** and **easy** to understand
- Prioritize **essential** points
- Drive **change**
- Main point:
  - **Avoid disconnected** facts
  - **Central** insight
- Explanatory context:
  - Understand **background** and audience
  - **Clarify** facts to that audience
- Linear sequence

> *A description of connected events that organizes information to engage the audience and make them care for the results or information shared*

Visuals



- Graphs should be:
  - simple
  - engaging
  - **not misleading**

Awareness

- **What do they know?**

How our model works

- **What do they need to understand?**

Why we chose our predictive variables

- **What level of information do they need?**

The correlation coefficients between variables

- **Adjust content**

Prediction's impact and limitations

- **Be conversational**

The context on which our model works

- **Serve audience**

The interactions between customer traits

ADEPT

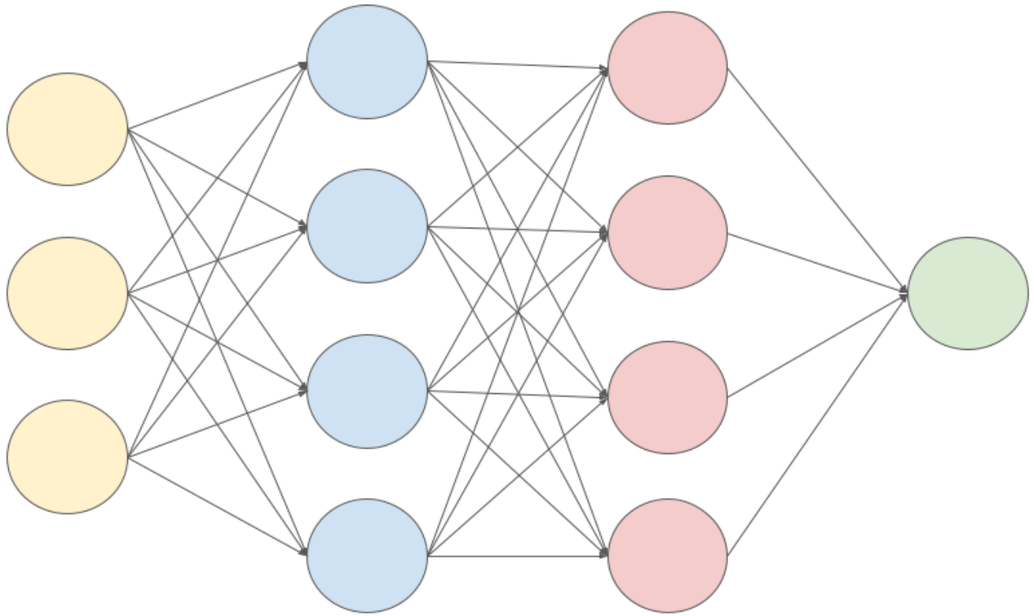
- **A**nalogy
- **D**iagram
- **E**xample
- **P**lain English
- **T**echnical definition

Technical jargon

- Use **acronyms** with caution
  - Can help or hurt communication
  - Introduce the term and acronym
- **Jargon**
  - Translate terminology
  - Simple terms
  - Guide
  - Definitions

Analogies

Instead of



Use

[1] Alpha, "Liam is an expert on the shape sorter", Creative Commons

Focus on impact

Instead of

- *Use a non-relational database to make efficient nested queries.*
- *Number of rooms shows correlation of 0.7 with a house price.*

Focus on

- *Changing the storage approach will save a lot of time.*
- *The more rooms in the house, the higher the price.*

Humility

- Be receptive
- Proactively ensure understanding
- Explain differently

Executive team

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

Project manager

- **Interest:** Project aligns with company objectives
- **Right data:**
  - Summary data: $2M cost of marketing campaign
  - Metrics

Tech team

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

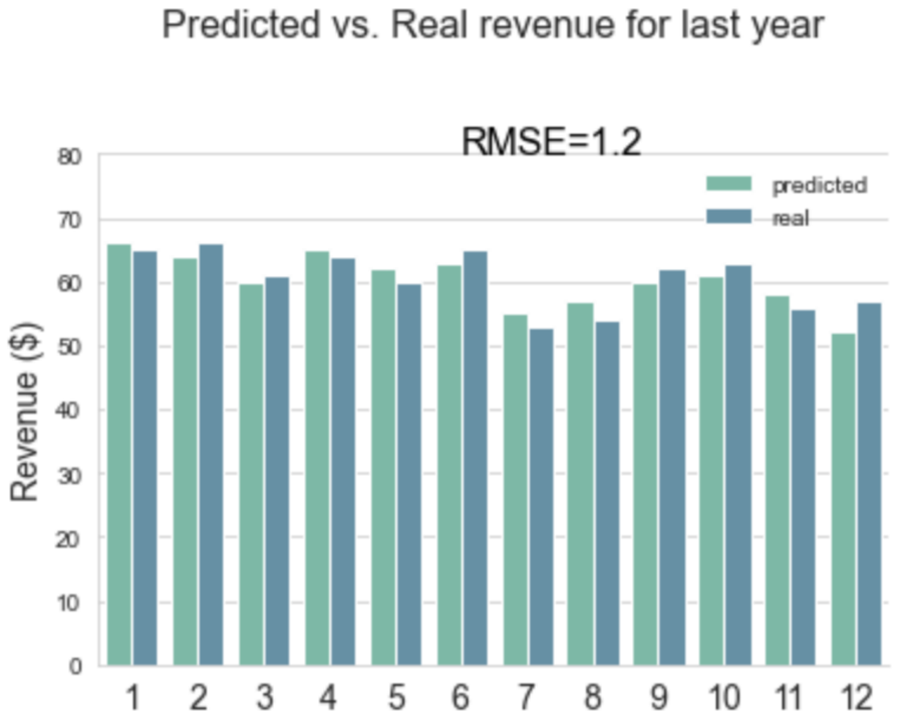
General audience

- **Role:**
  - Customer (external)
  - Other department staff (internal)
- **Knowledge:** Novice or generalist
- **Interests:**
  - To understand the general results and impact of the project

Directly linked to message

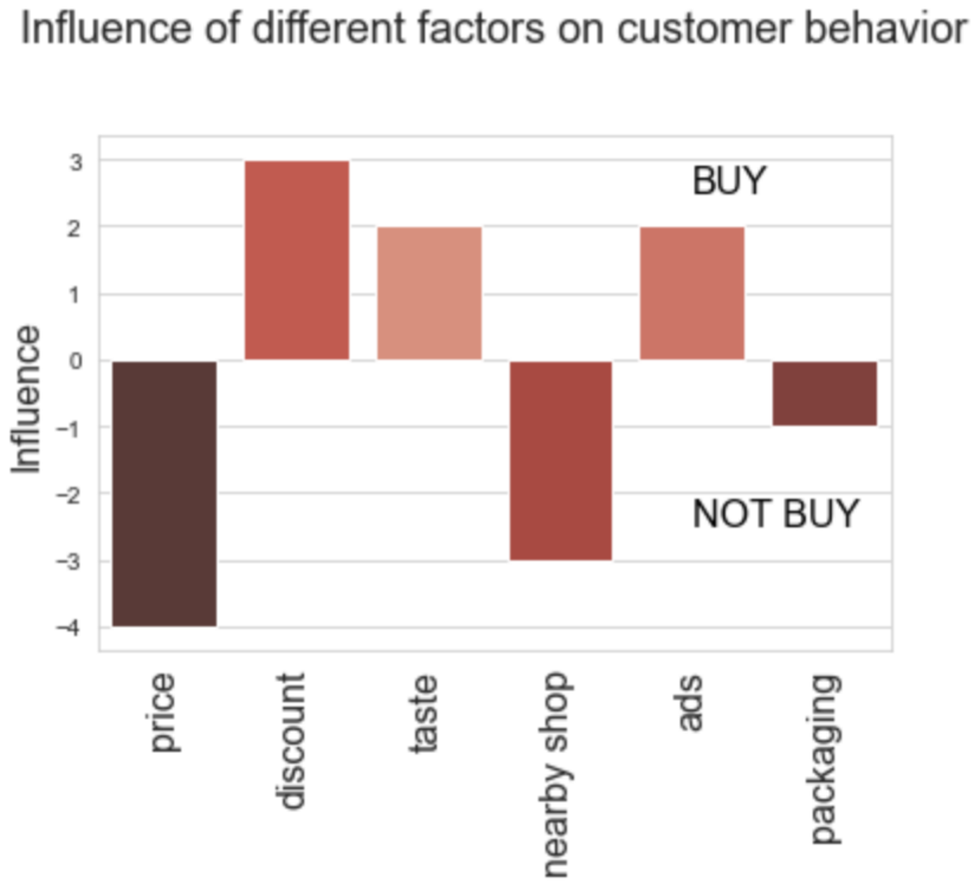
- **Investor**

- **Technical lead**

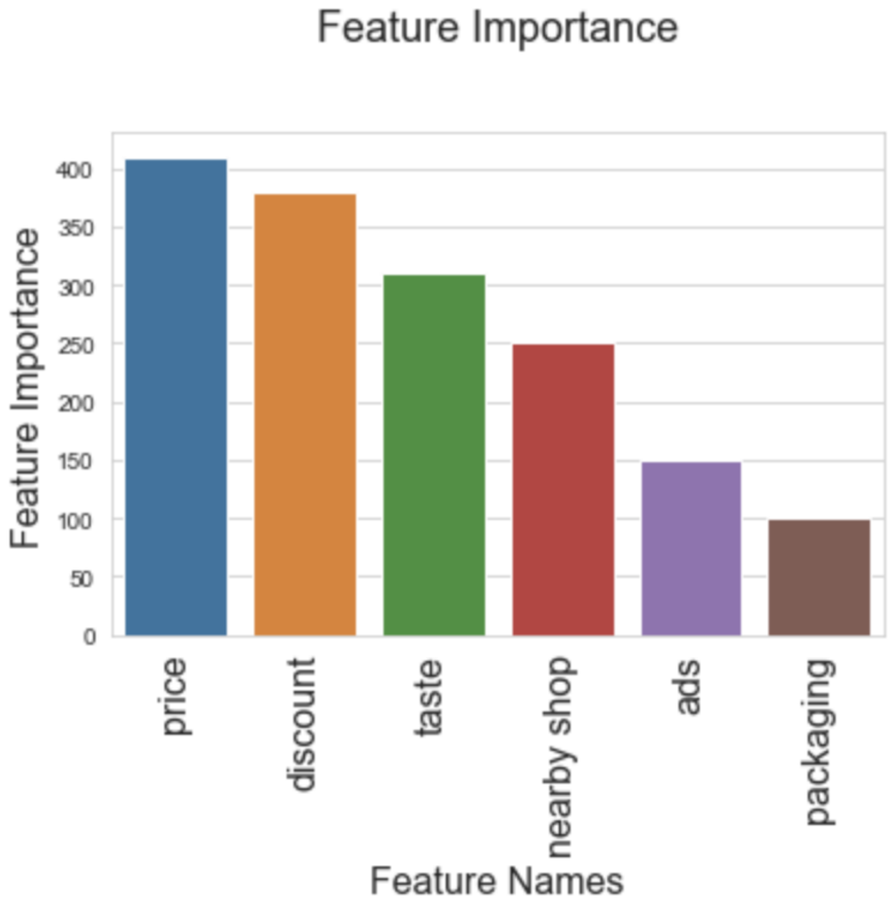


Provide context

- **Investor**



- **Technical lead**



Narrative structure

Background

- What **motived** the analysis?
- What **changed**?
- Who is the **focus** of the analysis?
  - Customers? Employees? Something else?

Our background: Total profit decreased

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

Narrative structure

- **What contributed to the problem?**
- Only relevant information

Our insight: Chips 20% increase. Sweets 30% decrease.

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

Narrative structure

- Add **supporting evidence**
- Help better explain the cause of problem

More insights: Most popular chocolate 50% decreased.

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

Narrative structure

- Central insight
- **What would happen if there is no change**

Our climax: Loss $10M next year.

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

Narrative structure

- Potential solutions
- Course of action
- Proactive

Our next steps: Rebrand chocolate.

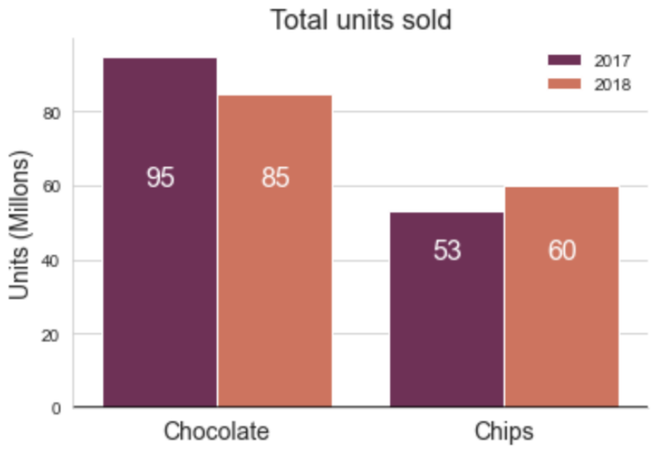
[1] Dykes, Brent. Effective Data Storytelling. Wiley.

Building narrative

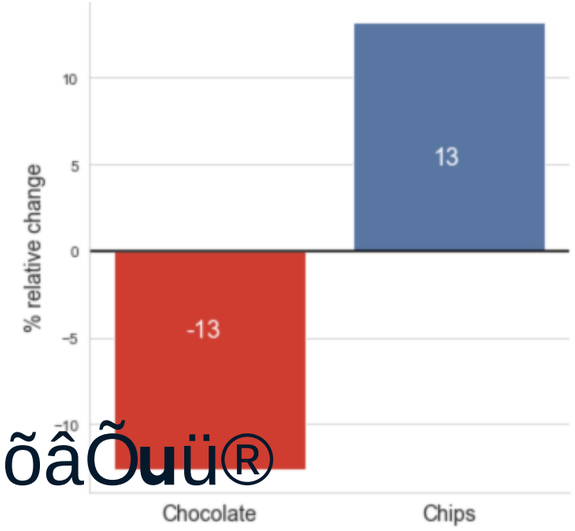
- **Change over time**: Chocolate lower in summer and higher in winter.
- **Correlation**: Chocolate rating vs. price
- **Comparison**: Two age groups vs. chocolate consumption
- **Clustering**: Groups with different coffee and chocolate consumption

Variations of data

Absolute



Relative

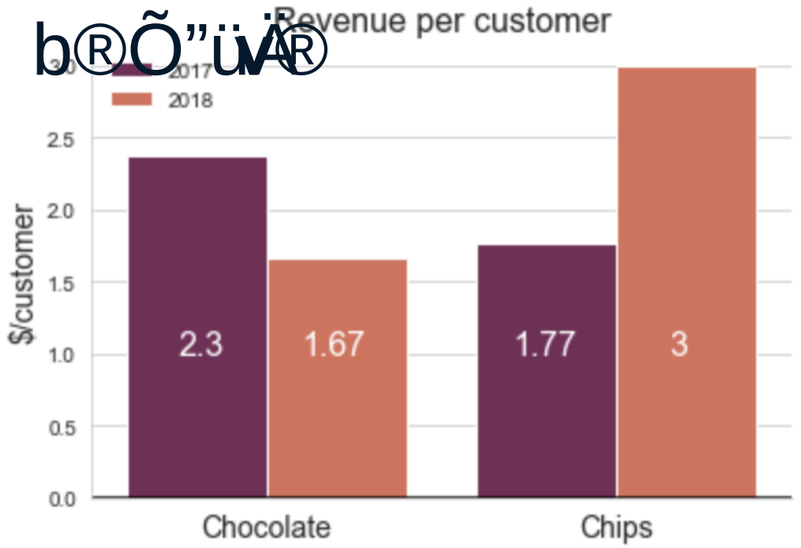


¢õâÕuü®

Ratio

- Quotient of two variables
  - Revenue per customer (**total product revenue/number customers**)
- Normalize values = **better comparisons**

b®Õ”üÄ®



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?

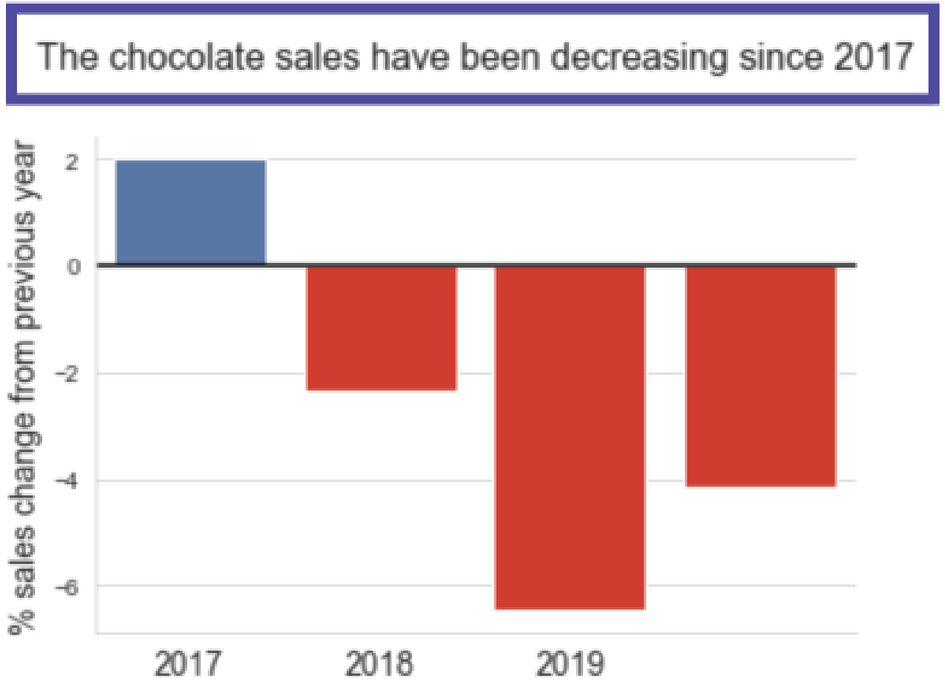
- Not proof of evidence
- Consider alternatives or complementary metrics

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
2. **Anticipate audience's questions**
   - Focus on story not on decoding graph
3. State insights
4. **Help the audience relate**
   - Importance
   - Action items



[1] https://artscience.blog/home/the-mccandless-method-of-data-presentation

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

Chapter 3

How to structure a written report?

- Types of reports
- Reproducibility
- Write precise and clear reports

Written reports

- Explain data analysis project
  - Sentiment analysis on product reviews

Written reports

- Explain data analysis project
  - Sentiment analysis on product reviews
- Communicate findings
  - 30% negative ratings for delayed shipping
  - Predict ratings with 90% accuracy
- Standards

Written reports

- Explain data analysis project
  - Sentiment analysis on product reviews
- Communicate findings
  - 30% negative ratings for delayed shipping
  - Predict ratings with 90% accuracy
- Standards
- Give recommendations to drive change

Types of reports

Informational

- Factual information
- Short
- Not strict structure
- Inform about facts

Analytical

- Analysis (relationships/recommendations)
- Varies (short or long)
- Strict structure
- Data-driven decisions

Final report

Elements

- Data analysis
- Findings and results
- Visuals

Format

- Long

Audience

- Details

Summary report

Elements

- Key findings and recommendations
- Visuals

Format

- Short (< 5 pages)
- Summary of final report
- Link to main document

Audience

- No need for details

Report structure

- Introduction
  - Purpose
    - Analysis of the product reviews gathered from website
    - Rating prediction based on review
  - Contextual information
    - Increase in negative reviews
  - Question of analysis
    - Factor affecting bad user experience

Report structure

- Introduction
- Body
  - Data
    - Description and tables
  - Methods
    - NLP and Random Forest
  - Analysis
    - Visuals
      - Graphs with most common words
  - Results
    - Description and visuals
      - 30% negative ratings associated with words "delayed" and "shipping".

Report structure

- Introduction
- Body
  - Data
  - Methods
  - Analysis
  - Results
- Conclusions
  - Restate question
  - Summarize important results
  - Add recommendations

Report structure

- Business context
- 1-3-25
  - 1 page of abstract
  - ? 3 pages of executive summary
  - ? 25 pages of detail

Audience

- People with little time
  - Introduction
  - Conclusion
  - Scan body

Audience

- Customer or internal collaborator
- Executive team
  - Scan introduction and conclusions
  - Recommendations

Audience

- Customer or internal collaborator
- Executive team
- Technical stakeholder
  - Body

Let's practice!

DATA COMMUNICATION CONCEPTS

Reproducibility and references

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

Written report

A report must be clear and **reproducible.**

Reproducibility example

- Baking a cake
  - Recipe
  - Raw ingredients
  - Our oven and kitchen measuring gadgets
  - Cake with a **similar flavor**
- Data project
  - Run analysis again - **same results**

Replicability example

- Baking a cake
  - Own utensils
  - Own ingredients
- Data project
  - Different environment

Reproducibility and replicability virtues

- Prevents duplication effort
- Build upon preexisting work
- Focus on new challenges
- Peer review
- Tool agnostic

Best practices

1. Keep track of how results were produced
   - Well document scripts
     - Comments in code
   - List packages and environment used
   - Version control

Best practices

1. Keep track of how results were produced
2. Avoid manual data manipulation
    - Data versioning
    - Store raw data and intermediate steps
    - Adapt and resolve problems
    - Example: data imputation
        - impute missing values with the mean
        - save and close editor
        - how to know which values were replaced in the first place?

Best practices

1. Keep track of how results were produced
2. Avoid manual data manipulation
3. Control randomness
    - Random seeds for ML pipelines
    - Controls confounding variables

Best practices

1. Keep track of how results were produced
2. Avoid manual data manipulation
3. Document randomness
4. Interpretability
    - Understand the cause of a decision or predict model results
    - Story with compelling narrative
    - Link with reproducibility

[1] Molnar C. Interpretable Machine Learning. 2019.

Best practices

1. Keep track of how results were produced
2. Avoid manual data manipulation
3. Document randomness
4. Interpretability
5. Cite bibliography correctly

References

- A **citation** is the basic information required to **identify** and **locate** a specific publication

References

- Different styles but same underlying logic
  - *Book*: Author Name (Year). Title. Publisher.
  - *Journal Article*: Author Name. (Year) 'Article Title.' Journal Title, Volume Number, Issue Number, Page Numbers.
  - *Website*: Author Name. Date of Publication, 'Title of Page/Work.' Title of Website, Location

- **APA style**:
  - In text citations (author, date)

Reference

- Reference management tools
  - Easier to keep track
  - Change between styles
  - Search for reference online
  - Options:
    - EndNote
    - Mendeley
    - RefWorks

References

- Business context
  - Less strict
  - Simpler (hyperlink)
  - ==> information available and retrievable

Let's practice!

DATA COMMUNICATION CONCEPTS

Write precise and clear reports

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

datacamp

Written report

A report must be **clear** and reproducible.

Write precise and clear reports

- Concise
- Precise
- Avoid misleading and confusion
- Meaningful message

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Empty phrases

- Contain no information
  - It is interesting to note that
  - The fact that
  - It should be pointed out that
  - It is well known that
  - It is obvious that

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Empty phrases

- Contain no information
- Distracting
- ==> should be removed

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Empty phrases

Negative ratings are associated with the words "delayed" and "shipping"

Another important point is the fact that negative ratings were associated with the words "delayed" and "shipping"

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Concrete nouns

- Write concrete nouns
- Avoid "this", "that", "it"
  - Adds cognitive load
  - Distracts them from insights

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Concrete nouns

This shows an accuracy of 80% when predicting customer churn.

The model shows an accuracy of 80% when predicting customer churn.

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

More pronouns

- Active voice: emphasis on the author
- Passive voice: stuffy and hard to read
- Academic vs business context

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Redundant adjectives and adverbs

- Phrases that say the same thing twice
  - Introduce a new
  - Done previously
- Eliminate redundant adjective and adverbs

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Run-on sentences

- Two or more independent clauses connected incorrectly
  - There is a correlation between delayed shipping and customer rating, the shipping delay is the cause for negative review.
- Correction
  - Make two sentences
  - Use dependent clause

Let's practice!

DATA COMMUNICATION CONCEPTS

Case study: report on credit risk

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

Credit risk

- Credit risk: probability of defaulting
- Loanme bank wants to predict if a customer is likely to default
- Raw data available
- Data Exploration Analysis
- Model training and evaluation

Audience

- Non-technical stakeholders
- Bank decision-makers

Story

- Background:
  - Increase in defaulting percentage over last 5 years.
  - Predicting which customers had a high probability of default.
- Insight: People with more unemployment periods tends to default more
- Insight: People with lower income tend to default more
- Climax: Possible to predict which people is more likely to default with an accuracy of 95%
- Next steps: Run a trial on a control population

Tech or non-tech

- Translate technical results

The right data

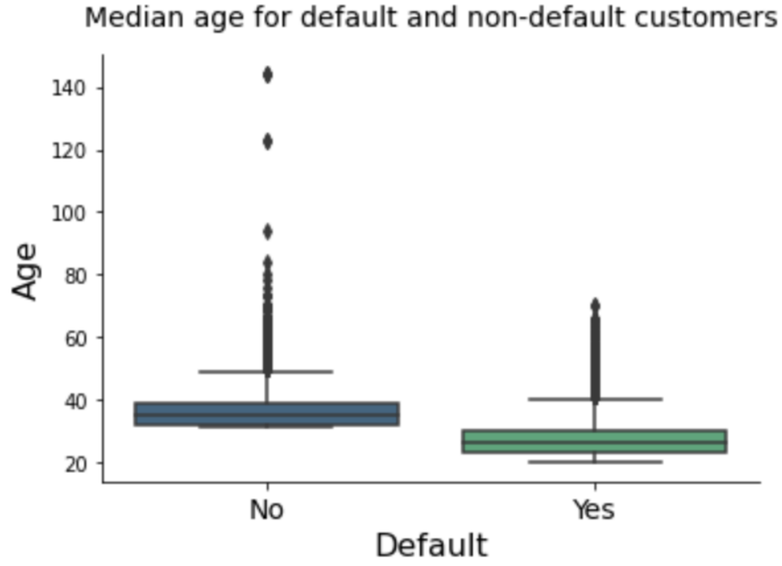
- Audience persona
  - **Role:** Financing Department Director
  - **Interest:** Decision on implementing an automated loan rejection system
  - **Appropriate data:**
    - Relationship between unemployment or income and loan default
    - Percentage customer defaulting over the next months

Statistics

- Median age and income
- Percentage of change

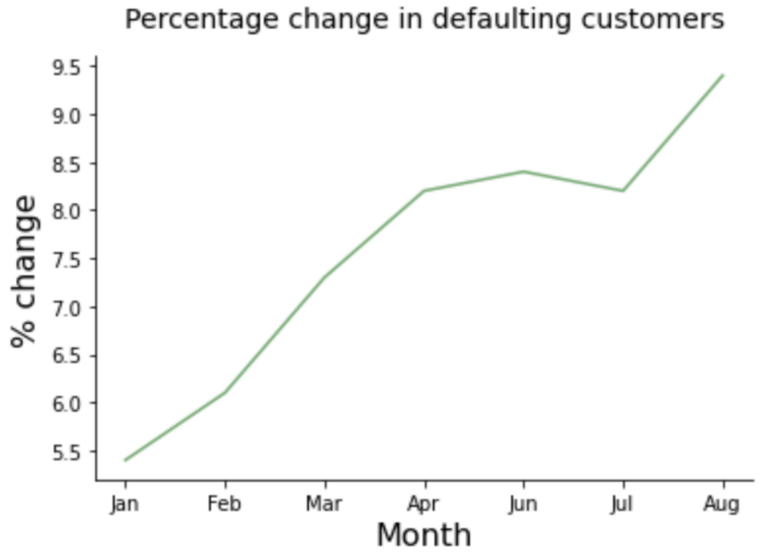
Visuals

- Boxplot with age vs. default condition



Visuals

- Boxplot with age vs. default condition
- Lineplot with % change defaulting customers



Correct format

- Who? **Financial Department director**
- Why? **Important decisions ahead**
- Content: **Key findings and recommendations**
- Channel: **Send the results before the meeting**

Report

- Written report
- Summary report or final report?

Report

- Summary report
- Informational report vs. analytical report?

Report

- Summary report
- Analytical report

Summary report structure

- Introduction
  - Purpose
  - Contextual information
  - Question of analysis
- Body
  - Data
  - Results: Key findings
- Conclusions
  - Restate question
  - Central insight
  - Add recommendations

Let's practice!

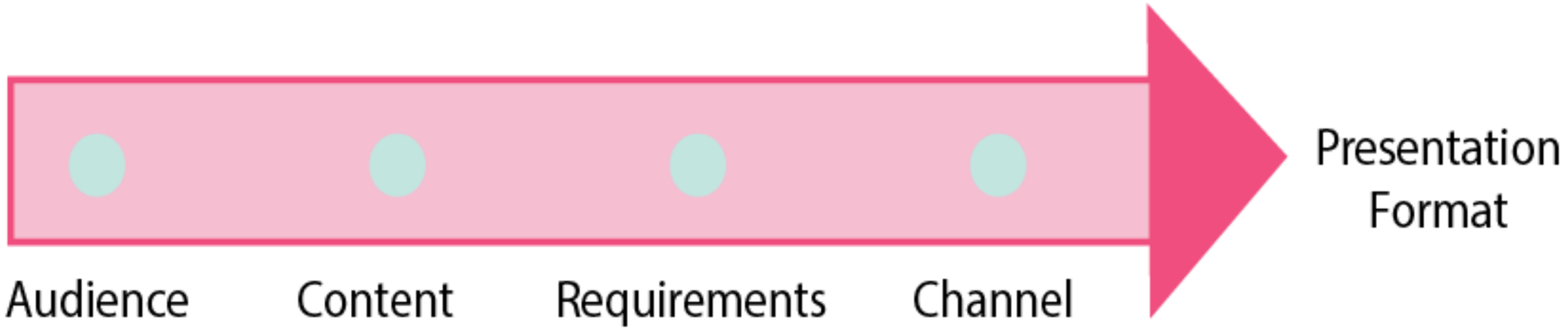
DATA COMMUNICATION CONCEPTS

Planning an oral presentation

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

Presentation strategy



Chapter 4

How to build a compelling oral presentation?

- Plan and build presentation slides
- Avoid common mistakes
- Present effectively

Plan a presentation

- Before building slides: **Plan!**
- Presentation structure
  - Purpose
  - Audience
  - Message

[1] Schwabish J. Better Presentations. Columbia University Press. 2017.

Purpose

What type of presentation?

- **Informative**
  - Current number of negative and positive ratings and words associated with negative reviews

Purpose

What type of presentation?

- Informative
- **Instructional**
  - How to build the model for sentiment analysis

Purpose

What type of presentation?

- Informative
- Instructional
- **Persuasive**
  - Follow-up actions to revert the current situation of high number of negative ratings

Audience

- **Who is the audience?**
  - Technical colleagues
  - Managers or executive team
  - Customer

Audience

- Who is the audience?
- **How big is the audience?**
  - Small meeting
    - Meeting with 10 members of the financial team
  - Conference or large meeting
    - Meeting with 100 employees from the software development department
  - Workshop
    - Technical workshop for 30 customer's IT employees

Message

What is the central message?

- After one week: 90% forgotten
- ==> What do we want to stick?
- **Opening statement**
  - Capture audience's attention
    - Negative ratings scare customers away from our website

Message

What is the central message?

- Open statement
- **Central message**
  - One sentence
    - Delayed shipping is the main cause of negative reviews and immediate actions are needed to revert the situation.

Message

What is the central message?

- Open statement
- Central message
- **Closing statement**
  - Sums up presentation and strengthens central message
    - There is a decrease in sales. Negative reviews have been increasing. Delayed shipping is causing negative ratings. Actions are needed to revert situation.

Structure

- **Introduction**
  - Provide background information
  - Catch audience attention
  - Glimpse of presentation content

Structure

- Introduction
- **Methods, analysis and model outputs**

Structure

- Introduction
- Methods, analysis and model outputs
- **Conclusions and takeaways**
  - Refers back to the introduction
  - Contains call-to-action statement or/and next steps

Outline

- Graphs and visuals
- Sections (five or less smaller parts)
  1. Reason for analysis
  2. Exploratory analysis
  3. Sentiment analysis
  4. Conclusions
  5. Follow-up actions

Keep time in mind!

- How long do you have?

Let's practice!

DATA COMMUNICATION CONCEPTS

Building presentation slides

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

From planning to building

- Slides
  - Support story
  - Short, dynamic

From planning to building

- Slides
  - Support story
  - Refined slides
  - **Slide count or timing = bad metric**

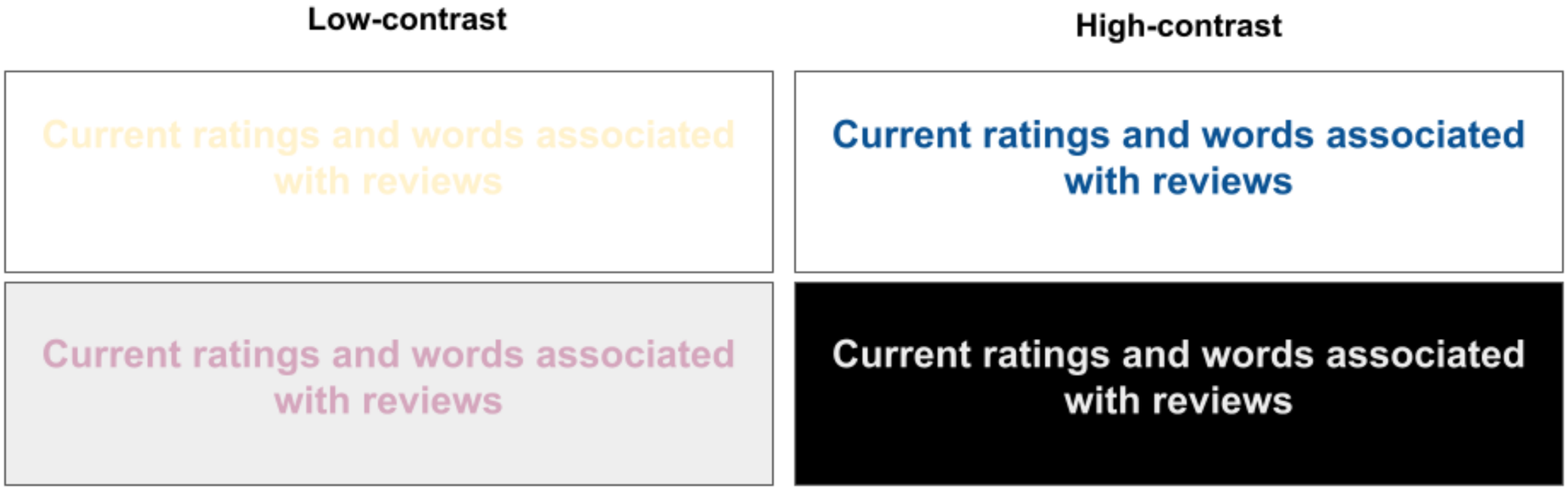
From planning to building

- Slides
  - Support story
  - Refined slides
  - Slide count or timing = bad metric
  - **One message per slide**

[1] https://www.slidecow.com/blog/how-many-slides

Color

- Convey meaning
- Readability

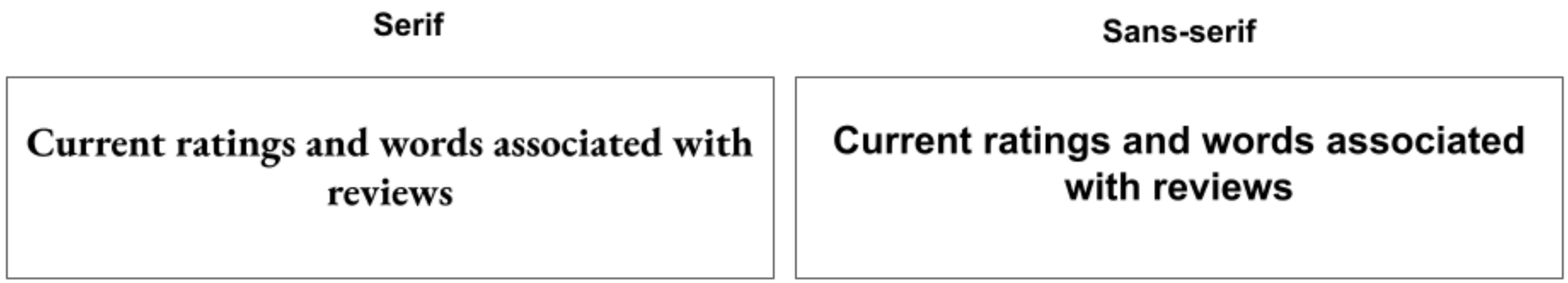


Color

- No more than **3 colors**
- Good **contrast** between words and background
- **Inclusive**
  - Color deficiency
  - Example: green and red

Fonts

- Serif vs sans-serif
- Context
- Support
- Size



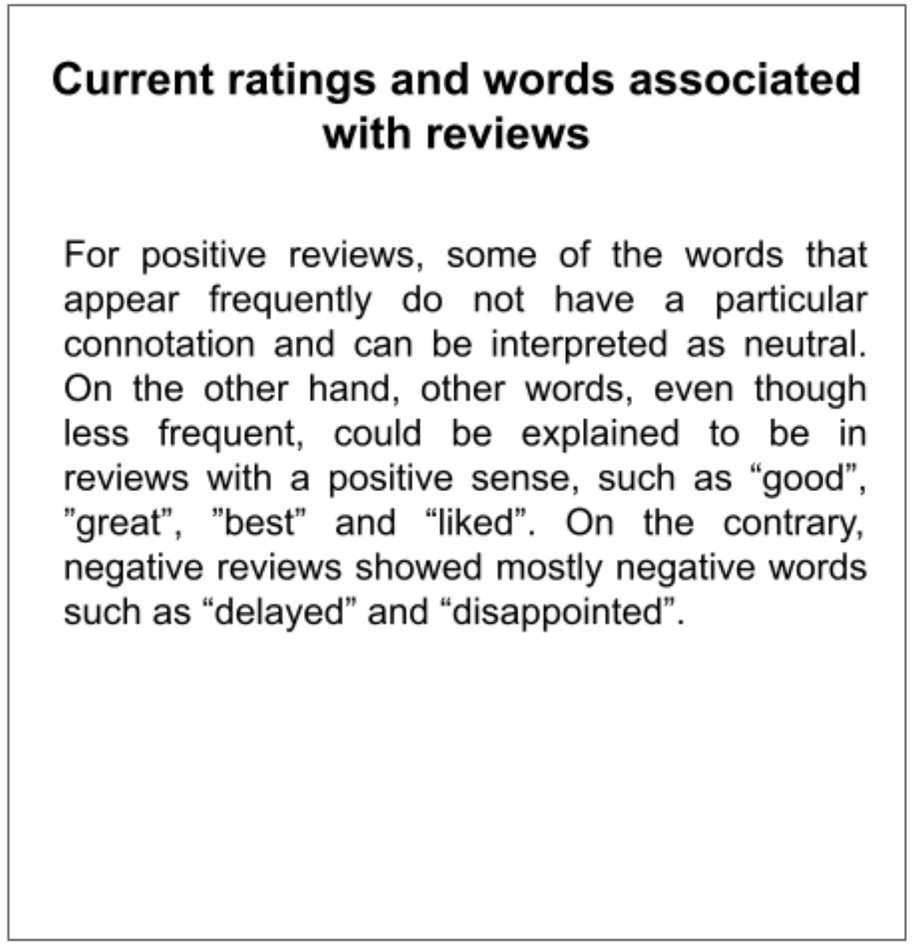
Fonts

- Several fonts
- Spacing of letters and lines
- **Bold**, italic and sizes

For **positive reviews**, some of the words that appear frequently do not have a particular connotation and can be interpreted as **neutral**.

Text slide

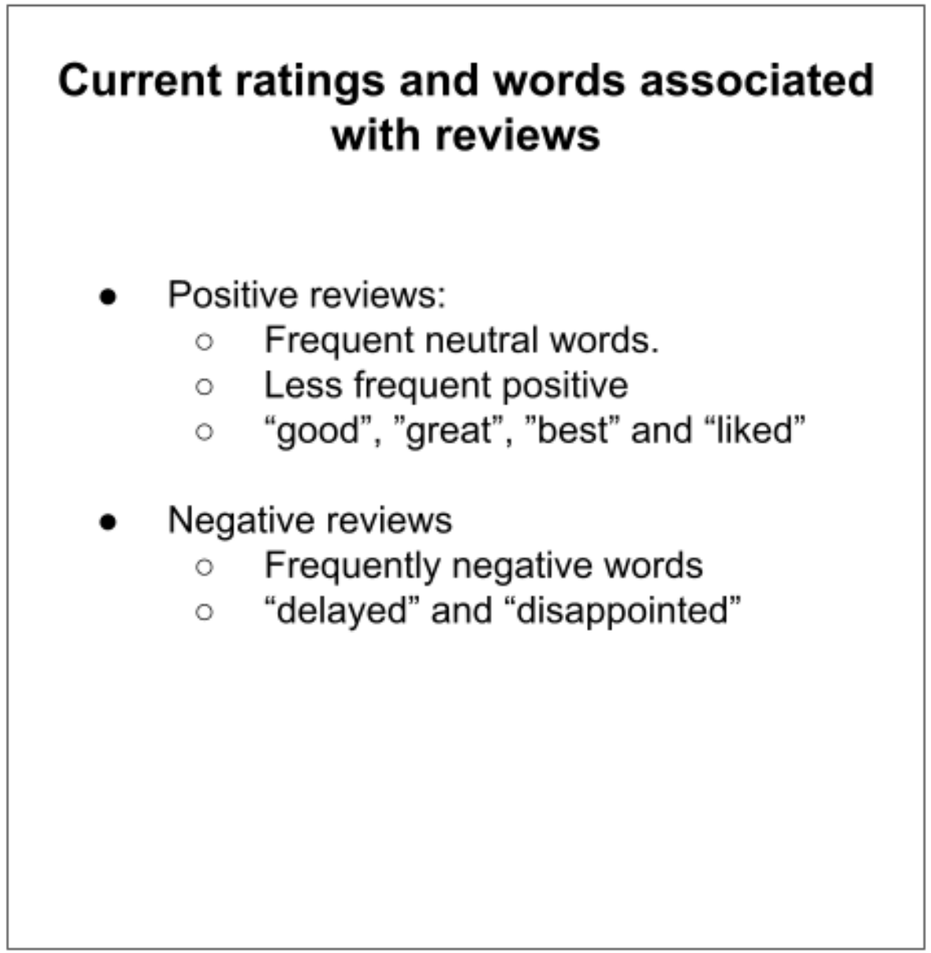
- Too much text
  - Audience reads instead of listening



[1] Schwabish, J. Better Presentations . Columbia University Press. 2017.

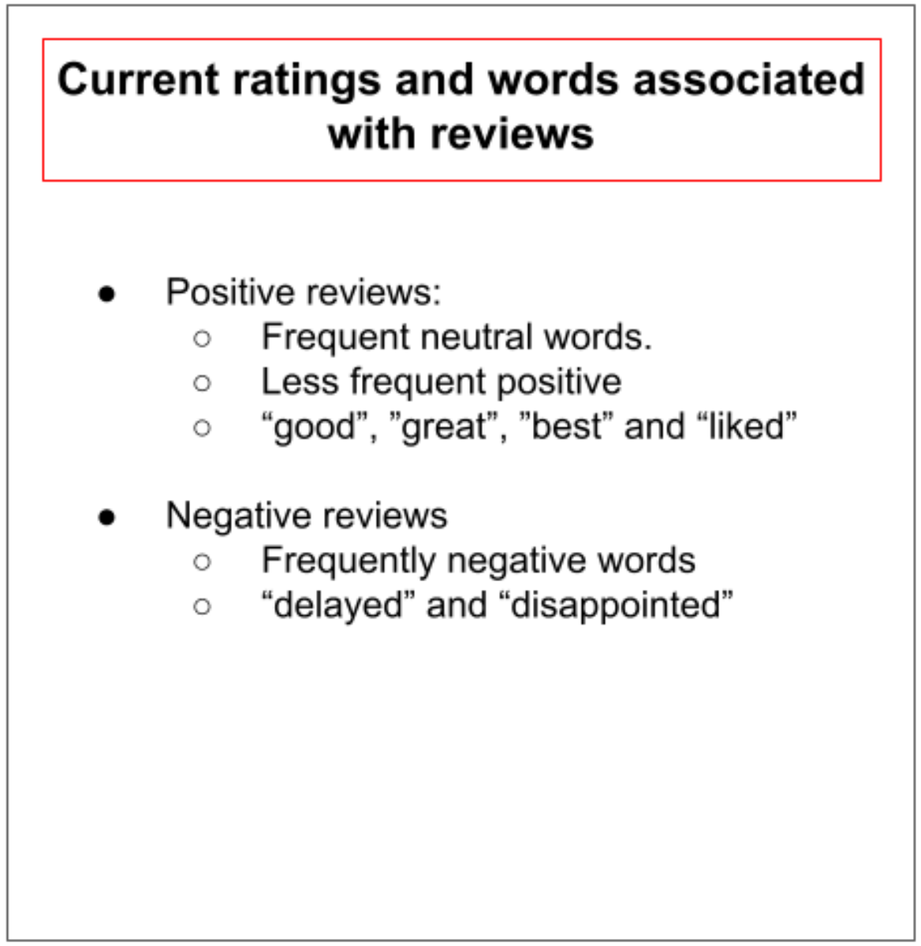
Text slide

- **Main points**
  - Don't dual purpose the slide deck



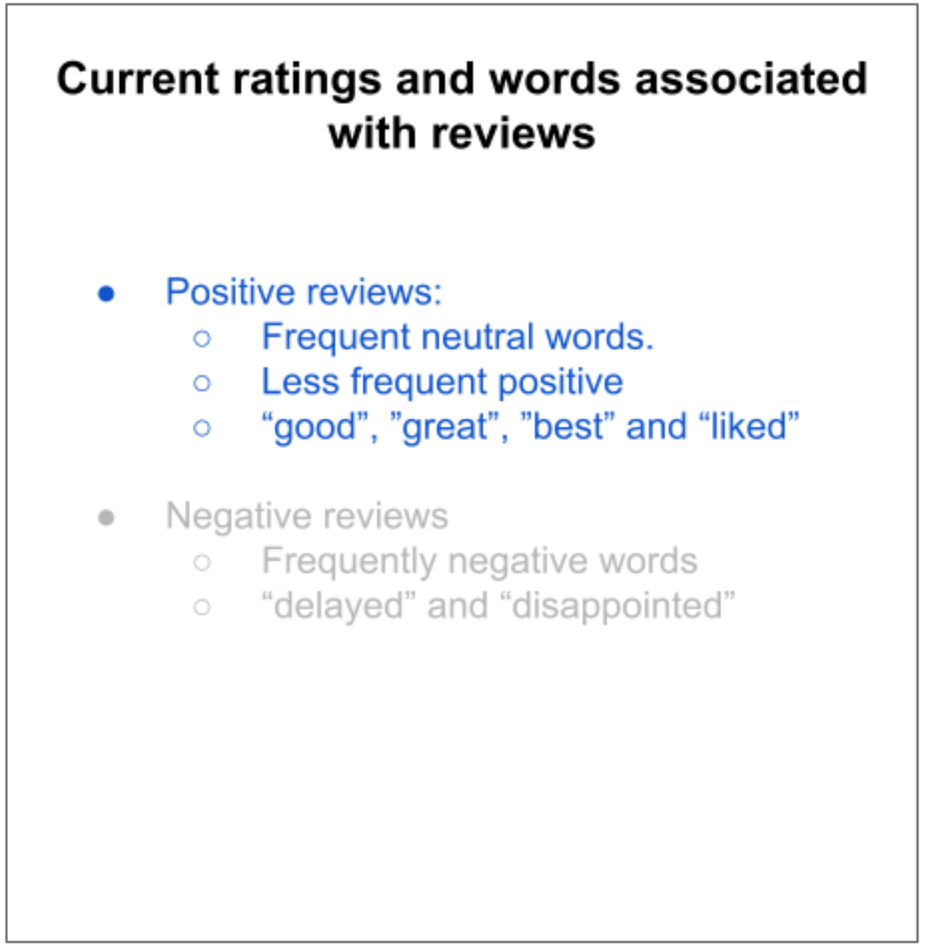
Text slide

- Less text
- **Headline**
  - Highlight main point
  - Specific and concise
  - Big size



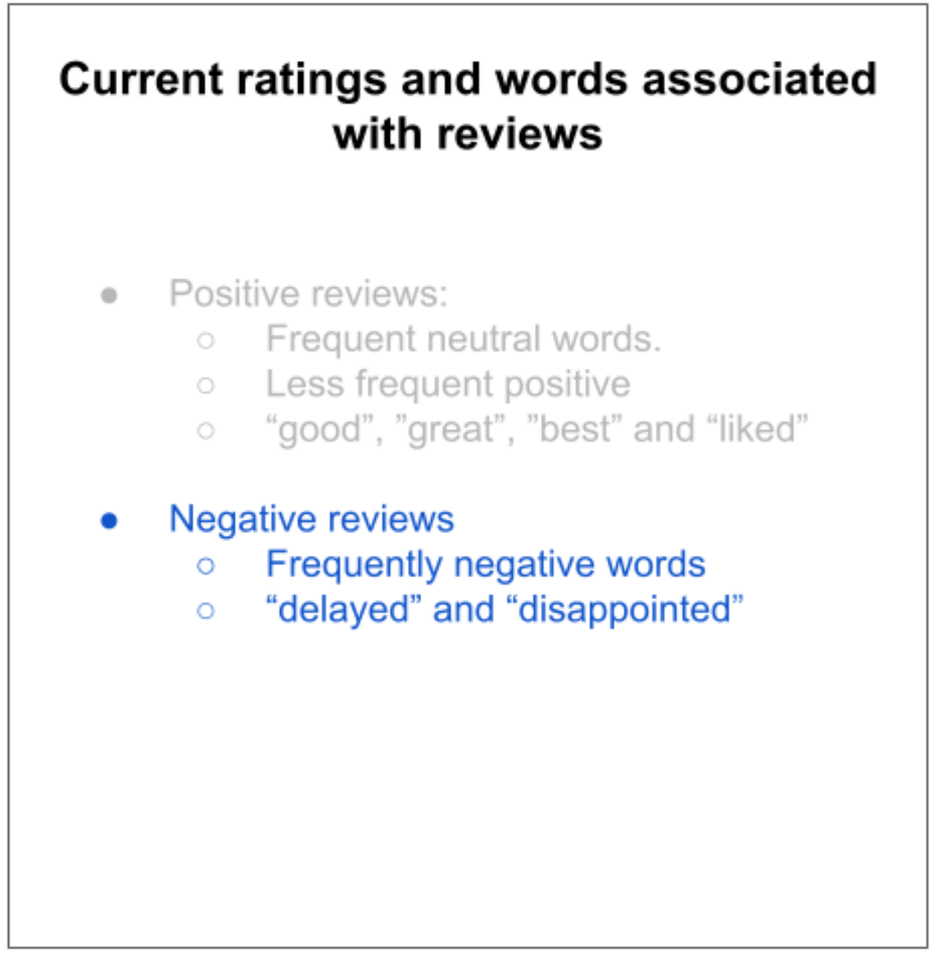
Text slide

- Less text
- Headline
- **Layering approach**
  - Breaks complex slide into smaller points



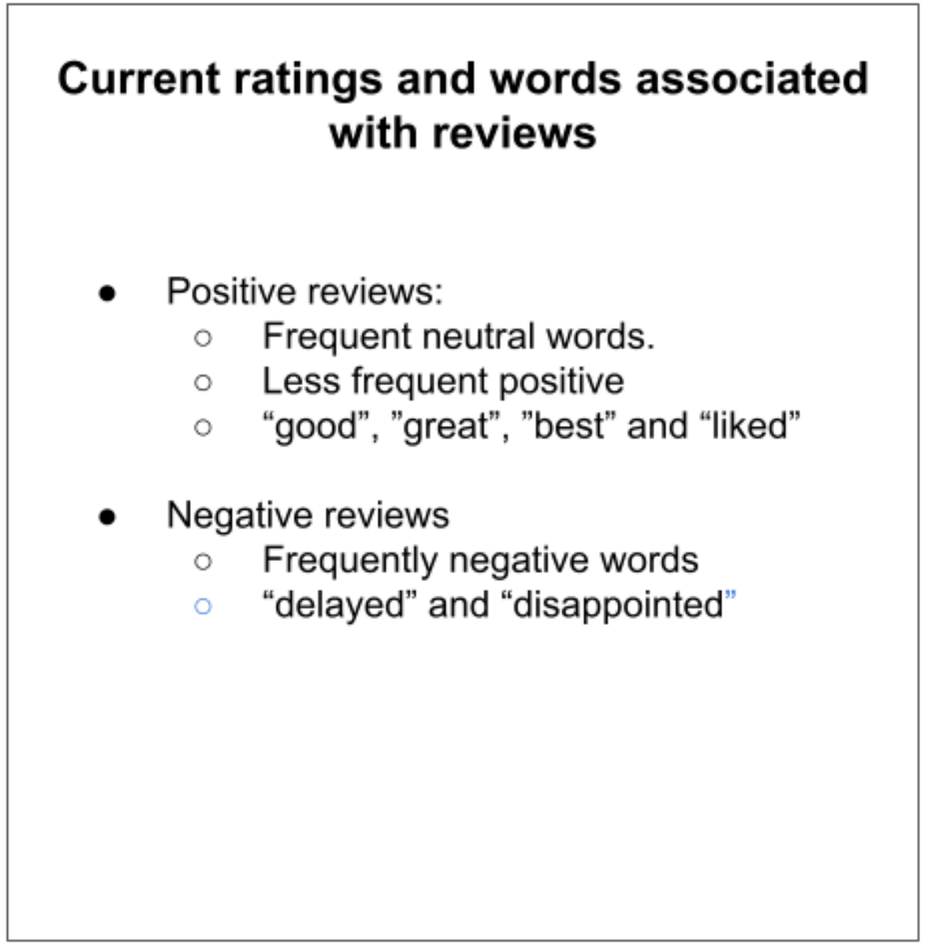
Text slide

- Less text
- Headline
- **Layering approach**
  - Breaks complex slide into smaller points
  - Present each point on its own



Text slide

- Less text
- Headline
- **Layering approach**
  - Breaks complex slide into smaller points
  - Present each point on its own
  - Displayed together



Visualization slide

- **Replace many sentences**

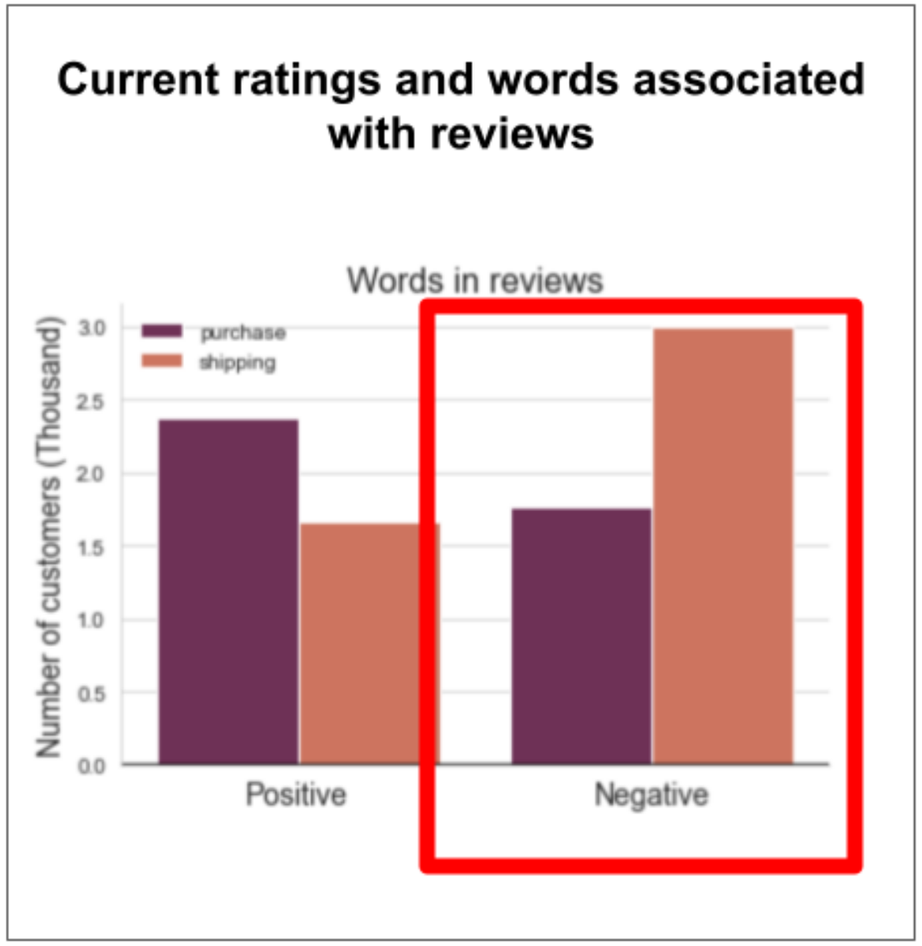
Current ratings and words associated with reviews

For positive reviews, some of the words that appear frequently do not have a particular connotation and can be interpreted as neutral. On the other hand, other words, even though less frequent, could be explained to be in reviews with a positive sense, such as “good”, ”great”, ”best” and “liked”. On the contrary, negative reviews showed mostly negative words such as “delayed” and “disappointed”.

Current ratings and words associated with reviews

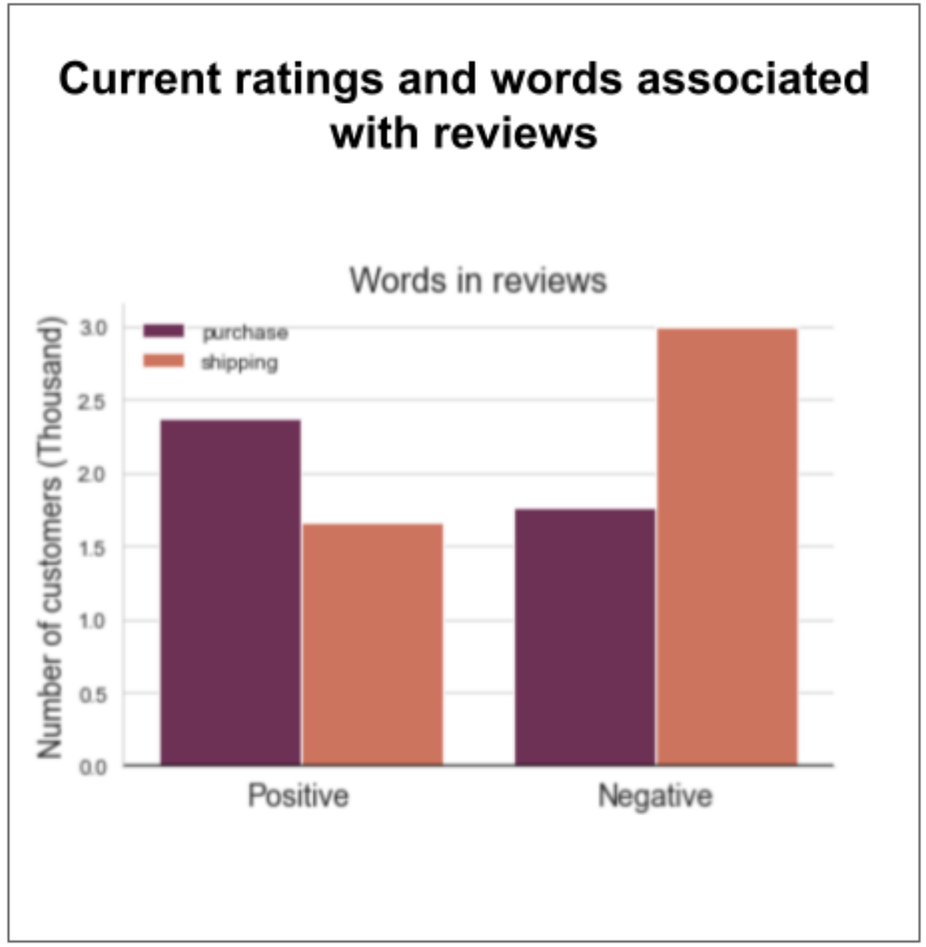
Visualization slide

- Replace many sentences
- **Use layering and highlighting**



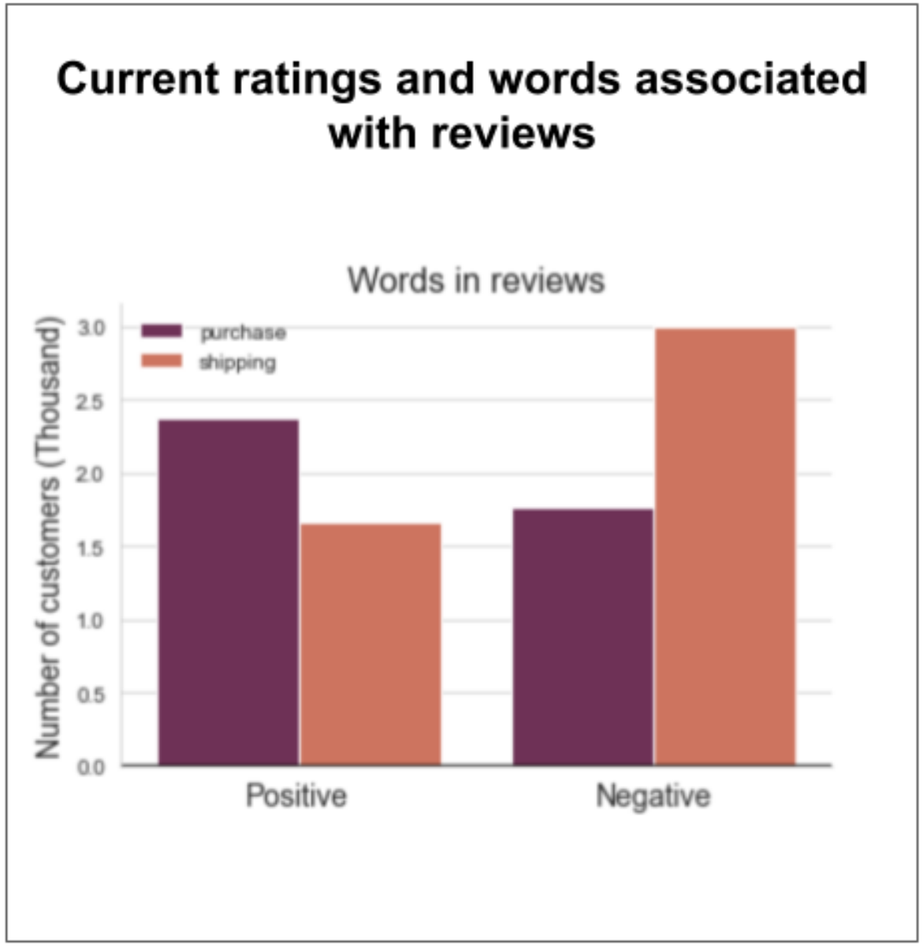
Visualization slide

- Replace many sentences
- Use layering and highlighting
- **Headline (if needed)**



Visualization slide

- Replace many sentences
- Use layering and highlighting
- Headline (If needed)
- **One or two full-size graphs**
  - One message per slide
  - No overcrowding



Let's practice!

DATA COMMUNICATION CONCEPTS

Delivering the presentation

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

datacamp

Recap

- Before building slides: **plan!**
- **Build** slides that support story

Practice

- Write script
- Don't memorize
- Become familiar with content
- Anticipate follow-up questions

Practice

- Prepare
- **Rehearsal**
  - Stand up
  - Use the slides
  - Speak out loud
  - Detect distracting patterns (um, so, like, basically, actually)
  - Find linking statements
  - Answer to Q&A

Deliver the presentation

- **Be aware of emotions**
  - Confidence vs. unsure

Deliver the presentation

- Short attention span
- Talk **to** audience (not _at_ them)
- Develop a relationship

Deliver the presentation

- Be aware of emotions
- Talk to audience
- **Timing**
  - Use allocated time

Deliver the presentation

- Be aware of emotions
- Talk to audience
- Timing
- **Pace**

Deliver the presentation

- Be aware of emotions
- Talk to audience
- Timing
- Pace
- **Open up for questions**
  - During or at the end of the presentation

An effective oral presentation

- Talk to the audience
- Less is more
- Consistent and persistent take-home message
- Practice
- Prepare to answer follow-up questions

Let's practice!

DATA COMMUNICATION CONCEPTS

Avoiding common errors

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

Recap

- Plan
- Prepare and practice
- Deliver

Length

- Ineffectively long presentations
- Attention span
- Around 20 minutes
- Leave time for Q&A

Purpose

- State the purpose at the beginning
- Better understanding
- Better story impact

Guide audience

- Sequence of information
- Keep audience's attention
- Do not leave all findings to the end

Audience involvement

- **Engage** and **involve** audience

Audience involvement

- Engage and involve audience
- **Strong introduction**
  - *Good morning! My name is Hadrien, and I'm here today to present how negative ratings are impacting the company profits.*

Audience involvement

- Engage and involve audience
- Strong introduction
- **State key assumptions**

Audience involvement

- Engage and involve audience
- Strong introduction
- State key assumptions
- **Ask questions**
  - Know answer
  - Hook for next slide

Audience involvement

- Engage and involve audience
- Strong introduction
- State key assumptions
- Ask questions
- **Reiterate** to main idea

Body language

- What matters is the **message**...
- ...but the **speaker** is at the center of the presentation

Body language

- What matters is the message...
- ...but the speaker is at the center of the presentation
- Emphasis by **natural gesture and movements**
  - Move hands or point at slide
  - Smile or make a facial expression

Body language

- What matters is the message...
- ...but the speaker is at the center of the presentation
- Emphasis by natural gesture and movements
- **Attracts attention**
  - Posture can convey confidence

Body language

- What matters is the message...
- ...but the speaker is at the center of the presentation
- Emphasis by natural gesture and movements
- Attracts attention
- **Supports message**

Voice tonality

- Use different voice tonalities

Voice tonality

- Use different voice tonalities
  - **Speed**
    - **Fast**: urgency, excitement, and emotion
    - **Slow**: importance, and new ideas introduction

Voice tonality

- Use different voice tonalities
  - Speed
  - **Volume**
    - Live: speak loud
    - Online: check mic

Voice tonality

- Use different voice tonalities
  - Speed
  - Volume
  - **Intonation**

Let's practice!

DATA COMMUNICATION CONCEPTS

Congratulations!

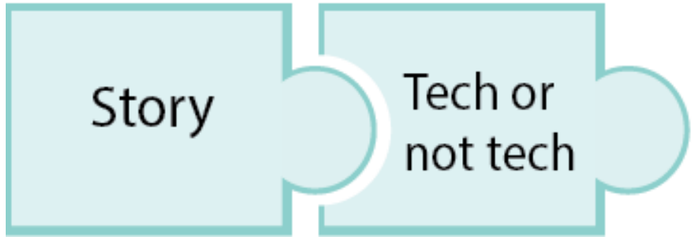
DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

What you've learned - chapter 1

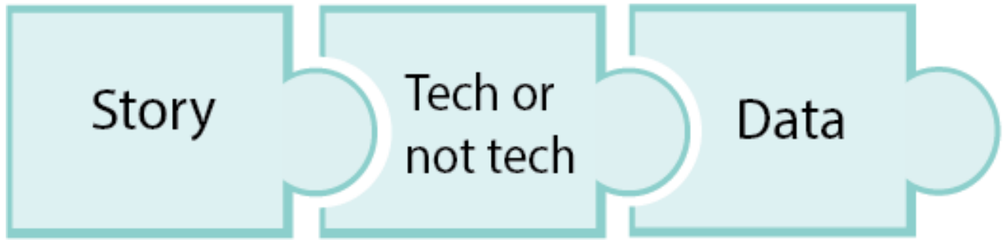
- Importance of data storytelling

What you've learned - chapter 1



- Importance of data storytelling
- Translate results for non-technical stakeholders
- Craft stories that impact the decision-making process

What you've learned - chapter 2



- Select right data and statistics
- Audience persona

What you've learned - chapter 2

- Select right data and statistics
- Audience persona
- Choose appropriate visualization

What you've learned - chapter 3

- Types of reports
- How to structure a clear report
- Reproducibility

What you've learned - chapter 4

- Planning and building a presentation
- Importance of practicing and rehearsing
- Best practices and common mistakes when delivering a presentation

Next piece

Congratulations!

DATA COMMUNICATION CONCEPTS