Information Visualization

# Course introduction

Lesson 1

**Marilena Daquino**
Assistant Professor

Department of Classical Philology and Italian Studies

marilena.daquino2@unibo.it

# Table of contents

# 01

# Overview

What we talk about, what you are interested in (topic-wise) and what you want to take form this course

01

# Data analysis

What do you want to learn?

Methods to query, process, analyse Linked Data with Python

# Data visualization

Methods and techniques to plot data with Python and Js

# Web communication

Presentation techniques for explanatory projects

# Background

02

What I wish you already knew

## Comp Think

Python (intermediate)

Install libraries, Jupyter notebook, read/write CSV and JSON data

## Web tech: UI / UX

HTML, CSS, JS (good)

JQuery for UI, modify DOM, interactivity

## Introductory methods

GitHub (good)

A short introduction
Github guides
sourcetree GUI

## Knowledge mgmt.

RDF, SPARQL, OWL (good)

Read RDF in several syntaxes, read/write SPARQL queries, understand basics of ontologies

# Background

02

What I will show you

## Comp Think

Python and Jupyter

Libraries for data exploration.
Jupyter to document your work

## Web tech: UI / UX

JS (good)

Libraries for data visualization,
Digital storytelling strategies

## Introductory methods

Github, Colab

Publish your work on github (data, software and website)

## Knowledge mgmt.

RDF, SPARQL, OWL (good)

Python APIs for RDF/SPARQL

# Classes overview

**1/4**

Introduction

**2/4**

Preliminaries

**8/4**

RDF

**9/4**

SPARQL

**15/4**

Data sense making

**16/4**

Data sense making (2)

**29/4**

Data visualisation

**30/4**

Storytelling

**6/5**

Workshop

**7/5**

Publication, review, and wrap up

# 02

# Project

A web project explaining what you have discovered while analysing selected data sources

# 03

What you'll do


## The project

Choose a **topic**, find data-driven **questions**, **analyse** and **visualize** data, produce a **notebook** with your code, and a **website** for presenting results.

# 03

Step by step

## MUST

### Topic

Pick a domain, and quantitative questions

### Data

Find or create data that support your analysis

### Group

Max 3-people group to share the pain

# Topics



Some suggestions

The course will make use of examples from the History of Art. The topic is free as long as you are able to find good questions, data, and a group by your own.

# Topics

03

Some suggestions

## Art history

**Artworks provenance**
E.g. What are the artworks that travelled most in Europe?

**Iconography**
E.g. What subjects are mostly represented in artworks of 16th century?

**Connoisseurship**
E.g. What are the most reliable criteria to justify an artwork attribution?

## History of Art history

**Art historians' relations**
E.g. Which countries are historians from and where did they work?

**The debate on research topics (artists, periods, movements)**
E.g. In which periods the research focused on Mannerism?

**Resources**
E.g. which types of resources in archival collections represent research topics?

# Topics

03

Some suggestions

## Photography of Art

**Photographers' relations**
E.g. Which photographers worked in Italy in 20th century?

**The most photographed artworks**
E.g. What art genres are mostly represented in professional photography?

**Archives and museums**
E.g. Which museums commissioned photographs of artworks?

## Gender in Arts

**Representativeness of female or non-binary gender in history**
E.g. How influential were female photographers in the 20th century?
E.g. Are female photographers under-represented in photo archives?

**Photographers' occupations**
E.g. What other occupations had female art historians?

**Women and market**
E.g. Who buys women's art?

# Ask for advice



Once you defined your research questions, send an email to me **marilena.daquino2@unibo.it** for feedback and suggestions.

Discussing the topic and the research questions will help your work to be sound, not to be abandoned right after the exam, and maybe it will be useful to somebody in the near future.

Best projects always find their way to get popular.

# 02

# The data

The course will make use of data from a few Linked Open Data for Art History and popular sources. Consider integrating multiple sources to answer your questions. You must use two data sources of which **at least one Linked Open Dataset**.

# Data

03

Some suggestions

## Artchives

**http://artchives.fondazionezeri.unibo.it**

ARTchives includes data created by cataloguers of art historical photo archives and reuses data harvested from Wikidata. Data includes information on art historians, archival collections, debated art genres, and keeping institutions.

## Zeri Photo archive

**http://data.fondazionezeri.unibo.it/**

Zeri & LODE includes data created by the Zeri Foundation and reuses data harvested from Wikidata, DBpedia, ICONCLASS, AAT Getty, VIAF. Data includes information on artworks and photographs of artworks collected by Federico Zeri, one of the most notable connoisseurs of last century. The dataset is limited to artworks of Modern Art.

# Data

03

Some suggestions

## Wikidata

**https://query.wikidata.org/**

Wikidata is a general purpose Linked Open Dataset, originally born to represent structured data of Wikipedia (the right-side boxes) in RDF. It includes a variety of information, such as people biographical data, scholarly publications, historical events, and so on.

## And more...

You can integrate data sources with other data, Linked Data or not.

You can work on other sources that you found being appropriate.

# 02

# The Group

3 people max. Justify your contribution to the project. Grades are individual (you are judged for your contribution).

You can work alone, but you need to achieve the same results. No discounts :(

# 03

# Evaluation

What to prepare, what happens the day of the exam, how I grade your work.

# 04

The day before the D-Day

## MUST HAVE

### Presentation

15’ **presentation** (optional slides) according to a template I give later

### Jupyter notebook

A **jupyter** notebook including the data processing on a Github repo

### Website

An **online webpage** presenting results of your work to a broader audience

The D-Day

# HOW IT GOES

## Presentation

You present your project and show the **website** (15’)

## Grading review

While you present I may review the grade I already assigned you

## Q&A

Discussion of methods, **results** and questions

# Jupyter notebook

04

Preparation checklist

## Install

Jupyter [1] or use Colab

## Create

1 notebook per project

## Abstract

Introduce the scope in the notebook

## Data

Manipulate data via python and save results as CSV/JSON

## Markdown

Document functions and operations with markdown [2]

## Clean up

Keep it short, group functions, use titles and Table of contents

# Website

04

Preparation checklist

## HTML / CSS /JS

Create a static webpage, you can use templates

## Data

Access your data from JS (CSV/JSON or APIs)

## Visualize

Create charts with your data

## Present

Add titles, sections, descriptions of charts, and discuss **results**

## Credits

Add people, **roles**, and **licenses** (of both data you used and that you created)

# GitHub repository

04

Preparation checklist

## Create

1 repo per project

## Upload

Notebook, website (optional) and data

## README file

Add project title, people, roles, licences

## Binder

[optional] Connect the notebook to Binder

## Open an issue on the course repo

Open or comment an issue on the repo of the course called “Exam DD/MM/YYYY” with: Project title, Website URL, Repository URL, People [here]

# Presentation

Preparation checklist

If you present with a slide presentation, **please make it 10 slides max** (No need to share it before the exam day).

Use the following template to prepare the **talk**.

1. Title
2. Background (the domain, the problem)
3. Goals, Research questions, Approach
4. Data sources, preparation and data analysis
5. Data communication strategies
6. Summary of results

NB. The projects will be evaluated the day before the exam. If the project is good enough, we may agree on an *offline evaluation* (I send you comments and grade via email). We can always have a chat during the exam session.

# An example

**Title: Trends in the study of artistic periods**

**Background.** Artistic periods are been studied by art historians over time with different granularities and degree of interest.

**Goals.** Discover trends of scholars' interest towards artistic periods

**Research questions.**

1. **When** there is **more interest** towards artistic periods?
2. **How** such an interest **evolves** over time?
3. **Which** artistic periods show a significant trend over time?

**Approach.** Aggregate dates of art historians' *flourit* (i.e. date ranges of their archival collections) and studied periods.

# An example

Pay attention!

**Data preparation and data analysis.** We use ARTchives and Wikidata. Data about archival collections, art historians' activity dates, and related artistic periods are collected from ARTchives. Descriptions and dates of artistic periods are collected from Wikidata.

We query ARTchives and Wikidata SPARQL endpoints, we reconcile the data, we prune duplicates (e.g. "Baroque" and "Baroque art").

We perform some preliminary analysis to understand the **distribution** of periods over the archival collections. We analyse the **trend** of artistic periods as subject of art historians' collections over time.

**Data visualisations selected and reasons.** We show trends in a line chart having on the x axis the dates of art historians' period of activity (corresponding to the time range of their collection) and on the y axis the artistic periods they studied. So doing a user can see all trends at the same time and figure if there are correlations in trends.

# An example

Pay attention!

**Data communication strategies.** We show preliminary exploratory visualisations about the distribution of periods so as to demonstrate the validity and representativity of sources, hence of our results.

We show results of our investigation as an interactive line chart where users can select/remove the period to be shown in the graph.

**Summary of results.** Most significant results show that artistic period XX received attention in the 19XXs, while period YY shows a significant loss of attention in the same period. Notably, period XX and ZZ have similar trend over time.

*NB. TO EXPLAIN **WHY** is difficult if you are not an art historian, but best projects are those that say more than stating the obvious and look for patterns somewhere else.*

# No panic, you can copy

Pay attention!

- You can reuse existing **templates** for the website (e.g. CMS, HTML templates)
- You can reuse any **py/JS libraries**
- Feel free to use other solutions than **GitHub** to publish the website, However, a git repo is mandatory to publish notebook and data.
- We will set up all the pieces of your project in dedicated hands-on classes

# Now you can panic, you cannot use chat GPT & co.

Pay attention!

- You must **produce original content**. You can use genAI to improve your text, but the content must be original, well written, and be precise. If you (or chatGPT) write hot air, I'll notice (or maybe I won't, but I'm sure I'll not like it).

- You **can** use genAI to help you in coding tasks. That depends on how proficient you are or want to become in some technology.

# Evaluation

My grades

## SKILLS

### Data

Correct SPARQL queries

Correct final data

Data integration

### Research

Soundness of research questions

Use of adequate visualizations

Graphic skills

### Presentation

Clarity of the presentation

Summarise complex issues

Surprise me!

26

28

30

30L

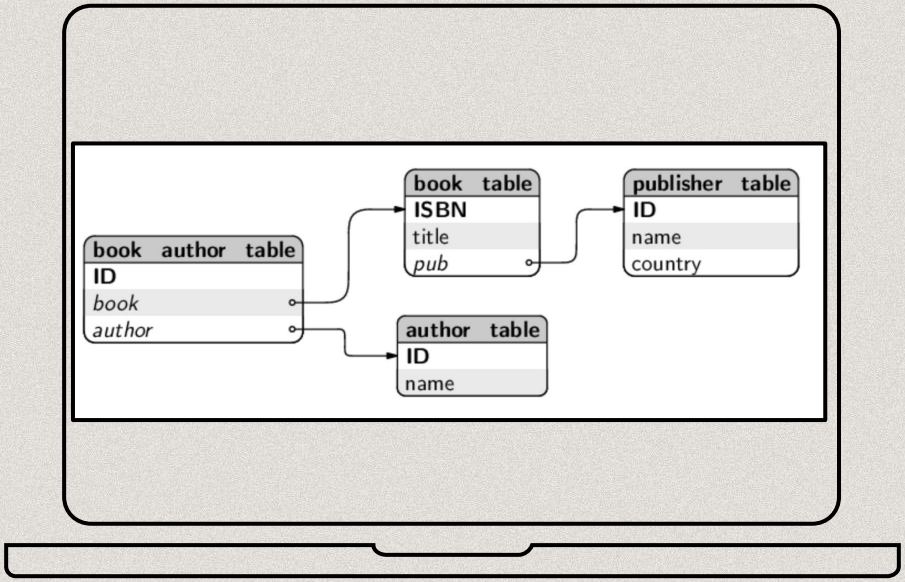
# 04

# Basic concepts

Data and visualization dimensions

01

Data attributes



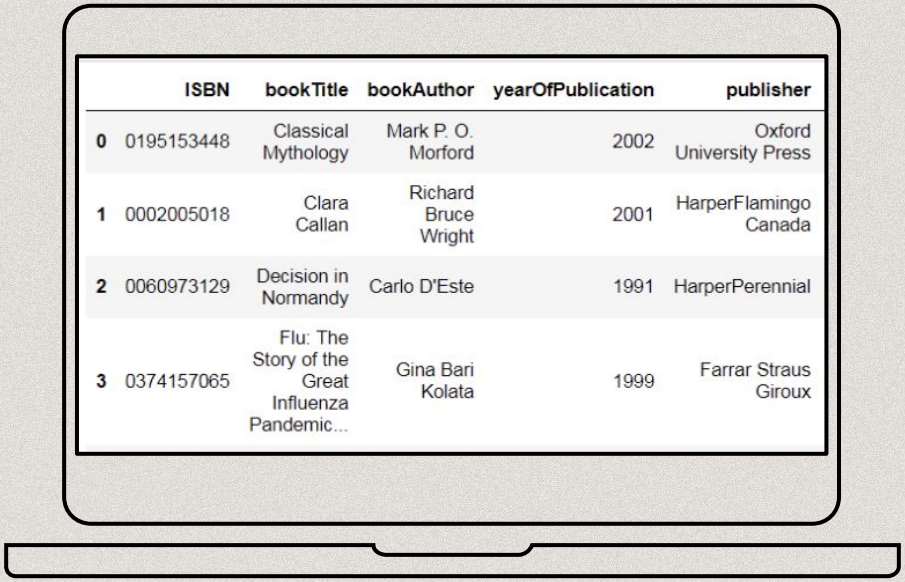

# Data

Data are abstractions of concepts and real-world **entities** (people, books, places)

Data have **attributes**, also called features or variables (title, author, ISBN are attributes of books)

01

Data attributes



| | ISBN | bookTitle | bookAuthor | yearOfPublication | publisher |
|---|---|---|---|---|---|
| 0 | 0195153448 | Classical Mythology | Mark P. O. Morford | 2002 | Oxford University Press |
| 1 | 0002005018 | Clara Callan | Richard Bruce Wright | 2001 | HarperFlamingo Canada |
| 2 | 0060973129 | Decision in Normandy | Carlo D'Este | 1991 | HarperPerennial |
| 3 | 0374157065 | Flu: The Story of the Great Influenza Pandemic... | Gina Bari Kolata | 1999 | Farrar Straus Giroux |

# Dataset

A dataset is a **collection** of entities with their attributes.

It can be seen as an **n*m matrix,** wherein n is the number of rows (entities) and m is the number of attributes (columns)

# 01

Data types

## Attributes

### Numeric

Measurable **quantities** (e.g. temperature, price)

### Categorical

[also called Nominal] **Names** from non-overlapping sets, classes, or states (e.g. title, sex)

### Ordinal

Nominal or numeric attributes that can be **ranked** (e.g. days, years, Likert: “strongly like” to “strongly dislike”).

# 01

Data types

## Attributes

### Numeric

Measurable **quantities** (e.g. temperature, price)

**Interval** attributes are measured on the basis of a scale with an interval and an origin (e.g. time, temperature). They have no true-zero.

**Ratio** attributes have a true-zero - meaning no quantity is measured in that point (e.g. price)

# 01

Operations on data types

## Operations

### Numeric

Interval can be **ordered**, but not multiplied/divided.

On Ratio you can do **arithmetic** operations.

### Categorical

Can be **sorted** (e.g. alphabetically) and counted.

Cannot be ordered and arithmetic operations cannot be performed.

### Ordinal

Can be naturally **ordered** and counted.

Arithmetic operations are not possible.

# What can you do with these data?

01

Test

Answer the questionnaire (10 minutes)

https://forms.gle/caSW8W4nPaavsir56

# Graphical properties



Some properties of a data visualization help making noticeable graphical elements

## Axes

## Layout

## Shape

## Colour

## Size

## Typography

# Axes

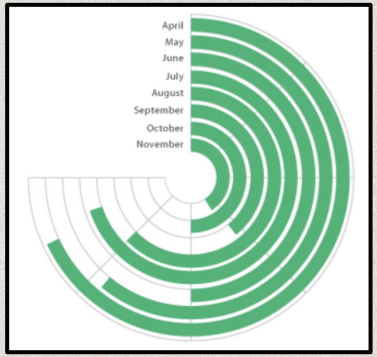


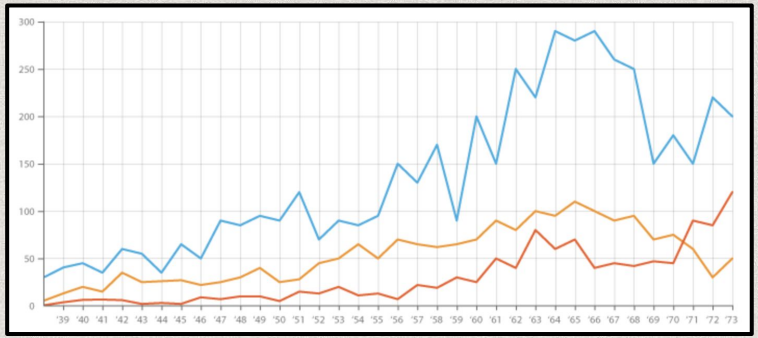
Graphical properties

## Cartesian/radial axes

A visual guide for the **placement** of elements composing the visualisation.

A visualization with axes is called **graph**, otherwise it is called chart (although chart is used for any kind of visual).



# Layout

02

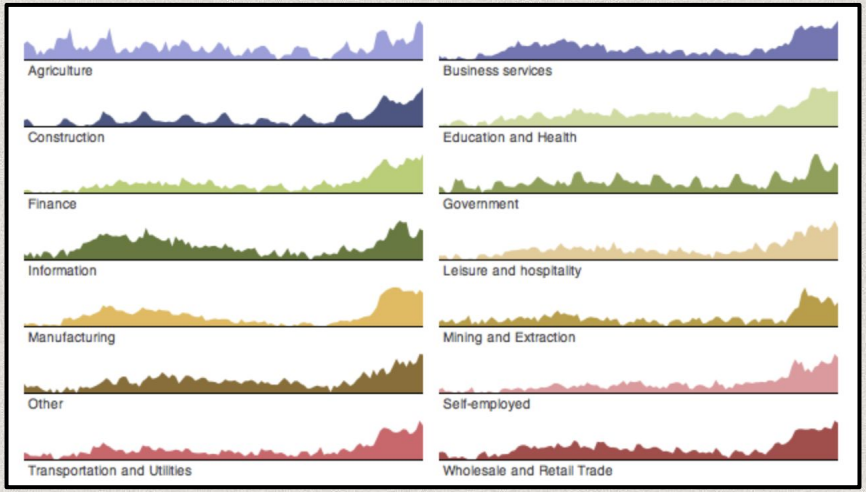
Graphical properties

## Single/Multiple canvas

The **format and symmetry** of the visualisation change according to the volume of data and the number of attributes to show.

More visualisations in the canvas help comparison, but hide the big picture!

A viz. should always fit the frame (e.g. a screen).



# Shape

02

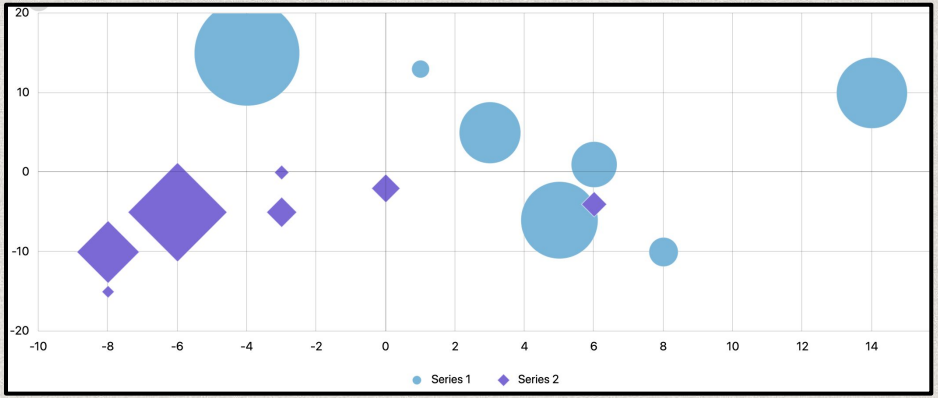
Graphical properties

## Shapes and glyphs

Realistic (icons) or abstract (circles) help to **distinguish** elements.

Shapes and glyphs demand more **attention** (icons can mitigate the effort)

**Lines** are effective shapes (highlight a trend)



# Colour

02

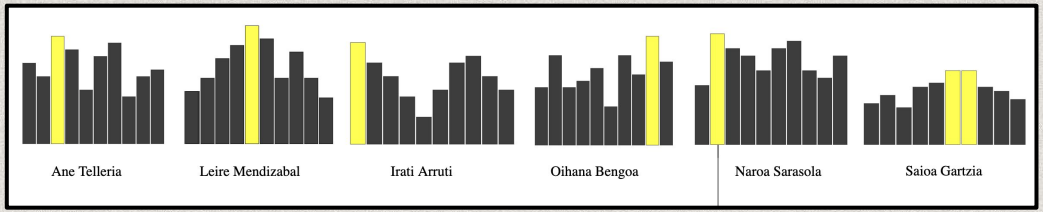
Graphical properties

## Palette and luminosity

Color differences are detected in <200 milliseconds (preattentive perception). Must be natural (e.g. found in nature).

Used to **distinguish elements or patterns** in big datasets, it’s less useful in small datasets (one color is sufficient).

**Luminosity** highlights relevant patterns (bright colors pop out, while dark colors recede).



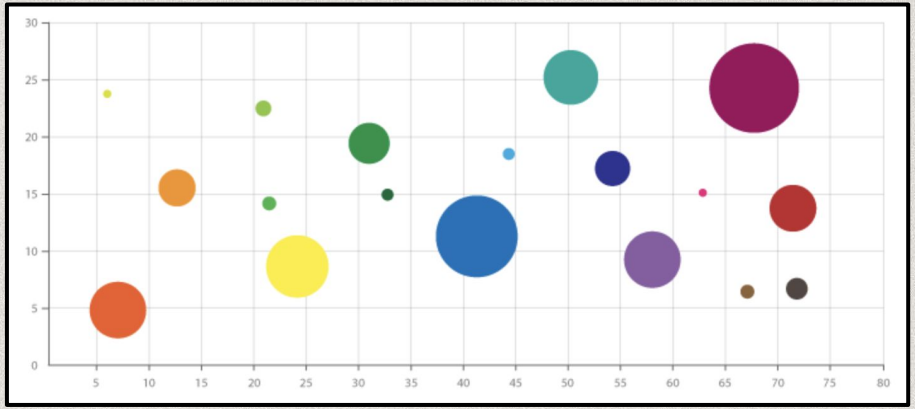
# Size



## Show the variety

Identify **variations between similar elements** by size (regardless of other dimensions) is the quickest way.

Still, it is not the most effective means to compare values (e.g. circle areas).



# Typography



Graphical properties

## Primary/secondary role

Can be a primary aspect (e.g. wordle) along with other dimensions (e.g. size) or secondary, to describe the visualization.

# Interaction and context



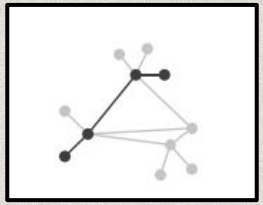
Graphical properties interact with each other to convey the message in different contexts.

## Location

**Placement** helps to familiarise with data (e.g. a map)

## Network

**Closeness** and **connections** of data points (e.g. a network graph)



## Sequence

Axes-based vis. create a **linear** reading (e.g. line graphs)

# Graphical properties + Context =

# Visual pattern

By means of graphical and contextual aspects, what should emerge from visualisations is a **visual pattern**. Several co-occurring dimensions can contribute to highlight patterns.

# How do visual patterns look like

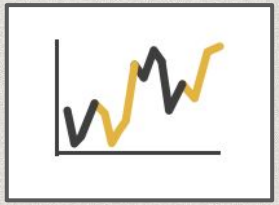
02

Visual patterns

## Differentiation

**Contrast** values

distribution



samples

similarity

# How do visual patterns look like

02

Visual patterns

## Gradation

**Continuous** values

trend

# How do visual patterns look like

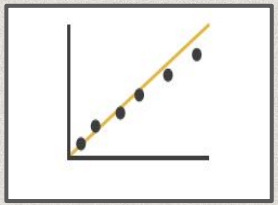
02

Visual patterns

## Anomaly

Break the **pattern.**

outliers



correlation

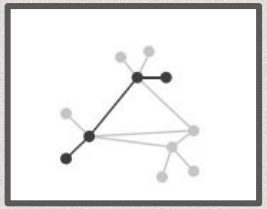
# How do visual patterns look like

02

Visual patterns

## Paths

**Position** and relations



Clusters and paths

# Perception

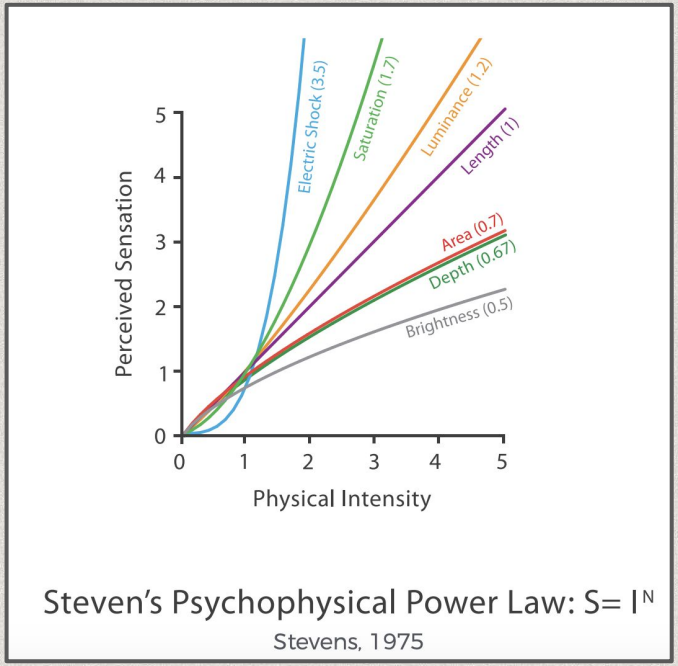
Preattentive variables

Visual aspects have a **functional** role in the interpretation of data rather than aesthetic.

The idea of visualising data in a graphical form is to **replace cognition with perception**.

The human eye perceives visual dimensions in different ways, with more or less effort, and users' **knowledge background** can affect the perception.

Nonetheless, some dimensions are more **intense** than others.



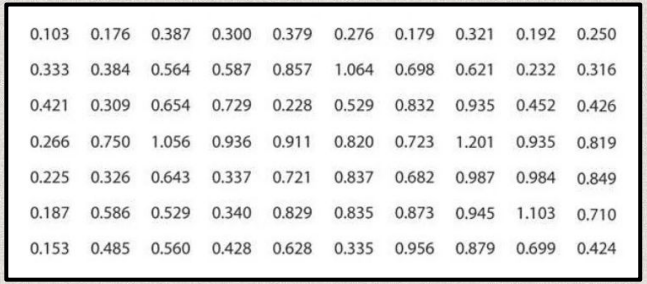

Steven's power law

# Tidwell's preattentive variables

02

Preattentive variables

Some visual aspects work **preattentively**, meaning they are able to communicate something before the user pays conscious attention to it.



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.103 | 0.176 | 0.387 | 0.300 | 0.379 | 0.276 | 0.179 | 0.321 | 0.192 | 0.250 |
| 0.333 | 0.384 | 0.564 | 0.587 | 0.857 | 1.064 | 0.698 | 0.621 | 0.232 | 0.316 |
| 0.421 | 0.309 | 0.654 | 0.729 | 0.228 | 0.529 | 0.832 | 0.935 | 0.452 | 0.426 |
| 0.266 | 0.750 | 1.056 | 0.936 | 0.911 | 0.820 | 0.723 | 1.201 | 0.935 | 0.819 |
| 0.225 | 0.326 | 0.643 | 0.337 | 0.721 | 0.837 | 0.682 | 0.987 | 0.984 | 0.849 |
| 0.187 | 0.586 | 0.529 | 0.340 | 0.829 | 0.835 | 0.873 | 0.945 | 1.103 | 0.710 |
| 0.153 | 0.485 | 0.560 | 0.428 | 0.628 | 0.335 | 0.956 | 0.879 | 0.699 | 0.424 |

Find numbers greater than 1

# Tidwell's preattentive variables



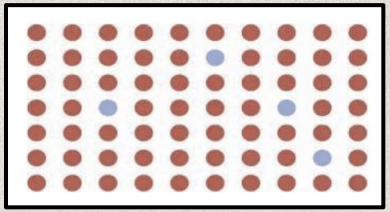
Preattentive variables

Effective and impactful data visualisations work extensively on these aspects.

Tidwell et al, Designing interfaces found 8 variables.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.103 | 0.176 | 0.387 | 0.300 | 0.379 | 0.276 | 0.179 | 0.321 | 0.192 | 0.250 |
| 0.333 | 0.384 | 0.564 | 0.587 | 0.857 | 1.064 | 0.698 | 0.621 | 0.232 | 0.316 |
| 0.421 | 0.309 | 0.654 | 0.729 | 0.228 | 0.529 | 0.832 | 0.935 | 0.452 | 0.426 |
| 0.266 | 0.750 | 1.056 | 0.936 | 0.911 | 0.820 | 0.723 | 1.201 | 0.935 | 0.819 |
| 0.225 | 0.326 | 0.643 | 0.337 | 0.721 | 0.837 | 0.682 | 0.987 | 0.984 | 0.849 |
| 0.187 | 0.586 | 0.529 | 0.340 | 0.829 | 0.835 | 0.873 | 0.945 | 1.103 | 0.710 |
| 0.153 | 0.485 | 0.560 | 0.428 | 0.628 | 0.335 | 0.956 | 0.879 | 0.699 | 0.424 |

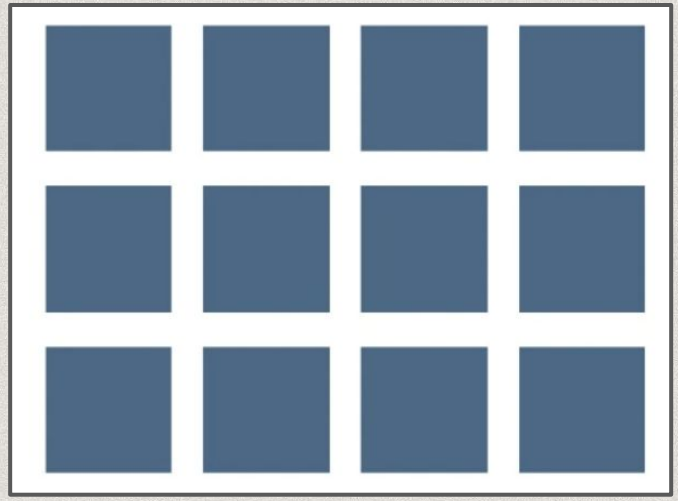
Find numbers greater than 1



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.103 | 0.176 | 0.387 | 0.300 | 0.379 | 0.276 | 0.179 | 0.321 | 0.192 | 0.250 |
| 0.333 | 0.384 | 0.564 | 0.587 | 0.857 | 1.064 | 0.698 | 0.621 | 0.232 | 0.316 |
| 0.421 | 0.309 | 0.654 | 0.729 | 0.228 | 0.529 | 0.832 | 0.935 | 0.452 | 0.426 |
| 0.266 | 0.750 | 1.056 | 0.936 | 0.911 | 0.820 | 0.723 | 1.201 | 0.935 | 0.819 |
| 0.225 | 0.326 | 0.643 | 0.337 | 0.721 | 0.837 | 0.682 | 0.987 | 0.984 | 0.849 |
| 0.187 | 0.586 | 0.529 | 0.340 | 0.829 | 0.835 | 0.873 | 0.945 | 1.103 | 0.710 |
| 0.153 | 0.485 | 0.560 | 0.428 | 0.628 | 0.335 | 0.956 | 0.879 | 0.699 | 0.424 |

# Tidwell's preattentive variables

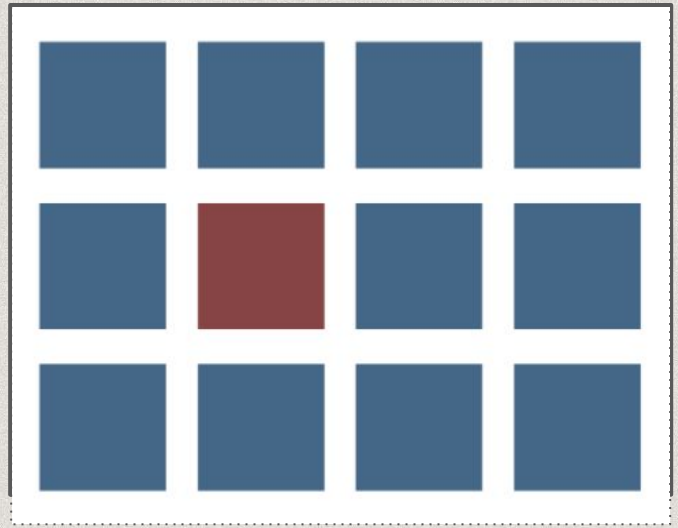




Let's start!

# Tidwell's preattentive variables



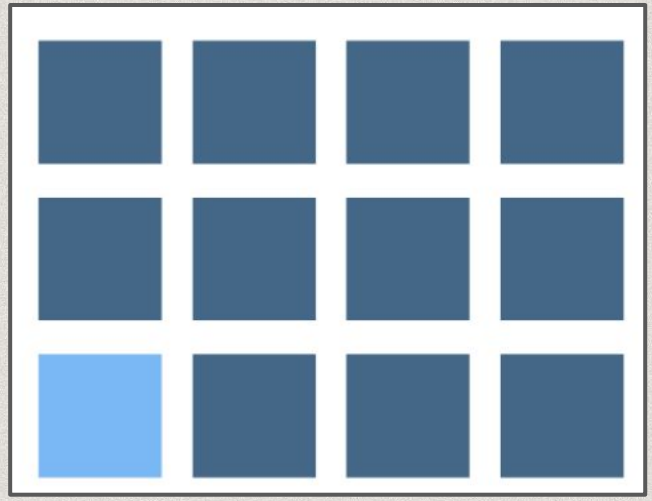


Color hue

# Tidwell's preattentive variables





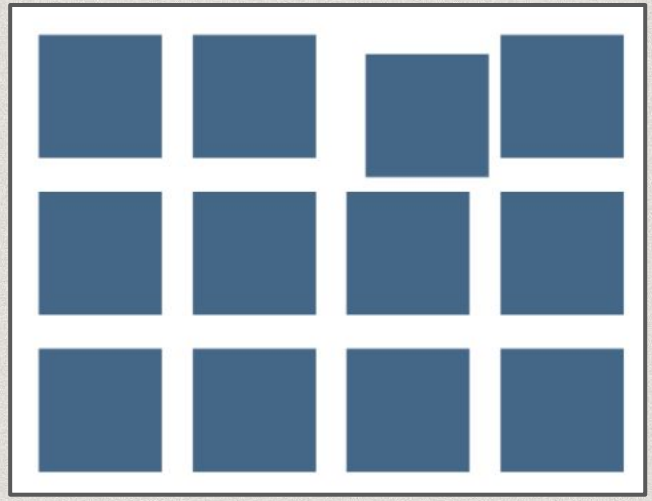


Color brightness

# Tidwell's preattentive variables





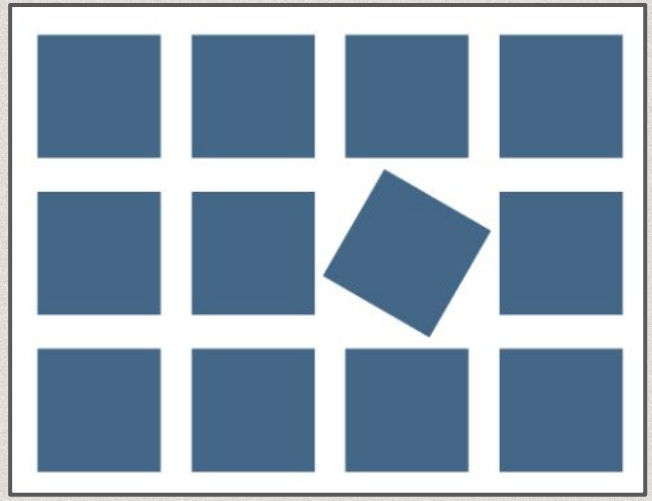


Position and alignment

# Tidwell's preattentive variables







Orientation

# Tidwell’s preattentive variables



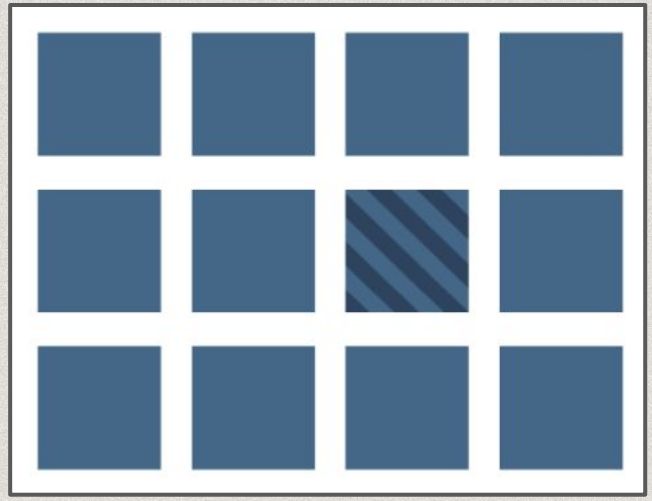
Preattentive variables

Color saturation

# Tidwell's preattentive variables





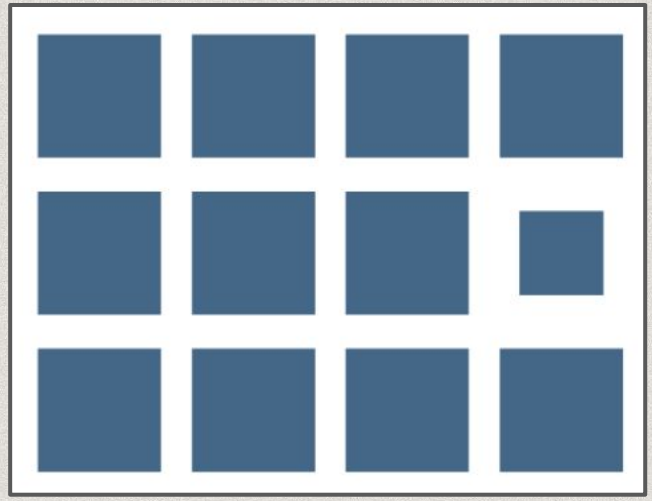


Texture

# Tidwell's preattentive variables





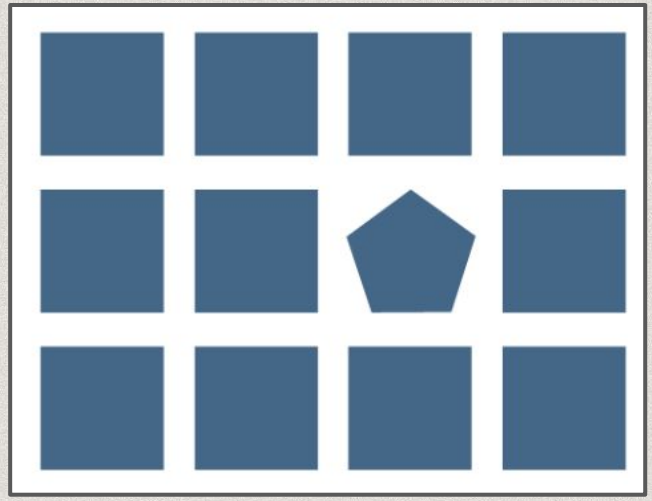


Size (or width, length)

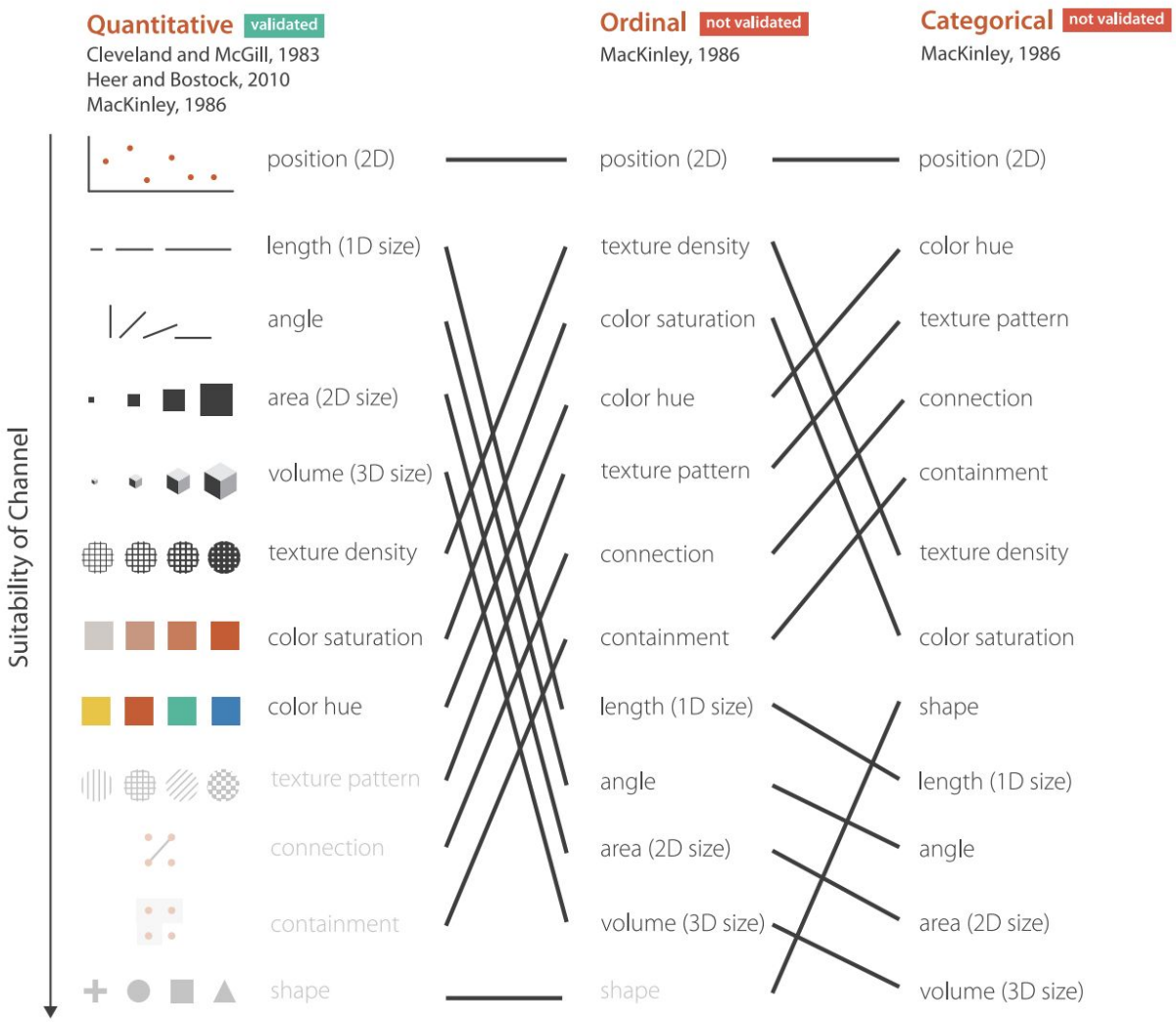
# Tidwell's preattentive variables







Shape

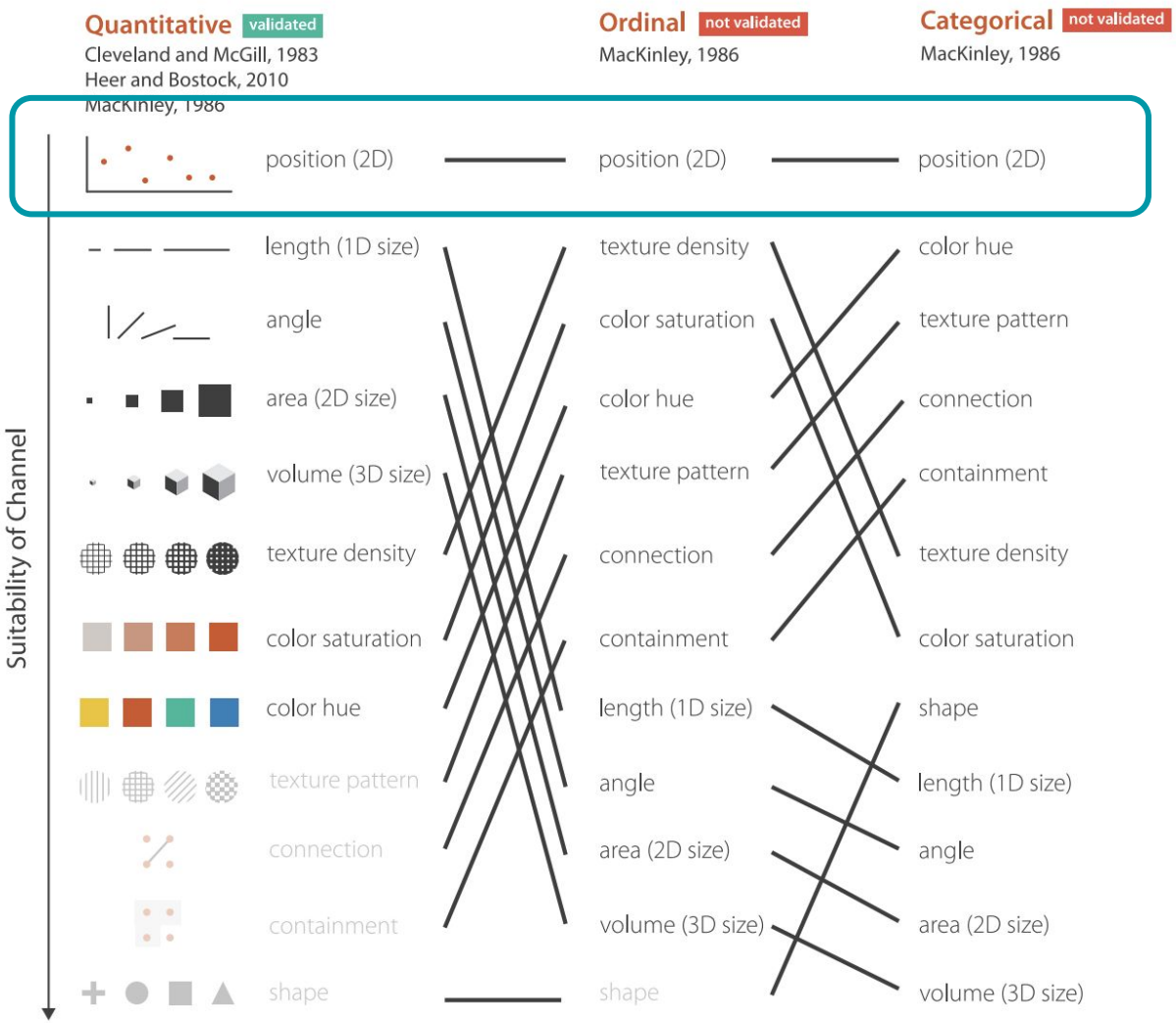



# 03

## Best practices

Studies demonstrated that according to the **type of data attributes** at hand, certain **graphical properties** work better than others.

In particular, properties are perceived with more or less **accuracy**.

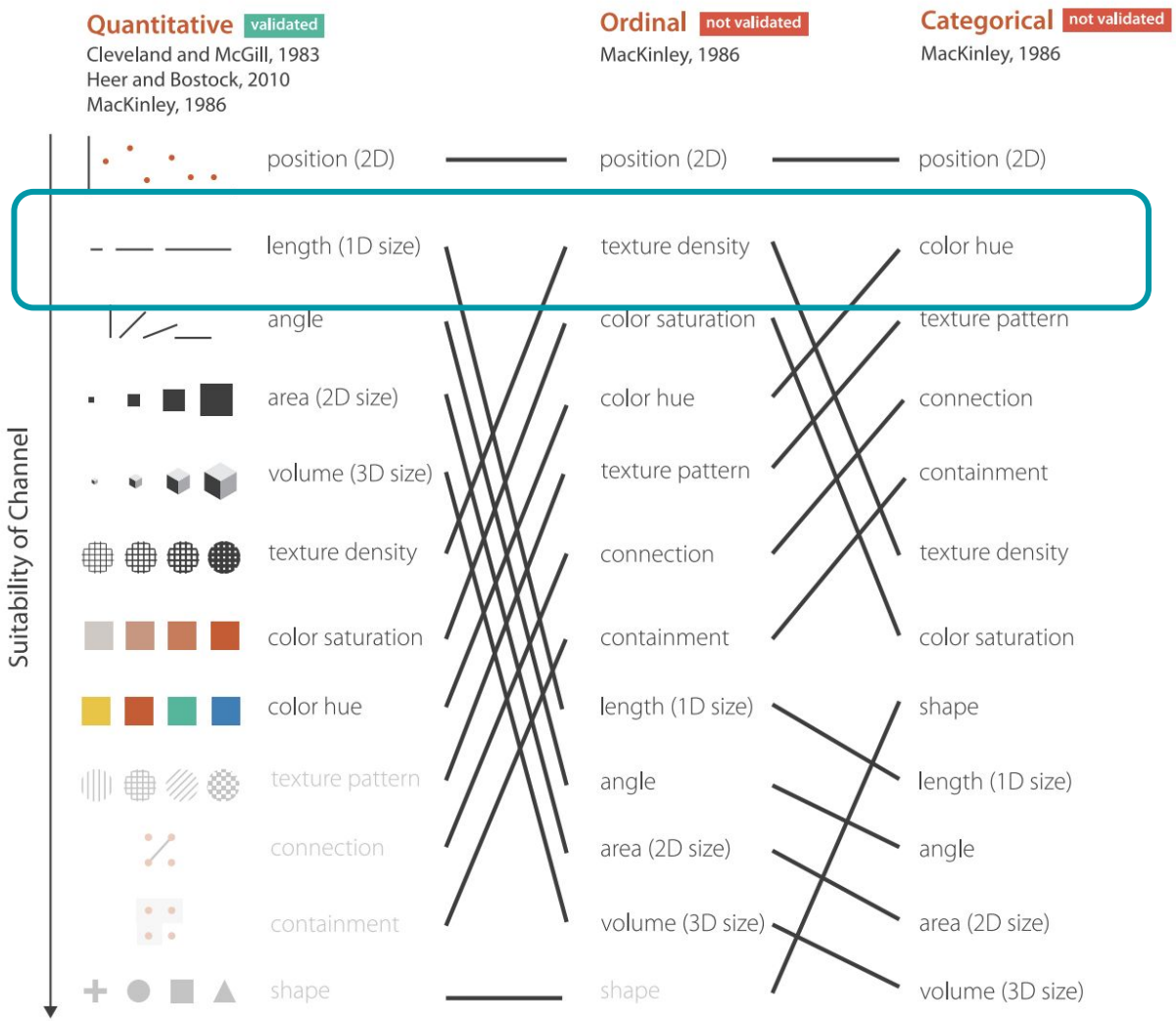



# 03

Best practices

**Relative location** (position) of elements using the same scale is the most effective property regardless of the data type

(maps, axes…)




# 03

## Best practices

**Length** works better than angles (e.g. bar chart VS pie chart) when comparing numeric values.

**Colors** help to discriminate non quantitative values, as well as texture (gradation) for ordinal values.

# What do you see in these charts?

02

Test

Answer the questionnaire (15 minutes)

https://forms.gle/s7mWD5mLtEyY6iL37

# Thanks!

Do you have any questions?

marilena.daquino2@unibo.it

https://github.com/marilenadaquino/information_visualization

**CREDITS:** This presentation template was created by **Slidesgo,** and includes icons by **Flaticon,** and infographics & images by **Freepik**