Information Visualization 1 : Foundations

Introduction

Growth in many different areas

- Data Journalism (New York Times amazing graphs…)
- Products : Tableau, Qlik / Open source : ggplot2, D3
- Companies : strong visualization groups (Netflix, Uber…)
- Independant Designer

Common idea :
Transforming data into something that enhances the comprehension of what is described by the data

Goal of specialization :
Design, evaluate, develop interactive visualisations to help people generate insights and then communicate these insights to other people as effectively as possible

5 main objectives :

- How to use graphs appropriately (eg. Choose the right graph)
- Evaluate visualisations designs
- Innovate : come up with new visualisations methods and techniques
- Code information visualisation
- How to go from specification of a problem to transforming this problem and the data into Information Visualization

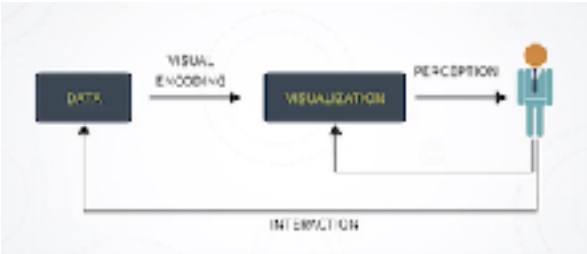
---

DEFINITION

Readings in Information Visualization:
Using Vision to Think, "the use of a computer-supported, interactive, visual representations of abstract data to amplify cognition."

- Computer-based
- Visual Representation
- Abstract Data
- Interactive
- Amplify Cognition

THE INFOVIZ PIPELINE (DIAGRAM OF Data Visualization PROCESS)



The interaction with the data AND the visualization enables him to answer different questions.

KEY CONCEPT : COMPUTER BASED GRAPHICAL REPRESENTATIONS AND VISUALIZING ABSTRACT DATA

- Abstract data
  - no obvious/natural visual representation
  - Time, duration, project name, type of activity

    - https://medium.com/@FILWD/quantifying-and-visualizing-deep-work-af4689a62423
        - Visualisation avec GS basique. ‘Deep work’ = seulement si enregistre dans GS. Insights: horaires lieux, productifs, pomodorros.
        - Benefice : permet d’avoir une idée de ses patterns (ex. Plus de sessions le matin mais courtes / moins de sessions l’aprem mais longues)
        - Savoir si travail efficace: mesurer un autre critère (ex. Nb mots/duree)
- ≠ Physical objects/phenomena :
    - Human skull, météo
    - realism

KEY CONCEPT : INTERACTIVITY

- What: variables, rows, columns
- How: per year

KEY CONCEPT : AMPLIFYING COGNITION

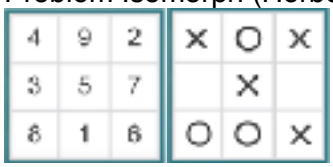
Cognitive artefacts : tools that help us think, for example an abacus.

Exercise : Try to multiply 34 x 72 using exclusively your mind
Why is it easier on paper than with the mind exclusively ? Because we can store intermediary results in the paper rather than keeping information in mind.

Distributed cognition : our cognition is also made of the artefacts in the world around us.

In visualization : a way to store information out of our mind and make it accessible through our eyes and through manipulation with interactive systems.

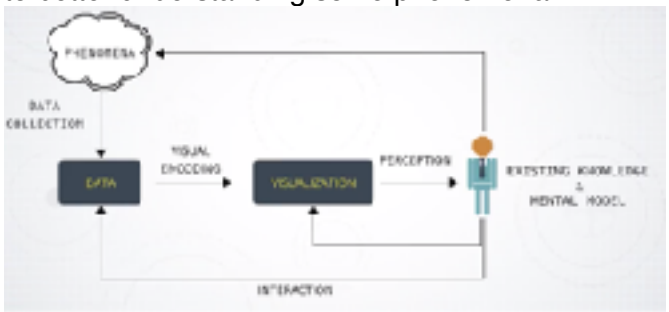
Problem Isomorph (Herbert Simon, Nobel Prize) : 2 problems are exactly the same



Game of 15”

WHY VISUALIZE DATA?

The ultimate interest is using data as **a way to understand some phenomena**.
People who are using visualization come with some pre-existing knowledge and some goal related to better understanding some phenomena.



WHY USE VISUALIZATION?

- **Explanatory**: main purpose is to explain something to somebody else, message with an appropriate design
- **Exploratory**: need to extract data but don’t know the content

- **Confirmatory**: hypothesis that needs to be checked out

EXAMPLE OF EXPLANATORY VISUALIZATION

https://www.nytimes.com/interactive/2017/03/21/climate/how-americans-think-about-climate-change-in-six-maps.html

---

EXAMPLES OF EXPLORATORY AND CONFIRMATORY VISUALIZATIONS

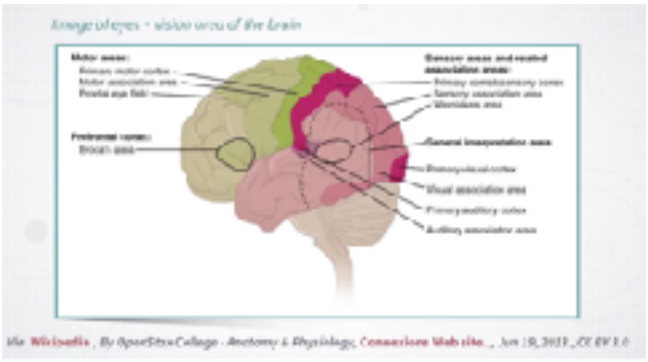
Data analysis.
We often start with a question and answer with a new question:

- Why do trafic collision go down in winter time.
- Hypothesis (confirmatory analysis): there are less cyclists and pedestrians in winter.

EXAMPLES OF EXPLORATORY VISUALIZATIONS AND TOOLS

https://driven-by-data.net
https://duckduckgo.com/?q=nicky+case+explorable+explanations&t=osx&ia=web
https://explorabl.es
https://pudding.cool/process/pudding-awards-2018/
http://polygraph.cool
https://www.propublica.org

WHY USE A GRAPHICAL REPRESENTATION?



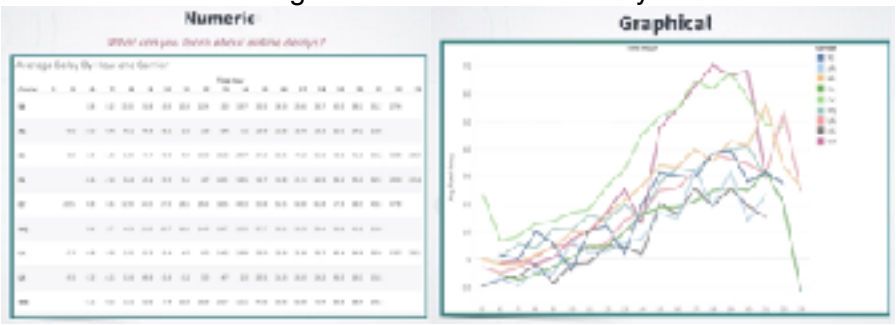
We are very good at visual

What are other ways :

- Verbally : we are forced to process the information sequentially
- Numerically : we have to process numbers that don't match directly to some quantities
- Graphically : we are processing information in a parallel fashion, we don't need to go through it sequentially necessarily

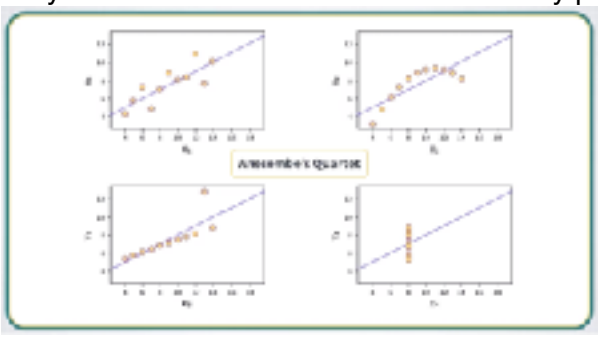
Verbal : "The average delay increases steadily throughout the day starting from 6am around 9 pm. Much higher delays take place between 3pm and 9 pm and among the airlines we considered EV and WN experience much steeper increases. EV experiences in general much higher delays than

all the other airlines regardless the time of the day…”



PROBLEMS WITH SUMMARY STATISTICS

Used very often to convey information, but problem with aggregate statistics.
Only when we visualise data we can readily perceive information that is hidden in the statistics.



WHY USE COMPUTERS TO VISUALIZE DATA?

Automatic vs Manual :

- A lot of data
- Interactivity

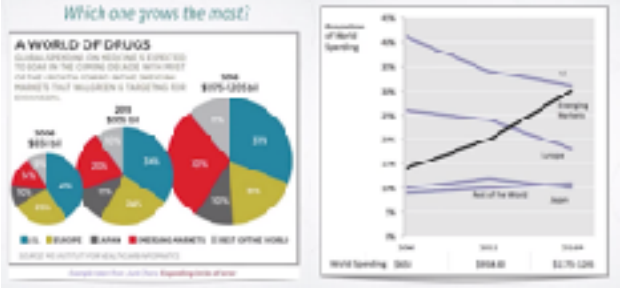
WHY USE INTERACTION?

Not all questions can be answered by looking at one visual representation.
Interaction helps the user answer multiple questions.

Interactivity in visualization is an active area of research.
How to interact with tablets, small phones, very large screen ...

ASSESSING THE QUALITY OF A VISUALIZATION

Some visual representations are better than others at solving particular problems…



Pie chart :

- We see the quantity as the area, but the real quantity is given by the angle
- Angle and size interfere...

When evaluating, assessing the quality of a visual representation:

- what is the intent
- what problem am I trying to solve
- what information am I trying to convey

Designing effective visualizations requires two main step :

1. **Explore the design space**

You typically come up with a first idea, a first design but if you don't know enough of the design space, it's very hard to create alternatives.
Being able to create a certain number of alternatives is a crucial skill for visualization design, because you need to see different solutions and start assessing them.

2. **Compare the solutions**

You have to be able to predict whether a given visual representation is actually going to be more effective than another. There are many ways, but the most important skill is to understand how human perception of graphical representation works.

Rest of the course :

- the design space,
- what visual representations are available for a certain type of data and problems,
- learning enough about human perception that you can start reasoning about whether and why a given visual representation may be more effective than another.

READING : A TOUR THROUGH THE VISUALIZATION ZOO

https://queue.acm.org/detail.cfm?id=1805128

Data Abstraction

REFLECTING ON DATA

Data definition:
"Factual information used as a basis for reasoning, calculation and discussion”
"General concept that refers to the fact that some existing information or knowledge is represented or coded in some form suitable for better usage or processing."

Data **processing**’s 3 main stages:

- Collection
- Transformation
- Encoding

WHAT IS DATA ABSTRACTION?

A method to describe data :

- help you decide what operations and encoding methods are available and also appropriate.
- helps you figure out what transformations are possible and what visual representations are also possible
  - Ex. Column -> borough = 1 column = 1 category
  - -> bar chart appropriate for categorical visual

A way to recognise common structures in data coming from different domains, hence the name “data abstraction”

- Ex. network
- Description : connections
- From different domains: social connections, between molecules, between people in a criminal organisation

---

DATASETS TYPES: TABLES AND NETWORKS

- **Tables**: all items have same number of attributes and each attribute column has the same type
- **Networks** & **Trees**: nodes (items) are connected by links, both can have attributes

---

ATTRIBUTE TYPES

- categorical
- ordinal
- quantitative

ATTRIBUTE SEMANTICS

- Spatial and temporal
  - ‘Region’ = categorical + spatial
  - ‘GPS Coordinates’ = quantitative + spatial
- Sequential
- Diverging: for a given quantity, it is possible to identify a middle value
- Cycling: months over many years
- Hierarchical
  - Categories, sub-categories

DATA PROFILING & QUALITY

Data **profiling** / familiarization : understand the meaning of data

Data quality and “**wrangling**” : for ex. fix outliners, can also take a lot of time

Data manipulation -> Not covered in the course

RECAP

What is Data?
Data in visualization
Data abstraction
Datasets = items + attributes
Dataset and attribute type

Fundamental Graphs and Data Transformation

HOW TO VISUALIZE?

1. What = Select/Transform
2. How = Choose/Design representation

FUNDAMENTAL GRAPHS

5 graphs :

- Solve very large percentage of vis problems
- Training ground for more sophisticated graphs

Are combinations of 2 or more attributes

1. Bar chart : how a quantity distributes across a set of categories
2. Line chart : how a quantity changes in relation to another quantity (typically time)
3. Scatter plot : how a quantity relates to another quantity
4. Matrix : how a quantity distributes across two categories
5. Symbol map : how a quantity distributes across two spatial coordinates

---

ALTERNATE REPRESENTATIONS

Very important skills :

- Being able to create different representations of the same data
- How to reason in terms of alternative strategies for the same dataset

Ex:
Pb with line chart and categorical data that conveys a pattern that is not meaningful
Effective replacement : dot chart

Area :

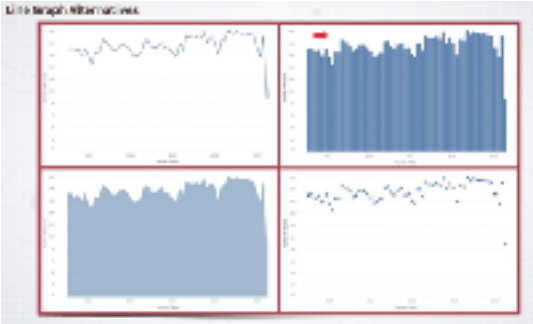
1. Reader less comfortable
2. Areas is not as effective to convey quantitative info than comparing the length and height

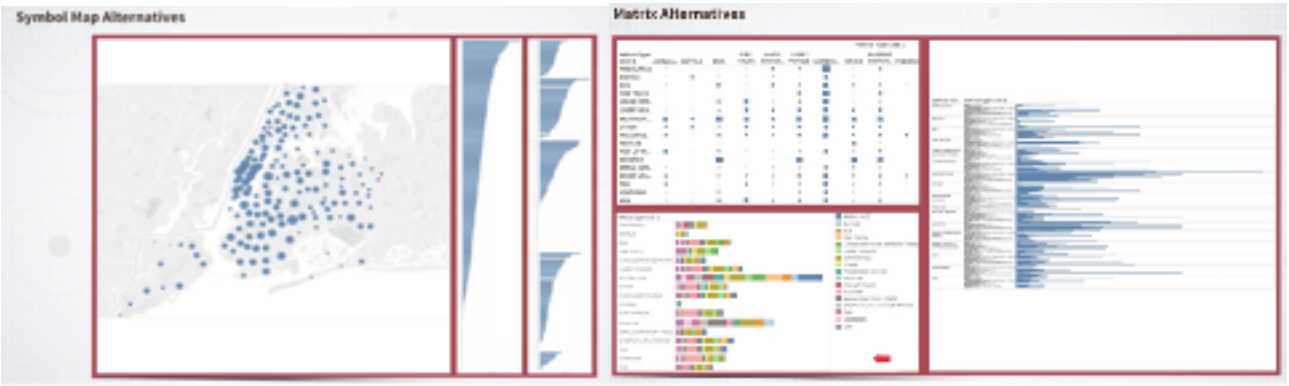
Slope Chart (scatter plot alternative):

- Useful to see when values are going in one direction
    - Line that goes up = teams that didn't have a good win/lose ratio but spent a lot of money hiring their players

Line Graphs alternatives:

- Bar chart : useful to read values accurately
- Area chart
- Simple dots





GOING BEYOND TWO ATTRIBUTES

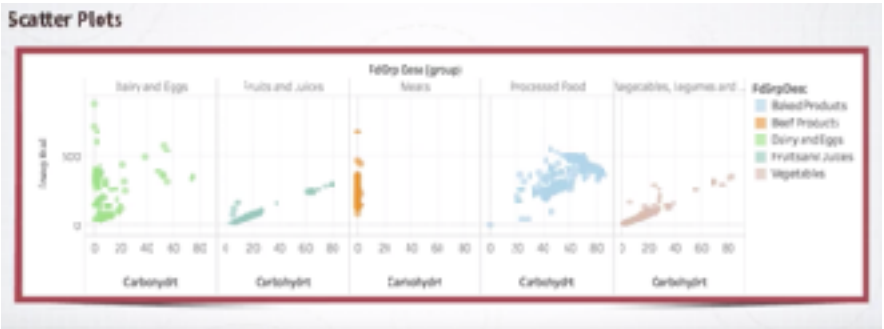
Vehicle type / main cause of accident / how many injured

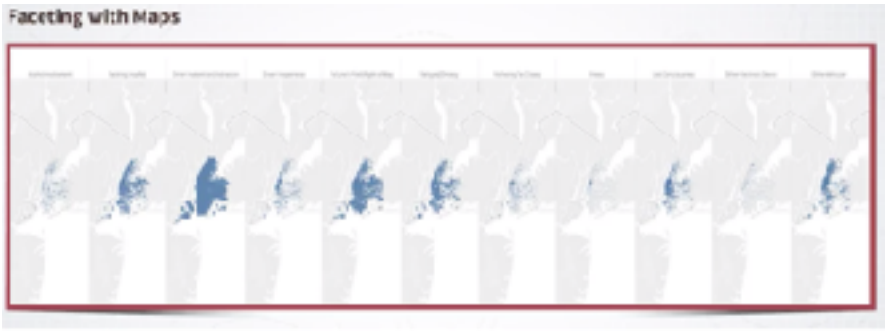
- Stacked : if main question regarding proportions (part to whole)
- Bar : comparing single value

Additional categorical attribute :

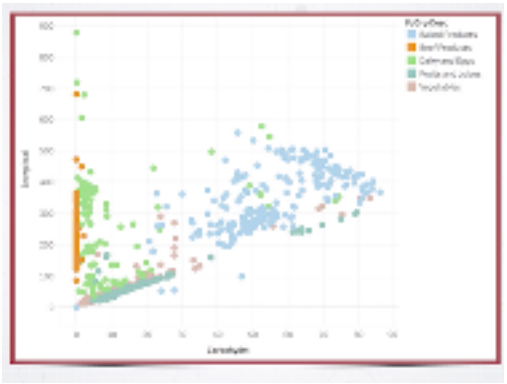
1. Color lines
2. Series of line charts on top of the other
3. Stacked line chart (since the values are stacked, the pattern that you see in the lines is affected by the shape of the line that is below the one that you are observing)
    - good for : if the only question that you have is how does the proportion of these values change over time?
    - not good for :
        - reading the individual values of these categories over time,
        - to compare the values of these categories over time.

SCATTER PLOTS + FACETING





To visualize another attribute : colour



Faceting / small multiples : can be applied to any graph

- Select one categorical or ordinal attribute
- Create as many sets as the number of values
- Create one plot for each of the values

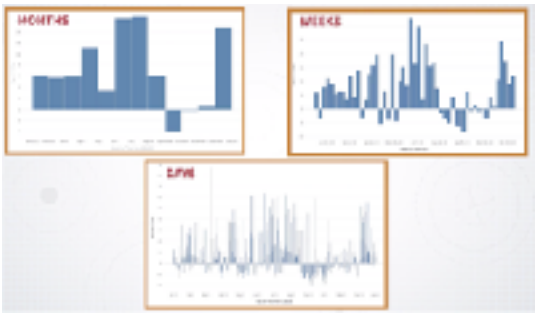
DATA TRANSFORMATION

Selection : selecting the attributes you need
Ex: visualize the relationship between carbohydrates and calories and see how it is affected by food categories, where every single item in you data set is a food product
Scatter plot
Transformation : intermediate step often needed : aggregate two columns together and make the mean

---

COMMON USEFUL DATA TRANSFORMATIONS

Part of the design process. Most projects require you to think creatively about how transformation may lead to better communication and understanding.

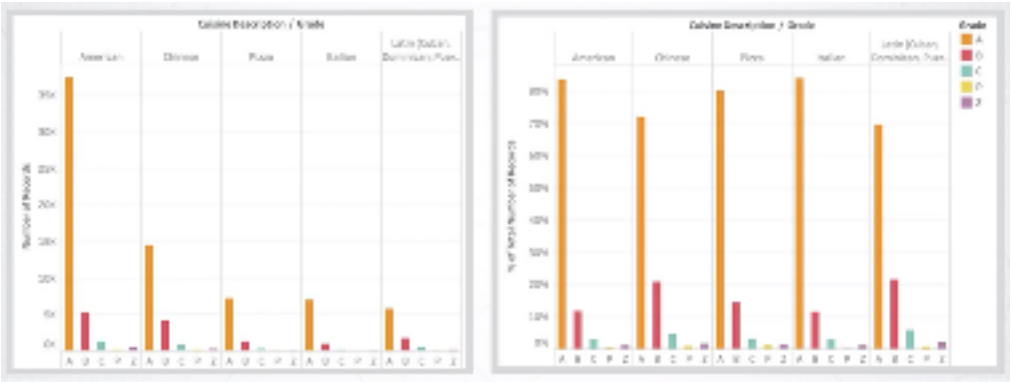


- Aggregation

Time and date at the same time; both have hierarchical structure; both can be aggregated at

different level. Aggregate by: Hours, days, weeks, month

- Geo coding / decoding
- Binning : transforming a quantitative attribute into an ordinal one
- Rescaling / re-expression:
    - Normalization: transform the range into [-1; +1]
    - Percentages : transforming quantities into %
    - Distance from reference : calculate the average and re-express in terms of distance from the average value



Visualization is not only how (graph) but what to visualize (data, quantities, percentages, etc).
Data transformation :

- Selection
- Aggregation

SUMMARY

Data abstraction is useful to design and use appropriate graphs.

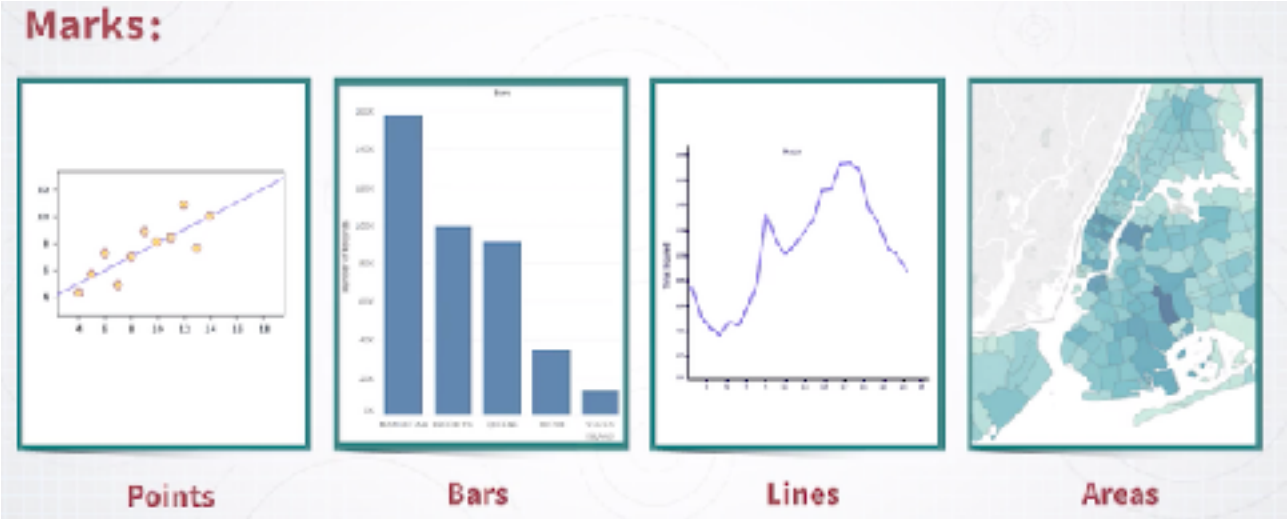
Marks & Channels

Data Items <-> visual Marks
Data Attributes <-> visual Channels

MARKS & CHANNELS

The alphabet

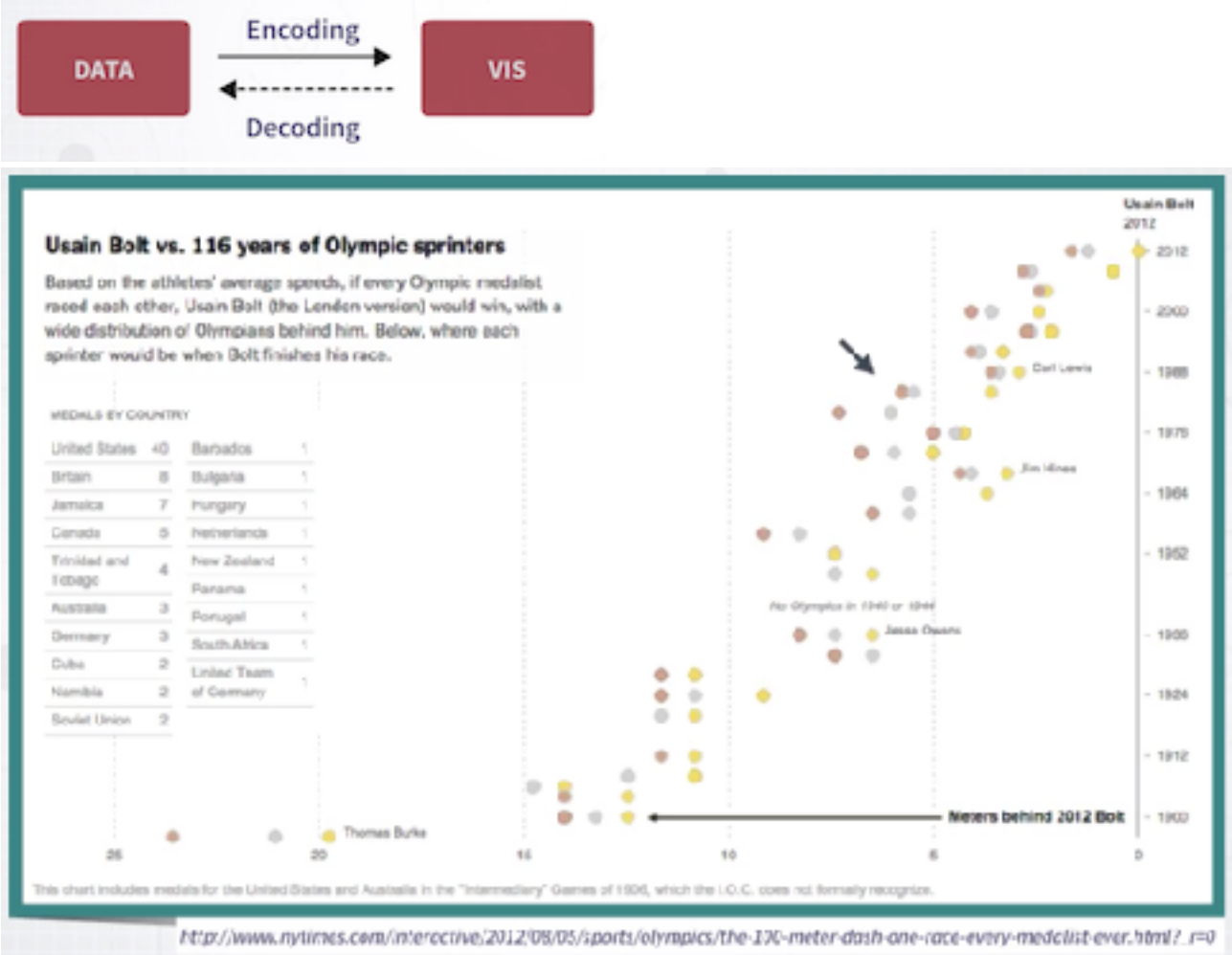
MARKS: Visual objects that represent data items.



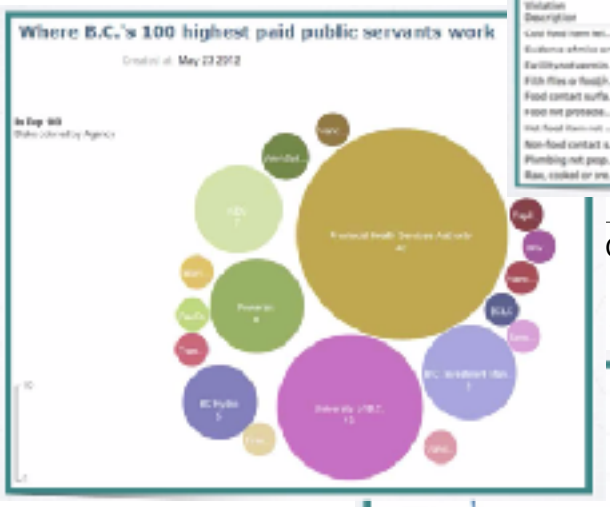
CHANNELS: encode data attribute

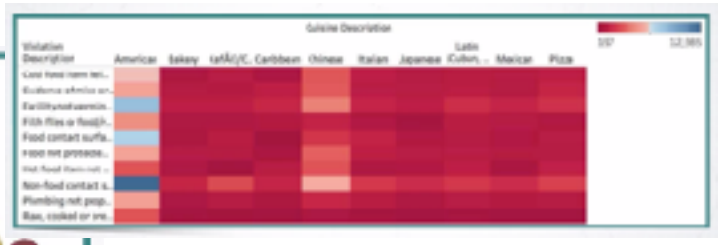
- Position
- Size (length, width, area)
- Angle and Slope
- Color (intensity and hue)
- Shape and Texture

GRAPHICAL “DECODING”



- Marks: medals
- Visual Channels : x, y, colour
  - // Data attributes : distance, time, type

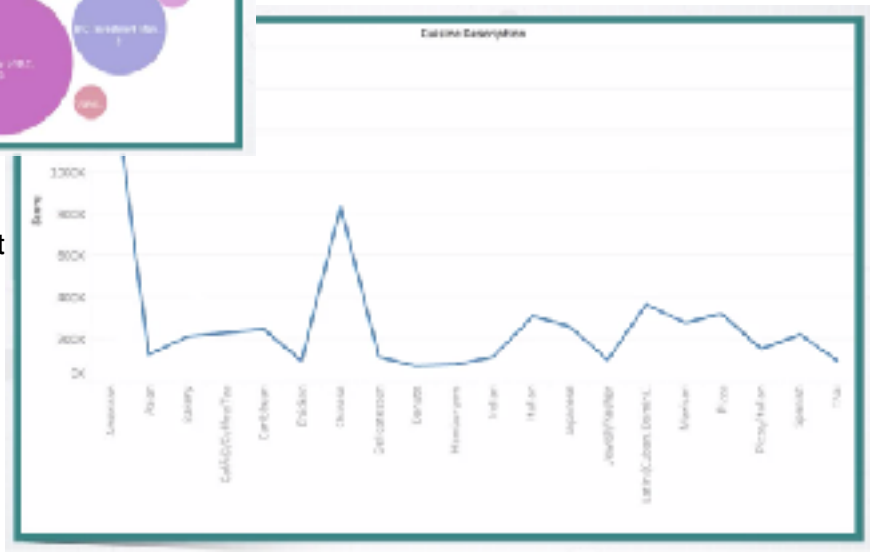




QUALITY OF VISUAL ENCODING

- **Expressiveness**:

Represent all and only the relationships that exist in the data (ex. Ordered data shouldn't appear unordered / Unordered shouldn't appear ordered …)



Line chart :

Something continuous / change over time.

Order here is not meaningful.

Information (pattern) is not contained in the data.

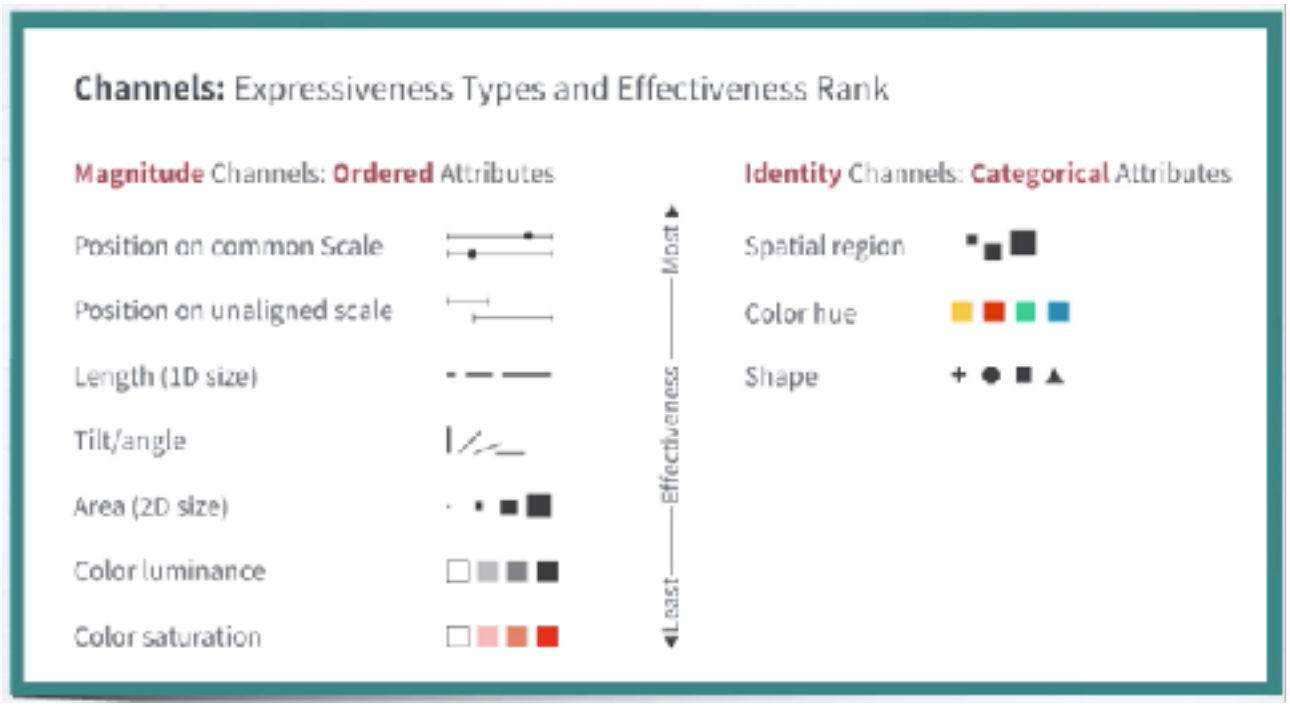
Divergent color map

Used when there is a middle value, but here there is only 3 blue cases so no middle value.

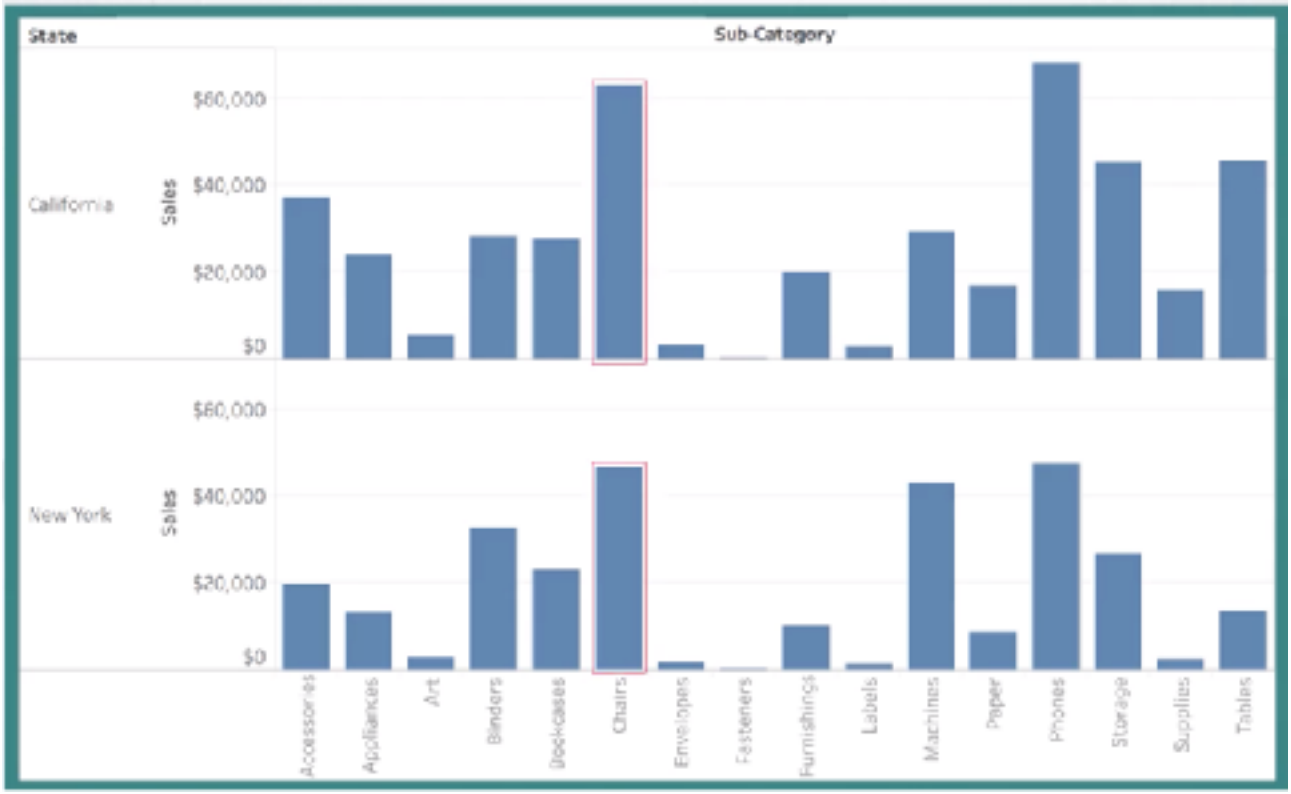
Bubble map where the colour doesn't mean anything.

- **Effectiveness**:

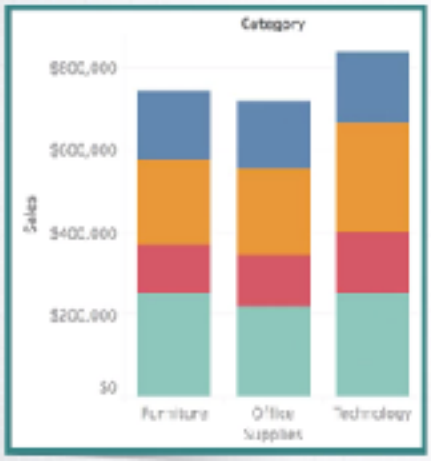
Relevance of information should match the effectiveness of the channel.



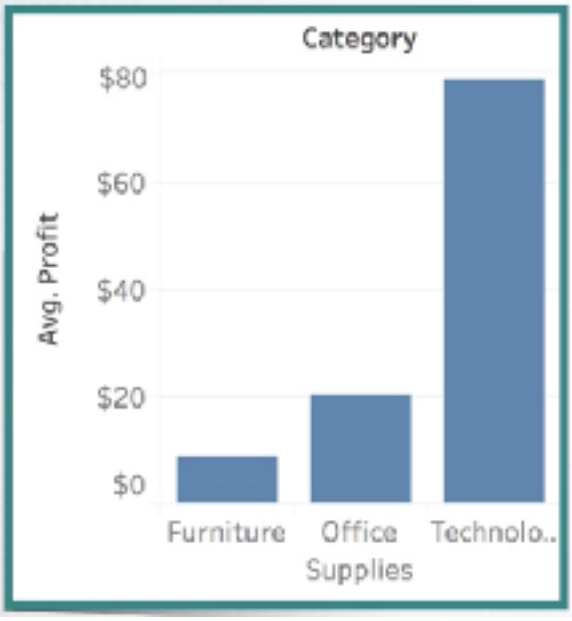
Position on unaligned scale

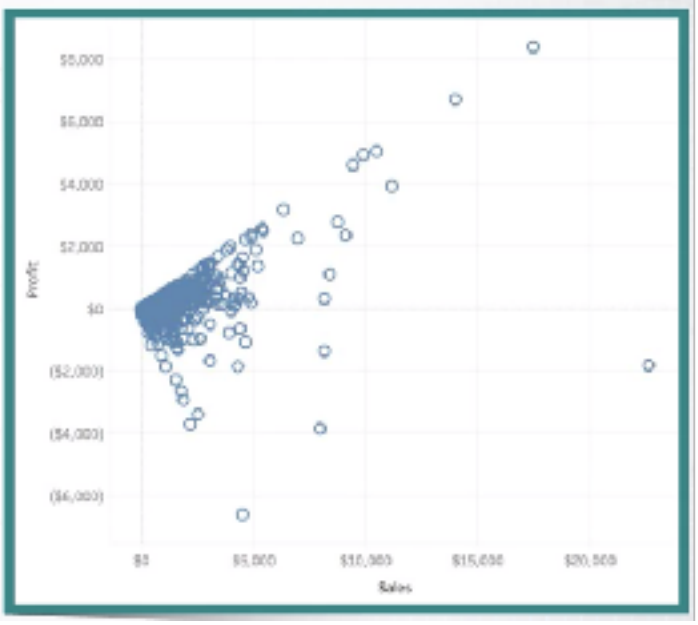


Length

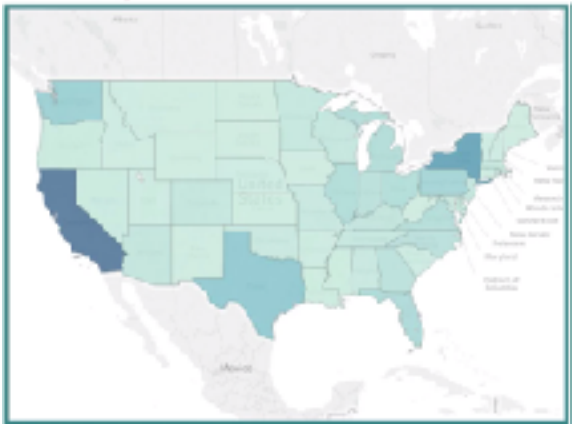


Position on common scale

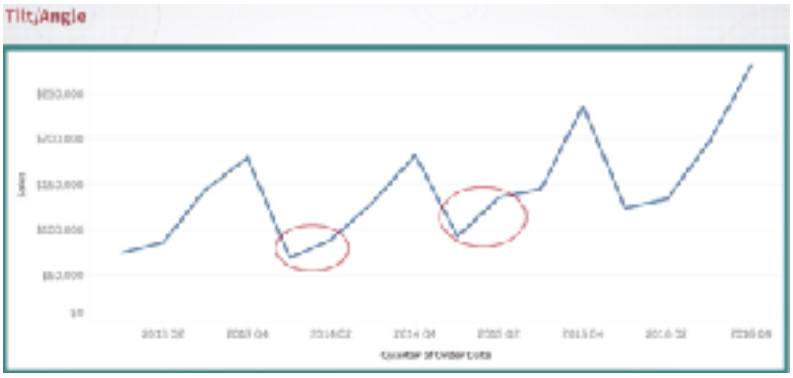




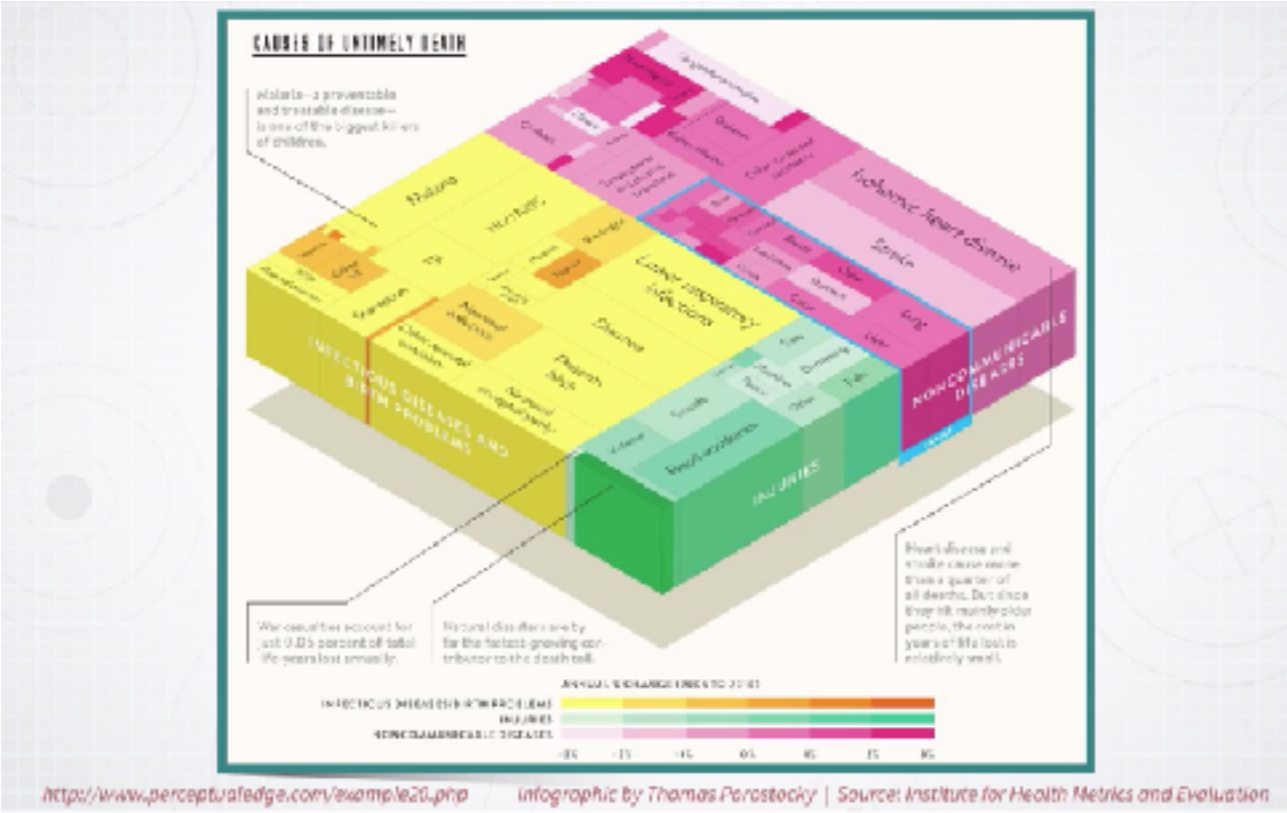
Color intensity



Area



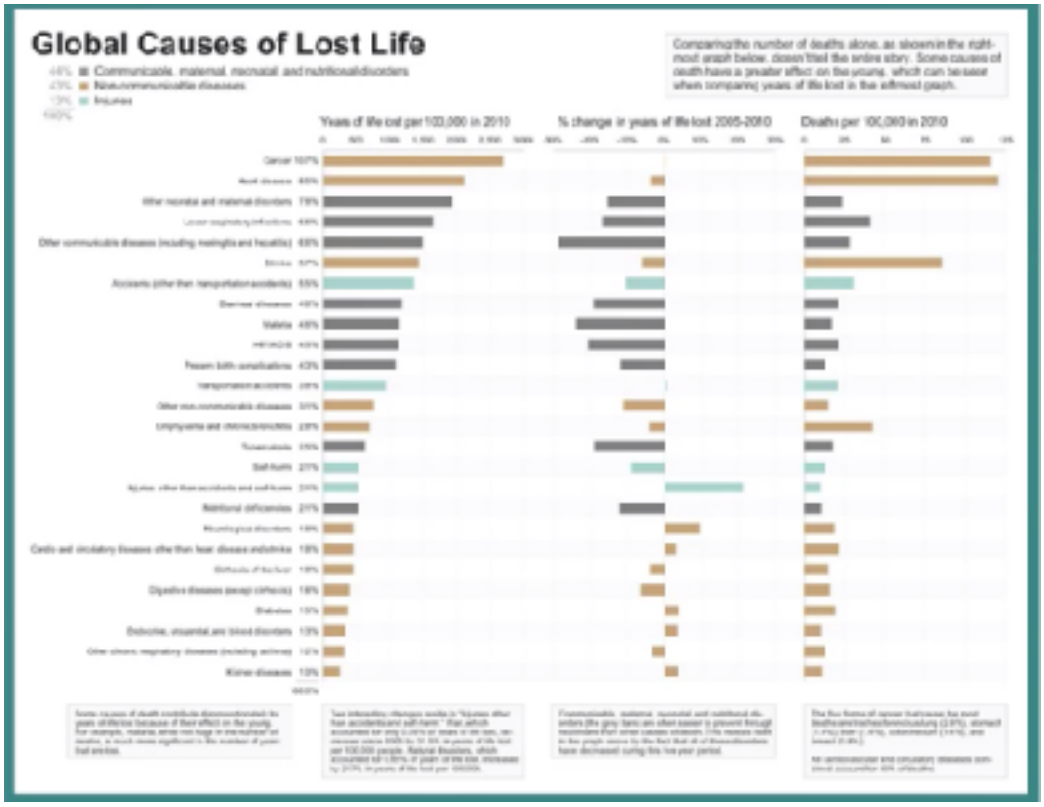
EVALUATE VISUALIZATIONS



- Area shows years of life lost, give more weight to cases when people die young.
- Categories: injuries, infectious diseases...
- Colour intensity: amount of change from one year to another.

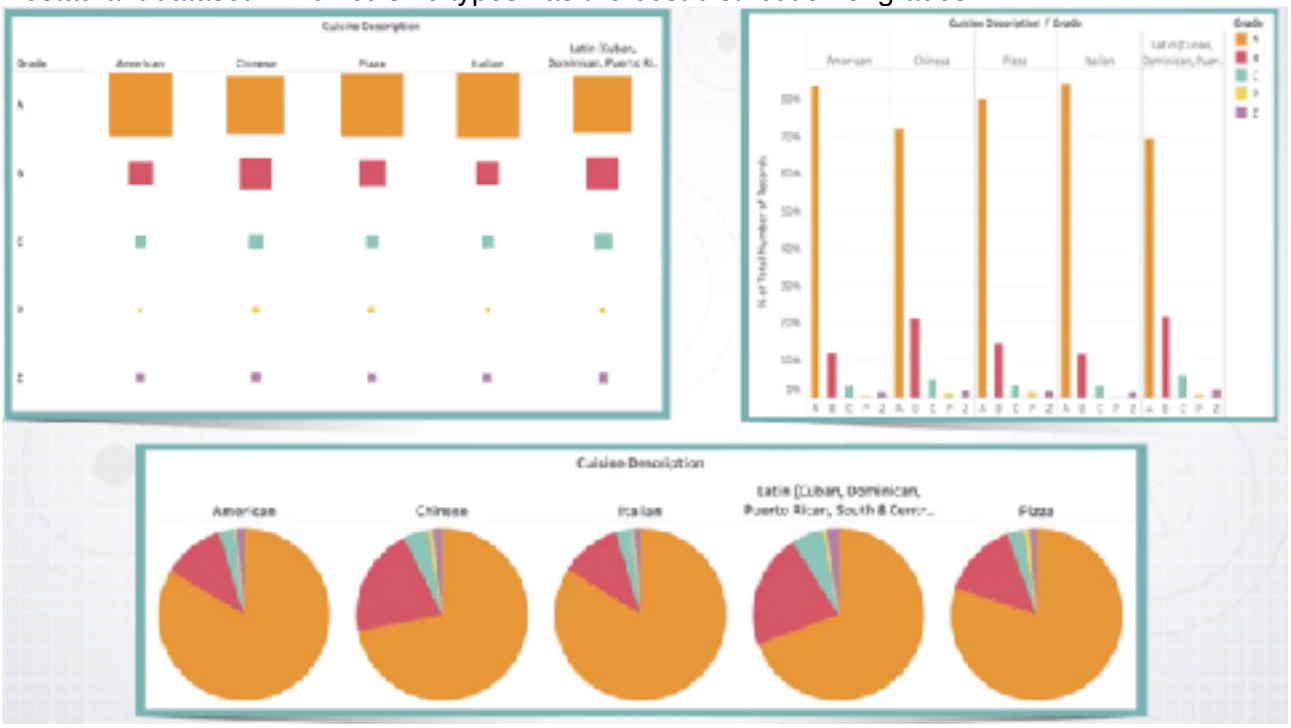
Pb:

- 3D projection:
  - every area is skewed = difficulty comparing
  - Area is not a good channel for quantities
  - Rectangles of different proportion between width and height
- Colour encoding to represent change not effective
  - Hard to distinguish positive changes to negative changes
- Some rectangles are really small to show the name

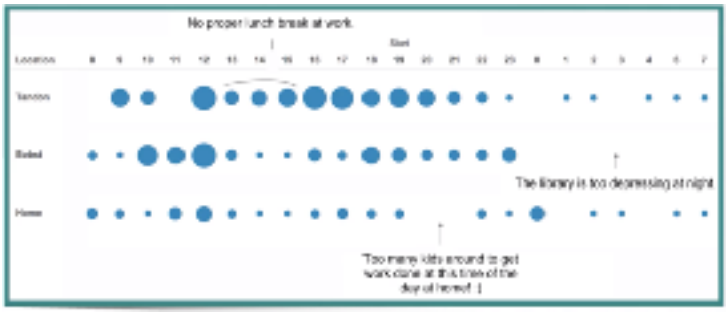


USING THE PRINCIPLES TO DESIGN VISUALIZATION

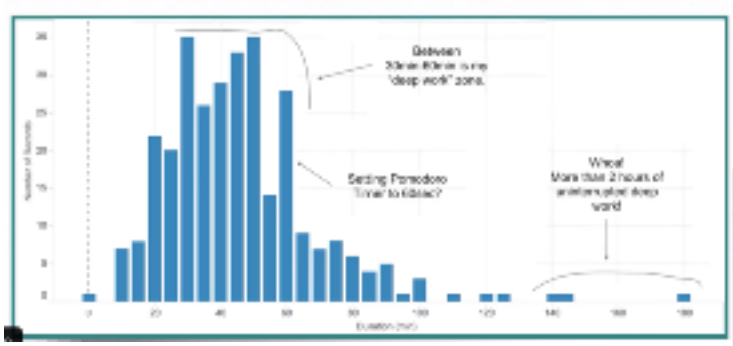
Restaurant dataset: which cuisine types has the best distribution of grades?



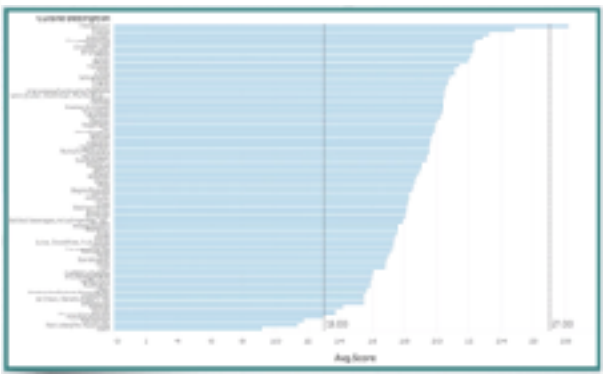
BEYOND MARKS & CHANNELS: CONTEXTUAL COMPONENTS



1 . LEGENDS, LABELS, ANNOTATIONS : enable interpretation of graphical elements, annotation guide attention to explain patterns of interest



2 . AXES, GRIDS, REFERENCE LINES: enable value reading and comparison



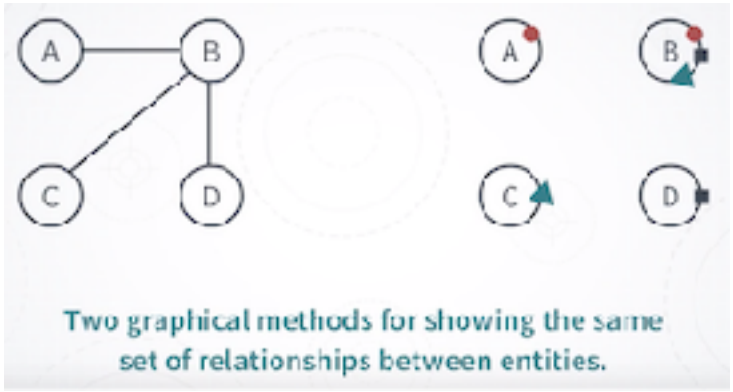
. The bar shows the limit between grade A and B

SUMMARY

Visual Encoding/Decoding
Marks + Channels
Expressiveness/Effectiveness
Channels Appropriateness and Ranking
Evaluation + Design
Contextual components (labels, legends, annotations, axes, grids, trend lines)

Information Visualization - 2 : Applied Perceptions

WHAT IS APPLIED PERCEPTION



Study of human perceptual/cognitive processing to make informed decisions about visualization design.

Work of the designer :

Space of solutions (both driven by principles of visual perception) :

1. Start from good enough
2. Improve on initial solution

Task : we cannot define effectiveness without a task. The role of task is essential in visualization design and evaluation. To choose the visualization you need to know what to accomplish. For a given set of tasks :

- Predict what works and what doesn't
- Explain why it works

Visual perception :

1. Visual/cognitive system
2. Effectiveness of visual channels
3. Color perception

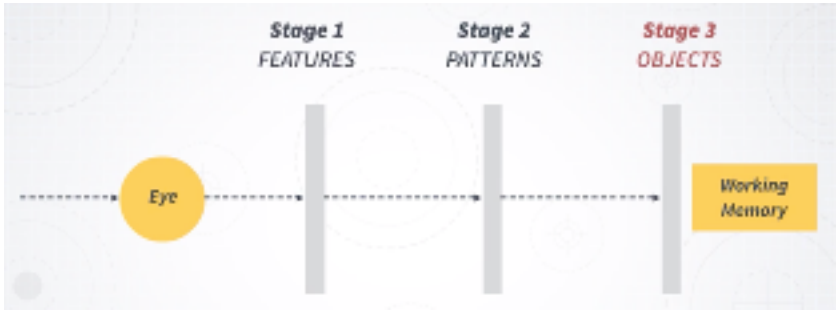
HUMAN VISUAL PROCESSING SYSTEM

Retina packed with 2 types of sensors :

- Rods: low light conditions
- Cones: normal conditions

Lossy process

1. Extraction of low level features
   - Orientation, Color, Texture, Movement
   - Parallel processing. Unconscious. Transitory
2. Patterns
   - Perception is sequential. Influenced by attention
3. Objects
   - Stored in memory
   - Conscious. Limited. (7 chunks)



SACCADIC EYE MOVEMENT

- Quickly twitching around the area of interest
- Image suppressed while the eye is moving
- Image reconstructed in the brain

Stephen Few *Show me the number* “We don’t see image with our eyes, we see them with our brains."

ROLE OF ATTENTION IN VISUAL PERCEPTION

Not only we see a small portion but we also retain very little information. What we retain is highly dependent on the task.

We encode what is relevant : the goal of vision is not to make a complete picture of the world but to make sense of the meaning of the world around us.

“Solving the ‘real’ mysteries of visual perception: the world as an outside memory”.
Kevin O’Regan

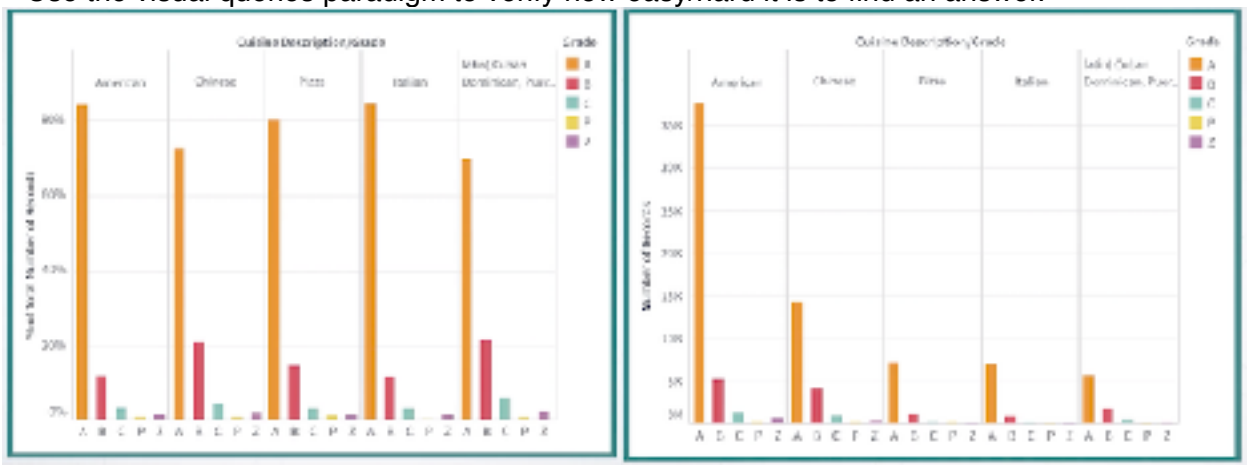
“It is more accurate to say that we are conscious of the field of information to which we have rapid access rather than we are immediately conscious of the world."

When we design we have to guide attention in a way that is useful to achieving the goal. In turn, you have to be aware how design choices influence the viewer’s attention.

VISUAL QUERIES

Acts of attention driving our eye-movements to find the information we need to accomplish a task.

Ex. Connect 2 subway stations
- Explicitly identify the tasks and visual queries your visualization is meant to help accomplish.
- Use the visual queries paradigm to verify how easy/hard it is to find an answer.



WHAT WE CAN EASILY SEE

V1 and V2 regions of neurones that respond to specific kind of properties: form, colour, orientation, motion, depth and sound.
There are some other visual characteristics that our brain is not capable of processing as efficiently as when we have pre-attentive features.

RECAPITULATION OF KEY CONCEPTS

Viz effectiveness and objective facts form vision science
Information retained and discarded in visual processing
Fovea, saccades and fast sampling of the world around us
Role of attention on focus and what is retained
Understanding visual queries and low-level vision

Make tasks and visual queries explicit when evaluating visualization design

Ask : *What questions does the*

Leverage pre-attentive (‘tunable’) features

Eyes beat memory
Moving your eyes is much faster than moving between views (scrolling, zooming, panning)

MAPPING BETWEEN DATA PROPERTIES

Each visualization can be described as a mapping between data properties to visual properties.

Visual

channels

: graphical properties to represent object.

EXPRESSIVENESS OF VISUAL CHANNEL

What (what type of information) can be expressed.
Types of measurements : nominal, ordinal, interval (percent, temperature), ratio (weight, height)
Quantitative information, sequential information, categorical information.

EXAMPLE 1: HOW TO EXPRESS QUANTITY WITH DIFFERENT CHANNELS

Position
Size (area, length)
Angle
Slope, direction
Colour intensity

EXAMPLE 2: HOW TO EXPRESS THE IDEA OF ORDER WITH DIFFERENT CHANNELS

Position (list…)
Texture / Density
Color Intensity

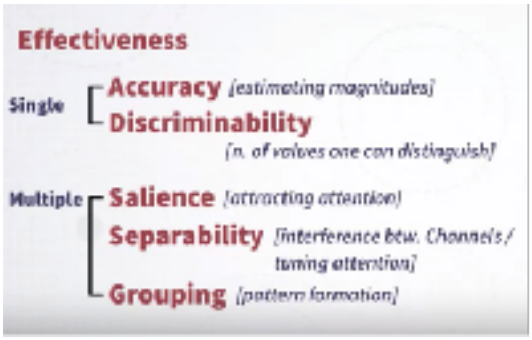
EXAMPLE 3: HOW TO EXPRESS CATEGORIES WITH DIFFERENT CHANNELS

Position (bar chart…)
Hue
Shape

Effectiveness of Visual Channels

INTRODUCTION

How well



EXPERIMENTS IN “GRAPHICAL PERCEPTION”

IMPLICATIONS FOR DESIGN (ACCURACY)

DISCRIMINABILITY

IMPLICATIONS FOR DESIGN (DISCRIMINABILITY)

SEPARABILITY

IMPLICATIONS FOR DESIGN (SEPARABILITY)

GROUPING: SIMILARITY & PROXIMITY

GROUPING: CONNECTION & ENCLOSURE

HIERARCHY WITHIN GROUPING TECHNIQUES

GROUPING: CLOSURE & CONTINUITY

RECAPITULATION & READING

Colour Perception & Colour Spaces

Using Colour in Visualization