

DATA VISUALIZATION

JAVIER GARCIA-BERNARDO

Department of Methodology & Statistics

Utrecht University



“Those hired into analytical roles typically have quantitative backgrounds that suit them well for the other steps (finding the data, pulling it together, analyzing it, building models), but not necessarily any formal training in design to help them when it comes to the communication of the analysis—which, by the way, is typically the only part of the analytical process that your audience ever sees”

Cole Nussbaumer Knaflic. “Storytelling with Data”.

WHAT IS A GOOD VISUALIZATION

Two conditions:

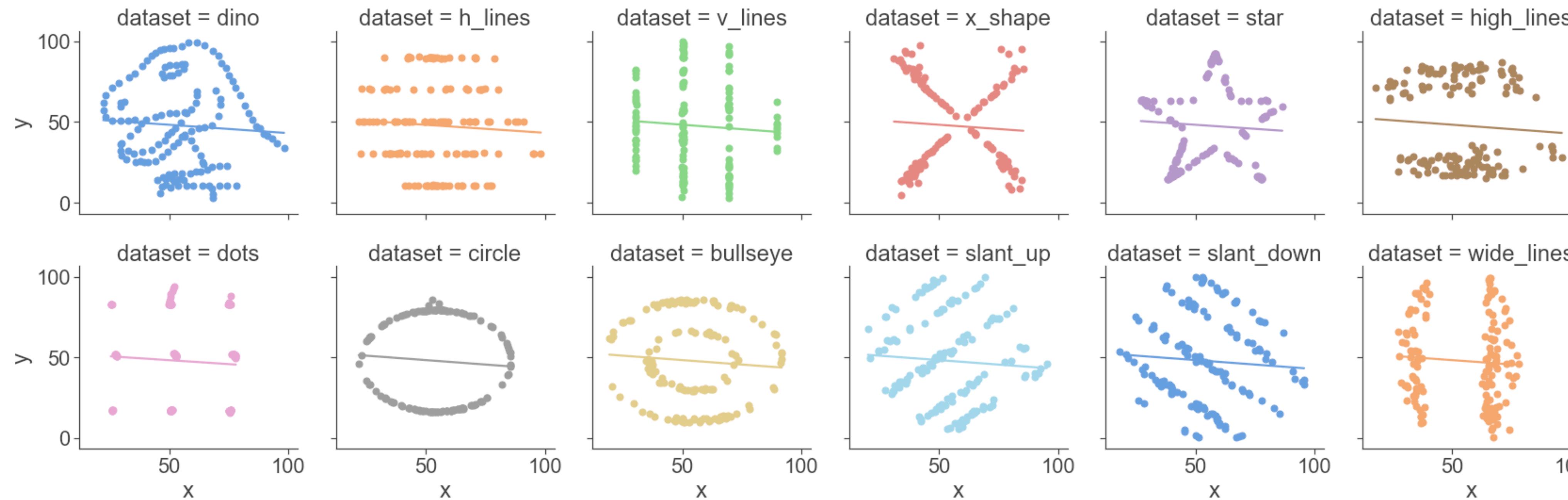
- ▶ **Fair** (e.g. correct data, not hiding important information)
- ▶ Effective and efficient: Reduce **cognitive load** and misinterpretations

Cognitive load

*The amount of working memory used to take in information
and consolidate it into long-term memory.*

INTRODUCTION

How much longer would it take you to understand that the datasets are different?



WHY DO WE WANT TO REDUCE COGNITIVE LOAD

- ▶ More willing to **read** your paper
- ▶ More likely to **understand** the data/results
- ▶ More willing to **accept** the results
- ▶ More likely to **remember** them

HOW TO REDUCE COGNITIVE LOAD

- ▶ Good use of perception principles → Ethical and efficient
- ▶ Good use of design and storytelling principles → Effective



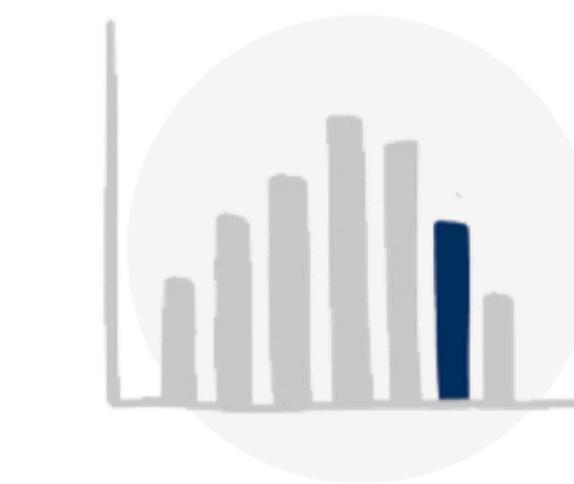
**understand the
context**



**choose an
effective visual**



**eliminate
clutter**



**focus
attention**



**tell a
story**

DESIGN

STORYTELLING

PART 1

UNDERSTAND

TWO MAIN QUESTIONS

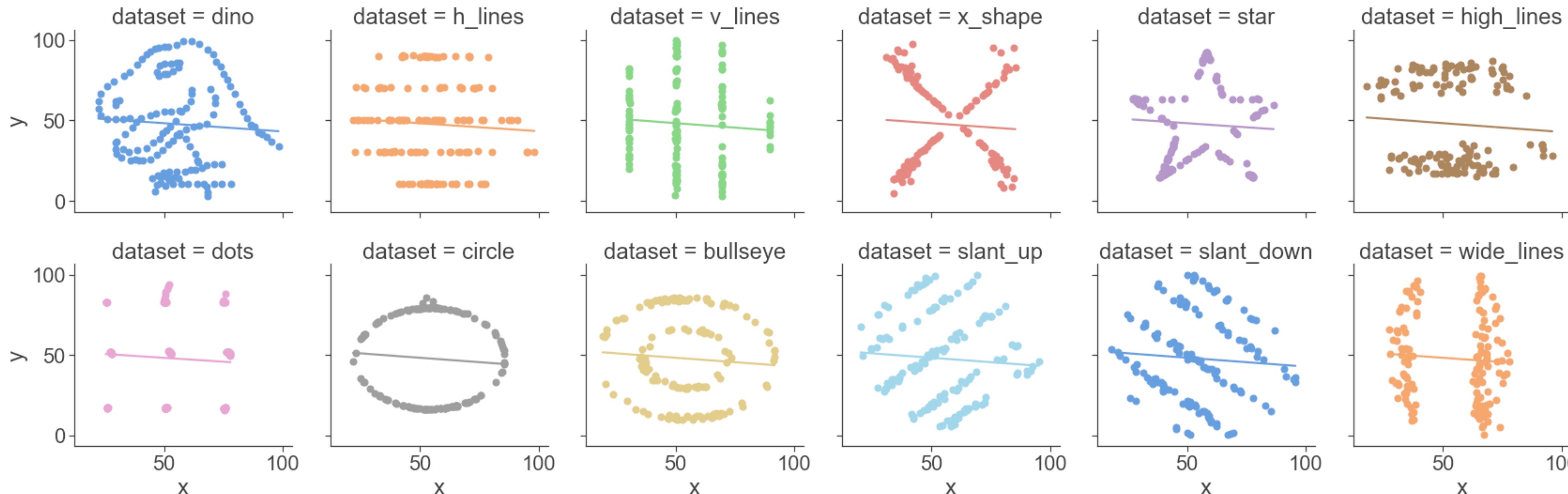
- (1) WHO IS IT FOR?
- (2) WHAT IS THE MAIN MESSAGE?

(1) WHO IS IT FOR? — EXPLORATORY VS EXPLANATORY VISUALIZATION

- Yourself (exploratory) vs **others (explanatory)**

- Academics vs journalists vs lay audience

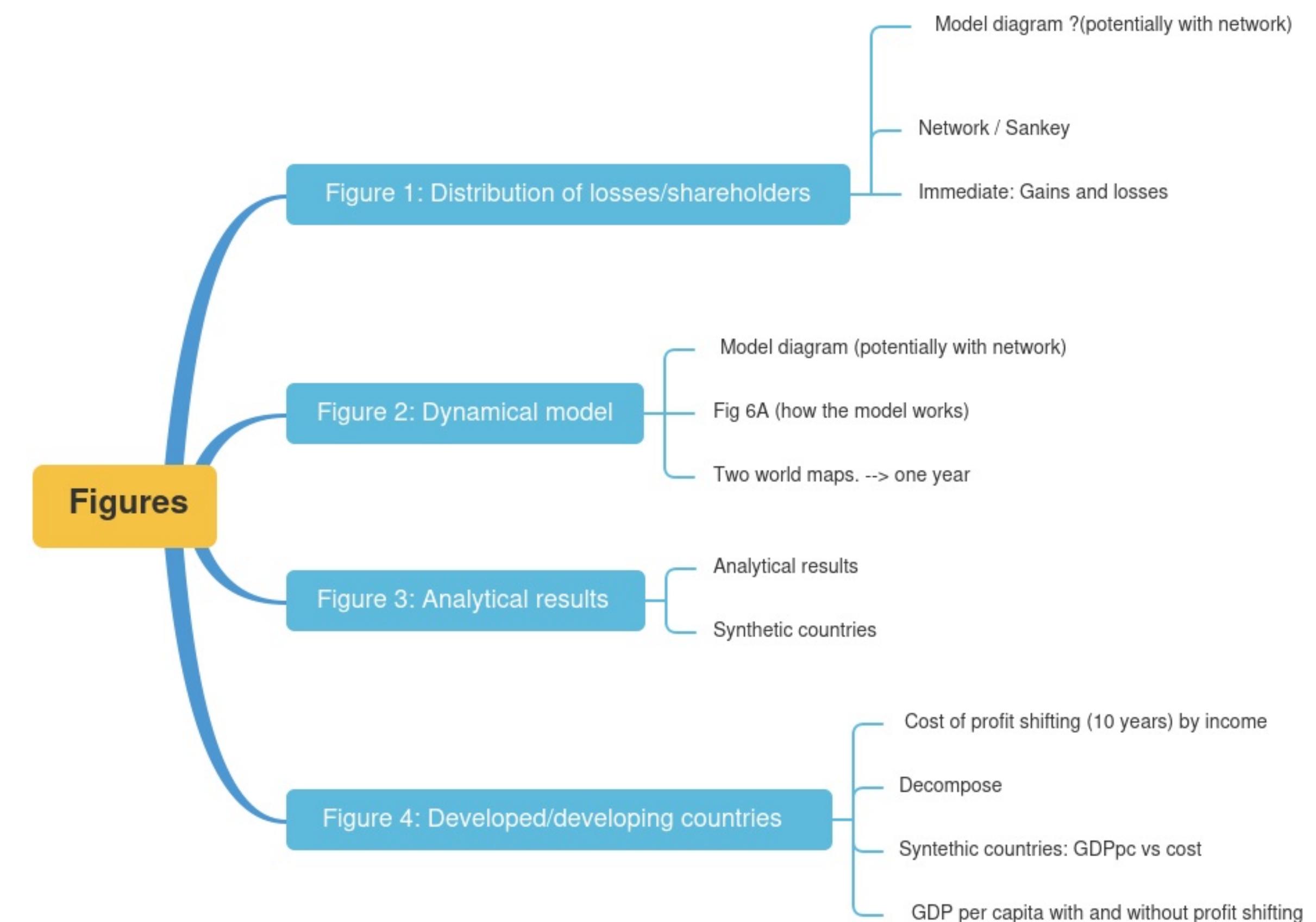
- Online audience (color/bw) vs presentation



(2) WHAT IS THE MAIN MESSAGE?

Each figure should have **one**
(and only one) main message

Write it down in one sentence,
be specific.



PART 1: UNDERSTAND THE CONTEXT

PRACTICE: EXPLORE DATASET

- ▶ Do some exploratory data visualization. Think about the main message and the audience
- ▶ Derek's:
 - ▶ Collection of (microbial DNA) samples of the inside of the dental unit water system
 - ▶ Columns: dosing of antimicrobial, amount of legionella, amount of DNA
 - ▶ Message idea: X and Y cleaning factors give the safest microbiome in the water system
- ▶ School Shooting 1: (ss.tsv)
 - ▶ School shooting per day
 - ▶ Columns: date, number of victims, state of attack
 - ▶ Message idea: the probability of a new attack increases after each attack
- ▶ School Shooting 2: (tweets_and_ss.tsv)
 - ▶ School shooting per day and number of tweets with the terms "school" and "shooting"
 - ▶ Columns: date, number of victims, number of tweets, state of attack
 - ▶ Message idea: the probability of a new attack increases with the buzz in social media

PART 2

EFFECTIVE VISUALS

PART 2: CHOOSE AN EFFECTIVE VISUAL

WHICH FIGURE SHALL I USE?



THE UNSPOKEN PITCH

GOALS:

Two conditions:

- ▶ Fair (e.g. correct data, not hiding important information)
- ▶ Effective and efficient: Reduce **cognitive load** and misinterpretations

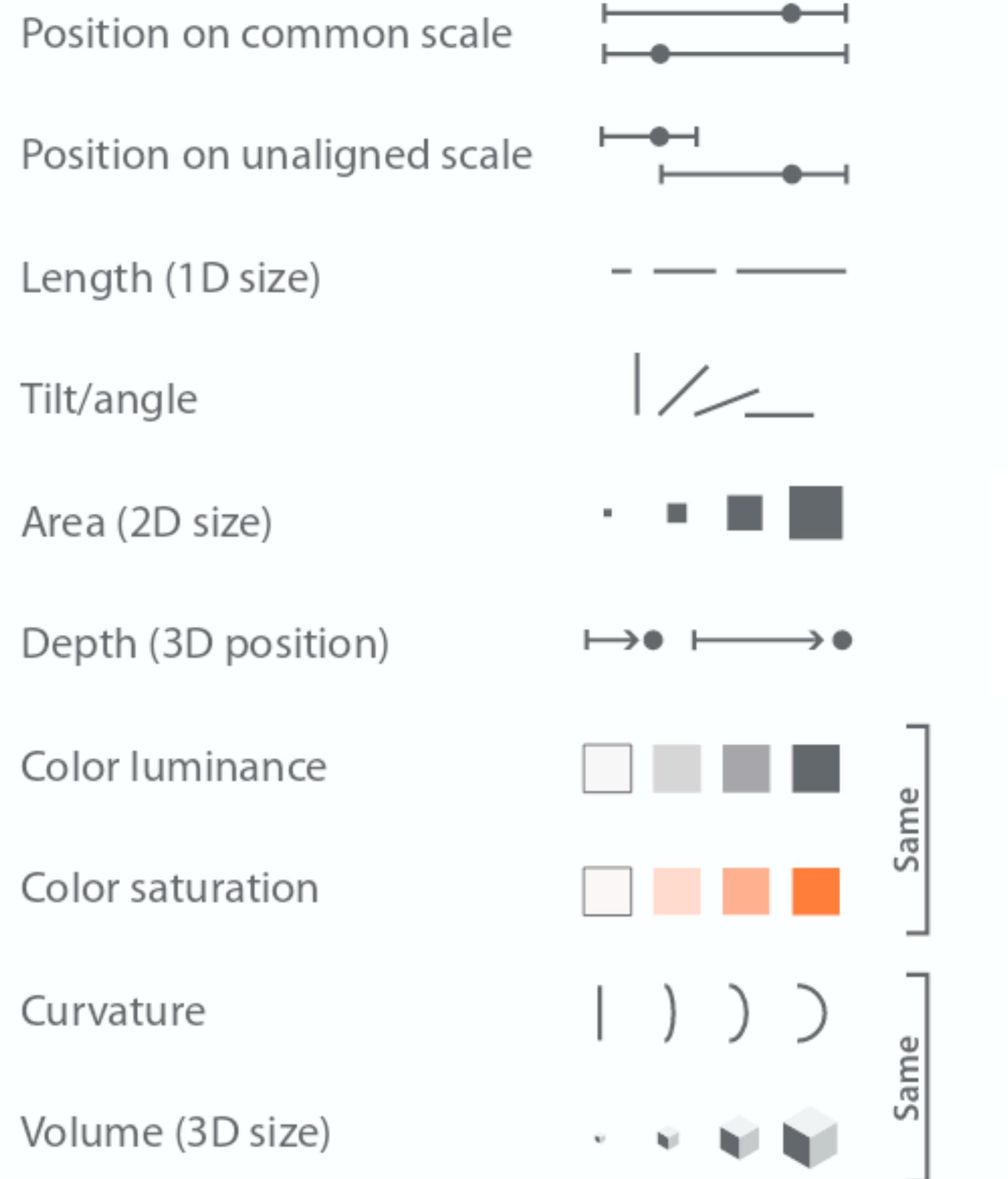
ELEMENTS OF A PLOT:

CHANNELS

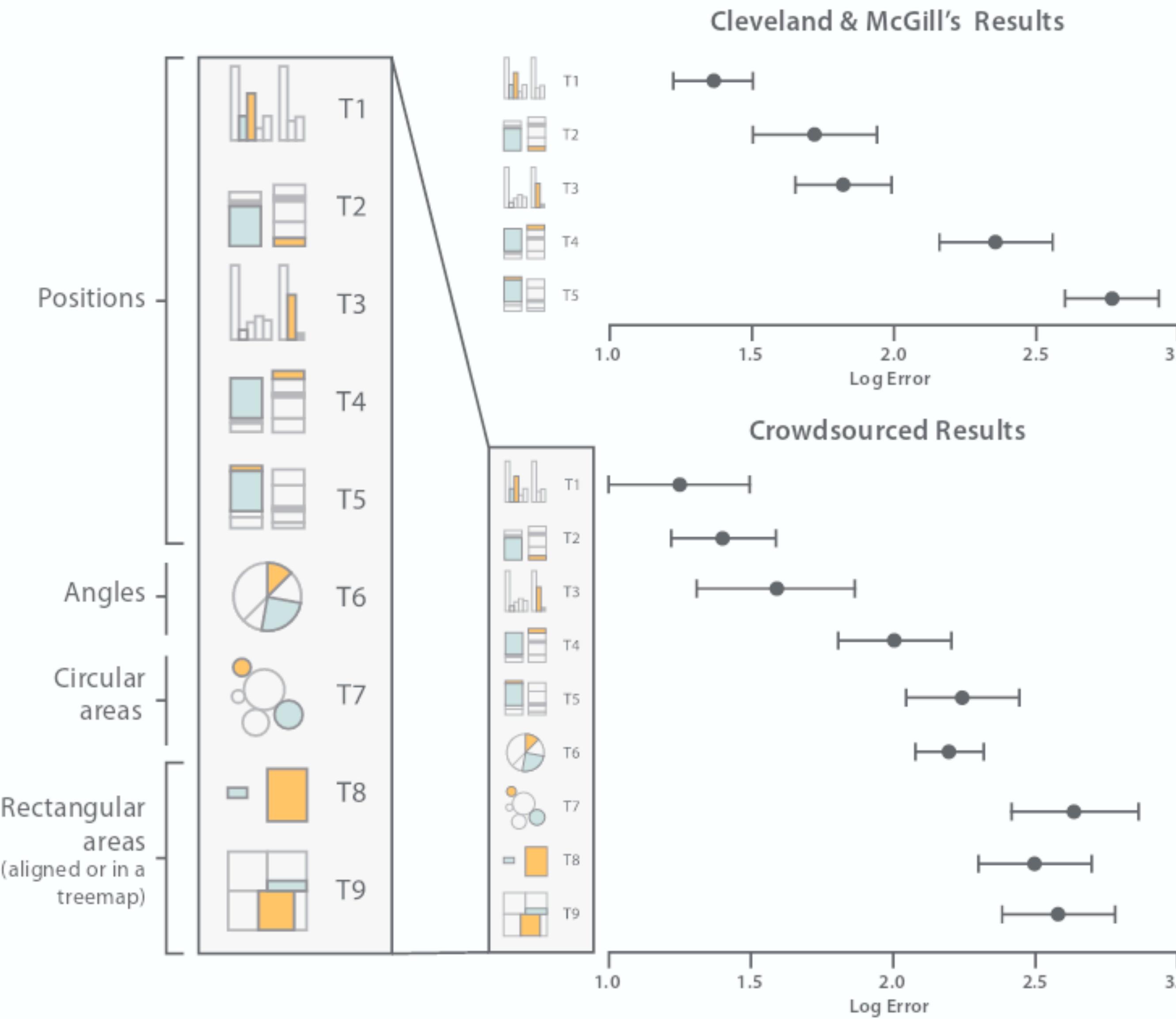
PART 2: CHOOSE AN EFFECTIVE VISUAL

Channels: Expressiveness Types and Effectiveness Ranks

④ **Magnitude** Channels: **Ordered Attributes**



PART 2: CHOOSE AN EFFECTIVE VISUAL



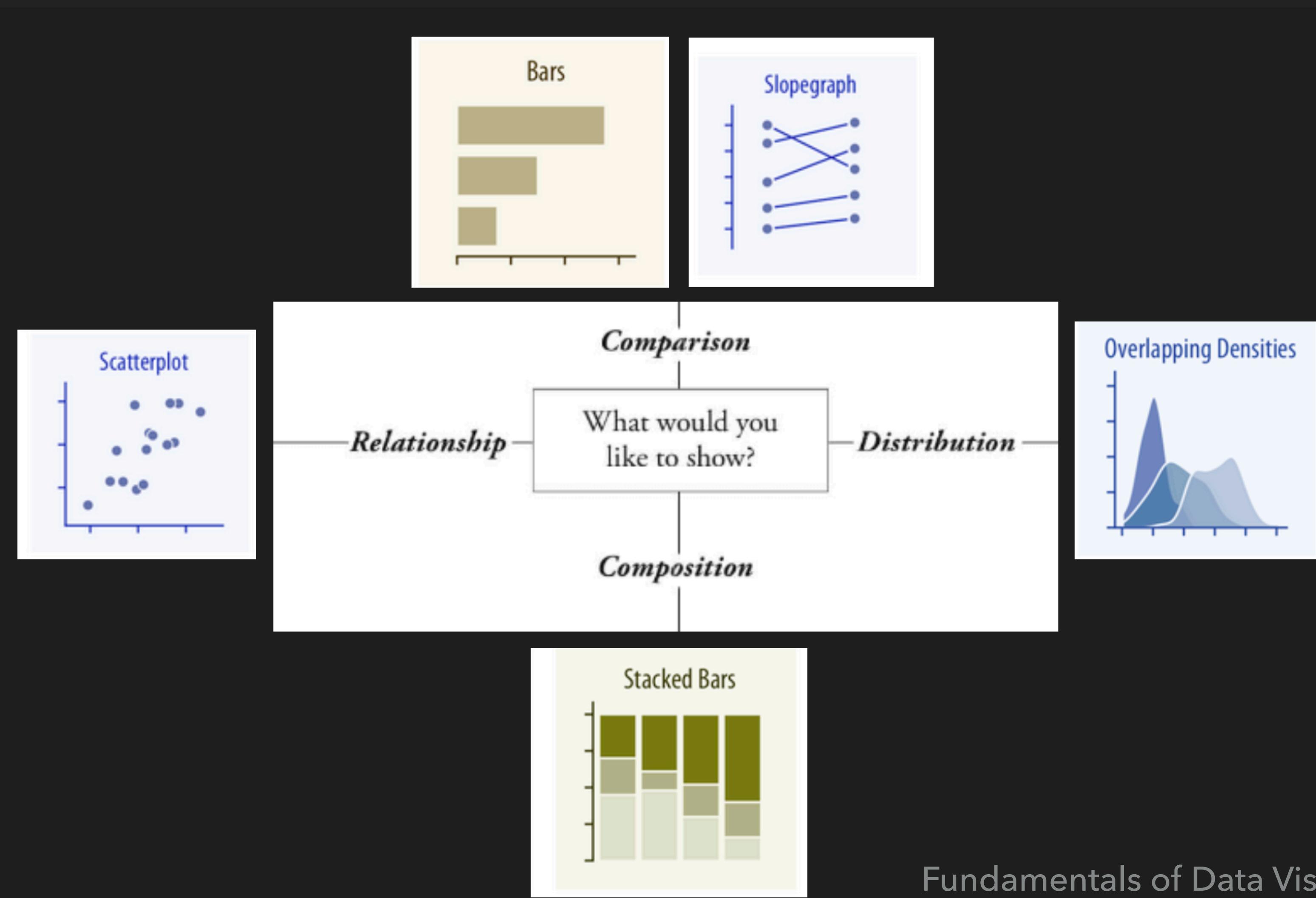
**NOT ALL CHANNELS
ARE EQUAL**

ELEMENTS OF A PLOT:

GRAPHICAL OBJECTS

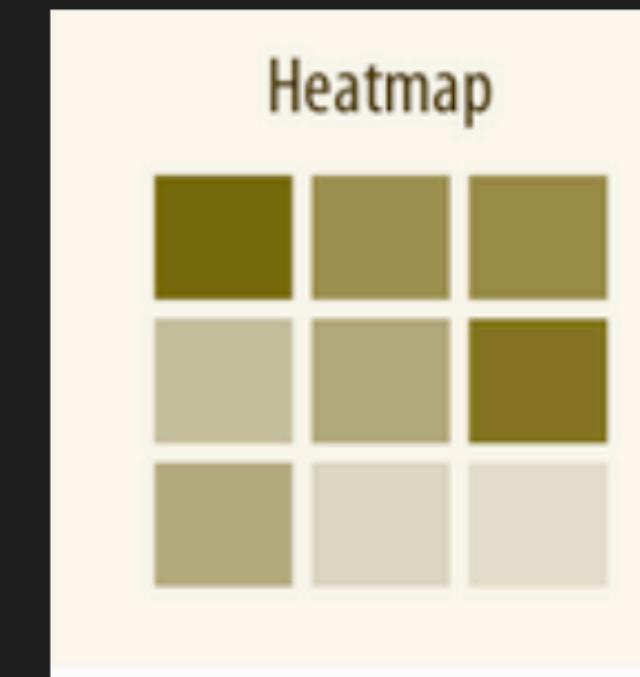
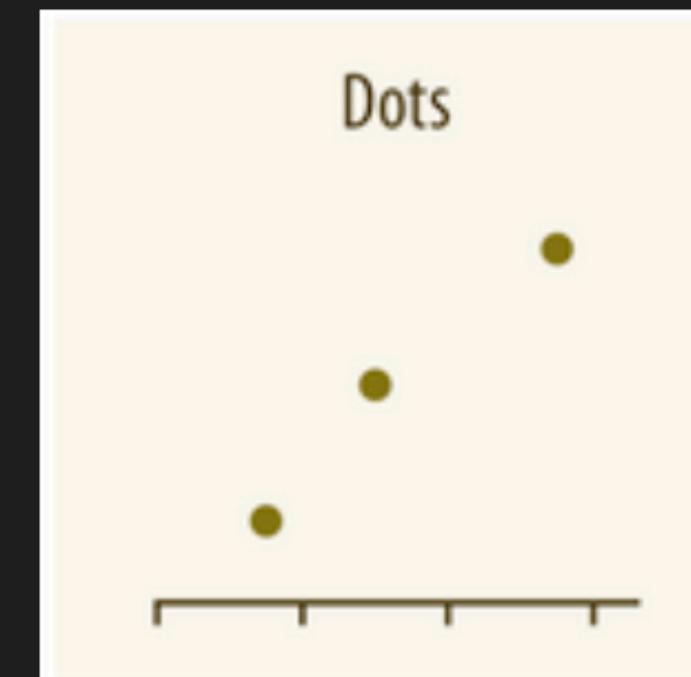
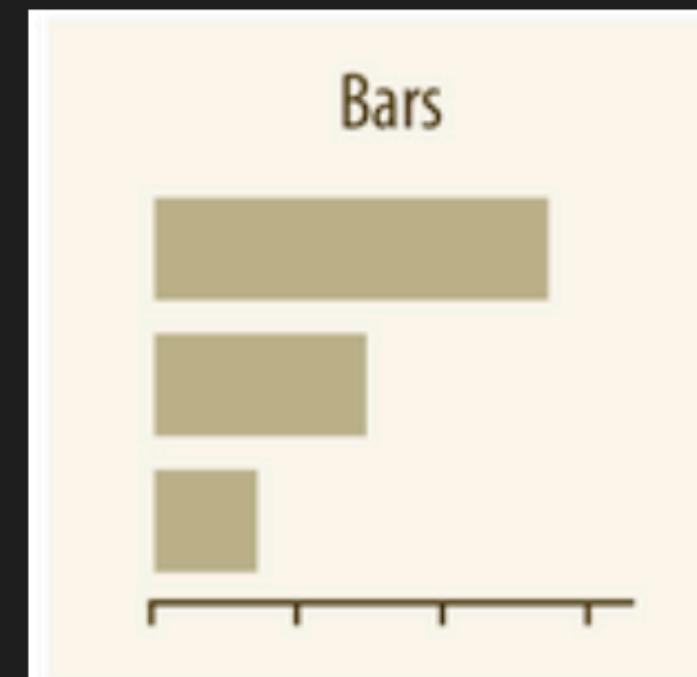
PART 2: CHOOSE AN EFFECTIVE VISUAL

THE TYPE OF GRAPH DEPENDS ON THE GOAL



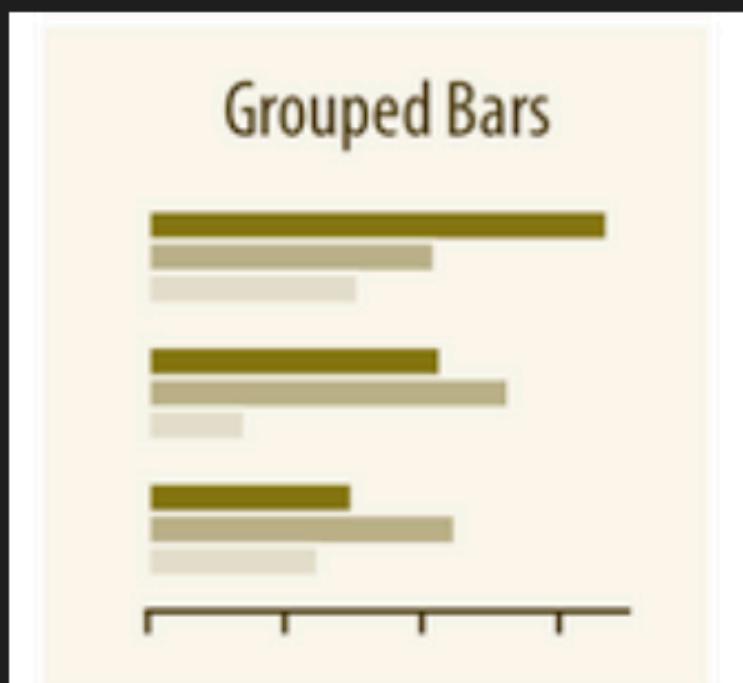
PART 2: CHOOSE AN EFFECTIVE VISUAL

AMOUNTS AND PROPORTIONS



Required with log-scales

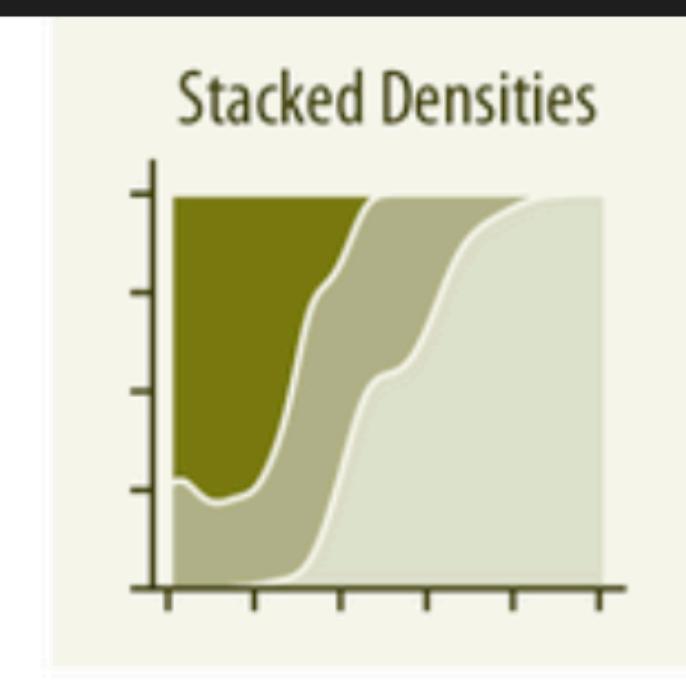
Trends



Differences within row



Proportions over x



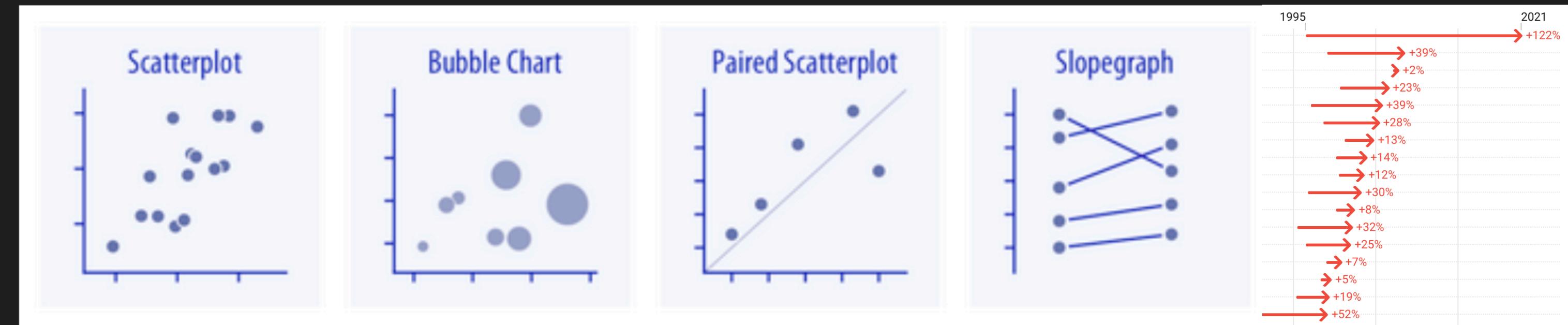
PART 2: CHOOSE AN EFFECTIVE VISUAL

DISTRIBUTIONS

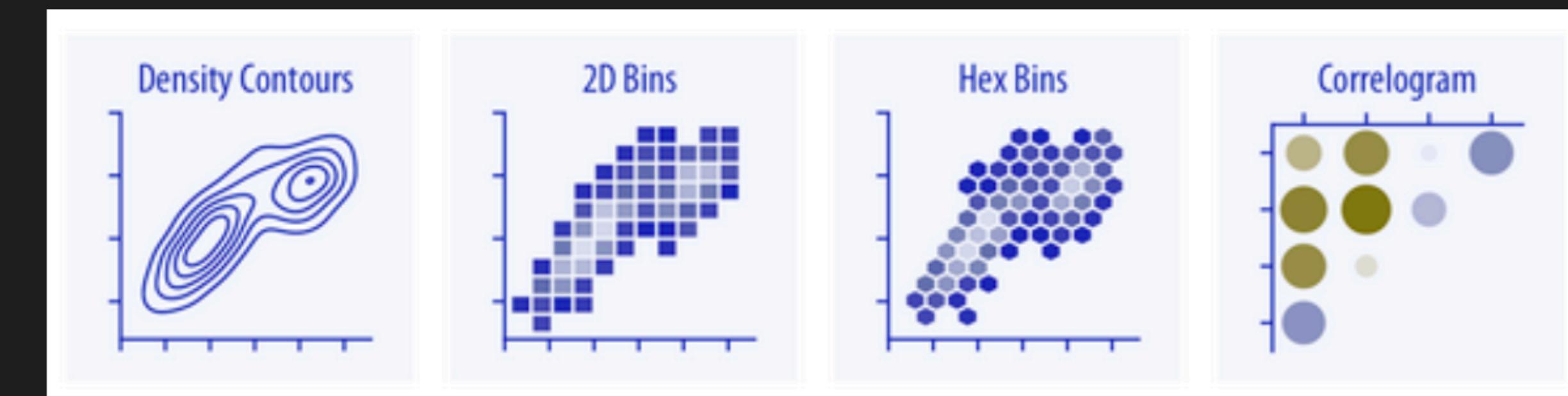


PART 2: CHOOSE AN EFFECTIVE VISUAL

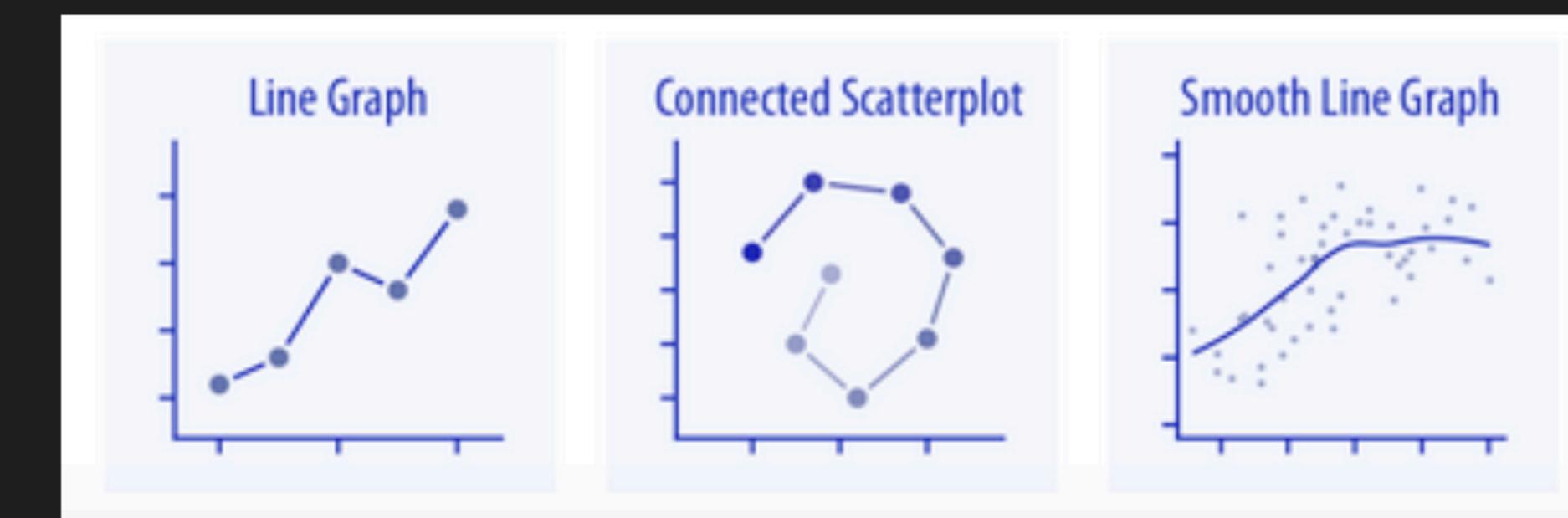
RELATIONSHIPS



Too many points?



Time series?

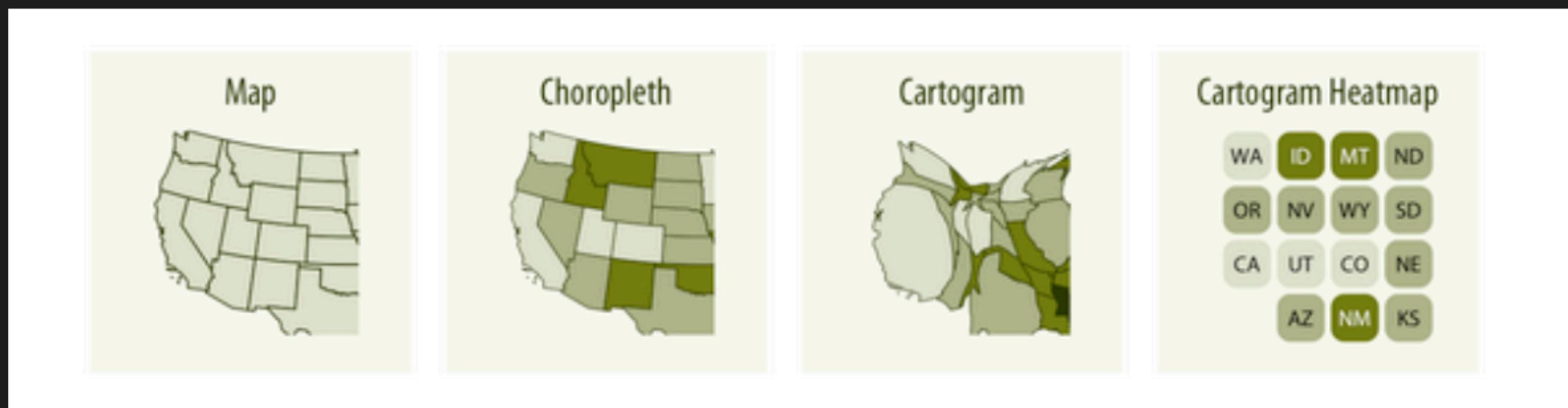


PART 2: CHOOSE AN EFFECTIVE VISUAL

GEOGRAPHICAL DATA

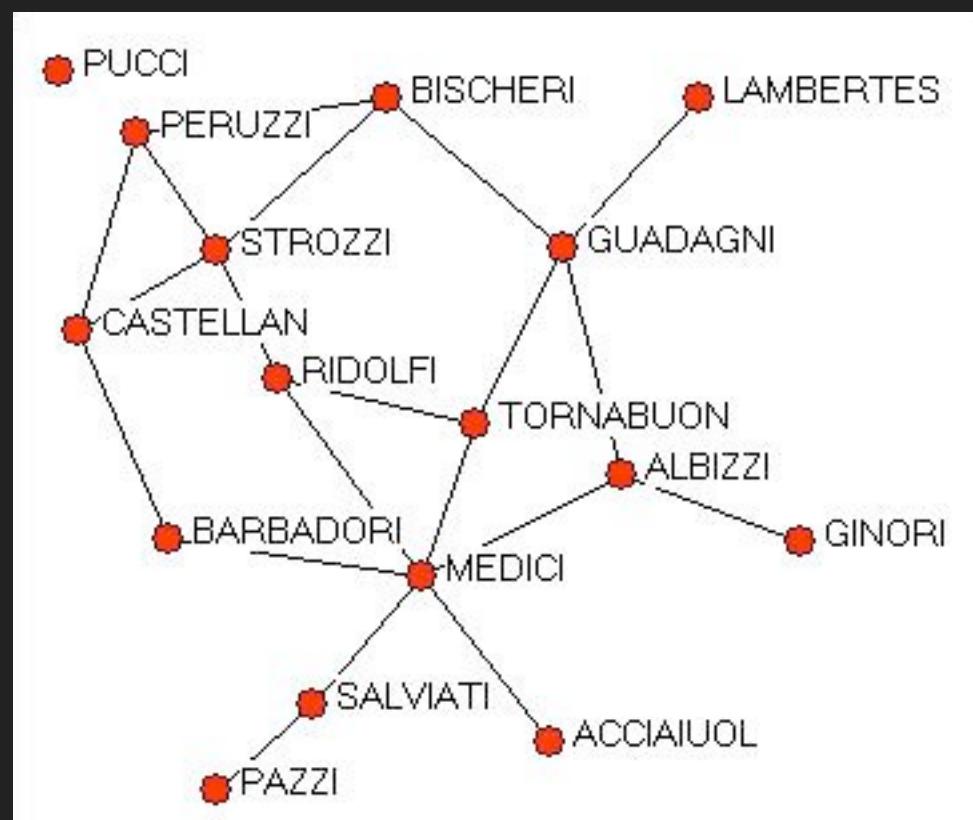
- ▶ Color is key (more on this later)
- ▶ Combine with a barplot or bubbles if the values are important

Do you *actually* need a map?

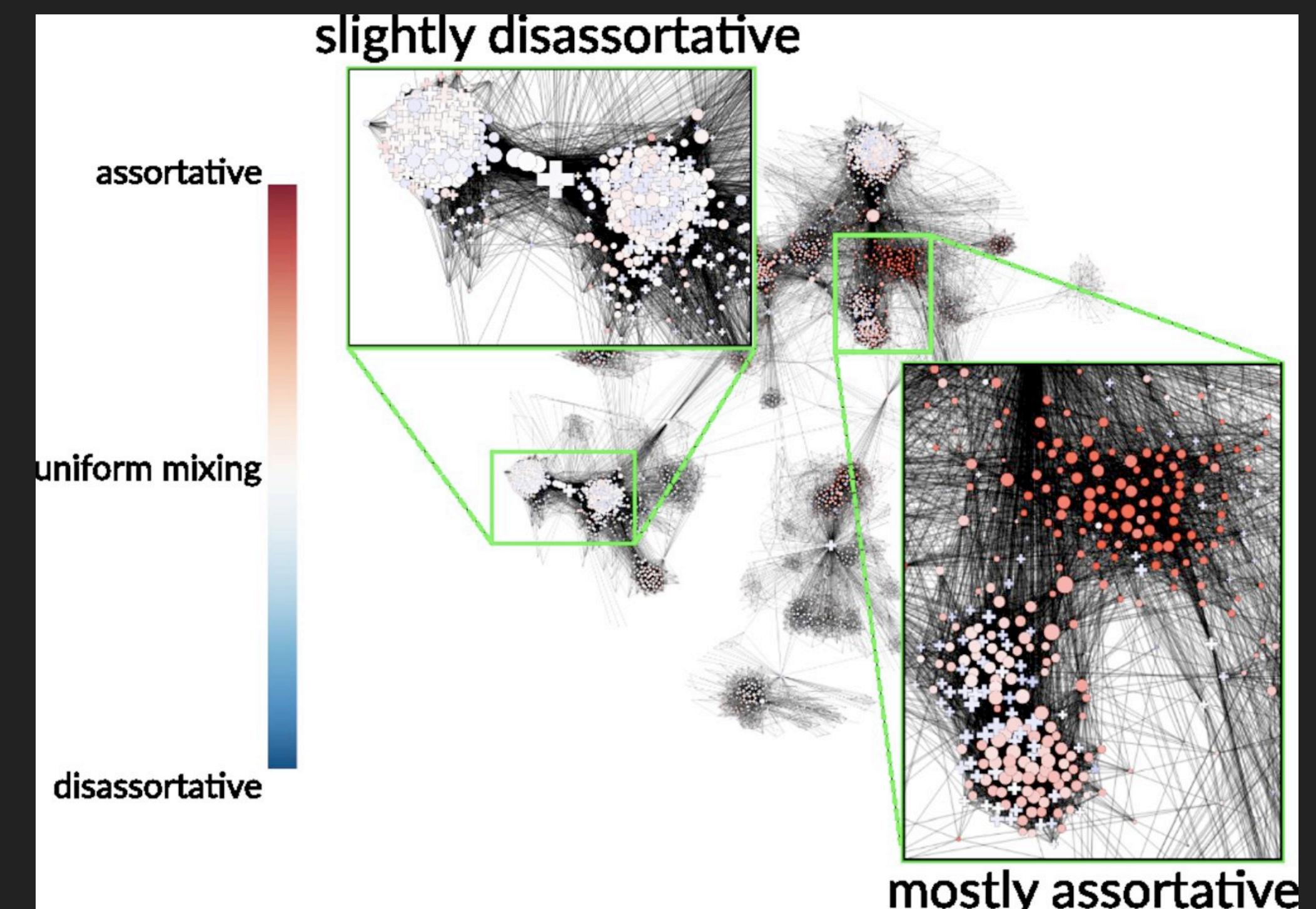


NETWORKS: DO YOU ACTUALLY NEED A NETWORK?

- ▶ Nobody wants to see your hairball
- ▶ Show if:
 - ▶ Small networks
 - ▶ Show a macro-pattern



Florentian families



Peel et al

ELEMENTS OF A PLOT:

SCALES, FACETS, TRANSFORMATIONS

SCALES, FACETS, TRANSFORMATIONS

- ▶ **Scales:** When to use logarithmic scale
 - ▶ Represent ratios or percentages (1:2 and 2:1 are equidistant from 1:1 in a log-scale)
 - ▶ Increase visibility (too many values with small values)
 - ▶ Show that our distribution follows a exponential (lin-log scale), lognormal (log-lin scale) or power-law (log-log scale) distribution
- ▶ **Facets** –> Comparing numbers next to each other is better
- ▶ **Transformations** –> Jitter, etc

PRACTICE: MAP TO CHANNELS AND SKETCH VISUALIZATION

- ▶ Link your message to the type of figure
 - ▶ Type of graph(s)
 - ▶ How are you mapping variables to channels?
 - ▶ Which transformations (if any)

PRACTICE: MAP TO CHANNELS AND SKETCH VISUALIZATION

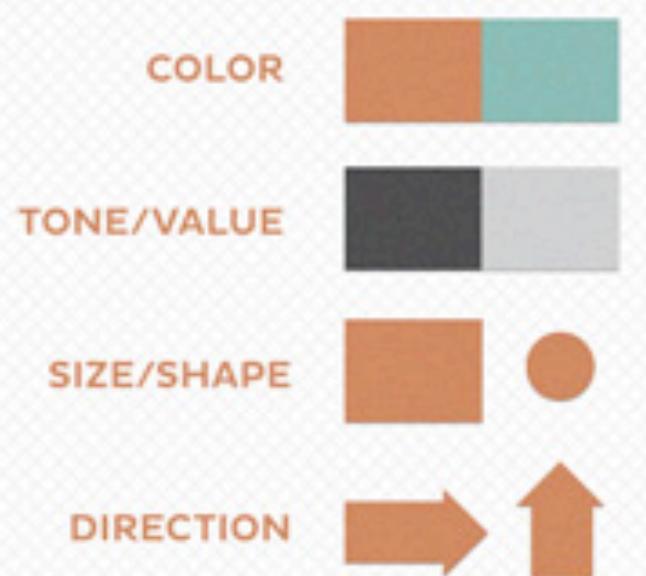
- ▶ Link your message to the type of figure
 - ▶ Type of graph(s)
 - ▶ How are you mapping variables to channels?
 - ▶ Which transformations (if any)

PART 3

DESIGN

PRINCIPLES OF DESIGN: CRAP

CONTRAST



Unique elements in a design should stand apart from one another. One way to do this is to use contrast. Good contrast in a design – which can be achieved using elements like color, tone, size, and more – allows the viewer's eye to flow naturally.

To the left, you can see 4 ways to create contrast in your design.

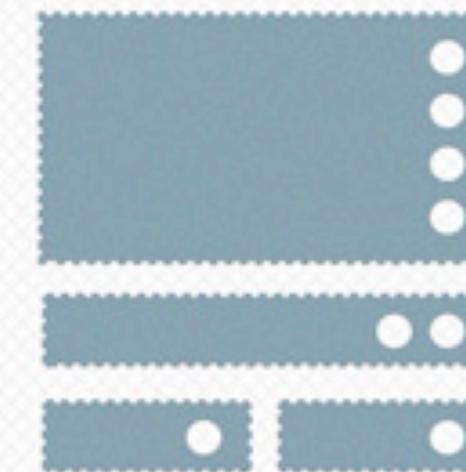
ALIGNMENT

Proper alignment in a design means that every element in it is visually connected to another element. Alignment allows for cohesiveness; nothing feels out of place or disconnected when alignment has been handled well.



REPETITION

Repetition breeds cohesiveness in a design. Once a design pattern has been established – for example, a dotted border or a specific typographic styling – repeat this pattern to establish consistency.



The short version?

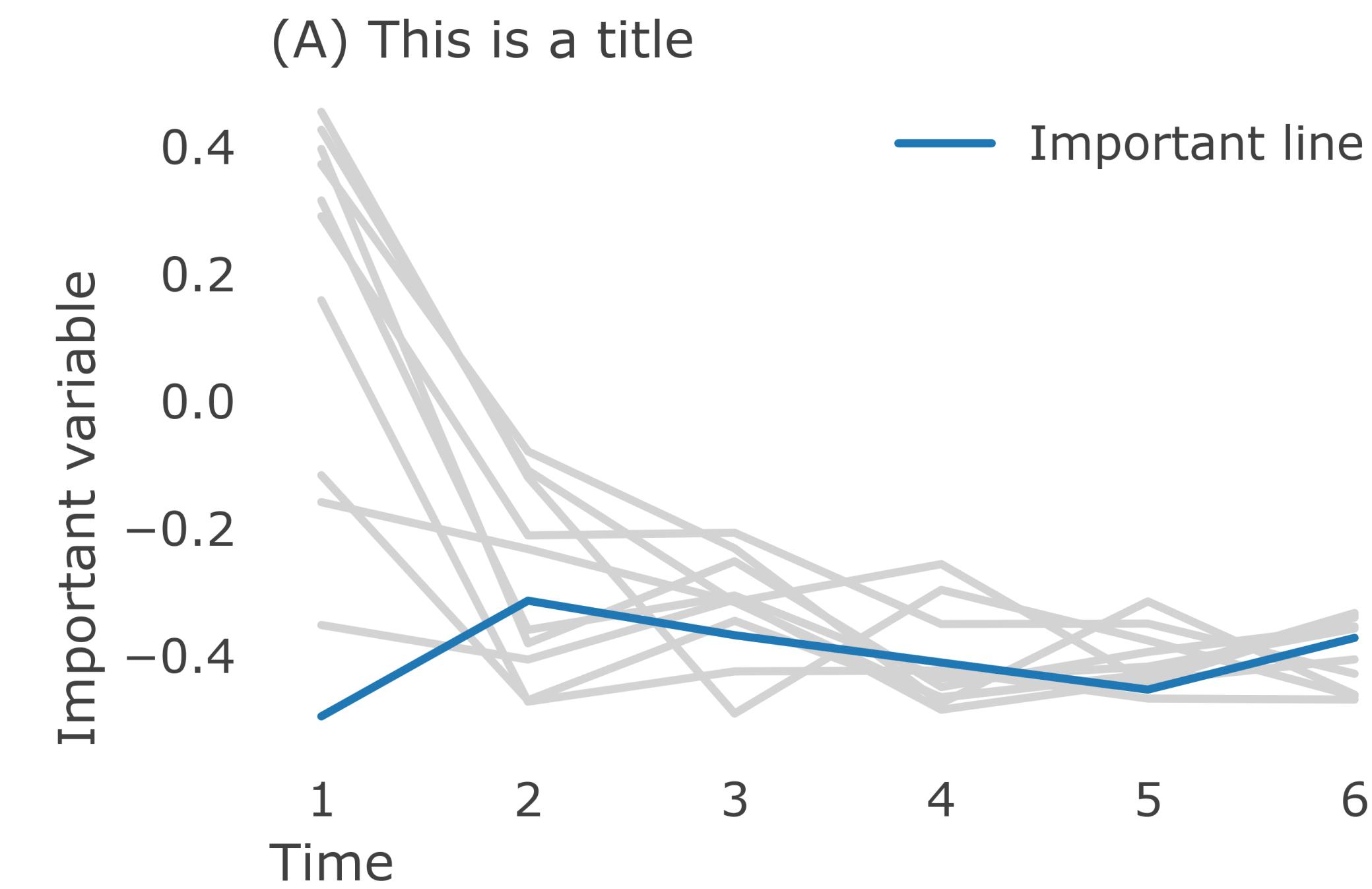
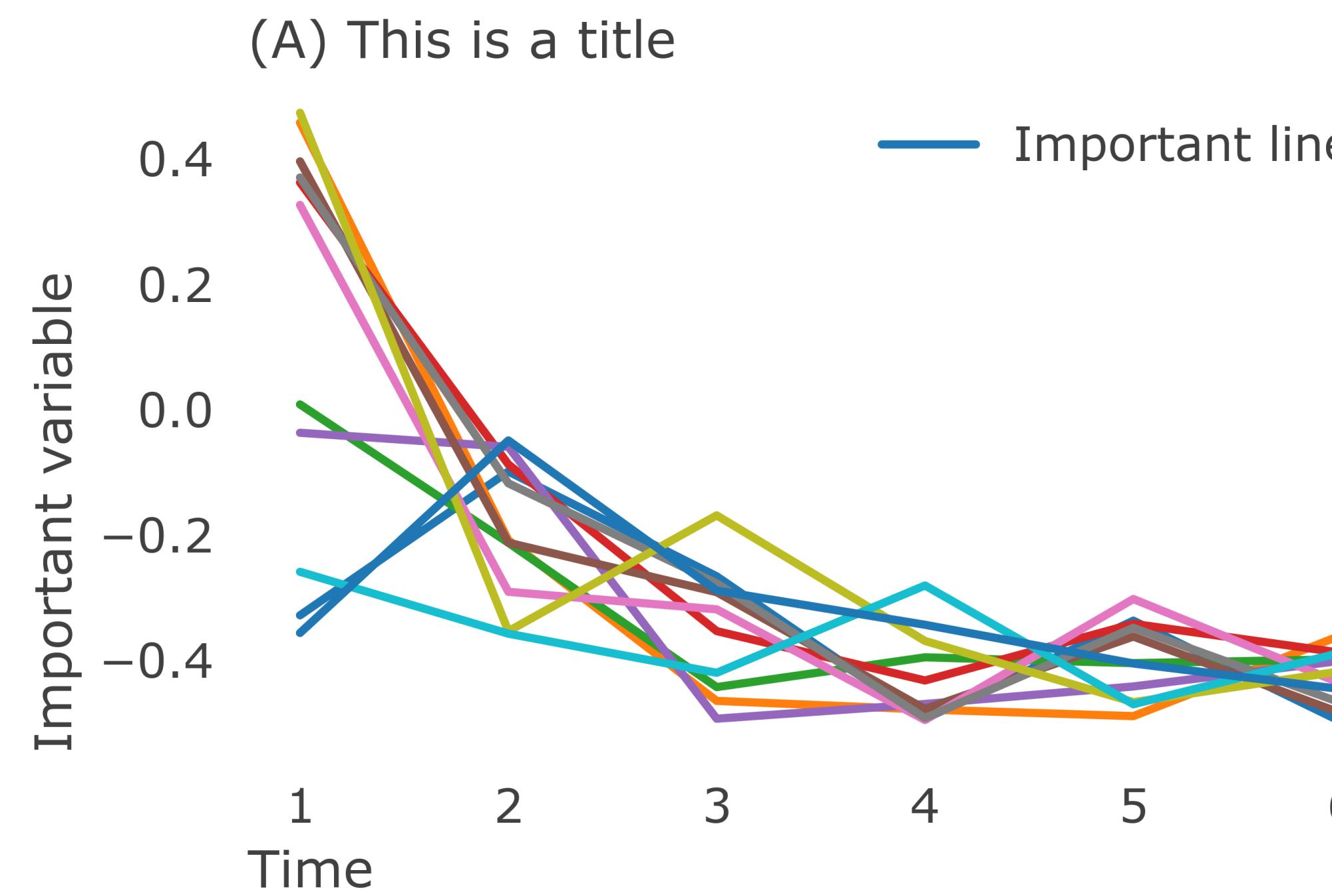
Establish a style for each element in a design and use it on similar elements.

PROXIMITY

Proximity allows for visual unity in a design. If two elements are related to each other, they should be placed in close proximity to one another. Doing so minimizes visual clutter, emphasizes organization, and increases viewer comprehension.

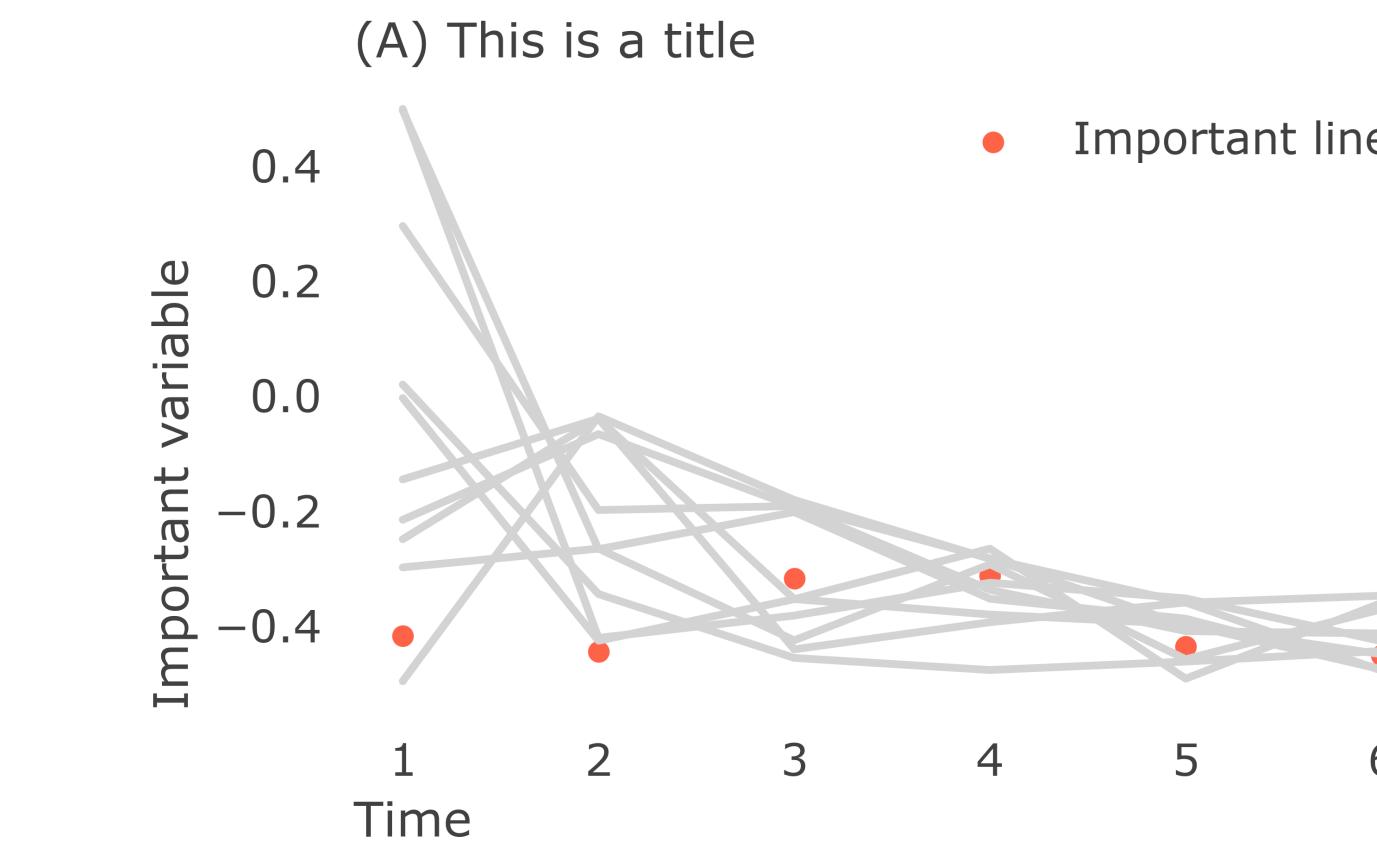
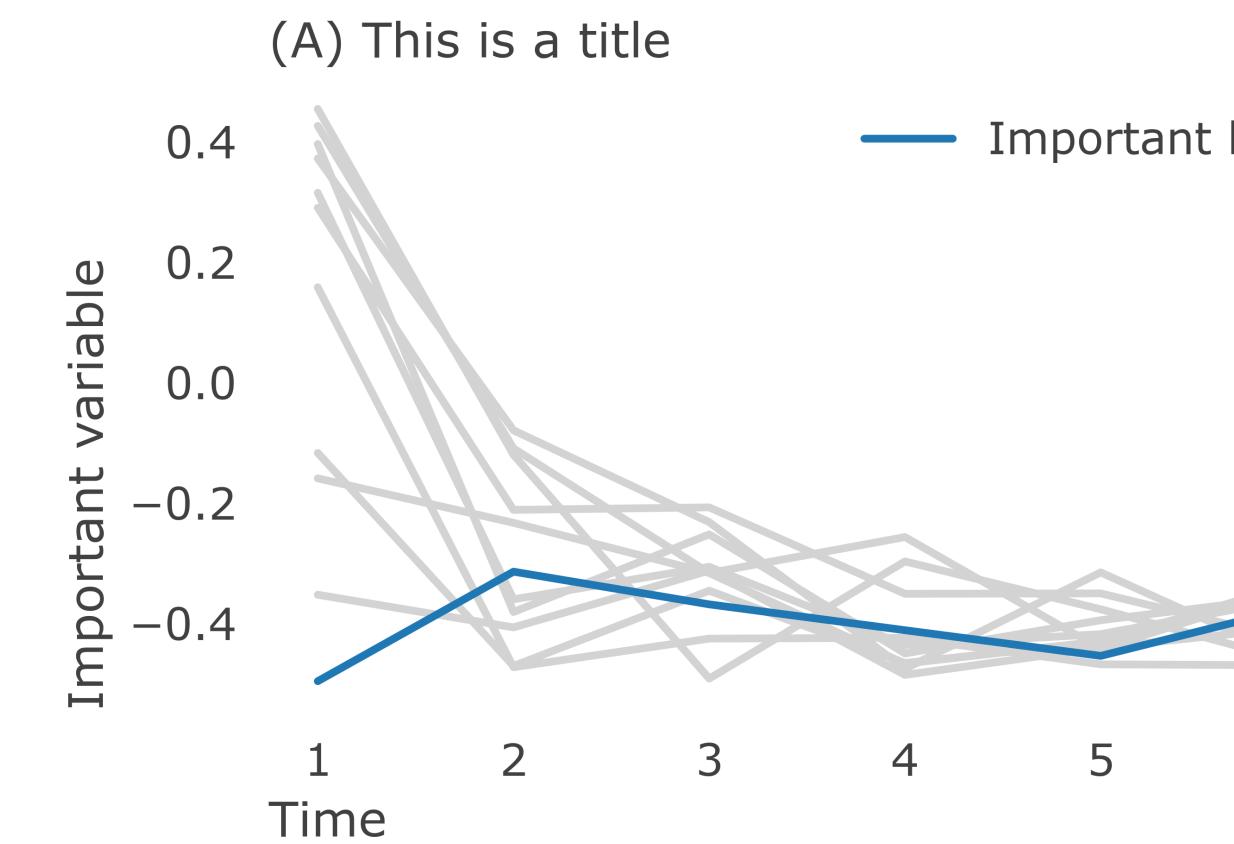
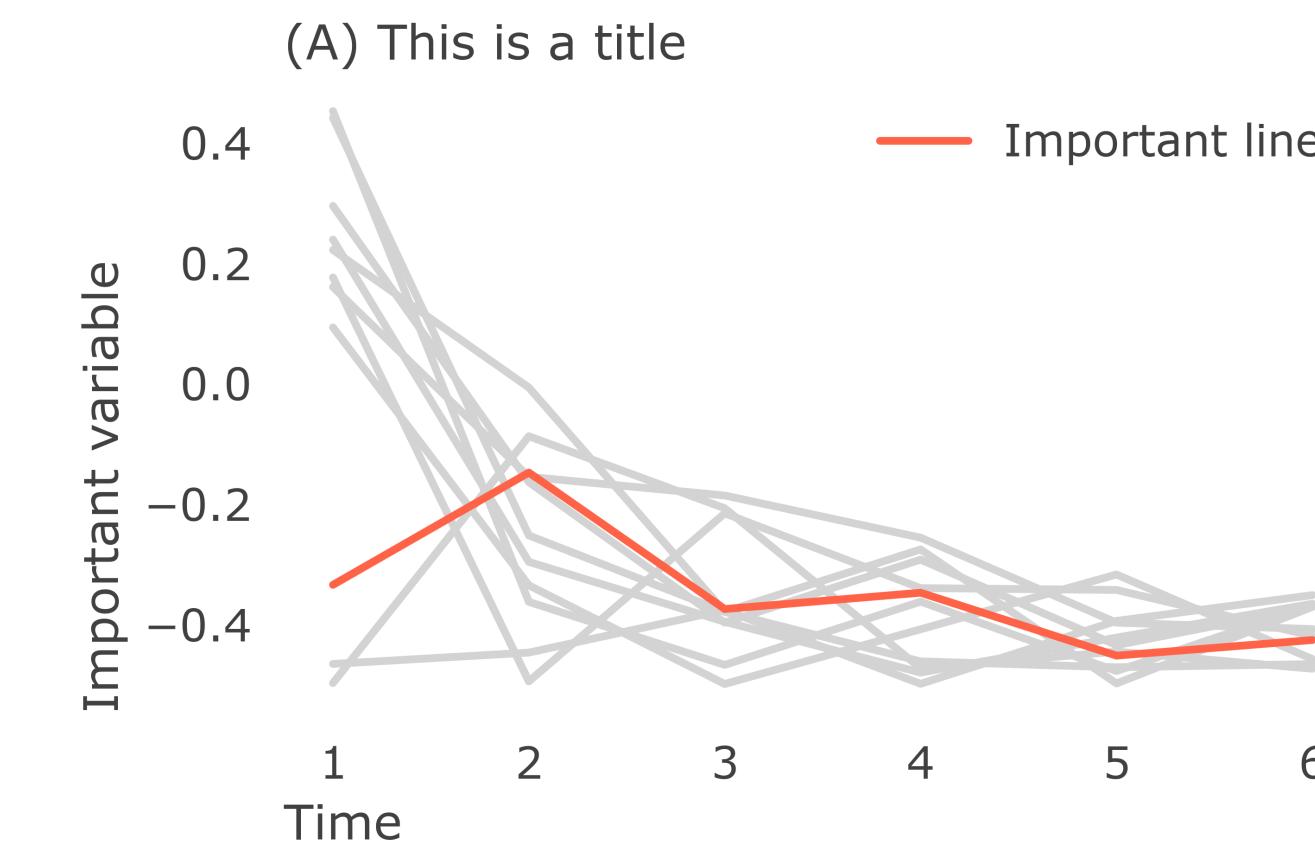
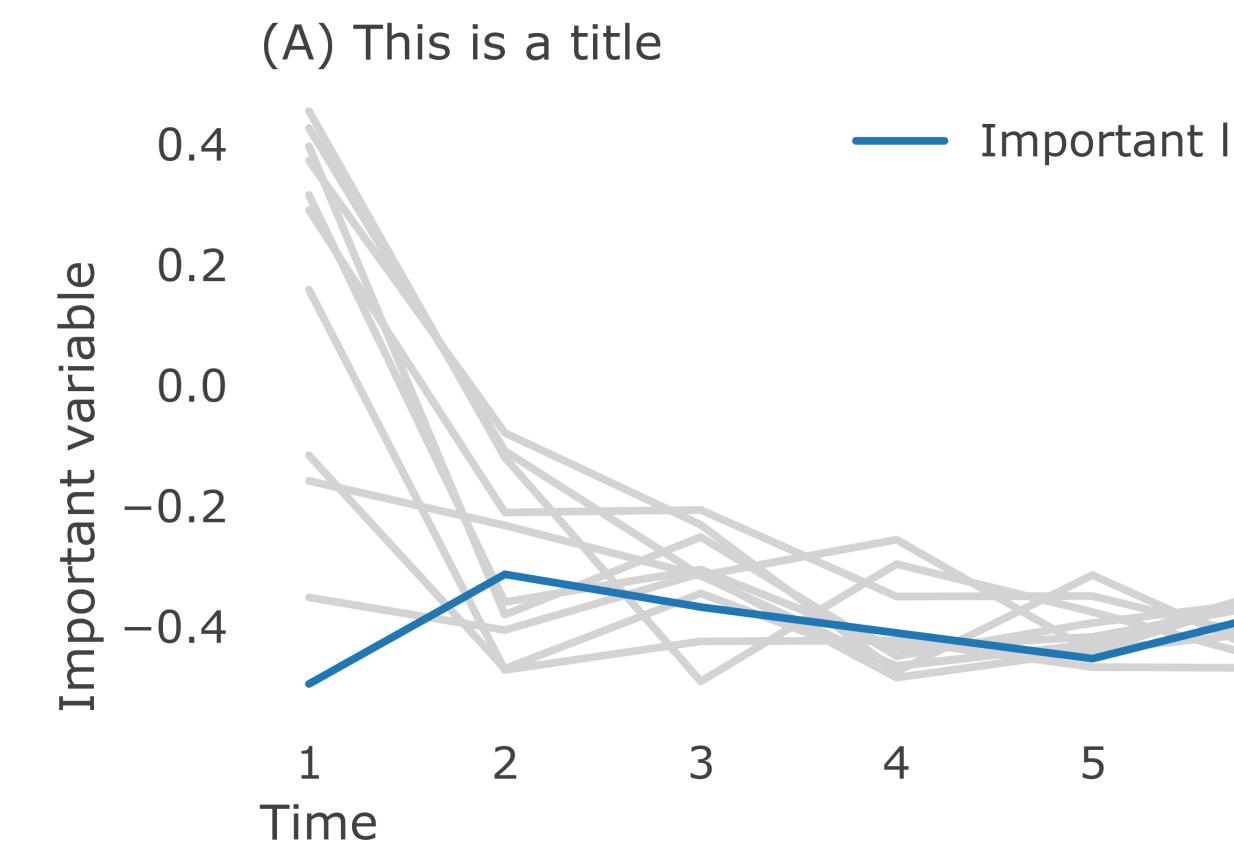


CONTRAST: ELIMINATE CLUTTER



PART 3: DESIGN

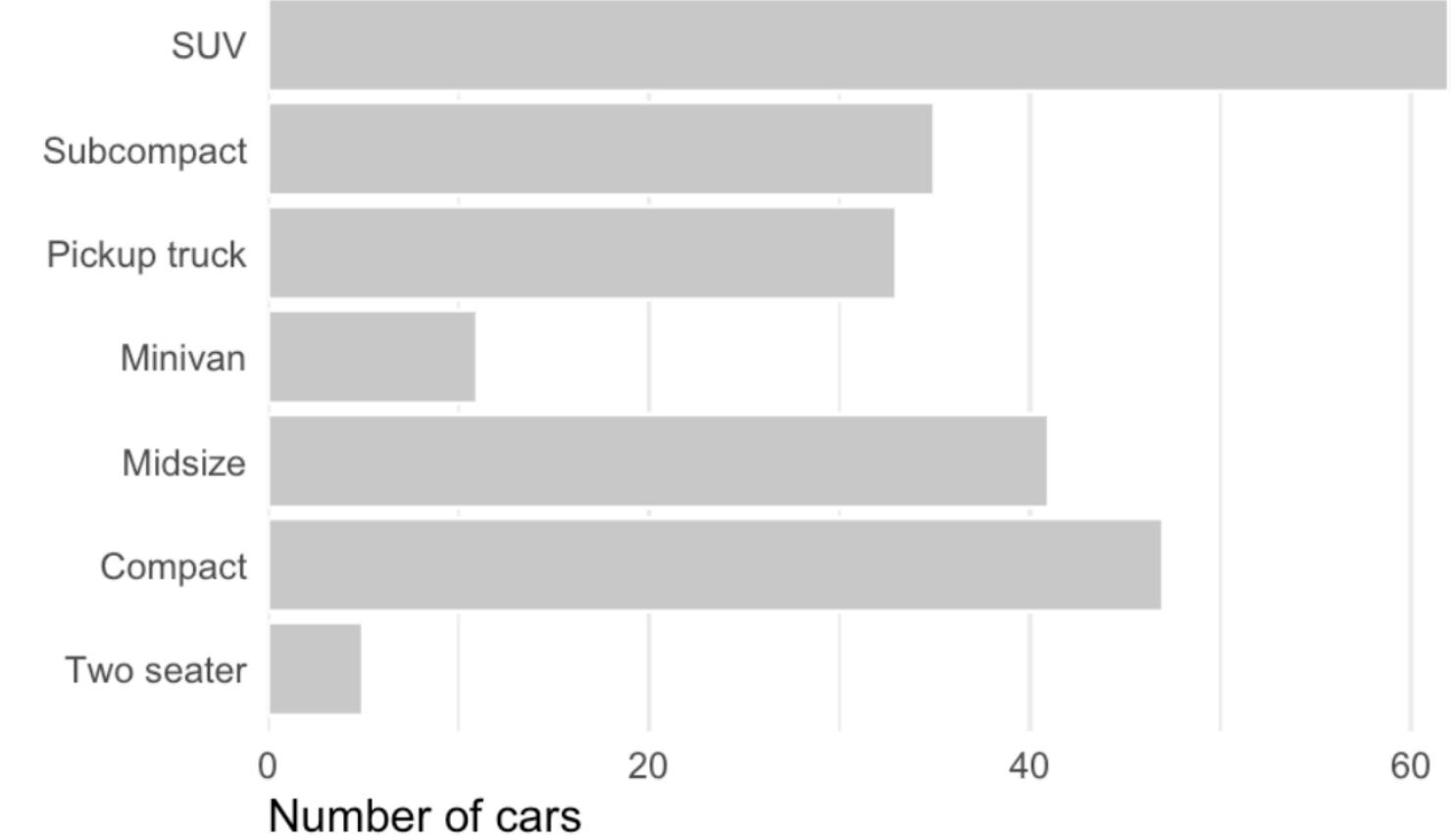
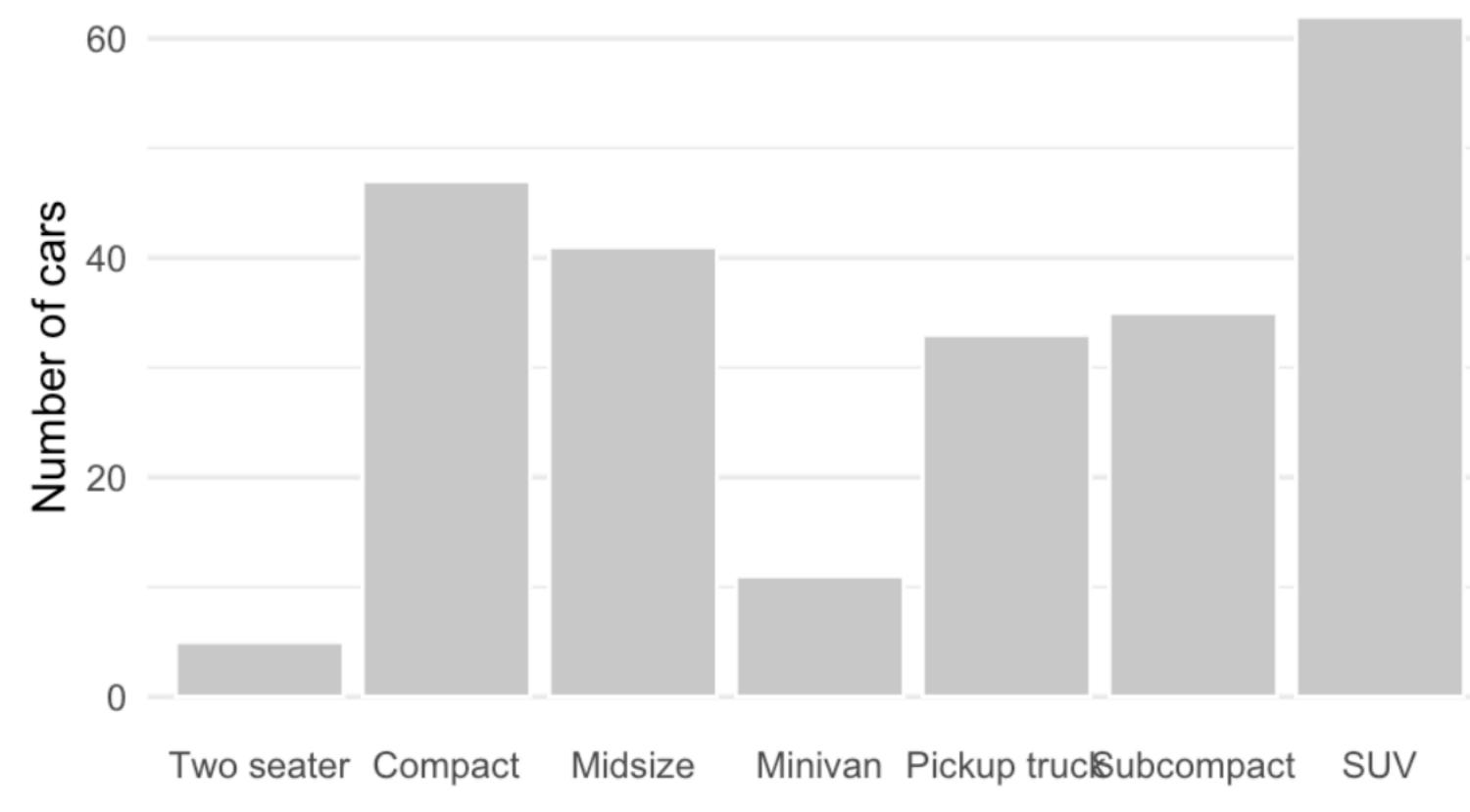
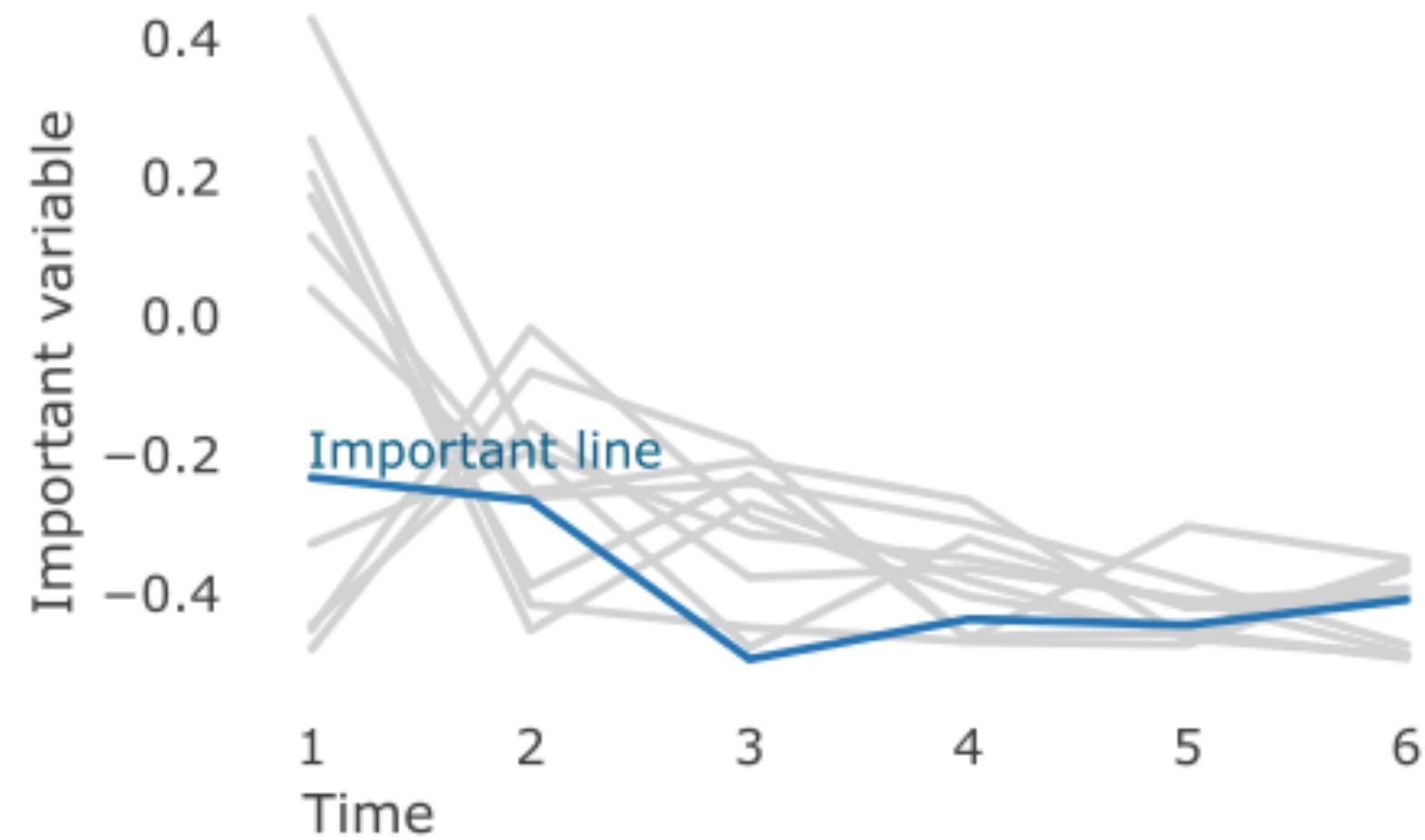
REPETITION



ALIGNMENT

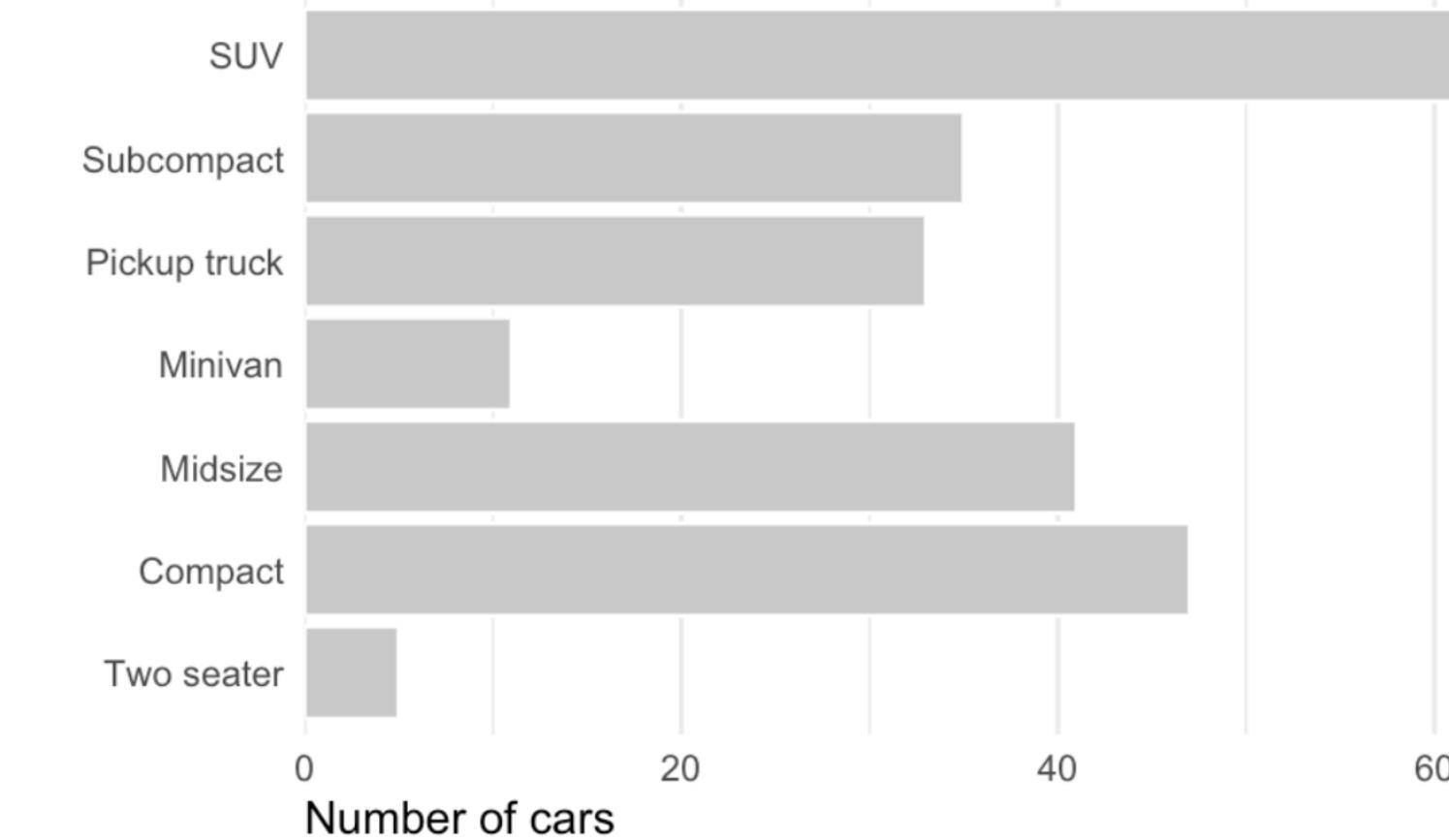
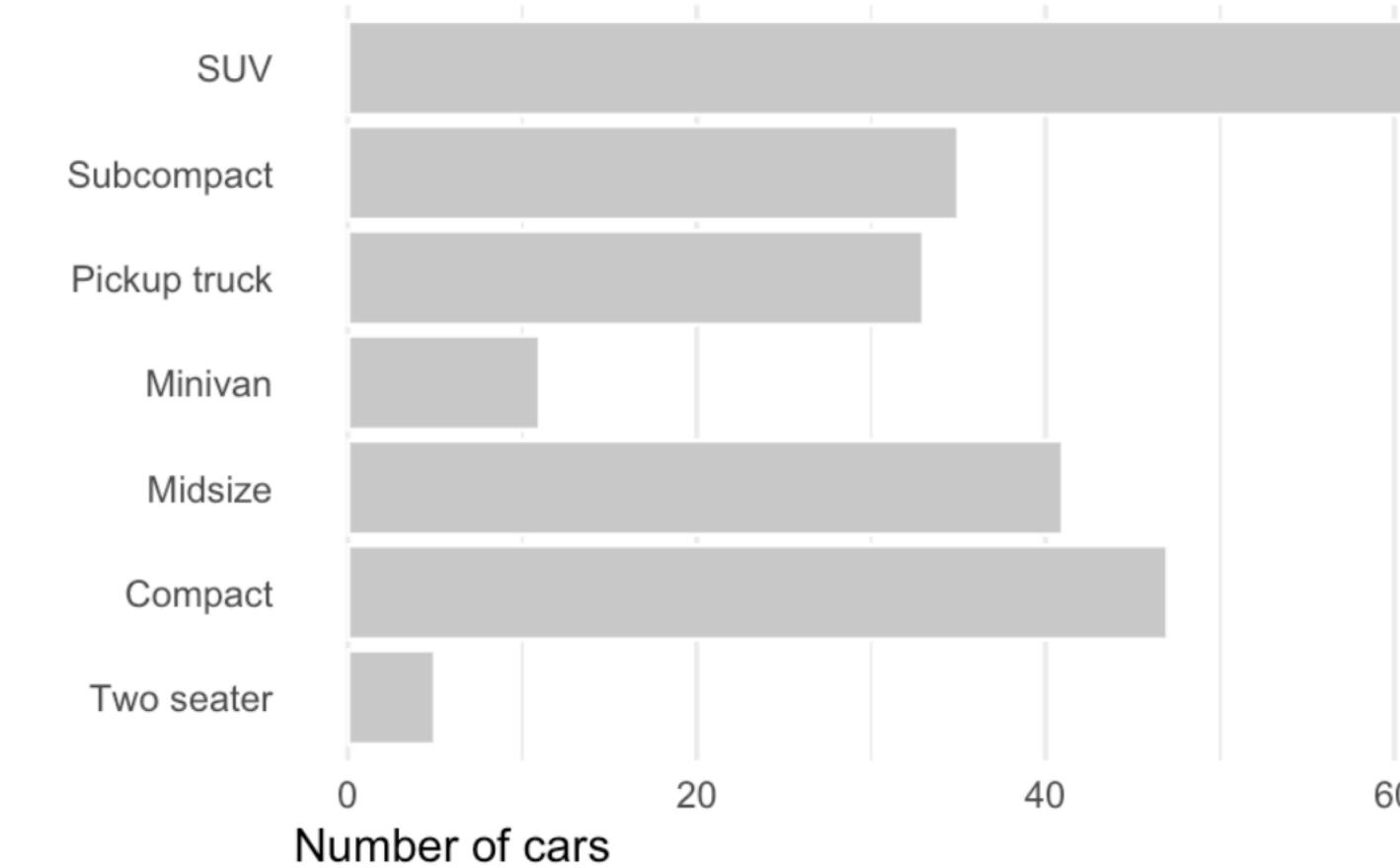
LOOKS
GOOD

DOES NOT
LOOK SO GOOD

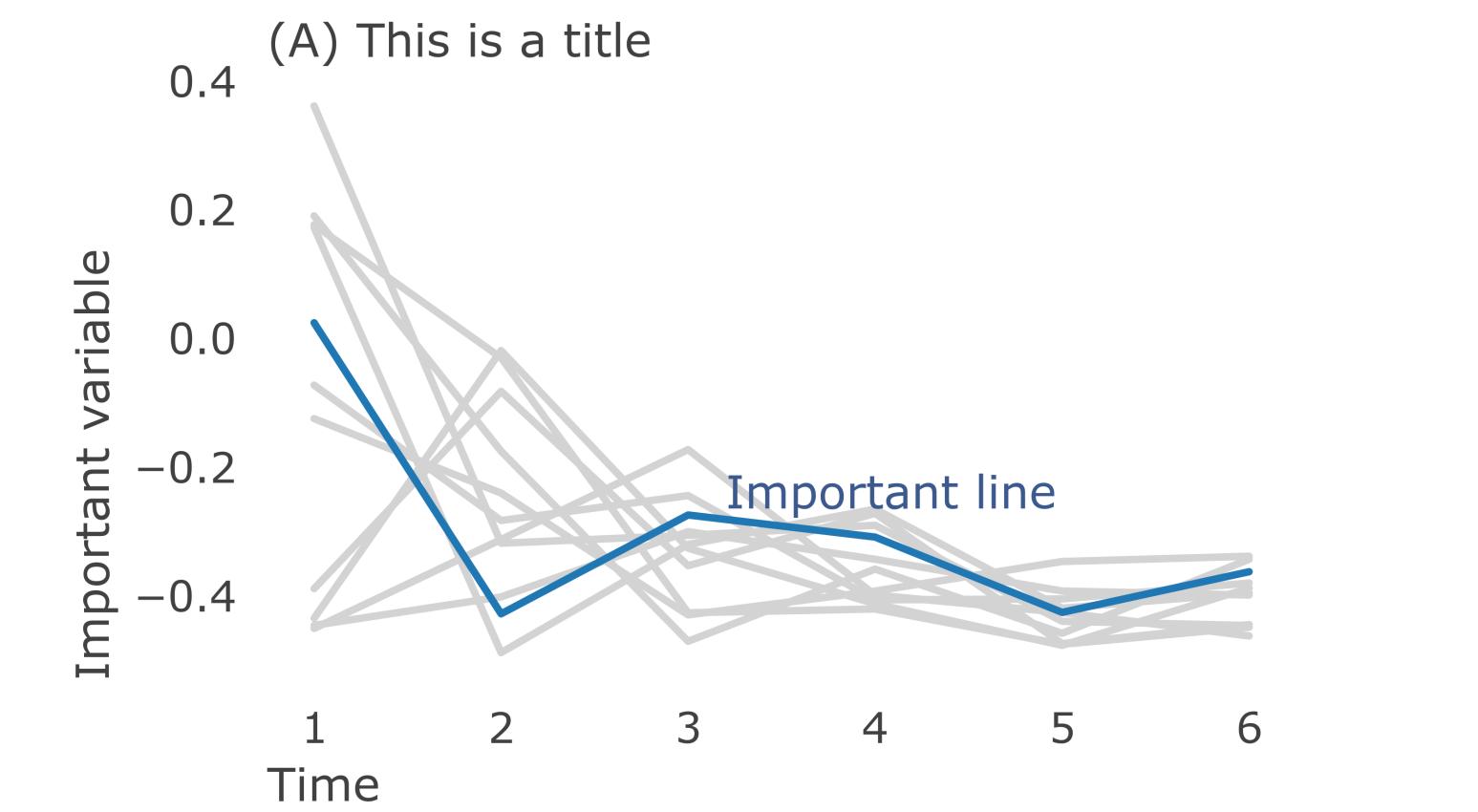
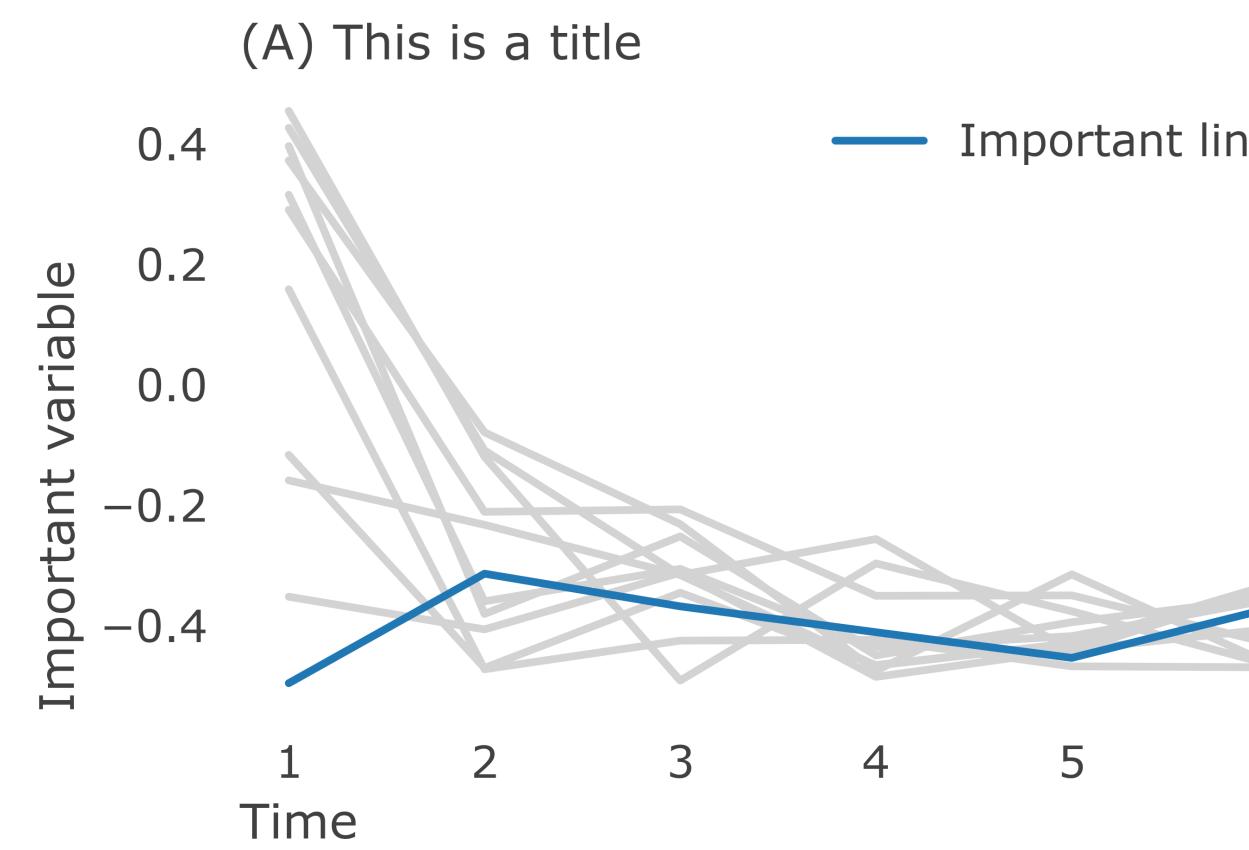
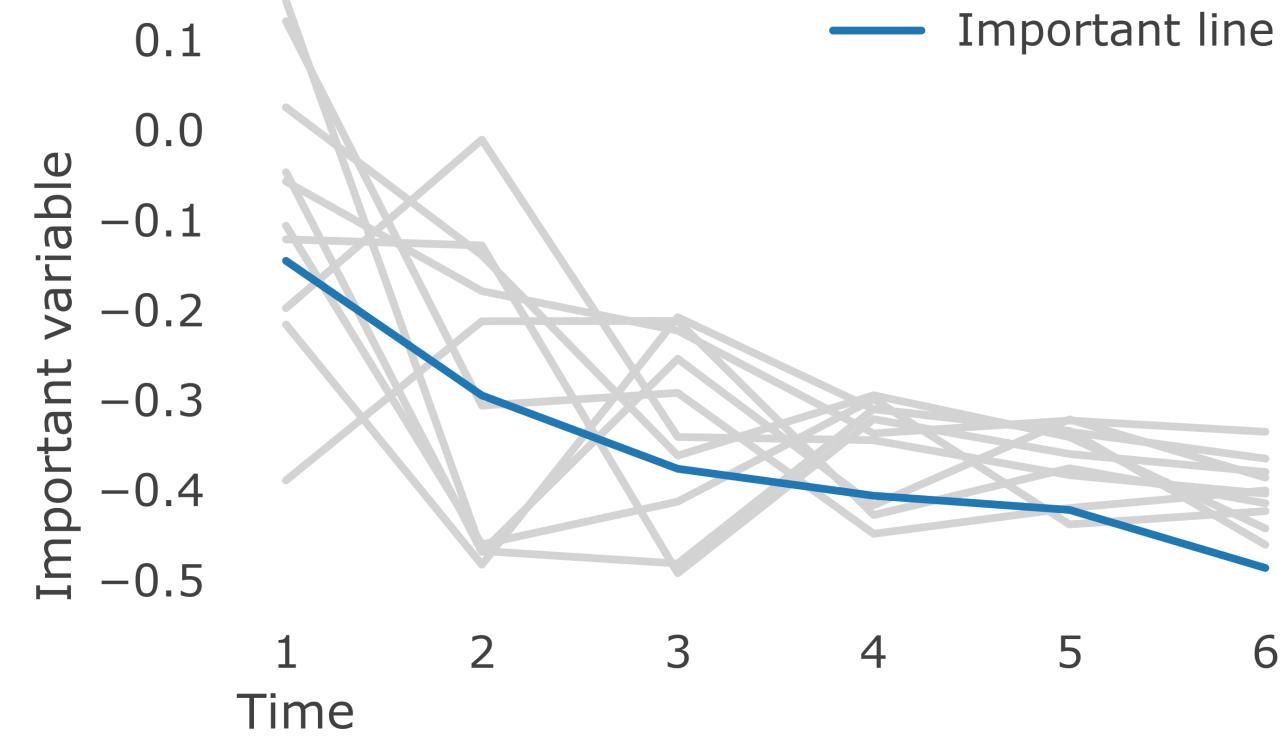


PART 3: DESIGN

PROXIMITY



(A) This is a title



PRINCIPLE 1

Contrast blah

PRINCIPLE 2

Repetition blah

Principle 3

Alignment blah

PRINCIPLE 4

Proximity blah

PRINCIPLE 1

Contrast blah

PRINCIPLE 2

Repetition blah

Principle 3

Alignment blah

PRINCIPLE 4

Proximity blah

PRINCIPLE 1

Contrast blah

PRINCIPLE 2

Repetition blah

PRINCIPLE 3

Alignment blah

PRINCIPLE 4

Proximity blah

PRINCIPLE 1

Contrast blah

PRINCIPLE 2

Repetition blah

PRINCIPLE 3

Alignment blah

PRINCIPLE 4

Proximity blah

PRACTICAL GUIDE

- Reduce cognitive load:
 - Removing unnecessary clutter
 - More professional/aesthetically pleasant
- Contrast:
 - Eliminate unnecessary lines (all frames, use gray grid lines, etc)
 - Don't use a gray background
 - White space is your friend (allows for "breathing")
 - Enlarge the labels
 - Use vector graphics (svg/pdf/eps) to avoid blurry figures -> Edit them in AI or Inkscape
- Repetition: Be consistent in different figures
- Alignment: Make sure you align subplots/labels
- Proximity: When possible, label data directly (instead of using legends)

PRINCIPLES OF VISUAL PERCEPTION

GESTALT PRINCIPLES OF VISUAL PERCEPTION

Principles/laws of human perception that describe how humans group similar elements, recognize patterns and simplify complex images when we perceive objects



PROXIMITY

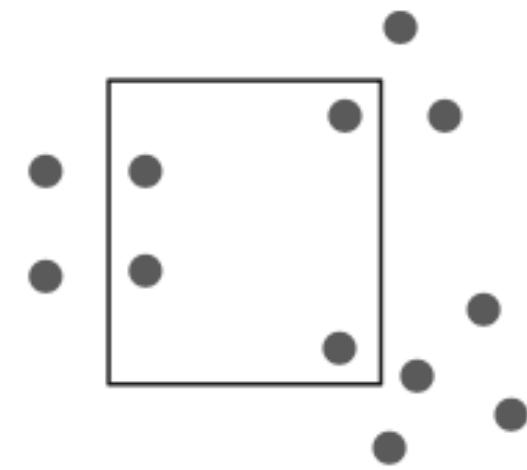
e.g. barplot: Bars next to each other are perceived as related



SIMILARITY

e.g. color across plots are perceived as related (be consistent)

GESTALT PRINCIPLES OF VISUAL PERCEPTION



ENCLOSURE

Enclosed areas contain related objects (highlight areas)



CLOSURE

No need to close frames

GESTALT PRINCIPLES OF VISUAL PERCEPTION



CONTINUITY

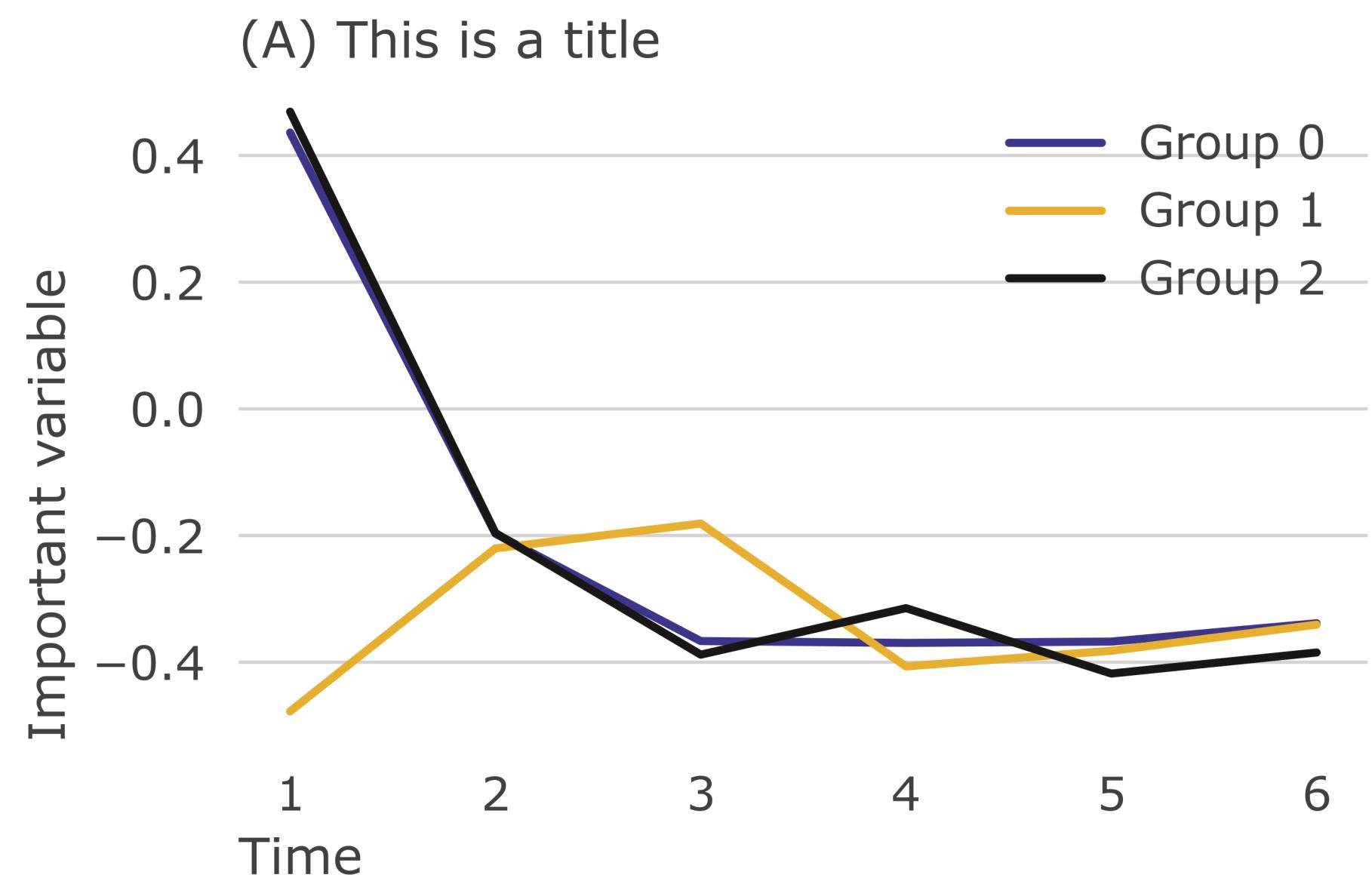
e.g., No need to have a left line



CONNECTION

e.g., Networks

PART 3: DESIGN



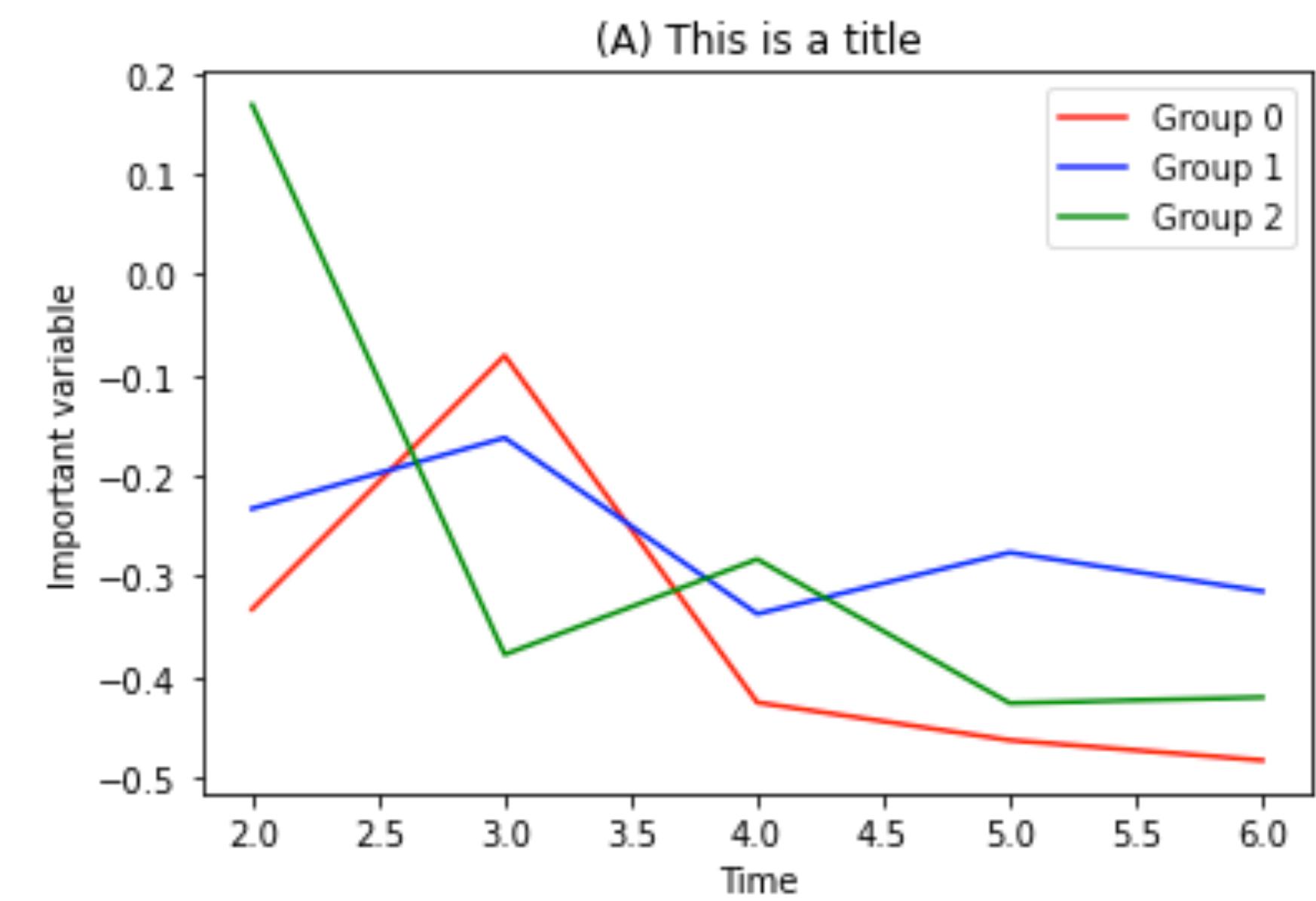
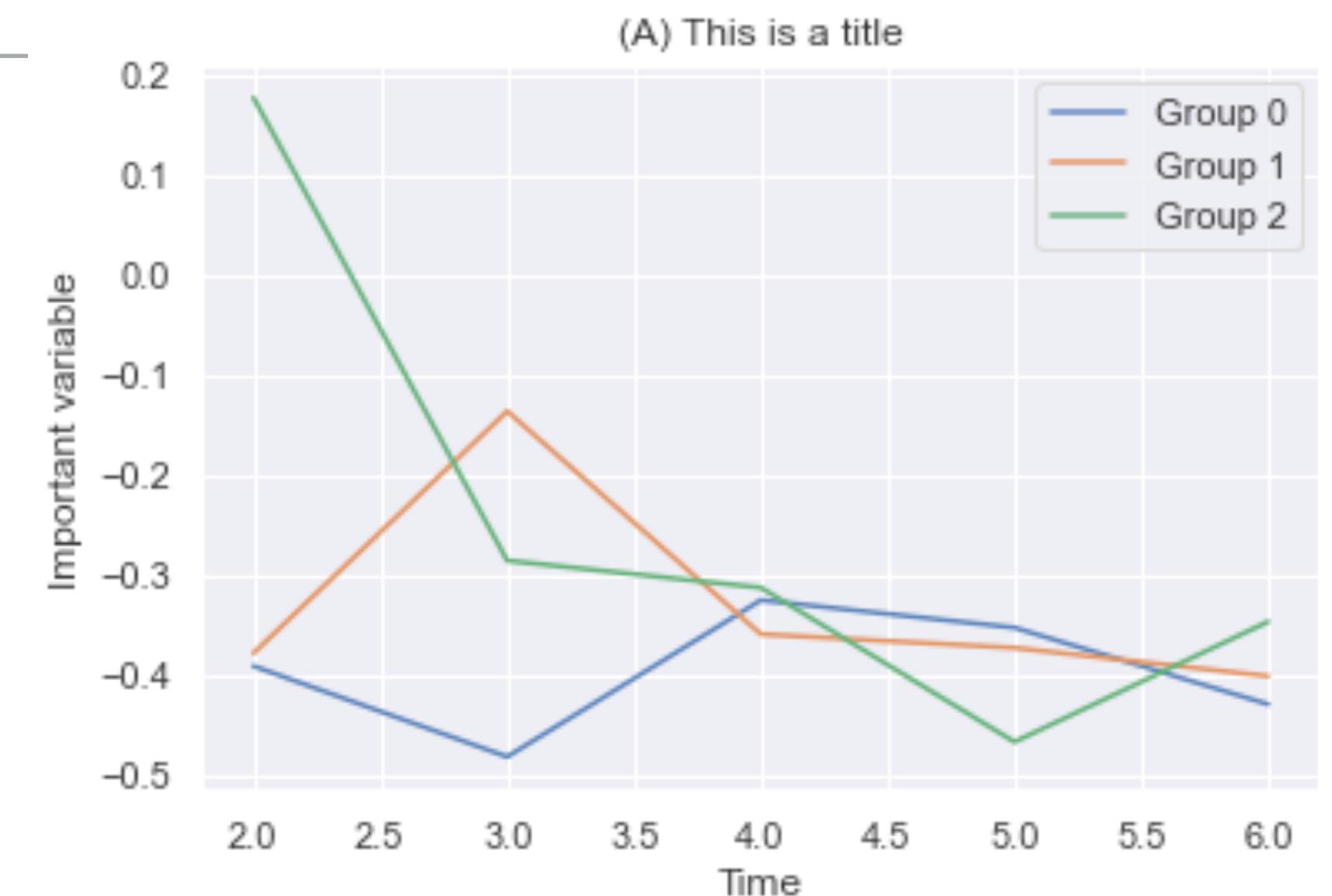
Closure: No need for a frame, we understand that it is one plot

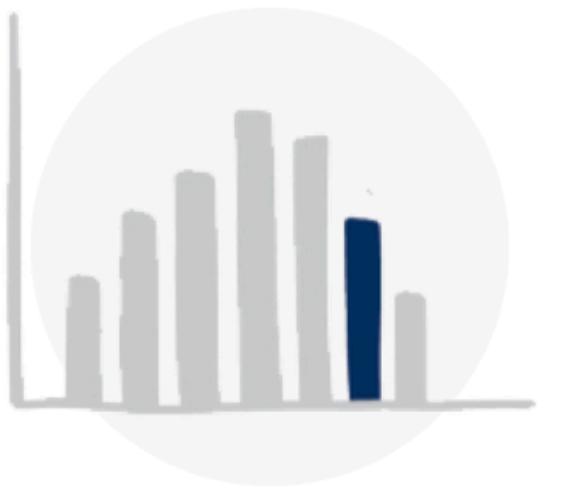
Proximity: We know that the labels belong to the axes because they are nearby

Continuity: 2, 3, 4, 5, 6 are perceived as connected

Connection: All dots within one line

Similarity: The three lines represent similar data





TENSION

focus

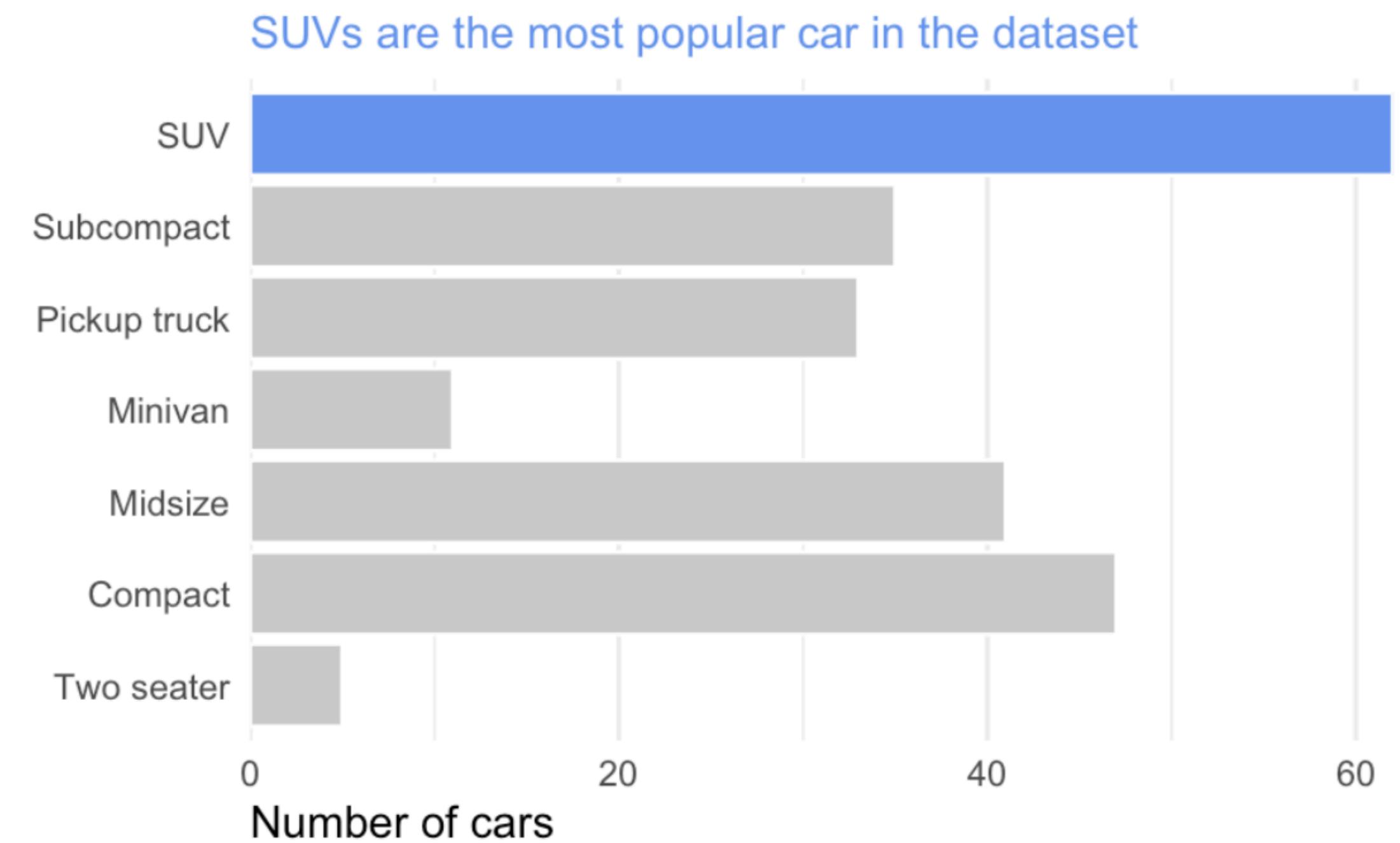
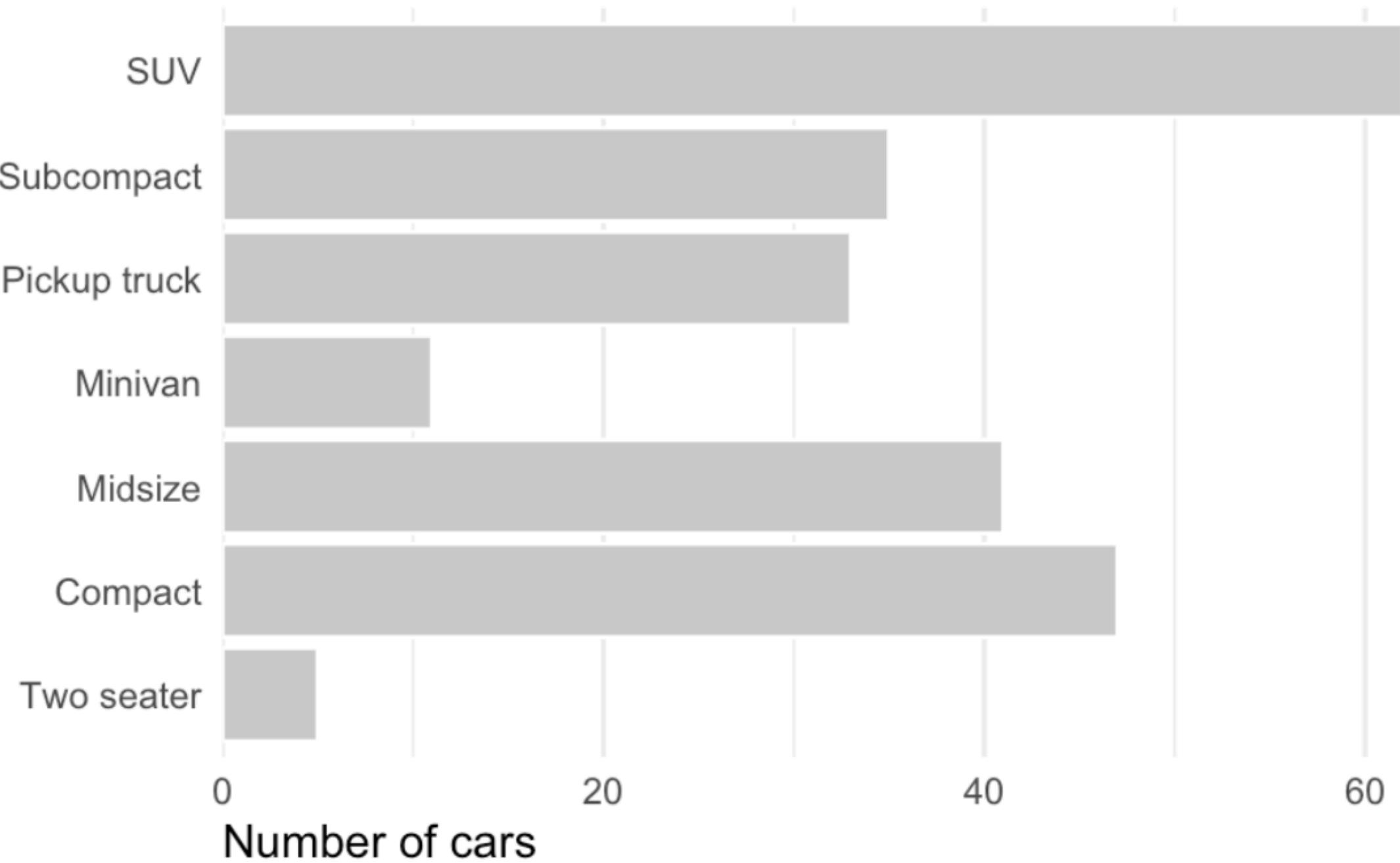
attention

GUIDE THE READER:

PRACTICAL GUIDES

TEXT

- Compare how you read the figure in the right and in the left



WHY DO WE NEED TO FOCUS ATTENTION?

Help the person interpret the plot (reduce cognitive load, make it enjoyable)

HOW DO WE FOCUS ATTENTION?

Using pre-attentive attributes

PART 3: DESIGN

756395068473
658663037576
860372658602
846589107830

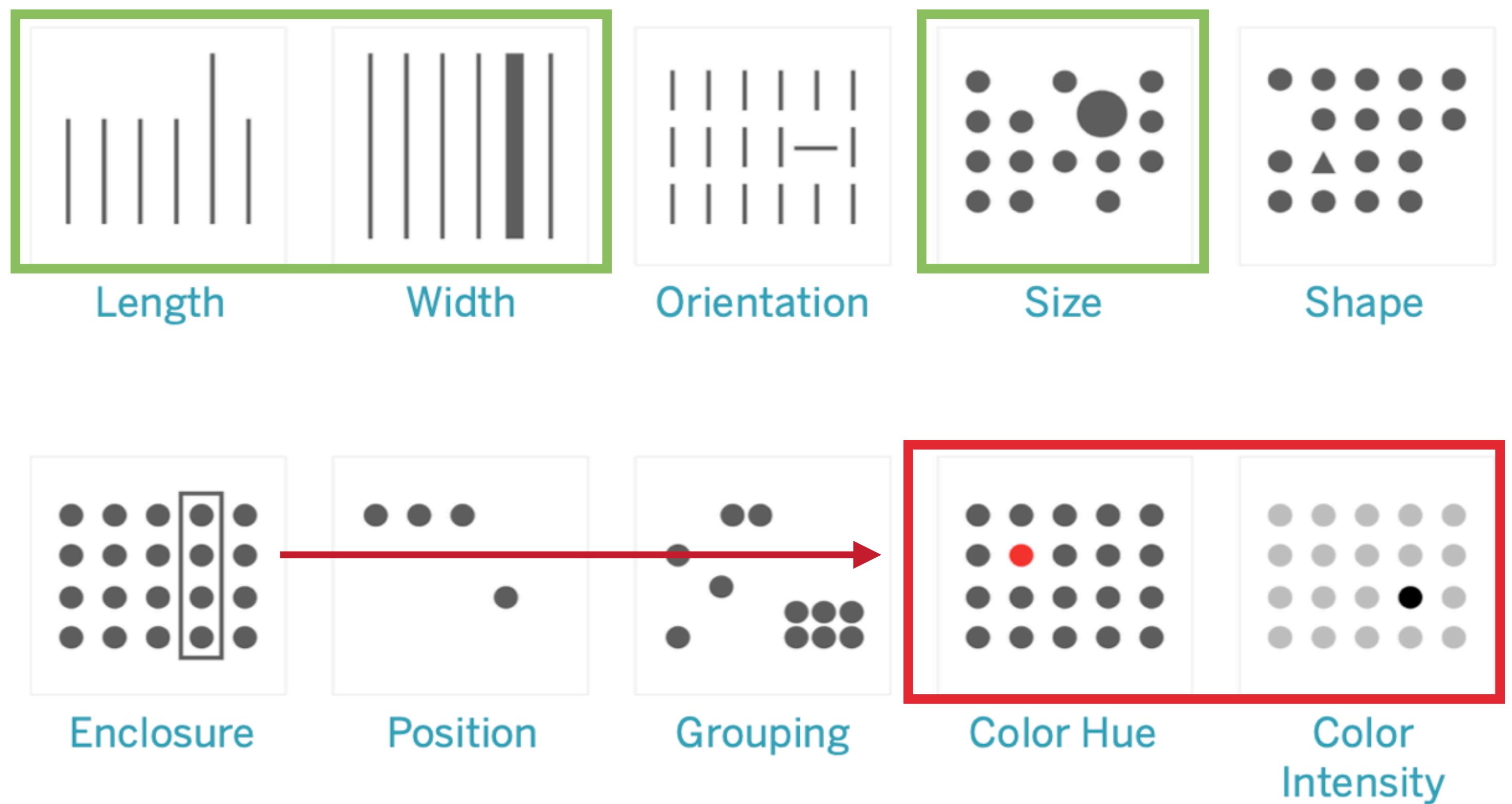
756**3**9506847**3**
65866**3**037576
860**3**72658602
8465891078**3**0

Story telling with data

IMPORTANT PRE ATTENTIVE ATTRIBUTES

What do we focus on:

- ▶ Large objects
- ▶ Bright objects
- ▶ Contrasting objects



Color makes ice cream taste sweeter,
veggies taste fresher,
and coffee taste richer

Ellen Lupton

COLOR

The most useful **pre-attentive attribute**

- ▶ Increases contrast
- ▶ Allows for consistency

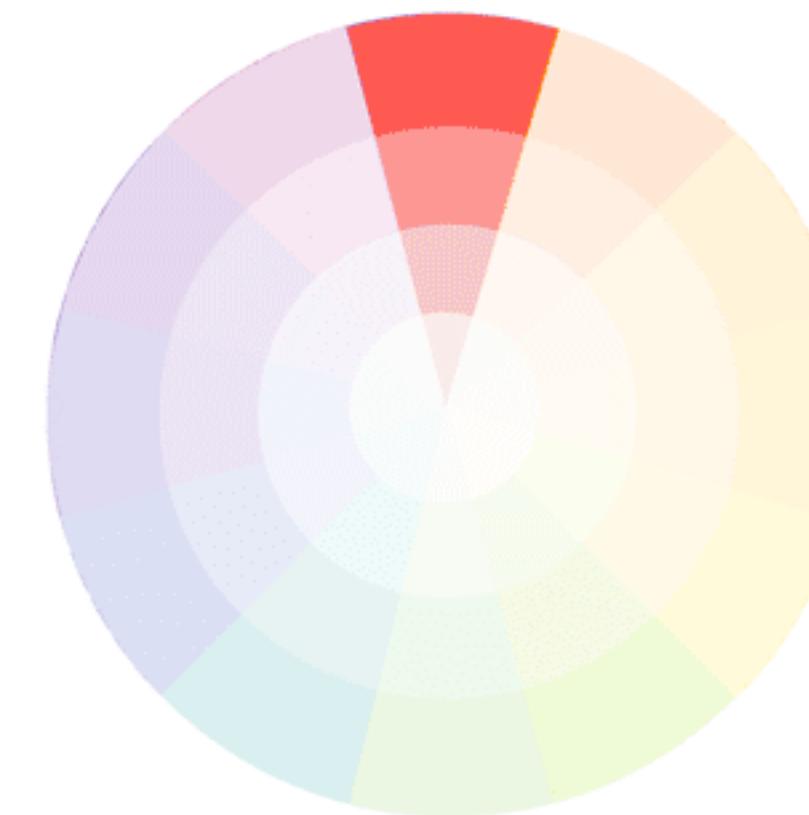
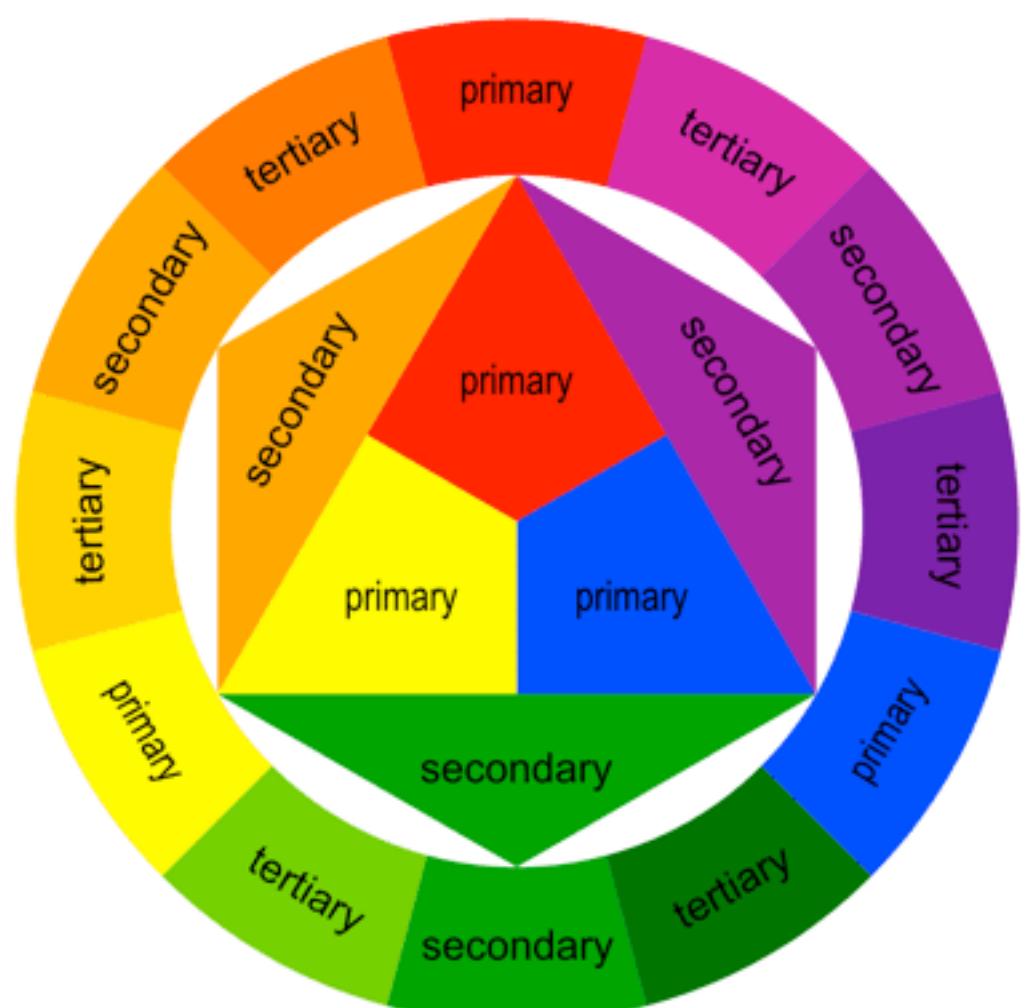
Color affect emotion, culture-dependent. But some responses are nearly universal

- ▶ Warm colors --> alive/alert
- ▶ Blue colors --> calming/focus

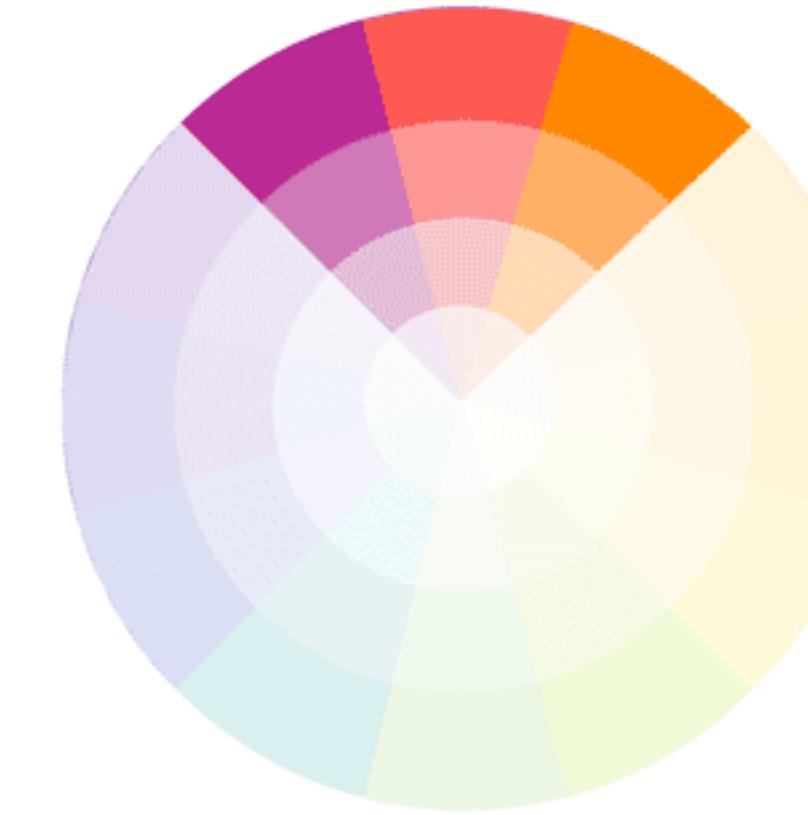
<https://blog.datawrapper.de/which-color-scale-to-use-in-data-vis/>

<https://davidmathlogic.com/colorblind/#%23D81B60-%231E88E5-%23FFC107-%23004D40>

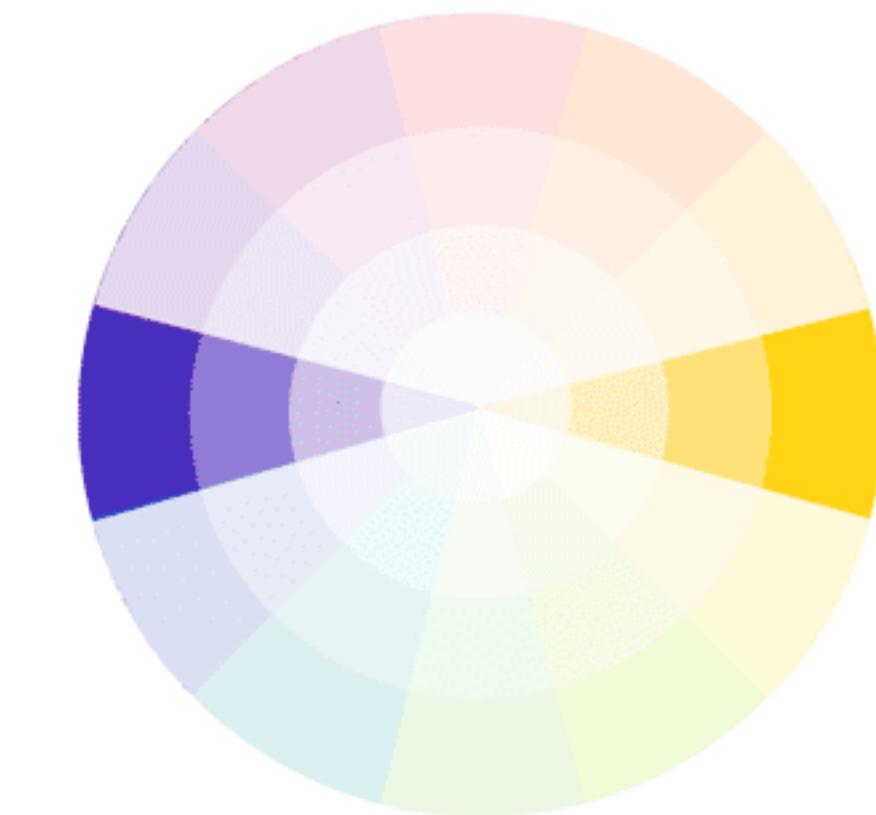
COLORS THAT LOOK GOOD TOGETHER



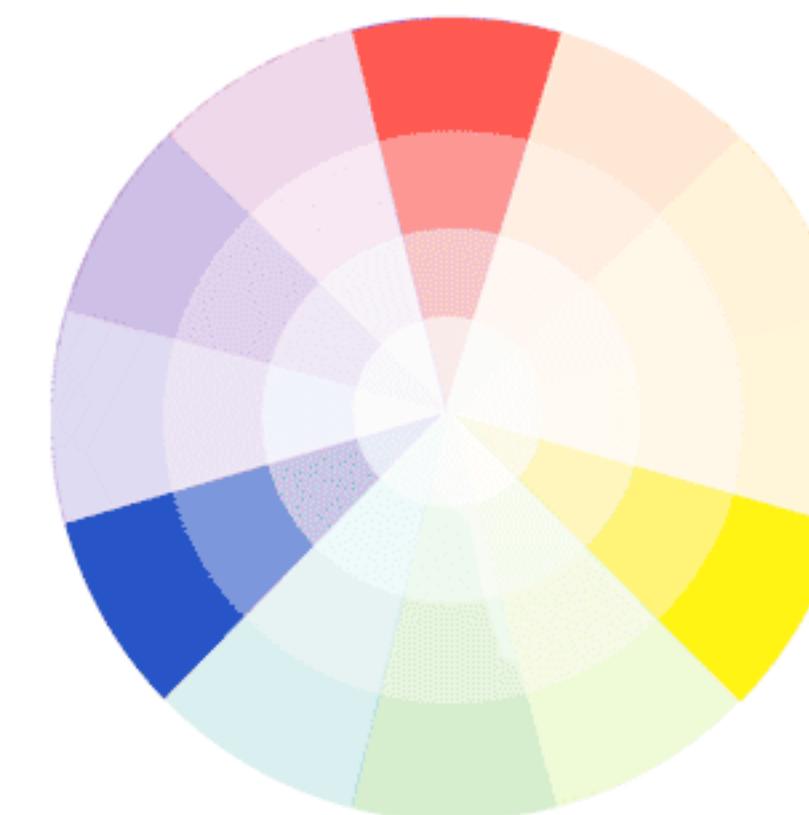
Monochromatic



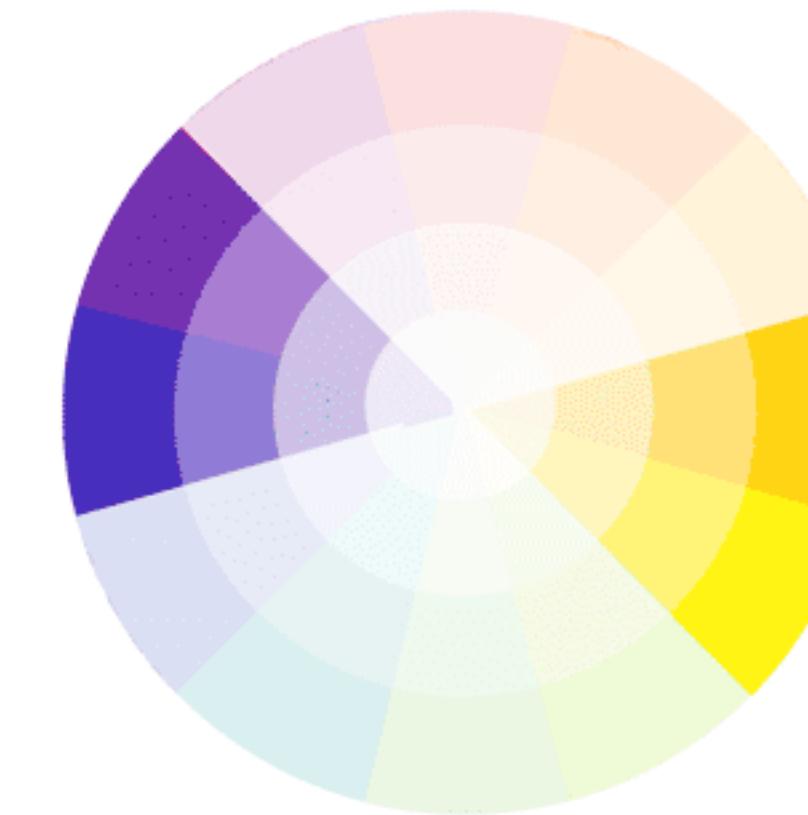
Analogous



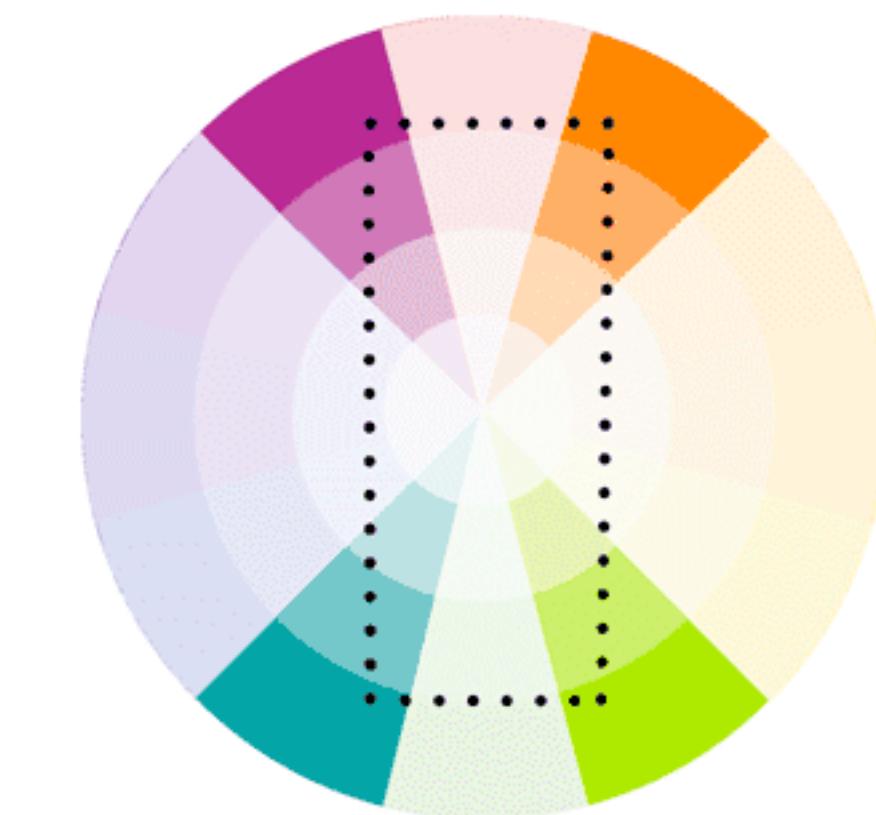
Complementary



Triad



Split-Complementary



Tetradic

COLOR PALETTES

SEQUENTIAL

Minimum is important



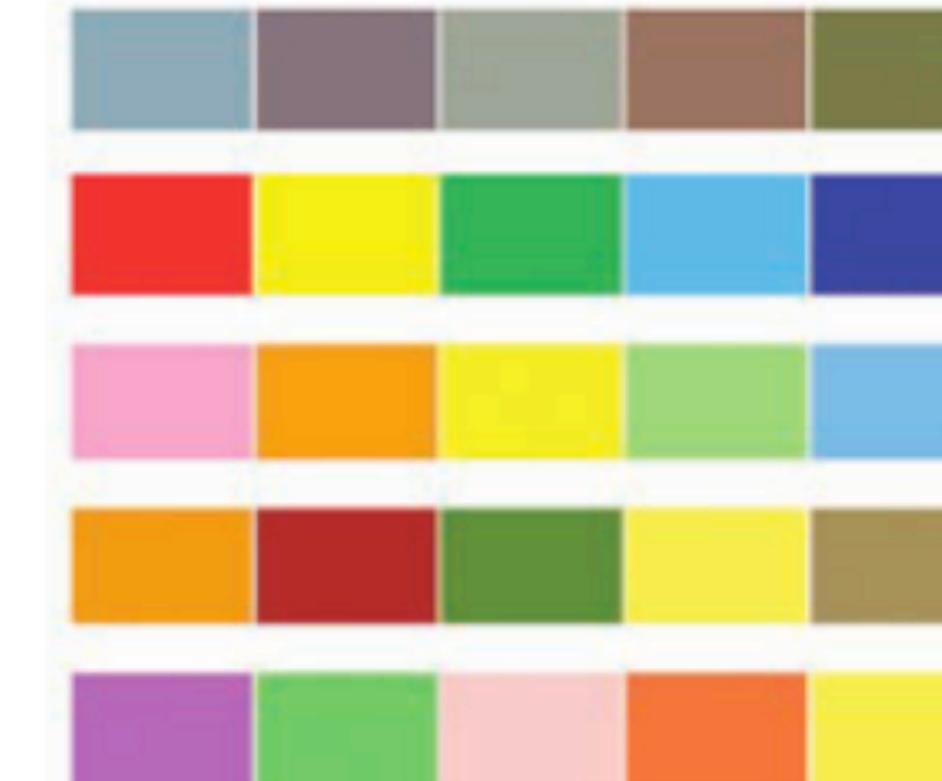
DIVERGING

Mean is important



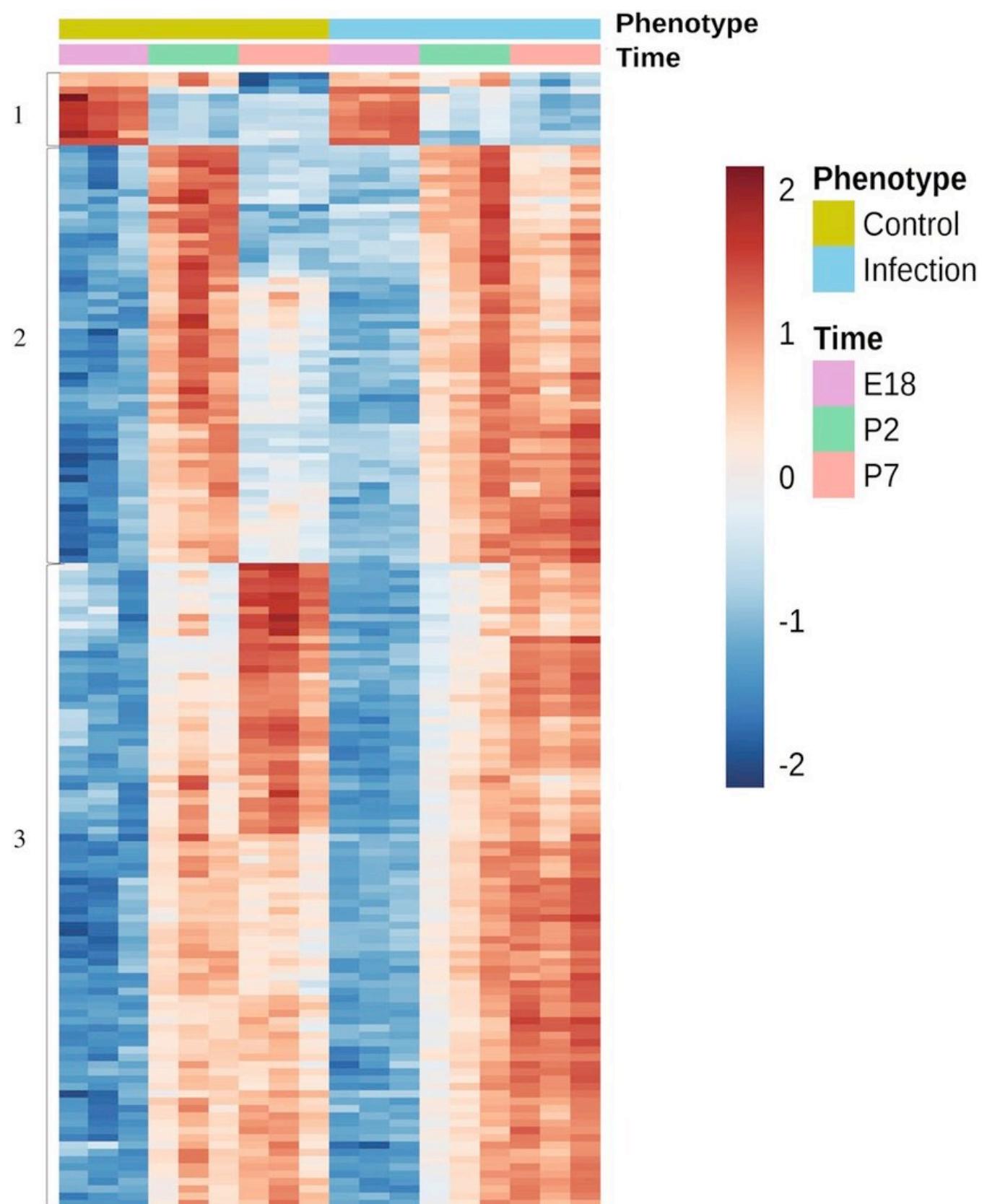
QUALITATIVE

Represent categories



COLOR TO REPRESENT VALUES

ONLY IF THE EXACT VALUES ARE NOT IMPORTANT



PRACTICE: MAKE YOUR VIZ BEAUTIFUL AND EFFECTIVE

- ▶ Think of the CRAP principles. Are you following them?
- ▶ Color:
 - ▶ Which reason? (guide the reader, represent categories, represent values)
 - ▶ Which palette?
- ▶ How are you using pre-attentive attributes?

PART 4

STORYTELLING

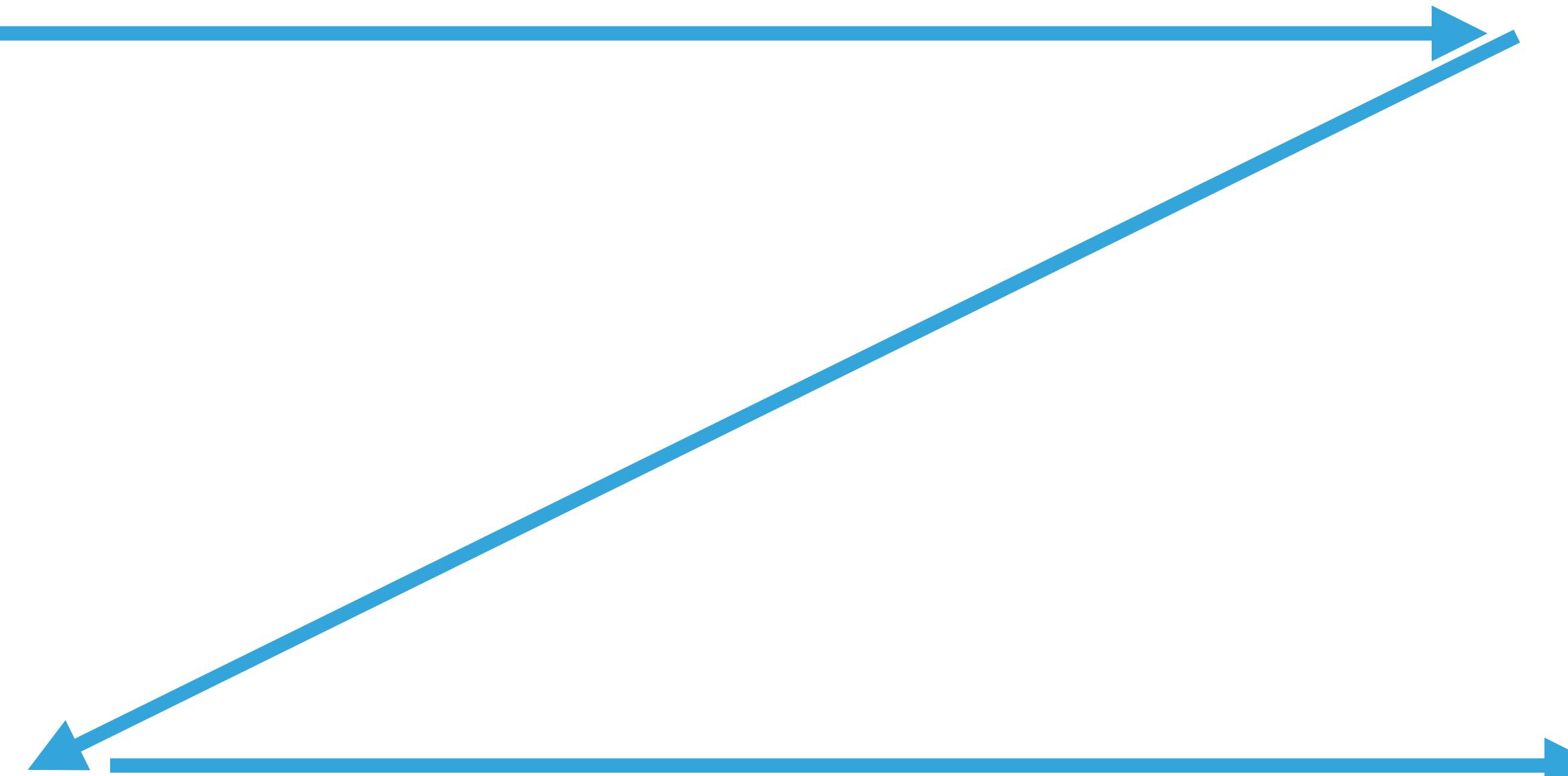
STORYTELLING

Storytelling is the **most effective tool** to make audiences enjoy a presentation, make them patient and curious to accept an idea, help them better understand an instruction, and keep them away in lectures. People love cute stuff – *Mengyan Li*

A great story does more than represent emotion from a distance. It makes us feel an emotional charge – *Ellen lupton*

You probably read this 1st

You probably read this 2nd



PART 4: STORYTELLING

HOW TO TELL A STORY

Narrative arc

Exposition: What does the reader need to know to understand the plot?

- Leverage how we read plots (Z)
- Call to action (tell the reader what do do)

Middle: What is the point of the plot?

- Use pre attentive attributes: Guide the attention
- Create emotions using color
- Based on conflict/danger

Conclusion:

- They should get the message of the plot

Russia has recorded more than 753,000 excess deaths during the pandemic, almost four times the official Covid death toll provided by state agencies

Daily **excess deaths** vs **reported deaths**, per million people

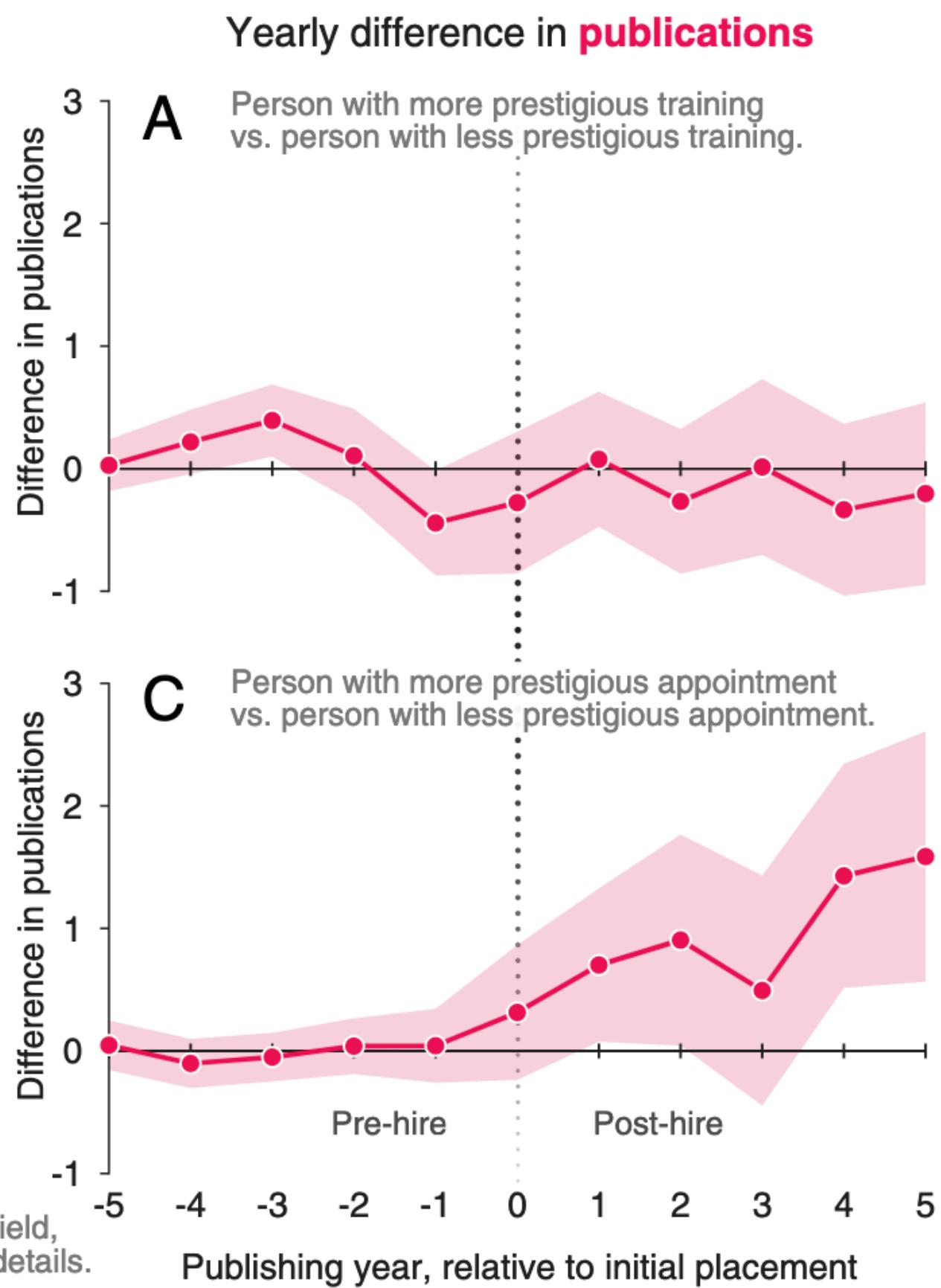
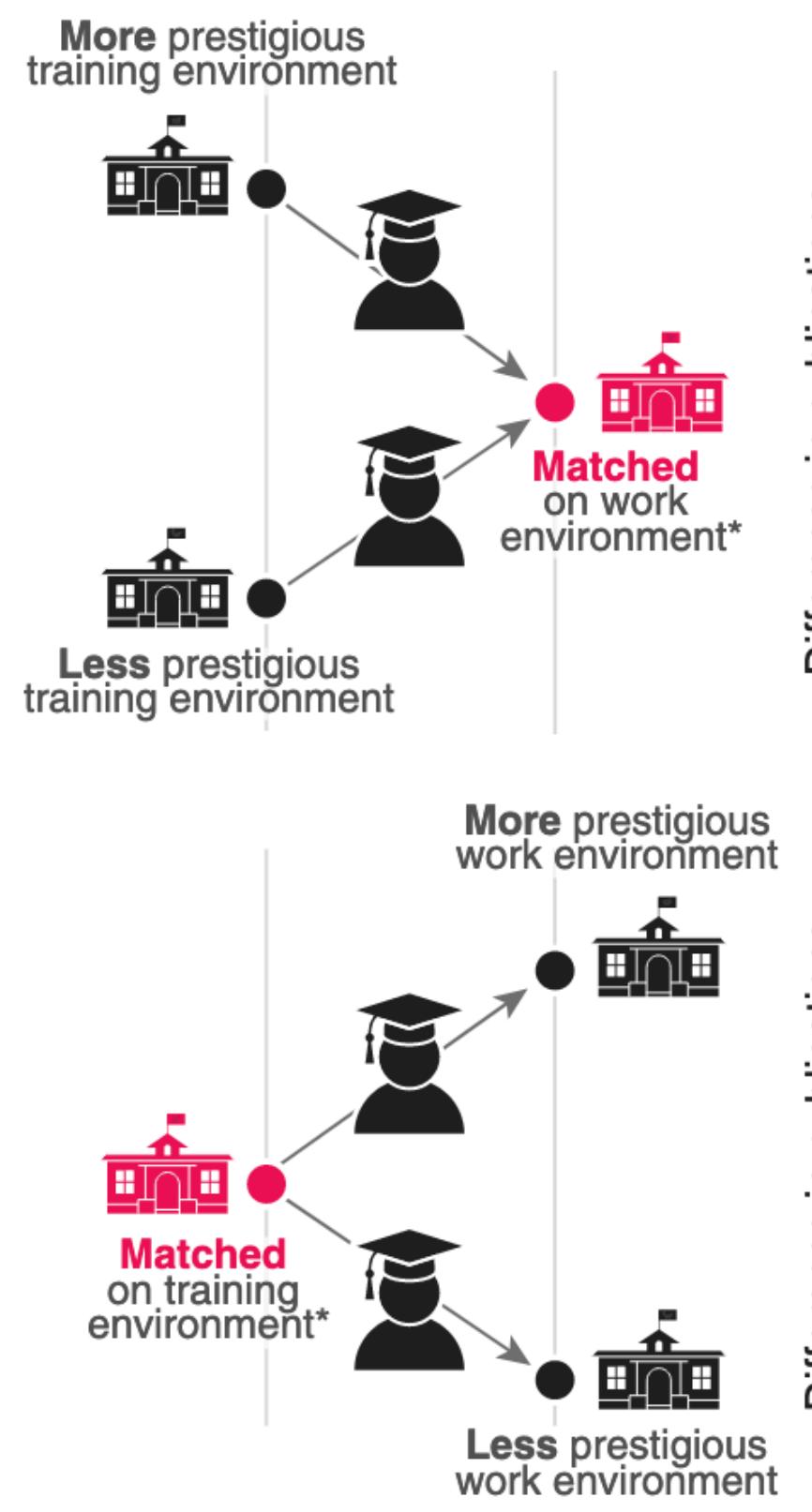


Source: Johns Hopkins CSSE; FT analysis of national mortality data and Karlinsky & Kobak's World Mortality Dataset
© FT

Exposition: Label on top-left corner, tells the reader what to do

Conflict: **Excess** and **reported** deaths are very different

Resolution: The reader has understood how it looks in different countries



Exposition: Label and drawings.

Conflict: Between both matched persons.

Resolution: Reader has learned that prestige helps get published.

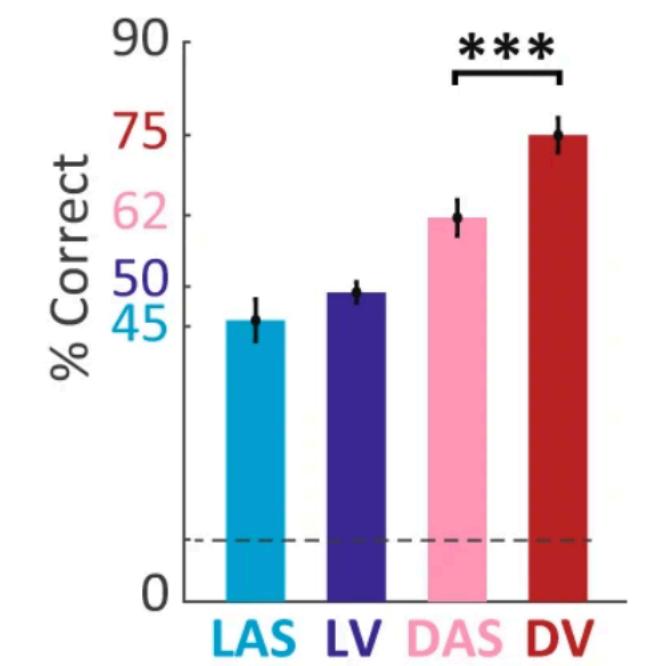
A. Behavioral experiment

Listen: Reconstructed digit sounds

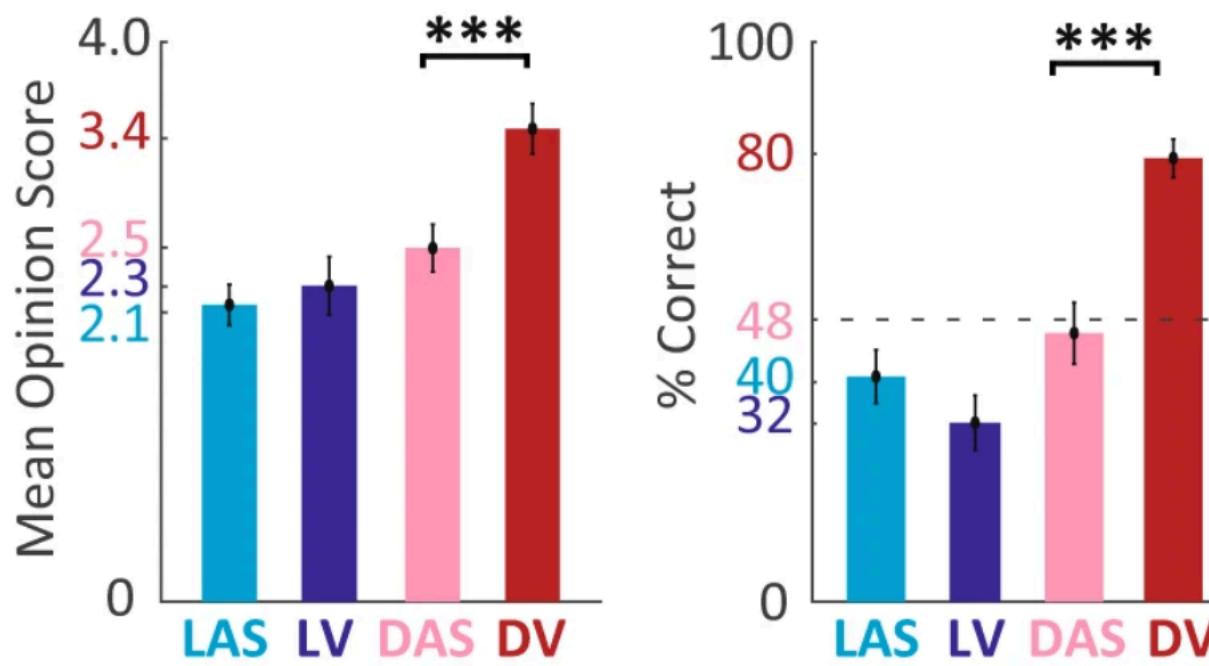
Report:

- Digit?
- Quality?
- Gender?

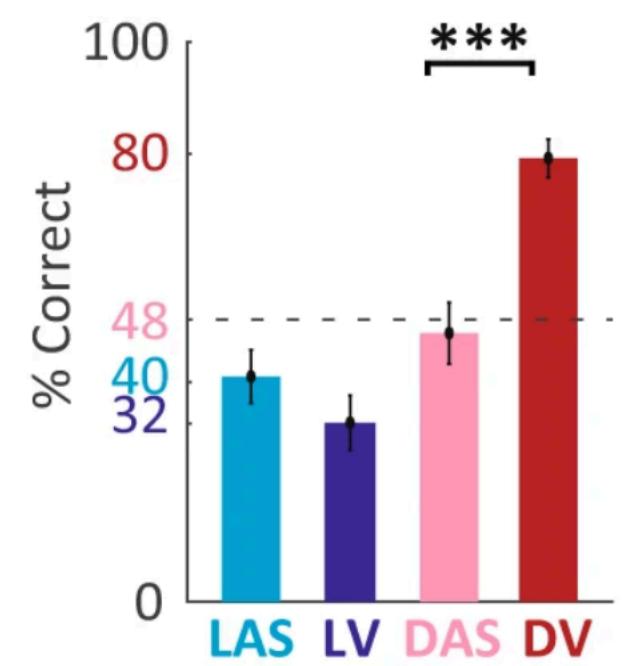
B. Digit intelligibility



C. Speech quality

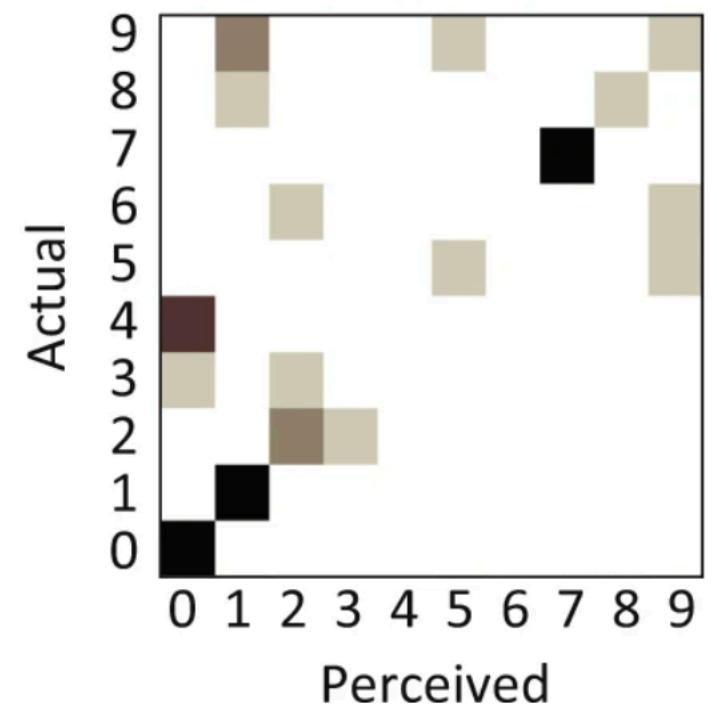


D. Gender identification

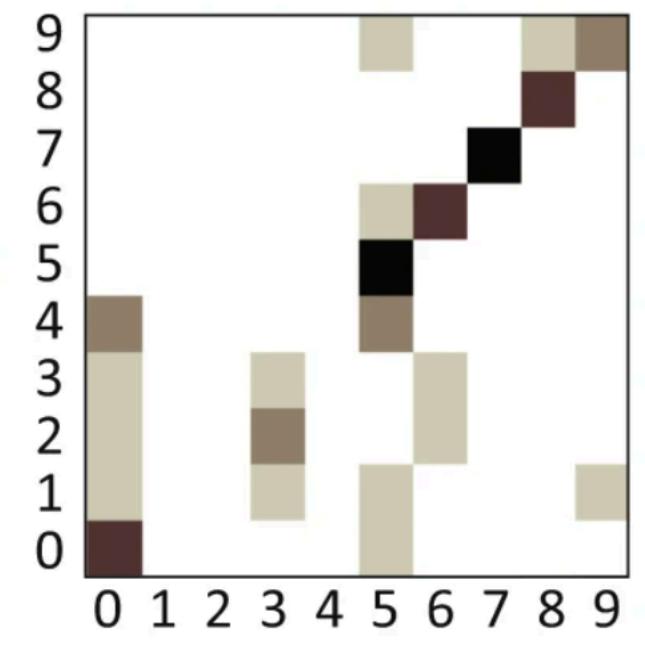


E. Digit confusion patterns

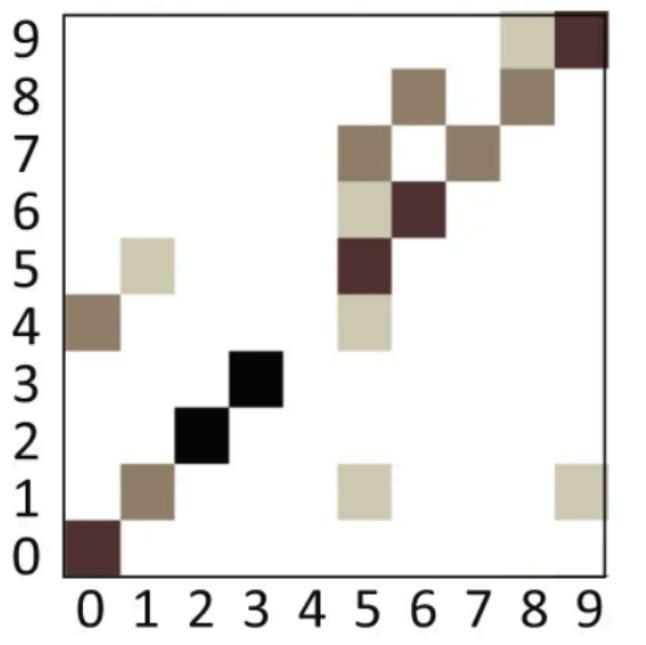
Lin Reg Aud Spec



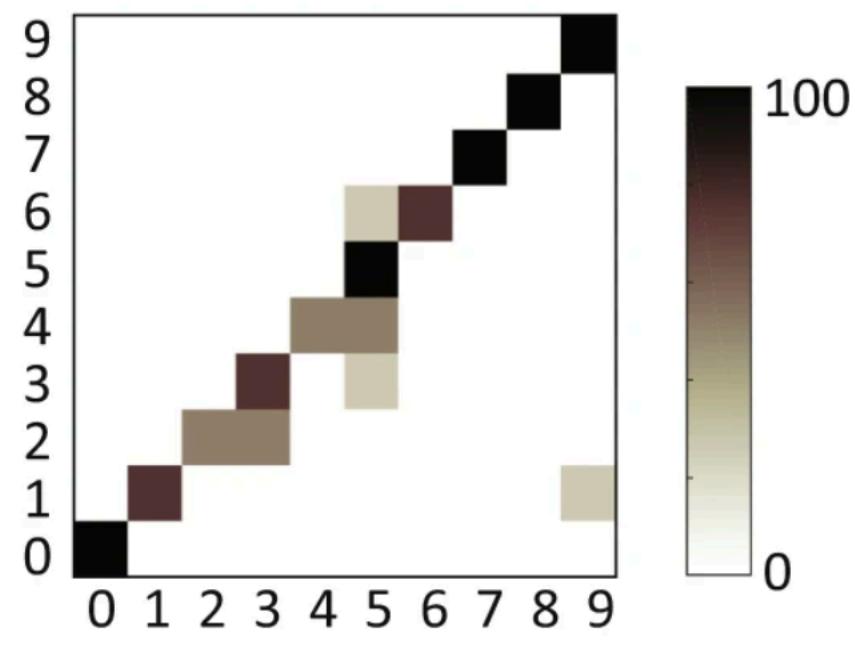
Lin Reg Vocoder



DNN Aud Spec



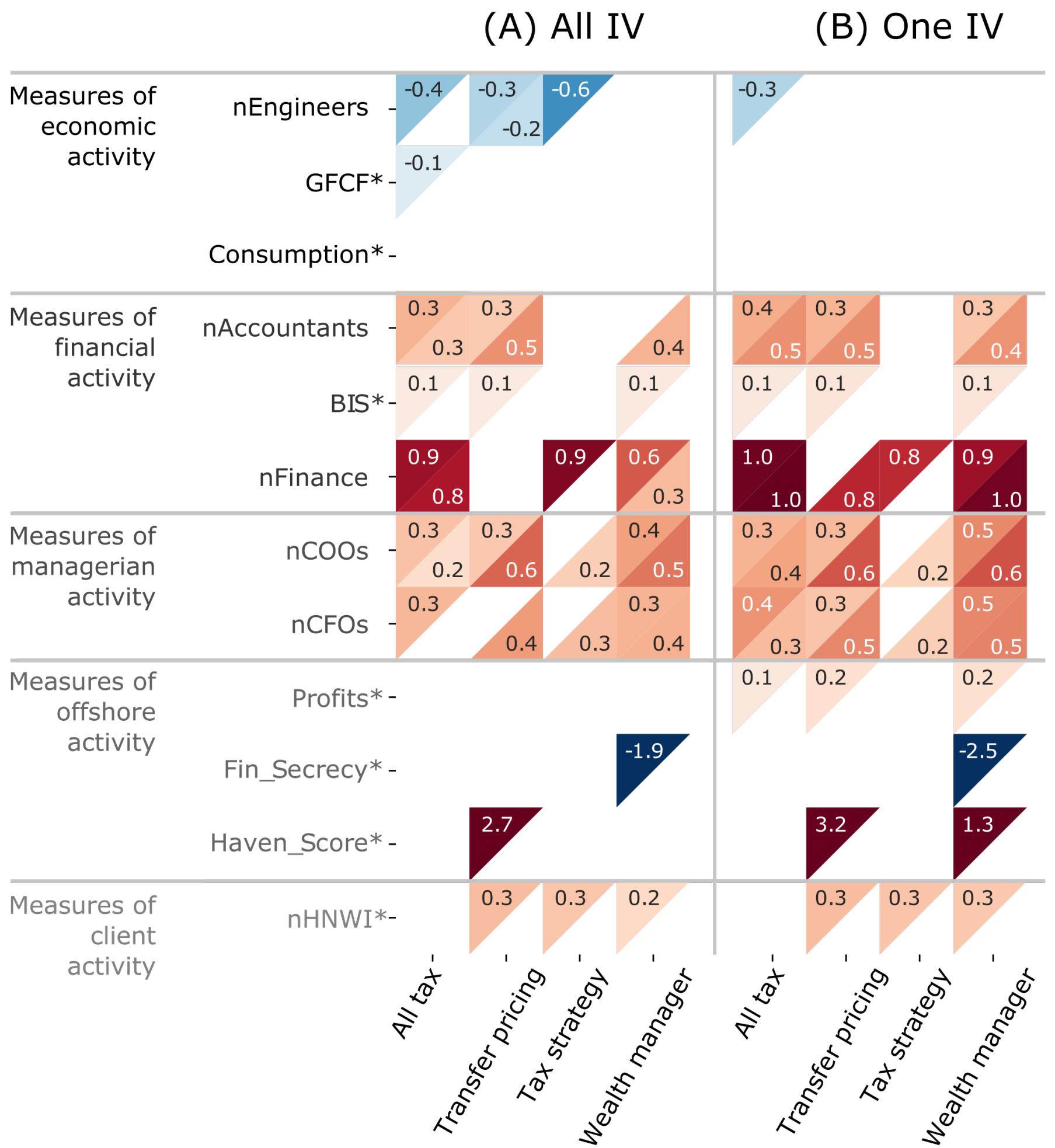
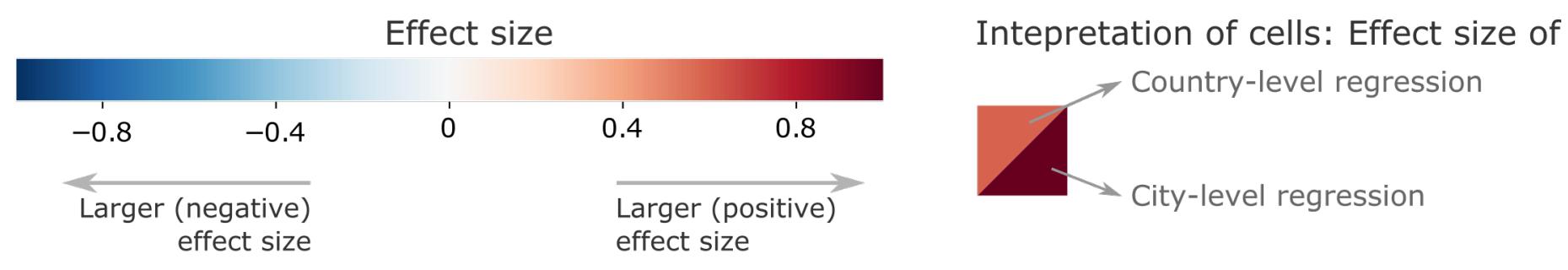
DNN Vocoder



Exposition: (A) tells you how to read the rest of the plot

Conflict: Between the four methods

Resolution: **DNN Vocoder** much better



Exposition: Top-left tells you how to read the rest of the plot

Conflict: Between economic, financial, managerial and offshore

Resolution: Financial and managerial activities overrepresented

PRACTICE: TELL A STORY

- ▶ What is the exposition, conflict and resolution in your graph?
- ▶ How will your graph be perceived by the audience? What moods and emotions might users experience as they engage with your work? (not frustrated!)
- ▶ How can you guide them in the interpretation?