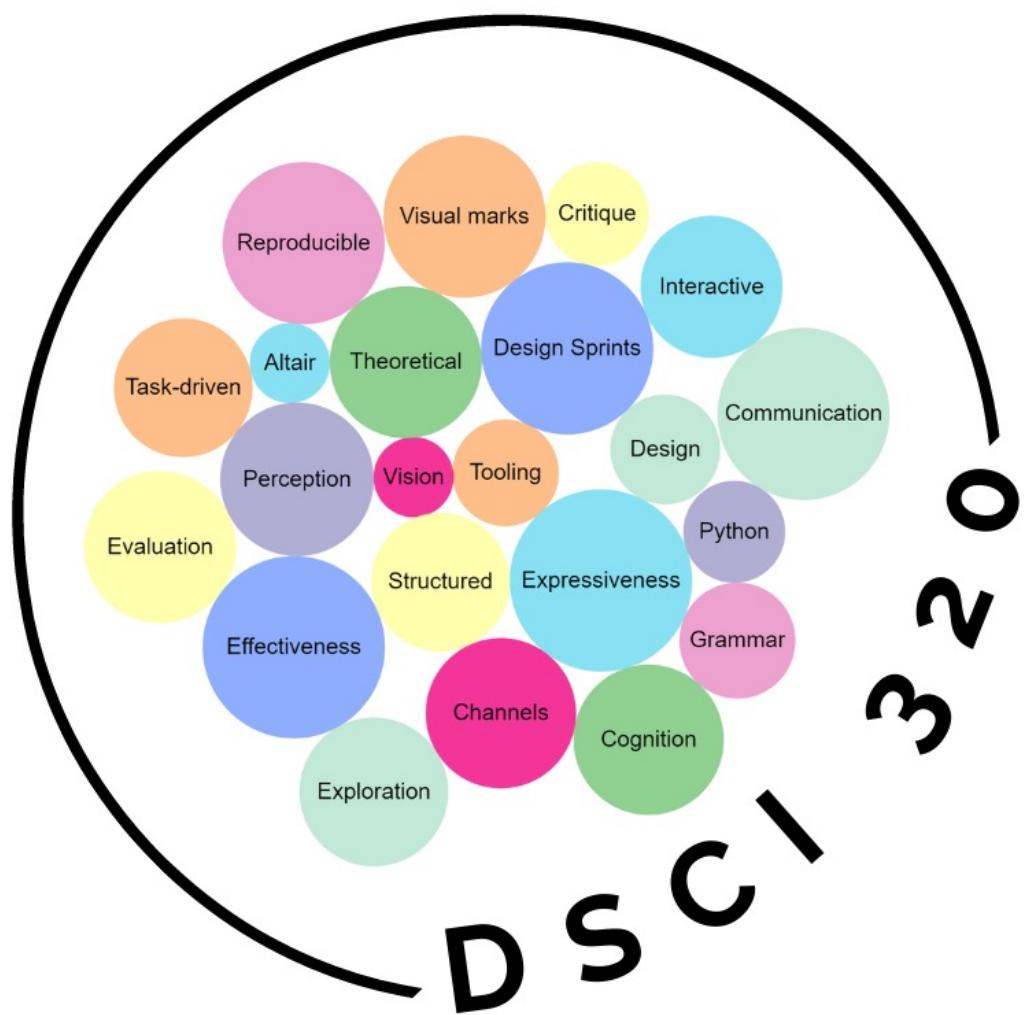


# Visualization for Data Science

## And it so begins ...

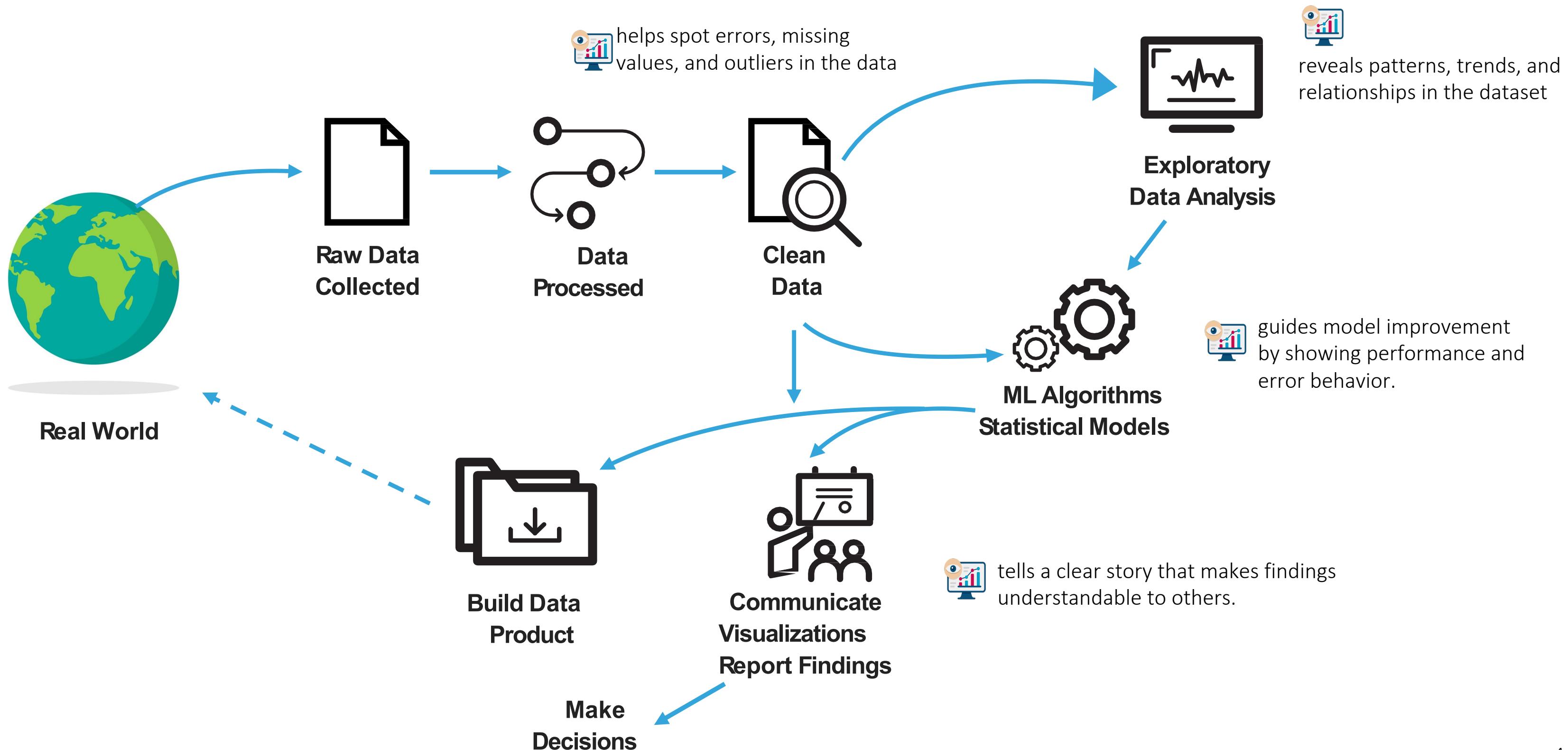


# Office Hours Next Week

- – ALL ONLINE. See Ed and Canvas for Zoom Links
- Sunday
  - 10am – 12pm
  - 2 – 4pm
- Monday
  - 1pm – 3pm
- Tuesday
  - 2 – 4pm

By the end of this course, my goal is for you to have a toolkit to design, create, critique, and interpret interactive visualizations.

# Role of Visualization in Data Science?



## Week 2 – 4: Data Abstraction, Vis. Grammar, Viz Deconstruction

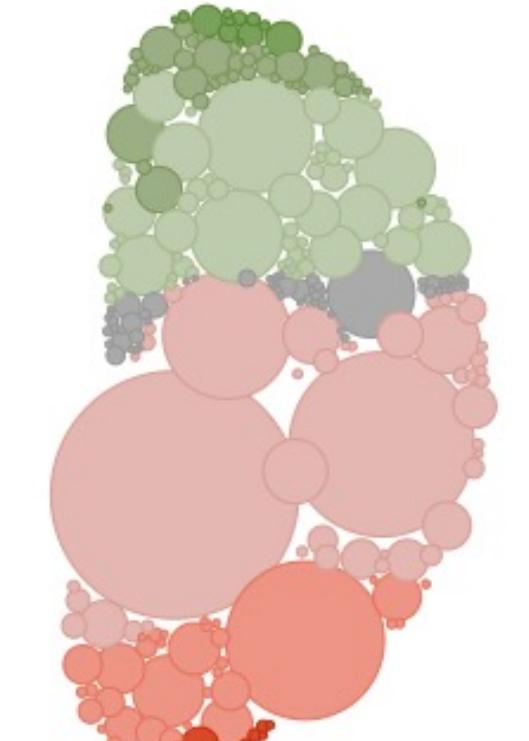
Develop a strong theoretical foundation for understanding how data characteristics align with visual encoding choices—learning to analyze, deconstruct, and critique visualizations by connecting data types, marks, and channels to their perceptual properties and effectiveness.

**Mandatory****\$2.5 trillion**

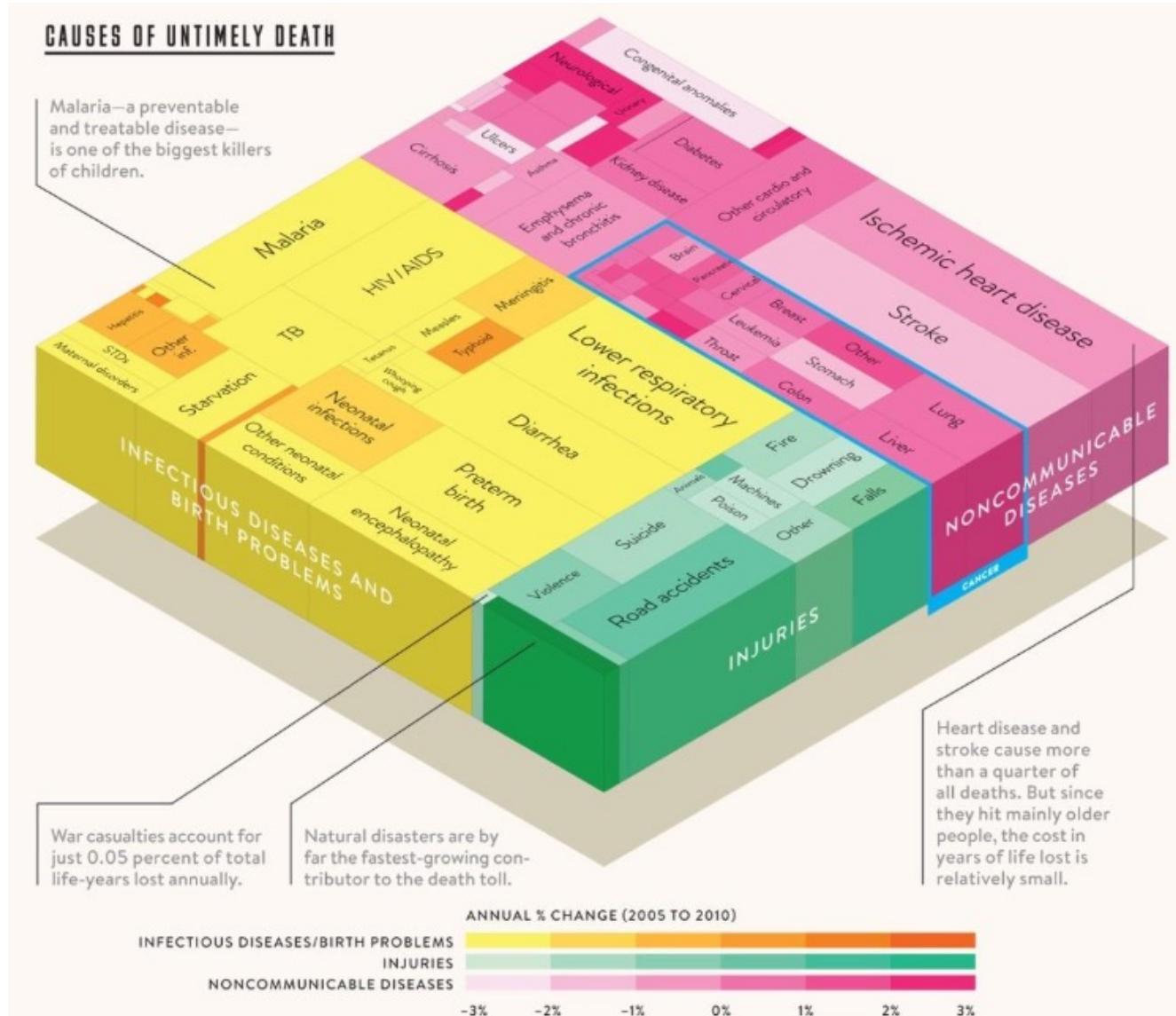
About 70 percent of budgetary spending is controlled by existing laws, including entitlements like Medicare, Medicaid and Social Security.

**Discretionary****\$1.1 trillion**

Only about 30 percent of the budget is controlled by the annual budget process. Last August, the White House and Congress agreed to a cap on this spending.



Describe the data and deconstruct the visualization based on the principles from Week 2 – 4 (data type, semantics, cardinality, mark, channel, channel characteristics)



Describe the data and deconstruct the visualization based on the principles from Week 2 – 4 (data type, semantics, cardinality, mark, channel, characteristics)

## Week 5 – 7: Exploratory Data Analysis

Build the ability to explore and diagnose datasets through visualization—identifying patterns, relationships, and data quality issues, and selecting the most appropriate visual forms for specific analytical tasks. This includes breaking down real-world questions into structured tasks and matching those tasks to effective exploratory visualizations.

- A \_\_\_\_\_ shows the probability density of data at different values, displaying the distribution as a smooth curve along a single axis.
- The width of a violin plot at any given point represents the \_\_\_\_\_ of observations at that value.
- What does Simpson's Paradox teach us about the importance of data stratification in exploratory data analysis?
- Explain what the missing matrix and UpsetPlot show, be able to deconstruct it.

## Week 8 – 10: Faceting & Interaction

Develop the ability to design, analyze, and critique complex visualizations by understanding how marks, channels, interaction techniques, and multi-view systems work together. Select appropriate visualization types, manage visual complexity, and apply interaction principles (including affordances and coordinated views) to enhance data exploration, communication, and usability.

# Interaction Exploration

Characterize and critique the viz based on the concepts you have been exposed to this term. Notice the focus is solely on interaction and NOT have the data is visualized.

1. [Life Expectancy](#)
2. [Global Oil Production and Consumption](#)
3. [Obama 2013 Spending](#)
4. [Homelessness – The Guardian](#)
5. [Gender Pay Gap](#)
6. [Demographics of Others](#)

Action event and semantics : click, hover, drag, etc

Reaction type – highlight, change over time (data, encoding, arrangement, alignment)

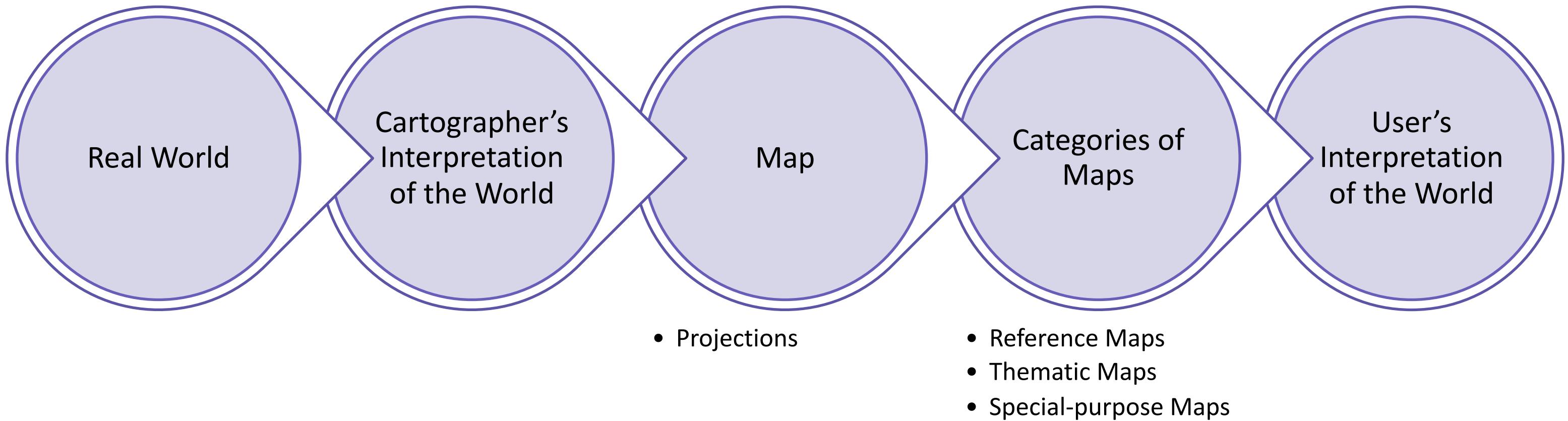
Views – juxtapose, partition, superimpose

Views data – share encoding, share data, share navigation

Interactivity – Action Elements (Focus, Presence)

Interactivity – Reaction Elements (Spread, Activation)

# Week 10B: Maps



A choropleth map shows unemployment rates by state using a sequential color scale from light yellow (low unemployment) to dark red (high unemployment). A viewer concludes that larger states (like Texas) have more unemployed people than smaller states (like Massachusetts) because they see more dark red area on the map. **Which of the following statements BEST describes this situation?**

- A. This is a Stage 2 error because the cartographer misunderstood the unemployment data
- B. This is a Stage 3 error because choropleth maps are inappropriate for displaying rates
- C. This is a Stage 4 error because the user is confusing visual area with the data attribute being encoded (rate, not total)
- D. This is a Stage 1 error because the underlying data was collected incorrectly
- E. This error occurs across multiple stages and cannot be attributed to any single stage

## Week 12 - 13

Develop a deep understanding of human perception and cognition—how we see, interpret, and sometimes misinterpret visual information—and apply this knowledge to create perceptually effective, accessible, and principled visualization designs. This includes leveraging Gestalt principles, color theory, and core design guidelines to analyze and improve visualizations.

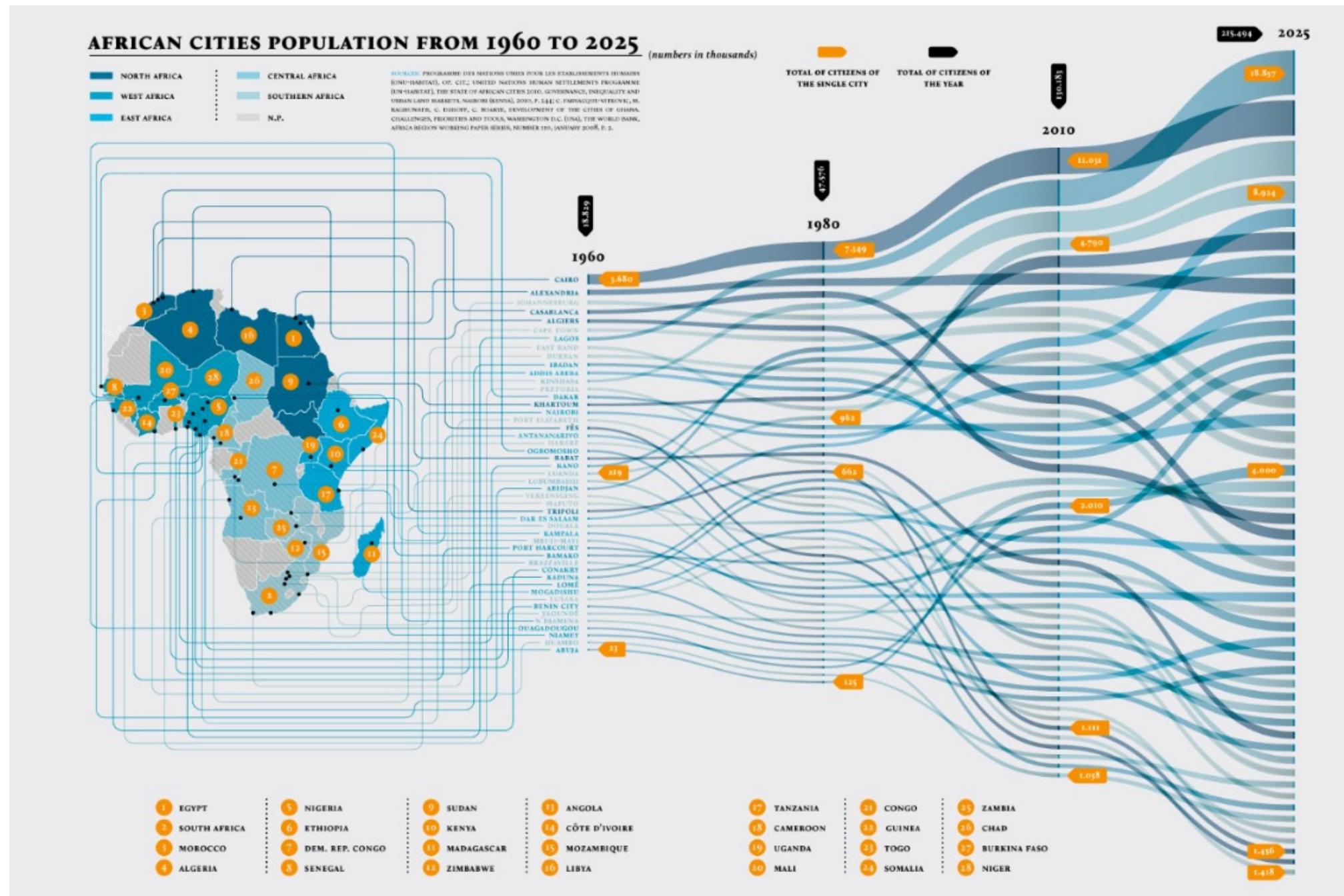
What Gestalt Principle is being exploited in the image below



## Select the statement that best describe the data-ink ratio principle

- A. The data-ink ratio advocates for the maximum use of decorative elements in a visualization to make the data more engaging.
- B. The data-ink ratio refers to the proportion of a visualization's ink dedicated to non-data elements, such as background images or decorative graphics.
- C. The data-ink ratio suggests that every mark on a visualization should convey new and useful information, and non-essential elements should be eliminated.
- D. The data-ink ratio is primarily concerned with the color choices in a visualization, ensuring a vibrant and diverse color palette.
- E. In data-ink ratio terms, it is essential to include as many labels and grid lines as possible to ensure the data is understood by the viewer.

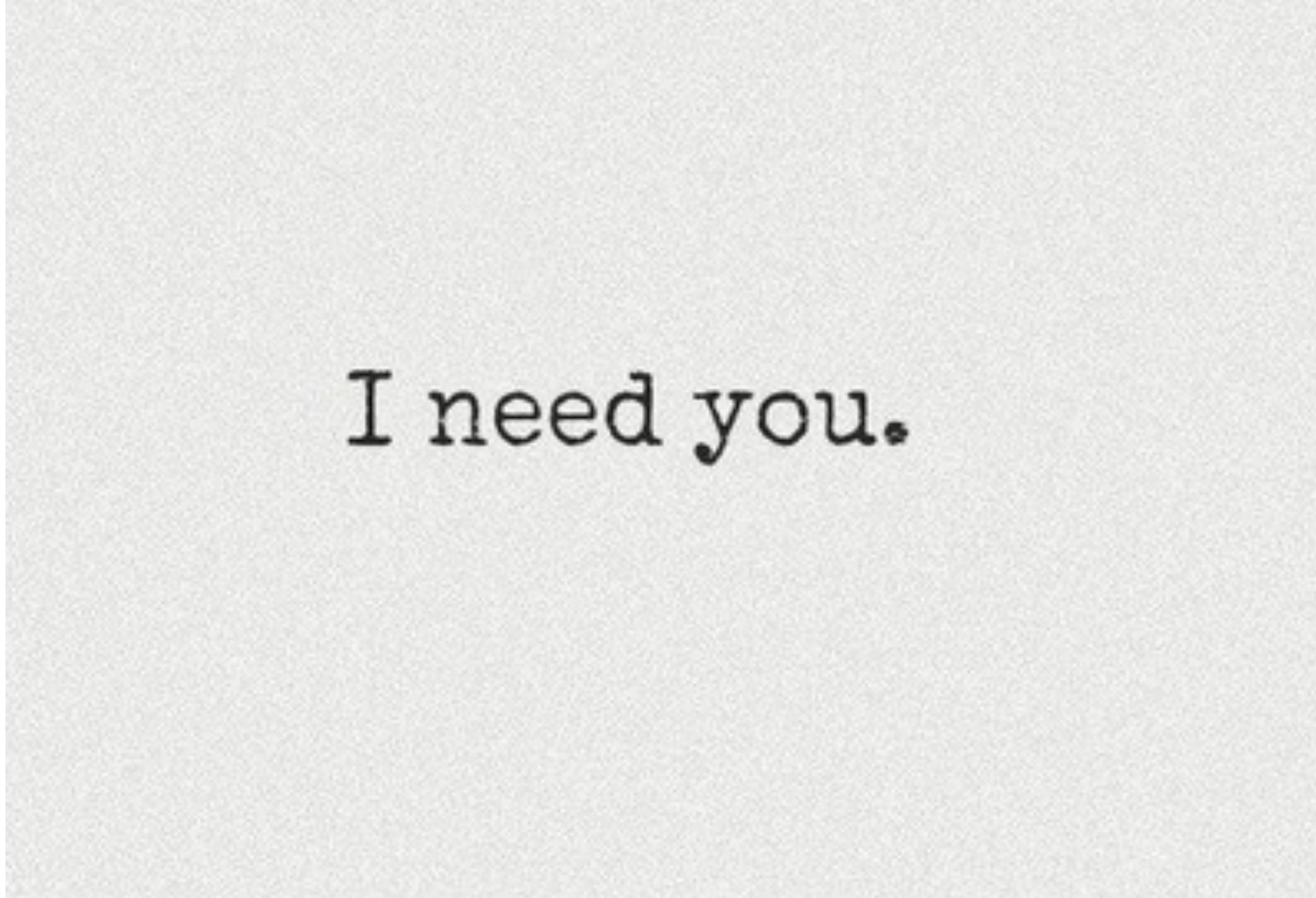
# For the visualization shown below, what principles does it violate?





KEEP  
CALM  
THE  
SEMESTER IS  
ABOUT TO END.

Send me an email if  
you would like to be  
a TA next year. We  
will have over 100  
students



I need you.

# *Your Feedback Student Questionnaires on Courses and Teaching*

your  Feedback  
matters!

The UNI one is on Canvas

# Course Learning Outcomes

- Justify the importance/value of visualizations for data analysis and presentation.
- Describe the foundational concepts that influence visualization design.
- Use the Vega-lite visualization grammar to categorize, describe, design, and create existing and novel visualizations.
- Engage in exploratory data analysis (EDA) and justify EDA's role in data science.
- Use Python to create reproducible visualizations.
- Select and tailor the appropriate plot for the given dataset, audience, and data analytical question at hand.
- Compare methods for visualizing data and explain the tradeoffs of different approaches.
- Interpret visualizations and be able to explain how data is encoded.
- Evaluate visualizations based on principles of effective design.
- Using a structured approach, design and implement effective visualizations for exploration and communication.

# What & How

- Posted on Website. Let's go there now
- <https://pages.github.ubc.ca/kemiola/DSCI320-2025W1/weeks/week15.html>

# How to Study

## Programming

- If you have not done so before make sure you review the tutorials as well.
- Review each lab and assignment
  - Focus on common mistakes that encounter
  - Focus on structure and not on memorizing and scribing content
    - Basic structure of an altair Chart
    - Specify properties for different channels
    - Define interactions
    - Attach interactions
    - Use facets etc
  - Make sure that for the common vizzes you know how to create in altair (what mark and channel should be used)
- Difficulty – as difficult as what you have encountered on quizzes

PRACTICE  
ISN'T THE THING YOU DO  
ONCE YOU'RE GOOD.  
IT'S THE THING YOU DO  
THAT MAKES YOU GOOD.

*Malcolm Gladwell*

# How to Study

## Design and Theory

- Review lecture slides
- Create your own questions using the learning outcomes.
- Discuss with your class mates your responses on different graded work in order to understand different points of view
- Create an extensive rubric, take any graded quiz or lecture activity, assignment and create an exhaustive rubric on how to describe and critique a visualization
  - E.g. Marks and Channels, Channel characteristics, Accuracy, Perception, Color, Interaction
- Use this created rubric to critique and describe various representations.
- Remember the viz I will give you is something we have NOT seen in class, so memorizing a critique is not helpful.

# Credits and Acknowledgements

- UW Interactive Data Lab @ University of Washington
  - Enrico Bertini @ NYU
  - Cody Dunne @ Northeastern
  - Hanspeter Pfister @ Harvard
  - Alexander Lex @ University of Utah
- At UBC
- Joel Ostblom
  - Firas Moosvi
  - MDS Teaching Team
  - Tamara Munzner

Data overload

discriminability

proximity

External cognition

aggregation

Perceptual ordering

effectiveness

Direct manipulation

affordances

grouping

choropleth

pareidolia

accuracy

Gestalt principles

Oluwakemi

Action -> target

juxtapose

Change alignment

Adefunmike

enclosure

Data abstraction

Oluwafifehan

Color palettes

items

Ayodele

Integral dimensions

Oluwafunmilayo

Visual idiom

dataset

Olamudzengi

Ames room

Thematic maps

Sensory symbols

accuracy

Reaction -> flow

Separability

cardinality

Scale distortion

Color hue

marks

expressiveness

emergence