WEEK 11: DEEP LEARNING AND INTRODUCTION TO INFOVIS

Dr. Kai Li
School of Information Sciences
University of Tennessee, Knoxville
Spring 2025

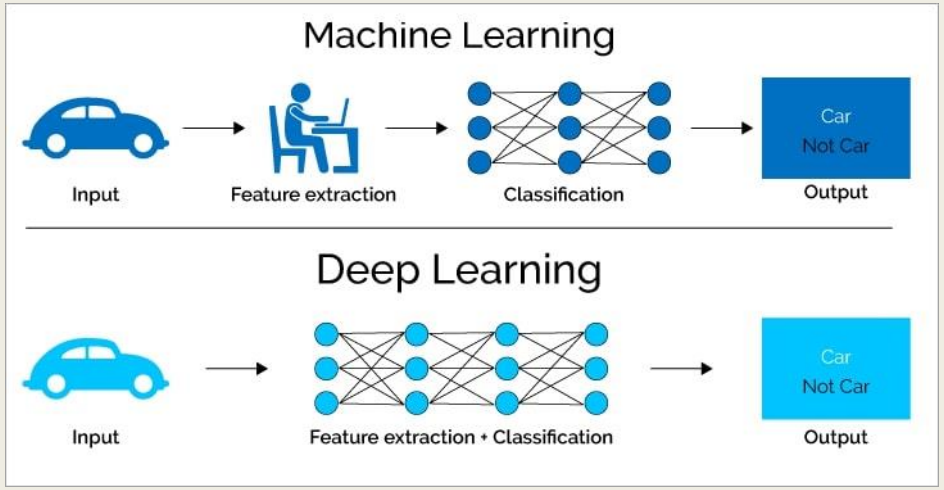
Review / Overview

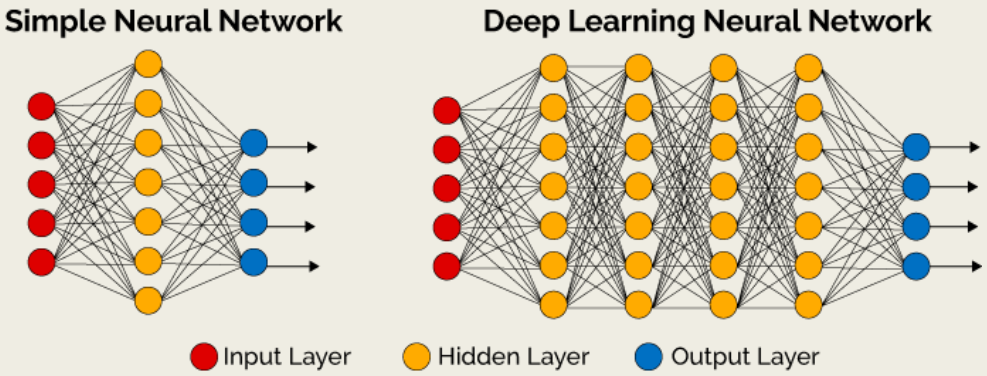
- Last week:
  - *Clustering*
- This week:
  - *(Very briefly) Deep learning*
  - *Intro to data/information visualization*

- Any questions about the previous weeks or assignments?

Deep learning vs. machine learning

- Deep learning uses neural network, a technology to simulate how human brain works.
- This technology creates hidden layers in the process to transform the data and achieve the best performance.
  - *The hidden layers are NOT totally invisible to us in many models, but they are still black boxes because how they are decided is unknown.*



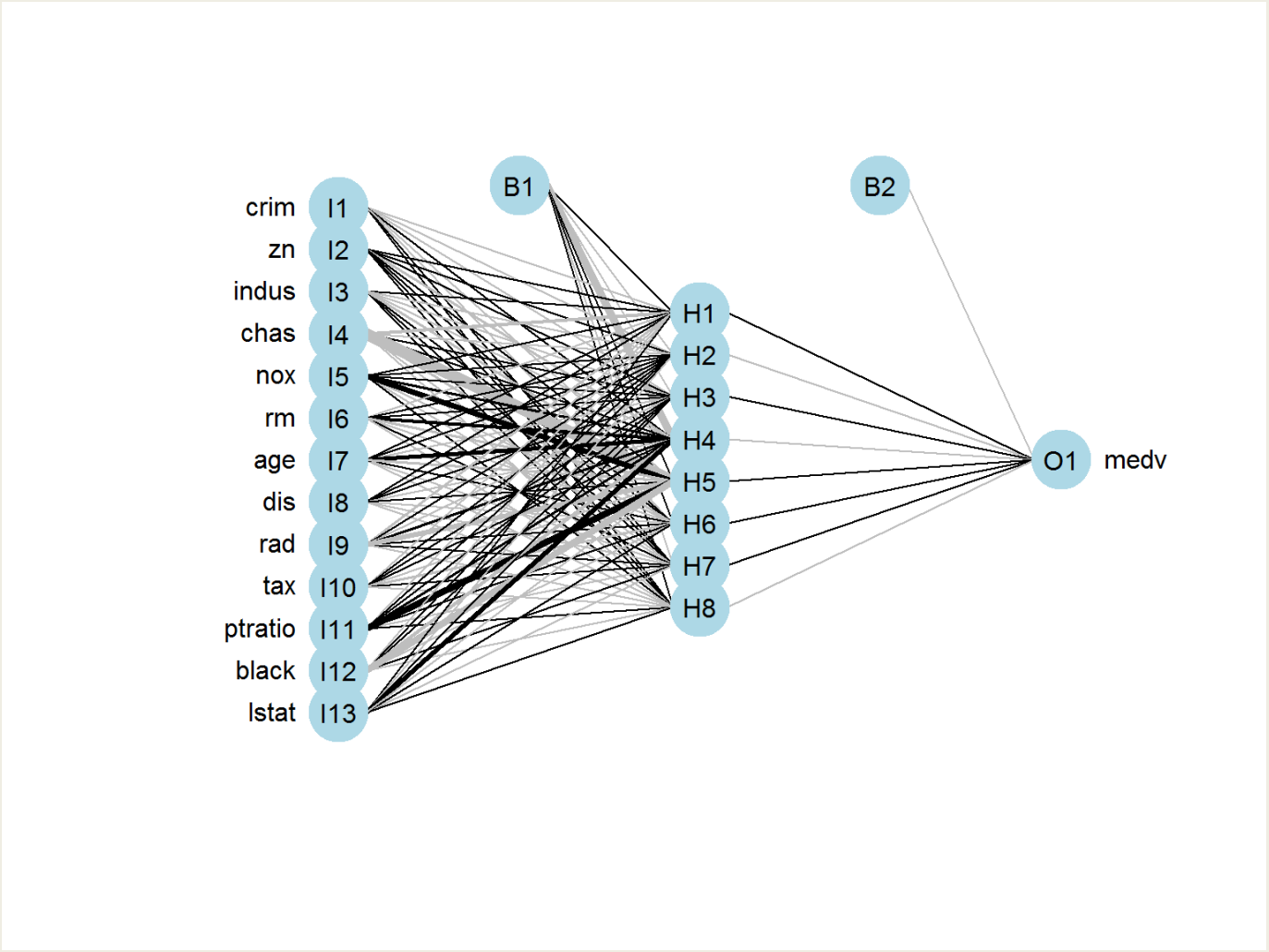


Artificial neural networks (ANN)

- ANN is a foundational deep learning model to implement the idea of deep learning.
- It is the foundation to many other more advanced DL models, such as CNN and RNN.
- It is also able to support major statistical models, such as regression and classification.

Demonstration

- Instead of directly using IVs to predict outcome, the model creates various hidden nodes (H1-8).
- B nodes stand for bias in each step.

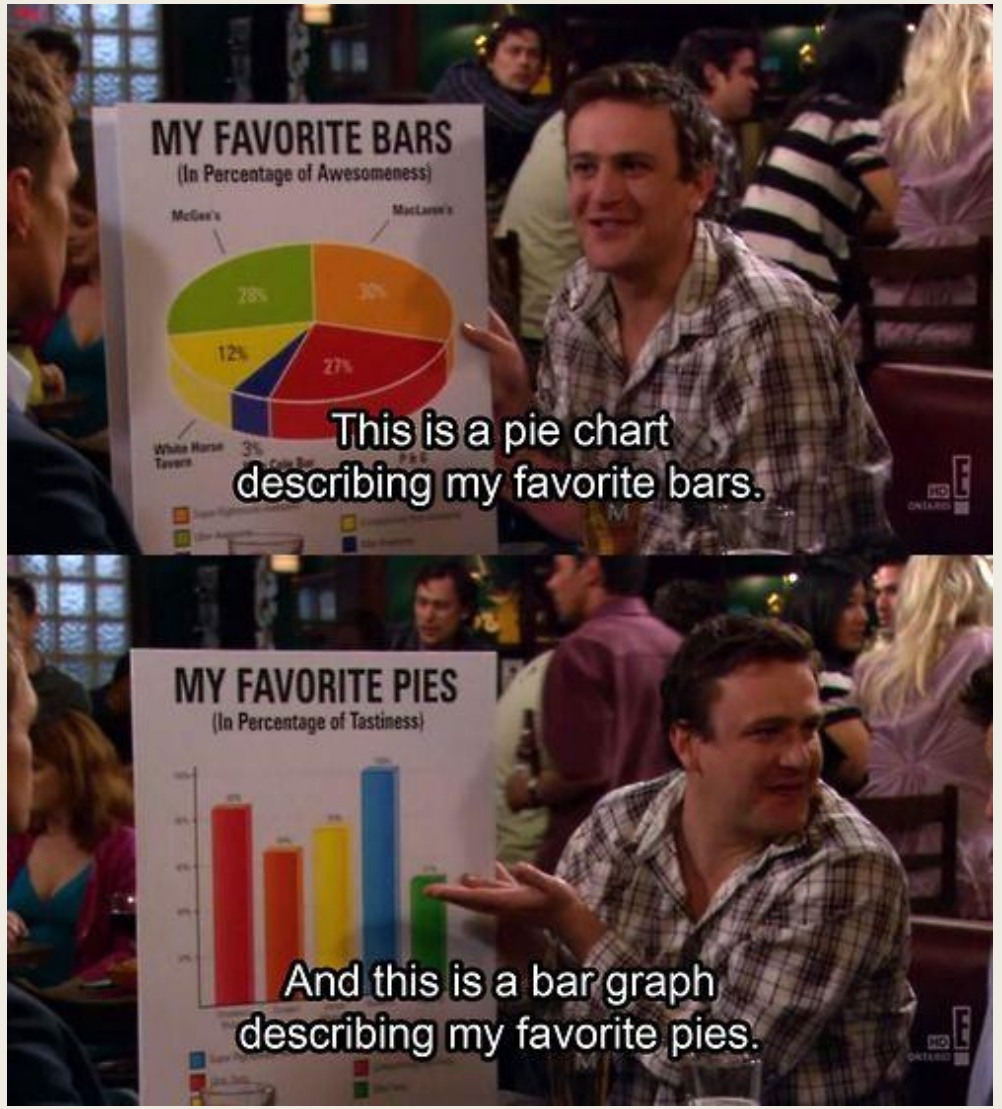


A summary

- Deep learning is not well supported by R. Python would be much more useful (faster and with more models and options)!
- For ANN@R, the implementation is not really that different from other statistical models.
- We can consider using deep learning if we are primarily focused on prediction rather than inference.

Information visualization

- Please consider taking our new 590 class (Information visualization) in the Fall semester if you are interested in this topic.



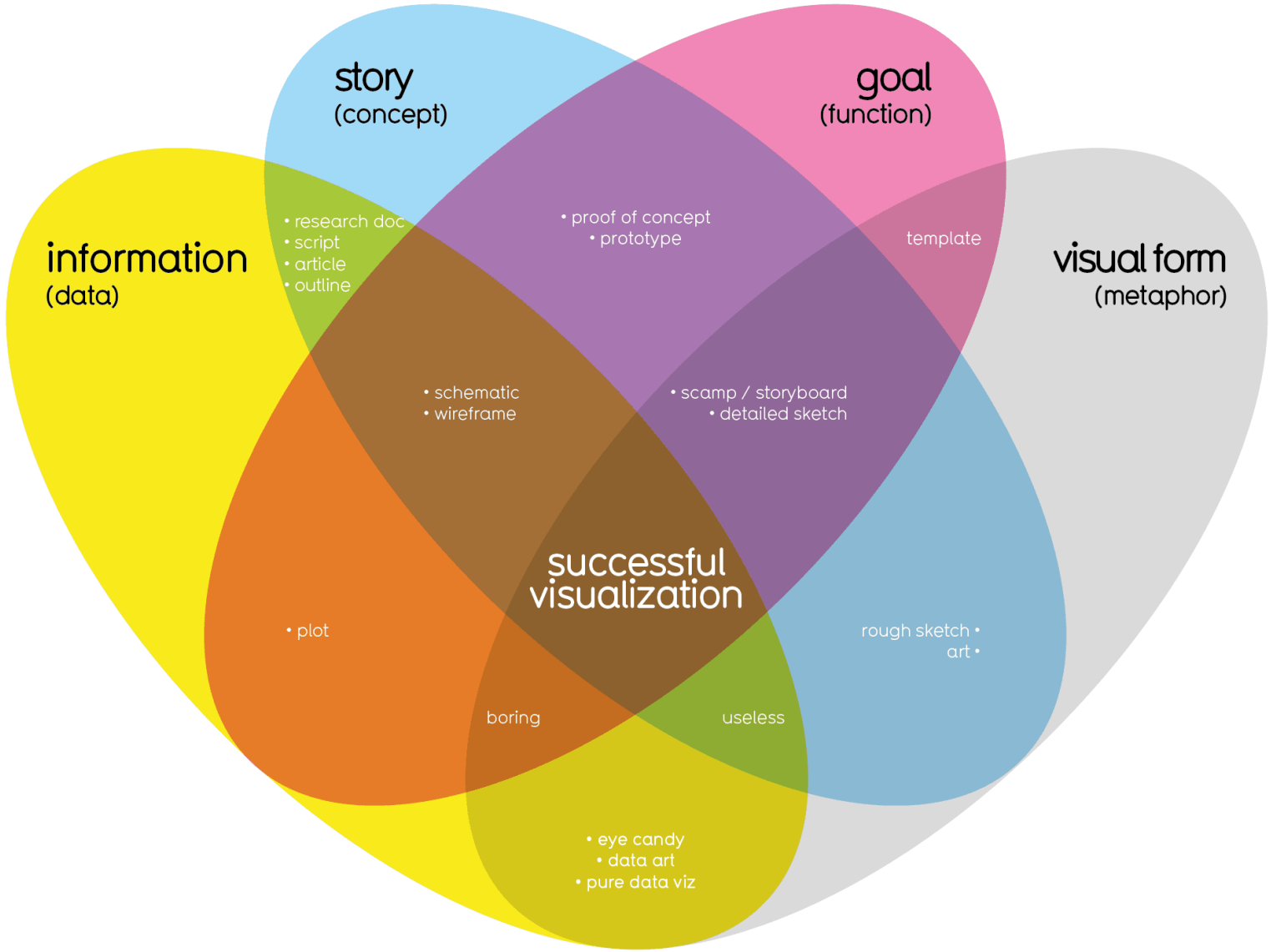
What is data visualization?

- In InfoVis, data is transformed to intuitive and meaningful graphical representations.
  - *The transformation is a creative process in which designers translate data into geometric shapes and give meanings to it.*
  - *Data visualization is far beyond just a technical topic: we need to consider data, users, design, and storytelling.*
- The ultimate criteria is whether a graph conveys the intended message to the viewers effectively.

rollover for more detail

What Makes a Good Visualization?

explicit (implicit)



David McCandless
InformationisBeautiful.net

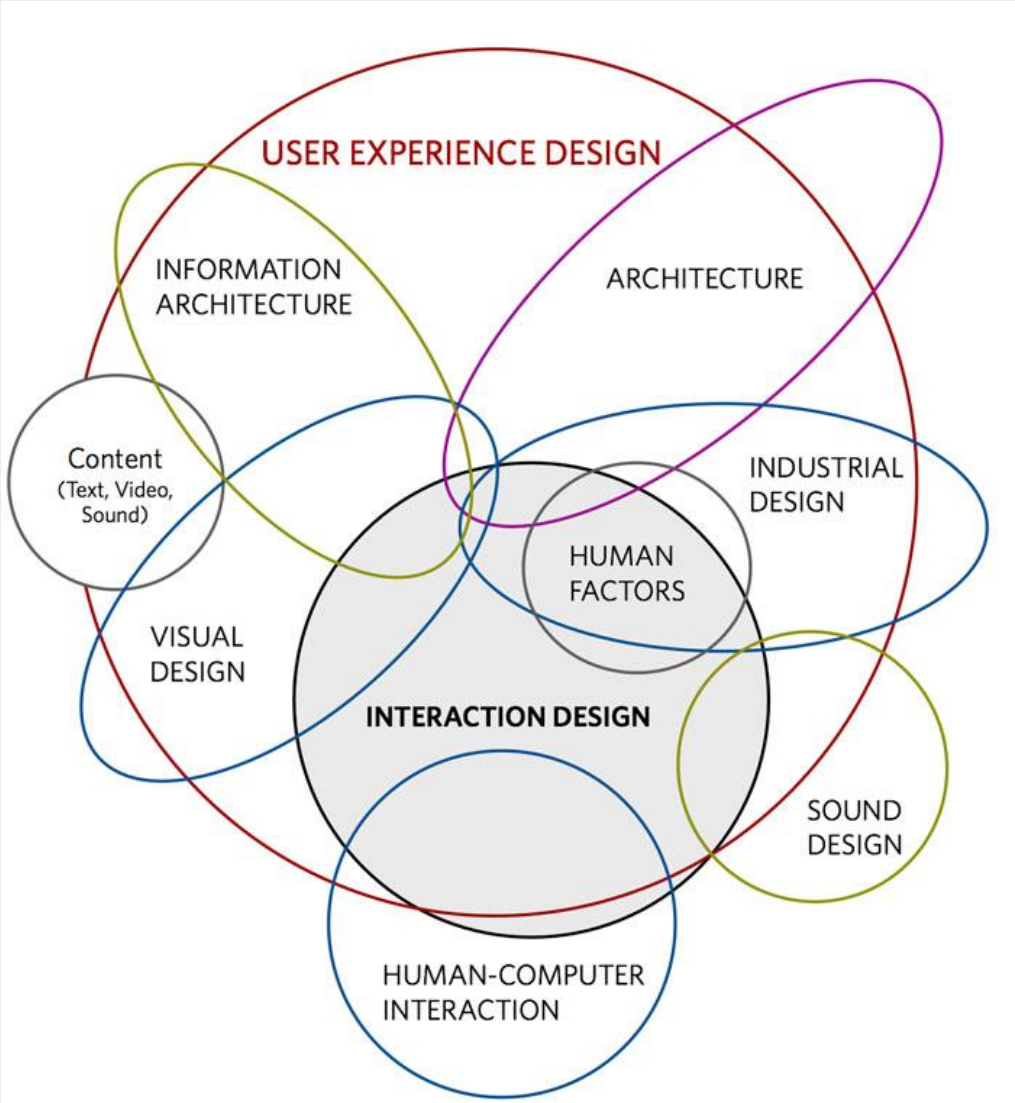
taken from new book
Knowledge is Beautiful

find out more
bit.ly/KIB_Books

https://informationisbeautiful.net/visualizations/what-makes-a-good-data-visualization/

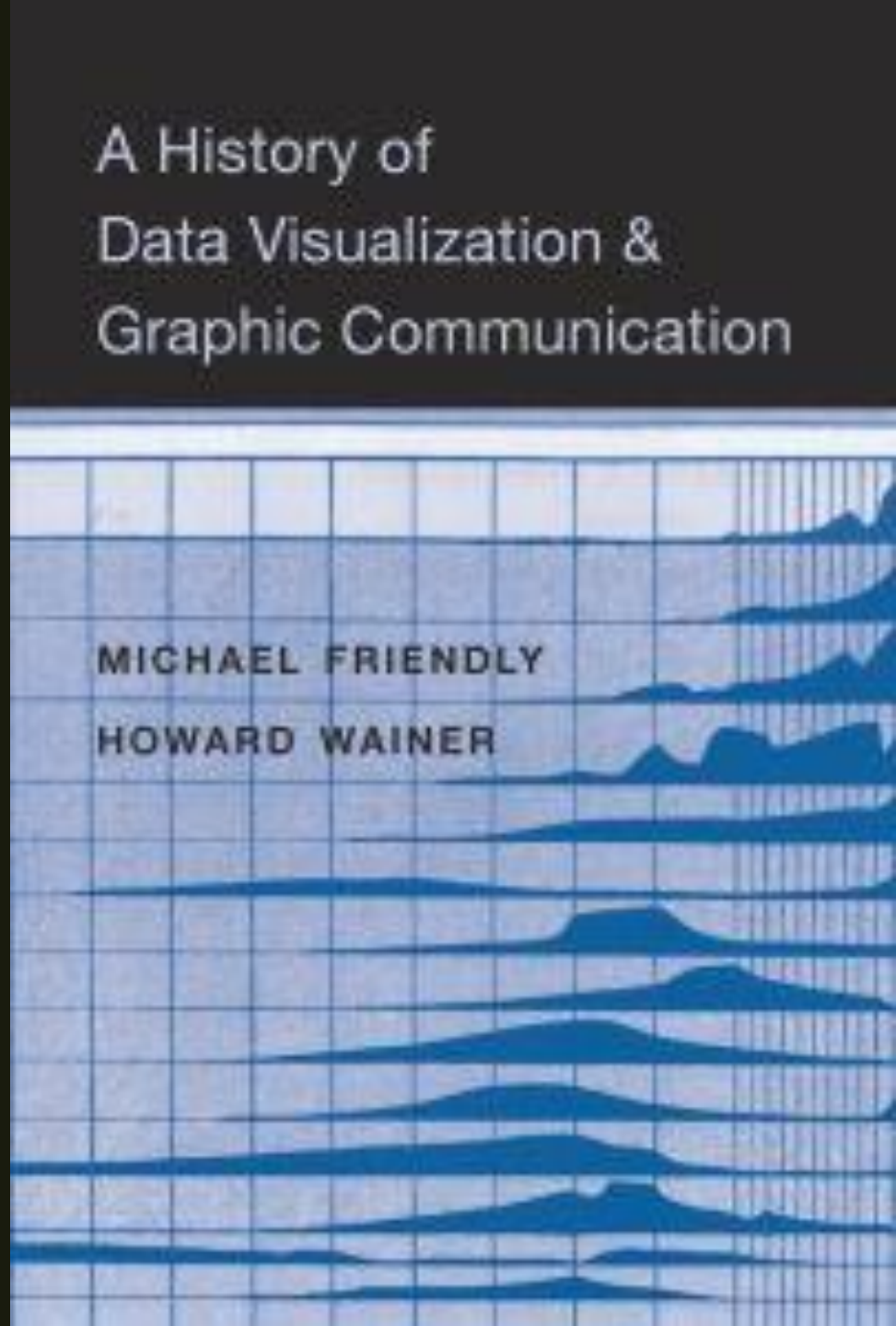
Story in data visualization

- The basic requirement: a good visualization is one that can effectively convey the message to its readers.
- Less is more!



History of data visualization

- The history is strongly paralleled with that of statistics.
- https://www.datavis.ca/



Type of visualization

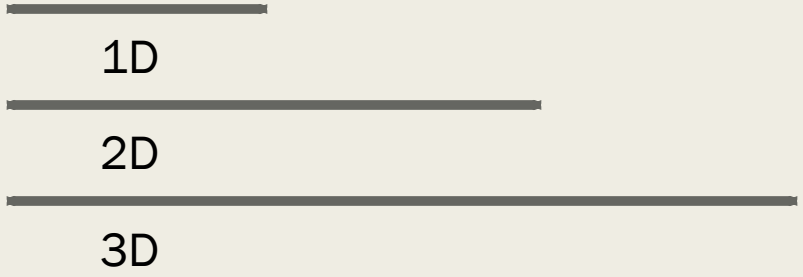
- Each type of visualization (1) can support different types of data values and stories and (2) will create different outcomes.
- For example, we can have visualization types based on the minimum number of categories required by the graph:
  - *1-D: pie chart, histogram…*
  - *2-D: bar chart, line chart, scatter plot… (using both x- and y-axis)*
  - *3-D:*
- The selection of the type of visualization to use is a critical first step.

1- to 3-dimensional data and visualization

ID	Name	Width	Height	Weight	Value
1	ABC	100	100	103	100
2	DEF	105	95	97	90

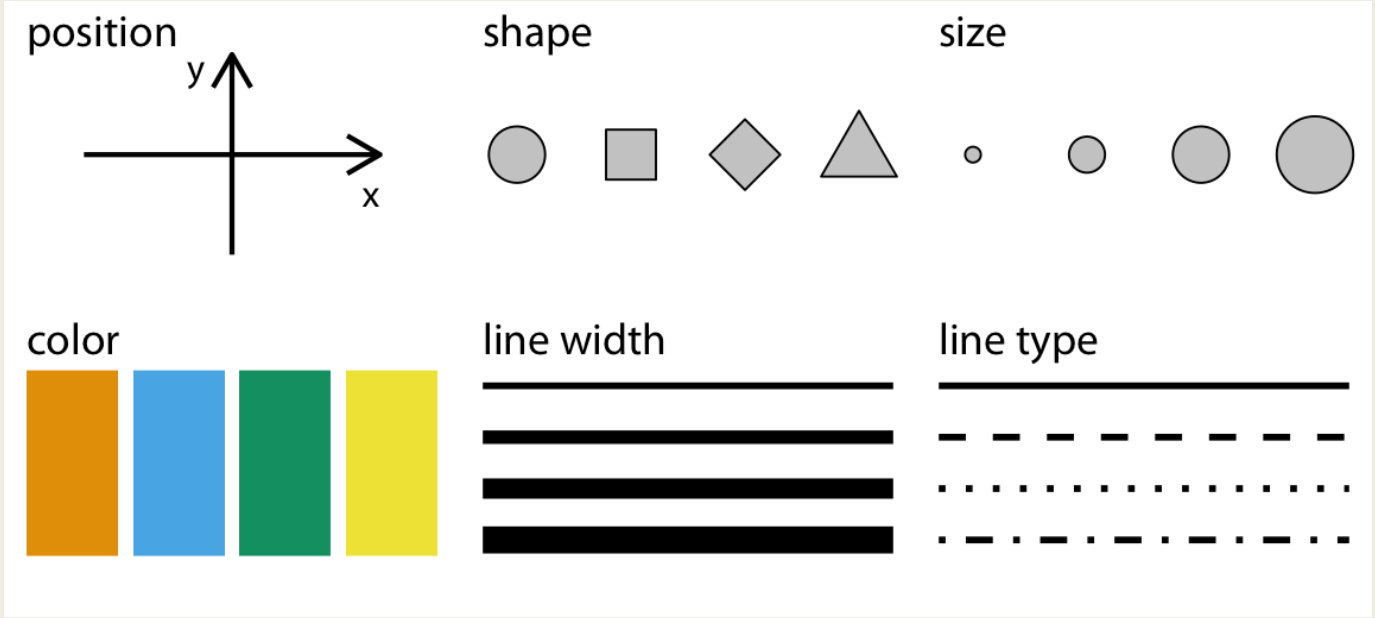
Most of the popular visualization types uses x- and y-axes and hence are 2-d graph.

That said, we can include other variables to the graph to add more information!



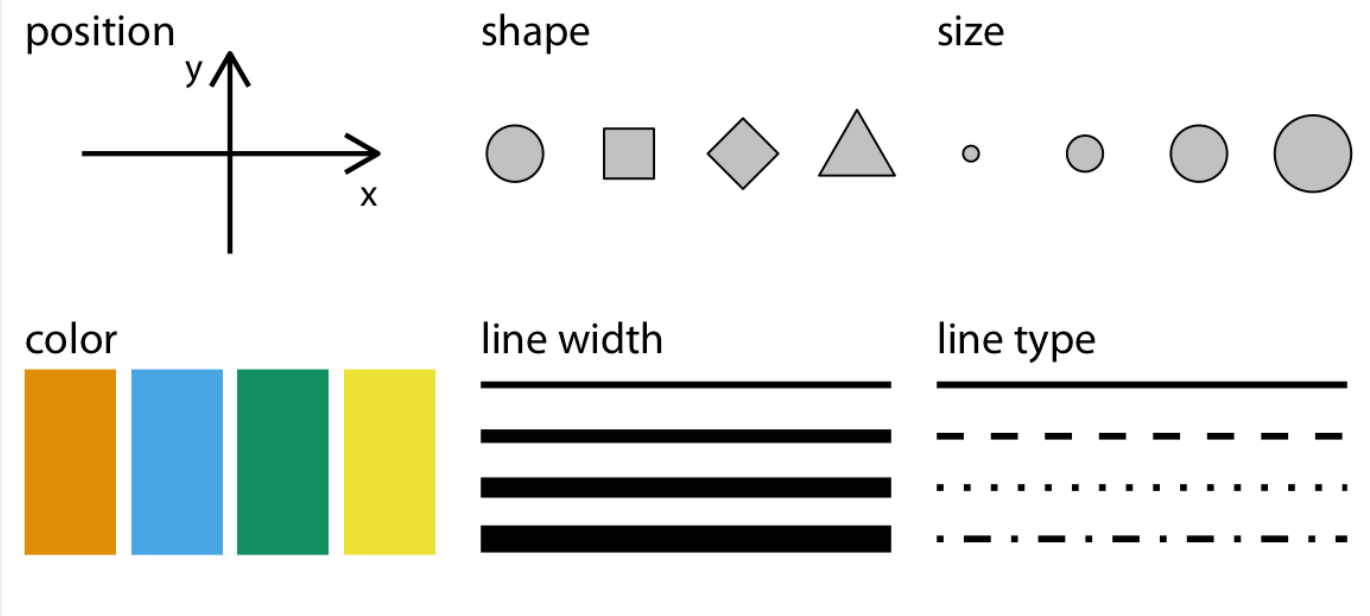
Visual features in data visualization

- Does a feature support the following type of data?
  - *Numeric*
  - *Categorical*
  - *Time*
- What is the best type of data supported by each feature?

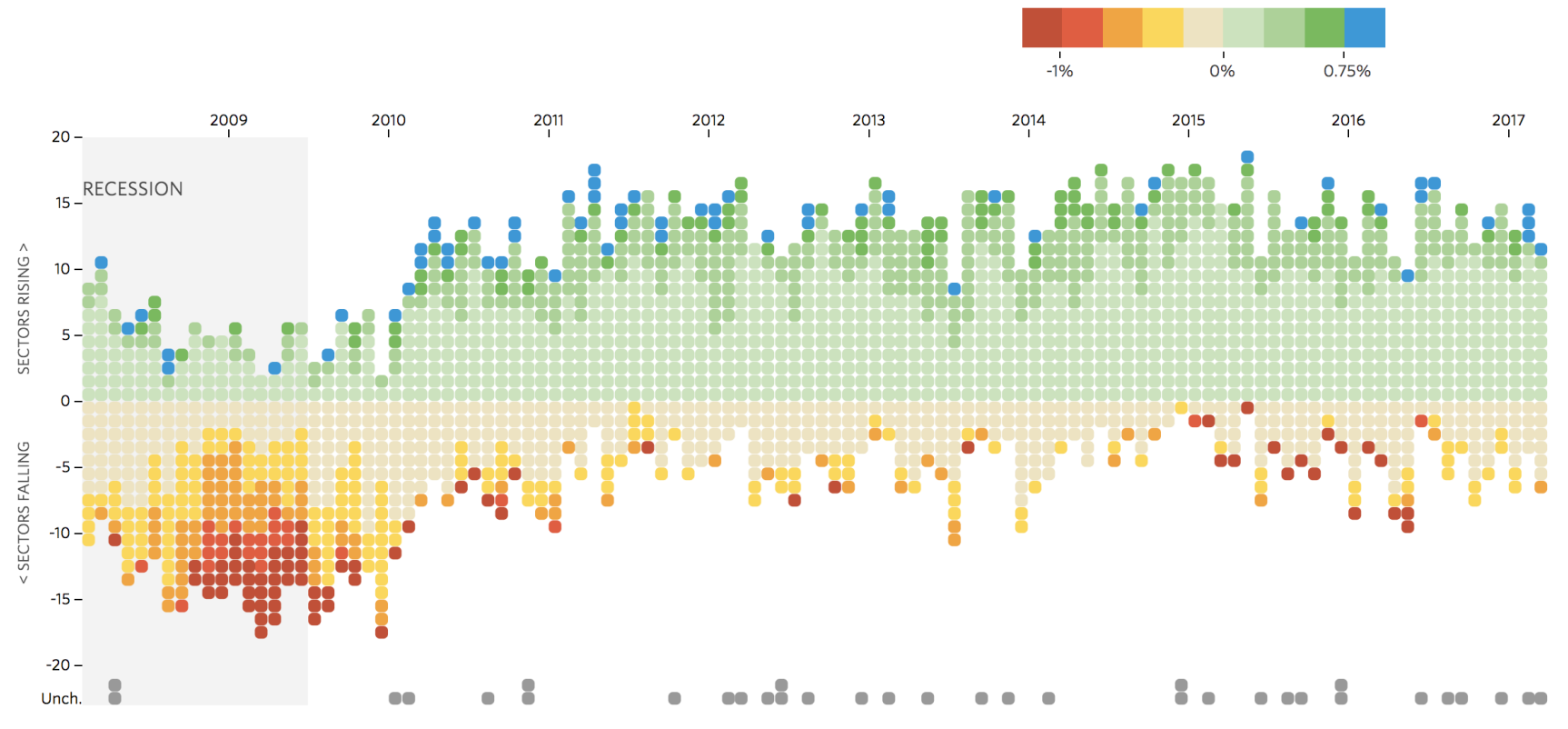


Visual cues in data visualization

- How well can we use each of the features to map:
  - *Numeric*
  - *Categorical*
  - *Time*
- Numeric values are ideally only used in x- and y-axis, if possible!
  - *Or, we can categorize the values.*



How many visual features are used in this graph? And what data type does each pattern represent?

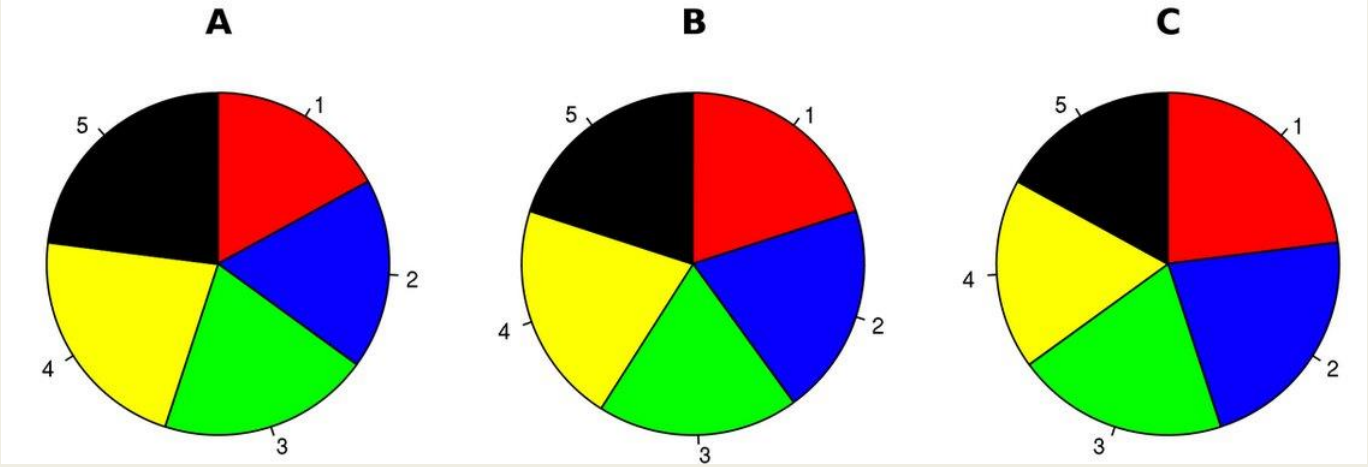


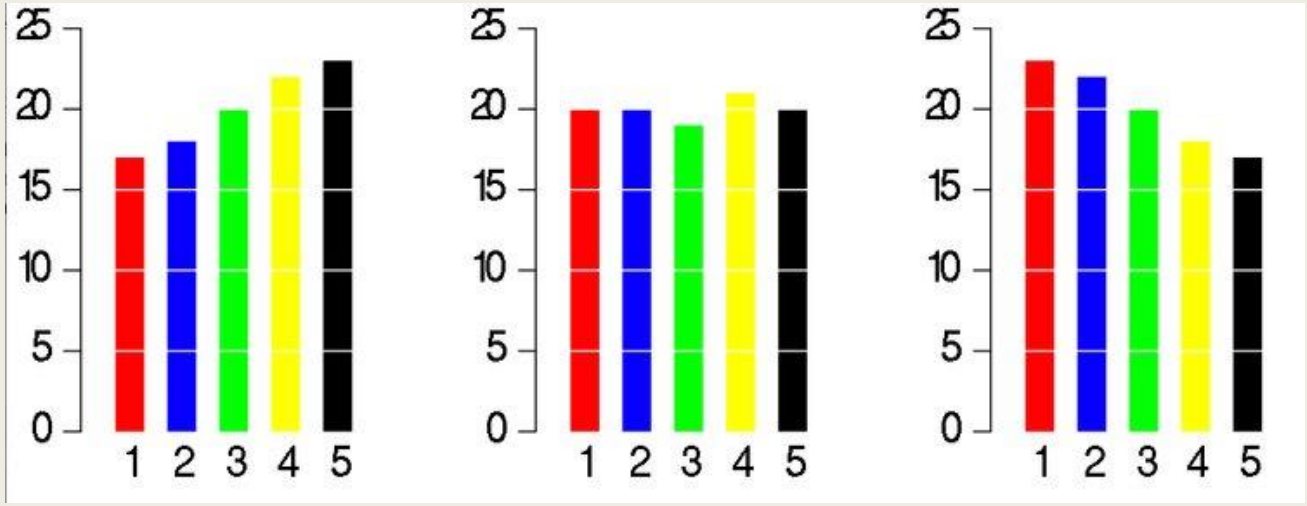
http://graphics.wsj.com/job-market-tracker/

Analysis of the previous graph

- X-axis: Year (Datetime)
- Y-axis: Rising or falling (Categorical)
- Color: The extent to which one market changes in a given year (Categorical) *
- Order of the y-axis: The order based on the amount of change of a market in a given year (Ordinal)

For example: which graph tells the story better?

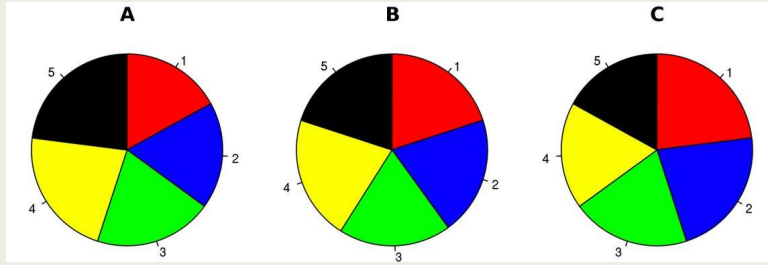


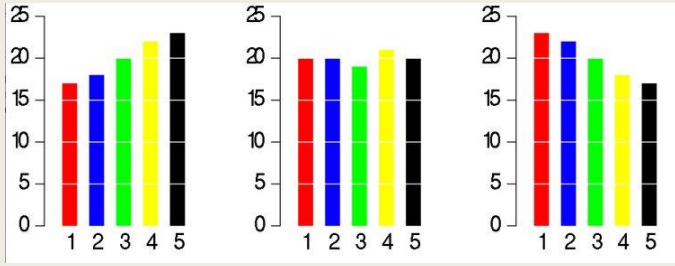


Problem with pie chart

- We can much more effectively and accurately observe the differences along the two axes (x- and y-) than those in the angle or area.
  - *If there is a key value we want to compare, we should present it in one of the two axes.*
  - *Table below shows the ranking of effectiveness of "pre-attentive features":*

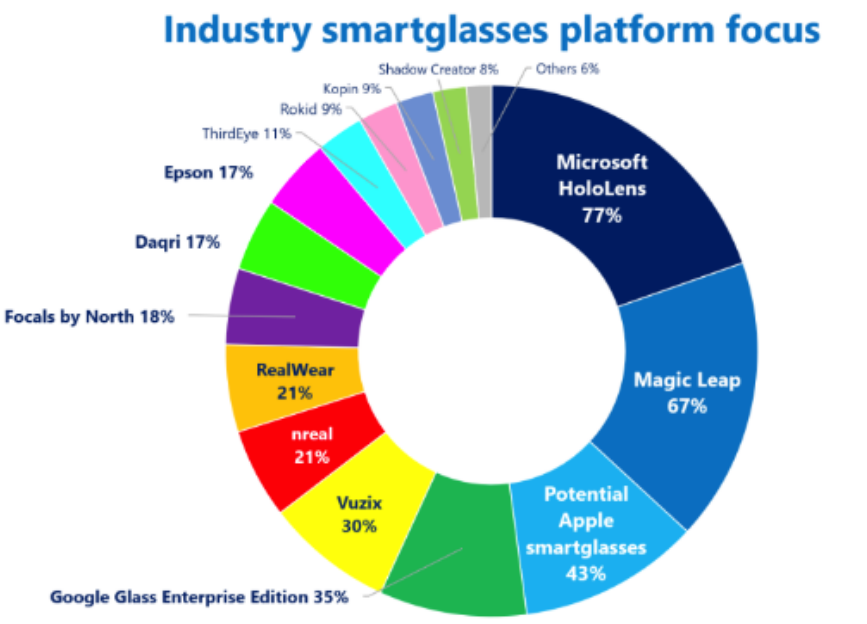
1. Position along a common scale
2. Position on identical but nonaligned scales (e.g., small multiples)
3. Length
4. Angle, Slope
5. Area
6. Volume, Density, Color Saturation
7. Color hue





Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531-554.

Pie Chart: Pitfalls

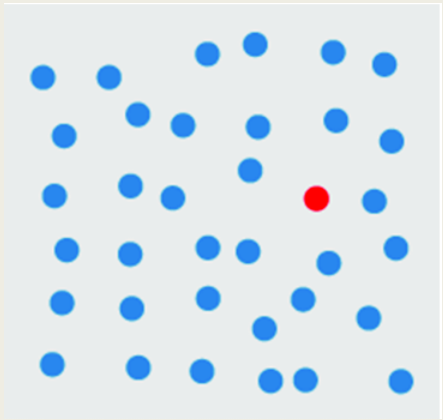


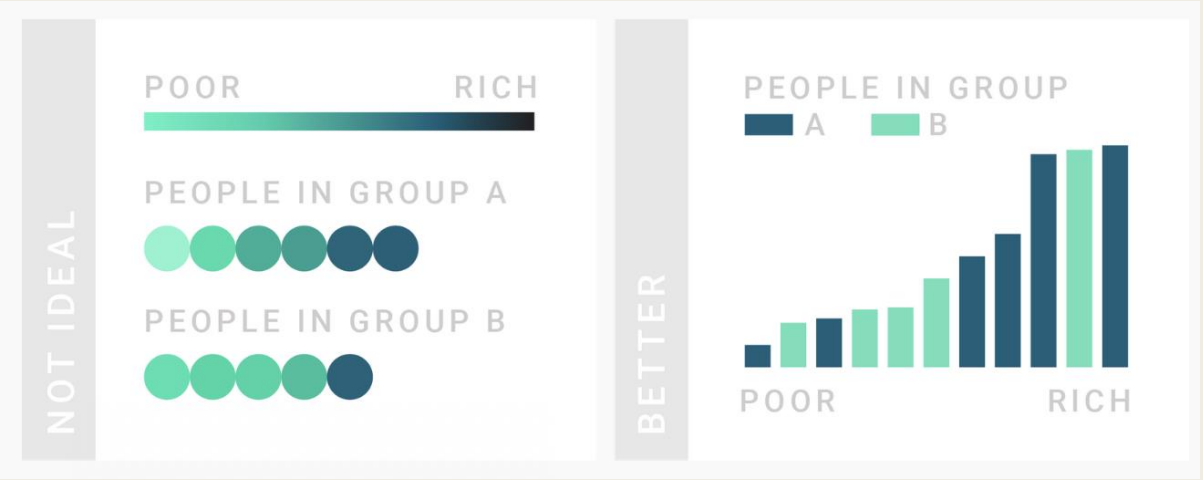
https://viz.wtf/

- Pie chart (or donut chart) is based on three assumptions:
  1. *Categories should be mutually-exclusive;*
  2. *All categories of the population are ideally included;*
  3. *Ratios of all categories should be added up to 100%.*
- If any assumption is violated, the data is better visualized in bar chart.
- Not to say that it is less effective!

Cognitive science behind visualization

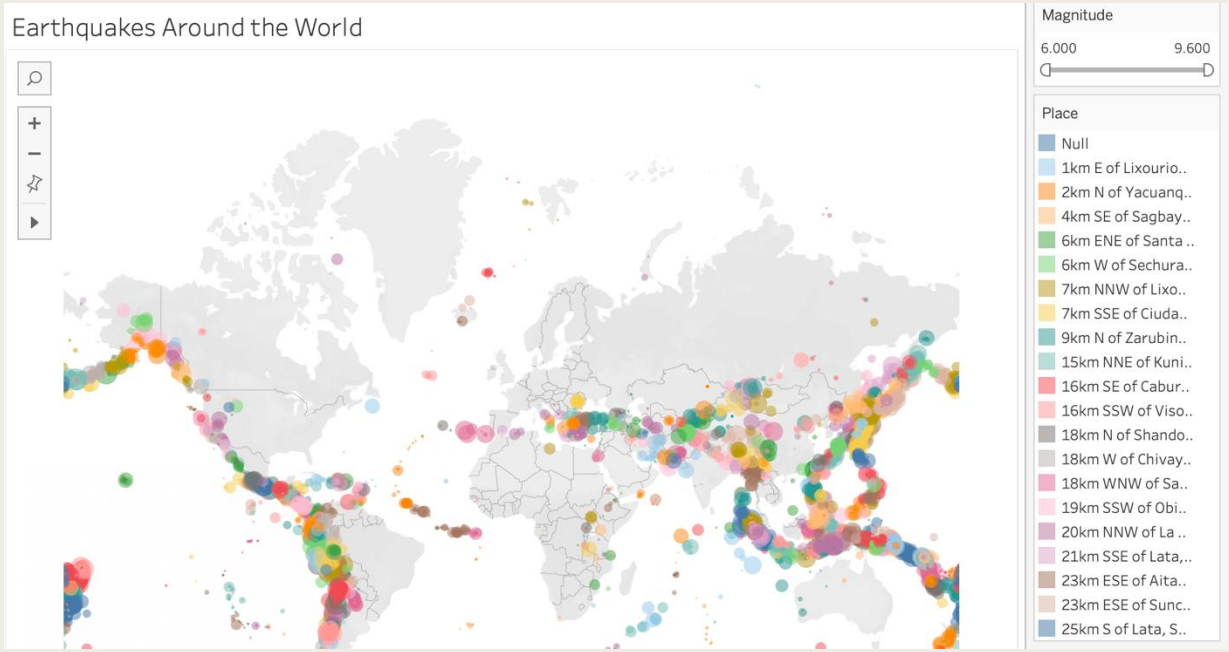
- Many important visualization principles are supported by cognitive science.
- For example, even though color is a very effective pre-attentive feature, it is much more useful to distinguish categories than comparing numbers.

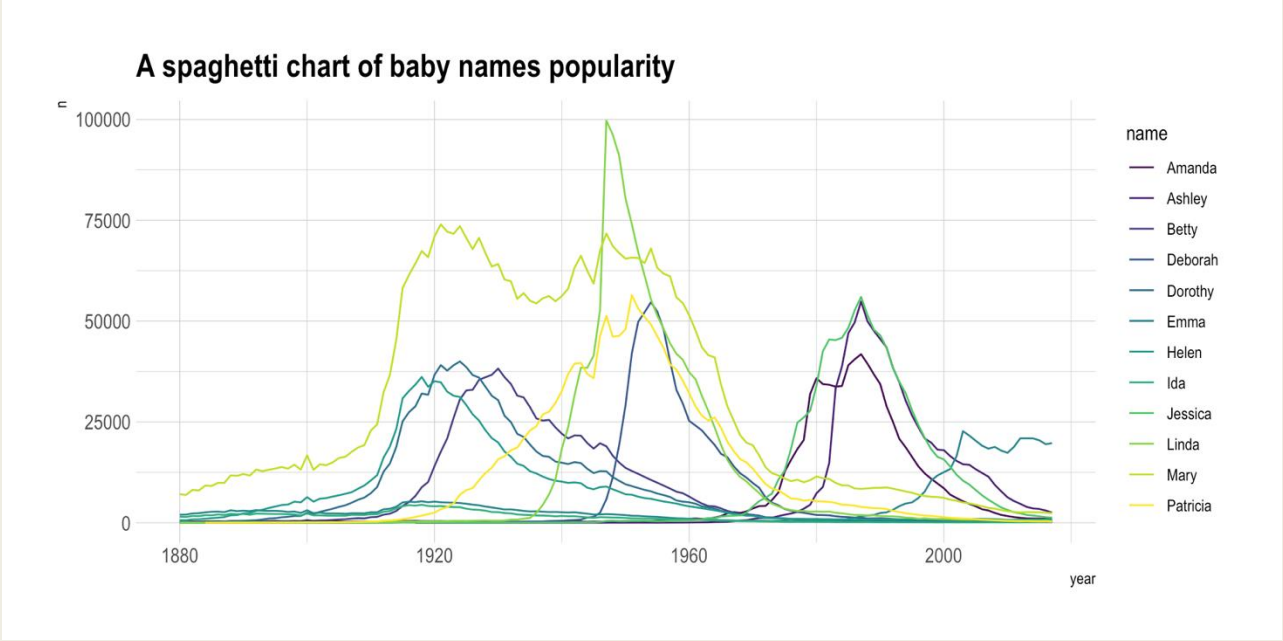




Magical number of seven

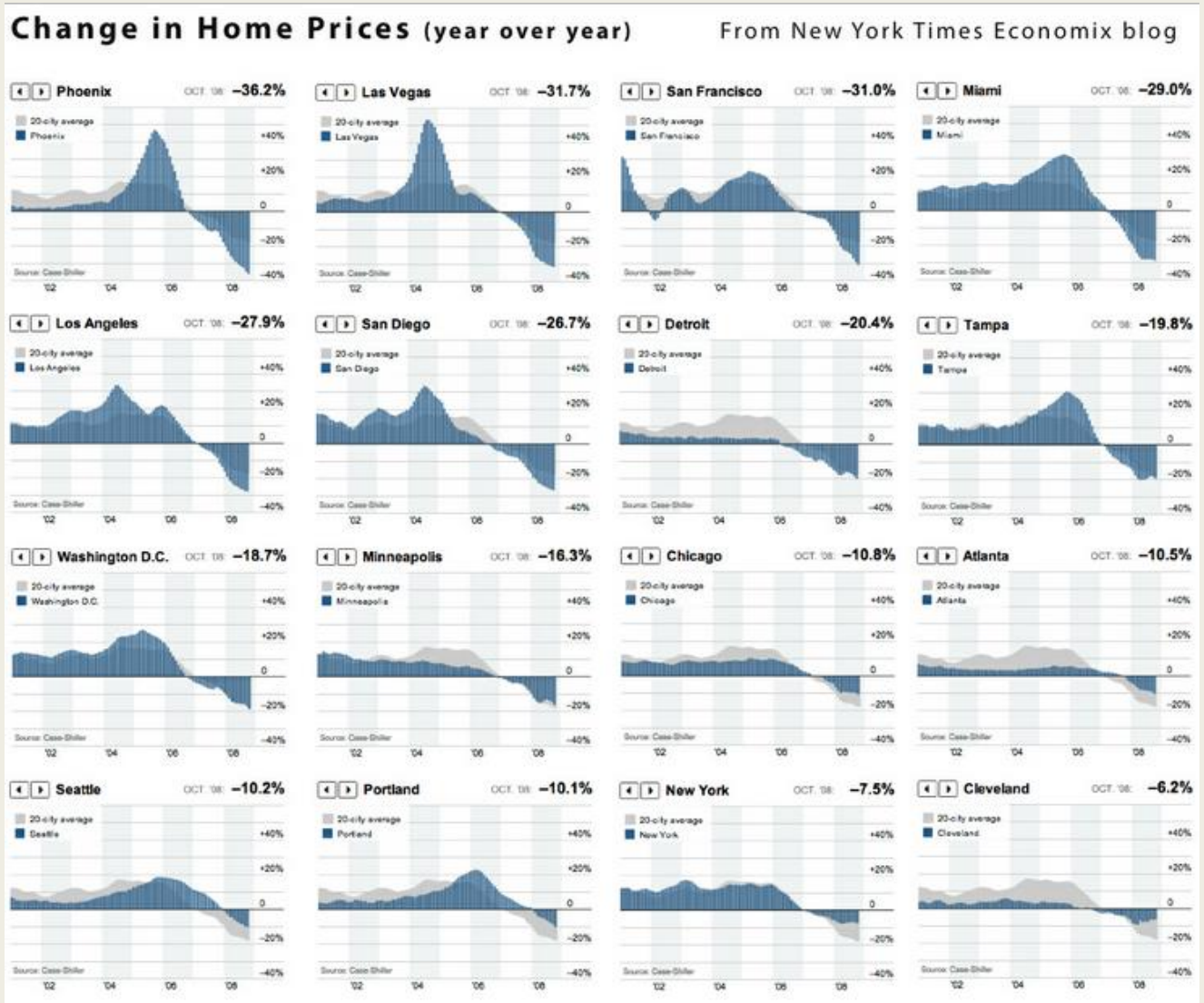
- Our short memory can only deal with a limited number of items.
- What can we do if we are dealing with a very large number of categories?
- https://www.data-to-viz.com/caveat/spaghetti.html





What can we do to reduce the number of categories?

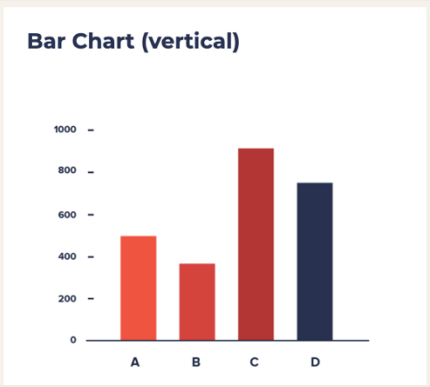
- Report the most important categories
- Combine categories
- Using small multiples



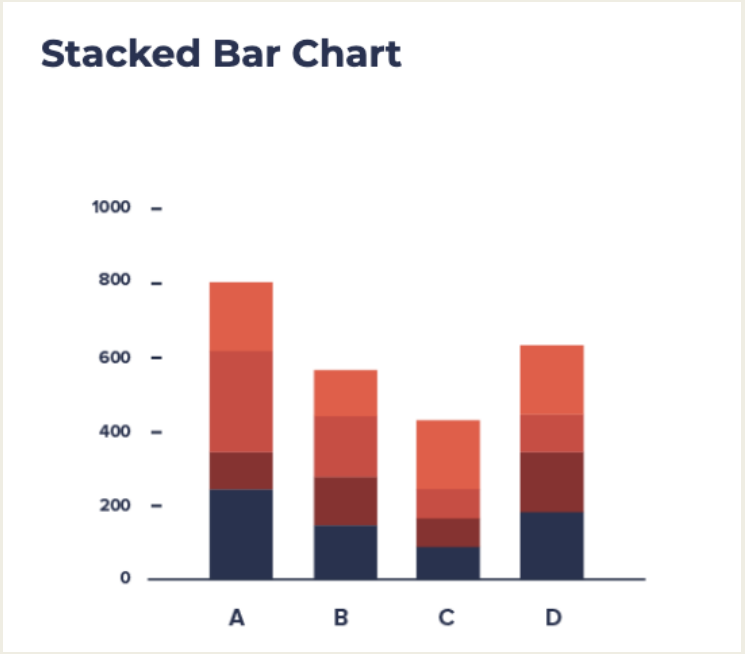
A comparison of bar, line, and scatterplot

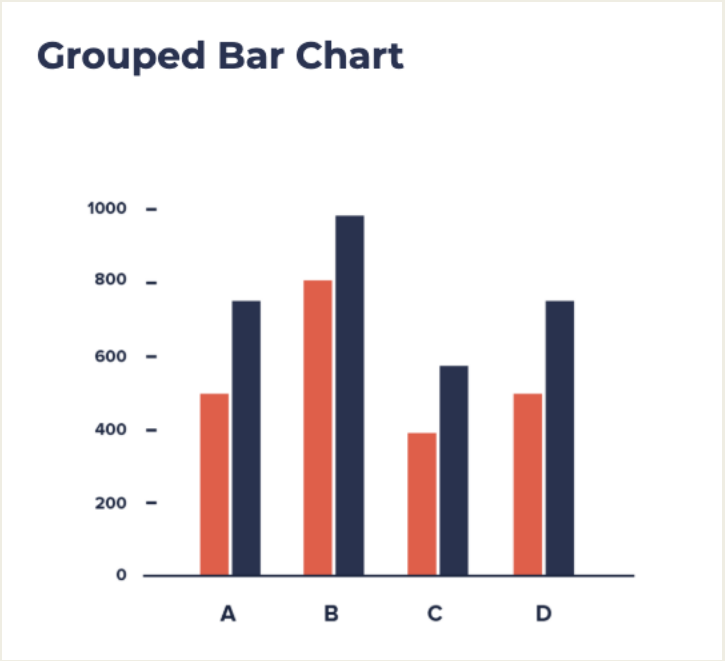
- Bar chart:
  - *Category (x) + Numeric (y)*
  - *To compare categories (normally need to be sorted by the value)*
- Line chart:
  - *Nominal/Datetime (x) + Numeric (y)*
  - *Can compare categories + lines*
  - *It is more frequently used for timeline data.*
- Scatterplot:
  - *Numeric (x + y)*
  - *To show the distribution, correlation, and scatter*

Paying attention to details



What different stories can the latter two graphs tell?





Why including extra information?

- The inclusion of extra variables can add important information to our story and make the story more interesting.
- But we will talk about what features are available next week under the Grammar of Graphics model.