CS 625: Data Visualization

Course Introduction

Presented by: Bhanuka Mahanama

Department of Computer Science
Old Dominion University
bhanuka@cs.odu.edu

Based on lecture materials by Dr. Michele Weigle, and Dr. Tamara Munzner

About Me

- Bhanuka Mahanama
  - Graduated from ODU in Summer 2025
  - Joined ODU in Fall 2025
  - Eye Tracking, Human-Computer Interaction, Machine/Deep Learning
- Contact
  - Office: Dragas, 1100H
  - Hours: Wednesday 10:00-12:00 PM
    - Appointments preferred
  - Email: bhanuka@cs.odu.edu
  - Web: https://www.cs.odu.edu/~bhanuka

An eye tracker
(https://pupil-labs.com)

Course Material

- **Canvas**
  - All course materials
  - Assignments and grades
  - Announcements
  - Discussions
  - https://canvas.odu.edu/courses/201491/
- **GitHub**
  - All course materials
  - Assignment instructions
  - https://github.com/odu-cs625-datavis/public-spring26-bhanuka
- No prerequisites, comfortable with,
  - Basic programming and statistics

What to do Today?

- Homework 0 is out
  - Setup your GitHub account
  - Setup your observables account
  - Instructions provided
- Discussion
  - Introduce yourself to the class!

Define Visualization

Define Visualization

- Computer-based visualization systems provide visual representations of **datasets** designed to help **people** carry out tasks more effectively

Why have a human in the loop?

- Visualization is suitable when there is a need to augment human capabilities
- Instead of replacing people with computational decision-making methods

When We Don’t Need Visualizations

- When fully automatic solution **exists and is trusted**
- Examples
  - Retrieve or sort matching records
    - Well known and proven algorithms
  - Choose optimal routes for packets
    - Proven algorithms
    - Routing algorithms
  - Stock trading algorithms
    - Buy/sell specific price/pattern reached
- No need for time consuming check from a human

Why Have a Human in the Loop?

- Many problems are ill specified
  - Don’t know how to approach, many possible questions
- Many possibilities for human in the loop visual data analysis
  - Long term use for end users
    - Genomic browsers: researchers use the same system over years
  - Presentation of known results
    - Election results: make them easy to understand
  - Stepping stone to assess requirements
  - Refine and develop parameters for automatic solutions
  - Help end users of automatic solution verify and build trust
    - Medical systems: model predictions with confidence
- Exploit the pattern detection capabilities of **Human Vision**

Why Visual Representation?

- Visual representations are part of external representation
  - Touchable objects (abacus)
  - Pencil and paper
  - Computer screen, …
- Offload internal cognition and memory usage
  - Allows us to surpass limitations of our own cognition
- Diagrams as external representations
  - Information can be organized by spatial location
  - Search - grouping items needed for problem-solving in one location
  - Recognition - grouping relevant info for one item in the same location

Computer-based visualization systems provide **visual representations** of datasets designed to help people carry out tasks more effectively

Visualizations Can Extend Your Memory: Example

- What is 7 x 8, 5 x 8, 4 x 7, 4 x 5?

Visualizations Can Extend Your Memory: Example

- What is 7 x 8, 5 x 8, 4 x 7, 4 x 5?
- What is 57 x 48?

Visualizations Can Extend Your Memory: Example

- What is 7 x 8, 5 x 8, 4 x 7, 4 x 5?
- What is 57 x 48?

Paper	Mental buffer
57 x 48 6	[7 x 8 = 56], carry 5

Example courtesy Tamara Munzner, Univ. of British Columbia

Visualizations Can Extend Your Memory: Example

- What is 7 x 8, 5 x 8, 4 x 7, 4 x 5?
- What is 57 x 48?

Paper	Mental buffer
57 x 48 456	[7 x 8 = 56], carry 5 [5 x 8 = 40] + [5] = 45

Example courtesy Tamara Munzner, Univ. of British Columbia

Visualizations Can Extend Your Memory: Example

- What is 7 x 8, 5 x 8, 4 x 7, 4 x 5?
- What is 57 x 48?

Paper	Mental buffer
57 x 48 456 80	[7 x 8 = 56], carry 5 [5 x 8 = 40] + [5] = 45 [4 x 7 = 28], carry 2

Example courtesy Tamara Munzner, Univ. of British Columbia

Visualizations Can Extend Your Memory: Example

- What is 7 x 8, 5 x 8, 4 x 7, 4 x 5?
- What is 57 x 48?

Paper	Mental buffer
57 x 48 456 2280	[7 x 8 = 56], carry 5 [5 x 8 = 40] + [5] = 45 [4 x 7 = 28], carry 2 [4 x 5 = 20] + [2] = 22

Example courtesy Tamara Munzner, Univ. of British Columbia

Visualizations Can Extend Your Memory: Example

- What is 7 x 8, 5 x 8, 4 x 7, 4 x 5?
- What is 57 x 48?

Paper	Mental buffer
57 x 48 456 2280 2736	[7 x 8 = 56], carry 5 [5 x 8 = 40] + [5] = 45 [4 x 7 = 28], carry 2 [4 x 5 = 20] + [2] = 22 [6 + 0 = 6], [5 + 8 = 13] carry 1, [4 + 2 = 6] + [1] = 7, ...

Example courtesy Tamara Munzner, Univ. of British Columbia

Why Human Vision?

- Human vision is a high bandwidth channel to brain
  - Significant amount of background processing
  - Brain combines information captured by fixations and saccades
  - Experience of seeing everything simultaneously
- Sound (or sonification): lower bandwidth
  - No support for simultaneous experience
  - Experience as a sequential stream
- Touch/haptics: poor record and replay
  - Low bandwidth
- Taste, smell: no viable record or replay devices

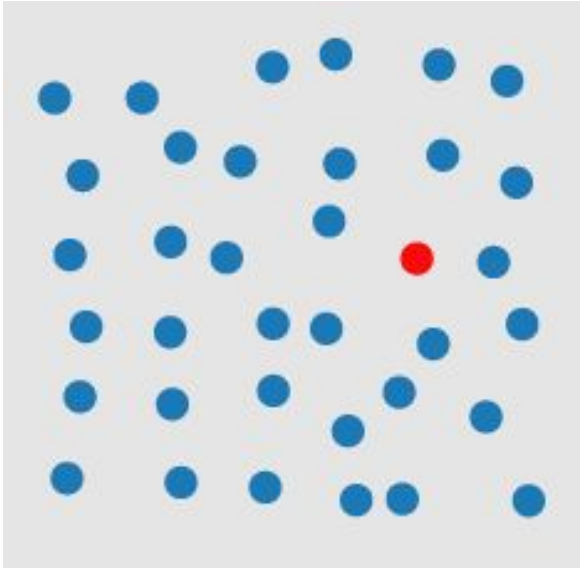
Computer-based visualization systems provide **visual** representations of datasets designed to help people carry out tasks more effectively

Preattentive Processing: Example

- Try to find the red dot

Preattentive Processing: Example

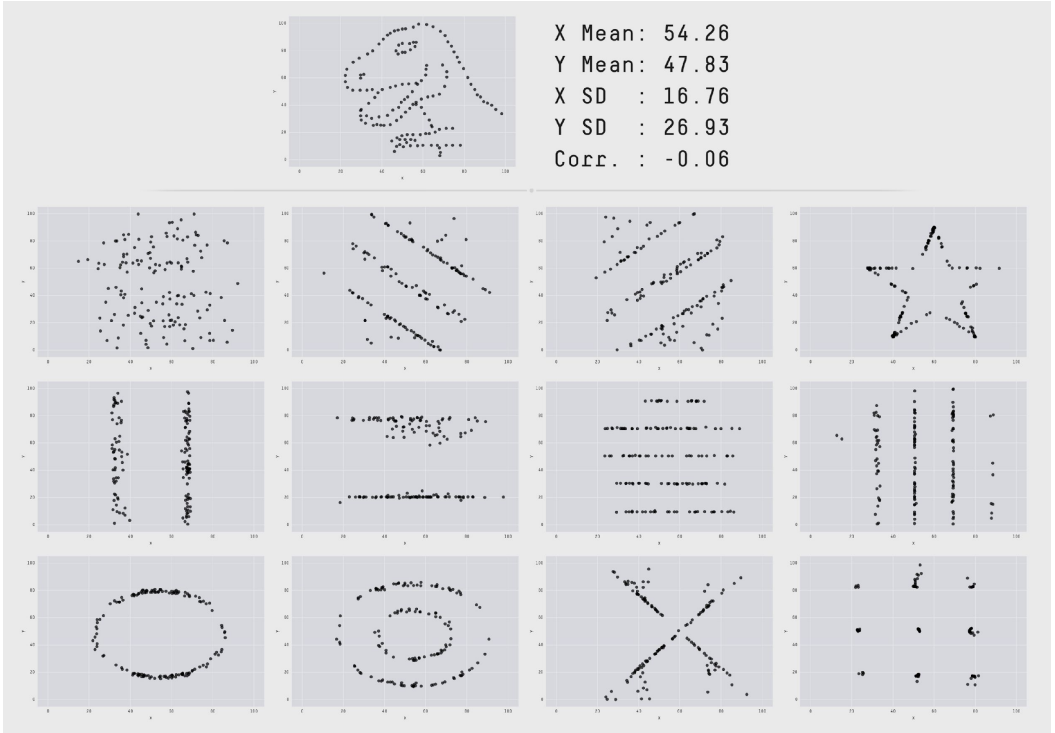
- Try to find the red dot



http://www.csc.ncsu.edu/faculty/healey/PP/index.html

Why Represent All the Data?

- Summaries lose information
  - Confirm expected patterns
  - Find unexpected patterns



J. Matejka and G. Fitzmaurice, "Same Stats, Different Graphs", SIGCHI 2017, https://www.research.autodesk.com/publications/same-stats-different-graphs/

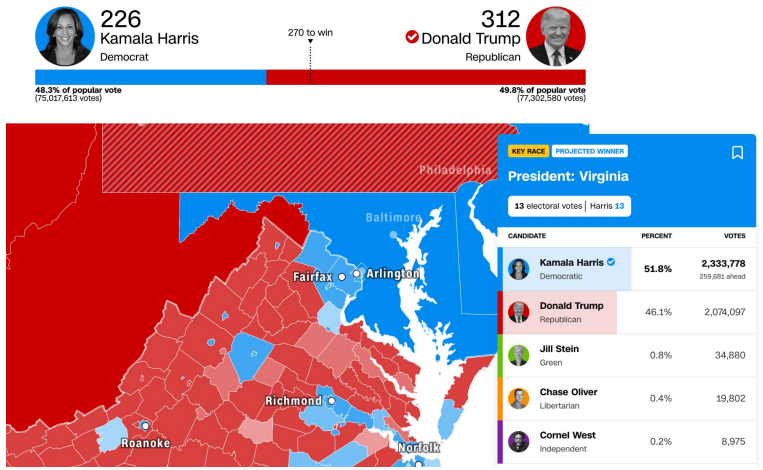
Computer-based visualization systems provide visual **representations of datasets** designed to help people carry out tasks more effectively

What are the Resource Limitations?

- We must consider three different kinds of resource limitations
- Computational limitations
  - Computation time
  - System memory
- Display limits
  - Pixels are precious and constrained
  - Information density
    - Ratio of space used for visualization vs unused
    - Clutter and wasting space
- Human limits
  - Human time, memory, and attention

Visualization Idiom Design Space

- A visualization idiom: distinct approach to **create and manipulate** a visual representation
- The design space of possible idioms is huge!
- Simple static idioms
  - Scatterplots, bar charts, line charts
- Complex idioms can link multiple charts through interaction
  - Visualization of election results

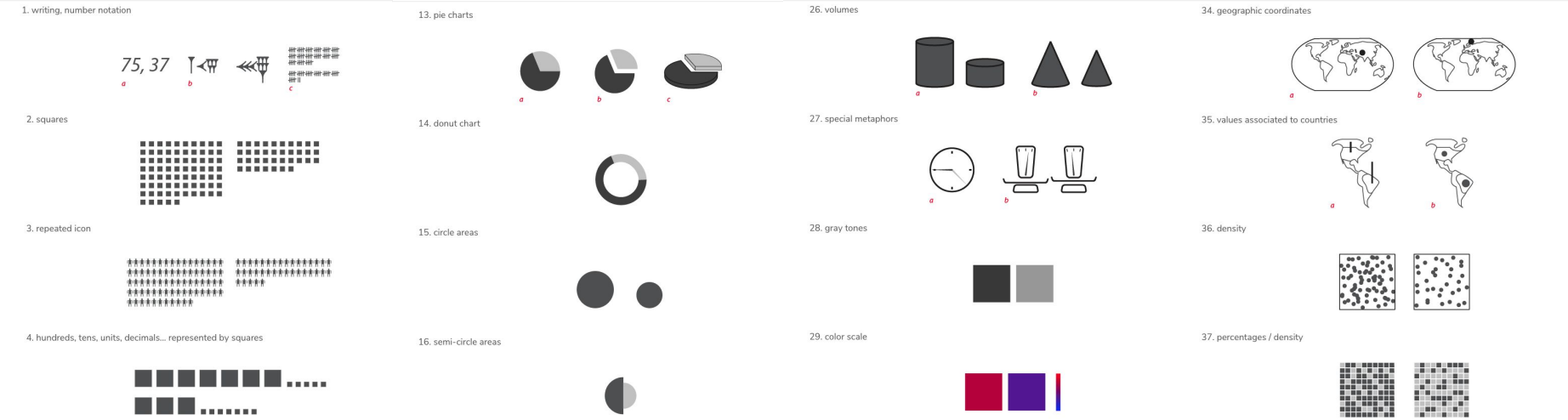


https://www.cnn.com/election/2024/results/

Idiom Design Space: Example

- Think of the numbers 75 and 37. How can you visualize them?

45 Ways to Visualize Two Quantities



https://web.archive.org/web/20201127013528/https://en.rockcontent.com/blog/45-ways-to-communicate-two-quantities/

Data Viz Project

https://datavizproject.com

Why Interactivity?

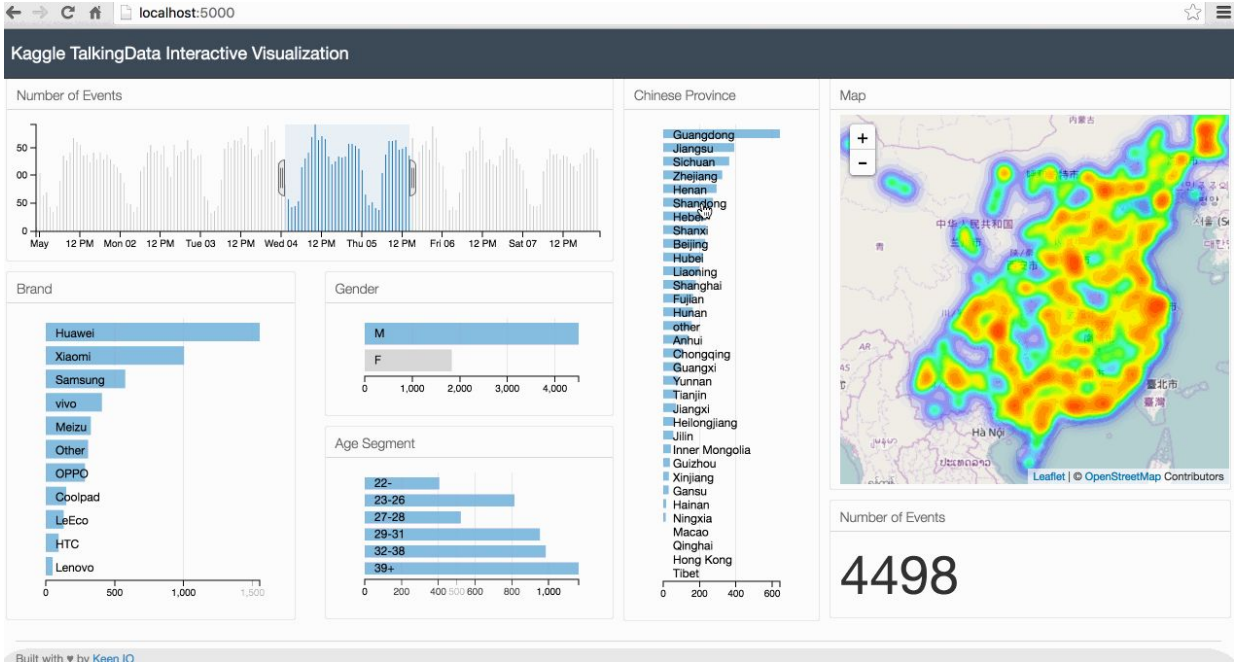
When datasets are large,

- Cannot show everything at once!
  - Computational, display, and human limitations
- Static view shows only one aspect of dataset
- Interactivity is important to handle complexity and show multiple aspects of the dataset

A visualization idiom: distinct approach to create and **manipulate** a visual representation

Interactivity Example

- Interactivity can show multiple aspects of the dataset
- Investigation at different levels



http://adilmoujahid.com/posts/2016/08/interactive-data-visualization-geospatial-d3-dc-leaflet-python/

Visualization Design Effectiveness

Visualization design is full of tradeoffs, and **most possibilities in the design space are ineffective** for a particular **task**, so **validating the effectiveness** of a design is both necessary and **difficult**

- Why focus on tasks?
- Why focus on effectiveness?
- Why are most design ineffective?
- Why validation difficult?

Why Focus on Tasks

- The intended task is important just important as the data to be visualized
- A tool that serves well for one task can be poorly suited for another task
  - Google Maps: works well for real-time traffic awareness not for urban planning
- Four categories of tasks
  - Presentation
  - Discovery
  - Enjoyment of information
  - Information of later use

Visualization design is full of tradeoffs, and most possibilities in the design space are ineffective for a particular **task**, so validating the effectiveness of a design is both necessary and difficult

Why Focus on Effectiveness

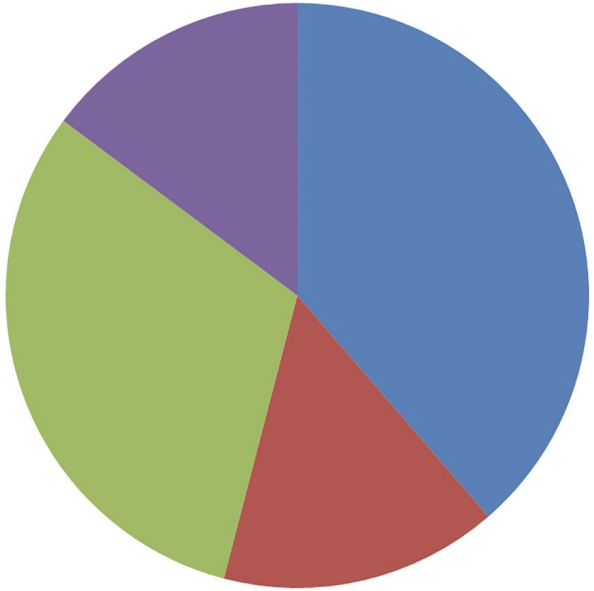
- Effectiveness indicates if a visualization supports what users need to do
  - A visually appealing chart is not necessarily a useful one
  - The goal of visualization is insight, not decoration

- But, no picture can tell the truth, the whole truth, and nothing but the truth

Visualization design is full of tradeoffs, and most possibilities in the design space are ineffective for a particular task, so validating the **effectiveness** of a design is both necessary and difficult

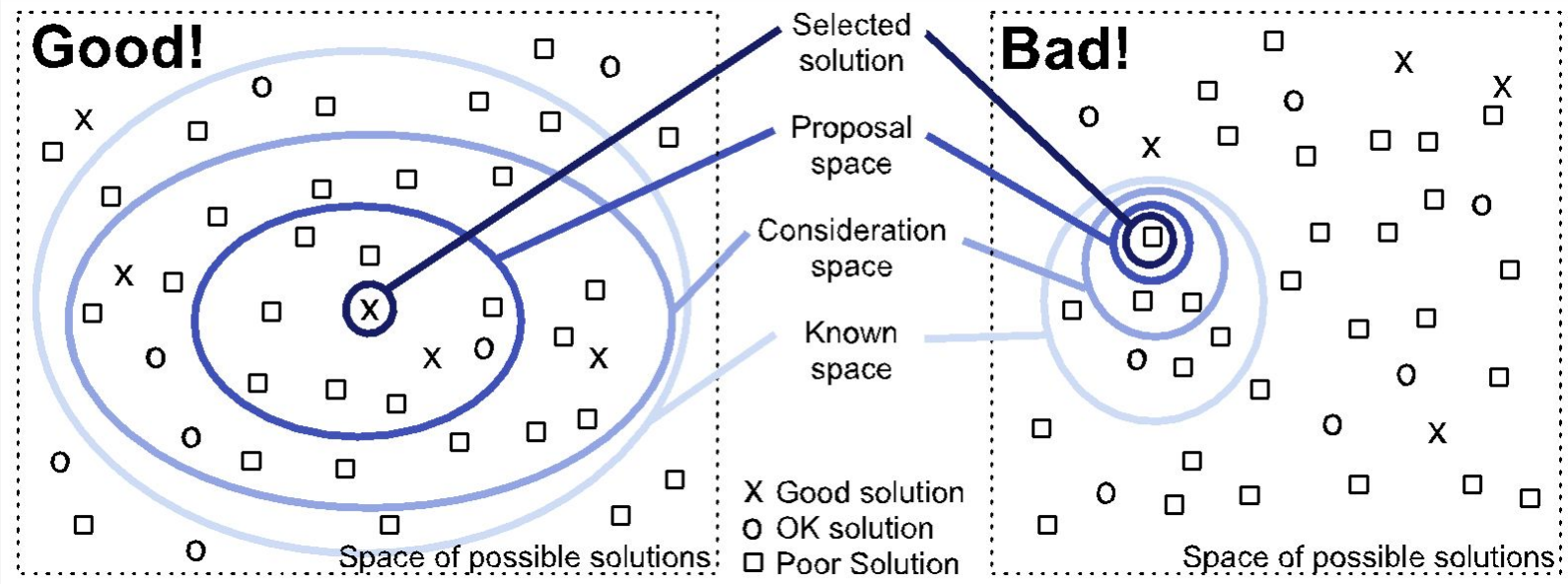
Why Are Most Designs Ineffective?

- Design may not match with human perception and/or intended task
- Example: Pie chart in a business report
- Human perception
  - Poor at accurately comparing angles and areas
  - Hard to judge small differences
- Intended task
  - Compare categories, Rank values



Visualization design is full of tradeoffs, and **most possibilities in the design space are ineffective** for a particular task, so validating the effectiveness of a design is both necessary and difficult

Search Space Design for Visualization Design



- Goal to satisfy, find one of many possible good solutions
- Large known and consideration space is a good strategy

Munzner, Figure 1.5

Why Validation Difficult?

- There are so many questions you could ask when considering whether a tool has met your design goals
  - Can users quickly identify the main pattern or trend? What is quick?
  - Can users easily do what they need to do? What is easily?
  - Are values encoded using perceptually effective channels (position, length)?
  - Is interaction intuitive or confusing?
  - How do you argue if a design is better?

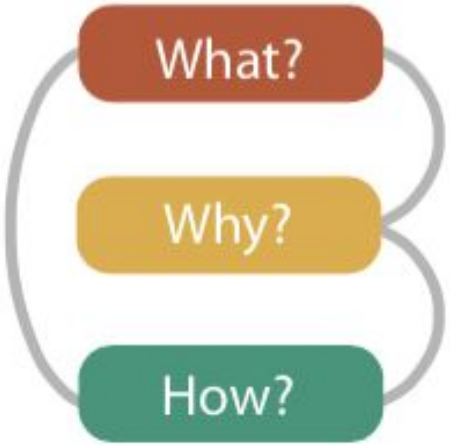
Visualization design is full of tradeoffs, and most possibilities in the design space are ineffective for a particular task, so **validating the effectiveness of a design is both necessary and difficult**

Analyzing Visualizations

- Analyzing existing systems is a good stepping stone to designing new ones.

High-level framework for analyzing visualization use

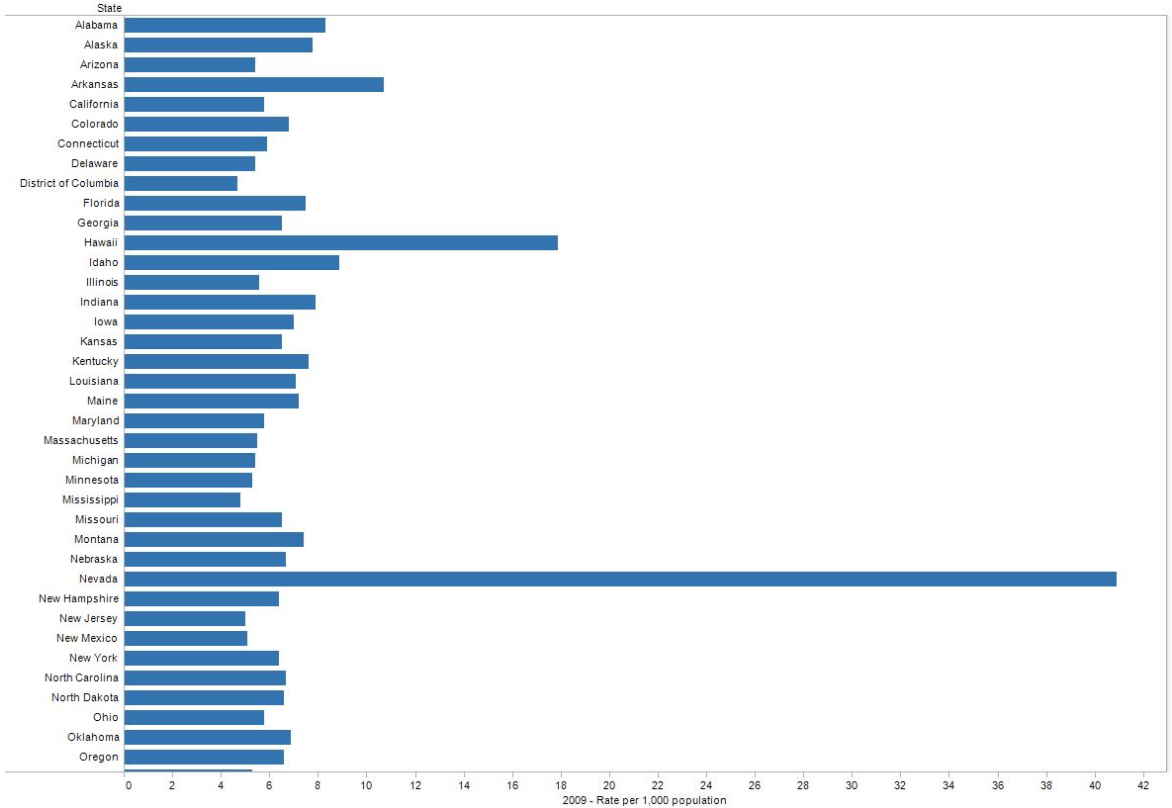
- **What** data the user sees
- **Why** the user intends to use a vis tool
- **How** the visual encoding and interaction idioms are constructed



Each *what-why-how* has a *data-task-idiom* answer

A Typical Workflow

- What
  - Data gathering
  - Data wrangling
- Why
  - Developing questions
  - Initial analysis
- How
  - Charts for analysis
  - Charts for presentation



Marriage rates in US 2009, Michele Weigle, ODU

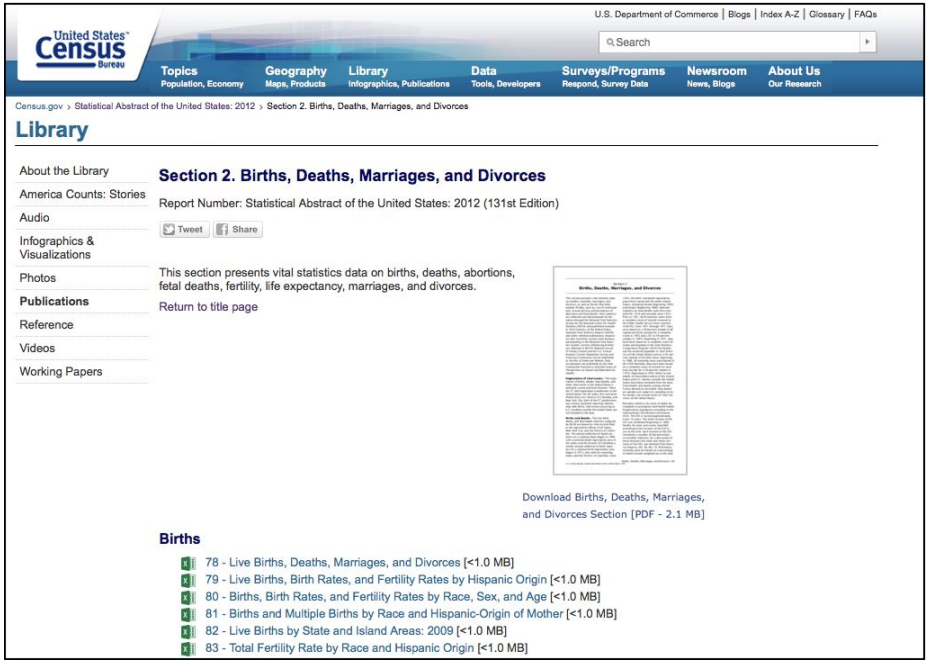
What: Data Gathering

Extract tables from PDFs

- Tabula: https://tabula.technology/

Extract data from webpages

- Beautiful Soup: https://pypi.org/project/beautifulsoup4/
- Scrapy: https://www.scrapy.org/



https://www.census.gov/library/publications/2011/compendia/statab/131ed/births-deaths-marriages-divorces.html

What: Data Wrangling

- Data is often messy
- Tools
  - Excel
  - OpenRefine – filter and clean data files
- Much more on this later

Each ***what***-*why*-*how* has a ***data***-*task*-*idiom* answer

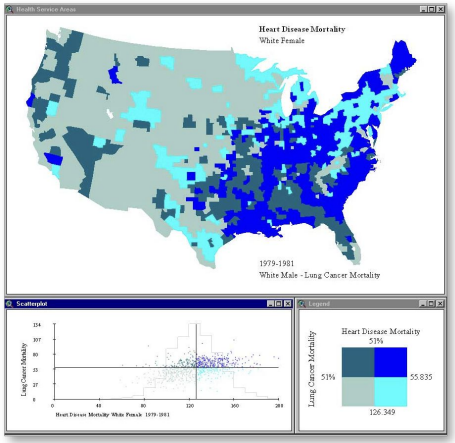
Why

- Developing questions
  - Which states have the highest marriage rates?
  - Which states have the highest divorce rates? is that correlated to marriage rate?
  - Which states have the highest birth rates? is that correlated to marriage rate?
- Initial analysis
  - Excel
  - Google Sheets, Google Charts
  - Tableau
  - Python matplotlib, seaborn
  - Vega-Lite

Each *what-**why**-how* has a *data-**task**-idiom* answer

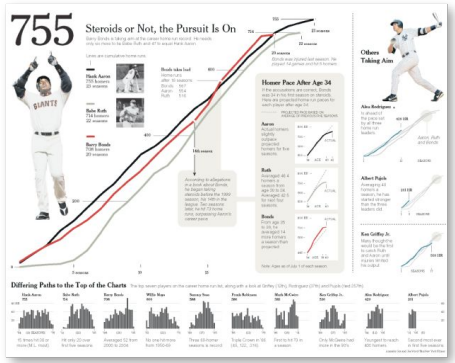
How

Charts for analysis



https://www.geovista.psu.edu/research/healthvisualization/

Charts for presentation



New York Times, 2006

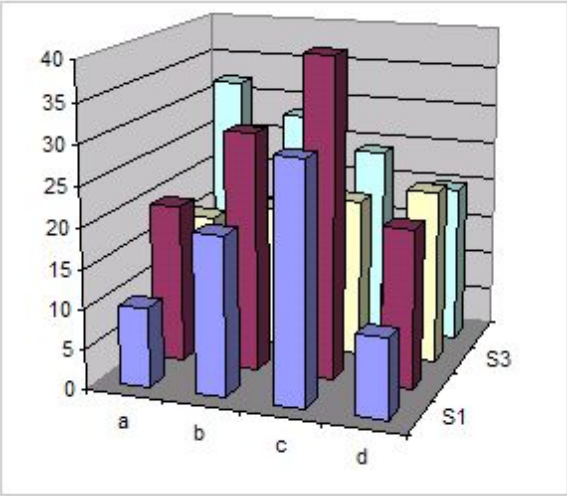
Each *what-why-**how*** has a *data-task-**idiom*** answer

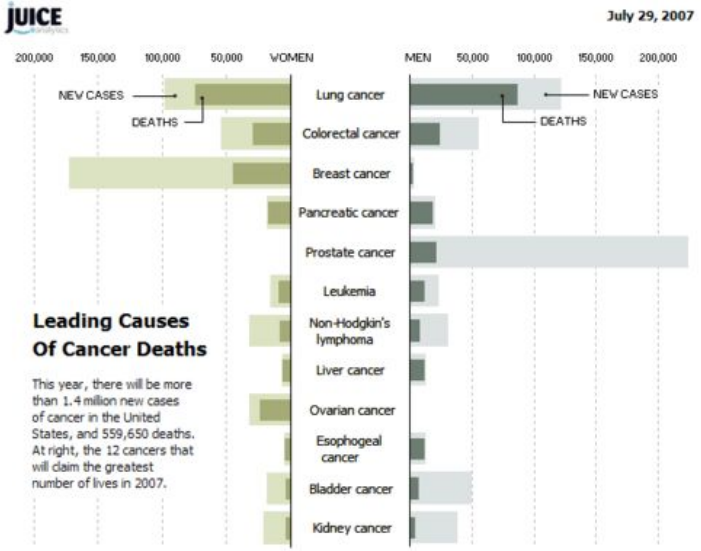
Some tools

- Excel
- Tableau
- Python
- D3
- Vega and Vega-lite
- Observable

Excel

- Good for quick exploring
  - Not familiar with programming
- Not scalable for large datasets



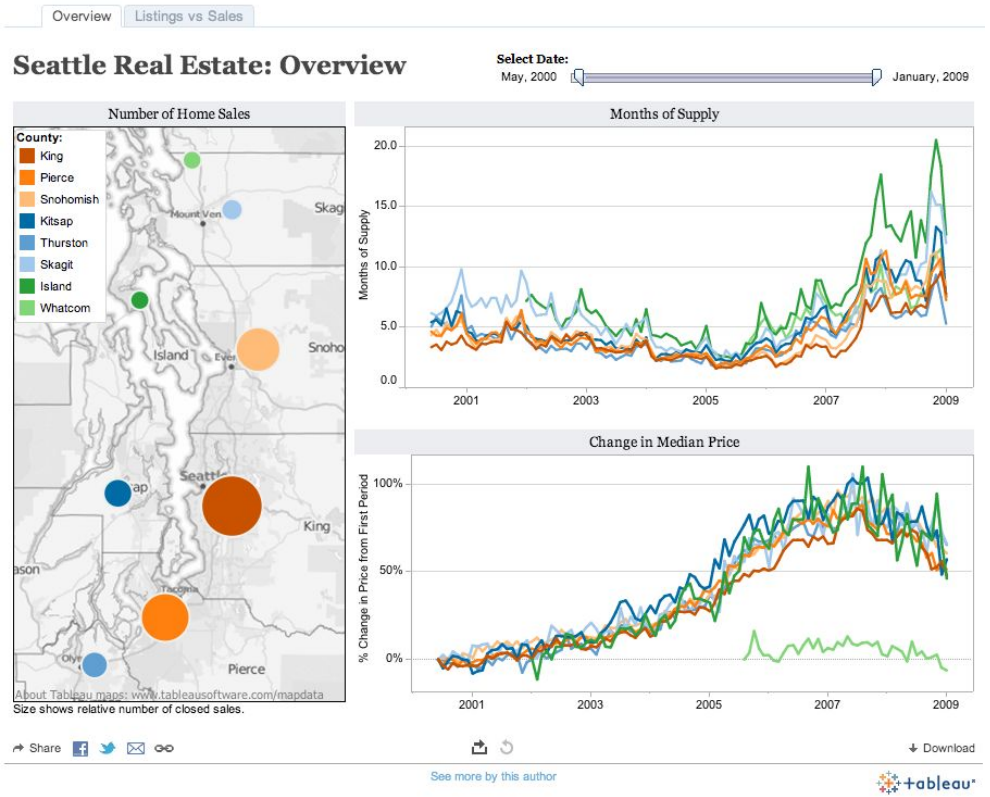


Recreating the NY Times Cancer Graph,
http://www.juiceanalytics.com/writing/recreating-ny-times-cancer-graph/

6 Excel charts you must never use,
http://chandoo.org/wp/2008/09/03/6-charts-to-never-use/

Tableau

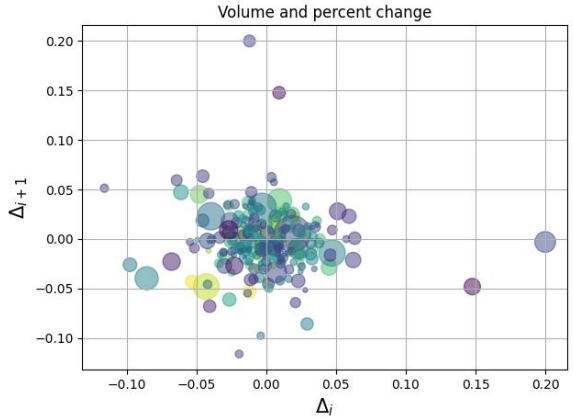
- Powerful commercial product used in business intelligence applications
  - Produce multiple visualizations
  - Create dashboard
- Tableau public is free
  - Make data and visualizations public



https://www.tableau.com/academic/students

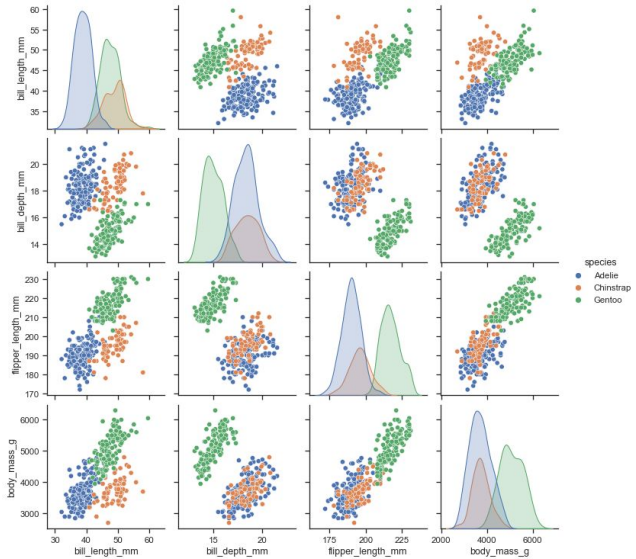
Python

- Some powerful visualization libraries
  - Matplotlib: https://matplotlib.org/
  - Seaborn: https://seaborn.pydata.org/
  - Plotly: https://plotly.com/python/
  - GGplot: https://pypi.org/project/ggplot/



matplotlib

seaborn



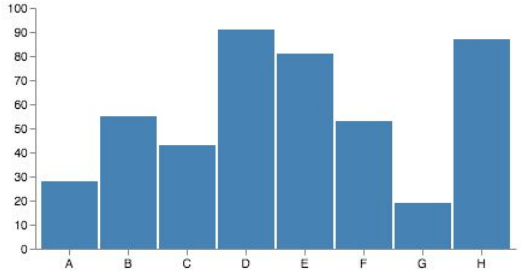
D3

- Web based interactive visualizations
  - Built with HTML, CSS and JavaScript
  - Focus on web based interactive visualizations

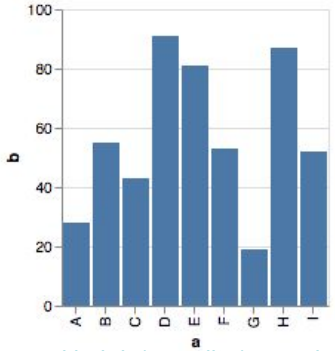
http://d3js.org

Vega and Vega-Lite

- Vega is a visualization grammar, built on top of D3
  - https://vega.github.io/vega/
- Vega-Lite is a higher-level language built on top of Vega
  - https://vega.github.io/vega-lite/
- Provides building blocks for web-based visualizations
  - Defined using JSON specifications



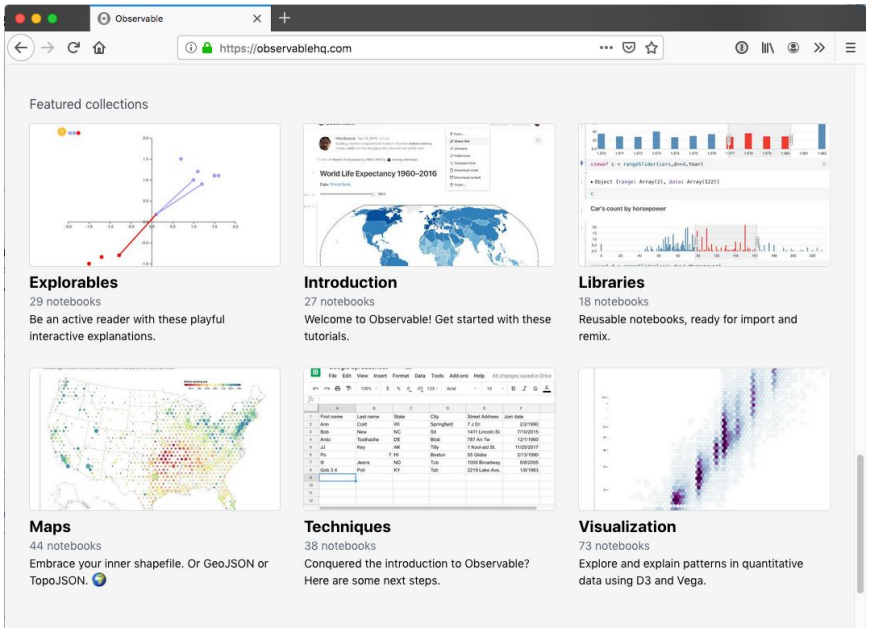
https://vega.github.io/vega/examples/bar-chart/



https://vega.github.io/vega-lite/examples/bar.html

Observable

- Notebook-based platform for JavaScript
  - Similar to Jupyter notebooks
  - JavaScript instead of Python
- Build analysis and visualizations using D3, Vega-Lite, or Observable Plot
  - https://observablehq.com/plot/

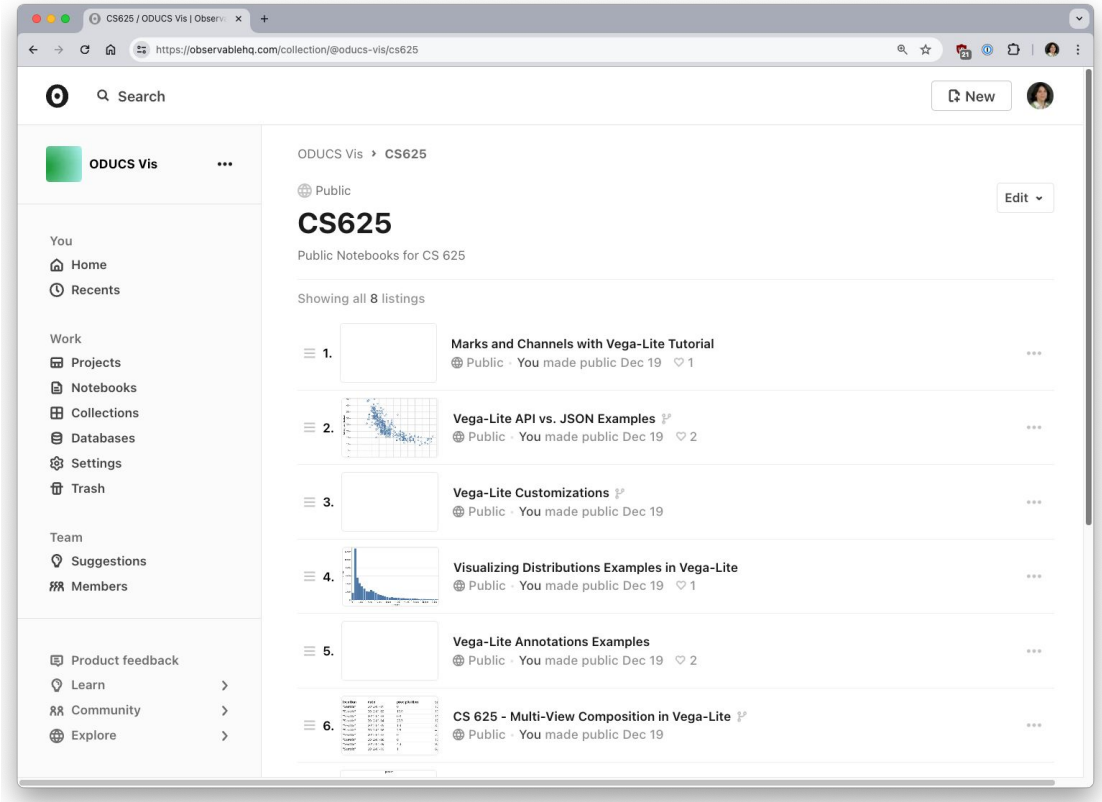


https://observablehq.com

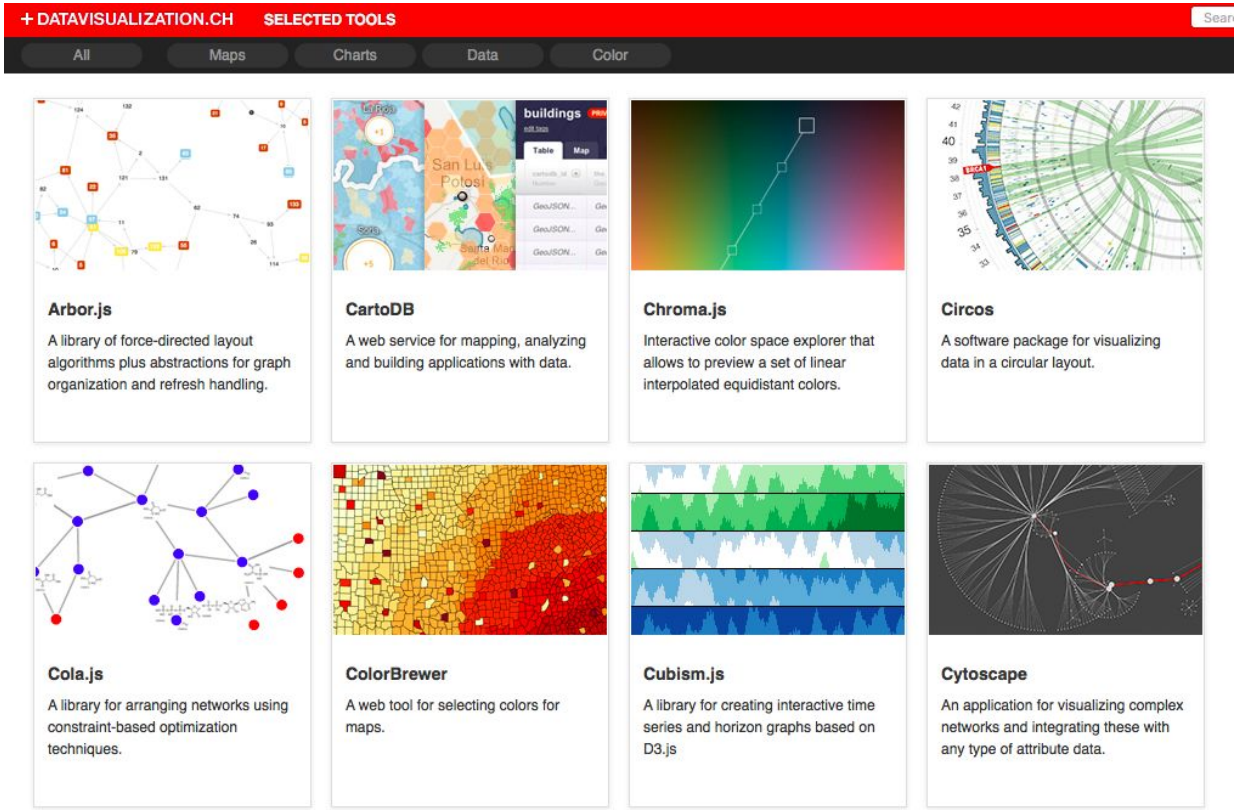
Class Observable Team

Private notebooks for assignments

https://observablehq.com/@oducs-vis

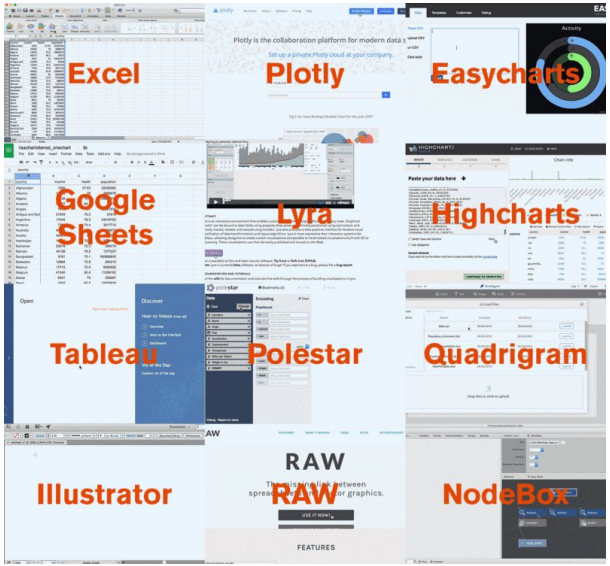


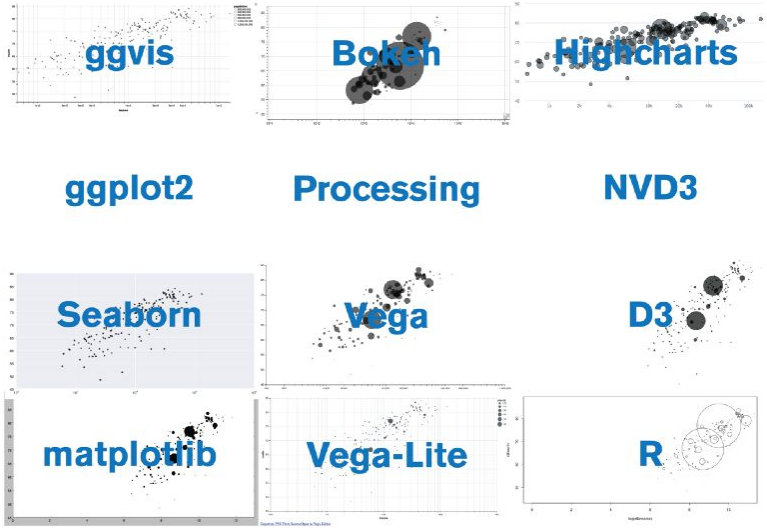
Choosing Tools



https://selection.datavisualization.ch/

What I Learned Recreating One Chart Using 24 Tools





https://source.opennews.org/en-US/articles/what-i-learned-recreating-one-chart-using-24-tools/

Tools for Analysis vs. Presentation

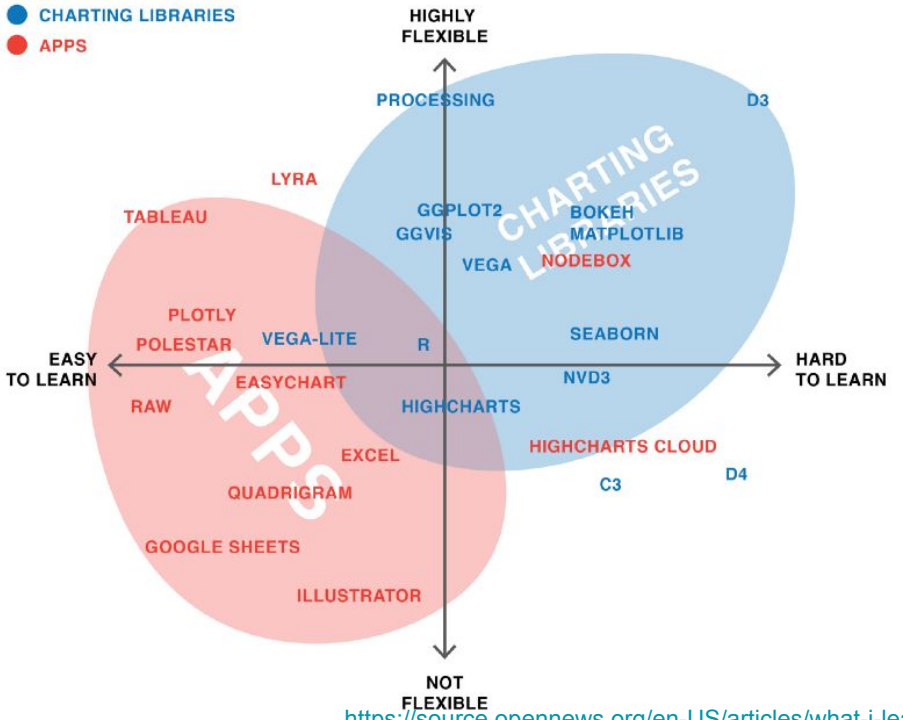
https://source.opennews.org/en-US/articles/what-i-learned-recreating-one-chart-using-24-tools/

Apps vs. Libraries and Static vs. Interactive

	STATIC	WEB - INTERACTIVE
APPS	ILLUSTRATOR, NODEBOX, EXCEL, POLESTAR, RAW	HIGHCHARTS CLOUD, QUADRIGRAM, EASYCHRT, DATAWRAPPER, TABLEAU, PLOTLY, GOOGLE SHEETS
CHARTING LIBRARIES	GGPLOT2, MATPLOTLIB, R, SEABORN, BOKEH, PROCESSING	D3, D4, C3, NVD3, GGVIS, HIGHCHARTS, SHINY, VEGA, VEGA-LITE

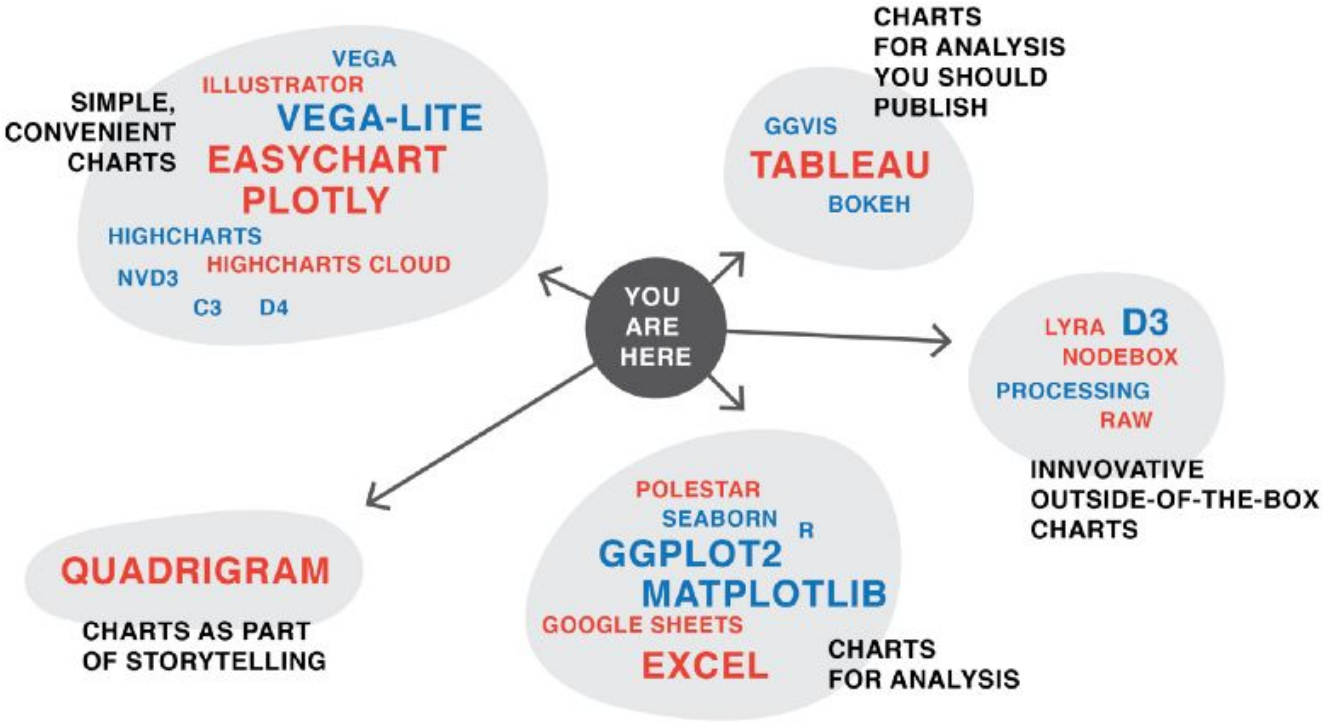
https://source.opennews.org/en-US/articles/what-i-learned-recreating-one-chart-using-24-tools/

Flexibility vs. Learning Curve



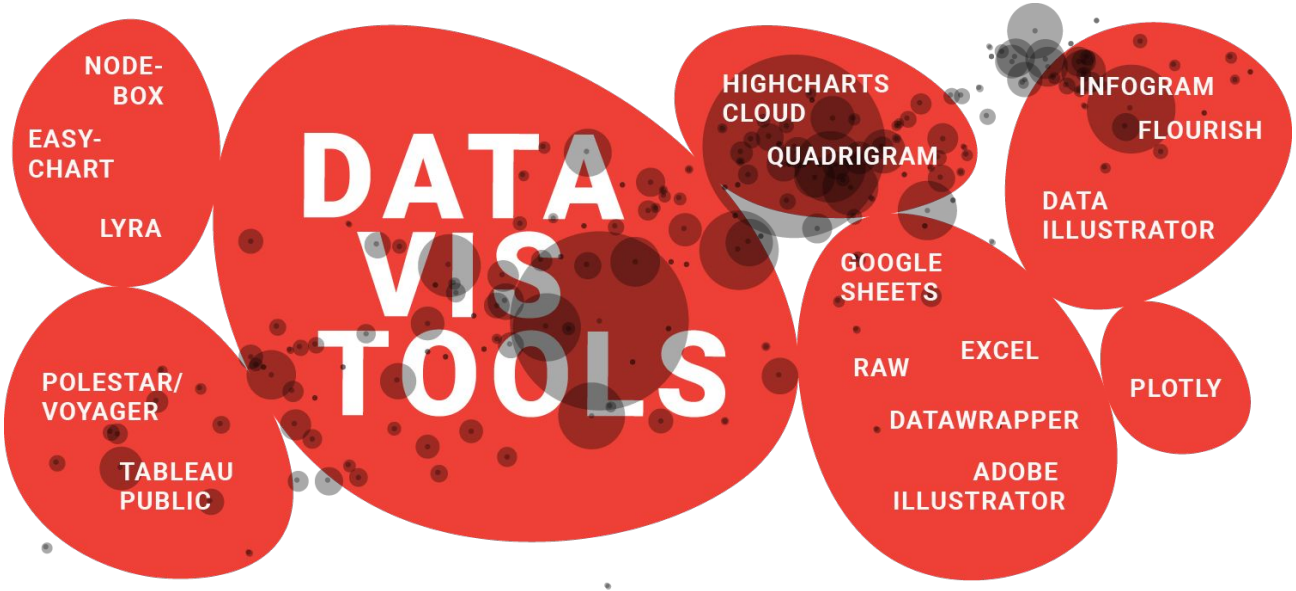
https://source.opennews.org/en-US/articles/what-i-learned-recreating-one-chart-using-24-tools/

Mental Map



https://source.opennews.org/en-US/articles/what-i-learned-recreating-one-chart-using-24-tools/

One Chart, Nine Tools – Revisited



New tools: Infogram, Flourish Data Illustrator

https://lisacharlotterost.github.io/datavistools-revisited