Creating Custom Visualisations

David Trye | COMPX532 | DSIGN532

1

Goals

1. To introduce you to some powerful **visualisation tools**

2. To demonstrate some **prototyping strategies** using examples from my research

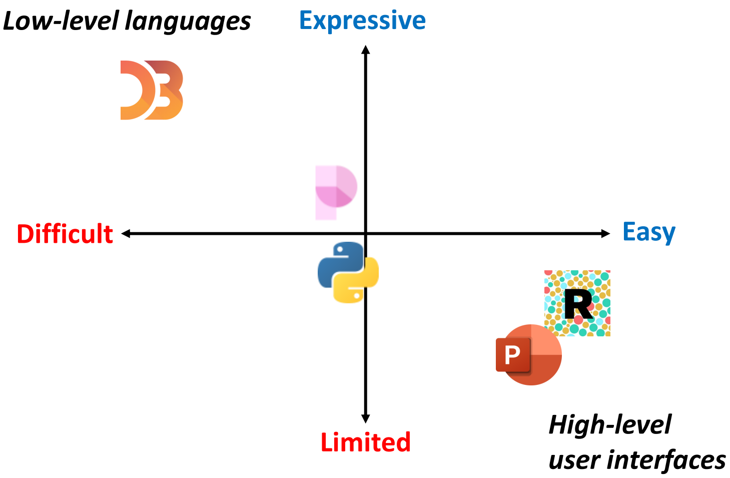
2

Disclaimer: A Subjective Selection

- There are many other powerful visualisation tools available!
- For instance, see Top 15 Data Visualization Tools For 2022

Expressiveness vs Usability

"Visualisation tools sacrifice *generality* for greater *efficiency* performing particular tasks.
Code has unparalleled expressiveness; it is the most general tool we have.
A medium for discovery *must* be general. Creativity requires composition."
– Mike Bostock, D3 Creator, 2017



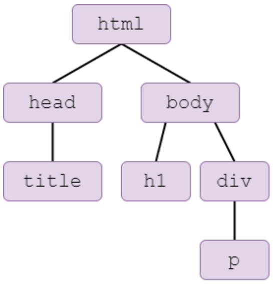
D3 via Observable

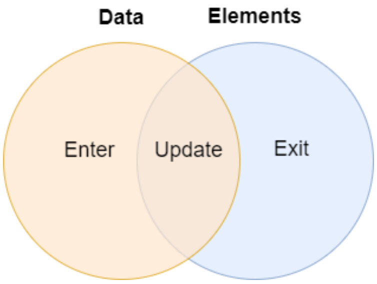
5

What is D3?

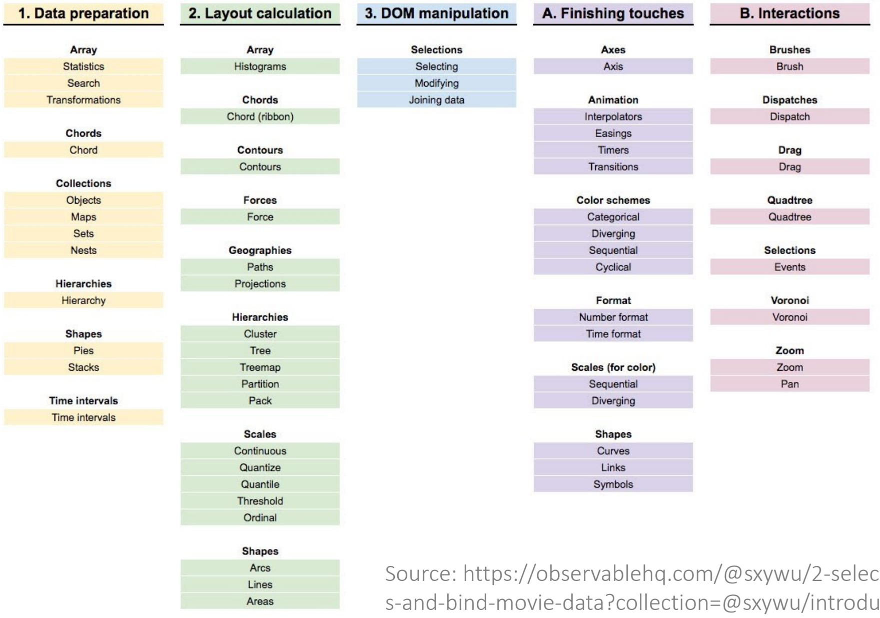
- D3 (or D3.js) is a JavaScript library that stands for **data-driven documents**
- Allows you to dynamically manipulate the **Document Object Model (DOM)**
  - Key concepts include selections, data-binding & the enter-update-exit pattern
  - Also supports data preparation, layout calculation, scales, shapes, animation, interaction and more
- Interactive, flexible, expressive and fast!
- See http://shirleywu.studio/d3intro/

Document Object Model





6



1. Data preparation

Array
- Statistics
- Search
- Transformations

Chords
- Chord

Collections
- Objects
- Maps
- Sets
- Nests

Hierarchies
- Hierarchy

Shapes
- Pies
- Stacks

Time intervals
- Time intervals

2. Layout calculation

Array
- Histograms

Chords
- Chord (ribbon)

Contours
- Contours

Forces
- Force

Geographies
- Paths
- Projections

Hierarchies
- Cluster
- Tree
- Treemap
- Partition
- Pack

Scales
- Continuous
- Quantize
- Quantile
- Threshold
- Ordinal

Shapes
- Arcs
- Lines
- Areas

3. DOM manipulation

Selections
- Selecting
- Modifying
- Joining data

A. Finishing touches

Axes
- Axis

Animation
- Interpolators
- Easings
- Timers
- Transitions

Color schemes
- Categorical
- Diverging
- Sequential
- Cyclical

Format
- Number format
- Time format

Scales (for color)
- Sequential
- Diverging

Shapes
- Curves
- Links
- Symbols

B. Interactions

Brushes
- Brush

Dispatches
- Dispatch

Drag
- Drag

Quadtree
- Quadtree

Selections
- Events

Voronoi
- Voronoi

Zoom
- Zoom
- Pan

Source: https://observablehq.com/@sxywu/2-select-existing-petal-s-and-bind-movie-data?collection=@sxywu/introduction-to-d3-js

Observable Notebooks

https://observablehq.com/explore

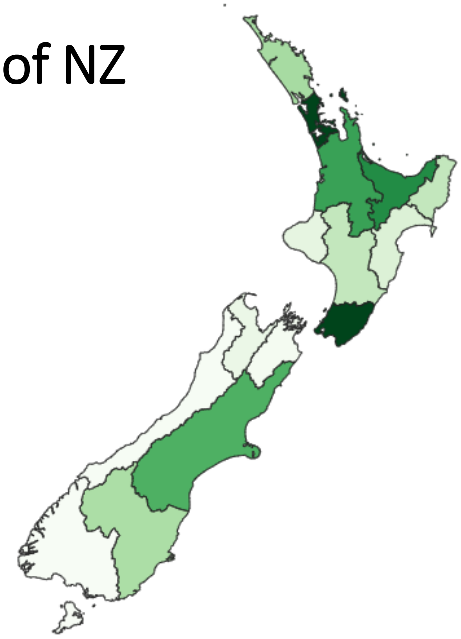
- "Observable helps you use data to think"
  - Explore and visualise data
  - Get ideas/inspiration from existing notebooks
  - Share your insights with the vis community
- Reactive (live) programming
  - Interactive JavaScript notebooks (similar to Jupyter Notebook for Python)
  - Experiment with (big) data in real-time
  - Quickly build and iterate prototypes, re-use code, etc.
  - Supports creation of highly interactive visualisations

Melody Meckfessel,
CEO & Co-founder

Mike Bostock,
CTO & Co-founder
D3 Creator

D3 Example: Choropleth Map of NZ

- Shows the distribution of Māori language tweeters across different regions of NZ
  - Self-reported location
  - Some users missing
- TopoJSON defines geospatial positioning
  - https://github.com/deldersveld/topojson
- Based on the following tutorial
  - https://www.dataviscourse.net/tutorials/lectures/lecture-maps/
- Could be enhanced with simple interaction
- Need to include a legend!



9

Demonstration: Tinkering with D3

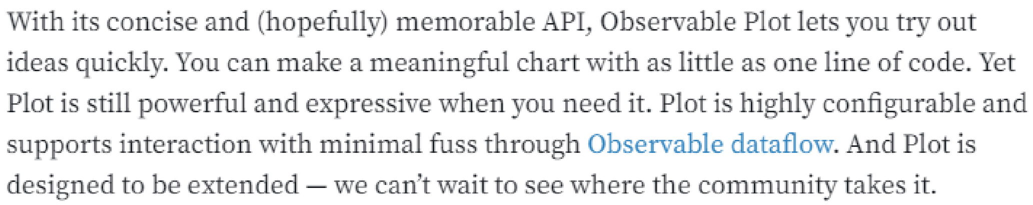
10

Plot via Observable

11

What is Plot?

- A **higher-level** JavaScript library for **exploratory** data visualisation
- Built on top of D3 (but more concise and convenient)
- Centred around the idea of "using vision to think"

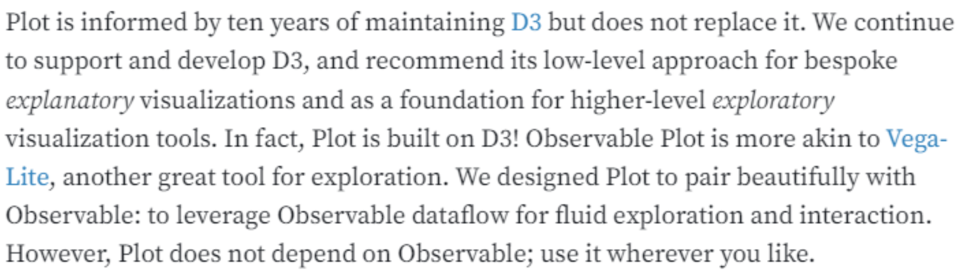


With its concise and (hopefully) memorable API, Observable Plot lets you try out ideas quickly. You can make a meaningful chart with as little as one line of code. Yet Plot is still powerful and expressive when you need it. Plot is highly configurable and supports interaction with minimal fuss through Observable dataflow. And Plot is designed to be extended — we can't wait to see where the community takes it.

12

Plot vs D3

- Plot not intended as a “replacement” for D3



Plot is informed by ten years of maintaining D3 but does not replace it. We continue to support and develop D3, and recommend its low-level approach for bespoke *explanatory* visualizations and as a foundation for higher-level *exploratory* visualization tools. In fact, Plot is built on D3! Observable Plot is more akin to Vega-Lite, another great tool for exploration. We designed Plot to pair beautifully with Observable: to leverage Observable dataflow for fluid exploration and interaction. However, Plot does not depend on Observable; use it wherever you like.

- See https://observablehq.com/@observablehq/introducing-observable-plot

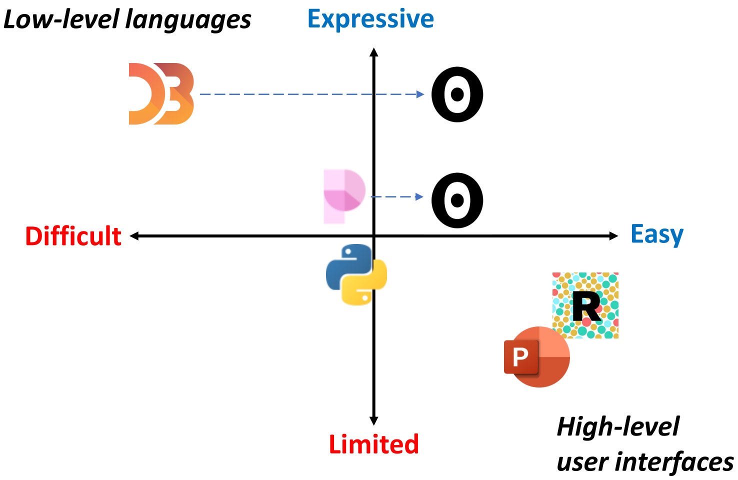
13

Demonstration: Prototyping with Plot

14

Expressiveness vs Usability Revisited

"Visualisation tools sacrifice *generality* for greater *efficiency* performing particular tasks.
Code has unparalleled expressiveness; it is the most general tool we have.
A medium for discovery *must* be general. Creativity requires composition."
– Mike Bostock, D3 Creator, 2017



Low-level languages

Expressive

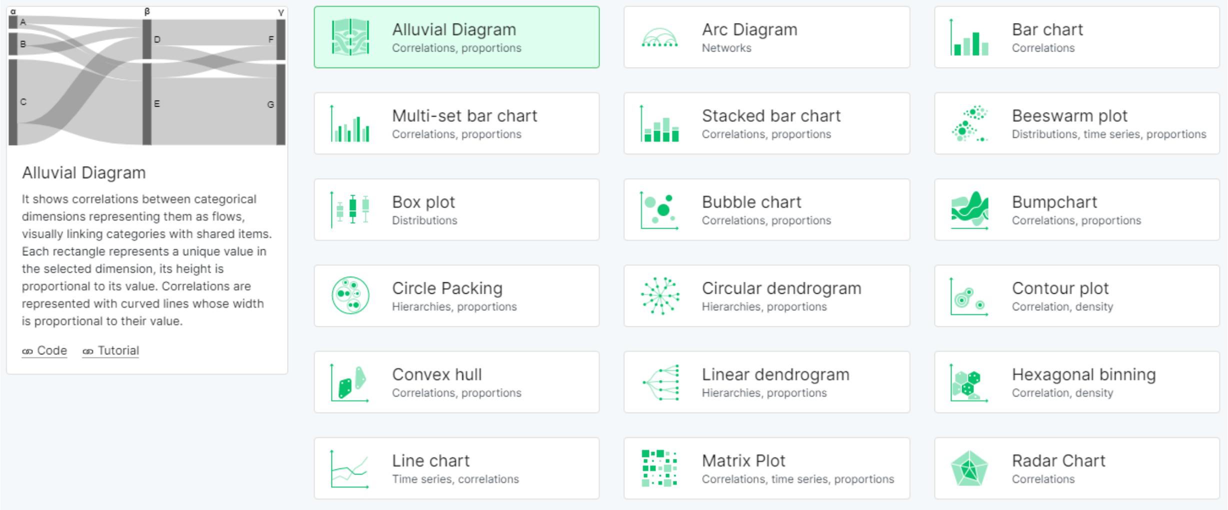
Difficult

Easy

Limited

High-level user interfaces

RAW Graphs



Alluvial Diagram

It shows correlations between categorical dimensions representing them as flows, visually linking categories with shared items. Each rectangle represents a unique value in the selected dimension, its height is proportional to its value. Correlations are represented with curved lines whose width is proportional to their value.

Code Tutorial

- Alluvial Diagram – Correlations, proportions
- Arc Diagram – Networks
- Bar chart – Correlations
- Multi-set bar chart – Correlations, proportions
- Stacked bar chart – Correlations, proportions
- Beeswarm plot – Distributions, time series, proportions
- Box plot – Distributions
- Bubble chart – Correlations, proportions
- Bumpchart – Correlations, proportions
- Circle Packing – Hierarchies, proportions
- Circular dendrogram – Hierarchies, proportions
- Contour plot – Correlation, density
- Convex hull – Correlations, proportions
- Linear dendrogram – Hierarchies, proportions
- Hexagonal binning – Correlation, density
- Line chart – Time series, correlations
- Matrix Plot – Correlations, time series, proportions
- Radar Chart – Correlations

RAWGraphs

- “The missing link between spreadsheets and data”
  - Instantly maps data variables to visual channels
  - Quick and convenient!
  - Uses D3 under the hood
- No programming knowledge required
  - Friendly user interface
  - However, limited layouts available
    - Doesn’t support detailed (low-level) customisation
    - BUT can build custom models with basic understanding of D3

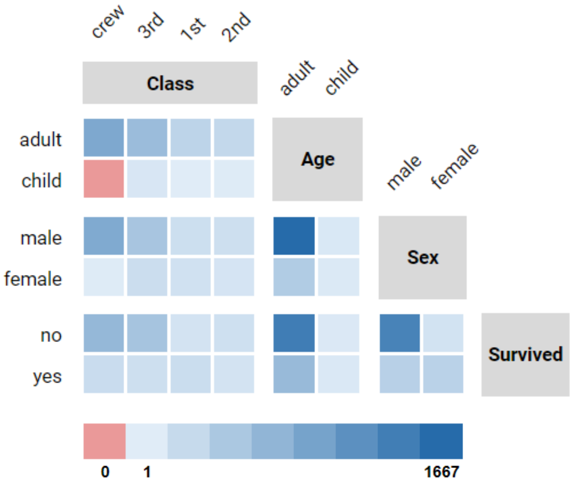
17

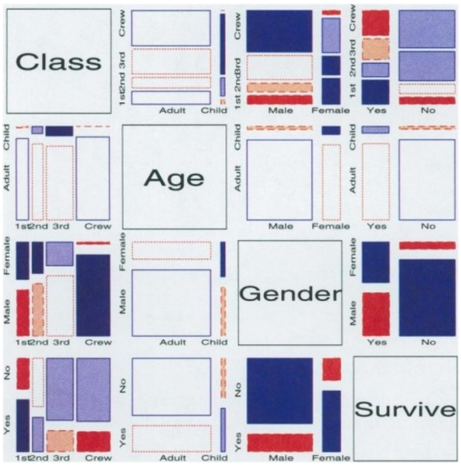
Demonstration: Prototyping with RAW Graphs

18

Prototyping with Python (Pandas), Sheets & PowerPoint

Creating Custom 'Staircase Plots'

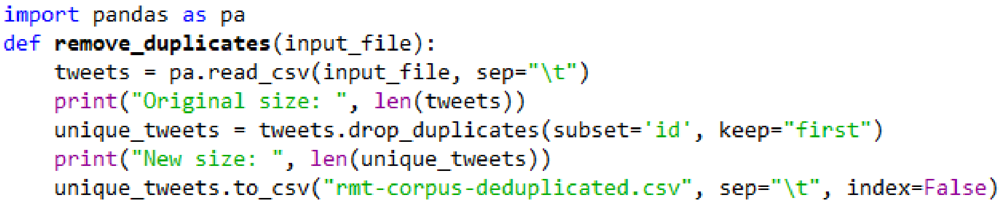




Friendly (1999)

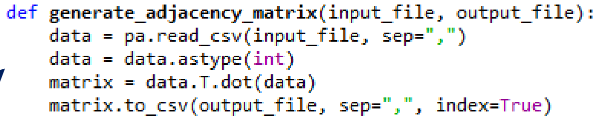
Data Wrangling with Python

- I tend to use Python for cleaning & pre-processing data
  - Popular alternatives include R and JavaScript
- Pandas *DataFrames* are especially useful for handling datasets with many variables (multi-dimensional data)
  - Rows = data items
  - Columns = variables



```
import pandas as pa
def remove_duplicates(input_file):
    tweets = pa.read_csv(input_file, sep="\t")
    print("Original size: ", len(tweets))
    unique_tweets = tweets.drop_duplicates(subset='id', keep="first")
    print("New size: ", len(unique_tweets))
    unique_tweets.to_csv("rmt-corpus-deduplicated.csv", sep="\t", index=False)
```

Process



```
def generate_adjacency_matrix(input_file, output_file):
    data = pa.read_csv(input_file, sep=",")
    data = data.astype(int)
    matrix = data.T.dot(data)
    matrix.to_csv(output_file, sep=",", index=True)
```

1. Download Titanic dataset
2. Compute pairwise adjacency matrix using **Pandas**, keeping categories grouped by variable
3. Open output file as spreadsheet
4. Remove redundant values
5. Create variable ‘boxes’
6. Colour cells using conditional formatting
7. Create legend
8. Remove numeric values
9. Fine-tune display (borders, etc.)
10. Add tooltips in **PowerPoint**

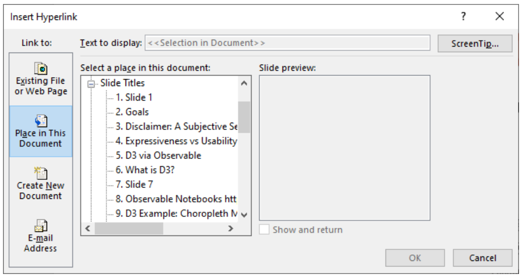
Sheets

Demonstration: Staircase Plots Spreadsheet

23

Adding Interaction with PowerPoint

- Can create Tooltips and Buttons
  - Buttons are just clickable regions that point to different slides
  - Simply create (transparent) shapes, then press ‘Ctrl + K’

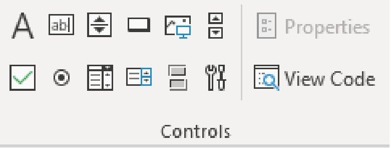


Text to display: <<Selection in Document>>

Select a place in this document:

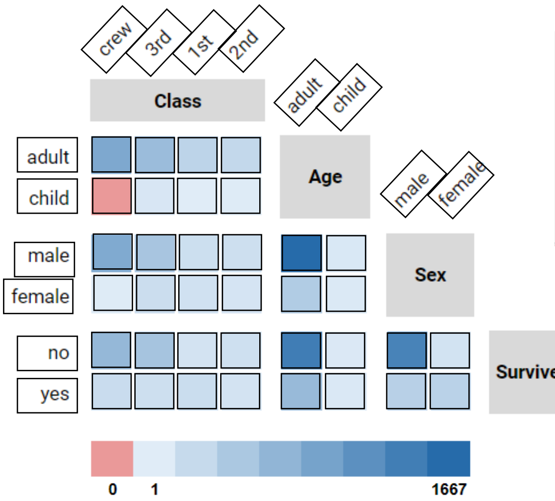
- Slide Titles
  - 1. Slide 1
  - 2. Goals
  - 3. Disclaimer: A Subjective Se
  - 4. Expressiveness vs Usability
  - 5. D3 via Observable
  - 6. What is D3?
  - 7. Slide 7
  - 8. Observable Notebooks htt
  - 9. D3 Example: Choropleth M

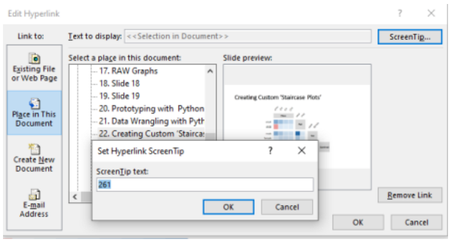
- Other controls in ‘Developer’ mode



24

Creating Custom 'Staircase Plots'





Text to display: <<Selection in Document>>

Select a place in this document:

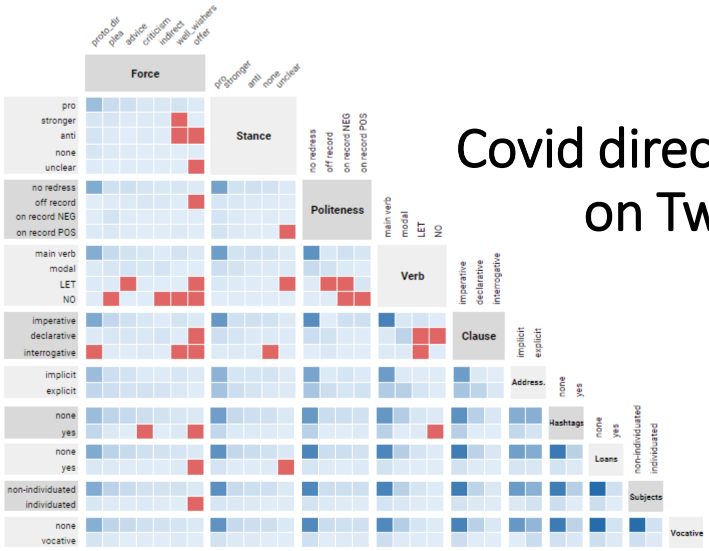
17. RAW Graphs
18. Slide 18
19. Slide 19
20. Prototyping with Python
21. Data Wrangling with Pyth
22. Creating Custom 'Stairca

Set Hyperlink ScreenTip

ScreenTip text: 261

25

A more complex dataset…



Covid directives
on Twitter

26

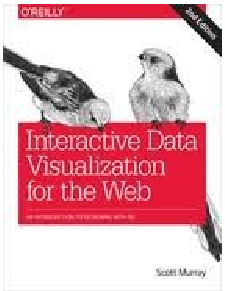
Resources

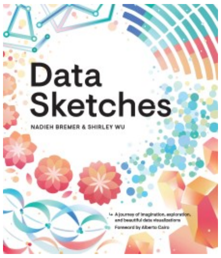
- Observable
  - Introduction
  - Twitter
  - YouTube Tutorials
- Plot
  - Introduction
  - Eric Lo’s Examples
- D3
  - Shirley Wu’s Introduction to SVG & D3
  - Mike Freeman’s University of Washington Tutorials
  - Mike Bostock’s D3 Gallery
  - Prototyping in D3 by Amelia Wattenberger

27

More Resources

- D3 (more generally)
  - HTML, CSS and JavaScript
  - D3 Graph Gallery
  - Scott Murray’s *Interactive Data Visualization for the Web*
  - *Data Sketches* by Nadieh Bremer & Shirley Wu
- Miscellaneous
  - Data Wrangling with Pandas
  - Interactive Prototypes with PowerPoint
  - Vega-Lite and Altair (akin to Plot)
  - NetworkX





28

References

- Bostock, M., Ogievetsky, V., & Heer, J. (2011). **D3: data-driven documents**. *IEEE transactions on visualization and computer graphics*, *17*(12), 2301-2309.
- Friendly, M. (1999). Extending **mosaic displays**: Marginal, conditional, and partial views of categorical data. *Journal of Computational and graphical Statistics*, 8(3), 373-395.
- Satyanarayan, A., Moritz, D., Wongsuphasawat, K., & Heer, J. (2016). **Vega-lite**: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics*, *23*(1), 341-350.
- Mauri, M., Elli, T., Caviglia, G., Uboldi, G., & Azzi, M. (2017). **RAWGraphs**: a visualisation platform to create open outputs. In *Proceedings of the 12th biannual conference on Italian SIGCHI chapter* (pp. 1-5).

Thank you!