Using Visualisation Toolkits

David Trye | COMPX532 | DSIGN532

1

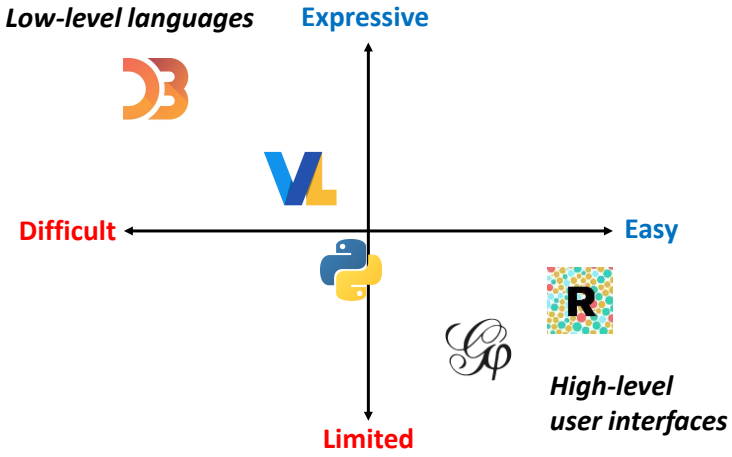
Goals

1. To introduce you to some popular visualisation tools
   - D3 & Vega-Lite via Observable
   - RAW Graphs
   - Python (+ vis libraries)
   - Gephi
2. To show you some visualisations I've been working on using real language data
   - Think about when/how you could use these tools (or similar) to create/customise your own visualisations

2

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



3

Disclaimer: A subjective selection

- There are many other powerful visualisation tools available!
- For instance, see this overview by Cameron Chapman

4

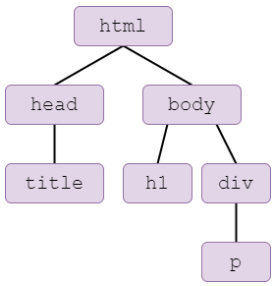
D3 via Observable

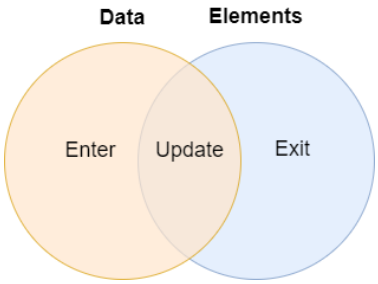
5

What is D3?

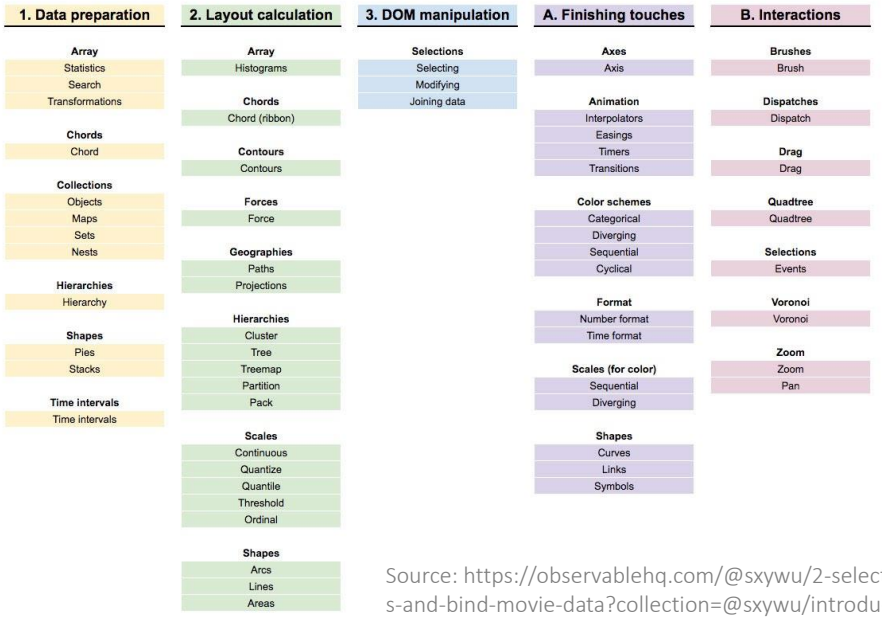
- D3 (or D3.js) is a JavaScript library that stands for *data-driven documents*
- Allows you to dynamically manipulate the Document Object Model (DOM)
  - Key concepts include selections, data-binding & the enter-update-exit pattern
  - Also supports data preparation, layout calculation, scales, shapes,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

Document Object Model





6



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

7

Observable Notebooks

https://observablehq.com/explore

- "Observable helps you use data to think"
  - Explore and visualise data
  - Get ideas/inspiration from existing notebooks
  - Share and publish 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

8

The future of visualisation design?

Introducing Observable Collaboration

We're excited to announce 🥳 some major improvements to collaboration on Observable: you can now invite up to 4 people to edit or view your notebooks! Collaborators can view each others' edits and cursors in real-time — so called *multiplayer* 🎮. It's like Google Docs or Figma, but with code.

We've been testing these features internally for the past few months, and now welcome your feedback 🙋 on the new editing experience. *Help us make the product better for everyone!* 🧡🙏

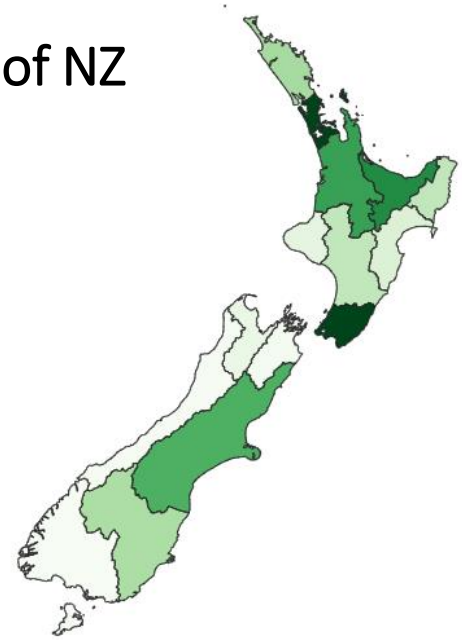
To try our improved collaboration, enable experimental features. To learn more, read about Observable Collaboration and browse the collection. 👊👊

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

9

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!



10

Demonstration: Tinkering with Observable

11

Vega-Lite via Observable

12

What is Vega-Lite?

- A high-level grammar of interactive graphics: http://vega.github.io/vega-lite/
- Built on top of Vega & D3 (more concise and convenient)
- Declarative: Specify *what* you want the visualisation to include rather than *how* to implement it
- Supports interactive, multi-view graphs
- Has an online editor but can also use Observable Notebooks
- See examples and tutorials

13

D3 vs Vega-Lite

- Vega(-Lite) not intended as a "replacement" for D3

> Vega provides a higher-level visualization specification language on top of D3. By design, D3 will maintain an "expressivity advantage" and in some cases will be better suited for novel design ideas. On the other hand, we intend Vega to be convenient for a wide range of common yet customizable visualizations. Vega's design builds on concepts we developed in both Protovis and D3, and is informed by years of research at Stanford and UW.
>
> As is always the case, the right tool for the job depends on the task at hand. We expect D3 will often be the tool of choice for realizing novel visualization design ideas. For common yet customizable chart types, programmatic generation, and flexible rendering, we believe Vega can further facilitate the use of data visualization across a variety of new tools and web applications.

- See https://vega.github.io/vega/about/vega-and-d3/

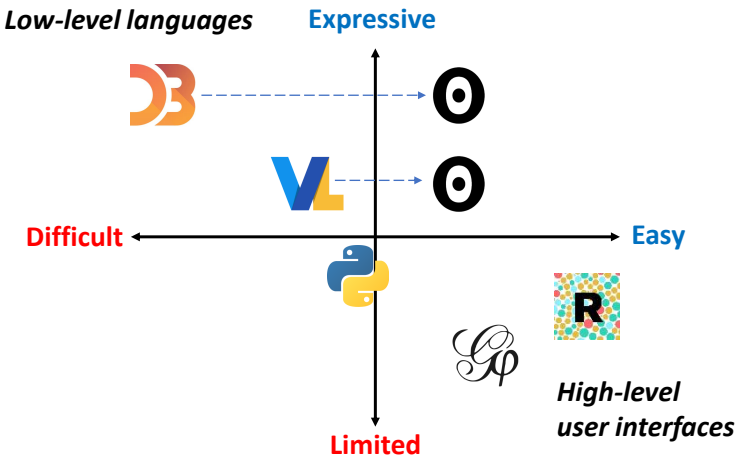
14

Demonstration: Prototyping with Vega-Lite

15

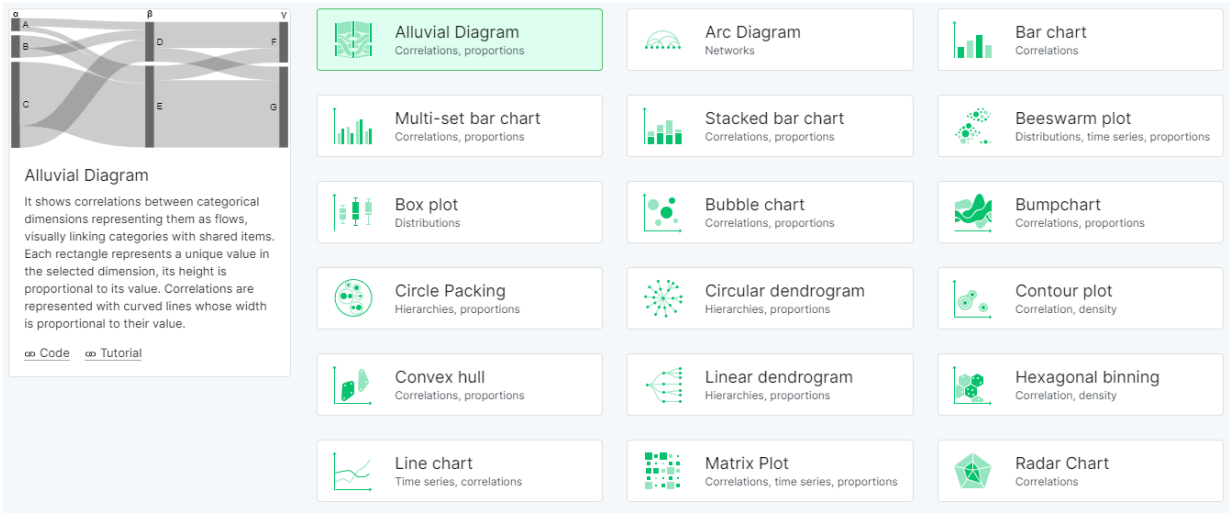
Expressiveness vs Usability Revisited

"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



16

RAW Graphs



17

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

18

Demonstration: Prototyping with RAW Graphs

19

Python + Gephi

20

Data Wrangling with Python

- I tend to use Python for cleaning & pre-processing data
  - Other popular alternatives include R and (more recently) JavaScript
- Pandas *data frames* 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)
```

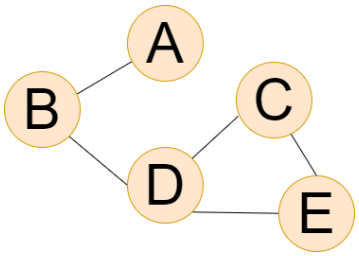
21

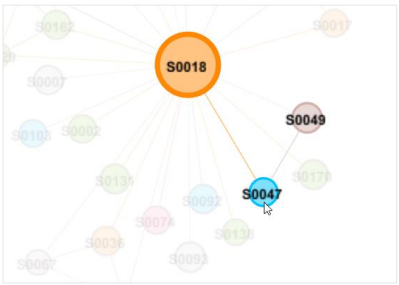
Python Pre-processing & Visualisation Libraries

22

Gephi

- Free, open-source software: https://gephi.org/users/download/
- Specifically for drawing *networks*
- Interaction improves readability
  - Hover over a node to see its neighbours
- Can also re-encode & filter attributes
  - Change the appearance (e.g. colour, size) of nodes & edges according to different attributes
  - Filter nodes by attribute values



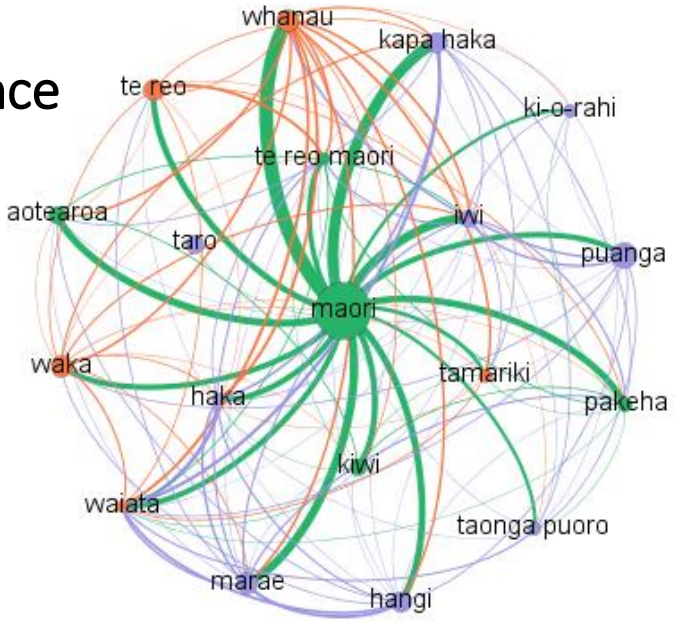


23

Network Example: Loanword Co-occurrence

- Loanword relationships within NZE newspaper articles/texts
- Limitation: Only shows pairwise relationships!

Nodes = Loanwords
Links = Text-level co-occurrence
Node Size = Frequency
Colour = Semantic type



24

Process

1. Extracted Māori loanwords in the corpus
2. Built a co-occurrence matrix
   - Rows = articles
   - Columns = loanwords (1 = present, 0 = absent)
3. Derived loanword pairs and their weights
4. Used Python's *NetworkX* library to export the data into Gephi's GEXF format
   - https://programminghistorian.org/en/lessons/exploring-and-analyzing-network-data-with-python
5. Visualised the network in Gephi

Loanwords

Texts	L1	L2	...	L140
T1	1	0	...	1
T2	0	1	...	0
...	...	...	...	...
TN	0	1	...	1

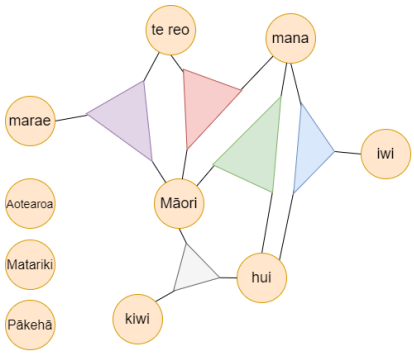
```
source,target,weight
aoraki,matariki,1
aoraki,puanga,1
aotearoa,aroha,1
aotearoa,hangi,2
aotearoa,harakeke,1
aotearoa,huia,2
aotearoa,iwi,2
aotearoa,kapa haka,2
```

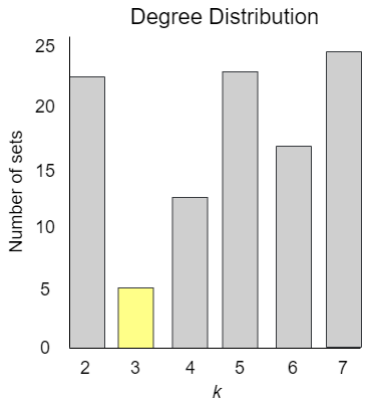
25

Demonstration: Creating a Network in Gephi

26

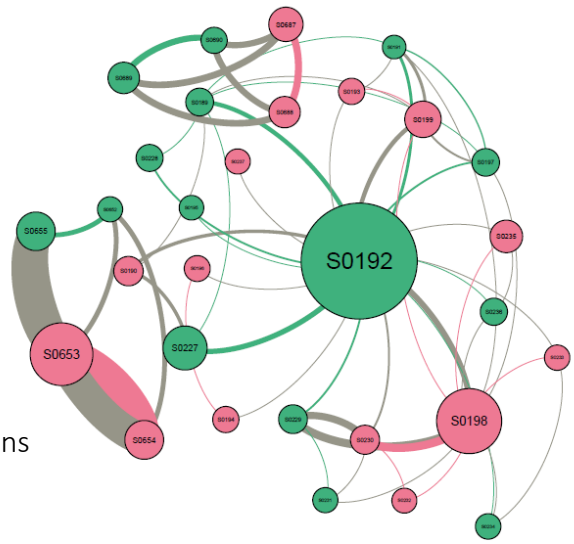
Beyond Gephi: *k*-Uniform Hypergraphs





27

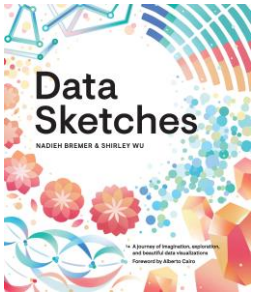
Network Example: Speakers in the BNC Corpus



Nodes = Speakers
Links = Shared conversations
Size = Total word count
Colour = Gender

28

Resources



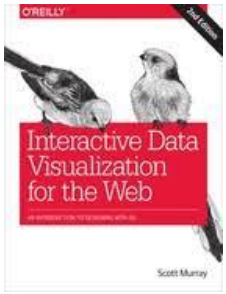
29

Observable & D3

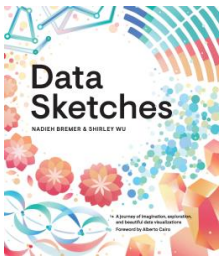
- Introduction: https://observablehq.com/@observablehq/a-taste-of-observable?collection=@observablehq/overview
- Twitter: https://twitter.com/observablehq
- Shirley Wu's Notebooks on SVG & D3: https://observablehq.com/@sxywu/introduction-to-svg-and-d3-js
- Mike Freeman's University of Washington Tutorials: https://observablehq.com/collection/@uw-info474/in-class
- Mike Bostock's D3 Gallery: https://observablehq.com/@d3/gallery
- Prototyping in D3 by Amelia Wattenberger: https://observablehq.com/@wattenberger/prototyping-in-d3

30

D3 (more generally)



- Scott Murray's *Interactive Data Visualization for the Web*: https://www.oreilly.com/library/view/interactive-data-visualization/9781449340223/
- *Data Sketches* by Nadieh Bremer & Shirley Wu: https://www.datasketch.es/
- D3 Graph Gallery: https://www.d3-graph-gallery.com/index.html
- Shirley Wu's YouTube videos: https://www.youtube.com/user/sxywu/videos



31

Miscellaneous

- HTML, CSS & JavaScript: https://observablehq.com/@jdev42092/week-1a-introduction-to-html-css-and-javascript
- Data Wrangling with Pandas: https://infovis.fh-potsdam.de/tutorials/infovis2data.html
- Vega-Lite and Altair: https://github.com/uwdata/visualization-curriculum
- Altair: https://altair-viz.github.io/user_guide/data.html
- NetworkX: https://programminghistorian.org/en/lessons/exploring-and-analyzing-network-data-with-python

32

References

- Bostock, M., Ogievetsky, V., & Heer, J. (2011). **D3: data-driven documents**. *IEEE transactions on visualization and computer graphics*, *17*(12), 2301-2309.
- 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).
- Bastian, M., Heymann, S., & Jacomy, M. (2009). **Gephi**: an open source software for exploring and manipulating networks. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 3, No. 1).

33

Thank you!

34