Topic 10

Text Visualisation and Ethics

The rationale for visualising text

Understand (also quick and informative)

- Summarise a large amount of text
- Make complex text data easy to understand

Group, Compare and Correlate

- For overview or classification
- Compare document collection or look at evolution overtime
- Discover hidden trends and patterns

Text?

Documents and Collection of documents

- Articles, books, emails, web reviews….etc
- Social Media networks, Comments…

Many text visualisations do not represent the text directly. They represent the output of a language model (word counts and sequence, etc.)

https://electionspeeches.moadoph.gov.au/speeches/2016-malcolm-turnbull

Example

Former Prime Minister Malcolm Turnbull's (2016) election speech

- What questions might you want to answer?
- What visualisation might help?

Tag Clouds: Word Count : Malcolm's speech

https://www.freewordcloudgenerator.com/generatewordcloud

Tag Clouds… (Heer n.d.)

Strengths

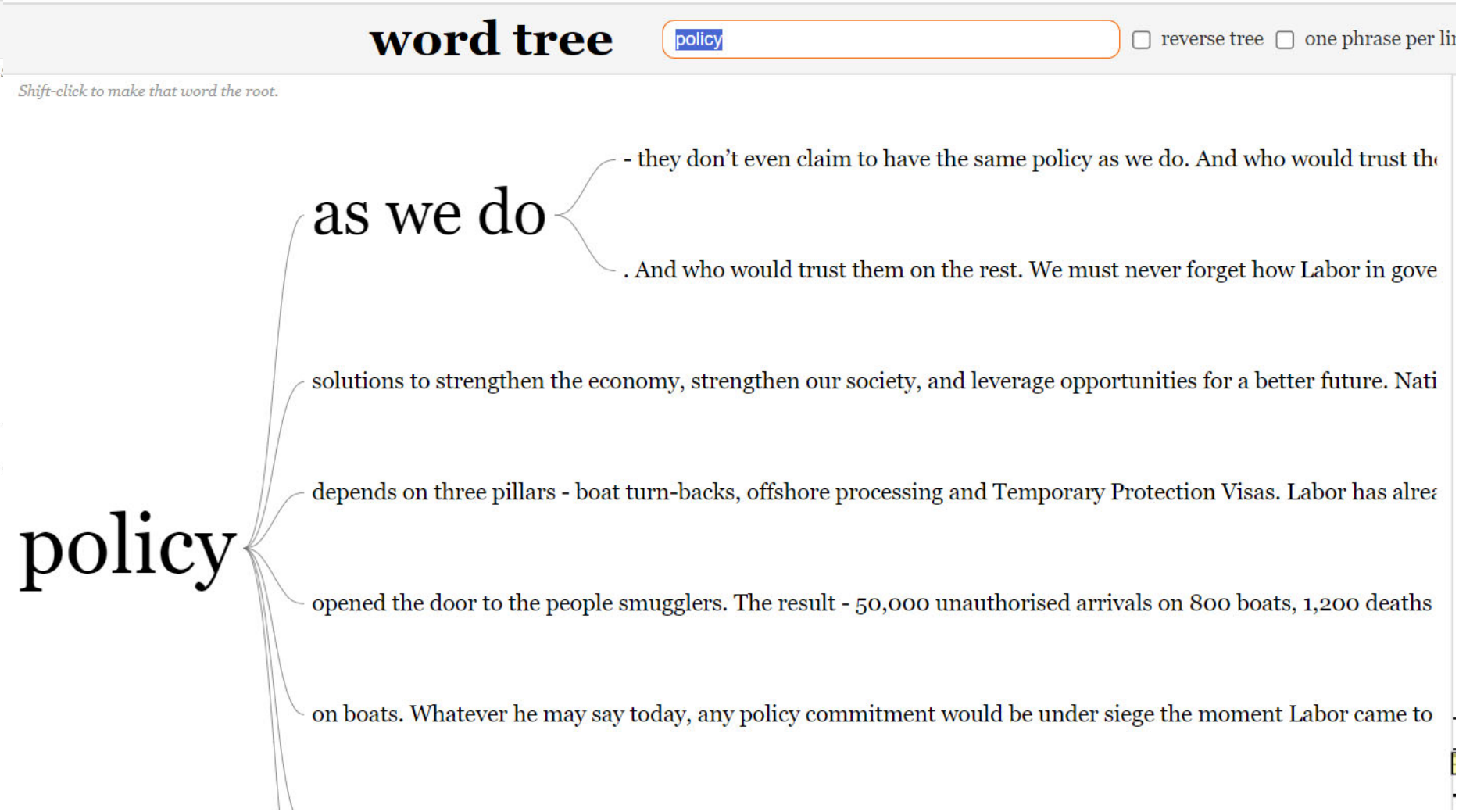
- Can help with inquiry and understanding

Weaknesses

- Sub-optimal visual encoding (size vs position)
- Inaccurate size encoding (long words are bigger)
- May not facilitate comparison (unstable layout)
- Does not show the structure of the text
- Term frequency may not be meaningful

Word tree: Word Sequence

https://www.jasondavies.com

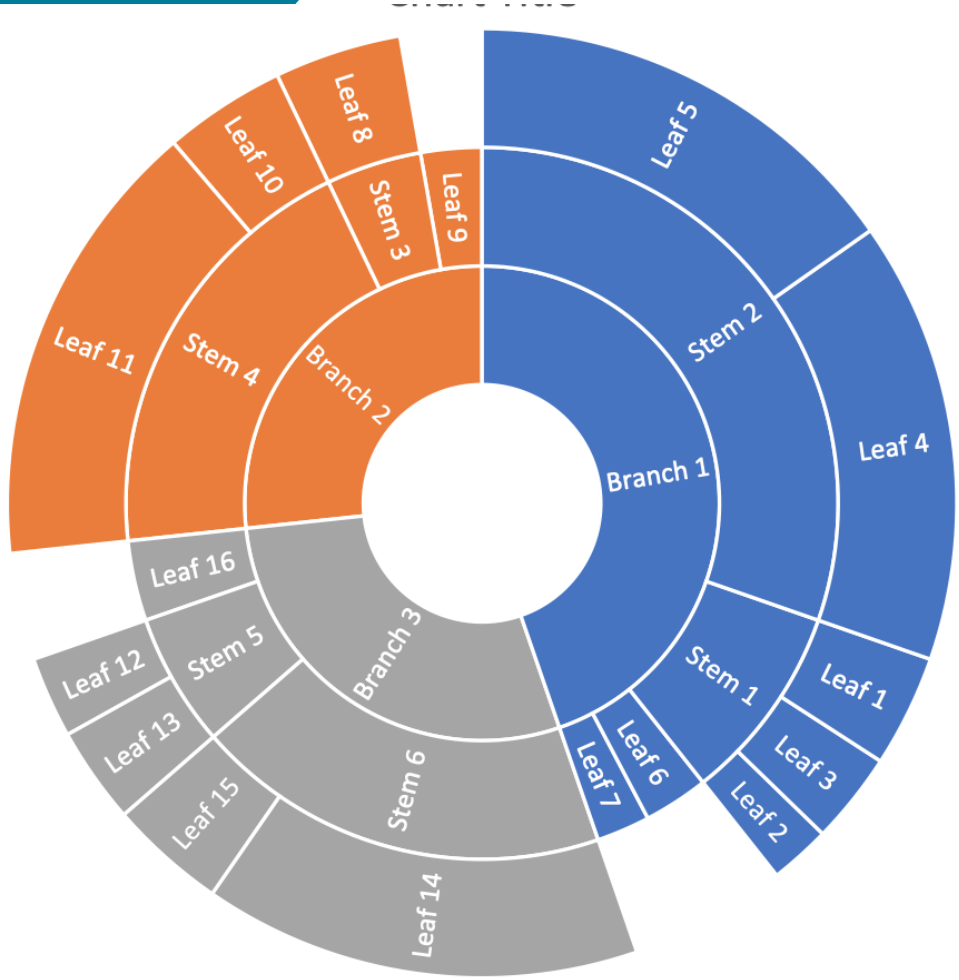


Text… (Heer n.d.)

Challenges

- High Dimensionality (Where possible, use text to represent text … which terms are the most descriptive?
- Context and Semantics (Provide relevant context to aid understating. Show the source of text)
- Modelling Abstraction (Determine your analysis task. Understand the abstraction of your language models. Match Analysis task appropriate tools and models)

Visual Thesaurus – Visual Maps



Concordance
What is the common local context of a term?

Search Entry

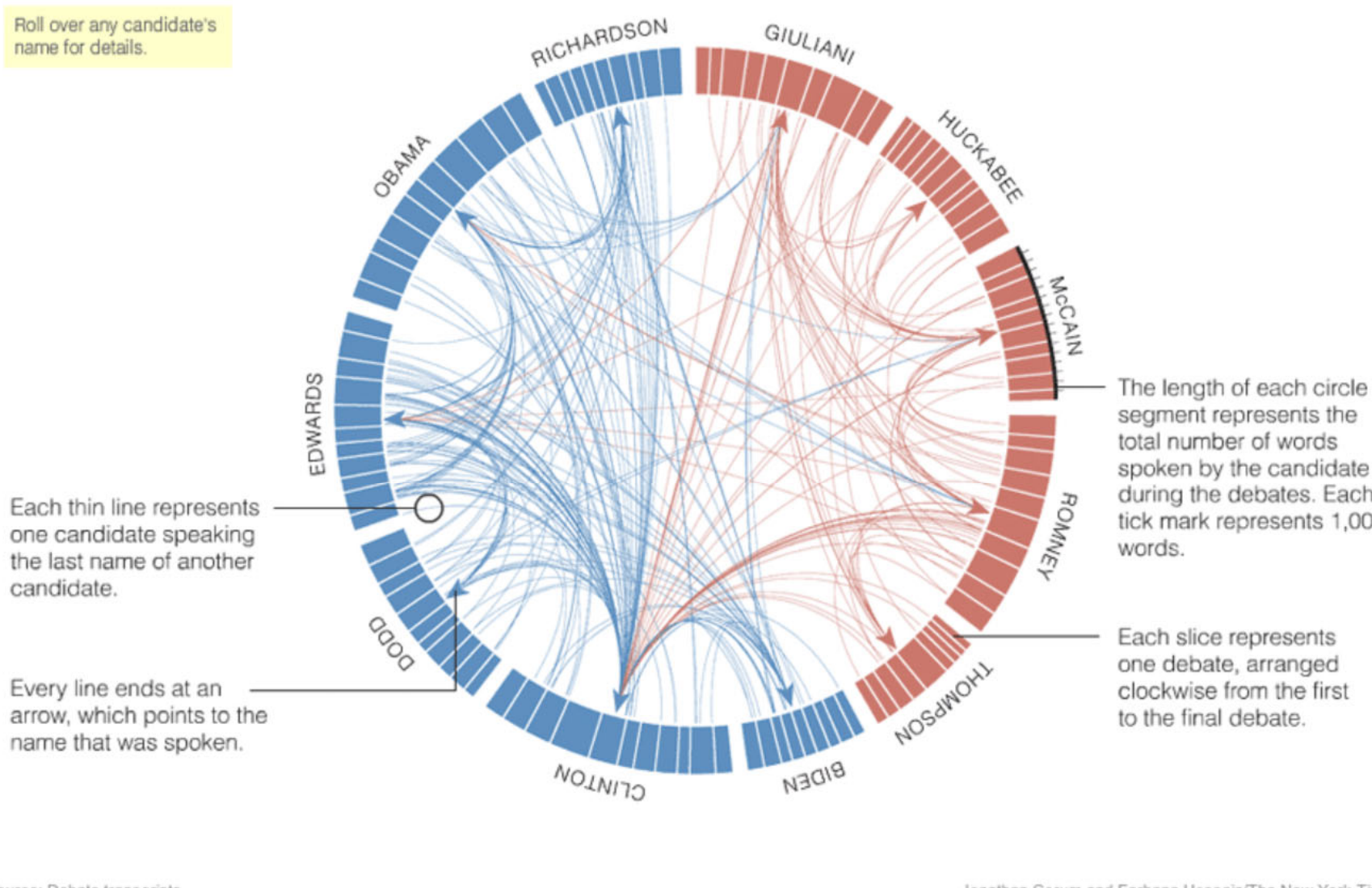
Word: Covid POS: Conj.: Additional Options Search

Sort 1: None Sort 2: None Sort 3: None (Retrieve LR 24 Words) Ready

Result

hich often , you know , it 's complicated because there was demand for some clients to help them with	COVID	and with the external environment , especially with technology related services , but in general , anythin
h I 'll talk about a little bit as well . So this is our managed service area predominantly . Okay . So how	covid	affected us ? I mean , initially there was the scramble obviously for to make sure that everyone was able
financial hit to us was not that great . I know in the wider NTT businesses during the first few months of	COVID	, particularly those that are providing infrastructure services to support just all kinds of networking capab
to make sure that their work forces were able to , to operate remotely . Since that time , obviously , as	covid	has really bitten into everything we see , not only within our organization , but also the wider organizatio
an ebb and flow . Much of the ebb and flow that we 've seen this year is I think less to do with	covid	itself . We 've seen some increased competition from a number , in a number of ways . So we 've been b
umber , in a number of ways . So we 've been buffeted by those winds . But not particularly because of	COVID	itself , ◇ Yeah , so really it affected government business . You know , it had a national impact and of
as all manual . You know , there were other services that , that were built and deployed just because of	COVID	and why we delivered information to people . So it 's kind of stream of things happening . The impact wa
federal government service delivery , broadly dipped the sentiment around service delivery dipped during	COVID	it really increased to pass previously . So you know , it 's probably at an all time high in terms of consur
tics , even though deaths in aged care were a big focus in the media . The deaths in aged care through	covid	are probably equal or possibly even less than the number of deaths , you would typically see an aged ca
d some staff in the beginning . So there was a 10 % head count reduction that was done at the start of	COVID	in anticipation of some challenges that we 're going to have and potentially some loss of revenue due to
in anticipation of some challenges that we 're going to have and potentially some loss of revenue due to	COVID	. ◇ Well you know , deacon has been always well-prepared for working online , working digitally , a big
ur , where delivery with students is of course , exclusively on working on top of digital platforms . What	COVID	has done to the business . Its a double effect on one side has expanded that digital only activity that we
, we have been discussing that many changes were required for the organization to adapt to the future .	COVID	has shown us that we needed to go there a lot quicker that is from , from the , from the deacon perspec
rse a big job for my team . But again , we were there already not at the the volume and dimension that	Covid	required but we were there already . So we , the thing is we did n't , we did n't need to change much
lenged , the operating model it challenged , the financial sustainability from , from the , I am discussing	COVID	of course only from the organizational perspective , many other consequences from , from the society p
rhaps we should talk I got the delivery responsibility , well , actually the digital response for government	COVID	response . So I was responsible for COVID safe . So and the coronavirus Australia app which has been
nsibility , well , actually the digital response for government COVID response . So I was responsible for	COVID	safe . So and the coronavirus Australia app which has been used to support you know , information to p
een responsible for all of those . ◇ We had to scale up there were and that was particularly around the	COVID	response because that was all new work . There were some pieces of work that we did have to , we bas
here was a large scaling up both for internal staff , contractors and companies to be able to deliver that	COVID	response . ◇ So one question I have there because Oracle is so big , what is the -LSB- inaudible -RSB
of care around our employees because there were different challenges with their health and safety in a	COVID	environment . ◇ Uh no , not really . I mean , it 's the nature of the work changed sometimes in
I mean , it 's the nature of the work changed sometimes in terms of , you know , if it was an immediate	COVID	response , but the usual processes and quality and risk really was I guess my observation was you just
some interesting experiences , but that was a big change . And I guess during the , during the time of	COVID	, we 've kind of also got better video conferencing and the processes that go with that and the etiquettes
things . So I work on my gov a lot , my work on digital identity scaled up . And I work on the	COVID	site response . So COVIDsafeAustralia.co.au and the COVIDsafe what 's up channel , all those things w
nce we were allowed to go back in . So different States have obviously at different levels of openness to	COVID	now . But in the beginning there was definitely you know , a very quick adjustment made to working fron
we 'd assign certain days so that we were able to maintain social distancing and keep them safe during	COVID	. So things like adjusting workshare , when people would engage with a piece of work , you know , the v
rying to work with the business , understanding what the requirements are from the government around	COVID	for a hospital environment and us trying to spin them up quickly for them . It took a while for , so we 'd
to spin them up quickly for them . It took a while for , so we 'd have to , Epworth put together a	COVID	task force team or governance group that took a beat for us to understand how we work with , with them
've had to manage them differently . Now the , the one that probably took the most impact was the non	COVID	related work . So this was the work that is just general government business type work . Because we er

Visual Conversations



▲ Names used by major presidential candidates in the series of Democratic and Republican debates leading up to the Iowa caucuses. By Jonathan Corum and Farhan Hossain (NYT).

Ethics

Ethical Challenges

- Visibility: To make the invisible visible (hard to balance comprehensibility of design and literacy of the audience)
- Privacy: To collect data with empathy (restricting the type of data used has implications for the impact and quality of the design)

Ethics

Consequentialist Ethics: An action is ethical or unethical based on its consequences. In utilitarianism – an action's ethics is based on *utility*.

So, what is the value of visualisation?

Issues with this approach:

- What do we measure?
- Who is doing the measuring?
- When are we done measuring?

"I'll give you a thousand words for it."

a measure for this. We simplify this by assuming that there is a homogeneous user community, consisting of n users which use a certain visualization V to visualize a data set m times each, where each session takes k explorative steps and time T. This is a crude simplification of course. In the real world, the user community will often be highly varied, with different K_0's and also with different aims. The costs associated with using V come at four different levels:

- $C_i(S_0)$: *Initial development costs*. The visualization method has to be developed and implemented, possibly new hardware has to be acquired.
- $C_u(S_0)$: *Initial costs per user*. The user has to spend time on selection and acquisition of V, understanding how to use it, and tailoring it to his particular needs.
- $C_s(S_0)$: *Initial costs per session*. Data have to be converted, and an initial specification of the visualization has to be made.
- C_e: *Perception and exploration costs*. The user has to spend time to watch the visualization and understand it, as well as in modification and tuning of the specification, thereby exploring the data set.

The total costs are now given by

$$C = C_i + nC_u + nmC_s + nmkC_e.$$

The return on these investments consists of the value $W(\Delta K)$ of the acquired knowledge $\Delta K = K(T) - K(0)$ per session, multiplied by the total number of sessions:

$$G = nmW(\Delta K)$$

and hence for the total profit $F = G - C$ we find

$$F = nm(W(\Delta K) - C_s - kC_e) - C_i - nC_u.$$

Ethics…

Deontological Ethics: An action is ethical or unethical based on whether or not it follows the *rules.*

Tufte's principles of graphical display?

Calculate a graph's Lie factor (divide the size of the effects shown in the graphic by the effect size in the data).

Issues with the Demonological approach:

- Where do the rules come from?
- Can the rules change?
- What about edge cases?

Ethics…

Virtue Ethics: Your actions are assessed by how much they cultivate your virtues. (Make the invisible visible, Collect data with empathy, and Challenge structures of power)…?

Issues with this approach:

- Where do the virtues come from?
- What do I do about conflict in virtues?
- What is never ok to do?

Ethics…

Role Ethics: Your actions are assessed by the extent to which you fulfil the obligations of your role.

Professional Ethics: ACS code of conduct

Issues with this approach:

- Who gives you your role?
- What happens if you have multiple roles?
- How big do our roles get?

What is a good visualisation?

Most cost-effective

Good design adherence

Virtue cultivation

Fulfilling designer duties

Self-evaluation

Can I identify ethical and unethical data visualisations?

Can I critique the moral character of the data visualisations?

Can I intervene or stop projects I think are unethical?

References

Cairo, Alberto. 2014. “Ethical infographics In data visualization, journalism meets engineering.” https://www.dropbox.com/s/pqgmg02yz0pgju4/EthicalInfographics.pdf.

https://mschermann.github.io/data_viz_reader/ethics.html

https://vega.github.io/vega-lite/

J. J. van Wijk, "The value of visualization," VIS 05. IEEE Visualization, 2005., Minneapolis, MN, USA, 2005, pp. 79-86, doi: 10.1109/VISUAL.2005.1532781.