



## Chapter 2

# The Importance of Visualisation: From Social Networks to the Atmosphere



*“If you really want to understand something, the best way is to try and explain it to someone else. That forces you to sort it out in your mind. ... By the time you’ve sorted out a complicated idea into little steps that even a stupid machine can deal with, you’ve learned something about it yourself.<sup>1</sup>”*

- Douglas Adams, *Dirk Gently’s Holistic Detective Agency*

---

<sup>1</sup>Omitted at ellipsis : “And the more slow and dim-witted your pupil, the more you have to break things down into more and more simple ideas. And that’s really the essence of programming.”



# Contents

<b>2 The Importance of Visualisation: From Social Networks to the Atmosphere</b>	<b>201</b>
2.1 Introduction . . . . .	207
2.1.1 Visualisation Design . . . . .	208
2.1.1.1 Storytelling . . . . .	208
2.1.1.2 Metaphor Selection . . . . .	209
2.1.2 Visualising Relationships . . . . .	215
2.1.2.1 The Dataset . . . . .	215
2.1.3 The social network . . . . .	216
2.1.4 Circular Edge Bundling . . . . .	218
2.1.4.1 Direct representation of the relational matrix . . . . .	219
2.1.4.2 The traditional network graph . . . . .	225
2.1.5 Summary . . . . .	228

## Background

In deciding on the best method in which to portray scientific information we must first look at how our ability to understand comprehend information has changed over time. This section looks at how an increase in neocortex size proves pivotal to the sudden rise to power sapiens experienced as a species, and in turn the development of language and visualisation in the context of information transfer.

In nature many animals rely on the propagation of DNA to encode important information essential to their survival. Examples of these are seen in hives (where an insect's role is defined by its genetic composition), or in Oscines (songbirds) which have an inherent predisposition to learn species specific songs, [De Smedt, 2013, Hunt and Gadau, 2017, Ackerman, 2016, Wada, 2010, Harari, 2015].

In sapiens, the vast and varied nature of the information we are required to process makes this process highly impractical. Instead we develop a predisposition to learning language at an early age. Such a skill allows for the effective communication of ideas, conditions and dangers between a large number of people<sup>2</sup>.

However since communicatory patterns are limited to only the people they have been taught to, problems of differing language and dialect greatly reduce the amount of information which may be passed between groups or tribes. One way to overcome falls in the use of pictographs, a form of visualisation colloquially known as cave paintings (e.g. Figure 2.1) which complements the sapien ability to both detect shapes and spot patterns within nature<sup>3</sup>. As communities increased in size, problems of accounting and resource management begin to emerge.

*REWRITE: Until this point an ability to store large amounts of numerical data had not been required by a hunter-gatherer species.* Around 3,500BC, the Samaritans solved this probelm with the creation of a system for coordinating affairs and storing information external to peoples brains.

This system is that of writing [Nissen et al., 1993, Schmandt-Besserat, 1992]. Here quantities and items are depicted using a numerical system of signs and visual system of shapes (cuneiform<sup>4</sup>). This external method of storing information, coupled with visual representation, allowed for a an effective and intuitive way for us to apply the pattern recognition and analytical parts of our brain whilst reducing the cognitive load by breaking up the problem into managable parts.



Figure 2.1: **An example pictograph:** a 2000 year old petroglyth in Utah titled ‘Tse Hane’ (rock that tells a story). Source: [ugc, 2019]

Throughout history, we have continued to apply this system of intertwining data information with visual artifacts to enable people to cope with the complexities of the infomation provided, [Tufte, 1983]. In the remainder of this chapter the use of visualisation as a means of enhancing the readers ability

---

<sup>2</sup>Several studies, exploring the ratio of the neocortex to the rest of the brain, suggest that the number of relationships a human can successfully monitor is limited to 150. It is suggested that ideas of gossip and common metaphysical beliefs are the reason for this [Harari, 2015, Aiello and Dunbar, 1993, Dunbar, 1997]. This limit is still seen in social networks today [Hill and Dunbar, 2003].

<sup>3</sup>It has been found that 10,000-year-old pictographs show hints of a common cultural background between spatially different groups of humans [M. Chazine, 2005].

<sup>4</sup>This is often mistaken for hieroglyphics. Although both are forms of logographic script, hieroglyphs are restricted to the ancient Egyptian sociolinguistic context.

to understand the large-scale complexities of the chemistry within the troposphere shall be explored.

**A note on terminology** Although the word *species* refers to the biological definition within this section, any further reference throughout this thesis refers to the chemical definition instead.

## 2.1 Introduction

The term ‘Big Data’ originated in the mid-1990s, appearing in several job adverts and a slide deck by Jon Mashey the chief scientist at Silicon Graphics (SGI), [Mashey, 1998, Diebold, 2012]. This term is used as a way to describe the ever-increasing amount we can generate each day. Such phenomena are seen in anything from the growing number of websites to the number of reactions within a near explicit gas-phase chemistry mechanism, Figure 2.2. Coupled with the ability to collect large amounts of data is our requirement to analyse and understand it. This has commonly be done through the use of visualisation, a topic in which careful consideration must be made to ensure the correct information is conveyed [Kirk, 2016]. In a paper on mining big data, [Fan and Bifet, 2013] explains that the main task of Big Data analysis is on deciding how to visualise the results- simply its size introduces complexity in uncovering a user-friendly method to represent the information required.

Although graphical representation has been an integral part of the data comprehension process, it is only relatively recently (1990’s) that it is recognised as a research field [Wybrow et al., 2014]. Even though we are not explicitly dealing with ‘big’ data, the number of species and reactions within the troposphere are still sufficiently large such that many of the same problems still exist. It is for this reason that this section<sup>5</sup>, will explore topics such as considerations that need to be made before selecting a visualisation (subsection 2.1.1) and the methods we can use to represent the complex relationships within the atmospheric chemistry domain (subsection 2.1.2).

---

<sup>5</sup>and consequently much of this thesis

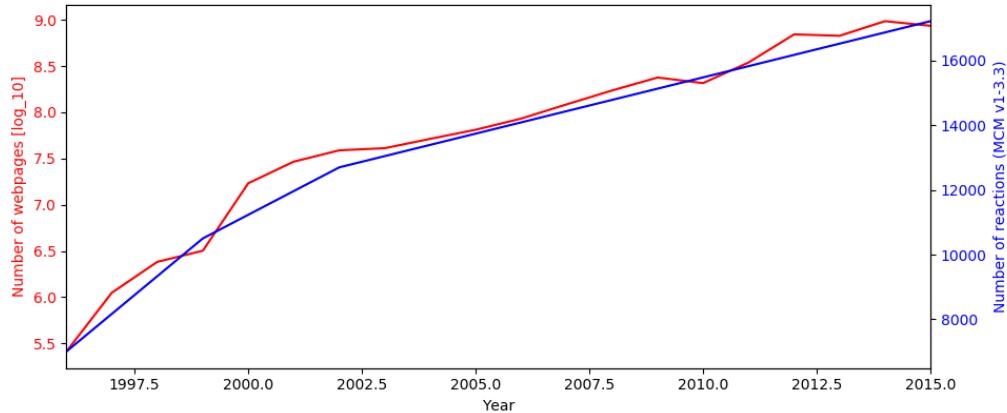


Figure 2.2: **Growing trends in data.** This figure shows the change in the number of websites ( $\log_{10}$  normalised) between 1996 and 2015 in red, and the changes of species for the number of reactions within the master chemical mechanism for versions 1 to 3.3.1 - This is introduced later on. *Note: The two lines are here to illustrate the growing trends between the subjects and are not incomparable due to their differing scales.* Source: [InternetLiveStats, 2020, Jenkins, 2002]

### 2.1.1 Visualisation Design

New ideas are developed and refined through an iterative process of cognition and discussion with other researchers[Roberts et al., 2014]. As we are constrained by our experience and knowledge, novel ideas often consist of an amalgamation of many existing concepts [Descartes et al., 1996, Johnson, 2010]. [Ziemkiewicz and Kosara, 2008] suggests that the process of understanding a visualisation depends heavily on the interaction between the user's internal knowledge and experience with the images and ideas depicted within the visualisation. This means that the careful selection of content and medium (of presentation) can directly influence what a reader takes away from the graphic.

Knowledge and experience dictate how we may perceive new data. This means that a data visualisation designed to be relate-able in content and representation to a reader will have a much higher chance in conveying the correct information successfully. As a consequence of this, the interface between quantitative science and the user (i.e. the visualisation) will have a higher knowledge transfer or throughput. Here two visualisation design considerations (storytelling and metaphor selection) are discussed.

#### 2.1.1.1 Storytelling

Storytelling is commonly used to highlight the process of cause and effect. It has applications in the education for both the explanation of scientific concepts ([Martin and Miller, 1988]) and the dangers of the world (fables such as Little Red Riding Hood and Goldilocks teach the dangers of speaking to strangers and the importance of respecting personal property). As human life is subject to the

conditions of the ‘arrow of time’, the inherent familiarity of linearly consequential events make the narrative a great way to inform the reader of new (unseen) concepts. Such methods are not limited to the education of young children, but also allow for the understanding of real-world events through the use of dreams, media and the news [Gottschall, 2012, Freud and Cronin, 2013].

With the use of a narrative, a visualisation can educate the user, providing them with a level of prerequisite knowledge that then enables them to draw their own (guided) conclusion. In exploring how people interpret visualisation, it is seen that readers apply a visual routine to create a narrative and compare values within a graph, [Michal and Franconeri, 2017]. This emphasizes the importance of designing a visualisation in such a way that the user is ‘guided’ through the correct narrative. The question-answer cause-effect structure not only satisfies the inquisitive nature of the human species, but it also educates the reader to the events which led to a certain conclusion. It is this process that makes ‘storytelling’ an effective method of communication to a non-expert audience [Dahlstrom, 2014].

### 2.1.1.2 Metaphor Selection

In addition to storytelling, the use of suitable metaphor is also important in designing a visualisation capable of appealing to the users’ knowledge and experience. Much like the use of narrative, metaphors have infiltrated our everyday lives. These range from the morals presented by fables, to the concept of money and belief to govern everyday life through the inter-subjective<sup>6</sup> [Harari, 2015].

In selecting a suitable metaphor for a visualisation we often have three options. These are natural, man-made and composite.

#### Nature-Inspired

Inspiration for a metaphor can come from a range of sources. The most effective of which, often have objects of events which are universal and inherently familiar to all readers. The most widely used of these are nature-inspired metaphors. Such metaphors do not only guarantee a basic level of item comprehension but also have an aesthetically pleasing quality and familiarity to the visualisation components.

Examples of these may be the use of ice and water to symbolise glacial melting, using trees to represent the binary decision tree of a random Forrest (Figure 2.3a), or rings within a trunk cross-section (Figure 2.3b) to show changes in conditions over time. The main benefit of using natural metaphors is the low learning curve required to understand them.

---

<sup>6</sup>The inter-subjective is something that exists within the communication network. It allows a fictional idea, such as a limited-liability company, to exist as a real physical entity with a bank account and subject to laws.

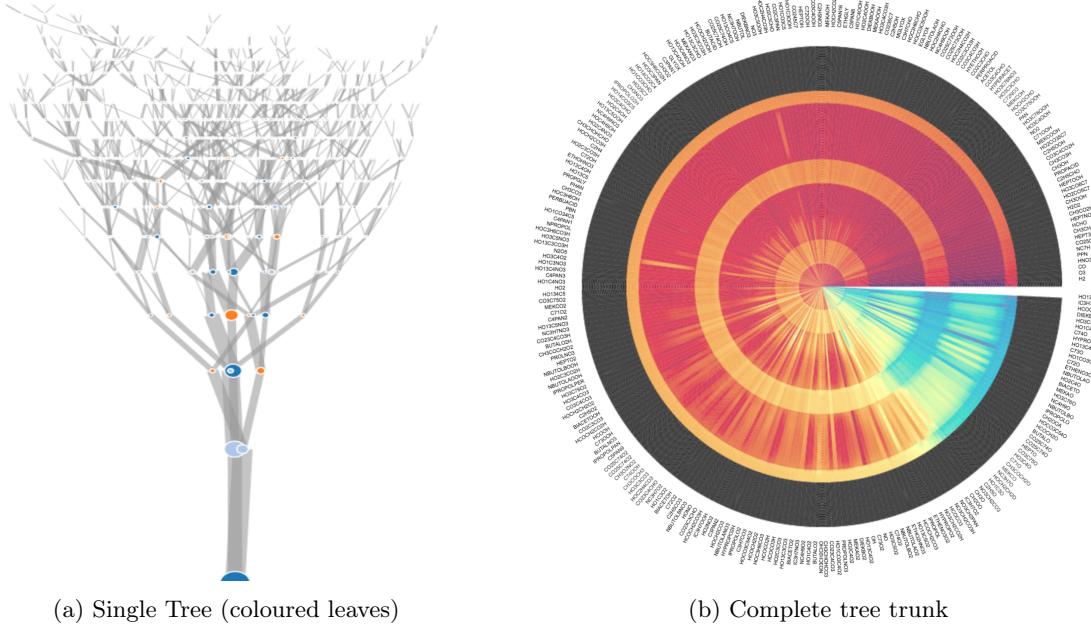


Figure 2.3: **A collection of tree-based visualisations.** (a) shows the decisions made on a tree within a random Forrest. Node/Leaf colour represents the property which is being decided, whilst node and branch size are symbolic of its importance (Larger leaves/branches are naturally stronger and more important).

(b) is a radial plot in the shape of a tree trunk. Much like that of a real tree, time is shown across the radial direction, starting at the origin. The colours represent the net flux of a series of species from a chemical simulation. Much like in a tree, large scale events produce a change in the ring colour (these are seen in avalanches and tsunami struck forests) - in this case, the abrupt changes are due to the diurnal cycle.

### Man-made

Similar to nature, another similarity between many readers would be their familiarity with the urban environment. Metaphor inspiration from man-made objects such as objects, buildings often contain characteristics of symmetry and manual/mechanical design. These allow the user to interpret any features presented using their pre-existing knowledge about an object. The most famous example of this is Henry Beck's tube map (1933) [Beck, 2017] where stations are positioned at 0,90 and 45-degree angles. This provides a clearer representation, much like a road map than the conventional space-specific location of each station. Although other designs, such as a concentric one, [Fisher, 2013], have been attempted, adaptations of Beck's design are still being used in the present day due to their intuitive nature.



Figure 2.4: **The classic tube map design.** Source: [Beck, 2017]

## Composite

Sometimes ideas may be combined to provide a composite metaphor. The classic example of this is seen in much of greek mythology with the concept of Pegasus. Here although we are familiar with the concepts of both wings and a horse, no one has encountered such a creature - yet we can assume the implications of its existence. Here the inspiration of new designs often stems from many different sources which are confluent and refined through the use of sketches and doodles [Roberts, 2011].

Similarly, it is possible to drive inspiration from a range of sources: natural, anthropogenic, other visualisations etc. Such techniques allow for the creation of xenographics, with the most famous example falling under the combination of a ‘pizza/pie chart’<sup>7</sup> with a bar chart to create a radial plot in Figure 2.5 - A novel representation which helped establish the methods of modern-day nursing by Florence Nightingale. Overcoming the problems presented by these composite designs involves a level of lateral thinking to produce a simpler, more effective, user-visual metaphor interaction [Ziemkiewicz and Kosara, 2008].

<sup>7</sup>There is a good amount of literature that suggests these are far from optimal methods of representing data.

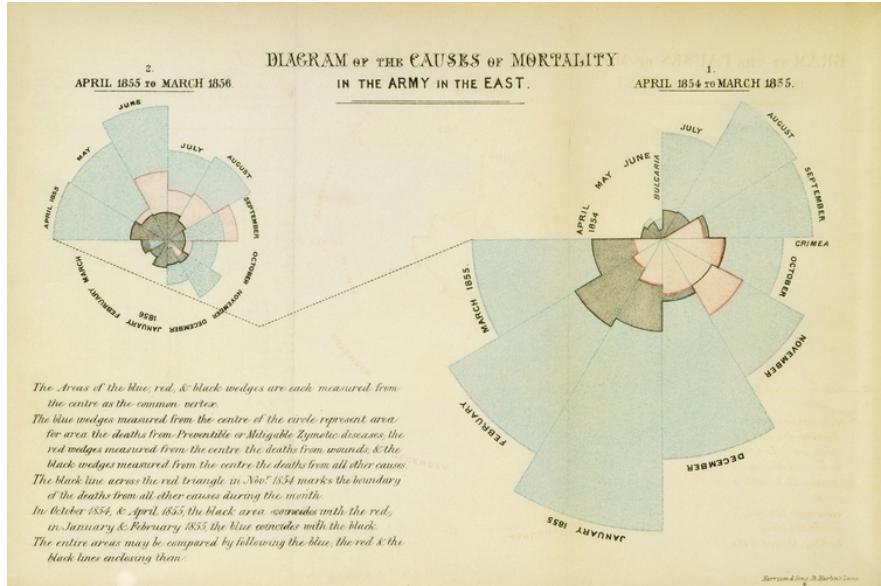


Figure 2.5: A (stacked) Radial bar-chart showing the causes of mortality in the British army<sup>8</sup>, [Nightingale, 1820].

### Domain Specific

Composite designs are not constrained to the combination of different metaphors. In specialised roles, it is common to find visualisations which draw on pre-existing knowledge either about the content of the figure. Although this involves an additional learning curve before information about the topic can be extracted, once this is obtained, the wealth and complexity that is portrayed by a single visualisation drastically increases. Two of the more common examples where prior knowledge is required to read a plot are connected scatter plots<sup>9</sup> and Tephigrams.

**T- $\phi$ -grams (Tephigrams)** are defined from their axis of temperature ( $T$ ), entropy( $\phi$ ) are used in the field of meteorology and weather forecasting. The combine grid lines for constant temperature and pressure, as well as allowing the user to calculate the equivalent potential temperature for saturated air parcels and the saturation mixing ratio concerning plane water surface, Figure 2.6b. This provides a good example where both scientific knowledge and ability to read the diagram are required to obtain meaningful information from it.

<sup>8</sup>Also known as the Nightingale Rose or Coxcomb plot.

<sup>9</sup>Occasionally known as snail trail chart (R users seem to like re-inventing the wheel).

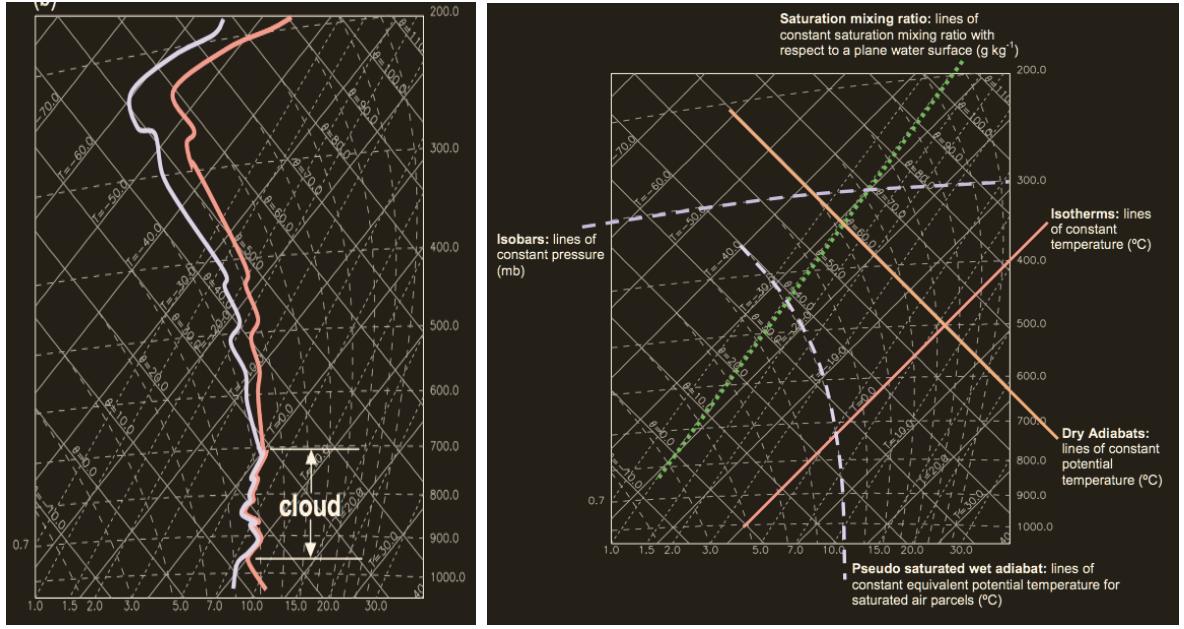


Figure 2.6: **Two T- $\phi$ -gram plots.** An example of the T- $\phi$ -gram data (a), and instructions on how to interpret it (b). Source: [Brooks, 2019]

**Connected scatter-plots** are useful in representing the temporal changes of correlation between two items. These are particularly useful in the field of economics due to their ability to highlight the different trends that can occur over time. Figure 2.7 shows how the number of vehicle-related fatalities changes with the number miles driven. In this situation, a simple  $x - y$  plot may highlight a decreasing inverse relationship between the number of fatalities and the miles driven per capita over time, but lose some of the many features presented by the additional dimension.

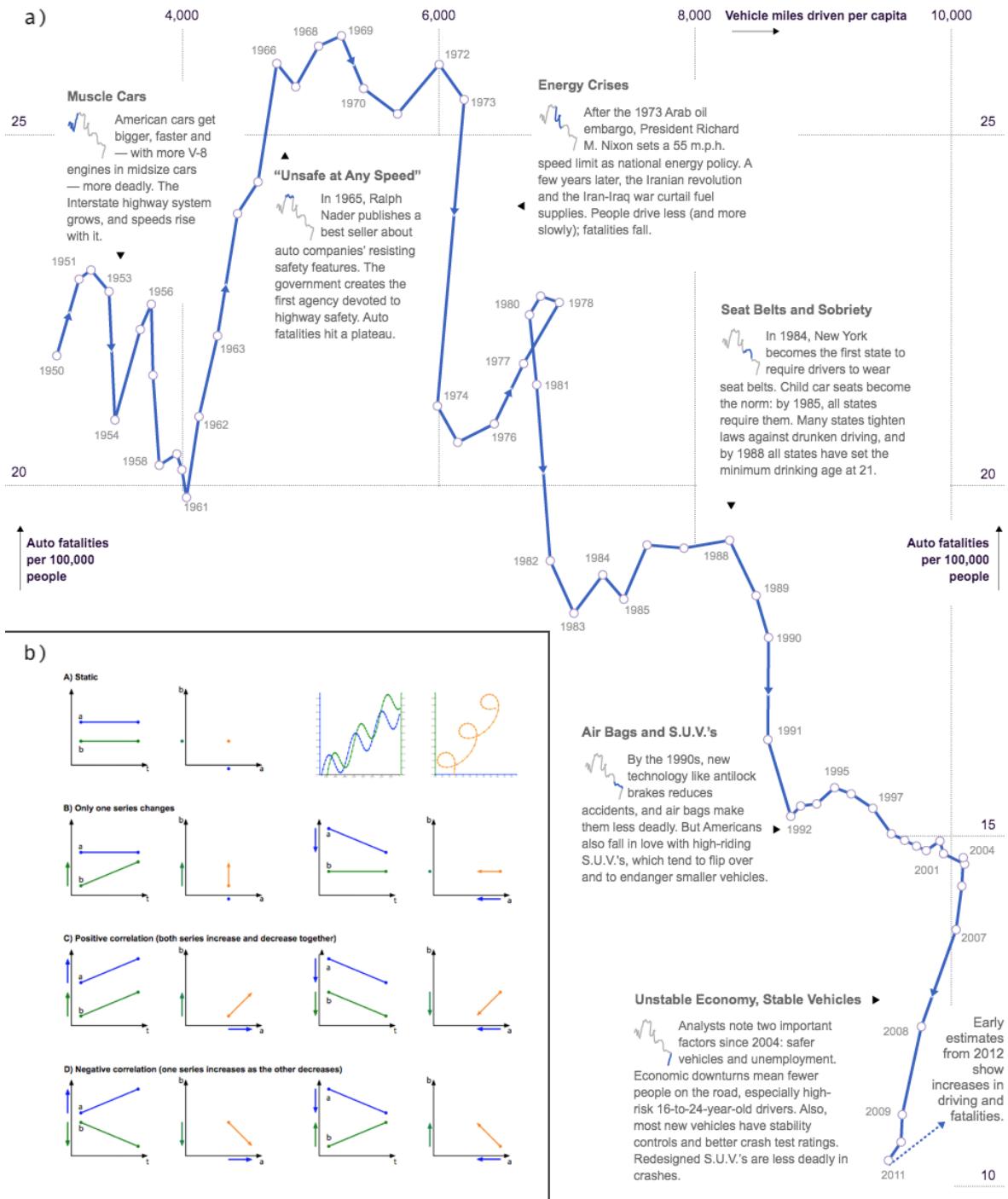


Figure 2.7: Examples of connected scatterplots (a) A connected scatter-plot comparing the the number of vehicle related fatalities with the number miles driven per capita [Fairfield, 2012] (b) a plot showing how the corellation of two spatial variables are transformed from a classical scatter/line plot into a connected scatter-plot [Haroz et al., 2016].

### 2.1.2 Visualising Relationships

Discerning important relationships is a vital part of both science and pattern recognition. Even in storytelling (subsubsection 2.1.1.1), we see readers trying to piece together a relationship between events, actions and outcomes. It is for this reason that we shall explore the different ways in which we may visually represent relationships within science. To do this we first define a dataset (subsubsection 2.1.2.1) - in this case, this is the master chemical mechanism (described below). We then proceed to identify a variety of different visualisation methods, before expanding on a particular style of visual analytics in ??.

#### 2.1.2.1 The Dataset

Since our topic of interest lies in the representation of chemistry within the atmosphere, we take a set of first-order ordinary differential equations and rates designed to describe how species react in the troposphere. This set of equations is known as a mechanism.

The equations we are to use are given by the master chemical mechanism (MCM). This is a near explicit representation of our understanding of the gas-phase chemistry within the troposphere. In its mathematical form, it describes how species are related and at what rate they react. As the aim of this section is to identify features within the chemical mechanism we begin by looking at the mechanisms construction protocol, Figure 2.8. In its design, this aims to mimic much of the reasoning and procedures that would be otherwise performed by an analytic chemist. The procedure is semi-automated and follows several iterations starting from a set of primary emitted VOCs. These are run from the protocol to build a set of reactions in which the selected species may undergo. Any new products are then introduced to the procedure until everything has been oxidised to produced carbon dioxide (via CO) and water. This produces the set of near explicit equations which are then used to describe the evolution of chemistry within an atmospheric model.

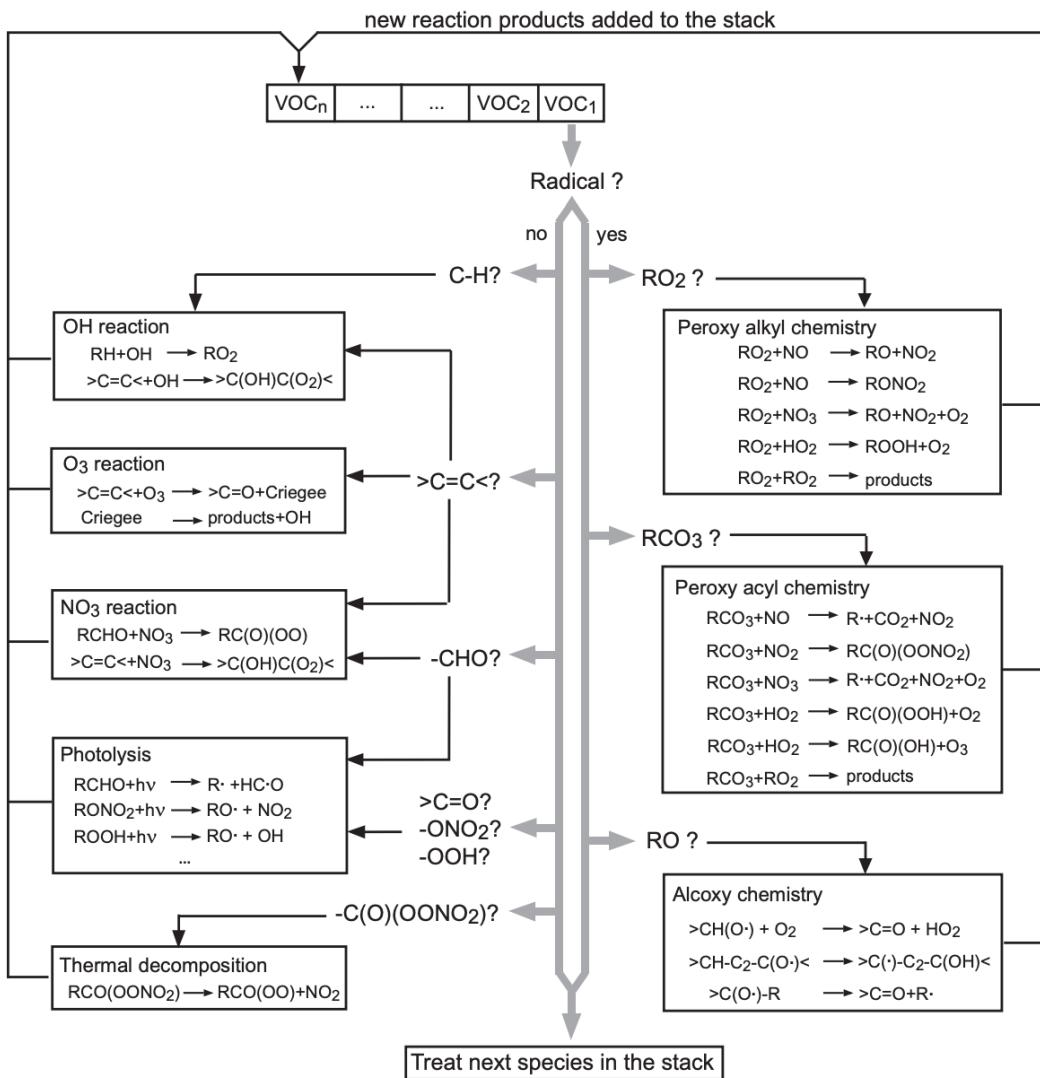


Figure 2.8: **The generation flowchart used within the MCM.** This shows the process undergone by a new species to generate its products. Any unseen products are then fed back into the flowchart until the entire mechanism has been produced. Source:[Aumont et al., 2005]

### 2.1.3 The social network

Social network analysis is a type of sociological work (sociometry) that aims to reveal the interweaving and interlocking relations between items or individuals, [Scott, 1988]. [Dunbar, 1997] argues that the evolution of language as a result of social grooming (gossip), and therefore its adaption for storytelling. It follows that using the social network construct to represent the latent patterns between objects makes use of both our inherent ability to discern complex patterns through the use of storytelling. In trying to depict the relationships between a social network, we employ the use of sociograms. These are a class of visualisations which reveal certain properties of a social network. The following section will look at the extraction of useful information from the relationships within the MCM through the use of several sociograms.

## The Chord Diagram

The chord diagram is a visual sociogram known for its use in summarising the overarching relations within a dense social network, [Jalali, 2016]. Here arcs are used to represent groups (or a node from a social network graph), and their length corresponds to the percentage of items they contain. Within Figure 2.9 we represent the different routes a species may react within the MCM protocol flowchart in Figure 2.8. The figure contains two sets of arcs. These represent the different level of classification of the chemistry - The first (outer) arc represents the split in channels between radical (red) and non-radical (orange) species which further separated into the finer categories within each branch (the labelled inner arcs). The inner arcs again show the probability a randomly chosen reaction on a species will fall under a certain category - if we convert an arc into a segment we have a pie chart of probability.

*It is worth noting that segment sizes do not represent the number of species undergoing a certain reaction pathway, but rather the percentage of all possible pathways which follow that route. This is because species often undergo a range of reactions, each of which counts as an individual weighting. It is for this reason that even though almost all<sup>10</sup> contain a C-H bond, does not consume the whole graph.*

From this, we see that hydroxy reactions are the most common with C-H bonds being in abundance<sup>11</sup>. We also see that having another type of reaction is also just as probable, with a third of the most utilised branches within the MCM protocol falling under species containing the Carbonyl group.

Next, we look at the co-occurrence of branches for different species. These are represented using the area of a circle connecting two arcs (a chord). Each chord has two edges connecting two arcs<sup>12</sup>. Comparing the width of each chord to its parent arc, it is possible to discern the percentage of items going between these and other branches. Here, for example, we see a roughly even split between species with a hydrogen bond (i.e. all species) and every other group. This suggests an even distribution of reaction types between species. This means that in comparing the arc length of each chord, we can visually determine the percentage of group A which relates to its partner group B. Finally it is also possible to determine the number of items in a group which contain themselves. Chemically these are species with multiples of one functional group that undergo a certain reaction pathway more than one time. Although these reactions will be combined within a mechanism (to avoid duplication), their rate would be increased accordingly.

<sup>10</sup>Except any inorganic species.

<sup>11</sup>This is seen within the graph layout ??

<sup>12</sup>With the exception of self loops, although these are addressed below.

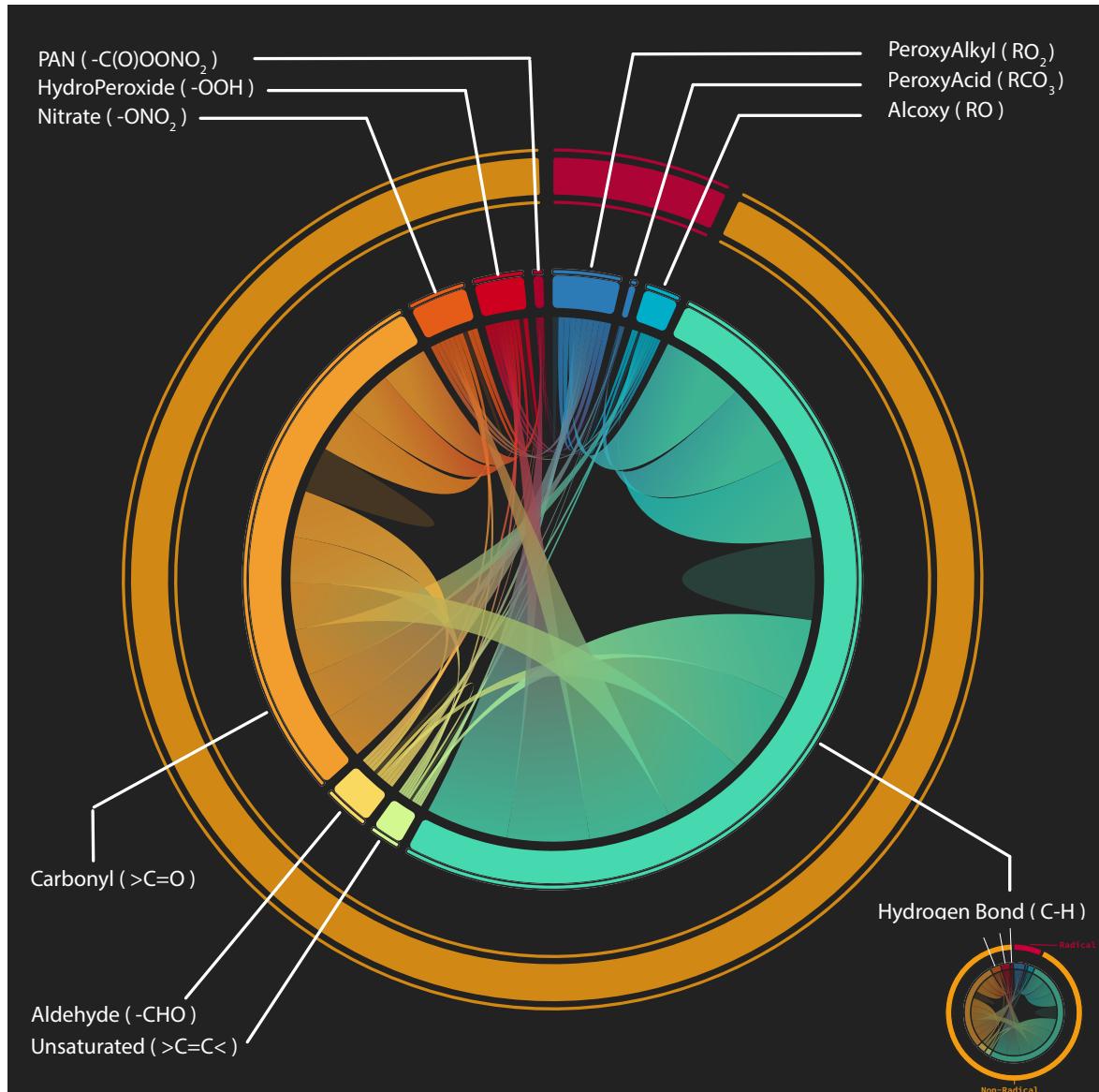


Figure 2.9: **A chord diagram.** This shows the proportional probability a species from the MCM will follow one or more of the paths presented in Figure 2.8.

The chord layout provides an easy way to calculate the percentage of items which contain multiple properties. It requires a relatively low learning curve and is intuitive to those with experience using pie charts (namely the Microsoft Office generation), however, this radial format can sometimes also make it more difficult to read. Special attention needs to be paid to the order arcs are drawn, as with some datasets, certain configurations may obfuscate trends. Finally, the chord diagram requires a certain amount of data munging before its employment. It is due to this that finer details about the system may be lost, especially if there is a coarse scale between chord sizes.

#### 2.1.4 Circular Edge Bundling

Show relationships (see zip on slack)

### 2.1.4.1 Direct representation of the relational matrix

Relational matrices, also known as the adjacency matrix, are a  $n \times n$  square matrix highlighting the relationships between items. Figure 2.10 shows the adjacency matrix representing the co-occurrence of characters in the Les-Miserables musical, as provided by [Knuth, 2019]. Through ordering, we can achieve a view into the structure of the overall network. Visualisations, such as those in Figure 2.10b are often used in the assessment and evaluation of network clustering algorithms since they clearly show the effect modularity has on the diagonal of the adjacency matrix.

The main downside of this type of visualisation is the sparsity of many graphs. If we take the complete MCM 3.3.1 network, as used above, and visualise it in this manner it will have a density of  $5.9^{-4}$ . This means that for an average  $600 \times 600$  pixel figure, only  $0.36^2$  of a pixel would be coloured in. And since we are not able to colour only parts of pixels our final plot will remain blank. Methods to circumvent this involve looking at subgraphs, or individual sections interactively or applying composite graph/adjacency techniques such as those presented in NodeTrix, a method for visualising sparse small world networks<sup>13</sup> [Agarwal et al., 2017]

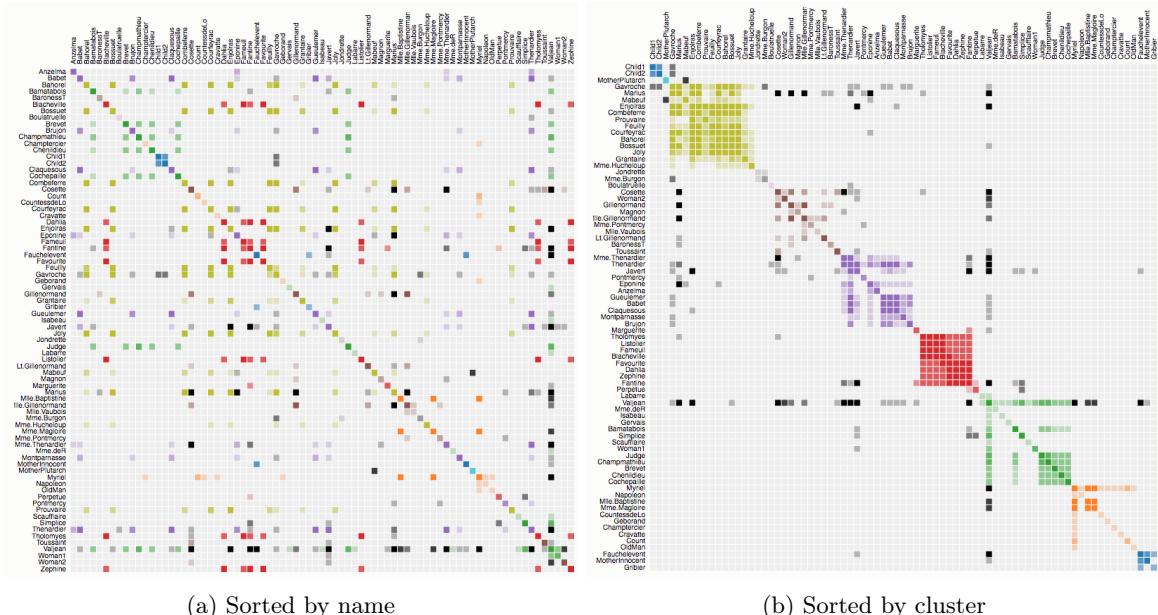


Figure 2.10: **The adjacency or relational matrix of the Les Misérables graph.** These show how changing the order of elements within a matrix can aid in the visual identification of clusters in the data. Squares features along the diagonal indicate a larger number of links between grouped items in the axis order, and less to those in other modules. Source: [Bostock, 2019]

<sup>13</sup>Small world networks are discussed in the following chapter.

## Arc Diagrams

Arc diagrams are a subset of sociographs where items are represented as nodes along the horizontal. Relationships between nodes are then shown through the use of curved links (arcs). It is this that makes them particularly suited for the highlighting of repetition between music or DNA sequences, [Wattenberg, 2002]. Using the data as above I use an arc diagram to explore how species containing the same combination of functional groups react. This is done by first determining all branch permutations from the protocol flowchart (Figure 2.8). This produces [179??????] groups which are positioned across the  $x$  axis in ascending group size (the more branches matched, the further to the right a group is positioned). The width of each group is determined from the number of species within it. Next links are added between the species based on what groups the product species from each reaction contain. Figure 2.11 discretizes reactions which produce products with an increased number of reaction pathways (positive arcs - top), from those which result in species with less (positive arcs - below). Here the cyclic nature of tropospheric chemistry can be seen, with many species producing larger, less stable, products, which then go on to react and dissociate back into smaller ones. In some cases a complete circle between two nodes can be indicative of a catalytic reaction. Using interaction and selective shading, it is possible to isolate certain types of reactions and determine which of the functional groups are responsible for the change experienced within a reaction.

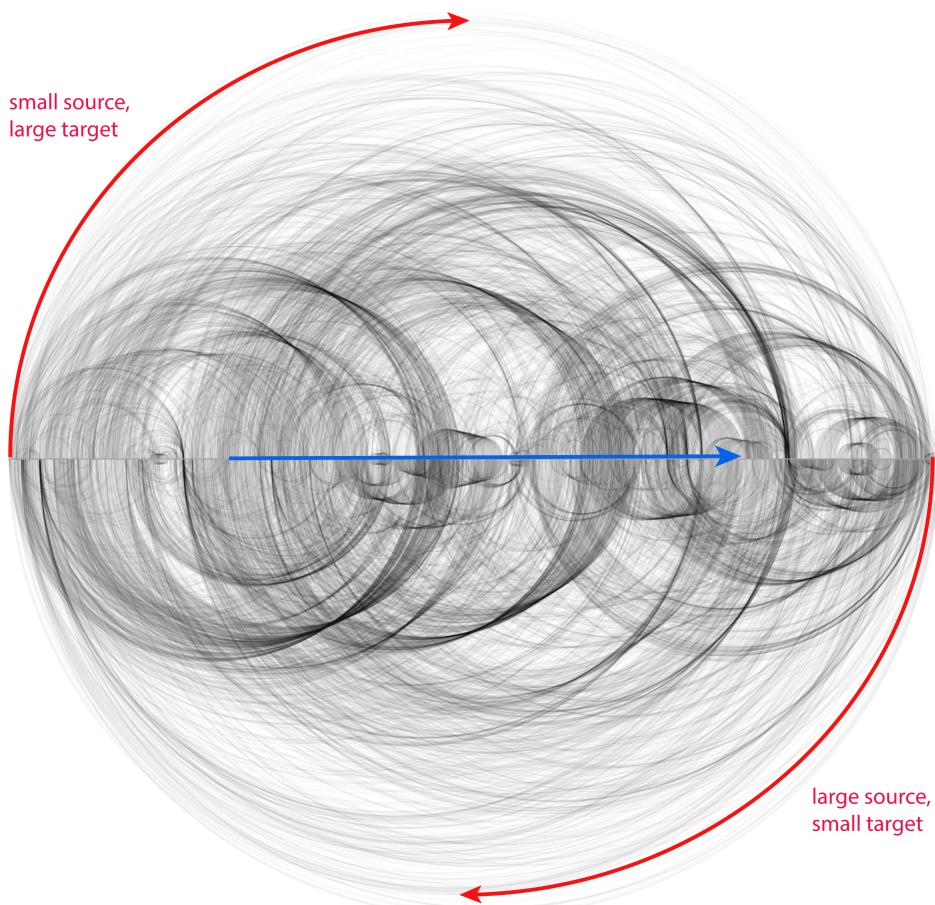
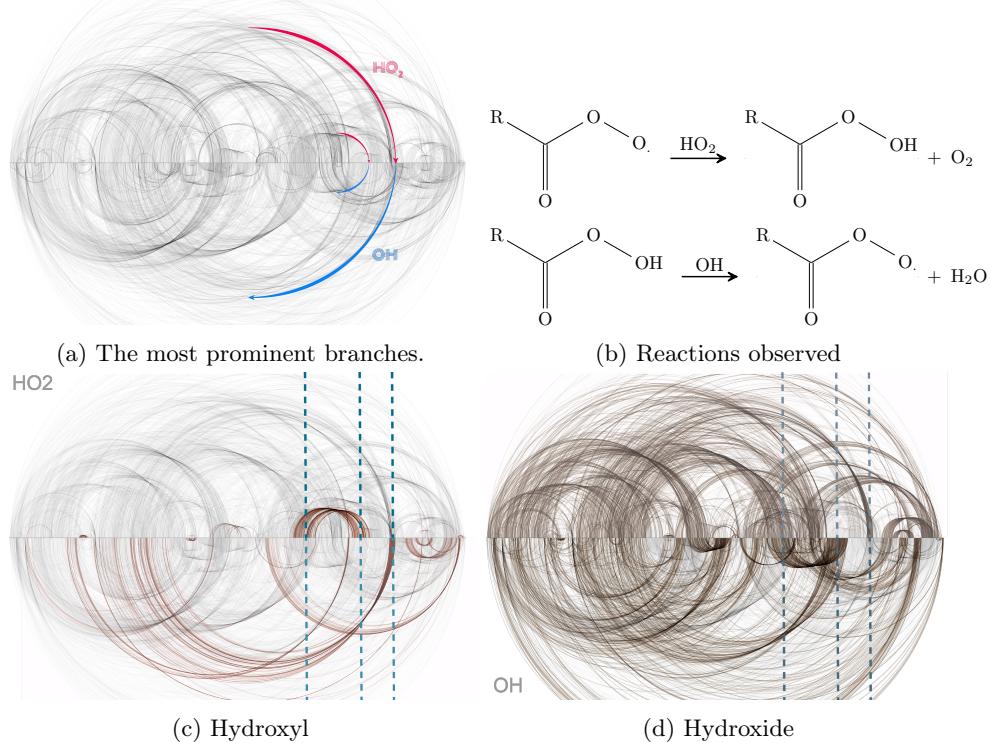


Figure 2.11: **An arc diagram constructed using the construction pathways presented in Figure 2.8.** Possible pathways for each species are extracted. These are then grouped and sorted by the number of functional group/pathways available. Species are shown along the blue line and sorted with the most multifunctional groups to the right. Reaction pairs (reactant-product) are then depicted through the use of arcs. The direction of these is shown using the location of the arcs.

Although such a layout may seem daunting at first, with many lines, in all directions, filtering by type of reactions can draw attention to several features from the chemistry. In REF CHAPTER/FIG the importance of  $HO_x$  cycle for the removal of VOCs and greenhouse gasses is shown. For this reason, I decided to explore the relationship between reactions with OH and  $HO_2$ , ???. These are represented using individual versions of the complete arc plot, Figure 2.11, where the relevant reactions (i.e. containing the radical of choice) are the only ones highlighted in colour and increased in opacity, Figure 2.13c, Figure 2.13d. Using this it is possible to see that several groups form arcs highlighting the cyclic OH  $\rightleftharpoons HO_2$  chemistry, Figure 2.12b. It is also visible that there are a significantly larger number of OH reactions, which corresponds to our knowledge of the mechanism, FIGURE MCM ALL.



**Figure 2.12: Arc diagram features for the Hydroperoxyl and Hydroxide radicals.** The  $HO_x$  cycle chemistry (b) can be seen within certain groups in the network. The main ones of these are highlighted in (a).

In addition to the reaction with the hydroxide radical, species may also be broken down with the aid of a photon - i.e. photolysis. If we compare all photolytic reactions, within those of OH, it can be seen that for certain groups, there exist two possible reaction pathways, ???. Since the reaction with OH serves to strip the hydrogen, the resultant species has a similar number of functional groups. Photolysis reactions, however, breaks the double bond, producing a species with a much smaller number of groups. This may be seen in the difference of length in the arcs emanating from these species, Figure 2.13.

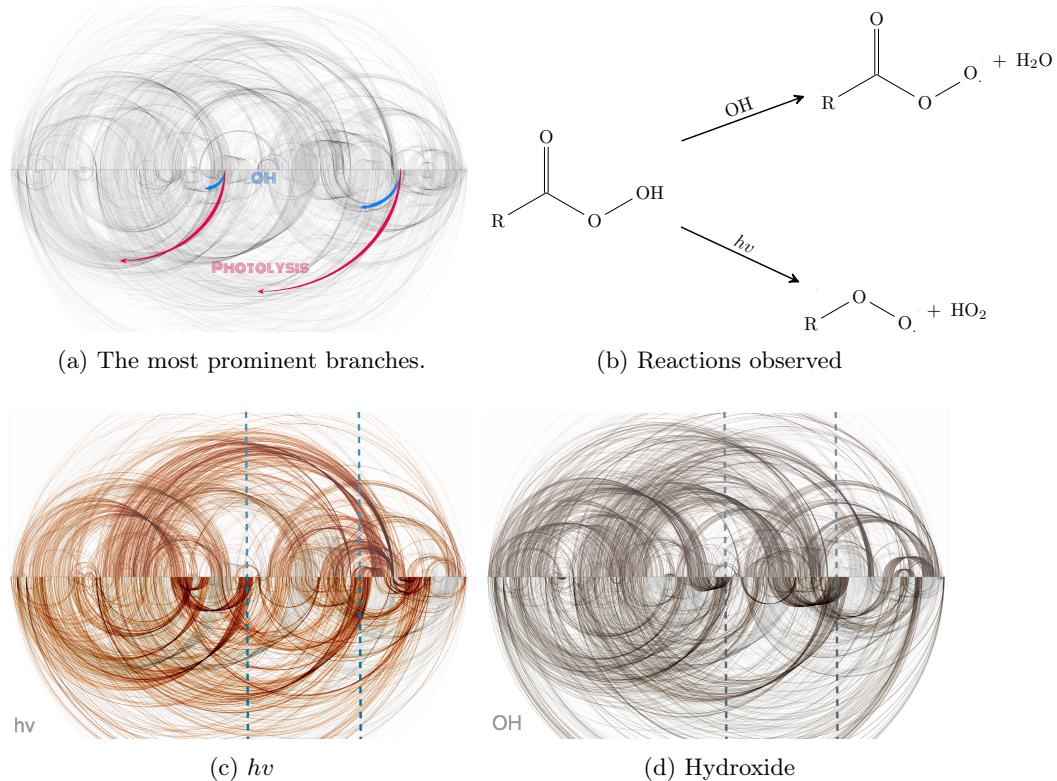


Figure 2.13: **Arc diagram features for photolysis and hydroxide reactions.** Photolysis results in species with a reduced number of functional groups, and therefore longer arcs. OH reactions for the same species do not produce such a drastic change on group number, and therefore have a smaller arc length.

Peroxy acetyl nitrates (PANs) play a key role in the modelling of photochemical smog (ozone events), [Singh, 2015]. PANs are species which are very stable at cold temperatures but can easily decompose and release  $\text{NO}_x$  when warmed. As this is an important species in the production of Ozone, this makes PANs an effective reservoir species with significant importance within atmospheric chemistry models (especially if transportation is involved). In the MCM the thermal decomposition of PANs is determined by the KBPAN rate constant. In comparing reactions of this, ??, with those of  $\text{NO}_2$  (at rate KFPAN), the emergence of the species cycle is seen.

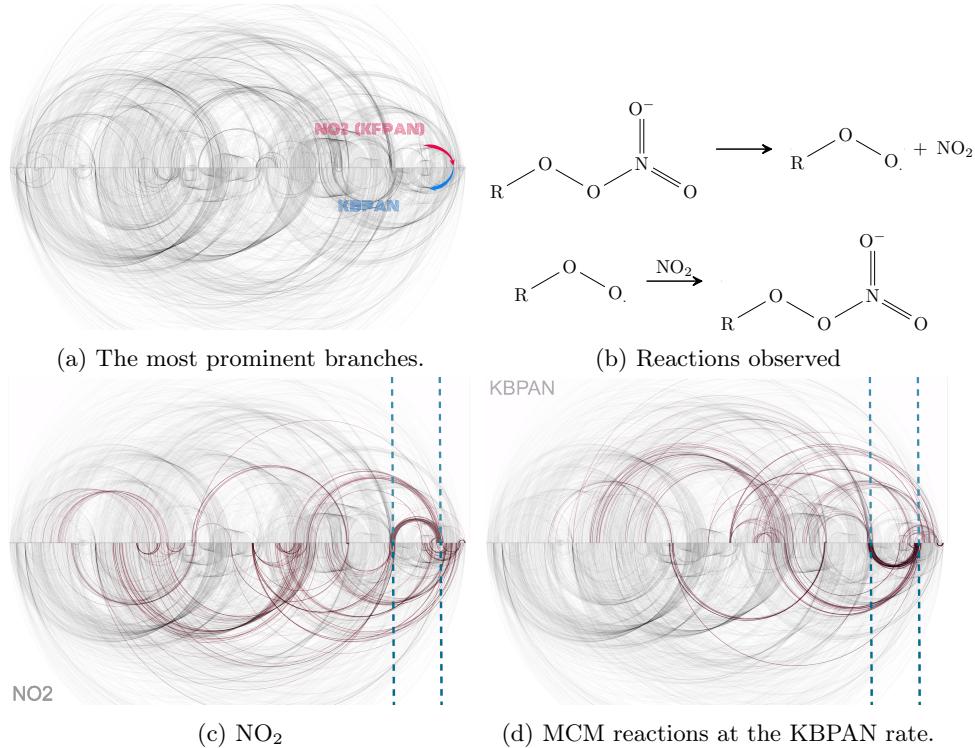


Figure 2.14: Arc diagram features for the PAN reactions.

Finally, patterns that exist within certain groups can be explored. We know that for RO<sub>2</sub>-RO<sub>2</sub> reactions there are three possible reaction pathways for the unpaired oxygen ... fill alter'

**NOTE:** A downside to the arc diagrams format that I have chosen is that for reactions between species of the same number of functional groups, there is no set direction. This can be seen in the positive arcs of Figure 2.15a.

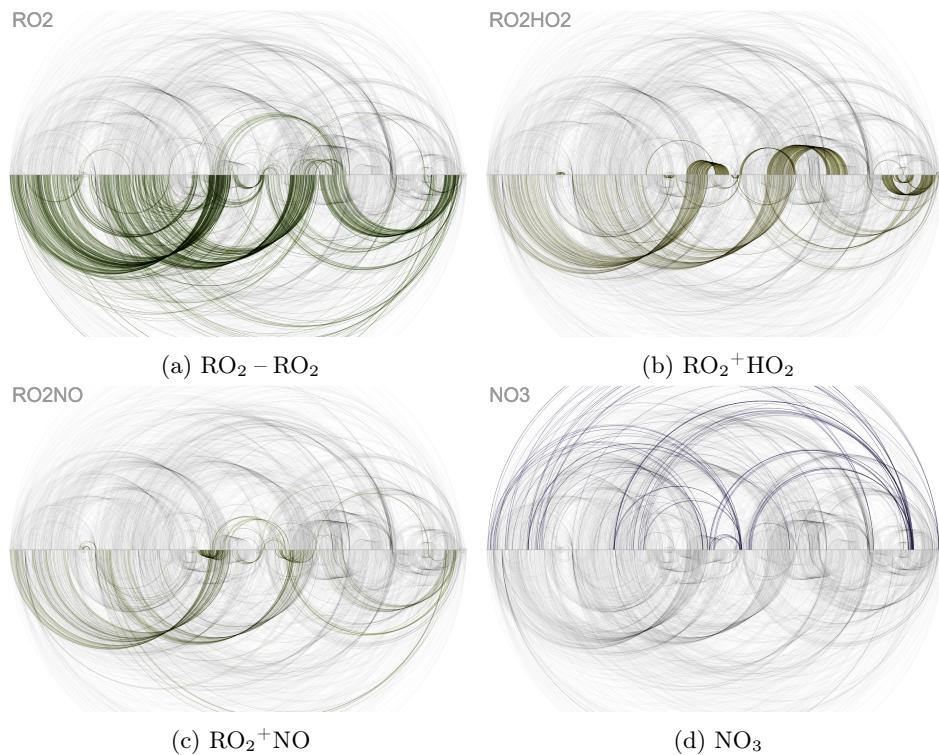


Figure 2.15: **Categories with a net direction/trend.** It is possible to see the patterns in which certain reactions function.

#### 2.1.4.2 The traditional network graph

Finally, we have the traditional network representation in the form of a mathematical graph.

Something about graphs. Intuitive. Blah

In ?? it is shown that the network shape is constrained by the construction protocol of the MCM, and thus the chemistry it contains. In ?? it is seen that in extracting an MCM subset to model different scenarios may result in the selection of sufficiently different mechanisms. This makes comparing the data collected difficult, as we do not know if a species' absence can be attributed to it having a very low concentration for a specific region, or simply that there was a lack of instrumentation capable of measuring it. A similarity in the mechanism subsets extracted as part of ?? (Mace Head) and Figure 2.16c (London) can be seen. Since both campaigns use a GC-FID (gas chromatography with flame ionisation detection), Table 2.1 the similarities in recorded species may be in-part attributed to this. Beyond the instrumentation, it the ClearFlo (London) campaign measures the most long-chain hydrocarbons; Mace Head records the most benzene based species possibly due to shipping emissions and Borneo (Figure 2.16b) has the most terpene species - it is a rainforest.

	Full MCM	Mace Head	London	Borneo
Species	XXXX	2363	1857	1587
Instrument	n/a	GC-FID	DC-GC-FID and	
GC×GC-FID	FAGE			
Observation Source	[?]	[Lewis et al., 1997]	[Bohnenstengel et al., 2015]	
[Dunmore et al., 2015]	[Edwards et al., 2013]			

Table 2.1: **The subset sizes, campaigns and instrumentation used for measurement.**

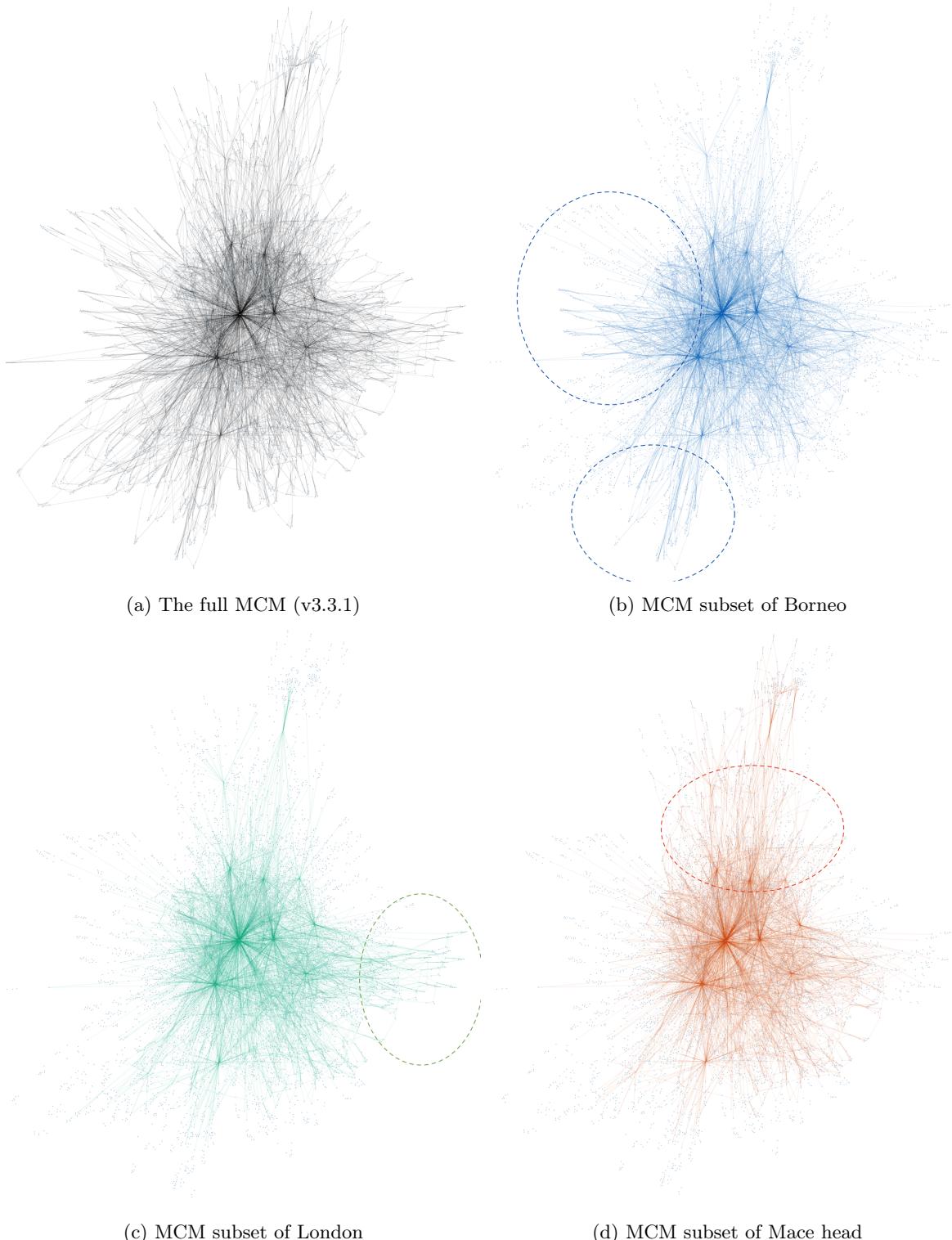


Figure 2.16: **Comparing MCM subsets using the traditional graph representation.** A graph (a) for the whole MCM is created. From this several mechanism subsets are extracted. The species and reactions are then extracted to produce the plots (b-d). All the species within the MCM are shown for comparison between plots, with only the links relevant to the mechanism being coloured.

### 2.1.5 Summary

Several different visualisation methods have been suggested. [select get filter statemen: donald knuth?]

Although each holds a different strength, node-link style methods tend to prevail in the representation of complex relationships for large systems. Other methods such as the chord diagram, tend to require a greater amount of pre-processing, which can obfuscates certain parts of the data. Matrix representation methods, although great for applying algorithms to, can often prove problematic for very large or sparse networks. Additionally these generally only have a single relationship between two items, meaning that information has to be grouped or averaged before use. Node-Link methods can not only allow easy filtration, by interactively hiding edges or reducing their opacity, but they show the entire state of the network, enabling to make informed choices about which parts are suitable for further exploration. Finally, the physical notion of a thread binding two items is intuitive and simple to comprehend, making these graph-like structures the perfect metaphor for representing relationships within the real world.

# Bibliography

- [Ackerman, 2016] Ackerman, J. (2016). *The Genius of Birds: The Intelligent Life of Birds*. Little, Brown Book Group.
- [Agarwal et al., 2017] Agarwal, S., Tomar, A., and Sreevalsan-Nair, J. (2017). NodeTrix-Multiplex: Visual Analytics of Multiplex Small World Networks. In *Complex Networks & Their Applications V*, pages 579–591. Springer International Publishing.
- [Aiello and Dunbar, 1993] Aiello, L. C. and Dunbar, R. I. M. (1993). Neocortex size, group size, and the evolution of language. *Current Anthropology*, 34(2):184–193.
- [Aumont et al., 2005] Aumont, B., Szopa, S., and Madronich, S. (2005). Modelling the evolution of organic carbon during its gas-phase tropospheric oxidation: development of an explicit model based on a self generating approach. *Atmospheric Chemistry and Physics*, 5:2497–2517.
- [Beck, 2017] Beck, H. (2017). London Icon: A History of Harry Beck’s Iconic Tube Map. <https://londontopia.net/site-news/featured/london-icon-tube-map/>. Accessed: 2019-10-6.
- [Bohnenstengel et al., 2015] Bohnenstengel, S., Belcher, S., Aiken, A., Allan, J., Allen, G., Bacak, A., Bannan, T., Barlow, J., Beddddows, D., Blossss, W., Booth, A., Chemel, C., Coceal, O., Di Marco, C., Dubey, M., Faloon, K., Flemiming, Z., Furger, M., Gietl, J., Graves, R., Green, D., Grimmimmimmond, C., Halios, C., Hamiamiamilton, J., Harrisison, R., Heal, M., Heard, D., Helfter, C., Herndon, S., Holmes, R., Hopkins, J., Jones, A., Kelly, F., Kotthaus, S., Langford, B., Lee, J., Leigh, R., Lewis, A., Lidster, R., Lopez-Hilfiker, F., McQuaidaid, J., Mohr, C., Monks, P., Nemitz, E., Ng, N., Percival, C., Prévôt, A., Ricketts, H., Sokhi, R., Stone, D., Thornton, J., Tremper, A., Valach, A., Vissisisser, S., Whalley, L., Williamsiamsiamsams, L., Xu, L., Young, D., and Zotter, P. (2015). Meteorology, air quality, and health in london: The clearflo project. *Bulletin of the American Meteorological Society*, 96(5):779–804.
- [Bostock, 2019] Bostock, M. (2019). Les Misérables Co-occurrence. <https://bostocks.org/mike/miserables/>. Accessed: 2019-12-4.

- [Brooks, 2019] Brooks, I. (2019). Tephigram.pdf.
- [Dahlstrom, 2014] Dahlstrom, M. F. (2014). Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the National Academy of Sciences of the United States of America*, 111 Suppl 4:13614–13620.
- [De Smedt, 2013] De Smedt, T. (2013). *Modeling Creativity: Case Studies in Python*. Proefschriften UA-LW : taalkunde. Universiteit Antwerpen, Faculteit Letteren en Wijsbegeerte, Departement Taalkunde.
- [Descartes et al., 1996] Descartes, R., Cottingham, J., and Williams, B. (1996). *Descartes: Meditations on First Philosophy: With Selections from the Objections and Replies*. Cambridge Texts in the History of Philosophy. Cambridge University Press.
- [Diebold, 2012] Diebold, F. X. (2012). A personal perspective on the origin(s) and development of big data: The phenomenon, the term, and the discipline, second version. PIER Working Paper Archive 13-003, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.
- [Dunbar, 1997] Dunbar, R. I. M. (1997). *Groups, Gossip, and the Evolution of Language*, pages 77–89. Springer US, Boston, MA.
- [Dunmore et al., 2015] Dunmore, R., Hopkins, J., Lidster, R., Lee, J., Evans, M., Rickard, A., Lewis, A., and Hamilton, J. (2015). Diesel-related hydrocarbons can dominate gas phase reactive carbon in megacities. *Atmospheric Chemistry and Physics*, 15(17):9983–9996.
- [Edwards et al., 2013] Edwards, P. M., Evans, M. J., Furneaux, K. L., Hopkins, J., Ingham, T., Jones, C., Lee, J. D., Lewis, A. C., Moller, S. J., Stone, D., Whalley, L. K., and Heard, D. E. (2013). OH reactivity in a South East Asian tropical rainforest during the Oxidant and Particle Photochemical Processes (OP3) project. *Atmospheric Chemistry and Physics*, 13(18):9497–9514.
- [Fairfield, 2012] Fairfield, H. (2012). Driving Safety, in Fits and Starts. *The New York Times*.
- [Fan and Bifet, 2013] Fan, W. and Bifet, A. (2013). Mining big data: Current status, and forecast to the future. *SIGKDD Explor. Newslett.*, 14(2):1–5.
- [Fisher, 2013] Fisher, J. (2013). Alternative Tube Maps: Circles Within Circles. <https://londonist.com/2013/01/alternative-tube-maps-circles-within-circles>. Accessed: 2019-10-6.
- [Freud and Cronin, 2013] Freud, S. and Cronin, A. (2013). *The Interpretation Of Dreams*. Read Books Limited.

- [Gottschall, 2012] Gottschall, J. (2012). *The Storytelling Animal: How Stories Make Us Human.* Houghton Mifflin Harcourt.
- [Harari, 2015] Harari, Y. (2015). *Sapiens: A Brief History of Humankind.* Harper.
- [Haroz et al., 2016] Haroz, S., Kosara, R., and Franconeri, S. L. (2016). The Connected Scatterplot for Presenting Paired Time Series. *IEEE transactions on visualization and computer graphics*, 22(9):2174–2186.
- [Hill and Dunbar, 2003] Hill, R. A. and Dunbar, R. I. M. (2003). Social network size in humans. *Human nature*, 14(1):53–72.
- [Hunt and Gadau, 2017] Hunt, G. and Gadau, J. (2017). *Advances in Genomics and Epigenomics of Social Insects.* Frontiers Research Topics. Frontiers Media SA.
- [InternetLiveStats, 2020] InternetLiveStats (2020). Total number of Websites - Internet Live Stats. <https://www.internetlivestats.com/total-number-of-websites/>. Accessed: 2020-1-3.
- [Jalali, 2016] Jalali, A. (2016). Reflections on the use of chord diagrams in social network visualization in process mining. In *2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS)*, pages 1–6.
- [Jenkins, 2002] Jenkins, M. (2002). History of the Master Chemical Mechanism (MCM) and its development protocols. slide deck. Presentation for the EPSR group, Imperial Collage.
- [Johnson, 2010] Johnson, S. (2010). *Where Good Ideas Come From.* Penguin Publishing Group.
- [Kirk, 2016] Kirk, A. (2016). *Data Visualisation: A Handbook for Data Driven Design.* SAGE Publications.
- [Knuth, 2019] Knuth, D. (2019). Knuth: The Stanford GraphBase. <https://www-cs-faculty.stanford.edu/~knuth/sgb.html>. Accessed: 2019-12-4.
- [Lewis et al., 1997] Lewis, A., Bartle, K., Heard, D., McQuaid, J., Pilling, M., and Seakins, P. (1997). In situ, gas chromatographic measurements of non-methane hydrocarbons and dimethyl sulfide at a remote coastal location (mace head, eire) july-august 1996. *JOURNAL OF THE CHEMICAL SOCIETY-FARADAY TRANSACTIONS*, 93(16):2921–2927.
- [M. Chazine, 2005] M. Chazine, J. (2005). Rock Art, Burials, and Habitations: Caves in East Kalimantan. *Asian Perspectives*, 44(1):219–230.
- [Martin and Miller, 1988] Martin, K. and Miller, E. (1988). Storytelling and science. *Language Arts*.
- [Mashey, 1998] Mashey, J. (1998). Big Data And The Next Wave of InfraStress .

- [Michal and Franconeri, 2017] Michal, A. L. and Franconeri, S. L. (2017). Visual routines are associated with specific graph interpretations. *Cognitive research: principles and implications*, 2(1):20.
- [Nightingale, 1820] Nightingale, F. (1820). Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army 1820-1910. <https://www.rct.uk/collection/1075240/notes-on-matters-affecting-the-health-efficiency-and-hospital-administration-of>. Accessed: 2019-10-10.
- [Nissen et al., 1993] Nissen, H., Damerow, P., Englund, R., Englund, R., Larsen, P., and Larsen, R. (1993). *Archaic Bookkeeping: Early Writing and Techniques of Economic Administration in the Ancient Near East*. University of Chicago Press.
- [Roberts, 2011] Roberts, J. C. (2011). The Five Design-Sheet (FdS) approach for Sketching Information Visualization Designs. In Maddock, S. and Jorge, J., editors, *Eurographics 2011 - Education Papers*. The Eurographics Association.
- [Roberts et al., 2014] Roberts, J. C., Yang, J., Kohlbacher, O., Ward, M. O., and Zhou, M. X. (2014). *Novel Visual Metaphors for Multivariate Networks*, pages 127–150. Springer International Publishing, Cham.
- [Schmandt-Besserat, 1992] Schmandt-Besserat, D. (1992). *Before Writing, Vol. I: From Counting to Cuneiform*. Before Writing. University of Texas Press.
- [Scott, 1988] Scott, J. (1988). Social network analysis. *Sociology*, 22(1):109–127.
- [Singh, 2015] Singh, H. (2015). Tropospheric chemistry and composition | peroxyacetyl nitrate. In North, G. R., Pyle, J., and Zhang, F., editors, *Encyclopedia of Atmospheric Sciences (Second Edition)*, pages 251 – 254. Academic Press, Oxford, second edition edition.
- [Tufte, 1983] Tufte, E. (1983). *The visual display of quantitative information*. Number v. 914 in The Visual Display of Quantitative Information. Graphics Press.
- [ugc, 2019] ugc (2019). Newspaper Rock. <http://www.atlasobscura.com/places/newspaper-rock>. Accessed: 2019-12-9.
- [Wada, 2010] Wada, H. (2010). The Development of Birdsong | Learn Science at Scitable. <https://www.nature.com/scitable/knowledge/library/the-development-of-birdsong-16133266>. Accessed: 2019-5-17.
- [Wattenberg, 2002] Wattenberg, M. (2002). Arc diagrams: visualizing structure in strings. In *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.*, pages 110–116.

- [Wybrow et al., 2014] Wybrow, M., Elmquist, N., Fekete, J.-D., von Landesberger, T., van Wijk, J. J., and Zimmer, B. (2014). *Interaction in the Visualization of Multivariate Networks*, pages 97–125. Springer International Publishing, Cham.
- [Ziemkiewicz and Kosara, 2008] Ziemkiewicz, C. and Kosara, R. (2008). The shaping of information by visual metaphors. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1269–1276.