

Module 6 Interactive data exploration & visualisation

Kimbal Marriott

25/01/2019

Contents

1	Interactive data visualisation	5
1.1	Analytics tasks	5
1.2	Manipulate a view	6
1.3	Visualise	7
1.4	Analytics	8
1.5	General usability and interaction guidelines	8
1.6	Summary	9
2	Immersive analytics	11
2.1	New technologies	11
2.2	Immersive analytics	11
2.3	Representative examples	14
2.4	Summary	15
3	New sources of data	17
3.1	Big data	17
3.2	Internet of Things (IoT)	17
3.3	Node-red	18
3.4	Summary	18
4	Activity: Visual analytics tool design	19

Chapter 1

Interactive data visualisation

As a data scientist you are comfortable using R and Python to interactively explore data. However, often it's better to use a less powerful but easier to use interactive visualisation tool. This is because programming is slow and error prone, and many analysts are not experienced programmers. Most data analysts find exploring data with Tableau a lot easier than using R.

In this topic we look at how to design such visual analytics tools. While it is unlikely that you will ever be involved in the design or implementation of a really generic tool like Tableau it is quite possible that you may be a part of a team developing a more specialised analytics tool. Scaffold Hunter is an example of such a visual analytics tool. It was originally developed for drug design. To find out more about Scaffold Hunter please take a look at the video below. Another good reason for learning about the design space is that it informs more critical evaluation of existing visual analytics tools.

1.1 Analytics tasks

The first part of any design process is to understand the user tasks: *what* do they want to be able to do with the tool. Once we know this, then we can decide *how* best to achieve this. There have been many different ways of classifying the user tasks in data exploration.

Ben Schneiderman introduced the very influential data visualisation mantra: *Overview first, zoom and filter, details on demand* as a summary of the key steps in data visualisation. An initial overview provides the user with a summary of the entire data set and a context in which to drill down.

Providing an initial overview makes sense for applications in which the relevant context is well defined and the data set is moderately sized. In many applications however providing an overview may not be beneficial. For instance, when using Google Maps it isn't that useful to always start with the whole Earth. For cases like this van Ham and Perer introduced an alternative mantra: *search, show context, expand on demand*. This is precisely how Google Maps works.

Other data visualisation researchers have attempted to provide a more fine-grained and multi-level analysis of the visual analytics tasks:

Discovery: The fundamental task that the analyst wants to achieve is to derive insight or knowledge from the data. The desired outcome might be to formulate a hypothesis ("discover the unexpected") or to test a hypothesis ("confirm the expected"). The hypotheses may take the form of cause-and-effect relationships, trends, correlations or clusters.

Presentation: This goal refers to presenting insight or knowledge that has already been found to some intended audience.

We shall focus on the discovery task. This relies on performing a series of *analytical tasks*:

- *Search* for elements that satisfy certain properties, if they exist. This might be locating a known data point, filtering the data, or finding outliers.
- *Identify* the properties of a single data item
- *Compare* or *rank* elements
- *Visually* identify patterns in some subset of elements. Examples include trends, correlations, clusters or categories.
- Calculate *derived* properties not originally in the data. These may be data transformations, data aggregations or may be statistical properties such as regression lines or clusters

Analytical tasks rely on *presentation* tasks:

- *Visualise* by mapping elements and their attributes to visual variables to create a view
- *Manipulate (or configure)* a view by navigating and selecting subsets of elements

There are other user tasks that support analytical tasks

- *Annotate* visual elements with text or graphical elements
- *Record* visualisation elements so that they can be preserved and accessed outside the analytics tool. For example, Tableau records a graphical history, with snapshots showing how the current visualisation was reached
- *Revisit* an earlier visualisation or relocate an element or pattern that was previously found by the analyst

We now look at various ways to support the analytical and presentation tasks.

1.2 Manipulate a view

One of the defining characteristics of a visual analytics tool is that it supports direct manipulation of views by the user. This is unlike say using **R** and **ggplot2** in which views are static and any modification (apart from zooming and panning) requires generating a new view from the command line. Please take a look at the Line Up system for a good example of view manipulation.

Common reasons for manipulating the view are to

- *Control a dynamic visualisation.* It is natural to visually encode data that changes over time as an animation. Effective understanding requires that the viewer can control the speed, pause and move backwards and forwards in time. If possible the changes should be staged and use animated transitions so that the viewer's attention is focused on salient changes and they preserve their mental map of the visualisation. Also remember that animation is not the only way, and often not the best, of showing dynamic data: small multiples are very effective and can be used to show differences as well as the actual values.
- *Modify the visual encoding.* It is common to allow the viewer to change the choice of colour palette (improving accessibility), or provide the ability to reorder or sort elements and to align elements.
- *Filter items.* A major reason for interaction is to reduce data complexity. Filtering data items is a widely used technique for reducing complexity. Typically the viewer can use slider bars, search terms etc, to filter data based on restricting attribute values.
- *Control the level of detail (LoD).* Interaction can be used to control whether sub-hierarchies or clusters are expanded/collapsed or the level of detail shown in the data items.
- *Filter attributes.* Attributes can also be filtered to reduce complexity. The most common approach is slicing which eliminates an attribute by only showing the items with the chosen value in that attribute. Slicing is widely used to show axis-aligned 2D slices or sections through 3D images. Closely related is the *cut*, which shows the data on one side of a cutting plane.
- *Navigate.* But undoubtedly the main use of interaction is to allow the user to navigate through the view, to change the viewpoint.

The most common navigation technique is to provide zooming and panning, exactly as in Google Maps. In *geometric zooming* the objects do not change their spatial appearance, they simply get larger or smaller. In *semantic zooming* the appearance changes at different scales. Typically, as the object becomes larger, the level of detail increases. For instance in maps, towns are shown as dots on large scale maps but when the viewer zooms in they reveal their spatial shape and then the buildings and streets that form the town.

It is useful to distinguish between *unconstrained navigation* in which the viewer can zoom and pan anywhere, and *constrained navigation* in which the allowed viewpoints are restricted. For instance, the user may not be able to zoom in or out too far, or may be able to select an item of interest or home view and the system will move to that viewpoint, often using an animated transition. Constrained navigation has the advantage that users are less likely to get lost, though they lose some flexibility.

An alternative to zooming and panning is to use what are called *focus+context* views in which data in the view is shown in more detail around the focal point on the view and in less detail elsewhere in the view, so as to provide context. Probably the best known focus+context view is the fisheye lens, in which the view is distorted so as to achieve an effect similar to the fish eye lens in photography. Other types of focus+context views remove detail for objects away from the focus or provide detail on a separate layer, for instance using the metaphor of a magnifying lens.

1.3 Visualise

Multiple views are required to understand all but the simplest data. These show complementary aspects of the data. For instance, they might show:

- the same data visualised at different levels of scale for instance an overview and a detailed view,
- different attributes of the data, or
- comprise small multiples showing data at different time periods or for different values of a categorical variable.

A good visual analytics tool provides the views and view manipulation tools that will be of most use to the typical user and controls to customise the view. The degree of customisability depends greatly on the application and the kind of users. Take a look at the VisGets tool for exploring posts on on-line forums and consider the choice of views it provides as well as the kinds of interaction provided.

Views can be organised in many different ways:

- *Temporal*: the viewer can toggle or move between different views, showing one at a time.
- *Side-by-side*: the views are shown next to one another
- *Layered*: the views sit on top of one another.

Each way of organising the views has advantages and disadvantages. Showing the views one at a time has a high cognitive cost as the user must remember information from previously shown views, layering can lead to clutter and less effective visual encodings because of the need to use different encodings for the different layers, while a side-by-side arrangement takes up more screen real estate and each view has less resolution. Nowadays, as display resolution is increasing, I think the side-by-side arrangement should be the first choice considered.

Linking different views of the data is important if they are to be effective. Some of the ways of linking different views are

- *Shared visual encodings*. This includes using the same color or shape encoding for an attribute shared between the different views.
- *Alignment and common scale*. This is particularly useful in small multiples so as to allow ready comparison between the different views. It is used very effectively in scatter plot matrices.
- *Shared filtering*. Filtering in one view affects all views. This allows filters to be linked with the appropriate views. Thus one view might show data organised by time with a time line for filtering

by date while another view might show the data organised by category with check boxes to select categories.

- *Shared selection.* Highlighting selected objects in all views is a powerful way of synchronising the views. This is often called brushing and typically occurs on mouse hover.
- *Shared navigation.* Zooming and panning affects all relevant views.

A crucial part of the view are the keys, labels and legends which explain what the view is of and how to read the visual encoding. It is surprising how often these are forgotten. Labelling and gridding should be consistent across different views.

A common use of multiple views is for navigation. A common approach is to provide the user with both a *detailed* and *overview* of the data, usually shown side-by-side but sometime overlaid. The two views are linked in that the position of the detailed view is always shown on the overview and sometimes can be controlled from the overview. Depending upon the application either the detailed or the overview may be larger.

1.4 Analytics

Visual analytics is not only about visualisation, it is also about analytics. The choice of these is application specific. Some common kinds of analytics are

- *Item aggregation, clustering and smoothing.* Aggregating multiple items into a new item is one way of reducing complexity and allowing the wood to emerge from the trees. Totals, averages, counts, clustering, curve and surface fitting, and spatial clustering are all ways to aggregate/combine items. The box plot is a good example of a visualisation that aggregates information about a distribution.
- *Attribute aggregation* reduces the number of attributes by combining attributes to give a new attribute. Dimension reduction techniques such as PCA or MDS are examples of attribute reduction.
- *Data transformation* allows the user to scale or combine data sets
- *Identifying interesting views* is less common but potentially very useful. This might for example analyse correlations between all attribute pairs and only show those with a correlation above some threshold.

One pitfall to be aware of when showing the result of some kind of analytics is to not mislead the viewer. It is good practice to allow the viewer to see the raw data together with the results of clustering or smoothing so that they can judge the efficacy of the fit.

1.5 General usability and interaction guidelines

The development of visual analytics tools is an application of human-centered design. The acronym PACT (People, Activities, Context and Technology) summarises this approach. The designer needs to understand the background and skills of the people who will use the system and take account of their perceptual and cognitive abilities. They need to involve these end-users in the design of the system from the very beginning and really listen to their advice and criticisms. They need to understand what they want to use the system for, the context in which it will be used and what sort of technology it will need to run on. It is worth emphasising again the need to design and test the application with real users all the way through implementation. For me this is the golden rule of HCI (Human Computer Interaction) design.

There are of course many other principles in effective HCI design. These include ensuring that new users can quickly learn how to use the tool. This requires using familiar conventions and using a consistent and predictable interaction model. The interface should allow the user to complete their tasks *flexibly* by not unnecessarily constraining the order of operations and by allowing them to customise it to their needs and wants. Finally the system should be robust. Visual feedback should let the user know in what state the system is in and it should be easy to recover from mistakes.

When developing an interactive system it is important to think about latency: how long will it take for the system to respond to the user. This significantly affects the user experience. The system will feel responsive if the system provides feedback in the following time scales (Table based on Table 6.1 from *Visualization Analysis and Design* by T. Munzner, 2014):

Level of Responsiveness	Response Time (in seconds)	Example
Perceptual processing	0.1	Screen update
Immediate response	1	Visual feedback on selection or animated transition between frames
Brief task	10	Sort or filter

From the user's perspective it is important for the system to provide feedback, usually visual, that the action they have requested has been completed. Highlighting selected elements is an example of visual feedback. If the operation is going to take longer than they might expect then some kind of visual indication of progress is a good idea.

1.6 Summary

In this topic we have investigated how to design visual analytics tools and interfaces. This is a complex area and requires understanding basic HCI design principles and methodologies. In my view the most important is to employ a participatory design process in which the end-users of the system are involved in the design from the very beginning and are frequently given the opportunity to provide feedback on the evolving design.

We have looked at the kinds of tasks the tool should support and also key elements in the interface such as view manipulation and navigation, linking different views, and kinds of analytics to provide.

REFERENCES AND FURTHER READING This topic is based upon

Pretorius, A. Johannes, Helen C. Purchase, and John T. Stasko. Tasks for multivariate network analysis. In *Multivariate Network Visualization*, pp. 77-95. Springer International Publishing, 2014.

Ward, Matthew O., Georges Grinstein, and Daniel Keim. Chapters 11-13 of *Interactive data visualization: foundations, techniques, and applications (2nd Ed)*. CRC Press, 2015.

Munzner, Tamara. Chapters 11-14 of *Visualization Analysis and Design*. CRC Press, 2014.

Further reading

- Chapters 11-14 of Munzner, 2014.

Chapter 2

Immersive analytics

In a Brief history of data visualisation we looked at the reasons for the rise of data visualisation and visual analytics. One of the most important was changing technology for producing and presenting graphics. Widespread use of information graphics could not have happened without paper and printing while interactive data visualisation and visual analytics was not possible before the computer. In this topic we look at what new and emerging display and interaction technologies might mean for data science.

2.1 New technologies

Virtually all tools for interactive data exploration and visual analytics are targeted towards a standard desktop computer: a single average-sized display, keyboard and mouse. This is not surprising as this is the standard computing environment in business, science and government. However this is changing.

On one hand, we have the development of expensive bespoke technologies, mainly for scientific visualisation. For example, ultra-high resolution virtual reality environments like the Monash CAVE2TM use wall-mounted displays and head tracking equipment to immerse users in 2D and 3D visualisations. A classic use-case for the CAVE2 is “walk-throughs” of human-scale architectural and engineering models.

To see the CAVE2 in action take a look at the above video.

On the other hand, we also have the recent entry of touch screens and touch tables into the mainstream marketplace as well as the arrival of technologies such as the Leap Motion, Kinect, zSpace, Oculus Rift and (soon) Microsoft HoloLens. These new devices potentially provide more natural user interfaces than the traditional mouse and keyboard and immersive 2D and 3D visualisations that move far beyond the current HD desktop monitor. Industry—in particular the entertainment industry—has driven this development and it has progressed rapidly, providing immersive visualisation for a fraction of the cost of the CAVE. As a first step in this direction Tableau have just released Vizable an iPad app for visual analytics.

The amount of new, low-cost interaction technology hitting the market is unprecedented. Given the rise of new manufacturing technologies (additive manufacturing) and business models (crowdfunding, etc.) the rate of innovation in display and interaction devices is only going to increase.

2.2 Immersive analytics

I believe it is likely that, like other innovations in graphics production and display technologies have done, the arrival of these new low-cost interaction and display technologies will reshape data visualisation.



Figure 2.1: Large tiled-displays with touch and gesture controlled interaction provide the possibility of deep collaboration between data scientists. License: Copyright © Monash University, unless otherwise stated. All Rights Reserved.

At Monash we have a new research initiative exploring how these new user interface and virtual reality technologies can be used in data analytics and data science. We call this initiative Immersive Analytics. Some of the things we are looking at are:

- *To 3D or not to 3D:* As we have discussed there are currently good reasons to limit the use of 3D when visualising abstract data: occlusion, distortion introduced by perspective and difficulty of interaction. However the new technologies potentially mitigate some of these problems and so the use of 3D for abstract data visualisation may well become more popular.
- *Mixed 2D and 3D visualisation:* 3D visualisation has been used in the physical sciences, engineering and design, and for geospatial analysis while 2D visualisations are used to show tabular and network data. In many larger, multidisciplinary projects, there is a need to combine both sorts of visualisations. For instance, in the life sciences different aspects of a cell are shown using 3D volumes, 2D images and 2D multivariate network data. Emerging technologies provide support for this by allowing holistic visualisations incorporating both 3D and 2D.
- *Interaction and interface design:* What are the interface ‘tricks’ and affordances such as high-resolution displays, sound, touch and responsive interaction that change the user perception from an allocentric view of the data to a more egocentric and immersive view of the data? What kinds of interaction and interaction devices will allow the user to work as comfortably and productively with 3D data in an AR and VR environment as they can currently work with 2D data in a desktop environment.
- *Visualisation platform:* While there many information visualisation libraries, like **D3** or **ggplot2**, these are not designed to work with immersive visualisation environments. On the other hand **Unity** is designed for gaming applications rather than analytics while the **Visualization Toolkit (VTK)** is targeted at scientific visualisation applications but not information visualisation. We need information visualisation toolkits that work with immersive visualisation platforms.

These new technologies also open up new opportunities for visual analytics:

- *Immersive narrative visualisation:* The first is to use immersive VR and AR for more effective and engaging presentations to stake-holders such as managers, policy makers and the general public. Immersive VR is routinely used to present architectural and engineering designs to clients but its use for presenting abstract data to stakeholders has not been considered. Imagine the power of a presentation about health risks or emergency response planning in which the city is modelled in the space around

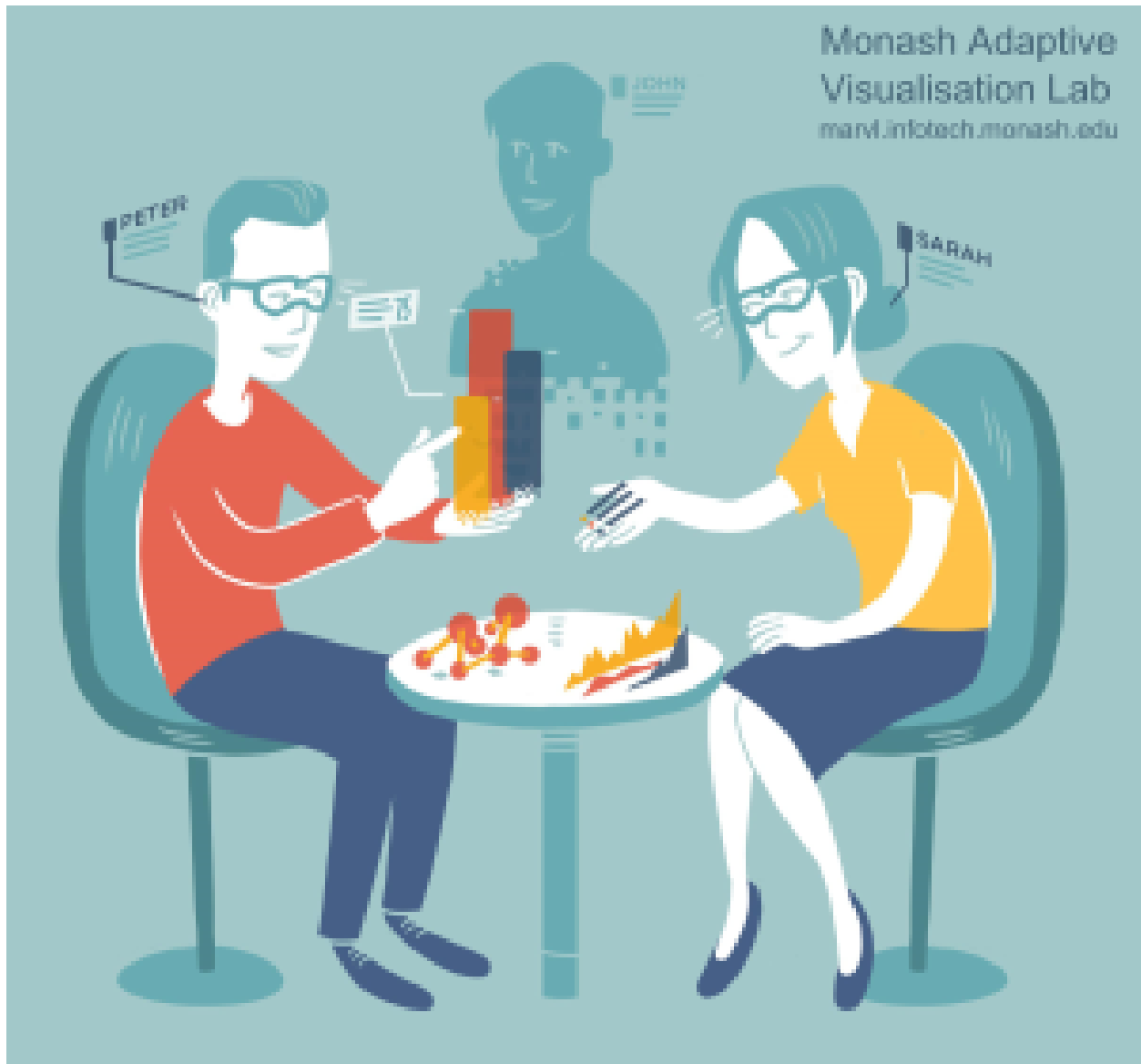


Figure 2.2: Augmented reality promises distributed collaboration between data scientists working in shared virtual environments. License: Copyright © Monash University, unless otherwise stated. All Rights Reserved.



Figure 2.3: ContextuWall distributed collaboration application running in the CAVE2 at Monash License: Copyright © Monash University, unless otherwise stated. All Rights Reserved.

the viewer and key problems are highlighted through data overlays directly on the model which the viewer can interactively explore.

- *Situated analytics:* Augmented reality (AR) displays potentially allow applications to overlay analytics onto objects in the actual physical environment. This could allow environmental scientists or farmers to see sensor data overlaid on top of the landscape they are standing in and perform data analysis in situ. Shoppers could overlay nutritional, environmental and ratings and comments from social media on products as they stroll down the shopping aisle. No longer must the visual analyst sit at a desk, AR allows visual analytics out into the real-world, potentially ubiquitous and part of everyday life.
- *Collaboration:* Traditionally data scientists perform data analysis individually, and then report back what they have found. Devices like the CAVE and Oculus Rift potentially support much deeper collaboration between data scientists.

2.3 Representative examples

Here are some examples showing the sort of data visualisation that immersive visualisation environments will support.

The first example is shown in the video clip above. It is a virtual reconstruction of the ancient Angkor temple complex in Cambodia displayed on Monash’s CAVE2 visualisation facility. The reconstruction is programmed in Unity, a popular gaming engine. Not only does this provide a 3D model of the site, but also uses rule-based using multi-agent simulation to reconstruct human activity in and around the complex. Data gathered from the simulation can then be explored using overlaid heatmaps of traffic flows. As well as supporting the evaluation of archeological hypotheses, e.g. on settlement density, detailed 3D interactive animations like that of Angkor are potentially a powerful way to communicate findings to stakeholders.

The second example also utilises the CAVE2. This time the application is ContextuWall, an application that supports distributed collaborative decision making. In this application the images are mirrored on the displays in different locations. Viewers at all locations can interact with the display through their favourite touch screen device by using a finger to “flip” new images on to the shared display or remove, annotate the images and also position them.

The third example is another example of collaborative data analysis, but this time with the Oculus Rift personal virtual reality environment. Here we see the two analysts looking at air traffic control data. They can see where the other person is looking and direct them to points of interest.



Figure 2.4: Virtual collaboration using the Oculus Rift at Monash SensiLab. License: Copyright © Monash University, unless otherwise stated. All Rights Reserved.

2.4 Summary

We can see that new display and interaction technologies will completely change the way in which we can interact with the computer. It is likely that these will also impact on visual analytics and lead to much greater use of immersive visualisation for more abstract data analysis, particularly for collaborative analysis.

CAVE2 is a trademark of the University of Illinois Board of Trustee

FURTHER READING

ElSayed, N., Thomas, B., Marriott, K., Piantadosi, J., & Smith, R. (2015, September). Situated Analytics. In *Big Data Visual Analytics (BDVA)*, 2015 (pp. 1-8). IEEE.

Chandler, T., Cordeil, M., Czauderna, T., Dwyer, T., Glowacki, J., Goncu, C., Klapperstueck, M., Klein, K., Marriott, K., Schreiber, F. and Wilson, E., 2015, Immersive Analytics. In *Big Data Visual Analytics (BDVA)*, 2015 (pp. 1-8). IEEE.

Chapter 3

New sources of data

3.1 Big data

We have talked about the future of data visualisation in terms of new kinds of display and interaction technologies. As we have discussed another driver for changes in data visualisation is the availability of data. And as data scientists we should be well aware that Big Data is here (see Wikipedia on Big Data).

Some of the reasons for the massive increase in data are social media, higher resolution scientific instruments such as radio and optical telescopes, RFID and other object tagging, remote sensing and wireless sensor networks, software logs, cameras and microphones. Data visualisation is often promoted as one of the ways of handling such Big Data. This certainly has some truth but the immense size of the data sources, terabytes or even petabytes, means that considerable aggregation, filtering and selection is required before any kind of visualisation is possible. Visual analytics with its human-in-the-loop exploration model using both analytics and visualisation seems one of the best approaches.

diagrams_datasets/section3/Viegas-UserActivityonWikipedia-300x226.gif

Visualisation of daily Wikipedia edits taken from Wikipedia entry on Big Data by Fernanda B. Viégas. Wikipedia is terabytes in size and an example of big data. License: Copyright © License: CC Attribution 2.0 Generic (CC BY 2.0)

3.2 Internet of Things (IoT)

One of the most interesting sources of Big Data is the *Internet of Things (IoT)* or more grandiosely the *Internet of Everything (IoE)*. This is “the network of physical objects—devices, vehicles, buildings and other items embedded with electronics, software, sensors, and network connectivity—that enables these objects to collect and exchange data.” (Wikipedia). A recent blog post on Data Science Central suggest that by 2020, this has the potential to connect 50 billion people, devices and things. That’s an awful lot of data and its coming very soon. Also see How Big Will The Internet of Things Be?

One of the characteristics of this kind of sensor data is that it is streamed (streamed data is V for Velocity in the The 4 Vs of Big Data (see the IBM Infographic)). Real time analysis and visualisation of large quantities of streamed data is almost certainly going to be a major preoccupation for data scientists in the next decade. A number of graphical tools based around the data flow programming paradigm are being developed to support this kind of analysis. Shown below are a couple of interactive tools that allow you to connect to various data sources and process them in a series of steps or a pipeline. Both node-red (IBM), and StreamTools (NYT) are free downloads if you would like to try either.

3.3 Node-red

Instructions for installing and running node-red (Windows) and for setting up a Twitter feed using node-red are here

If you just want to see node-red running:

This video shows a twitter feed running at a reasonable speed, seems people are twittering about Apple constantly (and bear in mind you only get access to about 1% of feeds for free, see firehose vs garden hose vs spritzer). The pipeline then feeds the tweets into a sentiment analysis node (essentially a dictionary that categorizes into positive, negative and neutral terms) which then splits into 3 output files. There's also a 'debug' siding, so you can see what's going on, output is far right. In the video debug is turned off and on, but note that the files will keep filling up even if it's off.

StreamTools

(on a Mac)

3.4 Summary

Not only will new kinds of display and interaction devices change data visualisation. Another driver for change is Big Data and in particular the new kinds of streamed data available from massive interconnected sensor networks like the Internet of Things.

Chapter 4

Activity: Visual analytics tool design

Work in pairs to come up with designs for a visual analytics tool for on-line exploration of the collection of a natural history museum (plants, animals, rocks, fossils etc).

The first step is to identify the tasks and audience the tool is intended to support.

Next come up with various designs for the tool. Mock these up in PowerPoint and show the steps in completing various tasks.

You should prepare designs for two platforms:

1. A standard desktop environment
2. An Oculus Rift equipped with gesture recognition

If you have time now think about an interface for tool running on a Head Mounted Display Augmented Reality display that is for use when you are wandering around the museum.