



Immersive Visual Analytics

A White Paper

November 2021

www.daden.co.uk

info@daden.co.uk

+44 (0)7811 266199

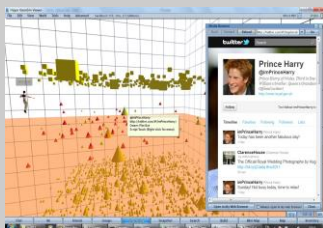
How did we get here?



At Daden we have a long history of using immersive spaces to visualise data. When Google Earth came out we were one of the first companies to bring RSS feeds into it. In virtual worlds like Second Life we created physical looking “command centres” in which to view 2d data (as above), and inherently 3D data (as below).



From here we progressed to platforms such as OpenSim where we had far more control about what we could plot and how we could display it, as in the visualisation of Twitter data during the Royal Wedding in 2011 below.



It was these experiences that led us to develop Datascape, taking the best parts of virtual worlds and applying them to visual analytics.

Introduction

Having worked in virtual worlds and immersive spaces for over 15 years, and with nearly 25 years' experience of the technology, you begin to spot long term trends, applications and aspirations. One of those trends has been the use of immersive spaces to create interesting data visualisations. Unlike traditional 2D visualisations, or even 3D data cubes, these immersive data experiences let you move into the data, viewing it from inside and out. Almost all of us who have had experience of developing and using immersive data visualisations have felt that they add a definite *something* to the experience, and more importantly have enabled us to spot trends and anomalies, and to remember what we saw, better than non-immersive visualisations.

This paper reflects on our experiences of data visualisation, following a journey from 2D and 2.5D visualisations to fully immersive 3D visual analytics. It builds not only the anecdotal experiences of ourselves and others, but also on the academic research in this area by third parties, and our own primary research with Aston University funded by the Defence Scientific and Technical Laboratory (Dstl) of the UK Ministry of Defence, a further project funded by InnovateUK, and our experiences in developing Datascape, the first general purpose immersive visual analytics application.

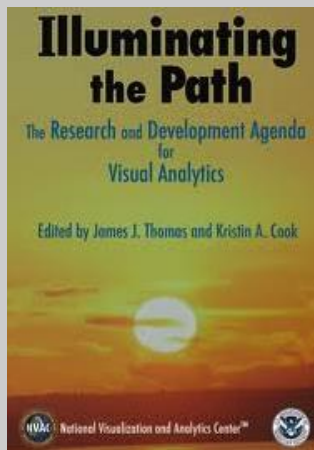
Terminology

First, a few words about terminology. Terms like “data visualisation”, “infographics”, “visual analytics” are often used interchangeably. But to us there are some clear differences, and our interest is far more about visual analytics than infographics. Our definitions would be:

- **Data Visualisation:** Any form or complexity of visualisation of information, but where the focus is likely to be more passively viewing the information rather than interacting with it and changing the display
- **Infographic:** A simplified data visualisation whose prime aim is to communicate AFTER the analysis has been completed
- **Visual Analytics:** An interactive visualisation of data used to perform an analysis and with the ability to alter the display in order to make better sense of the data

Visual Analytics

Illuminating the Path



The 2005 report by James Thomas and Kristin Cook of the National Visualization and Analytics Centre (NVAC) which marked the beginning of a new discipline - visual analytics.

You can read the report in full at:

http://vis.pnnl.gov/pdf/RD_Agenda_VisualAnalytics.pdf

Thomas, J.J. & Cook, K.A. (Eds.). Illuminating the path: The research and development agenda for visual analytics. IEEE Computer Society, 2005

In 2005 James Thomas and Kristin Cook of the National Visualization and Analytics Centre (NVAC) produced a report funded by the Department of Homeland Security and co-produced with the Institute of Electrical and Electronics Engineers entitled “Illuminating the Path: The Research and Development Agenda for Visual Analytics”. For many this call-to-arms marked the beginning of a new discipline - visual analytics. The report identified one key challenge:

“the analysis of overwhelming amounts of disparate, conflicting, and dynamic information to identify and prevent emerging threats, protect our borders, and respond in the event of an attack or other disaster. This analysis process requires human judgement to make the best possible evaluation of incomplete, inconsistent, and potentially deceptive information in the face of rapidly changing situations.”

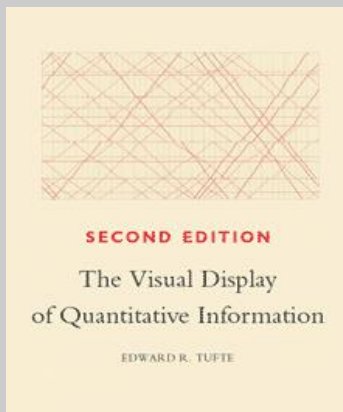
Whilst phrased in counter-terrorism terms in response to the attacks of 9/11 the sentiment applies to anyone currently faced with large amounts of complex, and often unstructured, data, to analyse in order to make the best scientific or commercial decision. Indeed the next paragraph is also likely to strike a chord with anyone involved in data analysis:

“Employing today’s best practices in these areas and using the best possible training, technologies, and processes will still not position the country to meet the growing needs The scale of data is staggering, and our ability to collect data is increasing at a faster rate than our ability to analyze it. Although massive amounts of information are available from multiple sources, the relevant information content exists in a few nuggets. New methods are required that will allow the analyst to examine this massive, multi-dimensional, multi-source, time-varying information stream to make decisions in a time critical manner.”

The report then positions Visual Analytics as a response to this challenge, and by extension to the wider problem:

“Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces. People use visual analytics tools and techniques to synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data; detect the expected and discover the unexpected; provide timely, defensible, and understandable assessments; and communicate assessment effectively for action.

Edward Tufte



You can't get far in talking about Data Visualisation without mentioning Edward Tufte – the American statistician and professor emeritus of political science, statistics, and computer science at Yale University.

His large format hardback books, including *The Visual Display of Quantitative Information* and *Envisioning Information* are just a joy to hold, look at and read. They are beautiful but informative artifacts, and a far cry from the typically over-long and over-hyped business book or dry academic tome.

There are many lessons you can learn from reading Tufte, and most apply to computer visualisations, whether 2D or 3D. One of our favourites is the concept of data-ink – making sure that as high a percentage as possible of ink on the page is actually showing data, rather than being used for features which are just eye-candy.

The report describes visual analytics as a multidisciplinary field that includes the following focus areas:

- *Analytical reasoning techniques that enable users to obtain deep insights that directly support assessment, planning, and decision making*
- ***Visual representations and interaction techniques that take advantage of the human eye's broad bandwidth pathway into the mind to allow users to see, explore, and understand large amounts of information at once***
- ***Data representations and transformations that convert all types of conflicting and dynamic data in ways that support visualization and analysis***
- *Techniques to support production, presentation, and dissemination of the results of an analysis to communicate information in the appropriate context to a variety of audiences."*

It is these second and third bullet points (our emphasis) which are of particular interest to us.

2D Visualisations

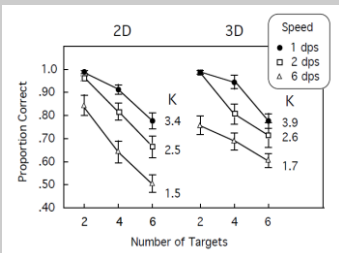
2D visualisation has a long history, as Edward Tufte's books (see sidebar) amply demonstrate. The advantages of 2D are:

- It is familiar and easily understood
- It can be easily produced in printed or drawn form
- You can easily apply linear measuring devices such as a ruler
- It is easy to "lasso" points for further analysis
- There are no perspective effects

However we are increasingly finding disadvantages:

- The drawing space (screen or paper) rapidly becomes too dense as we plot larger and larger amounts of data
- We only get one view/perspective on the data at a time
- We only have two primary axes to plot data against
- The flat image provides no "immersive" cues to help the brain identify trends/outliers or remember one plot from another
- If we have data of widely varying ranges the smaller set will get scrunched up in order to keep the larger set on the axis, or we lose the larger set in order to show more definition in the smaller

2D vs 3D



Some interesting work comparing 2D and 3D for air-traffic control operations was done in 2005 by Geniva Liu et al at the University of British Columbia [LIU05].

Subjects were asked to see how many moving objects they could track over time, first in 2D and then using a 3D non-immersive data cube.

The results showed that performance was enhanced by using the 3D display instead of the 2D display. More surprisingly when the cube was moved and rotated by a third party whilst the exercise was underway it did not have a detrimental impact on performance.

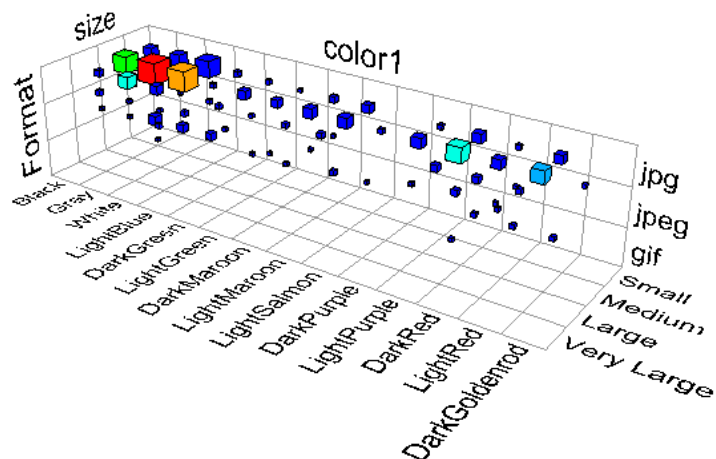
“These results confirm that allocentric object tracking depends on the perception of a coherent 3D scene. [...] These results also established that the addition of several pictorial depth cues (wire frame, checkerboard floor) and dynamic changes in relative size did not negatively affect tracking accuracy. If anything, there was a trend for improved tracking in the large target sizes in the 3D condition.”

- The number of marker types we can have is limited given their typically small size

As we have to deal with ever increasing amounts and complexities of data then does 3D offer us any advantages?

Data Cubes and Non-Immersive 3D

Many simple charting apps such as Excel provide a very basic pseudo-3d chart where they just use 3D bars instead of 2D bars to give the illusion of 3D (and Tufte has lots to say on the dangers of such plots). A slightly more developed version is where you do have 3 axes, but a fixed isometric perspective – often called 2.5D. However by 3D we mean something *more* than just this.

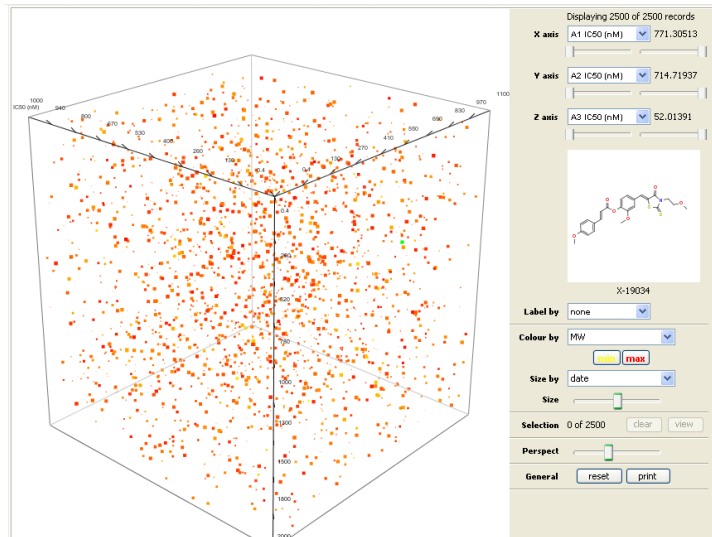


A 2.5D fixed perspective “3D” plot

<http://www.eecs.wsu.edu/~cook/dm/lectures/l14/3mde mo.htm>

Datacubes are the most common form of 3D plot. Instead of looking at a 2D axes on a screen we see a 3D axes, each axis forming the edge of a cube. The application then provides controls which let us rotate and pan this cube, and zoom onto it (and sometimes into it) in order to view the data in different ways.

The primary advantage is that we now have a third axis on which to plot our data. So for instance we can show time as well as latitude and longitude, or dividend as well as capitalisation and share price.



A datacube for scientific data visualisation

http://www.dotmatics.net/images/datacube_large.png

However in our experience, and those of many others, such data cubes have problems:

- It can be hard to maintain orientation as you twist and turn the data in front of you
- You always feel like you are looking in on the data, rather than being able to explore inside it (even when you can actually zoom into the cube)
- The view is often small as a lot of screen real estate is taken up by controls and subsidiary data
- The cube feels “claustrophobic” with the data packed inside a finite space with limited scope to extend the axes.

We refer to these data cubes as “non-Immersive” 3D as they provide no sense of immersion. The data cube is something separate from you, something you always feel outside of, and certainly not something you experience.

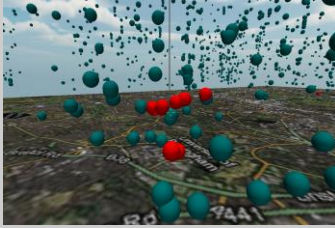
Immersive 3D

An alternate expression of data is in an immersive 3D environment.

We would characterise an immersive environment as one in which:

- You have the sense moving through the space, rather than it moving around you

Time and Geocoded Data



This image shows a sample (about 20,000 points) of Twitter data geocoded over a map of Birmingham. The vertical dimension represents time, oldest at the top, newest at the bottom (so new tweets appear directly on the map).

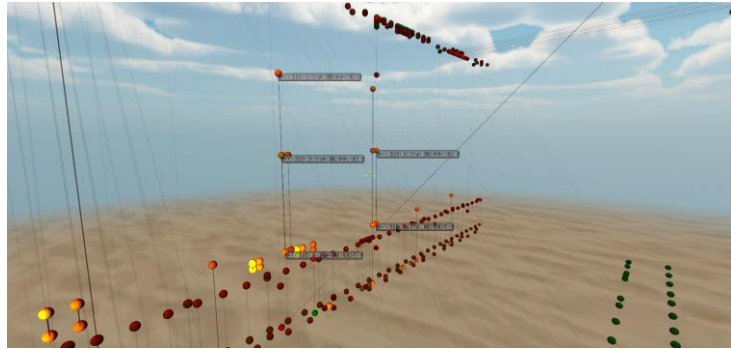
In the image above we have identified an interesting group of tweets, and then had Datascape highlight all tweets by the same use (in red).

If we then show only this users tweets (now in blue), and have Datascape link them in time order (the pink lines), we can rapidly begin to understand that users journey through the day.



- You can move where you want, taking up any viewpoint, inside or out, looking inwards or outwards

What is important about the first definition is that that is the way the physical world works – we move through it, not it moving around us.



Firewall traffic visualised in immersive 3D using Daden's Datascape application

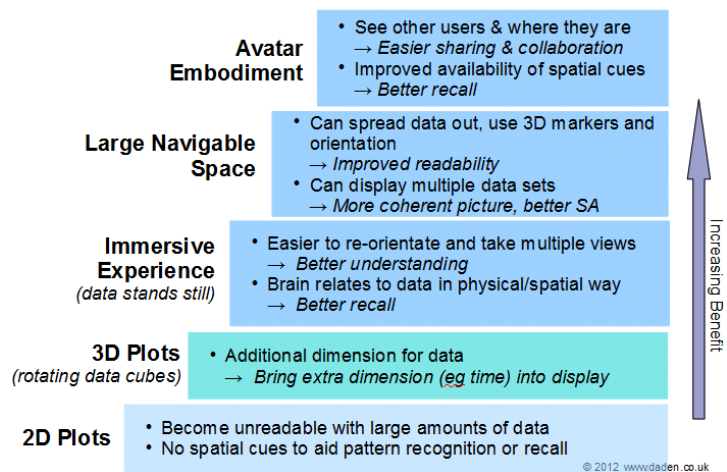
You will also note from the full-screen image capture above that in the immersive space we also minimise any additional text or displays so that the whole screen is focussed on the data and the immersive experience – in Tufte's terms we are maximising the “data ink”.

We think that these features are at the core of the improved performance that we describe below - when in an immersive system the data behaves in the same way as the physical world. It performs in the way that the brain expects, so we find it easier not only to maintain orientation as we move, but we also find spatial cues to aid in pattern detection and analysis, and even in the recall of the scene at a later date.

As the NVAC Thomas & Cook report mentioned at the beginning says, we are taking “**advantage of the human eye's broad bandwidth pathway into the mind to allow users to see, explore, and understand large amounts of information at once**”.

Benefits Escalator

From our experiences with 2D and 3D data visualisation we have developed a “benefits escalator” which illustrates the advantages to be gained as you move along a spectrum from standard 2D visualisations, to fully immersive and collaborative 3D visualisations.



The steps on the escalator are:

• 2D Plots

Conventional 2D displays rapidly become too dense as more points are plotted, or can be tedious to zoom and scroll through, and provide no real aids to recall other than the overall pattern.

• 3D Plots

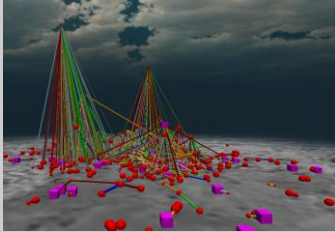
By making use of the third dimension in our plot we can immediately add another set of parameters to our data – common examples being to add time to a geographic display, or to add time to a sales-by-business unit plot. In conventional systems – and especially when targeting a printed output, a 2.5D and/or data-cube approach is taken, where the viewpoint is either fixed at an oblique angle, or the user can rotate the data cube around (and maybe even zoom), but typically cannot step “inside” the data.

Such data cubes have not been typically well received as they can be disorientating with familiarity and recognition issues when rotated to different angles, and often with no spatial cues as to which way it's facing. They can also feel claustrophobic, typically being squeezed into a small amount of screen real-estate.

• Immersive Experience

Having worked in immersive worlds we are used to not only being able to go anywhere in the space – usually by flying – but also to the space behaving like the real worlds – it's us that moves and the objects within the space stay static. Anecdotally it is this that people like about viewing data as an immersive experience – the data stays still but you are free to fly through and around it, rapidly taking up lots of different viewpoints without the same level of disorientation that people find with 3D

Protein Interaction Networks



This image shows the protein-protein interactions within a plant cell. About 700 proteins and several thousand interactions were identified, but when the scientists tried to plot the resulting interaction network in 2D they ended up with what they called a “ridiculum” - just a mess of lines with no meaning.

In Datascape we can plot the data over a large subjective space. The vertical axis is showing frequency – how many other proteins each protein interacts with. Each points size is its age – how long has it been in the cell, and each link is colour coded against the 6 tests which were carried out. A force graph was then used to lay the data out so that the large clusters were readily visible, as were the proteins not connected to the main clusters, and other smaller cluster.

Arabidopsis Interactome Mapping Consortium (2011) Evidence for network evolution in an Arabidopsis interactome map. Science 333(6042): 601-607.

data cubes. The brain likes immersive experiences *because* they mimic the way the real world works.

An additional benefit is that the brain likes using spatial cues as an aid to memory and recall – just think of the use of memory-palaces to enhance memory power. When we look at a 2D screen or sheet of paper, or even a 3D data cube, we don't get those spatial cues. But when we explore data in an immersive space we do. Even though we don't have an “avatar” in the space we think of ourselves as being at a particular point in the space and remember the arrangement of key data points around us.

In summary, experiencing data as a 3D immersive experience helps us to better understand the data through easier exploration, and to remember the data and analysis longer due to the increased use of visual cues.

• Large Navigable Space

We talked earlier about how data cubes are almost inherently “small” - as they need to be rotated within the confines of the screen. However immersive spaces are inherently large, possibly infinite. This has three benefits:

First, it means that we can spread data out so as to de-clutter the display or maximise the uses of the axes. This immediately helps us counter one of the biggest problems with 2D displays. We also find that the zoom and pan features in immersive worlds are typically far smoother, more responsive, and intuitive than similar controls on 2D displays.

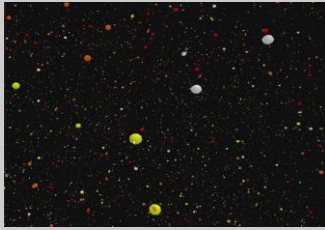
Second, we are no longer restricted to just one data display. We can fill the space with multiple displays of widely varying types – but having all that data in one place can significantly aid the analysis process, and for live activities our situational awareness.

Third, the 3D perspective within a large space allows us to keep distant data in sight, where it would have been scrolled out of the 2D view.

• Avatar Embodiment

It is only at the final stage, when we want to collaborate with others, that an avatar – a representation of ourselves in the immersive environment, becomes necessary. Having each user in the space represented by an avatar brings two major benefits. First we don't

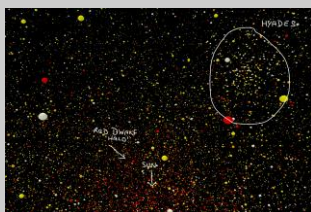
Astronomy Data



Plotting star (and exoplanet) data is a fairly natural application for a 3D tool such as Datascape as the data is inherently 3D. We've used the HYG catalogue which shows the brightest 120,000 or so stars visible from Earth and plotted them in Datascape, using the colour mapping to reflect the stars temperature/colour.

Within the data two big features stood out at once. First the Sun is surrounded by a dense halo of small red stars – red dwarves. These are common across the galaxy, but their dimness means that we only see the ones closest to us – so we actually need to image the rest of the plot filled with myriads of these cool red suns.

Second the Hyades group of hot young white stars is a very obvious cluster in the data.



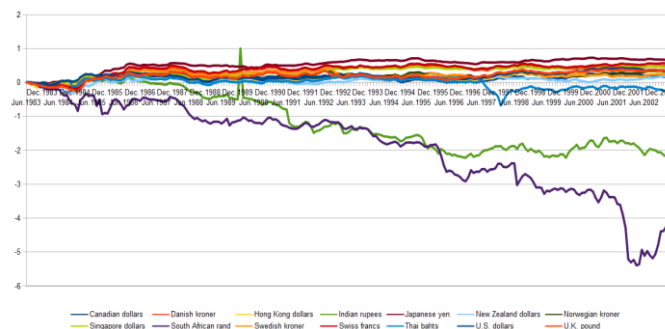
For the future we plan to start looking at Exoplanet data in Datascape.

need to explicitly communicate what we are looking at, or what we find interesting in the data, our colleagues can tell just by looking at where our avatar is going and where it lingers. Secondly when our avatars meet by an interesting piece of data the spatial cues of avatar positions and data positions again aids the retention of the discussion and the analysis.

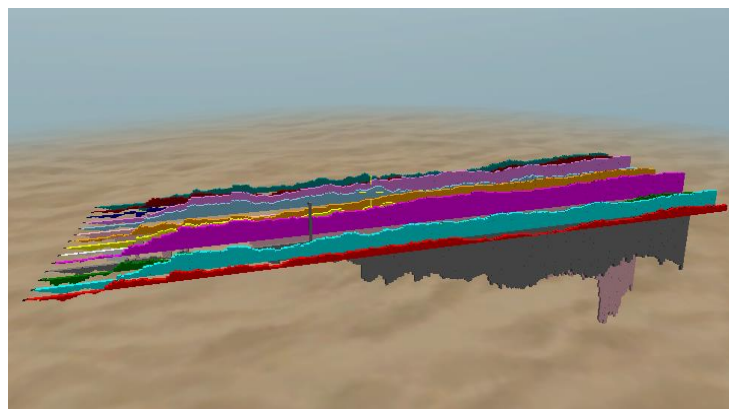
The avatar provides only the visual component of the collaboration. The space could also support voice and text chat for more conventional communication, and the visible tagging and commenting on features of interest as a way of further enhancing the collaboration.

Coping with Data Ranges

An example of another benefit that we have found is that in 3D you don't get the compression of data that you often see in 2D when one (or more) of the data sets needs an axis far greater than the others. For instance the compare the 2D graph from Excel for exchange rates with the same data in 3D:

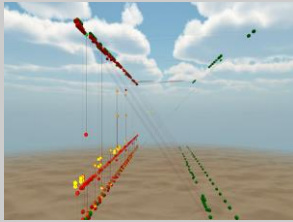


2D plot of 20 years of currency exchange rates for 20 countries



3D plot of 20 years of currency exchange rates for 20 countries

Firewall Network Traffic



This visualisation uses data from the log of a firewall – showing all the computer to computer interactions passing through it.

In this case time is “into” the display, and each point represents a computer sending or receiving a data packet. A single computer actually shows up as a row of points going “into” the tunnel – as it shows each computer transaction separately. A computer’s position on the X & Y axis is based on its IP and subnet address. The points are colour coded by the port in use, and the lines coloured by the protocol in use. Points could also be sized to reflect the amount of data, and textured or shaped to reflect whether they are sending or receiving the data.

With this sort of display it can be easier to identify the patterns of “normal” behaviour, and then the anomalies which may be worth further investigation – or scripting a routine for automatic monitoring – now you know what you are looking for.

In 2D the relatively stable currencies get “bunched” due to the wide variations in the others (in this case the Rand and Rupee). But in 3D we can still easily see the smallest variation in the stable currencies whilst also seeing the large variation of the Rupee and Rand.

Evidence and Research Base

There have been a number of previous studies that have looked at 2D vs 3D and even immersive vs non-immersive 3D visualisations.

The manner in which human visual systems analyse images is of paramount importance in the design of data visualizations. It was discovered that a limited set of visual properties are detected quickly and accurately by the visual system. These visual properties; shape, colour and position, are known as pre-attentive visual stimuli and play an important role in improving the quality and speed of visual comprehension. [TRIESMAN, 1985].

Klatzky identified a distinction between allocentric and egocentric visualisations [Klatzky, 1998]. Allocentric visualisation relies upon information about the location of one entity in relation to another entity. This contrasts the recent advent of egocentric visualisation which represents the location of entities in relation to some 'ego' or self. The 2D and 2.5D/3D data cube visualisations described previously are allocentric as the user is not granted any form of self with which to form egocentric views.

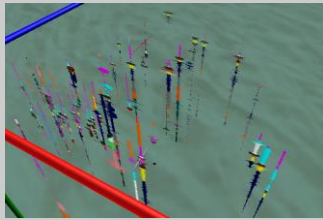
Klatzky's studies indicated that scene and data parameters can be computed more accurately in an egocentric view, suggesting that there may be benefits from egocentric visualisation.

Egocentric views are also known as the tethered perspective. McCormick identified that they gave advantages such that the user does not need to perform additional mental rotations or to interpolate position when navigating the visualization. [McCormick, 1998].

Bringing the research into the domain of game engines (as used by Daden's Datascape v1) the University of Auckland, used the Quake 3 game engine to create an immersive visualization of computer source code found that:

“The tool which was implemented in this project seems to work well as a source comprehension tool, and [provided] the ability to use one's spatial memory to remember the structure of the code.” [Kot, 2005]

Geophysical Data

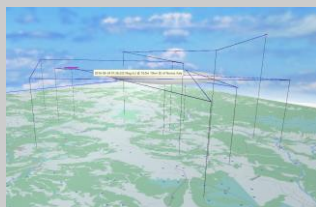


This visualisation was developed as part of an InnovateUK project where we worked with consultants IGI to visualise data from oil exploration bore holes.

Each column above is a separate bore hole, and data is plotted at the correct depth for each sample. The colour shows the rock strata found at each depth. The size of each point is then stretched in 3 dimensions according to 3 key oil indicators, the HI, OI and PI indicators.

The visualisation lets even a non-expert rapidly identify areas where one or more indicators are high. More interesting though are situations where there are big changes in the indicators, often in the same rock, from one sample to the one above or below it.

The example below shows the tremor data for the Aug 16 Italian Earthquake – here height in time, and tremors are linked in time sequence.



Research with Aston University and MOD

In 2010/11 we obtained funding from the Ministry of Defence's Centre for Defence Enterprise at the Defence Science and Technology Laboratory to work with human factors experts at Aston University to do an A-B comparison of data analysts in immersive and non-immersive 3D environments.

We felt that to directly compare 2D and 3D immersive was too great a step which would introduce too many uncontrolled variables (although it's something we'd like to do in a larger study). To further minimise the variables we also recreated a non-immersive data visualisation tool using an immersive environment, rather than using such a tool (e.g. Matlab) directly.

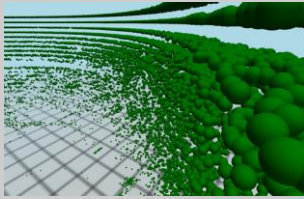
We used OpenSim to create both the immersive and non-immersive environment. The data was simulated scientific experiment data and students were asked to identify the centres of clusters, identify outliers, and characterise the changes in data (in terms of its clustering) between one data set and another. We had around 30 participants, all drawn from post-graduate and post-doctoral researchers who had some familiarity with data analysis. The group was split into two, one half doing the immersive visualisation and then the non-immersive visualisation, and the other group the opposite. Datasets for the tests were randomised from the same base characteristics, so whilst the generalities were similar the specifics that we (or rather the participants) were measuring, were different.

There were four main findings from the experiment:

- Accuracy

Participants were asked to identify the data point closest to the centre of a cluster of data points. The graph below shows the average error – the distance between the true centre and the centre chosen by the participants (remembering that they were choosing the closest data point and there was usually not a data point at the exact centre).

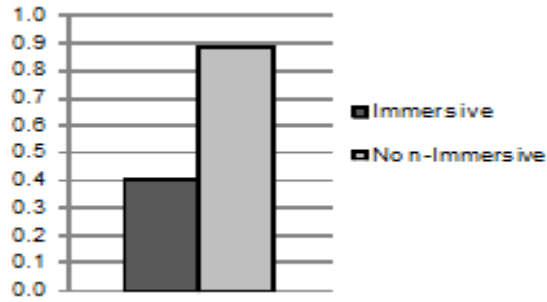
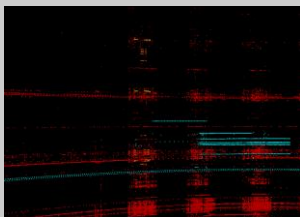
Cylinder Plots



We have found cylinder plots a very useful way to look at a variety of data. We usually wrap time around the cylinder, and then user or target in the vertical domain. This means that by placing yourself in the centre you can easily pan around the whole of the time period.

One useful trick (as in the above network traffic data) is to plot time around the cylinder independent of day – so an event that takes place at a particular time appears in the same place regardless of day – so you can see things like log on/off patterns or attacks made at particular repeating times.

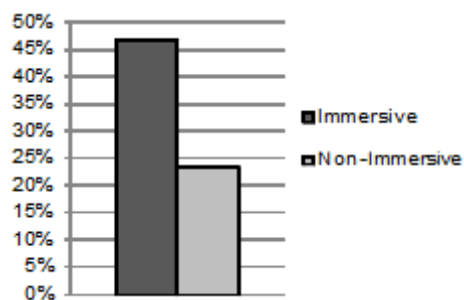
The image below shows network DNS traffic. Here several months of data is shown, with IP address in the vertical dimension, and points colour coded by DNS lookup type. The vertical “bands” are the 9-5 daily activity, grouped in batches of Monday-Friday.



The results showed that the average normalised error across all the tests in the non-immersive space was only 0.4, whereas in the immersive space it was nearly 0.9. In other words in the immersive space accuracy more than doubled. We postulate that this was due to the easier navigation in the immersive environment, with users being able to easily take up multiple views and to maintain orientation whilst doing so, and being able to centre the display on candidate points and rotate the other data around it.

- Change Detection

Participants were shown one data set and then a second, which had some sort of variation in the clustering – by number of clusters and/or density and definition – which they were asked to identify and characterise.

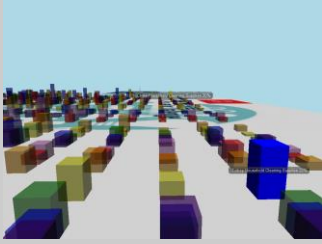


Here we plot the number of respondents making the correct identification of the change in the immersive and non-immersive spaces. In the immersive space over 45% of participants correctly identified the change as against only just under 25% in the non-immersive case – i.e. again the immersive environment resulted in almost a doubling of accuracy.

- Time

As well as capturing the decisions made by the users we also measured the time it took them to come to a decision.

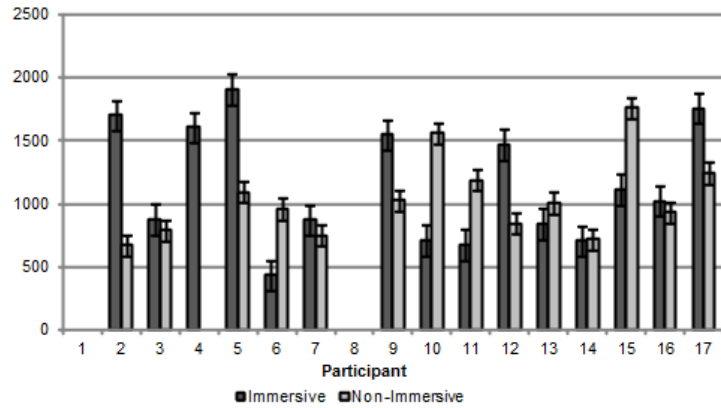
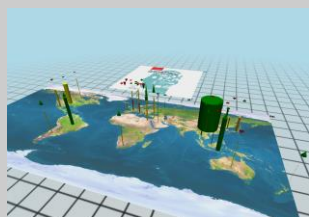
Retail and Consumer Data



With so many variables involved in consumer behaviour and market research data being able to bring that data into a 3D space can provide very powerful opportunities to gain powerful new customer insights – and to then communicate those insights in an equally powerful way to senior managers and clients.

The image above is a 3D bar chart showing consumer preferences across global markets and in a dozen or so product areas. The user can fly through the plot along any of its dimensions and readily highlight any important trends or anomalies. Datascape includes a bookmarking feature to save favourite viewpoints, and also lets you control the camera speed so you can slow it down when recording a video flythrough.

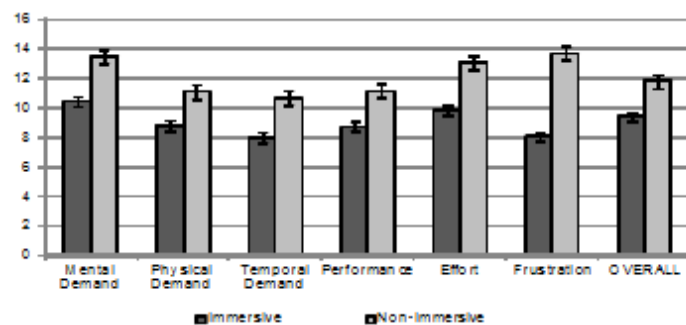
The image below shows the same dataset plotted on the global map.



The results showed that participants actually took *longer* to come to a decision when in the immersive environment compared to the non-immersive environment. Taken with the workload stress results (below) we postulate that because the users felt more at ease in the immersive environment they were happier to take time taking up different viewpoints when analysing the data. We would not take this as indicating that when performing a task within a time-sensitive environment that the immersive environment would necessarily lead to a slower analysis.

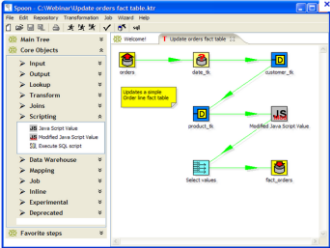
- Workload Stress

The final measure was a standardised workload stress survey done at the end of the task. This assesses how much stress the participants were under when doing the work.



Again the immersive environment shows a consistently better score with participants being on average under about 20% less stress. We take this as due to their being more “at ease” in the immersive environment, having a very high degree of freedom of movement and viewpoint, with very few limitations being put on them by the interface.

Kettle



We've found that Kettle is a useful tool when it comes to trying to get data out of one (or more) data repositories and bring it in to Datascape. Whilst Datascape will import from a generic CSV or Excel file, it expects to have one record per row, and if you want network data then you need to split a single network file into separate node and edges files (as is common with several network graph tools).

Whilst we could do all this in C# or Perl its often very messy and undocumented and hard to share. With Kettle you get a very visual drag-and-drop interface to build your transforms, and to do any data cleansing or standardisation.

We will be posting Kettle "recipes" to our blog to help in getting data from various systems (including Hadoop and social media feeds) through Kettle and in to Datascape.

Web site:
<http://kettle.pentaho.com/>

- Conclusion

Whilst we accept that this was only a very initial look at the comparison the results seem significant enough to suggest that there is definitely something in the difference between immersive and non-immersive visualisation which is worth exploring and exploiting.

Characteristics of an Immersive 3D Visual Analytics system

So what are the important characteristics of an immersive 3D visual analytics system? From our own research and experience we would suggest that the following are the top 10 characteristics:

- Offer a fully immersive experience, allowing the user to feel like they are flying and giving the experience of data standing still and the user moving
- Minimise the user-interface, again to maximise immersion
- Allow maximum flexibility in how fields are mapped to features – don't assume anything - and allow rapid remapping of fields to features, so that the user can quickly find the visualisation that is right for the data and the question
- Support scripting for complex mappings
- Allow the user to see the history of their transforms, and back up to a previous state when needed
- Support a wide variety of plot-types, included geocoded and network graphs, as well as traditional scatter and bubble plots
- Allow user defined images and 3D shapes for data points
- Allow users to see each other within the environment, to be able to interact and to share the analysis
- Make avatars optional, as many business users find they get in the way

Datascape v1

Datascape was an application built by Daden Limited to enable users to gain the benefits of immersive visual analytics. Datascape embodied the experiences of Daden and others in the visualisation of data within

Big Data

We should probably add a word of caution about “big data”.

Big data typically means having Terabyte (10^{12}) or Petabytes (10^{15}) worth of data.

Visualising all that much data at the same time is essentially impossible given the resolution of most computer screens (1440 x 900 is only around 10^6 pixels, full HD only twice that).

So to visualise big data we need to do some form of pre-processing to either just extract the data we want to see, or to combine data through techniques such as dimensional reduction (e.g. Principal component analysis) to get it down to a manageable size.

Interestingly our last (Sep 12) survey into data visualisation suggests that whilst many people are using datasets of 10^6 or more records they are actually only interested in plotting a subset of that data – 10,000 to 100,000 points at a time.

That said we've had datasets of over 15 million data points visualised in Datascape v2!

So visual analytics, immersive or otherwise, are only another tool in the big data armoury, and not a panacea, or even a tool which would stand alone.

immersive worlds, and built on the results of the MOD and InnovateUK research detailed above.

Datascape v1 was launched in 2012, and we used it on a wide variety of projects, visualising everything from cyberdata to windfarms. It was written using the Unity3D game engine as that was the sort of environment we were familiar with. However, in practice we found three big drawbacks of Unity for this sort of application:

- It was hard to integrate with “standard” business IT systems (not many people want to link SQL databases to Grand Theft Auto!)
- It was hard to use the Unity GUI tools to create an interface which was familiar to business IT users. It always felt like a game interface.
- We maxed out at around 100,000 data points, but many users said they needed to be able to visualise millions of points.

So part funded by an InnovateUK project looking at the visualisation of geochemical data we embarked on the creation of Datascape v2 – writing it in native C# and DirectX, and using the standard Windows WPF user interface. This would let us address all 3 of the main shortcomings of Datascape 1 – although at the cost of becoming Windows only.

Datascape v2

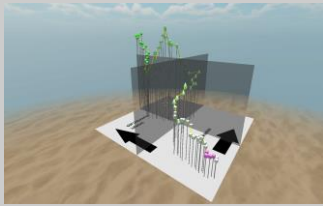
Datascape v2 was a downloadable application which installed quickly and simply on Windows PCs. It let you import data directly from Excel spreadsheets and CSV files (with options to link to SQL and soon NoSQL databases), and through a set of workspace and mapping templates and smart field mapping let you get your data on the screen in a 3D environment as quickly as possible.

Datascape let the user map any field from the data onto any of the “features” associated with a plottable data point. At the most basic those features included:

- X, Y and Z position (essential for 3D!)
- size (which you might want to vary in x,y,z)
- shape
- colour
- label

... essentially Triesman's pre-attentive stimuli.

Economics Data



Here we have used Kettle to merge data from three sources with a common key of financial quarter from 1984 to 2009 and then bring it into Datascape. The three dimensions are:

- Bankruptcy rate (X/across screen)
- GDP (Z/into screen)
- FTSE(Y/up screen).

We then used Datascape's network generation feature to draw the line through them, and added some panels to divide the data into quadrants for high/low bankruptcy, and high/low GDP. Data points are colour coded for quarter/time (dark/old to light/recent).

When we looked at the data we identified 6 distinct periods, reflecting different relationships between the variables, and different stages in the UK economy and then used manual point select to group these and show each period in a different shape.

This already gave us 8 potential “dimensions” with which to plot our data, plus the human readable label.

More advanced visualisations could then add:

- orientation (in 3 axis)
- image/texture

as well as adding additional “rich” information:

- x/y/z axis labels per point
- clickthrough URL for drilldown
- network graph “edges”/lines between nodes

With Datascape v2 we could plot over 15 million data points on screen, although this depended on the power of your machine, the density of the data and the complexity of the 3D objects that you were using for each data point.

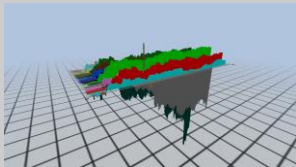
But getting the data onto the screen is only half the task – the key point about a visual analytics system is that you can manipulate the display in order to extract the sense, meaning and stories behind the data. With Datascape you could:

- Remap the data in seconds so as to get the best field-feature mapping to extract maximum value from the information
- Search using full boolean logic to highlight related data
- Join related data points with network lines
- Highlight selected points (from manual selection or search) by changing colour, shape and size
- Drill down to related data or information
- Bring in multiple data sets into the same space
- Adjust the amount of “3D” effect
- Scrub and animate data between times – or any other values
- Use planes to segment the display to help highlight quadrants or layers of interest
- Use OpenStreetMap to plot data against a map for anywhere in the world

Datascape v2 began to give us an idea of the functionality needed by an immersive visual analytics system, and again we used it on a wide variety of projects between 2014 and 2018.

Template Library

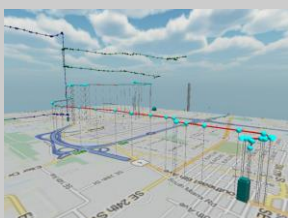
The images below show some of the standard templates mapping in Datascape – but you are always free to create your own.



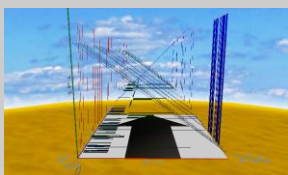
Bar Chart



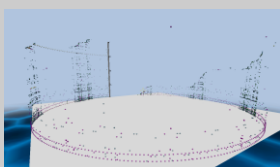
Globe



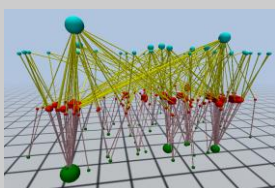
Geotime



Time Tunnel



Cylindrical



Manual/Network

3D Plot Types and Examples

The sidebars show a variety of different 3D data plots created using Daden's Datascape immersive visual analytics tool. They show the wide variety of plots which could be generated including:

- 2D and 3D spatial/geocoded
- 2D geocoded plus time
- 3D scatter plot – e.g. scientific data
- 3D bar charts – e.g. financial data
- 3D object mapped data – e.g. to globe or body
- 3D time lines and time-tunnels – e.g. computer network traffic
- Network graphs – e.g. from social graphs and protein data
- Layout charts – e.g. from DNA data

Publishing to the Web

One of the downsides of Datascape v1 and v2 was that it was hard to share the 3D visualisation unless you found someone else with Datascape and then gave them the data and the mappings. What we (and customers) have always wanted was to be able to share and experience immersive 3D visualisations on the web. The emergence of WebGL has made that possible, and with Datascape v2.GL we made it a reality.

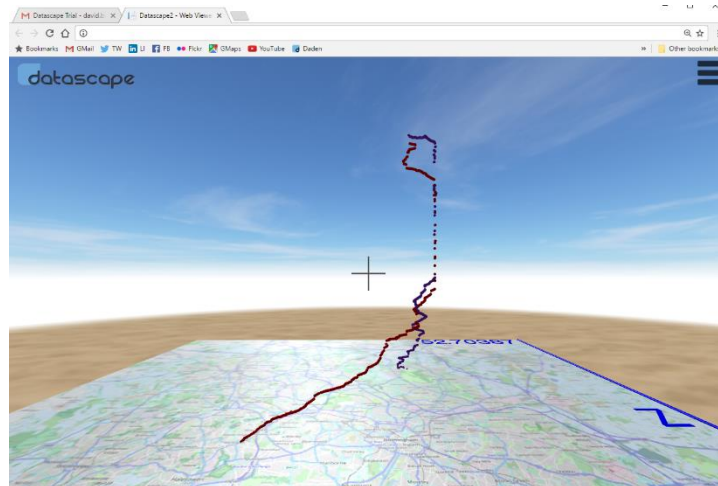
When you selected “Publish to Web” from Datascape your current visualisation (including any applied filters) was published to our web server, and you were given a URL from which you, and anyone you give the URL to, could access the visualisations. Pasting the URL into your browser loaded up a cut-down version of the Datascape viewing application, and then loaded the data into it. The user could then fly through the data and examine the hover text on points just as though they were using the desktop application. And since we were using WebGL you could use Safari on a Mac, and web browsers on tablets and even smartphones to use the data.

However WebGL is a less performance tuned environment and so the following limitations applied:

- The WebGL viewer would only support up to around 10,000 data points, depending on your device (possibly only 2,000 on laptops/phones). We put a hard cap on 10,000 data points for the export, to both preserve performance and to conserve the use of the free data storage we are providing.

- We only exported the data need to build the plot (x,y,z, shape colour etc), NOT the dataset itself – so the user only saw the visualisation you created – they couldn't change it and they couldn't drill down to see the data behind each point (so any key info needed to be in the hover text)

So it wasn't a complete solution, but it was a big step forward – letting you (and us) share 3D immersive visualisations as easily as we do 2D ones.



Datascope WebGL Viewer

Virtual Reality

One of the big changes in the last few years has been the increase in interest in, and accessibility to, virtual reality. We were an early adopter of Oculus DK1 and soon had it working with Datascope, and upgraded to the DK2 when that came out.

From our experiences with DK2 we'd say that using VR to view data is an interesting experience, but I can't say we're totally sold on it yet as an everyday analysis tool. The big issues as we see it are:

- Disorientation/motion sickness – you are moving around a lot in a space with a minimal ground, so not for those with a delicate stomach
- Your field of view is very constricted, so you're moving your head even more – so yet more chances for nausea!
- It's very hard to read text in VR, especially at the small font sizes used in data labelling and spreadsheet views – so good for the impression of the data but not for the fine detail.

But we're sure you'd like to be able to make your own mind up, so....

Datascape 2.VR

Having played with DK2 and Oculus CV1 we actually decided to change tack, and support Google Cardboard as our initial VR device. There were two key reasons for this:

- At £10, almost anyone can afford Google Cardboard and try 3D data out in VR if they want to, far more than could afford Oculus Rift.
- Since we were doing a WebGL publisher/viewer, making a mobile VR version was almost trivial.

What this meant was that when you publish a visualisation to the web as WebGL from Datascape it is instantly also available as MobileVR on Google Cardboard. Just select the VR option on the viewer application and it will go into split-screen mode, put your phone into Google Cardboard and away you go!



WebGL Viewer in VR mode on an iPhone

One of the challenges with Cardboard is how to navigate – you've only one button and selecting off a menu would be very cumbersome. So pressing the button cycled you through a move forward – stop – move backwards – stop sequence. It's not ideal, but worked pretty well with the way that people explore immersive data.

Interface Experiments



Whilst applications like Datascape just operate using conventional PCs, display screens and input devices, there is inevitable interest in how immersive visualisations can be brought out into the physical world. Some of the things we've already used over the past few years include:

- Using CAVE style 3D screens to create a true 3D display (with or without 3D goggles)
- Using Kinect type sensors to enable gesture control and navigation of the data
- Using sound as an additional plot feature and/or navigation device
- Using a MIDI Mixer (see above) as a tactile visualisation controller (with both hardware and tablet based controllers)

Our to-do list for the future includes using data-gloves and other haptic interfaces to make the data tactile (force/resistance as another data point feature), or as an alternative/addition to Kinect or VR for interaction.

DatascapeXR?

With Datascape we seemed to be a good 5 years ahead of the curve, and whilst there was always a lot of interest in it and the visualisations it produced, and a lot of “starter” projects we never really got the volume sales to maintain it. 3D Immersive Visual Analytics is still something I find fascinating though, and I think has real potential, so I'm currently working on DatascapeXR. This will be a web-native application, written in WebXR so that you can view the data either in DesktopVR mode (i.e. on an ordinary PC or tablet) or HMD-VR mode (e.g. with an Oculus Quest in the browser in full 3D).

My first tentative steps towards this are on a game related visualisation – show stars in a space sector – but the whole visualisation is data driven and so has many of the key building blocks I need. If you have a WebXR compatible headset (e.g. Oculus Quest) you can view it live at <https://www.daden.co.uk/webxr>. The plan is to have some sort of DatascapeXR demonstrator available in early 2022.

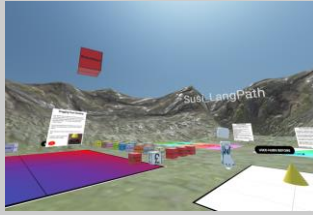


Visualising data in an early DatascapeXR prototype

Full Circle – Social Virtual Reality

We started this paper talking about our work visualising data in the Second Life virtual world. With Social Virtual Reality (SocialVR) we have come full circle. SocialVR are places accessed through VR where people can meet and collaborate in a fully realised 3D space – essentially Second Life + VR (or more realistically 3D chatroom + VR). One of the more interesting SocialVR systems was AltSpaceVR since in their original pre-Microsoft-takeover version they offered “enclosures”. An enclosure was a space within one of their environments – for instance the space above a board-room table, or the whole of the floor of an exhibition hall. As a user you could assign a URL

Emerging VR Spaces



We are currently tracking over 30 emerging and existing VR spaces – see our interactive table at <https://airtable.com/shrCMIXC2BysCng8j>.

Mozilla Hubs (above) is one of the most interesting as it runs in the browser – so no download needed. Although you can't dynamically load data yet it does make for a great ad-hoc collaborative meeting spaces. A similar offering is Frames (shown below).

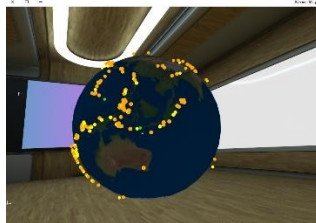


The worlds that do offer more capability (e.g. AltSpaceVR, VRChat) tend to be download base, need Unity for any world creation, and have avatars that are possibly too cartoon-like (see below) for many business users.



to that enclosure, and if that URL pointed to a WebGL or A-Frame (cut-down WebGL) visualisation, then your visualisation appeared inside the enclosure, in the AltSpaceVR space!

Well since we were already generating WebGL from Datascape2, creating an A-Frame version was almost trivial, and we demonstrated that we could take a Datascape visualisation and see it realised in AltSpaceVR and walk around and through it as a team and discuss it in audio – a truly immersive 3D data visualisation experience.



USGS Earthquake data in AltSpaceVR via Datascape



Collaborating between Spain and UK in AltSpaceVR

Trainingscapes, our (now discontinued) 3D immersive Learning environment was also very Second Life like, and a sort of SocialVR space so we showed how we could use that for social immersive 3D visual analytics too.



Country data plotted in Fieldscapes via Datascape



Visualising DNS data in Fieldscapes via Datascape

There is now (2021) a whole crop of emerging VR worlds (our database has 30+) and we are actively looking at which of them can be used for social 3D data visualisation. The big thing that most of them lack is any sort of scripting or API, and until those barriers are addressed then there's not too much we can do – hence the work on DatascapeXR – but the platforms we're really watching are again the web delivered ones such as Mozilla Hubs and Virbela's Frames.

Who are we?

Daden Limited is the consulting company of David Burden, specialising in virtual worlds and virtual humans.

David has been working with immersive worlds and artificial intelligence technologies for over 20 years, and his clients have included Government departments, city and local councils, educators and health providers and private sector organisations in the UK and abroad.

David is based in Birmingham, England, but can work virtually anywhere!

What's In It for Me?

So why should you be interested in immersive visual analytics?

The bottom line has to be that immersive visual analytics will help you understand your data better, and by understanding your data better **you will be able to make better decisions** – whether those are operational decisions, investment decisions or decisions about alternative hypothesis. And those decisions should save cost, increase revenues or reduce risk.

Whilst not every data analysis tasks may be enhanced by immersive visual analytics we believe that there are a considerable number that would be.

The “sweet spot” for immersive visual analytics (IVA) appears to be where the data, or the relationships that you are looking for, are relatively complex, and the data relatively unstructured. You *can* use an IVA tool to look at very structured financial data, and as the quantity of that data increases then the advantages of using an IVA tool also probably increase. But it is when you start combining things like social media data with geo-coding and social graphs, or looking at complex maps of protein interactions, or plotting multiple time and geo-coded data sets that IVA really gets into its stride.

Immersive Visual Analytics applications open up new possibilities as a presentation orientated data visualisation tool, allowing you to “walk” the audience through the data – again enhancing understanding and retention.

Any Immersive Visual Analytics system should be designed to augment your existing database and analytics tool-sets rather than replace them. It should work in harmony with your existing systems, drawing on your existing data cleansing/wrangling pipeline, your existing data repositories, and your existing upstream analysis tool, but then presenting the data in a way that gives the human analyst the best chance of making the right decision.

As the NVAC paper said at the very beginning, this is all about *“taking advantage of the human eye’s broad bandwidth pathway into the mind to allow users to see, explore, **and understand** large amounts of information at once”*.

Glossary

AFrame – a cut down version of WebGL

Allocentric – A view relating one entity to another. Contrast with Egocentric.

Avatar - A 2D or pseudo-3D representation of a person or character

Egocentric – A view relating an entity to your own location. Contrast with Allocentric.

Geo-coded – data with a set of spatial co-ordinates

Geo-spatial – data with a spatial position – typically related to its position on the Earth

Geo-temporal – data with both a time and geographic component

HMD-VR – VR using a headset

IVA – Immersive Visual Analytics

Open Sim – an open-source version of the Second Life virtual world

VW – Virtual World

VR – Virtual Reality

WebGL – Emerging 3D markup language for the web

WebXR – Standard for XR headsets on the web

XML - Extensible Markup Language – a human readable standard for representing data on computer systems

What Next?

You can see many more of our immersive data visualisations (including videos of them) on our web site at <https://www.daden.co.uk/datavisualisationgallery>, where we have also links to many client projects, and videos of some of our visualisations on YouTube at <https://www.youtube.com/playlist?list=PLDF3C6C647C1882E1>.

Any WebXR work-in-progress (whether IVA-related or other) we post to <https://www.daden.co.uk/webxr>.

We will post to the web site once we have a public prototype of DatascapeXR, and we blog about our immersive visual analytics work on virtual worlds (and DatascapeXR progress) at <http://dadenblog.blogspot.com/>.

If you would like to have a demonstration of Immersive Visual Analytics, or just discuss the concepts and issues raised in this paper in more detail, or then please contact me:

- By phone on +44 (0)7811 266199
- By email to: info@daden.co.uk
- On the web at: www.daden.co.uk

References

Klatzky, R.L. Allocentric and Egocentric Spatial Representations: Definitions, Distinctions and Interconnections. *Lecture Notes in Computer Science*, Volume 1404, 1998.

Kot, B. Wuensche, B., Grundy, J., Hosking, J. Information Visualisation Utilising 3D Computer Game Engines. Case Study: A source code comprehension tool. *CHINZ*, 2005.

Liu G, Austen EL, Booth KS, Fisher BD, Argue R, Rempel MI, Enns JT. Multiple-object tracking is based on scene, not retinal, coordinates. *J Exp Psychol Hum Percept Perform*. 2005 Apr;31(2):235-47.

McCormick, B. H., DeFanti, T. A, and Brown, M. D. Visualization in scientific computing. *Computer Graphics* 21, 6 (1987), 1–14.

Triesman, A. Preattentive processing in vision. *Computer Vision, Graphics and Image Processing*. 1985.

Tufte, E.R. *The Visual Display of Quantitative Information*, ISBN 978-0318029924, Graphics Press Connecticut, 1992.

Tufte, E.R. *Visual Explanations*, ISBN 978-0961392123, Graphics Press Connecticut, 1997.