

Towards “Visual Data Science“ with Images and Structured Data

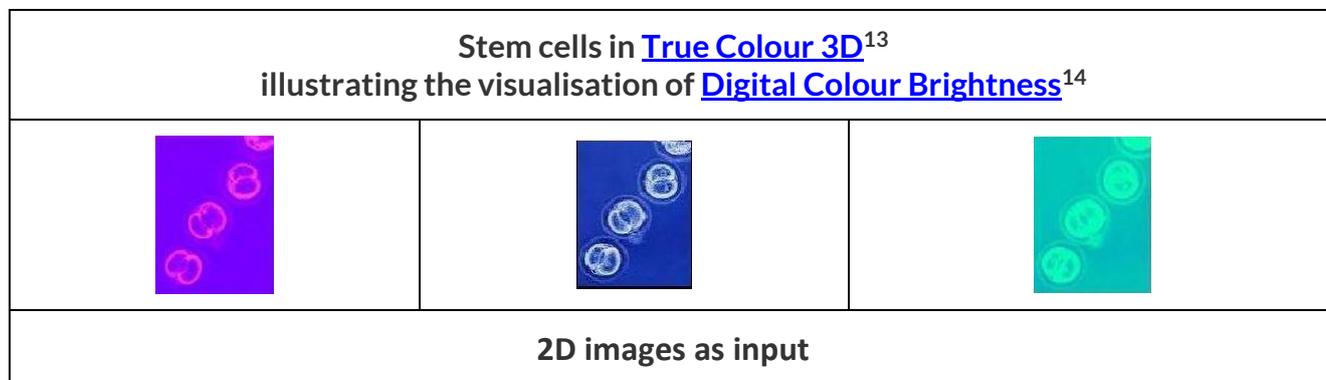
A Proposal to Learn from Data with New Visualisations

[Professor Robert Calderbank](#)¹ of Duke University in North Carolina recently gave the Turing Lecture [Fifty Years of Learning from Data](#)². I [blogged](#)³ about his great overview and the video will be [here](#)⁴.

At the end I introduced myself as a mathematician who used to diagnose software at CERN in Geneva and asked where he sees the trend going for courses on Data Science. He set [Data+](#)⁵ up in 2013 to train undergraduates in interdisciplinary teams. For the Turing Institute and its clients, as a hub for super international leaders, he answered most encouragingly to create data stories which would include the line “*we couldn’t have done this without this theoretical new idea.*”

My new theoretical input has not been published, but is being integrated into a [Smart Knowledge Engine](#)⁶ as work in progress and pre-startup SME. It is being described on [Smart Knowledge Space](#)⁷ in a number of [articles](#)⁸ and expressed in three new [visualisation styles](#)⁹:

1. [True Colour 3D](#)¹⁰ – for re-visualising images as 3D movable objects;
2. [Visual 3D](#)¹¹ – for layering multi-dimensional data along a ‘visual z’-axis;
3. [Chronometric 3D](#)¹² – for juxtaposing short-, medium and long-term trend periods.



¹ <http://ece.duke.edu/faculty/robert-calderbank>

² <https://www.turing.ac.uk/events/turing-lecture-fifty-years-learning-data/>

³ <https://3d-metrics.me/2017/10/19/fifty-years-of-learning-from-data-prof-robert-calderbank-turinglecture-turinginst/>

⁴ https://www.youtube.com/watch?v=y7ehrXZfIS0&list=PLuD_SqLtxSdVDcrCYIHayTL91DapuIHrO

⁵ <http://bigdata.duke.edu/data>

⁶ <http://www.smart-knowledge-portals.uk/>

⁷ <http://smartknowledge.space/>

⁸ <https://smartknowledge.space/articles/>

⁹ <https://smartknowledge.space/innovation/3-visualisation-styles-for-3-types-of-data/>

¹⁰ <https://smartknowledge.space/true-colour-3d/>

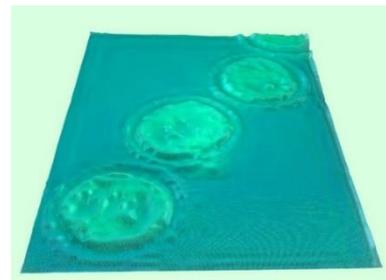
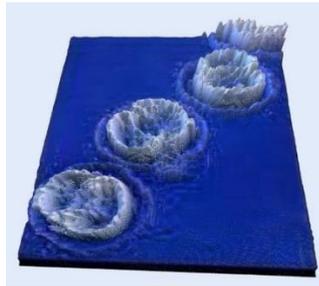
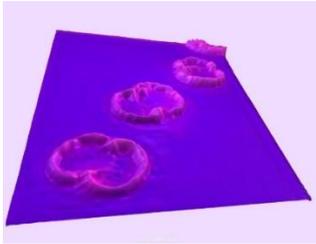
¹¹ <https://smartknowledge.space/investing/compare-contrast/layering-vs-excel-charts/>

¹² <https://smartknowledge.space/innovation/new-3d-visualisation-styles/chronometric-3d/>

¹³ <https://smartknowledge.space/true-colour-3d/>

¹⁴ <https://smartknowledge.space/digital-colour-brightness/>

3D movable objects as output

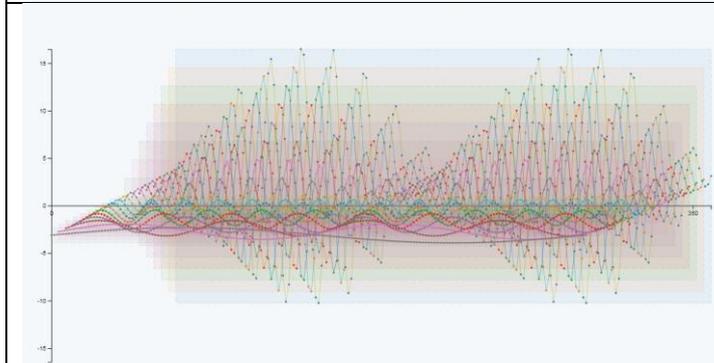
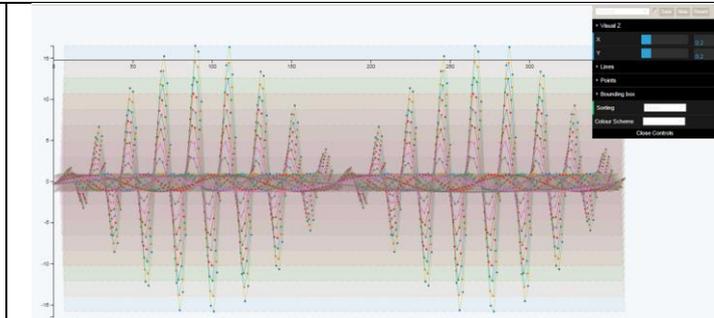
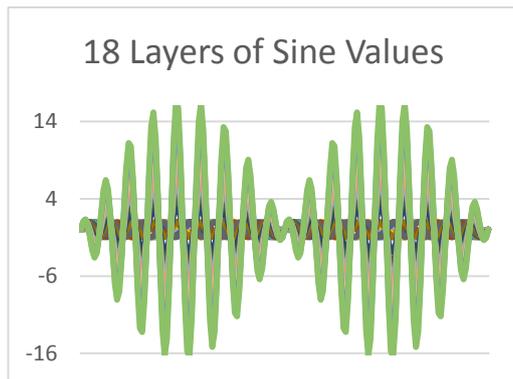


<http://www.smart-knowledge-portals.uk/projects/79>

<http://www.smart-knowledge-portals.uk/projects/77>

<http://www.smart-knowledge-portals.uk/projects/80>

In **Visual 3D**¹⁵ we unclutter into **Layers**¹⁶ the series that Excel draws on top of each other

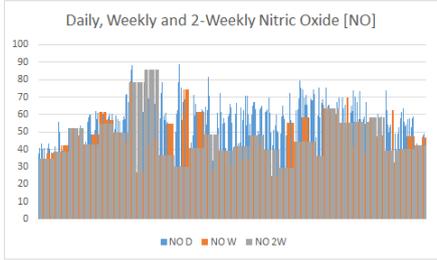
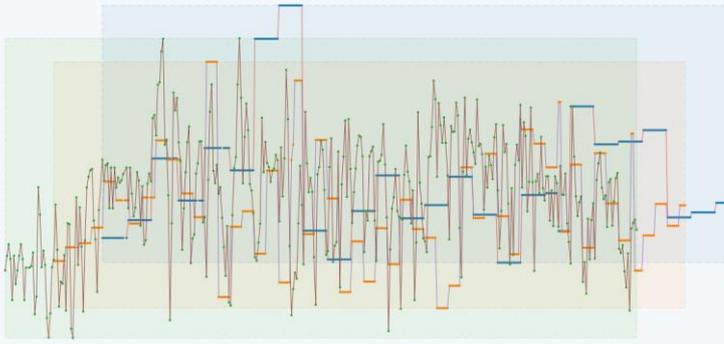
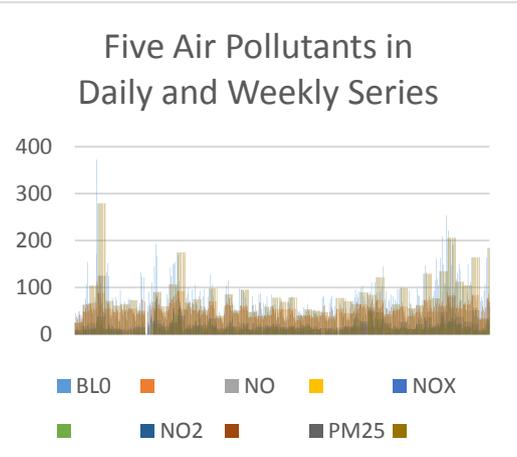
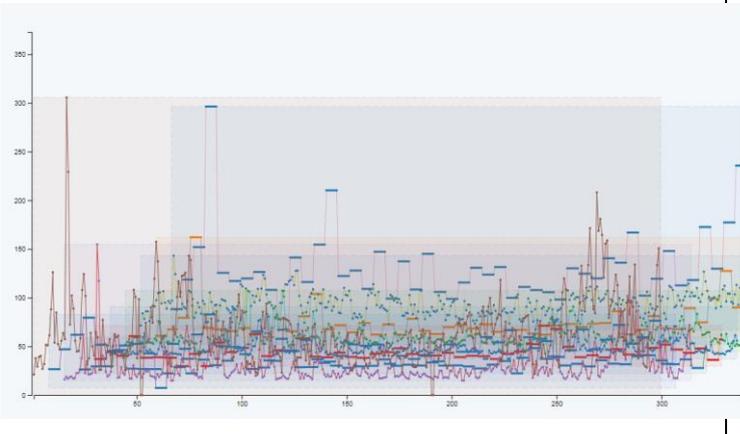


Excel

The same data layered – with controls to vary 'visual comparability'

¹⁵ <https://smartknowledge.space/investing/compare-contrast/layering-vs-excel-charts/>

¹⁶ <https://smartknowledge.space/investing/compare-contrast/layering-vs-excel-charts/>

	
<p><i>Three air pollutants in daily, weekly and 2-weekly series</i></p>	<p><i>The same data layered with controls to 'compare and contrast' visually</i></p>
	
<p><i>Five air pollutants in daily and weekly series</i></p>	<p><i>The same data in ten layers where the colour scheme can be varied</i></p>

To illustrate the interconnectedness of Data Science projects, Professor Calderbank groups them into projects that occupy a position in a matrix during Data+ courses. The results are so positive that I would like to propose them as a template for activities at the Turing Institute to specialise in research based on the use of our Smart Knowledge Engine.

The table below demonstrates the effect of the visualisation of numbers as generic building blocks. They require words and measuring units to communicate meaningful data stories. Using compelling and significant data sets, it will be possible to gain new insights by “compare and contrast”:

The left-hand column is taken from Professor Calderbank’s slide to answer “What is Data Science?” The right hand column contrasts “Visual Data Science”: generic visualisation styles looking for compelling data sets that make new correlations and priorities apparent and visible.

<p>Data Science</p>	<p>Visual Data Science for Images, Multi-Dimensional Data and Time Series</p>
----------------------------	--

1. Marshalling:	2. Domain Experts Select Vocabularies
1.1. Fast Checking 1.2. Imputation 1.3. Merging/Collecting Data	Topic e.g. <i>Smart Cities/Image Science/Financial Trends</i> Domain e.g. <i>Physics/Chemistry/Biology / Medicine</i> Scale e.g. <i>Microscopic / Camera / Satellite / Telescopic</i>
3. Analysing:	First Visual Impressions
1.4. Clustering 1.5. Image Analysis 1.6. Geospatial Analysis 1.7. Network Analysis 1.8. Statistical and Mathematical Models 1.9. Machine Learning	Experts can immediately see what's going on. Sorting similar data and images for finer comparisons will enable researchers and data scientists to make new discoveries. By comparing with 'reference images' and 'reference data', new standards can eventually be set.
4. Visualising and Communicating	Analysis and Preparation for Automation
1.10. Publications / Grants 1.11. Policy Briefs / White Paper 1.12. User Interface 1.13. Integrating into Systems.	Experts set boundaries so that operators can deal with events that are "off limits".

Considering the generic nature of these visualisation styles, the task is to find representative data sets and images that can be used as benchmarks and references, whether for the quality of data capturing, the essence of climate change or the trends of financial markets, besides the comparison between healthy and sick cells, the progress towards health or illness or faulty and counterfeited pills.

In terms of the science of imaging, the challenge is to find the best technology for any given scale. Proprietary techniques translate light and colour into digital values for digital images and videos. To find the best numerical representation of light and colour for any given scale and application will result in new equivalents of the [prototype meter](#)¹⁷. For the meter is now defined by the path travelled by light in a vacuum at 1 / 299 792 458 of a second.

In the spirit of my blog post about Professor Blum's lecture [Alan Turing and the Other Theory of Computation](#)¹⁸, our Smart Knowledge Engine is a new tool of investigation as a set of "software lenses" for "visual data science", moving [from computing to visualising or from numbers to pixels](#)¹⁹ – the [link between digital bits and physical atoms](#)²⁰.

¹⁷ https://en.wikipedia.org/wiki/History_of_the_metre#International_prototype_metre

¹⁸ <https://www.turing.ac.uk/events/theory-computing-can-machine-conscious/>

¹⁹ <https://3d-metrics.me/2017/09/27/from-computing-to-visualising-or-from-numbers-to-pixels-alanturinginst/>

²⁰ <https://smartknowledge.space/articles/visualisation-pixels-the-link-between-digital-bits-and-physical-atoms/>