CS2D7 Data Visualisation
Dr Claire Rocks & Dr Richard Kirk
July 2022
Lecture 1

Introduction
Introductions

• Claire Rocks: C.L.Rocks@warwick.ac.uk
• Richard Kirk: R.Kirk.1@warwick.ac.uk
Today

• What is Data Visualisation?
• Visualisation Design Processes
• Introduction to the module
• Introduction to the assessment
Why?

• Anyone can put data into a graphing application and create a graph
• The tools don’t know what the story is
A note on Tools

- Visualisation and the technology to create visualisations go hand in hand.
- There are many tools, some proprietary, some open source. Some easy to learn but perhaps lacking in functionality. Some are rich in functionality but take a long time to learn.
- This module will focus on the underlying craft of creating data visualisations and build an introductory toolkit for making visualisations using Python and Processing.
What is Data Visualisation?

- The visual representation and presentation of data to facilitate understanding (Andy Kirk)
- A graphical representation of data or concepts (Colin Ware (2013), Information Visualization: Perception for Design)
- The use of computer-supported, interactive, visual representations of abstract data to amplify cognition (Card et al (1999), Readings in Information Visualization: Using Vision to Think)
- Visual displays in which graphical approaches play a central role in communicating information in a meaningful way (Meirelles (2013), Design for Information)
What is Data Visualisation?

- The visual representation and presentation of data to facilitate understanding (Andy Kirk)
- A graphical representation of data or concepts (Colin Ware (2013), Information Visualization: Perception for Design)
- The use of computer-supported, interactive, visual representations of abstract data to amplify cognition (Card et al (1999), Readings in Information Visualization: Using Vision to Think)
- Visual displays in which graphical approaches play a central role in communicating information in a meaningful way (Meirelles (2013), Design for Information)
The visual representation and presentation of **data** to facilitate understanding

- Data is names and amounts, groupings, descriptions and measurements, dates and locations. Most of the data we will be considering will be in textual or numeric form but it could be in media form e.g. images, video or sound
- Data is the fundamental element driving decisions across the design process
We can work out that
- % sales for stores starts high and goes to nothing
- % sales online starts low and reaches 100%
- % sales via telephone is consistently small

What if we want to form our observations quickly? Or work out the first time % sales online > % sales in store?

Synthesising observations from multiple values across different rows and columns to perceive broader relationships does not exploit the capabilities of our visual system

This only gets worse as the volume of data increases

To really see the data we have to represent it in a different form

<table>
<thead>
<tr>
<th>REPORTING MONTH</th>
<th>STORES</th>
<th>ONLINE</th>
<th>TELEPHONE</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 2011</td>
<td>71</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>Jun 2011</td>
<td>72</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>Jul 2011</td>
<td>71</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>Dec 2011</td>
<td>71</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>Jun 2012</td>
<td>73</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>Jul 2012</td>
<td>77</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>Sep 2012</td>
<td>75</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>Nov 2012</td>
<td>75</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>Jun 2013</td>
<td>73</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>Nov 2013</td>
<td>73</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>Jan 2014</td>
<td>73</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>Jun 2014</td>
<td>72</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>Aug 2014</td>
<td>95</td>
<td>44</td>
<td>1</td>
</tr>
<tr>
<td>Sep 2014</td>
<td>60</td>
<td>38</td>
<td>2</td>
</tr>
<tr>
<td>Oct 2014</td>
<td>51</td>
<td>48</td>
<td>1</td>
</tr>
<tr>
<td>Nov 2014</td>
<td>44</td>
<td>55</td>
<td>1</td>
</tr>
<tr>
<td>Jan 2015</td>
<td>52</td>
<td>47</td>
<td>1</td>
</tr>
<tr>
<td>Mar 2015</td>
<td>50</td>
<td>48</td>
<td>2</td>
</tr>
<tr>
<td>Jun 2015</td>
<td>49</td>
<td>49</td>
<td>2</td>
</tr>
<tr>
<td>Jul 2015</td>
<td>37</td>
<td>61</td>
<td>2</td>
</tr>
<tr>
<td>Aug 2015</td>
<td>40</td>
<td>50</td>
<td>2</td>
</tr>
<tr>
<td>Nov 2015</td>
<td>40</td>
<td>59</td>
<td>1</td>
</tr>
<tr>
<td>Dec 2015</td>
<td>22</td>
<td>77</td>
<td>1</td>
</tr>
<tr>
<td>Jan 2016</td>
<td>21</td>
<td>77</td>
<td>1</td>
</tr>
<tr>
<td>Feb 2016</td>
<td>20</td>
<td>78</td>
<td>2</td>
</tr>
<tr>
<td>Apr 2016</td>
<td>14</td>
<td>84</td>
<td>2</td>
</tr>
<tr>
<td>Dec 2016</td>
<td>21</td>
<td>77</td>
<td>2</td>
</tr>
<tr>
<td>Jun 2018</td>
<td>6</td>
<td>93</td>
<td>1</td>
</tr>
<tr>
<td>Jul 2018</td>
<td>6</td>
<td>93</td>
<td>1</td>
</tr>
<tr>
<td>Dec 2018</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>
The **visual representation** and presentation of data to facilitate understanding

• This is the activity of data visualization

• The art of choosing the right representation to show the features of the data that are relevant
You can use this chart to form observations about the 3 sales channels, and you can compare them.

- The comparisons are especially relevant as together they make up 100%.

- We can process clusters of data points simultaneously and identify the slopes and the fats, the peaks and troughs, the gaps and crossovers.

- We lost precision in individual data points.
The visual representation and **presentation** of data to facilitate understanding

- This is everything else that makes up the anatomy of a visualization
  - Interactivity
  - Annotation
  - Colour usage
  - Composition
  - Medium or dissemination method
- Lots of small design decisions that add up to a big impact
Understanding the context

• Visualisation can be used for data analysis or data communication (exploratory/explanatory)

• Visualisation for data analysis is part of the process for creating visualisations for communication
  • The techniques don’t have to be visually polished or necessarily appealing.
  • The visualiser and the viewer are the same person

• When data visualisation is being used to communicate to others, many careful considerations are required regarding the requirements and interests of the audience
Who, What, Why and How

• To whom are you communicating?
• What do you want your audience to know or do?
• Why?
• How can you use your data to make your point?
  • Ignore the unsupporting data?
The visual representation and presentation of data to **facilitate understanding**

Perceiving

- What does it **show**?
  - Where is big, medium, small?
  - How do things compare?
  - What relationships exist?

Interpreting

- What does it **mean**?
  - What is good and bad?
  - Is it meaningful or insignificant?
  - Unusual or expected?

Comprehending

- What does it **mean to me**?
  - What are the main messages?
  - What have I learnt?
  - Any actions to take?
A bit of history

• The concept of using pictures to aid understanding has been around for centuries, from maps and graphs in the 17th century to the invention of the pie chart in the early 1800s.
William Playfair, trade-balance time-series chart, 1786

- Playfair invented several types of diagrams: in 1786 the line, area and bar chart of economic data, and in 1801 the pie chart.
- Chart shows the Time Series of Exports and Imports of Denmark and Norway.
Charles Dupin, choropleth map showing the distribution of illiteracy in France, 1826

- The earliest choropleth map is credited to Charles Dupin, a French mathematician.
- The map uses shadings to show the distribution of illiteracy in France.
The Broad Street cholera outbreak was a severe outbreak of cholera that occurred in 1854 near Broad Street (now Broadwick Street) in the Soho district of London.

Snow was sceptical of the prevailing miasma theory, and proposed that germ-contaminated water was the source of cholera.

By talking to local residents, Snow identified the source of the outbreak as the public water pump on Broad Street.

Snow used a dot map to illustrate how cases of cholera occurred around this pump.

https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak
Florence Nightingale, Causes of Mortality in the Crimean War, 1850s

- Nightingale is described as “a true pioneer in the graphical representation of statistics”, and is especially well-known for her usage of a polar area diagram.

- The chart illustrates seasonal sources of patient mortality in the military field hospital she managed.

- She used it to show that poor sanitary practices were the main culprit of high mortality in hospitals.

- Nightingale was elected the first female member of the Royal Statistical Society in 1874.

Prior to the Beck diagram, the various underground lines had been laid out geographically. This meant the centrally located stations were shown very close together and the out-of-town stations spaced far apart.

Beck believed that Underground passengers were not concerned with geographical accuracy and were more interested in how to get from one station to another and where to change trains.

Beck first submitted his idea in 1931 but it was considered too radical.

https://en.wikipedia.org/wiki/Harry_Beck
Charles Joseph Minard, The map of Napoleon's Russian campaign, 1869
Module Learning Outcomes

- Understand and demonstrate the core skills required to visualise information and processes.
- Explore procedures, data, and emergent systems as the subject of visualisations.
- Demonstrate an understanding of visualisations and their usage in a wide range of situations.
- Evaluate visualisations in terms of users and tasks and human factors.
- Assess visualisation needs and apply appropriate techniques to provide effective visualisation.
- Critically appraise visualisation approaches.
- Understand and be able to select and apply an appropriate range of visualisation and reporting techniques to data sets relevant to their workplace.
Our focus

• Data visualisation is not complicated (although it can be) but it is complex
• Lots of things to think about, many things to do and many things that need making
• Creative and journalistic sensibilities need to blend with analytical and scientific judgement
• Effective choices, consciously and efficiently made
The Visualisation Design Process

• There are good and bad visualisations but no perfect ones. People as recipients, introduce a diversity of need that cannot realistically be fulfilled

• Decision making is a key competency in data visualisation. Effective decisions, efficiently made.

• To accomplish this you need to follow a design process that organises your thinking and is underpinned by robust principles to optimise your thinking.
The 4 stages of the data visualization design process

The Seven Stages of Visualising Data

acquire — parse — filter — mine — represent — refine — interact

B Fry (2007), Visualizing Data, O’Reilly
Cairo’s creative methodology

• Define the focus – the story you want to tell and the key points you want to make. How will the visualization be useful to the viewer?
• Gather as much information as you can about the topic
• Choose the best graphic form(s)
• Complete your research – flesh out your ideas
• Think about the visual style – type faces, colour palettes etc
• Move the design to the computer

A Cairo (2012), Functional Art, The: An introduction to information graphics and visualization, New Riders
Fig. 2.1  The process of generating a graphical representation.

R Mazza (2009), Introduction to Information visualization, Springerlink
The Visualisation Design Process

• Process gives you a place to start and a way to know when you have finished
• Breaking down what can be an overwhelming prospect into a connected system of thinking can help progress your work and ensure cohesion between activities
• Incrementally leads you towards a solution with each stage building on the previous one and informing the next
• Should be used as a framework for thinking rather than a procedure to follow, offering adaptability and room for experimentation
Design Principles

• Where the process offers efficiency, design principles ensure effectiveness
• Most choices are relatively clear cut
• There are few universal rules
• But there are evidence-based useful suggestions about what you should and should not do
• We’ll look at what makes an effective visualisation in Lecture 4
<table>
<thead>
<tr>
<th></th>
<th>Morning Session</th>
<th>Lunch</th>
<th>Afternoon Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>9.00-10.00</td>
<td>10.15 - 12.15</td>
<td>12.15 - 13.00</td>
</tr>
<tr>
<td></td>
<td>Exam for a different module</td>
<td>Exam for a different module</td>
<td>Lab 1 Introduction to Visualisation</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td></td>
<td>Lab 1 Getting started</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Roundup</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Lecture 2</td>
<td>Lab 2</td>
<td>Lecture 3</td>
</tr>
<tr>
<td></td>
<td>Acquire, Parse, Filter and Mine</td>
<td>Data Acquisition and Pre-Processing</td>
<td>Visual Encoding and Different Types of Chart</td>
</tr>
<tr>
<td></td>
<td>Lecture 4</td>
<td>Lab 4</td>
<td>Lab 3</td>
</tr>
<tr>
<td></td>
<td>What makes an effective visualisation?</td>
<td>Text Analysis</td>
<td>Different Types of Static Plots</td>
</tr>
<tr>
<td></td>
<td>Lab 6</td>
<td></td>
<td>Lecture 5</td>
</tr>
<tr>
<td></td>
<td>Presentation and Influencing Factors</td>
<td>Interactivity</td>
<td>Evaluating Visualisations (Workshop)</td>
</tr>
<tr>
<td></td>
<td>Lab 7</td>
<td></td>
<td>Lab 5</td>
</tr>
<tr>
<td></td>
<td>Lecture 8</td>
<td>Lab 8</td>
<td>Lab 7</td>
</tr>
<tr>
<td></td>
<td>Visualising Processes and Systems</td>
<td>Particles and Particle Systems</td>
<td>Visualising Algorithms</td>
</tr>
<tr>
<td></td>
<td>Lecture 9</td>
<td>Lab 9</td>
<td>Lab 9</td>
</tr>
<tr>
<td></td>
<td>Bringing It All Together and Formulating a Brief</td>
<td>Cellular Automata</td>
<td></td>
</tr>
</tbody>
</table>
Labs

- All the labs will follow the same format. You will spend two hours working through a Jupyter notebook or lab sheet individually.
- You will be supported by Richard
- More information at https://warwick.ac.uk/fac/sci/dcs/people/claire_rocks/cs2d7/labs/
Coursework (100%)

- Visualise a data set, system, or algorithm of your choice.
- The module is 100% Coursework assessed, split into the following components:
  - Initial brief (20%)
  - Visualisation (50%)
  - Documentation (30%)
Initial Brief (20%)

• **Outline your brief in a 10-minute presentation, with a further 10 minutes for feedback.**

• For this part of the assessment, you will formulate and outline a brief for your final visualisation. This will be reported in a presentation which you should have **ready by 22 July 2022**, and we ask you to **submit your slides (via Tabula) by 12.00 on that day**.

• The module organisers will contact you to arrange a suitable time for you to give your presentation.

• Your presentation should include:
  - The motivating curiosity
  - The circumstances that will affect the project
  - A description of the dataset
  - A report on your initial understanding of the subject through the data
  - A description of your vision for the final visualisation
Initial Brief (20%)

Content (70% of this component)

- Knowledge/understanding
  - Demonstrates understanding of taught material around formulating a brief
  - The vision is well developed and supported through documentation of the process
- Individual research
  - Suitable background research conducted
  - Demonstrates an understanding of the subject through the data
- Use of evidence
  - Decisions are evidence-based

Communication Skills (30% of this component)

- Logical organisation/structure
- Well-prepared slides and presentation
- Audience awareness and engaging delivery
- Adherence to required time constraints
- Response to questions

- This is an opportunity for feedback on your ideas ahead of the main components of the assessment
Visualisation (50%)

• For this part of the assessment, you will visualise a data set, system, or algorithm of your choice. This should be submitted as a Jupyter notebook (.ipynb) or a Processing Sketchbook by 12.00 on 12 August 2022 (via Tabula).

• Note: if you are providing a Processing Sketchbook, please also provide a readme file with any relevant instructions.
Visualisation (50%)

**Final concept and Design (30% of this component)**
- The visualisation is intentional and well developed
- The design demonstrates an understanding of the principles underpinning good visualisations

**Technical proficiency (70% of this component)**
- Source code compiles
- Code is consistent, clean, formatted and well documented.
- Clearly shows where data/code has been borrowed from other sources
- Demonstrates a level of technical skill
For this part of the assessment, you are asked to provide a technical report describing your work in 2000 words. This should be submitted as a pdf file by 12.00 on 12 August 2022 (via Tabula).

Your report should include
- Introduction
- Design
- Implementation
- Resulting visualisation
- Conclusion
Documentation (30%)

Content (70% of this component)

- Knowledge/understanding
  - Demonstrates understanding of taught material
  - Relevant application of taught material
- Individual research
  - Suitable background research conducted
  - Demonstrates understanding of material researched
- Depth of reflection/analysis
  - Empathises with and analyses different perspectives
  - Demonstrates critical self-awareness
  - Critical analysis of sources and perspectives
  - Critical evaluation of final visualisation
  - Evaluates sources and constructs well-reasoned balanced arguments
- Use of evidence
  - Discussion is evidence-based

Communication Skills (30% of this component)

- All components present
- Logical organisation/structure
- Language and technical writing skills
- Readability
- Appropriate length
- References are used appropriately
General tips and tactics

• Document thoughts and capture sketches. Do not rely on your memory
  • Document information about the sources of data, details of any calculations or manipulations applied to the data, any assumptions you have made, technical properties of your data, issues or problems you have experienced or can foresee, wish lists of features or ideas you would like to explore, sources of inspiration, ideas you have or have rejected
  • This is about communicating information - it’s a two-way activity. Who are your audience? What do they want? What do they expect? What ideas do they have?
Lab

• Getting Started

Tomorrow morning

• Acquire, Parse, Filter, Mine