



MONASH
University

MONASH
INFORMATION
TECHNOLOGY

TURN ON RECORDING!



FIT5147

Data Exploration & Visualisation

Unit Overview

Exploring and understanding data using visualisation and analytics

- Introduction to visual analytics
- Different kinds of data visualisations and their uses
- Understanding what makes a graphic or visualisation effective
- Computer tools for creating visualisations (mainly **R & D3**)
- Hands on experience in developing your own graphics and visualisations

People

- Instructors:
 - Michael Niemann
 - Shirin Ghaffarian Maghool

We can be contacted through the private discussion forum.

- Chief Examiner: Prof Kim Marriott

Prerequisites

- Some of the material relies on a **basic knowledge of statistics** (mean, standard deviation, median) and a basic knowledge of geometry.
A secondary/high-school level understanding of these concepts is sufficient.
- Some **knowledge of programming with R** is required.

Unit Outcomes

On successful completion of this unit, you should be able to:

- Perform exploratory data analysis using a range of visualisation tools;
- Describe the role of data visualisation in data science and its limitations;
- Critically evaluate and interpret a data visualisation;
- Distinguish standard visualisations for qualitative, quantitative, temporal and spatial data;
- Choose an appropriate data visualisation;
- Implement interactive data visualisations using R and other tools.

Week	Material	Activities	Assignment
1	Visual Analytics; History of Data Vis; Tools for Data Exploration & Visualisation	Intro to common data exploration tools; Five design sheet visualisation design methodology	
2	Visualisation of tabular data; analysis of trends & patterns in tabular data	Advanced and interactive graphics with R;	Quiz 1
3	Exploring spatial data with data maps	Tools for creating data maps;	Project: Project Proposal
4	Network data analysis & visualization; Textual data analysis & visualisation	Tools for network and text analysis; Project progress interview	DataVis Design Exercise
5	Human visual system; Visual communication	Design of effective visualisations; Introduction to D3; Project progress interview;	Quiz 2
6	Interactive data visualisation; the future of data visualisation	Design of visual analytics tools; Selected design groups presentations	Project: Final product and report

Teaching style

- The e-textbook is the most important reference
 - text & activities
 - html and pdf formats
- Weekly meet-up session
 - Contains design critiques
 - Recordings, chat logs and slides will be posted on Moodle the next day
- We'll be slightly different from other GDDS units
 - expect more group activities and discussions in meet-ups.
- Discussion forums – general and private
- Consultations
 - see the Weekly Meet-ups section on Moodle to book appointments

Assessment

Assessment Task	Weight	Due Date
Quiz 1	20%	Monday Week 3
Assignment 1 Project Proposal	10%	Monday Week 4
Assignment 2 DataVis Design Exercise	10%	Monday Week 5
Quiz 2	15%	Monday Week 6
Assignment 3 Data Exploration and Visualisation Project Final	45%	Wednesday Week 7

- Easter holidays (Good Friday, Easter Monday, Easter Tuesday) come at the end of Week 6

Data Exploration & Visualisation Project

Runs the whole teaching period and is Assignments 1 and 3. It gives you the chance to investigate something you are interested in.

- **Stage 1:** Identify project (Purpose, data source, research questions);
- **Stage 2:** Collect data & wrangle it into a suitable form using whatever tools you like;
- **Stage 3:** Explore the data to find something interesting using **Tableau or R** and decide what you wish to communicate; **submit your project proposal and presentation via a YouTube video [Assignment 1]**
- **Stage 4:** Design an interactive narrative visualisation to communicate this using the **Five Sheet Design (FSD)** methodology;
- **Stage 5:** Implement your visualisation in **R or D3**;
- **Stage 6:** Prepare a video demo presentation (3-5 minutes) about your Project. **Submit your final report and code for your product, video presentation link [Assignment 3]**

You are required to attend an interview session, discuss your progress, findings and difficulties you encounter (Weeks 4-5).

An Example Project Idea

Title: Investigating causes of serious road accidents

Questions:

- Number of road accident fatalities has decreased in Australia over the last 20 years. Why?
- Recently the number of serious but non-fatal road accidents has increased. Why?
- What is the relationship between these two things? What government policy is working and what is not working?

Data:

<http://www.tac.vic.gov.au/road-safety/statistics/online-crash-database>

DataVis Design Exercise

Assignment 2 asks you to design ways to visualise some data

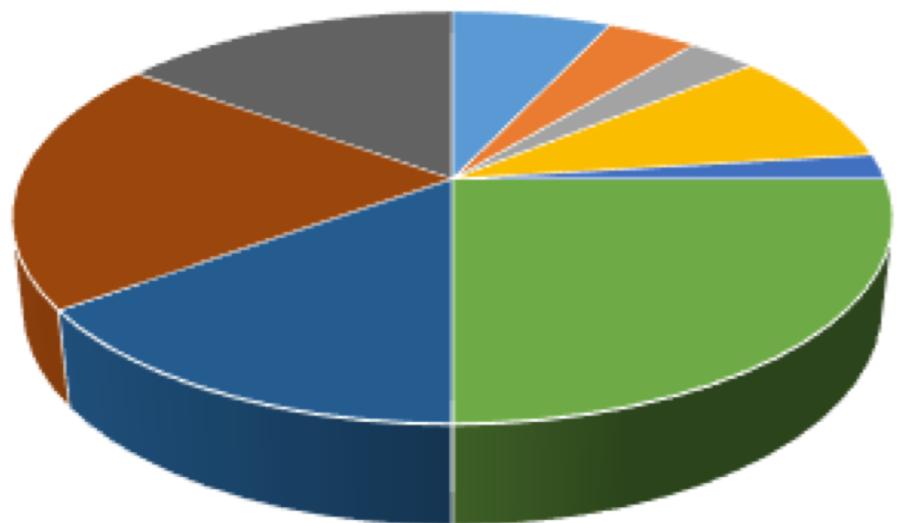
- The data relates to changes in the Australian population and migrant movements over time.
- You are using the **Five Design Sheet methodology (FdS)** to design and mock-up the visualisation
- There is no data analysis or coding required to be done by you.
- **We encourage you to do this in groups of 2 or 3 students**
- Due Monday Week 5

Quizzes

- 2 quizzes worth
 - **Quiz 1** (20% of unit mark)
 - Monday Week 3
 - Modules 1 & 2
 - 1 hour long
 - **Quiz 2** (15% of unit mark)
 - Monday Week 6 (before Easter)
 - Modules 1-5 (mainly 3-5)
 - 1.5 hours long
- Closed book
- Recorded using Respondus and webcam

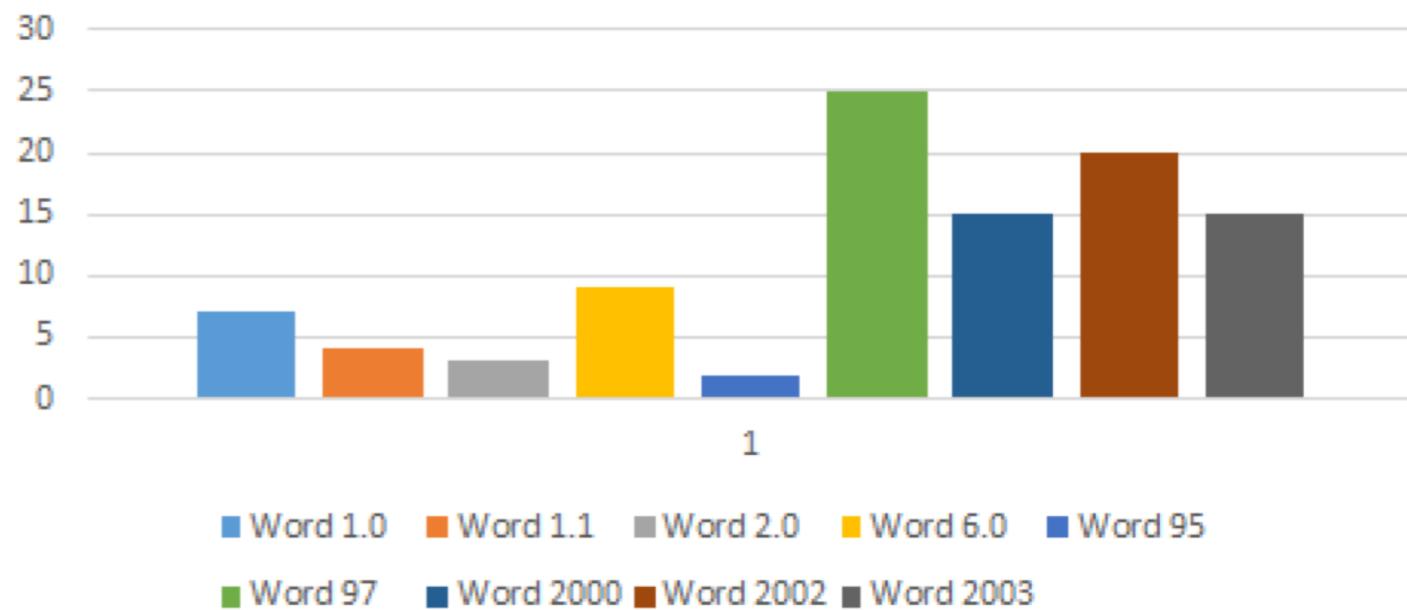
Which do you prefer?

Microsoft Word features by Version Added



- Word 1.0 ■ Word 1.1 ■ Word 2.0 ■ Word 6.0 ■ Word 95
- Word 97 ■ Word 2000 ■ Word 2002 ■ Word 2003

Microsoft Word features by Version
Added



Data Visualisation

- Not about creating cool InfoGraphics
- About helping you understand the data



FIT5147 Data Visualisation

Hall of Fame / Hall of Shame

Look online or in books to find data visualisations or information graphics that you **really like** or **really dislike**. Send us a link/image and a short reason why and I'll choose 1 or 2 to discuss at the start of each meet-up.



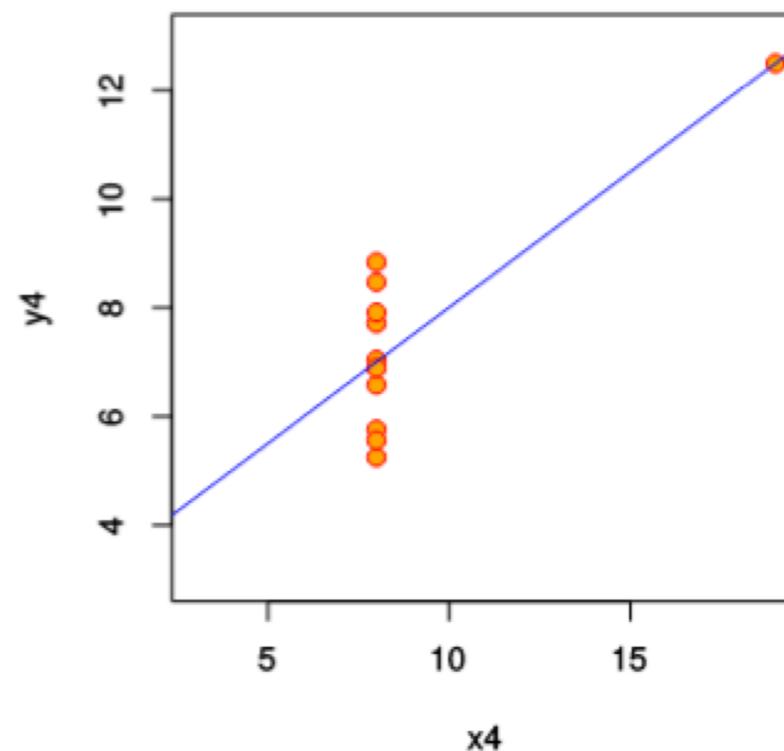
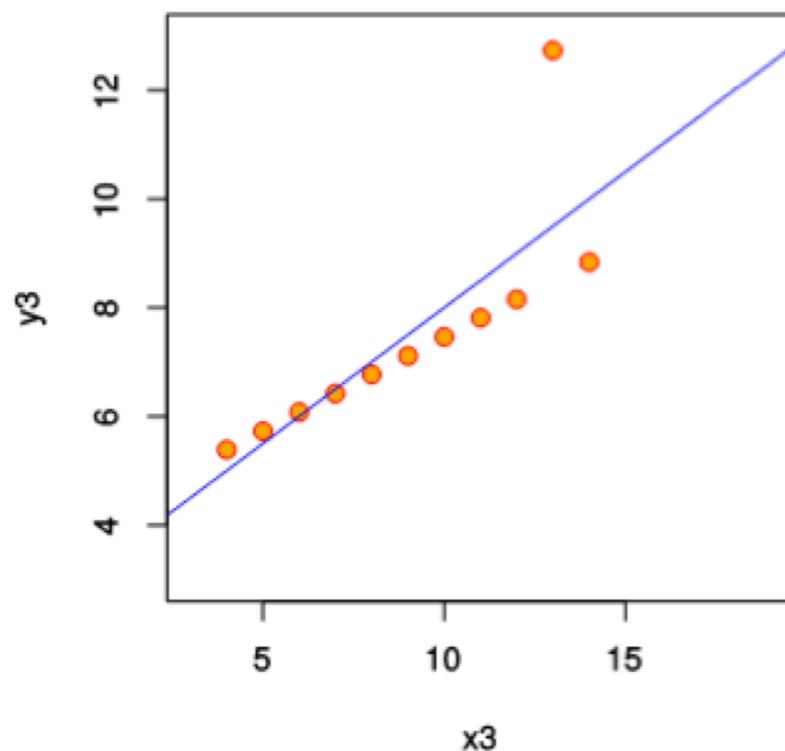
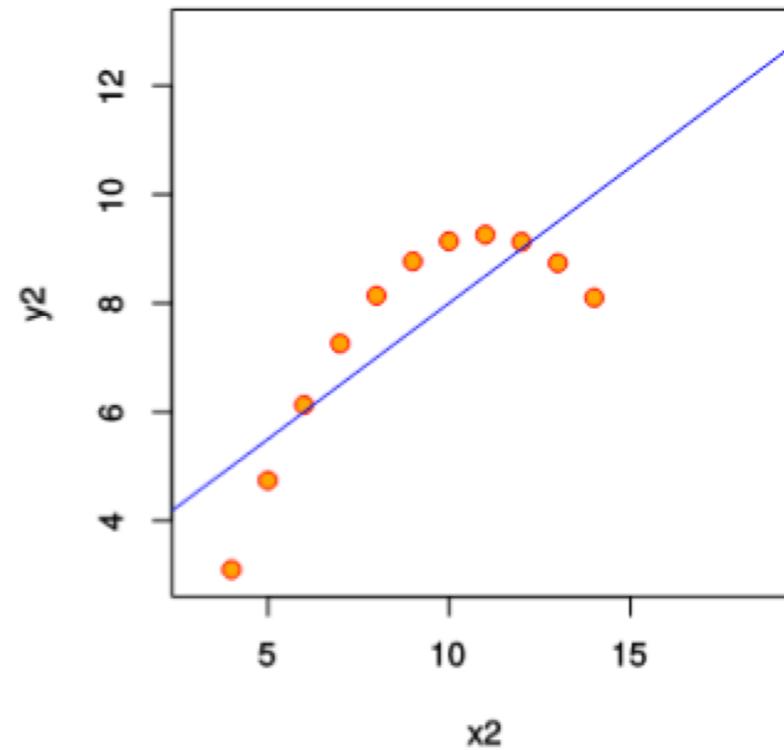
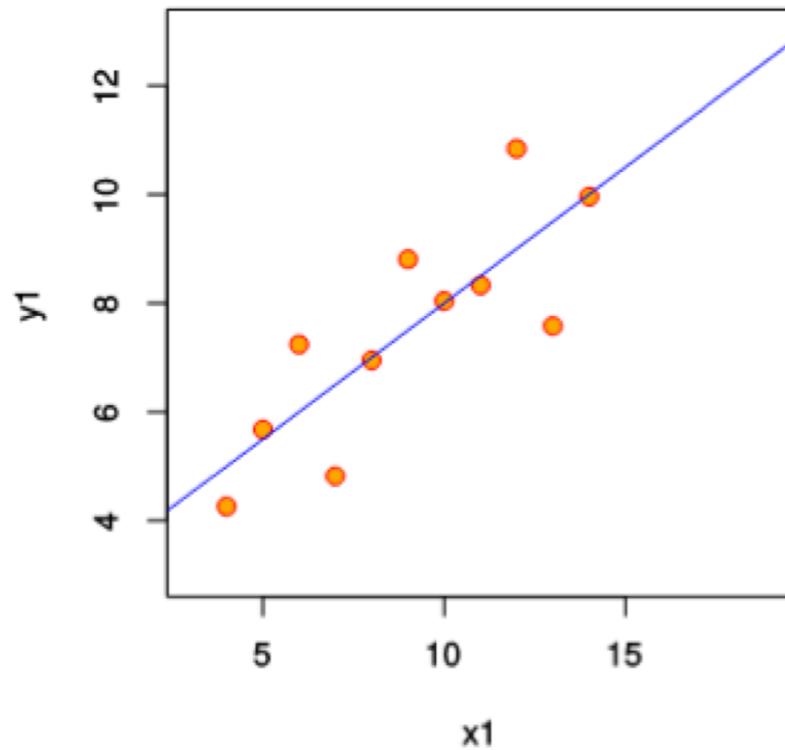
FIT5147
Data Exploration & Visualisation
MODULE 1

Discussion

Try and answer the following questions

1. What are the different uses of data visualisation in data science? Give an example for each use.
2. Why use visualisation?
3. What are the advantages and disadvantages of visualisation compared to statistical measures?
4. When should you use statistical measures and when should you use visualisation?
5. Can you combine them? How?

Anscombe's 4 Regression data sets



Role of Data Visualisation

Data visualisation is used in Data Science for

- Data checking and cleaning
- Exploration and discovery
- Presentation and communication of results

Interactive computer visualisations are one of the best ways we have to deal with Big Data.

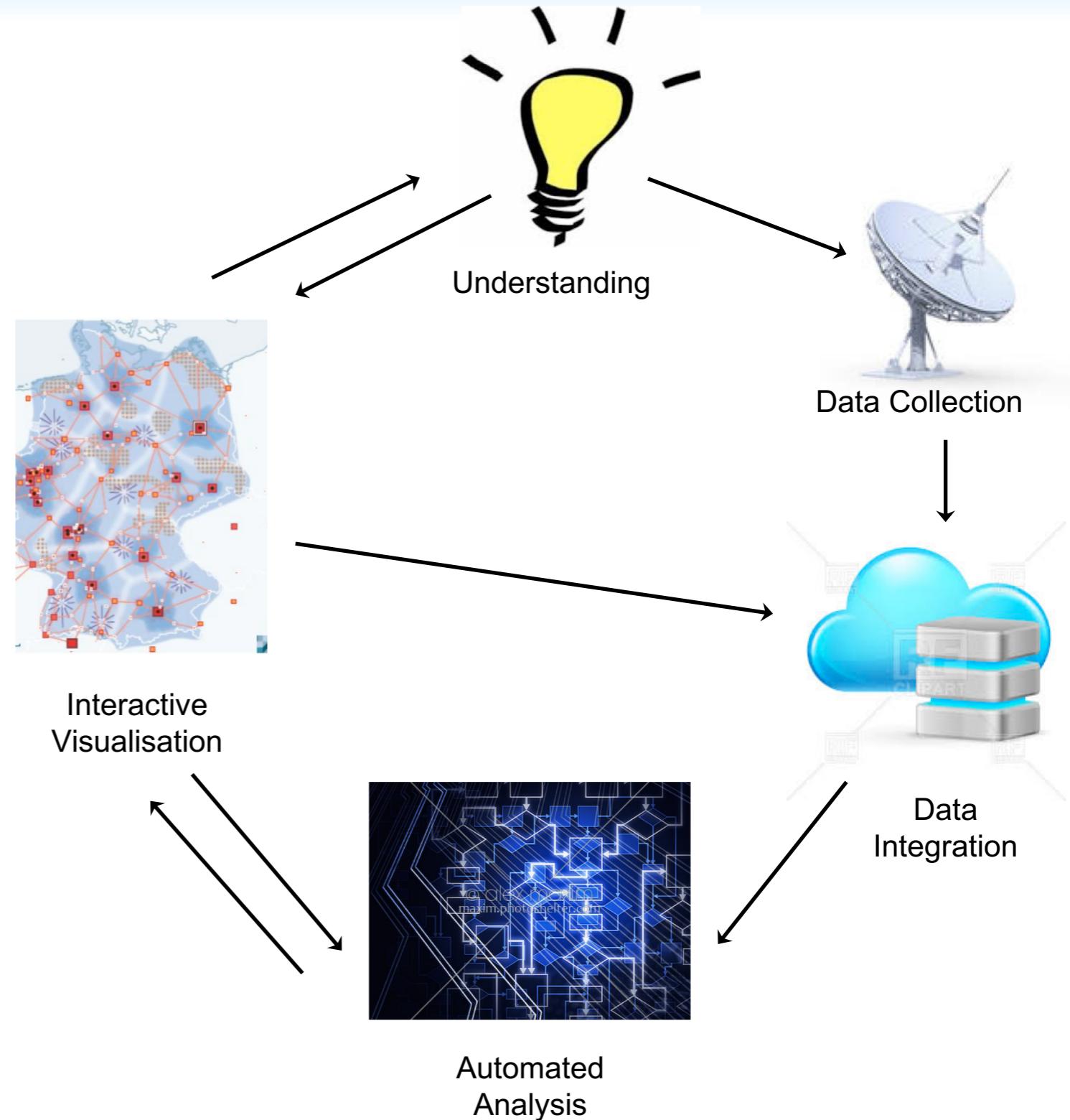
Human-in-the-Loop Analytics

You need the Human-in-the-Loop Analytics to

- Look at the data,
- Make some tentative hypothesis,
- Run appropriate analytics and visualise the results.
- Repeat this until you have found what you need.

Visual Analytics

- Visual analytics is a recent focus in data visualisation started by James Thomas and Kristin Cook in the early 2000s.
- It is "the science of analytical reasoning facilitated by interactive visual interfaces."
- The aim is to "detect the expected and discover the unexpected."

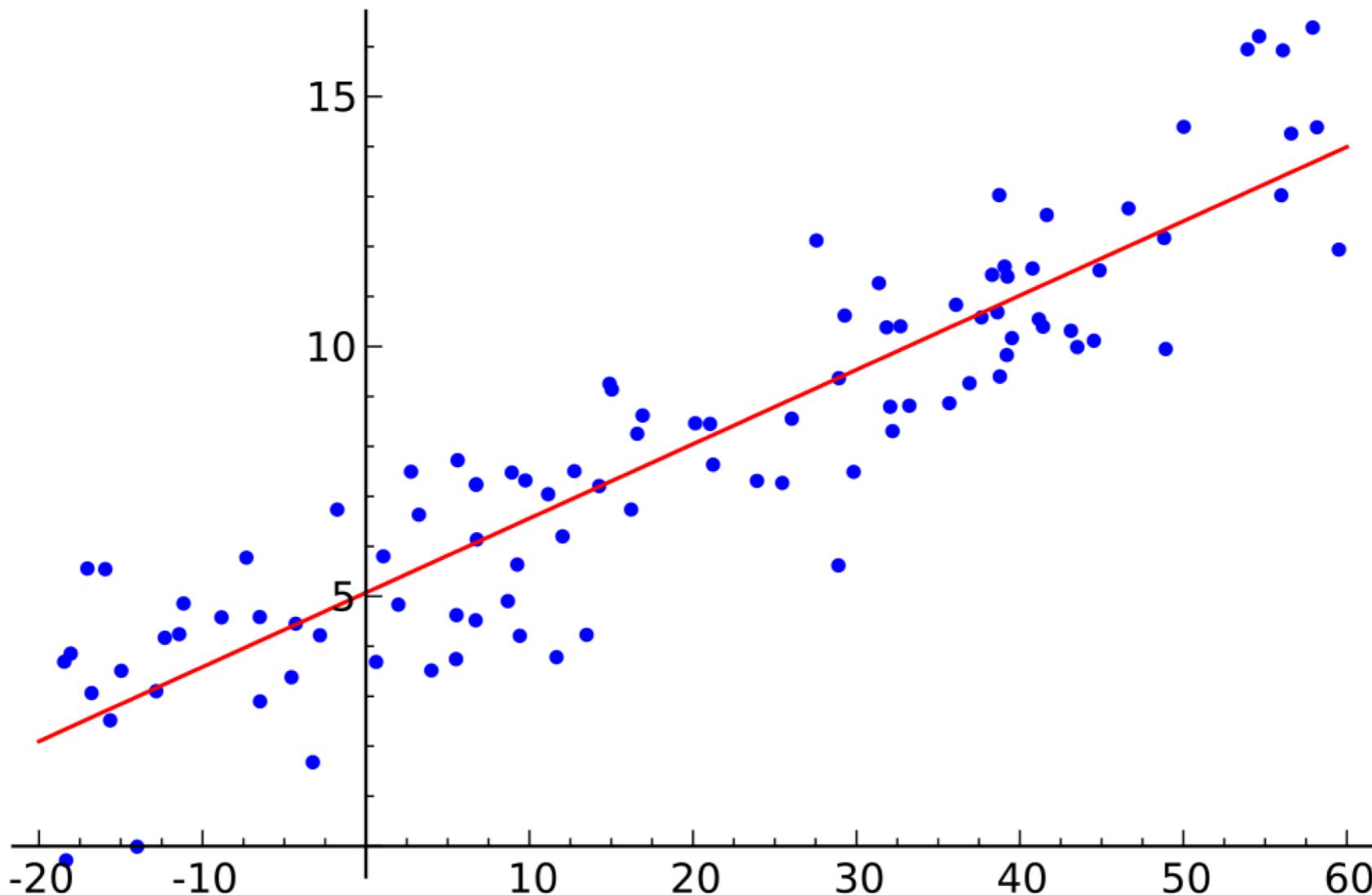


Exploration and Discovery

The heart of data science is exploring the data and discovering patterns and trends. Typical tasks are:

- 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

Curve Fitting

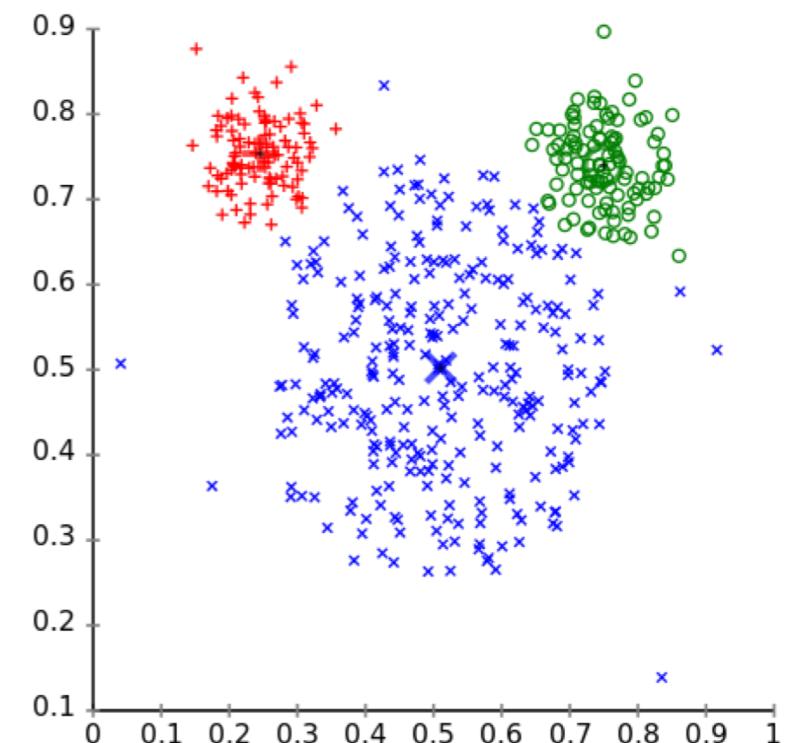
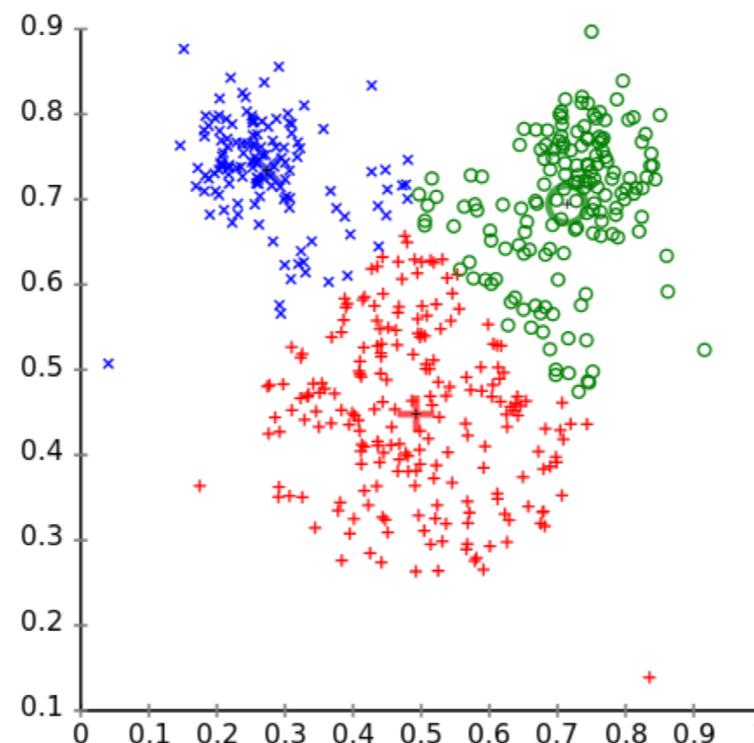
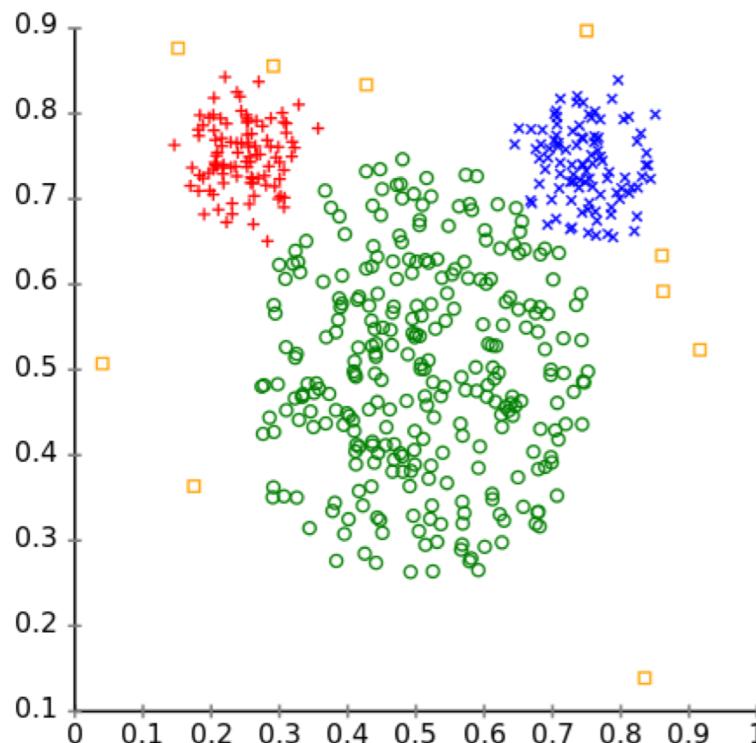


Curve fitting is an important part of data exploration.

Clustering

Different cluster analysis results on "mouse" data set:

Original Data k-Means Clustering EM Clustering



Clustering is another common exploration technique

Communication of Results

Data scientists often need to communicate their results to **stakeholders who are not data scientists**: managers, policy makers, students or the general public.

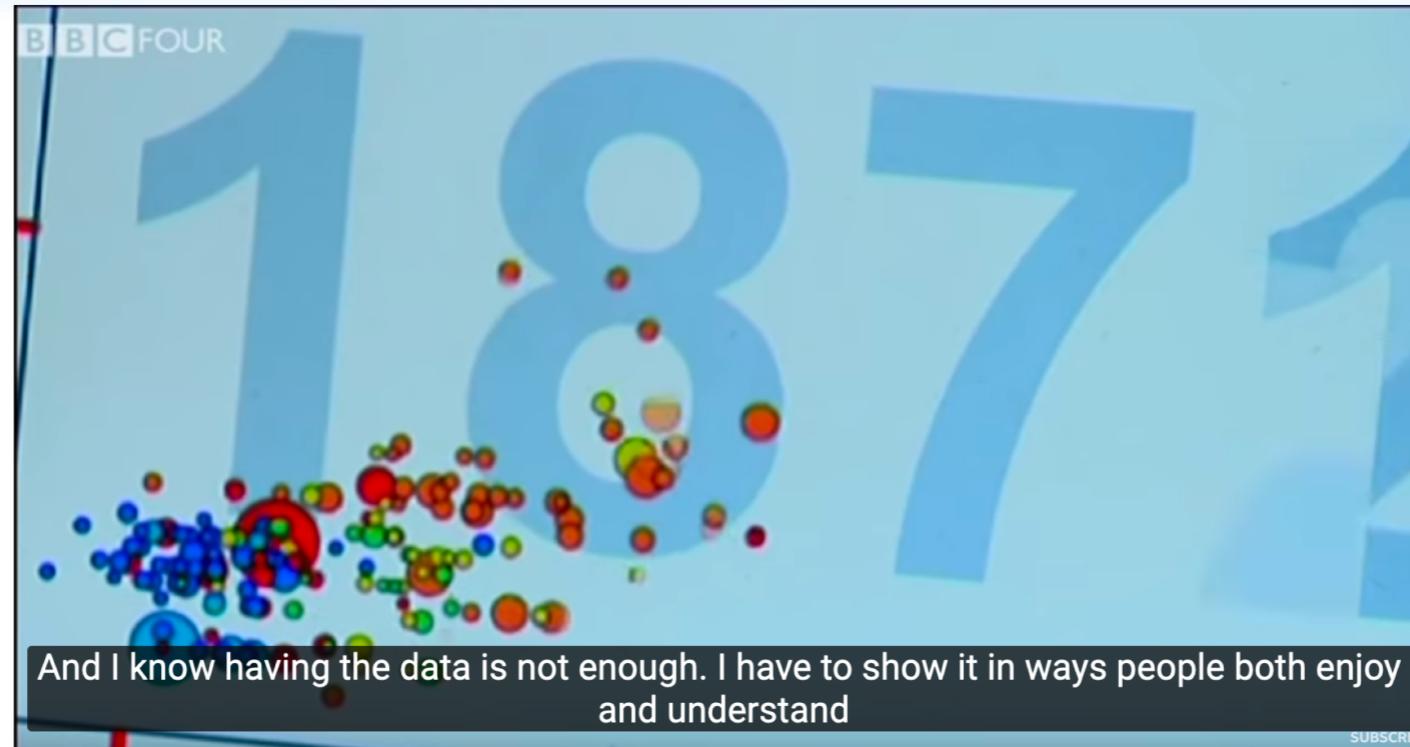
Such visualisations require considerable time to prepare

- Designed to communicate a particular message or narrative.
- Production values are very high: often touched up using graphics editing tools like Adobe Illustrator or Inkscape.

Once presentation graphics were static, printed on glossy paper or shown in PowerPoint presentations.

Now often interactive and published on the Web.

Communication of Results



Hans Rosling's 200 Countries, 200 Years, 4 Minutes - The Joy of Stats (YouTube – BBC Four)

Take a look at some Nathan Yau's books or his web site [Flowing Data](#)

History of Data Visualisation

Reasons for changes

- Improved technology for producing and presenting graphics.
- Changes in societal needs and attitudes to graphics.
- Availability of data.
- Invention of graphical notations and interaction techniques.

Visual Analytics Tools

- A wide variety of data exploration and visualisation tools are used in data science
- Programming languages
 - R
 - Python
 - JavaScript with the D3 visualisation library
- Generic Visual Analytics Tools
 - Tableau
 - Google Fusion Tables
 - Spotfire
- Application Specific Visual Analytics Tools
 - National Map
 - Scaffold Hunter

Main Kinds of Data Sets

- **Tabular data:** Data organised in tables, a row for each data item and a column for each of its attributes.
- **Spatial data:** Data which is naturally organised and understood in terms of its spatial location or extent.
- **Network data:** Nodes in the network are data items and links between the nodes are relations between. For instance, a social network.
- **Sequential data:** Data organised in a sequence. Text is an example of this.

Kinds of Attributes

Categorical: no natural ordering on the values.

Ordered: has a natural ranking or ordering

- **Ordinal:** only has a ranking
- **Quantitative:** has a magnitude

Can be **sequential**, **diverging** or **cyclic**

Tasks for this Week

- Go through the e-textbook content and activities
- In your private Moodle forum, tell us if you have a strong preference for a partner in Assignment 2 | DataVis Design Exercise – by noon of Friday Week 2. Otherwise, we will assume you want to work on your own.
- Start to think about the project for Assignments 1 & 3, and discuss your idea with us (either through the private forum or online consultation).

Group activity: Introductions

- Your name
- Why you are doing this unit & relevant background?
- What you want to learn/get out of the unit?

This week's Design Critique

- You have been approached by Monash to design an app that helps GDDS students plan their selection of units.
- **What:** You have access to all of the information in the handbook for units in GDDS including requirements for course completion, unit offerings sequence, unit dependencies & prerequisites as well as previous SETU evaluations of the units.
- **Why:** The app should help students select units and understand the dependencies between units and the perfect sequence of doing the units.
- **How:** Use the FdS methodology to work out How.
- You will be separated into groups on Zoom, and please work on the [corresponding group worksheet](#). You can use paper and pen as well! Just paste photos on the Google slide.

Five Design Sheet method

<https://5147.yalongyang.com/book/module01/activity-five-design-sheet-methodology.html>

Sheet 1: The Ideas Sheet

- Brainstorming
- Develop 3 ideas

Sheets 2, 3 & 4: Alternative Designs

- 5 panels: Information, Layout, Operations, Focus/Parti, Discussion

Sheet 5: Realisation

- Information so the final design can be implemented
- Description, Maths, Software, Estimates, Other requirements