Data Analysis and Visualisation
Data Analysis

“A process to obtain raw data and convert it into information that is useful for a decision maker”
Data Analysis

“A process to obtain raw data and convert it into information that is useful for a decision maker”

1. Data requirements
2. Data collection
3. Data processing
4. Data cleaning
5. Exploratory data analysis
6. Modelling and algorithms
7. Data Product
8. Communication
Data Analysis

“A process to obtain raw data and convert it into information that is useful for a decision maker”

1. Data requirements

<table>
<thead>
<tr>
<th>What is the type of entity the data will be collected on?</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g. an individual, a population, an area</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What variables will be considered?</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g. age, sex, lifestyle, measurements, opinions etc.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Will the data for analysis be numerical or categorical?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
Data Analysis

“A process to obtain raw data and convert it into information that is useful for a decision maker”

1. Data requirements
2. Data collection

Data can be collected through various methods:

- Interviews
- Recordings
- Environmental Sensors
- Instruments
- Downloaded
- Reading books
Data Analysis

“A process to obtain raw data and convert it into information that is useful for a decision maker”

1. Data requirements
2. Data collection
3. Data processing

Organising the data for analysis.

Classifying  Structuring
Sorting  Aggregating
Data Analysis

“A process to obtain raw data and convert it into information that is useful for a decision maker”

1. Data requirements
2. Data collection
3. Data processing
4. Data cleaning

Removing duplicates
Identifying gaps
Correcting typos
Data Analysis

“A process to obtain raw data and convert it into information that is useful for a decision maker”

1. Data requirements
2. Data collection
3. Data processing
4. Data cleaning
5. Exploratory data analysis

Initial Checks
Data Analysis

“A process to obtain raw data and convert it into information that is useful for a decision maker”

1. Data requirements
2. Data collection
3. Data processing
4. Data cleaning
5. Exploratory data analysis
6. Modelling and algorithms

Application of Mathematical Formulae or Models to the Data

- R-project
- SPSS
- Mathematica
- Matlab
- SolidWorks
- ANSYS
- Paraview
- BLAST
- QIIME
- Insight
- Maya

the list goes on... and on...
Data Analysis

“A process to obtain raw data and convert it into information that is useful for a decision maker”

1. Data requirements
2. Data collection
3. Data processing
4. Data cleaning
5. Exploratory data analysis
6. Modelling and algorithms
7. Communication

Data Visualisation
Tools

Microsoft Excel
Tools

Microsoft Excel

**Pros:**

Ubiquitously available

Spreadsheet (1,048,576 rows)

Many useful functions (approx. 400)

Syntax for using functions is fairly intuitive

Easy to create publication quality graphs
Tools

Microsoft Excel

Cons:

Time format is confusing

Excel Time ranges between 0 and 1
  0 = 12 midnight
  0.455555555555556 = 10:56 am
  0.5 = 12 noon
Tools

Microsoft Excel

**Cons:**

The Date formatting is diabolical:

Excel Dates range from 1 onwards:

1 = 01 Jan 1900 (PC) **OR** 1 = 01 Jan 1904 (Mac)

Transferring Excel files including dates from Mac to PC or to other packages can result in serious problems
Microsoft Excel

Cons:
The Date formatting is diabolical:
Excel Dates range from 1 onwards:
1 = 01 Jan 1900 (PC)
1 = 01 Jan 1904 (Mac)

Transferring Excel files including dates from Mac to PC or to other packages can result in serious problems.
# Tools

## Microsoft Excel

<table>
<thead>
<tr>
<th>Year</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*May 23, 2019*

---

Kathy Shein and A/Prof Julia Connell
Tools

Microsoft Excel

Kathy Shein and A/Prof Julia Connell
**Data Analysis**

**Microsoft Excel**

\[ f(x) = \text{AVERAGE}(\text{PersonA!E3}, \text{PersonB!E3}) \]

Kathy Shein and A/Prof Julia Connell
Data Analysis

Open Refine
Data Analysis

Open Refine

Free, open source software
**Data Analysis**

Open Refine

Identifies inconsistencies in data

<table>
<thead>
<tr>
<th>Address</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>45 Wooli Street</td>
<td>1</td>
</tr>
<tr>
<td>47 Wooli Street</td>
<td>1</td>
</tr>
<tr>
<td>57 Wingecarribee Street</td>
<td>1</td>
</tr>
<tr>
<td>55 Wingecarribee Street</td>
<td>1</td>
</tr>
<tr>
<td>60 Campbell Street</td>
<td>1</td>
</tr>
<tr>
<td>62 Campbell Street</td>
<td>1</td>
</tr>
<tr>
<td>129 Imlay Street</td>
<td>1</td>
</tr>
<tr>
<td>229 Imlay Street</td>
<td>1</td>
</tr>
<tr>
<td>93 Byron Street</td>
<td>1</td>
</tr>
<tr>
<td>95 Byron Street</td>
<td>1</td>
</tr>
</tbody>
</table>
Data Analysis

Open Refine

Supplement data through an “API”

...
Add column based on column Suburb

New column name: Google Geocoding information

On error: set to blank

Expression:
"http://maps.googleapis.com/maps/api/geocode/json?
sensor=false&address=\"+escape(value + ", New South Wales\", "url\")"

Preview:

<table>
<thead>
<tr>
<th>row</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Adaminaby</td>
</tr>
<tr>
<td>2.</td>
<td>Adamstown</td>
</tr>
<tr>
<td>3.</td>
<td>Adamstown</td>
</tr>
<tr>
<td>4.</td>
<td>Adelong</td>
</tr>
</tbody>
</table>

OK  Cancel
Data Analysis

Open Refine

Supplement data through an “API”

Google Geocoding API
Data Analysis

Open Refine

Supplement data through an “API”

“Online Mendelian Inheritance in Man”
API
Data Analysis

Google Fusion
Data Analysis

Google Fusion

A Google App

Connect to it your Google Drive

Data is uploaded to the cloud (if you have sensitive data, remove sensitive parts first!)
### Data Analysis

**Google Fusion**

**Merge 2 Tables**

<table>
<thead>
<tr>
<th>City</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYD</td>
<td>NSW</td>
</tr>
<tr>
<td>MEL</td>
<td>VIC</td>
</tr>
<tr>
<td>BNE</td>
<td>QLD</td>
</tr>
<tr>
<td>ADL</td>
<td>SA</td>
</tr>
<tr>
<td>CBR</td>
<td>ACT</td>
</tr>
<tr>
<td>PER</td>
<td>WA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>City</th>
<th>Pop</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADL</td>
<td>1.2m</td>
</tr>
<tr>
<td>MEL</td>
<td>4.1m</td>
</tr>
<tr>
<td>CBR</td>
<td>0.3m</td>
</tr>
<tr>
<td>BNE</td>
<td>2.0m</td>
</tr>
<tr>
<td>SYD</td>
<td>4.6m</td>
</tr>
<tr>
<td>PER</td>
<td>1.6m</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>City</th>
<th>Pop</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYD</td>
<td>4.6m</td>
<td>NSW</td>
</tr>
<tr>
<td>MEL</td>
<td>4.1m</td>
<td>VIC</td>
</tr>
<tr>
<td>BNE</td>
<td>2.0m</td>
<td>QLD</td>
</tr>
<tr>
<td>ADL</td>
<td>1.2m</td>
<td>SA</td>
</tr>
<tr>
<td>CBR</td>
<td>0.3m</td>
<td>ACT</td>
</tr>
<tr>
<td>PER</td>
<td>1.6m</td>
<td>WA</td>
</tr>
</tbody>
</table>
Data Analysis

Google Fusion

Create an interactive map and share it

The Guardian
Data Analysis

• At UTS we have free courses on how to use Excel, Google Fusion, Open Refine for research: [http://bit.ly/16iD1rV](http://bit.ly/16iD1rV)
Data Analysis

• At UTS we have free courses on how to use Excel, Google Fusion, Open Refine for research: http://bit.ly/16iD1rV

• Courses are advertised through UTS Staff Notices and the Graduate Research School email list
Data Analysis

• At UTS we have free courses on how to use Excel, Google Fusion, Open Refine for research: [http://bit.ly/16iD1rV](http://bit.ly/16iD1rV)

• Courses are advertised through UTS Staff Notices and the Graduate Research School email list

• The training material is free and available online
Data Visualisation
Data Visualisation

- UTS/Intersect presentation on Data Visualisation: http://bit.ly/18GhTg3

- UTS Library Blog on Data Visualisation: http://bit.ly/1CQung8
Want help or advice?

jeff@intersect.org.au

http://www.lib.uts.edu.au/answers

eresearch-it@uts.edu.au