

Overview and Introduction

SMI105: Week 1

Dr. Calum Webb

Sheffield Methods Institute, the University of Sheffield
c.j.webb@sheffield.ac.uk

Sign in



Learning outcomes

What will I learn?

By the end of this week you will:

- Lecture: How the module will operate, where to find more information, how to navigate the module resources, and what we will be doing each week.
- Lecture: What is data visualisation, why do we do it, why does it matter, and how will we approach it?
- Workshop: First steps in **ggplot2** and **R**: how do we create visualisations of data in **R**

Who are we?

Calum Webb

c.j.webb@sheffield.ac.uk

Research on poverty and the child welfare system in the UK using secondary survey and administrative data in R. My most recent project has been about how investment in services for children improves their welfare and safety.

Lowenna Jones

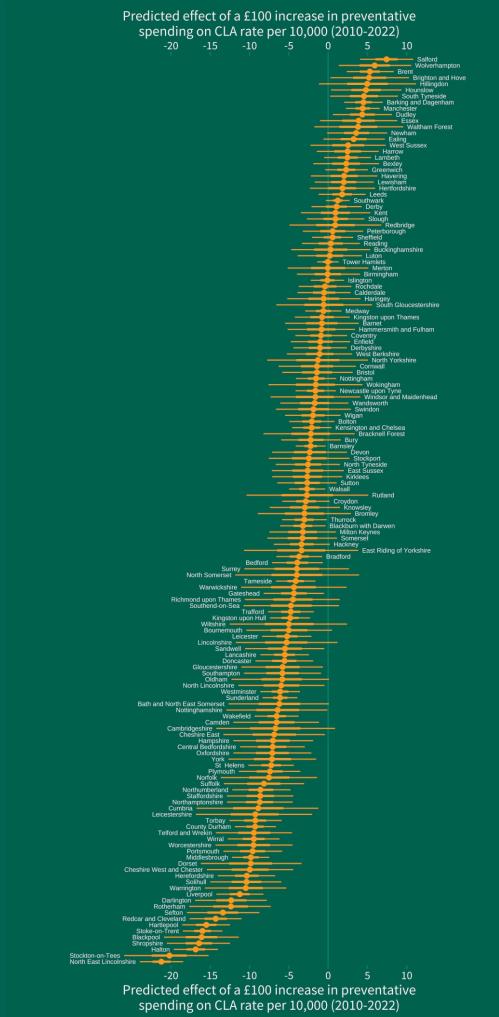
Lowenna's research focuses on regulation of chemical risk, particularly within the United Kingdom (UK) following the UK's departure from the European Union (EU) (i.e., Brexit).

Kate Montague-Hellen

Kate's PhD research project aims to explore the impact of an untreated chronic cancer diagnosis on patients, their expectations, and support needs in the first year following diagnosis.

Here's some data visualisation I've done in my own work
and hobbies





CWIP APP



Welcome to the CWIP App

Data Dashboard

Spending Waffle Plots

Side-by-side Trends

Side-by-side Maps

CLA Rate by Ethnic Group

LA Similarity Clusters

Neighbourhood Maps

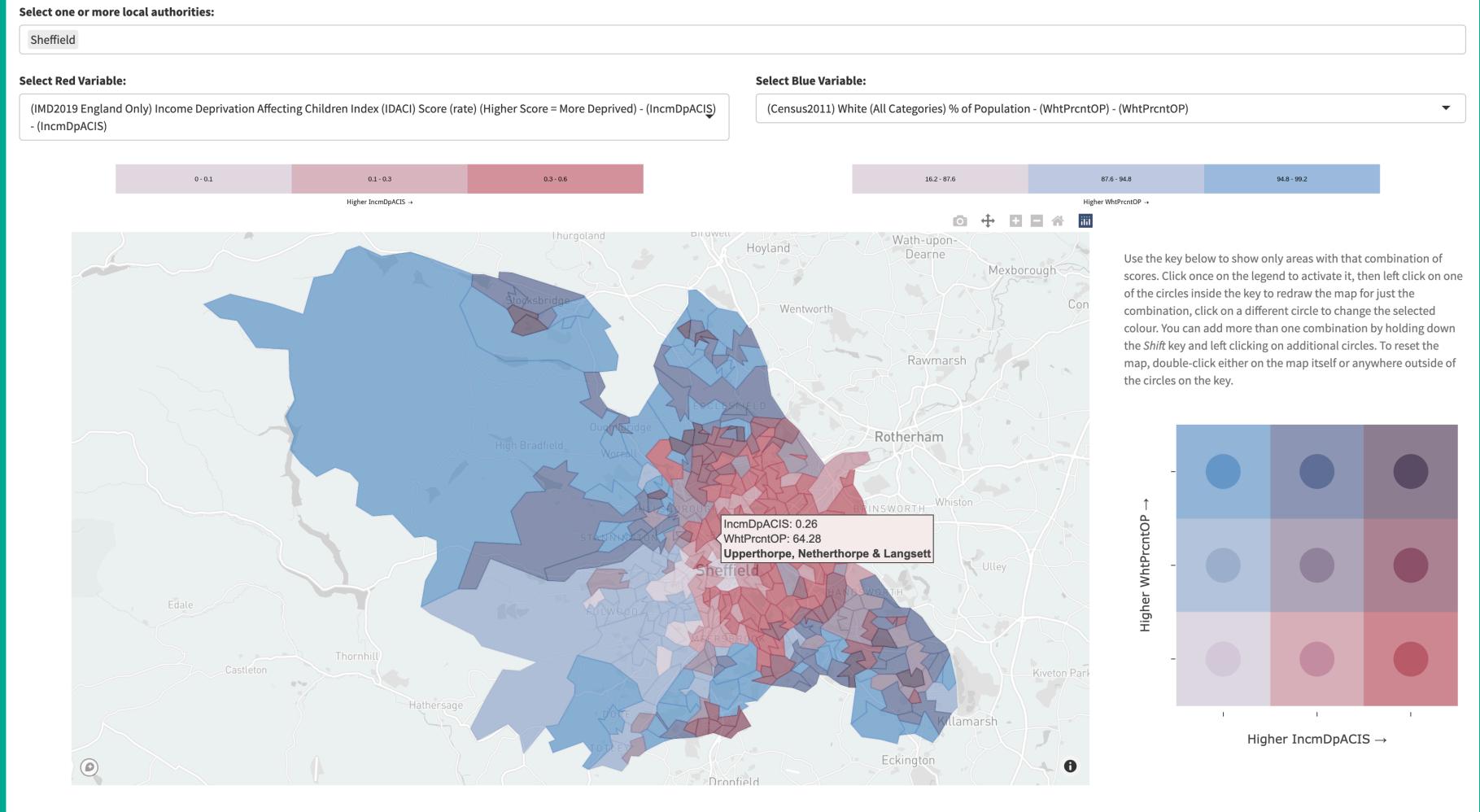
Neighbourhood Overlap Maps

Neighbourhood Correlations

LA ANOVA

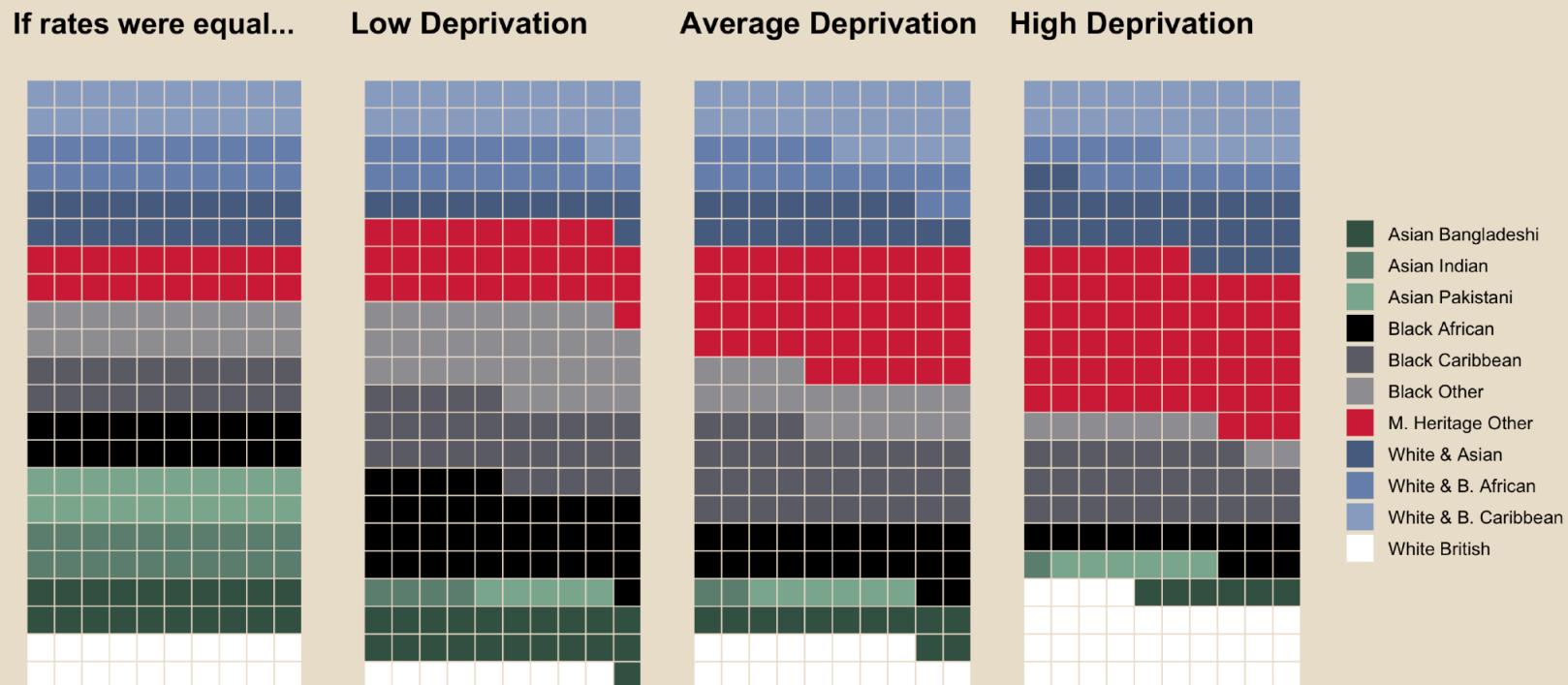
Neighbourhood Regression

Discussion Board



What would the ethnic profile of 220 Children Looked After look like in a Low, Average, and High Deprivation Neighbourhood if all neighbourhoods had identical ethnic population sizes?

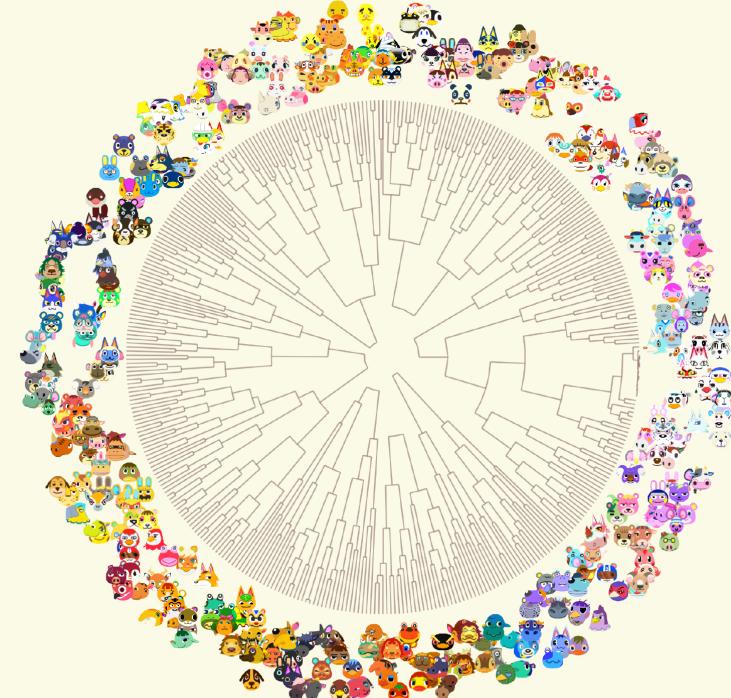
Calum Webb, the University of Sheffield



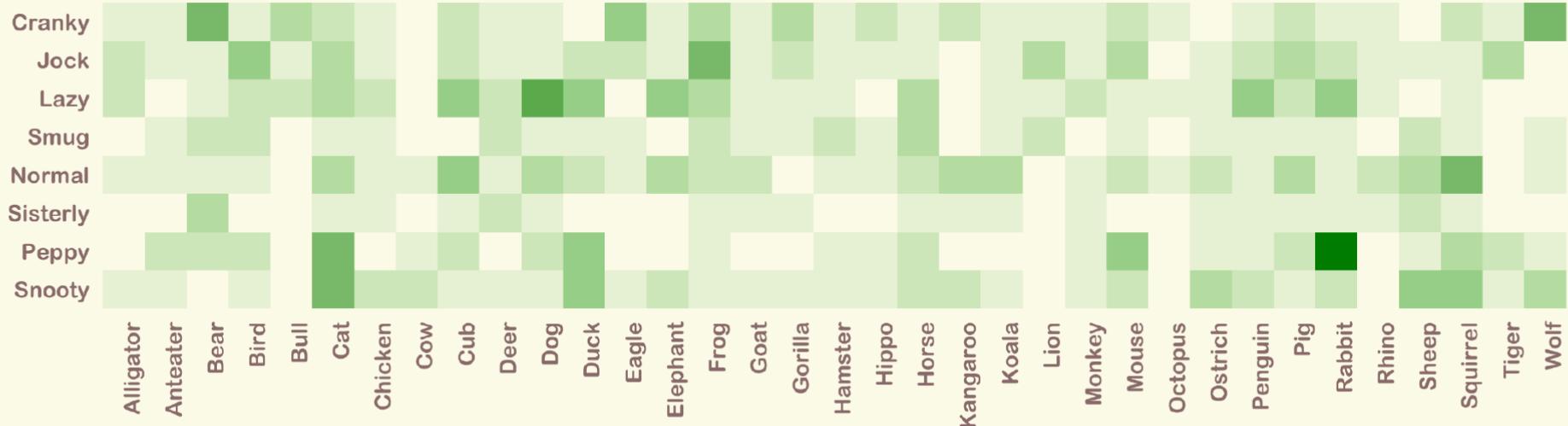
Visualising the Villagers

datavis meets Animal Crossing

Sources: Animal Crossing Icons (Nookipedia), species, and personality data (animalcrossing.fandom), Villager Popularity Rankings compiled by Pandoria, Mairen, & bloobelle from the Bell Tree Forums (28th March 2020)



I did a lot of Animal Crossing visualisation over the pandemic...

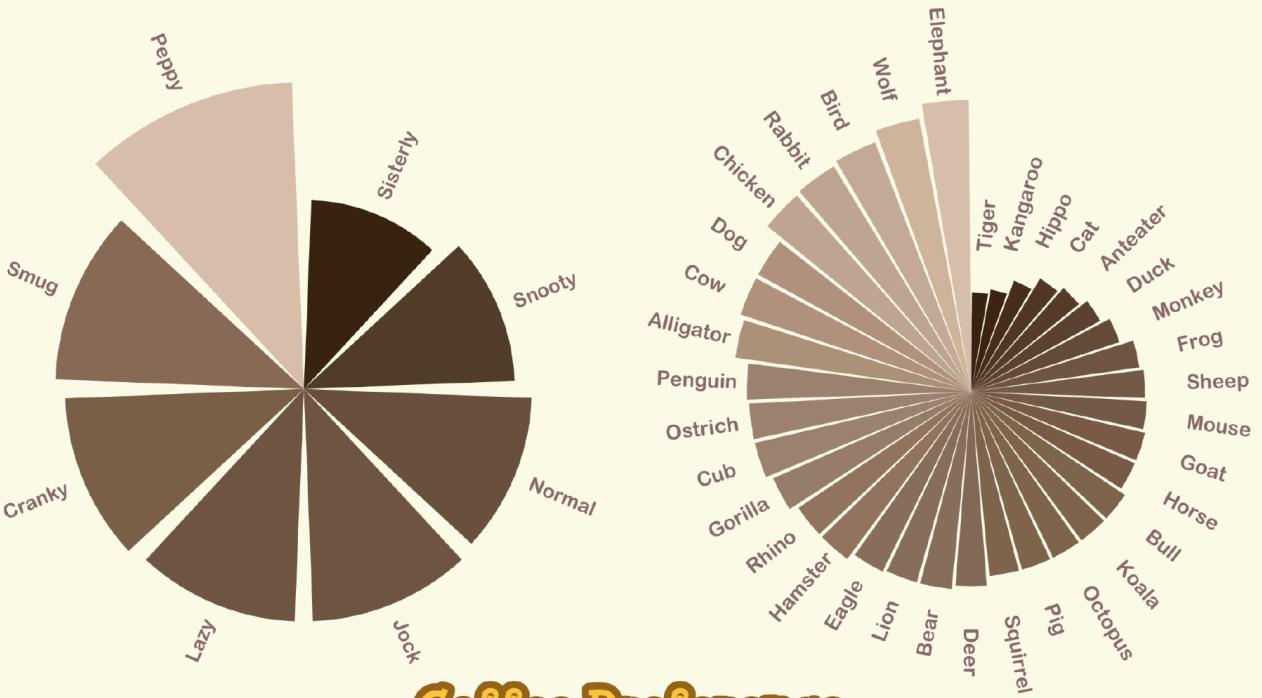


Personality x Species Combinations

Are some personality-species combinations more common than others?

A heatmap showing the numbers of each villager species and personality combination. Darker colours indicate more villagers of that combination (max = 8), completely transparent colours means no villagers of that combination. Cranky is the most common personality for Wolf and Bear villagers, Lazy is the most common for Dogs. The most common species and villager combination is Peppy Rabbits.

Like...



Coffee Preference

How do villagers like their coffee? (from New Leaf data)

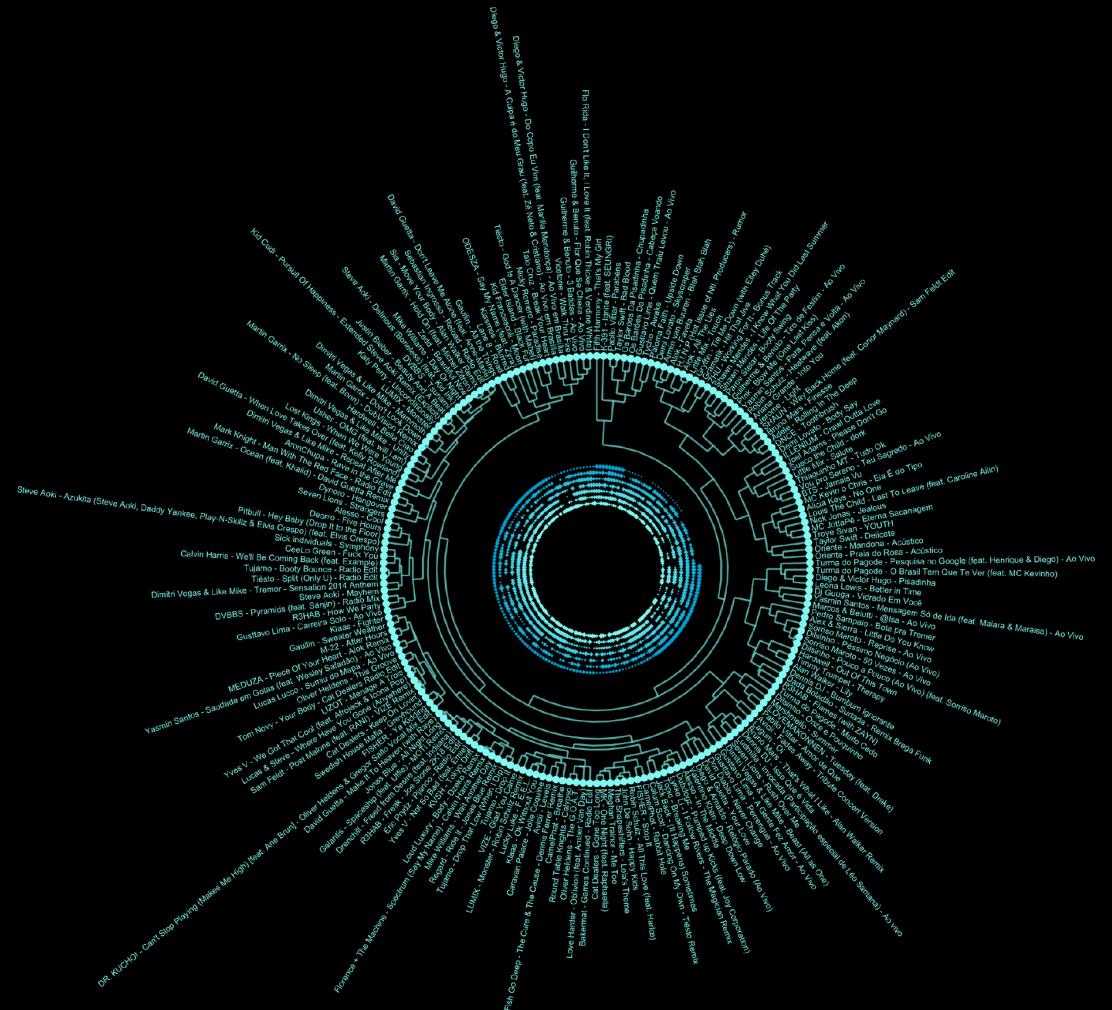
Sisterly/Uchi villagers are more partial to darker coffee with less sugar, while Peppy villagers prefer more sugar and milk in their brew. Sugar filled lattes are more common among elephants, wolf, bird, rabbit and chicken villagers while black coffee is more popular with tiger, kangaroo, hippo, and cat villagers.

A lot.

Most popular edm tracks, year of release weighted, by audio similarity

Audio features: tempo, valence, acousticness, speechiness, loudness, energy, danceability

data: sonify · data visualization: @sonify · created for Today Tuesday 21.01.2020



**Enough about me, let me
ask something about you.**

Go to Wooclap.com and enter code: **FNJKHC**



Part I

Overview of the module: How will it work?

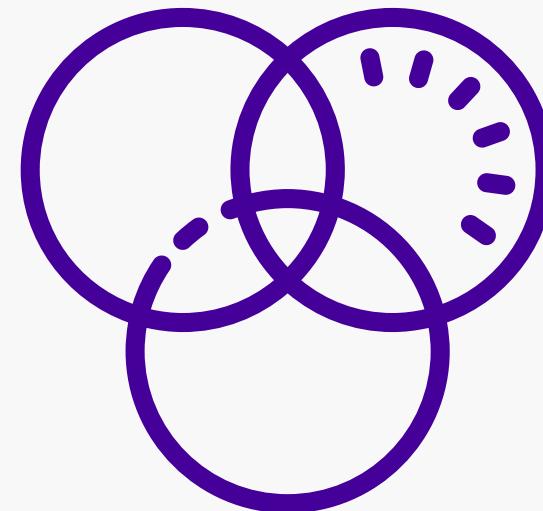


By the end of this module, you should know about:

- The **theory** of data visualisation
- The **practice** of data visualisation
- How to be **critical** about both the theory and practice of data visualisation

To do that, you will be learning how to *create* data visualisations using **R**, how to *design* data visualisations using best practice, and how to be *critical* of the data visualisations you make and see.

We don't assume you know anything about data visualisation, coding, or information design before starting this class!

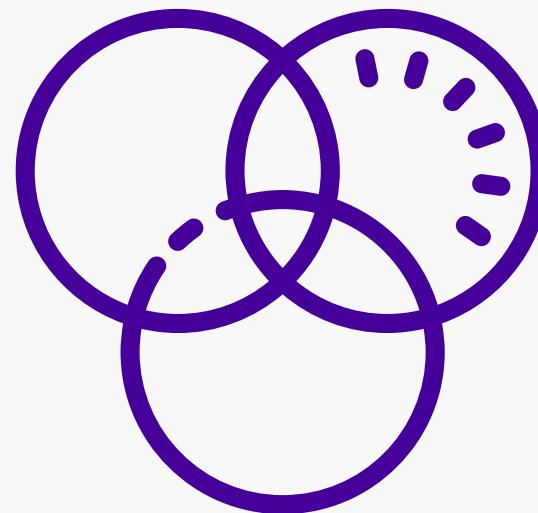


By the end of this module, you should know about:

- The **theory** of data visualisation
 - **Why is the data visualised in the way it is?**
- The **practice** of data visualisation
 - **How are key ideas implemented to create effective data visualisation?**
- How to be **critical** about both the theory and practice of data visualisation
 - **What is the impact of the things we visualise and the way we visualise them on society and culture?**

To do that, you will be learning how to *create* data visualisations using **R**, how to *design* data visualisations using best practice, and how to be *critical* of the data visualisations you make and see.

We don't assume you know anything about data visualisation, coding, or information design before starting this class!



Logistics

- **Two** teaching sessions per week: 1 hour lecture & 1.5 hour practical workshop

Logistics

- **Two** teaching sessions per week: 1 hour lecture & 1.5 hour practical workshop
- **Lectures** will mostly focus on theory and critical awareness
 - 4pm - 4:50pm Sheffield University Management School, MS Lowe Memorial LT03 (Uphill from the Wave) (Tuesdays, Week 1-5, 7-11)

Logistics

- **Two** teaching sessions per week: 1 hour lecture & 1.5 hour practical workshop
- **Lectures** will mostly focus on theory and critical awareness
 - 4pm - 4:50pm Sheffield University Management School, MS Lowe Memorial LT03 (Uphill from the Wave) (Tuesdays, Week 1-5, 7-11)
- **Workshops** will focus on the practice of creating data visualisations
 - 09:00-10:30/10:30-12:00/12:00-13:30 in The Wave, Computer Room 2A (Thursdays Week 1-5, 7-11), check your timetable for group allocation.

Logistics

- **Two** teaching sessions per week: 1 hour lecture & 1.5 hour practical workshop
- **Lectures** will mostly focus on theory and critical awareness
 - 4pm - 4:50pm Sheffield University Management School, MS Lowe Memorial LT03 (Uphill from the Wave) (Tuesdays, Week 1-5, 7-11)
- **Workshops** will focus on the practice of creating data visualisations
 - 09:00-10:30/10:30-12:00/12:00-13:30 in The Wave, Computer Room 2A (Thursdays Week 1-5, 7-11), check your timetable for group allocation.
- Slides and worksheets will be posted on Blackboard, usually worksheets will be posted after the workshops.
- Lectures will be recorded either via Encore Lecture Capture or Blackboard Collaborate. **We do our best to ensure all lectures are recorded, but cannot guarantee this.**

Lecture recordings and uploaded worksheets are not a replacement for attending lectures and worksheets.



If attendance at lectures and workshops becomes very poor, we will stop providing lecture recordings.



Weekly Tasks

- For most weeks there are **core** and **supplementary** tasks.
 - **You are required to complete core weekly tasks.** These are *formative* (ungraded) assessments; you will receive group feedback in the weekly workshops to help you improve and prepare for your *summative* (graded) assessments.
 - A full breakdown of all weekly tasks and topics is available [in the module handbook/resource list](#).
 - You are not required to complete the supplementary tasks, but you will probably do better in your assessment if you do.

This is a 10-credit module and therefore it is expected it will require 100 study hours per student. Formal teaching hours for this module will take 25 hours (lectures and workshops). Therefore, as a rough guide, this means you should expect to devote around 5 hours of study time per week to this module during a 15-week semester.



Assessment

There are two assessments:

- **Assessment 1:** due in November, is worth **30% of final grade**. You will be provided with some real contemporary social science data and asked to produce several different graphs and reflect on what they tell us about the topic.
- **Assessment 2:** due in January, is worth **70% of final grade**. You will be asked to critique and reflect on a published article that contains a data visualisation, and consider how it might be improved. Then, you will need to improve it using the same or related data, explaining the decisions you took.



Getting help with your studies

- **Consultation and Feedback hours**

I run two Consultation and Feedback hours per week, you can ask me questions about SMI105.

If you want to book an appointment with me, please use the calendar [here](#) or on Blackboard.

- **Piazza Discussion Board**

If you have any questions about the module or any tasks, including errors with **R** code, you can post for help on [the Piazza Discussion Board](#). This is accessible through Blackboard. Also, try to help other students with their questions — it will help you learn!



Additional Resources & Workshops from the University

Library

Research Skills and Critical Thinking (RSCT) workshops:

- Discovering Information: Getting Started
- Discovering Information: A masterclass

Online resources:

- How to read a journal article
- Active reading for understanding
- Tutorials A-Z

301 Skills Centre

- Academic skills refresher
- Reading and note-taking at university
- Managing your time and avoiding distractions
- Study Skills 1-to-1s

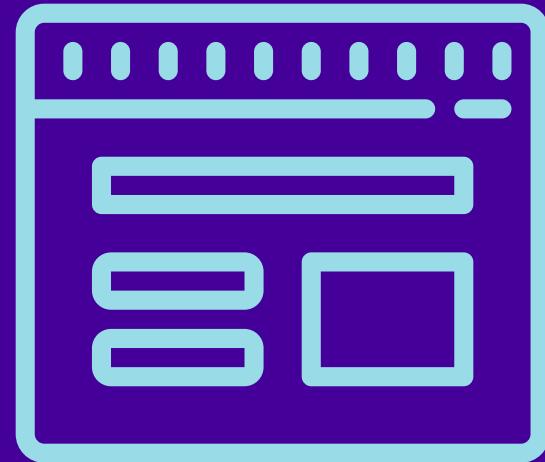
M.A.S.H.

Statistics resources:

- First steps in R
- R/Excel/SPSS for Ultra-Beginners
- 1-to-1 Support



Let's take a quick tour of the
Blackboard page for SMI105



Part II

Introduction to Data Visualisation



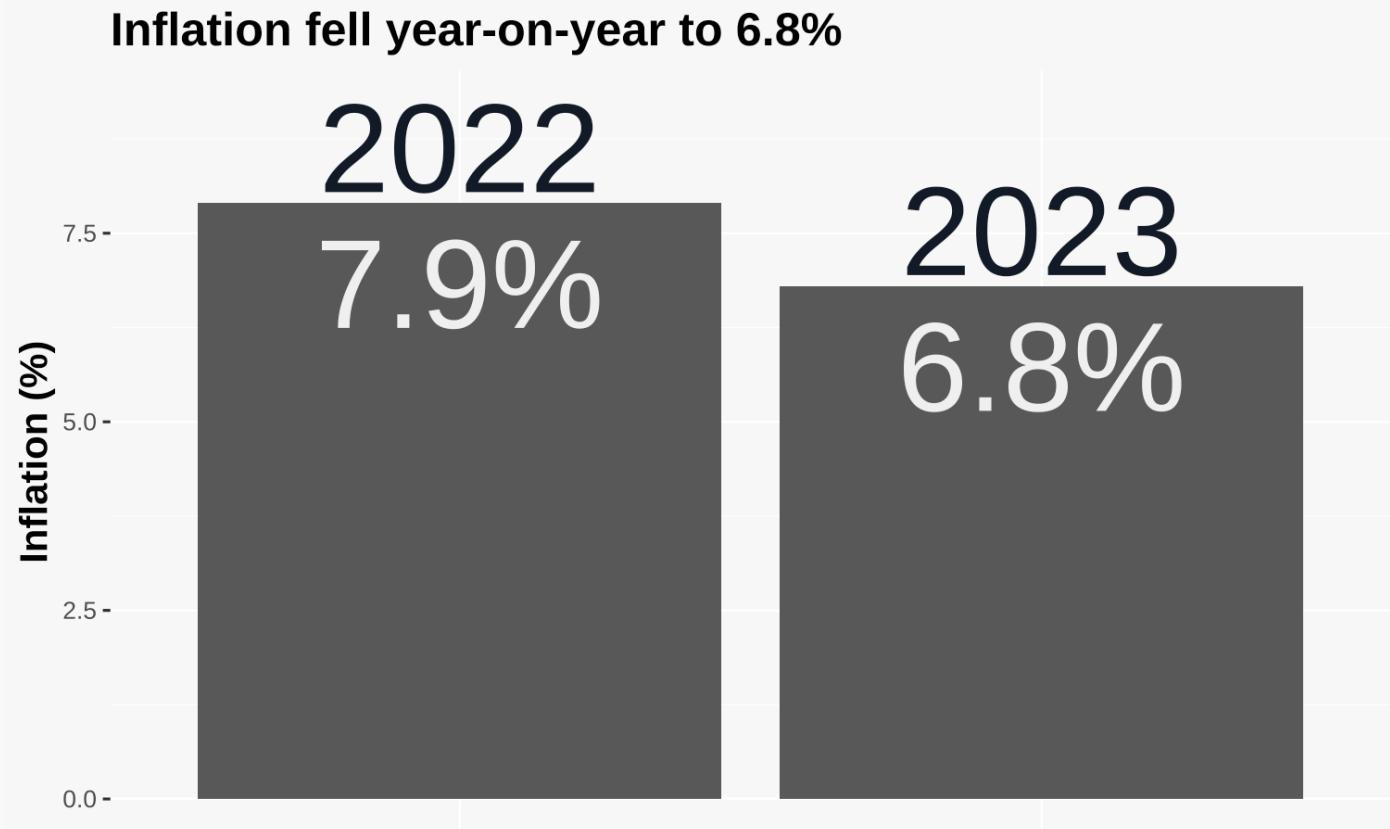
What is the purpose of data visualisation? Why do we do it?

- Re-open the wooclap.com app and go to Question/Activity 2



Data Visualisation 1

What is the main purpose?

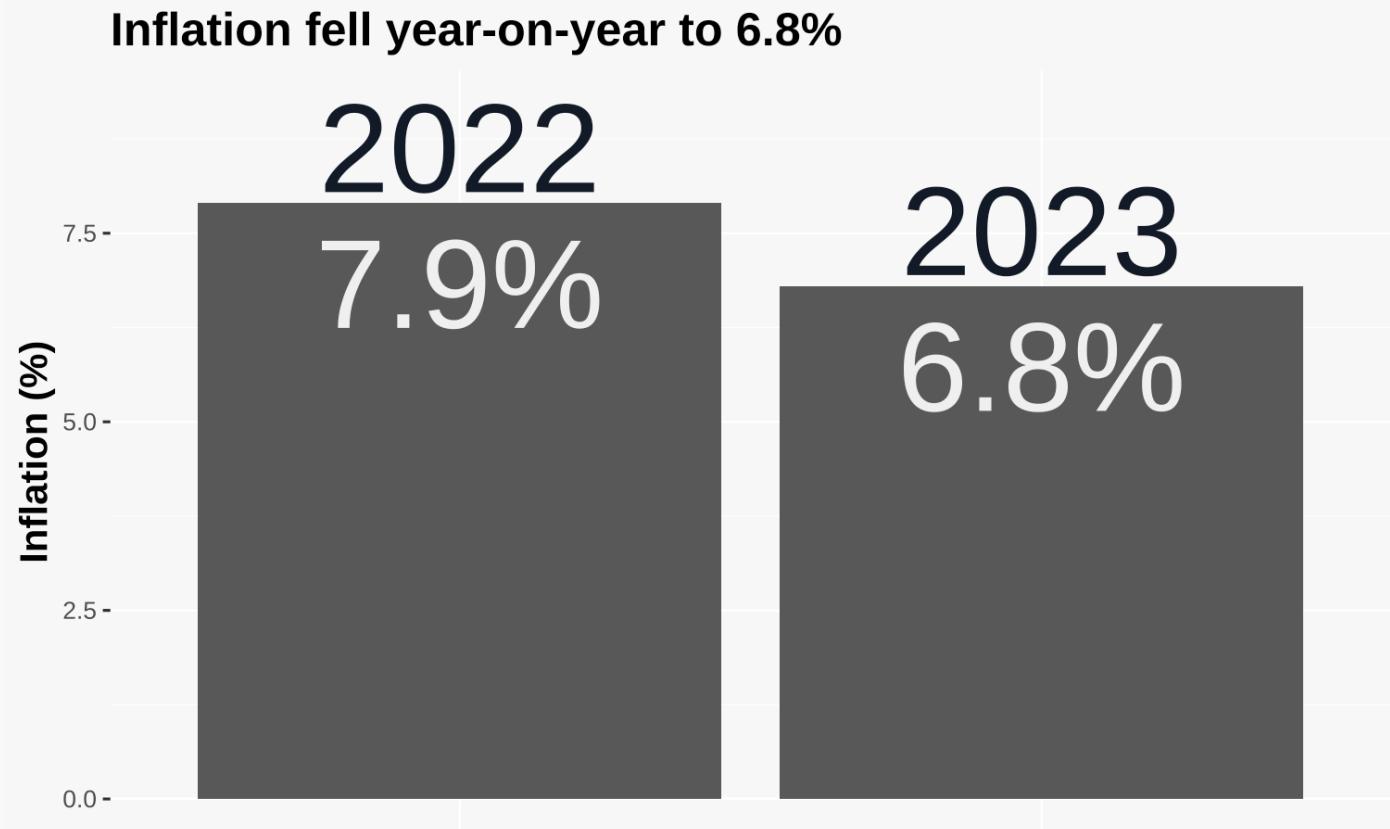


Woopla event code: app.woopla.com/FNJKHC (Activity 2)

Data Visualisation 1

What is the main purpose?

- To communicate information.

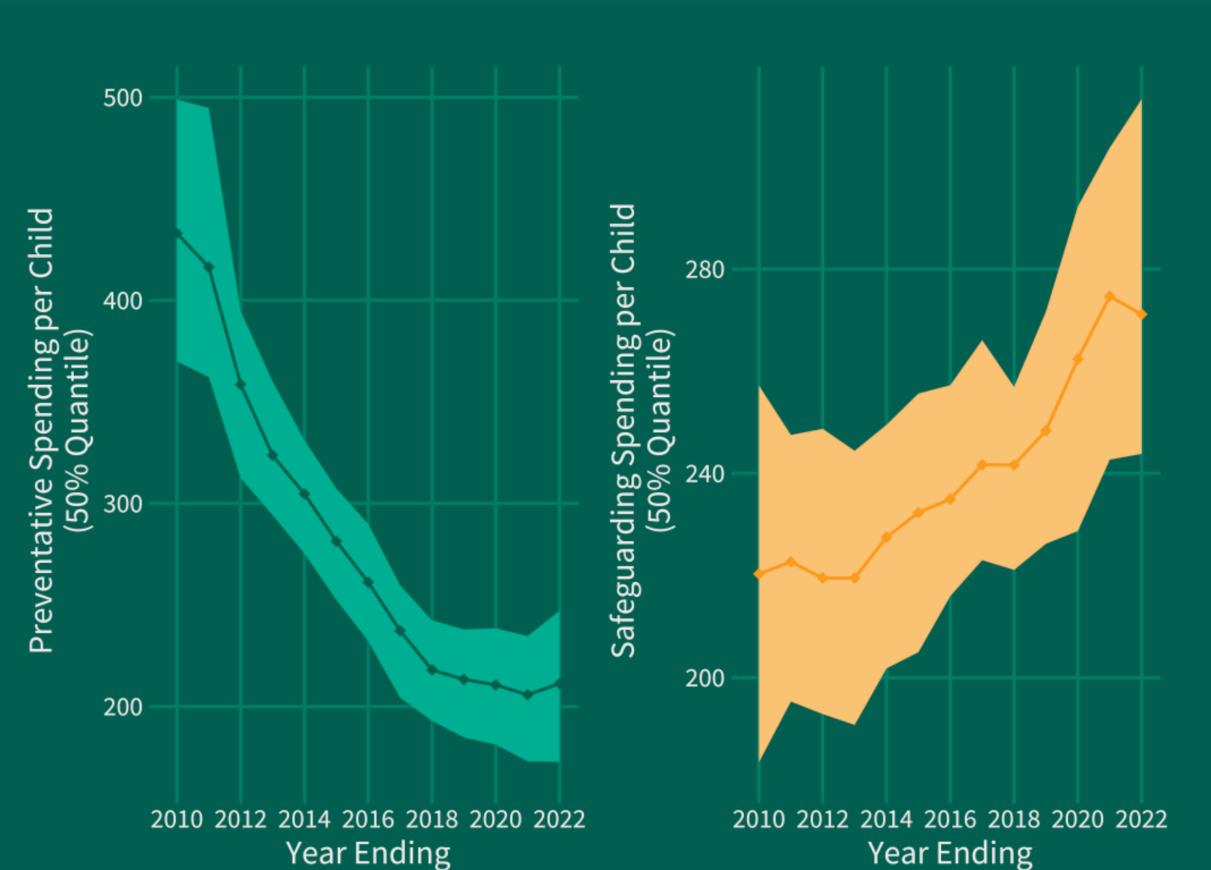


Woopla event code: app.woopla.com/FNJKHC (Activity 2)

Data Visualisation 2

What is the main purpose?

- To communicate information

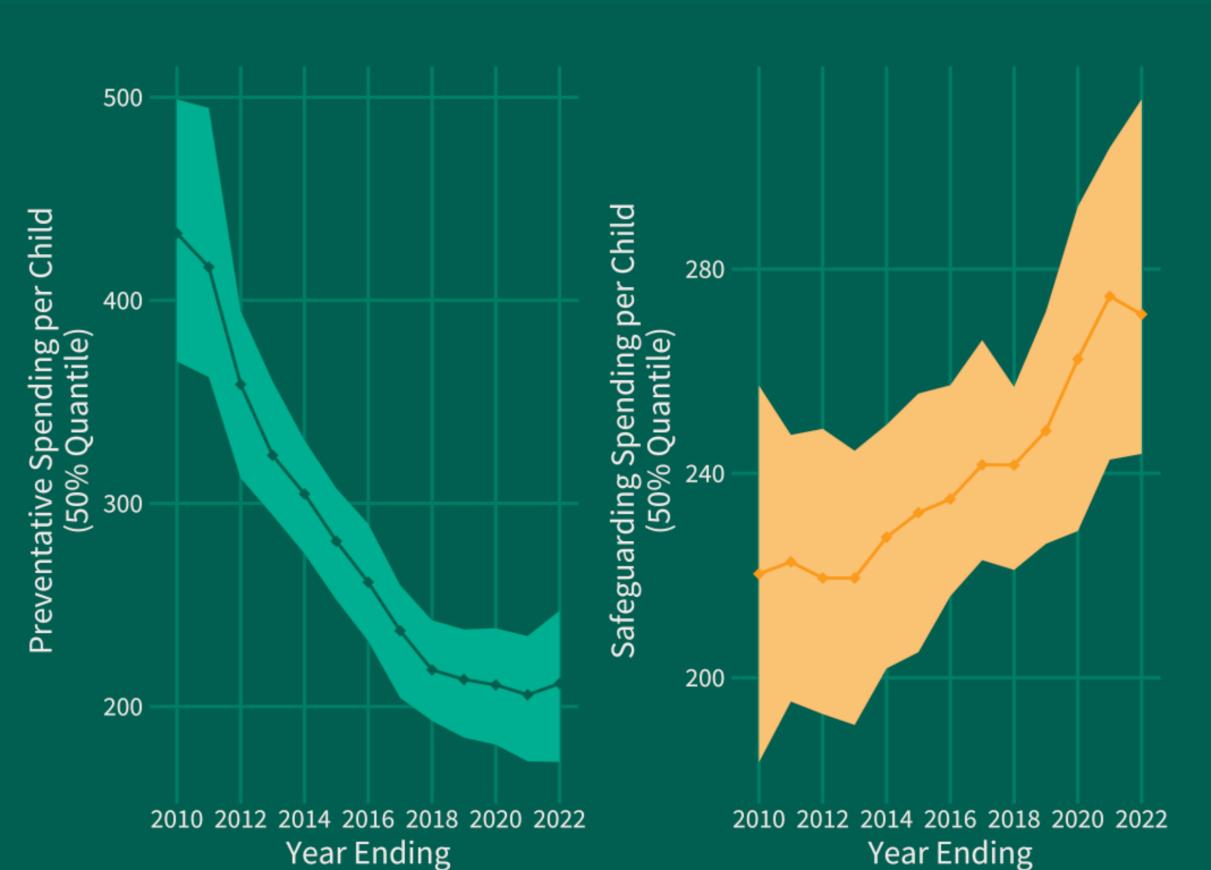


Woopla event code: app.woopla.com/FNJKHC (Activity 2)

Data Visualisation 2

What is the main purpose?

- To communicate information
- To tell a story



Woopla event code: app.woopla.com/FNJKHC (Activity 2)

Data Visualisation 3

What is the main purpose?

- To communicate information
- To tell a story

On a piece of paper in your notebook (if you have one), or by drawing with your finger or describing to your neighbour, draw a representation of the following table to illustrate the difference between "Party A"'s £2million commitment to spending and "Party B"'s £2billion commitment to spending.

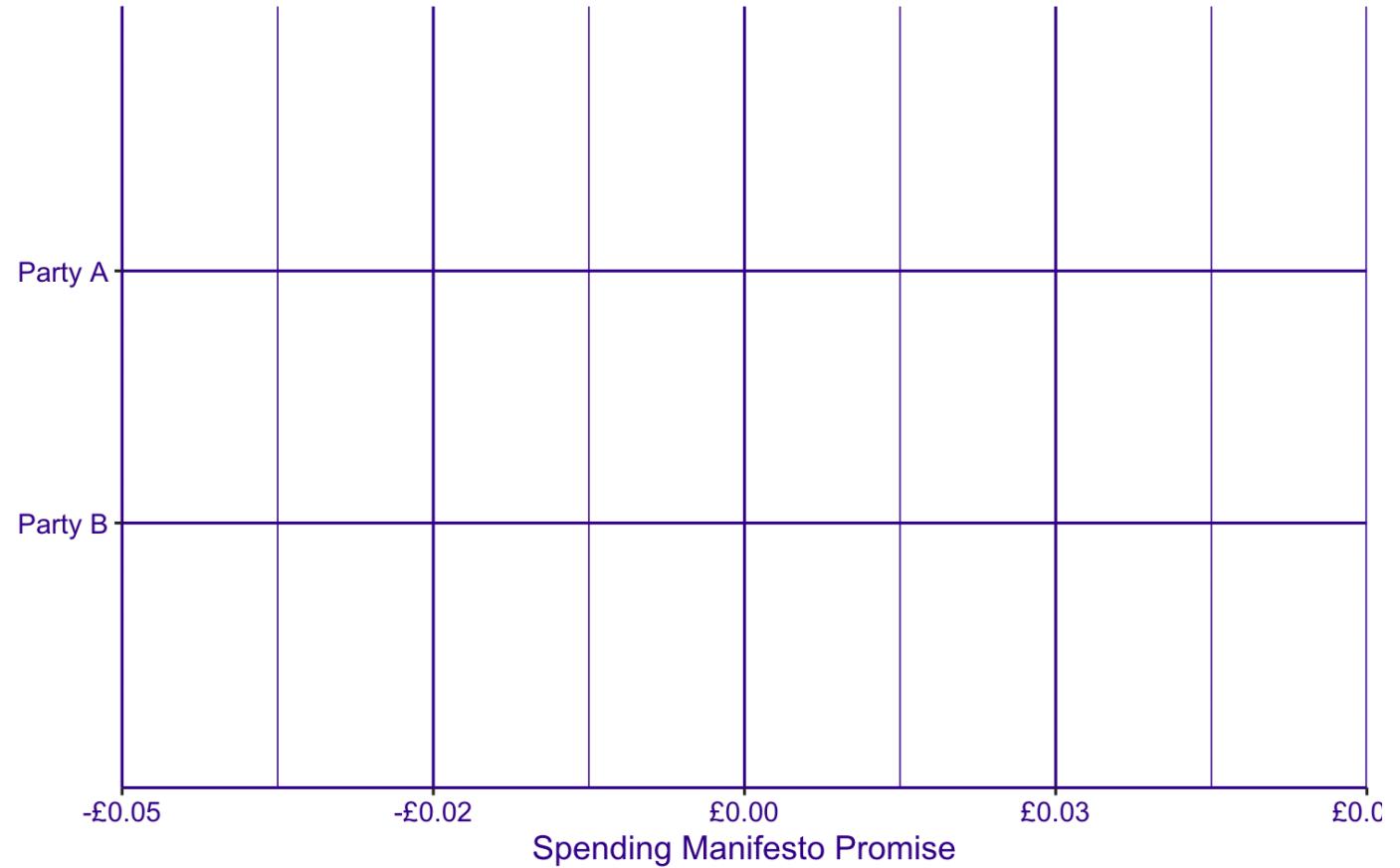
| Party | Spending Commitment |
|---------|---------------------|
| Party A | £2million |
| Party B | £2billion |

Woopla event code: app.wooplap.com/FNJKHC (Activity 2)

Data Visualisation 3

What is the main purpose?

- To communicate information
- To tell a story
- To use our visual sense to understand data, detect patterns, etc., rather than language and logical processing

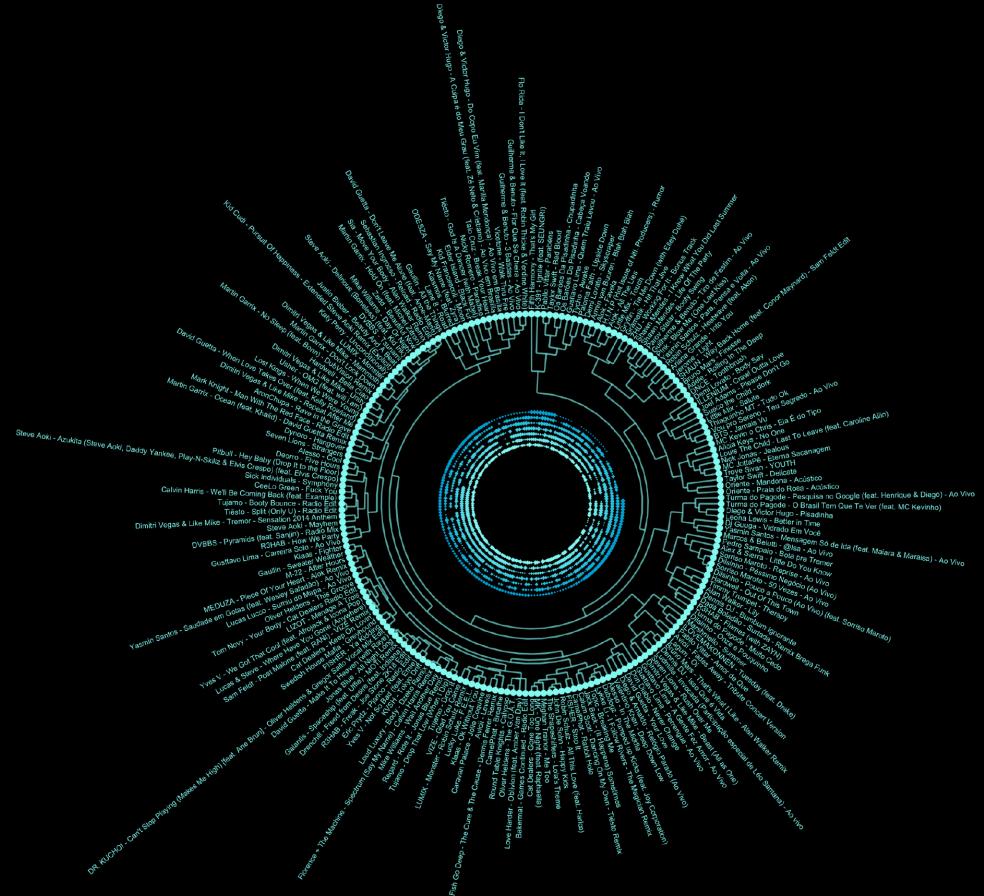


Woopla event code: app.woopla.com/FNJKHC (Activity 2)

Data Visualisation 4

What is the main purpose?

- To communicate information
- To tell a story
- To use our visual sense to understand data, detect patterns, etc., rather than language and logical processing



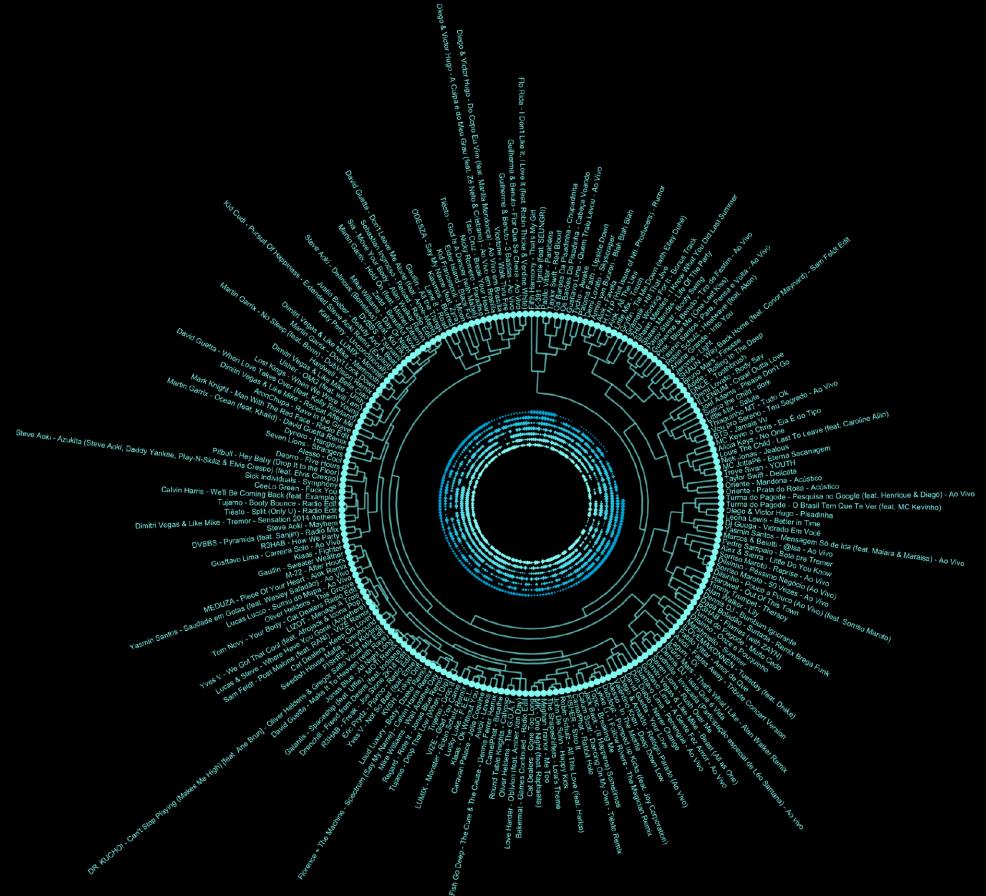
Wooclap event code: app.wooclap.com/FNJKHC (Activity 2)



Data Visualisation 4

What is the main purpose?

- To communicate information
 - To tell a story
 - To use our visual sense to understand data, detect patterns, etc., rather than language and logical processing
 - To look cool! Data visualisation can be art. E.g. see Generative Data Visualisation Art by Thomas Lin Pedersen and Danielle Navarro



Wooclap event code: app.wooclap.com/FNJKHC (Activity 2)



‘[Data visualisation is] the representation and presentation of data to facilitate understanding’

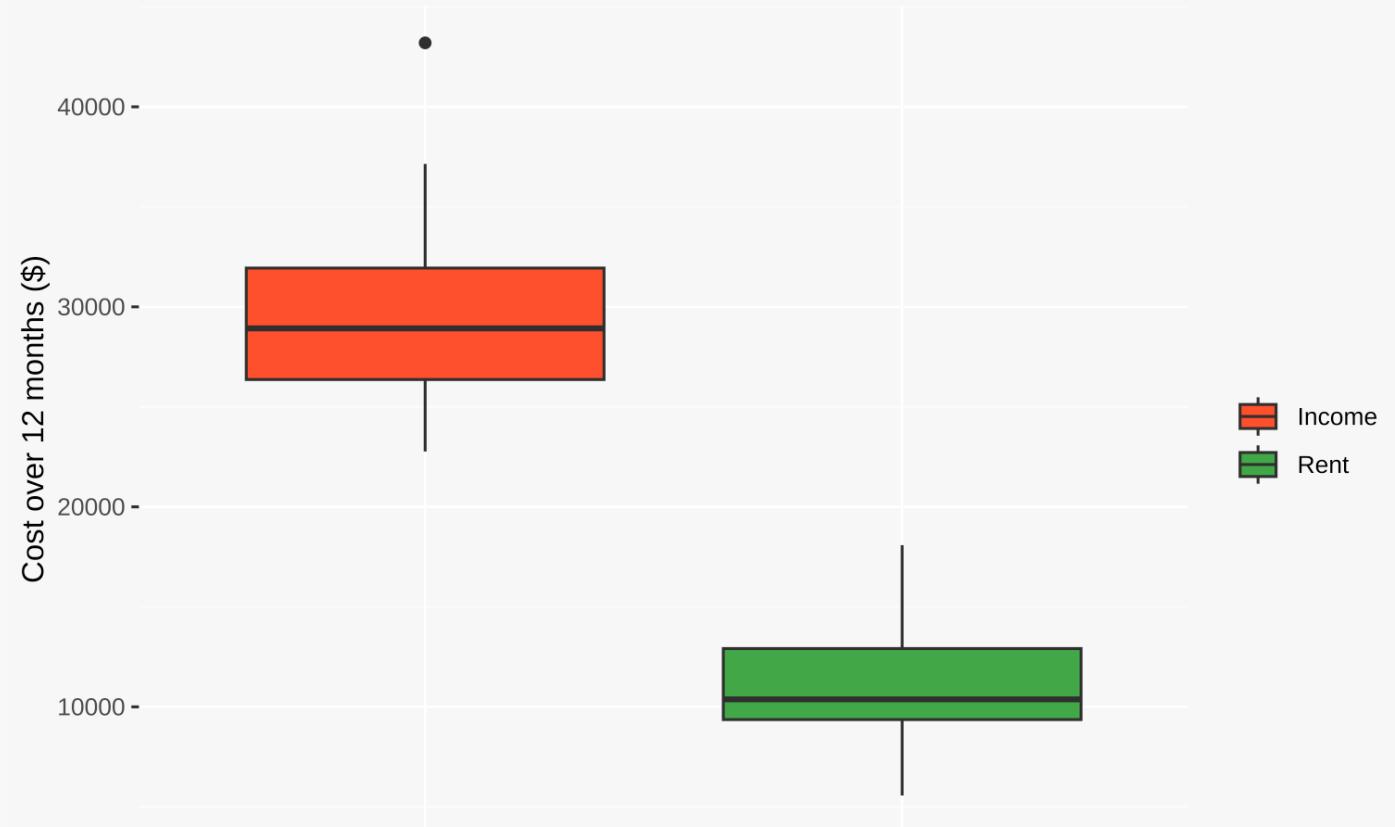
Andy Kirk, (2016), Data Visualisation: A Handbook for Data Driven Design



What assumptions do these data visualisations make, and reinforce, about society?



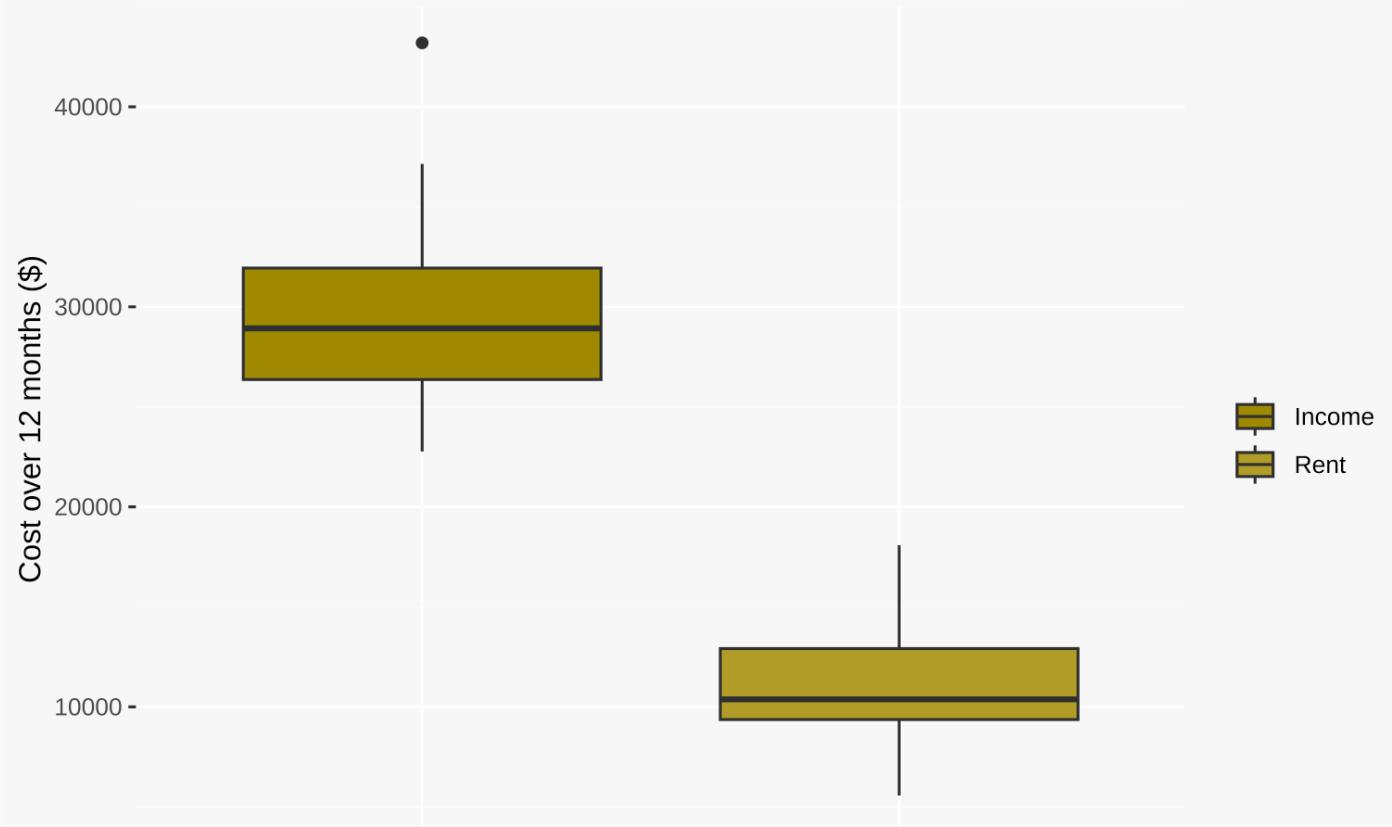
Data Visualisation 1



Woopla event code: app.woopla.com/FNJKHC (Activity 3)

Data Visualisation 1

- Assumes everyone is able to access red-green colour schemes



Woopla event code: app.woopla.com/FNJKHC (Activity 3)

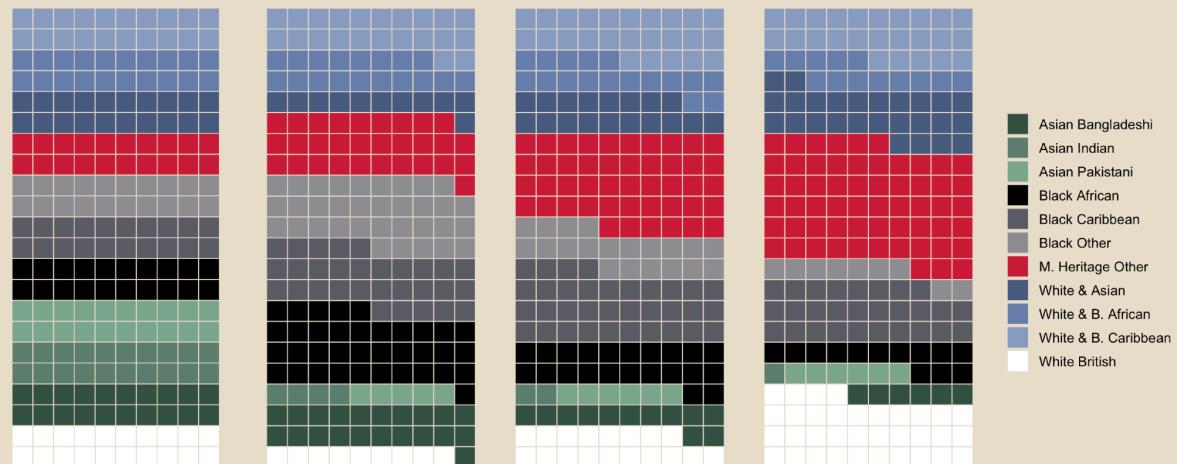
Data Visualisation 2

- Assumes everyone is able to access red-green colour schemes

What would the ethnic profile of 220 Children Looked After look like in a Low, Average, and High Deprivation Neighbourhood if all neighbourhoods had identical ethnic population sizes?

Calum Webb, the University of Sheffield

If rates were equal... Low Deprivation Average Deprivation High Deprivation



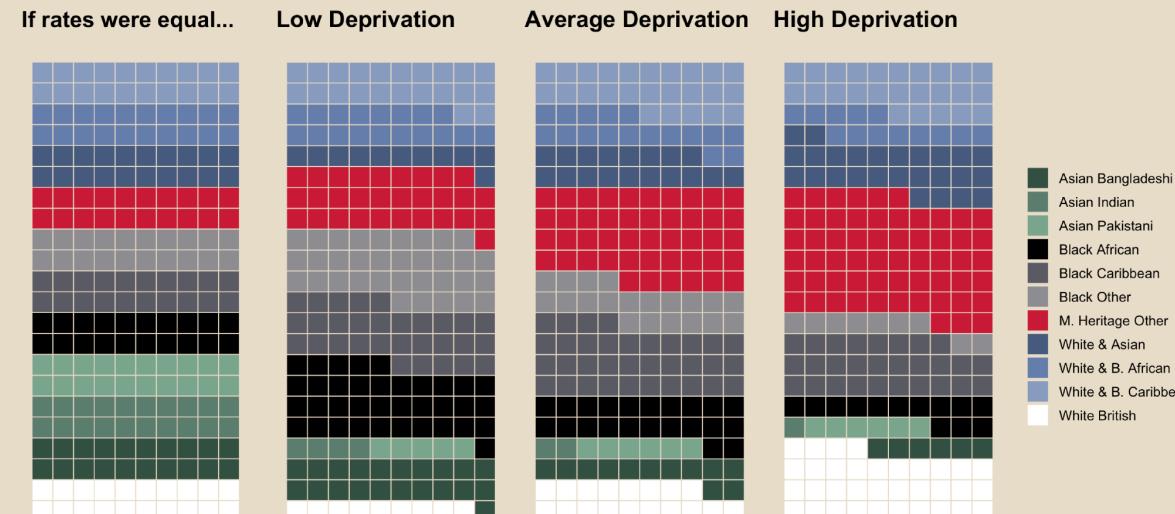
Woopla event code: app.woopla.com/FNJKHC (Activity 3)

Data Visualisation 2

- Assumes everyone is able to access red-green colour schemes
- Reinforces the idea that ethnicity and race can be categorised into distinct groups

What would the ethnic profile of 220 Children Looked After look like in a Low, Average, and High Deprivation Neighbourhood if all neighbourhoods had identical ethnic population sizes?

Calum Webb, the University of Sheffield



Woopla event code: app.woopla.com/FNJKHC (Activity 3)

Data Visualisation 3

- Assumes everyone is able to access red-green colour schemes
- Reinforces the idea that ethnicity and race can be categorised into distinct groups



Woopla event code: app.woopla.com/FNJKHC (Activity 3)

Data Visualisation 3

- Assumes everyone is able to access red-green colour schemes
- Reinforces the idea that ethnicity and race can be categorised into distinct groups
- Reinforces dominant (settler) forms of knowledge over other (indigenous) forms of knowledge



Woopla event code: app.woopla.com/FNJKHC (Activity 3)

“If you’re designing technology for society and you don’t know anything about society, you’re unqualified.”

Quote from Safiya Noble presenting her book Algorithms of Oppression, hosted by Open Data Manchester, 8th May 2018



“...some groups experience unearned advantages—because various systems have been designed by people like them and work for people like them—and other groups experience systematic disadvantages—because those same systems were not designed by them or with people like them in mind”

D'Ignazio, C. and Klein, L. (2020) Data Feminism (p.24)



Summary: So, what is data visualisation?

- Data visualisation is an **object** — a chart, a graph, a map — that follows a set of rules and principles, where parameters are defined by data

Summary: So, what is data visualisation?

- Data visualisation is an **object** — a chart, a graph, a map — that follows a set of rules and principles, where parameters are defined by data
- Data visualisation is also a **process** — an act of transformation and reproduction of best practice in order to tell a story or to make sense of the world visually

Summary: So, what is data visualisation?

- Data visualisation is an **object** — a chart, a graph, a map — that follows a set of rules and principles, where parameters are defined by data
- Data visualisation is also a **process** — an act of transformation and reproduction of best practice in order to tell a story or to make sense of the world visually
- Data visualisation is also a **social phenomenon** — it is related to power, can reinforce or challenges societal norms or assumptions, and shapes how people engage with information

The rest of this week:

This week's workshop:

- Computer Room 2a, SMI, The Wave
- You can take and use your own laptop if you like
- 09:00-10:30, 10:30-12:00, or 12:00-13:30

Do you know when and where you are going?

If not clear from your timetable which group you're in, email
smi@sheffield.ac.uk

You will not be permitted to move group without a very good reason.

The rest of this week:

This week's workshop:

- Computer Room 2a, SMI, The Wave
- You can take and use your own laptop if you like
- 09:00-10:30, 10:30-12:00, or 12:00-13:30

Do you know when and where you are going?

If not clear from your timetable which group you're in, email
smi@sheffield.ac.uk

You will not be permitted to move group without a very good reason.

Before week 2:

Core tasks:

- Before lecture: Read this module handbook in full
- Before workshop: If you've got a personal machine, get R and RStudio installed on it (instructions on how to do this will be in the Week 2 learning resources folder)
- Having done so, work through the workshop handout once again to check it works on your machine (if it doesn't, don't worry, we'll go through and troubleshoot it together)

Supplementary tasks:

- Before lecture: Read Top 5 things to look for in a visualisation on the Seeing Data website: <https://seeingdata.org/developing-visualisation-literacy/top-5-things-to-look-for-in-a-visualisation/>