MONASH University

MONASH
INFORMATION
TECHNOLOGY

FIT3179 Data Visualisation

Week 02: Visualisation Analysis & Design I

GROUP OF EIGHT AUSTRALIA

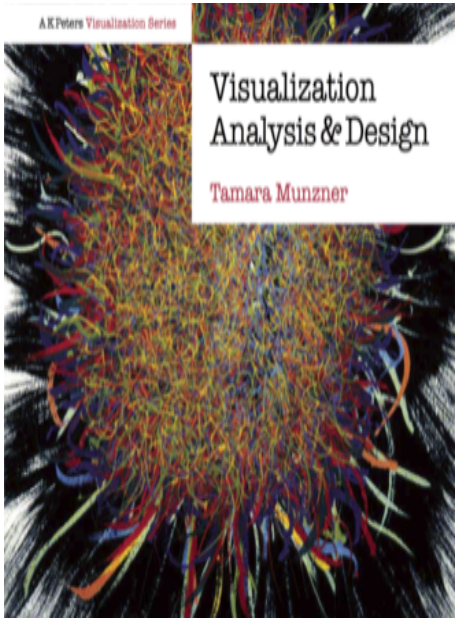
MONASH University

Lecture Overview

- Analysis Framework
  - What? – data abstraction
  - Why? – task abstraction
  - How? – intro to marks and channels
- Five design sheet methodology

Textbook

- For the next four lectures we will be focusing on the early chapters from the text:
  Visualization Analysis & Design
  by Tamara Munzner
- Available to read electronically for *free* from the Monash Library



Defining Visualisation

MONASH University

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Why have a human in the loop?

MONASH University

Computer-based visualization systems provide visual representations of datasets designed to **help people carry out tasks more effectively**.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

Why have a human in the loop?

MONASH University

Don't need viz when fully automatic solution exists and is trusted

- Many analysis problems ill-specified
  - don't know exactly what questions to ask in advance
- Possibilities
  - long-term use for end users (e.g. exploratory analysis of scientific data)
  - presentation of known results
  - stepping stone to better understanding of requirements before developing models
  - help developers of automatic solution refine/debug, determine parameters
  - help end users of automatic solutions verify, build trust

Why represent all the data?

MONASH University

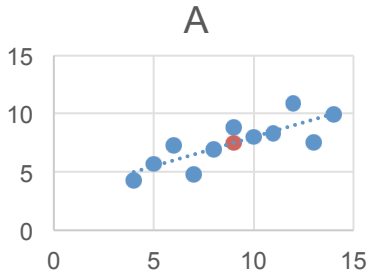
*Computer-based visualization systems provide visual **representations of datasets** designed to help people carry out tasks more effectively.*

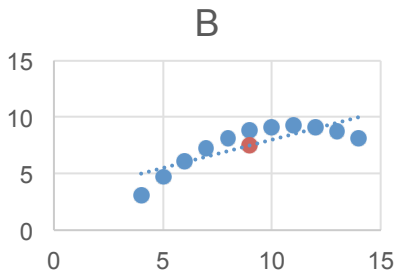
- summaries lose information, details matter
  - confirm expected and find unexpected patterns
  - assess validity of statistical model

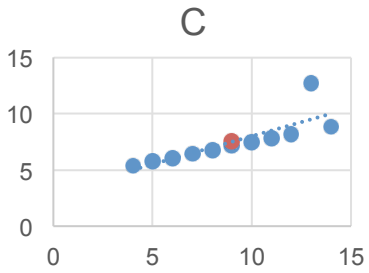
Anscomb's Quartet

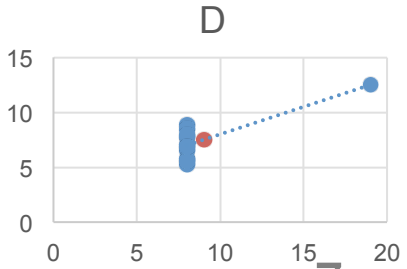
Dataset A		Dataset B		Dataset C		Dataset D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

mean	9	7.500	9	7.500	9	7.500	9	7.500
variance	11	4.127	11	4.127	11	4.127	11	4.127
x/y correlation		0.816		0.816		0.816		0.816
regression gradient		0.500		0.500		0.500		0.500

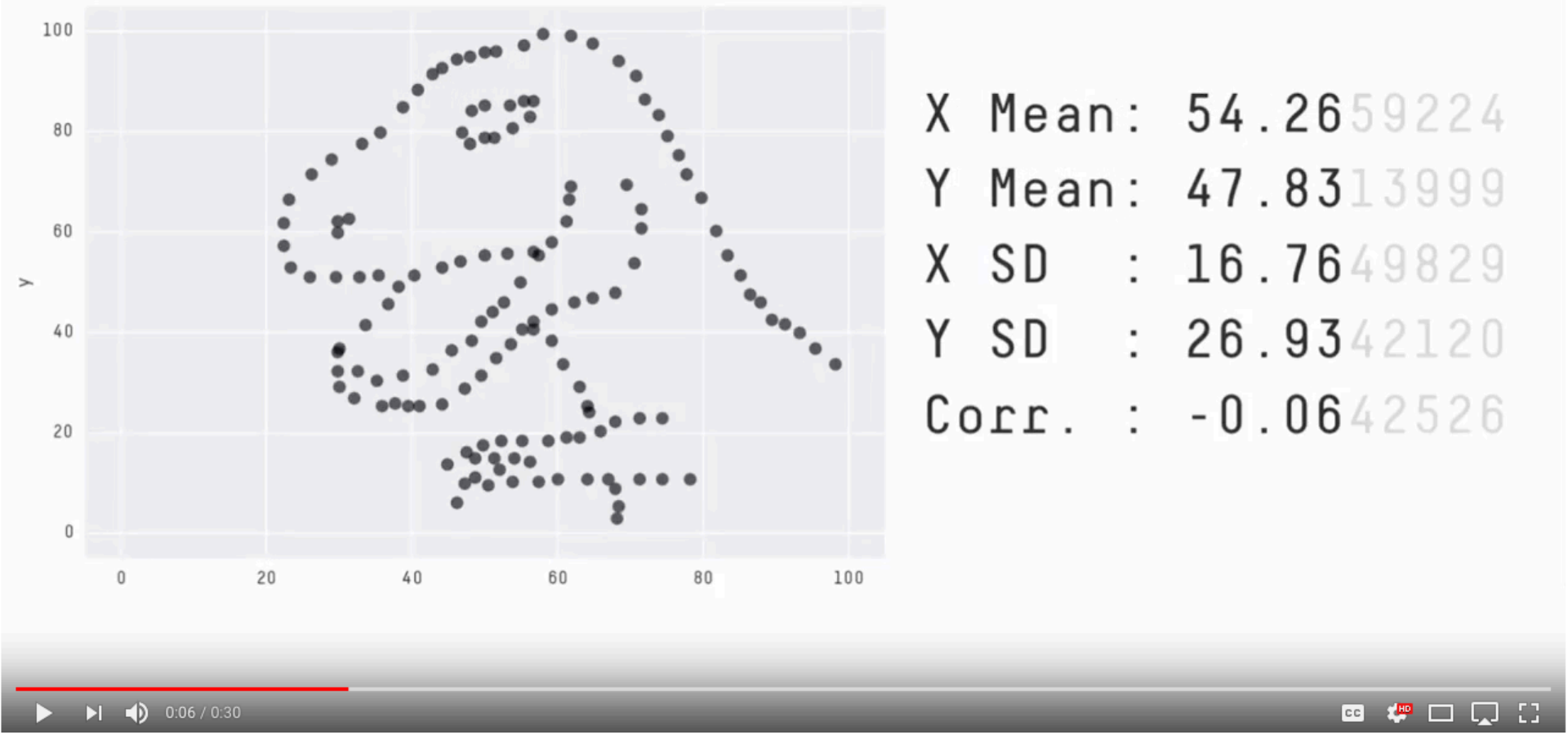








Why represent all the data?



https://youtu.be/It4UA75z_KQ

Matejka, J. and Fitzmaurice, G. 2017. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. CHI '17.

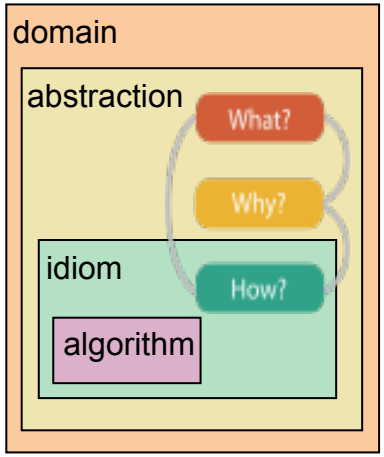
MONASH University

Analysis framework: Four levels, three questions

Textbook chapter 4

Analysis framework: Four levels, three questions

- *Domain: a field of interest of the users*
  - *domain-specific vocabulary, data and workflows*
  - who are the target users?
- *Task and data abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **Data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **Task abstraction** from domain-specific to generic task
- *Idiom for visual encoding and interaction*
  - **how** is it shown?
    - **visual encoding idiom**: how to draw
    - **interaction idiom**: how to manipulate
- *Algorithm*
  - efficient computation



A Multi-Level Typology of Abstract Visualization Tasks
Brehmer and Munzner. IEEE TVCG 19(12):2376–2385, 2013.

Textbook chapter 2

Data Attributes

MONASH University

Attribute Types

- Categorical
- Ordered
  - *Ordinal*
  - *Quantitative*

Ordering Direction

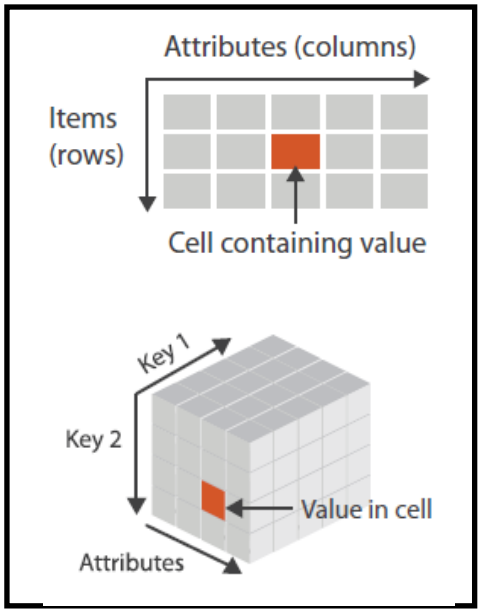
- Sequential
- Diverging
- Cyclic

VAD pg. 32

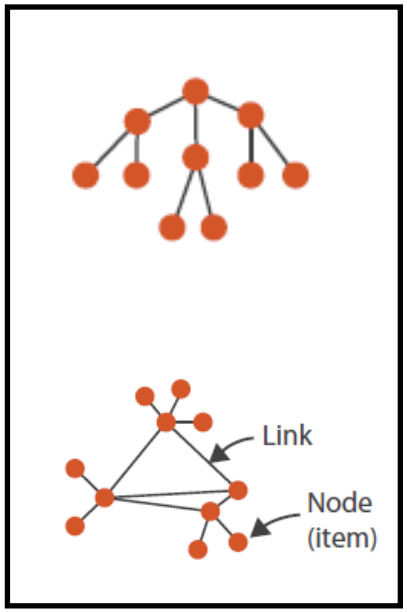
Dataset Types

MONASH University

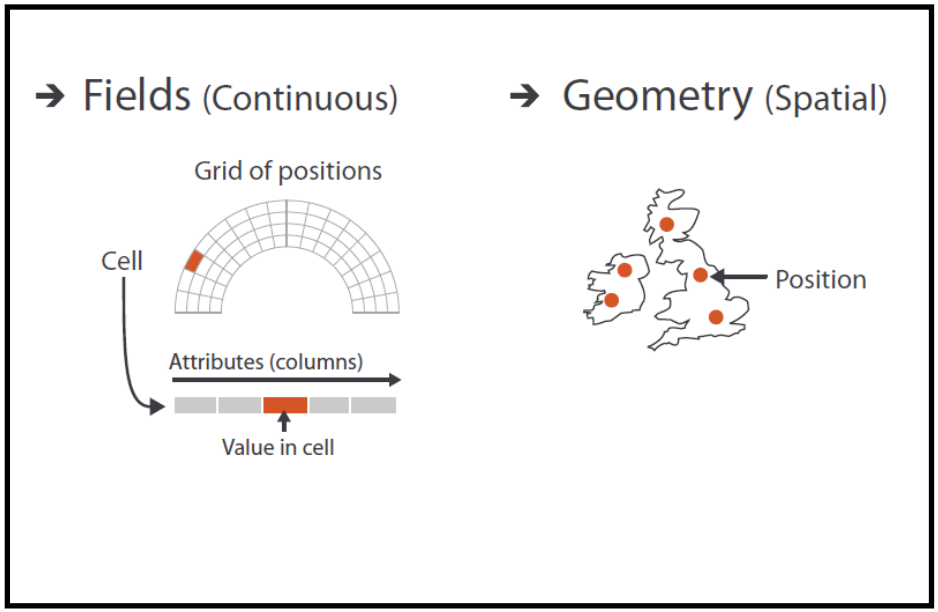
Tables



Trees & Networks



Spatial



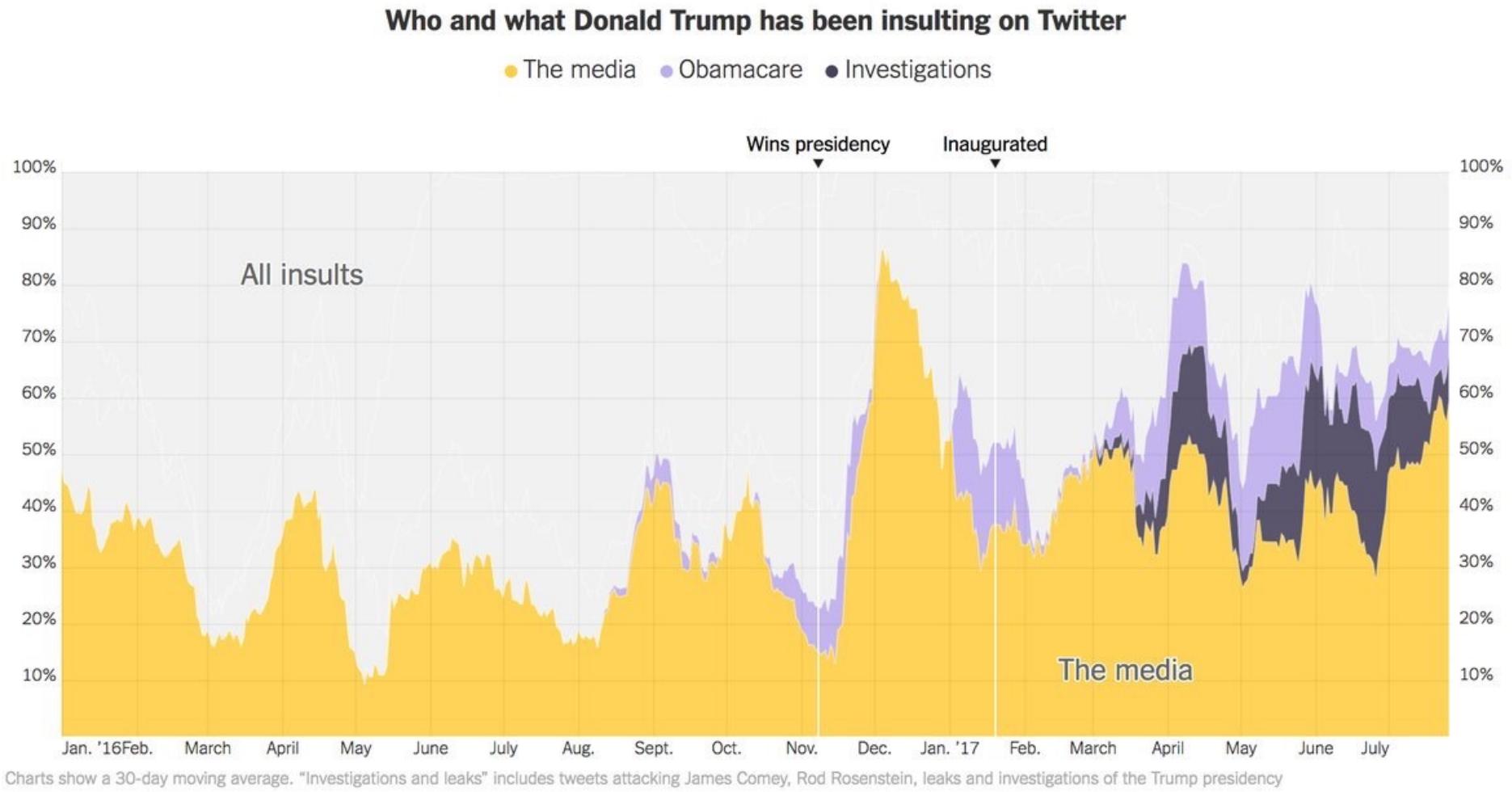
Network synonyms
network = graph
node = vertex
link = edge

Uniform grid: implicit geometry and topology
A grid cell can have one or more attributes.
Scalar field: 1 attribute per cell.
Vector field: 2 or more attributes with a direction
Tensor field: many attributes per cell.

A geometry item (point, line or area) can have attributes.

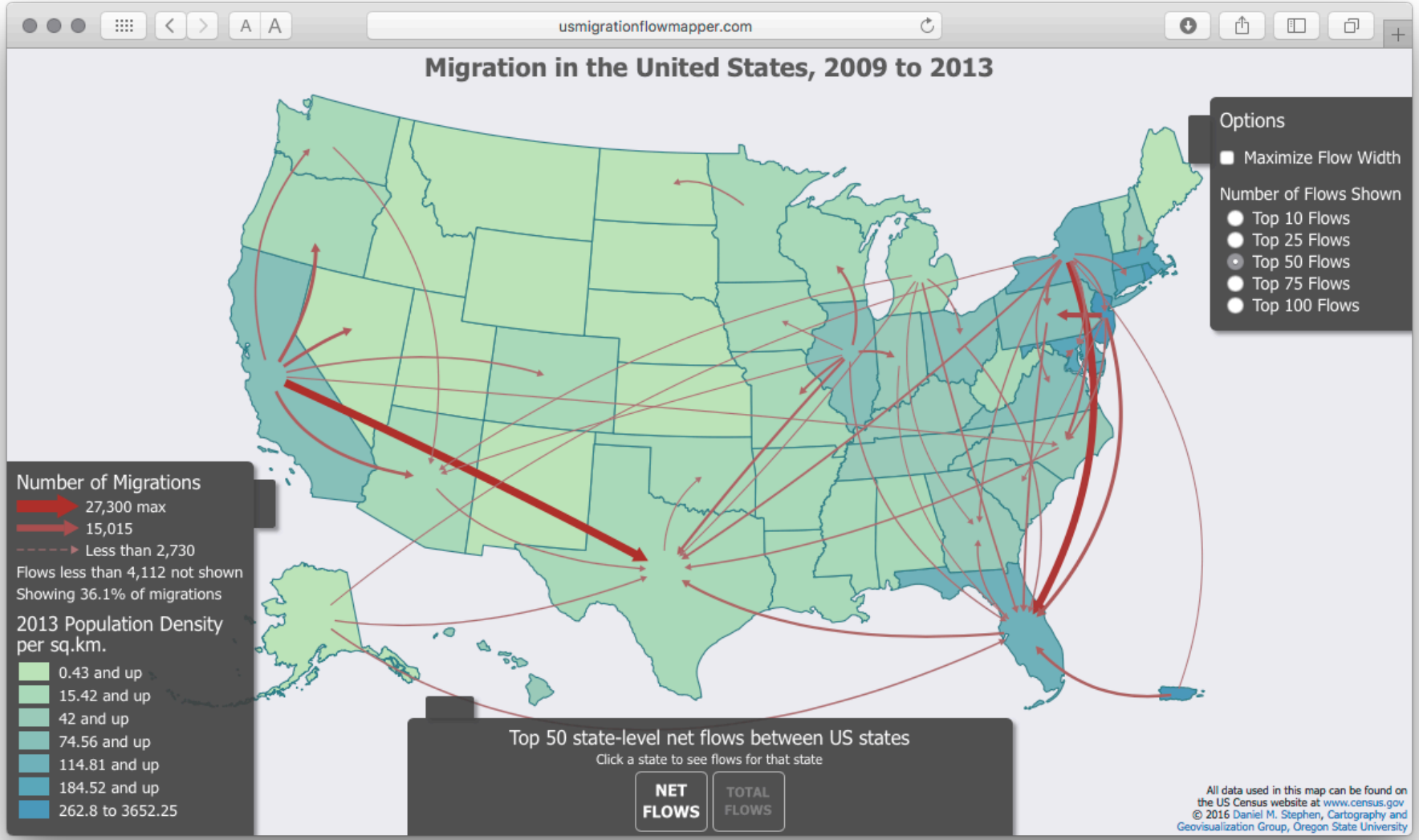
VAD pg. 25

Dataset type? Attribute type?



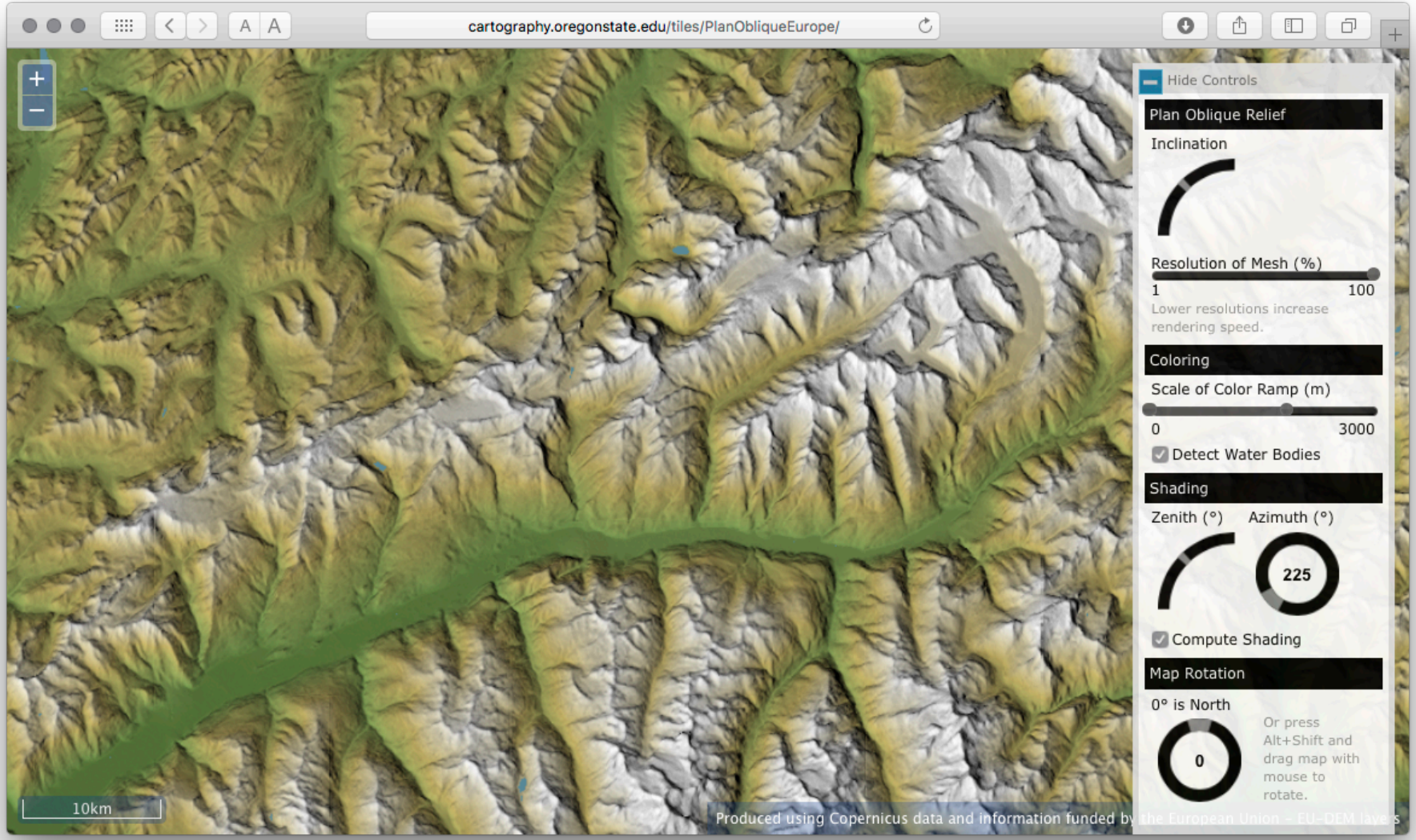
Source: New York Times, 26 July 2017
https://www.nytimes.com/interactive/2017/07/26/upshot/president-trumps-newest-focus-discrediting-the-news-media-obamacare.html

Dataset type? Attribute type?



Source: *http://usmigrationflowmapper.com*

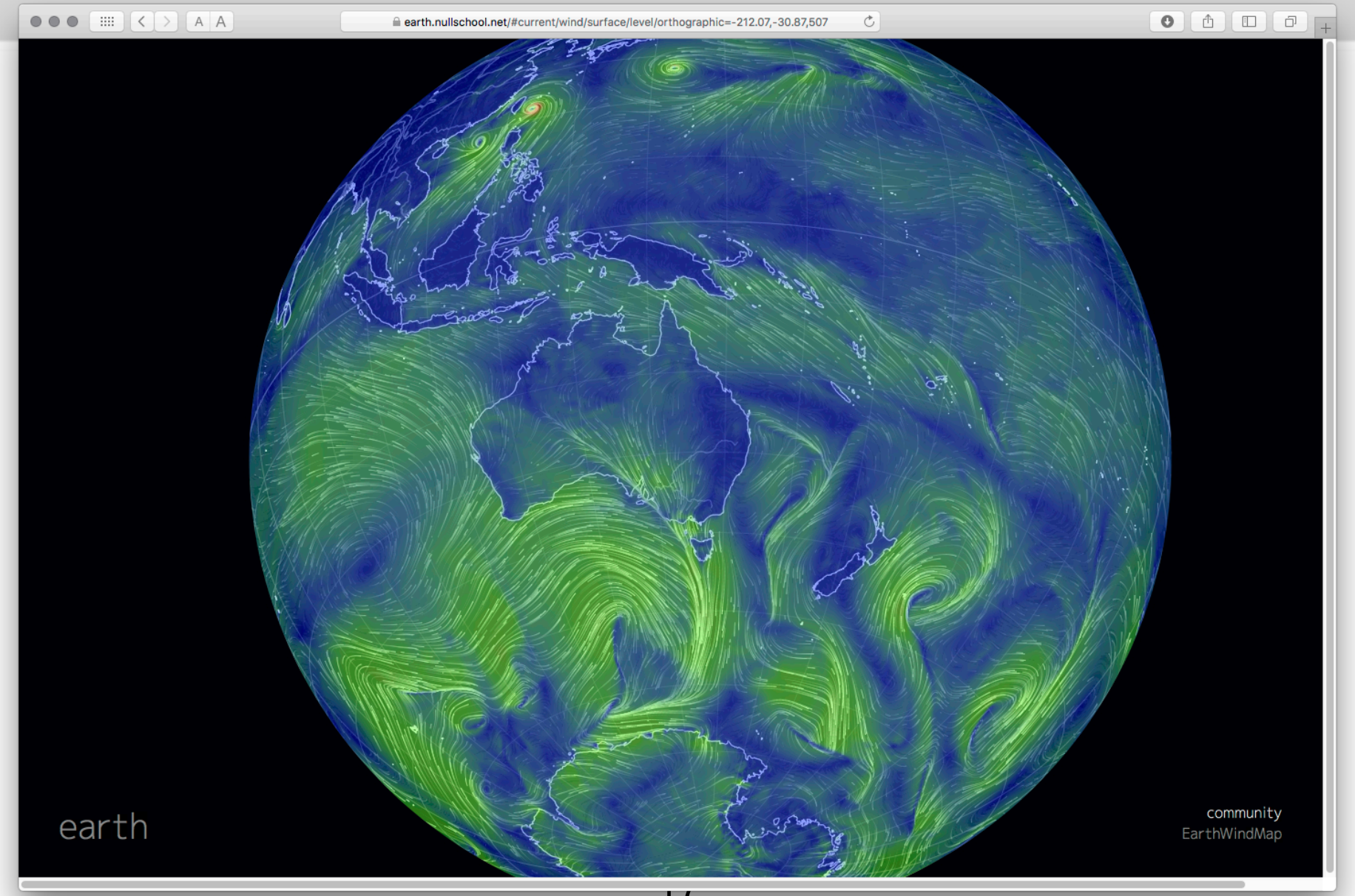
Dataset type? Attribute type?



Source: http://cartography.oregonstate.edu/tiles/PlanObliqueEurope/

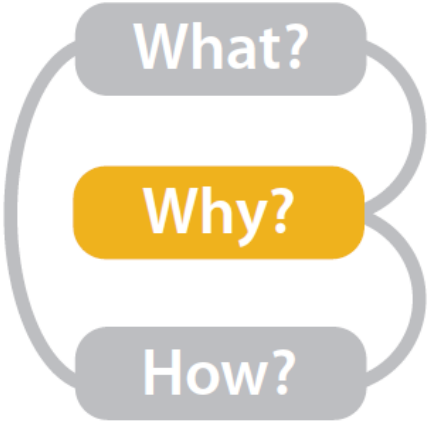
Dataset type? Attribute type?

MONASH University



Source: https://earth.nullschool.net

Why?



Textbook chapter 3

Why?

MONASH University

Actions

Analyze

- Consume
  - Discover
  - Present
  - Enjoy
- Produce
  - Annotate
  - Record
  - Derive

Search

	Target known	Target unknown
Location known	Lookup	Browse
Location unknown	Locate	Explore

Query

- Identify
- Compare
- Summarize

Targets

All Data

- Trends
- Outliers
- Features

Attributes

- One
  - Distribution
  - Extremes
- Many
  - Dependency
  - Correlation
  - Similarity

Network Data

- Topology
- Paths

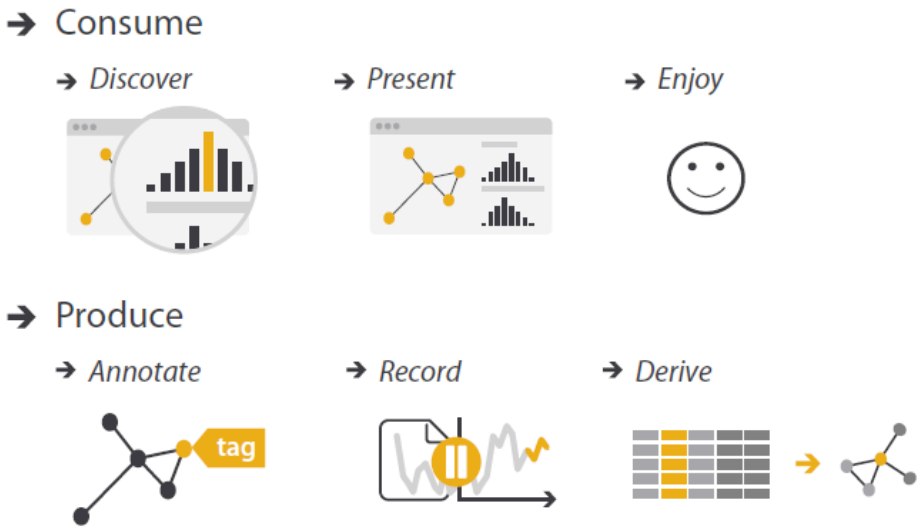
Spatial Data

- Shape

VAD pg. 42

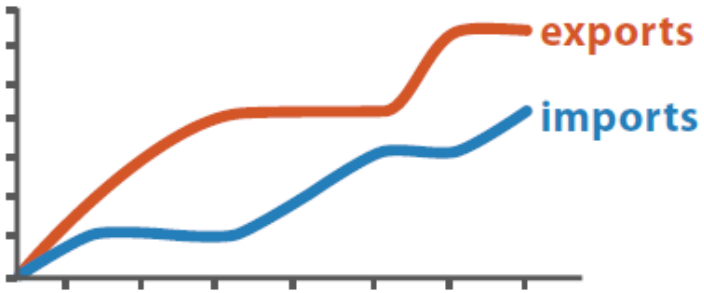
Why? Actions > Analyse

- Analyse > Consume
  - Discover vs. present
    AKA: explore vs. explain
  - Enjoy
    - AKA casual, social
    - AKA storytelling
- Analyse > Produce
  - Annotate, record
  - Derive/transform
    - Crucial design choice

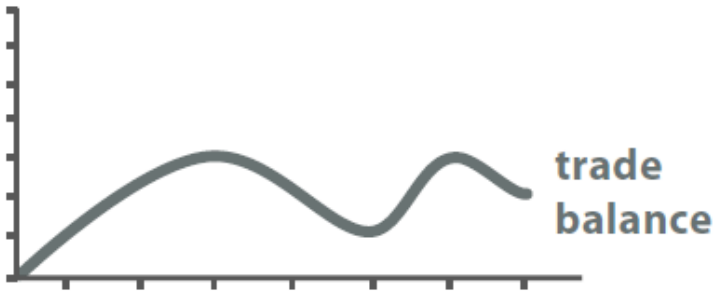


VAD pg. 46

- Don't just draw what you are given
  - Decide what is the right thing to show
  - Create it with a series of transformations from the original dataset
  - Draw that
- One of the main strategies for handling complexity



Original Data



trade balance = exports – imports

Derived Data

VAD pg. 52

MONASH University

- what does the user know?
  - target, location
- how much of the data matters?
  - one, some, all
- independent choices for each of these three levels
  - analyze, search, query
  - mix and match

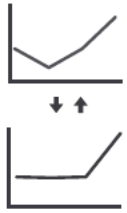
Search

	Target known	Target unknown
Location known	Lookup	Browse
Location unknown	Locate	Explore

Query

→ Identify

→ Compare



→ Summarize

VAD pg. 46

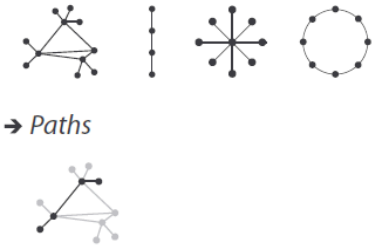
Why? Targets

Target: some aspect of the data that is of interest to the user.

- **All Data**
  - Trends
  - Outliers
  - Features

- **Attributes**
  - One
    - *Distribution*
    - *Extremes*
  - Many
    - *Dependency*
    - *Correlation*
    - *Similarity*

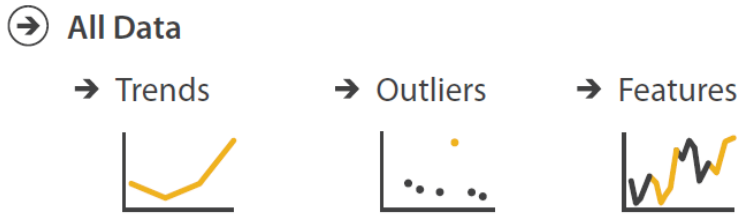
- **Network Data**
  - Topology
    - *Paths*



- **Spatial Data**
  - Shape

VAD pg. 56

Why? Targets



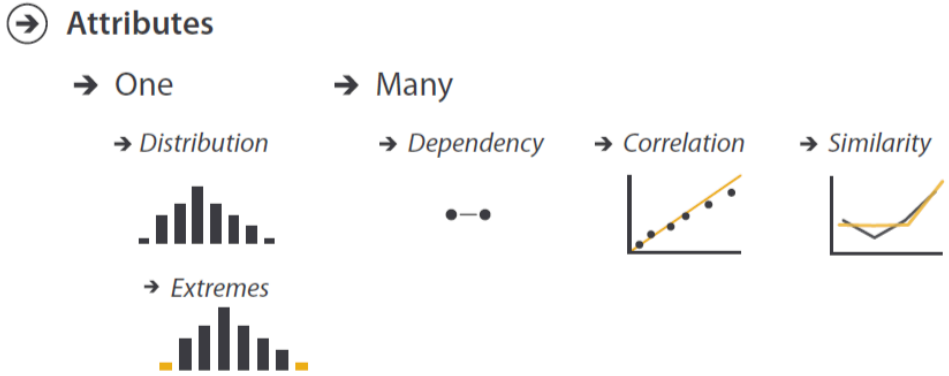
Trend: increases, decreases, peaks, etc.

Outlier: data not aligning with trend

Feature: task dependent

VAD pg. 56

Why? Targets



Attribute = property that can be visually encoded.

Dependency: values of first attribute directly depend on second attribute.

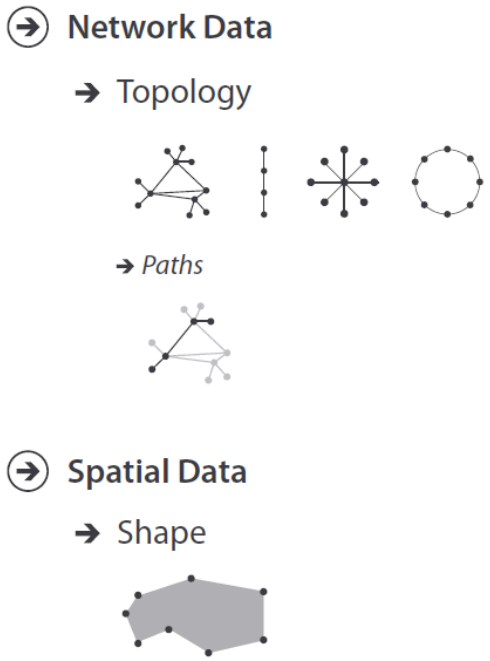
Correlation: tendency of first attribute to be tied to second attribute.

Similarity: quantitative measure of all values of two attributes.

VAD pg. 56

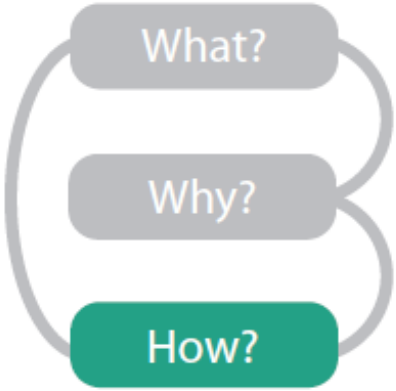
Why? Targets

Topology examples: adjacency of two nodes, shortest path between two nodes, number of connected edges, etc.



VAD pg. 56

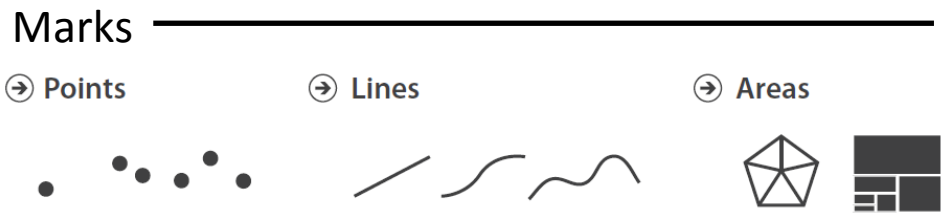
How?



VAD pg. 58

How? Marks and Channels

- Marks
  - Geometric primitives
- Channels
  - Control appearance of marks
  - Can redundantly code with multiple channels



?

Group work:
How can we vary the appearance of marks to encode quantitative and qualitative data? Compile a list of channels.

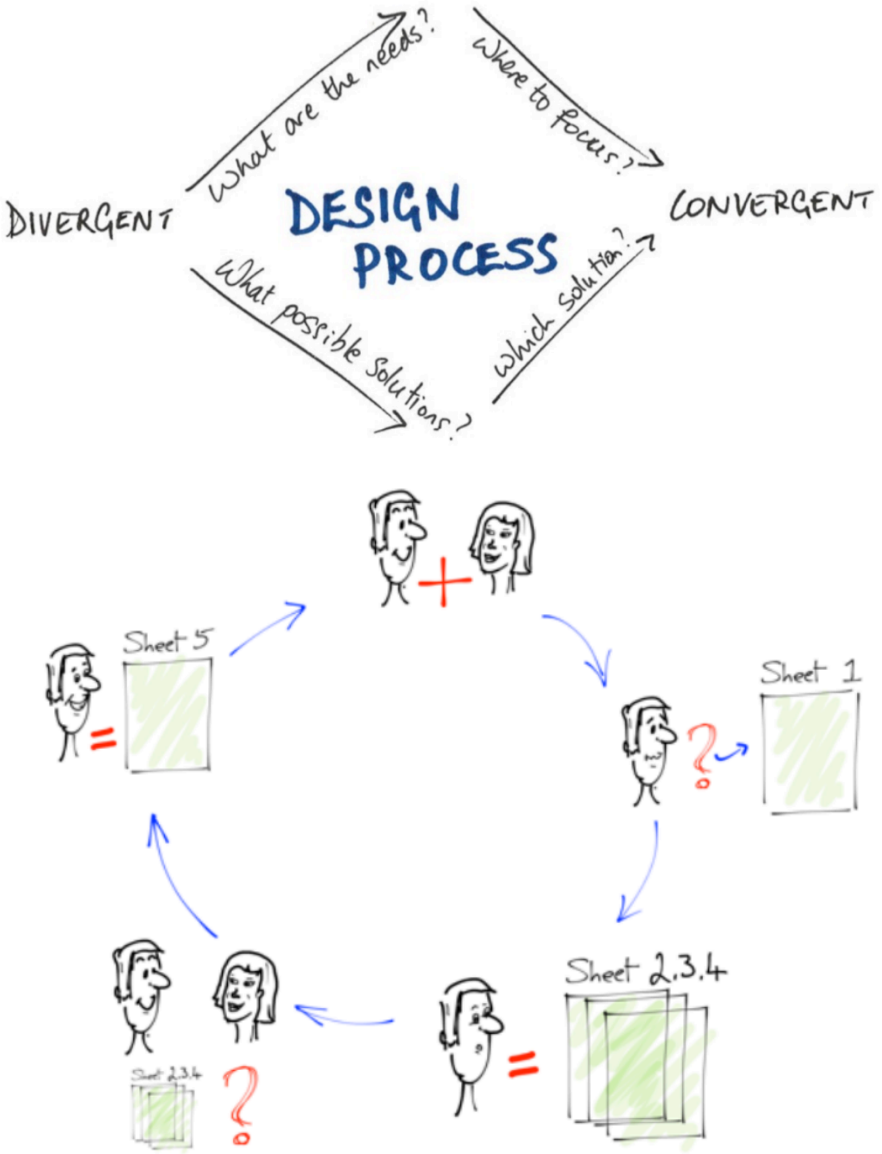
VAD pg. 96

MONASH University

Five Design Sheet Methodology

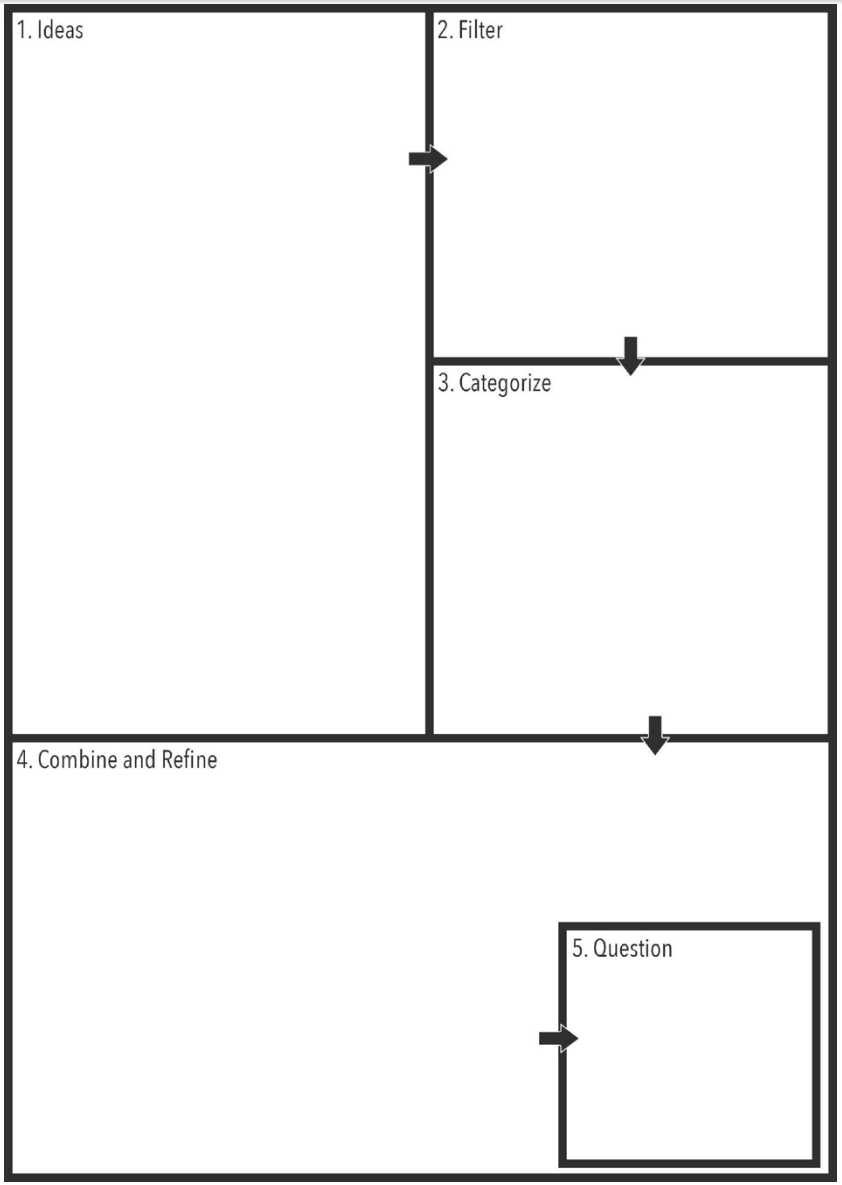
Five Design Sheet Methodology Goals

- The FDS methodology helps us structure our approach to ideation.
- It encourages us to freely, yet systematically explore a design space before converging on a solution.



Five Design Sheet Methodology
Sheet 1 - Brainstorm

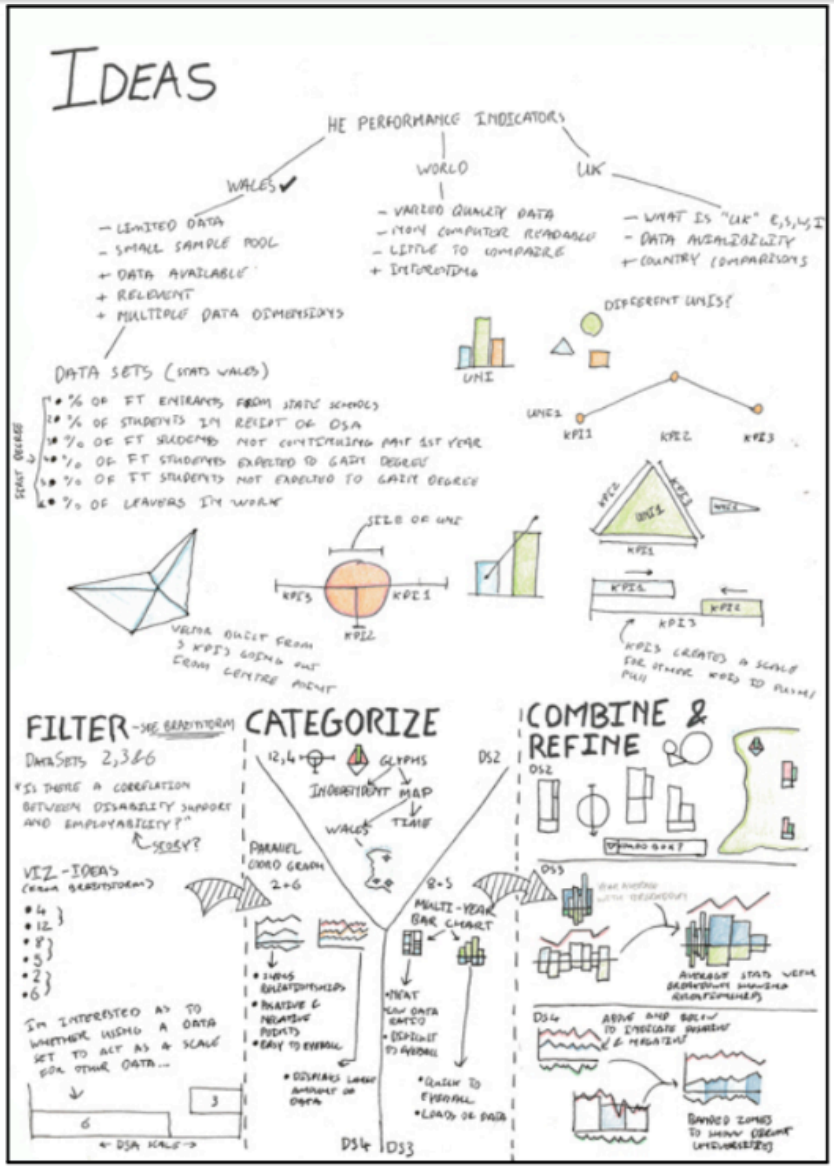
1. Sketch and draw as many ideas as you can think of
2. Remove duplication
3. Group similar ideas: sticky notes on a wall are helpful for this
4. Mini-ideas → bigger solutions… perhaps one system with multiple views?
5. Does this solution satisfy the *Why*?



Five Design Sheet Methodology
Sheet 1 - Brainstorm

MONASH University

1. Sketch and draw as many ideas as you can think of
2. Remove duplication
3. Group similar ideas: sticky notes on a wall are helpful for this
4. Mini-ideas → bigger solutions… perhaps one system with multiple views?
5. Does this solution satisfy the *Why*?



Five Design Sheet Methodology
Sheets 2-4 - Brainstorm

MONASH University

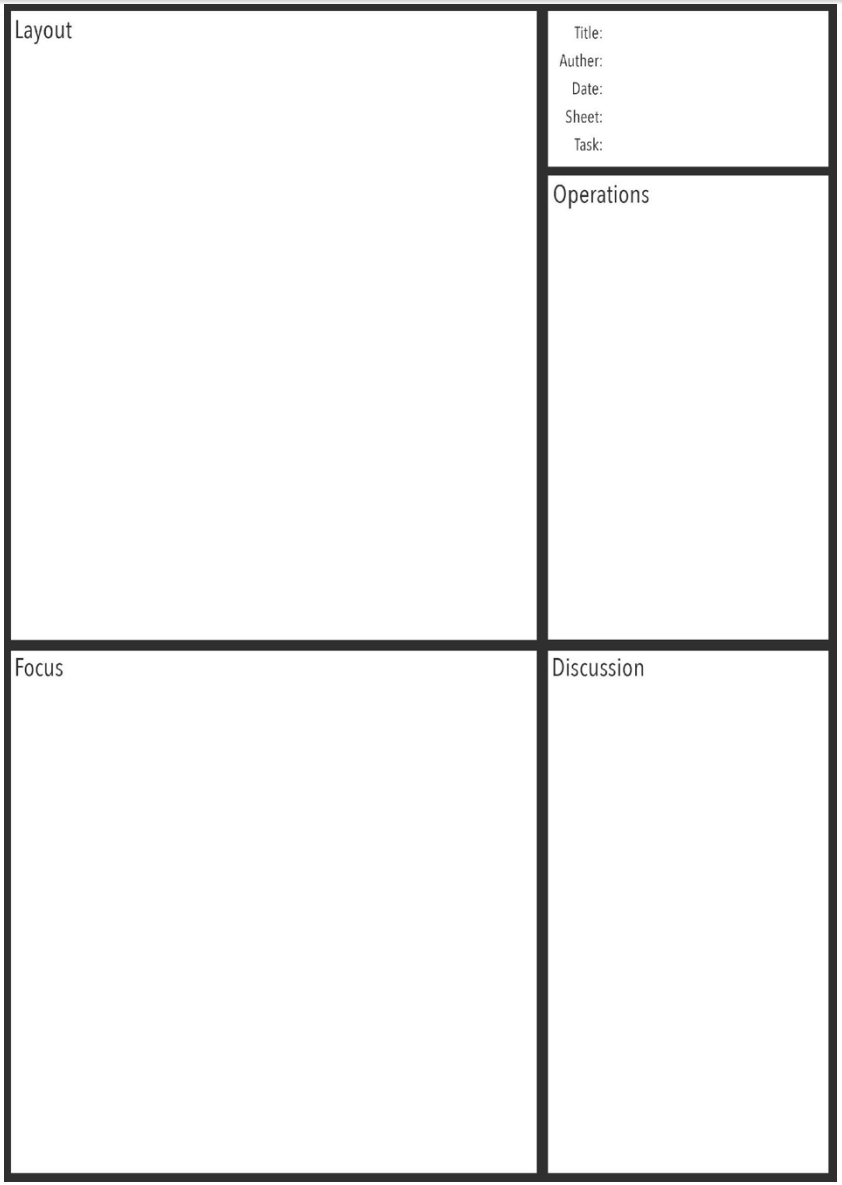
Layout: e.g. sketched screen-shot

Focus: explanations of key/ novel visualisation techniques

Operations: details of key interactions (e.g. a statechart)

Discussion: focus on advantages and disadvantages of the design

Meta-information: title/ author, date, sheet-number, task



Five Design Sheet Methodology
Sheets 2-4 - Brainstorm

MONASH University

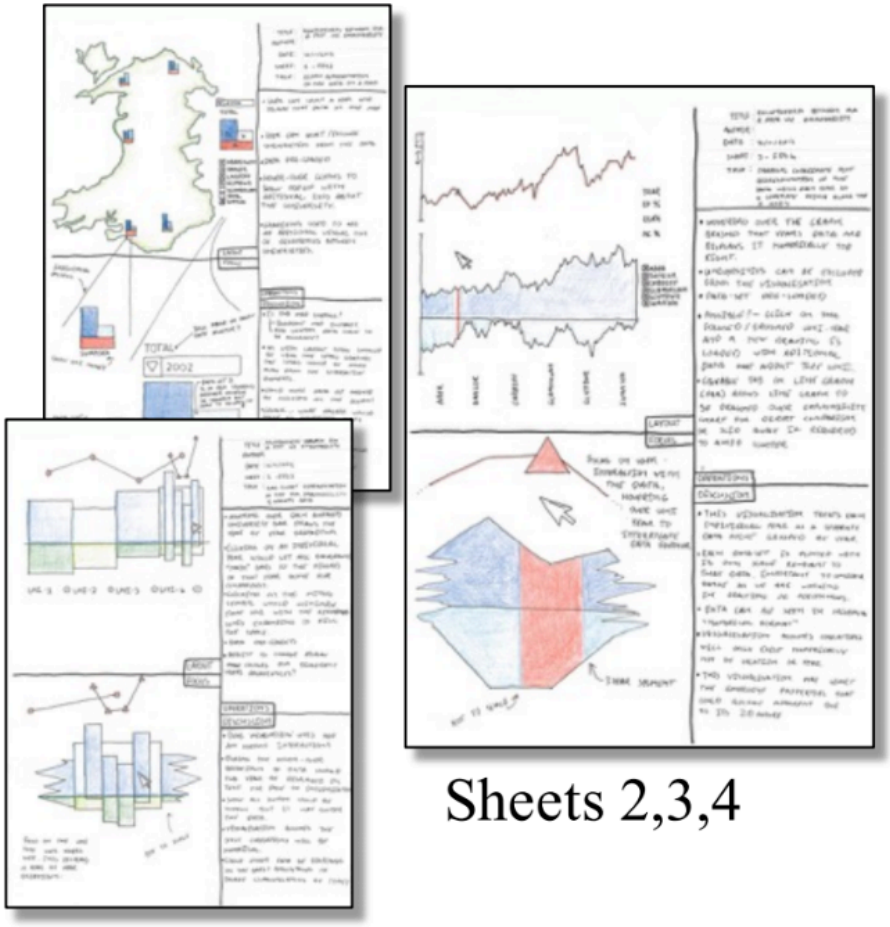
Layout: e.g. sketched screen-shot

Focus: explanations of key/ novel visualisation techniques

Operations: details of key interactions (e.g. a statechart)

Discussion: focus on advantages and disadvantages of the design

Meta-information: title/ author, date, sheet-number, task

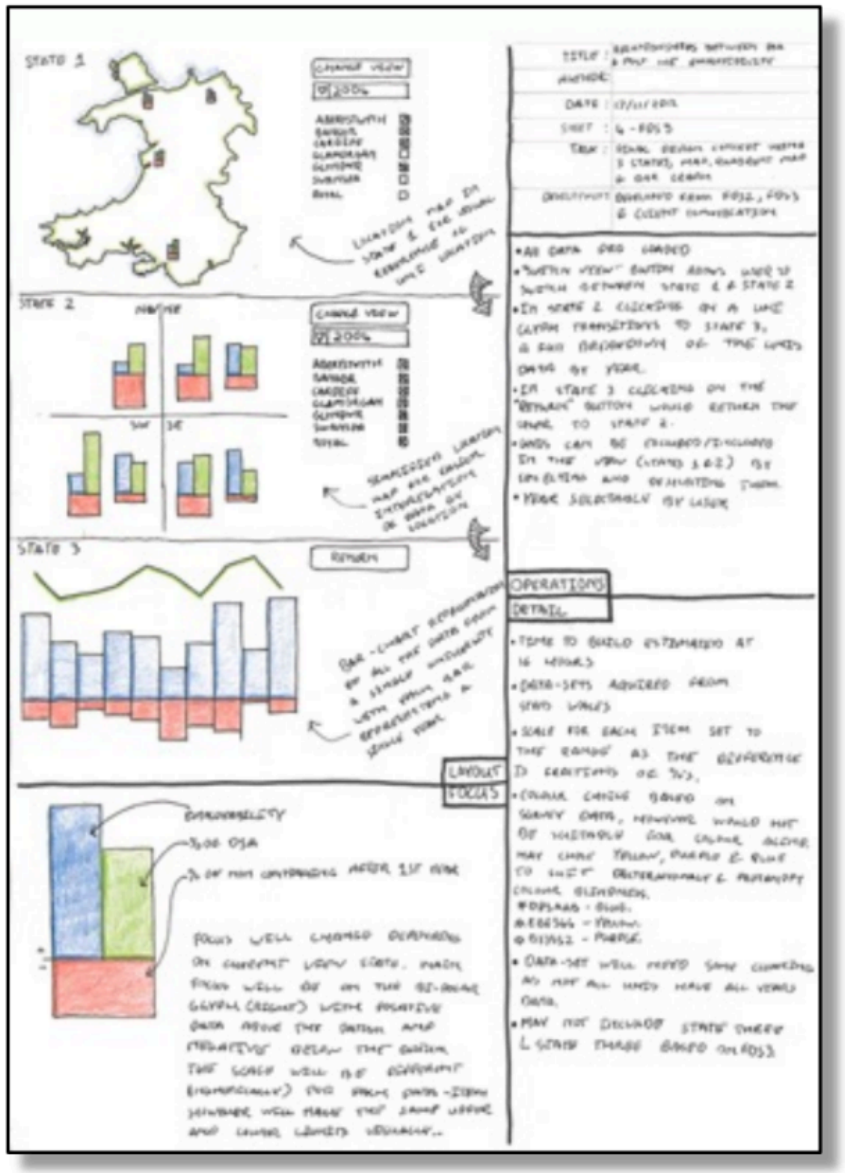


Sheets 2,3,4

Five Design Sheet Methodology
Sheet 5 - Realisation

Take the best of the previous designs and explore in greater detail. Focus on:

1. Description of algorithms / techniques
2. Dependencies: e.g. software libraries, compatibility, etc.
3. Estimate time and effort require to build the solution
4. Specific requirements of materials, hardware (desktop, tablet, phone, etc.)



MONASH University

Task for next Week!

MONASH University

Think about the **who**, **what** and **why** of your semester project.

- **Who** are your target users?
- **What** are you showing?
- **Why** is it useful?

Come with a project idea (and if possible data) to the lab in week 3. We will apply the Five Design Sheet Methodology to explore the **how** of your project.

MONASH University

Presentations next Week!

Presentation Theme

Based on the Allocate+ tutorial class list, the students in the CLAYTON Friday 10AM–12PM lab will be presenting next week.

All students: Please read the instructions on the weekly forum and post your viz. research and analysis there.

- Find a visualisation in a domain of your choice
- Discuss in a few words each the What / Why / How of your chosen Vis following Munzner's framework

Further reading

MONASH University

- Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, Nov 2014.
  - *Chap 1-5: Marks and Channels*
- *On the Theory of Scales of Measurement.* Stevens. Science 103:2684 (1946), 677–680.
- Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects. Stevens. Wiley, 1975.
- *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods.* Cleveland and McGill. Journ. American Statistical Association 79:387 (1984), 531–554.
- *Perception in Vision*. Healey. http://www.csc.ncsu.edu/faculty/healey/PP
- Visual Thinking for Design. Ware. Morgan Kaufmann, 2008.
- Information Visualization: Perception for Design, 3rd edition. Ware. Morgan Kaufmann /Academic Press, 2004.