

Bachelor in Politics And International Studies  
Bachelor in Sociology

## Digital Data, Data Collection (QS102)

### Week 3

The role of data in the social sciences  
before and since the digital times

+

Surveys in social and political research

### Summary of week 1

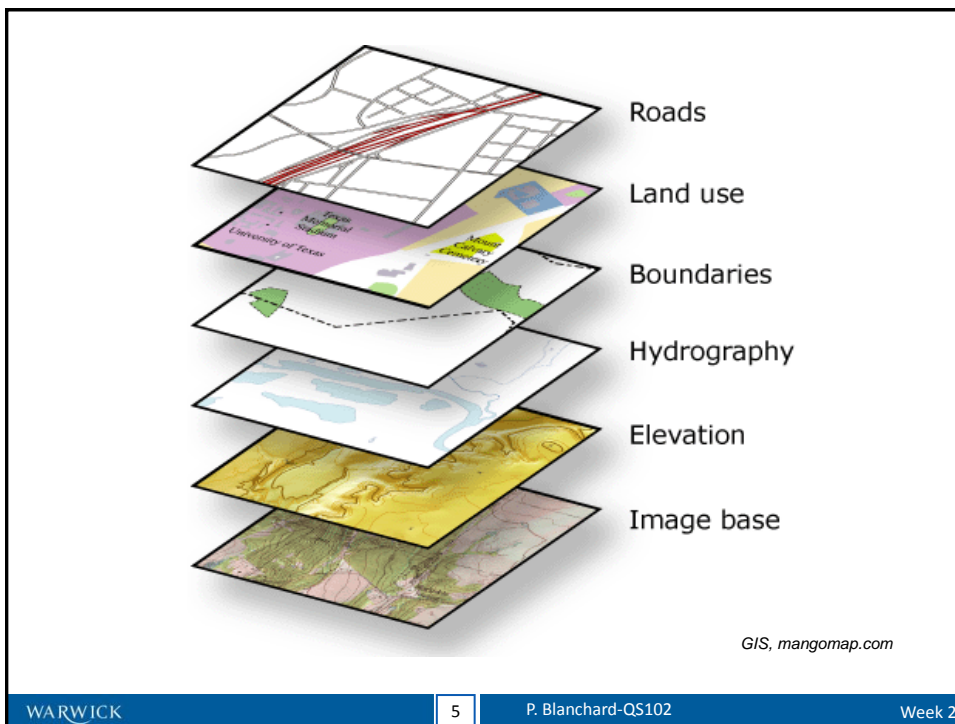
- Data are rising, including in the social sciences
- Data are part of a philosophical and epistemological debate



"The School of Athens", Raffaello Sanzio da Urbino 1510

## Summary of week 1

- Data are rising, including in the social sciences
- Data are part of a philosophical and epistemological debate
- There are hardly any pure/raw data, rather layers of transformation



### Summary of week 1

- Data are rising, including in the social sciences
- Data are part of a philosophical and epistemological debate
- There are hardly any pure/raw data, rather layers of transformation
- Data can be of many kinds and formats, but need standardisation

### Dataset definition

The diagram illustrates the structure of a dataset as a grid. A blue box at the top left is labeled 'dataset (table)'. A horizontal bar above the grid is labeled 'column properties (checks, limitation etc)'. A vertical bar on the left side of the grid is labeled 'dataset properties (some table properties)'. A horizontal bar on the right side of the grid is labeled 'columns definition'. A single horizontal row in the grid is labeled 'dataset row'. A single cell within the grid is labeled 'dataset value'. The source 'abantecart.com' is noted at the bottom right of the diagram.

abantecart.com

WARWICK
7
P. Blanchard-QS102
Week 2

The screenshot shows the IBM SPSS Statistics Data Editor interface. The main window displays a data table with the following columns: Serial, Region, Sex, Age, MarStat6, MarStat7, SupParty, ClosePty, and PartyIDN. The data is sorted by Serial number. The 'Sex' column has a yellow highlight on the value 'Male' in row 2. The status bar at the bottom indicates 'IBM SPSS Statistics Processor is ready' and 'Unicode: ON'. The title bar shows the file name '\*BSA2013.sav [DataSet1] - IBM SPSS Statistics Data Editor'.

Serial	Region	Sex	Age	MarStat6	MarStat7	SupParty	ClosePty	PartyIDN
1	England	Male	66	Married	In a couple	No	No	None
2	England	Male	41	Single (never ...	Not in a couple	No	No	None
3	England	Male	42	Married	In a couple	No	Yes	Labour
4	England	Female	35	Separated (aft...	Not in a couple	No	No	None
5	England	Female	61	Married	In a couple	No	Yes	Conservative
6	England	Male	24	Single (never ...	Not in a couple	No	No	Don't know
7	England	Male	75	Married	In a couple	No	Yes	Labour
8	England	Female	25	Single (never ...	Not in a couple	No	Yes	Liberal Dem...
9	England	Male	63	Married	In a couple	No	No	Labour
10	England	Male	69	Married	In a couple	Yes	Not applicable	Conservative
11	Scotland	Male	75	Married	In a couple	Yes	Not applicable	Conservative
12	England	Female	61	Married	In a couple	No	No	Labour
13	England	Female	34	Married	In a couple	No	Yes	Labour
14	Wales	Female	45	Divorced/diss...	Not in a couple	No	No	None
15	England	Female	56	Single (never ...	Not in a couple	Yes	Not applicable	Green Party
16	England	Male	54	Married	In a couple	Yes	Not applicable	Conservative
17	England	Female	65	Widowed/survi...	Not in a couple	Yes	Not applicable	Labour
18	England	Male	44	Married	In a couple	No	No	Don't know
19	England	Female	22	Single (never ...	Not in a couple	No	No	None
20	England	Female	61	Married	In a couple	Yes	Not applicable	Conservative
21	England	Male	67	Married	In a couple	No	No	Conservative
22	England	Male	54	Single (never ...	Not in a couple	Yes	Not applicable	Liberal Dem...
23	England	Female	40	Married	In a couple	No	No	Conservative
24	England	Female	69	Married	In a couple	No	Yes	Liberal Dem...

British Social Attitudes 2013

WARWICK
8
P. Blanchard-QS102
Week 2

### What about data for social and political sciences?

- Data can be of any kind: words, raw numbers, percentages, images, hyperlinks...
  - Data can describe many objects: events, opinions, ideological positions, attitudes, statuses, actions, relationships, etc.
  - Data are related to a unit of analysis: individuals, groups, regions, countries, organisations, families, events, actions, records, products, etc.
  - Most common: individuals x variables
- Data are **formatted** (standardised) in order to be used, reused, shared, replicated, combined, merged, compared, etc.

### What about data for social and political sciences?

- Data are **formatted** (standardised) in order to be used, reused, shared, replicated, combined, merged, compared.
- Formatting, standardisation, reuse, comparisons, etc. require **infrastructures**: data services, data software, data specialists, data servers, data conferences, data networks, data providers...
- Data norms, data procedures, data skills.

## Simple examples of social and political sciences data

### Counts

Victims of police violence

<http://www.innocentdown.org/>

[Regressin.deadspin.orgg.deadspin.org](http://Regressin.deadspin.orgg.deadspin.org)

[Fatal encounters](#)

Migrants to Europe

[The migrants files](#)

### Automated measures

[What the birth of a son can change to your life...](#) ("Quantified self")

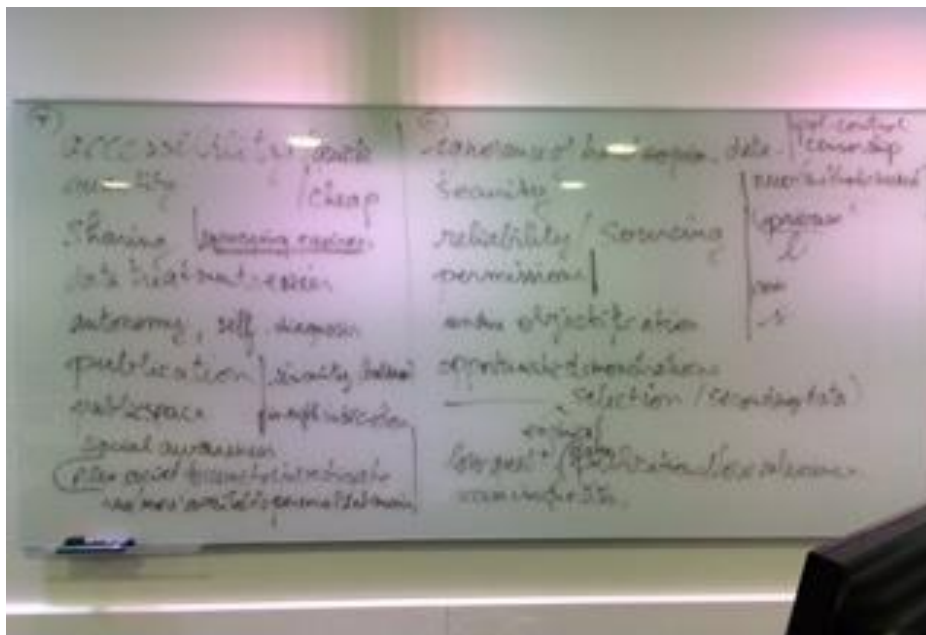
## What are good data (for social and political sciences)?

Criteria:

- Availability (technical, political, financial...)
- Relevance
- Cost
- Reliability
- Validity
- Richness
- Usability

→ **Data quality is variable. Seldom an easy work.**

## What does Internet change to the role of data?



## What does Internet change to the role of data?

- Opportunities
  - Standard medium, permanent, everywhere, mainly free, most people connected
  - Standardised and rationalised information
  - End of organisational gate-keeping
  - New activities (transnational and trans-class interactions, self-exhibition...), new social settings, new borders between private and public...
- Challenges
  - Profusion ('big data') → 'data bubble' (Golder and Macy in Gitelman (ed) 2013, chap. 7)
  - Distortions and fakes
  - Overlaps and contradictions
  - Sourcing omitted or wrong, theft
  - Quality
  - Prenotions, routines, seduction, conditioning, manipulation
  - Illusion of knowledge (see Kitchin 2014)

## What does Internet change to the role of data?

- Opportunities
- Challenges
- Internet is not so ideal
  - Without borders and beyond domestic cultures?
  - Anonymous?
  - A public space?
  - Exempt from geopolitical and geoeconomic constraints?
  - Free space for expression?
  - Easy technique?
  - Homogeneous space?

→ **A bundle of political and technical limitations!**

## What kinds of data on the Internet?

- Traces *or* experiments
  - Archives, automatic records, blogs and other webpages  
*versus* interviews, emails, experiments, participatory observation
- At individual *or* collective level
  - CVs, smartphones, blogs  
*versus* forums, social media, administrations
- Small *or* big
  - Targetted survey, family blog  
*versus* FaceBook, health insurances/administrations/laboratories, Amazon

→ **Already as diverse as pre-Internet data.**



## How do big data challenge social sciences?

- Kitchin 2014
  - Why is data mining opposed to theorising? Do you think the 'new empiricism' can work in the social sciences?
  - What are the 'digital humanities'? Can you give examples of 'big data' in history, archeology or literature, or another social science or humanities?