

# Week 1: Measurement - What, Why and How

POLS0013 Measurement in Data Science

---

Dr. Julia de Romémont

Academic Year 24-25

UCL Department of Political Science

Introduction and Course Outline

Motivating examples

Measuring social science concepts

Key concepts and distinctions

Perils of Social Science Measurement

## Introduction and Course Outline

---

- ▶ Originally designed for third year students in the Social Data Science degree programmes at UCL
  - Previous courses: POLS0008 & POLS0010
- ▶ Now also available as an elective in the BSc Politics and International Relations
  - Previous courses: POLS0060 & POLS0083
- ▶ Builds on materials covered in Social Data Science stream courses. In particular, it assumes you have (or are willing to gain) a good working knowledge of:
  - Linear regression
  - Sampling and statistical inference
  - R and Rstudio
  - **Regression models for binary and categorical outcomes**

The goal of quantitative empirical research is to **learn about how the world works from data.**

= Inference

To this end, we aim to **uncover patterns and regularities** (i.e. variation) in the data and

- ▶ Generalise them (**Population Inference**)
- ▶ Explain them (**Causal Inference**)
- ▶ Characterise them (**Measurement Inference**)

This course will provide you with:

- ▶ A framework to critically reflect on the connection between data and social science concepts that we may be interested in
- ▶ An overview of the different methods commonly used to link concepts and data and their intuitions
- ▶ The practical skills required to implement these methods using R

1. Measurement: What, Why and How
2. Measurement Theory and Error
3. Deriving Scales from Theory
4. Measuring Scales from Comparison Data
5. Supervised Scale Measurement I: Regression

## *Reading Week*

6. Supervised Scale Measurement II: Linear Indices
7. Supervised Class Measurement
8. Unsupervised Scale Measurement I: Interval-Level Indicators
9. Unsupervised Scale Measurement II: Categorical Indicators
10. Unsupervised Class Measurement

## Dr. Julia de Romémont

- ▶ E-mail: [j.romemont@ucl.ac.uk](mailto:j.romemont@ucl.ac.uk)
- ▶ Office Hours: Tuesdays 11.30am-12.30pm & Wednesdays 11.30am-12.30pm,  
Book [here](#)

Book office hours via MS Bookings (link on Moodle).



- ▶ Prof. Ben Lauderdale is writing a textbook on the content of this module, *Pragmatic Social Measurement*
  - Provides most of the required readings for this module
  - Available to download on the course website
- ▶ The course website and the syllabus on Moodle lists further readings for each week

## 1. Essay (50%)

- Deadline: 10<sup>th</sup> December 2024
- Involves finding and critiquing a measure of your choice
- More and more detailed info [here](#)
- You can (and should) start thinking about this now!

## 2. Coursework (50%)

- Deadline: 14<sup>th</sup> January 2025
- Involves conducting analyses with a provided data set and answering a series of questions in a style similar to the seminar tasks
- Questions will be released at the end of the term

## Course Website

### Lectures

- ▶ Tuesdays 9-11am
- ▶ Recording will be made available the day *after*

### Seminars

- ▶ Weekly one-hour practical session on lecture topic
- ▶ Tuesdays 2pm, 3pm, 4pm
- ▶ Solutions will be released Thursdays

This is not a course where you will do well simply by memorising some ideas and/or procedures

- ▶ You need to be able to figure out which procedures are appropriate
- ▶ You need to be able to communicate what you are doing and why
- ▶ You need to learn how to do some new things in R, but also apply things you already know to new problems

1. Complete required reading
2. Attend the lecture
3. Attempt the seminar assignments<sup>1</sup>
4. Attend the seminar
5. Go back through the seminar assignment and complete it by using the provided solutions
6. Make note of any unanswered questions, ask them in the seminars, lectures or SSF
7. Regularly go back through slides, assignments and readings of earlier weeks, and, *help each other!*

---

<sup>1</sup>At the very least, read the prompt and load the data into R

- ▶ I am happy to answer any questions
  - But I expect you to be up to date with the material, and read instructions and informational material diligently
  - I expect you to first genuinely try to find the answer to your question yourself
  - I expect you to know what your question is when you ask it!
- ▶ You can ask me questions in person (lectures, seminars, SSF) and via the Moodle forum
  - Only contact me via email if
    - ▶ I asked you to **OR**
    - ▶ Your question is urgent **AND**
    - ▶ Your question can be answered with two sentences or less

ChatGPT and many other GenAI tools (Large Language Models (LLMs)) *can* be useful to

- ▶ get a different explanation for some concept/method covered in the course than me
- ▶ get assistance for coding in the seminar tasks, especially for de-bugging
- ▶ get help for formatting (e.g. figures and tables) and writing clarity

But I would like to encourage you to think carefully about **how** and **how often** you rely on such tools, for...

## 1. Ethical reasons

- There are huge environmental costs of training and running these models
- CO2 emissions of any single ChatGPT query are (60x) higher than a Google search
- We should think carefully about who we empower by using these tools

## 2. Educational reasons

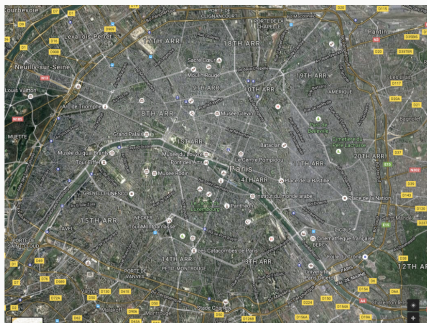
- Learning involves going through the same things over and over again, and using ChatGPT etc as a 'shortcut' is going to undermine this
- At UCL you have the chance to learn via human interaction
- Relying on these tools too much or *instead* of the course materials may be confusing



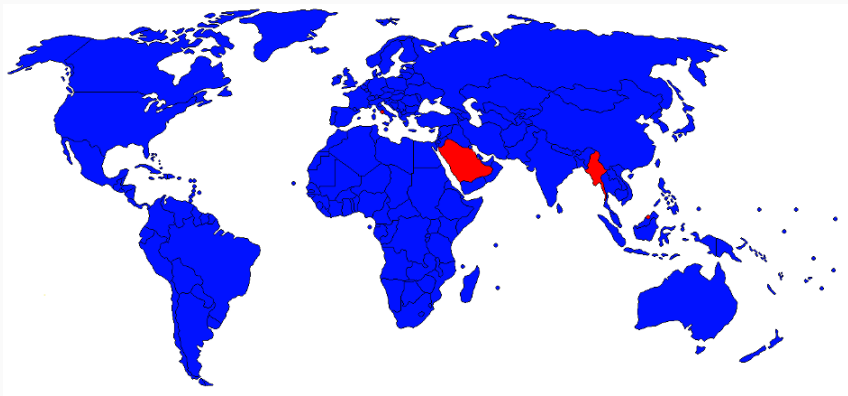
## Motivating examples

---

# Which is more useful?



## Misleading or necessary simplification?



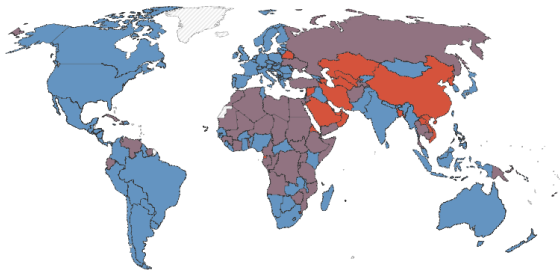
- ▶ What does this measure?
- ▶ What do we learn?
- ▶ What is 'filtered' out?

# Misleading or necessary simplification?

## Political regime, 2018

Based on the classification and assessment by Polity 5 (2021).

Our World  
in Data



No data    Autocracy    Anocracy    Democracy

Source: Polity 5 (2021)

Note: The Chart tab uses numeric values, ranging from 2 for democracies to 0 for autocracies.

OurWorldInData.org/democracy • CC BY

- ▶ What does this measure?
- ▶ What do we learn?
- ▶ What is 'filtered' out?

# Multiple personalities

"It's so incredible to finally be understood."

Only 10 minutes to get a "freakishly accurate" description of who you are and why you do things the way you do.

Take the Test →

194K+ Tests taken today

18M+ Tests taken in France

685M+ Total tests taken

91.2% Results rated as accurate or very accurate

What is being measured? Psychological concept: *human personality*

- ▶ People behave in different ways, when faced with similar situations
- ▶ Are there useful “types” with which we can describe this?

- ▶ One *conceptualisation* of the *concept* of human personality is the Myers–Briggs Personality Type.

Four dimensions, each of which contrasts two ideal types.

- ▶ (I)ntroversion vs (E)xtroversion
- ▶ (S)ensing vs i(N)tuition
- ▶ (T)hinking vs (F)eeling
- ▶ (J)udging vs (P)erceiving

Measurement is usually conducted via self-assessment using a series of agree-disagree test items:

- ▶ At social events, you rarely try to introduce yourself to new people and mostly talk to the ones you already know.
- ▶ Even a small mistake can cause you to doubt your overall abilities and knowledge.
- ▶ You enjoy participating in group activities.
- ▶ Your emotions control you more than you control them.
- ▶ etc etc



## “LOGISTICIAN”

ISTJ (-A/-T)

Practical and fact-minded individuals, whose reliability cannot be doubted.



## “PROTAGONIST”

ENFJ (-A/-T)

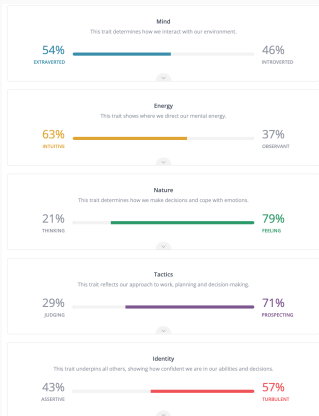
Charismatic and inspiring leaders, able to mesmerize their listeners.



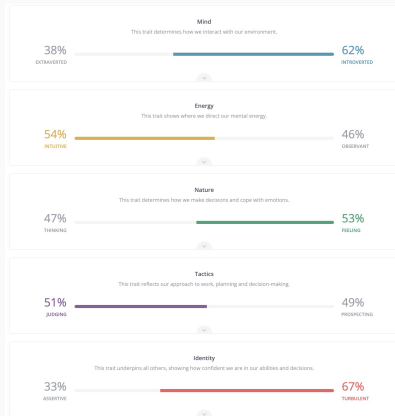
## Taking personality tests is fun(ish)

- ▶ If you are looking for some light entertainment, go take the personality quiz at <https://www.16personalities.com>
- ▶ I came out as “INFP-T” (Intraversion- Intuition- Feeling- Perceiving + Turbulent) when I last took this test and as “ENFP-T” (Extraversion- Intuition- Feeling- Perceiving + Turbulent) the time before
- ▶ This apparently makes me either a “Turbulent Advocate” or a “Turbulent Campaigner”) (“enthusiastic, creative and socialble free spirits, who can always find a reason to smile”)

# Now you see me?



October 2021



October 2022

“According to most personality type theories, the individual’s type is inborn and does not change.”

- ▶ Myers–Briggs tests are used in career counselling and hiring. (Pittenger 1993)
  - But not viewed as a serious measure of personality by most academic psychologists.
- ▶ Personality is a widely studied concept in the field of psychology.
  - See, for example, the “five factor” or “Big Five” model.
- ▶ The idea that humans think and reason and feel and act in ways that are *patterned* is central to this field.
  - Some of those patterns tend to be universal; others vary across different people.
  - Personality types are an attempt to measure some of this variation.

**GHS INDEX**  
GLOBAL HEALTH  
SECURITY INDEX

Index About Report & Model Resources for Action Data Stories News

Show Menu +

## Welcome to the 2019 Global Health Security Index

The GHS Index is the first comprehensive assessment of global health security capabilities in 195 countries. Read more [about](#) the Index and the [international panel](#) that helped develop it, download the [report](#) and [data model](#), and watch our introductory [video](#).

Get Started ▶

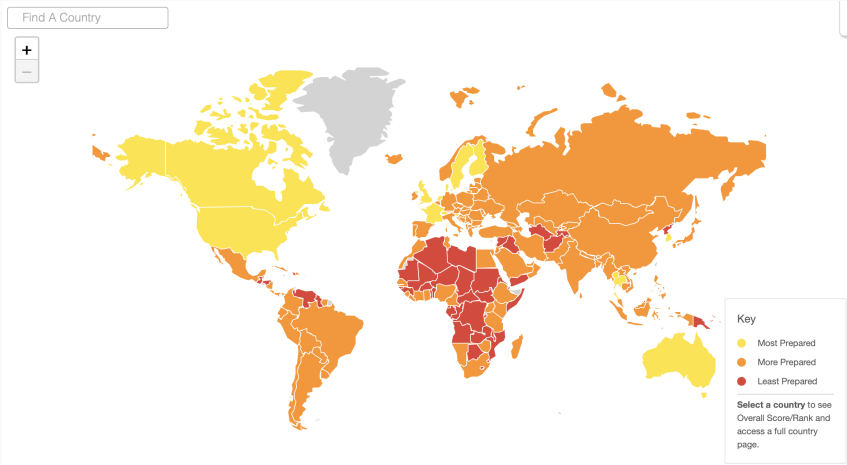
195  
Countries

## Goals

*The Global Health Security (GHS) Index is the first comprehensive assessment and benchmarking of health security and related capabilities across the 195 countries that make up the States Parties to the International Health Regulations (IHR [2005]). [...] the GHS Index will spur measurable changes in national health security and improve international capability to address one of the world's most omnipresent risks: infectious disease outbreaks that can lead to international epidemics and pandemics.*


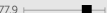

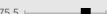

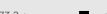

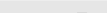
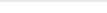

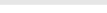
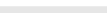
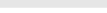


## Development

*To create the GHS Index, NTI, JHU, and the EIU project team—with generous grants from the Open Philanthropy Project, the Bill & Melinda Gates Foundation, and the Robertson Foundation—worked with an international advisory panel of 21 experts from 13 countries to create a detailed and comprehensive framework of 140 questions, organized across 6 categories, 34 indicators, and 85 subindicators to assess a country's capability to prevent and mitigate epidemics and pandemics.*



Source: GHS Index News (2020)

# 2019 GHS Top 15

| Rank | Country                        | Index Score ▼  | Region            | Population | Income              |
|------|--------------------------------|--|-------------------|------------|---------------------|
| 1    | <a href="#">United States</a>  | 83.5  | Northern America  | 100m+      | High income         |
| 2    | <a href="#">United Kingdom</a> | 77.9  | Europe            | 50-100m    | High income         |
| 3    | <a href="#">Netherlands</a>    | 75.6  | Europe            | 10-50m     | High income         |
| 4    | <a href="#">Australia</a>      | 75.5  | Oceania           | 10-50m     | High income         |
| 5    | <a href="#">Canada</a>         | 75.3  | Northern America  | 10-50m     | High income         |
| 6    | <a href="#">Thailand</a>       | 73.2  | Southeastern Asia | 50-100m    | Upper middle income |
| 7    | <a href="#">Sweden</a>         | 72.1  | Europe            | 1-10m      | High income         |
| 8    | <a href="#">Denmark</a>        | 70.4  | Europe            | 1-10m      | High income         |
| 9    | <a href="#">South Korea</a>    | 70.2  | Eastern Asia      | 50-100m    | High income         |
| 10   | <a href="#">Finland</a>        | 68.7  | Europe            | 1-10m      | High income         |
| 11   | <a href="#">France</a>         | 68.2  | Europe            | 50-100m    | High income         |
| 12   | <a href="#">Slovenia</a>       | 67.2  | Europe            | 1-10m      | High income         |
| 13   | <a href="#">Switzerland</a>    | 67.0  | Europe            | 1-10m      | High income         |
| 14   | <a href="#">Germany</a>        | 66.0  | Europe            | 50-100m    | High income         |
| 15   | <a href="#">Spain</a>          | 65.9  | Europe            | 10-50m     | High income         |

# A problem with the measure?

Amid the threat posed by COVID-19 the Global Health Security (GHS) Index has been used repeatedly as a reference for global preparedness during the pandemic. This is a key goal of the GHS Index, and the COVID-19 pandemic has become a proof point for its main finding: national health security is fundamentally weak around the world. No country is fully prepared for epidemics or pandemics, and every country has important gaps to address.

Given that overall finding, it is important to dispel misconceptions regarding the score of 83.5 (out of a possible 100) received by the United States. Although the United States received the top score of 195 countries assessed and was ranked number one, its score and rank do not indicate that the country is adequately prepared to respond to potentially catastrophic infectious disease outbreaks. Significant preparedness gaps remain, and some of those are playing out in the current crisis. The United States' response to the COVID-19 outbreak to date shows that capacity alone is insufficient if that capacity isn't fully leveraged. Strong health systems must be in place to serve all populations, and effective political leadership that instills confidence in the government's response is crucial.



## A problem with the measure?



The Atlantic: “This isn’t all Trump’s Fault (But he isn’t helping either)”

## Measuring social science concepts

---

- ▶ Psychology: Perception, Personality, Emotion
  - ▶ Sociology: Social Class, Mobility
  - ▶ Political Science: Ideology, Democracy
  - ▶ Economics: Productivity, Inflation
  - ▶ Geography: Distance, Composition, Distribution
- ⇒ These concepts—and many others like poverty and inequality and development—are used widely across the social sciences.

- ▶ These concepts are not “unproblematic” or “uncontested” in their definitions.
- ▶ Social scientists often disagree about the content of these core concepts.

## Is this because social science isn't really 'science'?

- ▶ Even Physical science concepts like length, mass, temperature, acidity, etc did not always have universally agreed definitions.
  - “Measurement: A Very Short Introduction” by David J Hand (2016)
  - “Inventing Temperature” by Hansok Chang (2004)

- ▶ Length, mass, etc are (more or less) uncontested (now) because
  1. They proved to be not just *useful* but *required* for scientists to communicate with one another.
  2. Scientists developed tools to measure them reliably in ways that could be understood and reproduced by other scientists.
- ▶ Sometimes what we think we know is wrong.
  - Einstein's development of special and general relativity in the early 20th century involved the reconceptualisation of length and mass.
  - And even Einstein and the properties of length and mass are still not uncontested!

*"In experiments conducted over the last 50 years, [the 2022 Nobel Prize winners] confirmed the reality of an effect that Albert Einstein had disdained as 'spooky action at a distance.' Measuring one of a widely separated pair of particles could instantaneously change the results of measuring the other particle, even if it was light-years away." (Source: NYT)*

- ⇒ Part of science **is** the process of conceptual clarification
- ▶ Critical engagement with concepts and measurement strategies is an important part of being a scientist.
  - ▶ This is particularly true (and crucial) in the social sciences
    - The conceptualisation and measurement of the things we are interested in are hard to separate from ideological, political and economic interests as well as normative and ethical considerations

- ▶ Measurement requires scientific understanding
  - A liquid-in-glass bulb thermometer relies on liquids expanding in volume when they get warmer.
  - What did we need to understand to make these work?
- ▶ Greater understanding and improved technology can both improve measurement
  - Infrared thermometers are much faster and more accurate.
  - What did we need to understand to make these work?

## Key concepts and distinctions

---



- ▶ This module is about the third of these:

## Population Inference

- = Inference from observed data to the data we would have measured if we had access to **a broader population of units**

## Causal Inference

- = Inference from observed data about the data we would have observed for **the same units given counterfactual circumstances**

## Measurement Inference

- = Inference from observed data to **unmeasured (latent) quantities describing the same units**

**Nominal:** Numeric values convey neither ordering nor distance

- Only allows tests of equality  $=$  and inequality  $\neq$

**Ordinal:** Numeric values convey ordering but not distance

- Also allows greater than  $>$  and less than  $<$

**Interval:** Numeric values convey ordering and distance

- Also allows addition  $+$  and subtraction  $-$

**Ratio:** Numeric values convey ordering, distance, and have a meaningful zero point

- Also allows multiplication  $\times$  and division  $/$

We will often use the terminology that

- ▶ measurements of nominal and ordinal quantities are *classes/ classifications* and
- ▶ measurements of interval and ratio quantities are *scales*.

## Representational vs. Pragmatic Measurement

- ▶ Cases where we think we are representing something that exists in the world and which causally shapes interactions
  - **Representational Measurement**
- ▶ Cases where we are creating a summary of things we observe in the world, inventing a concept to help us talk about patterns that we observe
  - **Pragmatic Measurement**

## Generative vs. Discriminative Measures

This is about the understanding behind the relationship between concept  $\mu$  and measure  $m$

$$\mu \leftarrow \overset{?}{-} - \rightarrow m$$

$\mu \rightarrow m$  Changes in the target concept *generate* changes in the measure through a direct causal pathway

→ **Generative Measures**

$\mu \leftarrow m$  Measure *discriminates* between different levels of the concept, implying that the concept is defined by the measurement strategy

→ **Discriminative Measures**

- ▶ Note that these sometimes yield the same numerical value, but our conceptual understanding of what we are entitled to claim about that number is different.

## Supervised vs. Unsupervised Measurement

- ▶ Cases where we use data and/our substantive knowledge of the relationship between observable data (indicators) and the concept we want to measure to relate the two
  - **Supervised Measurement**
- ▶ Cases where we use covariation in observable data (indicators) to try to discover relevant underlying concepts
  - **Unsupervised Measurement**
- ▶ Much more detail to come on all of these distinctions, in the textbook and in subsequent weeks of the module.

## Perils of Social Science Measurement

---

The concepts we have in our minds are often rich and unwieldy, typically measurement involves narrowing these towards more *minimalist* conceptualisations.

- ▶ A common criticism of quantitative social science measurements are that they are too narrow.
- ▶ More on this in Week 8

Quantitative measurement can create problems of fairness, or sometimes simply expose existing unfairness.

- ▶ About whether a measurement strategy systematically misrepresents the relative values of the underlying quantity that one aimed to measure when comparing different groups of units
- ▶ More on this in Week 2



## Problems of Unintended Consequences

- ▶ The use of quantitative measures can create incentives that lead to unintended consequences.

**Campbell's Law:** *"The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."*

**Goodhart's Law:** *"Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes."*

- ▶ This is not only the case for social science! Back to physics and Nobel prizes:

*"[Quantum mechanics'] centerpiece was the uncertainty principle, enunciated by Werner Heisenberg in 1927, which states that certain types of knowledge — of a particle's position and velocity, for example — are incompatible: The more precisely one property is measured, the blurrier and more uncertain the other becomes."*

(Source: NYT)

- ▶ Social measurements are used intentionally to shape behaviour
  - For example, your marks in this module and classification in your degree
- ▶ There is a long and depressing history of social measurement being used for oppression
  - However, if people have malign goals, they are likely to pursue those with or without the aid of quantitative measurement and the data analysis that it enables
  - The measurement is not really the problem in and of itself
- ▶ See the textbook for some discussion of UCL's history of developing social measurement for malign purposes.

# The stakes of social measurement are very high

- ▶ Social measurement shapes your lives in profound ways.
- ▶ It can be used to make peoples' lives better or worse.
- ▶ Bad measurement causes people to make bad decisions every day.

## Next week: More on Measurement Error

- ▶ Definition of measurement error
- ▶ Fairness in measurement
- ▶ Consequences of measurement error for subsequent analyses

- Chang, Hasok. 2004. *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press.
- Hand, David J. 2016. *Measurement: A Very Short Introduction*. Oxford University Press.
- Pittenger, David J. 1993. "Measuring the MBTI... And Coming up Short." *Journal of Career Planning & Placement* 54 (1): 48–52.