

Introduction  
ooooooooooooooo

Example  
ooo

Course Specifics  
oooo

Wrapping Up  
o

# Data Analysis in R

## Introduction: Data Analysis and Social Science

Ken Stiller

16th January 2026

# Data Analysis

*“When my information changes, I alter my conclusions.”*

- Paul Samuelson

# Why Analyse Data ?

- ▶ To know things - which is harder than you might think.
- ▶ To process (& replicate) everything from scientific studies to bogus claims.
- ▶ We all are exposed to these claims - this is inevitable!
- ▶ It will make you a better voter, informed citizen, allow you to make better decisions, and differentiate fake news.

# Why Do We Do Science?

- ▶ To *explain* things?
  - ▶ Why did something happen?
  - ▶  $? \rightarrow Y$
- ▶ To *predict* things?
  - ▶ What will happen?
  - ▶  $X \rightarrow ?$
- ▶ We want to formulate and test hypotheses to make sense of the world.
  - ▶ It's always about the relationship between two events (i.e. causation).
  - ▶  $X \rightarrow Y$

# Why Do We Do Science?

- ▶ To *explain* things?
  - ▶ Why did something happen?
  - ▶  $? \rightarrow Y$
- ▶ To *predict* things?
  - ▶ What will happen?
  - ▶  $X \rightarrow ?$
- ▶ We want to formulate and test hypotheses to make sense of the world.
  - ▶ It's always about the relationship between two events (i.e. causation).
  - ▶  $X \rightarrow Y$

# Why Do We Do Science?

- ▶ To *explain* things?
  - ▶ Why did something happen?
  - ▶  $? \rightarrow Y$
- ▶ To *predict* things?
  - ▶ What will happen?
  - ▶  $X \rightarrow ?$
- ▶ We want to formulate and test hypotheses to make sense of the world.
  - ▶ It's always about the relationship between two events (i.e. causation).
  - ▶  $X \rightarrow Y$

# Why Do We Do Science?

- ▶ To *explain* things?
  - ▶ Why did something happen?
  - ▶  $? \rightarrow Y$
- ▶ To *predict* things?
  - ▶ What will happen?
  - ▶  $X \rightarrow ?$
- ▶ We want to formulate and test hypotheses to make sense of the world.
  - ▶ It's always about the relationship between two events (i.e. causation).
  - ▶  $X \rightarrow Y$

# What Claim do We (Want to) Make?

- ▶ **Non-Empirical** claims (possibly unfounded)
  - ▶ E.g., faith, astrology, authority, 'because I/he/she said so', etc.
- ▶ **Descriptive empirical** claims
  - ▶ E.g., economic growth, unemployment rates, 'let's look at the numbers', ...
  - ▶ Oftentimes: credibility by creating a sense of authority
  - ▶ Also: (intentionally) misleading
- ▶ **Inferential** claims (observational analyses)
  - ▶ E.g., regression analyses, going beyond 'raw' data
  - ▶ Can be extremely useful - but careful with interpretation (e.g., coffee & health)
- ▶ **Causal** arguments (based on actual causal inference!)
  - ▶ E.g., experimental studies, causal identification strategies
  - ▶ 'We *know* that  $X \rightarrow Y$ '
  - ▶ While most arguments claim to be causal, the threshold for causality is incredibly high!

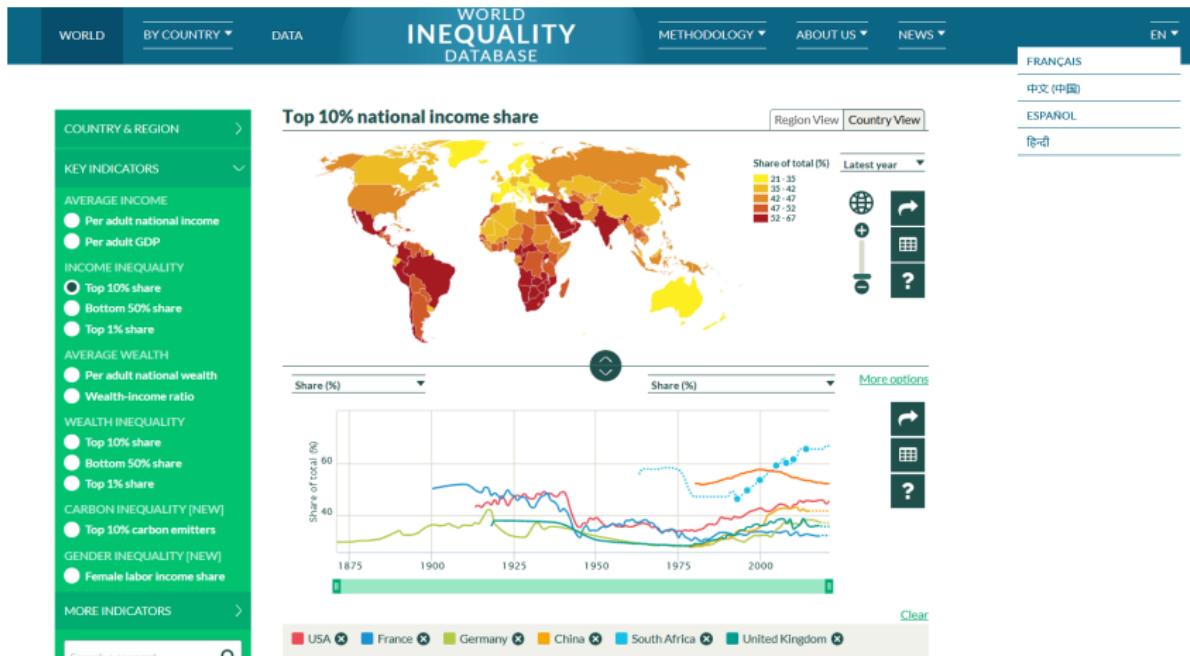
## Data Are Ubiquitous - But We Need to Make Sense of Them

## Data Are Ubiquitous - But We Need to Make Sense of Them

Data Analysis involves a 3-dimensional process (it's not statistics!)

- ▶ Research Design
- ▶ Methodology
- ▶ Implementation

# World Inequality Database



<https://wid.world>

# Our World in Data

Our World in Data

Articles by topic

Search...

Latest About Donate

All charts Sustainable Development Goals Tracker

COVID-19 vaccinations, cases, excess mortality, and much more Explore our COVID-19 data

Research and data to make progress against the world's largest problems

Scroll to all our articles

3268 charts across 297 topics  
All free: open access and open source

TRUSTED IN RESEARCH AND MEDIA

Science nature PNAS Royal Statistical Society BBC The New York Times CNN  
FT theguardian THE WALL STREET JOURNAL CNBC The Washington Post Vox

USED IN TEACHING

HARVARD UNIVERSITY Stanford BERKELEY UNIVERSITY OF CAMBRIDGE UNIVERSITY OF OXFORD MIT

<https://ourworldindata.org>

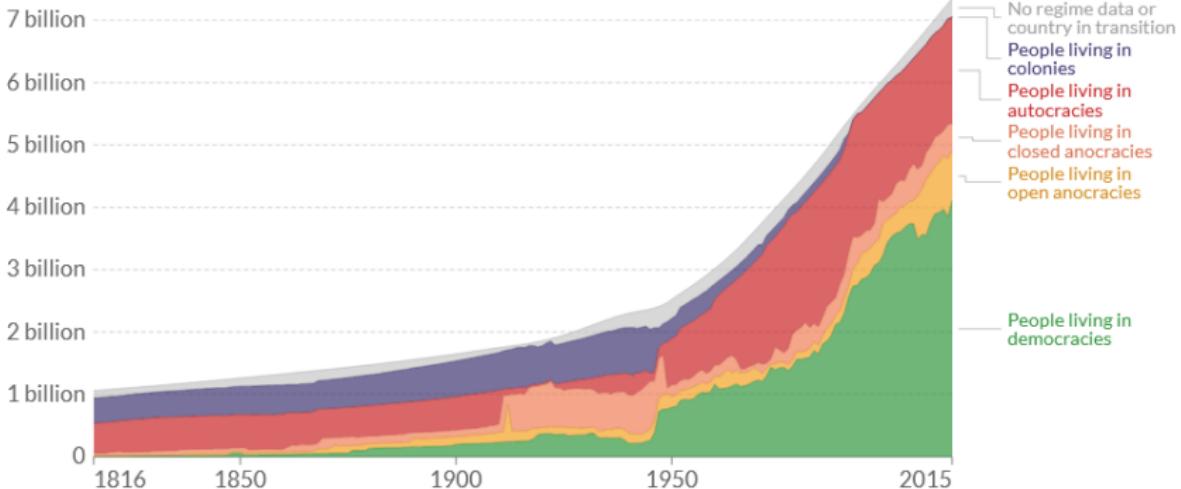
# Are the Days of Democracy Numbered?

People living in democracies and non-democracies

Political regimes based on the Polity IV data by the Center for Systemic Peace (2016).



Relative



Source: OWID based on Center for Systemic Peace (2016) and Wimmer and Min (2006)

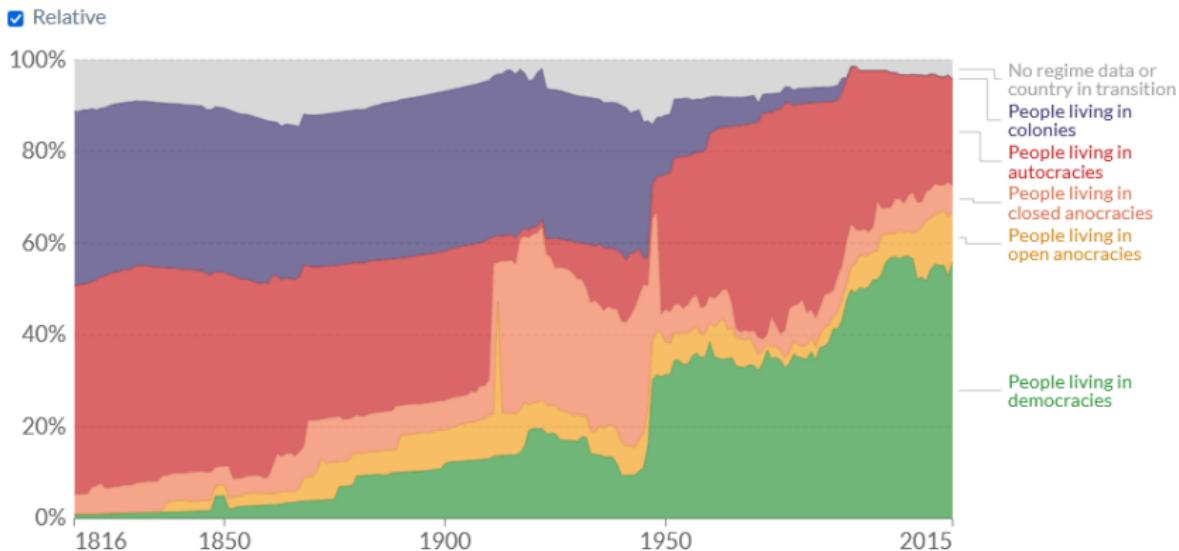
OurWorldInData.org/democracy • CC BY

# Are the Days of Democracy Numbered?

People living in democracies and non-democracies

Political regimes based on the Polity IV data by the Center for Systemic Peace (2016).

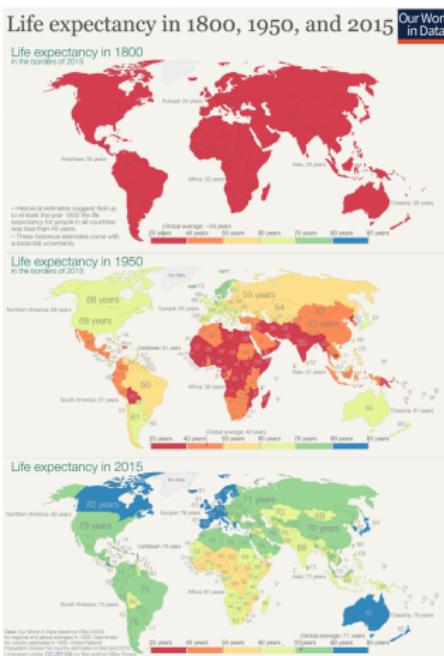
Our World  
in Data



Source: OWID based on Center for Systemic Peace (2016) and Wimmer and Min (2006)

OurWorldInData.org/democracy • CC BY

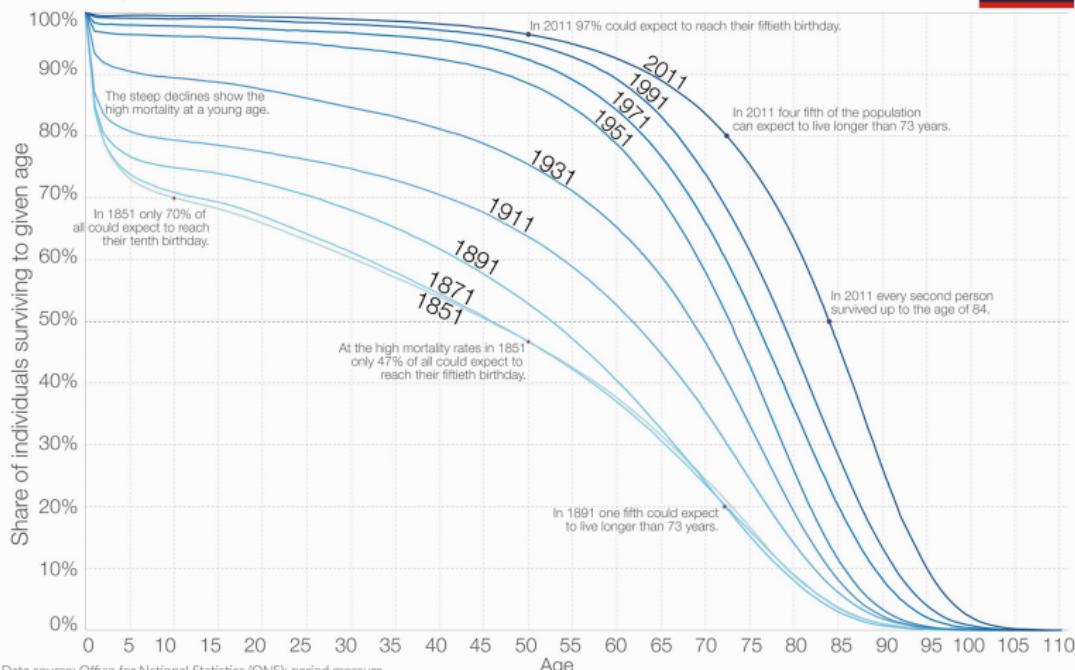
# Quantifying Human Progress I



# Quantifying Human Progress II

The survival curve for England & Wales – the share of individuals surviving up to a certain age  
Data from 1851 to 2011

Our World  
in Data



Data source: Office for National Statistics (ONS); period measure.

OurWorldInData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the author Max Roser

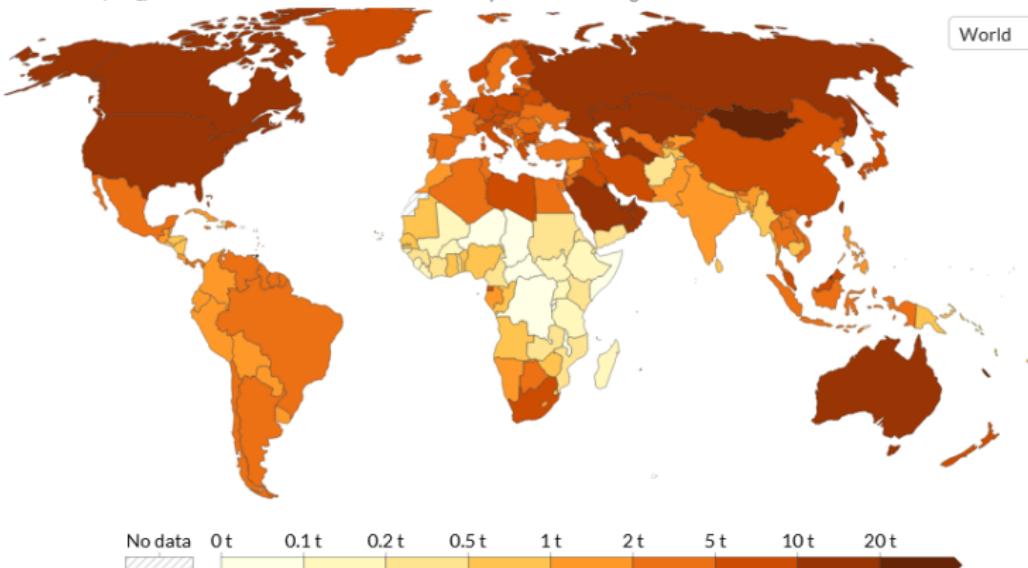
# Who's to Blame for Climate Change? - Cross-Section

## Per capita CO<sub>2</sub> emissions, 2020

Carbon dioxide (CO<sub>2</sub>) emissions from fossil fuels and industry. Land use change is not included.

Our World  
in Data

World 



Source: Our World in Data based on the Global Carbon Project

[OurWorldInData.org/co2-and-other-greenhouse-gas-emissions/](https://OurWorldInData.org/co2-and-other-greenhouse-gas-emissions/) • CC BY

# Who's to Blame for Climate Change? - Cross-Section II

## Who emits the most CO<sub>2</sub>?

Global carbon dioxide (CO<sub>2</sub>) emissions were 36.2 billion tonnes in 2017.

Our World  
in Data

### Asia

19 billion tonnes CO<sub>2</sub>,  
53% global emissions

#### China

9.8 billion tonnes CO<sub>2</sub>,  
27% global emissions

Japan  
1.2 billion tonnes  
3.3%

Saudi Arabia  
635 million tonnes  
1.8%

Thailand  
331M tonnes  
0.9%

UAE  
232M tonnes  
0.6%

Pakistan  
199M tonnes  
0.55%

Kazakhstan  
280M tonnes  
0.8%

South Korea  
616 million tonnes  
1.7%

Qatar  
227M tonnes  
0.6%

Taiwan  
272M tonnes  
0.8%

Philippines  
198M tonnes  
0.55%

Kuwait  
195M tonnes  
0.5%

Indonesia  
489 million tonnes  
1.4%

Malaysia  
255M tonnes  
0.7%

Uzbekistan  
132M tonnes  
0.37%

India  
2.5 billion tonnes  
6.8%

### North America

6.5 billion tonnes CO<sub>2</sub>,  
18% global emissions

#### USA

5.3 billion tonnes CO<sub>2</sub>,  
15% global emissions

Canada  
573M tonnes  
1.6%

Mexico  
490M tonnes  
1.4%

United States  
5.3 billion tonnes  
15% global emissions

South Africa  
456M tonnes  
1.3%

Nigeria  
379M tonnes  
1.1%

Morocco  
378M tonnes  
1.1%

Egypt  
278M tonnes  
0.8%

Argentina  
264M tonnes (0.7%)

Venezuela  
254M tonnes  
0.7%

Algeria  
151M tonnes (0.4%)

Brazil  
378M tonnes  
1.1%

Australia  
149M tonnes  
0.4%

International aviation & shipping  
1.15 billion tonnes  
3.2%

Chile  
55M tonnes (0.2%)

Oman  
52M tonnes (0.2%)

Peru  
51M tonnes (0.2%)

Greece  
50M tonnes (0.2%)

Malta  
49M tonnes (0.2%)

Other  
1.3 billion tonnes CO<sub>2</sub>,  
3.7% global emissions

### Europe

6.1 billion tonnes CO<sub>2</sub>,  
17% global emissions

#### EU-28

3.5 billion tonnes CO<sub>2</sub>,  
9.8% global emissions

Russia  
1.7 billion tonnes  
4.7%

Turkey  
149M tonnes  
1.2%

Ukraine  
212M tonnes  
0.6%

Belarus  
61M tonnes  
0.2%

Serbia  
50M tonnes  
0.1%

Northern Macedonia  
22M tonnes  
0.1%

Croatia  
21M tonnes  
0.1%

Montenegro  
17M tonnes  
0.1%

Bosnia and Herzegovina  
15M tonnes  
0.1%

North Macedonia  
14M tonnes  
0.1%

Albania  
12M tonnes  
0.1%

Other  
6.1 billion tonnes CO<sub>2</sub>,  
17% global emissions

Austria  
1.1 billion tonnes CO<sub>2</sub>,  
3.2% global emissions

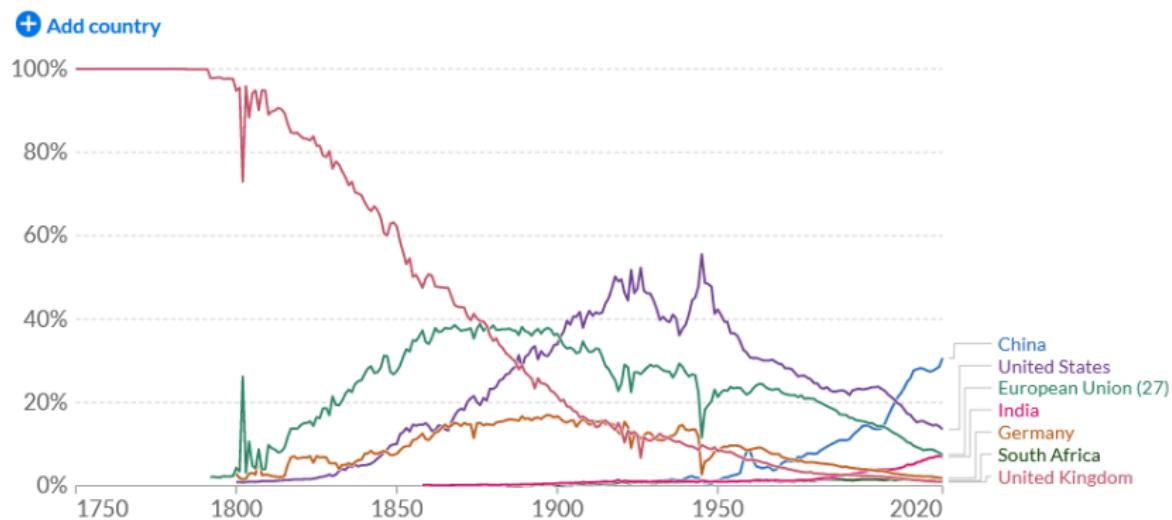
Oceania  
0.5 billion tonnes CO<sub>2</sub>,  
1.3% global emissions

# Who's to Blame for Climate Change? - Cross-Temporal

## Annual share of global CO<sub>2</sub> emissions

Carbon dioxide (CO<sub>2</sub>) emissions from fossil fuels and industry. Land use change is not included.

Our World  
in Data



Source: Our World in Data based on the Global Carbon Project

Note: This is measured as each country's emissions divided by the sum of all countries' emissions in a given year plus international aviation and shipping (known as 'bunkers') and 'statistical differences' in carbon accounts.

[OurWorldInData.org/co2-and-other-greenhouse-gas-emissions](http://OurWorldInData.org/co2-and-other-greenhouse-gas-emissions) • CC BY

# Who's to Blame for Climate Change? - Historic

Our World  
in Data

## Who has contributed most to global CO<sub>2</sub> emissions?

Cumulative carbon dioxide (CO<sub>2</sub>) emissions over the period from 1751 to 2017. Figures are based on production-based emissions which measure CO<sub>2</sub> produced domestically from fossil fuel combustion and cement, and do not correct for emissions embedded in trade (i.e. consumption-based). Emissions from international travel are not included.

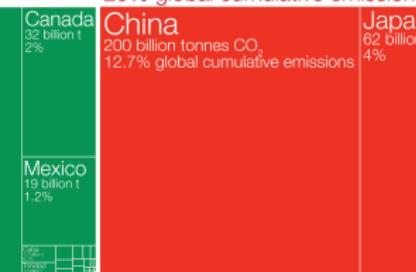
### North America

457 billion tonnes CO<sub>2</sub>  
29% global cumulative emissions



### Asia

457 billion tonnes CO<sub>2</sub>  
29% global cumulative emissions

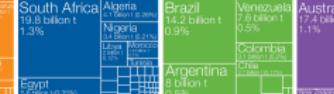
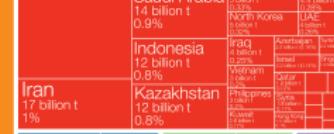


EU-28  
353 billion tonnes CO<sub>2</sub>  
22% global cumulative emissions

Russia  
101 billion tonnes  
6% global emissions



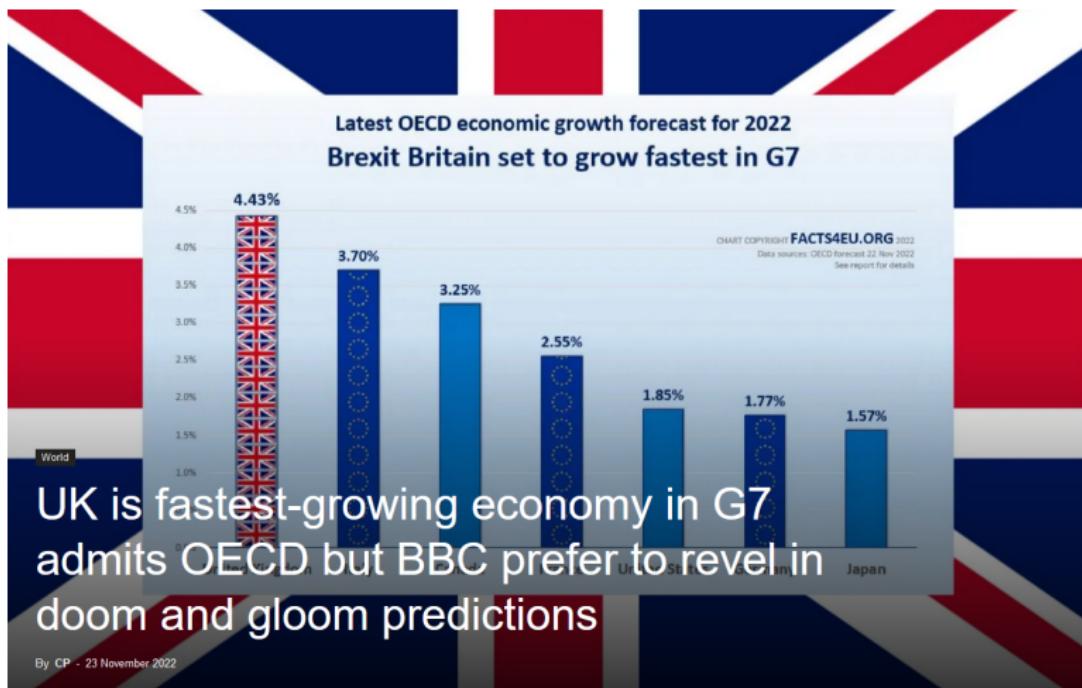
India  
48 billion t  
3%



Europe

Oceania  
20 billion tonnes CO<sub>2</sub>  
1.2% global emissions

## Example: Twisting the Numbers



## Example: Twisting the Numbers II

Bank of England + Add to myFT

### BoE warns UK set to enter worst recession for 300 years

Central bank predicts 30 per cent drop in output in first half of 2020 but opts against new stimulus

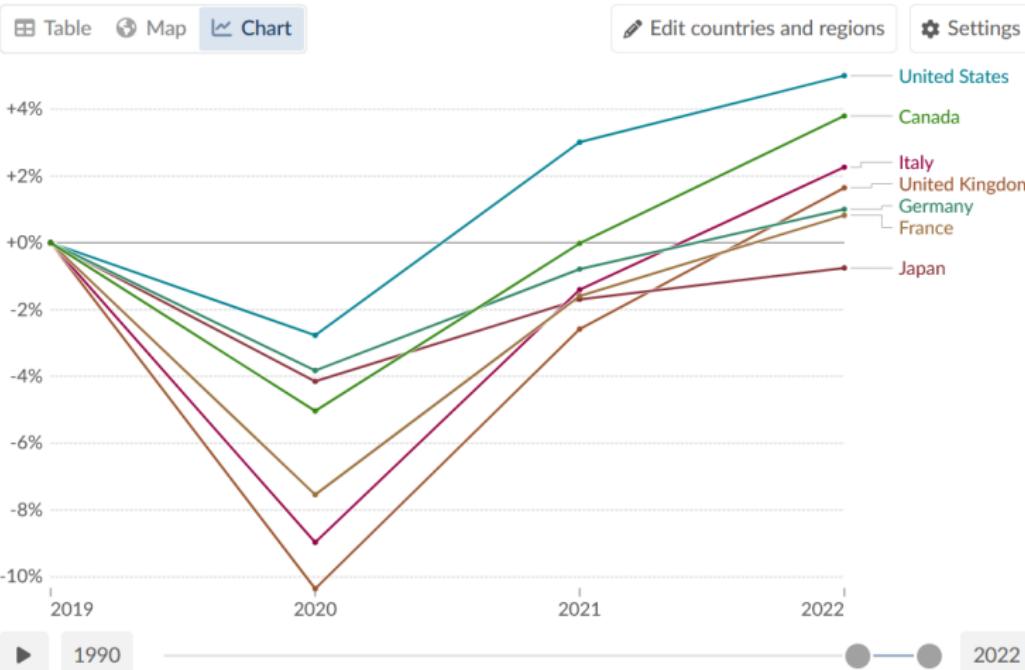


# Example: Twisting the Numbers III

## Change in gross domestic product (GDP), 2019 to 2022

This data is adjusted for inflation and differences in the cost of living between countries.

Our World  
in Data



Data source: World Bank (2023) – [Learn more about this data](#)

# Welcome to Data Analysis in R

Setup of the course:

- ▶ 3-day weekend seminar
- ▶ Subsequently, a period of self-study
- ▶ Then you'll work on a take-home exam
- ▶ You'll finally give a presentation on your results

There are three elements in this course:

- ▶ Theory (a bit)
- ▶ Learning to code in R (using real world data)
- ▶ Conduct your own statistical analyses
  
- ▶ **By end of term, you'll be able to use R conveniently and conduct your own statistical analysis (comes in handy for term papers)**
- ▶ **Keep in Mind: Data Analysis ≠ Statistics!**



## Ken Stiller

I'm a Lecturer at the University of Bayreuth. I'm also affiliated with the European University Institute and Nuffield College, University of Oxford, where I also obtained my PhD.

# Syllabus: Data Analysis in R

1. Introduction
2. Causality & Basics of Statistics
3. Visualisations & Sampling

---

4. Prediction
5. Multivariate Regression
6. Probability & Uncertainty

---

7. Hypothesis Testing
8. Assumptions & Limits of OLS
9. Interactions & Non-Linear Effects

# Assessments

- ▶ Contribution to discussions in class (10%)
- ▶ Take Home Exam (40%)
- ▶ Presentation (50%)
- ▶ Timeline:
  - ▶ **6 Feb:** Pick a dataset and run by me
  - ▶ **18 Feb:** THE distributed to you
  - ▶ **18 Mar:** THE due
  - ▶ **Apr:** Presentations (tbc)

**The use of LLMs and AI agents is absolutely prohibited!**

# Where to find our course materials

- ▶ You'll find all resources (slides, R code, exercises) on our course page:  
<https://bayreuth-politics.github.io/R2526/>
- ▶ A few times, I'll refer to textbooks. Those of interest to you are
  - ▶ Kosuke Imai's *Quantitative Social Science: An Introduction*
  - ▶ Agresti & Finlay's *Statistical Methods for the Social Sciences*