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!

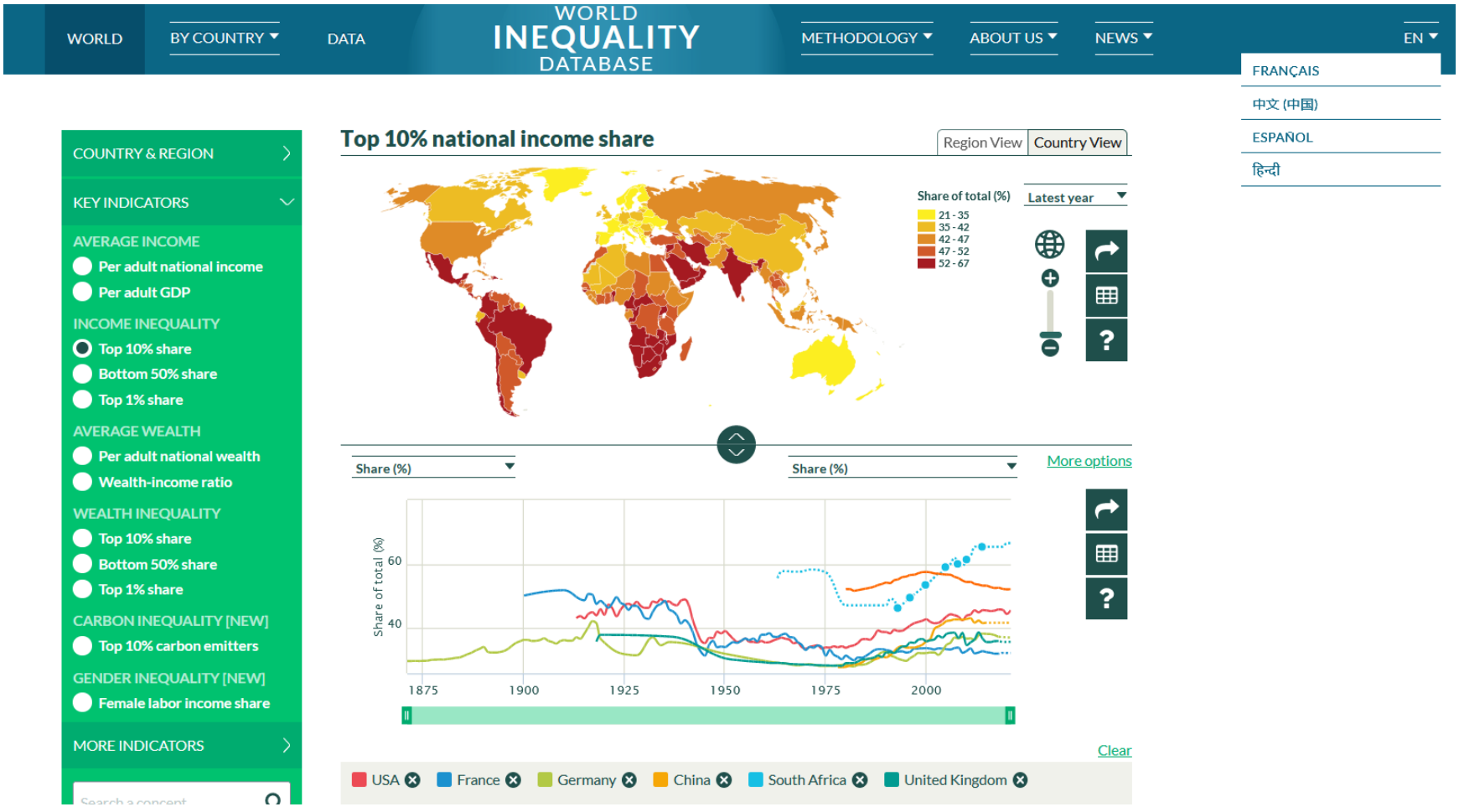
Data Are Ubiquitous - But We Need to Make Sense of Them

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

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

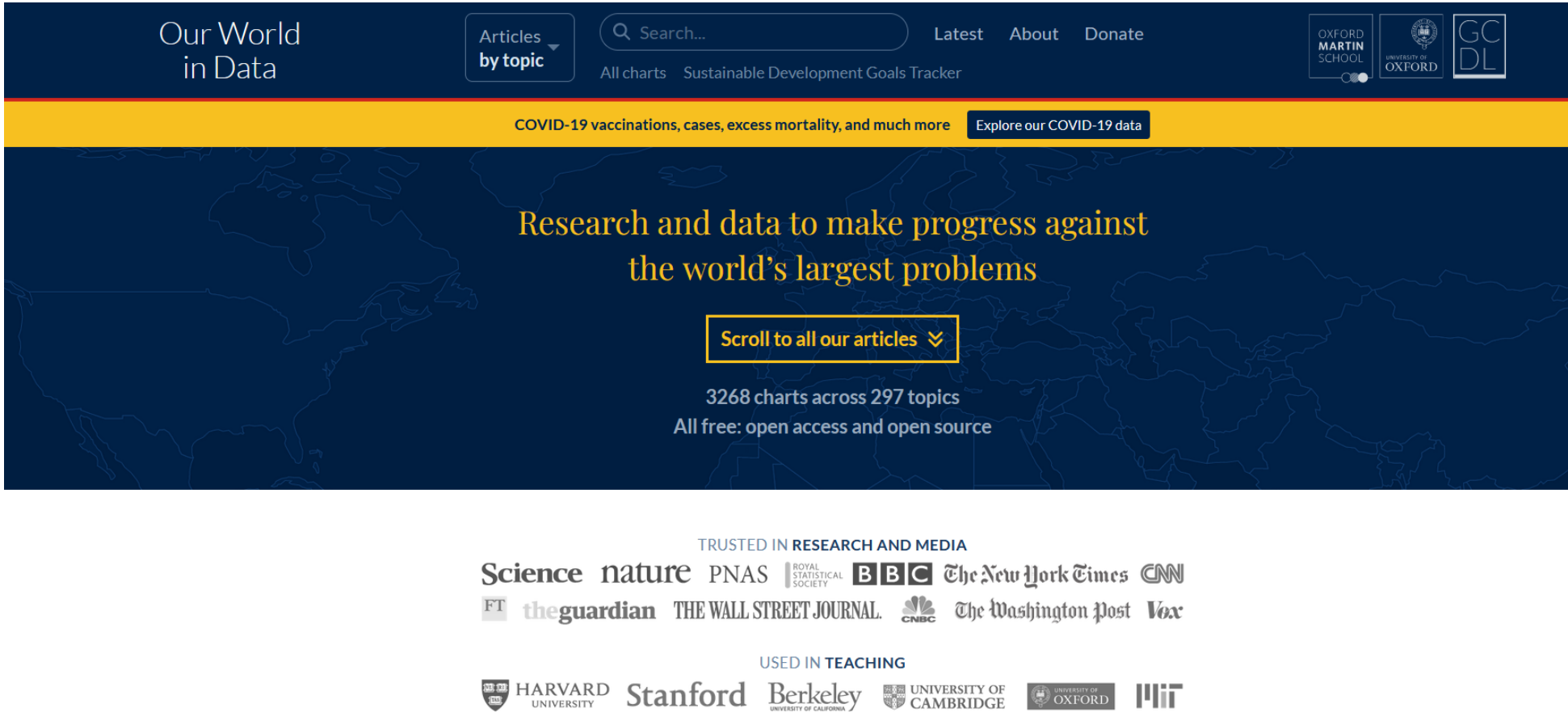
- Research Design
- Methodology
- Implementation

World Inequality Database



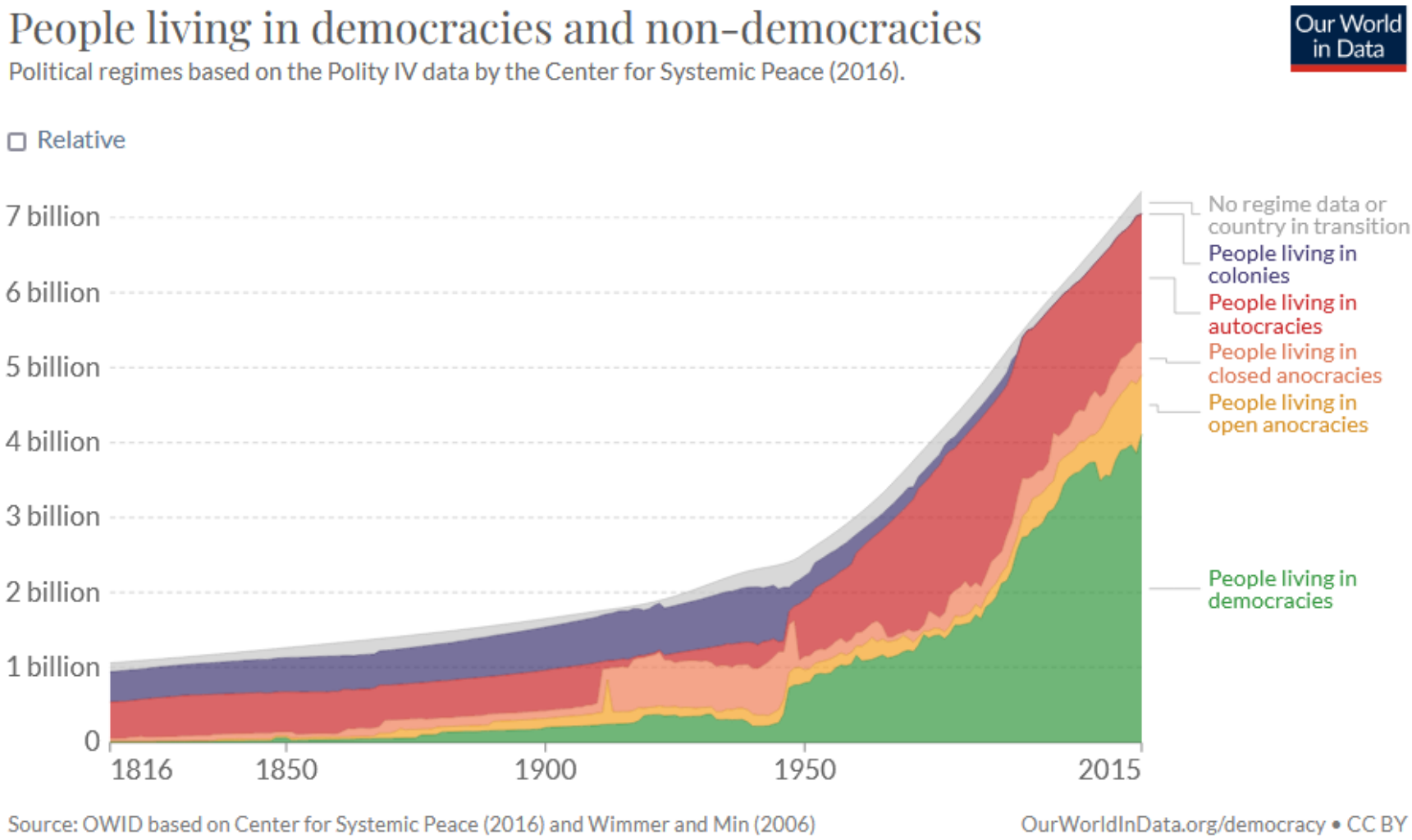
https://wid.world

Our World in Data

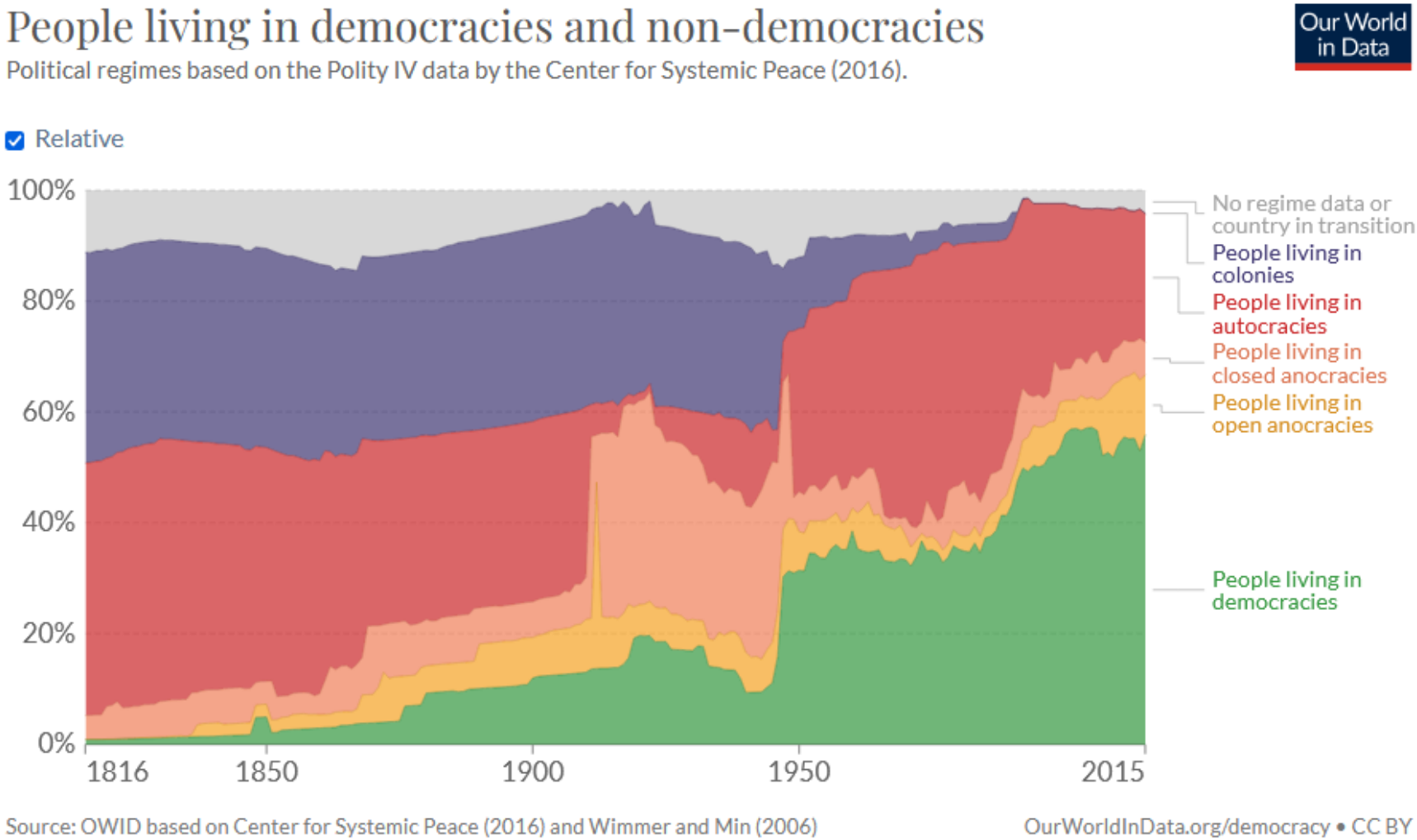


`https://ourworldindata.org`

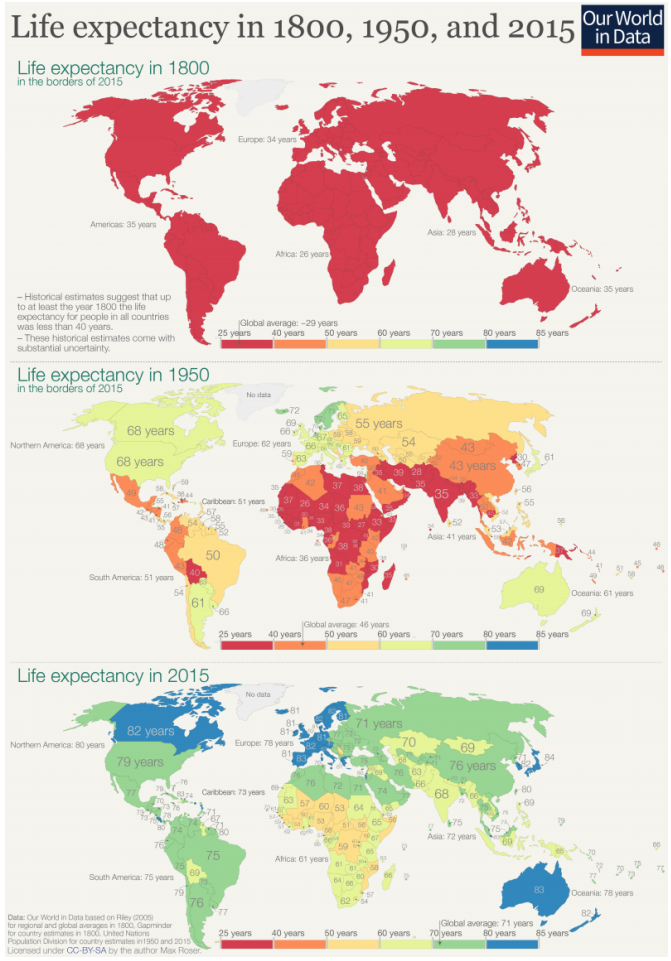
Are the Days of Democracy Numbered?



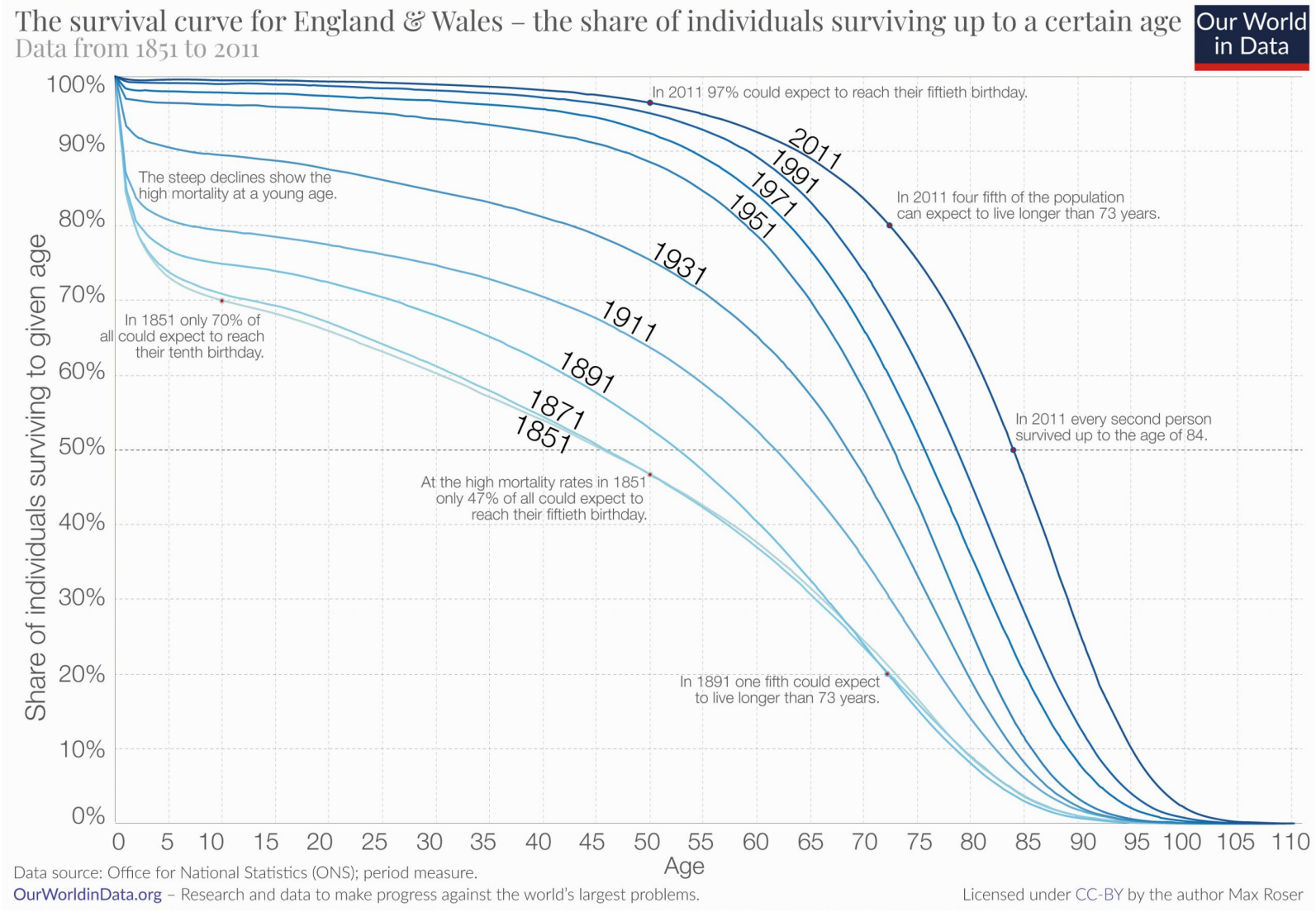
Are the Days of Democracy Numbered?



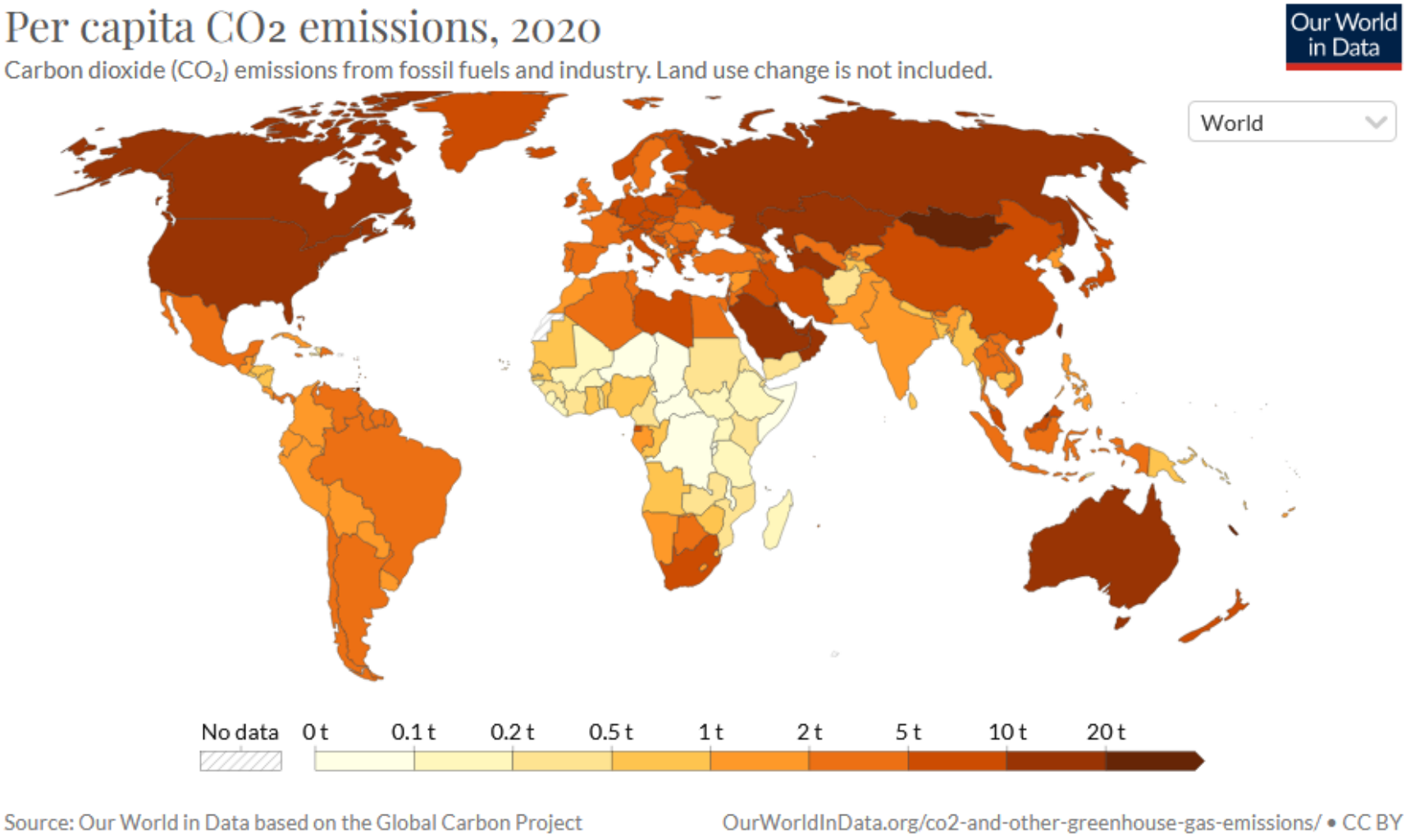
Quantifying Human Progress I



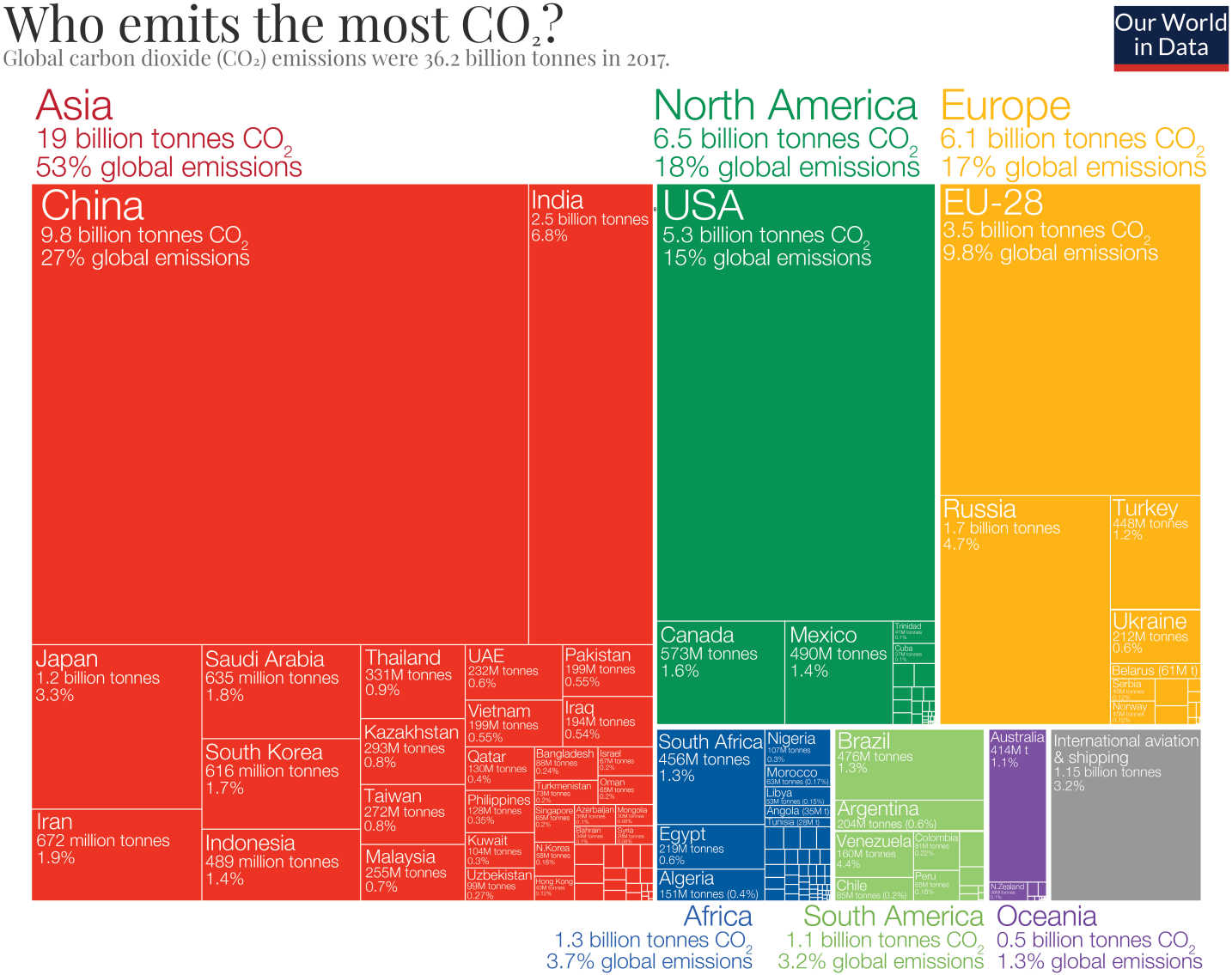
Quantifying Human Progress II



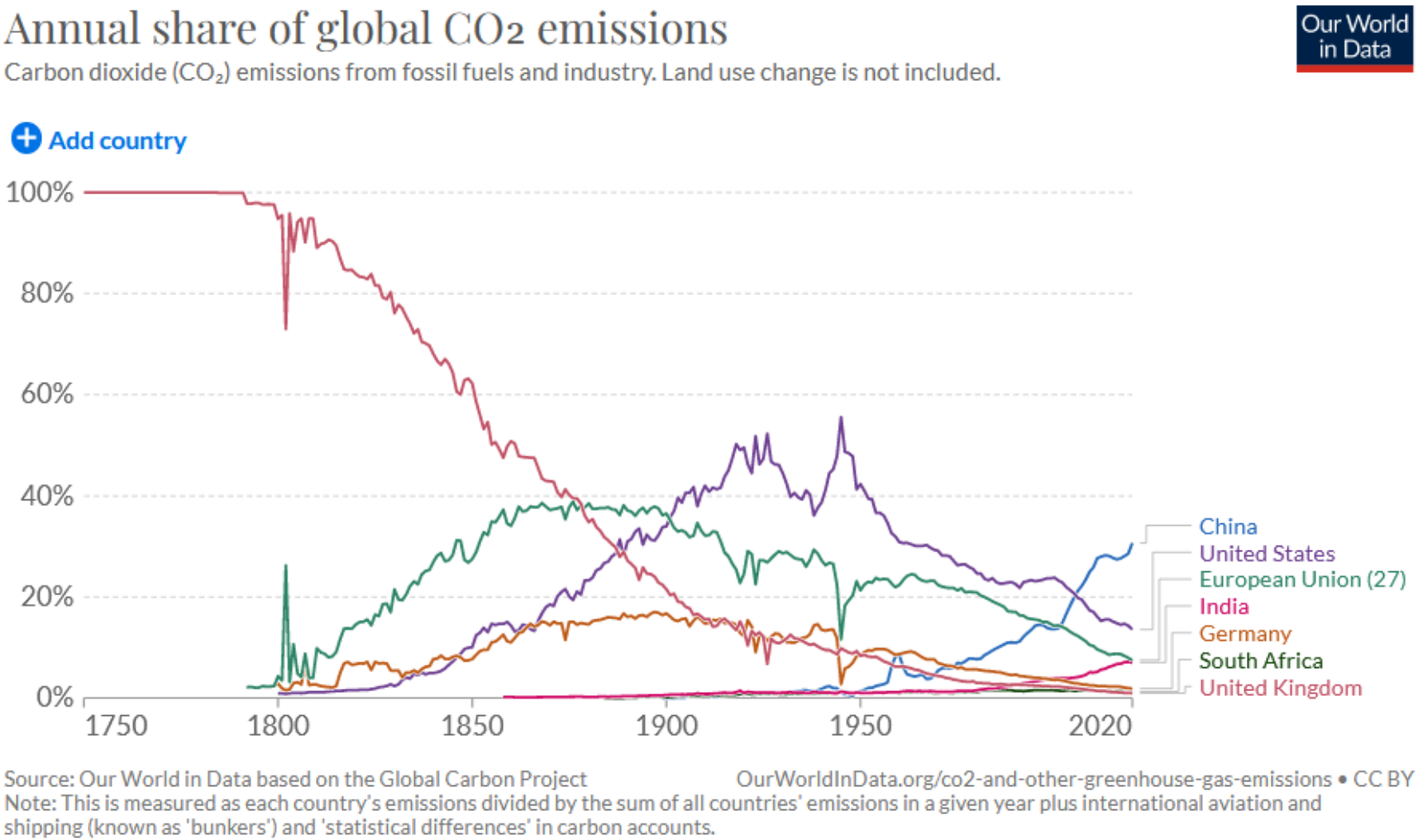
Who's to Blame for Climate Change? - Cross-Section



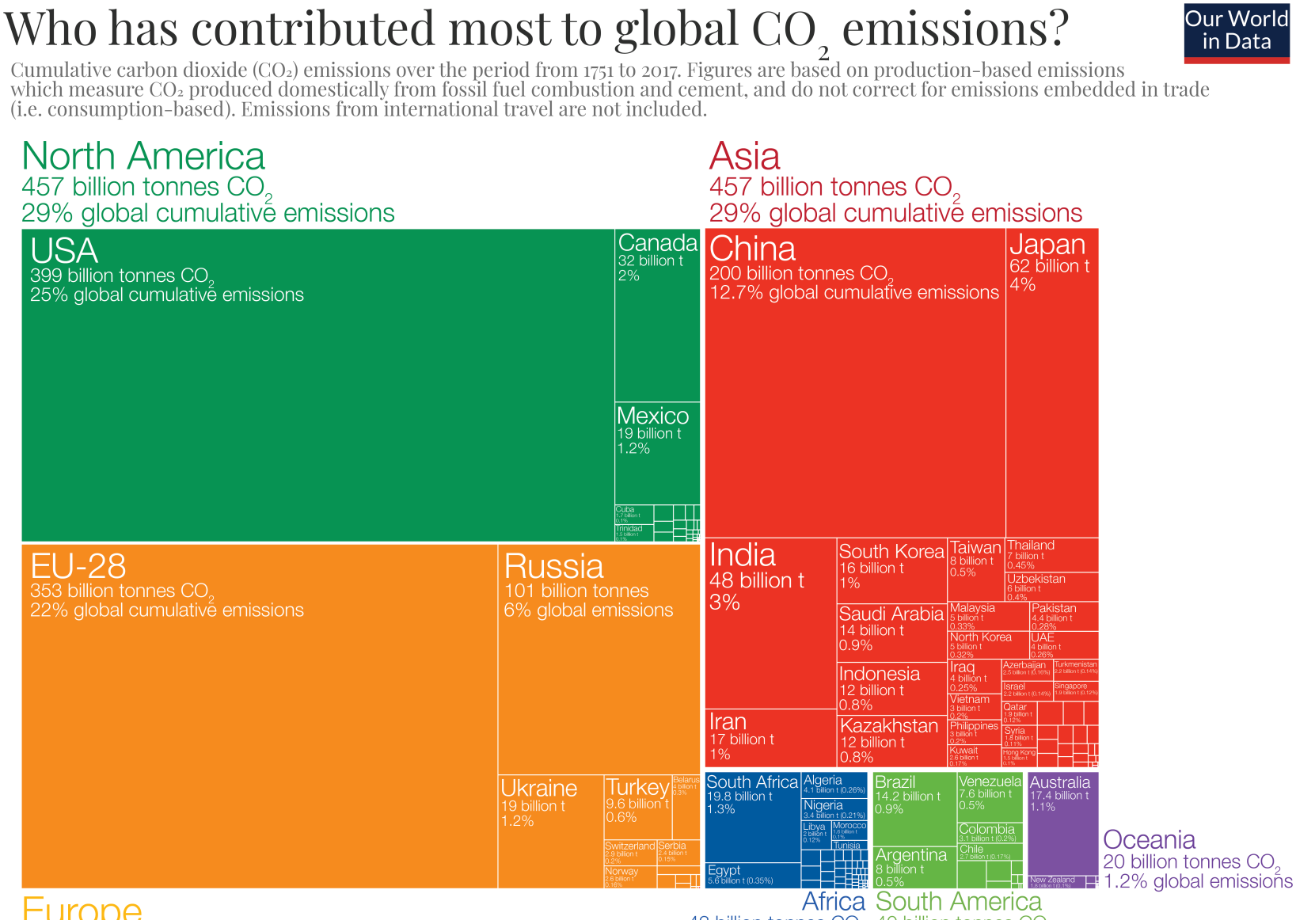
Who's to Blame for Climate Change? - Cross-Section II



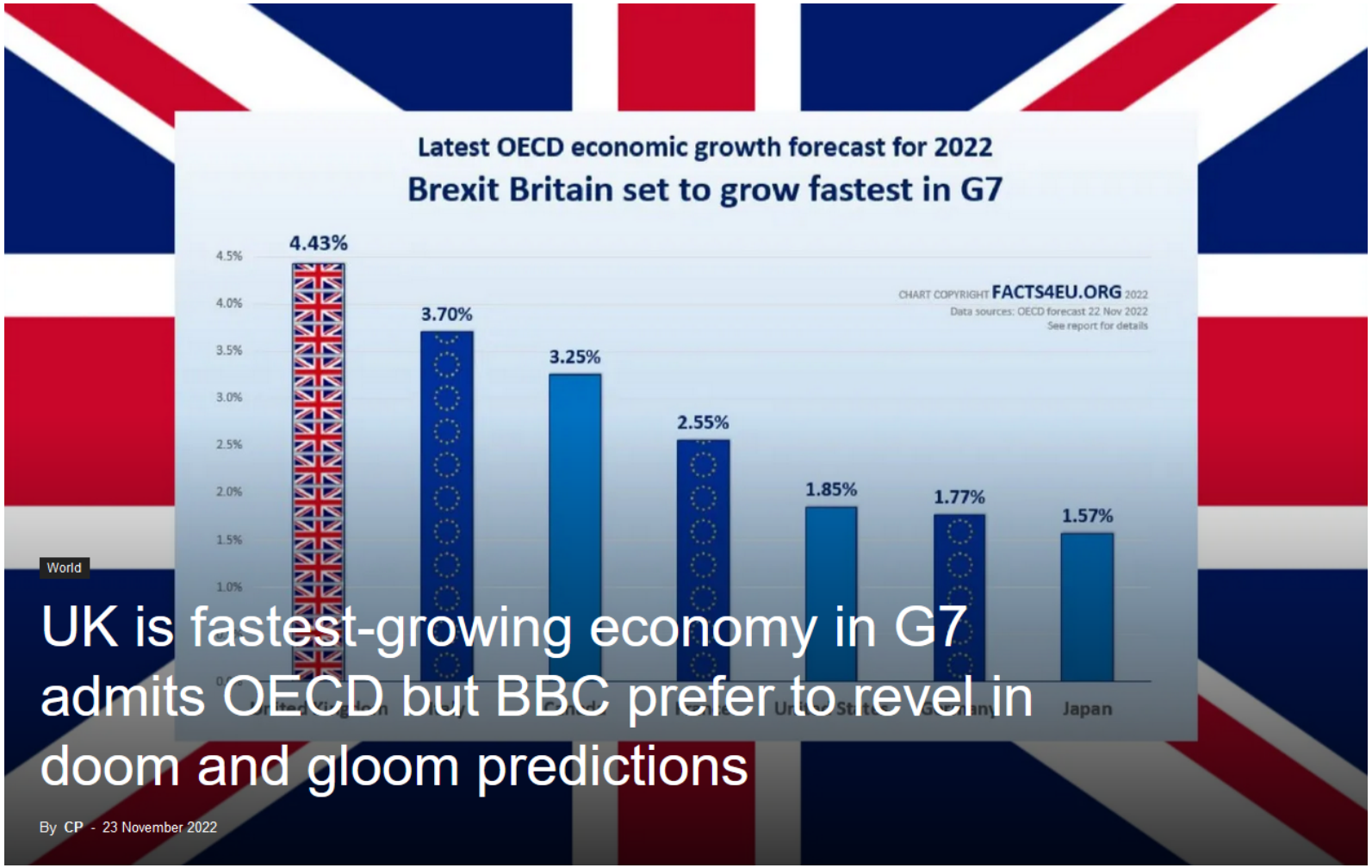
Who's to Blame for Climate Change? - Cross-Temporal



Who's to Blame for Climate Change? - Historic



Example: Twisting the Numbers



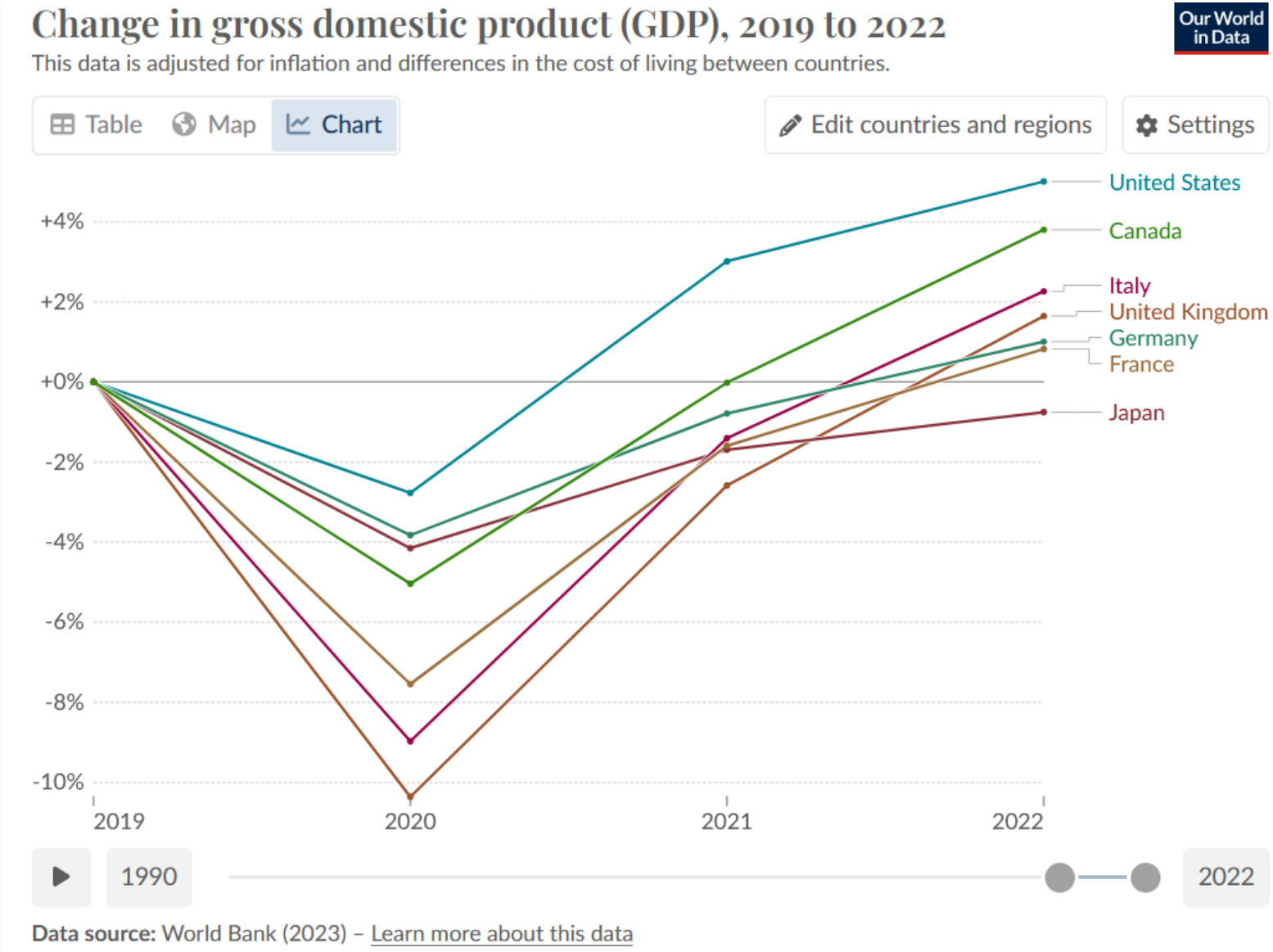
Example: Twisting the Numbers II

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



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 $\neq$ Statistics!**

Ken Stiller

I'm a Lecturer at the Univsersity 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*