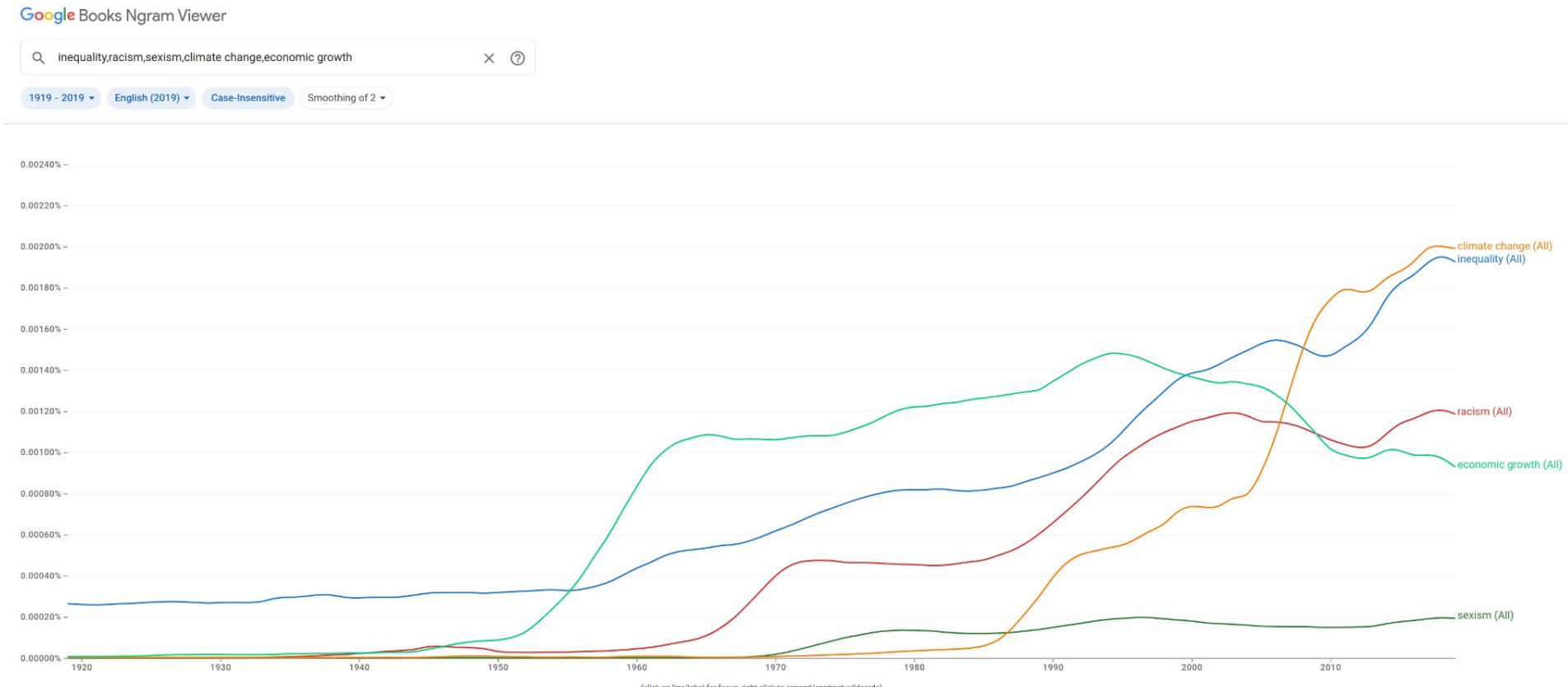
Week 7: How to approach inequality? The shifting nature of categories



This week

1. Admin: assessments (**quiz 1**; **learning journal feedback**), overview of 2nd half of the semester; any feedback from course reps;
2. Group research projects: the data, the groups, the process
    a. Group organisation and choosing a dataset
    b. Exploratory data analysis (EDA)
    c. Developing critical analysis (CDA)
    d. Statistical data analysis requirements (SDA)
3. Ideas/approaches/theory/concepts
    a. Critical analysis of categories and the role in thinking about inequalities (CDA)
    b. Estimates of uncertainty and **confidence intervals** (SDA)

week	eda	sda	cda	assessment	practical session
1	intro/setting up Rstudio	intro	intro		ryouwithme basics; cmu units 1
2	cleaning/ transforming data	kinds of quantitative data, the Big Picture	social facts and their significance		ryouwithme cleanitup; cmu unit 2, module 4
3	descriptive summaries	Descriptive statistics	standpoint in quantitative methods		ryouwithme cleanitup; cmu unit 2, module 5
4	Visualization 1 – grammar	Describing categorical data + probability	constructions of normality/average	wattle quiz (10%)	ryouwithme vizwhiz1; cmu unit 4, module 8
5	Visualization 2 – groups	Design/ random sampling	Methods make the social		ryouwithme vizwhiz2; cmu unit 5, module 11
6	Visualization 3 – combining plots	Parameters vs statistics	Groups and classes in social life	learning log due (25%)	ryouwithme vizwhiz3; cmu unit 5, module 12, 'inference for one variable'
7	Combining writing and data analysis	Kinds of statistical tests and hypotheses	Inequalities and their measurement		groupwork; cmu unit 5, module 13, 'inference for relationships C→Q'
8	Presentation and reporting techniques:	T-tests for comparing groups	Embodied differences in data		groupwork; ; cmu unit 5, module 14, 'inference for relationships C → Q'
9	Non-statistical ways of exploring differences	Chi-squared test for comparing groups using categorical data	The weight of the past in data		Groupwork ; cmu unit 5, module 15, "inference for relationships C→C;
10	Patterns and clusters	Introduction to models	Linear models of reality	**wattle quiz (wk 6-9) (10%)**	Groupwork; cmu consolidation of all work
11	Group presentations; online lecture-workship this week (TBC)			Presentation (15%)	Groupwork
12	Review – no new content				No practical session this week
13				Report (40%)	

Feedback from teaching team on learning journals/logs

1. Most people developing effective core strategies
    a. Paying close attention to the details in learning materials (CMU/ryouwithme)
    b. Practising specific EDA, SDA techniques
    c. Making notes to avoid having to ‘re-learn’
    d. Asking for help from each other or teaching team
    e. Using online help (web search, chatGPT)
2. What we liked in the learning logs
    a. Discussion of specifics about what you don’t like or do like
    b. Efforts to persevere in the face of difficulties
    c. Signs of working things out, experimenting or finding out for yourself
    d. Signs of bringing different course elements (EDA, SDA, CDA) into conversation
    e. Emerging awareness of learning something new or potentially interesting

Another view of learning logs

1. **Pay Close Attention to Details:** Carefully review materials, code, and instructions, as small errors can have significant impacts.
2. **Seek Help from Others:** Don't hesitate to ask tutors, peers, or online communities for assistance when stuck.
3. **Utilise Available Resources:** Make effective use of learning materials, online documentation, AI tools, and search engines.
4. **Experiment and Explore:** Try different approaches, run small tests, and play around with code or data to understand better.
5. **Break Down Problems:** Divide complex issues into smaller, more manageable parts to tackle them step by step.
6. **Develop a Consistent Workflow:** Establish a systematic approach to your work to improve efficiency and reduce errors.
7. **Track and Learn from Mistakes:** Note down errors and how you resolved them to avoid repeating them in the future.
8. **Be Persistent and Patient:** Problem-solving can be frustrating; persevere through difficulties and allow time for understanding.
9. **Actively Engage with the Material:** Don't just passively read; try to apply concepts, ask questions, and connect different ideas.
10. **Reflect on Your Learning Process:** Be aware of your problem-solving strategies and identify what works best for you.

https://forms.gle/BynbHdMCkNhQFdWd9 : feedback for us on where you are at with EDA, SDA, CDA

2nd half of the course

1. Where have we come from?
    a. Learning to work with data in its ‘real world’ messiness
    b. Learning techniques to select, transform, summarise and engage visually
    c. Understanding how different approaches (EDA, SDA, CDA) work with data and quantitative research methods **at various levels** (from practical to quite theoretical)
2. Where are we going?
    a. Consolidate all skills through a new project: EDA, SDA, CDA (**Group presentation, week 11)**
    b. Continue to learning SDA approaches – hypothesis testing, testing relations between variables (**Quiz week 10)**
    c. Cross-linking between different approaches (EDA, SDA, CDA) in both research project and final report (**Report week 13)**

Group research projects

Group allocations

- Each group has 1-5 people
- Groups are self-organised, and registered on wattle using the group self-selection tool (in Resources).
- Presentations are assessed i**n person** unless other arrangements are made in advance

Information about the group research projects

1. Wattle site:
    a. **Assessments** (including assessment criteria; submission of outcomes, etc).
    b. Data (see Resources)
2. Weekly worksheets (7-10) will provide guidance each week for the group work, including the presentation
3. Weekly slides will have some discussion of general topics or approaches

The datasets are posted on Wattle- Resources

1. International Social Survey Program (ISSP) on Environment and Climate, 2020-2023; 30 countries; approx 30k respondents
2. WGEA on ‘Gender Equality in Australian Workplaces’ 2023
3. World Bank Financial Inclusion Survey (Australia 2021)
4. Australian Voter Attitude Survey 2022
5. NSW Beaches
6. SOCY2038 2025 surveys
7. Other datasets, subject to tutor approval

The research problem statement

The anchoring question for whatever project you do will be: ****What does the data tell us about a social fact?****

In responding to the question, **which can be interpreted in various ways**, you will use an appropriate combination of exploratory (EDA), statistical (SDA) and critical data analysis (CDA) approaches.

What about CDA?

- Interact with some readings from weeks 1-8
- Perhaps use notebookLLM.google to engage with CDA readings and other course content?
- Develop the CDA angle by reviewing slides and readings

Categories and their calculation

Group exercise: estimating a proportion and expressing uncertainty about the estimate (‘confidence intervals’)

1. Use the class survey data from week 1-2. If you haven’t already, download the csv file from SOCY2038 Attitudes to Quantitative Methods (Responses) using File -> Download -> Comma Separated Values (csv) to your socy2038 data folder
2. In your group choose **one categorical variable** and calculate the proportions of responses (see next slide for R code to get you started).
3. Assuming that the **population** in question is socy2038 students, **estimate a confidence interval at 95%** certainty for the proportions. (CMU Unit 5 Module 12, p. 142 has guidance on this)
4. Interpret the results

Some code to get you started

```
library(tidyverse)

socy2038  <-  read_csv('data/socy2038_2025.csv')


# make it easier to work with car ownership question by creating a new car_owner variable that holds a logical TRUE/FALSE value

socy2038  <-  socy2038 %>% mutate(car_owner = tolower(do_you_own_a_car) == "yes")

# calculate proportion of students who own a car and a 95%  confidence interval 'manually'

socy2038 %>%

  group_by(car_owner) %>%

  summarise(n = n()) %>%

  mutate(p = n/sum(n)) %>%

  mutate(lower_ci = p - 1.96*sqrt(p*(1-p)/n),

                upper_ci = p + 1.96*sqrt(p*(1-p)/n))
```

```
# A tibble: 2 × 5
  car_owner     n     p lower_ci upper_ci
  <lgl>     <int> <dbl>    <dbl>    <dbl>
1 FALSE        43 0.462    0.313    0.611
2 TRUE         50 0.538    0.399    0.676
>
```

How would you interpret this result?

1. We are assuming that the population parameter of interest is ANU student car ownership.
2. The sample size is relatively small (<100)
3. The proportions are relatively close (0.46 vs 0.53)
4. The 95% confidence intervals overlap
5. The overlapping confidence intervals that the difference is not statistically significant

```
# A tibble: 2 × 5
  car_owner     n     p lower_ci upper_ci
  <lgl>     <int> <dbl>    <dbl>    <dbl>
1 FALSE        43 0.462    0.313    0.611
2 TRUE         50 0.538    0.399    0.676
>
```

Inequalities and their calculation

Great significance of estimates of inequality

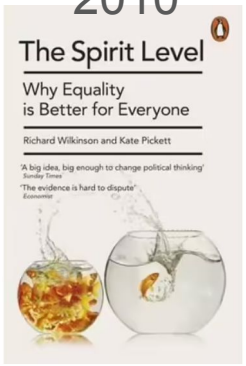
Google Books Ngram Viewer

This new big-picture social science provoked by inequality researchers met a huge public demand from large audiences trying to make sense of the rapidly changing world they were living in. Inequality provided the kind of overarching narrative that could string together economic boom and bust, globalization, political logjam and dysfunction, and wider feelings of malaise. The result has been a remarkable upturn of popular interest in social science. In the United Kingdom, the most prominent example is Richard Wilkinson and Kate Pickett's *The Sprit Level* (2010), which was a runaway best seller and prompted huge academic and policy debate through its message that unequal societies produced more social problems. Even though this book generated critical responses from some social scientists who saw it as too simplistic, this did not detract from its power to show that inequality mattered.

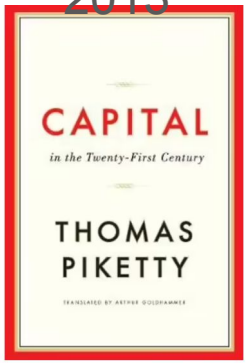
The same point became even more clear in 2013 with the publication of Thomas Piketty's *Capital in the Twenty-First Century*. A dense data-driven tome laying out long-term trends of income and wealth inequality in numerous nations was not an obvious candidate to sell two million hardbacks. Nonetheless, this book captured the public imagination to a remarkable extent and catapulted Thomas Piketty to superstar status. This was not the kind of attention that social scientists had normally enjoyed, and it easily exceeded the attention that any big data exponents were attracting.

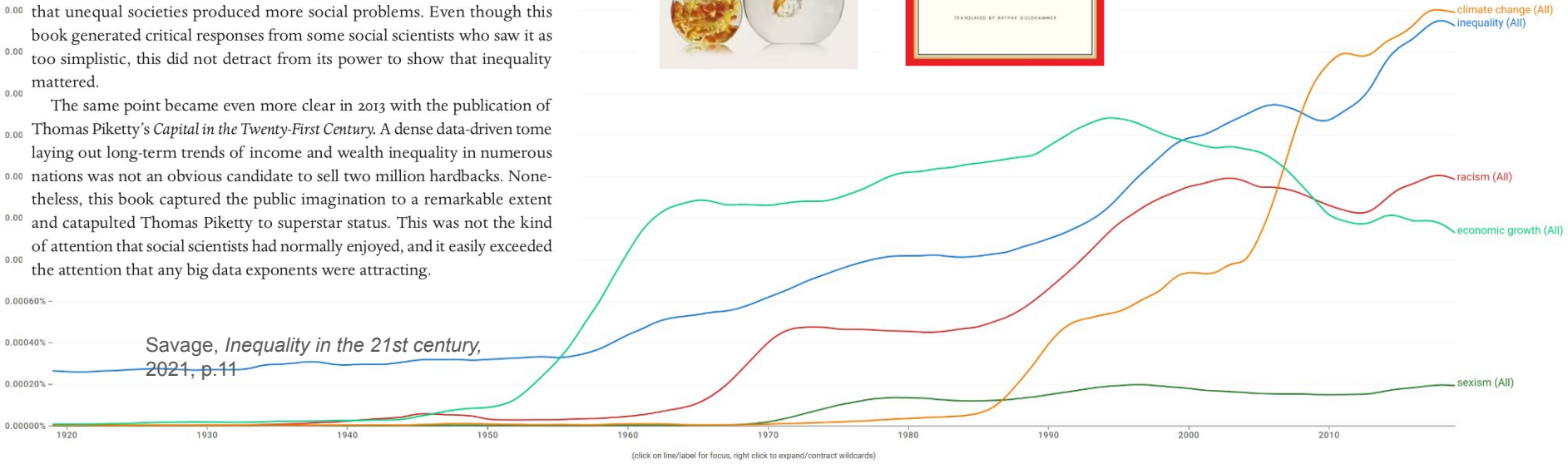
Savage, *Inequality in the 21st century*, 2021, p.11

2010



2013





How is inequality calculated?

‘It is common to separate out different axes of inequality, to prioritize some kinds over others, and to split them into different analytical registers—for instance, separating out inequalities of distribution, opportunity, and recognition’ (Savage, *The Return of Inequality: Social Change and the Weight of the Past*, 169)

Discussion exercise: what is the difference between the 2 figures?

Questions to discuss:

1. What is '**income inequality**?' How does income inequality differ from average income?
2. What is an '**index of** health and social problems'?
3. What is the **trendline** in Figure 2.2 showing?
4. Why does Figure 2.3 have **no trendline**?

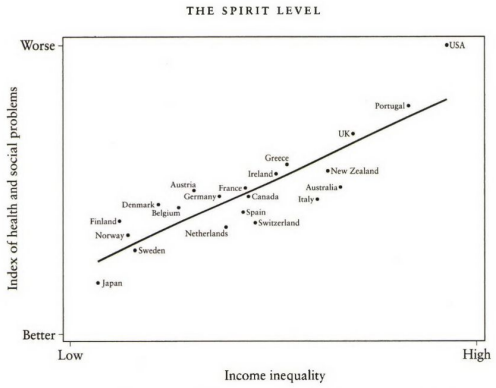


Figure 2.2 *Health and social problems are closely related to inequality among rich countries.*

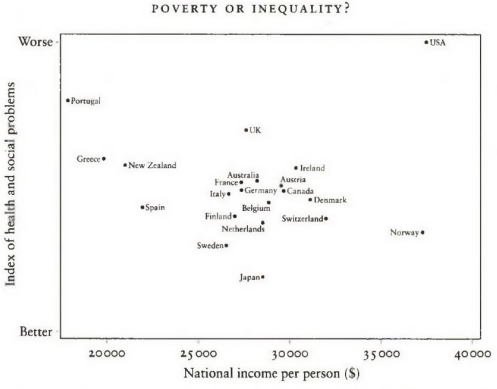


Figure 2.3 *Health and social problems are only weakly related to national average income among rich countries.*

Re-defining inequality *statistically*: from race, gender and class to ‘unequal chances’

During the twentieth century, as economic inequality declines within nation spaces, groups become defined *probabilistically*—that is to say, less as natural, moral, or ontological categories and more in terms of the differential life chances, advantages, and disadvantages they experience.

The very formation of nation spaces is bound up with mechanisms allowing the measurement of the differential life chances of its inhabitants, seen as located in its bounded territory. These construe field relations around national *parameters*. (Savage, *The Return of Inequality: Social Change and the Weight of the Past*, 170)

Thinking about inequality in 21st century

The shift toward lower inequality regimes during the twentieth century is to be understood as tied up with the rise of nation spaces and the probabilistic rendering of the stuff of inequality. This reflects a world mired by the relativization of categorical groups seen as comprising differing national opportunity structures.

Public services around the globe seek to identify and modulate (though not always with wholehearted commitment) these probabilistic inequalities. Corporations and employers of all kinds evaluate their members' categorical characteristics so as to manage equal opportunities—not always very successfully.

This current has ushered in a way of thinking about inequality between groups not in terms of their intrinsic properties, but in terms of the relative advantages or disadvantages that they experience.

This idea that inequality is to be understood probabilistically is now utterly ubiquitous, almost to the extent that it is a truism. However, we should not see this framing as inevitable or necessary. (Savage, *The Return of Inequality: Social Change and the Weight of the Past,* 194-195)

Visceral inequality today

‘Therefore, in recent decades, the stuff of inequality is increasingly rendered as visceral, physical, and embodied, rather than abstracted into relative categories. This is a far-reaching process that generates tensions and conflicts that are increasingly virulent—and which can be expected to become more so. Visceral inequality recognizes the historically different mechanisms by which the stuff of in equality is rendered, or more exactly, categorical distinctions are produced’ (Savage, *The Return of Inequality: Social Change and the Weight of the Past,* 2021, 227)

Review and next time

- Feedback on learning logs
- Group research projects – datasets, organisation, and next steps
- Inequalities and the changing nature of categories: from essence to probabilities
- Testing categories using probabilities: e.g. from socy2038

Week 8

- Group projects: continue EDA (Exploratory Data Analysis); develop CDA approach (Critical Data Analysis); plan some SDA (Statistical Data Analysis)