Applied Research Study Session 1: Visualization of Data and Frequencies

Dr. L. Sanna Stephan

2023

RUG Groningen

Learning objectives

You will learn

1. ... that data manipulation can be done in various ways, some of which are subtle and easy to overlook,
2. ... that the way data is graphically presented can easily lead to false conclusions and
3. ... that outliers can have enormous effect on estimates and conclusion.

Many shades of wrong: what governments do when they manipulate statistics

Introduction

- Authors: Roberto Aragão and Lukas Linsi (2022)
- Focus on three cases:
  - Greece's public deficit figures (2000s)
  - Argentina's inflation statistics (2007-2015)
  - Brazilian "fiscal pedaling" (2012-2015)
- Goal: examine how official statistics are being manipulated

Key messages

- Governments want to present statistics that make their performance look good and have a monopoly on national statistical data.
- Many studies find empirical evidence of data manipulation by governments.
- Economic indicators are much more ambiguous than typically considered.
- There are different types of manipulation.

- Most previous studies presume that there is **only one correct estimate** of each economic indicator, which governments **willingly choose to misreport**.
- This is an oversimplification!
  Two sources of ambiguity:
  1. International statistical standards have to accommodate a wide range of disparities between countries $\rightarrow$ intentional ambiguity
  2. (Unintentional) measurement inaccuracies

Four strategies for data manipulation

- **Type 1: outright manipulation**
  - "Correct" indicator known
  - Government pressures agencies to publish wrong figures
  - Drawbacks: strong executive control required, risky and could result in scandal

Four strategies for data manipulation

- **Type 2: politically convenient guesstimating**
  - Low level of statistical capacity
  - Uncertainty about the "true" value of statistical indicators
  - Choose (among several estimates of the indicator) the most flattering one
  - Drawback: less relevant when statistical capacity is higher

.

Four strategies for data manipulation

- **Type 3: opportunistic use of methodology space**
  - "Actual" value of the indicator known
  - Governments encourages use of alternative methodological approaches to produce more convenient figure
  - Drawback: may fail if statistical producers are independent and legally protected

Four strategies for data manipulation

- **Type 4: indicator-management through indirect means**
  - Adapt operational procedures in order to tweak the raw data in the politicians' favour
  - Drawback: limited room for maneuver

Four strategies for data manipulation: summary

	manipulation	guesstimating	methodology	indirect
do experts agree on the actual number?	Yes ✓	No ✗	Yes ✓	Yes ✓
do politicians pressure experts to change headline figures ?	Yes ✓	Yes ✓	No ✗	No ✗
Is it necessary to influence methodological choices	No ✗	No ✗	Yes ✓	No ✗
bluntness of intervention	++++	+++	++	+
typical statistical capacity	irrelevant	low	high	high
independence of statistical apparatus	very low	low	low	high

Table 1: Summary of the types of statistical manipulation

Case 1: Argentina's inflation statistics

- Outline: type 4 $\rightarrow$ type 3 $\rightarrow$ type 1.
- 2006: type 4 manipulation attempted
  - Statistical agency was pressured to reveal which prices from which shops were used for inflation index.
  - Potential goal: pressure shop owners to adjust the prices of products in the index
- Technicians in the statistical agency refused to collaborate.

Case 1: Argentina's inflation statistics

- Shift to type 3:
- Government officials argued that statistical office uses 'unpatriotic' methodology.
- Technicians refused to collaborate again.
- Eventually, director of department responsible for inflation statistics was pushed out.

Case 1: Argentina's inflation statistics

- Former director pushed out $\rightarrow$ statistical agency forced to cooperate
- The following type 3 strategies were used:
  - Using secondary data from government-controlled institutions
  - Excluding certain products from inflation calculations due to 'abnormal' price variation
  - As the above proved insufficient: systematically excluding all prices that increased by more than 15%.

Case 1: Argentina's inflation statistics

- Differences between official and personally experienced inflation rates became large enough to cause suspicion.
- Alternative inflation measures came into usage.
- Key takeaway: possibility to adjust macroeconomic statistics within a country, implementation challenging in particular in long run

Case 2: Brazilian "fiscal pedaling"

- Outline: type 3 + type 4
- Context: slowed economic growth in Brazil in the early 2010s as a result of the 2008 financial crisis and political uncertainties
- Upcoming election: government wants public finances to appear healthy
- Key adjustments:
  1. Changing the methodology for measuring public debt
  2. Removing expenditures (social programs and credit incentives) from calculation of public deficit

Case 2: Brazilian "fiscal pedaling"

- 2010: government approached the International Monetary Fund (IMF) to obtain approval for a methodological change (type 3 attempt) in calculating public debt.
- Argument: Brazilian monetary policy inflates debt but is beneficial to country's solvency
- Change accepted, but IMF keeps track of alternative numbers with old methodology

Case 2: Brazilian "fiscal pedaling"

- Second effort: manipulating public debt indicators through indirect means (type 4).
- Creative usage of public banks' balance sheets in order to conceal large amounts of public expenditures (outstanding payments not recorder until they are made)

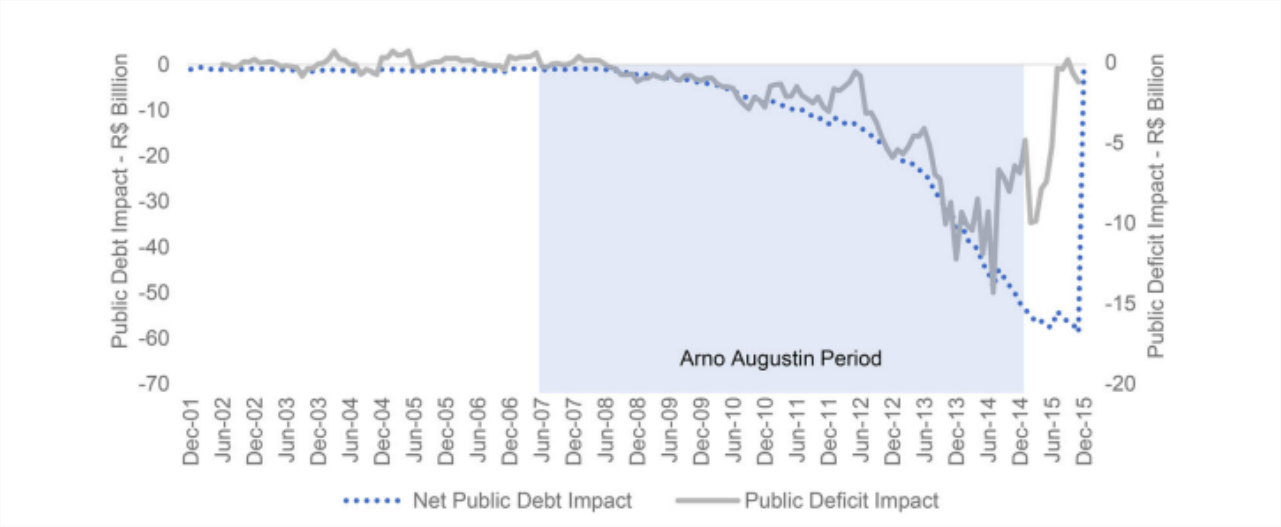


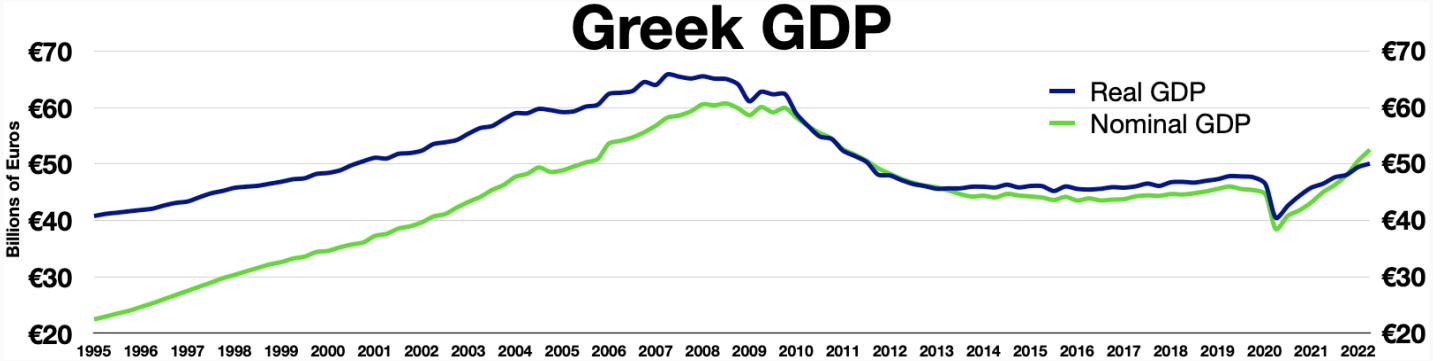
Figure 1: Brazilian public debt and deficit over time

Case 2: Brazilian "fiscal pedaling"

- Support for the government fell drastically $\rightarrow$ impeachment of the Brazilian president in 2016
- Key takeaways:
  1. Political power $\rightarrow$ international standards can be modified.
  2. Adjust economic indicators comes with substantial political risk.

Case 3: Greece's public deficit figures

- Context: sovereign debt crisis following the 2008 financial crisis



- Frequently cited cause: "deception" of the Greek government in producing official statistics (type 1 allegation), leading to a loss of confidence

Case 3: Greece's public deficit figures

- But: better classified as a type 2 situation, Greek officials take advantage of actual numbers not being known.
- Since 2004, various institutions (e.g. OECD, Eurostat) had expressed concern about validity of Greek macroeconomic statistics.
- Eurostat: official deficit figures are systematically understated (military and social security expenses excluded)

Case 3: Greece's public deficit figures

- Greece's auditors themselves deeply unsure of the "actual" state of Greece's deficit/debt

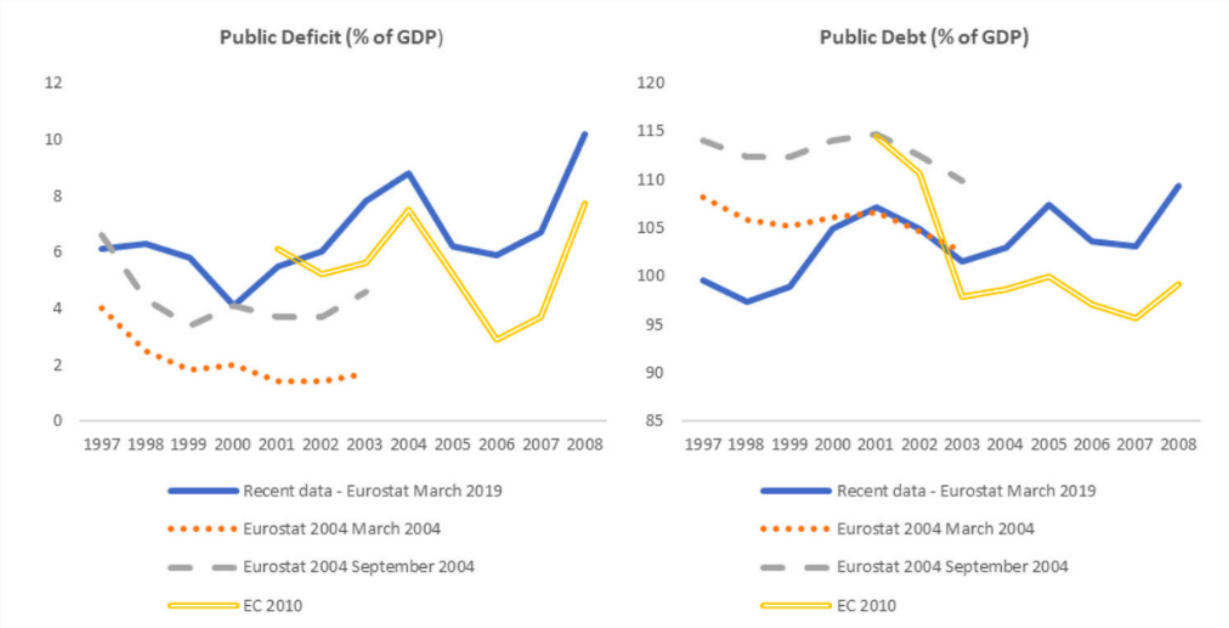


Figure 2: Eurostat estimations of Greece's public deficit and debt

Case 3: Greece's public deficit figures

- Greek officials knew forecasts were overly optimistic, but the extent of underestimations was not entirely known.
- Several pieces of evidence indicating that European institutions were aware of issues in Greek statistics long before the crisis took place.

Aragão and Linsi (2022): conclusion

- Distinguishing between "right" and "wrong" numbers is not clear in practice.
- Politics of data manipulation play out over "different shades of 'wrong'": four distinct types are distinguished.
- Cases analysed do not constitute type 1 outright headline figure fabrication.
- Extent to which manipulation can take place depends on
  - ... political power (local/international)
  - ... capacity and autonomy of statistical agency
  - ... compliance of public

Data (Mis)representation and COVID-19: Leveraging Misleading Data Visualizations For Developing Statistical Literacy Across Grades 6–16

Introduction

- Authors: Christopher Engledowl and Travis Weiland (2021)
- Data during COVID-19 pandemic has been (mis-)represented and (mis-)interpreted by governments and the media.
- Statistical education should be enhanced.

Context

- COVID-19 pandemic $\rightarrow$ people from all backgrounds want to consume statistics
- Lack of statistical literacy throughout society has resulted in widespread data misrepresentation and misinterpretation.

Case 1: Kansas department of health and environment

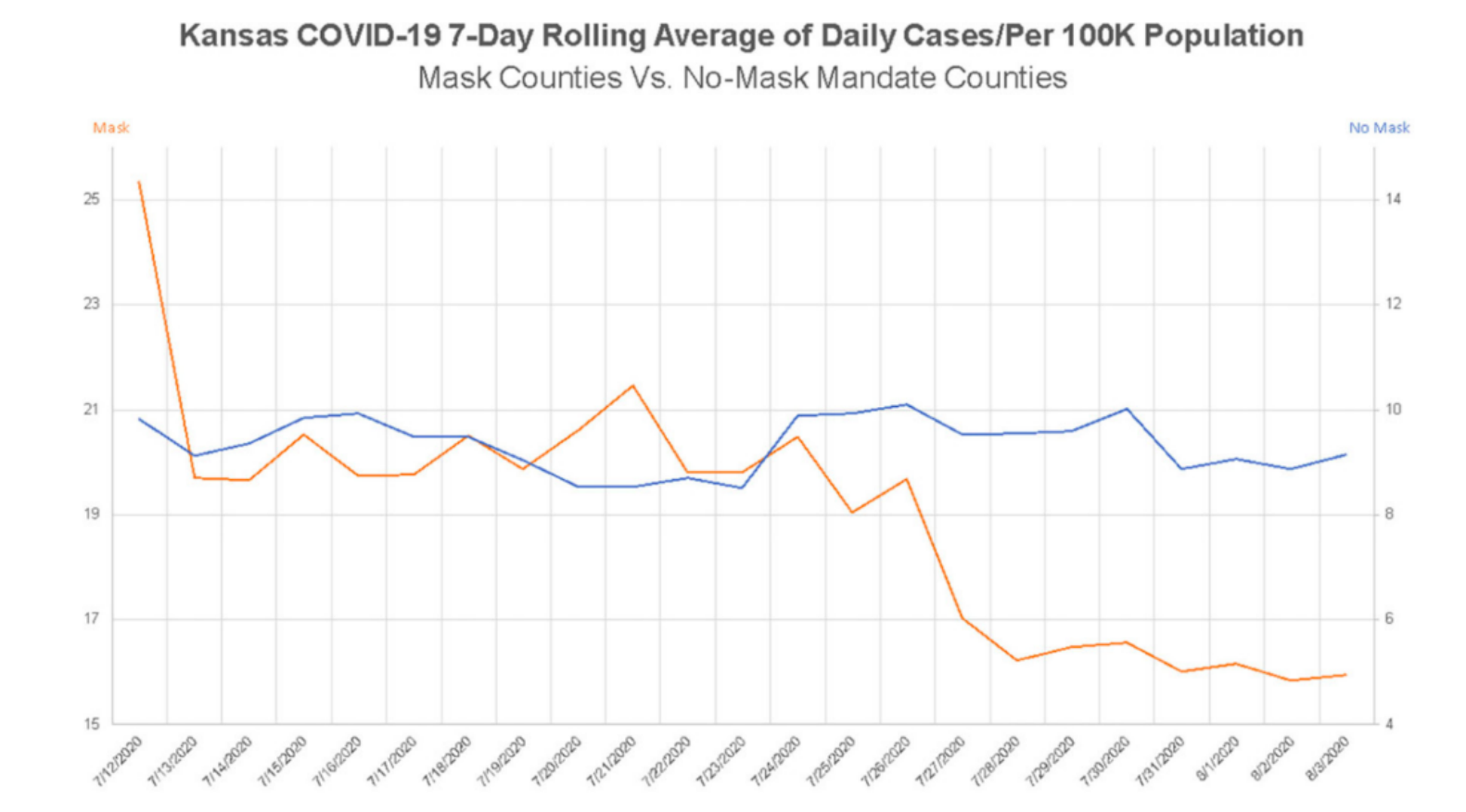


Figure 3: Rolling average of daily COVID-19 cases in Kansas counties with mask mandates (orange line) and without mask mandates (blue line)

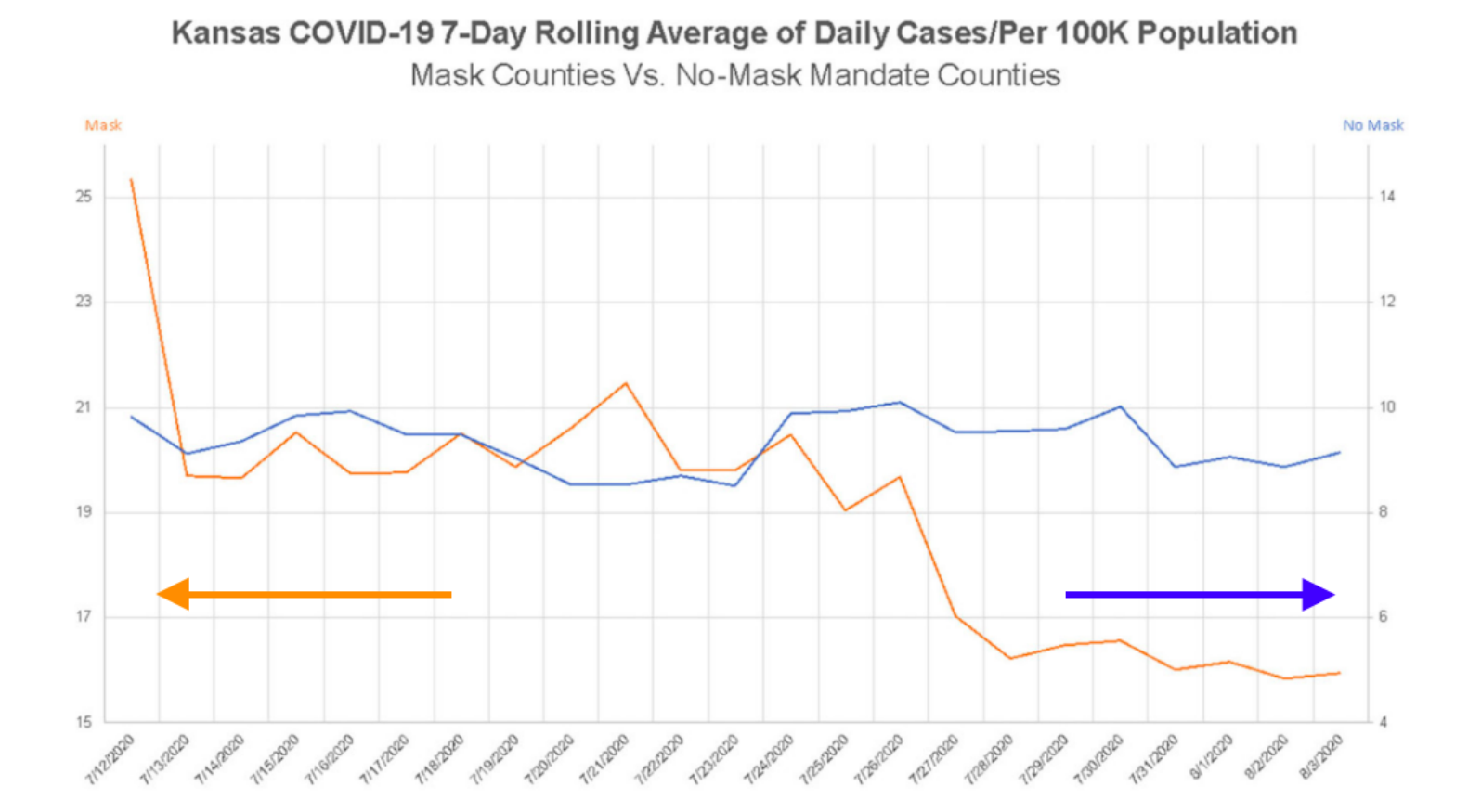
Case 1: Kansas department of health and environment

Intended message:

- Counties with mask mandates (orange line): stark decline in COVID-19 cases
- Counties without mask mandates (blue line): number of cases remained relatively

However, notice that the graph contains two vertical axes!

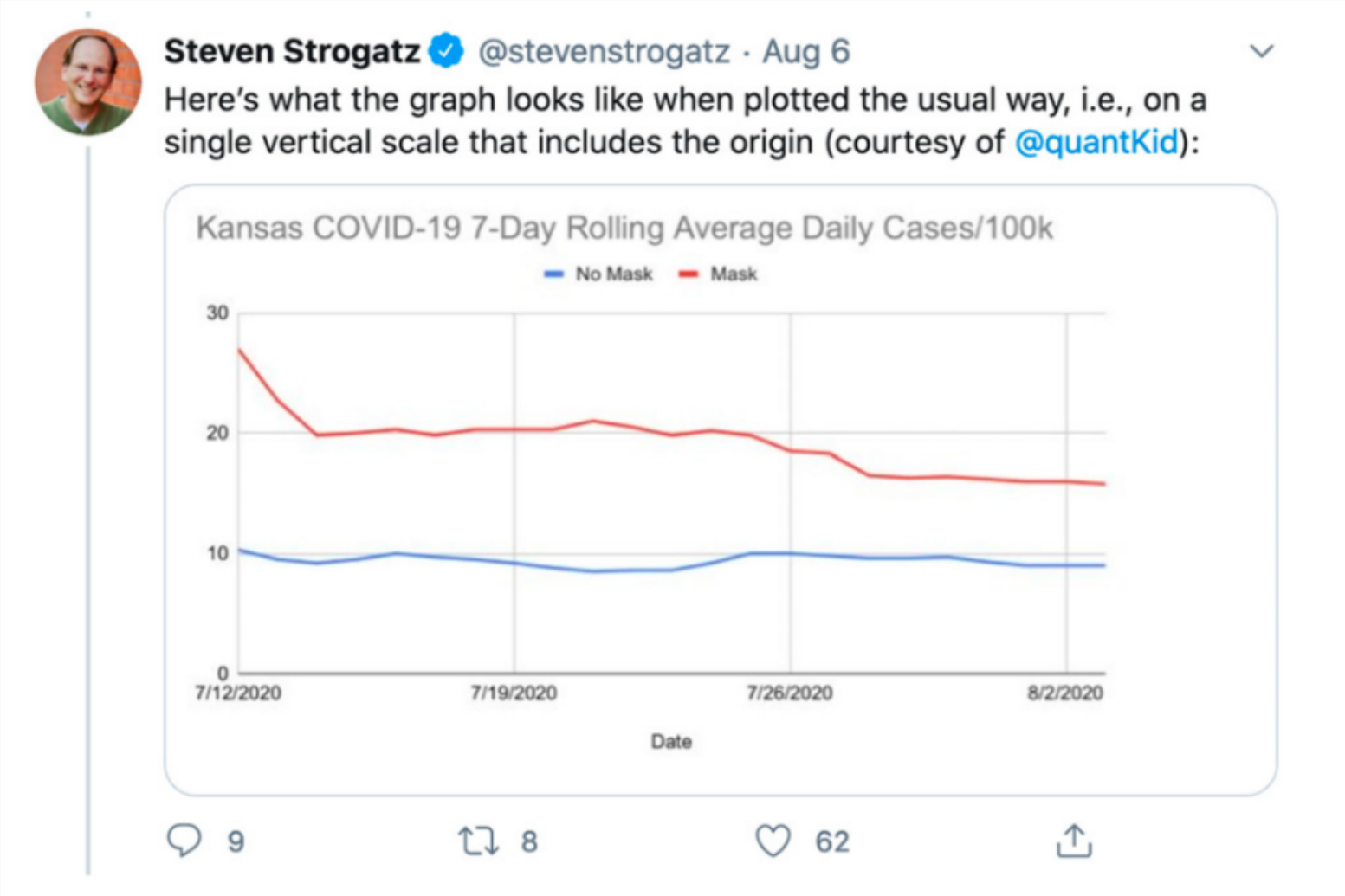
Case 1: Kansas department of health and environment



Case 1: Kansas department of health and environment

- Plot with two axes difficult to read!
- Here: axes are on different scales!

Case 1: Kansas department of health and environment



Case 1: Kansas department of health and environment

- Second plot shows only small decline in countries with mask mandate
- Case level in mask mandate countries higher!

Case 2: Georgia department of public health

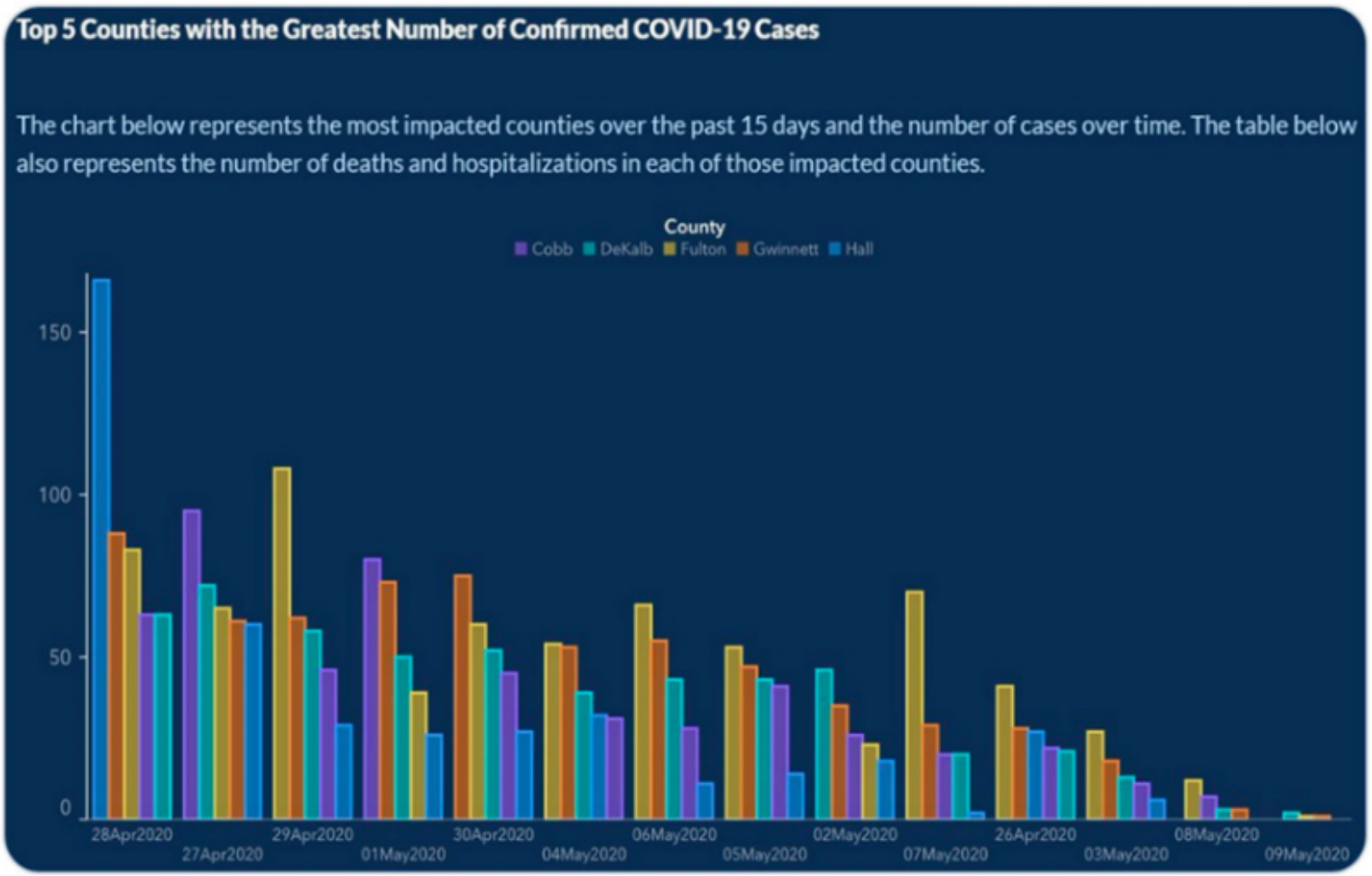


Figure 4: Top 5 counties with the greatest number of COVID-19 cases

Case 2: Georgia Department of Public Health

- Seems like COVID-19 cases have declined over time
- BUT, dates on horizontal axis not in order of time!
- instead, dates ordered such that numbers decrease!

Engledowl and Weiland (2021): Conclusion

Be critical!

- Was data arranged/scaled/modified?
- What conclusion can (not) be drawn?
- What conclusions can be drawn from the plot that do not make much sense?
- What is the intended purpose of the plot?

The Impact of Outliers on Income Inequality

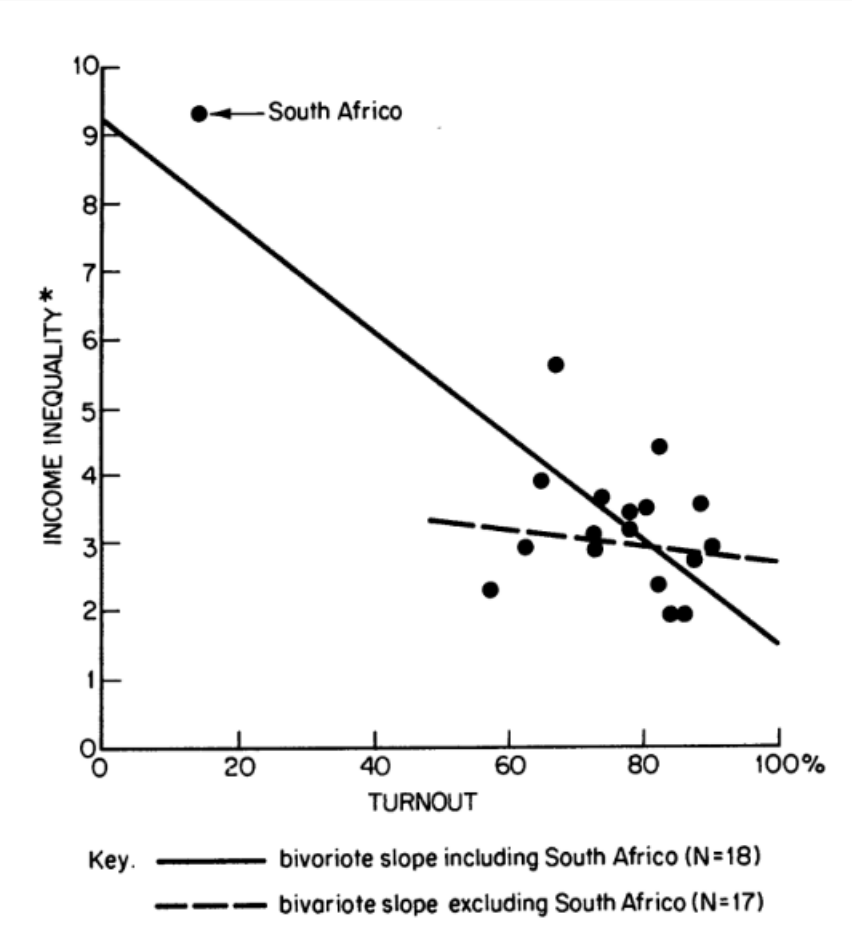
Introduction

- Author: Robert W. Jackman (1980)
- Investigate the impact of outlier observations when investigating the determinants of income inequality

Previous research debate

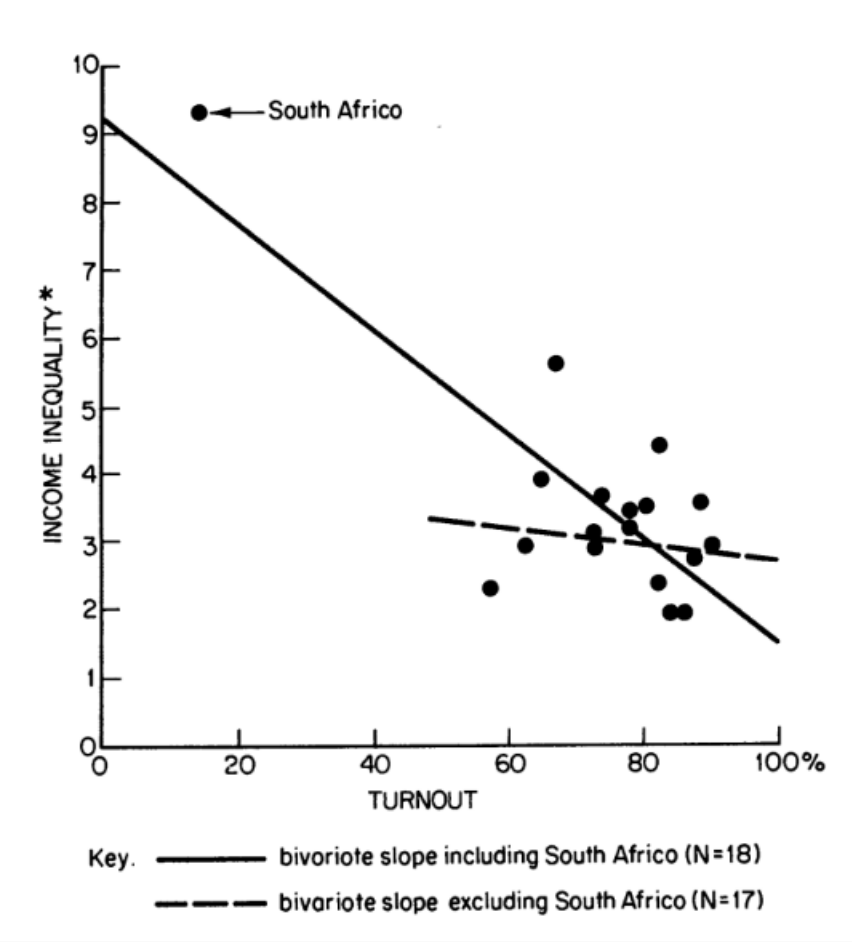
- Proposal: strong presence of socialist parties $\rightarrow$ more redistribution $\rightarrow$ more equality
- Counter-proposal: political participation rather than socialist party strength matters
- Jackman's proposal: both proposals unwarranted

Analysis



- (South Africa) represents outlier
- Suspicion: South Africa has large impact on estimates hence conclusion
- Does relationship still hold when the outlier is excluded ?

Analysis



- Excluding South Africa changes the results significantly.
- Political participation no longer a good predictor of the level of income inequality!
- No evidence is found for either the proposal or the counter-proposal when outliers are accounted for

Recap

We have seen

- ... various examples of subtle data manipulation done by governments,
- ... ways in which graphical data illustration has misguided the public during the COVID pandemic and
- ... the enormous impact one outlier can have on research findings.