The Good, the Bad, and the Ugly: Three Years of Mostly Bad Coronavirus Graphs

Jürgen Symanzik

Utah State University

Logan, UT, USA

symanzik@math.usu.edu

http://www.math.usu.edu/~symanzik

Data Science, Statistics & Visualisation 2022 (DSSV)

Tainan, Taiwan (Virtual)

June 27, 2022

Contents

- Motivation
- “How to Display Data Badly”
- The Bad and the Ugly Ones
- The Good Ones
- Summary & Discussion
- Good and Bad Covid-19 Graphs Collections on the Web
- Sources for Constructing Better Graphs

Motivation

Motivation

- As an expert in statistical graphics, it hurts to see bad graphs in the news, on the web, and produced by students.
- Even worse are bad graphs in science (including those from articles in peer-reviewed journals).
- As the former chair of a task force of the *Statistical Graphics Section of the American Statistical Association (ASA),* we reevaluated the winning posters of the annual ASA poster competition for children from kindergarten to grade 12 – and noticed many bad graphs, even among the winners.
- In the past three years, numerous bad graphs related to Covid-19 were published.

“How to Display Data Badly”

"How to Display Data Badly"

- From: Wainer, H. (1997), Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot, Copernicus/Springer, New York, NY:

- "The aim of good data graphics is to display data accurately and clearly. […]
- Thus, if we wish to display data badly, we have three avenues to follow.
  - A. Don't show much data.
  - B. Show the data inaccurately.
  - C. Obfuscate the data." [i.e., show the data unclearly]

A. Don’t show much data

- Rule 1: Show as little data as possible (minimize the data density).
- Rule 2: Hide what data you do show (minimize the data/ink ratio).

B. Show the data inaccurately

- Rule 3: Ignore the visual metaphor altogether.
- Rule 4: Only order matters.
- Rule 5: Graph data out of context.

C. Obfuscate the data

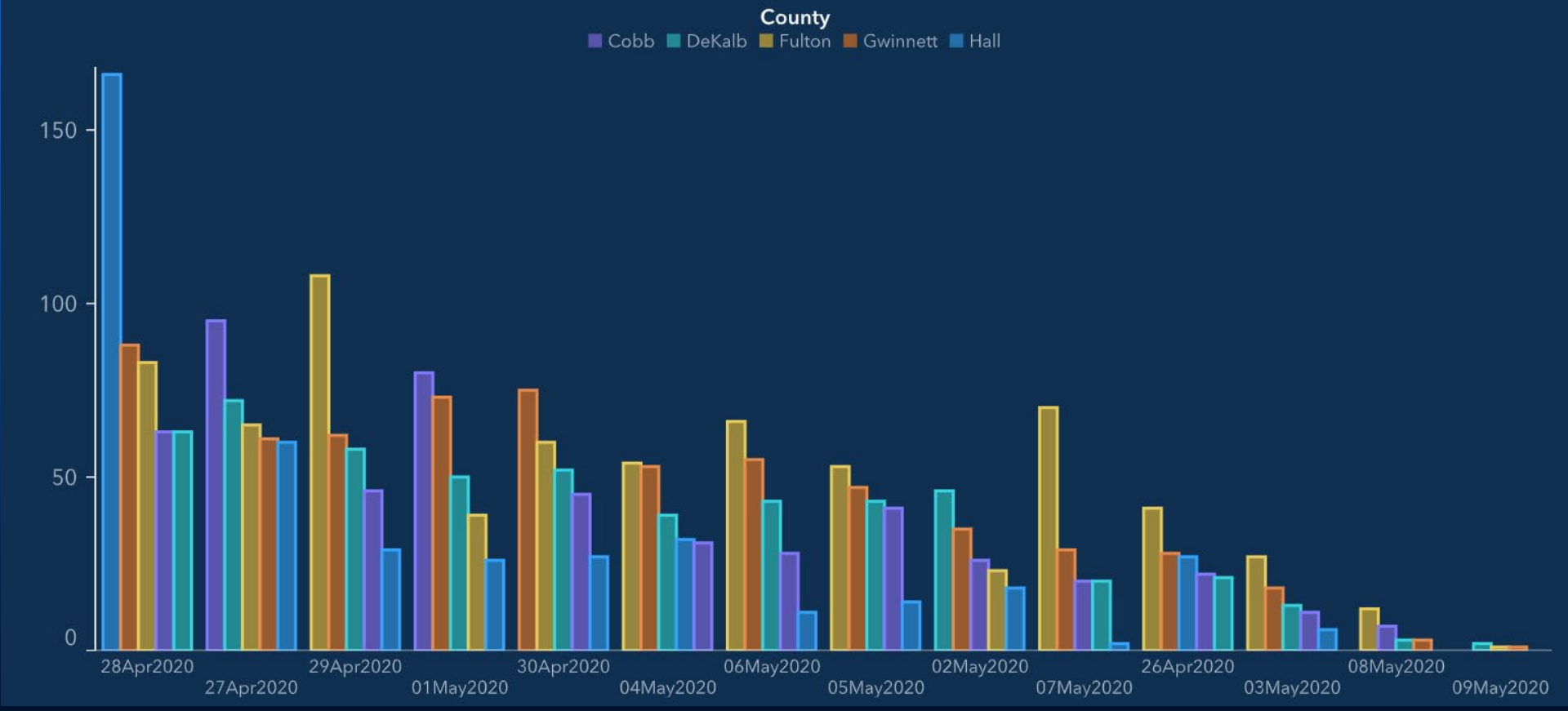
- Rule 6: Change scales in mid-axis.
- Rule 7: Emphasize the trivial (ignore the important).
- Rule 8: Jiggle the baseline.
- Rule 9: Alabama first!
- Rule 10: Label: (a) illegibly, (b) incompletely, (c) incorrectly, and (d) ambiguously.
- Rule 11: More is murkier: (a) more decimal places and (b) more dimensions.
- Rule 12: If it has been done well in the past, think of a new way to do it.

The Bad and the Ugly Ones

Georgia's Covid-19 Cases – The Original

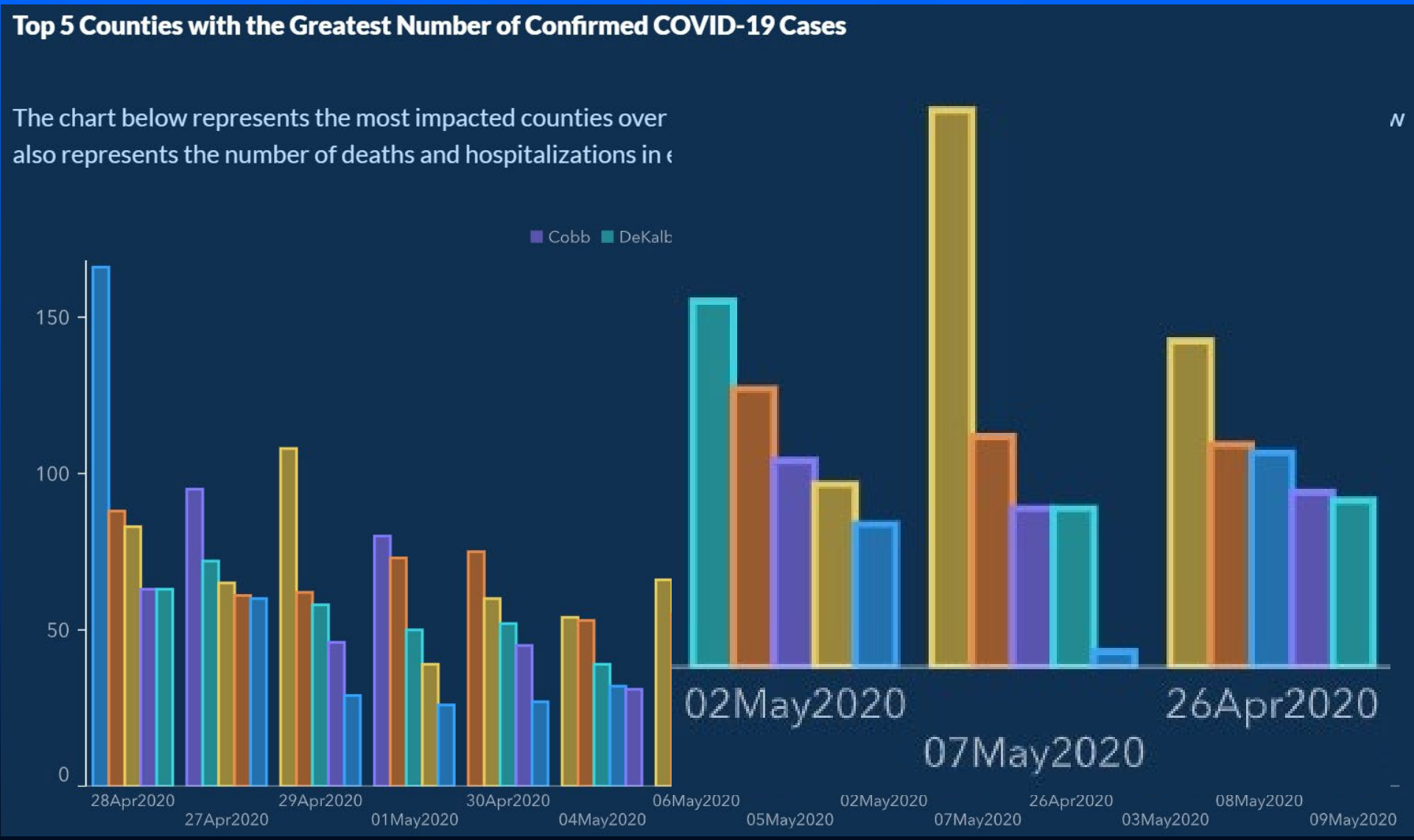
Top 5 Counties with the Greatest Number of Confirmed COVID-19 Cases

The chart below represents the most impacted counties over the past 15 days and the number of cases over time. The table below also represents the number of deaths and hospitalizations in each of those impacted counties.



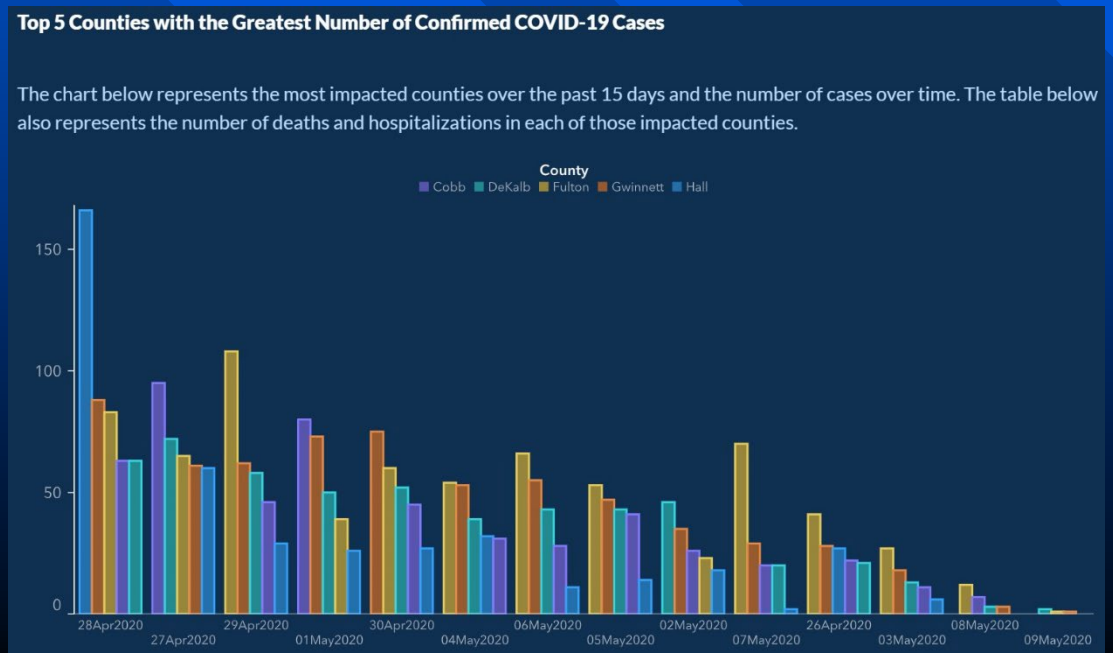
Georgia Department of Public Health (Graph no longer available from source)

Georgia's Covid-19 Cases – Close-Up

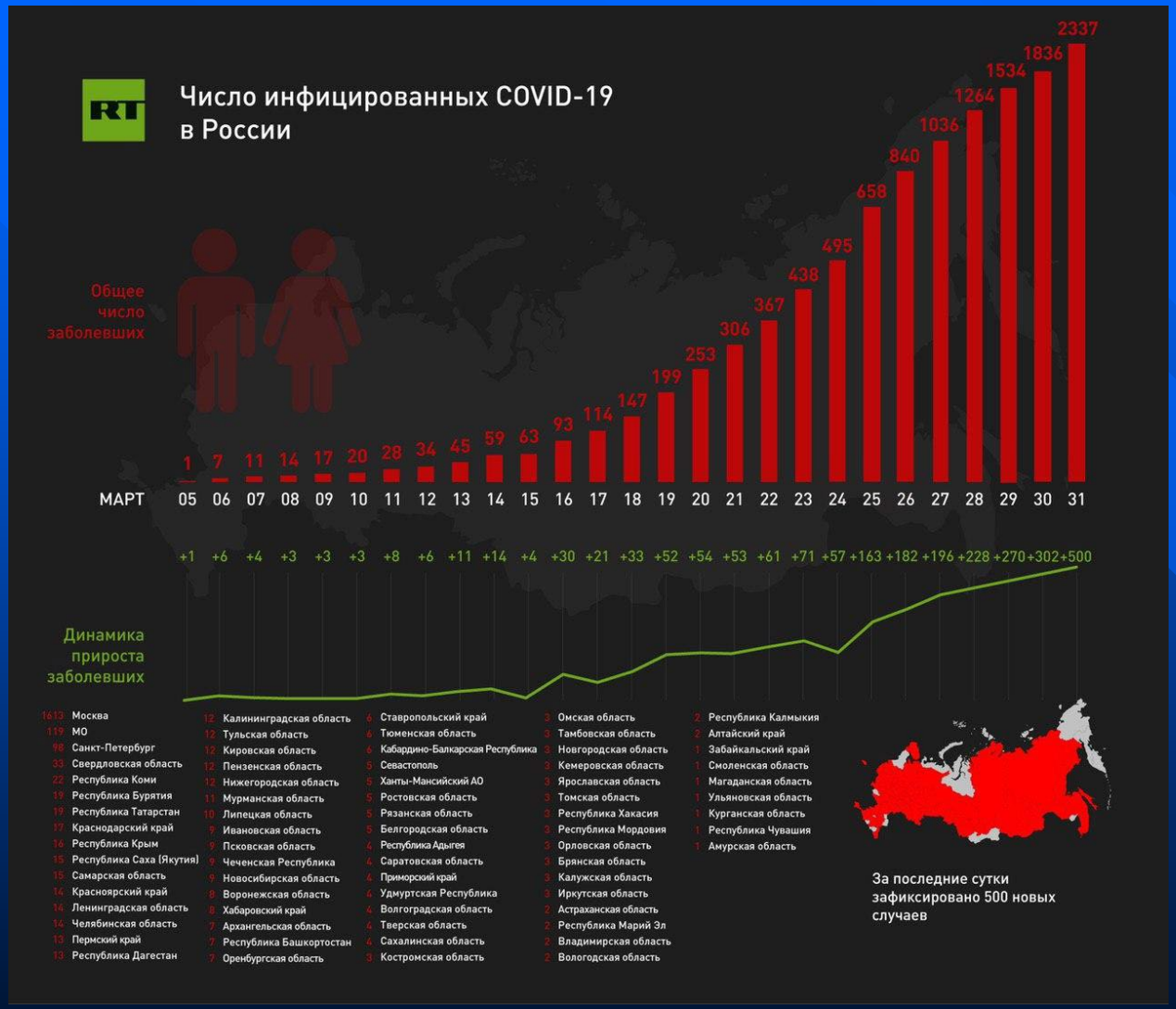


Georgia's Covid-19 Cases – Bad Graph Rules Followed

- Rule 3: Ignore the visual metaphor altogether: Dates arbitrarily arranged on horizontal axis.
- Rule 7: Emphasize the trivial (ignore the important): Different ordering of counties in each block of five, making it hard to compare / follow a county on various dates.



Russian Covid-19 Cases – The Original



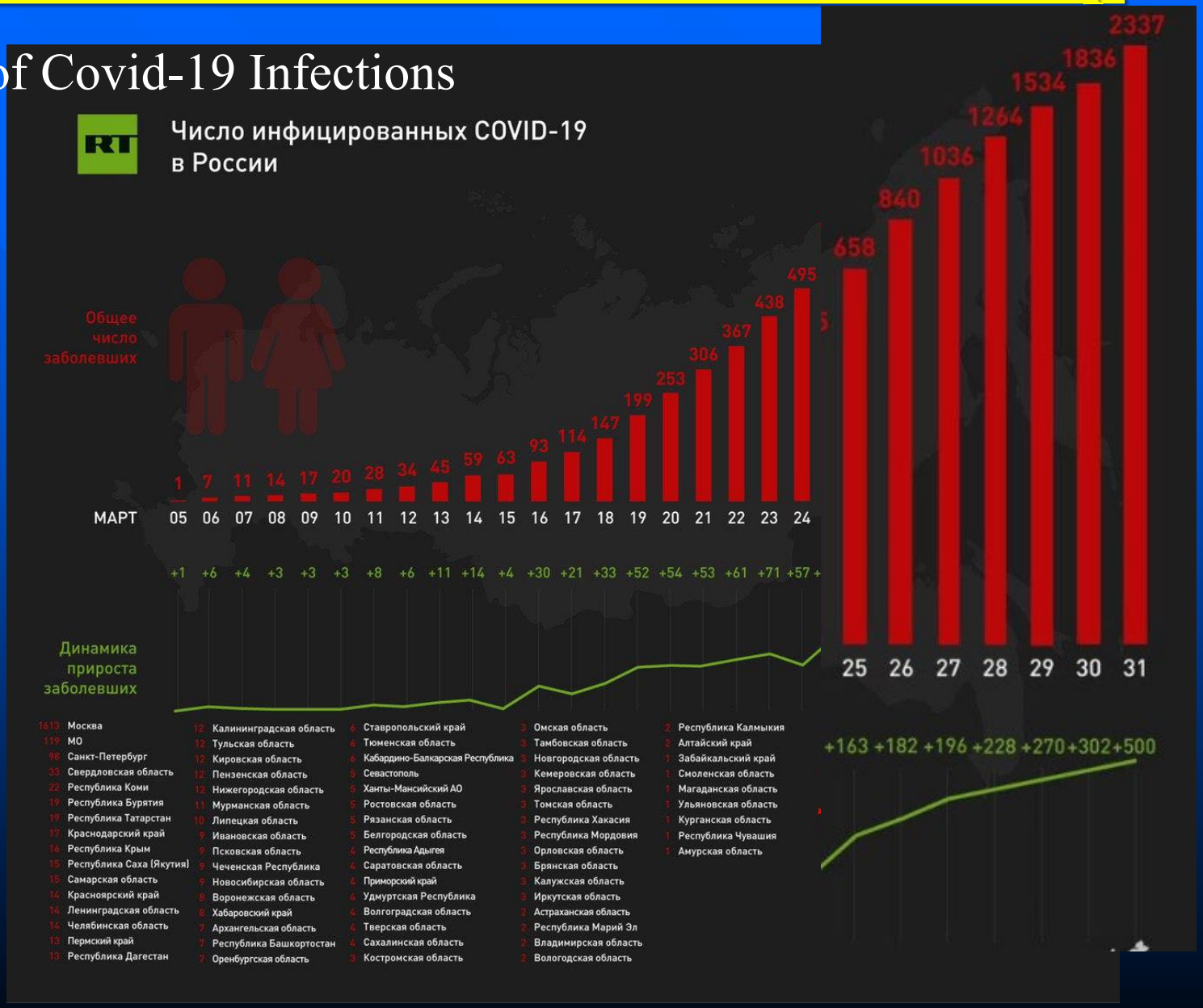
RT in Russian (Graph no longer available from source)

Russian Covid-19 Cases – Close-Up

Number of Covid-19 Infections in Russia

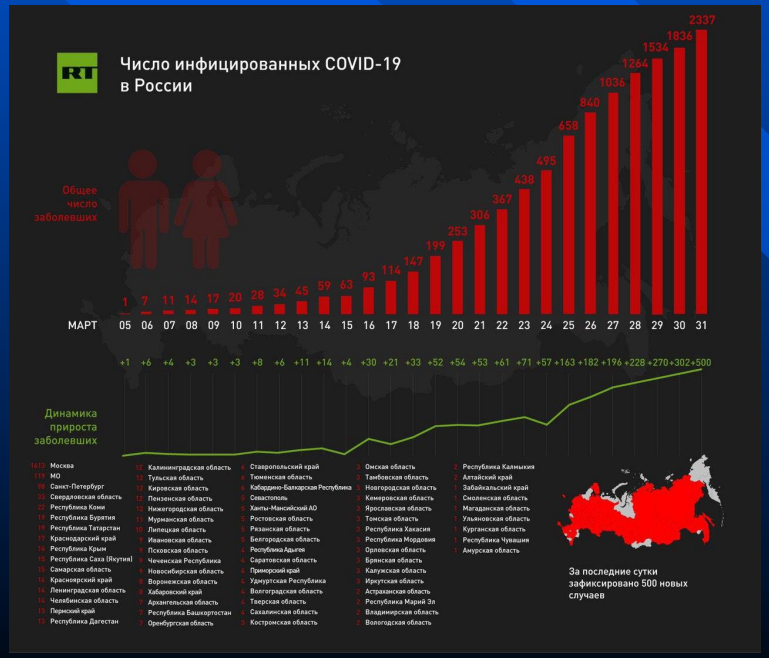
General Number of Sick People

Dynamic of Growth of Sick People

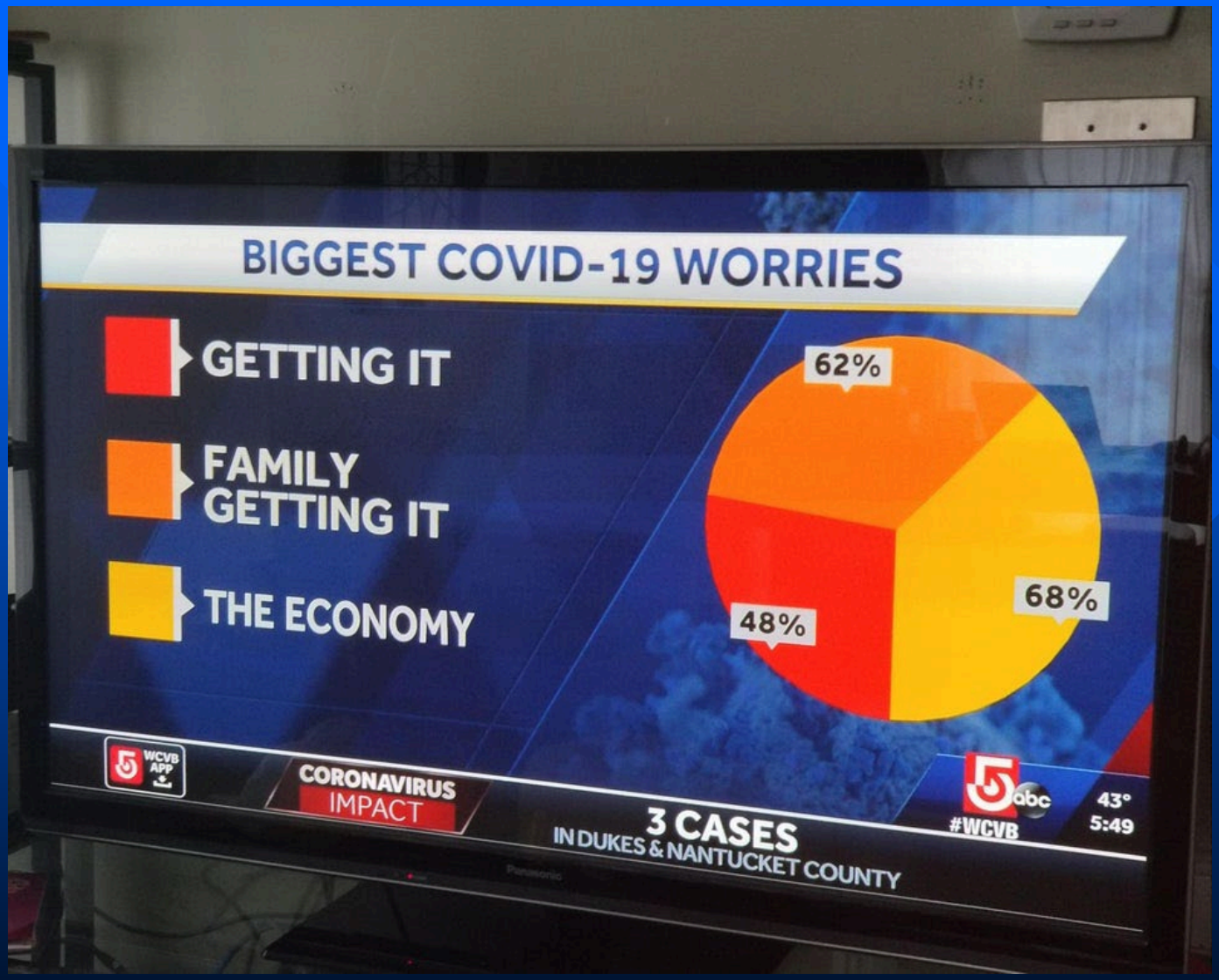


Russian Covid-19 Cases – Bad Graph Rules Followed

- Rule 4: Only order matters: Heights of red bars are not proportional to the numbers shown. Same for the green line.
- Rule 6: Change scales in mid-axis: In fact, no scale for the vertical axes are shown at all. Could these be log-scales?



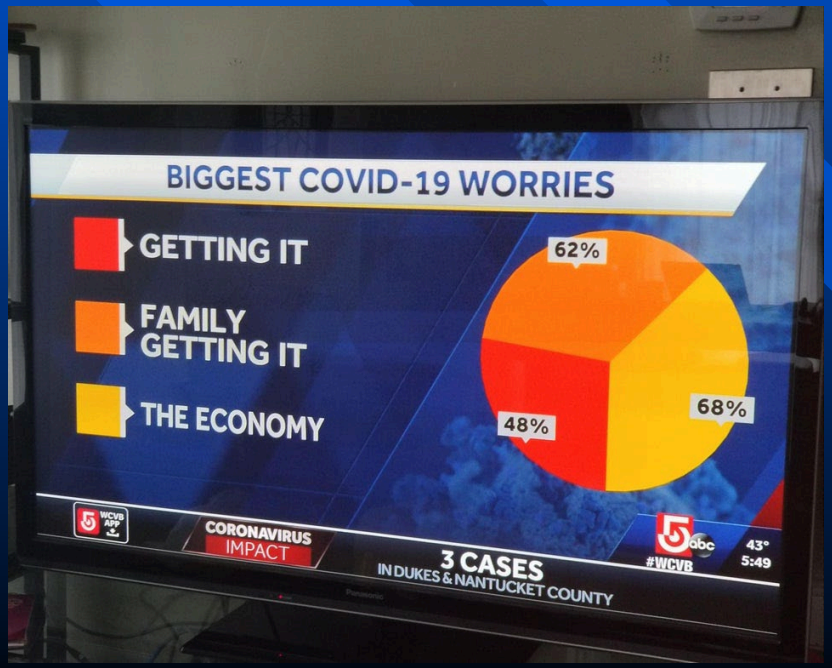
A Pie Chart out of 178% – The Original



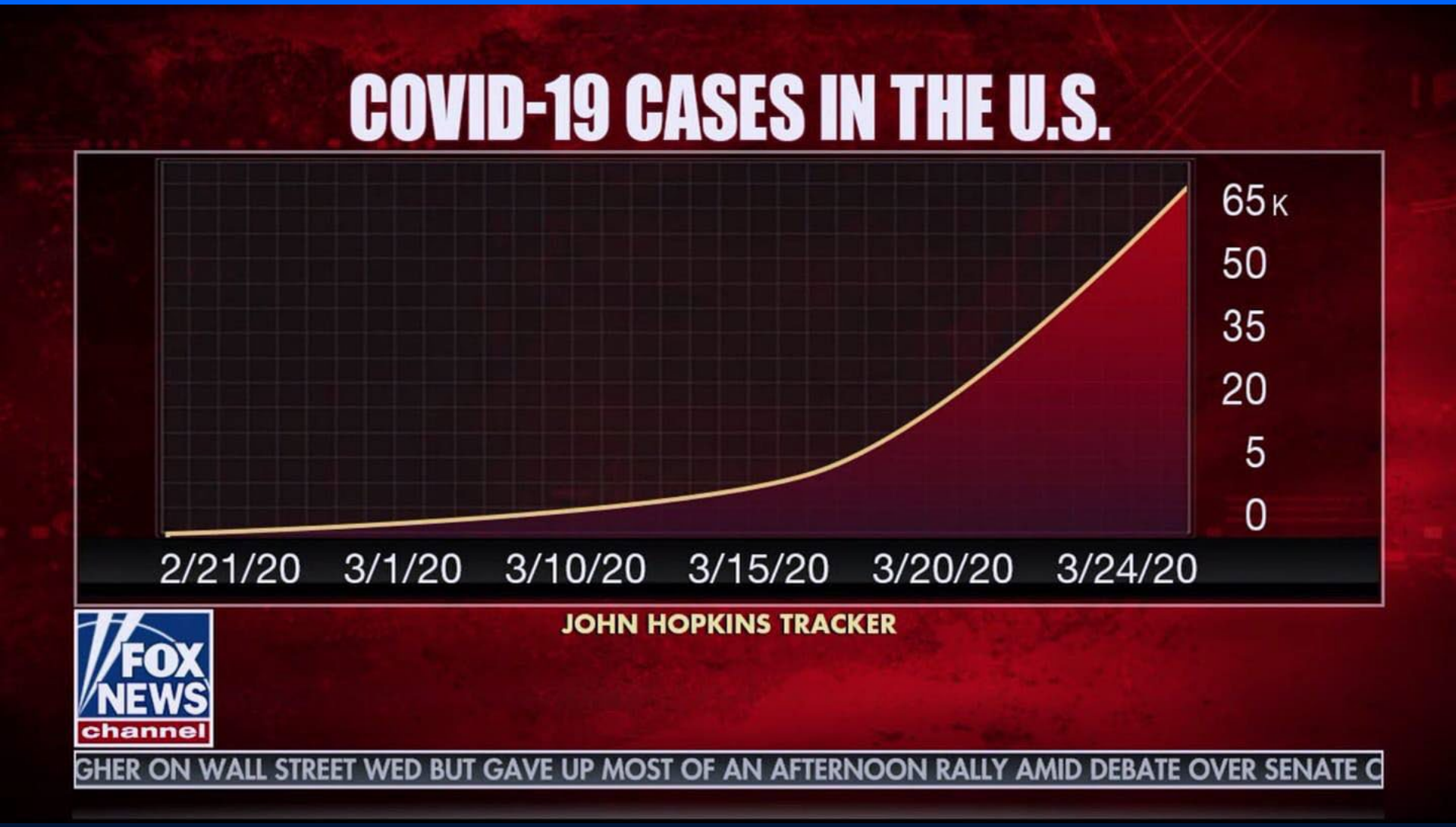
WCVB Channel 5 Boston (Graph no longer available from source)

A Pie Chart out of 178% – Bad Graph Rules Followed

- Rule 1: Show as little data as possible (minimize the data density): Only three possible outcomes are shown.
- Rule 12: If it has been done well in the past, think of a new way to do it: A pie chart is the wrong graph type here; why not a regular bar chart or Cleveland-style dot plot?



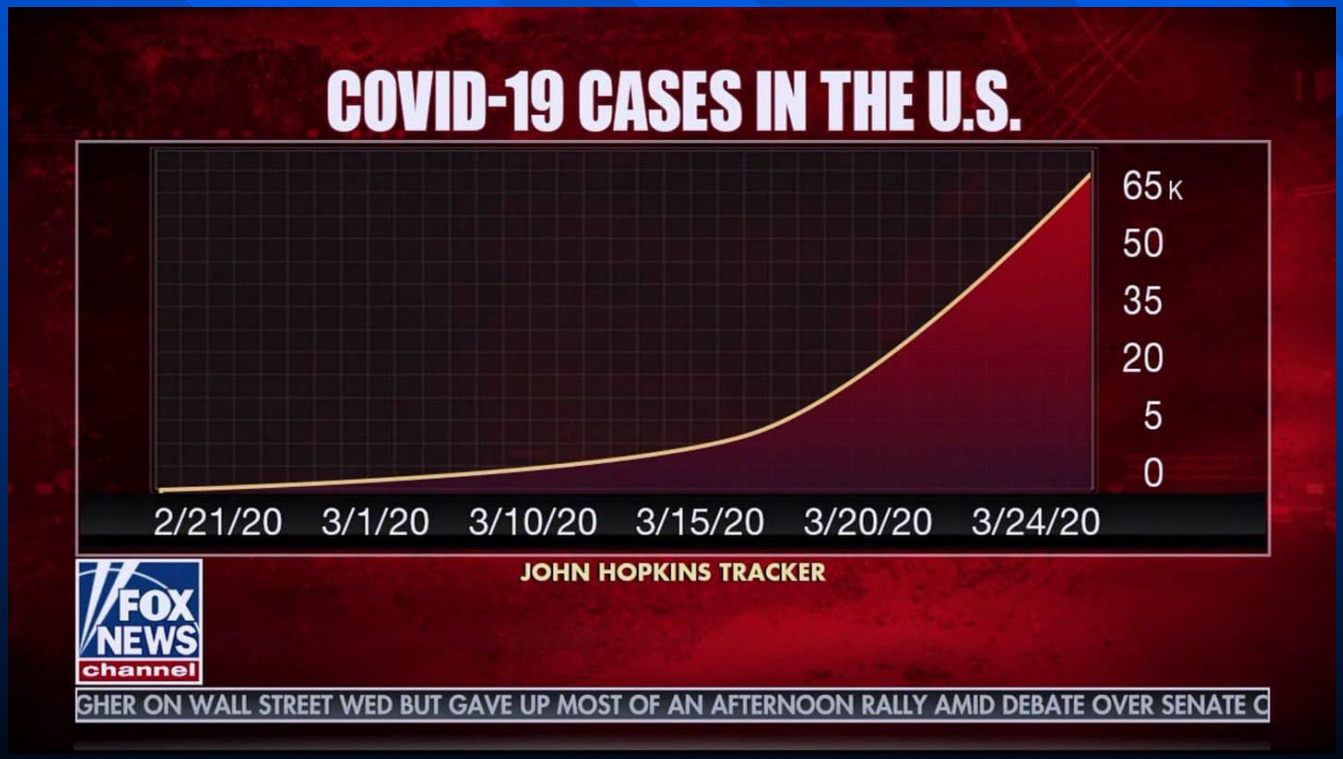
Covid-19 Cases in the US – The Original



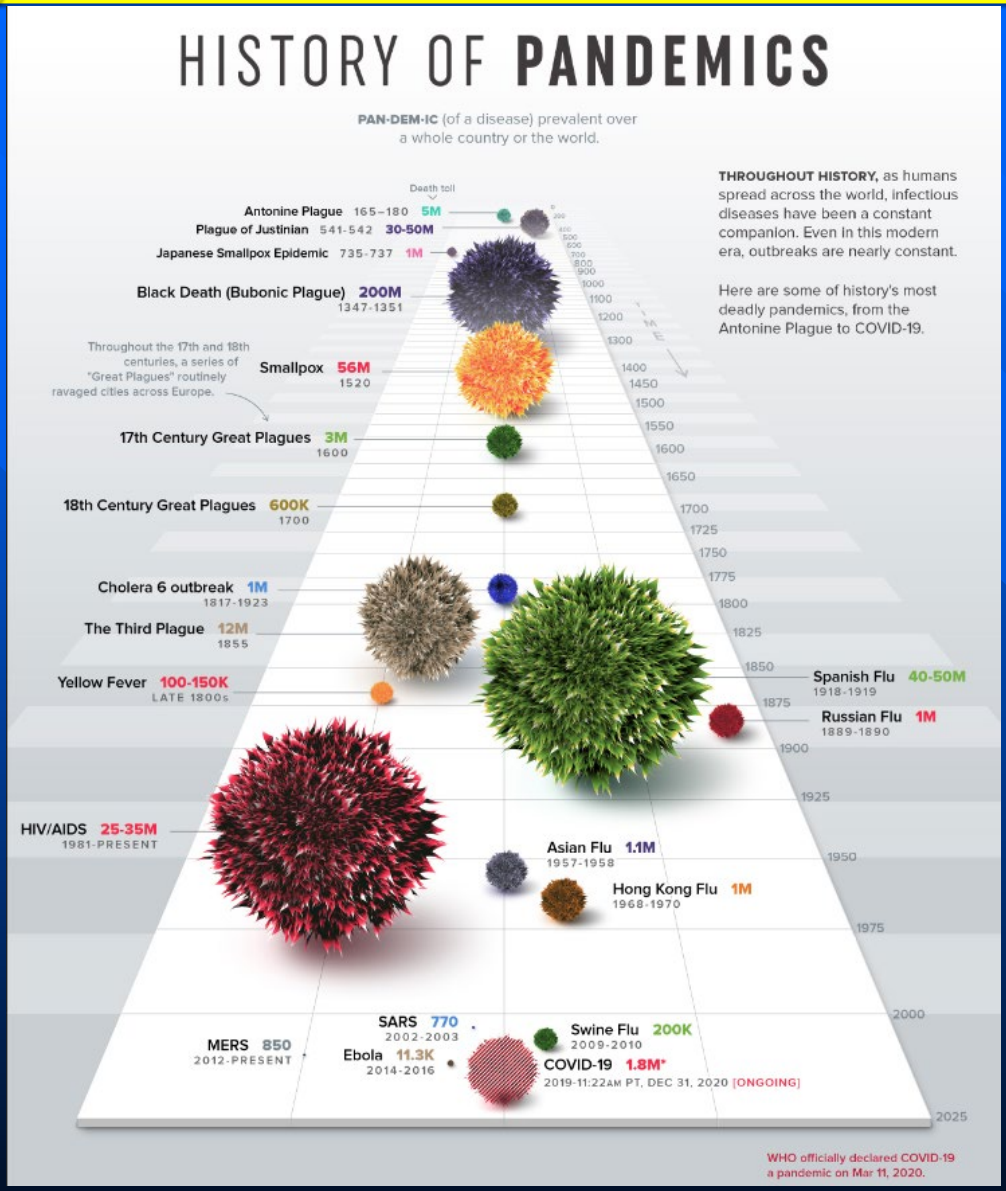
Fox News Channel (Graph no longer available from source)

Covid-19 Cases in the US – Bad Graph Rules Followed

- Rule 6: Change scales in mid-axis: Horizontal axis labels are placed after 9, 9, 5, 5, and 4 days.
- Rule 6: Change scales in mid-axis: Vertical axis labels are placed after 5K and then after every 15K.



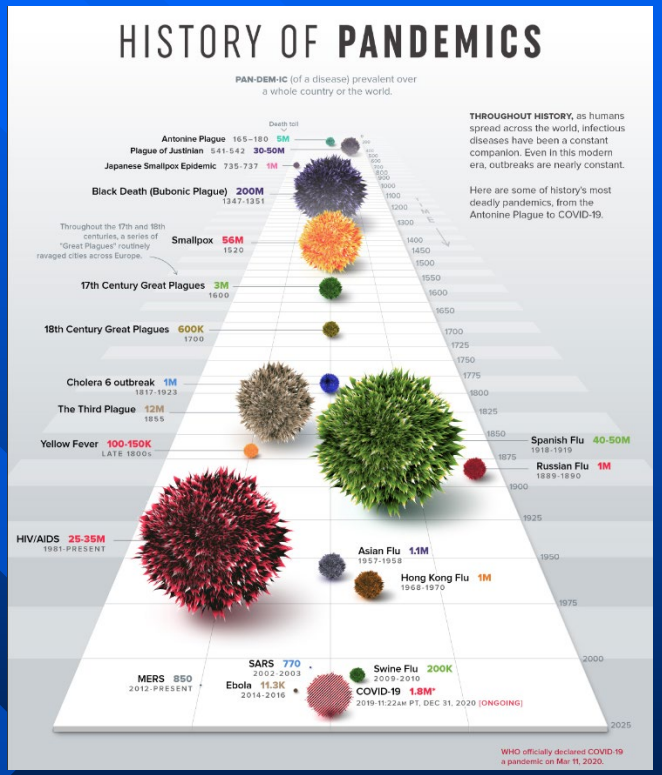
History of Pandemics – The Original



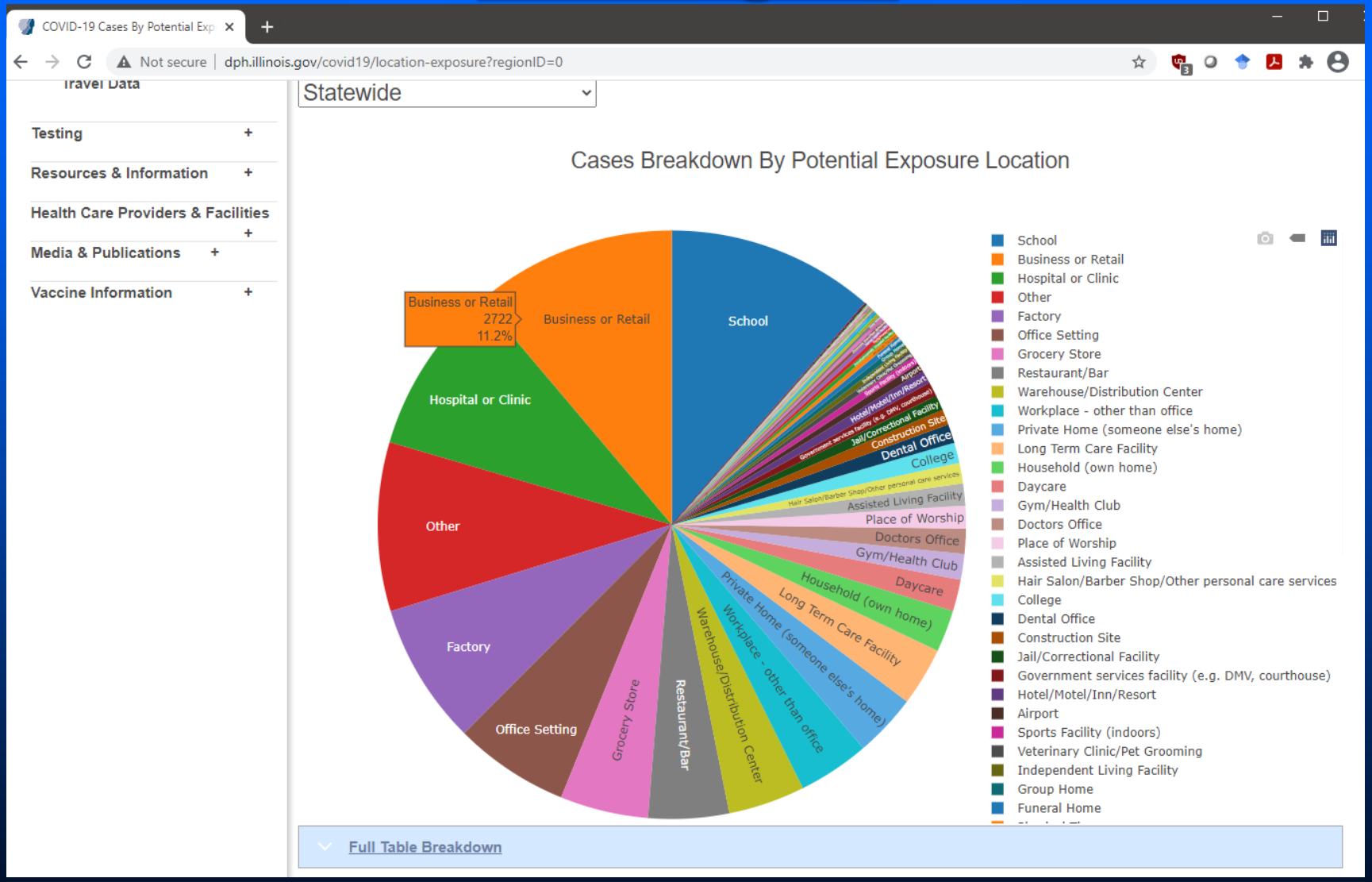
https://www.visualcapitalist.com/history-of-pandemics-deadliest/

History of Pandemics – Bad Graphs Rules Followed

- Rule 3: Ignore the visual metaphor altogether: Size of viruses not related to number of deaths.
- Rule 6: Change scales in mid-axis: Vertical time axis is condensed, due to the 3D-perspective view.
- Rule 10: Label: (a) illegibly, (b) incompletely, (c) incorrectly, and (d) ambiguously: “Death toll” label almost invisible.
- Rule 11: More is murkier: (a) more decimal places and (b) more dimensions: 3D representation makes some of the viruses/numbers look even bigger.



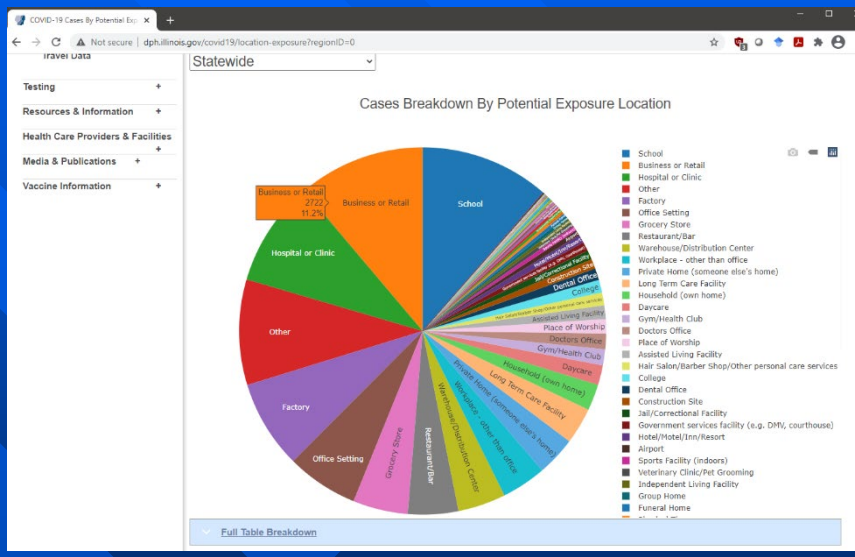
Cases by Potential Exposure Location – The Original



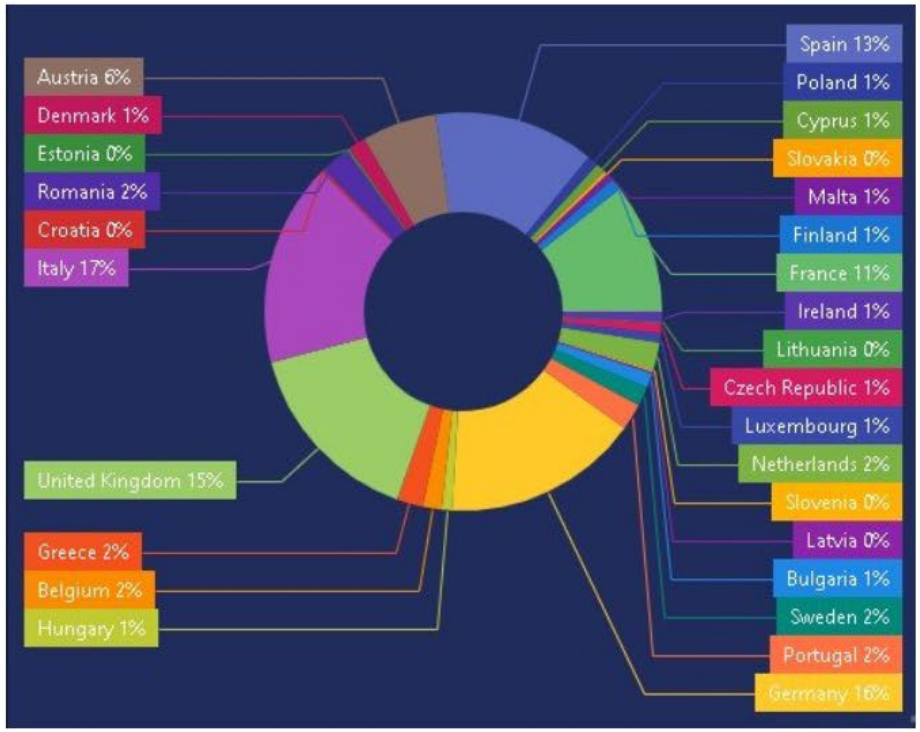
http://www.dph.illinois.gov/covid19/location-exposure?regionID=0

Cases by Potential Exposure Location – Bad Graph Rules Followed

- Rule 2: Hide what data you do show (minimize the data/ink ratio): Some of the categories cannot be identified at all.
- Rule 9: Alabama first!: Segments in the pie chart are not arranged in any meaningful order at all.
- Rule 10: Label: (a) illegibly, (b) incompletely, (c) incorrectly, and (d) ambiguously: Many of the labels are too small to read.



Media Coverage in Europe – The Original



Pie graph shows us the percentage of **European countries** in relation to the Coronavirus. How much were these countries mentioned with the COVID-19 **in the media**?

Top five countries:

Italy 17%

Germany 16%

England 15%

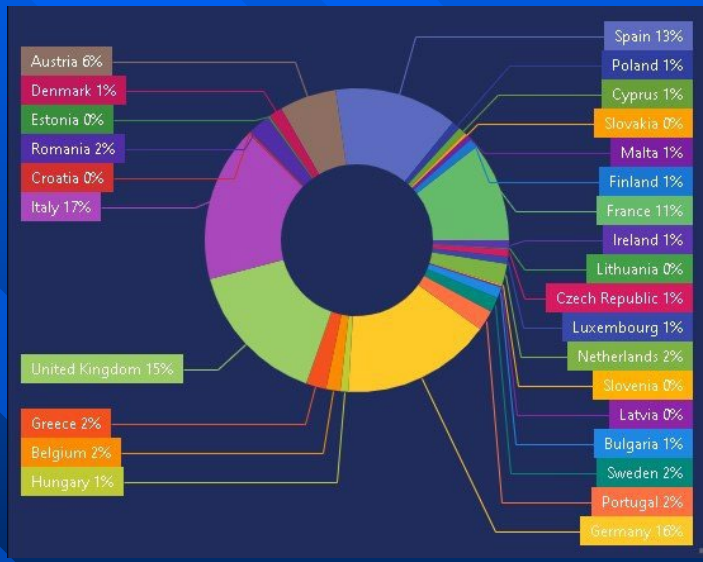
Spain 13%

France 11%

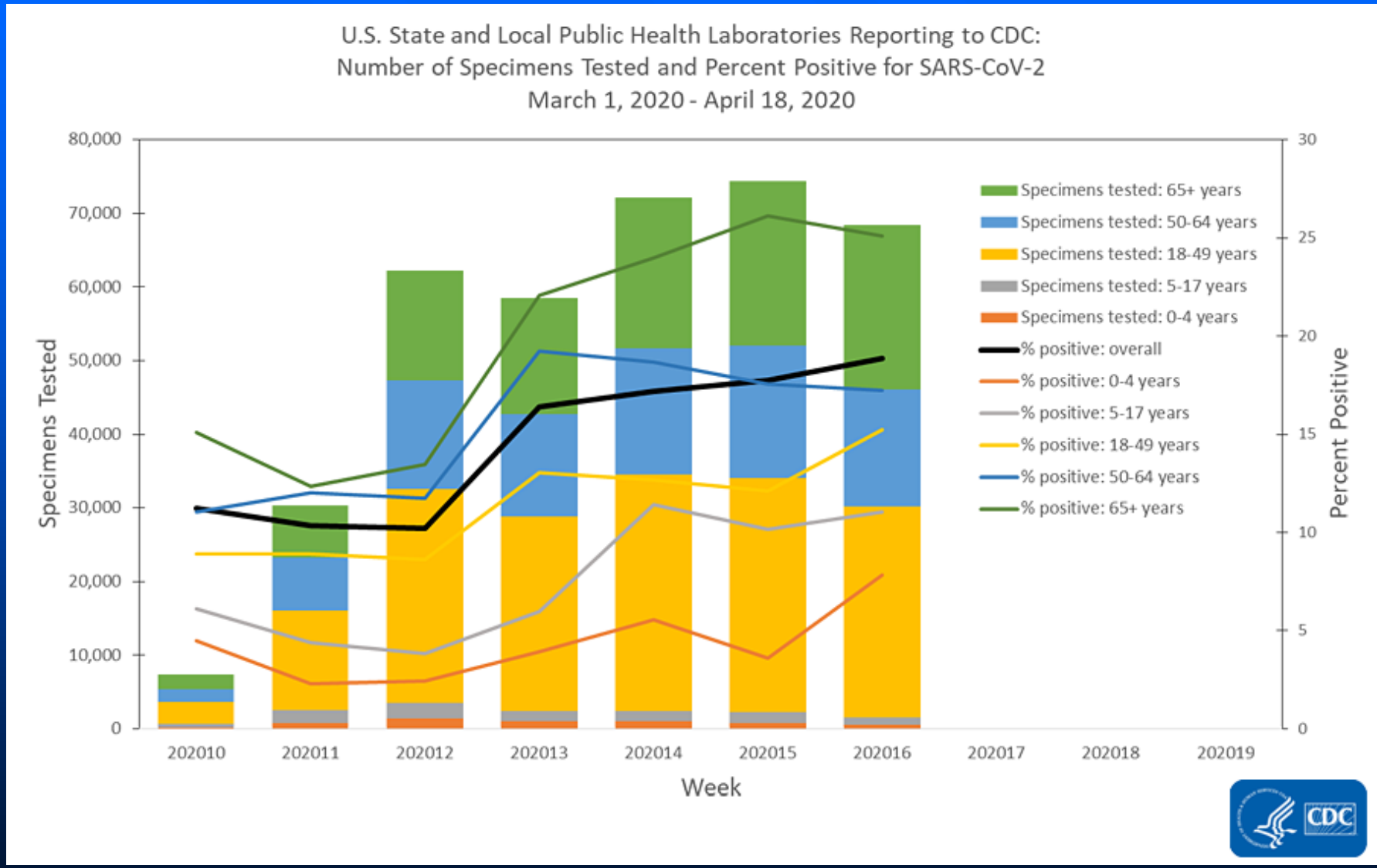
https://www.sail-labs.com/2020/04/15/coronavirus-covid-19-report-media-mining-client-for-media-monitoring-of-pandemic-outbreak/

Media Coverage in Europe – Bad Graph Rules Followed

- Rule 9: Alabama first!: Segments in the donut chart are not arranged in any meaningful order at all.
- Rule 10: Label: (a) illegibly, (b) incompletely, (c) incorrectly, and (d) ambiguously: Russia and Switzerland (and others) are not shown; these seem to be countries of the "European Union" (and not "European Countries").
- Rule 12: If it has been done well in the past, think of a new way to do it: Why showing 0% areas (i.e., countries) in the donut chart?



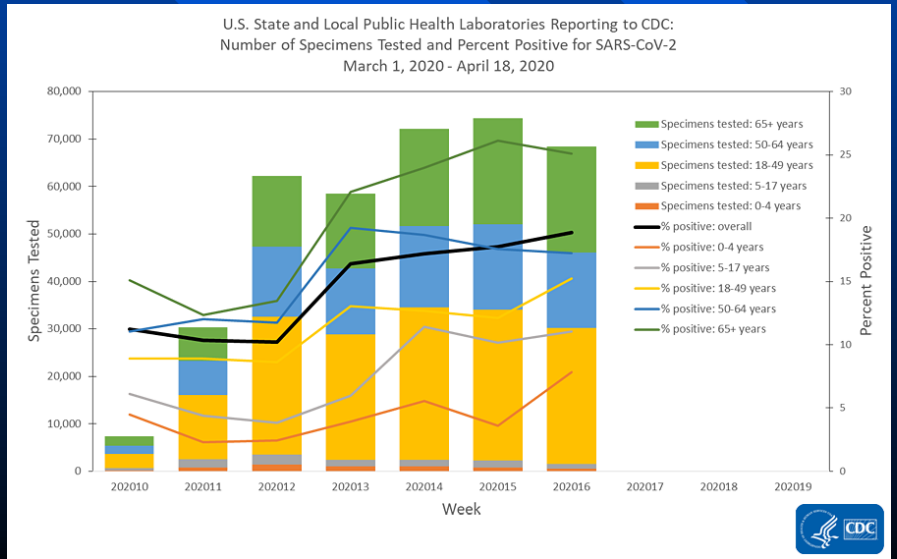
CDC Public Health Labs – The Original



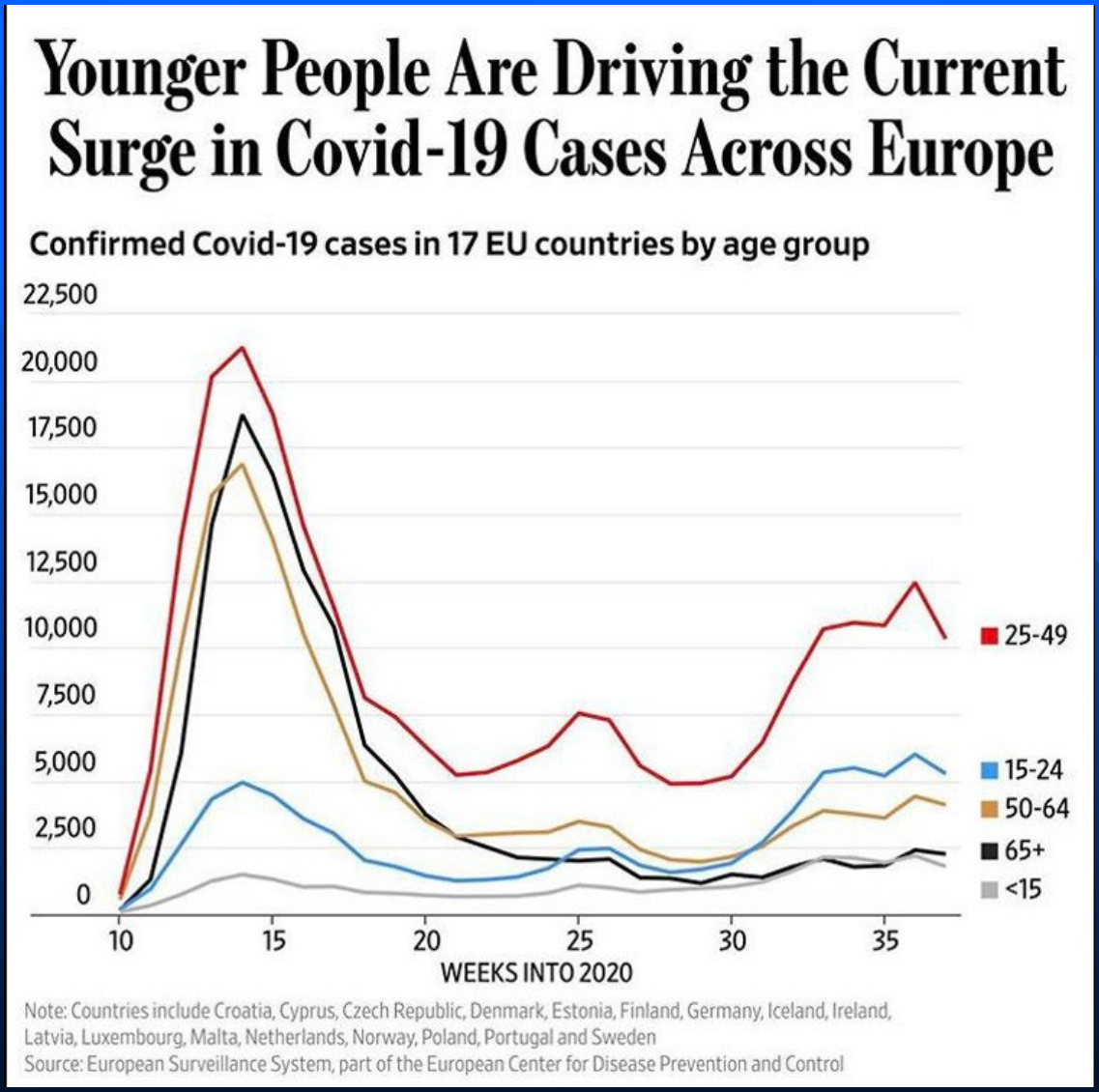
https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/past-reports/04242020.html

CDC Public Health Labs – Bad Graph Rules Followed

- Rule 8: Jiggle the baseline: The uneven (jiggled) base line makes it difficult to see trends for the 65+ (green), 50-64 (blue), and 18-49 (yellow) age groups.
- Rule 10: Label: (a) illegibly, (b) incompletely, (c) incorrectly, and (d) ambiguously: Instead of using labels such as "202010", separate month and week and label as "#10/2020" or "2020-10" for example.



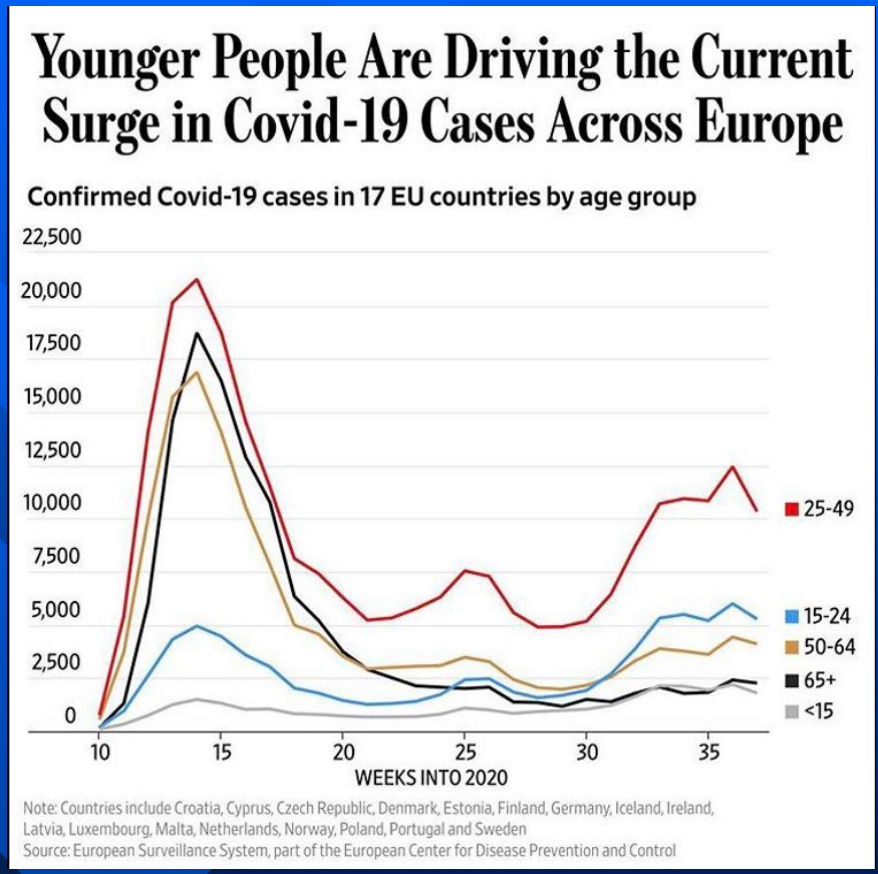
Younger People Surge – The Original



https://www.wsj.com/articles/europes-covid-fight-has-a-new-target-the-youth-11601553600

Younger People Surge – Bad Graph Rules Followed

- Rule 7: Emphasize the trivial (ignore the important): The 25-49 year category spans 25 years; no surprise that this contains more cases than the age groups that only span 10 or 15 years.



European Comparison – The Original

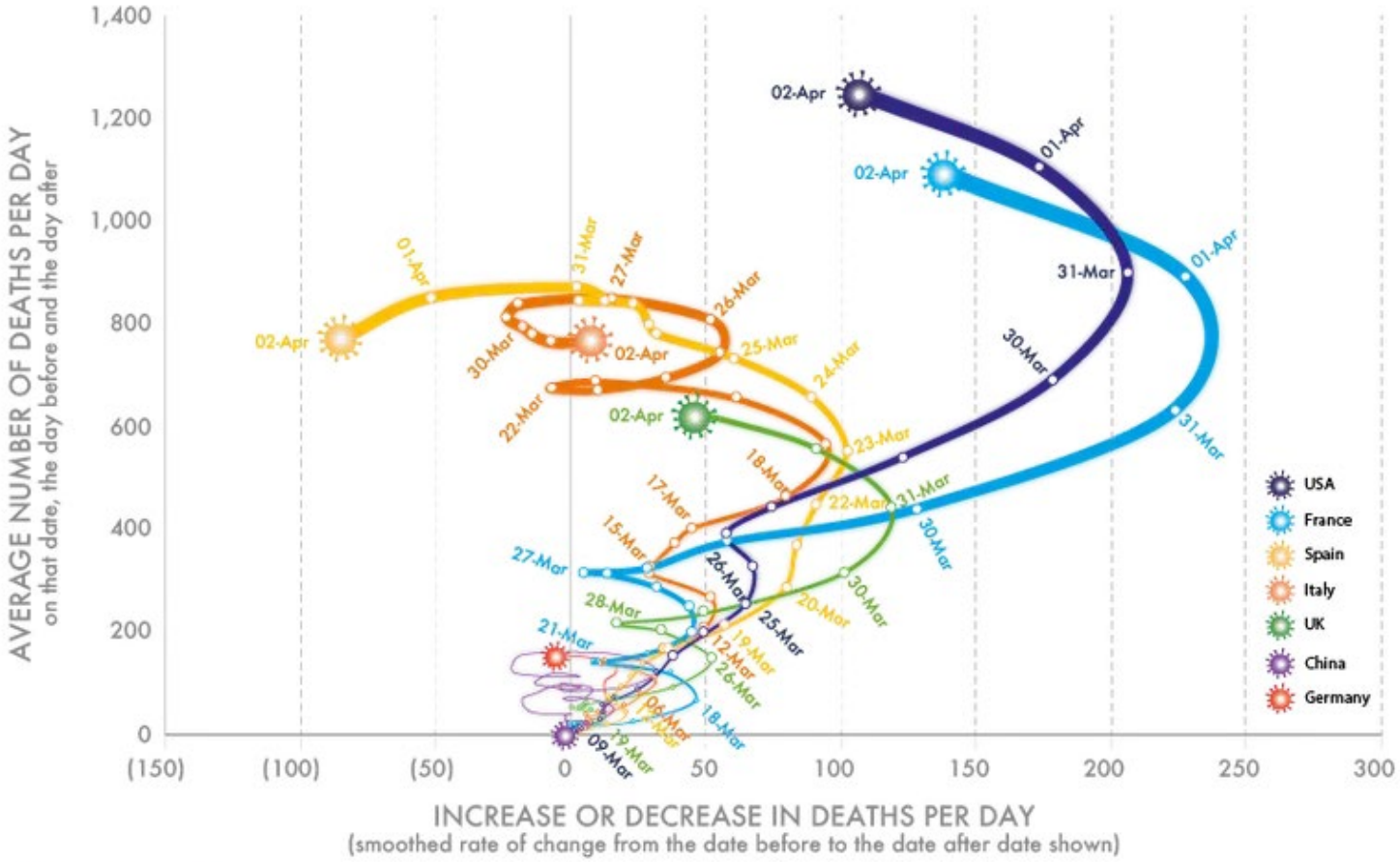
https://thefederalist.com/2020/10/29/these-12-graphs-show-mask-mandates-do-nothing-to-stop-covid/

European Comparison – Bad Graph Rules Followed

- Rule 5: Graph data out of context: Why these countries, but omitting the Netherlands, Russia, and Switzerland for example?
- Rule 5: Graph data out of context: No further context information on lockdowns, school & business closures, contact tracing, etc. provided.
- Rule 10: Label: (a) illegibly, (b) incompletely, (c) incorrectly, and (d) ambiguously: Impossible to identify which line belongs to which country.

Mortality due to Covid-19 – The Original (1)

Mortality in seven countries attributed to COVID-19 (January 23 to April 2, 2020).

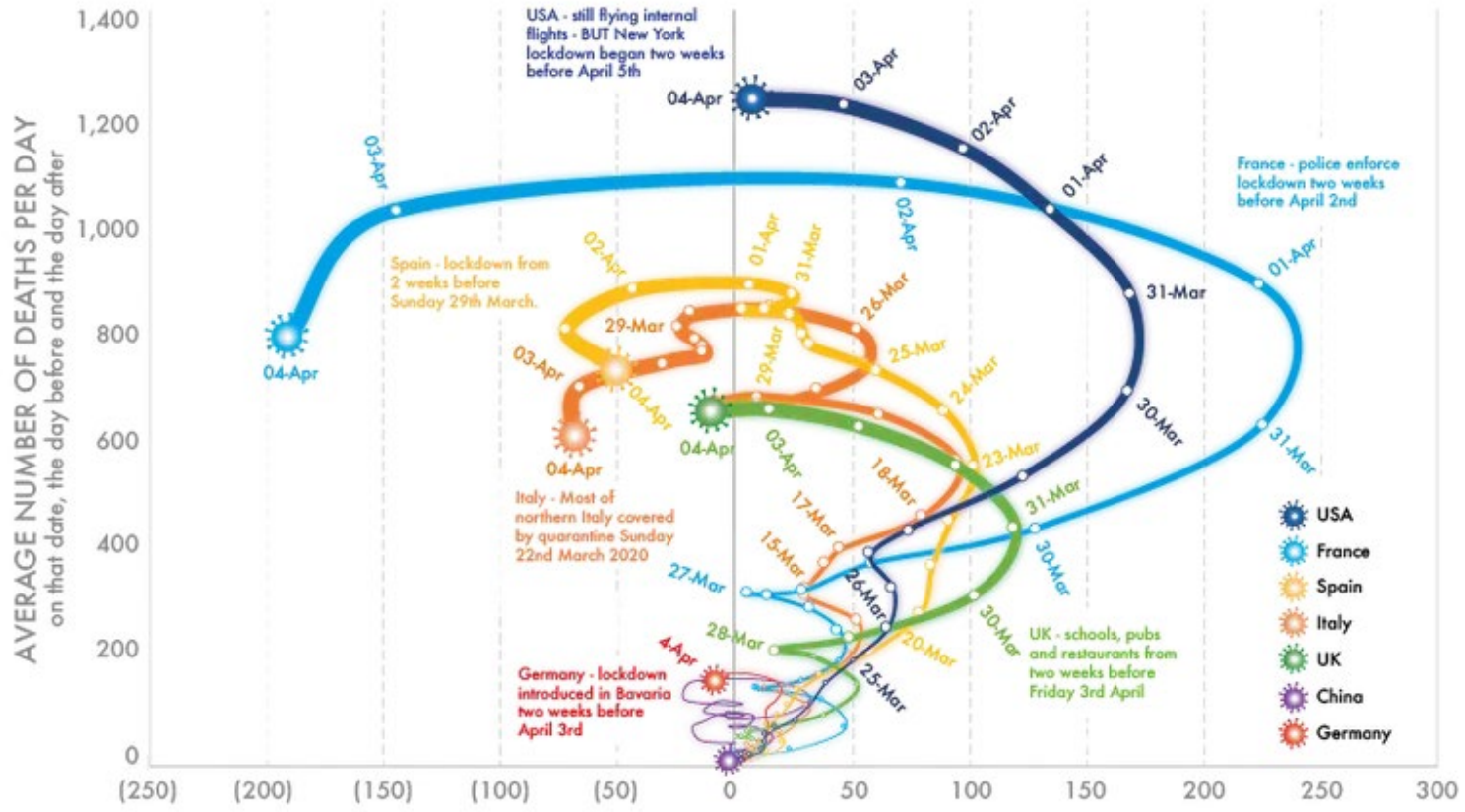


Danny Dorling/Kirsten McClure, Author provided

https://theconversation.com/three-graphs-that-show-a-global-slowdown-in-covid-19-deaths-135756

Mortality due to Covid-19 – The Original (2)

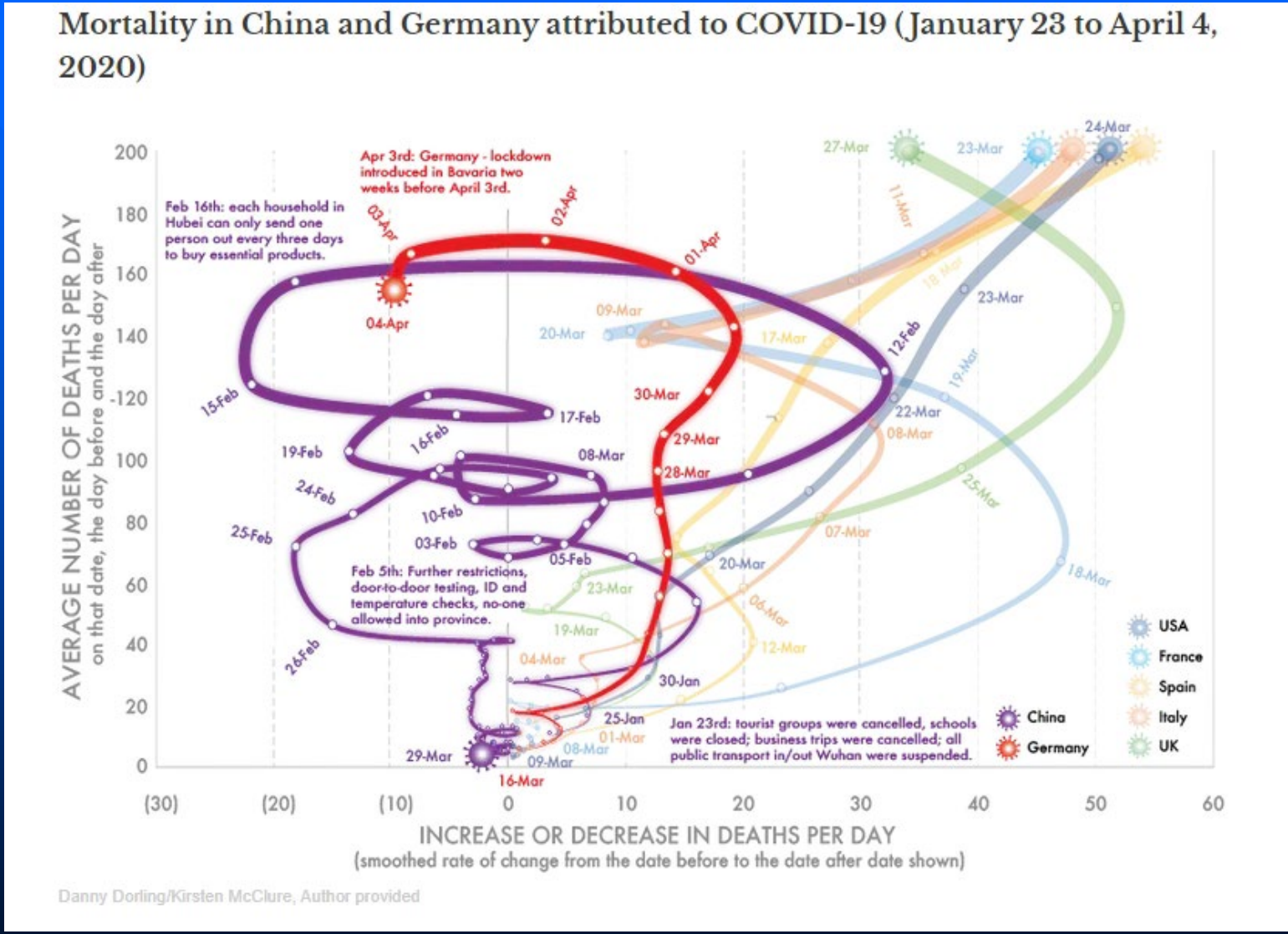
Mortality in seven countries attributed to COVID-19 (January 23 to April 4, 2020).



Danny Dorling/Kirsten McClure, Author provided

https://theconversation.com/three-graphs-that-show-a-global-slowdown-in-covid-19-deaths-135756

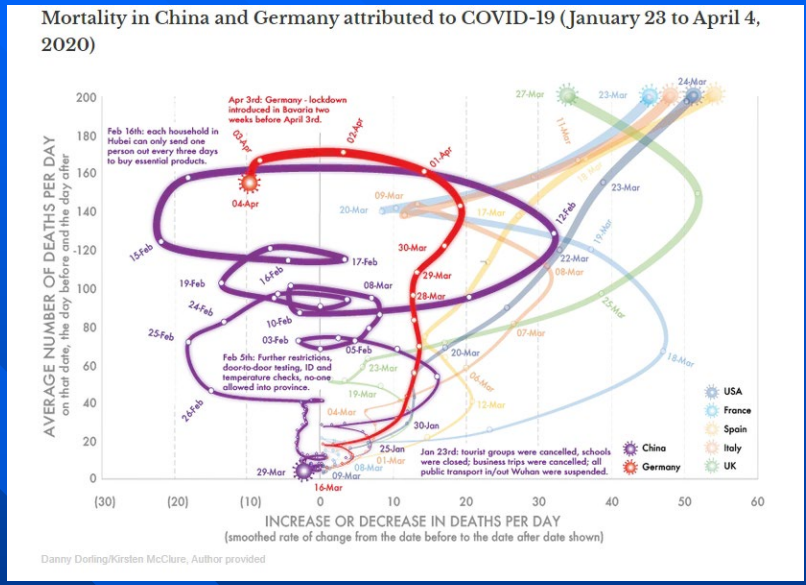
Mortality due to Covid-19 – The Original (3)



https://theconversation.com/three-graphs-that-show-a-global-slowdown-in-covid-19-deaths-135756

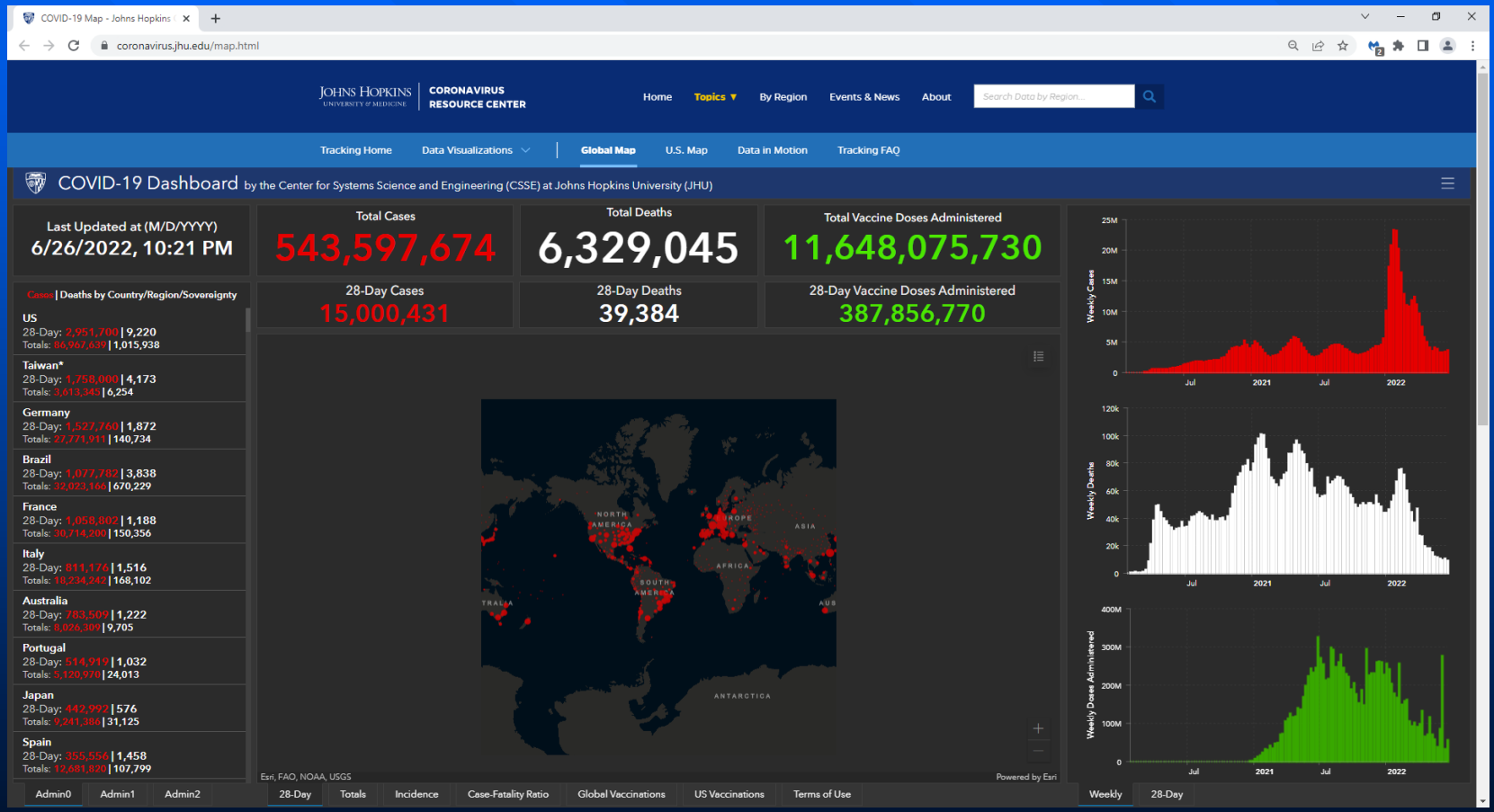
Mortality due to Covid-19 – Bad Graph Rules Followed

- Rule 2: Hide what data you do show (minimize the data/ink ratio): There is a lot of overplotting in the lower part of the graph
- Rule 3: Ignore the visual metaphor altogether: Why changing the line thickness? Or does this represent a fourth variable?
- Rule 5: Graph data out of context: Why these seven countries, but leaving out others with high numbers of cases in March (e.g., the Netherlands, Belgium, Iran)?
- Rule 10: Label: (a) illegibly, (b) incompletely, (c) incorrectly, and (d) ambiguously: Many labels difficult to read.
- Rule 12: If it has been done well in the past, think of a new way to do it: The graph tries to display three (or four?) variables (date, increase/decrease per day, average number of deaths per day) in a two-dimensional layout.



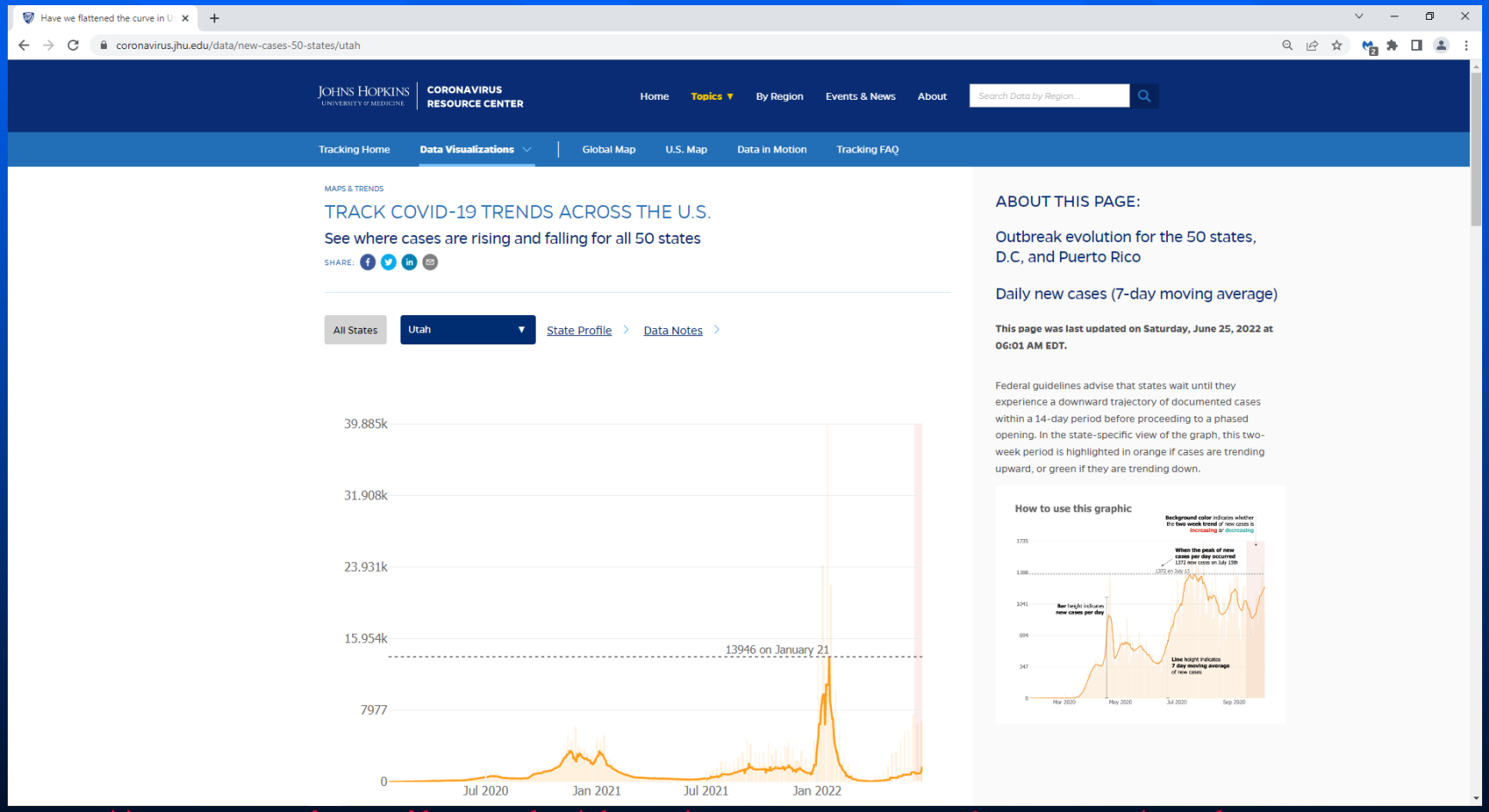
The Good Ones

COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)



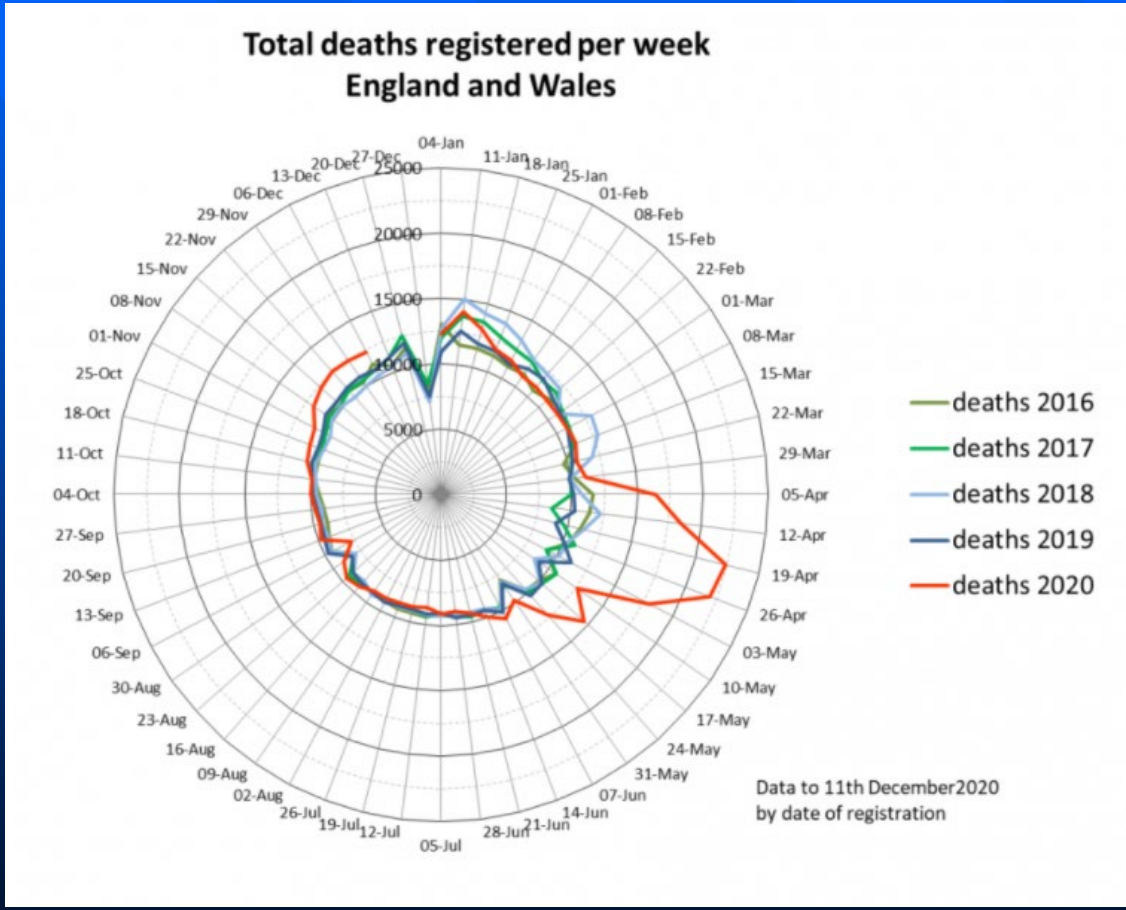
https://coronavirus.jhu.edu/map.html

Johns Hopkins University (JHU): Maps & Trends – Track Covid-19 Trends Across the U.S.



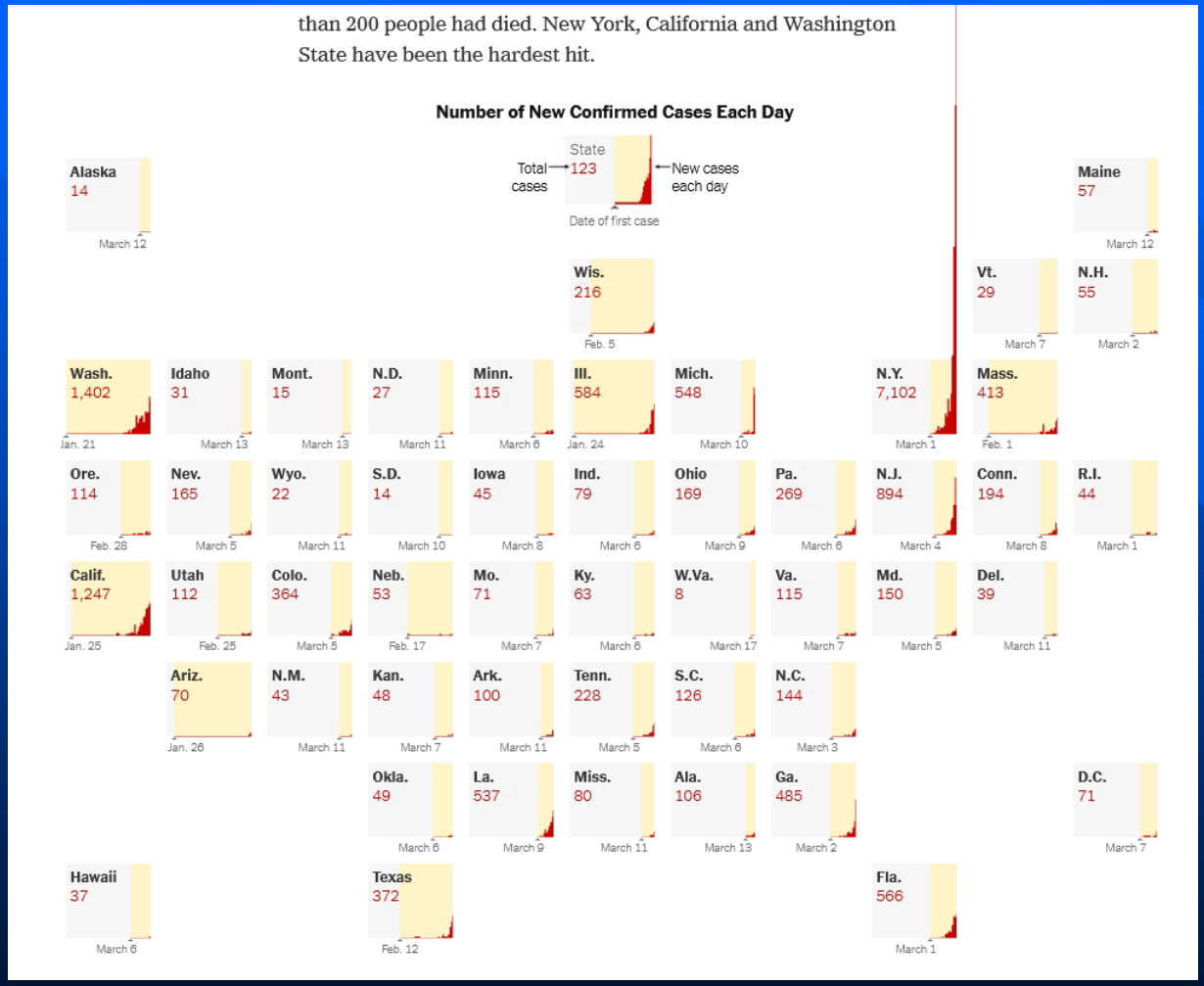
https://coronavirus.jhu.edu/data/new-cases-50-states/utah

The Centre for Evidence-Based Medicine: COVID-19 – Florence Nightingale Diagrams of Deaths in England & Wales



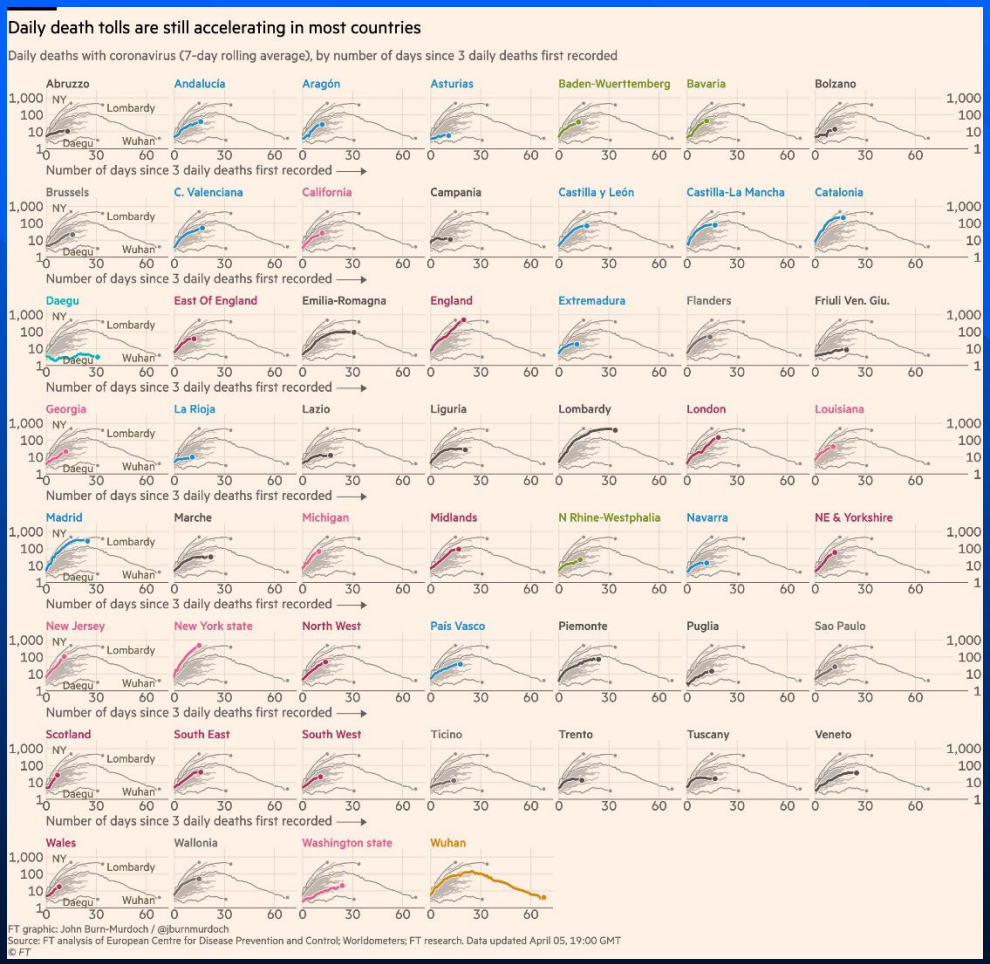
https://www.cebm.net/covid-19/covid-19-florence-nightingales-daigrams-for-deaths/

The New York Times: Watch How the Coronavirus Spread Across the United States



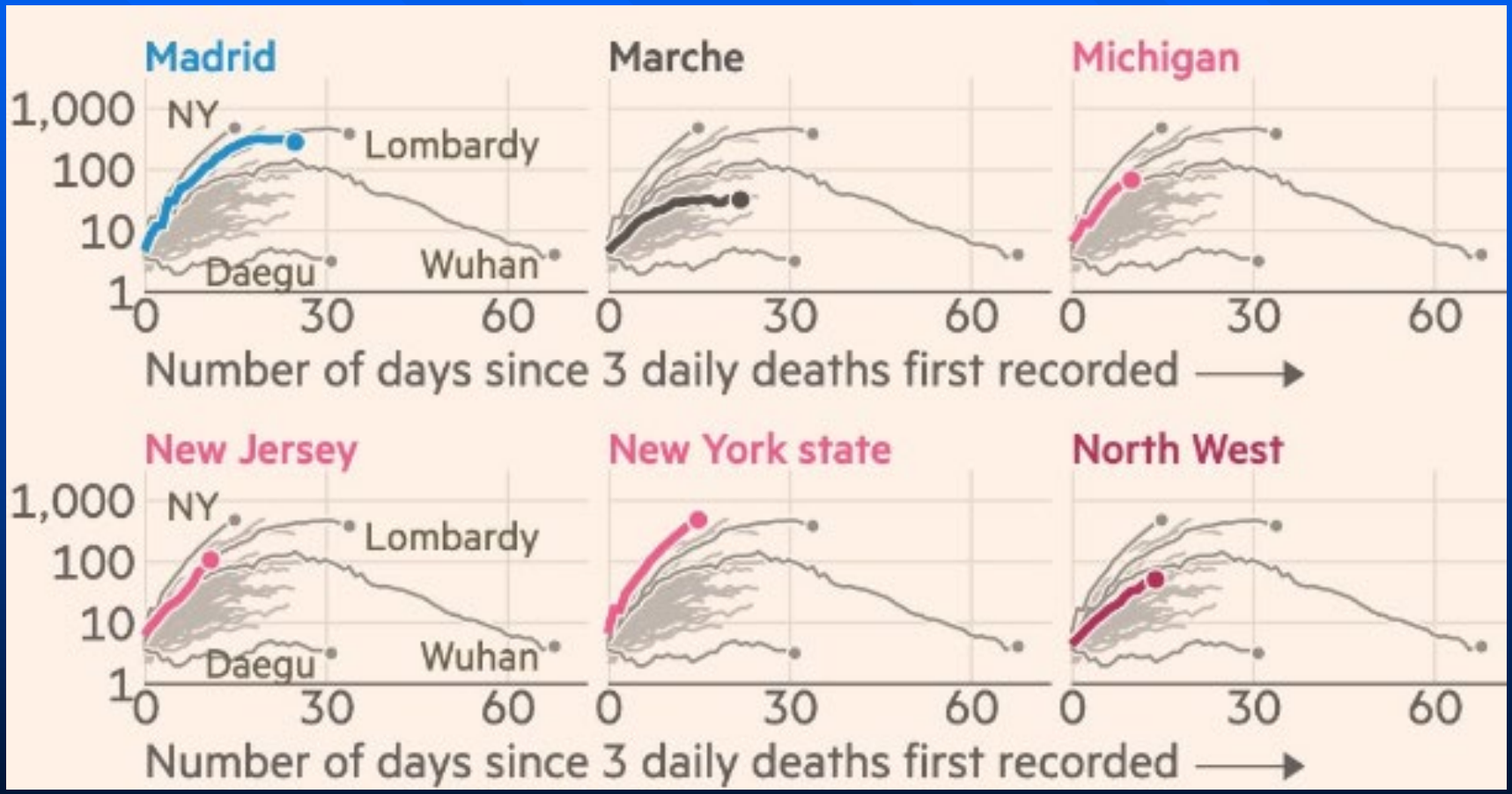
https://www.nytimes.com/interactive/2020/03/21/us/coronavirus-us-cases-spread.html

John Burn-Murdoch: Small Multiples of Daily Deaths in Subnational Regions



https://twitter.com/jburnmurdoch/status/1246904821170483201

John Burn-Murdoch: Small Multiples of Daily Deaths in Subnational Regions - Close-Up



Summary & Discussion

Possible Explanations for Misleading Graphs

- Covid-19 data are **complex**: space, time, rates vs. counts, by age group, deaths / hospitalizations / tested positive but no symptoms, etc.
- Some graphs should **intentionally** mislead!
- **Lack of Knowledge #1:**

  Graphs constructed by non-statisticians, e.g., graphic / web designers.
- **Lack of Knowledge #2:**

  Even teachers of statistics often do not know what the correct graphs should look like.

Good and Bad Covid-19 Graphs Collections on the Web

Good and Bad Covid-19 Graphs Collections on the Web (1)

- https://analythical.com/blog/covid19-in-charts
- https://chezvoila.com/blog/covid-19-best/
- https://junkcharts.typepad.com/junk_charts/covid-19/
- https://qz.com/1872980/how-bad-covid-19-data-visualizations-mislead-the-public/
- https://sloanreview.mit.edu/article/the-challenges-of-presenting-pandemic-data/
- https://theconversation.com/next-slide-please-data-visualisation-expert-on-whats-wrong-with-the-uk-governments-coronavirus-charts-149329
- https://towardsdatascience.com/stopping-covid-19-with-misleading-graphs-6812a61a57c9
- https://venngage.com/blog/misleading-graphs/#Misleading-Coronavirus-graphs

Good and Bad Covid-19 Graphs Collections on the Web (2)

- https://www.datasciencecentral.com/the-worst-covid-19-misleading-graphs/
- https://www.nature.com/articles/d41586-020-01136-8
- https://www.natureindex.com/news-blog/how-to-make-a-coronavirus-data-visualisation-that-counts
- https://www.popsci.com/story/health/misleading-covid-coronavirus-graphs-charts/
- https://www.popularmechanics.com/science/a32243926/best-graphs-coronavirus-statistics/
- https://slate.com/technology/2021/02/covid-19-data-dashboards-states.html
- https://www.thedailybeast.com/that-coronavirus-chart-might-be-lying-to-you
- https://www.theverge.com/2020/4/2/21201832/novel-coronavirus-covid-19-best-graphs-tracking-data
- …

Sources for Constructing Better Graphs

Sources for Constructing Better Graphs: Books & Articles

- Robbins, N. B. (2013). Creating More Effective Graphs. Chart House: Ramsey, NJ. (reprinted from Wiley 2005)
- Su, Y-S. (2008). "It's Easy to Produce Chartjunk using Microsoft Excel 2007, but Hard to Make Good Graphs." Computational Statistics and Data Analysis 52:4594-4601.
- Tufte, E. R. (2001). The Visual Display of Quantitative Information, 2nd edition. Graphics Press: Cheshire, CT. (first edition 1983)
- Wainer, H. (2000). Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot. Psychology Press: London, UK. (reprinted from Copernicus 1997)

Sources for Constructing Better Graphs: Blogs

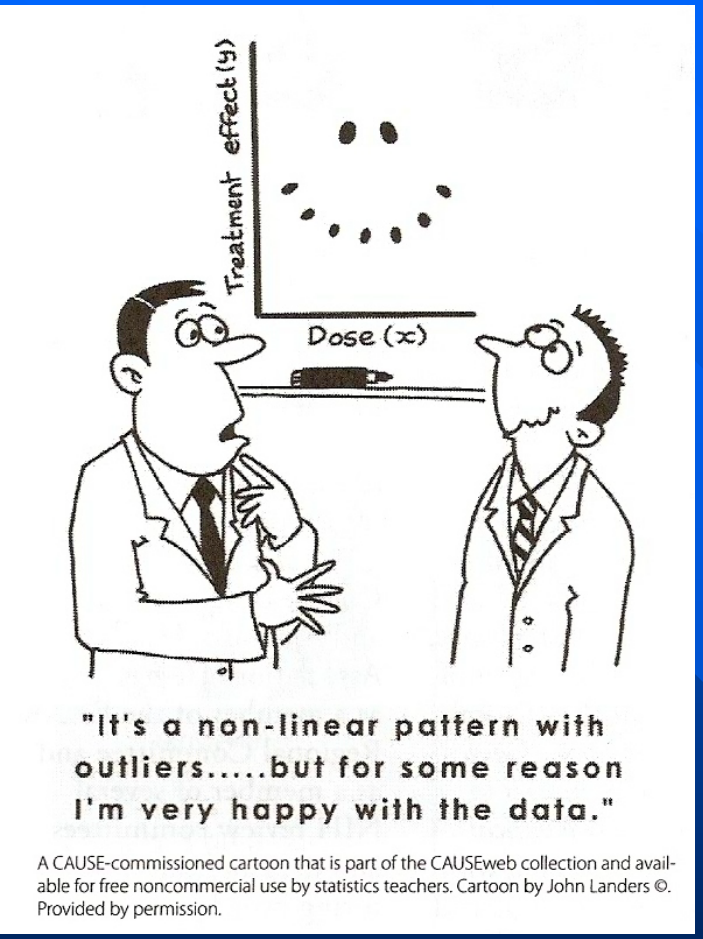
- Camoes, J. “ExcelCharts,” http://www.excelcharts.com/blog/
- Peltier, J. “Peltier Tech Blog— Peltier Tech Excel Charts and Programming Blog,” http://peltiertech.com/WordPress/
- Robbins, N. B. “Effective Graphs,” http://www.forbes.com/sites/naomirobbins/
- Robbins, N. B. (February 28, 2012). “Misleading Graphs: Displaying a Change in One Variable Using Area or Volume,” Forbes.com LLC. http://www.forbes.com/sites/naomirobbins/2012/02/28/misleading-graphs-displaying-a-change-in-one-variable-using-area-or-volume/
- Wyatt, S. (March 31, 2014). “Dishonest Fox Charts: Obamacare Enrollment Edition,” Media Matters for America. http://mediamatters.org/blog/2014/03/31/dishonest-fox-charts-obamacare-enrollment-editi/198679

Sources for Constructing Better Graphs: For Teachers

- Heiberger, R. M., Robbins, N. B., Symanzik, J. (2014). "Statistical Graphics Recommendations for the ASA/NCTM Annual Poster Competition and Project Competition", 2014 JSM Proceedings, American Statistical Association, Alexandria, VA.
- Symanzik, J., Robbins, N. B., Heiberger, R. M. (2014). "Observations from the Winners of the 2013 Statistics Poster Competition — Praise and Future Improvements." The Statistics Teacher Network, 83:2–5. https://ww2.amstat.org/education/stn/pdfs/stn83.pdf
- Symanzik, J., Robbins, N. B., Heiberger, R. M. (2016): "Observations on the Type and Quality of Graphs Used in the ASA/NCTM Annual Poster Competition during the Years 2013 to 2016", 2016 JSM Proceedings, American Statistical Association, Alexandria, Virginia.
- http://www.statlit.org/pdf/2016-Symanzik-Robbins-Heiberger-ASA.pdf

Acknowledgements

Thanks to my current and former students, my colleagues, and my collaborators who contributed to the collection of bad graphs and bad (and good) graph web pages shown in this presentation. I really appreciate obtaining links to bad graphs and bad graphs web pages even when not discussing bad graphs in the classroom or in a committee or task force.
Thank you!



From: Amstat News, January 2009, p. 25

Questions ???

– or –

send e-mail to:

symanzik@math.usu.edu