From scientific visualization to public engagement. Learning from a public archive of COVID-19 related visualizations

Paolo Ciuccarelli & Paul Kahn Northeastern University

One effect of the COVID-19 pandemic is the enormous production of visualizations by data journalists and information in the media and by life and social scientists in the scientific literature. In parallel with the global pandemic, we are experiencing an 'infodemic' – an excessive amount of information about a problem which, according to the World Health Organization, make it difficult to identify a solution[1]. The visualization are an integral component of the problem and the solution.

The spread of visualizations is truly global, with data visualization and information graphics appearing in publications throughout Europe, the Americas, Asia and Africa. Since March 2020, we have been part of an effort to collect a critical mass of these visualizations into an open and public repository, named COVIC (Covid-19 Online Visualization Collection)[2]. The aim of COVIC is to create an archive that can be used for research in data visualization technique and practice. To that end, we have thus far collected about 3,000 articles containing about 10,000 figures, with metadata attached to each article and figure identifying its source, date, subject, and visualization characteristics.

The lessons to be learned from this global effort fall into several categories: one can examine the strength and weakness of specific charts for carrying types of information or compare the presentation of pre-pandemic economic and social indicators with the same types of data generated by the pandemic event. We can see how new public health and public policy indices have been created, how new trends and relationships have been tracked. A visual language has emerged, evolving rapidly under tremendous pressure, largely made from the existing cloth of data visualization techniques.

The pandemic is reported each day in tables of localized time-series data. A broad sample of methods to visualize this kind of data – cases and deaths over time by location, availability of resources such as masks, hospital beds, and vaccines – presents us with an extraordinary opportunity to review the creative energies that have been brought to bear by the data visualization community. Likewise, illustrators and information designers have faced the challenge of explaining to a general public many aspects of microbiology, immunology, and virology – the life cycle of a virus, how our body responds to a previously unknown disease, the ways a vaccine defends us from infection – that are understood by a small scientific audience. Quite suddenly, these scientific hypotheses have leaped from their specialized literature onto the front page of newspapers, daily media broadcasts, and posters telling us where to stand and what not to touch. This shift of audience – from the scientific community to the public – and context – from academic papers to the media sphere – can be framed as a (crucial) act of translation; one that can make the difference between academic speculation and societal change. An effective communication, one that can engage the public with scientific content, and trigger the collective behavioral change we need to tackle the pandemic.

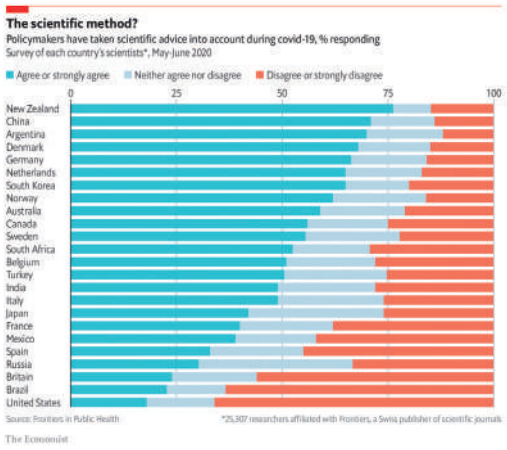
Translating from scientific visualization to data-driven, visual journalism

We have selected from COVIC archive seven examples, representing different levels of transformation or design interventions in the process of translating the data from visualization by life and social science domain experts to the public. COVIC is very strong in English-language examples from the Anglo-American media, and these examples are all drawn that pool of articles. We expect that similar results can be found in examples from other languages and countries.

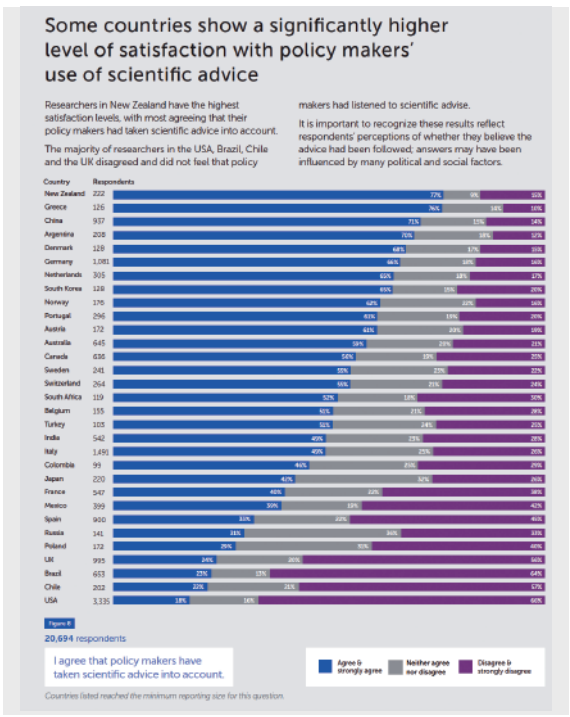
The changes from science to news range from visual design improvements like changing the color scheme to selection and compression of data to redraw the entire visualization. These examples were found by following references in news stories, where "a recent study" is mentioned or a scientific publication is identified as the source of a graphic. Discussing these examples anticipates the value that COVIC will provide to both educational and research activities, in the academic contexts as well as for professionals, in design, health and beyond.

Simple visual improvements

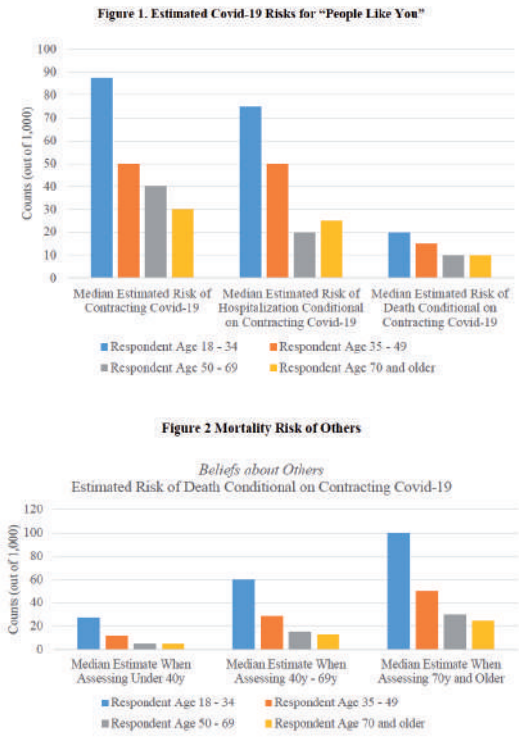
The stack of 100% bar charts that illustrated a survey measuring whether scientists feel that policy makers are following scientific advice is the source of a summary graphic in The Economist. We can see two striking visual design changes: a change in color scheme and the removal of numbers from the bars. The palate used throughout The Economist contributes to the color change, and so does the semiotics of brightness, along with the positive blue and negative orange values. But there is also simplification at work. The Frontiers chart shows the numbers of respondents to the survey for each country as well as repeating the numeric values represented by the parts of the bar. This justification and redundancy is unnecessary to tell the story in The Economist.



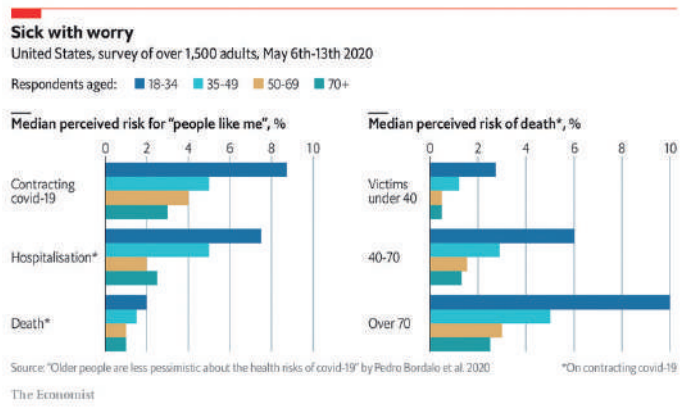
Daily chart - Are governments following the science on covid-19? The Economist [1938]



Frontiers | The academic response to COVID-19 | Frontiers in Public Health [1939] 3



Older People are Less Pessimistic about the Health Risks of Covid-19 NBER [1577-1, 1577-2]



Daily chart - Young people see covid-19 as a bigger threat than their elders do [1576-1]

Two charts summarizing survey data are selected from the National Bureau of Economic Research (NBER) report to support story in The Economist. The obvious transformation is rotating the bar charts and combining them into a single figure so that the reader takes in and compares both ratios. Where grey had been used for one bar in the original – an unhelpful signal of neutrality against other bright colors – the combined graphic substitute a more semantic range. Though the title of the original article led with "Older people are less pessimistic" the longer dark bars for the youngest age group stand out as a visual message that the youngest category is "Sick with worry".

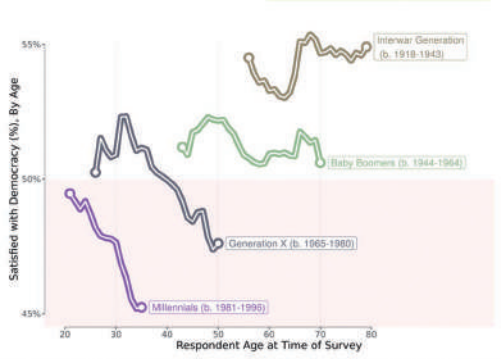
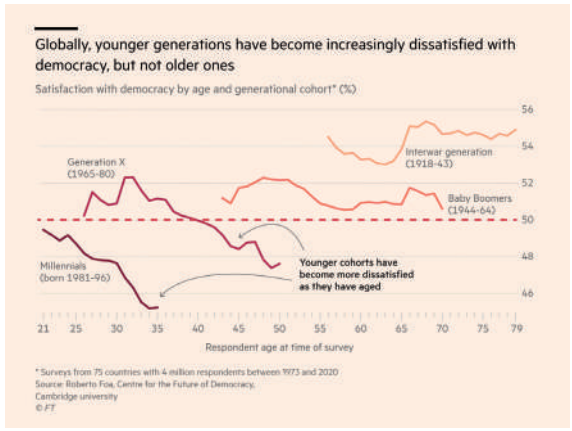


Figure 4: *Satisfaction with democracy by age and generational cohort, for 75 countries across the world in all regions. Data was first aggregated to the country–age level for each cohort, and then, aggregated globally based on country population-weighting in order to provide a representative estimate of the opinions of global democratic citizenry. A constant country sample is used in each age bracket. Each generational cohort is less satisfied with democracy than the prior cohort was at the same age in life. Contrary to the view that generational differences are merely a "life-cycle effect" – with people becoming less critical of democracy as they grow older – in fact, younger cohorts have become more dissatisfied as they have aged.*

Bennett Institute for Public Policy publications [2774]

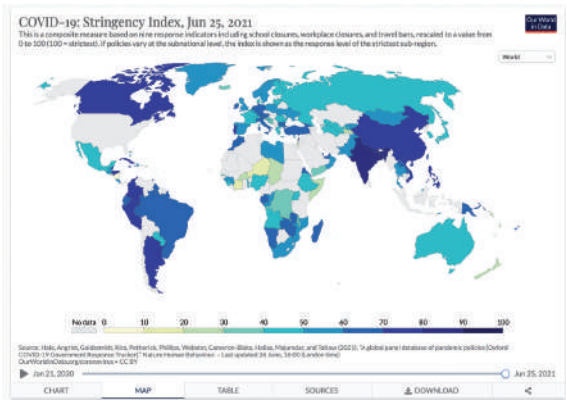


The kids aren't alright: How Generation Covid is losing out Financial Times [1959]

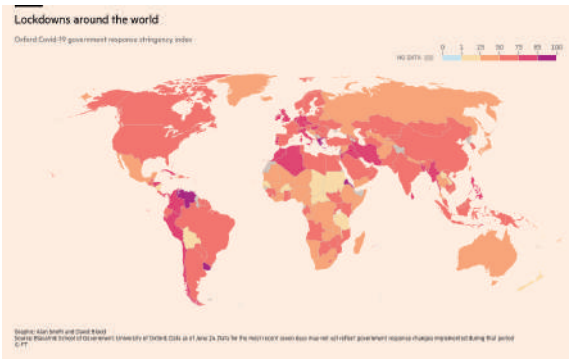
The transformation from the figure in the Bennett Institute and the Financial Times reporting on dissatisfaction with democracy is a good example translating a rigid graphic whose form seems to be defined by the software that drew it to a clear design combining data driven and hand-drawn touches. The Y-axis label is much easier to read horizontally at the top where a reader is expected to understand its relationship to the percentages on its right. Rather than hanging off an end-dot, the labels for the four cohorts are connected to their corresponding lines by proximity. The change in background color marking 50% is dramatized as a red dotted line. And perhaps most significantly a text label is added with arrows to "show" the conclusion: the two youngest cohorts become more dissatisfied as they age.

Translating choropleth map to heat map

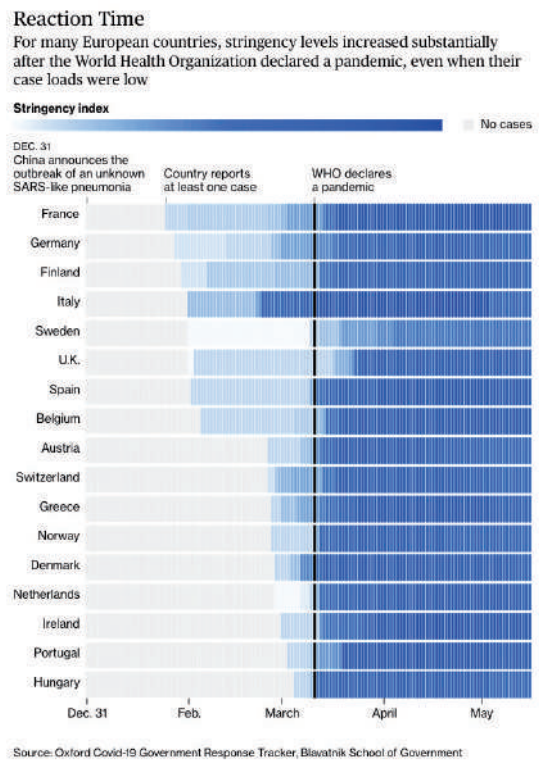
How can we compare the response of governments around the world? The Blavatnik School of Government at Oxford University developed the Government Response Stringency Index by rolling up a numerical value from many factors. The availability of this index made news that month it appeared and has been published daily by Our World In Data. The news stories in Bloomberg and the Financial Times translated the choropleth times series into a heat map. Both approaches visualize the change over time, but the heat map gives the viewer a summary of the change in one view. Bloomberg uses a continuous range of blue. It marks three event thresholds: China announcement, first case in country, pandemic declaration to emphasize when the 'dark' response begins. The Financial Times presents selected countries in continent groups, separating Middle East from Africa/Asia. Their choropleth and heat maps employ a non-continuous number scheme – 1, 25, 25, 25, 10, 15 – color scheme. The result is bright with emphasis on the bottom and top of the index. This change from linear to a custom scale may not be noticed by most viewers.



Our World in Data, COVID-19: Government Response Stringency Index, Blavatnik School of Government, Oxford University [2790-1]



Bloomberg, The Results of Europe's Lockdown Experiment Are In [0917-1]

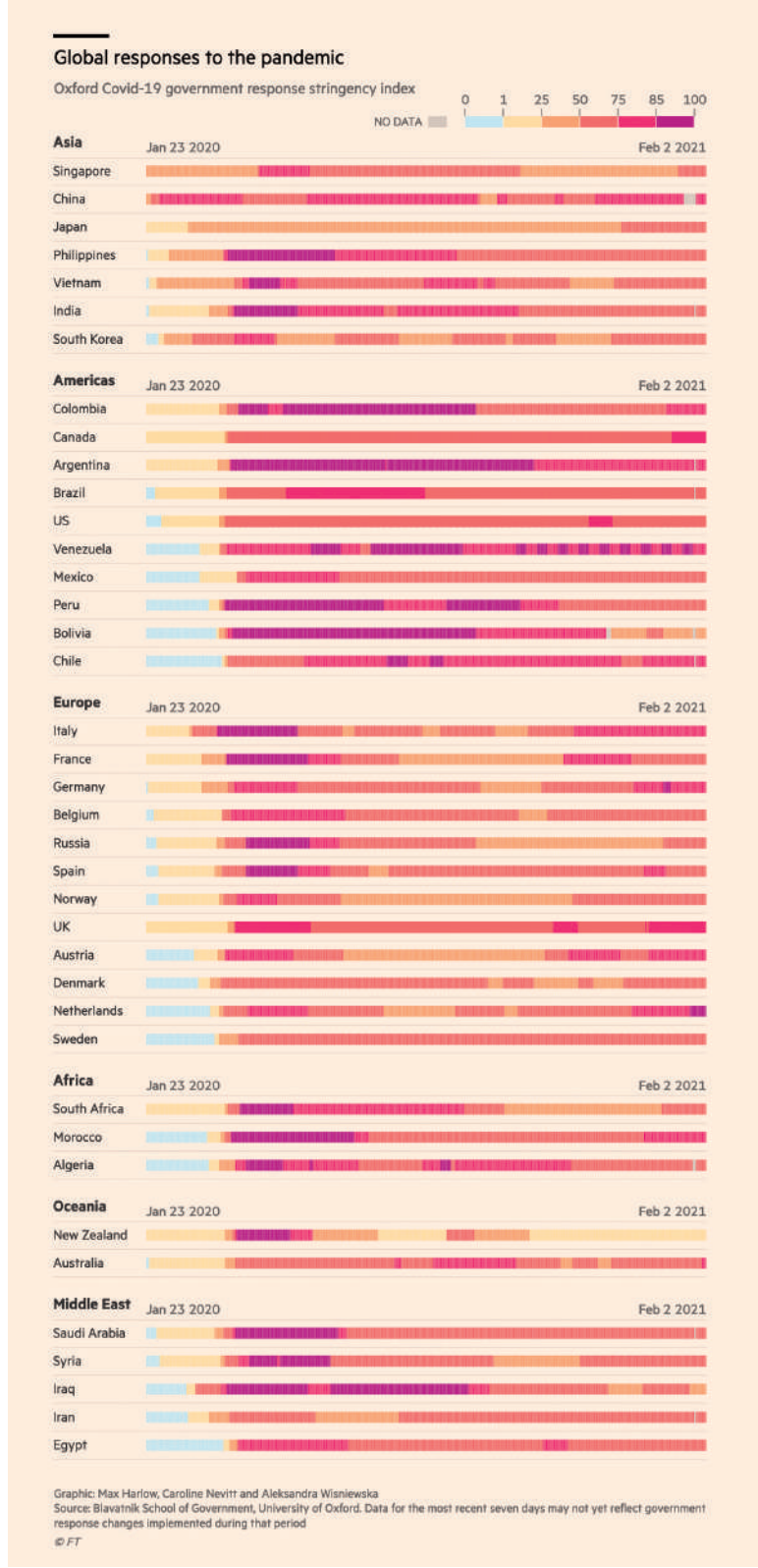


Financial Times, Exiting lockdowns: tracking governments' changing coronavirus responses [0824-2]

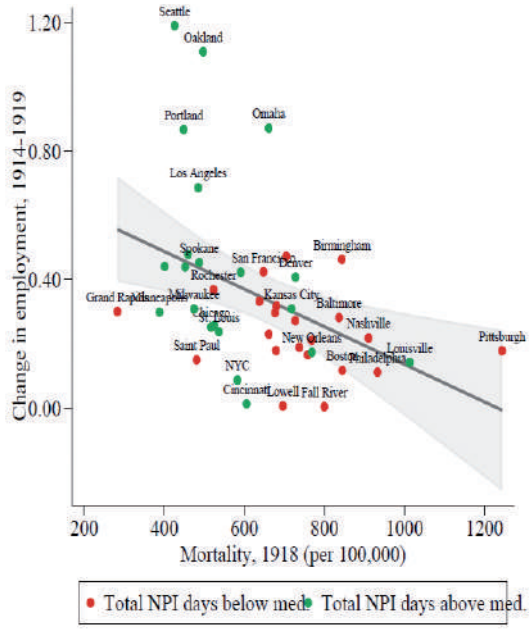
Making sense of scatterplots

As the pandemic spread into Europe and the Americas and governments responded by creating restrictions on gatherings, many people wondered if preventing public gatherings and wearing masks really works. A timely paper by researchers at the US Federal Reserve and the MIT Sloan School of Business addressed this question by combining 100-year-old economic and mortality data from the last US pandemic, the 1918 Flu. The conclusion – that non-pharmaceutical intervensions (NPIs) reduced deaths and over several years led to greater economic growth – was presented in a scatterplot. Cities were represented by above- or below-average color dot with the average shown as a slanting line with grey cones of uncertainty.

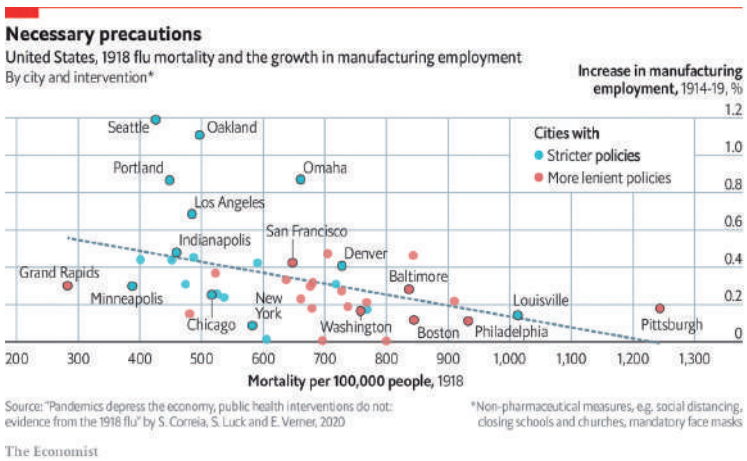
The Economist removed overlaps by spreading out X-axis and labelling only major cities. They simplified it by removing uncertainty, changing Y-axis label text and position and replacing stricter and lenient for above and below average. When a story based on this research appeared in The New York Times, the scatterplot was maintained but the entire orientation changed. The X-axis was changed from mortality to length of NPI, the direction of the average growth line reversed direction, making the line rise from left to right rather than fall. The color coding changed to group pairs cities described in story rather than classify them. In this version, the scatterplot could not be read independently from the narrative in the story.



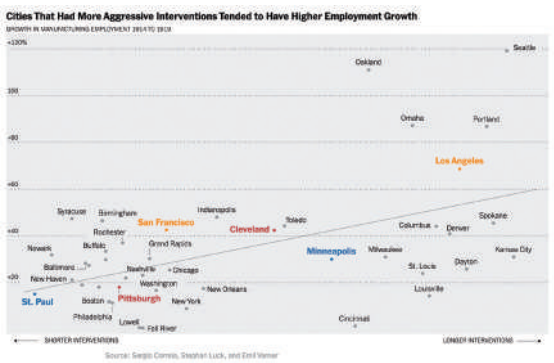
Financial Times, Exiting lockdowns: tracking governments' changing coronavirus responses [0824-3]



Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu, Federal Reserve & MIT Sloan School of Business, SSRN [0478-1]



The Economist, Lessons from the Spanish flu: social distancing can be good for the economy [0354-1]



The New York Times, Cities That Went All In on Social Distancing in 1918 Emerged Stronger for it [0199-1]

How does the disease affect different parts of our population?

As data was collected about the medical, economic and social impact of Covid-19 in the United States, several research groups published reports that focused on the differences among these numbers. By disaggregating numbers according to US racial classifications4, the research highlighted how the pandemic was having a disproportionate impact on specific groups. Two of these reports succeeded in getting the attention of the news media, resulting in a series of stories.

The report from APM Research Lab presented data on all groups in the form of dumbbell charts. This visualization technique is particularly strong at demonstrating the gap of the same variable (line) with two sets of data (dots). The length of the dumbbell in the state data for emphasizes the difference between the percentage of population and the percentage of deaths.

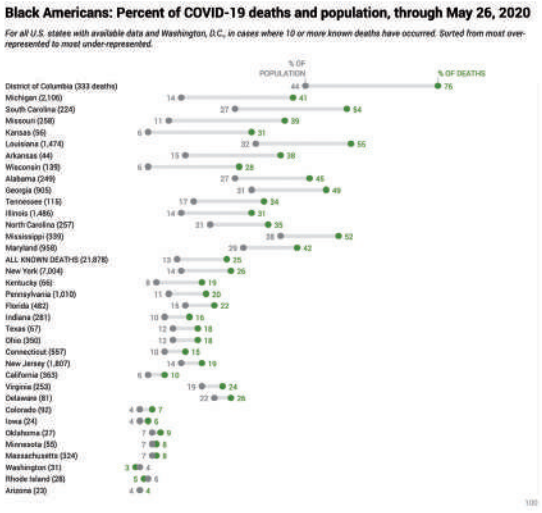
The report from the Economic Policy Institute (EPI) focused entirely on Black/African-American data. It presented data that corresponded to Black and White racial data in two shades of blue. The report consistently employed horizontal bar graphs, placing the Black bar above the White. The report was very thorough with data on a range of categories needed to understand a complex situation, including deaths, sick days, chronic illness, uninsured, population density. The result, however, was visually dull. Looking at the many categories in chart after chart introduced the challenge of understanding the significance of the lower or higher ratio of each category and the difficulty of seeing a relationship among them. For example, the length of bars showing the share of population against the share of Covid deaths tells several stories, none of them clearly.

A story on this subject in the Financial Times presented a selection of the same data in ways that were clearly influenced by their original presentation. They took the data from one bar chart on chronic illness, employed the dumbbell chart technique that was used to emphasize death disproportionate to population, and rounded up percentages to clarify the differences between Black and White population. They combined five of the EPI bar charts into three divergent bar chart, providing meaning in their titles while showing relationship among the categories with their proximity. The bars shift left to show share of population over share of deaths. Grouping potential benefits (paid sick leave, work at home) shows the balance on the right (white) side, while the risk of being uninsured is balanced on the left (black).

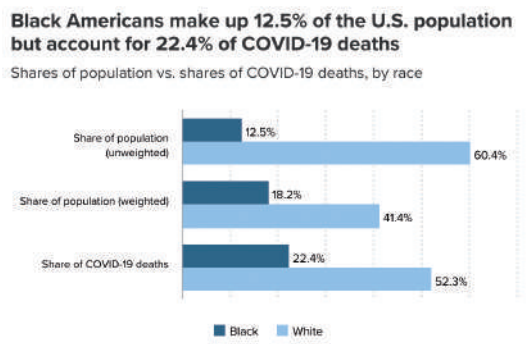
The different impacts on US population based on economic ranking was presented in PNAS. Income was presented in a set of line graphs showing the difference between five quintiles, with the X-axis shows four months (Jan-April), and the Y-axis being specific to each data category. Top income is highest in "Completely at home", while the bottom quintile is highest for "Retail and Recreation". Similar graphs for "Device Exposure" and "Dis-

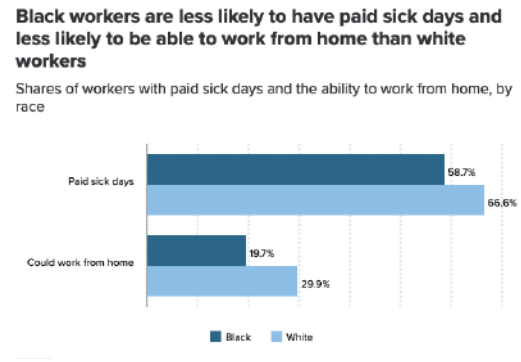
tance Travelled" show less difference or less difference between the top and bottom quintiles.

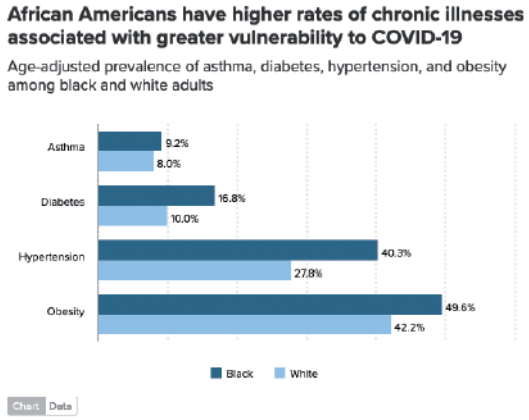
The New York Times story reduces the presentation to two lines representing only the top and bottom quintile. A chronological boundary (National emergency declared) is added as a yellow circle to provide context for the change in the slope of the lines. The labels on the gray line representing the wealthiest 20% tells the story. To further simplify the presentation there is no Y-axis labels. Instead, the final percentage value of rise or fall in the category is labeled at end of line.

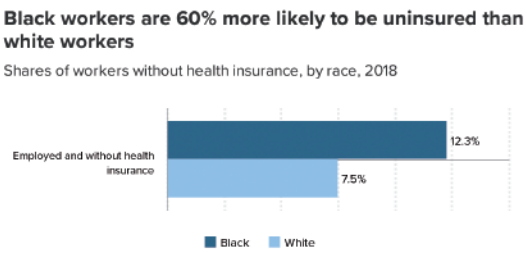


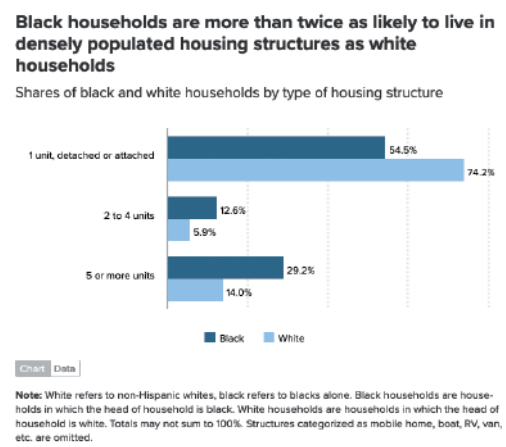
APM Research Lab, The Color of Coronavirus: COVID-19 Deaths by Race and Ethnicity in the U.S. [2219-16]



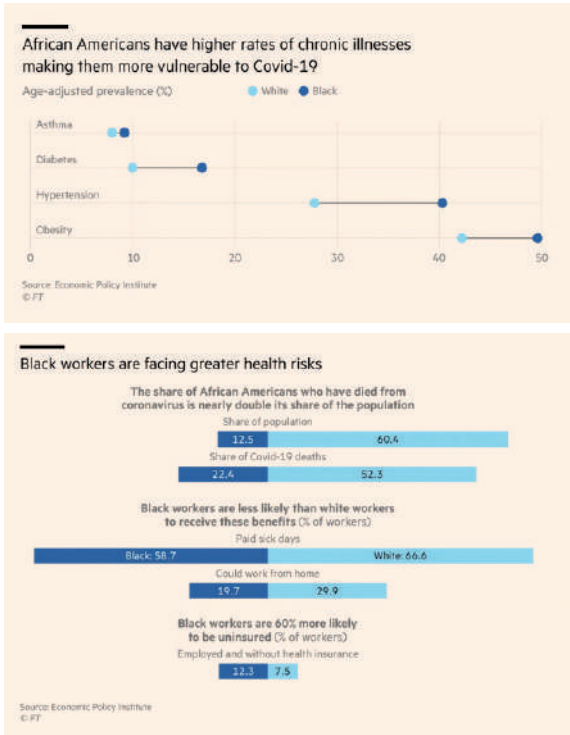








Economic Policy Institute, Black workers face two of the most lethal preexisting conditions for coronavirus—racism and economic inequality [1293-4, 7, 11, 12, 13]



Financial Times, Coronavirus fuels black America's sense of injustice

Combine, Simplify, Tell the Story

What we see in these examples are professionals in different domains – life and social science, news journalism – using data visualization to convince their respective audiences. The audience's expectations, the visual design sophistication, the space available and the time to publication are very different in these domains. There is also the need in data journalism to visually tell the story on the page, in the column, within a scroll range. Life and social science articles and reports have space for numerous figures and appendices. Daily and weekly journalism rarely has room for more than one or two.

A combination of the art direction, graphic design sophistication and data journalism teams in our publication examples are among the strongest in the US and UK. Many of the transformation here reply more on use of color, typography and editorial judgment than they do on data visualization software. Telling the story with data calls on visual sophistication more than on software libraries.

These examples also point to the practice of simplification when translating science into news stories. The choice to change a scale, omit a set of data, re-label axis are editorial judgments. Even our first and simplest example of how scientists feel about policy makers chooses to omit the number of respondents in each country. Simplification helps tell a visual story. It is also a judgment that comes with great responsibility.

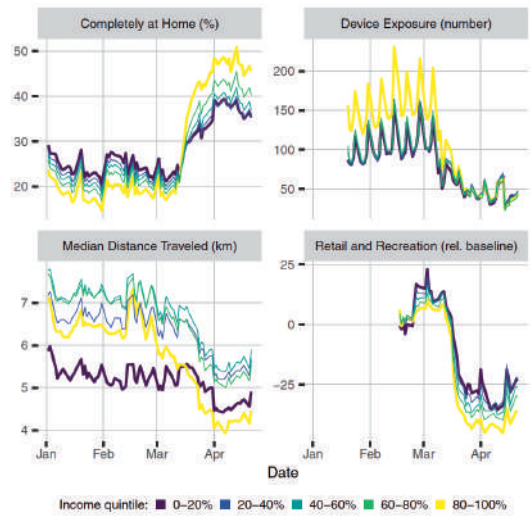
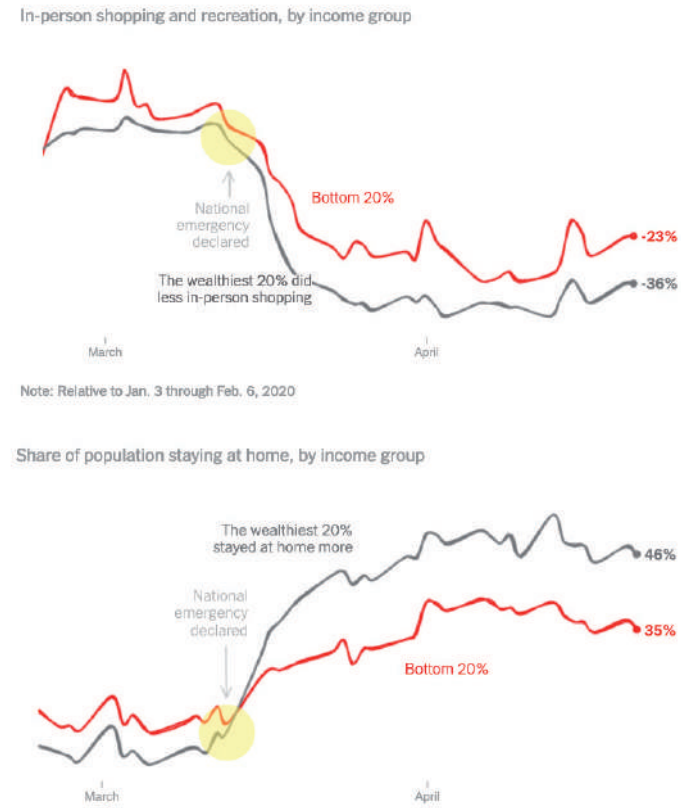


Fig. 1. Daily mean mobility measures in the United States from mobile devices for weekdays from January 1 to April 21, 2020 by quintiles of median income at the census tract (*Left*) or county (*Right*) level. Thicker lines indicate the top and bottom quintile. Each panel shows a different measure of social distancing behavior. Data are from SafeGraph, PlaceIQ, and Google.



PNAS, Social distancing responses to COVID-19 emergency declarations strongly differentiated by income [2461-1, 2]
New York Times, We Did Not Suffer Equally [2402- 2, 3]