What I Learned From COVID-19 Data Visualization

Ryu Sakai (Follow)

Apr 3, 2020 · 14 min read ★

History of Pandemics recreated based on the data from Visualcapitalist.com

Designers have tremendous power and responsibility to handle data and communicate the message visually through various media. The way two designers slice and present the data could make the stories in the two graphics look completely different. This hugely affects the perception and understanding of the viewers.

Today, we are flooded with all sorts of COVID-19 related information, visualized in various different ways. While it's important to stay up to date with the latest information, I felt many visualizations that I see are often short-sighted, and lack a holistic perspective. Among the majority of these, I was able to find a few really good data visualizations that helped me understand a broader perspective. I'd like to share those with you.

Editors note

The COVID-19 pandemic has been a trying time for all of us. The complexities of the new reality spreading across the world have forced us all to come to terms with information that is unfamiliar, scary, and confusing.

We applaud Ryu for sharing his process of learning about the virus to make sense of the world around him. Like many data visualization designers, he chose to understand the pandemic through visualization data. While we don't suggest he was "wrong" in his approach or in his logic, we would also like to offer some additional viewpoints throughout this article. We are thankful to Ryu for allowing us to collaborate with him on this article and hope the expertise of our guest editor Amanda Makulec can continue to guide the evolving best practices for public health and policy.

~ Jason Forrest

Volume comparison with influenza

While the COVID-19 outbreak is terrifying, I strongly felt a need to calm down and view data objectively, so that I can take actions effectively and efficiently without reacting irrationally by fears. I also felt I have a responsibility as a designer to contribute to the world in this respect, which is why I decided to write this article.

Infected annually: Influenza vs. COVID-19 (millions)

Number of people infected to influenza annually vs. number of people infected to COVID-19 as of 4/10/2020
Source: US National Library of Medicine, Worldmeter

The above chart shows there are up to 1 billion people who get infected with influenza every year, according to US National Library of Medicine. The latest data on the number of people infected to COVID-19 counts 1.7 million as of writing this today according to worldmeters.info. It's still April, and the number is continuously growing at a rapid pace, so I don't know how far it's going to go.

On the other hand, I usually don't pay much attention to the massive impact that influenza has every year. My typical attention on influenza has been nothing more than vaguely thinking "oh, it's another flu season again…".

This time, however, I paid more attention for the first time on how massive the effect of influenza has always been. Something that I probably never bothered to look at if COVID-19 never happened.

Death anually: Influenza vs. COVID-19 (thousands)

Death annually: Influenza vs. COVID-19 as of 4/10/2020
Source: US National Library of Medicine, Worldmeter

Editors note

From a public health perspective:

The comparison of COVID-19 to the seasonal flu started at the outset of the pandemic and continues to this day. One conclusion Ryu finds in these charts is worth holding on to: the seasonal flu should be taken seriously. Consider getting your flu shot!

The takeaway we've seen from the COVID-19 vs flu comparison by the numbers alone seems to be, "COVID-19 isn't that bad." As a result, someone may not take the various preventive measures as seriously, like social distancing, handwashing, minimizing trips to the store, and wearing a mask in public. Decisions like these, based on this comparison, is what worries us in public health.

Digging into the data, comparing these two indicators (cases and deaths in 2020) between flu and COVID-19 is misleading. Let's dig into the three big issues:

(1) Timeline: COVID-19 is an emerging pandemic with most known cases counted from January to April of this year. Seasonal influenza is a reportable disease, with <u>a dedicated seasonally-adjusted timeline for data collection</u> in the US. We would need a full year of (quality) data for both illnesses to make more direct comparisons, particularly given the seasonality of the flu.

(2) Data collection and data quality: In the US, COVID-19 cases and deaths are a function of testing (which we know has been inadequate to date to detect all cases). In addition, while there is a central CDC form to capture data about cases, how each state reports cases varies, making it hard to capture accurate source data and calculate aggregated counts. Based on the limitations around testing, we can expect we are undercounting the COVID-19 statistics. In contrast, there is <u>a dedicated, national reporting system in the US</u> and other countries for influenza and broader testing capacity.

(3) Knowledge of the virus, including treatment options: We know much more about the influenza virus and how to treat the associated disease (flu) than we do about SARS-CoV-2, the virus that causes COVID-19. As a result, I would be more concerned about infection with SARS-CoV-2, since treatment options are limited at this point in time, further underscoring why minimizing the impact of COVID-19 compared to flu could be misleading.

~ Amanda Makulec

Fatality rates visualized with Y-axis manipulation

When I looked at the death count of both COVID-19 and influenza above, the fatality rate in COVID-19 was drastically higher than influenza. The break down by age of the

fatality rates showed me a brutal reality that older age groups were significantly higher in COVID-19 compared to seasonal flu as seen in a chart below.
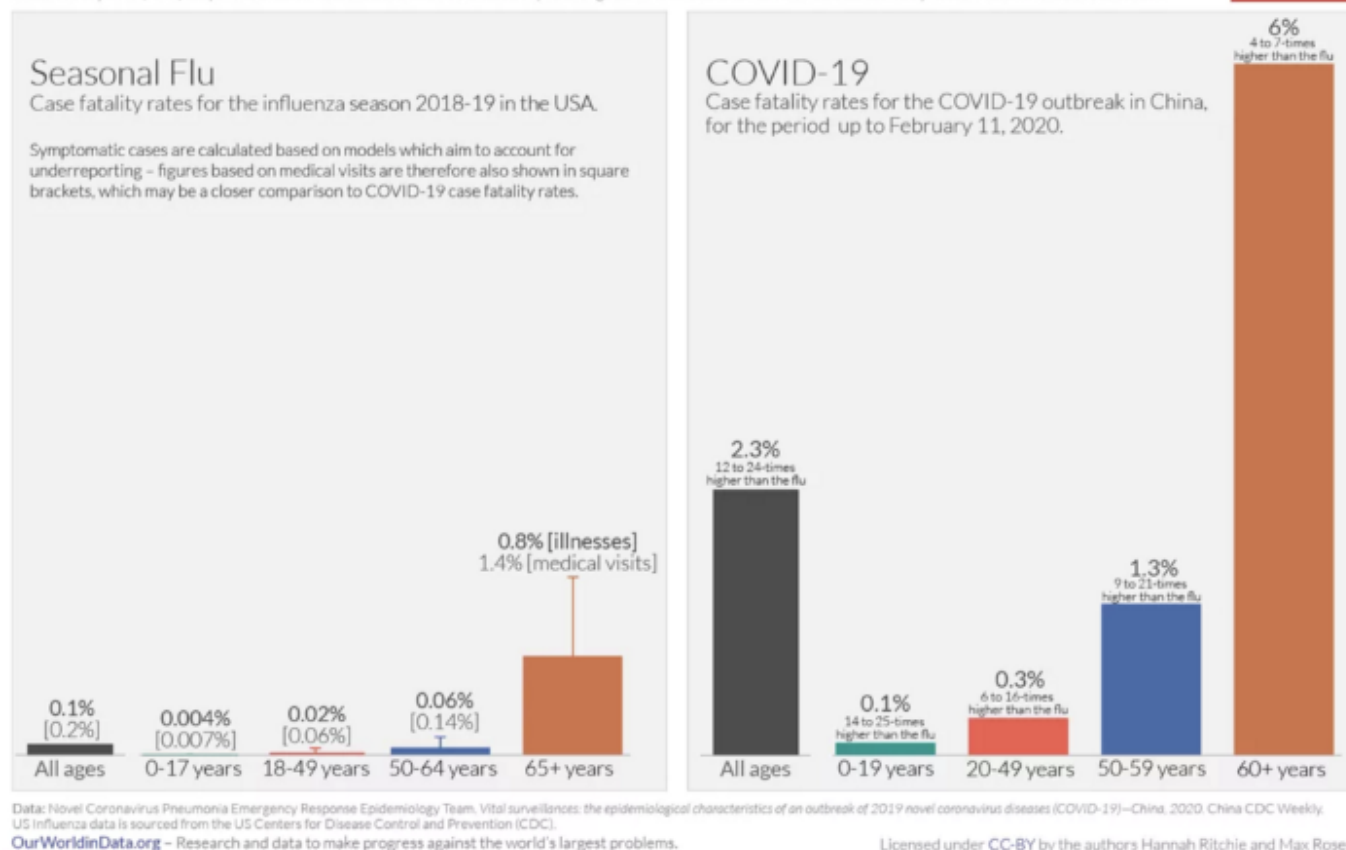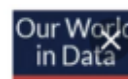
Source: OurWorldInData.org

But I noticed that the maximum value for Y-axis was set to 6%, mapped from the highest value which was the fatality rate of people with 60+ years for COVID-19. Since this graph shows percentages out of total, the maximum value should be set to 100% to give viewers the correct ratio of the data.

So I recreated a graph based on the same data, setting the maximum value of Y-axis to 100% (see below). It looked a lot different from the original chart. The magnitude of all the bars looked less threatening because the highest value was 6 out of 100%. This is how it should be rendered for a percentage-based graph in my opinion.

However, most of COVID-19 fatality rate charts that I see show the maximum value of Y-axis mapped to the highest value like the example above. This is not a data manipulation, but I think it ends up exaggerating the fatality rates of COVID-19 to be

more frightening than needed. This article written by Ryan Mccready describes these tactics as one of the common ways to manipulate viewers.
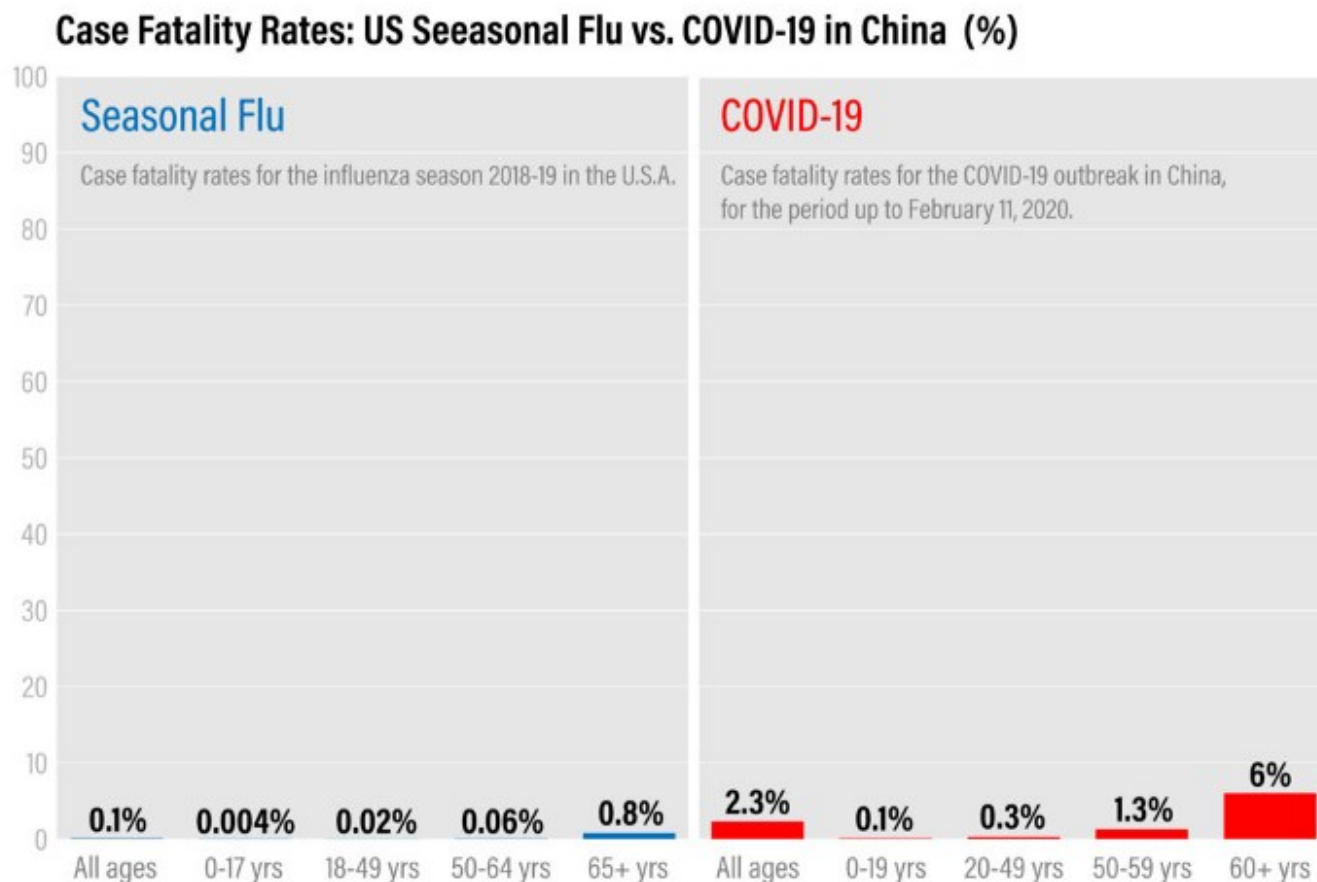
OurWorldInData.org's graph recreated with Y-axis maximum value set to 100%

Editors note

From a public health perspective:

The issues with comparing COVID-19 statistics to those about seasonal flu hold with these charts too.

*In addition, they compare **case fatality rates**, which are notoriously hard to calculate early in an epidemic. Early estimates are influenced by issues like deaths lagging behind cases (functionally, we still have cases in the denominator who will die from the disease) and inaccurate denominators (only including *confirmed* cases, which underrepresent total infections), I'd highly recommend this video interview with Dr. Ellie Murray, Assistant*

Professor of Epidemiology at the Boston University School of Public Health, who digs into issues around case fatality rates in more detail.
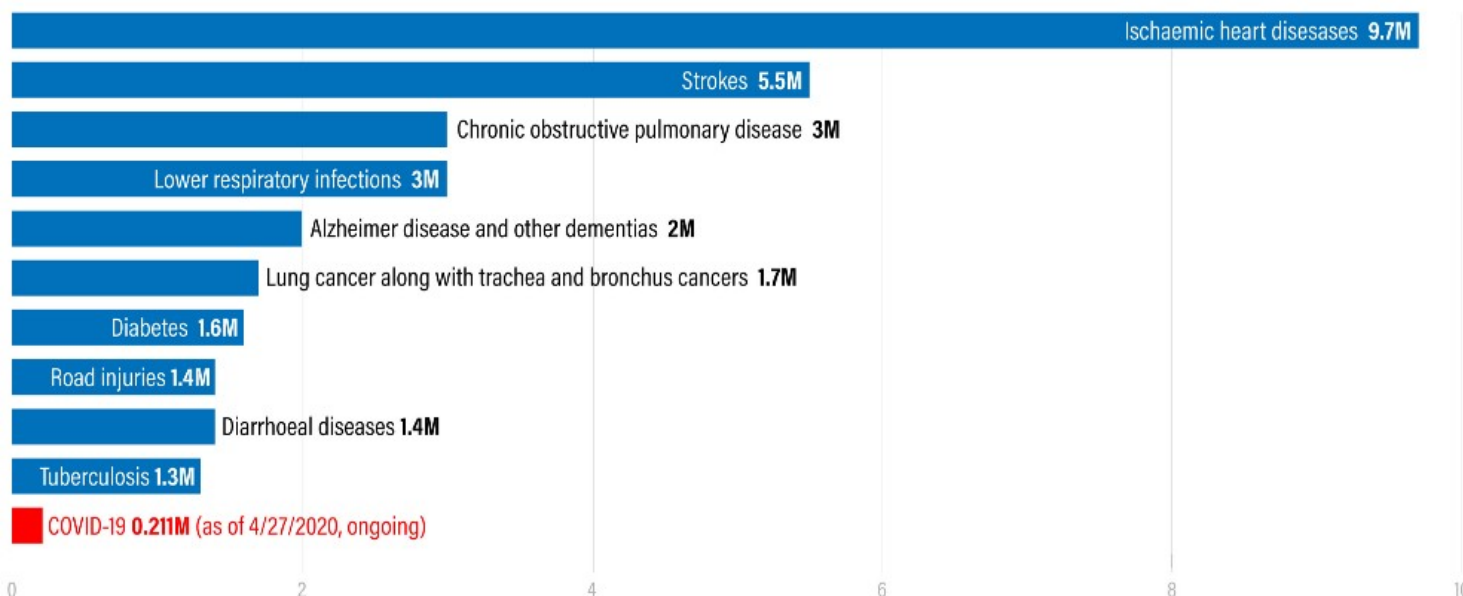
The charts further confuse readers with comparisons of data from different countries (the US and China) from very different periods (US flu data is for the full 2018–2019 flu season, while COVID data is for less than two months of data from China).

Finally, one criticism I've seen from the broader data viz community on some of the bar charts by age is that the age groups are inconsistently sized. This is intentional and common in public health, where the bin size is related more to risk (higher risk in older populations, so the bands are narrower). But in these charts, you'll note that the age ranges aren't equivalent, which further reduces the certainty of any side-by-side comparison.

~ Amanda Makulec

Top 10 cause of death worldwide

Top 10 cause of death in 2016 vs. COVID-19 death (millions)

Ischaemic heart disesases 9.7M
Strokes 5.5M
Chronic obstructive pulmonary disease 3M
Lower respiratory infections 3M
Alzheimer disease and other dementias 2M
Lung cancer along with trachea and bronchus cancers 1.7M
Diabetes 1.6M
Road injuries 1.4M
Diarrhoeal diseases 1.4M
Tuberculosis 1.3M
COVID-19 0.211M (as of 4/27/2020, ongoing)

Source: WHO, Worldmeter

While looking into COVID-19 statistics, I also wanted to look into how many people die with what kind of causes annually and globally. According to WHO, a total of 56.9

million people died in 2016 worldwide. This number blew me away. In another words, we already live in a world where so many people die every year for various causes.

This is absolutely not to say that we don't need to worry about COVID-19 by any means. But at the same time, I needed to understand the big picture from a global point of view. For example, 1.4 million people died from road injuries alone. Most of the time, I don't even have a chance to see such global statistics. COVID-19 pandemic gave me an opportunity to consider all these numbers around deaths in a global context. It also made me appreciate the fact that I still live healthily today.

Editors note

From a public health perspective:

The key takeaway Ryu landed on here — an appreciation for a healthy life — is lovely. I've seen others use this same chart (or something like it) to try to minimize the impact of COVID-19 though, so let's unpack why the numbers aren't all that comparable at this point in time.

*As noted above around making comparisons to seasonal flu, we expect the COVID-19 deaths to continue to grow due to the exponential spread of the virus. We also know that cases and deaths are likely undercounted, which is one reason the bar for COVID-19 would appear so small relative to other diseases. We don't have a firm estimate of **how** undercounted the cases are though, which makes it challenging to add uncertainty bands or bars to this kind of chart.*

If you want to see one set of estimates from epidemiologists on how much we're likely undercounting in various countries around the world, the London School of Hygiene and Tropical Medicine has started to try to create estimates.

> **Adam Kucharski** ✔
> @AdamJKucharski
>
> We've updated our (still very rough) estimates for the proportion of symptomatic COVID-19 cases being reported in different countries. This reflects potential reporting patterns as of couple of weeks ago. See thread below for caveats —

full details here: cmmid.github.io/topics/covid19...

Adam Kucharski ✔ @AdamJKucharski

New analysis from @cmmid_lshtm (led by @timwrussell) provides some rough initial estimates for the % of symptomatic COVID-19 cases that might have been detected/reported in different countries (focusing on those with >10 deaths)... 1/ cmmid.github.io/topics/covid19...

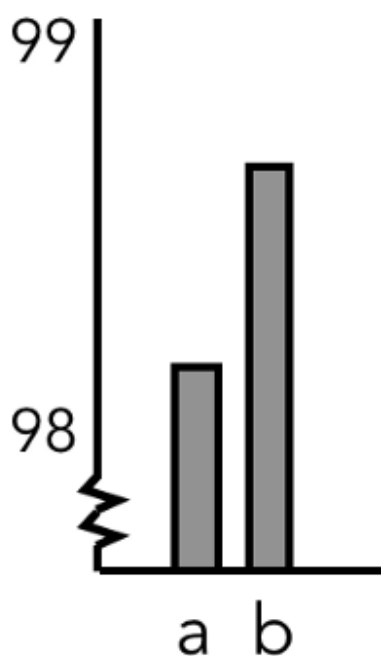11:06 AM · Mar 30, 2020   ⓘ

♡ 422   ⚡ See the latest COVID-19 information on Twitter

A second reason the comparison on a bar chart is misleading is the period represented for different diseases. While the chart calls out in text that the COVID-19 deaths only represent data through April 10, 2020 (approximately 14 weeks or ~27% of the year), the other diseases represent mortality from a full year. When plotting data, we need to be mindful that we're enabling accurate visual comparisons. Research has shown that even when someone tells us in text or with other visual cues that something doesn't quite fit our usual mental model, the speedy interpretations we can make from the visual pieces often mean visual brain beats rational brain.

Read more in the paper "Truncating the Y-axis: Threat or Menace?" (Source)

If you want to explore more about the global burden of diseases and associated mortality, the Institute for Health Metrics and Evaluation conducts a Global Burden of Disease study and has a number of associated articles and visualizations on the topic — but you likely won't see any comparisons with COVID-19 in the immediate future.

~ Amanda Makulec

History of pandemics

I came across a stunning data visualization of the history of pandemics created by visualcapitalist.com below. This introduced me to a history of pandemics. It gave me a broader perspective that humans have always lived in this continuum of constant pandemics. Below is a quote from visualcapitalist.com.

> *THROUGHOUT HISTORY, as humans spread across the world, infectious diseases have been a constant companion. Even in this modern era, outbreaks are nearly constant. Visualcapitalist.com*

Hong Kong Flu 1M
1968-1970

1975

2000

MERS 850
2012-PRESENT

SARS 770
2002-2003

Ebola 11.3K
2014-2016

Swine Flu 200K
2009-2010

COVID-19 23.1K
2019-MAR 26 2020* [ON-GOING]

2025

WHO officially declared COVID-19 a pandemic on Mar 11, 2020.

It is hard to calculate and forecast the impact of COVID-19 because the disease is new to medicine, and data is still coming in.

*As of 12pm PT, according to Johns Hopkins University estimates

DEATH TOLL
[HIGHEST TO LOWEST]

200M
Black Death (Bubonic Plague)
1347-1351

56M
Smallpox
1520

40-50M
Spanish Flu
1918-1919

30-50M
Plague of Justinian
541-542

The plague originated in rats and spread to humans via infected fleas.

The outbreak wiped out 30-50% of Europe's population. It took more than 200 years for the continent's population to recover.

Smallpox killed an estimated 90% of Native Americans. In Europe during the 1800s, an estimated 400,000 people were being killed by smallpox annually. The first ever vaccine was created to ward off smallpox.

The death toll of this plague is still under debate as new evidence is uncovered, but many think it may have helped hasten the fall of the Roman Empire.

A series of **Cholera** outbreaks spread around the world in the 1800s killing millions of people. There is no solid consensus on death tolls.

25-35M
HIV/AIDS
1981-PRESENT

12M
The Third Plague
1855

5M
Antonine Plague
165-180

3M
17th Century Great Plagues
1600

1.1M
Asian Flu
1957-1958

1M
Russian Flu
1889-1890

1M
Hong Kong Flu
1968-1970

1M
Cholera 6 outbreak
1817-1923

1M
Japanese Smallpox Epidemic
735-737

600K
18th Century Great Plagues
1700

200K
Swine Flu
2009-2010

100-150K
Yellow Fever
LATE 1800s

11.3K
Ebola
2014-2016

850
MERS
2012-PRESENT

770
SARS
2002-2003

23.1K
COVID-19
2019-MAR 26, 2020* [ONGOING]
*As of 12pm PT, according to Johns Hopkins University estimates

VISUAL CAPITALIST

Sources:
CDC, WHO, BBC,
Wikipedia,
Historical records,
Encyclopedia Britannica
Johns Hopkins University

f ▶ /visualcapitalist   🐦 📷 @visualcap   ➤ visualcapitalist.com

While this diagram was visually intriguing and the content presented was very good and informative, I found a few issues as stated below.

- Abstracted and generalized virus-like representation was visually interesting and drew my attention in the first place, but it was a more or less cosmetic effect.

- The three-dimensional depth of field towards Z-axis made it impossible to visually compare the sizes of virus-like spheres from different eras.

- It was unclear how the values were mapped to the sizes of the spheres. Was it radius, or the mass? If mapped to the mass, it was not visually intuitive.

- Color coding seemed arbitrary.

- Spherical form factor made it hard to plot precisely on a timeline.

- Because visual size comparison along the timeline was not possible, the designer duplicated another set of spheres just for the size comparison.

Different 3D visualization can make data easier to compare

Below is how I recreated the data visualization, which allowed a viewer to:

1. See a precise plotting of each pandemic on the timeline

2. Compare size of pandemics visually

Obviously it no longer had the colorful virus-like spheres that the original diagram had which drew my attention in the first place. Instead, those were replaced by a bunch of boring bars in a 3D space!

But it achieved successfully in presenting above two aspects without having to duplicate another set of data.
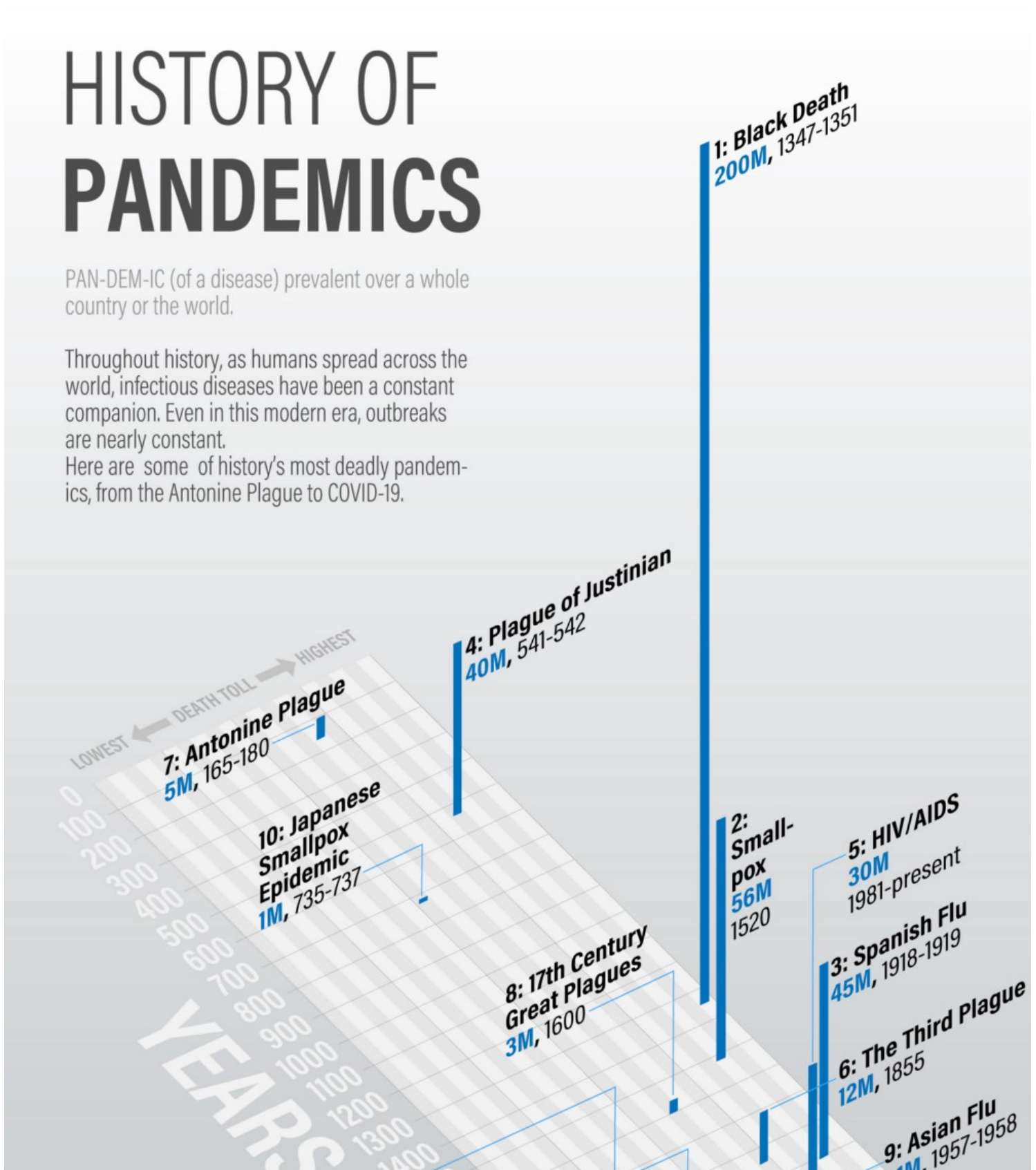
As you can see, a devastating impact and magnitude of Black Death was massive, which was missing in the original diagram as it was far back along the z-axis timeline thus rendered much smaller.
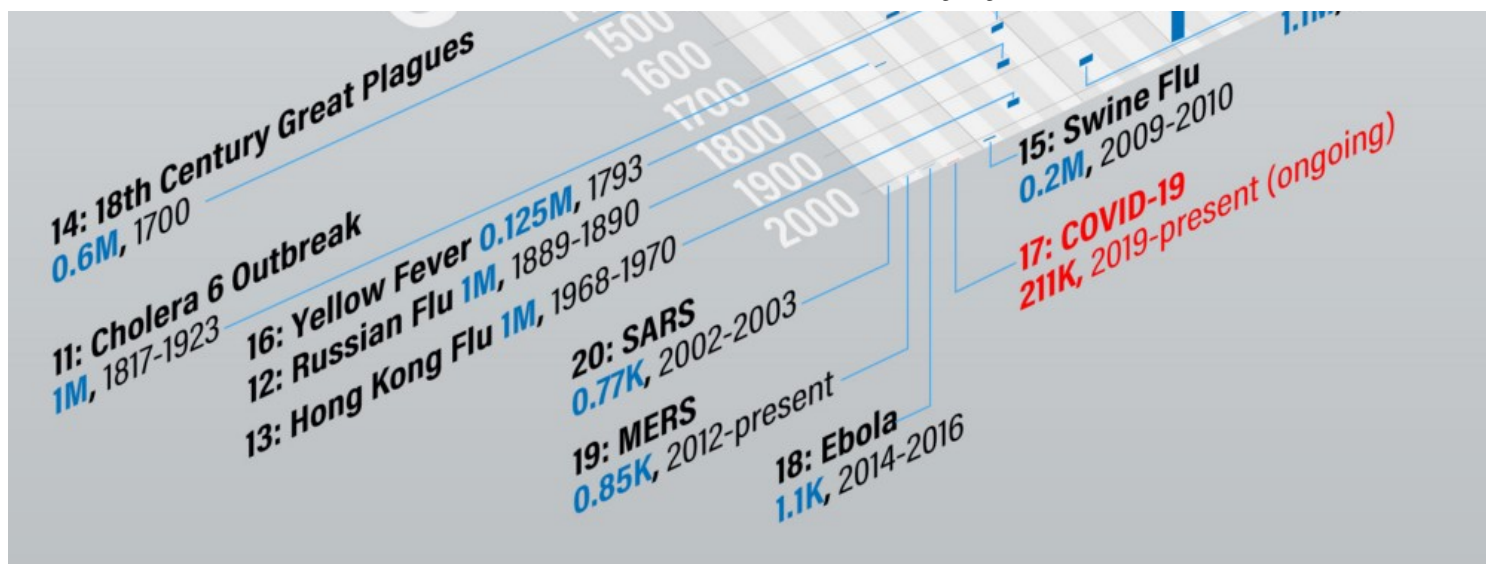

HISTORY OF
PANDEMICS

PAN-DEM-IC (of a disease) prevalent over a whole country or the world.

Throughout history, as humans spread across the world, infectious diseases have been a constant companion. Even in this modern era, outbreaks are nearly constant.
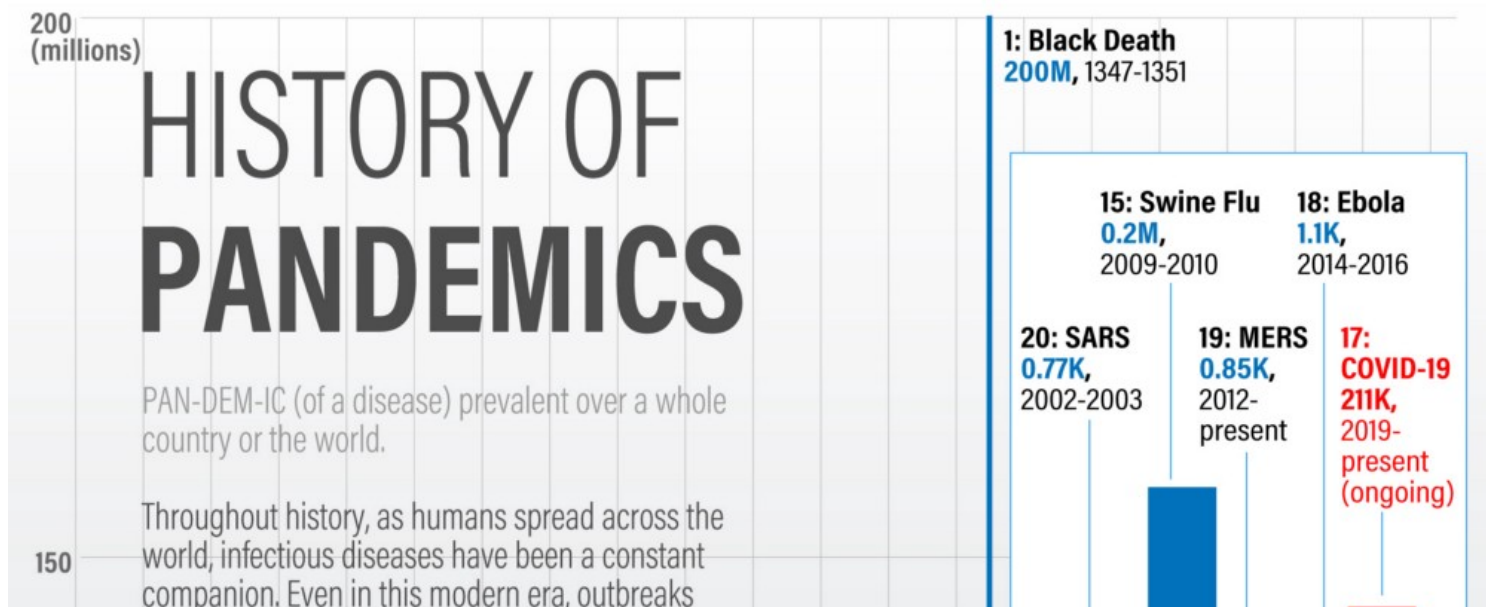Here are some of history's most deadly pandemics, from the Antonine Plague to COVID-19.

1: Black Death
200M, 1347-1351

4: Plague of Justinian
40M, 541-542

DEATH TOLL → HIGHEST
LOWEST ←

7: Antonine Plague
5M, 165-180

10: Japanese Smallpox Epidemic
1M, 735-737

2: Small-pox
56M
1520

5: HIV/AIDS
30M
1981-present

8: 17th Century Great Plagues
3M, 1600

3: Spanish Flu
45M, 1918-1919

6: The Third Plague
12M, 1855

9: Asian Flu
1M, 1957-1958

YEARS
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400

History of pandemics visualized differently for easier comparison — 3D bar graph
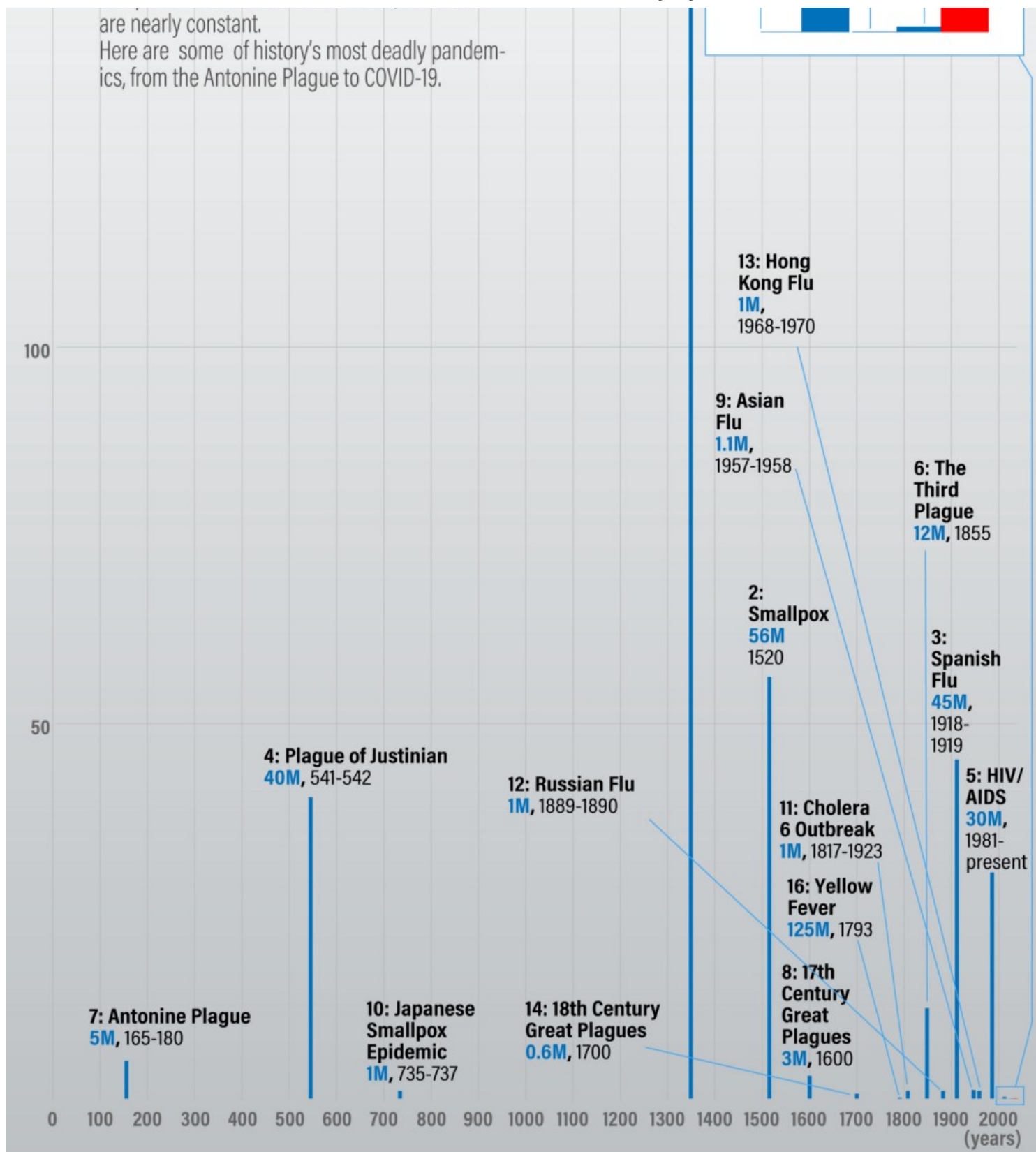Data is based on Visualcapitalist.com and Worldmeter

Interestingly, this 3D visualization allowed an overall holistic view of the data along the timeline while each data still being relatively comparable, as opposed to a 2D bar chart below which became a mess.

In the 2D bar chart, 1800–2000 area was too narrow to fit in all the pandemics that concentrated in that time period. As a result, too many callout lines were overlaid on top of each other and the bars, making it hard to track and read which one was which.

When it came to the most recent 5 pandemics, it was impossible to show on the same plane, which forced those to be separated in a callout box on the top right corner.

are nearly constant.
Here are some of history's most deadly pandem-
ics, from the Antonine Plague to COVID-19.

**13: Hong
Kong Flu
1M,
1968-1970**

**9: Asian
Flu
1.1M,
1957-1958**

**6: The
Third
Plague
12M,** 1855

**2:
Smallpox
56M**
1520

**3:
Spanish
Flu
45M,
1918-
1919**

**4: Plague of Justinian
40M,** 541-542

**12: Russian Flu
1M,** 1889-1890

**11: Cholera
6 Outbreak
1M,** 1817-1923

**5: HIV/
AIDS
30M,
1981-
present**

**16: Yellow
Fever
125M,** 1793

**8: 17th
Century
Great
Plagues
3M,** 1600

**7: Antonine Plague
5M,** 165-180

**10: Japanese
Smallpox
Epidemic
1M,** 735-737

**14: 18th Century
Great Plagues
0.6M,** 1700

100

50

0   100   200   300   400   500   600   700   800   900   1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000
(years)

History of pandemics visualized differently in a simple 2D bar graph, which ended up a mess.
Data is based on Visualcapitalist.com and Worldmeter

Editors note

From a public health perspective:

Comparing the overall toll of these different epidemics is hard to accurately represent in two-dimensional space, let alone moving into 3D where the perception of relative sizing can be skewed by the placement in space (two of the same sized objects will look different when one is placed further 'away' from the reader than the other).

Focusing on the public health concerns, the same issues of comparing an emerging pandemic with mediocre quality data to other diseases (here, pandemics in history) also apply. In some ways, the graphics are interesting to frame the historical perspective of our current situation.

But I've seen graphics like these used to justify resistance to social distancing and other prevention measures by minimizing the impact of COVID-19, which we cannot accurately estimate due to all of the unknowns. The level of uncertainty and the challenges of even quantifying the size of that uncertainty is a major challenge with any visualization of COVID-19 data.

~ Amanda Makulec

Cause and effect of social distancing clearly visualized

Here's another chart that I found extremely powerful and convincing in showing the effectiveness of social distancing, and the consequence of abandoning it too early. I found this in Vox's article titled *How we know ending social distancing will lead to more deaths, in one chart.*
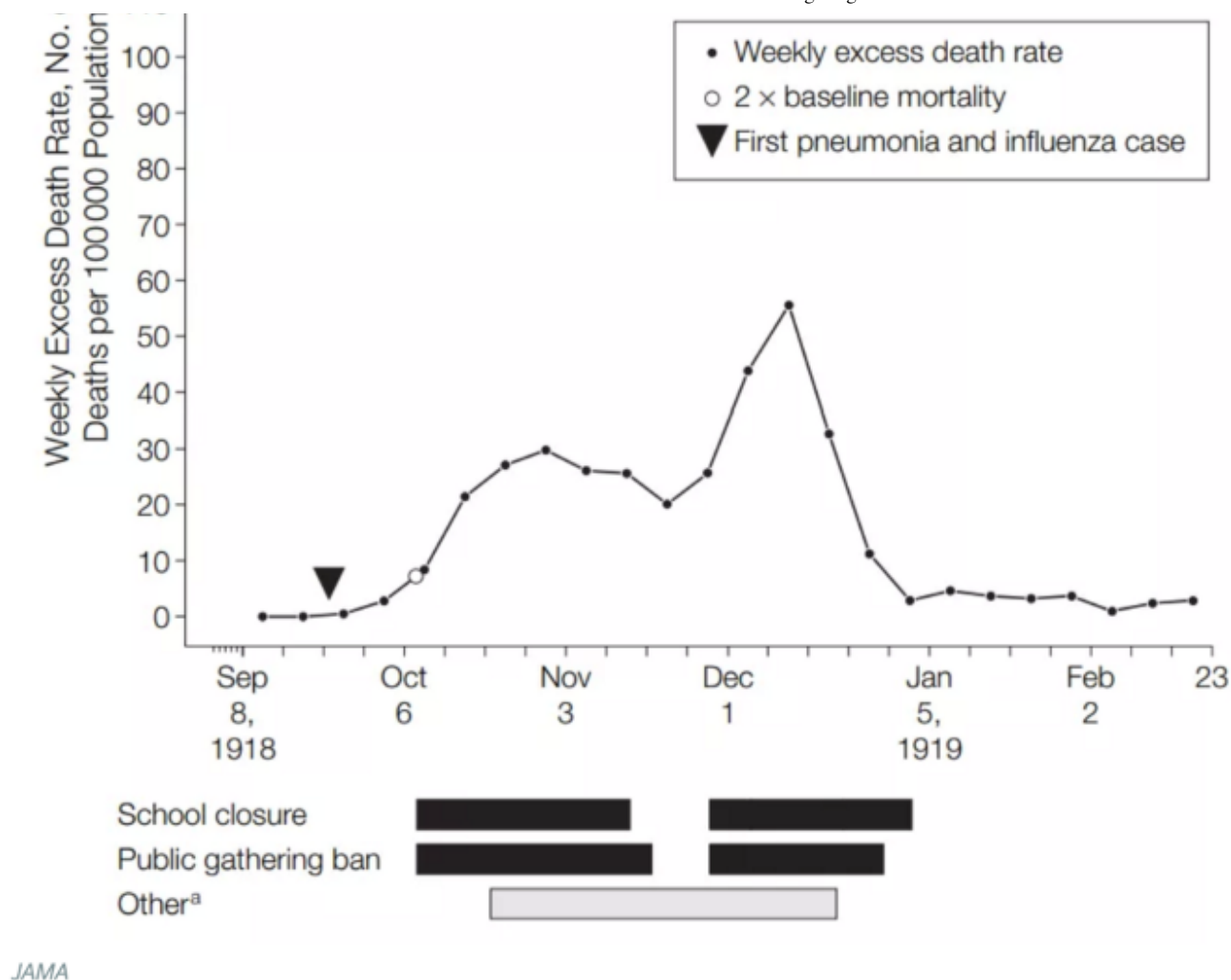
The highest peak comes after social distancing measures were lifted, with the death rate falling only after they were reinstituted.

Source: Vox

A St Louis, MO
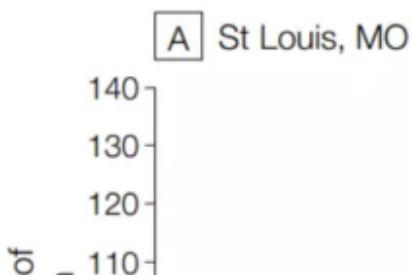
Total excess death rate: 358/100 000 population
Public health response time: +1 d
Total No. of days of nonpharmaceutical
   interventions: 143

140
130
120
110

Source: Vox

From visual design perspective, the chart above is nothing special. Rather, it's a very basic black and white line graph accompanied by bars on the bottom.

However, the content that it carries has significant value in my opinion. This is a clear evidence of how social distancing was effective in a pandemic situation. The way the death rate line graph was combined with the timeline of school closure and the public gathering ban was a brilliant execution to highlight a clear cause and effect relationship between those elements visually. I would say this is a masterpiece of infographics.

Instead of constantly showing fatality rates in Y-axis-manipulated graphs, a quality chart like this should be shared and circulated more across our news and social media. That would be far more informative, helpful and effective for viewers to understand the criticality of social distancing because it clearly illustrates its effectiveness without further explanation. That's the power of data visualization.

Editors note

From a public health perspective:

I love seeing charts from reputable, peer-reviewed journal articles getting shared. I wholly agreed that more of these 'boring' charts (no fuzzy balls representing viruses here!) can be tools to spread more quality information as we make sense of COVID-19, if shared and explained well.

~ Amanda Makulec

Understanding a big picture is an important first step to take actions

Confronting with loads of COVID-19 data visualizations while sheltering in place gave me an ironically interesting opportunity to think through these as a designer and it motivated me to write this article. It made me realize that asking good questions and understanding the big picture is crucial in order to grasp the situation unbiased, so that I can take the correct actions with confidence.

- *Can I get a big picture of past pandemics in our history?*

- *Do I know the impact of influenza every year in volume?*

- *How many people die every year worldwide for what causes?*

- *Why is social distancing important?*

- *What can I do to help myself, my family, and others?*

The whole situation reminded me of Rachel Carson's quote from her famous book, *Silent Spring*, which I borrowed in a Global Village animation that I wrote about in my previous article.

> *The human race is challenged more than ever before to demonstrate our mastery, not over nature but of ourselves.*
>
> *Rachel Carson, Silent Spring 1962*

I feel like we as the human race are challenged to not only care about ourselves but also care for others. To demonstrate our mastery, understanding is one of the first important steps. I took this step, and now I feel more prepared than before.

Editors note

Thanks again to Ryu for his interest in collaborating on this article!

It's challenging to navigate through so much information about COVID-19 in the world. Hopefully, this back and forth format helps you understand how an informed public health perspective can have such a huge impact on how we understand the data and communicate it effectively to others. To learn more:

Ten Considerations Before you Create another Chart about COVID-19

To sum it up — #vizresponsibly; which may mean not publishing your visualizations in the public domain at all.

medium.com

Sign up for The 'Gale

By Nightingale

Keep up with the latest from Nightingale, the journal of the Data Visualization Society Take a look.

Get this newsletter

Emails will be sent to nhayashi@bu.edu.
Not you?

Data Visualization   Data Ethics   Covid 19   Infographics   Design

Get the Medium app