1S 2022-2023

DEVC 127
LESSON 6

DATA STORIES

RIKKI LEE MENDIOLA
Lecturer, DDJ

today

data storytelling

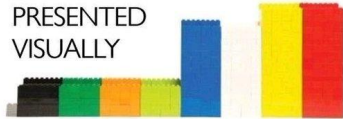developing data insight

tips

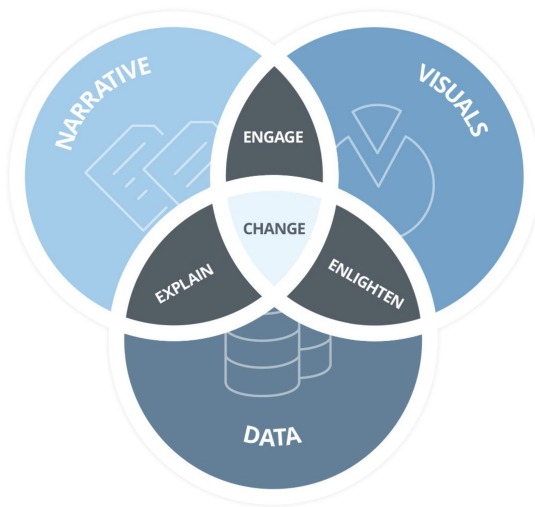next week

DATA

SORTED

ARRANGED

PRESENTED
VISUALLY

EXPLAINED
WITH A STORY

how do i put it all together?

data storytelling

- A structured approach for communicating data insights

- Three key elements: Data, Visuals, Narrative

data can...

- create deeper insights into what is happening around us and how it might affect us help analyze a complex situation

- tell us what a certain development might actually mean

- reveal trends

- debunk "common knowledge"

- reveal injustices perpetrated by public authorities and private corporations mislead or confuse

The Pudding

Calculating "Middle Income"

To understand movement between social classes, we need a quantitative metric: something we can measure to define the middle class. Many researchers use household income and thus numerically locate the families that are in the middle of the income distribution. Even then, there are many ways to numerically define the "middle" of an income range. One technique is to use the mathematical concept of quintiles, which requires equally dividing up the entire population into five groups based on their income. Here's how the concept of quintiles works:

STEP 1

Collect annual household income data from all families

$100k $20k

STEP 2

Sort all families in order of lowest household income to highest
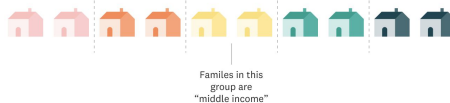
STEP 2

Sort all families in order of lowest household income to highest
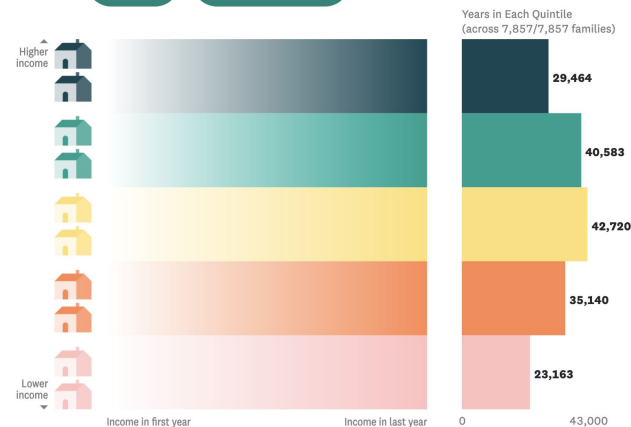
$20k $100k

STEP 3

Separate families into 5 groups with the same number of families in each group. These are called *quintiles*. Families in the 3rd or middle quintile can be thought of as being "middle income". For our essay, we consider "middle income" families to be the "middle class" and will use the two terms interchangeably.

Families in this group are "middle income"

REPLAY   SKIP TO THE END

Years in Each Quintile
(across 7,857/7,857 families)

Higher income

29,464

40,583

42,720

35,140

Lower income

23,163

Income in first year   Income in last year   0   43,000

https://pudding.cool/2020/08/income/

people hear statistics, but they feel stories

- According to USC neuroscientists, decisions are often based on emotion, not logic

- Packaging data insights as a story builds a bridge for your data to the influential, emotional side of the brain.

- Stories can share insights and ideas in a way that is memorable, persuasive, and engaging

studies on stories vs. statistics

- 63% of people remember stories, but only 5% could remember a single statistic – Stanford study

- A story-based brochure for Save the Children Foundation persuaded people to donate more than an infographic-based brochure did

- People drop their intellectual guard and are less critical and skeptical when listening to a story vs reading statistics

https://www.youtube.com/watch?v=QsBT5EQt348
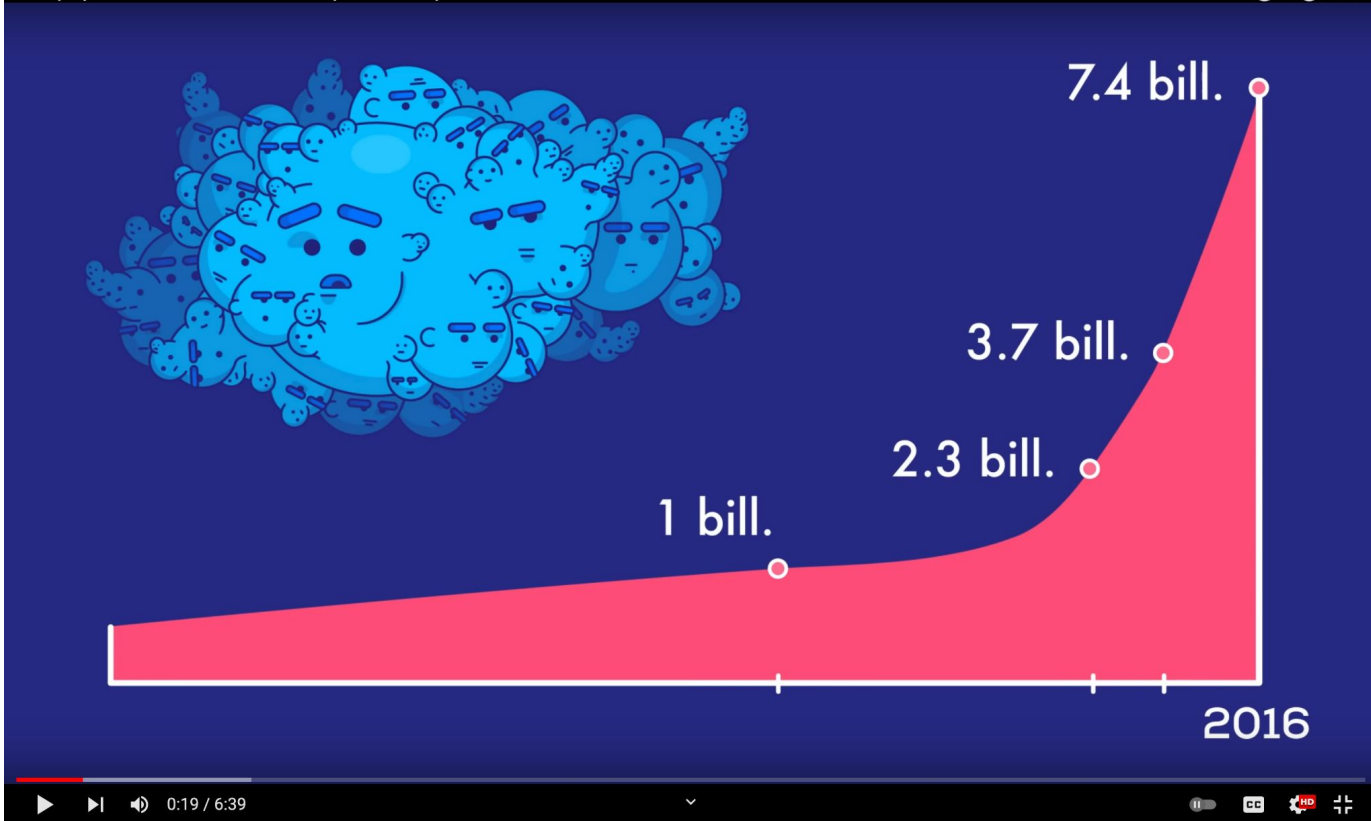
developing data insight

- Conceptual (What is it?)

- Processual (How does it function?)

- Procedural (How do you do it?)

- Structural (What are the parts?)

- Organizational (How is it organized?)

- Numerical (How many?)

How do death rates from COVID-19 differ between people who are vaccinated and those who are not?

by Edouard Mathieu and Max Roser

November 23, 2021

*Our World in Data presents the data and research to make progress against the world's largest problems. This data presentation is part of our work on the **Coronavirus Pandemic**.*

Reuse our work freely

To understand how the pandemic is evolving, it's crucial to know how death rates from COVID-19 are affected by vaccination status. The death rate is a key metric that can accurately show us how effective vaccines are against severe forms of the disease. This may change over time when there are changes in the prevalence of COVID-19, and because of factors such as waning immunity, new strains of the virus, and the use of boosters.

On this page, we explain why it is essential to look at death *rates* by vaccination status rather than the *absolute number* of deaths among vaccinated and unvaccinated people.

We also visualize this mortality data for the United States, England, and Chile.

Ideally we would produce a global dataset that compiles this data for countries around the world, but we do not have the capacity to do this in our team. As a minimum, we list country-specific sources where you can find similar data for other countries, and we describe how an ideal dataset would be formatted.

Then we learn that of the 10 who died from the virus, 50% were vaccinated.

5 of those who died were **unvaccinated**
5 of those who died were **vaccinated**

The newspaper may run the headline "Half of those who died from the virus were vaccinated". But this headline does not tell us *anything* about whether the vaccine is protecting people or not.

To be able to say anything, we also need to know about those who did *not* die: how many people in this population were vaccinated? And how many were not vaccinated?

10 were unvaccinated   50 were vaccinated

Now we have all the information we need and can calculate the death *rates*:

- of 10 unvaccinated people, 5 died → **the death rate among the unvaccinated is 50%**
- of 50 vaccinated people, 5 died → **the death rate among the vaccinated is 10%**

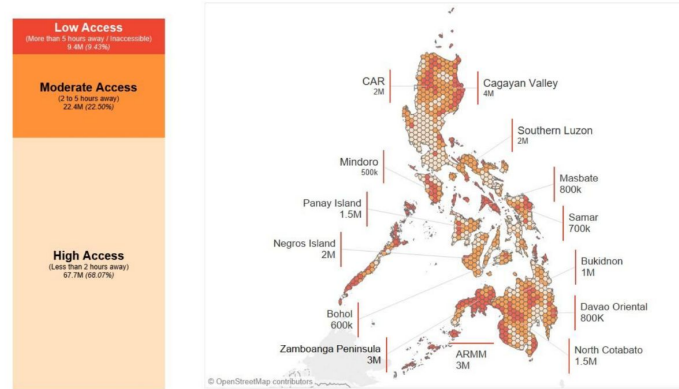https://ourworldindata.org/covid-deaths-by-vaccination

Mendiola, World Bank Philippines

Long distance: the struggle is real
High population areas where the nearest DOH-designated HIV Center is more than 2 hours away

Low Access
(More than 5 hours away / Inaccessible)
9.4M (9.43%)

Moderate Access
(2 to 5 hours away)
22.4M (22.50%)

High Access
(Less than 2 hours away)
67.7M (68.07%)

CAR 2M
Cagayan Valley 4M
Southern Luzon 2M
Mindoro 500k
Masbate 800k
Panay Island 1.5M
Samar 700k
Negros Island 2M
Bukidnon 1M
Bohol 600k
Davao Oriental 800K
Zamboanga Peninsula 3M
ARMM 3M
North Cotabato 1.5M

© OpenStreetMap contributors

Source: DOH HIV/AIDS & ART Registry of the Philippines (HARP) Report , July 2018

Accessibility of DOH HIV Centers in the Philippines
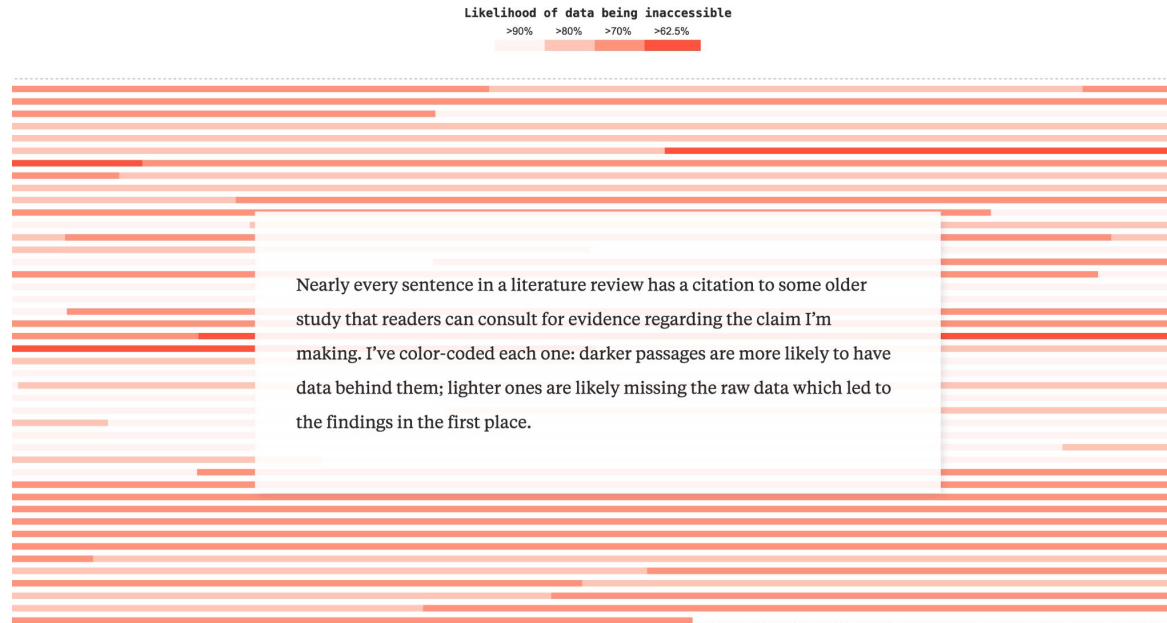stories.thinkingmachin.es

Thinking Machines

DOH HIV centers with high burdens

But even in areas within reasonable driving distance to an HIV center, are the existing facilities enough to serve the population?

We wanted to estimate, even in a simple way, how many people each HIV testing center might need to serve. Let's assume that even just 10% of everyone 15 years old and above (70% of the population according to the Philippine Statistics Authority) gets tested just once a year at their nearest DOH-designated HIV center. How

https://stories.thinkingmachin.es/doh-hiv-centers-access/

Nearly every sentence in a literature review has a citation to some older study that readers can consult for evidence regarding the claim I'm making. I've color-coded each one: darker passages are more likely to have data behind them; lighter ones are likely missing the raw data which led to the findings in the first place.

7/10

https://pudding.cool/2019/05/science_etc_1/

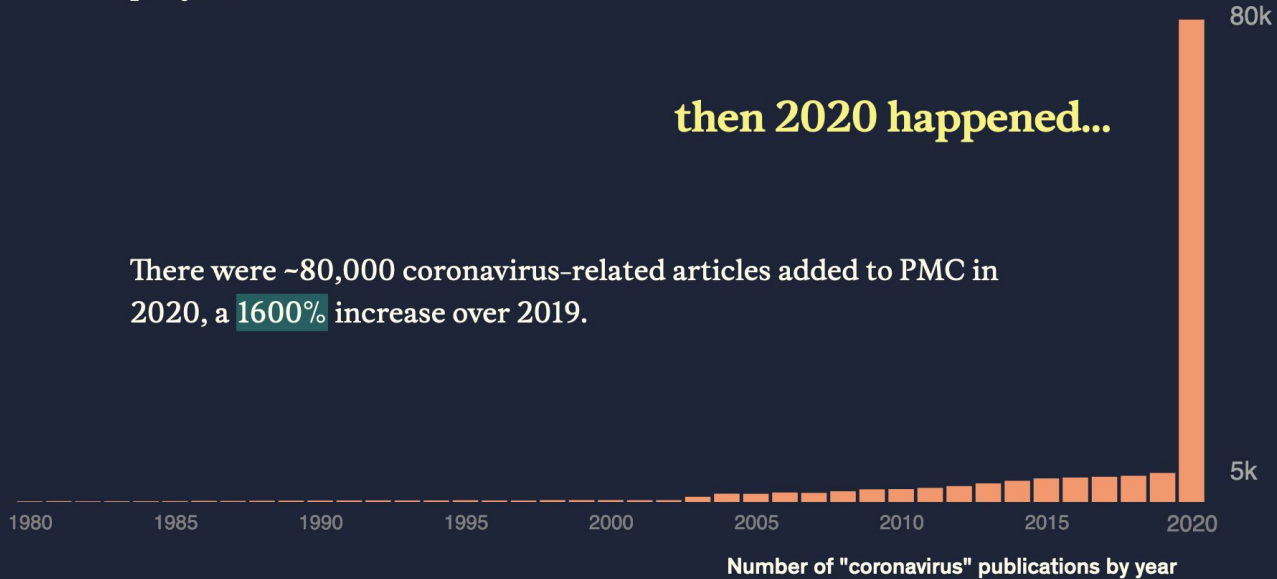# you don't need to discuss everything

tips for effective data storytelling

- Know and understand the audience

- Be timely

- Be clear

- Evoke emotions

- Visualize the data

- Provide the raw data alongside the story

By 2019, there were ~5000 publications on coronaviruses coming out per year.

then 2020 happened...

There were ~80,000 coronavirus-related articles added to PMC in 2020, a 1600% increase over 2019.

80k

5k

1980   1985   1990   1995   2000   2005   2010   2015   2020

Number of "coronavirus" publications by year

https://pudding.cool/2021/03/covid-science/
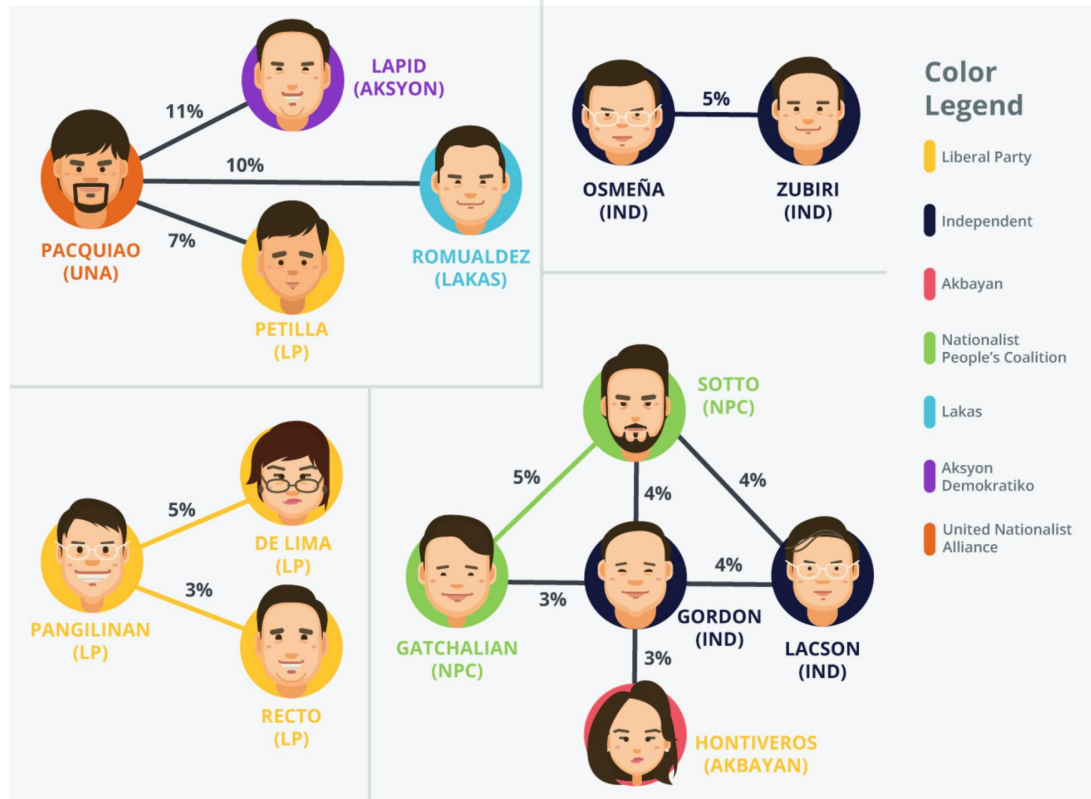
tips for effective data storytelling

- Have at most 3-4 main points and build the story around them—Keep it short and simple

- Use common words instead of jargon

- Show don't tell

- If you must use technical terms, translate and explain them

humanize the data

- People are interested in other people's lives

- Stories about people help readers relate to the data

- People make decisions based on emotions, not logic and reason

High-lift senatorial candidate pairs (Fig. 3)

These graphs show the pairs of candidates with the highest lift. These are candidates who appear in the top 12 together more frequently than expected based on how frequently they appear individually

LAPID (AKSYON)

11%

10%

7%

PACQUIAO (UNA)

PETILLA (LP)

ROMUALDEZ (LAKAS)

OSMEÑA (IND) — 5% — ZUBIRI (IND)

Color Legend

Liberal Party

Independent

Akbayan

Nationalist People's Coalition

Lakas

Aksyon Demokratiko

United Nationalist Alliance

PANGILINAN (LP)

5%

3%

DE LIMA (LP)

RECTO (LP)

SOTTO (NPC)

5%

4%

4%

3%

4%

GATCHALIAN (NPC)

GORDON (IND)

LACSON (IND)

3%

HONTIVEROS (AKBAYAN)

Reina Reyes

"I'm just waiting to reach ideal marrying age."

Being single in your twenties is nothing to worry about. More and more Filipinos are waiting until their 30s to get hitched, a pattern consistent with the global trend. In fact, while there are around 46% and 19% fewer Filipinos in their teens and twenties getting married today than there were about a decade ago, the number of marrying 30-somethings has grown by around 7%.

Changes in the number of marrying Filipinos per age group:
2006 vs. 2015

| Under 20 | 20s | 30s | 40 and Over |
| -46% | -19% | +7% | -15% |
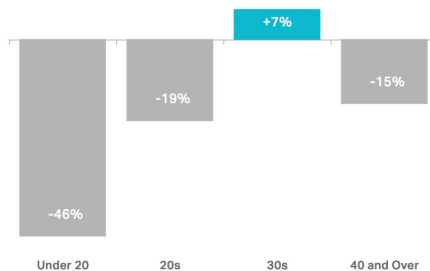
That said, most Filipinos still do get married in their 20s. The most usual marrying age for a Filipina is between 20 to 24, while that of a Filipino man ranges from 25 to 29.

And while plenty of Filipinos still get married outside their twenties, the age skew is different for men than for women. While there are four times as many teen brides as there are teen grooms, on the other end of the age spectrum, there are three times as many senior citizen grooms as there are brides. This suggests that "younger woman - older man" pairings are still pretty common.

https://stories.thinkingmachin.es/ph-marriage/

what's your anchor?

questions?
clarifications?

next week

- All materials are uploaded in our Github and YT playlists

- Feedback on Lesson 3 reflection will be released next week

- Consultation: book via Calendly, will open slots to discuss lecture project and cleaning/analysis techniques

- We'll maximize the sync sessions for next week. Last 4 (+2, tentatively) syncs! We can do thisssss!