Designing
COVIDcast 2.0

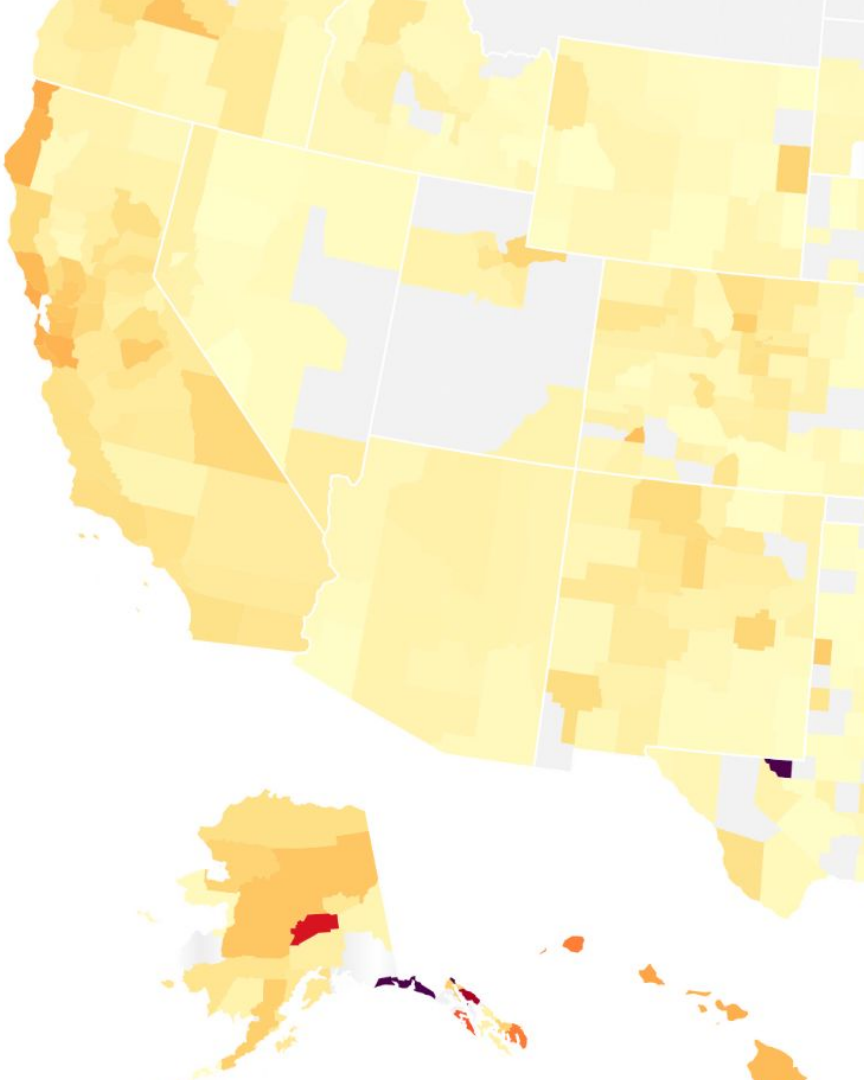
Implementing emergent data visualization designs from the COVID-19 pandemic

Carnegie Mellon University
DELPHI GROUP

Chris Scott & Raphael Hyde
Google.org Delphi Fellows

CS15 - Assessing and Evaluating Data with Visualizations

Thu, Jun 9, 1:15 PM - 2:45 PM



Speakers / Google.org Fellows

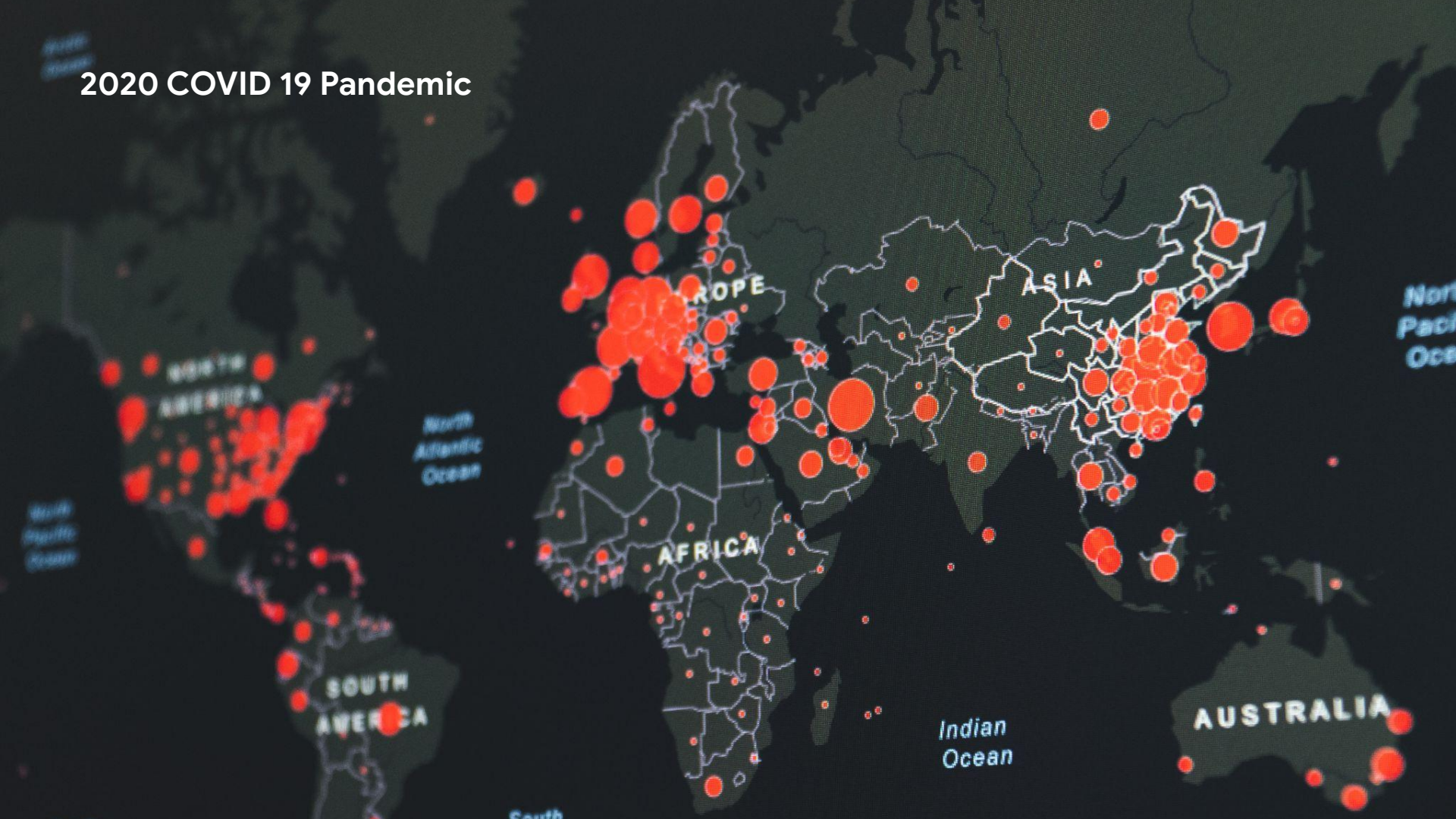
Raphael Hyde

Interaction Designer

Chris Scott

Solutions Engineer

2020 COVID 19 Pandemic



It’s March 2020 and you want to help public health officials make decisions, what do you do?

The Carnegie Mellon University Delphi Group was collecting and processing these data into more useful indicators, but the challenge is, *how can we make this useful to the people who need to make decisions now*?

Background

In May 2020, Google.org **donated $1 million to Carnegie Mellon University** (CMU) to support COVIDcast

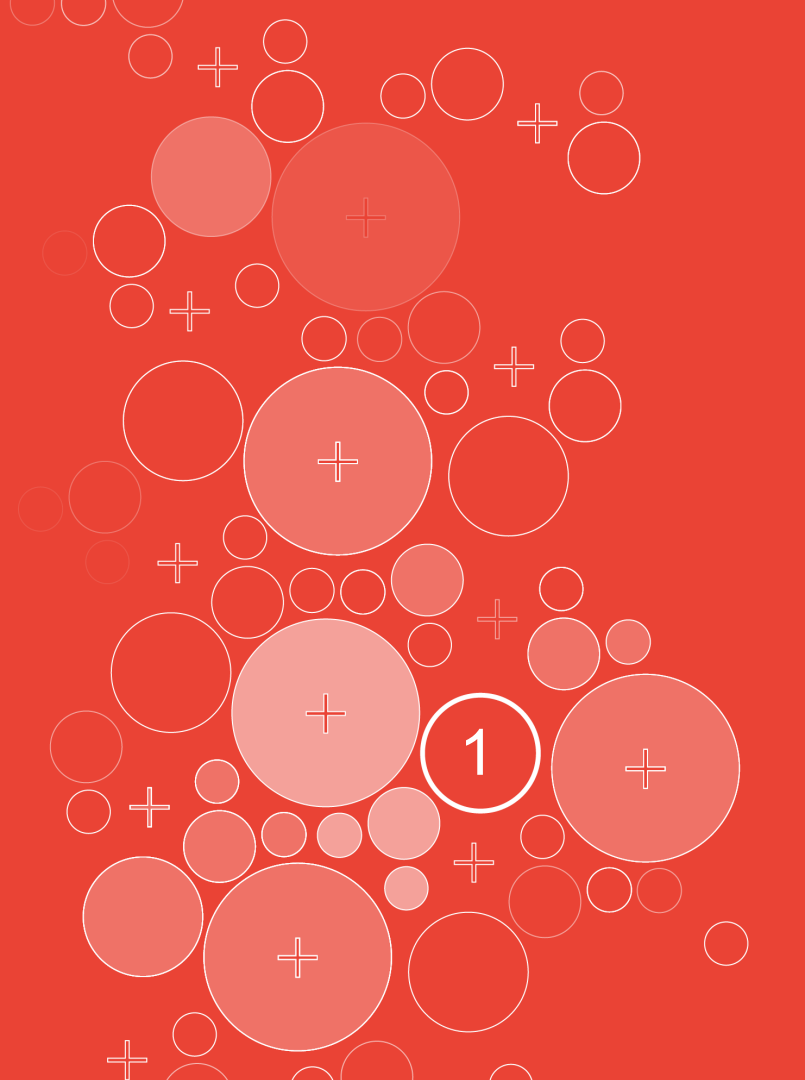
COVIDcast is CMU's effort to track and forecast localized COVID-19 activity nationwide.

Background

Google also provided a team of **12 Google.org Fellows** for a six month, full-time pro-bono engagement to work with the Delphi Group.

The Fellows were embedded in core workstreams to accelerate progress and introduce the Delphi team to leading industry product and software development techniques.

Source link



Speakers / Google.org Fellows

Raphael Hyde

~~Interaction Designer~~

Design Lead

Chris Scott

~~Solutions Engineer~~

Visualization Team Lead

Speakers / Google.org Fellows

Raphael Hyde

~~Interaction Designer~~

Design Lead

Chris Scott

~~Solutions Engineer~~

Visualization Team Lead

Not data scientist

Not statistician

Speakers / DELPHI Team members

Raphael Hyde

~~Interaction Designer~~

Design Lead

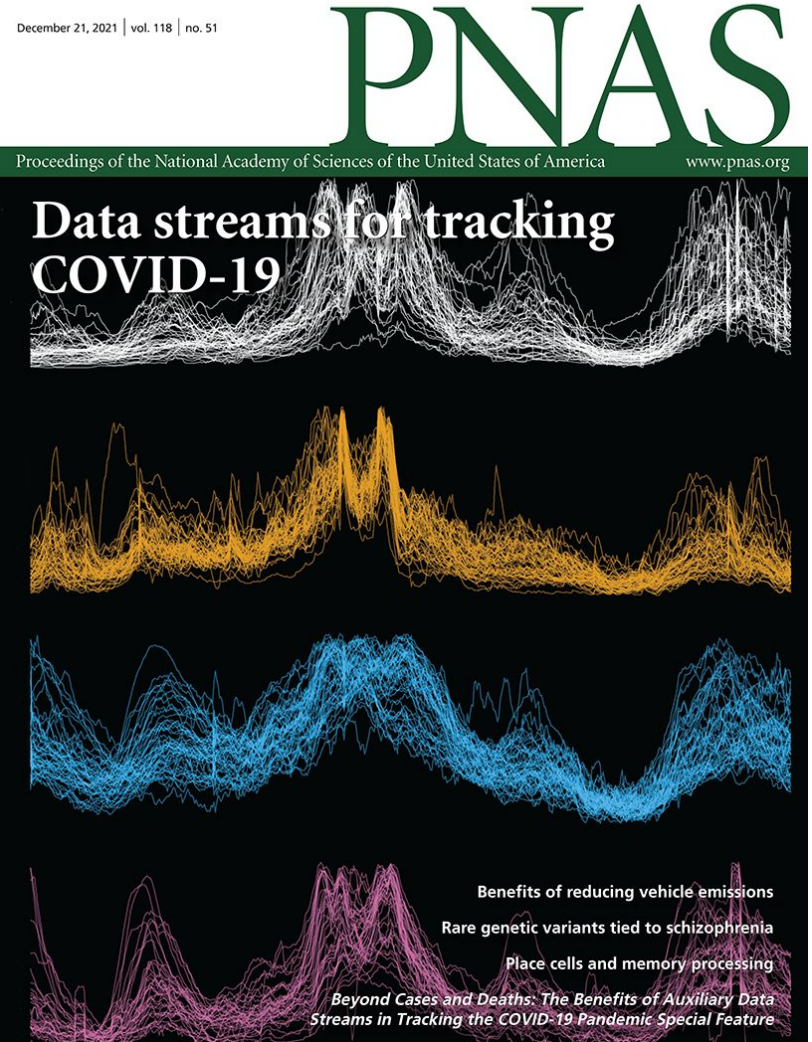
Chris Scott

~~Solutions Engineer~~

Visualization Team Lead

Alex Reinhart

CMU Statistics & Data Science



RESEARCH ARTICLE | MEDICAL SCIENCES | OPEN ACCESS

An open repository of real-time COVID-19 indicators

Alex Reinhart, Logan Brooks, Maria Jahja, +63, and Ryan J. Tibshirani Authors Info & Affiliations

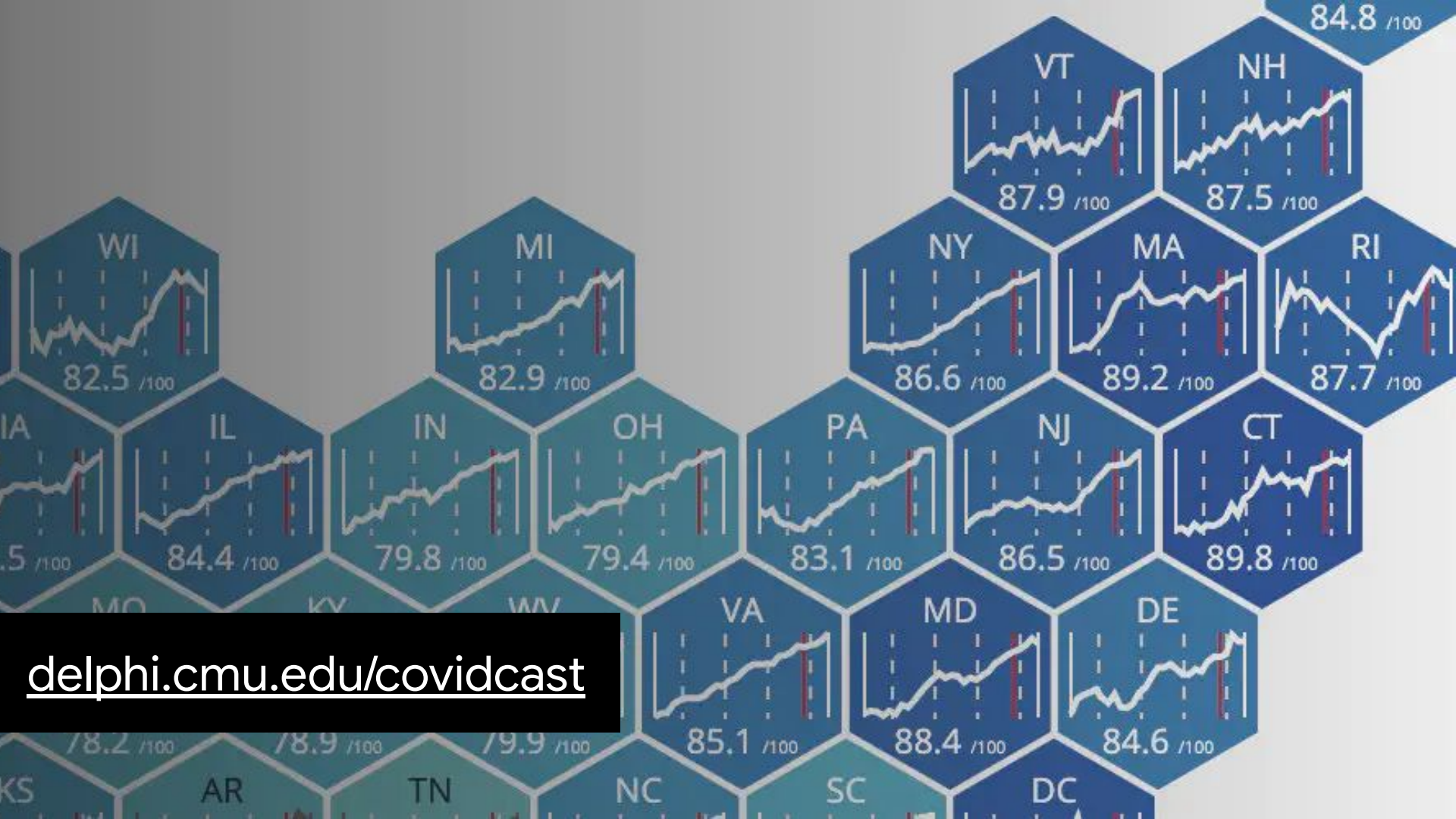
December 13, 2021 | 118(51)e2111452118 | https://doi.org/10.1073/pnas.2111452118

989 2

Significance

To study the COVID-19 pandemic, its effects on society, and measures for reducing its spread, researchers need detailed data on the course of the pandemic. Standard public health data streams suffer inconsistent reporting and frequent, unexpected revisions. They also miss other aspects of a population's behavior that are worthy of consideration. We present an open database of COVID signals in the United States, measured at the county level and updated daily. This includes traditionally reported COVID cases and deaths, and many others: measures of mobility, social distancing, internet search trends, self-reported symptoms, and patterns of COVID-related activity in deidentified medical insurance claims. The database provides all signals in a common, easy-to-use format, empowering both public health research and operational decision-making.

https://www.pnas.org/doi/abs/10.1073/pnas.2111452118



delphi.cmu.edu/covidcast

Overview

A case study in data visualization for a non-technical audience

- Our approach for re-designing a data intensive web application
- How we researched and found some emergent designs
- A component review of how we applied those patterns
- Generalization of those patterns that apply broadly to data visualization

A case study in data visualization for a non-technical audience

- Our approach for re-design a data intensive web application
- How we researched and found some emergent designs
- A component review of how we applied those patterns
- Generalization of those patterns that apply broadly to data visualization

A case study in data visualization for a non-technical audience

- Our approach for re-design a data intensive web application
- How we researched and found some emergent designs
- A component review of how we applied those patterns
- Generalization of those patterns that apply broadly to data visualization

A case study in data visualization for a non-technical audience

- Our approach for re-design a data intensive web application
- How we researched and found some emergent designs
- A component review of how we applied those patterns
- Generalization of those patterns that apply broadly to data visualization

COVIDcast Dashboard 1.0

COVIDcast Platform

- **Indicators**
  - Cases, deaths, hospitalizations
  - Mask wearing, COVID-like symptoms in community
- **Geography**
  - State, County, etc
- **Time**
  - Per day and as-of day

```r
library(covidcast)
library(dplyr)

cli <- suppressMessages(
  covidcast_signal(
    data_source = "fb-survey",
    signal = "smoothed_cli",
    start_day = "2020-05-01",
    end_day = "2020-05-07",
    geo_type = "county"
  )
)
knitr::kable(head(cli))
```

Source: github

Carnegie Mellon University

Search

Delphi | COVIDcast | Data | Blog | GitHub

COVIDcast | Real-time COVID-19 Indicators by the Delphi Group

Our mission • Data and methods • Surveys • Our team • Terms of use

Map Overview | Timelapse | Top 10 | Export Data

Displaying COVID-Related Doctor Visits for Counties on October 8, 2020

Allegheny County, PA

Percentage of daily doctor visits that are due to COVID-like symptoms

Seattle
Portland
San Jose
Fresno
Las Vegas
Los Angeles
San Diego
Phoenix
Tucson
Albuquerque
El Paso
Denver
Oklahoma City
Dallas
Austin
San Antonio
Houston
Milwaukee
Chicago
Detroit
Indianapolis
Columbus
Pittsburgh
Louisville/ Jefferson County
Nashville Davidson
Memphis
Charlotte
Jacksonville
Washington
Philadelphia
New York
Boston

Away from Home 6hr+ a Day

Away from Home 3-6hr a Day

People Wearing Masks

COVID Searches on Google

No data available

COVID-Related Doctor Visits

COVID-Like Symptoms

COVID-Like Symptoms in Community

COVID Indicator Combination

COVID Antigen Test Positivity (Quidel)

COVIDcast Dashboard 1.0

Features

- Optimized for power-users
- Powerful data visualizations
- **Unique indicators** (data not available anywhere else*)

Opportunities

- No mobile experience / Unresponsive
- Overly complex UI on landing page
- **Not approachable for non-technical users**
- Poor feature discoverability
- Inconsistent narrative
- Lack of summary data or at-a-glance features
- Backend required an overhaul

* See PNAS paper for more details

How could we make COVIDcast more useful?

Designing COVIDcast Dashboard 2.0

Audience Targeting

Questions

Who are our **users**?

What do they **want**?

What are our **goals**?

Who are our users

January 2021

- Primary: **Analyst/researcher**
  - “I want to see broad trends”
  - “...and then download the data”
- Secondary: **Decision maker**
  - “I want to see an overview in my area”
- Tertiary?: **Journalist**
  - I want to identify newsworthy trends

Potential users

General Public
Curious citizen data scientists and non-technical users

Public health officials
Political offices, State department, etc

Practitioners
Doctors, Hospital directors & Epidemiologists

Scientists
Researchers and lab partners

Data Journalists
Mainstream Media and Bloggers

Segmented audience

Sophisticated Data Consumers

Practitioners
Doctors, Hospital directors & Epidemiologists

Public health officials
Political offices, State department, etc

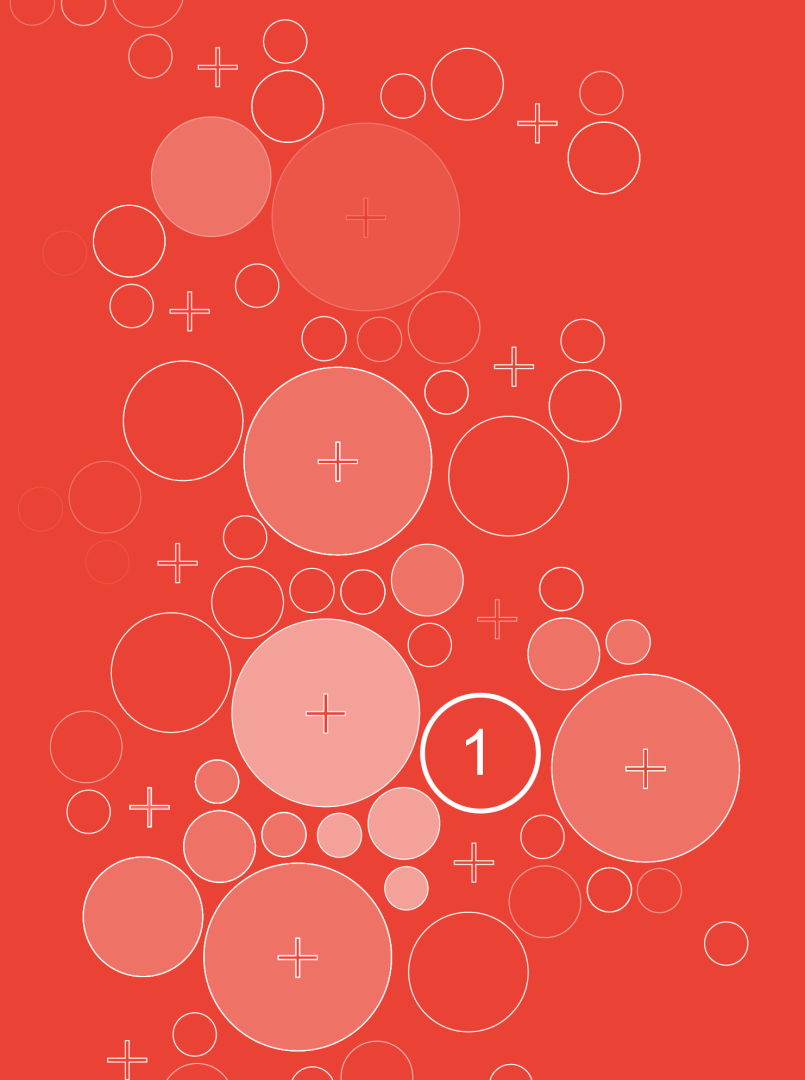
Scientists
Researchers and lab partners

Data Journalists
Mainstream Media and Bloggers

General Public

Curious citizen data scientists and non-technical users

...but generally, everyone became a more sophisticated data consumer



Design guidelines

- Approachable design, but we can **afford to be sophisticated**
- Introduce gradual complexity and **drill-down-ability**
- Afford both **global context** and **local trends**

Goals

- Complement, don’t compete with, other sites
- Other sites better serve the general public, we serve sophisticated data consumers
- Serve as the introduction to and reference implementation of the COVIDcast API
- Design for the long-term

Emergent Pattern Discovery

What exists today?

“We didn’t want to create yet another covid map”

January 2021

Comparative Design Review

Site Design Deep Dive by UX Design Team

Johns Hopkins COVID-19 Map

[Website]

The COVID-19 Testing Insights Initiative reflects an interdisciplinary collaboration between several groups at Johns Hopkins University, first publicly shared on January 22.

Core user: Researchers, public health authorities, and the general public

ANNOTATIONS / UX NOTES:

1. Johns Hopkins Coronavirus Resource Center logo placed on top left
2. Top navbar that contains general information
3. Secondary navbar with second home page dedicated to “Tracking” global cases and trends
4. Displays total cases/deaths in numerical value
5. Bottom tabs with different cases/rates

January 2021

CDC COVID Data Tracker

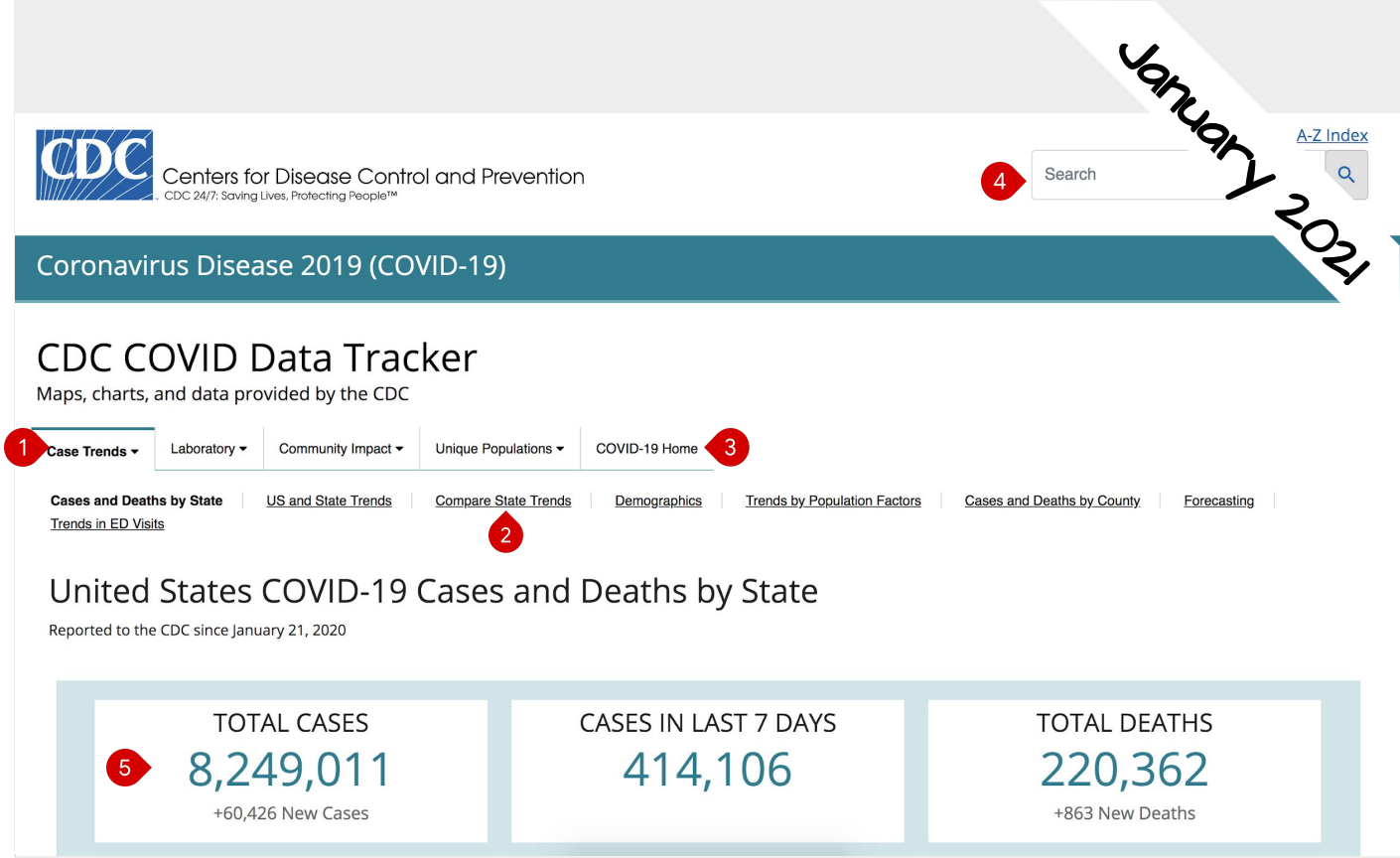
Website

A website that allows users to interact with a variety of data on COVID-19 that is updated daily. The website builds on other agency efforts - such as CDC's new weekly COVID-19 Surveillance report, COVIDview - to capture the impact the virus is having on the United States.

Core user: Healthcare workers, laboratories

ANNOTATIONS / UX NOTES:

1. Uses dropdown-like tabs to display numerous pages
2. Allows users to compare trends between states
3. Additional tab leads to CDC's main Home page
4. Identical search bar placement for the rest of other navigation pages
5. Displays total cases/deaths in numerical value



COVID Act Now

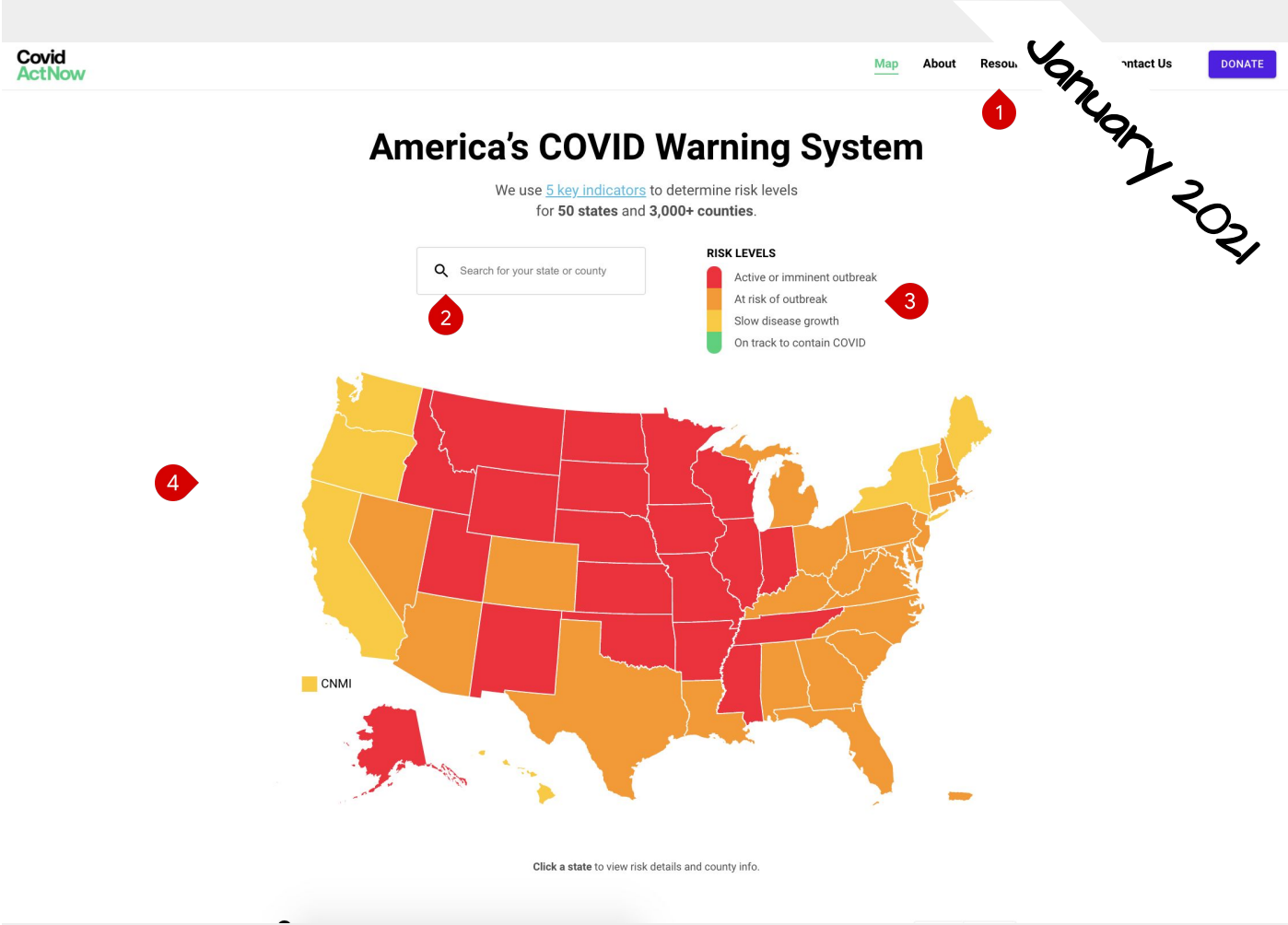
Website

A volunteer-driven 501c3 non-profit, which aims to provide the best-available local-level disease intelligence and data analysis on COVID in the U.S.

Core user: Technologists, epidemiologists, public health experts, and public policy leaders

ANNOTATIONS / UX NOTES:

1. Main nav bar placed on top is very user-friendly and easy to spot
2. Search bar helps to look up for state/county
3. Straightforward, easy to understand key legends placed on top
4. No state/county names shown on map



IHME COVID Forecast

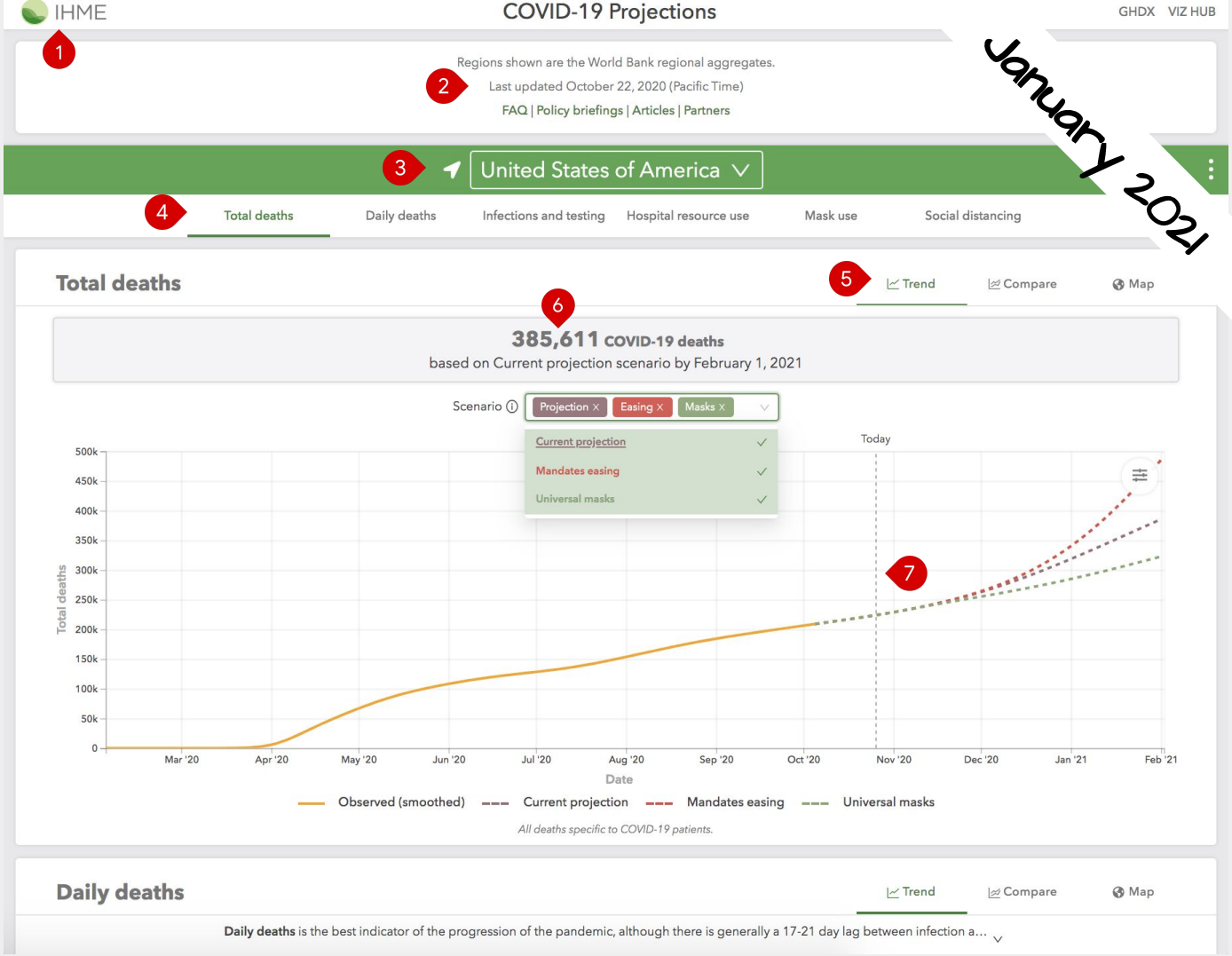
Website

Developed in response to requests from the University of Washington School of Medicine and other US hospital systems and state governments working to determine when COVID-19 would overwhelm their ability to care for patients.

Core user: Policymakers, US hospital systems, and state governments

ANNOTATIONS / UX NOTES:

1. IHME logo directs to main site
2. Displays last updated → builds trust and increases credibility
3. Search bar automatically set USA as default → when clicked, a dropdown appears with all global regions
4. When tab is clicked, the page automatically scrolls down to the designated section
5. Secondary tab designed for showing "Trend" charts, "Compare" graphs, and then "Map" (trend chart is the first view)
6. Total number of deaths displayed with bolded emphasis
7. Uses pointer to help viewer guide through different the dates on the chart



Bing COVID-19 Tracker

[Website]

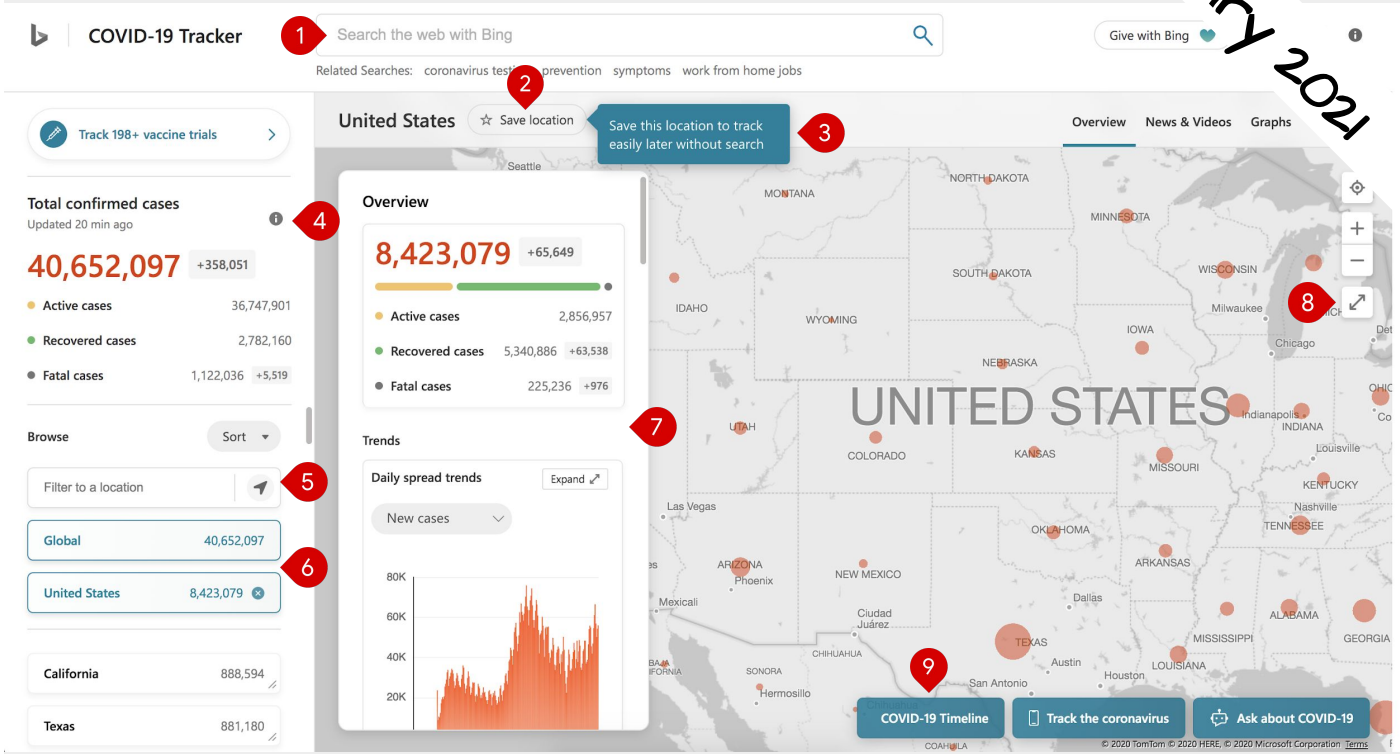
Launched by Microsoft, this tracker allows users to track COVID-19 local and global coronavirus cases with active, recoveries and death rate on the map, with daily news and video.

Core user: Public health officials, general public

ANNOTATIONS / UX NOTES:

1. The main search tool leads to the main bing website results
2. “Save location” button is interesting and something unique
3. Uses coachmarks for first time users to direct them
4. Location selection panel on left navbar
5. Search tool specifically for this tracker page is placed within the side navbar filter
6. Displays total confirmed cases along with location names
7. Displays a pop up of statistic overview, trends charts, as well as articles when scrolled down
8. Can hide side navbar (location selection panel) to see full map
9. When clicked COVID-19 Timeline, leads to a new screen (see next slide)

January 2021



World Health Organization (WHO)

Website

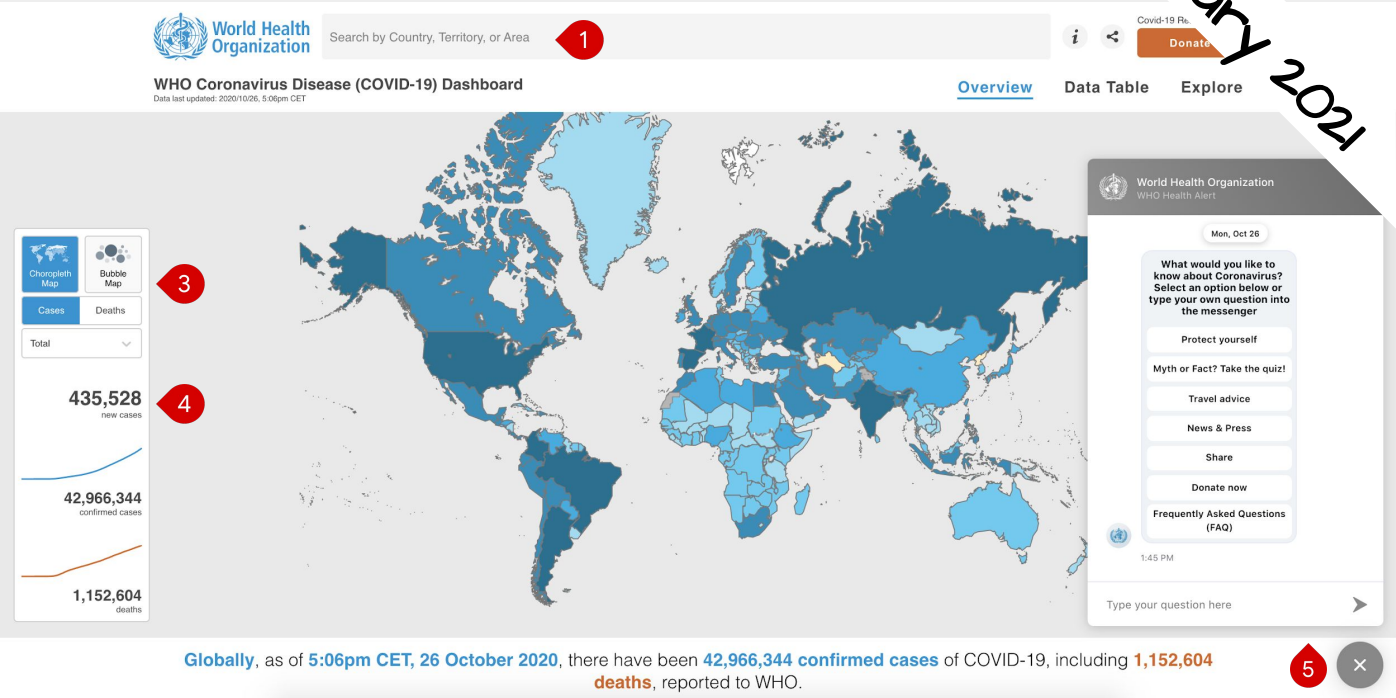
The WHO coronavirus disease (COVID-19) dashboard presents official daily counts of COVID-19 cases and deaths reported by countries, territories and areas. Through this dashboard, we aim to provide a frequently updated data visualization, data dissemination and data exploration resource, while linking users to other useful and informative resources.

Core user: National public health authorities

ANNOTATIONS / UX NOTES:

1. Search bar placed on very top to help users search country or are at any page navigation
2. Very simple/straightforward navbar
3. Left click-through menu(?) provided for users to view (either choropleth or bubble, cases or deaths)
4. Displays the total number of new cases and confirmed cases/deaths on the side
5. Friendly alert chatbot system implemented

January 2021

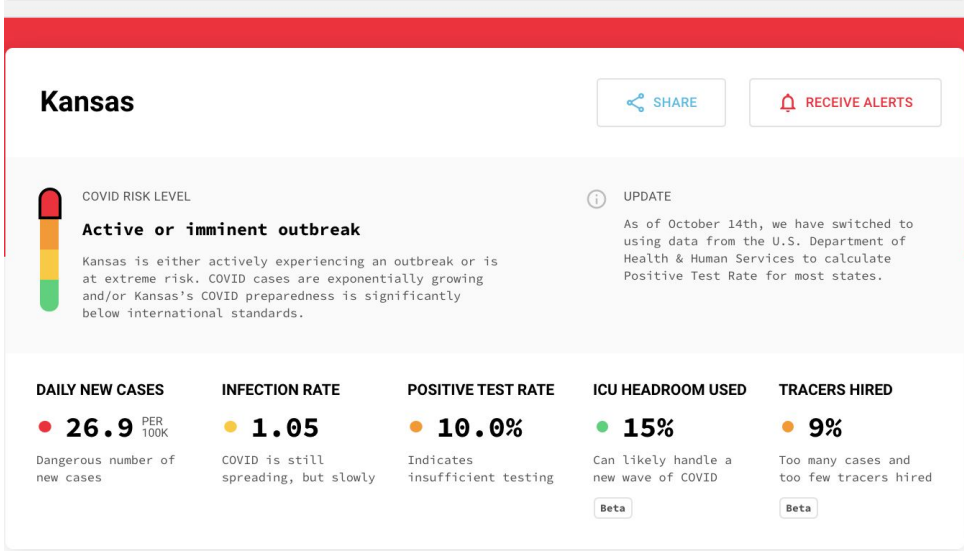


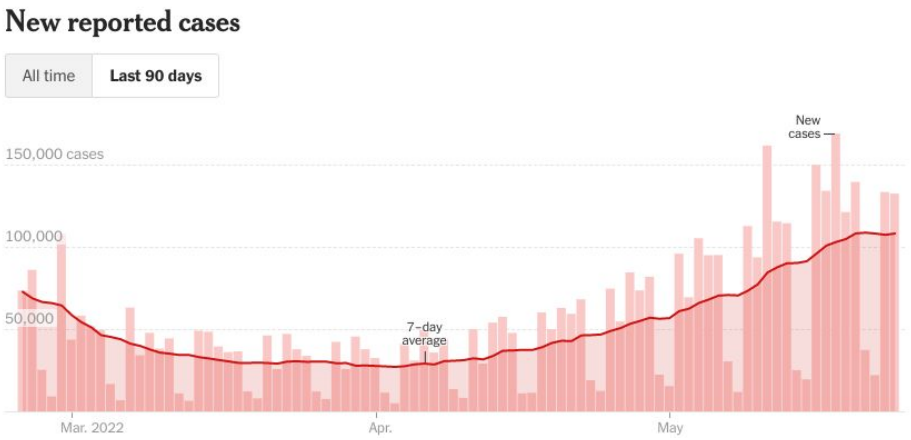
Emergent Patterns

Emergent Patterns

2 main exemplars

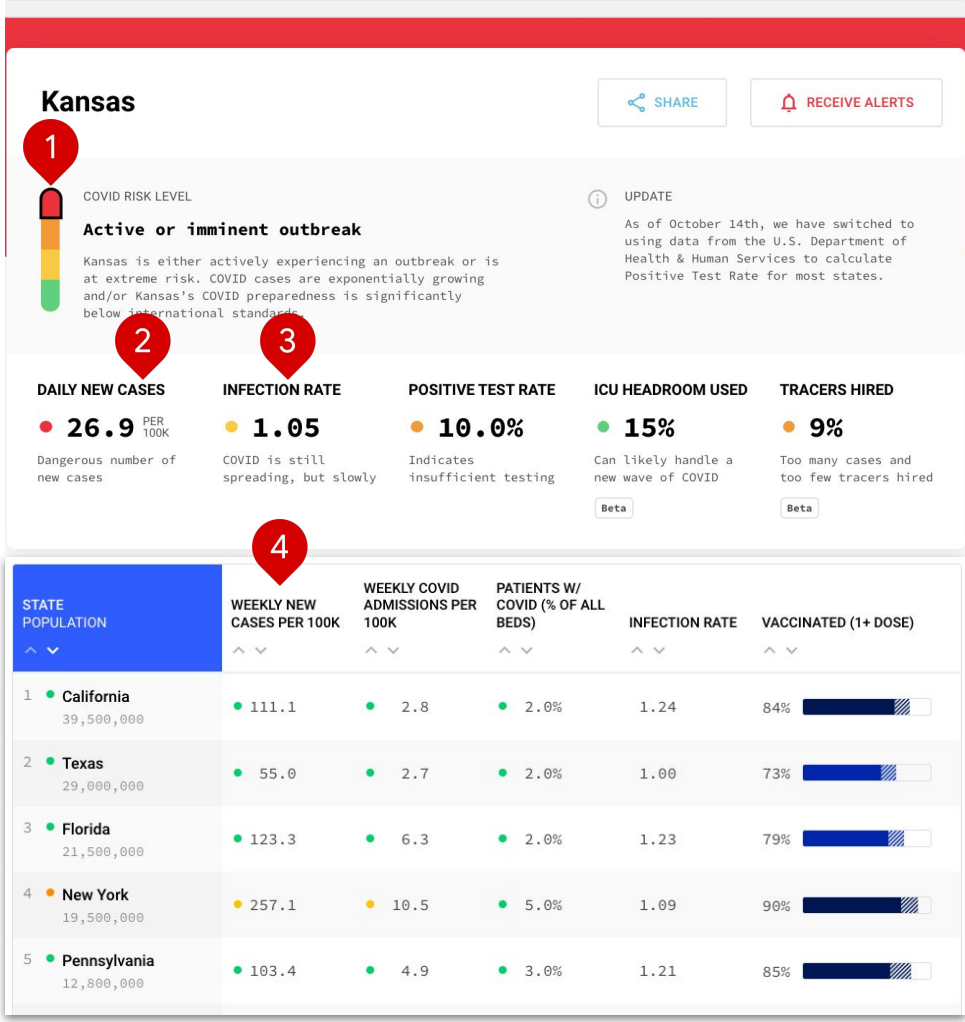
1. COVID Act Now
2. New York Times





Emergent Patterns

1. High-level summary
2. Case count in PER 100K
3. New metrics, e.g. Rt
4. Rich tables



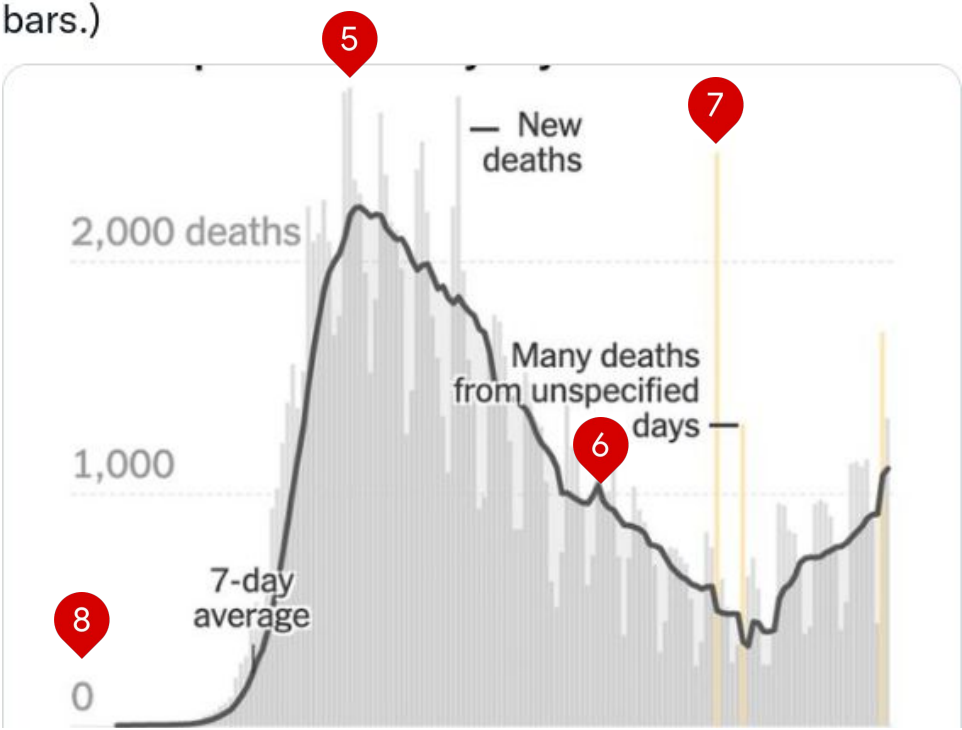
https://covidactnow.org

Emergent Patterns

1. High-level summary
2. Case count in PER 100K
3. New metrics
4. Rich tables
5. **Daily reporting**
6. **Smoothed Trends**
7. **Highlighting Outliers**
8. **Simplified charts**

John Lorinc
@JohnLorinc

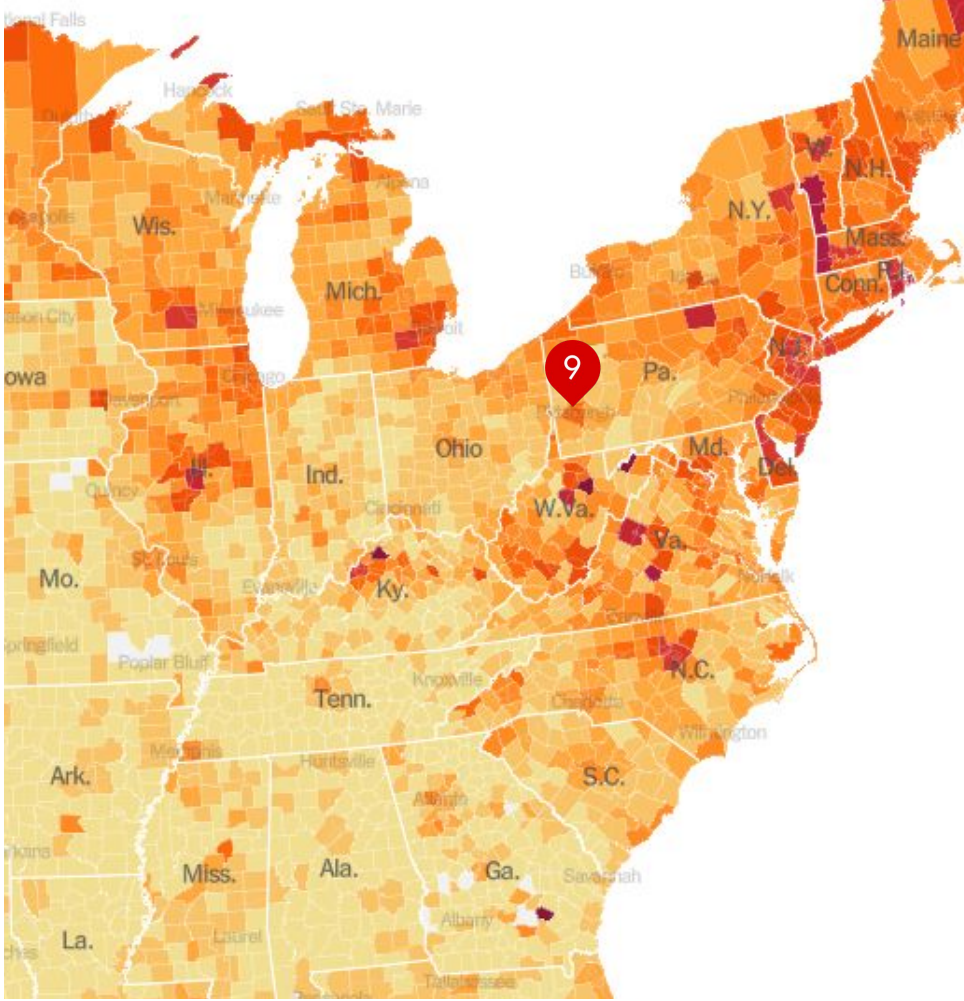
From @nytimes: US death toll, Covid. Anyone out there understand why there would be a weekly spike pattern on this chart? (Look at the light gray daily total bars.)



https://twitter.com/JohnLorinc/status/1288547655258341377

Emergent Patterns

1. High-level summary
2. Metrics in PER 100K
3. New metrics
4. Rich tables
5. Daily reporting
6. Smoothed Trends
7. Highlighting Outliers
8. Simplified charts
9. **Comparative geospatial PER 100K Choropleth**



Source: NYTimes

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

Global Cases
41,074,593

Cases by Country/Region/Sovereignty

- 8,319,271 US
- 7,651,107 India
- 5,273,954 Brazil
- 1,438,219 Russia
- 1,018,999 Argentina
- 1,005,295 Spain
- 999,744 France
- 974,139 Colombia
- 874,118 Peru
- 860,714 Mexico
- 792,194 United Kingdom

Admin0

Last Updated at (M/D/YYYY)
10/21/2020, 5:24 PM

Cumulative Cases | Active Cases | Incidence Rate | Case-Fatality Ratio | Testing Rate

189 countries/regions

Lancet Inf Dis Article: Here. Mobile Version: Here. Data sources: Full list. Downloadable database: GitHub, Feature Layer.
Lead by JHU CSSE. Technical Support: Esri Living Atlas team and JHU APL. Financial Support:

Global Deaths
1,128,226

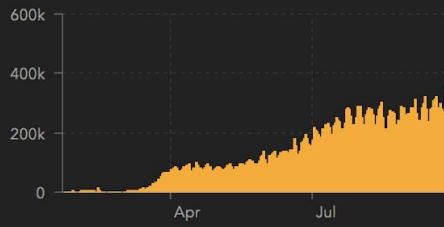
- 221,862 deaths US
- 154,837 deaths Brazil
- 115,914 deaths India
- 86,893 deaths Mexico
- 44,248 deaths United Kingdom

Global Deaths

US State Level
Deaths, Recov

- 33,371 deaths, 78, recovered New York US
- 17,649 deaths, 733 recovered Texas US
- 17,124 deaths, rec California US
- 16,245 deaths, 36, recovered New Jersey US

US Deaths, Rec



Daily Cases

Design Process

~~Overview~~ | Map | Time Trends | Nowcast | Survey | Export Data

Date: December 7, 2020 ▽

Location: U.S. National ▽

Compare to: ~~7 days ago~~ | 14 days ago | custom ...

Note: All numbers shown below are *estimates* of activity on the given day. Read more ...

X New COVID-19 Cases (?) ↗ Up A% from Nov 30, 2020	Y New COVID-19 Deaths (?) ↗ Up B% from Nov 30, 2020
W% of People Self-Reporting COVID-Like Illness (?) ↘ Down C% from Nov 30, 2020	Z% of People Reporting They Know Someone in Community w/ COVID-Like Ill (?) ↗ Up D% from Nov 30, 2020
U% of Doctor Visits w/ COVID-Like Illness (?) ↘ Up E% from ...	V% of Hospital Admissions w/ COVID Diagnosis ↗ Up F% from ...

Choose a state:

AK	AL	AR
AZ	CA	CO
	...	

Advanced Options

Summary: 5 Indicators Up (an Average of G%)
1 Indicator Down (an Average of H%)

Overview | Map | ~~Time Trends~~ | Nowcast | Survey | Export Data

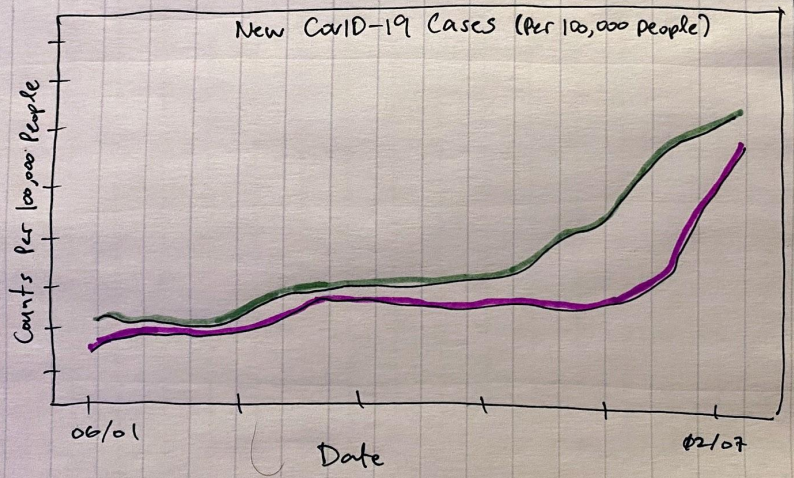
View: Grid | ~~Side Panel~~

Locations: California x | Pennsylvania x

Add location +

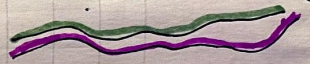
Date Range: June 1 2020 ▽ to Dec 7, 2020 ▽

Color legend
- California
- Pennsylvania



Advanced Options

⊕ New CoVID-19 Case Rates ?



⊕ New CoVID-19 Death Rates ?

⊕ % of People with CLI ?

...

⊕ % of People who know someone in community w/ CLI ?

...

⊕ % of Doctor Visits due to CLI ?

...

⊕ % of Hospital Admissions w/ CoVID Diagnosis ?

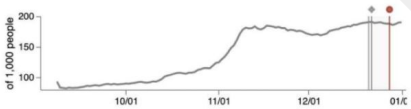
...

January 2021

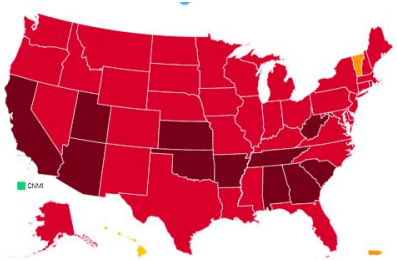
Visualization primitives

Metrics card

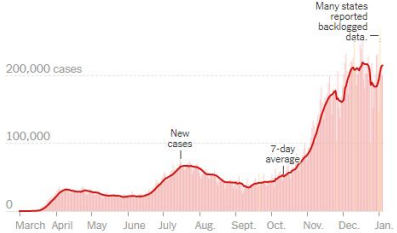
Simple time-series



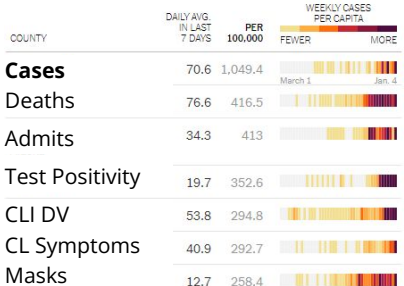
Choropleth Map



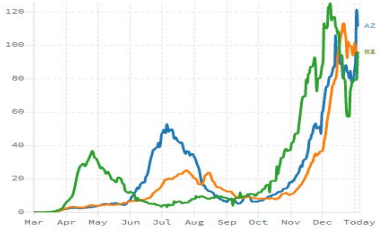
Combo bar/line Time-series



Rich Tables

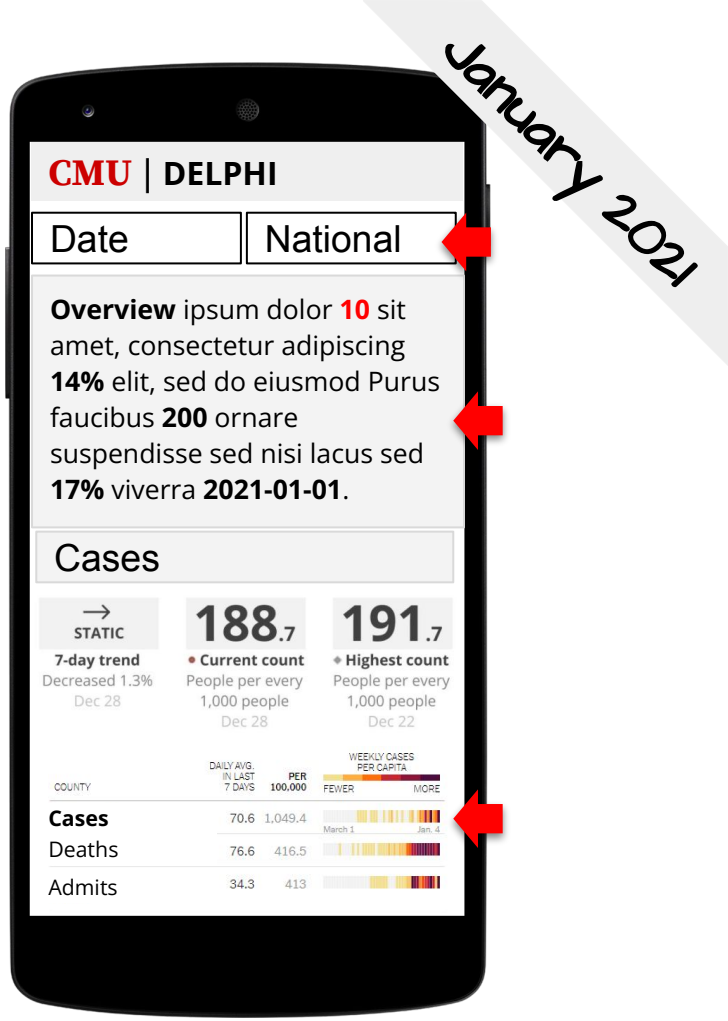


Comparison Time-series



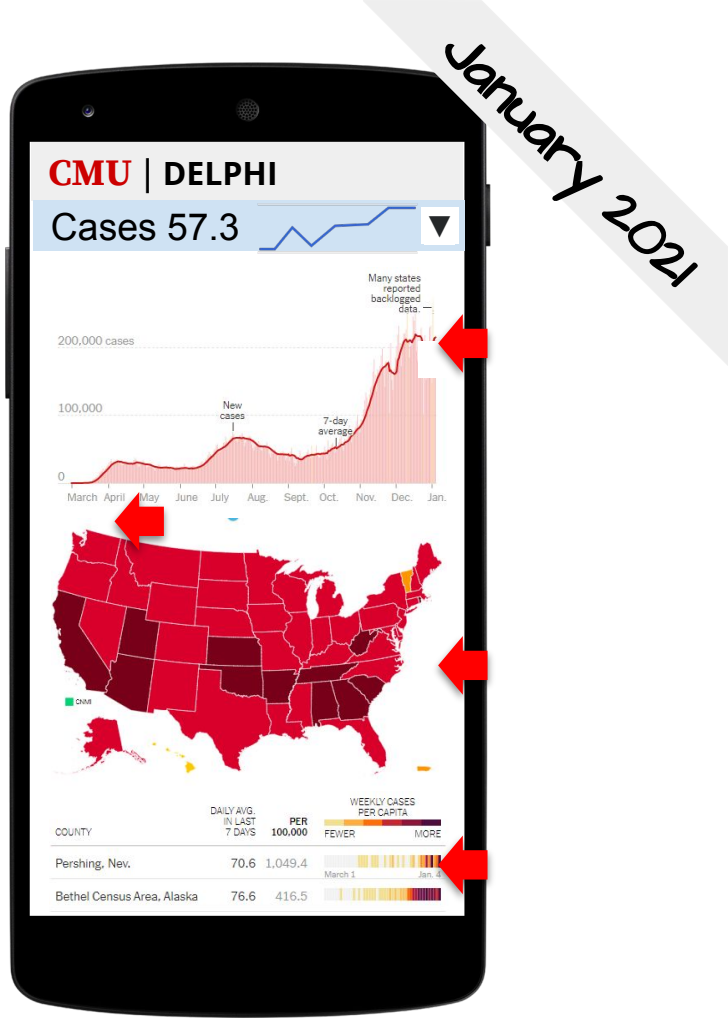
Signal Centric Overview

- At the top of the page, we have a simple overview describing the movement of all of our signals
- Next, we have our “signals” section
  - It is important to have a clear description what the signal is about and how to interpret the data
- **Design Rule**:
  - Include textual descriptive statistics



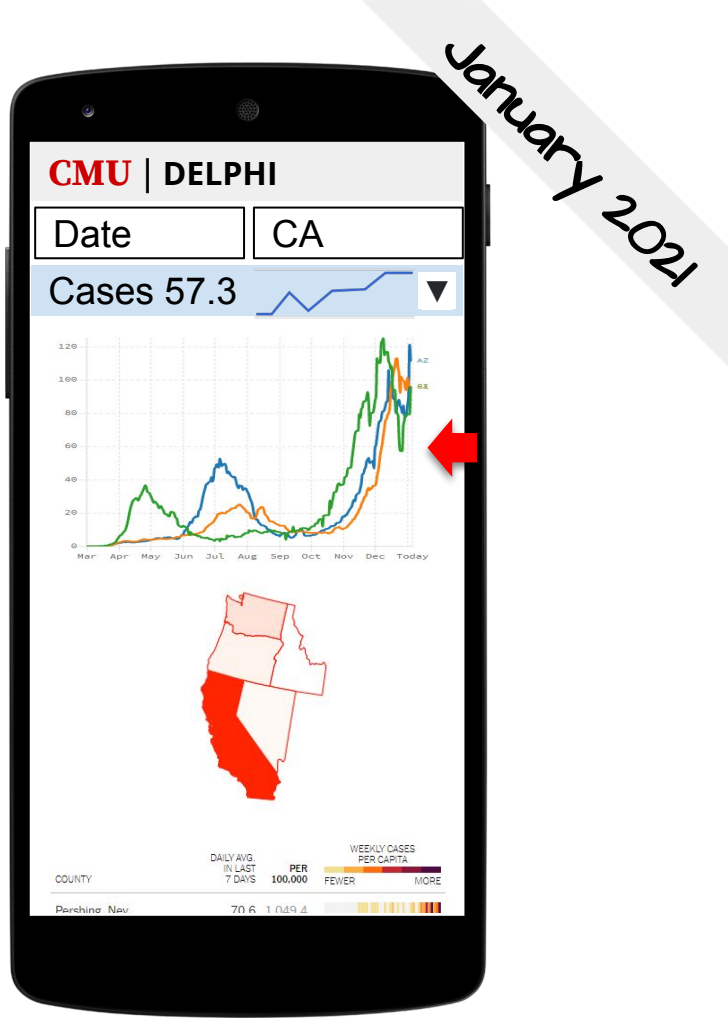
Geography comparison

- Time series shows overall movement
- Sub-geography is exposed as a Rich Table
- Timelapses can be accomplished by having a play button next to the time series, or simply by interacting with the plot



Sub-Geography

- **Design Rule**: All sub-geography types should start at a "surrounding areas" context
- Comparison time-series shows location vs. "surrounding" vs. parent
  - E.g. Allegheny County vs. Surrounding county average vs. PA
  - Also allow arbitrary comparisons via a location picker.

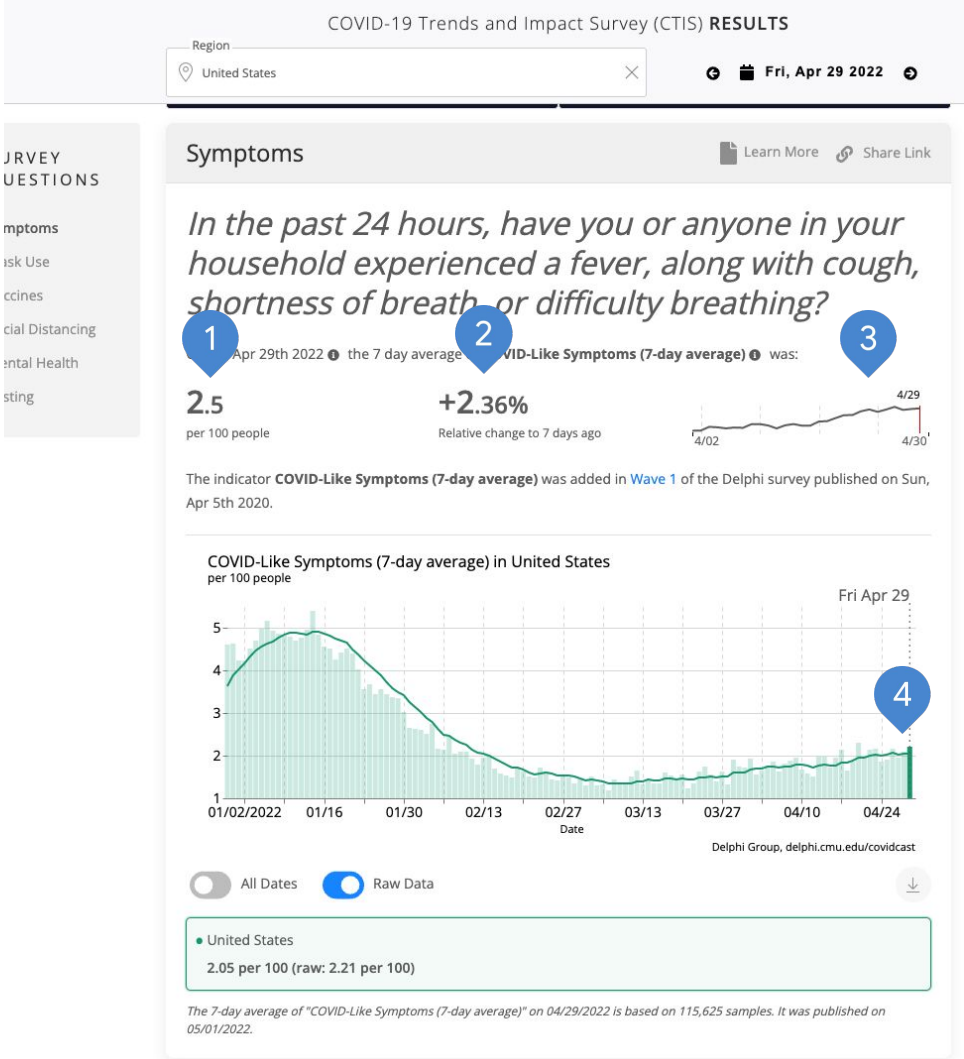


Component Analysis

Question Card

Question Card

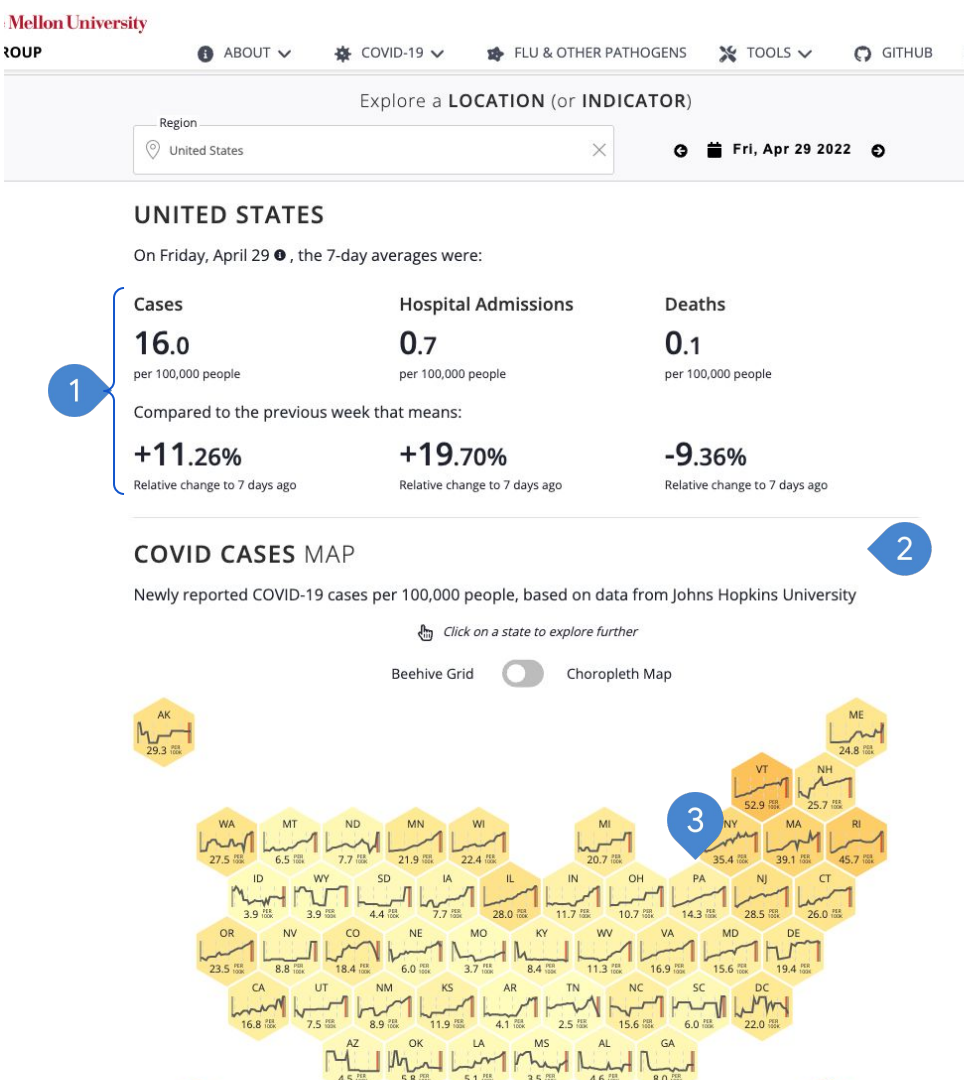
1. Value
2. Change over unit time
3. Sparkline provides context
4. Full time-series invites deeper investigation and comparisons



Location Overview

Location Overview

1. Re-use summary pattern from question card
2. Then start to provide deeper context
3. Hex-grid gives good high-level at-a-glance info



Location Overview

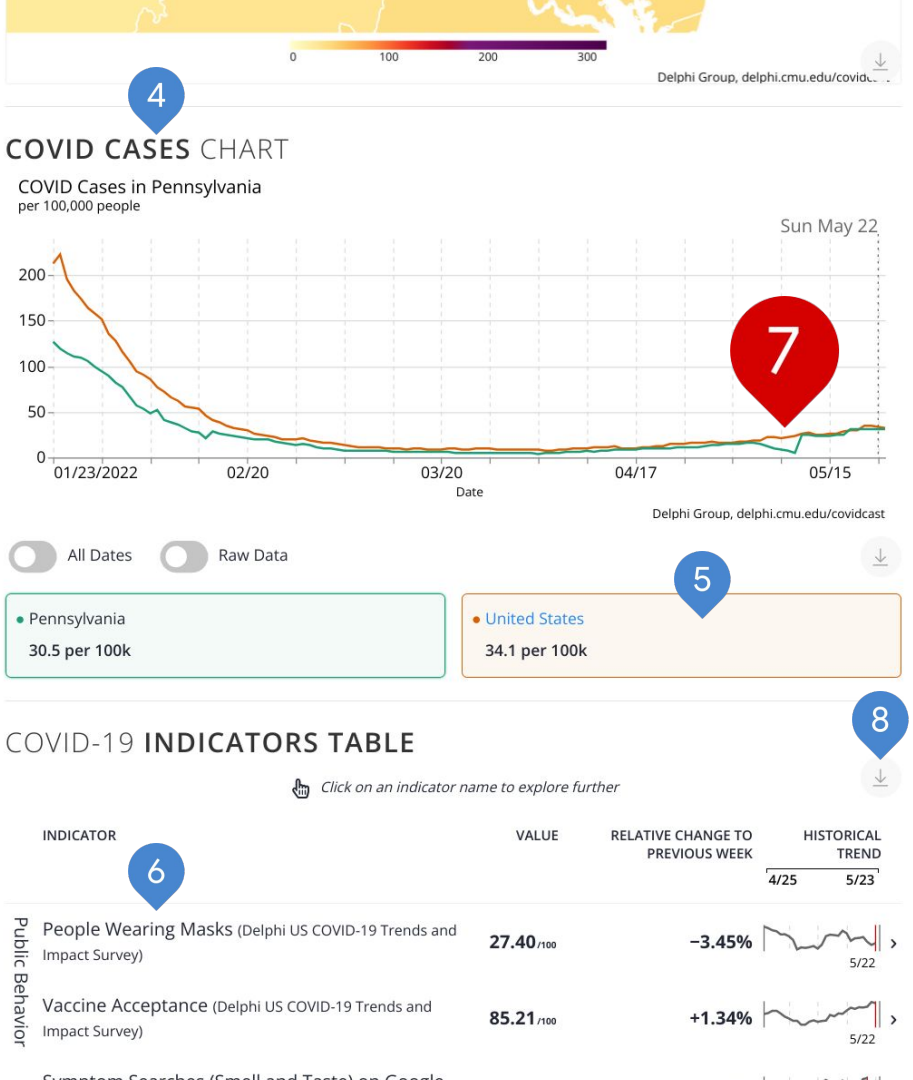
4. Then provide more precise temporal data

5. With geo-comparisons

6. Filled by drill-down links

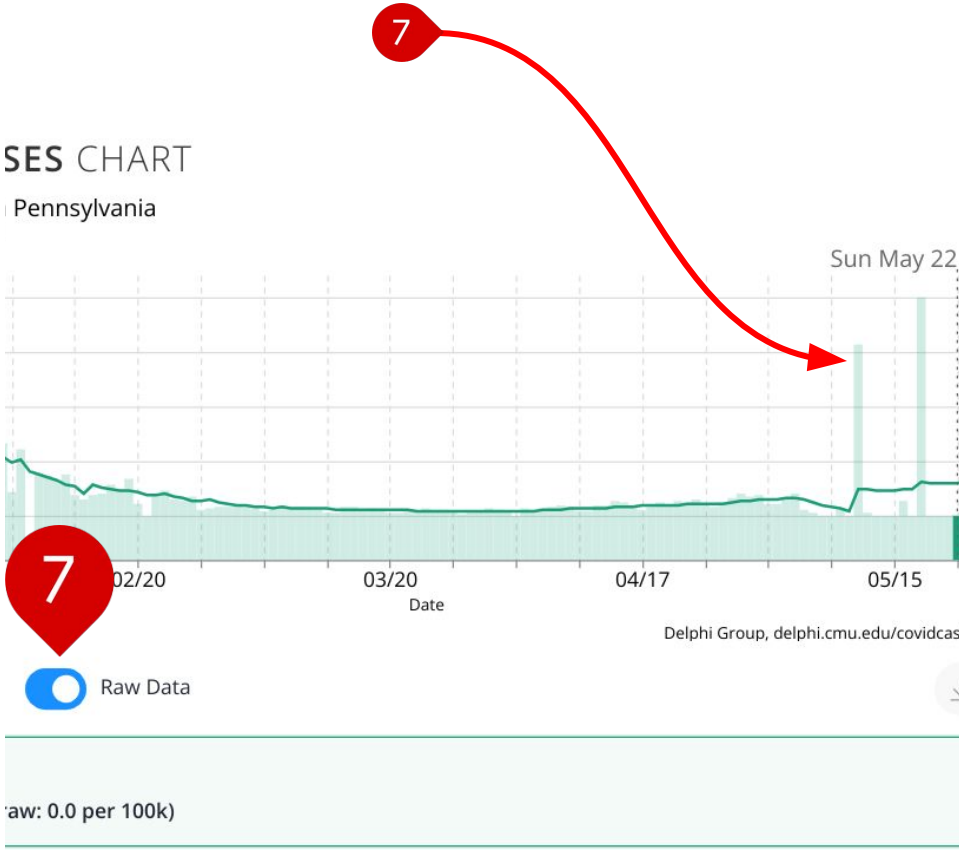
7. Note this odd feature

8. Download link!



Location Overview

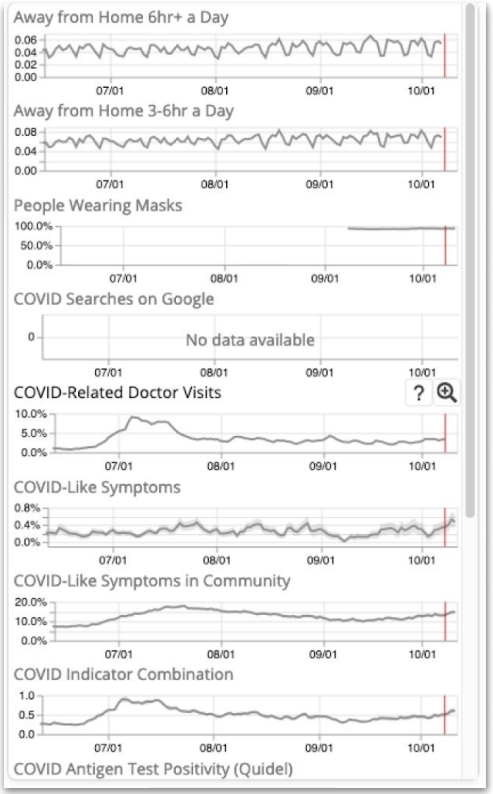
7. By switching to alternate views we can yield additional insights



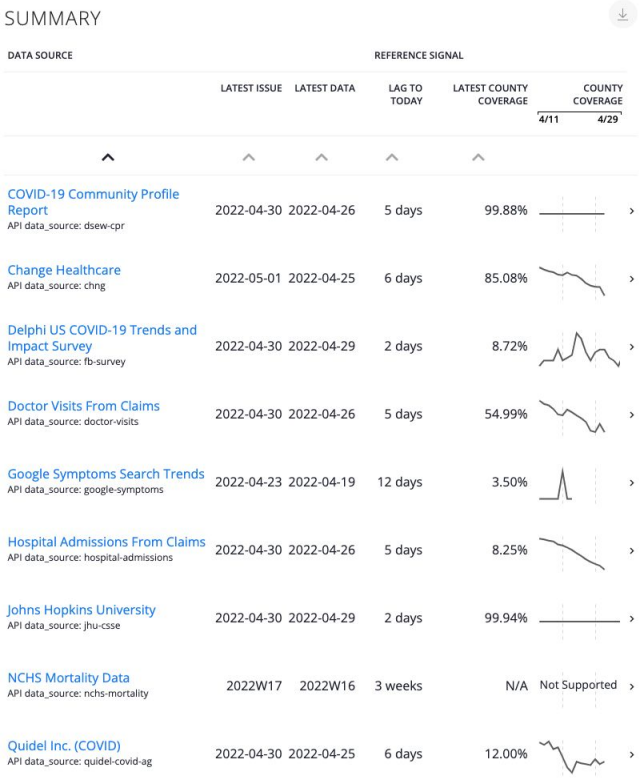
Indicator Table

Location Overview

Before



After

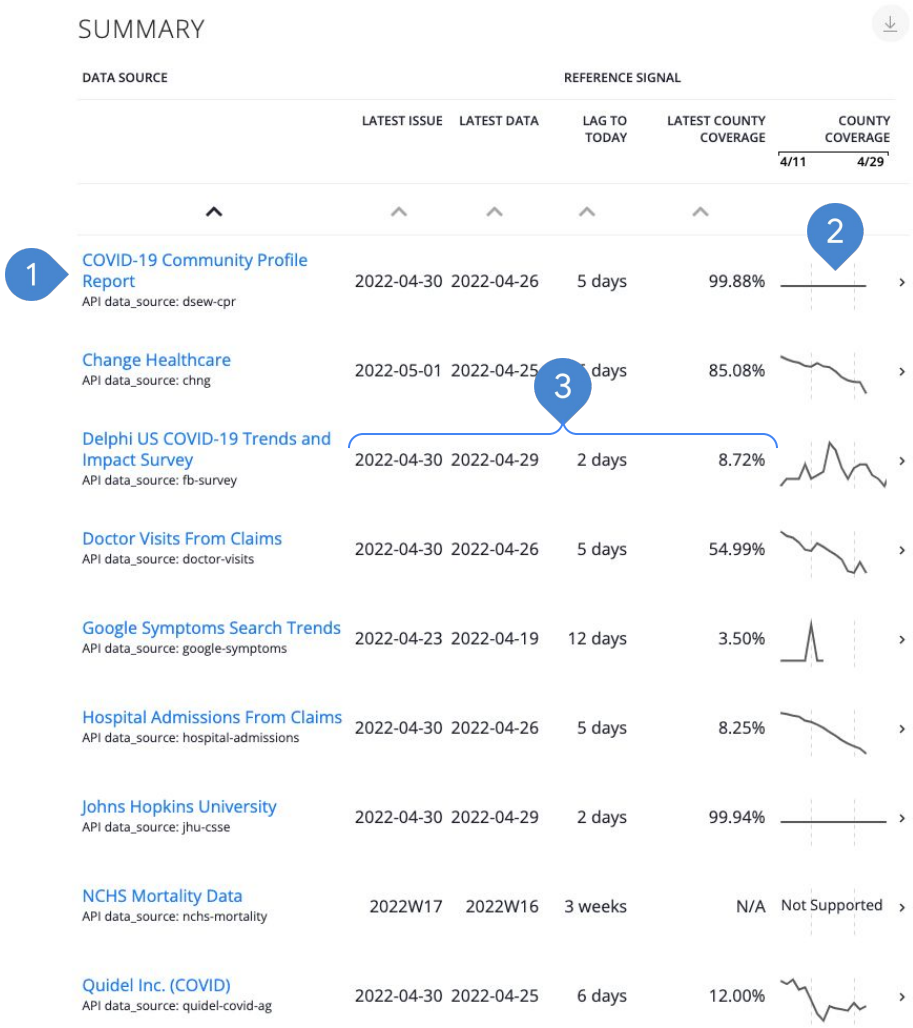


SUMMARY

DATA SOURCE	LATEST ISSUE	LATEST DATA	REFERENCE SIGNAL: LAG TO TODAY	REFERENCE SIGNAL: LATEST COUNTY COVERAGE	REFERENCE SIGNAL: COUNTY COVERAGE (4/11 – 4/29)
COVID-19 Community Profile Report API data_source: dsew-cpr	2022-04-30	2022-04-26	5 days	99.88%	
Change Healthcare API data_source: chng	2022-05-01	2022-04-25	6 days	85.08%	
Delphi US COVID-19 Trends and Impact Survey API data_source: fb-survey	2022-04-30	2022-04-29	2 days	8.72%	
Doctor Visits From Claims API data_source: doctor-visits	2022-04-30	2022-04-26	5 days	54.99%	
Google Symptoms Search Trends API data_source: google-symptoms	2022-04-23	2022-04-19	12 days	3.50%	
Hospital Admissions From Claims API data_source: hospital-admissions	2022-04-30	2022-04-26	5 days	8.25%	
Johns Hopkins University API data_source: jhu-csse	2022-04-30	2022-04-29	2 days	99.94%	
NCHS Mortality Data API data_source: nchs-mortality	2022W17	2022W16	3 weeks	N/A	Not Supported
Quidel Inc. (COVID) API data_source: quidel-covid-ag	2022-04-30	2022-04-25	6 days	12.00%	

Location Overview

1. Sortable list of indicators is easier to browse
2. Retain trend information with sparklines, bars, etc.
3. Introduce quality metrics

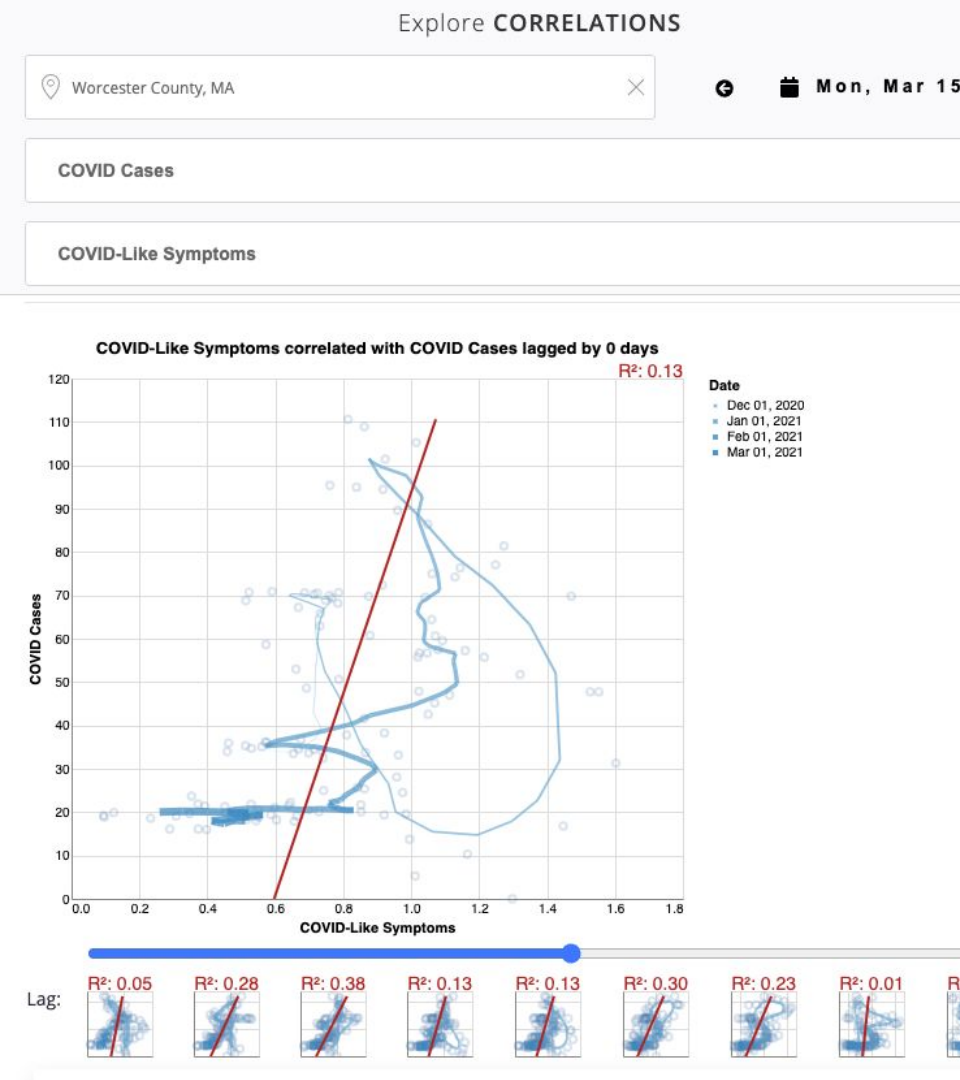


Correlation Explorer

Correlation Explorer

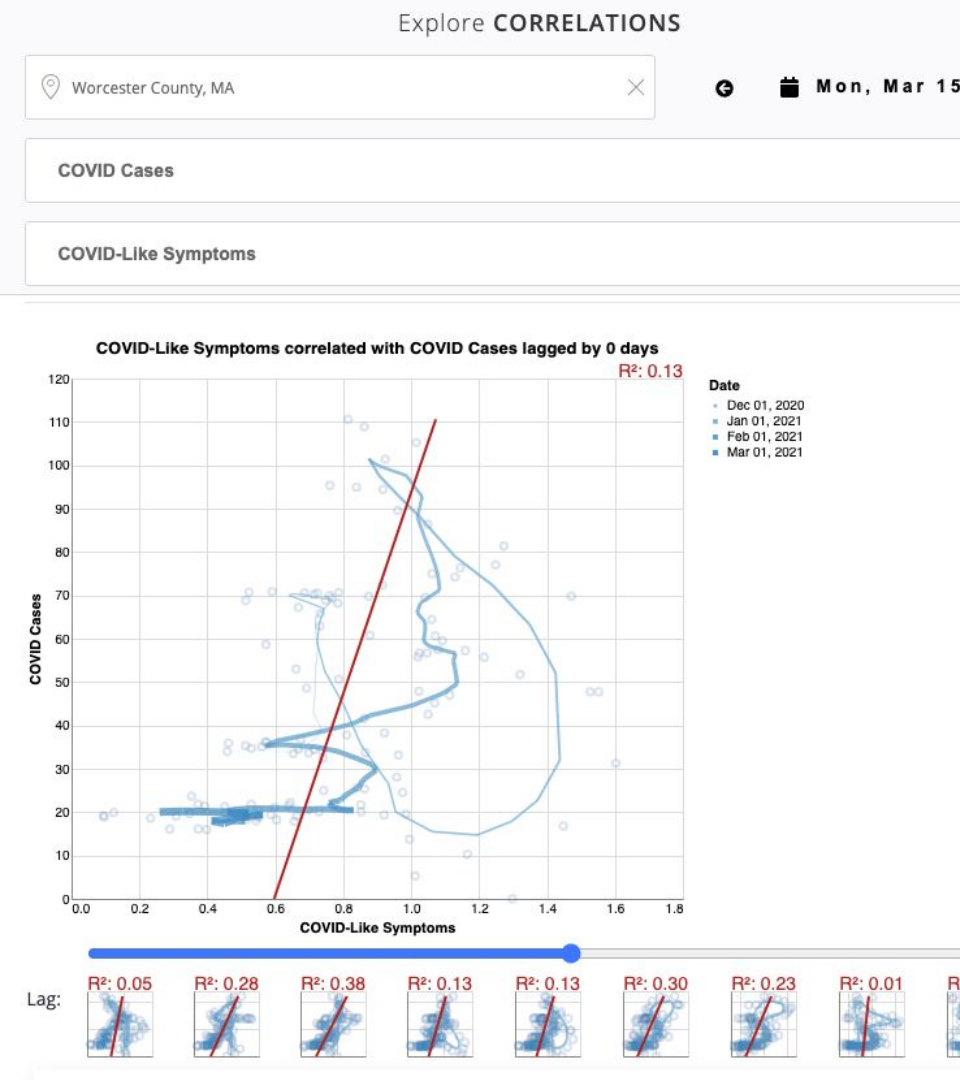
Motivation

- “How long after a spike in cases does do hospitalizations increase?”
- “What other leading indicators are there”
- “How does this relationship change over time and location?”



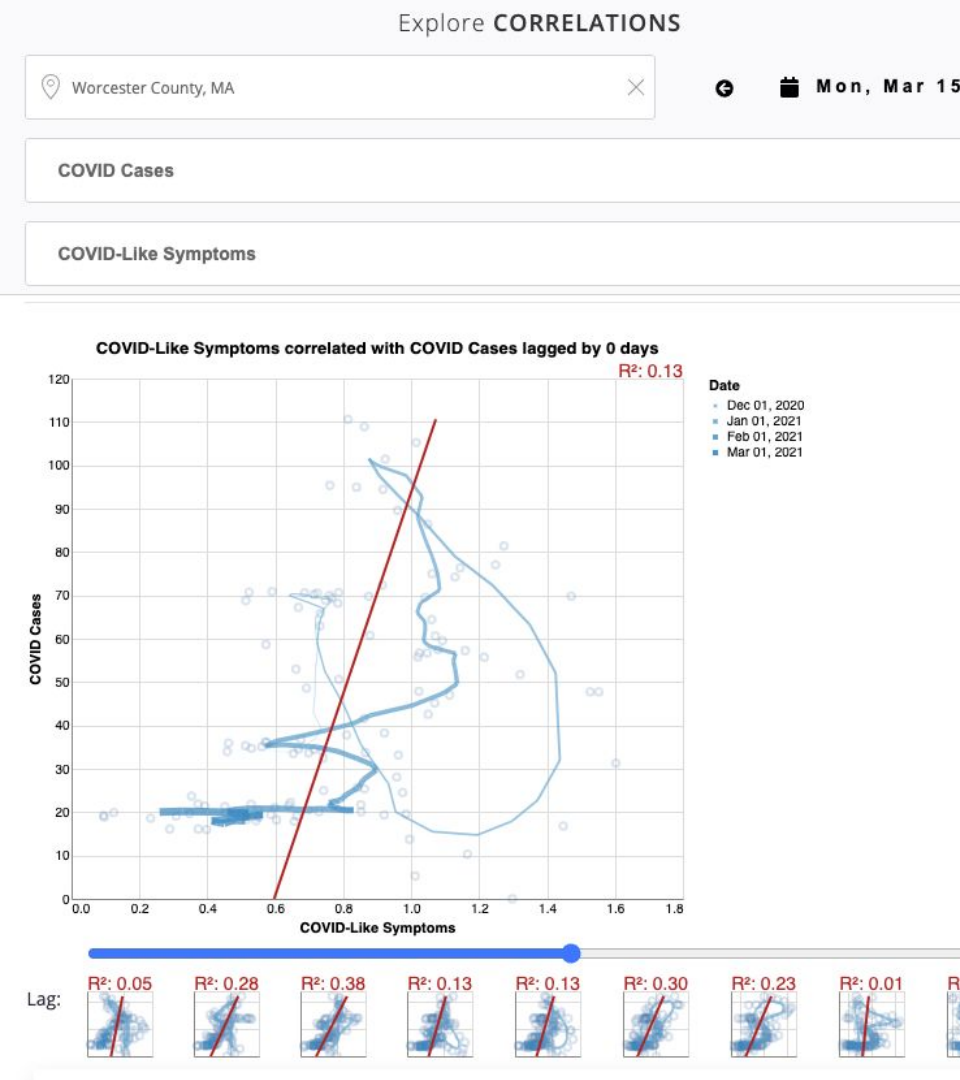
Correlation Explorer

- Interactive correlation plot
- By sliding the days-lagged control, you can see where R2 is maximized
- Days vs R2 shows peak



Correlation Explorer

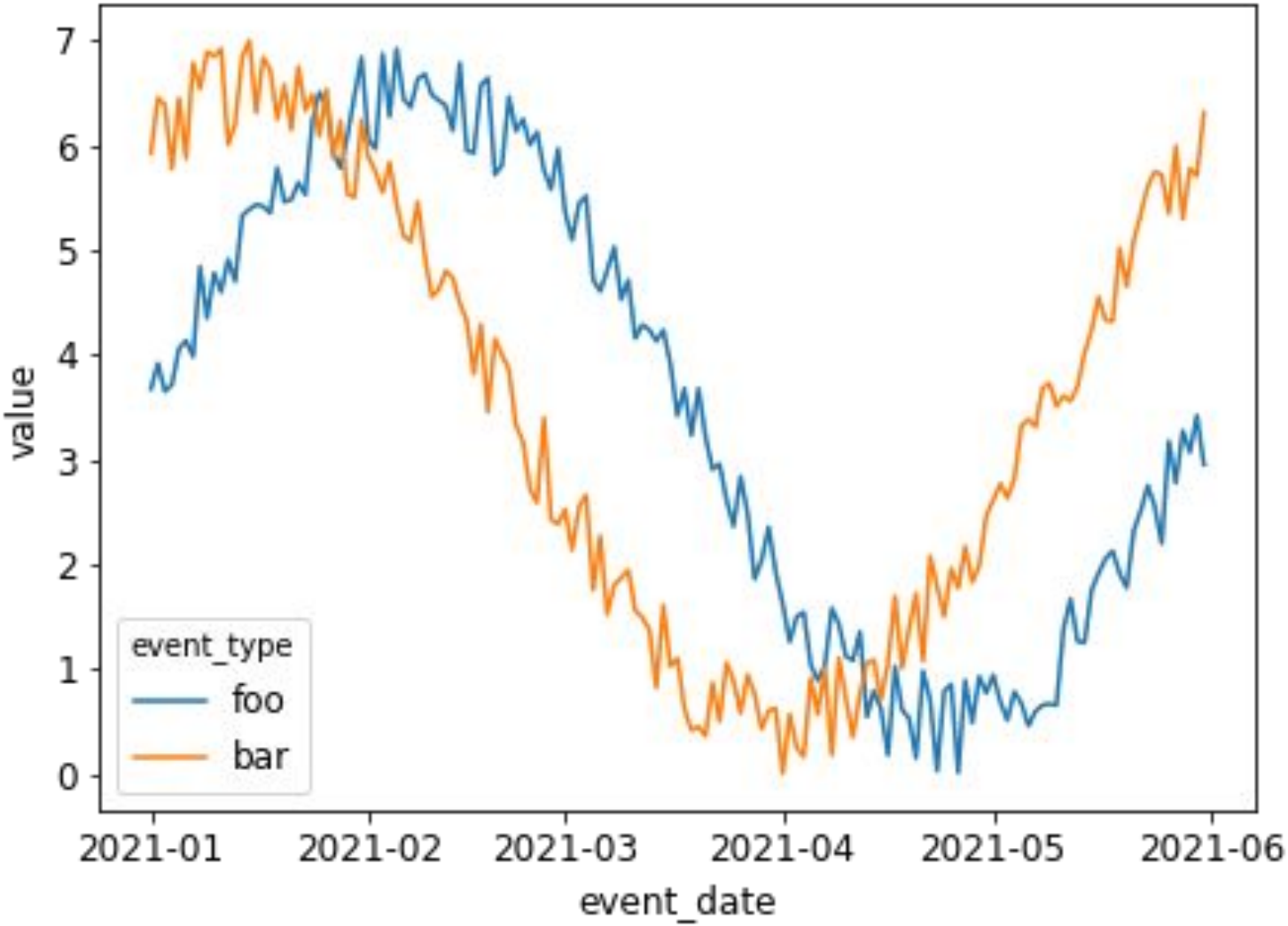
- Really neat!
- Hard to explain
- Hard to interpret
- Complexity can prevent utility



Patterns

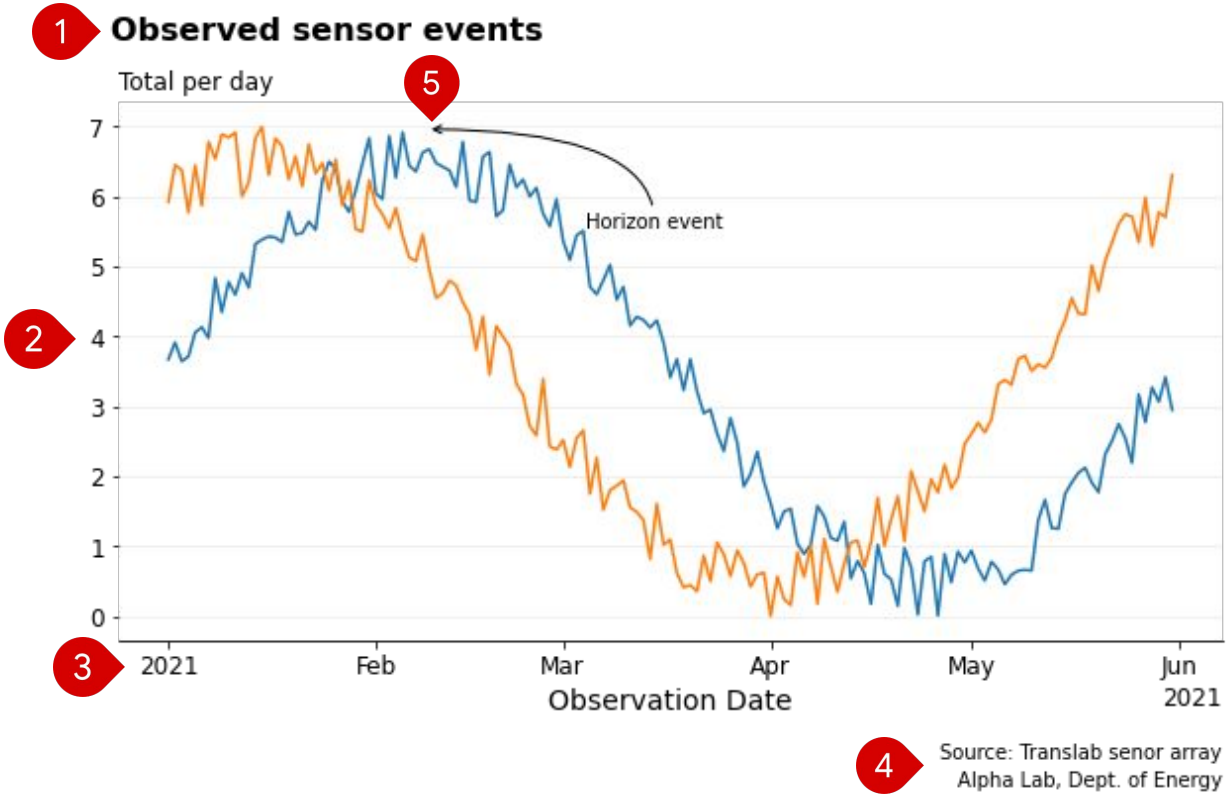
Design Patterns

- **Consider the data journalist style**
- Highlight features and data quality
- Use progressive complexity
- Provide summaries and context
- Design for comparison
- Embrace tabular data
- Unleash your data

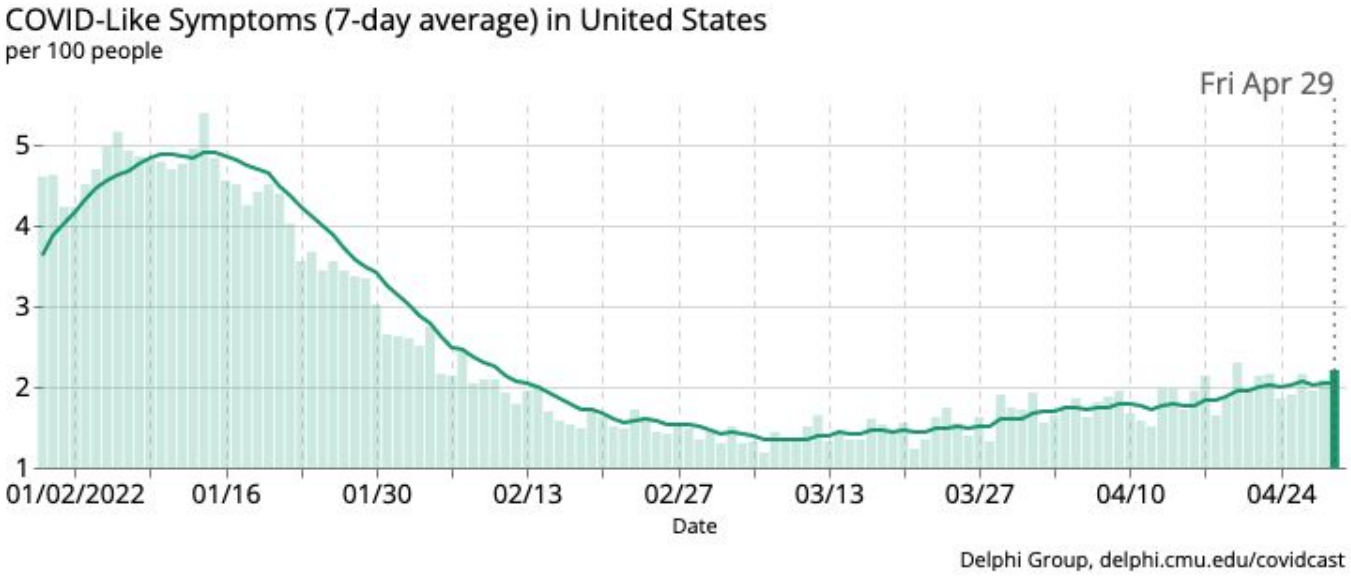

7
6
5
4
3
2
1
0
value
event_type
foo
bar
2021-01
2021-02
2021-03
2021-04
2021-05
2021-06
event_date

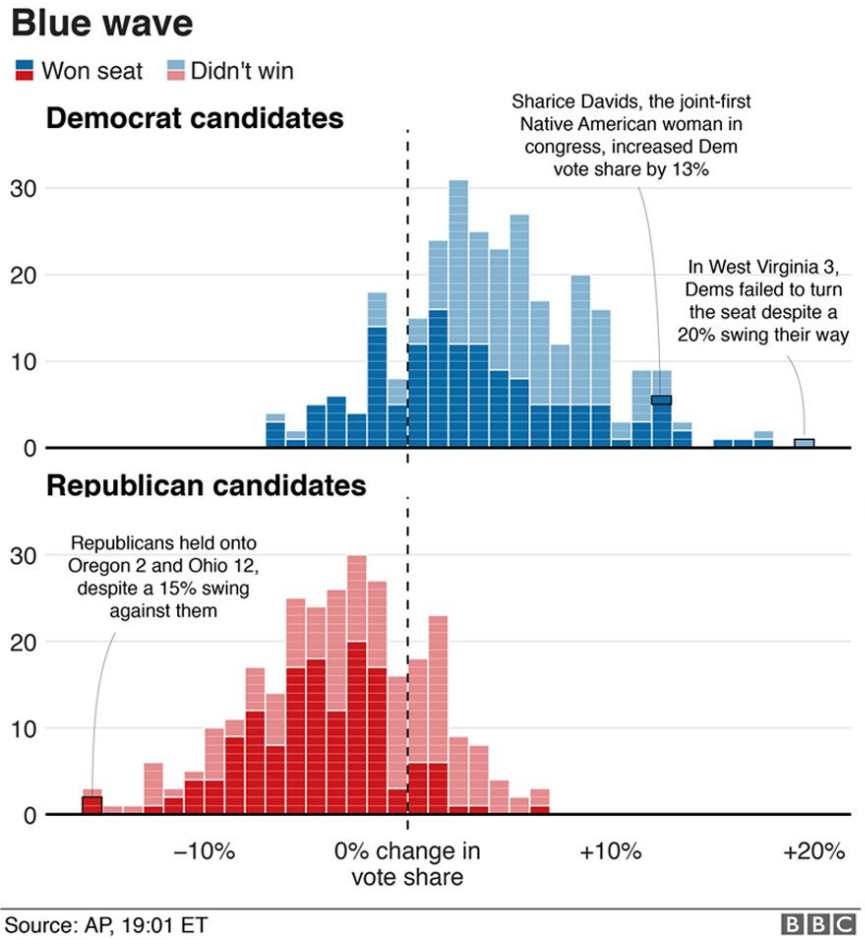
Data Journalist Style

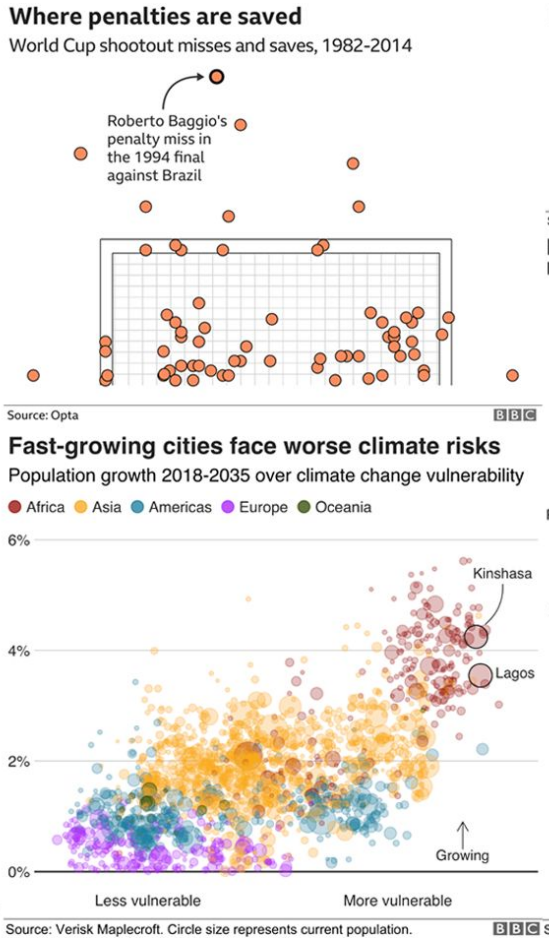
1. Move y-axis label to chart title
2. Simplify chart axes and grid lines
3. Format time-series
4. Attribute data source
5. Highlight major features

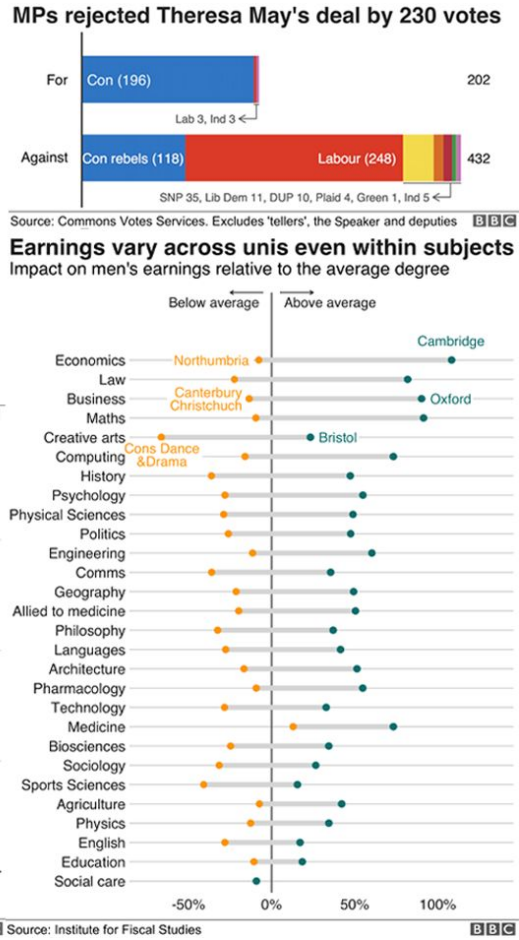


Data Journalist Style

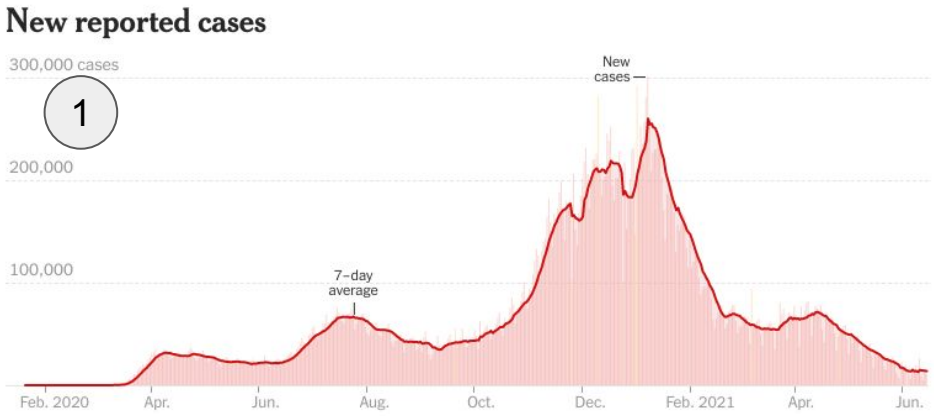


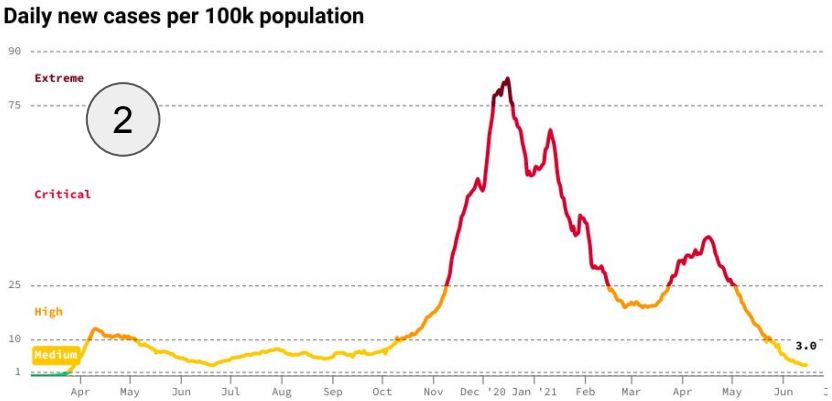




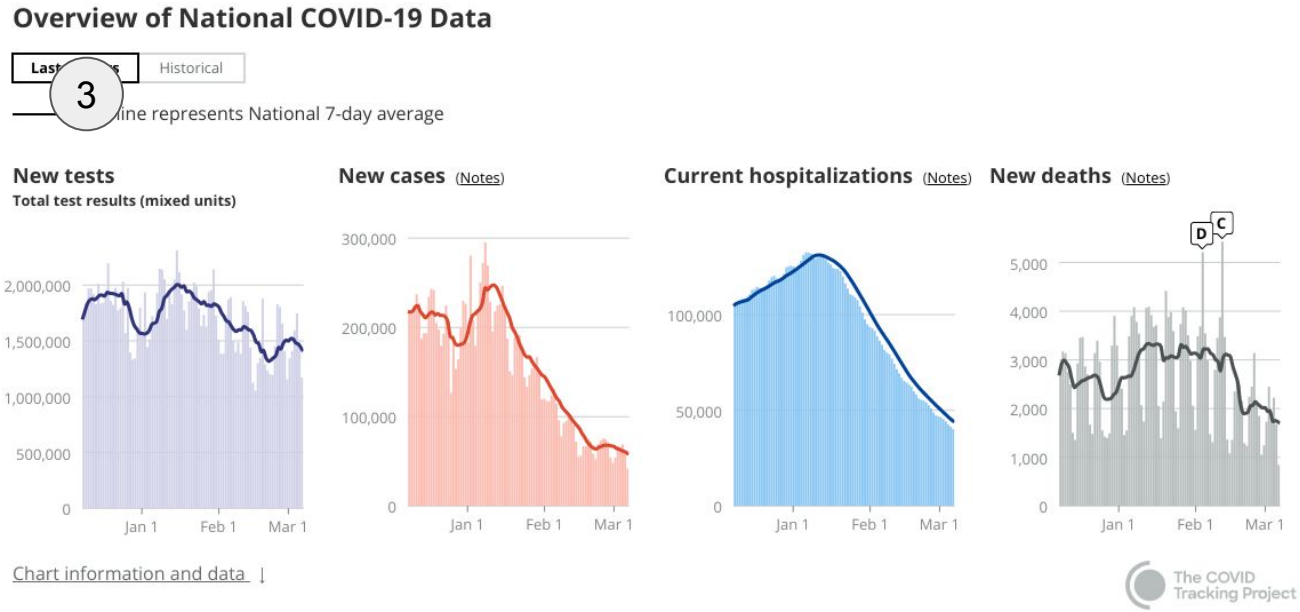


https://bbc.github.io/rcookbook/





1. New York Times
2. COVID Act Now
3. The COVID Tracking Project

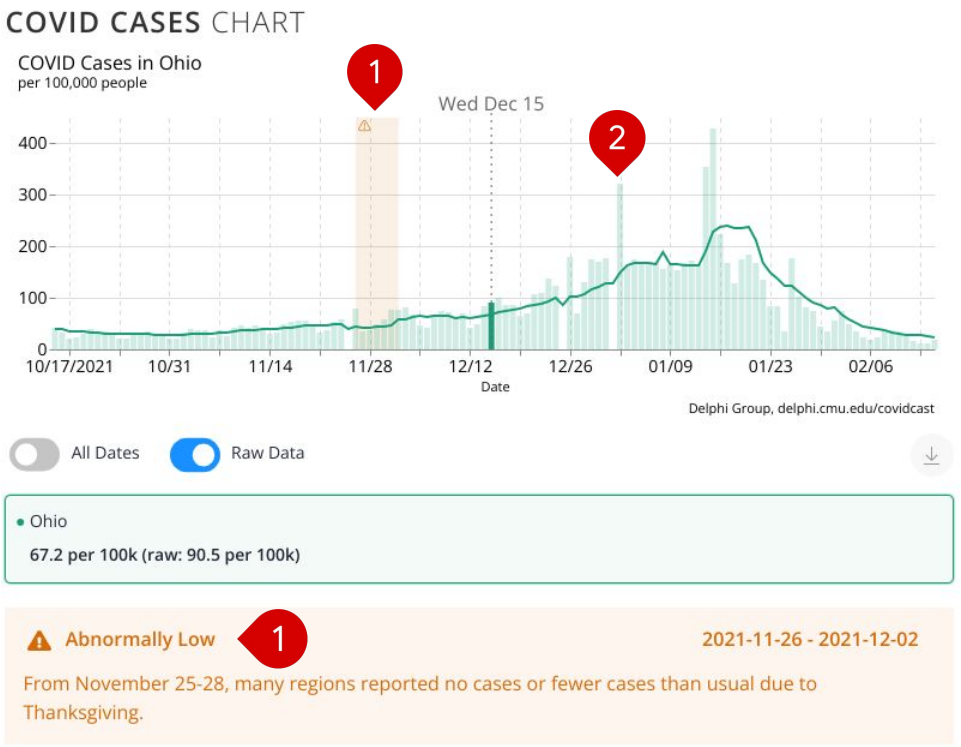


Design Patterns

- **Consider the data journalist style**
- **Highlight features and data quality**
- Use progressive complexity
- Provide summaries and context
- Design for comparison
- Embrace tables
- Unleash your data

Highlight features and data quality

1. Highlight notable features
2. Consider Bar+Line chart for time series with only one series

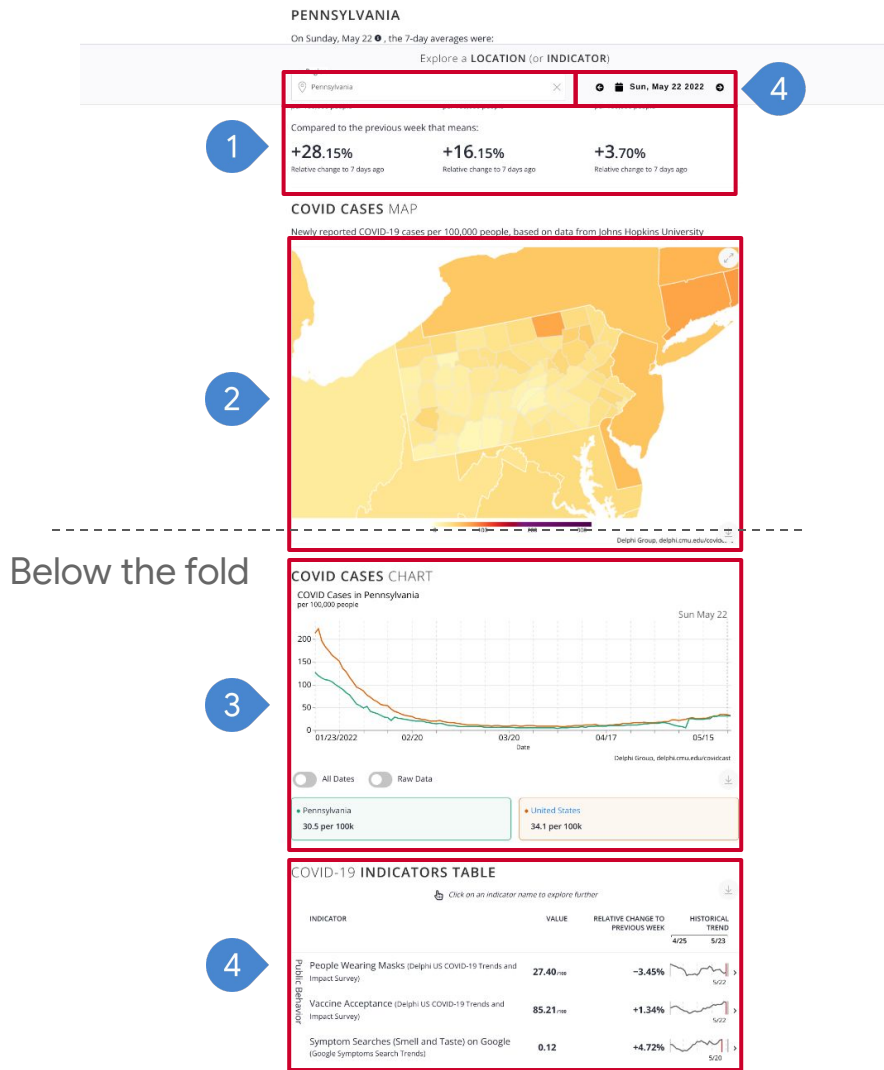


Design Patterns

- **Consider the data journalist style**
- **Highlight features**
- **Use progressive complexity**
- Provide summaries and context
- Design for comparison
- Embrace tables
- Unleash your data

Use Progressive Complexity

- Start simple
- Introduce concepts one at a time
- Manage complexity across two dimensions
- Vertical space
- Drill downs or links
  - Vertical space
  - Drill downs or links

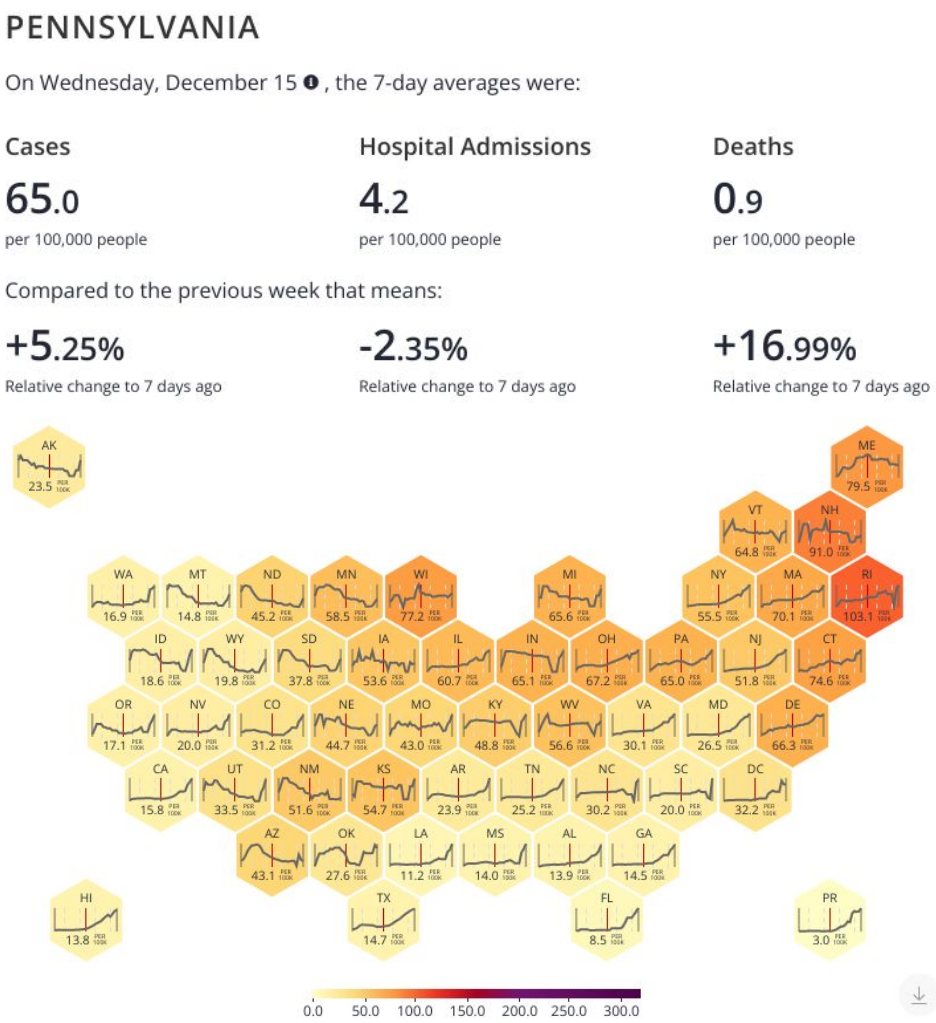


Design Patterns

- **Consider the data journalist style**
- **Highlight features**
- **Use progressive complexity**
- **Provide summaries and context**
- Design for comparison
- Embrace tables
- Unleash your data

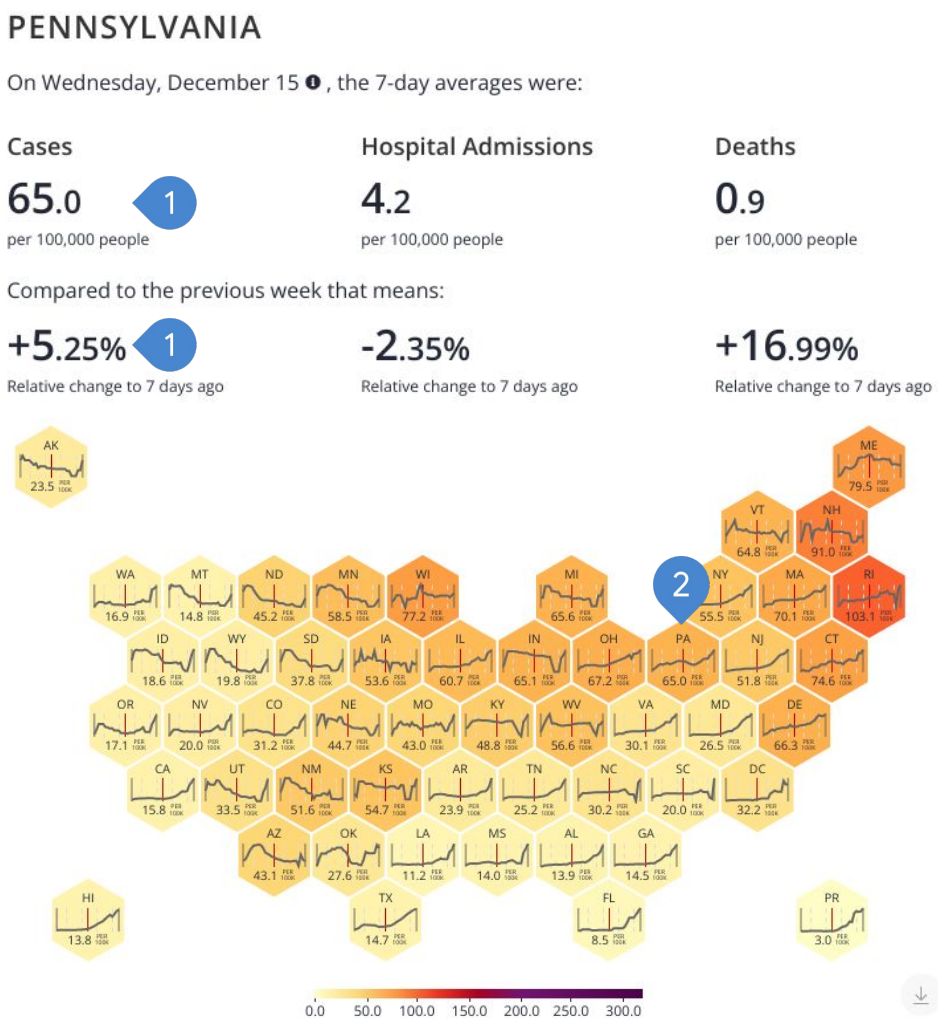
Provide summaries and context

- Data points without context are hard to interpret
- “Is 65 good or bad?”



Provide summaries and context

1. Consider “summary cards” for top-level metrics
   a. Name, Value, Unit + Change/Time
2. Consider equal-area plots like hex-grids for high-level geospatial data



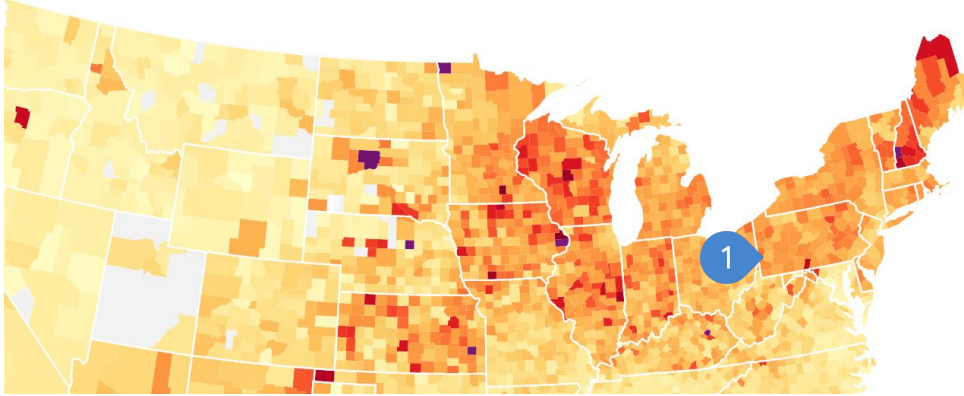
Design Patterns

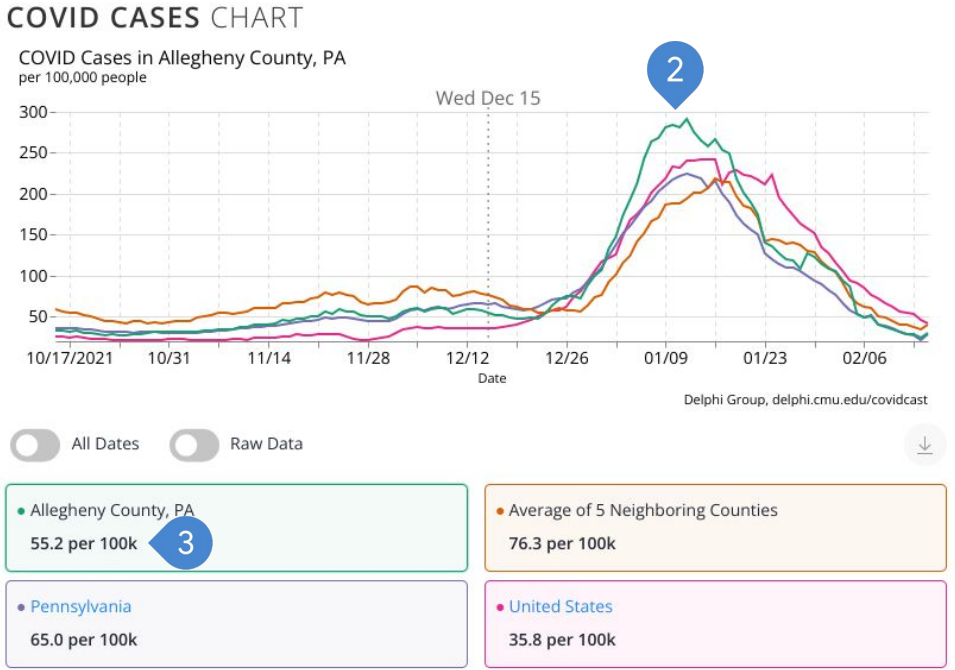
- **Consider the data journalist style**
- **Highlight features**
- **Use progressive complexity**
- **Provide summaries and context**
- **Design for comparison**
- Embrace tables
- Unleash your data

Design for Comparisons

Introduce several methods for comparisons, to provide context to metrics. This is especially use for geo-temporal data.

1. Choropleths expose high-level geospatial relationships
2. Pair with time series to provide temporal insights
3. Consistent use of units allows comparisons along visual and textual dimensions





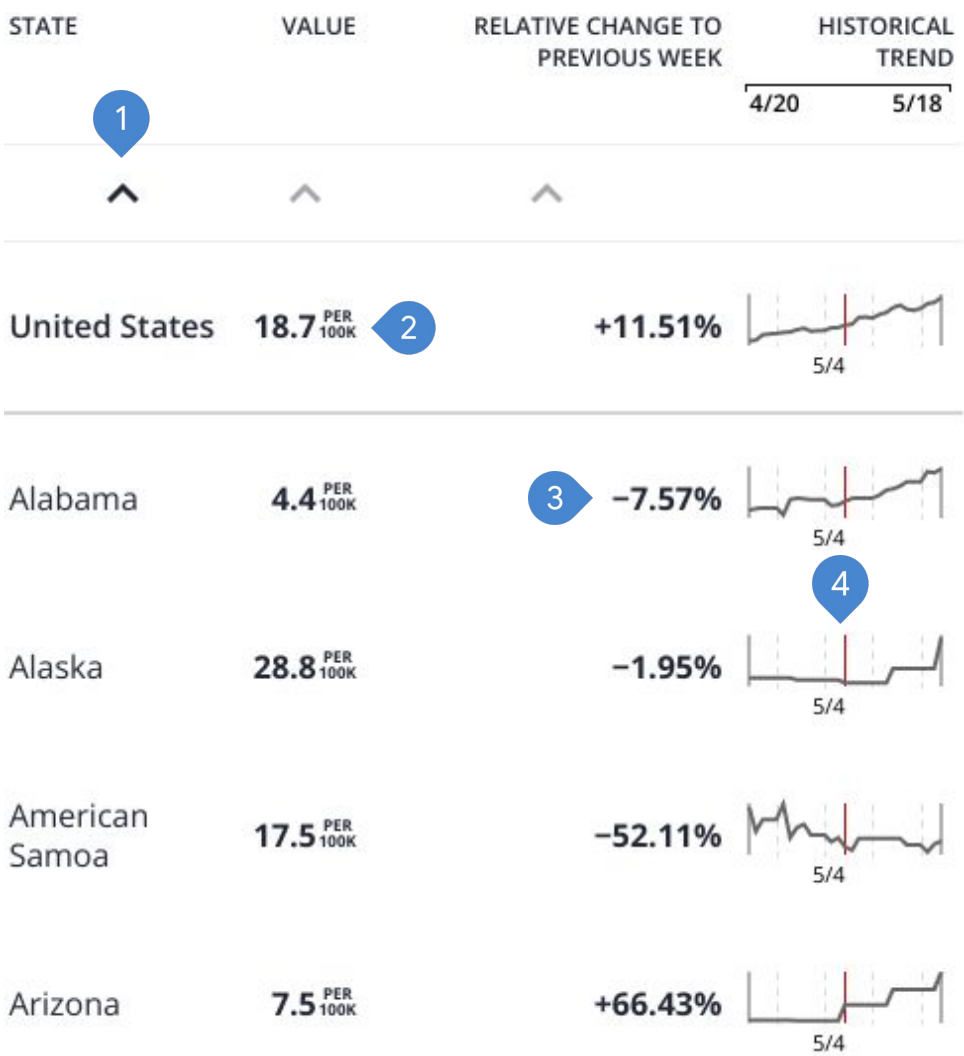
Design Patterns

- **Consider the data journalist style**
- **Highlight features**
- **Use progressive complexity**
- **Provide summaries and context**
- **Design for comparison**
- **Embrace tables**
- Unleash your data

Embrace Tables

Tabular visualizations are underutilized

1. Sorting allows quickly finding min and max values
2. Define your units precisely
3. Quantify reactive change
4. “Rich” cells like sparklines can be particularly effective
5. **Pair charts with the same data in tables to maximize utility**

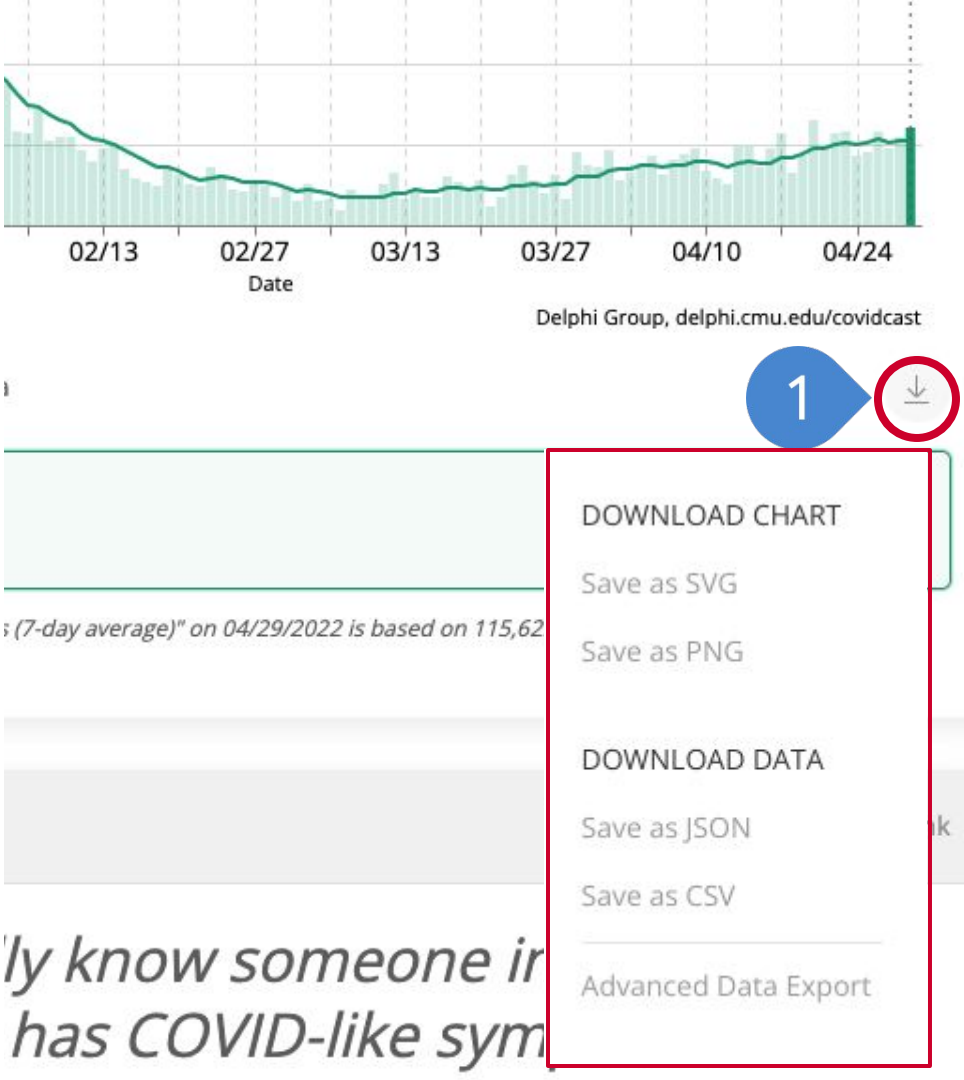


Design Patterns

- Consider the data journalist style
- Highlight features
- Use progressive complexity
- Provide summaries and context
- Design for comparison
- Embrace tables
- Unleash your data

Unleash your data

- Data always has greater utility than visualizations
- Let your users download it!
- Make your license easy to discover, use and cite
- **Tip: when creating visualizations aggregate first, then plot**



Summary

Design Patterns

- Use the data journalist style
- Highlight features
- Use progressive complexity
- Provide summaries and context
- Design for comparisons
- Embrace tables
- Unleash your data

Conclusion

Although this applies to a non-technical audience, we hope these principles increase the utility, clarity and usability of your work.

Thank you to all who contributed

Google fellows, Delphi Visualization team and the entire Delphi team

Chris Scott
Solutions Engineer
Google.org

Phil McGuinness
UX Research
Google.org

Raphael Hyde
Designer
Google.org

Spencer Whitman Product Manager
Google.org

Kate Harwood
Software Engineer
Google.org

Alex Reinhart
CMU/Stat

Ryan Tibshirani
Delphi Lead
CMU/MLD/Stat

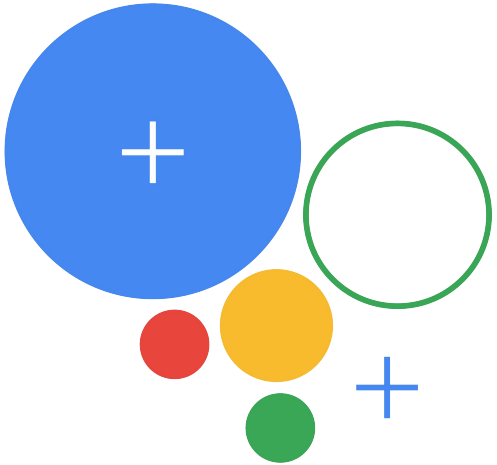
Roni Rosenfeld
Delphi Lead
CMU/MLD

Jenny Lee
Design Intern
CMU/MLD

Samuel Gratzl
Consultant

Daniel LaLiberte
Software Engineer
Google.org

Kathryn Mazaitis
CMU/MLD



Thank you