Demo

1. Describe your data set.
    a. Why was the data collected?
    b. How was it collected?
    c. What are the characteristics of the data?
2. Provide a rationale for a visualization tool.
    a. Why is a visualization tool necessary?
    b. What can we expect to learn?
3. Discuss your final design and justify your design decisions.
    a. Why did you choose your visual encodings?
    b. What kind of interactions did you implement and why?
    c. Does your visualization scale with the size of the data?
4. Provide a live demo that illustrates key features of your Shiny app

15 minutes presentation. 5 minutes Q&A

When did US states close their schools when the COVID-19 pandemic hit?

Jonathan Levin, Chenyue Lu, Kathleen Sucipto, Dany Thorpe

Dataset 1: COVID Tracking Project

- Case counts do not tell the full story
- State-level testing data (web scraping + manual annotation) from various testing facilities
- Update daily
- Has information on hospitalization, ICU admission, and ventilator usage but unreliable

NEW We just launched the COVID Racial Data Tracker

https://covidtracking.com/why-it-matters

Massachusetts

Current data quality grade: A

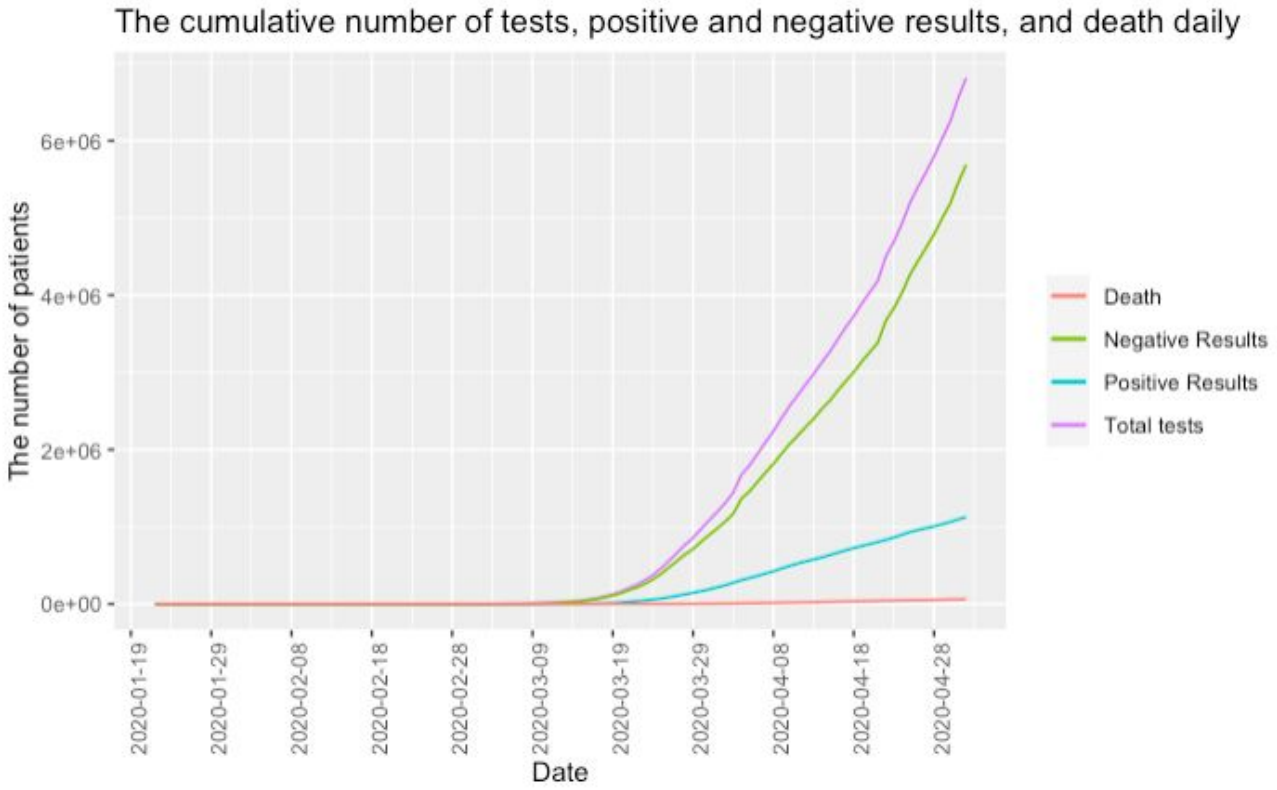
Tests			Hospitalized		In ICU		On Ventilator		Outcomes		Total Test Results
Positive	Negative	Pending	Currently	Cumulative	Currently	Cumulative	Currently	Cumulative	Recovered	Deaths	Positive + Negative
66,263	232,731	N/A	3,601	6,378	921	N/A	N/A	N/A	N/A	3,846	298,994

Last updated: Sat May 2 2020 12:00 pm EST

Massachusetts on Twitter
Best current data source for Massachusetts
Historical data for Massachusetts

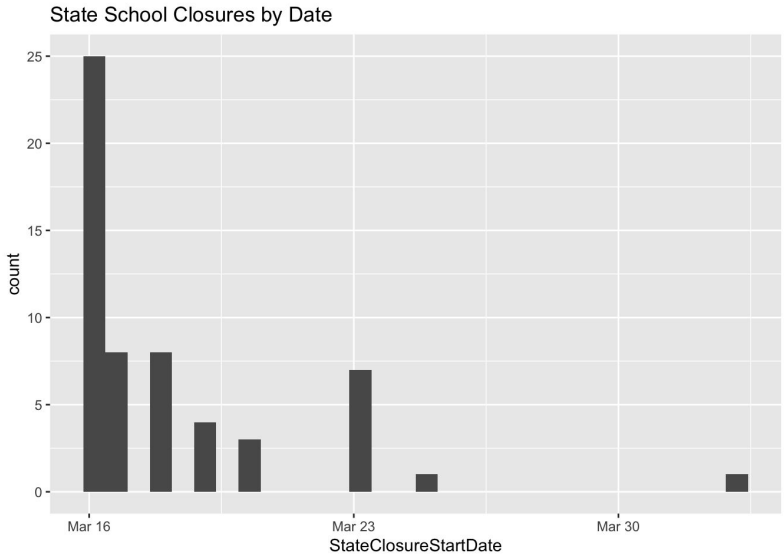
Positives include results from the CDC. Total tested includes repeat testing on individuals. State lab numbers are as of 8 am, other lab numbers are as of 11:30 am. Commercial lab numbers are subject to change. As of 4/20, Cumulative Hospitalized + ICU includes both confirmed and suspected cases.

Dataset 1: COVID Tracking Project



Dataset 2: School Closure Dates across US states

State	State Abbreviation	State Status	State Closure Status	State Closure Start Date	State Number of Public Schools	State Public School Enrollment
Alabama	AL	State ordered closure	Ordered closed for the academic year	3/19/20	1,513	744,930
Alaska	AK	State ordered closure	Ordered closed for the academic year	3/16/20	507	132,737
American Samoa	AS	State ordered closure	Ordered closed until further notice	3/23/20	28	n/a
Arizona	AZ	State ordered closure	Ordered closed for the academic year	3/16/20	2,308	1,123,137
Arkansas	AR	State ordered closure	Ordered closed for the academic year	3/17/20	1,089	493,447
Bureau of Indian Education	BIE	Closures determined at school/district levels	Varies by school/district	n/a	174	45,399
California	CA	State recommended closure	Recommended closed for the academic year	3/19/20	10,286	6,309,138
Colorado	CO	State ordered closure	Ordered closed for the academic year	3/23/20	1,888	905,019
Connecticut	CT	State ordered closure	Ordered closed through 5/20/2020	3/17/20	1,250	535,118
Delaware	DE	State ordered closure	Ordered closed for the academic year	3/16/20	228	136,264
Department of Defense Education Activity	DoD	Closures determined at school/district levels	Varies by school/district	3/25/20	163	70,852
District of Columbia	DC	State ordered closure	Ordered closed for the academic year	3/16/20	223	85,850
Florida	FL	State recommended closure	Recommended closed for the academic year	3/16/20	4,178	2,816,791
Georgia	GA	State ordered closure	Ordered closed for the academic year	3/18/20	2,300	1,764,346
Guam	GU	State ordered closure	Ordered closed for the academic year	3/16/20	41	30,758
Hawaii	HI	State ordered closure	Ordered closed for the academic year	3/23/20	290	181,550



https://www.edweek.org/ew/section/multimedia/map-coronavirus-and-school-closures.html

Additional Data Used

Henry J. Kaiser Family Foundation: Political Party affiliations

https://www.kff.org/other/state-indicator/state-political-parties/

CDC Census Data, 2019 estimates

https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/state/detail/

Why is a visualization tool necessary?

- Inform policy makers and public in enforcing regulations
- It is unclear of the transmissibility of coronavirus in children (CDC.gov)
  - Conversely, it is unclear what the risk is for parents of schoolchildren and their teachers
- School closings may mitigate spread of virus, but have their own risk as well
  - School lunch programs (NPR; 3/20/20)
  - Safety concerns at home (ABC News, 4/24/20)
- As first mitigation phase ends, whether or not school closings are an essential part of future mitigation efforts during repeat outbreaks will be important (Upshot NYTimes 3/17/20)
- Determining effect of school closure on “the curve”

Jim Lo Scalzo/EPA, via Shutterstock

What can we expect to learn?

- Explore relationship between state-level school closings and COVID-19 caseload across states
  - Specifically, how rates of new cases, % of positives, new deaths change *before* and *after* school closure by each state
  - Display each state's data not by date but centered around date of official school closure
- Compare characteristics across different states
  - Political party of governor, geographical regions, timing of school closure (tertile)
  - How these categories group together
  - Explore state's preparedness in COVID surveillance
- *Ultimately:*
  - What were characteristics of states that closed early vs late?
  - What was the effect of closing school on "the curve"?

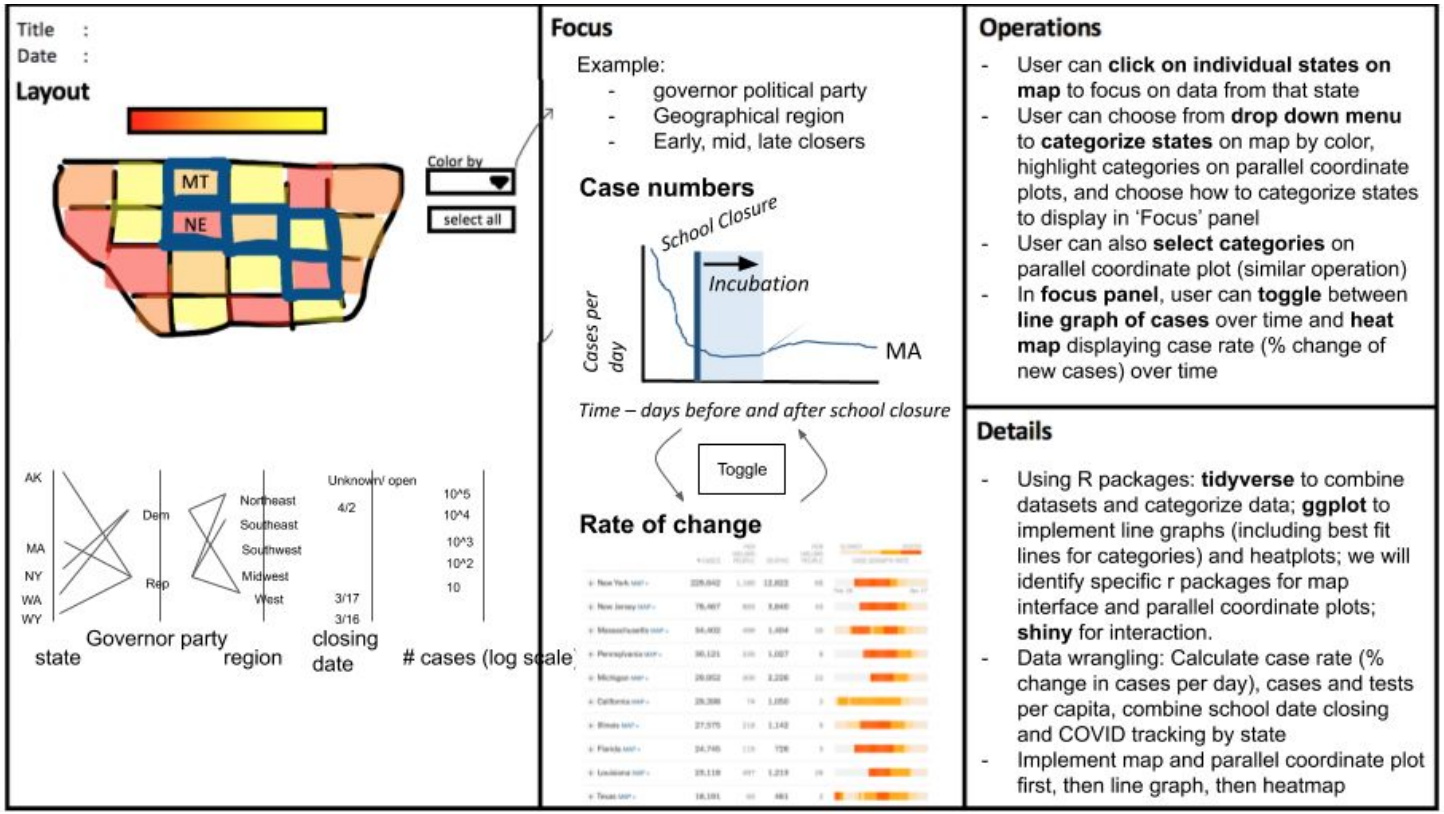
Potential sources of bias

- Like all datasets reporting number of positives, this is somewhat reflective of *total testing* as well as true case load
- In the early stages of the pandemic, many interventions were implemented at once (school closings, travel plans, social distancing, stay-at-home orders)
    - Difficult to tease out which one has the greatest effect
- Difficult to separate cause from effect
    - Confounders
- Incubation period - delay in any effect from a policy

Additional Bias in Datasets

- COVID Tracking Data
  - Incomplete hospitalization, ICU data due to incomplete Reporting
- School reporting
  - District level closing may have occurred before state level closure
  - Does not reflect parental behaviors or private schools

Visualization Design



Demo

1. Overview (including map)
2. Normalize data
3. Parallel coordinate plot
    a. Show linking: New England States, Dem vs republican, Northeast vs South, Region
        i. Change Category with each example; return to “None” at end
4. Heat Map - best for showing patterns of % change
    a. Positives, tests, deaths
5. Line graph - trend lines
    a. Overall trend in population, incubation slider
    b. Positive increase vs % change in positive, death increase vs % change in deaths
        i. Slide incubation bar - esp re death
    c. Political party and region (% change in new positives)
    d. Time of closure - highlight early vs others, % change in new positives and % change in deaths

Conclusions

- States with **Dem governors did more testing**, and tended to have **greater % of tests positive** than states with Rep governors; however ***time of school closure was similar***
- **Northeast states** also did more testing and had **greater % of tests positive**; ***no obvious difference in timing of school closure by region***
- No difference in case load (% of tests positive) or # of tests early vs late closure

Conclusions

- Timing of ***school closure could reflect changes from policy already implemented***
  - Overall, there was a **decrease in rate of new positive cases (% change) until ~5 days** after school closure and **new deaths (% change) ~8 days after school closure**, though **% of tests that were positive had already begun to decline before this**
  - No difference by political party of governor or region
  - In fact, states who **closed early** had **less of % change in new positives.**
  - **New deaths (by % change) decreased more in states who closed early**, but likely ***not enough incubation time*** for this to be a real effect
- Suggests that ***school closure timing was part of broader policy package and cultural shift, and more likely is a reflection of existing case load***

Conclusions

- Design experience
  - Allowing user to control assumptions about data, which data is most important
  - Multiple ways of visualizing data (heat map, line graph)
  - Interpretability vs readability (e.g. log scale)
  - Plotly has numerous usability and implementation drawbacks compared to ggplot
- Next steps
  - Extending data as schools re-open to look at effects of school openings
  - Can be used for other policies (business closure, stay at home order, etc)