



# Covid Analysis

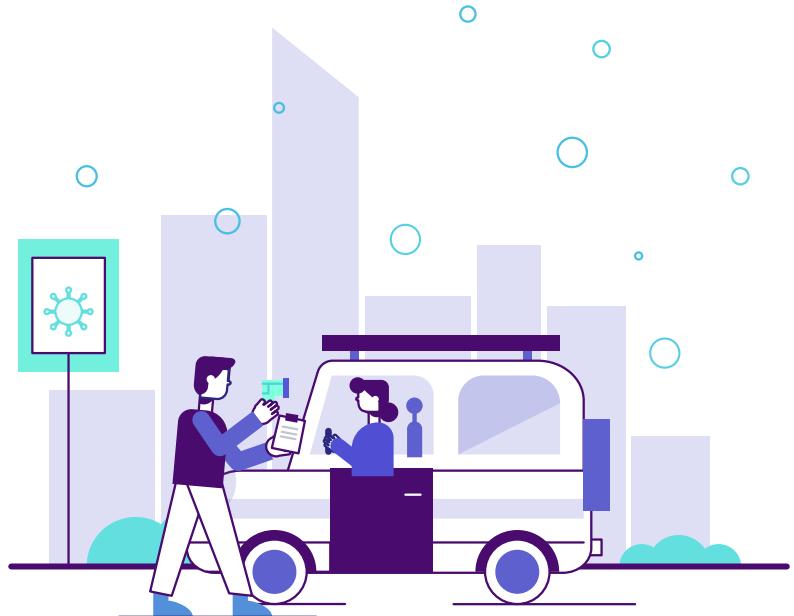
**MSCA 31012 Data Engineering Platforms for Analytics  
Final Project Presentation (Dec 12, 2020)**

Team: Jaird Meyer, Elly Yang, Mengyang Yu, XiaoJing Zhang



# Agenda

- Executive Summary
- Policy Insights
- Design Considerations
- Data tools
- Conclusions



# Executive Summary

- For this project, we will look at **Covid-19 data** in the United States, both live and historical, to assess the **effectiveness** of governmental policies.
- Our data visualization and analysis will allow us to provide **reflections** on the use of data analytics in the process of tackling this challenge, as well as **recommendations** for future steps in preventing deaths.

# Policy Insights

- **Apply quantitative insights on policy**
  - Find connections between real-time events to various metrics
  - It is important to recognize various lags when assuming time-dependent associations
    - Policy lag: announcement to implementation and effect
    - Infection lag: time of infection occurs before positive input
  - Quantify the effect of various policy initiatives
- **Takeaways**
  - These insights will be important to implement during the continuing battle with coronavirus, but will also provide guidance for future pandemics

# Design Considerations

- Data preparation
  - Aggregated data size: <1MB
  - Data type: structured data
  - Analysis granularity: county & state level
- Application type
  - OLAP application for ease of implementation, analysis and reporting
- Platform considerations
  - Google Cloud Platform (GCP) is compatible with MySQL and Tableau
  - Convenient for team members to share and transfer project data
  - Reliable cloud service provider choice with limited budget

# Data Collection

## 'aggregated by county.csv' Data Profile

- 3132 observations, 18 variables
- Joined by County Key
  - 'county.csv'
  - 'county\_mask.csv'
  - 'county\_pop.csv'

## 'aggregated by state.csv' Data Profile

- 50 observations, 38 variables
- Joined by State Key
  - 'state.csv'
  - 'state\_mask.csv'
  - 'state\_pop.csv'
  - 'state\_insurance.csv'
  - 'hosp\_util.csv'



### Covid-19 Mask Use by State & County:

Estimates of mask usage by state & county in the United States. This data comes from a large number of interviews conducted online by the global data and survey firm Dynata at the request of The New York Times.

### Covid-19 Live US-States/Covid-19 Live US-Counties:

Cumulative coronavirus cases and deaths for each geography up to Nov 13th. The counts include both laboratory confirmed and probable cases using criteria developed by states and the federal government.

### Annual Estimates of the Resident Population:

Annual Estimates of the Resident Population for Counties in the United States: April 1, 2010 to July 1, 2019.

### Health Insurance Coverage Status and Type of Coverage:

Health Insurance Coverage Status and Type of Coverage by State All Persons: 2008 to 2019. Type of coverage includes private (employer-based and direct-purchase) and public (Medicaid, Medicare,etc).

### Estimated U.S. Hospital Utilization:

Estimated hospital utilization data are available for the U.S. states, and territories. This data includes total inpatient bed, total ICU beds, percent of inpatient bed utilization, and percent of inpatients with covid, etc.



# Data Processing

- Drop duplicated columns
- Drop columns w/ many null values
- Create calculated fields for modeling

	county	state	fips	cases	deaths	confirmed_cases	confirmed_deaths	State_fips	death_rate_confirmed	death_rate
0	Autauga	Alabama	1001.0	2417	36.0	2165.0	33.0	1	0.015242	0.014894
1	Baldwin	Alabama	1003.0	7523	84.0	6309.0	80.0	1	0.012680	0.011166
2	Barbour	Alabama	1005.0	1117	9.0	765.0	9.0	1	0.011765	0.008057
3	Bibb	Alabama	1007.0	973	17.0	873.0	13.0	1	0.014891	0.017472
4	Blount	Alabama	1009.0	2488	34.0	1904.0	33.0	1	0.017332	0.013666



# Data Storage

- Loaded data into MySQL database
- Organized data into Entity Relation diagram

Google Cloud Platform DEP project1 Search products and resources

Storage Bucket details dep-final-proj

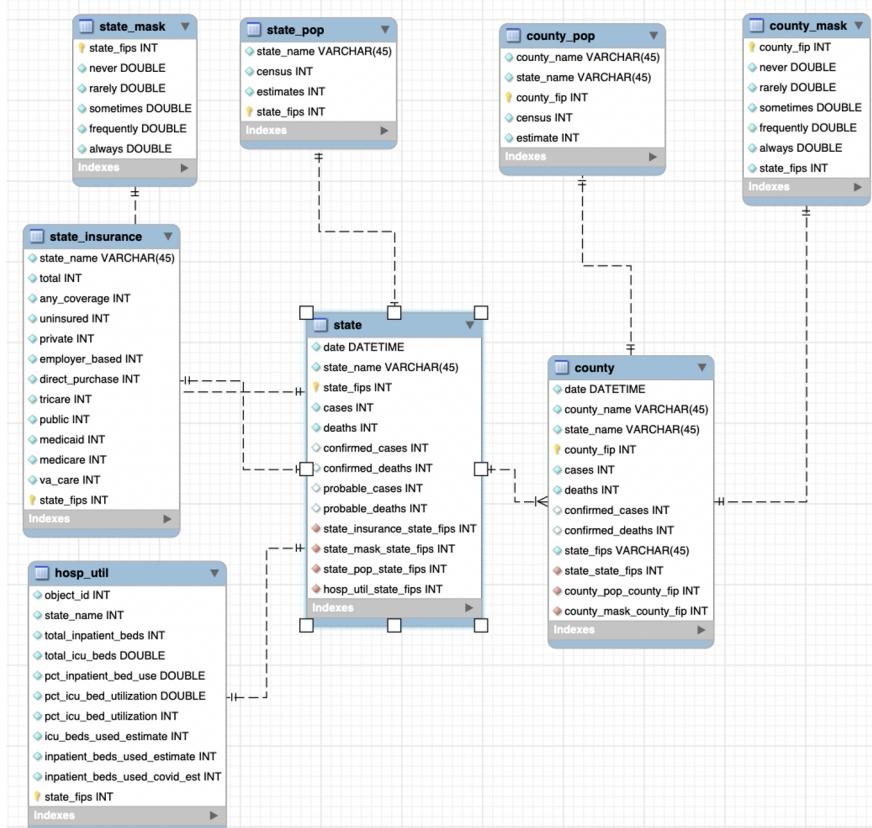
OBJECTS CONFIGURATION PERMISSIONS RETENTION LIFECYCLE

Buckets > dep-final-proj > DEPA Final Data

UPLOAD FILES UPLOAD FOLDER CREATE FOLDER MANAGE HOLDS DOWNLOAD DELETE

Filter Filter by object or folder name prefix

Name	Size	Type	Created time	Storage class	Last modified
DS_Store	6 KB	application/octet-stream	Dec 3, 2020, 1:...	Standard	Dec 3, 202...
County_Level_merged.csv	330.3 KB	text/csv	Dec 3, 2020, 1:...	Standard	Dec 3, 202...
County_Level_related_dates	—	Folder	—	—	—
R Exports/	—	Folder	—	—	—
State_Level_merged.csv	14.5 KB	text/csv	Dec 3, 2020, 1:...	Standard	Dec 3, 202...
State_Level_related_dates	—	Folder	—	—	—





- Multi-Criteria Decision Analysis
  - Combination of PCA and MFACT
  - Reduce collinearity and confounding variation among variables
- Rank States by the historic performance of attributes
- Measure influence of each variable through weighted linear combinations

# Data Analysis

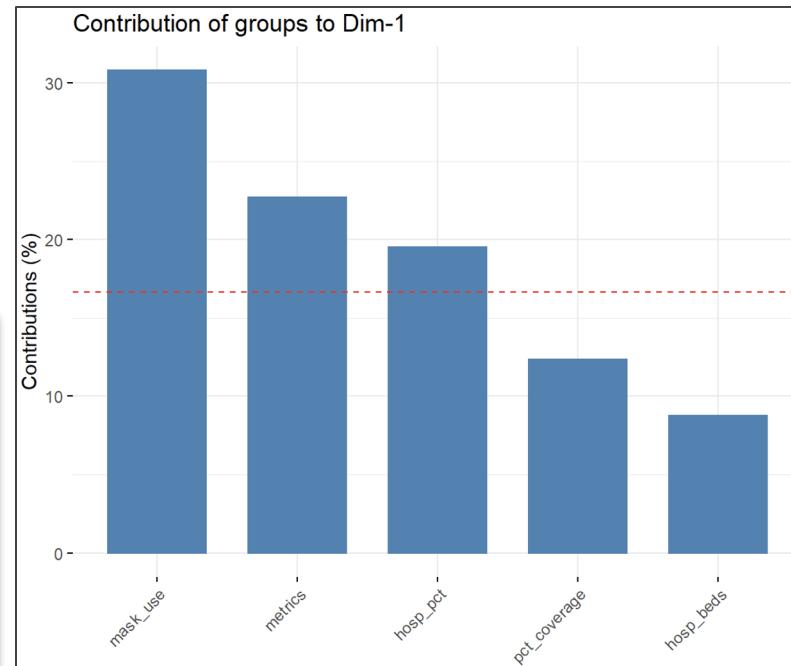
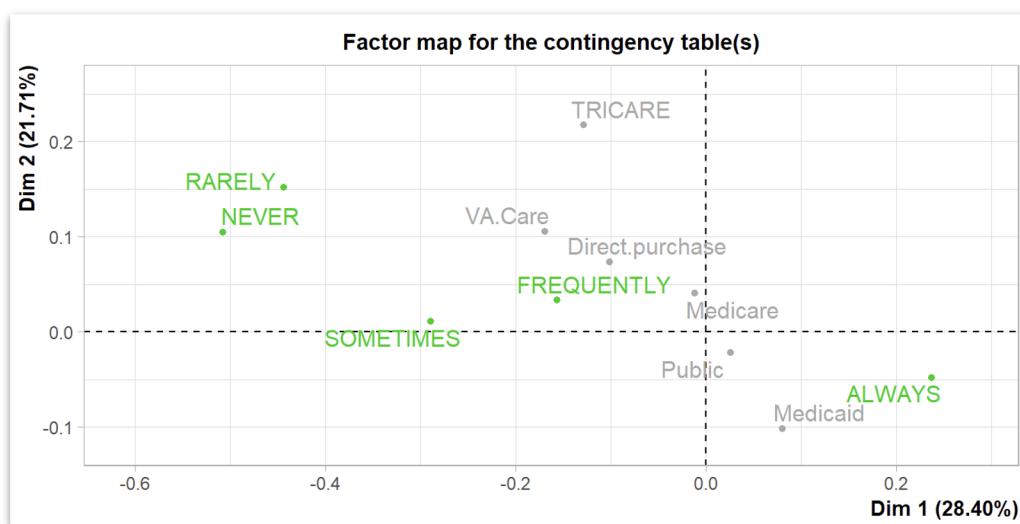
State	Score	Rank
Vermont	5.57589496	1
Hawaii	4.91919026	2
New Hampshire	3.58199386	3
Massachusetts	3.36986452	4
Washington	2.82677780	5
Oregon	2.34160193	6
Virginia	2.17984419	7
Maine	2.05900116	8
Connecticut	1.80476768	9
Utah	1.67333288	10
New Jersey	1.43901470	11
Wyoming	1.41117045	12
California	1.40881370	13
Maryland	1.31113702	14
Delaware	0.96233176	15
Minnesota	0.71441797	16
Colorado	0.71405012	17
Pennsylvania	0.53833939	18
Ohio	0.12651871	19
Wisconsin	0.07010251	20

...

# Model Recommendation

## Wear a Mask!

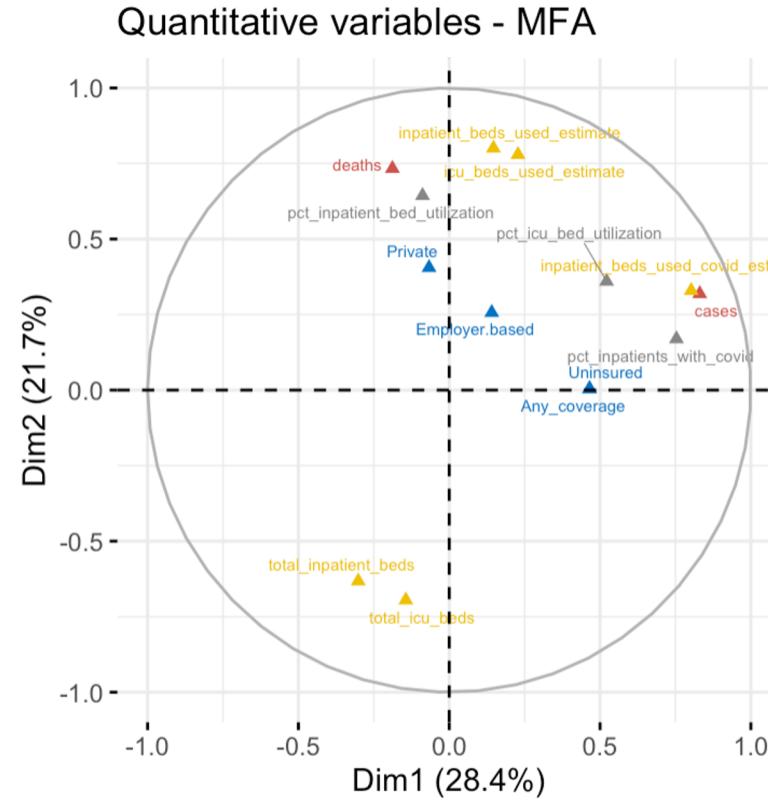
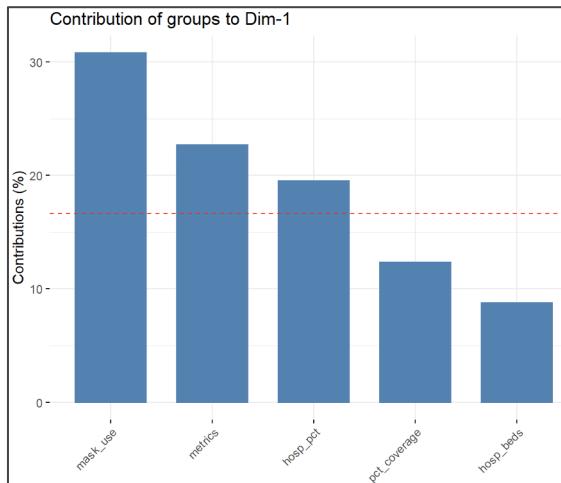
- Model indicates masks are only effective when **ALWAYS** worn
- Mask use has the greatest explanatory power of any group of variables in differentiating states



# Model Recommendation

## Social Distancing

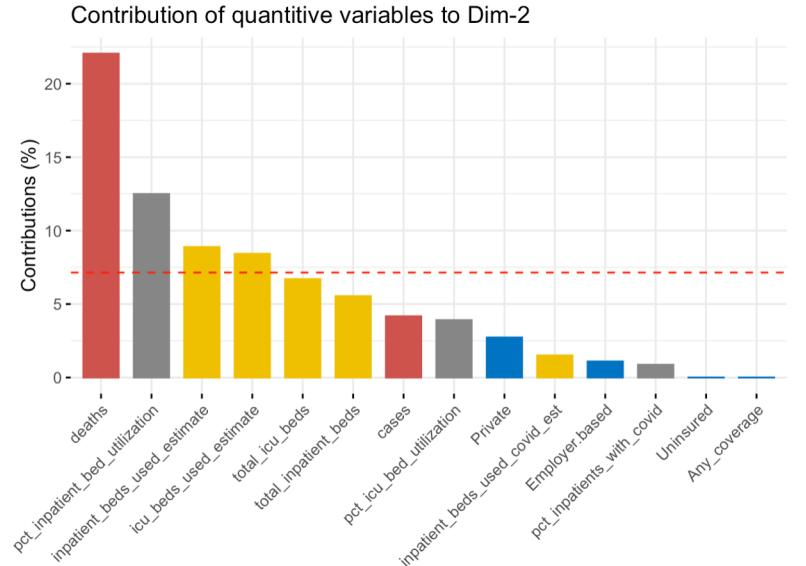
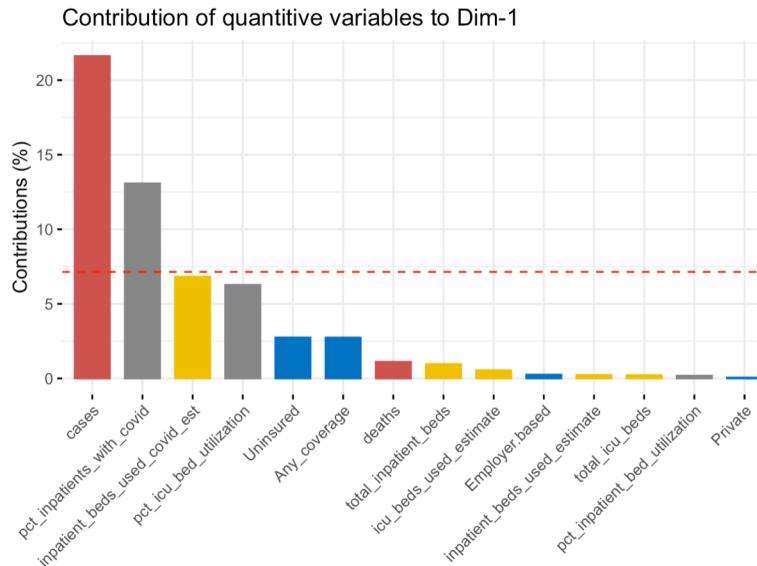
- Occupancy rates in hospitals is a more significant indicator than hospital capacity
- Flattening the curve is an important objective to maintain low occupancy rates



# Model Recommendation

## Quantitative Dimensional Comparison

- Temporal qualities of lagged variables
- Dim 1 prioritizes cases and early hospital intake
- Dim 2 reflects deaths and total hospital occupancy which are lagging



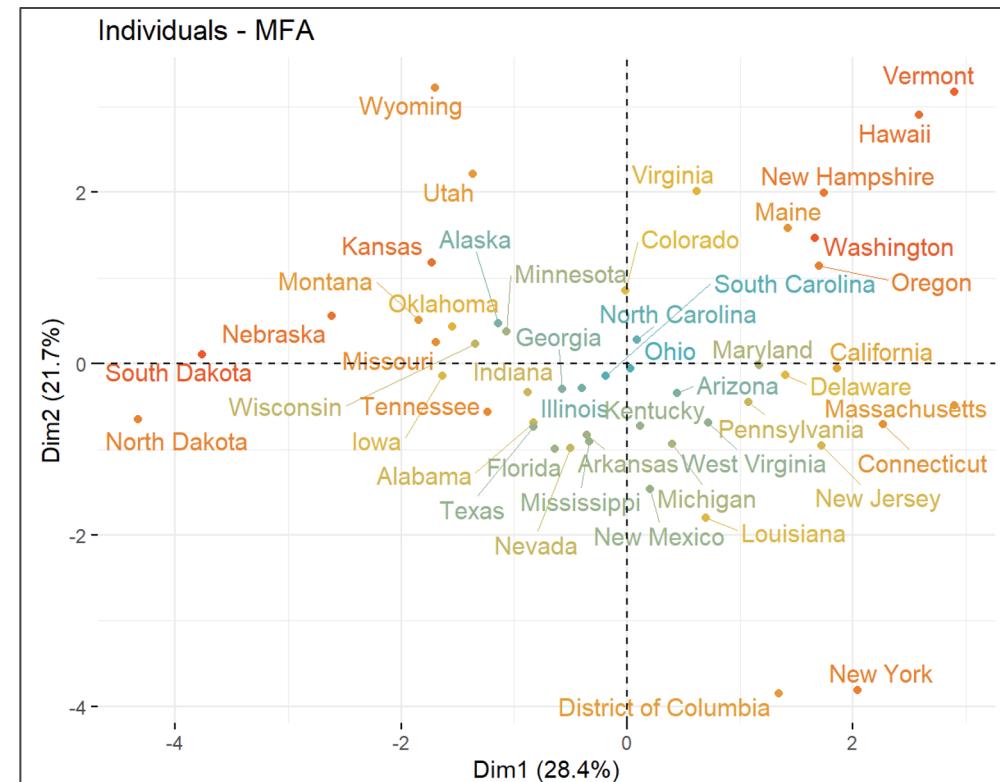
Groups

- coverage
- hosp\_beds
- hosp\_pct
- metrics

# Model Recommendation

## Social Distancing

- States with higher population densities tend to have higher death rates as shown along dimension 2 (possibly associated with lag in reporting)
- Groups per quadrant characterize a shift in affected areas as Dim 1 reflects recent developments
- Rural, less densely populated states are affected more by cases than deaths
- In mid-November case counts were rapidly exploding in these regions while deaths are lagged



# Case Study: Vermont

Vermont is situated near many states hit hardest by the onset of coronavirus, yet it has emerged as a model for state response to the pandemic

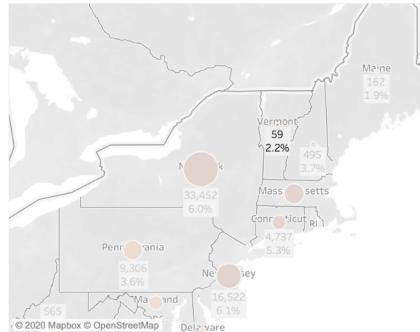
## Outcomes

- One of the lowest case and death rates of all states
- High rate of mask wearers chose always and few chose never

## Policy Implementation

- Quick to close and cautious to reopen
- Staggered reopening to monitor impact
- Strong centralized health authority to issue advice and guidance

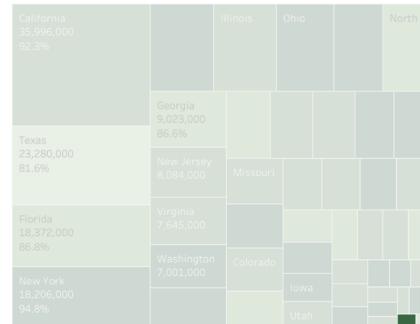
Death Number (size) & Rate (color)



Mask Use: Never, Rarely, Sometimes, Frequently, Always (Order by 'Always' DESC)

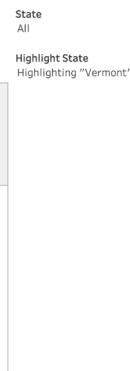
State	Never	Rarely	Sometimes	Frequently	Always
Delaware	9.6%	81.8%	80.6%	79.7%	78.0%
Hawaii	12.6%	80.5%	79.7%	77.2%	77.2%
Massachusetts	13.1%	79.7%	78.0%	74.8%	74.8%
Connecticut	12.6%	78.0%	77.2%	74.3%	74.3%
New York	13.4%	77.2%	76.0%	74.3%	74.3%
New Jersey	15.8%	74.8%	73.3%	71.5%	71.5%
District of Columbia	16.4%	74.3%	72.9%	71.5%	71.5%
Maryland	16.2%	74.3%	73.3%	71.5%	71.5%
California	15.9%	71.5%	70.0%	67.2%	67.2%
Pennsylvania	17.1%	67.2%	66.0%	65.5%	65.5%
New Mexico	16.0%	67.0%	65.5%	64.0%	64.0%
Washington	18.2%	67.0%	65.5%	63.8%	63.8%
<b>Vermont</b>	<b>13.8%</b>	<b>66.0%</b>	<b>65.5%</b>	<b>64.0%</b>	<b>64.0%</b>
Arizona	9.5%	14.7%	13.8%	12.9%	12.9%
Oregon	18.8%	64.0%	63.8%	62.9%	62.9%
Virginia	9.6%	17.8%	17.0%	16.2%	16.2%
Nevada	10.7%	19.4%	18.6%	17.8%	17.8%

Insurance Coverage Number (size) & Rate (color)



Pct Hospital Utilization (Order by 'Pct Inpatients With Covid' DESC)

State	Pct Icu Bed Utilization	Pct Inpatients With Covid
Washington	Pct Icu Bed Utilization	67.64
Washington	Pct Inpatients With Covid	6.11
Massachusetts	Pct Icu Bed Utilization	52.50
Massachusetts	Pct Inpatient Bed Utilization	68.04
New Hampshire	Pct Icu Bed Utilization	50.64
New Hampshire	Pct Inpatient Bed Utilization	68.05
New Hampshire	Pct Inpatients With Covid	4.02
Maine	Pct Icu Bed Utilization	74.92
Maine	Pct Inpatient Bed Utilization	69.32
Maine	Pct Inpatients With Covid	3.52
Hawaii	Pct Icu Bed Utilization	62.33
Hawaii	Pct Inpatient Bed Utilization	66.26
Hawaii	Pct Inpatients With Covid	3.35
<b>Vermont</b>	<b>Pct Icu Bed Utilization</b>	<b>46.77</b>
<b>Vermont</b>	<b>Pct Inpatient Bed Utilization</b>	<b>52.88</b>
<b>Vermont</b>	<b>Pct Inpatients With Covid</b>	<b>2.07</b>



# Case Study: The Dakotas

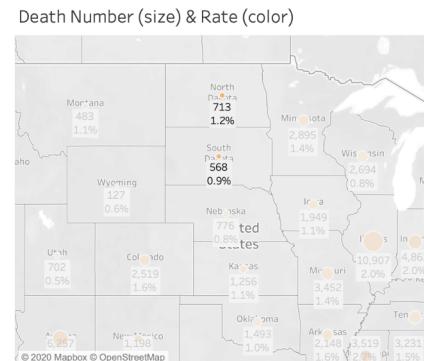
Among North and South Dakotas neighbors, these two states are performing much worse and stand out as examples of poor policy implementation and response

## Outcomes

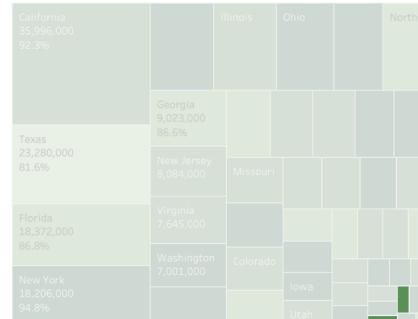
- Case rates are some of the highest in the country
- Death rates are relatively low indicating lag and a recent surge in those states

## Comparative Failures

- North Dakota has the lowest proportion of ‘always’ and the highest of ‘never’ for mask wearing by wide margins
- South Dakota has the highest percent of patients hospitalized with covid
- Both states neglected policy and proactive measures leaving defenses crippled



Insurance Coverage Number (size) & Rate (color)



Mask Use: Never, Rarely, Sometimes, Frequently, Always (Order by 'Always' DESC)

State	Never	Rarely	Sometimes	Frequently	Always
Arkansas	9.1%	8.7%	13.8%	22.1%	46.3%
Alaska	8.9%	—	32.9%	—	46.2%
Indiana	9.6%	9.6%	13.0%	23.7%	45.0%
Ohio	9.3%	—	15.0%	21.8%	43.9%
Wisconsin	8.7%	—	14.5%	23.2%	43.4%
Tennessee	11.0%	—	15.2%	21.0%	42.7%
Utah	9.7%	11.8%	—	27.2%	42.0%
Iowa	11.7%	—	16.7%	21.5%	39.6%
Kansas	11.7%	11.7%	15.1%	21.9%	39.6%
Oklahoma	12.1%	—	16.7%	22.6%	38.6%
Missouri	13.4%	12.7%	15.2%	21.1%	37.7%
Minnesota	12.2%	12.2%	15.6%	24.3%	37.6%
South Dakota	14.8%	16.7%	21.9%	—	36.1%
Nebraska	12.4%	11.8%	14.6%	26.1%	35.1%
Wyoming	15.2%	15.7%	12.3%	23.2%	33.6%
Montana	15.1%	17.2%	14.5%	21.5%	31.6%
North Dakota	18.1%	16.4%	14.6%	25.2%	25.7%

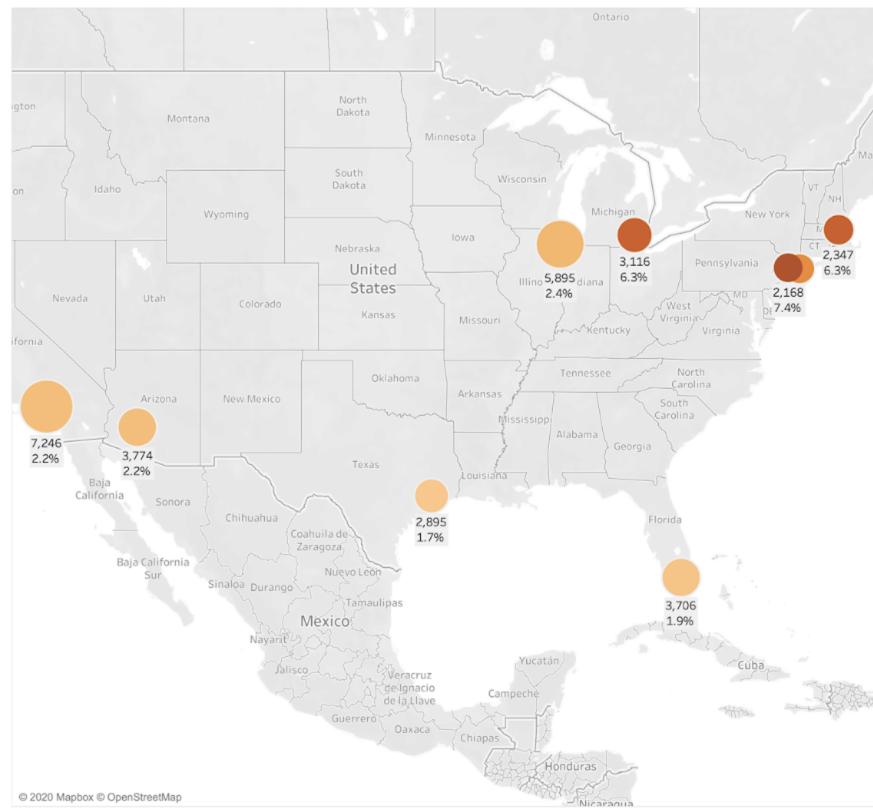
Pct Hospital Utilization (Order by 'Pct Inpatients With Covid' DESC)

State	Pct Icu Bed Utilization	Pct Inpatient Bed Utilizati...	Pct Inpatients With Covid
South Dakota	82.27	66.61	24.16
Nebraska	79.15	68.34	21.68
North Dakota	88.57	78.21	20.92
Wisconsin	70.98	71.52	20.76
Illinois	65.76	68.68	20.31
New Mexico	94.66	74.40	20.31

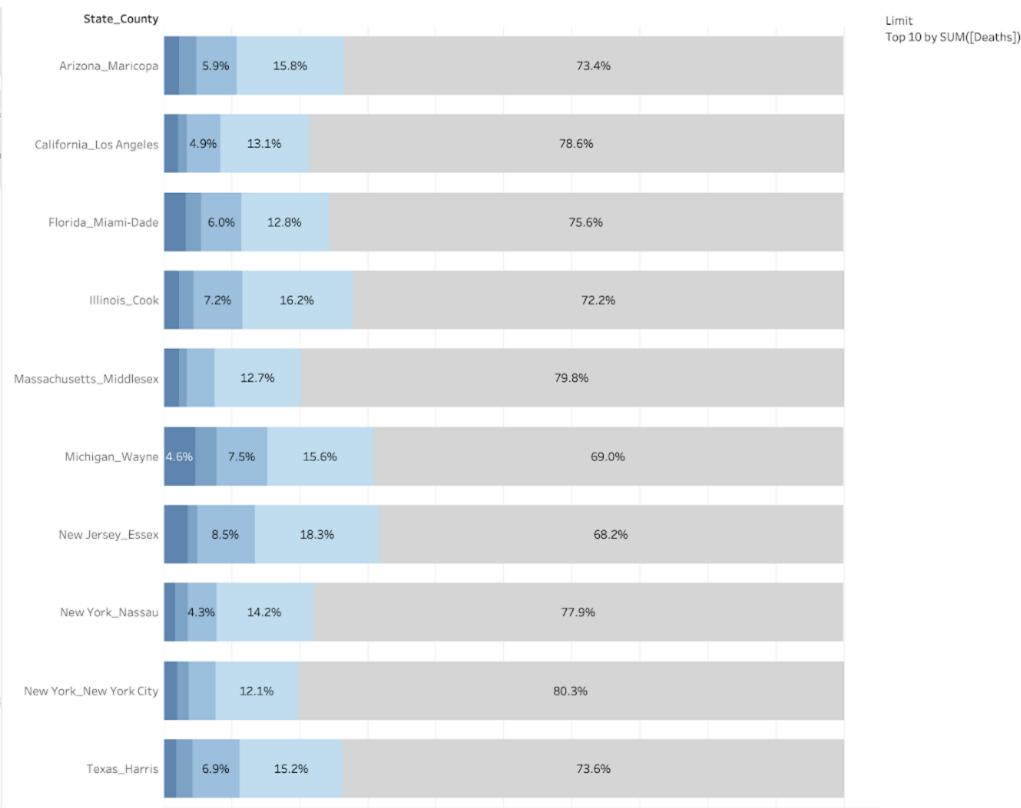
State All  
Highlight State  
Highlighting "Dakota"

# Tableau Dashboard: By County (Top 10 by Deaths)

Death Number (size) &amp; Rate (color)

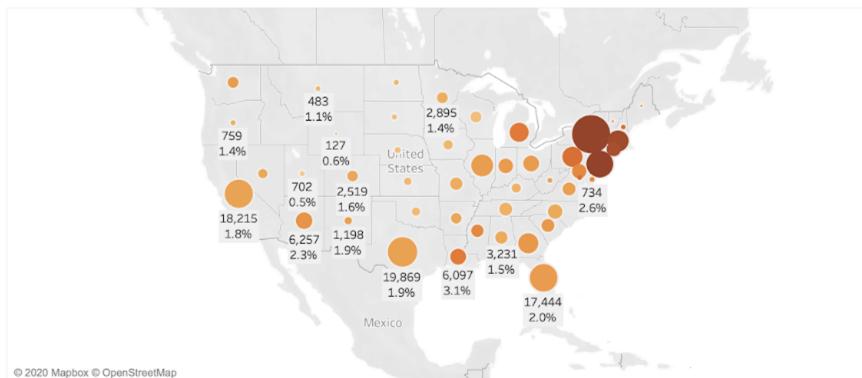


Mask Use: Never, Rarely, Sometimes, Frequently, Always



# Tableau Dashboard: By State

Death Number (size) &amp; Rate (color)



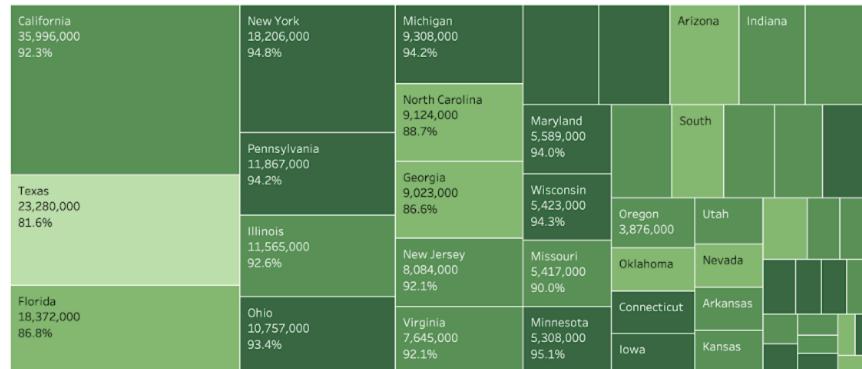
© 2020 Mapbox © OpenStreetMap

Mask Use: Never, Rarely, Sometimes, Frequently, Always (Order by 'Always' DESC)

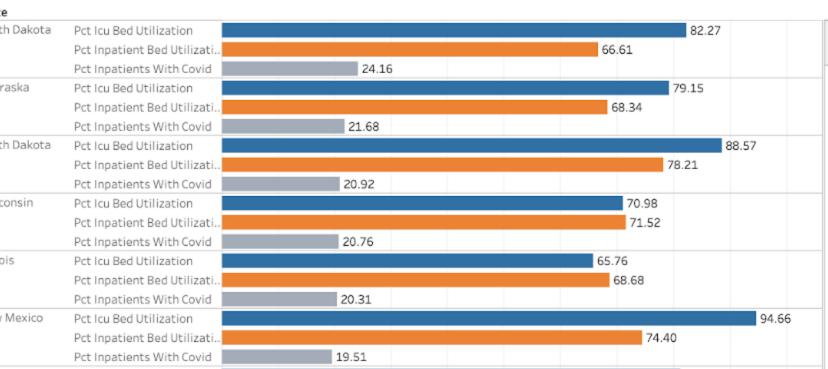
State	Never	Rarely	Sometimes	Frequently	Always
West Virginia	8.0%	7.7%	13.5%	21.8%	48.4%
Alabama	8.2%	8.1%	14.7%	21.5%	47.5%
Arkansas	9.1%	8.7%	13.8%	22.1%	46.3%
Alaska	5.6%	6.5%	8.9%	32.9%	46.2%
Indiana	8.5%	9.6%	13.0%	23.7%	45.0%
Ohio	9.1%	10.2%	15.0%	21.8%	43.9%
Wisconsin	8.7%	10.2%	14.5%	23.2%	43.4%
Tennessee	11.0%	10.1%	15.2%	21.0%	42.7%
Utah	9.3%	9.7%	11.8%	27.2%	42.0%
Iowa	10.4%	11.7%	16.7%	21.5%	39.6%
Kansas	11.7%	11.7%	15.1%	21.9%	39.6%
Oklahoma	10.0%	12.1%	16.7%	22.6%	38.6%
Missouri	13.4%	12.7%	15.2%	21.1%	37.7%
Minnesota	10.3%	12.2%	15.6%	24.3%	37.6%
South Dakota	14.8%	10.5%	16.7%	21.9%	36.1%
Nebraska	12.4%	11.8%	14.6%	26.1%	35.1%
Wyoming	15.2%	15.7%	12.3%	23.2%	33.6%
Montana	15.1%	17.2%	14.5%	21.5%	31.8%
North Dakota	18.1%	16.4%	14.6%	25.2%	25.7%

State  
AllHighlight State  
No items highlighted

Insurance Coverage Number (size) &amp; Rate (color)



Pct Hospital Utilization (Order by 'Pct Inpatients With Covid' DESC)



# Conclusions

## Lessons Learned

- Align information on collected data with all team members
  - Attribute definitions
  - Calculation procedures
  - Missing Data: the states of Idaho or Rhode Island not included
- Consider presentation of Tableau dashboards
  - Hard to present statistics for all states & counties simultaneously

## Future Scope

- Perform analysis on live Covid data
- Add stock data, industrial data and other metrics to get a comprehensive understanding Covid impacts
- Assess time-dependent associations between policy and infection

# Appendix: Our Team

Jaird Meyer



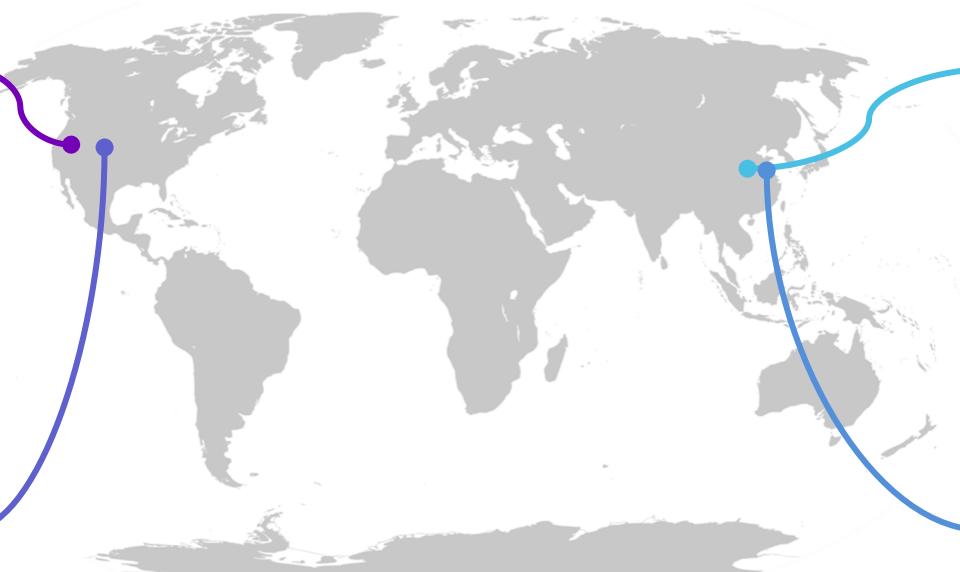
Elly Yang



Xiaojing Zhang



Mengyang Yu



# Appendix: Reference

