Social Distancing & Mobility

BLUEPRINT DATATHON
Solve Health Challenges with Data

Team Members: Ke Liu
Yimin Xiao

Main Models We Built

- **Infection Rate of COVID-19 *in US* v.s. *Social Distancing.***
  - -- By analyzing the US.csv from Data Set 5 & certain dates of social-distancing.csv from Data Set 1

- **Infection Rate of COVID-19 *in different states* in US v.s. State-level intervention**
  - -- By analyzing the US country and states level confirmed cases and death amount and rate, as well as the state -level intervention basically for social distancing intervention and public health actions.
  - Building of Trend Simulation Model

- **Infection Rate of COVID-19 *in different counties* in US v.s. *Mask-wearing situation***
  - -- By analyzing the us-counties.csv from Data Set 5, county_level_data.csv from Data Set 1 of Vertical two & mask-use-by-county.csv from Data Set 5

MODEL 1: Infection Rate of COVID-19 *in US* v.s. *Social Distancing.*

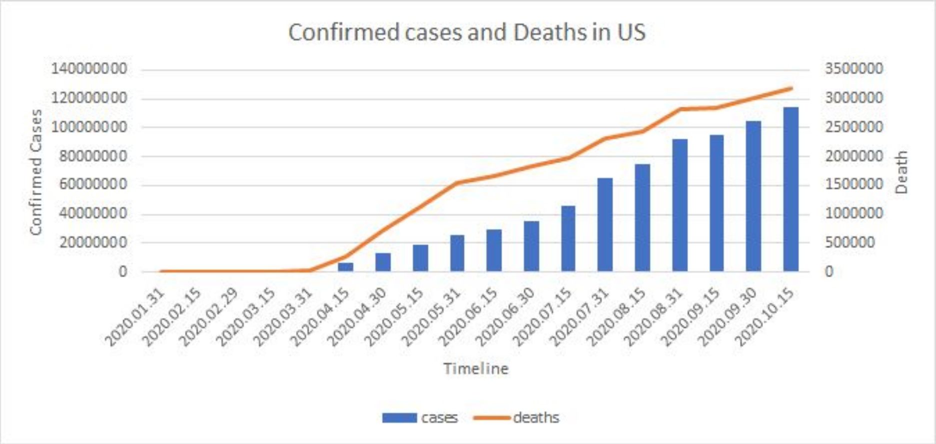


Figure 1.1

Main Ideas:

Analyzing the correlation between performance of Social Distancing and Infection and Death of COVID-19 in US.

Data usages:

Since the dataset is very large, we pick up sample data of certain dates (mid-month & end of month) to represent the trends of Infection and Death of COVID-19 in US.

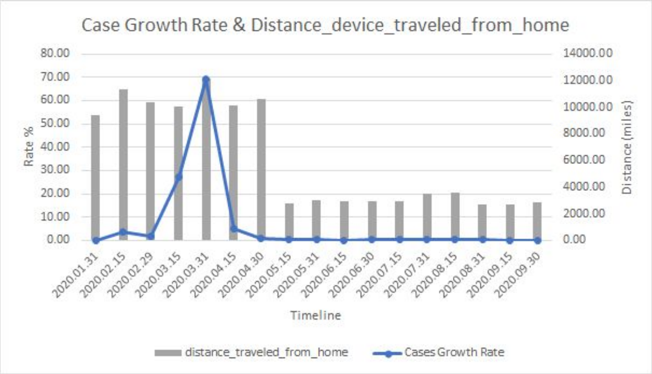


Figure 1.2

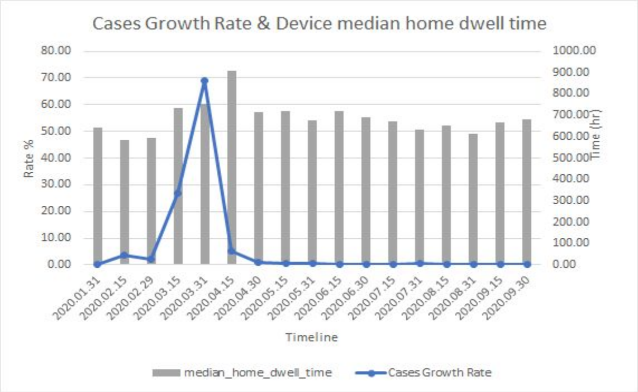


Figure 1.3

Interpretation:

The trend of convergence between the distance from home and the case growth rate can be observed from the graph.

From figure 1.2, "distance device traveled from home" can be interpreted as the distance of residents traveled from home. From figure 1.3, "device median home dwell time" can be interpreted as the time residents stay at home.

As several states declared state of emergency (Washington, Florida, California states, etc.) at the beginning of March, the distance of residents traveled has decreased, so did the cases growth rate.

MODEL 2: Infection Rate of COVID-19 *in different states* in US v.s. State-level intervention

Cases Trend Simulation Model

Location	Status of Reopening	Stay at Home Order
United States	Proceeding (19); Paused (5); New Restrictions Imposed (20); Reopened (7)	Original Stay at Home Order in Place (1); Stay at Home Order Eased or Lifted (44); No Action (6)

Table 2.1

Grading Scale (Set 'Status of Reopening' for example):
0: No response
1: Reopened
2: Proceeding Reopen
3: New Restrictions
4: Paused

Location	Waive Cost Sharing for COVID-19 Treatment	Free Cost Vaccine When Available
United States	State Requires (3); State-Insurer Agreement (3); No Action (45)	State Requires (9); State-Insurer Agreement (1); No Action (41)

Table 2.2

Grading Scale (Set 'Waive Cost Sharing for Covid-19 Treatment' for example):
0: No Action

2: State-Insurer Agreement

4: State Requires

Data Usage:

Based on the state-level intervention of 'social distancing actions' and 'covid-19 health policy actions';

Main Idea:

- Creating a scoring template from 0-4 degree to grade the actions in each states in terms of covid-19;
- Two-dimensional scoring: one is based on social distancing intervention; one is based on public health
- What we want: to see if the intervention factor in either social distancing or public health policy will have effect on the confirmed rate in different region

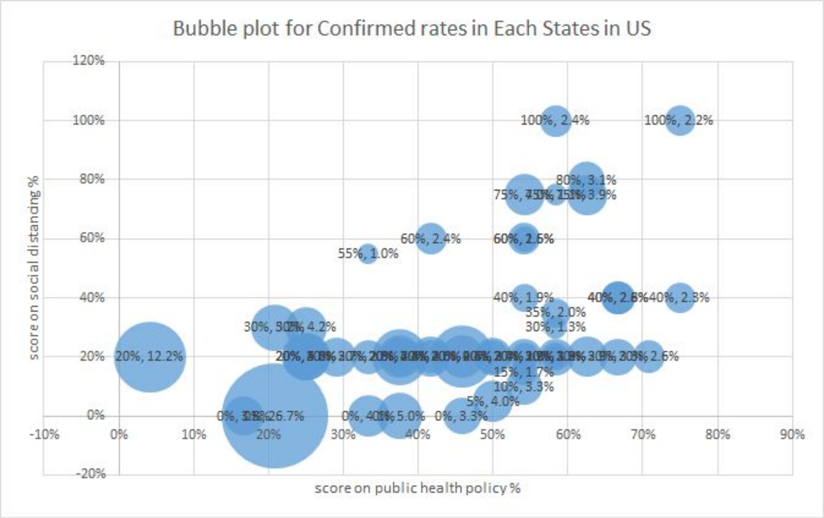


Figure 2.3

Data Analysis:

- The size of the bubbles indicates the Rate of Confirmed cases in each State.
- The vertical axis means score of social distancing intervention (%) and the horizontal axis means score of public health policy intervention (%).
- What we find: we can see that the less score the States has on social distancing, the larger the rate of the Confirmed cases the States will have; The less score the States has on public health policy intervention, the larger the rate of the Confirmed cases the States will have; vice versa.

Further Consideration:

Due to the information limit, we don’t have a specific timeline of the real-time reaction of the states. So it is not quite possible to trace back that how the state-level intervention will affect the increasing rate of confirmed or death data during this year.

However, by using the scoring template we create for all states, it is possible to predict an overall altitude of each states towards the pandemic, therefore, which can be related to the current accumulated confirmed cases. Thus , we can use it as a prediction to future trend of confirmed cases as well.

Confirmed Cases & Death Prediction (Next 6 months)

N	O	P
states	average_score	affected Increasing rate
Alabama	28%	0.847368118
Arizona	39%	0.496166123
Arkansa	29%	0.897238087
Californi	58%	1.011480445
Colorad	47%	0.645255404
Connect	39%	0.171986579
Delaware	57%	0.514140981
District (	79%	0.440924542
Florida	25%	0.736075418
Georgia	29%	0.837489724
Idaho	25%	1.084445133
Illinois	37%	0.583190476
Indiana	27%	0.672009849
Iowa	23%	0.71000873
Kansas	33%	0.979004414
Kentuck	69%	1.68460276
Louisian	65%	0.874964405
Maine	67%	0.498688014
Marylanc	51%	0.538947738
Massach	88%	0.229236269
Michigan	47%	0.570327166
Minneso	32%	0.731057112
Mississip	31%	0.865699946
Missouri	8%	0.96710997
Montana	33%	1.858417258

Figure 2.4

- According to the average scoring for each states, we can calculate out the affected rate that about to change the increase amount of both confirmed cases and death prediction per month. And we keep this affected increasing rate as a constant scalar for the next following 6 months as a prediction.
- We set the 0.5 as the normal score, the States which do better(higher intervention scoring), will have a predict rate higher than before, vice versa. (Refering to the formula in Figure 2.5)

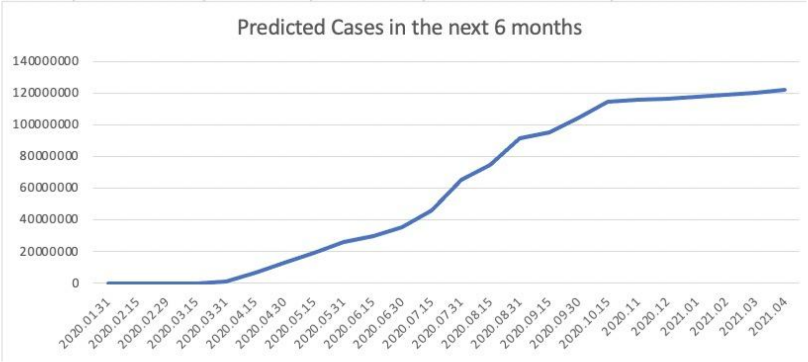
=M2*(1+O2-0.5)

N	O	P
states	average_score	affected Increasing rate
Alabama	28%	0.847368118
Arizona	39%	0.496166123
Arkansa:	29%	0.897238087
Californi	58%	1.011480445
Colorad	47%	0.645255404
Connect	39%	0.171986579

Figure 2.5

B	C	D	E	F	G	H	I	J	K	L	M	N	O
states	score_pe	score_	Rate	Rate of de	Confimed Cases in OCT	Confirmed Case in Sept	Confirmed per month	Death in OCT	Death in Sept	Death per Month	Increase	average_score	affected
Alabama	50%	5%	4.0%	1.6%	169162	140160	29002	2756	2387	369	1.0934	28%	0.8474
Arizona	58%	20%	3.3%	2.5%	228748	209215	19533	5789	5346	443	0.5564	39%	0.4962
Arkansa:	38%	20%	4.3%	1.7%	96524	71497	25027	1645	1150	495	1.1393	29%	0.8972
Californi	75%	40%	2.3%	1.9%	870380	769917	100463	16839	14606	2233	0.9409	58%	1.0115
Colorad	54%	40%	1.9%	2.4%	82443	62575	19868	2187	2006	181	0.6646	47%	0.6453
Connect	58%	20%	1.9%	6.9%	62028	55031	6997	4540	4485	55	0.1929	39%	0.172
Delaware	54%	60%	2.5%	2.9%	22560	19137	3423	661	618	43	0.4801	57%	0.5141
District (	58%	100%	2.4%	3.9%	16166	14687	1479	638	616	22	0.3414	79%	0.4409
Florida	29%	20%	3.7%	2.1%	744980	668838	76142	15735	12786	2949	0.9869	25%	0.7361
Georgia	38%	20%	3.4%	2.3%	321400	280356	41044	7320	6255	1065	1.0635	29%	0.8375
Idaho	21%	30%	5.2%	1.0%	51471	36306	15165	521	425	96	1.4379	25%	1.0844
Illinois	54%	20%	3.2%	2.6%	336211	267380	68831	9392	8589	803	0.6697	37%	0.5832
Indiana	33%	20%	2.8%	2.6%	143617	109445	34172	3864	3460	404	0.8765	27%	0.672
Iowa	25%	20%	5.6%	1.4%	104232	75722	28510	1516	1234	282	0.9793	23%	0.71
Kansas	46%	20%	4.5%	1.3%	70815	51006	19809	838	555	283	1.1807	33%	0.979
Kentuck	63%	75%	3.9%	1.5%	86410	61986	24424	1341	1112	229	1.4186	69%	1.6846
Louisian	54%	75%	4.0%	3.2%	176952	159253	17699	5707	5278	429	0.7636	65%	0.875
Maine	58%	75%	1.1%	2.4%	5836	4918	918	144	137	7	0.4274	67%	0.4987
Marylanc	42%	60%	2.4%	2.9%	134042	117745	16297	4028	3849	179	0.5345	51%	0.5389
Massach	75%	100%	2.2%	6.6%	141579	125393	16186	9672	9225	447	0.1667	88%	0.2292
Michigan	58%	35%	2.0%	4.4%	156926	124978	31948	7304	6933	371	0.59	47%	0.5703
Minneso	54%	10%	3.3%	1.8%	117159	85403	31756	2252	1979	273	0.8906	32%	0.7311
Mississip	42%	20%	4.0%	2.8%	108139	90523	17616	3152	2734	418	1.071	31%	0.8657
Missouri	17%	0%	3.8%	1.6%	156142	107687	48455	2531	1830	701	1.6579	8%	0.9671
Montana	46%	20%	9.4%	1.1%	21043	9281	11762	230	141	89	2.2413	33%	1.8584

Figure 2.6



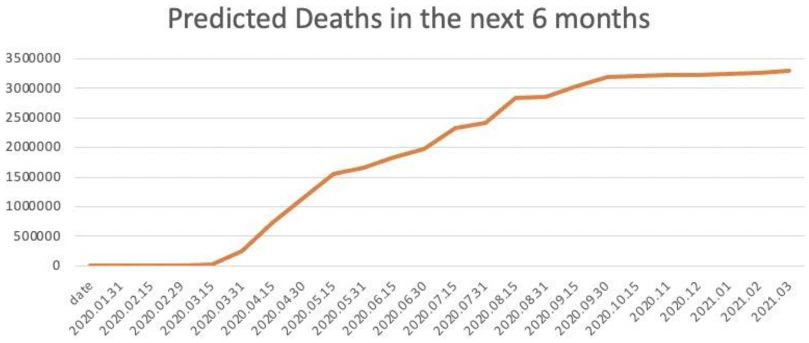


Figure 2.7

- Thus, this is the simulation of the predict cases in the next 6 months, which the rate of the increase will be slower than usual.

Deviation Analysis:

- The slower rate probably based on the scoring level of the built model, and the overall score are higher as normal, thus, this will cause the curve to bent down in prediction.

MODEL 3: Infection Rate of COVID-19 *in different counties* in US v.s. *Mask-wearing situation*

Main Ideas:

Analyzing the correlation between the frequency of population wearing masks in different counties in US and the infection rate(x10).

Data usages:

Since the us-counties.csv dataset is very large, we pick up sample data of three states(California, Michigan and Missouri), combining with the frequency of population wearing masks in different counties, to study the trends of infection rate of COVID-19 corresponding to mask-wearing situation.

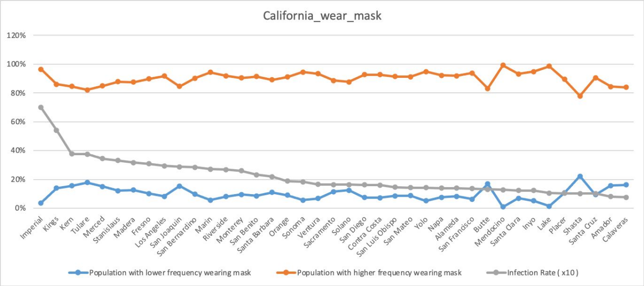
California is chosen because Stanford University is based in California, and is also a representative of western states.

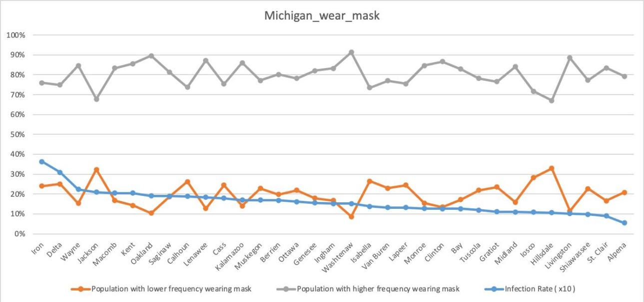
Michigan is chosen because our school, University of Michigan, is based in Michigan, and is also a representative of northeast states.

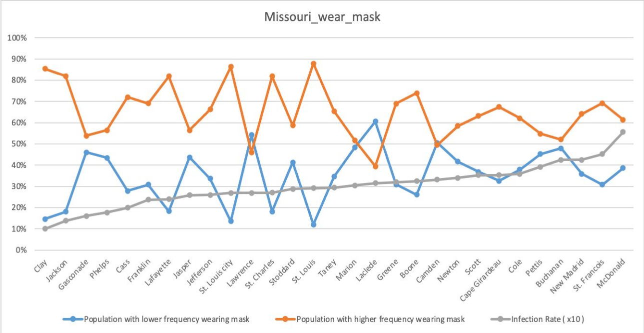
Missouri is chosen because it is a representative of midwestern states.

Since the infection rate is too low, we multiple it by 10 to make the changes more obvious.

We used the data of 'population estimate 2018' of each county from 'county_level_data.csv' of Vertical two to calculate the infection rate. However, this is the estimated data of 2018, not the current accurate population, and also some of the county’s population is not available, so we left out those counties and only assess the confirmed rate of those available. While calculating the population of states. We didn’t actually have all the counties but only some of the main counties, so there will be errors in the results.







Based on the figures on the left, there is no obvious relationship between the infection rate of COVID-19 and the frequency of people wearing masks.

However, according to CDC, it is recommended for people to wear masks in public and when around people who don't live in your household. Wearing a mask is medically proven to be effective in preventing the spread of the coronavirus. Why the frequency of wearing mask neither makes data increase slowly nor lower the infection rate, obvious relationship shown here according to the data?

We think the main reasons are that the surgical masks and KN95 masks should be the standard masks to prevent spreading of coronavirus, but since lots of people are not wearing a mask that meets the standard requirement, they did not protect themselves from being infected in fact.