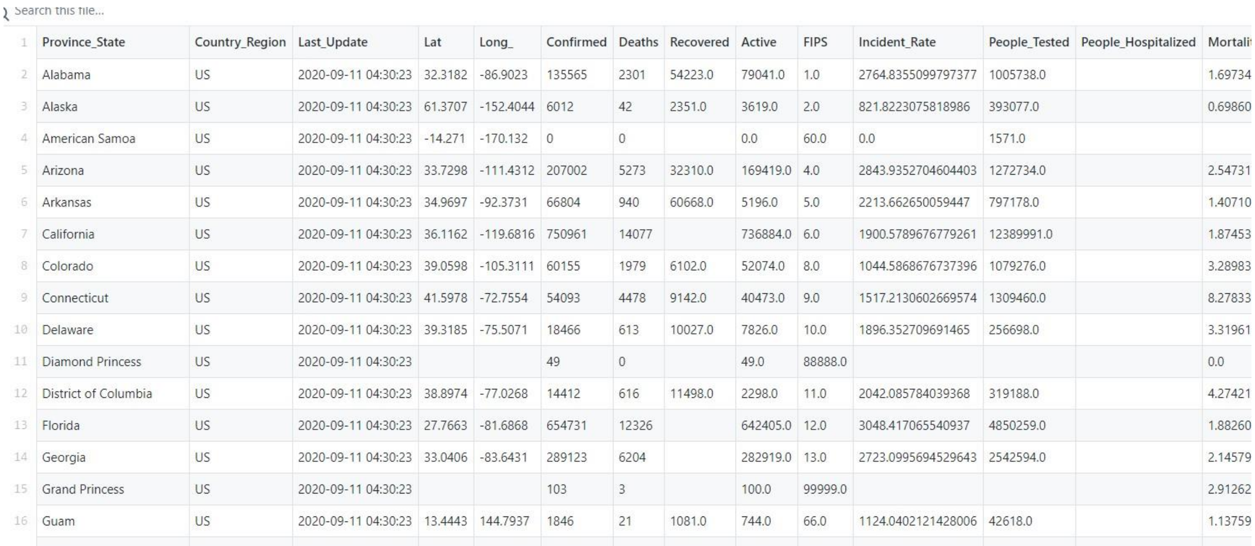
Covid-19 Cases for World and US

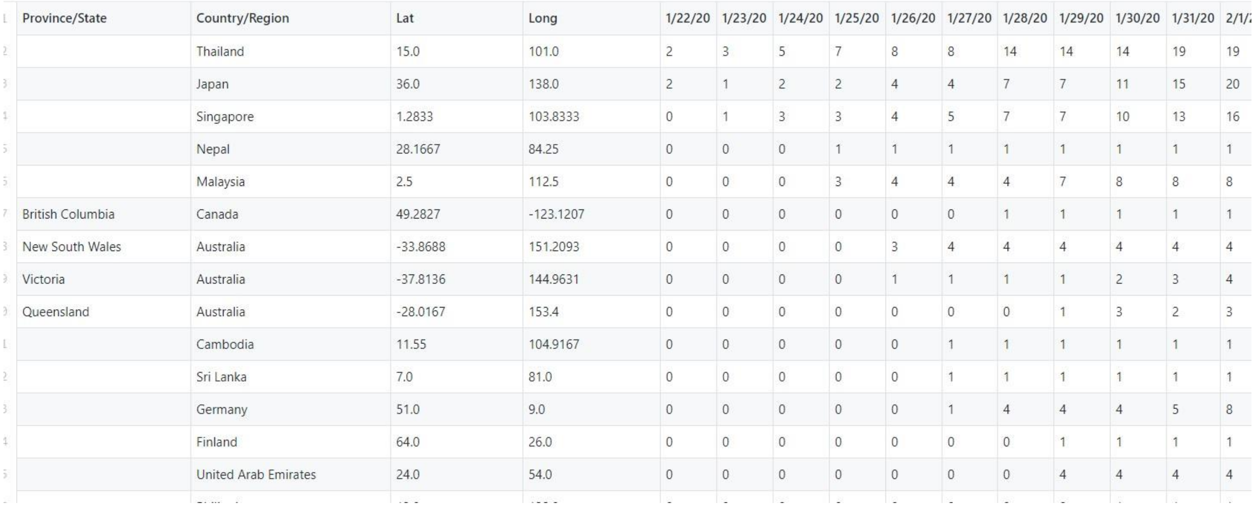
Zhou Sun

Jinfa Zhu

In this project, we are using the data from Johns Hopkins University about daily cases in each state in the United States and world cases on a specific date. Attached is a sample of one day of United State Cases and one day of World Cases. (at: https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_daily_reports_us/09-10-2020.csv)



Province_State	Country_Region	Last_Update	Lat	Long_	Confirmed	Deaths	Recovered	Active	FIPS	Incident_Rate	People_Tested	People_Hospitalized	Mortalit
Alabama	US	2020-09-11 04:30:23	32.3182	-86.9023	135565	2301	54223.0	79041.0	1.0	2764.8355099797377	1005738.0		1.69734
Alaska	US	2020-09-11 04:30:23	61.3707	-152.4044	6012	42	2351.0	3619.0	2.0	821.8223075818986	393077.0		0.69860
American Samoa	US	2020-09-11 04:30:23	-14.271	-170.132	0	0		0.0	60.0	0.0	1571.0		
Arizona	US	2020-09-11 04:30:23	33.7298	-111.4312	207002	5273	32310.0	169419.0	4.0	2843.9352704604403	1272734.0		2.54731
Arkansas	US	2020-09-11 04:30:23	34.9697	-92.3731	66804	940	60668.0	5196.0	5.0	2213.662650059447	797178.0		1.40710
California	US	2020-09-11 04:30:23	36.1162	-119.6816	750961	14077		736884.0	6.0	1900.5789676779261	12389991.0		1.87453
Colorado	US	2020-09-11 04:30:23	39.0598	-105.3111	60155	1979	6102.0	52074.0	8.0	1044.5868676737396	1079276.0		3.28983
Connecticut	US	2020-09-11 04:30:23	41.5978	-72.7554	54093	4478	9142.0	40473.0	9.0	1517.2130602669574	1309460.0		8.27833
Delaware	US	2020-09-11 04:30:23	39.3185	-75.5071	18466	613	10027.0	7826.0	10.0	1896.352709691465	256698.0		3.31961
Diamond Princess	US	2020-09-11 04:30:23			49	0		49.0	88888.0				0.0
District of Columbia	US	2020-09-11 04:30:23	38.8974	-77.0268	14412	616	11498.0	2298.0	11.0	2042.085784039368	319188.0		4.27421
Florida	US	2020-09-11 04:30:23	27.7663	-81.6868	654731	12326		642405.0	12.0	3048.417065540937	4850259.0		1.88260
Georgia	US	2020-09-11 04:30:23	33.0406	-83.6431	289123	6204		282919.0	13.0	2723.0995694529643	2542594.0		2.14579
Grand Princess	US	2020-09-11 04:30:23			103	3		100.0	99999.0				2.91262
Guam	US	2020-09-11 04:30:23	13.4443	144.7937	1846	21	1081.0	744.0	66.0	1124.0402121428006	42618.0		1.13759



Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	1/31/20	2/1/
	Thailand	15.0	101.0	2	3	5	7	8	8	14	14	14	19	19
	Japan	36.0	138.0	2	1	2	2	4	4	7	7	11	15	20
	Singapore	1.2833	103.8333	0	1	3	3	4	5	7	7	10	13	16
	Nepal	28.1667	84.25	0	0	0	1	1	1	1	1	1	1	1
	Malaysia	2.5	112.5	0	0	0	3	4	4	4	7	8	8	8
British Columbia	Canada	49.2827	-123.1207	0	0	0	0	0	0	1	1	1	1	1
New South Wales	Australia	-33.8688	151.2093	0	0	0	0	3	4	4	4	4	4	4
Victoria	Australia	-37.8136	144.9631	0	0	0	0	1	1	1	1	2	3	4
Queensland	Australia	-28.0167	153.4	0	0	0	0	0	0	0	1	3	2	3
	Cambodia	11.55	104.9167	0	0	0	0	0	1	1	1	1	1	1
	Sri Lanka	7.0	81.0	0	0	0	0	0	1	1	1	1	1	1
	Germany	51.0	9.0	0	0	0	0	0	1	4	4	4	5	8
	Finland	64.0	26.0	0	0	0	0	0	0	0	1	1	1	1
	United Arab Emirates	24.0	54.0	0	0	0	0	0	0	0	4	4	4	4

There are a couple problems with regard to the above databases. First, we will be using a huge dataset for United States Cases (every day from Apr. 13th until Sep. 9th). There are missing values in columns and some of the numbers are inconsistent. For instance, number of confirmed cases are reported daily, while number of death and recovery are reported accumutively. Therefore, when we analyze, we need to take those into account. We will use spark to speed up the preprocess of these data including data cleaning and extraction.

We are thinking about using PostgreSql as a Relational Database for US daily cases. All of our entries are numbers, and the only thing we are thinking about doing is normalization for better comparison. We will also be using MongoDB Atlas as nosql, as the world case is comparably small, thus one model (Country) with fields will satisfy the requirement.

For the interface, we are about to present the data with proper ways of visualization including forms for direct access of raw data and different kinds of charts for analyzing the selected data relative to population and regions.

Zhou is a senior in Mechanical Engineering with a good programming background. He wrote a website using ruby on rails and another using React and node.js. He is highly familiar with mongoDB and node js packages. He also wrote a Wechat Mini App using JS.

Jinfa is a second year master student majoring in computer science. His relevant skill sets include machine learning, python, spark, and sql.

Timeline:

Date	Things to accomplish
Sep 28th	Have data cleaning done
Oct 4th	Have firebase API and mongo DB done
Oct 21st	Have front-end done (either in Web or Wechat Mini App)
Nov 23rd	Final Presentation