SARS-CoV-2:
Past, Present, and Future

By
Onur Baser, Dan Brims and Samir Sidi

The Questions We Posed

With the Covid pandemic top on everyone's mind we decided explore the past, present, and future of this disease.
We utilized APIs to access public covid databases and correlate that data with additional metrics to gauge the past, present and future of this disease.

- Past (Samir): Covid data was correlated with economic data to see when and how disease prevalence affected the markets.
- Present (Onur): Covid disease metrics (cases, deaths), are correlated with healthcare spending, health care resources (doctors & nurses), health care capacity (doctor visits, acute bed and hospital utilization) and the immune system (vegetable usage, alcohol consumption), to assess which countries have dealt best with the pandemic.
- Future (Dan): Using historical data on death rate and rate of new cases, the number of new cases were projected using and SEIR model out to 6/30/2020, to determine, when it might be safe to start easing the current lockdowns.

Future

Data Used and Data Cleaning:

1. APIs or data sources
    a. The COVID Tracking Project
    b. Coronavirus COVID19 API
    c. DELPHI COVID project at MIT
2. Data processing
    a. When scanning through country data, detecting if the API call was successful and returned data or was unsuccessful or returned no data.
    b. Converting date from strings to a unified datetime format with truncation of hours.
    c. Determining if data fields are populated or None, and explicitly replacing the Nones with “Null” or “None”.

Data Analysis:

1. Historical data was formated and inputted into a SEIR model. This model stands for susceptible (S), exposed (E), infected (I), and recovered. A series of simultaneous differential equations are fitted to the historical data, with minimization of residuals using scipy. The fitted parameter are then used to predict cases and deaths out to 6/30/2020.
2. The shape of the daily new case curve is analyzed to determine when the cases peak in that geography.
3. One of the criteria proposed for safe reopening is a decline in new cases for at least 14 consecutive days. Based on the peak date this safe reopen data is proposed

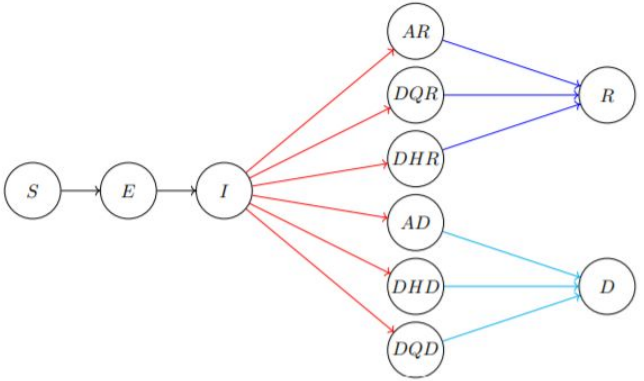


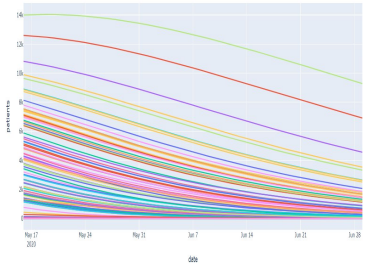
Figure 1: Simplified Flow Diagram of DELPHI (V2)

$$\frac{dS}{dt} = -\alpha\gamma(t)S(t)I(t)$$
$$\frac{dE}{dt} = \alpha\gamma(t)S(t)I(t)$$
$$\frac{dI}{dt} = r_i E(t) - r_d I(t)$$
$$\frac{dAR}{dt} = r_d(1-p_{dth})(1-p_d)I(t) - r_{ri}AR(t)$$
$$\frac{dDHR}{dt} = r_d(1-p_{dth})p_d p_h I(t) - r_{rh}DHR(t)$$
$$\frac{dDQR}{dt} = r_d(1-p_{dth})p_d(1-p_h)I(t) - r_{ri}DQR(t)$$
$$\frac{dAD}{dt} = r_d p_{dth}(1-p_d)I(t) - r_{dth}AD(t)$$
$$\frac{dDHD}{dt} = r_d p_{dth} p_d p_h I(t) - r_{dth}DHD(t)$$
$$\frac{dDQD}{dt} = r_d p_{dth} p_d(1-p_h)I(t) - r_{dth}DQD(t)$$
$$\frac{dTH}{dt} = r_d p_d p_h I(t)$$
$$\frac{dDD}{dt} = r_{dth}(DHD(t) + DQD(t))$$
$$\frac{dDT}{dt} = r_d p_d I(t)$$
$$\frac{dR}{dt} = r_{ri}(AR(t) + DQR(t)) + r_{rh}DHR(t)$$
$$\frac{dD}{dt} = r_{dth}(AD(t) + DQD(t) + DHD(t))$$

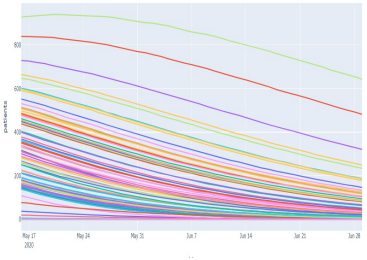
Discussion: The modeling was only done on a small subset of geographies, due to time it took each model to finish (~45 min). Below I am showing both MN and NY as representative examples of the data. Most geographies we looked at are on the downward slope and would be safe to reopen by 5/27/2020.

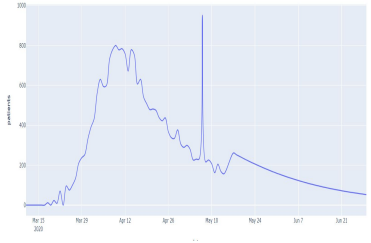
New York

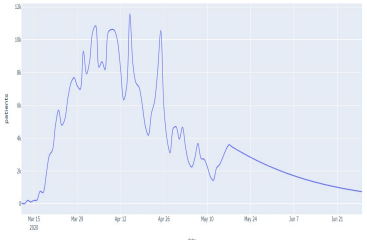
Deaths

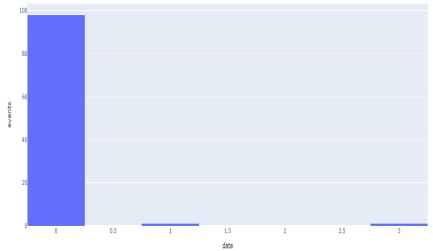


Cases



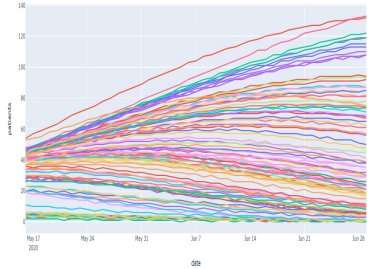




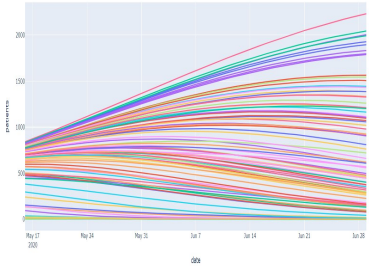


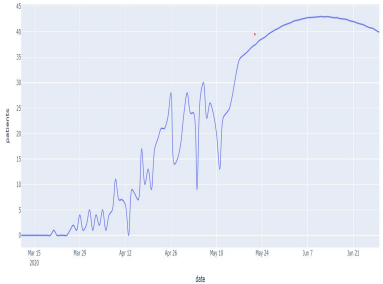
Minnesota

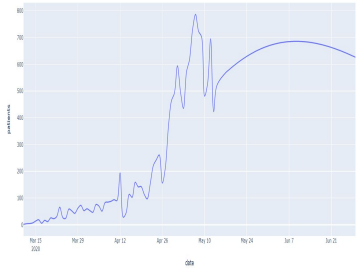
Deaths

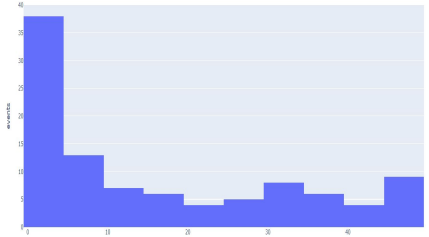


Cases









Most geographies we looked at are on the downward slope and based on this metric alone would be safe to reopen by 5/30/2020, but a few still have patient numbers increasing, should not even think of reopening until at least the middle of June.

	region	average time to peak	stdev	peak date	open date
0	CA	0	0.24	2020-05-16	2020-05-30
1	IA	0	0.54	2020-05-16	2020-05-30
2	NY	0	0.32	2020-05-16	2020-05-30
3	GA	0	0.41	2020-05-16	2020-05-30
4	NJ	0	0.54	2020-05-16	2020-05-30
5	NC	0	0.52	2020-05-16	2020-05-30
6	MN	15	16.26	2020-05-31	2020-06-14
7	FL	0	0.90	2020-05-16	2020-05-30
8	TX	0	0.30	2020-05-16	2020-05-30
9	Italy	0	0.52	2020-05-16	2020-05-30
10	Brazil	12	10.52	2020-05-28	2020-06-11
11	Singapore	10	15.16	2020-05-26	2020-06-09
12	Sweden	0	0.24	2020-05-16	2020-05-30
13	France	0	1.86	2020-05-16	2020-05-30
14	Canada	0	0.20	2020-05-16	2020-05-30
15	Turkey	0	0.50	2020-05-16	2020-05-30
16	India	14	13.07	2020-05-30	2020-06-13

Present

Data Used and Data Cleaning:

1. Data from OECD website and WHO Covid website downloaded
2. Data processing
    a. Y-axis was Covid Survival relative to OECD Mean
    b. X-axis changed depending on the explanatory variable
    c. Bubbles are used. Each buble is proportional to population size. Colors shows the pre-covid health of the population. (The classification of countries begin close to, better or worse than the OECD average is based on indicator’s standard deviation)

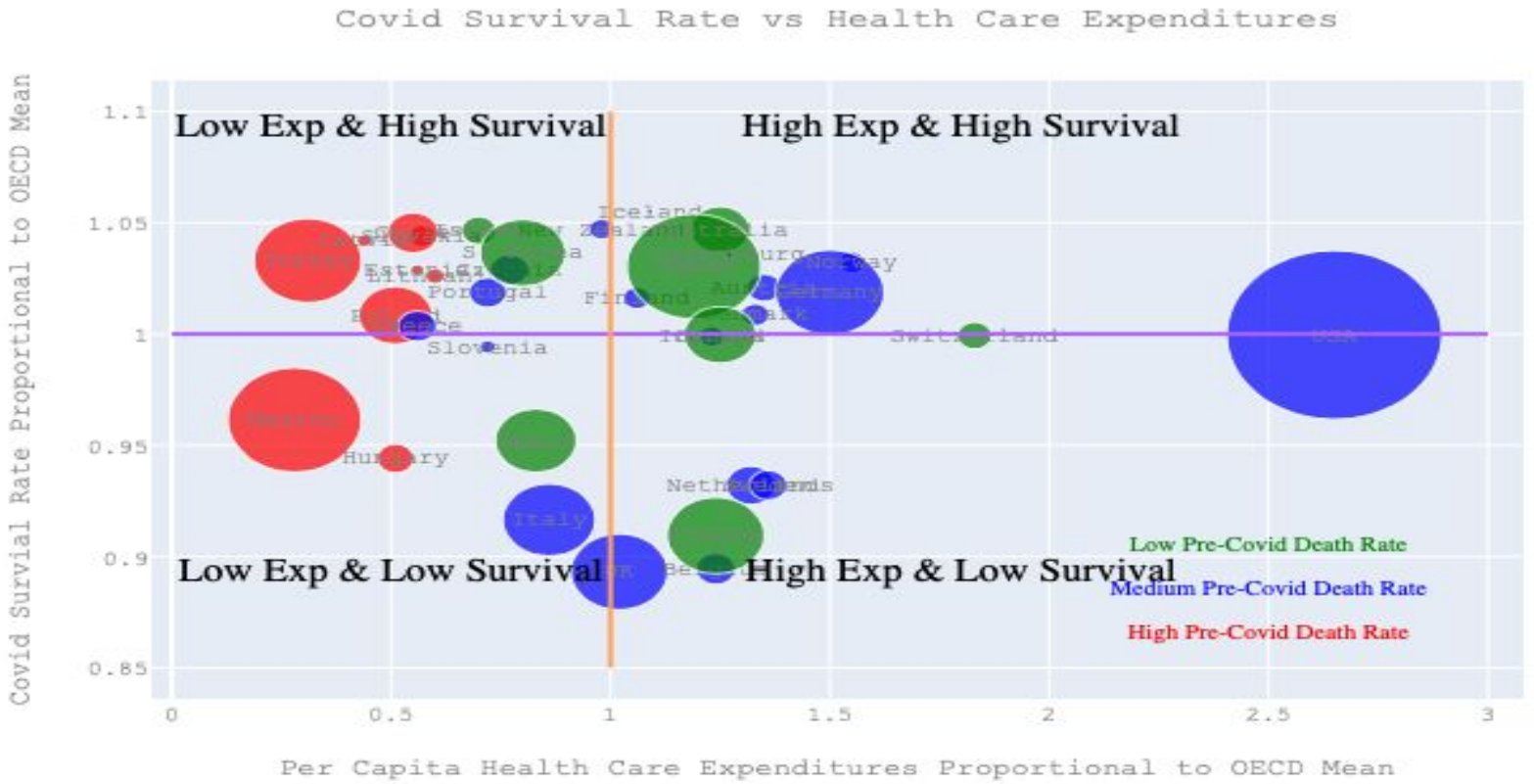
Quadrant Charts:

1. They will show simple associations at a macro level between the indicators rather than causal relationships.

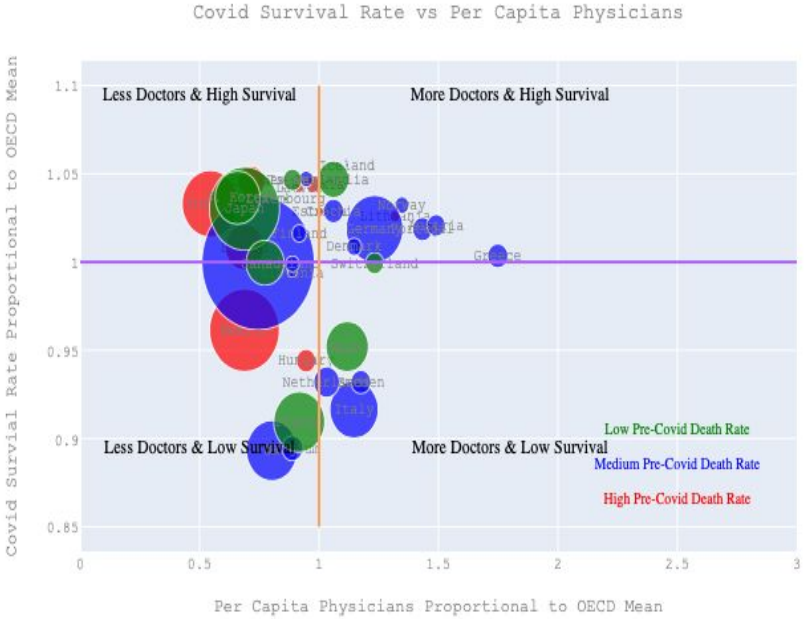
2. The purpose here is to stimulate deeper discussions on policy priority settings, by highlighting areas where countries could do better.

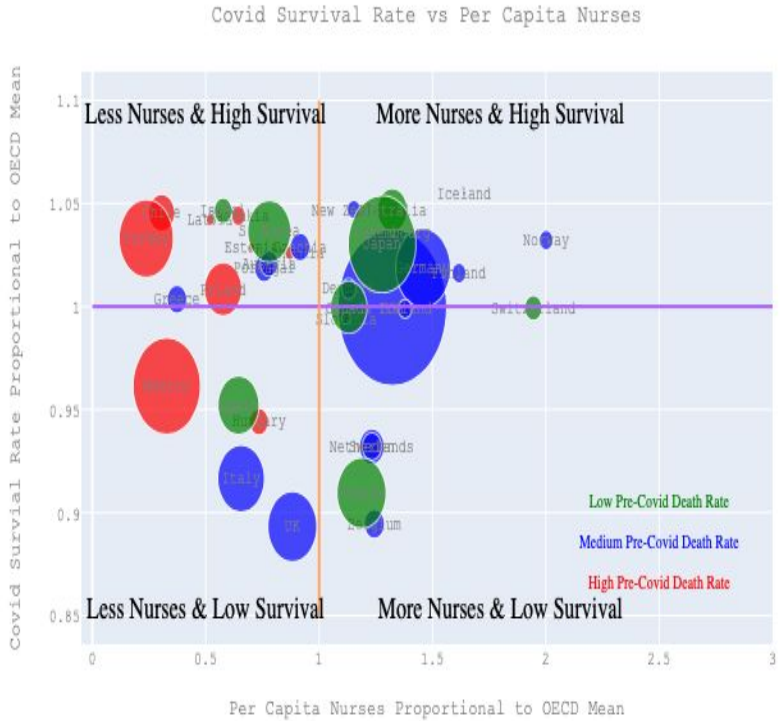
3. The center of each quadrant chart is the OECD average

Discussion: Countries like Mexico, Hungary were expected to have low survival rate. France unexpectedly have low survival rates. Germany, Japan are expected to have higher survival rates, and they did. Poland, Turkey, Israel, S.Korea did unexpectedly well.

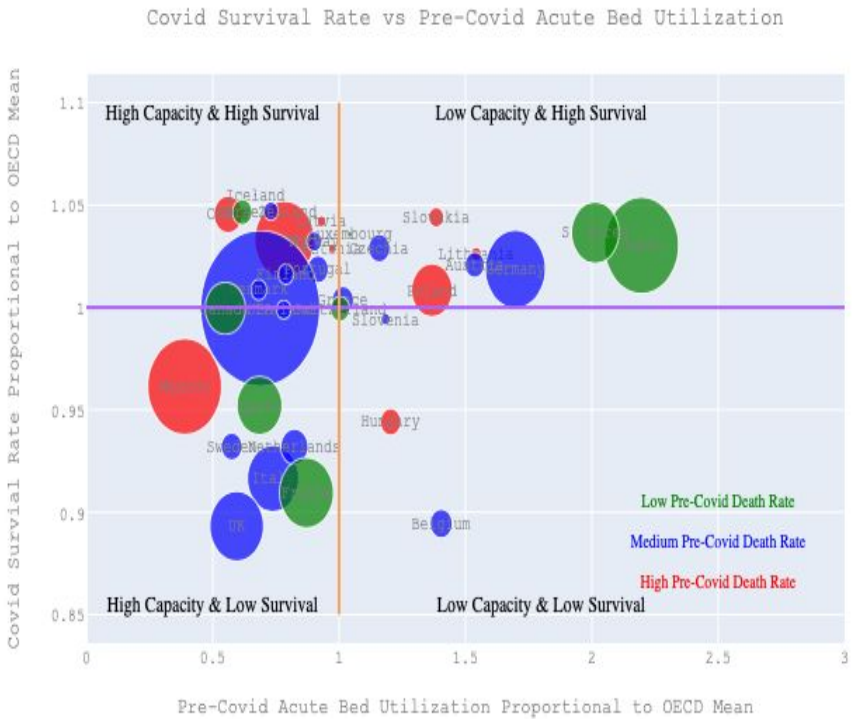


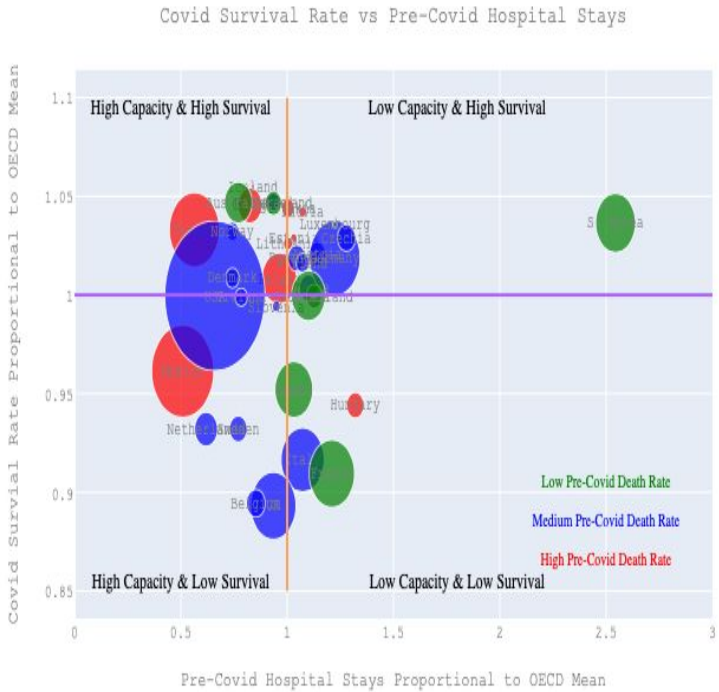
Resources: US spends the most money on health care but actually they have a fewer doctor than OECD average. Italy and Spain have more doctors, Sweden and Netherlands have more doctors and nurses but have lower survival rate.



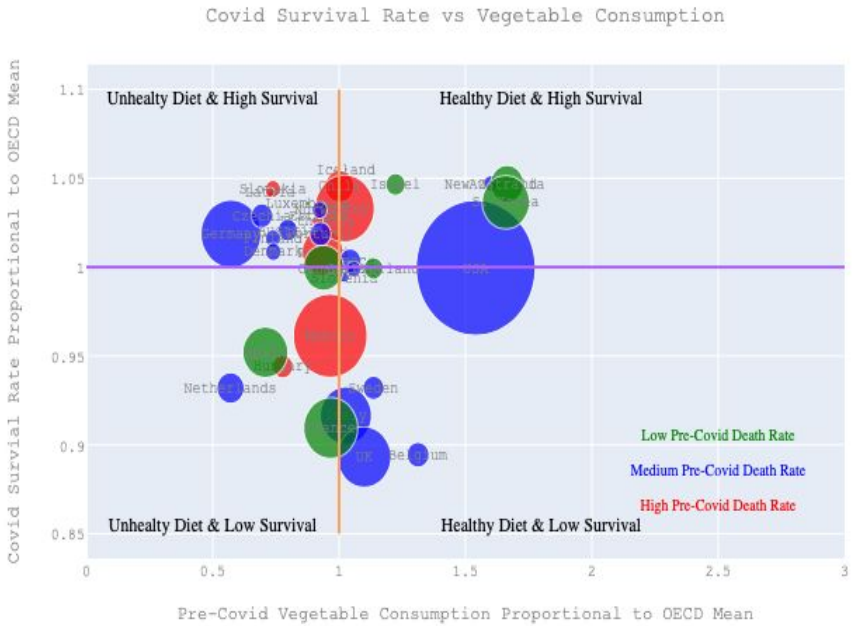


Capacity: Hospital Stays and Acute Bed Utilization. Japan and S.Korea has the highest utilization. Belgium and Italy had the low capacity pre-covid and as expected, they had low survival.





Immunity: Vegetable Consumption and Younger Population. Netherlands has the lowest vegetable consumption among OECD countries and also has older population. Australia, S. Korea and New Z. has the highest vegetable consumption and as expected high survival rates. Japan has the oldest population and did unexpectedly well.



Past

Data Used and Data Cleaning:

1. APIs or data sources
    a. Quandl API (Crude Oil (WTI), S&P, Gold)
    b. Human Data Org (COVID Csv)
2. Data processing
    a. Initially used Federal Reserve Data (FRED) and U.S. Energy Information Administration API but datasets returned errors in Python because of datetime and index incongruity.
    b. Quandl API very user friendly and easy to call.
    c. Quandl datasets indexed properly and current.

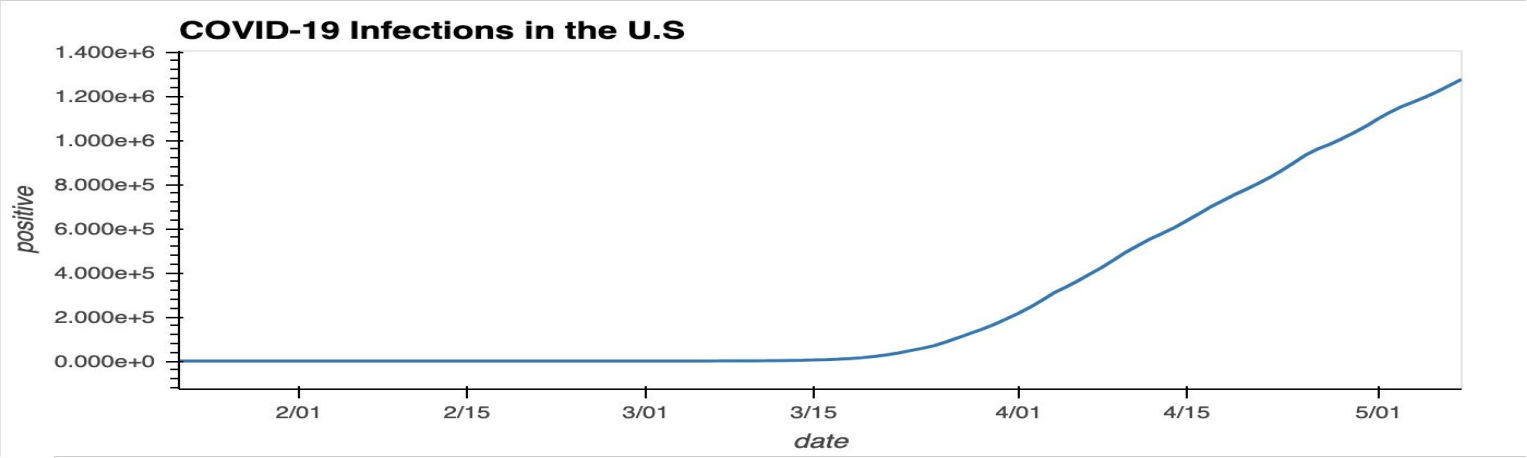
Data Analysis:

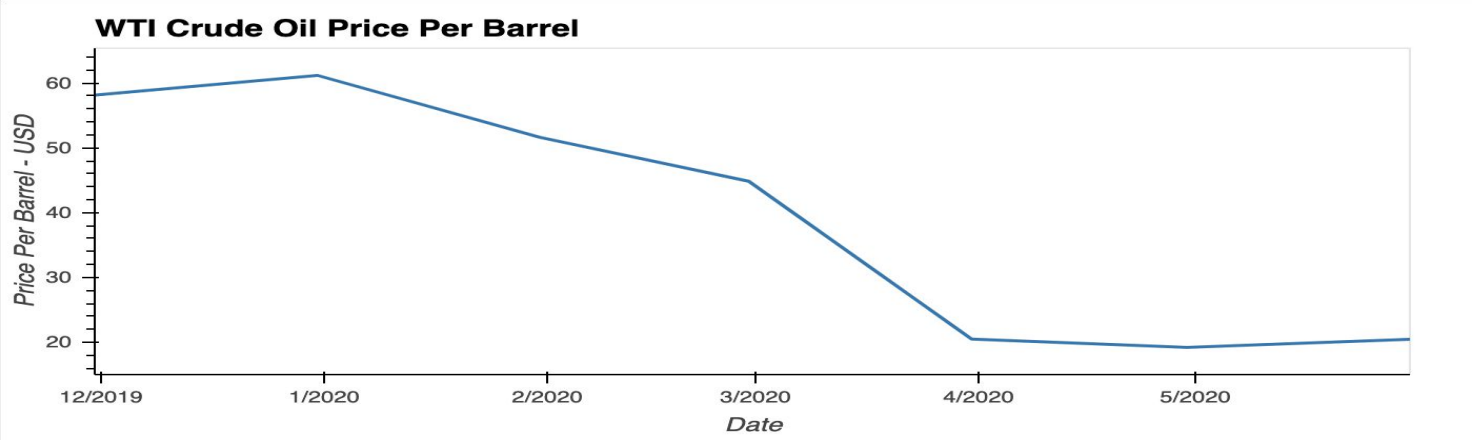
1. Covid Data
    - Downloaded csv file and saved in Github folder
    - Read in csv and set date and index
    - Grouped by month and plotted using hvplot

2. Macroeconomic Data
    - Did a pip install for Quandl API data
    - Imported quandl library in Python
    - Used quandl.get to call in datasets:
    - Crude oil (EIA/PET_RWTC_D)
    - Saudi oil production (EIA/STEO_COPR_SA_M)
    - Gold spot price (WGC/GOLD_DAILY_USD)
    - S&P 500 (MULTPL/SP500_REAL_PRICE_MONTH)

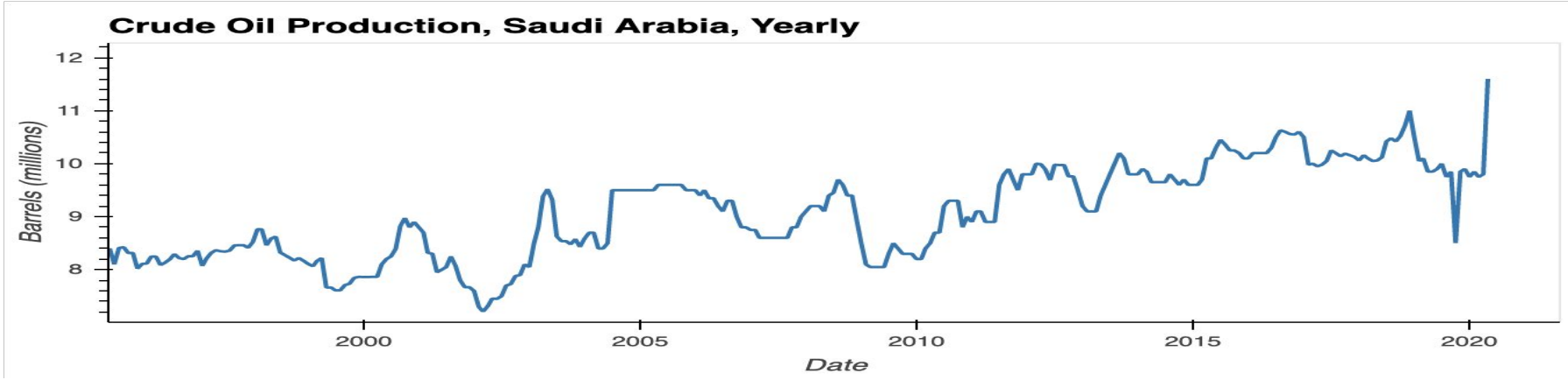
COVID-19 & Oil Prices

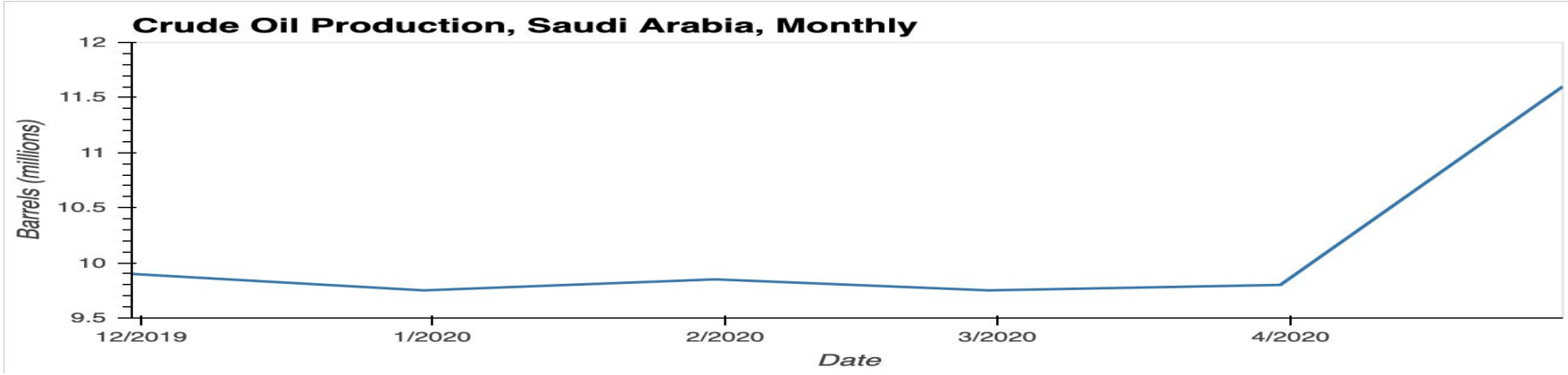
We conducted an investigation to determine whether the outbreak of COVID-19 caused the precipitous drop in WTI Crude Oil prices starting in March 2020. Our preliminary conclusion was that these events were correlated. But then we looked deeper….





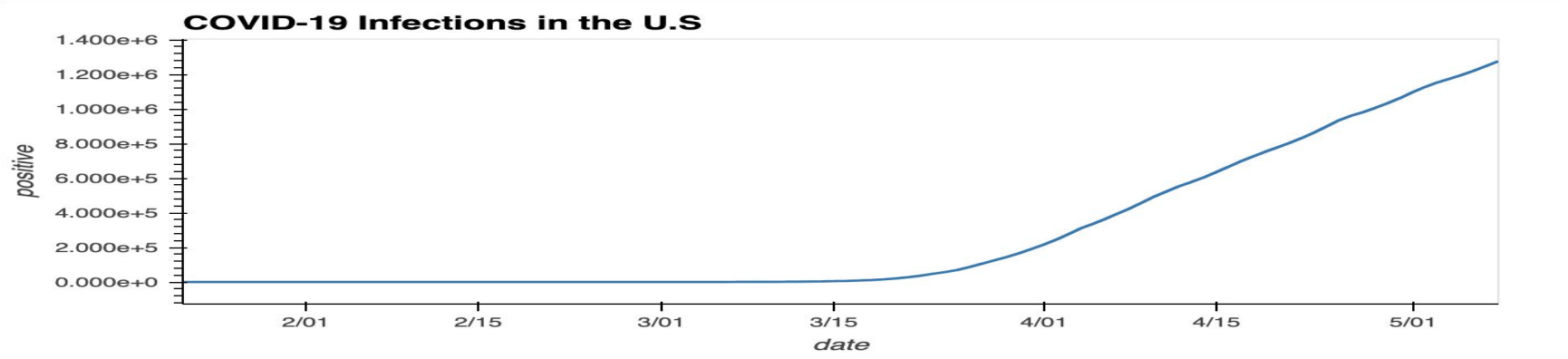
After the OPEC+ members failed to reach a production agreement in early March 2020, the Saudi government ordered Aramco to increase production to a record high of 13m barrels/day. According to CNBC, “The boost in exports comes against the backdrop of one of the most dramatic periods in oil market history: Record oil output from the world’s largest oil producers juxtaposed with eviscerated demand due to worldwide coronavirus lockdowns as economic activity and global commercial transport came to a screeching halt.”

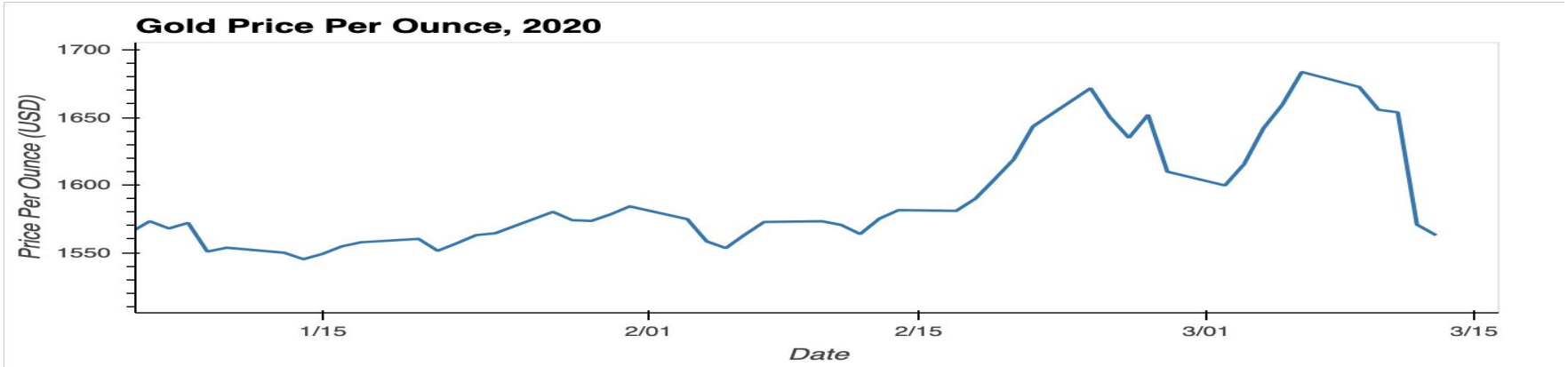




COVID-19 & Gold

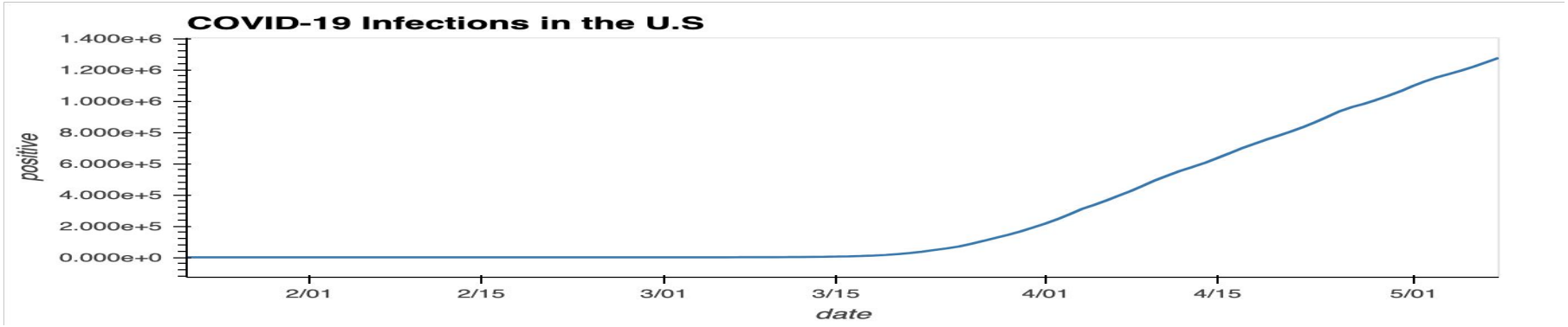
In early February, gold prices started going up as safe-haven buying increased. By mid-March, the ‘archetypal safe-haven asset’ saw a selloff as global markets realized COVID-19 would become a pandemic.

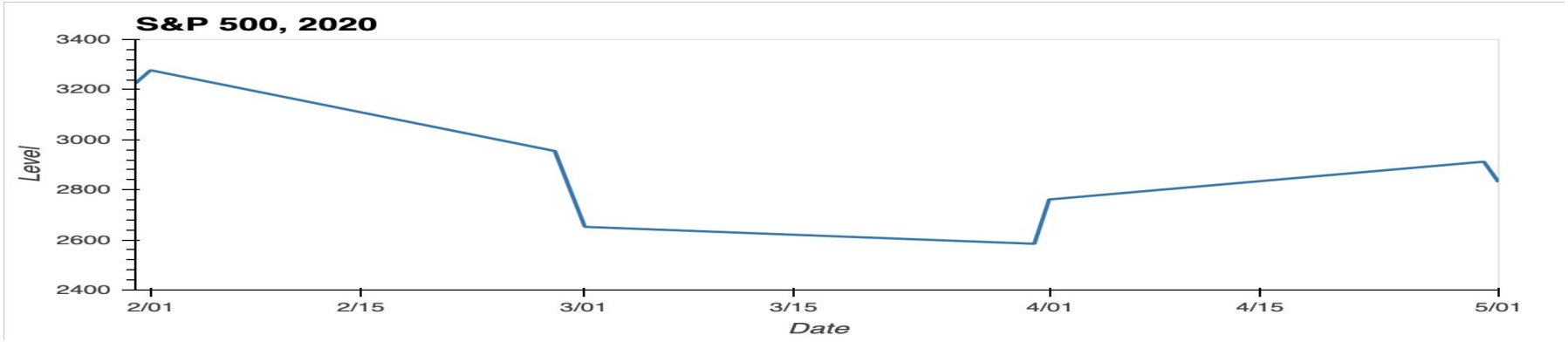




COVID-19 & S&P500

The S&P suffered its worst quarter since the Great Recession, wiping out $13 trillion from investors in just six weeks. In April, the S&P gained 23%, buoyed by economic stimulus programs launched by the federal government.





Quo vadis?

Past:

Challenges: Data from Energy Information Administration and Federal Reserve difficult to harmonize in Python given discrepancies in datetime and indexing.

Additional Metrics: Refine the commodities model by including additional supply and demand data to challenge conventional wisdom regarding asset class behaviour during downturns.

Present:

Future: Refine the model, utilizing anonymous cell phone position data to quantitate real extent of social distancing and feed that into the model as an additional parameter. The biggest hurdle was harmonizing data from multiple APIs. Also do to the length of time it took to generate models, it was hard to harmonize the date axis from different days, in a batch mode.

Thank You!

Questions?