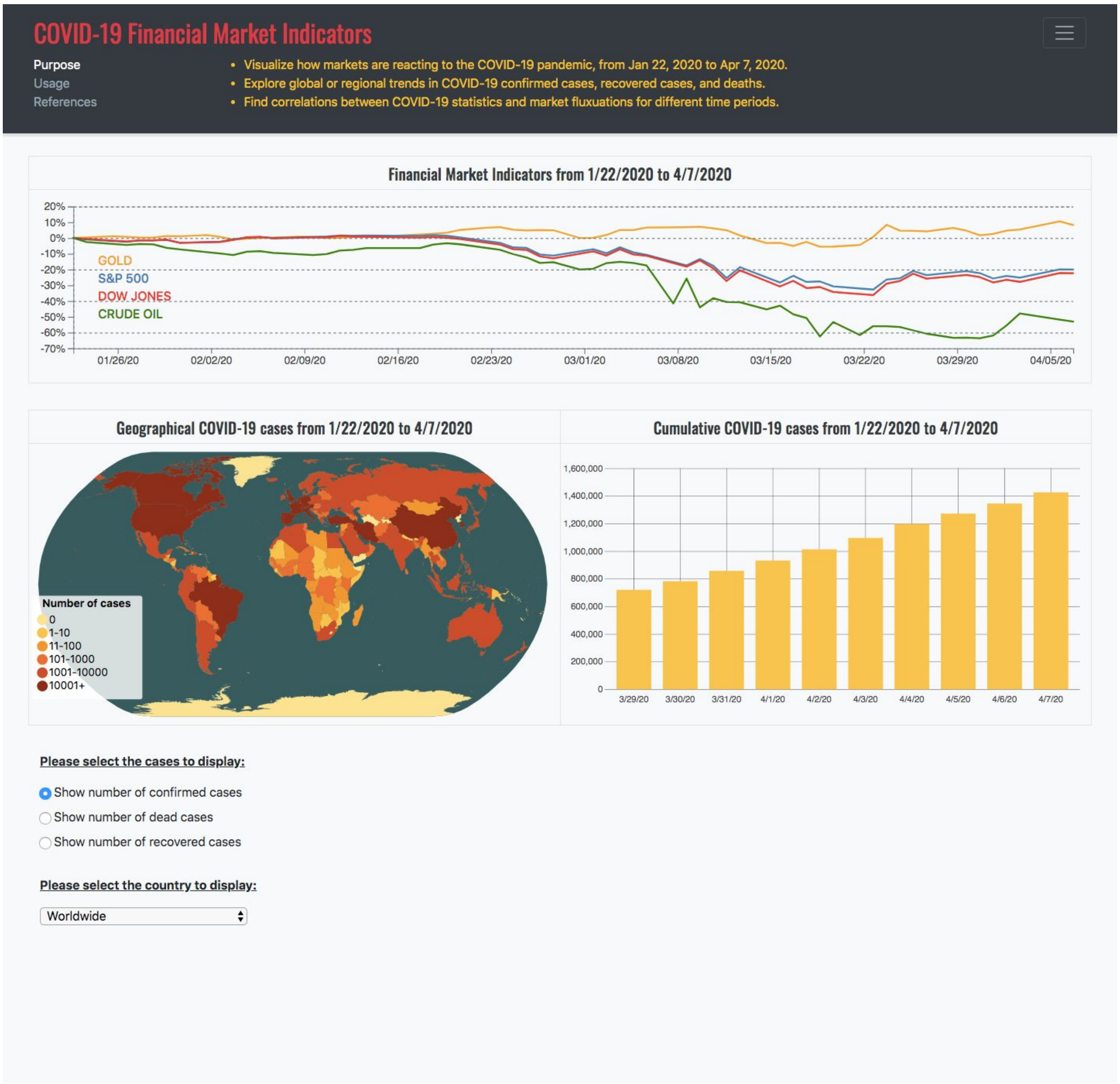
Project title: COVID-19 Financial Market Indicators
Team: Wise Donkey Mosquitoes
Team members: Alan Long (ala888), Brian White (bwrabbit), Jordan Chiu (u9y8)

Part 1) Overview



COVID-19 Financial Market Indicators

Purpose
Usage
References

- Visualize how markets are reacting to the COVID-19 pandemic, from Jan 22, 2020 to Apr 7, 2020.
- Explore global or regional trends in COVID-19 confirmed cases, recovered cases, and deaths.
- Find correlations between COVID-19 statistics and market fluxuations for different time periods.

Financial Market Indicators from 1/22/2020 to 4/7/2020

Geographical COVID-19 cases from 1/22/2020 to 4/7/2020

Cumulative COVID-19 cases from 1/22/2020 to 4/7/2020

Please select the cases to display:

● Show number of confirmed cases
○ Show number of dead cases
○ Show number of recovered cases

Please select the country to display:

Worldwide

Coronavirus disease (COVID-19) has spread rapidly to hundreds of countries around the world since its recent outbreak in China in late 2019. Being as the virus is novel, its effect on people is not well understood. The recent increase of cases arising in patients with no known exposure to infected individuals is an indication that community spread is occurring. This has led to widespread concern and has caused rapid, unpredictable changes in the US stock market.

To better understand the spread of COVID-19 and its effect on markets, our data visualization provides individuals with a synchronised view of geographical COVID-19 trends and fluctuations in various US stock market indicators. Our visualization consists of two main parts which in turn have multiple subcomponents:

1. Visualizing market fluctuations — Fluctuations in market indicators, including the S&P 500, Dow Jones Index, crude oil, and gold prices, are presented in an interactive multiple-line chart.

2. Visualizing spread of virus: — Geographical trends such as total cases per country, recovered patients, deaths are presented in an interactive choropleth map and an accompanying bar chart.

Part 2) Data

We utilized three datasets for our project:

1. Novel Corona Virus 2019 Dataset - Data on COVID-19 taken from the World Health Organization, split into a number of different CSV files:
    a. A primary dataset containing serial number, observation date and last update (ordinal, Jan 22, 2020 to Apr 7, 2020), Province/state (categorical, ~300 regions), Country/Region (categorical, ~200 countries), Confirmed, Deaths, and Recovered (quantitative, 0 to >100,000).
    b. Three time different time series files for confirmed, deaths, and recovered cases derived from the main dataset containing Province/State (categorical, ~300 regions), Country/Region (categorical, ~200 countries), Latitude (quantitative, -41.5 to 72), Longitude (quantitative, -158 to 178), and number of cases on a given date from Jan 22, 2020 to Apr 7, 2020 (quantitative, 0 to 22,000 for confirmed cases, 0 to ~700 for deaths, 0 to ~800 for recovered cases).
2. World Atlas TopoJSON - TopoJSON data used to render our choropleth map; countries rather than provinces were used.
3. Yahoo! Finance: S&P 500, Gold, Dow Jones, and Crude Oil consisting of Date (ordinal, Jan 22, 2019 to Apr 7, 2020), Opening, High, Low, Close, Adjusted Close amounts (quantitative, ~2000 to 3500), and Volume (~3 to 9 trillion).
   Note: Among these attributes, only the "Close" attribute was utilized to derive a new attribute showing percentage changes in these closing prices with respect to the first available date (quantitative, -65% to +15%).

We strived to keep our data processing pipeline simple and separate from our visualization. When each of our visualizations are constructed, we process different datasets for each visualization separately through the use of a separate dataset constructor for each chart.

Notably, the datasets used for COVID-19 data and for the topojson use different names for each country. For example, "United States of America" in the topojson is "US" in the COVID-19 data. To resolve this problem, we created a dictionary found in `js/map/map_dict` that maps country names between the two datasets. For the choropleth map, we additionally group the numbers for each country on a given day by using `d3.nest()` and `.rollup()` to allow for accurate case counts. Similarly, for the bar chart, we first clean our data and then group numbers by country instead of Province/State.

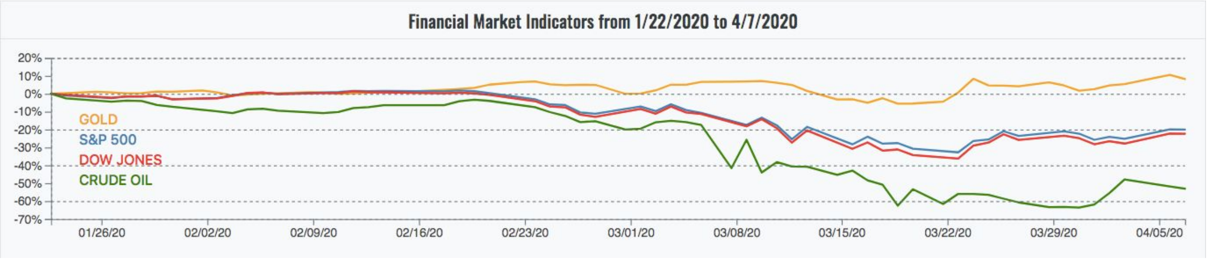
For the market indicator data, we chose to calculate percentage changes in each dataset's closing price with respect to its closing price on January 22, 2020. This was a necessary preprocessing step in order to normalize the datasets so we could use a single y-axis in the multiple-line chart and visualize the price changes on a common scale.

Finally, we had discussed the possibility of utilizing Crossfilter but ultimately decided against it, as it would add complexity to the project that would not significantly enhance our visualization to meet our goals.

Part 3) Goals and tasks

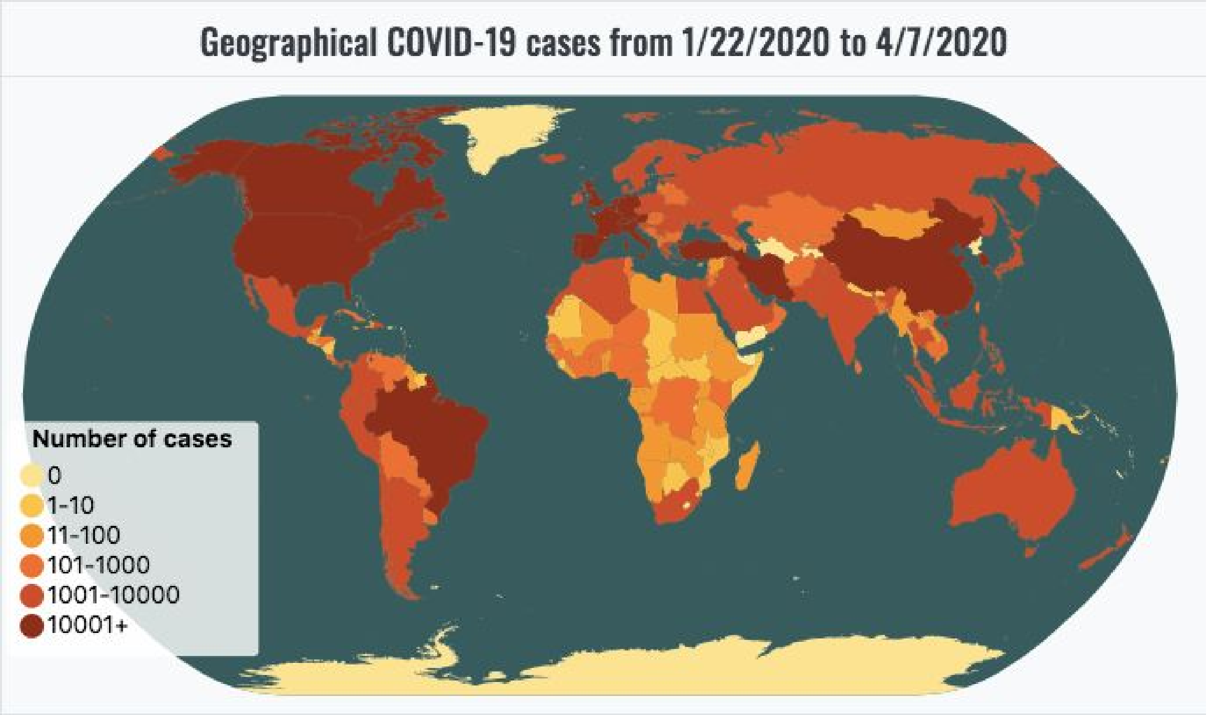
From the beginning, we knew that COVID-19 was going to be a popular topic for visualization, both in the context of this class and with sources in current events. Because of this, we wanted to produce a visualization that would approach the story from a different angle -- one in which we could visualize how markets are reacting to the pandemic and look for correlations between the spread of the virus and market activity. Additionally, we wanted to facilitate the more traditional tasks of tracking and comparing the rise and fall of market indicators on a daily basis, and allow users to explore and identify global trends of the spreading disease.

Part 4) Visualization



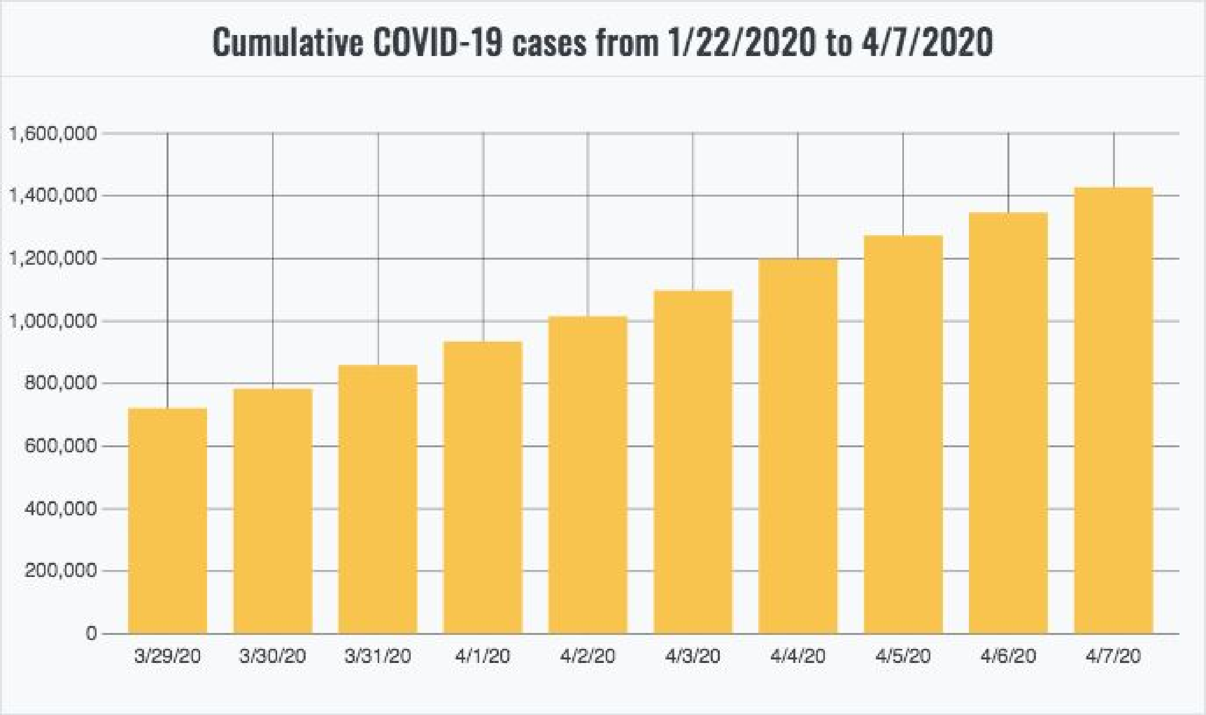
At the top of our visualization is a multiple-line chart, showing percent changes in various US market indicators from January 22 to April 7, 2020. Spanning across the full width of the page with a fixed scale on the x and y axes, this chart provides a holistic view of the story we aim to communicate and makes obvious to the user the extent of our supported date range. A one-dimensional d3 brush is placed on top of this chart to provide the user with an interactive medium for filtering the effective date range used to the COVID-19 statistics presented in the charts below.

When the page is loaded, a short animation is used to draw four lines from left to right via a d3 transition, emphasizing the chronological nature of the data and the downward trend of the market near the latter half of the period. The market indicators chosen to reflect the state or the US economy were simply chosen for their popularity and familiarity: S&P 500, Dow Jones index, crude oil, and gold. All values are shown in percentage increase / decrease from a 0% reference point on January 22, 2020. A categorical colour scale was chosen to easily distinguish between each of the market indicators displayed.



The next view we draw the user to is our choropleth map, which displays the number of confirmed cases in any given country, denoted by hue and luminance. We chose to use an ordinal colour scale where countries with more cases would be coloured more darkly, and countries with zero cases would be coloured more lightly. Depending on whether the user was choosing to visualize confirmed cases, recovered cases, or deaths, the base colour of the scale was changed to yellow, green, or red, respectively. These colour scales were chosen to be colorblind-safe, perceptually uniform, and monotonically increasing in luminance. We used an ordinal symlog rather than a continuous scale to ensure that multiple levels of severity in terms of infection would remain salient by the user. We chose to use a choropleth map rather than a cartogram to preserve the familiarity that users would have with world maps and to provide a low barrier of entry to our visualization.

A user can interact with this map in a number of different ways. First, when a user hovers over a country, detailed information about that country's case counts are shown in a tooltip with a responsive horizontal bar chart. Specifically, the number of confirmed cases, deaths, and recovered cases are shown in order to give users a fuller picture of that country's outbreak. A bar chart was chosen to help ensure users could accurately compare case counts within a particular country by aligning line marks along a common axis. The horizontal scale of the bar chart changes in range depending on which country is hovered over to make sure users are still able to perceive differences in data within a country with >50,000 cases and a country with <10 cases. The bars in the chart have three colours to allow users to distinguish between the colors: blue-green for recovered cases, red for deaths, and yellow for confirmed cases to match the colour scheme of our map. These colours were generated using Colorgorical with the goal of being distinct from the colours in our choropleth map while being distinct from each other.

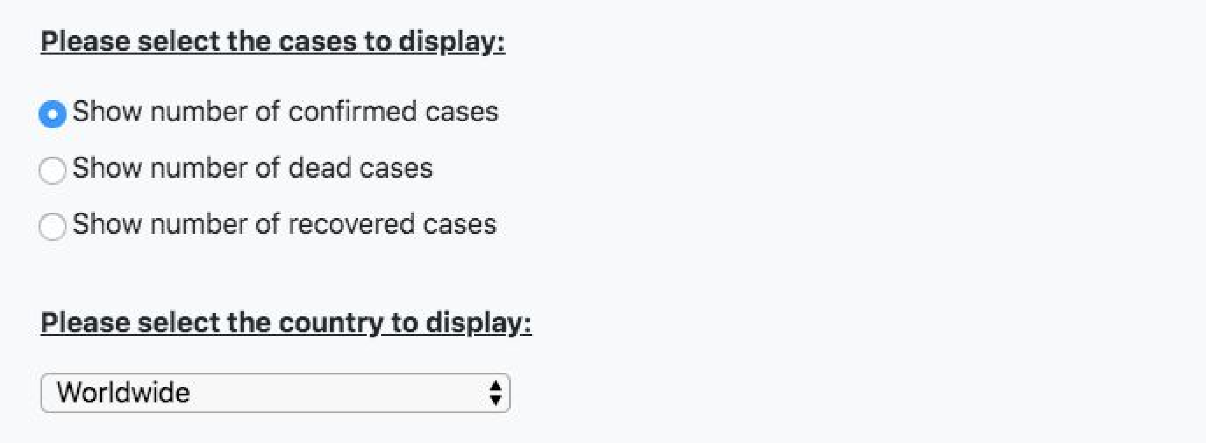


The final view in our project is this bar chart, which displays the cumulative number of cases worldwide or for a particular selected country. This chart was originally intended to be a grouped bar chart - allowing for the comparison between multiple countries. However, due to non-trivial issues between synchronizing multiple selections across different charts, the grouped bar chart was instead reduced to a standard bar chart. This bar chart displays the date on the x-axis and the number of cases on the y-axis. The scale on the y-axis is not static, and changes depending on the selected dates/country. This is because the number of cases varies wildly due to COVID-19's exponential growth. Attempting to maintain a static y-scale would've resulted in certain dates/countries early on in the pandemic being nearly indistinguishable from one and another.

When the user selects a certain date range via the stock chart at the top, this bar chart will adjust accordingly by either increasing or decreasing the number of bars displayed to accommodate the desired range up to a maximum of 10 days. When the user selects a different type of case to display (e.g. confirmed/dead/recovered), this bar chart will also change color to match the respective color schema of the aforementioned choropleth map.

Interactivity between views

We have several UI widgets: one to determine whether we are showing confirmed cases, deaths, or recoveries on the choropleth map and bar chart. The second is a dropdown selection of countries - this widget is described in greater detail below.



Please select the cases to display:

● Show number of confirmed cases

○ Show number of dead cases

○ Show number of recovered cases

Please select the country to display:

Worldwide

The map has a bidirectional link to the vertical bar chart through selecting countries. When a user clicks on a country on the map, that country is highlighted pink and its border becomes thick and dark to signify that it has been selected. Only one country can be selected at a time. Once a country is selected, its data is rendered in the vertical bar chart (countries can also be selected through the use of dropdown menus in the bar chart). If a selected country is clicked again (or if it is cleared from the dropdown), it is deselected and returns to its original colour determined by the ordinal scale described above.

The line chart is also linked unidirectionally to the map and the vertical bar chart. A user can select a date range on the chart by clicking and dragging a horizontal area on the line chart (called a brush). If a user has selected a date range using this brush, both the map and the bar chart respond by rendering only the data that is shown within that date range.

Part 5) Reflection

Describe how your project has developed from your initial proposal, through your first submission, to your final product

Our project has stayed largely true to our initial proposal with only a few minor changes. Instead of presenting pie chart glyphs for our "innovative view" for our choropleth map, we decided to show individual responsive barcharts as tooltips for two reasons. First, we wanted to present users with a more accurate rendering of the numerical data by using line marks on a common axis rather than radial area marks in a pie chart. Secondly, we felt that having pie chart glyphs displayed on every country would make for a cluttered and unfriendly graph. We also changed our market indicators graph from an area to a line chart, and our virus graph from a line chart to a vertical bar chart.

How have your visualization and technical goals changed?

Our visualization and technical goals did not change significantly from our project proposal.

How realistic was your original proposal in D3?

Our original proposal was largely achievable using D3. The more difficult technical challenges (ex. embedding a responsive bar chart inside a tooltip) were time-consuming but achievable with the technologies provided in D3. However, we did find it helpful to rely on some external technologies: our visual layouts were greatly improved with the addition of Bootstrap.

Was there anything you wanted to implement that you ultimately couldn't figure out how to do? If so, then what workarounds did you employ, or did you abandon your original idea?

At one point we discussed the possibility of using Crossfilter to allow for more complicated data manipulation (for example, allowing users to select the top three countries in terms of number of recovered cases within a given date range). However, we ultimately decided against it, since we felt that it would introduce unnecessary complexity to our project given our tight deadline. Instead, we decided to make smart use of `d3.nest()` and `.rollup()`, as well as standard javascript array manipulations to help mix and munge our data as needed.

If you were to make the project again from scratch (or any other interactive visualization), what would you do differently?

Time-permitting, it would have been incredibly valuable to create a draft visualization before deciding on a final direction for our project. Data visualization (and all software projects) are iterative by nature, and trying to create a fully fleshed-out proposal without first experimenting on our own was difficult. It was challenging to come up with a very detailed plan or schedule without having the knowledge of what was possible and what wasn't.

Now that we are more comfortable with d3, it would be interesting to take advantage of more complex data manipulation technologies like Crossfilter - there are many avenues that we were unable to explore in this project. Furthermore, our project works with a static dataset with data going up to March 11th, 2020. Accessing a dynamic dataset in real time would provide a more accurate and interesting visualization, as our numbers would reflect the most recent data in both stocks and COVID-19. However, this would come with its own challenges and our project would probably need to be updated on a daily basis (specifically to account for the discrepancies in country names).

There are a few minor details we were unable to address. Namely, when a user selects a date range by clicking and dragging the edges of the brush very quickly, we sometimes get strange graphical artifacts in our bar chart, where bars are not properly aligned or have odd widths. We suspect that this may have to do with the relatively intense processing that our charts must do in order to update themselves each time the brush is changed, as this bug is not consistently reproducible when selecting specific date ranges. Optimizing our data processing would likely resolve this issue, but would require more time.

Finally, it would be beneficial to rely on other back-end technologies through node. For example, our project's (bidirectional) interactivity relies on a single `state` object declared on the top level. This was a practical solution for this project, but more large-scale projects may require a state management tool to help facilitate interaction, particularly if asynchronous operations were to be involved. Using a back-end would also allow us to utilize database technologies rather than relying on static .csv files, allowing us to manage our data and our front-end code separately.

Part 6) Team Assessment

- Jordan: Set up scaffolding of project, implemented choropleth map and complex tooltips, created state object to facilitate interactivity, wrote first passes of written m2 and m3 writeups.

- Brian: Helped design initial scaffolding of project (implemented by Jordan), market indicator multiple-line chart implementation, responsive UI refactor with Bootstrap 4 including navbar and grid logic, and contributed to m2/m3 report contribution.

- Alan: Virus bar chart implementation, some dataset munging, some m2 and m3 writeup contribution. Overhauled the final visualization by standardizing colors and fonts. Also aligned page elements and chart axes to help improve aesthetics.