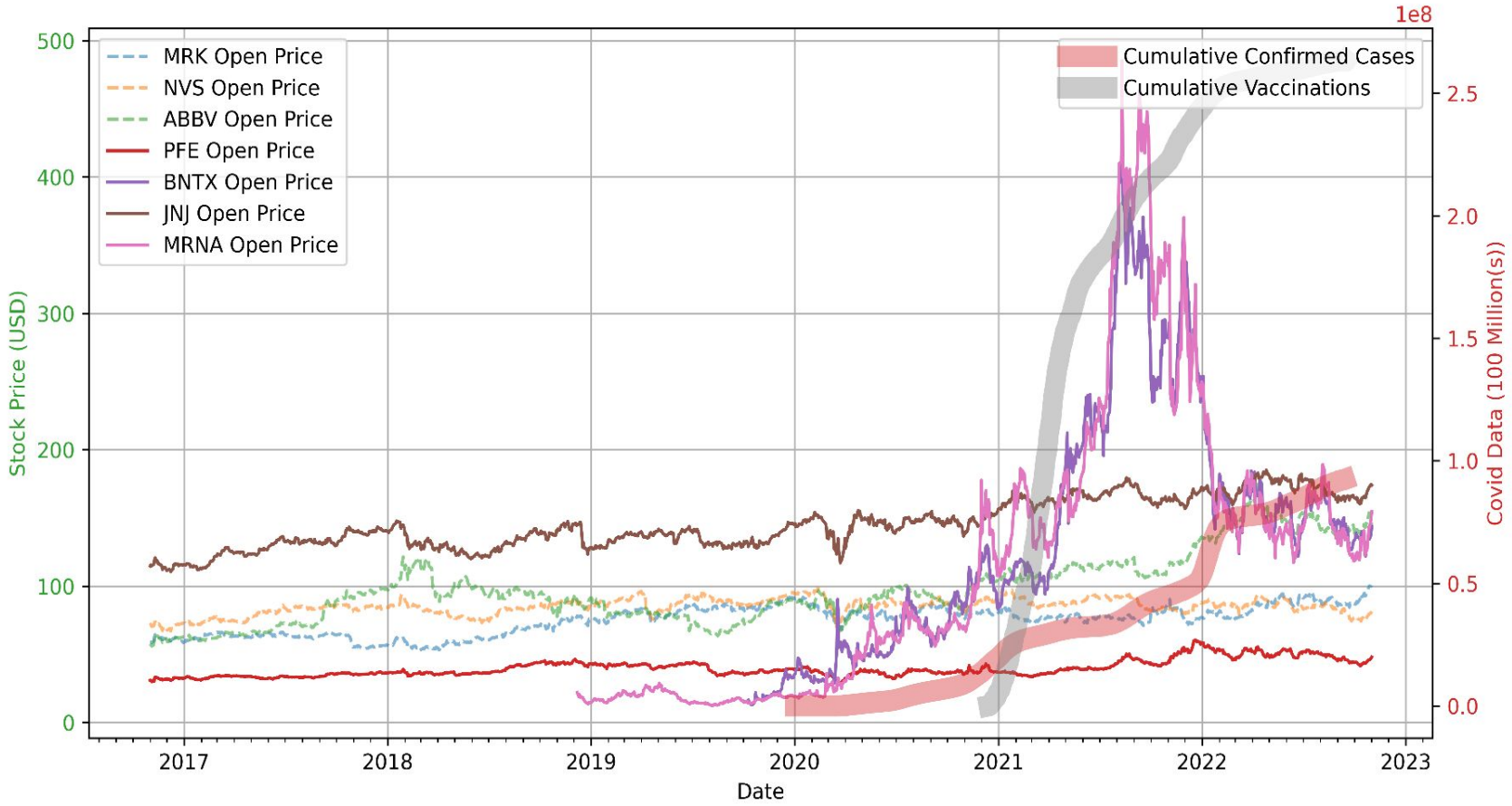
Pandemic VS Health Adjacent Stock Prices

Joseph Trybala,
Manas Bharadwaaj Subramanian,
Miko Wieczorek,
Ritu Kukreja

Introduction and Purpose of Dataset

- COVID19 Pandemic and Health Insurance & Pharmaceutical companies
- Equities traders and wealth managers who want to take advantage of stock price
- Acquire publicly available statistical data
- Process it in python to identify trends

Data Summary: Trend of Big Pharma Stock Prices and COVID-19 cases



label	date
First COVID-19 case	2019-12-31
First US COVID-19 case	2020-01-21
First EU COVID-19 case	2020-01-24
FDA EUA for hydroxychloroquine	2020-03-28
First Vaccine Dev Annoucement	2020-04-01
FDA EUA for remdesivir	2020-05-01
First Vaccine Trial	2020-07-27
FDA EUA for convalescent plasma	2020-08-23
FDA EUA for first vaccine	2020-12-11
FDA Booster approval	2021-08-12

Potential Users and Applications

- Pharmaceutical, health care and insurance companies
- Potential investors
- Correlations between public health and the company valuation

Source of Data

- MarketStack.com API
- Google Covid-19 Open Data, Various sub-sources

Approach to Data Acquisition

- Public Health APIs such as CDC
  - Google Covid-19 Open Data
    - Data we used came from CDC and Covid Tracking Project

- Various Financial APIs
  - MarketStack API
    - Simple JSON output from API request
    - Date, Symbol were the necessary specified metrics
    - Using multiple get requests to the market stack that were constrained by day limitations.
    - JSON to csv > csv saved > csv to dataframe > stitch together dataframes of differing times and stock symbols.

Acquisition - Covid Data - Epidemiology

date	location_key	new_confirmed	new_deceased	new_recovered	new_tested	cumulative_confirmed	cumulative_deceased	cumulative_recovered	cumulativ
7/19/2020	BR_AL_270360	0	0			111	3		
7/20/2020	BR_AL_270360	0	0	0	2	111	3	0	226
7/21/2020	BR_AL_270360	0	0	0	1	111	3	0	227
7/22/2020	BR_AL_270360	1	0	0	8	112	3	0	235
7/23/2020	BR_AL_270360	4	0			116	3		
7/24/2020	BR_AL_270360	0	1			116	4		
7/25/2020	BR_AL_270360	0	0	0	2	116	4	0	237
7/26/2020	BR_AL_270360	0	0			116	4		
7/27/2020	BR_AL_270360	2	0	0	3	118	4	0	240
7/28/2020	BR_AL_270360	0	0	0	1	118	4	0	241
7/29/2020	BR_AL_270360	0	0			118	4		
7/30/2020	BR_AL_270360	0	0	0	4	118	4	0	245
7/31/2020	BR_AL_270360	6	0			124	4		
8/1/2020	BR_AL_270360	0	0	0	3	124	4	0	248
8/2/2020	BR_AL_270360	0	0			124	4		
8/3/2020	BR_AL_270360	0	0	0	10	124	4	0	258
8/4/2020	BR_AL_270360	0	0	0	1	124	4	0	259
8/5/2020	BR_AL_270360	0	0			124	4		
8/6/2020	BR_AL_270360	3	0	0	3	127	4	0	262

Acquisition - Covid Data - Vaccinations

date	location_key	new_persons_vaccinated	cumulative_persons_vaccinated	new_persons_fully_vaccinated	cumulative_persons_fully_vaccinated	new_vaccine_doses_administered	cumulative_vaccine_doses_administered
5/9/2021	PL_24_16	707	25523	429	7433	1136	32956
5/10/2021	PL_24_16	0	25525	153	7587	153	33112
5/11/2021	PL_24_16	68	25788	8	7626	76	33414
5/12/2021	PL_24_16	621	26544	44	7674	665	34218
5/13/2021	PL_24_16	839	27655	410	8087	1249	35742
5/14/2021	PL_24_16	491	28349	800	8959	1291	37308
5/15/2021	PL_24_16	338	28821	614	9655	952	38476
5/16/2021	PL_24_16	343	29231	527	10211	870	39442
5/17/2021	PL_24_16	87	29452	66	10336	153	39788
5/18/2021	PL_24_16	100	29765	38	10479	138	40244
5/19/2021	PL_24_16	586	30580	167	10755	753	41335
5/20/2021	PL_24_16	1102	31785	309	11071	1411	42856
5/21/2021	PL_24_16	819	32646	559	11668	1378	44314
5/22/2021	PL_24_16	639	33339	331	12000	970	45339
5/23/2021	PL_24_16	367	33801	504	12608	871	46409
5/24/2021	PL_24_16	0	33903	51	12773	51	46676

Approach to Data Acquisition

- Public Health APIs such as CDC
  - Google Covid-19 Open Data
    - Data we used came from CDC and Covid Tracking Project

- Various Financial APIs
  - MarketStack API
    - Simple JSON output from API request
    - Date, Symbol were the necessary specified metrics
    - Using multiple get requests to the market stack that were constrained by day limitations.
    - JSON to csv > csv saved > csv to dataframe > stitch together dataframes of differing times and stock symbols

Acquisition - Stock Data

open	high	low	close	volume
212.57	213.36	211.79	212.7	11862705
209.62	212	208.72	212	10341300
208.94	211.03	208.93	210.1	7545901
204.06	205.84	203.7	205.36	910863
204.81	206.33	204.12	204.3	3893474
207.13	207.22	204	204	5670493
209.29	209.82	207.1	207.49	3387023
208.95	209.98	208.91	209.01	2722997
207.28	209.85	206.65	209.36	3287544
209.89	210.9	207.68	208.05	2568920
209.91	210.67	209	209.52	1755091
209.16	210.61	209.02	209.33	1526797
209.42	210.1159	208.56	209.36	1980801
208.62	209.7	207.871	209.23	3498370
207.91	209.75	207.71	208.84	3392595
202.62	207.13	202.15	207.05	4648657
198.68	200.3	198.68	199.33	1111154
200.97	201.54	197.81	198.21	1853113
199.02	200.66	198.42	200.14	1820597

AET	XNYS	2016-11-08T00:00:0
AET	XNYS	2016-11-07T00:00:0
AET	XNYS	2016-11-04T00:00:0
AET	XNYS	2016-11-03T00:00:0
AET	XNYS	2016-11-02T00:00:0
AET	XNYS	2016-11-01T00:00:0
MET	XNYS	2022-11-01T00:00:0
MET	XNYS	2022-10-31T00:00:0
MET	XNYS	2022-10-28T00:00:0
MET	XNYS	2022-10-27T00:00:0
MET	XNYS	2022-10-26T00:00:0
MET	XNYS	2022-10-25T00:00:0

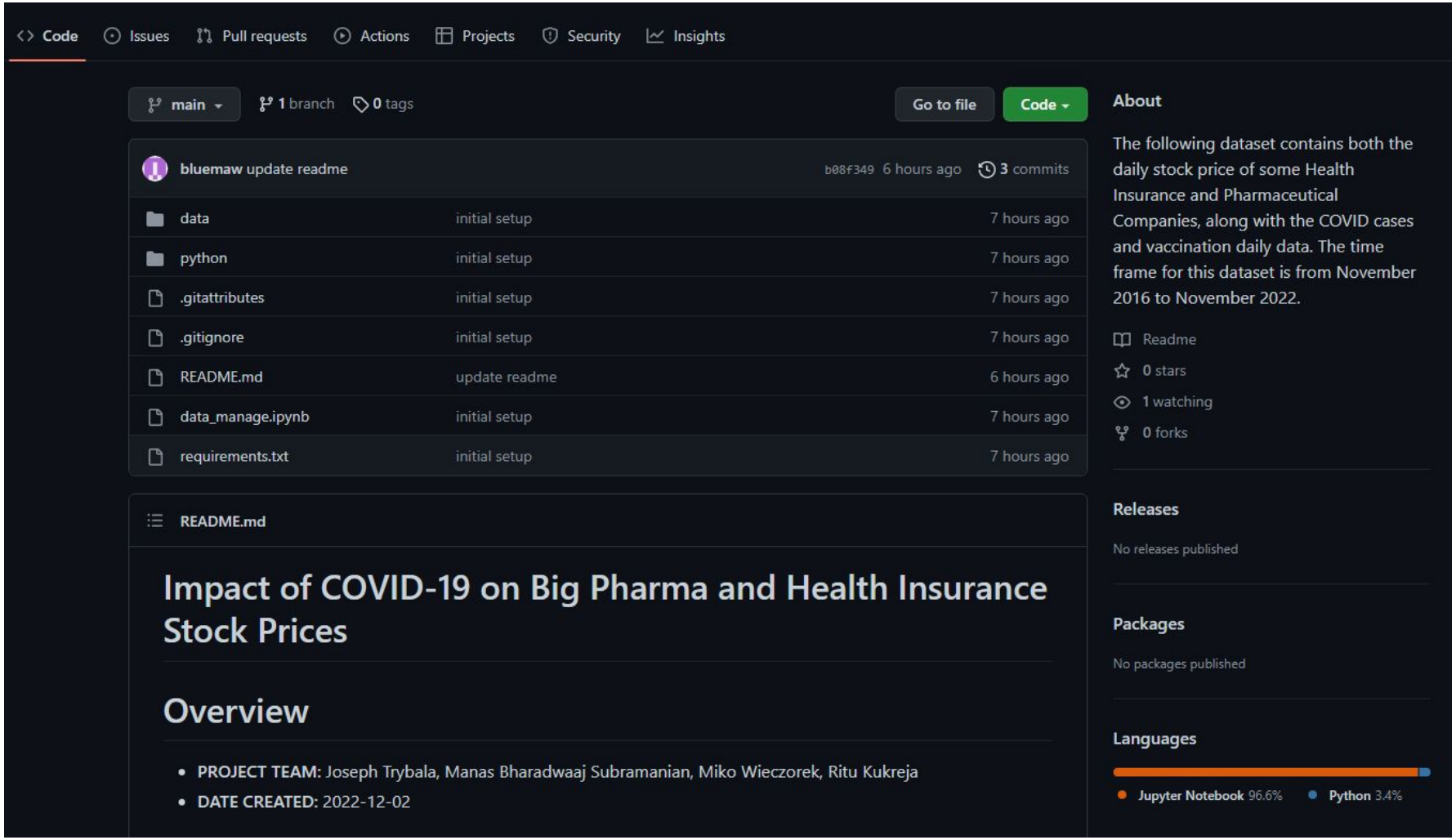
Approach to Preprocessing Data

- Cutting down Covid data after identification of interesting metrics. Specifically “cumulative” over “new”
  - US Data Only
  - Only data that was well represented without gaps
- Selecting, cutting down, and organizing stock data. Selection was an upfront and ongoing process
  - MarketStack provided their own adjusted metrics which, while valuable, were ultimately not utilized

Distribution Approach

- Distributed as .csv for ease of use
- The dataset will undergo further modifications and updated versions will be released
- Code for acquiring, slicing, and organizing may also be released
- We released the dataset on GitHub: github.com/bluemaw/covid-stock-data

Distribution Approach - Github



Discussion of Access Rights

- The dataset can be accessed via the following github.com/bluemaw/covid-stock-data
- The Stock data is acquired from an free online service (MarketStack.com) that provides an ability to extract data using an API request
- The COVID-19 dataset is acquired from Google Covid-19 Open Data, COVID Tracking Project, and the CDC
- The final merged dataset is released publicly under a Creative Commons 4.0 Attribution (CC BY 4.0) license for interested parties

Issues and Limitations

- Our stock data is limited to the United States stock exchange data
- The issue with Covid Data is that reporting is not uniform across all parties involved
- The API we are utilizing for stock data acquisition has a limit for API calls, which limits our data width

Team and Contributions

- Joseph Trybala
  - Topic Identification, Data Acquisition Coding, Data Processing Coding, Editing, Writing
- Manas Bharadwaaj Subramanian
  - Topic identification, Preprocessing Data, Structuring Notebook, Data Visualization, Markdown writing
- Miko Wieczorek
  - Modularizing Code, Structuring Repository, GitHub Setup, Project Reproducibility with virtualenv, and Healthcare Domain Expert
- Ritu Kukreja
  - Topic Identification, Preprocessing Data, Data Visualization, Editing, Writing