How the COVID-19 pandemic has affected both air quality and single use plastic waste

DS4A: Women's Summit 2020 - Team #21

Laura Collett, Mirella Koleva, Chiara Sotis, Bailey Wellen

[Github Repository]

Contents

Introduction

The COVID-19 pandemic has had severe repercussions not only on how people lead their day-to-day life, but also on the environment. On the one hand, newspapers and the media stress the reduction in air pollution and how the pandemic benefitted urban areas, typically surrounded by clouds of smoke. On the other hand, the restrictions imposed and the days in lockdown increased the consumption of plastic exponentially: people order more groceries and deliveries and, mostly, make large use of disposable plastic-based personal protection equipment (PPE).

The UN Environment [1] lists greenhouse gases (GHGs, and in particular CO_2) and plastic pollution as two of the main environmental challenges we face as humanity. In this project we seek to understand the impact of the recent pandemic on these two aspects of environmental quality and to assess the truthfulness of media reporting around them. In fact, while several newspapers and sectorial magazines (e.g. the Economist [2], NewScientist [3], Wired [4] and the BBC [5], to name a few) have commented on the environmental consequences of the pandemic, many of these articles have failed to specify the data used to make their inference and/or the underlying assumptions. We hope to shed light on these two key environmental issues and quantify the impact that COVID-19 has had on them.

To do so, we analyze the trends before and after the beginning of the pandemic to identify whether there was a significant gain in air quality and the magnitude of this new amount of plastic. To address these questions we start by using exploratory data analysis (EDA), then study the issues through geospatial analysis, time-series analysis, and significance testing. We constrain our analysis geographically to the United States, so any inferences found can only be applied to this country.

Aims and objectives

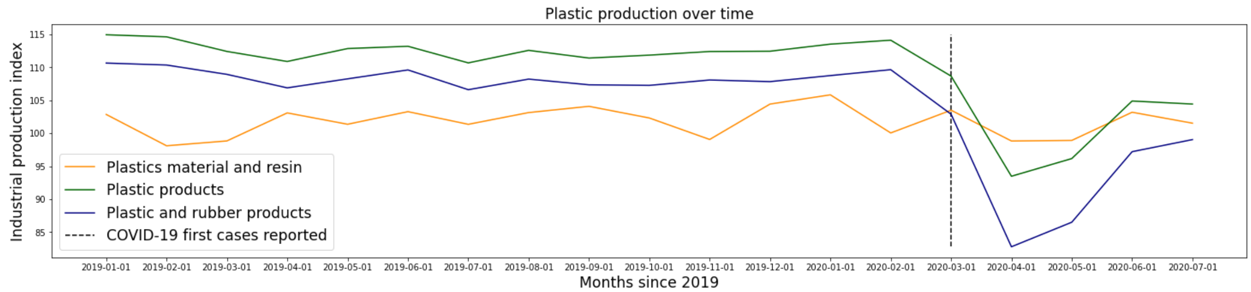
Our specific aims were as follows:

1) Identify the trends in plastic production in the United States over the last 2 years and whether there has been an increase in the production due to the pandemic
2) Identify the increase in plastic due to the higher need for PPE
3) Map air pollution and air quality in the United States in the last 2 years and assessing whether there has been an increase in air quality since the pandemic
4) Identify whether the increases in air quality imply lower CO_2 concentrations (i.e. contribute to a long-term reduction in pollution)

Results

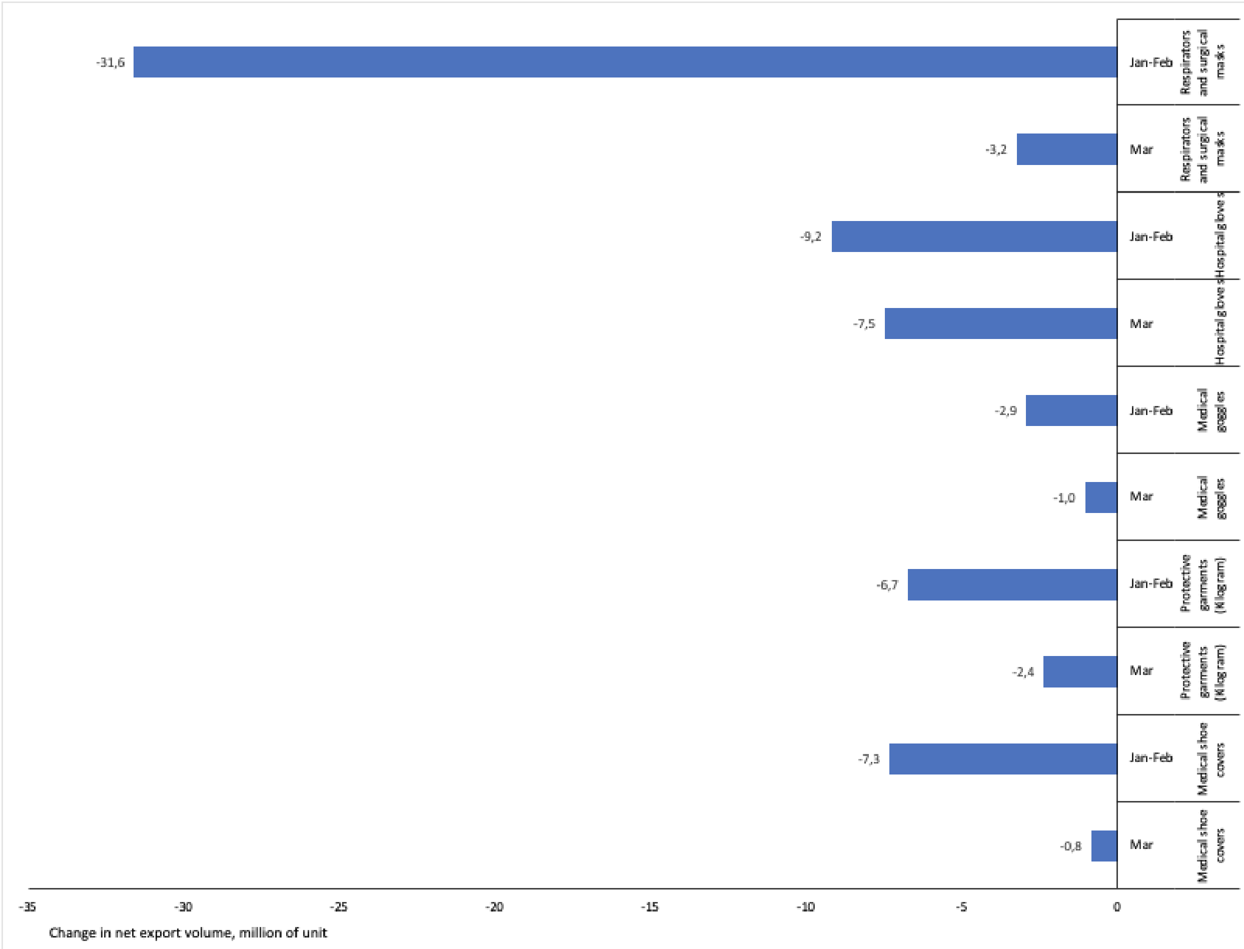
Initial reduction in plastic production

To investigate changes in plastic production over the past two years, and specifically since the pandemic, we plotted this as a time series for three plastic types. From this we see roughly constant plastic production since January 2019, with a significant drop seen just after COVID hit that seems to rise again a few months following this drop. A drop may have been observed due to lockdown and businesses stopping production, reducing the demand for plastic production, but this is in contrary to our hypothesis that there may have been an increase due to the pandemic. However, this type of drop in manufacturing is also a characteristic feature of periods of recession and the economic downturn in the US might help explain it. The rise may then have been observed due to the subsequent opening up of businesses again and the added demands of PPE and extra hygienic packaging, as hypothesised at the outset. In short, many factors might explain the pattern we have observed, but the pattern is not as hypothesised and portrayed strongly in the media.



In order to see if this drop after COVID was statistically significant across the three plastic types, 2019 and 2020 were compared using the differences between the levels of plastic production in the different types of plastics for every month from April to July. Levels in 2019 were compared with levels in 2020 and a difference was calculated. These differences across months and plastics types seemed to be normally distributed so a paired t-test was carried out on these data to test the null hypothesis of no difference between this calculated difference and zero (a null hypothesis of no change between the years). This test gave a test statistic of -4.21 and a p-value of 0.001, again showing strong evidence at the 5% level of a reduction in plastic production levels from the spring of 2019 to the spring of 2020.

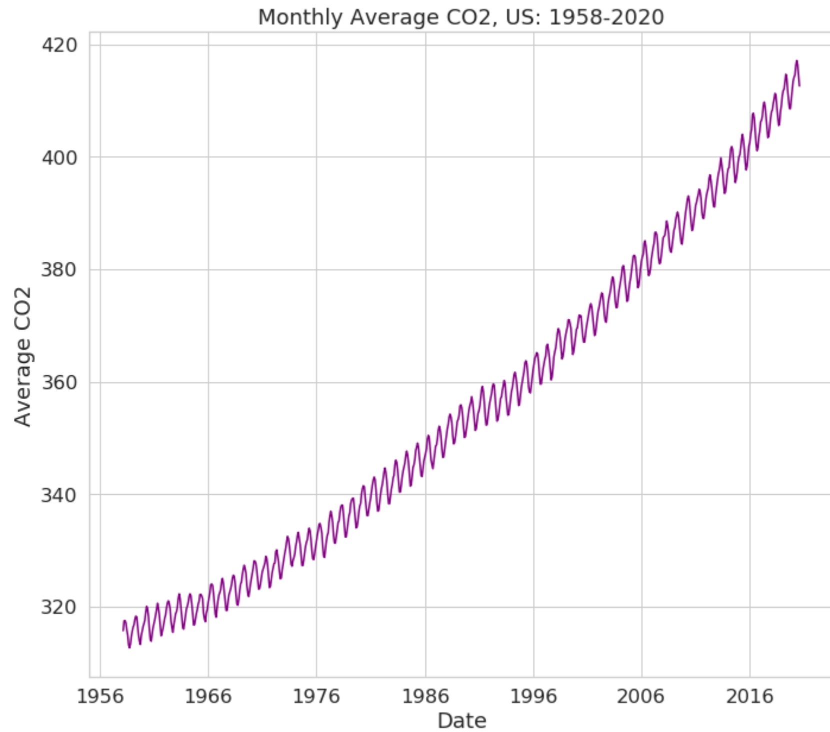
Furthermore, to assess whether the reduction in manufacturing corresponded to a reduction in plastic consumption (and hence waste) we use the data provided by the UN Comtrade accessed via World Integrated Trade Solution (WITS). This allows us to look at United States exports and imports of PPE and compare it to their exports in 2019. We observe negative net exports for the period ranging from January to March 2020 for all PPE items listed in the dataset.



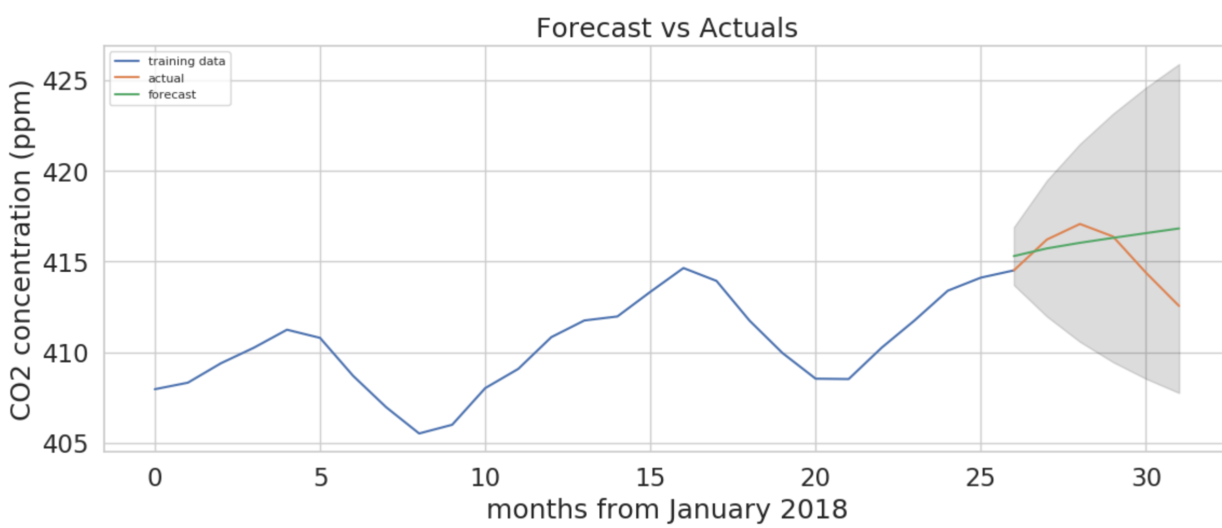
From the negative net exports, we know that even if the initial plastic production has decreased, the huge quantities of PPE being imported will eventually end up in the U.S. landfills. This indicates no real long-term benefits to the environment.

No change in persistent pollutants

A long-lasting improvement in air quality should reduce the CO2 concentrations in the atmosphere, as many of the activities leading to air pollution (e.g. travel) lead to high emissions. In order to detect whether there has been any change in CO2 over time and since the COVID-19 pandemic, we plotted a time series of the CO2 levels and we observed a general linear increase in CO2 in the atmosphere, with some seasonality.



To formally analyze these trends we first look at an ARIMA model to forecast the last 6 months (from March 2020) of CO2 concentrations from past data. We train the ARIMA(1,1,1) model with data from January 2008 to accurately forecast the concentration of CO2 from the beginning of the pandemic. We test whether the values observed are statistically lower from the ones predicted by the model.

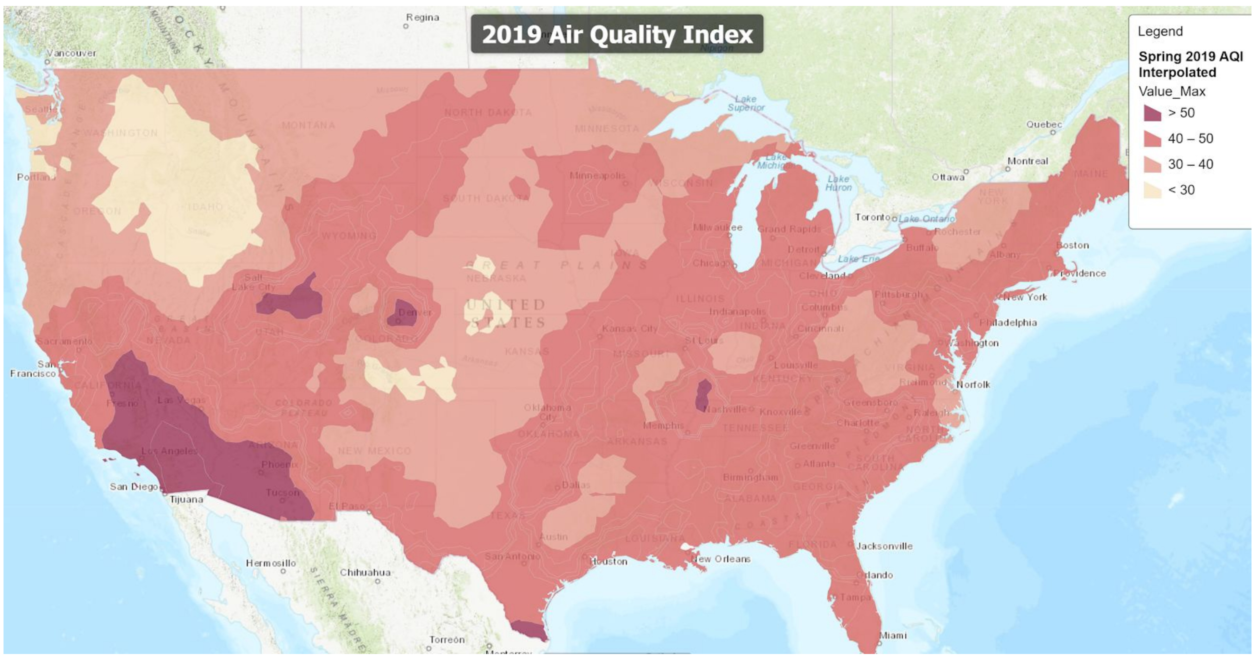


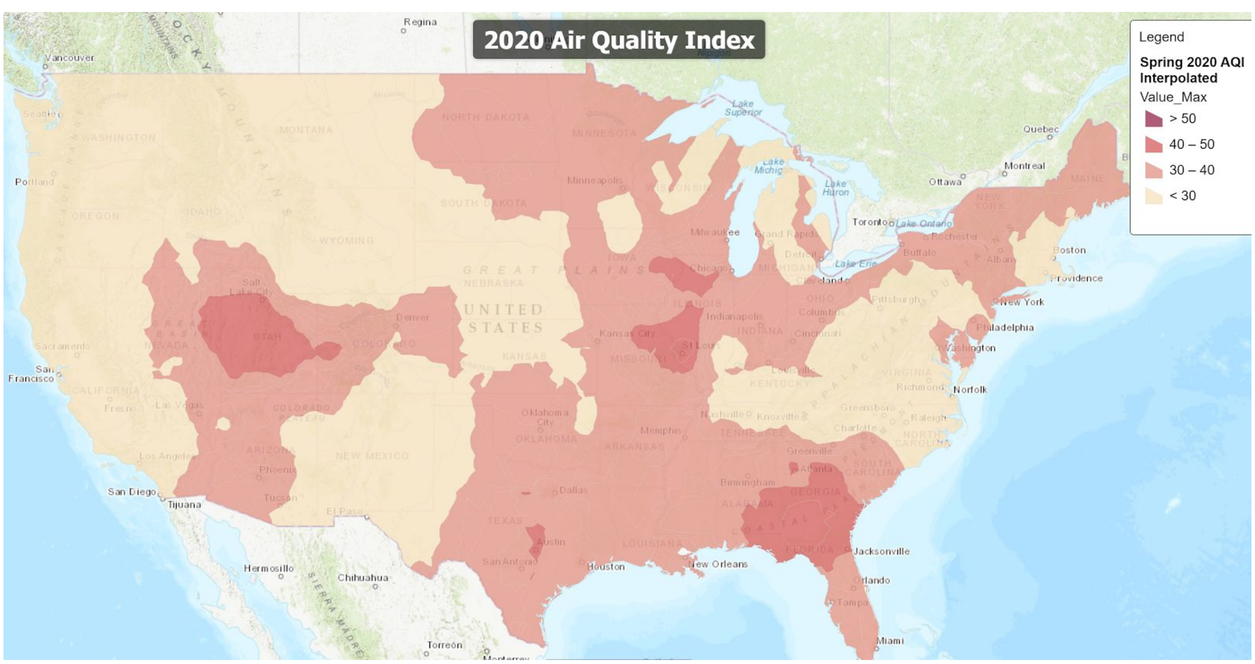
We observe that the realized concentrations of CO2 fall within the 95% confidence interval (the grey area on the figure) of the values predicted by the ARIMA(1,1,1) model, and hence there is no evidence that the pandemic has caused a statistically significant reduction in the main gas responsible for climate change.

Short-term air quality improvements

While we don't observe a significant decrease in CO_2 emissions, we do have evidence that air quality in regards to short-term pollutants has improved drastically from Spring 2019 to Spring 2020. We quantify this by using the US EPA's Air Quality Index, which considers ground-level ozone, particle pollution (both PM2.5 and PM10), carbon monoxide, sulfur dioxide, and nitrogen dioxide.

The data behind the maps below is the mean AQI by county for March to May of 2019 and 2020. Then, we used the individual testing site averages and ArcGISOnline's interpolation tool [14] to display approximate air quality beyond state and county borders.





As you can see above, the spatial data displays a stark difference between spring 2019 and spring 2020, and our statistical analysis backs up this assertion. Analyzing only the testing sites that have observations in both spring 2019 and spring 2020, we find the mean AQI in 2019 to be 40.975 and the mean AQI in 2020 to be 30.333. Performing a one sample t-test on the difference between these means (assuming that these two datasets are independent), we find a p-value of < 0.001, which is strong evidence at the 5% significance level that the mean AQI in spring 2019 differs from that in spring 2020.

Conclusions and Future Work

Overall we have seen that there have been a variety of changes (some statistically significant) in plastic production, air quality, CO_2 emissions and transport use that may be attributable to the COVID-19 pandemic.

In terms of plastic production, we have shown by way of two different statistical methods that plastic production has experienced a significant reduction since the pandemic first afflicted the US. Conversely, plastic imports in the form of PPE have risen. However, further data are required in order to determine whether production will return to previous levels quickly. Meanwhile, the current short term trend is upwards. Future work may be needed to better understand whether the fall in plastic production is due to the inclement economic conditions created by the COVID-19 pandemic or to changes in lifestyle invoked by the crisis.

Looking at air quality, it appears that the COVID-19 pandemic has had a significant beneficial effect on air pollution in the United States. Comparisons between spring this year and spring last year show that the average air pollution level in the country has fallen from near the top of the range considered as low pollution to the middle. Areas in the southwest US have enjoyed particularly large improvements in air quality, and it remains to be seen whether this will translate into better respiratory health for the affected populations.

After analysing rental bike ridership, our preliminary analyses have shown that it would be worthwhile to investigate the effect of weather on rental bike ridership and air pollution. In the case of the latter, it could have a significant effect on the measured levels of particulate matter, since rain tends to bring these particulates down to the ground and wind redistributes their concentrations over large areas that can span hundreds of miles [13]. In addition, finding data on personal bike riding patterns would be advantageous to obtain a complete picture of bike use and enable definitively answering the question of whether people have turned away from motor vehicles and embraced a more green means of transport since the start of the pandemic.

In terms of CO_2 levels, overall, our forecasting analysis determined that the pandemic has not diminished the rate at which CO_2 is output into the atmosphere, or at least not in a statistically significant way. The bold claims made by the media on the positive effect of the pandemic on the environment are likely to be overestimating the short term effects on local air pollution and discarding the issue of long-term determinants of climate change.

In general, we have seen that those factors that may react quickly to changes in human behaviour, for example plastic production and air quality, have been reduced due to the pandemic, and society shutting down temporarily, but this hasn't impacted the more pressing and endemic problem of CO_2 emissions in the atmosphere. It is a positive sign that human behaviour can shift some factors, but more lasting change of behaviour is required if CO_2 emissions are to not continue to increase and bring about further global warming.

References

[1] United Nations Environment Programme 2019 Annual Report, https://www.unenvironment.org/annualreport/2019/index.php

[2] The Economist, 'COVID-19 has led to a pandemic of plastic pollution', published on 22 June 2020, https://www.economist.com/international/2020/06/22/covid-19-has-led-to-a-pandemic-of-plastic-pollution?gclsrc=aw.ds&gclid=CjwKCAjw2Jb7BRBHEiwAXTR4jd2yhNjS2v129IbiPiF_KBbP-ob3riLCEASdewJ9ONFcDOeGwMkpJRoCN_wQAvD_BwE&gclsrc=aw.ds

[3] New Scientist, 'How COVID-19 spawned a plastic pandemic - and what we can do about it', published on 7 August 2020, https://www.newscientist.com/article/2251045-how-covid-19-spawned-a-plastic-pandemic-and-what-we-can-do-about-it/

[4] Wired, 'The coronavirus pandemic has totally derailed the war on plastic', published on 30 June 2020, https://www.wired.co.uk/article/coronavirus-plastic-pollution-environment

[5] BBC, 'Coronavirus: Pandemic causing new wave of plastic pollution', published on 31 August 2020, https://www.bbc.co.uk/news/uk-england-bristol-53947161

[6] Federal Bank of St. Louis, https://fred.stlouisfed.org/searchresults/?st=plastic&t=plastics%3Bmonthly%3Btrade&rt=trade&ob=sr

[7] U.S. Energy Information Administration, https://www.eia.gov/environment/emissions/state/

[8] NASA Global Climate Change Research, https://climate.nasa.gov/vital-signs/carbon-dioxide/

[9] U.S. Bureau of Transportation Statistics, https://www.bts.gov

[10] U.S. Office of Highway Policy Information, https://www.fhwa.dot.gov/policyinformation/

[11] U.S. Environmental Protection Agency, https://www.epa.gov/outdoor-air-quality-data

[12] USA Facts, https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/

[13] Official website for Pima County, Arizona, accessed on 8 October 2020, https://webcms.pima.gov/UserFiles/Servers/Server_6/File/Government/Environmental%20Quality/Air/Air%20Monitoring/AWhatisParticulateMatter1.pdf

[14] ArcGIS Interpolate Points Feature Documentation, https://doc.arcgis.com/en/arcgis-online/analyze/interpolate-points.htm

[15] Harvard School of Public Health, https://www.hsph.harvard.edu/news/hsph-in-the-news/air-pollution-linked-with-higher-covid-19-death-rates/

Acknowledgments

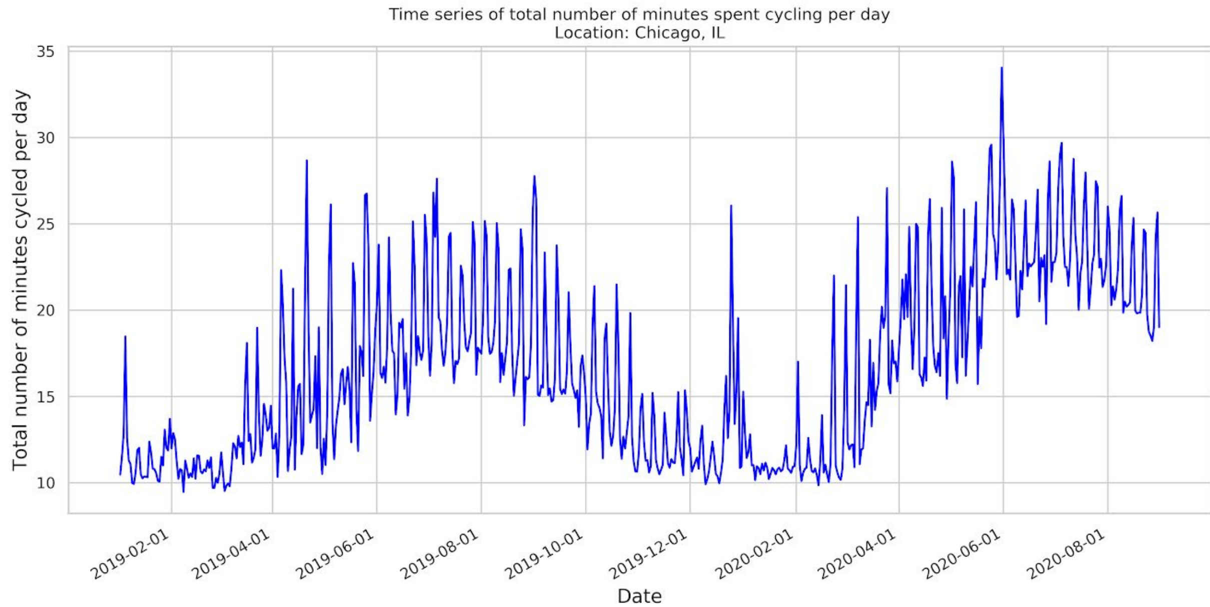
Many thanks to our mentors Partha Bose and Elisa Omodei, as well as our TA Kessie Zhang and the entire Correlation One team for their support and guidance throughout this project.

Appendix

Possible increase in bike ridership and negative correlation with air pollution

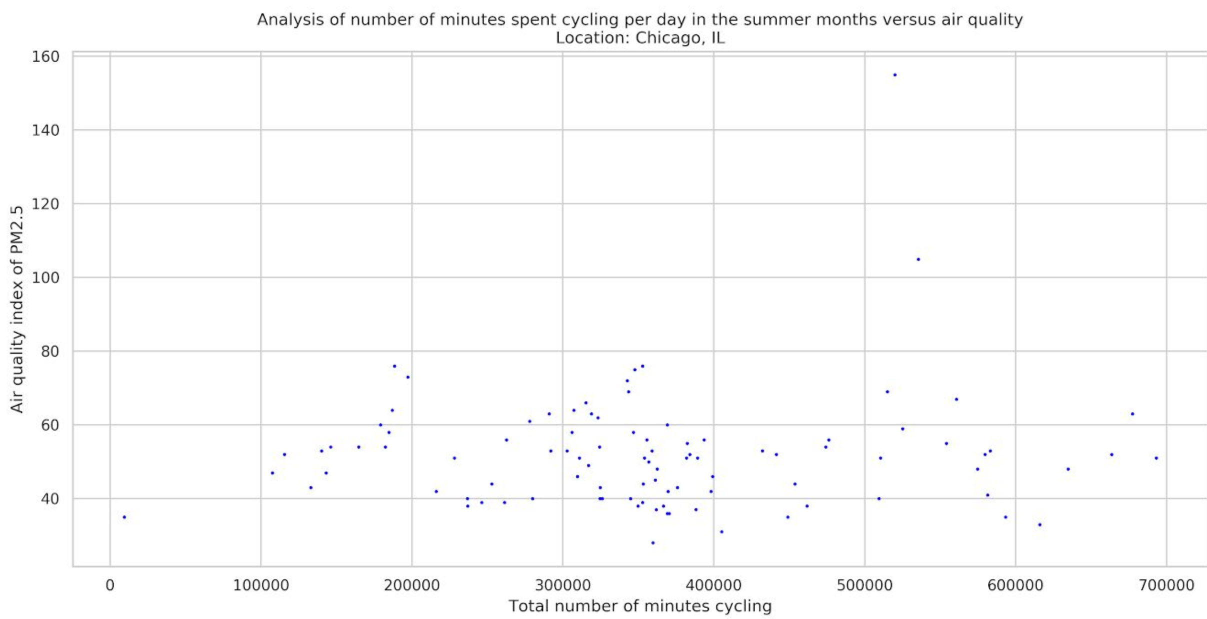
We wished to investigate whether the pandemic has changed people's habits with regard to transport, causing them to switch from motor vehicles to rented pedal bikes, and if so, whether this has had an impact on local air quality.

We investigated two US cities for changes in the pattern of their cycling behaviour before and after the pandemic. Results for Chicago, IL are shown below. (Results for New York are displayed in the appendix.) The following chart shows the total number of minutes cycled per day on rental bikes in Chicago, IL over time:



Ridership appears to have gone up slightly in the latter half of 2020, compared to the same period the year before.

To investigate whether there is an improvement in air quality as cycling duration increases, we produced a 2D scatter plot of air quality index for particulates of type PM2.5 versus total number of minutes cycled:



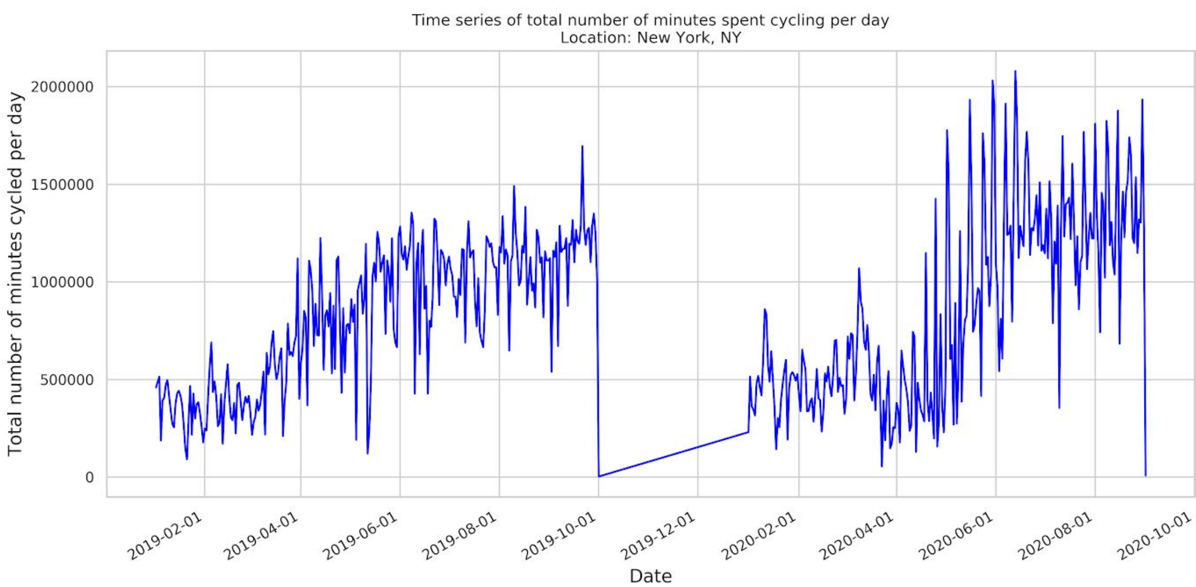
The scatter plot includes data points for dates in the range 26 May 2020 - 31 August 2020, which is when weather in Chicago is most permitting for cycling. The plot does not show any obvious signs of correlation between air quality and cycling activity. (We obtained similar random-looking results for New York, NY as well - see the tables below.)

Before embarking on analysing the data, we removed the two outliers in the top right-hand side since they occur on 4th and 5th of July, so the spikes in air pollution observed then must be due to fireworks, not transport.

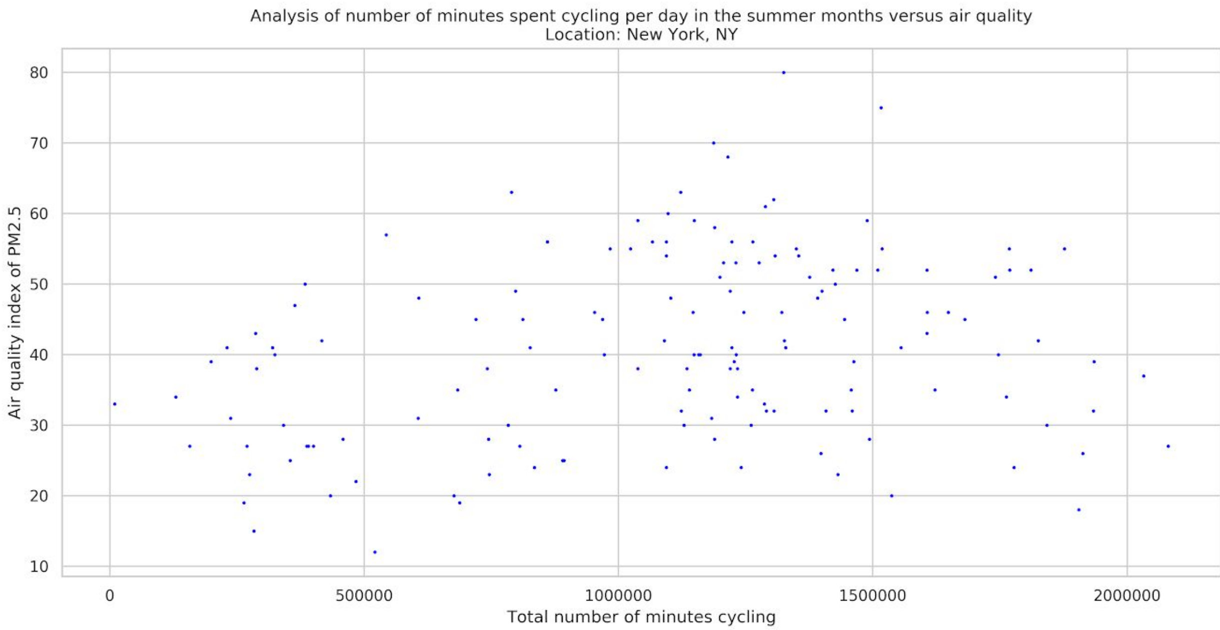
To verify that there is no correlation between the two quantities, we explored this relationship statistically using regression analysis and calculated Pearson's r correlation coefficient. Pearson's r correlation coefficient came to be -0.0544, which suggests a very slight negative correlation, and is a trend that we expect to find. Performing linear regression analysis on the data, however, yielded an R-squared value of 0.003, a very weak negative coefficient of -4.291 x 10^-6, and a high p-value of 0.602. (The full table of regression results is shown in the tables below.)

Taken together, these results indicate that there isn't sufficient statistical evidence to support the conjecture that air quality depends on the amount of cycling that people do, and as a result we are not able to conclude definitively that cycling is having any impact on air quality in Chicago. It is possible, however, that there is a correlation between the two that is being obscured by other intervening factors - such as the weather, for example. Further investigations are warranted to explore such possible confounding factors.

Variation over time of number of minutes spent cycling on hire bikes in New York



Dependence of air quality in New York on number of minutes spent cycling on hire bikes



Tables of Regression Results for Air Quality as a Dependent Variable of Total Number of Minutes Cycled in Chicago

Outcome including outliers:

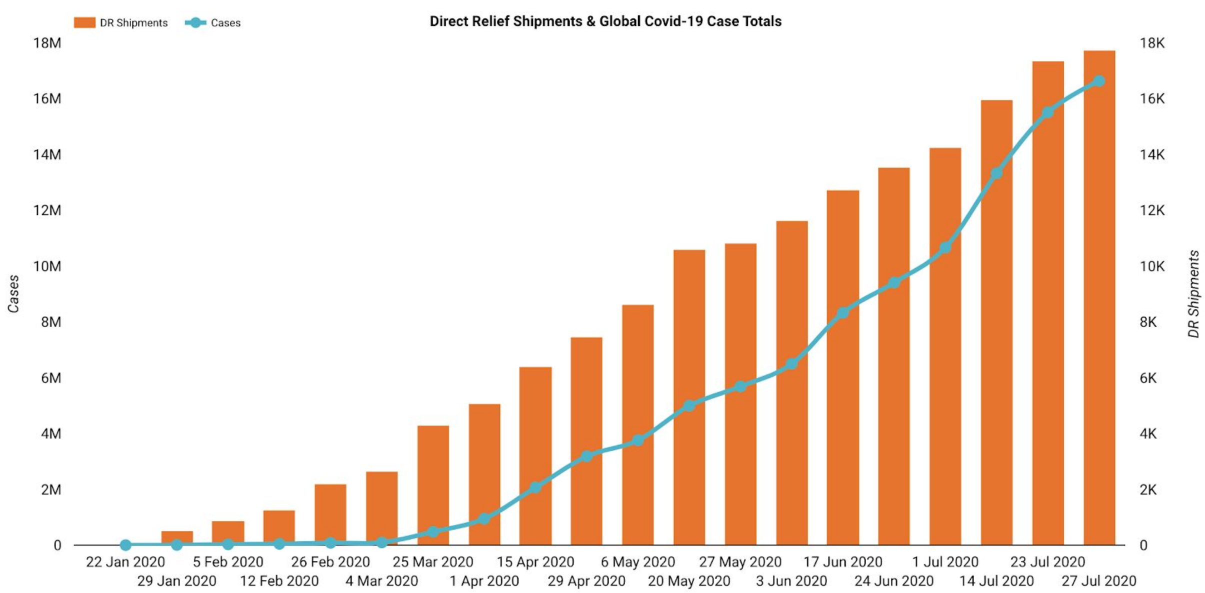
```
                            OLS Regression Results
==============================================================================
Dep. Variable:          PM2p5_AQI_Value   R-squared:                       0.008
Model:                              OLS   Adj. R-squared:                 -0.003
Method:                   Least Squares   F-statistic:                    0.7371
Date:                  Thu, 15 Oct 2020   Prob (F-statistic):              0.393
Time:                          00:13:29   Log-Likelihood:                -402.10
No. Observations:                    96   AIC:                             808.2
Df Residuals:                        94   BIC:                             813.3
Df Model:                             1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     48.2664      4.615     10.458      0.000      39.102      57.430
sum_min     1.026e-05    1.2e-05      0.859      0.393   -1.35e-05     3.4e-05
==============================================================================
Omnibus:                         92.047   Durbin-Watson:                   0.872
Prob(Omnibus):                    0.000   Jarque-Bera (JB):             1127.960
Skew:                             2.997   Prob(JB):                    1.17e-245
Kurtosis:                        18.686   Cond. No.                     1.08e+06
==============================================================================
```

Outcome excluding outliers:

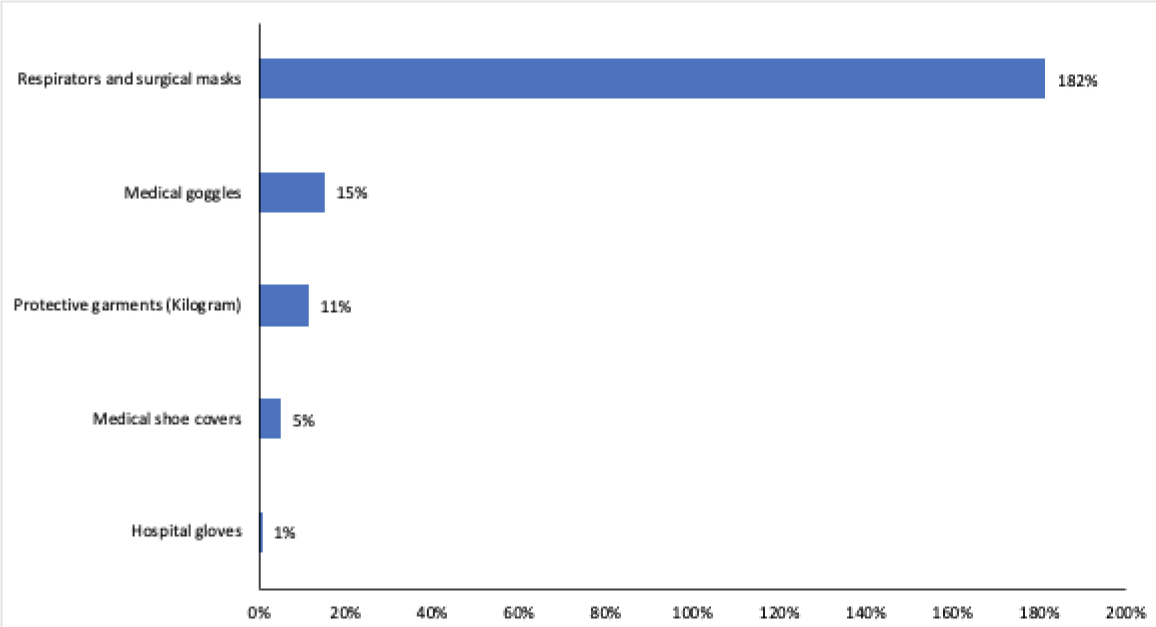
```
                            OLS Regression Results
==============================================================================
Dep. Variable:          PM2p5_AQI_Value   R-squared:                       0.003
Model:                              OLS   Adj. R-squared:                 -0.008
Method:                   Least Squares   F-statistic:                    0.2738
Date:                  Thu, 15 Oct 2020   Prob (F-statistic):              0.602
Time:                          02:27:10   Log-Likelihood:                -356.79
No. Observations:                    94   AIC:                             717.6
Df Residuals:                        92   BIC:                             722.7
Df Model:                             1
Covariance Type:              nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     51.8415      3.137     16.525      0.000      45.611      58.072
sum_min    -4.291e-06    8.2e-06     -0.523      0.602   -2.06e-05     1.2e-05
==============================================================================
Omnibus:                          2.689   Durbin-Watson:                   0.898
Prob(Omnibus):                    0.261   Jarque-Bera (JB):                2.654
Skew:                             0.362   Prob(JB):                        0.265
Kurtosis:                         2.608   Cond. No.                     1.07e+06
==============================================================================
```

Additional analysis on PPE and Plastic Waste

The percentage reductions in net exports from the year before are likely due to the need to import more PPE than the ones produced locally. Some evidence of this can also be found in the increased donations of PPE to hospitals from charities such as the Direct Relief, confirming the needs for higher consumption than the stocks consented. Similar evidence can be found in the reports made by [FEMA](#).

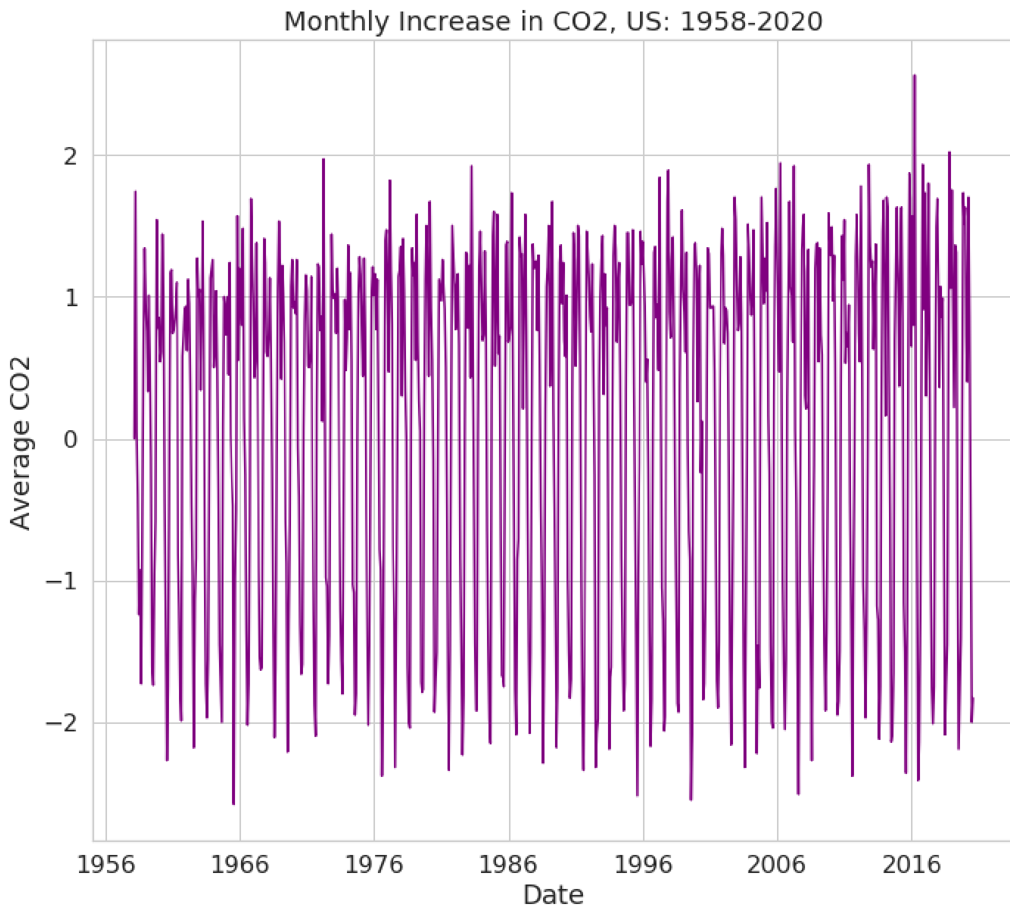


In fact, the worldwide increase in demand contributed to the economic recession in the US, as not only the volume, but also the value of imports rose.



Additional analysis on CO2 Emissions

As further confirmation that CO2 levels have not changed since COVID, we consider differences from the average concentration of CO2 from the month before. To do so we create a new variable differencing the n-th month's CO2 average and the n-1-th. We observe no different pattern after the beginning of the pandemic.



If the lockdown period had caused a decrease in emissions we should observe all 6 last observations below zero, showing lower concentrations of CO2 than the month before, but this is not the case.

While CO2 emissions have not changed, it is correlated with transportation

We find a significant correlation between multiple transportation variables and CO2 levels. However, some recent transportation data such as air traffic post-COVID and highway vehicle mileage 2018 - 2019 show no significant correlation. These findings that recent transportation data is not correlated with CO2 are consistent with our conclusion that CO2 levels have not changed since the lockdown in March.

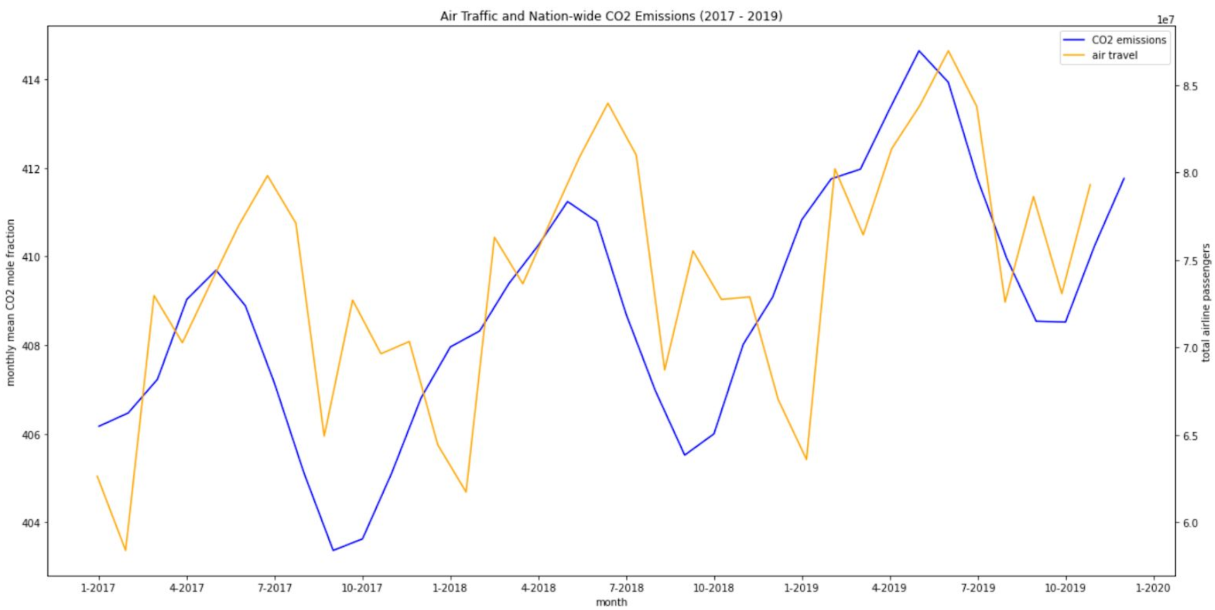
Looking at transportation and CO2 data visually, it appears that vehicle miles traveled and air travel are correlated with CO2 emissions. From this initial EDA (as well as from a basic understanding of the effects of burning fuel) it seemed that CO2 emissions would drop when COVID hit because there were fewer planes in the sky and fewer cars on the road. Using Pearson and Spearman correlation coefficients, we assessed these assumptions. The three transportation variables in the table below have statistically significant correlations with CO2 measures.

Transportation and CO2 - Significant Correlations

Transportation Measure	Correlation Coefficient	p-value
Air Traffic (2017 - 2019)	0.45	0.0059
Annual Performed Flights (1990 - 2019)	0.414	0.023
Vehicle Highway Miles Traveled (1990 - 2009)	0.968	2.64 * 10^-12

Monthly Air Traffic

The initial EDA shown above using time series analysis has peaks and drops around the same time and thus appears to have a correlation, but is not entirely clear.



Both the Air Traffic and the CO2 data over 2017 - 2019 are normally distributed and have a linear relationship, so the Pearson Test was the best choice. These two variables have a correlation coefficient of 0.45 with a p-value of 0.0059, which indicates that monthly air traffic and CO2 emissions have a statistically significant, moderate, positive correlation over the years 2017-2019.

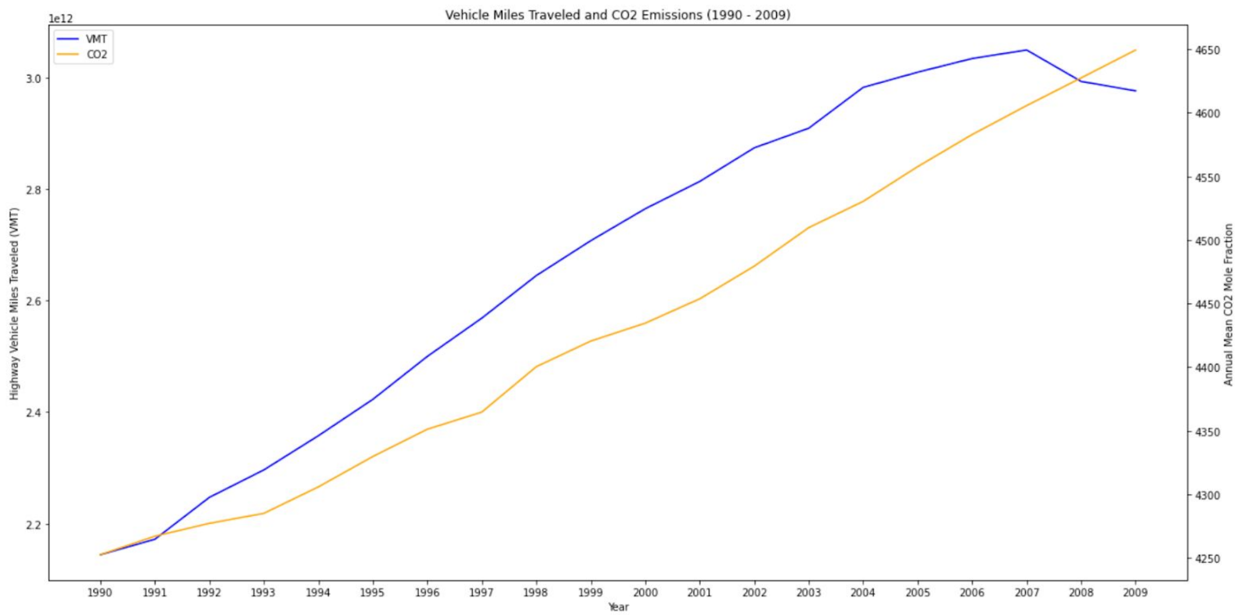
While CO2 and Air Traffic before 2020 have a significant correlation, the four observations post-COVID (March 2020 - June 2020) do not. While this small of a dataset presents a concern, the more recent data does not show any statistically significant correlation.

Annual Performed Flights

The Annual Performed Flights from 1990 - 2019 (also reported from the U.S. Bureau of Transportation Statistics) also displays a significant correlation with annual CO2 emissions. The two have a moderate, positive correlation, with a correlation coefficient of 0.414 at a p-value of 0.023.

Vehicle Mileage

An exploratory time-series analysis of Highway Vehicle Miles Traveled (VMT) and CO2 emissions in the U.S. between 1990 and 2009 also displays what appears to be a strong, positive correlation between the two over time.



Neither VMT nor CO2 have a normal distribution over this time, so we used the Spearman test for this data. The test yields a correlation coefficient of 0.968 at a p-value of 2.64*10^-12, which confirms a highly significant, strong, positive correlation. While this correlation is significant, more recent monthly data does not tell the same story. An analysis of monthly CO2 and VMT data from January 2018 - December 2019 shows no significant correlation.

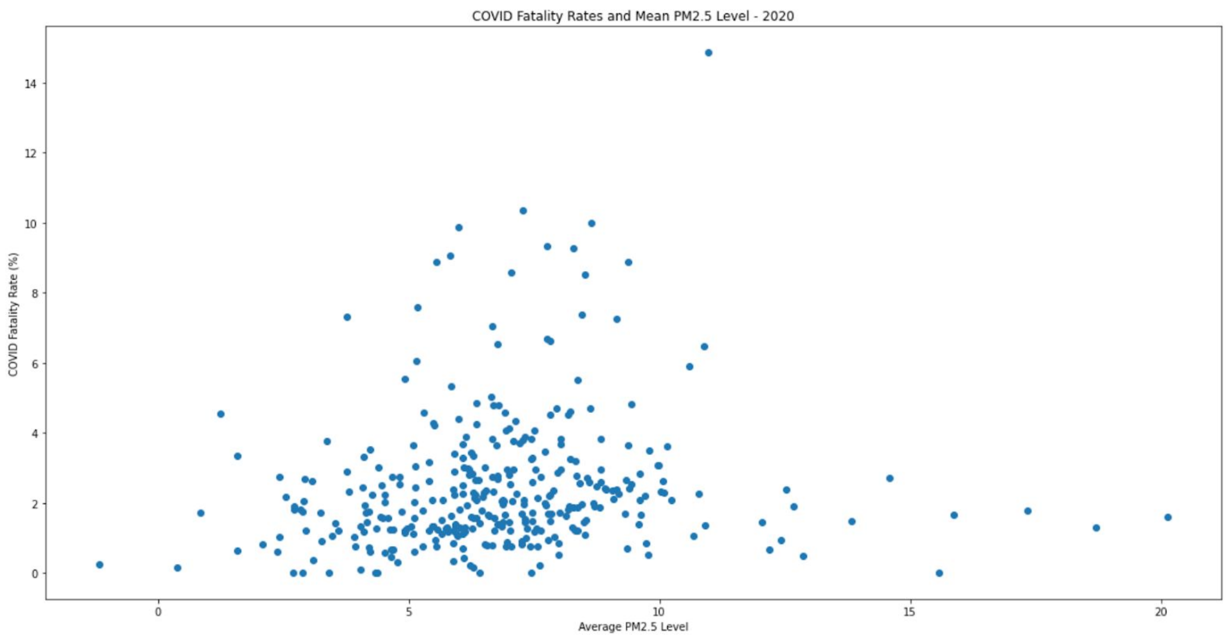
With the clear and strong evidence that CO2 and Transportation levels are correlated, we believed that CO2 levels would drop when COVID halted transportation. However, our analysis (as shown above) proved this assumption to be untrue.

Does poor air quality correlate with COVID severity?

With Harvard and other articles reporting that air quality has an effect on the severity of COVID cases, we wanted to assess how PM2.5, PM10, and NO2 may be correlated with COVID severity (measured as number of deaths / number of cases) by county.

PM2.5

- The correlation between average county PM2.5 levels over the past 3 years and severity of COVID cases is a weak, positive, statistically significant correlation. The correlation coefficient is 0.131 with a p-value of 0.0013.
- The correlation between average county PM2.5 levels in 2020 and severity of COVID cases (plotted below) is a weak, positive, highly statistically significant correlation. The correlation coefficient is 0.211 with a p-value of < 0.001.

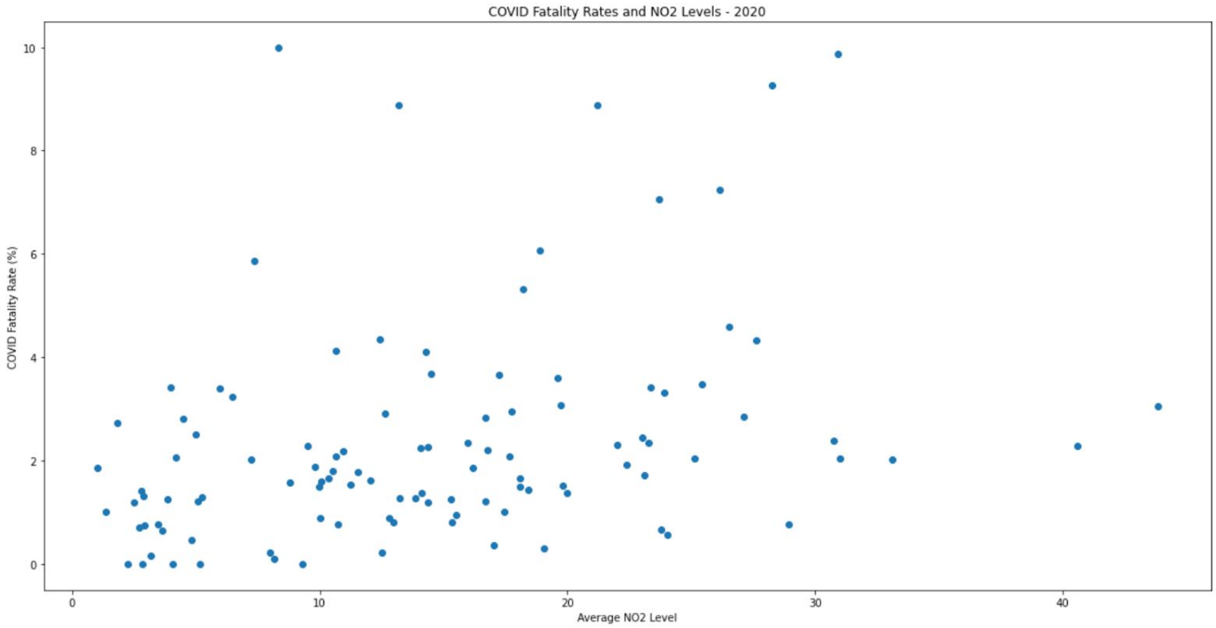


PM10

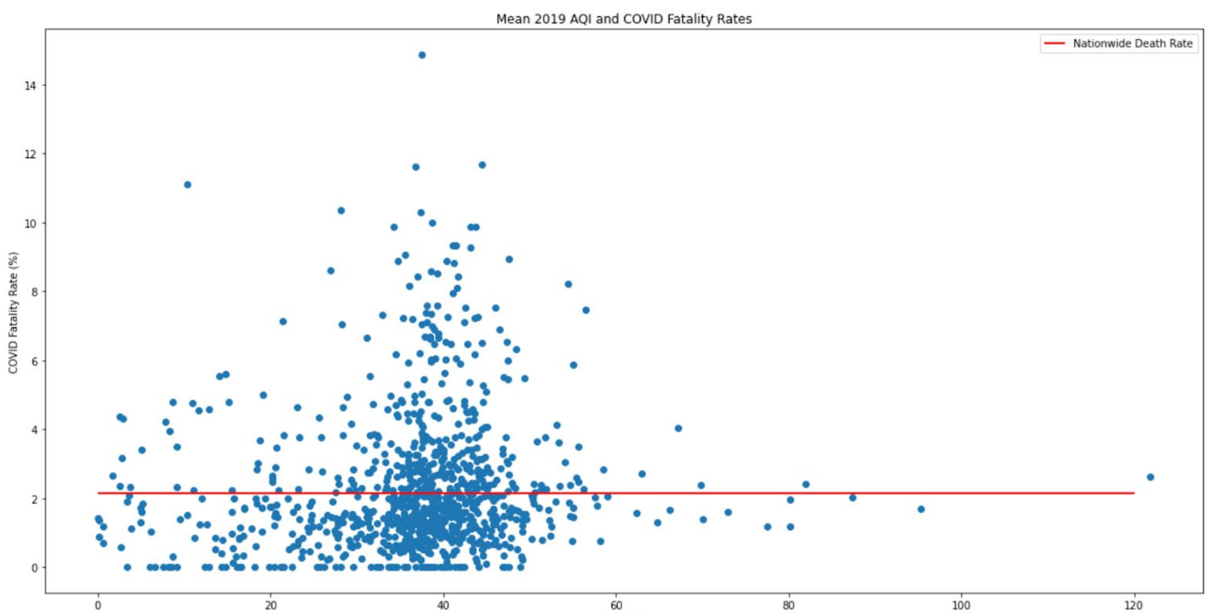
We find no evidence of a correlation between average annual PM10 and COVID severity in either 2018 - 2020 or in 2020 alone.

NO2

- The correlation between average county NO2 levels over the past 3 years and COVID severity is a medium, positive, statistically significant correlation The correlation coefficient is 0.377 with a p-value of < 0.001.
- The correlation between average county NO2 levels in 2020 alone and COVID severity (shown below) is a medium, positive, statistically significant correlation. The correlation coefficient is 0.438 with a p-value of < 0.001.



AQI



The average air quality index in 2019 -- which was also displayed geospatially in the report above -- has a significant correlation with COVID severity as well, with a correlation coefficient of 0.16 at a p-value of < 0.001.

The analysis of these three pollutants and AQI display significant correlations between the severity of COVID cases and different measurements of air quality in the region. While this is an important finding, we caveat this by pointing out that county population and demographics may be confounding variables. The counties that are more highly populated likely have worse air quality, and they may have busier hospitals where they received less attention.

Data and Methods

To attempt to answer the questions set out in our objectives, we needed multiple data sources. For each specific combination of factors, below, we discuss the data we required and obtained:

COVID and plastic use data

We found monthly data on plastic production spanning 1972 until July/August 2020 for 'plastic products', 'plastic and rubber products' and 'plastic material and resin' for the whole US from the Federal Reserve Bank of St. Louis [6]. We used the data from January 2019 up to July 2020 for the analysis and used all three plastic types. We understand that plastic production in these three categories may not include all plastic used in the USA, as imports and exports might alter this slightly, and this also may not correspond exactly to plastic waste, as some plastic may be used, reused and recycled, so there may be a lag or reduction in the amount of waste in relation to this. With these caveats in mind, our assumptions are that production can be used as a proxy for waste in this analysis, understanding that this might impact the interpretation of the results.

COVID and CO2 emissions data

We use the monthly CO2 data at the country and the state level from the U.S. Energy Information Administration [7] and monthly concentration levels from NASA's Global Climate Change Research [8]. To address the source of CO2 emissions we merge the country level dataset with datasets about monthly and annual rental bike, automobile, and airplane travel from the Bureau of Transportation Statistics [9] and the Office of Highway Policy Information [10] to compare their trends with the CO2 emissions. We then proceed to test whether the claimed improvements in air quality reduced the concentration of CO2 in the atmosphere.

COVID case counts and air quality data

In order to observe any possible correlation between the severity and frequency of COVID cases and air quality, we used Air Quality Index (AQI) data from the Environmental Protection Agency [11] at the city and county level and COVID data at the county level from USAFacts [12]. To clean these data, we grouped the case count and death count datasets by county and divided death counts by case counts to calculate a death rate per county which we can compare to the national average.

For analysis of the Air Quality data, we grouped AQI observations by county and by state and took the average for each region. These values allowed us to observe air quality trends across the country and compare them with how the pandemic has affected individual counties.

Methods used for spatial analysis

The AQI data was fed into ArcGISOnline as a point layer. Then, we used the Interpolation tool to predict the air quality at the locations for which we did not have data. This tool is best for point layers with a numeric field which expects a gradual change between locations, so the AQI dataset meets all of the requirements. When running this tool, we compromised speed of execution and accuracy at 50/50 and then clipped the predictions to a map layer of U.S. counties. The Interpolate Points feature outputs a maximum and minimum predicted value for each geographic point, and the maps displayed in our research both depict the maximum predicted value. The errors for both years are available in the Github repository.

Under the hood, ArcGIS's Interpolate Points feature uses the Empirical Bayesian Kriging geoprocessing tool - you can find more details about this process on ArcGIS's website.