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

DS4A: Women's Summit - Team #21

Laura Collett, Mirella Koleva, Chiara Sotis, Bailey Wellen

The Problem: A Tale of Two Environments

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).

In this project we try and analyze the trends before and after the beginning of the pandemic to identify the gain in air quality and the impact of this new amount of plastic waste.

Impact

Aims

1) Mapping the evolution of air pollution and air quality in the last 2 years
2) Identifying the reductions of CO_2 and PM2.5 due to a reduction in travel and general improvements in air quality imputable to the pandemic
3) Identifying the trends in plastic consumption and waste over the last 2 years
4) Identifying the increase in production and consumption of plastic due to the pandemic
5) (Ambitious:) identifying whether the increase in plastic also constitutes a shift from recyclable to non-recyclable plastic
6) (Ambitious:) identifying whether the pandemic induced changes in the form of travel, e.g. more bikes than cars/public transport
7) (Ambitious:) identifying if the locations with lower air quality have higher rates of COVID cases

The UN Environment 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, NewScientist, Wired and the BBC 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 the two issues and quantify the impact of COVID-19 on these two key environmental issues.

Feasibility

We started with a broad topic, and have worked hard to narrow this down into tangible, workable ideas and questions. We hope this project now fits into the available feasibility parameters. Throughout the project we will assess the time scale and if certain questions are taking too long to understand and get to the bottom of, we will narrow the scope. Likewise, if we see other areas of interest arise in the data, we will increase the scope in that direction so that interesting findings can be presented. Our aims above highlight our understanding of the ambition of each strand of the project.

Data

Data for CO2 Analysis:

U.S. Energy Information Administration : "State energy-related carbon dioxide emissions by year, unadjusted (1990-2017)"

Combined total energy-related carbon dioxide emissions levels by state. Fields are:

`State(string)` **:** State names

`1990, 1991… 2017 (float)` **:** Energy-related CO2 levels per year

U.S. Energy Information Administration : "Transportation energy-related carbon dioxide emissions"

Transportation-related carbon dioxide emissions levels by state for years 1980 - 2017. Fields are:

`State(string)` **:** State names

`1990, 1991… 2017 (float)` **:** Energy-related CO2 levels per year

NASA Carbon Dioxide Levels

Text file providing cO2 levels by month from March 1958 - August 2020. Fields are:

`Year (string)` **:** year of observation

`Month (string)` **:** month of observation

`Average (float)` **:** The "average" column contains the monthly mean CO2 mole fraction determined from daily averages. The mole fraction of CO2, expressed as parts per million (ppm) is the number of molecules of CO2 in every one million molecules of dried air (water vapor removed). If there are missing days concentrated either early or late in the month, the monthly mean is corrected to the middle of the month using the average seasonal cycle. Missing months are denoted by -99.99.

`Interpolated (float)` **:** The "interpolated" column includes average values from the preceding column and interpolated values where data are missing. Interpolated values are computed in two steps. First, we compute for each month the average seasonal cycle in a 7-year window around each monthly value. In this way the seasonal cycle is allowed to change slowly over time.

`Trend (float)` **:** We then determine the "trend" value for each month by removing the seasonal cycle; this result is shown in the "trend" column. Trend values are linearly interpolated for missing months. The interpolated monthly mean is then the sum of the average seasonal cycle value and the trend value for the missing month.

U.S. Department of Transportation: Monthly Transportation Statistics

The dataset consists of 138 columns representing different measures of U.S. transportation. The fields that we plan to use are:

`Date (datetime)` **:** date of collection

`U.S. Airline Traffic - Total - Non Seasonally Adjusted (float)` : total air traffic data based on monthly reports from commercial U.S. air carriers

`U.S. Airline Traffic - International - Non Seasonally Adjusted (float)` : international air traffic data based on monthly reports from commercial U.S. air carriers

`U.S. Airline Traffic - Domestic - Non Seasonally Adjusted (float)` : domestic air traffic data based on monthly reports from commercial U.S. air carriers

`Highway Vehicle Miles Traveled - All Systems (float)` : vehicle miles traveled on all roads and streets each month estimated by the Federal Highway Administration

United States Environmental Protection Agency Air Quality Tracker Dataset

This dataset only contains information on Ozone or PM2.5 or Ozone and PM2.5 combined, so won't specifically be able to look at CO2, but will give other insights into the air quality. We will be able to choose data from the city or county level.

`Date (date)` : Daily data available from 01/01/2020 to today

`Value (float)` : Air quality

`Main Pollutant (string)` : Ozone or PM2.5

`Site name (string)` : detection site within city or county

`5 year average (float)` : average value from the last 5 years on that specific date

Data for plastic analysis:

Federal Reserve Bank of St. Louis Plastic Manufacturing Dataset:

These datasets contain data from 1972 until July/August 2020.

`DATE (date)` : First of month

`IPN325211S (float)` : Plastic material and resin production 'industrial production index'

`IPG326S (float)` : Plastics and rubber products production 'industrial production index'

`IPG3261S (float)` : Plastic products production 'industrial production index'

Data for COVID Analysis:

U.S. Census : County Population Data

2010 census data of population by county and predictions for populations through 2019. Fields are:

`STNAME (string)` : State name

`CTYNAME(string)` : County name

`POPESTIMATE2019 (int)` : Estimate of population in 2019 based on birth and deaths since 2010

U.S.A. Facts : US Coronavirus Cases

Coronavirus case counts in the U.S. by county starting 1/22/20. Fields are:

`County Name (string)` : County name

`State (string)` : State name

`1/22/20, 1/23/20 … 9/25/20 (int)` : Number of reported COVID cases per day per county

U.S.A. Facts : US Coronavirus Deaths

Coronavirus death counts in the U.S. by county starting 1/22/20. Fields are:

`County Name (string)` : County name

`State (string)` : State name

`1/22/20, 1/23/20 … 9/25/20 (int)` : Number of reported deaths due to COVID per day per county

Methods

EDA

Before any formal analysis we will conduct some exploratory analysis of the data to understand the evolution in consumption and production of plastic and what might be driving it and decompose air quality into different pollutants and emitting sources. To take a closer look at what might be affecting these variables we will also try and find relevant scientific evidence on the causes behind different trends, e.g. policy reforms.

Geospatial Visualization

We will heavily rely on spatial visualizations and maps to show the trends in air quality and graphically decompose the main drivers of pollution. We will use maps to clearly display differences in pollution levels across the United States. In addition, these maps will be useful in determining how COVID hotspots may also be areas with low air quality.

Time Series Regressions

To explore the air quality and amount of plastic waste over time we will use time-series analyses. We will focus on deviations from the trends in trying to estimate how large the difference from the trend is and whether this is statistically significant.
We will also use time series plots to show the evolution of plastic consumption and production and, if the data allows it, to decompose whether there are different trends in

consumption and production of different types of plastic, with a focus on whether the material is recyclable or not.

Statistical Significance Testing
In order to determine whether any of the changes may be due to the effect we are looking at or due to chance, we will test some of our hypotheses using statistical significance testing. We will possibly look at paired tests for before and after COVID-19 to see whether there are any statistically significant changes. We won't be able to test for confounding factors, either whether a factor we have identified is a confounder, or whether there are confounding factors in the effect we are looking at that we haven't identified, but this will be an untestable assumption we make in these analyses, knowing that this may be a limitation to the project.

Timeline / Project Breakdown

Task	Team Member
Spatial Analysis / Maps	Bailey
EDA	Mirella
Time Series Analysis	Chiara
Significance Testing	Laura
Putting Together Poster	ALL
Writing Report	ALL

Concerns

The main concern we share is that we all have somewhat limited experience coding and come from different backgrounds/are used to different softwares. We are hoping that online guides (Medium, Youtube, etc) and the material we cover in DS4A (along with our sheer patience!) will allow us to go beyond simple EDA and shed more light on these two drivers of environmental quality during the pandemic.

Additionally, we are glad to have a lot of data, but are concerned about deciding and narrowing down which of our datasets will best help us to our conclusions. Much of the data is similar but at different levels (both geographically and by time) so we need to investigate which levels have the right amount of precision to allow us to draw conclusions about COVID's effects on these environmental issues. One specific concern we may have when trying to make inferences on the impact of plastic waste is that we lack specific plastic waste data and may have to make assumptions on the waste based on the production, which may not be robust.