

## Team # 21

### **Proposal #1: How the COVID-19 pandemic has affected both air quality and single use plastic waste**

**Problem Statement:** *(please describe the project in 1-2 sentences)*

Air quality has a major impact on human health and plastic waste has an even more major impact on the health of the planet. Now with the covid-19 pandemic grinding businesses, airlines and other CO2 producers to a halt, alongside the enforced use of disposable personal protective equipment and otherwise avoidable single use plastic, levels may have changed.

**Which question(s) do you want to explore? Why do you think this particular question is interesting?:** *(1-2 paragraphs elaborating on the project's relevance)*

Using data from the US, we intend to investigate air quality over time. We want to discover whether there has been a change in air quality since covid-19 hit, given the abrupt shift the pandemic has forced on the daily lives of citizens and businesses alike. We'd similarly like to explore plastic production for the same period, as it's possible that these two things are linked. We will incorporate CO2 emissions data from an earlier period and data on transportation to determine whether these may be associated with air quality, and whether there has been any similar shift in plastic production given covid-19.

We think these topics are both highly important and highly relevant to the present time, given how air quality can impact human health, in particular in the presence of a respiratory illness, and how plastic pollution has been highlighted as a growing problem that needs to be solved, but may have been set back given the covid-19 crisis and need for single-use items in the interests of hygiene. We think that any insight in these crucial areas would be important.

**Which datasets do you plan to use? Why? Are there any data sources that you have failed to find?** *(List relevant datasets)*

Federal Reserve Bank of St. Louis Plastic Manufacturing Dataset

<https://fred.stlouisfed.org/searchresults/?st=plastic&t=plastics%3Bmonthly%3Btrade&rt=trade&ob=sr>

United States Environmental Protection Agency Air Quality Tracker Dataset

<https://www.epa.gov/outdoor-air-quality-data/air-data-daily-air-quality-tracker>

U.S. Energy Information Administration: Energy-Related CO2 Emission Data Tables (1990 - 2017)

<https://www.eia.gov/environment/emissions/state/>

U.S. Dept of Transportation: Monthly Transportation Statistics

<https://data.bts.gov/Research-and-Statistics/Monthly-Transportation-Statistics/crem-w557>

**Please describe the plan or methodology that you will use to answer your question** *(1-2 sentence description of statistical analysis techniques)*

To explore the air quality and amount of plastic waste over time we will use exploratory data analysis, data visualisation and time-series analyses. We will focus on the explanatory variables mentioned to see the effect of these on the outcome variables air quality and plastic production.

## **Proposal #2: Water quality project**

**Problem Statement:** *(please describe the project in 1-2 sentences)*

Access to water and sanitation is recognised as a human right and yet many people still live without access to clean water. Research done by the WHO has shown that those with access to clean water have better health outcomes, so we intend to look at what might affect water quality in the United States.

**Which question(s) do you want to explore? Why do you think this particular question is interesting?:** *(1-2 paragraphs elaborating on the project's relevance)*

We intend to explore how water quality has changed over time and across the geographical counties of the US, as well as questions such as: how does water quality relate to political affiliation and how does it relate to common industries in each county?

**Which datasets do you plan to use? Why? Are there any data sources that you have failed to find?** *(List relevant datasets)*

- chemicals data containing the measured mean concentration of a particular chemical or substance in community water systems throughout the counties in the United States from 2000 – 2016. (CDC)
- droughts data containing the particular percentage of various range of drought severities, indexed by counties for particular start-end periods throughout the United States.
- political affiliation MIT dataset on political affiliation containing county-level returns for presidential elections from 2000 to 2016.
- Industry\_occupation data containing the estimated working population (16 years and over) for the various industries indexed by counties, taken from 2010 – 2016.
- County business patterns United States Census Bureau dataset on number of establishments, people in employment and total payroll per US county, from 2000 - 2016.

**Please describe the plan or methodology that you will use to answer your question** *(1-2 sentence description of statistical analysis techniques)*

We will initially look at the trends in water quality throughout the counties in time. To do so we will heavily rely on easy-to-interpret maps. In the same way we'll use graphs to easily convey the political affiliation of the counties considered. To think about the determinants of water quality we will mostly consider regression analysis.