### Documentation for this file: tidy\_mobility\_pollutants.RData

**Date range:** 03/01/2020 - 04/18/2020

**df** = dataframe with all variables of interest

* Counties we’re interested in
* Aqi values
* Different pollutant values (not available for some counties)
* Mobility information

**increase** = dataframe with only the counties that had increased mobility during the date range in the original article

**decrease** = dataframe with only the counties that had decreased mobility during the date range in the original article

### Data sources:

**Mobility:**

*Source & Documentation*: <https://github.com/descarteslabs/DL-COVID-19>

*Website of their analyses:* https://www.descarteslabs.com/mobility/#overview

**AQI and pollutants:**

*Source:* <https://www.epa.gov/outdoor-air-quality-data/air-quality-index-daily-values-report>

*Documentation:* <https://www.epa.gov/outdoor-air-quality-data/about-air-data-reports#aqidaily>

### Data Definitions (yellow means what data we need to run analyses):

**m50:** The median of the max-distance mobility for all samples in the specified region.

**m50\_index:** The percent of normal m50 in the region, with normal m50 defined during 2020-02-17 to 2020-03-07.

**Aqi:** average air quality index for that day

**Pollutant names:** each have their average daily value

**Main\_pollutant:** from the original data, represented which pollutant was the main pollutant affecting AQI

Data Prep Documentation

First, we took a look at the tidy data sets we prepared before analysis. There are three datasets we will be working with:

\* df: dataset containing the AQI, mobility index, and pollutant values for the 10 counties of interest

\* high\_mobility: decreased mobility 30% or less between Feb. 28 to Mar. 27

\* low\_mobility: dataset with counties that decreased mobility 98-100% between Feb. 28 to Mar. 27

### Data Sources

#### Air Quality Data

We collected our air quality data from the EPA's free API service. https://www.epa.gov/outdoor-air-quality-data/air-quality-index-daily-values-report

#### Mobility Data

Write blurb about what this data set is.

https://github.com/descarteslabs/DL-COVID-19

### Definitions

\* m50: The median of the max-distance mobility for all samples in the specified region.

\* m50\_index: The percent of normal m50 in the region, with normal m50 defined during 2020-02-17 to 2020-03-07.

\* AQI: average air quality index for that day

\* Pollutant names (e.g. Ozone, PM25): each have their average daily value

\* main\_pollutant: from the original data, represented which pollutant was the main pollutant affecting AQI

Because the mobility dataset we used only has data from March 1st we chose to remove analysis of air quality data from the months of January and February. Thus, our analyses only covers from March 1st to April 18th.