

Exploratory Data Final Project

Overview

Fine particulate matter (PM_{2.5}) is an ambient air pollutant for which there is strong evidence that it is harmful to human health. In the United States, the Environmental Protection Agency (EPA) is tasked with setting national ambient air quality standards for fine PM and for tracking the emissions of this pollutant into the atmosphere. Approximately every 3 years, the EPA releases its database on emissions of PM_{2.5}. This database is known as the National Emissions Inventory (NEI). You can read more information about the NEI at the EPA National Emissions Inventory web site: [<http://www.epa.gov/ttn/chief/eiinformation.html>] For each year and for each type of PM source, the NEI records how many tons of PM_{2.5} were emitted from that source over the course of the entire year. The data that you will use for this assignment are for 1999, 2002, 2005, and 2008.

Data

The data for this assignment are available from the course web site as a single zip file: [https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2FNEI_data.zip] The zip file contains two files:

PM_{2.5} Emissions Data (summarySCC_PM25.rds): This file contains a data frame with all of the PM_{2.5} emissions data for 1999, 2002, 2005, and 2008. For each year, the table contains number of tons of PM_{2.5} emitted from a specific type of source for the entire year. Here are the first few rows.

```
fips      SCC Pollutant Emissions  type year
4 09001 10100401 PM25-PRI 15.714 POINT 1999 8 09001 10100404 PM25-PRI 234.178 POINT 1999 12
09001 10100501 PM25-PRI 0.128 POINT 1999 16 09001 10200401 PM25-PRI 2.036 POINT 1999 20 09001
10200504 PM25-PRI 0.388 POINT 1999 24 09001 10200602 PM25-PRI 1.490 POINT 1999
```

- fips: A five-digit number (represented as a string) indicating the U.S. county
- SCC: The name of the source as indicated by a digit string (see source code classification table)
- Pollutant: A string indicating the pollutant
- Emissions: Amount of PM_{2.5} emitted, in tons
- type: The type of source (point, non-point, on-road, or non-road)
- year: The year of emissions recorded

Source Classification Code Table (Source_Classification_Code.rds): This table provides a mapping from the SCC digit strings in the Emissions table to the actual name of the PM_{2.5} source. The sources are categorized in a few different ways from more general to more specific and you may choose to explore whatever categories you think are most useful. For example, source “10100101” is known as “Ext Comb /Electric Gen /Anthracite Coal /Pulverized Coal”.

You can read each of the two files using the readRDS() function in R. For example, reading in each file can be done with the following code:

```
NEI <- readRDS(“summarySCC_PM25.rds”) SCC <- readRDS(“Source_Classification_Code.rds”)
```

Assignment

The overall goal of this assignment is to explore the National Emissions Inventory database and see what it says about fine particulate matter pollution in the United States over the 10-year period 1999–2008. You may use any R package you want to support your analysis.

Questions

You must address the following questions and tasks in your exploratory analysis. For each question/task you will need to make a single plot. Unless specified, you can use any plotting system in R to make your plot.

- Have total emissions from PM2.5 decreased in the United States from 1999 to 2008? Using the base plotting system, make a plot showing the total PM2.5 emission from all sources for each of the years 1999, 2002, 2005, and 2008.
- Have total emissions from PM2.5 decreased in the Baltimore City, Maryland (fips == “24510”) from 1999 to 2008? Use the base plotting system to make a plot answering this question.
- Of the four types of sources indicated by the type (point, nonpoint, onroad, nonroad) variable, which of these four sources have seen decreases in emissions from 1999–2008 for Baltimore City? Which have seen increases in emissions from 1999–2008? Use the ggplot2 plotting system to make a plot answer this question.
- Across the United States, how have emissions from coal combustion-related sources changed from 1999–2008?
- How have emissions from motor vehicle sources changed from 1999–2008 in Baltimore City?
- Compare emissions from motor vehicle sources in Baltimore City with emissions from motor vehicle sources in Los Angeles County, California (fips == “06037”). Which city has seen greater changes over time in motor vehicle emissions?

Making and Submitting Plots

For each plot you should

- Construct the plot and save it to a PNG file.
- Create a separate R code file (plot1.R, plot2.R, etc.) that constructs the corresponding plot, i.e. code in plot1.R constructs the plot1.png plot. Your code file should include code for reading the data so that the plot can be fully reproduced. You must also include the code that creates the PNG file. Only include the code for a single plot (i.e. plot1.R should only include code for producing plot1.png)
- Upload the PNG file on the Assignment submission page
- Copy and paste the R code from the corresponding R file into the text box at the appropriate point in the peer assessment.

About the Code

Please note that the Emissions are measure on THOUSAND TONS scale, observed from 1999 to 2008. A separated file, namely ‘fetch-data.R’ is used specifically for the purpose of download, unzipping and setting some general variables. ## Some variables

- NEI: The NEI dataset, with 6,497,651 observations of 6 variables - SCC: The database used to map the SCC code with 11717 observations - merged_dat: The dataset with SCC mapped to corresponding description in the SCC database. - totalpm25: The total PM2.5 emissions across the US from 1999 to 2008. - pm25_baltimore: The total PM2.5 emissions in the Baltimore City, Maryland (fips=‘24510’) without taking into account the four types. - baltimore_dat: The total PM2.5 emissions in the Baltimore City, consisting of four types (point, nonpoint, onroad, nonroad). - coal_source: The PM2.5 emissions from coal combustion-related sources accross the US. - coal_agg: The aggregated dataset of coal_source by the sum of the Emissions variable. - motor_emission: Total PM2.5 emissions from motor vehicle in Baltimore City. - motor_sources_baltiLA: Total PM2.5 emissions from motor vehicle in Baltimore City and LA (fips=‘06037’).