Exploratory Data Analysis
by Roger D. Peng, PhD, Jeff Leek, PhD, Brian Caffo, PhD

Course Home Page

COURSE
- Announcements
- Course Books
- Video Lectures
- Discussion Forums

EXERCISES
- Quizzes
- Course Projects
- swirl Exercises

ABOUT THE COURSE
- Syllabus
- About the Instructor

COMMUNITY
- DSS Community Site

Help Center

Peer Assessments / Course Project 2 Help Cer

Submission Phase | Evaluation Phase | Results Phase

1. Do assignment | 2. Evaluate peers | 3. See results

No work was submitted before the submission deadline. You will not be able to evaluate the work of your peers or receive an evaluation.

Introduction

Fine particulate matter (PM2.5) 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 wit setting national ambient air quality standards for fine PM and for tracking the emissions of this pollutant into the atmosphere. Approximatly every 3 years, the EPA releases its database on emissions of PM2.5 This database is known as the National Emissions Inventory (NEI). You can read more information abou the NEI at the EPA National Emissions Inventory web site.

For each year and for each type of PM source, the NEI records how many tons of PM2.5 were emitted from that source over the course of the entire year. The data that you will use for this assignment are fo 1999, 2002, 2005, and 2008.

Data

The data for this assignment are available from the course web site as a single zip file:

- Data for Peer Assessment [29Mb]

The zip file contains two files:

PM2.5 Emissions Data (`summarySCC_PM25.rds`): This file contains a data frame with all of the PM2.5 emissions data for 1999, 2002, 2005, and 2008. For each year, the table contains number of **tons** of PM2.5 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 PM2.5 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 fro the SCC digit strings in the Emissions table to the actual name of the PM2.5 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”.

```
## This first line will likely take a few seconds. Be patient!
NEI <- readRDS("summarySCC_PM25.rds")
SCC <- readRDS("Source_Classification_Code.rds")
```

as long as each of those files is in your current working directory (check by calling `dir()` and see if tho files are in the listing).

Assignment

The overall goal of this assignment is to explore the National Emissions Inventory database and see wh it say about fine particulate matter pollution in the United states over the 10-year period 1999–2008. Yo 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/tas you will need to make a single plot. Unless specified, you can use any plotting system in R to make your plot.

1. 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.
2. Have total emissions from PM2.5 decreased in the **Baltimore City**, Maryland (`fips == "24510"`) fro 1999 to 2008? Use the **base** plotting system to make a plot answering this question.
3. Of the four types of sources indicated by the `type` (point, nonpoint, onroad, nonroad) variable, whi of these four sources have seen decreases in emissions from 1999–2008 for **Baltimore City**? Whic have seen increases in emissions from 1999–2008? Use the **ggplot2** plotting system to make a plo answer this question.
4. Across the United States, how have emissions from coal combustion-related sources changed from 1999–2008?
5. How have emissions from motor vehicle sources changed from 1999–2008 in **Baltimore City**?
6. 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 chang 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 i the peer assessment.

Have total emissions from $PM_{2.5}$ decreased in the United States from 1999 to 2008? Using the **base**plotting system, make a plot showing the *total* $PM_{2.5}$ emission from all sources for each of the years 1999, 2002, 2005, and 2008.
Upload a PNG file containing your plot addressing this question.

Evaluation/feedback on the above work

Please view the plot for this question. Does the plot appear to address the question being asked? In other words, can you answer the question using the information shown in the plot?

Copy and paste the R code file for the plot uploaded in the previous question.

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

Have total emissions from $PM_{2.5}$ decreased in the **Baltimore City**, Maryland (`fips == 24510`) from 199 to 2008? Use the **base** plotting system to make a plot answering this question.

Upload a PNG file containing your plot addressing this question.

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Please view the plot for this question. Does the plot appear to address the question being asked? In other words, can you answer the question using the information shown in the plot?

Copy and paste the R code file for the plot uploaded in the previous question.

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Of the four types of sources indicated by the `type` (point, nonpoint, onroad, nonroad) variable, which o 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.

Upload a PNG file containing your plot addressing this question.

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Please view the plot for this question. Does the plot appear to address the question being asked? In other words, can you answer the question using the information shown in the plot?

Copy and paste the R code file for the plot uploaded in the previous question.

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

Across the United States, how have emissions from coal combustion-related sources changed from 1999–2008?

Upload a PNG file containing your plot addressing this question.

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

shown in the plot?

Copy and paste the R code file for the plot uploaded in the previous question.

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

How have emissions from motor vehicle sources changed from 1999–2008 in **Baltimore City**?

Upload a PNG file containing your plot addressing this question.

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Please view the plot for this question. Does the plot appear to address the question being asked? In other words, can you answer the question using the information shown in the plot?

Copy and paste the R code file for the plot uploaded in the previous question.

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

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 ove time in motor vehicle emissions?

Upload a PNG file containing your plot addressing this question.

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Please view the plot for this question. Does the plot appear to address the question being asked? In other words, can you answer the question using the information shown in the plot?

Copy and paste the R code file for the plot uploaded in the previous question.

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

Overall evaluation/feedback

Note: this section can only be filled out during the evaluation phase.

Please use the space below to provide constructive feedback to the student who submitted the work. Point out the submission's strengths as well as areas in need of improvement. You may also use this space to explain your grading decisions.

You've written 0 words