CSCI 3900C Project III (75 points)

Write an R script to accomplish the tasks described below. Please put all code in a single script file, and comment your code appropriately. Then compose a report using the results generated by your script. Submit BOTH (R script and report pdf file) to the Project III D2L drop box.

Background

Fine particulate matter (PM2.5) is a pollutant present in the air. There is strong evidence that PM2.5 is harmful to human health. In the U. S., the Environmental Protection Agency (EPA) sets national air quality standards for fine PM and tracks the emissions of PM2.5 into the atmosphere. About every 3 years, the EPA releases its database on emissions of PM2.5. This database is known as the National Emissions Inventory (NEI). See the EPA National Emissions Inventory web site for more details.

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.

Getting Started

Download Project3.zip and extract the following 2 files:

summarySCC_PM25.rds	*This file contains the data you will need to analyze*
Source_Classification_Code.rds	*This file may be useful as a reference*

The data file contains PM2.5 emissions data for 1999, 2002, 2005, and 2008.

The appendix at the end of this document explains the content and format of these files.

Program Tasks

Use the data file provided to investigate PM2.5 emissions data for the designated years.
The first step in your program should be to read the data file with the readRDS() function as follows:

```
NEI <- readRDS("summarySCC_PM25.rds")
```

NOTES:

a. *This command assumes that the file is in your current working directory.*
b. *The file is very large, so this command may take several seconds*

Answer EACH question below by constructing ***only one*** plot that addresses the question. (Your plot may have multiple facets, but there should only be one overall plot to address each question.) Generate all plots using the ggplot2 package. Annotate plots with appropriate labels and attributes.

1. Have total PM2.5 emissions decreased in the United States from 1999 to 2008? Your plot(s) should show 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 area of Baltimore City, Maryland from 1999 to 2008? (Note the **`fips`** code for this area is "24510".)
3. The type variable identifies four types of pollutant sources (point, nonpoint, onroad, nonroad). Of these four types of sources, which have seen emission increases in the Baltimore City area from 1999 to 2008? Which have seen emission decreases in that area during the same time period?
4. Compare emissions from onroad sources in Baltimore City with emissions from onroad sources in Los Angeles County, California (**`fips`** code "06037"). Which city has seen greater changes over time in onroad emissions?

Report

- Include one section for each question.
- In each section:
    - Repeat the entire question
    - Show your plot
    - In 2-3 sentences, explain what your plot means. Be sure your explanation answers or addresses the question!

Appendix

The format and content of the data files are described below.

`summarySCC_PM25.rds` – *the PM2.5 Emissions Data file*

The 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
```

The variables (columns) in the data frame are

- fips — A string containing a 5-digit code representing the county
- SCC — A digit string representing the specific pollutant source
- Pollutant — A string identifying the pollutant
- Emissions — Amount of PM2.5 emitted (tons)
- type — The type of the source (point, non-point, on-road, non-road)
- year — The year the emissions were recorded

`Source_Classification_Code.rds` – *a reference file*

The file contains a table with information about each pollutant source. It can be used to look up the SCC digit string in the data file above. For each SCC digit string, this file pro vides a detailed name for the pollutant source, as well as multiple classifications. (This information will not be explicitly needed for the current project, but it will give you a better sense of the codes in the main data file and what they mean.) Here is ***part*** of the first row in the table:

SCC	Data. Category	Short.Name	EI. Sector	SCC. Level.One	SCC. Level.Two
10100101	Point	Ext Comb /Electric Gen /Anthracite Coal /Pulverized Coal	Fuel Comb - Electric Generation - Coal	External Combustion Boilers	Electric Generation