National Emissions Inventory Analysis

Exploratory Data Analysis Course Project

Andrew Burns

Quarter 1, 2016

Introduction

This project was used as a course project for the Exploratory Data Analysis Course by Johns Hopkins. I used the features of the R programming language that I have learned in this course and previous courses. The task was to look at NEI(National Emissions Inventory) and explore the data through graphical features of R. I am not making any conclusions or drawing any inferences from the data, but rather, I am exploring the data to see why it contains. The data included record of readings from emission sensors throughout the country and over a number of years.
The values for each record were:

- `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

More can be read about PM2.5 and Emissions here :
https://www.epa.gov/air-emissions-inventories

Processes

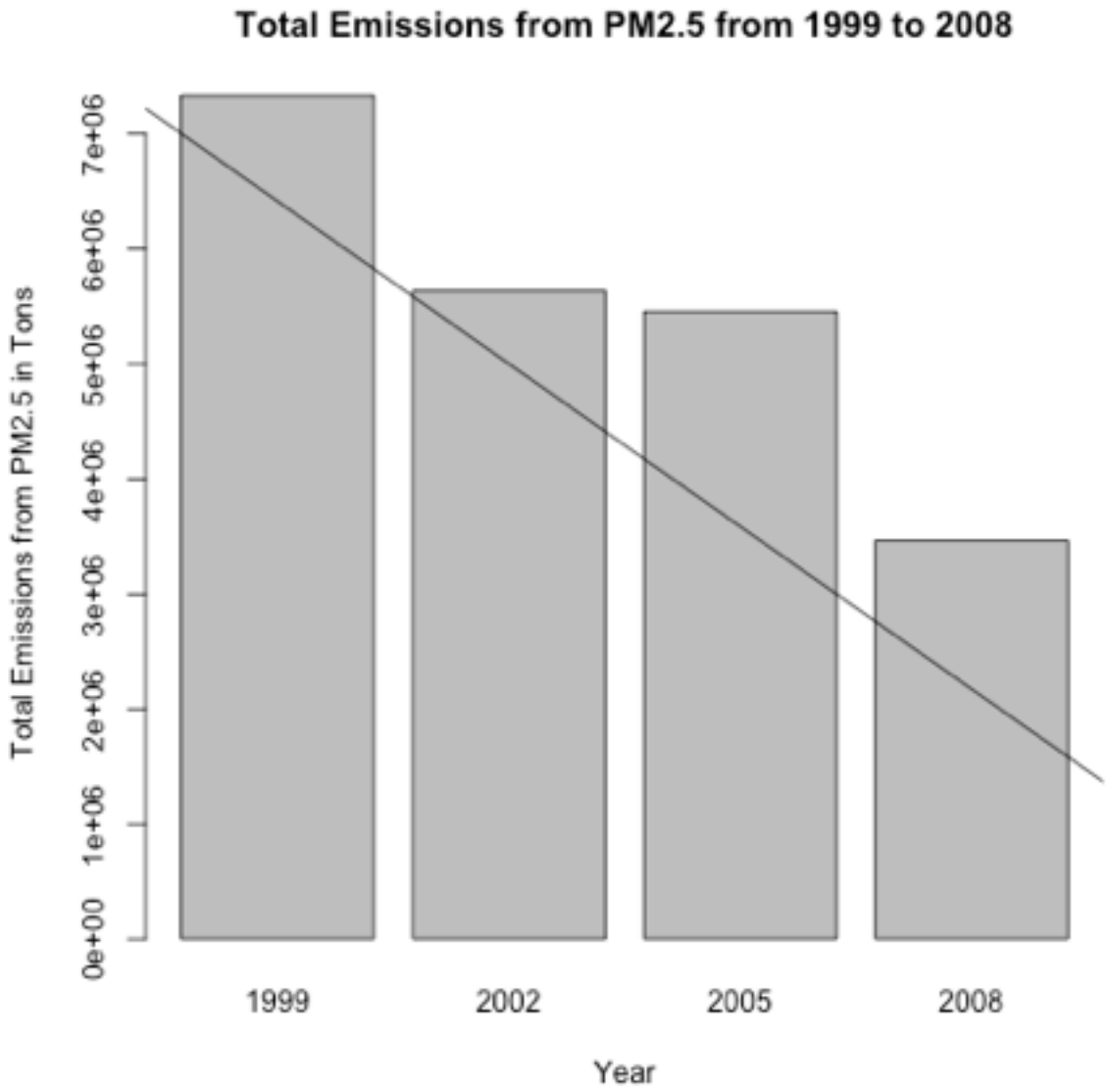
I used the R language and RStudio to conduct this project. Some of the skills I used in the project were:

- ggplot2 : A graphing package of R
- Dataframes + Vectors + Lists
- For loops
- Subsetting : Selecting part of a data frame based on the parameters I am searching
- Barplots
- Box plots + Line graphs (Not used in the final results)
- Factors : A data type in R that has limited values (ex: a person can only be assigned Male or Female)
- Grepl : Search list or data frame for matches

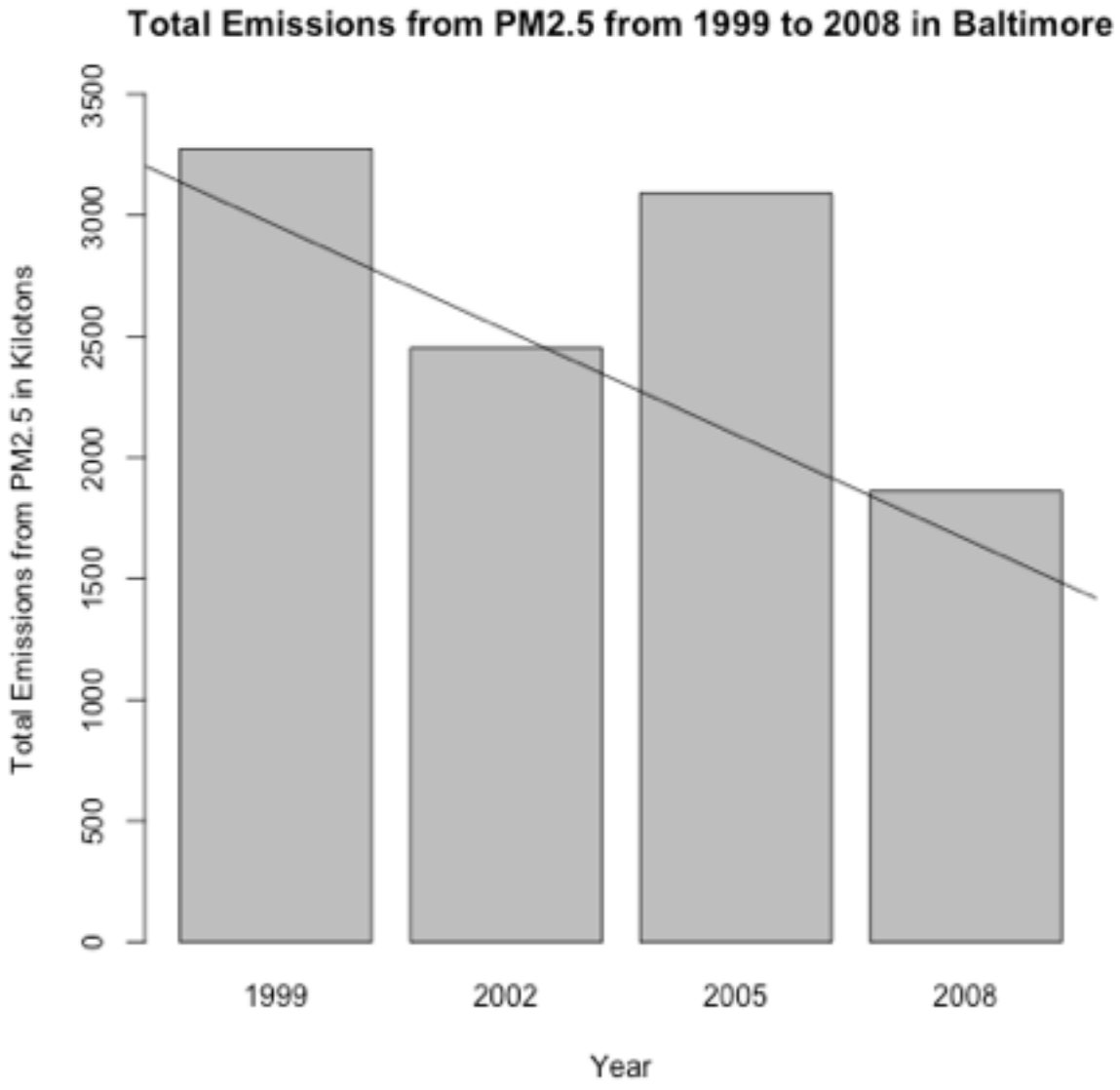
The idea of the code was to make it as concise as possible and in the fewest lines of code possible. I did this by combing functions and assigning variables on the fly within a function. In the first graph, I didn't use the easy way to subset the data, but I took a longer route because I wanted to try out some things in R that would be useful to practice.

Analysis

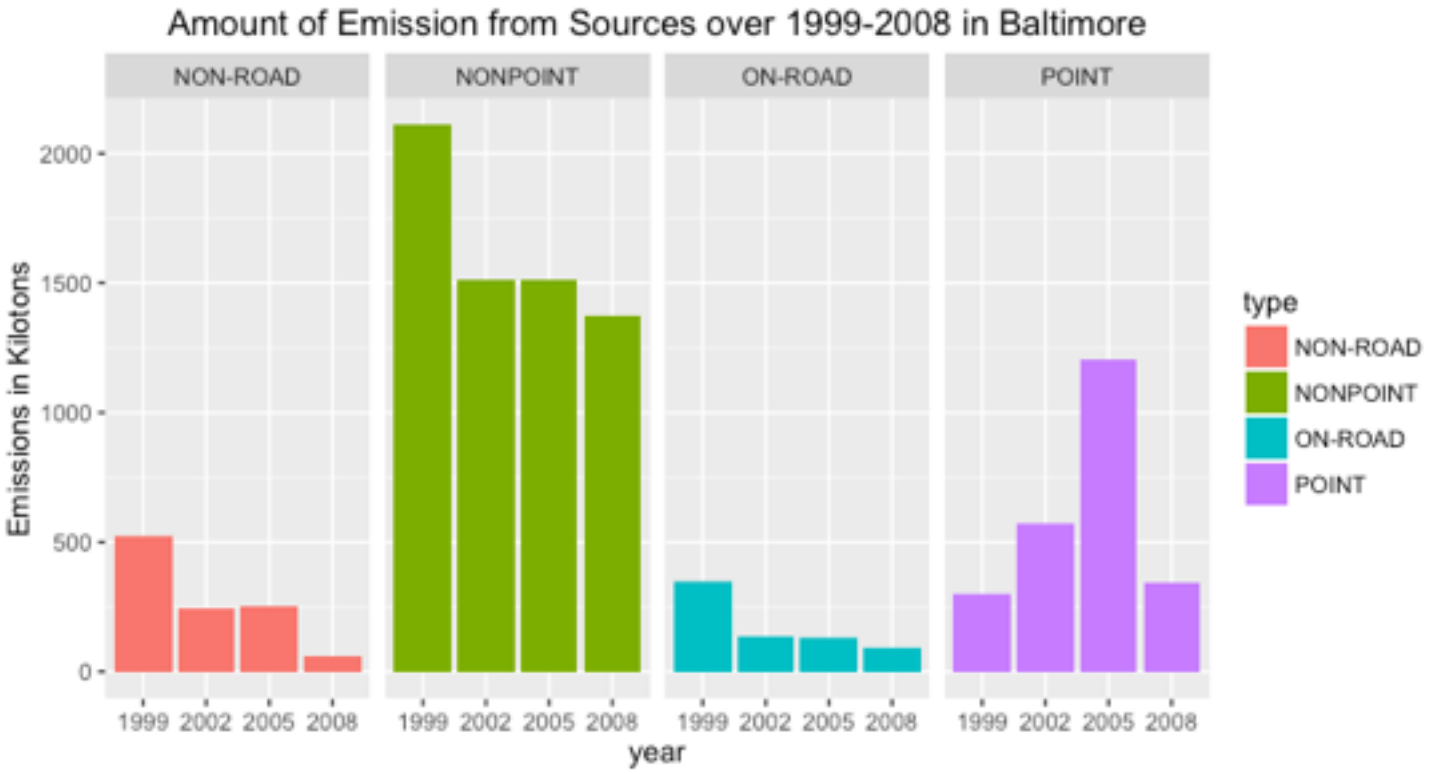
1. Have total emissions from PM2.5 decreased in the United States from 1999 to 2008? Using the base plotting system, make a plot showing the total PM2.5 emission from all sources for each of the years 1999, 2002, 2005, and 2008.



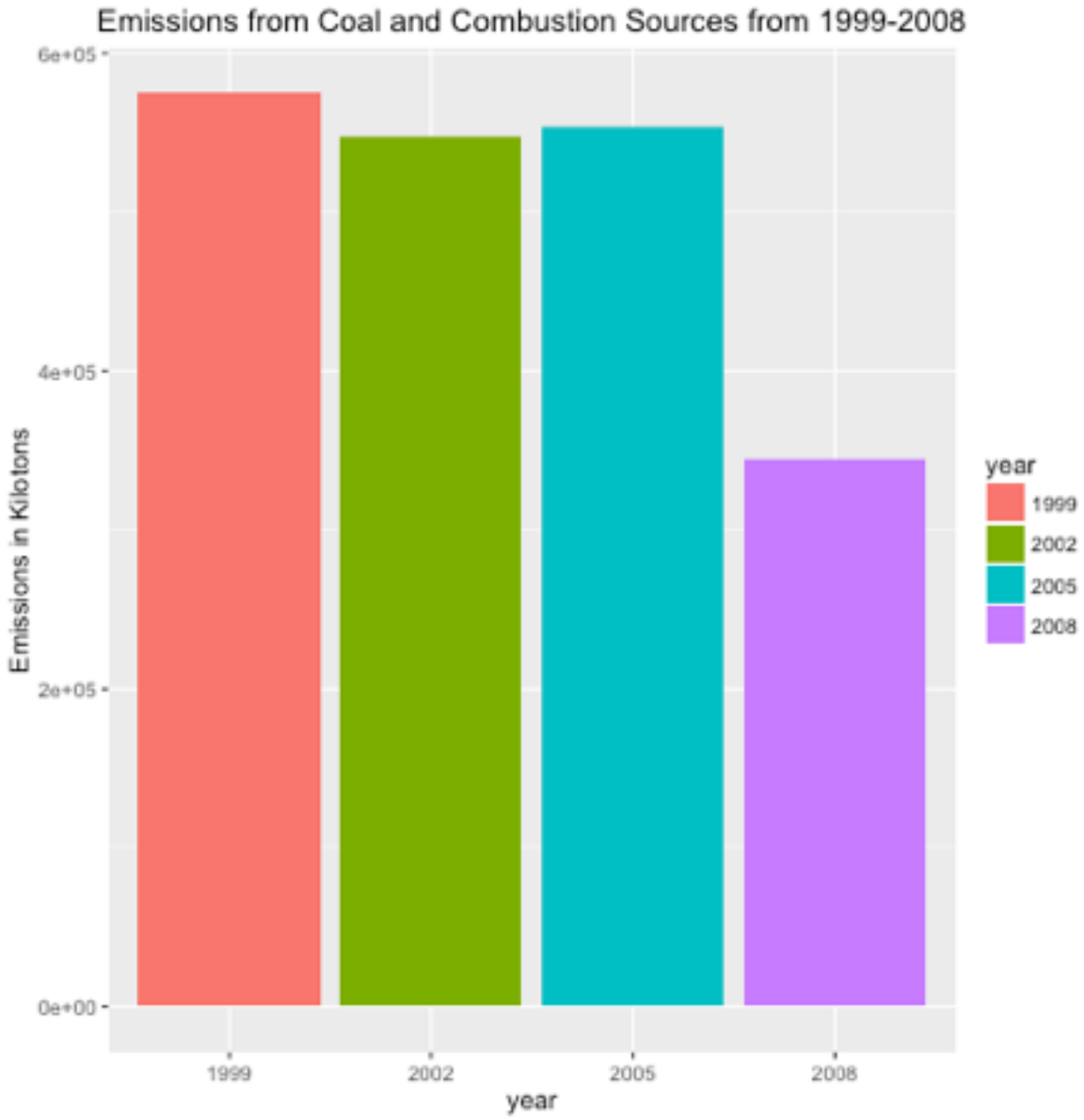
2. Have total emissions from PM2.5 decreased in the Baltimore City, Maryland (`fips == "24510"`) from 1999 to 2008? Use the base plotting system to make a plot answering this question.



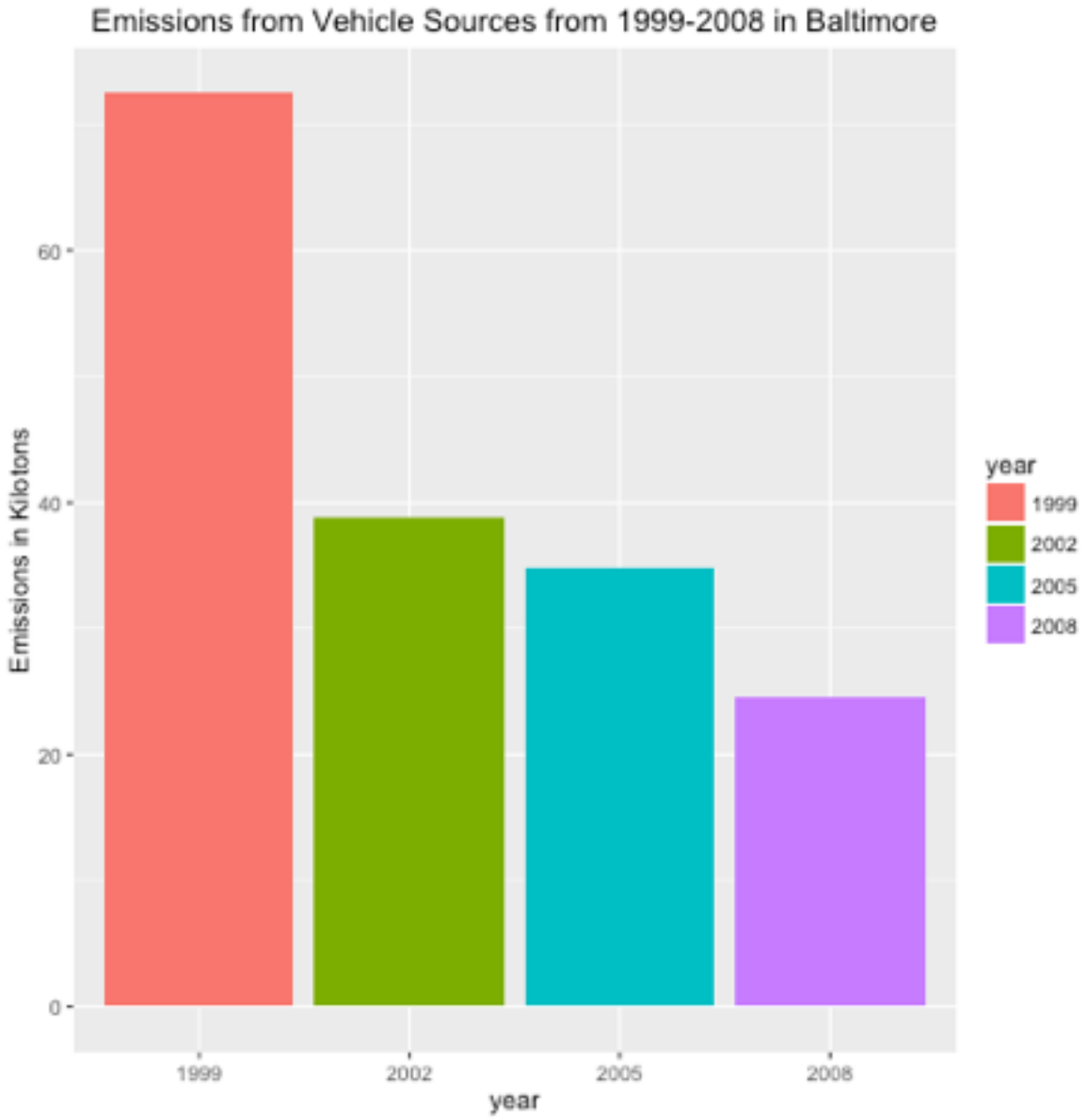
3. Of the four types of sources indicated by the `type` (point, nonpoint, onroad, nonroad) variable, which of these four sources have seen decreases in emissions from 1999–2008 for Baltimore City? Which have seen increases in emissions from 1999–2008? Use the ggplot2 plotting system to make a plot answer this question.



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 changes over time in motor vehicle emissions?