Plot3

Jerez

11/29/2020

Course: Exploratory Data Analysis

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 with 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 about 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 for 1999, 2002, 2005, and 2008.

Library

```
setwd("~/Google Drive/Coursera/EDA")

# Install and load required packages
if (!require('ggplot2'))     {install.packages('ggplot2')}
```

```
## Loading required package: ggplot2
```

```
if (!require('dplyr'))      {install.packages('dplyr')}
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
if (!require('data.table')) {install.packages('data.table')}
```

```
## Loading required package: data.table
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```r
library(ggplot2)
library(dplyr)
library(data.table)
```

Download the dataset

```r
# name for zip file
setwd("~/Google Drive/Coursera/EDA")
file.zip <- 'EDA_Final.zip'

# Cheking if zip file exists
if (!file.exists(file.zip)){
 file.URL <- 'https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2FNEI_data.zip'
 download.file(file.URL, file.zip, method = 'curl')
 unzip(file.zip, exdir = '.')}

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

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.

```r
head(NEI)
```

```
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

```r
Baltimore <- NEI %>% subset(fips == '24510') %>%
 group_by(year, type) %>%
 summarize(Total = sum(Emissions, na.rm = TRUE)) # Removing the NA's
```
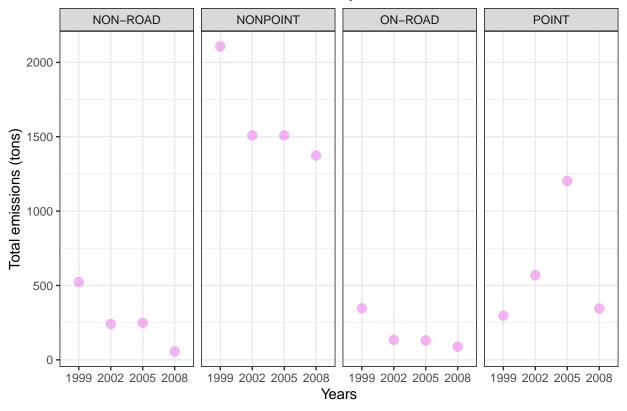
Baltimore

A tibble: 16 x 3

Groups: year [4]

year type   Total

1 1999 NON-ROAD 523. 2 1999 NONPOINT 2108. 3 1999 ON-ROAD 347. 4 1999 POINT 297. 5 2002 NON-ROAD 241. 6 2002 NONPOINT 1510. 7 2002 ON-ROAD 134. 8 2002 POINT 569. 9 2005 NON-ROAD 249. 10 2005 NONPOINT 1510. 11 2005 ON-ROAD 130. 12 2005 POINT 1202. 13 2008 NON-ROAD 55.8 14 2008 NONPOINT 1373. 15 2008 ON-ROAD 88.3 16 2008 POINT 345.

```
# Plot using ggplot

Plot3 <- ggplot(Baltimore, aes(as.factor(year), Total, color = type,  alpha = 0.9))+
 geom_point(size = 3, alpha = 0.6, color = 'violet')+
 #geom_boxplot()+
 #geom_jitter(alpha = 0.3)+
 facet_grid(. ~ type)+
 labs(x = 'Years', y = 'Total emissions (tons)',
      title = 'Total Annual Emissions in Baltimore by Year')+

 theme_bw()

Plot3
```

Total Annual Emissions in Baltimore by Year

```r
ggsave('Plot3.png', plot = Plot3, width = 18, height = 7, units = 'cm')
```

Create Plot3.R

```r
#library(knitr)
#purl('Plot3.Rmd')
```