Plot1

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 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.

```r
# Total PM2.5 emission from 1999 to 2008

# First, we look inside the dataset
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
# Then, wee group by year, and summarise the total emissions
Emiss.tot <- NEI %>%
 group_by(year) %>%
 summarize(Total = sum(Emissions, na.rm = TRUE)) # Removing the NA's
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
Emiss.tot
```

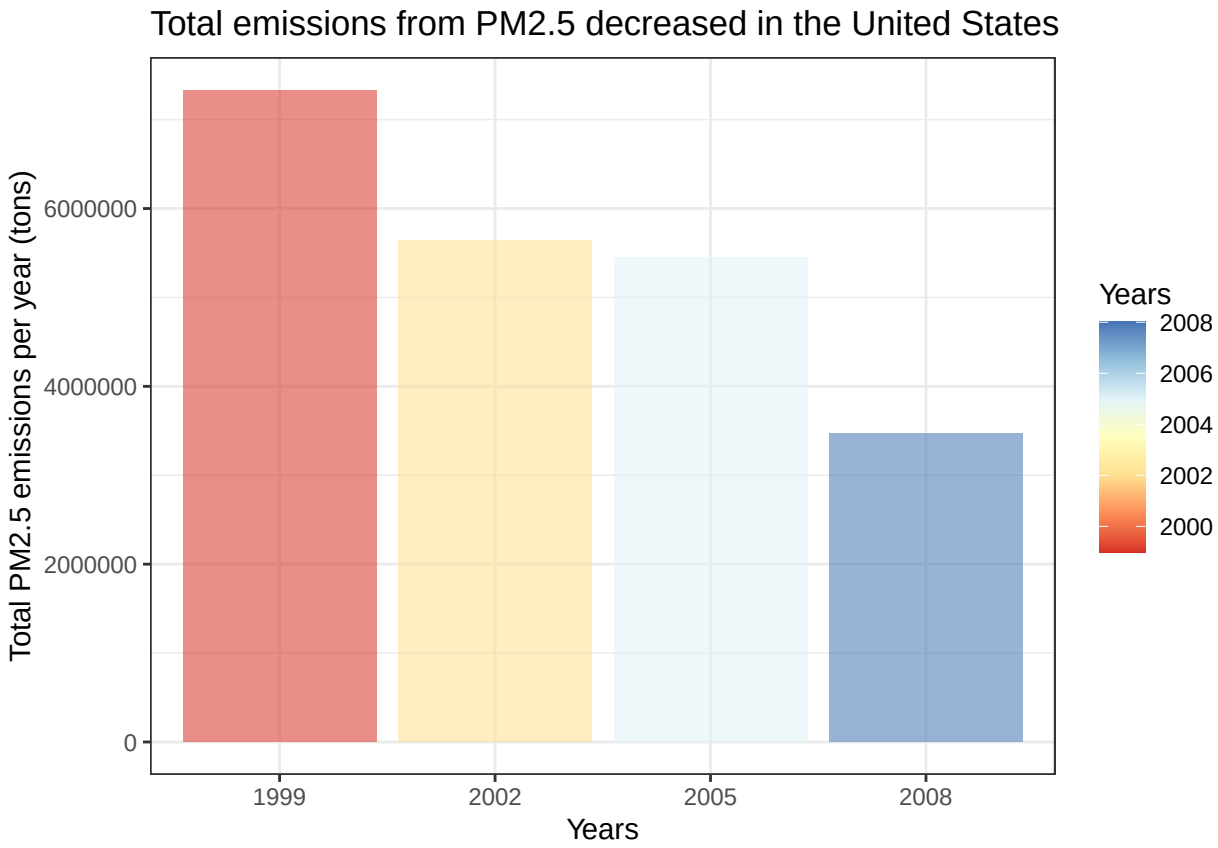
A tibble: 4 x 2

year Total 1 1999 7332967. 2 2002 5635780. 3 2005 5454703. 4 2008 3464206.

```r
# Plot emission using base plotting system


png("Plot1A.png", width = 1000, height = 800, units = 'px')
barplot(Emiss.tot$Total, Emiss.tot$year,
        col = c("lightblue", "mistyrose","lightcyan", "lavender"),
        log = 'y', ylab = 'Total PM2.5 emissions per year (tons)',
        xlab = "Years", names.arg = Emiss.tot$year,
        main = "Total emissions from PM2.5 decreased in the United States")
dev.off()
```

pdf 2

```r
# Plot emission using ggplot2

plot1B <- ggplot(data = Emiss.tot)+
 geom_col(aes(x = as.factor(year), y = Total, fill = year, alpha = 0.9))+
 scale_y_continuous(labels=function(n){format(n, scientific = FALSE)})+
 labs(x = 'Years', y = 'Total PM2.5 emissions per year (tons)',
      title = 'Total emissions from PM2.5 decreased in the United States')+
 scale_fill_distiller(name = 'Years',palette = "RdYlBu", direction = 1)+
 guides(alpha = 'none')+
 theme_bw()

plot1B
```




```
ggsave('Plot1B.png', plot = plot1B, width = 10, height = 7, units = 'cm')
```

Create Plot1.R

```
#library(knitr)
#purl('Plot1.Rmd')
```