Exploratory Data Analysis Project 2

Ceyda

14 09 2021

Loading Packages and Data:

```
library(RColorBrewer)
library(dplyr)
library(ggplot2)
library(rmarkdown)

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

Sys.setlocale("LC_TIME", "English")
```

```
## [1] "English_United States.1252"
```

```
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
```

```
head(SCC)
```

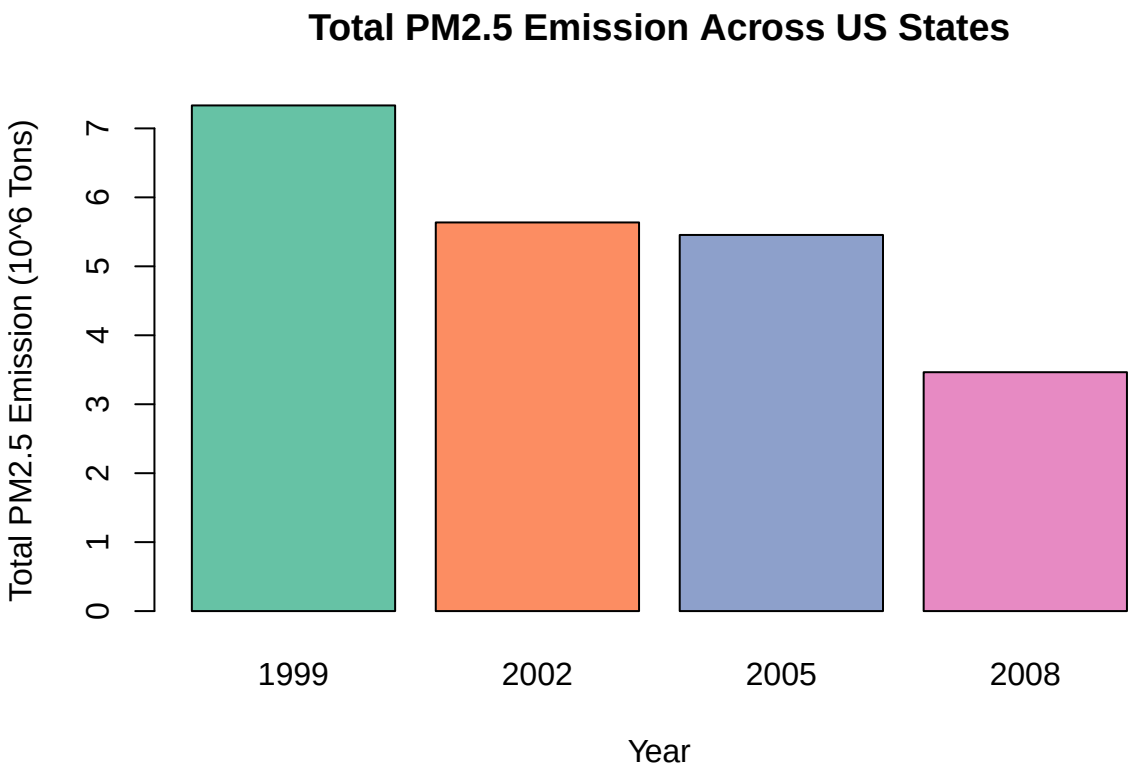
```
##        SCC Data.Category
## 1 10100101         Point
## 2 10100102         Point
## 3 10100201         Point
## 4 10100202         Point
## 5 10100203         Point
## 6 10100204         Point
##                                                                Short.Name
## 1                    Ext Comb /Electric Gen /Anthracite Coal /Pulverized Coal
## 2 Ext Comb /Electric Gen /Anthracite Coal /Traveling Grate (Overfeed) Stoker
## 3       Ext Comb /Electric Gen /Bituminous Coal /Pulverized Coal: Wet Bottom
## 4       Ext Comb /Electric Gen /Bituminous Coal /Pulverized Coal: Dry Bottom
## 5                    Ext Comb /Electric Gen /Bituminous Coal /Cyclone Furnace
```

```
## 6                   Ext Comb /Electric Gen /Bituminous Coal /Spreader Stoker
##                                 EI.Sector Option.Group Option.Set
## 1 Fuel Comb - Electric Generation - Coal
## 2 Fuel Comb - Electric Generation - Coal
## 3 Fuel Comb - Electric Generation - Coal
## 4 Fuel Comb - Electric Generation - Coal
## 5 Fuel Comb - Electric Generation - Coal
## 6 Fuel Comb - Electric Generation - Coal
##                 SCC.Level.One       SCC.Level.Two               SCC.Level.Three
## 1 External Combustion Boilers Electric Generation               Anthracite Coal
## 2 External Combustion Boilers Electric Generation               Anthracite Coal
## 3 External Combustion Boilers Electric Generation Bituminous/Subbituminous Coal
## 4 External Combustion Boilers Electric Generation Bituminous/Subbituminous Coal
## 5 External Combustion Boilers Electric Generation Bituminous/Subbituminous Coal
## 6 External Combustion Boilers Electric Generation Bituminous/Subbituminous Coal
##                                  SCC.Level.Four Map.To Last.Inventory.Year
## 1                                Pulverized Coal     NA                  NA
## 2             Traveling Grate (Overfeed) Stoker     NA                  NA
## 3 Pulverized Coal: Wet Bottom (Bituminous Coal)     NA                  NA
## 4 Pulverized Coal: Dry Bottom (Bituminous Coal)     NA                  NA
## 5             Cyclone Furnace (Bituminous Coal)     NA                  NA
## 6             Spreader Stoker (Bituminous Coal)     NA                  NA
##   Created_Date Revised_Date Usage.Notes
## 1
## 2
## 3
## 4
## 5
## 6
```

Question 1 (plot1.R)

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_emission <- aggregate(Emissions ~ year, NEI, FUN = sum)

coul <- brewer.pal(5, "Set2")
barplot(
    (height = total_emission$Emissions)/10^6,
    names.arg = total_emission$year,
    col= coul,
    main = "Total PM2.5 Emission Across US States",
    xlab = "Year",
    ylab = "Total PM2.5 Emission (10^6 Tons)"
)
```




```
dev.copy(png, file="plot1.png", height=480, width=640)
```

```
## png
##   3
```

```
dev.off()
```

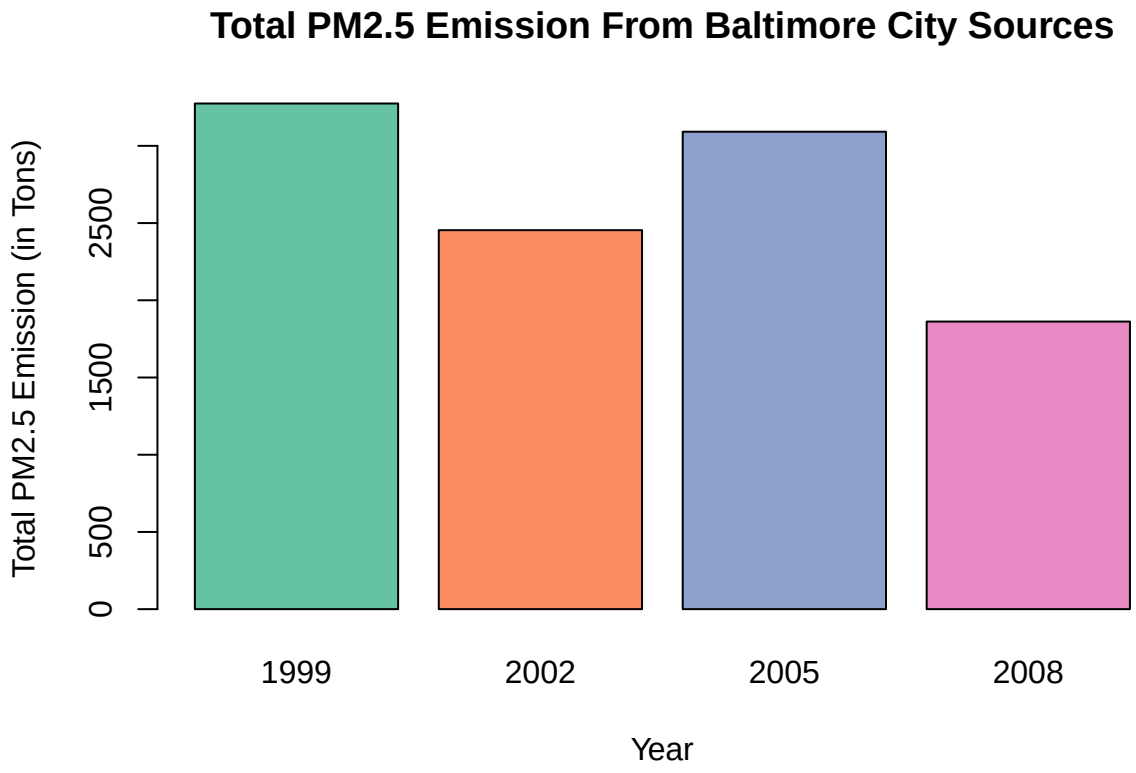
```
## pdf
##   2
```

Question 2 (plot2.R)

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.

```
total_emission_baltimore <- NEI %>%
    filter(fips == "24510")

total_emission <- aggregate(Emissions ~ year, total_emission_baltimore, FUN = sum)

coul <- brewer.pal(5, "Set2")
barplot(
    height = total_emission$Emissions,
    names.arg = total_emission$year,
```

```
    col= coul,
    main = "Total PM2.5 Emission From Baltimore City Sources",
    xlab = "Year",
    ylab = "Total PM2.5 Emission (in Tons)"
)
```




```
dev.copy(png, file="plot2.png", height=480, width=640)
```

```
## png
##   3
```

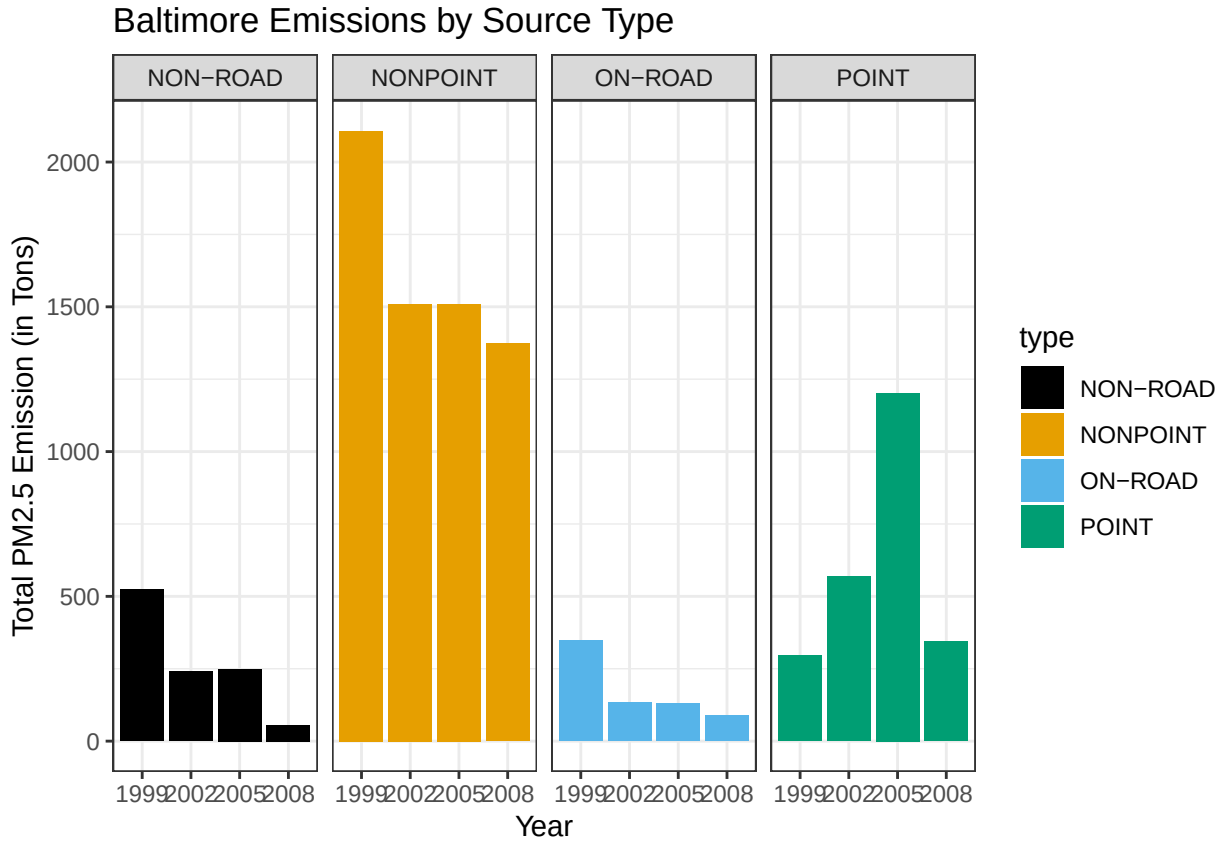
```
dev.off()
```

```
## pdf
##   2
```

Question 3 (plot3.R)

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.

```
total_emission_baltimore <- NEI %>%
    filter(fips == "24510") %>%
    group_by(year, type) %>%
    summarise(baltimore_emission = sum(Emissions))

cbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73")
BaltimoreEmission <- ggplot(data = total_emission_baltimore,
                            aes(x = factor(year), y= baltimore_emission,
                            fill = type, colore = "black")) +
    geom_bar(stat = "identity") +
    scale_fill_manual(values=cbPalette) +
    facet_grid(. ~ type) +
    xlab("Year") +
    ylab("Total PM2.5 Emission (in Tons)") +
    ggtitle("Baltimore Emissions by Source Type") +
    theme_bw()

print(BaltimoreEmission)
```




```
dev.copy(png, file="plot3.png", height=480, width=640)
```

```
## png
##   3
```

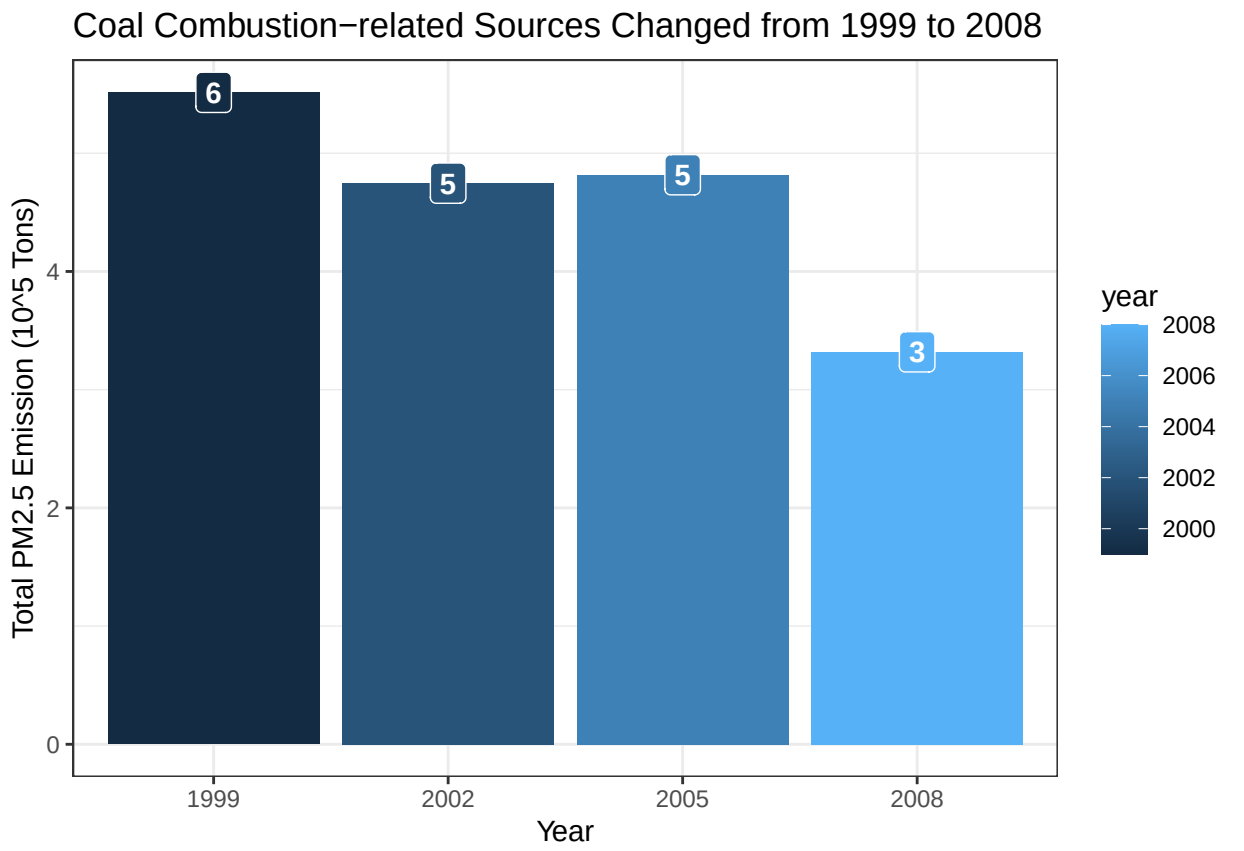
```
dev.off()
```

```
## pdf
##   2
```

Question 4 (plot4.R)

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

```
combustionRelated <- grepl("comb", SCC$SCC.Level.One, ignore.case=TRUE)
coalRelated <- grepl("coal", SCC$SCC.Level.Four, ignore.case=TRUE)
coalCombustion <- (combustionRelated & coalRelated)
combustionSCC <- SCC[coalCombustion,]$SCC
combustionNEI <- NEI[NEI$SCC %in% combustionSCC,]

coalcomb_emissions_sum <- combustionNEI %>%
    group_by(year) %>%
    summarise(coalcomb_emission = sum(Emissions))

CoalCombEmission <- ggplot(data = coalcomb_emissions_sum,
                           aes(x = factor(year), y= coalcomb_emission/10^5,
                               fill = year, label = round(coalcomb_emission/10^5))) +
    geom_bar(stat = "identity") +
    geom_label(aes(fill = year),colour = "white", fontface = "bold") +
    xlab("Year") +
    ylab("Total PM2.5 Emission (10^5 Tons)") +
    ggtitle("Coal Combustion-related Sources Changed from 1999 to 2008") +
    theme_bw()

print(CoalCombEmission)
```




```
dev.copy(png, file="plot4.png", height=480, width=640)
```

```
## png
##   3
```

```
dev.off()
```

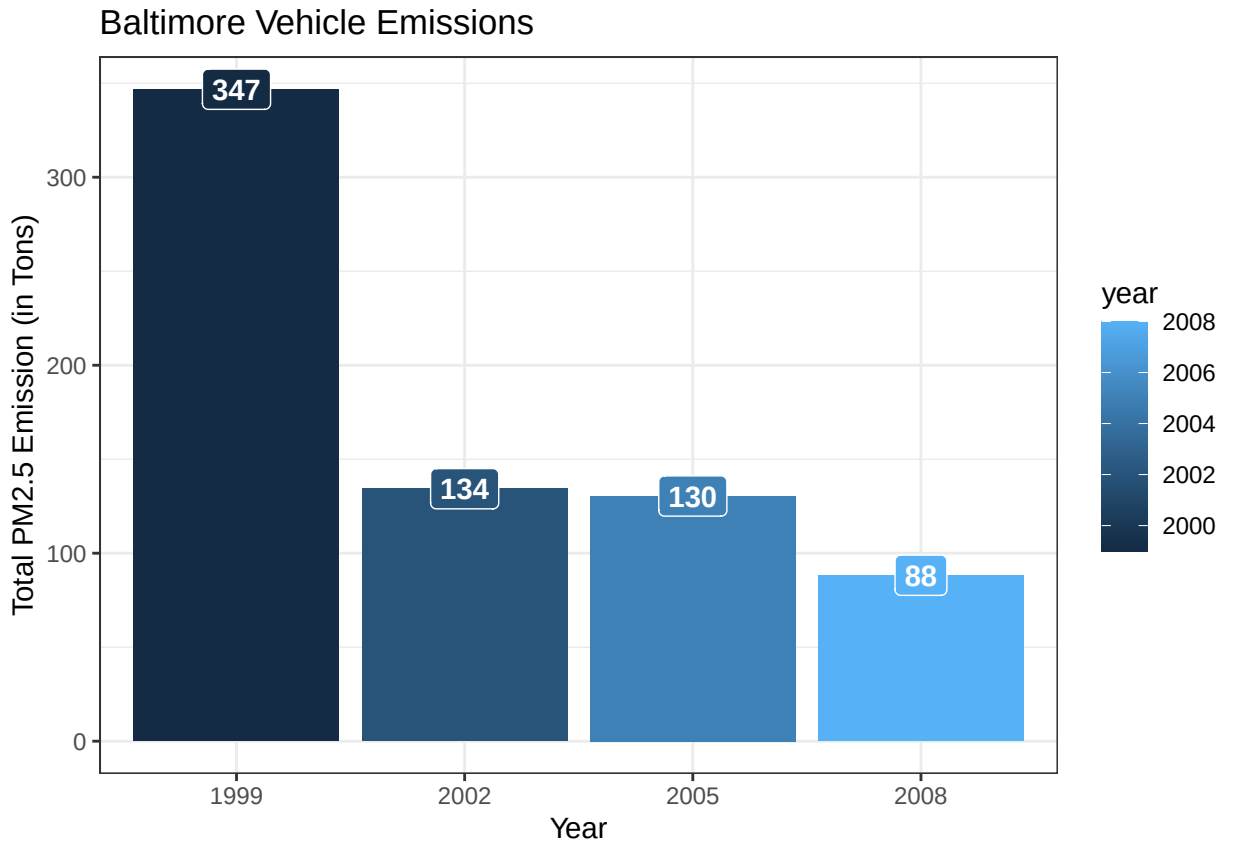
```
## pdf
##   2
```

Question 5 (plot5.R)

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

```
motorsources <- SCC[grep("[Vv]ehicle", SCC$EI.Sector), "SCC"]
vehicle_emissions <- NEI %>%
    filter(SCC %in% motorsources & fips == "24510") %>%
    group_by(year) %>%
    summarise(BaltimoreVehicleEmissions = sum(Emissions))

BaltimoreEmission <- ggplot(data = vehicle_emissions,
                            aes(x = factor(year), y = BaltimoreVehicleEmissions,
                                fill = year, label = round(BaltimoreVehicleEmissions))) +
    geom_bar(stat = "identity") +
```

```
    geom_label(aes(fill = year),colour = "white", fontface = "bold") +
    xlab("Year") +
    ylab("Total PM2.5 Emission (in Tons)") +
    ggtitle("Baltimore Vehicle Emissions") +
    theme_bw()

print(BaltimoreEmission)
```




```
dev.copy(png, file="plot5.png", height=480, width=640)
```

```
## png
##   3
```

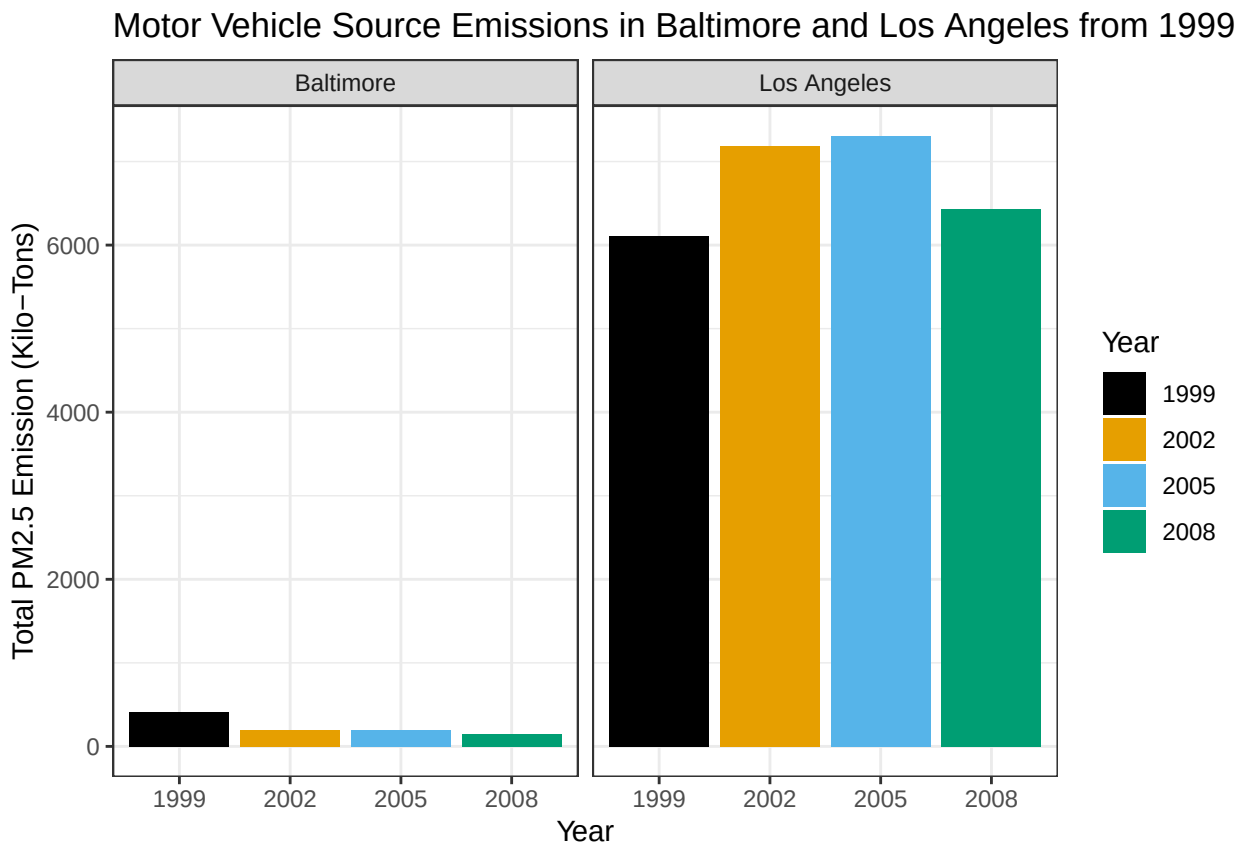
```
dev.off()
```

```
## pdf
##   2
```

Question 6 (plot6.R)

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?

```r
vehicles <- grepl("vehicle", SCC$SCC.Level.Two, ignore.case=TRUE)
vehiclesSCC <- SCC[vehicles,]$SCC
vehiclesNEI <- NEI[NEI$SCC %in% vehiclesSCC,]

baltimore <- vehiclesNEI %>%
    filter(fips == "24510") %>%
    group_by(year) %>%
    mutate(city = "Baltimore")

los_angeles <- vehiclesNEI %>%
    filter(fips == "06037") %>%
    group_by(year) %>%
    mutate(city = "Los Angeles")

bothcities <- rbind(baltimore,los_angeles)

cbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73")
BaltimoreLosAngelesEmission <- ggplot(data = bothcities,
                           aes(x = factor(year), y= Emissions,
                               fill = factor(year))) +
    geom_bar(stat = "identity") +
    scale_fill_manual(values=cbPalette) +
    facet_grid(scales="free", space="free", .~city) +
    xlab("Year") +
    ylab("Total PM2.5 Emission (Kilo-Tons)") +
    labs(fill="Year") +
    ggtitle("Motor Vehicle Source Emissions in Baltimore and Los Angeles from 1999 to 2008") +
    theme_bw()

print(BaltimoreLosAngelesEmission)
```




```
dev.copy(png, file="plot6.png", height=480, width=640)
```

```
## png
##   3
```

```
dev.off()
```

```
## pdf
##   2
```