Exploratory Analysis Project

Siti Salwani Yaacob

1/19/2020

Assignment

The overall goal of this assignment is to explore the National Emissions Inventory database and see what it say about fine particulate matter pollution in the United states over the 10-year period 1999–2008. You may use any R package you want to support your analysis.

Questions

You must address the following questions and tasks in your exploratory analysis. For each question/task you will need to make a single plot. Unless specified, you can use any plotting system in R to make your plot.

(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?

Data Preparation

Data is now downloaded and extracted. Now, load the .rds data and assign into 'SummarySCC_PM25' and 'Source_Classification_Code' dataframe.

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

View the imported data.

```
head(summarySCC_PM25)
```

```
##     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
```

View the imported data

```
head(Source_Classification_Code)
```

```
##        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
## 1 External Combustion Boilers Electric Generation
## 2 External Combustion Boilers Electric Generation
## 3 External Combustion Boilers Electric Generation
## 4 External Combustion Boilers Electric Generation
## 5 External Combustion Boilers Electric Generation
## 6 External Combustion Boilers Electric Generation
##                 SCC.Level.Three
## 1               Anthracite Coal
## 2               Anthracite Coal
## 3 Bituminous/Subbituminous Coal
## 4 Bituminous/Subbituminous Coal
## 5 Bituminous/Subbituminous Coal
## 6 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
```

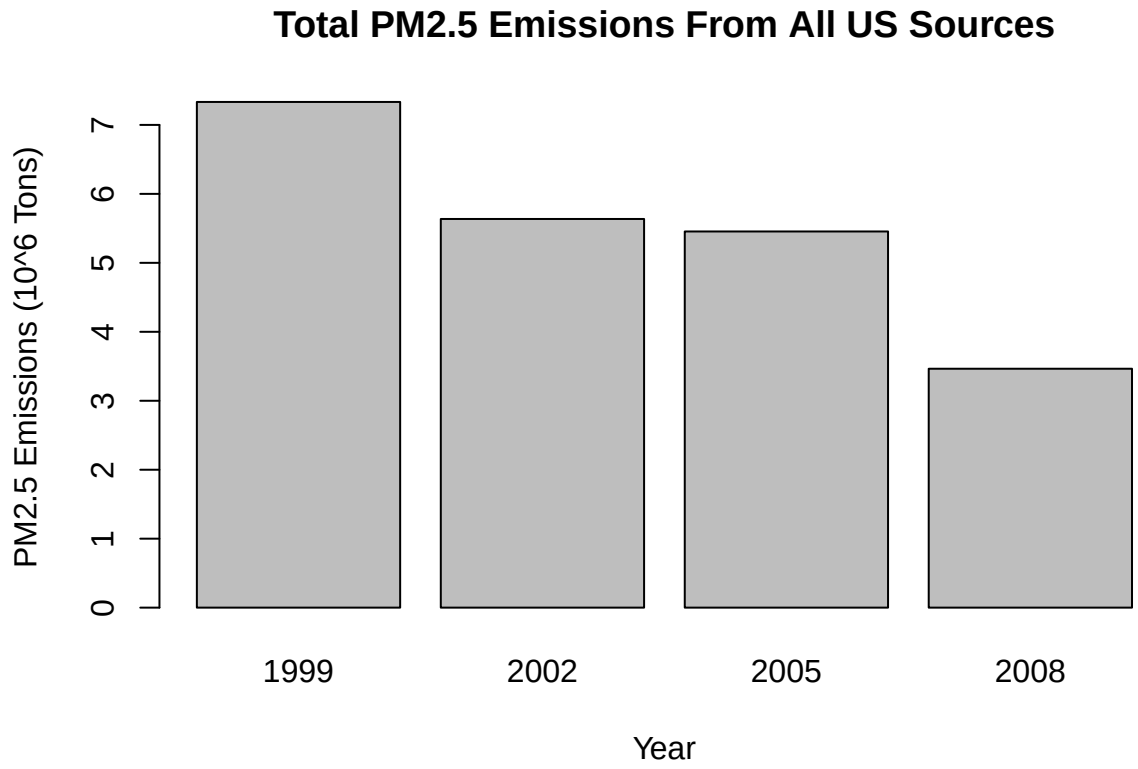
Question 1 : Have total emissions from PM2.5 decreased in the United States from 1999 to 2008? Using the base plotting system, make a plot showing the total PM2.5 emission from all sources for each of the years 1999, 2002, 2005, and 2008.

First we will aggregate the total PM2.5 emission from all sources for each the years 1999, 2002, 2005 and 2008.

```
aggregate_total <- aggregate(Emissions ~ year, summarySCC_PM25, sum)
```

Using the base plotting system, now we plot the total PM2.5 Emission from all sources,

```
barplot(
  (aggregate_total$Emissions)/10^6,
  names.arg=aggregate_total$year,
  xlab="Year",
  ylab="PM2.5 Emissions (10^6 Tons)",
  main="Total PM2.5 Emissions From All US Sources"
)
```




```
dev.copy(png,'plot1.png')
```

```
## quartz_off_screen
##                 3
```

```
dev.off()
```

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

Question : Have total emissions from PM2.5 decreased in the United States from 1999 to 2008?
As we can see from the plot, total emissions have decreased in the US from 1999 to 2008.

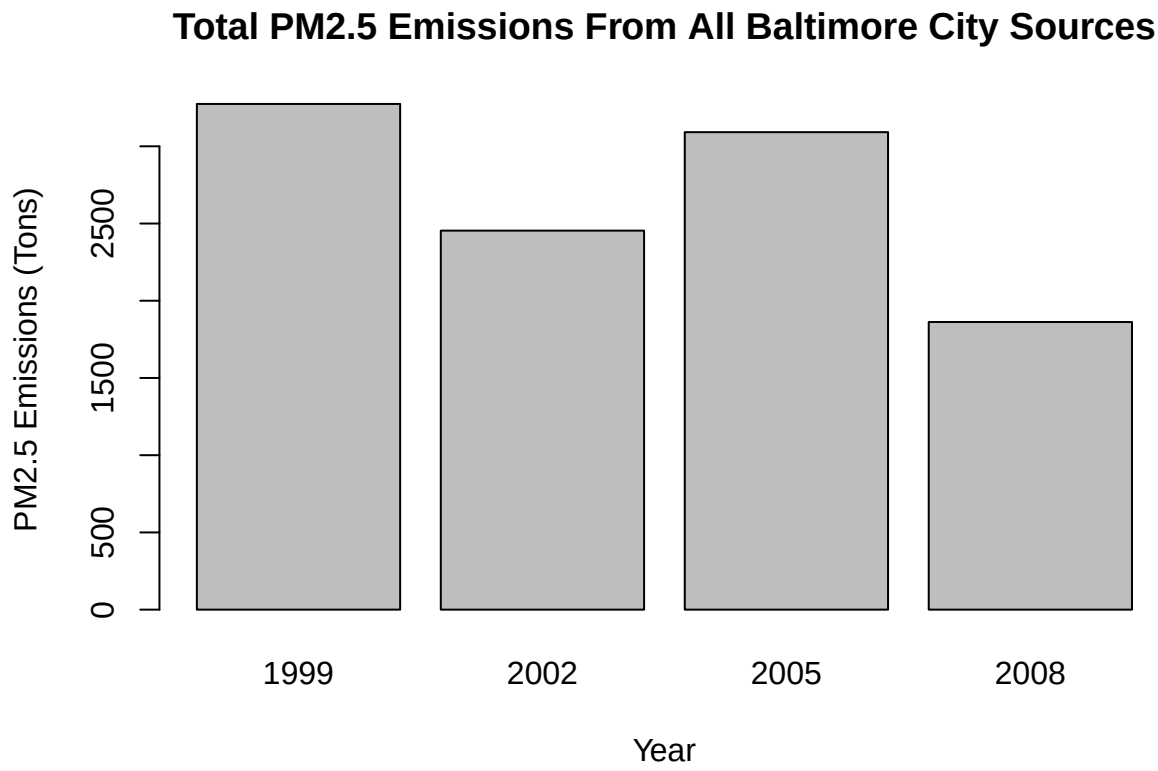
Questions 2 : Have total emissions from PM2.5 decreased in the Baltimore City, Maryland (fips == "24510") from 1999 to 2008? Use the base plotting system to make a plot answering this question.

Subset the data for fips == "24510" and then aggregate them by summing the Emissions per years

```
baltimore_summarySCC_PM25 <- summarySCC_PM25[summarySCC_PM25$fips=="24510",]
aggregate_total_baltimore <- aggregate(Emissions ~ year, baltimore_summarySCC_PM25,sum)
```

Now we use the base plotting system to make a plot of this data

```
barplot(
  aggregate_total_baltimore$Emissions,
  names.arg=aggregate_total_baltimore$year,
  xlab="Year",
  ylab="PM2.5 Emissions (Tons)",
  main="Total PM2.5 Emissions From All Baltimore City Sources"
)
```




```
dev.copy(png,'plot2.png')
```

```
## quartz_off_screen
##                 3
```

```
dev.off()
```

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

Have total emissions from PM2.5 decreased in Baltimore City, Maryland from 1999 to 2008? Overall total emissions from PM2.5 have decreased in Baltimore City, Maryland from 1999 to 2008.

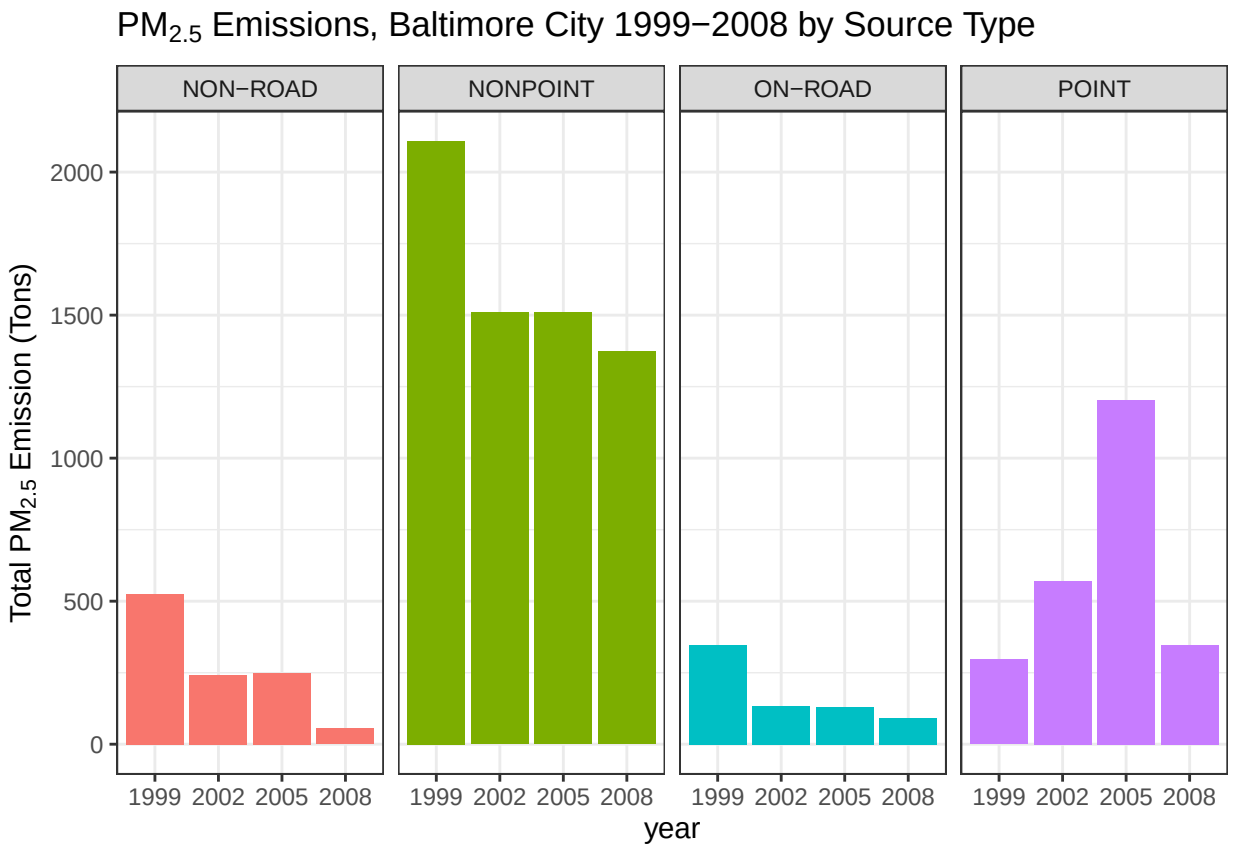
Question 3 : Of the four types of sources indicated by the type (point, nonpoint, onroad, nonroad) variable, which of these four sources have seen decreases in emissions from 1999–2008 for Baltimore City? Which have seen increases in emissions from 1999–2008? Use the ggplot2 plotting system to make a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

Using ggplot2 plotting system.

```
library(ggplot2)

plotting_baltimore <- ggplot(baltimore_summarySCC_PM25,aes(factor(year),Emissions,fill=type)) +
  geom_bar(stat="identity") +
  theme_bw() + guides(fill=FALSE)+
  facet_grid(.~type,scales = "free",space="free") +
  labs(x="year", y=expression("Total PM"[2.5]*" Emission (Tons)")) +
  labs(title=expression("PM"[2.5]*" Emissions, Baltimore City 1999-2008 by Source Type"))

print(plotting_baltimore)
```




```
dev.copy(png,'plot3.png')
```

```
## quartz_off_screen
##                 3
```

```
dev.off()
```

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

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?

The non-road, nonpoint, on-road sorces types have all seen decreased emissions overall from 1999-2008 in Baltimore City.

Which have seen increases in emissions from 1999-2008?
The point sources saw a slight increase overall from 1999-2008. Also note that the point source saw a significant increase until 2005 at which point it decreases again by 2008 to just above the starting values.

Question 4 : Across the United States, how have emissions from coal combustion-related sources changed from 1999–2008?

First we subset coal combustion source factors NEI data.

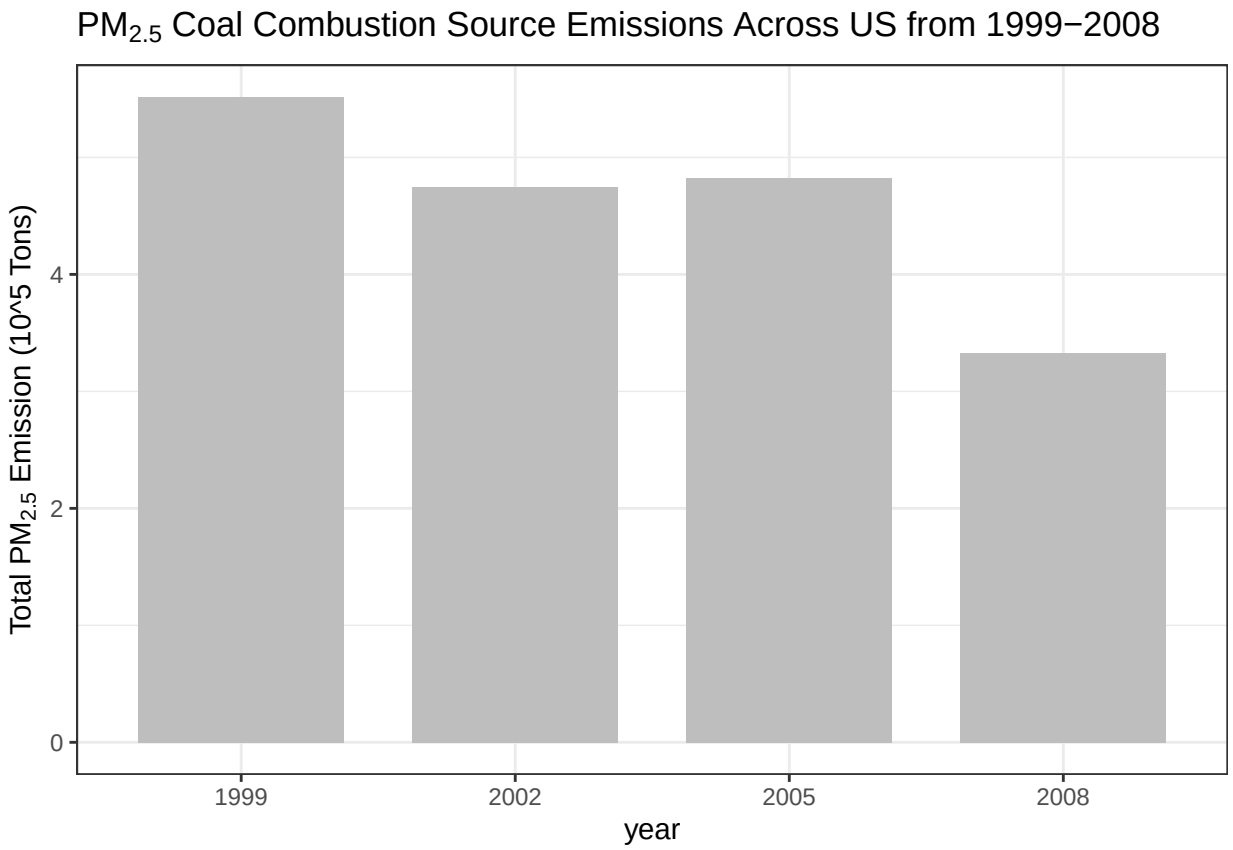
```
# Subset coal combustion related NEI data
combustion_related <- grepl("comb", Source_Classification_Code$SCC.Level.One, ignore.case=TRUE)
coal_related <- grepl("coal", Source_Classification_Code$SCC.Level.Four, ignore.case=TRUE)
coal_combustion <- (combustion_related & coal_related)
combustion_Source_Classification_Code <- Source_Classification_Code[coal_combustion,]$SCC
combustion_summarySCC_PM25 <- summarySCC_PM25[summarySCC_PM25$SCC %in% combustion_Source_Classification_
```

Note: The SCC levels go from generic to specific. We assume that coal combustion related SCC records are those where SCC.Level.One contains the subctring 'comb' and SCC.Level.Four contains the substring 'coal'.

```
library(ggplot2)

emission_changed_coal <- ggplot(combustion_summarySCC_PM25,aes(factor(year),Emissions/10^5)) +
  geom_bar(stat="identity",fill="grey",width=0.75) +
  theme_bw() +  guides(fill=FALSE) +
  labs(x="year", y=expression("Total PM"[2.5]*" Emission (10^5 Tons)")) +
  labs(title=expression("PM"[2.5]*" Coal Combustion Source Emissions Across US from 1999-2008"))

print(emission_changed_coal)
```




```
dev.copy(png,'plot4.png')
```

```
## quartz_off_screen
##                 3
```

```
dev.off()
```

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

Accross the United States, how have emissions from coal combustion-related sources changed from 1999-2008? Emission from coal combustion related sources have decreased from 6*10^6 to below 4*10^6 from 1999-2008. Eg. Emissions from coal combustion related sources have decreased about 1/3 from 1999-2008.

Question 5 : How have emissions from motor vehicle sources changed from 1999–2008 in Baltimore City?

First we subset the motor vehicles, which we assume is anything like Motor Vehicle in SCC.Level.Two.

```
library(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
```

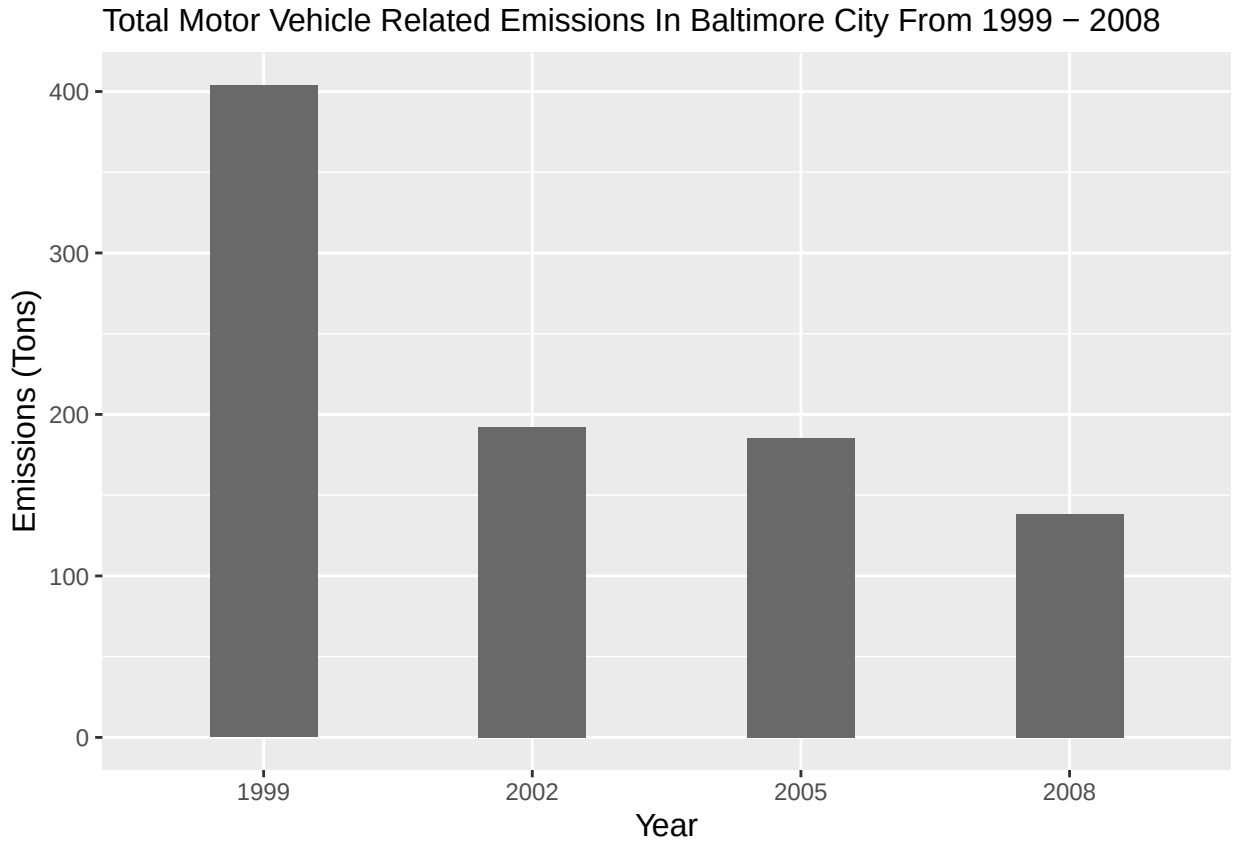
```
vehicle_Source_Classification_Code <- Source_Classification_Code %>%
  filter(grepl('[Vv]ehicle', SCC.Level.Two)) %>%
  select(SCC, SCC.Level.Two)

total_emission_24510_vehicle <- summarySCC_PM25 %>%
  filter(fips == "24510") %>%
  select(SCC, fips, Emissions, year) %>%
  inner_join(vehicle_Source_Classification_Code, by = "SCC") %>%
  group_by(year) %>%
  summarise(Total_Emissions = sum(Emissions, na.rm = TRUE)) %>%
  select(Total_Emissions, year)
```

```
## Warning: Column `SCC` joining character vector and factor, coercing into
## character vector
```

```
baltimore_vehicle_plot <- ggplot(total_emission_24510_vehicle, aes(factor(year), Total_Emissions)) +
  geom_bar(stat = "identity", fill = "dimgray", width = 0.4) +
  labs(x = "Year", y = "Emissions (Tons)", title = "Total Motor Vehicle Related Emissions In Baltimore C
  theme(plot.title = element_text(size = 12),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12)) +
  ggsave("plot5.png", width = 30, height = 30, units = "cm")

print(baltimore_vehicle_plot)
```




```
dev.copy(png,'plot5.png')
```

```
## quartz_off_screen
##                 3
```

```
dev.off()
```

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

How have emissions from motor vehicle sources changed from 1999-2008 in Baltimore City?
Emissions from motor vehicle sources have dropped from 1999-2008 in Baltimore City

Question 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?

Comparing emissions from motor vehicle sources in Baltimore City (fips==“24510”) with emissions from motor vehicle sources in Los Angeles County, California (fips==“06037”)

```
library(dplyr)

vehicles_Source_Classification_Code <- Source_Classification_Code %>%
  filter(grepl('[Vv]ehicle', SCC.Level.Two)) %>%
  select(SCC, SCC.Level.Two)

total_emission_two_location <- summarySCC_PM25 %>%
  filter(fips == "24510" | fips == "06037") %>%
  select(fips, SCC, Emissions, year) %>%
  inner_join(vehicles_Source_Classification_Code, by = "SCC") %>%
  group_by(fips, year) %>%
  summarise(Total_Emissions = sum(Emissions, na.rm = TRUE)) %>%
  select(Total_Emissions, fips, year)
```

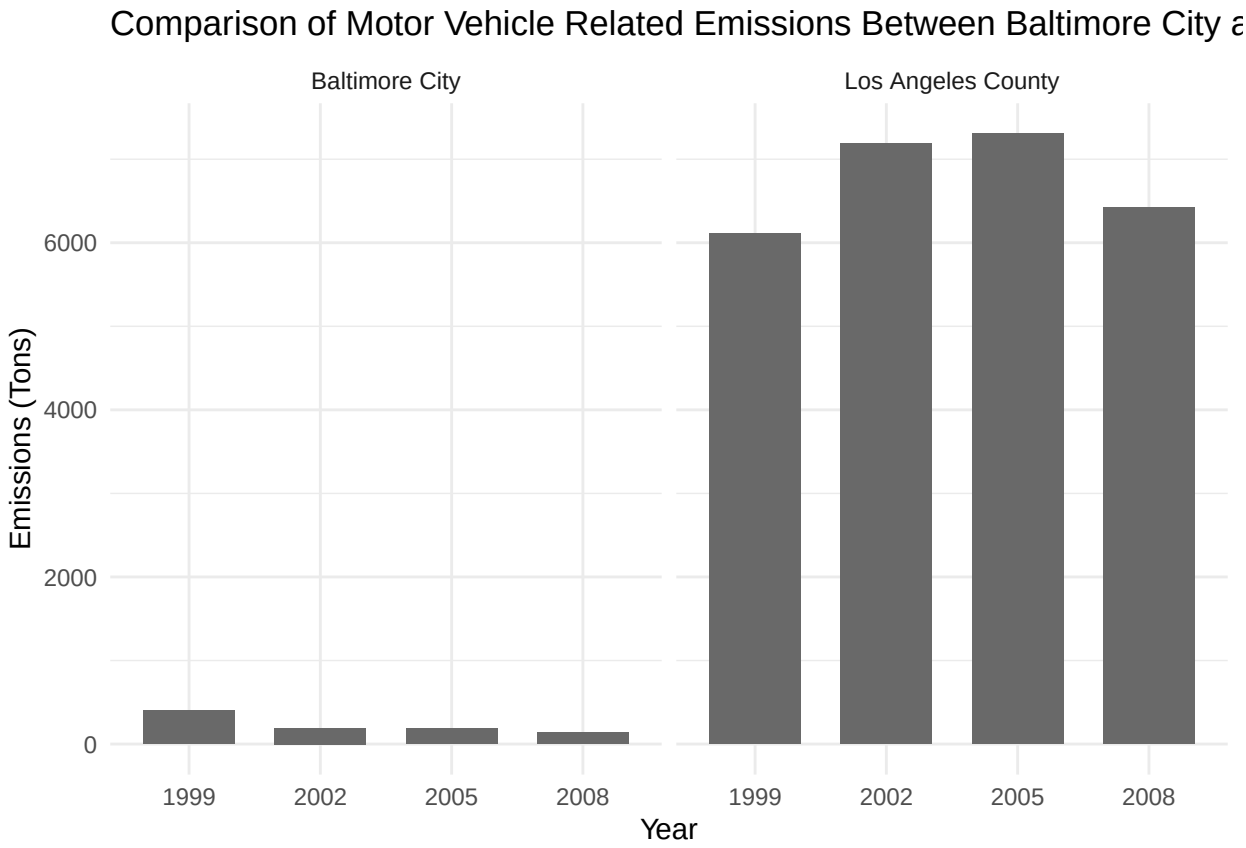
```
## Warning: Column `SCC` joining character vector and factor, coercing into
## character vector
```

Now we plot using the ggplot2 system,

```
total_emission_two_location$fips <- gsub("24510", "Baltimore City", total_emission_two_location$fips)
total_emission_two_location$fips <- gsub("06037", "Los Angeles County", total_emission_two_location$fips
```

```
plotting_two_location <- ggplot(total_emission_two_location, aes(x = factor(year), y = Total_Emissions,
  geom_bar(stat = "identity", fill = "dimgray", width = 0.7) +
  facet_grid(.~fips) +
  labs(x = "Year", y = "Emissions (Tons)", title = "Comparison of Motor Vehicle Related Emissions Betwe
  theme(plot.title = element_text(size = 14),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        strip.text.x = element_text(size = 12)) +
  theme_minimal() +
  ggsave("plot6.png", width = 30, height = 30, units = "cm")

print(plotting_two_location)
```



Which city has seen greater changes over time in motor vehicle emissions?
Los Angeles County has seen the greatest changes over time in motor vehicle emissions.