Coursera - Exploratory Data Analysis - Project 2

Ali Magzari

09/07/2021

Description

This project uses National Emissions Inventory records to investigate particulate matter emissions across the United States over a 10-year period.

In an effort to answer the project questions through exploratory data analysis, practical use of commands in data manipulation and plotting is put in action.

https://www.coursera.org/learn/exploratory-data-analysis/peer/b5Ecl/course-project-2

Procedure

Load necessary packages

```
library(dplyr)
```

```
##
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':
##
##     filter, lag

## Les objets suivants sont masqués depuis 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

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.

Import data from rds file

```
file_name <- "summarySCC_PM25.rds"
data <- readRDS(file_name)
```

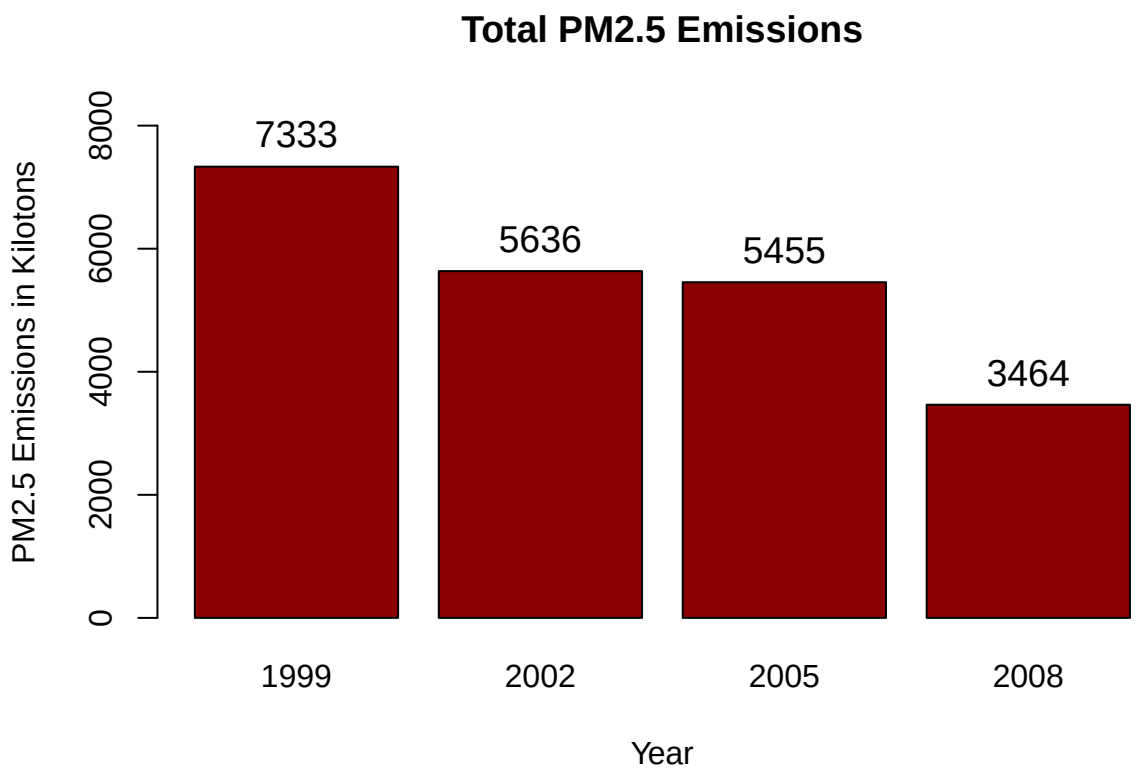
Calculate total PM2.5 emissions for each year

```
emi_year <- data %>% group_by(year) %>% summarise(total = sum(Emissions))
```

Plot the data accordingly

```
plot1 <- barplot(emi_year$total/1000, main = "Total PM2.5 Emissions",
                 xlab = "Year", ylab = "PM2.5 Emissions in Kilotons",
                 names.arg = emi_year$year, col = "darkred", ylim = c(0,8300))

text(plot1, round(emi_year$total/1000), label = round(emi_year$total/1000),
     pos = 3, cex = 1.2)
```



From the graph above, it is clear that PM2.5 emissions did decrease from 1999 to 2008 with an overall decrease of 53%.

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.

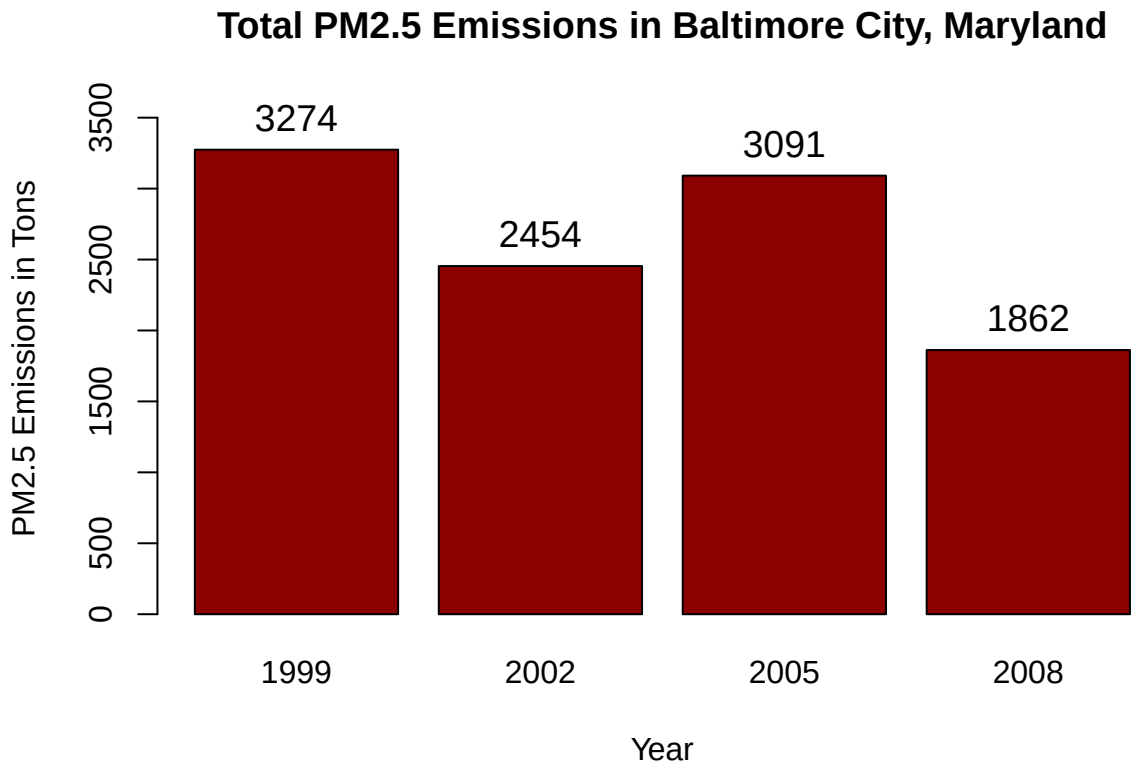
Calculate total emissions for Baltimore City

```
emi_balt <- data %>% group_by(year) %>% filter(fips == "24510") %>%
  summarise(total = sum(Emissions))
```

Plot Baltimore City emissions per year

```
plot2 <- barplot(emi_balt$total,
                 main = "Total PM2.5 Emissions in Baltimore City, Maryland",
                 xlab = "Year", ylab = "PM2.5 Emissions in Tons",
                 names.arg = emi_balt$year, col = "darkred", ylim = c(0,3600))

text(plot2, round(emi_balt$total), label = round(emi_balt$total),
     pos = 3, cex = 1.2)
```



Overall, Baltimore city PM2.5 emissions did decrease from 1999 to 2008 by 43%, although the year 2005 noticed a spike at 3091 tons.

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.

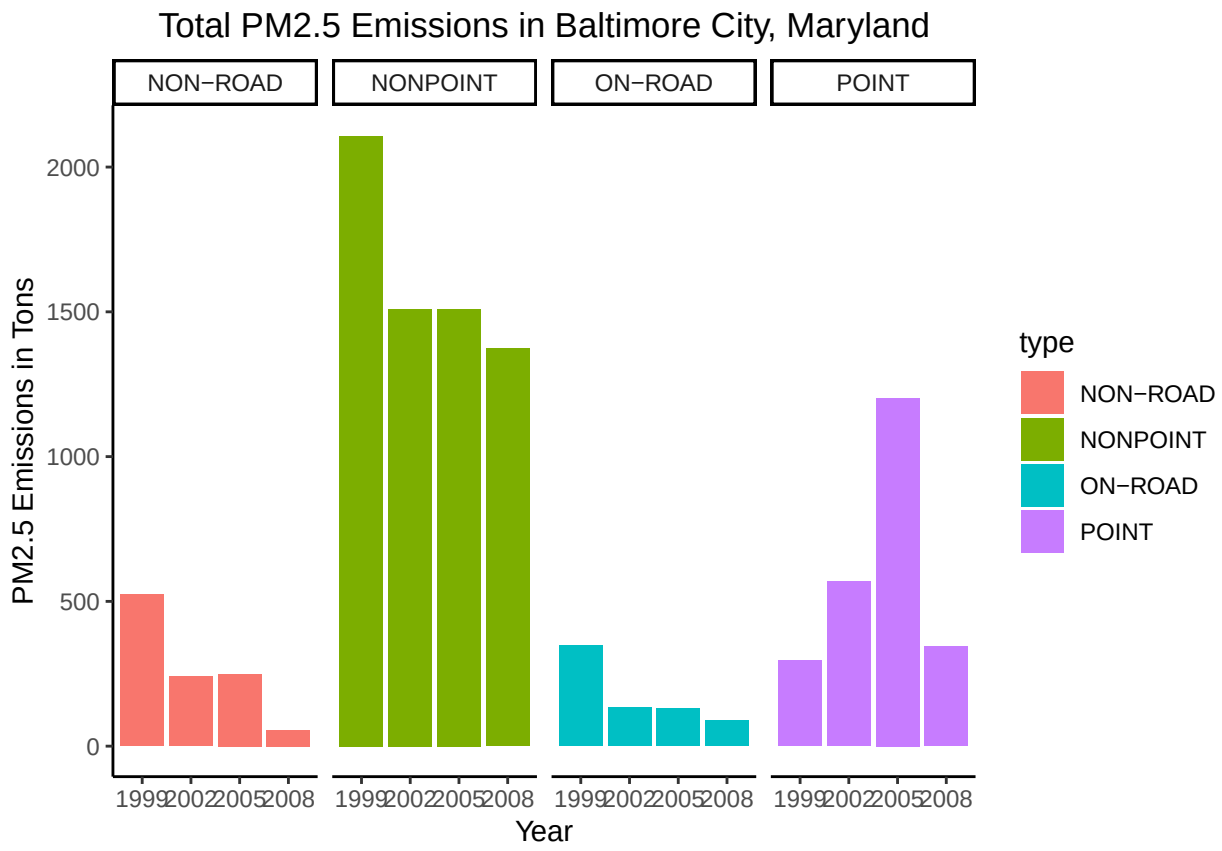
Calculate total emissions by year and source type for Baltimore City

```
emi_balt_t <- data %>% group_by(type, year) %>% filter(fips == "24510") %>%
  summarise(total = sum(Emissions))
```

```
## `summarise()` has grouped output by 'type'. You can override using the `.groups` argument.
```

Plot Baltimore City emissions by type for each year

```
ggplot(emi_balt_t, aes(x = factor(year),
                       y = total, fill = type, label = round(total))) +
  geom_bar(stat = "identity") + facet_grid(. ~ type) +
  ggtitle("Total PM2.5 Emissions in Baltimore City, Maryland") +
  xlab("Year")+ ylab("PM2.5 Emissions in Tons") +
  theme_classic() + theme(plot.title = element_text(hjust = 0.5))
```




```
  scale_fill_brewer(palette = "Set1")
```

```
## <ggproto object: Class ScaleDiscrete, Scale, gg>
##     aesthetics: fill
##     axis_order: function
##     break_info: function
##     break_positions: function
##     breaks: waiver
##     call: call
##     clone: function
##     dimension: function
##     drop: TRUE
##     expand: waiver
##     get_breaks: function
##     get_breaks_minor: function
##     get_labels: function
##     get_limits: function
##     guide: legend
```

```
##     is_discrete: function
##     is_empty: function
##     labels: waiver
##     limits: NULL
##     make_sec_title: function
##     make_title: function
##     map: function
##     map_df: function
##     n.breaks.cache: NULL
##     na.translate: TRUE
##     na.value: NA
##     name: waiver
##     palette: function
##     palette.cache: NULL
##     position: left
##     range: <ggproto object: Class RangeDiscrete, Range, gg>
##         range: NULL
##         reset: function
##         train: function
##         super:  <ggproto object: Class RangeDiscrete, Range, gg>
##     rescale: function
##     reset: function
##     scale_name: brewer
##     train: function
##     train_df: function
##     transform: function
##     transform_df: function
##     super:  <ggproto object: Class ScaleDiscrete, Scale, gg>
```

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

Import source classification code data

```
ssc_file <- "Source_Classification_Code.rds"
data_ssc <- readRDS(ssc_file)
```

Select coal combustion_related sources using keywords from the Ei.Sector column

```
data_coal <- data_ssc[grepl("Comb.*Coal", data_ssc$EI.Sector), ]
```

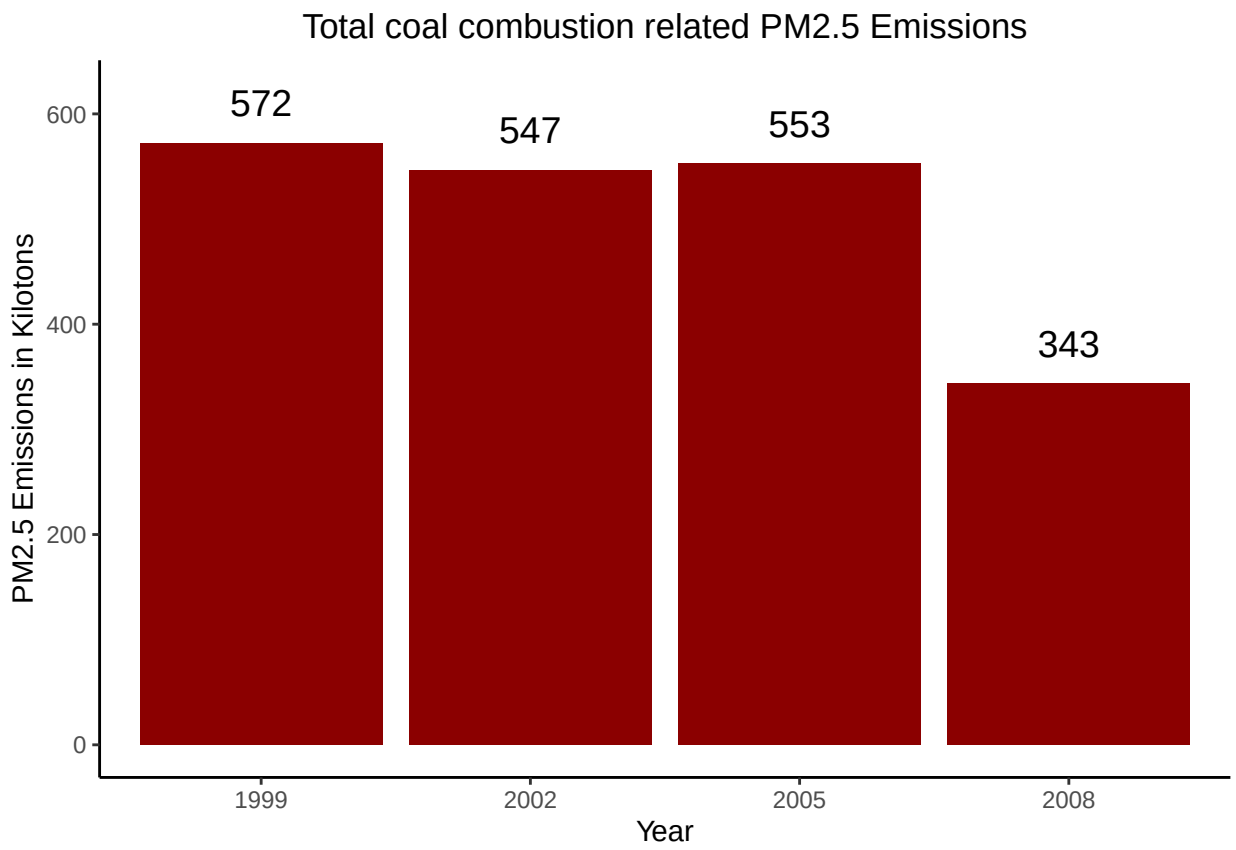
Calculate total coal combustion-related emissions

```
coal_scc <- unique(data_coal$SCC)
coal_emi <- data[(data$SCC %in% coal_scc), ]
coal_year <- coal_emi %>% group_by(year) %>% summarise(total = sum(Emissions))
```

Plot total coal combustion-related emissions for each year

```
ggplot(coal_year, aes(factor(year), total/1000, label = round(total/1000))) +
  geom_bar(stat = "identity", fill = "darkred") +
  ggtitle("Total coal combustion related PM2.5 Emissions") +
```

```
  xlab("Year") + ylab("PM2.5 Emissions in Kilotons") +
  ylim(c(0, 620)) + theme_classic()+ geom_text(size = 5, vjust = -1) +
  theme(plot.title = element_text(hjust = 0.5))
```



The graph above shows that combustion-related emissions decreased from 1999 to 2008 with an overall decrease of 40%, although they are quasi-equal for the years 1999, 2002 and 2005.

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

Select records related to motor vehicle sources

```
data_motor <- data_ssc[grepl("Vehicle", data_ssc$SCC.Level.Two), ]
```

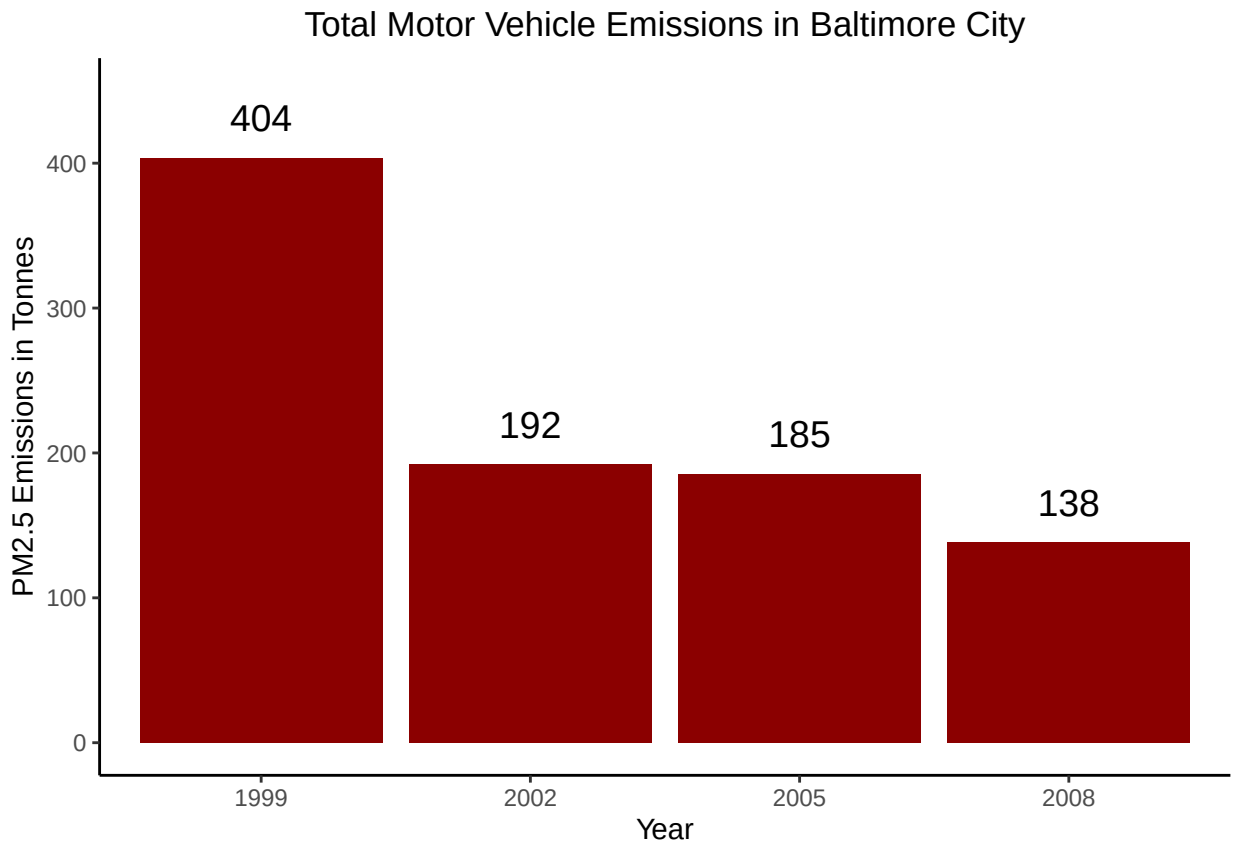
Calculate total emissions from motor vehicle sources in Baltimore City

```
motor_scc <- unique(data_motor$SCC)
motor_emi <- data[(data$SCC %in% motor_scc), ]
motor_year <- motor_emi %>% filter(fips == "24510") %>% group_by(year) %>%
  summarise(total = sum(Emissions))
```

Plot total emissions from motor vehicle sources in Baltimore City

```
ggplot(motor_year, aes(factor(year), total, label = round(total))) +
  geom_bar(stat = "identity", fill = "darkred") +
```

```
  ggtitle("Total Motor Vehicle Emissions in Baltimore City") +
  xlab("Year") + ylab("PM2.5 Emissions in Tonnes") +
  ylim(c(0, 450)) + theme_classic()+ geom_text(size = 5, vjust = -1) +
  theme(plot.title = element_text(hjust = 0.5))
```



The graph above shows that Baltimore City witnessed a clear decrease in emissions from motor vehicle sources (by 66%)

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?

Select motor vehicle emissions in Baltimore City and Los Angeles County

```
balti_la_year <- motor_emi %>% filter(fips == "24510" | fips == "06037") %>%
  group_by(fips, year) %>% summarise(total = sum(Emissions))
```

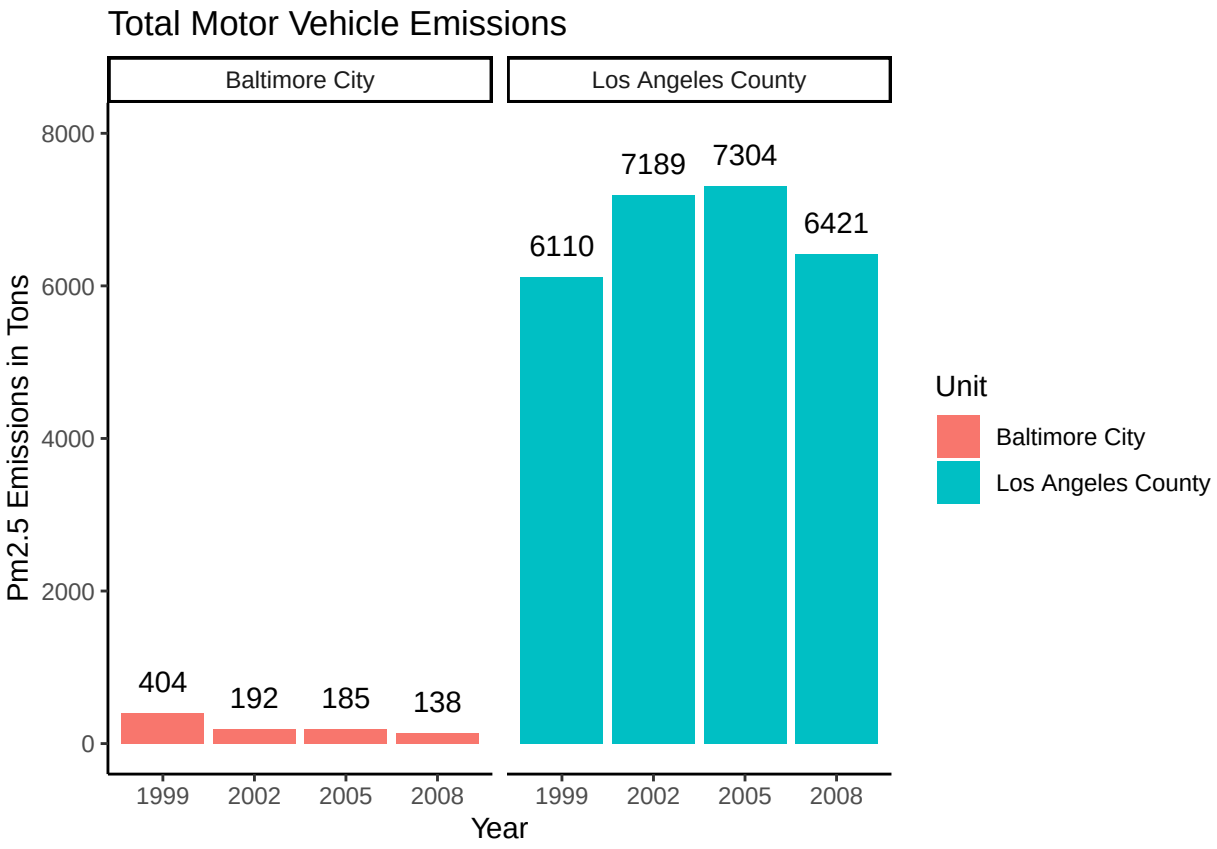
```
## `summarise()` has grouped output by 'fips'. You can override using the `.groups` argument.
```

Add a column referring unit (Baltimore City or LA County)

```
balti_la_year <- mutate(balti_la_year,
                    Unit = ifelse(fips == "24510", "Baltimore City",
                           ifelse(fips == "06037", "Los Angeles County")))
```

PLot total motor vehicle emissions in Baltimore City and Los Angeles County

```
ggplot(balti_la_year, aes(factor(year), total,
  fill = Unit, label = round(total))) +
  geom_bar(stat = "identity") + facet_grid(. ~ Unit) +
  ggtitle("Total Motor Vehicle Emissions") +
  xlab("Year") + ylab("Pm2.5 Emissions in Tons") +
  theme(plot.title = element_text(hjust = 0.5)) + ylim(c(0, 8000)) +
  theme_classic() + geom_text(size = 4, vjust = -1)
```



The graph above shows that while Baltimore City witnessed a decrease in motor vehicle emissions (MVE), Los Angeles County on the other hand shows an increase in MVEs with a staggering value of 6421 tons in 2008.