CourseProjecy-ExploratoryDataAanlysis

TianyiFang

August 29, 2017

Course Project 2 for Exploratory Data Analysis

Introduction

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. *fips: A five-digit number (represented as a string) indicating the U.S. county* SCC: The name of the source as indicated by a digit string (see source code classification table) *Pollutant: A string indicating the pollutant* Emissions: Amount of PM2.5 emitted, in tons *type: The type of source (point, non-point, on-road, or non-road)* year: The year of emissions recorded ###Analysis ####1.Preparing Data

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.1
```

```
library(lattice)
library(data.table)
library(gmodels)
```

```
## Warning: package 'gmodels' was built under R version 3.4.1
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.1
```

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

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

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

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

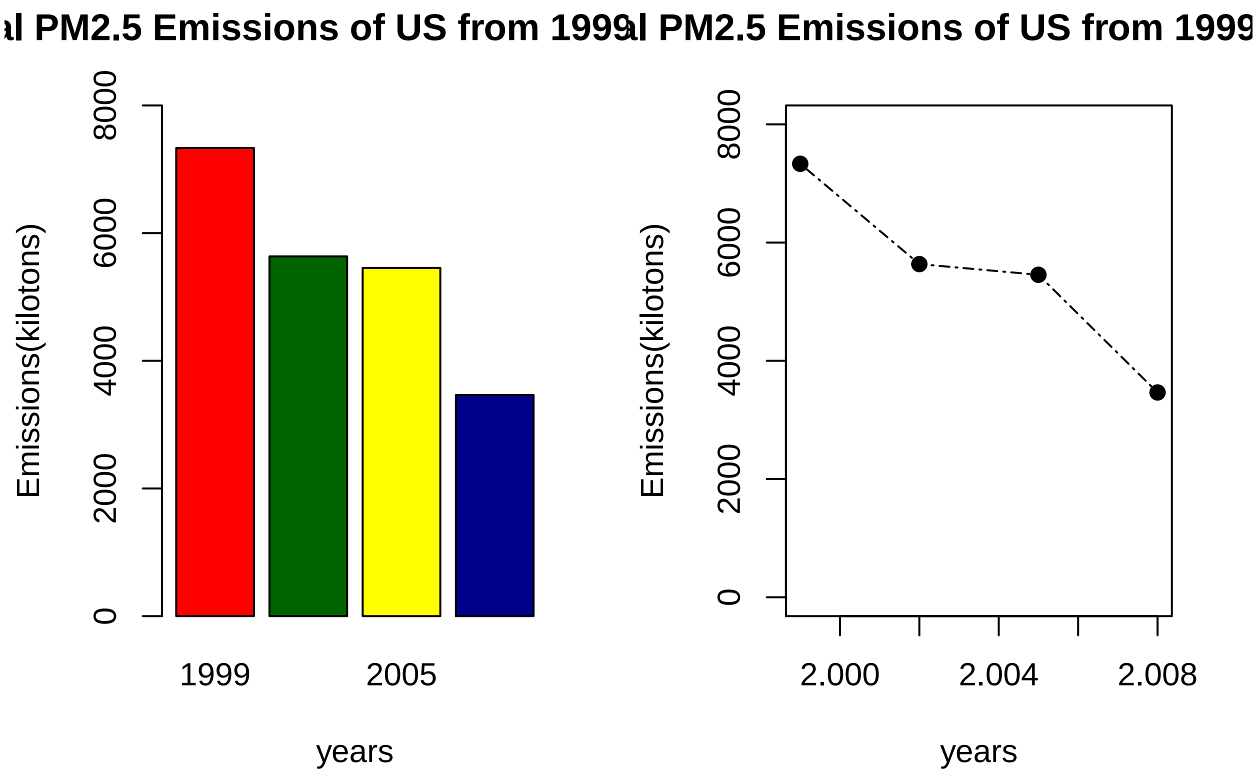
```
setwd("E:\\R\\Exploratory Data Porject2")
nei <- readRDS("summarySCC_PM25.rds")
SCC <- readRDS("Source_Classification_Code.rds")
```

2.Total Emissions Trend in US from 1999 to 2008

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 subset method with group_by
total_emission <- summarize(group_by(nei, year), Emissions = sum(Emissions))
#second subset method with aggregate function
total_emission_2 <- with(nei, aggregate(Emissions, by= list(year), sum))
#using barplot
collabel <- c("red","darkgreen","yellow","darkblue")
par(mfrow = c(1,2))
q1 <- barplot(height = total_emission$Emissions/1000,
              names.arg = factor(total_emission$year),
              col = collabel,
              main = "Total PM2.5 Emissions of US from 1999 to 2008",
              xlab= "years", ylab="Emissions(kilotons)",
              ylim = c(0,8000))
#dev.copy(png, "Plot1.png",width = 480, height = 480, unit = "px")
#dev.off()
plot(total_emission_2/1000, type = "o",
     main = "Total PM2.5 Emissions of US from 1999 to 2008",
     xlab= "years", ylab="Emissions(kilotons)", pch = 19,
     ol = "blue", lty = 6, ylim=c(0,8000))
```

```
## Warning in plot.window(...): "ol" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "ol" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "ol" is not a
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "ol" is not a
## graphical parameter

## Warning in box(...): "ol" is not a graphical parameter

## Warning in title(...): "ol" is not a graphical parameter
```




```r
#dev.copy(png, "Plot1.png",width = 480, height = 480, unit = "px")
#dev.off()
dev.copy(png, "Plot1_2.png",width = 480, height = 480, unit = "px")
```

```
## Warning in dev.copy(png, "Plot1_2.png", width = 480, height = 480, unit =
## "px"): "ol" is not a graphical parameter

## Warning in dev.copy(png, "Plot1_2.png", width = 480, height = 480, unit =
## "px"): "ol" is not a graphical parameter

## Warning in dev.copy(png, "Plot1_2.png", width = 480, height = 480, unit =
## "px"): "ol" is not a graphical parameter

## Warning in dev.copy(png, "Plot1_2.png", width = 480, height = 480, unit =
## "px"): "ol" is not a graphical parameter

## Warning in dev.copy(png, "Plot1_2.png", width = 480, height = 480, unit =
## "px"): "ol" is not a graphical parameter

## Warning in dev.copy(png, "Plot1_2.png", width = 480, height = 480, unit =
## "px"): "ol" is not a graphical parameter

## png
##   3
```

```r
dev.off()
```

```
## pdf
```

```
##   2
```

*From the plot, we can see the total emission from PM2.5 decreased in US from 1999 to 2008. Espeically during the year from 1999 to 2002, and 2005 to 2008, which have larger decrease rates.

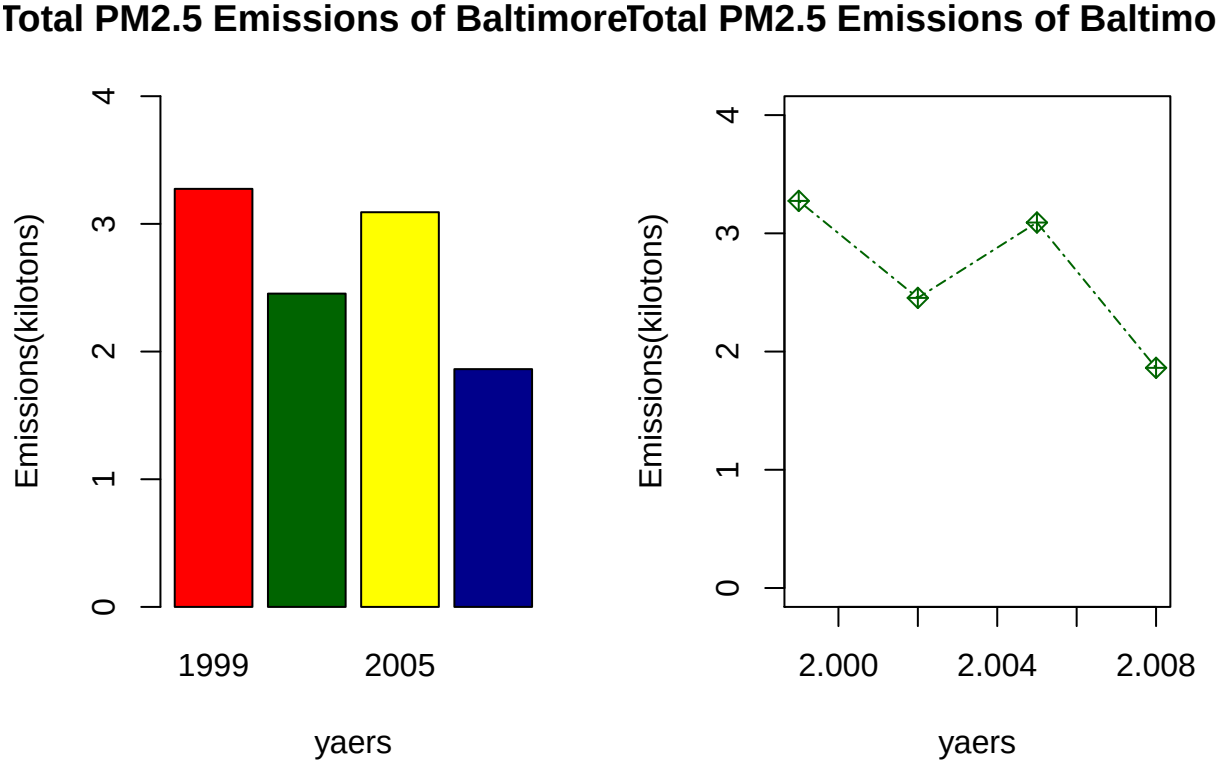
3.Emissions of Baltimore City

Have total emissions from PM2.5 decreased in the Baltimore City, Maryland (fips == "24510") from 1999 to 2008? Use the base plotting system.

```
#subset Baltimore by its fip code
baltimore <- filter(nei, fips == "24510")
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.1
```

```
t_b <- summarize(group_by(baltimore, year), Emissions = sum(Emissions))
t_b_2 <- with(baltimore, aggregate(Emissions, by=list(year), sum))
#using barplot/plot
par(mfrow = c(1,2))
q2 <- barplot(height = t_b$Emissions/1000, names.arg=t_b$year,
              col=collabel,
              main ="Total PM2.5 Emissions of Baltimore City",
              xlab = "yaers", ylab = "Emissions(kilotons)",
              ylim = c(0,4))
plot(t_b_2/1000, type = "o", col="darkgreen",
     main ="Total PM2.5 Emissions of Baltimore City",
     xlab = "yaers", ylab = "Emissions(kilotons)",
     ylim = c(0,4),pch=9, lty=6)
```



```
dev.copy(png, "Plot2_2.png",width = 480, height = 480, unit = "px")
```

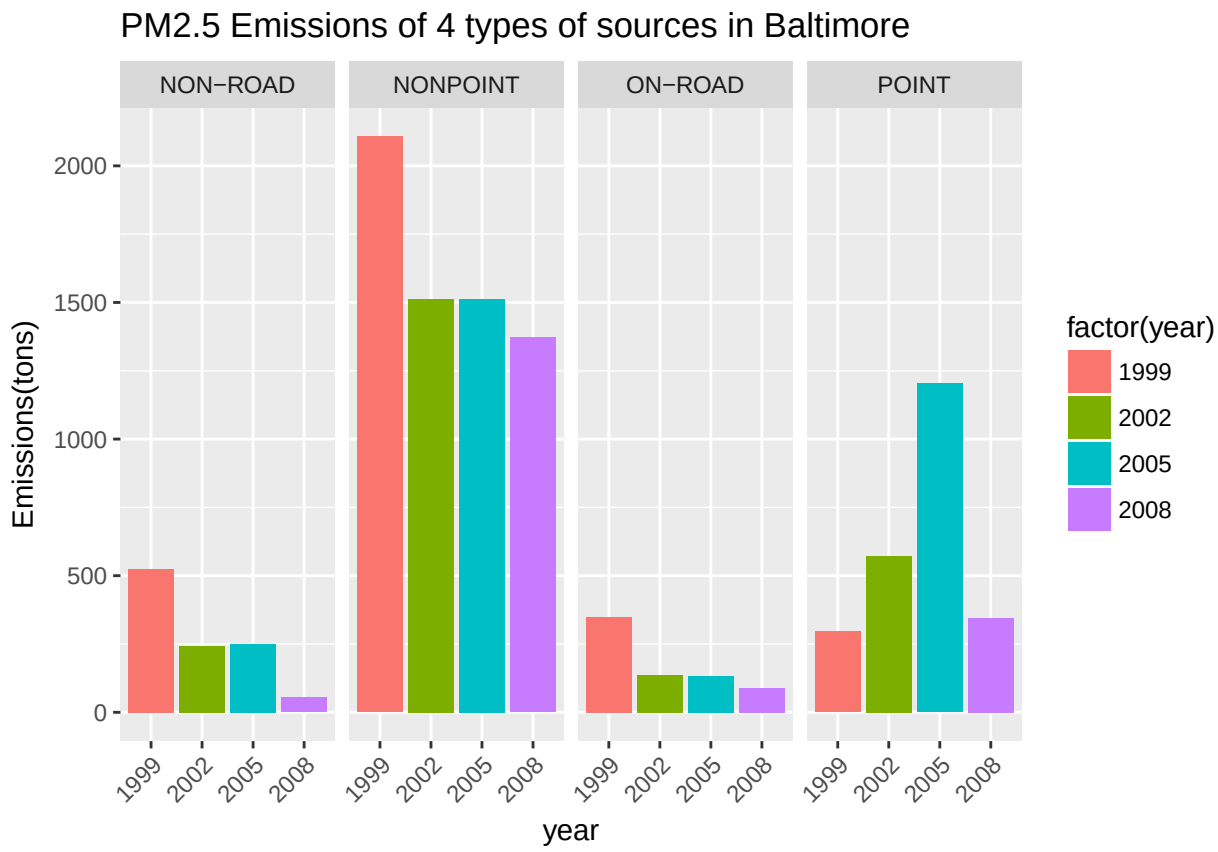
```
## png
##   3
```

```
dev.off()
```

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

*From the plots we can easily find that the total PM2.5 emission decreased from 1999 to 2002, while increased from 2002 to 2005, and then decreased again, but the total trend is decreasing. ####4.Emissions of 4 types of Sources 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.

```
g <- ggplot(baltimore, aes(x=as.factor(year), y=Emissions, fill = factor(year)))
g +
  geom_bar(stat="identity") +
  facet_grid(.~type) +
  theme(axis.text.x = element_text(angle = 45,hjust = 1)) +
  labs(x="year", y="Emissions(tons)",
       title="PM2.5 Emissions of 4 types of sources in Baltimore")
```




```
dev.copy(png, "Plot3.png")
```

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

```
dev.off()
```

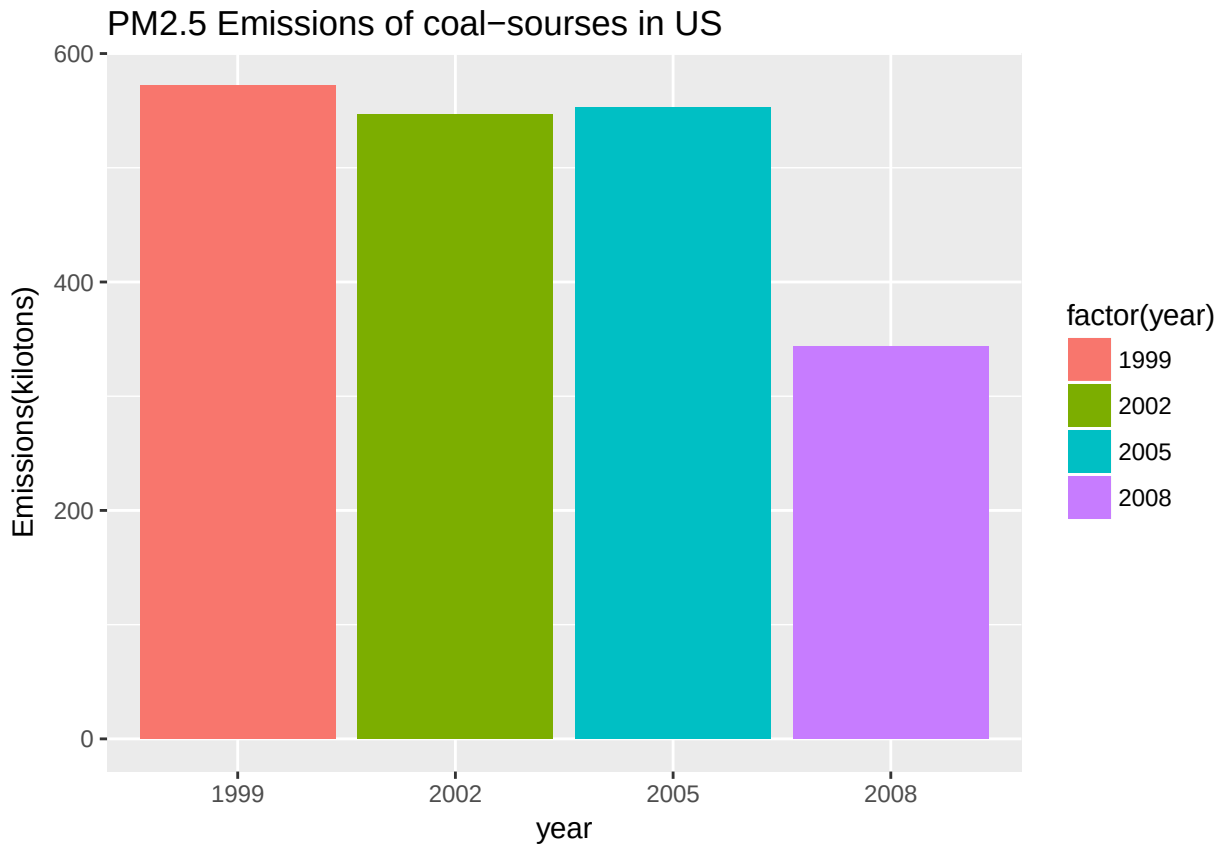
```
## pdf
##   2
```

*From the plots we can see there are three types shown decreasing trend of PM2.5 emission. All of them shown larger decrease from 1999 to 2002, while during 2002 to 2005, there are plateau, which means the decrease rates are relatively lower than former year. However, for the POINT type, there was an great increase trend from 1999 to 2005, and suddenly droped. Maybe that is something we should learn more in the future analysis.

5.Emissions of Coal combustion sources

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

```
coal <- grep("Coal$", SCC$EI.Sector)
coal_data <- SCC[coal,]
coal_nei <- nei[(nei$SCC %in% coal_data$SCC),]
coal_total <- summarize(group_by(coal_nei, year), Emissions = sum(Emissions))
gg <- ggplot(coal_total, aes(as.factor(year), Emissions/1000, fill = factor(year)))
gg + geom_bar(stat = "identity") +
  labs(x="year", y="Emissions(kilotons)",
       title="PM2.5 Emissions of coal-sourses in US")
```




```
dev.copy(png, "Plot4.png")
```

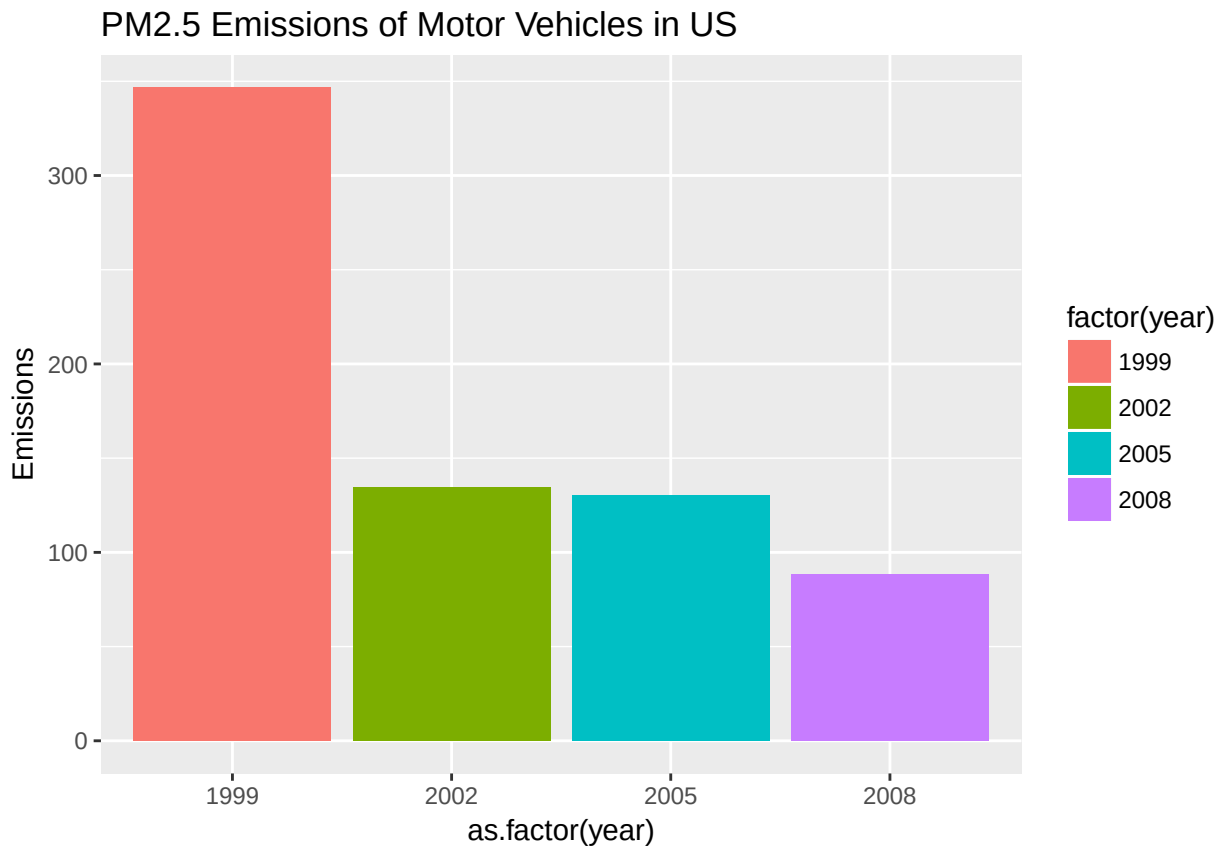
```
## png
##   3
```

```
dev.off()
```

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

*Also shown decrease trend. ####6.Motor vehicle emissions of Baltimore How have emissions from motor vehicle sources changed from 1999-2008 in Baltimore City?

```
motor_bal <- subset(baltimore, type == "ON-ROAD")
motor_bal_nei <- summarize(group_by(motor_bal, year), Emissions= sum(Emissions))
ww <- ggplot(motor_bal_nei, aes(as.factor(year), Emissions, fill = factor(year)))
ww + geom_bar(stat = "identity") +
  labs(xlab = "year", ylab = "Emissions(tons)", title="PM2.5 Emissions of Motor Vehicles in US")
```




```
dev.copy(png, "Plot5.png")
```

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

```
dev.off()
```

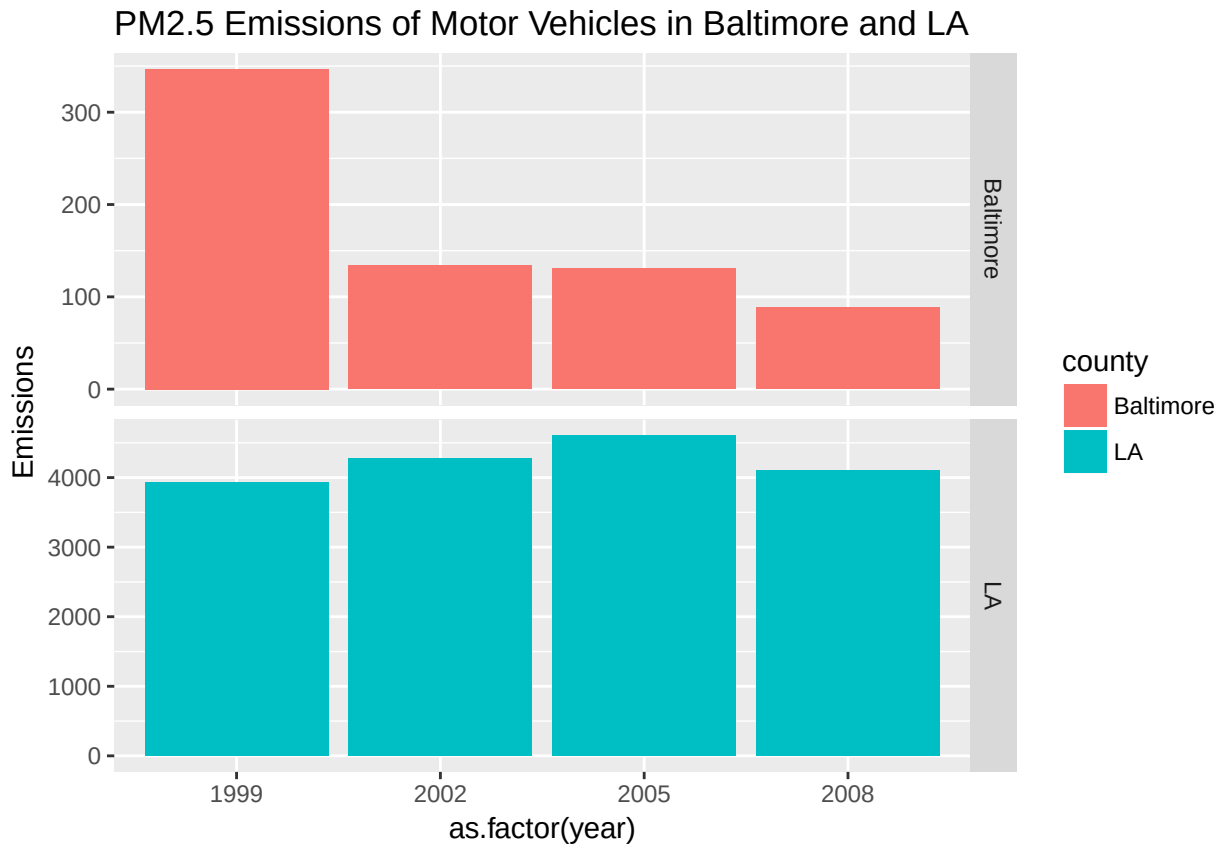
```
## pdf
##   2
```

7.Emissions of Baltimore and LA

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?

```
#subset data
motor_bal_nei <- summarize(group_by(
  filter(nei, fips=="24510"&type=="ON-ROAD"), year),
  Emissions = sum(Emissions))
motor_la_nei <- summarize(group_by(
  filter(nei, fips=="06037"&type=="ON-ROAD"), year),
  Emissions = sum(Emissions))
#add new variables as the compare factor
motor_bal_nei$county <- "Baltimore"
motor_la_nei$county <- "LA"
motor_bal_la <- rbind(motor_bal_nei, motor_la_nei)
ggplot(motor_bal_la, aes(as.factor(year), Emissions, fill = county)) +
  geom_bar(stat="identity") +
  facet_grid(county~., scales="free") +
  labs(xlab = "year", ylab = "Emissions(tons)",
       title="PM2.5 Emissions of Motor Vehicles in Baltimore and LA")
```




```
dev.copy(png,"Plot6.png")
```

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

```
dev.off()
```

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

*Baltimore shown decrease of PM2.5 emission, while on the other hand, LA has a different story. Not only did LA have much higher amount of emission, compared to Baltimore, but also it shown an increase trend from 1999 to 2008.