**Exploratory Data Analysis Project 2 (JHU) Coursera**

**Unzipping and Loading Files**

library("data.table")

path <- getwd()

download.file(url = "https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2FNEI\_data.zip"

, destfile = paste(path, "dataFiles.zip", sep = "/"))

unzip(zipfile = "dataFiles.zip")

SCC <- data.table::as.data.table(x = readRDS(file = "Source\_Classification\_Code.rds"))

NEI <- data.table::as.data.table(x = readRDS(file = "summarySCC\_PM25.rds"))

**Question 1 (**[**plot1.R**](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/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.

# Prevents histogram from printing in scientific notation

NEI[, Emissions := lapply(.SD, as.numeric), .SDcols = c("Emissions")]

totalNEI <- NEI[, lapply(.SD, sum, na.rm = TRUE), .SDcols = c("Emissions"), by = year]

png(filename='plot1.png')

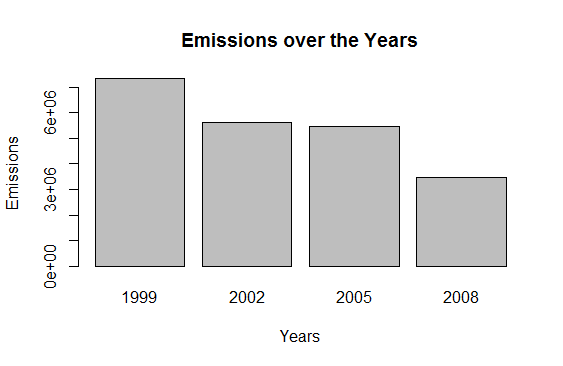
barplot(totalNEI[, Emissions]

, names = totalNEI[, year]

, xlab = "Years", ylab = "Emissions"

, main = "Emissions over the Years")

dev.off()

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/plot1.png)

**Question 2 (**[**plot2.R**](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/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.

NEI[, Emissions := lapply(.SD, as.numeric), .SDcols = c("Emissions")]

totalNEI <- NEI[fips=='24510', lapply(.SD, sum, na.rm = TRUE)

, .SDcols = c("Emissions")

, by = year]

png(filename='plot2.png')

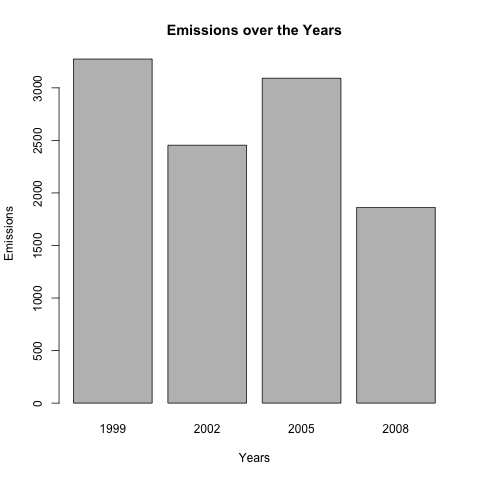
barplot(totalNEI[, Emissions]

, names = totalNEI[, year]

, xlab = "Years", ylab = "Emissions"

, main = "Emissions over the Years")

dev.off()

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/plot2.png)

**Question 3 (**[**plot3.R**](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/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.

# Subset NEI data by Baltimore

baltimoreNEI <- NEI[fips=="24510",]

png("plot3.png")

ggplot(baltimoreNEI,aes(factor(year),Emissions,fill=type)) +

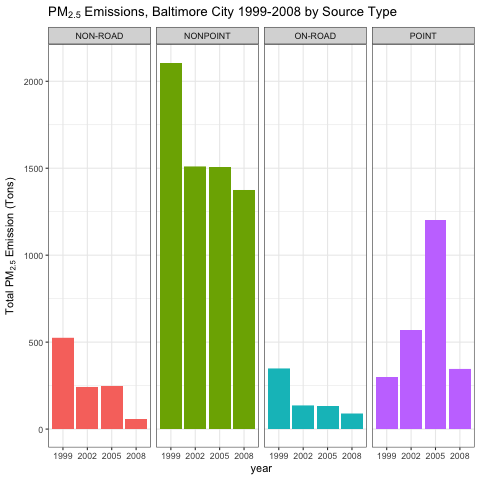
geom\_bar(stat="identity") +

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"))

dev.off()

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/plot3.png)

**Question 4 (**[**plot4.R**](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/plot4.R)**)**

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

# Subset coal combustion related NEI data

combustionRelated <- grepl("comb", SCC[, SCC.Level.One], ignore.case=TRUE)

coalRelated <- grepl("coal", SCC[, SCC.Level.Four], ignore.case=TRUE)

combustionSCC <- SCC[combustionRelated & coalRelated, SCC]

combustionNEI <- NEI[NEI[,SCC] %in% combustionSCC]

png("plot4.png")

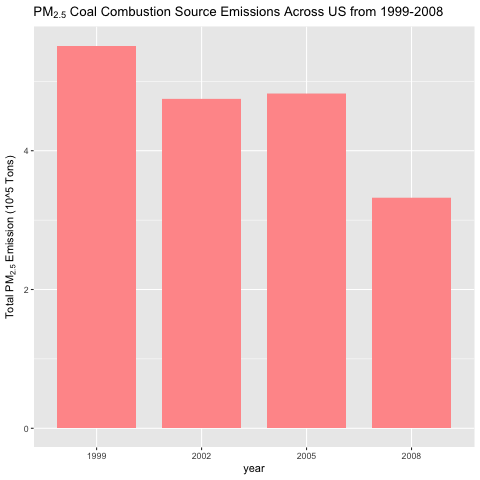
ggplot(combustionNEI,aes(x = factor(year),y = Emissions/10^5)) +

geom\_bar(stat="identity", fill ="#FF9999", width=0.75) +

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"))

dev.off()

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/plot4.png)

**Question 5 (**[**plot5.R**](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/plot5.R)**)**

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

# Gather the subset of the NEI data which corresponds to vehicles

condition <- grepl("vehicle", SCC[, SCC.Level.Two], ignore.case=TRUE)

vehiclesSCC <- SCC[condition, SCC]

vehiclesNEI <- NEI[NEI[, SCC] %in% vehiclesSCC,]

# Subset the vehicles NEI data to Baltimore's fip

baltimoreVehiclesNEI <- vehiclesNEI[fips=="24510",]

png("plot5.png")

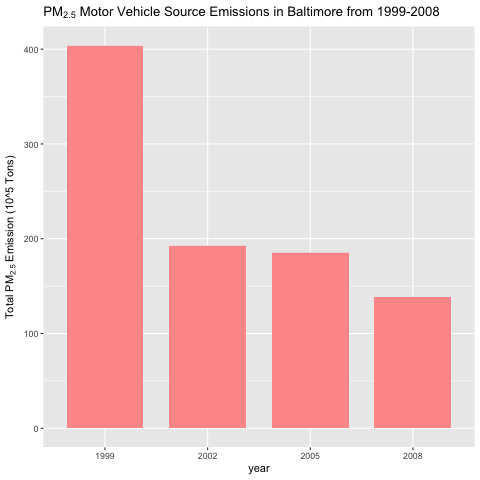
ggplot(baltimoreVehiclesNEI,aes(factor(year),Emissions)) +

geom\_bar(stat="identity", fill ="#FF9999" ,width=0.75) +

labs(x="year", y=expression("Total PM"[2.5]\*" Emission (10^5 Tons)")) +

labs(title=expression("PM"[2.5]\*" Motor Vehicle Source Emissions in Baltimore from 1999-2008"))

dev.off()

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/plot5.png)

**Question 6 (**[**plot6.R**](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/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?

# Gather the subset of the NEI data which corresponds to vehicles

condition <- grepl("vehicle", SCC[, SCC.Level.Two], ignore.case=TRUE)

vehiclesSCC <- SCC[condition, SCC]

vehiclesNEI <- NEI[NEI[, SCC] %in% vehiclesSCC,]

# Subset the vehicles NEI data by each city's fip and add city name.

vehiclesBaltimoreNEI <- vehiclesNEI[fips == "24510",]

vehiclesBaltimoreNEI[, city := c("Baltimore City")]

vehiclesLANEI <- vehiclesNEI[fips == "06037",]

vehiclesLANEI[, city := c("Los Angeles")]

# Combine data.tables into one data.table

bothNEI <- rbind(vehiclesBaltimoreNEI,vehiclesLANEI)

png("plot6.png")

ggplot(bothNEI, aes(x=factor(year), y=Emissions, fill=city)) +

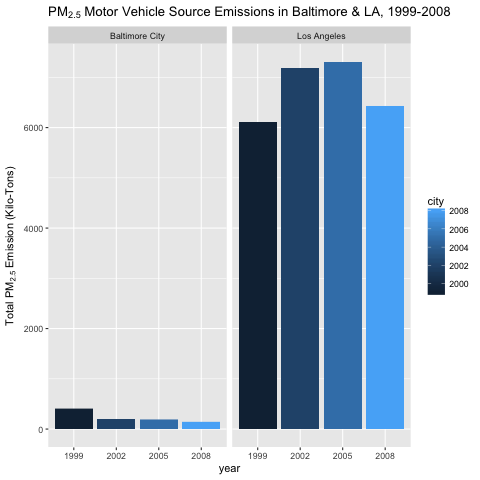
geom\_bar(aes(fill=year),stat="identity") +

facet\_grid(scales="free", space="free", .~city) +

labs(x="year", y=expression("Total PM"[2.5]\*" Emission (Kilo-Tons)")) +

labs(title=expression("PM"[2.5]\*" Motor Vehicle Source Emissions in Baltimore & LA, 1999-2008"))

dev.off()

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project2/plot6.png)