(https://courserahelp.zendesk.com/hc/en-us/requests/new)

Peer Assessments (https://class.coursera.org/exdata-007/human_grading/) / Course Project 2

Help (https://class.coursera.org/exdata-007/help/peergrading?url=https%3A%2F%2Fclass.coursera.org%2Fexdata-007%2Fhuman_grading%2Fview%2Fcourses%2F972594%2Fassessments%2F4%2Fsubmissions)

due in 4day 22h

Submission Phase

1. Do assignment ☑ (/exdata-007/human_grading/view/courses/972594/assessments/4/submissions)

Evaluation Phase

2. Evaluate peers 🔒 (/exdata-007/human_grading/view/courses/972594/assessments/4/peerGradingSets)

Results Phase

3. See results 🔒 (/exdata-007/human_grading/view/courses/972594/assessments/4/results/mine)

✔ Submitted. You can still make changes and re-submit before the deadline.

☑ In accordance with the Honor Code, I certify that my answers here are my own work, and that I have appropriately acknowledged all external sources (if any) that were used in this work.

Re-submit for grading

Introduction

Fine particulate matter ($PM_{2.5}$) 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 $PM_{2.5}$. 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 (http://www.epa.gov/ttn/chief/eiinformation.html).

For each year and for each type of PM source, the NEI records how many tons of $PM_{2.5}$ 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.

Data

The data for this assignment are available from the course web site as a single zip file:

- Data for Peer Assessment (https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2FNEI_data.zip) [29Mb]

The zip file contains two files:

$PM_{2.5}$ Emissions Data (`summarySCC_PM25.rds`): This file contains a data frame with all of the $PM_{2.5}$ emissions data for 1999, 2002, 2005, and 2008. For each year, the table contains number of **tons** of $PM_{2.5}$ emitted from a specific type of source for the entire year. Here are the first few rows.

```
##     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
```

- `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 $PM_{2.5}$ emitted, in tons
- `type`: The type of source (point, non-point, on-road, or non-road)
- `year`: The year of emissions recorded

Source Classification Code Table (`Source_Classification_Code.rds`): This table provides a mapping from the SCC digit strings int he Emissions table to the actual name of the $PM_{2.5}$ source. The sources are categorized in a few different ways from more general to more specific and you may choose to explore whatever categories you think are most useful. For example, source “10100101” is known as “Ext Comb /Electric Gen /Anthracite Coal /Pulverized Coal”.

You can read each of the two files using the `readRDS()` function in R. For example, reading in each file can be done with the following code:

```
## This first line will likely take a few seconds. Be patient!
NEI <- readRDS("summarySCC_PM25.rds")
SCC <- readRDS("Source_Classification_Code.rds")
```

as long as each of those files is in your current working directory (check by calling `dir()` and see if those files are in the listing).

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 $PM_{2.5}$ decreased in the United States from 1999 to 2008? Using the **base** plotting system, make a plot showing the *total* $PM_{2.5}$ emission from all sources for each of the years 1999, 2002, 2005, and 2008.

2. Have total emissions from $PM_{2.5}$ 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?

Making and Submitting Plots

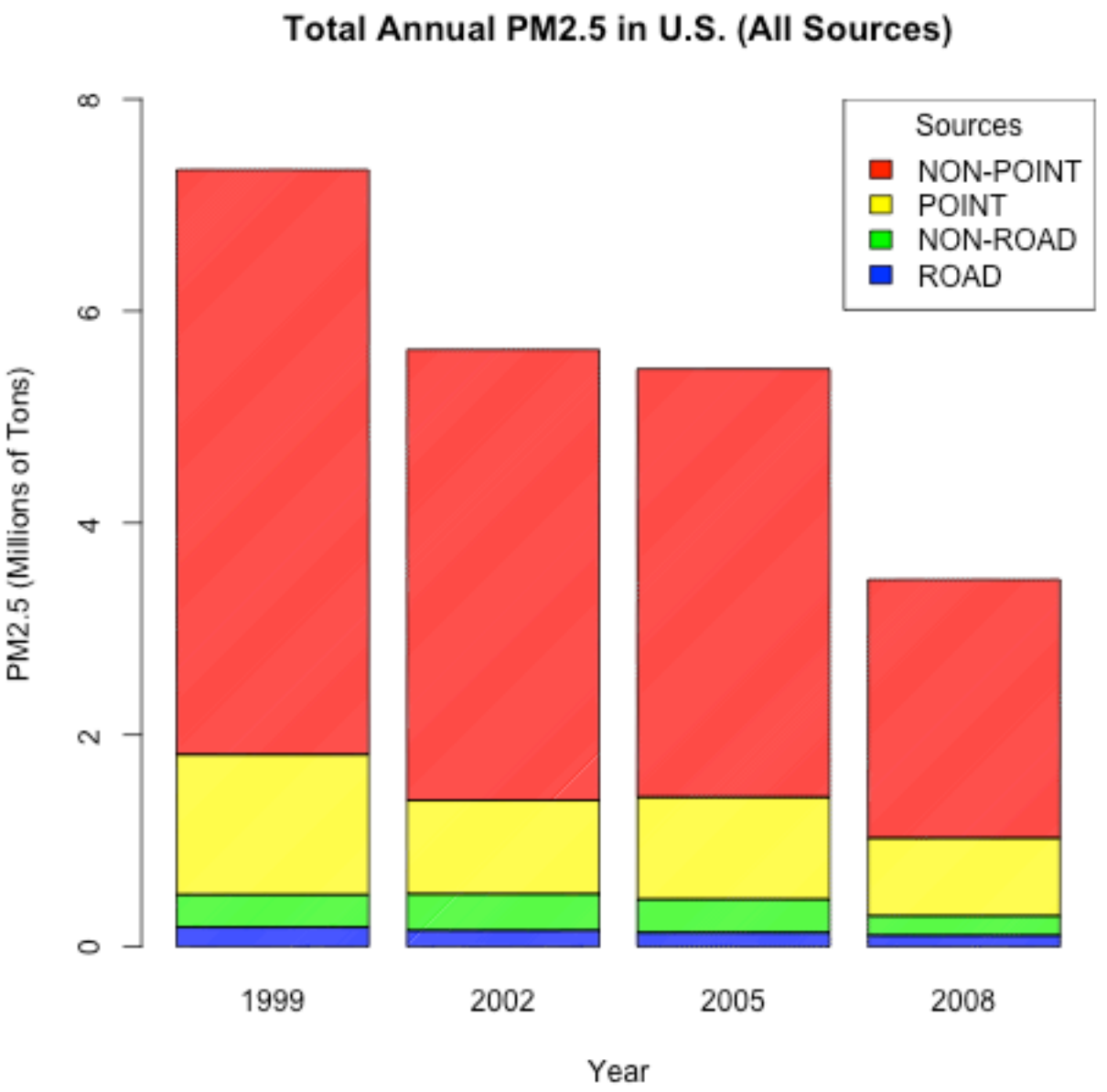
For each plot you should

- Construct the plot and save it to a **PNG file**.
- Create a separate R code file (`plot1.R` , `plot2.R` , etc.) that constructs the corresponding plot, i.e. code in plot1.R constructs the plot1.png plot. Your code file should include code for reading the data so that the plot can be fully reproduced. You should also include the code that creates the PNG file. Only include the code for a single plot (i.e. `plot1.R` should only include code for producing `plot1.png`)
- Upload the PNG file on the Assignment submission page
- Copy and paste the R code from the corresponding R file into the text box at the appropriate point in the peer assessment.

Have total emissions from $PM_{2.5}$ decreased in the United States from 1999 to 2008? Using the **base** plotting system, make a plot showing the *total* $PM_{2.5}$ emission from all sources for each of the years 1999, 2002, 2005, and 2008.
Upload a PNG file containing your plot addressing this question.

B | *I* | Link | <code> | Math | Edit: Rich | Preview

Total PM2.5 emissions have decreased by almost half in the United States from 1999 - 2008.



Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Please view the plot for this question. Does the plot appear to address the question being asked? In other words, can you answer the question using the information shown in the plot?

Upload the R code file for the plot uploaded in the previous question.

B | I | Link | <code> | Math | Edit: Rich | Preview

```
# function plot1()
#
# This function creates a stacked bar chart of NEI data for the years 1999 - 2008,
# summarizing the total PM25 emissions across the nation for each year.
#
# The function uses the barplot function from the base graphics package
#
# It assumes the presence of a data frame with NEI data in the parent environment.
#
# If that is not there, it calls getAndCleanData() to read it in.
#



plot1 <- function() {

    library(dplyr)

# If source data is not already in parent environment, read in the data
if (!exists ("NEI")) {

    getAndCleanData()

}

# For this plot we only need year, type and Emissions from NEI
NEI2 <- tbl_df(subset(NEI, select = c(year, type, Emissions)))

# Change year into a factor for grouping and plotting
NEI2 <- transform(NEI2, year = factor(year))

# Group by type, then year
NEI2 <- NEI2 %>% group_by(type, year)

# Summarise by type, then year
NEI2 <- summarise(NEI2, sum(Emissions))

# Relabel columns
colnames(NEI2) <- c("Source", "Year", "totPM25")

# Create a height matrix for stacked barcharts
a <- rbind(NEI2[1:4,3], NEI2[5:8,3], NEI2[9:12,3], NEI2[13:16,3])
```

```
    # Order the matrix rows from low to high (first row plotted first)
    row <- order(a[,1])
    a <- a[row,]

    # Pull out the date factors
    b <- unique(NEI2$Year)

    # Open image file
    png (filename = "./plot1.png")

    # Plot the PM2.5 Totals
    barplot(height = a/1.e6, names.arg = b, col = c("blue", "green", "yellow", "red"),
            density = 100, angle = 45, ylim = c(0,8), xlab = "Year",
            ylab = "PM2.5 (Millions of Tons)",
            main = "Total Annual PM2.5 in U.S. (All Sources)", beside = FALSE,
            legend.text = c("ROAD", "NON-ROAD", "POINT", "NON-POINT"))

    # Close file
    dev.off()
}

# function getAndCleanData ()
#
# This function checks for the presence of NEI data in the working directory.
# If it does not exist, it downloads and unzips the source .zip file.
# It then reads the source RDS data into NEI and SCC data frames and stores
# That data in the parent environment for used by other functions.
#

getAndCleanData <- function () {

    # If the project source data does not exist, then
    if (!file.exists("./summarySCC_PM25.rds")) {

        url <- "https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2FNEI_data.zip"
        destFile <- "exdata-data-NEI_data.zip"

        # Download File
        download.file(url, destfile = destFile, method = "curl")

        # Unzip File
        unzip(destFile)

        # Remove Zip File
        file.remove(destFile)

    }
```

```
    # Read In Data and save data frames to the parent environment
    NEI <<- readRDS("summarySCC_PM25.rds")
    SCC <<- readRDS("Source_Classification_Code.rds")

}

plot1 ()
```

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

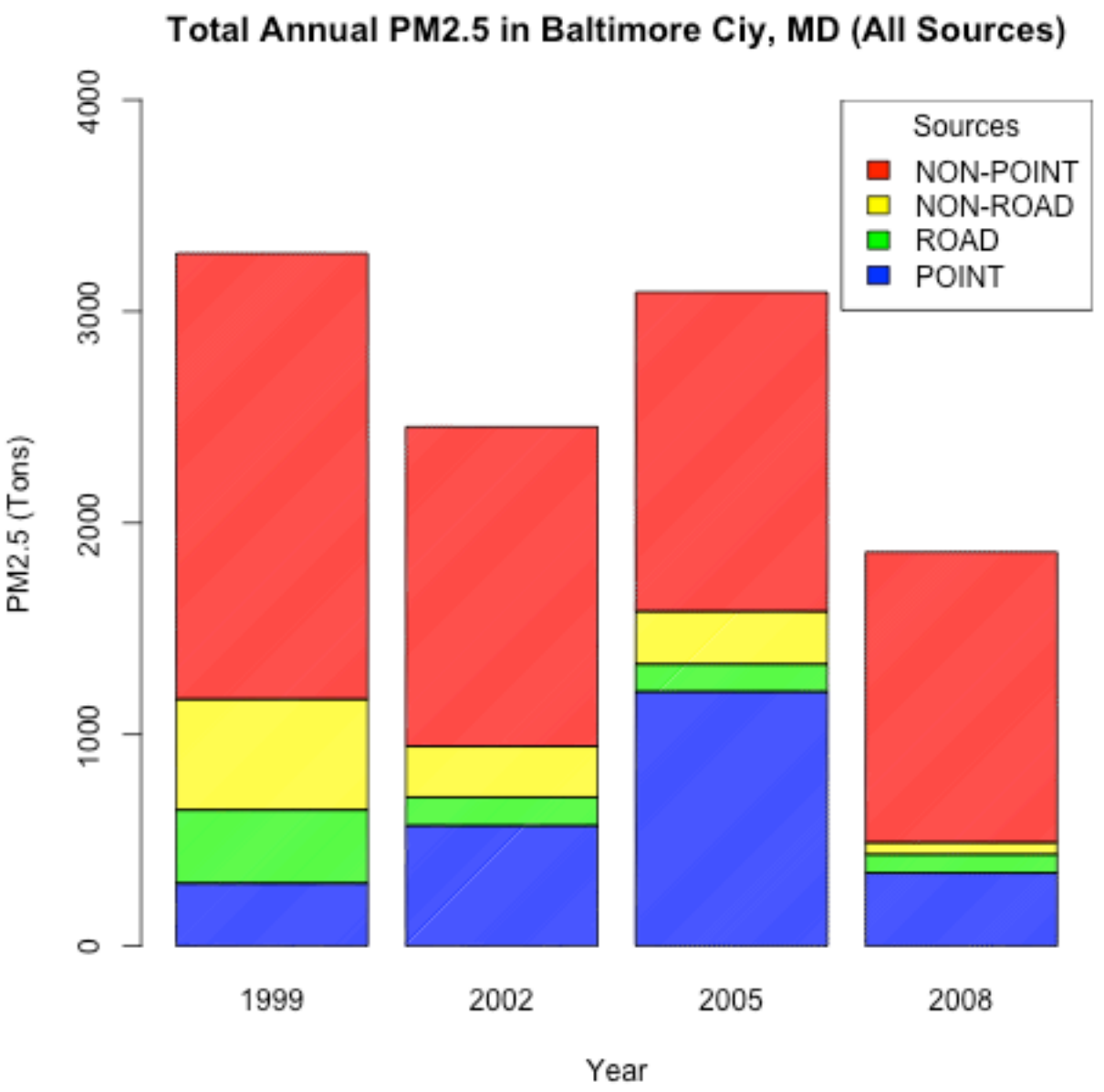
Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

Have total emissions from $PM_{2.5}$ decreased in the **Baltimore City**, Maryland (`fips == 24510`) from 1999 to 2008? Use the **base** plotting system to make a plot answering this question.

Upload a PNG file containing your plot addressing this question.

B | *I* | Link | <code> | Math | Edit: Rich | Preview

Overall, PM2.5 levels have decreased in Baltimore City, MD from 1999-2008. There was a bit of an anomaly in 2005, where levels increased mostly because of an increase in POINT data. However, the sum of all sources was still below the 1999 levels. In 2008, the total level was sharply lower.



Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Please view the plot for this question. Does the plot appear to address the question being asked? In other words, can you answer the question using the information shown in the plot?

Upload the R code file for the plot uploaded in the previous question.

B | I | Link | <code> | Math | Edit: Rich | Preview

```
# function plot2()
#
# This function creates a stacked bar chart of NEI data for the years 1999 - 2008,
# summarizing the total PM25 emissions for Baltimore City, MD for each year.
#
# The function uses the barplot function from the base graphics package
#
# It assumes the presence of a data frame with NEI data in the parent environment.
#
# If that is not there, it calls getAndCleanData() to read it in.
#
# Dale Wickizer
# 10/19/2014

plot2 <- function() {

    library(dplyr)

# If source data is not already in parent environment, read in the data
if (!exists ("NEI")) {

    getAndCleanData()

}

# For this plot we only need year, type and Emissions from NEI for Baltimore City, MD
# fips == 24510
NEI2 <- tbl_df(subset(NEI, fips == "24510", select = c(year, type, Emissions)))

# Change year into a factor for grouping and plotting
NEI2 <- transform(NEI2, year = factor(year))

# Group by type, then year
NEI2 <- NEI2 %>% group_by(type, year)

# Summarise by type, then year
NEI2 <- summarise(NEI2, sum(Emissions))

# Relabel columns
colnames(NEI2) <- c("Source", "Year", "totPM25")

# Create a height matrix for stacked barcharts
```

```
    # Create a height matrix for stacked barcharts
    a <- rbind(NEI2[1:4,3], NEI2[5:8,3], NEI2[9:12,3], NEI2[13:16,3])

    # Order the matrix rows from low to high (first row plotted first)
    row <- order(a[,1])
    a <- a[row,]

    # Pull out the date factors
    b <- unique(NEI2$Year)

    # Open image file
    png (filename = "./plot2.png")

    # Plot the PM2.5 Totals
    barplot(height = a, names.arg = b, col = c("blue", "green", "yellow", "red"),
            density = 100, angle = 45, ylim = c(0,4000), xlab = "Year", ylab = "PM2.5 (Tons)",
            main = "Total Annual PM2.5 in Baltimore Ciy, MD (All Sources)", beside = FALSE,
            legend.text = c("POINT", "ROAD", "NON-ROAD", "NON-POINT"))

    # Close file
    dev.off()
}

# function getAndCleanData ()
#
# This function checks for the presence of NEI data in the working directory.
# If it does not exist, it downloads and unzips the source .zip file.
# It then reads the source RDS data into NEI and SCC data frames and stores
# That data in the parent environment for used by other functions.
#
# Dale Wickizer
# 10/19/2014
#
getAndCleanData <- function () {

    # If the project source data does not exist, then
    if (!file.exists("./summarySCC_PM25.rds")) {

        url <- "https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2FNEI_data.zip"
        destFile <- "exdata-data-NEI_data.zip"

        # Download File
        download.file(url, destfile = destFile, method = "curl")

        # Unzip File
        unzip(destFile)

        # Remove Zip File
        file.remove(destFile)
```

```
    }

    # Read In Data and save data frames to the parent environment
    NEI <<- readRDS("summarySCC_PM25.rds")
    SCC <<- readRDS("Source_Classification_Code.rds")

}

plot2 ()
```

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

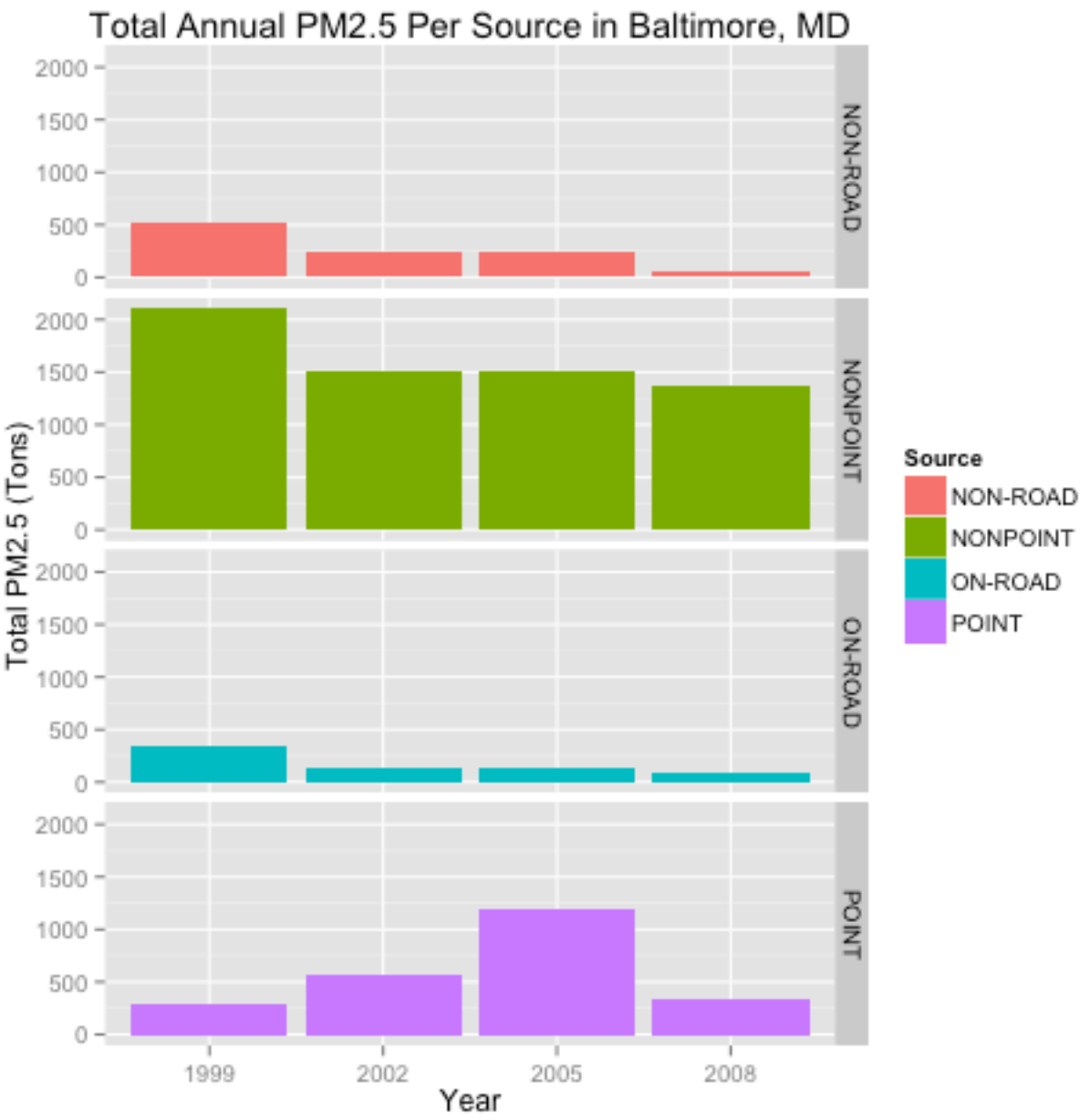
Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

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.

Upload a PNG file containing your plot addressing this question.

Three of the four sources (non point, non road and inroad) show a decrease in the total measured PM2.5 levels in Baltimore City, MD for the years 1999-2008. One source (point) showed an increase the total measured PM2.5 levels in 2005, decreasing in 2008 to a level just above the 1999 level.

The attached graph shows bar charts of each of the sources as confirmation.



Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Please view the plot for this question. Does the plot appear to address the question being asked? In other words, can you answer the question using the information shown in the plot?

Upload the R code file for the plot uploaded in the previous question.

B | I | Link | <code> | Math | Edit: Rich | Preview

```
# function plot3()
#
# This function creates bar charts for the various sources of NEI data for the years
# 1999 - 2008, summarizing the total PM25 emissions for Baltimore City, MD for each year.
#
# The function uses the qplot function from the ggplot2 package
#
# It assumes the presence of a data frame with NEI data in the parent environment.
#
# If that is not there, it calls getAndCleanData() to read it in.
#

plot3 <- function() {

    library(dplyr)
    library(ggplot2)

    # If source data is not already in parent environment, read in the data
    if (!exists ("NEI")) {

        getAndCleanData()

    }

    # For this plot we only need year, type and Emissions from NEI for Baltimore City, MD
    # fips == 24510
    NEI2 <- tbl_df(subset(NEI, fips == "24510", select = c(year, type, Emissions)))

    # Change year into a factor for grouping and plotting
    NEI2 <- transform(NEI2, year = factor(year))

    # Group by year then type
    NEI2 <- NEI2 %>% group_by(year, type)

    # Summarise by year and type
    NEI2 <- summarise(NEI2, sum(Emissions))
```

```
    # Relabel columns
    colnames(NEI2) <- c("Year", "Source", "totPM25")

    # Open image file
    png (filename = "./plot3.png")


    # Plot the Total Annual PM2.5 data by Source and Year for Baltimore City, MD
    bar <- qplot(x = Year, y = totPM25, data = NEI2, fill = Source, facets = Source ~ .,
         ylab = "Total PM2.5 (Tons)", main = "Total Annual PM2.5 Per Source in Baltimore, MD",
         stat = "identity", geom = c("bar"))

    print(bar)

    # Close file
    dev.off()
}



# function getAndCleanData ()
#
# This function checks for the presence of NEI data in the working directory.
# If it does not exist, it downloads and unzips the source .zip file.
# It then reads the source RDS data into NEI and SCC data frames and stores
# That data in the parent environment for used by other functions.
#

getAndCleanData <- function () {

    # If the project source data does not exist, then
    if (!file.exists("./summarySCC_PM25.rds")) {

        url <- "https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2FNEI_data.zip"
        destFile <- "exdata-data-NEI_data.zip"

        # Download File
        download.file(url, destfile = destFile, method = "curl")

        # Unzip File
        unzip(destFile)

        # Remove Zip File
        file.remove(destFile)

    }

    # Read In Data and save data frames to the parent environment
```

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

}

plot3 ()
```

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

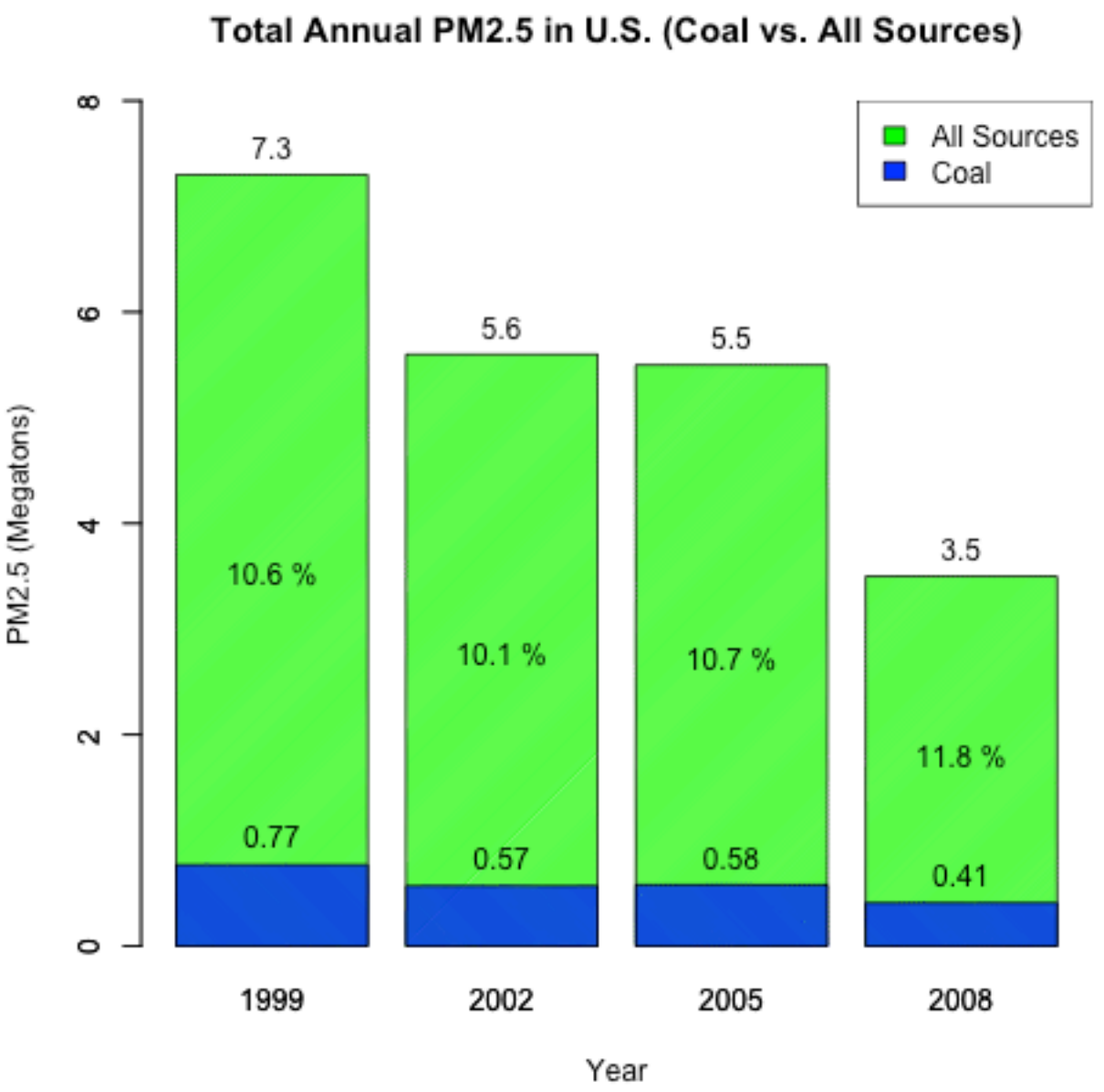
Note: this section can only be filled out during the evaluation phase.

Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

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

Upload a PNG file containing your plot addressing this question.

The emissions of PM2.5 from coal combustion-related sources have decreased from 1999 - 2008. However, total emissions from other non-coal sources have decreased faster. So, while absolute levels are lower since 1999, coal combustion-related sources now make up a higher percentage of the total US PM2.5 levels. This information is summarized in the graph below. **NOTE:** In the graph below, the Coal Data is an overlay for comparison.



Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Please view the plot for this question. Does the plot appear to address the question being asked? In other words, can you answer the question using the information shown in the plot?

Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

B | I | Link | <code> | Math | Edit: Rich | Preview

```
# function plot4()
#
# This function compares the coal combustion sources of PM2.5 to the US totals from
# 1999 - 2008. It summarizes the results of that comparison in a table and graph.
#
# The graph is created using the base graphics package.
#
# It assumes the presence of a data frame with NEI data in the parent environment.
#
# If that is not there, it calls getAndCleanData() to read it in.
#

plot4 <- function() {

    library(dplyr)

    # If source data is not already in parent environment, read in the data
    if (!exists ("NEI")) {

        getAndCleanData()

    }

    # Calculate Total PM2.5 Emissions for US by year using the NEI data from parent env
    NEIsm <- subset(NEI, select = c(Emissions, year))
    NEIsm <- NEIsm %>% group_by(year)
    totUS <- summarize(NEIsm, sum(Emissions))
    colnames(totUS) <- c("Year", "US.PM2.5")

    # Create a subset of NEI data from Coal Combustion Sources
    NEI2 <- inner_join(SCC, NEI, by = "SCC")
    NEI2 <- subset(NEI2, select = c(fips, Short.Name, Emissions, year))

    # Find the Combustion Rows
    row <- grep ("Ext Comb", NEI2$Short.Name, ignore.case = TRUE)
```

```
NEI2 <- NEI2[row,]

# Find the Combustion Rows that are from Coal
row <- grep ("Ext Comb", NEI2$Short.Name, ignore.case = TRUE)
NEI2 <- NEI2[row,]

colnames(NEI2) <- c("fips", "Short.Name", "Emissions", "Year")

# Calculate Total PM2.5 Emissions for US by Coal by year
NEI2sm <- subset(NEI2, select = c(Emissions, Year))
NEI2sm <- NEI2sm %>% group_by(Year)
totCoal <- summarize(NEI2sm, sum(Emissions))
colnames(totCoal) <- c("Year", "from.Coal")

# Compare and summarize Sources
totSum <- inner_join(totUS, totCoal, by = "Year")

# Scale to Megatons
totSum$US.Megatons <- totSum$US.PM2.5/1.e6
totSum$Coal.Megatons <- totSum$from.Coal/1.e6

# Find percentage of Coal vs. US total
totSum$Percent <- with(totSum, from.Coal/US.PM2.5*100)
```

```
# Reformat result
totSum <- subset(totSum, select = c(Year, US.Megatons, Coal.Megatons, Percent))
totSum <- transform(totSum, US.Megatons = round(US.Megatons,1),
                    Coal.Megatons = round(Coal.Megatons,2),
                    Percent = round(Percent,1))

print(totSum, row.names = FALSE)

# Calculate y-offsets for text labels and generate labels
us.ty <- totSum$US.Megatons
us.labels <- as.character(totSum$US.Megatons)

coal.ty <- totSum$Coal.Megatons
coal.labels <- as.character(totSum$Coal.Megatons)

per.ty <- (us.ty - coal.ty)/2
per.labels <- as.character(totSum$Percent)
per.labels <- paste(per.labels, "%")

# Open image file
png (filename = "./plot4.png")



# Plot the total US data first
```

```
    bp1 <- barplot(height = totSum$US.Megatons, names.arg = totSum$Year, col = "green",
            density = 100, angle = 45, ylim = c(0,8), xlab = "Year", ylab = "PM2.5 (Megatons)",
            main = "Total Annual PM2.5 in U.S. (Coal vs. All Sources)" )

    # Plot the Coal data next; It will overlay Total US, The part left
    # represents the Non-Coal sources
    bp2 <- barplot(height = totSum$Coal.Megatons, names.arg = totSum$Year, col = "blue",
            density = 100, angle = -45, ylim = c(0,8), beside = TRUE, add = TRUE )

    legend("topright", legend = c("All Sources", "Coal"), fill = c("green", "blue"))

    text(bp1, us.ty, labels = us.labels, col = "black", pos = 3)
    text(bp2, coal.ty, labels = coal.labels, col = "black", pos = 3)
    text(bp1, per.ty, labels = per.labels, col = "black", pos = 3)


    # Close file
    dev.off()
}


# function getAndCleanData ()
#
# This function checks for the presence of NEI data in the working directory.
# If it does not exist, it downloads and unzips the source .zip file.
# It then reads the source RDS data into NEI and SCC data frames and stores
# That data in the parent environment for used by other functions.
#

getAndCleanData <- function () {

    # If the project source data does not exist, then
    if (!file.exists("./summarySCC_PM25.rds")) {

        url <- "https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2FNEI_data.zip"
        destFile <- "exdata-data-NEI_data.zip"

        # Download File
        download.file(url, destfile = destFile, method = "curl")

        # Unzip File
        unzip(destFile)

        # Remove Zip File
        file.remove(destFile)

    }

    # Read In Data and save data frames to the parent environment
```

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

}

plot4 ()
```

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

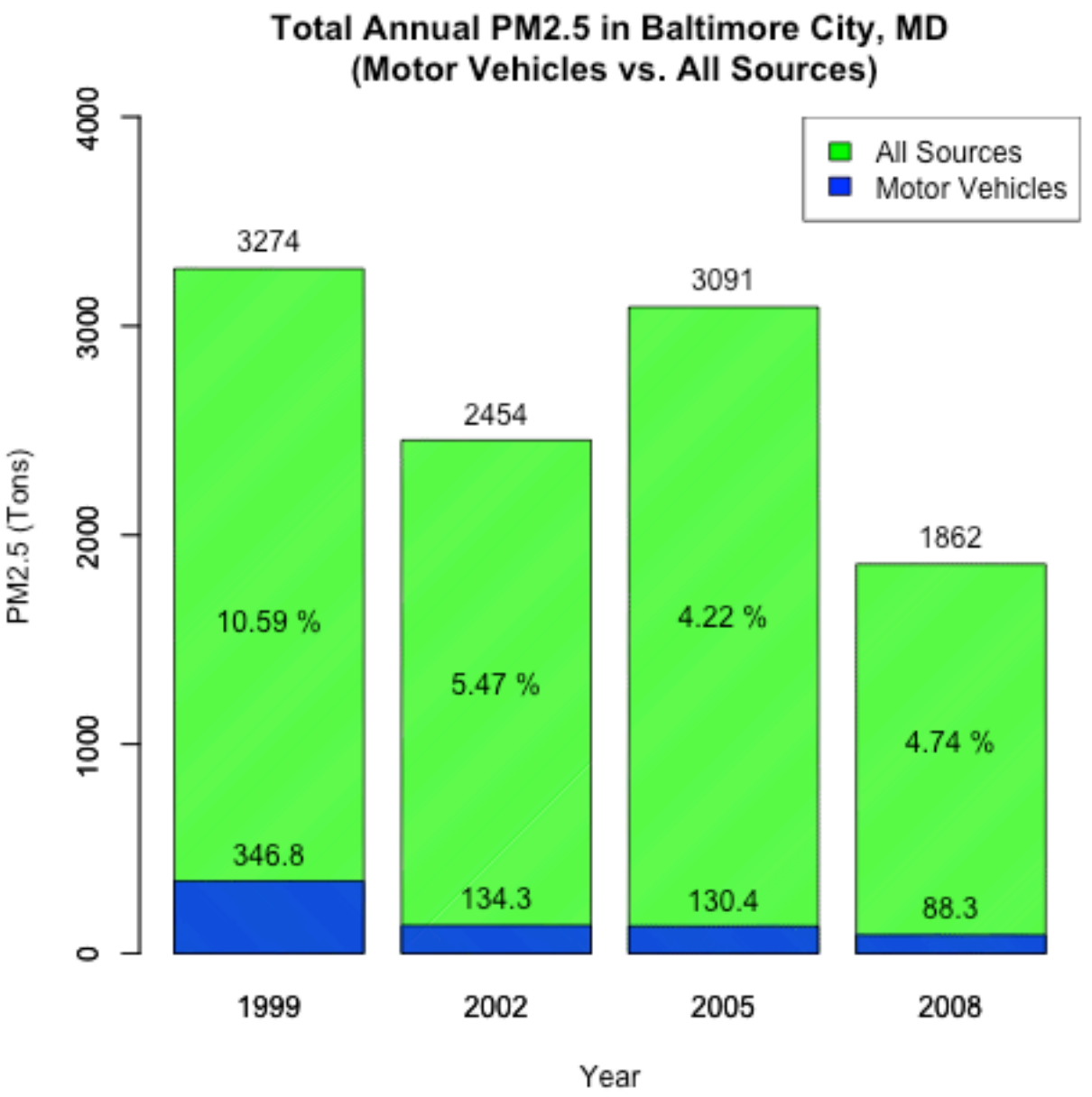
Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

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

Upload a PNG file containing your plot addressing this question.

B | *I* | Link | <code> | Math | Edit: Rich | Preview

Motor vehicle emissions of PM2.5 have been cut by almost a factor of 4 (3.93) in Baltimore City, MD from 1999 - 2008, decreasing faster than other sources over that time. As a result, motor vehicle emissions are a much smaller percentage of the overall PM2.5 emissions in Baltimore City. This information is summarized in the following graph, where the Motor Vehicle data is overlaid for comparison to the total.



Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Please view the plot for this question. Does the plot appear to address the question being asked? In other words, can you answer the question using the information shown in the plot?

Upload the R code file for the plot uploaded in the previous question.

B | I | Link | <code> | Math | Edit: Rich | Preview

```
# function plot5()
#
# This function compares the motor vehicle sources of PM2.5 to all sources from
# 1999 - 2008 in Baltimore City, MD. It summarizes the results of that comparison in
# a table and graph.
#
# The graph is created using the base graphics package.
#
# It assumes the presence of a data frame with NEI data in the parent environment.
#
# If that is not there, it calls getAndCleanData() to read it in.
#

plot5 <- function() {

    library(dplyr)

# If source data is not already in parent environment, read in the data
if (!exists ("NEI")) {

    getAndCleanData()

}

# Create a subset of NEI data for Baltimore CIty, MD using NEI data from parent env
NEI.BC <- inner_join(SCC, NEI, by = "SCC")
NEI.BC <- subset(NEI.BC, fips == "24510", select = c(year, Short.Name, Emissions))
colnames(NEI.BC) <- c("Year", "Short.Name", "Emissions")

# Which rows are motor vehicle? (Assumption: Motor Vehicle = Highway Vehicle)
row <- grep ("Highway Veh", NEI.BC$Short.Name, ignore.case = TRUE)
NEI.BC.mv <- NEI.BC[row,]

# Calculate Total PM2.5 Emissions Baltimore City by year
totBC <- subset(NEI.BC, select = c(Year, Emissions))
totBC <- totBC %>% group_by(Year)
totBC <- summarize(totBC, sum(Emissions))
colnames(totBC) <- c("Year", "BC.PM2.5")

# Calculate Total PM2.5 Motor Vehicle Emissions for Baltimore CIty by year
```

```
totMV <- subset(NEI.BC.mv, select = c(Year, Emissions))
totMV <- totMV %>% group_by(Year)
totMV <- summarize(totMV, sum(Emissions))
colnames(totMV) <- c("Year", "from.MV")

# Compare and summarize Sources
totSum <- inner_join(totBC, totMV, by = "Year")
totSum$Percent <- with(totSum, from.MV/BC.PM2.5*100)
totSum <- transform(totSum, BC.PM2.5 = round(BC.PM2.5,), from.MV = round(from.MV,1),
                    Percent = round(Percent,2))

print(totSum, row.names = FALSE)

# Calculate y-offsets for text labels and generate labels
tot.ty <- totSum$BC.PM2.5
tot.labels <- as.character(totSum$BC.PM2.5)

mv.ty <- totSum$from.MV
mv.labels <- as.character(totSum$from.MV)

per.ty <- (tot.ty - mv.ty)/2
per.labels <- as.character(totSum$Percent)
per.labels <- paste(per.labels,"%")

    # Open image file
    png (filename = "./plot5.png")

    # Plot total emissions in Baltimore City
    bp1 <- barplot(height = totSum$BC.PM2.5, names.arg = totSum$Year, col = "green",
            density = 100, angle = 45, ylim = c(0,4000), xlab = "Year", ylab = "PM2.5 (Tons)",
            main = "Total Annual PM2.5 in Baltimore City, MD\n (Motor Vehicles vs. All Sources)" )

    # Plot motor vehicle emissions in Baltimore City
    bp2 <- barplot(height = totSum$from.MV, names.arg = totSum$Year, col = "blue",
            density = 100, angle = -45, ylim = c(0,4000), beside = TRUE, add = TRUE )

    legend("topright", legend = c("All Sources", "Motor Vehicles"),
           fill = c("green", "blue"))

    text(bp1, tot.ty, labels = tot.labels, col = "black", pos = 3)
    text(bp2, mv.ty, labels = mv.labels, col = "black", pos = 3)
    text(bp1, per.ty, labels = per.labels, col = "black", pos = 3)

    # Close file
    dev.off()
}


# function getAndCleanData ()
```

```
# function getAndCleanData ()
#
# This function checks for the presence of NEI data in the working directory.
# If it does not exist, it downloads and unzips the source .zip file.
# It then reads the source RDS data into NEI and SCC data frames and stores
# That data in the parent environment for used by other functions.
#

getAndCleanData <- function () {

    # If the project source data does not exist, then
    if (!file.exists("./summarySCC_PM25.rds")) {

        url <- "https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2FNEI_data.zip"
        destFile <- "exdata-data-NEI_data.zip"

        # Download File
        download.file(url, destfile = destFile, method = "curl")

        # Unzip File
        unzip(destFile)

        # Remove Zip File
        file.remove(destFile)

    }

    # Read In Data and save data frames to the parent environment
    NEI <<- readRDS("summarySCC_PM25.rds")
    SCC <<- readRDS("Source_Classification_Code.rds")

}

plot5 ()
```

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

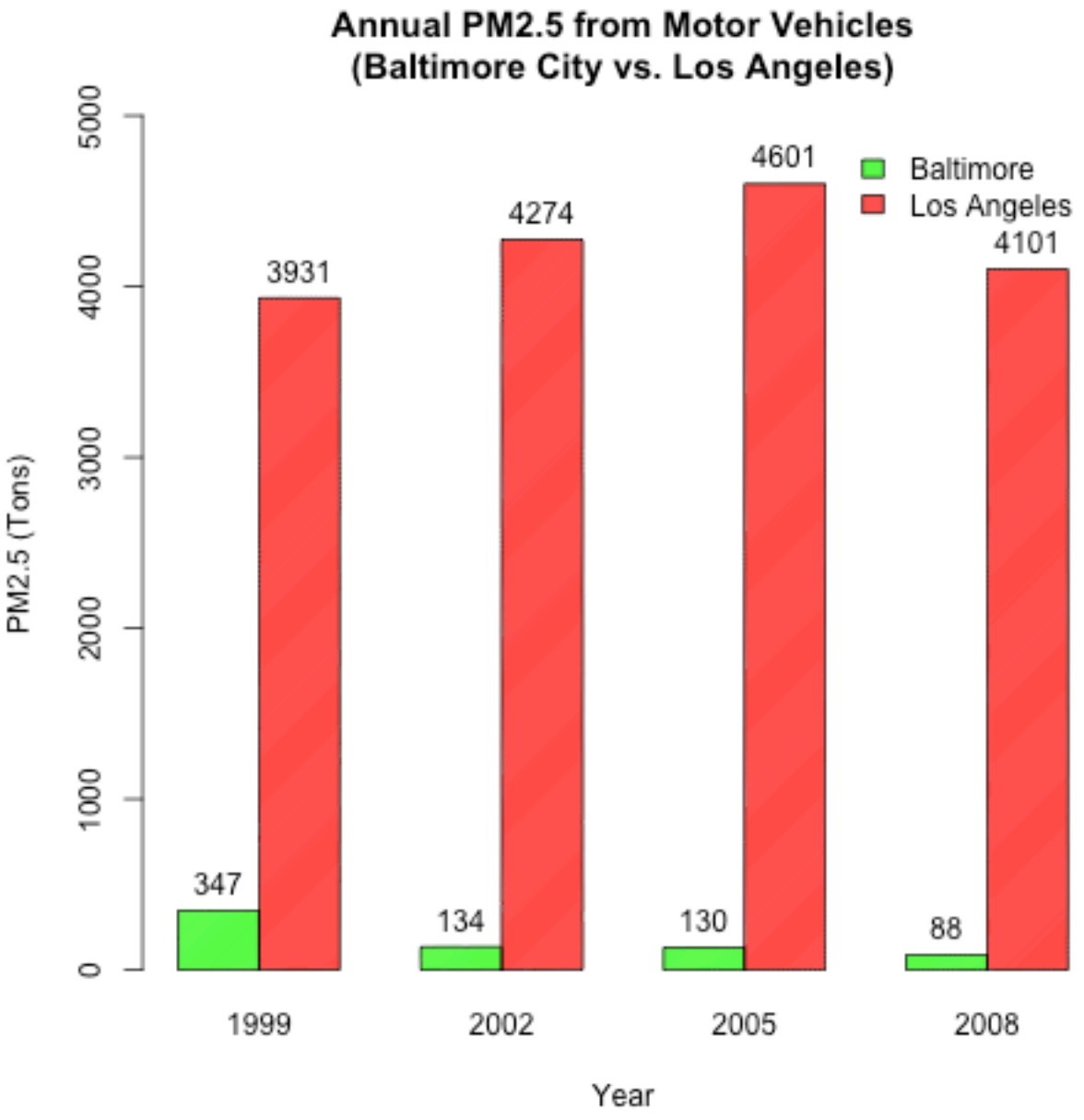
Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

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?

Upload a PNG file containing your plot addressing this question.

B | *I* | Link | <code> | Math | Edit: Rich | Preview

Analysis of NEI data from 1999 - 2008 shows that PM2.5 emissions from motor vehicles was **reduced by almost a factor 4 (3.94)** in Baltimore City, MD, while in Los Angeles, CA, the emissions actually rose by as much as 15% in 2005, dropping to 4% above the 1999 levels in 2008, *despite California's ridiculously more stringent emissions requirements*. The supporting information is shown figure below.



Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Please view the plot for this question. Does the plot appear to address the question being asked? In other words, can you answer the question using the information shown in the plot?

Upload the R code file for the plot uploaded in the previous question.

B | I | Link | <code> | Math | Edit: Rich | Preview

```
# function plot6()
#
# This function compares the motor vehicle sources of PM2.5 in Baltimore City, MD from
# 1999 - 2008 to those of Los Angeles County, CA. It summarizes the results of that
# comparison in a table and graph.
#
# The graph is created using the base graphics package.
#
# It assumes the presence of a data frame with NEI data in the parent environment.
#
# If that is not there, it calls getAndCleanData() to read it in.
#

plot6 <- function() {

    library(dplyr)

    # If source data is not already in parent environment, read in the data
    if (!exists ("NEI")) {

        getAndCleanData()

    }

    # Create a subset of NEI data for Baltimore CIty, MD using NEI data from parent env
    NEI.BC <- inner_join(SCC, NEI, by = "SCC")
    NEI.BC <- subset(NEI.BC, fips == "24510", select = c(year, Short Name, Emissions))
```

```
colnames(NEI.BC) <- c("Year", "Short.Name", "Emissions")
NEI.BC <- transform(NEI.BC, Year = factor(Year))

# Which rows are motor vehicle?
row <- grep ("Highway Veh", NEI.BC$Short.Name, ignore.case = TRUE)
NEI.BC <- NEI.BC[row,]

# Calculate Total PM2.5 Motor Vehicle Emissions for Baltimore CIty by year
totBC <- subset(NEI.BC, select = c(Year, Emissions))
totBC <- totBC %>% group_by(Year)
totBC <- summarize(totBC, sum(Emissions))
colnames(totBC) <- c("Year", "from.BC")

# Create a subset of NEI data for Los Angeles County, CA using NEI data from parent env
NEI.LA <- inner_join(SCC, NEI, by = "SCC")
NEI.LA <- subset(NEI.LA, fips == "06037", select = c(year, Short.Name, Emissions))
colnames(NEI.LA) <- c("Year", "Short.Name", "Emissions")
NEI.LA <- transform(NEI.LA, Year = factor(Year))

# Which rows are motor vehicle?
row <- grep ("Highway Veh", NEI.LA$Short.Name, ignore.case = TRUE)
NEI.LA <- NEI.LA[row,]

# Calculate Total PM2.5 Motor Vehicle Emissions for Los Angeles County by year
totLA <- subset(NEI.LA, select = c(Year, Emissions))
totLA <- totLA %>% group_by(Year)
totLA <- summarize(totLA, sum(Emissions))
colnames(totLA) <- c("Year", "from.LA")

# Compare and summarize sources
totSum <- inner_join(totBC, totLA, by = "Year")
totSum <- transform(totSum, from.BC = round(from.BC,0), from.LA = round(from.LA,0))

print(totSum, row.names = FALSE)

# Create matrix for plot heights
a <- matrix(c(totSum$from.BC, totSum$from.LA), nrow = 4, ncol = 2)
a <- t(a)

bc.ty <- totSum$from.BC
bc.labels <- as.character(totSum$from.BC)

la.ty <- totSum$from.LA
la.labels <- as.character(totSum$from.LA)

# Open image file
png (filename = "./plot6.png")
```

```
    # Plot
    bp <- barplot(height = a, names.arg = totSum$Year, col = c("green","red"), density = 100,
            angle = 45, xlab = "Year", ylab = "PM2.5 (Tons)", ylim = c(0,5000),
            main = "Annual PM2.5 from Motor Vehicles\n(Baltimore City vs. Los Angeles)",
            beside = TRUE, legend.text = c("Baltimore", "Los Angeles"),
            args.legend = list(bty = "n"))

    text(bp[1,], y = bc.ty, labels = bc.labels, pos = 3)
    text(bp[2,], y = la.ty, labels = la.labels, pos = 3)

    # Close file
    dev.off()
}


# function getAndCleanData ()
#
# This function checks for the presence of NEI data in the working directory.
# If it does not exist, it downloads and unzips the source .zip file.
# It then reads the source RDS data into NEI and SCC data frames and stores
# That data in the parent environment for used by other functions.
#

getAndCleanData <- function () {

    # If the project source data does not exist, then
    if (!file.exists("./summarySCC_PM25.rds")) {

        url <- "https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2FNEI_data.zip"
        destFile <- "exdata-data-NEI_data.zip"

        # Download File
        download.file(url, destfile = destFile, method = "curl")

        # Unzip File
        unzip(destFile)

        # Remove Zip File
        file.remove(destFile)

    }

    # Read In Data and save data frames to the parent environment
    NEI <<- readRDS("summarySCC_PM25.rds")
    SCC <<- readRDS("Source_Classification_Code.rds")

}
```

plot6 ()

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Examine the submitted R code file. Does the R code appear to construct the plot shown in the previous question? NOTE: Do not run the code on your own computer.

Overall evaluation/feedback

Note: this section can only be filled out during the evaluation phase.

Please use the space below to provide constructive feedback to the student who submitted the work. Point out the submission's strengths as well as areas in need of improvement. You may also use this space to explain your grading decisions.

You've written 0 words

✔ Submitted. You can still make changes and re-submit before the deadline.

☑ In accordance with the Honor Code, I certify that my answers here are my own work, and that I have appropriately acknowledged all external sources (if any) that were used in this work.

Re-submit for grading