due in 4day 10h

Submission Phase

1. Do assignment ☑ (/exdata-011/human_grading/view/courses/973505/assessments/4/submissions)

Evaluation Phase

2. Evaluate peers 🔒 (/exdata-011/human_grading/view/courses/973505/assessments/4/peerGradingSets)

Results Phase

3. See results 🔒 (/exdata-011/human_grading/view/courses/973505/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 (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 (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 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.

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:

PM2.5 Emissions Data (`summarySCC_PM25.rds`): This file contains a data frame with all of the PM2.5 emissions data for 1999, 2002, 2005, and 2008. For each year, the table contains number of **tons** of PM2.5 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 PM2.5 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 in the Emissions table to the actual name of the PM2.5 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 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?

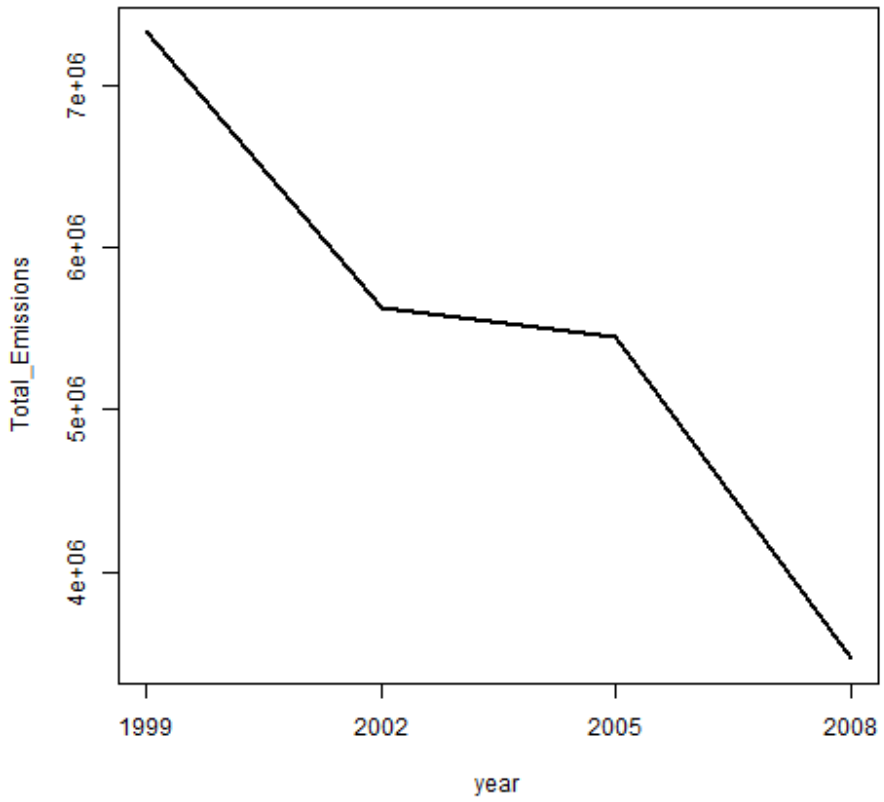
Making and Submitting Plots

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 must 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



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

```
plot1 <- function() {
    library(dplyr)
    library(data.table)

    # Read the files into local variables
    NEI <- readRDS("summarySCC_PM25.rds")
    SCC <- readRDS("Source_Classification_Code.rds")

    NEI <- data.table(NEI)
    #make a new data frame with just year and Emissions from all sources
    #for the years 1999, 2002, 2005 and 2008
    mydf <- subset(NEI, select=c(Emissions, year))

    #total PM2.5 emissions from all sources
    sumdt <- summarise(group_by(mydf, year), sum(Emissions))

    setnames(sumdt, old="sum(Emissions)", new="Total_Emissions")
    png(filename="plot1.png", width=480, height=480)
    #par(mar=c(4, 4, 2, 2))
    with(sumdt, plot(year, Total_Emissions, type="l", lwd=2, axes=FALSE))
    axis(side=1, at=c(1999, 2002, 2005, 2008))
    axis(side=2)
    box()
    dev.off()
}
```

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

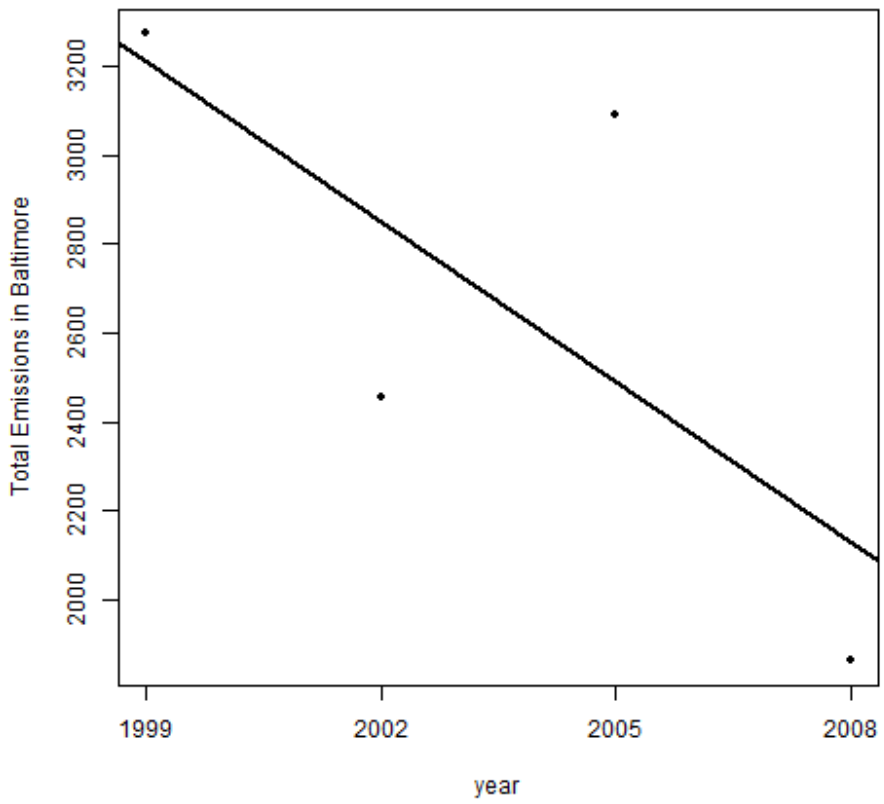
Evaluation/feedback on the above work

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.



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.

```
plot2 <- function() {
   library(dplyr)
   library(data.table)

   # Read the files into local variables
   NEI <- readRDS("summarySCC_PM25.rds")
   SCC <- readRDS("Source_Classification_Code.rds")

   NEI <- data.table(NEI)

   #get data for Balitmore i.e. fips = 24510
   mydf <- subset(NEI, fips=="24510", select=c(Emissions, year))

   sumdt <- summarise(group_by(mydf, year), sum(Emissions))

   setnames(sumdt, old="sum(Emissions)", new="Total_Emissions_in_Baltimore")
   png(filename="plot2.png", width=480, height=480)
   #plot with points and draw a smoother
   with(sumdt, plot(year, Total_Emissions_in_Baltimore, pch=20,
                    axes=FALSE, ylab="Total Emissions in Baltimore"))

   axis(side=1, at=c(1999, 2002, 2005, 2008))
   axis(side=2)
   box()
   model <- lm(Total_Emissions_in_Baltimore ~ year, sumdt)
   abline(model, lwd=2)
   dev.off()

}
```

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

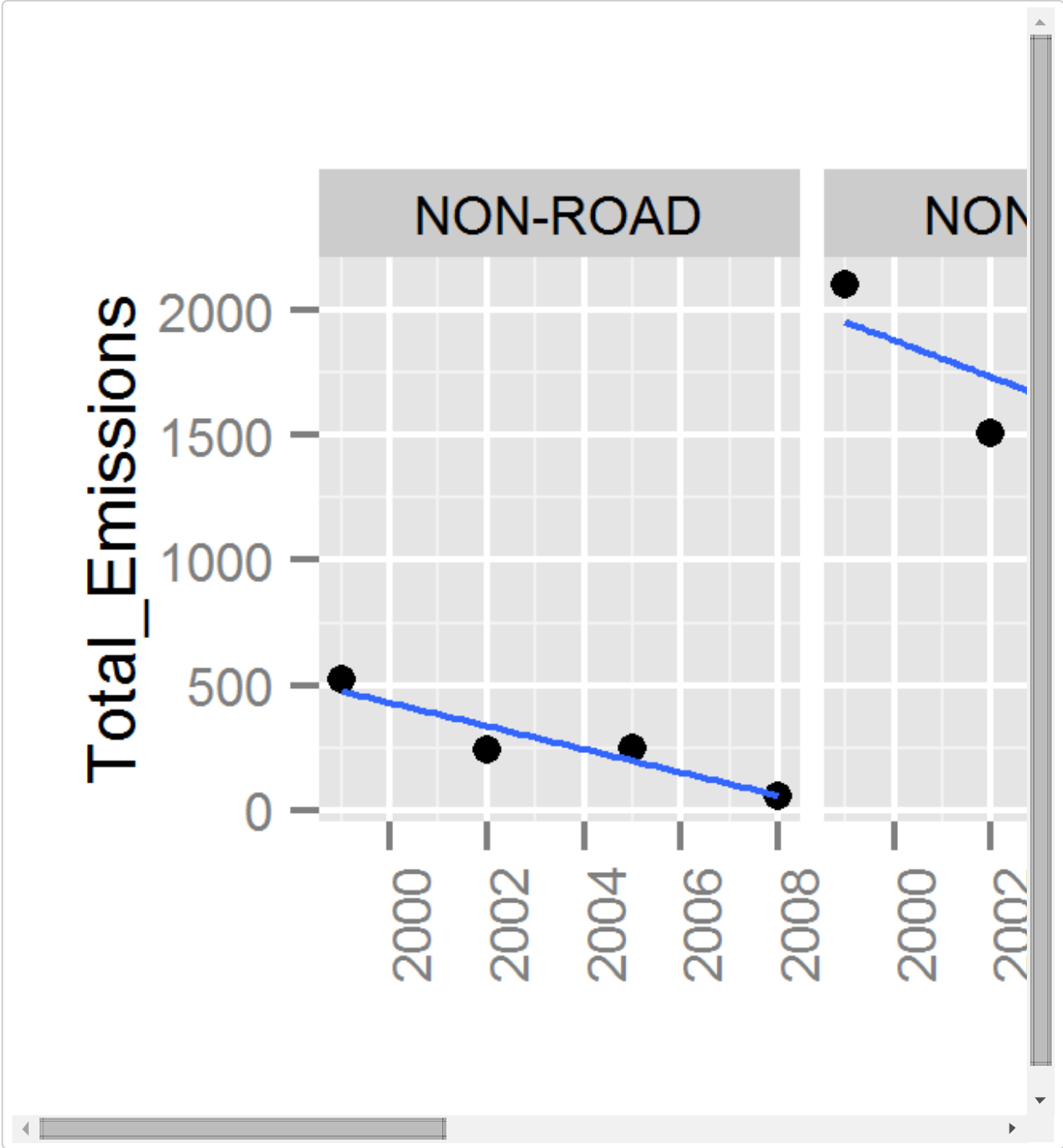
Evaluation/feedback on the above work

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.



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

```
plot3 <- function() {
    library(dplyr)
    library(data.table)
    library(ggplot2)

    # Read the files into local variables
    NEI <- readRDS("summarySCC_PM25.rds")
    SCC <- readRDS("Source_Classification_Code.rds")

    NEI <- data.table(NEI)

    #select Emissions, year and type for Baltimore
    mydf <- subset(NEI, fips=="24510", select=c(Emissions, year, type))

    #sum by year and type, so that we get the total emissions across
    #various sources (or types)
    sumdt <- summarise(group_by(mydf, year, type), sum(Emissions))
    setnames(sumdt, old="sum(Emissions)", new="Total_Emissions")

    #p <- qplot(year, Total_Emissions, data=sumdt, facets=.~type,
    #           geom=c("point", "smooth"), method=lm, se=FALSE)

    #draw the plot, add the facets and smoother.
    p <- ggplot(sumdt, aes(x=year, y=Total_Emissions))
    p <- p + geom_point() + facet_grid(.~type) +
        geom_smooth(method="lm", se=FALSE)
    # change the orientation of xlab to make it more readable
    p <- p + theme(text = element_text(size=10),
                   axis.text.x=element_text(angle=90, vjust=1))

    ggsave(filename="plot3.png", plot=p, width=5, height=2.2)

}
```

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

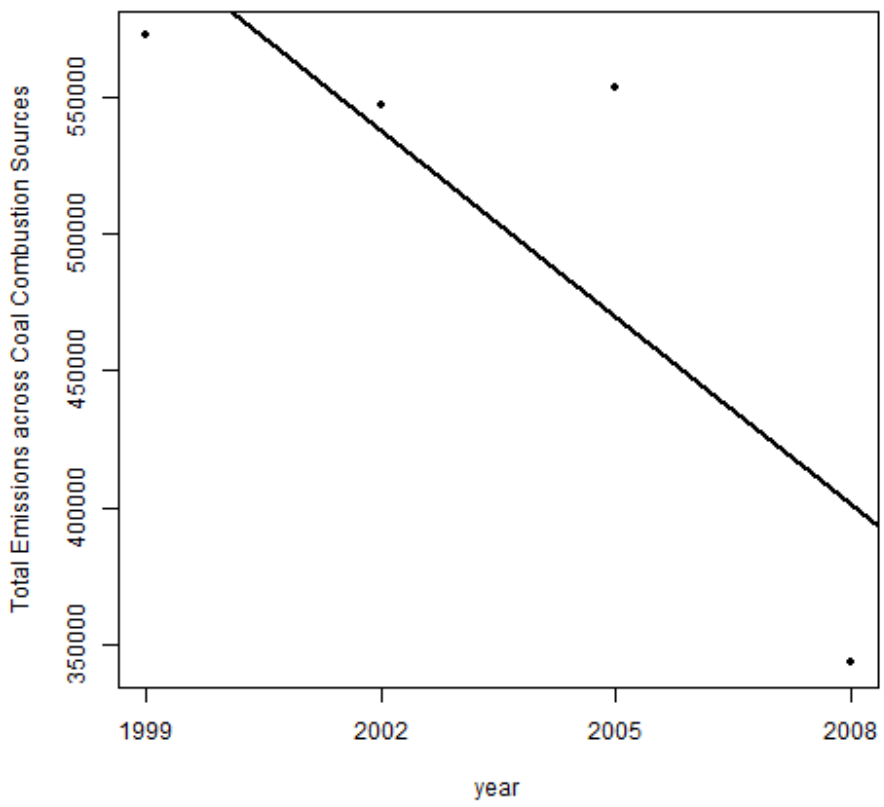
Evaluation/feedback on the above work

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.



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.

```
plot4 <- function() {
    library(dplyr)
    library(data.table)
    library(ggplot2)

    # Read the files into local variables
    NEI <- readRDS("summarySCC_PM25.rds")
    SCC <- readRDS("Source_Classification_Code.rds")

    NEI <- data.table(NEI)


    #Problem : Across the United States, how have emissions from
    #coal combustion-related sources changed from 1999–2008?
    #Here we need to define what "coal combustion-related sources" are
    #When I look at the SCC$EI.Sector column, I see that the most relevant
    #ones are the records "Fuel Comb - Electric Generation - Coal"
    #"Fuel Comb - Industrial Boilers, ICEs - Coal" and
    #"Fuel Comb - Comm/Institutional - Coal"

    #So first I will filter out these records from SCC, there are 99 such
    #records

    ccSCC <- SCC[SCC$EI.Sector == "Fuel Comb - Electric Generation - Coal"
                 |SCC$EI.Sector == "Fuel Comb - Industrial Boilers, ICEs -
Coal"
                 |SCC$EI.Sector == "Fuel Comb - Comm/Institutional - Coal",
 ]



    #now filter NEI based on ccSCC$SCC to get data only for coal combustion
    #related sources
    ccNEI <- NEI[NEI$SCC %in% ccSCC$SCC, ]

    #Now get the total of emissions across the year
    sumdt <- summarise(group_by(ccNEI, year), sum(Emissions))
    setnames(sumdt, old="sum(Emissions)", new="Total_Emissions")

    #draw a base plot showing the data
    png(filename="plot4.png", width=480, height=480)
    #plot with points and draw a smoother
    with(sumdt, plot(year, Total_Emissions, pch=20,
                     axes=FALSE, ylab="Total Emissions across Coal Combusti
on Sources"))

    axis(side=1, at=c(1999, 2002, 2005, 2008))
    axis(side=2)
```

```
    box()
    model <- lm(Total_Emissions ~ year, sumdt)
    abline(model, lwd=2)
    dev.off()
}
```

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

Evaluation/feedback on the above work

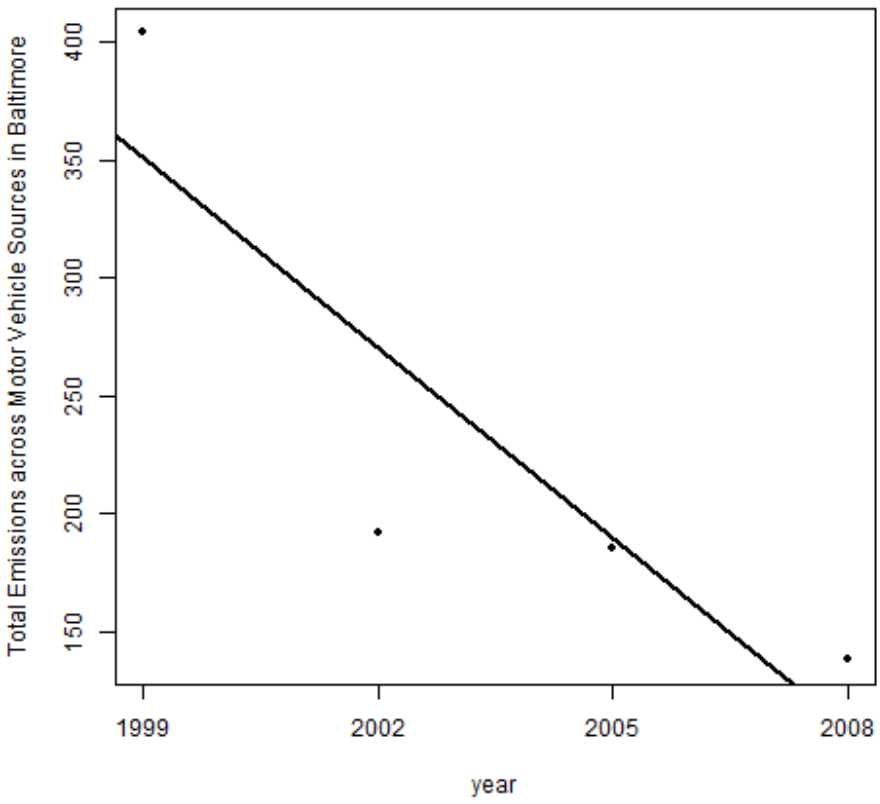
Note: this section can only be filled out during the evaluation phase.

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



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.

```
plot5 <- function() {
    library(dplyr)
    library(data.table)
    library(ggplot2)

    # Read the files into local variables
    NEI <- readRDS("summarySCC_PM25.rds")
    SCC <- readRDS("Source_Classification_Code.rds")

    NEI <- data.table(NEI)

    #Problem: How have emissions from motor vehicle sources changed from
    #1999–2008 in Baltimore City?
    #We first get the SCC codes relevant to Motor Vehicle sources
    #To do this, we see that SCC.Level.One should be "Mobile Sources"
    #and SCC.Level.Two should be any highway or off-highway vehicle and we
    #should exclude aircraft etc. Lets do this now
    msSCC <- SCC[SCC$SCC.Level.One == "Mobile Sources", ]

    mvSCC <- msSCC[msSCC$SCC.Level.Two == "Highway Vehicles - Gasoline"
                   |msSCC$SCC.Level.Two == "Highway Vehicles - Diesel"
                   |msSCC$SCC.Level.Two == "Off-highway Vehicle Gasoline, 2
-Stroke"
                   |msSCC$SCC.Level.Two == "Off-highway Vehicle Gasoline, 4
-Stroke"
                   |msSCC$SCC.Level.Two == "Off-highway Vehicle Diesel", ]

    #Now, we will filter NEI based the SCC in the motor vehicle sources
    mvNEI <- NEI[NEI$SCC %in% mvSCC$SCC, ]

    #Next we filter mvNEI to have data only for Baltimore
    mydf <- subset(mvNEI, fips=="24510", select=c(Emissions, year))

    #Now get the total of emissions across the years
    sumdt <- summarise(group_by(mydf, year), sum(Emissions))
    setnames(sumdt, old="sum(Emissions)", new="Total_Emissions")

    #draw a base plot showing the data
    png(filename="plot5.png", width=480, height=480)
    #plot with points and draw a smoother
    with(sumdt, plot(year, Total_Emissions, pch=20,
                     axes=FALSE,
                     ylab="Total Emissions across Motor Vehicle Sources in
Baltimore"))

    axis(side=1, at=c(1999, 2002, 2005, 2008))
    axis(side=2)
```

```
    box()
    model <- lm(Total_Emissions ~ year, sumdt)
    abline(model, lwd=2)
    dev.off()
}
```

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

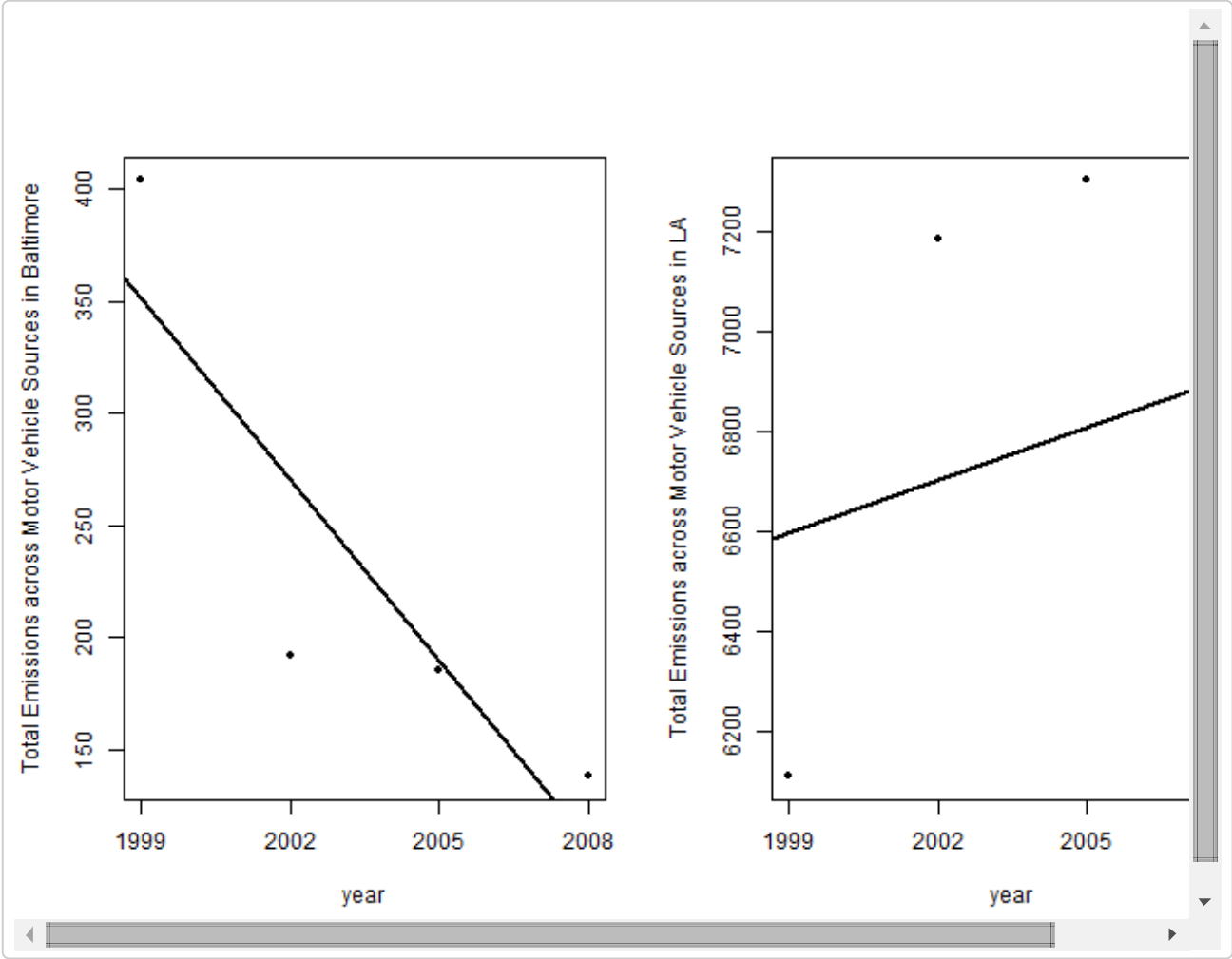
Evaluation/feedback on the above work

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.



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.

```
plot6 <- function() {
    library(dplyr)
    library(data.table)

    # Read the files into local variables
    NEI <- readRDS("summarySCC_PM25.rds")
    SCC <- readRDS("Source_Classification_Code.rds")

    NEI <- data.table(NEI)

    #Problem: Compare emissions from motor vehicle sources in Baltimore Cit
y
    #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?
    #We first get the SCC codes relevant to Motor Vehicle sources
    #To do this, we see that SCC.Level.One should be "Mobile Sources"
    #and SCC.Level.Two should be any highway or off-highway vehicle and we
    #should exclude aircraft etc. Lets do this now
    msSCC <- SCC[SCC$SCC.Level.One == "Mobile Sources", ]

    mvSCC <- msSCC[msSCC$SCC.Level.Two == "Highway Vehicles - Gasoline"
                   |msSCC$SCC.Level.Two == "Highway Vehicles - Diesel"
                   |msSCC$SCC.Level.Two == "Off-highway Vehicle Gasoline, 2
-Stroke"
                   |msSCC$SCC.Level.Two == "Off-highway Vehicle Gasoline, 4
-Stroke"
                   |msSCC$SCC.Level.Two == "Off-highway Vehicle Diesel", ]

    #Now, we will filter NEI based the SCC in the motor vehicle sources
    mvNEI <- NEI[NEI$SCC %in% mvSCC$SCC, ]

    #Next we filter mvNEI to have data only for Baltimore
    mvNEI_b <- subset(mvNEI, fips=="24510", select=c(Emissions, year))

    #Next we filter mvNEI to have data only for Log Angeles
    mvNEI_la <- subset(mvNEI, fips=="06037", select=c(Emissions, year))

    #Now get the total of emissions across the years in baltimore
    sumdt_b <- summarise(group_by(mvNEI_b, year), sum(Emissions))
    setnames(sumdt_b, old="sum(Emissions)", new="Total_Emissions_in_Baltimo
re")

    #Now get the total of emissions across the years in los angeles
    sumdt_la <- summarise(group_by(mvNEI_la, year), sum(Emissions))
    setnames(sumdt_la, old="sum(Emissions)", new="Total_Emissions_in_LA")
```

```
    # To show the changes for the two counties, we will draw 2 plots
    # and draw a smoother. Visually you can make out which county has seen
    # greater changes in the emissions

    png(filename="plot6.png", width=700, height=480)
    #since we will be drawing 2 plots, set mfrow
    par(mfrow=c(1,2))

    #plot with points and draw a smoother for Baltimore
    with(sumdt_b, plot(year, Total_Emissions_in_Baltimore, pch=20,
                       axes=FALSE,
                       ylab="Total Emissions across Motor Vehicle Sources in
Baltimore"))

    axis(side=1, at=c(1999, 2002, 2005, 2008))
    axis(side=2)
    box()
    model <- lm(Total_Emissions_in_Baltimore ~ year, sumdt_b)
    abline(model, lwd=2)

    #plot with points and draw a smoother for LA
    with(sumdt_la, plot(year, Total_Emissions_in_LA, pch=20,
                        axes=FALSE,
                        ylab="Total Emissions across Motor Vehicle Sources i
n LA"))

    axis(side=1, at=c(1999, 2002, 2005, 2008))
    axis(side=2)
    box()
    model <- lm(Total_Emissions_in_LA ~ year, sumdt_la)
    abline(model, lwd=2)
```

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