

CS4048 Data Science

Friday, November 26, 2021

Course Instructor

Mr. Adeel Ashraf Cheema

Mr. Muhammad Usman Joyia

Serial No:

2nd Mid Term Exam

Total Time: 2 Hour

Total Marks: 60

Signature of Invigilator

17F8148	7C	Sania Kashif
Roll No	Section	Signature

DO NOT OPEN THE QUESTION BOOK OR START UNTIL INSTRUCTED.

Instructions:

1. Read the questions carefully for clarity of context and understanding of meaning and make assumptions wherever required, for neither the invigilator will address your queries, nor the teacher/examiner will come to the examination hall for any assistance.
2. Fit in all your answers in the provided space. You may use extra space on the last page if required. If you do so, clearly mark question/part number on that page to avoid confusion.
3. Use only your own stationery and calculator. If you do not have your own calculator, use manual calculations.
4. Use only permanent ink-pens. Only the questions attempted with permanent ink-pens will be considered. Any part of paper done in lead pencil cannot be claimed for checking/rechecking.

	Q-1	Q-2	Q-3	Q-4	Q-5	Q-6	Total
Total Marks	10	10	10	10	10	10	60
Marks Obtained							

Vetted By: _____ **Vetter Signature:** _____

University Answer Sheet Required: No ☒ Yes ☐

Question 1: Have total emissions from PM_{2.5} decreased in the United States from 1999 to 2008? Using the 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.

Code:

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

totalNEI <- aggregate(Emissions ~ year, NEI, sum)

plot(totalNEI$year, totalNEI$Emissions, type = "o", col = "steelblue3", main = expression("Total US
"~ PM[2.5]~ "Emissions by Year"), ylab = expression("Total US "~ PM[2.5] ~ "Emissions"), xlab =
"Year")
```

Answer:

Yes, total emissions from PM_{2.5} have decreased in the range 1999-2008.

Output plots

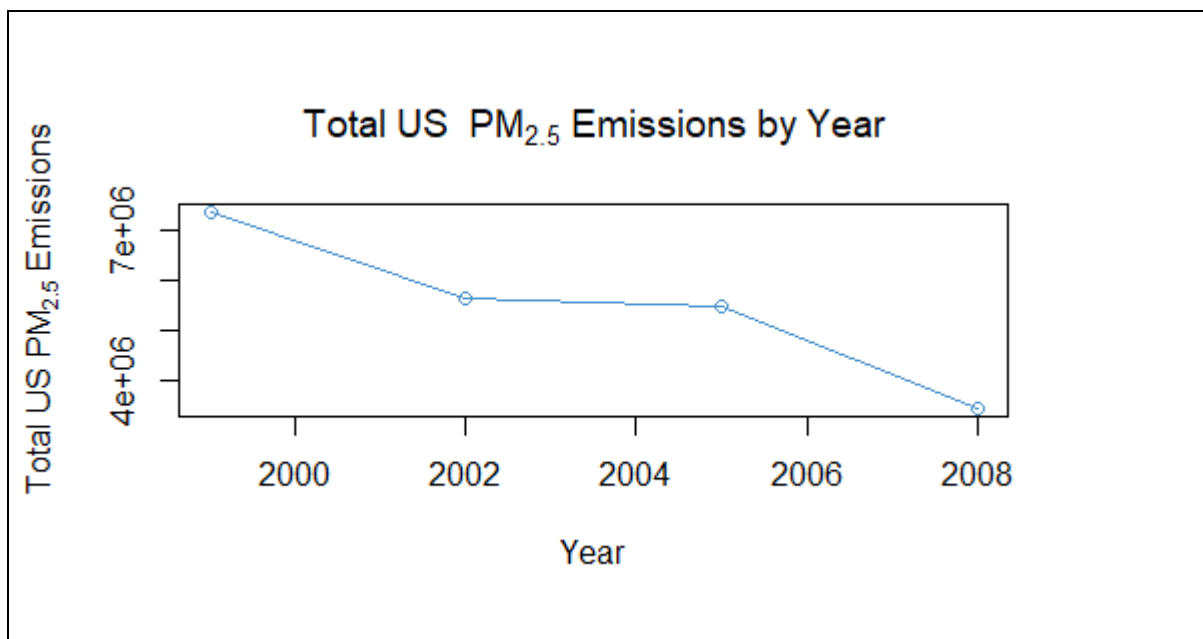


Figure 1 Total US Emissions by year

Question 2: Have total emissions from PM_{2.5} decreased in the **Baltimore City**, Maryland (**fips == "24510"**) from 1999 to 2008? Use the plotting system to make a plot answering this question.

Code:

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

baltimore <- subset(NEI, NEI$fips == "24510")
```

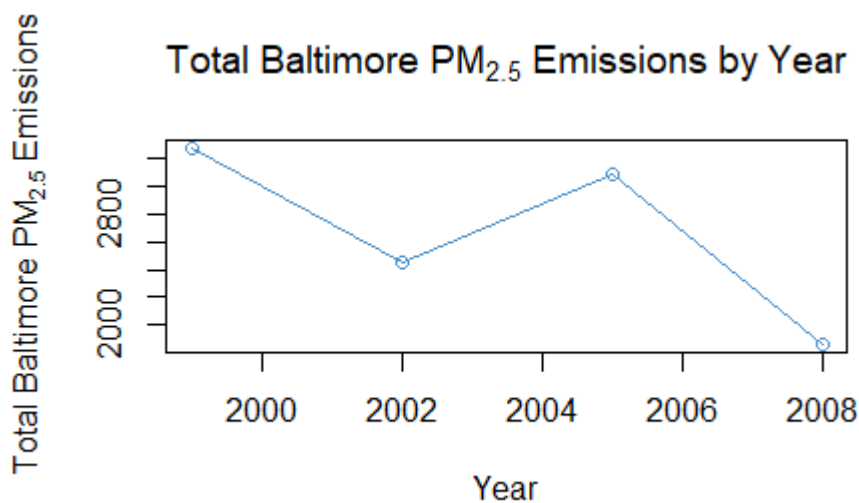
Answer:

Yes, total emissions from PM_{2.5} in the Baltimore region have also decreased, though the trend is less clear.

```
totalBaltimore <- aggregate(Emissions ~ year, baltimore, sum)
```

```
plot(totalBaltimore$year, totalBaltimore$Emissions, type = "o", main = expression("Total  
Baltimore" ~ PM[2.5] ~ "Emissions by Year"), xlab = "Year", ylab = expression("Total Baltimore "~  
PM[2.5] ~ "Emissions"), col = "steelblue3")
```

Output plots



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.

Code:

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

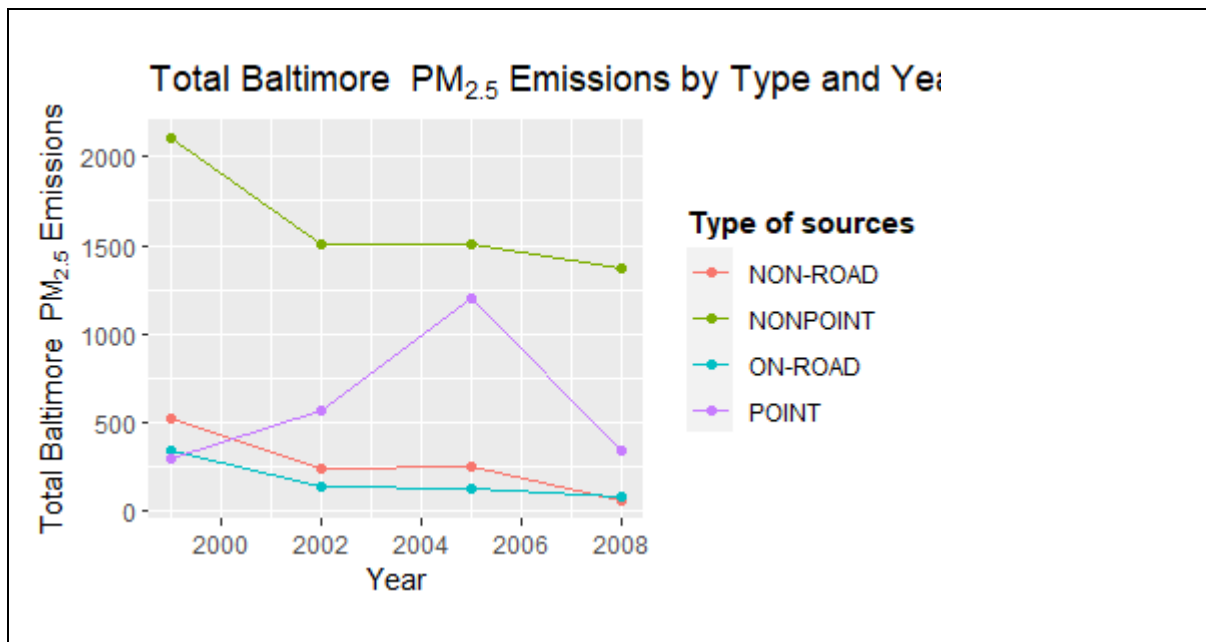
baltimore <- subset(NEI, NEI$fips == "24510")
baltimoreType <- aggregate(Emissions ~ year + type, baltimore, sum)

ggplot(baltimoreType, aes(year, Emissions, col = type)) +
  geom_line() +
  geom_point() +
  ggtitle(expression("Total Baltimore " ~ PM[2.5] ~ "Emissions by Type and Year")) +
  ylab(expression("Total Baltimore " ~ PM[2.5] ~ "Emissions")) +
  xlab("Year") +
  scale_colour_discrete(name = "Type of sources") +
  theme(legend.title = element_text(face = "bold"))
```

Answer:

Nonpoint and non-road emissions have consistently decreased in Baltimore during the period 1999–2008. On-road emissions are consistently low over the time period. Point emissions are inconsistent over the time period.

Output plots



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

Code:

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

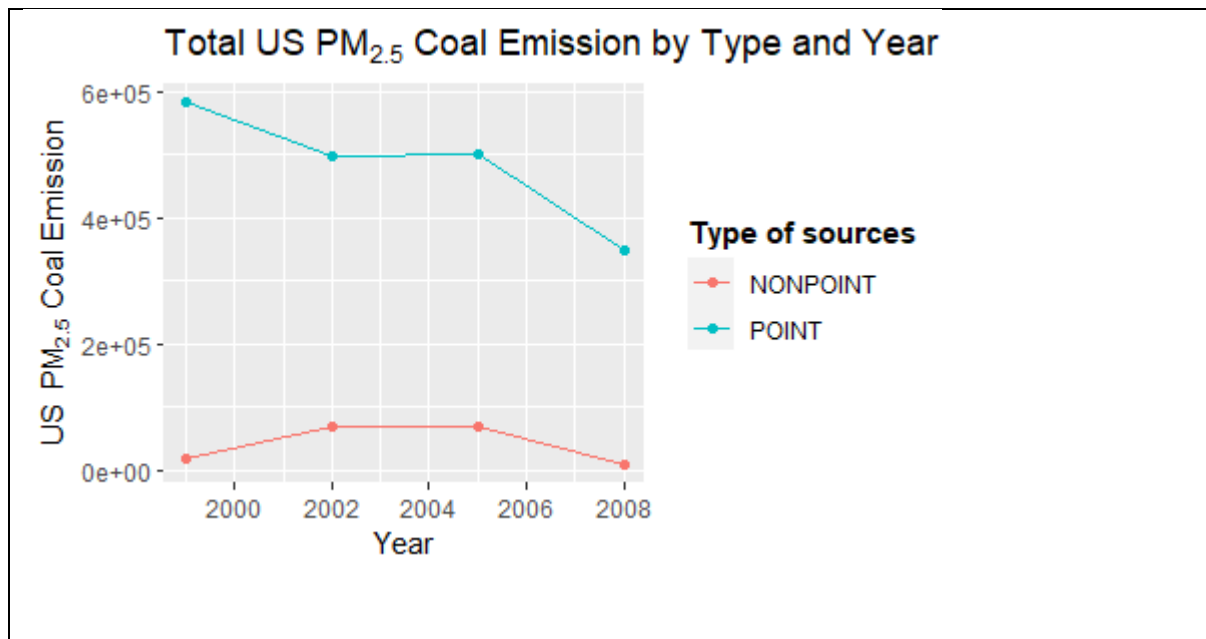
SCCcoal <- SCC[grepl("coal", SCC$Short.Name, ignore.case = T),]
NEIcoal <- NEI[NEI$SCC %in% SCCcoal$SCC,]
totalCoal <- aggregate(Emissions ~ year + type, NEIcoal, sum)

ggplot(totalCoal, aes(year, Emissions, col = type)) +
  geom_line() +
  geom_point() +
  ggtitle(expression("Total US" ~ PM[2.5] ~ "Coal Emission by Type and Year")) +
  xlab("Year") +
  ylab(expression("US " ~ PM[2.5] ~ "Coal Emission")) +
  scale_colour_discrete(name = "Type of sources") +
  theme(legend.title = element_text(face = "bold"))
```

Answer:

Point-source coal emissions have decreased in 2008 compared to previous years, which may be indicative of a downward trend. Non-point sources are inconsistent along years and more difficult to interpret, but a sharp decline in 2008 may indicate a downward trend as well.

Output plots



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

Code:

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

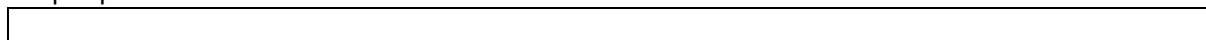
baltimoreMotor <- subset(NEI, NEI$fips == "24510" & NEI$type == "ON-ROAD")
baltimoreMotorAGG <- aggregate(Emissions ~ year, baltimoreMotor, sum)

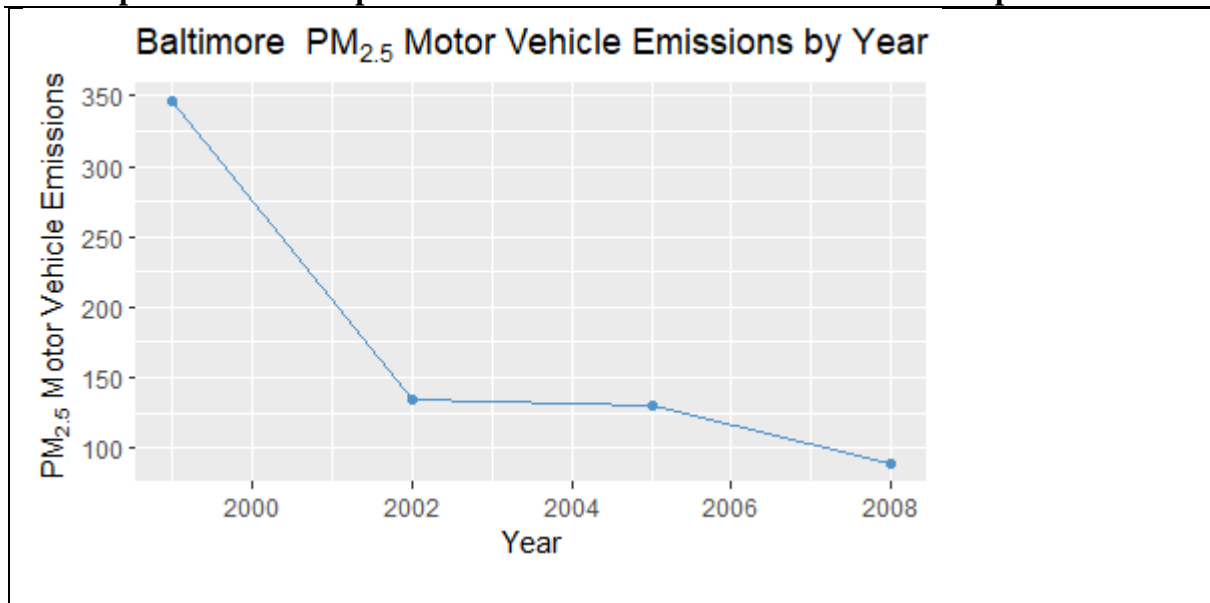
ggplot(baltimoreMotorAGG, aes(year, Emissions)) +
  geom_line(col = "steelblue3") +
  geom_point(col = "steelblue3") +
  ggtitle(expression("Baltimore " ~ PM[2.5] ~ "Motor Vehicle Emissions by Year")) +
  xlab("Year") +
  ylab(expression(~PM[2.5]~ "Motor Vehicle Emissions"))
```

Answer:

Heavy-duty diesel vehicles have shown a sharp decline in PM_{2.5} emissions over the time period. Heavy-duty gasoline vehicle emissions have also declined. Light-duty vehicles of both types remain consistently low.

Output plots





Question 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?

Code:

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

baltLosAngelesMotors <- subset(NEI, NEI$fips %in% c("24510", "06037") & NEI$type == "ON-ROAD")
baltLosAngelesMotorsAGG <- aggregate(Emissions ~ year + fips, baltLosAngelesMotors, sum)

ggplot(baltLosAngelesMotorsAGG, aes(year, Emissions, col = fips)) +
  geom_line() +
  geom_point() +
  ggtitle(expression("Baltimore and Los Angeles" ~ PM[2.5] ~ "Motor Vehicle Emissions by Year"))
+
  labs(x = "Year", y = expression(~PM[2.5]~ "Motor Vehicle Emissions")) +
  scale_colour_discrete(name = "City", labels = c("Los Angeles", "Baltimore")) +
  theme(legend.title = element_text(face = "bold"))
```

Answer:

Los Angeles has seen greater changes over time.

Output plots

