pm2.5 eda

2023-11-13

EDA done based on questions for Course Project 2 for Exploratory Data Analysis course by John Hopkins University

Loading data

```
dt <- fread("data/merged.csv") # run prep_data to create this file
```

Data consists of a summary data frame and a codetable data frame, which have been merged into merged.csv. Summary contains all of the PM2.5 commissions data for 1999, 2002, 2005, 2008. The code table provides a mapping from the SCC digit strings in the summary table to the actual name of the PM2.5 source.

```
summary(dt)
```

```
##      SCC                fips            Pollutant           Emissions
##  Length:6497651     Length:6497651     Length:6497651     Min.   :     0.0
##  Class :character   Class :character   Class :character   1st Qu.:     0.0
##  Mode  :character   Mode  :character   Mode  :character   Median :     0.0
##                                                           Mean   :     3.4
##                                                           3rd Qu.:     0.1
##                                                           Max.   :646952.0
##      type                year       Short.Name         EI.Sector
##  Length:6497651     Min.   :1999   Length:6497651     Length:6497651
##  Class :character   1st Qu.:2002   Class :character   Class :character
##  Mode  :character   Median :2005   Mode  :character   Mode  :character
##                     Mean   :2004
##                     3rd Qu.:2008
##                     Max.   :2008
```

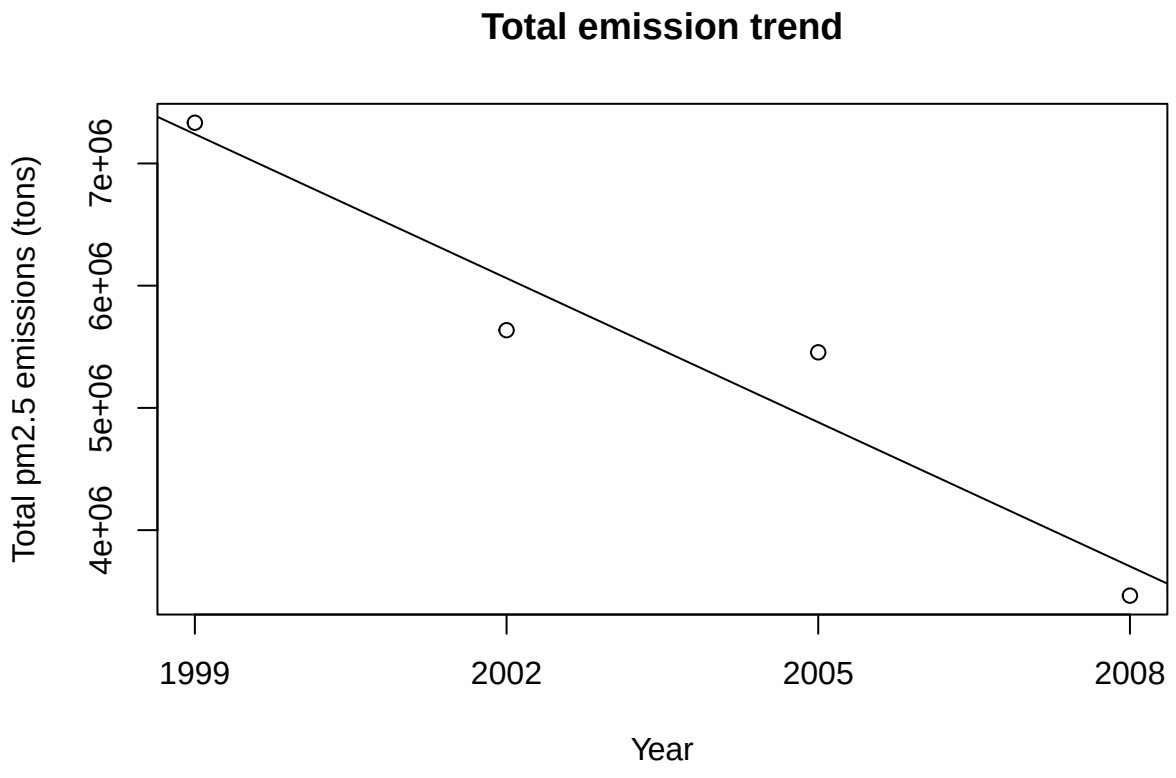
```
str(dt)
```

```
## Classes 'data.table' and 'data.frame':   6497651 obs. of  8 variables:
##  $ SCC       : chr  "10100101" "10100101" "10100101" "10100101" ...
##  $ fips      : chr  "34017" "01123" "08041" "42109" ...
##  $ Pollutant : chr  "PM25-PRI" "PM25-PRI" "PM25-PRI" "PM25-PRI" ...
##  $ Emissions : num  898.42 0.08 2.48 58.61 131.8 ...
##  $ type      : chr  "POINT" "POINT" "POINT" "POINT" ...
##  $ year      : int  1999 2002 1999 2002 2005 1999 2005 2005 2008 1999 ...
##  $ Short.Name: chr  "Ext Comb /Electric Gen /Anthracite Coal /Pulverized Coal" "Ext Comb /Electric G
##  $ EI.Sector : chr  "Fuel Comb - Electric Generation - Coal" "Fuel Comb - Electric Generation - Coal
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
slice_sample(dt, n=5)
```

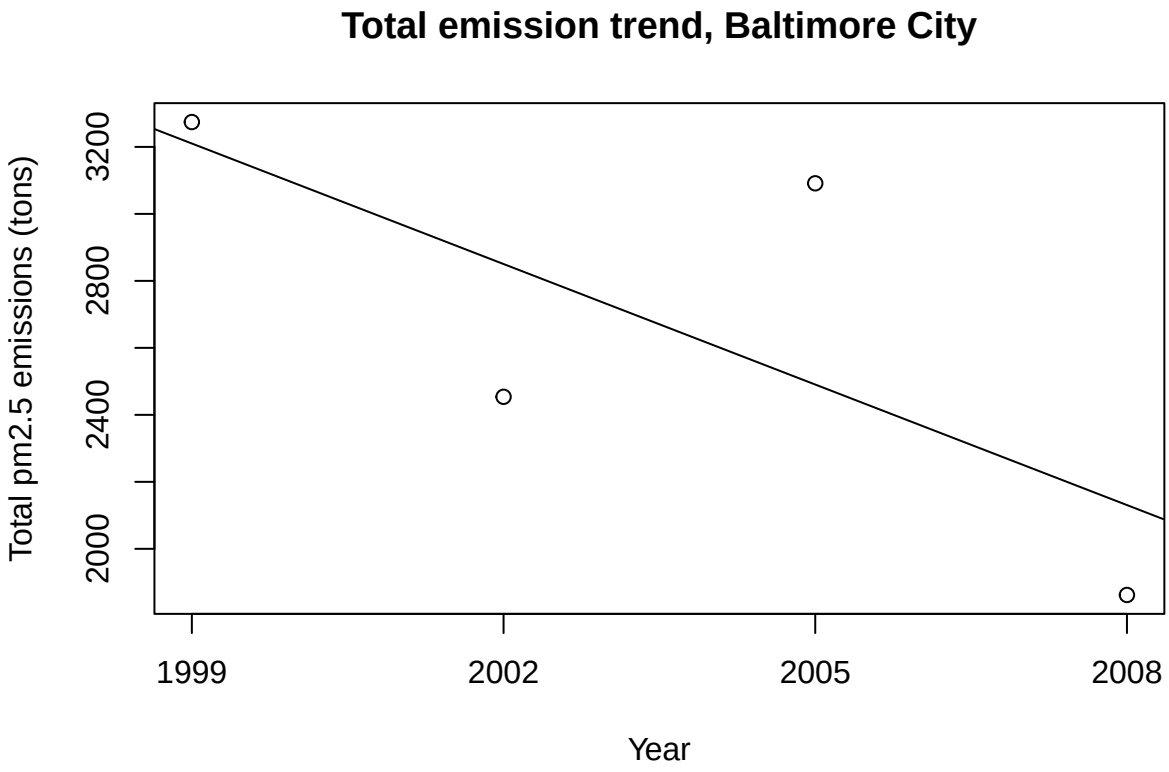
```
##           SCC  fips Pollutant  Emissions     type year
## 1: 2265001030 22053  PM25-PRI 0.26000000 NON-ROAD 2002
## 2: 2265002081 35047  PM25-PRI 0.00000000 NON-ROAD 1999
## 3: 2230072150 41043  PM25-PRI 0.00217850  ON-ROAD 2008
## 4: 2230074170 19037  PM25-PRI 0.26330000  ON-ROAD 2008
## 5: 2201080170 01125  PM25-PRI 0.00579317  ON-ROAD 2008
##                                                                                Short.Name
## 1:               Off-highway Gasoline, 4-Stroke /Recreational Equipt /All Terrain Vehicles
## 2: Off-highway Gasoline, 4-Stroke /Construction & Mining Equipt /Other Construction Equipt
## 3:  Highway Veh - Diesel - Heavy Duty (HDDV) Class 3, 4, & 5 - Rural Minor Arterial: Total
## 4:   Highway Veh - Diesel - Heavy Duty (HDDV) Class 8A & 8B - Rural Major Collector: Total
## 5:               Highway Veh - Gasoline - Motorcycles (MC) - Rural Major Collector: Total
##                                    EI.Sector
## 1:        Mobile - Non-Road Equipment - Gasoline
## 2:        Mobile - Non-Road Equipment - Gasoline
## 3:   Mobile - On-Road Diesel Heavy Duty Vehicles
## 4:   Mobile - On-Road Diesel Heavy Duty Vehicles
## 5: Mobile - On-Road Gasoline Light Duty Vehicles
```

Question 1: Have total emissions from PM2.5 decreased from 1999 to 2008?



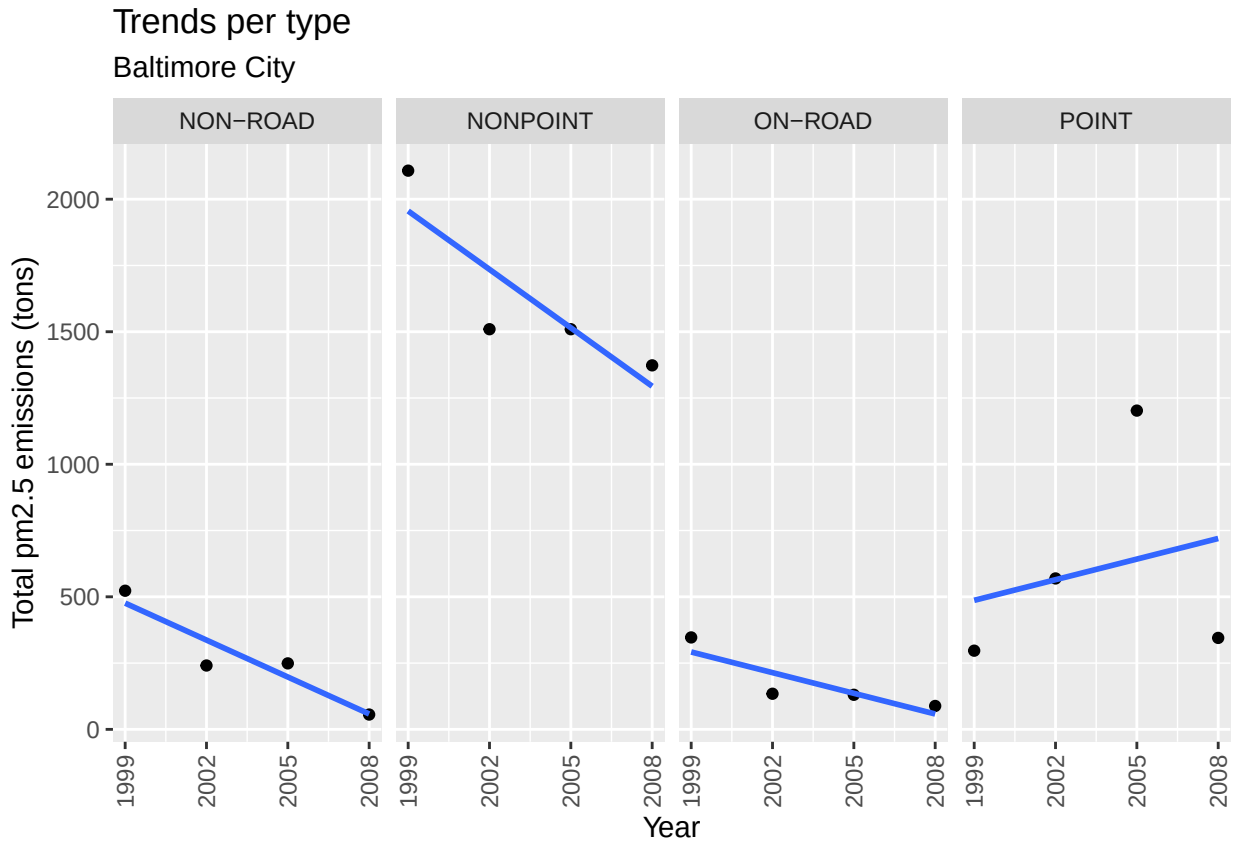
This plot shows a clear negative trend, meaning that emissions have decreased from 1999 to 2008.

Question 2: Have total emissions from PM2.5 decreased in Baltimore City, Maryland



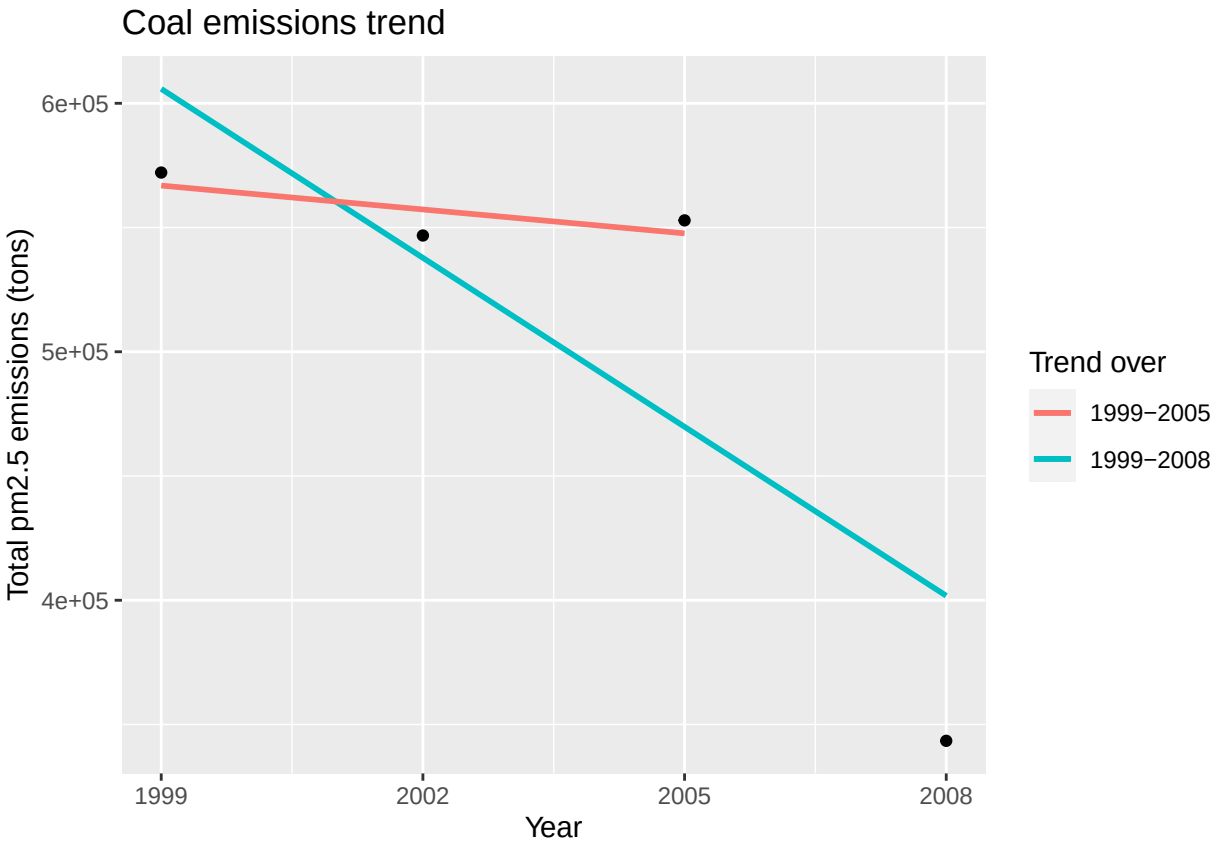
By subsetting data to Baltimore City, Maryland (fips == "24510") we can we that total emissions have raised here.

Question 3: Of the 4 types of sources, which see decreases in emissions for Baltimore City



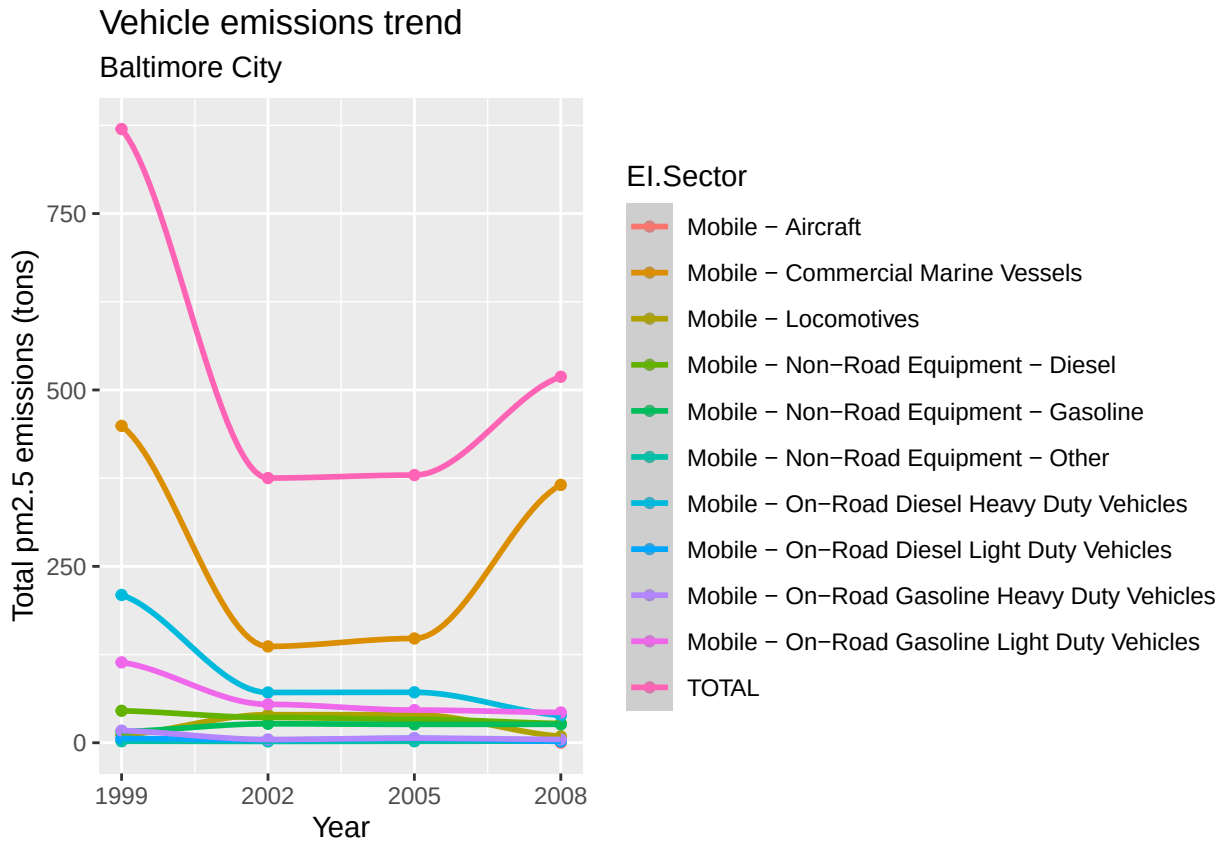
From this plot we can see taht the pm2.5 trends downwards for NON-ROAD, NON-POINT, and ROAD. But it is trending up for the POINT type.

Question 4: How have emissions from coal combustion sources changed?



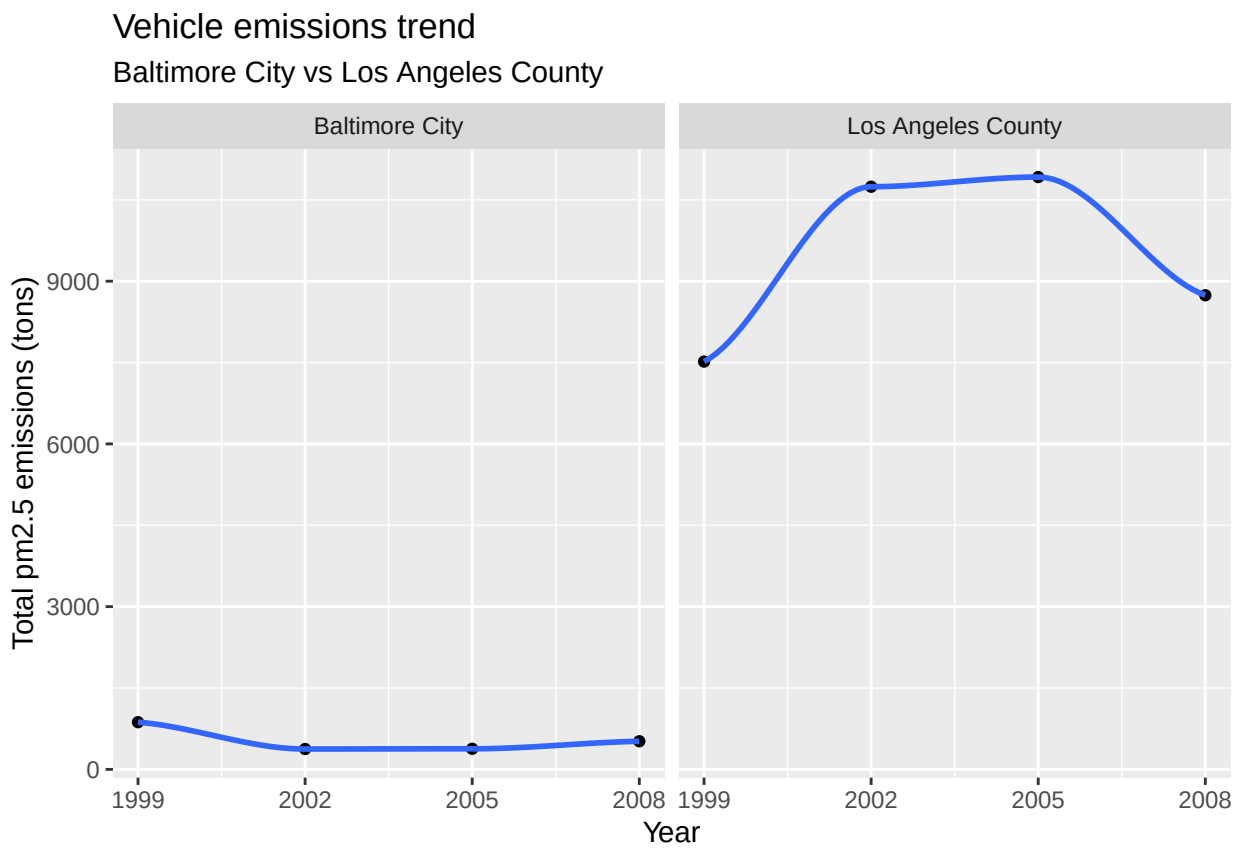
Yes! Although the trend for 1999-2005 (shown in red) seems to be fairly consistent, the only real decrease being 2008.

Question 5: How have emissions from motor vehicle sources changed?



Overall emissions from vehicles are lower than in 1999, but have risen in 2008 mainly due to commercial marine vessels, which is the main contributor.

Question 6: Has Baltimore City emissions from motor vehicles changed greater than in Los Angeles County?



The change in Los Angeles County is much greater, presumably in part due to the significant population difference (9.78 million vs 560 thousand as of 2008).