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
```

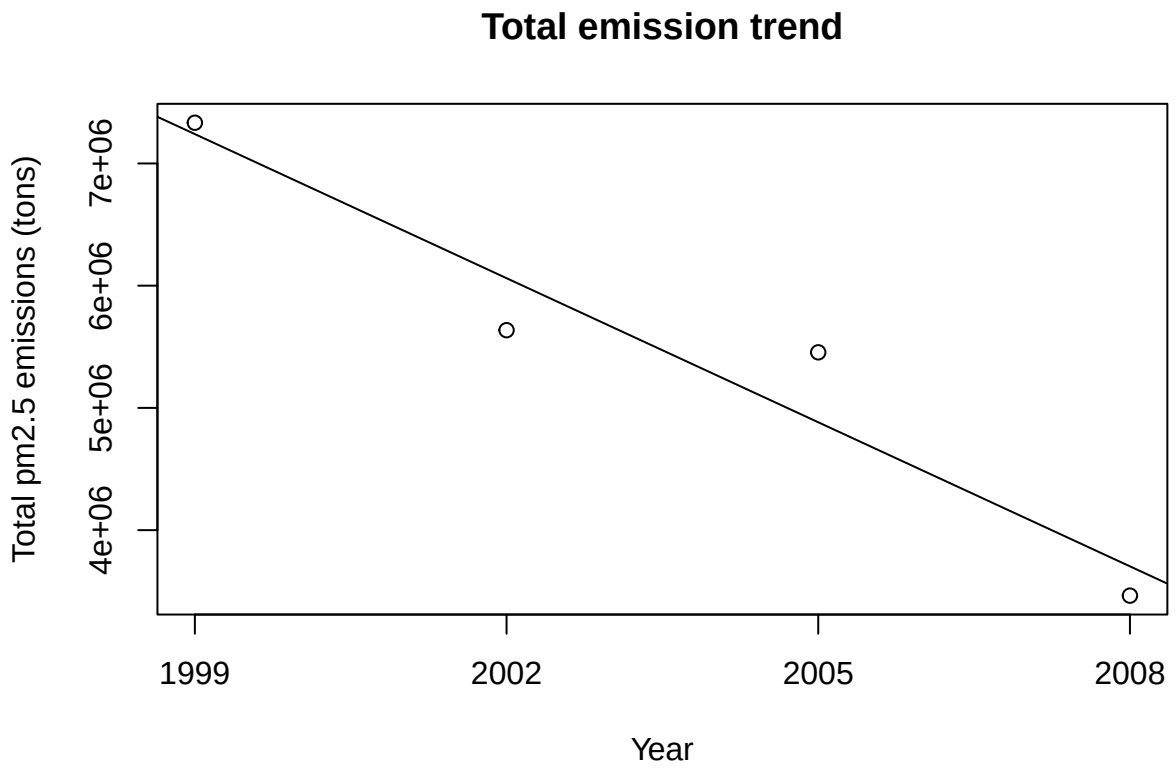
```
str(dt)
```

```
## 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>
```

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

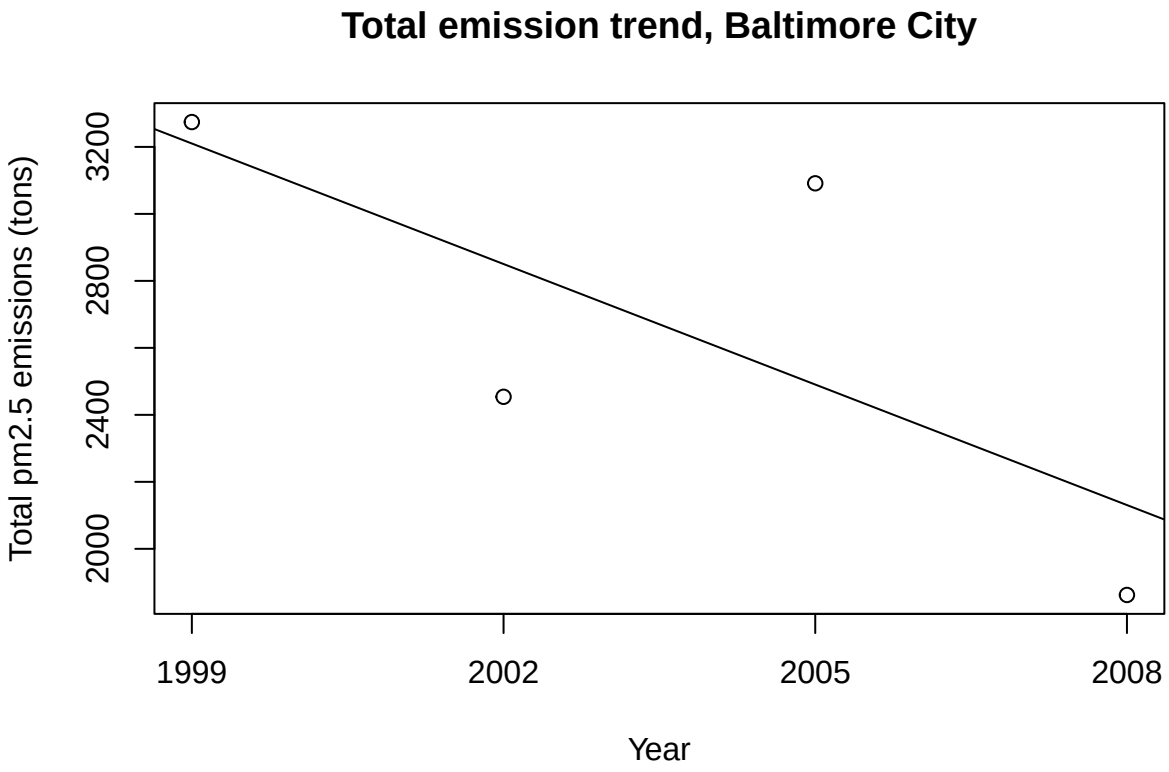
```
##           SCC  fips Pollutant   Emissions     type year
## 1: 220104013T 37141  PM25-PRI 2.00000e-02  ON-ROAD 1999
## 2: 2270003040 42045  PM25-PRI 1.73000e+00 NON-ROAD 1999
## 3: 2270002003 16073  PM25-PRI 1.23023e-02 NON-ROAD 2008
## 4: 2270007015 17189  PM25-PRI 7.00000e-02 NON-ROAD 2005
## 5: 2265004026 08125  PM25-PRI 6.50008e-05 NON-ROAD 2008
##                                                                                                Short.Name
## 1:     Highway Veh - Gasoline - Light Duty Trucks 3 & 4 - Rural Other Principal Arterial: Tire Wear
## 2:                           Off-highway Diesel /Industrial Equipt /Other General Industrial Equipt
## 3:                                      Off-highway Diesel /Construction & Mining Equipt /Pavers
## 4:                          Off-highway Diesel /Logging Equipt /Forest Equipt - Feller/Bunch/Skidder
## 5: Off-highway Gasoline, 4-Stroke /Lawn & Garden Equipt /Trimmers/Edgers/Brush Cutters (Commercial)
##                                    EI.Sector
## 1: Mobile - On-Road Gasoline Light Duty Vehicles
## 2:          Mobile - Non-Road Equipment - Diesel
## 3:          Mobile - Non-Road Equipment - Diesel
## 4:          Mobile - Non-Road Equipment - Diesel
## 5:        Mobile - Non-Road Equipment - Gasoline
```

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



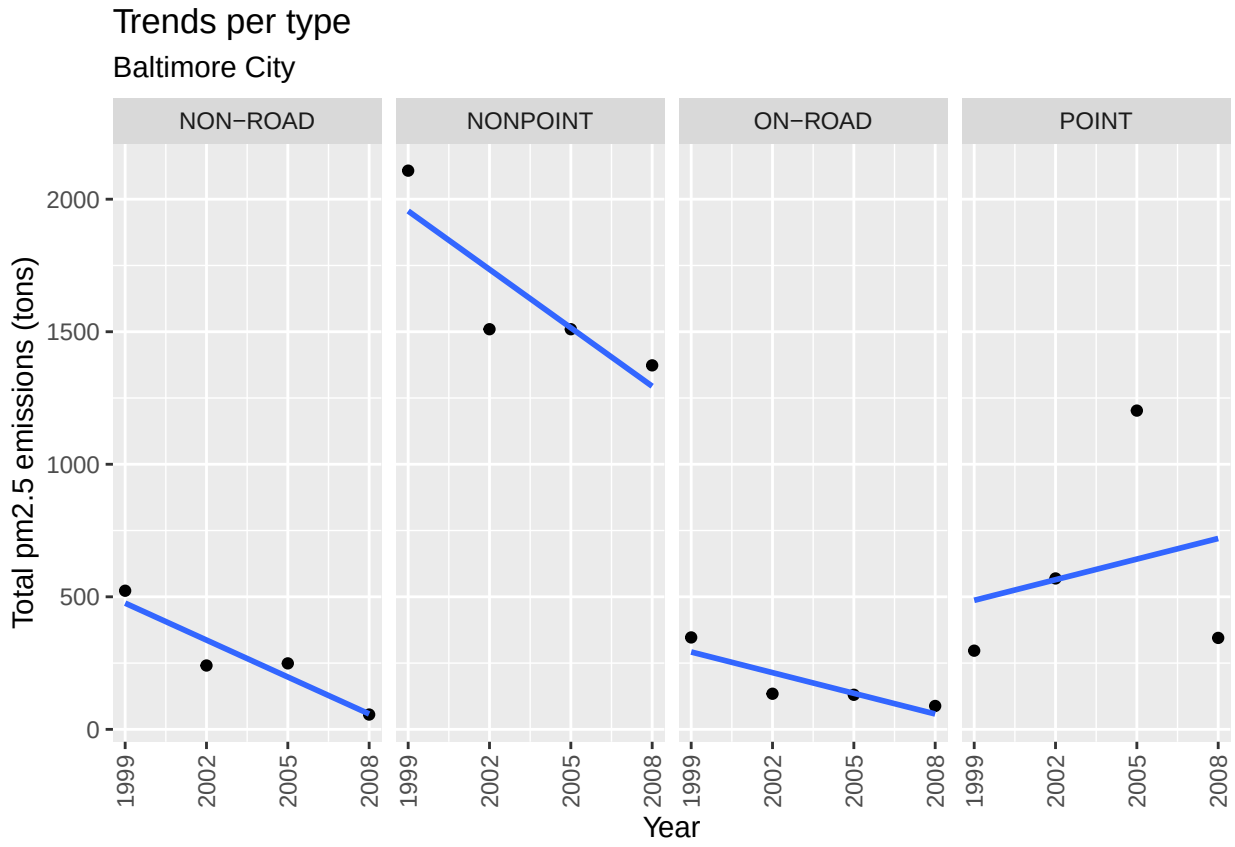
This plot shows a clear negative trend, meaning that emissions have decreased from 1999 to 2008.

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



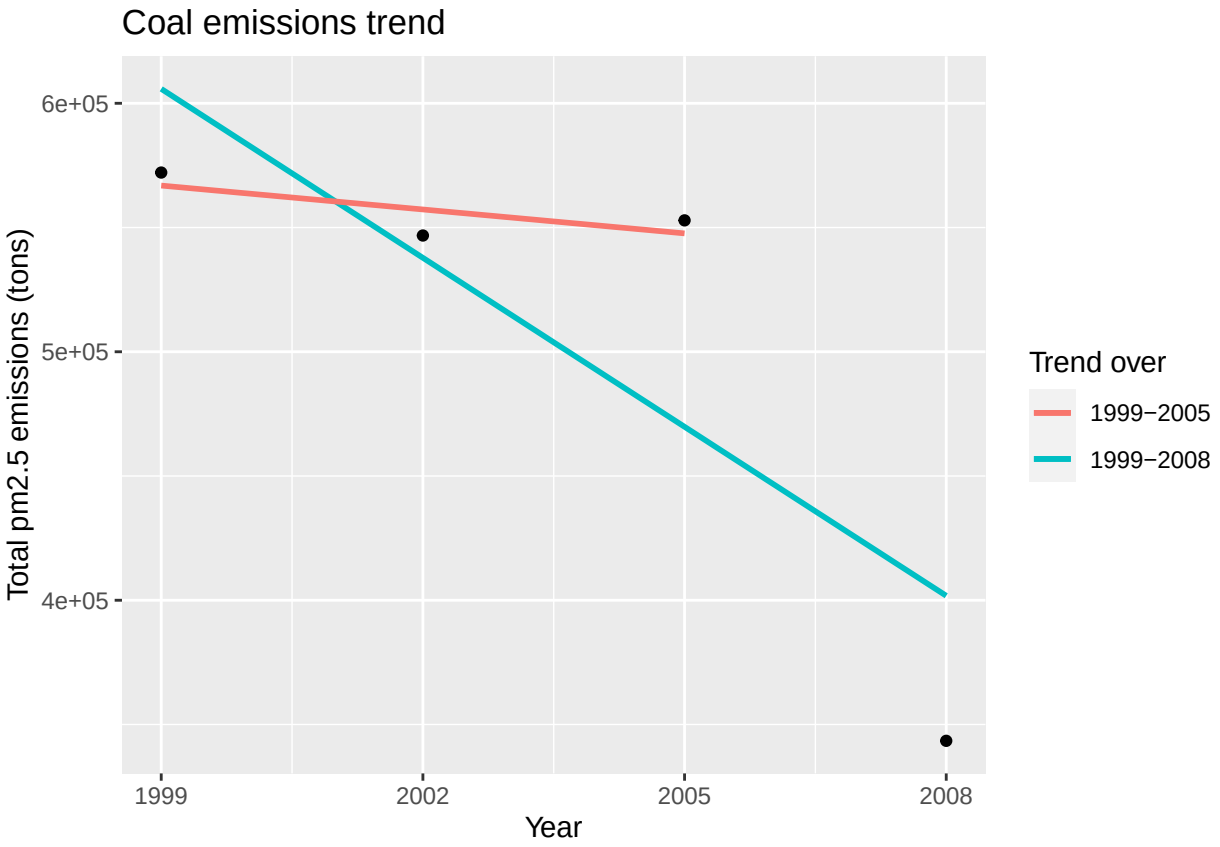
By subsetting data to Baltimore City, Maryland (fips == "24510") we can we that total emissions have raised here.

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



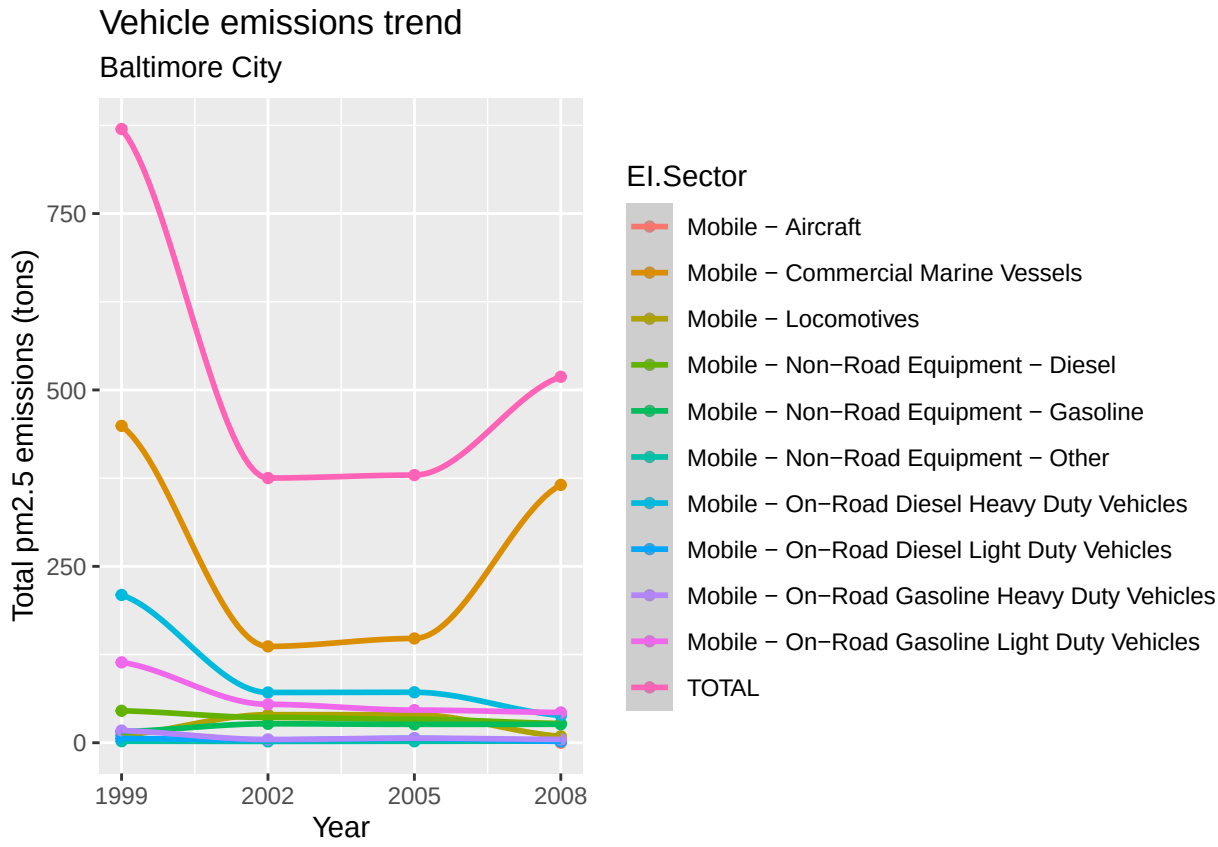
From this plot we can see taht the pm2.5 trends downwards for NON-ROAD, NON-POINT, and ROAD. But it is trending up for the POINT type.

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



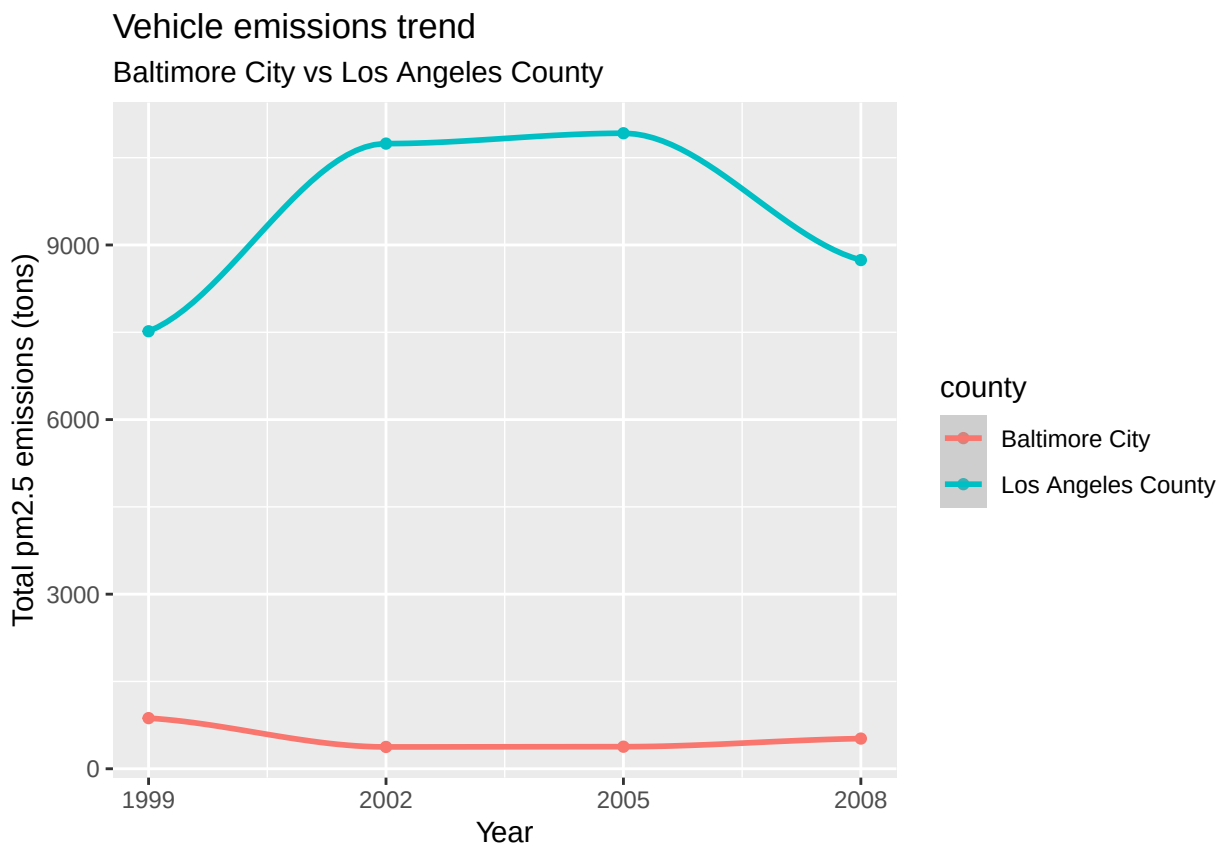
Yes! Although the trend for 1999-2005 (shown in red) seems to be fairly consistent, the only real decrease being 2008.

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 due to the significant population difference (9.78 million vs 560 thousand as of 2008).