Pollution

Jeffrey Strickland

12/25/2021

# Source Classification Codes (SCCs)

The U.S. EPA uses Source Classification Codes (SCCs) to classify different types of activities that generate emissions. Each SCC represents a unique source category-specific process or function that emits air pollutants. The SCCs are used as a primary identifying data element in EPA’s WebFIRE (where SCCs are used to link emissions factors to an emission process), the National Emissions Inventory (NEI), and other EPA databases. The SCCs are also used by many regional, state, local and tribal agency emissions data systems. Examples of processes described by SCCs and some of the emissions they generate include: \* Burning fuel in a boiler produces oxides of nitrogen (NOx) and other criteria and hazardous air pollutants (HAP). \* An industrial process such as paint coating produces volatile organic compounds (VOC). \* Fires produce particulate matter (PM). Sources in the SCC table are classified into the following five broad types: point, non-point, events, non-road and on-road (see EPA.GOV, *Introduction to Source Classification Codes and their Use for EIS Submissions*)

## Question 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.

## Data Wrangling

Getting, Cleaning, and Manipulating for analysis

library(ggplot2)  
library(RColorBrewer)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggformula)

## Warning: package 'ggformula' was built under R version 4.1.2

## Loading required package: ggstance

## Warning: package 'ggstance' was built under R version 4.1.2

##   
## Attaching package: 'ggstance'

## The following objects are masked from 'package:ggplot2':  
##   
## geom\_errorbarh, GeomErrorbarh

## Loading required package: scales

## Loading required package: ggridges

## Warning: package 'ggridges' was built under R version 4.1.2

##   
## New to ggformula? Try the tutorials:   
## learnr::run\_tutorial("introduction", package = "ggformula")  
## learnr::run\_tutorial("refining", package = "ggformula")

### Check the working directory

getwd()

## [1] "C:/Users/jeff/Documents/R/EPA\_emissions"

### Download and unzip the file:

* Set the filename to match “summarySCC\_PM25.rd”
* Check to see if already downloaded and unzipped working directory (WD)
* If filename is missing, download to the WD from the given URL
* Unzip the air\_pollution.zip into the WD
* summarySCC\_PM25.rds & Source\_Classification\_Code.rds appear in the WD

filename = "summarySCC\_PM25.rds"if (!file.exists(filename)){ urlzip <- "https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2FNEI\_data.zip" download.file(urlzip, destfile = “./air\_pollution.zip” ) unzip("./air\_pollution.zip", exdir = "." )}

## Load the data:

NEI <- readRDS("./summarySCC\_PM25.rds")  
SCC <- readRDS("./Source\_Classification\_Code.rds")

### Check NEI data structure

summary(NEI)

## fips SCC 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   
## Length:6497651 Min. :1999   
## Class :character 1st Qu.:2002   
## Mode :character Median :2005   
## Mean :2004   
## 3rd Qu.:2008   
## Max. :2008

### Check SCC data structure

ls.str(SCC)

## Created\_Date : Factor w/ 57 levels "","1/27/2000 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...  
## Data.Category : Factor w/ 6 levels "Biogenic","Event",..: 6 6 6 6 6 6 6 6 6 6 ...  
## EI.Sector : Factor w/ 59 levels "Agriculture - Crops & Livestock Dust",..: 18 18 18 18 18 18 18 18 18 18 ...  
## Last.Inventory.Year : int [1:11717] NA NA NA NA NA NA NA NA NA NA ...  
## Map.To : num [1:11717] NA NA NA NA NA NA NA NA NA NA ...  
## Option.Group : Factor w/ 25 levels "","C/I Kerosene",..: 1 1 1 1 1 1 1 1 1 1 ...  
## Option.Set : Factor w/ 18 levels "","A","B","B1A",..: 1 1 1 1 1 1 1 1 1 1 ...  
## Revised\_Date : Factor w/ 44 levels "","1/27/2000 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...  
## SCC : Factor w/ 11717 levels "10100101","10100102",..: 1 2 3 4 5 6 7 8 9 10 ...  
## SCC.Level.Four : Factor w/ 6084 levels "","(NH4)2 SO4 Acid Bath System and Evaporator",..: 4455 5583 4466 4458 1341 5246 5584 5983 4461 776 ...  
## SCC.Level.One : Factor w/ 17 levels "Brick Kilns",..: 3 3 3 3 3 3 3 3 3 3 ...  
## SCC.Level.Three : Factor w/ 1061 levels "","100% Biosolids (e.g., sewage sludge, manure, mixtures of these matls)",..: 88 88 156 156 156 156 156 156 156 156 ...  
## SCC.Level.Two : Factor w/ 146 levels "","Agricultural Chemicals Production",..: 32 32 32 32 32 32 32 32 32 32 ...  
## Short.Name : Factor w/ 11238 levels "","2,4-D Salts and Esters Prod /Process Vents, 2,4-D Recovery: Filtration",..: 3283 3284 3293 3291 3290 3294 3295 3296 3292 3289 ...  
## Usage.Notes : Factor w/ 21 levels ""," ","includes bleaching towers, washer hoods, filtrate tanks, vacuum pump exhausts",..: 1 1 1 1 1 1 1 1 1 1 ...

## Define dataset for analysis and plotting

### Aggregate the Emissions data from NEI by year and sum them

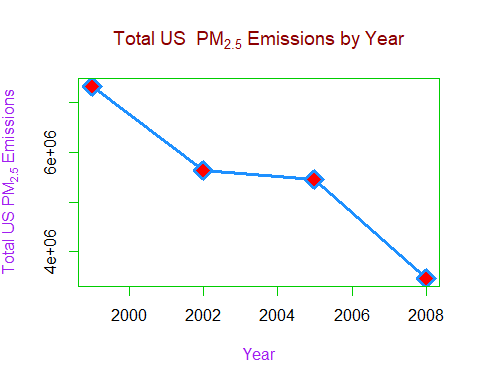
totalNEI <- aggregate(Emissions ~ year, NEI, sum)

## Plot Construction

Construct a combined line-point plot showing emissions from the given years and save it as a PNG graphic.

### Base R Plot

plot(totalNEI$year, totalNEI$Emissions, # Plot variables  
 type = "o", # Plot type is overlaid (line & points)  
 col = "dodgerblue", # Plot line color  
 lwd=3, # Plot line thickness   
 font.main = 3, # Main title font size  
 col.main = "darkred", # Main title color  
 main = expression("Total US "~ PM[2.5]~ "Emissions by Year"),   
 # Title label  
 col.lab = "purple", # Axes labels text color  
 ylab = expression("Total US "~ PM[2.5] ~ "Emissions"),   
 # y-axis label text   
 xlab = "Year", # x-axis label text  
 fg = "green3", # plot axes color  
 pch = 23, # plot point type  
 bg = "red", # plot point fill color  
 cex = 2) # Plot point size



#dev.off()

### Answer to Example 1:

Total emissions from PM2.5 have decreased in the United States from 1999 to 2008.

## Example 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.

### Extract Baltimore City Data

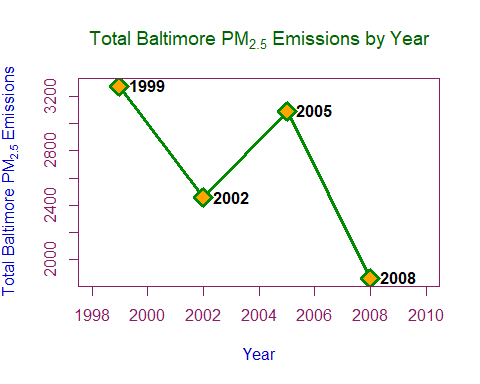
Extract Baltimore City (FIPS Code 24510) Emissions data for the corresponding years and sum them for each year. Then aggregate the data, summing by year.

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

### Construct a Combinaton Line-Point Plot

Construct a combined line-point plot showing emissions for Baltimore City for the given years 1999 to 2008

p1 <- plot(totalBaltimore$year, totalBaltimore$Emissions,   
 type = "o", # Plot type is overlaid (line & points)  
 col = "green4", # Plot line color  
 lwd=3, # Plot line thickness   
 font.main = 3, # Main title font size  
 col.main = "darkgreen", # Main title color  
 main = expression("Total Baltimore" ~ PM[2.5] ~ "Emissions by Year"),   
 xlim = c(1998,2010),  
 xlab = "Year", # x-axis label text  
 ylab = expression("Total Baltimore "~ PM[2.5] ~ "Emissions"),   
 col.lab = "blue3",  
 fg = "maroon4", # plot axes color  
 col.axis = "maroon4", # axis labels color  
 pch = 23, # plot point type  
 bg = "orange", # plot point fill color  
 cex = 2) # Plot point size)  
p1 + text(Emissions ~year, labels=totalBaltimore$year,data=totalBaltimore, cex=1, font=2, pos=4)



## integer(0)

# Answer to Example 2:

The total emissions from PM2.5 did decrease in the Baltimore City, Maryland from 1999 to 2008. However, this is not a complete picture. Although decreasing from 1999 to 2002, the was a steep increase from 2002 to 2005, nearly at the 1999 value. After that there was another drastic decline from 205 to 2008.

## Example 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.

### Data Wrangling

Construct the data setcomprised of emissions by source (type) in Baltimore City (FIPS Code 24510) from 1999 to 2008

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

### Build the Plot

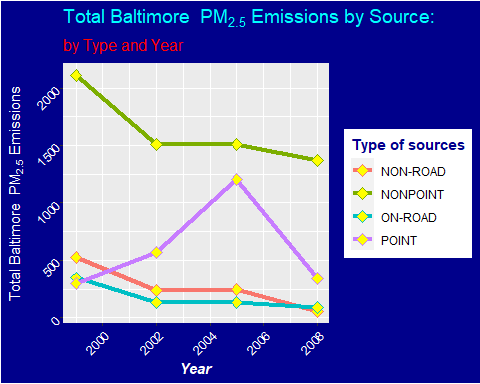
Plot the four types of sources indicated by the type (point, nonpoint, onroad, nonroad) variable and the corresponding emissions fro 1999 to 2008

#### Base graphic object

g <- ggplot(baltimoreType, aes(year, Emissions, col = type))

#### Add plot as a line type with 50% increase in size

g + geom\_line(size = 1.5) +   
 # Plot is a line type with 50% increase in size  
 geom\_point(size = 3, pch =23, fill="yellow") +  
 # Main title label text  
 ggtitle(expression("Total Baltimore " ~ PM[2.5] ~ "Emissions by Source:")) + #Title label  
 # Main subtitle label text   
 labs(subtitle=expression("by Type and Year")) +   
 # y-axis label text  
 ylab(expression("Total Baltimore " ~ PM[2.5] ~ "Emissions")) +  
 # x-axis lable text  
 xlab("Year") +  
 # plot line colors varied by pollution sources  
 scale\_colour\_discrete(name = "Type of sources") +  
 # plot themes including: background color  
 theme(plot.background = element\_rect(fill = "darkblue"),   
 # colors of axis titles   
 axis.title=element\_text(face = "bold.italic", color = "white"),  
 # color of legend title  
 legend.title = element\_text(face = "bold", color="blue4"),  
 # orientation, size, and color of x-axis labels  
 axis.text.x = element\_text(angle = 45, hjust = 1, color = "white"),  
 # orientation, size, and color of x-axis labels  
 axis.text.y = element\_text(angle = 45, hjust = 1, color = "white"),  
 # font size, color, and facing of main title  
 plot.title= element\_text(size=14,  
 color="cyan",  
 face="bold"),  
 # font size, color, and facing of main subtitle  
 plot.subtitle = element\_text(size=12,  
 color="red",  
 face="bold"))



### Answer to Example 3:

The emissions from point sources is higher in 2008 that it was in 1999, even though it has risen and fell. For the sources non-point, road, and on-road the 2008 levels are lower that those of 1999, and the overall trend is decreasing.

## Example 4.

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

### Data Wrangling

Construct the data set comprised of emissions from coal related combustion across the United States from 1999 to 2008. This requires aggregating data from both base data sets and use of the pattern-matching and replacement function grepl() for “coal”.

SCCcoal <- SCC[grepl("coal", SCC$Short.Name, ignore.case = T),]  
NEIcoal <- NEI[NEI$SCC %in% SCCcoal$SCC,]  
totalCoal <- aggregate(Emissions ~ year + type, NEIcoal, sum)  
totalCoal2 <- aggregate(Emissions ~ year, NEIcoal, sum)

### Construct the Plot

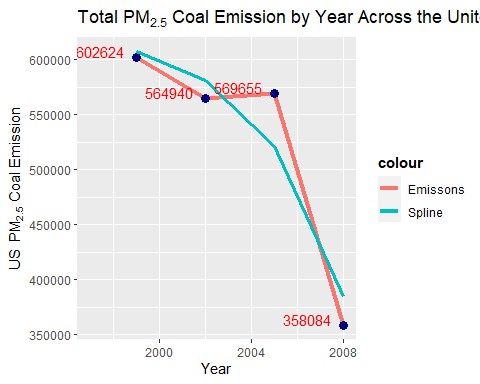
Here, we’ll use a combined line-point plot as we did in examples 1 and 2, but will add a spline (fitted line) to the plot.

First, we define a ggplot combining line-point plots for coal related emissions

g= ggplot(totalCoal2, aes(year, Emissions, col = "Emissons"))

Next we’ll add the line-point-spline with aesthetcics. The spline function projects points bettween data years in the context of the total interval (1999-2008) to form a trend-line. This is different than fittine a line for the entire plot from 1999-2008.

g+geom\_line(size=1.5) +  
 geom\_point(col = "navyblue", size = 3) + xlim(1997, 2008) +  
 geom\_spline(aes(x = year, y = Emissions, color = "Spline"), size = 1.25) +  
 geom\_text(aes(label=round(Emissions,0)), hjust=1.25, vjust=0, col = "red") +  
 ggtitle(expression("Total" ~ PM[2.5] ~ "Coal Emission by Year Across the United States")) +  
 xlab("Year") +  
 ylab(expression("US " ~ PM[2.5] ~ "Coal Emission")) +  
 theme(legend.title = element\_text(face = "bold"))



### Answer to Example 4:

Across the United States, emissions from coal combustion-related sources has decreased from 1999-2008, and the overall trend is negative (decreasing) and seen by the spline in plot4.png.

## Example 5. How have emissions from motor vehicle sources changed from 1999-2008 in Baltimore City?

### Data Wrangling

Construct a dataset comprised of emissions data for Baltimore City (FIPS Code 24510) from 199 to 2008, aggregated by year and summed

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

### Build the Plot

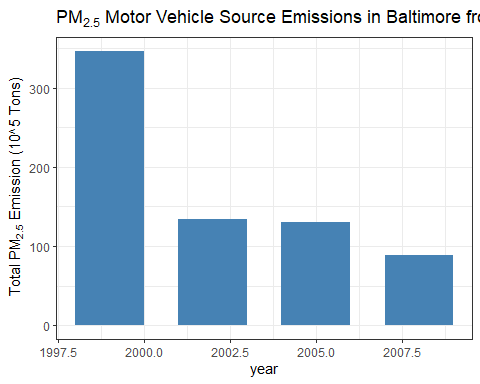
Construct a bar chart comprised of four years (one bar for each) with summed emissions data for Baltimore City #### Define a base plotting object “g” using ggplot

g <- ggplot(baltimoreMotorAGG, aes(year, Emissions))

#### Add geom functions to augment the plot and add aesthetics

g+geom\_bar(stat="identity",fill="steelblue",width=2) +  
 # Add theme  
 theme\_bw() +   
 # Set guides for scale to FALSE  
 guides(fill=FALSE) +  
 # Add x and y axes labels  
 labs(x="year", y=expression("Total PM"[2.5]\*" Emission (10^5 Tons)")) +   
 # Add main titlle  
 labs(title=expression("PM"[2.5]\*" Motor Vehicle Source Emissions in Baltimore from 1999-2008"))

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

 ### Answer to Example 5: Total emissions from motor vehicle sources in Baltimore City have decreased from 1999-2008, with a clear downward trend.

## Example 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?

vehiclesNEI <- subset(NEI, NEI$fips %in% c("24510","06037") & NEI$type == "O`N-ROAD")

Use a filter to get vehicle data from SCC. Level 2 is classified as industrial category sources, like industrial solid waste disposal or commercial marine vessels.

SCC\_Vehicles <- SCC %>%  
 filter(grepl('[Vv]ehicle', SCC.Level.Two)) %>%  
 select(SCC, SCC.Level.Two)

Get Location Emissions Data \* Filter the data to get Baltimore cCity (FIPS code = “24510”) and Los Angeles County (FIPS Code “06037”) \* Select the variables “fips”, “SCC”, “Emissions”, and “year” Inner-join the data set with SCC\_Vehicle \* Group cities (fips) by year (four gropps for each city) \* Select the variables Total Emissions, fips (Cities), and year.

Balt\_LA\_Emissions <- NEI %>%  
 filter(fips == "24510" | fips == "06037") %>%  
 select(fips, SCC, Emissions, year) %>%  
 inner\_join(SCC\_Vehicles, by = "SCC") %>%  
 group\_by(fips, year) %>%  
 summarise(Total\_Emissions = sum(Emissions, na.rm = TRUE)) %>%  
 select(Total\_Emissions, fips, year)

## `summarise()` has grouped output by 'fips'. You can override using the `.groups` argument.

Merge emission data into one superset

Balt\_LA\_Emissions$fips <- gsub("24510", "Baltimore City", Balt\_LA\_Emissions$fips)  
Balt\_LA\_Emissions$fips <- gsub("06037", "Los Angeles County", Balt\_LA\_Emissions$fips)

## Plot Construction

Construct a set of bar charts one for Baltimore City and one for Los Angeles County comprised of four years (one bar for each) with summed emissions. Define a base plotting object “g” using ggplot().

g <- ggplot(Balt\_LA\_Emissions, aes(x = factor(year), y = Total\_Emissions, fill = fips))

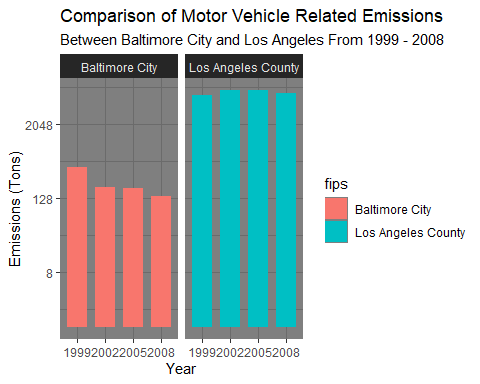
Get a description of the structure of the base graphic object

str(g)

## List of 9  
## $ data : grouped\_df [8 x 4] (S3: grouped\_df/tbl\_df/tbl/data.frame)  
## ..$ Total\_Emissions: num [1:8] 6110 7189 7304 6421 404 ...  
## ..$ fips : chr [1:8] "Los Angeles County" "Los Angeles County" "Los Angeles County" "Los Angeles County" ...  
## ..$ year : int [1:8] 1999 2002 2005 2008 1999 2002 2005 2008  
## ..$ .group : int [1:8] 2 2 2 2 1 1 1 1  
## ..- attr(\*, "groups")= tibble [2 x 2] (S3: tbl\_df/tbl/data.frame)  
## .. ..$ fips : chr [1:2] "Baltimore City" "Los Angeles County"  
## .. ..$ .rows: list<int> [1:2]   
## .. .. ..$ : int [1:4] 5 6 7 8  
## .. .. ..$ : int [1:4] 1 2 3 4  
## .. .. ..@ ptype: int(0)   
## .. ..- attr(\*, ".drop")= logi TRUE  
## $ layers : list()  
## $ scales :Classes 'ScalesList', 'ggproto', 'gg' <ggproto object: Class ScalesList, gg>  
## add: function  
## clone: function  
## find: function  
## get\_scales: function  
## has\_scale: function  
## input: function  
## n: function  
## non\_position\_scales: function  
## scales: NULL  
## super: <ggproto object: Class ScalesList, gg>   
## $ mapping :List of 3  
## ..$ x : language ~factor(year)  
## .. ..- attr(\*, ".Environment")=<environment: R\_GlobalEnv>   
## ..$ y : language ~Total\_Emissions  
## .. ..- attr(\*, ".Environment")=<environment: R\_GlobalEnv>   
## ..$ fill: language ~fips  
## .. ..- attr(\*, ".Environment")=<environment: R\_GlobalEnv>   
## ..- attr(\*, "class")= chr "uneval"  
## $ theme : list()  
## $ coordinates:Classes 'CoordCartesian', 'Coord', 'ggproto', 'gg' <ggproto object: Class CoordCartesian, Coord, gg>  
## aspect: function  
## backtransform\_range: function  
## clip: on  
## default: TRUE  
## distance: function  
## expand: TRUE  
## is\_free: function  
## is\_linear: function  
## labels: function  
## limits: list  
## modify\_scales: function  
## range: function  
## render\_axis\_h: function  
## render\_axis\_v: function  
## render\_bg: function  
## render\_fg: function  
## setup\_data: function  
## setup\_layout: function  
## setup\_panel\_guides: function  
## setup\_panel\_params: function  
## setup\_params: function  
## train\_panel\_guides: function  
## transform: function  
## super: <ggproto object: Class CoordCartesian, Coord, gg>   
## $ facet :Classes 'FacetNull', 'Facet', 'ggproto', 'gg' <ggproto object: Class FacetNull, Facet, gg>  
## compute\_layout: function  
## draw\_back: function  
## draw\_front: function  
## draw\_labels: function  
## draw\_panels: function  
## finish\_data: function  
## init\_scales: function  
## map\_data: function  
## params: list  
## setup\_data: function  
## setup\_params: function  
## shrink: TRUE  
## train\_scales: function  
## vars: function  
## super: <ggproto object: Class FacetNull, Facet, gg>   
## $ plot\_env :<environment: R\_GlobalEnv>   
## $ labels :List of 3  
## ..$ x : chr "factor(year)"  
## ..$ y : chr "Total\_Emissions"  
## ..$ fill: chr "fips"  
## - attr(\*, "class")= chr [1:2] "gg" "ggplot"

Use to the basic ggplot to generate a bar chart for emission from each location,and add aesthetics to the basic plot, including: \* plot theme \* main title label \* main subtitle label \* x-axis and y-axis labels \* facet\_grid() forms a matrix of panels defined by row and column faceting variables. \* binary logarithm transformation of the y-axis for scaling effect

g + geom\_bar(stat = "identity", width = 0.7) +  
 facet\_grid(.~fips) + scale\_y\_continuous(trans='log2') +  
 labs(x = "Year", y = "Emissions (Tons)", title = "Comparison of Motor Vehicle Related Emissions", subtitle = "Between Baltimore City and Los Angeles From 1999 - 2008") +  
 theme(plot.title = element\_text(size = 14),  
 plot.subtitle = element\_text(size = 14),  
 axis.title.x = element\_text(size = 12),  
 axis.title.y = element\_text(size = 12),  
 strip.text.x = element\_text(size = 12)) +  
 theme\_dark()



Save the plots using ggsave()

ggsave("plot6.png", width = 20, height = 10, units = "cm")  
ggsave("plot6\_log.png", width = 20, height = 10, units = "cm")

## Answer to Example 6:

Comparing emissions from motor vehicle sources in Baltimore City with emissions from motor vehicle sources in Los Angeles County, California are as follows: 1. Los Angeles County has seen significant change in scale that Baltimore City. 2. However, while not at the same scale, Baltimore city has seen a decline in vehicle emissions emissions. 3. While Los Angeles County has seen greater changes over time in motor vehicle emissions, the 2008 level is high than the 1999 level. 4. The picture is clearer in plot6\_log.png with the y-axis logarithm transformation, with Baltimore City seeing greater downward changes over time.