```
########################

### PM25_USA_EPA_NEI ###

########################

# 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?

########################

###    Resources    ###

########################

# EPA Government references
# Reference 1. 2011 National Emissions Inventory, version 1 Technical Support Document November 2013 - DRAFT
# http://www.epa.gov/ttn/chief/net/2008neiv3/2008_neiv3_tsd_draft.pdf
# Reference 2. 2008 National Emissions Inventory, version 3 Technical Support Document September 2013 - DRAFT
# http://www.epa.gov/ttn/chief/net/2011nei/2011_neiv1_tsd_draft.pdf

# STATE Government references
# Reference 3. Methodologies for U.S. Greenhouse Gas Emissions Projections: Non-CO2 and Non-Energy CO2 Sources DECEMBER, 2013
# http://www.state.gov/documents/organization/219472.pdf

# MOVES Government references
# Reference 4. MOVES (Motor Vehicle Emission Simulator) Website
# http://www.epa.gov/otaq/models/moves/index.htm
# Reference 5. MOVES FAQ - http://www.epa.gov/otaq/models/moves/420f09073.pdf

# Definition of a Motor Vehicle from MOVES:
# MOVES2010 is the state-of-the-art upgrade to EPA's modeling tools
# for estimating emissions from highway vehicles,
# based on analysis of millions of emission test results
# and considerable advances in the Agency's understanding of vehicle emissions.
# MOVES2010 replaces the previous model for estimating
# on-road mobile source emissions, MOBILE6.2.

# reshape2
# 1. CRAN - http://cran.r-project.org/web/packages/reshape2/index.html
# 2. Sean C. Anderson Blog - An Introduction to reshape2 - http://seananderson.ca/2013/10/19/reshape.html

# ggplot2
# 1. GGPLOT2 - http://docs.ggplot2.org
# 2. GGPLOT2 - http://www.cookbook-r.com/Graphs/Axes_(ggplot2)/
# 3. GGPLOT2 - http://acaird.github.io/computers/r/2013/11/27/slopegraphs-ggplot/


########################

# Data sets inspection #

########################

# Inspect the SCC file for the requested data
```

```
# > str(SCC)
# levels(SCC$Data.Category)
# [1] "Biogenic" "Event"    "Nonpoint" "Nonroad"  "Onroad"   "Point"
# interesting result: "Onroad" Data Category

# > levels(SCC$EI.Sector)
# interesting results: 4 Source Category Sectors are corresponding to the question
# [49] "Mobile - On-Road Diesel Heavy Duty Vehicles"
# [50] "Mobile - On-Road Diesel Light Duty Vehicles"
# [51] "Mobile - On-Road Gasoline Heavy Duty Vehicles"
# [52] "Mobile - On-Road Gasoline Light Duty Vehicles"

# > levels(SCC$SCC.Level.Two)
# interesting results: 2 Source Category Levels 2 are corresponding to the question
# [48] "Highway Vehicles - Diesel"
# [49] "Highway Vehicles - Gasoline"

########################

###   plot6 R code   ###

########################

# Create a function called
plot6_MOVES_TotalEmissionsPM2.5_BA_LA_Source_MotorVehicule_1999_to_2008() to do the
requested plot

plot6_MOVES_TotalEmissionsPM2.5_BA_LA_Source_MotorVehicule_1999_to_2008 = function()
{
  library(reshape2) # use reshape2 to clean and prepare the data
  library(ggplot2) # use ggplot2 to plot
  library(scales) # use scale functions to modify aesthetics to legend's plot

  # Set the working directory on my local machine
  setwd("~/Desktop/Data Science Specialization/Exploratory Data Analysis/Course project 2")

  # Read the PM2.5 Emissions Data in summarySCC_PM25.rds file with readRDS() function
  NEI <- readRDS("summarySCC_PM25.rds")

  # Read the Source Classification Code Table in Source_Classification_Code.rds file with readRDS()
function
  SCC <- readRDS("Source_Classification_Code.rds")

  ###### BALTIMORE DATASET #########

  # Define the Baltimore dataset
  # Subset of PM2.5 Emissions Data with NEI$fips == "24510"
  NEI_Baltimore <- subset(NEI, fips == "24510")

  # Find the Baltimore Motor Vehicule sources
  # Inspect Source Classification Code Table (SCC) file: column SCC.Level.Three
  # Use of regular expressions and grep() function
  Baltimore_Motor_Vehicule_Related_Source <- grep("Highway Vehicles(*.)", SCC$SCC.Level.Two,
ignore.case = TRUE)

  # Subset Motor Vehicule sources for Baltimore
  SCC_Baltimore_Motor_Vehicule <- SCC[Baltimore_Motor_Vehicule_Related_Source ,"SCC"]

  # Subset PM2.5 Emissions with Motor Vehicle sources for Baltimore
  NEI_Baltimore_Motor_Vehicule_Combustion_Related_Source <- subset(NEI_Baltimore, SCC %in%
```

```
SCC_Baltimore_Motor_Vehicule)

  # Split the PM2.5 emissions for coal combustion sources
  # into two column frame year and type, with melt() function
  NEI_Baltimore_Motor_Vehicule_Combustion_Related_Source_Year_Type <-
melt( NEI_Baltimore_Motor_Vehicule_Combustion_Related_Source, id.vars = c("year","type"),
measure.vars="Emissions")

  # Sum the PM2.5 emmmissions in Baltimore resulting dataset by year variable with dcast() function
  NEI_Baltimore_Motor_Vehicule_Combustion_Related_Source_Year_Type_Sum <-
dcast(NEI_Baltimore_Motor_Vehicule_Combustion_Related_Source_Year_Type, year ~ variable,
fun.aggregate = sum, na.rm = TRUE)

  ###### LOS ANGELES DATASET #########

  # Define the Los Angeles dataset
  # Subset of PM2.5 Emissions Data with NEI$fips == "06037"
  NEI_LosAngeles <- subset(NEI, fips == "06037")

  # Find the Baltimore Motor VehiculeS sources
  # Inspect Source Classification Code Table (SCC) file: column SCC.Level.Three
  # Use of regular expressions and grep() function
  LosAngeles_Motor_Vehicule_Related_Source <- grep("Highway Vehicles(*.)", SCC
$SCC.Level.Two, ignore.case = TRUE)

  # Subset Motor Vehicule sources for Los Angeles
  SCC_LosAngeles_Motor_Vehicule <- SCC[LosAngeles_Motor_Vehicule_Related_Source ,"SCC"]

  # Subset PM2.5 Emissions with Motor Vehicle sources for Los Angeles
  NEI_LosAngeles_Motor_Vehicule_Combustion_Related_Source <- subset(NEI_LosAngeles, SCC
%in% SCC_LosAngeles_Motor_Vehicule)

  # Split the PM2.5 emissions for Motor Vehicle sources
  # into two column frame year and type, with melt() function
  NEI_LosAngeles_Motor_Vehicule_Combustion_Related_Source_Year_Type <-
melt( NEI_LosAngeles_Motor_Vehicule_Combustion_Related_Source, id.vars = c("year","type"),
measure.vars="Emissions")

  # Sum the PM2.5 emmmissions in Baltimore resulting dataset by year variable with dcast() function
  NEI_LosAngeles_Motor_Vehicule_Combustion_Related_Source_Year_Type_Sum <-
dcast(NEI_LosAngeles_Motor_Vehicule_Combustion_Related_Source_Year_Type, year ~ variable,
fun.aggregate = sum, na.rm = TRUE)

  ############ MERGING Baltimore & Los Angeles datasets #############

  # Add a column with character variable "Los Angeles" to identify data origin before merging datasets
  NEI_LosAngeles_Motor_Vehicule_Combustion_Related_Source_Year_Type_Sum
$CityStateCountry <- "Los Angeles - California - USA"

  # Add a column with character variable to identify data origin before merging datasets
  NEI_Baltimore_Motor_Vehicule_Combustion_Related_Source_Year_Type_Sum$CityStateCountry
<- "Baltimore - Maryland - USA"

  # Combine the Baltimore & Los Angeles datasets with rbind() function
  NEI_Baltimore_LosAngeles <-
rbind(NEI_Baltimore_Motor_Vehicule_Combustion_Related_Source_Year_Type_Sum,
NEI_LosAngeles_Motor_Vehicule_Combustion_Related_Source_Year_Type_Sum)

  ############ PLOT with facets #############
```

```
  # Create another multiple barplot with a colour per year : modify year variable from integer to factor

  # Modify the merged dataset with passing year variable from integer to factor
  NEI_Baltimore_LosAngeles$year <- factor(NEI_Baltimore_LosAngeles$year, levels=c('1999', '2002',
'2005', '2008'))

  # Begin the plot with ggplot() funtion
  g <- ggplot(data = NEI_Baltimore_LosAngeles, aes(x = year, y = Emissions))
  # Use bar for the plot with scaling color for variable year
  g <- g + geom_bar(aes(fill = year), stat = "identity")
  # Add a title to the plot
  g <- g + ggtitle("PM2.5 Emissions for Motor Vehicle Sources type in Baltimore & Los Angeles
(USA)")
  # Add text values in bars for results
  g <- g + geom_text(aes(label = round(Emissions,0)), size = 6, hjust = 0.5, vjust = -0.5, colour =
"darkblue")
  # Add a legend on x with "Year"
  g <- g + xlab("Year")
  # Add a legend on y with "PM2.5 Emissions (Tons)"
  g <- g + ylab("PM2.5 Emissions  (Tons)")
  # Use facet function to draw the both barplots
  g <- g + facet_grid(. ~ CityStateCountry)
  # Modify themes of the plot
  g <- g + theme_bw(base_family = "Times", base_size = 12)

  # print the plot()
  print(g)

  # Save png file in working directory
  dev.copy(png, filename = "plot6.png", height = 600, width = 800, unit = "px", bg = "transparent")

  # Release screen
  dev.off()
}

plot6_MOVES_TotalEmissionsPM2.5_BA_LA_Source_MotorVehicule_1999_to_2008()

# Answer 6:
# 1. PM2.5 Emissions from Motor Vehicle sources in Baltimore have decreased
# from 1999 to 2009, and especially between 1999 to 2002 with an important
# negative slope on graph.
# 2. On the contrary, PM2.5 Emissions from Motor Vehicle sources in Los Angeles
# have increased from 1999 to 20008.
# To be more precise, the PM2.5 Emissions have increased from 1999 to 2005 with
# a positive slope, an then decreased with a more significative importance from 2005
# to 2008 (realtively importante negative slope).
```