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

### PM25_USA_EPA_NEI ###

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

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

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

###    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

# 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"

# > levels(SCC$EI.Sector)
# interesting results: indentify the judicious Source Category sectors type
# [13] "Fuel Comb - Comm/Institutional - Coal"
# [18] "Fuel Comb - Electric Generation - Coal"
# [51] "Fuel Comb - Industrial Boilers, ICEs - Coal"
# [52] "Fuel Comb - Residential - Other"
# NOTA - Reference 1 - 3.13 Fuel Combustion – Residential – Natural Gas, Oil, and Other:
# "Fuel Comb - Residential – Other" which includes the fuels:
# (1) coal,
```

```
# (2) liquid petroleum gas and
# (3) "Biomass; all except Wood".
# (Anthracite Coal & Bituminous/Subbituminous Coal)

# > levels(SCC$SCC.Level.Three)
# interesting results: indentify the judicious Source Category Level 3 type
# Main idea : find the Source Category Levels 3 types corresponding to:
## - Source Category sectors type retained ;
## - With a judicious label.
# Results:
# [88] "Anthracite Coal"
# = YES
# [156] "Bituminous/Subbituminous Coal"
# = YES
# [267] "Commercial/Industrial" = NO (not considered for the coal combustion process & negligible)
# [591] "Lignite"
# = YES
# [531] "Industrial"
# = NO (not considered for the coal combustion process & negligible)
# [592] "Lignite Coal"
# = YES
# [1032] "Waste Coal"
# = NO (not considered as a combustion process & negligible)
#
# NOTA: "Commercial/Industrial" and "Industrial" Source Category Level 3 types
# are not exclusive to coal source - check for the adequate Source Category sectors

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

###  plot4 R code  ###

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

# Create a function called
plot4_TotalEmissionsPM2.5_USA_Source_CoalCombustion_1999_to_2008() to do the requested plot

plot4_TotalEmissionsPM2.5_USA_Source_CoalCombustion_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")

  # Find the USA coal combustion sources : see Data
  # Inspect Source Classification Code Table (SCC) file: column SCC.Level.Three
  # Use of regular expressions and grep() function
  USA_Coal_Combustion_Related_Source <- grep("Anthracite Coal|Bituminous/Subbituminous Coal|
Lignite|Lignite Coal", SCC$SCC.Level.Three, ignore.case = TRUE)
  # Result dataset of 103 observations

  # Subset Coal Combustion sources
```

```r
  SCC.Coal.Combustion <- SCC[USA_Coal_Combustion_Related_Source ,"SCC"]

  # Subset PM2.5 Emissions with Coal Combustion sources
  NEI_USA_Coal_Combustion_Related_Source <- subset(NEI, SCC %in% SCC.Coal.Combustion)

  # Split the PM2.5 emissions for coal combustion sources
  # into two column frame year and type, with melt() function
  NEI_USA_Coal_Combustion_Related_Source_Year_Type <-
melt(NEI_USA_Coal_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_USA_Coal_Combustion_Related_Source_Year_Type_Sum <-
dcast(NEI_USA_Coal_Combustion_Related_Source_Year_Type, year ~ variable, fun.aggregate =
sum, na.rm = TRUE )

  # Create a function g with ggplot() function with NEI_baltimore_Year_Type_Sum dataset
  g <- ggplot(data = NEI_USA_Coal_Combustion_Related_Source_Year_Type_Sum, aes(x = year, y
= Emissions))
  # Add line
  g <- g + geom_line()
  # Add points
  g <- g + geom_point(aes(colour = Emissions), size = 6, alpha = 1/2) + scale_colour_gradient(low =
"blue")
  # Add text for Emissions values, with integer Emissions values well positionned
  g <- g + geom_text(aes(label = round(Emissions,0)), size = 5, hjust = 0.5, vjust = -2, position =
"stack", colour = "orangered")
  # Add a title to the plot
  g <- g + ggtitle("PM2.5 Emissions for Coal Combustion Source Type in USA")
  # Use scale function to modify legend title "Type" by "Source Type"
  # g <- g + scale_color_discrete(name = "PM2.5 Emissions/Coal Comb.")
  # Add a legend on x with "Year"
  g <- g + xlab("Year") + ylab("PM2.5 Emissions (Tons)")
  # Modify x axis limit with scale_y_continuous() function
  g <- g + scale_y_continuous(limits = c(330000, 600000))
  g <- g + scale_x_continuous(limits = c(1999,2008+1))
  # 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 = "plot4.png", height = 600, width = 800, unit = "px", bg = "transparent")

  # Release screen
  dev.off()
}

plot4_TotalEmissionsPM2.5_USA_Source_CoalCombustion_1999_to_2008()

# Answer 4: PM2.5 Emissions from coal combustion related sources have decreased from
1999-2008.
```