Lab 3

Kyra Liu

ESP 106 Lab 3

Reading, merging, and presenting data on economic development and indoor and outdoor air pollution

1. Reading in the three csv files: gdppercapitaandgini and airpollution

```
#read in air pollution data and set as data frame
airpol <- read.csv("airpollution.csv")
#read in GDP data and set as data frame
gdpcap <- read.csv("gdppercapiandgini.csv")
```

Both data sets are from Our World in Data: ourworldindata.org

The GDP dataset has GDP per capita and the GINI index (a measure of income inequality: https://en.wikipedia.org/wiki/Gini_coefficient)

The air pollution data set has death rates from indoor and outdoor air pollution - units are in deaths per 100,000 people

Note that Indoor air pollution is the Household Air Pollution from Solid Fuels and Outdoor air pollution is split into particulate matter and ozone

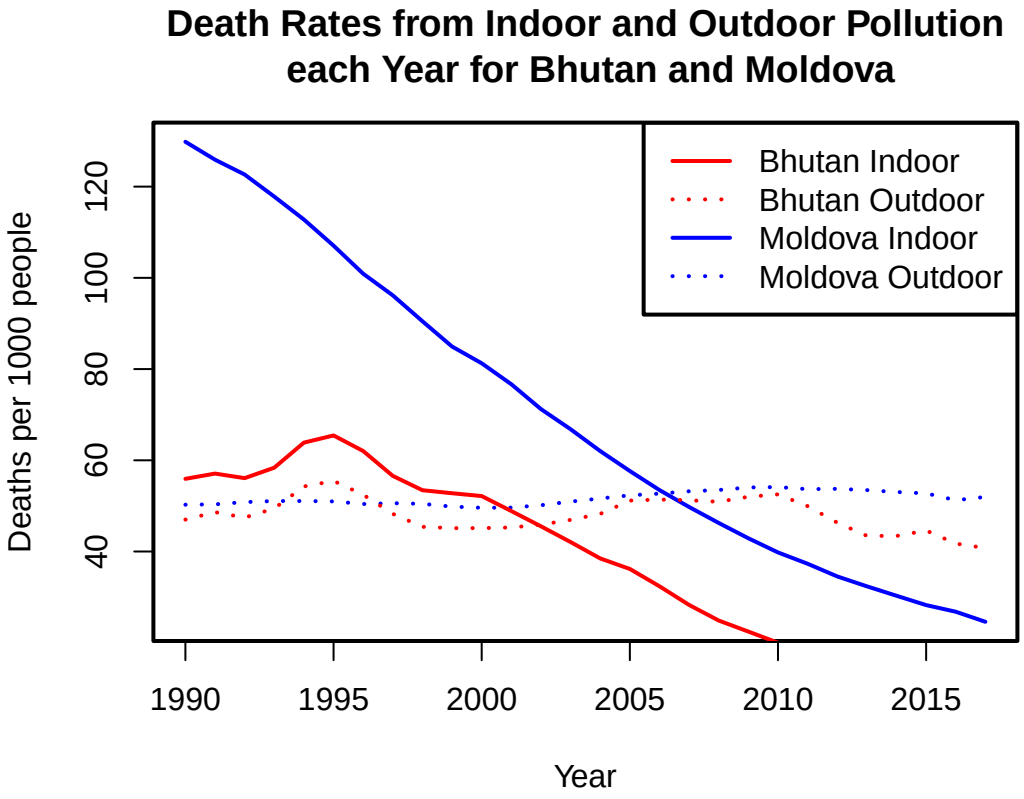
renaming the column names for clarity and ease of use

```
namesair = c("entity","code","year","deaths.ambientpmp", "deaths.householdfuels","deaths.ambientozone",

colnames(airpol) <- namesair


namesgdpcap = c("entity", "code","year","population","continent","ginicoef","gdppercap")

colnames(gdpcap) <- namesgdpcap
```

2. Plotting the death rates from indoor air pollution and outdoor air pollution (sum of particulate matter and ozone) over time for Bhutan and Moldova

```
bhind <-match(airpol$entity, "Bhutan") #returns 1 for values where the country is Bhutan
isnabh = !is.na(bhind)
bhutanair = airpol[isnabh, ]
```

```r
bhind2 = !is.na((match(gdpcap$entity, "Bhutan")))
bhutapgdp = gdpcap[bhind2, ] #bhutan gdp extracted from all gdp

bhutan = merge(bhutanair,bhutapgdp) #merged data of bhutan

#repeating above steps for country 2 (moldova)
mdind<-(match(airpol$entity, "Moldova"))
isnamo = !is.na(mdind)
moldovaair = airpol[isnamo, ]


mdind2<-(match(gdpcap$entity, "Moldova"))
isnamo2 = !is.na(mdind2)
moldovagdp = gdpcap[isnamo2, ]
moldova = merge(moldovaair,moldovagdp)
# defining variables to be used for plotting
yr.bhu = bhutan$year #years that we have data for bhutan
in.bhu = bhutan$deaths.householdfuels # indoor pol deaths for bh
out.bhu = bhutan$deaths.ambientozone+ bhutan$deaths.ambientpmp # outdoor pol deaths
#sum of ozone and pmp deaths

#repeating steps for 2nd country
yr.mol = moldova$year
in.mol = moldova$deaths.householdfuels
out.mol = moldova$deaths.ambientpmp+moldova$deaths.ambientozone

#creating plot with different line types and colors
par(mar = c(5,6,4,4),lwd = 2) #margins and line width
#bhutan indoor deaths
plot(yr.bhu,in.bhu, type ="l", col = "blue",
       xlab = "Year", ylab = "Deaths per 1000 people",
     main = "Death Rates from Indoor and Outdoor Pollution\n each Year for Bhutan and Moldova")
lines(yr.bhu,out.bhu,type ="l", col = "blue", lty = 3) #bh out deaths
lines(yr.mol,in.mol,type="l",col="red") #mol in deaths
lines(yr.mol,out.mol,type = "l",col = "red",lty = 3) #mol out deaths
#create legend
legend("topright",c("Bhutan Indoor","Bhutan Outdoor","Moldova Indoor","Moldova Outdoor"),
      col = c("red","red","blue","blue"), lty=c(1,3,1,3),lwd = 2 )
```



3. Merging the air pollution data with the gdp data

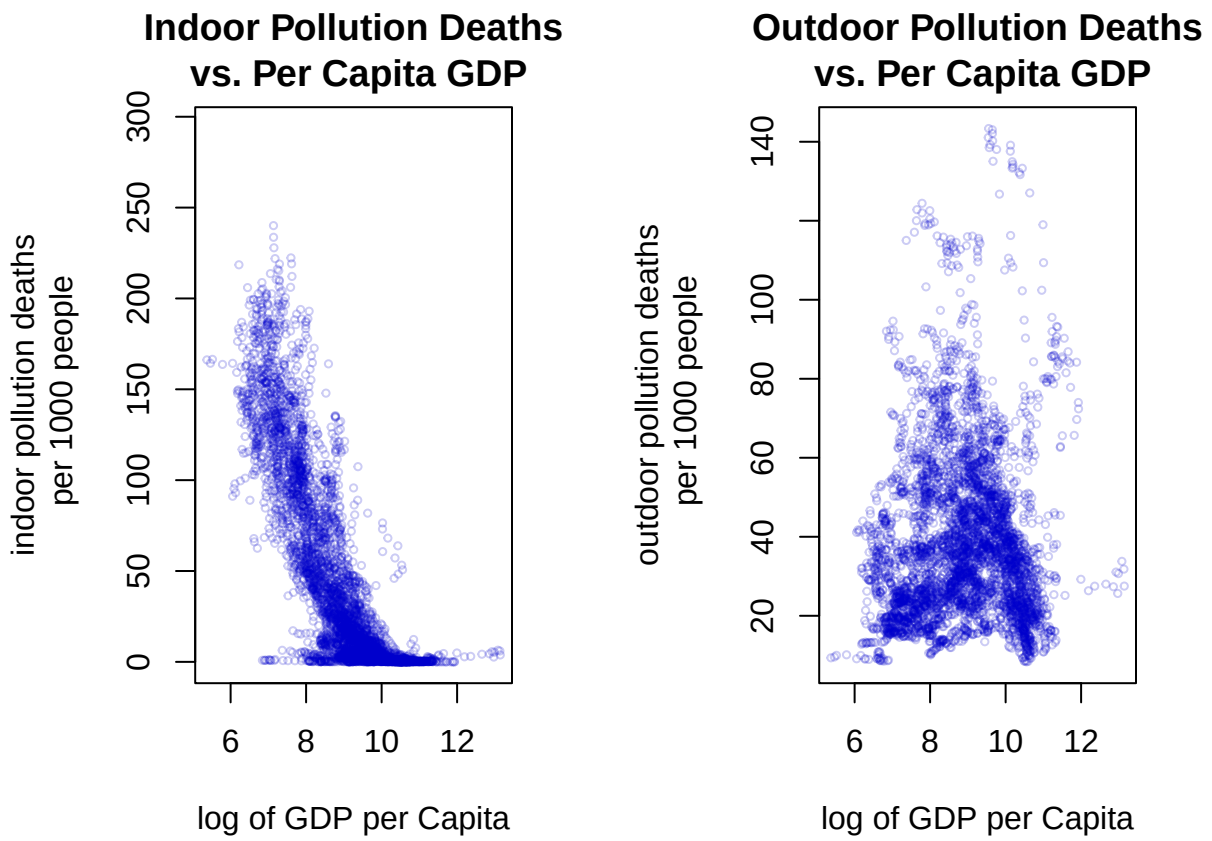
Use of merge() function allows the data to be merged by specified row and column names that they have in common

Merging by "entity" and "year" allows us to have the data that we need organized by the name of entity and the year to be used for plotting

```
all = merge(gdpcap,airpol,by = c("entity", "year" ),no.dups = 1)
```

4. Plotting the relationship between log of per-capita GDP (x axis) and indoor air pollution death rate (y axis) and one showing log of per-capita GDP (x axis) and outdoor air pollution (y axis)

```
par(mfrow = c(1,2),mar = c(4.5,5.5,3,2.5)) #figure with 2 subplots
#all indoor deaths v gdp
plot(log(all$gdppercap),all$deaths.householdfuels,
     main = "Indoor Pollution Deaths\n vs. Per Capita GDP",
     ylab = "indoor pollution deaths\n per 1000 people", xlab= "log of GDP per Capita", col =rgb(0,0,0.8
#all outdoor deaths v gdp
plot(log(all$gdppercap),all$deaths.ambientozone+all$deaths.ambientpmp,
     main = "Outdoor Pollution Deaths\n vs. Per Capita GDP",
     ylab = "outdoor pollution deaths\n per 1000 people", xlab= "log of GDP per Capita",col =rgb(0,0,0.8
```




```
# cont.in = all$continent != ""
# continents = all$continent[cont.in]
# t15.in = all$year == 2015
#
# z = all[t15.in ,c("continent","entity")]
# q = merge(z,all,all= 1)
```

4. Stretch Challenge

Plotting the points color-coded by continent

```
cont.in = all$continent != "" #taking index of nonempty continent values
#continents = all$continent[cont.in]
#t15.in = all$year == 2015
z = all[cont.in ,c("continent","entity")] #creating new df with index
all$continent = NULL #taking out the continent column with the empty values
all = merge(all, z, by="entity") #redefining all as a new merged df with z
colorm = rainbow(6,alpha = 0.5) #creating vector of colors
f = as.factor(all$continent)
cont.names = levels(f) # taking categories
i = as.integer(f) # making factor into integers
coli = colorm[i] #making colors to put into plot command
par(mfrow = c(1,2),mar = c(3,2,3,1.5))
plot(log(all$gdppercap),all$deaths.householdfuels,
     main = "Indoor Pollution Deaths\n vs. Per Capita GDP",
```

```
     ylab = "indoor pollution deaths", xlab= "log of GDP per Capita", col =coli,cex=0.5)

legend("topright",cont.names,pch = 1,col = colorm,pt.cex = 0.5,cex = 0.6)
plot(log(all$gdppercap),all$deaths.ambientozone+all$deaths.ambientpmp,
     main = "Outdoor Pollution Deaths\n vs. Per Capita GDP",
     ylab = "outdoor pollution deaths", xlab= "log of GDP per Capita",col =coli,cex=0.5)

legend("topright",cont.names,pch = 1,col = colorm,pt.cex = 0.5,cex = 0.6)
```