

# Ryan Tillis - R Programming - Functional Programming - Coursera

Ryan Tillis (<http://www.ryantillis.com>)

6/25/2016

## R Programming Project 1

### Part 1

Write a function named 'pollutantmean' that calculates the mean of a pollutant (sulfate or nitrate) across a specified list of monitors. The function 'pollutantmean' takes three arguments: 'directory', 'pollutant', and 'id'. Given a vector monitor ID numbers, 'pollutantmean' reads that monitors' particulate matter data from the directory specified in the 'directory' argument and returns the mean of the pollutant across all of the monitors, ignoring any missing values coded as NA.

```
pollutantmean <- function(directory, pollutant, id = 1:332){
  ##directory is a character vector of length 1               indicating the location of the csv file

  ##id is an integer vector indicating which monitor ids to use

  ##Pollutant is the character vector of length 1 indicating which pollutant to calculate the mean for; either "sulfate" or "nitrate"

  ##Return the mean of the pollutant across all id monitors ignoring NA values

  names <- list.files(directory)[id]
  read <- lapply(paste(directory,"/",names, sep = ""),read.csv)
  return(mean(unlist(lapply(read, function(x){x[,pollutant]})), na.rm=T))
}
```

# 1. What value is returned by the following call to pollutantmean()? You should round your output to 3 digits.

```
pollutantmean("specdata", "sulfate", 1:10)
```

```
## [1] 4.064128
```

# 2. What value is returned by the following call to pollutantmean()? You should round your output to 3 digits.

```
pollutantmean("specdata", "nitrate", 70:72)
```

```
## [1] 1.706047
```

# 3. What value is returned by the following call to `pollutantmean()`? You should round your output to 3 digits.

```
pollutantmean("specdata", "sulfate", 34)
```

```
## [1] 1.477143
```

# 4. What value is returned by the following call to `pollutantmean()`? You should round your output to 3 digits.

```
pollutantmean("specdata", "nitrate")
```

```
## [1] 1.702932
```

## Part 2

Write a function that reads a directory full of files and reports the number of completely observed cases in each data file. The function should return a data frame where the first column is the name of the file and the second column is the number of complete cases. A prototype of this function follows

```
## [1] 228 148 124 165 104 460 232
```

```
## [1] 219
```

```
## [1] 711 135 74 445 178 73 49 0 687 237
```

## Part 3

Write a function that takes a directory of data files and a threshold for complete cases and calculates the correlation between sulfate and nitrate for monitor locations where the number of completely observed cases (on all variables) is greater than the threshold. The function should return a vector of correlations for the monitors that meet the threshold requirement. If no monitors meet the threshold requirement, then the function should return a numeric vector of length 0. A prototype of this function follows

For this function you will need to use the 'cor' function in R which calculates the correlation between two vectors. Please read the help page for this function via '?cor' and make sure that you know how to use it.

```

corr <- function(directory, threshold = 0) {
  ##threshold is a numeric vector length 1 which indicates the n number of completely ob
served observations required to compute the correlation between nitrate and sulfate; d
efault is 0

  ##Return a numeric vector of correlation

  all <- complete(directory)

  if (sum(which(all$nobs>threshold))==0){

    NULL

  }

  else {
    thresh_index <- as.numeric(all[which(all$nobs>threshold),]$id)

    names <- list.files(directory)[thresh_index]
    read <- lapply(paste(directory,"/",names, sep = ""),read.csv)

    return(unlist(lapply(read, function(x)
{cor(x[,2],x[,3],use="pairwise.complete.obs")})))
  }
}

# 8. What value is printed at end of the following code?
cr <- corr("specdata")
cr <- sort(cr)
set.seed(868)
out <- round(cr[sample(length(cr), 5)], 4)
print(out)

```

```
## [1] 0.2688 0.1127 -0.0085 0.4586 0.0447
```

```

# 9. What value is printed at end of the following code?
cr <- corr("specdata", 129)
cr <- sort(cr)
n <- length(cr)
set.seed(197)
out <- c(n, round(cr[sample(n, 5)], 4))
print(out)

```

```
## [1] 243.0000 0.2540 0.0504 -0.1462 -0.1680 0.5969
```

```

# 10. What value is printed at end of the following code?

cr <- corr("specdata", 2000)
n <- length(cr)
cr <- corr("specdata", 1000)
cr <- sort(cr)
print(c(n, round(cr, 4)))

```

```
## [1] 0.0000 -0.0190 0.0419 0.1901
```