

Intro to R for Epidemiologists

Lab 4 (2/5/15)

Many of these questions go beyond the information provided in the lecture. Therefore, you may need to use R help files and the internet to search for answers. Feel free to ask questions of the instructor, the TAs, or your classmates, but try to work through as much as you can independently.

For the lab, you are expected to create an R script (.R file in the R editor) with your code corresponding to each question. Begin each question with a commented line of code indicating the question. As an example:

```
# Jenna Krall  
  
# Question 1.  
head(iris)
```

Part 1. Conditional Statements and For Loops

We will use the `airquality` dataset to explore conditional statements and `for` loops. This dataset contains daily meteorological measurements and ozone levels from 1973. Recall that you can learn more about the R dataset by typing `?airquality`.

1. How many times did the daily ozone measurement exceed 75 parts per billion (ppb), the National Ambient Air Quality Standard for ground-level ozone set by the EPA?
2. Determine if any measurements from January, March, or May are in the dataset (Hint: use `%in%`).
3. How many days had wind speed greater than 7.5 miles per hour and temperature greater than 80 degrees F?
4. For each day in the dataset, use `ifelse` statements within a `for` loop to create an ozone warning variable (see the ozone Air Quality Index <http://www.airnow.gov/index.cfm?action=pubs.aqiguideozone> for more information). Hint: Use `is.na` and the `and` operator in each of your `ifelse` statements.
 - a. If the daily ozone concentration is above 75 ppb, the warning should be "Ozone Alert: Unhealthy levels of ambient ozone today!"
 - b. If the daily ozone concentration is between 60 and 75 ppb, the warning should be "Moderate ozone levels today". Additionally for this case, print the ozone concentration value (i.e., "Ozone level is XXX ppb").
 - c. If the daily ozone concentration is less than 60 ppb, the warning should be "Low ozone level today".
 - d. If the daily ozone concentration is missing, the warning should indicate "Ozone is missing today".
 - e. Add a "catchall" condition for `else` and have this print an error (e.g. "ERROR: this case is not defined!").

```
# Part 1  
#1 Use table to determine number of days ozone exceeds 75  
test <- airquality$Ozone > 75  
table(test)
```

```
## test  
## FALSE TRUE  
## 94 22
```

```
#2. Do we have any measurements for January, March, or May?
```

```
jan_mar_may <- c(1, 3, 5)
jan_mar_may %in% airquality$Month
```

```
## [1] FALSE FALSE TRUE
```

```
#3. How many days had wind speed over 7.5 and Temperature over 80 degrees F?
```

```
wind <- (airquality$Temp > 80) & (airquality$Wind > 7.5)
table(wind)
```

```
## wind
## FALSE TRUE
## 114 39
```

```
#4. Define an ozone warning for each day
```

```
ozone_warning <- vector(length = nrow(airquality))
for (i in 1 : nrow(airquality)) {
  ozone1 <- airquality$Ozone[i]
  if (!is.na(ozone1) && ozone1 >= 75){
    ozone_warning[i] <- "Ozone Alert: Unhealthy level of ozone today!"
  } else if (!is.na(ozone1) && ozone1 >= 60 && ozone1 < 75) {
    ozone_warning[i] <- "Moderate ozone level today."
    cat(c("Ozone level is", ozone1, "ppb\n"))
  } else if (!is.na(ozone1) && ozone1 < 60) {
    ozone_warning[i] <- "Low ozone level today"
  } else if (is.na(ozone1)) {
    ozone_warning[i] <- "Ozone is missing today"
  } else {
    cat("ERROR: this case is not defined!\n")
  }
}
```

```
## Ozone level is 71 ppb
## Ozone level is 64 ppb
## Ozone level is 61 ppb
## Ozone level is 63 ppb
## Ozone level is 64 ppb
## Ozone level is 66 ppb
## Ozone level is 65 ppb
## Ozone level is 73 ppb
## Ozone level is 73 ppb
```

```
table(ozone_warning)
```

```
## ozone_warning
## Low ozone level today
## 85
## Moderate ozone level today.
## 9
## Ozone Alert: Unhealthy level of ozone today!
## 22
## Ozone is missing today
## 37
```

Part 2. Apply statements

Use the `airquality` dataset above to explore apply statements. **For each of these questions, you should use only one line of code.**

1. Find the mean of ozone concentration, solar radiation, wind speed, and temperature.
2. Find the 25th, 50th, and 75th percentiles of ozone concentration for each month.
3. Using the R object you created in (2), find the median of each element in (2). What value does this return?

```
#Part 2
```

```
#1. Find mean ozone concentration, solar radiation, wind speed, and temperature  
apply(airquality[, 1 : 4], 2, mean, na.rm = T)
```

```
##      Ozone      Solar.R      Wind      Temp  
## 42.129310 185.931507   9.957516  77.882353
```

```
#2. 20th and 80th percentiles of ozone by month
```

```
qn_28 <- tapply(airquality$Ozone, airquality$Month, quantile, probs = c(0.25,  
  0.5, 0.75), na.rm = T)
```

```
#3. Find the median of each element in qn_28
```

```
# this is the median ozone by month!
```

```
sapply(qn_28, median)
```

```
##  5  6  7  8  9  
## 18 23 60 52 23
```