# Rclub exercise 2

*Matt Boone*

*October 13, 2015*

We're back again, time for your weekly practice, practice, practice. Again, there's more than one way to do anything in R (as we've seen in our dplyr lesson taught by Des). Work your way as far down as you can go before you start struggling. **DONT GIVE UP**. If things get too hard, feel free to google or check the help files for various functions. R is about googling as much as anything else!

Today we focus on **subsetting** otherwise known as taking a data set and filtering it to our choosing. There's **many** different ways to do this

## Easy

I want to buy a car. Say we have a data set of cars and their various performance statistics. Hey wait we do have that, it comes free with R! Its called the 'mtcars' data set. It's automatically loaded when you type in 'mtcars'

Now say I want to use this data set to help me decide what car to buy. I want to subset this list so I can see the 6 cylinder cars (cyl), but I'm only interested in seeing each cars horsepower (hp) and miles per gallon (mpg). Subset the 'mtcars' data set so I only have 6 cylinder cars and only the two columns 'hp' and 'mpg'

```
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

---

---

## Medium

Now I'm a picky person and I have a lot of needs and wants. What I really want is a car that satisfys these 4 critera.

1. I do not want a heavy car ('wt' is less than 4kg)
2. I want a 4 or a 6 cylinder engine ('cyl' means cylinder)
3. I want it to get atleast 20 miles to the gallon ('mpg')
4. Because I am a scientist, I want the car that has the highest value of this metric I call the 'awesome' metric, which is obviously:

   mpg^(hp) /((10*wt)^(20*cyl))

At the end I want it to **only** show me the car that fits all of these criteria. Subset the data set so it shows me what car I'm going to buy.

**Bonus (Medium-Hard difficulty):** I want a Mercedes (the 'Merc' means Mercedes ) and instead of my awesome metric I want the highest horsepower possible ('hp'). In this situation I want my final dataframe to only contain **ONE** row, my wining entry. So you must choose the name 'Merc' as well, **WITHOUT** manually typing in the number of the row for the Mercedes!

---

## Hard

Part 1. Say we have a 10 x 10 gridded wetland. In each cell we've measured two habitat characteristics water height and percent cover of grasses. We want to characterize each cell as a pond, mudflat, wetland, wet field, or upland based on our two measurements and put this into **one** categorical column called 'HabitatType'.

- pond = water height > 90mm and percent cover of grass<= 0.5
- mudflat = water height <= 90mm and percent cover of grass <= 0.5
- wetland = water height > 50mm and percent cover of grass > 0.5
- wetfield = water height <= 50mm and percent cover of grass > 0.5
- upland = water height <= .1 and percent cover of grass > 0.5
- bare = water height <=.1 and percent cover of grass <0.5

Heres your data set:

```
data <- data.frame(
        lat=rep(seq(30,30.9,.1),each=10),
        long=rep(seq(100,100.9,.1), by=10),
        water = sample(seq(1,120,.1), 100, replace=T),
        PerCover = sample(seq(.3,1,.1), 100, replace=T))

head(data)
```

```
##   lat  long water PerCover
## 1  30 100.0  93.4      0.6
## 2  30 100.1  46.3      0.3
## 3  30 100.2 118.3      1.0
## 4  30 100.3  85.0      0.5
## 5  30 100.4  73.6      0.5
## 6  30 100.5 102.1      0.3
```

**Bonus 1. (Very Hard)** During my field work I noticed that wetlands tend to occur next to other wetlands. I want to figure out how many of my wetland cells have atleast 1 neighbor that is also a wetland (This is a loose measure of contagion in landscape ecology).

Count neighbors as only their north/south/east/west neighbor (4 neighbor rule). Assume anything outside our study area is a 0.

Hint - In this instance you'll have to make a 10x10 matrix with latitude as rows, columns as longitude and wetland is coded as 1 or 0.

**Bonus 2. (INSANITY)  THE COORDINATES ARE POLAR COORDINATES!**

```
data1 <- data.frame(
        range = rep(seq(1,10,1), each=10),
        azimuth = rep(seq(0,350, 36), by=10),
        water = sample(seq(1,120,.1), 100, replace=T),
        PerCover = sample(seq(.3,1,.1), 100, replace=T))

head(data1)
```

```
##   range azimuth water PerCover
## 1     1       0  32.1      0.4
## 2     1      36 102.8      0.6
## 3     1      72  59.8      0.4
## 4     1     108  83.0      0.9
## 5     1     144 109.6      1.0
## 6     1     180  44.8      0.9
```