Getting Started with R

Create a numeric vector containing the numbers 2.23, 3.45, 1.87, 2.11, 7.33, 18.34, 19.23. What is the average of these numbers?

```
a <- c(2.23, 3.45, 1.87, 2.11, 7.33, 18.34, 19.23)
mean(a)
```

```
## [1] 7.794286
```

Use a for loop to determine the value of $\sum_{i=1}^{25} i^2$

```
sum <- 0
for(i in 1:25)
  sum <- sum + i^2
sum
```

```
## [1] 5525
```

The cars dataset is available in base R. You can type cars to see it. Use the class function to determine what type of object is cars.

```
class(cars)
```

```
## [1] "data.frame"
```

How many rows does the cars object have?

```
nrow(cars)
```

```
## [1] 50
```

What is the name of the second column of cars?

```
names(cars)[2]
```

```
## [1] "dist"
```

The simplest way to extract the columns of a matrix or data.frame is using [. For example you can access the second column with cars[,2]. What is the average distance traveled in this dataset?

```
mean(cars[,2])
```

```
## [1] 42.98
```

Familiarize yourself with the which function. What row of cars has a a distance of 85?

```
which(cars[,2]==85)
```

```
## [1] 50
```

Here we will test some of the basics of R data manipulation which you should know or should have learned by following the tutorials above. You will need to have the file femaleMiceWeights.csv in your working directory. As we showed above, one way to do this is by using the downloader package:

```
library(downloader)
url <- "https://raw.githubusercontent.com/genomicsclass/dagdata/master/inst/extdata/femaleMiceWeights.cs
filename <- "femaleMiceWeights.csv"
download(url, destfile=filename)
```

Read in the file femaleMiceWeights.csv and report the exact name of the column containing the weights.

```
micewt <- read.csv(filename)
colnames(micewt)
```

```
## [1] "Diet"       "Bodyweight"
```

The [and] symbols can be used to extract specific rows and specific columns of the table. What is the entry in the 12th row and second column?

```
micewt[12,2]
```

```
## [1] 26.25
```

You should have learned how to use the $ character to extract a column from a table and return it as a vector. Use $ to extract the weight column and report the weight of the mouse in the 11th row.

```
weights <- micewt$Bodyweight
weights[11]
```

```
## [1] 26.91
```

The length function returns the number of elements in a vector. How many mice are included in our dataset?

```
length(micewt$Bodyweight)
```

```
## [1] 24
```

To create a vector with the numbers 3 to 7, we can use seq(3,7) or, because they are consecutive, 3:7. View the data and determine what rows are associated with the high fat or hf diet. Then use the mean function to compute the average weight of these mice.

```
hfdiet<-micewt[micewt$Diet=="hf",]
mean(hfdiet$Bodyweight)
```

```
## [1] 26.83417
```

One of the functions we will be using often is sample. Read the help file for sample using ?sample. Now take a random sample of size 1 from the numbers 13 to 24 and report back the weight of the mouse represented by that row. Make sure to type set.seed(1) to ensure that everybody gets the same answer.

```r
weights <- hfdiet$Bodyweight
mean( weights[ 13:24 ])
```

```
## [1] NA
```

Or

```r
set.seed(1)
i <- sample( 13:24, 1)
micewt$Bodyweight[i]
```

```
## [1] 25.34
```

Read in the msleep_ggplot2.csv file with the function read.csv and use the function class to determine what type of object is returned.

```r
library(downloader)
url="https://raw.githubusercontent.com/genomicsclass/dagdata/master/inst/extdata/msleep_ggplot2.csv"
filename <- basename(url)
download(url,filename)

mammalian<-read.csv("msleep_ggplot2.csv")
class(mammalian)
```

```
## [1] "data.frame"
```

Now use the filter function to select only the primates. How many animals in the table are primates? Hint: the nrow function gives you the number of rows of a data frame or matrix.

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
primates <- filter(mammalian, order=="Primates") #keep only the ones with Primates order
nrow(primates)
```

```
## [1] 12
```

What is the class of the object you obtain after subsetting the table to only include primates?

```
class(primates)
```

```
## [1] "data.frame"
```

Now use the select function to extract the sleep (total) for the primates. What class is this object? Hint: use %>% to pipe the results of the filter function to select.

```
primateSleep<-filter(mammalian, order=="Primates") %>% select(sleep_total)
class(primateSleep)
```

```
## [1] "data.frame"
```

Now we want to calculate the average amount of sleep for primates (the average of the numbers computed above). One challenge is that the mean function requires a vector so, if we simply apply it to the output above, we get an error. Look at the help file for unlist and use it to compute the desired average.

```
y <- filter(mammalian, order=="Primates") %>% select(sleep_total) %>% unlist
mean( y )
```

```
## [1] 10.5
```

For the last exercise, we could also use the dplyr summarize function. We have not introduced this function, but you can read the help file and repeat exercise 5, this time using just filter and summarize to get the answer.

```
y <- filter(mammalian, order=="Primates") %>% select(sleep_total) %>% summarise(mean(sleep_total))
y
```

```
##   mean(sleep_total)
## 1              10.5
```