Max Hammond "Lists and Dataframes"

30 January, 2023

Write the R code to answer the following questions. You have until the beginning of next class to answer all of the questions below and commit to GitHub, both the .Rmd file and the .pdf.

Question 1 - Lists

Run this code, then answer the questions below.

```
data(cars)
mylm <- unclass(lm(dist ~ speed, data = cars))
```

1a

Is `mylm` a list? yes, a list of 12

```
str(mylm)
```

```
## List of 12
##  $ coefficients : Named num [1:2] -17.58 3.93
##   ..- attr(*, "names")= chr [1:2] "(Intercept)" "speed"
##  $ residuals    : Named num [1:50] 3.85 11.85 -5.95 12.05 2.12 ...
##   ..- attr(*, "names")= chr [1:50] "1" "2" "3" "4" ...
##  $ effects      : Named num [1:50] -303.914 145.552 -8.115 9.885 0.194 ...
##   ..- attr(*, "names")= chr [1:50] "(Intercept)" "speed" "" "" ...
##  $ rank         : int 2
##  $ fitted.values: Named num [1:50] -1.85 -1.85 9.95 9.95 13.88 ...
##   ..- attr(*, "names")= chr [1:50] "1" "2" "3" "4" ...
##  $ assign       : int [1:2] 0 1
##  $ qr           :List of 5
##   ..$ qr   : num [1:50, 1:2] -7.071 0.141 0.141 0.141 0.141 ...
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:50] "1" "2" "3" "4" ...
##   .. .. ..$ : chr [1:2] "(Intercept)" "speed"
##   .. ..- attr(*, "assign")= int [1:2] 0 1
##   ..$ qraux: num [1:2] 1.14 1.27
##   ..$ pivot: int [1:2] 1 2
##   ..$ tol  : num 1e-07
##   ..$ rank : int 2
##   ..- attr(*, "class")= chr "qr"
##  $ df.residual  : int 48
##  $ xlevels      : Named list()
##  $ call         : language lm(formula = dist ~ speed, data = cars)
```

```
## $ terms        :Classes 'terms', 'formula'  language dist ~ speed
##   .. ..- attr(*, "variables")= language list(dist, speed)
##   .. ..- attr(*, "factors")= int [1:2, 1] 0 1
##   .. .. ..- attr(*, "dimnames")=List of 2
##   .. .. .. ..$ : chr [1:2] "dist" "speed"
##   .. .. .. ..$ : chr "speed"
##   .. ..- attr(*, "term.labels")= chr "speed"
##   .. ..- attr(*, "order")= int 1
##   .. ..- attr(*, "intercept")= int 1
##   .. ..- attr(*, "response")= int 1
##   .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
##   .. ..- attr(*, "predvars")= language list(dist, speed)
##   .. ..- attr(*, "dataClasses")= Named chr [1:2] "numeric" "numeric"
##   .. .. ..- attr(*, "names")= chr [1:2] "dist" "speed"
## $ model        :'data.frame':   50 obs. of  2 variables:
##   ..$ dist : num [1:50] 2 10 4 22 16 10 18 26 34 17 ...
##   ..$ speed: num [1:50] 4 4 7 7 8 9 10 10 10 11 ...
##   ..- attr(*, "terms")=Classes 'terms', 'formula'  language dist ~ speed
##   .. .. ..- attr(*, "variables")= language list(dist, speed)
##   .. .. ..- attr(*, "factors")= int [1:2, 1] 0 1
##   .. .. .. ..- attr(*, "dimnames")=List of 2
##   .. .. .. .. ..$ : chr [1:2] "dist" "speed"
##   .. .. .. .. ..$ : chr "speed"
##   .. .. ..- attr(*, "term.labels")= chr "speed"
##   .. .. ..- attr(*, "order")= int 1
##   .. .. ..- attr(*, "intercept")= int 1
##   .. .. ..- attr(*, "response")= int 1
##   .. .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
##   .. .. ..- attr(*, "predvars")= language list(dist, speed)
##   .. .. ..- attr(*, "dataClasses")= Named chr [1:2] "numeric" "numeric"
##   .. .. .. ..- attr(*, "names")= chr [1:2] "dist" "speed"
```

1b

What is the class of the element named `model` in the `mylm` object? 'model' = data.frame

```
class(mylm$model)
```

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

1c

Change the column names of the element named `model` in the `mylm` object to `Y` and `X`.

```
colnames(mylm$model)
```

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

```
colnames(mylm$model) <- c("Y", "X")
```

1d

Assign the element named `model` in the `mylm` object to a new object called `df`. What is the class of this object? Is it the same class or different than the object in 1b? Why?

Data frame. Because we took the data frame from model and just and put it into another object called df which will also just be a dataframe

```
df <- mylm$model
class(df)
```

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

Question 2 - Dataframes

In this exercise, you will create a new column in the data frame based on values in an existing column. Use the following code to create the data frame.

```
set.seed(123)
df <- data.frame(Y = runif(10, 0, 1),
                 X = rbinom(10, 1, .5))
df$condition <- NA
```

2a

Populate the entries of the variable `condition` with the string "Treatment" if the corresponding value of X is 1 and "Control" otherwise.

```
df[df$X == 1, "condition"] <- "Treatment"
df[df$X == 0, "condition"] <- "Control"
```

2b

What is the mean value of `Y` for the observations where `condition` is "Treatment"?

```
mean(df[df$condition == "Treatment", "Y"])
```

```
## [1] 0.4163486
```

Question 3

In a few words, explain a few reasons you might decide to use a list vs. a dataframe (or vice versa) for a programming task. What are some different use cases for each?

We would use a list if we want a list of numbers and we would use a dataframe if we want to look at a matrix with more variables. Lists can be good if we need to view just a list of numbers or strings just to get the basic data layout. They are preferable for smaller datasets. For larger datasets it would be better to use dataframe and then View them laid out. We can access columns in that data set using "$."