Week 2 RHMI Exercise Answers

Andrew Perfors

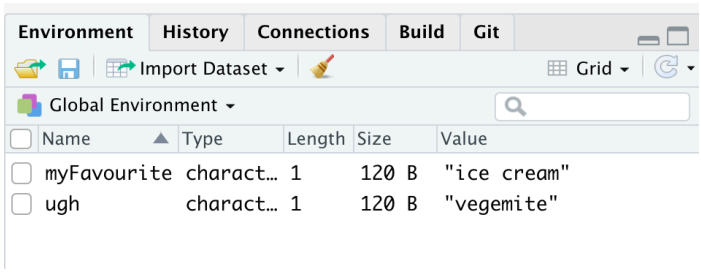
Day 1

Packages

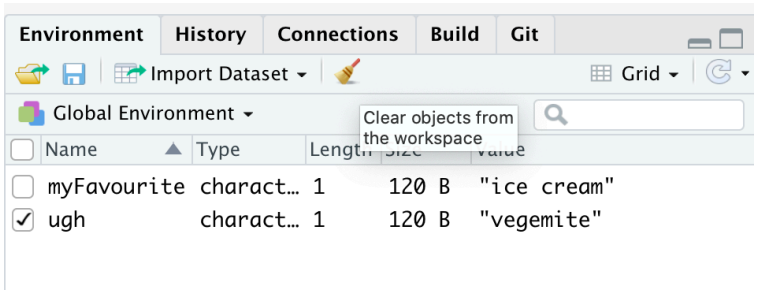
1. Make a variable called `myFavourite` with the name of your favourite food, and another called `ugh` with one of your least favourites. Check in the environment panel to make sure they are in your workspace, and then remove `ugh`.

```
> myFavourite <- "ice cream"
> ugh <- "vegemite"
```

Your environment panel should look something like this (in Grid view):



Then remove `ugh` by doing this (with the icon on the broom):



2. Install and load the package called `lsr`. Use the `who()` function in it to figure out what's in your environment. Use the `rm()`function to delete the variable `myFavourite`. (hint: remember to try the `help()` function when you don't know what to do!)

```
> install.packages("lsr")
trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.6/
lsr_0.5.tgz'
Content type 'application/x-gzip' length 217978 bytes (212 KB)
==================================================
downloaded 212 KB

The downloaded binary packages are in
/var/folders/wj/4y3l7vrx3w7__c6qzr4psrs80000gn/T//RtmpLISqFI/
downloaded_packages
```

Then we need to remember to load it and use who() to figure out what's there:

```
> library(lsr)
> who()
   -- Name --      -- Class --    -- Size --
   myFavourite     character      1
```

Apparently there is one variable, a character variable (size is one because it's equivalent to a vector with one item in it). Now let's get rid of it:

```
> rm(myFavourite)
> who()
No variables found
```

Manipulation

1. Load the `bunnysurvey.csv` dataset. Make a new dataframe/tibble called `d` which is just a copy of `data`.

   To load it use the following commands, having made sure you opened R by clicking on the Rproj file in the directory with it in there, and also loaded the libraries you need:

   ```
   > library(here)
   > library(tidyverse)
   > loc <- here("bunnysurvey.csv")
   > data <- read_csv(file=loc)
   ```

   Then to create a new one type:

   ```
   > d <- data
   ```

2. In `d`, add 1 to every entry for height. Then subtract it again.

   ```
   > d$height <- d$height+1
   > d$height
   [1] 21 19 21 21 25 25 18
   > d$height <- d$height-1
   > d$height
   [1] 20 18 20 20 24 24 17
   ```

3. Create a new variable in `d` called `dislikesDogs` which is `TRUE` if that person ranked dogs as #3, and `FALSE` otherwise.

   ```
   > d$dislikesDogs <- d$doggyrank==3
   ```

   We can check if this works by looking at both variables and seeing if the `TRUE` values in `dislikesDogs` correspond only to the 3 values in `doggyrank`:

   ```
   > d$dislikesDogs
   [1] FALSE FALSE  TRUE FALSE  TRUE    NA FALSE
   > d$doggyrank
   [1]  2  2  3  2  3 NA  1
   ```

4. Create a new variable in `d` called `inches` which gives the height in inches (hint: inches is cm divided by 2.54).

```
> d$inches <- d$height/2.54
```

And to check:

```
> d$height
[1] 20 18 20 20 24 24 17
> d$inches
[1] 7.874016 7.086614 7.874016 7.874016 9.448819 9.448819 6.692913
```

5. Select the first three rows (with all columns) out of `d`.

```
> d[1:3,]
    name colour height bunnyrank bearrank doggyrank
1  bunny   grey     20         1        3         2
2 gladly purple     18         3        1         2
3 flopsy  black     20         1        2         3
```

6. Select only the rows of `d` that contain an `NA` for the `colour` variable. (hint: use the `is.na()` function). For an extra challenge, try to select only the rows of `d` that do not contain an `NA` for the `colour` variable. (Remember our logical operators from last week).

```
> d[is.na(d$colour),]
        name colour height bunnyrank bearrank doggyrank likesBunnies   inches
6 cuddly paws   <NA>     24        NA       NA        NA           NA 9.448819
```

To select the ones that don't have `NA`, we use the `!` operator, which means "not":

```
> d[!is.na(d$colour),]
    name colour height bunnyrank bearrank doggyrank likesBunnies   inches
1  bunny   grey     20         1        3         2         TRUE 7.874016
2 gladly purple     18         3        1         2        FALSE 7.086614
3 flopsy  black     20         1        2         3         TRUE 7.874016
4 shadow    red     20         1        3         2         TRUE 7.874016
5    lfb purple     24         1        2         3         TRUE 9.448819
7 doggie   blue     17         2        3         1        FALSE 6.692913
```