homework2

Claire Whittington

2023-10-06

(1) Exercise 1 − 5 pts. In exercise 1:

- (1 pts) Import the correct sheet from Example_5, without modifying it in any way.

```
# grab sheet 2 and put it into my.trees
my.trees <- read_excel("~/Downloads/Example_5.xls", sheet = 2)
```

```
## New names:
## * '' -> '...2'
## * '' -> '...3'
## * '' -> '...4'
## * '' -> '...5'
```

```
my.trees
```

```
## # A tibble: 35 x 5
##    A data set of 31 Cherry trees. This is a commonly u~1 ...2  ...3  ...4  ...5
##    <chr>                                                 <chr> <chr> <lgl> <lgl>
##  1 I have no idea where the orignial data set came fro~  <NA>  <NA>  NA    NA
##  2 Girth is diameter at breast height (in inches), Heig~ <NA>  <NA>  NA    NA
##  3 <NA>                                                  <NA>  <NA>  NA    NA
##  4 Girth                                                 Heig~ Volu~ NA    NA
##  5 8.3000000000000007                                    70    10.3~ NA    NA
##  6 8.5999999999999996                                    65    10.3~ NA    NA
##  7 8.8000000000000007                                    63    10.1~ NA    NA
##  8 10.5                                                  72    16.3~ NA    NA
##  9 10.699999999999999                                    81    18.8~ NA    NA
## 10 10.800000000000001                                    83    19.6~ NA    NA
## # i 25 more rows
## # i abbreviated name:
## #   1: 'A data set of 31 Cherry trees. This is a commonly used set of data that is available in R.'
```

- (½ pt) How many blank columns were imported?

  - Two blank columns

- (½ pt) Show the structure of the structure of the data using str(yourdataframe'sname)

```
str(my.trees)
```

```
## tibble [35 x 5] (S3: tbl_df/tbl/data.frame)
##  $ A data set of 31 Cherry trees. This is a commonly used set of data that is available in R.: chr [
##  $ ...2                                                                                      : chr [
##  $ ...3                                                                                      : chr [
##  $ ...4                                                                                      : logi
##  $ ...5                                                                                      : logi
```

- (3 pts) Modify the data frame so that it only has 3 columns of 31 observations. Rename the columns to have the appropriate names of "Girth", "Height," and "Volume" if your step above did not automatically rename the columns.

```
# remove rows 1,2,3 and cols 4 and 5
cleaned_data <- my.trees[-c(0:4),-c(4,5)]

# rename cols
colnames(cleaned_data)[1] <- "Girth"
colnames(cleaned_data)[2] <- "Height"
colnames(cleaned_data)[3] <- "Volume"

# show structure summary
str(cleaned_data)
```

```
## tibble [31 x 3] (S3: tbl_df/tbl/data.frame)
##  $ Girth : chr [1:31] "8.3000000000000007" "8.5999999999999996" "8.8000000000000007" "10.5" ...
##  $ Height: chr [1:31] "70" "65" "63" "72" ...
##  $ Volume: chr [1:31] "10.300000000000001" "10.300000000000001" "10.199999999999999" "16.399999999999
```

```
# show data
head(cleaned_data)
```

```
## # A tibble: 6 x 3
##    Girth              Height Volume
##    <chr>              <chr>  <chr>
## 1 8.3000000000000007 70     10.300000000000001
## 2 8.5999999999999996 65     10.300000000000001
## 3 8.8000000000000007 63     10.199999999999999
## 4 10.5               72     16.399999999999999
## 5 10.699999999999999 81     18.800000000000001
## 6 10.800000000000001 83     19.699999999999999
```

- Your output should have the answer to question in the 2nd bulletin point, the structure of the data frame, and the final data frame. Include all code used to clean up the data frame.

(2) Exercise 2 − 4 pts. In exercise 2,

- (1 pts) Import the correct sheet from Example_5, without modifying it in any way.

```r
my.cars <- read_excel("~/Downloads/Example_3.xls", sheet = 2)
```

```
## New names:
## * `` -> `...13`
## * `` -> `...14`
## * `` -> `...15`
## * `` -> `...16`
## * `` -> `...17`
## * `` -> `...18`
## * `` -> `...19`
## * `` -> `...20`
## * `` -> `...21`
## * `` -> `...22`
## * `` -> `...23`
## * `` -> `...24`
## * `` -> `...25`
## * `` -> `...26`
## * `` -> `...27`
## * `` -> `...28`
## * `` -> `...29`
## * `` -> `...30`
## * `` -> `...31`
## * `` -> `...32`
## * `` -> `...33`
## * `` -> `...34`
## * `` -> `...35`
## * `` -> `...36`
## * `` -> `...37`
## * `` -> `...38`
## * `` -> `...39`
## * `` -> `...40`
## * `` -> `...41`
## * `` -> `...42`
## * `` -> `...43`
## * `` -> `...44`
## * `` -> `...45`
## * `` -> `...46`
## * `` -> `...47`
## * `` -> `...48`
## * `` -> `...49`
## * `` -> `...50`
## * `` -> `...51`
## * `` -> `...52`
## * `` -> `...53`
## * `` -> `...54`
## * `` -> `...55`
## * `` -> `...56`
## * `` -> `...57`
## * `` -> `...58`
## * `` -> `...59`
## * `` -> `...60`
## * `` -> `...61`
## * `` -> `...62`
```

```
## * `` -> `...63`
## * `` -> `...64`
## * `` -> `...65`
## * `` -> `...66`
## * `` -> `...67`
## * `` -> `...68`
## * `` -> `...69`
## * `` -> `...70`
## * `` -> `...71`
## * `` -> `...72`
## * `` -> `...73`
## * `` -> `...74`
## * `` -> `...75`
## * `` -> `...76`
## * `` -> `...77`
## * `` -> `...78`
## * `` -> `...79`
## * `` -> `...80`
## * `` -> `...81`
## * `` -> `...82`
```

```
head(my.cars)
```

```
## # A tibble: 6 x 82
##    model    mpg cyl  disp    hp drat    wt qsec vs      am gear carb  ...13
##    <chr>  <dbl> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <chr> <lgl>
## 1 Mazda~  21   6     160    110  3.9  2.62 16.5 0       1    4 4     NA
## 2 Mazda~  21   6     160    110  3.9  2.88 17.0 0       1    4 4     NA
## 3 Datsu~  22.8 4     108     93  3.85 2.32 18.6 1       1    4 1     NA
## 4 Horne~  21.4 6     258    110  3.08 3.22 19.4 1       0    3 1     NA
## 5 Horne~  18.7 8     360    175  3.15 3.44 17.0 0       0    3 2     NA
## 6 Valia~  18.1 6     225    105  2.76 3.46 20.2 1       0    3 1     NA
## # i 69 more variables: ...14 <lgl>, ...15 <lgl>, ...16 <lgl>, ...17 <lgl>,
## #   ...18 <lgl>, ...19 <lgl>, ...20 <lgl>, ...21 <lgl>, ...22 <lgl>,
## #   ...23 <lgl>, ...24 <lgl>, ...25 <lgl>, ...26 <lgl>, ...27 <lgl>,
## #   ...28 <lgl>, ...29 <lgl>, ...30 <lgl>, ...31 <lgl>, ...32 <lgl>,
## #   ...33 <lgl>, ...34 <lgl>, ...35 <lgl>, ...36 <lgl>, ...37 <lgl>,
## #   ...38 <lgl>, ...39 <lgl>, ...40 <lgl>, ...41 <lgl>, ...42 <lgl>,
## #   ...43 <lgl>, ...44 <lgl>, ...45 <lgl>, ...46 <lgl>, ...47 <lgl>, ...
```

- (2 pts) Use View(yourdataframe'sname). Write 2 – 3 sentences contrasting what you saw in the Excel file and what the data frame looks like to R.

```
head(my.cars)
```

```
## # A tibble: 6 x 82
##    model    mpg cyl  disp    hp drat    wt qsec vs      am gear carb  ...13
##    <chr>  <dbl> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <chr> <lgl>
## 1 Mazda~  21   6     160    110  3.9  2.62 16.5 0       1    4 4     NA
## 2 Mazda~  21   6     160    110  3.9  2.88 17.0 0       1    4 4     NA
## 3 Datsu~  22.8 4     108     93  3.85 2.32 18.6 1       1    4 1     NA
## 4 Horne~  21.4 6     258    110  3.08 3.22 19.4 1       0    3 1     NA
```

```
## 5 Horne~  18.7 8     360     175  3.15  3.44  17.0 0        0     3 2     NA
## 6 Valia~  18.1 6     225     105  2.76  3.46  20.2 1        0     3 1     NA
## # i 69 more variables: ...14 <lgl>, ...15 <lgl>, ...16 <lgl>, ...17 <lgl>,
## #   ...18 <lgl>, ...19 <lgl>, ...20 <lgl>, ...21 <lgl>, ...22 <lgl>,
## #   ...23 <lgl>, ...24 <lgl>, ...25 <lgl>, ...26 <lgl>, ...27 <lgl>,
## #   ...28 <lgl>, ...29 <lgl>, ...30 <lgl>, ...31 <lgl>, ...32 <lgl>,
## #   ...33 <lgl>, ...34 <lgl>, ...35 <lgl>, ...36 <lgl>, ...37 <lgl>,
## #   ...38 <lgl>, ...39 <lgl>, ...40 <lgl>, ...41 <lgl>, ...42 <lgl>,
## #   ...43 <lgl>, ...44 <lgl>, ...45 <lgl>, ...46 <lgl>, ...47 <lgl>, ...
```

In the excel file, I just see 12 columns all properly named with their corresponding data. In R, there are 82 columns and like 1600 rows with NA values

- (1 pt) Modify the data frame so that it only has the 12 columns x 33 rows of data.

```r
cleaned_data <- my.cars[1:33,1:12]
cleaned_data
```

```
## # A tibble: 33 x 12
##    model         mpg cyl   disp     hp  drat    wt  qsec vs       am  gear carb
##    <chr>       <dbl> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <chr>
## 1 Mazda RX4      21  6     160     110  3.9   2.62  16.5 0        1     4 4
## 2 Mazda RX4 ~    21  6     160     110  3.9   2.88  17.0 0        1     4 4
## 3 Datsun 710    22.8 4     108      93  3.85  2.32  18.6 1        1     4 1
## 4 Hornet 4 D~   21.4 6     258     110  3.08  3.22  19.4 1        0     3 1
## 5 Hornet Spo~   18.7 8     360     175  3.15  3.44  17.0 0        0     3 2
## 6 Valiant       18.1 6     225     105  2.76  3.46  20.2 1        0     3 1
## 7 Duster 360    14.3 8     360     245  3.21  3.57  15.8 0        0     3 4
## 8 Merc 240D     24.4 4     146.~    62  3.69  3.19  20   1        0     4 2
## 9 Merc 230      22.8 4     140.~    95  3.92  3.15  22.9 1        0     4 2
## 10 Merc 280     19.2 6     167.~   123  3.92  3.44  18.3 1        0     4 4
## # i 23 more rows
```

(3) Turned in by the due date/time – 1 point