

Lab: Data I/O

36-600

Fall 2022

Question 1

Download `simple.txt` from the Canvas site. It is in the `DATA` directory. Use an external viewer (your choice) to look at the file. Then apply an appropriate function to read the file's contents into `R`. Show the names of the columns. Make sure the names are correct, and that there are eight columns. (Use `ncol()` to determine the number of columns.) (Note: you may find that your first choice of function does not provide optimal results. If so, try another function. Note that `read_delim()` with a properly formatted regex might help here. If you use base `R` functionality, consider including the argument `stringsAsFactors=FALSE`.)

```
# Base R
df = read.table("DATA/simple.txt",header=TRUE,stringsAsFactors=FALSE)
names(df)
```

```
## [1] "name"      "u"         "g"         "r"         "i"         "z"         "y"         "red
shift"
```

```
ncol(df)
```

```
## [1] 8
```

```
# readr - read_table doesn't work, guesses one column total
suppressMessages(library(tidyverse))
tbl = read_delim("DATA/simple.txt",delim=" ")
```

```
##

Rows: 9 Columns: 8
## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;
##

## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;

— Column specification —


---


## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;

Delimiter: " "
## chr (1): name
## dbl (7): u, g, r, i, z, y, redshift
##

## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;
##

##

## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;

i Use `spec()` to retrieve the full column specification for this data.
## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;

i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;
```

```
names(tbl)
```

```
## [1] "name"      "u"         "g"         "r"         "i"         "z"         "y"         "redshift"
```

```
ncol(tbl)
```

```
## [1] 8
```

Question 2

Read in the data file from Question 1 but skip the header. Display the names that `R` gives to the columns. For `readr`, you may find sub-optimal results; if so, you would need to specify the column names as an argument. (You need not actually do that here.)

```
# Base R
df = read.table("DATA/simple.txt", skip=1)
names(df)
```

```
## [1] "v1" "v2" "v3" "v4" "v5" "v6" "v7" "v8"
```

```
# readr
tbl = read_delim("DATA/simple.txt", delim=" ", skip=1)
```

```
##

Rows: 8 Columns: 8
## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;
##

## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;

— Column specification —


---



## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;

Delimiter: " "
## chr (1): galaxy.A
## dbl (7): 17.8313, 16.9077, 16.4431, 16.2099, 16.0613, 15.8732, 0.038356
##

## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;
##

##

## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;

i Use `spec()` to retrieve the full column specification for this data.
## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;

i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## [36m i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.Rmd [34m Lab_03T_Sol.Rmd [39m ]8;; into ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32m Lab_03T_Sol.html [39m ]8;;
```

```
names(tbl)
```

```
## [1] "galaxy.A" "17.8313" "16.9077" "16.4431" "16.2099" "16.0613" "15.8732" "0.038356"
```

Question 3

Read in the data file from Question 1 but only read in the first four lines, while retaining the header.

```
# Base R
df = read.table("DATA/simple.txt",header=TRUE,nrows=4)
df
```

```
##      name      u      g      r      i      z      y redshift
## 1 galaxy.A 17.8313 16.9077 16.4431 16.2099 16.0613 15.8732 0.038356
## 2 galaxy.B 19.0731 17.7448 16.9789 16.5288 16.2551 15.9531 0.058309
## 3 galaxy.C 21.6380 21.0106 20.8286 20.6283 20.6552 20.5280 0.063701
## 4 galaxy.D 20.5474 19.5542 19.2387 19.0568 19.0887 18.9865 0.059006
```

```
# readr
tbl = read_delim("DATA/simple.txt",delim=" ",n_max=4)
```

```
##

Rows: 4 Columns: 8
## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;
##

## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

— Column specification —————

## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

Delimiter: " "
## chr (1): name
## dbl (7): u, g, r, i, z, y, redshift
##

## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;
##

##

## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

i Use `spec()` to retrieve the full column specification for this data.
## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;
```

```
tbl
```

```
## # A tibble: 4 × 8
##   name          u          g          r          i          z          y redshift
##   <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>    <dbl>
## 1 galaxy.A    17.8   16.9   16.4   16.2   16.1   15.9     0.0384
## 2 galaxy.B    19.1   17.7   17.0   16.5   16.3   16.0     0.0583
## 3 galaxy.C    21.6   21.0   20.8   20.6   20.7   20.5     0.0637
## 4 galaxy.D    20.5   19.6   19.2   19.1   19.1   19.0     0.0590
```

Question 4

Download `planets_small.csv` from the Canvas site. It is in the `DATA` directory. Use an external viewer (your choice) to look at the file. Then apply an appropriate function to read the file's contents into `R`. Note that here, you have one column that should be rendered as character strings (`pl_hostname`, the first column), while the rest should be rendered as factor variables. Thus you do not want to use the `stringsAsFactors` argument here, as it is too coarse. You need to explicitly specify the types of each column. (There's an example of how to do this in the notes, with base `R` functionality! Or you can try using the argument `col_types` in `read_csv()`.)

```
# Base R
df = read.csv("DATA/planets_small.csv", colClasses=c("character", "factor", "factor", "factor"))
df
```

```
##           pl_hostname pl_letter  pl_discmethod pl_pnum
## 1             11 Com          b Radial Velocity      1
## 2             11 UMi          b Radial Velocity      1
## 3             14 And          b Radial Velocity      1
## 4             14 Her          b Radial Velocity      1
## 5             16 Cyg B          b Radial Velocity      1
## 6             18 Del          b Radial Velocity      1
## 7 1RXS J160929.1-210524          b      Imaging      1
## 8             24 Sex          b Radial Velocity      2
## 9             24 Sex          c Radial Velocity      2
## 10 2MASS J01225093-2439505          b      Imaging      1
```

```
# readr
tbl = read_csv("DATA/planets_small.csv", col_types="cfff")
tbl
```

```
## # A tibble: 10 × 4
##   pl_hostname      pl_letter pl_discmethod  pl_pnum
##   <chr>          <fct>      <fct>          <fct>
## 1 11 Com         b          Radial Velocity 1
## 2 11 UMi        b          Radial Velocity 1
## 3 14 And         b          Radial Velocity 1
## 4 14 Her         b          Radial Velocity 1
## 5 16 Cyg B       b          Radial Velocity 1
## 6 18 Del         b          Radial Velocity 1
## 7 1RXS J160929.1-210524 b          Imaging         1
## 8 24 Sex         b          Radial Velocity 2
## 9 24 Sex         c          Radial Velocity 2
## 10 2MASS J01225093-2439505 b          Imaging         1
```

Question 5

Download `students.txt` from the Canvas site. It is in the `DATA` directory. Use an external viewer (your choice) to look at the file. Then apply an appropriate function to read the file's contents into `R`. Use an appropriate argument that ensures that each non-numerical column is treated as a vector of strings, and not a factor variable. (You may have done this back up in Question 1.) Show that you've done this by displaying the type of the variable the column `Last.Name`. (Use `typeof()`.)

```
# Base R
df = read.table("DATA/students.txt",header=TRUE,stringsAsFactors=FALSE)
typeof(df$Last.Name)
```

```
## [1] "character"
```

```
# readr
tbl = read_delim("DATA/students.txt",delim=" ")
```



```
##

Rows: 3 Columns: 3
## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;
##

## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

— Column specification —————

## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

Delimiter: " "
## chr (3): Last.Name, First.Name, Andrew.ID
##
## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;
##

##
## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

i Use `spec()` to retrieve the full column specification for this data.
## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## [36m[i [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;
```

```
typeof(tbl$Last.Name)
```

```
## [1] "character"
```

Question 6

Download `emline.csv` from the Canvas site. It is in the `DATA` directory. Use an external viewer (your choice) to look at the file. Then apply an appropriate function to read the file's contents into `R`. When you are done, show the mean and median values of the `sfr` column. Hint: if they are wildly different, you may need to adjust how you read in the data. Hint: look for numbers that represent missing data, and use an appropriate argument to tell `R` that those numbers should be converted to `NA`.

```
# Base R
df = read.csv("DATA/emline.csv", na.strings="-9999")
mean(df$sfr, na.rm=TRUE)
```

```
## [1] -0.71412
```

```
median(df$sfr, na.rm=TRUE)
```

```
## [1] -0.6436
```

```
# readr
tbl = read_csv("DATA/emline.csv", na=c("-9999"))
```

```
##

Rows: 9 Columns: 12
## [36m] [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;
##

## [36m] [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

— Column specification —————

## [36m] [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

Delimiter: ","
## dbl (12): O_II_3729, H_gamma, H_beta, O_III_4959, O_III_5007, N_II_6548, H_ALPHA, N_I
I_6584, S_II_6717, S_II_6731, mass, sfr
##

## [36m] [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;
##

##

## [36m] [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

i Use `spec()` to retrieve the full column specification for this data.
## [36m] [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;

i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## [36m] [39m Rendering ]8;;file:///Users/peterfreeman/Dropbox/36-600/FALL_22/NOTES/WEE
K_03/Lab_03T_Sol.Rmd [34mLab_03T_Sol.Rmd [39m]8;; into ]8;;file:///Users/peterfreeman/D
ropbox/36-600/FALL_22/NOTES/WEEK_03/Lab_03T_Sol.html [32mLab_03T_Sol.html [39m]8;;
```

```
mean(tbl$sfr, na.rm=TRUE)
```

```
## [1] -0.71412
```

```
median(tbl$sfr, na.rm=TRUE)
```

```
## [1] -0.6436
```

Question 7

Download `students.csv` from the Canvas site. It is in the `DATA` directory. Use an external viewer (your choice) to look at the file. Then apply an appropriate function to read the file's contents into `R`. Apply appropriate arguments or other function calls after the fact to ensure that the first three columns of the final data frame (or tibble) are character vectors, and the fourth column is a factor variable. (Simply print out the data frame or tibble; the variable type will be shown.) (Hint: if you use a `readr` function, then the argument `col_types` can be a single string with one letter for each column: "c" for a character vector, "d" for double, "i" for integer, "f" for factor, etc.). Hint: if using base `R`, it is simplest to start with `stringsAsFactors=FALSE` and then convert one column to a factor variable.

```
# Base R
df = read.csv("DATA/students.csv", stringsAsFactors=FALSE, skip=6)
df$Vegetarian = factor(df$Vegetarian)
df
```

```
##   Last.Name First.Name Andrew.ID Vegetarian
## 1   Baggins      Bilbo  bilbobag          N
## 2     Bar        Foo    fbar1          Y
## 3   Umlaut      Dieter  dumlaut          N
```

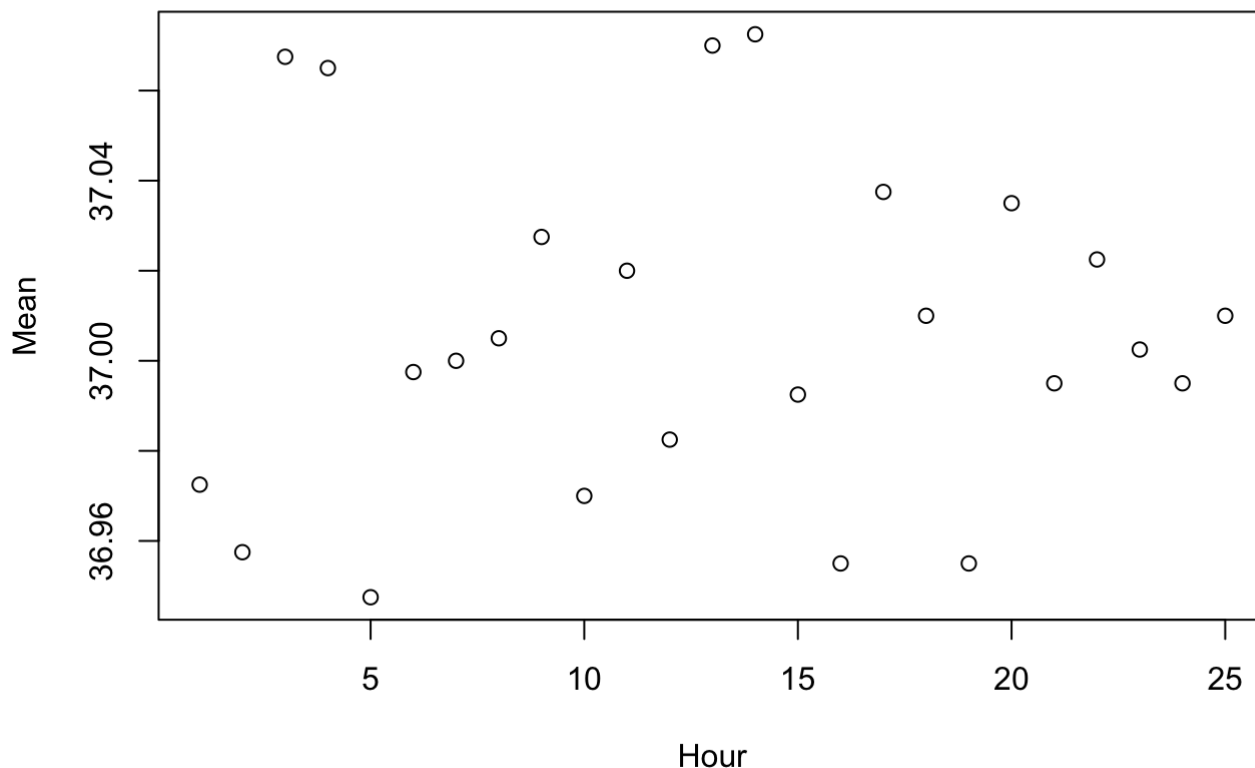
```
# readr
df = read_csv("DATA/students.csv", col_types="cccf", skip=6)
df
```

```
## # A tibble: 3 × 4
##   `Last Name` `First Name` `Andrew ID` Vegetarian
##   <chr>      <chr>      <chr>      <fct>
## 1 Baggins    Bilbo      bilbobag    N
## 2 Bar        Foo        fbar1       Y
## 3 Umlaut     Dieter     dumlaut     N
```

Question 8

Download `bolts.xls` from the Canvas site. It is in the `DATA` directory. If you cannot view the file with Excel, there are 100 rows and two columns: `Hour` and `Length`. Read the file into `R`, compute for average `Length` for each value of `Hour`, then plot that average versus `Hour`. (Remember: the variable before the "versus" is the y variable.) Note that if you pass a two-column data frame to `plot()`, it will know what to do. Hint: perhaps `group_by()` and `summarize()` here.

```
library(readxl)
tbl = read_excel("DATA/bolts.xls")
tbl %>% group_by(Hour) %>% summarize(Mean=mean(Length)) %>% plot()
```



Question 9

Download `data.xlsx` from the Canvas site. It is in the `DATA` directory. Also in the `DATA` directory is `data.png`; download and open it. It is a screen shot of a portion of the data that indicates possibly problematic data. Read the file into `R` while properly dealing with these problematic data. Note that data actually exists in the field(s) marked "`#####`"; you need not actually do anything about that marker. Display just the four columns in which problematic data existed in the screen shot (use `select()`); those data should be replaced with `NA`.

```
tbl = read_excel("DATA/data.xlsx", na=c("999", "999.00"))
tbl %>% select(Sex, RaceEthnicity, ADHD, Presentation)
```

```
## # A tibble: 3 × 4
##   Sex RaceEthnicity ADHD Presentation
##   <dbl>         <dbl> <dbl> <chr>
## 1     2             4     1 3
## 2    NA             NA    NA <NA>
## 3     2             1     0 4
```