Homework Week 1 V2

Steven Simonsen

2024-01-18

```
#Calculator commands
2+2 #addition
```

```
## [1] 4
```

```
3-5 #subtraction
```

```
## [1] -2
```

```
3*2 #multiplication
```

```
## [1] 6
```

```
4/3 #division
```

```
## [1] 1.333333
```

```
(2+2) ^ (3/3.5)
```

```
## [1] 3.281341
```

```
#Equality/inequality operations
12==(6/2) #test for equality
```

```
## [1] FALSE
```

```
(3*4) != (18-7) #test for inequality
```

```
## [1] TRUE
```

```
3 < 10 #less than
```

```
## [1] TRUE
```

```
(2*5) <= 10 #less than or equal
```

```
## [1] TRUE
```

```
#Logical values and logical operations
TRUE & FALSE #logical and
```

```
## [1] FALSE
```

```
TRUE | FALSE #logical or
```

```
## [1] TRUE
```

```
!TRUE #logical not
```

```
## [1] FALSE
```

```r
(TRUE | !TRUE) & !FALSE
```

```
## [1] TRUE
```

```r
#Variables and assignment
(12/3.5)^2 + (1/2.5)^3 + (1+2+3)^0.33
```

```
## [1] 13.6254
```

```r
x <- (12/3.5)^2 + (1/2.5)^3 + (1+2+3)^0.33
x
```

```
## [1] 13.6254
```

```r
x^2
```

```
## [1] 185.6516
```

```r
x * 3.6
```

```
## [1] 49.05145
```

```r
#Vectors
primes <- c(2, 3, 5, 7, 11, 13)
primes + 1
```

```
## [1]  3  4  6  8 12 14
```

```r
primes / 2
```

```
## [1] 1.0 1.5 2.5 3.5 5.5 6.5
```

```r
primes == 3
```

```
## [1] FALSE  TRUE FALSE FALSE FALSE FALSE
```

```r
primes == 7
```

```
## [1] FALSE FALSE FALSE  TRUE FALSE FALSE
```

```r
#Indexing Vectors
primes[1]
```

```
## [1] 2
```

```r
primes[5]
```

```
## [1] 11
```

```r
primes[c(3, 5, 2)]
```

```
## [1]  5 11  3
```

```r
primes[-1]
```

```
## [1]  3  5  7 11 13
```

```r
primes[-2]
```

```
## [1]  2  5  7 11 13
```

```r
#Vector types
nation <- c('ireland', 'england', 'scotland', 'wales')
nation[1]
```

```
## [1] "ireland"
```

```
nation[2:3]
```

```
## [1] "england"  "scotland"
```

```
nation == 'ireland'
```

```
## [1]  TRUE FALSE FALSE FALSE
```

```
class(primes)
```

```
## [1] "numeric"
```

```
class(nation)
```

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

```
class(nation == 'ireland')
```

```
## [1] "logical"
```

```
#Data Frames
Df <- data.frame(name = c('billy', 'joe', 'bob'),
                 age = c(21, 29, 23))
Df
```

```
##    name age
## 1 billy  21
## 2   joe  29
## 3   bob  23
```

```
#Indexing data frames
Df[3,2] #row 3, col 2
```

```
## [1] 23
```

```
Df[c(1, 3), 2] #rows 1 and 3, col 2
```

```
## [1] 21 23
```

```
Df[1,] #row 1, all cols
```

```
##    name age
## 1 billy  21
```

```
Df[, 2] #all rows, col 2
```

```
## [1] 21 29 23
```

```
Df$age
```

```
## [1] 21 29 23
```

```
Df[['age']]
```

```
## [1] 21 29 23
```

```
Df['age']
```

```
##   age
## 1  21
## 2  29
## 3  23
```

```r
#Functions
length(primes)
```

```
## [1] 6
```

```r
sum(primes)
```

```
## [1] 41
```

```r
mean(primes)
```

```
## [1] 6.833333
```

```r
median(primes)
```

```
## [1] 6
```

```r
sd(primes)
```

```
## [1] 4.400758
```

```r
var(primes)
```

```
## [1] 19.36667
```

```r
#Custom functions
my_mean <- function(x){sum(x)/length(x)}
my_mean(primes)
```

```
## [1] 6.833333
```

```r
#Writing R scripts and code comments
#Here is a data frame with four variables
#The variables are name, age, sex, and occupation
composites <- c(4, 6, 8, 9, 10, 12)
composites_plus_one <- composites + 1
composites_minus_one <- composites - 1
Df2 <- data.frame(name = c('jane', 'joe', 'billy'),
                  age = c(23, 27, 24),
                  sex = c('female', 'male', 'male'),
                  occupation = c('tinker', 'tailor', 'spy')
                  )

#Packages
#install.packages("dplyr")
#install.packages(c("dplyr", "tidyr", "ggplot2"))
#install.packages("tidyverse")
library("tidyverse")
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```r
#Reading in data
library(readr)
getwd()
```

```
## [1] "C:/Users/steve/OneDrive/Documents/School/DSE5001 Intro to Data Science and Stats/Week 1"
```

```r
test_data <- read_csv("weight.csv")
```

```
## Rows: 6068 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): gender
## dbl (7): subjectid, height, height_selfreport, weight, weight_selfreport, ag...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
test_data
```

```
## # A tibble: 6,068 x 8
##    subjectid gender height height_selfreport weight weight_selfreport   age
##        <dbl> <chr>   <dbl>             <dbl>  <dbl>             <dbl> <dbl>
##  1     10027 Male     178.              180.   81.5              81.7    41
##  2     10032 Male     170.              173.   72.6              72.6    35
##  3     10033 Male     174.              173.   92.9              93.0    42
##  4     10092 Male     166.              168.   79.4              79.4    31
##  5     10093 Male     191.              196.   94.6              96.6    21
##  6     10115 Male     172               175.   80.2              79.4    39
##  7     10117 Male     181               183.  116.              113.     32
##  8     10237 Male     185               188.   95.4              95.7    23
##  9     10242 Male     178.              178.   99.5              99.8    36
## 10     10244 Male     181.              183.   70.2              72.6    23
## # i 6,058 more rows
## # i 1 more variable: race <dbl>
```

```r
glimpse(test_data)
```

```
## Rows: 6,068
## Columns: 8
## $ subjectid         <dbl> 10027, 10032, 10033, 10092, 10093, 10115, 10117, 102~
## $ gender            <chr> "Male", "Male", "Male", "Male", "Male", "Male", "Mal~
## $ height            <dbl> 177.6, 170.2, 173.5, 165.5, 191.4, 172.0, 181.0, 185~
## $ height_selfreport <dbl> 180.34, 172.72, 172.72, 167.64, 195.58, 175.26, 182.~
## $ weight            <dbl> 81.5, 72.6, 92.9, 79.4, 94.6, 80.2, 116.2, 95.4, 99.~
## $ weight_selfreport <dbl> 81.66969, 72.59528, 93.01270, 79.40109, 96.64247, 79~
## $ age               <dbl> 41, 35, 42, 31, 21, 39, 32, 23, 36, 23, 32, 28, 36, ~
## $ race              <dbl> 1, 1, 2, 1, 2, 1, 2, 1, 1, 1, 1, 1, 2, 1, 1, 2, 1, 1~
```