Homework 1

Solutions

2026-01-08

Here's your chance to demonstrate that you can integrate the topics and skills you learned so far into a literate report.

When the question asks you to perform a coding task, insert a code chunk after each question where you will write the code to answer that question. Knitting after each completed code chunk will help you to ensure your final product works as intended! That way if it breaks, you know exactly where the error lies. It's like saving after every answer!

The first question is done for you as an example.

Object assignments

0. Calculate 3 + 4. Put the answer in the grey area below and knit the document. Make sure you can find this code and output in the resulting HTML file.

```
3+4
```

```
[1] 7
```

1. Calculate 2^5 by typing this mathematical expression in the code chunk below and then knit the document.

```
2^5
```

```
[1] 32
```

2. What are the values after each statement in the following?

```r
(mass <- 47.5)            # a
```

```
[1] 47.5
```

```r
(age  <- 122 )            # b
```

```
[1] 122
```

```r
(mass <- mass * 2.0)      # c
```

```
[1] 95
```

```r
(age  <- age - 20)        # d
```

```
[1] 102
```

```r
(mass_index <- mass/age)  # e
```

```
[1] 0.9313725
```

a. `mass` = 47.5
b. `age` = 122
c. `mass` = 95
d. `age` = 102
e. `mass_index` = 0.9313725

3. Assign a numeric value to the variable `my_apples`, assign a different numeric value to the variable `my_oranges`. Add these two together and assign the result to the variable `my_fruit`. Print the result of `my_fruit` to the report.

```r
my_apples <- 6
my_oranges <- 19
my_fruit <- my_apples + my_oranges
my_fruit
```

```
[1] 25
```

4. What is the data type of the variable `my_fruit`?

```r
class(my_fruit)
```

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

Its class is "numeric"

5. Render this document then remove this question once you have confirmed that the PDF was created successfully.

Vectors

1. Use the `class()` function to explore what happens when you have vectors of different data types. The first one has been done for you. Report your answer for each in a sentence below.

```r
num_only <- c(1,2,3,4)                    #a
char_only <- c("a","b","c","d")           #b
num_char <- c(1, 2, 3, "a")               #c
num_logical <- c(1, 2, 3, TRUE)           #d
char_logical <- c("a", "b", "c", TRUE)    #e
tricky <- c(1, 2, 3, "4")                 #f
class(num_only)
```

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

```r
class(char_only)
```

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

```r
class(num_char)
```

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

```r
class(num_logical)
```

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

```r
class(char_logical)
```

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

```r
class(tricky)
```

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

- a. The `num_only` vector contains only numeric values, and it's class is numeric.
- b. The `char_only` vector contains only character values, and it's class is character.
- c. The `num_char` vector contains numeric and character values, and it's class is character.
- d. The `num_logical` vector contains numeric and Boolean values, and it's class is numeric.
- e. The `char_logical` vector contains character and Boolean values, and it's class is character.
- f. The `tricky` vector contains three numeric value and one character value("4"), and it's class is character.

Let's go to Vegas!

2. Create three vectors (`weekday`, `poker`, `roulette`) using the `c()` operator to describe the following outcome.

- `weekday`: The 5 weekdays. Use these three letter codes: "mon", "tue", "wed", "thu", "fri."
- `poker`:
  - On Monday you won $125
  - Tuesday you lost $37
  - Wednesday you won $18
  - Thursday you lost $91
  - Friday you won $221
- `roulette`:
  - On Monday you lost $18
  - Tuesday you lost $45
  - Wednesday you won $136
  - Thursday you lost $321
  - Friday you won $9

Hint: If I won $30 on monday, lost $20 on tuesday and broke even every other day, my vector would look like `c(30, -20, 0, 0, 0)`

```r
weekday <- c("mon", "tue", "wed", "thu", "fri")
poker <- c(125, -37, 18, -91, 221)
roulette <- c(-18, -45, 136, -321, 9)
```

3. Use the `sum()` function to calculate your net winnings in each game by the end of the week.

```r
sum(poker)
```

```
[1] 236
```

```r
sum(roulette)
```

```
[1] -239
```

We gained `$230` on poker and lost `$239` on roulette.

4. Add the daily results of both vectors together, then calculate your net gain over the weekend. Did you come out ahead?

```r
sum(poker+roulette)
```

```
[1] -3
```

We came out behind.

5. On which days did you make money on poker? *Hint: subset the* `weekday` *vector using a logical statement about the* `poker` *vector*

```r
weekday[poker>0]
```

```
[1] "mon" "wed" "fri"
```

I made money on MWF.

6. On which days did you do better on poker than roulette? *Hint: subset the* `weekday` *vector using a logical comparison between the individual game vectors*

```
weekday[poker > roulette]
```

```
[1] "mon" "tue" "thu" "fri"
```

I did better in poker on every day except wednesday.

7. Render this document then remove this question once you have confirmed that the PDF was created successfully.

Data Frames

Run the code chunk below to read in the Ames data set from the web. This data set is on all residential home sales in Ames, Iowa between 2006 and 2010. The data set contains many explanatory variables on the quality and quantity of physical attributes of residential homes in Iowa sold between 2006 and 2010. Most of the variables describe information a typical home buyer would like to know about a property (square footage, number of bedrooms and bathrooms, size of lot, etc.). A detailed discussion of variables can be found in the original paper: De Cock D. 2011. Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project. Journal of Statistics Education; 19(3).

```
ames <- openintro::ames
```

1. Use the `names()` and `class()` functions to identify the data types for three different variables in the ames data set. List the variable names and data types in a bulleted list.

```
names(ames)
```

```
 [1] "Order"           "PID"             "area"            "price"
 [5] "MS.SubClass"     "MS.Zoning"       "Lot.Frontage"    "Lot.Area"
 [9] "Street"          "Alley"           "Lot.Shape"       "Land.Contour"
[13] "Utilities"       "Lot.Config"      "Land.Slope"      "Neighborhood"
[17] "Condition.1"     "Condition.2"     "Bldg.Type"       "House.Style"
[21] "Overall.Qual"    "Overall.Cond"    "Year.Built"      "Year.Remod.Add"
[25] "Roof.Style"      "Roof.Matl"       "Exterior.1st"    "Exterior.2nd"
[29] "Mas.Vnr.Type"    "Mas.Vnr.Area"    "Exter.Qual"      "Exter.Cond"
[33] "Foundation"      "Bsmt.Qual"       "Bsmt.Cond"       "Bsmt.Exposure"
[37] "BsmtFin.Type.1"  "BsmtFin.SF.1"    "BsmtFin.Type.2"  "BsmtFin.SF.2"
[41] "Bsmt.Unf.SF"     "Total.Bsmt.SF"   "Heating"         "Heating.QC"
[45] "Central.Air"     "Electrical"      "X1st.Flr.SF"     "X2nd.Flr.SF"
[49] "Low.Qual.Fin.SF" "Bsmt.Full.Bath"  "Bsmt.Half.Bath"  "Full.Bath"
```

```
[53] "Half.Bath"       "Bedroom.AbvGr"   "Kitchen.AbvGr"   "Kitchen.Qual"
[57] "TotRms.AbvGrd"   "Functional"      "Fireplaces"      "Fireplace.Qu"
[61] "Garage.Type"     "Garage.Yr.Blt"   "Garage.Finish"   "Garage.Cars"
[65] "Garage.Area"     "Garage.Qual"     "Garage.Cond"     "Paved.Drive"
[69] "Wood.Deck.SF"    "Open.Porch.SF"   "Enclosed.Porch"  "X3Ssn.Porch"
[73] "Screen.Porch"    "Pool.Area"       "Pool.QC"         "Fence"
[77] "Misc.Feature"    "Misc.Val"        "Mo.Sold"         "Yr.Sold"
[81] "Sale.Type"       "Sale.Condition"
```

```r
class(ames$MS.SubClass)
```

```
[1] "integer"
```

```r
class(ames$MS.Zoning)
```

```
[1] "factor"
```

```r
class(ames$Lot.Frontage)
```

```
[1] "integer"
```

- `MS.Subclass`, integer
- `MS.Zoning`, character
- `Lot.Frontage`, integer

2. How many observations does the ames data set have? How many variables? Explain where you got this information from.

```r
dim(ames)
```

```
[1] 2930   82
```

2930 observations and 82 variables. I found this using the `dim` function.

3. Extract the variable that measures the overall condition of the house (`Lot.Area`) by position (using bracket index notation `[]`), and by variable names (using `$` notation).

```r
# grading note - not run due to output length
ames[,8]
ames$Lot.Area
```

4. What is the maximum number of Full bathrooms (`Full.bath`) in this housing data set?

```
max(ames$Full.Bath)
```

```
[1] 4
```

5. Do any houses have more than 2 fireplaces (`Fireplaces`)? *Hint: Use the `summary()` function and examine the output to answer this question.*

```
summary(ames$Fireplaces)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.0000  0.0000  1.0000  0.5993  1.0000  4.0000 
```

Since the max is 4, that means there are houses that have more than 2 fireplaces.

6. What is the average sale price (`price`) for houses sold in 2010? (`Yr.Sold`). *Hint: Subset the data on the correct year first, then calculate the mean of the variable of interest. Be sure to match the data type of your logical statement with the data type of `Yr.Sold`. The year should be written as 2010, not "2010".*

```
mean(ames$price[ames$Yr.Sold==2010])
```

```
[1] 172597.6
```