Quiz: Intro to R and RStudio

Carey Kopeikin & Steph Neul

8/25/2021

The following quiz is *open note.* You may not give or receive help from any other students, teachers, or anyone else. You may not use the Internet except to read about the data set (link is below) and submit your work when you are finished. When you are finished please knit the file to pdf and upload it to Canvas. Every question should be answered in the R markdown file. *All answers should be in complete sentences OUTSIDE OF THE CODE CHUNK unless the questions only asks for a value to be printed by R.*

If you are stuck because of an error or other issue you may raise your hand. I will come by and see what the issue is and depending on the problem I may be able to give you some assistance but you may lose partial credit.

##Part A

1. Find the square root of 52.

```r
52^(1/2)
```

```
## [1] 7.211103
```

2. Use R to print every *even* integer between -11 and 91.

```r
seq(from = -11, to = 91, by = 2)
```

```
##  [1] -11  -9  -7  -5  -3  -1   1   3   5   7   9  11  13  15  17  19  21  23  25
## [20]  27  29  31  33  35  37  39  41  43  45  47  49  51  53  55  57  59  61  63
## [39]  65  67  69  71  73  75  77  79  81  83  85  87  89  91
```

3.Create a character vector of your three favorite colors and then call it so R prints it out.

```r
three.favorite.colors = c("Blue", "Black", "Pink")
three.favorite.colors
```

```
## [1] "Blue"  "Black" "Pink"
```

4. Use R to print out your vector from problem 3 sorted alphabetically A-Z. If you created the vector already in alphabetical order, print it out sorted Z-A.

```r
sort(three.favorite.colors, decreasing = FALSE)
```

```
## [1] "Black" "Blue"  "Pink"
```

Part B

1. Load up the data set `teacher.csv` and display the first 7 rows. You can find more information about the data set here: https://www.openintro.org/data/index.php?data=teacher *You may use the Internet here to read about the data set using the link provided!*

```r
data.set = read.csv("teacher.csv")
head(data.set, 7)
```

```
##   id degree fte years  base     fica retirement    total
## 1  1     BA   1   5.0 45388 3472.18    7688.73 56548.91
## 2  2     MA   1  15.0 60649 4639.65   10273.94 75562.59
## 3  3     MA   1  16.0 60649 4639.65   10273.94 75562.59
## 4  4     BA   1  10.0 54466 4166.65    9226.54 67859.19
## 5  5     BA   1  26.0 65360 5000.04   11071.98 81432.02
## 6  6     BA   1  28.5 65360 5000.04   11071.98 81432.02
## 7  7     BA   1  12.0 58097 4444.42    9841.63 72383.05
```

2. For the variables *id, degree, and years* write down what type of variable they are (quantitative or categorical), what class (in R) they *should* be, and *why*. If you think it could be more than one class, choose the class that you think BEST suits that variable, and explain why.

id:

Categorical (doesn't pass the average test)

ID should be character (or numerical, I just think character is better). ID has no units and you wouldn't average it (hence why it shouldn't really be numerical), yet each person has a unique ID, so it's not like a team that you belong to (henec why it shouldn't be factor).

degree:

Categorical (doesn't pass the average test)

Degree should be factor because there is a limited amount of options and because having a certain degree means the same thing across multiple people.

years:

Quantitative (passes the average test)

Years should be numerical as it has units (time) and can be averaged.

3. For each variable that is not the correct class (according to your answers above) change it to the correct class. Print the first 6 rows of the updated data frame to show that the variables' classes changed.

```
data.set$id = as.character(data.set$id)
class(data.set$id)
```

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

```
data.set$degree = as.factor(data.set$degree)
class(data.set$degree)
```

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

```
head(data.set, 6)
```

```
##   id degree fte years  base     fica retirement    total
## 1  1     BA   1   5.0 45388 3472.18    7688.73 56548.91
## 2  2     MA   1  15.0 60649 4639.65   10273.94 75562.59
## 3  3     MA   1  16.0 60649 4639.65   10273.94 75562.59
## 4  4     BA   1  10.0 54466 4166.65    9226.54 67859.19
## 5  5     BA   1  26.0 65360 5000.04   11071.98 81432.02
## 6  6     BA   1  28.5 65360 5000.04   11071.98 81432.02
```

4. What does the variable *fte* tell us *and* why might it be best as a logical variable?

fte tells us whether a teacher is employed full time or part time. It might be best as a logical variable because there are only two options (full or part), which can be represented as true or false.

5. Make *fte* a logical variable.

```
data.set$fte = data.set$fte == 1
data.set$fte
```

```
##  [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [25]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
## [37]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [49]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [61]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
```

6. How many observations and variables does this data frame have? How do you know, and show how you used R to find these values.

```
nrow(data.set)
```

```
## [1] 71
```

```
ncol(data.set)
```

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

7. Create a new vector in the data frame that tells us how much the teachers make in base pay and retirement combined and then print the first 6 rows of the data frame to show that this worked.

```
data.set$combined = data.set$base + data.set$retirement
head(data.set, 6)
```

```
##   id degree  fte years  base    fica retirement    total combined
## 1  1     BA TRUE   5.0 45388 3472.18    7688.73 56548.91 53076.73
## 2  2     MA TRUE  15.0 60649 4639.65   10273.94 75562.59 70922.94
## 3  3     MA TRUE  16.0 60649 4639.65   10273.94 75562.59 70922.94
## 4  4     BA TRUE  10.0 54466 4166.65    9226.54 67859.19 63692.54
## 5  5     BA TRUE  26.0 65360 5000.04   11071.98 81432.02 76431.98
## 6  6     BA TRUE  28.5 65360 5000.04   11071.98 81432.02 76431.98
```

8. Create a *new vector in the data frame* that tells us how much `total` compensation the teachers would get if they got a 5% raise next year and print the first 6 rows of the data frame. (Make sure you use the `total` variable for this calculation, you don't need to do anything fancy with the other columns to answer this question.)

```
data.set$raise = data.set$total * 1.05
head(data.set)
```

```
##   id degree  fte years  base    fica retirement    total combined    raise
## 1  1     BA TRUE   5.0 45388 3472.18    7688.73 56548.91 53076.73 59376.36
## 2  2     MA TRUE  15.0 60649 4639.65   10273.94 75562.59 70922.94 79340.72
## 3  3     MA TRUE  16.0 60649 4639.65   10273.94 75562.59 70922.94 79340.72
## 4  4     BA TRUE  10.0 54466 4166.65    9226.54 67859.19 63692.54 71252.15
## 5  5     BA TRUE  26.0 65360 5000.04   11071.98 81432.02 76431.98 85503.62
## 6  6     BA TRUE  28.5 65360 5000.04   11071.98 81432.02 76431.98 85503.62
```

9. Find the amount of `total` compensation that is earned by the third highest compensated teacher and the third lowest compensated teacher. For full credit you must use indexing and element extraction.

```
sort(data.set$total, decreasing = TRUE)[3]
```

```
## [1] 85007.76
```

```
sort(data.set$total, decreasing = FALSE)[3]
```

```
## [1] 44138.5
```