Khan_Ali_Yawar_hmk5.qmd

Answers:

Q1:

- Use `read_csv()` to read the file 'sample_data.csv' to an R data frame. Follow the instructions in Ch 7 to format it properly. Make sure the following are true:
  - Column names should be *syntactic*, meaning they don't contain spaces.
  - N/A values should be represented with the R value `NA`, not the character "N/A".
  - Data types (character vs factor vs numeric) should be appropriate.

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.6
v forcats   1.0.1     v stringr   1.6.0
v ggplot2   4.0.1     v tibble    3.3.1
v lubridate 1.9.4     v tidyr     1.3.2
v purrr     1.2.1
-- 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 beco
```

```
library(janitor)
```

```
Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```
df <- read_csv("../data/sample_data.csv",
               na = c("N/A"))
```

```
Rows: 6 Columns: 5
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (4): Full Name, favourite.food, mealPlan, AGE
dbl (1): Student ID

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.
```

```
df <- df %>%
  clean_names() %>%
  type_convert()
```

```
-- Column specification --------------------------------------------------------
cols(
  full_name = col_character(),
  favourite_food = col_character(),
  meal_plan = col_character(),
  age = col_character()
)
```

```
df <- df %>%
  mutate(across(where(is.character), as.factor))

glimpse(df)
```

```
Rows: 6
Columns: 5
$ student_id     <dbl> 1, 2, 3, 4, 5, 6
$ full_name      <fct> Sunil Huffmann, Barclay Lynn, Jayendra Lyne, Leon Rossi~
$ favourite_food <fct> Strawberry yoghurt, French fries, NA, Anchovies, Pizza,~
$ meal_plan      <fct> Lunch only, Lunch only, Breakfast and lunch, Lunch only~
$ age            <fct> 4, 5, 7, NA, five, 6
```

Q2

Next, use `read_csv()` to import the file colloquium_assessment.csv into an R dataframe. Note that you'll need to make use of some of the optional arguments. Use `?read_csv` to see what they are.

As we discussed in class, the correct shape depends on what you want to do with the data. Use `pivot_longer()` to make the data frame longer, in a way that makes sense.

```
library(tidyverse)

colloquium_raw <- read_csv("../data/colloquium_assessment.csv")
```

```
Rows: 36 Columns: 25
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (25): StartDate, EndDate, Status, Progress, Duration (in seconds), Finis...

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.
```

```
header_labels <- as.character(colloquium_raw[2, ])
names(colloquium_raw) <- paste(names(colloquium_raw), header_labels, sep = "_")


colloquium_clean <- colloquium_raw[-5, ]


colloquium_long <- colloquium_clean %>%

  filter(!is.na(names(colloquium_clean)[1])) %>%
  filter(!str_detect(.[[1]], "Start Date|ImportId")) %>%
  pivot_longer(
    cols = starts_with("Q"),
    names_to = "question_info",
    values_to = "response"
  )
head(colloquium_long)
```

```
# A tibble: 6 x 18
  `StartDate_Start Date` `EndDate_End Date` `Status_Response Type`
  <chr>                  <chr>              <chr>
```

```
1 11/11/22 9:01          11/11/22 9:02      0
2 11/11/22 9:01          11/11/22 9:02      0
3 11/11/22 9:01          11/11/22 9:02      0
4 11/11/22 9:01          11/11/22 9:02      0
5 11/11/22 9:01          11/11/22 9:02      0
6 11/11/22 9:01          11/11/22 9:02      0
# i 15 more variables: Progress_Progress <chr>,
#   `Duration (in seconds)_Duration (in seconds)` <chr>,
#   Finished_Finished <chr>, `RecordedDate_Recorded Date` <chr>,
#   `ResponseId_Response ID` <chr>,
#   `RecipientLastName_Recipient Last Name` <chr>,
#   `RecipientFirstName_Recipient First Name` <chr>,
#   `RecipientEmail_Recipient Email` <chr>, ...
```

Q3

Finally, calculate this student's average score for each of questions 7-10.

```r
library(tidyverse)


colloquium_raw <- read_csv("../data/colloquium_assessment.csv")
```

```
Rows: 36 Columns: 25
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (25): StartDate, EndDate, Status, Progress, Duration (in seconds), Finis...

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
descriptions <- as.character(colloquium_raw[2, ])


colnames(colloquium_raw) <- paste(colnames(colloquium_raw), descriptions, sep = "_")


colloquium_clean <- colloquium_raw %>%
  slice(-4) %>%
  filter(!is.na(`StartDate_Start Date`) & !str_detect(`StartDate_Start Date`, "Start Date"))
```

```r
average_scores <- colloquium_clean %>%

  pivot_longer(
    cols = starts_with(c("Q7", "Q8", "Q9", "Q10")),
    names_to = "Question",
    values_to = "Score"
  ) %>%

  mutate(Score = as.numeric(Score)) %>%

  group_by(Question) %>%
  summarise(Average_Score = mean(Score, na.rm = TRUE))


print(average_scores)
```

```
# A tibble: 4 x 2
  Question                                                      Average_Score
  <chr>                                                                 <dbl>
1 Q10_Limitations/alternatives. Were the limitations of the exper~       4.44
2 Q7_Justification/rationale. Was the presenter presenting convin~       4.5
3 Q8_Experimental design. Was experimental set-up well presented?~      4.62
4 Q9_Results. Were the results clearly presented? Were rigorous s~      4.31
```