collaborative_notebook

Keya Chinnam and Gabriella Bann

2024-02-21

Abstraction

1 Heading 1

1.1 Heading 2

1.1.1 Heading 3

hello

- List item
  - List item 2
    - Listen item 3

Best K-pop Groups:

1. BTS (3^{rd} generation)
2. Stray Kids (4^{th} generation)
   1. 3RACHA [^note1]
   2. (Malcolm, Polanco, and Barsuglia 2018)

click here for the [best choreography] (https://www.youtube.com/watch?v=FKJk2Lr7Sck)

[^note1]: Stray Kids subunit

```
#|echo: false

#loading packages and libraries
#quarto::quarto_use_template("wjschne/apaquarto", no_prompt = T)
#install.packages("apaquarto")
#library(apaquarto)
#quarto not running properly, commenting out for now
library(tidyverse)
```

```
Warning: package 'tidyverse' was built under R version 4.4.3
```

```
Warning: package 'ggplot2' was built under R version 4.4.3
```

```
Warning: package 'tidyr' was built under R version 4.4.3
```

```
Warning: package 'dplyr' was built under R version 4.4.3
```

```
Warning: package 'stringr' was built under R version 4.4.3
```

```
Warning: package 'forcats' was built under R version 4.4.3
```

```
-- 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   3.5.1     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
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 beco
```

```
library(dplyr)
library(ggplot2)
library(kableExtra)
```

```
Warning: package 'kableExtra' was built under R version 4.4.3
```

```
Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

    group_rows
```

```r
#setting seed
set.seed(123)

#loading in data
data <- read_csv("data.csv")
```

```
Rows: 9 Columns: 1
-- Column specification --------------------------------------------------------
Delimiter: ","
dbl (1): numbers

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
#sourcing R script
#if (file.exists("data_preprocessing.R")) {
  source("data_preprocessing.R")
#} else {
#  print("'data_preprocessing.R' not found. please make sure it exists in your working direct
```

```r
#assigned values to objects
x <- 10
y <- 20
z <- x + y

#constucted a data frame
df <- tibble(
  Name = c("Andres", "Blanca", "Carlos", "Dino", "Edgar"),
  Age = c(23, 19, 45, 33, 27),
  Score = c(90, 85, 99, 86, 94),
  Group = c("B", "A", "B", "A", "B"))

#defined functions
calculate_mean <- function(numbers) {
```

```
  mean(numbers)}

mean_score <- calculate_mean(df$Score)

average_age <- mean(df$Age, na.rm = TRUE)
```

2 Results

```
summary(df)
```

```
     Name                Age           Score         Group
 Length:5           Min.   :19.0   Min.   :85.0   Length:5
 Class :character   1st Qu.:23.0   1st Qu.:86.0   Class :character
 Mode  :character   Median :27.0   Median :90.0   Mode  :character
                    Mean   :29.4   Mean   :90.8
                    3rd Qu.:33.0   3rd Qu.:94.0
                    Max.   :45.0   Max.   :99.0
```

```
#if needed, filtered data frame
#df_filtered

#table of score staus if needed
#table(df$Status)

#summary statistics for numeric variables
summary_statistics <- df %>%
  summarise(
    Mean_Age = mean(Age),
    SD_Age = sd(Age),
    mean_score = mean(Score),
    SD_Score = sd(Score))

# numeric_summary %>%
#   kable("html", caption = "Summary Statistics for Numeric Variables") %>%
#   kable_styling(bootstrap_options = c("striped", "hover", "condensed"), full_width = FALSE)

#summary statistics for character variables #ERROR HERE
group_summary_statistics <- df %>%
```

```
  count(Group) %>%
  mutate(Proportion = n / sum(n))
```

```
c("\nSummary Statistics for Non-Numeric Variables:\n")
```

```
[1] "\nSummary Statistics for Non-Numeric Variables:\n"
```

```
print(group_summary_statistics)
```

```
# A tibble: 2 x 3
  Group     n Proportion
  <chr> <int>      <dbl>
1 A         2        0.4
2 B         3        0.6
```

```
#printing the average and mean age
c("The average age of participants in this data set is", average_age, " .\n")
```

```
[1] "The average age of participants in this data set is"
[2] "29.4"
[3] " .\n"
```

```
c(
  "The mean age of participants in this data set is", round(mean(df$Age), 2),
  "years (SD =", round(sd(df$Age), 2), ").\n",
  "The mean percent score participants recieved on the exam is", round(mean(df$Score), 2),
  "points (SD =", round(sd(df$Score), 2), ").\n")
```

```
 [1] "The mean age of participants in this data set is"
 [2] "29.4"
 [3] "years (SD ="
 [4] "10.14"
 [5] ").\n"
 [6] "The mean percent score participants recieved on the exam is"
 [7] "90.8"
 [8] "points (SD ="
 [9] "5.81"
[10] ").\n"
```

Table 1: Data Frame of Test Scores

Name	Age	Score	Group
Andres	23	90	B
Blanca	19	85	A
Carlos	45	99	B
Dino	33	86	A
Edgar	27	94	B

```
#printing the frequency in each course session
c(
  "The frequency of participants in Group A of the course session is", group_summary_statisti
  "(", round(group_summary_statistics$Proportion[group_summary_statistics$Group == "A"] * 100
  "The frequency of participants in Group B of the course session is", group_summary_statisti
  "(", round(group_summary_statistics$Proportion[group_summary_statistics$Group == "B"] * 100
)
```

```
 [1] "The frequency of participants in Group A of the course session is"
 [2] "2"
 [3] "("
 [4] "40"
 [5] "% of the sample).\n"
 [6] "The frequency of participants in Group B of the course session is"
 [7] "3"
 [8] "("
 [9] "60"
[10] "% of the sample).\n"
```

Table 1

```
`geom_smooth()` using formula = 'y ~ x'
```

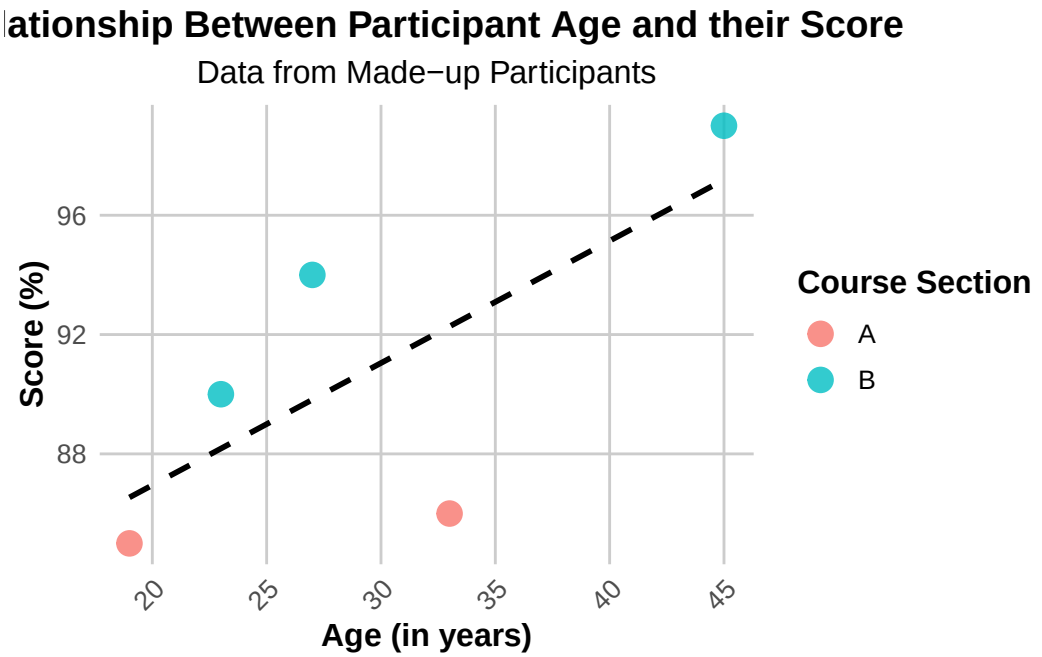

Figure 1: Figure 1

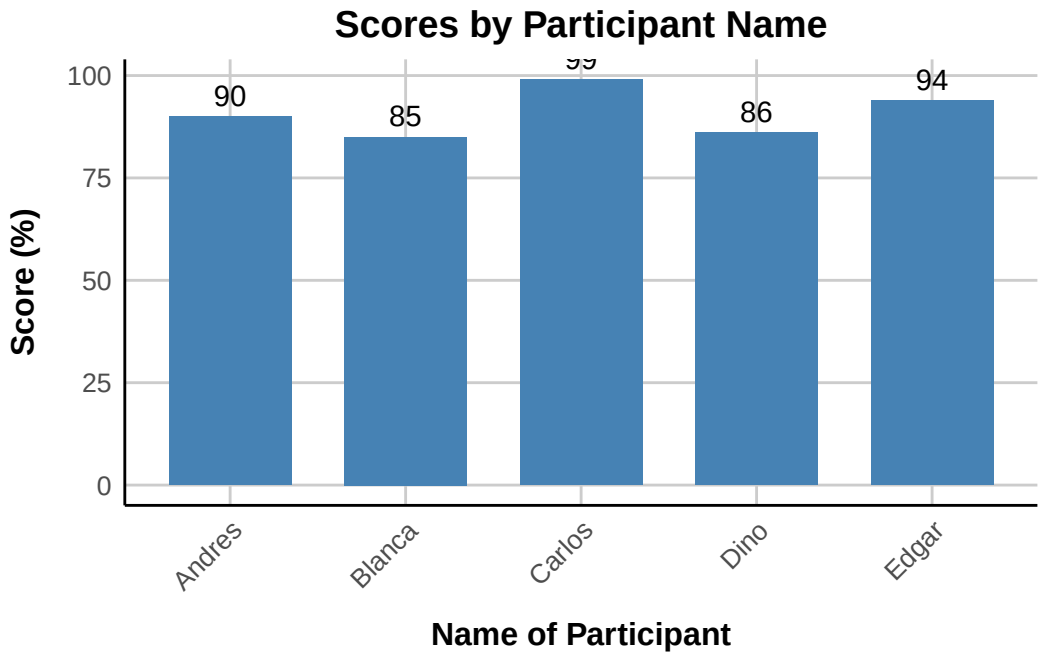

Figure 2: Figure 2

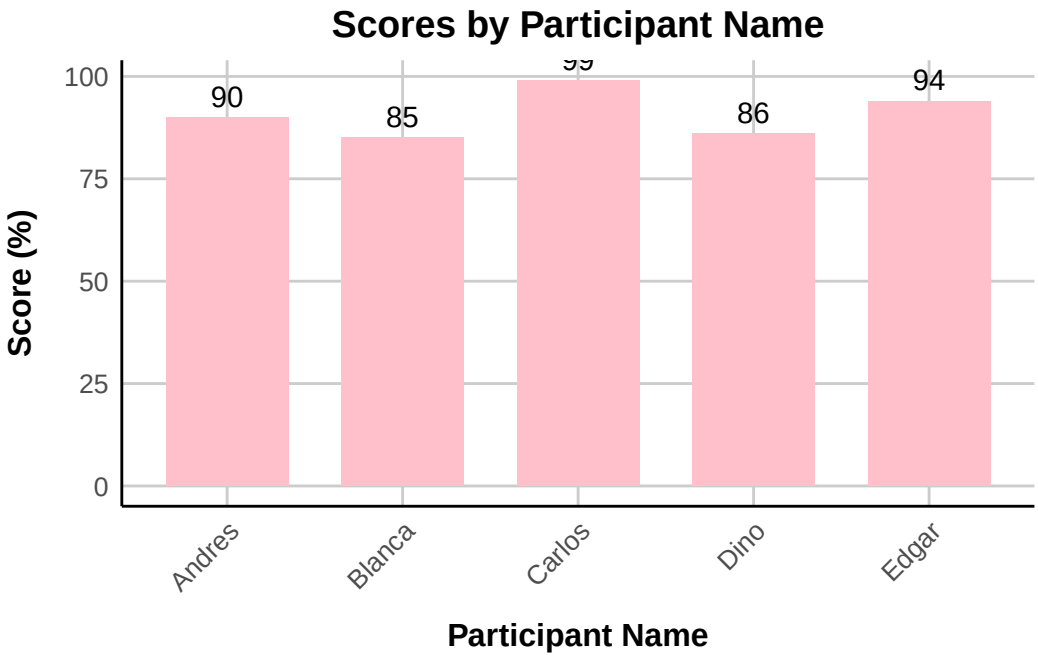

Figure 3: Figure 3

Figure 4: Image 1

The mean test score is 90.8 .

The highest test score in the data set of given participants is 99 .

The range of participant ages in the data set is from 19 to 45 years old. The median age in the data set is 27 years old. The youngest person in the data set who took the exam is Blanca, who is currently 19 years old.

The data set contains 5 observations, meaning that 5 individuals took the exam; with the scores ranging from `rmin(df$Score)` to `rmax(df$Score)`. The difference between the highest achieved score and the lowest is 14 percent. In regards to the course sections enrolled, there were 2 in session A and 3 in session B. In comparing the two groups, there was an average acore of 85.5 in session A while session B had an average score of 94.33 percent.

Figure 1 shows that there is a relationship between the participant's age and their test score. Further, the scores themselves are displayed by their respective participant name in **Figure 2**. More information on this data set is given in **Table 1**.

3 References

Malcolm, Benjamin J., Martin Polanco, and Joseph P. Barsuglia. 2018. "Changes in Withdrawal and Craving Scores in Participants Undergoing Opioid Detoxification Utilizing Ibogaine." *Journal of Psychoactive Drugs* 50 (3): 256–65. https://doi.org/10.1080/02791072.2018.1447175.