Age_and_gender_analysis

Kriss Tesink

2023-06-20

Introduction

In this file, the Mean, range, and standard deviation of ages for rounds 1 and 2 are computed. Additionally, the gender distribution for these two rounds are computed.

```r
# Set the file path of the CSV file
round1_first_half <- "prolific_export_dutch_first_half_anonym.csv"
round1_second_half <- "prolific_export_dutch_second_half_anonym.csv"
round2 <- "prolific_export_dutch_round2_anonym.csv"

# Import the CSV files
round1_first_half_data <- read_csv(round1_first_half)
```

```
## Rows: 30 Columns: 5
## -- Column specification --------------------------------------------------
## Delimiter: ","
## chr (4): Completion code, Gender, Highest education level completed, rand_id
## dbl (1): Age
##
## 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
round1_second_half_data <- read_csv(round1_second_half)
```

```
## Rows: 30 Columns: 5
## -- Column specification --------------------------------------------------
## Delimiter: ","
## chr (4): Completion code, Gender, Highest education level completed, rand_id
## dbl (1): Age
##
## 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
round2_data <- read_csv(round2)
```

```
## Rows: 30 Columns: 5
## -- Column specification --------------------------------------------------
## Delimiter: ","
## chr (4): Completion code, Gender, Highest education level completed, rand_id
## dbl (1): Age
```

```
##
## 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.

# Merge both half of round 1
round1_data <- rbind(round1_first_half_data, round1_second_half_data)

# Extract the age and gender
ages_round_1 <- round1_data[[4]]
genders_round_1 <- round1_data[[2]]

# Calculate the mean
mean_value_round_1 <- mean(ages_round_1)

# Calculate the standard deviation
sd_value_round_1 <- sd(ages_round_1)

# Extract the age and gender
ages_round_2 <- round2_data[[4]]
genders_round_2 <- round2_data[[2]]

# Calculate the mean
mean_value_round_2 <- mean(ages_round_2)

# Calculate the standard deviation
sd_value_round_2 <- sd(ages_round_2)


male1 <- 0
female1 <- 0
non_binary1 <- 0
##Count the number of men, women, and non-binary people
for(i in 1:length(genders_round_1)) {
  if(grepl("Man (", genders_round_1[i], fixed=TRUE)) {
    male1 <- male1 + 1
  } else if(grepl("Woman (", genders_round_1[i], fixed=TRUE)) {
    female1 <- female1 + 1
  } else {
    non_binary1 <- non_binary1 + 1
  }
}

male2 <- 0
female2 <- 0
non_binary2 <- 0
##Count the number of men, women, and non-binary people
for(i in 1:length(genders_round_2)) {
  if(grepl("Man (", genders_round_2[i], fixed=TRUE)) {
    male2 <- male2 + 1
  } else if(grepl("Woman (", genders_round_2[i], fixed=TRUE)) {
    female2 <- female2 + 1
  } else {
    non_binary2 <- non_binary2 + 1
  }
```

```
}
##Count the ratios of men, women, and non-binary people
male_ratio_round_1 <- male1/(male1+female1+non_binary1)
female_ratio_round_1 <- female1/(male1+female1+non_binary1)
non_binary_ratio_round1 <- non_binary1/(male1+female1+non_binary1)

male_ratio_round_2 <- male2/(male2+female2+non_binary2)
female_ratio_round_2 <- female2/(male2+female2+non_binary2)
non_binary_ratio_round_2 <- non_binary2/(male2+female2+non_binary2)


# Print the results
cat(" Age range round 1:", min(ages_round_1), "-", max(ages_round_1), "\n",
"Age mean round 1:", mean_value_round_1, "\n",
"Age standard deviation round 1:", sd_value_round_1, "\n",
"Male ratio round 1:", male_ratio_round_1, "\n",
"Female ratio round 1:", female_ratio_round_1, "\n",
"Non binary ratio round 1:", non_binary_ratio_round1, "\n",
"Age range round 2:", min(ages_round_2), "-", max(ages_round_2), "\n \n",
"Age mean round 2:", mean_value_round_2, "\n",
"Age standard deviation round 2:", sd_value_round_2, "\n",
"Male ratio round 2:", male_ratio_round_2, "\n",
"Female ratio round 2:", female_ratio_round_2, "\n",
"Non binary ratio round 2:", non_binary_ratio_round_2, "\n")
```

```
##  Age range round 1: 18 - 63
##  Age mean round 1: 30.78333
##  Age standard deviation round 1: 9.70251
##  Male ratio round 1: 0.5
##  Female ratio round 1: 0.5
##  Non binary ratio round 1: 0
##  Age range round 2: 20 - 74
##
##  Age mean round 2: 31
##  Age standard deviation round 2: 12.00862
##  Male ratio round 2: 0.4666667
##  Female ratio round 2: 0.5
##  Non binary ratio round 2: 0.03333333
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.