BIP Summer-Intro to R Workshop Activity

Joscelin Rocha-Hidalgo

2023-08-21

I am super excited to walk you through our activity report while I show you how you can use some functions and even create some basic visualizations!

How did I choose the questions?

You may be wondering why I asked you the questions you saw in the online survey. I am a developmental scientist and most of my current research is with children between the ages of 4 and 6 years. Some of them helped me come up with some of the "funner" questions such as the ones asking about superpowers. And of course, I also chose some questions that could give me some variability of responses and one that is very poorly stated so we can do some data cleaning on its content.

Let's load some packages

For the purposes of this workshop, most of our functions will come from packages part of the bigger package called *"Tidyverse"*.

```
#Load the `tidyverse` package.
library(tidyverse)

#Load the `janitor` package.
library(janitor)
```

Let's import our dataset

In theory, we could have imported the googlesheets directly to R (definitely possible!) but for simplicity and lack of time, I downloaded the responses as a csv and saved them to our repo under the "data" folder.

```
#Import the data.csv file into R and assign it to an object called "data"
#FYI: you could have called it anything!
data <- read_csv("data/data.csv")

#After you run it, check the environment panel. It should be saved there!
```

Let's clean & rename the variables

The variables exported from the googlesheet are too long and we need to `rename()` them to shorter ones.

```r
#let's clean the variables names using the package `janitor()

data <- data %>%
  janitor::clean_names()

## You can run this code to get the column names so it is easier to copy and paste
## them when changing them
#colnames(data)


data <- data %>%
  rename( "books_2023" = "how_many_books_have_you_read_so_far_this_2023",
          "enjoy_cooking" = "on_a_scale_of_1_to_10_how_much_do_you_enjoy_cooking",
          "hrs_sleep" = "how_many_hours_of_sleep_do_you_get_on_average_per_night",
          "superpower" = "which_superpower_would_you_choose",
          "chosen_weekend_activity" = "if_traveling_was_not_an_issue_you_can_teleport_anywhere_in_a_bli
          "least_fav_movie" = "whats_your_least_favorite_type_of_movie",
          "age_group" = "what_is_your_age_group",
          "gender" = "what_is_your_gender",
          "state" = "which_state_do_you_currently_reside_in",
          "race" = "what_is_your_racial_background",
          "latin_status" = "are_you_latino_a_e_x"  )
```

Let's separate the demo questions and the "fun" questions into 2 new dataframes

To separate the demo questions and the fun questions into 2 separate dfs, we will use the function `select()`.

1. Assign the fun questions to a df called "fun_data"
2. Assign the demo question to a df called "demo_data"

```r
# you can list them one by one
data %>%
  select(age_group,race)
```

```
## # A tibble: 43 x 2
##    age_group race
##    <chr>     <chr>
##  1 18-24     Black/African American
##  2 35-44     Black/African American
##  3 18-24     Prefer not to say
##  4 35-44     White/Caucasian, Black/African American
##  5 18-24     Black/African American
##  6 18-24     Prefer not to say
##  7 35-44     Black/African American
##  8 18-24     Prefer not to say
##  9 18-24     Black/African American
## 10 35-44     White/Caucasian, Black/African American
## # i 33 more rows
```

```r
#another way:
data %>%
  select(age_group:latin_status)
```

```
## # A tibble: 43 x 5
##    age_group gender            state      race                     latin_status
##    <chr>     <chr>             <chr>      <chr>                    <chr>
##  1 18-24     Female            PA         Black/African American   No
##  2 35-44     Prefer not to say California Black/African American   Yes
##  3 18-24     Female            Arizona    Prefer not to say        No
##  4 35-44     Female            TX         White/Caucasian, Black/A~ Yes
##  5 18-24     Male              Alabama    Black/African American   No
##  6 18-24     Male              PA         Prefer not to say        Yes
##  7 35-44     Prefer not to say TX         Black/African American   No
##  8 18-24     Male              California Prefer not to say        Yes
##  9 18-24     Female            NY         Black/African American   No
## 10 35-44     Prefer not to say NJ         White/Caucasian, Black/A~ No
## # i 33 more rows
```

```r
#another way
data %>%
  select(age_group:last_col())
```

```
## # A tibble: 43 x 5
##    age_group gender            state      race                     latin_status
##    <chr>     <chr>             <chr>      <chr>                    <chr>
##  1 18-24     Female            PA         Black/African American   No
##  2 35-44     Prefer not to say California Black/African American   Yes
##  3 18-24     Female            Arizona    Prefer not to say        No
##  4 35-44     Female            TX         White/Caucasian, Black/A~ Yes
##  5 18-24     Male              Alabama    Black/African American   No
##  6 18-24     Male              PA         Prefer not to say        Yes
##  7 35-44     Prefer not to say TX         Black/African American   No
##  8 18-24     Male              California Prefer not to say        Yes
##  9 18-24     Female            NY         Black/African American   No
## 10 35-44     Prefer not to say NJ         White/Caucasian, Black/A~ No
## # i 33 more rows
```