R for Psychology Research

Week 4 - Exercises

Marcus Lindskog

1. Importing and preparing data for analyses.

For all these exercises, please use packages and functions from the `tidyverse`.

1. Read the `iqitems.csv` file.
2. Read the first 10 lines from the `iqitems.csv` file.
3. Read the `iqitems.csv` file, but skip the first 10 lines.
4. Read the `iqitems_challenge.csv` file. This file has couple of challenges in it, so be sure to manage those to get a `tibble` that is the same as the one from `iqitems.csv`.
5. Assign the `tibble` from reading `iqitems.csv` to a variable name, e.g., `iqitems_data`.
6. Select the first three columns of the dataset using their column names.
7. Select all the columns of the dataset except "letter.7".
8. Select all columns of the dataset that start with the character string "matrix".
9. Filter the rows of the dataset for `rotate.3 >= 3` and `matrix.45 >= 4`.
10. Pipe the data frame to the function that will select two columns (`letter.58` and `matrix.55`).
11. Arrange rows by a particular column, such as the `rotate.3`.
12. Select three columns from the data set, arrange the rows by `rotate.4`, then arrange the rows by `letter.33`.
13. Create a new column called proportion, which is the ratio of `matrix.45` to `matrix.55`.
14. Compute the average of `reason.19`, apply the mean() function to the column `reason.19`, and call the summary value `avg_reason.19`.
15. Split the data frame by the values in `reason.4`, then ask for the same summary statistics as in 14.
16. The `mtcars` data set is available in your R session. Type `str(mtcars)` to get an overview of the it.
17. Use the gather function to gather cyl, and disp columns (from the mtcars data) into key-kalue pairs.
18. Use the function to spread a key-value pair across cyl, and disp.
19. Use the unite function to converge the cylinder column "cyl" and transmission "am" column in a single column called "cyl_am".
20. Use the seperate function to turn revert the operation in 19. If you feel like it, pipe the results from 19 to separate.

2. Examination Exercises

The solution to the exercises in this section should be handed in as a part of the examination. Your solution should be contained in a single R-script that is emailed to `marcus.lindskog@psyk.uu.se`. Your code should be well commented and easy to follow. Answers to any questions below should be written as a comment in the R-script after the code that produces the answer.

Use, as far as possible, a workflow with pipes and use functions from the `tidyverse`.

1. Read the `exercise_data.csv` file into a tibble. The file contains data from a fictional eye-tracking experiment. The eight variables in the file are ID - participant id, Sex - participant sex (0- boy, 1-girl), Age - age in months,trial - trial number, AOI_A - looking time to Area of Interest A, AOI_B - looking time to Area of Interest B;AOI_C - looking time to Area of Interest C, AOI_D - looking time to Area of Interest C.
2. We require participants to have looked at least .5 seconds in each of the four AOIs. Filter the data set based on this criterion.
3. Create two new variables called `prop_a_b`, which is the proportion of looking time in AOI_A to that of AOI_B, and `prop_c_d`, which is the proportion of looking time in AOI_C to that of AOI_D.
4. Remove the four AOI columns from the data set.
5. Create summaries (mean, standard deviation, count, standard error) of `prop_a_b` and `prop_c_d` for the boys and girls respectively. Note that you will need to creat a mean for each participant first.
6. Read the `exercise_data.csv` file into a new tibble.
7. Gather the four AOI_columns into a key-value pair with column names `AOI` and `looking_time`.
8. Create a new variable `z_looking_time` which is z-transformation of `looking_time`. Hint: the `scale` function in base R might be useful.
9. Filter the data set such that -2.5 $<$ z_looking_time $<$ 2.5.
10. Remove the `z_looking_time` variable.
11. Spread the key-value pair `AOI` and `looking_time` into four variables `AOI_A`, `AOI_B`, `AOI_C`, and `AOI_D`
12. Create two new variables called `prop_a_b`, which is the proportion of looking time in AOI_A to that of AOI_B, and `prop_c_d`, which is the proportion of looking time in AOI_C to that of AOI_D.
13. Create summaries (mean, standard deviation, count, standard error) of `prop_a_b` and `prop_c_d` for the three age groups respectively. Note that you will need to creat a mean for each participant first.