
Lecture 7: Dataframes in R

For each question below, write the R code you would use and record the result when appropriate. Answer in complete sentences when asked to explain.

Part 1: Creating Dataframes

1. Create the following vectors:

```
name <- c("Adam", "Brad", "Claire", "Donald", "Elaine", "Fiona")
age <- c(30, 49, 32, 28, 27, 33)
role <- c("attendee", "speaker", "org.", "attendee", "attendee", "org.")
registered <- c(TRUE, TRUE, TRUE, FALSE, FALSE, TRUE)
```

Then answer the following:

- (a) Combine the vectors into a dataframe called `conference_df`.
- (b) Display the dataframe.
- (c) In one sentence, explain what one row of this dataframe represents.

2. Create the following vectors:

```
city <- c("Albany", "Buffalo", "Syracuse", "Rochester", "Ithaca")
population <- c(97500, 255000, 142000, 210000, 32000)
capital <- c(TRUE, FALSE, FALSE, FALSE, FALSE)
```

Then answer the following:

- (a) Combine the vectors into a dataframe called `city_df`.

(b) Display the dataframe.

(c) What do the rows represent in this dataframe?

Part 2: Structure and Properties

3. Using `conference_df`, answer the following:

(a) Display the structure of the dataframe.

(b) Determine the dimension of the dataframe.

(c) Determine the number of rows and columns for the dataframe.

4. Using `city_df`, answer the following:

(a) Display the structure of the dataframe.

(b) Determine the dimension of the dataframe (two ways).

Part 3: Column Names and Index Selection

5. Using `conference_df`, answer the following:

- (a) Display the current column names.
- (b) Change the column names to
`"participant_name", "participant_age", "conference_role", "paid"`
- (c) Display the new column names.

6. Using the updated `conference_df`, answer the following:

- (a) Display just the `participant_name` column using the dollar sign.
- (b) Display the 4th observation.
- (c) Display the 3rd column.
- (d) Display the element in the 2nd row and 1st column.
- (e) Display rows 2 through 5 and columns 1 and 3.

7. Using the updated `conference_df`, answer the following with logical indexing:

- (a) Display all people who are speakers.
- (b) Display all people who are under the age of 35.
- (c) Display all people who are attendees and have paid.

Part 4: Editing a Dataframe

8. Using `conference_df`, add a new row for Gracie, who is 39, a speaker, and has not paid. Save the result as `conference_df2`.

- (a) Write the code to add the new row.
- (b) Display the new dataframe.
- (c) Look at the structure of the dataframe. Do any of the data types change?
- (d) What is the dimension of `conference_df2`?

9. Using `conference_df2`, answer the following:

(a) Permanently remove the `participant_age` column and save the resulting dataframe as `conference_df3`. This can be done 2 different ways.

(b) Remove the 1st observation from `conference_df3`.

(c) Display the resulting dataframe.

(d) What is the structure of `conference_df3` after these edits?

10. In one or two sentences, explain what the following line of code does:

```
conference_df3 <- conference_df3[-1, ]
```

Part 5: Reflection

11. In one or two sentences, explain why it is important to know the structure and dimensions of a dataframe before working with it.