

Analysing Data in R

Assignment 2.

The assignment should be done in an R script. Please create a new R script and do the exercises inside it, adding comments that explain what you are doing. Then upload the script to the Kampus platform. Use tidyverse verbs and syntax where possible (exceptions are: `unique()`, `table()`). Use the pipe in your code.

Exercise 1. Tidying data.

- 1) Prepare your script and working directory. Load **Ling_Van_data**. Note that it's an Excel file! Solve it using your imagination or R documentation (help).

The dataset is a subset of 4 speakers from this paper:

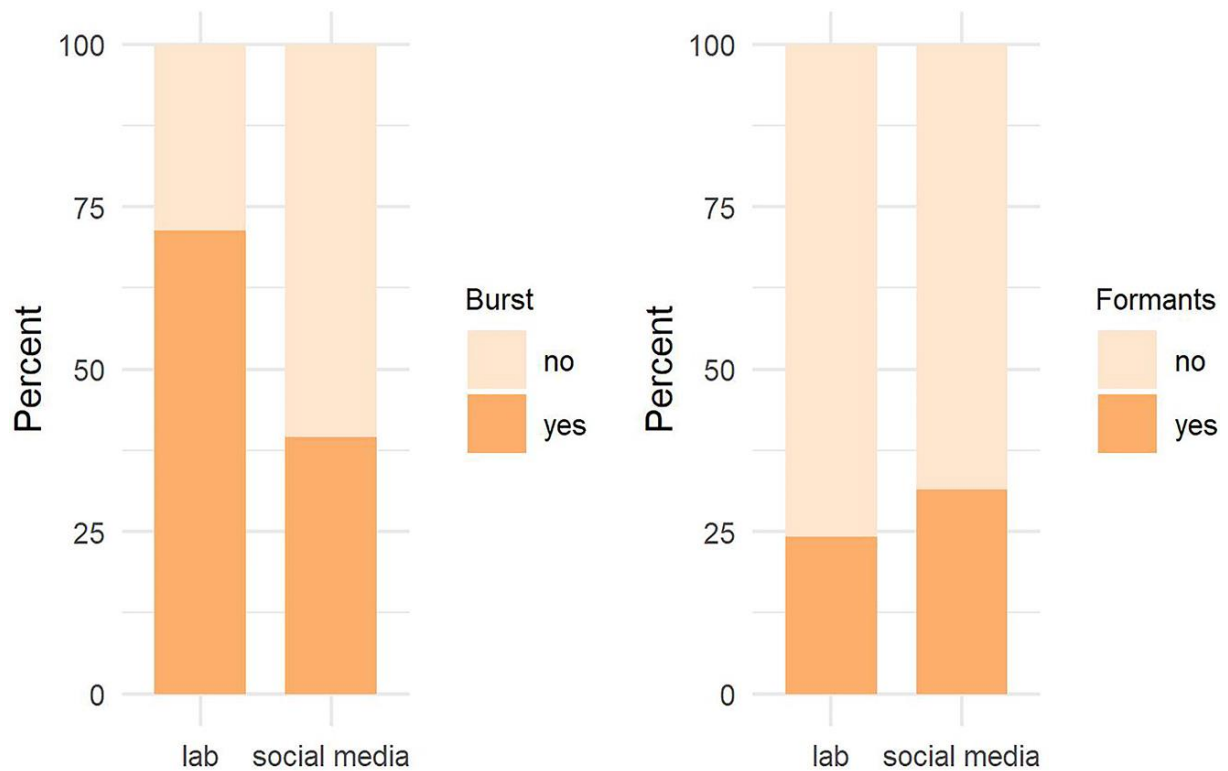
<https://www.degruyter.com/document/doi/10.1515/lingvan-2021-0153/html>

- 2) How many columns and how many rows does the database have?
- 3) What are the ages of the participants?
- 4) How many observations do we have per participant?
- 5) Select only "p" sounds and only in the laboratory setting, then provide the mean intensity difference for each participant from this subdatabase.
- 6) Select only the numeric columns and provide a summary of descriptive statistics for all of them.
- 7) Change column names to uppercase.
- 8) Create a new column called pronunciation based on 'annotated_sound'. In this new column, [p t k] should be named 'voiceless stops', [b d g] should be named 'voiced stops' and the remaining sounds should be named 'approximants'.
- 9) Create a new column called 'lenition' in which:
 - a. /p/ annotated as [p] has a value of '0' (no lenition)
 - b. /t/ annotated as [t] has a value of '0' (no lenition)
 - c. /k/ annotated as [k] has a value of '0' (no lenition)
 - d. /p/ annotated as [b] has a value of '1' (one change)
 - e. /p/ annotated as [d] has a value of '1' (one change)
 - f. /p/ annotated as [g] has a value of '1' (one change)
 - g. /p/ annotated as [B] has a value of '2' (two changes)
 - h. /p/ annotated as [D] has a value of '2' (two changes)
 - i. /p/ annotated as [G] has a value of '2' (two changes)

Hints: you may want to revise the difference between `summary()` and `summarise()`

Exercise 2. Plotting data.

- 1) Plot a histogram of intensity difference depending on each level of lenition. Embellish and name the plot. Save it to your computer.
- 2) Try to recreate this plot: (you can use different colours)



- 3) Create a scatterplot that shows whether intensity difference differs depending on sound duration (relative_C_dur). Add pronunciation as an additional variable marked by different colours. Embellish and save the plot.

Hint: get rid of the obvious outlier so that the plot makes sense.

- 4) Create a boxplot showing relative sound duration in seconds ('relative_C_dur') depending on the level of lenition. Does the degree of lenition influence the duration of the consonant? Embellish and save the plot.