QML - Formative Assessment 1

YOUR EXAM NUMBER

2023-09-08

1   Instructions

PLEASE READ CAREFULLY

DUE Week 5 - Thu 19 October at noon

You must include your **exam number as the author** in the document preamble above.

You'll notice that line 10 of the preamble says `mainfont: DejaVu Sans`. We would appreciate it if you'd use this font (or at least some other sans-serif font). **Try to render this Rmd file now, before making any changes, to see if you have this font installed.**

- If you get an error message saying that DejaVu Sans could not be found, you can download it here: http://sourceforge.net/projects/dejavu/files/dejavu/2.37/dejavu-fonts-ttf-2.37.zip.
- Then, install it appropriately for your operating system.
    - Here's a guide for Windows: https://www.digitaltrends.com/computing/how-to-install-fonts-in-windows-10/.
    - Here's a guide for Mac: https://support.apple.com/en-gb/HT201749.

Do each exercise by completing tasks, answering questions and/or providing code if required. Please **keep your written answers as concise as possible.**

Feel free to **add as many code chunks as you want** throughout.

When you are ready to submit:

1. Render the Rmd file to **PDF**.
2. **Rename** the PDF to your exam number only.
3. **Upload** the PDF file to Learn.

2 Exercises

2.1 Exercise 1: Creating plots

The next three exercises below require you to read in a particular file, filter/transform the data as needed, and create one or more plots that appropriately illustrate the described aspects of the data. Please also include a concise written description of the patterns you notice in each plot you make.

2.1.1 Exercise 1.1

Based on `data_e1_1.csv`: Plot speech rate against midpoint f0, colouring by condition and faceting by vowel. Describe what you see.

2.1.2 Exercise 1.2

Based on `data_e1_2.csv`: Create a single plot with logged reaction times by language, by environment, and by age (use any means). Describe what you see.

2.1.3 Exercise 1.3

Based on `data_e1_3.csv`: Plot the proportion of correct responses for each trial in the easy and difficult conditions, faceting by priming setting. Describe what you see.

## 2.2 Exercise 2: Critiquing and correcting plots

The following two plots are not appropriate for the type of data they show. Briefly describe what is wrong with each plot, try to figure out what the plots might be aiming to visualise, and write your own code to create a more appropriate plot for the exact same data. (If you're unsure about the kind of data you're dealing with, having a look at the data frame might help.)

2.2.1 Exercise 2.1

```
data_e2_1 <- read_csv("data/data_e2_1.csv")
```

```
## Rows: 600 Columns: 2
## -- Column specification ------------------------------------------------
## Delimiter: ","
## chr (2): reponse, condition
##
## 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.
```

```
data_e2_1 %>%
  ggplot(aes(reponse, condition)) +
  geom_point()
```
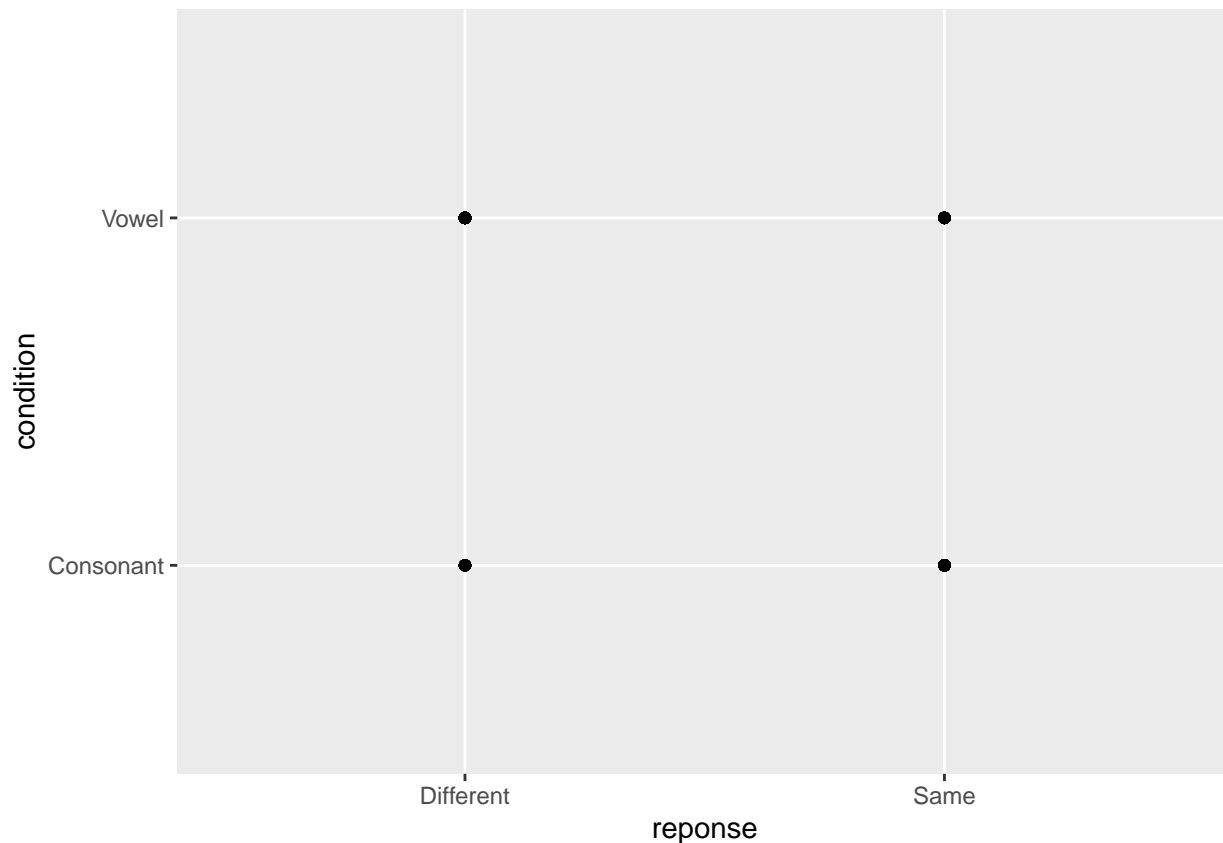

2.2.2 Exercise 2.2

```
data_e2_2 <- read_csv("data/data_e2_2.csv")
```

3

```
## Rows: 1035 Columns: 2
## -- Column specification ---------------------------------------------------
## Delimiter: ","
## chr (1): voicing
## dbl (1): vot
##
## 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.
```
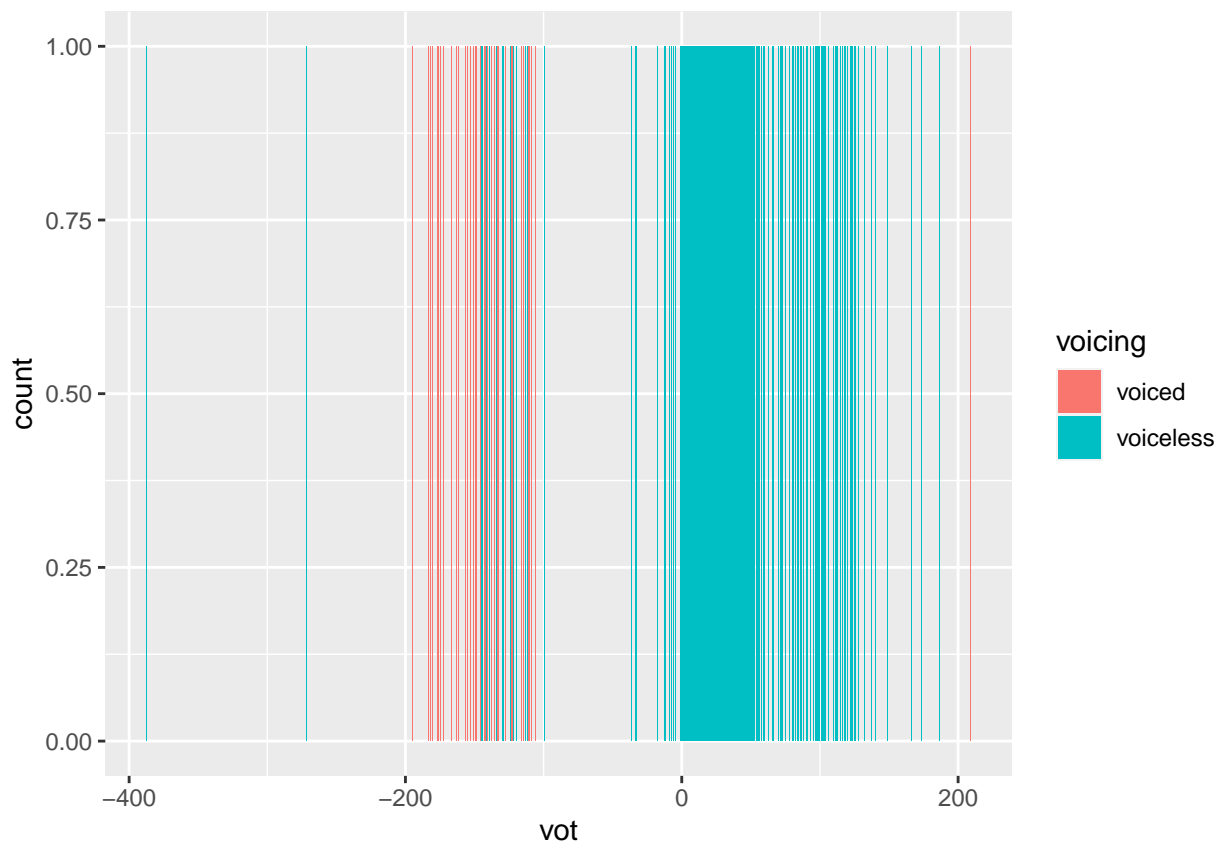
```r
data_e2_2 %>%
  ggplot(aes(vot, fill = voicing)) +
  geom_bar(width=0.5)
```

2.3 Exercise 3: Choosing appropriate summary measures

Read in `data_e3.csv` and obtain summary measures (central tendency: mean, median, or mode; dispersion: standard deviation or range) for each variable in the data. Make sure to pick the correct measure(s) for the respective variable type.

Then, for each variable, briefly state the type of variable, the chosen measure(s), and report the value of the measure(s).

2.4 Exercise 4: Identifying probability distributions

For each variable in the table below, specify in the "Probability distribution" column whether it's (in principle) distributed according to a Gaussian, a log-normal, or a Bernoulli distribution, or according to some different one (put "other" in this case).

If you have doubts about any of the variables, you can write about it briefly below the table.

	Variable	Probability distribution
1	Vowel duration (ms)	
2	Formant values (hz)	
3	Accuracy (binary)	
4	Readability (0-100)	
5	Reaction times (ms)	
6	Number of relative clauses	
7	Scots vs English	
8	Counts of infant gestures	
9	Logged reaction times	
10	Ratio of 1st vs non-1st pronouns	

2.5 Exercise 5: Running a Bayesian linear model