Lab 11: Linear mixed-effect models

1. Vowel reduction in Russian

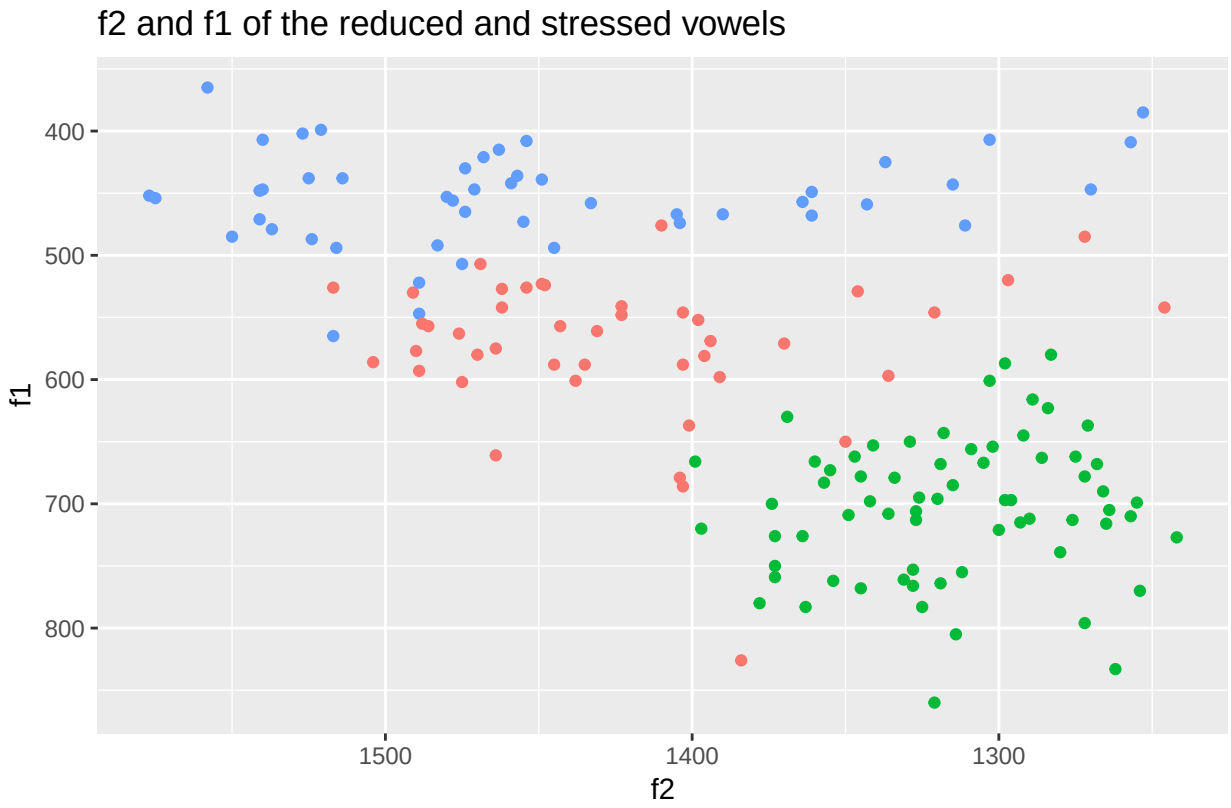
Pavel Duryagin ran an experiment on perception of vowel reduction in Russian language. The dataset `shva` includes the following variables:

- `time1` - reaction time 1
- `duration` - duration of the vowel in the stimuly (in milliseconds, ms)
- `time2` - reaction time 2
- `f1`, `f2`, `f3` - the 1st, 2nd and 3rd formant of the vowel measured in Hz (for a short introduction into formants, see here)
- `vowel` - vowel classified according the 3-fold classification (*A* - *a* under stress, *a* - *a/o* as in the first syllable before the stressed one, *y* (stands for shva) - *a/o* as in the second etc. syllable before the stressed one or after the stressed syllable, cf. *g*[*y*]*g*[*a*]*t*[*A*]*l*[*y*] *gogotala* 'guffawed').
  In this part, we will ask you to analyse correlation between f1, f2, and duration. The dataset is available https://raw.githubusercontent.com/LingData2019/LingData/master/data/duryagin_ReductionRussian.txt.

1.0 Read the data from file to the variable `shva`.

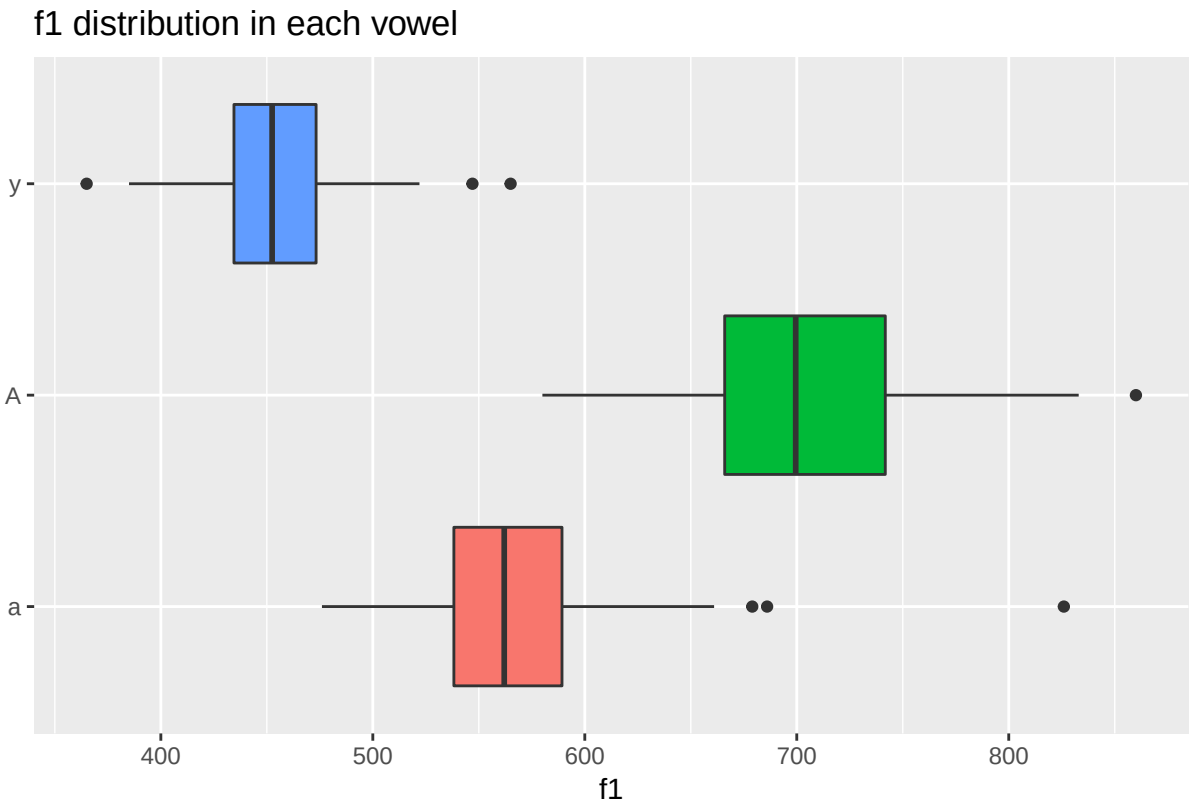
1.1 Scatterplot `f1` and `f2` using `ggplot()`.

Design it to look like the following:

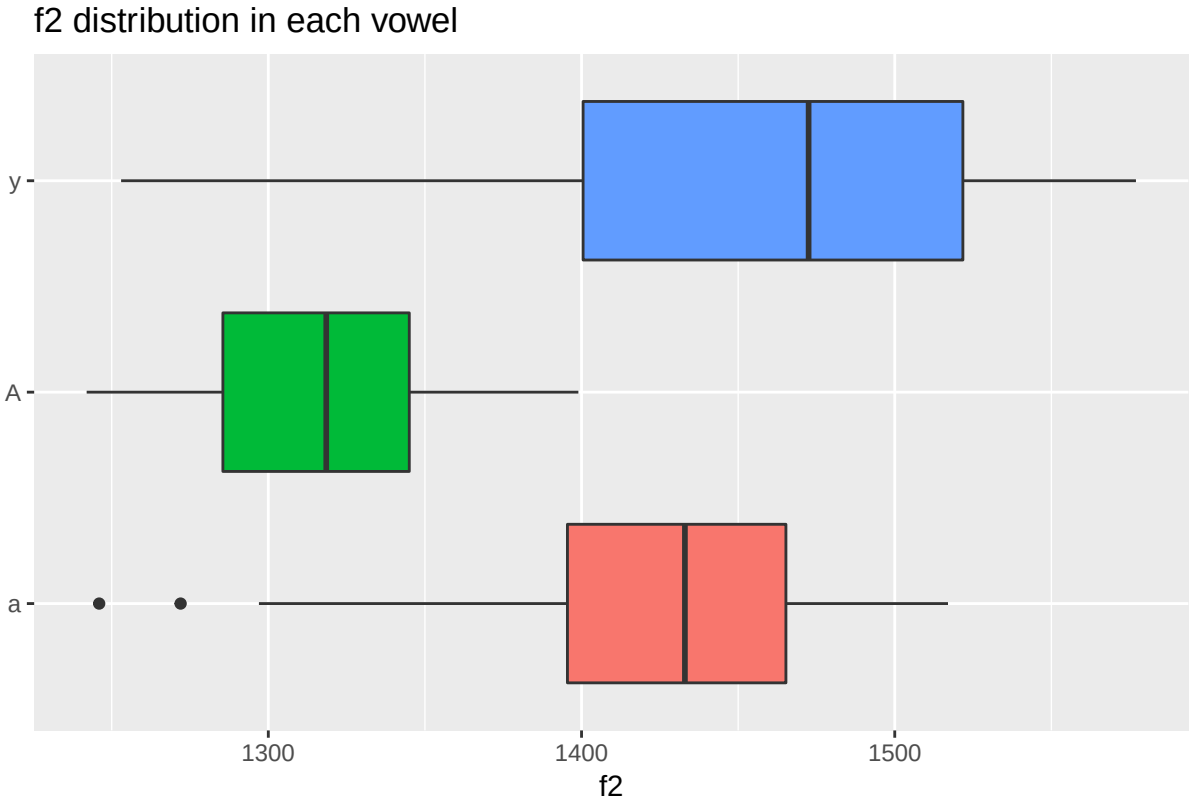


Data from Duryagin 2018

1.2 Plot the boxplots of `f1` and `f2` for each vowel using `ggplot()`.



Data from Duryagin 2018



Data from Duryagin 2018

1.3 Which `f1` can be considered outliers in *a* vowel?

We assume outliers to be those observations that lie outside 1.5 * IQR, where IQR, the 'Inter Quartile Range', is the difference between the 1st and the 3rd quartile (= 25% and 75% percentile).

1.4 Calculate Pearson's correlation of `f1` and `f2` (all data)

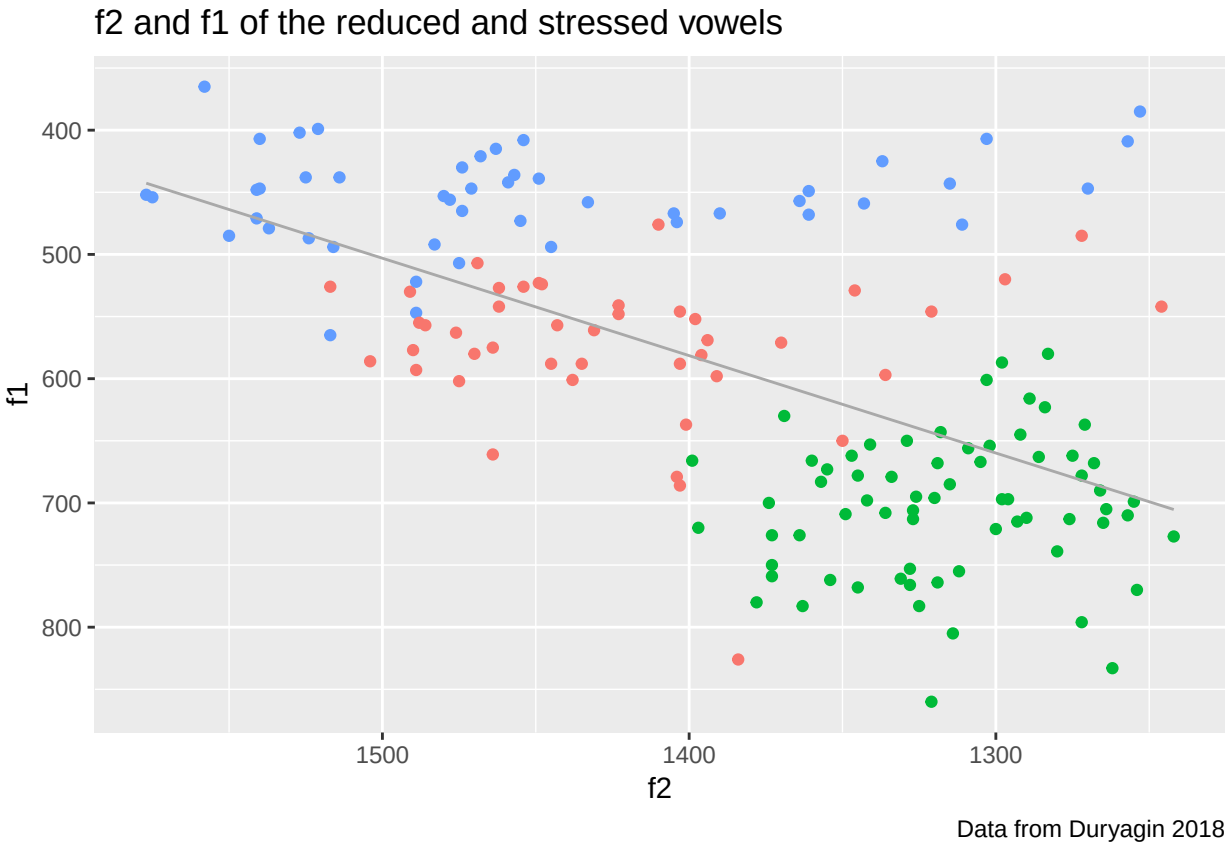
1.5 Calculate Pearson's correlation of `f1` and `f2` for each vowel

1.6 Use the linear regression model to predict `f2` by `f1`.

1.6.1 Provide the result regression formula

1.6.2 Provide the adjusted R^2

1.6.3 Add the regression line in scatterplot 1.1

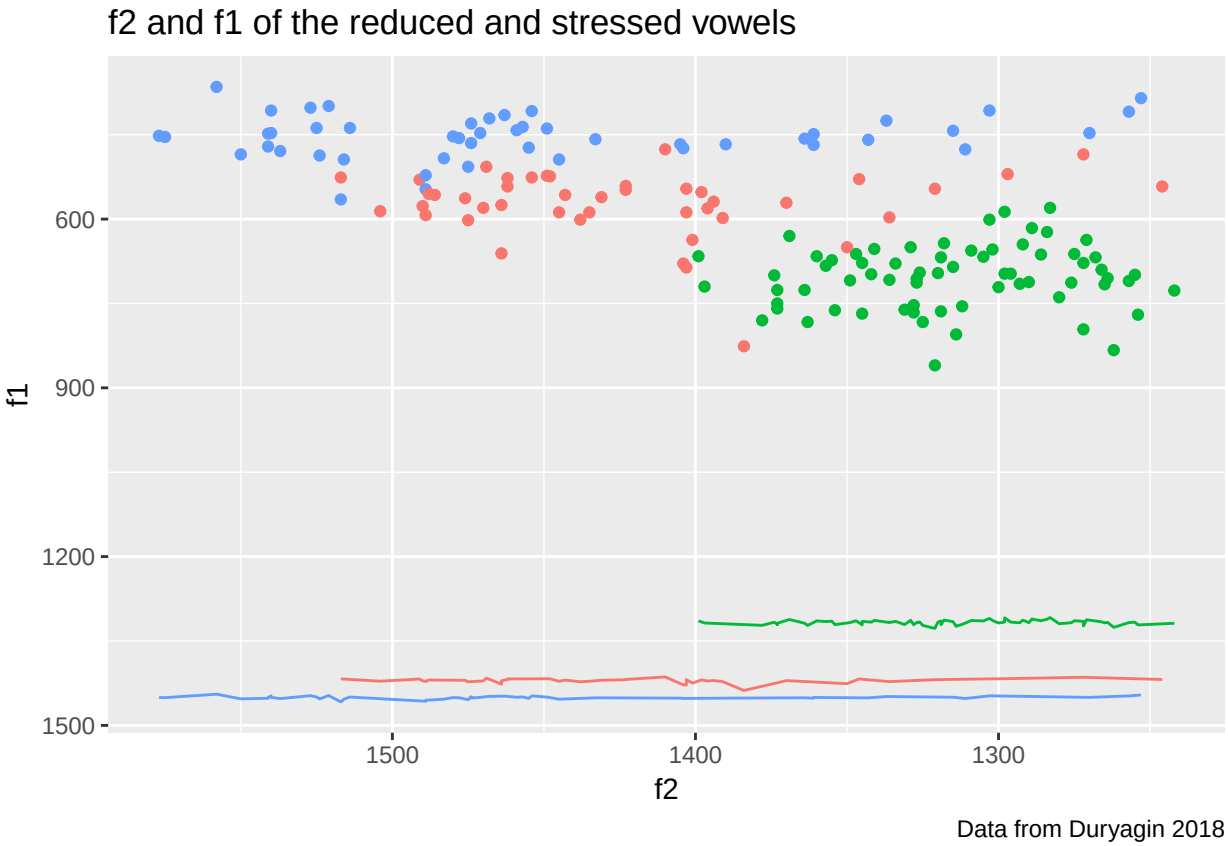


1.7 Use the mixed-efects model to predict `f2` by `f1` using `vowel` intercept as a random effect

1.7.1 Provide the fixed effects formula

1.7.2 Provide the variance for intercept argument for vowel random effects

1.7.3 Add the regression line in scatterplot 1.1



2. English Lexicon Project data

880 nouns, adjectives and verbs from the English Lexicon Project data (Balota et al. 2007).

- `Format` – A data frame with 880 observations on the following 5 variables.
- `Word` – a factor with lexical stimuli.
- `Length` – a numeric vector with word lengths.
- `SUBTLWF` – a numeric vector with frequencies in film subtitles.
- `POS` – a factor with levels JJ (adjective) NN (noun) VB (verb)
- `Mean_RT` – a numeric vector with mean reaction times in a lexical decision task

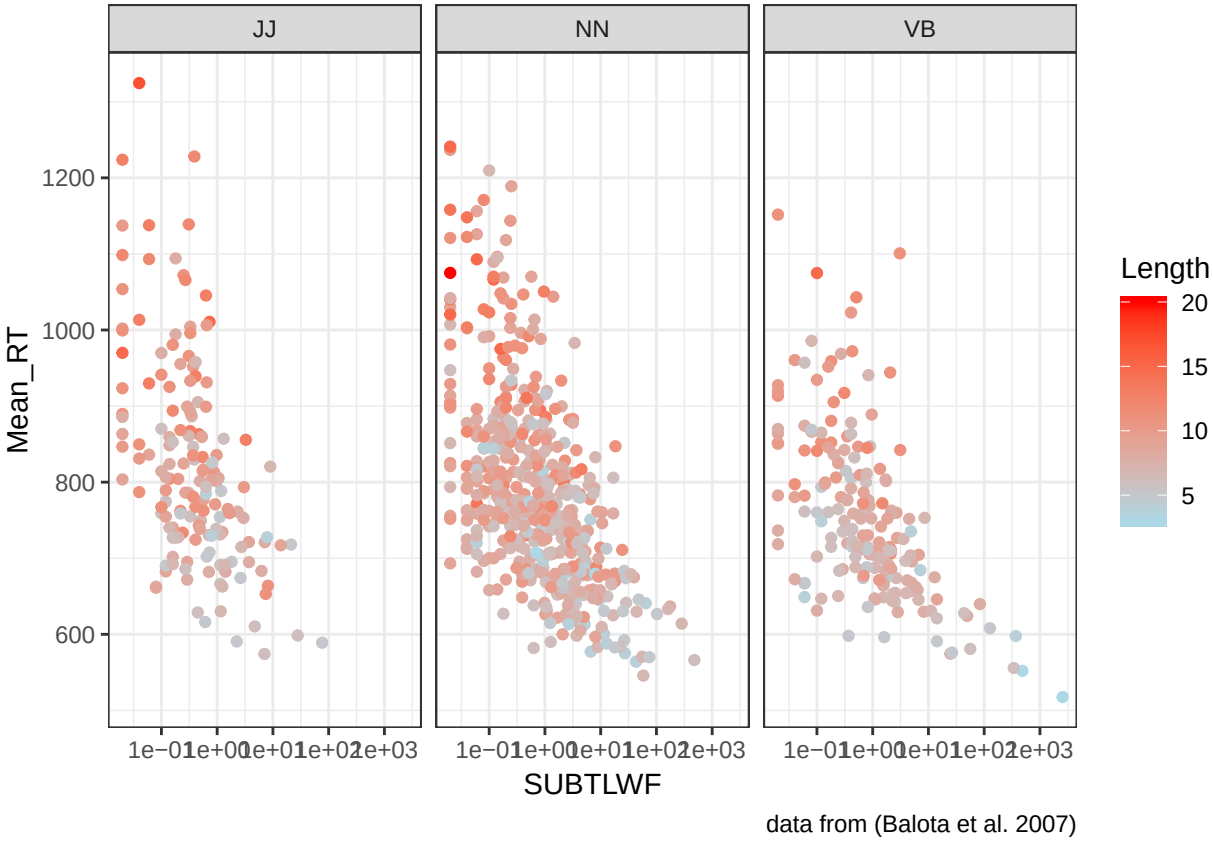
Source (http://elexicon.wustl.edu/WordStart.asp)

Data from Natalya Levshina's `RLing` package available (here)[https://raw.githubusercontent.com/LingData2019/LingData/master/data/Levshina_ELP.csv]

2.0 Read the data from file to the variable `elp`.

2.1 Which two variables have the highest Pearson's correlaton value.

2.2 Group your data by parts of speech and make a scatterplot of SUBTLWF and Mean_RT.



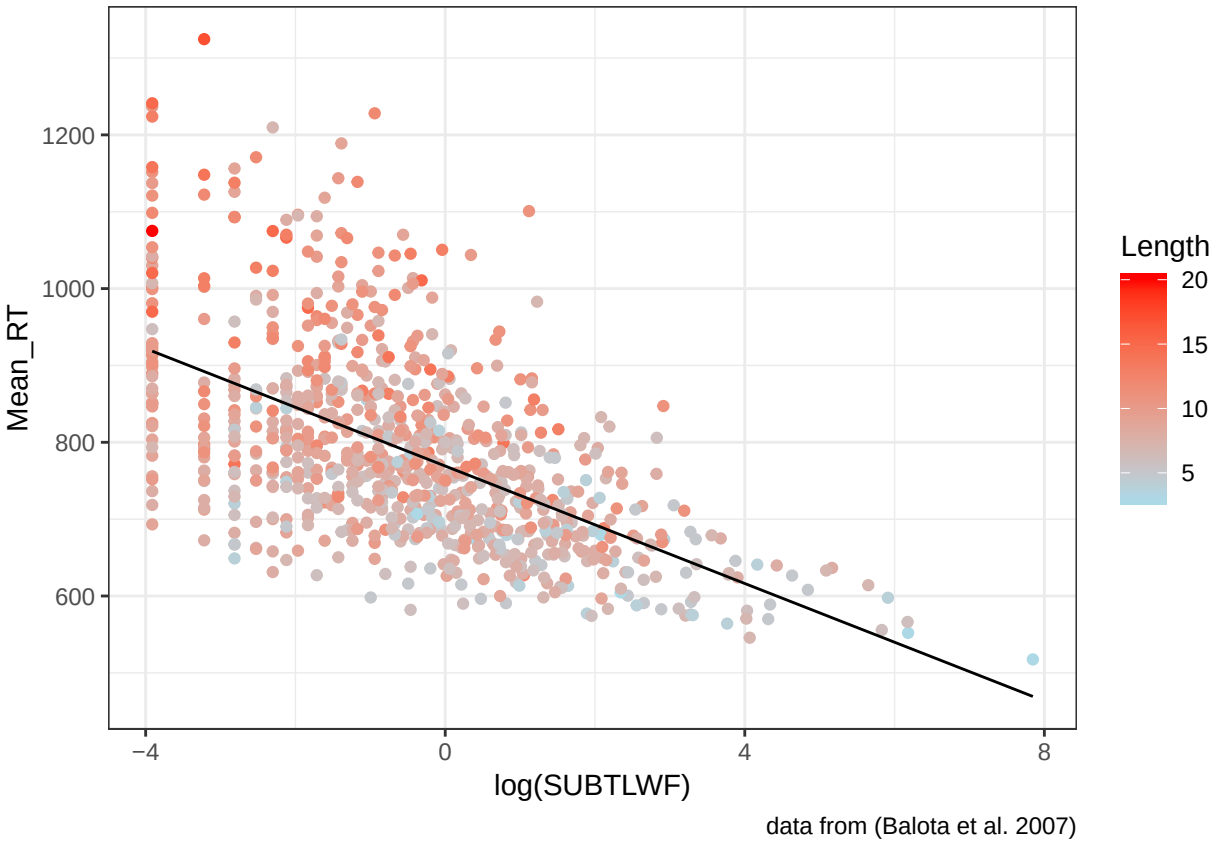
I've used `scale_color_continuous(low = "lightblue", high = "red")`

2.3 Use the linear regression model to predict Mean_RT by log(SUBTLWF) and POS.

2.3.1 Provide the result regression formula

2.3.2 Provide the adjusted R^2

2.3.3 Add the regression line in scatterplot 1.1

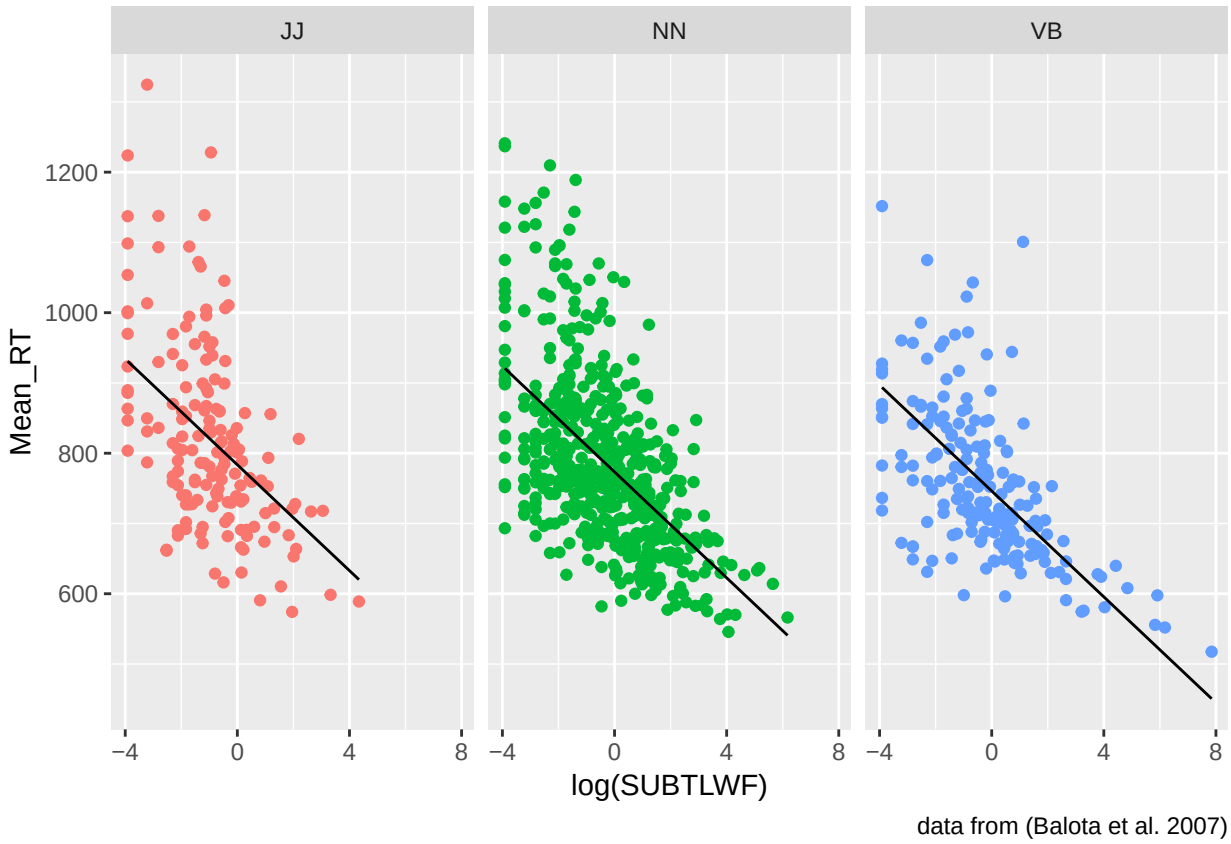


2.4 Use the mixed-efects model to predict `Mean_RT` by `log(SUBTLWF)` using POS intercept as a random effect

2.4.1 Provide the fixed effects formula

2.4.2 Provide the variance for intercept argument for POS random effects

2.4.3 Add the regression line to scatterplot



data from (Balota et al. 2007)

3.8 Why is it not recommended to run multiple Chisq tests of independence on different variables within your dataset whithout adjusting for the multiplicity? (i.e. just testing all the pairs of variables one by one)

3.9 Provide a short text (300 words) describing the hypothesis on this study and the results of your analysis.

—>