***Statistics for Humanities Researchers – day3***

**Read the following instructions with the Day3Exercise1.Rmd file open.**

**The exercises should have 2 outputs:**

* **This document with answers to the questions. Cut and paste any plot you’re asked to produce.**
* **The Rmd file with your code written in and “knitted”**

*Background:* Previous research conducted in the USA has shown that people with depression tend to speak differently from comparison groups (see Pennebaker, *The secret life of pronouns*). In particular, they tend to use fewer words, fewer positive words, more negative words and more past tense. A Danish student has decided to replicate these findings in the Danish population. The data you have come from his research.

*Exercise 1)* When you look at depression severity, is it normally distributed? Discuss whether the full dataset can be analyzed or whether you should only select a subset of the participants.

*Exercise 2)* Write a short paragraph reporting the results + plots per each of the dependent variables.

*Exercise 3)* Go to <http://guessthecorrelation.com/> and play a game. How much did you score?

*Exercise 4)* Go to <http://www.tylervigen.com/spurious-correlations> and have a bit of fun.

*Optional Exercise 5)* Explore other linguistic indexes

*Optional Exercise 6)* Add gender and age to the model