**Exploratory Data Analysis Assignment 2**

*You can do this assignment individually or in groups of two*

The attached survey (*sportrisks.docx*) and data (*sportsrisks.Rds*) investigates people’s perceptions of how risky or dangerous different sports are, and how “reckless” people are when they engage in these activities. People participating in the survey are shown a short written story or “scenario” describing a person who is engaging in some kind of sporting activity. The stories change slightly to vary the kind of sport the person is doing, whether the person doing the activity is female or male, has dependents or not, is doing the activity for charity or not, is experienced or not, etc.

The goals of the study are:

1. To investigate the relationships between three measures of risk: perceptions of danger, perceptions of recklessness, and whether someone should pay an insurance premium to cover their participation in the activity (and if so, how much). For example, is someone who does a dangerous activity always reckless, or are there some conditions where it is more acceptable to engage in a dangerous activity? If you are being more reckless, how much insurance should you pay? Is that amount the same regardless of what sport you do?
2. To investigate those factors that influence people’s judgments of danger and recklessness. Just for example, are some sports considered more risky than others, even when the probability of something bad happening is the same? Are people with dependents judged more reckless than people without dependents, even when they do the same activity?

The goal of this assignment is to expore the dataset recording the survey responses, and to use exploratory analyses and visualizations (like what you have covered in the course so far) to address the two goals above (you can include other goals if you like). These questions can and should of course eventually be answered by formal statistical modelling, but that is not part of the assignment.

Write up your work in the form of a short report (max 10 pages) in whatever word processing software you like (e.g. word, latex). The report should contain an introduction to the problem but the majority of the document should be about your results. Your code should be included, either embedded in the document if an Rmd file, or as a separate .R or .Rmd file. The key thing is that I need to be able to run your code and reproduce your results, so there should be clear instructions on how to use your script(s). The code itself should not be displayed in the final typeset document (use “echo = FALSE” for Rmd files) and not pasted as an appendix in your report.

Assignment hand-in is via the "Assignments" tab on Vula, and you should submit a single .zip file containing your report and all code. The submission deadline is on or before 23:59:59 on 3 March 2020.