Assignment 6

Due Thursday October 31 at 11:59pm on Quercus

The assignment is due on the date shown above. An assignment handed in after the deadline is late, and may or may not be accepted (see course outline). My solutions to the assignment questions will be available when everyone has handed in their assignment.

You are reminded that work handed in with your name on it must be *entirely your own work*.

Assignments are to be handed in on Quercus. See `https://www.utsc.utoronto.ca/~butler/c32/quercus1.nb.html` for instructions on handing in assignments in Quercus. Markers' comments and grades will be available there as well.

Begin with the usual:

```
library(tidyverse)
```

Hand in question 2 below.

1. Work through chapter 13 of PASIAS.

2. A cosmetic company created a small trial of a new cream for treating skin blemishes. It measured the effectiveness of the new cream (`New`) compared to the leading cream on the market (`Old`) and a placebo (`Control`). Thirty people were put into three groups of 10 at random, although just before the trial began 2 people from the control group and 1 person from the test group for the existing cream dropped out. The data as originally recorded are in `http://www.utsc.utoronto.ca/~butler/assgt_data/cosmetic-wide.txt`. The values in each column are the number of blemishes removed from each person during the trial (so a higher number is better).

   For this question, write a report, with a title, an Introduction, an Analysis and a Conclusion, that would make sense to someone who didn't know about this data set or what you were trying to do with it. Your report needs to read like something *you* wrote, not like some questions that I gave you. Parts (a), (b), etc, below, are to guide you about what to include in the report; *do not* include the part names in your report. (The grader will check that you have discussed the appropriate things somewhere in your report.) The grader also has 3 discretionary marks to award for the quality of writing, on the following scale: 3: excellent writing with few to no spelling or grammatical errors; 2: writing that is easy to read but contains grammatical errors; 1: writing that is hard to understand; 0: no attempt was made to turn this assignment into a report.

   (a) (4 marks) Begin your report with an Introduction that describes, *in your own words*, what the data set is, what is being measured, and what you hope to find out.

   (b) (2 marks) Begin your Analysis section. Read in and display the data. Why are there some `NA` values?

   (c) (2 marks) These data are not in a suitable layout for making plots or for running any kind of analysis of variance. Why not? Explain briefly.

(d) (3 marks) Use something from the `tidyverse` to get the data into the right format, and show that this is what you have done. Find out how to get rid of the missing values, so that your final data frame has 27 rows for the 27 non-missing observations.

(e) (3 marks) Make a suitable plot that will enable you to assess the distributions of values in each treatment group. What do you see?

(f) (4 marks) Carry out a suitable analysis to compare the number of blemishes among the treatment groups. Do an overall test, and follow up with comparisons between the groups if it is a sensible thing to do. If your comparisons between groups seem to be inconsistent with your overall test, discuss briefly what seems to have happened.

(g) (4 marks) End your report with a Conclusion that summarizes what you found out in the analysis, and how it answers your questions from the Introduction (or fails to do so). If you were able to do this study a second time, would you change anything? Why?