R Workshop Session 4 Exercise

Exercise 1:

Check out the following resources

a) R graph gallery (https://www.r-graph-gallery.com/). Which graph sparks the most joy for you?

b) Bioconductor course materials (https://bioconductor.org/help/course-materials/). Can you find anything of interest to you?

c) R-bloggers site (https://www.r-bloggers.com/). This site aggregates some of the best blogs about R. You can start by checking out the "most visited articles of the week" section on right hand side. You can also sign up and you will get a daily email with some of the best blogs. This covers a very wide range of topics so don't worry if 99% of them are irrelevant. We are looking for the needle in the haystack.

There is nothing for you to return back to us.

Exercise 2:

The data is a simulation of a longitudinal birth cohort. The two attached files contain the child length (i.e. height when the baby is laid down) and weight at two timepoints: delivery and month 3.

1) Download both files and eyeball scan for anything suspicious. Report anything that is odd.

2) Set the working directory, load the packages and read in both files.

3) Make the necessary changes (e.g. changing column names and units) to merge the two files.

4) Check for overlap in Subject IDs. Fix any errors in R (hint: `gsub()` might be useful) or Excel (in which case you need to re-read the data).

5) Merge the two files. Calculate the gain in weight and length.

6) Investigate the gain in length. Identify anything suspicious. You don't need to fix the suspicious data point here.

7) Investigate the gain in weight. Identify anything suspicious. You don't need to fix the suspicious data point here.

8) It is known in literature that babies who are born small for their gestational age tend to have a faster growth in the first few months. Can you able to test this with the available information?