Assignment 3. ST661 2019
Catherine Hurley
Due Monday November 11, 6pm

You should complete this assignment in Rmarkdown. You should knit the file to html and upload the html file to Moodle. The upload must be completed by time and date given above.

1. Type in

   ```
   library(ggplot2)
   library(MASS)
   head(Pima.tr)
   ```

   This code loads the data set `Pima.tr`, and its help file. Use ggplot2 for all plots. Using this dataset:

   (a) Make a scatterplot plot of `bp` versus `npreg`.

   (b) Using the function `cut_interval`, construct a factor version of `npreg` with n=4 levels. Call this new variable `npregf`. Add this variable to dataset `Pima.tr`.

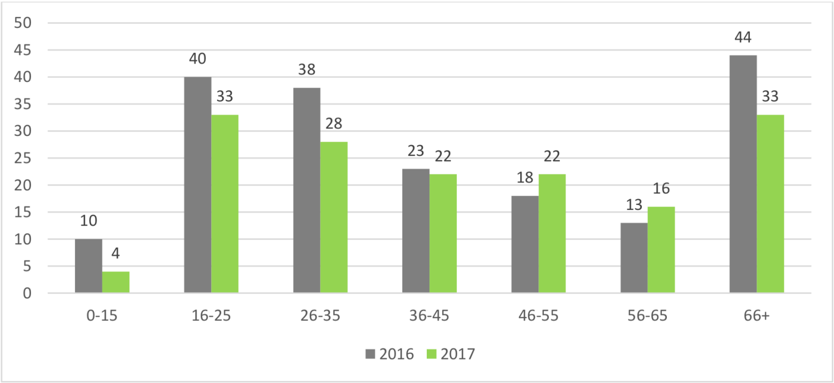
   (c) Plot boxplots of `bp` for each `npregf` level.

   (d) Make a scatterplot of `glu` versus `age`. Use colour to show variable `type` and add smooths for the two groups.

   (e) Redo the previous plot, separating out the two types. (Colour is not now needed)

2. (Optional) The Irish road safety authority recently produced a report with the graph below. Reconstruct it using ggplot. (Use google for help!)

   Figure 8. Deaths by age group, January to December 31st 2017 vs 2016



3. The data Hep2012.csv contains data with the points athletes received from each event in the Heptathlon in the 2012 Olympics. Access the data with

   ```
   hep <- read.csv("Hep2012.csv")
   ```

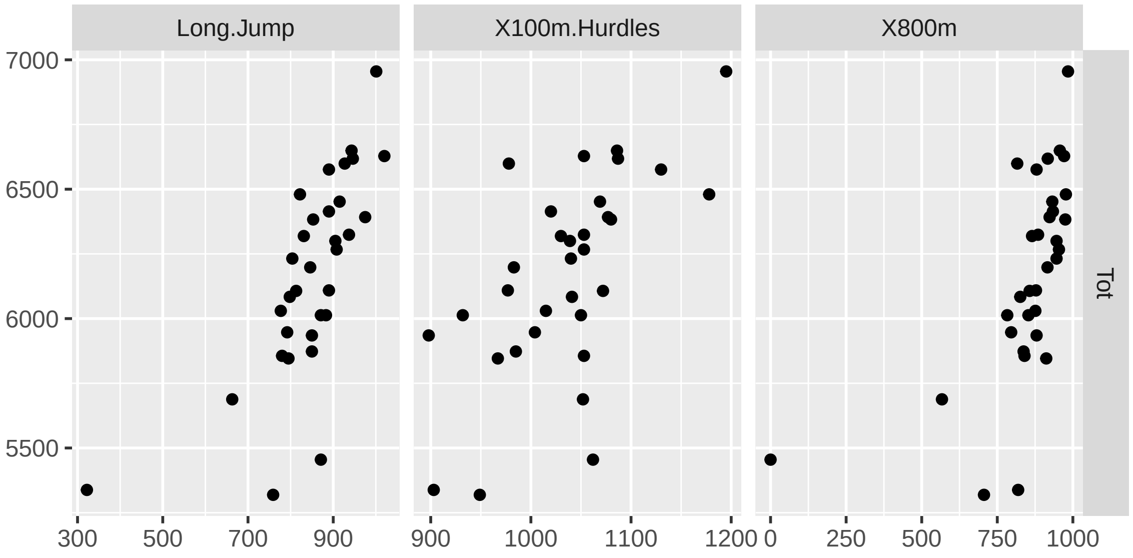
   (a) Use `na.omit` to leave out athletes with NAs as they did not complete the event.

   (b) Calculate the points total for the athletes and add it as an extra variable to the data frame.

   (c) Obtain a vector of the names of the athletes with the best five results overall. Your answer should use two lines of code. Hint: use `order`.

   (d) Which three of the 7 events scores are most highly correlated with the overall points total? Answer with one line of code. Hint: use `sapply` and `sort`.

   (e) Use ggduo in package GGally to plot points total versus the points in three events identified in part (d). Your result should look like


Long.Jump
X100m.Hurdles
X800m
Tot
7000
6500
6000
5500
300
500
700
900
900
1000
1100
1200
0
250
500
750
1000