[10 points] Instructions: You will have 100 minutes to complete this open book, open notes exam. Internet usage is strictly forbidden other than for Zoom and accessing the Canvas course page. Combine all solutions and upload, in Canvas, a single PDF or self-contained html file, starting each new problem on a new page, including all pertinent work and R code for full credit. Partial credit will only be given where work is shown. Your answers should be written sentences referencing R code and graphs as necessary. **R code and output by itself is not sufficient**.
For the online section, you are required to be on Zoom, **with your camera on**, for the duration of the exam. Good luck!

The file `midterm1data.csv` contains data representing $n = 104$ possums in Australia and New Guinea. Use the `read.csv()` function in R to read in the data set.

40 Pts

1. Building the model. In this problem we will focus on predicting `head_l`, the head length of the possum in mm, by using `skull_w`, the width of the skull in mm.
   (a) Create a scatterplot of the data, along with marginal histograms (seperate plots are ok), and comment on the validity of the usual assumptions based only on these plots.
   (b) Explore the use of higher order polynomial models and appropriate transformation(s) to correct for any problems you found in part (a); normality will be addressed later so you should not address that here. You should verify the adequacy of your new potential model visually with relevant plots. [Reminder: if you choose to use the poly() function, be sure to use the "raw=TRUE" argument]
   (c) Create a QQ plot of the residuals and comment on the appropriate assumptions.
   (d) Regardless of your conclusions in part (c), write out the full model to be estimated (the population model to be estimated).
   (e) Using the rules of thumb discussed in lecture, determine which points, if any, are considered high leverage points and/or outliers. Show the code along with relevant plot(s) and provide a discussion of your findings along with how you would address any issues you find.

50 Pts

2. Inference
   (a) Test to see if the variable(s) in your model are useful in explaining the response. Show all steps. Report and interpret the R^2 value.
   (b) Write out the estimated regression equation.
   (c) Interpret the estimated intercept, $\hat{\beta}_0$, in the context of the problem. Is this interpretation meaningful? Explain.
   (d) Give the point estimate for the head length of a possum whose skull width is 55mm (be sure to give your answer in the original scale of the data). Show the formula you use in your calculation. Verify your results in R by using the predict() function.
   (e) Use R to find a 97% prediction interval for the possum that has a skull width of 55mm. Interpret this interval as it relates to the population in terms of the problem.