ANALYSIS OF VARIANCE

There are four rules that apply to all projects so far:
   a) Follow instructions *precisely*. If I do not tell you what to write on a particular line, leave it blank.
   b) Do not use any functions or approaches to problems that we have not yet learned in this course.
   c) All code must be *scalable by sample size* unless specifically noted otherwise. This means your code should work equally well on a dataset with N=10 as N=1000.
   d) Any code using *magrittr* should contain a max of one verb per line.

Part 1 – Set up a new R Studio Project with one R script called week8.R

You can find a [description of the overall database we'll be working with here](#) and a link to download the [specific dataset we need here](#), the *National Longitudinal Study of Adolescent to Adult Health (Add Health), Wave IV, 2008*. Download it as a **tab-delimited file**. You do not need to give truthful information about yourself when asked. Also download [this codebook](#) and put it in an appropriate location.

Part 2 – Data Import and Cleaning

1. **Lines 1-3:** Write a comment that says: **R Studio API Code**, and set the wd as usual.
2. **Line 5-15:** Write a comment that says: **Libraries**, and import any libraries needed.
3. **Line 17-24:** Write a comment that says: **Data Import and Cleaning**, and using a single pipe, import the dataset as *health_tbl* such that: (a) it includes the *IMONTH4*, *BIO_SEX4*, *H4WP1*, and *H4LM29* variables, (b) rename those variables: *admin_month*, *gender*, *living_mother*, *fiw*, (c) change *gender* and *living_mother* to a more appropriate variable type, (d) add value labels to any factors, (e) change any non-scale values for *fiw* to NA [using **replace()**](#).

Part 3 – Visualization and Analysis

4. **Line 26-27:** Write a comment that says: **Visualization** and display density plots, boxplots, bar charts, and scatterplots, as appropriate to variable types, of all variables in *health_tbl.*
5. **Line 29:** Write a comment that says: **Analysis: FIW on Gender and Living Mother**
6. **Line 30+:** Create a model to test the hypothesis that *living_mother* and *gender* interact to predict *fiw,* controlling for *admin_month* **using a Type III sum of squares**, by completing the following tasks in this order:
      a. Create any needed models.
      b. Display your final model summary in ANOVA table format.
      c. Display model diagnostic plots plus RDI plots of the DV on the two IVs.
      d. Display a marginal means line plot of the interaction and
      e. Display all Tukey post-hoc tests, regardless of earlier results.
      f. Display partial effect size estimates for each component of the model.
      g. Create an ANOVA summary table in APA format that is saved to a file called "analysis.doc".
      h. Within your existing model, complete a planned contrast test, testing the hypothesis that those who don't know if their mother is alive or not have greater *wif* than those that do while controlling for *admin_month*.
      i. Write a multi-line comment stating whether and which of your two hypotheses were supported, including which statistics and p-values you used to conclude that.