

GENERAL AND GENERALIZED LINEAR MODELS

There are four rules that apply to all projects so far:

- Follow instructions *precisely*. If I do not tell you what to write on a particular line, leave it blank.
- Do not use any functions or approaches to problems that we have not yet learned in this course.
- All code must be *scalable by sample size* unless specifically noted otherwise.
- 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 week9.R

Same dataset as last time. You can still find a [description here](#) and a link to [download here](#) the *National Longitudinal Study of Adolescent to Adult Health (Add Health), Wave IV, 2008*. Download it as a **tab-delimited file**. Be sure the [codebook](#) is in an appropriate location.

Part 2 – Data Import and Cleaning

- Lines 1-3:** Write a comment that says: **R Studio API Code**, and set the wd as usual.
- Line 5-11:** Write a comment that says: **Libraries**, and import any libraries needed.
- Line 13-30:** 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 *H4MH3*, *H4MH4*, *H4MH5*, *H4MH6*, *H4CJ20*, and *BIO_SEX4* variables, (b) rename those variables: *mh1-mh4*, *jailage*, and *gender*, (c) change any non-scale values for *mh1-mh4* to NA, (d) reverse-code *mh1-mh4* as necessary to ensure more positive values represent better mental health, (e) calculate a mean score of *mh1-mh4* called *mh*, (f) change *gender* to a more appropriate variable type with appropriate labels, and (g) remove any cases where the participant has not been to jail or does not remember when they first went.

Part 3 – Visualization and Analysis

- Line 32-34:** Write a comment that says: **Visualization** and display density plots, boxplots, bar charts, and scatterplots, as appropriate to variable types, of the *jailage*, *gender* and *mh* variables.
- Line 34:** Display a pairwise correlation matrix of all numeric variables.
- Line 36:** Write a comment that says: **Analysis: Mental Health on Age of First Jailing and Gender**
- Line 37+:** Create a model to test the hypothesis that *jailage* and *gender* interact to predict *mh* by completing the following tasks in this order:
 - Create a linear model called *model1* containing main effects only.
 - Create a 2x2 panel of diagnostic plots (i.e., all plots should appear in one figure).
 - Write a comment summarizing your findings regarding linearity, homoscedasticity, normality, and influence.
 - Display model summary information, including R^2 and unstandardized coefficients.
 - Create a ggplot depicting the main effects model.
 - Create a new linear model called *model2* containing the full interactive model.
 - Display a 2x2 panel of diagnostic plots for this model, followed by a summary comment.
 - Display model summary information for this new model.
 - Create a ggplot depicting this new model.
 - Display ΔR^2 , F, and a p-value to compare the two models.
 - Write a multi-line comment stating which model you should retain and why.