EDA Homework 8

Due: Friday, March 31, 5pm.

The dataset `roaches`, which can be obtained either by installing the package `rstanarm` and then running `library(rstanarm); data(roaches)` or by downloading the file from http://jfukuyama.github.io/teaching/stat670/assignments/roaches.rda and running `load('roaches.rda')`.

The dataset has information about a roach treatment program. The columns are

- `y`: The number of roaches that were trapped.

- `roach1`: The number of roaches before treatment.

- `treatment`: Whether the residence was treated or not (0/1).

- `senior`: Whether the building was for seniors only (0/1).

- `exposure2`: How long the traps were set out for.

Notice that the number of roaches trapped should be some multiple of `exposure2`, and so when you fit a model you should use `log(exposure2)` as an offset.

- Fit a Poisson regression where the number of roaches trapped is a function of `roach1`, `treatment`, and `senior`. Make sure to include an offset for the `exposure2` variable.

- Using `expand.grid` to make a grid of predictor variables at which to get fited values and `augment` to get the fitted values at the grid of predictors, make a plot that shows the model you just fit. What does your plot show you about the effect of treatment? Of the effect of a seniors-only building?

- Make a residuals-vs.-fitted values plot (it might be easier to visualize if you plot the log-fitted values instead of the raw fitted values). Does the residual plot suggest that overdispersion is a problem? What would you do to solve this issue (no need to run a new model)?