# WU #5 - Residual Plots

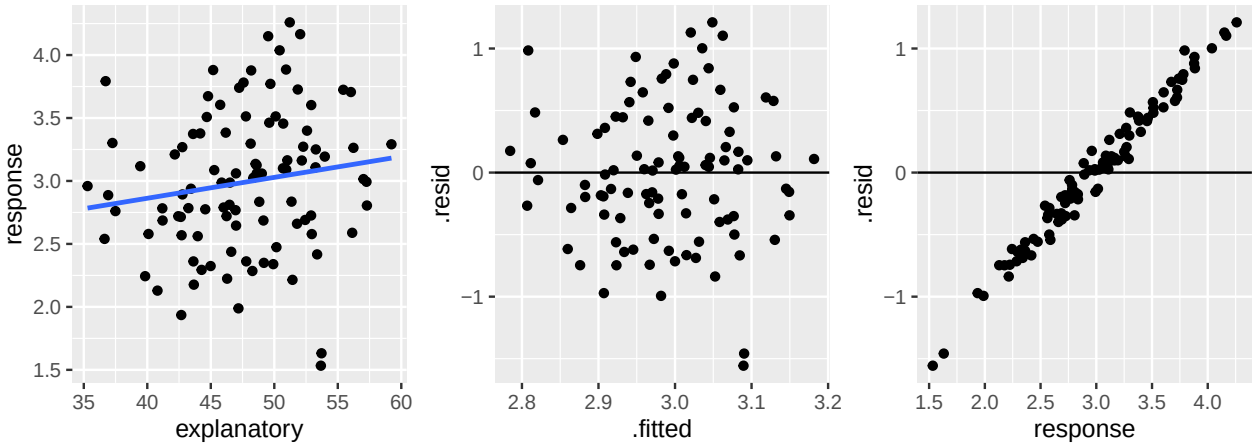
Math 158 - Jo Hardin

Thursday 2/3/2022

Name: _____

Names of people you worked with: _____

Consider the following made up data. The explanatory and response variable are not correlated. Pay attention to the labels on each axis.




1. Convince yourself that you understand the first two plots. The first one is response vs explanatory. The second is a standard residual plot, residual vs fitted value. You might be curious to know that due to linear algebra, the residual will always be uncorrelated with the fitted value under our model. [Nothing to report for #1.]
2. **Argue** for why the third plot is so strongly correlated. Your argument will go something like this:

- If we go from the center plot to the right plot, the variable that changes (shifts) is... (provide the formula for the change)
- If the residual itself is positive then the change (shift) ...
- If the residual itself is negative then the change (shift) ...
- ...

**Solution:**

We know that $e_i$ and $\hat{y}_i$ are uncorrelated (this can be shown using linear algebra, also note that $\sum e_i \hat{y}_i = 0$). So, if we go from the resid vs. fitted scatterplot to the resid vs. observed scatter plot, we shift each point in the x-direction (only) by an amount equal to the residual. If the residual is negative, then the point will shift to the left. If the residual is positive, then the point will shift to the right. We thus create a positively correlated relationship (between resid and observed). The degree of the shift will depend on the relative magnitudes of the residuals and predicted values.

$\Rightarrow e_i$ and $y_i$ are correlated and therefore **not** independent. Therefore, we use the fitted value instead of the response variable to check residuals.