Consider the Old Faithful Guyser (Yellowstone National Park) dataset in R.

> ?faithful

Perform a simple linear regression to determine if the expected waiting time to the next eruption can be modeled by the length of the previous eruption.

The estimated regression equation is given by

$\hat{Y}_i =$ 33.47 + 10.73 X_i

Even though both the slope and intercept terms are significant, the intercept term is not of interest. This is because we are not interested in eruptions that last for zero minutes. However, the slope term is of interest. The slope tells us the increase in the expected waiting time to the next eruption for every 1 minute increase in the length of the current eruption. Specifically, for every one minute increase in the length of the current eruption, we would expect to wait 10.73 minutes longer (on average) until the next eruption.

Verify that this linear regression is appropriate by making the residuals versus fitted-values and normal Q-Q Plot of the residuals in R. Select the appropriate statements for these plots from the list below. (Check all that apply.)

- [x] It appears safe to assume that the variance of the errors is constant.
- [ ] The errors do not appear to be normally distributed.
- [x] It appears safe to assume that the errors are normally distributed.
- [ ] Using this regression model for interpretation is not appropriate.
- [x] The regression model appears to be appropriate to use for interpretation.
- [ ] The variance of the errors does not appear to be constant.

Produce a scatterplot in R of this linear regression.

Say the most recent eruption of the Old Faithful Guyser lasted for 3.5 minutes. Visitors should expect to wait 71.025 minutes until the next eruption, on average.