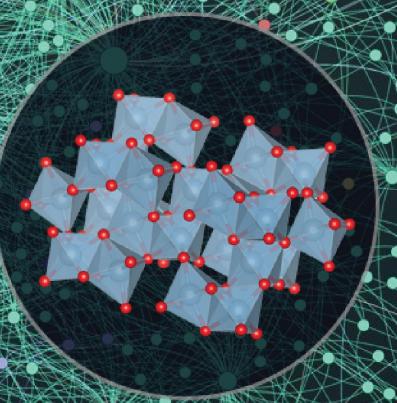
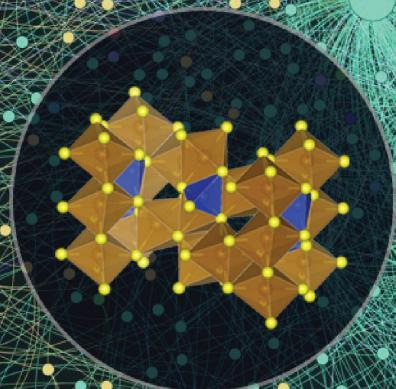
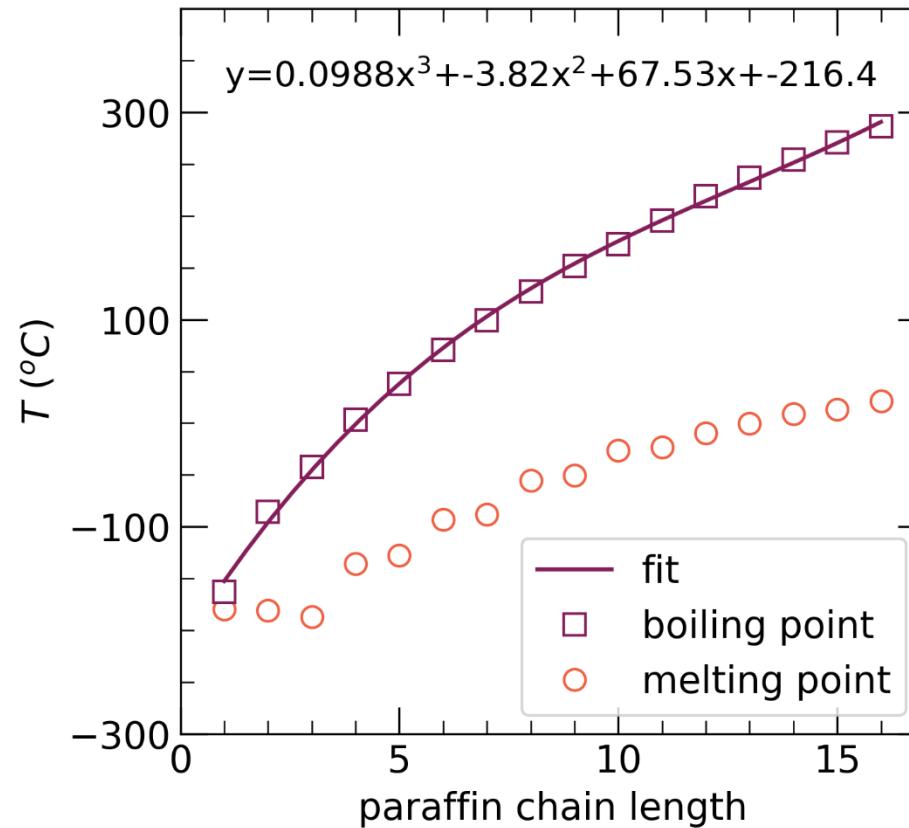


# metrics and evaluation



# How good is our fit? What metrics can we rely on?





# Most popular is Coefficient of Determination, R<sup>2</sup>

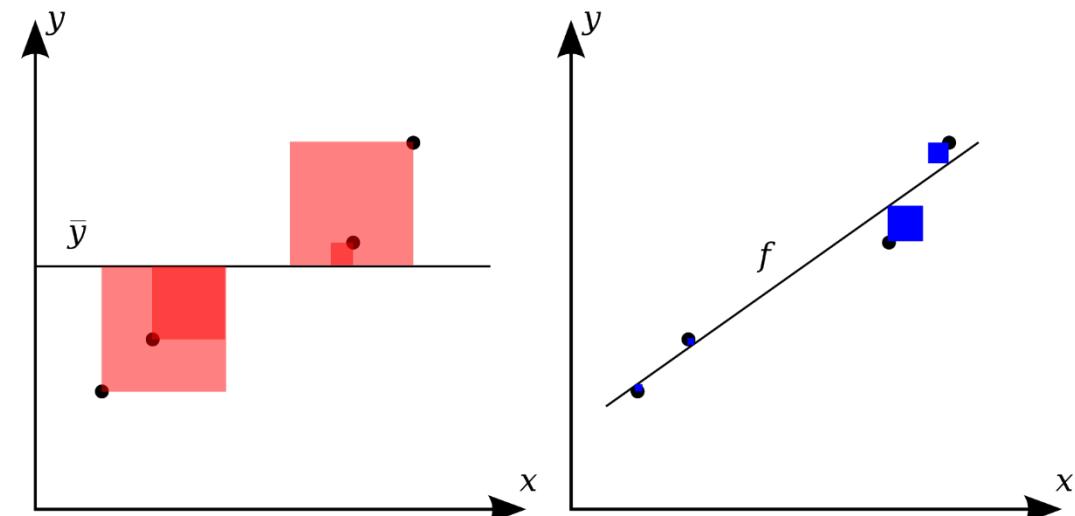
R<sup>2</sup> is the proportion of the variation in the dependent variable that is attributable to the independent variables

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$SS_{res} = \sum_i (y_i - f_i)^2$$

$$SS_{total} = \sum_i (y_i - \bar{y})^2$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$$



# $R^2$ does not work for nonlinear models!!

$R^2$  is the proportion of the variation in the dependent variable that is attributable to the independent variables

Total variance = error variance + explained variance

But for nonlinear models...

Total variance is NOT equal to error variance + explained variance

This leads to  $R^2$  being not between 0% and 100%

# $R^2$ does not work for nonlinear models!!

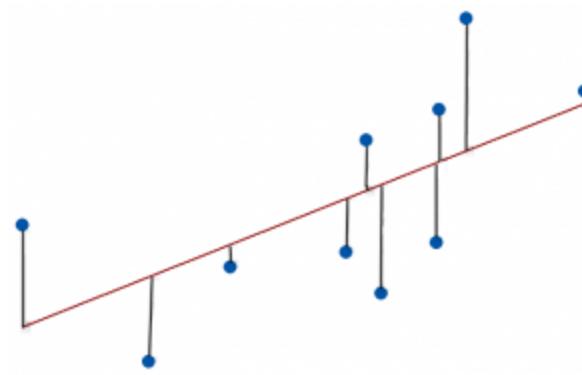
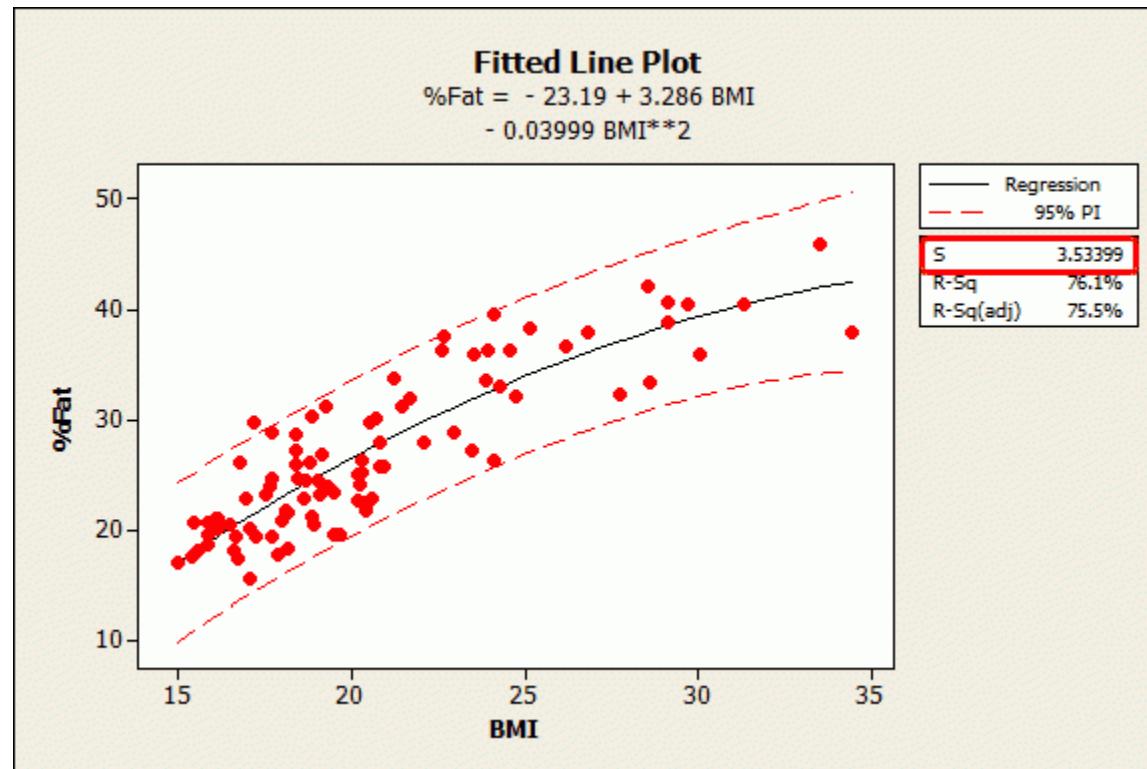
Statisticians carefully explored  $R^2$  on linear vs nonlinear models and found:

- R-squared is consistently high for both excellent and appalling models.
- R-squared will not rise for better models all of the time.
- If you use R-squared to pick the best model, it leads to the proper model only 28-43% of the time.

“ In the field of biochemical and pharmacological literature there is a reasonably high occurrence in the use of  $R^2$  as the basis of arguing against or in favor of a certain model. . . . Additionally, almost all of the commercially available statistical software packages calculate  $R^2$  values for nonlinear fits, which is bound to unintentionally corroborate its frequent use. . . . As a result from this work, we would like to advocate that  $R^2$  should not be reported or demanded in pharmacological and biochemical literature when discussing nonlinear data analysis.

# We have better options than R<sup>2</sup> for nonlinear models

One alternative is to simply quantify how far off the points are from the regression prediction. The average of this distance is the “standard error of the regression” and will have the same units as the dependent variable and will not vary between 0% and 100%.





# There are lots of other useful metrics

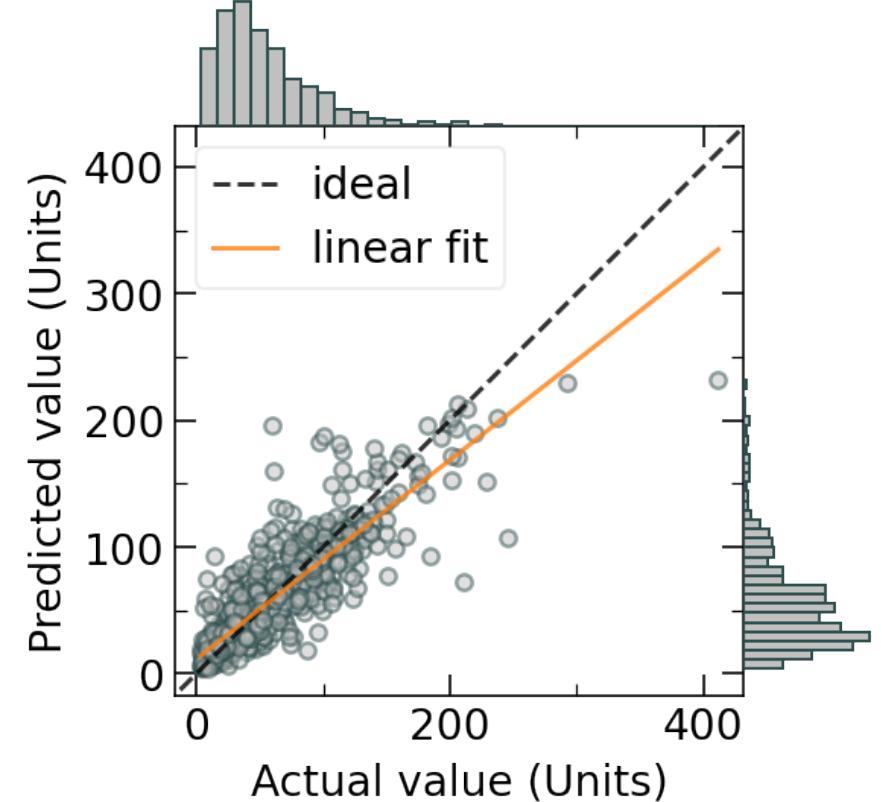
## Parity plots

- Include histograms on axes to show point density
- Semi-transparent points also helps
- Add 1:1 line for perfect fit, plus a best fit

$$\text{Mean absolute error, } MAE = \frac{\sum_{i=1}^n |y_i - f_i|}{n}$$

$$\text{Mean squared error, } MSE = \frac{\sum_{i=1}^n (y_i - f_i)^2}{n}$$

$$\text{Root mean squared error, } RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - f_i)^2}{n}}$$



# splitting data into test, train, validation

