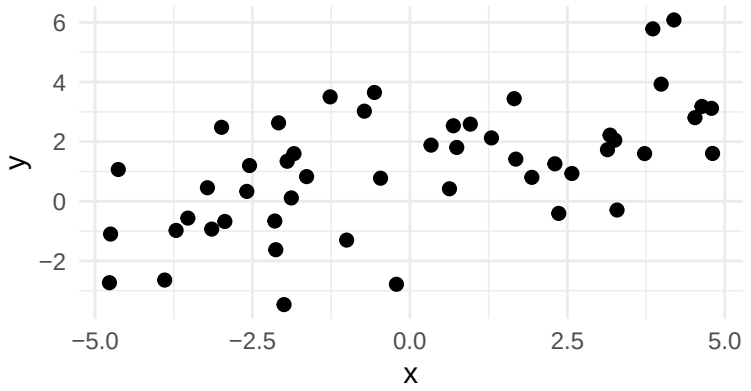
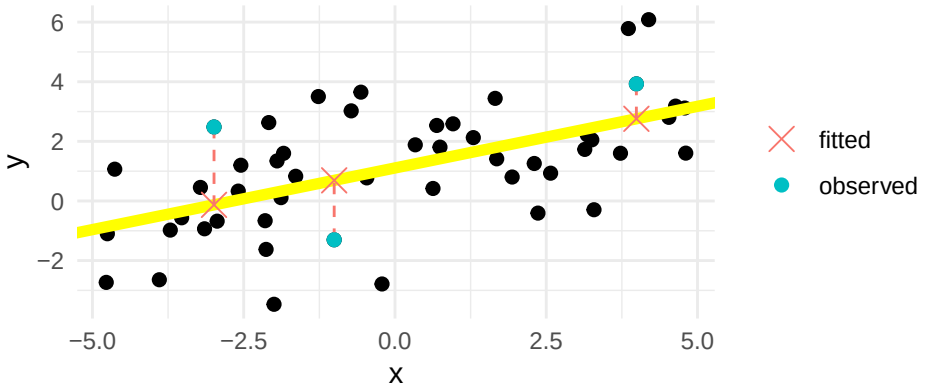
Introduction to Simple Linear Regression

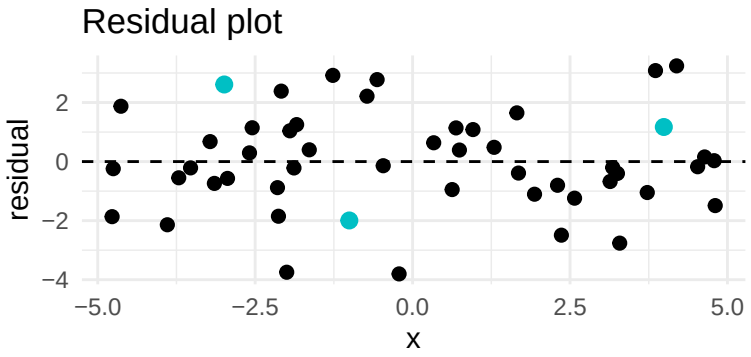
Some simulated data that we might want to fit a linear regression model to:



The yellow line is the line "fit" to the data above. We highlight three specific points in blue to compare the observed values to their fitted values:



What are the residuals of the highlighted observations?



Blue dots = residuals for specific points from previous plot

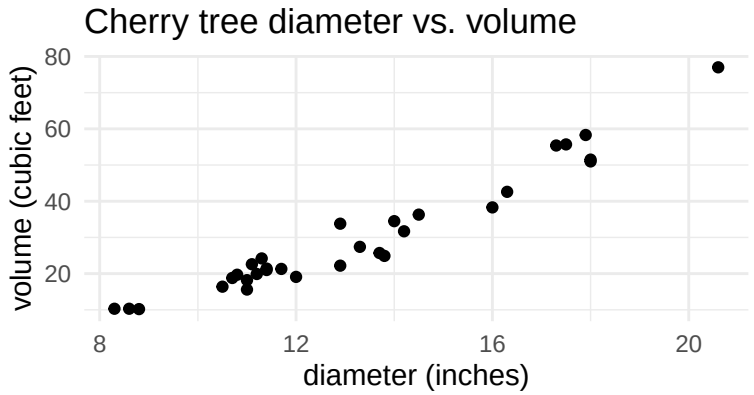
Running example

Using the `cherry` data from `openintro`: we will see if a linear regression model is appropriate to model the relationship between cherry tree volume (response) and diameter (explanatory).

Linear regression model (in context):

We check two conditions *before* fitting the the model.

Condition 1: **L**inearity



Does the linearity condition appear to be met?

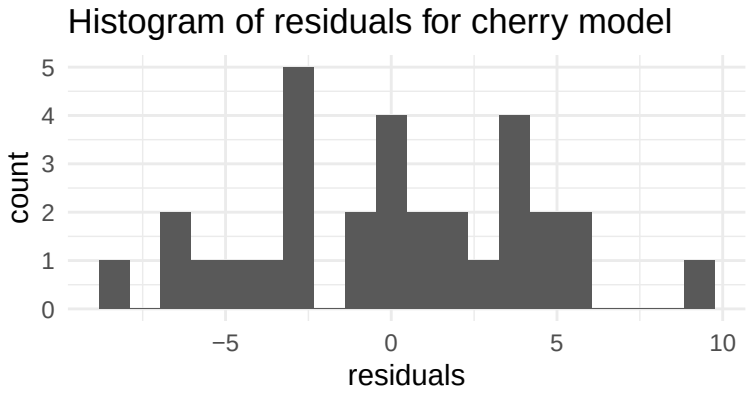
Condition 2: **I**ndependence

Does the independence condition appear to be met?

Fitted model (in context):

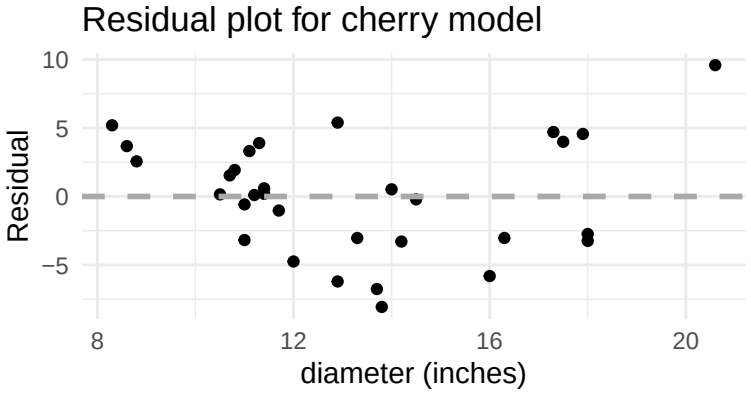
After obtaining the fitted model, we have access to the residuals which we use to assess the remaining two conditions.

Condition 3: **N**ormality



Does the Normality condition appear to be met?

Condition 4: Equal variance



Does the equal variance condition appear to be met?

Given our checks of the conditions, should we believe the linear regression model is appropriate for these data?