CHECKING THE NORMALITY ASSUMPTION

1. Theoretical considerations (e.g., CLT) might indicate conditional distributions are normal.

2. We can also check (*not* definitively) empirically:

Recall the error formulation of the model:

$Y|x = \eta_0 + \eta_1 x + e|x$
$e|\ x \sim N(0, \sigma^2)$, independent of x

Since the residuals $\hat{e}_i = y_i - \hat{y}_i$ approximate $e|x_i$, they approximate a sample from e|x. Thus a normal plot can give us some check on whether or not the errors might be normally distributed.

Cautions:

- The usual cautions in interpreting normal plots
- Since $\sum \hat{e}_i = 0$, the $\hat{e}_i$'s are *not* independent.

  (Thus only severe departures from a line should be taken as evidence of non-normality.)

Example: Forbes data

In arc: Residuals are automatically computed when doing regression.