Data

❑ y = variable you are trying to predict.
  ◦ Called many names: Dependent variable, response variable, target, regressand, …
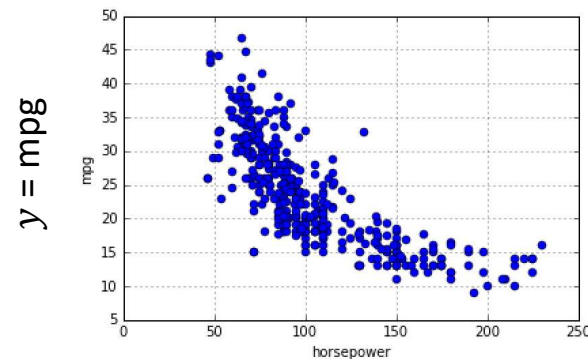
❑ x = what you are using to predict:
  ◦ Predictor, attribute, covariate, regressor, …

❑ Data: Set of points, $(x_i, y_i), i = 1, …, n$
  ◦ Each data point is called a sample.

❑ Scatter plot

x = horsepower

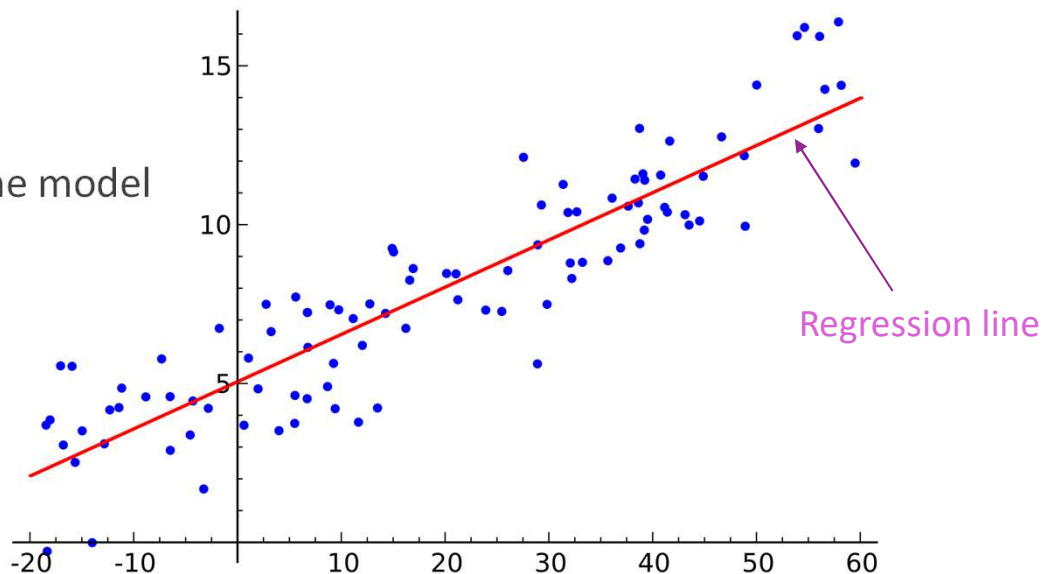# Linear Model

□ Assume a linear relation
$$y \approx \beta_0 + \beta_1 x$$
- ○ β_0 = intercept
- ○ β_1 = slope

□ $\beta = (\beta_0, \beta_1)$ are the parameters of the model

□ What are the units of β_0, β_1?

□ When is this model good?

Regression line

Why Use a Linear Model?

❑Many natural phenomena have linear relationship

❑Predictor has small variation
- Suppose $y = f(x)$
- If variation of x is small around some value x_0, then

$$y \approx f(x_0) + f'(x_0)(x - x_0) = \beta_0 + \beta_1 x,$$

$$\beta_0 = f(x_0) - f'(x_0)x_0, \qquad \beta_1 = f'(x_0)$$

❑Simple to compute

❑Easy to interpret relation

❑Gaussian random variables: If x and y were Gaussian, optimal estimator of y is linear in x