- The equation for a line is $y = mx + b$. The m is the "slope" or "coefficient of x" and the b is the intercept or bias.
- Let x_i and y_i for stand in for our feature values and target values respectively.
  - For example, if our feature values are $[1, 2, 3, 4, 5]$ then x_0 is 1, and x_1 is 2, and so on.
- Let $\hat{y}_i$ stand in for our line's values at each x_i.
  - So if our line is $y = 2x + 1$, then

    $\hat{y}_0 = 2x_0 + 1 = 2 \times 1 + 1 = 3$.

    y_3

    $\hat{y}_3$

    $\hat{y}_3 - y_3$

- To find the line of best fit, we minimize the "residual sum of squares".
- A residual is the difference between the true value of a data point, and our model's predicted value.

- Residuals are usually defined as $y_i - \hat{y}_i$.
- To save us from negative residuals, we square them, so each residual in our "sum of squares" is

  $(y_i - \hat{y}_i)^2$.
- And finally we sum them up for each i to get our "residual sum of squares".

- Let n be the number of real data points that we have.
- Thus our "loss function" is as follows:

  $$RSS = (y_0 - \hat{y}_0)^2 + (y_1 - \hat{y}_1)^2 + \ldots + (y_n - \hat{y}_n)^2$$
- Some fancy calculus allows us to find out exactly what our model's coefficient and intercept need to be.

- Let p equal our number of features.
- The equation for our linear regression in higher dimensional space:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

- R^2