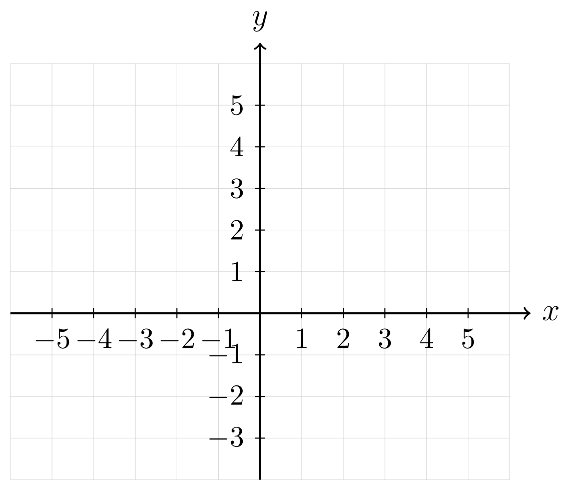
Warm up ***(find and work with a partner)***

1. Say we have the following linear model:

$$y = 1 + \frac{x}{3}.$$

Sketch a graph of this line on the coordinate plane below.



2. What is the slope? What is the y-intercept?

3. What parameters do the slope and y-intercept correspond to in our linear model?

4. Suppose we have the following five points. For each point, calculate the model's prediction and the residual. Plot each observed point on the graph above and draw a vertical line representing its residual.

$$e_i = y_i - \hat{y}_i$$

	x_i	y_i	$\hat{y}_i = 1 + \frac{x_i}{3}$	Residual $e_i = y_i - \hat{y}_i$
A	0	2		
B	3	1		
C	6	3		
D	9	3		
E	12	6		

5. What is the sum of the five residuals?

Chain Rule

Recall that if $h(x) = f(g(x))$, then $h'(x) = f'(g(x))g'(x)$.
In words: *differentiate the outside, leave the inside alone, and multiply by the derivative of the inside.*

1. Find $\frac{dy}{dx}$: $y = (3x + 2)^2$

2. Find $\frac{dy}{dx}$: $y = \frac{1}{2}(5 - x)^2$

3. Find $\frac{dy}{dx}$: $y = \frac{1}{2}(7 - 2x)^2$

Parial Derivatives

When a function has multiple variables, a partial derivative tells us how the function changes with respect to one variable while treating the other variables as constants.
For example, if $f(x, y) = x^2 + 3y$, then $\frac{\partial f}{\partial x} = 2x$ and $\frac{\partial f}{\partial y} = 3$.

Find the following partial derivatives.

1. Let $f(w_0, w_1) = 3w_0 + w_1^2$. Find $\frac{\partial f}{\partial w_0}$ and $\frac{\partial f}{\partial w_1}$.

2. Let $f(w_0, w_1) = (4 - w_0 - w_1)^2$. Find $\frac{\partial f}{\partial w_0}$.

3. Let $f(x, y) = x^2 + 3xy + y^2$. Find $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$.

4. Let $f(w_0, w_1) = \frac{1}{2}(4 - w_0 - 3w_1)^2$. Find $\frac{\partial f}{\partial w_0}$ and $\frac{\partial f}{\partial w_1}$.

Linear Regression *(find and work with a partner)*

In the case of *simple* linear regression, for each example we have a single feature x and associated response (output) variable y. Our model is:

$$h_{\boldsymbol{w}}(x) = w_0 + w_1 x$$

1. Name one goal of linear regression. Why would we want to fit a linear model?

2. In simple linear regression, how many features does each example have? How many parameters does our linear model have in this case?

   - # features (p) =
   - # model params =

3. What cost/loss function are we trying to minimize in simple linear regression?

4. *Small example.* Let $n = 2$ and $p = 1$, with the following data:

   $$\boldsymbol{X} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

   Plot these two points – what should $\hat{w}_0$ and $\hat{w}_1$ be?

5. This week we derived the solution for simple linear regression:

   $$\hat{w}_1 = \frac{\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y})}{\mathrm{Var}(\boldsymbol{x})} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \hat{w}_0 = \bar{y} - \hat{w}_1 \bar{x}$$

   where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. Use these equations to compute $\hat{w}_0$ and $\hat{w}_1$ and verify your answer to the previous question.