5.4/5.7 Best Fitting Lines

Usually there is **NO** single line that passes through all of the data points, so we try to find the line that **best** fits the data. This is called the **best-fit line**. We can do this by drawing a line on a graph.
Example: On the scatter plot, draw a best-fit line.

- Choose 2 points on the line: 
  \[(2,1) \quad (8,8)\]

- Find the slope between the 2 points:
  \[
  \frac{y_2-y_1}{x_2-x_1} = \frac{8-1}{8-2} = \frac{7}{6}
  \]

- Use slope-intercept or point-slope form to find the equation of the line.
  \[
  y-1 = \frac{7}{6}(x-2) \\
  y-1 = \frac{7}{6}x - \frac{14}{6} + \frac{6}{6} \\
  y = \frac{7}{6}x - \frac{8}{6} \\
  y = \frac{7}{6}x - \frac{4}{3}
  \]
The number $r$, which indicates how well a set of data can be estimated by a straight line, is \underline{correlation}.

Correlation $r$ is always \underline{between -1 and 1}.

When the points on a scatter plot can be approximated by a line with \underline{positive} slope, $x$ and $y$ have a \underline{positive} correlation. When the points can be approximated by a line with \underline{negative} slope, $x$ and $y$ have a \underline{negative} correlation. When points CANNOT be approximated by a straight line, there is \underline{no linear correlation}. 
Example: Label each graph with its correlation. (positive, negative, no corr.)

- None
- Negative Linear
- Positive Linear
We can decide when a **linear** model can be used to represent real-life data.

We do this by **creating** a scatter plot of the data.

On the graph, are the points almost in a straight line? **Yes**

If the points look like they’re almost in a straight line on the graph, then the data can be modeled with a **linear equation**.