The Geometry of Correlation

While a scatterplot allows us to *see* a relationship, the **Correlation Coefficient** (r) allows us to **measure** it. r is a bounded decimal between -1.00 and 1.00 that tells us two things: the **direction** of the relationship and how **tightly** the data hugs a straight line.

1. Interactive Friction: The Morphing r

Scan the QR code below to open the **Desmos Correlation Engine**.

https://www.desmos.com/calculator/gv3ypw9ouh

In **Section 5.3** of your book, find the exploration which matches this exploration. Once you find it, ensure you toggle the **Visual** switch to **ON**. Follow the instructions to drag the points until you have seen plots for the r-values listed in the book. **Pay attention to where the dashed line exits the darker purple rectangle.**

Complete the exploration in your textbook first, then return here to answer the reflection questions below.

Reflection: Reflecting on the Geometry

Based on your engineering work in Desmos and the textbook, answer the following:

1. **The Cloud:** As you moved the dots what happened to the "tightness" of the points around the line as you went from $r = 1$ to $r = 0$ and eventually $r = -1$?

2. **The Magnitude:** You engineered a weak positive ($r = 0.40$) and a moderate negative ($r = -0.70$). Which one actually represents a **stronger** linear relationship?

CHAPTER 5

2. The Shared Story (r^2)

Bill the Statypus says: Don't let the decimal fool you. While researchers report r to show direction, they use r^2 (The Coefficient of Determination) to explain **power**. I call this the **"Strength"** percentage.

Statypus Insight: Conversion

Take a moderate negative correlation ($r = -0.85$) and square it.

$$(-0.85)^2 = \_\_\_\_\_ = \_\_\_\_\_\%$$

Even though the relationship was negative, why is it mathematically impossible for the "Strength" percentage to be a negative number?

While the **tightness** ($|r|$) tells us how close the points are to the line, the **strength** (r^2) tells us the proportion of the story that is actually shared. An r of 0.50 is tight, but its true predictive strength is only 25%.

3. The Engine: Fish Market Analysis

We are moving away from the Catchment to look at a new dataset: `FishMarket` in the Chapter 5 .RData file. This data contains measurements for 159 individual fish.

Bill the Statypus says: (sighs) ...beginning of Chapter 5... look for New R Functions... find the .RData file for chapter 5

Coding Corner: Finding the Value of r

Look through Section 5.3 of your textbook to find the function to find the correlation coefficient between the **Height** and **Weight** variables of **FishMarket** and use the function to calculate r.

Sally the Statypus says: We try to always follow the **Statypus Reporting Standards.**
When reporting **Correlation (r)**, always use a decimal (e.g., -0.85).
When reporting **Determination (r^2)**, always use a percentage (e.g., 72.25%).
We will explain why we call r^2 the "Strength" of the relationship later.

Record the values of r and r^2 below following the **Statypus Reporting Standards.**