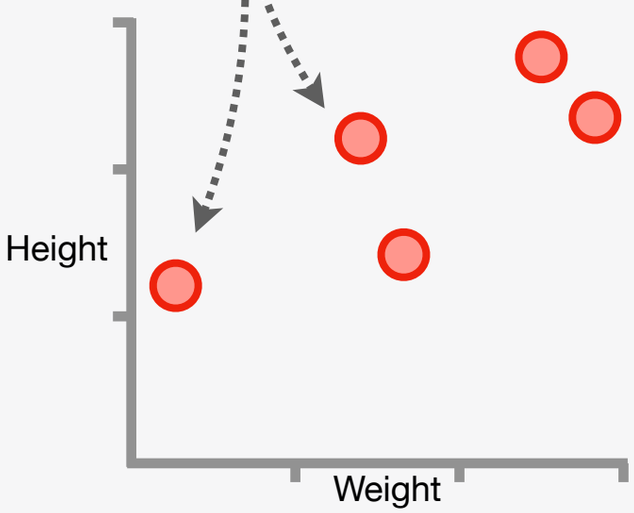
Chapter 04

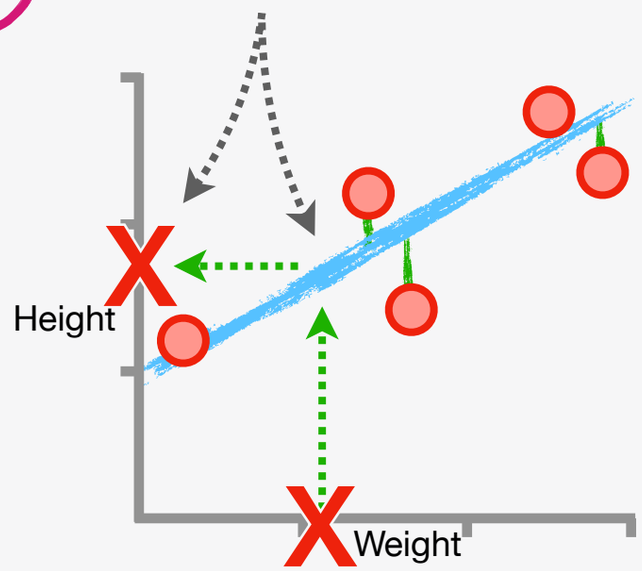
Linear Regression!!!

Linear Regression: Main Ideas

1 **The Problem:** We've collected Weight and Height measurements from **5** people, and we want to use Weight to predict Height, which is ***continuous***…



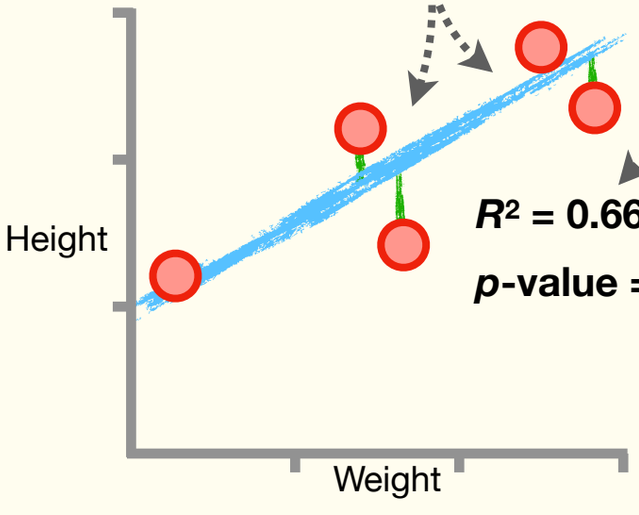
…and in **Chapter 3**, we learned that we could fit a **line** to the data and use it to make predictions.



However, **1)** we didn't talk about how we fit a **line** to the data…

…and **2)** we didn't calculate a ***p*-value** for the **fitted line**, which would quantify how much confidence we should have in its predictions compared to just using the **mean** y-axis value.

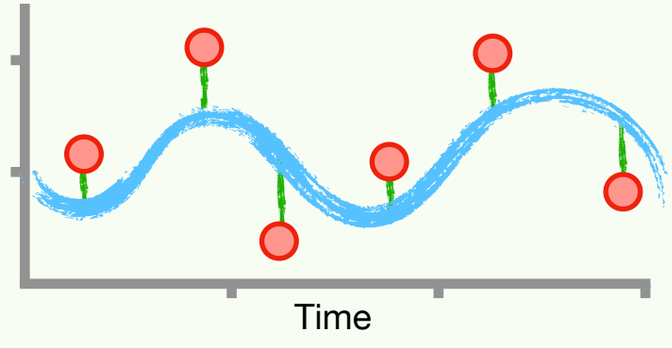
2 **A Solution: Linear Regression** fits a **line** to the data that *minimizes* the **Sum of the Squared Residuals** (**SSR**)…



…and once we fit the line to the data, we can easily calculate R^2, which gives us a sense of how accurate our predictions will be…

…and **Linear Regression** provides us with a ***p*-value** for the R^2 value, so we can get a sense of how confident we should be that the predictions made with the **fitted line** are better than predictions made with the **mean** of the y-axis coordinates for the data.

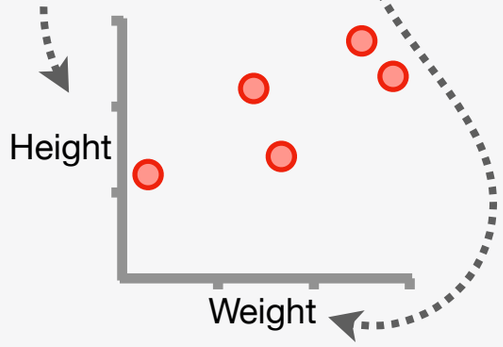
NOTE: Linear Regression is the gateway to a general technique called **Linear Models**, which can be used to create and evaluate models that go way beyond fitting simple straight lines to data!!!



BAM!!!

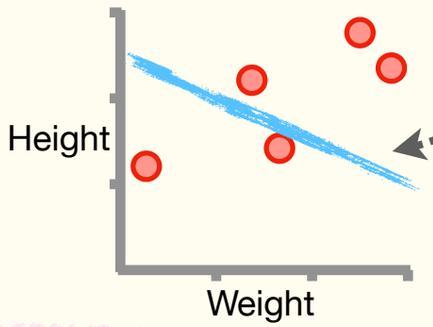
Fitting a Line to Data: Main Ideas

1 Imagine we had Height and Weight data on a graph…

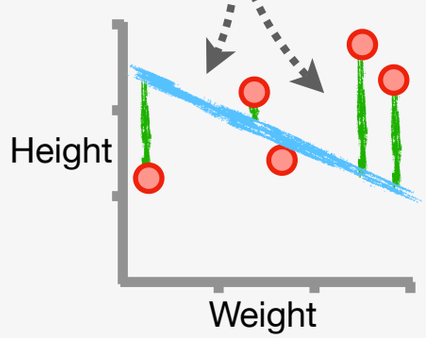


…and we wanted to predict Height from Weight.

2 Because the heavier Weights are paired with taller Heights, this **line** makes terrible predictions.



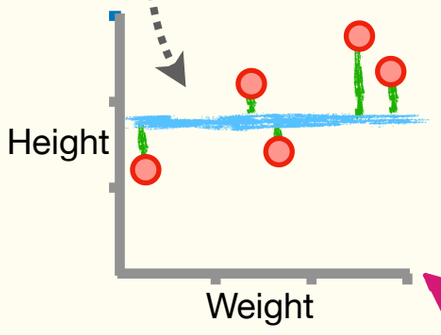
3 We can quantify how bad these predictions are by calculating the **Residuals**, which are the differences between the Observed and Predicted heights…



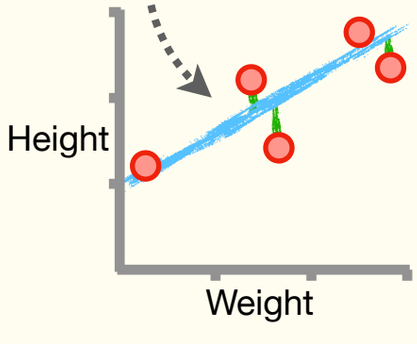
…and using the **Residuals** to calculate the **Sum of the Squared Residuals** (**SSR**).

Then we can plot the **SSR** on this graph that has the **SSR** on the y-axis, and different lines fit to the data on the x-axis.

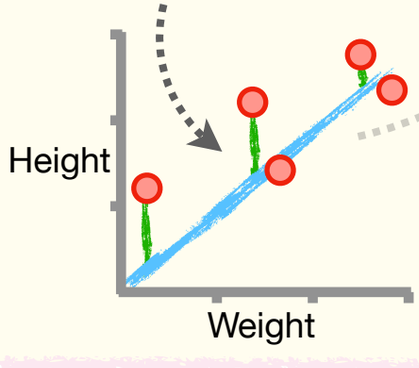
4 This **line**, which has a different y-axis intercept and slope, gives us slightly smaller residuals and a smaller **SSR**…

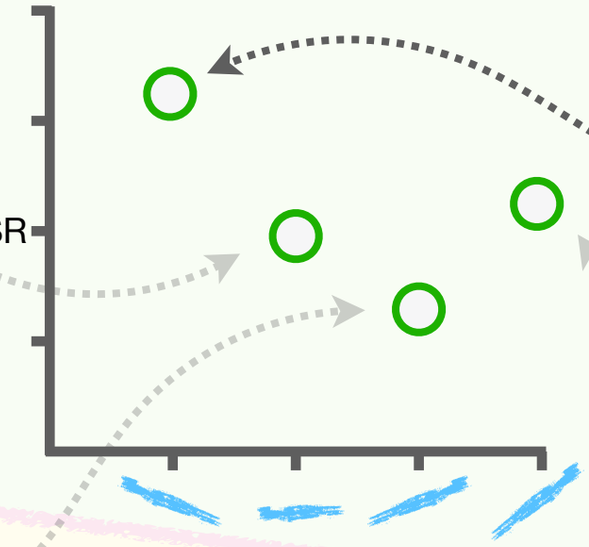


…and this **line** has even smaller residuals and a smaller **SSR**…



…and this **line** has larger residuals and a larger **SSR**.





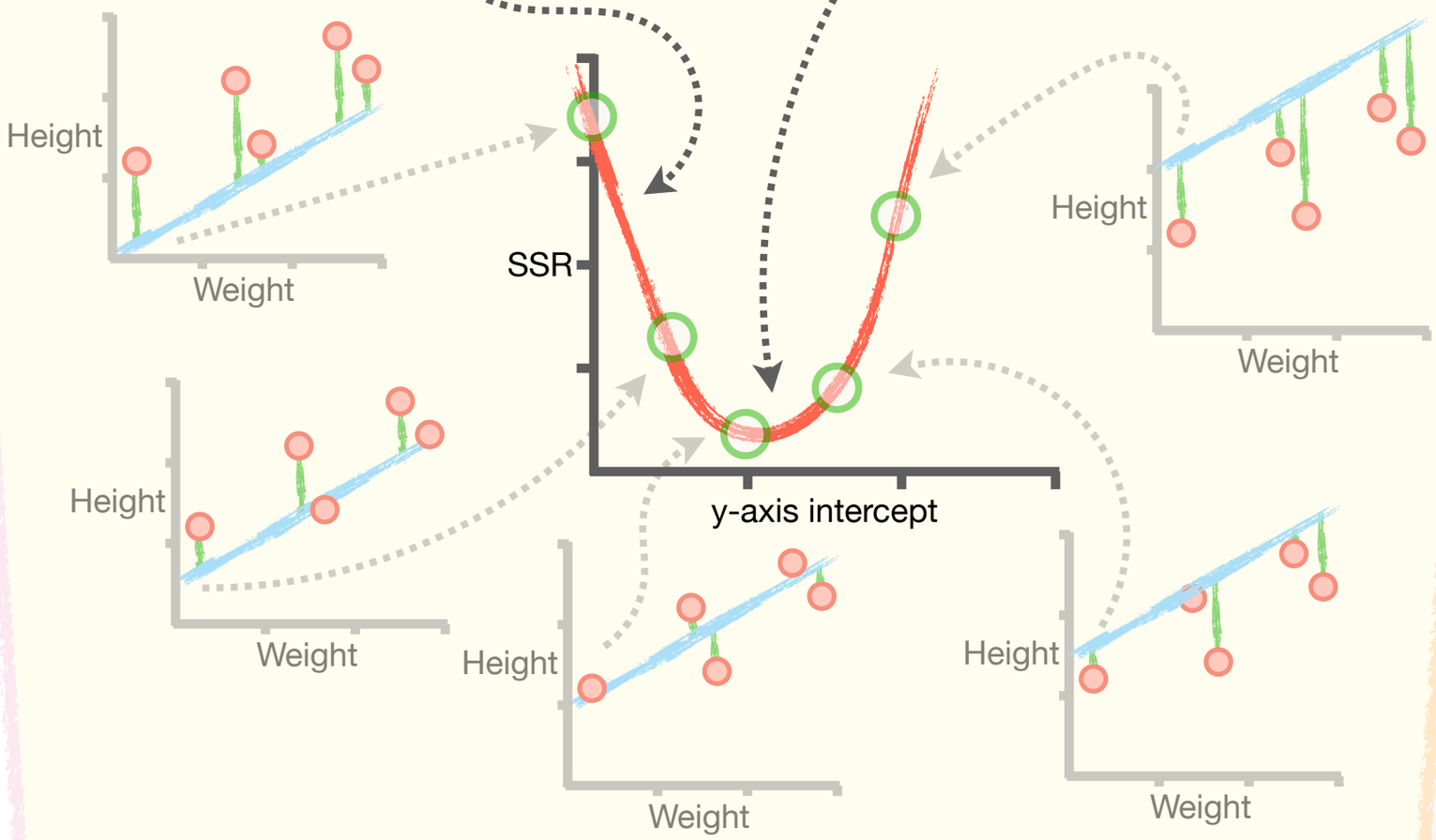
5 As we can see on the graph, different values for a **line**'s y-axis intercept and slope, shown on the x-axis, change the **SSR**, shown on the y-axis. **Linear Regression** selects the **line**, the y-axis intercept and slope, that results in the minimum **SSR**.

BAM!!!

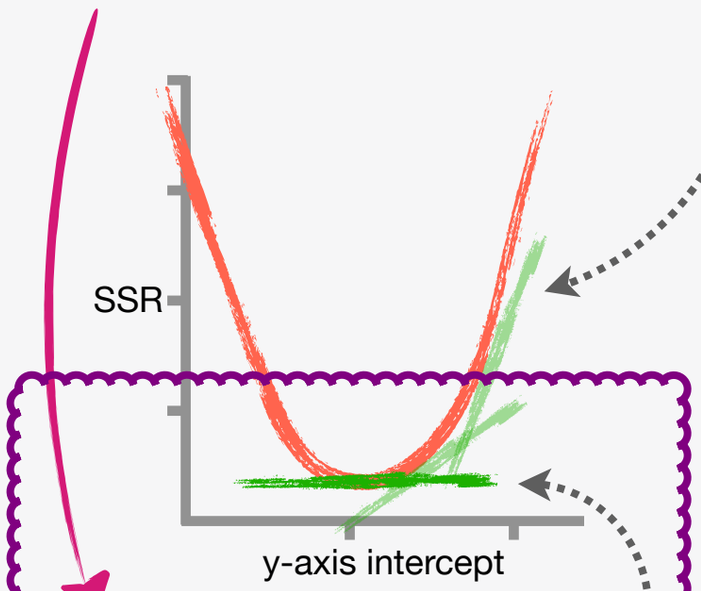
Fitting a Line to Data: Intuition

1 If we don't change the slope, we can see how the **SSR** changes for different y-axis intercept values…

…and, in this case, the goal of **Linear Regression** would be to find the y-axis intercept that results in the lowest **SSR** at the bottom of this curve.



2 One way to find the lowest point in the **curve** is to calculate the **derivative** of the **curve** (**NOTE:** If you're not familiar with derivatives, see **Appendix D**).



…and solve for where the **derivative** is equal to **0**, at the bottom of the **curve**.

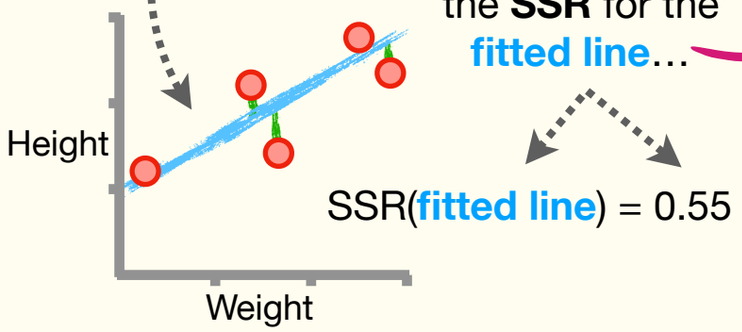
Solving this equation results in an **Analytical Solution**, meaning, we end up with a formula that we can plug our data into, and the output is the optimal value. Analytical solutions are awesome when you can find them (like for **Linear Regression**), but they're rare and only work in very specific situations.

3 Another way to find an optimal slope and y-axis intercept is to use an **Iterative Method** called **Gradient Descent**. In contrast to an **Analytical Solution**, an **Iterative Method** starts with a guess for the value and then goes into a loop that improves the guess one small step at a time. Although **Gradient Descent** takes longer than an analytical solution, it's one of the most important tools in machine learning because it can be used in a wide variety of situations where there are no analytical solutions, including **Logistic Regression**, **Neural Networks**, and many more.

Because **Gradient Descent** is so important, we'll spend all of **Chapter 5** on it. **GET EXCITED!!!**

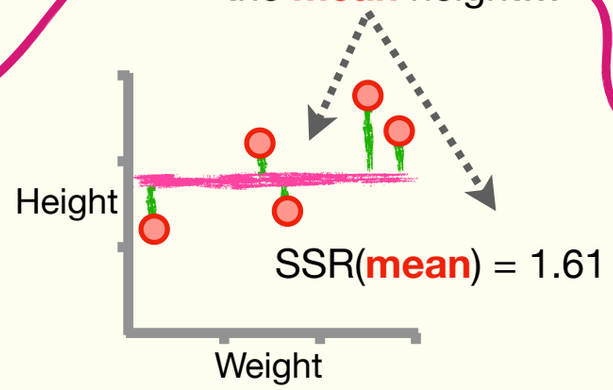
p-values for Linear Regression and R^2: Main Ideas

1 Now, assuming that we've fit a line to the data that minimizes the **SSR** using an analytical solution or **Gradient Descent**, we calculate R^2 with the **SSR** for the **fitted line**…



SSR(**fitted line**) = 0.55

…and the **SSR** for the **mean** height…



SSR(**mean**) = 1.61

…and plug them into the equation for R^2 and get **0.66**.

Gentle Reminder:

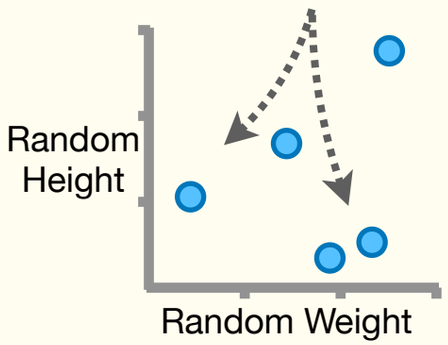
$$R^2 = \frac{\text{SSR(mean)} - \text{SSR(fitted line)}}{\text{SSR(mean)}}$$

$$R^2 = \frac{1.61 - 0.55}{1.61} = 0.66$$

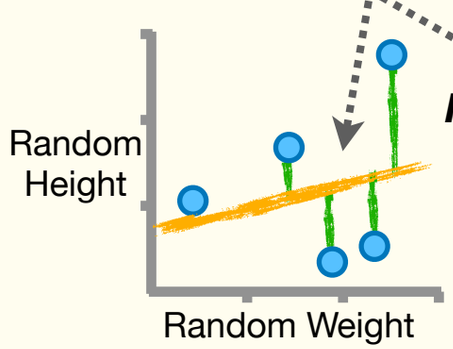
2 The R^2 value, **0.66**, suggests that using Weight to predict Height will be useful, but now we need to calculate the ***p*-value** to make sure that this result isn't due to random chance.

In this context, a ***p*-value** tells us the probability that random data could result in a similar R^2 value or a better one. In other words, the ***p*-value** will tell us the probability that random data could result in an $R^2 \geq 0.66$.

3 Because the original dataset has **5** pairs of measurements, one way* to calculate a ***p*-value** is to pair **5** *random* values for Height with **5** *random* values for Weight and plot them on a graph…

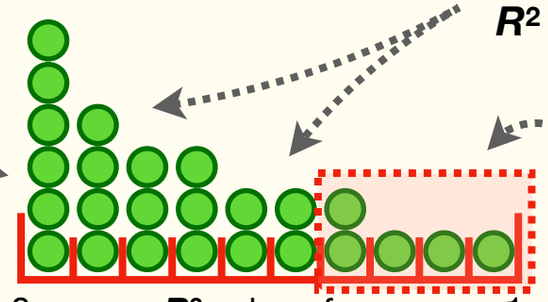


…then use **Linear Regression** to fit a line to the random data and calculate R^2…



$R^2 = 0.03$

…and then add that R^2 to a histogram…



…and then create **>10,000** more sets of random data and add their R^2 values to the histogram…

…and use the histogram to calculate the probability that random data will give us an $R^2 \geq 0.66$.

4 In the end, we get ***p*-value = 0.1**, meaning there's a **10%** chance that random data could give us an $R^2 \geq 0.66$. That's a relatively *high* ***p*-value**, so we might not have a lot of confidence in the predictions, which makes sense because we didn't have much data to begin with.

small bam.

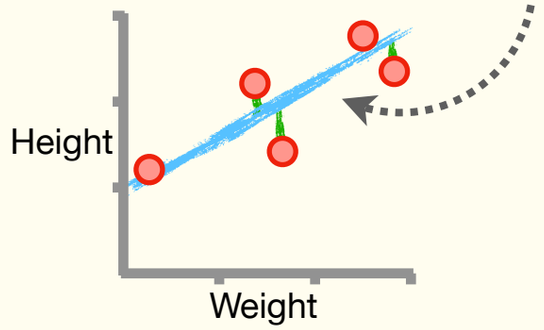
* **NOTE:** Because **Linear Regression** was invented before computers could quickly generate random data, **this is not the traditional way to calculate *p*-values**, but it works!!!

Multiple Linear Regression: Main Ideas

1 So far, the example we've used demonstrates something called **Simple Linear Regression** because we use one variable, Weight, to predict Height…

Height = 1.1 + 0.5 x **Weight**

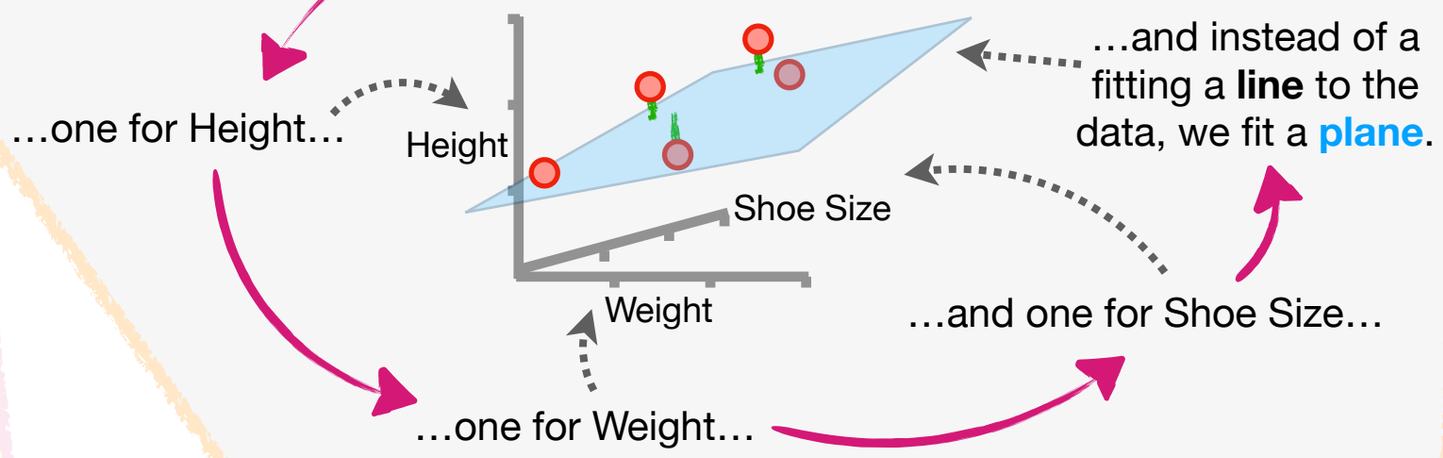
…and, as we've seen, **Simple Linear Regression** fits a line to the data that we can use to make predictions.



2 However, it's just as easy to use **2** or more variables, like Weight and Shoe Size, to predict Height.

Height = 1.1 + 0.5 x **Weight** + 0.3 x **Shoe Size**

This is called **Multiple Linear Regression**, and in this example, we end up with a **3**-dimensional graph of the data, which has **3** axes…



3 Just like for **Simple Linear Regression**, **Multiple Linear Regression** calculates R^2 and ***p*-values** from the **Sum of the Squared Residuals** (**SSR**). And the **Residuals** are still the difference between the **Observed Height** and the **Predicted Height**.

The only difference is that now we calculate **Residuals** around the **fitted plane** instead of a line.

$$R^2 = \frac{\text{SSR(mean)} - \text{SSR(fitted plane)}}{\text{SSR(mean)}}$$

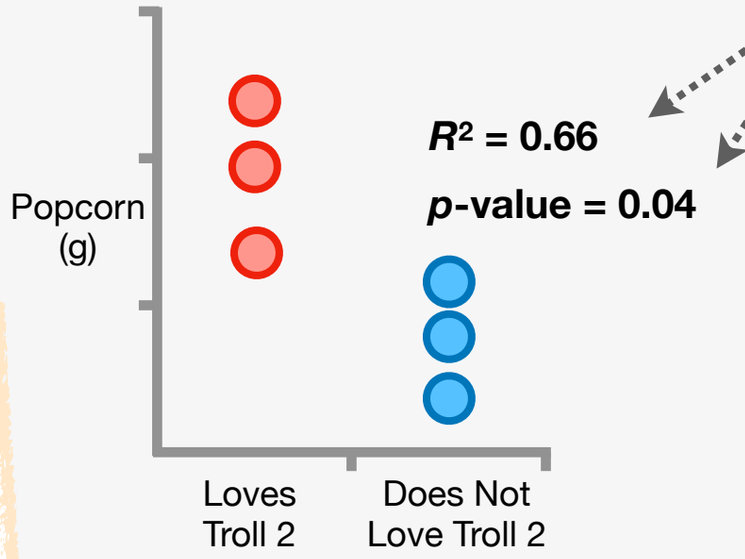
4 And when we use **3** or more variables to make a prediction, we can't draw the graph, but we can still do the math to calculate the **Residuals** for R^2 and its ***p*-value**.

Bam.

Beyond Linear Regression

1 As mentioned at the start of this chapter, **Linear Regression** is the gateway to something called **Linear Models**, which are incredibly flexible and powerful.

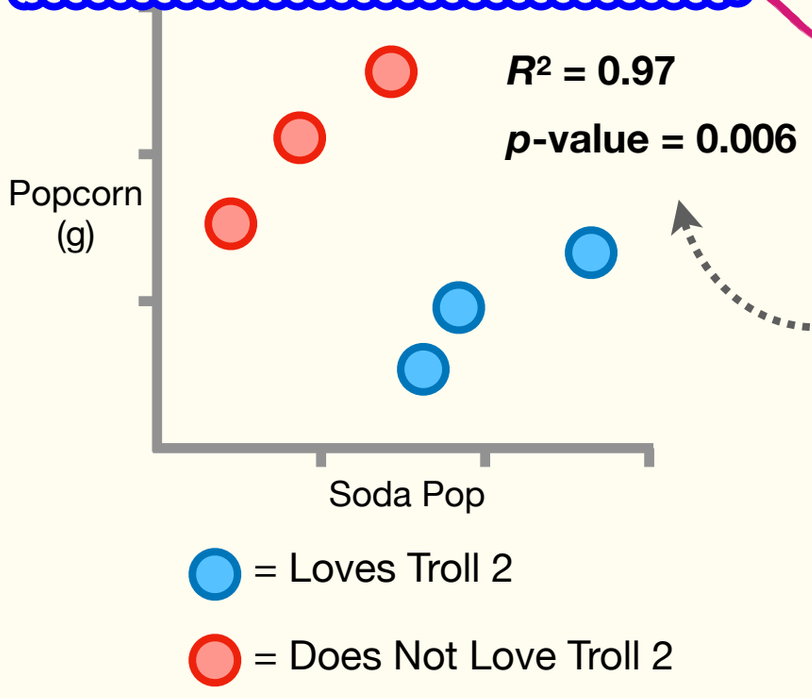
2 **Linear Models** allow us to use ***discrete*** data, like whether or not someone loves the movie Troll 2, to predict something ***continuous***, like how many grams of Popcorn they eat each day.



3 Just like when we used Weight to predict Height, **Linear Models** will give us an R^2 for this prediction, which gives us a sense of how accurate the predictions will be, and a ***p*-value** that lets us know how much confidence we should have in the predictions.

In this case, the ***p*-value** of **0.04** is relatively *small*, which suggests that it would be unlikely for random data to give us the same result or something more extreme. In other words, we can be confident that knowing whether or not someone loves Troll 2 will improve our prediction of how much Popcorn they will eat.

4 **Linear Models** also easily combine ***discrete*** data, like whether or not someone loves Troll 2**,** with ***continuous*** data, like how much Soda Pop they drink, to predict something ***continuous***, like how much Popcorn they will eat.

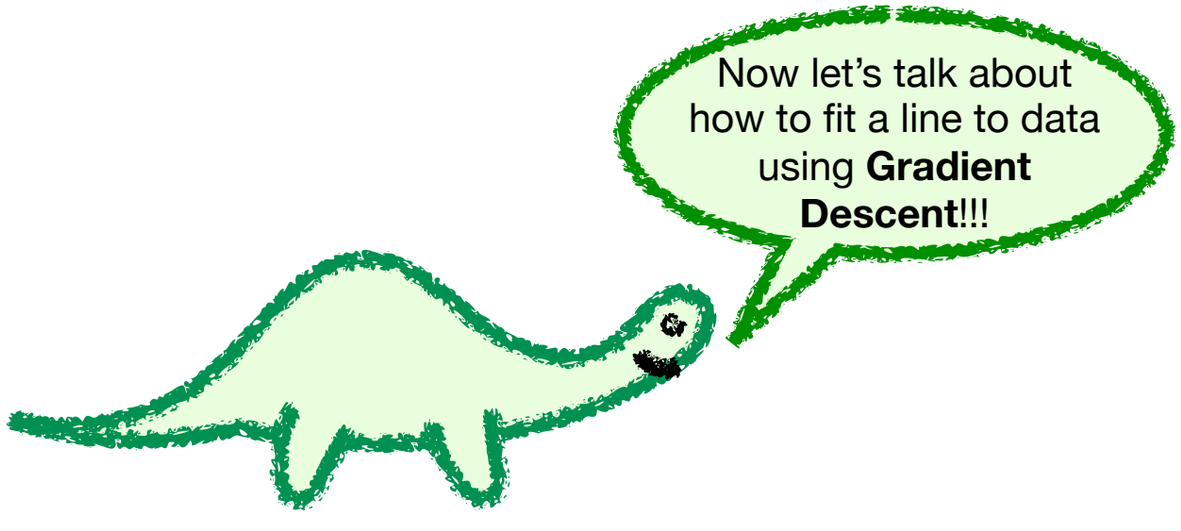


In this case, adding how much Soda Pop someone drinks to the model dramatically increased the R^2 value, which means the predictions will be more accurate, and reduced the ***p*-value**, suggesting we can have more confidence in the predictions.

DOUBLE BAM!!!

5 If you'd like to learn more about **Linear Models**, scan, click, or tap this QR code to check out the **'Quests** on YouTube!!!

bam.


Now let's talk about how to fit a line to data using **Gradient Descent**!!!