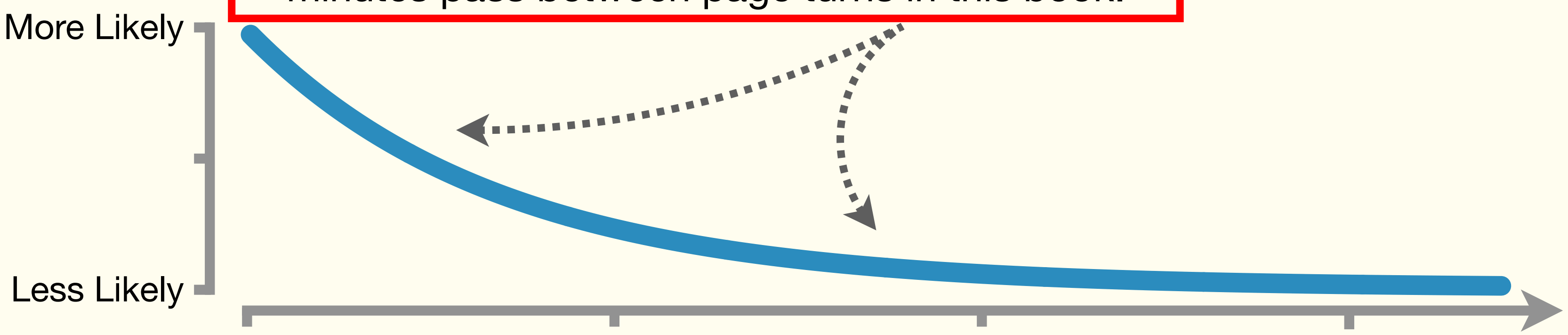
Other Continuous Distributions: Main Ideas

1 **Exponential Distributions** are commonly used when we're interested in how much time passes between events. For example, we could measure how many minutes pass between page turns in this book.



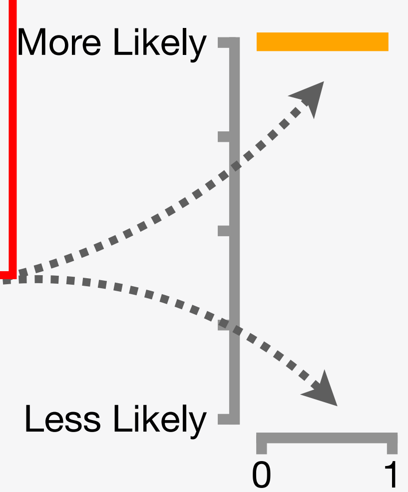
Using Distributions To Generate Random Numbers

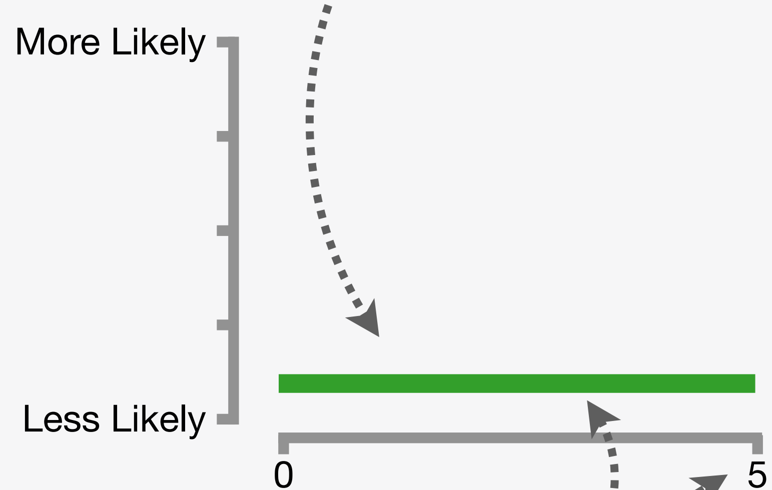
We can get a computer to generate numbers that reflect the likelihoods of any distribution. In machine learning, we usually need to generate random numbers to initialize algorithms before training them with **Training Data**. Random numbers are also useful for randomizing the order of our data, which is useful for the same reasons we shuffle a deck of cards before playing a game. We want to make sure everything is randomized.

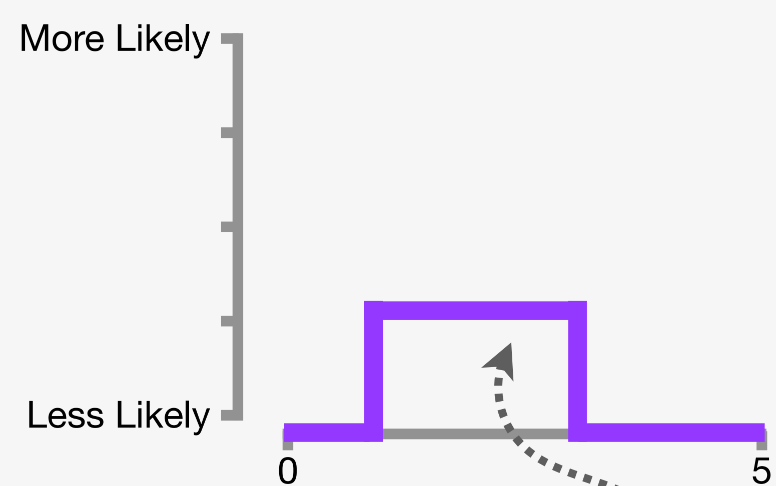
2 **Uniform Distributions** are commonly used to generate random numbers that are equally likely to occur.

NOTE: Because there are fewer values between **0** and **1** than between **0** and **5**, we see that the corresponding likelihood for any specific number is higher for the **Uniform 0,1 Distribution** than the **Uniform 0,5 Distribution**.

For example, if I want to select random numbers between **0** and **1**, then I would use a **Uniform Distribution** that goes from **0** to **1**, which is called a **Uniform 0,1 Distribution**, because it ensures that every value between **0** and **1** is equally likely to occur.





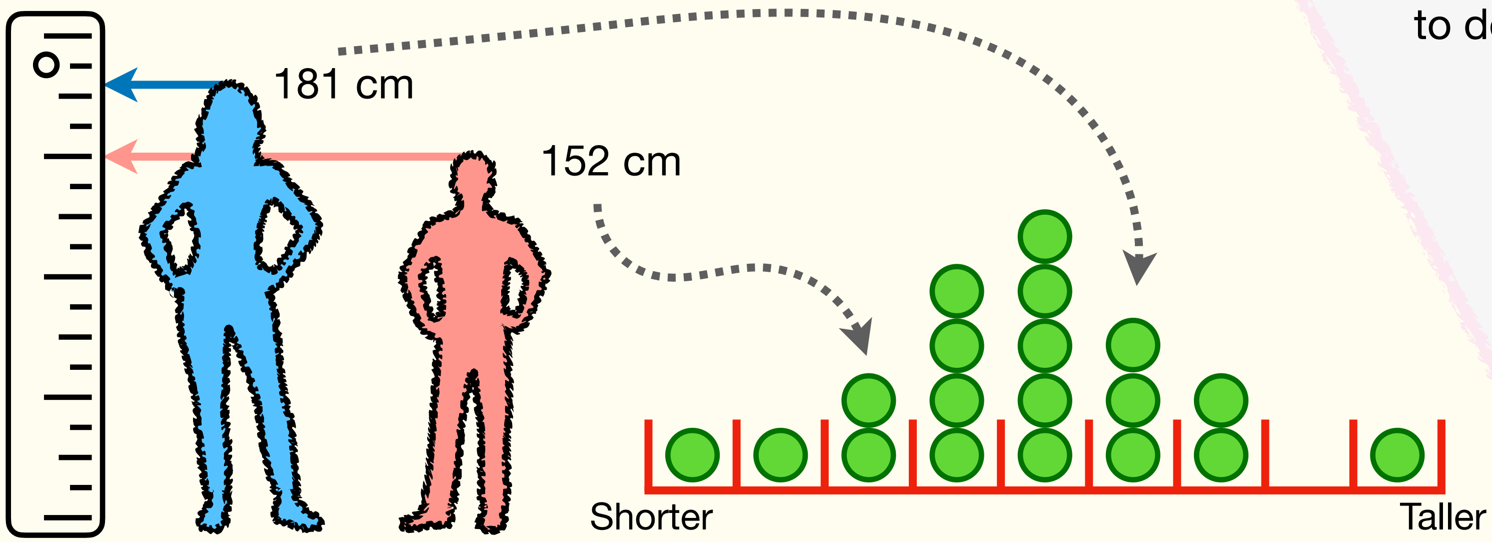


In contrast, if I wanted to generate random numbers between **0** and **5**, then I would use a **Uniform Distribution** that goes from **0** to **5**, which is called a **Uniform 0,5 Distribution**.

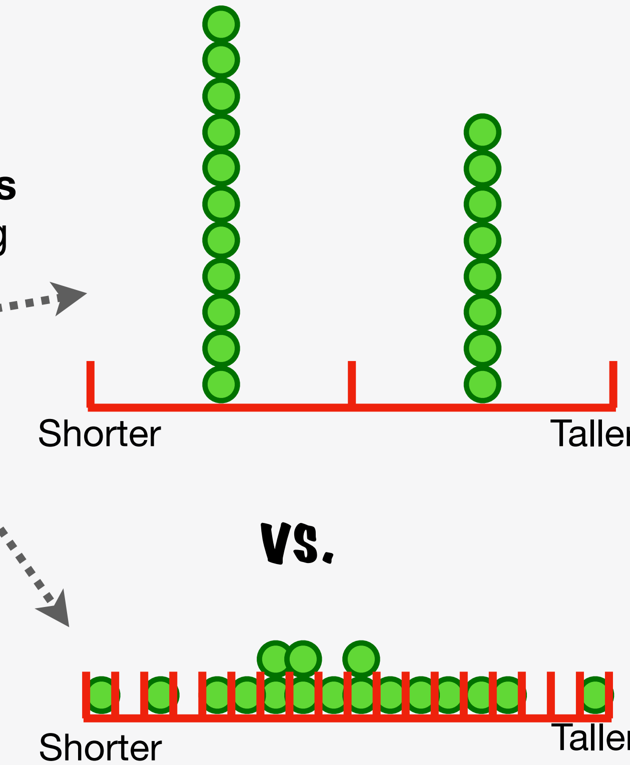
Uniform Distributions can span any **2** numbers, so we could have a **Uniform 1,3.5 Distribution** if we wanted one.

Continuous Probability Distributions: Summary

1 Just like **Discrete Distributions**, **Continuous Distributions** spare us from having to gather tons of data for a histogram…

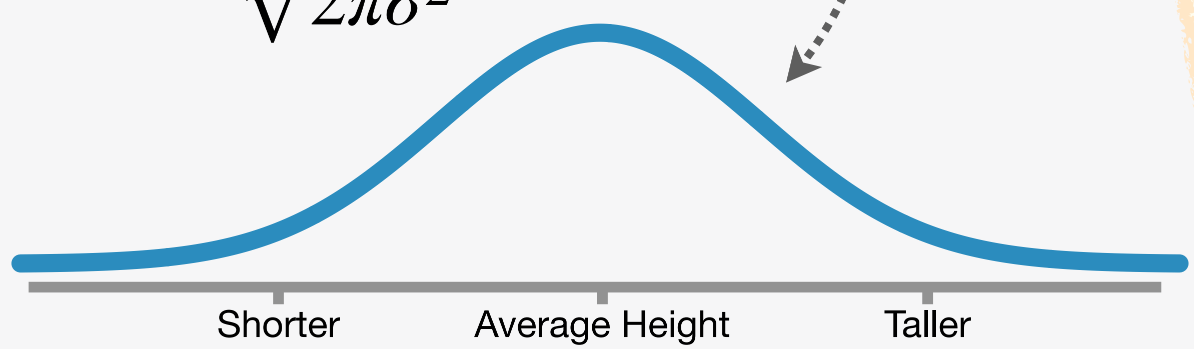


2 …and, additionally, **Continuous Distributions** also spare us from having to decide how to bin the data.



3 Instead, **Continuous Distributions** use equations that represent smooth curves and can provide likelihoods and probabilities for all possible measurements.

$$f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$



4 Like **Discrete Distributions**, there are **Continuous Distributions** for all kinds of data, like the values we get from measuring people's height or timing how long it takes you to read this page.

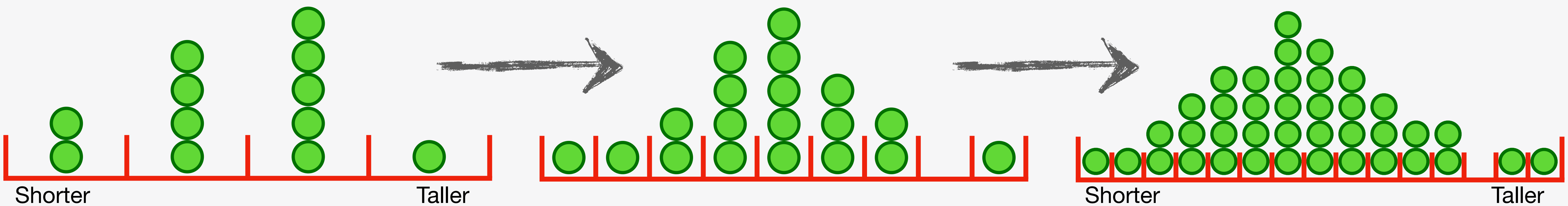
In the context of machine learning, both types of distributions allow us to create **Models** that can predict what will happen next.

So, let's talk about what **Models** are and how to use them.

(small but mighty) **BAM!!!**

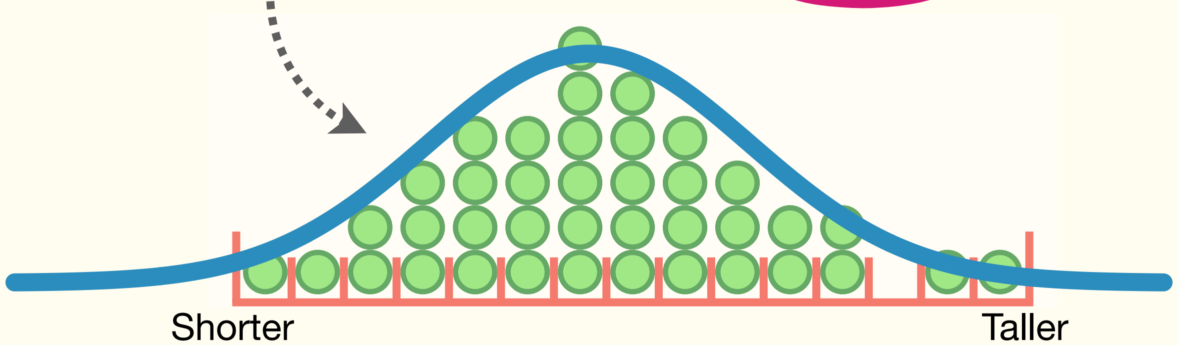
Models: Main Ideas Part 1

1 **The Problem:** Although we could spend a lot of time and money to build a precise histogram… …collecting *all* of the data in the world is usually impossible.

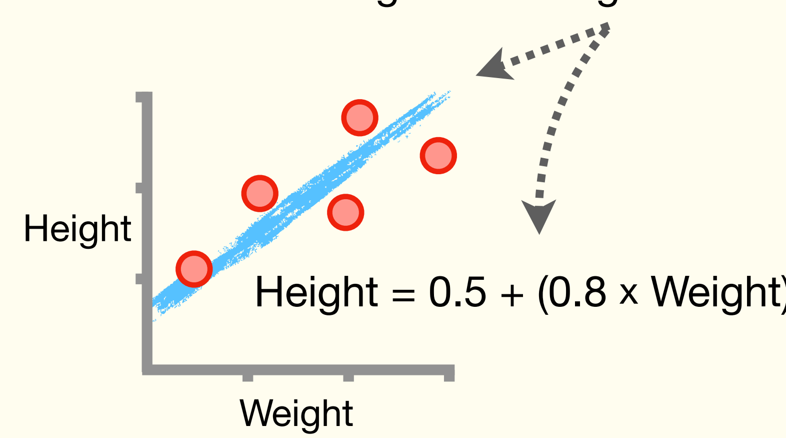


2 **A Solution:** A statistical, mathematical, or machine learning **Model** provides an *approximation* of reality that we can use in a wide variety of ways.

A **Probability Distribution** is a type of model that approximates a histogram with an infinite amount of data.

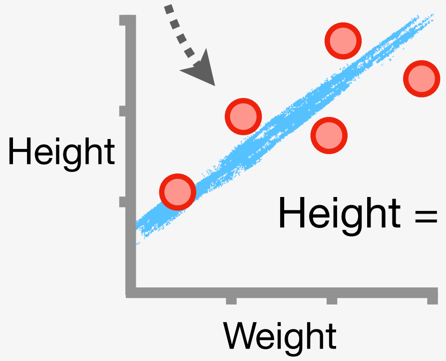


Another commonly used model is the equation for a straight line. Here, we're using a **blue line** to model a relationship between Weight and Height.



Models: Main Ideas Part 2

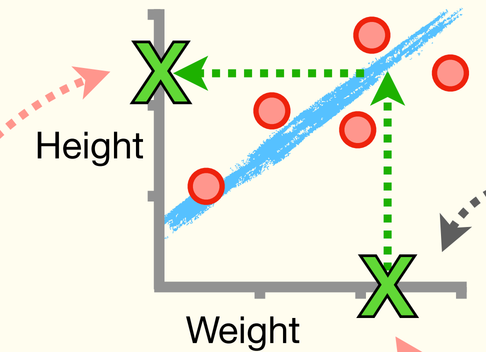
3 As we saw in **Chapter 1**, models need **Training Data**. Using machine learning lingo, we say that we *build models by training machine learning algorithms*.



Height = 0.5 + (0.8 x Weight)

4 Models, or equations, can tell us about people we haven't measured yet.

For example, if we want to know how tall someone is who weighs this much…



…we plug the Weight into the equation and solve for Height…

Height = 0.5 + (0.8 x Weight)

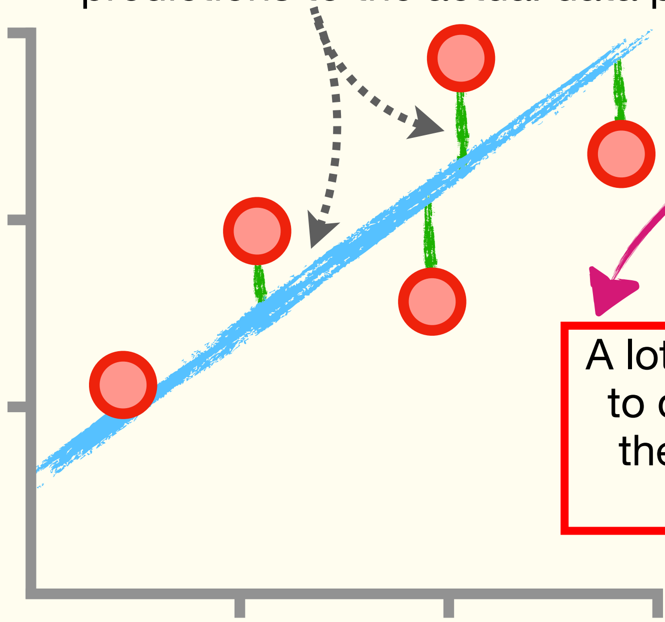
Height = 0.5 + (0.8 x 2.1)

Height = 2.18

…and get **2.18**.

5 Because models are only approximations, it's important that we're able to measure the quality of their predictions.

These **green lines** show the distances from the model's predictions to the actual data points.



A lot of statistics is dedicated to quantifying the quality of the predictions made by a model.

6 In summary:

1) Models approximate reality to let us explore relationships and make predictions.

2) In machine learning, we build models by training machine learning algorithms with **Training Data**.

3) Statistics can be used to determine if a model is useful or believable.

Bam!

Now let's talk about how statistics can quantify the quality of a model. The first step is to learn about the **Sum of the Squared Residuals**, which is something we'll use throughout this book.