



# **The Scientific Method**

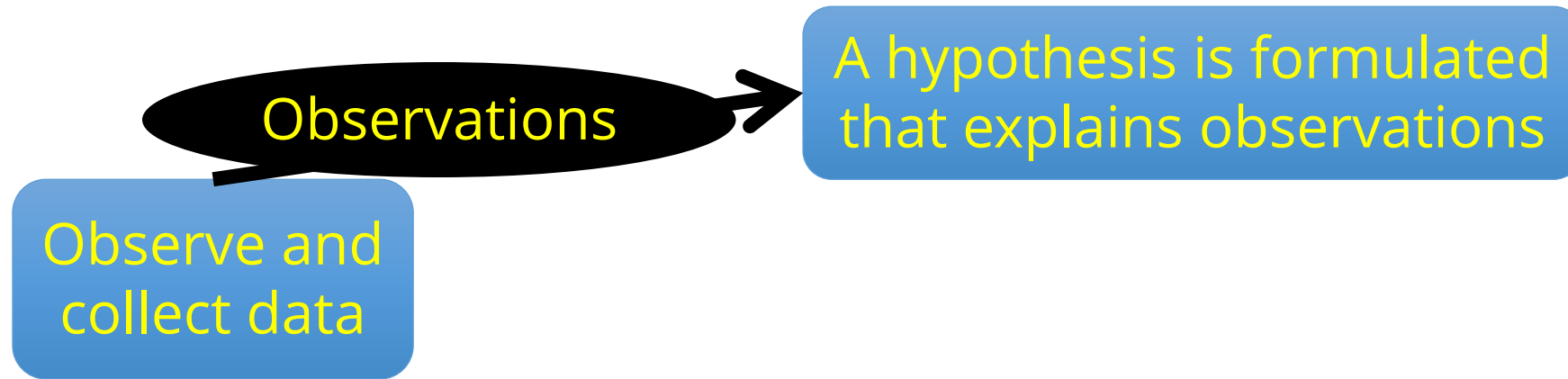
## The Science in Data Science



# The Scientific Method

Observe and  
collect data

# The Scientific Method



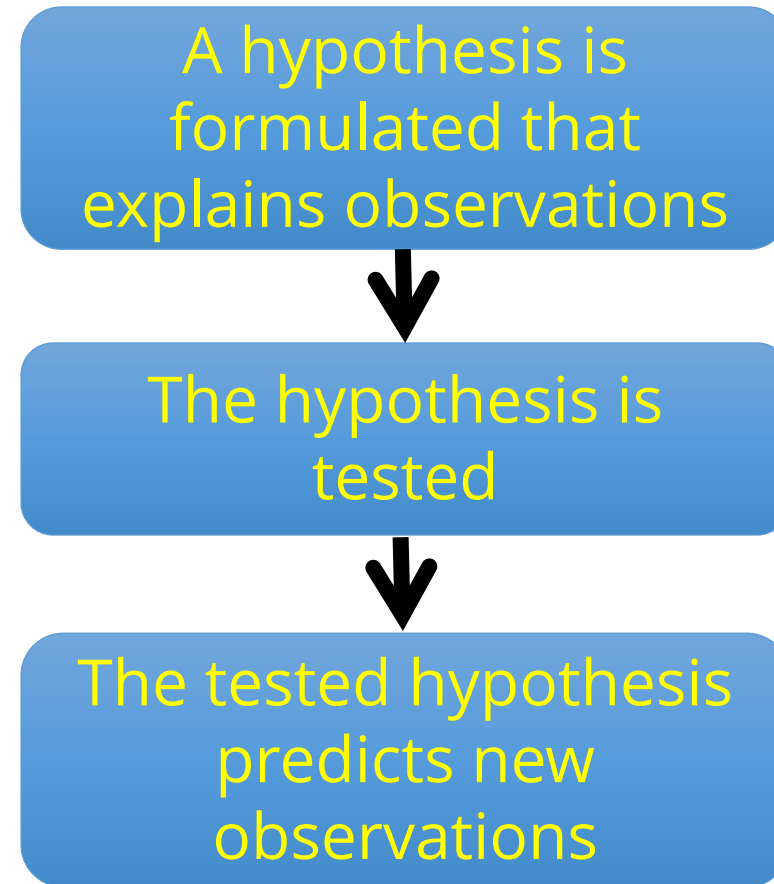
# The Scientific Method

A hypothesis is  
formulated that  
explains observations

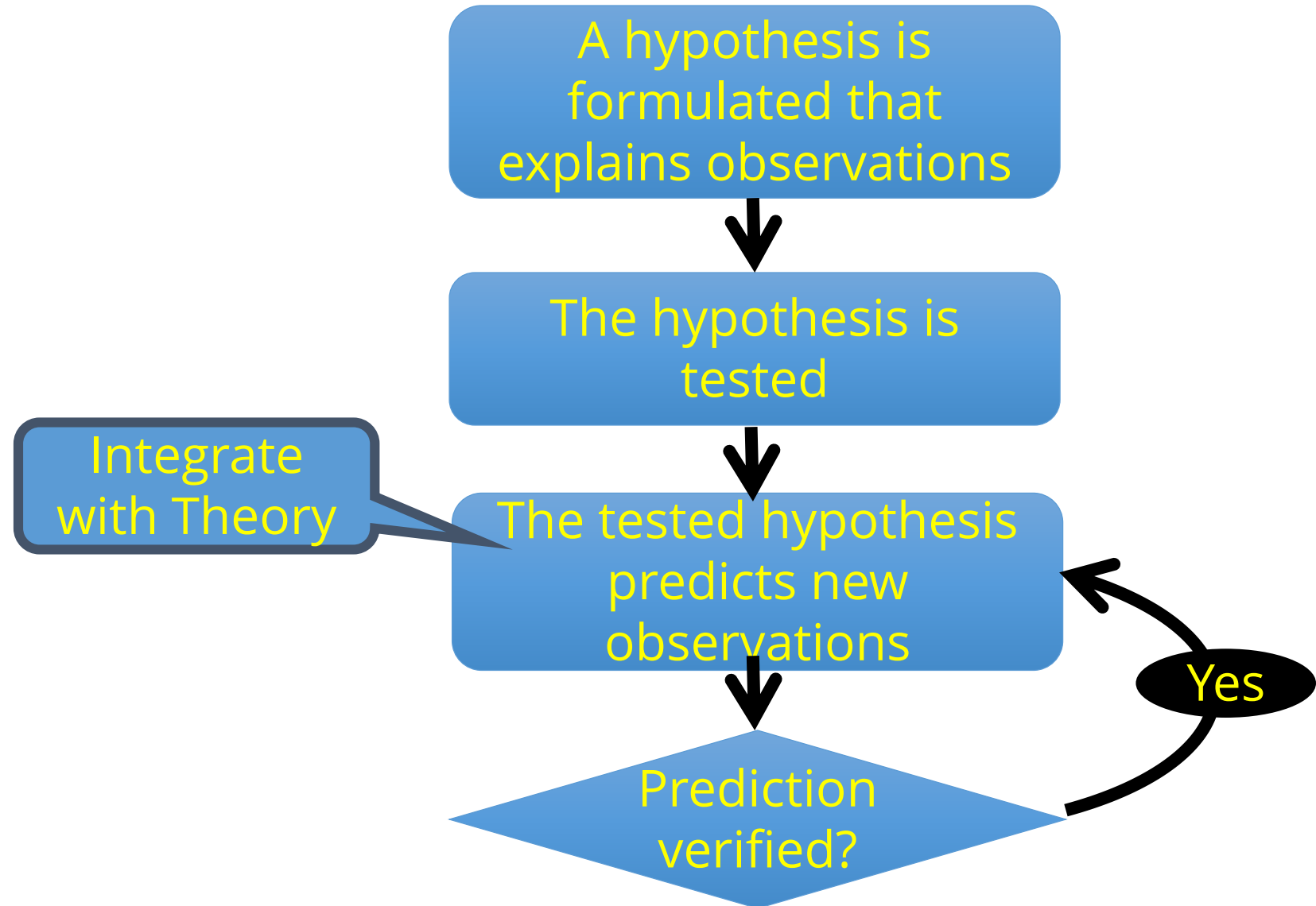


The hypothesis is  
tested

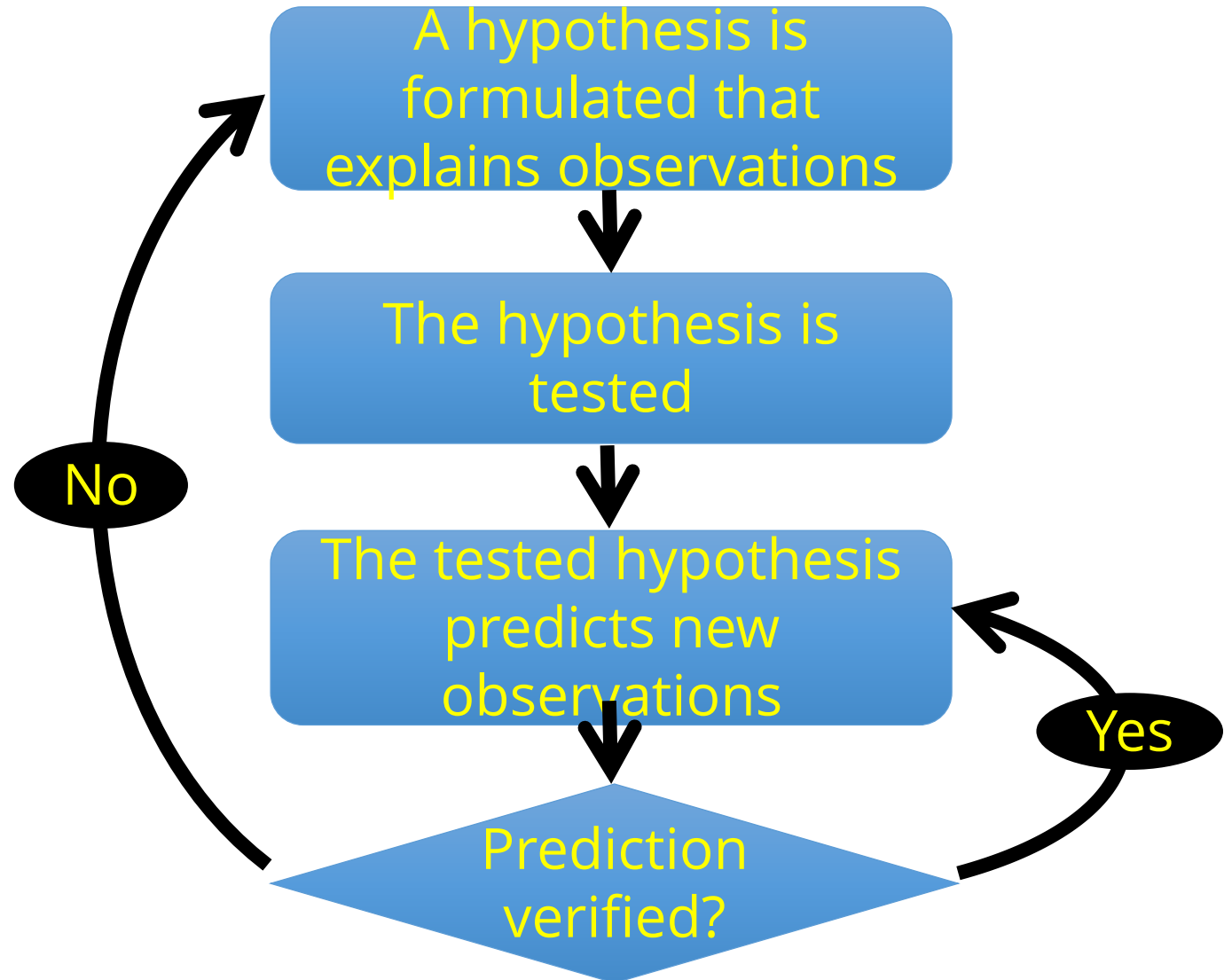
# The Scientific Method



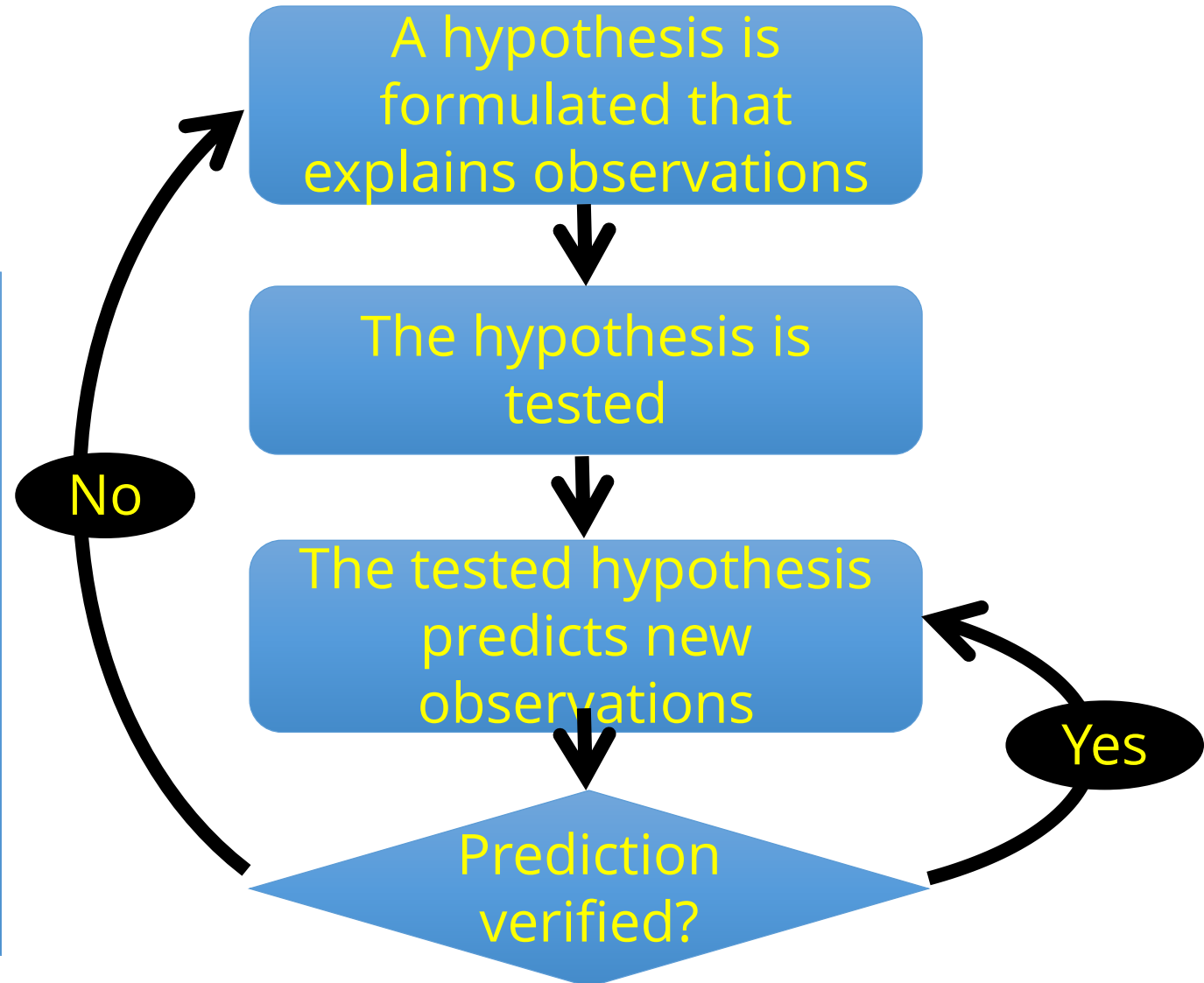
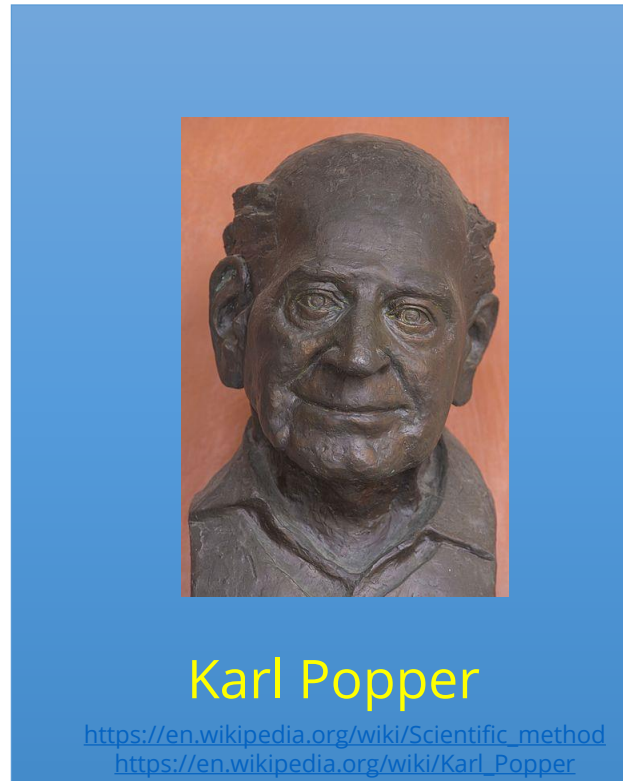
# The Scientific Method



# The Scientific Method



# The Scientific Method



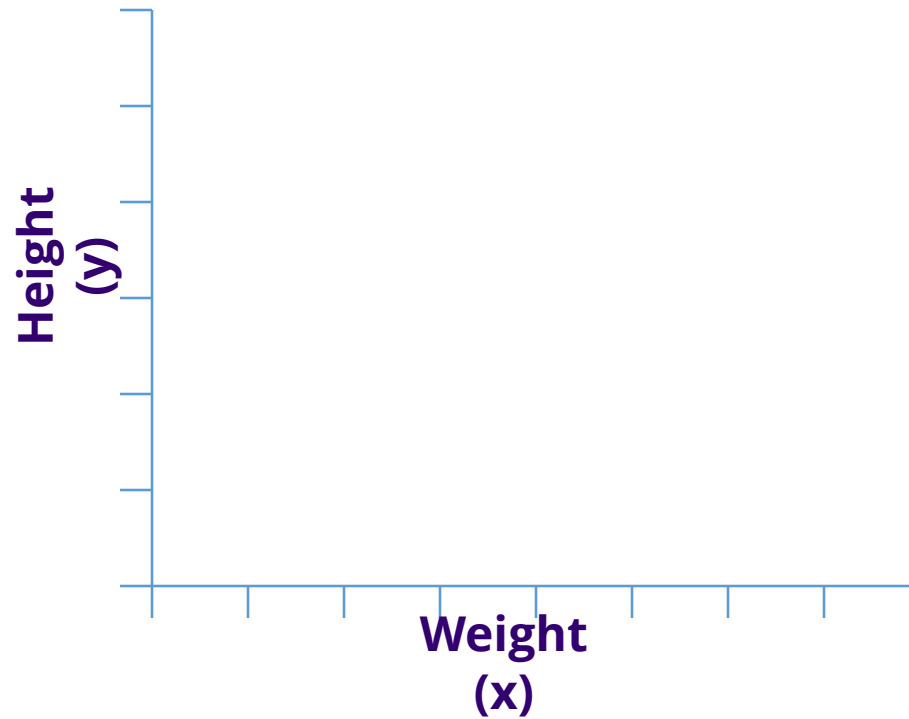


# The Scientific Method

**Model as a  
Hypothesis**

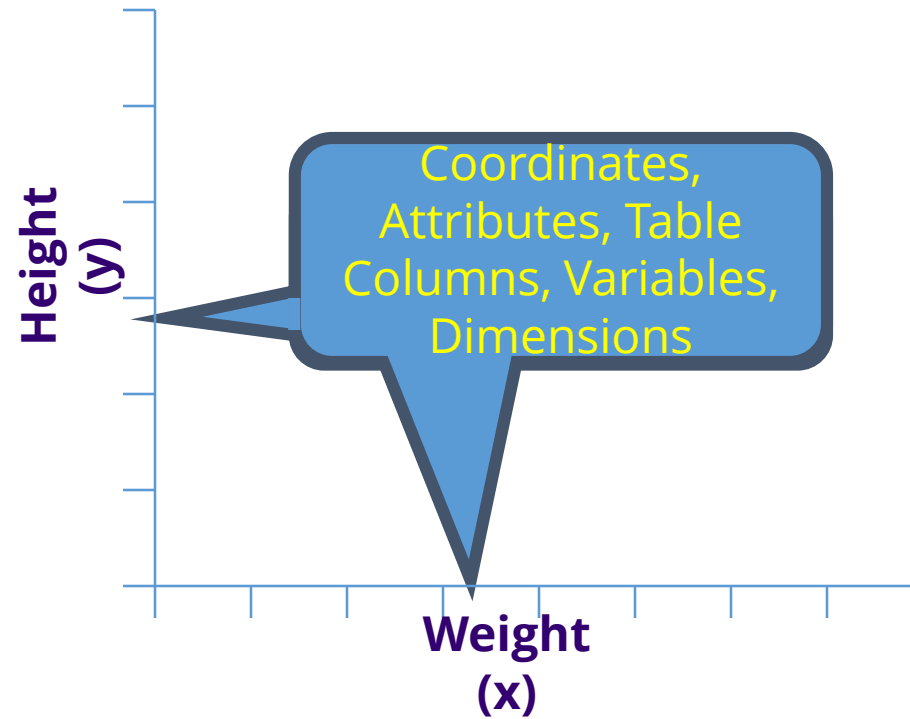
# The Scientific Method

## Model as a Hypothesis



# The Scientific Method

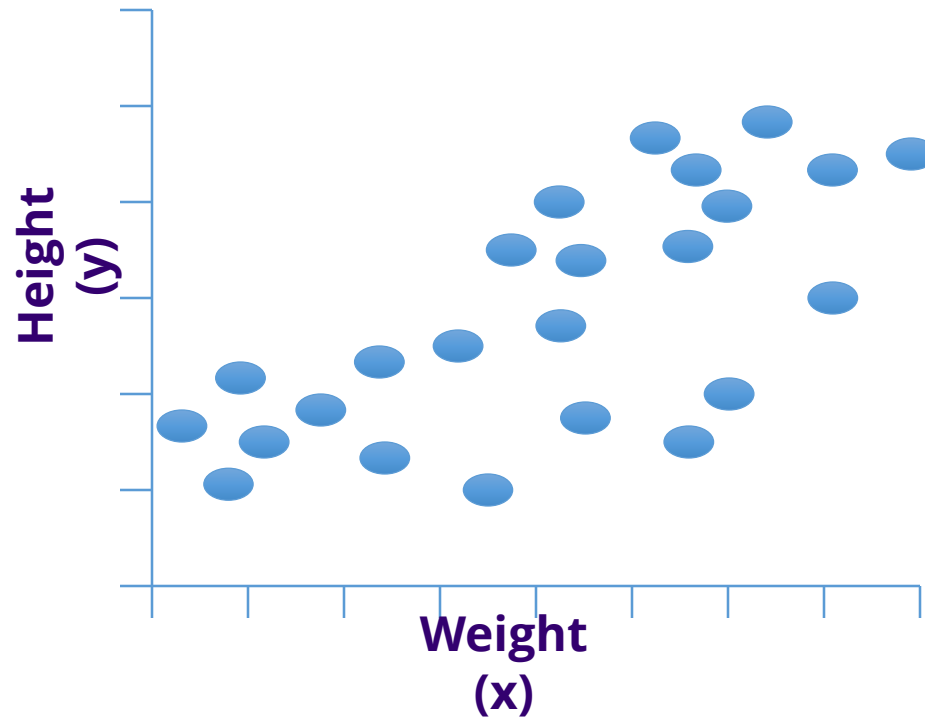
## Model as a Hypothesis



Schema, Space;

# The Scientific Method

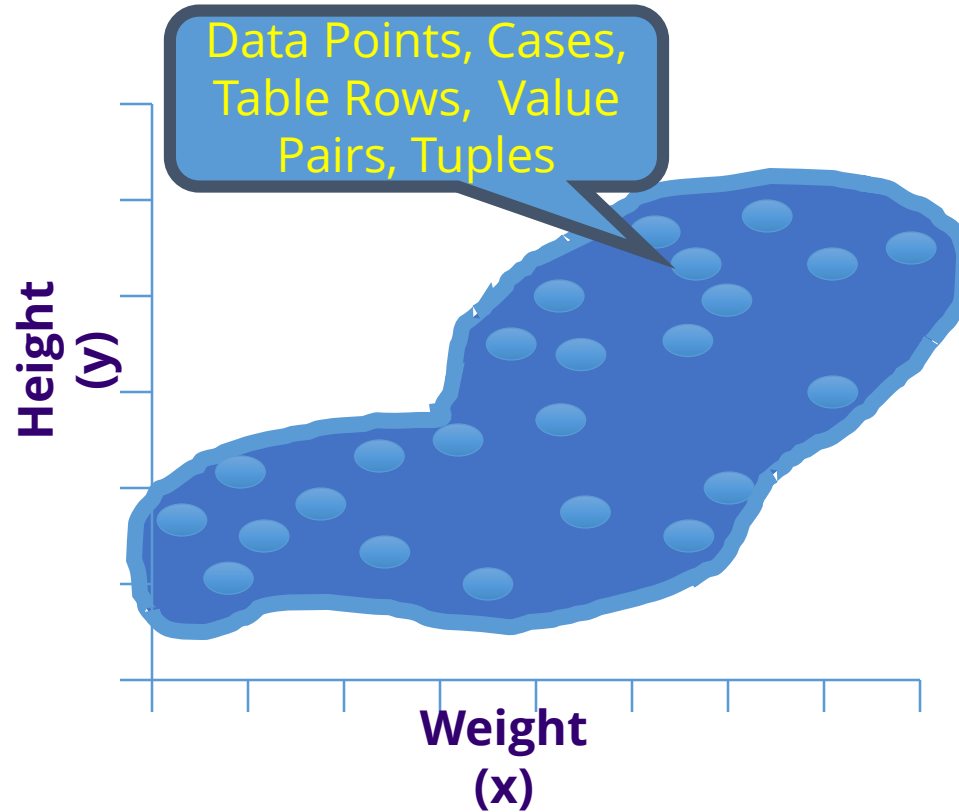
## Model as a Hypothesis



Schema, Space;  
This is a two dimensional  
space of weight and height

# The Scientific Method

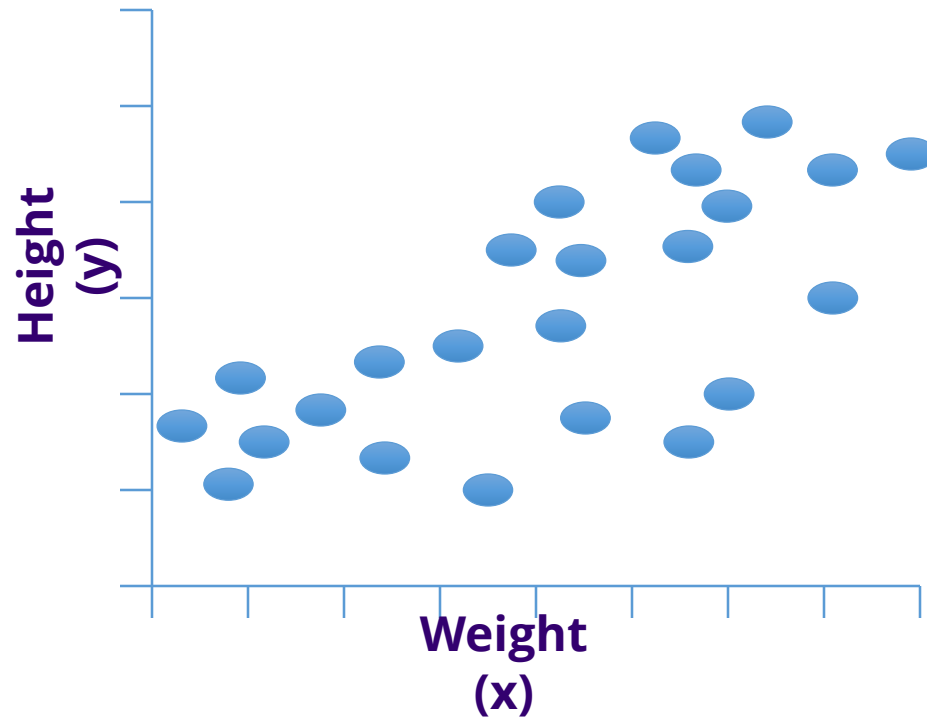
## Model as a Hypothesis



Schema, Space;  
This is a two dimensional  
space of weight and height

# The Scientific Method

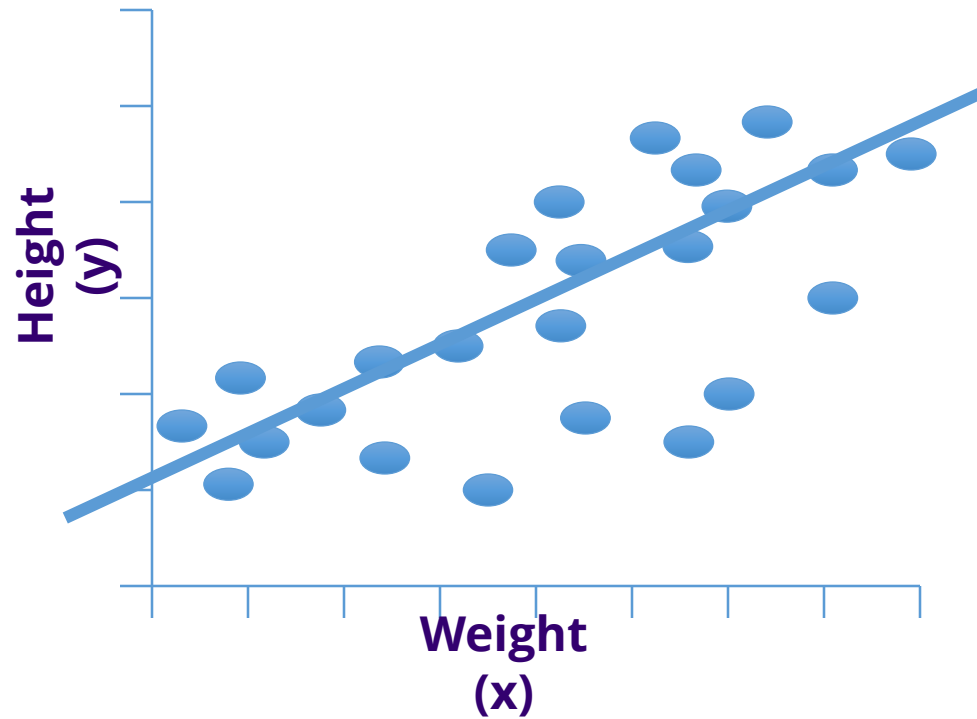
## Model as a Hypothesis



Use these data to create a model.

# The Scientific Method

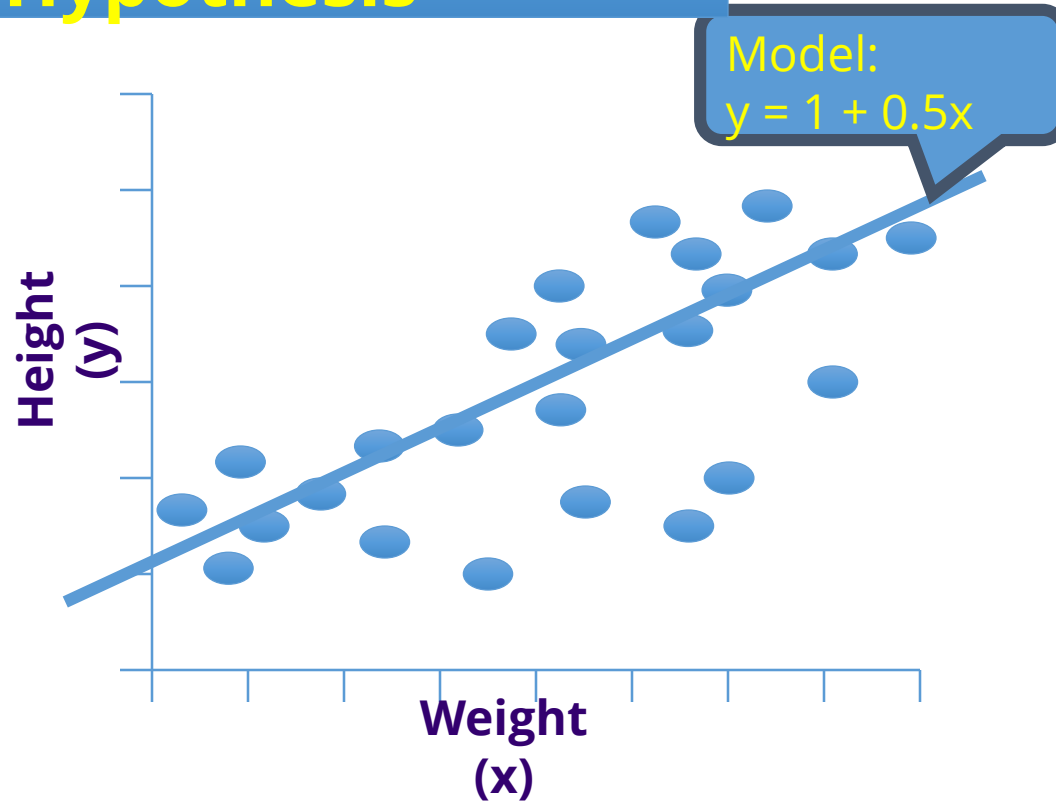
## Model as a Hypothesis



The thick line is a model that represents the ideal data

# The Scientific Method

## Model as a Hypothesis

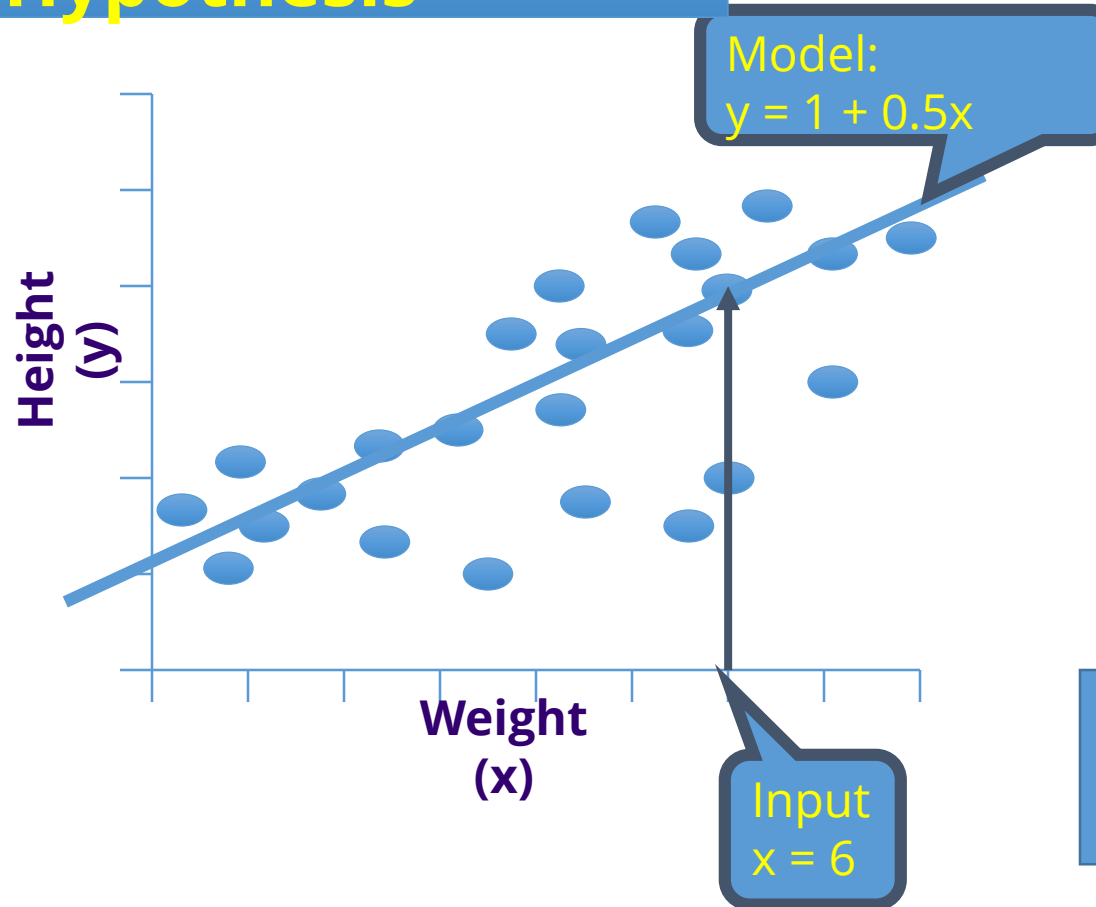


The line can be formalized.  
The model is the formalized  
statement.



# The Scientific Method

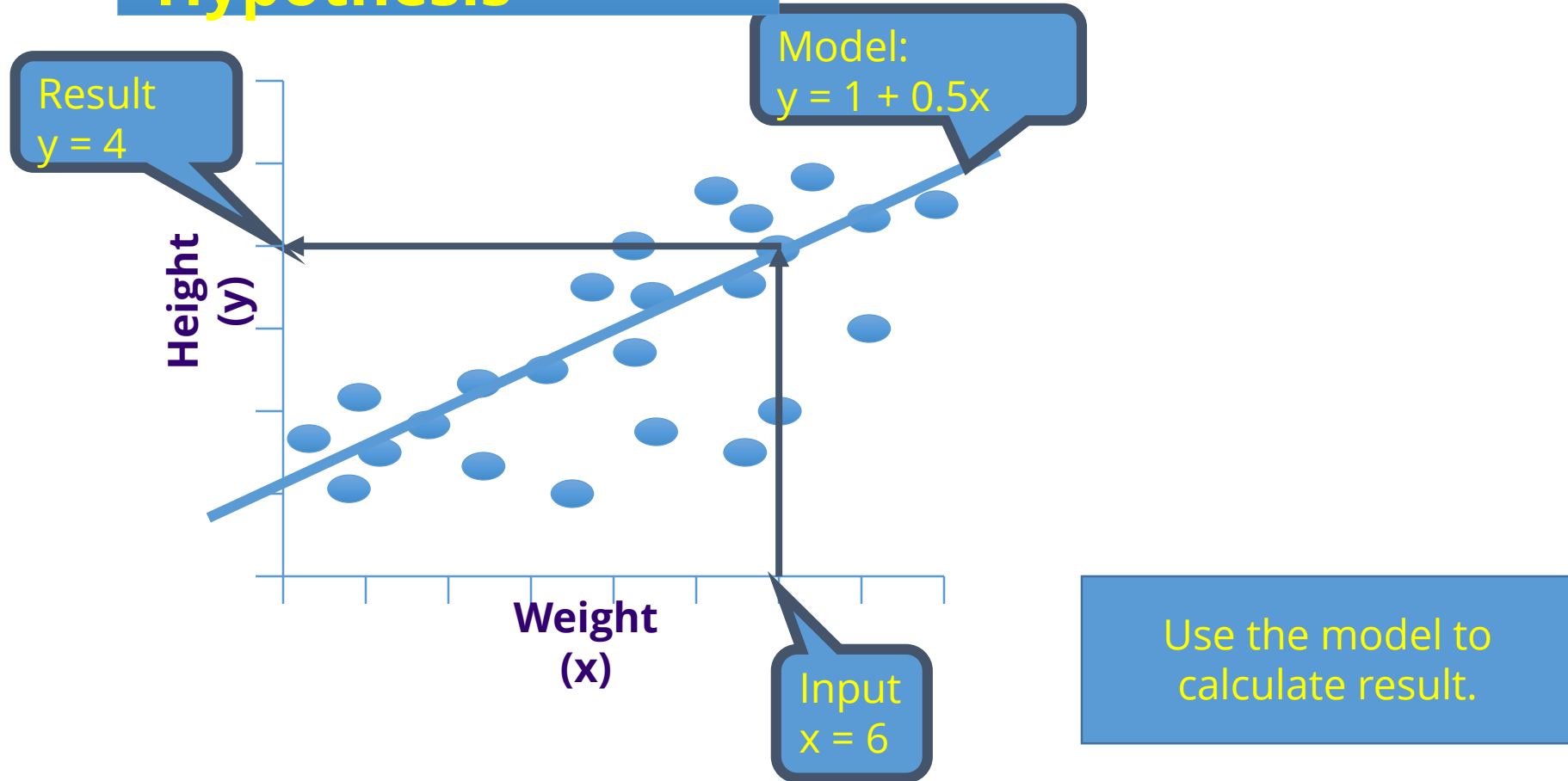
## Model as a Hypothesis



The model requires an input.

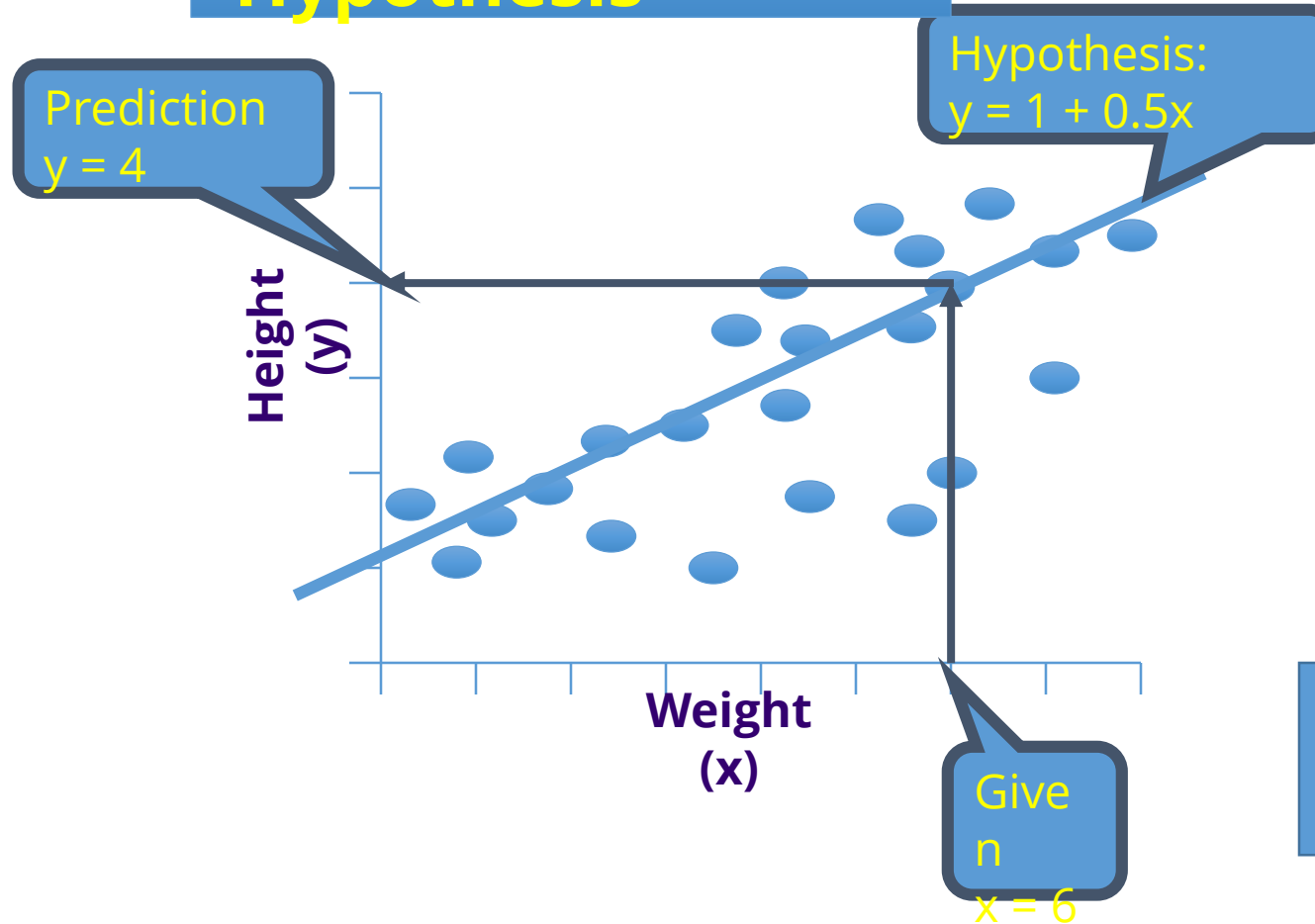
# The Scientific Method

## Model as a Hypothesis



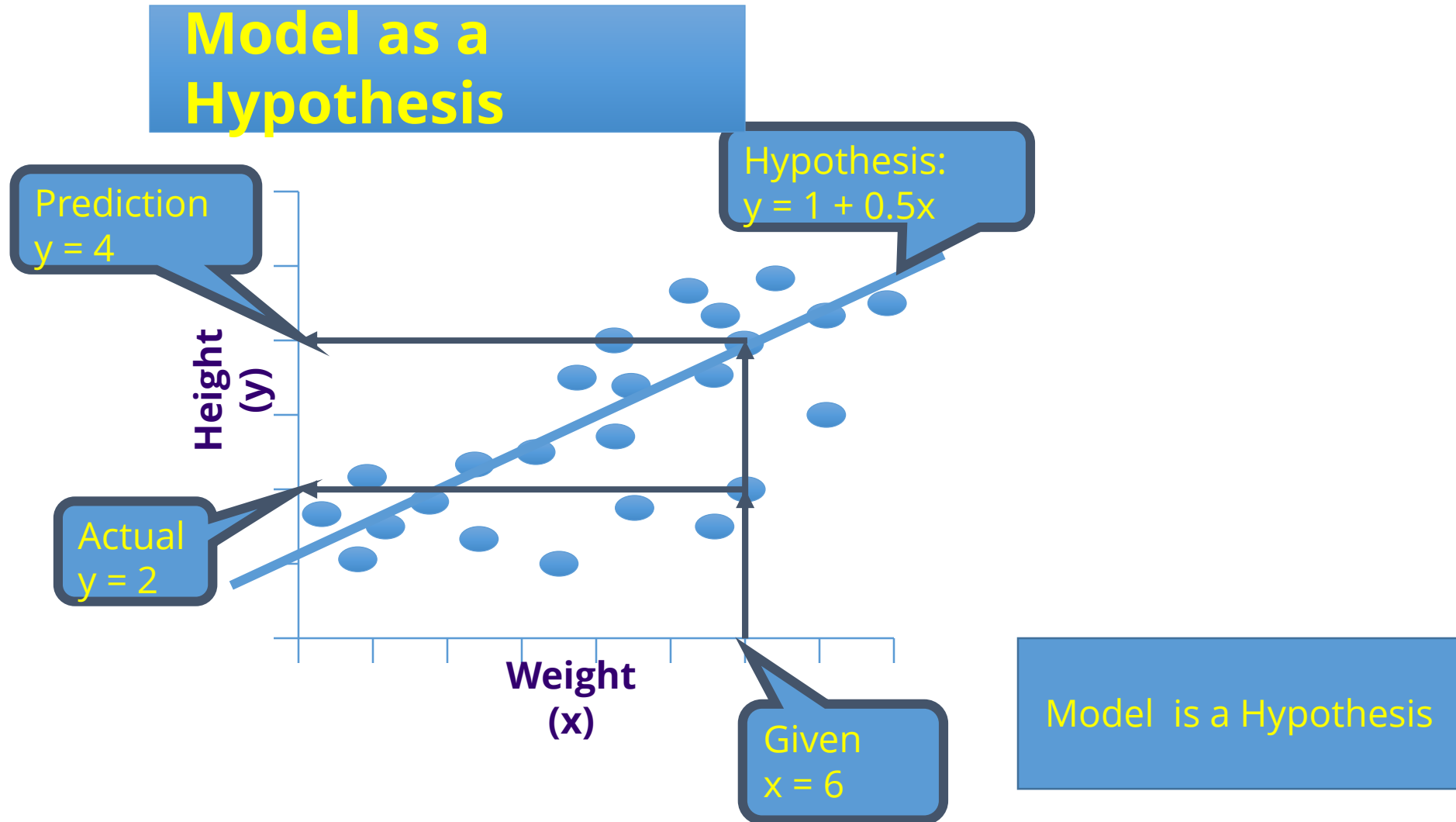
# The Scientific Method

## Model as a Hypothesis



The hypothesis makes a prediction.

# The Scientific Method



# The Scientific Method

## Traditional Science

Examples of data in Traditional Sciences:

- Planetary motions
- Reaction Rate of Chemicals
- Genome Sequence

Examples of models in Traditional Sciences:

- $F = m \cdot a$
- Molecular Models
- Breast Cancer Probability

## Data Science

Examples of data in Data Science:

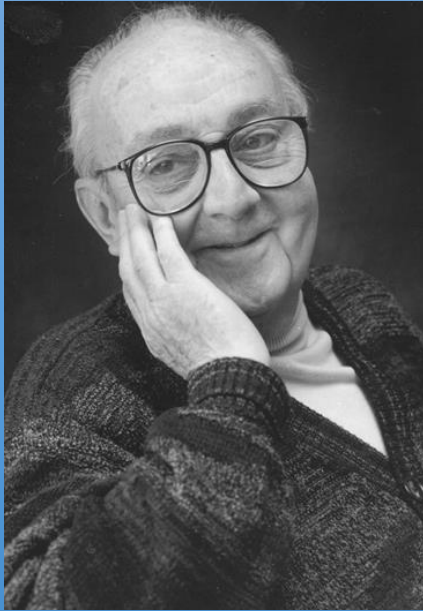
- Twitter Feeds
- Emails
- Patient Records

Examples of models in Data Science:

- Movie Sentiment
- Spam Filter
- Hospital Readmission

# The Scientific Method

"All models are wrong but some are useful"



George Box

[https://en.wikipedia.org/wiki/George\\_E.\\_P.\\_Box](https://en.wikipedia.org/wiki/George_E._P._Box)

# The Scientific Method

## Statistical Modeling: The Two Cultures

[https://projecteuclid.org/download/pdf\\_1/euclid.ss/1009213726](https://projecteuclid.org/download/pdf_1/euclid.ss/1009213726)



Leo Breiman

[https://en.wikipedia.org/wiki/Leo\\_Breiman](https://en.wikipedia.org/wiki/Leo_Breiman)

# The Scientific Method

"Predictive accuracy on test sets is the criterion for how good the model is"

[https://projecteuclid.org/download/pdf\\_1/euclid.ss/1009213726](https://projecteuclid.org/download/pdf_1/euclid.ss/1009213726)



Leo Breiman

[https://en.wikipedia.org/wiki/Leo\\_Breiman](https://en.wikipedia.org/wiki/Leo_Breiman)



# Summary: Some Terminology

---

## >Scientific Method

- Body of techniques

## >Model

- Based on data & method

## >Hypothesis

- Explanation of a dataset

## >Falsification

- Process to disprove hypothesis

## >Theory

- Well-tested hypothesis

## >Karl Popper

- Predictive hypotheses

## >George Box

- Some models are useful

## >Leo Breiman

- 2 cultures in statistics