

# INTRODUCTION TO MODELING

---

## LEARNING OBJECTIVES

---

# TODAY'S LEARNING OBJECTIVES

At the end of today's lesson, you will be able to:

- Explain what a model is and how models are used
- Identify different types of models
- Understand, implement, and explain a simple linear regression model

---

## INTRODUCTION

---

# Modeling Basics

---

# MODELING BASICS

---

- **Modeling** is something humans do naturally.
- We **collect data** more or less constantly.
  - We **abstract and simplify** that data.
    - And we (hopefully!) **update** our models with new information.

---

# MODELING BASICS

---

“Essentially, all models are wrong, but some are useful.”

- George Box, 1987

---

## MODELING BASICS

---

“Because, if this was not the case ... human beings, faced with the daily wondrousness of everything, would go around wearing big, stupid grins ... and say ‘Wow!’ a lot.”

- Terry Pratchett  
*Small Gods*

---

## WHY DO DATA SCIENTISTS MODEL?

---

- We model to make predictions
  - How long will it take me to get to class?
  - How much money should I expect to make as a new Data Scientist?
- We model to make inferences
  - What is the effect of temperature on traffic?
  - How does education affect salary?

---

## TWO KINDS OF MODELING

---

- **Computational modeling** is the process of writing simulations to estimate or predict how a system will behave.
- **Machine learning modeling** applies an algorithm to data in order to ‘teach’ a computer to do something without writing that something explicitly.
- We do the second kind in this course; but it’s good to remember that the first kind exists because it’s really cool.



---

## TERMINOLOGY

---

- **X** the independent/explanatory variables we use to predict
- **Y** the dependent/target variable we want to predict
- $\hat{Y}$** : (pronounced 'y-hat') is our predicted value of **Y**

---

# GOALS

---

- Model the relationship between **X** and **Y**
- Build a model that gets **Y** as close as possible to **y-hat**
- Use our observed values of **X** and a model to generate **y-hat**