

Model Basics

READING:

0.1

EXERCISES:

ALL CHAPTER 0

ASSIGNED:

HW 1

PRODUCER:

DR. MARIO



IMG CREDIT: ALEX RIEGL

Course Website / Syllabus

- Access Course Website Through Canvas
- Cover Syllabus
 - Office Hours
 - Grading and Curving
 - Attendance: UNC Check-In App
 - Homework
 - Quizzes
 - Exams
 - PDFs and Gradescope
 - Grade Disputes
 - Honor Code
- Usage of Course Website and Canvas

Preview of Dataset

LEGO Dataset (n=1304)

| Set_Name | Theme | Pieces | Price | Amazon_Price | Year | Ages | Minifigures | Unique_Pieces |
|-------------------------|----------------|--------|-------|--------------|-------|------------|-------------|---------------|
| <chr> | <chr> | <dbl> | <dbl> | <dbl> | <dbl> | <chr> | <dbl> | <dbl> |
| Imperial Star Destroyer | Star Wars | 4784 | 700. | 700. | 2019 | Ages_16+ | 2 | 445 |
| Betrayal at Cloud City | Star Wars | 2812 | 350. | 668. | 2018 | Ages_14+ | 19 | 676 |
| Liebherr R 9800 | Powered UP | 4108 | 450. | 443. | 2019 | Ages_12+ | NA | 221 |
| NINJAGO City Docks | NINJAGO | 3553 | 230. | 440 | 2018 | Ages_12+ | 14 | 690 |
| Hogwarts Castle | Harry Potter | 6020 | 400. | 400. | 2018 | Ages_16+ | 28 | 624 |
| Voltron | Ideas | 2321 | 180. | 389. | 2018 | Ages_16+ | NA | 471 |
| Roller Coaster | Creator Expert | 4124 | 380. | 380. | 2018 | Ages_16+ | 11 | 556 |
| Bugatti Chiron | Technic | 3599 | 350. | 340. | 2018 | Ages_16+ | NA | 306 |
| Mack Anthem | Technic | 2595 | 180. | 330 | 2018 | Ages_11-16 | NA | 253 |
| Farm Adventures | DUPLO | 104 | 60.0 | 300. | 2018 | Ages_2-5 | 3 | 42 |

Any interesting questions about LEGO we may want to answer?

Models Help Us...

- Answer Questions
- Make Predictions or Classifications
- Evaluate Treatments or Test Theories
- Understand Relationships

Architecture of a Model

$$Y = f(X) + \varepsilon$$

Response Variable

Explanatory Variable(s)

Error or Deviation from the Model

Generates an Expectation about Y given X

The diagram illustrates the architecture of a model using the equation $Y = f(X) + \varepsilon$. The components are annotated as follows:

- Y is labeled as the "Response Variable" with an arrow pointing to it from the left.
- $f(X)$ is labeled as "Explanatory Variable(s)" with an arrow pointing down to X . A bracket underneath $f(X)$ points to the text "Generates an Expectation about Y given X" at the bottom.
- ε is labeled as "Error or Deviation from the Model" with an arrow pointing to it from the right.

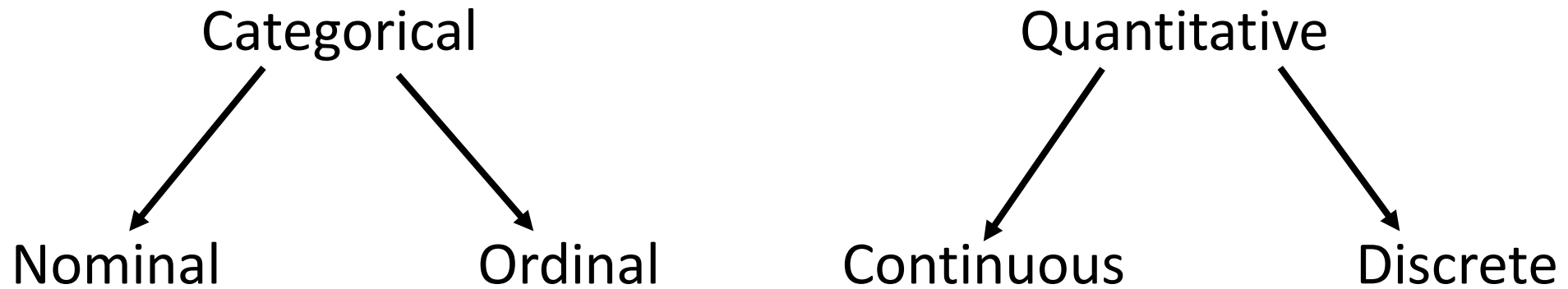
Statistical Modeling

Statistical Modeling is the Process of ...

Defining the Function $f(X)$ and then **Fitting that Function $f(X)$** to a sample dataset by **Minimizing the Error** the best we possibly can

Methodology we Use Depends on the Types of Variables

Variable Types

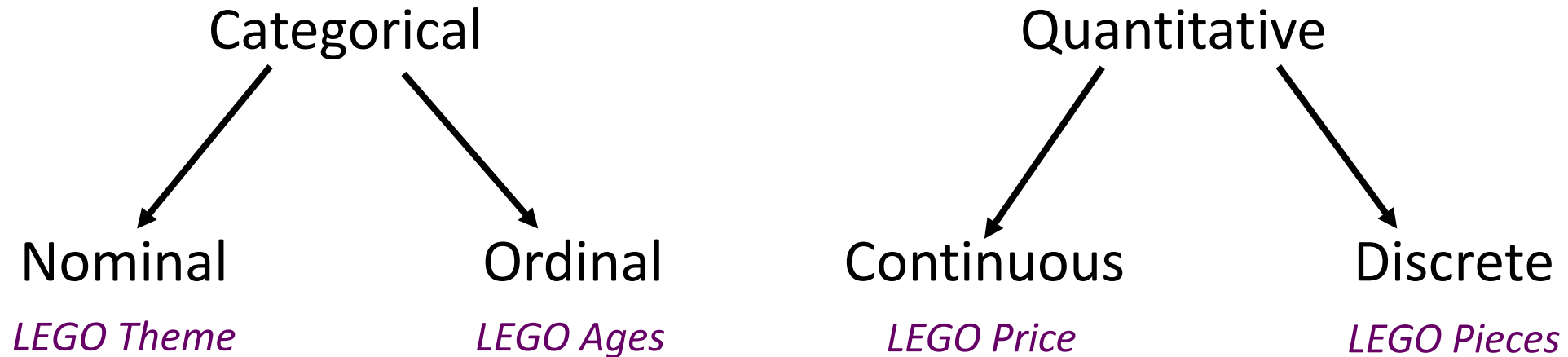


Preview of Dataset

Can we find all the variable types in this dataset?

| Set_Name | Theme | Pieces | Price | Amazon_Price | Year | Ages | Minifigures | Unique_Pieces |
|-------------------------|----------------|--------|-------|--------------|-------|------------|-------------|---------------|
| <chr> | <chr> | <dbl> | <dbl> | <dbl> | <dbl> | <chr> | <dbl> | <dbl> |
| Imperial Star Destroyer | Star Wars | 4784 | 700. | 700. | 2019 | Ages_16+ | 2 | 445 |
| Betrayal at Cloud City | Star Wars | 2812 | 350. | 668. | 2018 | Ages_14+ | 19 | 676 |
| Liebherr R 9800 | Powered UP | 4108 | 450. | 443. | 2019 | Ages_12+ | NA | 221 |
| NINJAGO City Docks | NINJAGO | 3553 | 230. | 440 | 2018 | Ages_12+ | 14 | 690 |
| Hogwarts Castle | Harry Potter | 6020 | 400. | 400. | 2018 | Ages_16+ | 28 | 624 |
| Voltron | Ideas | 2321 | 180. | 389. | 2018 | Ages_16+ | NA | 471 |
| Roller Coaster | Creator Expert | 4124 | 380. | 380. | 2018 | Ages_16+ | 11 | 556 |
| Bugatti Chiron | Technic | 3599 | 350. | 340. | 2018 | Ages_16+ | NA | 306 |
| Mack Anthem | Technic | 2595 | 180. | 330 | 2018 | Ages_11-16 | NA | 253 |
| Farm Adventures | DUPLO | 104 | 60.0 | 300. | 2018 | Ages_2-5 | 3 | 42 |

Example of Variable Types Using LEGO



Families of Models

Response Variable

Predictor Variable

Categorical

Categorical

Quantitative

Quantitative

*Complicated by **multiple** predictor variables and/or response variables*

Technology We Will Use

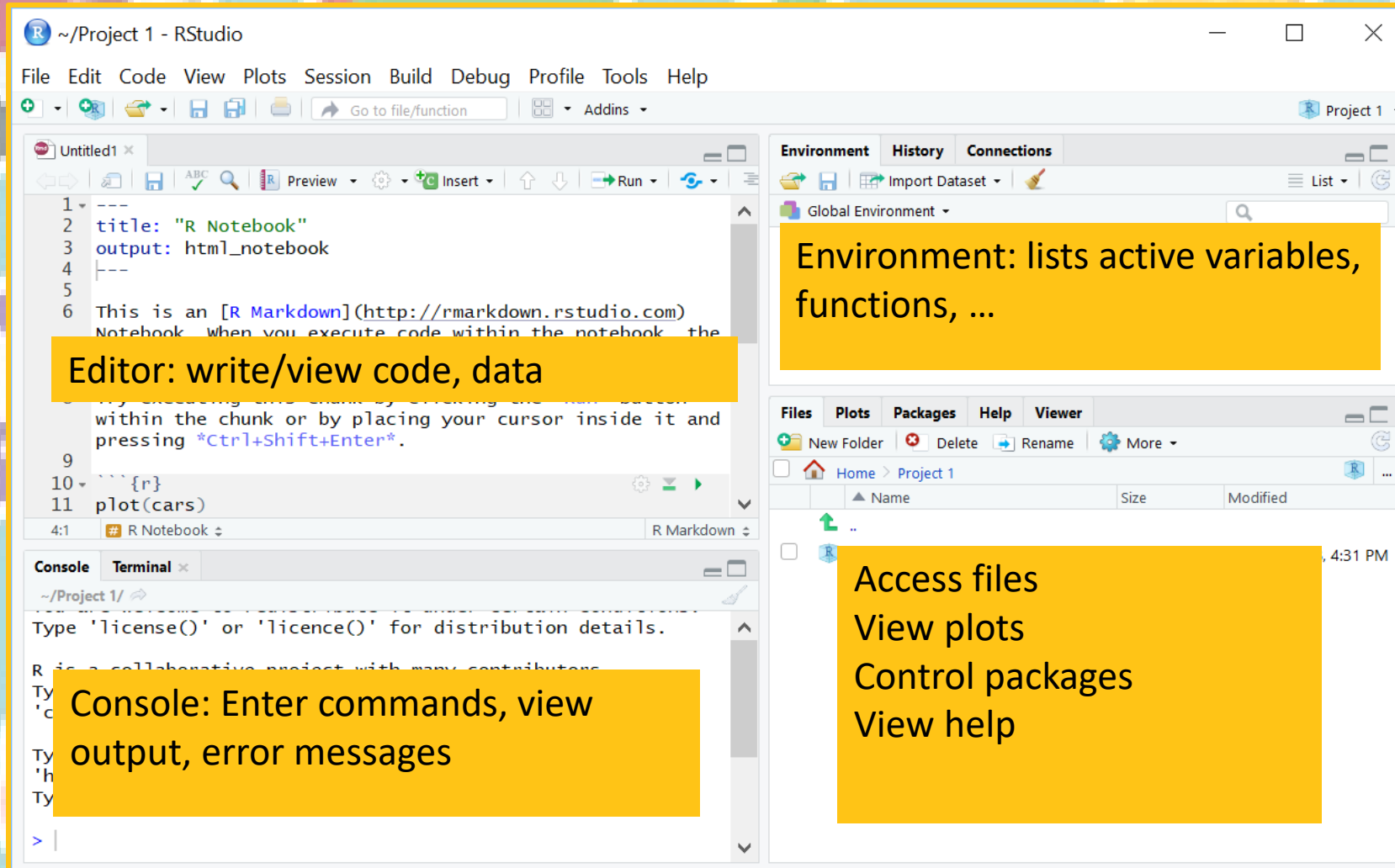
- *R* = a free, widely used, open source, language and environment for statistical computing and graphics
- *RStudio* = an interface for *R* (Integrated Development Environment)
- *RMarkdown* = a tool in *R* for creating documents that combine *R* code with text

Download R and RStudio to Your Computer

Supplement for Lecture 1

- Download Zip Folder on Course Website for Supplement
- Unzip Folder on Your Computer
- Open the Template.rmd File from the Unzipped Folder
- RMD Files Should Automatically Open in RStudio

Quick Look at R Studio



The screenshot shows the R Studio interface with the following components and annotations:

- Editor:** The central area for writing and viewing code and data. It contains a notebook with R Markdown code:

```
1 ---  
2 title: "R Notebook"  
3 output: html_notebook  
4 ---  
5  
6 This is an [R Markdown](http://rmarkdown.rstudio.com)  
7 Notebook. When you execute code within the notebook, the  
8 output is displayed within the chunk or by placing your cursor inside it and  
9 pressing *Ctrl+Shift+Enter*.  
10  
11 {r}  
12 plot(cars)
```
- Environment:** Lists active variables, functions, and data. It shows the "Global Environment" with a search bar.
- Files:** Access files, view plots, control packages, and view help. It shows a file browser with a "Home" button and a list of files.
- Console:** Enter commands, view output, and error messages. It shows the R startup message: "Type 'license()' or 'licence()' for distribution details. R is a collaborative project with many contributors. Type 'c' for copyright, 'l' for license, 'h' for help, 'T' for trademarks, and 'S' for source code. Type 'q' to quit R. >".

Learning Objectives for R

- Install Packages
- Load Packages
- Read CSV Files
- Create Objects
- Columns and Rows
- Subset Data
- Mathematical Calculations
- Statistical Functions

Thank You

Make Reasonable Decisions

