# **Teacher Notes for Cars and Housing Lesson**

## **Motivation and Essential Understandings**

The goal of this lesson is to demonstrate the potential uses of linear regression while highlighting the importance of practicing good data science.

When and how can I use linear regression?

## **Context and Dataset**

Students will examine two datasets: one about cars, and one about the cost of housing. Students will determine whether it is appropriate to apply linear regression to these datasets.

## **Learning Objectives**

* Learn what linear regression is
* Learn how to evaluate where strong correlations exist
* Learn what requirements a dataset must meet in order to use linear regression
* Learn the importance of practicing good data science

## **Data Science Concepts and Skills**

1. Linear Regression – Use and Assumptions
2. Data Wrangling

## **Students**

This lesson is ideal for students who do not have previous exposure to linear regression. A basic understanding of statistics is useful, but not necessary. A knowledge of programming is necessary to complete the bonus assignment.

## **Time to Teach this Lesson**

2 Class Periods

Class 1 – Students Walk Through RNotebook of their Choice (Beginner, Intermediate, Advanced)

Class 2 – Discussion

## **Lesson Materials**

Link to Git Repository: <https://github.com/generationai/Johns_Sociology_2020>

|  |  |  |
| --- | --- | --- |
| **Materials** | **File** | **Description** |
| Beginner R notebook | RNotebook\_Beginner\_R\_Soc\_2020 | R notebook for the beginner lesson |
| Beginner R notebook pdf | RNotebook\_Beginner\_pdf\_Soc\_2020 | pdf of R notebook for the beginner lesson |
| Intermediate R notebook | RNotebook\_Intermediate\_R\_Soc\_2020 | R notebook for the intermediate lesson |
| Intermediate R notebook pdf | RNotebook\_Intermediate\_pdf\_Soc\_2020 | pdf of R notebook for the intermediate lesson |
| Advanced R notebook | RNotebook\_Advanced\_R\_Soc\_2020 | R notebook for the advanced lesson |
| Advanced R notebook pdf | RNotebook\_Advanced\_pdf\_Soc\_2020 | pdf of R notebook for the advanced lesson |
| Dataset | housePractice.csv | Data on the cost of housing in Seattle |
| Lesson template | Lesson\_Template\_Soc\_2020 | Lesson planner |

## **Teaching Strategies**

* Students can choose the level at which they feel comfortable learning
* If you have a good sense of the skill level of your class, you can have all of the students work from the same level notebook

## **Lesson Narrative**

|  |
| --- |
| **Module 1: Introduction to AI** |

**Activity** – Have students work through their selected level of the R notebook. Students can work together or in groups.

**Assignment –** Have students answer the questions in the notebook and turn them in at the end of the class.

|  |
| --- |
| **Module 2: AI Applications and Ethics** |

**Discussion** – Discuss the following questions with the class:

1. What are some potential uses of linear regression in our field?oHow Why is using linear regression beneficial in these examples?
2. Why is it important to practice good data science? Discuss the importance of addressing:
   1. Missing data
   2. Outliers and high leverage points
   3. Linear regression assumptions

**Bonus Assignment** – Students comfortable with programming can change the code in the R notebook to read in a new dataset. Find a dataset relevant to your field, especially one where linear regression is already applied. See if missing data and outliers were addressed, and if the linear regression assumptions were met. If not, what does this say about the validity of the linear regression model?