# **Teacher Notes for Environmental Science (with [which dataset?) \*DRAFT\***

## **Motivation and Essential Understandings**

Many researchers in Ecotoxicology need to be proficient with assessing, interpreting, and communicating information about environmental hazards and their effects on life expectancy. Of particular concern in Texas is the plethora of superfund sites.

* Is age, gender, and/or population density predictors of life expectancy?
* Does living near an EPA Superfund site affect life expectancy?

## **Context and Dataset**

Students will interpret visualizations using data from two demographic datasets: Superfund Sites (published by EPS) and Life Expectancy at Birth (published by Texas Department of State Health Services (DSHS)). Students will explore correlations of various features regarding life expectancy in the state of Texas.

## **Learning Objectives**

Students will be able to:

## Explore predictors for life expectancy in Texas

## **Data Science Concepts and Skills**

1. Summary statistics
2. Exploratory data analysis; Static data visualization
3. Data wrangling
4. Hypothesis testing
5. Predictive modeling with Decision tree; Logistic Regression

## **Students**

This lesson is for late undergraduate students. Students should be familiar with statistical concepts, basic data visualizations, and have worked in Excel.

## **Time to Teach this Lesson**

This lesson can be taught in 2 sessions using a worksheet

**First Week**: 1-hour prep, 2-hour class session; guided data analysis and review of Statistics

**Second Week**: 1-hour class session to review decision trees; guided & independent work

## **Lesson Materials**

You will find all the lesson materials in the GenAI GitHub repository. The R notebook is not necessary to teach this lesson but is available to those who wish to teach more hands-on Data Science.

|  |  |  |
| --- | --- | --- |
| **Materials** | **File** | **Description** |
| Lecture |  | PPT lecture |
| Handout 1 | Handout\_LEinTX\_EnSci\_2020.pdf | Life Expectancy in Texas Report |
| Handout 2 |  |  |
| Worksheet |  | Lesson worksheet |
| Dataset 1 | tx\_count\_life\_data1 | Life Expectancy |
| Dataset 2 | tx\_epa\_data2 | EPA Superfund sites |
| Data dictionary | data\_dictionary | pdf of data dictionary explaining the column headings (data fields) in the datasets |
| R notebook | RNotebook\_R\_EnvSci\_2020 | R notebook |
| R notebook pdf | RNotebook\_pdf\_EnvSci\_2020 | pdf of annotated R notebook |
| Template | Lesson\_Template\_EnvSci\_2020.docx | Lesson planner \*DRAFT |

## **Teaching Strategies**

* Review concepts in Statistics such as descriptive stats, correlation, and hypothesis testing
* Pose **challenge questions** for engagement and allow students to interpret visualizations and hypothesize. Students may have difficulty limiting inferences to within the scope of the dataset, so discuss over-hypothesizing beyond the data.
* Discuss the difference between Inferential vs. Predictive modeling

## **Lesson Narrative**

|  |
| --- |
| **Module 0: Pre-lesson** |

Review concepts of descriptive Statistics, correlation, and hypothesis testing.

|  |
| --- |
| **Module 1: Dataset and Exploratory Data Analysis** |

This lesson narrative follows the lesson worksheet. Interpretive questions are posed for each step.

Illustrate **Descriptive Stats**: erroneous data; distributions; correlations between features

|  |
| --- |
| **Module 2: Explain Predictive Modeling** |

Present and demonstrate two models for prediction.

* **Logistic Regression:** Confusion matrix; binary dependent variable (yes/no?)
* **Decision tree**: Demonstrate over-correlation – binary variable for diabetes prediction; Confusion matrix: Precision, Accuracy, Recall

|  |
| --- |
| **ASSIGNMENT:** Ask students about feature importance and how to identify predictors. Students will discuss and write up conclusions. |

|  |
| --- |
| **Module 3: Close Out** |

Post-assessment questions.

What are some of the limitations of using this dataset?

What risks are inherent in making predictions using multi-sourced datasets?