



# Rensselaer

why not change the world?®

## Lab2: Linear Regression

**Ahmed Eleish**

**February 5th, 2024**

**ITWS-4600/ITWS-6600/MATP-4450/CSCI-4960**

Tetherless World Constellation  
Rensselaer Polytechnic Institute



# Lab 02

Files:

<https://rpi.box.com/s/iby6k1sdugnmixd63uxrrgkm3dyro7y>



# Exercise 1: linear models

- Your exercise: Examine the influence of various variables on property price
- Dataset: NY-House-Dataset.csv



# Using the code in NYHousing\_Linear\_Regression.R do the following:

- Fit 3 linear models with Price as the response variable and combinations of PropertySqFt, Beds, and Bath as predictors
- Do any data cleaning (outliers, etc.) to get the best possible models. Remember that you may need to transform a variable for a better fit
- For each model print the model summary stats and plot the most significant variable vs Price with the best fit line as well as a scatter plot of the residuals for the model.
- Compare the 3 models and decide which you believe is most useful. You don't need to tell me which one.



Please push to your github repository:

1. All your code in a \*.R or \*.MD file
2. Summary stats and select plots from 3 linear models



Thanks!  
Have a great weekend!