**Final Project**

**Machine Learning**

This “Pokemon” set includes 721 Pokemon, including their number, name, first and second type, and basic stats: HP, Attack, Defense, Special Attack, Special Defense, and Speed. It has been of great use when teaching statistics to kids. With certain types you can also give a geeky introduction to machine learning.

This are the raw attributes that are used for calculating how much damage an attack will do in the games. This dataset is about the pokemon games (*NOT* pokemon cards or Pokemon Go).

The data as described by:

* **#**: ID for each pokemon
* **Name**: Name of each pokemon
* **Type 1**: Each pokemon has a type, this determines weakness/resistance to attacks
* **Type 2**: Some pokemon are dual type and have 2
* **Total**: sum of all stats that come after this, a general guide to how strong a pokemon is
* **HP**: hit points, or health, defines how much damage a pokemon can withstand before fainting
* **Attack**: the base modifier for normal attacks (eg. Scratch, Punch)
* **Defense**: the base damage resistance against normal attacks
* **SP Atk**: special attack, the base modifier for special attacks (e.g. fire blast, bubble beam)
* **SP Def**: the base damage resistance against special attacks
* **Speed**: determines which pokemon attacks first each round

**Tasks:**

* Visualize the data and get familiar with the observations.

1. Supervised machine learning (linear regression with one variable)

In this task, the independent variable X will represent ‘HP’ and output (dependent) variable represents ‘Attack’

* Implement the cost function for linear regression.
* Implement the gradient descent algorithm.
* Invoke these functions to optimize the parameters (theta) for the optimization problem J.
* Given X i= 30, predict the output Yi using the optimized model found above.
* Optional task:
  + Plot the history of executed cost functions stored in **result$J\_history**.
  + Plot the fitting lines outputted simple linear regression and the one computed through the gradient descent. Then compare analyze the accuracy of both fitting lines.

1. Unsupervised machine learning (clustering)

* Create and initialize a variable x that will contain column data from the Pokemon dataset, Let’s assign the observations for Hit Point (HP) to x. Then, Create the K-means model, called km.out with 4 clusters.
* Visualize, plot, and interpret the clustering output

1. Decision tree

* Use the R in-built package for decision tree to describe the ‘Legendary’ observations with regard to ‘Sp. Attack’ and ‘Sp. Defense’. Specifically, you need to describe how the legendary observation (true or false) affects the choice of both ‘Sp. Attack’ and ‘Sp. Defense’.