1. <https://bioconnector.github.io/workshops/r-predictive-modeling.html> (Flu Forecast)
2. <https://www.kaggle.com/shivamb/real-or-fake-fake-jobposting-prediction> (Job Posting Fraud)

The proposal should clearly state the question, the data to be used, the dependent variable and the explanatory variable of interest. (5 points)

Question: What is the likelihood of recovery for those affected by the flu given factors of age, gender, province, and hospitalization length?

Data To Be Used: bioconnector from Github

Dependent Variable: Death and Recovery outcome  
  
Explanatory Variable: age, gender, length of hospitalization, province

Our project will focus on the question: what is the likelihood of recovery for those affected by the flu given factors of age, gender, province, and hospitalization length? To achieve this goal, we will be using a dataset from github which contains explanatory variables about the onset of the flu, hospitalization length, the province in which it occurred, age of individuals, and gender of individuals. Our dependent variable will be the outcome of death or recovery from those affected.

The causal effect we are looking to establish is the effect of age, gender, province, and hospitalization length on the outcome of the patient (death or recovery). Without looking deeply into the data we expect younger people and people with shorter hospitalization lengths to recover more often than those who are older and in the hospital for longer. We will do more exploration into the gender and province factors to see if and by how much they are relevant to the health outcome of the patient.