CS Project

Student Name: Ghis Alhasan

Student Number: 215413205

Department: Computer Engineering

Class: CS

Instructure Name: Mehmet Baysan

Date: 10/5/2017

In this project, first, I was asked to choose one of the given files to study. The file I chose was Master as it is well organized, and has valuable information that I had good plans for.

The columns I chose were birthyear, weight, and height, and with these, I managed to think of three questions that would make a good hypothesis.
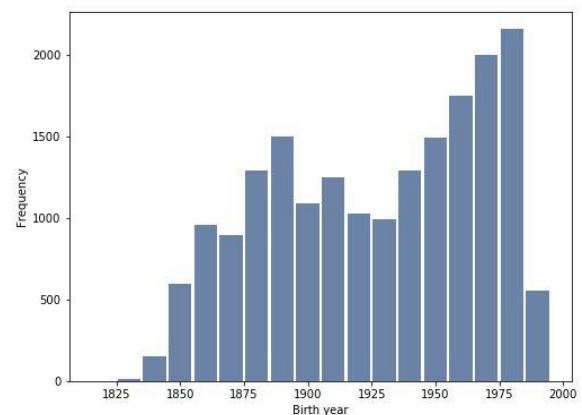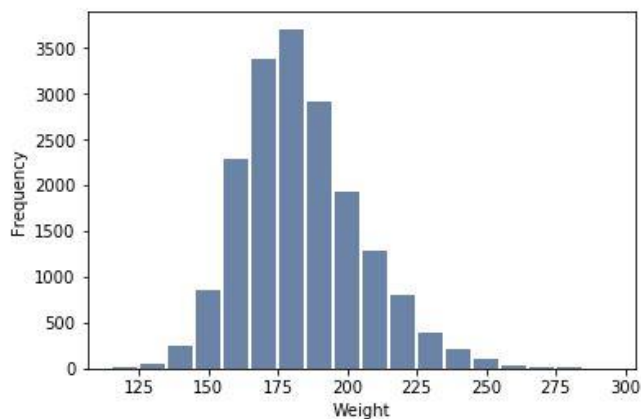
1. Do players born in the recent decades tend to care less about their weights?

2. Is the height and weight of players inversely proportional?

3. Did Teams began hiring tall players selectively at the last decades?

The question's theory suggests a good debate. In the recent years, there was a rise in the number fast food restaurants everywhere. The old folks didn't grow up with such bad eating habits, and diets, but the new generation came to accept it, and suffer from higher fat percentages. This project will attempt to check the validity of this point.
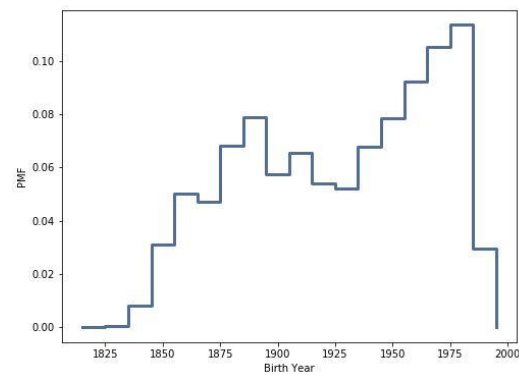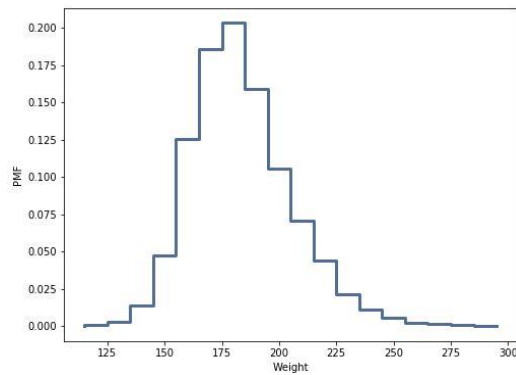
We begin our code by fetching the important information from the master file, clean it, and finally, make a dataframe of it making up our necessary columns. Aside from that, we create multiple lists and dictionaries about the columns important for us. weights, and heights.

Later, we are asked to plot a histogram, cdf, and pmf for our variables. We do it for weights, and the birthyears to get
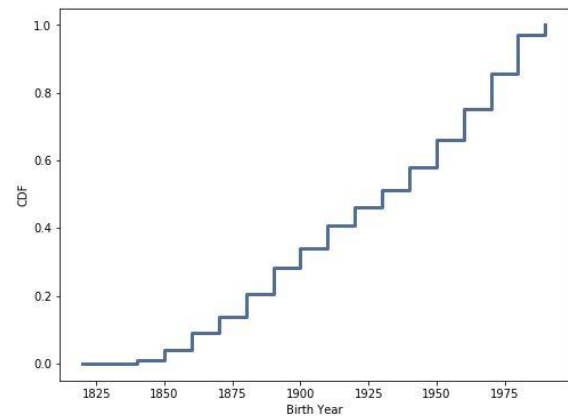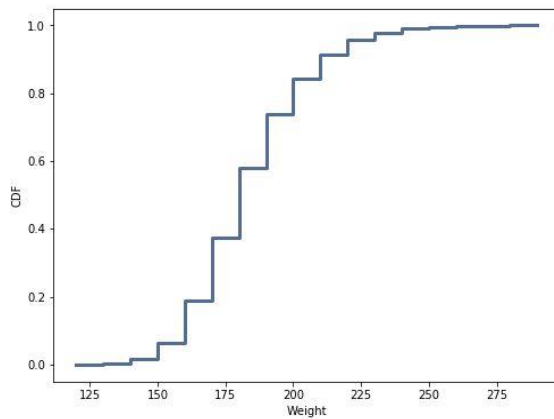
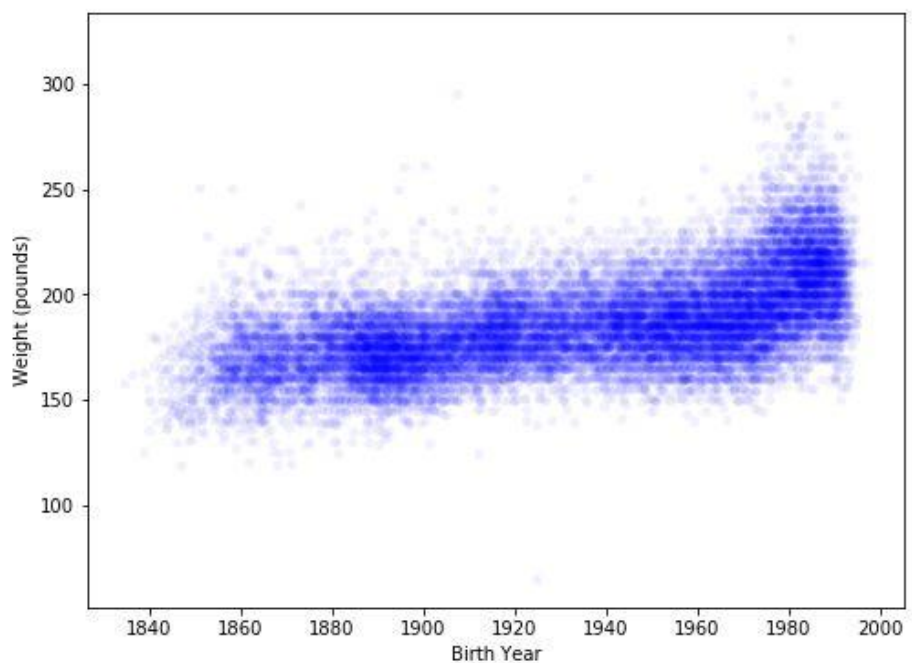1. Histograms

2. CDF's

3. PMF's

From these plots, we can tell the concentration points of the

samples, and their probabilities. This makes it easier to thoroughly

study the players.

We finally try the modeling distribution of the cumulative

distribution function, but instead, take the complement of it

making it a complementary cumulative distribution function,

another important model which can be studied.

Then we attempt to create a correlation which attempts to

decide the answer for the question. We use a scatterplot for the

weights and birthyears, and jitter it making it easier to visualize,

and understand.

As we can see, the scatter plot goes diagonally tilted up from left

to right, showing that the weights increase for players who were born

in the closer decades, verifying that the answer to the question "Do

players born in the recent decades tend to care less about their

weights?" is yes.

This study must not be avoided, nor taken for granted. It must be

considered as a sign that we need to change our lifestyles, and not fall

for the marketing of the fast food industry, or any types of unhealthy

foods.