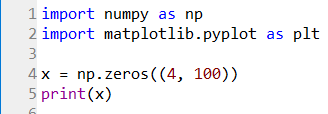
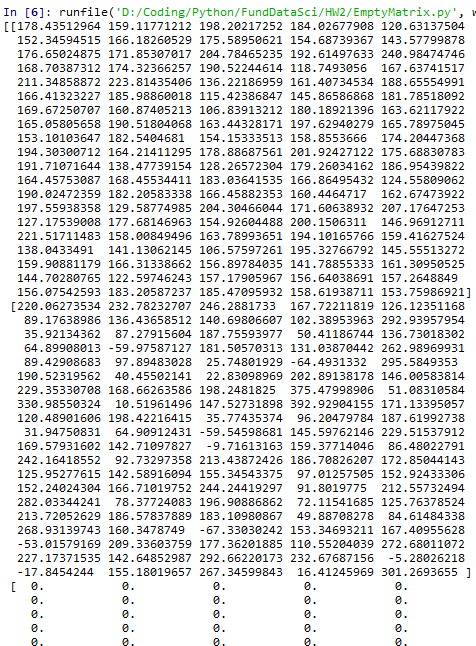
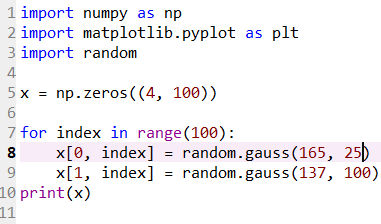
**Homework 2: Data Simulation**

1.





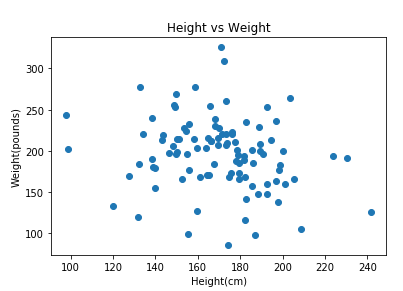
2.

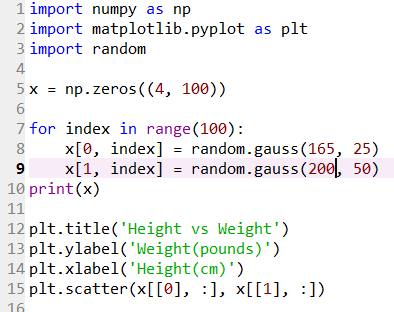


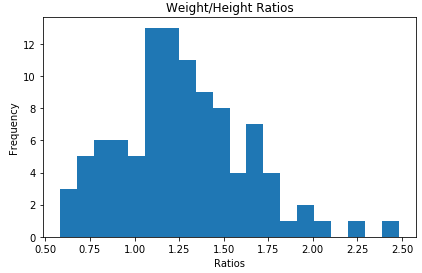
3. Assuming the heights are measured in centimeters, the data is valid on its own. The shortest and tallest person are within range of human height. However, the weights on their own are not valid. There cannot be people of negative weight. Also, the heights and weights considered together give some very unrealistic body sizes. Assuming weights are measured in pounds, some weights are too low to be possible with the correlating height. For example, there are height/weight pairs: 176.6cm/35.9pounds, 168.7cm/64.8pounds, and 174.3cm/-59.9pounds. It is unreasonable for weights to get as low as 0 (and below) when the range of heights suggests that there are no children.

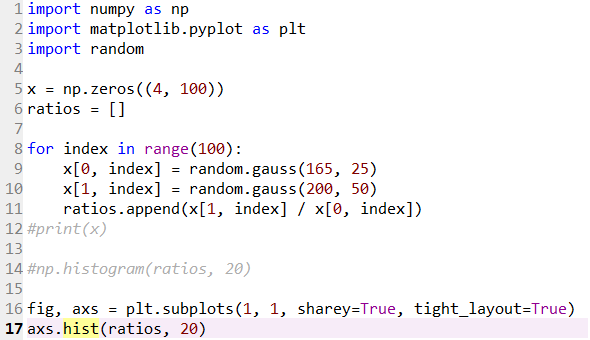
If randomness is to remain in data generation, solutions involve improving the data ranges to make them more realistic. If the height range and distribution are to stay the same, the weight data would have to tighten (decrease standard deviation) and overall increase to represent more accurate adult weights.

Alternatively, the current weight distribution could stay the same except for added preprocessing of negative weights while the height distribution and range are loosened and decreased to represent children’s heights. However, this solution is still problematic because of unrealistic height/weight pairs created from high standard deviations.

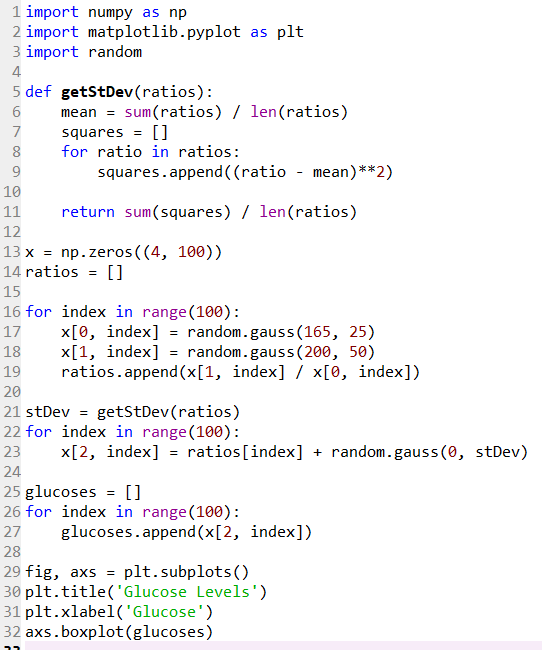
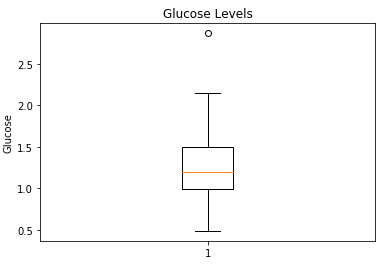
4.



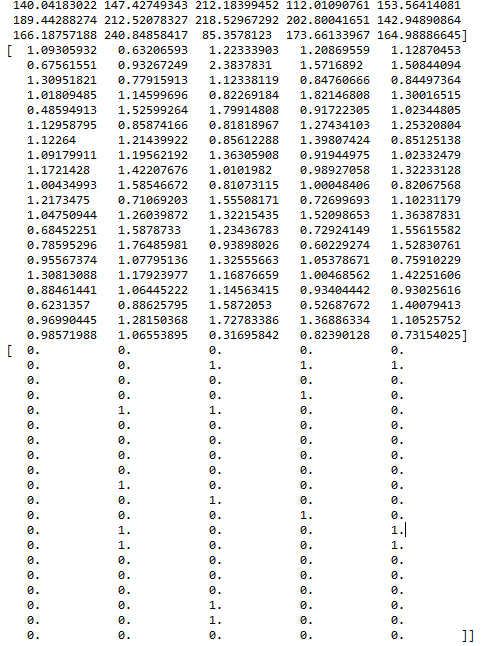
5.



6.

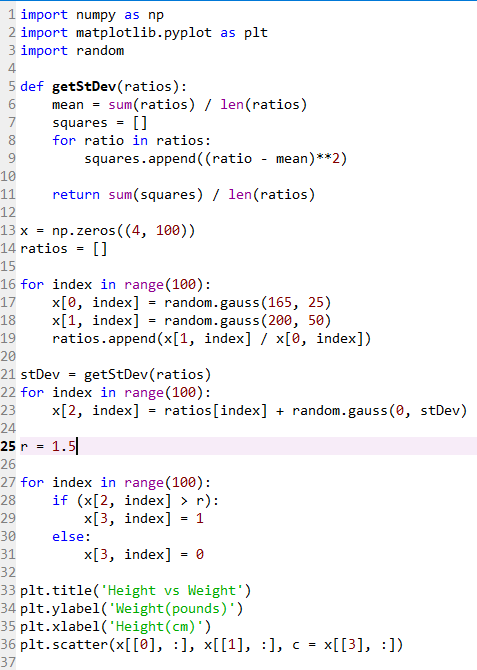
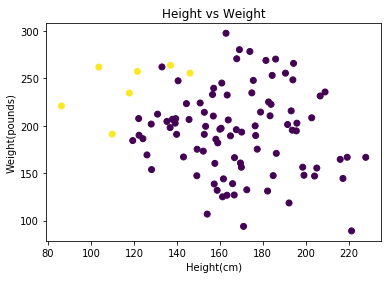
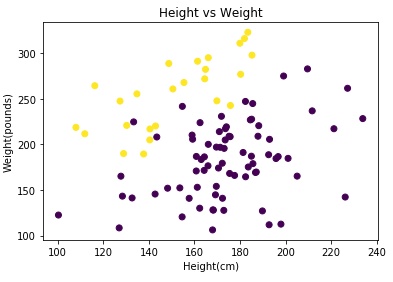
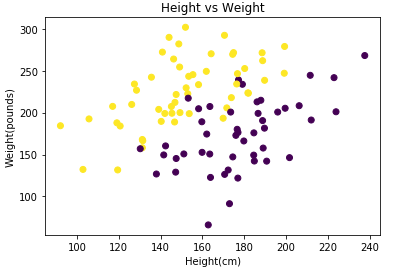
 

7.



8. Black = healthy, Yellow = diabetic

From top to bottom, r = 1.2, 1.5, and 1.8



9. Variance and the threshold number have a large effect on the sample data’s glucose levels and diabetes. The larger the variance of the data, the more spread out the range of glucose levels is. A smaller variance leads to more homogenous glucose levels with a smaller range. Also, a smaller threshold number means more people who end up having diabetes while a larger number results with a smaller portion of the sample size having diabetes.