The following is the summary of the dataset based on our last discussion. I tried to filter out the number of people who are diagnosed with diabetes and high blood pressure and calculated their probabilities as follows. (N.b. all information regarding the diagnosis history is based on the answers from the questionnaire)

1. Total dataset size: 535,169
2. People who are diagnosed with diabetes: 18,332
3. People who are on medication for diabetes: 16,082
4. People who are diagnosed with diabetes but not on medication: 2,461
5. Probability of being diabetes
   * p(d) = 18,332/535,169 = 0.034
6. People diagnosed with high blood pressure (HBP): 48,805
7. People who are on medication for high blood pressure: 43,617
8. People who are diagnosed with HBP but not on medication: 5,882
9. probability of being HBP
   * p(hbp) = 48,805/535,169 = 0.091
10. People who are diagnosed with both HBP and diabetes: 7,936
11. Probability of being diabetes and have HBP at the same time
    * p( d n hbp ) = 7,936/535,169 = 0.014
12. Probability of being diabetes given the person have hbp
    * p(d/hbp)= p(d n hbp)/p(hbp) = 0.014/0.091 = 0.162
13. Probability of being hbp given the person have diabetes
    * p(hbp/d) = p(d n hbp)/p(d) = 0.014/0.034 = **0.432**