DS 100: Principles and Techniques of Data Science **Date: February 2, 2018**

Discussion #2

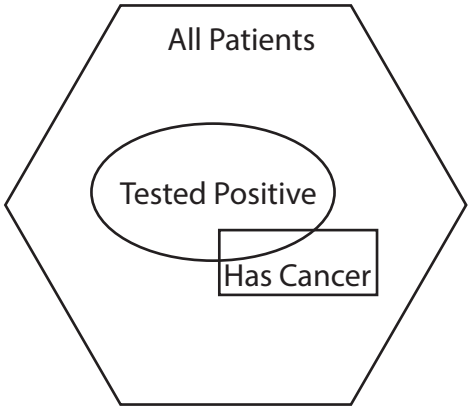
Name:

Bayes' Rule

1. Are you smarter than a doctor? ONLY 46% OF DOCTORS GOT THIS QUESTION RIGHT.

   100 out of 10,000 women at age forty who participate in routine screening have breast cancer. 80 of those 100 women with breast cancer test positive. 950 out of 9,900 women without breast cancer also test positive. If 10,000 women in this age group undergo a routine screening, about what fraction of these women with positive tests will actually have breast cancer?

2. Consider the following Venn Diagram of the same problem.



   Discuss as a class how we might represent conditional probabilities, chain rule, and Bayes' rule using the Venn Diagram. Use the space below for your notes.

3. HEALTH PROFESSIONALS HATE THIS.

   Only 1% of women at age forty who participate in a routine mammography test have breast cancer. 80% of women who have breast cancer will test positive, but 9.6% of women of women who dont have breast cancer will also get positive tests. A woman of this age tested positive in a routine screening. What is the probability that she actually has breast cancer?

Data Visualization and Scope

This part of the discussion will be centered on this video

https://tinyurl.com/data100-rosling

4. Answer the following questions about the quality of the visualization in the video.

   (a) What variables are being represented in the graphic?

   (b) How are the variables being represented visually?

   (c) How do we interpret the visual qualities? In other words, how can we look at the image and know how to interpret the properties of the plot into data?

   (d) Does it look like the raw values of the data were plotted or were they (numerically) transformed before plotting?

   (e) Is there any information present that is not represented visually?

   (f) How good do you think the data are for this visualization?