STA 100 Discussion Week 3

Book Problems

1. If a women takes an early pregnancy test, she will either test positive or negative. Suppose that if a women is really pregnant, there is a 98% chance that she will test positive. If a women is not really pregnant, there is a 99% chance she will test negative. Lastly, assume the overall probability that a women who takes the test is pregnant is 0.10.

   (a) Find the probability that a women is both pregnant and tests positive

   (b) Find the probability that women is both pregnant and does not test positive

   (c) Find the probability a women tests positive

   (d) Find the probability a women is either pregnant, or tests negative, or both (the union)

   (e) Find the probability that if a women tests positive, she is pregnant.

   (f) Find the probability that if a women does not test positive, she is not pregnant.

   (g) Are testing positive and being pregnant independent?

2. Observe the following frequency table of income and smoking status:

	(L)	(M)	(H)
Smoked (S)	63	32	24
Do Mot Smoke (S^C)	184	162	186

Where L = low income, M = medium income, and H = high income.

(a) Estimate the probability that someone smokes.

(b) Estimate the probability that someone both smokes and is high income.

(c) If someone is high income, estimate the probability they smoke?

(d) If someone does not smoke, estimate the probability they are not low income.

(e) Estimate the probability that someone does not smoke, or has medium income, or both.

R Portion

1. Online you will find the file "GSK.csv". The csv file has the following columns:

   Column 1. `sysbp`: The systolic blood pressure of the subject (mmHg).

   Column 2. `gender`: The gender, with levels F and M.

   Column 3. `married`: Y if the subject was married, N if not.

   Column 4. `exercise`: With levels L = low, M = medium, H = high.

   Column 5. `age`: The age of the subject in years.

   Column 6. `stress`: With levels LS = low, MS = medium, HS = high.

   Column 7. `educatn`: With levels LE = low, ME = medium, HE = high.

   (a) Create a histogram of systolic blood pressure. Which interval has the highest number of subjects in it?

   (b) Create a boxplot of systolic blood pressure by exercise group. Which group has the highest median?

   (c) Create a barplot for the stress levels. Which group has the least subjects in it?

   (d) Create a mosaic plot between stress and education. Of those who have "medium" education, what is the most common stress level?