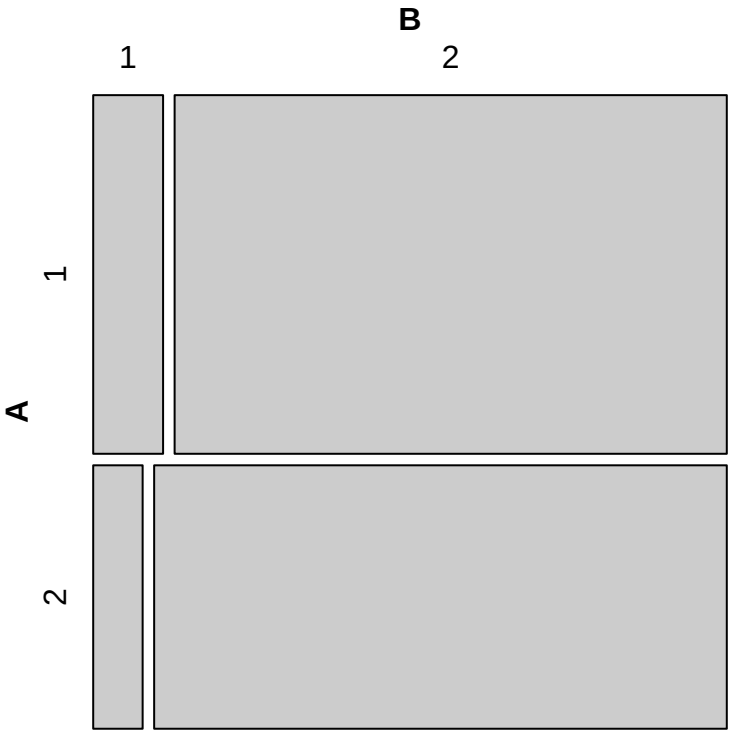
ST 518 - Homework 2

Nora Quick

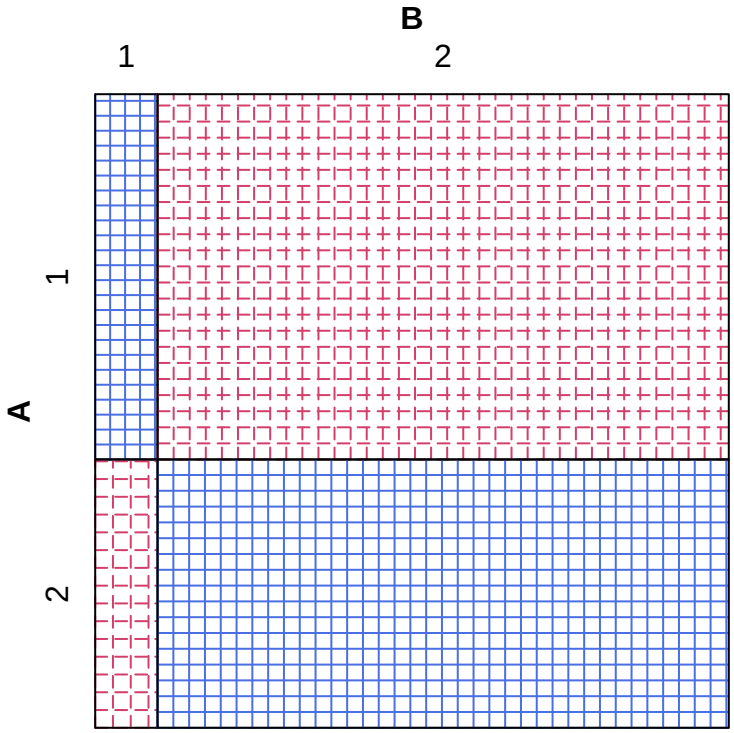
R Question:

1. Conduct your own analysis of the data and comment on this conclusion. Is there evidence of a difference between the two groups? Is the strength of the causal implication of this statement justified by the data from this study?

```
tab1 <- matrix(c(102,806,53,614),2,2,byrow=TRUE)
mosaic(tab1)
```




```
sieve(tab1, shade = TRUE)
```



Based on the work we did in lab I decided to concude a test using graphs. Based on these graphs there is no evidence that there is a diffence between the two groups. There is no justification for the strength of the researchers casual implication of the statement. In fact, the language they use seems not only incorrect, but extremely incorrect.

Conceptual Questions:

2. During an investigation of the U.S. space shuttle Challenger disaster, it was learned that project managers had judged the probability of mission failure to be 0.00001, whereas engineers working on the project had estimated failure probability at 0.005. The difference between these two probabilities, 0.00499, was discounted as being too small to worry about. Is a different picture provided by considering odds? How is that interpreted?

Yes, a different picture is provided becuase while the odds are very small there is a significant amount of weight in the 0.00499 that is discounted. If the two calculated odds were closer it would possibly be more reasonable to negate it. But with such a big different between the two there should have been more calculations done because clearly one of the groups was off their calculation or the two calculations are not similar enough to be correlated. And when it comes to lives that difference should have been looked at more.

3. Suppose that 90% of orange tabby cats are male. Determine if the following statements are true or false, and explain your reasoning.

a. True, with such a small population the outcome would be skewed.

b. False, it would reduce it but not by half.
c. True, with a sample size that is not too large or two small it should be relatively normal.
d. False, from what I have learned in the class on big data the larger your samples get the less fluxuation within the data there will be. In addition it would likely start to show the true 90% with that number.

4. A 2010 survey asked 827 randomly sampled registered voters in California, "Do you support or do you oppose drilling for oil and natural gas off the coast of California? Or do you not know enough to say?" Below is the distribution of responses, separated based on whether or not the respondent graduated from college.

a. About 24% of college graduates do not know enough and about 34% of non-college graduates do not know enough.
b. The hypothesis is $\mu 1 = \mu 2$. To test with without R (as it seems we are supposed to) we can test the equality in proportion of all three. As we see there is 35% & 34%, 41% & 32%, and 24% & 34%. We can see that support is about equar, however, the next two are not close to equlelent. Therefore we can say that the proportion of college graduates who do not have an opinion on the issue is different from that of non-college graduates.

5. A study of British male physicians noted that the proportion who died from lung cancer was 0.00140 per year for cigarette smokers and 0.00010 per year for nonsmokers. Additionally, the proportion who died from heart disease was 0.00669 for smokers and 0.00413 for nonsmokers. Which response (lung cancer or heart disease) is more strongly related to cigarette smoking, in terms of the reduction in deaths that could occur with the absence of smoking?

Lung cancer has a stronger reponse to cigarette smoking. We can see this because there is a 0.0013 differece between smoker and non-smoker versus 0.00256 difference. In other words, there is a larger difference in lung cancer between groups than heart disease.