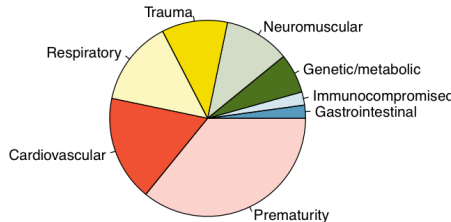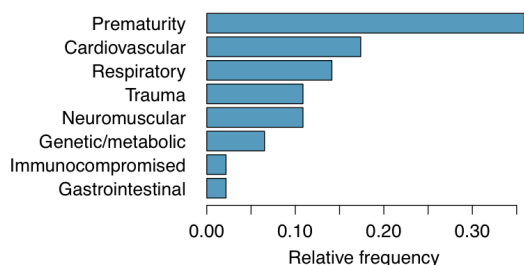Mat 1372 HW3

2.21 Antibiotic use in children. The bar plot and the pie chart below show the distribution of pre-existing medical conditions of children involved in a study on the optimal duration of antibiotic use in treatment of tracheitis, which is an upper respiratory infection.

(a) What features are apparent in the bar plot but not in the pie chart?
(b) What features are apparent in the pie chart but not in the bar plot?
(c) Which graph would you prefer to use for displaying these categorical data?

(a) The relative frequency

(b) There are no features that are apparent in the pie chart but not in bar plot

(c) Bar plot.

2.22 Views on immigration. 910 randomly sampled registered voters from Tampa, FL were asked if they thought workers who have illegally entered the US should be (i) allowed to keep their jobs and apply for US citizenship, (ii) allowed to keep their jobs as temporary guest workers but not allowed to apply for US citizenship, or (iii) lose their jobs and have to leave the country. The results of the survey by political ideology are shown below.[22]

		Political ideology			
		Conservative	Moderate	Liberal	Total
	(i) Apply for citizenship	57	120	101	278
	(ii) Guest worker	121	113	28	262
Response	(iii) Leave the country	179	126	45	350
	(iv) Not sure	15	4	1	20
	Total	372	363	175	910

(a) What percent of these Tampa, FL voters identify themselves as conservatives?
(b) What percent of these Tampa, FL voters are in favor of the citizenship option?
(c) What percent of these Tampa, FL voters identify themselves as conservatives and are in favor of the citizenship option?
(d) What percent of these Tampa, FL voters who identify themselves as conservatives are also in favor of the citizenship option? What percent of moderates share this view? What percent of liberals share this view?
(e) Do political ideology and views on immigration appear to be independent? Explain your reasoning.

(a) $\frac{372}{910} = 41\%$

(b) $\frac{278}{910} = 31\%$

(c) $\frac{57}{910} = 6\%$

(d) $\frac{\# \text{ citizenship and conservative}}{\# \text{ conservative}} = \frac{57}{372} = 15\%$

$\frac{\# \text{ citizenship and moderates}}{\# \text{ moderates}} = \frac{120}{363} = 33\%$

$\frac{\# \text{ citizenship and liberals}}{\# \text{ liberals}} = \frac{101}{175} = 58\%$

(e) The percentages in (d) cross conservative, moderates, and liberals are quite different from one another. Therefore, the two variables apear to be dependent.

A random sample of registered voters from Tampa, FL were asked if they support the DREAM Act, a proposed law which would provide a path to citizenship for people brought illegally to the US as children. The survey also collected information on the political ideology of the respondents. Based on the mosaic plot shown below, do views on the DREAM Act and political ideology appear to be independent? Explain your reasoning.[23]
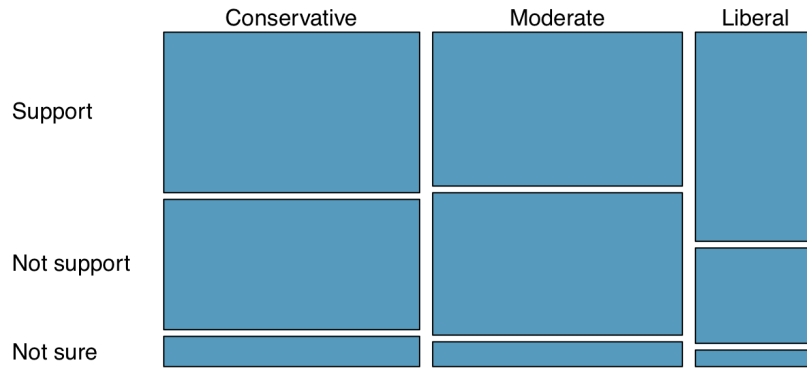
There are three categories for the Act : Support, Not support, Not sure. Only 30% ~40% of the people in the first two ideological groups support the Act, but over 60% of the people in liberal group support it. This shows that the two variables may be dependent.

2.26   **Heart transplants.** The Stanford University Heart Transplant Study was conducted to determine whether an experimental heart transplant program increased lifespan. Each patient entering the program was designated an official heart transplant candidate, meaning that he was gravely ill and would most likely benefit from a new heart. Some patients got a transplant and some did not. The variable `transplant` indicates which group the patients were in; patients in the treatment group got a transplant and those in the control group did not. Of the 34 patients in the control group, 30 died. Of the 69 people in the treatment group, 45 died. Another variable called `survived` was used to indicate whether or not the patient was alive at the end of the study.[30]
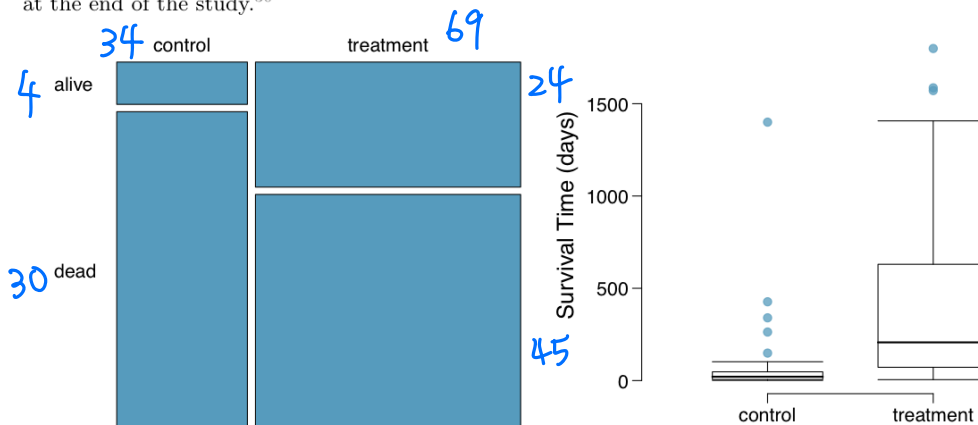
(a) Based on the mosaic plot, is survival independent of whether or not the patient got a transplant? Explain your reasoning.

(b) What do the box plots below suggest about the efficacy (effectiveness) of the heart transplant treatment.

(b) Both are right skewed, but the median in treatment group is higher than control group and the patients who didn't get transplant tended to consistently die quite early relative to those who got the transplant.
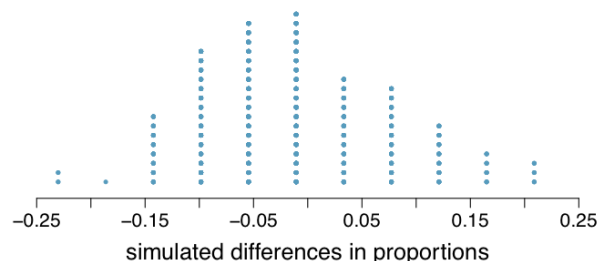
(a) Alive rate from control group is $\frac{4}{34} = 11.76\%$ and Alive rate from treatment group is $\frac{24}{69} = 34.78\%$. This shows that the survival may be dependent of whether or not the patient got a transplant.

(c) What proportion of patients in the treatment group and what proportion of patients in the control group died?

(d) One approach for investigating whether or not the treatment is effective is to use a randomization technique.

    i. What are the claims being tested?

    ii. The paragraph below describes the set up for such approach, if we were to do it without using statistical software. Fill in the blanks with a number or phrase, whichever is appropriate.

> We write *alive* on ___28___ cards representing patients who were alive at the end of the study, and *dead* on ___75___ cards representing patients who were not. Then, we shuffle these cards and split them into two groups: one group of size ___69___ representing treatment, and another group of size ___34___ representing control. We calculate the difference between the proportion of *dead* cards in the treatment and control groups (treatment - control) and record this value. We repeat this 100 times to build a distribution centered at ___0___. Lastly, we calculate the fraction of simulations where the simulated differences in proportions are ~23.20% If this fraction is low, _or lower_ we conclude that it is unlikely to have observed such an outcome by chance and that the null hypothesis should be rejected in favor of the alternative.

    iii. What do the simulation results shown below suggest about the effectiveness of the transplant program?

simulated differences in proportions

(handwritten annotation, right side):

in (a)

Alive in control minus
Alive in treatment
= 11.67% - 34.87%
= -23.20%

(handwritten solutions):

(c) Dead rate from control group is $\frac{30}{34} = 88.23\%$ and Dead rate from treatment group is $\frac{45}{69} = 65\%$.

(d) (i) H_0: The variables group and outcome are independent: The difference of survival rates in two groups was due to chance.

H_A: The variables group and outcome are NOT independent. The difference of survival rates in two groups was not due to chance and the transplant is effective

(ii) alive: 4 + 24 = 28
dead: 30 + 45 = 75
treatment group size = 69
control group size = 34

(iii) Under the independence hypothesis (H_0 assumes to be correct), only 2 out of 100 times (2%) did we get a difference of

-23.2% or lower between the proportions of patients that alive in treatment and control groups.

Since this is a low probability (2%), we can reject H_0 in favor of the alternate model H_A.