Problem 1:

Introduction:

The goal of this report is to find the relationship between year, group and success of the treatment, which helps patients with a particular conditions to decide whether or not to be treated.

Our analysis is solely based on the current dataset *Trial.csv* which recorded patients' conditions after receiving treatments for two years. We start off calculating the estimated rate of success rate for treatment overall, followed by the estimated probability of a successful treatment for group A and B, respectively. Then we calculate the estimated probability of a successful treatment for year one and two. Note that *confidence intervals* for each sample estimate are also created, to capture how much they differ from true population value.

Summary:

Below is the partition table that we used to find our sample estimates.

Year One		A	B	Total
	No	26	27	53
	Yes	86	83	169
	Totals	112	110	222
Year Two		A	B	Total
	No	27	28	55
	Yes	45	45	90
	Totals	72	73	145

By applying 1-sample proportions test in R (*see Appendix*), we found that the following estimate probabilities and their confidence intervals.

The sample estimate for the Probability of a successful treatment overall is

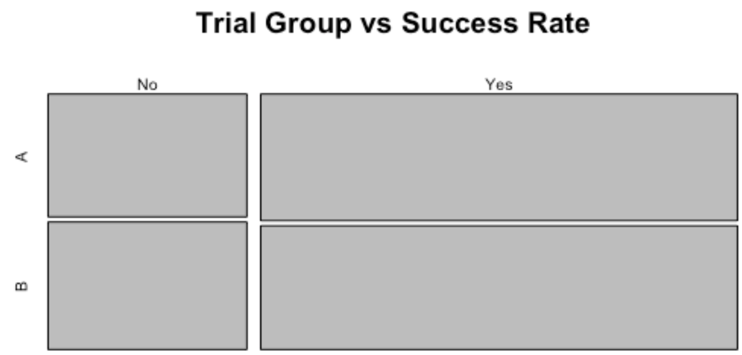
=0.7057, and its 95 percent confidence interval is (0.6558, 0.7513). It means that we are 95% confident that the population probability of a successful treatment overall falls into (0.6558, 0.7513).

The sample estimate for the Probability of success for group A is = 0.7120, and its 95 percent confidence interval is (0.6399, 0.7750). It means that we are 95% confident that the population probability of a successful treatment for A falls into (0.6399, 0.7750).

The sample estimate for the probability of success for group B is = 0.6995, and its 95 percent confidence interval is (0.6266, 0.7637). It means that we are 95% confident that the population probability of a successful treatment for B falls into (0.6266, 0.7637).

The sample estimate for the Probability of a successful treatment, for year 1 is = 0.7613, and its 95 percent confidence interval is (0.6987, 0.8146). It means that we are 95% confident that the population probability of a successful treatment for year 1 is (0.6987, 0.8146).

The sample estimate for the Probability of a successful treatment, for year 2 is = 0.6207, and its 95 percent confidence interval is (0.5360, 0.6988). It means that we are 95% confident that the population probability of a successful treatment for year 2 is(0.5360, 0.6988)



These two mosaic plots show that the probability of success in both is higher than being unsuccessful. In the *Trial Year vs Success Rate* graph, we can see that the success rate of year one seems to be greater than that of year two. From the *Trial Group vs Success Rate* mosaic plot, we can see that the success rates for group A is slightly higher than that of group B.

Analysis:

We will now analyze whether or not successes are independent from groups, with and without year.

First, we consider the confidence interval for relative risk between for each group, **without the year**. If 1 falls within the confidence interval, which means that there is no significant difference between these two variables. It also proves that they are independent.

Formula is below: $Ln(R) \pm Z(1\text{-}\alpha/2)\sqrt{\frac{(1-Pi(1)}{Yi} + \frac{(1-Pi(2)}{Y2}}$

As we have above:

	A	B	Total
No	53	55	108
Yes	131	128	259
Totals	184	183	367

R = $\frac{(131/184)}{(128/183)} = \frac{(0.7120)}{(0.6995)}$ =1.0179, *Ln(R)* = 0.0177

$$\sqrt{\frac{(1-Pi(1)}{Yi} + \frac{(1-Pi(2)}{Y2}} = \sqrt{\frac{(1-0.7120)}{(131)} + \frac{(1-0.6995)}{(128)}} = 0.0674,$$

Then,
95% Confidence Interval = 0.0177 ± 1.96* 0.0674
= 0.0177 ± 0.1321 = (-.1144, .1498)

After exponentiating 95% confidence interval would be (0.8919, 1.1616)0

Next, we perform the calculations for the confidence intervals of relative risks between for each group for **each year**, assuming the level of significance is 5%(**α** = 5%).

For year one, *R* = $\frac{(86/112)}{(83/110)} = \frac{(0.7679)}{(0.7545)}$ =1.0176, *Ln(R)* = 0.0175

$$\sqrt{\frac{(1-Pi(1)}{Yi} + \frac{(1-Pi(2)}{Y2}} = \sqrt{\frac{(1-0.7679)}{(86)} + \frac{(1-0.7545)}{(83)}} = 0.0752$$

Then, $Z(1-\alpha/2g)=1.78$
95% Confidence Interval = 0.0176 ± $Z(1-\alpha/2g)$* 0.0752
= 0.0176 ± 0.1339 = (-0.1163, 0.1515)
After exponentiating 95% confidence interval would be (0.8902, 1.1635)

For year two, *R* = $\frac{(45/72)}{(45/73)} = \frac{(0.625)}{(0.6164)}$ =1.0139, *Ln(R)* = 0.0138

$$\sqrt{\frac{(1-Pi(1)}{Yi} + \frac{(1-Pi(2)}{Y2}} = \sqrt{\frac{(1-0.625)}{(45)} + \frac{(1-0.6164)}{(45)}} = 0.1298$$

Then,
95% Confidence Interval = 0.0138 ± $Z(1-\alpha/2g)$* 0.1298
= 0.0138 ± 0. 231= (-.2172, .2448)
After exponentiating 95% confidence interval would be (0.8047, 1.2774)

Interpretation:

Not considering the year, the confidence interval of the relative risk between for each group is (0.8966, 1.1554). Since 1 is within this 95% confidence interval, it means that there is evidence to suggest that there is NO difference for the success of treatments between group A and B. This means that regardless of if you are in Group A or Group B the treatment has the same probability of being successful. The probability of success of treatments for group A and the one for group B are independent.

The results remain the same, when considering the year separately.

Conclusion:

In conclusion, the success of treatments for group A is independent from Group B. It means regardless of what team the patients are in, their treatments have the same probability of success. Since the treatment overall receives over 60% rate of success, patients might want to consider taking the treatments.