Section 1:

1. Identify the control group and the treatment group

Treatment group: Group that is given Pfizer-BioNTech's trial COVID-19 vaccine
Control group: Group that is given a placebo treatment and not the vaccine

2. Write a hypothesis for what you think the effect of receiving the Pfizer-BioNTech COVID-19 vaccine was on whether individuals were infected with COVID-19

Words:

Null Hypothesis: Receiving the Pfizer-BioNTech COVID-19 vaccine did not have any effect on the probability of individuals being infected with COVID-19.

Research/Alternative Hypothesis: Receiving the Pfizer-BioNTech COVID-19 vaccine decreased an individual's chance of being infected with COVID-19 compared to those who did not receive the vaccine.

Math:
μ = population proportion
μ_0 = hypothesized value proportion
α(alpha) = 0.05

Null Hypothesis - H_0: $\mu >= \mu\_0$, if p-value >/= α(0.05)
Research/Alternative Hypothesis - H_A: $\mu\_0 > \mu$, if p-value < α(0.05)

3. The alpha value I chose is 0.05 ($\alpha = 0.05$)

4. Results:

Words:
After doing the Left-tailed Binomial Test, I came to the conclusion that I reject the null hypothesis since the p-value (0.00000001276) I got is less than the alpha value of 0.05 and therefore we accept the alternative hypothesis.

Math:
Reject H_0, since p-value(0.00000001276) < α(0.05)
Accept H_A, since p-value(0.00000001276) < α(0.05)

5.Was receiving the Pfizer-BioNTech vaccine correlated with reduced risk (lower probability) of getting COVID-19?

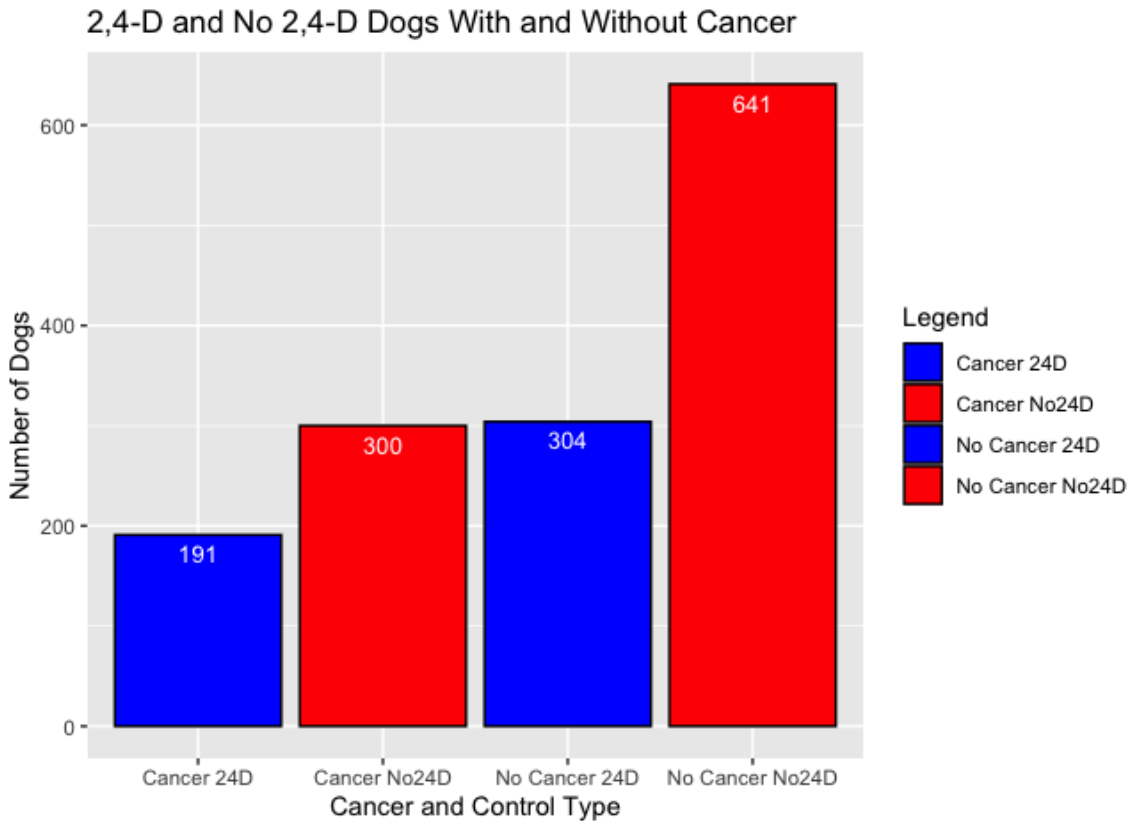
Since we rejected the null hypothesis because our p-value was less than our alpha value we can say that we have evidence that receiving the Pfizer-BioNTech vaccine correlated with reduced risk or lower probability of getting COVID-19.

Does this study make it possible to prove causation?

We cannot ultimately say that the study makes it possible to prove causation since proving causation heavily relies on the study design. First, we need a controlled study, and the vaccine study is a controlled study for the most part, as we have a treatment and placebo group. However, we must ensure that one group is kept unchanged and the other we can change. In this study, we had to make sure all the participants had the same genetic background, genetic risk factors, diets, and other risk factors, and things they did have to be the same and controlled. In addition, the study must be monitored for a very extended period, as in decades, and every aspect of their life must be controlled. If all these components are factored in, then can we possibly imply causation, but this study does not satisfy all these components to say it proves causation. Even if there is a correlation, it does mean causation.

Section 2:

Graph 1:

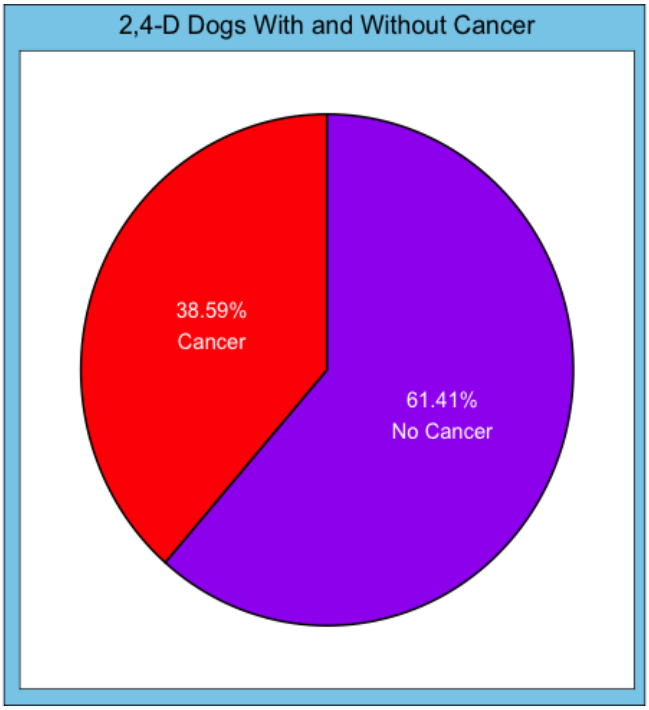


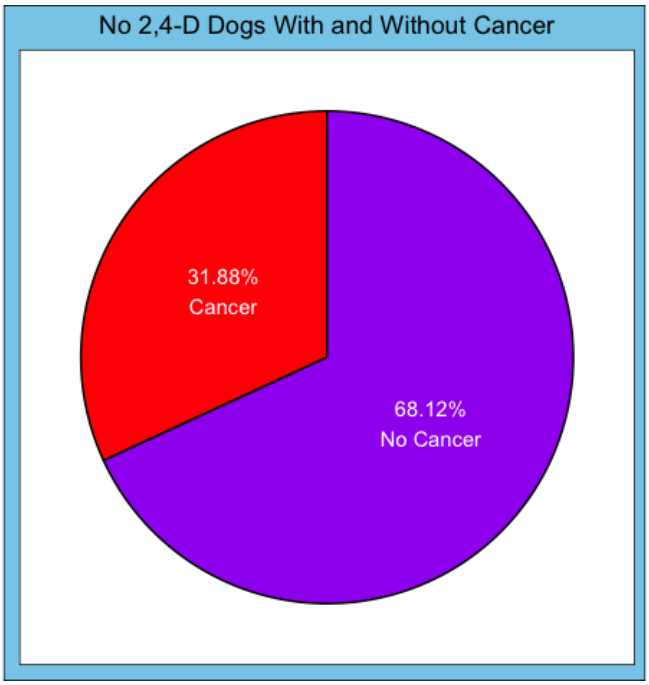
What was graphed:
The dataset I chose to work on is the "Cancer in dogs" dataset. The first graph illustrated above is a bar chart, the first one I decided to use to represent the data. The bar chart shows the data on how many dogs developed and did not develop cancer in the two groups into which they were separated. The red bars represent the group of No 2,4-D dogs or the dogs who were not exposed to the herbicide, while the blue bars represent the dogs who were exposed to the herbicide or the 2,4-D dogs. Then each bar in the graph shows how many dogs in those groups developed or did not develop cancer. Looking at the x-axis, you can see which groups represent each of the bars.

Why you chose those graphs:
The reason why I chose a bar chart is because it is very good at representing data visually. With a bar chart, you can use the bars to distinguish easily between the different groups by color coding, as I have done. You can also quickly compare the data side by side by using the height of the bars to see how the data compares. For example, if you look at the blue "No Cancer

24D" bar and the red "No Cancer No24D" bar, you can quickly see that the "No Cancer No24D" bar dominates compared to the blue one by looking at the height of the bars. Another reason why I decided to use a bar chart is because bar charts allow data that are comprised of words to be easily graphed, as the "Cancer in dogs" dataset did not contain data with any numbers but words. If we look at the y-axis, I do not have numbers but words that are split into different categories. In addition, when it comes to large datasets, you can easily fit them into a bar chart. In my graph, the y-axis ranges from 0 to about 700, as the dataset I used had more than 1000 entries.

Graph 2:





What was graphed:
For my second graph, I decided to use two pie graphs to represent the data from the “Cancer in dogs” dataset. The first pie chart illustrates the comparison between the percentage of dogs who got cancer and did not get cancer in the 2,4-D group or the group exposed to the herbicide. The second pie chart illustrates the comparison between the percentage of dogs who got cancer and did not get cancer in the No 2,4-D group or, in other words, the group of dogs that were not exposed to the herbicide.

Why you chose those graphs:
The reason why I chose a pie chart is because pie charts are very useful when you are trying to convey a part to a whole relationship or when you want to show that one part of the data is small or large in comparison to the rest. In both pie charts, I wanted to show and figure out whether there are more dogs with cancer or no cancer in each group. The pie chart makes this very easy and quick to visualize. If you look at both of the pie charts, we can see that in both groups, there are more dogs with no cancer than there were with cancer. As the dogs with cancer section of the pie is bigger in size in both pie charts than that of the dogs with no cancer and which is also emphasized by the percentage number depicted in each segment. In addition, we can also color code segments of pie charts to emphasize more meaning and importance to the data, hence why I used red to represent cancer and purple to represent no cancer. Red is usually a color that symbolizes a sense of importance, and cancer is a severe issue, so I chose red to highlight cancer in dogs.