UTM STA Data Activity

Module 1 Tutorial Presentation: Exploring Categorical Data

Learning Objective

- Use R to explore data through RStudio in laptop/desktop or RStudio via U of T JupyterHub

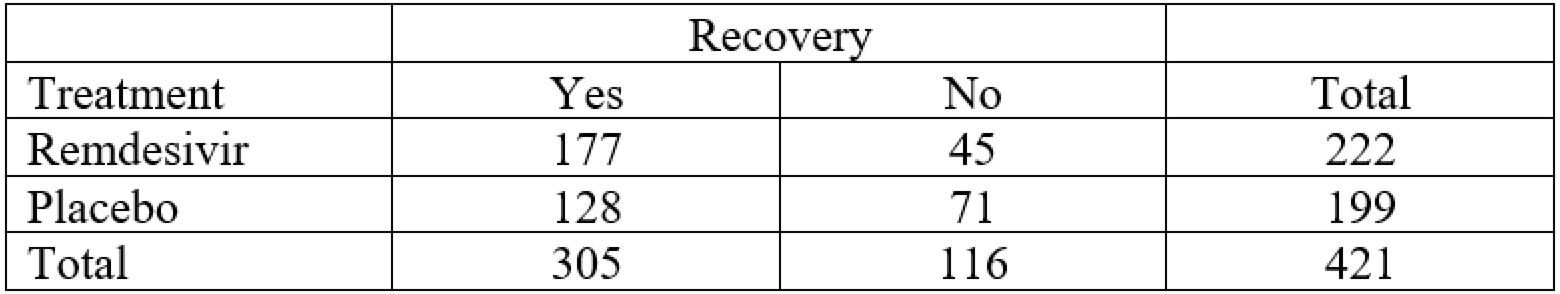
- Construct plots and frequency tables for describing categorical data

- Assess association between variables

Get into Small Groups ☺
Discuss the Tutorial Worksheet Questions

Tutorial Worksheet Activity

In December 2019, a novel coronavirus (COVID-19) was identified that would soon spread around the world. An existing drug called Remdesivir was identified early on as a potential treatment against the symptoms of COVID-19, which includes severe respiratory infections. In the spring of 2020, a clinical trial assessed the efficacy of Remdesivir versus a placebo treatment in terms of the recovery of patients infected and hospitalized with the COVID-19 virus. Here, we look at data from this trial concerning patients with severe COVID-19 (those that needed oxygen supply). Of 222 such patients randomly assigned to the Remdesivir treatment group, 177 recovered after 28 days. Of 199 such patients randomly assigned to the placebo group, 128 recovered after 28 days. The contingency table below shows these results.

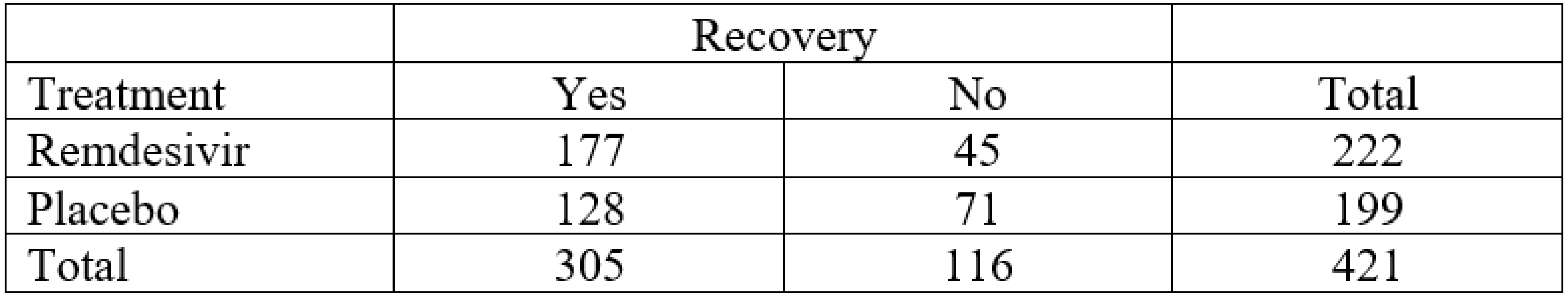


	Recovery		
Treatment	Yes	No	Total
Remdesivir	177	45	222
Placebo	128	71	199
Total	305	116	421

Use the context of this study to answer the related questions.

a. Identify the response and explanatory variable.

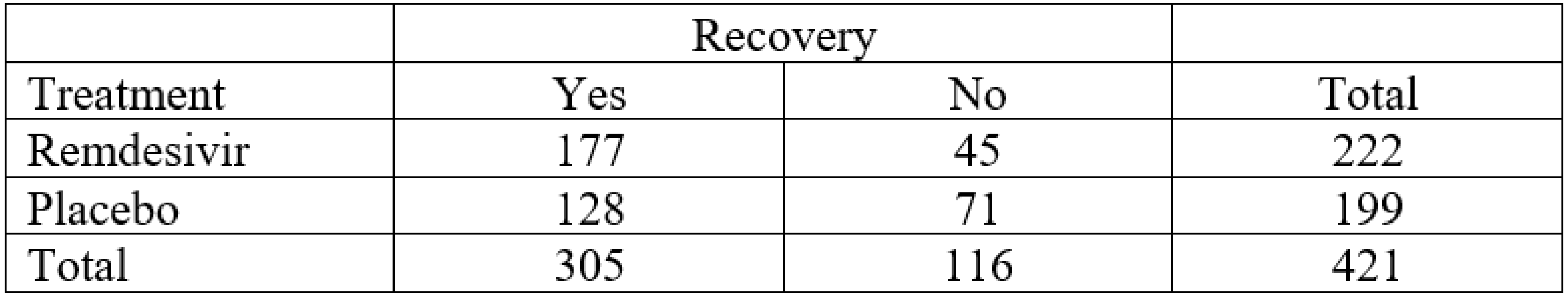
Tutorial Worksheet Activity



	Recovery		
Treatment	Yes	No	Total
Remdesivir	177	45	222
Placebo	128	71	199
Total	305	116	421

b. Identify the response and explanatory variable.

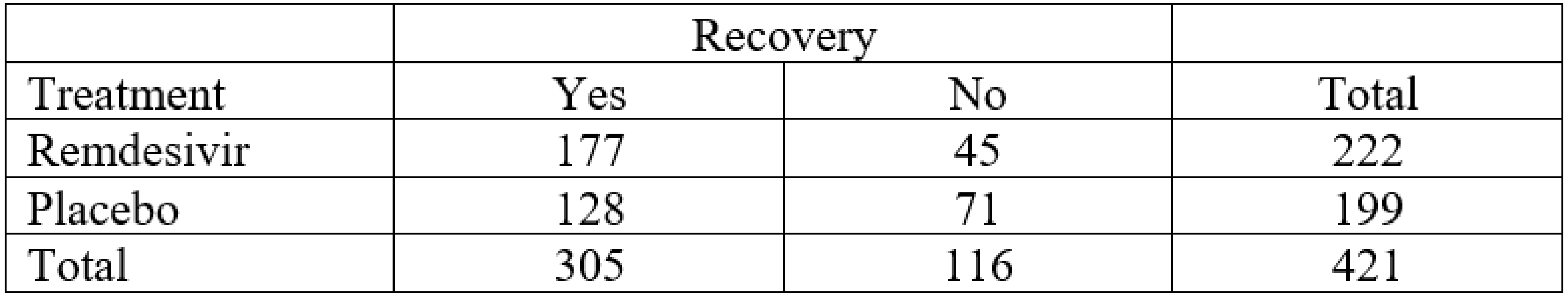
Tutorial Worksheet Activity



	Recovery		
Treatment	Yes	No	Total
Remdesivir	177	45	222
Placebo	128	71	199
Total	305	116	421

c. Compute the difference between the proportion of recovery for Remdesivir and placebo and interpret this difference in context.

Tutorial Worksheet Activity



	Recovery		
Treatment	Yes	No	Total
Remdesivir	177	45	222
Placebo	128	71	199
Total	305	116	421

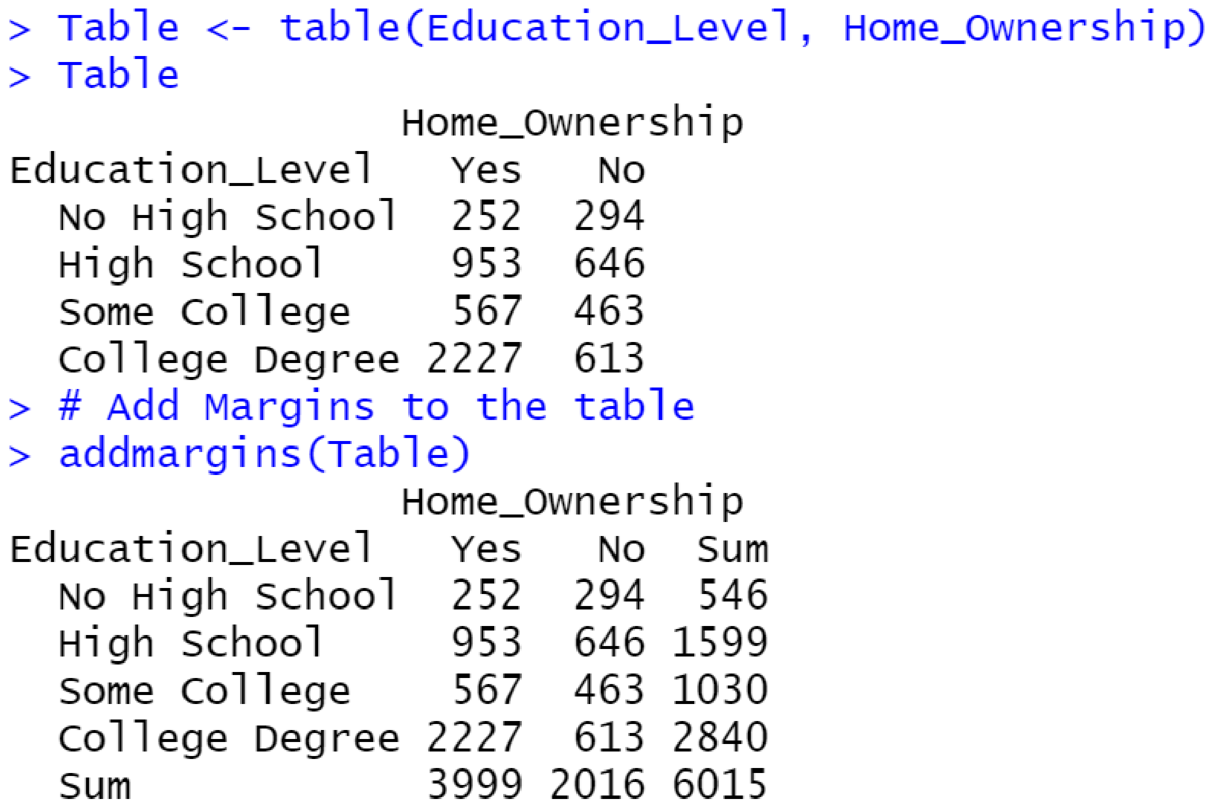
d. Compute the ratio between the proportion of recovery for Remdesivir and placebo and interpret this ratio in context.

Pre-Task Tutorial Activity using R

You used R and investigated an association between two variables in **Home_SCF2013** data file.

Recall: The Survey of Consumer Finances (SCF, 2013) took a random sample of 6015 adult Canadians and collected information on their level of education and whether or not they own a home.

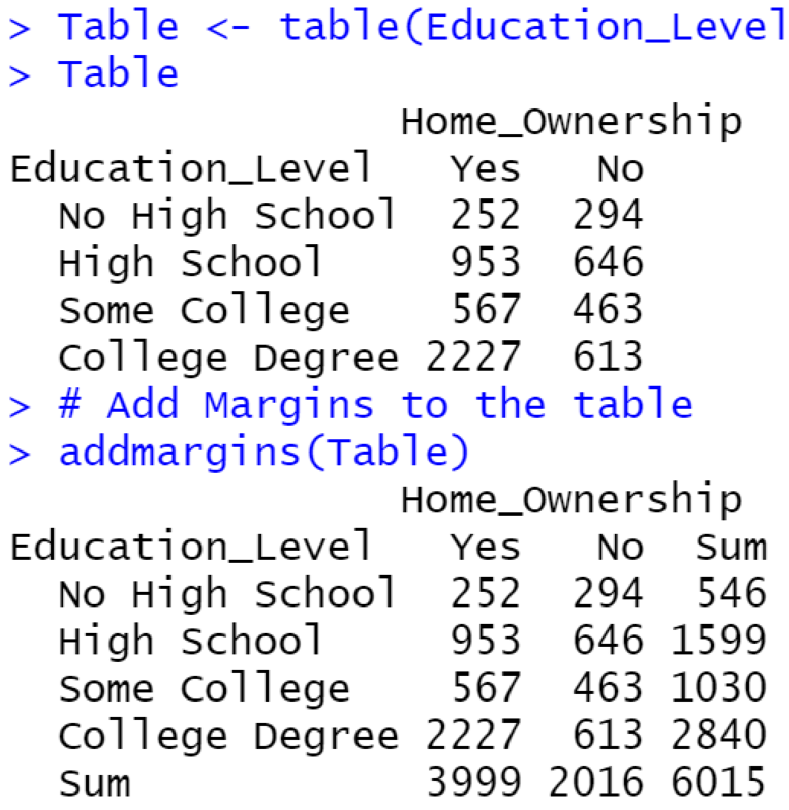
The contingency table in the right margin shows the results.



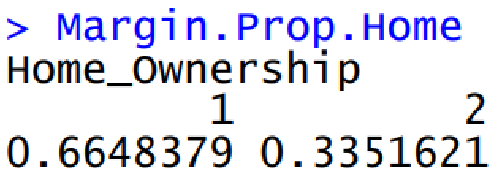
```
> Table <- table(Education_Level, Home_Ownership)
> Table
                Home_Ownership
Education_Level   Yes   No
  No High School  252  294
  High School     953  646
  Some College    567  463
  College Degree 2227  613
> # Add Margins to the table
> addmargins(Table)
                Home_Ownership
Education_Level   Yes   No  Sum
  No High School  252  294  546
  High School     953  646 1599
  Some College    567  463 1030
  College Degree 2227  613 2840
  Sum            3999 2016 6015
```

Pre-Task Tutorial Activity using R

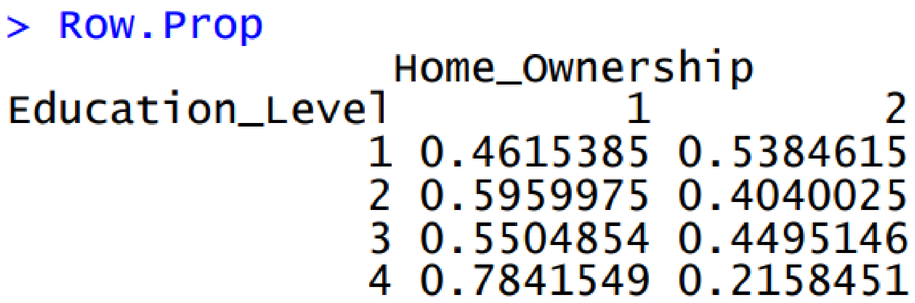
Using data-based arguments (i.e., comparing conditional proportions) describe the relationship between the variables **“home ownership”** and **“education level”**.



```
> Table <- table(Education_Level, Home_Ownership)
> Table
              Home_Ownership
Education_Level   Yes   No
  No High School  252  294
  High School     953  646
  Some College    567  463
  College Degree 2227  613
> # Add Margins to the table
> addmargins(Table)
              Home_Ownership
Education_Level   Yes   No  Sum
  No High School  252  294  546
  High School     953  646 1599
  Some College    567  463 1030
  College Degree 2227  613 2840
  Sum            3999 2016 6015
```



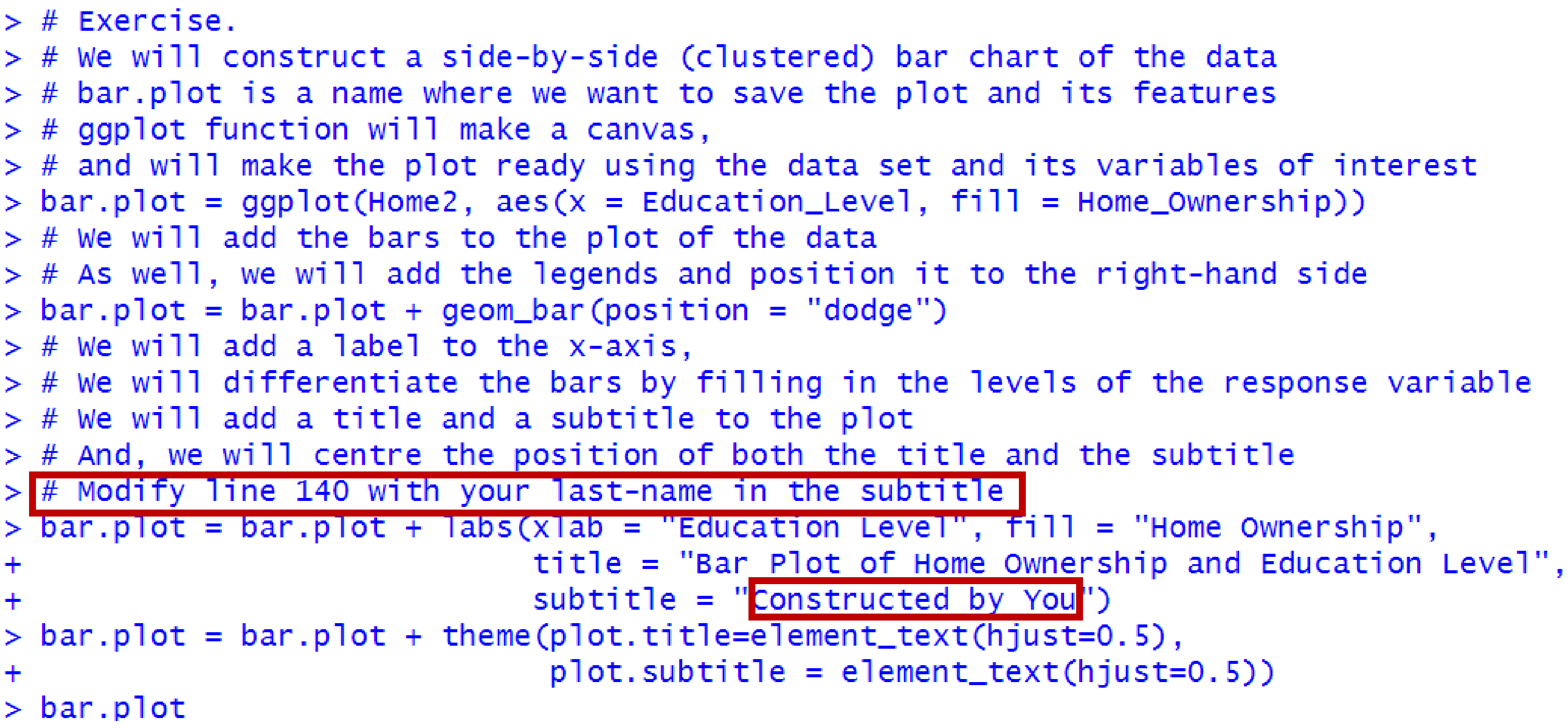
```
> Margin.Prop.Home
Home_Ownership
        1         2 
0.6648379 0.3351621 
```



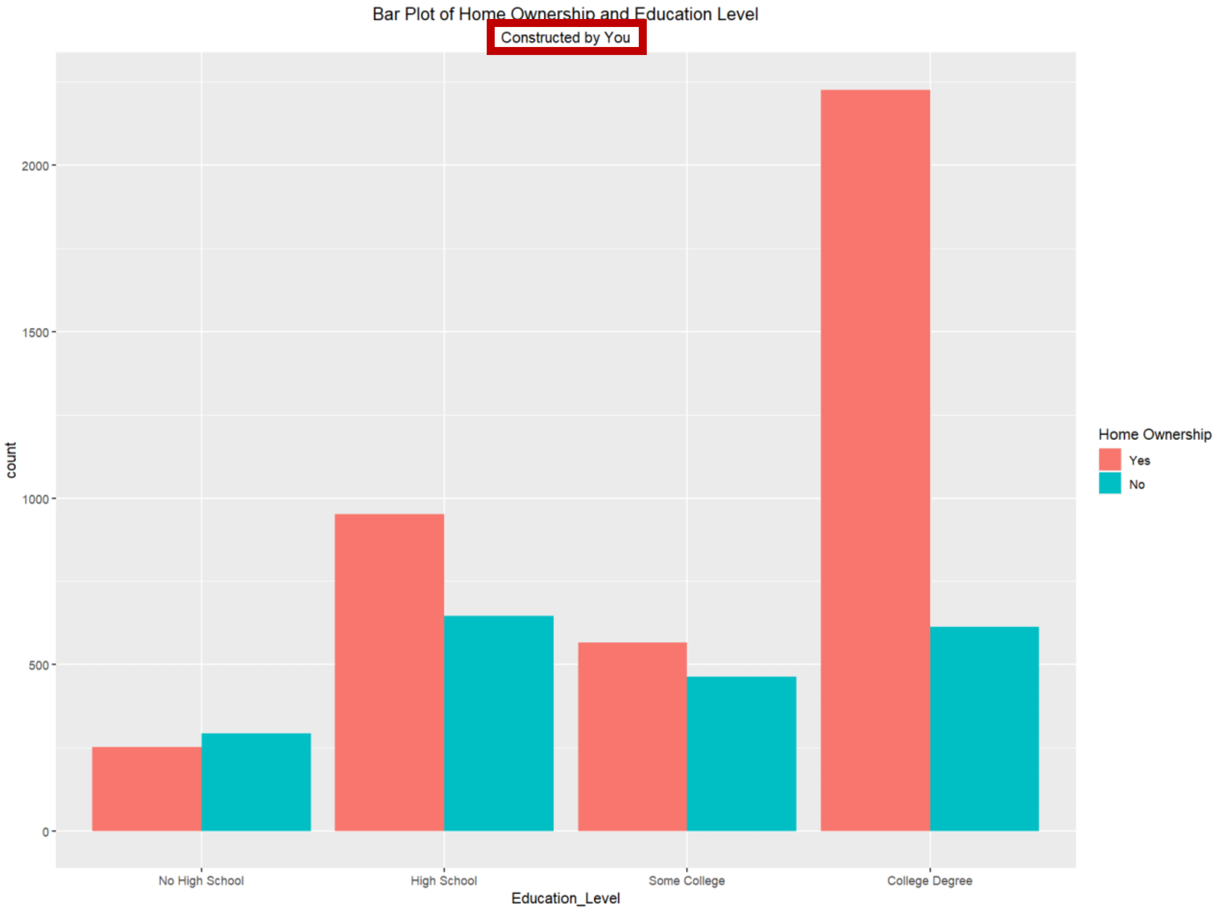
```
> Row.Prop
               Home_Ownership
Education_Level         1         2
              1 0.4615385 0.5384615
              2 0.5959975 0.4040025
              3 0.5504854 0.4495146
              4 0.7841549 0.2158451
```

Pre-Task Tutorial Activity using R

b. Submit through Quercus your bar plot of home ownership by education level (see Module 2 Tutorial page for the online submission instruction).

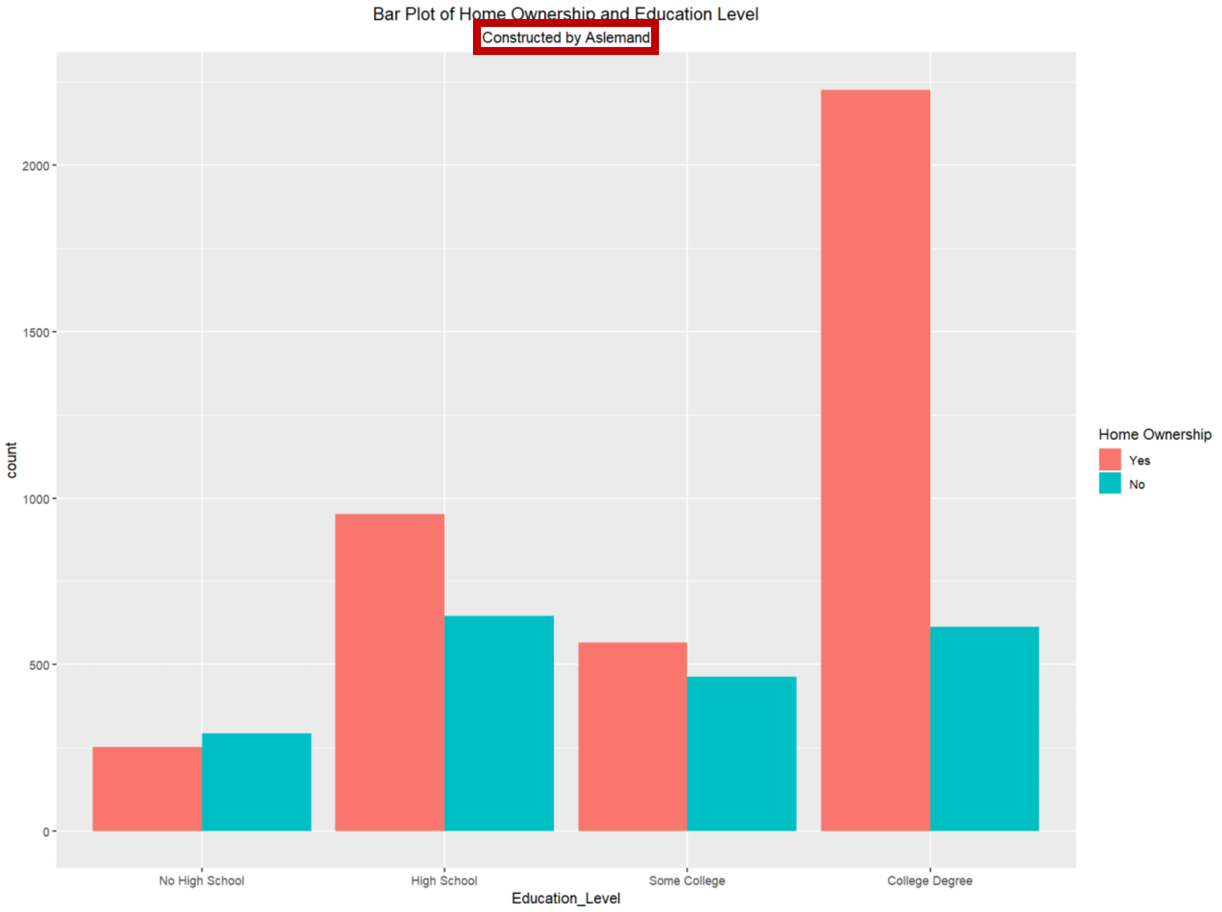


```
> # Exercise.
> # We will construct a side-by-side (clustered) bar chart of the data
> # bar.plot is a name where we want to save the plot and its features
> # ggplot function will make a canvas,
> # and will make the plot ready using the data set and its variables of interest
> bar.plot = ggplot(Home2, aes(x = Education_Level, fill = Home_Ownership))
> # We will add the bars to the plot of the data
> # As well, we will add the legends and position it to the right-hand side
> bar.plot = bar.plot + geom_bar(position = "dodge")
> # We will add a label to the x-axis,
> # We will differentiate the bars by filling in the levels of the response variable
> # We will add a title and a subtitle to the plot
> # And, we will centre the position of both the title and the subtitle
> # Modify line 140 with your last-name in the subtitle
> bar.plot = bar.plot + labs(xlab = "Education Level", fill = "Home Ownership",
+                            title = "Bar Plot of Home Ownership and Education Level",
+                            subtitle = "Constructed by You")
> bar.plot = bar.plot + theme(plot.title=element_text(hjust=0.5),
+                             plot.subtitle = element_text(hjust=0.5))
> bar.plot
```



Bar Plot of Home Ownership and Education Level

Constructed by You



Bar Plot of Home Ownership and Education Level

Constructed by Aslemand

Tutorial Reflection

Based on your participation in the tutorial session today, please reflect on the following three questions.

You may choose to answer to each of the following question individually or answer all three globally/generally in the space provided in your worksheet😊

1. What statistical contents were you able to understand?
2. What statistical contents were challenging for you to grasp?
3. What are your strategies/plans to improve and/or expand your knowledge on this week’s contents?