lab-3

Vivian Yeh

Part I: Understanding the Context of the Data

```
library(stat20data)
data(class_survey)
library(tidyverse) #add the library package
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.3.6      v purrr   0.3.4
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.0      v stringr 1.4.1
v readr   2.1.2      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
```

1. **What is the unit of observation in the survey of students in Stat 20?**

   A student is currently enrolled in Stat 20 and took the survey

2. **Where in the Taxonomy of Data are the variables that correspond to each of the following survey questions (i.e., what is their type)?**

   - **Question 6: What is your favorite thing about Cal?**

     Nominal Categorical

   - **Question 13: Where would you rather be?**

     Nominal Categorical

- **Question 22: What is the chance that a fire alarm goes off in one of the dorms this week?**

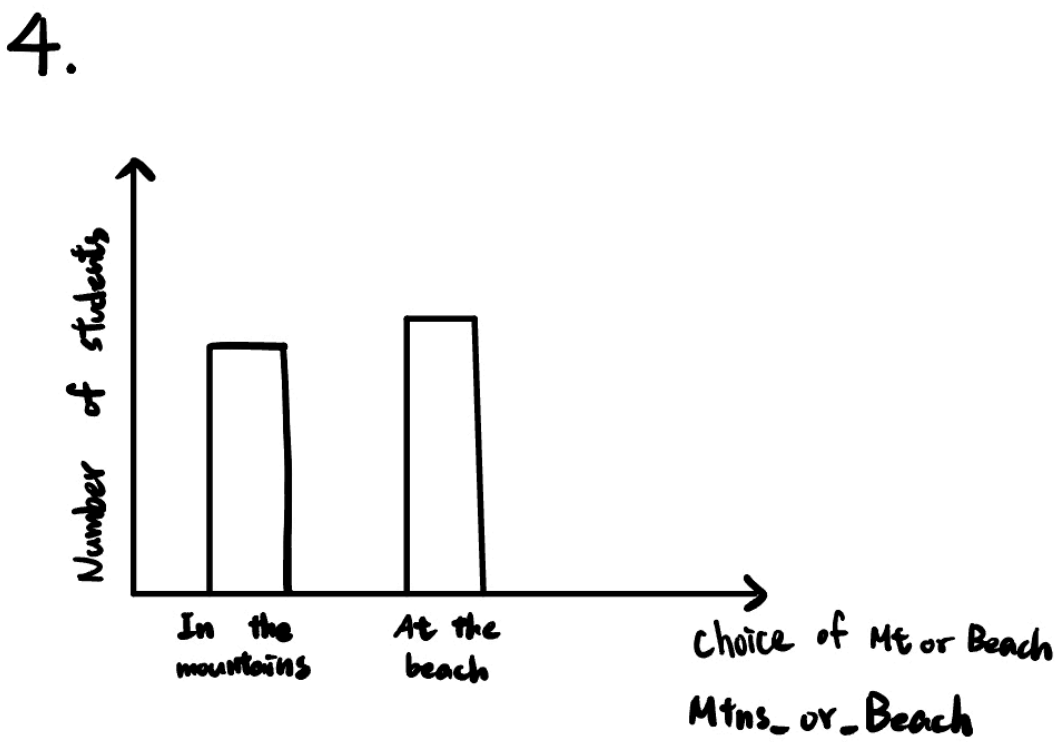
  Continuous Numerical

3. **What values would you expect `Major` to take? Do you anticipate any challenges when analyzing this data?**

   Character

   The challenge would be how the students answer this question. People would answer "Econ", "econ", and "Economics", and they are all have the same meaning, which is Economics major". Therefore, this will be a challenge when analyzing this data.

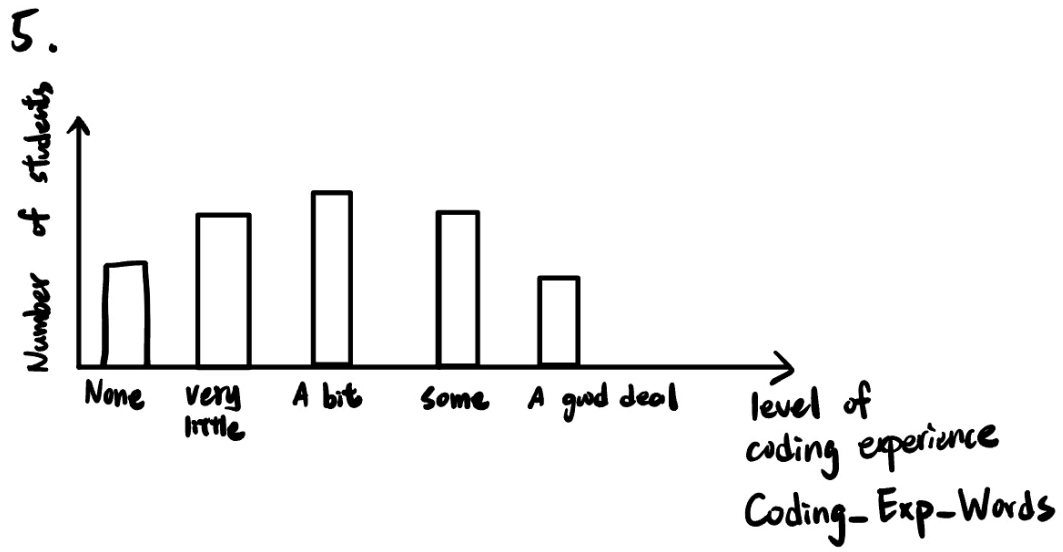
4. **Do students prefer to spend time at the beach or in the mountains?**



   Aesthetic mappings: x-axis: Mtns_or_Beach

   Geometry:Bar chart

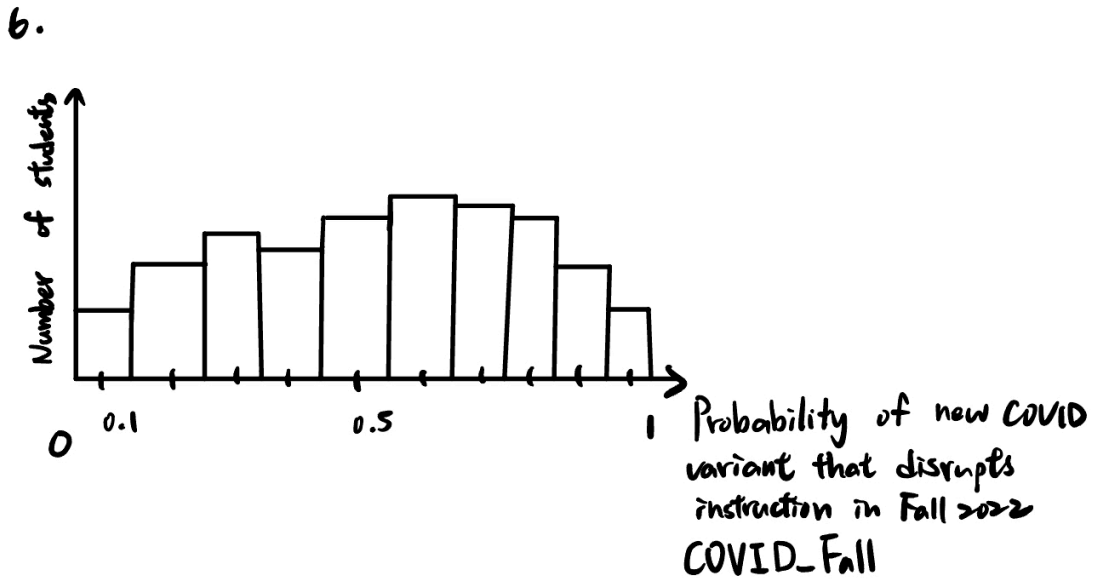
5. **How much coding experience do students have?**



Aesthetic mappings: x-axis: Coding_Exp_Words

Geometry:Bar chart

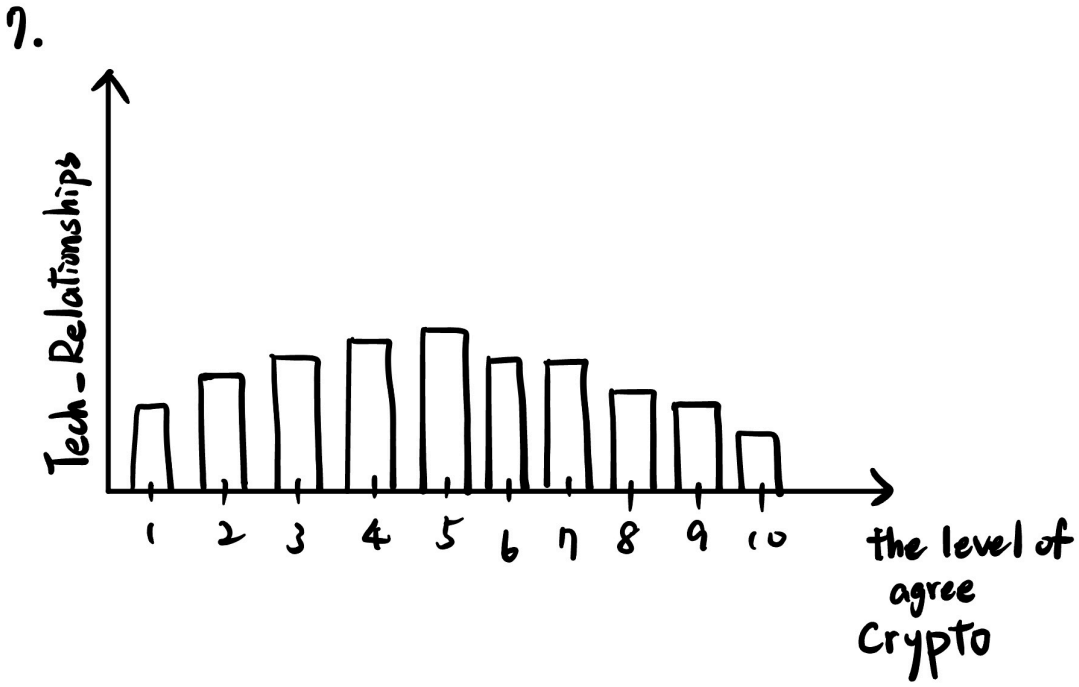
6. **What are students' perceptions of the chance that there is a new COVID variant that disrupts instruction in Fall 2022?**



Aesthetic mappings: x-axis: COVID_Fall

Geometry: Histogram

7. **What is the relationship between students' optimism for cryptocurrency and their skepticism of the effect of technology on interpersonal relationships?**



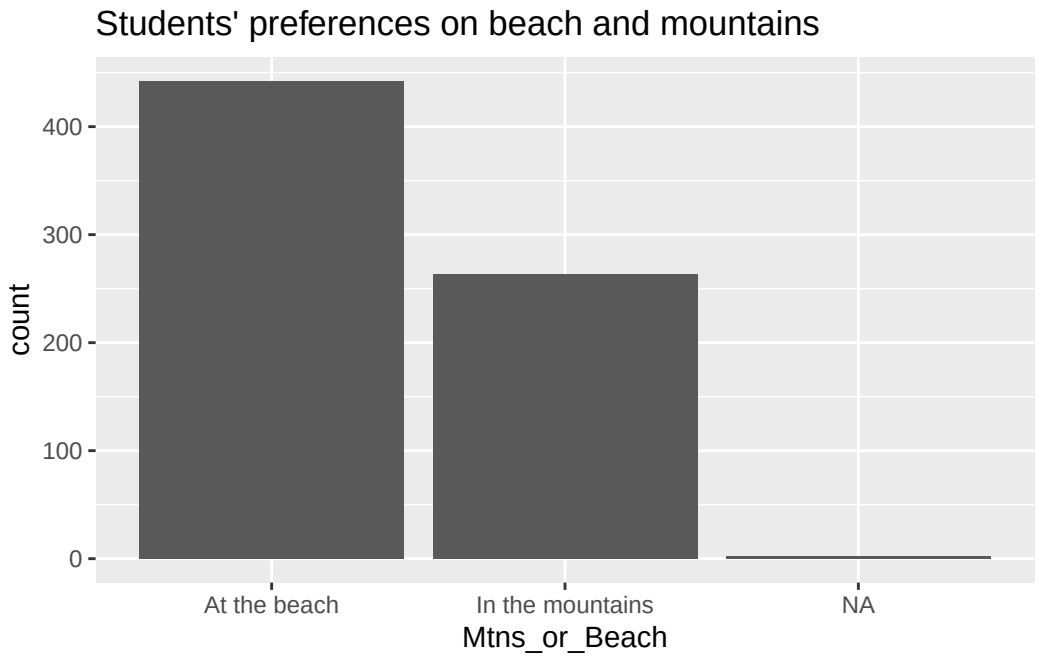
Aesthetic mappings: x-axis: Crypto/ y-axis: Tech_Relationships

Geometry:Bar chart

Part II: Computing on the Data

8. **Do students prefer to spend time at the beach or in the mountains?**

```
ggplot(class_survey) + ggtitle("Students' preferences on beach and mountains") +
  geom_bar(aes(x = Mtns_or_Beach))
```




```
total <- nrow(select(class_survey, Mtns_or_Beach)) #707
num_Mnt <- nrow(filter(class_survey, Mtns_or_Beach == "In the mountains")) #263
num_Beach <- nrow(filter(class_survey, Mtns_or_Beach == "At the beach")) #442

find_mode <- function(x) {
  u <- unique(x)
  tab <- tabulate(match(x, u))
  u[tab == max(tab)]
}
mode_Mnt_or_Beach <- find_mode(class_survey$Mtns_or_Beach)
```

mean: Cannot find because the response of this question is nominal categorical

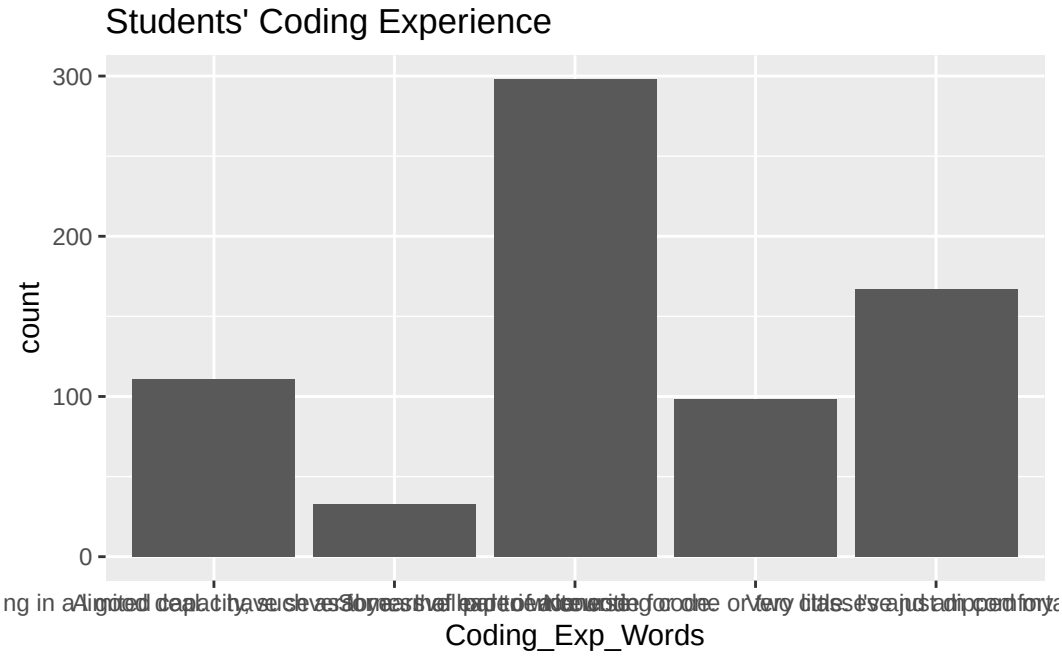
median: "At the beach"

mode:"At the beach"

Ans: Students prefer to spend time at the beach

9. **How much coding experience do students have?**

```
ggplot(class_survey) + ggtitle("Students' Coding Experience")+
  geom_bar(aes(x = Coding_Exp_Words))
```




```
#mean_code <- mean(class_survey$Coding_Exp_Words)
#median_code <- median(class_survey$Coding_Exp_Words)
mode_Coding_Exp_Words <- find_mode(class_survey$Coding_Exp_Words)
```

mean: Cannot find because the response of this question is ordinal categorical

median: Cannot find because the response of this question is ordinal categorical, and we do not know how the data be arranged in order

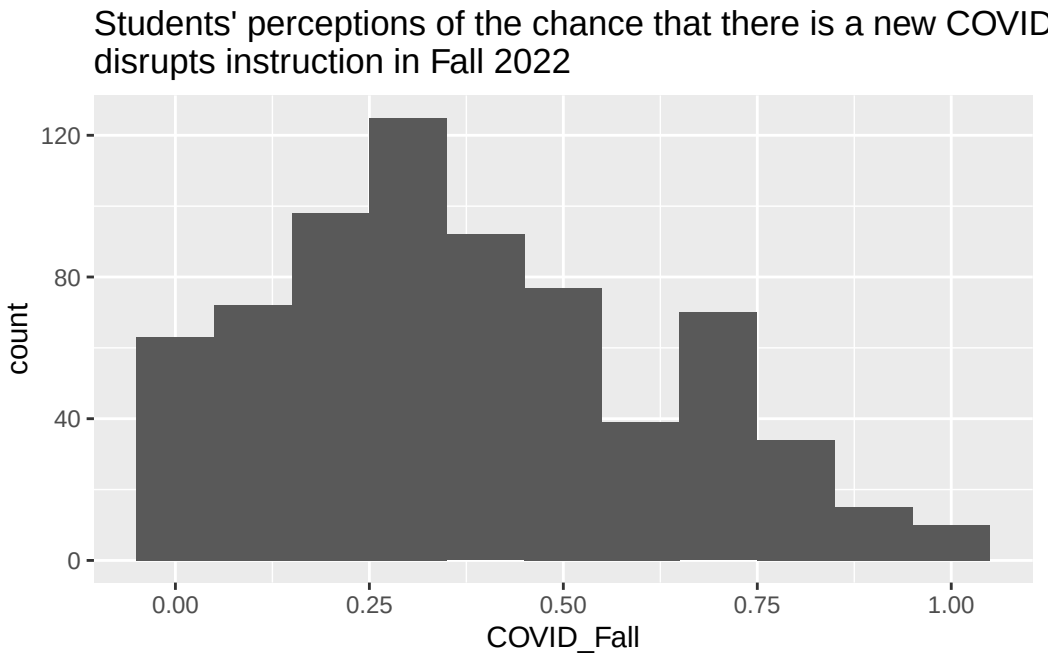
mode: "None"

Ans: Most students have none coding experience

10. **What are students' perceptions of the chance that there is a new COVID variant that disrupts instruction in Fall 2022?**

```
#ggplot(class_survey, aes(x = COVID_Fall)) + ggtitle("Students' perceptions of the #c

COVID = filter(class_survey, COVID_Fall <= 1)
ggplot(COVID, aes(x=COVID_Fall)) + ggtitle("Students' perceptions of the chance that
```




```
mean_Covid <- mean(class_survey$COVID_Fall)
median_Covid <- median(class_survey$COVID_Fall)
mode_Covid <- find_mode(class_survey$COVID_Fall)
```

mean: 14.773

median: 0.35

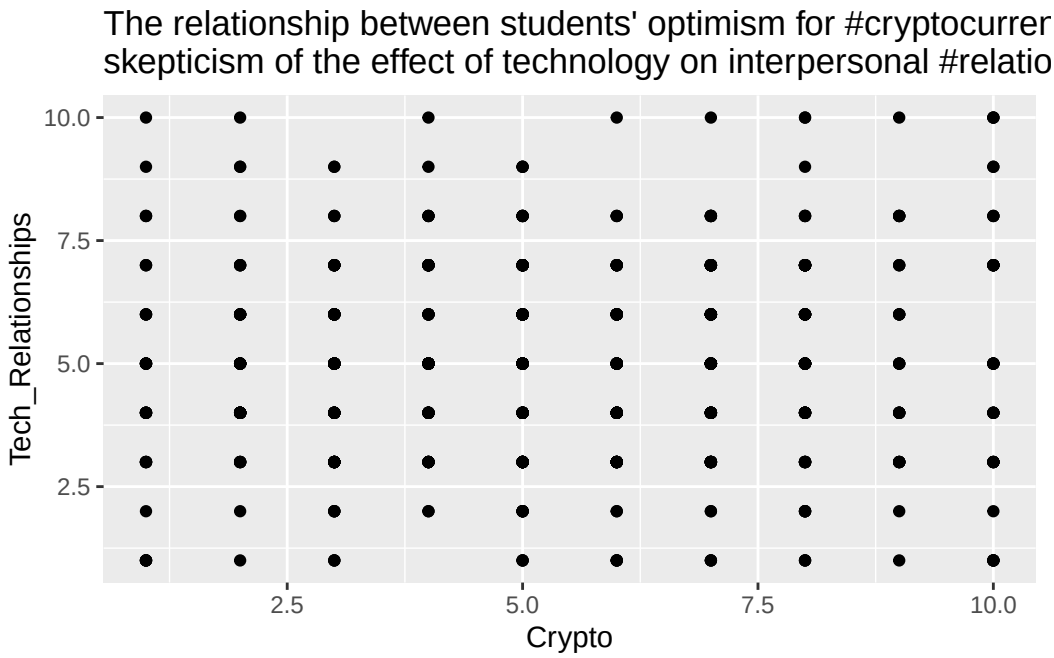
mode: 0.3

Ans: Most students' think that the probability that there is a new COVID variant that disrupts instruction in Fall 2022 is 0.3, which is not a high point.

11. **What is the relationship between students' optimism for cryptocurrency and their skepticism of the effect of technology on interpersonal relationships?**

```
ggplot(class_survey, aes(x = Crypto, y = Tech_Relationships)) + ggtitle("The relation
  geom_point()
```

```
Warning: Removed 2 rows containing missing values (geom_point).
```



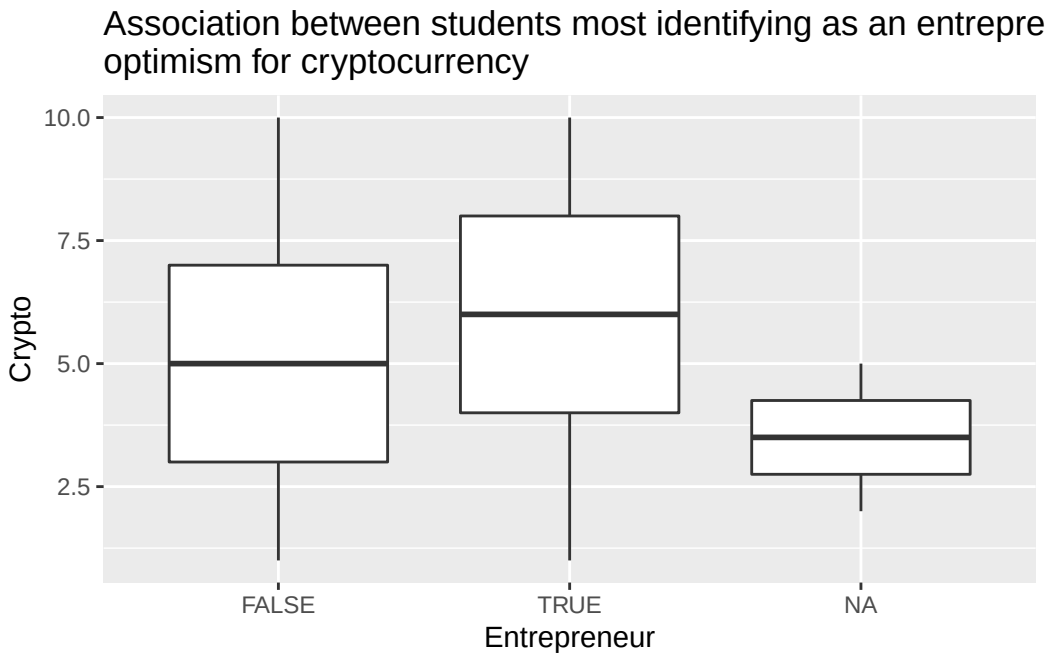
Ans: In this scatter plot, the dots are spread evenly, so we can conclude that students' optimism for cryptocurrency and their skepticism of the effect of technology on interpersonal relationships is low-related.

Part III: Extensions

12. **Is there an association between students most identifying as an entrepreneur and their optimism for cryptocurrency?**

```
ggplot(class_survey, aes(x = Entrepreneur, y = Crypto)) + ggtitle("Association betwee
```

```
Warning: Removed 2 rows containing non-finite values (stat_boxplot).
```

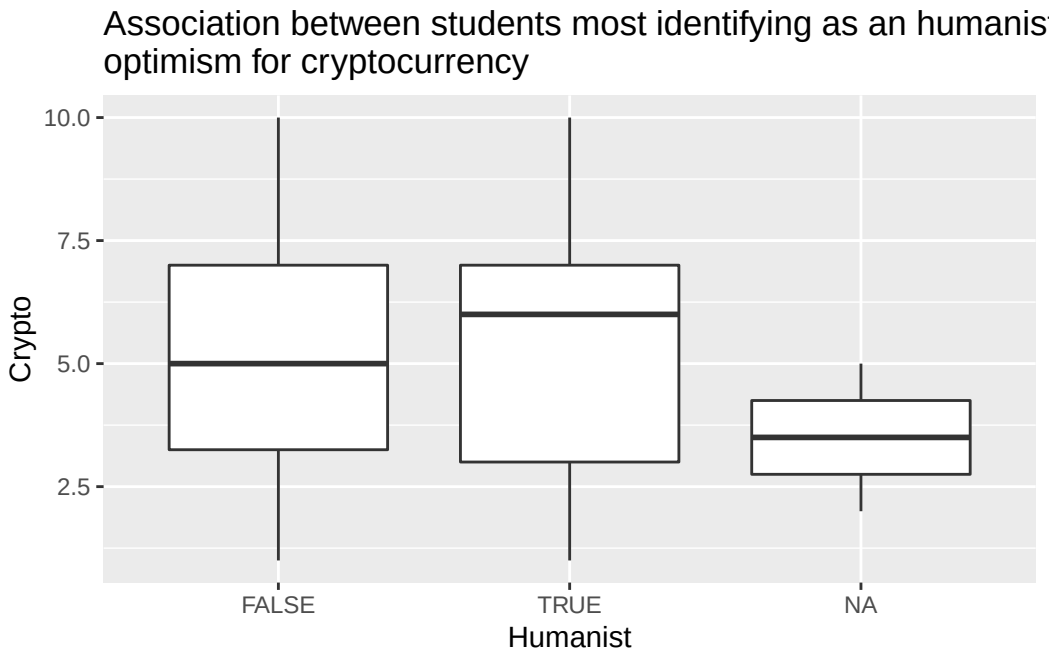


Yes, we can conclude that students who most identifying as a entrepreneur have more optimism for cryptocurrency because the box and median are higher than the box of students who are not entrepreneur.

13. **Is there an association between students most identifying as a humanist and their optimism for cryptocurrency?**

```
ggplot(class_survey, aes(x = Humanist, y = Crypto)) + ggtitle("Association between st
```

```
Warning: Removed 2 rows containing non-finite values (stat_boxplot).
```

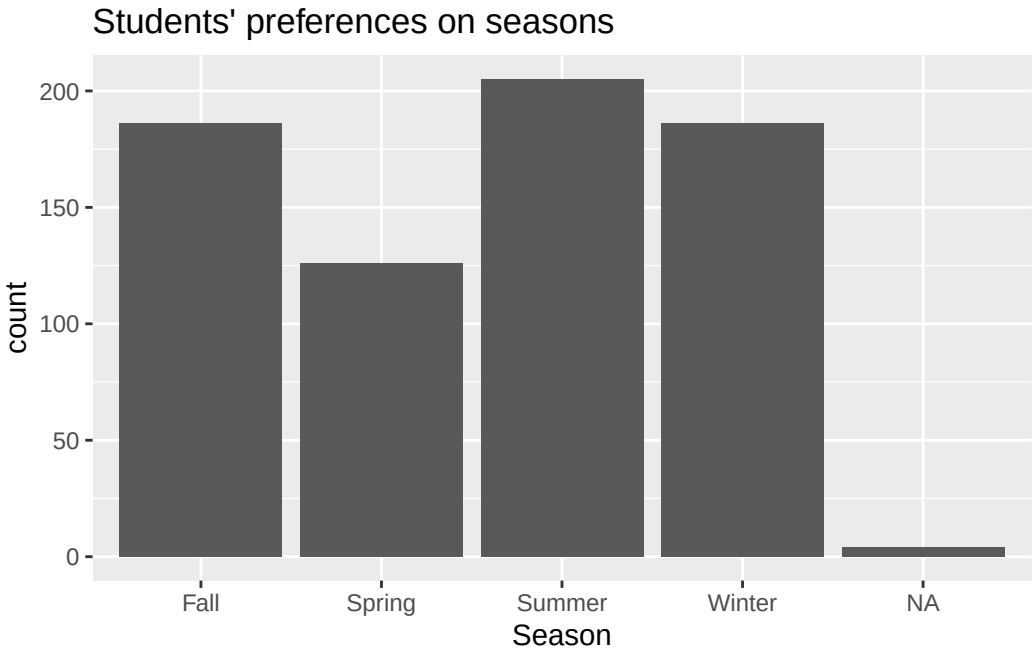


Yes but not high-related, we can see that only a few students who most identifying as a humanist have higher optimism for cryptocurrency because the median is higher than the box of students who are not humanist . The two boxes looked as the same, but the box of students who most identifying as a humanist has a higher median line.

14. **Propose your own question involving one or two variables and answer it using a plot with a written interpretation.**

    Which season did the students like the most?

```
ggplot(class_survey) + ggtitle("Students' preferences on seasons") +
  geom_bar(aes(x = Season))
```



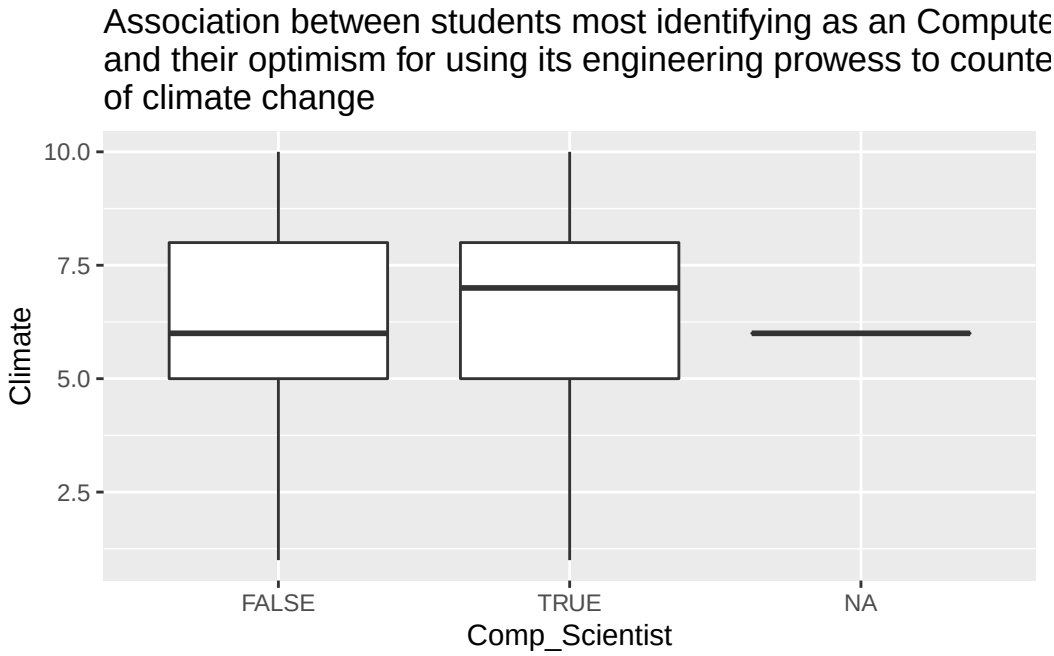
In this plot, we can say that students who took this survey like Summer the most.

15. **Propose your own question involving two or three variables and answer it using a plot with a written interpretation.**

    Is there an association between students most identifying as a Computer Scientist and their optimism for that humanity will be able to use its engineering prowess to counteract the effects of climate change?

```
ggplot(class_survey, aes(x = Comp_Scientist , y = Climate)) + ggtitle("Association be
```

```
Warning: Removed 3 rows containing non-finite values (stat_boxplot).
```



Yes but not high-related, both the boxes for Computer Scientist students and non-Computer Scientist students are at the same high. Even though the high are the same, the median are quite different, and this means that relatively more Computer Scientist students believe that humanity will be able to use its engineering prowess to counteract the effects of climate change.