Social Anxiety Survey

2023-10-23

Introduction

Social anxiety is a very common fear that appears to be growing, or at least getting more attention in these recent years. This is a concerning development due to the increasingly social nature of the academic field. The further we progress in our academic careers, the more we are made to socialize with colleagues, advisers, and made to give public presentations and speeches. This is an increasingly difficult task depending on the severity of someone's social anxiety. With how common social anxiety has become, it is important to look at factors that could lead to an increase in anxiety. Bringing this knowledge into the academic community could lead to better support for those suffering with social anxiety. The progressive nature of academic careers poses a unique challenge, as individuals may find themselves navigating a complex web of social engagements that can prove to be particularly difficult for those grappling with the effects of social anxiety. Furthering this knowledge could help students progress in certain academic avenues that were previously closed off due to fear of speaking in social settings. The data we will be looking at is from kaggle, with the link attached here: https://www.kaggle.com/datasets/carlsonhoo/university-social-anxiety-survey/data This results from this data set are not generalizable due to the sample of students being a group of college students recruited in Malaysia. However, the questions posed in the survey is a good starting point in how we can learn more about anxiety in varying social situations, as well as examining other factors that could lead to increased social anxiety. The dataset will give us a reference point to begin. We will see what information the current dataset can provide, and hope to reapply the questions posed here to make the results more applicable in the future. The question we will look at in this set is follows: Is there is a difference in the severity of social situations depending on ones age group? The age groups in this data set are separated into three groups: 20 or below, 21-30, and 31-40. I expect to find that students anxiety level will decrease by age. I believe that students 20 or below will have the most anxiety and students aged 31-40 will have the least anxiety with ages 21-30 being a middle ground between the two age groups.

Methods:

For my methods I will talk about my data journey that I compiled earlier in the semester. The data journey of the "Student Social Anxiety Survey" dataset that I retrieved from Kaggle.com is a relatively short one. The person that created this dataset was a university student from Malaysia. They created this dataset for their class and out of their own curiosity developed from past research articles. According to the Kaggle website, the inspiration for building this dataset was as follows: "More and more people reply on social media or mobile phone to communicate with others. It is interested to find out if the social media makes people develop social anxiety disorder or a platform for people with social anxiety people to feel more comfortable to communicate there" (Kaggle.com). They created an online survey and posted the QR code at multiple locations around their campus. The questions that the author used are based off the "Liebowitz Social Anxiety Scale (LSAS-SR)" (Hoo, Carlson). The online survey was conducted via google forms and exported to a csv file. The original intended use of the data was for the creator to determine whether the conveniences of social media and smartphones increased social anxiety in students, or if the lack of face-to-face communication attracted the attention of students who were already introverted or had social anxiety. I cannot find on Kaggle how the data might have otherwise been used. From what I can determine the data has not been changed or filtered. The creator took the information he received from google forms, put the

data into a spreadsheet, and uploaded it to Kaggle, where I believe it remains unchanged. It is important to know that we are working with the raw data the author collected. The first step I will have to take in cleaning the data is to remove the answers of the participants who labeled themselves as non-students. Then all I will need to do is run analysis on the rates of certain variables with other variables to get the information I wanted for my project. From what I can see, this will not create friction. Since my research focus is on students, then removing the non-students will not affect my work in any way. Some of the variables have been organized into graphs already which will facilitate the work I need to do in order to compare certain variables. The values that have been crystallized are gender, student status, marital status, and age range. A few of the other values are made clear with the responses being on a scale of one to five, with one being the least anxious in the situation asked from the question and 5 being the most anxious in the situation presented by the question. I do not believe there were any power relations that shaped what was included or not included in this data. This data does not appear to be fit for purpose, but instead will be used to inform my question. My research question is what factors/situations are most contributable to a student's social anxiety level, and which age population is more affected by these factors. The data provides information on anxiety levels in multiple common situations that students may encounter, as well as their age range and a few other background details that I may decide to include in my research. Overall, the data collected in "Student Social Anxiety Survey" will provide useful information in gathering a general idea in which situations students find themselves feeling more anxious, and who is more susceptible to these situations. Due to the small number of respondents in the dataset, we were unable to find any significance in any of the social situations when comparing age groups. However, we will be examining visual representations shown by bar graphs, and discussing any apparent trends and potential reasons for these trends. We will examine each relevant question presented in the survey in order, and do a comparison of answers across age groups.

Results:

We will be first giving an overview of the data to denote any missing data points, discrepancies, or anything that could potentially be concerning in analyzing the small pool of participants we have. I will conduct a few analysis per section of code and then break down each piece of information obtained.

```
skim(data_f)
```

Table 1: Data summary

Name	data_f
Number of rows	66
Number of columns	12
Column type frequency:	
character	6
numeric	6
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Timestamp	0	1	14	19	0	65	0
Gender	0	1	4	6	0	2	0
Student	0	1	2	3	0	2	0
Age	0	1	7	11	0	3	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Marital	0	1	6	7	0	2	0
Q1	0	1	25	25	0	5	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Q2	0	1	3.30	1.07	1	3.00	3	4	5	
Q3	0	1	3.89	1.12	1	3.00	4	5	5	
Q4	0	1	2.98	1.06	1	2.00	3	4	5	
Q5	0	1	3.45	1.04	1	3.00	4	4	5	
Q6	0	1	3.85	1.06	1	3.00	4	5	5	
Q7	0	1	3.18	1.20	1	2.25	3	4	5	

```
data_fr <- read_csv("Social_Anxiety_Survey_Master.csv") |>
  filter(Student == 'Yes')
```

```
## Rows: 66 Columns: 12
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (6): Timestamp, Gender, Student, Age, Marital, Q1
## dbl (6): Q2, Q3, Q4, Q5, Q6, Q7
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

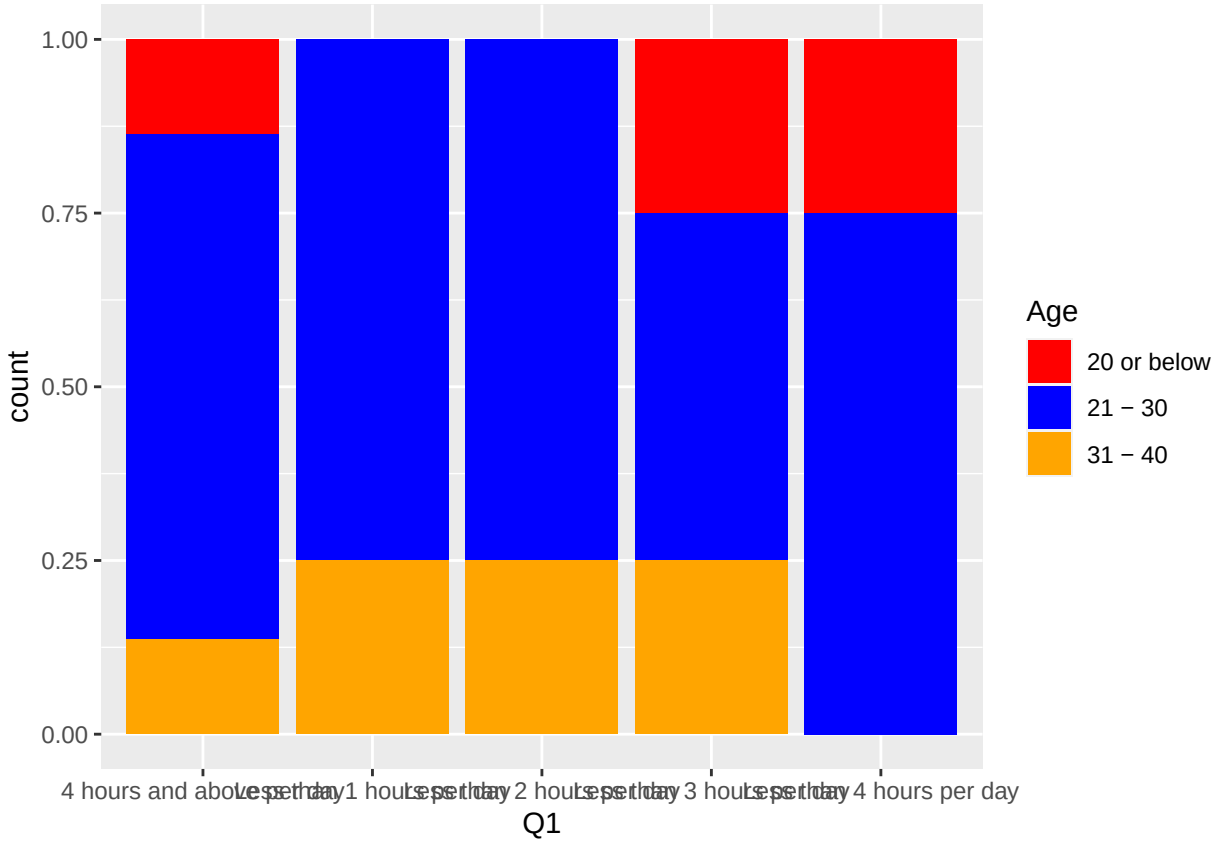
```
data_fr |>
    count(Age)
```

```
## # A tibble: 3 x 2
##   Age             n
##   <chr>       <int>
## 1 20 or below     6
## 2 21 - 30        29
## 3 31 - 40         7
```

The skim function gives an overview of the data allowing us to check for any missing data points, as well as any large discrepancies that we could attend to for future examination of the data. As we can view from the table output above, there are no missing values in any categories. However, we do notice that there are respondents to the survey who claimed they are not students. Due to my research question focusing on students, we will be filtering out any respondents that said no to the "are you a student" question from now on and naming that new data frame data_fr (previously called data_f). After filtering out the non-students, we have a distribution of 6 respondents who are 20 or below, 29 that are in the age range of 21-30, and 7 who are 31-40 years of age for a total of 42 respondents (previously 60). As discussed, this is an extremely small pool of data, so we will be mostly focusing on trends instead of any significance. If you would like to see some more general data that is not pertinent to the research question, there is code in the script to view summary, head and tail data, and a visualization denoting that participants responded to all questions. Next we compared the accuracy of the dataset to an external source for validity. It was found that Carlson Hoo the creator of the data set, also created a blog post detailing his own analysis and interpretations of the data.

In this blog post it was noticed that they had 69 student participants while we only have 42 in the uploaded dataset. There was no description left by Carlson as to why there is a data discrepancy. Due to this I have inferred that more people answered his survey after he already uploaded the dataset, and did not see a reason to update his spreadsheet. We will continue with the current dataset, as an updated one is not provided. The blogsite can be viewed here: (https://carlson-hoo.medium.com/what-type-of-social-media-users-are-more-likely-to-have-social-anxiety-disorder-65194323f8e1). In the attached script, I provided count visualizations for the total amount of respondents to each question. Since this section will be repeated later separating respondents by age group, I will not go over these visualizations, however you can view them for personal interest. Next we will set the color bar for the results as well as begin examine each question presented. This will provide the percentage of respondents from each age group for each potential answer to the question. We will go into potential interpretations in the discussion section. The first question is "How many hours per day do you spend on social media?" We can view the distribution in the code chunk below. This chunk of code also creates a percentage function so we can see clearly see the percentage of each answer separated by the age group. If no value appears then that means there were 0 respondents for that answer in a specific age group.

```
color_scale = scale_fill_manual(
    values = c('red', 'blue', 'orange', 'black', 'purple'))
#This sets the color scale for each of my graphs.
data_fr |>
  ggplot(aes(Q1, fill = Age)) +
  geom_bar(position = position_fill()) +
  color_scale
```



```
#Due to the x-axis being indiscernible in the view, I will list the answers from left to right.
  #4 hours and above, less than 1 hour per day, less than 2 hours per day, less than 3 hours per day, l
percentage = function(question) {
    data_fr |>
        count(Age, {{ question }}) |>
        group_by(Age) |>
        mutate(share = n / sum(n)) |>
        ungroup()|>
        arrange(desc({{ question }}))|>
        mutate(share = scales::percent(share, accuracy = 1))
}
```

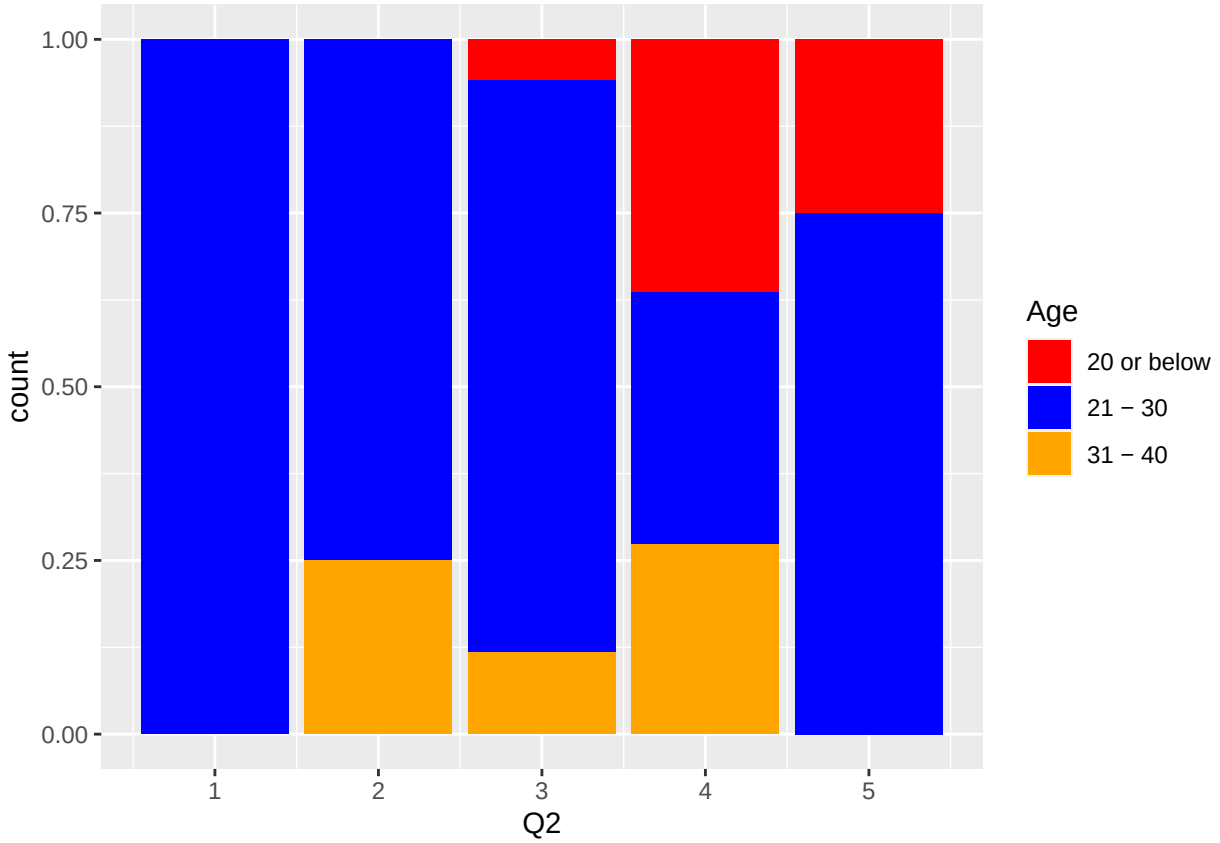
```
percentage(Q1)
```

```
## # A tibble: 12 x 4
##    Age         Q1                            n share
##    <chr>       <chr>                     <int> <chr>
##  1 20 or below Less than 4 hours per day     1 17%
##  2 21 - 30     Less than 4 hours per day     3 10%
##  3 20 or below Less than 3 hours per day     2 33%
##  4 21 - 30     Less than 3 hours per day     4 14%
##  5 31 - 40     Less than 3 hours per day     2 29%
##  6 21 - 30     Less than 2 hours per day     3 10%
##  7 31 - 40     Less than 2 hours per day     1 14%
##  8 21 - 30     Less than 1 hours per day     3 10%
##  9 31 - 40     Less than 1 hours per day     1 14%
## 10 20 or below 4 hours and above per day     3 50%
## 11 21 - 30     4 hours and above per day    16 55%
## 12 31 - 40     4 hours and above per day     3 43%
```

The x-axis are listed from left to right in the code chunk above. As we can see above, the ages of 21-30 and 31-40 are pretty widespread across all categories. However, 20 and below have 50% of their answers in 4 hours or more per day with the rest of their distribution in 4 hours or less, and 3 hours or less.

The next question is "How do you feel about being the center of attention?" The rest of these responses are based of a scale of 1-5, where 1 would be the least anxious, and 5 would be the most anxious.

```
data_fr |>
  ggplot(aes(Q2, fill = Age)) +
  geom_bar(position = position_fill()) +
  color_scale
```




```
percentage(Q2)
```

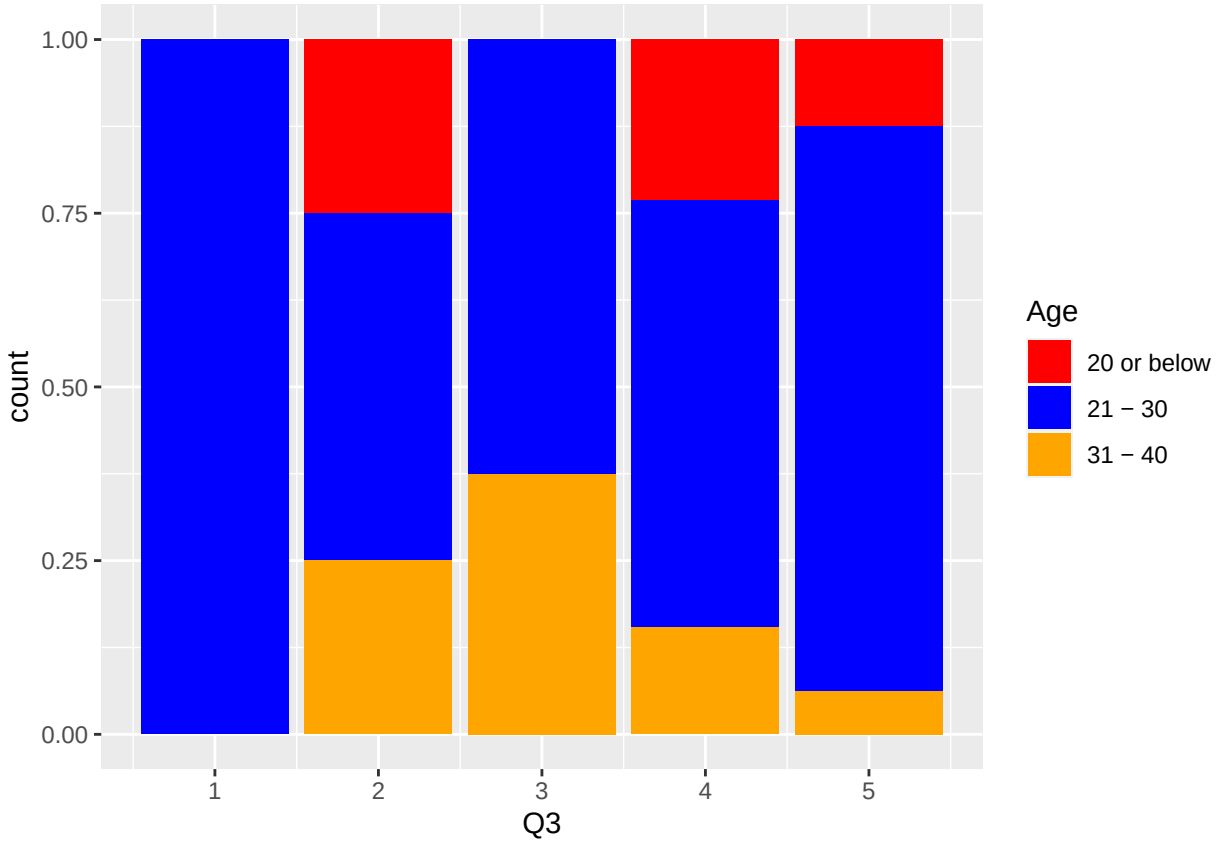
```
## # A tibble: 11 x 4
##    Age            Q2     n share
##    <chr>       <dbl> <int> <chr>
##  1 20 or below     5     1 17%
##  2 21 - 30         5     3 10%
##  3 20 or below     4     4 67%
##  4 21 - 30         4     4 14%
##  5 31 - 40         4     3 43%
##  6 20 or below     3     1 17%
##  7 21 - 30         3    14 48%
##  8 31 - 40         3     2 29%
##  9 21 - 30         2     6 21%
## 10 31 - 40         2     2 29%
## 11 21 - 30         1     2 7%
```

Ages 31-40 chose neither extreme, and are split across the middle sections with 41% answering exactly in the middle(3). Ages 20 and below had the 4 response which was 67% of that age group. Ages 21-30 also had the majority of their answers as 3.

Question 3 asks the anxiety level you feel in "Working while being observed"?

```
data_fr |>
  ggplot(aes(Q3, fill = Age)) +
```

```
  geom_bar(position = position_fill()) +
  color_scale
```



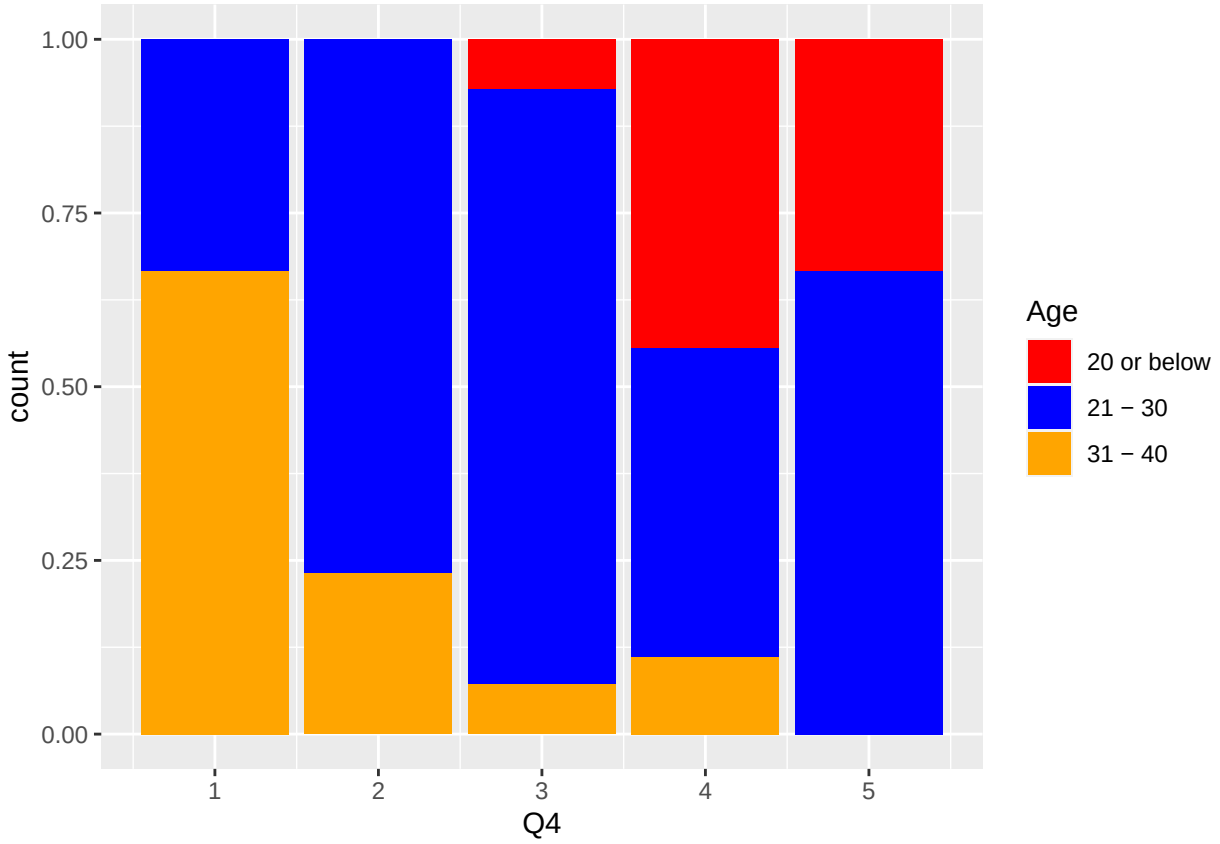

```
percentage(Q3)
```

```
## # A tibble: 12 x 4
##    Age            Q3     n share
##    <chr>       <dbl> <int> <chr>
##  1 20 or below     5     2 33%
##  2 21 - 30         5    13 45%
##  3 31 - 40         5     1 14%
##  4 20 or below     4     3 50%
##  5 21 - 30         4     8 28%
##  6 31 - 40         4     2 29%
##  7 21 - 30         3     5 17%
##  8 31 - 40         3     3 43%
##  9 20 or below     2     1 17%
## 10 21 - 30         2     2 7%
## 11 31 - 40         2     1 14%
## 12 21 - 30         1     1 3%
```

Ages 21-30 had the majority of their answers in the 4 or 5 category with a 45% of respondents saying they feel the most anxious in this situation. 20 or below had similar results with a majority responding a 4. 31-40 was spread across the scale with a majority in responding 3.

Question 4 asks the anxiety level you fear when "Talking face to face with someone you don't know very well"?

```
data_fr |>
  ggplot(aes(Q4, fill = Age)) +
  geom_bar(position = position_fill()) +
  color_scale
```



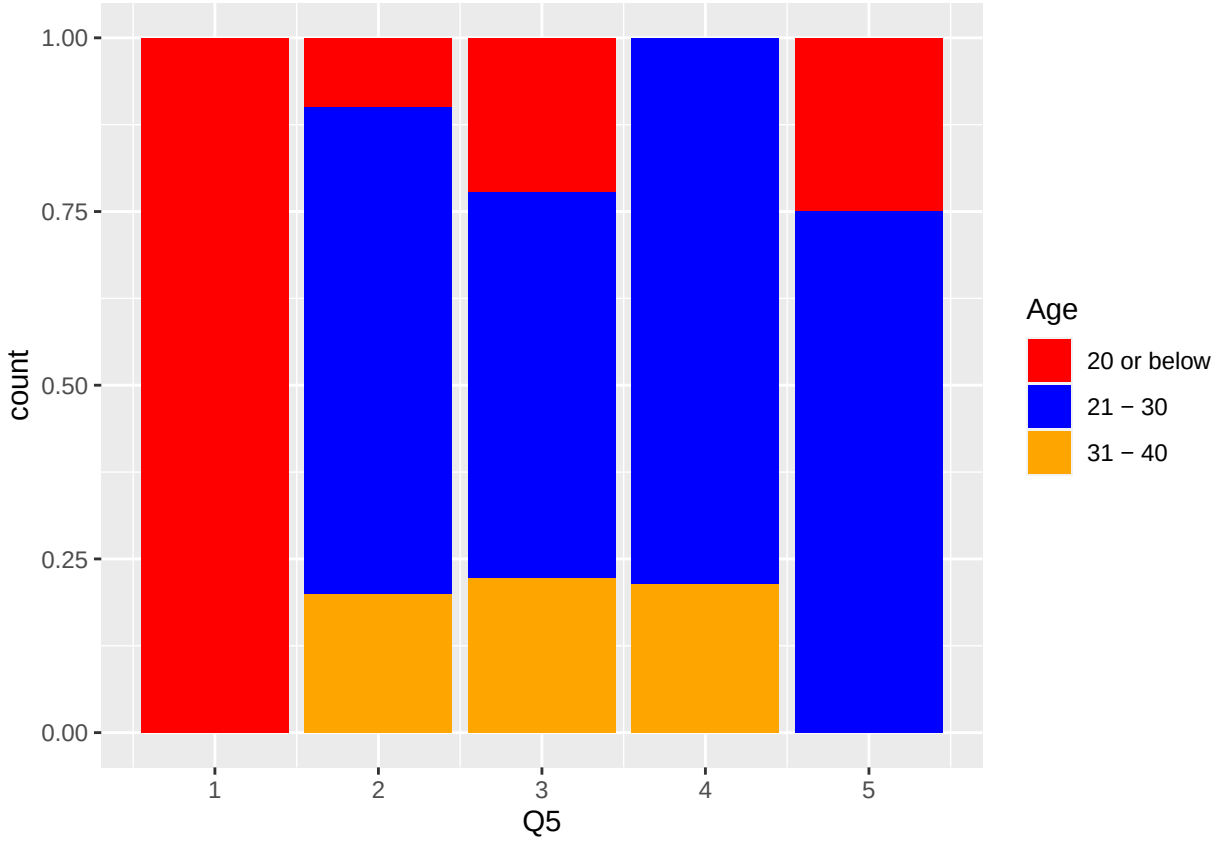

```
percentage(Q4)
```

```
## # A tibble: 12 x 4
##    Age            Q4     n share
##    <chr>       <dbl> <int> <chr>
##  1 20 or below     5     1 17%
##  2 21 - 30         5     2 7%
##  3 20 or below     4     4 67%
##  4 21 - 30         4     4 14%
##  5 31 - 40         4     1 14%
##  6 20 or below     3     1 17%
##  7 21 - 30         3    12 41%
##  8 31 - 40         3     1 14%
##  9 21 - 30         2    10 34%
## 10 31 - 40         2     3 43%
## 11 21 - 30         1     1 3%
## 12 31 - 40         1     2 29%
```

For question 4, ages 31-40 are skewed to the left side of the scale, with over 70% of responses being 1 or 2. Ages 20 and below have a total 84% of responses in the 4 and 5 section. 21-30 was mostly skewed across the left middle range of 2-3, totaling about 75% for that age range.

Next, question 5 asks to report anxiety levels when "Expressing disagreement or disapproval to someone you don't know very well".

```
data_fr |>
  ggplot(aes(Q5, fill = Age)) +
  geom_bar(position = position_fill()) +
  color_scale
```




```
percentage(Q5)
```

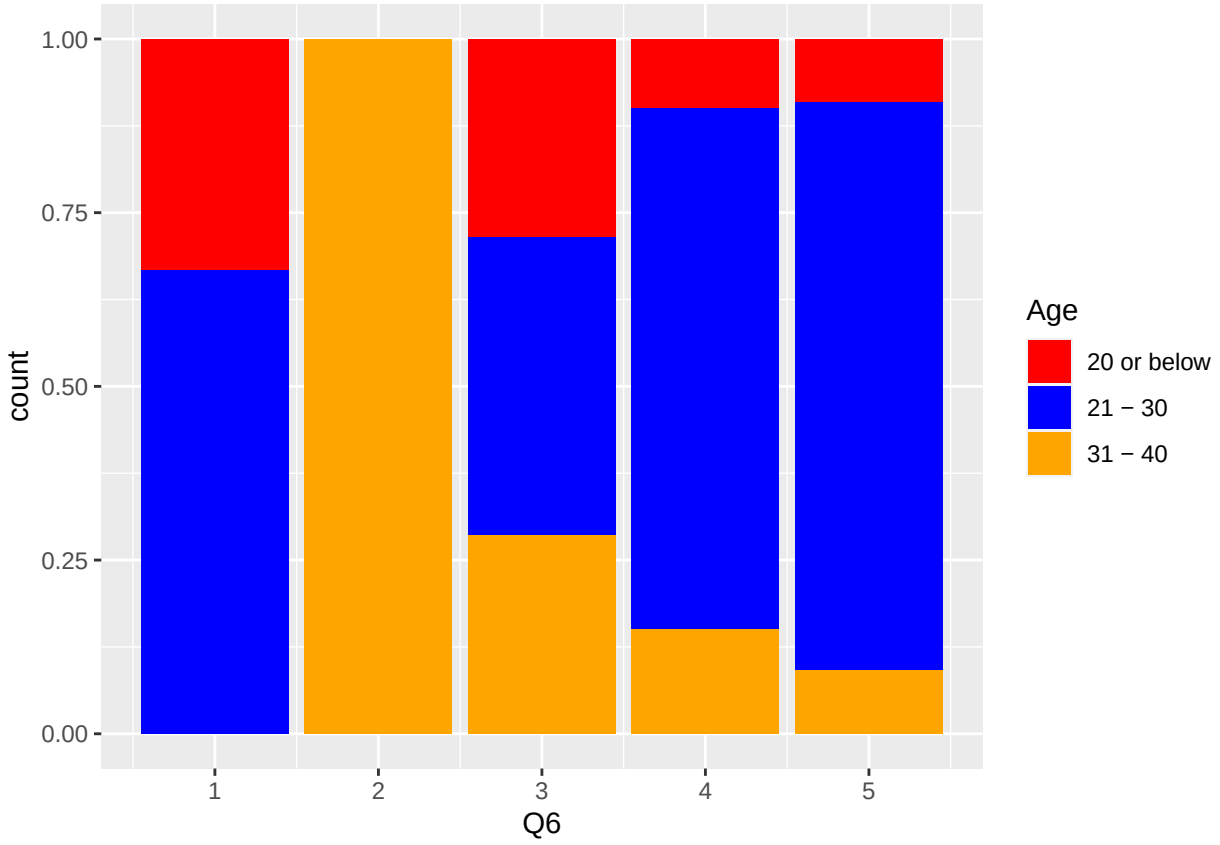
```
## # A tibble: 11 x 4
##    Age             Q5     n share
##    <chr>        <dbl> <int> <chr>
##  1 20 or below     5     2 33%
##  2 21 - 30         5     6 21%
##  3 21 - 30         4    11 38%
##  4 31 - 40         4     3 43%
##  5 20 or below     3     2 33%
##  6 21 - 30         3     5 17%
##  7 31 - 40         3     2 29%
##  8 20 or below     2     1 17%
```

```
##  9 21 - 30          2     7 24%
## 10 31 - 40          2     2 29%
## 11 20 or below      1     1 17%
```

As shown above, the distribution of all age groups are pretty split across the scale for this response. The largest noticeable trend from this section is that 43% of the 31-40 age group chose an anxiety level of 4. The other 2 groups do not have any largely skewed responses for this section.

Question 6 asks ""Are you extremely conscious of your actions when in social settings because you fear they might offend someone or you could be rejected?"

```
data_fr |>
  ggplot(aes(Q6, fill = Age)) +
  geom_bar(position = position_fill()) +
  color_scale
```




```
percentage(Q6)
```

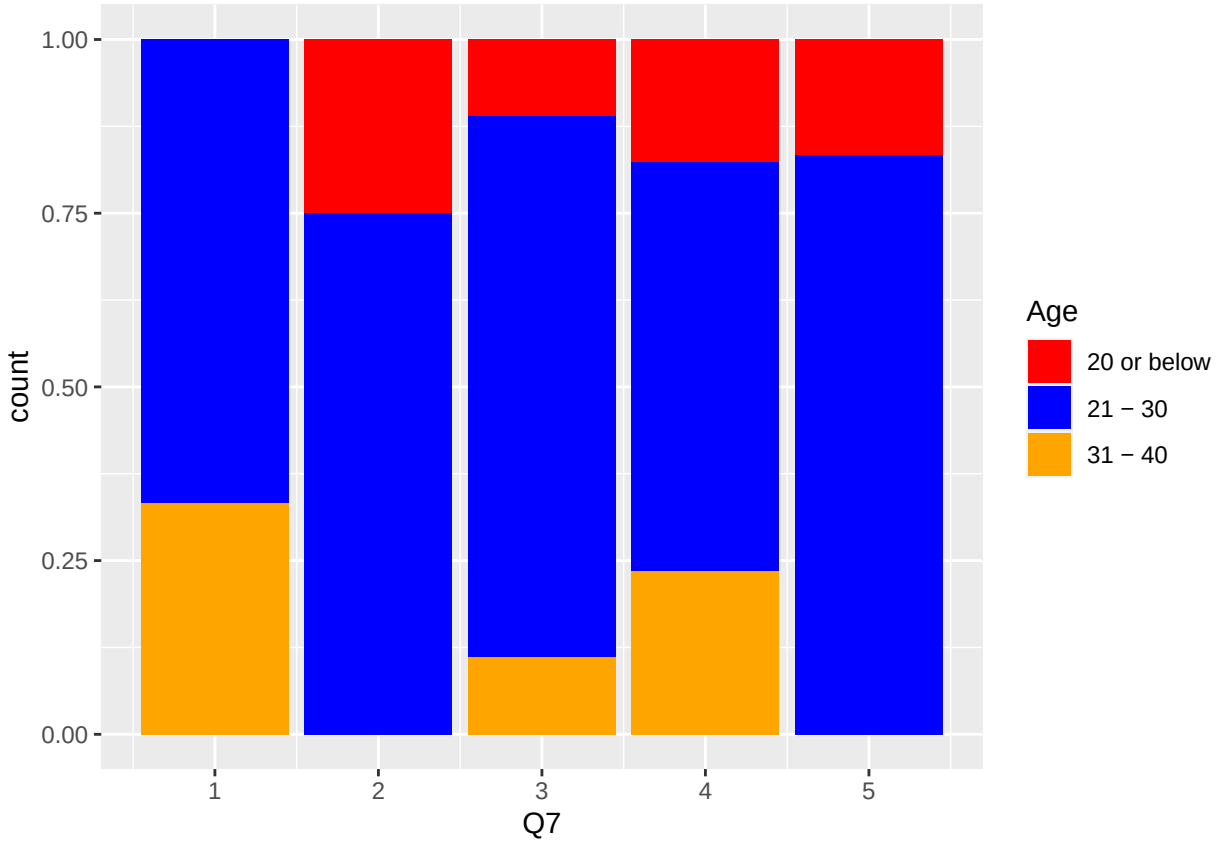
```
## # A tibble: 12 x 4
##    Age            Q6     n share
##    <chr>       <dbl> <int> <chr>
##  1 20 or below     5     1 17%
##  2 21 - 30         5     9 31%
##  3 31 - 40         5     1 14%
##  4 20 or below     4     2 33%
```

```
##  5 21 - 30          4    15 52%
##  6 31 - 40          4     3 43%
##  7 20 or below      3     2 33%
##  8 21 - 30          3     3 10%
##  9 31 - 40          3     2 29%
## 10 31 - 40          2     1 14%
## 11 20 or below      1     1 17%
## 12 21 - 30          1     2 7%
```

For the 21-30 age group, they had a majority of 83% responding either 4 or 5 to this question. 20 or below had a middle to right skew of their responses being 66% for the 3 and 4 anxiety levels.31-40 had the majority respond either a 2 or 3.

Lastly, question 7 asked the repondents to report "Do you feel anxious or panicky before social situations?"

```
data_fr |>
  ggplot(aes(Q7, fill = Age)) +
  geom_bar(position = position_fill()) +
  color_scale
```




```
percentage(Q7)
```

```
## # A tibble: 12 x 4
##    Age            Q7     n share
##    <chr>       <dbl> <int> <chr>
```

```
##  1 20 or below     5     1 17%
##  2 21 - 30         5     5 17%
##  3 20 or below     4     3 50%
##  4 21 - 30         4    10 34%
##  5 31 - 40         4     4 57%
##  6 20 or below     3     1 17%
##  7 21 - 30         3     7 24%
##  8 31 - 40         3     1 14%
##  9 20 or below     2     1 17%
## 10 21 - 30         2     3 10%
## 11 21 - 30         1     4 14%
## 12 31 - 40         1     2 29%
```

For the final question, the 21-30 age group was split across all levels. 31-40 had the majority of 57% that answered 4. The 20 or below age group also had 50% respond to 4 for this question.

Q1 is ("How many hours per day do you spend on social media?")

Q2 is ("How do you feel being the center of attention")

Q3 is ("Working while being observed?")

Q4 is ("Talking face to face with someone you don't know very well?")

Q5 is ("Expressing disagreement or disapproval to someone you don't know very well")

Q6 is ("Are you extremely conscious of your actions when in social settings because you fear they might offend someone or you could be rejected?")

Q7 is ("Do you feel anxious or panicky before social situations?")

Discussion:

I repeated the questions in order above for an easy reference. While the results are not meant to present significance in any categories due to the small number of participants, we will discuss the trends that were mentioned and the potential reasons for them. I will be grouping questions in pairs for each section to improve readability, with question 1 being responded to in this section, as it is somewhat of an outlier question. For question 1 we only had the 20 and below age group have a majority spending 4 or more hours per day on social media. Though this question does not reflect anxiety level, it will be interesting to see in future studies where we can measure statistical significance, how time spent on social media could correlate to anxiety levels in these varying social situations, and how it could impact the younger generation that spends more time using social media than others. The most notable trend in question 2 was for the age group 20 and below with a majority responding with an anxiety level of 4. Based on the trend of the rest of the age groups being skewed across the middle, this could be interpreted as lack of experience in this situation. This does follow the trend of my hypothesis that anxiety levels in social situations will decrease with age. For question 3, the 21-30 age group had the highest anxiety response. This could be interpreted in multiple fashions. However, this is the age when respondents may be entering their first serious jobs, as well as being later in their academic career, whereas the 20 and below age group may not have that experience yet, and the 31-40 age group has more experience in this situation. Since the 21-30 age group reported more anxiety than the 20 and below group for this question, the trend does not fit my hypothesis. Question 4 follows my hypothesis quite well with anxiety levels decreasing as age increases, with the 20 and below having a large majority answering 4 or 5. This is one of the questions that would be interesting to see the correlational effects of time spent on social media. It could be interpreted that the longer time spent on social media in this age group has had a negative impact on one's ability to comfortably talk to others in real life conversations. In question 5 the only noticeable trend was in the 31-40 age group having reportedly higher anxiety in this situation. This goes against my hypothesis. The only interpretable idea that comes

to mind for this trend, is that the current older age group was not very well received my older generations when expressing their emotions, specifically when they were unhappy. This reasoning could be correlated to the trend witnessed. In question 6, the noticeable trend was in the 21-30 age group having an 83% response to the 4 and 5 categories, with 20 and below having a similar skew at 66%. Both younger groups responded to higher anxiety than the 31-40 age group which does follow the hypothesis. However, I lack any current insight as to why this trend occurs. Perhaps with more data and a few additional questions, there would be more interpretations that could be made. Lastly question 7, had a surprising trend, and did not follow the hypothesis. The 31-40 age group and the 20 and below age group responded on average higher anxiety in this situation than the 21-30 group. For the 20 and below age group it would again be interesting to compare social media time with this trend. This could not be interpreted as a lack of experience in social situations as the 31-40 age group also reported high anxiety. Overall, the data was not fit for purpose. I believe the questions posed would make for an interesting comparison if we had a much higher and diverse participant pool. However, my hypothesis did prove correct on about 50% of the questions. There are multiple interpretations for each of these situations, and I attempted to correlate those responses in a fit to the trends of the data. Without more participants, measuring any real significance is difficult, as well as drawing any definitive conclusions other than the noticeable trends observed. For future studies, I hope more researchers conduct these kinds of surveys related to anxiety to increase our awareness as students, faculty, and scholars of the problems of anxiety in every age group, and how it can impact our personal and academic career. By shedding light on the intricacies of social anxiety in academic settings, we pave the way for a more empathetic and supportive academic community. This knowledge can support the establishment of developing tailored interventions that not only acknowledge the prevalence of social anxiety but also actively work towards dismantling the barriers it poses. Ultimately, creating an environment that prioritizes mental health and well-being can empower students to overcome social anxiety, unlocking pathways to academic success and ensuring that a student's potential is not unreached due to the constraints that anxiety can place on the individual.