# DIG HUM 101 Project Analysis

2025-10-13

Potential Research Question: How did mental health during the COVID-19 pandemic impact academic development?

```
library(ggplot2)
students_2 <- read.csv("~Alan Ruiz/Desktop/COVID-19 Survey Student Responses (1) (1).csv")
colnames(students_2)
```

```
##  [1] "ID"
##  [2] "Region.of.residence"
##  [3] "Age.of.Subject"
##  [4] "Time.spent.on.Online.Class"
##  [5] "Rating.of.Online.Class.experience"
##  [6] "Medium.for.online.class"
##  [7] "Time.spent.on.self.study"
##  [8] "Time.spent.on.fitness"
##  [9] "Time.spent.on.sleep"
## [10] "Time.spent.on.social.media"
## [11] "Prefered.social.media.platform"
## [12] "Time.spent.on.TV"
## [13] "Number.of.meals.per.day"
## [14] "Change.in.your.weight"
## [15] "Health.issue.during.lockdown"
## [16] "Stress.busters"
## [17] "Time.utilized"
## [18] "Do.you.find.yourself.more.connected.with.your.family..close.friends...relatives..."
## [19] "What.you.miss.the.most"
```

- There are two categories for residence. Inside Delhi-NCR and outside Delhi-NCR.
- What kind of city or region is Delhi-NCR? Culture, economy, population, geography, access to technology?
- Are there any significant differences with outside Delhi-NCR that would impact our interpretation of the data?

```
unique(students_2[["Region.of.residence"]])
```

```
## [1] "Delhi-NCR"         "Outside Delhi-NCR"
```

- These students we are analyzing aren't all one age.
- Minimum age is 7 and maximum is 59.
- Do we want to filter data to focus on a certain age group?
- Age may affect their views on the pandemic. Something to consider.

```r
min(students_2$Age.of.Subject)
```

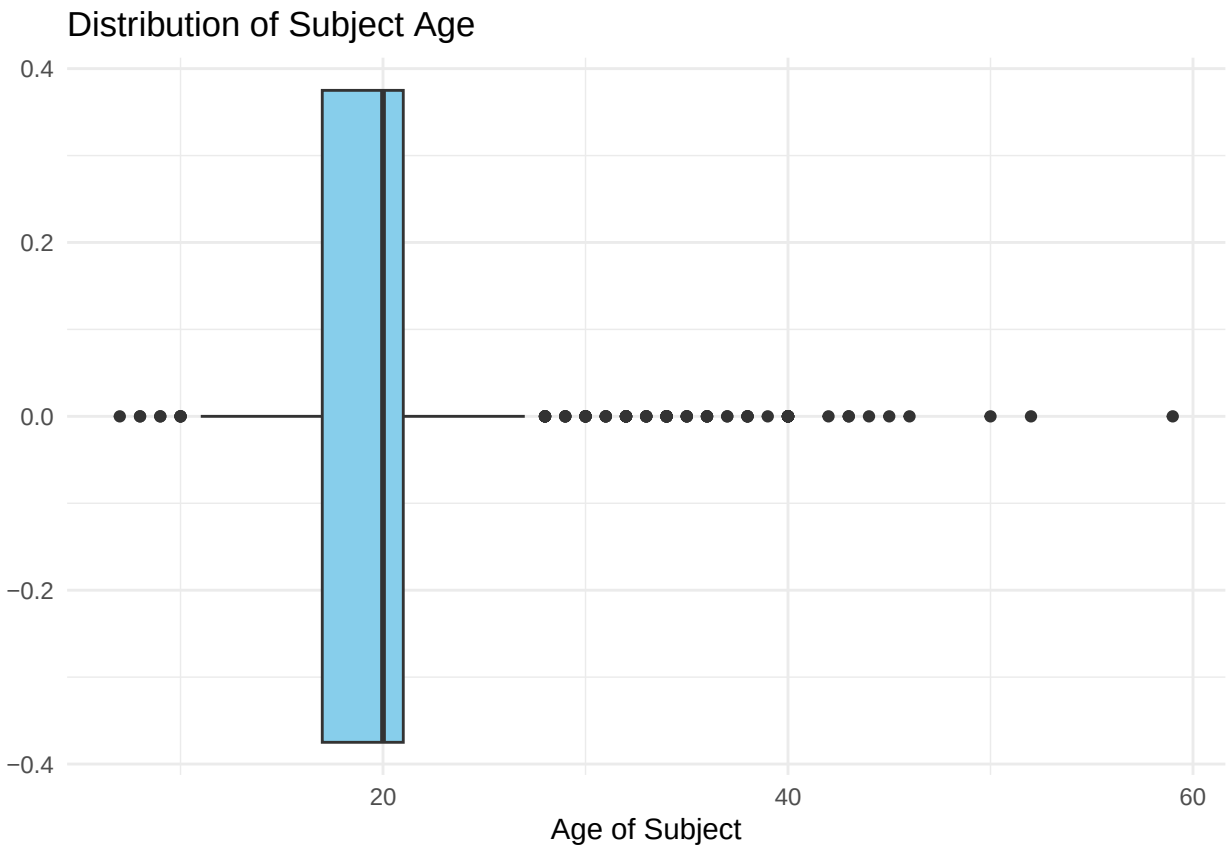
```
## [1] 7
```

```r
max(students_2$Age.of.Subject)
```

```
## [1] 59
```

- The majority of students represented in this dataset is around 17-21 (high school to college)
- How does that impact our analysis of their responses?
- What do we do about the outliers? Should we primarily focus on students found within IQR?

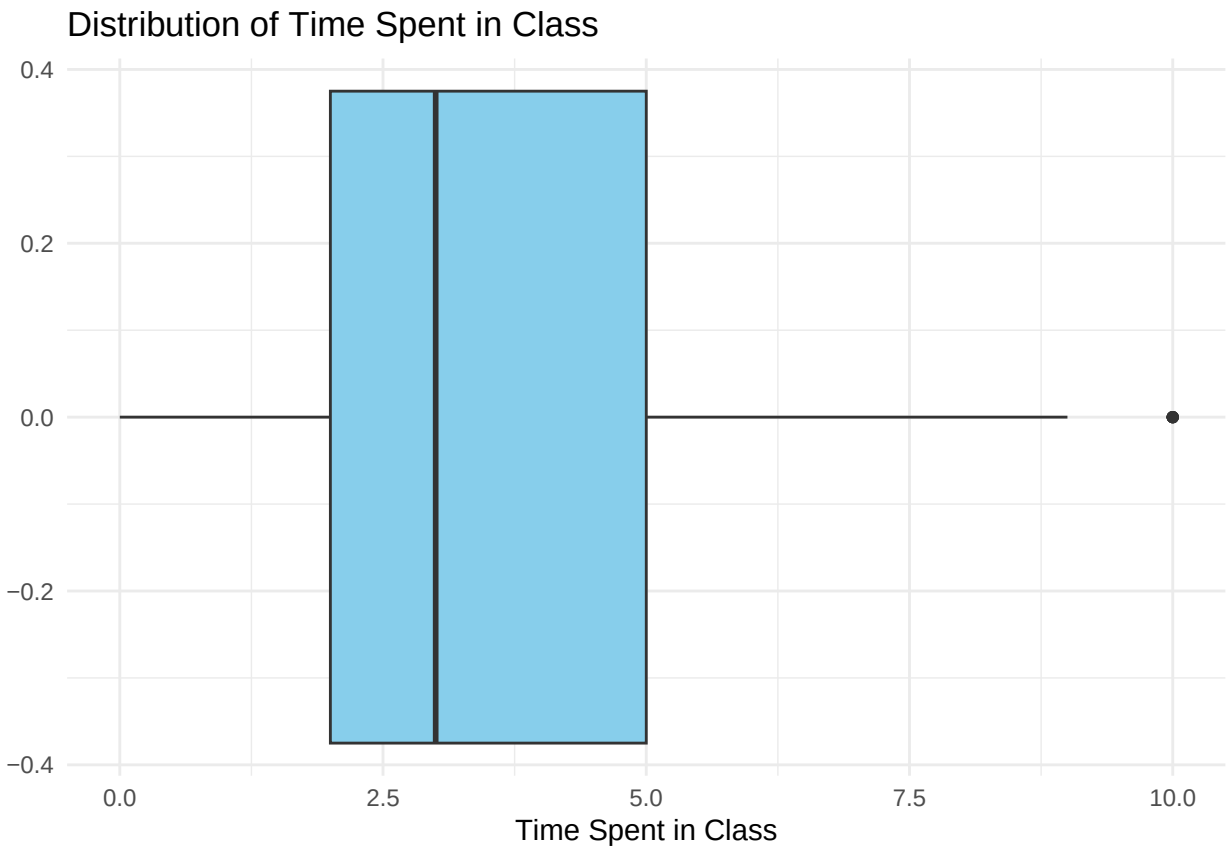
```r
ggplot(data = students_2) + aes(x = Age.of.Subject) +
        geom_boxplot(fill = "skyblue") +
        labs(
          title = "Distribution of Subject Age",
          x = "Age of Subject") +
        theme_minimal()
```




- This variable is about time spent on class. Does that mean per week?
- Also, I'm assuming the unit is hours
- How many classes did the students have? That would affect how much time they spent on online class

- Keep a close eye on the 132 students who spent 0 time on class. What was their rating of the experience? What were they doing instead? How was their emotional health?

```
ggplot(data = students_2, aes(x = Time.spent.on.Online.Class)) +
         geom_boxplot(fill = "skyblue") +
         labs(
           title = "Distribution of Time Spent in Class",
           x = "Time Spent in Class") +
         theme_minimal()
```

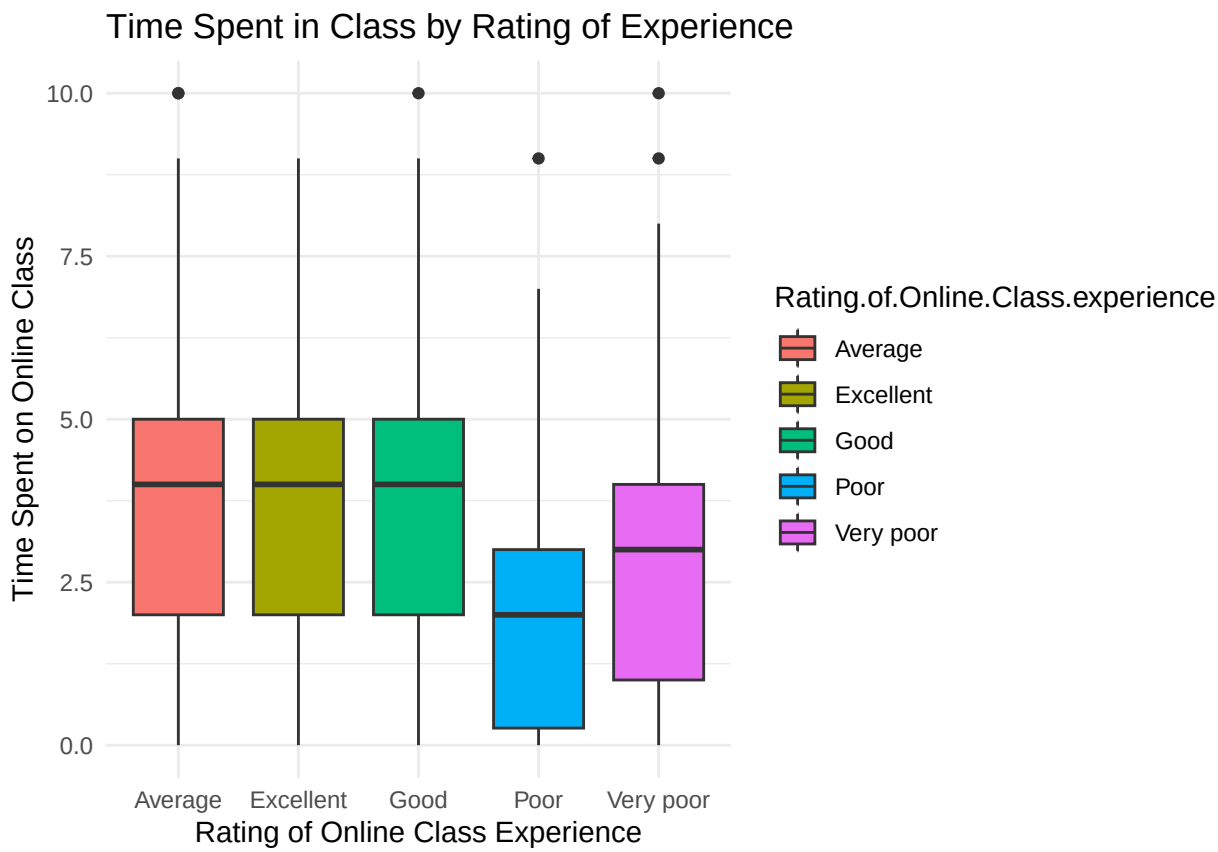



```
# Dataframe for students who spent 0 hours on online class
sum(students_2$Time.spent.on.Online.Class == 0)
```

```
## [1] 132
```

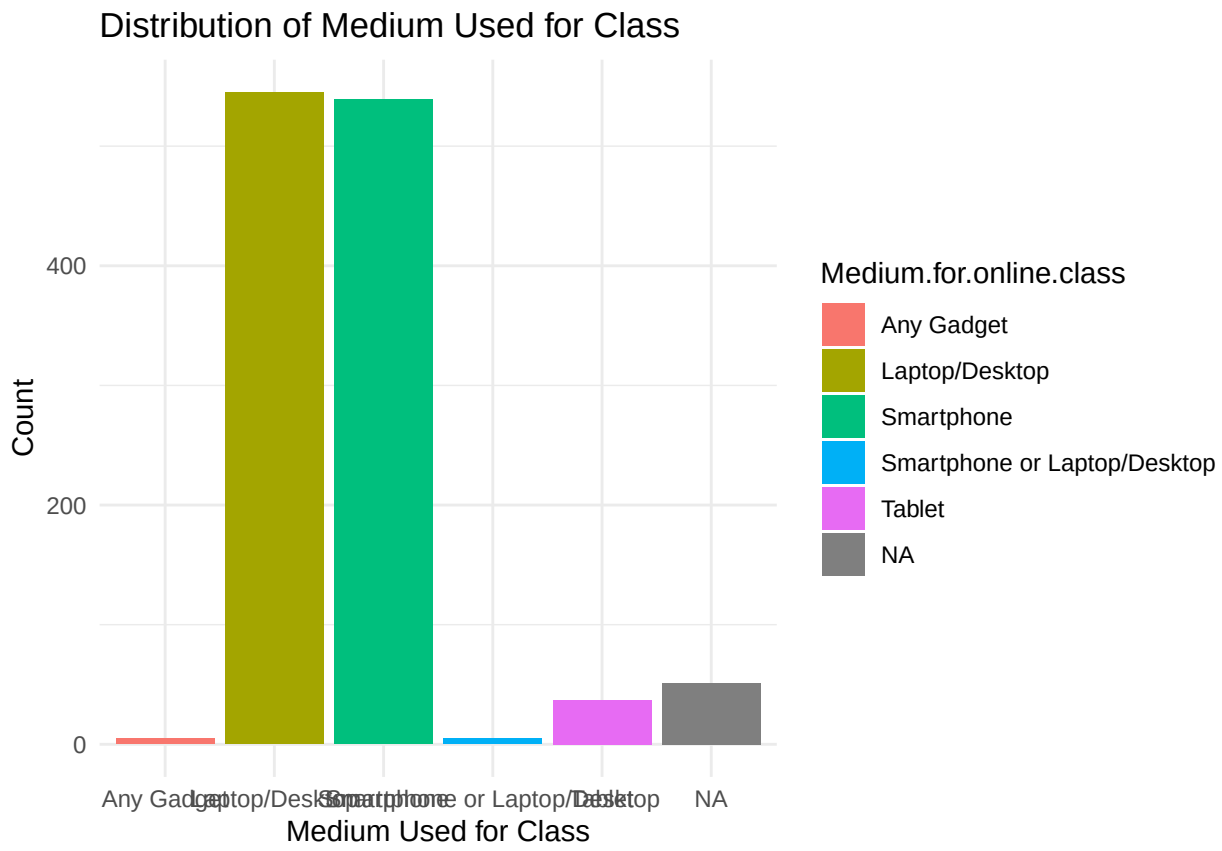
- It's notable that the students who gave a negative sentiment about their experience spent the least time in online class.
- Did they give a negative review because they didn't spend enough time in class?
- Were they not in class as much because they didn't like the format, or do they just not like school in general?
- If they just don't like school in general, was their rating of the experience really because of the online format?
- Note: I bypassed the students who put NA for their experience review, but we could include them if it adds more context

```r
ggplot(students_2[!is.na(students_2$Rating.of.Online.Class.experience), ]) +
      aes(x = Rating.of.Online.Class.experience,
                          y = Time.spent.on.Online.Class,
                          fill = Rating.of.Online.Class.experience) +
    geom_boxplot(na.rm = TRUE) +
    labs(
      title = "Time Spent in Class by Rating of Experience",
      x = "Rating of Online Class Experience",
      y = "Time Spent on Online Class"
    ) +
    theme_minimal()
```



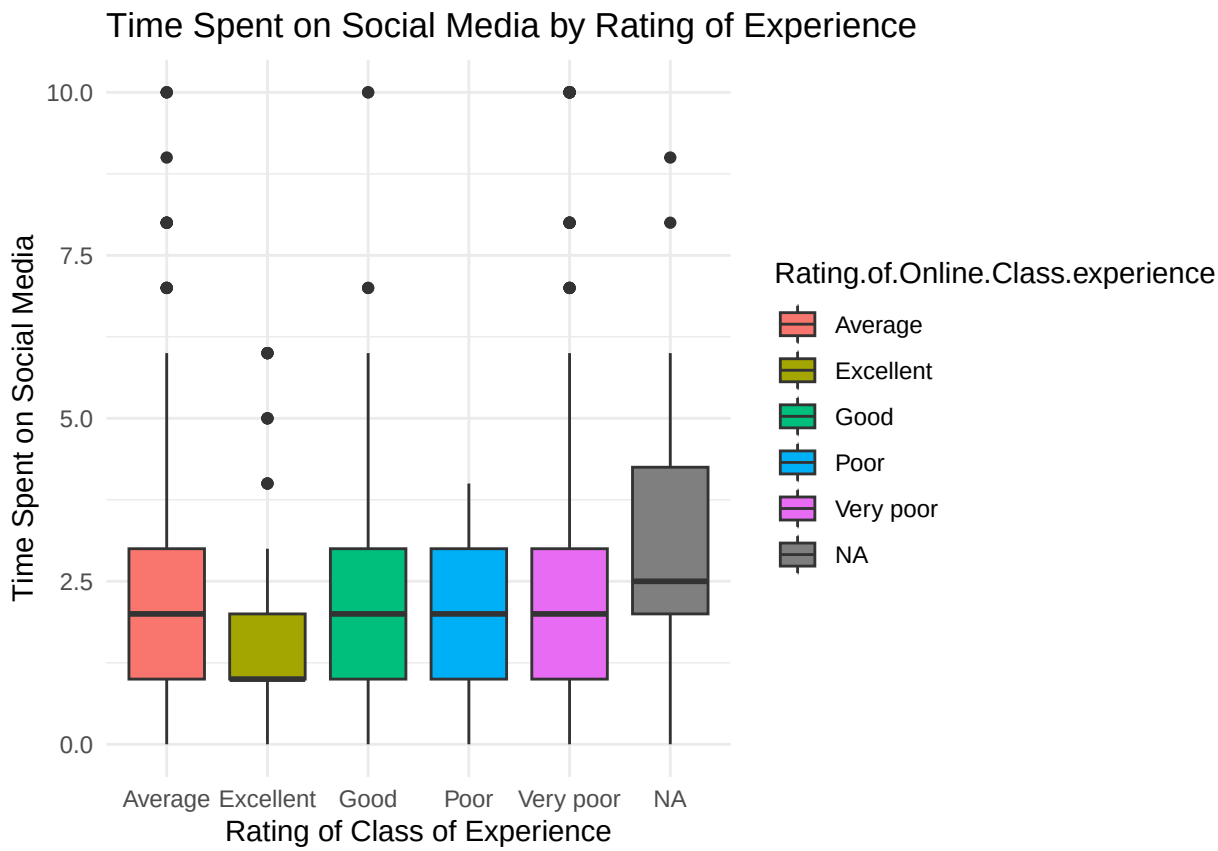

- The main tools for class were laptop/desktop and smartphone.
- There is a category called Laptop/Desktop or Smartphone. Why are these set separately? How come they weren't put in the distinct categories for either Laptop/Desktop or Smartphone.

```r
ggplot(students_2) + aes(x = Medium.for.online.class, fill = Medium.for.online.class) +
    geom_bar() +
    labs(
      title = "Distribution of Medium Used for Class",
      x = "Medium Used for Class",
      y = "Count"
    ) +
    theme_minimal()
```




- Not a big correlation between time spent on social media and experience rating
- Notable: Out of the 122 students who had an "Excellent" experience, the time spent on social media is less

```
ggplot(students_2) + aes(x = Rating.of.Online.Class.experience,
                         y = Time.spent.on.social.media,
                         fill = Rating.of.Online.Class.experience) +
    geom_boxplot() +
    labs(
      title = "Time Spent on Social Media by Rating of Experience",
      x = "Rating of Class of Experience",
      y = "Time Spent on Social Media"
    ) +
    theme_minimal()
```



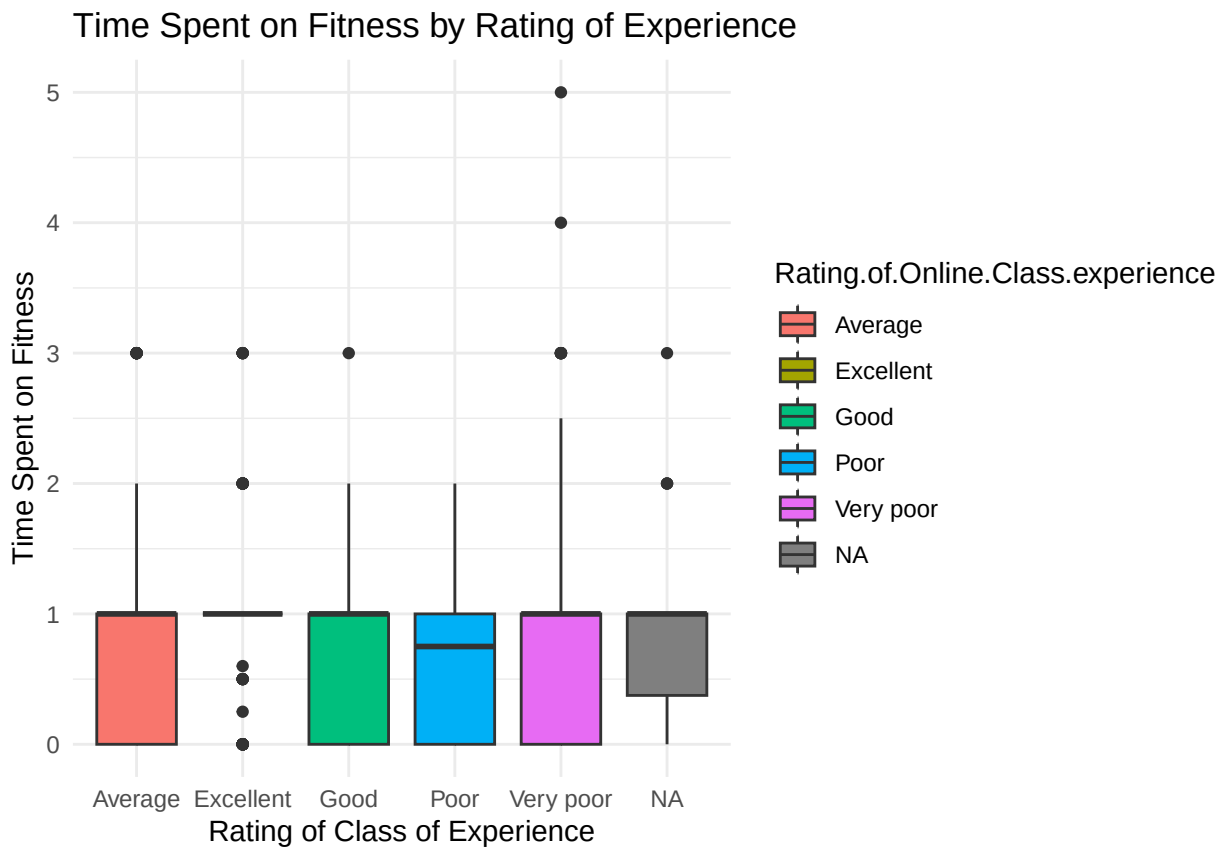

- There are 122 students who stated having an "Excellent" online class experience.

```
length(students_2$Rating.of.Online.Class.experience[students_2$"Rating.of.Online.Class.experience" == "
```

```
## [1] 122
```

- Very similar medians for all these categories when it comes to how much time they spent on fitness.
- So the health benefits from fitness may or may not have had an impact on their outlook on class. Need more research.

```
ggplot(students_2) + aes(x = Rating.of.Online.Class.experience,
                         y = Time.spent.on.fitness,
                         fill = Rating.of.Online.Class.experience) +
    geom_boxplot() +
    labs(
      title = "Time Spent on Fitness by Rating of Experience",
      x = "Rating of Class of Experience",
      y = "Time Spent on Fitness"
    ) +
    theme_minimal()
```




- Not much distinction between how much sleep the students got either.
- Probably not a factor in how favorably they viewed the online experience

```
ggplot(students_2) + aes(x = Rating.of.Online.Class.experience,
                         y = Time.spent.on.sleep,
                         fill = Rating.of.Online.Class.experience) +
    geom_boxplot() +
    labs(
      title = "Time Spent on Sleep by Rating of Experience",
      x = "Rating of Class of Experience",
      y = "Time Spent on Sleep"
    ) +
    theme_minimal()
```

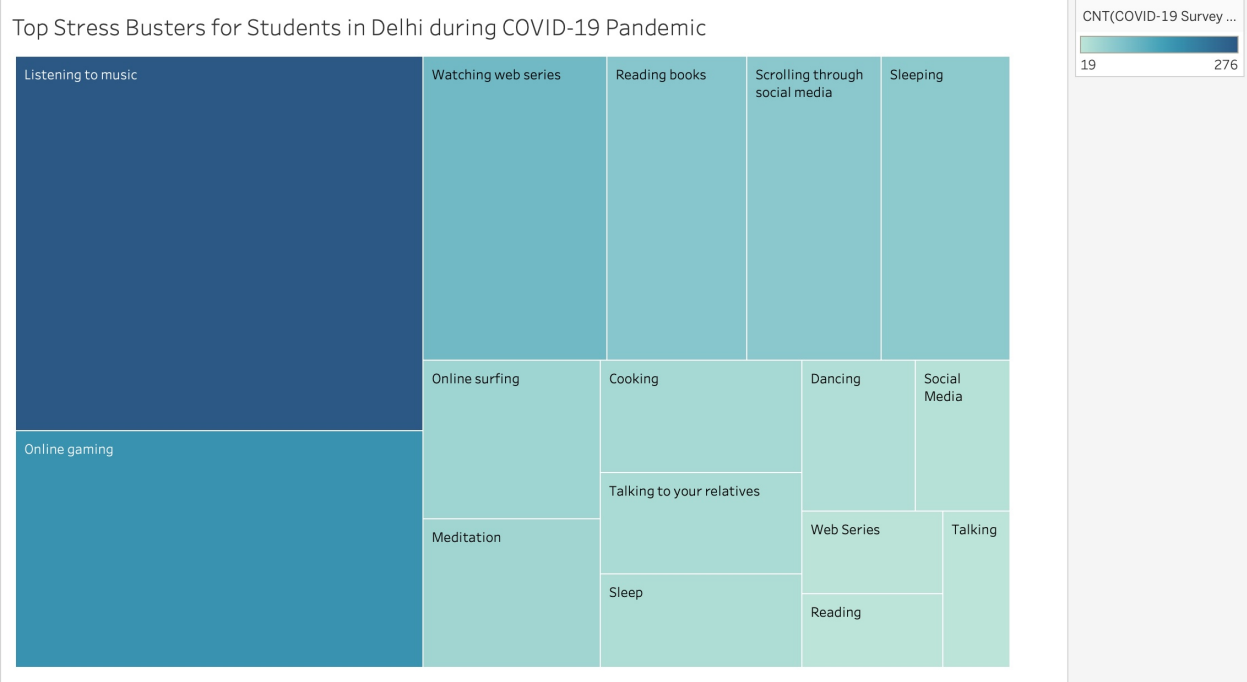

### Notes from Tableau graphics:

- What does the time utilized column mean? The options are YES and NO
- The main thing the students missed was School/college. Close seconds were Friends/relatives, travelling, and roaming around freely.

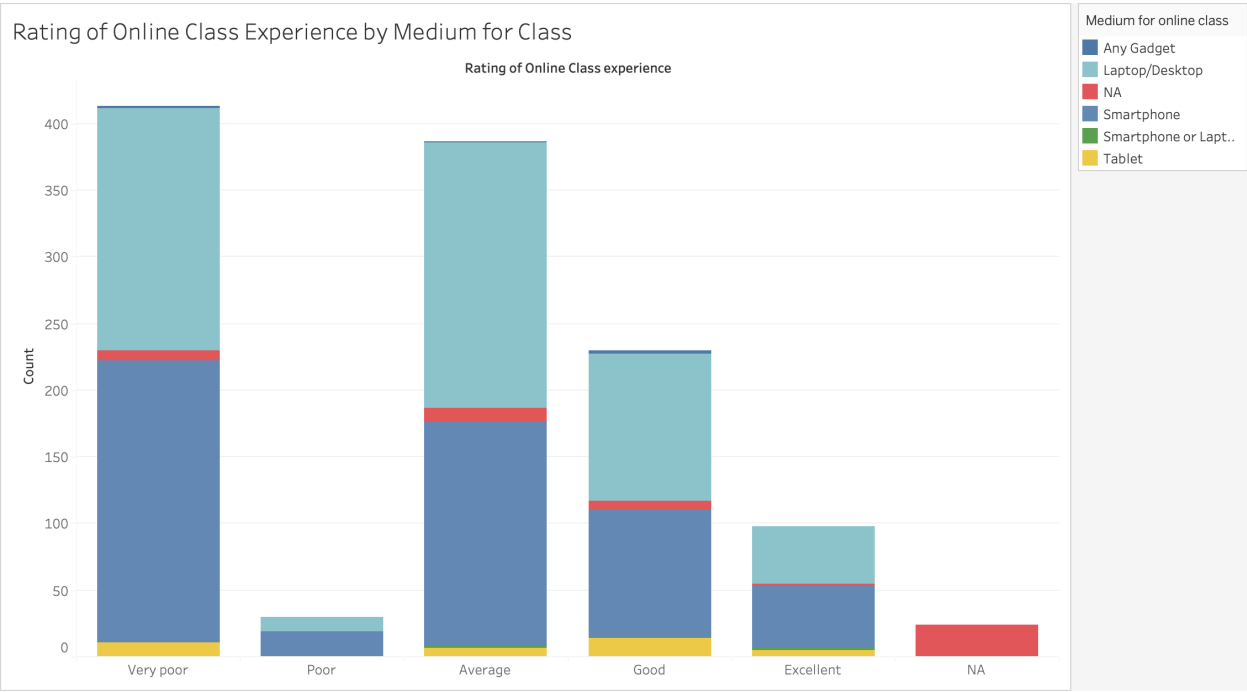
```
library(magick)
```

```
## Linking to ImageMagick 6.9.13.29
## Enabled features: cairo, fontconfig, freetype, heic, lcms, pango, raw, rsvg, webp
## Disabled features: fftw, ghostscript, x11
```

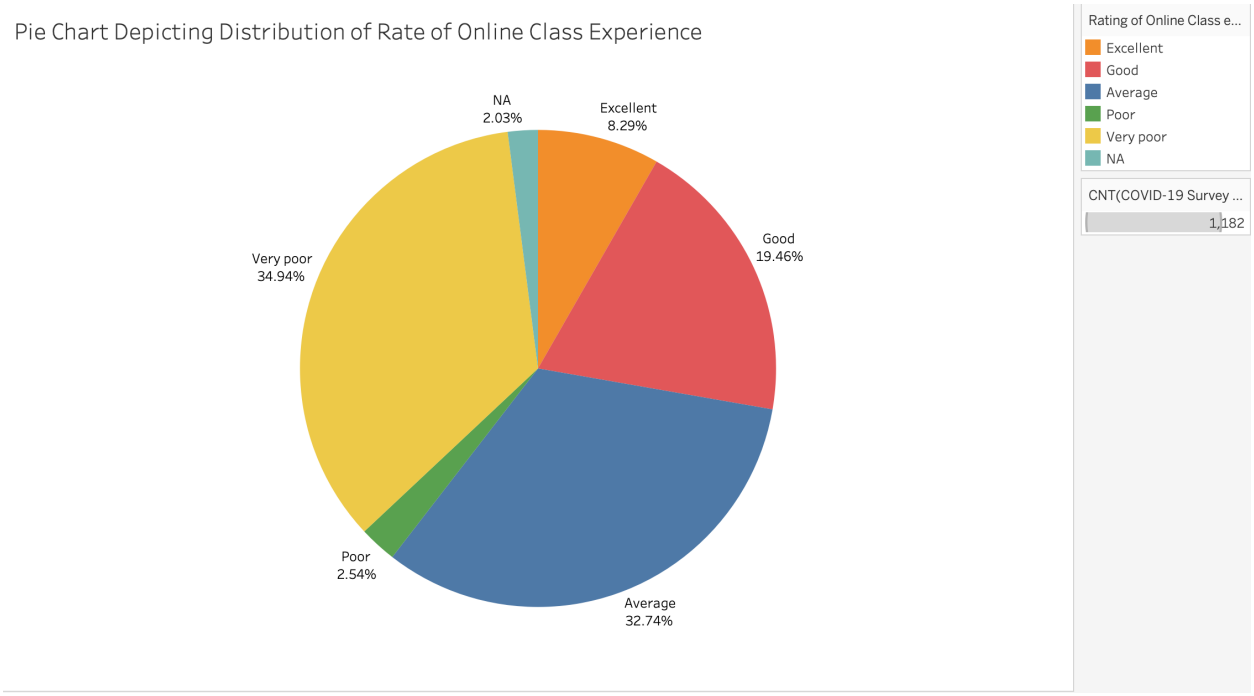
```
image_path <- "~Alan Ruiz/Desktop/graph1.jpeg"
my_image <- image_read(image_path)
knitr::include_graphics(image_path)
```




```
image_path_2 <- "~Alan Ruiz/Desktop/graph2.jpeg"
my_image_2 <- image_read(image_path_2)
knitr::include_graphics(image_path_2)
```




```
image_path_3 <- "~Alan Ruiz/Desktop/graph3.jpeg"
my_image_3 <- image_read(image_path_3)
knitr::include_graphics(image_path_3)
```




```
image_path_4 <- "~Alan Ruiz/Desktop/graph4.jpeg"
my_image_4 <- image_read(image_path_4)
knitr::include_graphics(image_path_4)
```