STA302H1 – Final Project Descriptive Statistics

Danny Chen

August 10, 2021

Import STA302H1 Study Time and COVID Contemplation Time vs. Quiz Performance Dataset

Data Cleaning

First, I'll clean my data.

```
cleaned_sta302_performance_data <- sta302_performance_data %>%
    # Create a new "country" column, which is just "Country" but whose entries are factors.
    mutate(country = as.factor(Country)) %>%

    # Remove the "X" column: it's simply the row number, which isn't very useful.
    # Remove the "Country" column: column "country" already exists
    select(-X, -Country) %>%

    # Group student overall quiz 4 scores from highest to lowest.
    arrange(desc(Quiz_4_score)) %>%

    # Rearrange similar columns side-by-side.
    relocate(country,
             COVID.hours..W1., COVID.hours..W2.,
             COVID.hours..W3., COVID.hours..W4.,
             STA302.hours..W1., STA302.hours..W2.,
             STA302.hours..W3., STA302.hours..W4.,
             Quiz_1_score, Quiz_2_score,
             Quiz_3_score, Quiz_4_score)

    # TODO: Make sure all country names are lowercase.
    # e.g. "Canada" and "canada" are the same country.
    # 1. Consider running a for loop that makes all rows in column "Country" lowercase,
    # 2. Consider string replacement on "Canada" -> "canada"?

    # TODO: Make sure all STA302H1 hours and COVID contemplation hours are
    # all in numeric form.
    # 1. use as.numeric()?

    # Identify rows with no quiz 4.
    # These indicate students who have dropped STA302H1, and who
    # should be excluded from the final data.
```

```r
head(cleaned_sta302_performance_data, n = 15)
```

```
##    country COVID.hours..W1. COVID.hours..W2. COVID.hours..W3. COVID.hours..W4.
## 1   Canada              2.0            3.000              1.0              2.0
## 2    China              1.0            0.500              1.0              2.0
## 3    China              5.0            4.000              5.0             12.0
## 4    China              0.0            0.000              0.5              0.5
## 5   Canada              1.0            0.000              0.0               NA
## 6    China              0.5            0.500              0.0              2.0
## 7   Canada              2.0            1.000              0.5              2.0
## 8    China              0.5            1.000              0.0              1.0
## 9    China              2.0            2.000              1.5              2.0
## 10   China              0.1            0.000              1.0              1.0
## 11   China              3.0            2.000              1.0               NA
## 12   China              1.0            2.000              1.0              5.0
## 13   China              2.0            2.000              2.0              2.0
## 14   China              1.0            0.500              0.5              0.5
## 15   China              0.0            0.333               NA              1.0
##    STA302.hours..W1. STA302.hours..W2. STA302.hours..W3. STA302.hours..W4.
## 1                  3               7.0                6                6
## 2                  3               3.0                3                3
## 3                 18               6.0               12               15
## 4                  6               6.0                3                4
## 5                  5               4.0                6               NA
## 6                  6               8.0               11               17
## 7                  9               9.0               15                9
## 8                 20              11.0               10                8
## 9                  8              10.0               11               12
## 10                 6               9.0                8               14
## 11                 6               8.0                7               NA
## 12                 8              10.0               10               16
## 13                10              14.0               14               24
## 14                 6               5.0                8               18
## 15                 3               3.5               NA               20
##    Quiz_1_score Quiz_2_score Quiz_3_score Quiz_4_score
## 1            10          7.8            9           10
## 2             8          2.8            9           10
## 3             9          9.4            9           10
## 4             9         10.0            9           10
## 5             9         10.0            9           10
## 6             8          5.2           10           10
## 7             8          5.8            5           10
## 8             6         10.0            9           10
## 9             7          2.8            9           10
## 10            5          9.0            9           10
## 11            9           NA            8           10
## 12            9         10.0            9           10
## 13            6          8.2            8           10
## 14            7          8.2            9           10
## 15            6         10.0            9           10
```

Identifying Anomalous Data

Rows with at least one NA deserve closer examination. Some of the rows might only have 1 - 2 NAs and are therefore salvageable, which is OK. Other rows may contain 3 or more NAs, and might indicate students who have dropped STA302H1. We'd like to exclude them from our analysis.

```
at_least_one_NA = function(data) {
  return (rowSums(is.na(cleaned_sta302_performance_data)) >= 1)
}

rows_with_some_NAs = cleaned_sta302_performance_data[
  at_least_one_NA(cleaned_sta302_performance_data),
]
rows_with_some_NAs
```

```
##      country COVID.hours..W1. COVID.hours..W2. COVID.hours..W3.
## 5     Canada             1.00            0.000           0.0000
## 11     China             3.00            2.000           1.0000
## 15     China             0.00            0.333               NA
## 27    Canada             1.00            1.000           1.0000
## 28      <NA>               NA            2.000           3.0000
## 29      <NA>               NA               NA           2.0000
## 30      <NA>               NA               NA               NA
## 31      <NA>               NA               NA               NA
## 36     China             0.50               NA           1.0000
## 39    Canada             1.50               NA           1.0000
## 48    Canada            30.00           40.000               NA
## 51     China             7.00               NA           4.0000
## 53     China             1.00            1.500           1.0000
## 54     Japan            10.00            3.000           4.0000
## 58    Canada             1.00               NA           0.5000
## 59      <NA>               NA            1.000           1.0000
## 60      <NA>               NA               NA               NA
## 61      <NA>               NA            1.000           3.0000
## 62     China             0.20            0.100           0.0000
## 63     China            84.00               NA          10.0000
## 93    Canada             1.00           40.000           0.5000
## 95      <NA>               NA            5.000           8.0000
## 96      <NA>               NA            0.000           0.0000
## 97      <NA>               NA            1.000           0.5000
## 98      <NA>               NA            1.000           4.0000
## 99      <NA>               NA               NA           0.1000
## 100     <NA>               NA               NA               NA
## 107    China            20.00               NA               NA
## 116   Canada             8.00               NA           1.0000
## 119   Canada            15.00            2.000           1.0000
## 126    China             1.25            1.000           2.0000
## 132     <NA>               NA            2.000           2.0000
## 133     <NA>               NA            1.000           1.0000
## 134     <NA>               NA               NA               NA
## 139   Canada             2.00            2.000           1.0000
## 140    China             1.00            1.000           1.0000
## 141    China             0.00               NA           2.0000
```

```
## 150   Canada             2.00            1.000           2.0000
## 152     <NA>               NA               NA           2.0000
## 153     <NA>               NA            3.000           2.0000
## 154     <NA>               NA            2.000           1.0000
## 155     <NA>               NA            0.500           3.0000
## 162   Canada             1.00            2.000           2.0000
## 167   Canada             3.00            1.000           1.0000
## 176     <NA>               NA            0.000           2.0000
## 177 Pakistan             2.00               NA              NA
## 186    China             2.00            1.500              NA
## 190     <NA>               NA               NA           1.0000
## 191   Canada             3.00            3.000              NA
## 193      UAE             6.00            6.000          10.0000
## 194   Canada             2.00            2.000              NA
## 195   Canada             1.00            3.000           3.0000
## 196   Canada             6.00            2.000           3.0000
## 197    China             3.00            3.000           4.0000
## 198    China             1.00               NA           1.0000
## 199   Canada             2.00            1.000           4.0000
## 200    China             1.00            2.000           2.0000
## 201   Taiwan            14.00            2.000           0.5000
## 202   Canada             1.00            3.000           1.0000
## 203   Canada             0.50            3.000           1.0000
## 204   Canada             1.00            1.000           1.0000
## 205   Canada             3.00           12.000              NA
## 206   Canada             7.00           14.000          16.0000
## 207   Canada             1.00            3.000              NA
## 208    China             3.00               NA          10.0000
## 209    China             1.00            1.000              NA
## 210   Canada             5.00            2.000              NA
## 211      USA             0.00               NA           1.0000
## 212   Canada             1.00            1.000           2.0000
## 213   Canada             1.00            1.000              NA
## 214    China             1.00            0.500           1.0000
## 215   Canada             1.00            3.000           1.0000
## 216    China             1.00            1.000           1.0000
## 217   Canada             2.00            2.000              NA
## 218   Taiwan             7.00            3.000           5.0000
## 219    China             0.50            0.300              NA
## 220   Canada             1.00            2.000           1.0000
## 221   Canada             3.50            4.000           4.0000
## 222   Taiwan             1.00            1.000           0.0833
## 223    China             1.00            2.000           1.0000
## 224     <NA>               NA            1.000           0.5000
## 225     <NA>               NA            2.000           5.0000
## 226     <NA>               NA               NA              NA
## 227     <NA>               NA            4.000              NA
##     COVID.hours..W4. STA302.hours..W1. STA302.hours..W2. STA302.hours..W3.
## 5                NA               5.0               4.0               6.0
## 11               NA               6.0               8.0               7.0
## 15            1.000               3.0               3.5                NA
## 27               NA               6.0               5.0               5.0
## 28            3.000                NA               8.0              10.0
## 29            3.000                NA                NA               4.0
```

```
## 30              NA              NA              NA              NA
## 31          10.000              NA              NA              NA
## 36           8.000             3.0              NA             2.0
## 39           1.500             7.0              NA             8.5
## 48          40.000             7.0            12.0              NA
## 51          10.000             3.0              NA            10.0
## 53           2.000             7.0             6.5             6.0
## 54           3.000            10.0             6.0            10.0
## 58           0.000             4.0              NA             4.0
## 59           2.000              NA            10.0             6.0
## 60              NA              NA              NA              NA
## 61           3.000              NA             8.0             9.0
## 62              NA             5.0             4.0             0.0
## 63          10.000             8.0              NA            10.0
## 93              NA             8.0             8.0            15.0
## 95           2.000              NA             6.0             7.0
## 96           0.000              NA             6.0             6.0
## 97           0.500              NA             8.0             7.0
## 98           4.000              NA             5.0            10.0
## 99           0.200              NA              NA            10.0
## 100          3.000              NA              NA              NA
## 107          5.000            20.0              NA              NA
## 116          1.000             2.0              NA             5.0
## 119          5.000             4.0             2.0             2.0
## 126             NA             3.0             5.0             7.0
## 132          1.000              NA             9.0            12.0
## 133          1.000              NA             6.0             7.0
## 134          5.000              NA              NA              NA
## 139             NA            16.5            12.0            10.5
## 140             NA             5.0             7.0             8.5
## 141          1.000             1.5              NA             3.0
## 150             NA             9.0             6.0             7.0
## 152          1.000              NA              NA             6.0
## 153          0.500              NA             7.0            10.0
## 154          1.000              NA             6.0             7.0
## 155          6.000              NA             8.0             7.0
## 162          3.000             5.0             8.0             8.0
## 167          1.000             9.0             7.0             9.0
## 176          1.000              NA             8.0             6.0
## 177          3.000             3.0              NA             4.0
## 186          6.000            11.0            12.0              NA
## 190          1.000              NA              NA            10.0
## 191          3.000             2.0             3.0              NA
## 193             NA            10.0             8.0             6.0
## 194             NA             6.0            10.0              NA
## 195             NA             1.0             3.0             4.0
## 196             NA             6.0             8.0             6.0
## 197             NA            21.0            20.0            21.0
## 198          1.000            10.0              NA             6.0
## 199             NA             0.5             2.0             2.0
## 200          3.000             5.0             7.0             6.0
## 201             NA             3.5             6.0             2.0
## 202             NA             3.0             5.0             4.0
## 203             NA             7.0             3.0             1.5
```

```
## 204                NA             3.0             5.0             5.0
## 205                NA             7.0             8.0              NA
## 206                NA             8.0            12.0            14.0
## 207                NA             3.0             4.0              NA
## 208                NA            20.0              NA            20.0
## 209                NA             8.0            11.0              NA
## 210                NA             3.0             4.0              NA
## 211             1.000             1.0              NA             1.0
## 212             2.000             9.0             9.5            11.0
## 213                NA             9.0            10.0              NA
## 214                NA             8.0             9.0             8.0
## 215                NA            10.0            10.0            12.0
## 216             1.000            10.0            10.0            10.0
## 217                NA             3.0             2.0              NA
## 218                NA            13.0            10.0            10.0
## 219                NA             7.0             6.5              NA
## 220                NA             4.0            10.0            11.0
## 221                NA             6.0             7.0             7.5
## 222             0.167             3.5            12.0            10.0
## 223                NA             3.0             6.0             8.0
## 224                NA              NA             6.0             8.0
## 225                NA              NA             5.0             4.0
## 226                NA              NA              NA              NA
## 227                NA              NA            16.0              NA
##     STA302.hours..W4. Quiz_1_score Quiz_2_score Quiz_3_score Quiz_4_score
## 5                  NA            9          10.0           9          10
## 11                 NA            9            NA           8          10
## 15               20.0            6          10.0           9          10
## 27                 NA           NA          10.0           9          10
## 28               12.0            7          10.0           9          10
## 29                5.0           10            NA           8          10
## 30                 NA           10          10.0          10          10
## 31               10.0           10          10.0          10          10
## 36               23.0            8           9.4          10           9
## 39               10.0           10           1.2           9           9
## 48               16.0            9           2.8           9           9
## 51               21.0           10           9.4           9           9
## 53                6.5            8           9.4          NA           9
## 54               20.0           NA           2.8           2           9
## 58                6.0           NA          10.0          10           9
## 59                7.0           10           5.2           8           9
## 60                 NA            4          10.0           9           9
## 61               15.0            4           4.8           9           9
## 62                 NA            6            NA           6           8
## 63               10.0            8           9.4          10           8
## 93                 NA            9           9.4           2           8
## 95               15.0            9           8.8           5           8
## 96                5.0            5           5.8           8           8
## 97               15.0           10           9.4           9           8
## 98                5.0           NA          10.0           8           8
## 99                7.0           NA           9.4           9           8
## 100              10.0           NA            NA          NA           8
## 107               6.0           NA          10.0           5           7
## 116               6.0            3           1.2           8           7
```

```
## 119   5.0   7   NA   5   7
## 126    NA   7  3.4   5   7
## 132  30.0   7  8.2   5   7
## 133   6.0  10 10.0   9   7
## 134  12.0  NA   NA   5   7
## 139    NA   8  8.4   5   6
## 140    NA   5  5.8   8   6
## 141   3.0   2  7.8   8   6
## 150    NA   7  8.8   5   6
## 152   8.0   8 10.0  10   6
## 153  12.0   7  0.0   5   6
## 154  14.0   9  6.4   3   6
## 155   8.0   3  2.8   5   6
## 162   7.0  NA  5.8   6   5
## 167  20.0   7   NA   6   5
## 176  10.0   4  5.8   8   5
## 177   5.0   9   NA  NA   4
## 186  12.0   6  9.4   4   4
## 190   6.0   3  5.8   5   4
## 191   3.0   6  5.4   5   3
## 193    NA   3   NA   3   1
## 194    NA  10  5.8  NA  NA
## 195    NA   6   NA  NA  NA
## 196    NA   8   NA  NA  NA
## 197    NA  10   NA  NA  NA
## 198   8.0   8  9.4   4  NA
## 199    NA   3   NA  NA  NA
## 200   5.0  10 10.0   9  NA
## 201    NA   8   NA  NA  NA
## 202    NA   6   NA  NA  NA
## 203    NA   9   NA  NA  NA
## 204    NA  NA   NA  NA  NA
## 205    NA   5   NA  NA  NA
## 206    NA   7   NA  NA  NA
## 207    NA   5   NA  NA  NA
## 208    NA   8   NA  NA  NA
## 209    NA  NA   NA  NA  NA
## 210    NA   5   NA  NA  NA
## 211   3.0   8  2.8   4  NA
## 212  18.0   8  9.4   3  NA
## 213    NA   5   NA  NA  NA
## 214    NA   4   NA  NA  NA
## 215    NA   6   NA  NA  NA
## 216   7.0   6  2.2   9  NA
## 217    NA  NA   NA  NA  NA
## 218    NA   5   NA  NA  NA
## 219    NA  10   NA  NA  NA
## 220    NA   6   NA  NA  NA
## 221    NA   7   NA  NA  NA
## 222  16.0   7 10.0  10  NA
## 223    NA   9   NA  NA  NA
## 224    NA   6   NA  NA  NA
## 225    NA   7   NA  NA  NA
## 226    NA   7   NA  NA  NA
```

```
## 227                NA           NA           NA           NA           NA
```

Rows with Mistyped Columns

Rows whose columns are mis-typed may need to be corrected via imputation.

```
##     country COVID.hours..W1. COVID.hours..W2. COVID.hours..W3. COVID.hours..W4.
## 38    China               0             0.5             1.0             0.5
## 83   Canada             168            40.0            20.0            12.0
## 84   Canada               1             1.0             2.0             1.0
## 117  Taiwan               1             1.0             0.5             0.5
##     STA302.hours..W1. STA302.hours..W2. STA302.hours..W3. STA302.hours..W4.
## 38                  4               5.5               5.5               6.0
## 83                  8               6.0               6.0              20.0
## 84                  9               8.0              12.0              15.0
## 117                 7               8.0               7.0               7.5
##     Quiz_1_score Quiz_2_score Quiz_3_score Quiz_4_score
## 38             9         10.0           10            9
## 83            10          9.4            9            8
## 84             9          5.4            9            8
## 117            6          8.8            8            7
```

Select Predictor Variables, Find Their Significance

```
# use week 5b slides -- choose model selection criterion to pick predictor variables.


# use lm() on a bunch of predictor variables to determine significant
# predictor variables.
```

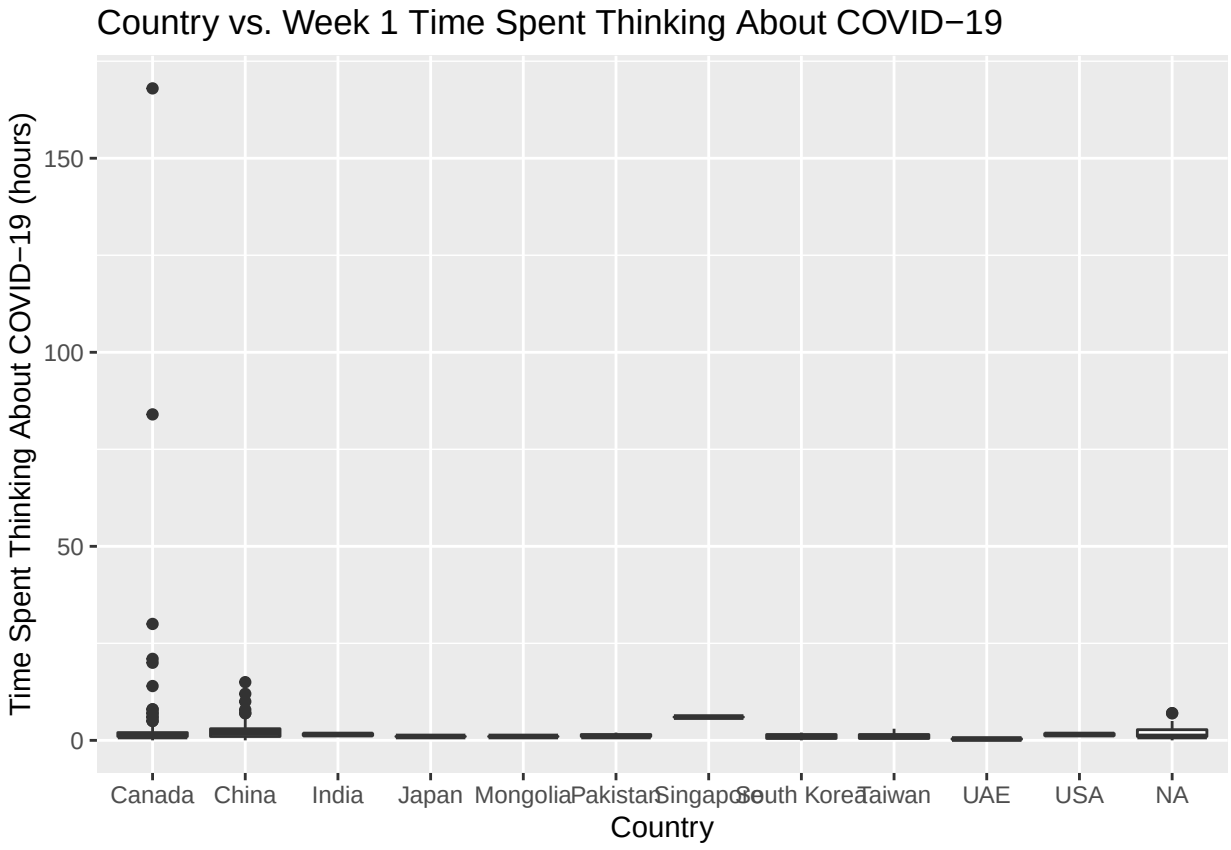
Histograms

```
# TODO: See Demo 1 to figure out how to add histograms in a matrix format.

# TODO: create histograms with ggplot, and then grid.arrange them together.
```

Boxplots

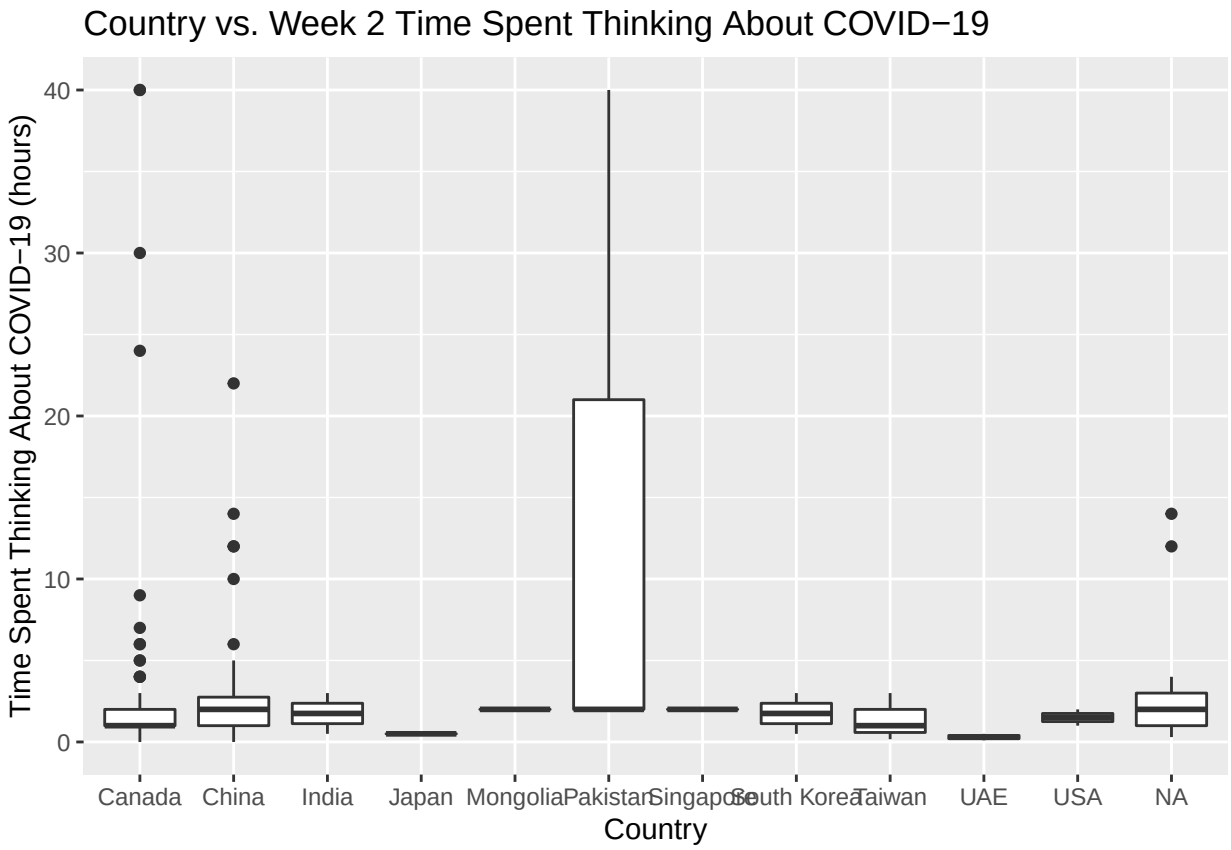
```
# TODO: See STA248H1 notes to figure out how to create boxplots. -- DONE
# TODO: See toy program of boxplots to see how to color them by factor

ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = COVID.hours..W1.)) +
  labs(title = "Country vs. Week 1 Time Spent Thinking About COVID-19",
       x = "Country",
       y = "Time Spent Thinking About COVID-19 (hours)")
```

```
## Warning: Removed 26 rows containing non-finite values (stat_boxplot).
```



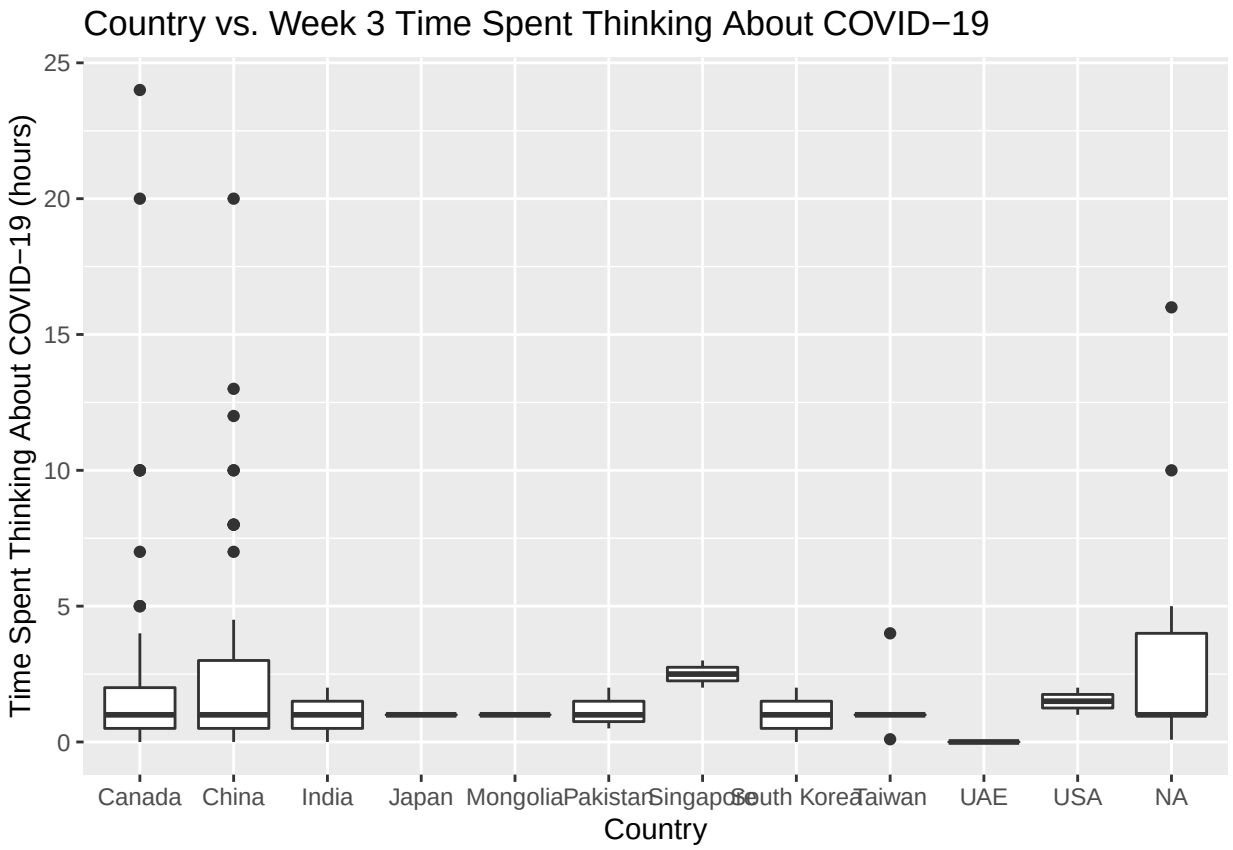

```
ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = COVID.hours..W2.)) +
  labs(title = "Country vs. Week 2 Time Spent Thinking About COVID-19",
       x = "Country",
       y = "Time Spent Thinking About COVID-19 (hours)")
```

```
## Warning: Removed 22 rows containing non-finite values (stat_boxplot).
```



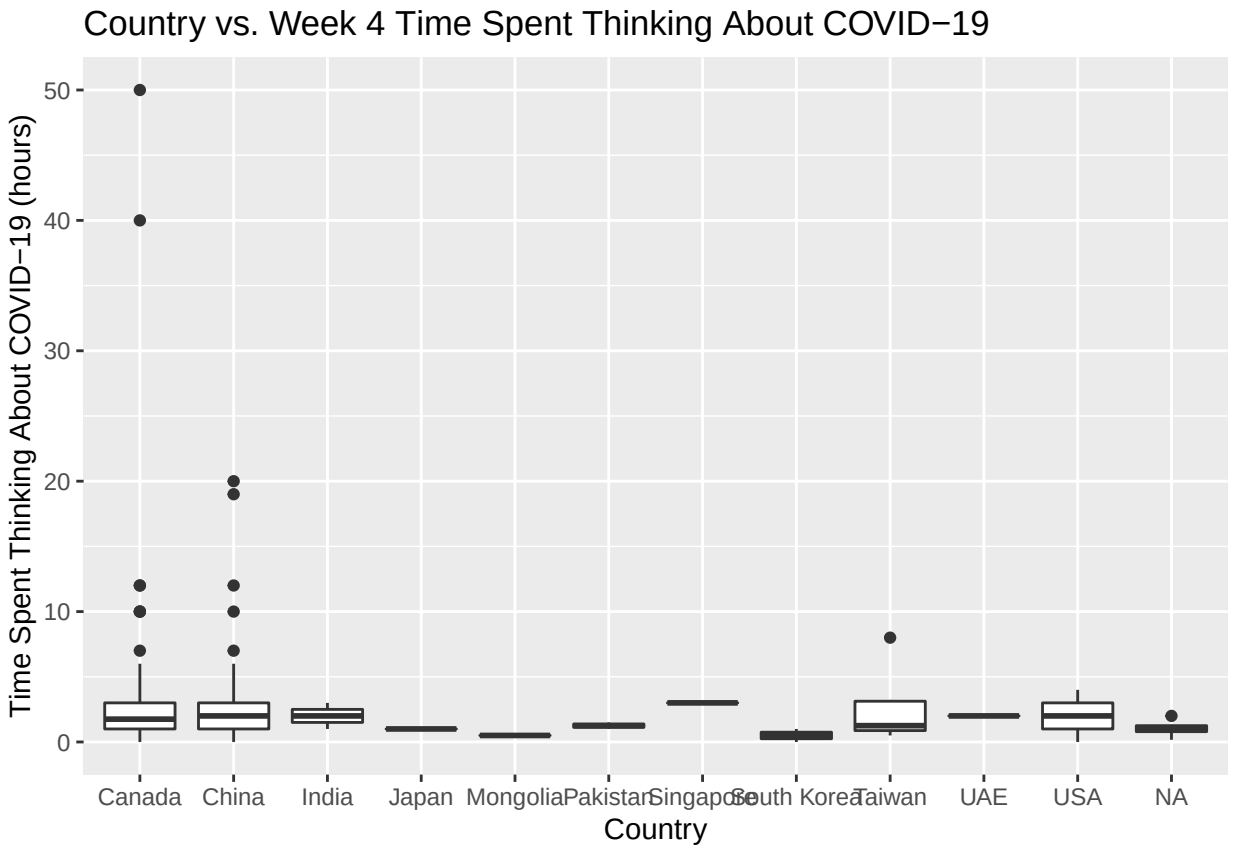

```
ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = COVID.hours..W3.)) +
  labs(title = "Country vs. Week 3 Time Spent Thinking About COVID-19",
       x = "Country",
       y = "Time Spent Thinking About COVID-19 (hours)")
```

```
## Warning: Removed 21 rows containing non-finite values (stat_boxplot).
```



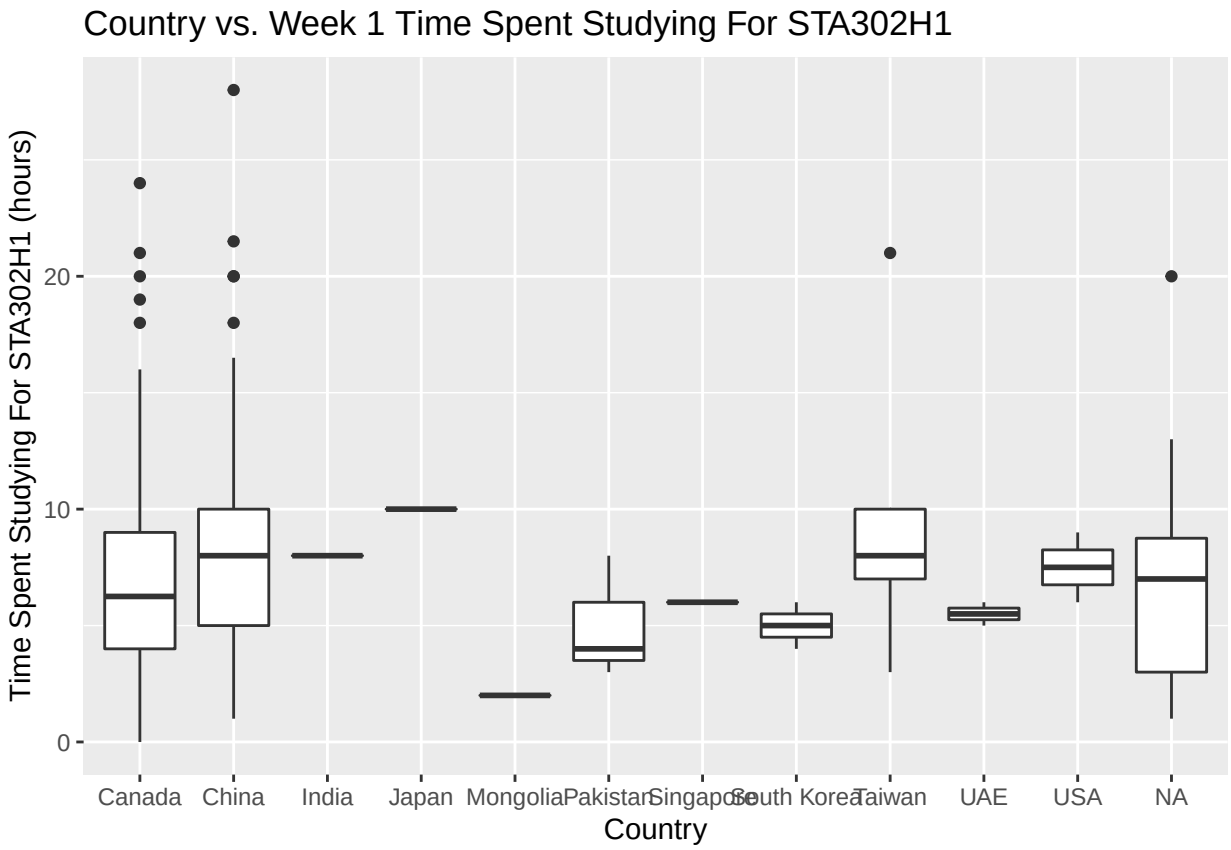

```
ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = COVID.hours..W4.)) +
  labs(title = "Country vs. Week 4 Time Spent Thinking About COVID-19",
       x = "Country",
       y = "Time Spent Thinking About COVID-19 (hours)")
```

```
## Warning: Removed 40 rows containing non-finite values (stat_boxplot).
```



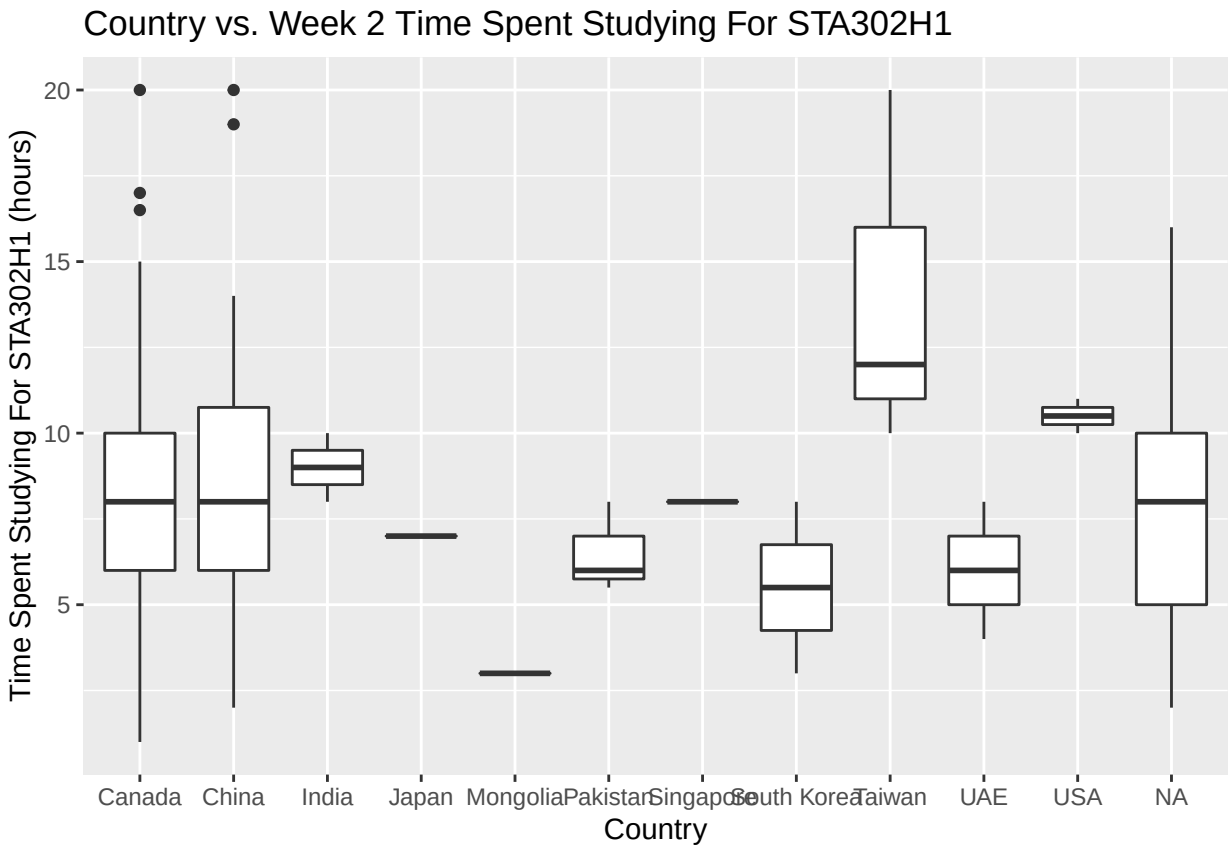

```
ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = STA302.hours..W1.)) +
  labs(title = "Country vs. Week 1 Time Spent Studying For STA302H1",
       x = "Country",
       y = "Time Spent Studying For STA302H1 (hours)")
```

```
## Warning: Removed 26 rows containing non-finite values (stat_boxplot).
```



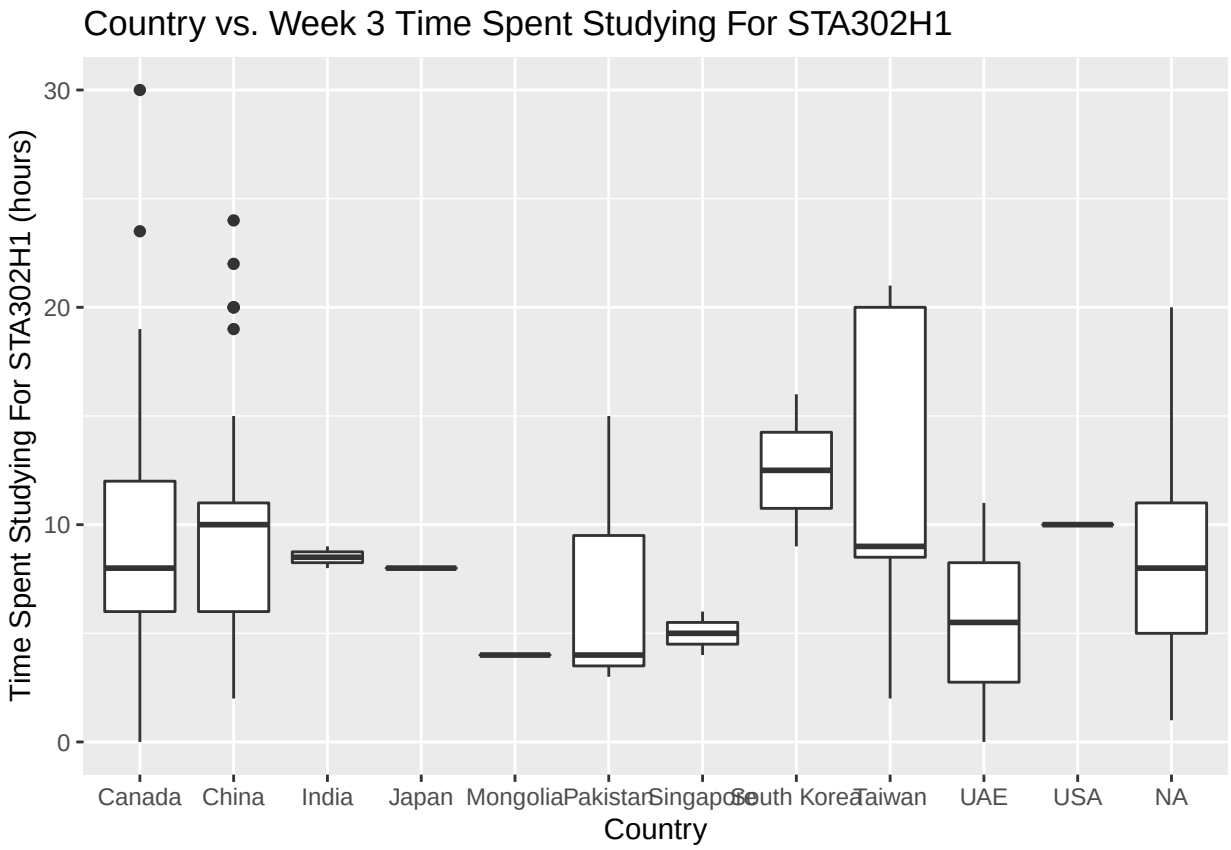

```
ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = STA302.hours..W2.)) +
  labs(title = "Country vs. Week 2 Time Spent Studying For STA302H1",
       x = "Country",
       y = "Time Spent Studying For STA302H1 (hours)")
```

```
## Warning: Removed 22 rows containing non-finite values (stat_boxplot).
```



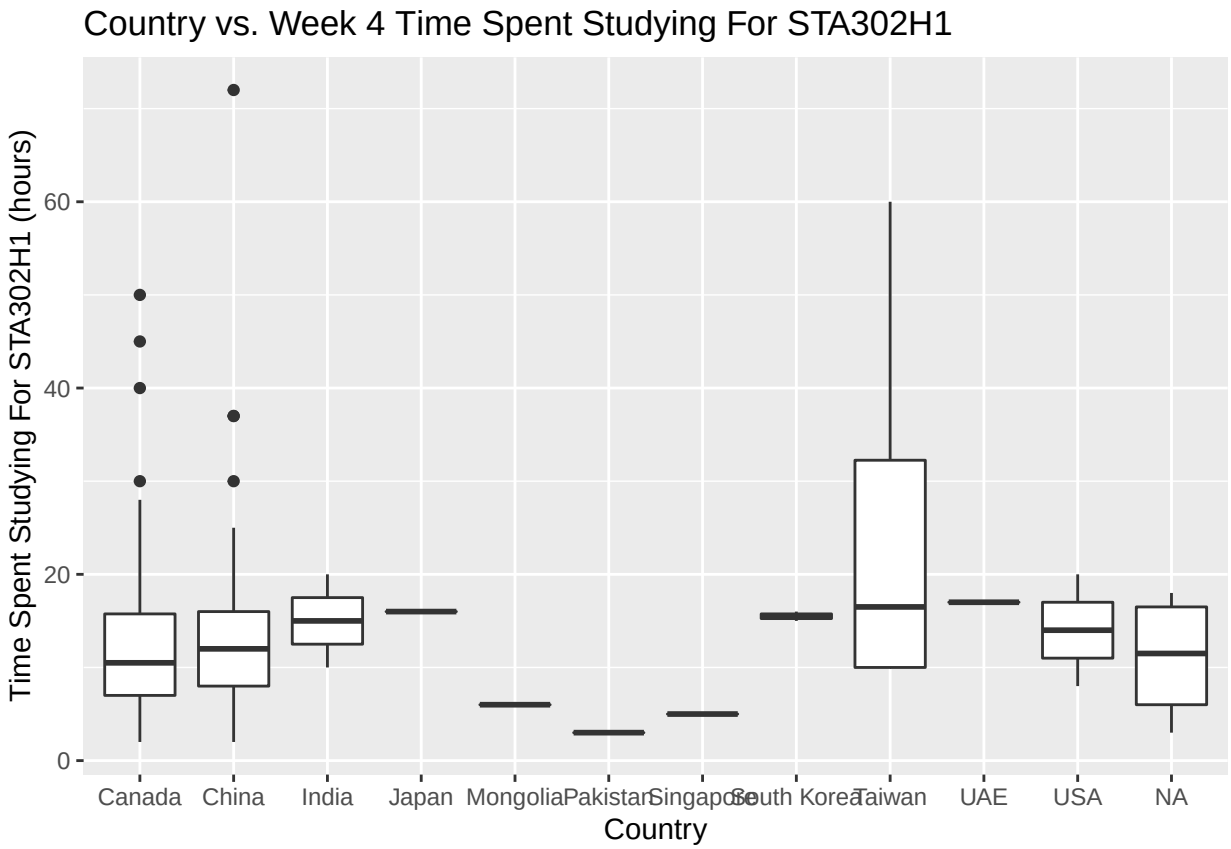

```
ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = STA302.hours..W3.)) +
  labs(title = "Country vs. Week 3 Time Spent Studying For STA302H1",
       x = "Country",
       y = "Time Spent Studying For STA302H1 (hours)")
```

```
## Warning: Removed 20 rows containing non-finite values (stat_boxplot).
```



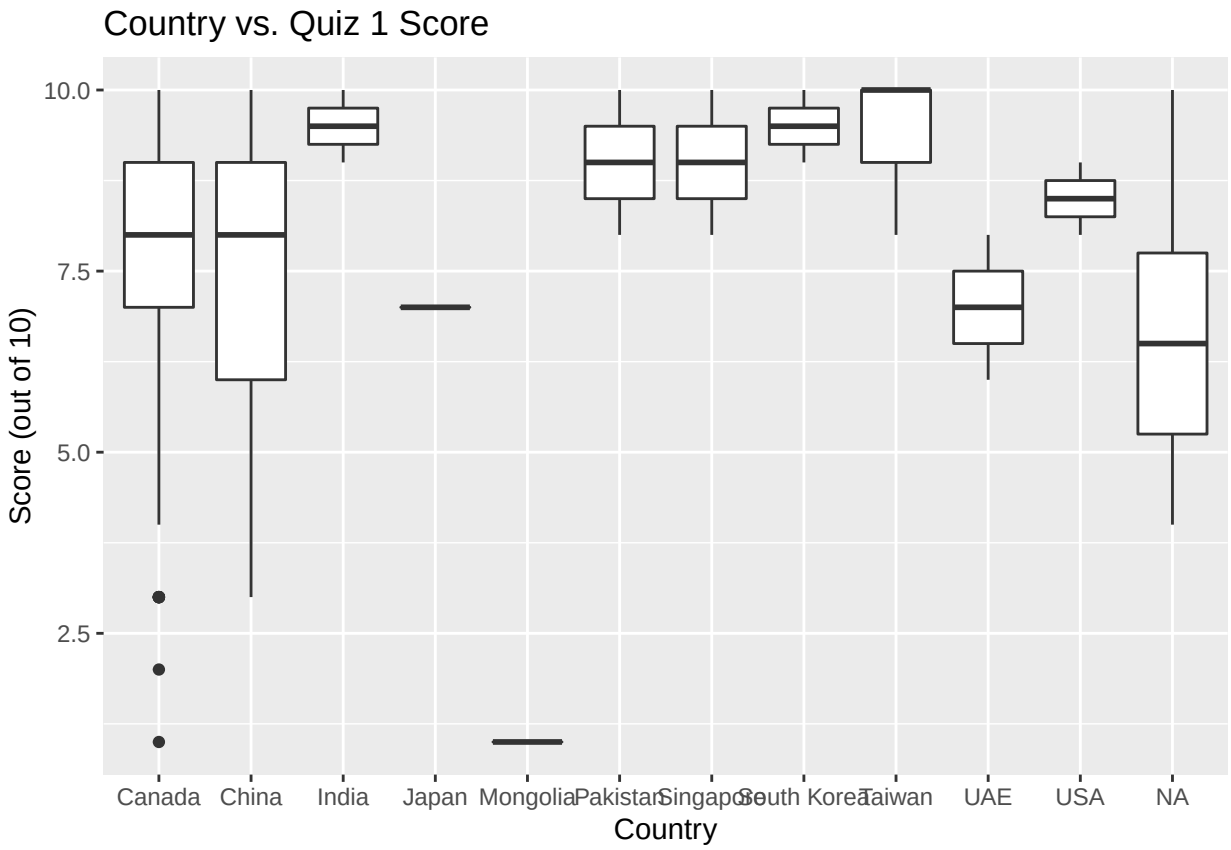

```
ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = STA302.hours..W4.)) +
  labs(title = "Country vs. Week 4 Time Spent Studying For STA302H1",
       x = "Country",
       y = "Time Spent Studying For STA302H1 (hours)")
```

```
## Warning: Removed 40 rows containing non-finite values (stat_boxplot).
```



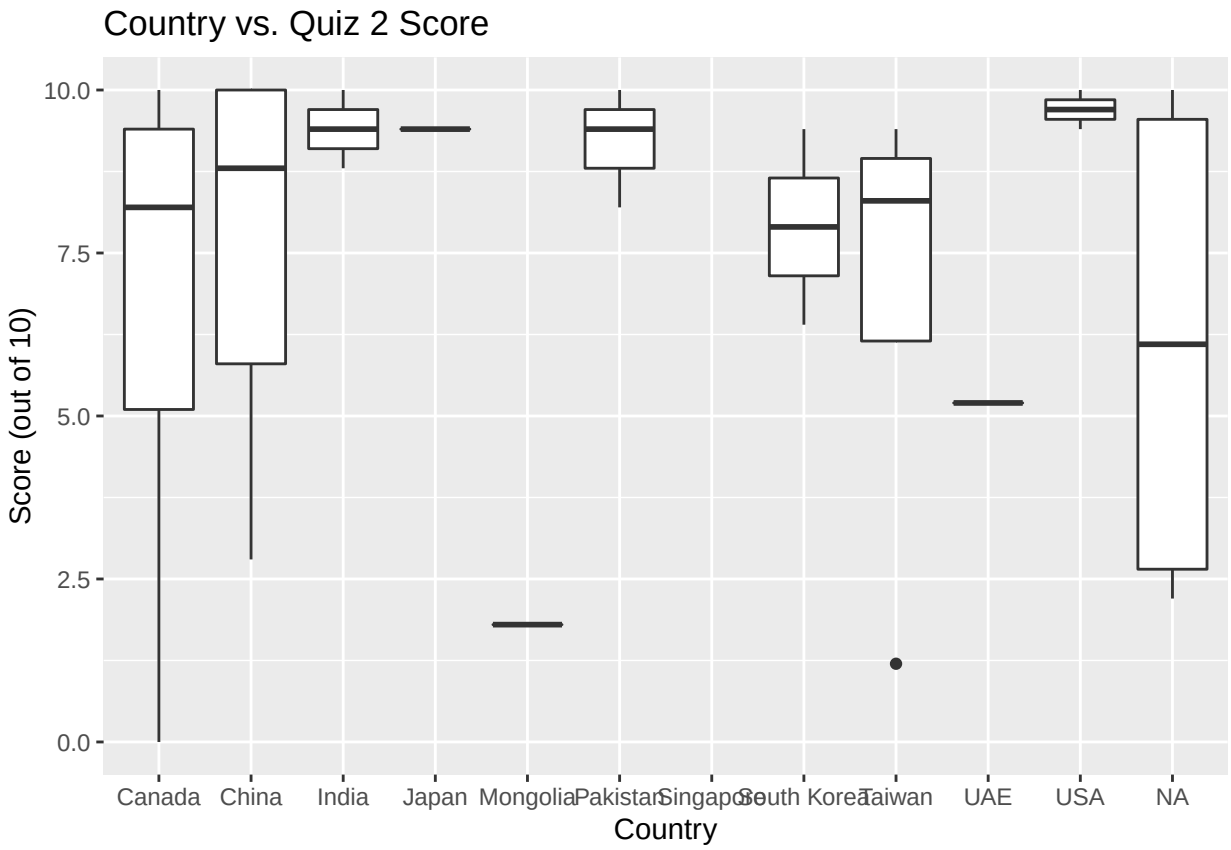

```
ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = Quiz_1_score)) +
  labs(title = "Country vs. Quiz 1 Score",
       x = "Country",
       y = "Score (out of 10)")
```

```
## Warning: Removed 13 rows containing non-finite values (stat_boxplot).
```



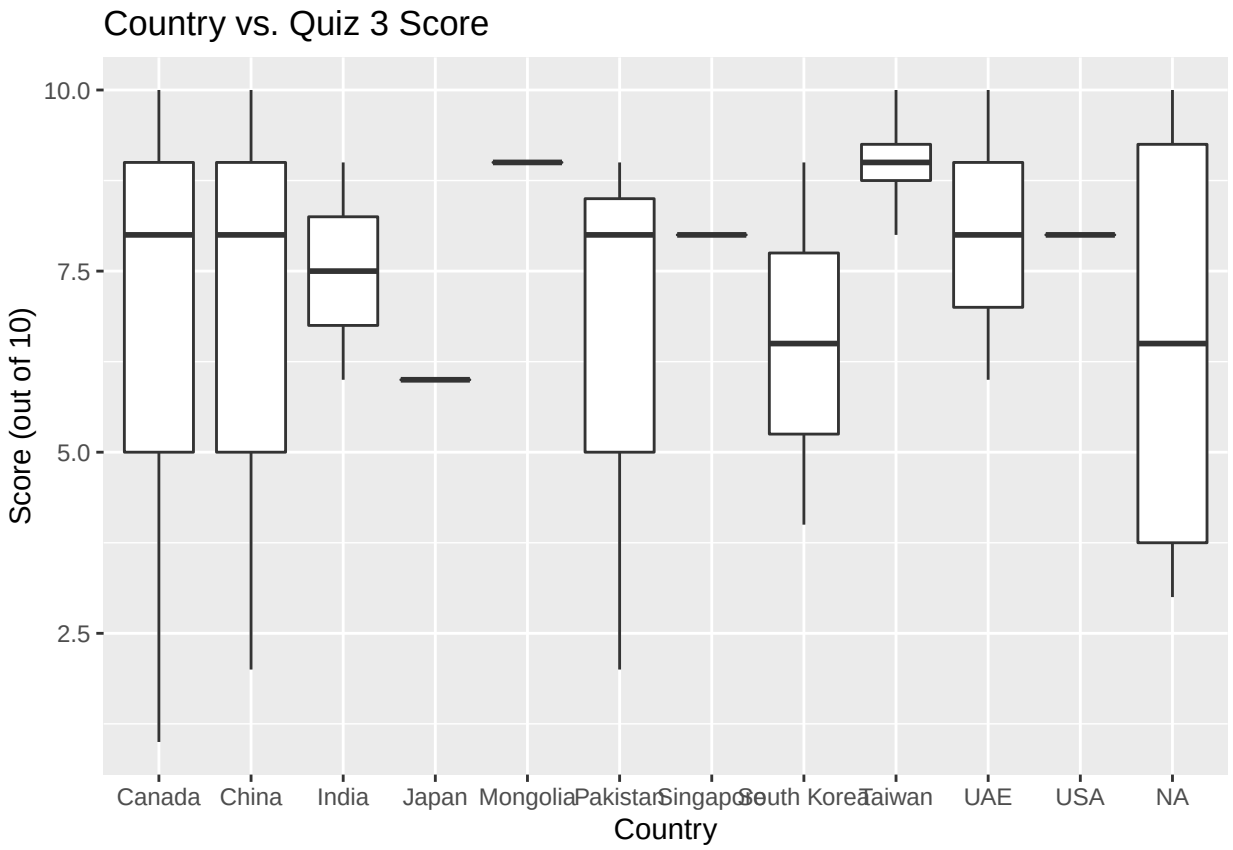

```
ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = Quiz_2_score)) +
  labs(title = "Country vs. Quiz 2 Score",
       x = "Country",
       y = "Score (out of 10)")
```

```
## Warning: Removed 36 rows containing non-finite values (stat_boxplot).
```



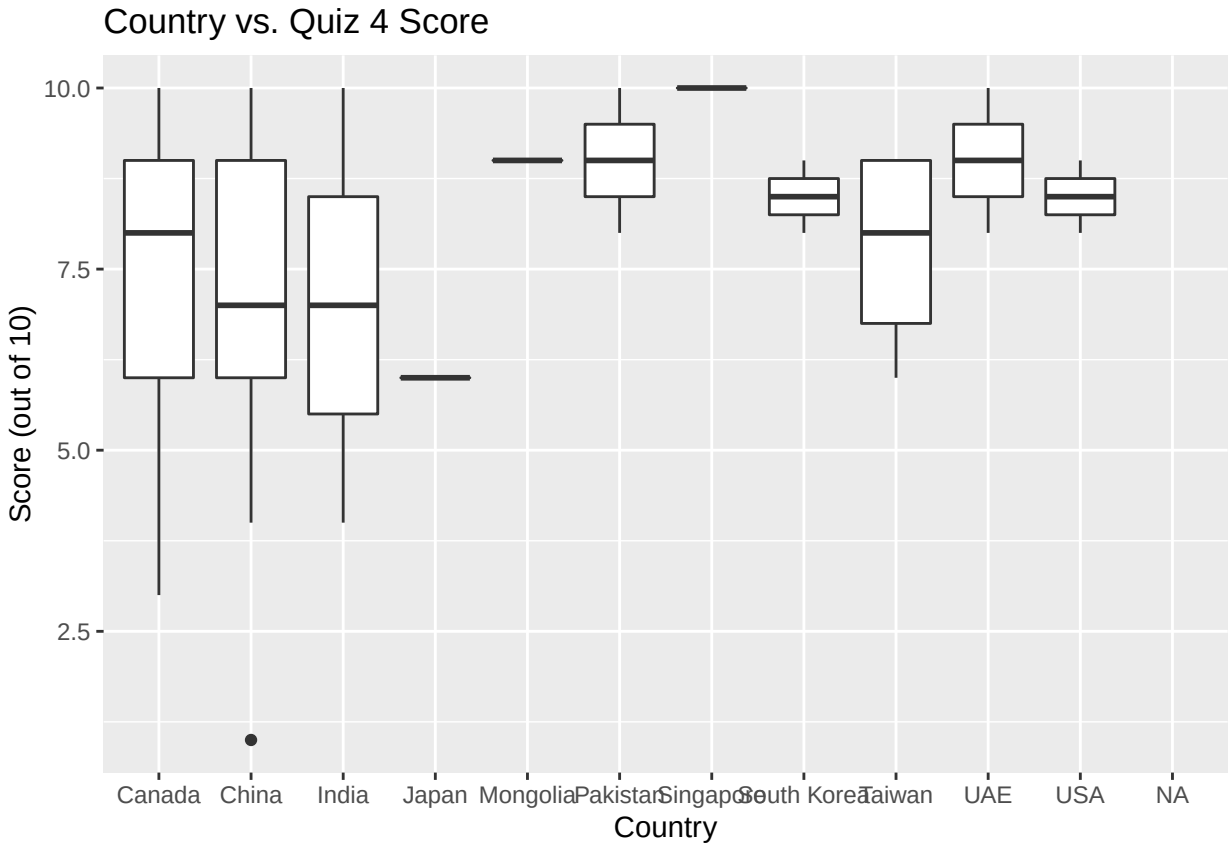

```
ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = Quiz_3_score)) +
  labs(title = "Country vs. Quiz 3 Score",
       x = "Country",
       y = "Score (out of 10)")
```

```
## Warning: Removed 31 rows containing non-finite values (stat_boxplot).
```



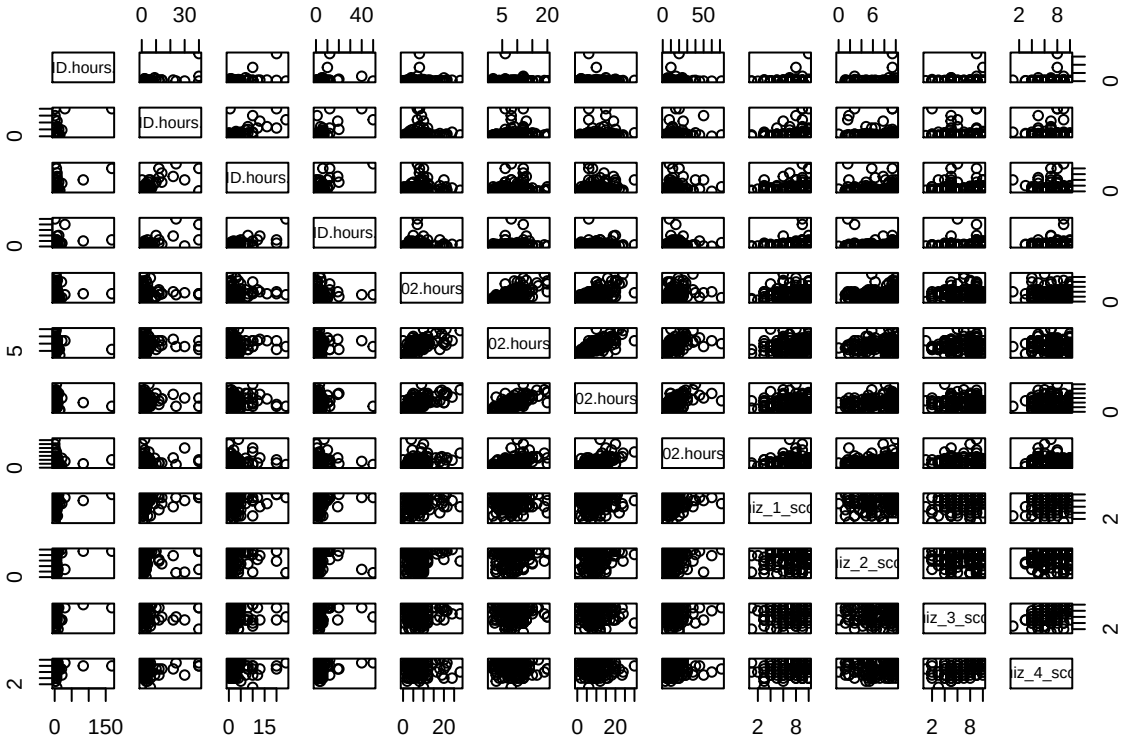

```
ggplot(data = cleaned_sta302_performance_data) +
  geom_boxplot(mapping = aes(x = Country, y = Quiz_4_score)) +
  labs(title = "Country vs. Quiz 4 Score",
       x = "Country",
       y = "Score (out of 10)")
```

```
## Warning: Removed 34 rows containing non-finite values (stat_boxplot).
```


Country vs. Quiz 4 Score
Score (out of 10)
10.0
7.5
5.0
2.5
Canada
China
India
Japan
Mongolia
Pakistan
Singapore
South Korea
Taiwan
UAE
USA
NA
Country

Scatterplots

```
# pairwise scatterplot
pairs(~COVID.hours..W1. + COVID.hours..W2. + COVID.hours..W3. + COVID.hours..W4. +
      STA302.hours..W1. + STA302.hours..W2. + STA302.hours..W3. + STA302.hours..W4. +
      Quiz_1_score + Quiz_2_score + Quiz_3_score + Quiz_4_score,
      data = cleaned_sta302_performance_data)
```



Correlation Matrix

```
# take out country column
# TODO: Or create separate correlation matrices for each country?
no_country = cleaned_sta302_performance_data %>%
  select(-country)
```

```
# Find correlation matrix to determine candidate significant predictor values.
# library(GGally)
colnames(no_country) <- c("W1COV", "W2COV", "W3COV", "W4COV",
                          "W1302", "W2302", "W3302", "W4302",
                          "Q1", "Q2", "Q3", "Q4")
# ggcorr(no_country, label = TRUE, label_round = 2)
round(cor(no_country, use = "complete.obs"), 2)  # TODO: na.rm = true
```

```
##       W1COV W2COV W3COV W4COV W1302 W2302 W3302 W4302    Q1    Q2    Q3    Q4
## W1COV  1.00  0.66  0.46  0.20  0.02 -0.04 -0.02  0.06  0.10  0.07  0.05  0.01
## W2COV  0.66  1.00  0.82  0.60  0.06  0.05  0.13  0.21  0.11 -0.10 -0.08 -0.06
## W3COV  0.46  0.82  1.00  0.73  0.06  0.09  0.14  0.13  0.13 -0.10 -0.11 -0.06
## W4COV  0.20  0.60  0.73  1.00  0.02  0.04  0.09  0.07  0.10 -0.09 -0.03  0.01
## W1302  0.02  0.06  0.06  0.02  1.00  0.61  0.57  0.31  0.02  0.11  0.03 -0.07
## W2302 -0.04  0.05  0.09  0.04  0.61  1.00  0.70  0.49 -0.04  0.08 -0.09 -0.12
## W3302 -0.02  0.13  0.14  0.09  0.57  0.70  1.00  0.62 -0.07  0.08 -0.14 -0.09
## W4302  0.06  0.21  0.13  0.07  0.31  0.49  0.62  1.00 -0.07  0.02 -0.05 -0.11
## Q1     0.10  0.11  0.13  0.10  0.02 -0.04 -0.07 -0.07  1.00  0.22  0.33  0.21
## Q2     0.07 -0.10 -0.10 -0.09  0.11  0.08  0.08  0.02  0.22  1.00  0.22  0.16
## Q3     0.05 -0.08 -0.11 -0.03  0.03 -0.09 -0.14 -0.05  0.33  0.22  1.00  0.54
## Q4     0.01 -0.06 -0.06  0.01 -0.07 -0.12 -0.09 -0.11  0.21  0.16  0.54  1.00
```

Summary Statistics

Mean STA302H1 study time

```
mean_STA302H1_study_times <- data.frame(
  week1 = mean(sta302_performance_data$STA302.hours..W1., na.rm = TRUE),
  week2 = mean(sta302_performance_data$STA302.hours..W2., na.rm = TRUE),
  week3 = mean(sta302_performance_data$STA302.hours..W3., na.rm = TRUE),
  week4 = mean(sta302_performance_data$STA302.hours..W4., na.rm = TRUE)
)
mean_STA302H1_study_times
```

```
##      week1    week2    week3    week4
## 1 7.457711 8.297561 9.224638 13.41711
```

Mean COVID contemplation time

```
mean_COVID_contemplation_times <- data.frame(
  week1 = mean(sta302_performance_data$COVID.hours..W1., na.rm = TRUE),
  week2 = mean(sta302_performance_data$COVID.hours..W2., na.rm = TRUE),
  week3 = mean(sta302_performance_data$COVID.hours..W3., na.rm = TRUE),
  week4 = mean(sta302_performance_data$COVID.hours..W4., na.rm = TRUE)
)
mean_COVID_contemplation_times
```

```
##      week1    week2    week3    week4
## 1 3.607163 2.884312 2.333171 2.917717
```

Median STA302H1 study time

```
median_STA302H1_study_times <- data.frame(
  week1 = median(sta302_performance_data$STA302.hours..W1., na.rm = TRUE),
  week2 = median(sta302_performance_data$STA302.hours..W2., na.rm = TRUE),
  week3 = median(sta302_performance_data$STA302.hours..W3., na.rm = TRUE),
  week4 = median(sta302_performance_data$STA302.hours..W4., na.rm = TRUE)
)
median_STA302H1_study_times
```

```
##   week1 week2 week3 week4
## 1     7     8     9    11
```

Median COVID contemplation time

```
median_COVID_contemplation_times <- data.frame(
  week1 = median(sta302_performance_data$COVID.hours..W1., na.rm = TRUE),
  week2 = median(sta302_performance_data$COVID.hours..W2., na.rm = TRUE),
  week3 = median(sta302_performance_data$COVID.hours..W3., na.rm = TRUE),
  week4 = median(sta302_performance_data$COVID.hours..W4., na.rm = TRUE)
)
median_COVID_contemplation_times
```

```
##   week1 week2 week3 week4
## 1     1     1     1   1.5
```

Country summary statistics

```
length(which(cleaned_sta302_performance_data$Country == "Canada")) + 2
```

```
## [1] 2
```

```
length(which(is.na(cleaned_sta302_performance_data$Country)))
```

```
## [1] 0
```

Study hours summary statistics

```
summary(sta302_performance_data$STA302.hours..W1.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   4.000   7.000   7.458   9.000  28.000      26
```

```
summary(sta302_performance_data$STA302.hours..W2.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   6.000   8.000   8.298  10.000  20.000      22
```

```
summary(sta302_performance_data$STA302.hours..W3.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   6.000   9.000   9.225  11.500  30.000      20
```

```
summary(sta302_performance_data$STA302.hours..W4.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    2.00    7.00   11.00   13.42   16.00   72.00      40
```

COVID hours summary statistics

```
summary(sta302_performance_data$COVID.hours..W1.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   1.000   1.000   3.607   2.000 168.000      26
```

```
summary(sta302_performance_data$COVID.hours..W2.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   1.000   1.000   2.884   2.000  40.000      22
```

```r
summary(sta302_performance_data$COVID.hours..W3.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   0.500   1.000   2.333   2.000  24.000      21
```

```r
summary(sta302_performance_data$COVID.hours..W4.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   1.000   1.500   2.918   3.000  50.000      40
```