TITLE OF YOUR PROJECT

Group 8

November 12th, 2024

List your group members, including their student numbers, here:

«««< HEAD

- Aditya Chauhan (169027493)
- Andy Ferrer Lorenzo (169081356)
- Oscar Saul (169057882)
- Taras Zakydalsky (169099616)
- Ezra Furman (#########) »»»> 5f7fee6584554860c6964f24cabaeb60755a72fe

You **must** be in a group in MyLS in order to see the DropBox used for submission. Even if you're alone, you must join a group by yourself.

You **must** be in a group with people from the same section as you. MyLS does not allow for groups including students from both Data100A and Data100B.

Instructions

You are encouraged to remove this instruction section prior to submission.

It is recommended that you follow the structure of this template. The text is all placeholder - you are free to change any/all wording as you please, but it is very helpful for the grading process if you keep the same structure. Anything in <> definitely needs to be changed, but you are free to change any/all sentences!

Note that all of the code is *hidden* by default. This file will be graded based on the insights, not the code.

You will only submit the PDF version of this document. To knit to PDF, you'll need to run `install.packages("tinytex")` in the console, followed by `tinytex::install_tinytex()` (DO NOT PUT THESE COMMANDS IN AN RMD FILE!!!). If you encounter errors in "Knit to PDF", you can "knit to html" and then print the html file to PDF using your operating system's PDF view (e.g. Adobe Acrobat). Only standalone PDF files will be accepted by MyLS.

Abstract

Introduction

Climate change is something that has been studied. Here's some relevant information about the context of our study.

If needed, this paragraph is more information about the context.

In this report, we are going to explore some aspects climate change and the impact and/or perceptions of it by using exploratory techniques. We'll explore <> using <>.

By the end of this report, we will have shown ...

Data Description

«Cyclones + Climate Awareness»

```
## Warning: There was 1 warning in `summarise()`.
## i In argument: `across(where(is.numeric), mean, na.rm = TRUE)`.
## Caused by warning:
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
## Supply arguments directly to `.fns` through an anonymous function instead.
##
##   # Previously
##   across(a:b, mean, na.rm = TRUE)
##
##   # Now
##   across(a:b, \(x) mean(x, na.rm = TRUE))
```

The data come from <> and detail <>.

In order to clean the data, we «steps to clean the data, concise but precise enough that a reader could follow your steps without seeing your code»

«Data Set 3»

The data come from <> and detail <>.

In order to clean the data, we «steps to clean the data, concise but precise enough that a reader could follow your steps without seeing your code»

«Data Set 4»

«Data Set 5»

Combining the Data

Explain how any combinations of data were performed. Explain what kind of join was needed, whether columns had to be modified (for example, matching "country" names.)

Exploratory Data Analysis

To achieve our goals, we explored the data by...

We explored many aspects of the data, but will demonstrate three. These are «Social Support VS Life Expectancy», «Perception of Corruption VS Covid Cases», and «Freedom to Make Life Choices & Climate Awareness»

The first aspect that we found interesting is shown in @ref(fig:insight1). The insight should be specific to the data shown, not a general statement beyond the data (leave that for the conclusion).

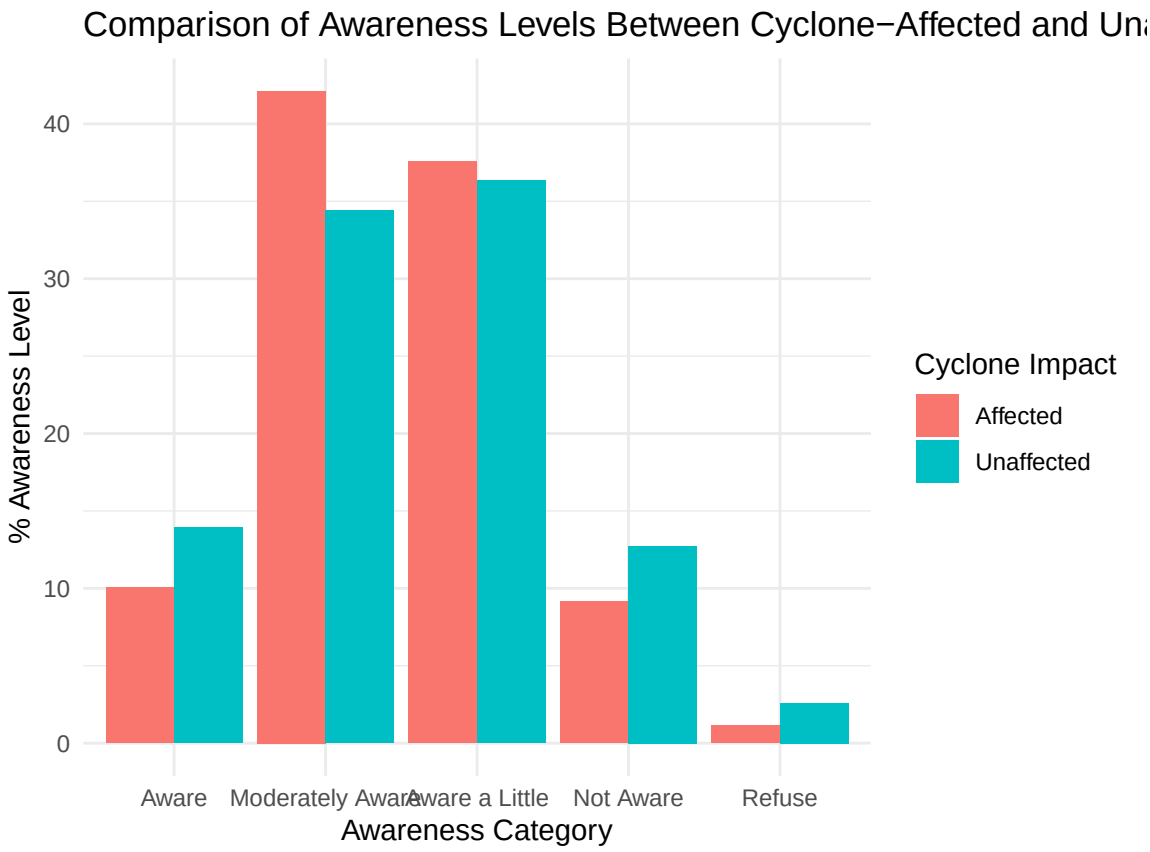


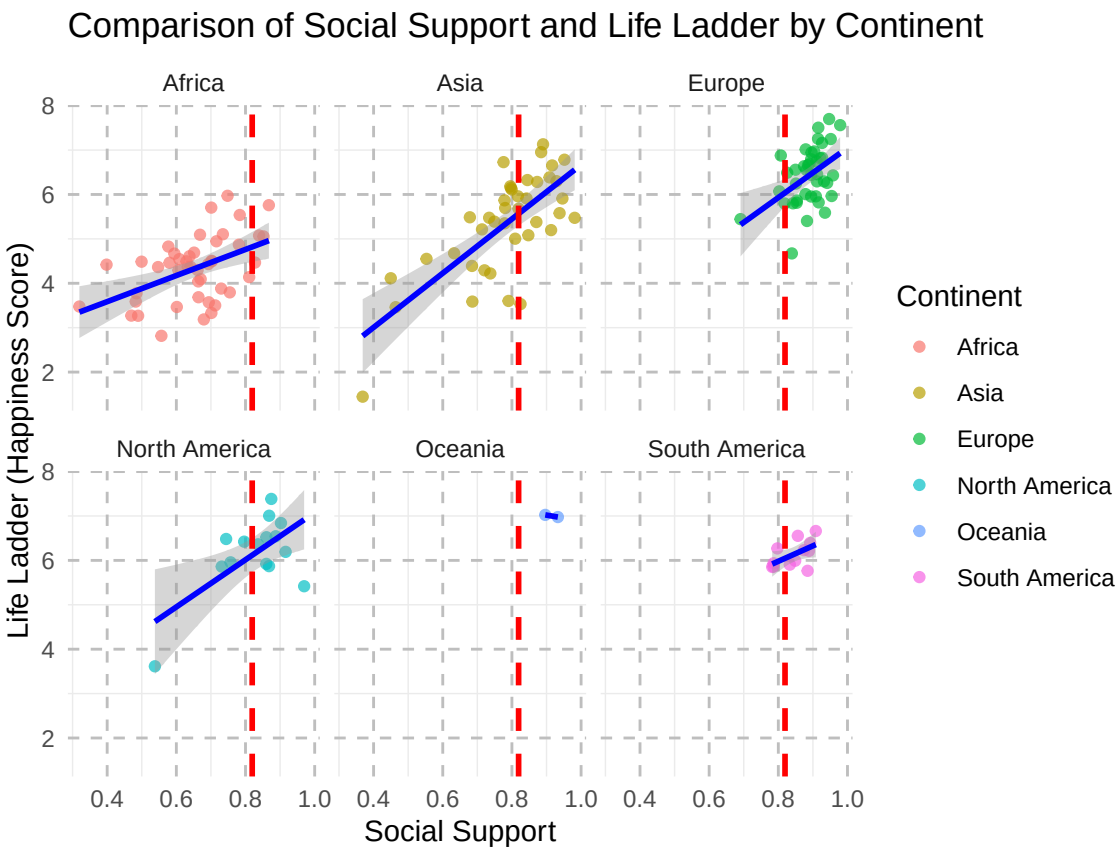
Figure 1: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.

```
## Warning: Removed 2 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning in qt((1 - level)/2, df): NaNs produced
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```

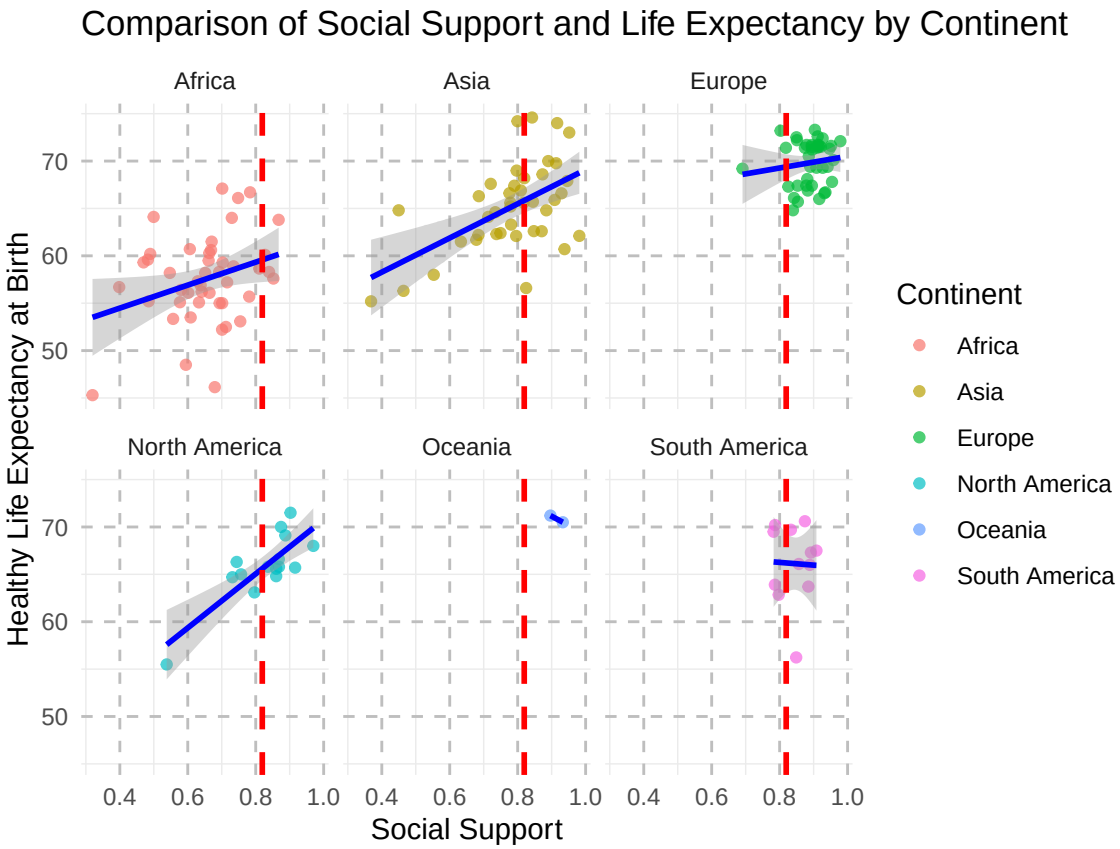



```
## Warning: Removed 3 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

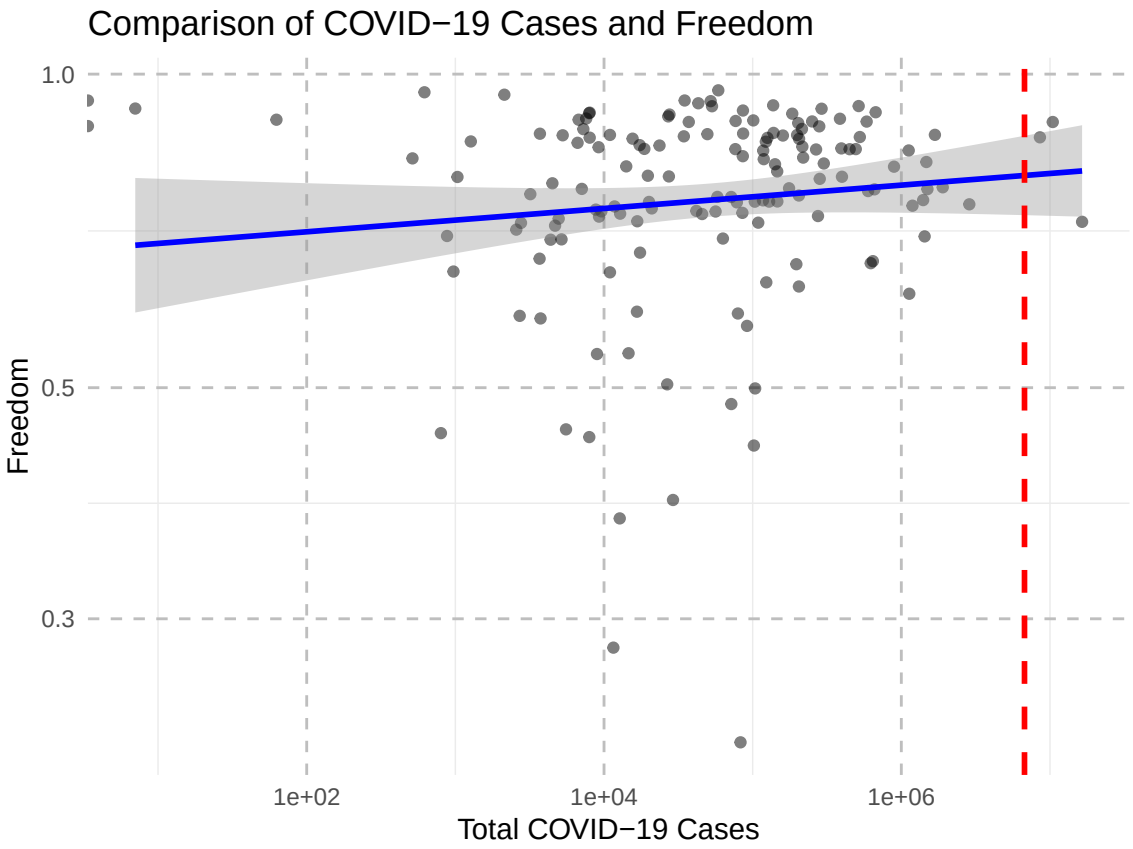
```
## Warning in qt((1 - level)/2, df): NaNs produced
```

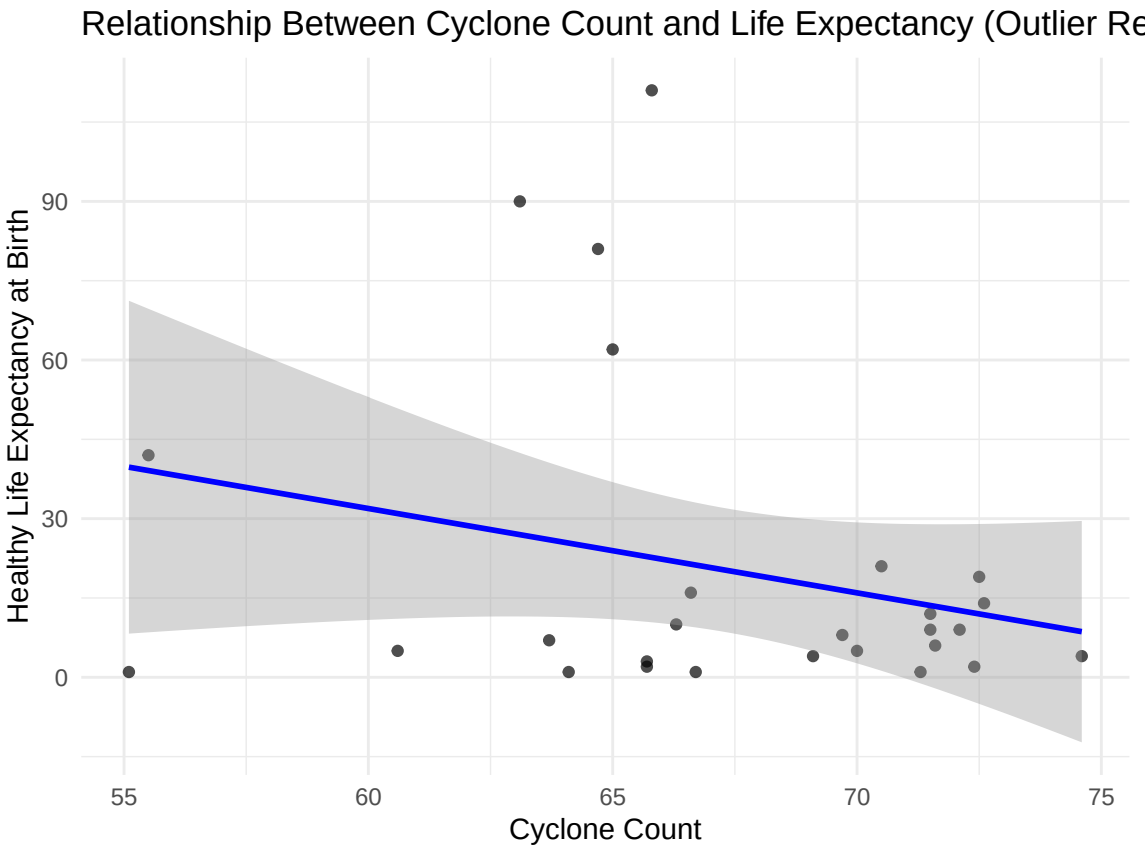
```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```

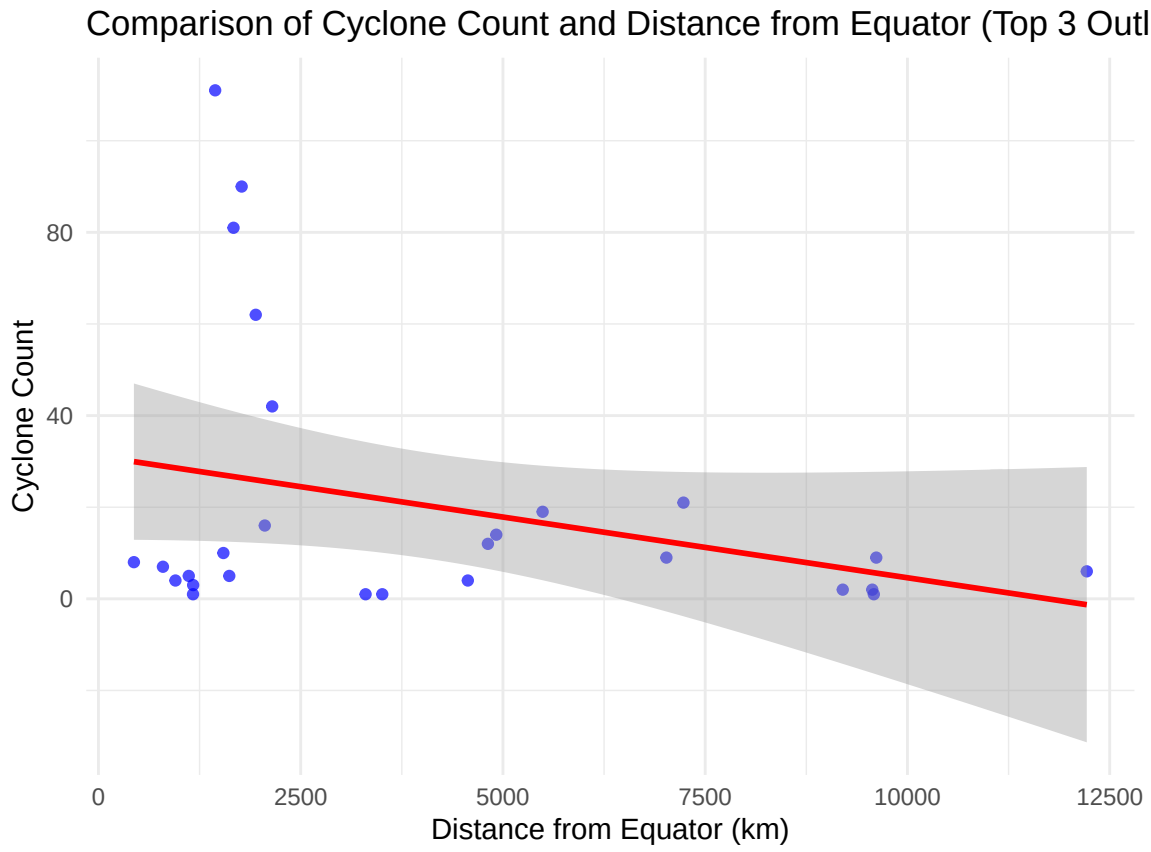



```
## Warning in scale_x_log10(): log-10 transformation introduced infinite values.

## Warning in scale_x_log10(): log-10 transformation introduced infinite values.

## Warning: Removed 8 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 6 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

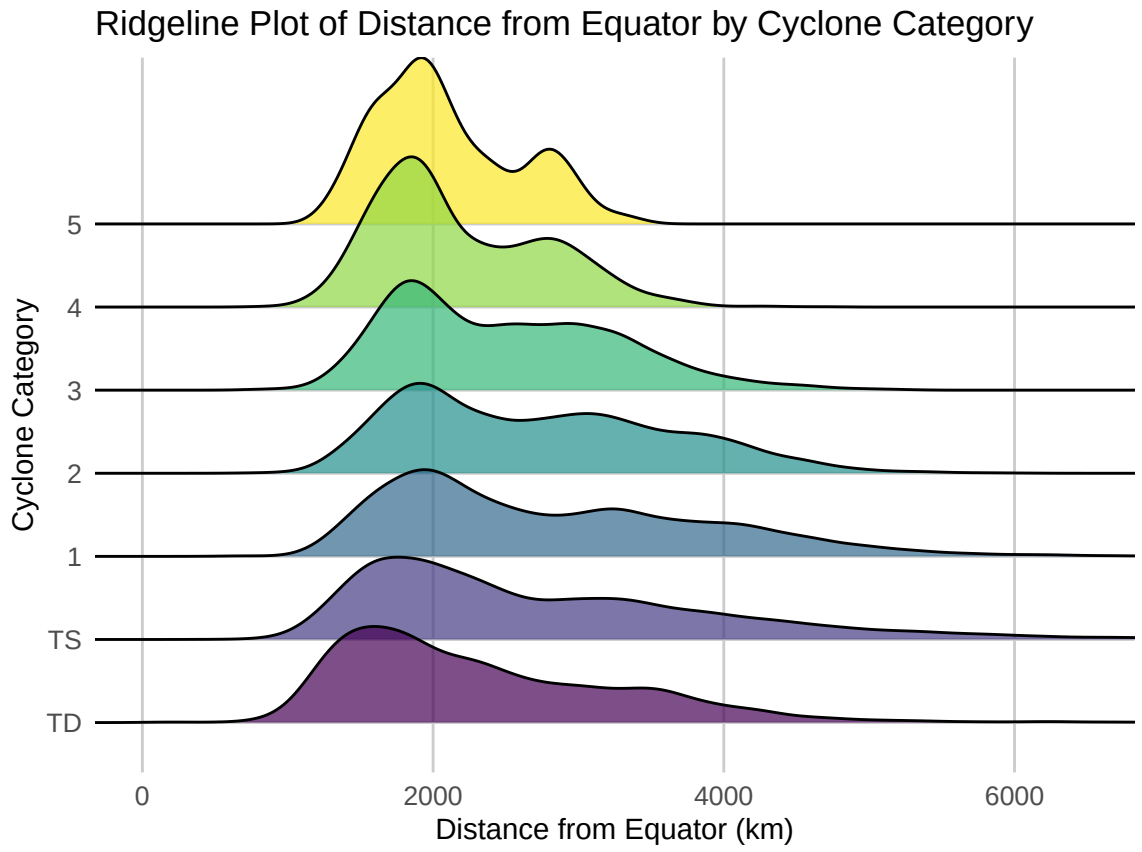

Comparison of COVID−19 Cases and Freedom
1.0
0.5
0.3
Freedom
1e+02
1e+04
1e+06
Total COVID−19 Cases



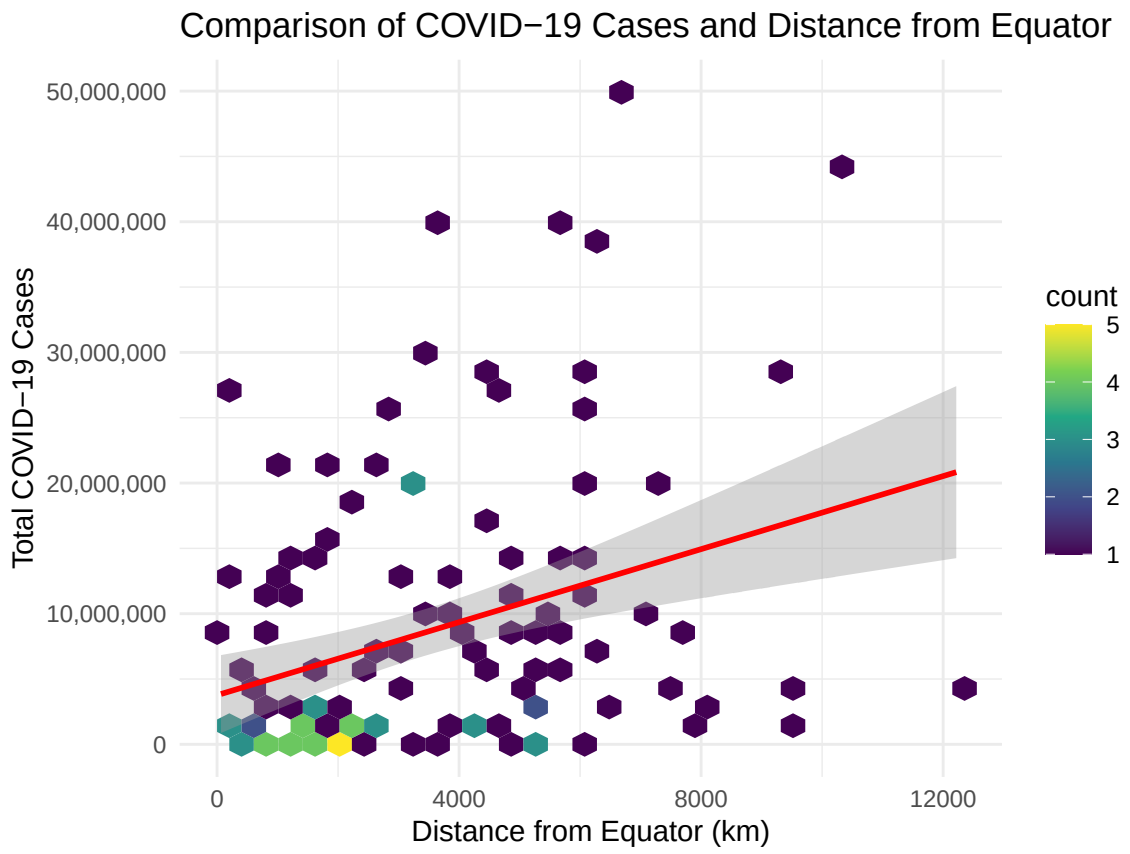

```
##  [1] "country"                          "year"
##  [3] "life_ladder"                      "log_gdp_per_capita"
##  [5] "social_support"                   "healthy_life_expectancy_at_birth"
##  [7] "freedom_to_make_life_choices"     "generosity"
##  [9] "perceptions_of_corruption"        "positive_affect"
## [11] "negative_affect"                  "hemisphere"
## [13] "distance_to_equator_km"           "geometry"
```



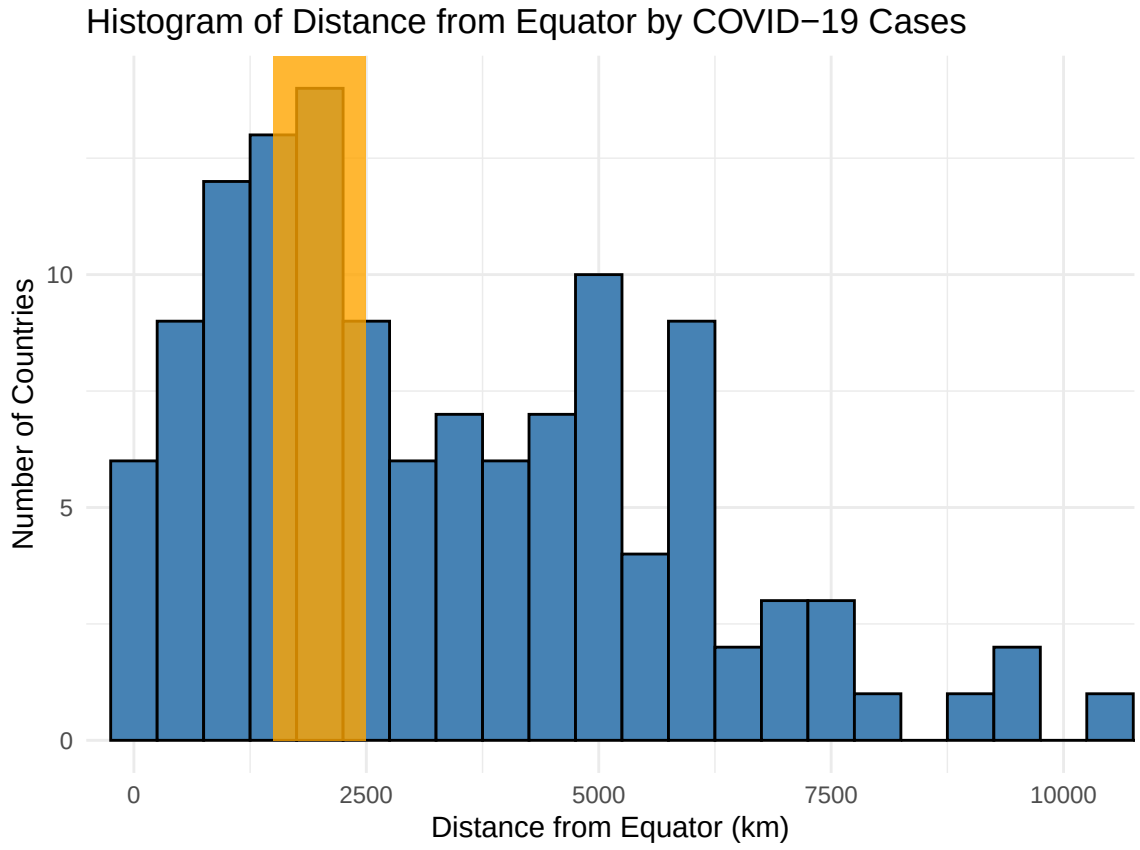

```
## Warning in geom_density_ridges(scale = 2, alpha = 0.7, color = "black", :
## Ignoring unknown parameters: ‘size‘
```

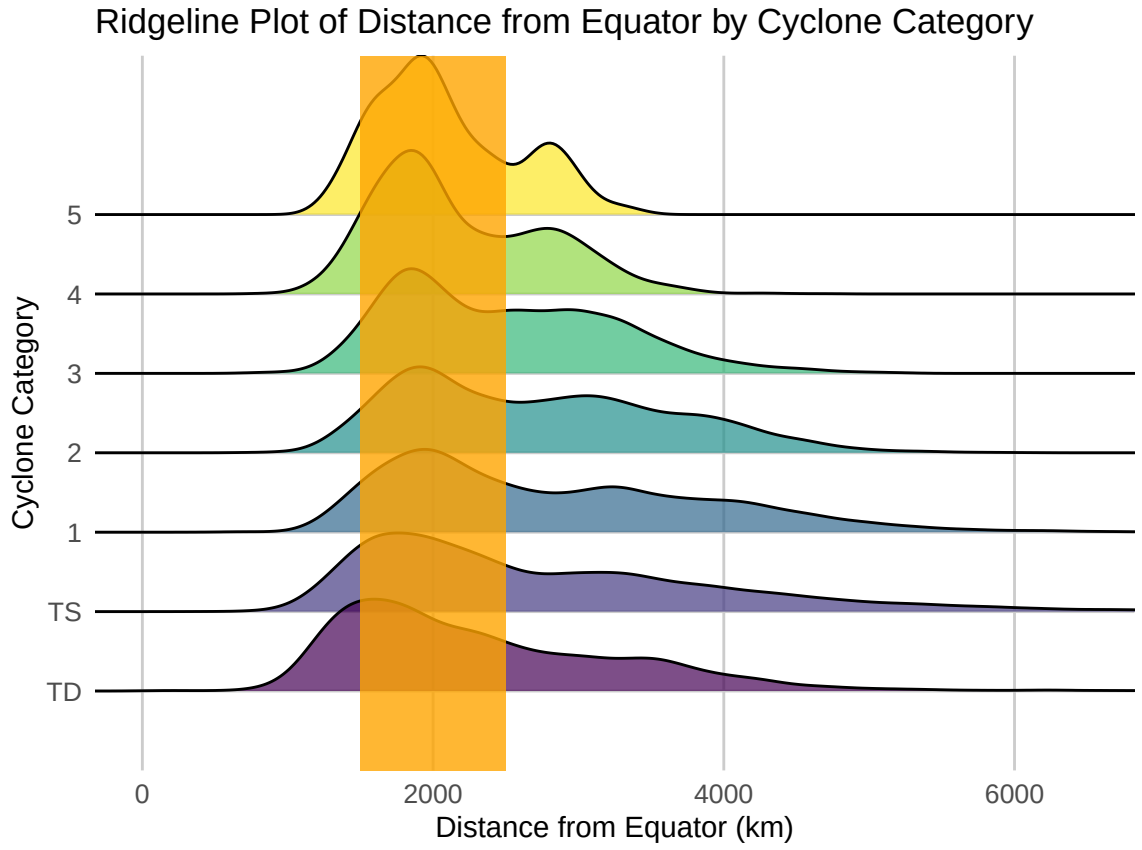



```
## # A tibble: 126 x 3
## # Groups:   country [126]
##    country     total_cases distance_to_equator_km
##    <chr>             <dbl>                  <dbl>
##  1 Afghanistan     8284992                  4017.
##  2 Albania         3444855                  5035.
##  3 Algeria        10188569                  3302.
##  4 Angola          1275899                  1386.
##  5 Armenia        14136861                  4909.
##  6 Australia       4960098                  3028.
##  7 Austria        19892748                  6040.
##  8 Azerbaijan     11877460                  4910.
##  9 Bahrain        12565889                  3004.
## 10 Belarus        19679206                  7091.
## # i 116 more rows
```

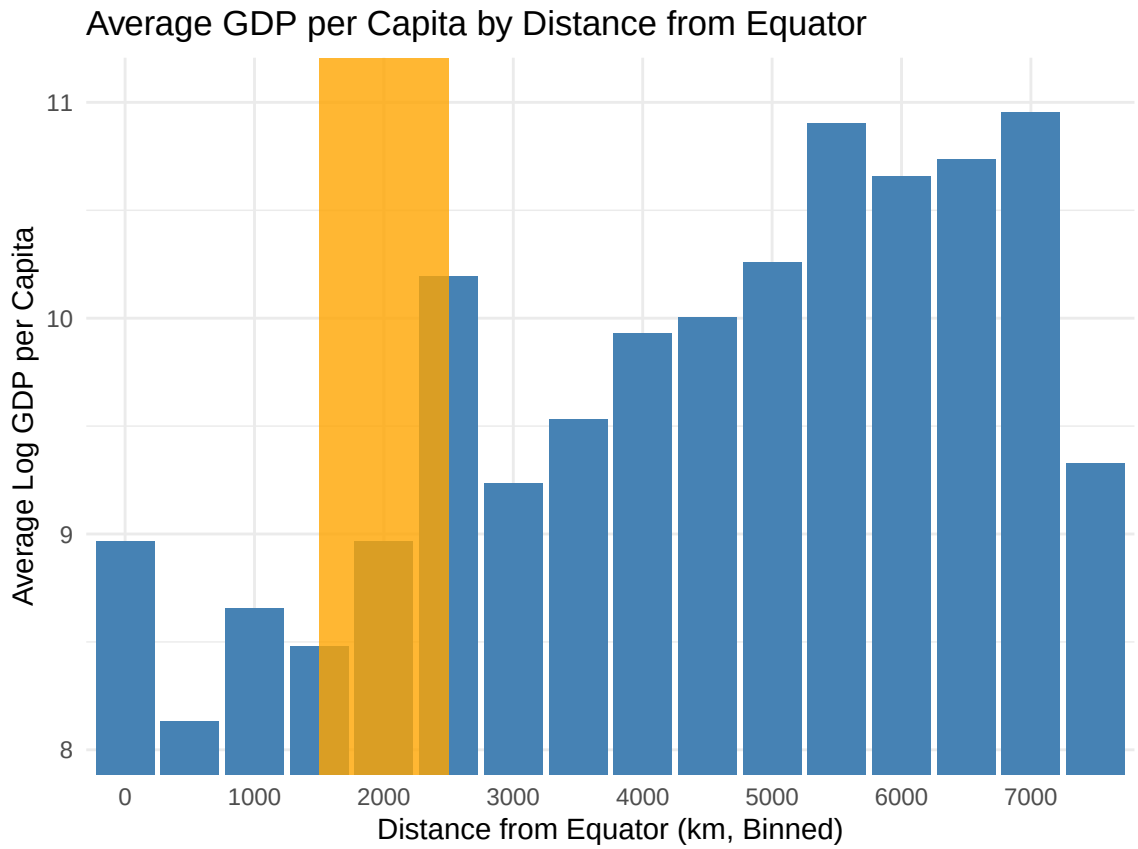


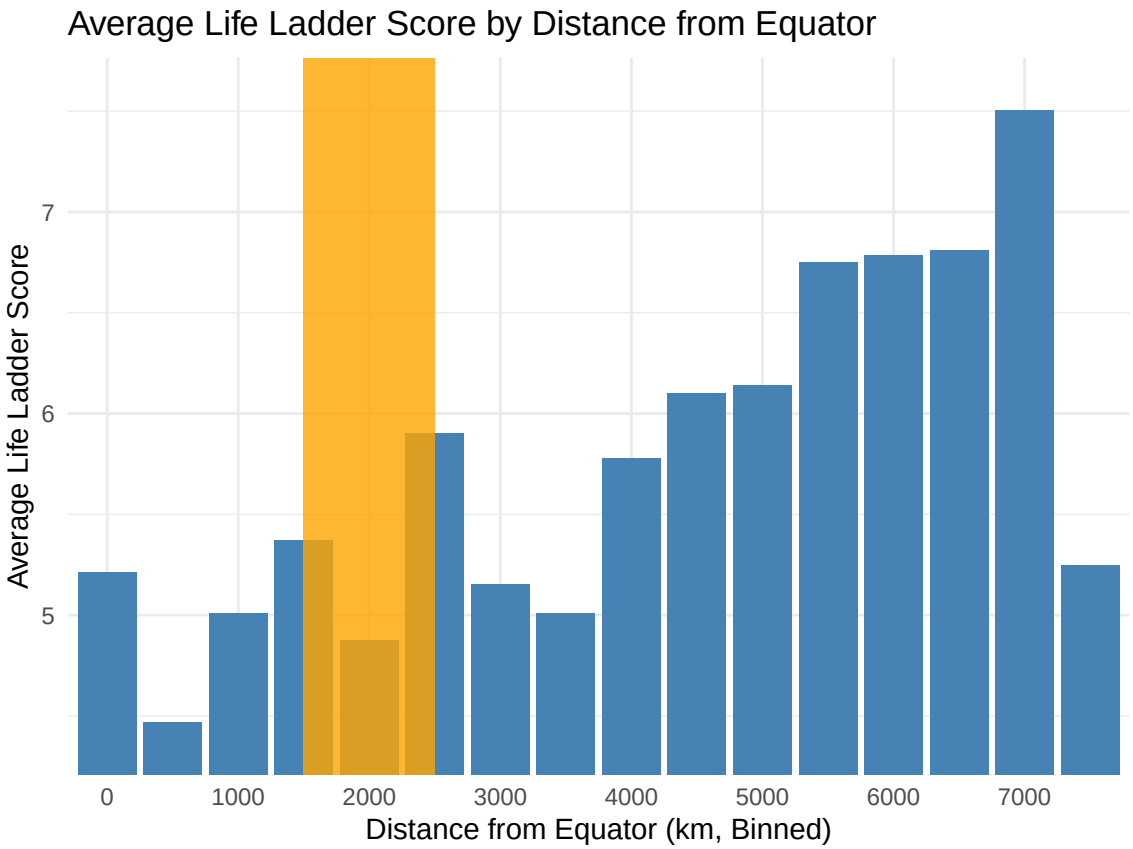

```
##  [1] "country"                          "year"
##  [3] "life_ladder"                      "log_gdp_per_capita"
##  [5] "social_support"                   "healthy_life_expectancy_at_birth"
##  [7] "freedom_to_make_life_choices"     "generosity"
##  [9] "perceptions_of_corruption"        "positive_affect"
## [11] "negative_affect"                  "hemisphere"
## [13] "distance_to_equator_km"           "geometry"
```


Histogram of Distance from Equator by COVID−19 Cases
Number of Countries
0
5
10
0
2500
5000
7500
10000
Distance from Equator (km)



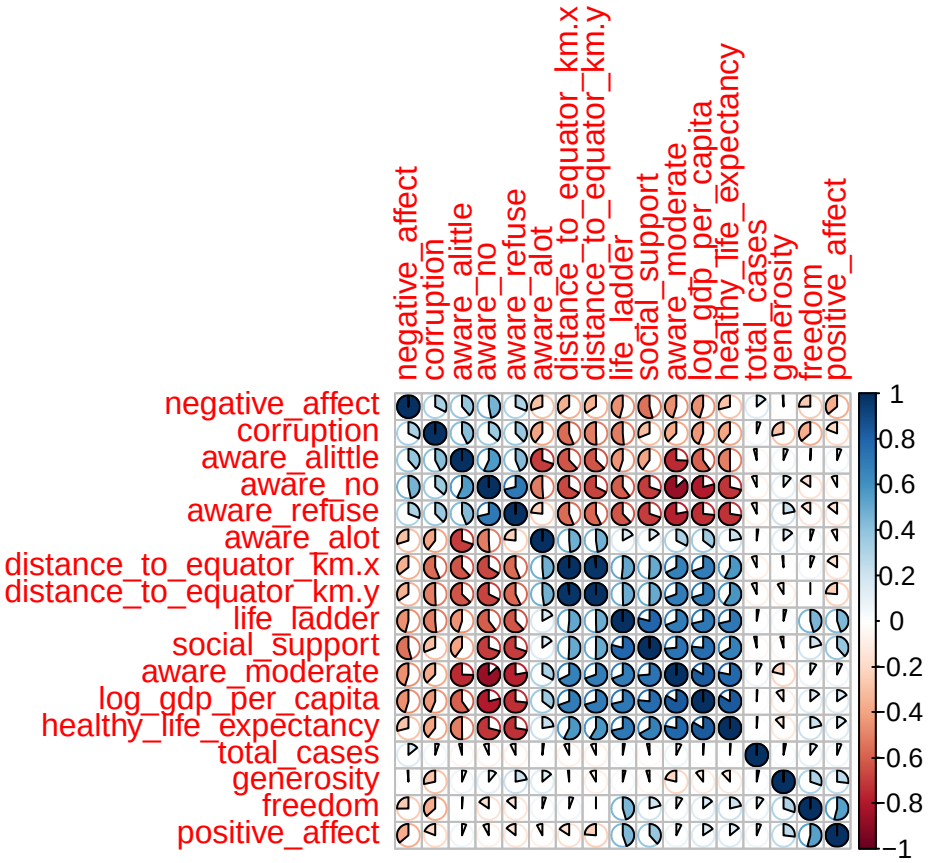

```
## # A tibble: 165 x 16
## # Groups:   country [165]
##    country      year life_ladder log_gdp_per_capita social_support
##    <chr>       <dbl>       <dbl>              <dbl>          <dbl>
##  1 Afghanistan  2023        1.45              NA             0.368
##  2 Albania      2023        5.44               9.69          0.691
##  3 Algeria      2022        5.54               9.32          0.783
##  4 Angola       2014        3.79               9.01          0.755
##  5 Argentina    2023        6.39               9.99          0.892
##  6 Armenia      2023        5.68               9.73          0.819
##  7 Australia    2023        7.02              10.8           0.896
##  8 Austria      2023        6.64              10.9           0.874
##  9 Azerbaijan   2023        5.21               9.64          0.713
## 10 Bahrain      2023        5.96              10.9           0.817
## # i 155 more rows
## # i 11 more variables: healthy_life_expectancy_at_birth <dbl>,
## #   freedom_to_make_life_choices <dbl>, generosity <dbl>,
## #   perceptions_of_corruption <dbl>, positive_affect <dbl>,
## #   negative_affect <dbl>, hemisphere <chr>, distance_to_equator_km.x <dbl>,
## #   geometry.x <MULTIPOLYGON [m]>, distance_to_equator_km.y <dbl>,
## #   geometry.y <MULTIPOLYGON [°]>
```


Average GDP per Capita by Distance from Equator
Average Log GDP per Capita
11
10
9
8
0
1000
2000
3000
4000
5000
6000
7000
Distance from Equator (km, Binned)



This insight is supported by the summary statistics in table @ref(tab:summary_stats)

«««< HEAD




```
## # A tibble: 6 x 2
##   continent     Social_Support_Life_Ladder_correlation
##   <chr>                                          <dbl>
## 1 Africa                                         0.480
## 2 Asia                                           0.725
## 3 Europe                                         0.461
## 4 North America                                  0.638
## 5 Oceania                                       -1
## 6 South America                                  0.550
```

The next insight that we found is shown in @ref(fig:insight2).

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
## Warning: Removed 12 rows containing non-finite outside the scale range
## (‘stat_smooth()‘).
```

```
## Warning: Removed 12 rows containing missing values or values outside the scale range
## (‘geom_point()‘).
```

Finally, @ref(fig:insight3) shows ...

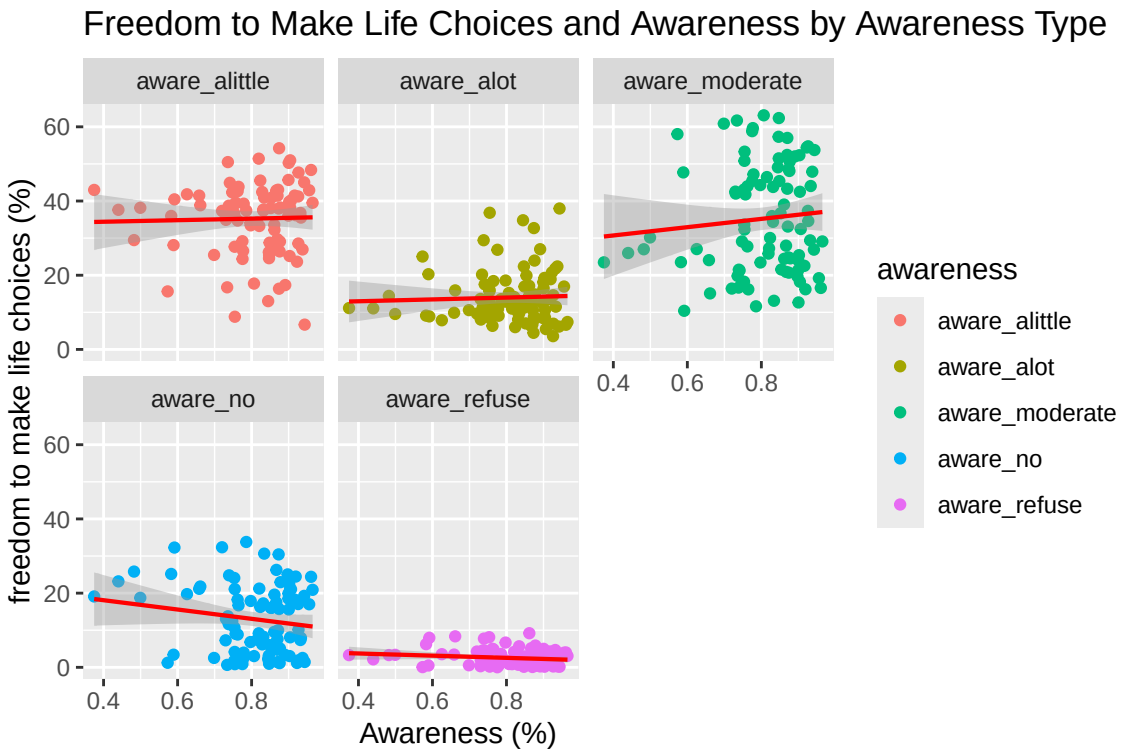
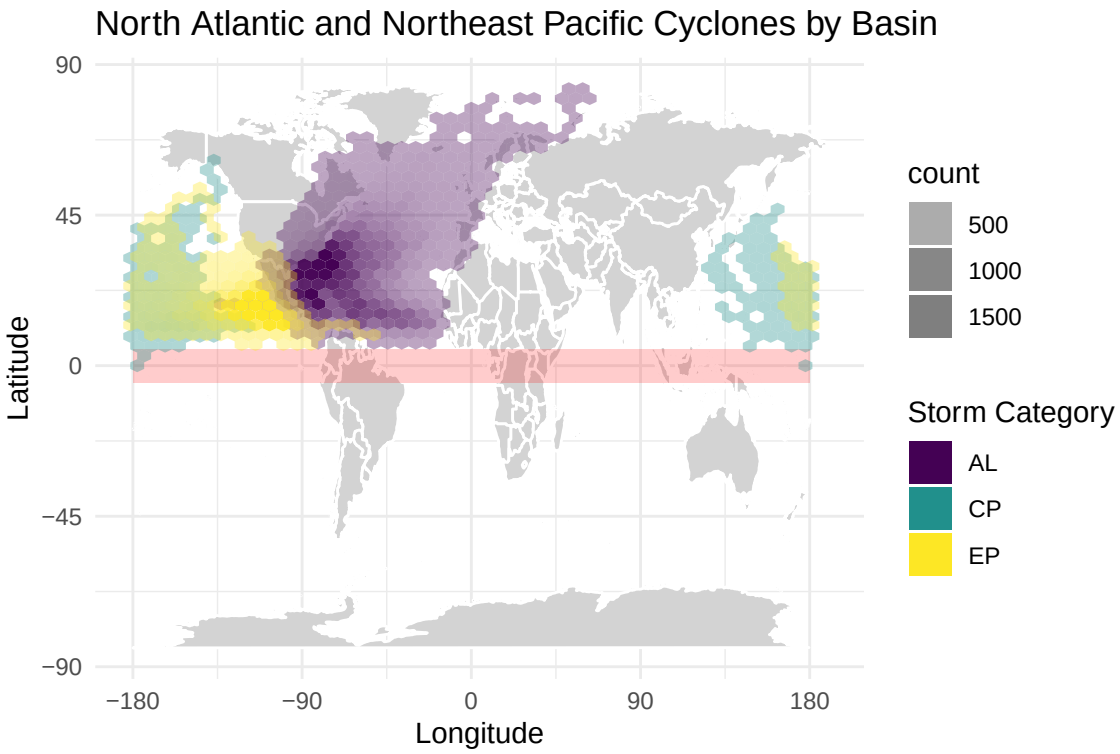
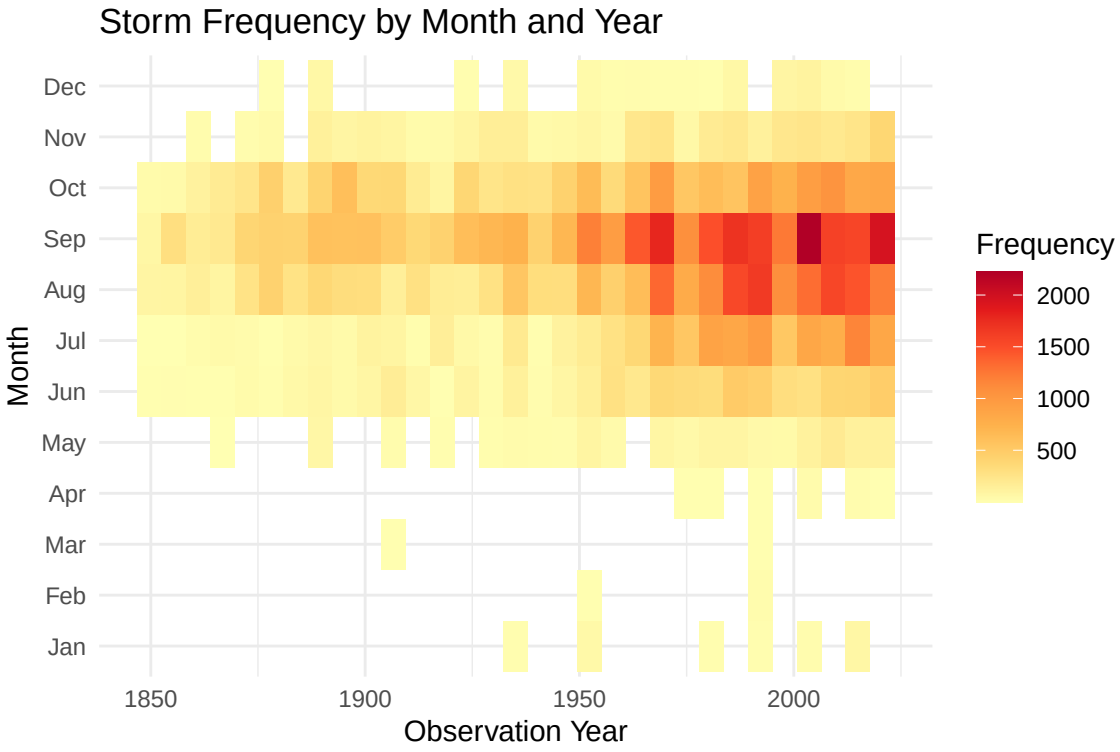
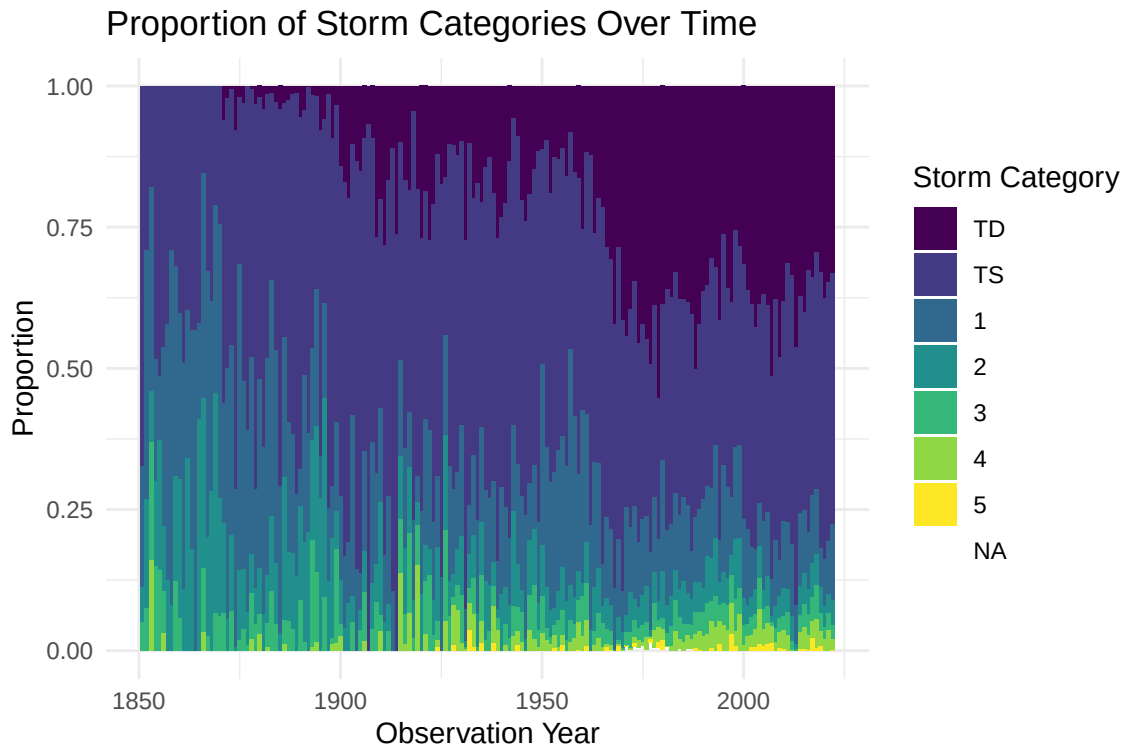


Figure 2: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.

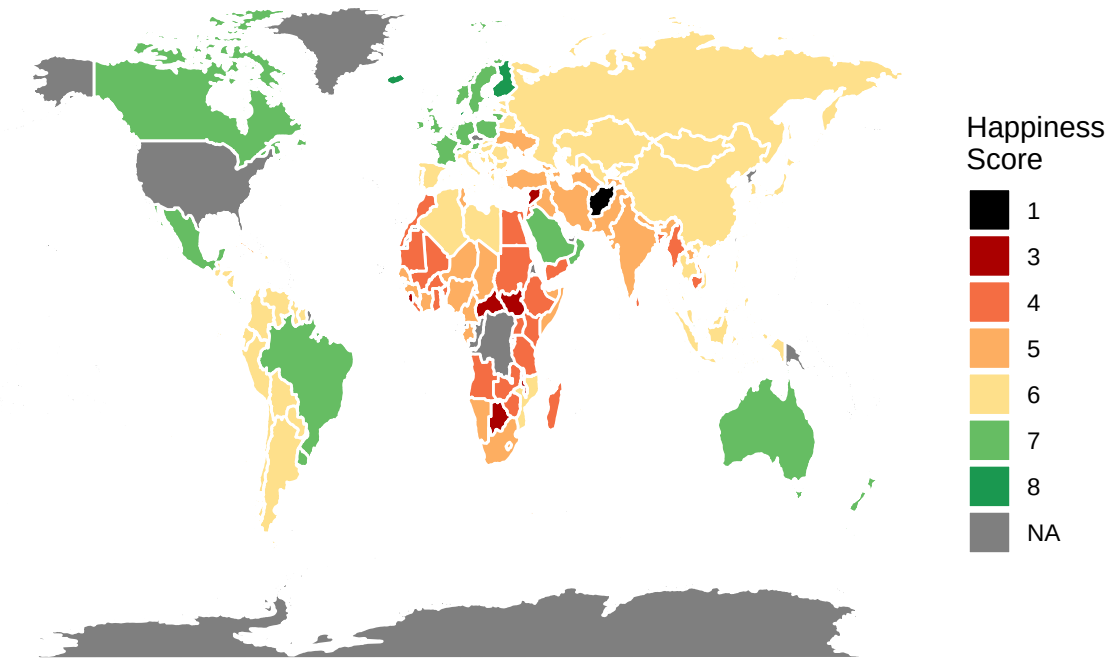
```
## Warning: The dot-dot notation (‘..count..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(count)‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```


North Atlantic and Northeast Pacific Cyclones by Basin
90
45
0
−45
−90
Latitude
−180
−90
0
90
180
Longitude
count
500
1000
1500
Storm Category
AL
CP
EP


Storm Frequency by Month and Year
Dec
Nov
Oct
Sep
Aug
Jul
Jun
May
Apr
Mar
Feb
Jan
Month
1850
1900
1950
2000
Observation Year
Frequency
2000
1500
1000
500



World Happiness Scores by Country



Conclusion and Future Work

Overall, we found <>.

A second paragraph about our findings.

The next steps in this analysis are...

The limitations of this analysis are as follows. (Do not simply list potential issues with sampling, but relate them to your analysis and how they affect your conclusions. An honest and complete acknowledgement of the limitations makes the analysis more trustworthy.)

References

I am not strict about MLA or APA style or anything like that. For this report, I would much rather have your citations be easy to match to your insights.

The easiest way is to use Rmd's footnote syntax. This will put a number beside the word where the footnote appears, and the full text of the footnote at the bottom of the page (pdf) or end of the document (html). The syntax is:[1], where I suggest that you put in something like this[2] to make references for this assignment.

Alternatively, you could make a list of citations with their main arguments and why they're relevent to your insights, methods, etc.

The link above also references "bibtex" files. These are also extremely convenient, but have a steep learning curve and they make it difficult to tie them to an insight. If you use bibtext, then make sure that you provide a sentence to describe the source and it's relevance when you cite it - don't just add citations to the end of a sentence (this is common practice in academia, but I want to know that your citations are directly relevant for this assignmnet).

[1] See the source view to see this footnote

[2] The relevance to the insight is ... . From <>, published on <>, url: <>