Investigation on World Happiness, Covid Deaths and Vaccination

fa21-prj-shiyuan8-sw20-zihanz12-zl32
Shiyuan Wang, Songyuan Wang, Zihan Zhou, Kimmy Liu

Project Outline

1. Visualize the life satisfaction (happiness index) across the globe.
2. Visualize the trend of Covid-19 cases and mortality rates, and combine the world happiness with total Covid-19 cases and deaths.
3. Combine the world happiness with vaccination rate, and visualize the correlation between Covid-19 vaccination rates and world happiness.
4. Use Anova analysis to figure the impacts of Covid-19 cases, deaths, vaccination rates and different factors across fields, such as public policy, public health, economics etc. on the happiness index.
5. Explore as many R packages as we can.

Data Source

- Covid-19 daily cases and excess deaths
- Covid-19 daily vaccination rates
- World happiness

For above datasets, we remove all the missing values in the dataset of Covid-19 daily cases, excess deaths, and vaccination rates. Then we calculate the total number of Covid-19 daily cases, excess deaths, and vaccination rates in 2020. After that, we merge them with the world happiness data, on the column of country name.

Analysis on world-happiness dataset

Data cleaning and data wrangling:

First, we included libraries such as :data.table, tidyverse, ggrepel, sf, dplyr, ggplot2, etc. for data cleaning, data visuilazation;

Then, we used Package(countrycode) to assign each country to according region, and add the column to data.table(happiness)

```
head(happiness,4) # Before assigning each country to according region
```

```
##         Entity Year
## 1: Afghanistan 2008
## 2: Afghanistan 2009
## 3: Afghanistan 2010
## 4: Afghanistan 2011
##    Life satisfaction in Cantril Ladder (World Happiness Report 2019)
## 1:                                                             3.724
## 2:                                                             4.402
## 3:                                                             4.758
## 4:                                                             3.832
```

```
head(happiness,4) # After assigning each country to according region
```

```
##         Entity Year life_satisfaction     region
## 1: Afghanistan 2008            3.724 South Asia
## 2: Afghanistan 2009            4.402 South Asia
## 3: Afghanistan 2010            4.758 South Asia
## 4: Afghanistan 2011            3.832 South Asia
```

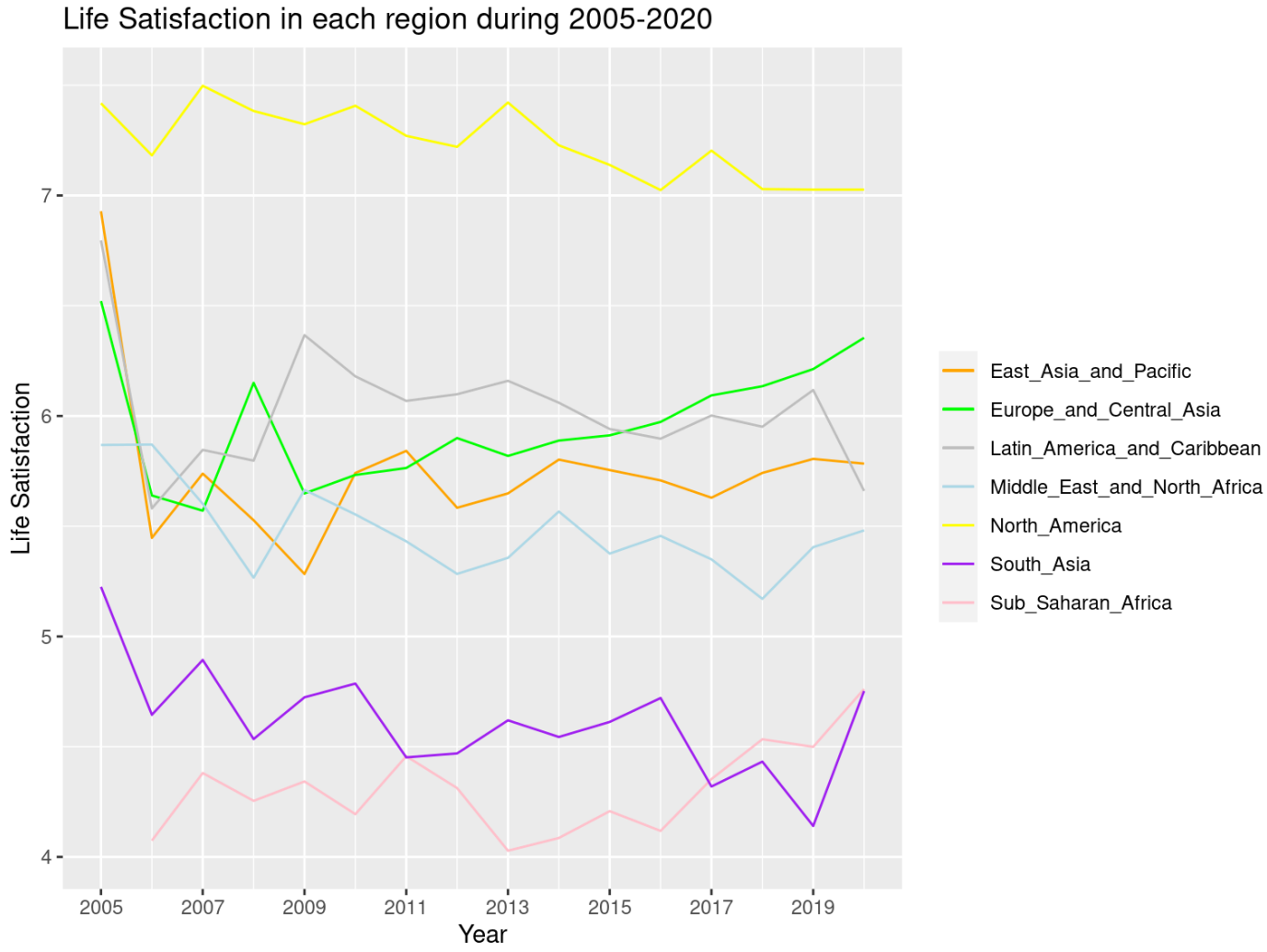
After that, we implemented 'dcast' method to create desired data.table, and change the columnnames to avoid "&" and blank space. For example:

```
head(happiness_1,2)
```

```
##    Year East_Asia_and_Pacific Europe_and_Central_Asia
## 1: 2005              6.928500                6.521267
## 2: 2006              5.447154                5.639417
##    Latin_America_and_Caribbean Middle_East_and_North_Africa North_America
## 1:                       6.796                       5.8684         7.418
## 2:                       5.581                       5.8704         7.182
##    South_Asia Sub_Saharan_Africa
## 1:    5.22500                NaN
## 2:    4.64475           4.074182
```

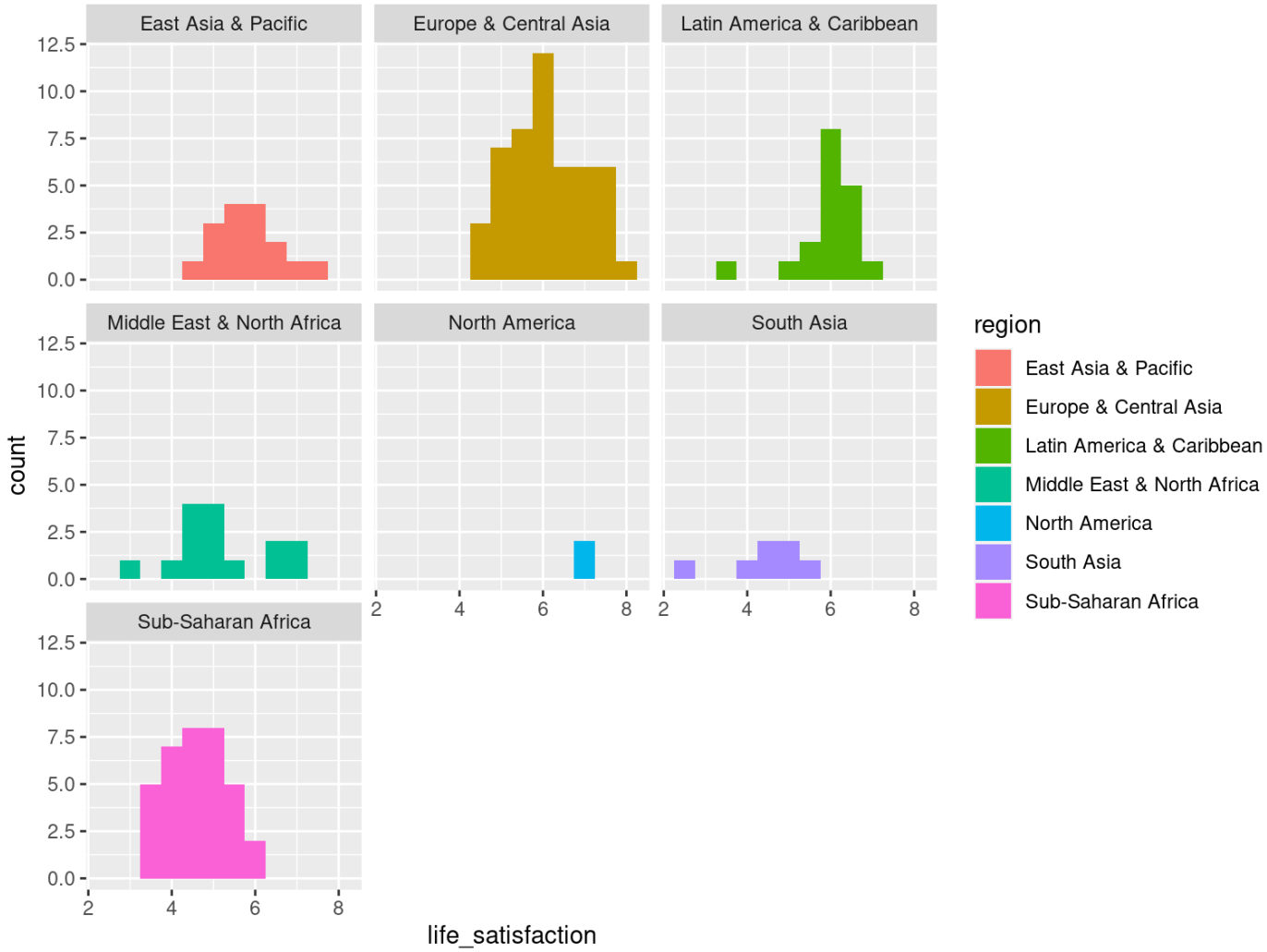
Data visulization

First, we plot "Life Satisfaction in each region during 2005-2020":



From this graph, we can see that: Among all these 7 regions, North_America has the highest level of life-satisfaction, and sub_saharan_africa has the lowest level of life-satisfaction for almost each year except year 2017-2019. There is also an decrease for each region(except sub_saharan_africa) during 2005-2020.

Then we plot plot "Life Satisfaction in 2018" for each country:



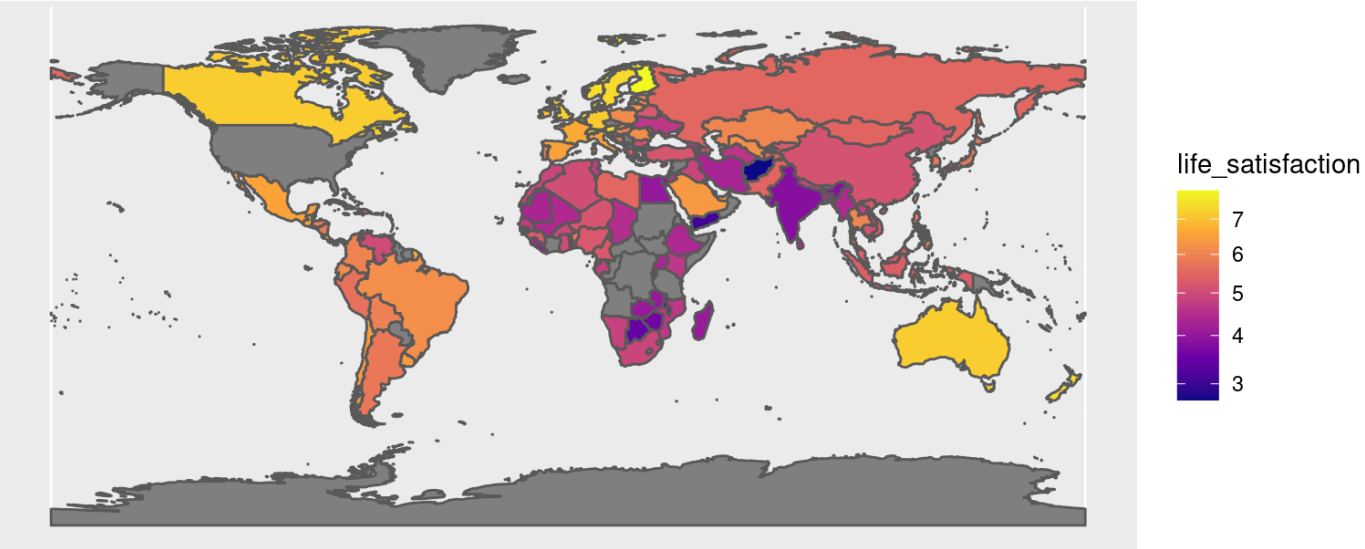
From the histogram graph above, we can see that in year 2018, most countries in Europe&central asia, Lation America& Caribbran, and East Asia&Pacific has life_satisfaction over 5, while most of coutries in Sub-Saharan Africa and South Asia has life_satisfaction less than 5.There exits huge disparities worldwide.

After that, we Create world map for "life satisfaction" in each country in year 2018:

First, we merge dataset "world" and dataset "h1", and transform the output(world_) from data.frame to sf using method 'st_as_sf';

Then,we plot the map:

```
ggplot(data = world_) + geom_sf(aes(fill = life_satisfaction)) +
scale_fill_viridis_c(option = "plasma", trans = "sqrt")
```

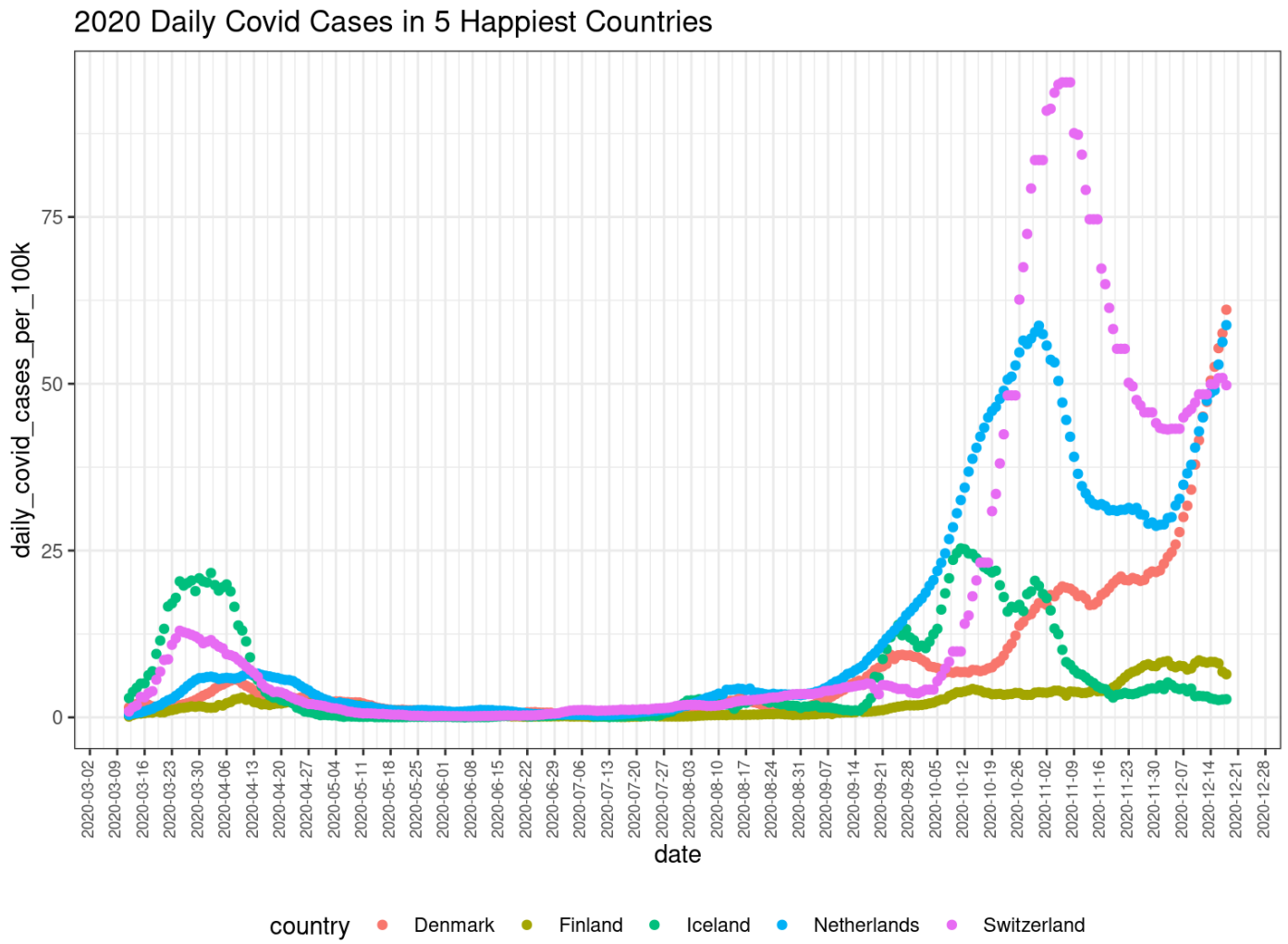


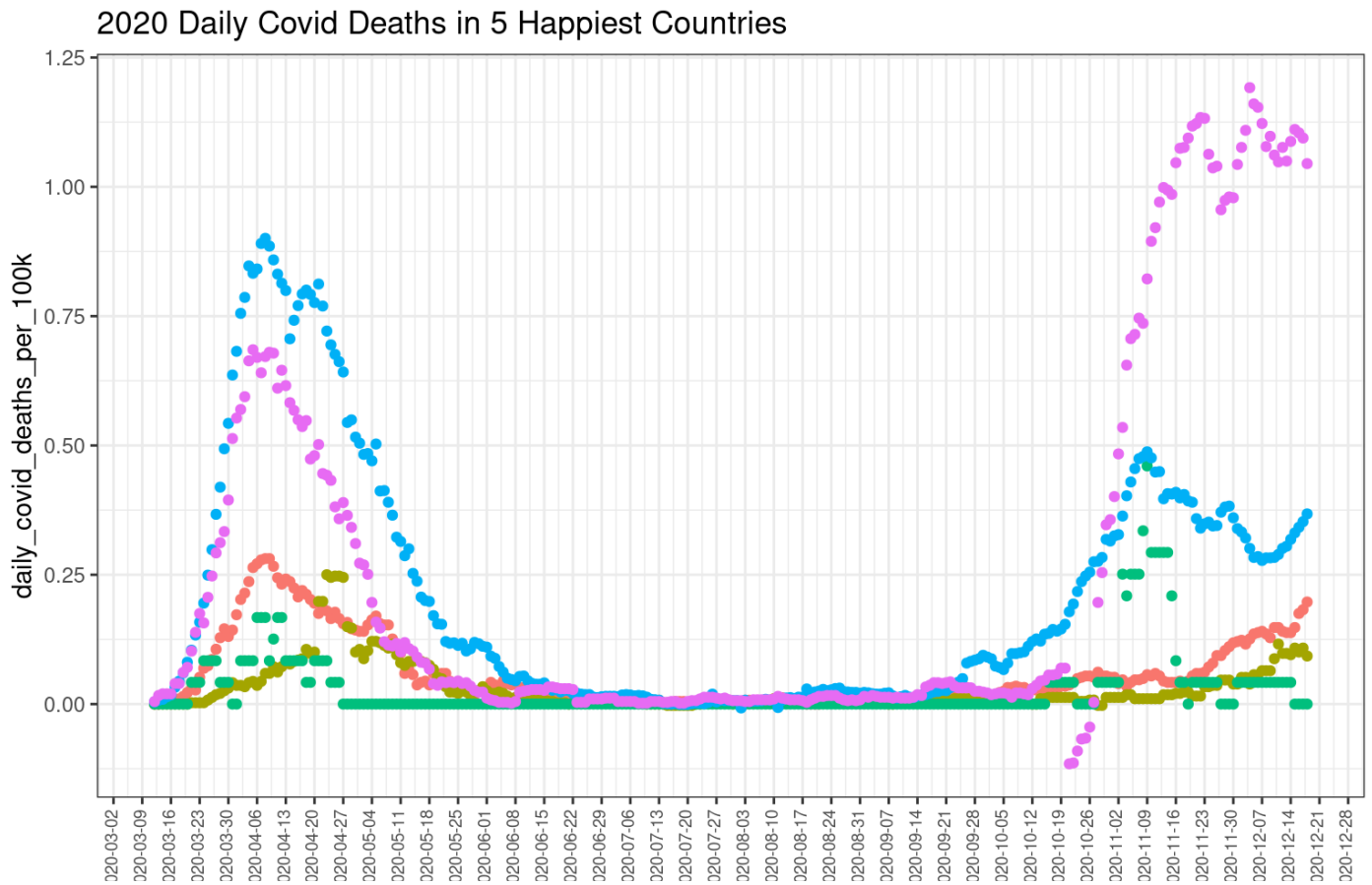
From the map above, we can see that: grey parts stand for countries without data recorded in year 2018; and for other parts in this world map,from colour yellow to blue, the darker the colour is for each country, the lower life_satisfaction score the coutry has. We can see that Canada and Australia and most parts of Europe has high level of life_satisfaction, while most parts of Africa has low level of life_satisfaction.

Analysis on the trend of Covid-19 daily cases and excess deaths.

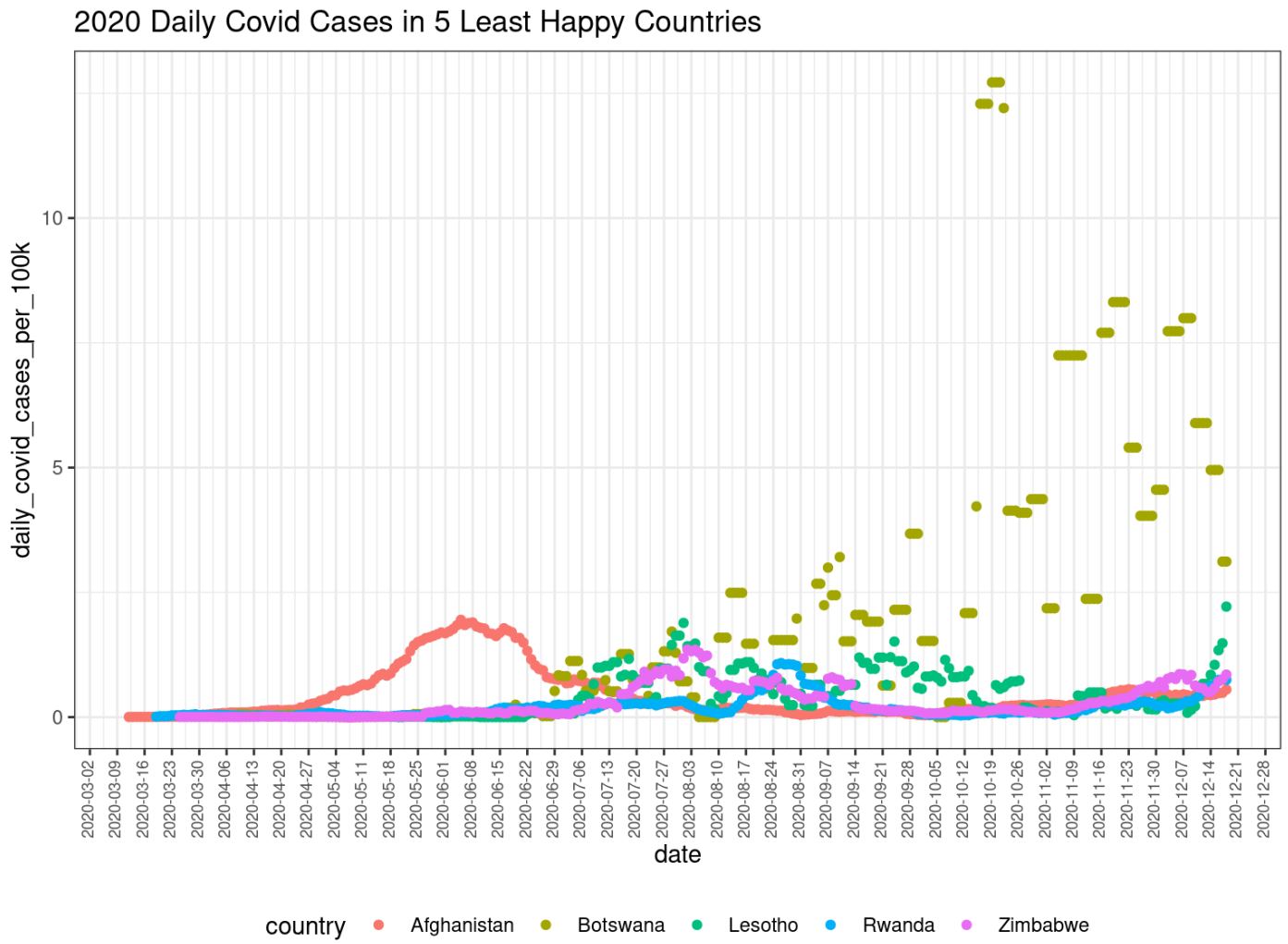
First, we only want to analyze the countries which have a happiness index.

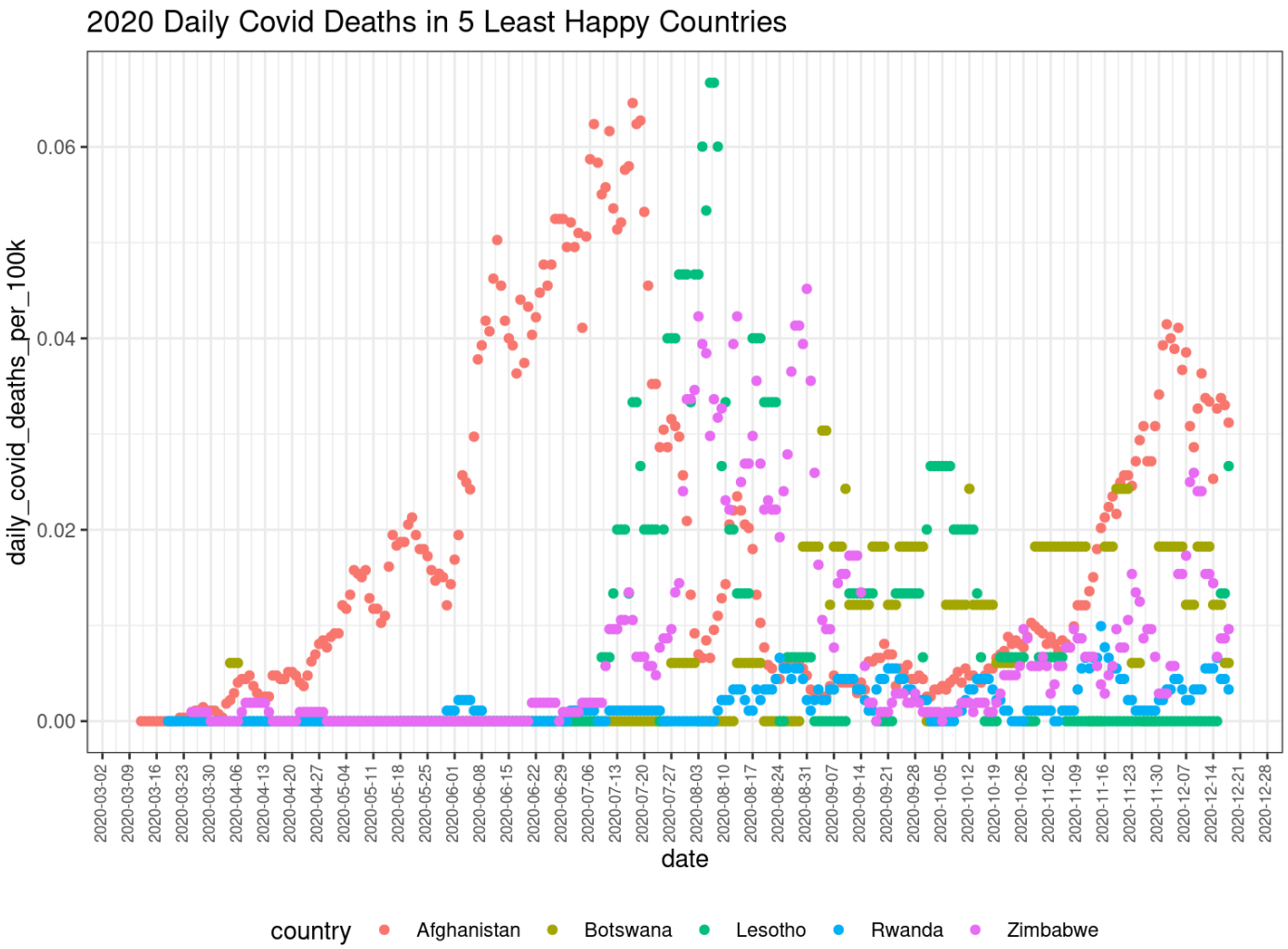
Then we want to visualize the Covid-19 cases and deaths in the happiest and least happiest country in the world, and try to see whether there exists a specific pattern.


2020 Daily Covid Cases in 5 Happiest Countries
daily_covid_cases_per_100k
75
50
25
0
2020-03-02
2020-03-09
2020-03-16
2020-03-23
2020-03-30
2020-04-06
2020-04-13
2020-04-20
2020-04-27
2020-05-04
2020-05-11
2020-05-18
2020-05-25
2020-06-01
2020-06-08
2020-06-15
2020-06-22
2020-06-29
2020-07-06
2020-07-13
2020-07-20
2020-07-27
2020-08-03
2020-08-10
2020-08-17
2020-08-24
2020-08-31
2020-09-07
2020-09-14
2020-09-21
2020-09-28
2020-10-05
2020-10-12
2020-10-19
2020-10-26
2020-11-02
2020-11-09
2020-11-16
2020-11-23
2020-11-30
2020-12-07
2020-12-14
2020-12-21
2020-12-28
date
country
Denmark
Finland
Iceland
Netherlands
Switzerland


2020 Daily Covid Deaths in 5 Happiest Countries
daily_covid_deaths_per_100k
1.25
1.00
0.75
0.50
0.25
0.00


2021
date
country
Denmark
Finland
Iceland
Netherlands
Switzerland

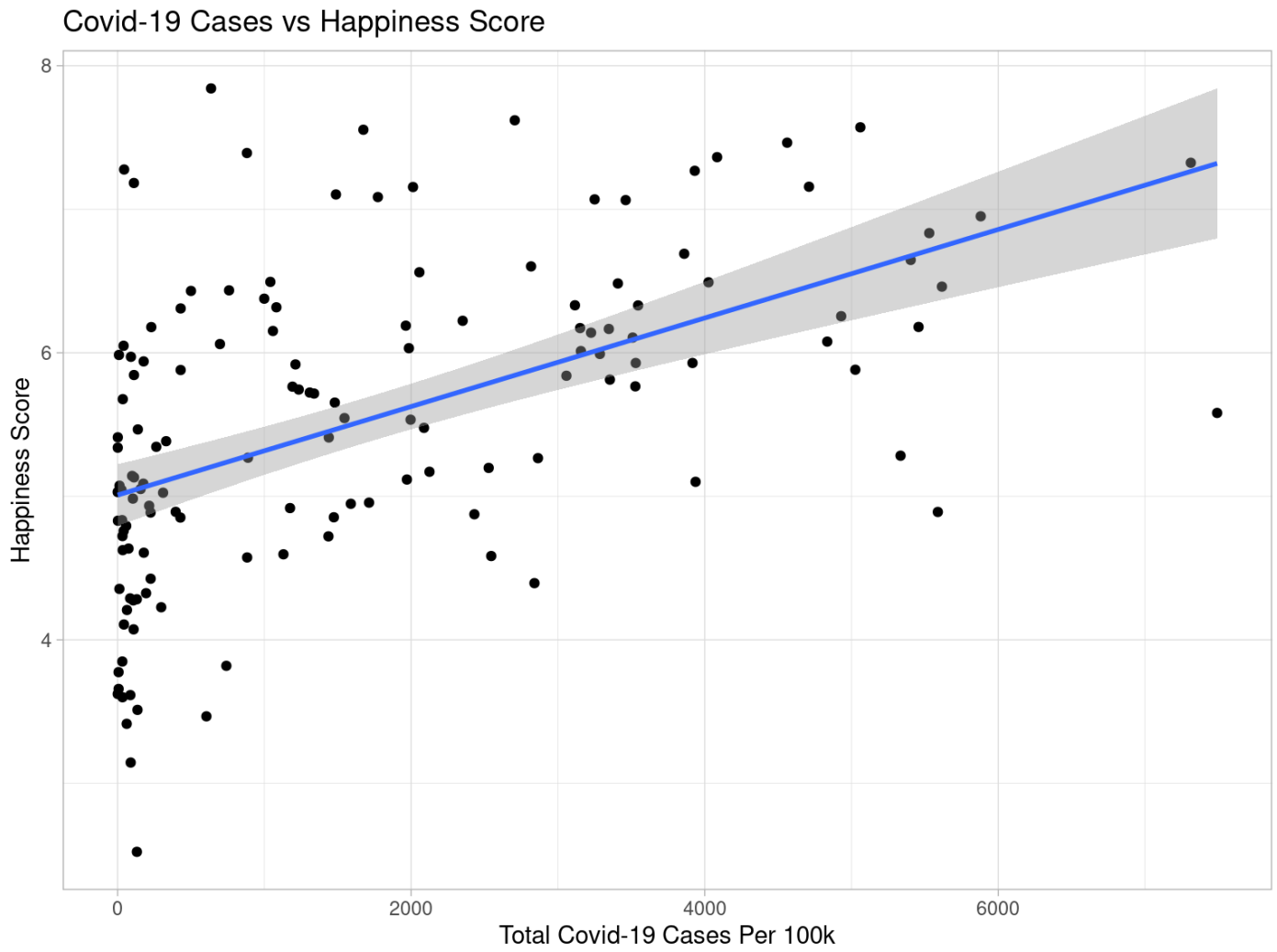

2020 Daily Covid Cases in 5 Least Happy Countries
daily_covid_cases_per_100k
10
5
0
2020-03-02
2020-03-09
2020-03-16
2020-03-23
2020-03-30
2020-04-06
2020-04-13
2020-04-20
2020-04-27
2020-05-04
2020-05-11
2020-05-18
2020-05-25
2020-06-01
2020-06-08
2020-06-15
2020-06-22
2020-06-29
2020-07-06
2020-07-13
2020-07-20
2020-07-27
2020-08-03
2020-08-10
2020-08-17
2020-08-24
2020-08-31
2020-09-07
2020-09-14
2020-09-21
2020-09-28
2020-10-05
2020-10-12
2020-10-19
2020-10-26
2020-11-02
2020-11-09
2020-11-16
2020-11-23
2020-11-30
2020-12-07
2020-12-14
2020-12-21
2020-12-28
date
country
Afghanistan
Botswana
Lesotho
Rwanda
Zimbabwe



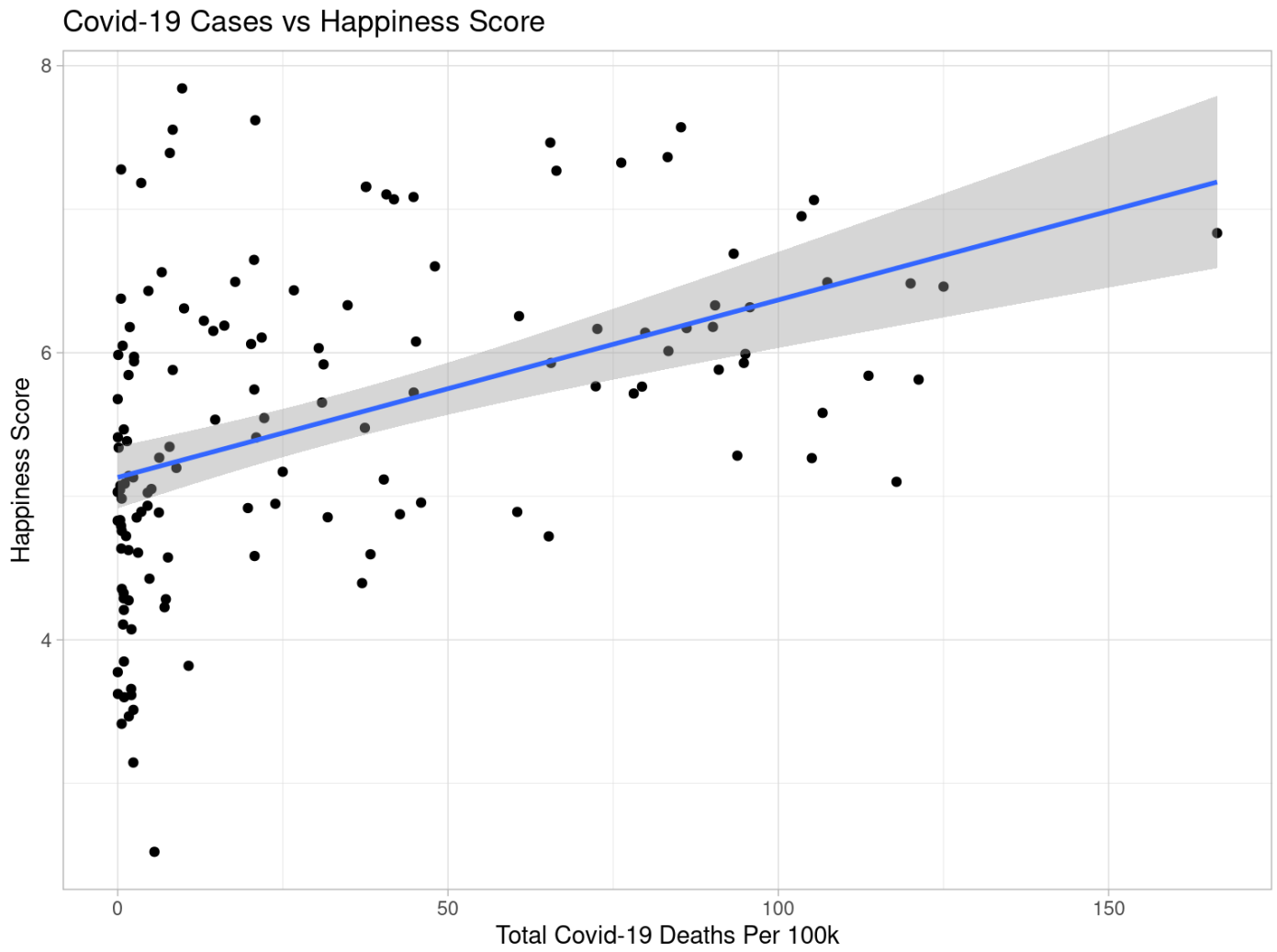
From those plots we can see that, the happiest countries had two spikes both in Covid-19 cases and deaths, one during the beginning of March, and one around November; while the least happiest countries witnessed rises in cases and deaths in the middle and the end of 2020. Overall, the happiest countries had more cases and excess deaths than the least happiest ones.

It's rather counter-intuitive that the countries with more Covid-19 deaths are happier, and thus we want to see the relationship between Covid-19 cases, deaths and happiness scores.

```
## `geom_smooth()` using formula 'y ~ x'
```




```
## `geom_smooth()` using formula 'y ~ x'
```



From this plot, we surprisingly found that the Covid-19 cases and deaths seem have a positive relationship with the happiness score.

After that, we want total values for each country, so we combine happiness data with covid data. In this step, we calculate the death_case_ratio, which the ratio of the number of deaths and the number of cases. The lower the ratio means there would be fewer people dying from the covid when they are infected, and intuitively, we would assume a country with a lower ratio would have a higher happiness index.

```
##           country total_cases total_deaths total_cases_per_100k
## 1:  United States    19463947    342637.00            5880.2995
## 2:         Brazil     7537457    192213.86            3546.0474
## 3:          India    10223903    148133.44             740.8602
## 4:         Mexico     1394837    123384.73            1081.8333
## 5:          Italy     2060800     72566.71            3408.4311
## 6: United Kingdom     2349819     71549.43            3461.4187
##    total_deaths_per_100k death_case_ratio
## 1:             103.51488       0.01760367
## 2:              90.42830       0.02550115
```

```
## 3:          10.73427      0.01448893
## 4:          95.69696      0.08845814
## 5:         120.02071      0.03521289
## 6:         105.39643      0.03044891
```

Analysis on the relationship between each country's happiness score and vaccination rate.

We are surprised to find that there is a strong positive relationship between the two.

First we do some data cleaning job. Since the raw data is a daily reported dataset, we get the latest vaccination rate by groupby each country and select the last rows from each group.

Get 2021 data by select from last row.

```
## # A tibble: 6 × 5
## # Groups:   location [6]
##   location  people_vaccinated people_fully_va… people_vaccinat… people_fully_va…
##   <chr>                 <dbl>            <dbl>             <dbl>            <dbl>
## 1 Afghanis…           4397449          3566192              11.0             8.95
## 2 Africa            160680802        105893196              11.7             7.71
## 3 Albania             1087187           978333              37.8            34.0
## 4 Algeria             6875003          5391232              15.4            12.1
## 5 Andorra               54999            49535              71.1            64.0
## 6 Angola              6737721          3280340              19.9             9.67
```

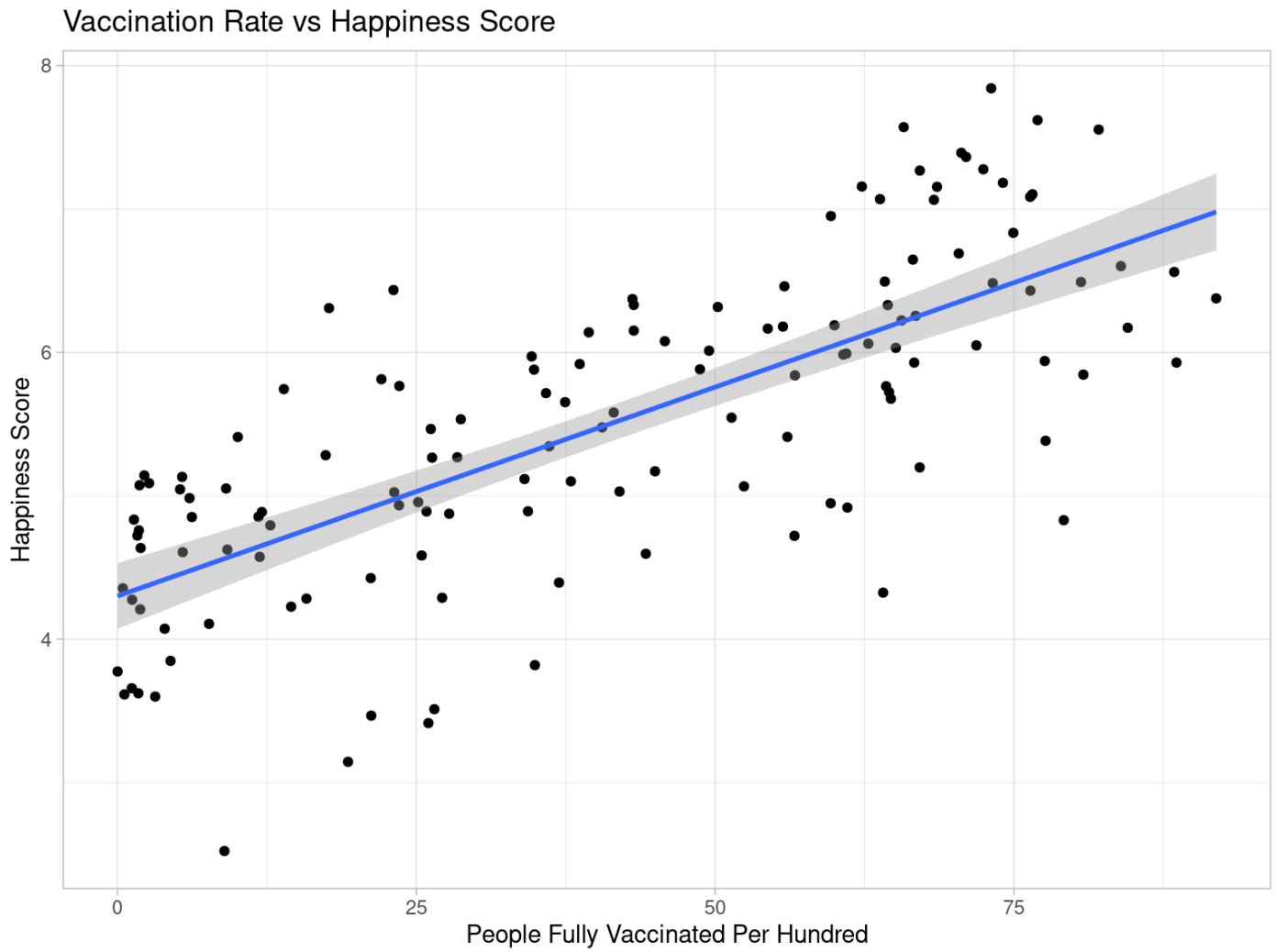
Get hapiness data.

Merge vaccination data with world happiness data.

```
## # A tibble: 6 × 3
## # Groups:   location [6]
##   location    people_fully_vaccinated_per_hundred Ladder.score
##   <chr>                                     <dbl>        <dbl>
## 1 Afghanistan                                8.95         2.52
## 2 Albania                                   34.0          5.12
## 3 Algeria                                   12.1          4.89
## 4 Argentina                                 66.6          5.93
## 5 Armenia                                   17.4          5.28
## 6 Australia                                 74.1          7.18
```

Plot scatter plot and see correlation between vaccination rate and happiness score for countries where both data are available:

```
## `geom_smooth()` using formula 'y ~ x'
```



This plot reveals the relationship between country vaccination rate and country happiness scores. We can see from the plot that they are positive related, and countries with higher vaccination rate generally have higher happiness score.

Next, we prepare data for world map vaccination rate:

Vaccination is not available in some countries 2020, so we need to set these countries's vaccination rate to zero.

Get location where vaccination available in 2020.

```
## # A tibble: 6 × 6
## # Groups:   location [6]
##   location date       people_vaccinated people_fully_vaccinated people_vaccinat…
##   <chr>    <chr>                  <dbl>                   <dbl>            <dbl>
## 1 Asia     2020-12-31           1056736                       1             0.02
## 2 Austria  2020-12-31              5185                      NA             0.06
## 3 Bahrain  2020-12-31             58643                      NA             3.35
```

```
## 4 Belarus  2020-12-28                0                      NA             0
## 5 Belgium  2020-12-31              929                      21             0.01
## 6 Bulgaria 2020-12-30             4608                      NA             0.07
## # … with 1 more variable: people_fully_vaccinated_per_hundred <dbl>
```

Get location where vaccination not available in 2020 by select first available date reported in each country and compare it with 2020-12-31:

```
## # A tibble: 6 × 6
## # Groups:   location [6]
##   location    date       people_vaccinated people_fully_vacc… people_vaccinated…
##   <chr>       <chr>                  <dbl>              <dbl>               <dbl>
## 1 Afghanistan 2021-02-22                 0                 NA                   0
## 2 Africa      2021-01-09                 0                 NA                   0
## 3 Albania     2021-01-10                 0                 NA                   0
## 4 Algeria     2021-01-29                 0                 NA                   0
## 5 Andorra     2021-01-25               576                 NA                   0.74
## 6 Angola      2021-03-01                 0                 NA                   0
## # … with 1 more variable: people_fully_vaccinated_per_hundred <dbl>
```

Create 2020 zero data for countries that don't have vaccination in 2020:

```
## # A tibble: 6 × 6
## # Groups:   location [6]
##   location    date       people_vaccinated people_fully_vacc… people_vaccinated…
##   <chr>       <chr>                  <dbl>              <dbl>               <dbl>
## 1 Afghanistan 2020-12-31                 0                  0                   0
## 2 Africa      2020-12-31                 0                  0                   0
## 3 Albania     2020-12-31                 0                  0                   0
## 4 Algeria     2020-12-31                 0                  0                   0
## 5 Andorra     2020-12-31                 0                  0                   0
## 6 Angola      2020-12-31                 0                  0                   0
## # … with 1 more variable: people_fully_vaccinated_per_hundred <dbl>
```

Combine the two datasets into world_vacci_data_2020:

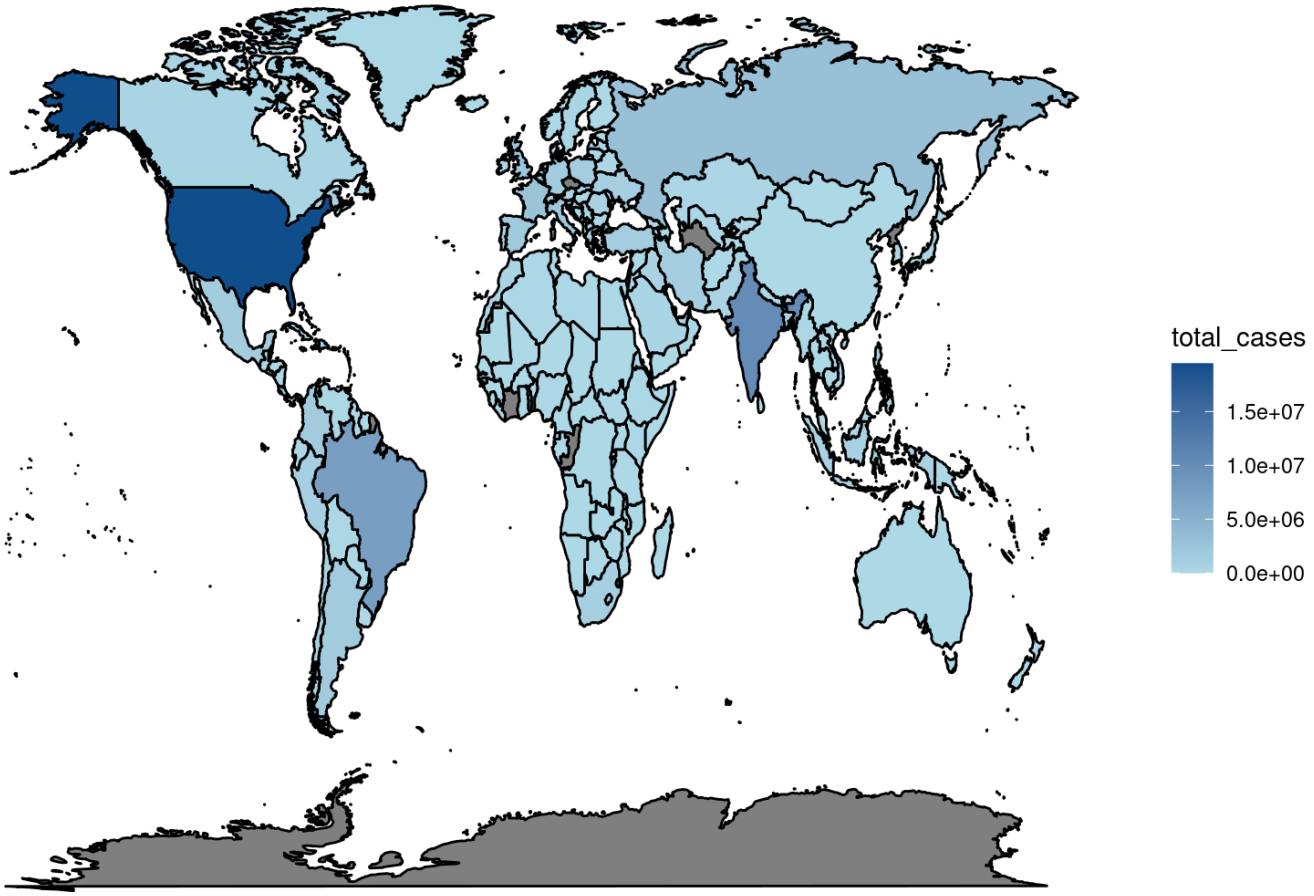
```
## # A tibble: 6 × 5
## # Groups:   location [6]
##   location people_vaccinated people_fully_vac… people_vaccinat… people_fully_va…
##   <chr>               <dbl>              <dbl>            <dbl>             <dbl>
## 1 Asia              1056736                  1             0.02                 0
## 2 Austria              5185                 NA             0.06                NA
## 3 Bahrain             58643                 NA             3.35                NA
## 4 Belarus                 0                 NA             0                   NA
## 5 Belgium               929                 21             0.01                 0
## 6 Bulgaria             4608                 NA             0.07                NA
```

Now, we have finished data cleaning and preparation for world maps vaccination rate.

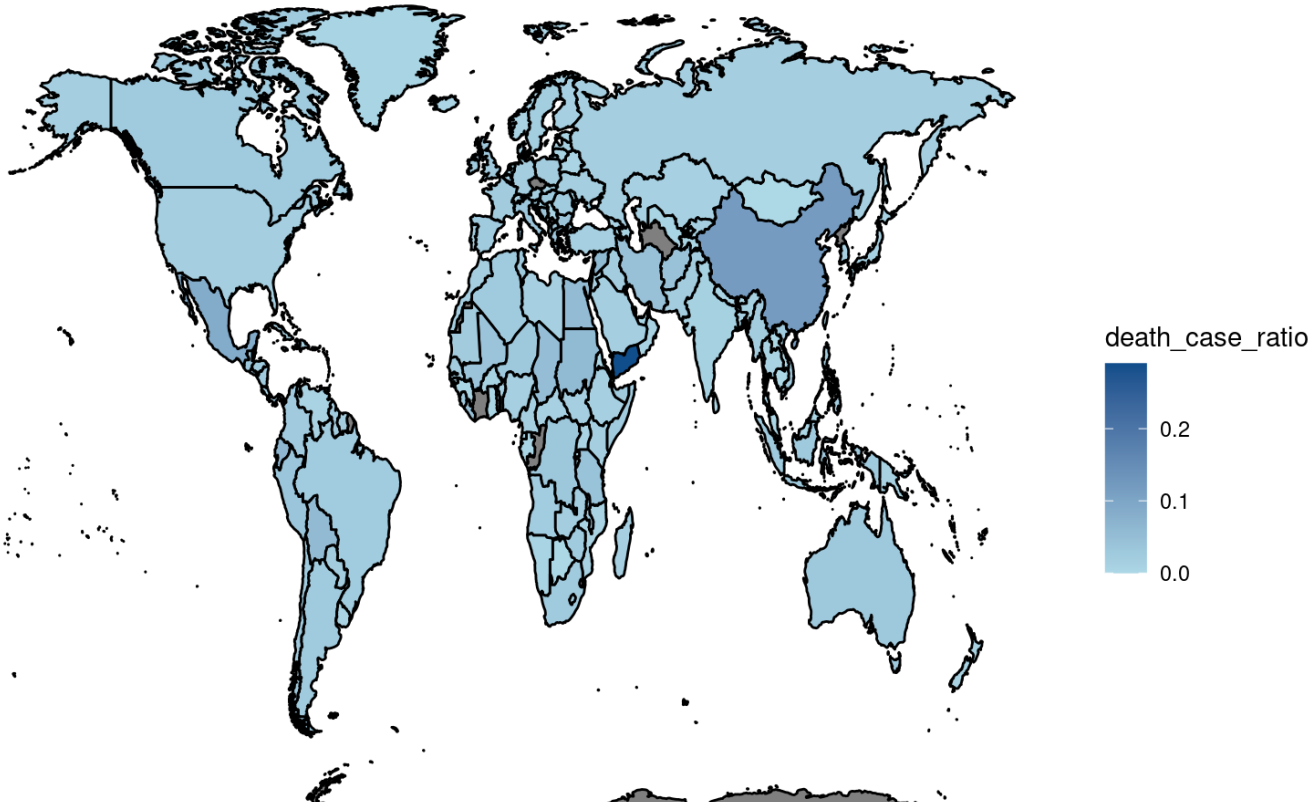
Investigation on World Happiness, Covid Deaths and Vaccination

World maps for covid-related data and analysis

total number of covid cases world map



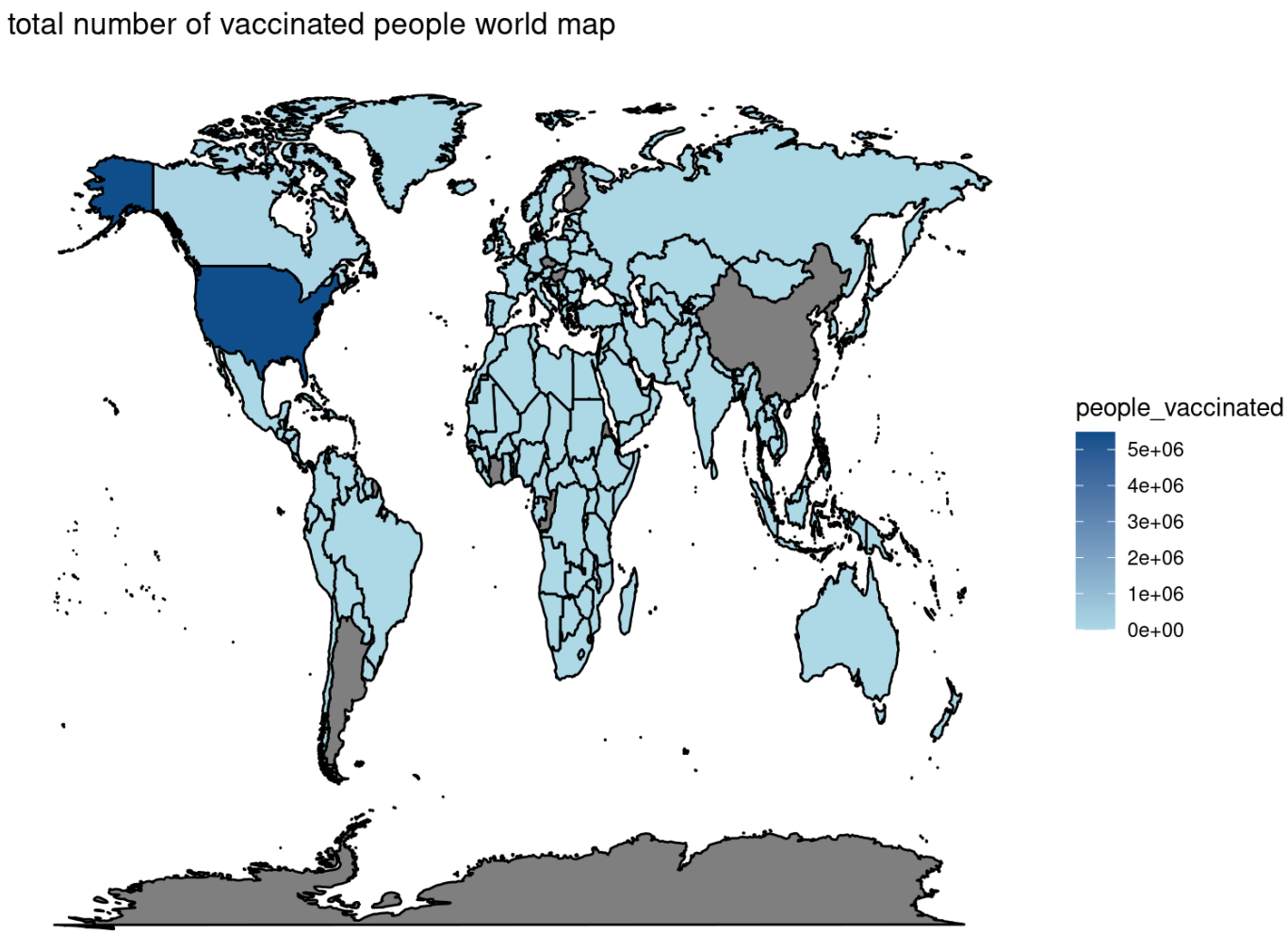
covid death ratio world map



From the **total number of covid cases world map**, we can observe that countries such as United States, India and Brazil have very high number of covid infection rate. On the other hand, countries like China have lower covid infection rate.

However, from the **covid death ratio world map**, we can observe that countries like United States, India and Brazil, which have very high number of covid infection rate, now have a relatively lower death ratio. This interesting phenomenon could be due to effective vaccination or unreported/misrecorded death rate.

World maps for vaccination-related data and analysis



From the **total number of vaccinated people world map**, we can

observe that countries such as United States have very high vaccination rate. This justifies my previous assumption about United State's low death rate ratio. On the other hand, vaccination information is not provided for countries like China in the raw data we find.

The low vaccination rate in countries like India, Brazil and Russia could be one of the key factors that contributed to their high covid-related death incidents.

ANOVA Analysis and Determine significant factors

To examine the whether the differences between the covid data, vaccination data and world happiness data are statistically significant, we wish to conduct ANOVA.

Now we would like to merge covid data, vaccination data and world happiness data together to continue our ANOVA analysis.

```
## New names:
## * `` -> ...1
```

```
## # A tibble: 149 × 32
##    `Country name` `Regional indica… `Ladder score` `Standard error… upperwhisker
##    <chr>          <chr>                      <dbl>            <dbl>        <dbl>
##  1 Finland        Western Europe              7.84            0.032         7.90
##  2 Denmark        Western Europe              7.62            0.035         7.69
##  3 Switzerland    Western Europe              7.57            0.036         7.64
##  4 Iceland        Western Europe              7.55            0.059         7.67
##  5 Netherlands    Western Europe              7.46            0.027         7.52
##  6 Norway         Western Europe              7.39            0.035         7.46
##  7 Sweden         Western Europe              7.36            0.036         7.43
##  8 Luxembourg     Western Europe              7.32            0.037         7.40
##  9 New Zealand    North America an…           7.28            0.04          7.36
## 10 Austria        Western Europe              7.27            0.036         7.34
## # … with 139 more rows, and 27 more variables: lowerwhisker <dbl>,
## #   Logged GDP per capita <dbl>, Social support <dbl>,
## #   Healthy life expectancy <dbl>, Freedom to make life choices <dbl>,
## #   Generosity <dbl>, Perceptions of corruption <dbl>,
## #   Ladder score in Dystopia <dbl>, Explained by: Log GDP per capita <dbl>,
## #   Explained by: Social support <dbl>,
## #   Explained by: Healthy life expectancy <dbl>, …
```

In this analysis we consider all the data and want to examine happiness variables, covid-related variables and vaccination-related variables, explaining the happiness ladder score.

```
## Analysis of Variance Table
##
## Response: Ladder score
##                                    Df  Sum Sq Mean Sq F value    Pr(>F)
## Generosity                          1  0.1860  0.1860  0.6467 0.4232497
## `Perceptions of corruption`         1 26.4690 26.4690 92.0241 1.015e-15 ***
## total_cases                         1  2.6000  2.6000  9.0392 0.0033645 **
## total_deaths                        1  3.8032  3.8032 13.2227 0.0004451 ***
## people_fully_vaccinated_per_hundred 1  0.6554  0.6554  2.2787 0.1344131
## total_cases_per_100k                1 20.9248 20.9248 72.7487 1.989e-13 ***
## total_deaths_per_100k               1  0.4888  0.4888  1.6994 0.1954529
## death_case_ratio                    1  1.7860  1.7860  6.2093 0.0144044 *
## people_vaccinated                   1  2.5987  2.5987  9.0347 0.0033722 **
## people_fully_vaccinated             1  0.6114  0.6114  2.1257 0.1480725
## people_vaccinated_per_hundred       1  0.1577  0.1577  0.5483 0.4607941
## `Freedom to make life choices`      1 19.0947 19.0947 66.3861 1.301e-12 ***
## `Logged GDP per capita`             1 17.9702 17.9702 62.4768 4.285e-12 ***
## `Social support`                    1  4.3030  4.3030 14.9603 0.0001988 ***
## `Healthy life expectancy`           1  0.9482  0.9482  3.2964 0.0725201 .
## Residuals                          97 27.9002  0.2876
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For this data, we will take p-value = 0.05 as a bench mark and we will run our analysis on significant factors based on this p-value.

According to the Analysis of Variance Table, we can conclude the following:

In conclusion, we can conclude that perceptions of corruption, total covid cases, total covid deaths, total covid cases per 100k, covid death case ratio, number of vaccinated people, freedom to make life choices, logged GDP per capita and social support are significant predictor of happiness of a country's citizens.