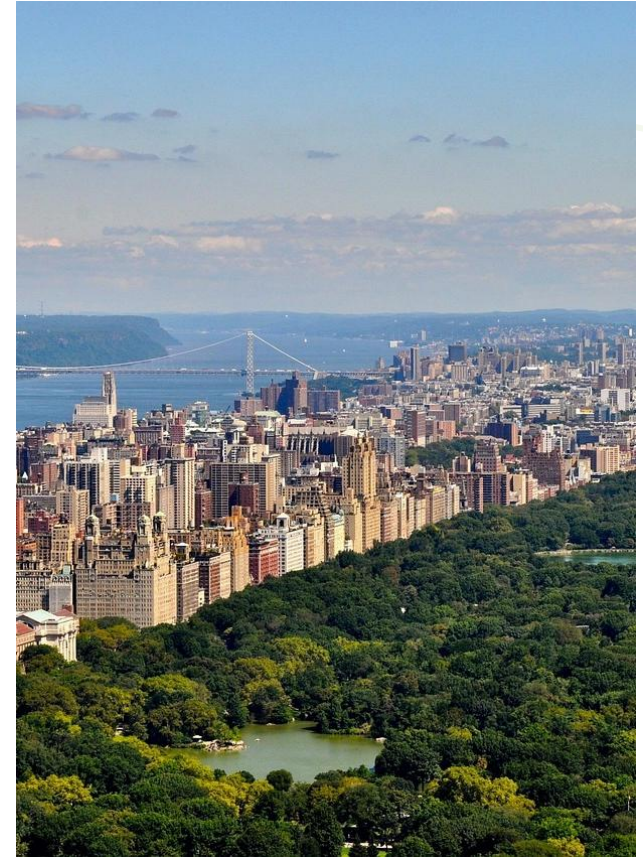


# Is there a significant relationship between the percentage of forest area and happiness?

COMP 2501 Project  
Minseo Yoon (3035759117)





# 01

## Project Background

Purpose & Challenges





# Just Preference?



**Me**

*Love **City** the most!  
Prefer to live in **Central** of  
Hong Kong, Seoul, Singapore...*



**My Sister**

*Love **Nature** the most!  
Prefer to live in **Rural Area** of  
Switzerland, Malaysia, Australia ...*



# Why is this question important?

As the world has developed, **interest in quality of life has drastically increased** over the past few decades.

What makes people to be happier?

This question is important and potentially insightful because it is **closely related to our daily lives.**

*Even serve as a **small guideline** for **selecting a country to live in!***





# Related Studies

Navarrete-Hernandez, P., Kiarostami, N., Yang, D., & Ozcakir, A. (2024). **Green Enough? A dose-response curve of the impact of street greenery levels and types on perceived happiness.** Landscape and Urban Planning, 251, 105130. <https://doi.org/10.1016/j.landurbplan.2024.105130>

Patino, J. E., Martinez, L., Valencia, I., & Duque, J. C. (2023). **Happiness, life satisfaction, and the greenness of urban surroundings.** Landscape and Urban Planning, 237, 104811. <https://doi.org/10.1016/j.landurbplan.2023.104811>

Behera, D. K., Rahut, D. B., Padmaja, M., & Dash, A. K. (2024). **Socioeconomic Determinants of Happiness: Empirical Evidence from Developed and Developing Countries.** Journal of Behavioral and Experimental Economics, 109, 102187. <https://doi.org/10.1016/j.socrec.2024.102187>

*There are many related studies,  
but **NO** studies with standard data from public institutions!*





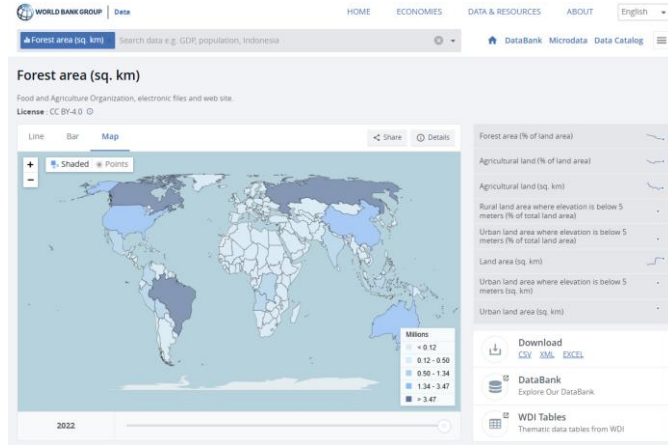
# 02

## Methods & Results

The research outcome and correlations



# Key Sources of Information



## WorldBank

*International Financial Institution*  
Got 'Land/Forest Area by country' &  
'GDP per capita' datasets



ABOUT FAQ DATA

## Data Sharing

The World Happiness Report is powered by data from the Gallup World Poll.

Since creating the World Poll in 2005, Gallup has conducted studies in more than 160 countries and territories that are home to more than 96% of the world's adult population.

The World Poll survey includes more than 100 global questions as well as region-specific items. Gallup asks residents from Australia to Zimbabwe the same questions, every time, in the same way. This makes it possible to trend data from year to year and make direct country comparisons.

We also utilise data from many other international, national, regional, and other datasets which are generally available for public use.

**Data available for free**

## SDSN

*Sustainable Development Solutions Network*  
launched by **United Nations (UN)**  
Got 'Happiness Index' dataset



# Collecting Data

forest\_area.csv  
gdp\_per\_capita.csv  
happiness\_data.csv  
land\_area.csv

파일

홈

삽입

레이아웃

데이터

검토

보기

자동화

개발

도구

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

*In this project, all datasets were available to download by .csv format.  
For stable connection, placed datasets in local directory.*



# Environmental Setup

```
48 ▾ ### 1. Environmental Setup
49
50 ▾ ```{r}
51
52 #Load the packages.
53 library(dplyr)
54 library(ggplot2)
55
56 ▴ ```
```

다음의 패키지를 부착합니다: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

경고: 패키지 'ggplot2'는 R 버전 4.4.3에서 작성되었습니다

*Before begin the project, setup the programming environment.*



# Load & Check Datasets

```
59 - ### 2. Load CSV Files from Local Folder
60 -
61 - ```{r}
62 - |
63 - | # Read the CSV files into R.
64 - | # 'skip = 4' removes metadata rows from World Bank files.
65 - | land_area <- read.csv("land_area.csv", skip = 4)
66 - | forest_area <- read.csv("forest_area.csv", skip = 4)
67 - | happiness <- read.csv("happiness_data.csv", encoding = "latin1")
68 - |
69 - | ...
70 - |
71 - |
72 - | ### 3. Check Column Names in Each Dataset
73 - |
74 - | ```{r}
75 - |
76 - | # Display the column names to confirm the correct year and country name columns are present
77 - | colnames(land_area)
78 - | colnames(forest_area)
79 - | colnames(happiness)
80 - |
81 - | ...
```

[1]	"Country.Name"	"Country.Code"	"Indicator.Name"	"Indicator.Code"	"x1960"	"x1961"	"x1962"	"x1963"	"x1964"	"x1965"
[11]	"x1966"	"x1967"	"x1968"	"x1969"	"x1970"	"x1971"	"x1972"	"x1973"	"x1974"	"x1975"
[21]	"x1976"	"x1977"	"x1978"	"x1979"	"x1980"	"x1981"	"x1982"	"x1983"	"x1984"	"x1985"
[31]	"x1986"	"x1987"	"x1988"	"x1989"	"x1990"	"x1991"	"x1992"	"x1993"	"x1994"	"x1995"
[41]	"x1996"	"x1997"	"x1998"	"x1999"	"x2000"	"x2001"	"x2002"	"x2003"	"x2004"	"x2005"
[51]	"x2006"	"x2007"	"x2008"	"x2009"	"x2010"	"x2011"	"x2012"	"x2013"	"x2014"	"x2015"
[61]	"x2016"	"x2017"	"x2018"	"x2019"	"x2020"	"x2021"	"x2022"	"x2023"	"x2024"	"x"
[1]	"Country.Name"	"Country.Code"	"Indicator.Name"	"Indicator.Code"	"x1960"	"x1961"	"x1962"	"x1963"	"x1964"	"x1965"
[11]	"x1966"	"x1967"	"x1968"	"x1969"	"x1970"	"x1971"	"x1972"	"x1973"	"x1974"	"x1975"
[21]	"x1976"	"x1977"	"x1978"	"x1979"	"x1980"	"x1981"	"x1982"	"x1983"	"x1984"	"x1985"
[31]	"x1986"	"x1987"	"x1988"	"x1989"	"x1990"	"x1991"	"x1992"	"x1993"	"x1994"	"x1995"
[41]	"x1996"	"x1997"	"x1998"	"x1999"	"x2000"	"x2001"	"x2002"	"x2003"	"x2004"	"x2005"
[51]	"x2006"	"x2007"	"x2008"	"x2009"	"x2010"	"x2011"	"x2012"	"x2013"	"x2014"	"x2015"
[61]	"x2016"	"x2017"	"x2018"	"x2019"	"x2020"	"x2021"	"x2022"	"x2023"	"x2024"	"x"
[1]	"Year"		"Rank"		"Country.name"					
[4]	"Ladder.score"		"upperwhisker"		"lowerwhisker"					
[7]	"Explained.by..Log.GDP.per.capita"		"Explained.by..Social.support"		"Explained.by..Healthy.life.expectancy"					
[10]	"Explained.by..Freedom.to.make.life.choices"		"Explained.by..Generosity"		"Explained.by..Perceptions.of.corruption"					
[13]	"Dystopia...residual"		"x"		"x.1"					
[16]	"x.2"		"x.3"		"x.4"					
[19]	"x.5"		"x.6"		"x.7"					
[22]	"x.8"		"x.9"		"x.10"					
[25]	"x.11"		"x.12"		"x.13"					
[28]	"x.14"									

*Checked the datasets whether data is well loaded.*



# Select Data

```
84 ▾ ### 4. Select Relevant Columns from Each Dataset
85
86 ▾ ```{r}
87
88 # Extract only the country name and 2022 values from land and forest data
89 land_2022 <- land_area %>%
90   select(Country = `Country.Name`, Land2022 = X2022)
91
92 forest_2022 <- forest_area %>%
93   select(Country = `Country.Name`, Forest2022 = X2022)
94
95
96 # Filter happiness index data for the year 2022 and select key columns
97 happiness_2022 <- happiness %>%
98   filter(Year == 2022) %>%
99   select(Country = `Country.name`, HappinessScore = `Ladder.score`)
100
101 ▴ ```
```

*Since the latest version of land/forest area datasets are 2022, this project incorporated Happiness Index in 2022 for accuracy and consistency.*



# Merge Data

```
104 - ### 5. Merge Datasets & Calculate Forest Area (% of Land Area)
105
106 - ...{r}
107
108 # Merge land and forest area data, and compute forest coverage as a % of total land
109 area_data <- left_join(land_2022, forest_2022, by = "Country") %>%
110   mutate(ForestPercent = Forest2022 / Land2022 * 100)
111
112 # Merge with the happiness index dataset
113 merged_data <- inner_join(area_data, happiness_2022, by = "Country")
114
115 # Remove entries that represent regions or aggregates instead of individual countries
116 merged_data <- merged_data %>%
117   filter(!grepl("world|Europe|Africa|Asia|America|OECD|Arab", country))
118
119 #Print merged dataset
120 print(merged_data)
121
122 - ...
```

Description: df [119 x 5]

Country <chr>	Land2022 <dbl>	Forest2022 <dbl>	ForestPercent <dbl>	HappinessScore <dbl>
Afghanistan	652230.00	12084.400	1.8527820	1.859
Albania	27400.00	7889.000	28.7919708	5.277
Argentina	2736690.00	283552.222	10.3611378	6.024
Armenia	28470.00	3280.500	11.5226554	5.342
Australia	7692020.00	1340051.000	17.4213146	7.095
Austria	82520.00	38954.500	47.2061318	7.097
Belgium	30494.00	6893.000	22.6044468	6.859
Benin	112760.00	30351.500	26.9169032	4.374
Burkina Faso	273600.00	61164.000	22.3552632	4.638
Bangladesh	130170.00	18834.000	14.4687716	4.282

1-10 of 119 rows

Previous 1 2 3 4 5 6 ... 12 Next

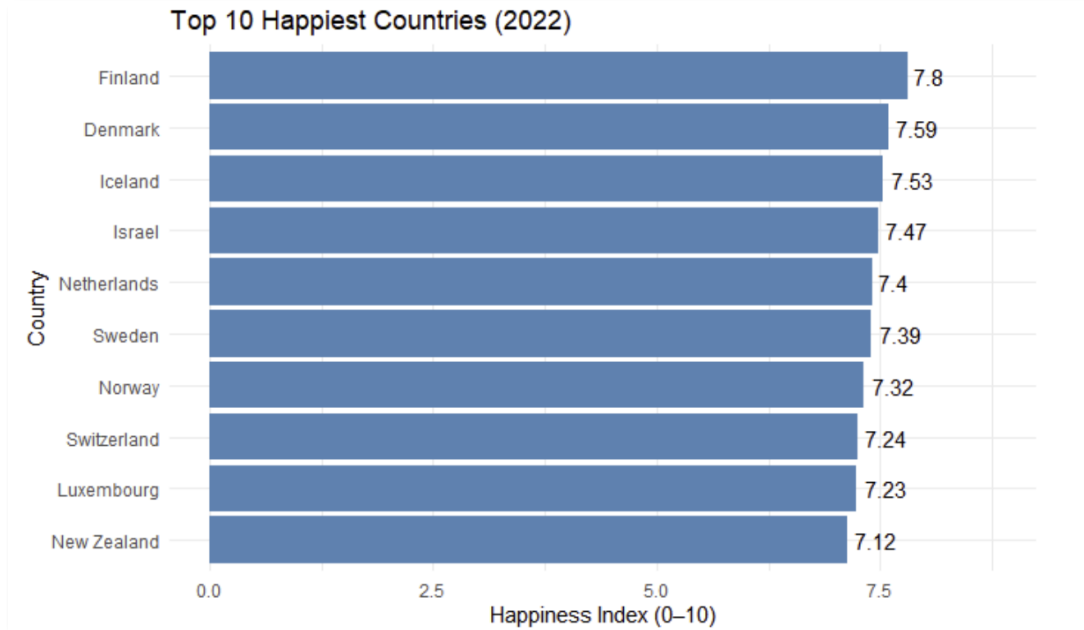
*Merge datasets into one, with filtered columns in previous steps.  
Check the merged dataset by printing it.*



# [Visualization] Top 10 Countries

```
# TOP 10 Countries - Bar Chart
top10 <- happiness_ranked %>%
  arrange(desc(HappinessScore)) %>%
  slice(1:10)

ggplot(top10, aes(x = fct_reorder(Country, HappinessScore), y = HappinessScore)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  geom_text(aes(label = round(HappinessScore, 2)), hjust = -0.2, size = 4) +
  coord_flip() +
  labs(title = "Top 10 Happiest Countries (2022)",
       x = "Country",
       y = "Happiness Index (0-10)") +
  theme_minimal() +
  ylim(0, max(top10$HappinessScore) + 1)
```

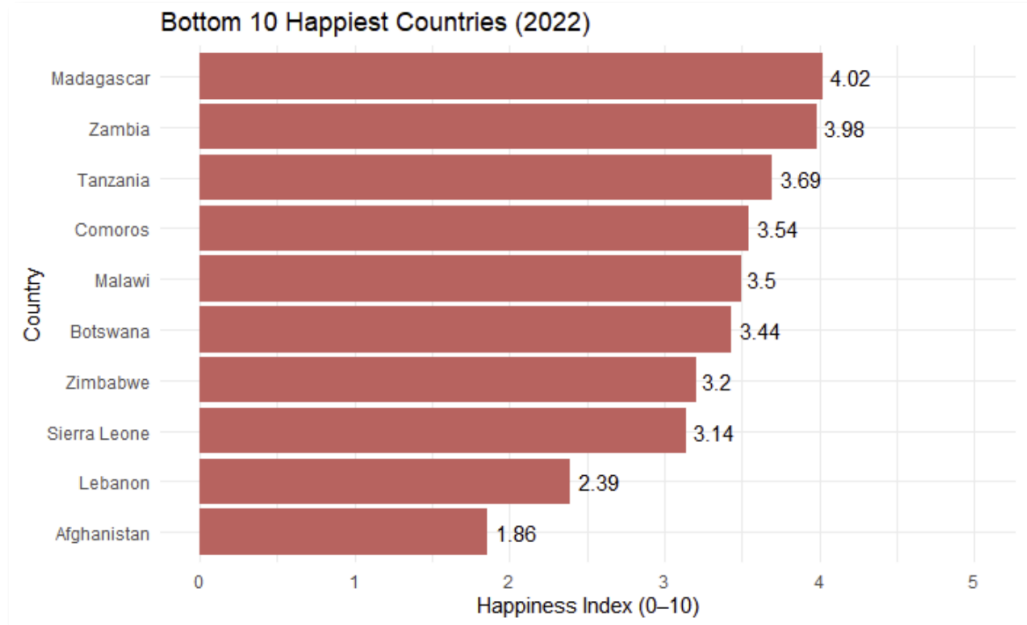




# [Visualization] Bottom 10 Countries

```
# BOTTOM 10 Countries - bar Chart
bottom10 <- happiness_ranked %>%
  arrange(HappinessScore) %>%
  slice(1:10)

ggplot(bottom10, aes(x = fct_reorder(Country, HappinessScore), y = HappinessScore)) +
  geom_bar(stat = "identity", fill = "indianred") +
  geom_text(aes(label = round(HappinessScore, 2)), hjust = -0.2, size = 4) +
  coord_flip() +
  labs(title = "Bottom 10 Happiest Countries (2022)",
       x = "Country",
       y = "Happiness Index (0-10)") +
  theme_minimal() +
  ylim(0, max(bottom10$HappinessScore) + 1)
```





# Correlation Test

```
126 ▾ ### 6. Correlation Test (Forest Area & Happiness Index)
127
128 ▾ {r}
129
130 # Perform a statistical test to check the correlation between forest coverage and happiness
131 cor.test(merged_data$ForestPercent, merged_data$HappinessScore)
132
133 ▲
```

Pearson's product-moment correlation

data: merged\_data\$ForestPercent and merged\_data\$HappinessScore  
t = 1.8263, df = 115, p-value = 0.07041  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.01407985 0.33908130  
sample estimates:  
cor  
0.1678822

*Calculate the correlation value to check the relationship in numeric.*



# Visualization the Result

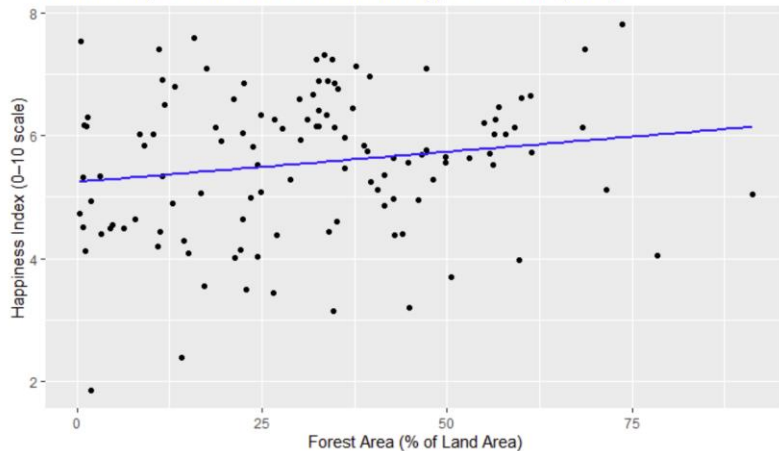
```
136 - ## 7. Visualization the result (Scatterplot)
```

```
137 - {r}
138 - # Scatterplot with a linear regression line to illustrate the relationship
139 - ggplot(merged_data, aes(x = ForestPercent, y = HappinessScore)) +
140 -   geom_point() +
141 -   geom_smooth(method = "lm", se = FALSE, color = "blue") +
142 -   labs(title = "Relationship Between Forest Area and Happiness Score (2022)",
143 -        x = "Forest Area (% of Land Area)",
144 -        y = "Happiness Index (0-10 scale)")
145 -
146 -
```

[-38;5;232m'geom\_smooth()' using formula = 'y ~ x'[-39m

경고: [-38;5;232mRemoved 2 rows containing non-finite outside the scale range ('stat\_smooth()').[-39m  
경고: [-38;5;232mRemoved 2 rows containing missing values or values outside the scale range ('geom\_point()').[-39m

Relationship Between Forest Area and Happiness Score (2022)



Correlated, but **NOT**  
**strongly correlated**  
**than I thought.**

*Maybe we can check  
the other factor  
together.*

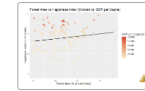
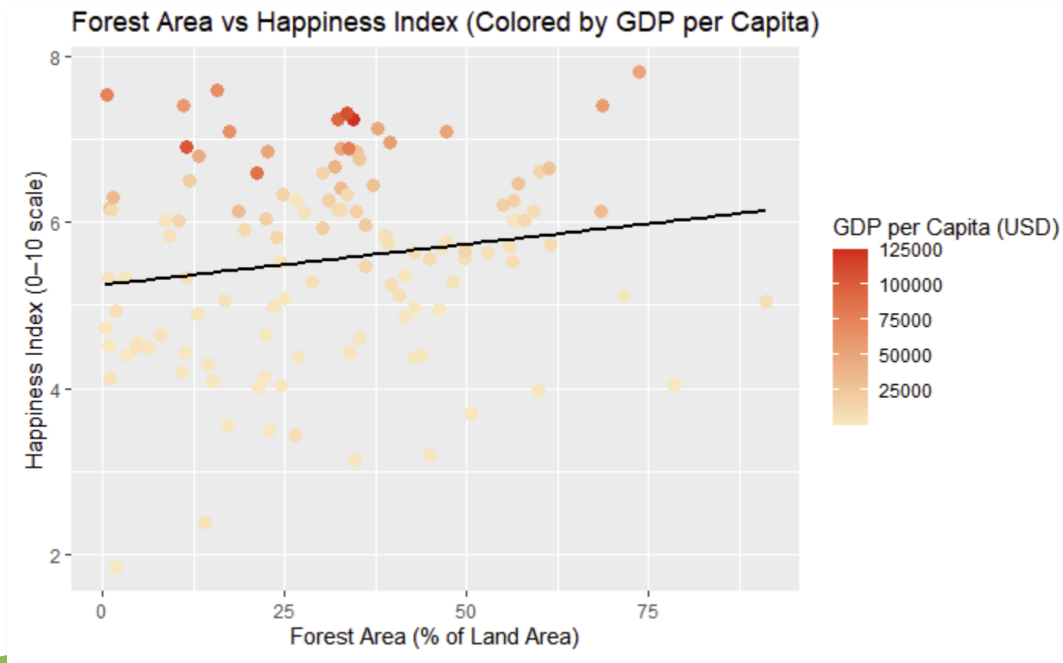


# [Further Analysis] GDP per Capita

```
150 ▾ ### Further Analysis: Consider economic condition
151 According to the result above, there are some exceptions.
152 To clarify this exceptions, I did further analysis to examine an additional hypothesis.
153 My new hypothesis is "National economic environment will also hugely affect people's happiness."
154
155 I used GDP per capita, to improve my analysis.
156
157 ▾ ```{r}
158
159 # 1. Load the file
160 gdp <- read.csv("gdp_per_capita.csv", skip = 4)
161
162 # 2. Select only country and 2022 GDP
163 gdp_2022 <- gdp %>%
164   select(Country = `Country.Name`, GDPperCapita = x2022)
165
166 # 3. Merge with existing dataset
167 merged_data <- left_join(merged_data, gdp_2022, by = "Country")
168
169 ggplot(merged_data, aes(x = ForestPercent, y = HappinessScore, color = GDPperCapita)) +
170   geom_point(size = 3) +
171   geom_smooth(method = "lm", se = FALSE, color = "black") +
172   scale_color_gradient(low = "wheat1", high = "red2") +
173   labs(title = "Forest Area vs Happiness Index (Colored by GDP per Capita)",
174        x = "Forest Area (% of Land Area)",
175        y = "Happiness Index (0-10 scale)",
176        color = "GDP per Capita (USD)")
177
178
179 # 4. Calculate correlation between GDP Per Capita Vs. Happiness Index
180 cor.test(merged_data$GDPperCapita, merged_data$HappinessScore)
181
182 ▾ ```
```



# [Further Analysis] GDP per Capita



PEARSON'S PRODUCT-MOMENT CORRELATION

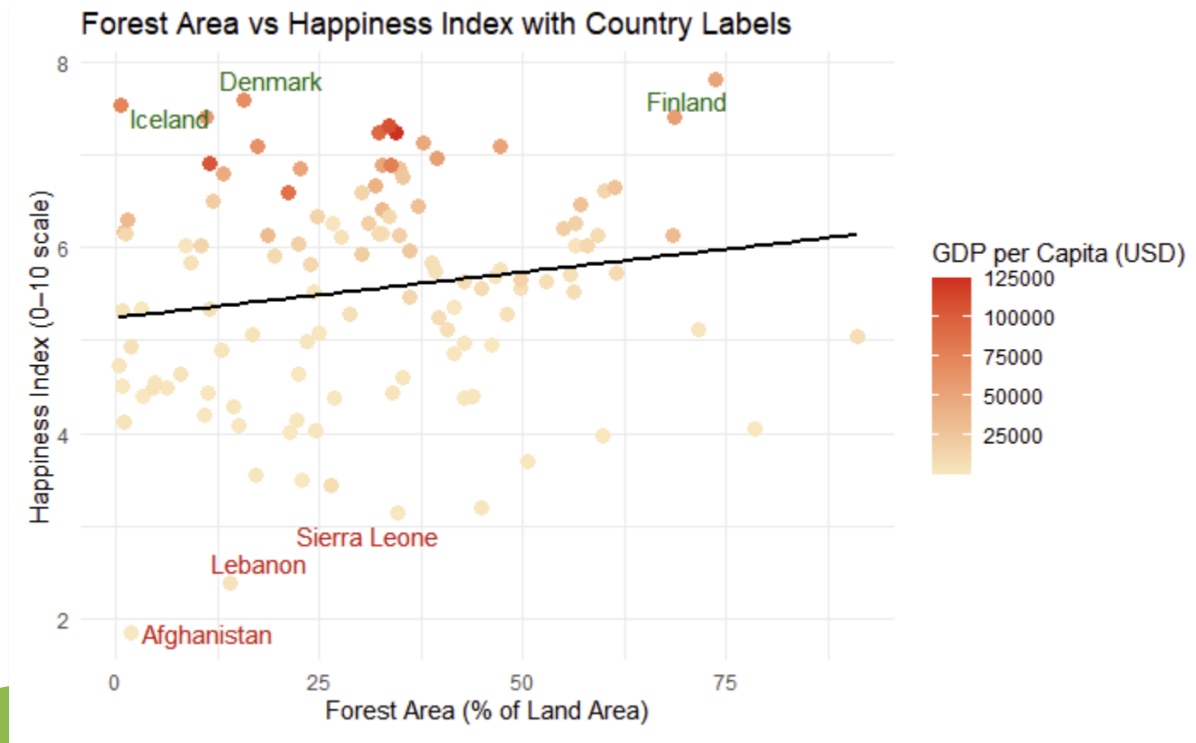
data: merged\_data\$GDPperCapita and merged\_data\$HappinessScore  
t = 10.564, df = 117, p-value < 2.2e-16  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
0.5933159 0.7805295  
sample estimates:  
cor  
0.6986948

Pearson's product-moment correlation

```
data: merged_data$GDPperCapita and merged_data$HappinessScore
t = 10.564, df = 117, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.5933159 0.7805295
sample estimates:
cor
0.6986948
```



# [Further Analysis] GDP per Capita



Label **Top 3 & Bottom 3 Countries**, to visualize the data effectively.



# 03

## Conclusions

Answers &  
Future Research Suggestions





# Conclusion

Is there a correlation between forest area percentage and happiness index?

**YES.** But it does **not mean** that it is **absolute factor** to determine happiness.

**Economic Situation of Countries** also largely contribute to happiness index.

There are so **many factors that we should consider** to measure people's happiness.

## WHY?

### Regression Line Slope

- Forest Area (% of Land) Vs. Happiness Index: 0.1678822
- GDP per Capita Vs. Happiness Index: **0.6986948** → *much higher*

### Additional Considerations & Assumptions:

- Better national economic situation → Higher probability to enjoy **better living infrastructure & welfare**
- Forest Area doesn't always mean green spaces in living regions



# Limitations & Future Improvements

01

## Definition of Forest Area

Forest Area includes area of mountain as well, not only limited to region where people can access.

02

## Lack of Specific Data

There was no data which takes green area only, where people can access.

03

## Personal Opinion

Happiness Index is based on people's survey. Everyone has different base lines and perceptions.

04

## Complex Factors

Happiness can be determine with various factors. Hard to digitize the analysis with few factors only.



# Future Suggestions



## Collect Green Area Data with Specific Criteria

Green infrastructure in living area would give more specific result. Thus, collect data from Google Earth or Maps by image analysis to measure green area (e.g. park, hiking course...) with specific criterias where closely related to people's living environment.



## Take Scientific Data to Measure Happiness

Although reliable organization has conducted happiness measurement, it is based on surveys. Therefore, take scientific data (e.g. medical data related to happiness) which can be measure without personal opinion.



# References

Caring and sharing: Global analysis of happiness and kindness | The World Happiness Report. (2025). <https://worldhappiness.report/ed/2025/caring-and-sharing-global-analysis-of-happiness-and-kindness/#happiness-ranking>

Data sharing | The World Happiness Report. (2025). <https://worldhappiness.report/data-sharing/>

World Bank Open Data. (2023a). World Bank Open Data. <https://data.worldbank.org/indicator/AG.LND.TOTL.K2>

World Bank Open Data. (2023b). World Bank Open Data. <https://data.worldbank.org/indicator/AG.LND.FRST.K2?view=map>

World Bank Open Data. (2024). World Bank Open Data. <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

***All datasets and rmd file are uploaded on OneDrive:***

[https://connecthkuhk-my.sharepoint.com/:f:/g/personal/mincow\\_connect\\_hku\\_hk/ErigrNoGCgVDu\\_YyJNn5hFUBT97tqJ1fgiCHNq9wtoouZQ?e=ogWZSZ](https://connecthkuhk-my.sharepoint.com/:f:/g/personal/mincow_connect_hku_hk/ErigrNoGCgVDu_YyJNn5hFUBT97tqJ1fgiCHNq9wtoouZQ?e=ogWZSZ)





**Thank You!**