

KEYWORDS – Happiness Index, Typst, Quarto

I. INTRODUCTION

i. Objective

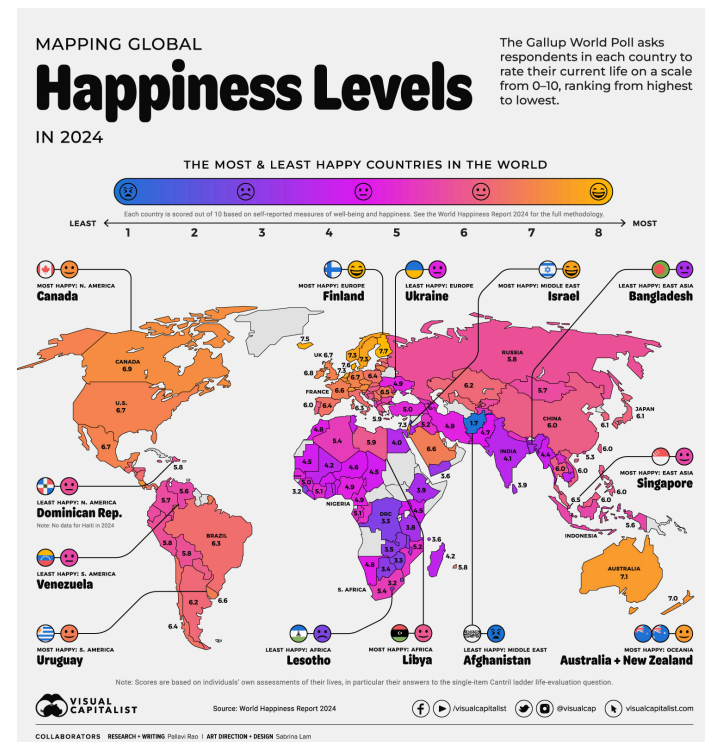
To critically assess and enhance a static visualization of global happiness levels, identifying its strengths and weaknesses, and proposing improvements for a more interactive and insightful presentation.

ii. Visualcapitalist’s Global Happiness Levels Report

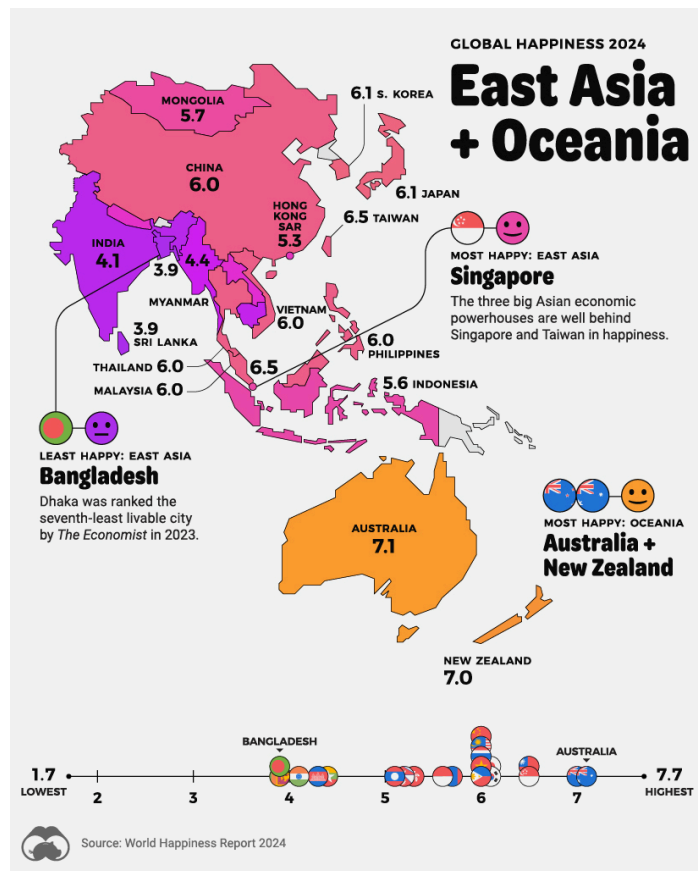
Happiness levels across the globe are a complex interplay of economic, social, psychological, and environmental factors that contribute to an individual’s sense of well-being. To quantitatively assess and compare these levels internationally, the **World Happiness Report** (WHR) serves as a vital tool. Initiated by the **United Nations Sustainable Development Solutions Network**, the WHR annually ranks countries based on the self-reported happiness of their citizens.

The report utilizes data primarily sourced from the **Gallup World Poll** (GWP), a sub-dataset of the World Happiness Report, which asks respondents to evaluate their current lives on a scale from 0 to 10, known as the **Cantril Ladder**. This evaluation is supported by key variables that are believed to contribute to life satisfaction: GDP per capita, social support, healthy life expectancy, personal freedom, generosity, and perceptions of corruption.

The following Figure 1 from Visualcapitalist presents the global happiness levels in 2024, as reported in the **World Happiness Report**. The static world map provides a visual representation of the happiness scores across different countries, with the color gradient indicating the relative happiness levels. The visualization aims to highlight the disparities in happiness levels worldwide and the factors that contribute to these variations. However, through our critical analysis, we identified several strengths and weaknesses in the original visualization that we aim to improve as detailed below.



(a) Global Overview World map of Happiness Levels in 2024



(b) Map of Happiness Levels of East Asia + Oceania in 2024

Figure 1: Visualcapitalist Global Happiness Levels Report

iii. Weakness

The original visualization has several weaknesses that limit its effectiveness in conveying the information to the audience. These weaknesses include:

- Lack of Interactivity:** The static nature of the world map limits the audience’s ability to interact with the data and explore specific countries or regions in more detail. An interactive visualization would allow users to hover over countries to see detailed information and compare happiness scores across different years.
- Limited Context:** While the color gradient provides a general overview of happiness levels, the visualization lacks additional context or explanations of the factors contributing to these scores. Including annotations or tool-tips that explain the key variables influencing happiness levels would enhance the audience’s understanding of the data.

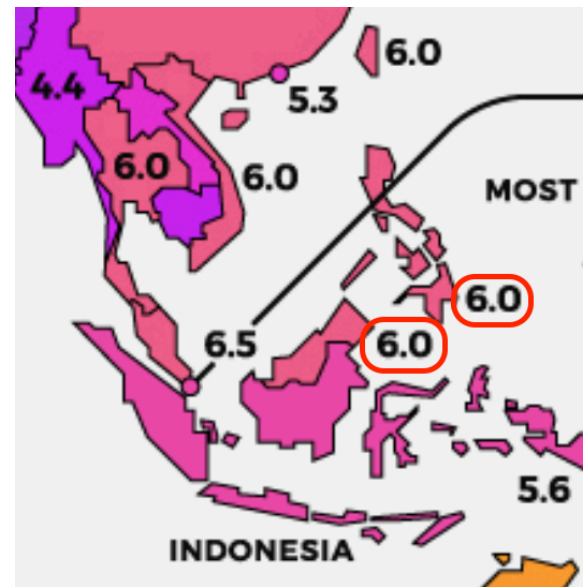


Figure 2: Inconsistent Labeling example

3. Inconsistent Labeling:

- The labeling of happiness scores and name of countries are inconsistent in placement or are not displayed, making it challenging for the audience to quickly identify and compare scores across countries.
 - A standardized labeling format would improve the readability and clarity of the visualization. In addition to this, there are countries’ happiness scores that are not labeled due to the lack of space for example Cambodia and Laos, which may lead to confusion for the audience.
 - Furthermore, the lack of pointers and inconsistent formatting severely affect readability for countries such Philippines and Indonesia have confusing unclear placements as to which country the label belongs to.
4. **Limited Historical Comparison:** The visualization only presents the happiness scores for the year 2024, without providing any historical comparison or trend analysis. Including data from previous years or a time slider feature would allow users to track changes in happiness levels over time and identify emerging patterns or trends.
5. **Limited Data Insights:** The visualization focuses solely on the happiness scores of countries without providing additional insights or analysis of the data. Including annotations, data summaries, or comparisons with other relevant metrics would enrich the audience’s understanding of the factors influencing happiness levels.
6. **Color Blindness:** The color gradient used in the visualization may not be accessible to color-blind users, as certain color combinations may be difficult to distinguish. Using color schemes that are accessible to users with color vision deficiencies would improve the inclusivity of the visualization.



Figure 3: Linear scale of Happiness score of each country by region

7. **Linear Scale:** The linear scale at the bottom of the region map is poorly visualized and does not provide a distinct separation and vivid depiction between the countries. This further worsened by the clustering and overlapping of icons, makes it difficult for the audience to interpret and compare the happiness scores of each country within the region.
8. **Incorrect Date for the Visualization:** The visualization is titled “A Map of Global Happiness by Country in 2024” but the data is from 2023. This inconsistency may lead to confusion among the audience and affect the credibility of the visualization.

II. METHODS

i. Data Sources

World Happiness Report (WHR) 2021-2023: Contains happiness scores and key variables for countries worldwide. Gallup World Poll (GWP): Provides self-reported happiness evaluations on a scale from 0 to 10.

ii. Data Understanding and Cleaning

iii. Load Data

```
df <- read_excel("data/DataForTable2.1.xls")
```

iv. List and Identify Missing Countries

```
# how many countries are missing in the dataset use Country name
missing_countries <- setdiff(countries, unique(df$Country name))
missing_countries
```

```
[1] "Andorra"
[3] "Bahamas"
[5] "Brunei"
[7] "Congo"
[9] "East Timor"
[11] "Eritrea"
[13] "Grenada"
[15] "Kiribati"
[17] "Marshall Islands"
[19] "Monaco"
[21] "North Korea"
[23] "Papua New Guinea"
[25] "Saint Lucia"
[27] "Samoa"
[29] "Sao Tome and Principe"
[31] "Solomon Islands"
[33] "Tuvalu"
[35] "Vatican City"

"Antigua and Barbuda"
"Barbados"
"Cabo Verde"
"Dominica"
"Equatorial Guinea"
"Fiji"
"Guinea-Bissau"
"Liechtenstein"
"Micronesia"
"Nauru"
"Palau"
"Saint Kitts and Nevis"
"Saint Vincent and the Grenadines"
"San Marino"
"Seychelles"
"Tonga"
"Vanuatu"
```

We can see that there are 34 countries missing from the dataset.

v. Data Cleaning and Transformation

As we will be only using data from 2021 - 2023, we will filter out the data from other years. We will also check for missing data and fill in the missing values with the mean of the available data from each country.

1) Filter data for years 2021 - 2023:

```
df <- df %>%
  filter(year %in% c(2021, 2022, 2023))
# head(df)
```

2) Renaming headers and keeping only the necessary columns:

```
# Rename the columns
colnames(df) <- c("country", "year", "score", colnames(df)[4:ncol(df)])

# Keep only the specified columns
df_filtered <- df[, c("country", "year", "score")]
# head(df_filtered)
```

3) Include Missing Countries with NA Scores:

```
# List of all countries in your data
all_countries <- unique(df_filtered$country)

# List of years to ensure each country has entries for each year
years <- c(2021, 2022, 2023)

# Create a data frame with all combinations of countries and years
complete_df <- expand_grid(country = all_countries, year = years)

# Add country codes to the complete data frame
complete_df$country_code <- countrycode(complete_df$country, origin = "country.name",
destination = "iso3c")
```

Warning: Some values were not matched unambiguously: Kosovo

```
# Merge the complete data frame with the original data
df_combined <- left_join(complete_df, df_filtered, by = c("country", "year"))
```



```
# Calculate the mean score for each country
country_means <- df_combined %>%
  group_by(country) %>%
  summarize(mean_score = mean(score, na.rm = TRUE))

# Merge the country means back into the combined data
df_combined <- left_join(df_combined, country_means, by = "country")

# Fill any remaining NA values with the mean score of the respective country
df_combined <- df_combined %>%
  mutate(score = ifelse(is.na(score), mean_score, score))

# Check for any remaining missing values
missing_scores <- df_combined %>% filter(is.na(score))
if (nrow(missing_scores) > 0) {
  warning("There are still countries with missing scores.")
}

# Inspect the data for missing scores
# print(missing_scores)
```

```
region <- read_csv("data/world-regions-according-to-the-world-bank.csv")
```

```
Rows: 217 Columns: 4
— Column specification —————
Delimiter: ","
chr (3): Entity, Code, World Region according to the World Bank
dbl (1): Year
```

```
i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# head(region)

df_combined_region <- left_join(df_combined, region, by = c("country_code" = "Code"))
# head(df_combined_region)
# find empty rows in "World Region according to the World Bank"

missing_region <- df_combined_region %>% filter(is.na(`World Region according to the World Bank`))
# print(missing_region)

df_combined_region$`World Region according to the World Bank`[which(df_combined_region$country == "Taiwan Province of China")] <- "East Asia and Pacific"

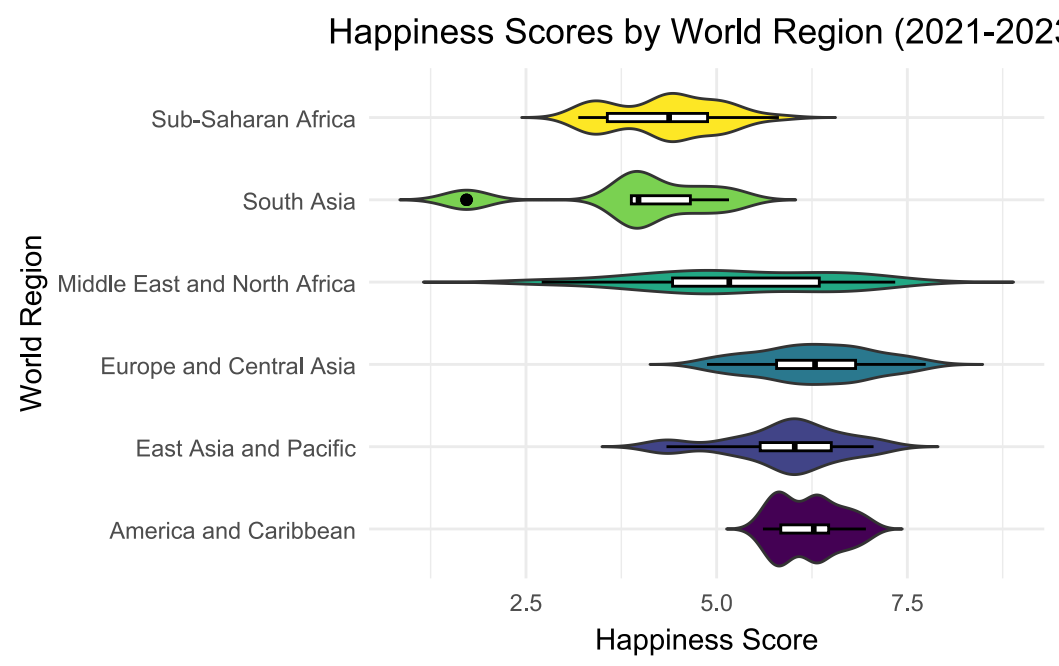
missing_region <- df_combined_region %>% filter(is.na(`World Region according to the World Bank`))
# print(missing_region)

df_combined_region$`World Region according to the World Bank`[which(df_combined_region$`World Region according to the World Bank` == "North America")] <- "Latin America and Caribbean"
df_combined_region$`World Region according to the World Bank`[which(df_combined_region$`World Region according to the World Bank` == "Latin America and Caribbean")] <- "America and Caribbean"
```

#Create violin plot

```
ggplot(df_combined_region, aes(y = `World Region according to the World Bank`, x = mean_score, fill = `World Region according to the World Bank`)) +
  geom_violin(trim = FALSE) + # Create horizontal violin plots
  geom_boxplot(width = 0.1, fill = "white", color = "black", orientation = "y") + # Overlay boxplots on the violins horizontally
  scale_fill_viridis(discrete = TRUE, guide = FALSE) + # Use viridis color scale for discrete data and remove legend
  labs(title = "Happiness Scores by World Region (2021-2023)", y = "World Region", x = "Happiness Score") +
  theme_minimal() + # Use a minimal theme
  theme(
    plot.title = element_text(hjust = 0.5) # Center the plot title
  )
```

```
Warning: The `guide` argument in `scale_*()` cannot be `FALSE`. This was deprecated in ggplot2 3.3.4.
i Please use "none" instead.
```



vi. Data Visualization

1) World Map:

```
# Get world data
world <- ne_countries(scale = "small", returnclass = "sf")

# Filter data for each year
df_2021 <- df_combined %>% filter(year == 2021)
df_2022 <- df_combined %>% filter(year == 2022)
df_2023 <- df_combined %>% filter(year == 2023)

# Convert country names to country codes
df_2021$country_code <- countrycode(df_2021$country, origin = "country.name", destination = "iso3c")
```

```
Warning: Some values were not matched unambiguously: Kosovo
```

```
df_2022$country_code <- countrycode(df_2022$country, origin = "country.name", destination = "iso3c")
```

```
Warning: Some values were not matched unambiguously: Kosovo
```

```
df_2023$country_code <- countrycode(df_2023$country, origin = "country.name", destination = "iso3c")
```

```
Warning: Some values were not matched unambiguously: Kosovo
```

```
# Merge world data with each year's data
world_data_2021 <- left_join(world, df_2021, by = c("iso_a3" = "country_code"))
world_data_2022 <- left_join(world, df_2022, by = c("iso_a3" = "country_code"))
world_data_2023 <- left_join(world, df_2023, by = c("iso_a3" = "country_code"))

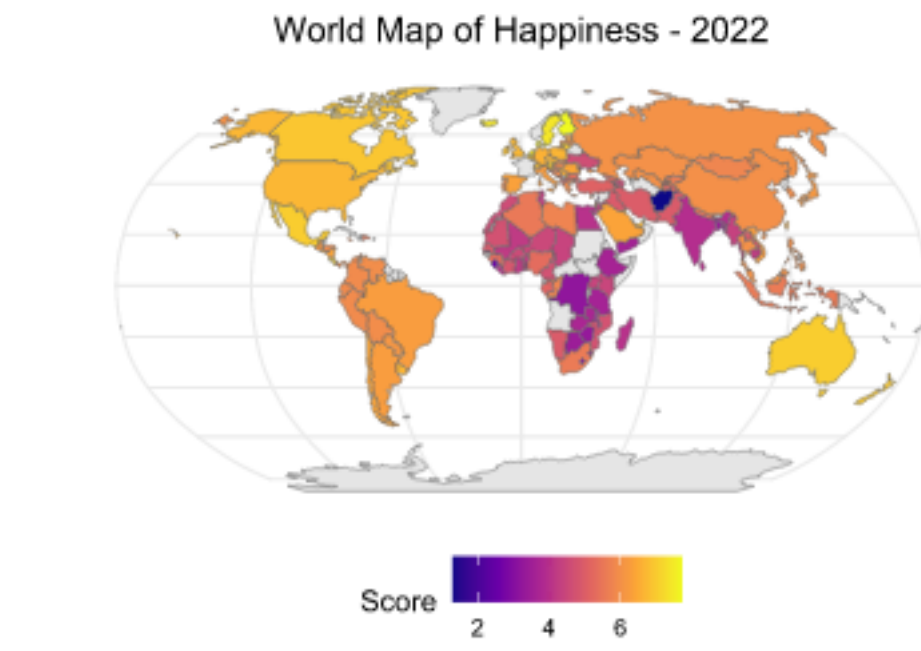
# Check for missing data after the join
missing_world_2021 <- world_data_2021 %>% filter(is.na(score))
missing_world_2022 <- world_data_2022 %>% filter(is.na(score))
missing_world_2023 <- world_data_2023 %>% filter(is.na(score))
```

```
# Plot data for 2021
ggplot() +
  geom_sf(data = world_data_2021, aes(fill = score)) +
  scale_fill_viridis_c(option = "plasma", na.value = "grey90") +
  theme_minimal() +
  theme(legend.position = "bottom") +
  labs(title = "World Map of Happiness - 2021", fill = "Score") +
  theme(plot.title = element_text(hjust = 0.5)) +
  coord_sf(crs = "+proj=robin")
```

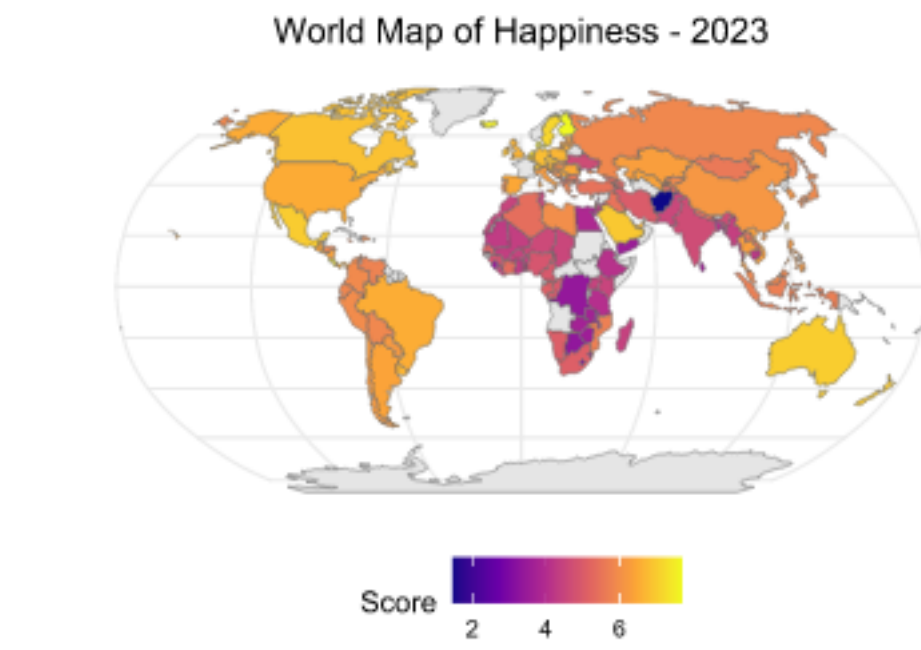


```
# Plot data for 2022
ggplot() +
```

```
geom_sf(data = world_data_2022, aes(fill = score)) +
scale_fill_viridis_c(option = "plasma", na.value = "grey90") +
theme_minimal() +
theme(legend.position = "bottom") +
labs(title = "World Map of Happiness - 2022", fill = "Score") +
theme(plot.title = element_text(hjust = 0.5)) +
coord_sf(crs = "+proj=robin")
```



```
# Plot data for 2023
ggplot() +
  geom_sf(data = world_data_2023, aes(fill = score)) +
  scale_fill_viridis_c(option = "plasma", na.value = "grey90") +
  theme_minimal() +
  theme(legend.position = "bottom") +
  labs(title = "World Map of Happiness - 2023", fill = "Score") +
  theme(plot.title = element_text(hjust = 0.5)) +
  coord_sf(crs = "+proj=robin")
```



```
# Create bins for the scores with 0.5 intervals
bins <- seq(0, 10, by = 0.5)
labels <- paste0(bins[-length(bins)], "-", bins[-1])
df_combined <- df_combined %>%
  mutate(score_bin = cut(score, breaks = bins, labels = labels, include.lowest = TRUE, right = FALSE))

# Check for any remaining missing values
missing_scores <- df_combined %>% filter(is.na(score))
if (nrow(missing_scores) > 0) {
  warning("There are still countries with missing scores.")
}
```

```
# Inspect the data for missing scores
# print(missing_scores)
```

```
# Get world data
world <- ne_countries(scale = "small", returnclass = "sf")
```

```
# Filter data for each year
df_2021 <- df_combined %>% filter(year == 2021)
df_2022 <- df_combined %>% filter(year == 2022)
df_2023 <- df_combined %>% filter(year == 2023)
```

```
# Merge world data with each year's data
world_data_2021 <- left_join(world, df_2021, by = c("iso_a3" = "country_code"))
world_data_2022 <- left_join(world, df_2022, by = c("iso_a3" = "country_code"))
world_data_2023 <- left_join(world, df_2023, by = c("iso_a3" = "country_code"))
```

```
# Check for missing data after the join
missing_world_2021 <- world_data_2021 %>% filter(is.na(score))
missing_world_2022 <- world_data_2022 %>% filter(is.na(score))
missing_world_2023 <- world_data_2023 %>% filter(is.na(score))
```



```
# print(missing_world_2021)
# print(missing_world_2022)
# print(missing_world_2023)

# Plot data for 2021
ggplot() +
  geom_sf(data = world_data_2021, aes(fill = score), color = "black") +
  scale_fill_viridis_c(option = "plasma", na.value = "grey90", name = "Score") +
  theme_minimal() +
  theme(legend.position = "bottom") +
  labs(title = "World Map of Happiness - 2021", fill = "Score") +
  theme(plot.title = element_text(hjust = 0.5)) +
  coord_sf(crs = "+proj=robin")
```



```
# Plot data for 2022
ggplot() +
  geom_sf(data = world_data_2022, aes(fill = score), color = "black") +
  scale_fill_viridis_c(option = "plasma", na.value = "grey90", name = "Score") +
  theme_minimal() +
  theme(legend.position = "bottom") +
  labs(title = "World Map of Happiness - 2022", fill = "Score") +
  theme(plot.title = element_text(hjust = 0.5)) +
  coord_sf(crs = "+proj=robin")
```



```
# Plot data for 2023
ggplot() +
  geom_sf(data = world_data_2023, aes(fill = score), color = "black") +
  scale_fill_viridis_c(option = "plasma", na.value = "grey90", name = "Score") +
  theme_minimal() +
  theme(legend.position = "bottom") +
  labs(title = "World Map of Happiness - 2023", fill = "Score") +
  theme(plot.title = element_text(hjust = 0.5)) +
  coord_sf(crs = "+proj=robin")
```



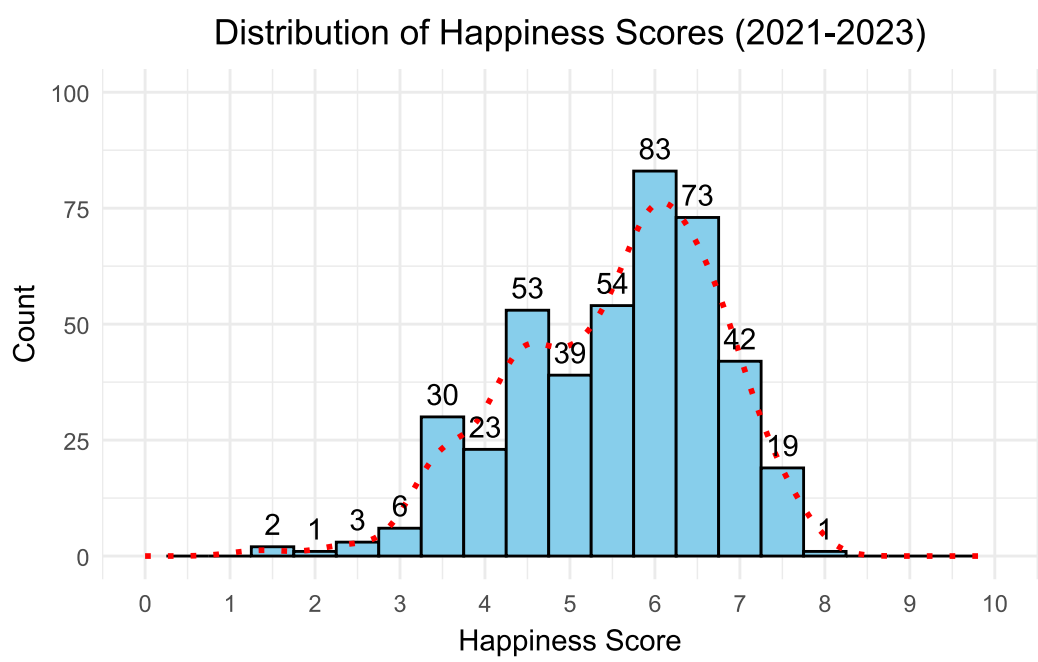
vii. Distribution Chart of Scores

```
ggplot(df_combined, aes(x = score)) +
  geom_histogram(binwidth = 0.5, fill = "skyblue", color = "black") +
  geom_density(aes(y = 0.5 * ..count..), color = "red", size = 1, linetype = "dotted") +
  stat_bin(binwidth = 0.5, geom = "text", aes(label = ifelse(..count.. > 0, ..count.., "")),
  vjust = -0.5, color = "black", position = position_stack(vjust = 1.0)) +
  scale_x_continuous(limits = c(0, 10), breaks = seq(0, 10, by = 1)) +
  scale_y_continuous(limits = c(0, 100)) +
  labs(title = "Distribution of Happiness Scores (2021-2023)",
  x = "Happiness Score",
  y = "Count") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.

Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
i Please use `after_stat(count)` instead.

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



viii. Critical Assessment and Proposed Improvements

- Interactive Data Visualization:
 - Implement hover-over information for detailed country data.
 - Add filtering and sorting capabilities.
 - Include historical comparison using a time slider.
- Consistent Labeling and Annotations:
 - Standardize labeling format for readability.
 - Provide annotations explaining key variables influencing happiness.
- Accessible Design:
 - Ensure a user-friendly interface.
 - Use color schemes accessible to color-blind users.
- Data Insights and Analysis:
 - Provide summaries and comparisons with relevant metrics.
 - Enhance linear scale visualization for regional comparisons.

These methods aim to create an improved, interactive visualization of global happiness levels, offering better insights and user engagement.

III. FINDINGS

i. New World Map

ii. Happiness Score By World Region

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequae doleamus animo, cum corpore dolemus, fieri tamen per-

- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.
- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.
- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequae doleamus animo, cum corpore dolemus, fieri tamen per-

magna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

Table 1: Lorem ipsum dolor sit amet

Lorem ipsum dolor sit.	Lorem ipsum.	Lorem ipsum.
Lorem ipsum dolor.	Lorem ipsum.	α
Lorem ipsum.	Lorem ipsum.	β
Lorem.	Lorem.	γ
Lorem ipsum dolor.	Lorem.	θ

Lorem Table 1 ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequae doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguere possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defenza et.

iii. Lorem ipsum dolor sit amet

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequae doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguere possit, augeri amplificarique non possit. At etiam Athenis, ut e patre.

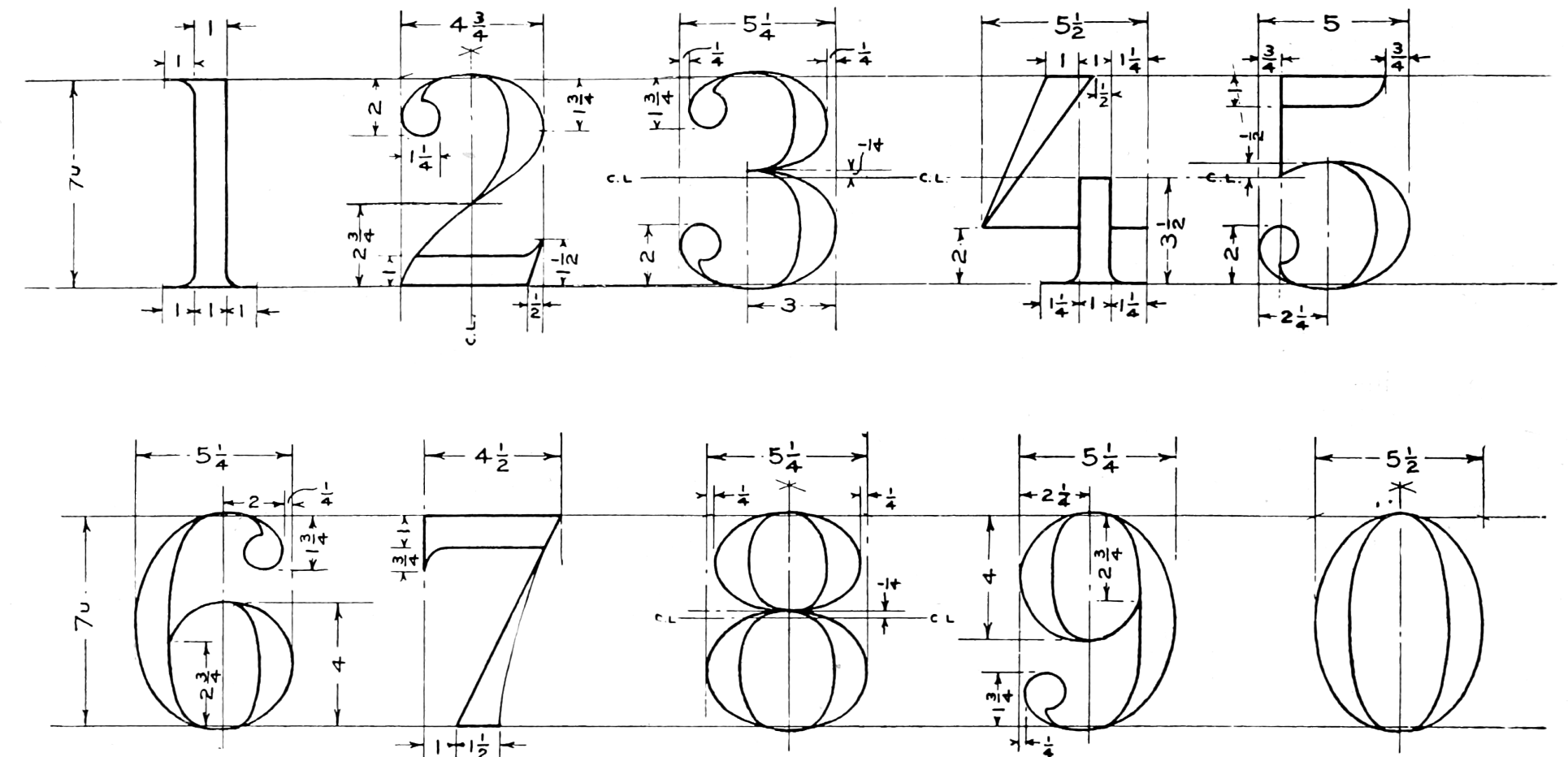


Figure 4: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.

IV. FINDINGS

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequae doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguere possit, augeri amplifi- carique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defenza et.

- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.
- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.
- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.

i. Lorem ipsum etiam

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequae doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur.

V. CONCLUSION

Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguere possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defenza et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, om- nis dolor repellendus. Temporibus autem quibusdam et.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequale doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur.

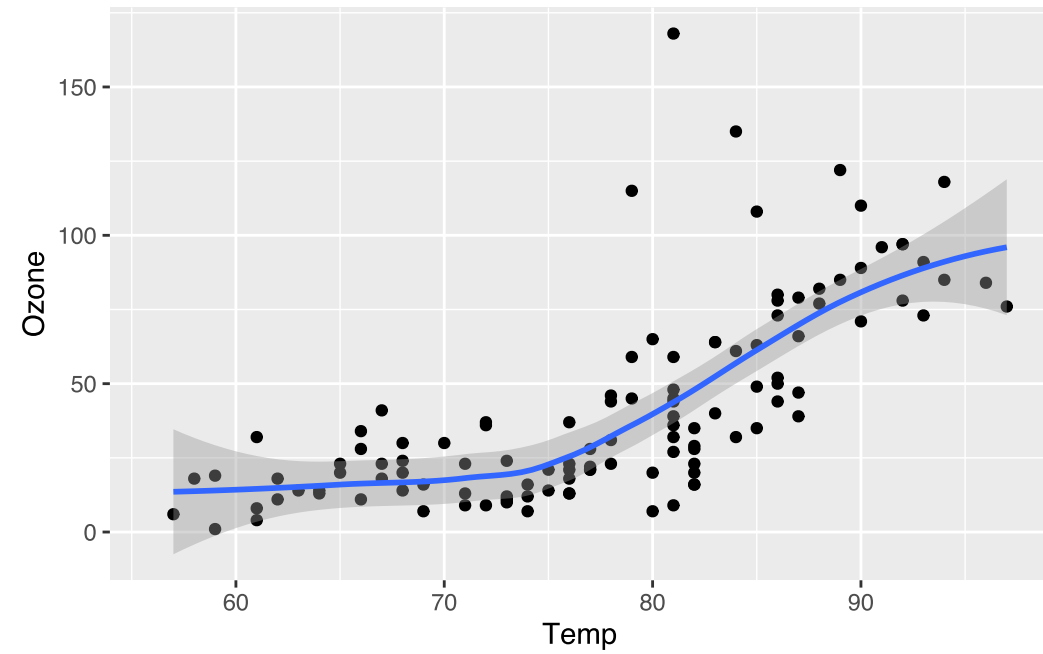


Figure 5: Temperature and ozone level

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequale doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.
 - Lorem ipsum dolor sit amet.
 - Lorem ipsum dolor sit amet, consectetur adipiscing elit.
- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore.
 - Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed.
 - Lorem ipsum dolor sit amet, consectetur adipiscing.

$$\sum_{(k=1)}^n k = \frac{n(n+1)}{2} = \frac{n^2+n}{2} \tag{1}$$

VI. ACKNOWLEDGEMENT

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequale doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.