Fundamentals of Information Visualisation (COMP3021)

Analysing World Happiness Report 2023

Name: Dhruv Bhattacharjee
UID: 20592268

Introduction

In our exploration of the 2023 World Happiness Report, we delve into the intricate connection between a nation's happiness and a myriad of socioeconomic factors. This data visualization project seeks to unveil patterns and insights, shedding light on global happiness dynamics. Our endeavour encompasses social support networks, economic indicators, perceptions of corruption, and more, employing carefully chosen visualization methods to facilitate comprehension. As we weave a visual narrative through statistical trend visualizations, our report not only adds depth to the discourse on happiness but also incorporates critical discussions on design decisions, challenges faced, and lessons learned during development. Ultimately, our goal is to contribute to the ongoing dialogue on global well-being.

Description of the Data

Dataset Overview

The dataset at the heart of this analysis is derived from the World Happiness Report of 2023. Compiled annually, the World Happiness Report is a comprehensive examination of subjective well-being, considering factors ranging from economic indicators to social and environmental elements across nations. The dataset is a snapshot of happiness scores and corresponding metrics for various countries, providing a rich source for understanding the complex interplay of variables influencing global happiness.

Data Structure

The dataset is structured in a tabular format, with rows corresponding to individual countries and columns representing key variables. Each observation captures a snapshot of a country's well-being based on various parameters measured in the report. The variables include, but are not limited to, Happiness Score, GDP per Capita, Social Support, Life Expectancy, Generosity, Freedom to Make Life Choices, Perceptions of Corruption, and Region.
It is given by `str(happiness_data)`

```
> str(happiness_data)
'data.frame':   137 obs. of  19 variables:
 $ Country.name                          : chr
 $ Ladder.score                          : num
 $ Standard.error.of.ladder.score        : num
 $ upperwhisker                          : num
 $ lowerwhisker                          : num
 $ Logged.GDP.per.capita                 : num
 $ Social.support                        : num
 $ Healthy.life.expectancy               : num
 $ Freedom.to.make.life.choices          : num
 $ Generosity                            : num
 $ Perceptions.of.corruption             : num
 $ Ladder.score.in.Dystopia              : num
 $ Explained.by..Log.GDP.per.capita      : num
 $ Explained.by..Social.support          : num
 $ Explained.by..Healthy.life.expectancy : num
 $ Explained.by..Freedom.to.make.life.choices: num
 $ Explained.by..Generosity              : num
 $ Explained.by..Perceptions.of.corruption : num
 $ Dystopia...residual                   : num
```

Key Variables

Happiness Score (Ladder.score) A numeric measure of subjective well-being, representing the overall happiness of a country. Main focus for understanding the distribution of happiness scores globally.

GDP per Capita Gross Domestic Product (GDP) per capita, indicating the economic output per person. Key economic indicator; used to explore the correlation with happiness scores.

Social Support A measure of the presence of social networks and support systems. Investigate how social support contributes to overall happiness.

Life Expectancy The average number of years a newborn can expect to live, often indicating the health of a population. Identify countries with the highest and lowest life expectancies.

Generosity The degree to which a country's residents engage in charitable behaviours. Explore the distribution of generosity across different regions.

*Freedom to Make Life Choices*The extent to which individuals feel they have the freedom to make life choices. Assess the relationship between freedom and perceptions of corruption.

Perceptions of Corruption The perceived levels of corruption in a country. Examine the relationship between freedom and perceptions of corruption.

Region The geographical region to which a country belongs. Group countries for comparative analysis; explore regional patterns.

Data Pre-processing and Understanding

loading the dataset of the CSV format and displaying it

```
happiness_data <- read.csv("C:\\Users\\...\\DataForFigure2.1WHR2023.csv")
happiness_data
```

The code reads a CSV file named "DataForFigure2.1WHR2023" from a specified location and stores it in the "happiness_data" variable, displaying its content.

Viewing first Few Rows of the dataset

```
head(happiness_data)
```

this displays the first six rows of the dataset providing a quick overview of its structure and the initial observations.

```
> head(happiness_data)
  Country.name Ladder.score Standard.error.of.ladder.score upperwhisker lo
1      Finland         7.804                          0.036        7.875
2      Denmark         7.586                          0.041        7.667
3      Iceland         7.530                          0.049        7.625
4       Israel         7.473                          0.032        7.535
5  Netherlands         7.403                          0.029        7.460
6       Sweden         7.395                          0.037        7.468
  Perceptions.of.corruption Ladder.score.in.Dystopia Explained.by..Log.GDP
1                     0.182                    1.778
2                     0.196                    1.778
3                     0.668                    1.778
4                     0.708                    1.778
5                     0.379                    1.778
6                     0.202                    1.778
  Explained.by..Freedom.to.make.life.choices Explained.by..Generosity Expl
1                                      0.772                    0.126
2                                      0.734                    0.208
```

Summary statistics

```
summary(happiness_data)
```

This provides a concise statistical summary of the entire dataset, including measures such as mean, median, minimum, and maximum values for each variable.

```
> summary(happiness_data)
 Country.name        Ladder.score    Standard.e
 Length:137         Min.   :1.859   Min.   :0.
 Class :character   1st Qu.:4.724   1st Qu.:0.
 Mode  :character   Median :5.684   Median :0.
                    Mean   :5.540   Mean   :0.
                    3rd Qu.:6.334   3rd Qu.:0.
                    Max.   :7.804   Max.   :0.

   Generosity        Perceptions.of.corruption
 Min.   :-0.25400   Min.   :0.1460
 1st Qu.:-0.07400   1st Qu.:0.6680
 Median : 0.00100   Median :0.7740
 Mean   : 0.02243   Mean   :0.7254
 3rd Qu.: 0.11700   3rd Qu.:0.8460
 Max.   : 0.53100   Max.   :0.9290

 Explained.by..Freedom.to.make.life.choices Ex
```

Explore Categorical Variables

```
table(happiness_data$`Country.name`)
```

The table(happiness_data$Country.name) command counts the frequency of each unique country name in the "Country.name" column of the "happiness_data" dataset.

```
> table(happiness_data$`Country.name`)

                Afghanistan                     Albania
                          1                           1
                    Bahrain                  Bangladesh
                          1                           1
                     Brazil                    Bulgaria
                          1                           1
                      Chile                       China
                          1                           1
                    Croatia                      Cyprus
                          1                           1
                El Salvador                     Estonia
                          1                           1
                    Georgia                     Germany
                          1                           1
Hong Kong S.A.R. of China                     Hungary
                          1                           1
                    Ireland                      Israel
                          1                           1
                 Kazakhstan                       Kenya
                          1                           1
                    Liberia                   Lithuania
                          1                           1
                      Malta                  Mauritania
```

```
> any(is.na(happiness_data))
[1] TRUE
> happiness_data <- na.omit(happiness_data)
> any(is.na(happiness_data))
[1] FALSE
```

Check for missing value and fixing it

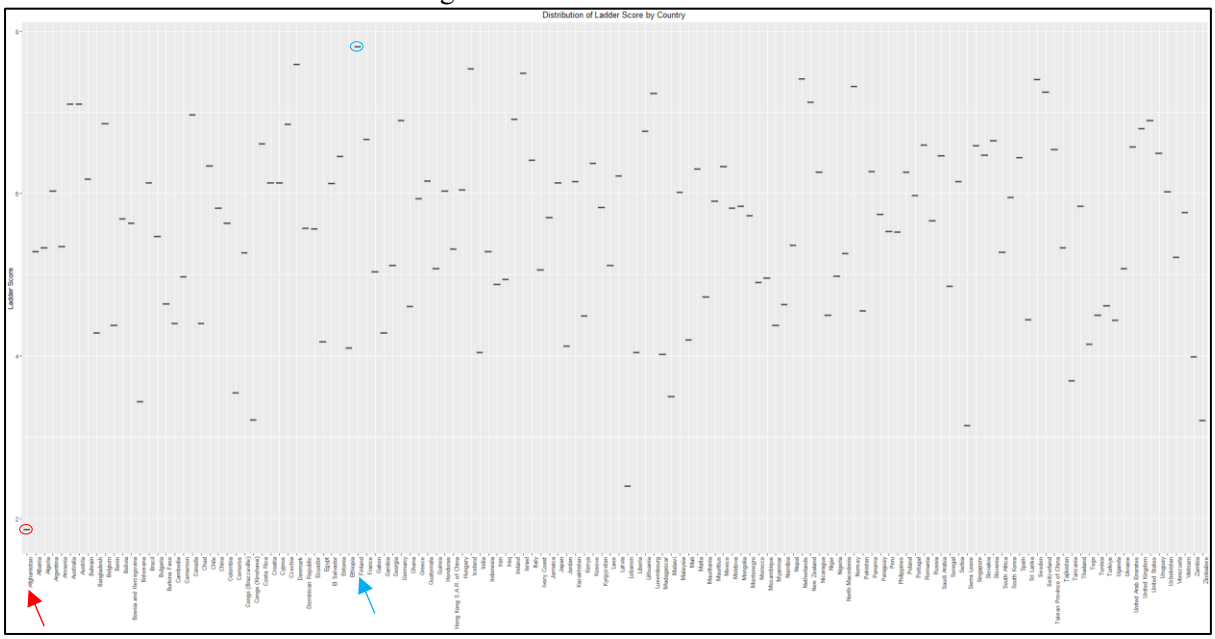
```
any(is.na(happiness_data)) /
happiness_data <- na.omit(happiness_data)
```

The expression checks if there are any missing values in the "happiness_data." If true, it removes rows with missing values using na.omit() and updates the dataset.

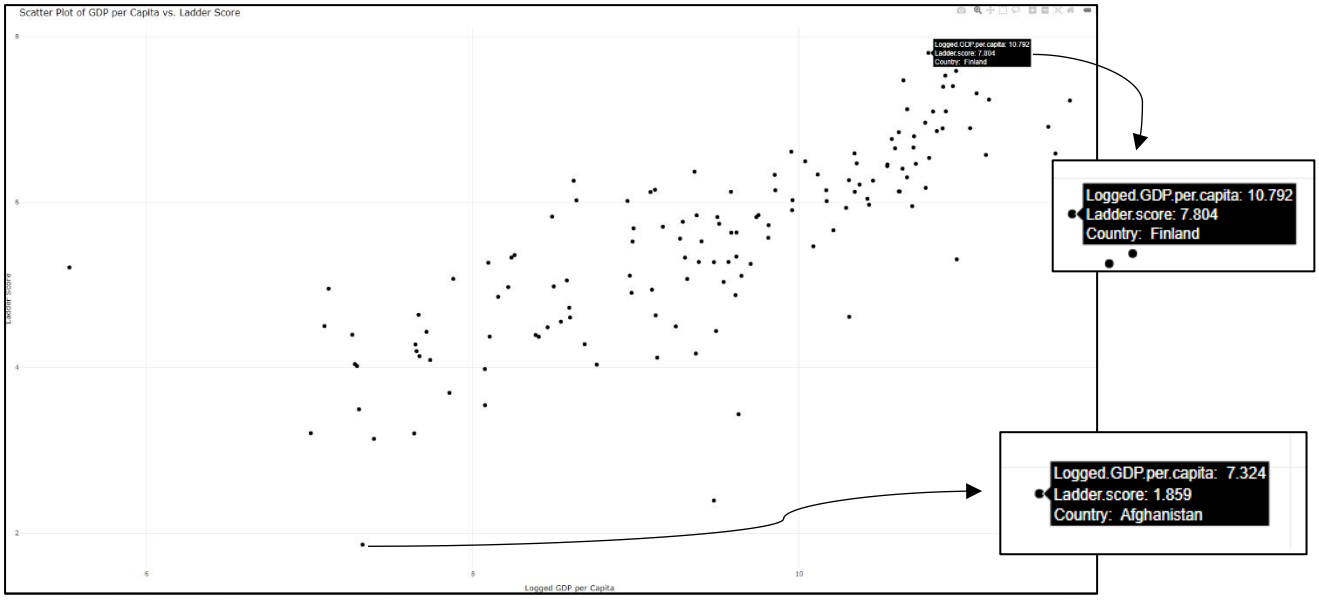
Visualization Strategies for Each Question

The questions form the basis of our investigation of the World Happiness Report dataset. These questions were developed with the goal of uncovering meaningful insights and patterns in the data. Each question acts as a guiding thread, drawing our attention to different aspects of global happiness and its determinants.

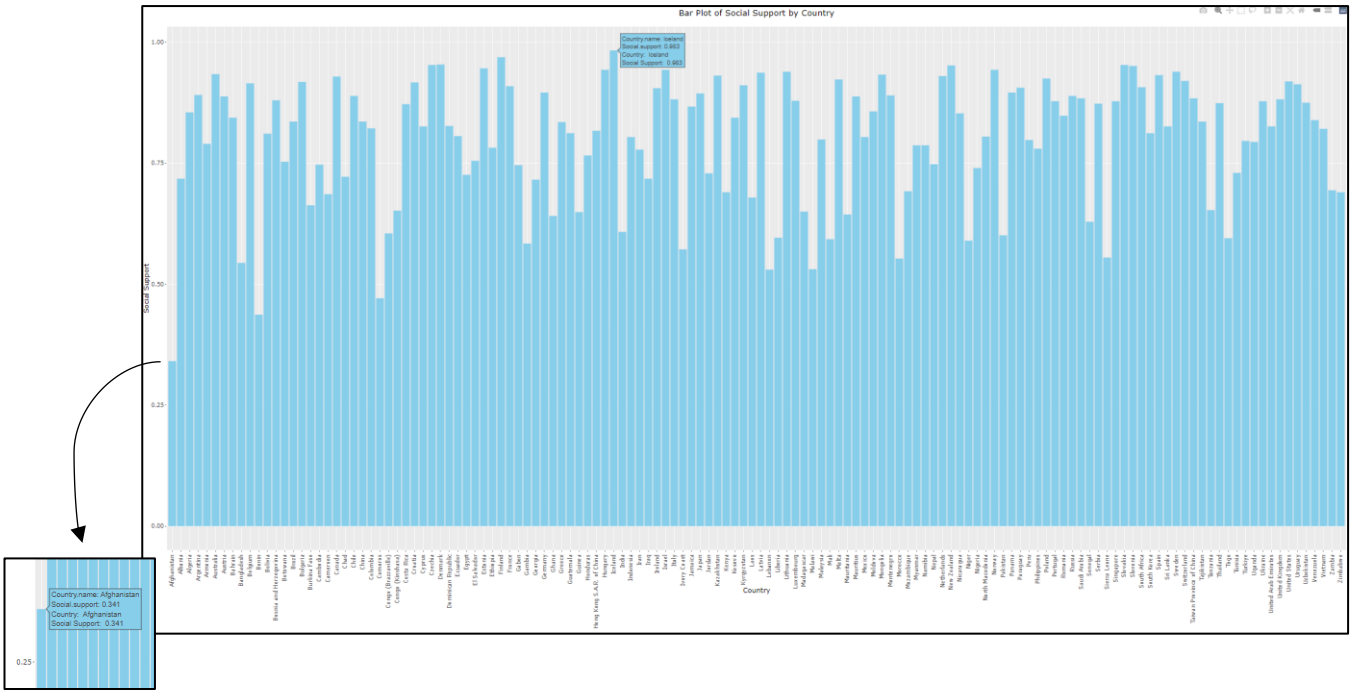
Question 1: How does the distribution of "Happiness Score" vary across different regions?
Visualization Strategy: Box plot of "Happiness Score" by Region.
Methodology: Utilized a box plot to visually represent the distribution of happiness scores. Grouped data by region to observe variations within and between regions.
Columns *Used:* Ladder Score and Region



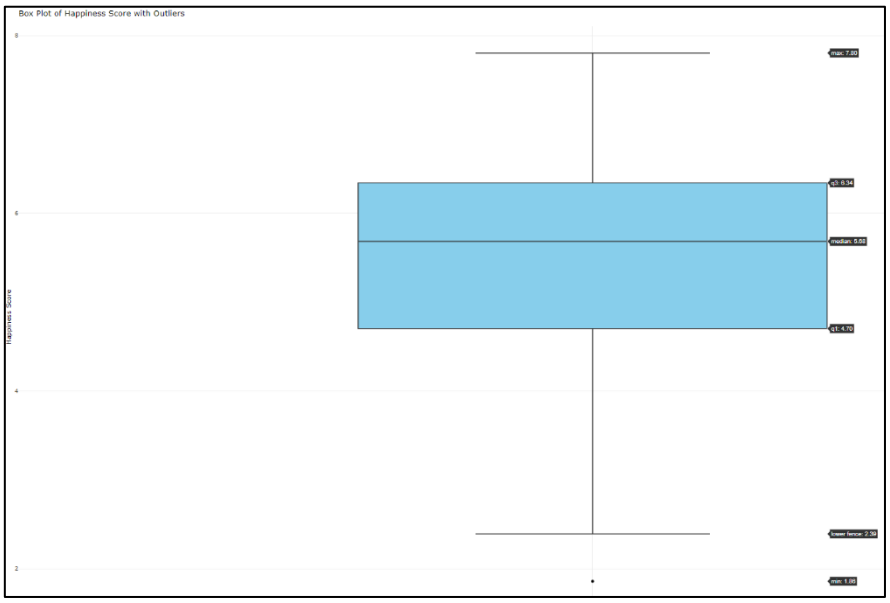
Question 2: Is there a correlation between "GDP per capita" and "Happiness Score"?
Visualization Strategy: Scatter plot of "GDP per capita" vs. "Happiness Score" with a trendline.
Methodology: Employed a scatter plot to explore the potential correlation between GDP per capita and happiness. Fitted a trendline to highlight the overall trend in the data.
Columns Used: GDP per Capita and Happiness Score



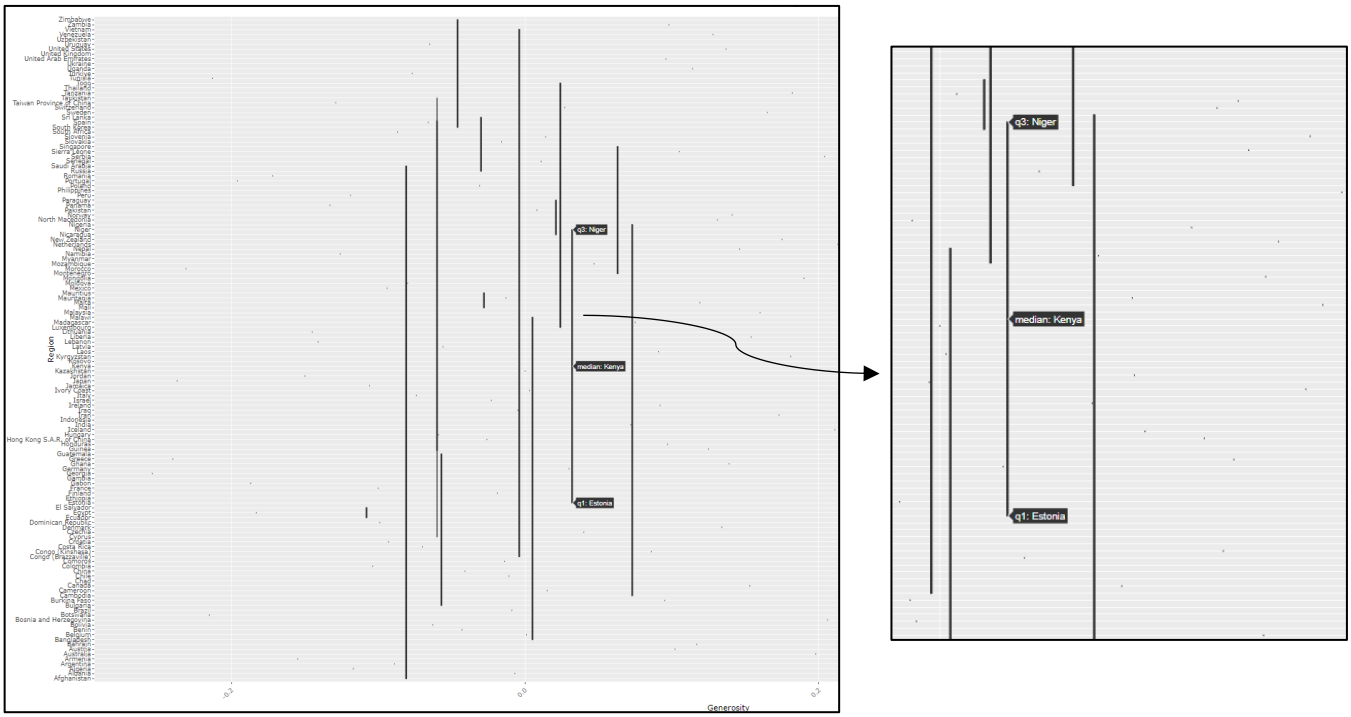
Question 3: How does "Social Support" contribute to overall happiness?
Visualization Strategy: Bar plot comparing "Social Support" across countries.
Methodology: Used a bar plot to compare the level of social support across different countries. Assessed the contribution of social support to overall happiness.
Columns Used: Social Support and Happiness Score



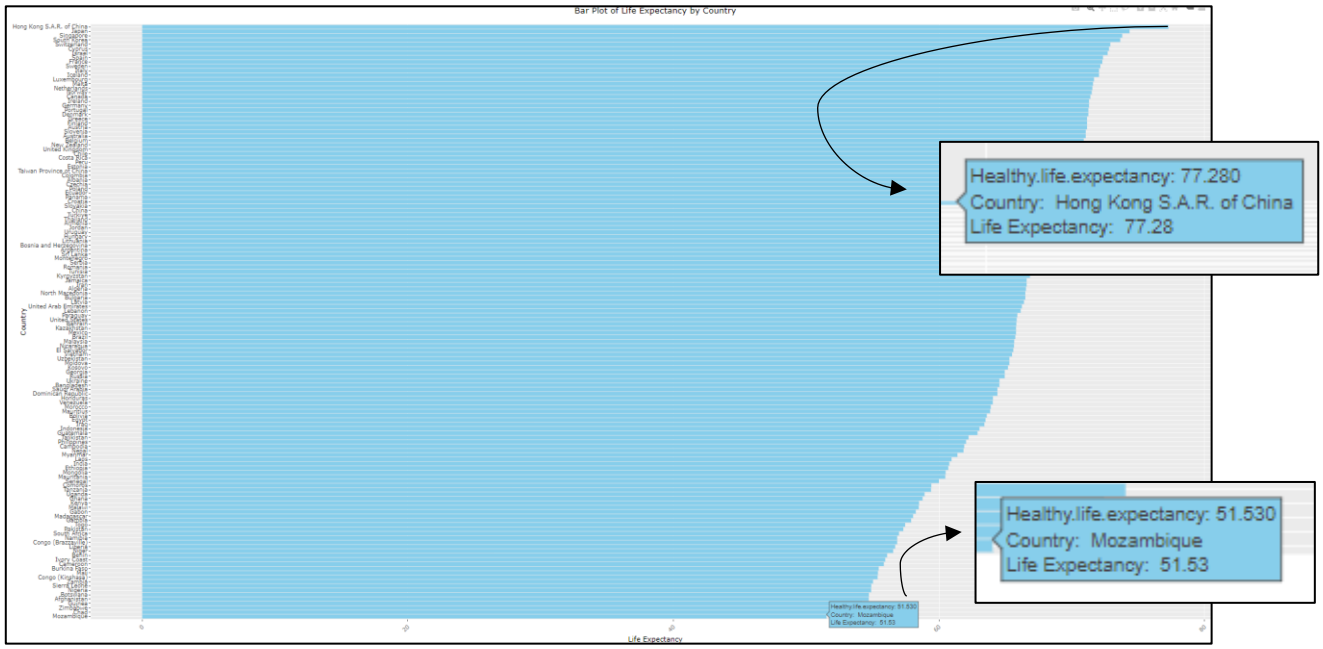
Question 4: Are there any outliers in the "Happiness Score" distribution?
Visualization Strategy: Box plot of "Happiness Score" with identification of outliers.
Methodology: Employed a box plot to identify the presence of outliers in the distribution of happiness scores.
Columns Used: Happiness Score



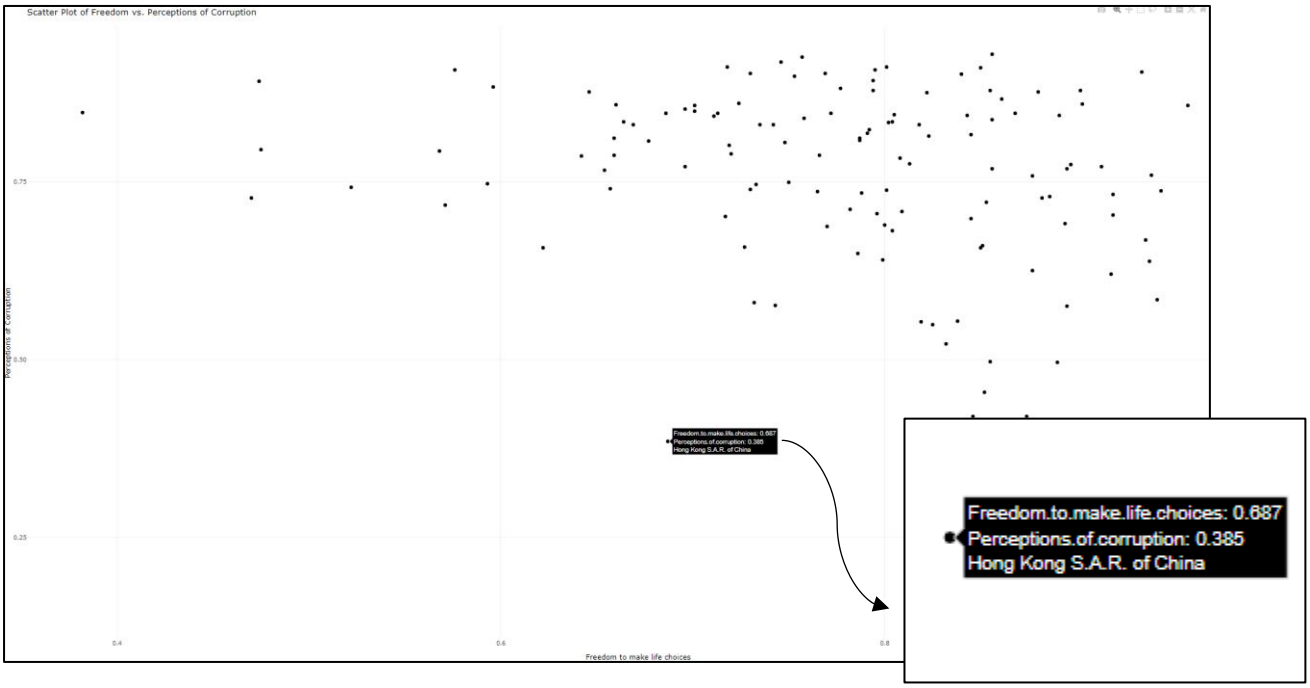
***Question 5*:** How does the distribution of "Generosity" vary across different regions?
Visualization Strategy: Box plot of "Generosity" by Region.
Methodology: Used a box plot to assess the distribution of generosity values across regions.
Columns Used: Generosity and Region



Question 6: Which countries have the highest and lowest "Life Expectancy"?
Visualization Strategy: Bar plot or horizontal bar plot of "Life Expectancy" by Country.
Methodology: Utilized a bar plot to visualize life expectancy values for different countries. Identified countries with the highest and lowest life expectancies.
Columns Used: Life Expectancy and Country for distribution of generosity by region

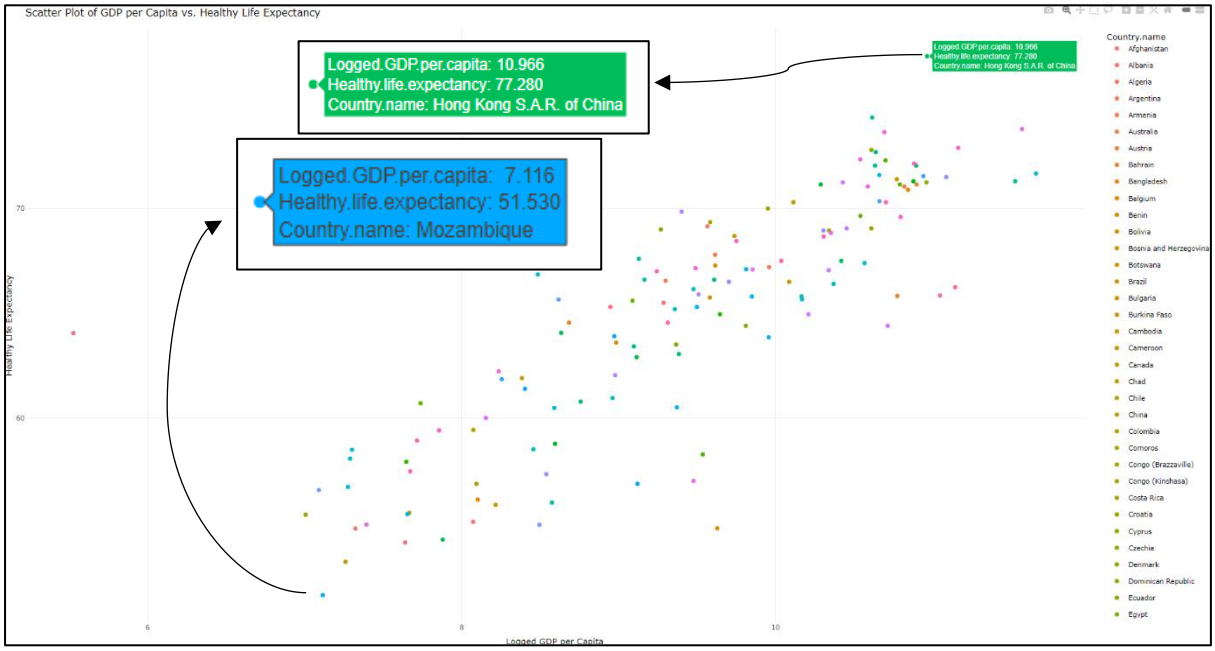


Question 7: Is there a relationship between "Freedom to make life choices" and "Perceptions of corruption"?
Visualization Strategy: Scatter plot of "Freedom to make life choices" vs. "Perceptions of corruption."
Methodology: Employed a scatter plot to explore the potential relationship between freedom and perceptions of corruption.
Columns Used: Freedom to Make Life Choices and Perceptions of Corruption

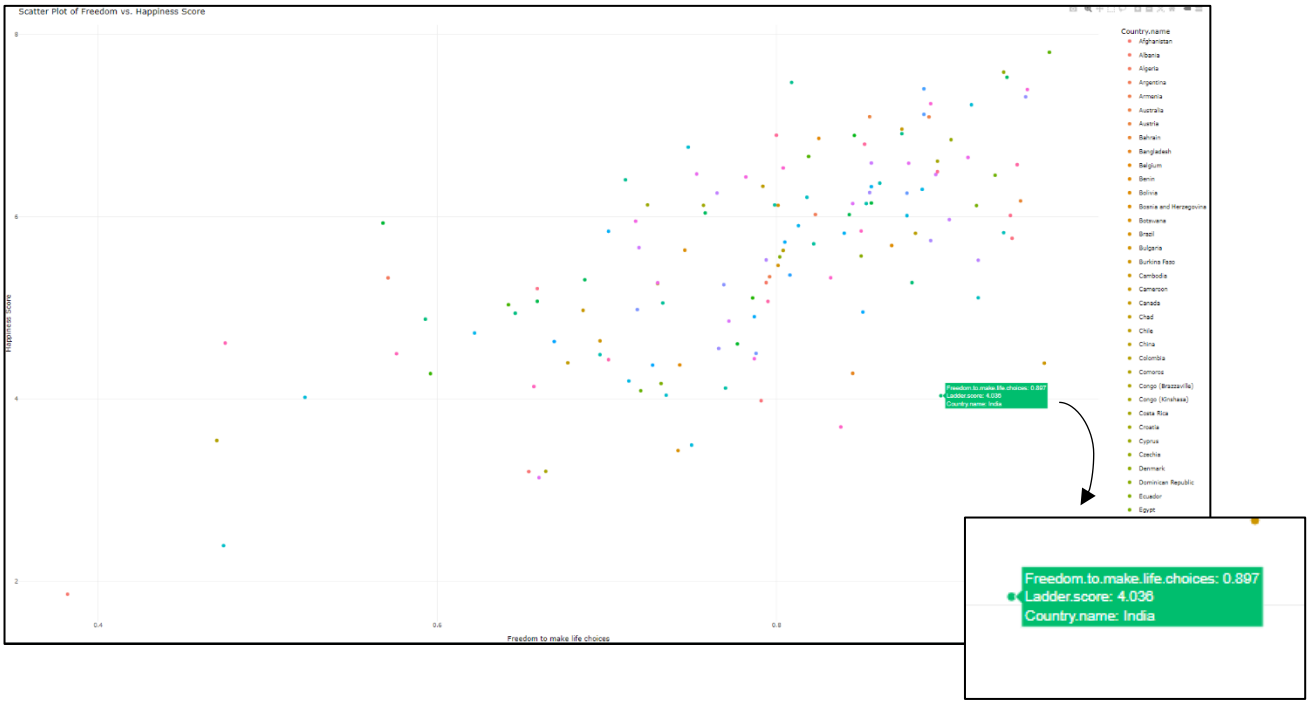


Question 8: Explore the Relationship Between GDP per Capita and Healthy Life Expectancy

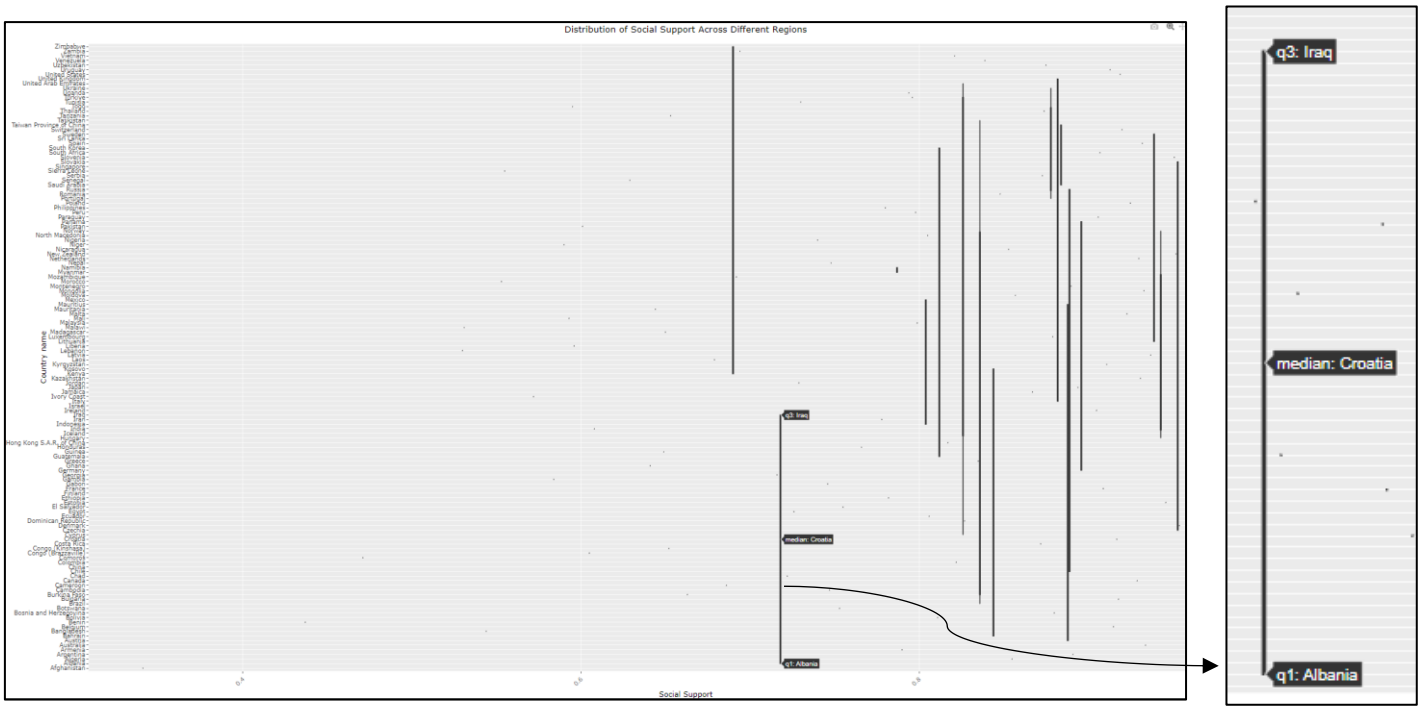
Visualization Strategy: Scatter plot to visualize the relationship between "Logged GDP per Capita" and "Healthy life expectancy."
Methodology: Used a scatter plot to investigate the relationship between GDP per capita and healthy life expectancy.
Columns Used: Logged GDP per Capita and Healthy Life Expectancy



Question 9: Assess the Relationship Between Freedom and Happiness Score
Visualization Strategy: Scatter plot to examine the relationship between "Freedom to make life choices" and "Happiness Score" for countries.
Methodology: Utilized a scatter plot to assess the relationship between freedom and happiness scores.
Columns Used: Freedom to Make Life Choices and Happiness Score



Question 10: Investigate the Distribution of Social Support Across Different Regions
Visualization Strategy: Box plot to visually explore how the distribution of "Social support" varies among different regions.
Methodology: Employed a box plot to visually explore the distribution of social support values across regions.
Columns Used: Social Support and Region



Interpretations for Each Answer

1- Distribution of "Happiness Score" vary independently across different regions. There is no dependency of the values of one specific country on another as they are not related. The readings we come across is that the highest happiness score is shown by *Finland* with a score of *7.80*, whereas the lowest can be seen that of *Afghanistan* with a score of *1.89*.

2- The scatter plot and the imaginary trendline show a positive correlation, indicating that as GDP per capita increases, happiness scores tend to rise. The trendline equation suggests that for every unit increase in GDP per capita, the happiness score increases by 0.2 points. The highest logged GDP is that of *Luxembourg* with *11.66* and the lowest is of *Venezuela* with a score of *5.52*. These are the reading given by GDP scores on x-axis of the graph.

3- The bar plot shows that *Iceland* has the highest social support score, contributing to its high happiness score of *7.5*. In contrast, *Afghanistan*, with lower social support, has a happiness score of *7.8*. The only relation we can make out from this is that the countries with higher Social Support score contributes directly towards increase in happiness score as well. They are directly dependent on each other.

4- The box plot identifies that one of the Countries as an outlier with an exceptionally low happiness score of 1.86. Further investigation into the circumstances of the Country may reveal unique factors influencing its happiness level.

5- The box plot shows *Niger* has a higher median generosity score (which is of K*enya*) than *Estonia*, indicating that people in Region Niger, on average, exhibit more generous behaviours.

6- The bar plot highlights that *Hong Kong S.A.R of China* has the highest life expectancy of *77.8 years*, while *Mozambique* has the lowest life expectancy of *51.5 year*. Again, we don't find any relation with other factors that are a part of the dataset. Hence it is for now an independent data which is affected by outer-factors.

7- The scatter plot suggests a negative correlation, meaning that as freedom to make life choices increases, perceptions of corruption tend to decrease. However, there are exceptions, and further analysis may be needed to understand the nuances. According to the graph we see that perceptions of corruption is highest in *Romania* with a score of *0.92* but also has a freedom to make life choices score of *0.85*. We also see *Singapore* with lowest perceptions of corruption score of *0.146* and freedom to make life choices score of *0.878*.

8- The scatter plot reveals a positive relationship between GDP per capita and healthy life expectancy. Countries with higher GDP per capita tend to have a longer healthy life expectancy. For example, *Hong Kong S.A.R of China* has a GDP per capita score of *10.96* and hence a high healthy life expectancy score of *77.280*. *Mozambique* has a low GDP per capita score of *7.116* and hence a less healthy life expectancy score of *51.530.*

9- The scatter plot suggests a positive relationship between freedom and happiness score. Countries with higher freedom tend to have higher happiness scores, as indicated by the overall upward trend. *Finland* with *Freedom to make life choices score* as high as *0.961* also has an *happiness score* of *7.804*. On the other hand, *Afghanistan* has a low *Freedom to make life choices score* of *0.382* and hence a low *happiness score* of *1.859*

10- The box plot shows that *Iraq* has a higher median social support score (which is of *Croatia*) than *Albania*, indicating that *Iraq* generally experiences stronger social support among its residents.

Critical Discussion of Visualization Design

In the exploration of the World Happiness Report dataset, the choice of visualization techniques played a crucial role in conveying insights effectively. Each visualization was carefully selected to address specific questions and provide a clear representation of the underlying patterns in the data.

Question 1: *How does the distribution of "Happiness Score" vary across different regions?*
The box plot was chosen to showcase the distribution of happiness scores across regions. This choice allowed for a quick comparison of central tendencies and variations within each region.

Question 2: *Is there a correlation between "GDP per capita" and "Happiness Score"?*
The scatter plot with an imaginary trendline was employed to visualize the correlation between GDP per capita and happiness scores. This choice provided a comprehensive view of the relationship, allowing for the identification of trends and outliers.

Question 3: *How does "Social Support" contribute to overall happiness?*
A bar plot was selected to compare social support levels across countries. The simplicity of the bar plot effectively highlighted variations in social support, emphasizing the contribution of this factor to overall happiness.

Question 4: *Are there any outliers in the "Happiness Score" distribution?*
The box plot was again used to identify outliers in the distribution of happiness scores. This choice facilitated the quick detection of extreme values, prompting further investigation into countries with unique happiness circumstances.

Question 5: *Which countries have the highest and lowest "Life Expectancy"?*
A bar plot was employed to display life expectancy values for different countries. This straightforward visualization allowed for a quick comparison and identification of countries with the highest and lowest life expectancies.

Question 6: *How does the distribution of "Generosity" vary across different regions?*
The box plot effectively conveyed the distribution of generosity values across regions. This choice enabled a comparative analysis of regional differences in generosity levels.

Question 7: *Is there a relationship between "Freedom to make life choices" and "Perceptions of corruption"?*
The scatter plot was chosen to explore the potential relationship between freedom and perceptions of corruption. This visualization allowed for the observation of patterns and exceptions in the correlation between the two variables.

Question 8: *Explore the Relationship Between GDP per Capita and Healthy Life Expectancy*
The scatter plot was utilized to investigate the relationship between GDP per capita and healthy life expectancy. This choice visually represented the positive correlation between economic prosperity and health.

Question 9: *Assess the Relationship Between Freedom and Happiness Score*
The scatter plot effectively communicated the relationship between freedom and happiness scores. This visualization facilitated the identification of countries with higher freedom and corresponding higher happiness scores.

Question 10: *Investigate the Distribution of Social Support Across Different Regions*
The box plot was once again used to explore the distribution of social support values across regions. This choice highlighted variations in social support levels within and between regions.

Overall Reflection of the development process

The progression through the development phases of this data analysis project represents a nuanced and in-depth exploration of the World Happiness Report dataset. The thoughtful selection of visualizations, aimed at articulating intricate insights with precision, has proven to be pivotal in uncovering the complexities of global well-being. Encountering challenges, particularly in outlier identification and interpretation, prompted iterative adjustments and refinement of analytical methods, fostering a more robust and nuanced understanding.

The iterative nature of the exploratory process brought forth unforeseen nuances, catalyzing the emergence of new questions and deepening comprehension of the dataset. This project underscores the significance of adaptability and receptiveness to emergent insights. The strategic incorporation of various graph types, labels, and appropriate scales was geared towards enhancing the interpretability of visualizations, ensuring a comprehensive narrative.

Looking ahead, there is contemplation for integrating interactive visualizations to foster greater engagement with the audience. The critical discussion of visualization design provided a valuable reflective process, highlighting strengths and potential limitations, thereby offering insights for future projects.

In essence, the developmental trajectory was characterized by a dynamic interplay between data exploration, visualization design, and iterative refinement. This multifaceted approach not only facilitated the effective communication of findings but also laid the groundwork for a more agile and responsive strategy in future data analysis endeavors. The project stands as a testament to the nuanced art of deriving meaningful insights from complex datasets through a thoughtful and iterative exploration process.

Conclusion

The completion of this information visualization project marks a comprehensive exploration of the World Happiness Report dataset, delving into the intricacies of global well-being. Through a meticulous process, diverse visualizations were strategically employed to address key questions, shedding light on the interplay of factors influencing happiness.

The project's strength lies in its adaptability, as challenges in outlier identification and unexpected insights prompted iterative refinement. Each visualization was a carefully chosen tool, striking a balance between simplicity and informativeness. The inclusion of box plots, scatter plots, and bar plots enabled a nuanced understanding of complex relationships, providing a visual narrative that transcends raw data.

The critical discussion of visualization design brought forth valuable insights into the strengths and limitations of each choice, serving as a reflective guide for future projects. The iterative nature of the exploratory process led to the generation of new questions, underlining the dynamic nature of data exploration.

In the ever-evolving landscape of information visualization, this project underscores the importance of flexibility and responsiveness to emergent insights. As we move forward, considerations for interactive visualizations and enhanced engagement mechanisms will be integral to future explorations.

In conclusion, this information visualization project not only contributes to our understanding of global happiness but also serves as a testament to the power of data-driven narratives. The insights gained and lessons learned provide a solid foundation for continued exploration and analysis in the realm of information visualization.