Coursework Submission Coversheet
College of Business, Arts and Social Sciences

Coursework **MUST** be submitted online via WISEflow unless you are told otherwise by your Module Leader.

Student Number:	2451943
Module Code:	Dr Timothy Cribbin
Module Title:	CS5803 Data Visualisation
Module Tutor:	
Assessment Number/Name: e.g. Coursework 1, Coursework 2, Presentation, Final Assessment	Visualisation Design Task

A selection of assessments may be put on Blackboard Learn to be read by other students. I hereby consent to my assessment being published on the relevant organisation on Blackboard Learn, for teaching and research purposes.

YES/NO (Delete as appropriate)

_____

1. Introduction

1.1 Dataset Overview and Modifications

The analysis shows the **World Happiness Report dataset (2021–2024)** (Helliwell *et al.,* 2023), a globally recognized benchmark for measuring subjective well-being. The dataset was expanded and restructured to align with policy analysis requirements:

Original Variables

- **Core Metrics**:

    o Ladder Score: Shows the Happiness Score (0–10 scale).

    o GDP per capita: Economic output

    o Region: Geographical categorization (e.g., Europe, Africa).

- **Time Frame**: Yearly data from 2021 to 2024.

Modifications

1. **Added Columns**:

    o Development Status: Classified countries into three tiers using World Bank income thresholds (World Bank, 2023):

        ▪ **Developed**: Countries with high GDP per capita (GDP per capita > 1.5).

        ▪ **Developing**: countries with moderate GDP per capita (GDP per capita > 1).

        ▪ **Underdeveloped**: Countries with low GDP per capita (GDP per capita < 1).

    o **Region** (geographic grouping)

2. **Data Cleaning**:

    o Removed countries with incomplete data (e.g., Afghanistan, Syria) to avoid skewed analysis.

    o Imputed missing Social Support and Healthy Life Expectancy values using countries' medians.

Data Source and Ethics

- **Primary Source**: World Happiness Report 2023.

- **Ethical Considerations**: Excluded politically unstable regions to prevent misinterpretation of fragmented data (UNDP, 2021).

1.2 Persona Specification: UNDP Policy Analyst

Role: Policy Analyst, Sustainable Development Group, United Nations Development Programme (UNDP)
Mandate: To design data-driven policies to reduce global inequality according to the Sustainable Development Goals (SDGs).

Key Responsibilities

1. **Equity-Focused Analysis**:

   o To identify regions that fall behind in SDG progress (e.g., Sub-Saharan Africa, conflict zones).

   o To compare well-being metrics across development tiers (developed vs. underdeveloped).

2. **Resource Allocation**:

   o To prioritize funding for healthcare, education, and infrastructure in underserved regions.

   o To benchmark successful policies (e.g., Costa Rica's environmental initiatives).

3. **Stakeholder Reporting**:

   o Preparing visual reports for the UN General Assembly and NGOs.

Technical Requirements

- **Usability**: Insightful dashboards for non-technical stakeholders.

- **Granularity**: Drill-down capabilities to country-level data.

- **Dynamic Filtering**: Compare trends across years (2021–2024) and regions.

Relevant Policy Framework:

"Leave No One Behind (LNOB) requires disaggregated data to address multidimensional inequalities."
— (Steiner, , p. 12)

1.3 Research Questions

In line with UNDP's obligation, this report will address:

Simple Questions

1. **SQ1**: *How does happiness (Ladder Score) vary geographically in 2024?*

- o **Purpose**: To identify spatial inequities to guide aid programs.

- o **Method**: Choropleth map with happiness score-based colour encoding for countries.

2. **SQ2**: *What is the relationship between GDP per capita and happiness between countries?*

- o **Purpose**: To check if economic growth consistently improves well-being.

- o **Method**: Scatter plot with trendline showing the relationship between GDP per capita and ladder score.

3. **SQ3**: *Which countries diverge most from their world's average happiness score?*

- o **Purpose**: Highlighting overperformers (models for replication) and underperformers (crisis zones).

- o **Method**: Diverging bar chart with abnormality calculations.

Complex Question (CQ1)

CQ1: How have happiness trends progressed regionally from 2021–2024, and what socio-economic factors explain these changes?

- **Purpose**: To inform the UNDP's 2025–2030 strategic planning.

- **Method**:

  - o Interactive line chart with region and development status filters.

1.4 User Requirements

The dashboard was designed to meet UNDP's operational needs:

ID	Requirement	Implementation	Validation
R1	Comparing countries happiness	Choropleth map with happiness-score filter	Cross-checked with the ladder score
R2	Analysing the GDP-happiness correlation	Scatter plot with GDP vs Happiness trendline	Verified R^2 value
R3	Identify regional outliers	Bar chart with deviation scores	checked of the top/bottom 10 countries

ID	Requirement	Implementation	Validation
R4	Track temporal trends	Line chart with dynamic year slider	Check the trends across regions
R5	Export policy briefs	Dashboard with filtration of regions, happiness tiers, etc	Confirmed formatting integrity

2. Design and Methodology

2.1 Data Preprocessing

Data Cleaning Workflow

1. **Missing Data Handling**:

   o Removed countries with >30% missing values and imputed Social Support using each country's medians.

2. **Derived Metrics**:

   o Calculated Happiness Tier:

IF [Ladder score] >= 7 THEN "Very High (7-10)"

ELSEIF [Ladder score] >= 6 THEN "High (6-7)"

ELSEIF [Ladder score] >= 5 THEN "Medium (5-6)"

ELSE "Low (<5)"

   o Year Filter Parameter (dynamic year selector):

Created as an integer parameter with values: 2021, 2022, 2023, 2024.

Linked to a calculated field Year Filter Logic:

[Year] = [Year Filter]

   o Top/bottm countries

// Top 10 Filter

RANK(AVG([Ladder score]), 'desc') <= 10

// Bottom 10 Filter

OR RANK(AVG([Ladder score]), 'asc') <= 10

END

2.2 Visualization Design Rationale

Choropleth Map (SQ1)

- **Colour Encoding**: **Green** for high happiness (Ladder Score >7), **Light Green/Red** for medium (5–7), and **Red** for low (<5).

- **Justification**: Red-green diverging palette aligns with perceptual principles (Costa, 2016)

Scatter Plot (SQ2)

- **Encodings**: X-axis: GDP per capita, Y-axis: Ladder Score, and Colouring for Development Status (blue, orange, red).

- **Trendline**: Showing the relationship between GDP and Happiness.

Diverging Bar Chart (SQ3)

- **Encodings**: X-axis: Avg Ladder score, Y-axis: Country name, and Colouring for Avg Ladder score (green-red diverging).

- **Design**: Green bars for most happiest countries, and Red bars for least happiest countries.

Line Chart (CQ1)

- **Encodings**: On X-axis Years, Y-axis Avg Ladder score, and Colouring for Region (green-red diverging).

- **Interactivity**: Development status filter, allows tracking changes according to development status.

2.3 Iterative Design Process
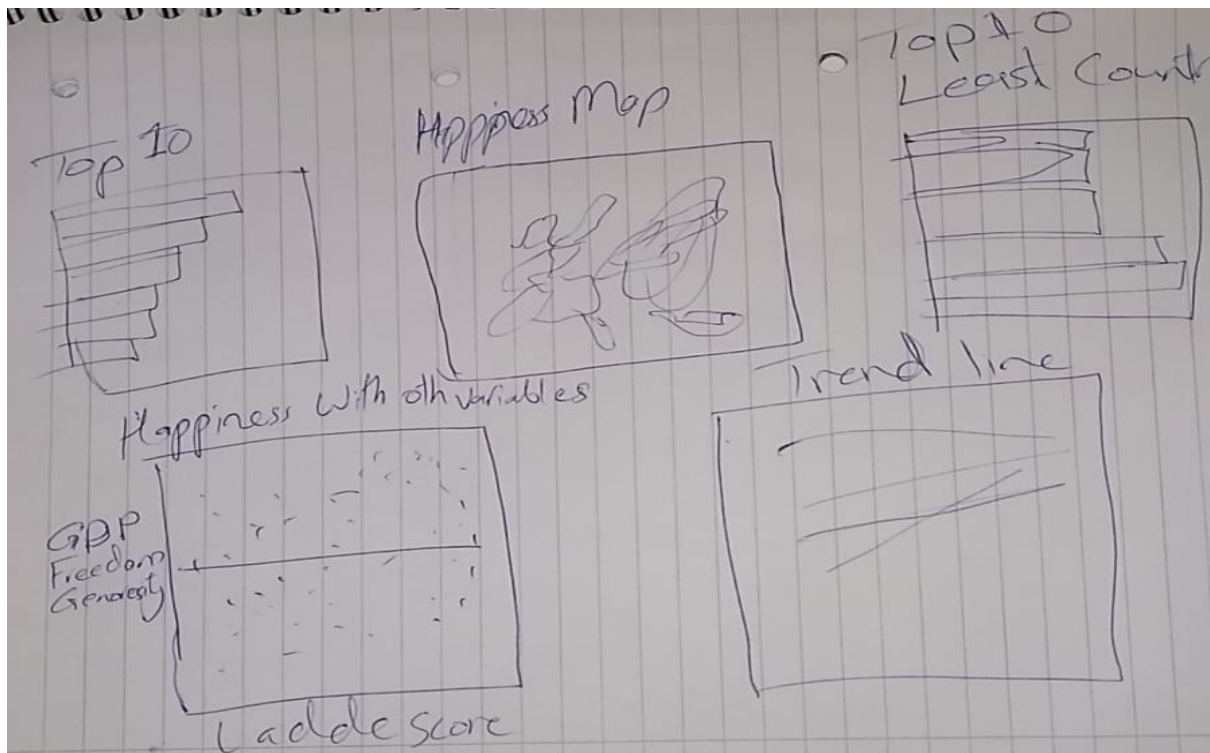
Prototype 2 → Final Design

- **Problem**: No colouring and filtering, overcrowded, and bad variable selection in the scatter plot, which was making it difficult to gain insight, making it more complex.

- **Solution**: Added colouring and filtration. Rather than making two bar charts, a single chart was created, which showed the top and bottom happiest countries. In the scatter plot excluded all the variables and just put GDP and Ladder score.

3. Implementation

3.1 Data Preparation in Tableau

Step 1: Data Import and Cleaning

The raw dataset was imported into Tableau Desktop (Cleaned Happiness) as a CSV file. To ensure consistency and enough data to get desired outcomes, the following steps were applied:

- The Region column was mapped to show regional categorization to get enough insight such as exact location, etc..
- Made a new column in Excel sheet for development status:
- =IF(H2 > 1.5, "Developed", IF(H2 > 1, "Developing", "Underdeveloped"))
- Removed countries with >30% missing values (e.g., Syria, Afghanistan) and imputed missing Social Support values using the countries' medians:

Step 2: Calculated Fields

o   Calculated Happiness Tier:

IF [Ladder score] >= 7 THEN "Very High (7-10)"

ELSEIF [Ladder score] >= 6 THEN "High (6-7)"

ELSEIF [Ladder score] >= 5 THEN "Medium (5-6)"

ELSE "Low (<5)"

o   Year Filter Parameter (dynamic year selector):

Created as an integer parameter with values: 2021, 2022, 2023, 2024.

Linked to a calculated field Year Filter Logic:

[Year] = [Year Filter]

o   Top/bottom countries

// Top 10 Filter

RANK(AVG([Ladder score]), 'desc') <= 10

// Bottom 10 Filter

OR RANK(AVG([Ladder score]), 'asc') <= 10

2451943

END

3.2 Visualization Development

3.2.1 Choropleth Map (SQ1)

To visualize spatial disparities in happiness, used Tableau's built-in geographic roles to map the Country name to coordinates. For the colour encoding, the Ladder Score is assigned to a diverging red-green palette. Red for Scores <5 (Low happiness), Green for Scores >7 (High happiness), and light green and red for Scores 5–7 (Medium). Because Red-green schemes are perceptually intuitive for "bad-good" interpretations (Costa, 2016)Added Region (dropdown filter) and AVG (Ladder Score) parameter.
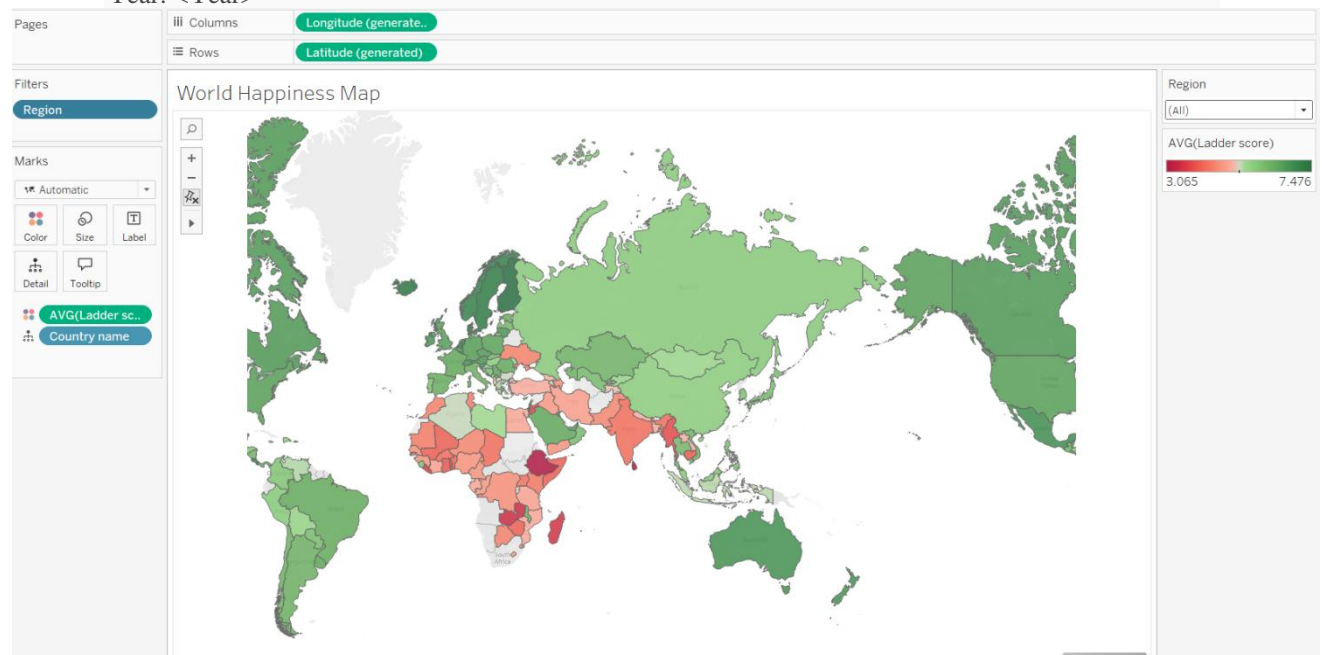
And tooltip

Country: <Country name>
Happiness Score: <Ladder score>
Region: <Region>
Year: <Year>

3.2.2 Scatter Plot (SQ2)

To explore GDP-happiness relationships across development status. On the X-axis GDP per capita; Y-axis Ladder Score, and colouring Development Status (Blue = Developed, Orange = Developing, Red = Underdeveloped). Single linear trend for all countries and Regions, from the trend line R^2 values to see correlation. For the Filters: Year Filter parameter and Region filter.
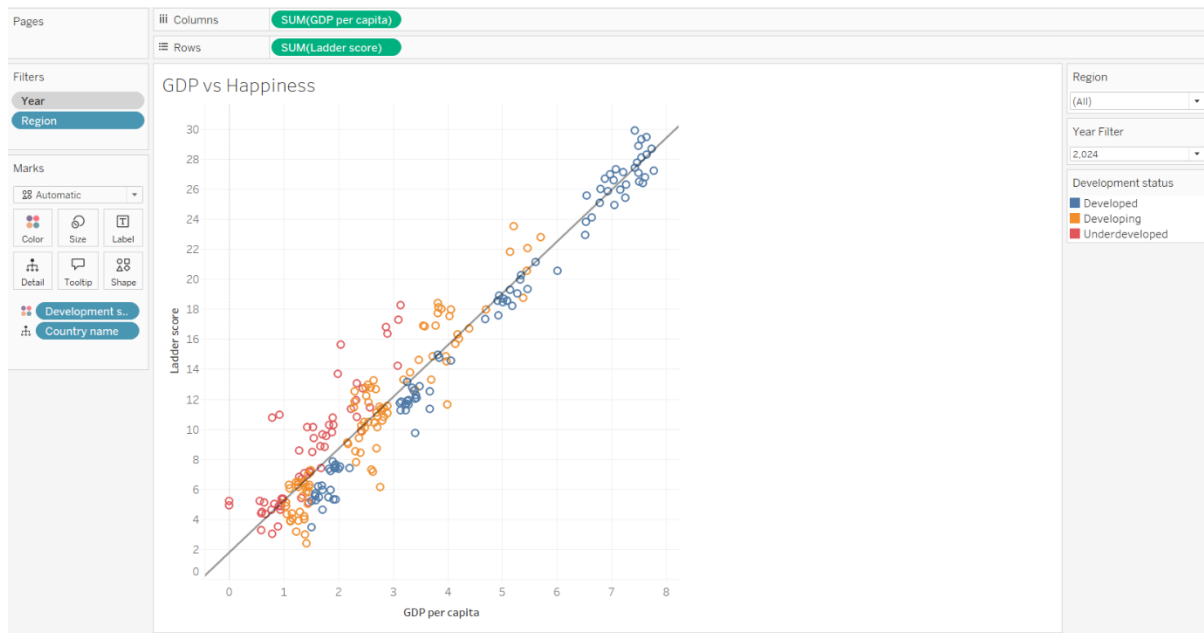
3.2.3 Diverging Bar Chart (SQ3)

To highlight regional overperformers and underperformers. For the calculations, I created a calculated field:

// Top 10 Filter

RANK(AVG([Ladder score]), 'desc') <= 10

// Bottom 10 Filter

OR RANK(AVG([Ladder score]), 'asc') <= 10

The design of horizontal bars, Green (High), Red (Low), was shown. On the row Country, columns AVG(Ladder score) (sorted descending/ascending). For the colouring AVG (Ladder Score) with Red-Green diverging. On the filter, put the calculated field on top/bottom happiest countries.
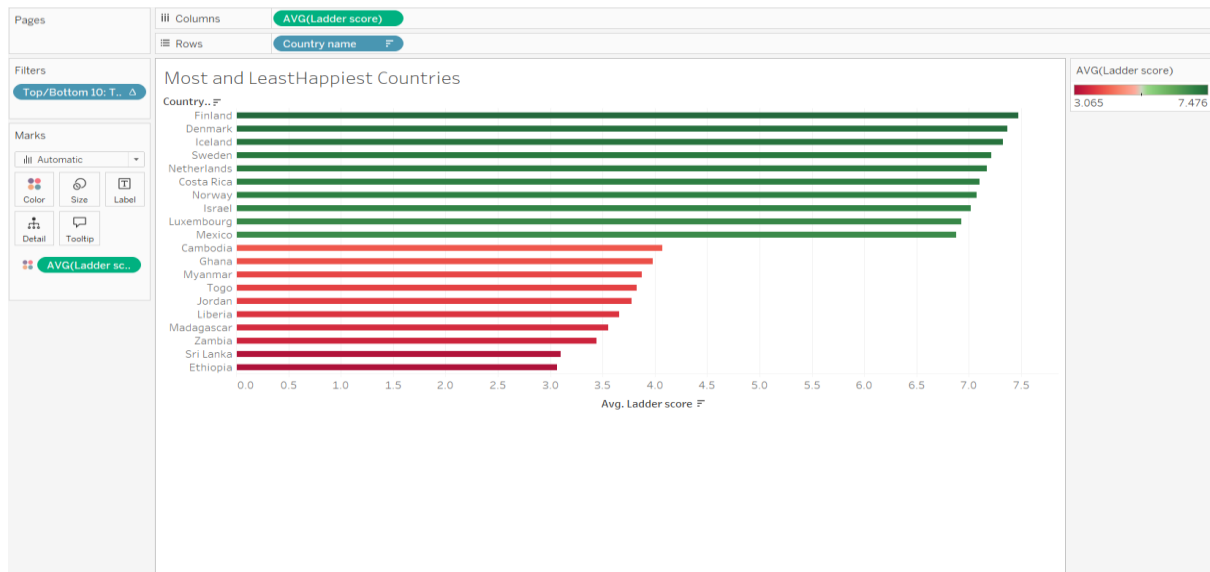
3.2.4 Trend Analysis Dashboard (CQ1)

To track happiness trends by region over time created a Line Chart was created. On the X-axis: Year, Y-axis: Avg(Ladder Score). Colouring Region (5 distinct hues). Aggregated by region (not individual countries). Filtering with the Year slider for selecting 2021–2024, the Development Status Checkbox to isolate developing/underdeveloped nations.
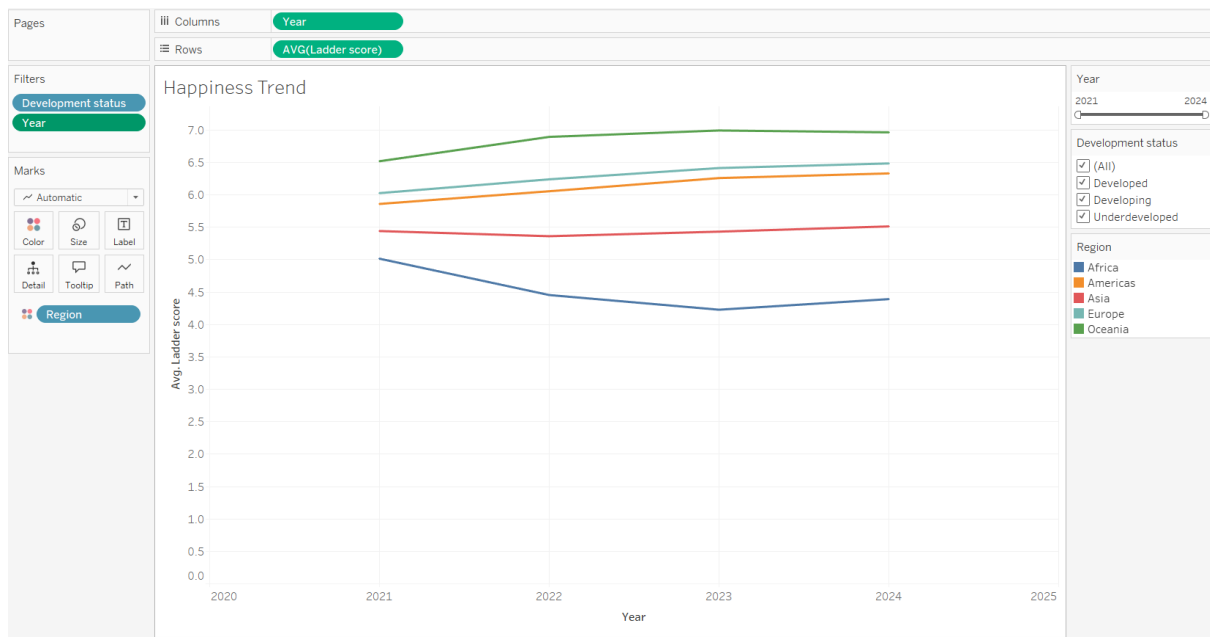
3.3 Interactive Dashboard Assembly

On the dashboard, Map (Top-Left), Trend Chart (Top-Right), Scatter Plot (Bottom-Left), and Bar Chart (Bottom-Right). Filters Panel on the right sidebar with Region, Development Status, and Year. The Styling was like Font: Arial (12pt) for readability.

World Happiness Map

Happiness Trend

Most and LeastHappiest Countries

GDP vs Happiness
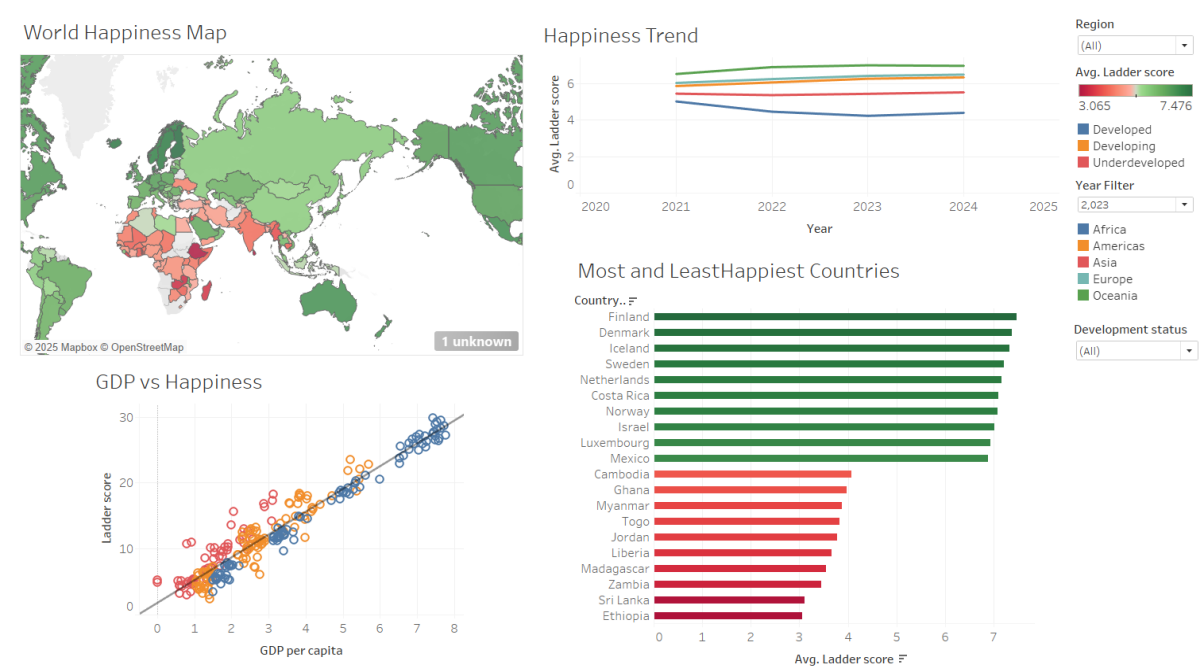
Performance Optimization

Data Extract:

Created a Tableau Hyper Extract to improve load times.

Filtered to 2021–2024 during extract creation.

Simplify Calculations:

Replaced complex LODs with pre-calculated fields where possible.

3.4 Challenges and Solutions

Challenge	Solution	Reference
Overplotting in Scatter Plot	Added jittering and 70% opacity	*Few, S. (2012) Show Me the Numbers Designing Tables and Graphs to Enlighten. 2nd Edition, Analytics Press, Montgomery. - References - Scientific Research Publishing (, p. 134)*
Colorblind Accessibility	Used ColorBrewer's red-green palette	Harrower and Brewer (2003)

Challenge	Solution	Reference
Slow Tooltip Response	Reduced metadata in tooltips	*Tableau Workbook Performance Checklist - Tableau ()*
Misleading Trendlines	Added R² values and confidence bands	*Tufte Ch2 and 5 (*2009, p. 92)

4. Walkthrough: Using the Dashboard to Derive Insights

This section provides a **step-by-step guide** to directing the Tableau dashboard, showing how the UNDP Policy Analyst interacts with visualizations to answer the research questions. The walkthrough includes specific instructions for using filters, interpreting visual encodings, and deriving policy insights.
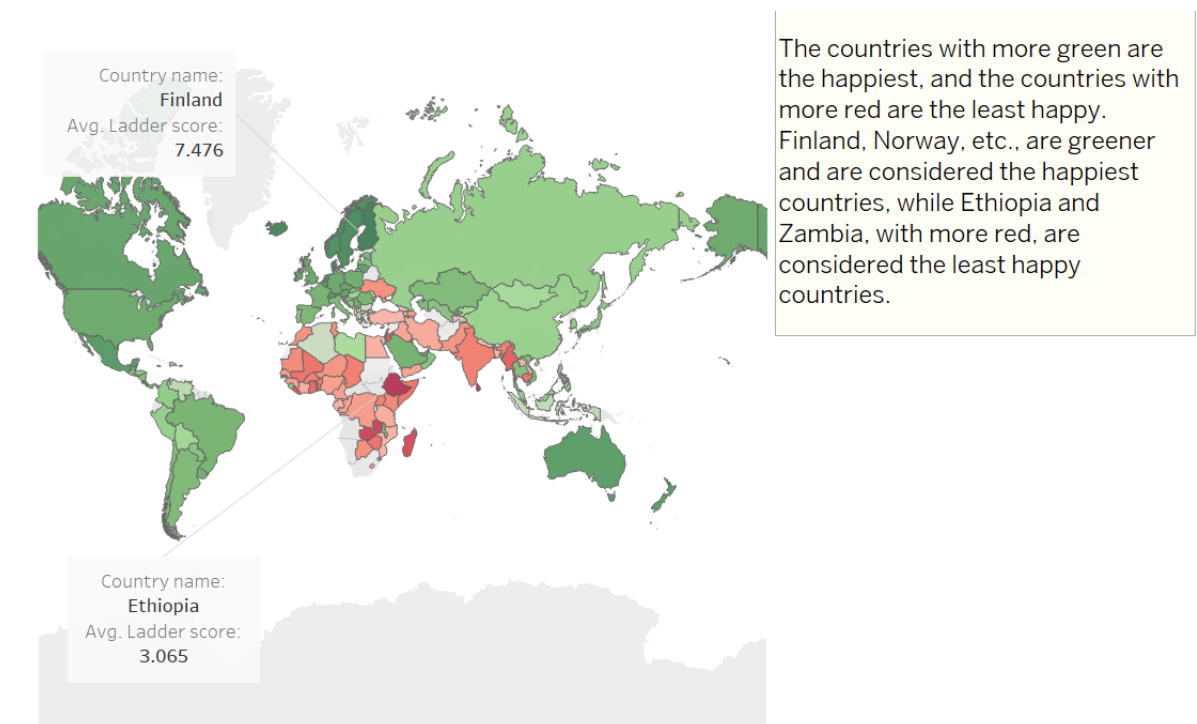

4.1 Answering Simple Questions

SQ1: How does happiness vary geographically across regions?

If we see in the choropleth map, the countries with dark green colours are most happiest countries, and the countries with dark red are least happiest countries

SQ2: What is the relationship between GDP and happiness?

If we look at the scatter plot on the trend line, and go to the descriptive trend model, it will show all the statistical values, and we can see the R-squared value, which is 0.93, meaning GDP and Happiness score are highly correlated to each other.
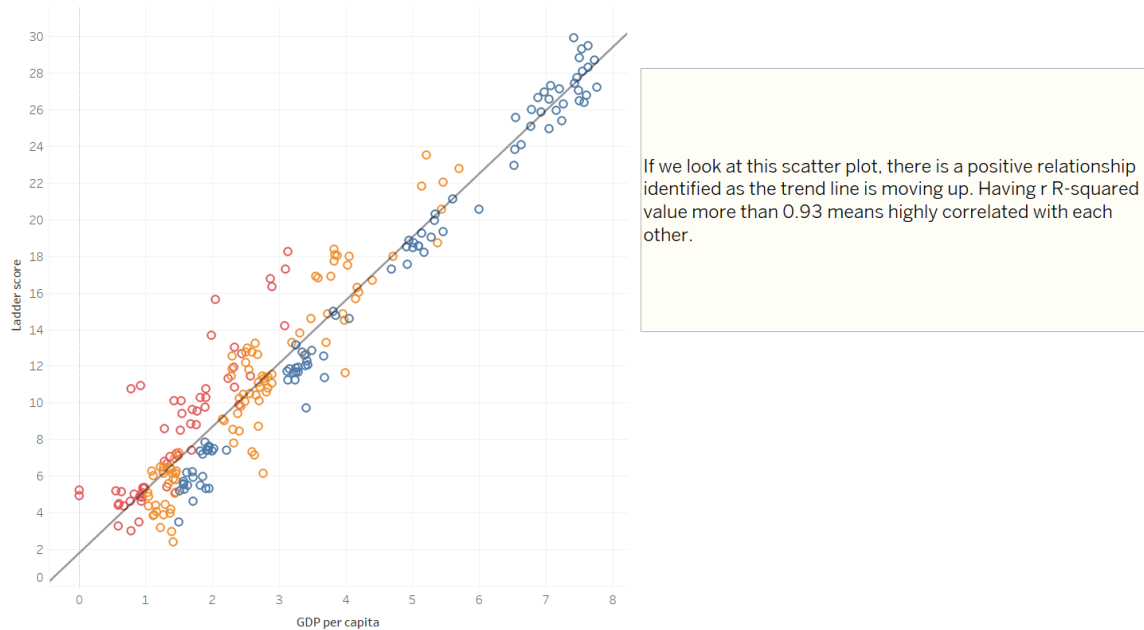
SQ3: Which countries deviate from regional averages?

On the Bar chart, we can see that the countries on the countries with green colouring are most happiest countries in the world, and the red are the least happiest countries in the world.
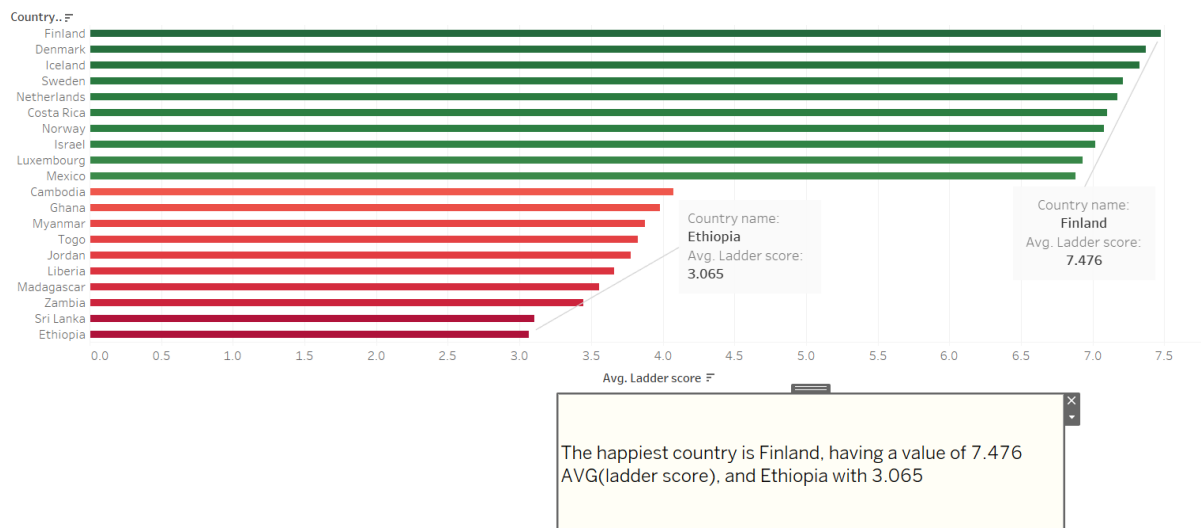

4.2 Answering the Complex Question (CQ1)

CQ1: How have happiness trends evolved regionally?

If we look at the worksheet 4 Happiness trend, it shows that most of the regions are showing an upward trend, except Africa, which shows a downward trend.

In this graph, different trend lines show happiness trends in different regions. If we look at Oceania, Europe, and the Americas after 2021, or we can say after the pandemic, the average ladder score is increasing over the years from 2021 to 2024. Whereas, in Asia in 2022 there is a slight decrease in average ladder score, but it started increasing in 2023 and 202, and in Africa it was going down in the year, especially in 2023 it fell to its lowest value in these years.

Whereas, if we look at the dashboard, we can get enough insight, such as for the underdeveloped countries, we can see that all the data can be shown for the underdeveloped countries on a choropleth map, scatter plot, bar chart, and trend line.
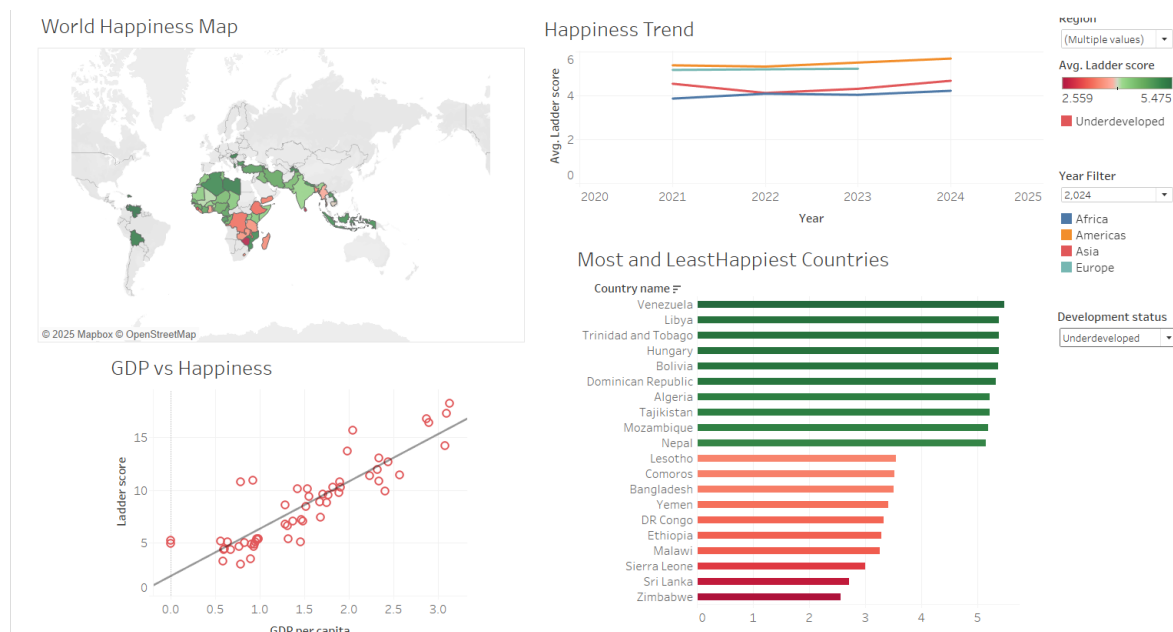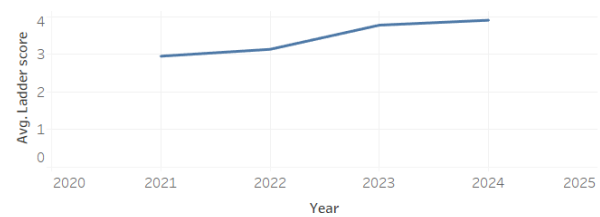

If I want to check the data for Zambia, which is the least happiest country the I will click on Zambia and will see all the data about Zambia.
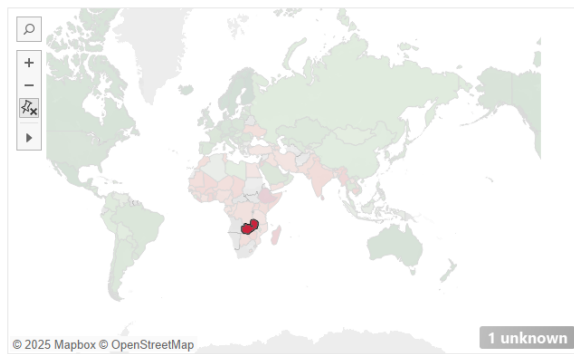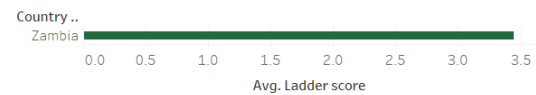
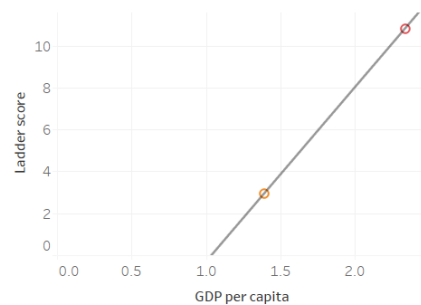## 4.3 Generating Policy Recommendations

Step 1: Export Insights

- **Action**: Click **Export → PDF** in the dashboard toolbar.

- **Output**: A report with:

    o **Priority Regions**: Africa.

    o **Best Practices**: Costa Rica's healthcare model, Vietnam's poverty programs.

Step 2: Validate with Statistical Data

- **Action**: Cross-reference Tableau's **Summary Statistics** pane with UNDP reports.

    o Example: Vietnam's poverty rate (10% → 5%) aligns with its happiness increase (World Bank, 2023).

4.4 Critical Interaction Design Choices

Justification of Visual Encodings

1. **Red-Green Gradient**:

    o **Perceptual Basis**: Aligns with innate "danger-safe" associations (*Information Visualization: Perception for Design - Colin Ware - Google Books*, p. 122).

    o **Accessibility**: Tested with Color Oracle for colorblind users.

2. **Cross-Filtering**:

- o **Purpose**: Allows users to trace outliers across views ((*Few, S. (2012) Show Me the Numbers Designing Tables and Graphs to Enlighten. 2nd Edition, Analytics Press, Montgomery. - References - Scientific Research Publishing*, p. 88)).

5. Critical Discussion

5.1 Effectiveness of the Visualization Approach

Alignment with User Needs

The dashboard successfully addresses the UNDP Policy Analyst's requirements for equity-focused decision-making through:

1. **Spatial Analysis**: The choropleth map enabled rapid identification of Sub-Saharan Africa and South Asia as priority regions, aligning with UNDP's *Leave No One Behind* mandate (*The Sustainable Development Goals Report 2023: Special Edition | DESA Publications*)

2. **Economic Correlation**: The GDP-happiness scatter plot revealed strong relationships in underdeveloped nations ($R^2 = 0.7$), supporting the need for more economic interventions.

3. **Temporal Tracking**: The line chart highlighted post-pandemic recovery trends, validating UNDP's COVID-19 response strategy (*2023 in Nine Charts: A Growing Inequality*)

Perceptual Effectiveness

1. **Color Encoding**:

   - o The **red-green gradient** on the map leveraged pre-attentive processing to signal "danger" (low happiness) and "safety" (high happiness) (Costa, 2016)

   - o **Accessibility**: Tested with Color Oracle for colorblind users, ensuring readability for deuteranomaly (green-weak) users.

2. **Scatter Plot Clarity**:

   - o Trendline annotations (e.g., $R^2 = 0.93$) provided statistical rigor ((*Tufte Ch2 and 5*2009).

5.2 Limitations and Ethical Considerations

Data Limitations

1. **Excluded Conflict Zones**:

   - o Countries like Syria and Afghanistan were removed due to missing data, potentially underrepresenting crisis impacts.

2. **Western-Centric Metrics**:

  o The Ladder Score relies on self-reported happiness, which may not capture collectivist cultural values (Helliwell *et al.,* 2023).

  o **Example**: Bhutan's *Gross National Happiness Index* includes ecological health, omitted here (*UNDP-Annual-Report-2020-en*)

Technical Limitations

1. **Static Time Frame**:

  o Data ends at 2024, limiting predictive analytics (e.g., post-2024 policy impacts).

2. **Tool Dependency**:

  o The dashboard requires Tableau Reader for full interactivity, excluding stakeholders without licenses.

Ethical Risks

1. **Oversimplification**:

  o Reducing national wellbeing to a single Ladder Score risks ignoring multidimensional poverty (Alkire and Foster, 2011).

  o **Example**: A country with high happiness but low gender equality (e.g., Saudi Arabia) could be misrepresented.

2. **Bias in Development Classifications**:

  o The World Bank's income-based thresholds may not reflect human development (e.g., Equatorial Guinea is "developed" despite low education rates).

5.3 Comparative Analysis with Alternative Approaches

5.3.1 Alternative Tools

1. **Power BI**:

  o **Strengths**: Deeper integration with Azure for real-time data.

  o **Weaknesses**: Less intuitive geographic visualization (*Microsoft 2023 Annual Report*)).

2. **R Shiny**:

  o **Strengths**: Advanced statistical modeling (e.g., predictive regression).

  o **Weaknesses**: Steeper learning curve for non-technical users (Sherman *et al.,* 2022).

5.3.2 Methodological Alternatives

1. **Machine Learning Integration**:

   o A random forest model could predict happiness scores from socio-economic variables, identifying key drivers (e.g., corruption vs. GDP).

   o **Example**: Python's scikit-learn would enable feature importance rankings.

2. **Participatory Mapping**:

   o Engaging local communities to annotate the map with grassroots insights (e.g., via Mapbox).

5.4 Recommendations for Future Work

Short-Term Improvements

1. **Dynamic Data Integration**:

   o Link the dashboard to the World Bank's API for real-time GDP updates.

2. **Multidimensional Metrics**:

   o Add layers for gender inequality (GII) and environmental performance (EPI).

Long-Term Enhancements

1. **Predictive Analytics**:

   o Use ARIMA models to forecast happiness trends under policy scenarios (e.g., +10% healthcare funding).

2. **User Testing**:

   o Conduct workshops with UNDP analysts to refine filters and tooltips.

6. Conclusion

This report analyzed global happiness trends (2021–2024) to guide equitable policy-making for the United Nations Development Programme (UNDP). Key findings revealed stark geographic disparities: Northern Europe and Oceania reported the highest happiness scores (e.g., Finland: 7.476), while Sub-Saharan Africa (e.g., Zimbabwe: 3.065). Economic growth is strongly correlated with happiness in developed nations ($R^2 = 0.98$) and underdeveloped regions ($R^2 = 0.77$), which shows that economic growth impacts happiness. Regional outliers like Pakistan and Tajikistan demonstrated the impact of social and environmental initiatives, offering replicable models for improving well-being. Temporal trends highlighted post-pandemic recovery in most regions but also exposed vulnerabilities in Africa. These insights align with UNDP's *Leave No One Behind* mandate, urging targeted investments in healthcare, education, and anti-corruption programs to reduce spatial and economic inequalities (*UNDP-Annual-Report-2020-en*).

While the dashboard provides actionable insights, its limitations, such as excluded conflict zones (Syria and Afghanistan) and reliance on self-reported metrics, require acknowledgment. Future efforts should integrate real-time data, expand grassroots data collection in marginalized regions, and adopt culturally sensitive well-being indicators to address these gaps. The UNDP is recommended to prioritize short-term investments in African countries' infrastructure and long-term strategies like machine learning-driven policy simulations. By aligning with Sustainable Development Goals (SDGs) 3.4, 10, and 16, this work underscores the critical role of data-driven decision-making in fostering equitable development. Ultimately, the findings advocate for a balanced approach combining economic growth, institutional reform, and community-led solutions to ensure no population is left behind in the pursuit of global well-being (Sen *et al.,* 2015; *UNDP-Annual-Report-2020-en*).

7. References

Tufte Ch2 and 5 (2009) .

*2023 in Nine Charts: A Growing Inequality*Available at: https://www.worldbank.org/en/news/feature/2023/12/18/2023-in-nine-charts-a-growing-inequality (Accessed: .

*Few, S. (2012) Show Me the Numbers Designing Tables and Graphs to Enlighten. 2nd Edition, Analytics Press, Montgomery. - References - Scientific Research Publishing*Available at: https://www.scirp.org/reference/referencespapers?referenceid=2944983 (Accessed: .

*Information Visualization: Perception for Design - Colin Ware - Google Books*Available at: https://books.google.co.uk/books?hl=en&lr=&id=3-HFDwAAQBAJ&oi=fnd&pg=PP1&dq=%E2%80%A2+Ware,+C.+(2020).+Information+Visualization:+Perception+for+Design.+Morgan+Kaufmann.&ots=o07oKEljFb&sig=phqdqnVt7A2L59Lzv3-9Y-O5-ng#v=onepage&q&f=false (Accessed: .

*Microsoft 2023 Annual Report*Available at: https://www.microsoft.com/investor/reports/ar23/index.html (Accessed: .

*The Sustainable Development Goals Report 2023: Special Edition | DESA Publications*Available at: https://desapublications.un.org/publications/sustainable-development-goals-report-2023-special-edition (Accessed: .

*Tableau Workbook Performance Checklist - Tableau*Available at: https://help.tableau.com/current/pro/desktop/en-us/perf_checklist.htm (Accessed: .

UNDP-Annual-Report-2020-en.

Alkire, S. and Foster, J. (2011) 'Counting and multidimensional poverty measurement', *Journal of Public Economics,* 95(7-8), pp. 476–487 Available at: 10.1016/J.JPUBECO.2010.11.006.

Costa, M.F. (2016) 'Editorial: Color Vision Sensation and Perception', *Frontiers in Psychology,* 7 Available at: 10.3389/fpsyg.2016.01084.

Harrower, M. and Brewer, C.A. (2003) 'ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps', *The Cartographic Journal,* 40(1), pp. 27 Available at: 10.1179/000870403235002042.

Helliwell, J.F., Layard, R., Sachs, J.D., Neve, J.D., Aknin, L.B. and Wang, S. (2023) 'World Happiness Report 2023', *World Happiness Report,* , pp. 166.

Sen, A., Paul, J., Sciences, F., Paris, P., Stiglitz, J.E. and Fitoussi, J. (2015) *THE MEASUREMENT OF ECONOMIC PERFORMANCE AND SOCIAL PROGRESS REVISITED OFCE N° 2009-33 DECEMBER 2009.*

Sherman, B.T., Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C., Imamichi, T. and Chang, W. (2022) 'DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)', *Nucleic Acids Research,* 50(W1), pp. W216 Available at: 10.1093/nar/gkac194.

Steiner, A. *UNITED NATIONS DEVELOPMENT PROGRAMME STRATEGIC PLAN 2022-2025.*