

# FIT5147 DATA EXPLORATION AND VISUALIZATION

Data Visualization Project Report



# In the Pursuit of Happiness

State of global happiness and the potential factors affecting it



#### **Tutor**

Farah Tasnuba Kabir Tutorial Number 04-P1

**Submitted by** 

Manali Prakash Choudhary 30151198

### Table of Contents

1.	. Introduction	3
2.	. Design	4
	□ Sheet 1 – Ideas	4
	□ Sheet 2 – Design	5
	□ Sheet 3 – Design	5
	□ Sheet 4 – Design	5
	☐ Sheet 5 - Realization	6
3.	. Implementation	7
	3.1 Data	7
	A) The World Happiness Report Statistical Data for year 2015, 2016, 2017, 2018 and 2019	97
	B) Data from Survey on Mental Health	7
	C) Data for Shape file – World	7
	3.2 Shiny Interface	
	3.3 Choropleth Maps and user interactions	8
	3.4 Other Plots and user interactions	8
	3.5 Data operations	9
4.	. User guide	10
	4.1 Tab 1 - Home page	10
	4.2 Tab 2 - Global State of Happiness	10
	4.3 Tab 3 - Happiness Factors	13
	4.4 Tab 4 - Mental Health Well Being	15
5.		
6.	. Bibliography	18
7.	. Appendix	19
	☐ Five Design Sheets	19
	Sheet 1	19
	Sheet 2	20
	Sheet 3	21
	Sheet 4	22
	Sheet 5	23

## Table of Figures

Figure 1. Home Page	. 10
Figure 2. Tab 2 - Global State of Happiness	. 11
Figure 3. Slider for selecting a year	
Figure 4. Choropleth map with tooltip	. 11
Figure 5. Zoom in functionality with tooltip	. 12
Figure 6. Tabs showing ranks of countries for the year selected	
Figure 7. Tab 3 – Happiness Factors	. 13
Figure 8. Dropdown for selecting Happiness factors	. 14
Figure 9. Click on country to see the Polar area chart	. 14
Figure 10. Functionalities of the streamgraph	. 15
Figure 11. Tab 4 - Mental health Wellbeing	. 15
Figure 12. Diverging bar plot interactions	

#### 1. Introduction

The world is facing a tough time recently, ever since the ending of 2019 due to the ongoing crisis like bush fires, earthquake, corona virus outbreak, riots and many more. All this has led to issues like unemployment, losses of lives, health issues, instability in the emotional and mental well-being. Clearly, 2020 has not been so great for the whole world until now and happiness has been an urgent need to make this world a happy place again. This gives us the opportunity to look back over the past few years and analyse what can be improved to kick start with a new beginning. This gives each one of us an opportunity to stop for a while, define and explore what happiness really means to us so that we can rectify our mistakes from the past. 'In the Pursuit of Happiness' allows each one of us to realize the state of global happiness until now based on different subjective well-being factors and also the importance of acknowledging the vital role of mental health well-being for a happy life(Hall & Helliwell, 2014). It builds a standard fair platform or measures and study of happiness factors in all countries around the globe. It, hence, tries to answer some of our questions regarding the perceptions of happiness based on the findings from the data exploration done in the previous stage of this project. For example,

- Nation wise happiness state of the globe. Considering the race, history, geography, and the development of nations.
- The role of Money/ Economy as a critical subjective well-being happiness factor along with the others.
- The importance of mental health well-being in relation to the top ranked happiest countries.

This narrative visualization can be the first step for the general public, social activists, policy makers, communities, governments to come up with a solution for a change in the society and make it a better place.

#### 2. Design

Over the past years, the importance of understanding the end user experience on a product or design outcome has been give more importance. The portrayal of the findings to the end user can influence the design and the tools used for it. A well understood visualization which portrays the correct information it aims to is considered as a successful design. A new design with proper guidance and instructions can be well received by the end users on contrast to an unclear traditional design (Dahl, Chattopadhyay & Gorn, 2001). Hence a design process is much needed to evaluate and refine the design to make it user centric. A design process builds systematic steps which allows us to implement the planning, refining, and designing of a product to achieve its maximum efficiency. Hence, the more we refine our design, less are the issues occurring during and post implementation of it. Prototyping is one of the techniques where we can focus on our conceptual designs as an attempt to refine it which also gives us a deep idea of the end product before even actually implementing it (Rettig, 1994). The Five Design Sheet Methodology is one such technique which has been used for 5 years now for web designing. The Five Design Sheet (FDS) is one of the crucial components of designing involving the clients, brainstorming ideas, design creation and continuous refining of the design as a cycle (Roberts, 2011).

The Five Design Sheet includes five sheets, each one with a specific purpose. Also, each sheet is divided into five parts. The sheets referenced henceforward are included in the appendix section of the report.

#### • Sheet 1 – Ideas

This sheet is also called the Ideas sheet. This sheet encourages the user to accumulate as many ideas as possible regarding the design after finalizing the topics focused and the audience of the design. We can then filter the ideas according to our data, user experience or duplicate ideas. We then categorize the ideas based on the topics or types of visualization and combine them to form three stand-alone designs which can satisfy all our questions or focused topics. The five divisions of this sheets are as follows,

- 1. Ideas
- 2. Filter
- 3. Categorize
- 4. Combine/Refine
- 5. Question

According to the above methodology, for our visualization project, we have focused our audience as general public, societies, communities, policy makers and questions as discussed in the introduction. In our sheet 1 we have analysed our data and came up with plots like, bar plot, choropleth map, heat map, bubble plot, diverging bar plot, polar area chart, lollipop chart, area chart, table and line/ connected scatter plot and use of shapes for distinguishing categories.

We have further filtered these plots and avoided some visualizations like the heat map, different shapes, table, area chart based on the user experience, readability, and its intended use of data type.

Further, the selected plots were categorized according to its data type, like, spatial, categorical, ranks, trends over time and correlational.

In the part combine and refine, these visualizations are grouped together forming different designs as shown in the sheet 1 which will then be implemented in sheets 2,3 and 4.

#### • Sheet 2 – Design

In the sheet 2, 3 and 4 we design the possible design results or how the proposed product would be like. The layout is an overall layout of the application that we have designed. The focus and zoom includes the plots or visualizations which will be available on user interactions or drill down. The info includes details like the title, author, and the date the sheet was designed. The operations describe the operations of the applications or the visualizations available on user interactions. In the discussion, we talk about the positive and the negative sides of the visualizations created in the design based on the user experience and effective narration of the story. We include the following divisions,

- Layout
- Focus/zoom
- Info
- Operations
- Discussion

In our second sheet, we create the design using choropleth map for showing happiness ranks of the nations, a bar plot for the happiness factors when the user clicks on the country on the map. A streamgraph can be used to show the trend of the factors over the years 2015-19. A scatterplot is used to show the relation between the high ranked countries and the Mental health support percent from the employers of those countries. In the discussion, scatterplot is realized to be a difficult to decode plot of the users along with some other thoughts.

#### • Sheet 3 – Design

Moving on to the second design, here we used the famous bubble plot to show the trend of the countries for happiness rank over the years. We have also introduced the idea of implementing a dropdown for changing the x-axis of the bubble plot on user input. Also, the happiness factors are visualized using the polar area chart, which is the improvised version of the radical charts by keeping the angles equal. This chart also plots the mental health support percent as an additional factor along with the happiness factors. We have used the lollipop charts which are a modern version of the bar plots ideally as used for rankings.

The positive and the negatives of the design are discussed in the discussion section of the sheet. Though the sheet answers all the questions, it fails to clearly show the relation of the happiest countries with high mental support percent to the user. Also, the bubble chart makes it difficult for the user to locate a country and takes much effort from the user to decode the summary the plot it aims to show. The lollipop chart and the polar area chart on the other hand is a modern, easy to understand, creative and, effective way of conveying the story the data wants to tell the user.

#### • Sheet 4 – Design

Our last design also tries to answer all the potential questions and is the third design sheet. It plots the happiness rank around the globe using the circular bar plot which has the world map as a globe in the centre. The user can move the globe round to find his/her country and its respective rank. Also, the user can view the ranks of the countries using the colors based on the regions they belong to. The user can

click on the country in the dropdown to plot the bar + connected scatterplot for the happiness factors. The bar and scatterplot together can help the user to understand the trend of the factors.

Here, we try to show the relation of the happiest countries and the mental health support percent using a diverging bar plot as shown in the figure. The right and the left sides of the axis have plots differentiated by colour. The positive side of this layout is the comparison of two categories using the diverging plot. As it clearly defines the categories and shows a quick comparison of in a glance, even without looking at the tooltips. The choropleth map definitely proves easy to understand as compared to the circular bar plot and globe map. Plotting the countries on such a small map is inconvenient and finding them in the headings of the bar plot is also inconvenient.

The colour differentiation of the lollipop plots based on the regions can be misleading especially for the ranks. The regions can repeat in the list; hence the user might be confused on how to come up to a conclusion. The bar plot with connected scatter plot for trends is difficult to decode. The user might confuse it to be two different plots.

#### • Sheet 5 - Realization

The sheet 5 is the realization sheet which has the final design of the application. After great consideration of all the positives and the negatives of each design, we have taken plots from all three sheets into this final sheet as shown in the sheet 5. The important criteria followed here was the user experience. The idea of the design is to create a narrative visualization where any layman, without any knowledge of the visualizations should be involved with the story trying to be narrated by the application and understand it in its real sense with manipulating any plots.

The choropleth map is the traditional and the most effective way for displaying the change in values over time for all countries. It is easy to locate the countries as people are familiar with the geography of the world. Hence, we use it to show the happiness ranks of countries from 2015-19. The lollipop chart is eye catching and creative version of the bar plot. Hence, we use it for the top 10 and bottom 10 ranks of countries for happiness ranks. On the next tab, we can have a choropleth map for the happiness factors, where the country can be selected by clicking on the map. The polar chart of the country will then be visible plotting the relative distribution of the happiness factors. The user can check the distribution of the factors over the years using the streamgraph. Hence, the user can select the countries from the dropdown where the countries will be ordered alphabetically. For showing the relation between the happiest countries and the mental health support percent by the employers, the diverging bar plot is on point. On hovering, the tooltips showing detailed information will be accessible to the user for all the plots including the maps. The design shows the overall picture of the state of global happiness drilling down to the happiness factors, again for the whole world and further drilling down to the happiness factor distribution of each country. After analysing these subjective wellness factors, the visualization then further conveys the importance of the mental health support contributing to the happiness. Hence, the story should motivate the user to realize the missing and lacking attributes of the definition of happiness for their nation or for him too.

The details section of the sheet 5 contains the technology used to implement this. i.e. R shiny. Also, this project is estimated to take 80 hours of effort to implement the design. The respective cost and the schedule are also mentioned.

#### 3. Implementation

#### 3.1 Data

# A) The World Happiness Report Statistical Data for year 2015, 2016, 2017, 2018 and 2019

The World Happiness Report is a publication of the Sustainable Development Solutions Network, powered by data from the Gallup World Poll for around 157 countries and is released by United Nations (UN) every year since 2012, with the aim of helping the world governments and policy makers to enable such policies benefiting the quality of human development. This data basically contains the happiness scores of different countries and their rankings, based on the answers acquired to the main life evaluation questions on subjective wellbeing raised in the poll. The happiness score ranges from 0-10 and estimates the dependencies of the score on the six factors that potentially contributes to the happiness index of the countries. These scores are evaluated in comparison to the scores of an imaginary country "Dystopia", that has world's least-happy people or least possible scores for those six factors. The idea of having "Dystopia" is to have standard set against all countries for favorable comparison and positive values for each key factor. The links for the three data sources for year 2015, 2016, 2017, 2018 and 2019 are shared in this report. The data from all the years is combined into one single data and used for the visualization.

#### B) Data from Survey on Mental Health

Data from Survey on Mental Health that measures the attitude of mental health and frequency of mental health disorders in the Tech Workplace in 2014-15 from all over the globe, acquired from Open Sourcing Mental Illness under CC BY-SA 4.0. This dataset is based on the responses acquired from the respondents where each column states each question asked regarding the workplace attitude towards mental health and frequency of mental health disorders. The questions demand a column showing the percentage of categorical responses from the total responses from the country for all the countries. Hence a column named 'Percent' is calculated and added in this dataset and this dataset is later merged with the previous one.

#### C) Data for Shape file – World

Data for mapping of the world has been downloaded as shape file of the world from ThematicMapping.org as guided from the R-graph-gallery for mapping choropleth maps.

#### 3.2 Shiny Interface

The implementation of this design is done using R shiny. R has few very powerful packages which are ideal for data visualizations. *Shiny* is a package in R which allows interactive web applications with visualizations. As the application demands high user interaction, Shiny provides open source, powerful and sophisticated framework for building applications with R. The shiny has few in-built themes implemented through the *shinythemes* package. This package allows to set the overall appearance of the application which contains inbuilt CSS to use with Shiny instead of the traditional bootstrap. The application is divided into four tabs,

- 1. Home page
- 2. Global State of Happiness
- 3. Happiness factors
- 4. Mental Health Wellbeing

#### 3.3 Choropleth Maps and user interactions

The choropleth maps are implemented using <u>Leaflet</u>. Leaflet is a very famous and powerful open source JavaScript library for interactive maps. This package in r is useful for integrating the leaflet maps in R. We use leaflet as it allows to add layers to the maps e.g. addMarkers, addTiles, addPolygons etc. For the choropleth map, we will need the layers, addTiles and addPolygons. Also, it allows interactions like zooming, mouse events, rendering of spatial objects for longitude and latitude and most important nonspherical maps.

In order to plot the world map, we use the shape file data of the world with countries and its respective latitudes and longitudes. To use this data further, we convert this data frame into spatial object using the <u>sf: simple features</u> package. This package allows to encode spatial vector data. Hence, now we can use the data from shape file with leaflet for plotting the maps.

We have created bins and used colorbins function for coloring the map which slices the input domain with values. It uses the numeric data to map the output colors fixed to it. We have opted to create separate bins for different factors as the data of these factors have behaved differently from each other and have showed variation in the trends. Hence, keeping a standard bin for all the categories is unfair and would depict a wrong story. Bins and scaling play a very important role in choropleth as improper bins can showcase a completely different story in contrast to the truth. Choropleth map has been a topic for showing misleading information or incorrect results which we should be specific and careful about (Choosing Map Bins, 2017). Hence, for the same reason, we have focused on keeping the bin intervals of the maps based on the data classification by trends seen during the exploration and not a standardized one (Equal interval bins). Keeping in mind the golden rules of user experience for choropleth map, the bin size has not been any more than 5 bins, so that the variation in the colors is pretty evident and easily understood by the user.

We have added the zoom functionality for the map on single click so that the user gets a closer view of the country selected. On single click also allows selection of the country on map in tab 3. The addTiles allows us to customize the base tiles of the map. Here we have used the provider 'Wikimedia' as the map is simple with countries as white tiles which allows the user to recognize in the choropleth map, if the data is unavailable for that particular country.

We have played with different color palettes like <u>viridis</u>, <u>RColorBrewer</u>, and theme library <u>hrbrthemes</u> for some modern and crisp quality visualizations which enhance the appearance and quality of the visualizations along with adding some utilities.

#### 3.4 Other Plots and user interactions

Plots like Polar Area Chart, Diverging bar plot, Lollipop plot have been implemented using the basic ggplot. <u>Ggplot2</u> is again another powerful package for crating complex plots through simple means using the plot aesthetics. Though the plots are modern, the method of plotting them is basic. These plots are rendered through <u>plotly</u> to make them interactive. Just using the <u>plotly</u> package for rendering the basic plot outputs through plotly allows user interactive graphics with high graphic quality. Polar Area chart has been advised to be kept simple with a smaller number of categories for its effective use. As this best suit our implementation, we chose to go with it and link the country selection functionality by on-click event with it. We have hence, used this chart to show the relative distribution of subjective wellbeing happiness factors of the selected country.

Package *ggiraph* allows to create dynamic ggplot. We can add tooltips and other interactive actions on the ggplot graphs using this package. Here we have used this library to add actions to such elements which cannot be rendered using plotly.

We have plotted streamgraph in our application with the help of <u>devtools</u> to install packages from github. The <u>streamgraph</u> package is a html widget from Javascript/d3 library. As it is available live on github, we need to install it from there. The advantage of this package is that the output of this plot is interactive. Streamgraphs are famous and loved by the users, keeping in mind they should be easy to understand and clear. As we are plotting only 6 categories here, and looking at the trends, the use of streamgraphs is very effective in this case as also special care has been taken about the user interactions like filter dropdown and tooltips for this graph.

#### 3.5 Data operations

The <u>dplyr</u> package is a very important package in R for faster data modification and manipulation. It provides functions like filter, select, mutate etc. which allows quick operation on data to make it ready to be plotted. We have used it for the same purpose in our application as well.

#### 4. User guide

The application is divided into four tabs,

- 1. Home page
- 2. Global State of Happiness
- 3. Happiness factors
- 4. Mental Health Wellbeing

#### 4.1 Tab 1 - Home page

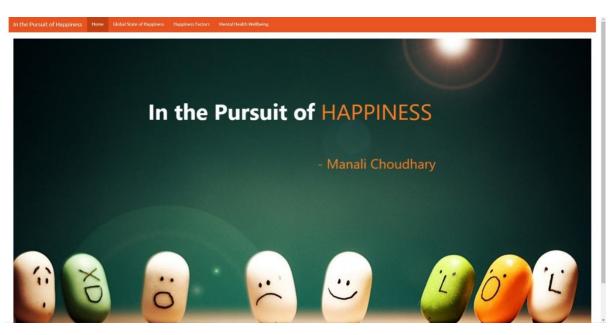


Figure 1. Home Page

The Application starts with a home page. The Home page invites the user with a title relevant to the topic and a picture where the user can relate to about the topic- "In the Pursuits of Happiness"

The user can click on any 4 of the tabs in the menu bar to navigate further. Let's start from the second Tab.

#### 4.2 Tab 2 - Global State of Happiness

The Tab 2 gives an overall picture of the state of happiness around the globe since 2015 as seen in the picture. It also shows the rankings of the top 10 and bottom 10 countries for that year. The map shows a choropleth map which can be zoomed. The colors are based on the happiness scores of the countries.

Narratives are provided to guide the user through the page. The page loads as following, and the year selected is 2015 by default.



Figure 2. Tab 2 - Global State of Happiness

The user can select any year by moving the slider through the years as shown.

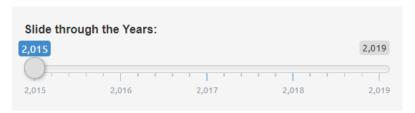


Figure 3. Slider for selecting a year

The choropleth map will update as the slider moves through the years. On hover on the country, the map will show tooltips as follows.

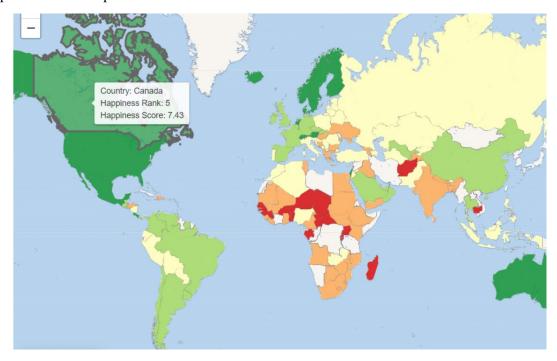


Figure 4. Choropleth map with tooltip

If the user further clicks on the map it will zoom in for that country as shown in the below picture.



Figure 5. Zoom in functionality with tooltip

As the year changes, the country rankings change for that year as well.

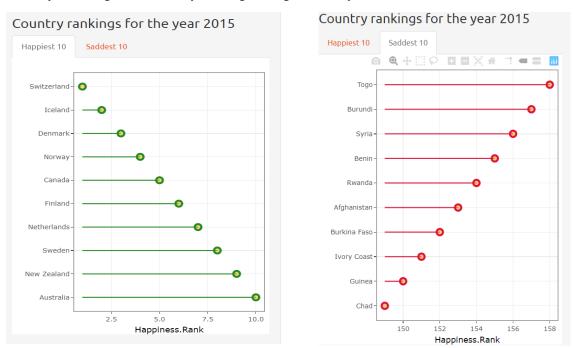


Figure 6. Tabs showing ranks of countries for the year selected

In the sidebar panel, the user can click on the tabs- Happiest 10 (Green) and Saddest 10 (Red) to check the rankings. Appropriate tooltips are added to them as well.

#### 4.3 Tab 3 - Happiness Factors

The third tab shows the Happiness factors. A narrative is shown explaining what we mean by the happiness factors. This page consists of a choropleth map for the distribution of the factors, a polar area chart and a streamgraph as shown if the below pictures.



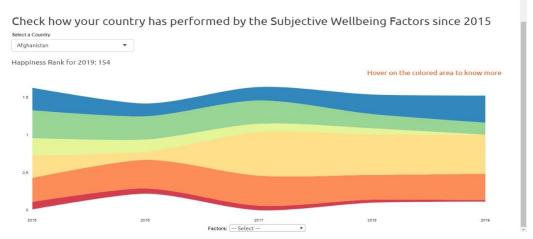


Figure 7. Tab 3 – Happiness Factors

The user can select the Happiness factors from the dropdown and the choropleth map will update as follows. Also, "Economy" is the default factor chosen when the user first navigates to this page.

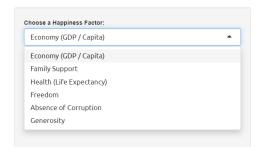


Figure 8. Dropdown for selecting Happiness factors

The choropleth map shows the similar zoom in and tooltip functionality as the map on previous tab. When we click on the country, it zooms in to that country, showing the tooltip and also the polar area chart with appropriate legend and scale in the side bar panel for that country as shown in the picture.

The user can further change the factors are select countries to visualize the updated plots.

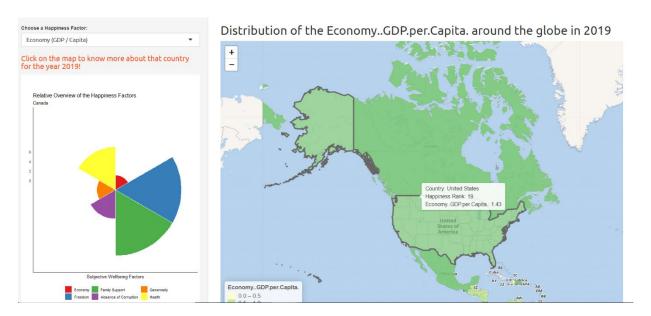


Figure 9. Click on country to see the Polar area chart

The user can further scroll down the page to check the streamgraph according to the country for years 2015-19. The user can select the country by a dropdown of countries as shown below. The streamgraph is updated as the country is selected. The happiness rank of the country is also displayed. On hover on any part of the streamgraph, a tooltip showing the value is displayed on the top left of the graph.

A filter is also provided as seen in the below image, where the user can select the factor which will then be highlighted in the streamgraph.

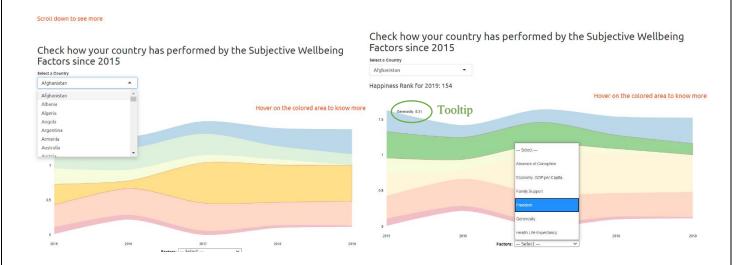


Figure 10. Functionalities of the streamgraph

#### 4.4 Tab 4 - Mental Health Well Being

The last tab talks about the relation of mental health well being as a comparison with the happiest countries in the world. It is seen that the countries providing mental heath support by the employer are also the countries which rank high for the happiness rank. This relation is shown by the diverging bar chart as shown. The same is explained in a narrative in this page.

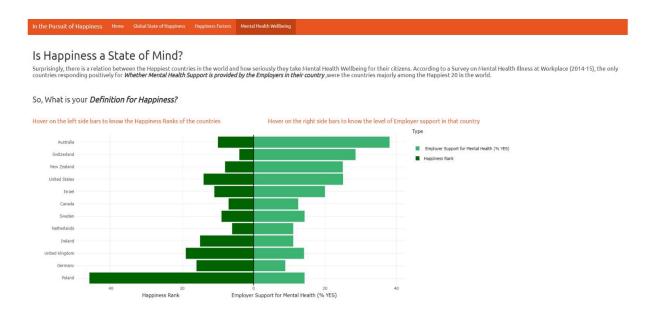


Figure 11. Tab 4 - Mental health Wellbeing

The user can click on the legend to check the visualization for that topic and appropriate tooltips are provided as well as shown in the figure.

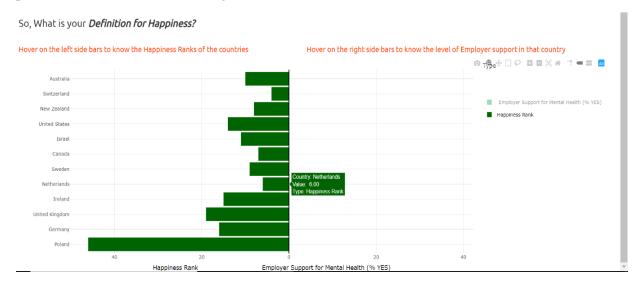


Figure 12. Diverging bar plot interactions

#### 5. Conclusion

The topic of Happiness is very broad and may mean different for different users. Hence, to get an application which can treat all users on the same page was challenging. Understanding the topic from user's perspective was a key here. Also, as some countries are happier than others, this might de motivate some users if we narrate the story in an improper way. Hence the narration of such a topic should not be biased. Also, motivating users and changing or moulding their perspective on subjective wellbeing factors and mental health well being was an important motive of this application which was successfully achieved. In this application, we have used modern visualizations and their use in the most advised sense to avoid the drawbacks. Hence, use of such visualizations in an informed and researched manner was a challenge for this project. Technically this project allowed us to explore the updated research and advancements in R. It has taught the power and importance of various libraries in R. Here we have also explored some html widgets, d3 packages and their use in R. The use of libraries has taught us how some libraries cannot be used with some few and has encouraged us to find the solutions. The implementation of the Five Design Sheet Methodology was an important stage in this project. It was informative to brainstorm and at the end come up with visualizations in the final design that one may have never thought of while starting the process. In the end thought the application involves visualizations, coding, and data, we have learned it is further about the story telling and the user experience all in one.

In the hindsight, exploring more and more about the factors contributing to happiness can be thought of. Further exploration in handling different types of data is also fascinating to play with coding different types of visualizations. As R supports developer tools, coding/inventing a visualization of our own from scratch can be a further step to convey this narrative in the most effective manner.

#### 6. Bibliography

Choosing Map Bins (2017, November). [Web blog post]. *PolicyViz.com*. Retrieved from <a href="https://policyviz.com/2017/11/02/choosing-map-bins/">https://policyviz.com/2017/11/02/choosing-map-bins/</a>

Dahl, D. W., Chattopadhyay, A., & Gorn, G. J. (2001). The importance of visualisation in concept design. Design Studies, 22(1), 5-26.

Hall, J., & Helliwell, J. F. (2014). Happiness and human development. UNDP Human Development Report: Occasional Paper. New York, USA: United Nations Development Programme

Helliwell, J., Layard, R., & Sachs, J. (2015). World Happiness Report 2015, Update (Vol. I). New York: Sustainable Development Solutions Network. Retrieved from <a href="https://worldhappiness.report/ed/2015/">https://worldhappiness.report/ed/2015/</a>

Helliwell, J., Layard, R., & Sachs, J. (2016). World Happiness Report 2016, Update (Vol. I). New York: Sustainable Development Solutions Network. Retrieved from <a href="https://worldhappiness.report/ed/2016/">https://worldhappiness.report/ed/2016/</a>

Helliwell, J., Layard, R., & Sachs, J. (2017). World Happiness Report 2017, Update (Vol. I). New York: Sustainable Development Solutions Network. Retrieved from <a href="https://worldhappiness.report/ed/2017/">https://worldhappiness.report/ed/2017/</a>

Helliwell, J., Layard, R., & Sachs, J. (2018). World Happiness Report 2018, Update (Vol. I). New York: Sustainable Development Solutions Network. Retrieved from <a href="https://worldhappiness.report/ed/2018/">https://worldhappiness.report/ed/2018/</a>

Helliwell, J., Layard, R., & Sachs, J. (2019). World Happiness Report 2019, Update (Vol. I). New York: Sustainable Development Solutions Network. Retrieved from <a href="https://worldhappiness.report/ed/2019/">https://worldhappiness.report/ed/2019/</a>

Osmihelp.org. 2020. Research :: Open Sourcing Mental Illness - Changing How We Talk About Mental Health In The Tech Community - Stronger Than Fear. [online] Retrieved from <a href="https://osmihelp.org/research">https://osmihelp.org/research</a>

Rettig, M. (1994). Prototyping for tiny fingers. Communications of the ACM, 37(4), 21-27.

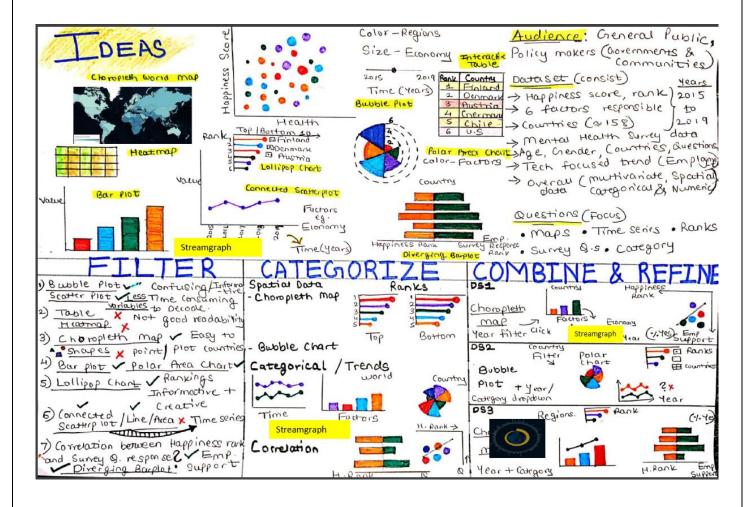
Roberts, J. C. (2011). The Five Design-Sheet (FdS) approach for Sketching Information Visualization Designs. In Eurographics (Education Papers) (pp. 29-36).

Thematic Mapping.org. 2009. World Borders Dataset. [online] Retrieved from <a href="http://thematicmapping.org/downloads/world\_borders.php">http://thematicmapping.org/downloads/world\_borders.php</a>

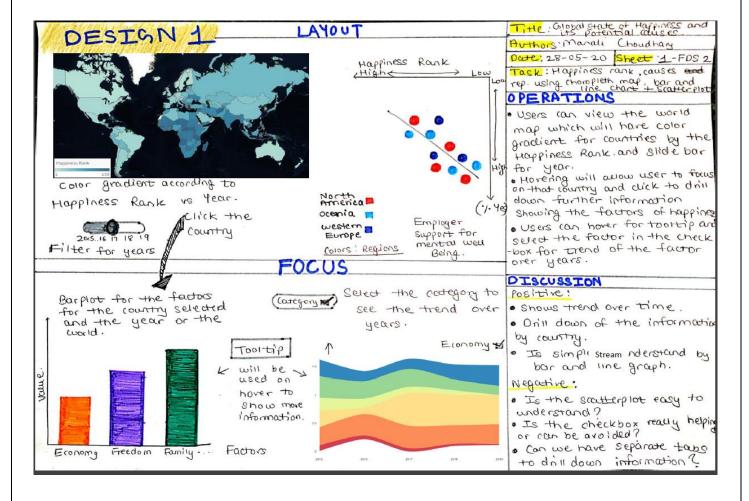
#### 7. Appendix

• Five Design Sheets

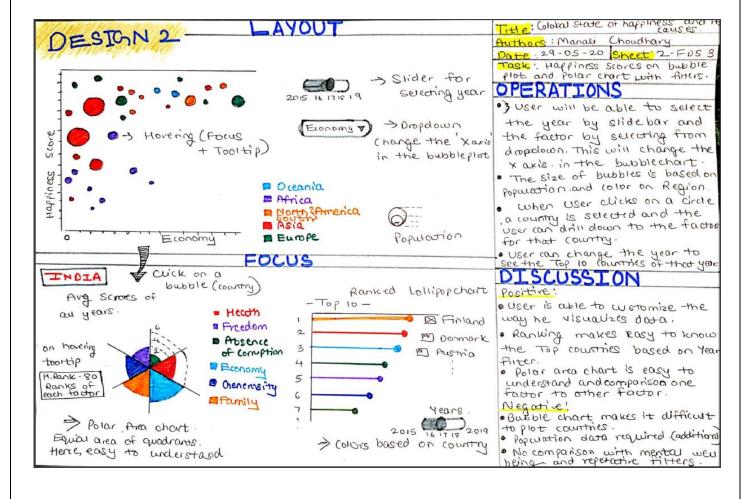
Sheet 1



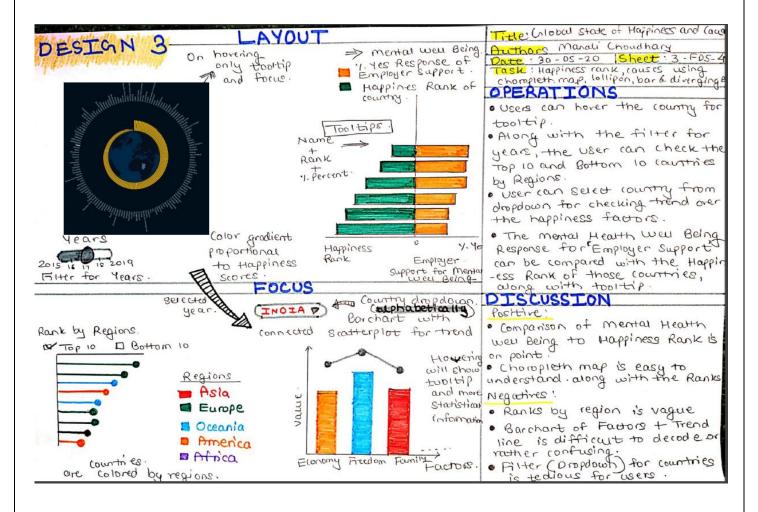
Sheet 2



Sheet 3



Sheet 4



Sheet 5

