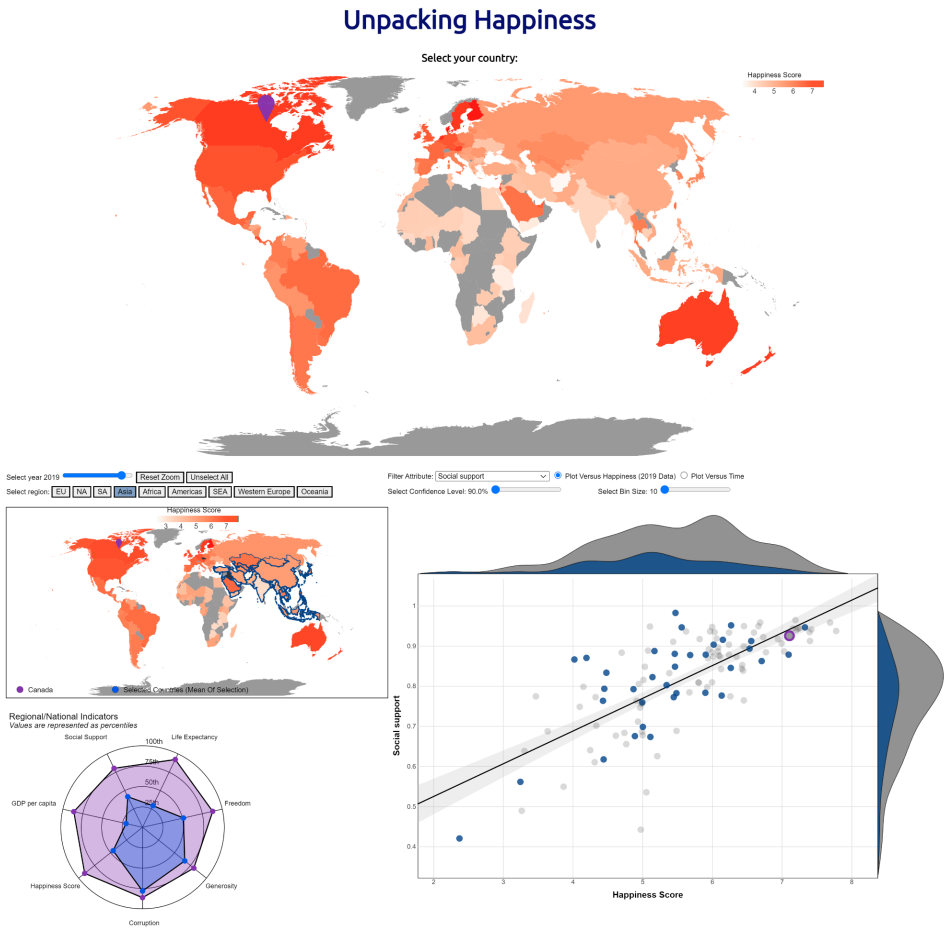
CPSC 436V Milestone 3

Unpacking Happiness

Group 14 Team Members: Anju Sethu (l1t1b), Himanshu Goyal (r7w1b), Kevin Cai (p4u9a)

Overview/Teaser



To date, we have used economic production (i.e. GDP) as the primary indicator of a country's well-being. The Covid-19 pandemic has demonstrated that GDP is not synonymous with the health of a nation. While economists and policymakers have experimented with using national happiness measures, they can be opaque since they solely provide a happiness score, as opposed to investigating the factors that lead to the score. To address this challenge, we propose building a data visualization that allows policymakers to visually explore the happiness levels of countries and regions around the world and how it is influenced by different key attributes (GDP per capita, life expectancy etc.). On a practical level, our visualization will help policymakers see gaps for improvement in their own country and allow economists to understand geographic and other patterns that contribute to national happiness. Beyond that, our visualization aims to get economists and policymakers thinking about what really matters and the indicators we use to measure national well-being. Our visualization will allow the user to select countries and regions and explore its happiness levels and the factors that contribute to a country's happiness level. Users can explore different aspects of this data by comparing their country of study to a basket of multiple countries and a region, examining changes over time (2011-2020) and filtering specific attributes to examine its impact on happiness levels.

Data and Preprocessing

We will be visualizing the World Happiness Index dataset from the World Happiness Report, which includes a happiness score for countries worldwide. Also included in this dataset are a mix of socioeconomic attributes that are hypothesized to be correlated with happiness. Our data is sourced from the most recent World Happiness Report from the organization's website (Appendix 1), which includes data up to 2020. We decided to focus on the most recent 10 years, spanning from 2011 to 2020. This decision was made in order to choose years for which most countries had data (pre 2011 data was sparse).

The dataset was incomplete, with some countries missing data for multiple years across the selected time range. We used a script (Appendix 4) to filter out countries missing data for more than 1 year across our selected time range, since requiring full data reduced the number of eligible countries by 40%. In order to best compare how countries ranked in terms of their quantitative attributes, we added 7 derived attributes corresponding to the percentile rankings of each country, listed in the table below (Appendix 5). Our processed dataset contains 17 attributes and 1149 rows of data for 119 countries from 2011 to 2020.

The following is a data abstraction based on our processed dataset:
Dataset type: Multi-dimensional table (keys: country + year)

Field name	**Attribute type**	**Cardinality/Range**
Country name	Categorical	C: 119
Country ID	Categorical	C: 119
Year	Ordered	C: 10
Happiness Score	Quantitative	R: [2.375, 7.889]
Logged GDP per capita	Quantitative	R: [6.885, 11.648]
Healthy Life Expectancy	Quantitative	R: [45.42, 75.2]
Social support	Quantitative	R: [0.42, 0.987]
Freedom to make life choices	Quantitative	R: [0.315, 0.985]
Generosity	Quantitative	R: [-0.335, 0.698]
Corruption Perception	Quantitative	R: [0.078, 0.977]
Happiness score percentile	Quantitative	R: [0, 100]
GDP per capita percentile	Quantitative	R: [0, 100]
Life expectancy percentile	Quantitative	R: [0, 100]
Social support percentile	Quantitative	R: [0, 100]
Freedom percentile	Quantitative	R: [0, 100]
Generosity percentile	Quantitative	R: [0, 100]
Corruption perception percentile	Quantitative	R: [0, 100]

We added a script to include ISO 3166-1 Numeric Codes to the dataset (Appendix 2) to be used as a primary key when merging other datasets (to account for country name differences). We wanted policy makers and economists to be able to easily select common predefined regions, such as continents and some geographic regions like South-East Asia and Western Europe. Therefore, we added an additional dataset to get the country's regional data (Appendix 3). This dataset contained 249 countries with their ISO-3166-1 country name, numeric code, continent, intermediate region and subregion.

The following is a data abstraction based of the regional dataset:
Dataset type: One-dimensional table

Field name	**Attribute type**	**Cardinality/Range**
Country name	Categorical	C: 249
ISO 3166-1 Numeric Code	Categorical	C: 249
Alpha-2 Code	Categorical	C: 249
Alpha-3 Code	Categorical	C: 249
Region	Categorical	C: 249
Sub-region	Categorical	C: 249
Intermediate region	Categorical	C: 249
Region code	Categorical	C: 249
Sub-region code	Categorical	C: 249
Intermediate region code	Categorical	C: 249

Link Appendix:
1 - Data source: https://happiness-report.s3.amazonaws.com/2021/DataForFigure2.1WHR2021C2.xls
2 - Numeric code script: https://pastebin.com/CQJUSWQB
3 - Regional data source:
https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv
4 - Filtering data script: https://pastebin.com/vLQP7UYi
5 - Finding percentiles script: https://pastebin.com/nVsteKY7

Goals and Tasks

The main goal of this project is to allow users to understand how happiness scores are distributed based on geographic location, understand which factors may contribute to those happiness scores, understand how strongly each attribute correlates to happiness, and see how they change over time.

We have economists and policymakers in mind as our target users. An economist may be interested in the following: They want to be able to [explore] a dataset in order to [compare] the effect of different variables on happiness levels in order to [discover] the impact of a specific attribute, e.g. GDP on happiness and [identify] the most relevant variables. The goal of this user may be to understand which factors are correlated with high happiness or come up with a formula to predict happiness levels for a country/in a certain period of time.

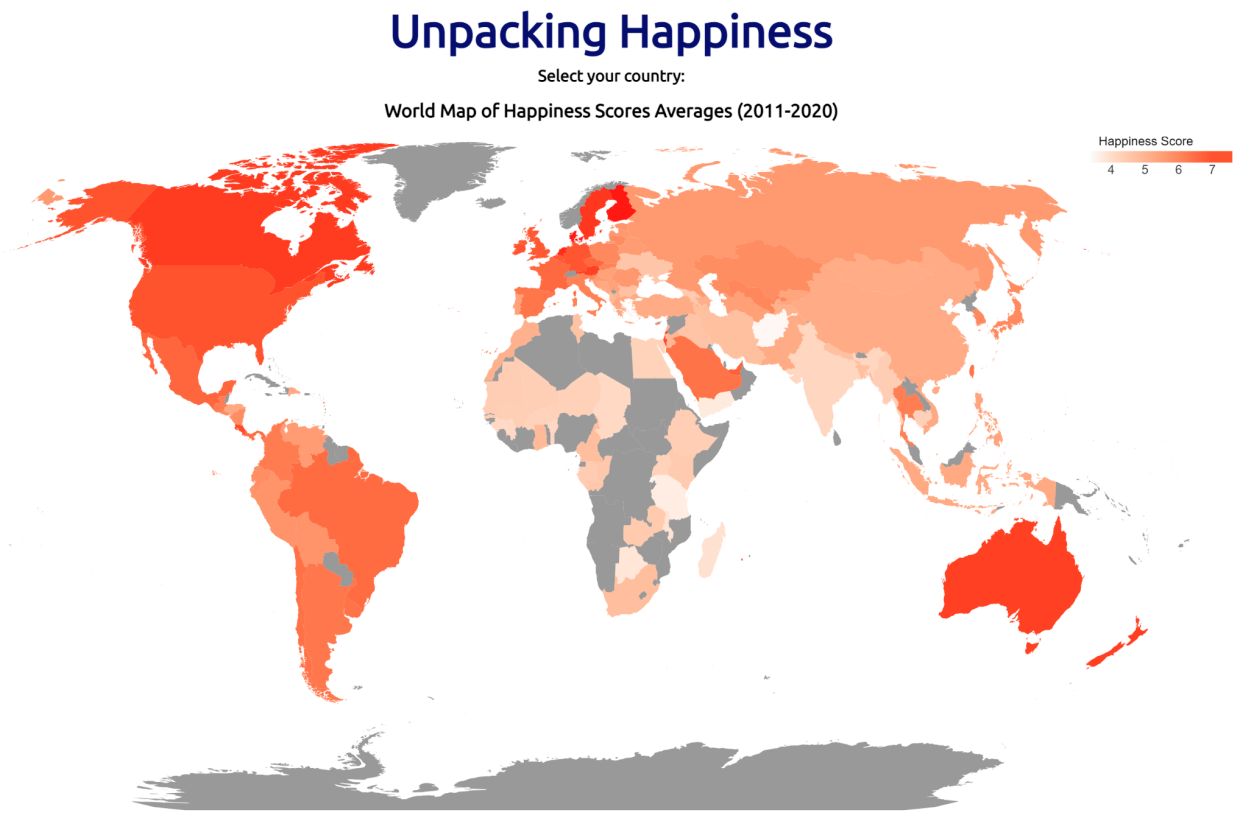
A policymaker may be interested in the following tasks: They want to be able to [query] happiness levels of their country and [compare] it with other similar/better countries to [identify] attributes in which their country can improve. The goal of this user may be to work on policies with others in government and present the findings to other policymakers or the public during election time.

Here are some of the tasks our visualization will help the user accomplish (using the {Action, Target} format):

- {Lookup, values}
    - For example, the user can look up the happiness and social/economic scores for specific countries at specific years.
- {Discover, distribution}
    - For example, the user can see that countries that have high GDP per capita generally have high happiness scores and vice versa.
- {Compare, trends}
    - For example, the user can see that happiness levels for African countries are increasing over time, which is the opposite of EU countries.
- {Locate, outliers}
    - For example, the user can look up countries with a high GDP but low happiness scores.
- {Compare, similarity}
    - For example, the user can look at countries that may have similar happiness scores to see which other attributes they are similar in and which attributes they differ in.

Visualization

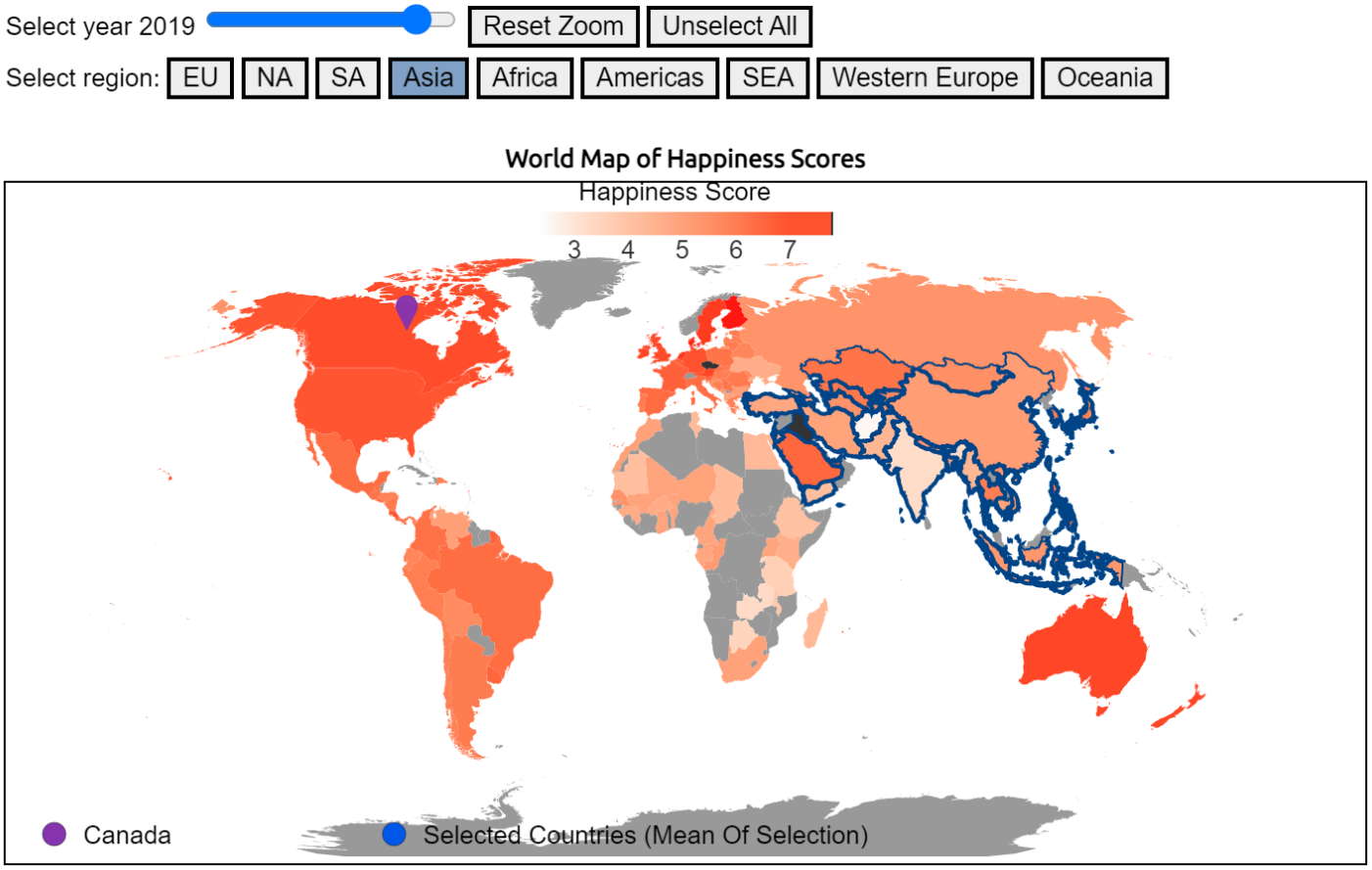
Choropleth Map:



On the first screen, our visualization opens up to a choropleth map of the average world happiness from 2011 to 2020. Our choropleth map allows economists and policy makers to identify trends and outliers related to the geographic distribution of happiness indices. We thought that policy makers and economists would most likely want to identify trends in certain regions, like comparing their region to other regions and the change of their region over time, so we felt a map would be most suitable for the task. We used a choropleth map as opposed to other maps because of user familiarity. Since we are color coding each country, it may be argued that a grid cartogram is more effective due to the difference in sizes of each country. However, we expect our target audience to be most familiar with ‘regular’ maps in the commonly used Earth projection. We wanted to retain this and leverage the familiarity, particularly since screenshots from the visualization may be used in reporting. As mentioned, we used luminance to encode the happiness score. We used a sequential color scheme because there was a clear start to end in happiness score with no meaningful middle point. It is also intuitive and the graduating warm tones were appropriate to represent happiness. There is also a legend to depict the color scheme and its related score.

The visualization prompts the user to select a country. We felt this was an important feature because most policymakers are representatives of a single country or most economists may be affiliated with a single country/have a focus on a single country for their study. This would give the user a starting point and act as a primary country so that all the other countries they select for comparison can be aggregated/bucketed against this country. If we just had a small map as on the second view, we noticed that it is easy to unclick the primary country by mistake. Most users will always want their primary country to be selected and just change the other country/region selections. This visual encoding is also a great starting point to tell the user what they will be looking at. For this reason, we encoded the average happiness score for all the years in our visualization (2011-2020) to provide an overview/general happiness score for different countries.

As the user hovers through the map, a tooltip depicts the country name and the mean happiness score for 2011-2020. As the user selects their primary country, the country gets marked with a purple map marker icon and moves to the next page. We used purple because it provided the best contrast with the graduating warm orange/red tones that was already encoded on the map.



On the second screen, we added the same choropleth map on the top-left side. We used the same visual encoding to retain user familiarity. We decided to have two maps because we still wanted a map to be available so the user could perceive geographic trends and select nearby countries. We also felt it was important for the user to see real time updates and bidirectional linking with the other plots without having to scroll up and down.

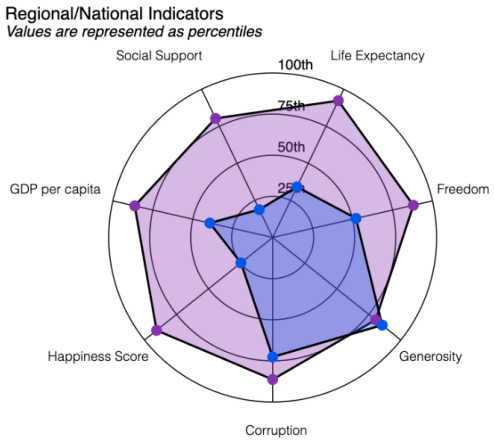
Interactions:

- Primary interaction: Select country by clicking, deselect country by clicking on an already selected country
- Color-coded tooltips: As a country/region gets selected, it becomes outlined in dark blue. We chose dark blue because it was a contrast to the sequential orange hues on the map but was not too bright as to be distracting because a user could technically select all the countries. There is a legend at the bottom showing the primary country and the selected countries and their representative purple/blue icons.
- Year slider: There is a slider to select the year so the user can easily scroll and identify changes in happiness scores over time. As you slide through the year slider, this provides an effective animation to show changes in happiness over time but also lets you pause when there is a year of interest. There may be an issue with high cognitive load due to having to remember the color before and after changing the year. To work around this, we included the option of a time series chart (more on that later).
- Region selection buttons: The user can select regions by clicking the top buttons to identify how countries in certain regions are distributed and identify outliers in those regions to understand what policies in those countries seem to be working especially well or are especially detrimental to the happiness of their populations. Only one region can be selected at a time. To swap to another region, the user can click the

desired region or to remove all countries from the bucket, the user can click “Unselect All”. We felt these were extremely intuitive and in particular, the button format would be very familiar to any Internet user.
- Country Marker: Clicking the marker scrolls the user to the top map. Clicking the marker on the top map or selecting a different country scrolls the user to the bottom view.

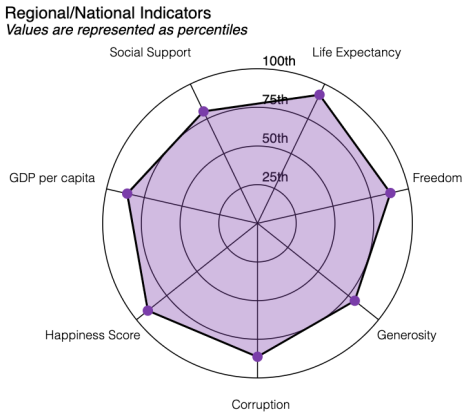
Linking:
- Selected year on year slider updates data to that year for not just the map, but also all the other visual encodings (scatterplot, line chart, radar chart).
- Selected region on the button panel outlines all countries in that region on the map, but also selects and colors those countries on the scatterplot and adds the countries in that region to the bucket for the radar chart and line chart.
- Selecting a country on the map outlines the country on the map, but also selects and colors those countries on the scatterplot and adds the countries in that region to the bucket for the radar chart and line chart.

Radar Chart

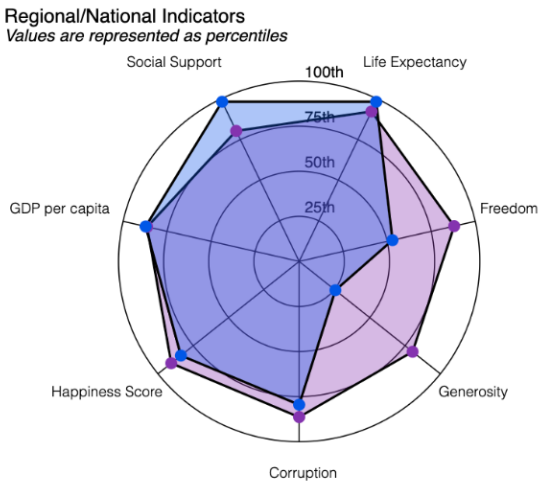


Below the choropleth map is a radar chart. Based on TA feedback, we decided to switch from using a stellar chart to a radar chart. The area/fill aspect allows the user to easily gauge countries that outperform/underperform and overlap. Comparing angles across circles is harder than comparison across a straight line, but we felt this solution to be better than e.g. using a grouped bar chart. The grouped bar chart may be better at comparing scores accurately across countries, but the radial axis orientation is better at comparing across seven attributes. Ultimately, it best suits the task we are looking to support: allowing the user to see patterns across multiple attributes, e.g. when a country has high scores in GDP and social support, they generally have a high happiness score. This multivariate comparison is difficult with rectilinear axis orientations that best support two axes or parallel axis orientations that are easiest to compare with neighboring axes.

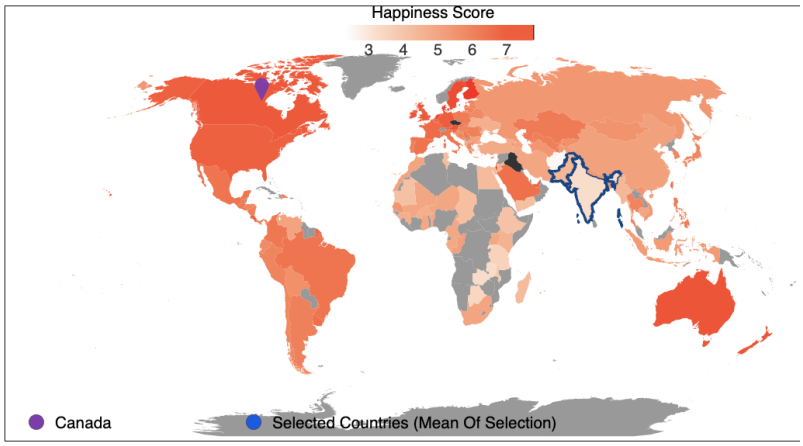
The radar chart is updated by default when a primary country is selected on the first page.

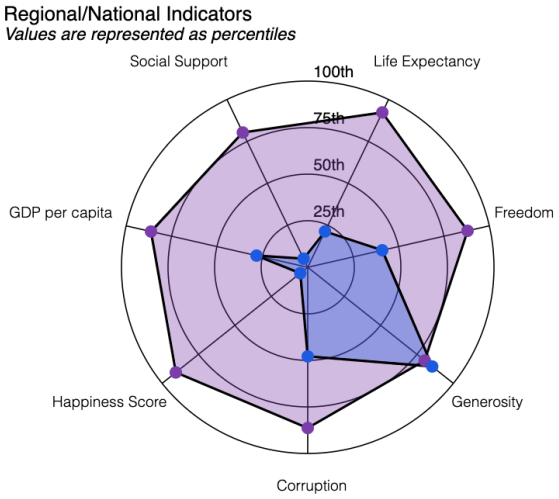


If another country/region is selected, the radar chart is updated to compare the primary country with the other selected country/region.



If multiple countries are selected, those countries aggregate into a bucket. The radar plot then displays the mean of the selected countries in the bucket.





Although our choice of visual encodings are scalable enough to accommodate for all the countries (119), we found that it created too much clutter and aggregation would better support our task abstraction. While the primary country needs a detailed focus, the other countries need not require in depth analysis. An exception to this is comparing one country to another country/region, e.g. Canada to USA or Canada to North America. Therefore, we still support this functionality. However, based on a TA suggestion, we realized an economist/policymaker may be more interested in composite scores and aggregation. This would best support

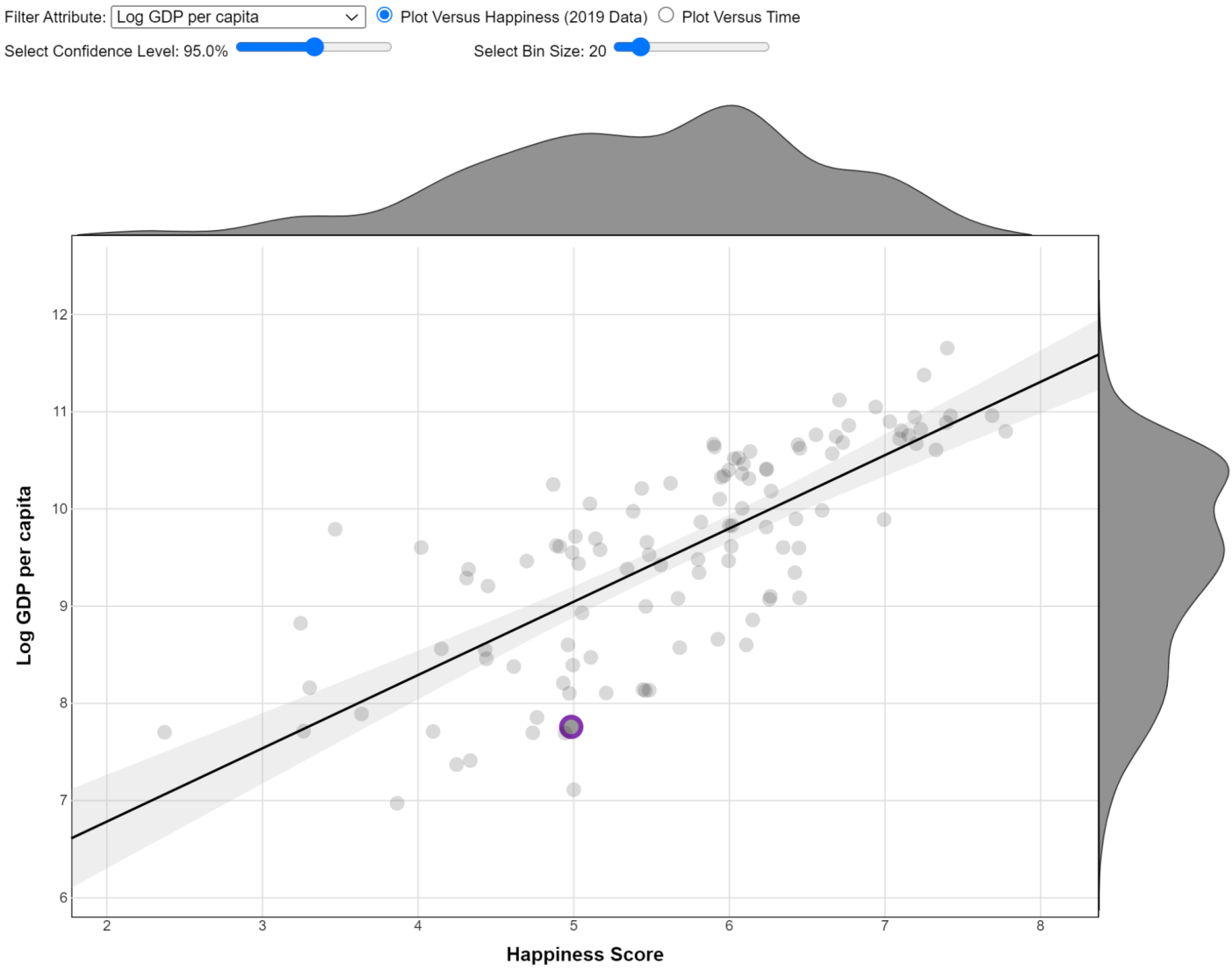
the user's goal of seeing how their country of focus is doing in comparison to others. We then presented a mean score. This allows the user to focus on general patterns as opposed to 1V1 comparison with each selected country, which would require a high cognitive load of remembering each country level comparison. The scatterplot (more on that later) will still be updated to show individual countries that are selected in case the user is interested in having more detail or finding outliers - those tasks will still be supported through the scatterplot.

As mentioned, we are using percentile data above to compare all attributes on the same scale. This is important because otherwise each attribute has large differences in scale. We have clearly communicated this in the visualization's description. We anticipate this being very useful to our target user because country comparisons are largely done relative to one another in percentiles. This should then be intuitive and familiar for the user.

The points are colour coded according to whether it is the primary country or another country or whether it is part of the aggregated bucket of selected countries or regions. Using the color hue channel allows the user to effectively see which group it belongs to and we also used the hue channel (at a lower opacity) in creating an area effect between the points to show overlap. They are two bright and different hues so there is popout and distinguishability. Clicking on a point on the radar chart opens up a color-coded tooltip (purple/blue as per the legend). Having access to exact figures may be important to our target user so those have been detailed on the tooltip alongside the percentile value. The tooltip also includes the other selection countries values in decreasing order for easy comparison. However, the focus of this visual encoding is on [discovering] trends and patterns so high accuracy in terms of being able to easily read accurate values is not the priority, unlike the scatterplot and time-series line chart.

Note: Clicking on the radar chart does not update any of the other charts. This visual encoding's primary purpose is for viewing as opposed to interaction. The map and scatterplot are unidirectionally linked to the radar plot, which gets updated when countries or regions are selected.

Scatterplot with density distribution (Innovative view)

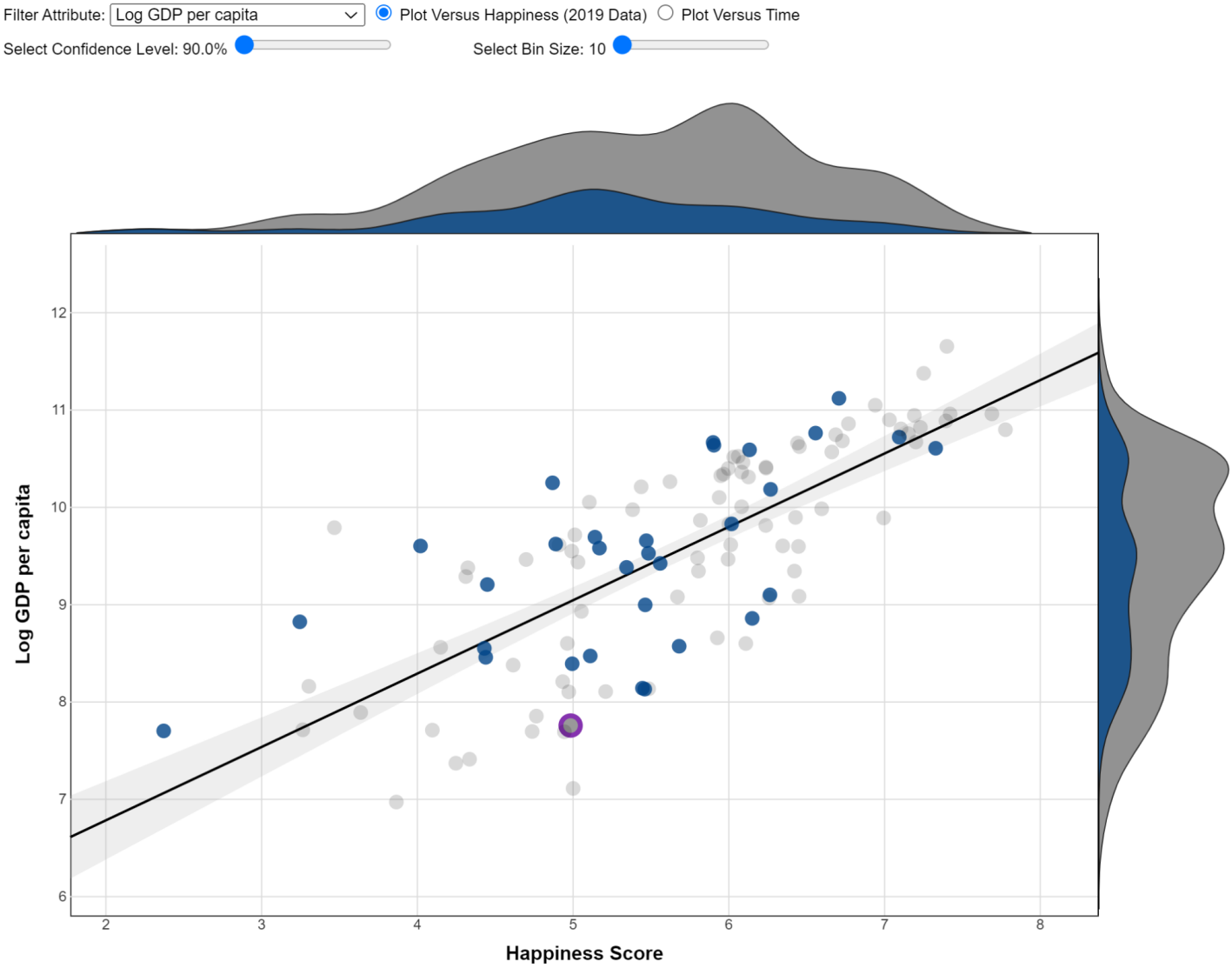


Marks and Channels

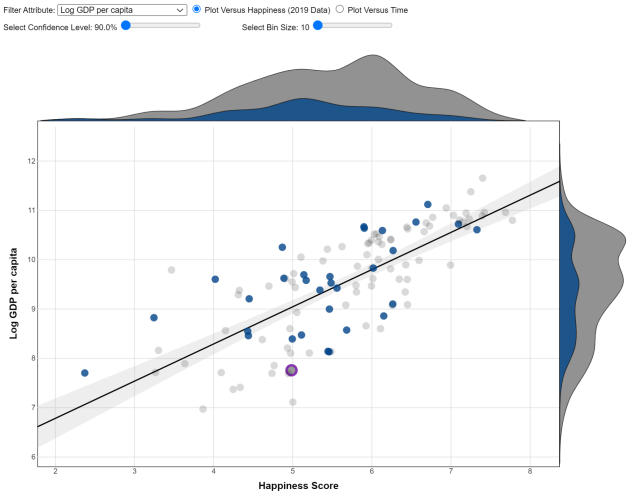
- Scatterplot:
  - Marks:
    - Each point represents a country item
  - Channels:
    - Horizontal position encodes happiness
    - Vertical position encodes selected attribute (can be GDP, Corruption, Social support, etc.)
    - Saturation represents whether the country is selected (unselected points are grey while selected points are blue)
    - Hue represents type of selection (region is encoded in blue while primary country is encoded in purple)
  - Marks:
    - Line mark represents the regression line
  - Channels:
    - Tilt of the line represents direction that the points are most strongly correlated

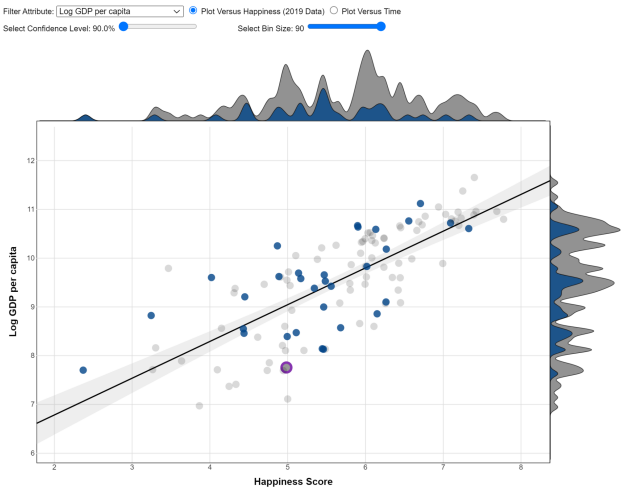
    - Marks:
        - Line marks denotes the confidence interval of the regression line at each point along the line
    - Channels:
        - Vertical length represents the spread of the regression line based on probability of the selected confidence level
        - Horizontal position encodes happiness
        - Vertical position encodes selected attribute (can be GDP, Corruption, Social support, etc.)
- Marginal histograms:
    - Mark
        - Line marks represent buckets of values
    - Channels
        - 1D length represents the frequency of occurrence of the value
        - Color hue represents whether all countries (grey) or countries in selected basket (blue)
        - Horizontal position for the top histogram represents happiness score and vertical position for the right histogram represent social support, corresponding to the axes of the scatterplot
- Confidence interval:
    - Mark
        - Interlocking area represents confidence
    - Channels
        - 2D area representing the probability that a parameter will fall around the regression line

On the right, we have a scatterplot (with some extensions that will be discussed later) to compare the six attributes that may contribute to national well-being plotted against happiness scores. This scatterplot is shown when “Plot versus Happiness” is selected (which is the default option). We chose this visual encoding because it is standard and familiar for users. It is extremely scalable and allows us to depict all countries clearly along with any overlap (through decreased opacity). It uses a rectilinear axis orientation and that, along with the large size, allows for high accuracy and effectiveness through horizontal and vertical 2D position. It also enables us to add a linear regression line (discussed further later).



We also included smoothed density histograms. The distributions along the x and y-axes are displayed in histograms to the top and right of the scatterplot for selected countries and juxtaposed with the overall distribution. This supports the important task of allowing users to {discover, distributions} to understand the distribution of attributes across different countries. They act like scented widgets. When a region or additional countries are selected for comparison, the density histograms show in the same blue the distribution for the bucket of selected countries/region. For example, an economist may find that most European countries have high generosity levels but Asian countries are all over the place with differing generosity levels. We chose to add this because it allows the user to get a sense of where the countries lie, provides a useful means of finding patterns, and saves screen pixel space by fitting it in with the scatterplot. We also provided a slider widget for changing the number of bins on the histogram. This allows the user to increase/decrease granularity as per their desire, to support seeing general trends or discovering outlier clusters.





We wanted our innovative view to be able to offer economists and policy makers a way to understand the data in depth, so we decided to add statistical elements, like the distribution along each axis, linear regression line and confidence intervals. The regression line and the confidence intervals offer insight as to how strongly happiness is correlated to each additional attribute, so that they can make policies that affect attributes that offer a better return on the overall happiness of the population. We also added a slider widget for changing the confidence level as this may be important for economists who are making forecasting/prediction calculations. We expect our target user, economists and policymakers, to have the education and experience necessary for leveraging these added statistical functionality without any explanation. We used luminance to have reduced opacity for the confidence interval to distinguish between the confidence interval and the linear regression line.

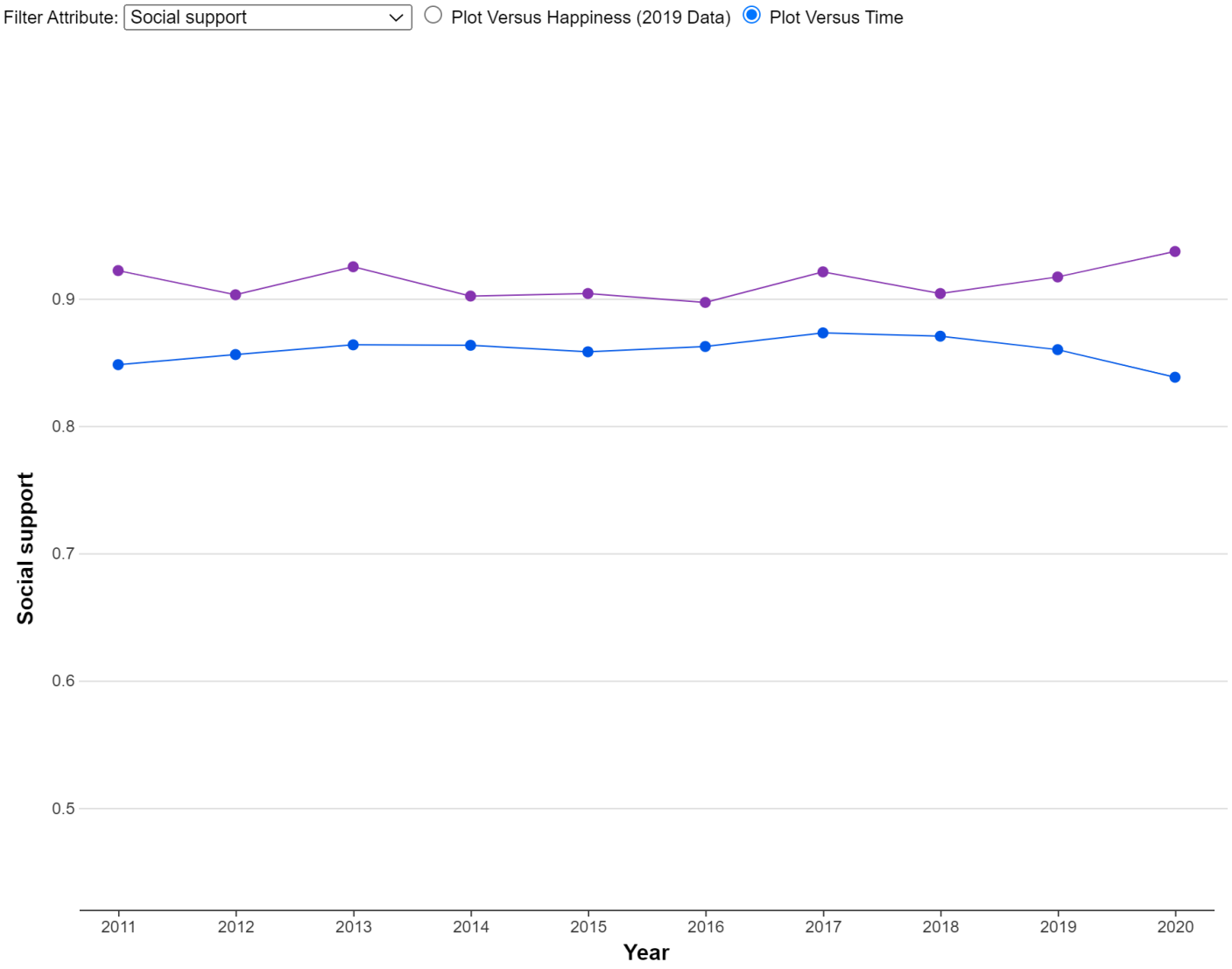
Other interactions:

- Primary interaction: Click on a point to select a country. Click on an already selected point to deselect. Click on an empty part of the plot to deselect all countries.
- Brushing: Drag to highlight and select countries. This choice was made because we want the user to easily select multiple countries in a particular area, without having to click each and every single one of them. For example, the user may be interested in adding all countries with happiness scores above 7 to the bucket for inclusion in comparison.
- Selection of “Plot versus Happiness” to display scatterplot or “Plot versus Time” tp display time-series line (further discussed later).

Linking:

- Bidirectional linking with map: Selecting a country on the scatterplot selects the country, adds it to the comparison bucket, fills it in blue, and also outlines the country on the map.
- Unidirectional linking with radar plot and line chart: Selecting a country on the scatterplot adds the country to the comparison bucket so it updates the radar plot and line chart with that country’s data accordingly.

Time-series line chart

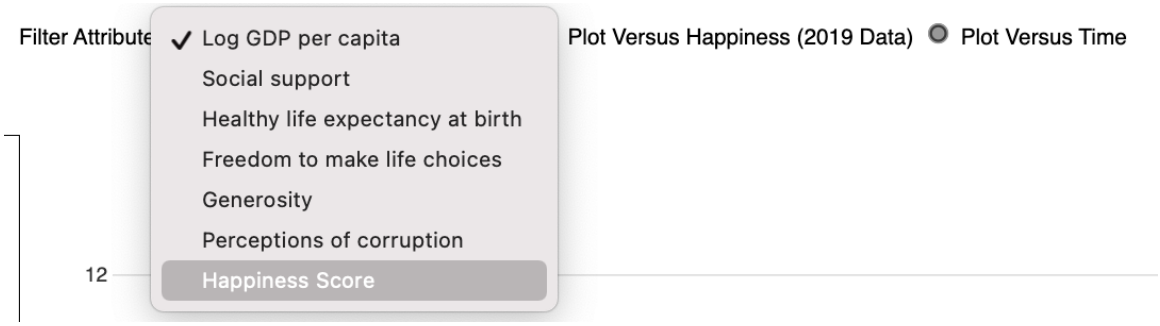


When “Plot Versus Time” is selected above the scatterplot, it opens up a line chart for the attribute selected in the “Filter Attribute” dropdown. The line chart helps analyze individual attributes over time. The color hue channel encodes the lines to show whether it is the primary country or another country or whether it is part of the aggregated bucket of selected countries/regions. Line chart was chosen for the visual encoding because it is familiar to economists and policymakers for such time-series tasks and the rectilinear orientation using horizontal and vertical 2D position allows for high accuracy.

We were planning on using small multiples to showcase different attributes at once. However, we decided to use the dropdown selection to filter and select a single attribute to depict. Initially, we thought it might be useful for the user to compare multiple attributes over time at once. However, doing this and using small multiples would require a high cognitive load because you would still have to remember the trend/figures before looking at the next graph. Given our screen space constraints, we thought small multiples would be far too small to read for a task where accuracy is important. We decided not to graph multiple attributes on the same line graph because the different metrics for each attribute makes the scale vastly different so it would be less effective and accurate. Accuracy was a factor we felt to be most important in time-series tasks because changes even in months could be crucial to note, e.g. GDP increased in the same year that lower interest rates were

announced. Therefore, we prioritized accuracy over comparing multiple attributes on one screen based on calculated assumptions about our target user. We anticipate that our target user will be most interested in the task of comparing a single attribute over time (particularly happiness).

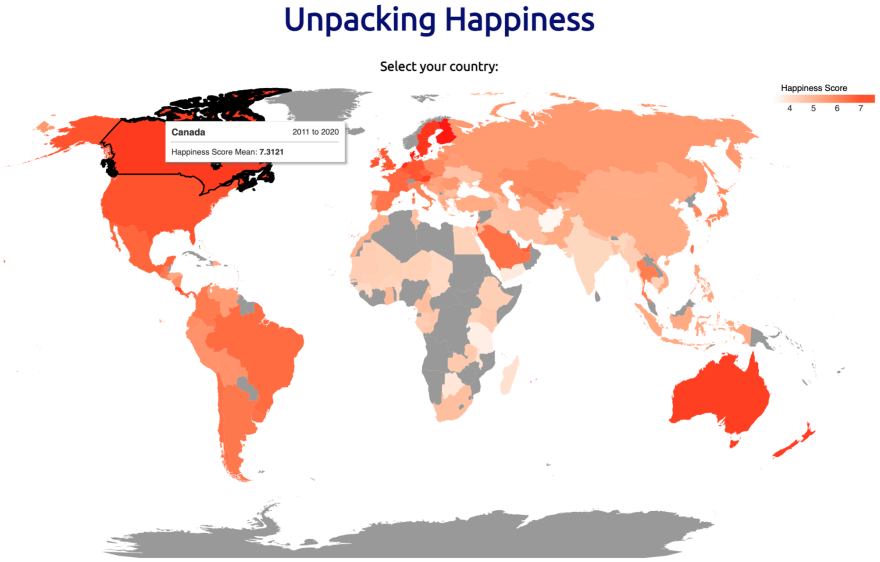
Note: Clicking on the time-series line chart does not update any of the other charts. This visual encoding's primary purpose is for viewing as opposed to interaction. The map and scatterplot are unidirectionally linked to the line chart, which gets updated when countries/regions are selected. Clicking on a point on the line chart does open up a color-coded tooltip (purple/blue as per the legend). Having access to exact figures will definitely be important to our target user so those have been detailed on the tooltip. The tooltip also includes the other selection countries values in decreasing order for easy comparison. As with the scatterplot, the attribute to be graphed can be selected using the "Filter Attribute" dropdown at the top-right of the screen. When "Plot versus time"/ the line chart is selected, the user additionally has the option of selecting the happiness score attribute. This is not available for the scatterplot so users don't graph happiness against happiness, but can still view changes in the happiness level over time. Based on M2 TA feedback, our previous line charts felt unintegrated with the rest of the project. Using the same dropdown attribute filter helps integrate it with other plots and saves space and also both the line chart and scatterplot to be big for easy viewing/reading.



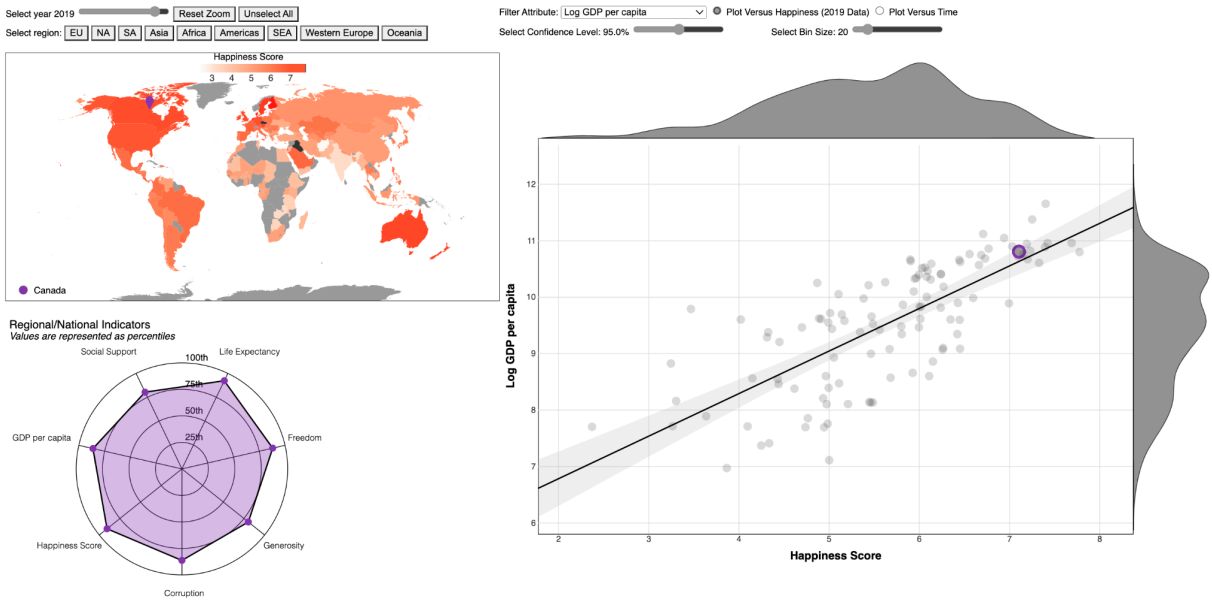
Usage scenarios

Jose - Canadian Ministry:

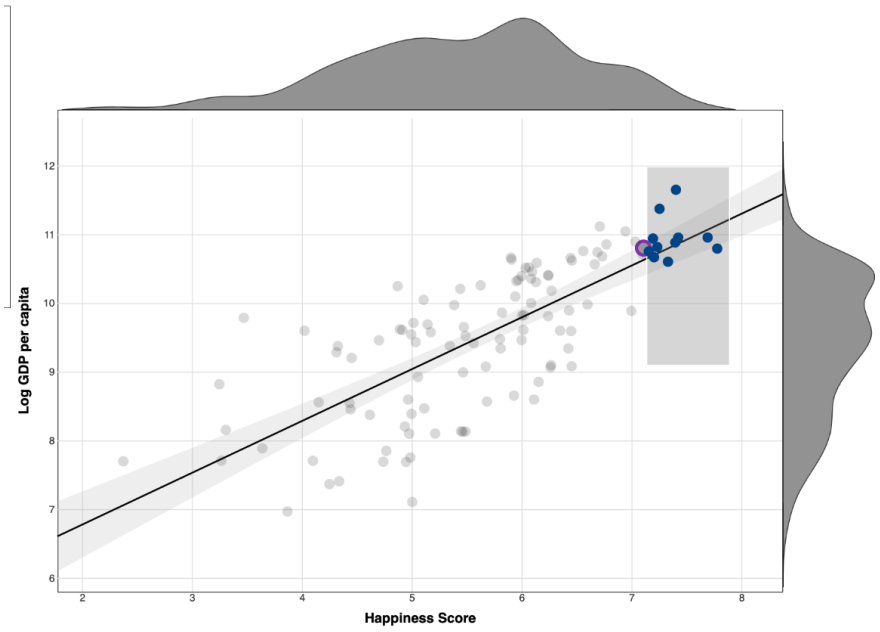
Jose is a part of the Canadian ministry and wants to understand room for improvement for Canada in order to set federal policies and priorities. When Jose opens up the visualization, he will see a world map where each country is filled with a different color depending on its happiness score. As he hovers over Canada, a tooltip pops up confirming that it is Canada and stating Canada's happiness score out of 10.



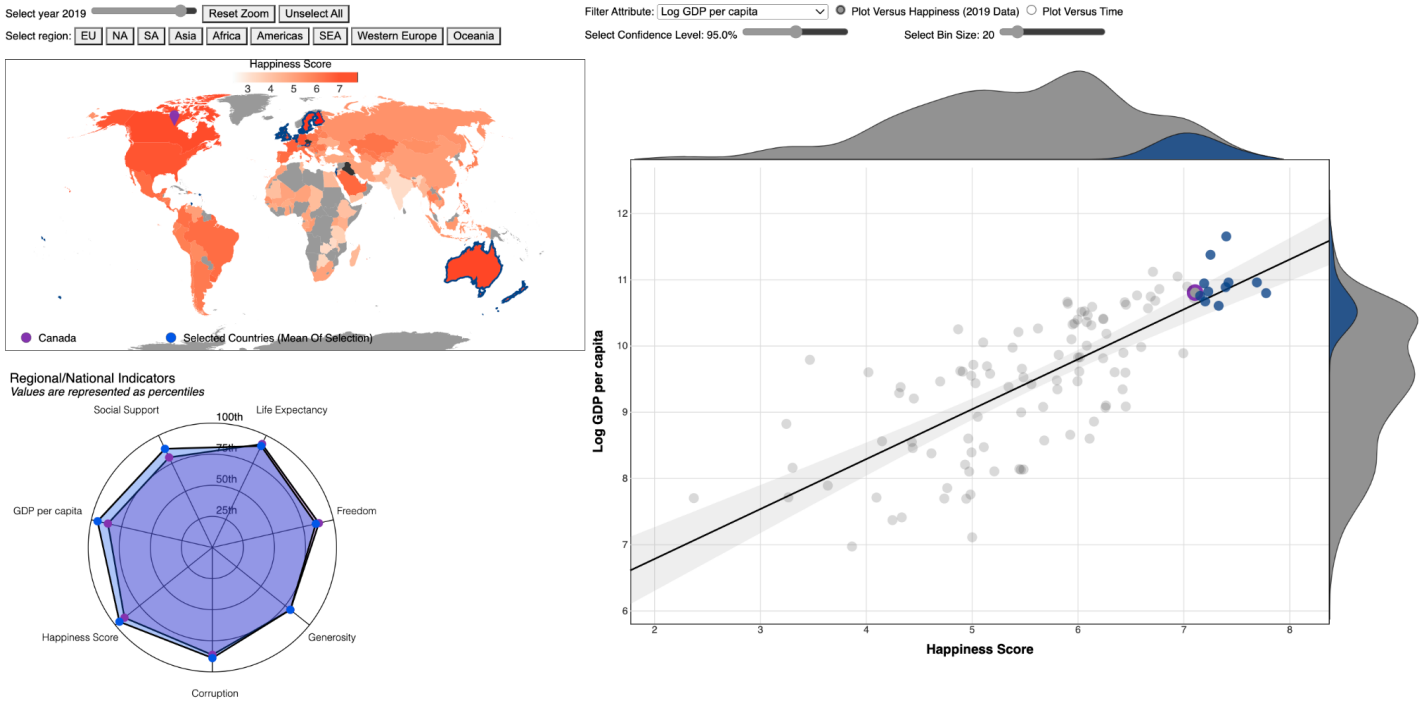
As he clicks on Canada, the country on the map has a purple map marker icon added to it and scrolls down to the next page.



The scatterplot shows all the countries graphed to show GDP and happiness score. Canada is represented with a purple outline. Jose sees that Canada has a high happiness score and GDP compared to the rest of the country. He notes the positive correlation between GDP and happiness. He notices some countries that have a higher happiness score than Canada, like Sweden, and selects them by dragging.



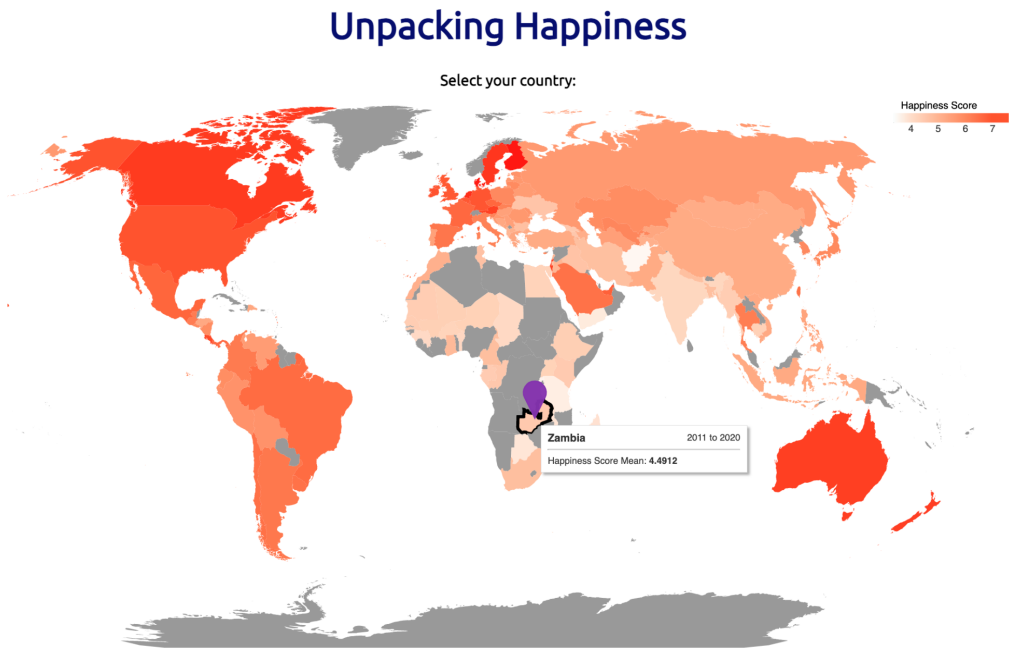
These countries get added to an aggregated bucket and the plots get updated to account for these countries in blue.



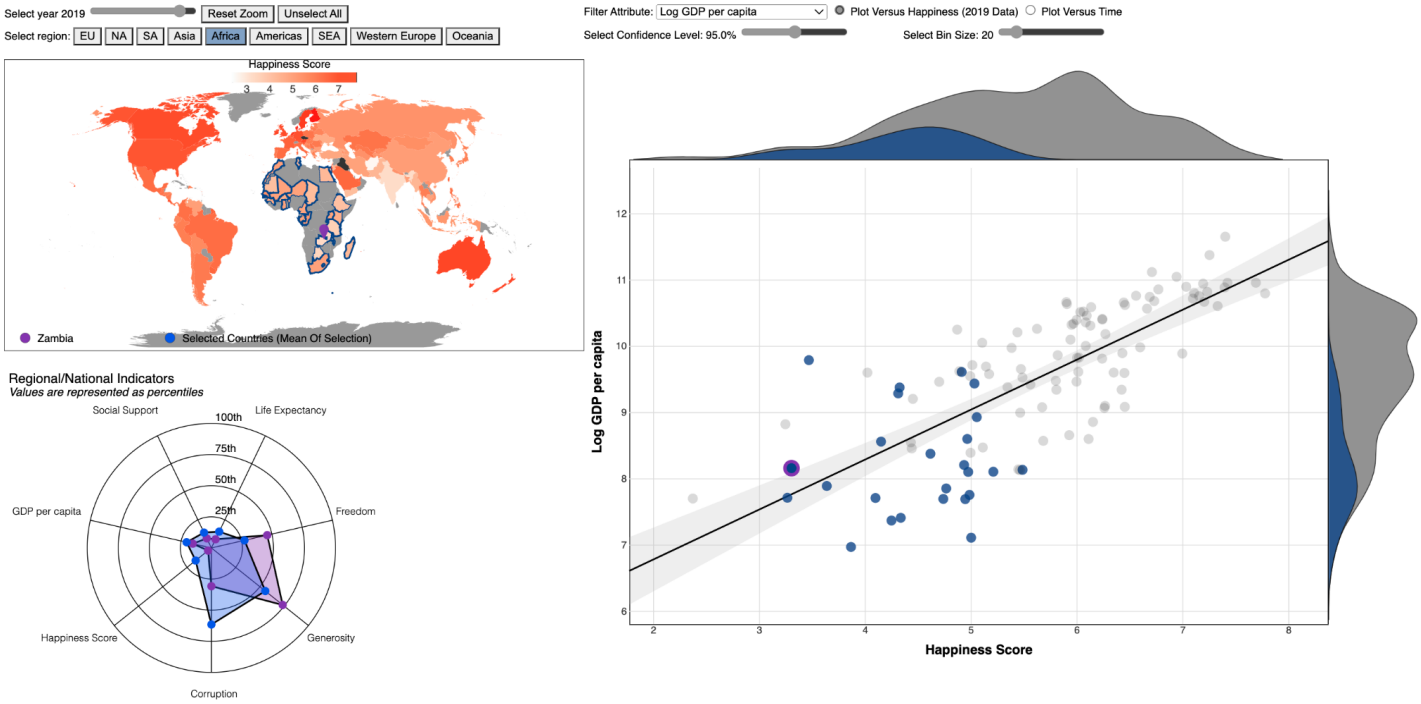
He changes the x-axis variable from GDP to other factors using the dropdown box. On the radar chart, he sees that Canada does best on some of the factors, but the selected countries have better social support scores and higher GDP per capita than Canada. Jose hypothesizes that having a high GDP makes a country happy, but social support is important too. As a next step, Jose decides to look into Scandinavian policies to improve social support.

Tara - Economist:

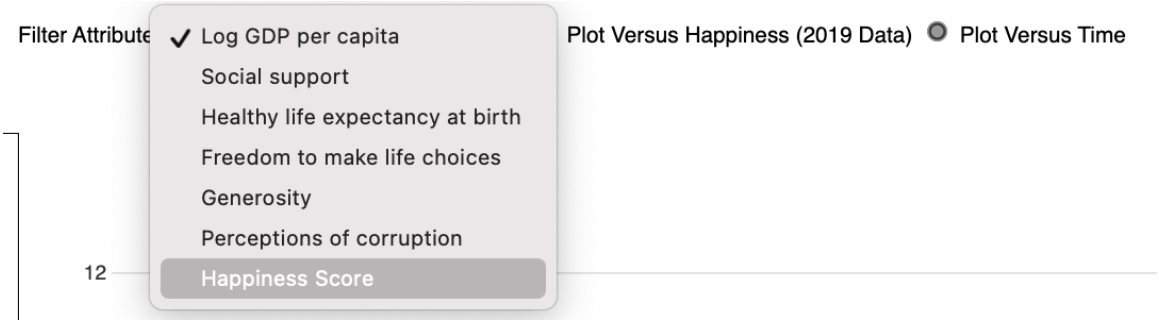
Tara is a Zambian economist. After opening up the visualization, Tara selects Zambia and is navigated to the next page.



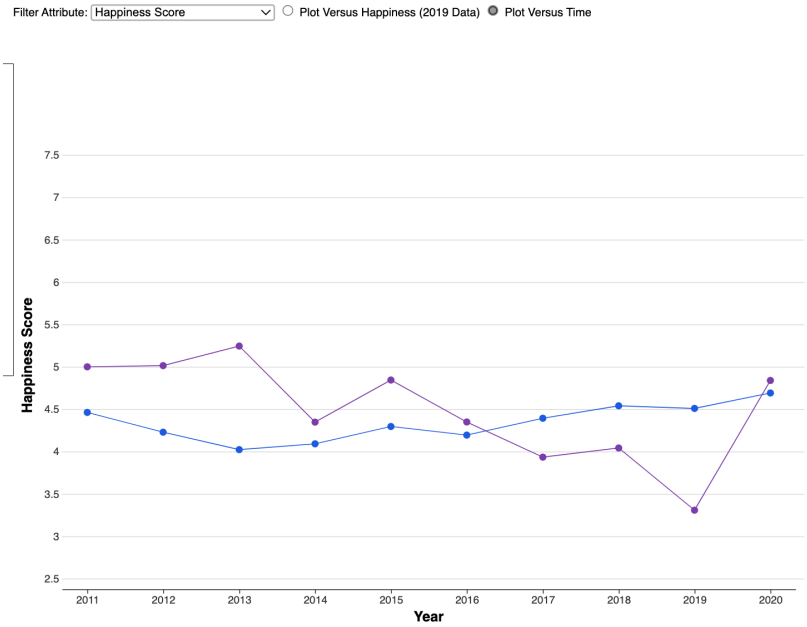
She also selects “Africa” as a region for comparison with Zambia.



On the radar plot, she notices that Zambia has a lower happiness score than Africa on average. Those countries don’t have higher GDP per capita on average, but have far better mean life expectancy scores. Her peer in Canada, Jose, argued that GDP and social support are the attributes that have the highest correlation with happiness. She wants to explore her hypothesis that for African economies, life expectancy matters more than GDP. After exploring the scatterplot and radar chart, she clicks “Plot versus time” to see a line graph that maps a factor’s score over time. Tara selects happiness as the attribute to filter by.



She sees that there was a drastic drop in Zambian happiness score for 2019 and a big hike for 2020.



She jots this down to look into why that might be. As a next step, Tara will look into her learnings further and compile more research to come up with a forecast to predict happiness scores for Zambia in the next 5 years.

Credits

Gradient Legend - bl.ocks.org

Problem: Displaying a legend that contains a gradient of colors

Modification: Added offset and stop-color to a rect at intervals of our color scale to create a gradient scale

Source: bl.ocks.org: Gradient Legend - bl.ocks.org

Radar Chart d3 - github.com

Problem: Creating radar chart paths

Modification: Math equation for calculating line angles for each attribute axis on a radar chart

Source: github.com: alangrafu/radar-chart-d3: Simple radar chart in D3.js

---

LineChart d3

Problem: Creating a plot for average over all selected countries

Modification: Rollup data based on year, find the mean and add it to the chart

Sources:

- Interactive line charts
- Overvable d3 rollups
- Geeks for Geeks

Scatterplot d3

Problems:

- Drag to select multiple countries
- Plot the line of best fit
- Draw the confidence bands

Sources:

- Scatterplot based on example from tutorials
- Used D3 Brushing to understand how to perform brushing, how to calculate if a point is contained in the selected area, and how to change the style of the selected points
- Used Scatter plot + Brush to know how to hide the brushed area on mouse up

- Used Simple Linear Regression to know how to calculate and draw a regression line using Simple Statistics
- Used Linear Regression with Confidence Bands to know how to draw the confidence bands (functions for statistical calculations are copied with no modifications)

Marginal Histograms

Problems:

- Calculate histogram binning
- Draw continuous histograms

Sources:

- Used Basic Density Graph and Density Graph From Integers to understand how to draw a smoothed curve from a list of integers
- Used Histogram to understand the d3.histogram() function and how to use it to bin the data

---

Choropleth Map

Problems:

- Draw choropleth world map
- Zoom + pan functionality
- Select predefined regions
- Reset zoom button
- Draw marker on primary country

Sources:

- Used Choropleth Map Tutorial Example to know how to use topojson to draw a map using geojson
- Used D3 Choropleth Map to understand how to add zooming + panning interactions
- Used Pan & Zoom Axes to reset the map zoom on button click
- Used Map Markers to get the path shape for the marker and Country Centroids to learn how to calculate the centroids of country paths to position the marker

---

Data Sources and Libraries

Sources:

Datasets:

- Happiness Dataset
- World Map GeoJSON (50m)
- Geographic Regions

Libraries:

- Simple Statistics
- TopoJson

Reflection

Changes in visualization goals

- Switching target user to economists/policymakers

Based on TA feedback from M1, we decided to switch our target user from the general public to economists and policymakers. We felt it was contrived that the general public may need such detail on global happiness scores and other relevant factors for personal decisions, e.g. voting. We have a more logical use case for economists/policymakers. Additionally, this allowed us to be ambitious with our innovative view and include more complex visual encodings, such as the confidence interval, that require some experience with graph reading.

- Including regional data

When we thought about what other data might be interesting to our new target user, we felt that it would be particularly useful for economists and policymakers to have regional data. They would likely want to identify trends in certain regions, compare a country to its/other regions and the change of a region over time. Hence, as described in the Data section, we chose to integrate regional data into our dataset.

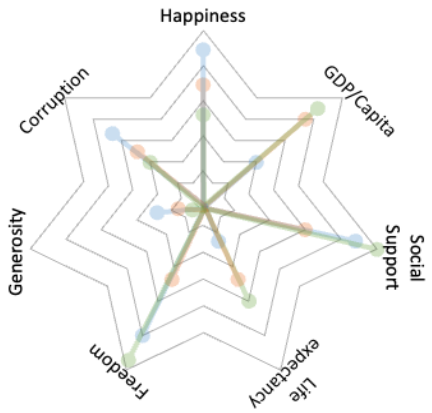
Changes in technical goals

- Updated innovative view

Based on the feedback from M1, our innovative view of the stellar chart was not novel enough. After discussing with our TA, we decided to make a significant extension of our scatter plot instead. In addition to the regression line which we planned for in M1, we decided to also display the confidence interval. We also allowed users to edit the confidence level (95% being the default). This would be of particular interest to economists who may be interested in quantitative patterns for use in forecasting. The distributions along the x and y-axes are displayed in smoothed histograms to the top and right of the scatterplot for selected countries and juxtaposed with the overall distribution.

- Switch from stellar chart -> radar chart

Based on TA feedback, we decided to switch from using a stellar chart to a radar chart. The area/fill aspect allows the user to easily gauge countries that outperform/underperform and overlap. The stellar chart was solely relying on length that was angled which is far less effective than a radar chart. This is what it would have looked like before our change:



- Bucket for comparison

We had initially created a 3 country limit for comparison, in order to reduce clutter. However, we felt it was important for economists and policymakers to compare a host of countries for them to understand trends. We still didn't want ~100 points on the radar and line chart because the clutter took the focus away. We looked to aggregation as a way of solving this problem. This would let the user compare how many ever countries they wanted by looking at the mean of the selected bucket. This would also let the user create their own regions that aren't geographic regions, e.g. NATO countries. They can still compare one country to another.

How realistic was your original proposal in terms of what is technically possible in D3?

We felt that our original proposal was really quite realistic in the sense that it was all technically possible in D3. Nevertheless, as noted in TA feedback for M2, our innovative view along with our features were quite extensive. They required far more time than we anticipated and a lot more trial and error. In general, we spent far more time on most things than we initially predicted. We found D3 to be exceptionally powerful and we could really make our visualizations look how we wanted them to (given a lot of time and error!). Rather than D3, we found HTML/CSS to be the hardest for us to figure out.

Was there anything you wanted to implement that you ultimately couldn't figure out how to do? If so, then what workarounds did you employ, or did you abandon your original idea?

This meant that there was nothing we wanted to implement that we couldn't figure out how to. However, there were a lot of additional features and finishing touches we wanted to add but this project taught us that, given the time constraints, we needed to prioritize.

If you were to make the project again from scratch (or any other interactive visualization), what would you do differently?

Significant amounts of time was spent building out plots and features that we didn't end up using. As we spent hours of time every weekend in discussion with each other about what the visualization would look like, we understood the importance of all the sketching in the first milestone. We all agreed that we wished we had consulted TAs on our project sooner, because if we were executing a design we'd already had in mind, we would have a lot more time for polishing, interactions and styling. We underestimated the importance of documentation. For example, we hadn't really described the interactions so we spent a lot of time understanding which interactions to add.

Project Management & Team Assessment

Project Schedule

Milestone 1: Data cleaning and preprocessing - Target completion: March 10; Completed: March 25

Work	**Est. hours**	**Actual hours**	**Completed by**
Project proposal and sketching (this includes changes made post M1 and M2 feedback)	6	20	Himanshu + Anju + Kevin
Filter data to only include 2011-2020	1.5	1.5	Kevin
Filter countries with missing data	2.5	2.5	Kevin
Find the percentile score for each happiness factor attribute	4	4.5	Himanshu
Total	**14**	**28.5**	-

Milestone 2: Create Static Views - Target completion: March 26; Completed: March 29

Work	**Est. hours**	**Actual hours**	**Assigned to**
Choropleth world map with happiness scores	25	20	Kevin
Scatterplot with histograms	35	40	Himanshu
Linear regression line on scatterplot	6	6	Kevin
Radar chart	30	35	Anju
Line chart	5	6	Himanshu
Innovative view redesign	10	10	Himanshu + Anju + Kevin
M2 Writeup	6	10	Anju
M2 Writeup Editing	2	2	Himanshu + Anju + Kevin
Total	**121**	**129**	-

Milestone 3: Add Interactions - Target completion: April 2; Completed: April 11

Work	**Est. hours**	**Actual hours**	**Assigned to**
Outline/highlight the selected country on the map and highlight corresponding points on the scatterplot and vice versa	12	14	Anju
Year selector - Slide the year to filter data on both the scatterplot and the map	3	2	Kevin
Dropdown for attribute selection for the scatterplot	3	4	Himanshu
Calculate mean percentiles	8	10	Himanshu
Regional filter for scatterplot	5	6	Kevin

Display data for the selected countries on the radar chart and line chart	2.5	4	Anju
Add tooltips - Map, scatterplot, radar chart and line graph	7	12	Anju
Add drag to select points on scatterplot	3	4	Kevin
Confidence interval on scatterplot + confidence level selector	10	14	Kevin
Draw marker on primary country + click interaction for marker	5	8	Kevin
Continuous marginal histograms + calculate binning	10	10	Kevin
Add zoom/pan interactions on map	10	4	Kevin
Total	**78.5**	**92**	-

Milestone 4: Styling and Editing - Target completion: April 11; Completed: April 12

Work	**Est. hours**	**Actual hours**	**Assigned to**
Legends: Happiness index (map), countries + color legend	5	9	Anju
Select + integrate color scheme	3	3	Anju
Integrate scatter and line charts into the same space	3	4	Himanshu + Kevin
Layout + aligning plots	4	6	Himanshu + Kevin
Auto scroll after selecting main country	4	6	Himanshu
Interactions (highlight point on click, clicker pointer etc.)	5	6	Anju
Blue background on selected region button	2	3	Kevin
M3 Writeup	20	15	Anju
Final styling touch ups	3	5	Himanshu + Anju + Kevin
Total	**46**	**52**	-

<u>Team Members & Responsibilities</u>

All members contributed to each aspect of the project and there were often overlaps between responsibilities. More detailed breakdown can be found above, but responsibilities was divided as follows:

- Himanshu - responsible for the scatterplot and line chart
- Anju - responsible for the radar plot, interactions, styling, writeups
- Kevin - responsible for the choropleth map and regression line/confidence interval

We thank our TAs, Francis Nguyen and Mara Solen, and our instructor Tamara Munzner for their invaluable contributions.