

# Visualizing The Natural Capital Project Training Program

Varun Mathur

[varun.mathur@epfl.ch](mailto:varun.mathur@epfl.ch)

Shikhar Dhar

[shikhar.dhar@epfl.ch](mailto:shikhar.dhar@epfl.ch)

Doga Tekin

[doga.tekin@epfl.ch](mailto:doga.tekin@epfl.ch)

December 21, 2018

## 1 Introduction

In this process book, we hope to show our approach to the design and implementation of the website we created for the Natural Capital Project. In the following sections, we will specifically explain the different ideas we had when it came to representing the data in a meaningful way, guiding the user through the information on the page, and linking separate piece of data to tell a story. The design principals, story flow, and visualization are all heavily inspired by the information we learned in COM-408 (Data Visualization).

### 1.1 Overview

The Natural Capital Project is a collective of academics, software engineers, and real-world professionals who share a common goal to shine a light on the intimate connections between people and nature, and to reveal, test, and scale ways of securing the well-being of both. With that being said, our goal is to help the Natural Capital Project with not only understanding overarching data facts, but to also allow them to explore their data and organization on their own. Building a tool that allows them to explore their data in an intuitive way will be more useful because it will provide them with the opportunity to look into specific questions they might have about their operation. Our overall goal is to help the organization answer any questions they might have with

regards to their trainings and overall figures, while also building a well designed website that will help inform new viewers about the Natural Capital Project Symposium and help them understand more facts about the organization.

## **1.2 Motivation**

The motivation behind this project is help the Natural Capital Project with understanding more about their data so that they can make more informed decisions about future trainings. We are also trying to get new people to understand more about the Natural Capital Project and more specifically understand the trainings that they have had over the years. We hope to achieve two goals with our website, the first is to help employees understand more about the organization and the other to help new users learn more about organization.

## **1.3 Target Audience**

Our target audience can be split into two separate groups. The first includes people that work for the Natural Capital Project. With our visualizations, we hope to help the employees understand more about their own organization so that they can make better informed decisions and understand more about their global trainings. The second group of people that the audience can be split into includes people that want to learn more about the Natural Capital Project. We hope that our visualizations will attract more people to attend symposiums and other events hosted by the Natural Capital Project. We also hope that our visualizations will provide new people with a better insight about the Natural Capital Project and their trainings worldwide.

## **1.4 Inspiration**

We were inspired to work on this project because of the Natural Capital Project. After we spent time reading through the site and understanding more about what the organization works on, we decided we wanted to take part in helping the Natural Capital Project better understand the data they currently have so that they can make a bigger difference on the global landscape. We also were inspired to help shed new insights about the organization to better inform the public about the impact that the Natural Capital Project has in the worldwide community.

## **2 Natural Capital Project Training Program Data**

The data we are using for our project is about the participants and instructors that took part in the Natural Capital Project Training Program (NCPTP). The NCPTP teaches ecosystem services approaches to multiple audiences around the world. They break up their courses into 3 levels: Intro, Intermediate, and Advanced. The intro course teaches basic concepts and serves as an introduction to natural capital approaches and tools. The intermediate course is more about hands-on application of the tools learned in the intro class. The advanced course is a more individualized tailored workshop session with project-specific support with local data.

### **2.1 Dataset**

The dataset we used was an excel file with information about all of the participants in the training program and all of the instructors for all the years from 2013 to 2018. The data was split into two main parts: information about the person, (email, affiliation, and country), and information about the courses they took (type, country, number of days, and class level).

### **2.2 Data Preprocessing**

For the preprocessing we had four main tasks: cleaning the data, formatting it properly, calculating general statistics, and generating a different dataset for each of the different visualizations. For the cleaning aspect, there were some entries in the 'Country' column which were not countries or they were misspelled. For example, there were entries for both UK and United Kingdom and there was an entry for 'Virtual'. After fixing these issues, we had to convert the 'Country' column entries from full country names into country codes. This is because the map library we decided to use requires the input to be in country codes.

Next we generated all of the datasets. The first dataset was used for the world map. It entailed grouping the number of trainee data by country and then grouping the result by year. This allowed for easy filtering based on country and by year, which was perfect for the map. The next dataset was used for both the area chart about the symposium attendees, and also the bar graph showing the symposium attendees vs others. In order to get this data, we had to take the trainee data, group it by year and then filter out the trainees that attended a symposium and get a count. The next dataset was used for the bar graph showing the distribution of the different types of trainings. We grouped the trainee data by country, then by year, then by the type of trainings and then got the count. This made for easy filtering by country and year, which was extremely useful for linking the bar graph to the selections on the map.

The next dataset was used for the line graph and the bar graph that showed the distribution of the different levels of trainings. Similarly to the previous dataset, we grouped by country, then by year, and then by the levels of trainings and got the count.

## **2.3 Exploratory Data Analysis**

After we were done with the cleaning, formatting, and generating the datasets, we computed some general statistics. First we split up the dataset into data about the trainees and data about the trainers. Then we calculated the following general statistics: total number of people trained over the years, total number of trainees over the years, total number of countries trained in over the years, total number of countries that trainees come from, and the total number of symposium attendees. We also looked at the breakdown of the different types and different levels of trainings over the years. After filtering and grouping the levels and types together, we used a line graph to visualize the resulting data to see if we could notice any trends.

# **3 Designs**

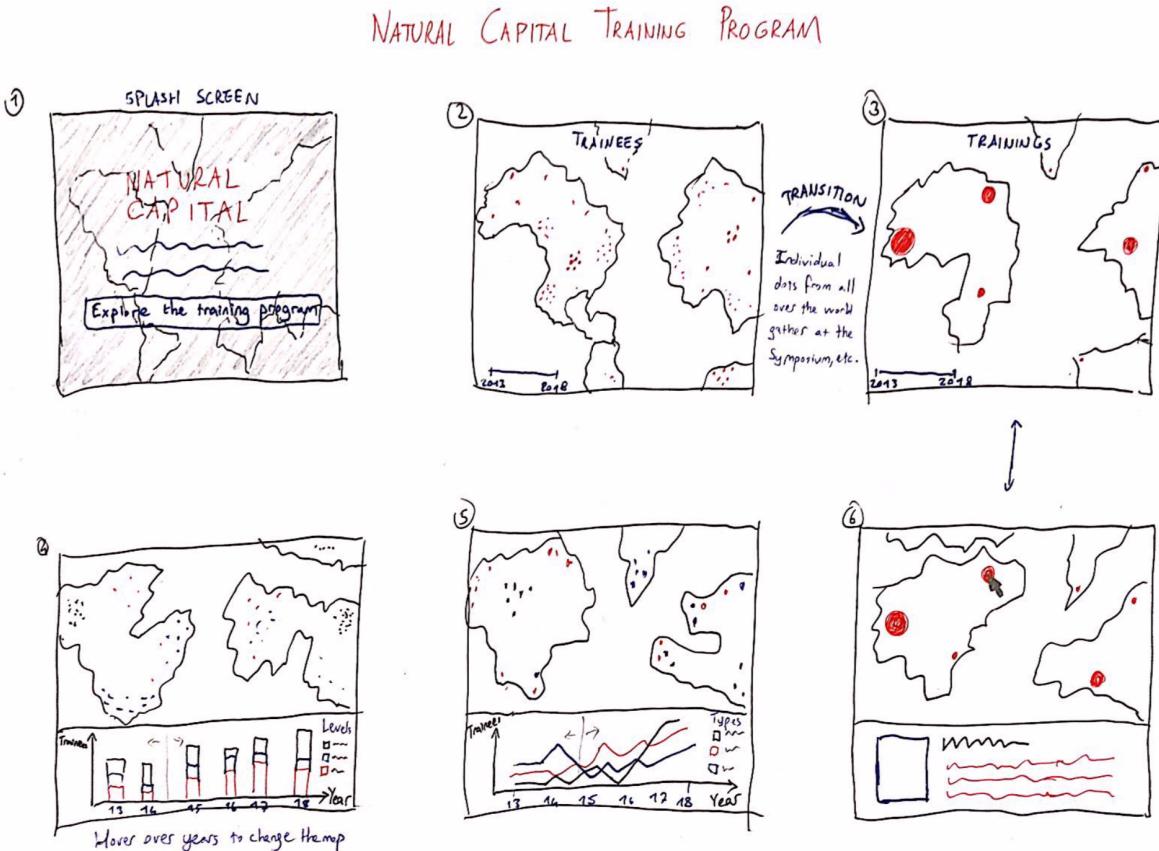
In this section, we discuss our design process and show how our designs changed over time to better engage the user, tell a better story, and account for changing circumstances. We also talk about our different ideas when it came to designing the layout of the website to best engage the user.

## **3.1 Desired Flow of the Design**

When designing the website, we had to first understand who our target audience was and what each of the different users wanted to see when they were using the website. As mentioned above, our target audience is split into two categories with the first being employees of the Natural Capital Project and the other being new users. Taking this under consideration, we wanted to design the website to display both general information to be useful for the new users and the specific and customizable visualizations that will be useful for the employees that want to explore more about their trainings. To accomplish this, we decided to segment the information into two separate sections

### 3.2 Initial Sketch

The first thing we did at the start of the project was sketch our ideas with pen and paper, as shown in Figure 1. As can be seen, the initial idea had a full-screen map with the trainees scattered around the map and the line chart and bar chart were planned to be in a panel on the bottom.



**QUESTIONS TO ANSWER :**

- Which countries are the trainees from?
- How many people have been trained (by type, by level, etc.)
- Which countries are the training events being held?
- How did all of these change between 2013 - 2018?

**BONUS:** Qualitative data on specific trainings that can be turned into a story (images, etc.).

Figure 1: Initial sketch for our project.

### 3.3 Wireframe

The logical next step for us was to put all our data in Tableau and create a quick wireframe for a potential dashboard layout. The result is in Figure 2.

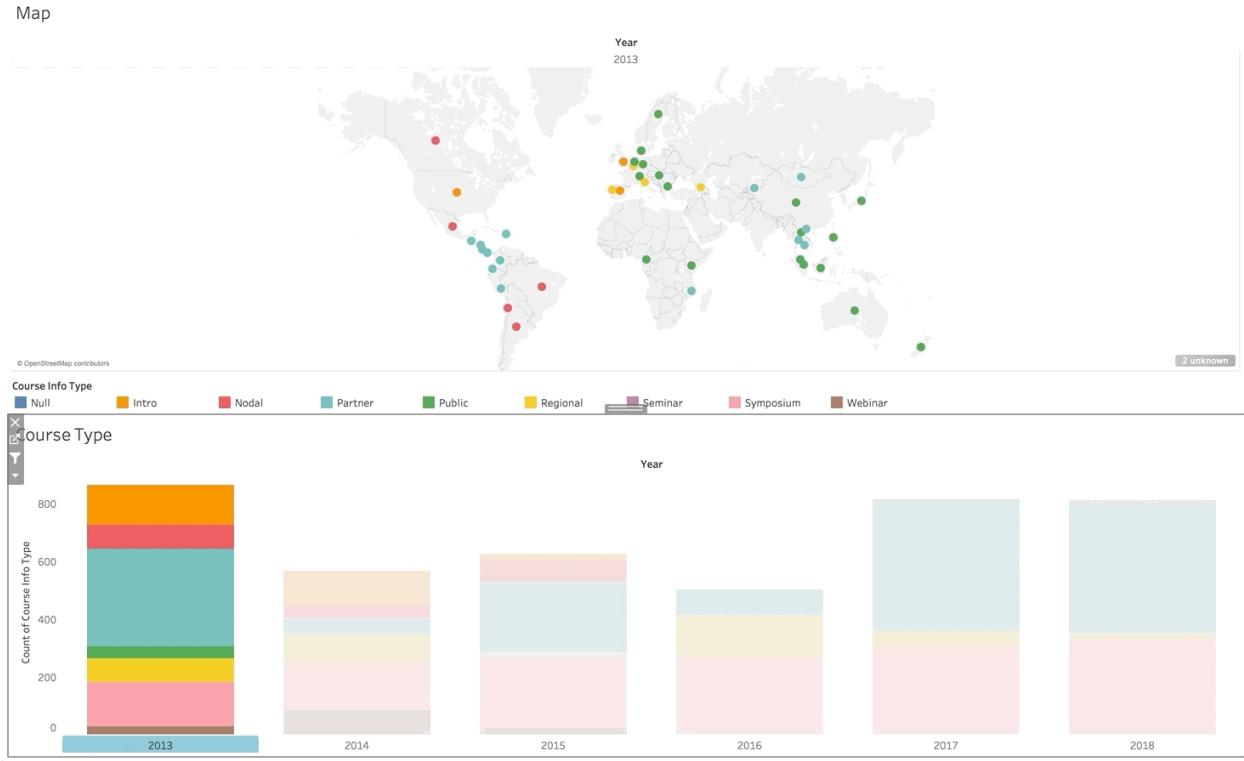


Figure 2: Wireframe made in Tableau for our initial design.

### 3.4 Prototype 1: Fullscreen Map

Based on our sketch and wireframe, our first implemented prototype was a full-screen map, the only functionality being a bunch of buttons that changes the data displayed on the map based on the selected year.

Although this prototype did not have much functionality, it allowed us to get feedback and critically think about our decisions. Namely, we decided we needed something more visually appealing and more practical than a full-screen map, since the map is not the main part of the visualization; the statistics are equally important so they should not be hidden away in a panel.

### 3.5 Prototype 2: Bootstrap Template

After our first prototype, we changed our plans to go with a dashboard layout instead of a full-screen map. For this, the first thing we tried was a Bootstrap template that we modified with our data. This is shown in Figure 4.

As seen from the copyright statement at the bottom, we of course could not use this for our actual visualization, but this inspired us to make our own dashboard layout

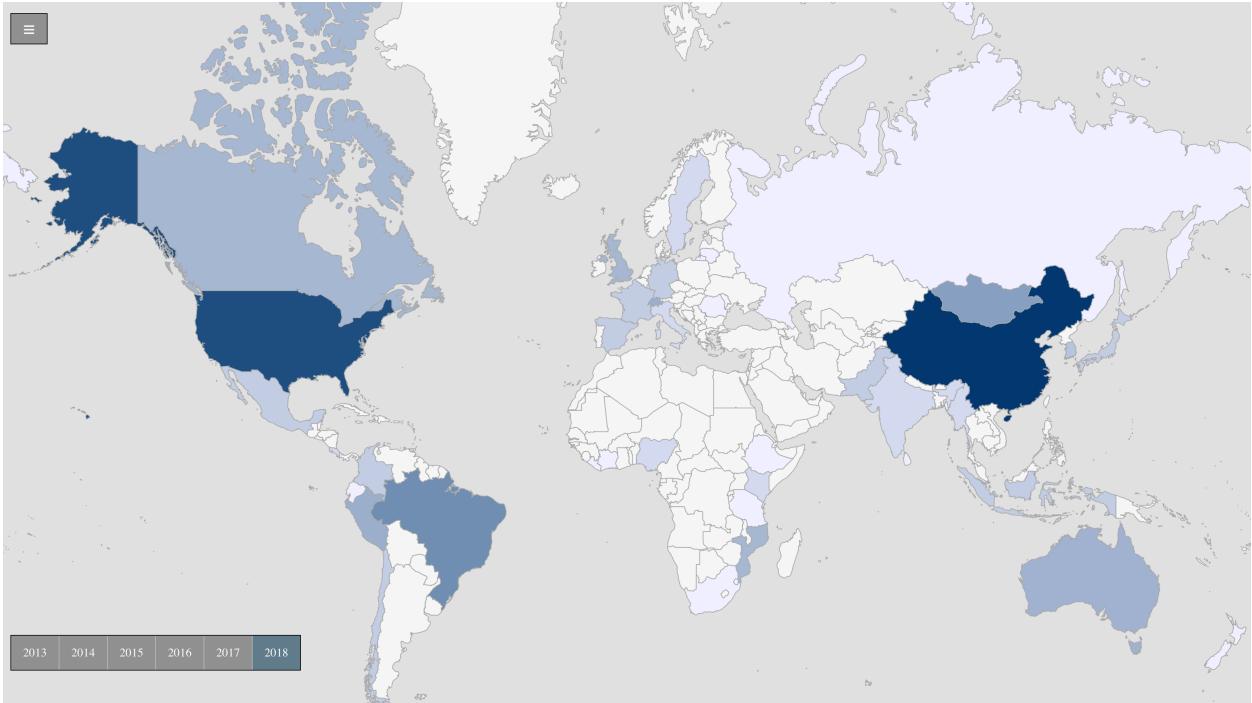


Figure 3: The initial map we made with the data.

from scratch. We also changed our map library at this prototype from DataMaps to JVectorMaps. The map is pannable and zoomable now as a result, with faster loading and possible selection/click events.

### 3.6 Prototype 3: Materialize Dashboard

We used cards from Materialize to make the layout in Figure 5 for our visualization.

Our first thought was to keep the whole visualization in one page and just link all the visualizations together. The cards at the top give the most relevant information directly to the user in the simplest way possible, and the charts and map allow further exploration. We were much happier with this design, but now we were lacking the story aspect of the project.

### 3.7 Final Design

After thinking long and hard about the story, we decided to follow the Martini Glass way of narrative storytelling. We turned the viz into a two page layout, the first page being author-driven and the second page allowing free exploration in a dashboard. We kept the informative cards and focused the first page on the main part of the training program, the symposium. Some textual description and a video give the whole context

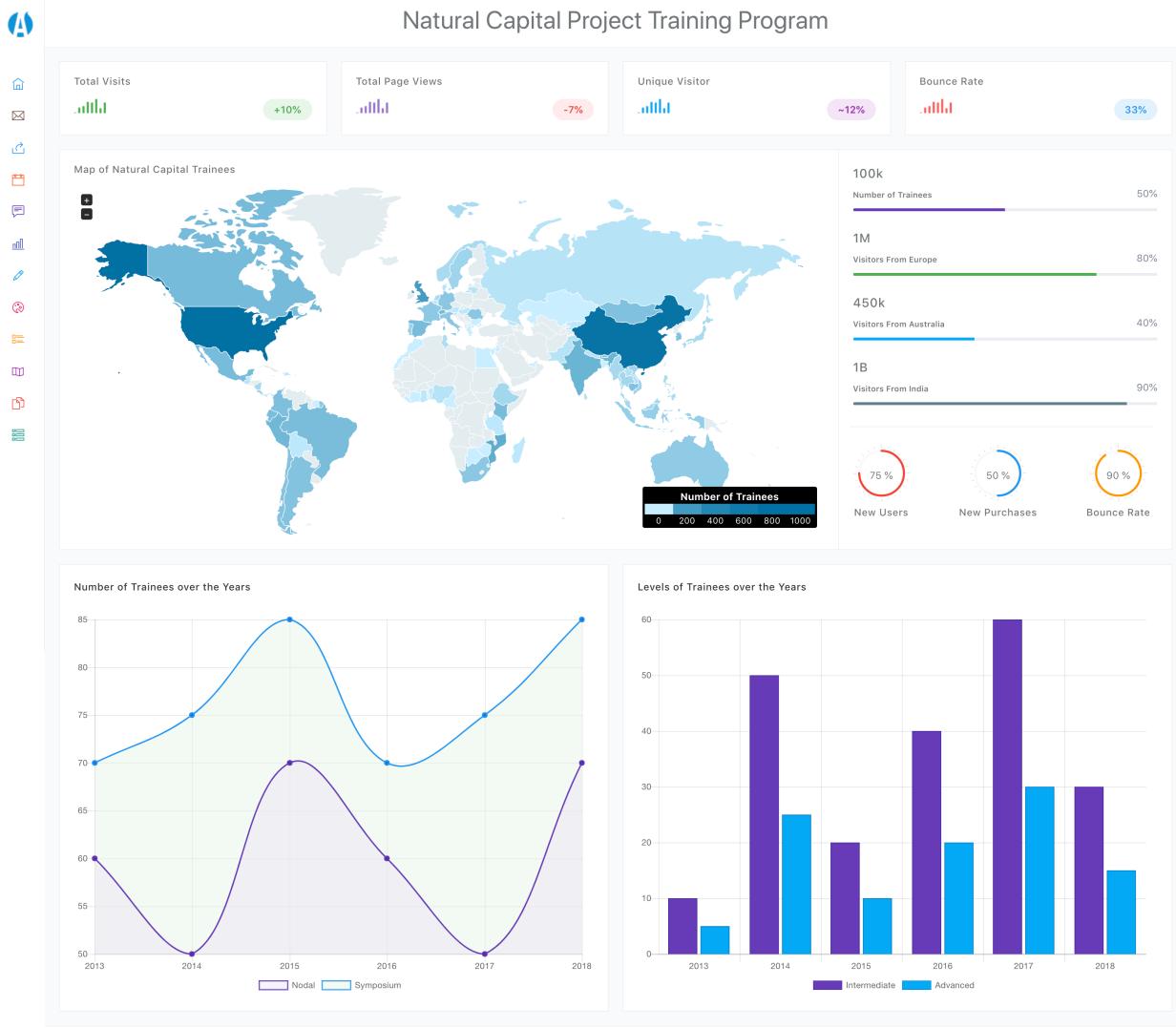


Figure 4: The bootstrap template we used.



Figure 5: Our initial dashboard design.

to the user, while our animated area chart shows the user the growth of the symposium over the years while also adding movement to the screen. The chart card has another tab that shows a stacked bar chart about the symposium and non-symposium trainees, showing another perspective of the data. The tabs also give the user a taste of the interactivity to come. After they take in the story, simply scrolling down leads to the interactive exploration dashboard.

In the second page, the user is greeted by a nice and large world map, two informative charts and an instruction card to help them make sense of everything that is going on. Every visual element seen on this webpage is linked and allows some interaction. The map allows clicking on different countries to automatically change the plots on the right side of the screen. The toggle switch under the map allows the user to choose between exploring the whole aggregated data and exploring its evolution over the years. When exploration by years is active, a slider appears that can be moved around to change the years, with visual cues to indicate the currently selected year. The data in the map and the charts automatically change as the slider is being moved. The final layout can be seen in Figure 6.

### **3.8 Visualizations We Considered**

In place of the stacked bar chart, we considered making a stream graph or a stacked area chart, but decided against it as the stacked bar chart was specifically requested and the stream graph idea was not well received by Natural Capital.

## **4 Implementation**

In this section, we explain the key design and interaction elements of our visualization, namely our card layout, the world map, different types of charts and the linking between these elements.

### **4.1 Layout**

We spent a considerable amount of time on making the layout both visually appealing and practical/informative. We found the Materialize library to be the most suitable for the task, it allowed us to create a functional dashboard that is flexible enough for all our purposes and easy to use, not to mention easy to make responsive. The final layout was shown above in Figure 6.

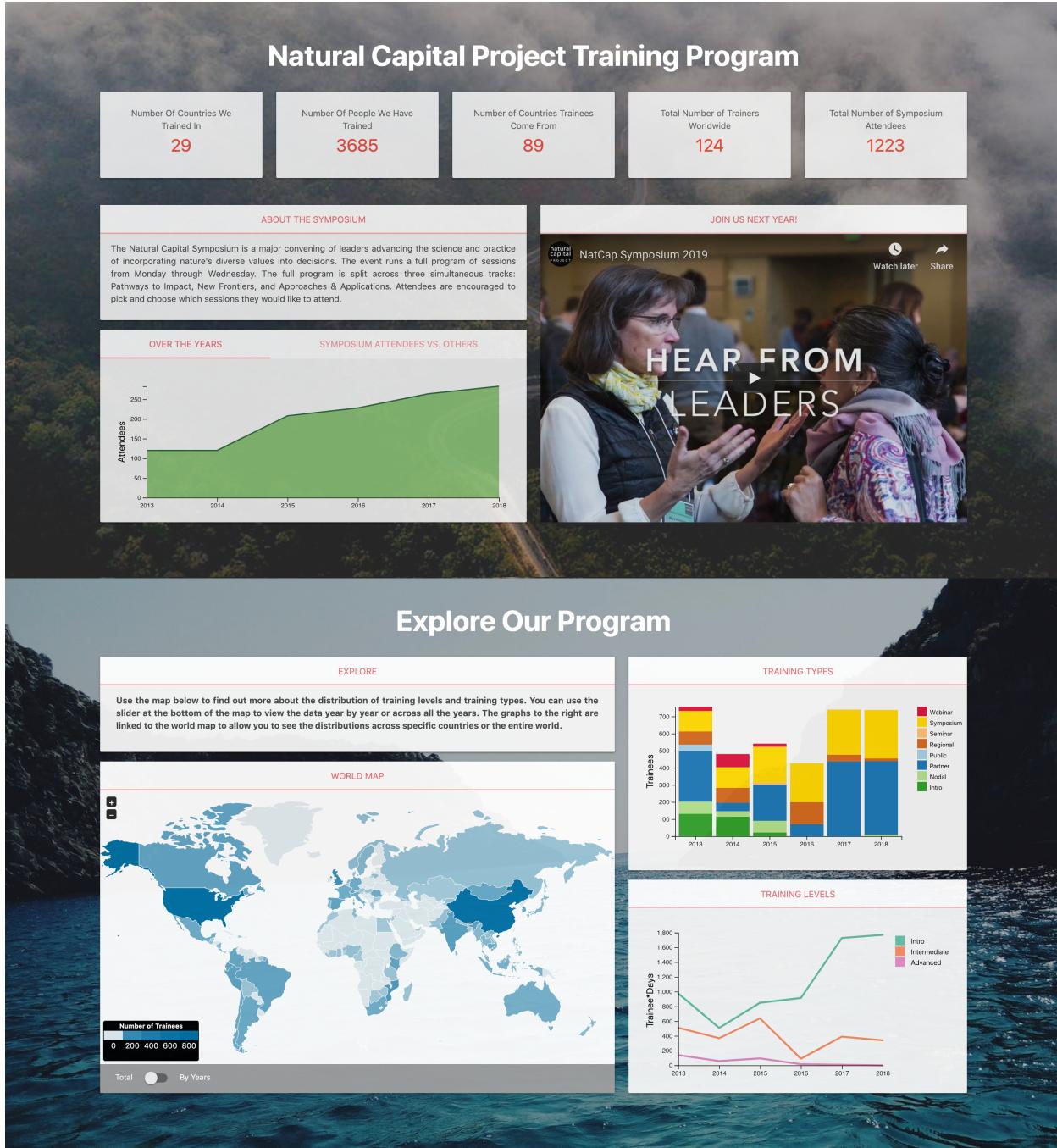


Figure 6: The final layout of our website.

## 4.2 Map

For the map, we used a library called JVectorMap. The Miller projection looks good in our card layout without being as extremely distorted as the Mercator projection. The library allowed us to make a Choropleth map of the trainees in each country, and also create events on clicking the countries. Since the data distribution is very

skewed many countries with only a few trainees and a few countries with a lot of trainees (e.g. USA and China), we opted to use a polynomial scale when coloring the countries. Even though this makes it hard to instantly tell the exact numerical difference between two countries, it makes it much easier to see relative differences between the countries. With a linear scale, most of the countries get painted one of two colors, which we wanted to avoid. The map also allows zooming and panning, as demonstrated in Figure 7.



Figure 7: The map zoomed into Europe.

## 4.3 Charts

In our visualization, we have an area chart, two stacked bar charts, two regular bar charts and a line chart (not all of them are on screen at the same time).

### 4.3.1 Area Chart

The area chart is the first chart that meets the user's eyes and it shows the growth of the Symposium over the last 6 years. Filling the area under the chart and the rising animation gives the feeling of growth as the user absorbs the information about the main training project of Natural Capital. Representative images can be found in Figure 8. The color was chosen as green to fit the theme of the project and the background of the first page.

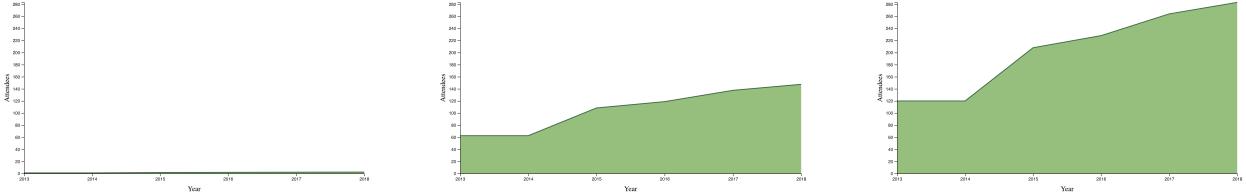


Figure 8: Rising animation of the area chart.

#### 4.3.2 Bar Charts

Right after the area chart, the user is presented with a stacked bar chart emphasizing the place of the Symposium among Natural Capital's events. The chart can be seen in Figure 9. The stacked area chart is useful in the sense that it makes it easy to see the evolution of both the ratio and the total numbers through the years. Since this plot is also on the first page, we kept the colormap green for visual consistency.

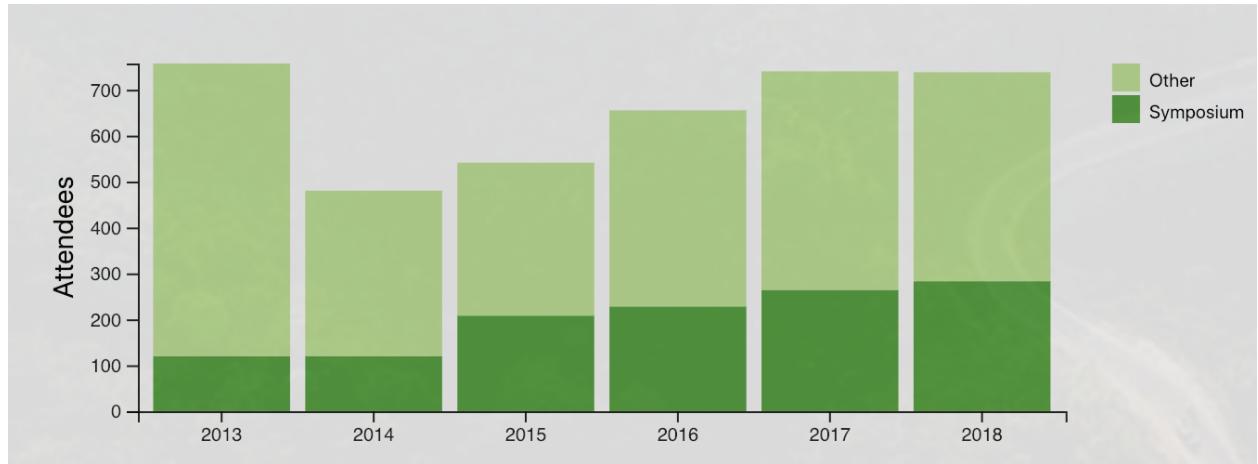


Figure 9: Stacked bar chart that compares symposium/non-symposium attendees.

The second bar chart on our visualization is more ambitious, it shows the proportions and evolution of all the training types in a condensed manner. This chart was specifically requested by Natural Capital, so we include it even though it does not follow best practices regarding the number of colors and groups shown in one plot. It can be found in Figure 10. To try and make the comparisons easier for the user, we also show the values of the part the user hovers on.

When visualizing the data for a single year, the stacked format loses its usefulness and regular bar charts are used instead. This allows much easier comparison between the numbers of trainees attending different types of classes. An example is shown in Figure 11.

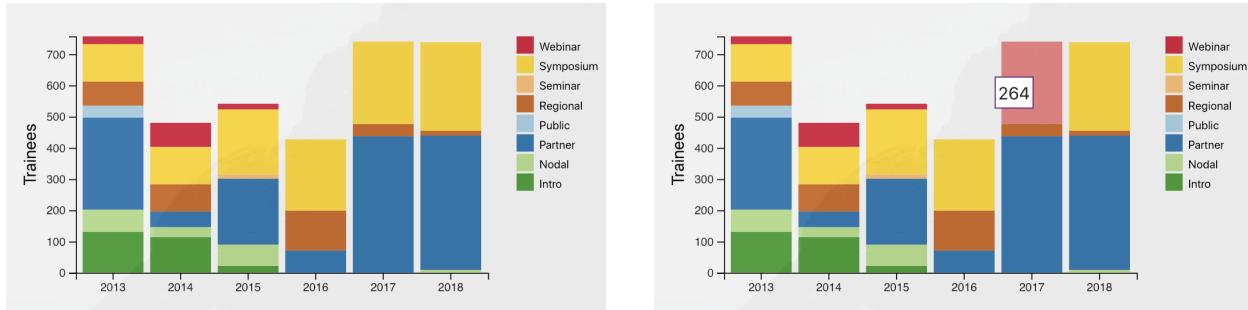


Figure 10: Stacked bar chart that shows all the training and its hover tooltip.

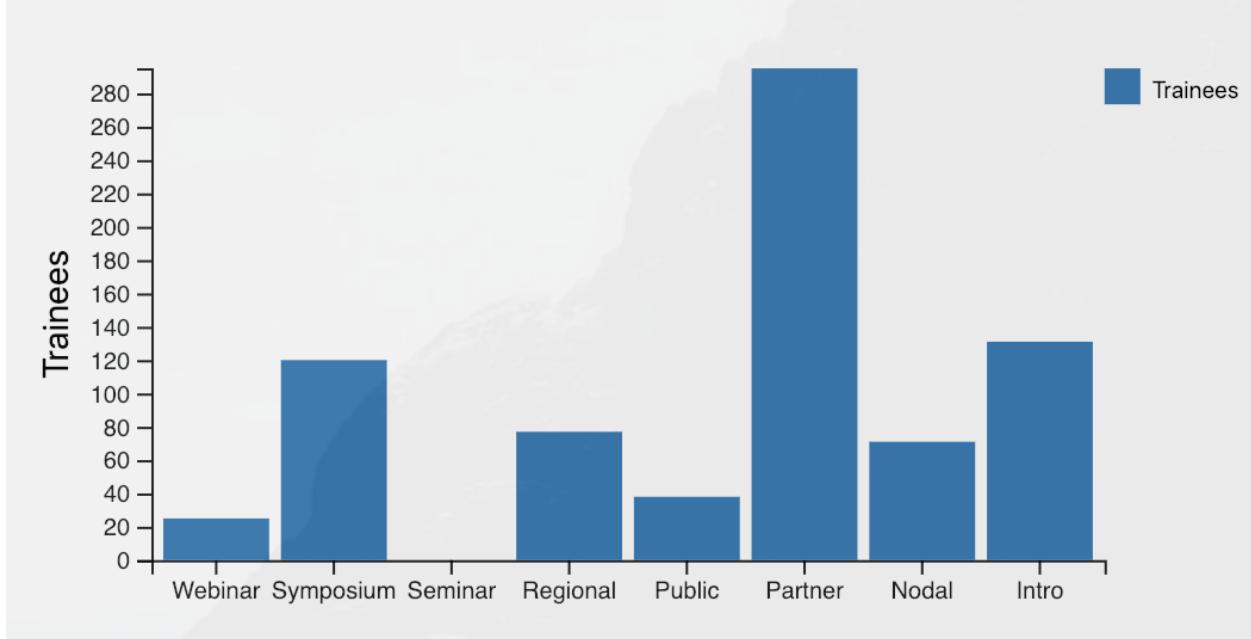


Figure 11: Regular bar chart that shows the training types over the year 2013.

### 4.3.3 Line Chart

The line chart is used to show the evolution of the levels of trainings attended by trainees over the years. There are only three levels so the three lines are not hard to follow by eye, but to maximize convenience we also show the datapoint values on hover and fade the other lines to help the user focus on the level they are interested in. This effect is demonstrated in Figure 12.

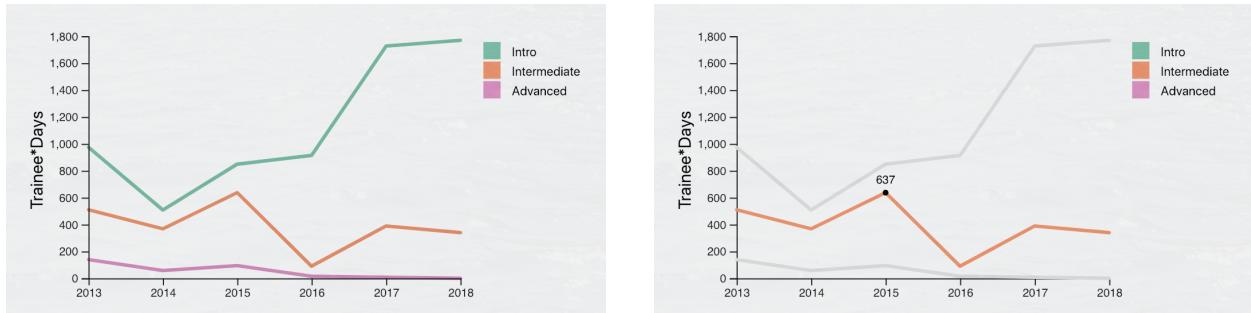


Figure 12: Line chart that shows all the training levels and its hover effects.

It should be noted here that this chart is not on the same scale as the bar chart above, as indicated by the y-axis label. Trainee\*days is a measure requested by Natural Capital that shows the total number of days attended by trainees. In this measure, if a trainee attended a training for 3 days, they count as 3.

## 4.4 Interactions

The dataset allows two degrees of freedom for exploration: location and time. For location, we extracted statistical data for each country separately so that we could allow the user to see the distribution of trainees over training types and levels and over the years. For time, we have data for each country for 6 years from 2013 to 2018.

Our intention was to utilize both of these degrees of freedom by letting the user change countries at will by clicking on the countries on the world map and change years by using a slider that spans the 2013-2018 range.

### 4.4.1 Country Interaction

Clicking on a country (that has trainees) on the map triggers a redraw on the two charts on the side. The training type and level distributions are still displayed over the years, but only for the country selected. Clicking on the selected country again deselects it. The change can be seen in Figures 13 and 14.

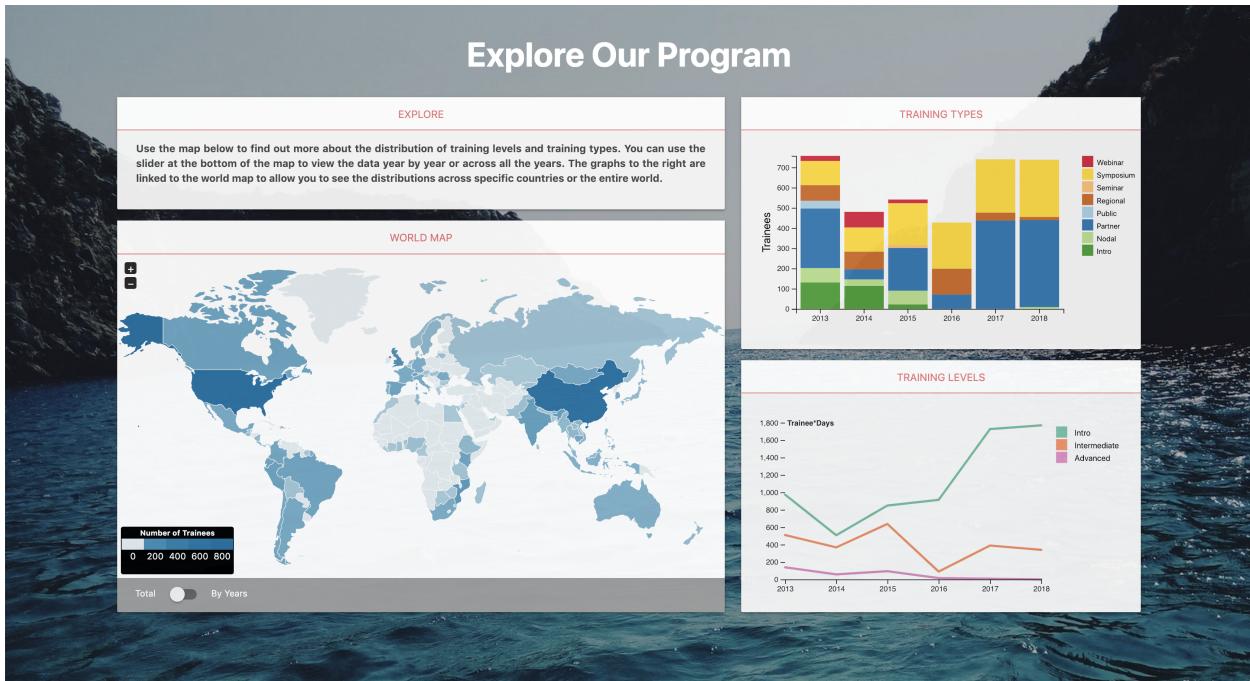


Figure 13: The exploration page when no country is selected.

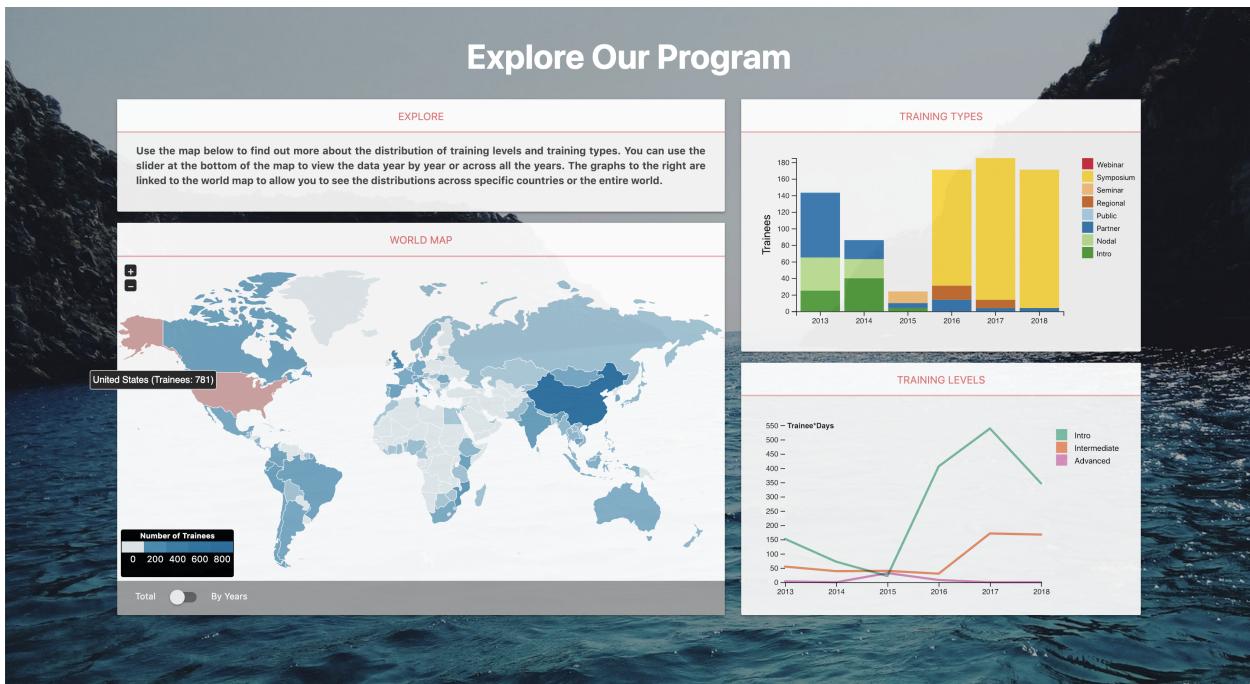


Figure 14: The exploration page when USA is selected.

#### 4.4.2 Time Interaction

If the user wants to observe the changes in the program over the years, they can simply toggle on the 'By Years' switch and a year slider will conveniently appear before them. By default, this toggle will set the year to the most recent one, 2018, and update the map and charts accordingly. The currently selected year is also displayed to the user to avoid any confusion.

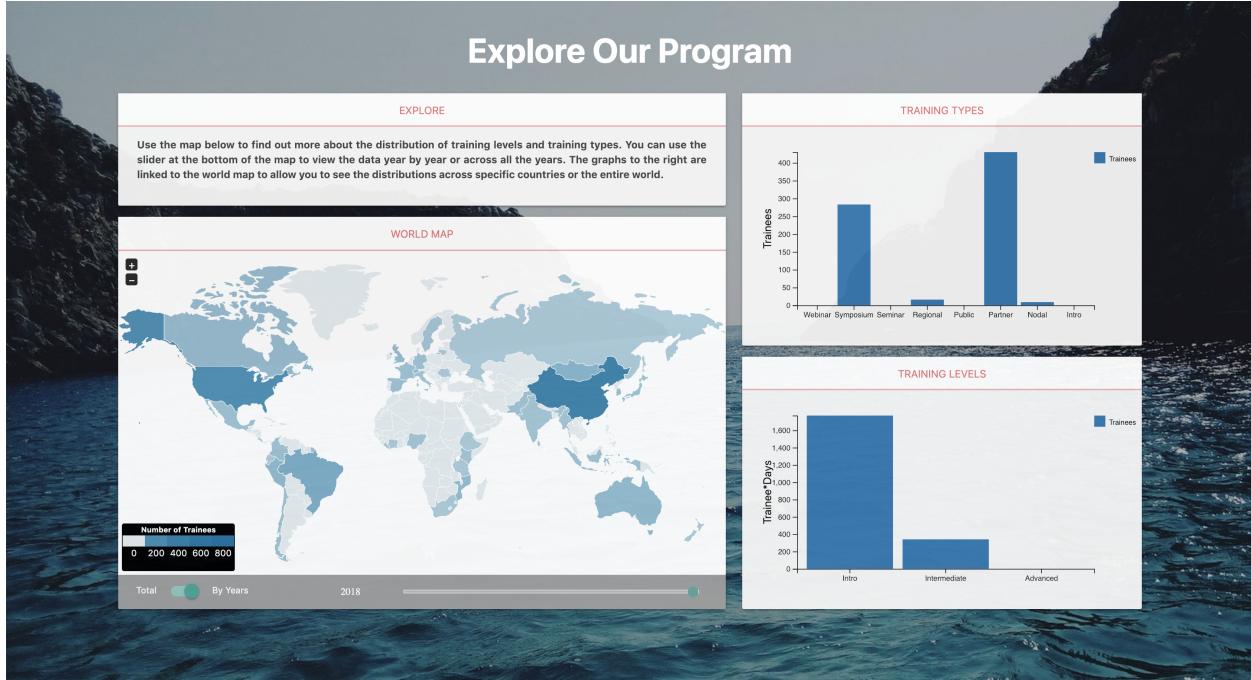


Figure 15: The exploration page when the 'By Years' toggle is switched on.

The important thing to note here is the change in chart types. When the user selects a year to explore, our data is filtered in a way that loses the time component, so the line chart and stacked bar chart do not make sense anymore. Instead, the bars that make up that year's stack are plotted as a regular bar chart. Similarly, the data in the line chart that corresponds to the selected year are displayed as a bar chart. The change can be seen in Figure 15.

#### 4.4.3 At the Same Time

We also allow the user to filter both by country and year simultaneously. This draws bar charts for that country and only for that year. This effect is demonstrated in Figure 16.

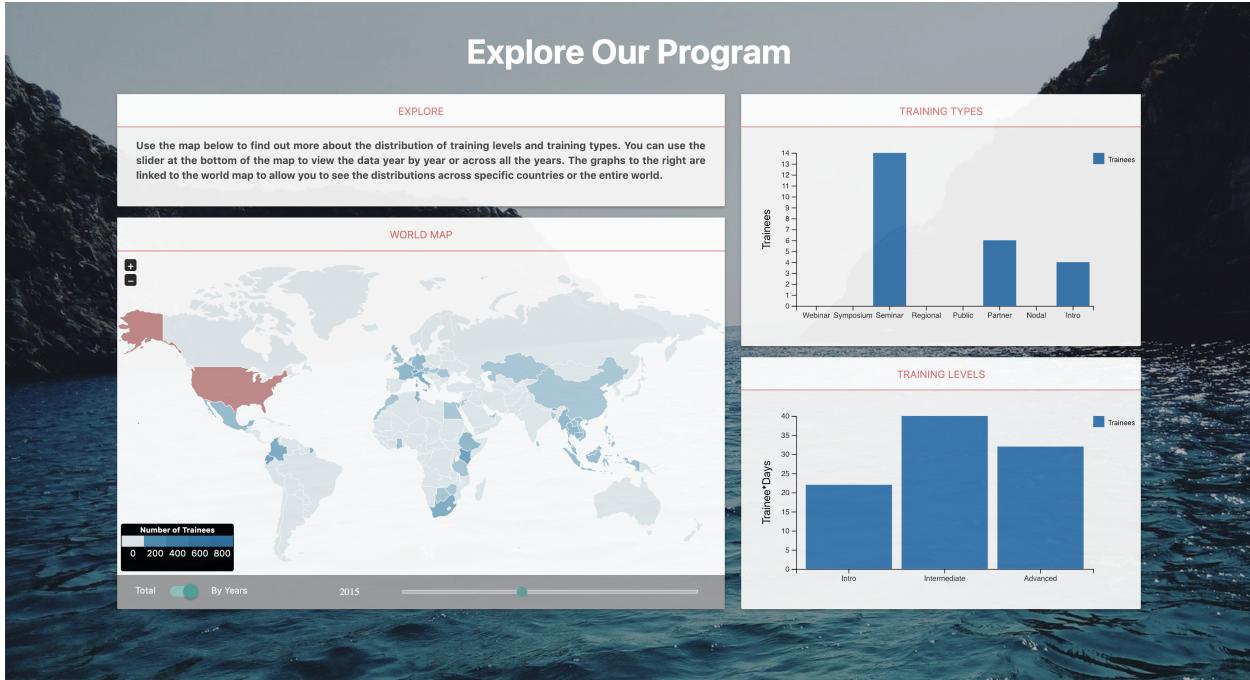


Figure 16: The exploration page when USA and 2015 are selected.

## 5 Evaluation

### 5.1 Data Insights

By looking at the distribution of levels of trainings over time, we saw that there was an increase in the overall number of trainings taken. There was clear growth in this area demonstrating that there was an increase in interest for taking part in the training sessions. We also took a look at the total number of symposium attendees over the years and also saw consistent growth over time.

### 5.2 Future Work

In the future, if we are provided more data, we would suggest adding to the world map as well as including more qualitative data. With regards to adding to the world map, we believe that some meaningful additions include geographically accurate bubbles (indicating the location of trainings) and the option to include directed arrows to intuitively show the global movement of trainees. With regards to the qualitative data, we believe a good place to show pictures of trainings and events held by the Natural Capital Project is an interactive pop-up card that is displayed once the user clicks on a training at a particular location. This would act a small window into the actual training that is occurring at that specific location. We believe that these small

additions to website, once the data is collected, will bring tremendous value to the experience of the website.

## **6 Peer Assessment**

In terms of preparation, respect for others' ideas, and flexibility, every one of our members was respectful of others' opinions and ideas and we worked together very well. The following is the breakdown of the contributions:

Doga - Worked on the map on the website, the line graphs on the website, and the linking of the map and the graphs.

Shikhar - Worked on the preprocessing of the data, the bar graphs on the website, and the linking of the map and the graphs.

Varun - Worked on the preprocessing of the data and on the layout and design of the website.