Process Book

Course Project of Data Visualization, COM 480, EPFL

Dashboard Viewer of Water Yield Model results in Myanmar

Cai Fengyu, Chen Liangwei, Loureiro Claudio Leao

Fall 2018

Human nature is like water. It takes the shape of its container.

Wallace Stevens

Introduction

The main objective of this process book is to provide the readers with a brief overview about a web-hosted visualization tool about the season water yield evolution and state-wise distribution in Myanmar, a southeast Asian country. Also, the readers could find out the further prediction about the water yield based on the current climate information.

Also, it is part of the course project of COM 480 Data Visualization at EPFL, 2018. And the practical skills and the discipline of data visualization in this course creates the possibility for us to design and implement our product, both technically and theoretically. We hope this project could be a milestone to record our growth in data visualization from zero.

Background

As human beings are part of ecosphere, many of our life and production can be vitally influenced by the chronological and regional variation of ecosystem. And many non-profits and international lending institutions and corporations are focusing on managing natural resources and evaluating the tradeoffs, and Natural Capital (NatCap) Project mainly led by Stanford University, the Nature Conservancy, and World Wildlife Fund is one of the pioneers. They are working to integrate the value nature providing to society into all major decision making, through the establishment of analytical modeling softwares, which is indispensable. One of its products, Integrated Value of Ecosystem Services and Tradeoffs (InVEST) works as a suite-of-free, open-source software models used to map and value the goods and services from the nature that sustain and fulfill human life.

Motivation

The NatCap training program has achieved great success to share the research result beneficial for decision making. However, most of the time the trainees have to utilize InVEST and other GIS to create and view the output, which is difficult and complicated for them to get familiar with these highly-integrated tools in a short time. Therefore, it is necessary to develop some educational tools for specific topics to simplify the training process and improve the educational result. And this tool should hold the following properties:

1. Easy to learn and control
2. Convenient to access
3. Clear visualization of the input data
4. Display the analytical visualization to indicate the potential insight

Therefore, we decided to choose data of specific natural model and implement a fully-usable web-hosted dashboard. On the one hand, it could offer the opportunity for users to comfortably control and interactively manipulate the map. On the other hand, the users also can find out the insightful statistical analysis through data visualization.

Data Selection

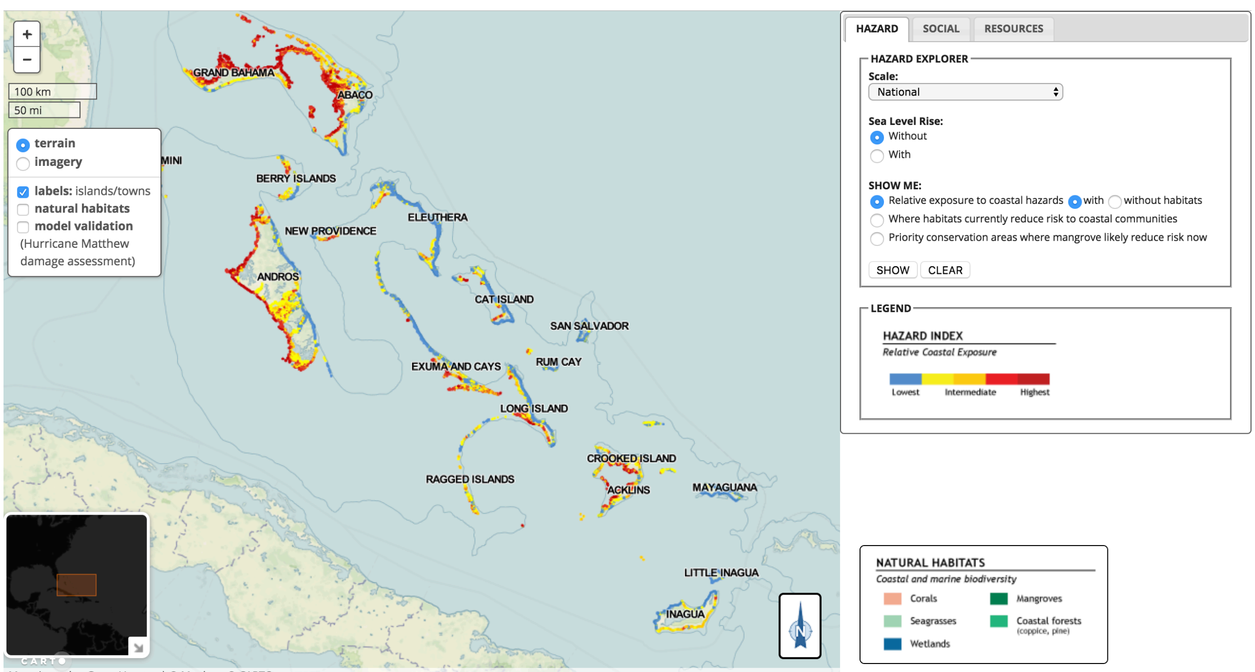
From the multiple datasets provided by NatCap, we finally chose the Seasonal Water Yield (SWY) Model which has been used in Myanmar national ecosystem service assessment technical report (We will briefly name it as 'report'). We will briefly introduce the data format later.

Target Audience

As mentioned above, our tool is purposed to be the substitution of other highly-integrated tool like InVEST in training program of NatCap. Therefore, our target audience are the trainees in studying the seasonal water yield of Myanmar. Through interactively manipulating our product, we hope they can quickly be familiar with the SWY model, and directly find out the result.

Inspirations and Related Work

After our search about the related works, we have got several inspirations from them. Also, we also found some drawbacks of their design which we need to avoid in our own. Here, we briefly mention two of them provided by NatCap.

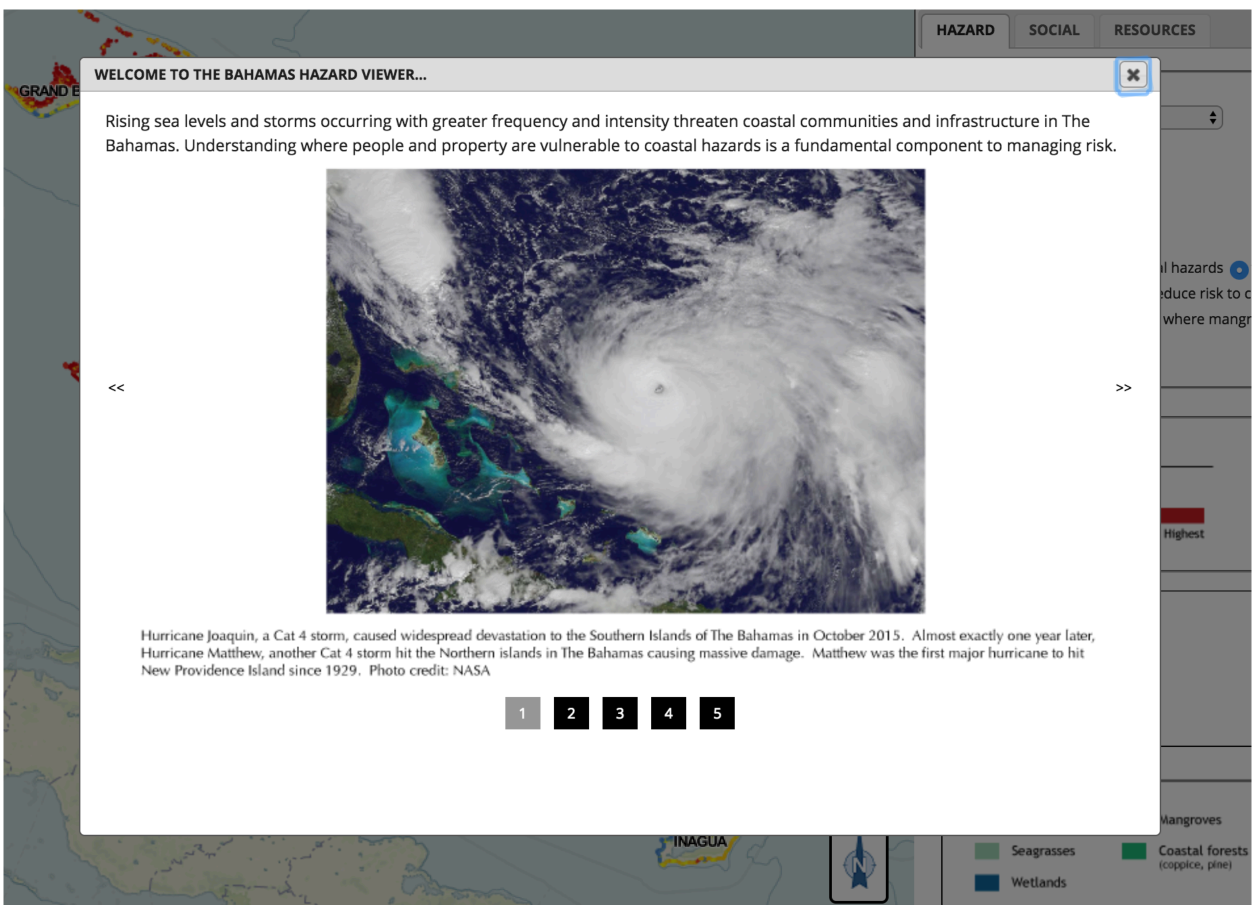


1. The natural resource study in Bahamas. (figure above)

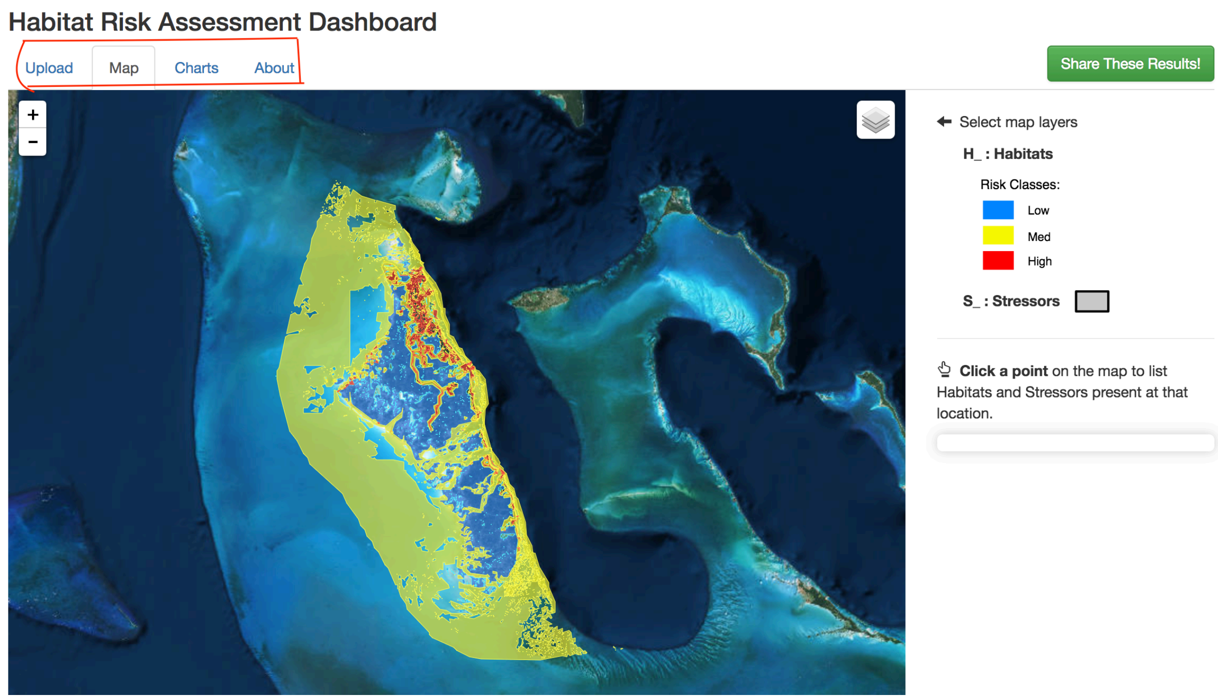
In the NatCap visualization tool, the numerical value (Hazard Index) has been categorically displayed at the map. The legend describes the measurement of the value.

When we look back to our data, we find this kind of design may not be suitable enough, as our data is continuous and widely ranged. Therefore, according to the content in the class, we decided to use the heap map to represent our values on the map.

Also, another inspiration we attained from this work is the guideline window. For the conveniency of the users, we need to provide the background introduction and utility instruction. However, there is the drawback in this project, which is that after users close this guideline, it can not be opened any longer, except refreshing the whole website.



2. Habitat Risk Assessment Dashboard



When trying on this one, we found that they put the chart analysis and the map in different tabs, which is not possible to let users do the map observation, panel controlling, and get the statistical information in the same time. We would like to provide this in our design.

What is more, both of their targets are relatively small, a region, or a small-sized country like Bahamas. However, our research target, Myanmar, is the second largest country in southeast Asia, just smaller than Indonesia. Studying it as a whole is not suitable What is more, from the preprocessing of the data, we find that the similar position shows quite similar tendency and result. Therefore, we decided to add the function to narrow down our visualization to *state-wise* analysis. We added the state label on the data based on the administrational division of Myanmar.

Conception and Dataset Description

Main Conceptions

Three main indices can be computed by the model, which are quickflow (QF), baseflow (B), and local recharge (R), and only first two have been used in the report, and this will continue the same for our project.

Definitions of QF, B and related conceptions are defined as following:

1. Quickflow represents the amount of precipitation, that is converted to direct runoff, entering stream soon after a rain event.
2. Baseflow represents the amount of precipitation that enters stream under subsurface flow.
3. Land Use is commonly defined as a series of operations on land, carried out by humans, with the intention to obtain products and/or benefits through using land resources.
4. Land Cover is commonly defined as the vegetation (natural or planted) or man-made constructions (buildings, etc.) which occur on the earth surface. Water, ice, bare rock, sand and similar surfaces also count as land cover.
5. The abbreviation of 'land use and land cover' is LULC

Dataset content

1. Baseflow historical change data

Baseflow results, were presented as the difference between the baseflow provided by the current LULC and the all-agriculture LULC.

2. Quickflow data

Quickflow data is more complex because it is given for each month for given parameters. The input parameters are the following:

a. Representative concentration pathway (RCP): two values are available in our dataset 8.5 and 2.5 where RCP 8.5 scenario corresponds of high greenhouse gas emissions and 4.5 to lower greenhouse gas emissions.
b. Percentile: Represents the values of climate model distribution, two percentiles are available in our dataset. 25th and 75th percentiles to show respectively low and high values of precipitation, evapotranspiration and rain events predicted from climate models.
c. Time slice: Represents the period of the measured quickflow.
d. Allag: The presence of the abbreviation means that we consider the all-agriculture LULC for the current dataset.

For example, qf_rcp_85_2020_25 means the 25 percent quantile value among the predicted quickflow in the year of 2020 assuming the RCP value is 8.5, meaning the greenhouse gas emissions is high at that time.

In our data, we provide historical quickflow data, and the predictions based the climate model have been shown as followed. One of the main task of our project is to visualize these distribution geographically and chronologically.

I. Historical QF
II. 25% quantile of Predicted QF in 2020 with RCP value 4.5
III. 75% quantile of Predicted QF in 2020 with RCP value 4.5
IV. 25% quantile of Predicted QF in 2020 with RCP value 4.5 considering all LULC
V. 25% quantile of Predicted QF in 2040 with RCP value 4.5

VI. 25% quantile of Predicted QF in 2040 with RCP value 4.5 considering all LULC
VII. 75% quantile of Predicted QF in 2020 with RCP value 8.5 considering all LULC
VIII. 75% quantile of Predicted QF in 2040 with RCP value 8.5 considering all LULC

Dataset structure

A first look at the dataset tells us that it is mostly composed of GEOTIFF/TIFF, TFW and XML files.

-GeoTIFF: This is the main data file. This format allows georeferencing to be embedded with a TIF file (computer file format for storing raster graphics images).

-TFW: Some tiff files are not directly georeferenced thus this format file is needed to indicate the reference on the map.

-XML: In our dataset these files indicate basic statistics of the current geotiff/tiff file.

Preprocessing steps

For processing our GEOTIFF files we used GDAL (Geospatial Data Abstraction Library). It is a translator library for raster and vector geospatial data formats. This allowed us to convert to others coordinate system since the actual one was not correct for our map visualization.

No preprocessing

As a first step, we started by trying to work with data without preprocessing it. But since each geotiff file representing our main raster file, is very large (around 340MB). It was almost impossible (the website crashed) to display it on a map on our website (in the correct georeferencing positions) without preprocessing it. By loading the raw data, it was almost impossible to move the map without waiting several seconds. A solution needed to be produced to make it smoother.

Map geotiff preprocessing

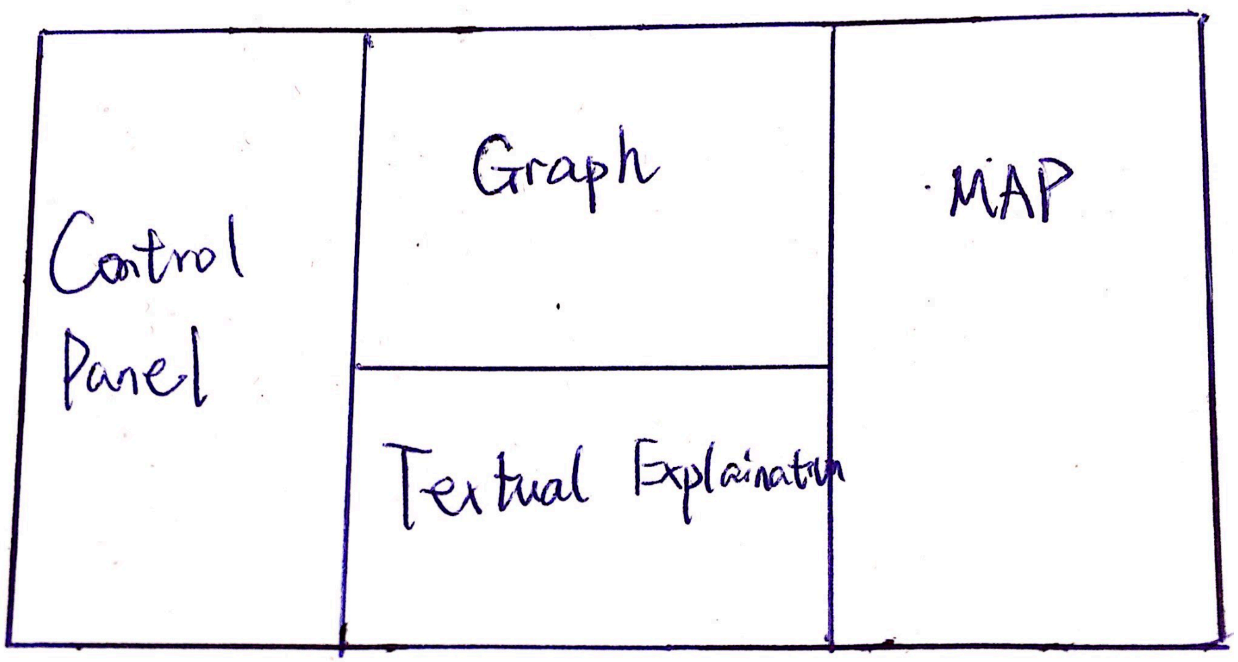
Using this library, we were able to decompose our geotiff file into several tiles. This allowed us to have different tiles for different zooms level. This make our map smooth and it was now possible to move map zoom/dezoom without any problem or making it crash.

Heat-map display preprocessing

We are displaying heatmaps for quickflow values. But as before we cannot use the raw data directly. To implement this task, we first down-sampled the tiff files using an average method (provided by GDAL). Then the downsample data was converted this into CSV format (thus the important task of down-sampling it) otherwise the file is too large. The next step was to convert this into an array that Javascript can handle.

Design

In order to reach our objective we set previously, the whole webpage would contain 3 parts: a control panel, a map representation, and a statistical analysis window (with text explanation, it is straightforward, we won't mention later). Considering the shape of Myanmar, in order to utilize the space efficiently, we let the height of the map larger than its map.



And the control panel is focusing on the user interface, and statistics analysis will be the part of visualization. Therefore, in the part we step with the following workflow:

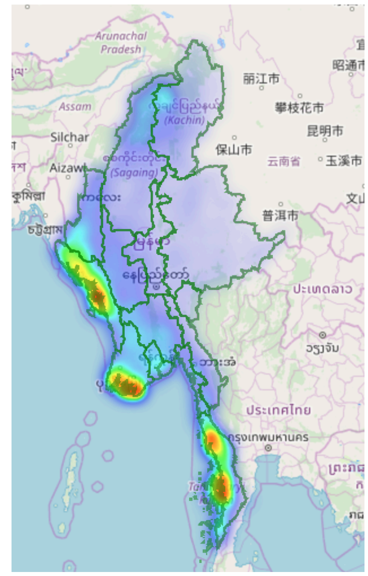
1. Visualization design: how the data can be represented on the map and the statistics window.
2. User Interface design: how users could interact with the visualization through the control panel or directly on the visualization.
3. Other Designs: some minor and user-friendly design details.

Visualization design

After our observation, each data point contains its numerical value with the position, longitude and latitude. And also, the QF data always contains 12 months. Basically speaking, our data has two dimensions, in both space and time. Therefore, we can combine them two by two.

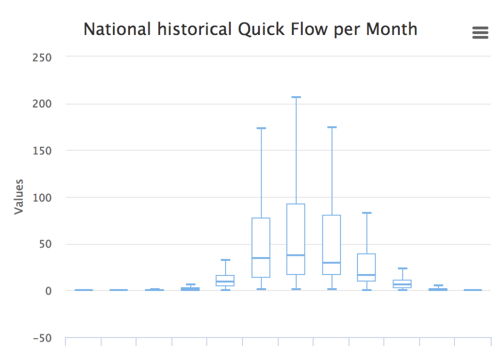
1. spacial display - map

After our analysis, the choice for this pair should be very straightforward. The normal methods for the map display include: the scatter map, the bubble map, and the heat map. The scatter map is mainly focusing on the visualization of distribution, while the bubble map is good at the clustering analysis. Therefore, neither of them are adequate for our goal. Meanwhile, the heap map can show the numeral distribution based on the geometrical variation.



2. temporal display - statistics analysis window

There are a lot of format to show the chronological variation. For example, line chart is the most direct one. However, for each month, what we get is a huge set of values, instead of a single. It is biased and over-simplified to use some statistics like median and mean to represent the monthly distribution. Therefore, in this case, the box-plot will be good alternative. Besides the single statistics, the box-plot can also offer the rough structure of the distribution.



The two pairs mentioned above are comparatively simple and straightforward, because as instincts, map is expert in representing the geometrical information, and axis is good at the chronological representation. However, spatial - statistical analysis window and temporal - map pairs are slightly much more complicated.

3. spatial display - statistical analysis window

The histogram is the first method popping up to our mind. Using the histogram, users are able to figure out the numerical distribution in the whole country or in the specific region. However, we found that there may exist outliers making the graph not readable, therefore, we will set the histogram range limited to a specific region.

4. temporal display - map

It will be really difficult to show the temporal variation on the static map. Therefore, we decided to add an animation on the heat-map. The map will move from January to December automatically. And it shows that the combination of heat map and timeline animation works perfectly.

To summary, based on the properties of our dataset and functions, we could finalize our data visualization as the following table.

	Temporal Domain	Spatial Domain
Statistical Analysis Window	Boxplot	Histogram
Map	Animation	Heat map

What is more, as mentioned, we added the visualization on the state-wise comparison. We still used the box-plot, with x-axis the list of states.

User Interface design

From the visualization design above, we could basically have a draft of ideas about how to design our user-interface site.

1. Firstly, we talk about the main part of User Interface, which is control panel. The main components and its corresponding function has been summarized in the following table.

Control Panel Design	Sub-components	Format	Main Function
Data Selection	Baseflow	Button	Switch mode to Baseflow
	Quickflow	Button	Switch mode to Quickflow
	Non Data	Button	Switch to pure map
State Selection	Selection	Dropdown	Choose the canton
	Search Input	Input	Input state name
	Search	Button	Switch Stats Window and Map to the searching state, zoom in the map and place the marker
	Summary	Button	Compare the states based on the current data
	Reset	Button	Reset and back to nation mode
Time line	Tileline	The combination of SVG	Display the month in a line
	Play/Pause	Button	Start/Pause the animation
	Reset	Image	Back to last month

And we will arrange all of these component downwards.

2. For the statistical window, we should make the samples in the plot interactive with the mouseover. Also, under it, some textual explanation is also needed.

3. We need to add the several features on the map.

a. The border between the states should be drawn.

b. Once the event of mouthover, the region of state should be highlighted.

c. Under the Baseflow mode, the information box should be placed to indicate the value, the current state, and position.
d. The legend should be able to respond the change of the data.

Other minor and user-friendly design details

1. In case the user get lost in our tool, we developed a guideline popup. It will pop up when the website is initialized. And if the user want to double check it, one button should be clickable for this function.
2. For the state search function, an autocompleting function should be designed for the user's convenience.

Implementation

In this part, we will briefly introduce the implementation of our tool. Similarly with the previous part, we will separately talk about the control panel, the statistical analysis window, and the map. For the further reference, a Github repository is provide here (https://github.com/TRUMANCFY/Data_Viz_2018). And at the end, we will briefly talked some difficulties met during our project, and some attempts we have tried.

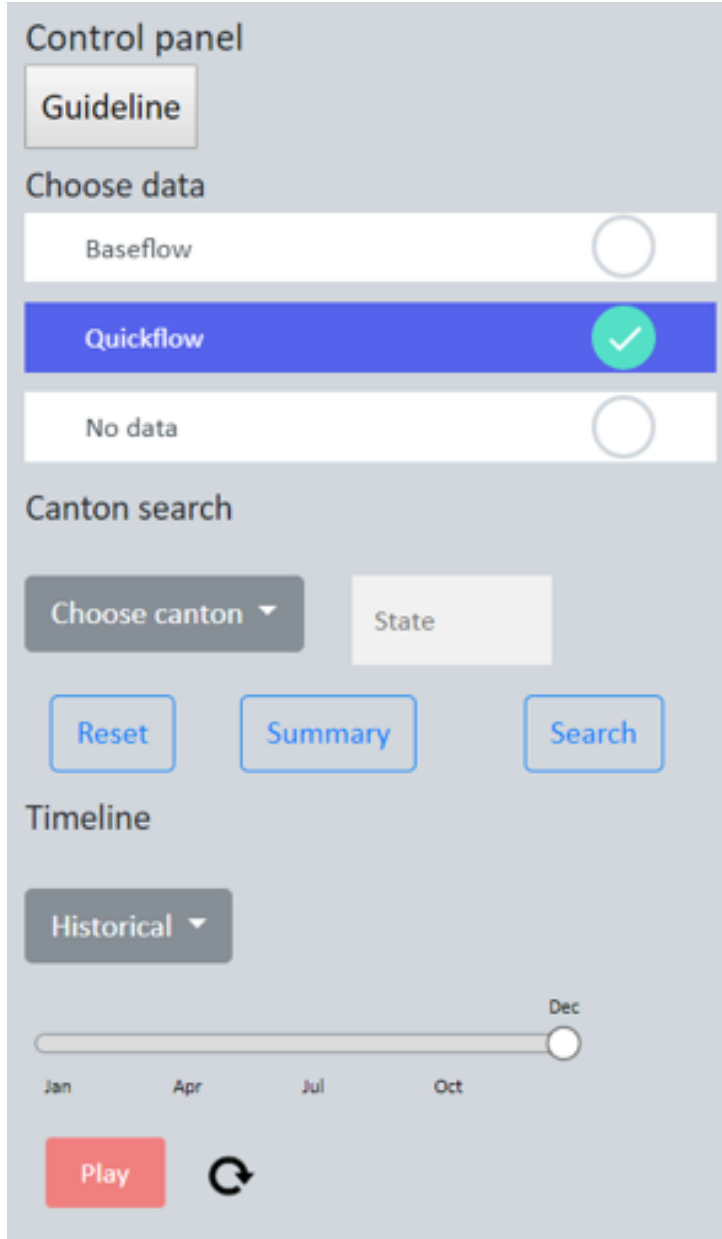
Control Panel

In this part, we use the style libraries like Bootstraps, and the visualization library like D3.js

1. Guideline: the guideline button opens a user guideline popup using jQuery.
2. Data Button: The style of the button is from Bootstrap
3. State selection: two methods of state selection have been provided, and the users can both search and select from dropdown directly.

3.1. An autocomplete tool has been implemented for the typing input.

3.2. Once search button is clicked, if the input is illegal, namely, the input cell is empty or the state name does not exist, it will pop up the alert. Otherwise, it will zoom the map to the corresponding state and place a marker on it, and also the statistical part will respond to this event.



4. The timeline tool is only available when selecting the quickflow dataset.

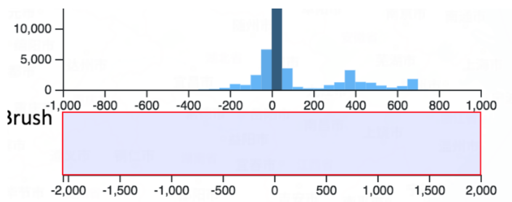
4.1. It seems this part is most complicated SVG part in our project. Its base is a slider object. It owns a movable circle and a line, where the circle could move through the line. Through the scale transformation, we can get the month from the circle's position

4.2. the Play button will let the iteratively call an update function to let the circle move forward (using setInterval()). And once clicked, the play button will be rewritten to the Pause button, which can work similar to the Play button

4.3. Reset button just stop the circle movement, and set the circle's position at the end of the line.

Statistical Analysis Window

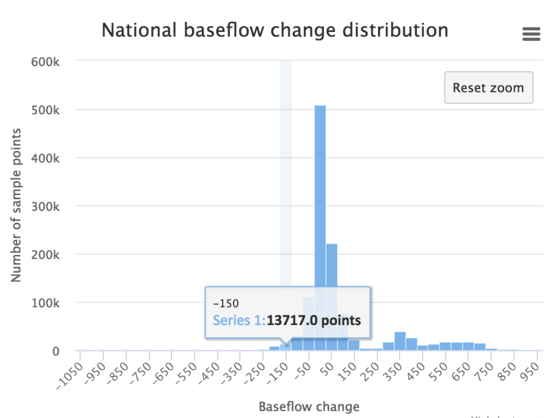
At first, we do this part of visualization using d3.js with a brush (just like the figure shown below,). However, it looks meaningless in the practical usage. And we just quitted the application of the brush. Later, We found an external library called HighCharts, which can help us produce more neat and precise interactive chart.



And here data preprocessing for the statistical analysis is still heavy buttoned, based on the huge size of the dataset. Therefore, instead of javascript, we used the packages in python, like numpy and pandas, to do the preprocessing like filtering the illegal data.

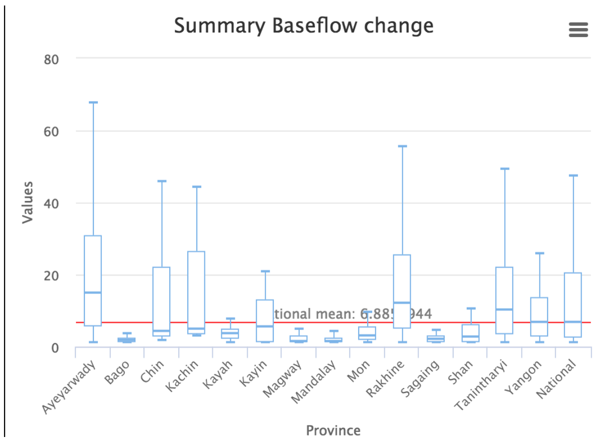
We would like to show some of our screen shot of this part, and discuss in details what we have specially implemented.

1. Baseflow:
    1.1. National histogram of baseflow is look like below, and the reset zoom button is designed to free the histogram back to the original distribution.



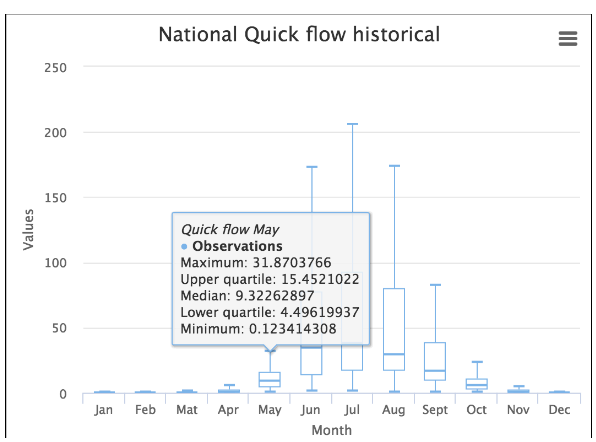
1.2 The state-wise baseflow histogram looks similar to the national wide plot.

1.3 The state-wise comparison also applies the boxplot to indicate the rough distribution in different states. And what is more, we also added the mean value as red line to indicate the comparative relationship between an individual and the mass.



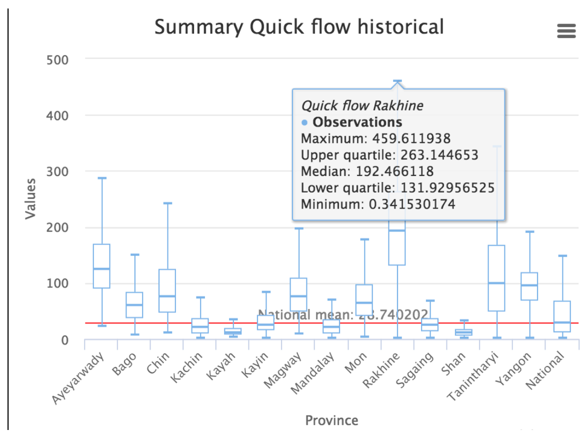
2. Quickflow:

2.1 National wide quickflow boxplot is shown as followed , as there is only 12 months, we do not need the interval zoom any longer. The statistical information will be shown if the mouse is floating over the box.



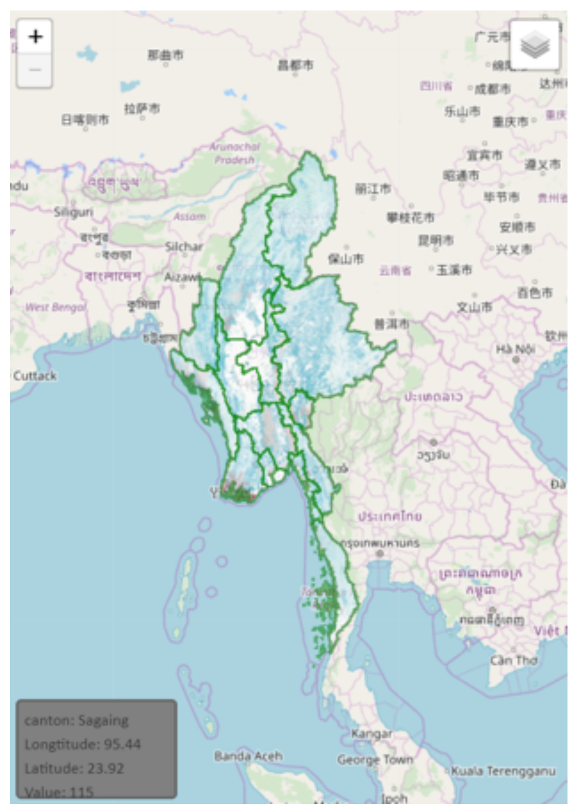
2.2 State-wise quickflow boxplot is quite similar to the national one.

2.3 Similarly, the red line showing as the mean make the state-wise comparison of quickflow more clearly.



Map

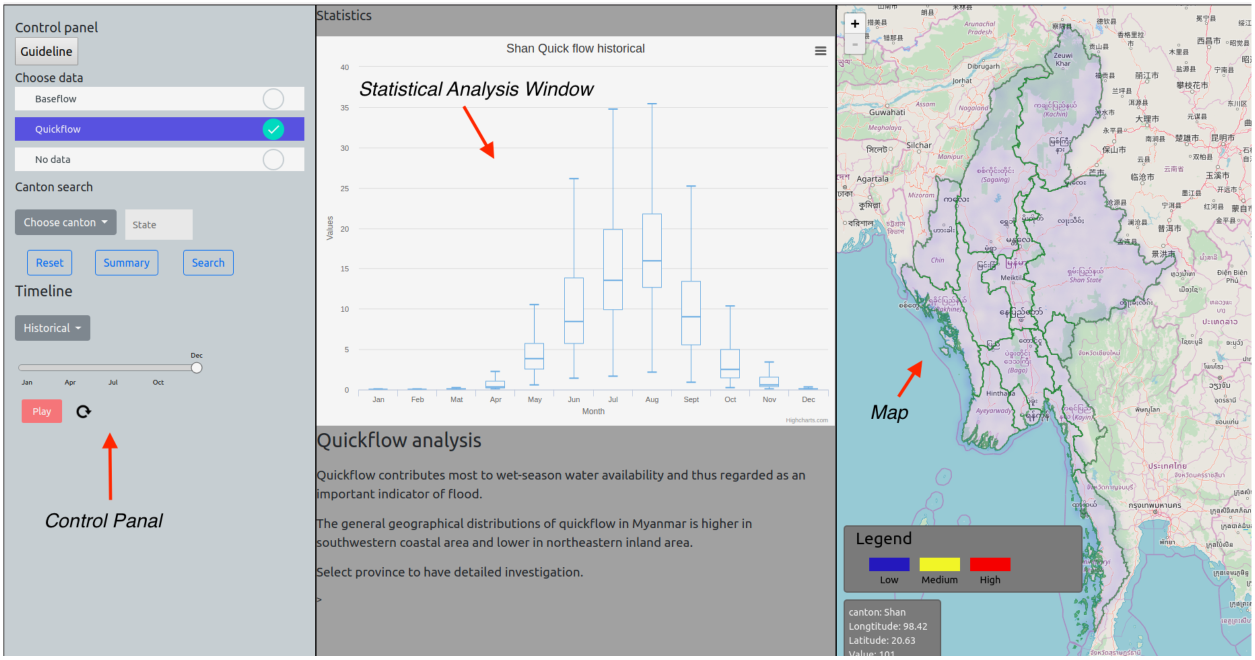
The map was generated using leaflet library. The layer loaded into the map is the one chosen from the control panel. The canton borders are drawn with D3.



1. The map could be zoomed in or zoomed out by the default setting of leaflet.js
2. The highlighted effect on the canton of the map is completed by the combination of two function (mouseover and mouseout)
3. Under the mode of Baseflow, the left-bottom controller will show the state name, position, and Baseflow value. It is implemented to clickOnevent to update the html inside a div element. Its

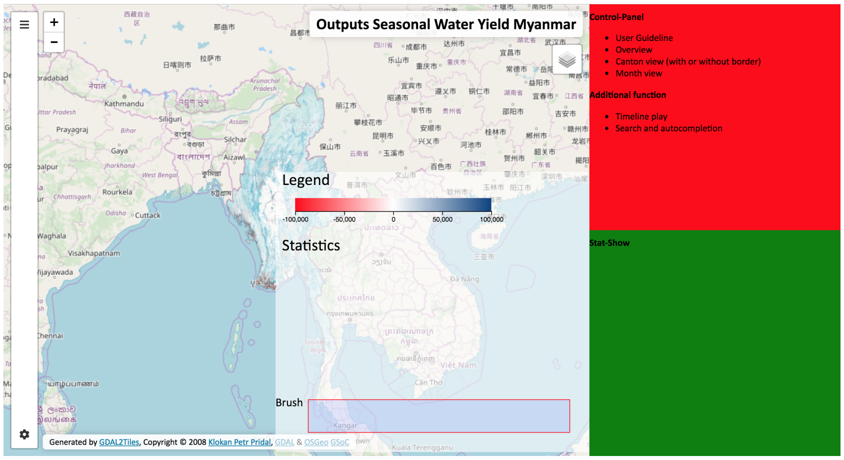
appearance and disappearance are controlled by the mode switch button.

Final Version Screenshot:



Detour and Difficulties

1. At first, as the dataset is really huge, we can not handle it locally. Therefore, we opened a remote server to do the data manipulation. The reacting time is quite long, and github.io is not compatible with http server. We did struggled with that in the beginning of the project. Therefore, we later chose to downsample the data, and found there is really similar to the previous version.
2. The previous stretch design looks like the following. As mentioned, we think this may cause the waste of the space, because the shape of map and the shape of country does not match with each other.



Evaluation

In this part, we would like to share our thoughts about the data insights, limitations and future works. Basically, we have met up with the goal set in the proposal. The main function s work stably.

Data Insights

Data generation (Recall):

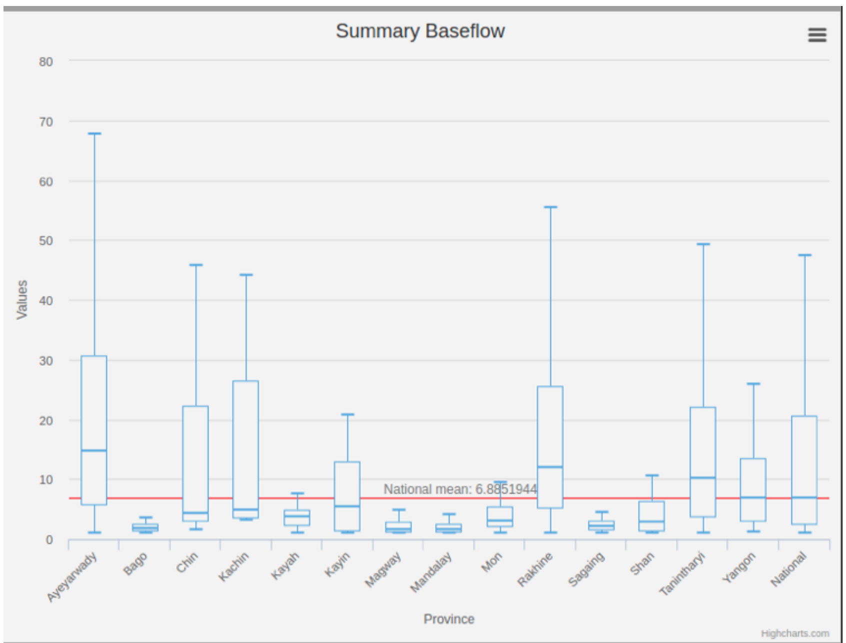
The baseflow data are downsampled and discretized into bins to enable real time interaction. The response speed was more than several seconds before preprocessing and became almost real time after that.

To extract the general pattern of data and enhance the responding time, the quickflow data as well as baseflow data are mapped into five representative numbers (low value, first quarter, median, second quarter, high value).

1. Baseflow_chang:

Baseflow, which represents the amount of precipitation that enters streams through subsurface flow and sustaining streamflow between events, contributing to dry-season water availability. Thus baseflow change value is regarded as an important indicator of sufficiency of water supply.

The geographical distribution of baseflow change in Myanmar is 'High among coastal area and low among inland area' in general. One interesting point is that the middle provinces such as Magway, Mandalay and Bago have even lower baseflow value than the ones that are more 'inland' such as Kachin. One possible explanation resides in the altitude and temperature difference. While provinces like Kachin have higher altitude and lower temperature, the above mentioned middle provinces are located in the center plains of Myanmar which have higher temperature and thus have more evaporation.



2. Quickflow analysis:

Quickflow is greatly related to precipitation and contributes most to wet-season water availability and thus regarded as an important indicator of flood.

The general geographical distributions of quickflow in Myanmar is higher in southwestern coastal area and lower in northeastern inland area. This is the natural results of monsoon climate in Myanmar. To be precise, the southwestern area are strongly influenced by the summer monsoon from Indian ocean which bring abundant precipitation. As the wet air move into the inland area, the humidity decreases and thus reduce the amount of rainfall.

Moreover, the wet air become cold when climbing the Arakan mountain in the south western part of Myanmar and thus transformed into liquid water and rainfall eventually. This causes the provinces around that mountain like Chin to have similar quick flow value to some of coastal provinces like Mon.

Deeper investigation into the distribution of quick flow in each province is enable by taking the boxplot of quick flow for the specific province into account.

Limitations and Future Task

First of all, as we do the downsampling on the data, the information loss is unavoidable, and therefore, in the future we hope to still host the data on the server, and figure out the solution of web communication block.

Secondly, our ignorance about the seasonal water yield and the climate model limited our further study in this model. In the future, we would like to collect more information about the background knowledge in order to have a deeper understanding about the data and model.

Last but not least, from our perspective, the UI design is very poor, as none of us have the related experience. We hope in the future, we could have a better and neater UI design.

Special Thanks: Instructor: Dr. Kirell Benzi, and supervisor from NatCap: Ms. Stacie Wolny

Peer assessment

1. Preparation: before each meeting, each of us prepared quite well
2. Contribution: Equally contributed to the team work
3. Respect for others' ideas: Yes, we did encourage to exchange different ideas for the project
4. Flexibility: Yes