

**INTERACTIVE VISUALIZATION SYSTEM OF OIL & GAS WELLS SOURCE-ROCK  
GEOCHEMICAL DATA  
TO AID HYDROCARBON EXPLORATION**

# **PROCESS BOOK**

By Skylar Shyu & Pablo Napan

## Overview and Motivation

There has been a recent convergence of conventional and unconventional hydrocarbon resources at the source-rock level and the question of why rich source rocks are located, and where they are in time and space, has become crucial for the global budget of petroleum resources. By having a global/detailed overview of source rocks geochemical data, explorationists can reduce the risk of charge factor and hydrocarbon generation as part of the overall petroleum system analysis.

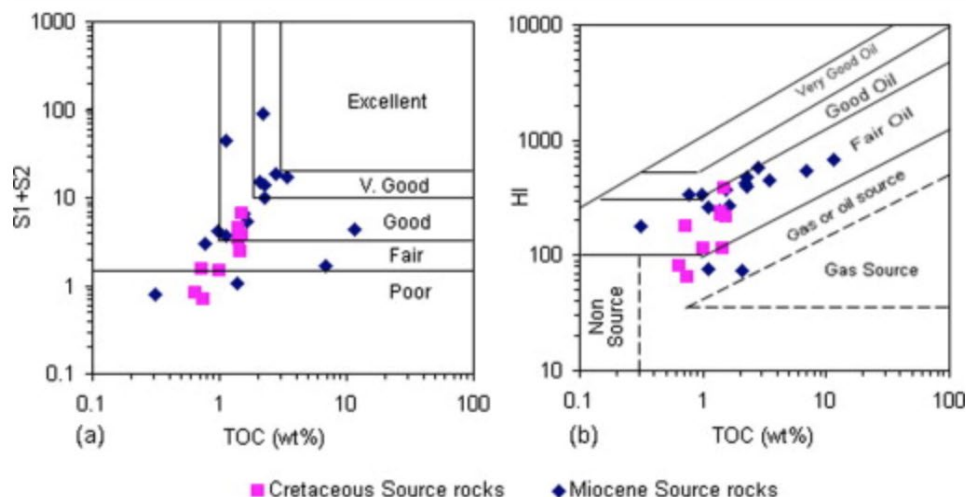
To aid this research frontier, the E&G institute has been compiling, standardizing and visualizing geospatial & geochemical data, however, the visualization tools that oil companies and research institutes have, are not appropriate for data visualization but mostly for the purpose of geographical data management e.g. ArcGIS, QGIS, WebGIS, etc. which are insufficient for visually interactive data exploration/visualization. More powerful data exploration tools like Spotfire not only have limitations, but also, they lack specialized diagrams/scatterplots rendering for geochemistry data analysis like Van-Krevelen, HI/OI plot, etc.

The main motivation of this project is the need of having an appropriate, powerful and interactive visualization system that allows visual data exploration to aid oil & gas prospection purposes. The geospatial focus of this project will be in the U.S. basins.

## Related Work

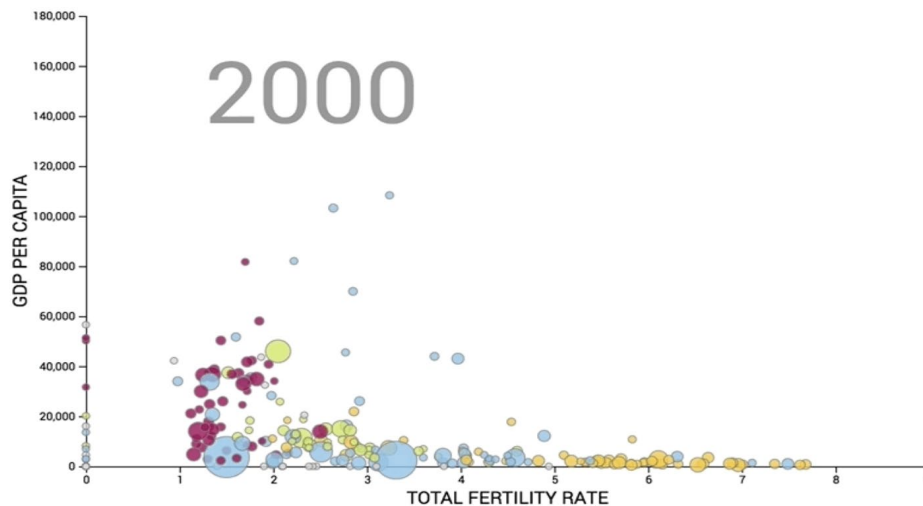
Most of the inspiration on the *technical design* (which plots or charts to display) came from the paper (1) **Evaluation of organic matters, hydrocarbon potential and thermal maturity of source rocks based on geochemical and statistical methods: Case study of source rocks in Ras Gharib oilfield, central Gulf of Suez, Egypt.**

This paper describes the flowchart in the analysis and evaluation of source rock organic matter based on geochemical and statistical methods. This paper was great giving the analysis flowchart but was not too great in visualizing the data.

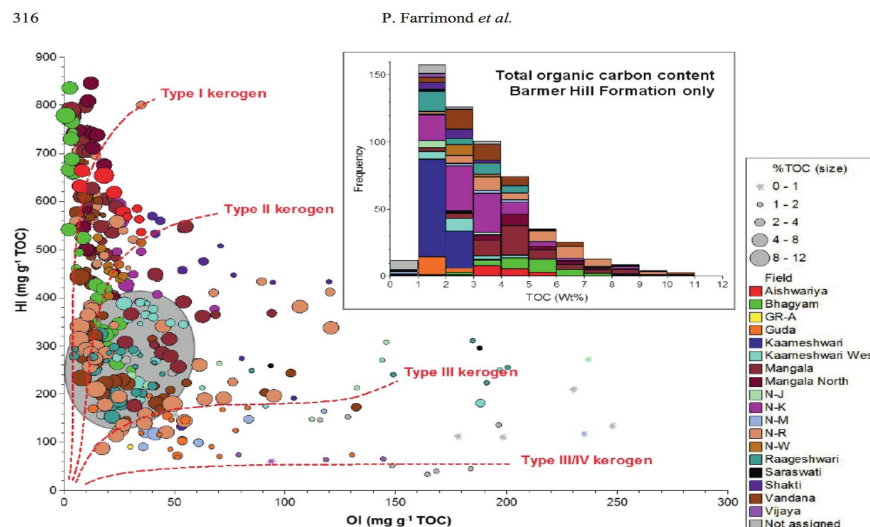


As seen in the image above, the data is poorly visualized despite the scientific correctness of the data display.

The inspiration for the *visual design* came from HW4. Visualizing and comparing world data in the screen led to the idea of applying a similar visualization flow to display U.S. source rock data and compare geochemical parameters across basins or formations (instead of countries). See image below for reference on what we are trying to build.



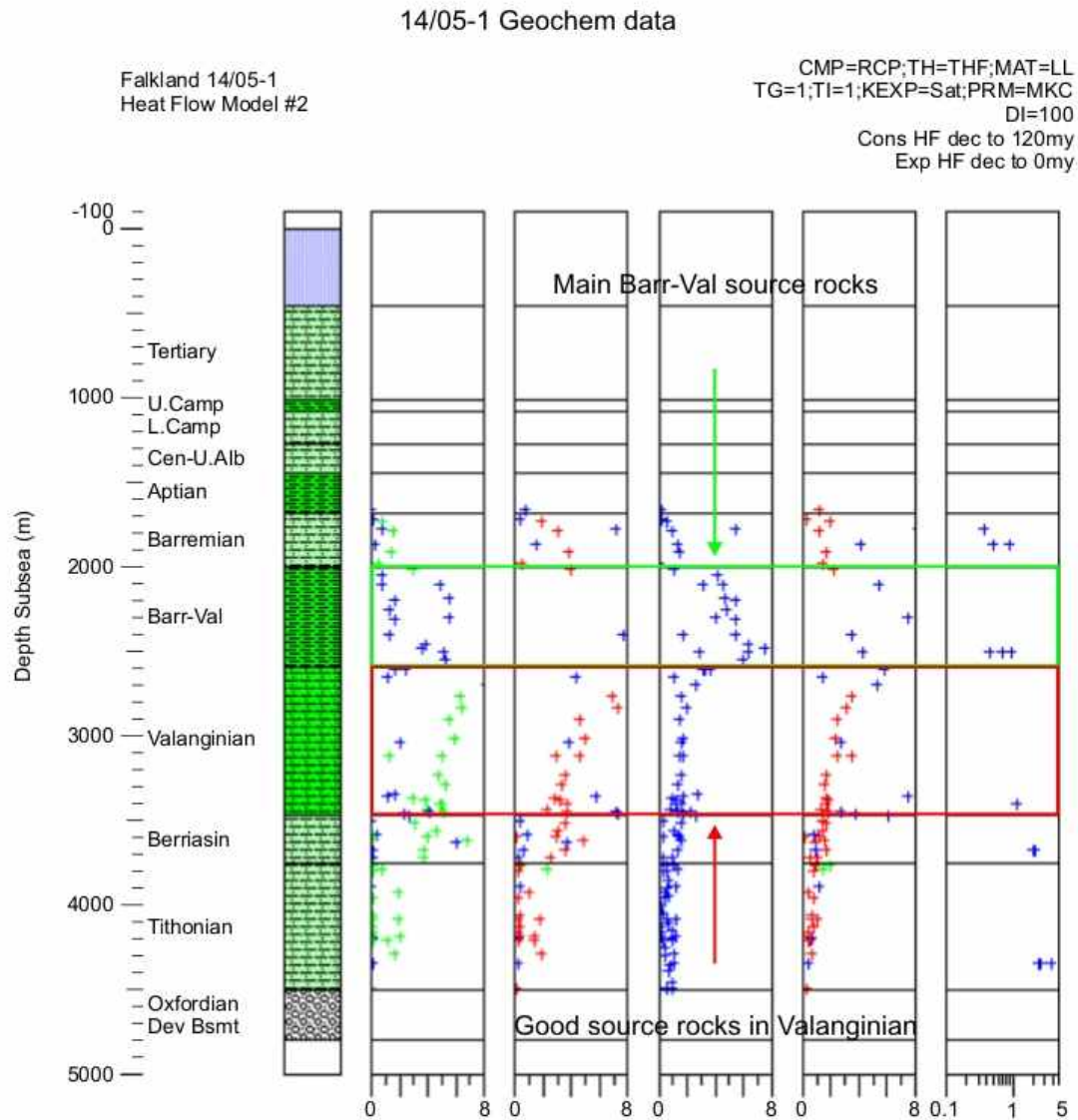
Also, more inspiration on the *visual design* of the bar chart and scatter plots came from the paper (2) **Geochemical characterization of oils and their source rocks in the Barmer Basin, Rajasthan, India** where the visualizations are much better than the previously mentioned one. See image below:

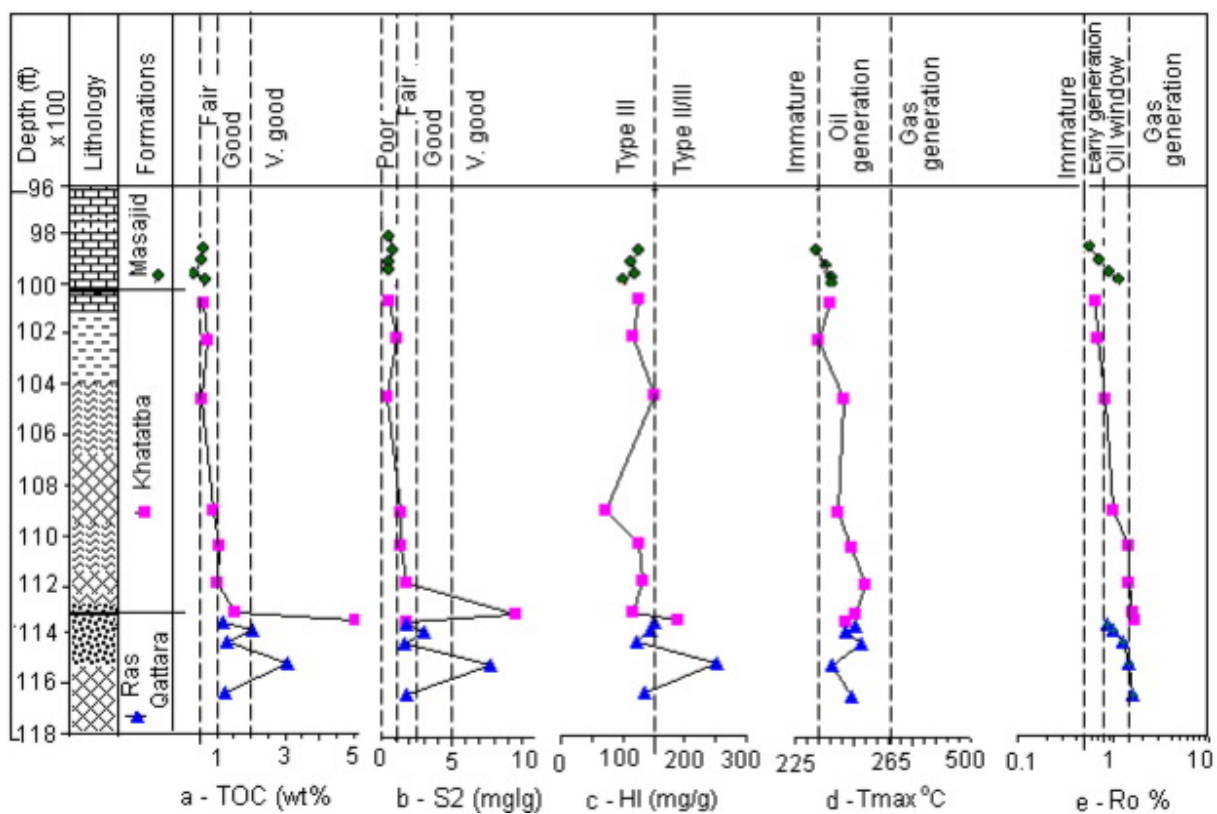


This image extracted from the mentioned paper (2). Notice the scatter plot data is nicely displayed. On the other hand, we did not take the bar chart as inspiration. In our opinion, this is poorly

visualized as there are too many colors to track. That is not our case since we will not go to the intra-formation level where we will need to differentiate formation members.

Lastly, the following two images extracted from scientific journals will also guide our nice-to-have depth chart where we will plot our geochemical parameters vs our Z dimension.





## Questions

We are building a visualization system that allows the user to explore objectively geochemical data and to take their own geoscientific conclusions. The objective of this visualization is to let the user discover his/her own data trends and not to impose any idea or bias.

At first we considered the question: "*What are the general source rock geochemical characteristics of X basin?*" But soon we realized that summarizing the data could be a complex task as the data has some real-world problems: 1) missing values; 2) there are many data values that a statistician would consider as 'outliers', however they are not outliers from a geological point of view, etc. So, if we showed only summarized data per basin, we would be not only dealing with a difficult task but also not fully seizing the richness of the data.

Then, we decided to go one more layer into the data and show data per source rock formation and not per basin, as before. Consequently, the question we are trying to answer changed to "*What are the geochemical characteristics of X source rock formation in X basin?*"

Other questions considered and that may be still floating around are:

- "*What are the spatial distribution of the geochemical samples in the Z dimension (depth)*"
- "*What are good/the best source rocks for recommending exploration*"

About the last question, although answering this question would make of this visualization the ultimate tool, we cannot take these conclusions by ourselves. Despite the fact that the analysis of the source rock geochemical data is a standard task and that there is a consensus in terms of the data parameters cut-offs, lower bounds, etc., the interpretations and conclusions about the prospectivity of source rocks may vary across geoscientists since there are so many other geological factors that can influence the production of hydrocarbons from a source rock so that it would be hard to tell with certainty which source rocks are the best for the users.

## Data

The data has been being collected and compiled over the last 5 years by the Energy & Geoscience Institute at the University of Utah which data values have been modified for confidentiality purposes.

The basinal data has also been obtained from the EGI project geodatabase, however, it is also possible to download it from this public source:

- <https://www.arcgis.com/home/item.html?id=4769216bf0234324881a6764f2979bd5>

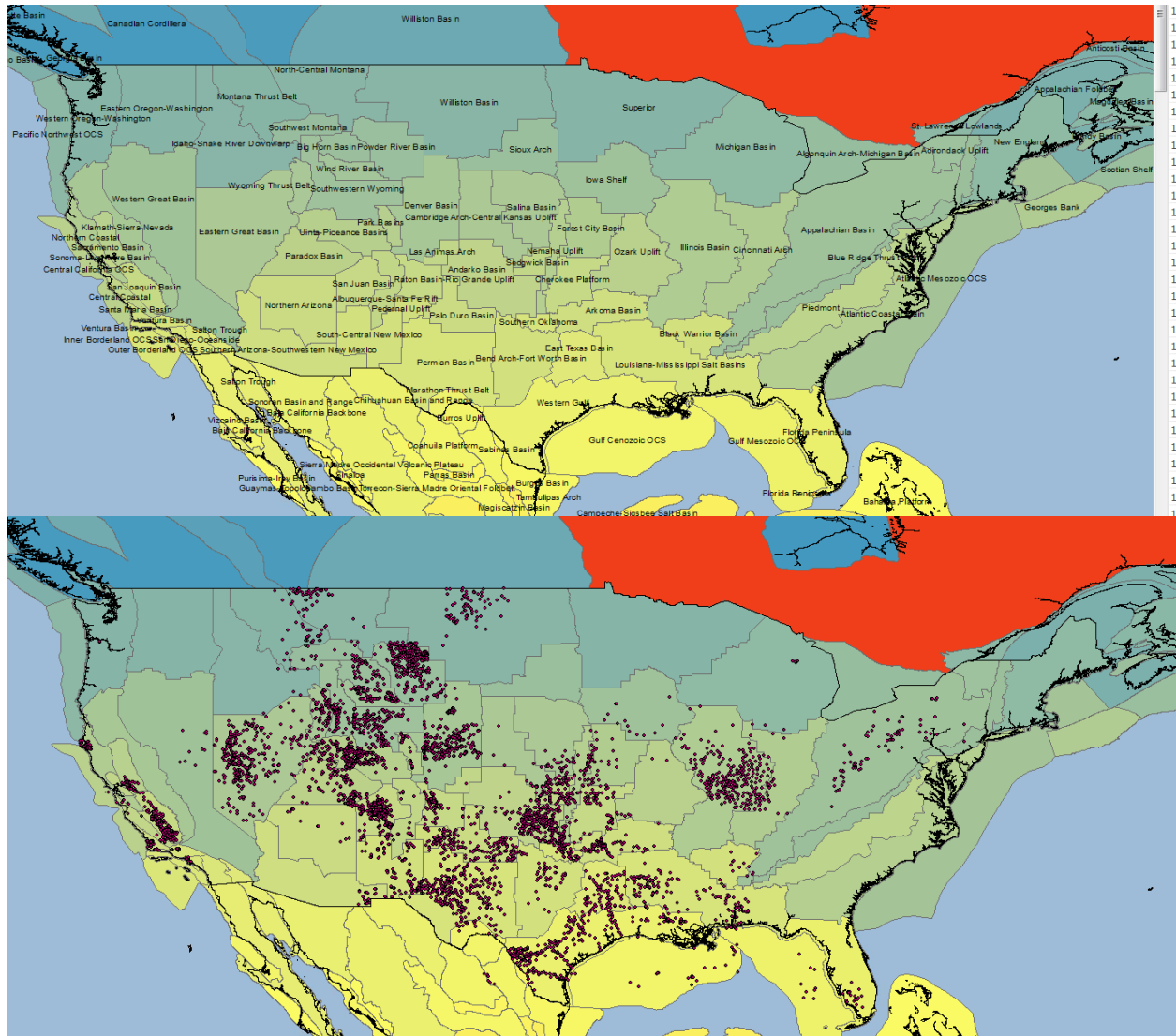
We expect to do a fair amount of data cleanup, as there's some inconsistencies regarding the availability/accuracy of the data for certain basins, groups, and wells and the samples that are extracted from them. This process requires some domain knowledge and discussion to ensure data and results significance.

From the tables, we are planning to derive:









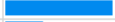












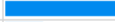























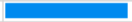




- Average age of the samples.
- Coordinate transformations.
- Production index.
- Averages (TOC, Ro, HI, OI, etc.).

## Exploratory Data Analysis

To initially look at our data we displayed our basins and well locations in ArcGIS to gain insights of their spatial distribution to decide the spatial projection which was set to d3.Albers. See two images below



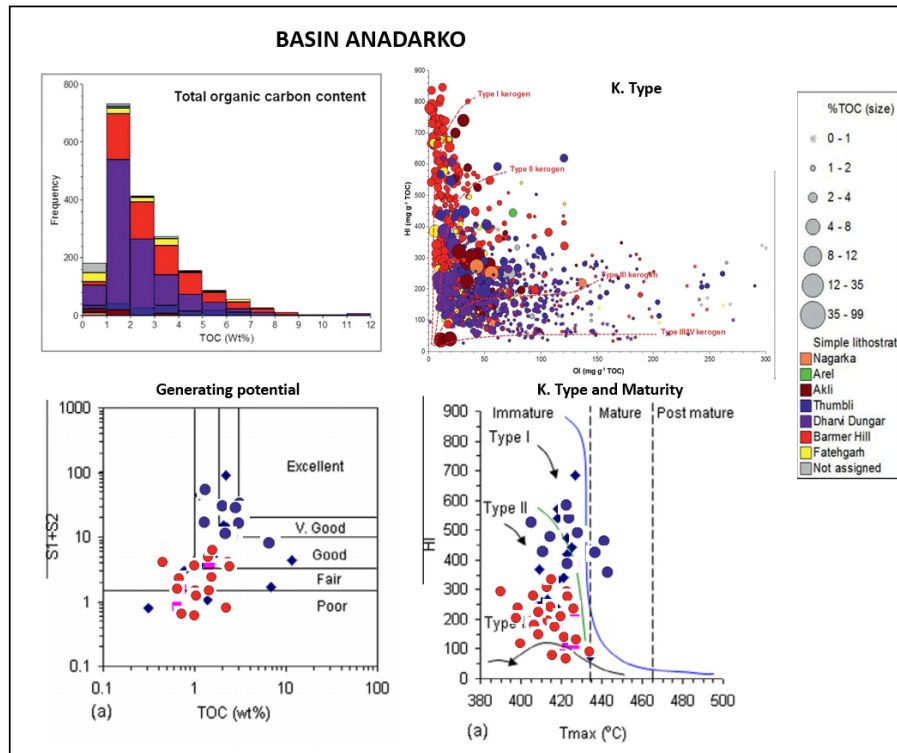
Also, the data, due to its tabular nature, was explored with Microsoft Excel, essentially by doing multiple and consecutive filters to gain insights on the maximum number of data points that would be displayed on screen at each time. We discovered that in some cases, there would be >90 formations displayed at a time in our Van-Krevelen scatterplot which would require 90 different colors to categorize and differentiate them from each other. This insight informed the decision to set a cut-off in the number of formations that will be displayed to <20. On the other hand, 20 formations would be too much for other chart (the stacked bar chart) so we decided to do a bigger modification of the project setting and display (in the same charts) the data by formation and not by basin. This, as well, informed the change in the question-to-answer section explained in the previous section. See image below.

Mowry Shale	Shale	Shale	Deltaic to Shallow marine		1.7				
Mowry Shale	Shale	Shale	Deltaic to Shallow marine		2.8				
Minnelusa	Shale; Car	Black shak	Marine		0.07				
Minnelusa	Shale; Car	Black shak	Marine		1.16				
Mowry Shale	Shale	Shale	Deltaic to Shallow marine		2.2				
Mowry Shale	Shale	Shale	Deltaic to Shallow marine		1.6				
Mowry Shale	Shale	Shale	Deltaic to Shallow marine		1.6				
Minnelusa	Shale; Car	Black shak	Marine		10.44				
Minnelusa	Shale; Car	Black shak	Marine		2.54				
Minnelusa /Sd	Shale	Shale			25.66				
Minnelusa /Sd	Shale	Shale			8.86				
Minnelusa /Sd	Shale	Shale			16.95				
Minnelusa /Sd	Shale	Shale			20.87				
Minnelusa /Sd	Shale	Shale			11.11				
Minnelusa /Sd	Shale	Shale			2.84				
Minnelusa /Sd	Shale	Shale			15.46				
Minnelusa /Sd	Shale	Shale			8.69				
Minnelusa /Sd	Shale	Shale			2.6				
Minnelusa /Sd	Shale	Shale			16.28				
Minnelusa /Sd	Shale	Shale			23.52				
Minnelusa /Sd	Shale	Shale			0.15				
Minnelusa	Shale; Car	Black shak	Marine		10.54				
Minnelusa	Shale; Car	Black shak	Marine		10.4				
Minnelusa	Shale; Car	Black shak	Marine		27.04				
Minnelusa	Shale; Car	Black shak	Marine		9				
Minnelusa	Shale; Car	Black shak	Marine		3.8				
Minnelusa	Shale; Car	Black shak	Marine		3.14				
Minnelusa	Shale; Car	Black shak	Marine		3.17				
Minnelusa	Shale; Car	Black shak	Marine		2.4				
Minnelusa	Shale; Car	Black shak	Marine		0.36				
Minnelusa	Shale; Car	Black shak	Marine		0.65				
Minnelusa	Shale; Car	Black shak	Marine		18.02				
Minnelusa	Shale; Car	Black shak	Marine		7.4				
Minnelusa	Shale; Car	Black shak	Marine		13.94				
Minnelusa	Shale; Car	Black shak	Marine		9.4				
Minnelusa	Shale; Car	Black shak	Marine		15.01				
Minnelusa	Shale; Car	Black shak	Marine		0.72				
Minnelusa	Shale; Car	Black shak	Marine		1.44				
Minnelusa	Shale; Car	Black shak	Marine		2.12				
Minnelusa	Shale; Car	Black shak	Marine		7.68				
Minnelusa	Shale; Car	Black shak	Marine		1.59				
Minnelusa	Shale; Car	Black shak	Marine		8.37				
Minnelusa	Shale; Car	Black shak	Marine		11.83				
Minnelusa	Shale; Car	Black shak	Marine		0.78				
Minnelusa	Shale; Car	Black shak	Marine		0.12				
Minnelusa	Shale; Car	Black shak	Marine		1.75				
Minnelusa	Shale; Car	Black shak	Marine		0.16				
Mowry Shale	Shale	Shale	Deltaic to Shallow marine		1.5				
Mowry Shale	Shale	Shale	Deltaic to Shallow marine		1.4				
Minnelusa /Sd	Marlstone	Marlstone			29.13				
Minnelusa /Sd	Marlstone	Marlstone			14.95				
Minnelusa /Sd	Marlstone	Marlstone			8.43				
Minnelusa	Shale; Car	Black shak	Marine		0.4				
Mowry Shale	Shale	Shale	Deltaic to Shallow marine		4.2				
Poison Canyon	Shale	Shale			0.06				
Poison Canyon		Sandstone			0.01				
Poison Canyon					0.00				

Also, barcharts and scatterplots were displayed (image below) to explore for outliers. With this analysis. This helped us realize that we should go for fixed rather than dynamic axis in all our plots.

# Design Evolution

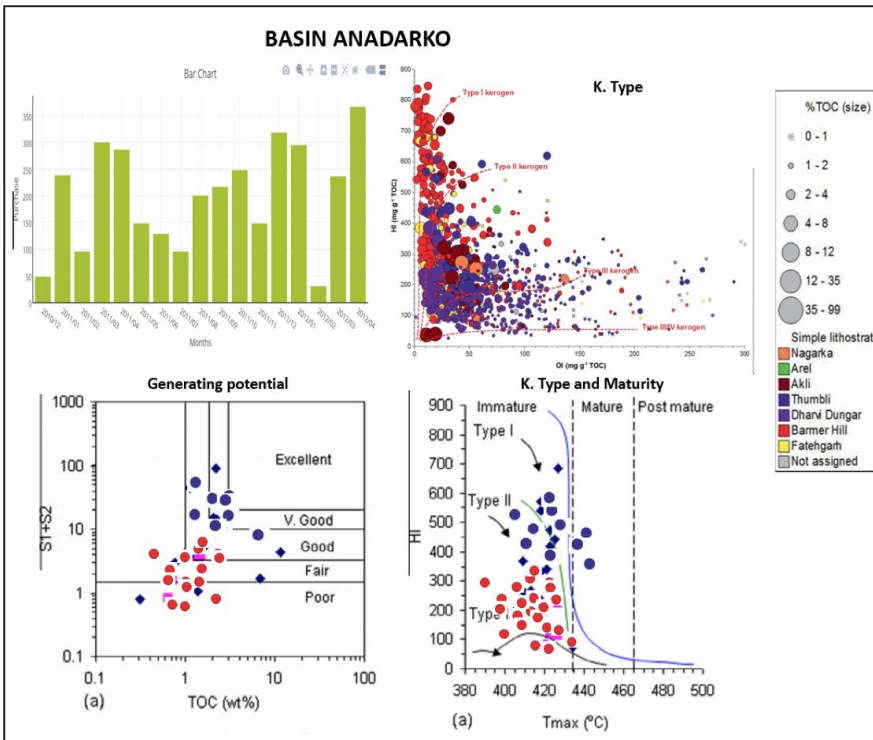
The group of visualizations proposed at first are shown below.



## Some thoughts

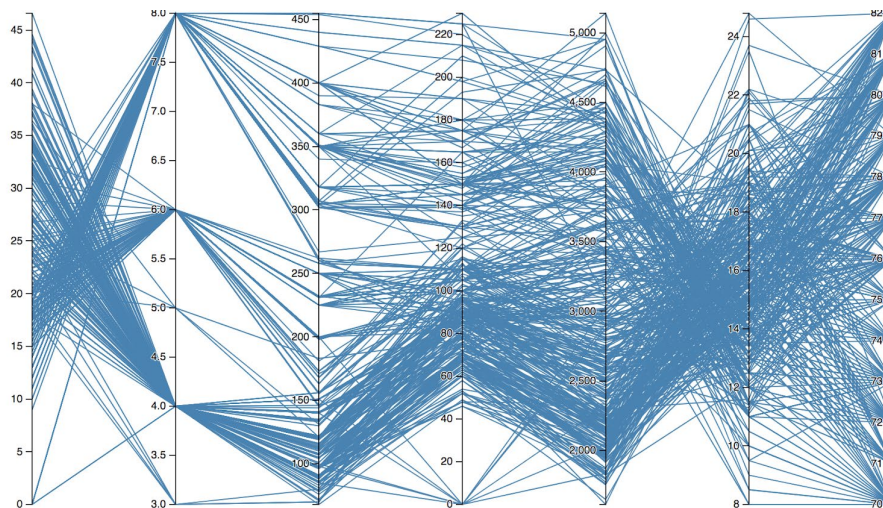
- All axis fixed. Values falling out of the graph should be greyed out/deleted.
- Colors will encode the column 'formation\_name'. In each basin there may be 10 formations or maybe more. The number of formations change from basin to basin. So, the scale and color palette should be dynamic (if this too hard, we can think something else).
- Color palette doesn't matter as long as they can be differentiated from each other.

Then, after exploring the data and realizing the possibility of >90 colors in our charts (explained in the previous section) we decided to go for a simple bar chart that would update to its respective formation whenever a sample is clicked in any of the other 3 plots.

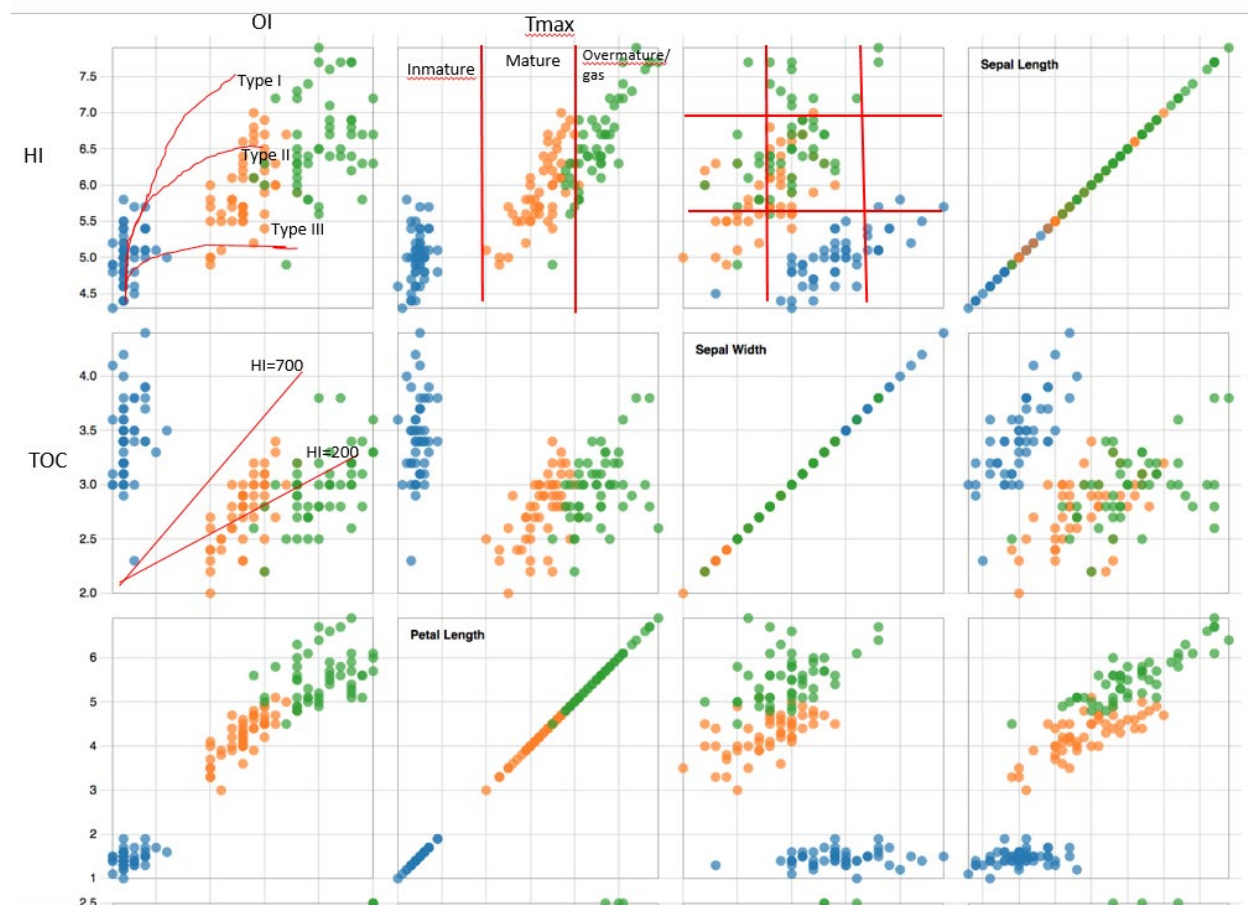


The second modification came after the peer-review session where our visualization was suggested to have less simplistic data visualizations, more interactive content, or a combination of both.

We tried redesigning our plots with a somewhat more sophisticated plot (parallel coordinates), however, despite its beauty and how eye-catching this chart is, it failed to encode the data as properly as a scatterplot where positions (X,Y) are better to analyze visually rather than slopes. Also, it's limited. One attribute can only be related with two other attributes (left and right) which are not applicable to our visualization needs. This visualization is great when we want to explore our data and look for patterns. In our case, we are certain of which relationships we need to see in the visualization for the geochemical analysis.



So, after deciding that we had to go with scatter plots, we analyzed the possibility of implementing a scatterplot matrix with synchronized brushing. This idea was dropped quickly as, since we need to plot all of our parameters with themselves, there will be some scatter plots in the matrix showing relationships (e.g. Ro vs. S2) with none or little geochemical significance.

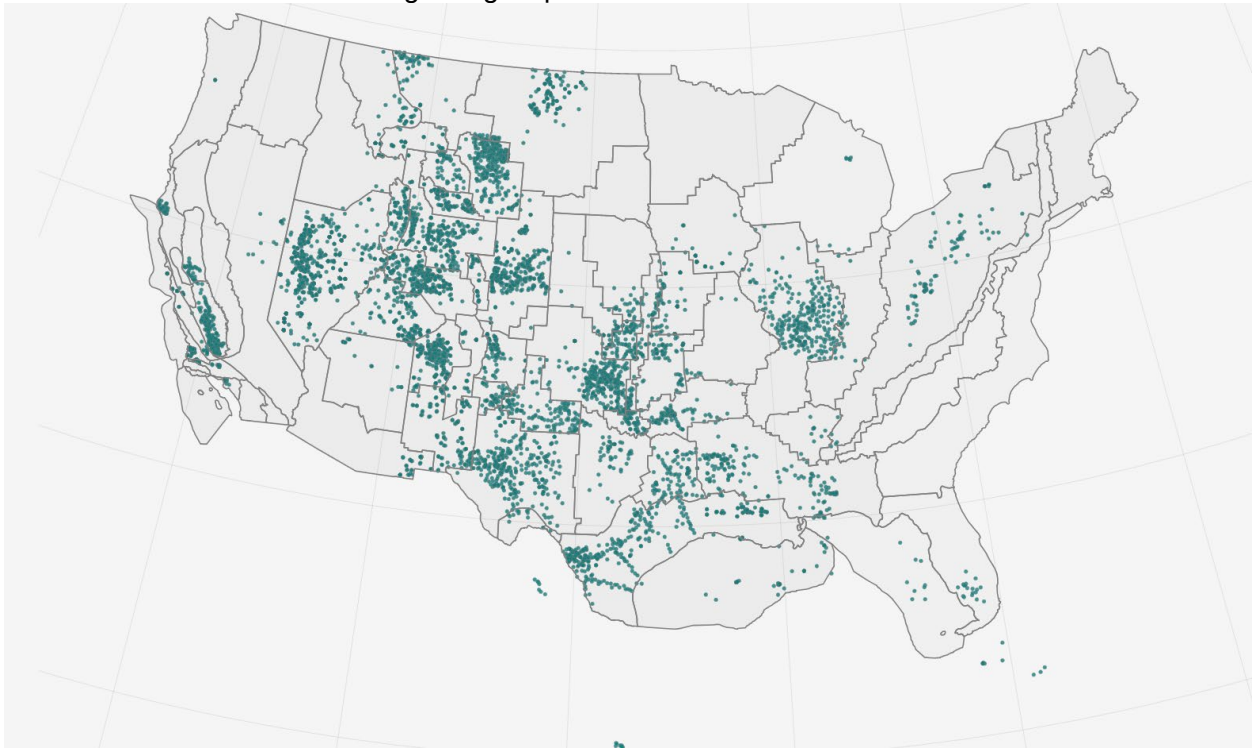


After these considerations of visualization redesigns, we opted to continue with our proposed 4 charts (TOC bar chart, Van Krevelen plot, inverse Van Krevelen plot and potential plot) which are justified. Additionally, it's important to mention that these plots/charts are industry-standard and proposing a new way of visualizing these data may disturb the geological interpretation of it. So, unless newer visualization plots get available and they prove to be helpful and suitable for the visualization/interpretation/analysis of geochemical data, visualizing the data with standard charts would be a sensible decision.

However, to respond to the peer-feedback, it was decided to either add more interactivity/functionality (search bar, dropdown, etc.) or draw an extra plot (depth plot) which would add one extra layer of data exploration. At this point, we have not decided which of these two options to take.

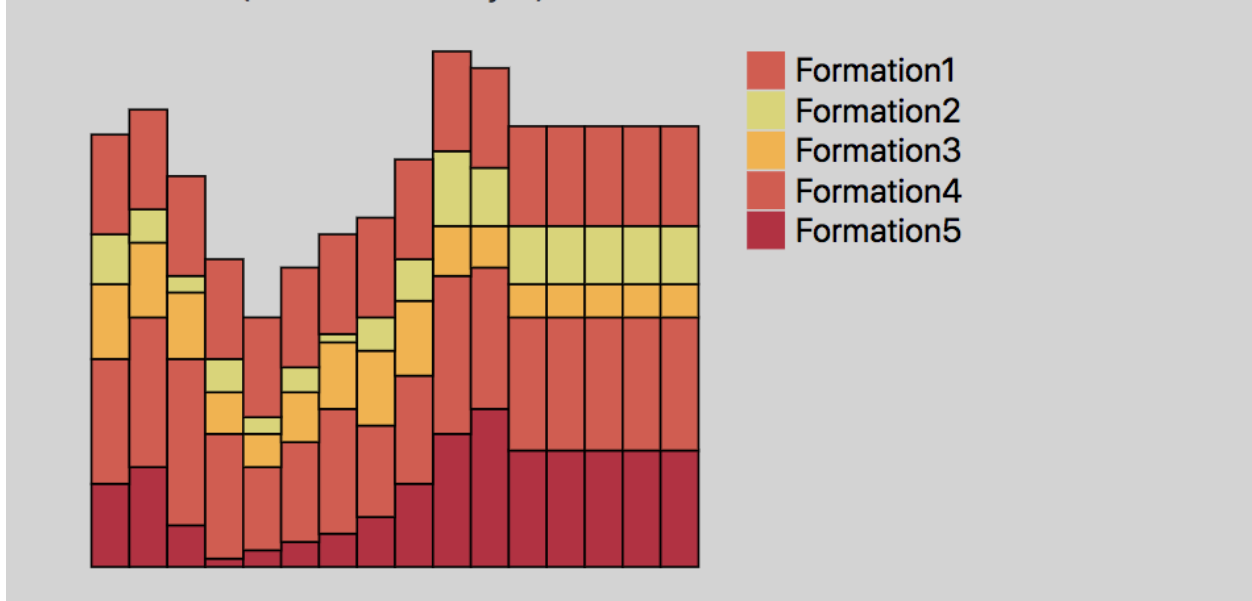
## Implementation:

First screen: U.S. basins showing well geospatial data.



So far, we have our spatial well data and basins (polygons) displayed on a map. The basins are clickable which update the bar chart below:

TOC bar chart (not interactive yet)



Next steps in the implementation of the project are: implement the other 3 charts and add interactivity and synchrony.

### Entry Nov. 1:

Implemented boilerplate code and basic bootstrap of the site's page

### Entry Nov. 3:

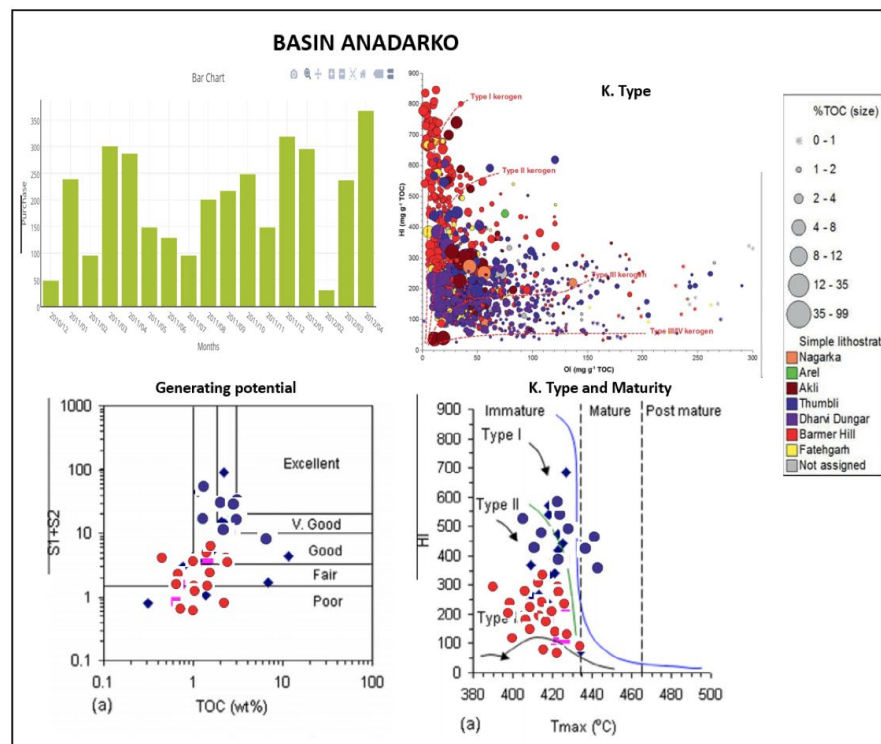
Set up initial div layers for all the components: plots, legends, map, etc. Added a dummy map.

### Entry Nov. 7:

Messed around with the basins. The zoom-in on basin selection (click) is proving to be a complete nightmare. There's too much proportion calculations, not to mention that's not including the transition that will need to be applied. Back to the drawing board.

### Entry Nov 9:

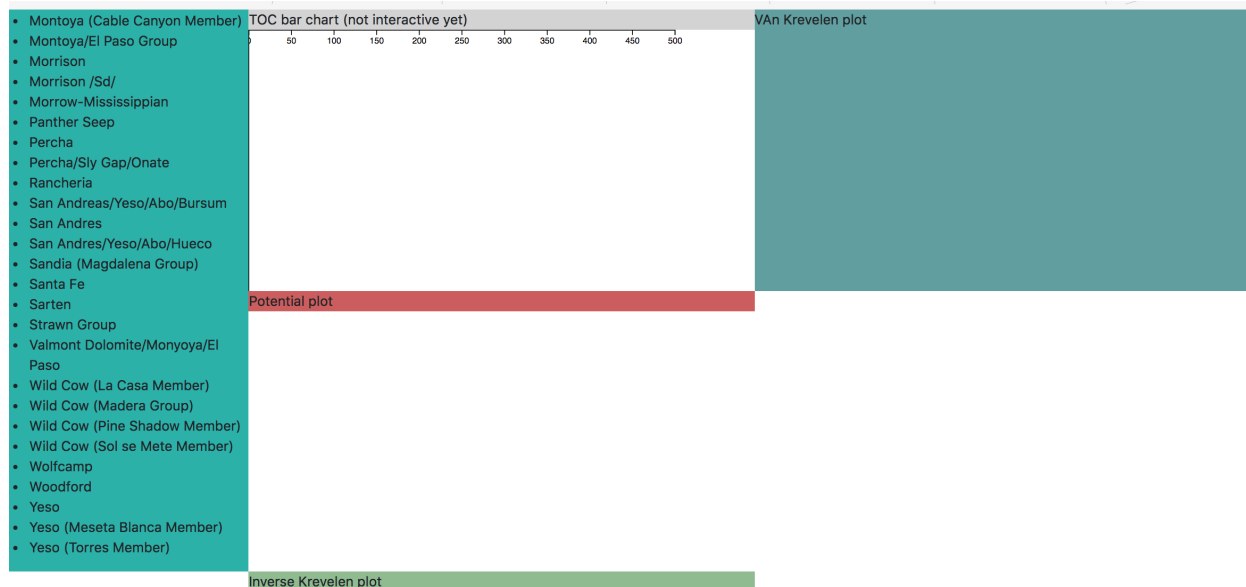
Redesigned the focus of the project from basin-wise to formation-wise



#### - Some thoughts

- All axis fixed. Values falling out of the graph should be greyed out/deleted.
- Colors will encode the column 'formation\_name'. In each basin there may be 10 formations or maybe more. The number of formations change from basin to basin. So, the scale and color palette should be dynamic (if this too hard, we can think something else).
- Color palette doesn't matter as long as they can be differentiated from each other.

The image above represents the new scope of the project. Note the chart in the top-left changed from a stacked bar-chart to a histogram.



## Entry Nov 17

### Meeting with TA

#### Feedback:

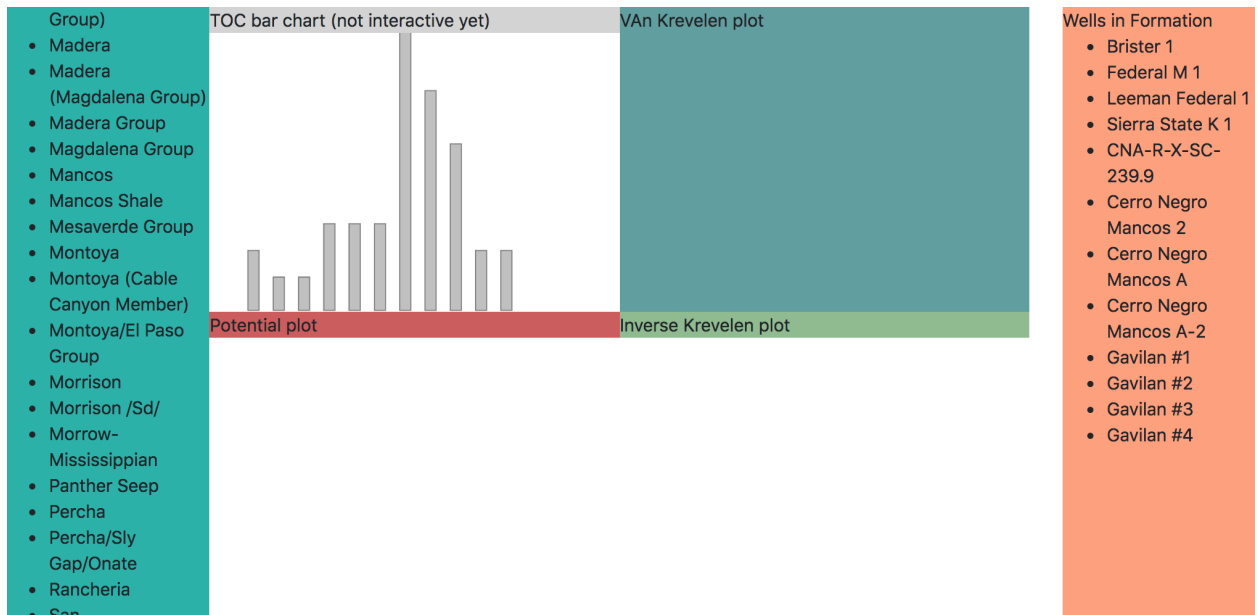
We were suggested to add an explanatory paragraph in the beginning of the visualization for the layman users.

#### Entry Nov 17 – pt2:

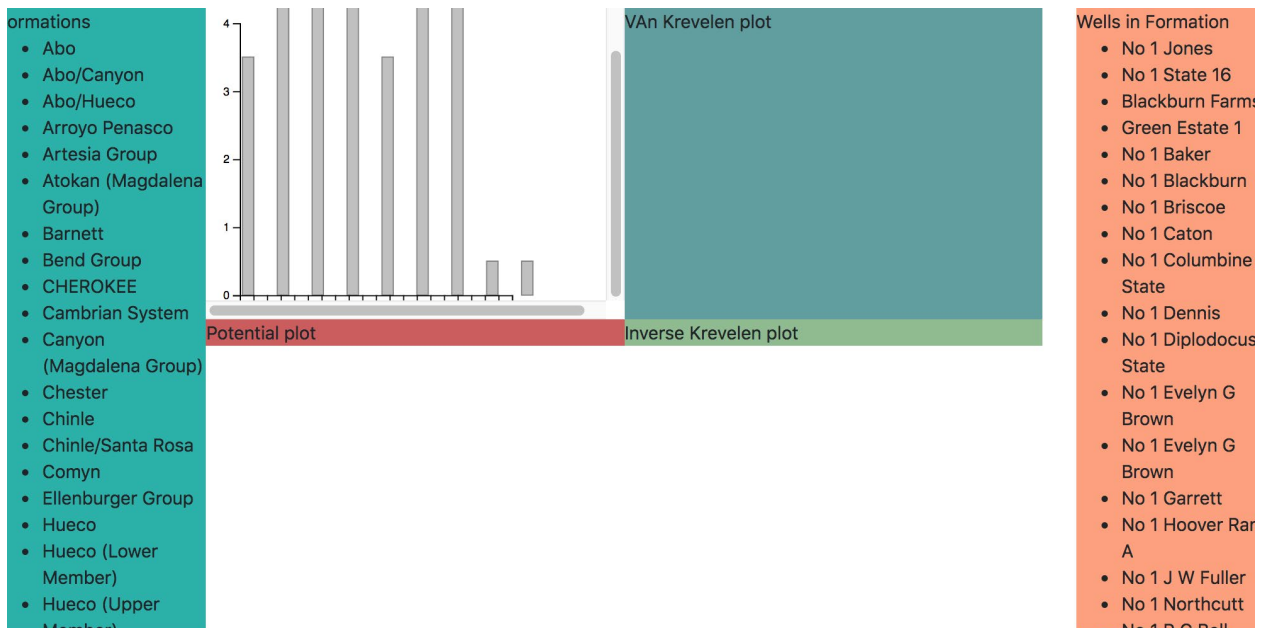
Populated the div layers with formation and well names. There was some messiness with async calls, but maintaining separation of concern seems to have fixed it.

#### Entry nov 18:

Working with histogram layout. Hard to configure. Seems like it does not give the option to hard configure the bin size, it recalculates it everytime. Most probably will implement this partially or from scratch if necessary.



Some decision needed to be made with respect to this chart because of its dynamic Y axis. Decided that Y axis will be adaptive for values <10 and fixed for values >10. This, due to the y axis can get too crowded with ticks. On the other hand the X axis will show the values from 0 to 10 and will not display values over this limit.



AutoSave OFF

SRCPPhase2GeochemUSA2

Search Sheet

Share

HomeInsertDrawPage LayoutFormulasDataReviewView

Still working with data:

There were unknown formations which were filter out and also it was notices that one space in the formation name gets recognized as a different names for javascript. E.g. ‘Sligo’ and ‘Sligo ’ are the same for excel but different for JS. Some data formatting was done here.

## Entry Nov. 18:

The axes seem a little botched for the TOC chart. Attempted to fix it, but it still seems wonky. Played with the overflow attribute to try and keep all the SVG elements visible in the div layers (for the plots). Expanded the map so that it’s bigger and takes up more of the screen.

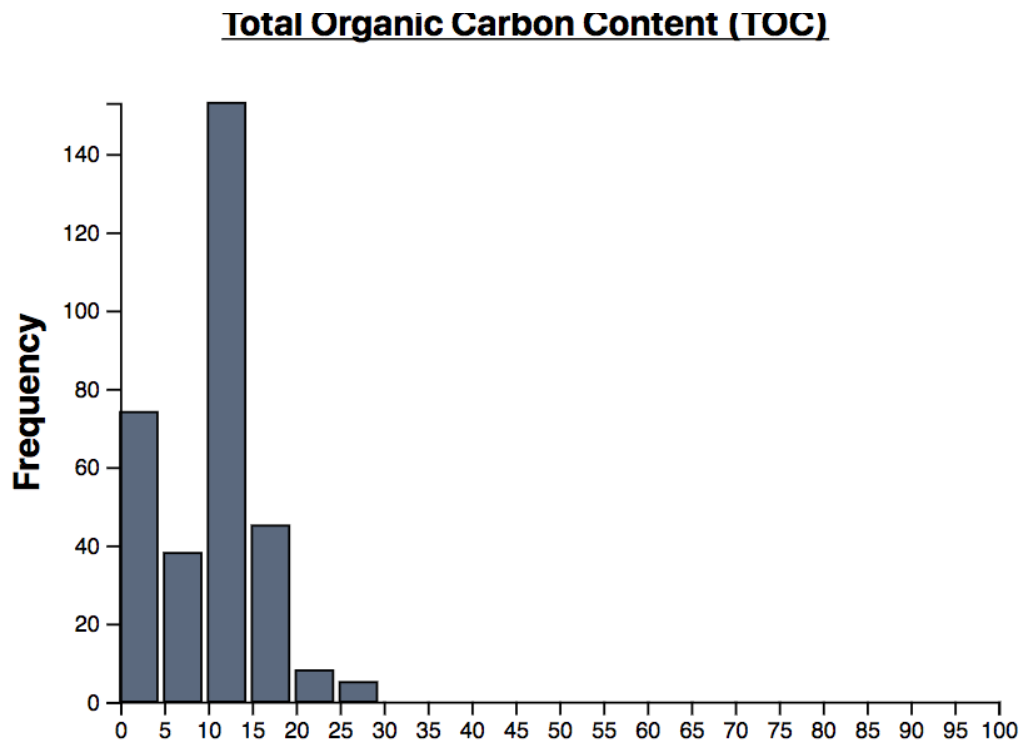
## Entry Nov. 21:

Implemented the Van Krevelen scatterplot constructor such that a static graph with all the wells of a selected formation will appear. Also implemented the updated function such that a new chart is rendered when a different formation from the same basin or a different basin is selected. All the points look the same, meaning that it’s impossible to distinguish with formation they’re from.

Also implemented the Inverse Krevelen plot such that it’s at the same stage as the Van Krevelen plot. Also figured out how to use the *transition()* method to *append()* and *remove()* datapoints when toggling between formations, as well as rescaling the axis such that the plot points are all displayed properly.

## Entry Nov. 24:

Implemented the Potential plot. There was a bit of data wrangling involved since the y-axis is the sum of two data elements (S1 and S2), but very doable and easy. I ended up appending the sum as another data element so we don’t have to keep recalculating it.

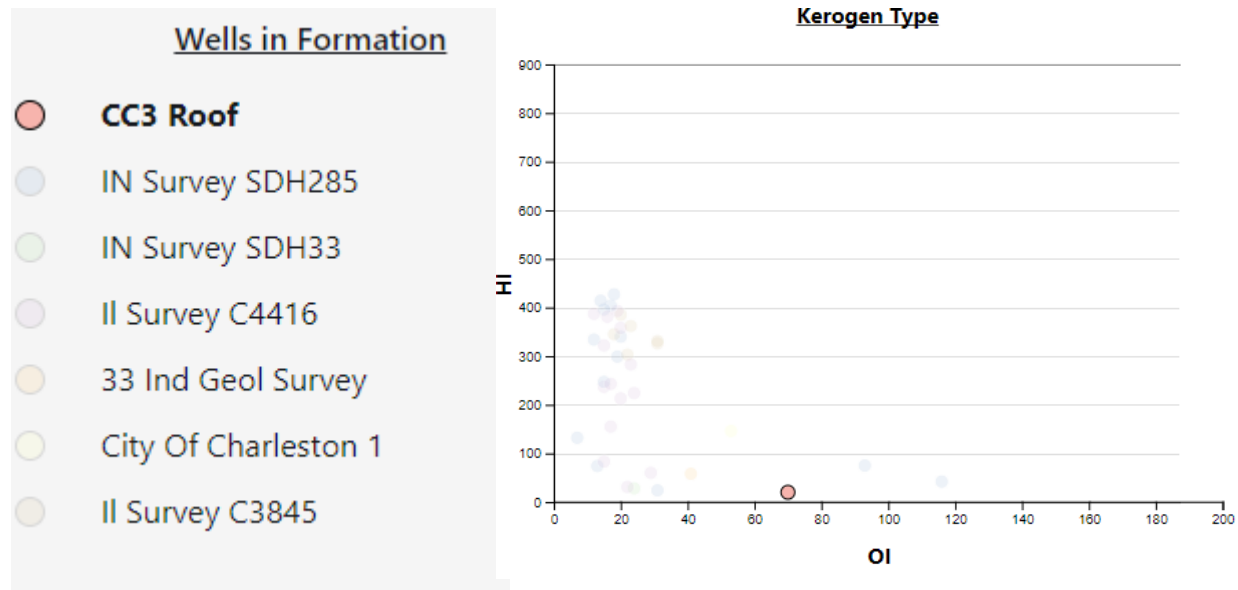


The TOC histogram was finally finished.

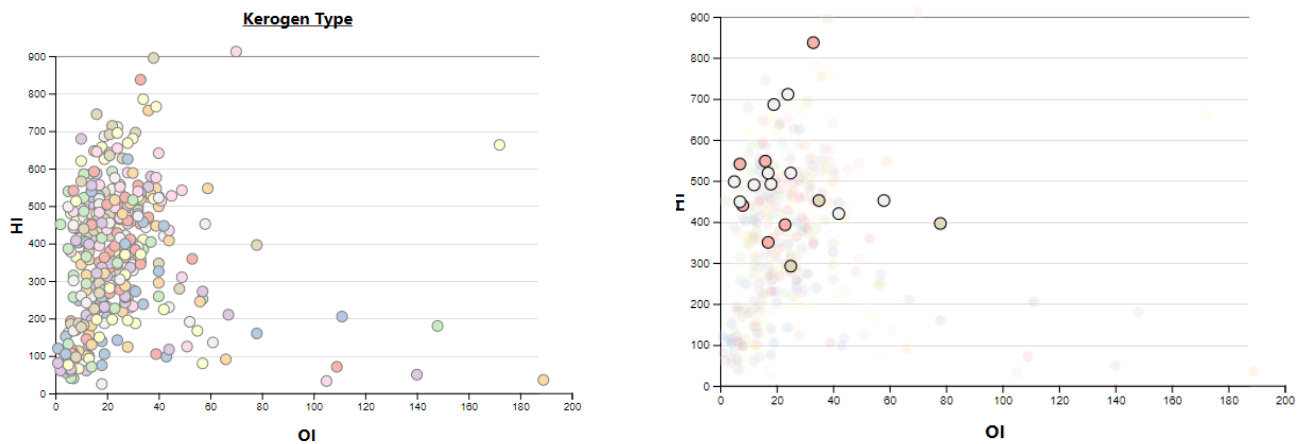
#### **Entry Nov.25:**

Skylar is pulling an all-nighter on this. Since the Van Krevelen, Inverse Van Krevelen, and Potential plots are all scatterplots and are at the same stage in terms of progress, everything described in his entry will apply to all three graphs:

- Figured out the color scale for each formation! *d3.schemePastell1* it is. Originally, I wanted *d3.schemePastell1* as the 'unselected' color and *d3.schemeSet1* as the 'selectedColor'. Ended up forgoing the latter and instead implemented opacity and stroke-color when it comes to accenting the points that are associated with the user's formation selection.
  - This also extends to the formations legend. Unselected formations will have the legend slightly faded, while selected formations will have the circle's stroke changed from gray to black for emphasis. This works well as it mirrors the chart selections.
  - Clicking once on a formation selects it. Clicking it again deselects it.



- There was a formation with 300 datasets. Today Skylar has learned that no matter how low the fill opacity is dialed down on a plot point, piling 100+ of them where they all somewhat overlap essentially nullifies the whole opacity effect. In the end, the selected elements were just brought to the front using d3's *selection.raise()* method.



- Implemented the formation selection! Both the lists for Formations and Wells were unordered lists. They've been both replaced with tables and rows. This allows easier manipulation of CSS (padding, margins, etc.), not to mention implementation of the formation legend is so much easier.
- Fixed the axes for the scatterplots. There used to be a weird little gap between the x-axis and y-axis— said gap has been eliminated and now it looks like a proper axis. This lead me to playing around with the parameters passed to the axes' *transform()* call. I had a difficult time abstaining from hardcoding dimensions such as height, width, and margins. In the end, I just set the height to be the height of the window and took a certain percentage of that; same goes for the width.

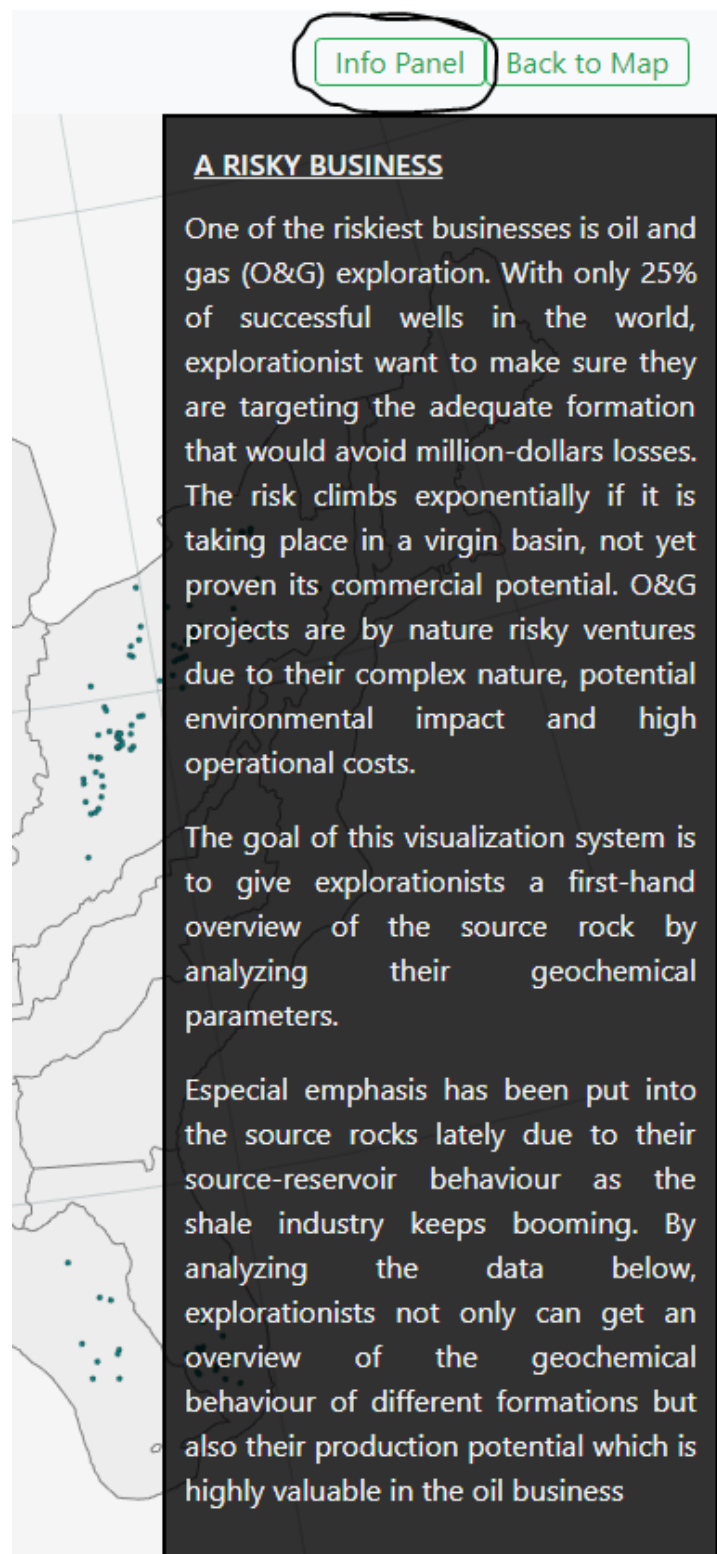
- Added gridlines and axis labels! Pretty much followed the same recipe for the axes. Didn't know that rotating `<text>` elements was a thing.

Outside of the three scatterplots, toggling between the two 'screens' has been implemented and is working. It was done changing the `style="display: value"` attribute for screen1 and screen2. When `value = none`, the div layer will be hidden; when it's `block`, the div layer will be displayed.

### Entry Nov. 27:

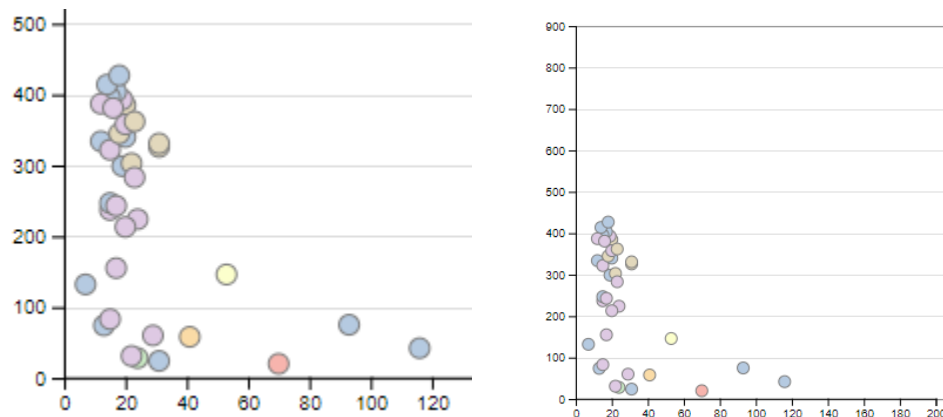
Stylized the tooltip for basin-hovering! D3's tooltip library was proving to be a bit challenging, so I decided to implement my own. Also tried to extend the use of tooltips to the van Krev plot. It's promising, but I need to figure out how to unbind events when certain formations are deselected.

An information panel has now been added to give users an intro as to what the smattering of circles are trying to describe upon first glance; it is toggleable:

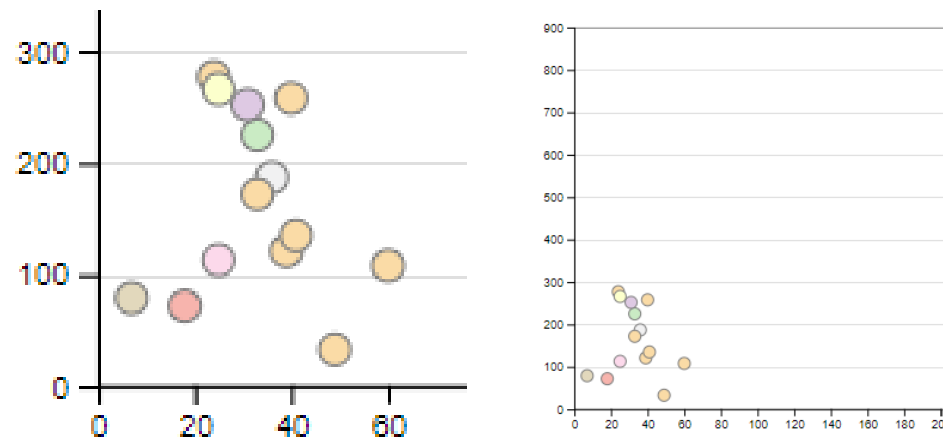


A decision was also made to change the axes for the scatter plot charts to be static, since those static domains sensibly responds to data domain-knowledge and geoscientist are accustomed to see the data points moving across a plot but not to rescaling on change.

For example, below are the Van Krevelen charts for formation A, with dynamic axes on the left and fixed axes on the right.

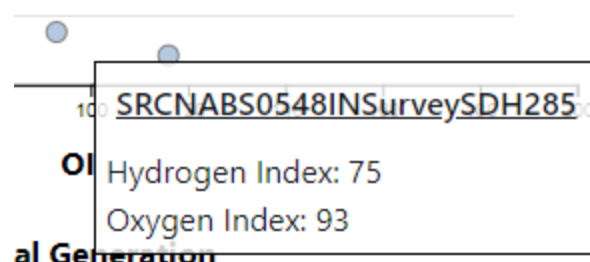


Below are the Van Krevelen charts for formation B, with the same axes as above.

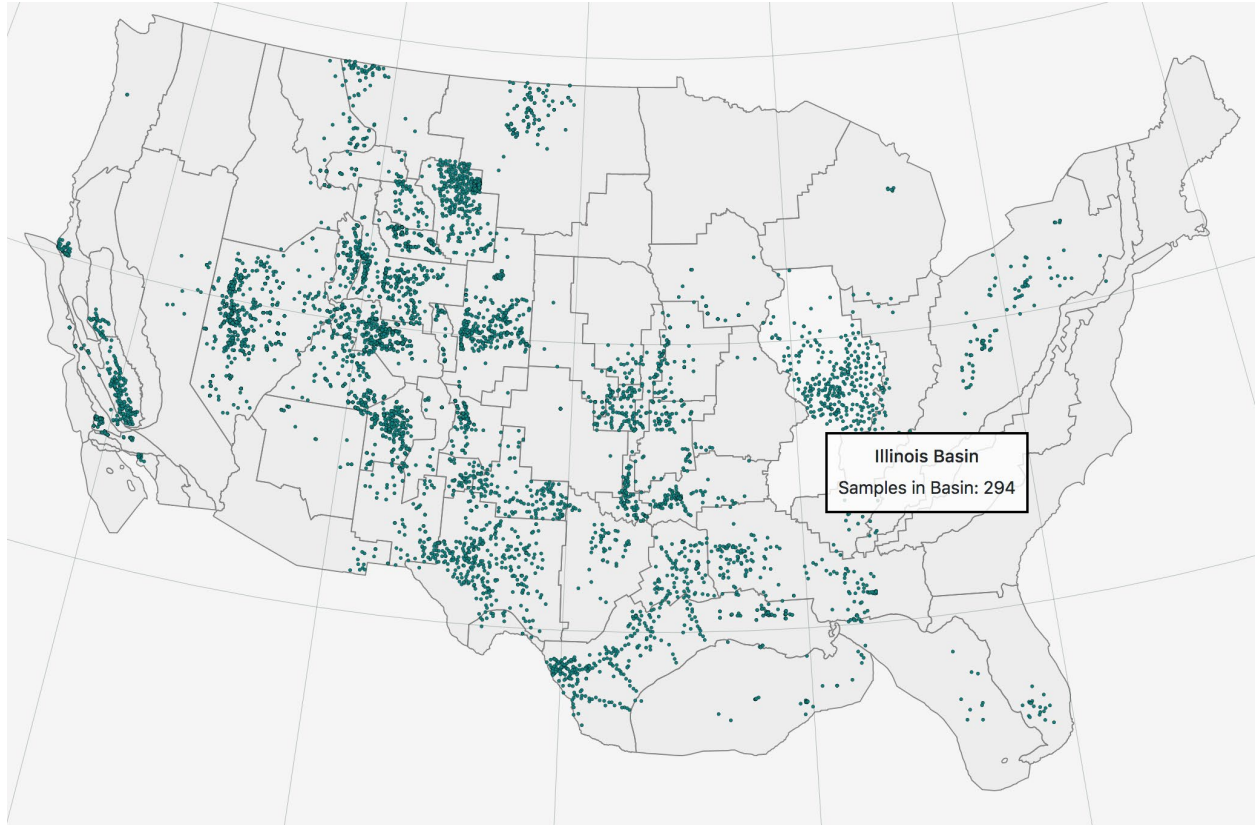


It is easier to compare the charts with the fixed axes and draw conclusions regarding the spread of samples.

Also managed to get the tooltip for each point to work. This way, users will be able to obtain accurate data regarding each point rather than eyeballing the approximation of the point on the plot if they wish.



The basin hover over now also works!

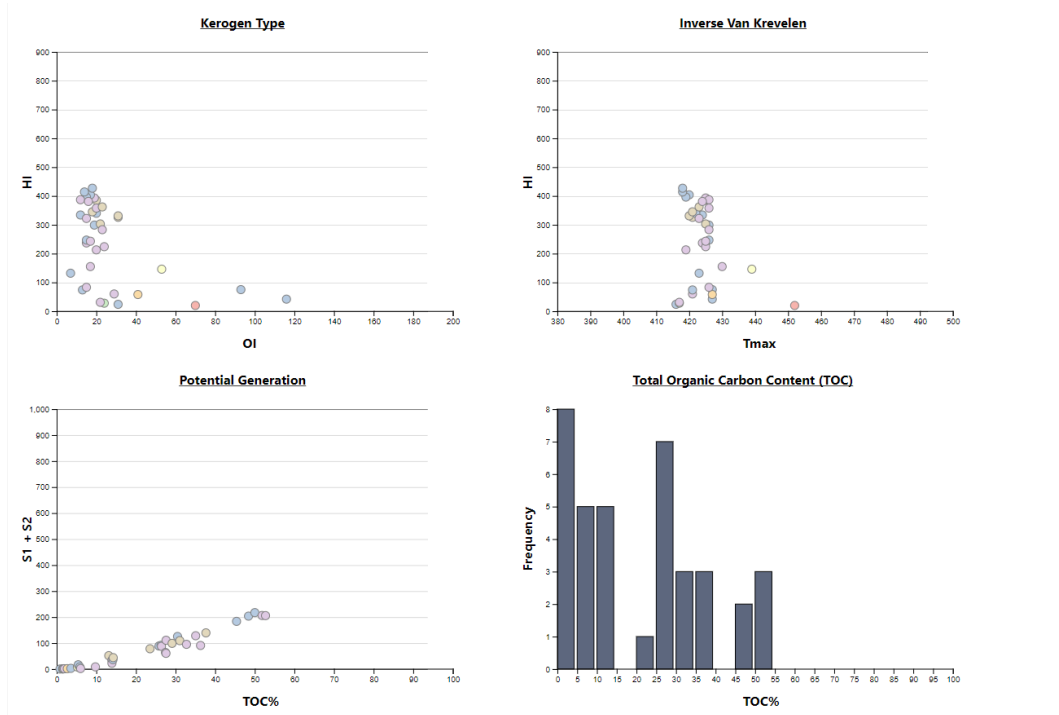


**Entry Nov.29:**

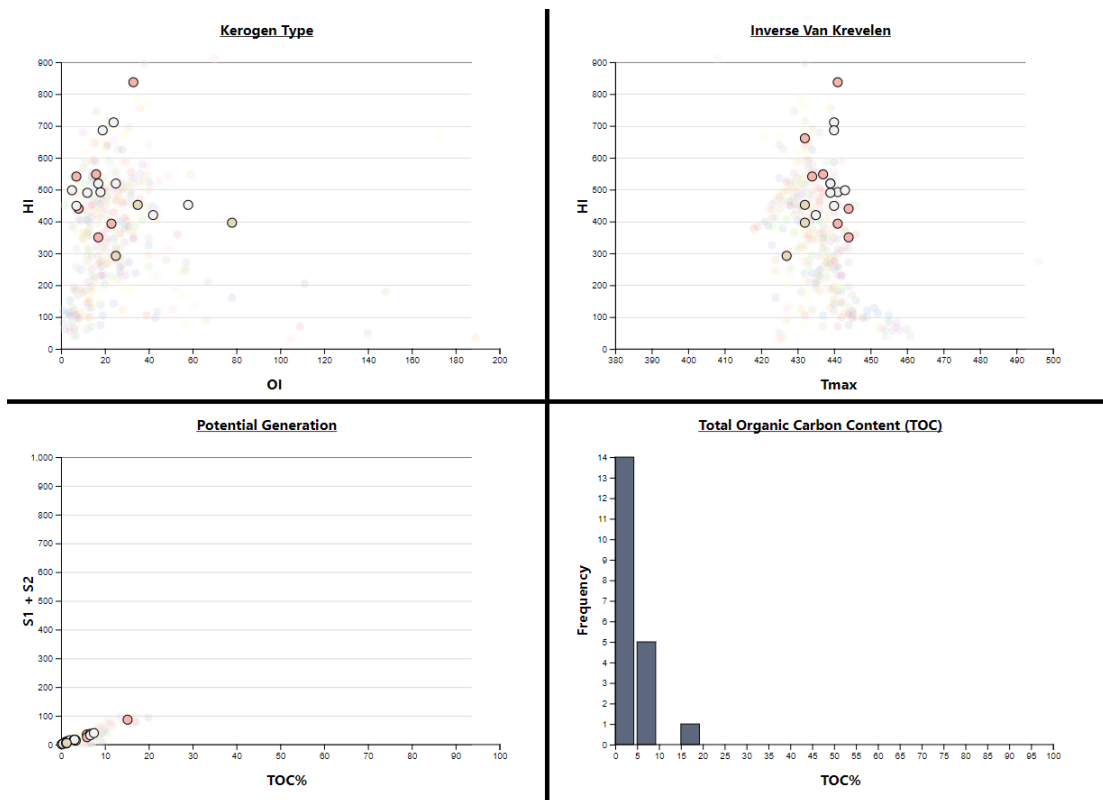
Last minute clean up! Fixed the living daylights out of the TOC Chart (it wasn't scaling properly, and there were issues with redundant rendering of axis, etc.).

Below are the final four charts. I think borders could be drawn around each of the plots' div layers, but it felt extremely busy upon a test implementation.

No borders:

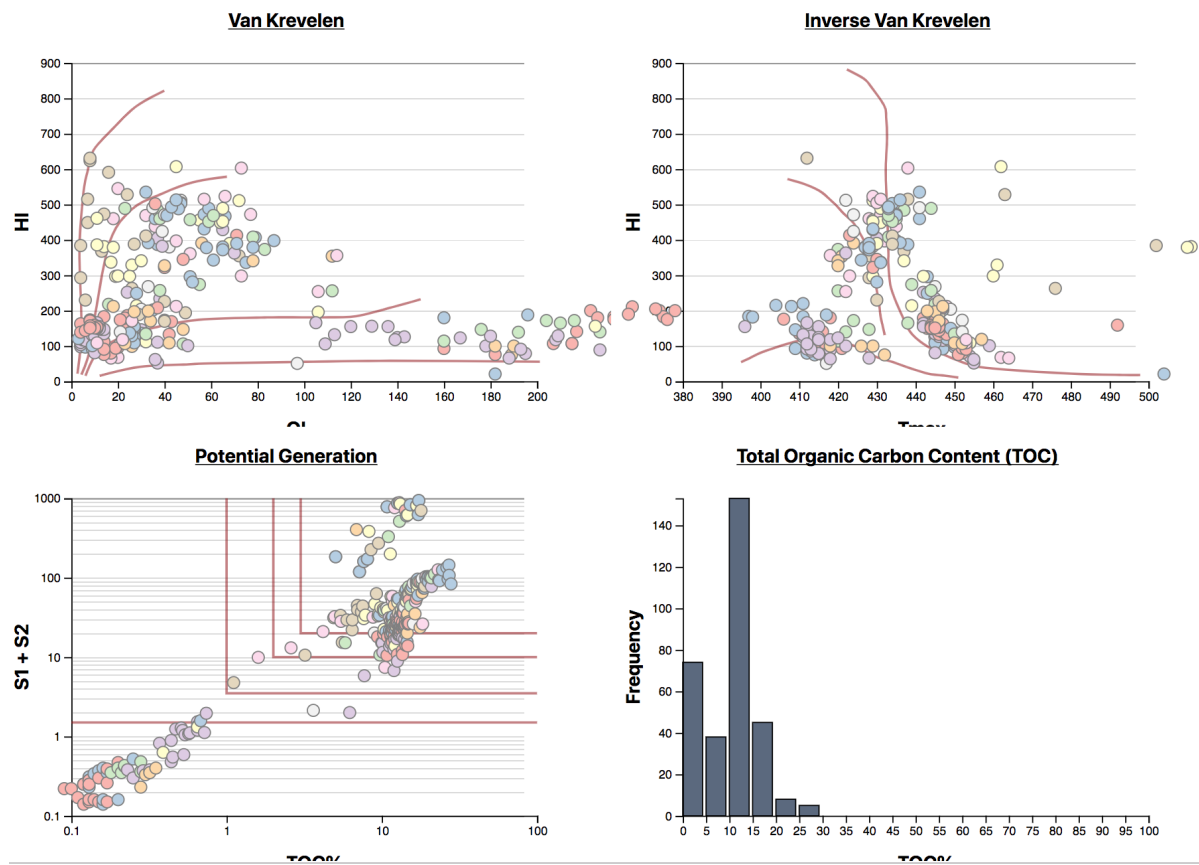


With borders:



Thus we ultimately decided to leave the borders off.

Williston Basin



We added the delimiter lines required and a title.

## Evaluation:

**What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?**

We learned that the geochemical data can have many angles to look at it, and deciding which way is the best for display takes time and a constant/ongoing redesign approach.

We answered the questions by giving the user the power of filtering the formation data by basin, formation and well. This is very important to be able to explore the data in different abstraction levels.

We would further improve the visualization by adding synchrony across plots. Also, adding a deeper layer of visualization (by including a depth chart) in which a user would be able to see the data spatially at a well-depth level.

Adding a dedicated comparator that allows users to select a number of formations and compare them side-by-side would also be an option.