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

PROCESS BOOK

By Skylar Shyu & Pablo Napan

Basic Information

Github Repo: https://github.com/psshyu/dataviscourse-GeochemOilandGas

Team Information:

Name	uNID	emails
Skylar Shyu	u1039726	skylar.shyu@utah.edu
Pablo Napan	u1077731	pablo.napan@utah.edu

Background 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 analisis 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.

Project Objectives

We are trying to accomplish to give our future visualization system users a birds-eye/detailed view of the geochemical system data of source rocks by displaying the data taken from hydrocarbon wells in the U.S.

Also, provide the users with a tool where they can interactively explore the data at different geospatial scales: data point, oil well and basin; and different time spans: from the paleozoic to the Cenozoic. Researchers and operator companies should be able to draw conclusions about: potential conventional/unconventional shale plays in marginal basins, production potential of source rocks, etc. by exploring the data visually and analyzing the specialized geochemical data plots presented.

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

Data Processing

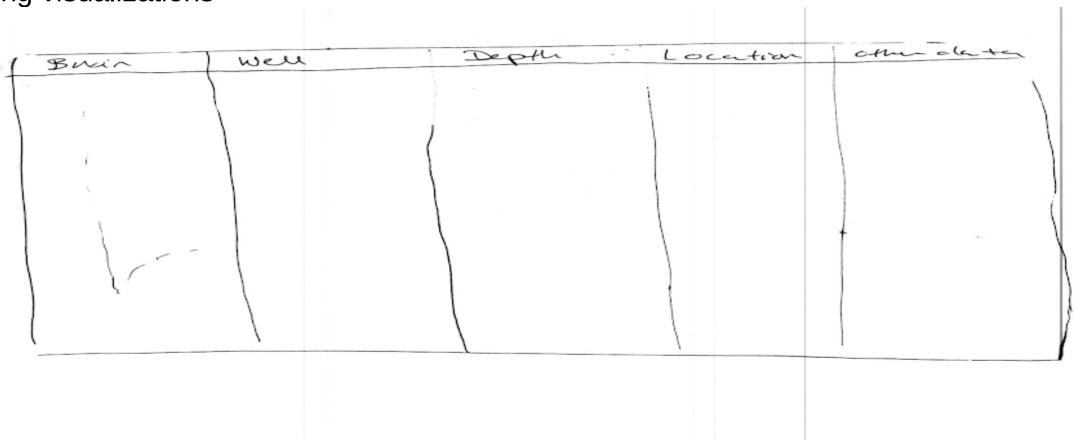
We expect to do a fair amount of data cleanup, as there's some inconsistencies regarding the availability/accuracy of the data for certain basins, groups, and wells and the samples that are extracted from them. This process requires some domain knowledge and discussion to ensure data and results significance.

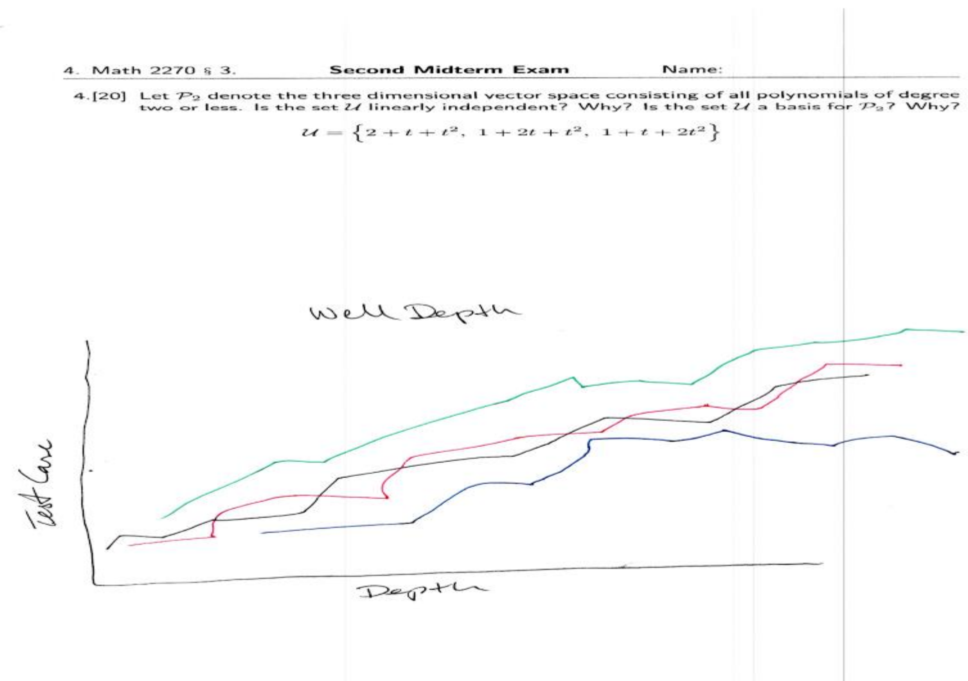
From the tables, we are planning to derive:

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

Visualization Designs

Brainstorming visualizations





2. Math 2270 § 3. **Second Midterm Exam** Name:

2.[20] Let $T : \mathbf{R}^2 \to \mathbf{R}^3$ be a linear transformation which takes

$$T\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1\\0\\-1\end{pmatrix}, \quad T\begin{pmatrix}1\\-1\end{pmatrix} = \begin{pmatrix}-1\\0\\-1\end{pmatrix}.$$

Find the matrix of the transformation. What is $T\begin{pmatrix}4\\3\end{pmatrix}$?

$$T\begin{pmatrix}4\\3\end{pmatrix} = \begin{pmatrix}3\\0\\-4\end{pmatrix}$$

3. Math 2270 § 3. **Second Midterm Exam** Name:

3.[20] Does the set S span $\mathbf{R}^3$? Why?

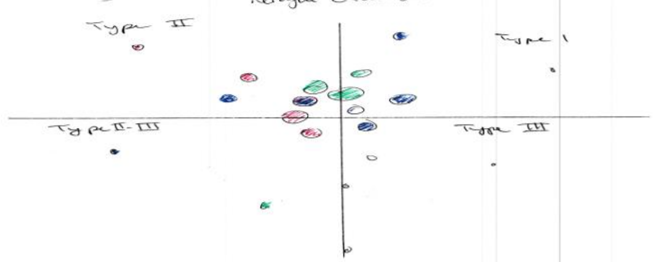
$$S = \left\{ \begin{pmatrix}1\\1\\-1\end{pmatrix}, \begin{pmatrix}1\\2\\1\end{pmatrix}, \begin{pmatrix}1\\4\\5\end{pmatrix}, \begin{pmatrix}5\\3\\-9\end{pmatrix} \right\}$$

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 5\\ 1 & 2 & 4 & 3\\ -1 & 1 & 5 & -9 \end{array}\right] \xrightarrow[R_3+R_1]{R_2-R_1} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 5\\ 0 & 1 & 3 & -2\\ 0 & 2 & 6 & -4 \end{array}\right] \xrightarrow{R_3-2R_2}$$

$$\begin{bmatrix} 1 & 1 & 1 & 5\\ & 1 & 3 & -2\\ & 0 & 0 & 0 \end{bmatrix}$$

The set S doesn't span $\mathbb{R}^3$

Keragen Classification



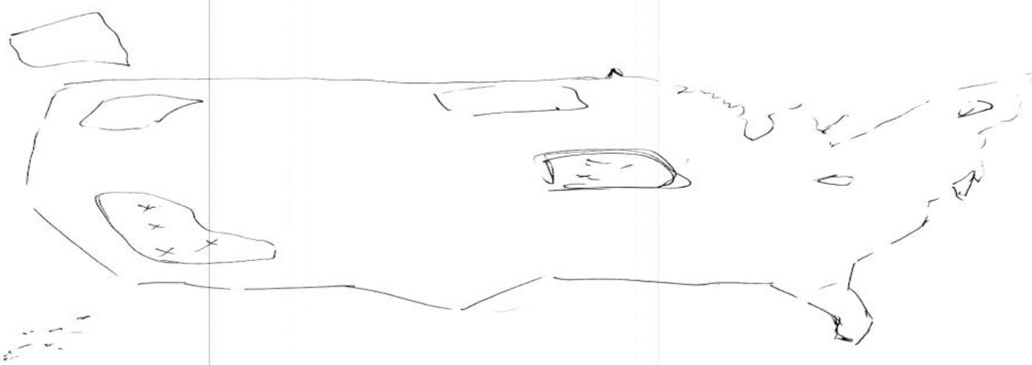
2. Math 2270 § 3. **Second Midterm Exam** Name:

2.[20] Let $T : \mathbf{R}^2 \to \mathbf{R}^3$ be a linear transformation which takes

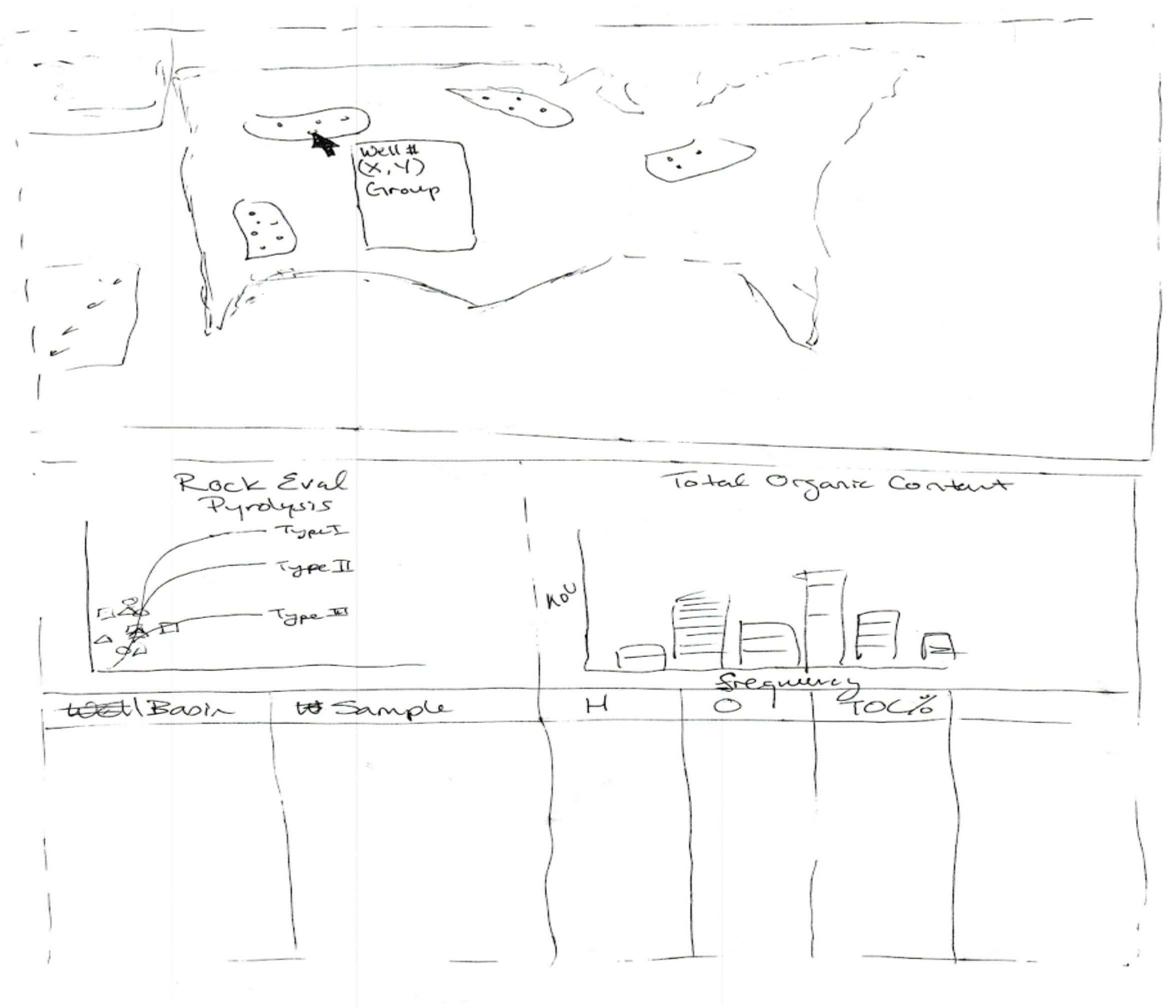
$$T\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1\\0\\-1\end{pmatrix}, \quad T\begin{pmatrix}1\\-1\end{pmatrix} = \begin{pmatrix}-1\\0\\-1\end{pmatrix}.$$

Find the matrix of the transformation. What is $T\begin{pmatrix}4\\3\end{pmatrix}$?

$$T\begin{pmatrix}4\\3\end{pmatrix} = \begin{pmatrix}3\\0\\-4\end{pmatrix}$$

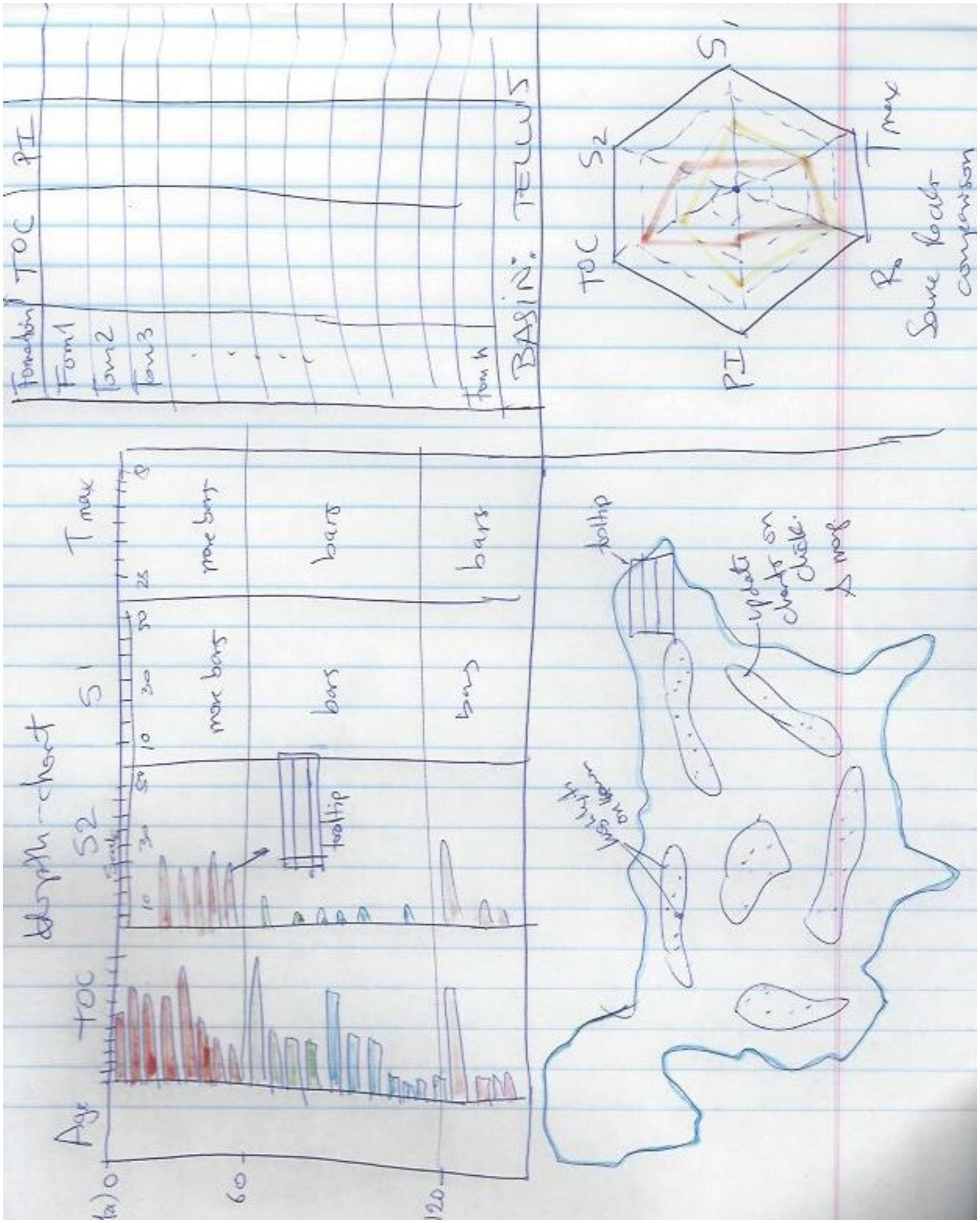


Sheet 1 (Not using):



We are not going with this design since it's fairly static. The table format creates a lot of redundant information, as one basin may have many wells that share similar features. While the tooltip is great for elaborating on a well's information, we are concerned that the dots may be too dense for users to hover over accurately. Furthermore, there isn't a lot of space here for us to elaborate on data specific to each basin.

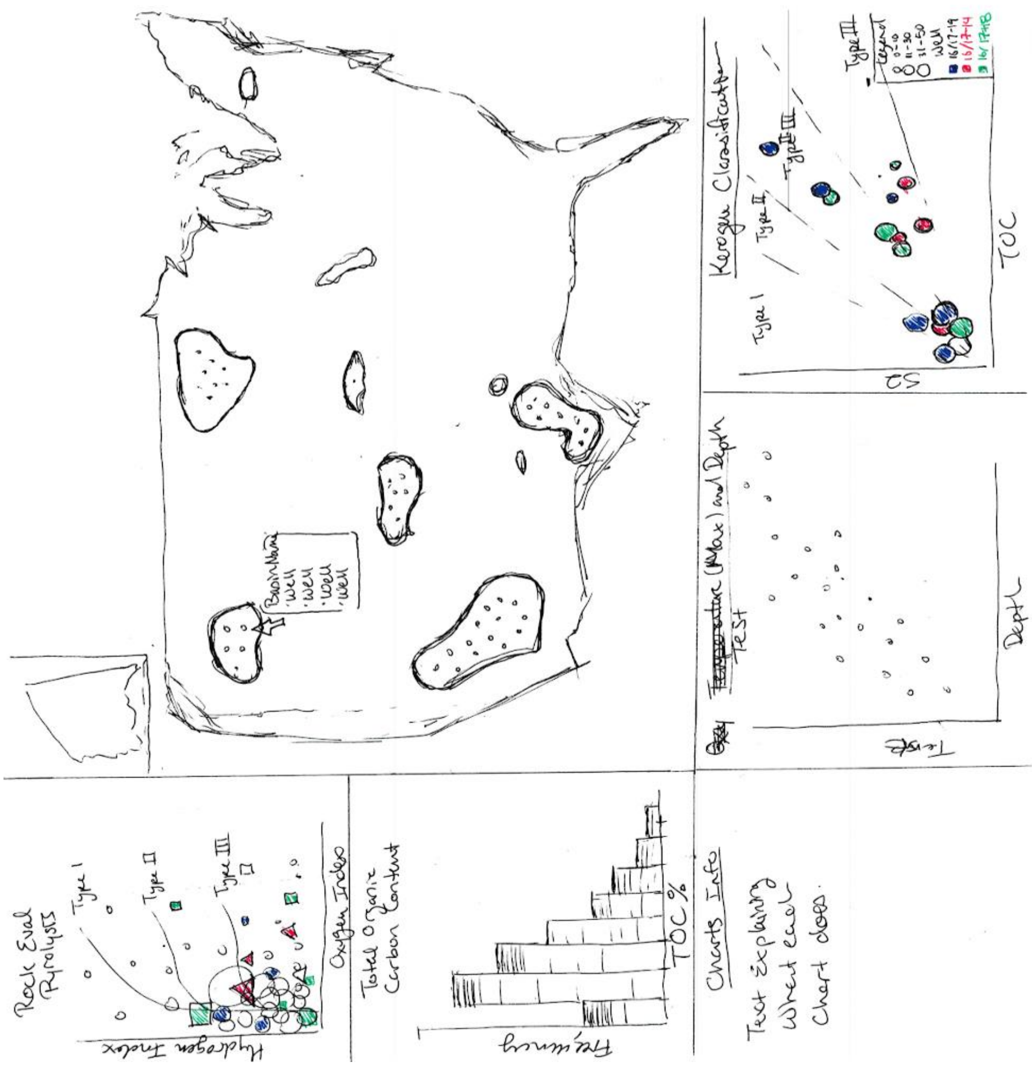
Sheet 2 (Not using):



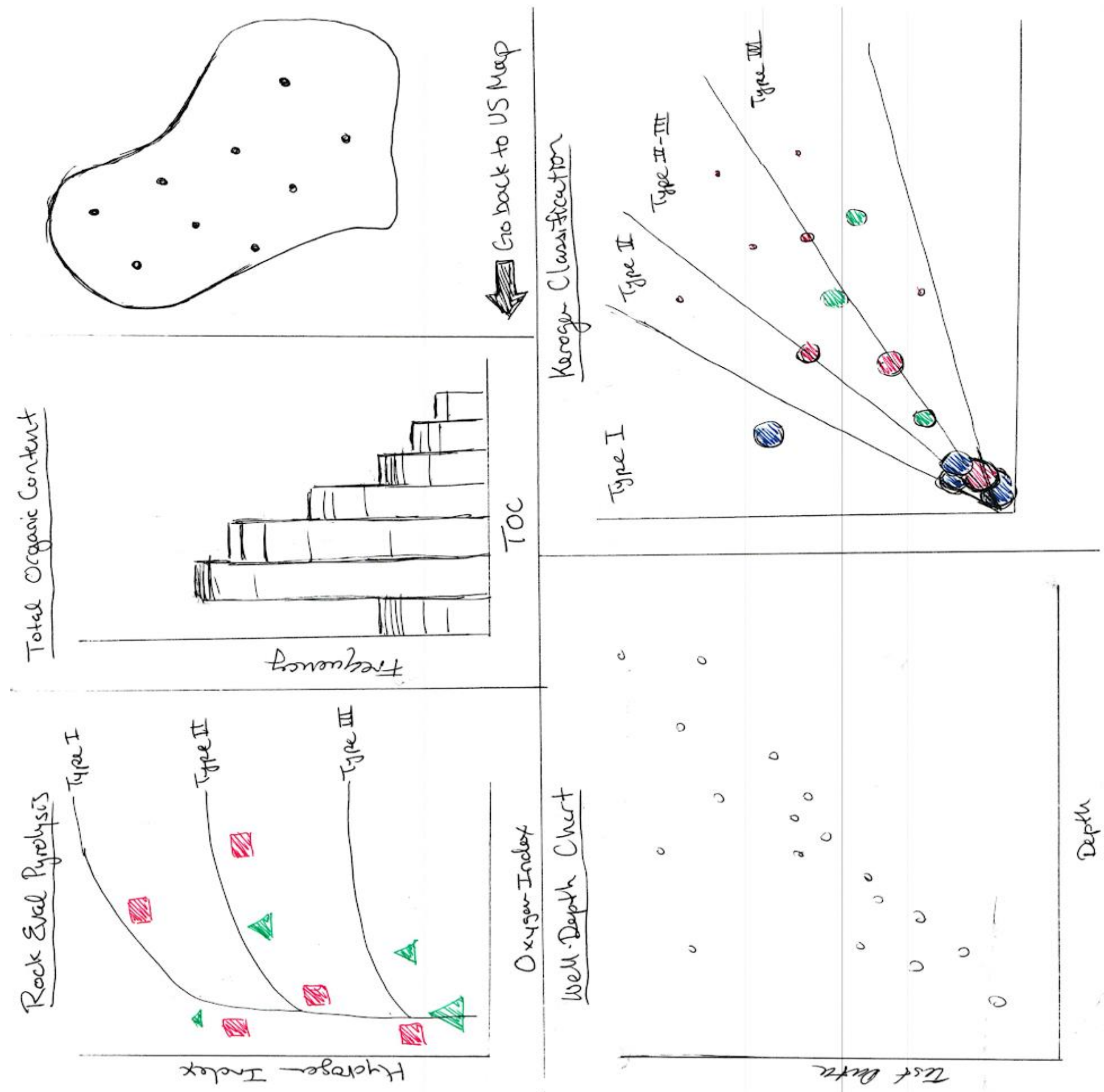
We are not going with most of this design either, because there are wells with so many data points that the bars would look too dense. Also, the hexagonal comparative chart, although it would have been nice to have, it is complicated to generate as it would need different scales/angles. And the summary table space, as Skylar suggested, could be use that for another visualization, so we have 4 final visualizations. Also, the map may be too small and may show very dense data points.

Sheet 3 (Using):

Screen 1: US Map



Screen 2: Charts showing basin information (after clicking on a basin in Screen 1)



***

We're opting to go with Sheet 3 where we can manipulate the sizes of charts with zoom-in shifting translations. Since the data for each chart is relatively dense and a tiny map isn't much use to anyone, we wanted our design to allow the audience to focus on one or the other.

Screen 1: US Map

To do so, our design will initially present a topological map of the US with basin/well overlays as the focal point, taking up about 75% of the screen. This will allow users to select the basin they wish to explore further; upon doing so, the map is no longer as vital since the focus is on the selected basin. Thus, we will zoom in on the selected basin while minimizing the size of the map chart as a whole, then move it to a corner, taking up about ⅙ of the screen. Other charts can be small as they're meant to represent the total data set.

Screen 2: Charts showing basin information

Whenever a basin gets clicked, 4 charts will show up:

- **Van-Krevelen diagram**: HI vs OI scatterplot (TOC encoded in plotpoint size).
- **TOC-chart**: stacked bar chart showing the contribution of each formation to the overall TOC.
- **S2 vs TOC:** scatter plot: showing the kerogen classification.
- **Well-depth chart**: Important for visualizing the how the geochemical varies in the Z dimension.

The channels in each plot (bars, circles) will highlight whenever a well (inside a basin) gets hovered over. Tooltips will be used when necessary, especially when explaining what certain abbreviations stand for, the significance of selected metrics, etc.

Must-Have Features

- An interactive U.S. map that allows zooming in of basins and hover-over of individual oil wells.
- Van-Krevelen diagram: a scatterplot to display Rock Eval Pyrolysis (hydrogen index & oxygen index)
- A stacked barchart: Total Organic Carbon (TOC) chart describing the frequency of the amount of total organic carbon.
- Well-depth charts for each individual well; these would including information such TOC, Tmax, vitrinite reflectance, S2, hydrogen index, etc.

Optional Features

- Provided that we have enough time, we will also allow implement a depth-chart that shows the geochemical data scaled in the Z (depth) axis and/or a slider that allows users to analyze the temporal evolution of the most potential source rocks in the U.S.

Project Schedule

Week	Tasks/Goals
Sat. Oct. 27th - Fri. Nov 2nd	Pablo • Data gathering, wrangling, integrity assurance, refining, outlier removal and standardizing. • Initialize Google Doc for Process Book Skylar • Spin up a basic boilerplate project with empty svgs (representing charts) that can downsize and move to correct places upon clicking spots on a map svg

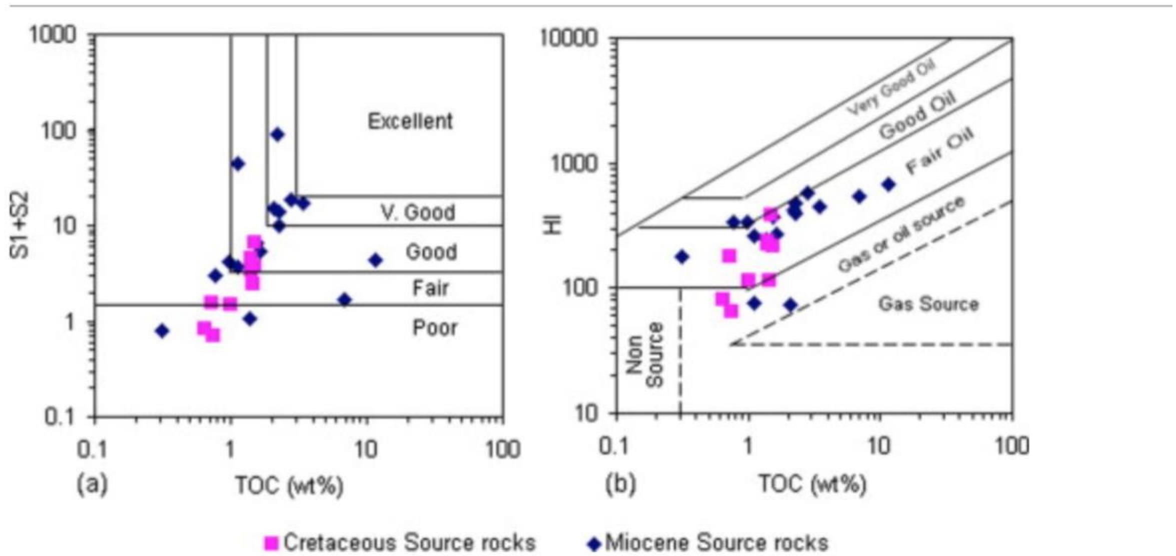
Sat. Nov. 3rd - Fri. Nov. 9th	(Project Milestone is due on the 9th)
	Pablo
	• Download basins shapefiles, filter and convert them to geojson or topojson.
	• Map implementation
	
	Skylar
	• Implement the topological map
	    ○ Basin overlays as clickable regions
	    ○ Ensure that the coordinates of wells can be returned and are accurate when clicked

Sat. Nov. 10th - Fri. Nov 16th	Pablo • Basin rendering and first scatter plots implementation (V-K and depth-Chart) Skylar • Start implementing graphs (TOC-Chart & S2 vs. TOC)
Sat. Nov. 17th - Fri. Nov. 23rd	Pablo • Heavy work on implementation of any remaining chart and refine details. • If there’s time, slider implementation. Skylar • Finish implementing graphs
Sat. Nov. 24th - Fri. Nov. 30th	Skylar & Pablo • Fix any last minute bugs

Related Work

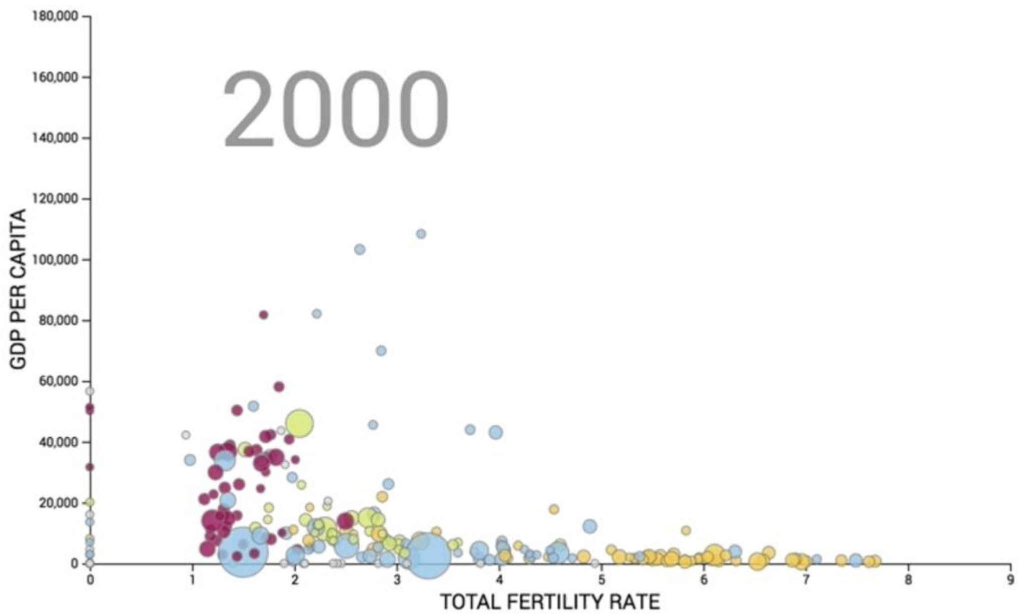
Most of the inspiration on the *technical design* (which plots or charts to display) came from the paper (1) **Evaluation of organic matters, hydrocarbon potential and thermal maturity of source rocks based on geochemical and statistical methods: Case study of source rocks in Ras Gharib oilfield, central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

316 P. Farrimond *et al.*

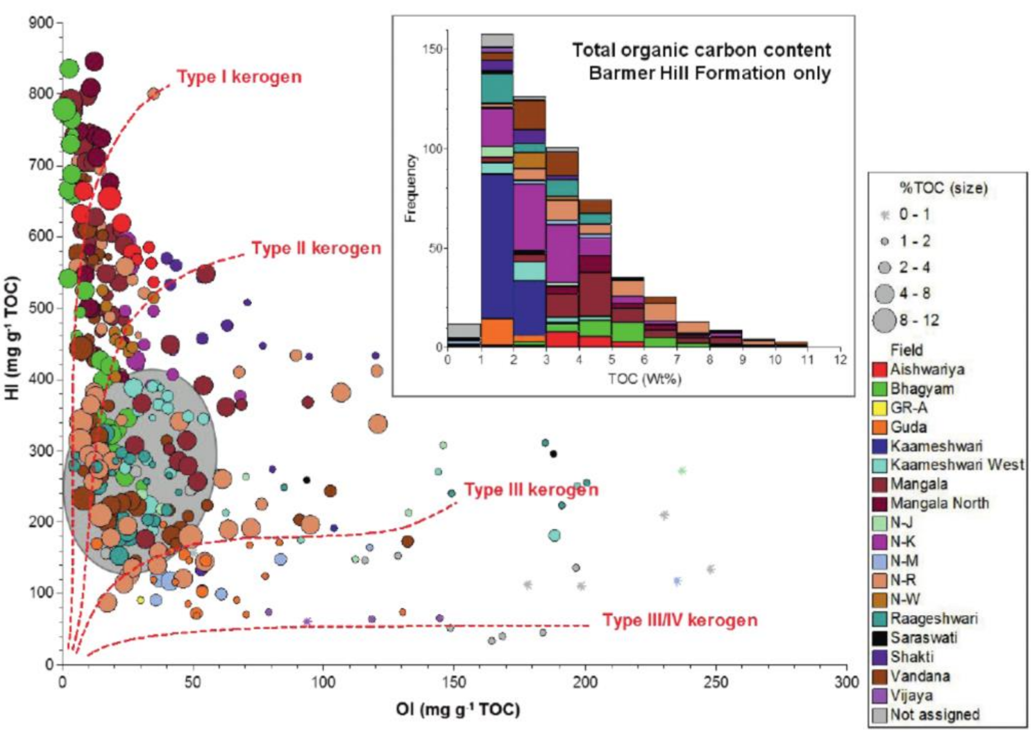
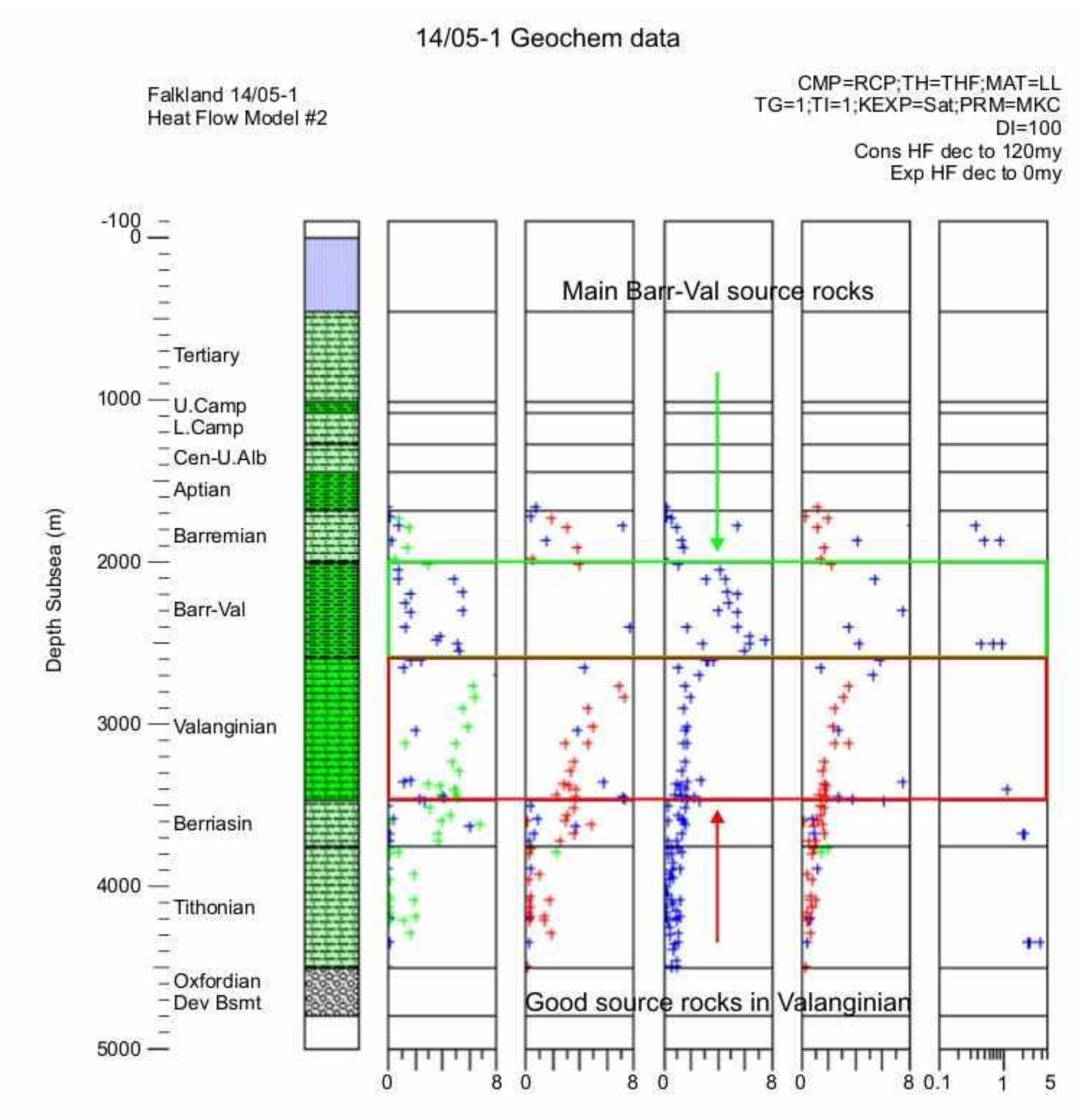
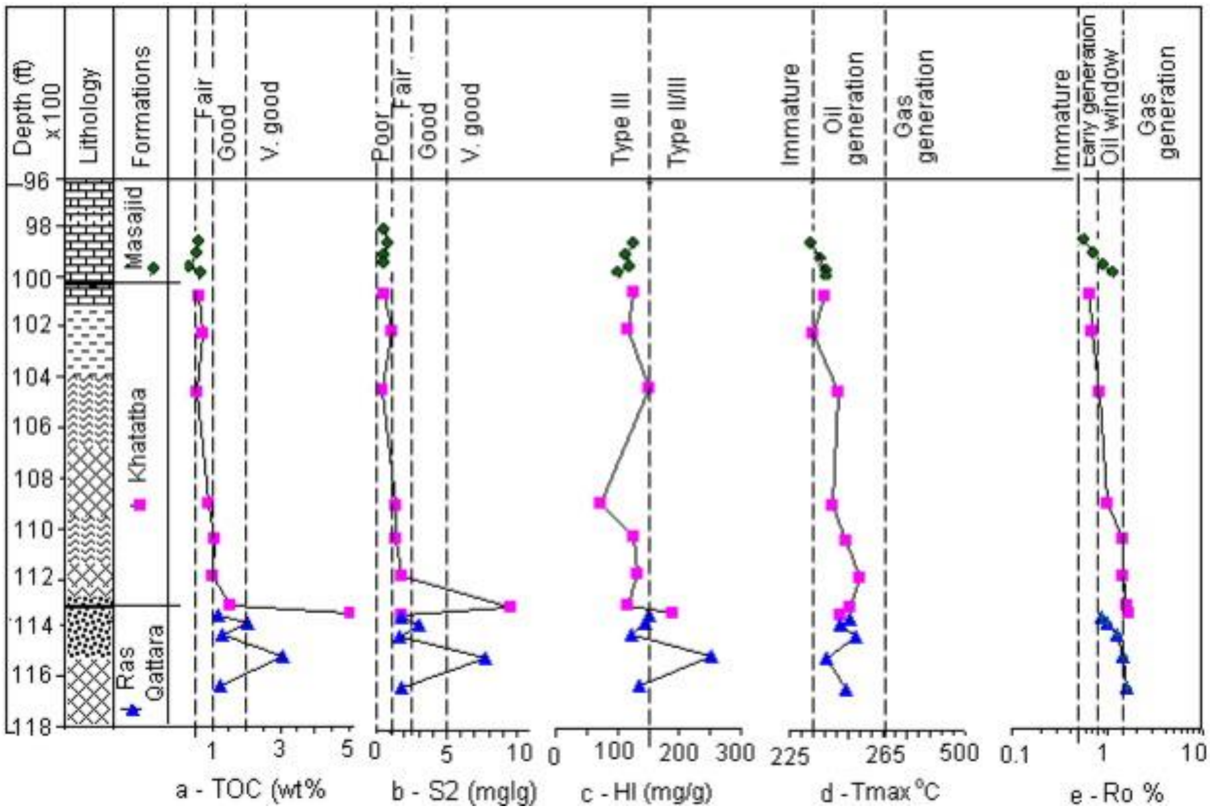


Fig. 17. A histogram of total organic carbon (TOC) content and pseudo Van Krevelen diagram for Barmer Hill Formation samples only. The grey shaded area indicates where samples from the southern Barmer Basin mainly plot (Kaameshwari West, Raageshwari and Guda).

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

What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

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*?"

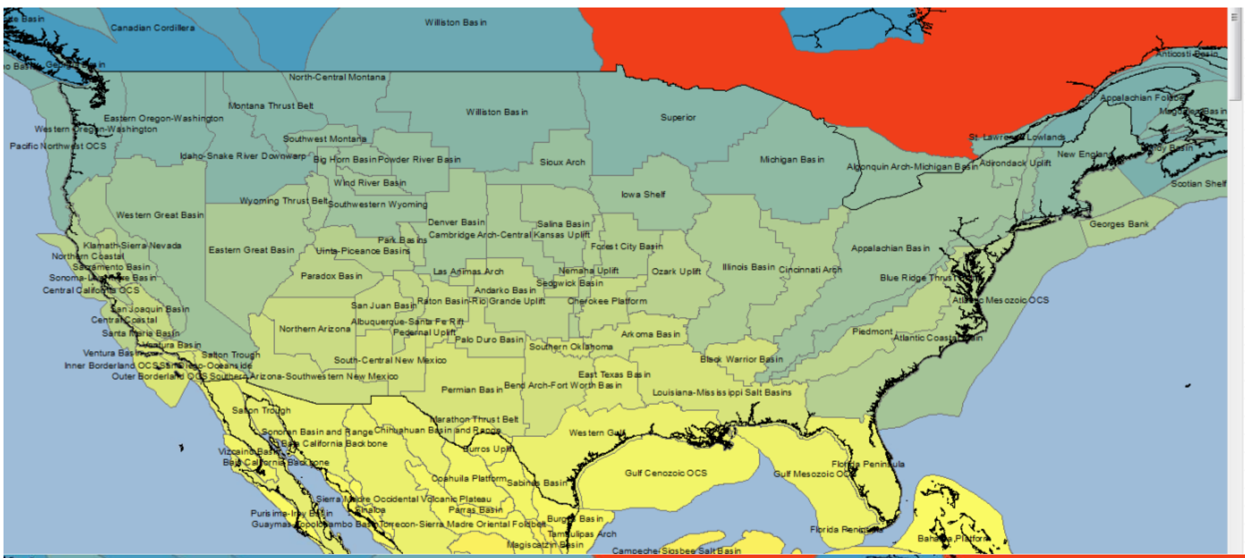
Other questions considered and that may be still floating around are:

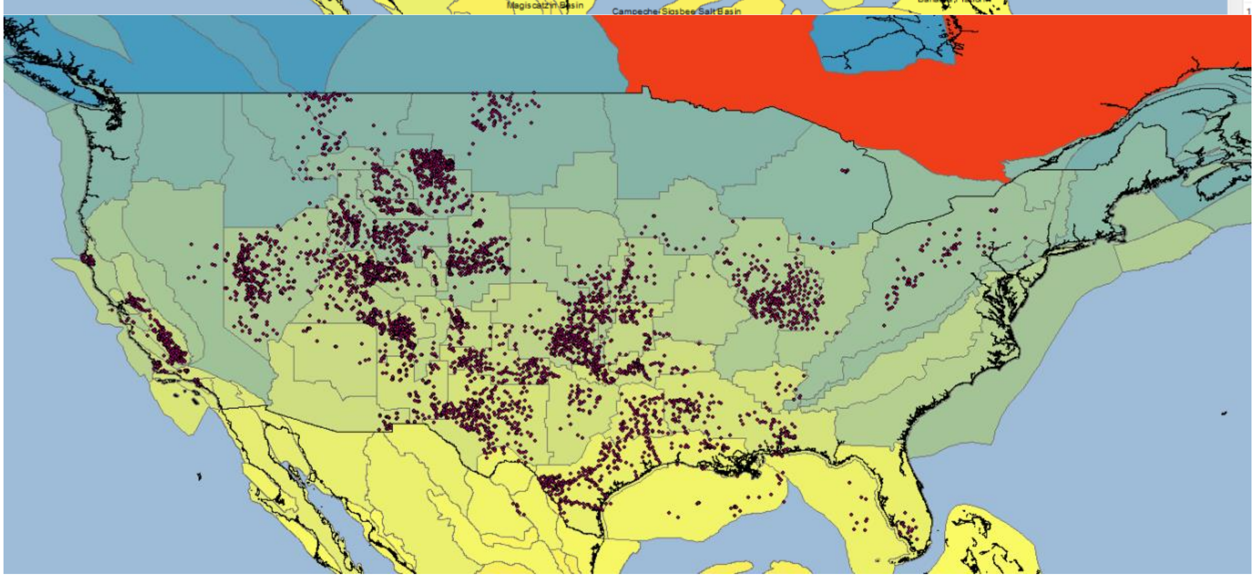
"*What are the spatial distribution of the geochemical samples in the Z dimension (depth)*"
"*What are good/the best source rocks for recommending exploration*"

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.

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





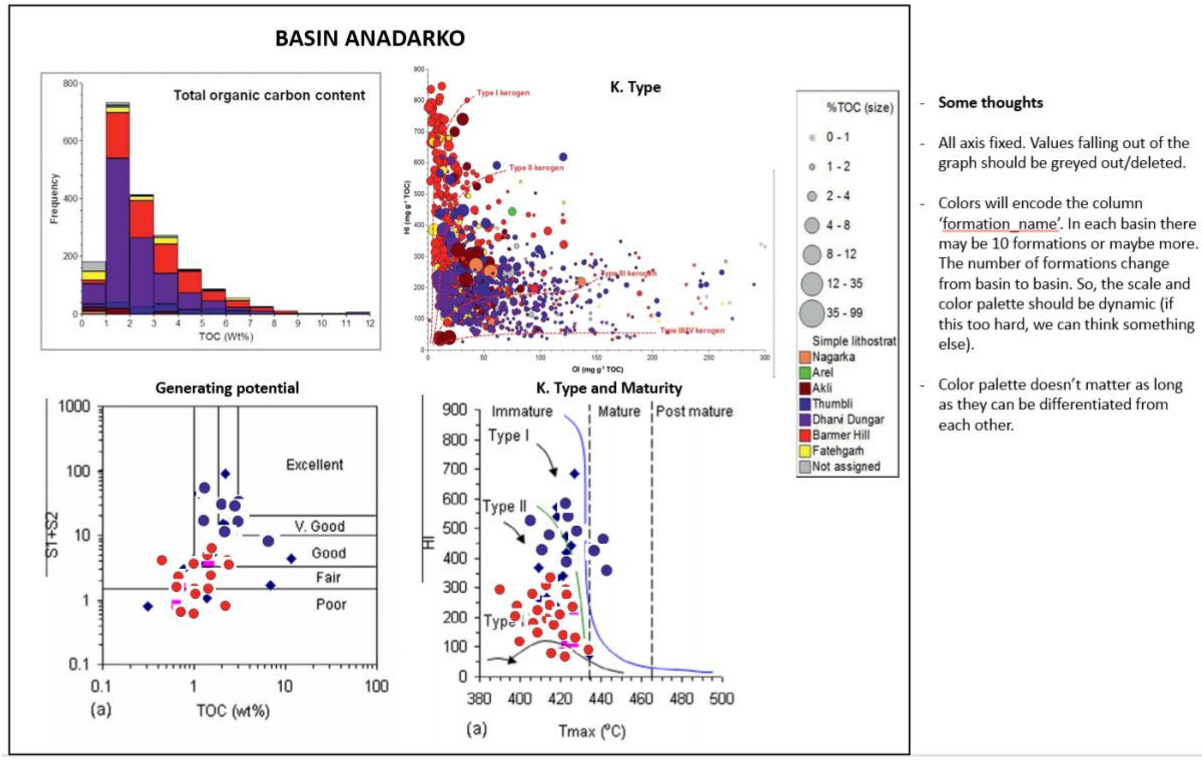
Also, the data, due to its tabular nature, was explored with Microsoft Excel, essentially by doing multiple and consecutive filters to gain insights on the maximum number of data points that would be display on screen at each time. We discovered that in some cases, there would be >90 formations displayed 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 shal	Marine	0.07
Minnelusa	Shale; Car	Black shal	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 shal	Marine	10.44
Minnelusa	Shale; Car	Black shal	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 shal	Marine	10.54
Minnelusa	Shale; Car	Black shal	Marine	10.4
Minnelusa	Shale; Car	Black shal	Marine	27.04
Minnelusa	Shale; Car	Black shal	Marine	9
Minnelusa	Shale; Car	Black shal	Marine	3.8
Minnelusa	Shale; Car	Black shal	Marine	3.14
Minnelusa	Shale; Car	Black shal	Marine	3.17
Minnelusa	Shale; Car	Black shal	Marine	2.4
Minnelusa	Shale; Car	Black shal	Marine	0.36
Minnelusa	Shale; Car	Black shal	Marine	0.65
Minnelusa	Shale; Car	Black shal	Marine	18.02
Minnelusa	Shale; Car	Black shal	Marine	7.4
Minnelusa	Shale; Car	Black shal	Marine	13.94
Minnelusa	Shale; Car	Black shal	Marine	9.4
Minnelusa	Shale; Car	Black shal	Marine	15.01
Minnelusa	Shale; Car	Black shal	Marine	0.72
Minnelusa	Shale; Car	Black shal	Marine	1.44
Minnelusa	Shale; Car	Black shal	Marine	2.12
Minnelusa	Shale; Car	Black shal	Marine	7.68
Minnelusa	Shale; Car	Black shal	Marine	1.59
Minnelusa	Shale; Car	Black shal	Marine	8.37
Minnelusa	Shale; Car	Black shal	Marine	11.83
Minnelusa	Shale; Car	Black shal	Marine	0.78
Minnelusa	Shale; Car	Black shal	Marine	0.12
Minnelusa	Shale; Car	Black shal	Marine	1.75
Minnelusa	Shale; Car	Black shal	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 shal	Marine	0.4
Mowry Shale	Shale	Shale	Deltaic to Shallow marine	4.2
Poison Canyo	Shale	Shale		0.06
Poison Canyon		Sandstone		0.01

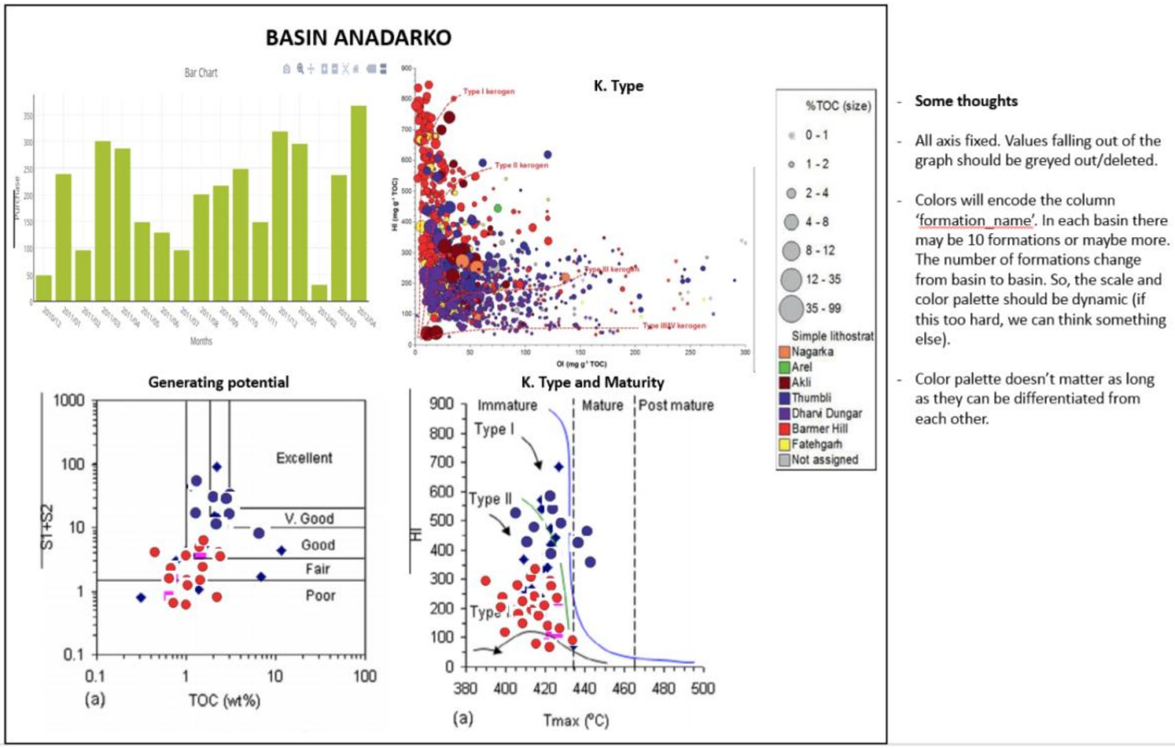
Also, barcharts and scatterplots were displayed (image below) to explore for outliers. With this analysis. This helped us realize that we should go for fixed rather than dynamic axis in all of our plots.

Design Evolution

The group of visualizations proposed at first are shown below.

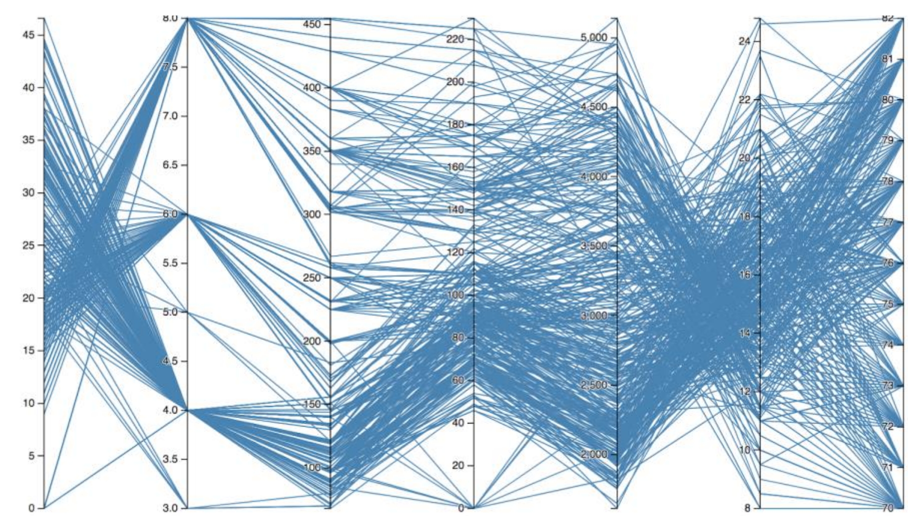


Then, after exploring the data and realizing the possibility of >90 colors in our charts (explained in the previous section) we decided to go for a simple bar chart that would update to its respective formation whenever a sample is clicked in any of the other 3 plots.



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-catchy 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.



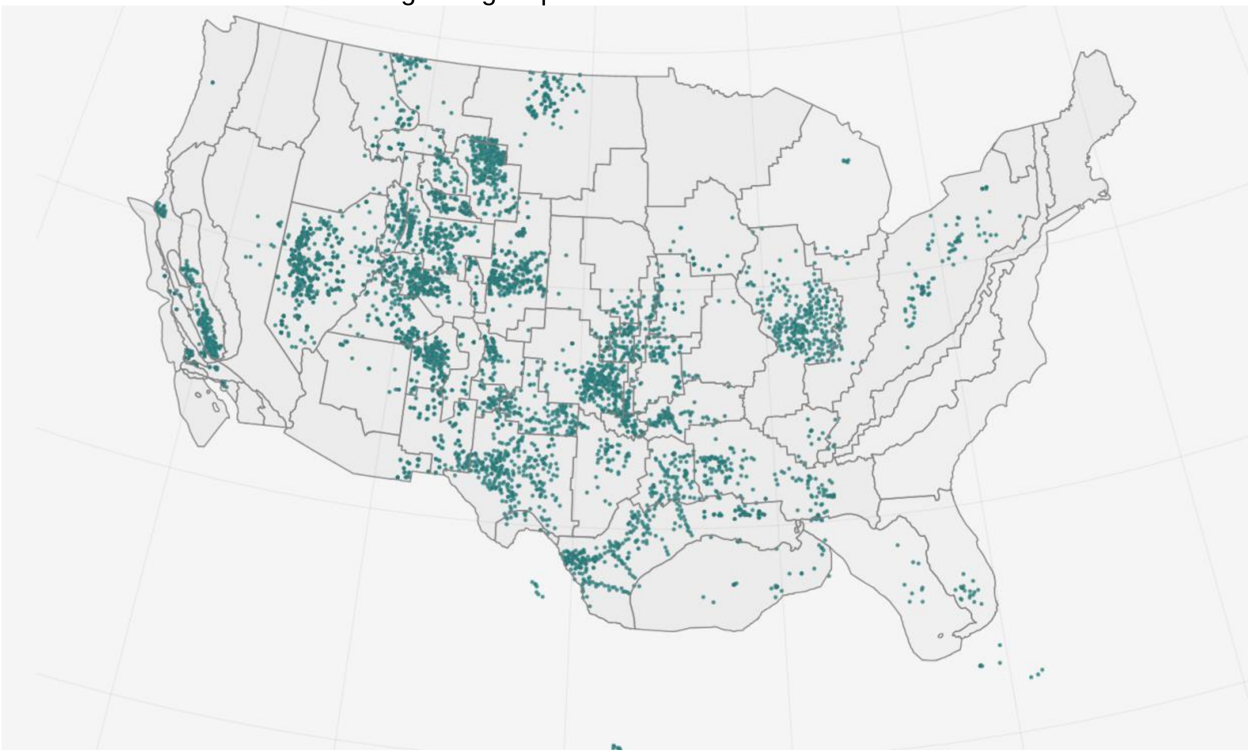
So, after deciding that we had to go with scatter plots, we analyzed the possibility of implementing a scatterplot matrix with synchronized brushing. This idea was dropped quickly as, since we need to plot all of our parameters with themselves, there will be some scatter plots in the matrix showing relationships (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.

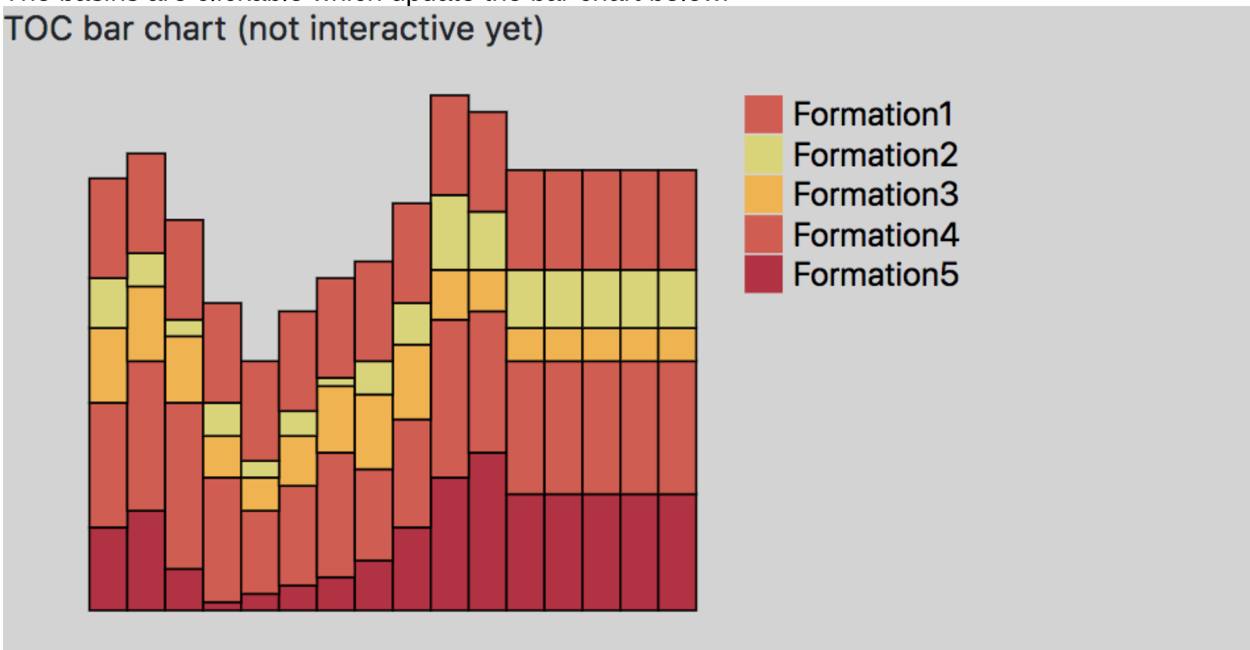
However, to respond to the peer-feedback, it was decided to either add more interactivity/functionality (search bar, dropdown, etc.) or draw an extra plot (depth plot) which would add one extra layer of data exploration. At this point, we have not decided which of these two options to take.

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:



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

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?

In progress