RaCAViz: Interactive Visualizations for Rapid Carbon Assessment

1nd Du Nguyen
Computer Science Department
Sam Houston State University
Huntsville, Texas, United States
dhn005@shsu.edu

2st Vung Pham
Computer Science Department
Sam Houston State University
Huntsville, Texas, United States
vung.pham@shsu.edu

2022 IEEE International Conference on Big Data (Big Data) | 978-1-6654-8045-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigData55660.2022.10020463

***Abstract*—Soil organic carbon is an essential element of environmental quality assessment because carbon dioxide is the main greenhouse gas causing global warming and climate change. Trees and plants acquire a lot of carbon in their lifetimes, and soil helps keep this portion of carbon from releasing to the environment. Thus, understanding the current carbon storage in the soil and the relationship with tree richness is important. The soil organic carbon datasets and corresponding tree information are available. However, the current visualizations for these datasets are either static visualizations that report initial findings to the public or visualizations that are too complicated to be useful for a broad audience. Therefore, this work proposes an interactive visualization solution for analyzing soil organic carbon values distribution and their relationship with tree heights and diameters at breast heights, called RaCAViz. The design of this interactive visualization is based on the analytical evaluation of visual and interaction idioms to tackle typical analysis tasks on this type of dataset. However, this paper also provides a specific use case to illustrate its usefulness in bringing different perspectives on soil organic carbon datasets to a broad audience.**

***Index Terms*—visual analytics, interactive visualizations, rapid carbon assessment**

I. Introduction

Soil organic carbon (SOC) is critical in the environmental quality assessment process [1]. Trees and plants acquire a large amount of carbon in their lifetimes, and soil helps to keep this amount of carbon from releasing to the environment when the trees and plants are broken-down. Weather conditions may impact this carbon storage in the soil [2]. For instance, colder regions preserve this carbon storage for longer than warmer ones. However, urbanization and deforestation disturb the soil in all areas and release this carbon storage as carbon dioxide (CO_2) into the air. On the other hand, agricultural activities may help or harm this carbon storage. For instance, abusive tillage activities harm, and proper types of trees and plants help.

CO_2 is the main greenhouse gas causing global warming and climate change [1]. Therefore, the United States Department of Agriculture (USDA), Natural Resources Conservation Service (NRCS) initiated the Rapid Carbon Assessment (RaCA) project [3] to statistically estimate the amounts and distributions of carbon stocks for the U.S. soils under various land covers and types of agricultural activities. Analyzing this dataset helps to plan land usage change, agricultural management, conservation practices, and climate change.

This project is funded by the College of Science and Engineering Technology Undergraduate Research Award from Sam Houston State University.

This project produced a dataset with 144,833 soil samples from 32,084 soil profiles. These profiles are uniformly distributed in 6,017 locations over the conterminous United States (CONUS). Besides the inorganic and organic carbon storage, this dataset also provides information about soil morphology, landscape characteristics, vegetation, and agricultural activities.

Though this dataset is important and helpful, the visualizations used for analysis and assessment are limited. Specifically, limited types of visualizations exist for this dataset (mostly heatmap, Choropleth map, and point map). Furthermore, these generated visualizations are static. The lack of interactions hinders analysis tasks that require interactions to support looking at the data from different perspectives [4]. Therefore, the contributions of this project include:

- Provide an easy-to-use visualization tool on the web interface to improve the accessibility of soil/plant carbon sequestration analysis.
- Incorporate interactions into the visualization tool to support different analysis scenarios.
- Educate and answer questions for stakeholders about carbon inventory and its correlation to tree richness over the conterminous United States.

We hypothesize that this solution helps the analysis of the RaCA data and provides further insights into the carbon storage distributions and their correlation with the trees and cultural activities. These insights help recommend proper activities such as what plants/trees to grow and tillage activities. These activities, in turn, help to combat climate change.

II. Related Work

There are various popular interactive data visualizations used to analyze soil profiles exist [4]–[8]. However, this work focuses on the carbon storage in soil assessment using RaCA dataset. The dataset is the first and foremost important part of any data analytic project. Interested readers can refer to the

RaCA project page [3] for the project description, including why it is important, the measurement methods utilized, and the data available as the results. Furthermore, Willis at al [9] also provides an excellent overview of this project, including sampling design, initial summary, and uncertainty estimates.

In terms of visualizations of this dataset, the RaCA project has some initial visualizations. However, these visualizations are for reporting findings to the public, and these findings include factual data overview and initial statistical analyses. All of the provided visualizations from the RaCA project are static. This limitation is reasonable because the primary purpose of these visualizations is to report the results to the public. However, interactive visualizations are necessary to further analyze the dataset from different views and domains.

Another data visualization project that works on the RaCA dataset is called CarbonScapes[1]. This project aims to improve access to and display the USDA carbon dataset. Its target users include USDA personnel, researchers and scientists from other agencies, and the public. The great strength of this project is its extensive work in setting up the requirements for a visualization toolkit to work on this type of dataset. Specifically, USDA Team assembled to formulate requirements and review the final product. Here are the main requirements found by the team [10]:

- A quick snapshot of different carbon pools by particular geographies
- A visual and investigative view of USDA terrestrial models and data
- Access to current USDA terrestrial carbon models and data
- Educational resources for K-12/STEM classrooms or scientific communities about terrestrial carbon landscapes.

Unfortunately, the resulting website for CarbonScapes is heavy because it needs to accommodate many different needs of different types of users and requirements. Furthermore, it is error-prone. Therefore, this project is also built upon the found requirements by CarbonScapes. However, it is a lightweight, fast, and focused solution to display the overview of carbon storage over CONUS at different soil horizons. Furthermore, we focus on the correlation of the carbon stock amounts and the tree information over CONUS.

III. Data and Analysis Tasks

A. RaCA dataset

From the USDA RaCA database, 142,445 samples from 31,565 unique Pedon IDs and data from 6,411 sites across the United States are downloaded for analyzing soil organic carbon (SOC). Pedons have 14 different soil horizons, which are used in the visualization. After filtering 14 soil horizons used in this project, 27,040 samples remain. Besides the samples, tree heights and tree diameter at breast height values are extracted within 1,634 different RaCA sites.

With each unique Pedon ID, the sample and site dataset is joined together to attain one of the resulting datasets that include SOC measurements corresponding to a unique soil horizon category from each state with its coordinates. Out of 27,040 samples of 14 soil horizons, 76 samples contain missing values for the state names. Therefore, the remaining 26,964 samples are used for visualizing point maps.

After joining the two original ones, the second dataset is the average SOC measurements from each soil horizon within that specific state. After grouping samples on each soil horizon category and state, the resulting dataset includes 465 rows where each state has the average SOC for its associated soil horizontal name. Choropleth maps and bar charts are created based on this specific dataset.

The next dataset results from joining the sites and trees dataset on the corresponding Site ID. The resulting 52,938 rows of data contain 23,653 missing values from both diameter at breast height (DBH) and height. Thus, 29,285 rows of data are being grouped by each state, and the averages for each DBH and height values from that particular state are calculated. From there, the Choropleth maps and bar charts for DBH and height information at each sampled location is generated.

B. Analysis tasks

Different users would like to work with different aspects of the dataset. For instance, some users emphasize states, while others work with counties, Major Land Resource Areas (MLRA), Ecoregions, or basins. We want to make the website lightweight and focus on specific tasks in this project. Therefore, we focus on analyzing at the state level. However, the concepts proposed in this work remain the same for other levels, such as counties or MLRA. Also, the attributes of concern include the distribution of soil organic carbon (SOC) and corresponding tree heights and tree DBH values. These attributes are called “analyzing attributes” hereafter. Specifically, there are three main analysis tasks that this proposed solution targets:

1. Qualitative Overview (**T1**)
2. Quantitative Analysis (**T2**)
3. Correlation Analysis (**T3**)

The following sections give details about the proposed interactive visualization components to support these analysis tasks in this current work.

IV. Proposed Visualization Solution

A. Schematic overview

The proposed visualization tool (RaCAViz) has three main components as depicted in Figure 1. They are the data querying and processing component, qualitative analysis component, and quantitative analysis component. The qualitative and quantitative analysis components are supplemented with interactions to support exploring and viewing data from different perspectives to find and report insights.

The first component is data querying and processing. The querying sub-component helps to query data from the RaCA project [3]. The processing sub-component helps aggregate data at different resolutions (e.g., per location, per site, or per

[1]http://carbonscapes.org

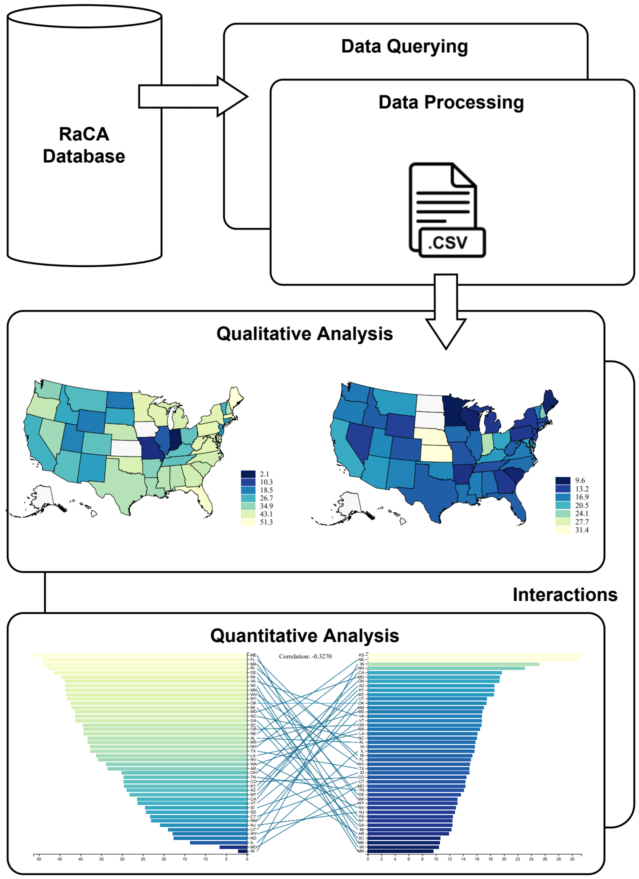

Fig. 1. RaCAViz Schematic Overview with three main components: (1) The data querying and processing component, (2) the qualitative analysis component, and (3) the qualitative analysis component. Furthermore, interactions are incorporated into the qualitative and quantitative analysis components to support exploratory analysis.

state). Details of this component are discussed in Section III-A. The following sections describe the other two components (Qualitative Analysis and Quantitative Analysis components). Specifically, Figure 2 shows the main view of the proposed solution with the qualitative and quantitative views. Furthermore, they are backed with interactive options to offer different perspectives of the underlying data.

B. The qualitative component

The qualitative analysis component provides views for qualitatively displaying the overall distribution of the carbon storage at a selected soil horizon (or the overall average over all horizons) on the left view (analysis task **T1**). Similarly, there is a corresponding map on its right view to display the tree information at the sampled locations. Specifically, users can select the diameter at breast heights (DBH, 1.3 m or 4.3 ft above ground [11], [12]) or the measured heights of the trees.

For these qualitative views, there are options for different display styles. They are Point Map, Choropleth Map (default), and Symbol Map. Point Map is a natural option for displaying the distribution of spatial data. It uses the points as marks and spatial channels to encode the sampled sites' x and y locations. Furthermore, the color channel encodes the value of the measured SOC (on the left view) or tree information (on the right view) of a site. Figure 3 shows an example of the Point Map to represent the overall distribution of the average SOC values over the 6,017 sampled locations. Observably, the east side has higher SOC values.

The Point Map is useful to show the overall view of the sampled locations and the distribution of corresponding SOC values. However, this visualization idiom becomes too cluttered in this case due to many sites (6,017 locations). This cluttered issue means points are covering one another and hinder the view of the color channel, making it hard for the users to view the encoded values. Therefore, an alternative view is the Choropleth Map. In Choropleth Map, the shapes, sizes, and locations of the analysis regions (states in this case) are fixed. Therefore, it uses the color channel to encode the analyzing quantitative attributes (average SOC, the tree height, or the tree DBH values per state).

Choropleth Map is advantageous compared to the Point Map that it aggregates the data at the state level and does not suffer from cluttering issues. These advantages come with the trade-off of losing details about sampled site distribution. However, as the result of data aggregation, this view option provides an excellent way for the users to look up (due to users' familiarity with the map) or view the qualitative distribution of the analyzing quantitative values. Figure 4 shows the Choropleth Map of the SOC distribution at the 'Oi' soil horizon aggregated at the state level. Observably, the visualization is cleaner and can depict the overview distribution well.

Choropleth Map also has some limitations. Specifically, using the color channel to encode quantitative data is less effective (e.g., compared to position, length, or size) because of the non-linear relationship between the color channel's intensities and corresponding humans' perceptions. Furthermore, it also has interference, such as the small region's size making the color less perceivable or the color of one region might be interfered with by the colors of the surrounding regions [13].

Therefore, this solution proposes an alternative view using Symbol Map [14] Symbol Map, as the name indicates, uses symbols (circles in this specific case) to represent data on top of the map. Instead of using the color channel, Symbol Map uses the size channel (sizes of the circles) to encode the analyzing quantitative attributes (i.e., SOC, tree height, or tree DBH). This size channel is more effective in representing the quantitative attribute with the trade-off of being less aesthetic compared to the Choropleth Map.

Furthermore, in this specific case, the sizes of the circles for some states with high values still overlap. Therefore, the transparency channel is used to mitigate this issue. Figure 5 is the Symbol Map to visualize the same information as the Choropleth Map in Figure 4. It still has the ease of looking up as the Choropleth, but the circle sizes are more efficient than the color in representing the SOC values.

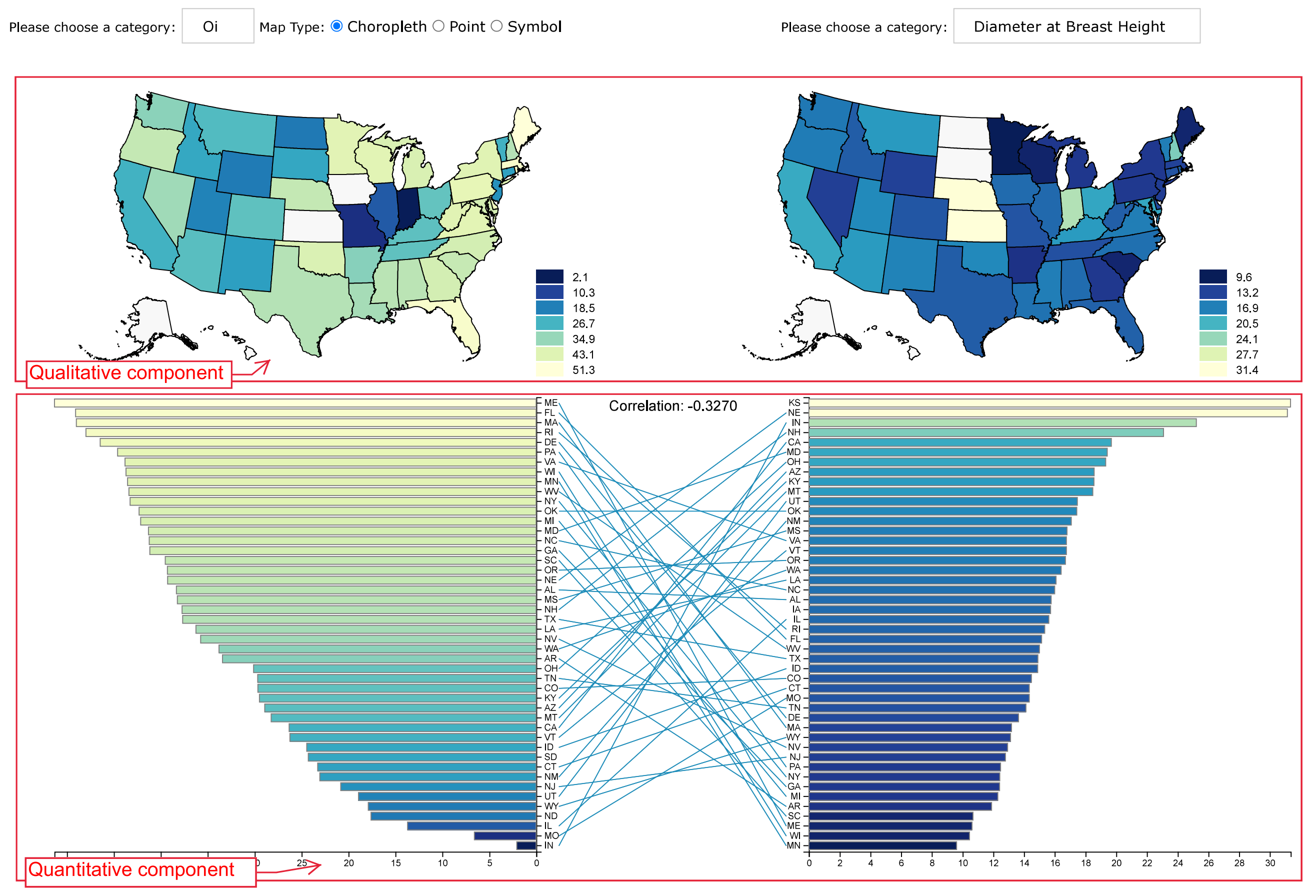

Fig. 2. Main view of RaCAViz with interaction options (top), qualitative views (top panel), and the quantitative views (bottom panel).

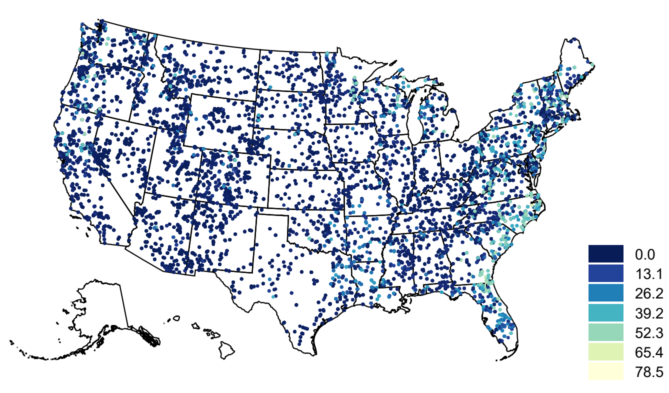

Fig. 3. Point Map of 6,017 sampled locations. This visualization helps to show the overall sampling distributions.

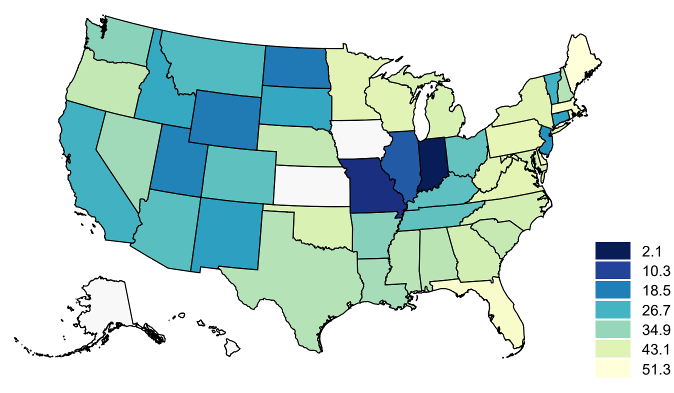

Fig. 4. Choropleth Map for visualizing average soil organic carbon (SOC) value at the "Oi" soil horizon. The base map helps users look for a state of interest quickly.

C. The quantitative component

The qualitative component serves the purpose of giving an overall view of the distribution of the analyzing quantitative attributes well. Furthermore, these quantitative views are also easy to look up due to the users' familiarity with the base map. However, they suffer from one issue. That issue is that using the color and size channels to encode quantitative attributes is less effective compared to the length channel [13]. Therefore, when users would like to analyze these quantitative attributes quantitatively, the quantitative component is necessary (anal-

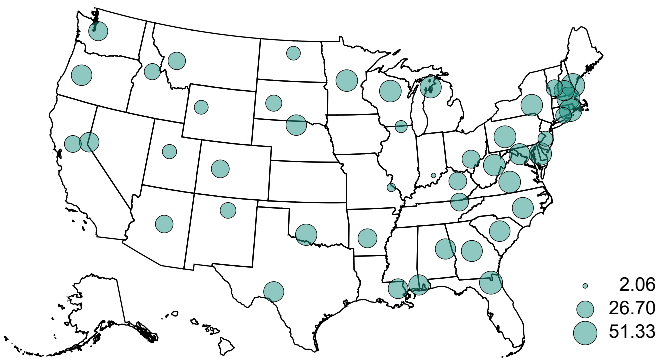

Fig. 5. Symbol Map for visualizing average soil organic carbon (SOC) value at the “Oi” soil horizon. It is still easy to look up, thanks to the base map. Additionally, the size channel is more efficient than the color channel in representing quantitative data (SOC value).

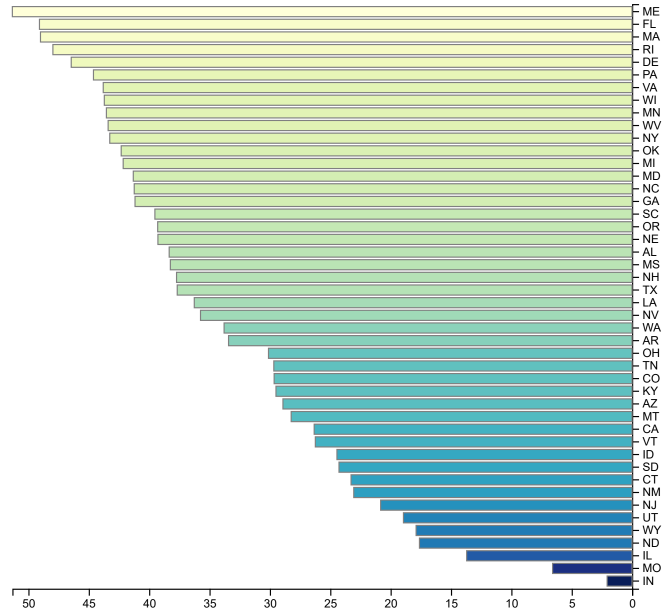

Fig. 6. Bar chart for visualizing average soil organic carbon (SOC) value at the “Oi” soil horizon. The color channel is a redundant (doubled-encoding) channel because it is the same as in the Choropleth Map. The redundant encoding assists in referencing the bars back to the state regions in the base map from the qualitative view. Lengths of the bars help differentiate the SOC values between states efficiently with the trade-off of losing the spatial relationship.

ysis task **T2**).

Views in this component use bar charts to encode quantitative attributes. Specifically, the marks are the bars for the regions (states in this case). The lengths of the bars encode the analyzing quantitative attributes (i.e., SOC, tree height, and tree DBH). According to Steven’s psychological power law [15], the length channel has a linear relationship between its change in intensity and the change in humans’ perceived sensation. Thus, the length channel is more effective than the size and color channels.

Thanks to the effectiveness of the channels used, these bar charts help the users look up and compare quantitative differences more accurately. Furthermore, while analyzing these quantitative attributes (i.e., SOC, tree height, and tree DBH), users are interested in particular values, such as the regions with the highest or the smallest measurements. Therefore, these views sort the bars in descending order of lengths from top to bottom to support lookup and comparison.

Another critical task is to analyze the correlation between the SOC values over the soil horizons or locations and the tree heights or tree DBH values (analysis task **T3**). A typical visualization idiom for analyzing the correlation between a pair of variables is to use a scatterplot. The scatterplot requires a rectilinear layout with two axes orthogonal to one another. One axis is for the SOC values for each state, and another is for each state’s average tree heights or DBHs.

In this proposed solution, the y-axis (SOC values) is available on the bar chart, as shown in Figure 2 bottom left corner, with the SOC values for the states ordered in descending order. Similarly, the tree heights or DBHs for the states are also available on the bar chart shown in Figure 2 bottom right corner. Therefore, it makes perfect sense to adapt parallel coordinate view [16] using these two available axes to depict the correlation between SOC values and tree heights or DBHs in this case. Therefore, we do not need to create a new axis as if a scatterplot is used and it connects different views of this proposed solution into one integrated system with interconnected views.

Parallel coordinate visual idiom, as the name indicates, organizes its axes in parallel to one another instead of being orthogonal to one another as in other common visual idioms with rectilinear layouts. This visualization idiom uses line marks to represent the correlation between a pair of variables. The start of a line segment represents the value of the first variable, and the end of this segment encodes the value of the second variable. It uses the tilt/angle channel to represent the relative change or difference between two values of the corresponding two variables.

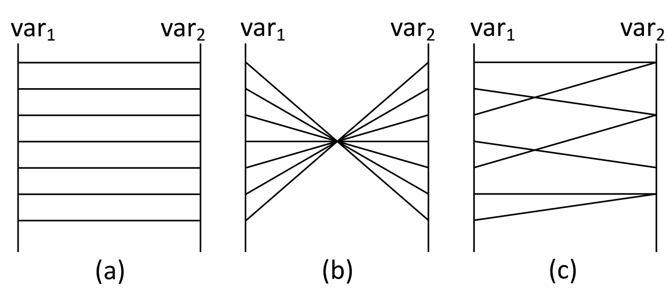

Fig. 7. Typical patterns of correlations between a pair of variables in parallel coordinate: (a) two variables are perfectly positively correlated, (b) two variables are perfectly negatively correlated, (c) two variables are not correlated.

Similar to the scatterplot, the parallel coordinate visual idiom also provides a good overview and ways to explore the correlations of the measured attributes. However, the parallel coordinate visual idiom is relatively difficult to read and interpret for beginners. Thus, it requires a little training time for users to read and interpret the encoded patterns. Figure 7 shows three typical patterns inside every pair of variables (var_1 and var_2, in this case). The first one (a) means there is a perfect correlation between these two variables because high values of var_1 correspond with high values of var_2 and vice versa. Conversely, in the second case (b), these two variables are perfectly negatively correlated. Specifically, high values of one variable correspond to low values of another variable and vice versa. Concomitantly, the third pattern (c) indicates that there is no significant correlation between these two variables.

Figure 8 depicts an example of using parallel coordinate visualization to depict the correlation between SOC values at the 'Oi' horizons all over all U.S. states (variable 1, left axis) versus corresponding DBH values (variable 2, right axis). Notably, these two axes are reused from the y-axes of the two bar charts for these two variables, respectively, as shown in Figure 2. Additionally, to precisely quantify the correlation, this view computes the Pearson correlation coefficient [17] of these two variables and displays it at the top of the view.

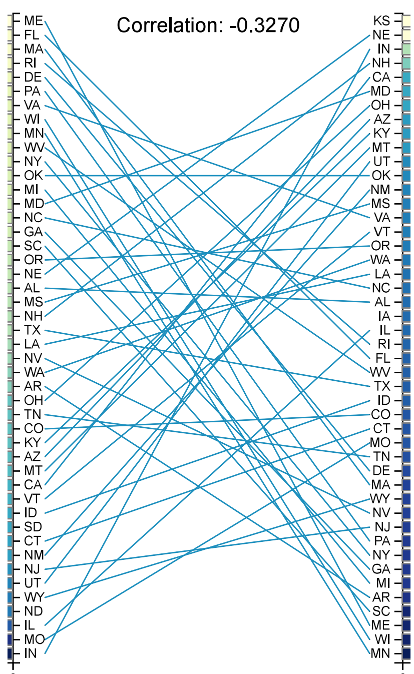

Fig. 8. Parallel coordinate encodes the correlation between Soil Organic Carbon values over 'Oi' soil horizons of the U.S. states (left, variable 1) versus their corresponding diameter at breast height values (right, variable 2). The Pearson correlation coefficient is also computed and displayed at the top to quantify the correlation between two variables.

D. The interactions

Computer visualization design has limitations to consider. These limitations include computation resources (e.g., processing a large amount of data at a time), human cognitive abilities, and display resolutions (screen sizes). For instance, we are dealing with a huge amount of data in this case, and it is impossible to display all types of data at once on the screen. Furthermore, computer screens have limited resolutions, and thus we cannot display all the data simultaneously. Even if we could display all the types of data at once with different visualizations on the screen, it would not be easy for humans to consume the resulting visualization because the resulting visualization would overwhelm users' cognitive loads.

Therefore, in this case, interactions come in handy. It helps to filter out a single type of data at a time to display. For instance, there are options for users to select SOC values over a specific soil horizon (out of 14 horizons) or the overall averages per state or per site (as shown on the left menu of Figure 9). Then there are also options for users to choose to compare the SOC value type to either tree heights or DBHs (as shown on the right menu of Figure 9).

Additionally, as discussed, different views support different analysis tasks. Therefore, there are also options for the users to select appropriate map types such as Choropleth Map, Point Map, and Symbol Map (as shown on the right menu of Figure 9). Furthermore, when users would like to focus on information about a specific state, users can mouse over a state, a state abbreviation, or a bar. The system then displays the related data for that state and also fade-out the displays of other states to let the user focus on analyzing a state of interest. Figure 9 shows how the system highlights the data of interest for Texas when the user moves the mouse over Texas state.

V. USE-CASES AND DISCUSSIONS

The design is based on the analysis of the appropriateness of visual encoding and interaction idioms of each visualization and interaction used for the targeted tasks. However, we also consider the validation of the design decisions at the data and task abstraction level and domain situation level [13]. Specifically, to show the usefulness of RaCAViz, we also present a typical use case of this interactive visualization in solving the three analysis tasks on the RaCA dataset as specified in Section III-B. Notably, this use case reuses the presented visualizations (figures) whenever possible for the space efficiency of this paper. However, other use cases can be performed with similar steps specified in this specific use case.

When the user first lands on the project's home page, the page displays a default view, as shown in Figure 2. The default map view type is Choropleth Map because it is aesthetic and gives users a quick overview of the attributes of interest over CONUS. Furthermore, Choropleth allows users to quickly look up the value of a specific state of interest, thanks to their familiarity with the map.

Next, the user wants to see the overall distribution of all the sampled locations to know where on CONUS these samples were taken. The user can accomplish this task using RaCAViz by selecting the data category for the left map view as 'All' (abbreviated for all locations) and the map view type as Point Map (from the radio buttons as shown on the top-left corner of Figure 2). The resulting map view is shown in Figure 3. It helps the user validate the overall sampling strategy in this RaCA project and have a sense of the measured

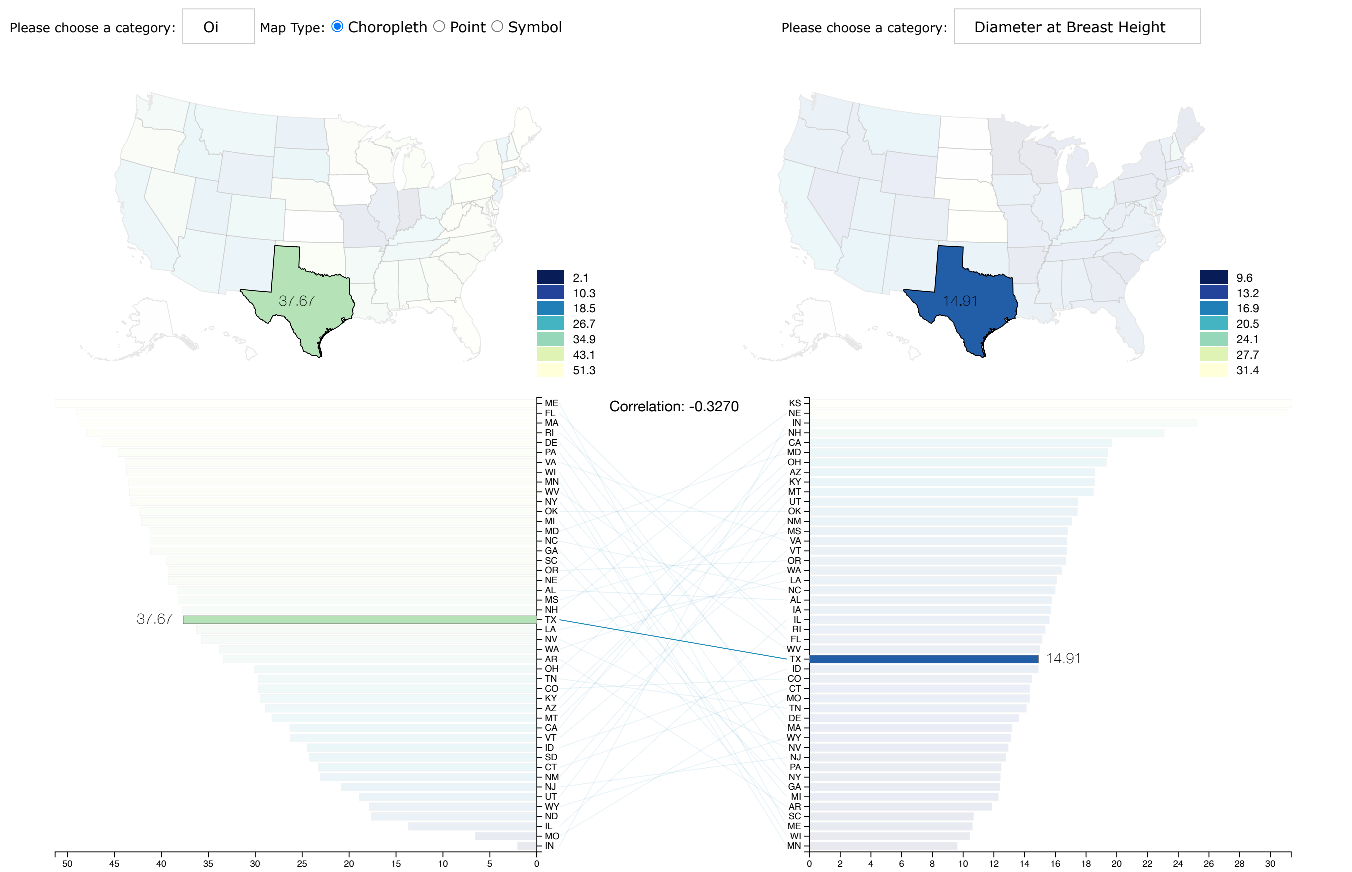

Fig. 9. Examples of RaCAViz's interactive options: (1) there are drop-down lists to select appropriate types of data to display, such as soil organic carbon values per horizon (left) versus tree heights or diameters at breast height (right); (2) there are options to change the map types to Choropleth Map, Point Map, or Symbol Map; (3) when users mouse over a state, a bar, a line, or a state name, information about this state is displayed, and other states are faded-out to let users focus on information of the current state (Texas in this specific case).

attributes (SOC values per horizon per location, tree heights, or tree DBHs). Users can also mouse over individual sampled locations or sites, and the system will display the measured values of the location/site at the mouse pointer.

After validating the sampled locations, the user would like to focus on analyzing SOC values at a specific soil horizon named 'Oi.' Thus, the user selects the 'Oi' soil horizon from the top left menu. The current map view type (Point Map) seems overwhelming for the user. Thus, the user switches the map type view to Symbol Map to have a clearer picture of how the SOC storage is distributed over all the CONUS states, as shown in Figure 3.

Once knowing how the SOC measured values are distributed over CONUS, the user wants to analyze how SOC values correlate with tree information, tree DBH in this specific case. Thus, the user uses the drop-down menu on the top right corner and selects "Diameter at Breast Height" for the map view on the right. The resulting visualization is shown in Figure 2. Using the two maps on top, the user can qualitatively compare the correlation between SOC values at the 'Oi' soil horizon and tree DBHs. However, the user also wants to analyze this case quantitatively. Specifically, the user wants to know which states have higher SOC or DBH values or to compare these values quantitatively over different states. In this case, the user attends to the two bar charts provided, such as those at the bottom of Figure 2.

Furthermore, the user also looks at the pattern displayed by the parallel coordinate view at the bottom of Figure 2 or a closed-up version of it in Figure 8. This view indicates a negative correlation between SOC stored at this soil horizon ('Oi') and tree DBHs. Quantitatively, the user can also view the Pearson correlation coefficient as -0.3270 at the top of the parallel coordinate view to confirm this negative correlation. Notably, other works performed in different regions also demonstrate this negative correlation. For instance, Saimun at al [18] reported that soil carbon was negatively correlated with tree richness, height, and basal area.

Lastly, the user would like to focus on analyzing information from a specific state (e.g., Texas, in this case). Therefore, the user moves the mouse over this state. As shown in Figure 9. The system then fade-outs all the visual marks that encode information from other states to highlight Texas-related information. Furthermore, specific measured SOC values and tree DBH values for Texas are laid over the maps and the bars to provide details about these measured attributes for Texas.

This use case is one typical one for this proposed system. However, users can use the provided data category, map view type, and interactions to perform the other typical analysis tasks on this dataset. Observably, this project is not as complicated as CarbonScape. However, it is lightweight, focused, and easy to use. Thus, it is more suitable for the general audience. Therefore, it should be a good candidate as an educational resource for K-12/STEM classrooms or scientific communities about terrestrial carbon landscapes.

In the future, we would like to collaborate with USDA and the RaCA project's personnel to validate the usefulness of this project. Furthermore, we could elicit further requirements, complete the prototype and incorporate this project into the USDA RaCA project's page. This direction should help bring the collected data to a broader audience.

VI. Implementation

The data collection and data processing component is implemented in Python. However, for portability, ease of use, and multiple platform compatibility, RaCAViz is implemented as JavaScript-based web application using D3.js [19] visualization library. The source codes and the web prototype of this proposed interactive visualization tool are available on the project's GitHub page https://github.com/mdptlab/racaviz.

VII. Conclusions

This paper presents an interactive visualization solution called RaCAViz for rapid assessment of carbon storage in soil and its relationship with tree richness. The solution includes three components: the data querying and processing component, the qualitative analysis component, and the quantitative analysis component. The quantitative and qualitative views are incorporated with interactions to support exploring and viewing the datasets from different perspectives. Different from other existing solutions, which are either merely static graphics to report initial insights to the public or too complicated to be helpful for a broad audience. RaCAViz focuses on main analysis tasks and provides interactive options built on the web platform to help bring insights about this type of dataset to a broad audience.

RaCAViz is designed using an analytical approach by evaluation of visualization and interaction idioms for three focused analysis tasks: (1) Qualitative Overview, (2) Quantitative Analysis, and (3) Correlation Analysis. It is then applied to a soil organic carbon storage dataset with corresponding tree heights and tree diameters at breast height over the conterminous United States as a typical use case to validate its usefulness. However, RaCAViz can be used to analyze any similar dataset collected over other regions. The future direction of this work should be collaborating with the United States Department of Agriculture (USDA), Natural Resources Conservation Service (NRCS) personnel to validate the usability of RaCAViz, followed by incorporating it as a built-in interactive visualization solution for the USDA, NRCS Rapid Carbon Assessment (RaCA) project. This step should help bring insights about the RaCA dataset to a broad audience.

References

[1] A. S. Devi, "Influence of trees and associated variables on soil organic carbon: a review," *Journal of Ecology and Environment*, vol. 45, no. 1, pp. 1–14, 2021.

[2] A. B. McBratney, U. Stockmann, D. A. Angers, B. Minasny, and D. J. Field, "Challenges for soil organic carbon research," in *Soil carbon*. Springer, 2014, pp. 3–16.

[3] Soil Survey Staff, "Rapid Carbon Assessment (RaCA) project," June 1, 2013 (FY2013 official release), Available online.

[4] V. Pham, C. M. Jordan, M. G. Siebecker, D. C. Weindorf, and T. Dang, "iDVS: interactive 2D and 3D visualizations of proximal sensor data for rapid characterization of soil profiles," *Precision Agriculture*, pp. 1–20, 2022.

[5] V. Pham and T. Dang, "Soaviz: visualization for portable x-ray fluorescence soil profiles," in *7th Workshop on Visualisation in Environmental Sciences, EnvirVis 2019 at EuroVis 2019*. The Eurographics Association, 2019, pp. 33–40.

[6] V. Pham, D. Weindorf, and T. Dang, "Soilscanner: 3d visualization for soil profiling using portable x-ray fluorescence," in *8th Workshop on Visualisation in Environmental Sciences, EnvirVis 2020 at Eurographics/EuroVis 2020*. The Eurographics Association, 2020, pp. 17–24.

[7] F. Sun, N. Bakr, T. Dang, V. Pham, D. C. Weindorf, Z. Jiang, H. Li, and Q.-B. Wang, "Enhanced soil profile visualization using portable x-ray fluorescence (pxrf) spectrometry," *Geoderma*, vol. 358, p. 113997, 2020.

[8] V. Pham, D. C. Weindorf, and T. Dang, "Soil profile analysis using interactive visualizations, machine learning, and deep learning," *Computers and Electronics in Agriculture*, vol. 191, p. 106539, 2021.

[9] S. Wills, T. Loecke, C. Sequeira, G. Teachman, S. Grunwald, and L. T. West, "Overview of the us rapid carbon assessment project: sampling design, initial summary and uncertainty estimates," in *Soil carbon*. Springer, 2014, pp. 95–104.

[10] S. Waltman, "USDA CarbonScapes," Available online: http://ftp.wvgis.wvu.edu/pub/temp/USDA/2015-01_CarbonScapes_USDA_GCTF_01062015_Report.pdf; Accessed date: August 18, 2022.

[11] C. Brack, "Standard point on tree bole for measurement," *Forest measurement and modelling, Computer-based course resources for Forest Measurement and Modeling (FSTY2009) at the Australian National University, Canberra, Australia*, 2009.

[12] T. R. Feldpausch, L. Banin, O. L. Phillips, T. R. Baker, S. L. Lewis, C. A. Quesada, K. Affum-Baffoe, E. J. Arets, N. J. Berry, M. Bird *et al.*, "Height-diameter allometry of tropical forest trees," *Biogeosciences*, vol. 8, no. 5, pp. 1081–1106, 2011.

[13] T. Munzner, *Visualization analysis and design*. CRC press, 2014.

[14] P. Cybulski, "Spatial distance and cartographic background complexity in graduated point symbol map-reading task," *Cartography and Geographic Information Science*, vol. 47, no. 3, pp. 244–260, 2020.

[15] S. S. Stevens, "On the psychophysical law." *Psychological review*, vol. 64, no. 3, p. 153, 1957.

[16] A. Inselberg and B. Dimsdale, "Parallel coordinates: A tool for visualizing multivariate relations (chapter 9)," 1991.

[17] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*. Springer, 2009, pp. 1–4.

[18] M. S. R. Saimun, M. R. Karim, F. Sultana, and M. A. Arfin-Khan, "Multiple drivers of tree and soil carbon stock in the tropical forest ecosystems of bangladesh," *Trees, Forests and People*, vol. 5, p. 100108, 2021.

[19] M. Bostock, V. Ogievetsky, and J. Heer, "D^3 data-driven documents," *IEEE transactions on visualization and computer graphics*, vol. 17, no. 12, pp. 2301–2309, 2011.