

SET-STAT-MAP: Extending Parallel Sets for Visualizing Mixed Data

Shisong Wang* Debajyoti Mondal† Sara Sadri‡ Chanchal K. Roy§ James S. Famiglietti¶
Kevin A. Schneider||

University of Saskatchewan, Canada

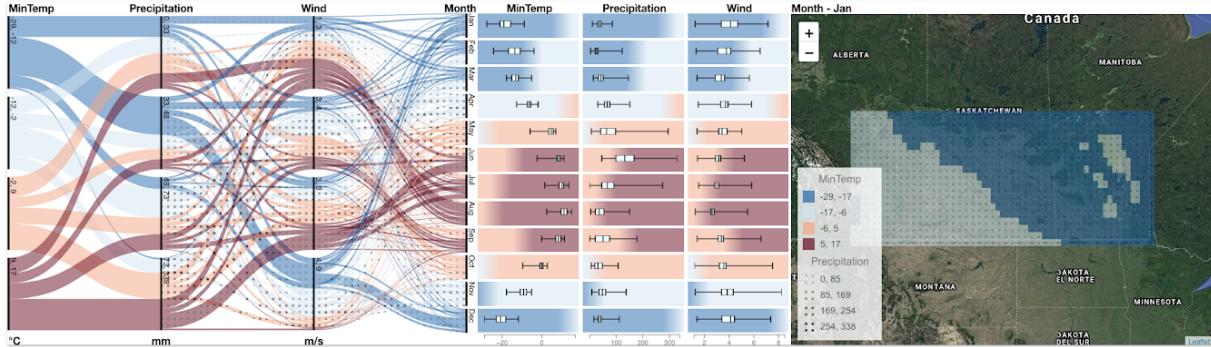


Figure 1: A SET-STAT-MAP visualization for a meteorological dataset over Canadian Prairies.

ABSTRACT

Multi-attribute dataset visualizations are often designed based on attribute types, i.e., whether the attributes are categorical or numerical. Parallel Sets and Parallel Coordinates are two well-known techniques to visualize categorical and numerical data, respectively. A common strategy to visualize mixed data is to use multiple information linked view, e.g., Parallel Coordinates are often augmented with maps to explore spatial data with numeric attributes. In this paper, we design visualizations for mixed data, where the dataset may include numerical, categorical, and spatial attributes. The proposed solution SET-STAT-MAP is a harmonious combination of three interactive components: Parallel Sets (visualizes sets determined by the combination of categories or numeric ranges), statistics columns (visualizes numerical summaries of the sets), and a geospatial map view (visualizes the spatial information). We augment these components with colors and textures to enhance users’ capability of analyzing distributions of pairs of attribute combinations. To improve scalability, we merge the sets to limit the number of possible combinations to be rendered on the display. We demonstrate the use of SET-STAT-MAP using two different types of datasets: a meteorological dataset and an online vacation rental dataset (Airbnb). To examine the potential of the system, we collaborated with the meteorologists, which revealed both challenges and opportunities for SET-STAT-MAP to be used for real-life visual analytics.

Index Terms: Human-centered computing—Visualization—Visualization techniques; Information Interfaces and Presentation—Miscellaneous

1 INTRODUCTION

Multi-attribute datasets are common in real-life data analytics scenarios and also challenging to analyze. Such data are often explored visually due to humans’ ability to rapidly comprehend and detect patterns in visual scenes. The preferred choice of visualization varies depending on the data attributes and tasks. For example, to visualize the pairwise relationships among categorical dimensions a simple way is to use a frequency matrix, where each entry contains a value representing the number of data records in the row and column category. Here an attribute may only have a few discrete values or categories. For numerical dimensions, which is either continuous or have many discrete values over a range, one can use a scatter plot matrix, where each cell of the matrix contains a scatterplot for corresponding row and column attribute. Some more recent approaches to examine multi-attribute datasets include Parallel Coordinates [25] and Parallel Sets [27] visualizations. These visualizations cleverly utilize the display space to show individual attribute information in a compact way, as well as provide interactivity to explore attribute relationship patterns.

1.1 Parallel Coordinate

A Parallel Coordinates visualization represents each attribute using a vertical axis and each data record as an x -monotone polygonal line or curve intersecting these axes according to the values present in the record (Figure 2(a)). For a numeric dataset, a Parallel Coordinates visualization may provide detailed information about individual attribute as well as their pairwise relationships. For example, one can examine the value distribution of an attribute by examining the intersection points along the vertical axis. The axes are usually reorderable and one can examine potential relationship among two adjacent attributes by examining the visual pattern appearing between the corresponding vertical axes.

Dataset with categorical attributes poses a problem to Parallel Coordinates since the lines can overlap each other. For example, consider a dataset with three attributes (X, Y, Z), each consisting of two categories. A Parallel Coordinates visualization of such a dataset will create at most $2 \times 2 \times 2 = 8$ distinct polygonal lines (Figure 2(b)), even when the dataset contains hundreds of records.

*e-mail: sean.w@usask.ca

†e-mail: d.mondal@usask.ca

‡e-mail: sara.sadri@usask.ca

§e-mail: chanchal.roy@usask.ca

¶e-mail: jay.famiglietti@usask.ca

||e-mail: kevin.schneider@usask.ca

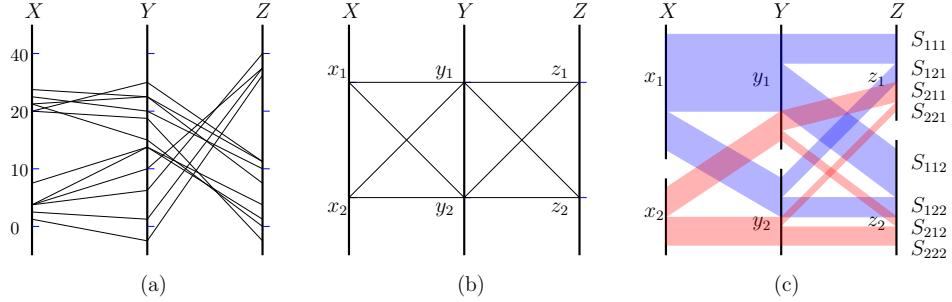


Figure 2: Parallel Coordinates visualizations with (a) numerical attributes and (b) categorical attributes. (c) A Parallel Sets visualization.

1.2 Parallel Sets

The Parallel Sets visualization is especially designed for categorical data. Instead of showing individual data points, a Parallel Sets visualization reveals data frequencies by drawing parallelograms between pairs of categories (Figure 2(c)). The width of the parallelogram is relative to the number of data records that share those categories. This idea gives rise to the notion of sets. For example, let (x_1, x_2) , (y_1, y_2) and (z_1, z_2) be the categories for the attributes X, Y, Z we considered earlier. Then for each i, j, k , where $1 \leq i, j, k \leq 2$, a Parallel Sets visualization forms a set S_{ijk} by collecting the data records that belong to all three categories x_i, y_j and z_k . Assume that the axis X, Y, Z are horizontally aligned. Then the set S_{ijk} is represented using two parallelograms: one corresponding to S_{ij} and the other to S_{jk} . Here S_{ij} is a predecessor of S_{jk} . We refer to such a sequence of contiguous parallelograms as a *ribbon*. Based on the context, sometimes a pair of ribbons may be interpreted as one ribbon being partitioned into two ribbons. For example, in Figure 2(c), the two ribbons corresponding to S_{221} and S_{222} may be interpreted as the ribbon $S_{x_2y_2}$ being partitioned into ribbons $S_{y_2z_1}$ and $S_{y_2z_2}$.

If the attributes are numeric, then a Parallel Sets visualization can be designed by transforming each numeric attribute into a categorical attribute, where the categories are defined as value ranges [27]. For example, the numeric attribute body weight can be transformed into categories: Underweight, Healthy weight, Overweight and Obese.

To explore mixed data, which may contain numerical, categorical and even spatiotemporal attributes, a common approach is to use information linked multiple views, where different visualizations of a dataset are visualized across different views simultaneously and interactions (selection or filtering) in one view is reflected across multiple views. However, an ill-designed visualization may significantly increase users' cognitive load, and mislead users showing ambiguous or unclear information.

1.3 Background and Motivation

In an Open Science Meeting of the Global Water Futures, we were intrigued by observing how scientists analyze various environmental variables through side-by-side contour-plots of the map to understand their yearly pattern and spatial correlations. The number of combinations that needs to be analyzed can become very large fairly easily, e.g., consider analyzing pairwise relations among different variables in different seasons in different spatial locations. This restricts the ability of researchers to gain quick insights into the data. Such a large number of combinations does not readily allow an efficient way to visualize all the information in multiple linked views. For a specific example, consider a hydrologist examining an evapotranspiration process (evaporation from water and soil and transpiration from plants), which is influenced by a combination of solar radiation, air temperature, wind speed and humidity. The hydrologist also needs to examine the annual trends of these weather parameters across different geographic landscapes: mountainous, forest, urban, and rural areas.

In a discussion with the researchers, we realized that for a scientific analysis one often needs to go through various data processing and visualization steps, e.g., the capability of selecting variables, dissecting datasets, filtering based on temporal ranges, visualizing statistics such as quantiles or boxplots, and plotting data on a map, etc. This motivated us to investigate different ways to design multiple-linked views that would help gain quick insights into such mixed data.

To explore a meteorological dataset, a widespread approach is to visualize a topographic map where a selection on the map generates a Parallel Coordinates visualization of the selected data in a secondary view. While investigating whether such a Parallel Set and map combination could be adopted to aid in visualizing such mixed data, we realized that a Parallel Sets visualization are more suitable in this context as it allows users to select various sets or combinations on the fly, and creates an opportunity to augment statistical plots on rows when the months are placed as the last vertical axis (e.g., see Figure 1). Although the discovery of SET-STAT-MAP was inspired by the analytics of meteorological dataset, we soon realized that similar scenarios may appear in many other real-life context. For example, consider a scenario where a policymaker is exploring a vacation rental database to examine the distributions of room price (numeric) and room type (categorical) across different neighbourhoods (spatial), and understand potential interactions among these attributes [20].

1.4 Contribution

We propose SET-STAT-MAP that extends Parallel Sets to visualize mixed data, i.e., the data points may contain categorical attributes, numerical attributes and possibly spatial information.

SET-STAT-MAP combines the idea of Parallel Sets, statistical columns and multiple coordinated view techniques. The Parallel Sets view displays the data records by grouping them into sets or ribbons, which provides information about the relationships among data attributes and the number of data records in each set. We design the Parallel Sets visualization to link the sets determined by the last axis with the rows of the statistics columns. Each row of the statistics columns provide statistical summaries of the numeric attributes for each category of the last axis of the Parallel Set. By selecting a row of the statistics columns, one can examine the data records on the map view. The interactive features (set selection, transformation of numeric attributes into categorical ranges, axis reordering, color and texture selection, filtering based on rectangle selection on maps) allow users to explore the dataset by generating visual configurations that best suit their needs.

We propose two enhancements for the Parallel Sets view. The original Parallel Sets visualization only uses colors for ribbons to help users quickly distinguish categories in the first axis. Our first enhancement adds textures relating to the second-axis categories to help users further analyse a pair of attribute combinations. Our second enhancement to the Parallel Sets view is merging unim-

portant sets into one in order to limit the number of possible set combinations. We also show how the selection of a category on an axis can provide users with an interesting spatial distribution of the corresponding data records on the map.

We demonstrated the usability of SET-STAT-MAP with use cases from two different domains. We show both the strengths and weaknesses of SET-STAT-MAP when visualizing mixed data and provide guidelines to effectively use such a visualization.

2 RELATED WORK

In this section we discuss the related research under four themes — axis based, set based, table based, and map based techniques.

2.1 Axis Based Techniques

Coordinate systems have been widely used to visualize two and three dimensional data for years, which proved their value of visualizing accurate details of such data. In general, for k attributes, one can construct $\binom{k}{2}$ scatterplots to examine pairwise relations. Radar chart [31] extends the idea of coordinate systems, where each axis of which is put in a circular manner sharing the same origin point to overcome the limitation of the standard 2D or 3D Cartesian coordinate systems. Each data record is represented as a polygon with vertices lying on various axes depending on the record's value at that dimension. Overlaps of the polygons can show the similarities between different data records. In a Parallel Coordinates visualization [25], axes are placed parallel to each other, where each data record is presented as a polygonal line. Compared to radar charts, a Parallel Coordinates visualization better utilizes the space use, i.e., it provides uniform space (instead of a cone shaped region) between a pair of adjacent axes, and the line patterns between adjacent axes are easily comparable since they are parallel to each other (instead of being arbitrarily rotated). Line overplotting (when the number of data records is large) and axis ordering (choosing a linear ordering that represents most interesting pairwise relations) are two common problems in a Parallel Coordinates visualization. A number of research works attempted to leverage Parallel Coordinate to effectively visualize categorical attributes. We refer interested readers to [24] for a survey on various extensions of Parallel Coordinates visualization.

2.2 Set Based Techniques

Set based techniques for visualizing multi-attribute data construct sets based on various-category combinations. Parallel Sets is an interesting example, where it is possible to define a set for any category combination that contains one category from each attribute. Since Parallel Sets visualization use parallelograms to connect adjacent axes, the visualization quickly becomes occluded as the number of categories grow. Therefore, Parallel Sets are most suited for categorical attributes with only a few categories.

Spatial data are often visualized on maps. Bubble set [15] is an example that visualizes contours (implicit surface) for every category (set) that encloses all its elements and excludes all other elements if possible. Thus multiple category combinations can be visualized directly on a map showing how the elements of each category are spatially distributed. However, such an approach is not as suitable as Parallel Sets in revealing frequency information and relationships among two categorical attributes. A region based set visualization approach [9, 10] may create a cluttered visualization only for a few categories for irregular and meandering regions. MetroSets [26] visualizes set systems using the metro map metaphor.

2.3 Table Based Techniques

A spreadsheet provides a tabular visualization for multi-attribute dataset, where the rows correspond to the data records, columns correspond to the attributes, and each table entry is either a value or a category of its column attribute. If the precise numerical values are not important, then replacing the numerical values using horizontal

bars and sorting the rows by the attribute of interest gives a quick overview of the value distribution. However, sometimes tabular visualizations appear to be sparse and therefore, are not scalable with increasing data size. LineUp [19] takes a step further and allows users to examine the ranks of data records, where a rank of a record is determined by a weighted sum of its attribute values. To improve the scalability of the visualization, LineUp allows users to collapse columns, shows only the top 100 rows at a time whereas the rest of the rows are shown in a compressed manner which can be explored using fish-eye selection [34].

UpSet [28] is a set visualization approach that can be used to examine complex multi-attribute datasets such as movie datasets. The user can construct a query using a visual query builder, e.g., visualize movies that must include the category adventure, but not comedy. UpSet first constructs all possible category combinations (sets) that satisfy the query and then uses those as rows. The statistical summaries (e.g., average rating, release date) of each set is then visualized using a separate statistics column. UpSet supports multiple views, which are collapsible. For example, the users can choose to visualize query records in a scatter plot or simply as a data table. A number of matrix and network based visualizations [11] have been explored to further improve the scalability of set visualizations.

2.4 Map Based Techniques

Choropleth maps and proportional symbol maps [17] are commonly used methods to visualize spatial information on a map. Although they are well-suited for univariate data, color blending and texture patterns (for choropleth), and glyphs (for symbols) can be leveraged to extend these techniques for bivariate visualization. Color blending is widely used in spatial analysis to find probable correlations between two geospatial variables [18] and researchers have examined different ways to construct trivariate choropleth maps [35, 36] and even 4-variate [39] maps. Overlaying glyphs (icons, pie charts, and even complex textures) on a map is a popular way to visualize multi-attribute geospatial information [14]. Typically, glyphs are designed to encode data into features that can be perceived through preattentive visual channels [38]. Pixels, introduced by Healey and Enns [22], encode multi-attribute datasets into perceptual textures with height, density, and regularity properties. In general, texture of varied contrast, density, and directionality [29, 37] are often used to visualize multivariate data. For more details on multivariate spatial data visualization, we refer the reader to [8, 21, 40].

The use of multiple coordinated views is becoming increasingly common in heterogeneous data analytics, e.g., environmental data exploration [7, 13] and multivariate network visualization [30]. One of the most common examples is a geographic map coordinated with charts or Parallel Coordinates view, where users can select a region of a map and examine the updated charts for the selected dataset. Since multiple coordinated views can visualize different aspects of the same dataset simultaneously, users can better comprehend the data. Such visualizations often lack methodological design efforts and the final visualization becomes a collection of charts, plots, tables and maps that do not look like an integral part of the overall visualization and may immensely increase users' cognitive load. We refer the readers to [12, 32] for a rich body of research that examines design principles and usability issues of multiple coordinated views.

3 DESIGN OF SET-STAT-MAP

The SET-STAT-MAP visualization is partitioned into three views: (left) Parallel Sets, (middle) statistics columns and (right) map view. We first describe the details of each view and then the coordination among these views.

3.1 Parallel Sets View

The visualization constructs a Parallel Sets view with vertical axes. The first, second and the last axes (when ordered from left to right)

are of most importance in our visualization.

Transforming numeric attributes to categorical attributes: The original Parallel Sets [27] showed how value ranges can be used to adopt numeric attributes in a Parallel Sets visualization by allowing users to specify value ranges of interest. We allow two options to automatically create such value ranges: one is partitioning the range of the attribute into some equal parts, and the other is to partition the range using quartiles. Both these two options (Figure 3) can be useful in practice. For example, quantile ranges of a variable are often used to examine regional extreme, and uniform variable ranges are often used to understand amount of data points at each interval.

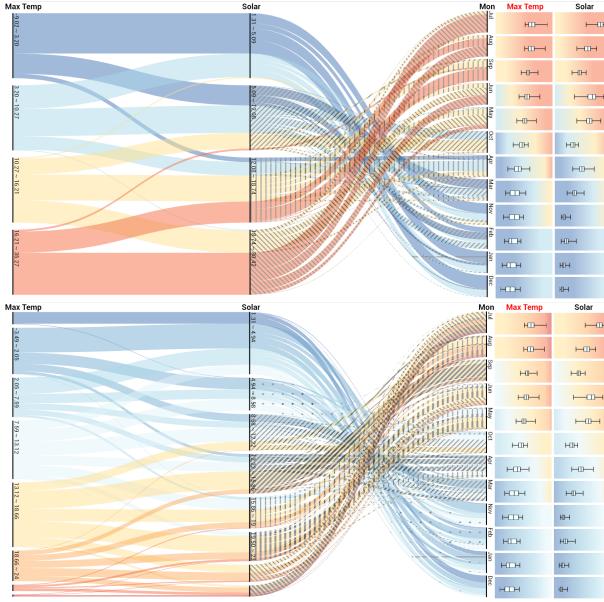


Figure 3: Users can explore two SET-STAT-MAP design by placing them one top of each other. For example, the top view shows an attribute transformation with quartile ranges, and the bottom view uses eight equal bins.

Limiting the number of categories: A large number of categories for each attribute may create a large number of ribbons in the original Parallel Sets visualization [27], which dramatically reduces the width of the ribbons. It not only consumes expensive resources for rendering, but also makes some ribbons hard or even impossible to see. To improve this situation, we apply a configurable ‘maximum category’ limitation for each axis and merge the remaining categories. This greatly reduces the number of ribbons that need to be rendered. Figure 4 illustrates such an example.

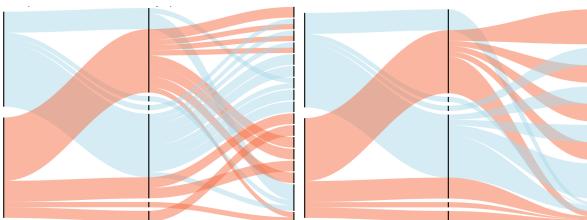


Figure 4: Parallel Sets (left) without any merged set and (right) with merged set. The number of categories at the last axis is limited to six, where the bottommost category is the merged one.

In a default setting, we limit the number of categories per attribute to be displayed to a small number $\alpha \leq 8$. The rest of the categories

are merged together and labelled as ‘other’. We set a small fixed width for the merged section, which improves the available space for the categories that have not been merged. Since not all the categories are now visible, an obvious remedy is to filter the categories based on users’ needs, however, it is also possible to scroll at an axis to add more categories from the bottom end (‘other’ section) whereas the categories from the top join the ‘other’ section.

In a worst case scenario, this may require rendering α^k ribbons, where k is the number of attributes. Hence the performance (rendering speed and interactivity) may decrease for five or more attributes depending on the data. If more attributes are needed to be displayed, one can define a smaller value for α (even can customize the parameter for each axis).

Color encoding for the ribbons: We color a ribbon to represent its category on the first axis, which follows original Parallel Sets design [27]. More specifically, we assign each category on the first axis using a distinct color. One can choose a color map based on the categories (qualitative color map for discrete options, or diverging color maps for numeric ranges). Let σ be the ordering of the colors from top to bottom. We then color each ribbon based on which category it belongs to on the first axis. For a category (for any attribute), there exists a vertical line interval intersected by a number of ribbons. The ribbons are ordered based on σ such that while examining such an interval, users can quickly recall where these ribbons are located at the first axis.

Texture encoding for the ribbons: We augment the original Parallel Sets [27] by adding texture to the ribbons. The textures are rendered on the ribbons from the second axis. We assign each category on the second axis using a distinct texture pattern. We then render a texture on each ribbon based on the category it belongs to on the second axis.

We allowed different texture patterns: one is a dotted pattern with different opacity (Figure 1), a second one with oriented lines (Figure 5), and we also examined arbitrarily chosen texture (Figure 6(left)). The first one is intuitive to understand, but does not work well if the category contains more than 4 attributes. The rationale behind allowing the second one (line-orientation pattern) is that for numeric ranges, the higher value ranges will naturally correspond to a larger angle of inclination of the lines with the x -axis. If the number of categories is at most four, then we use the first, third, fifth and seventh patterns. Even when the categories correspond to some discrete non-numeric options, this natural texture ordering is likely to allow users to easily recall their top to bottom ordering on the second axis. Figure 6 compares our texture pattern (right) with a texture pattern (left), which was chosen arbitrarily.

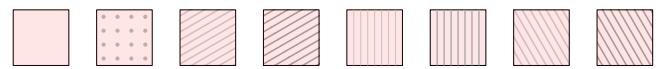


Figure 5: Illustration for the texture mapping for increasing data values (from left to right).

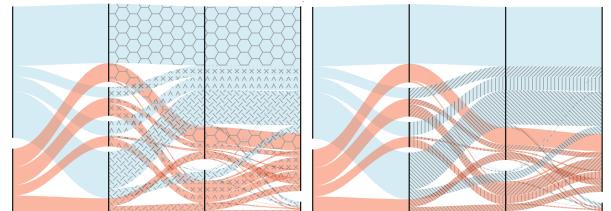


Figure 6: Parallel Sets (left) with arbitrary textures, and (right) with line-orientation based textures.

3.2 Statistics Columns View

The statistics column view consists of a set of numerical attributes. Each attribute is visualized in a column, where the rows are determined by the last axis of the Parallel Sets design (e.g., the months in Figure 1 create 12 rows in the statistics columns view). The j th row in a column (i.e., a cell) can be a box plot or a bar representing the summary statistic of the j th category of the last axis of the Parallel Sets. The box plot gives information about the quartiles and data range, and the bar plot shows the average value of the set (determined by the j th category) for the corresponding attribute. The background of a cell is filled using a linear color gradient using the colors present in the j th category. Let I_j be the vertical interval corresponding to the j th category. Then the width of a color range in the j th row is determined by the proportion of interval taken by the corresponding ribbon at I_j . We use the color order as it appears on I_j and set the height of a cell to be equal to the length of I_j .

The rationale behind coloring the background of the cells is to provide users with the composition details for the categories of the first axis in the Parallel Sets design. Sometimes, the interval I_j may be small and thus the ribbons may be hard to inspect. The gradient fill leverages the cell space to ease such inspection. The uniform cell space and linear ordering of the color gradients make it easier to compare various cell compositions visually with each other.

3.3 Map View

The third component of SET-STAT-MAP is a map view. The data is geolocated (i.e., gridded with latitude and longitude information) at about 36 km spatial resolution and displayed using color gradient heat map on a Google map. The heat map relates to a specific category at the last axis of the Parallel Sets view, equivalently, a row at the statistics column. Once the user selects a specific row, the heat map cells containing those data records are displayed. Let D be the set of data records for a specific cell. If the first axis categories are numeric ranges, then each heat map cell is colored based on the average of the attribute values. If the categories are non-numeric, then the color is chosen based on the most frequent first axis category in D . Similarly, we determine the textures for the cells by analyzing the second axis attribute.

While putting density plots or heat maps on map is self-explanatory for geolocated datasets, however, the encoding used in our heat map requires further explanation. The heat map provides users with an idea of where the data points are located. One can change the heat map resolution to visualize the distribution with different levels of granularity. The rationale behind using the colors and textures is to provide a sense of connection with the other components, as well as to provide users with a distribution of the colors and textures. Such a distribution over a spatial domain allows users to examine how various pairwise combinations of categories from the first and second axis relate in spatially.

3.4 Coordination and Interaction Support

We coordinate the views by synchronizing the updates on different views. The standard interactions on the Parallel Sets [27] have also been implemented in our design. Since the colors and textures on the ribbons are based on the first and second axis, one can reorder the axis and choose the first or second axis based on their need. Such an interaction simultaneously updates the colors of the statistics columns and heat map. Similarly, changing the last axis of the Parallel Sets view changes the number of rows in the statistics column based on the categories on the new last axis. For example, changing the last axis of Figure 1 from monthly to seasonal (spring, summer, fall and winter) would create 4 rows in the statistics columns view instead of 12 rows that it currently shows. Removal of the month axis and using Wind as the last axis would also create 4 rows corresponding to the 4 intervals at the Wind axis. This happens also for the reordering of axes, e.g., see Figure 10-11.

A statistics column can be sorted by clicking on the column header. Sorting a column will reorder both the rows and the intervals of the last axis of the Parallel Sets view. For example, in Figure 3, the MaxTemp column is sorted and the months are reordered based on this sorting (instead of being chronologically ordered from Jan. to Dec.). The users can filter the data records by drawing rectangles on the map view. This updates the ribbons in the Parallel Sets view and the plots on the statistics columns.

A novel interaction that SET-STAT-MAP introduces is interval selection. By selecting a vertical interval of an axis in the Parallel Sets view, the user can render the information (data records) of all the ribbons passing through that interval on the map. For the weather dataset, selecting such an interval (e.g., May) on the last axis (Month) would render all the tiles corresponding to this month on the map. We can thus see the spatial distribution of the first and second axis through color and texture distribution, respectively. This is illustrated in Figure 7(top), where by selecting the interval May at the Month axis, we can observe higher MaxTemp and precipitation at the south-west region and the opposite trend at the north-east region. In fact, this selection operation generalizes to the intervals of all the axes of the Parallel Sets, e.g., selection of the second interval of the Relative Humidity axis (Figure 7(bottom)) blends all the tiles corresponding to that interval over all months to depict a yearly trend. The map view of Figure 7(bottom) illustrates the yearly distribution of the first axis (MaxTemp) through the color distribution on the map, where we can immediately see a high variation of temperature within the selected humidity range at the north-east map region. However, using such an interval selection in the intermediate axis would require using simple textures such as dots to avoid visual clutter due to texture overlap.

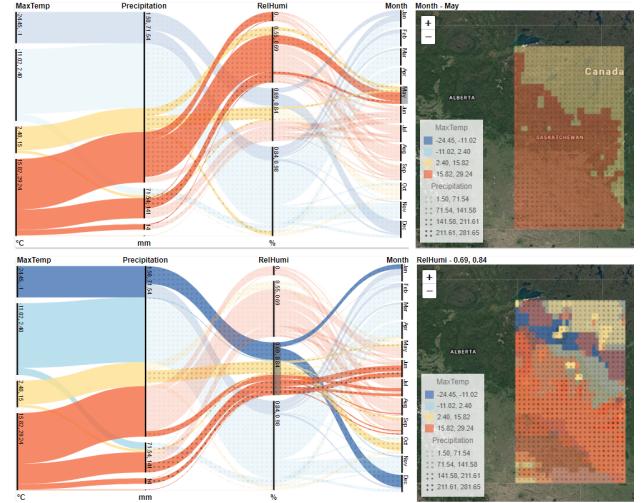


Figure 7: Interval selection in Parallel Sets: (top) selection of a temporal range, and (bottom) selection of a variable value range. The statistics columns are omitted for space constraints.

3.5 Design Rationale

The SET-STAT-MAP design follows Baldonado's [12] guidelines for designing multiple coordinated views. The three fundamentally different types of information (categorical, numeric, and geolocation) justifies leveraging the three views (Parallel Sets, statistics columns, and map). This follows the rules of *decomposition*, *complementarity* and *parsimony*, which suggests to partition complex data into multiple views to ease user comprehension, to provide complementary information to best leverage each view, and to use the views minimally (avoiding adding a view unless there is a significant benefit that justifies the cost).

The SET-STAT-MAP design also follows the rule of *resource optimization*. We organize the views side-by-side in a linear fashion, which creates a potential to provide (static or user interaction based) text summary beneath the display. In addition, a user can compare two SET-STAT-MAP visualizations by placing one on top of the other (Figure 3), which is useful to compare two map selections or two different datasets. Furthermore, we limit the number of categories visible simultaneously to maintain the system's interactivity. The flow of colour and texture from ribbons to the rows and finally, to the map make relationships among multiple views apparent to the user, which is aligned with the rule of *self-evidence*. The rule of *consistency* can be observed from the synchronized view updates upon user interactions (axis reordering, row sorting and map selection).

4 DATA PREPROCESSING AND IMPLEMENTATION

We implemented our proposed solution as a JavaScript based web component, which is cross-platform, easy-to-distribute, and easily embeddable in any in any modern web application (e.g., in a React app or even in a pure HTML/CSS/JavaScript project) irrespective of whether it is using a front-end framework or not. We used D3.js [2] and Textures.js [4] JavaScript library to build interactive visualizations. We also used a tool chain called Stencil [3] to build framework independent web components. Our Parallel Sets design is a standalone web component. Parallel Sets and the statistics columns together form another web component. The map view can pass data to those views by properties, and obtain data from them by events. This makes it easy to synchronize the updates to all the views during an user interaction.

The data is maintained using a single SQLite database file, containing a table with columns for all attributes (including latitude and longitude). We then use sql.js to query the SQLite file directly on the browser. We use SQL based queries to filter data based on users' map selection and compute necessary statistics on the fly. Thus the rendering time becomes directly proportional to the time to compute the statistics and the number of elements to be rendered. For large datasets, one can precompute the summary statistics and heatmap values into specialized data structures to improve scalability.

5 USE CASES

We demonstrate the SET-STAT-MAP technique using two use cases with datasets from totally different areas.

5.1 Meteorological Dataset

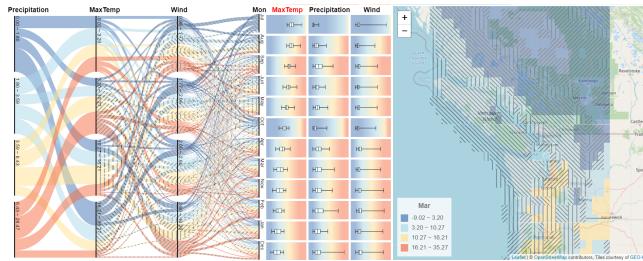


Figure 8: Illustration for the components of a SET-STAT-MAP visualization before the collaborative development.

Dataset description: This use case shows how an environmental scientist explored climate data using SET-STAT-MAP visualization. We use a Climate Forecast System Reanalysis (CFSR) database from the National Centers for Environmental Prediction (NCEP). Each record in the data consists of date, longitude, latitude, max temperature, min temperature, solar radiation, precipitation, relative humidity and wind speed. The longitude and latitudes of each data point is approximately 36 km away from the other data point (data

are at 36 km spatial resolution) Most attributes are numeric values for each date and geolocation, and hence ideal to test the functionality of SET-STAT-MAP on a predominately numeric dataset. We examined a dataset over the year 2013 and aggregated (by taking average or summery depending on attribute type) the data records by month, which transformed the Date attribute into a Month attribute.

Task Demonstration: To examine whether SET-STAT-MAP can reveal any useful information on this dataset, we initiated a collaboration with two environmental scientists: one within the university and another scientist from the Global Institute for Water Security. To describe the system to the scientists we planned to demonstrate the following tasks.

Examining Pairwise Relation: Examine the Parallel Sets view and examine whether there is a potential relation between the attributes X and Y.

Comparing Categories: Examine the last axis of the Parallel Sets view and compare these categories by sorting the statics columns.

Spatial Comparison: Load two SET-STAT-MAP view and select two areas to examine the similarity and differences between them.

Examining Attribute Composition: Consider a category on the last axis of the Parallel Sets view and examine the composition of the ribbons by examining colors and textures.

Examining Spatial Distribution: Consider the map view and examine how the categories of the first two axes are distributed over the map.

5.2 Collaborative Development

After we demonstrated the system to the scientists (Figure 8), we observed the need for several meetings so that they become comfortable in using the systems. We set weekly one-hour meetings over three months so that they can suggest modifications to SET-STAT-MAP.

Some major challenges that we noticed is that the experts took about three to four meetings before they could comprehend the system well enough to reason about it spontaneously. Initially, they had several questions on why the number of crossings increase as we move from the first to the later axes, and whether these crossings mean anything, and how they relate to the statistics column, what the colors in the statistics columns mean and so on. The Parallel Set design was not something that they were familiar with, and it was not straightforward for them to understand the connections among different views. It also took some time for the experts to recall the system interactions and effectively use SET-STAT-MAP.

Based on the experts' suggestion we refined the system to be used in a real-life setting (Figure 1). The experts wanted to add numerical axis at the bottom of the statistics columns, so that they can make the scientific reasoning and understand whether the data is being visualized meaningfully. They provided us with commonly used units for different meteorological variables and also they suggested color maps they are comfortable using. They preferred quartiles and simpler textures such as dotted ones over the line based textures as they felt it easier to understand (due to the smaller number of ribbons, less clutter and faster interactions).

We modified the legends on the map view to include texture information based on their suggestions. They also suggested that the color maps for temperature should be different from precipitation and follow standard color maps for the system to be used as they expect. Hence we created a separate control panel along with a split view so that the comparison between two different spatial regions becomes easier (Figure 9).

5.3 Experts' Analysis

We illustrate the usage of the system by describing a use case that was run by the experts (Figure 1). The experts used a hydroclimatic data from 2013 for the Canadian Prairie Ecozone (CPE). The CPE is cropped by a polygon covering an area of approx. 700,000 sq km (270,000 sq miles).

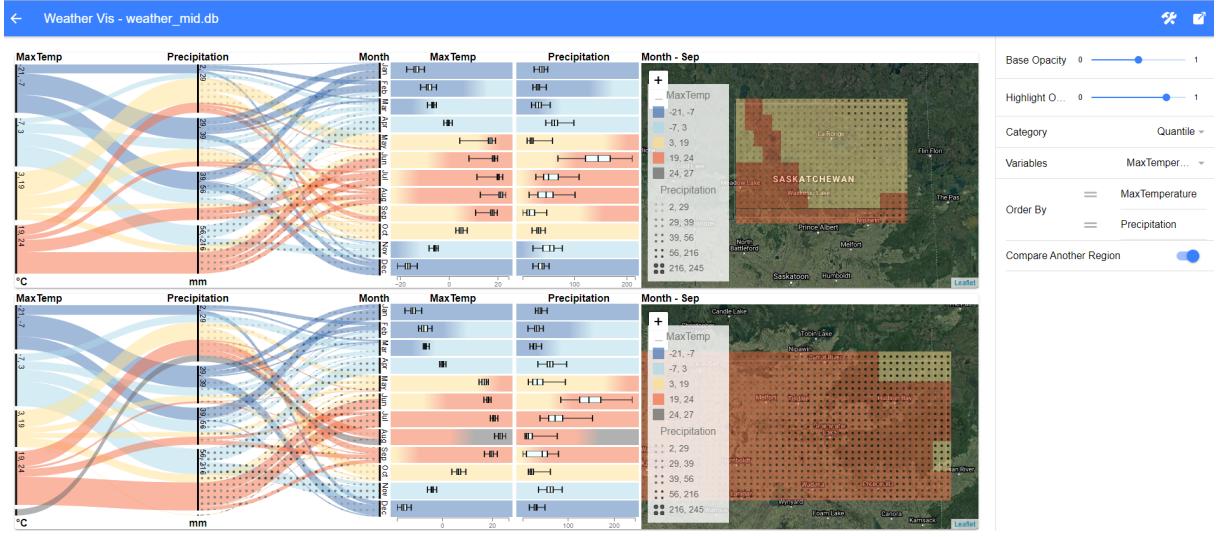


Figure 9: Comparing two different regions in a split view where the month September is selected in both views. The gray attribute in the Max Temp category at the bottom map denotes that, there exist a high temperature region in the selected area that falls outside of any temperature range shown on the top view.

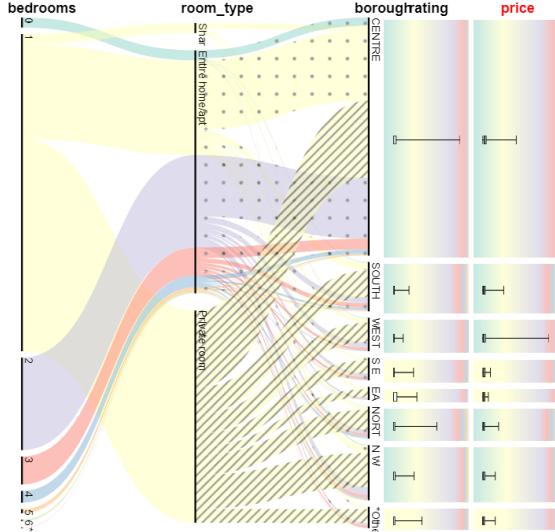


Figure 10: Exploration of Airbnb dataset where the city division attribute is chosen as the last axis of the Parallel Sets.

The experts selected the min temp, precipitation, and wind. Since the order of the variables in this study matters, they focused on the first two variables selected (i.e. min temp and precipitation). They chose the quartiles view as this is more common in hydrology to look at the highest and lowest quartiles. Looking at the Parallel Sets, they could immediately view that the lowest and highest quartiles of min temp are related to lower and higher amount of precipitation, respectively. They could also tell that higher min temp that corresponds with warmer months are associated with less wind. While colder months are expected to be with stronger wind gusts.

The texture showed more precipitation is associated with less windy scenarios and warmer months. The bar plot showed the monthly distribution of each selected variable. Precipitation is generally higher in the month of July but its daily values are with the highest variability. The average monthly wind gust and its daily variability in the months of January and November also are the highest.

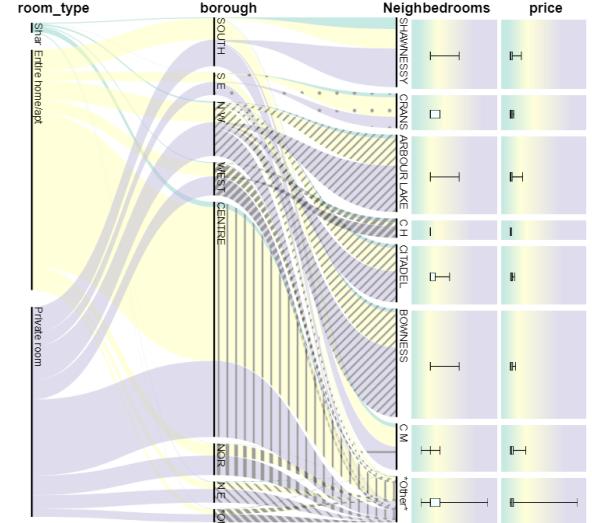


Figure 11: Exploration of Airbnb dataset where the neighborhood attribute is chosen as the last axis of the Parallel Sets.

The background of the bar plots in each column was identical and showed the distribution of the first selected variable, in this case, min temp with blue being the lower quantile or dark blue being the most dominant color in January and December and dark brown being the highest minimum temperature in the summer months and September. Finally, the Google overlay map showed the spatial distribution of the min temperature in color and precipitation in texture for a given month. They selected the month January and noticed that for 2013, Southern Manitoba, east, and north Saskatchewan experienced lower temperatures than the south Saskatchewan and Alberta.

The experts mentioned that the series of visualization provided by the system can be valuable for comparing key hydroclimatic variables. The also mentioned that the system, however, has its spatial and temporal limitations: (a) The data here is only collected for the year 2013, which makes it difficult to have meaningful comparisons with other years, comparing trends, and changes, which would be

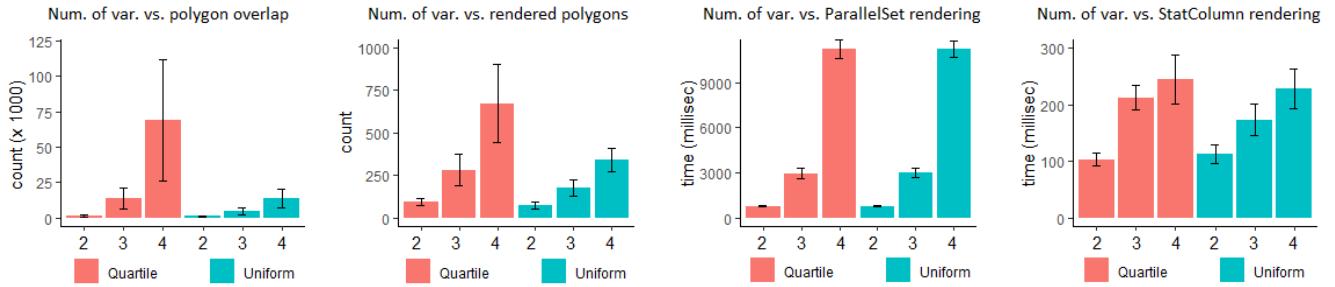


Figure 12: Performances of SET-STAT-MAP as the number of variables increase from 2 to 4. For i variables, where $i \in \{2, 3, 4\}$, the Parallel Sets view contains $(i+1)$ axes (i axes for variables and the last axis for Month).

of great interest for hydrologists and for climate change research. (b) The spatial resolution of the data is coarse (36 km) which is a challenge in understanding changes and shifts in variables over farms or urbanized areas.

We asked the experts about what tools they generally use to visualize geo-analysis results and how SET-STAT-MAP would compare to them. They mentioned that a typical output involves a single variable color-coded map, generated in Python or R, that can be also overlaid on a static or dynamic Google Maps. Generally, a couple of such maps, from different periods of time with the same variable or different variables from the same period of time, are compared side by side for decision making. In comparison, SET-STAT-MAP is capable of showing more than one variable in one geographic location at once making the relations among variables more detectable. The experts were enthusiastic about the prospect of SET-STAT-MAP and using it as a tool to display various standardized hydro-climate indices at once. These indices are built over long-term hydrological variables and traditionally, they are more indicative of drought and pluvial conditions [33]. Therefore, for future work, the experts wanted to extend our focus on using SET-STAT-MAP to present and compare the long-term indices, such as Standardized Precipitation Index (SPI) and Soil Moisture Deficit Index (SMDI), on one map. Another opportunity is to integrate SET-STAT-MAP with existing hydrological models so that one can validate the models’ prediction outputs against observed data.

5.4 Online Vacation Rental Dataset (Airbnb)

While SET-STAT-MAP was useful to reason about meteorological datasets, here we show a potential use case to demonstrate its use in an entirely different setting.

Dataset description: In this use case we examine an Airbnb dataset, which is an online vacation rental dataset. We use a dataset for listings over Calgary [1, 5], which contained 2521 listings from Dec 12, 2019. The final processed data contained many categorical and numeric attributes, e.g., room type, longitude, latitude, bedrooms, administrative division, neighborhood, overall satisfaction, price, etc. Since this dataset contains both categorical and numerical data, it provides a good opportunity to demonstrate the functionality of SET-STAT-MAP in handling mixed data.

Workflow: In this case the dataset was explored by a PhD student, who was familiar with both Parallel Coordinate and Parallel Sets visualization. He was interested in examining the Airbnb listings to better understand how the rental listing types and price vary across various divisions of the city.

He selected the attributes room type (shared, private room, or entire place), city division (borough), and bedrooms (number of rooms) to be visualized in the Parallel Sets view. Then he selected ratings and price attributes to be visualized in the statistics column. He first observed that there are only a few shared accommodation options compared to the private room or entire place option. He also noticed that private room always comes with a single bedroom as

expected, but the number of bedrooms varies widely for the entire place option. He then attempted to examine the median price over various division of the city. Therefore, he reordered the axis to set the city division to be the last axis (Figure 10), and then sorted the price column in descending order.

He noticed that the listings in the Calgary center and southern part had the higher median price, compared to the northern part of Calgary. He then examined the composition of room types in the Calgary center. By examining the color and texture, he found that about two-third of the listings are offering the entire place and among those listings the 1 and 2 bedroom types are almost equally frequent. To examine the dataset further the participant added the neighborhood, and found that a large number of neighborhoods are merged together. To find most affordable listings he sorted the price column in ascending order (Figure 11) and found the neighborhoods (Shawnessy, Crans, Arbour Lake, etc.) with low price listings. By examining the texture, he noticed that these places are far away from the Calgary center.

6 SCALABILITY OF SET-STAT-MAP

To test the limit of SET-STAT-MAP we analyzed its performance with the meteorological dataset [6] on an Intel i5-1135G7 machine with 16GB RAM, where the visualization was loaded with MS Edge 97.0.1072.55 on a 4K monitor. Figure 12 illustrates the performances of SET-STAT-MAP as the number of axes increases. The terms Quartile and Uniform correspond to the cases when the variables have been categorized based on quartile (e.g., Figure 3(top)) and equal value ranges (e.g., Figure 3(bottom)), respectively. Figure 12 shows that both the number of polygons rendered in the Parallel Sets view to draw the ribbons (and hence the rendering time), and the number of pairwise ribbon crossings increase drastically for five axes (4 variable axes and the Month axis). The rendering of statistics column takes negligible time (approx. 3% of total time) compared to the Parallel Sets rendering. The Parallel Sets view with Uniform categorization appears to be much more scalable than the Quartile based categorization.

An increased number of axes increases the pairwise ribbon crossings, and affects the readability of the visualization. Here the interval selection (Section 3.4) and opacity controls (Figure 9) come into play. As illustrated in Figure 7, the user can select an interval to highlight the ribbons passing through that interval to deal with overplotting. However, the overplotting problem remains if one desires to create a static picture or infographics.

7 DISCUSSION AND FUTURE WORK

From our experience, we have seen SET-STAT-MAP to perform very well if both the number of axes (variables) and the number of intervals (categories) per variable are bounded by 5. For an axis to be useful, it needs to have at least two intervals, each corresponding to a reasonable number of data records. Since the ribbons are

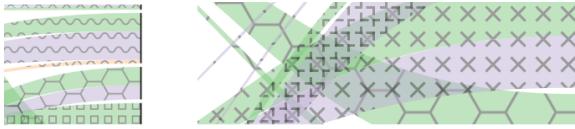


Figure 13: (left) Texture cut-off issues. (right) noisy pattern created by a combination of overlapping textures.

likely to get split in every subsequent axes, bringing the axes with few categories earlier in the axes sequence may reduce the overplotting problem. It would be interesting to further investigate such conditions for a dataset to be effectively visualized using SET-STAT-MAP. In addition, we have identified a number of key factors that needs to be carefully considered when designing a SET-STAT-MAP visualization.

Choice of axis: To effectively use SET-STAT-MAP, users need to understand the concept that the color and texture comes from the first and second axis, respectively, and the categories to be compared should be chosen as the last axis of the Parallel Sets. One option to ease the learning process, especially for non-scientific users, is to ask to identify important categories. For example, after loading the data, the system can ask to choose a categorical attribute (independent variable) over which the users would like to see the data pattern. This will be the last axis of the Parallel Sets view. After the user chooses the attribute, the system can ask to select the categorical and numerical attributes (dependent variables) of interest to populate the Parallel Sets and statistics column views. There can still be a steep learning curve to effectively use the numerical attribute as a categorical one. In the future, it would be interesting to investigate how the user's learning rate affects knowledge discovery.

Colors and opacity: One must add some transparency to the ribbons to provide users with a sense of connectedness. This also allows users to better perceive the line patterns that appears between adjacent axes. In a poorly designed color map some ribbon colors may interfere with its texture. A systematic way to choose colors would be to choose a dark color for texture and light colors for the ribbon color map, or vice versa. In our work we color the texture in gray and choose at most 8 other light colors from a perceptual color map. Ordering of the colors is also important as it helps users better understand the distribution of numeric ranges.

Choice of texture: Ribbon texture is an important aspect of SET-STAT-MAP design. While examining the ribbons intersecting a line interval, the textures of the ribbons can reveal the composition of the second axis categories. In a poorly designed texture map, thin ribbons may entirely miss the pattern or provide a partial pattern creating ambiguity. The problem of cut-off is illustrated in Figure 13(left), where the top ribbon can be interpreted as a cut-off cross or wave pattern. Similarly, The fourth ribbon from the top can be seen as a cut-off circle or wave pattern. Another problem of choosing complex patterns is that they are difficult to relate with numeric ranges, and overlapping ribbons makes the visualization cluttered by creating noisy patterns, as illustrated in Figure 13(right), which increase users' cognitive load.

Adaptive label size and mouse over: A SET-STAT-MAP design creates intervals of widely varied lengths. Therefore, some interval may be very thin, creating a problem for label placement, e.g., see Figure 14(left). To best leverage the vertical space, a careful design must attempt to limit the number of category per axis that are visible simultaneously, and provide functionality such that the users can sort and scroll the category list. Furthermore, the label size should be adaptive to the length of the line to avoid visual occlusion (Figure 14(middle)), and allow users to inspect detailed information on mouse over (Figure 14(right)).

Linear gradient color: The background of each cell in the statistics columns view is colored with a linear gradient. Such a linear



Figure 14: (left) Label overlapping. (middle) Adaptive labelling. (right) Mouseover interaction.

gradient may not create a color range with a proper perceptual balance, i.e., people may perceive the width or strength of the colors differently than the underlying numeric values. One problem with multiple colors appears to be with numeric ranges, where the preferred color map is sequential or diverging. Since the adjacent colors in such a color map is perceptually close to each other, the linear gradient may make it harder to distinguish the color boundaries (Figure 15). For categorical color maps, one may obtain a background with multiple vertical stripes. We suggest minimal use of the number of colors since multiple stripes may produce unwanted visual effects (e.g., McCollough effect [16] or optical illusion) over prolonged use. Further investigation on how human perceive such linear gradient in a visual analytics context, or whether the benefit of gradient exceeds the potential risk can be an exciting avenue of research.

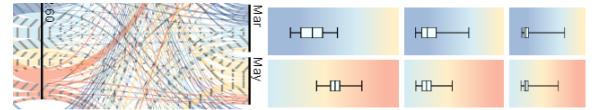


Figure 15: Linear gradient makes the boundaries harder to distinguish.

Adaptive heatmap: The SET-STAT-MAP visualization uses a heatmap view on the map, where the cells are colored and textured based on the first and second axis. If both these axes are numerical, then for each cell one can choose the color and text based on the average value. However, for categorical attributes, it needs to be based on a summary statistic. In our design, we pick the color and texture based on the high frequent categories. However, based on the data (e.g., when all categories are equally frequent) this may create misleading representation. One may partially improve the problem by using an adaptive quadtree-based heatmap or by providing complex glyphs (e.g., pie charts). However, a conscious investigation is required to shed more light into this situation.

Perceptually-motivated Parallel Sets: Although we currently used the first two axes for color and texture augmentation, it would be interesting to investigate ways to extend this to more axes. We believe that our effort for stylizing ribbons to analyze the distribution of a combination of attributes may potentially be extended further by leveraging the idea of perceptually-motivated visualization [23] by mapping the categories of different axes with different perceptual properties. An effective perceptually-motivated visualization is harder to design and also to interpret. However, such a visualization can still be valuable in an infographic setting, as well as to examine or extend the boundary of human perception.

8 CONCLUSION

Although Parallel Coordinates and Parallel Sets visualizations are widely used to interactively explore numerical and categorical datasets, respectively, there still exist opportunities to explore new methods to visualize spatial data with mixed numeric and categorical attributes. In this paper we propose SET-STAT-MAP that leverages the idea of multiple coordinated views. The design harmoniously blend a Parallel Sets view, a statistics columns view and a map view to visualize mixed data in a linear fashion. Our design allows users to explore the pairwise attribute relationships, statistical summaries and

trends, and spatial distribution of various attributes. We demonstrate the functionality of SET-STAT-MAP using real-life datasets.

Although SET-STAT-MAP is currently a research prototype, we plan to make our web-based implementation available to the developers through open-source software repositories. We believe our approach to visualizing mixed data will inspire future work in visualization and interface design to explore heterogeneous datasets.

ACKNOWLEDGMENTS

This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund. D. Mondal would like to thank Purbasha Mistry for stimulating discussion.

REFERENCES

- [1] airbnb-data – tom slee. <https://tomslee.net/category/airbnb-data>.
- [2] D3.js - data-driven documents. <https://d3js.org/>.
- [3] Stencil. <https://stenciljs.com/>.
- [4] Textures.js. <https://riccardoscalco.it/textures/>.
- [5] tomslee/airbnb-data-collection: Data collection for airbnb listings. <https://github.com/tomslee/airbnb-data-collection>.
- [6] Weather data (SWAT). <https://globalweather.tamu.edu/>.
- [7] Geo-coordinated parallel coordinates (GPCP): Field trial studies of environmental data analysis. *Visual Informatics*, 2(2):111 – 124, 2018.
- [8] Z. Ahmed, M. Beyene, D. Mondal, C. K. Roy, C. Dutchyn, and K. A. Schneider. Contourdiff: Revealing differential trends in spatiotemporal data. In *Proc. of Information Visualisation (IV)*, pp. 35–41. IEEE, 2021. doi: 10.1109/IV53921.2021.00016
- [9] B. Alper, N. H. Riche, G. A. Ramos, and M. Czerwinski. Design study of linesets, a novel set visualization technique. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2259–2267, 2011. doi: 10.1109/TVCG.2011.186
- [10] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. J. Rodgers. Visualizing sets and set-typed data: State-of-the-art and future challenges. In *Proc. of Eurographics Conference on Visualization (EuroVis), Swansea, UK, June 9–13, 2014*. Eurographics Association, 2014. doi: 10.2312/eurovisstar.20141170
- [11] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. J. Rodgers. The state-of-the-art of set visualization. *Comput. Graph. Forum*, 35(1):234–260, 2016. doi: 10.1111/cgf.12722
- [12] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the working conference on Advanced visual interfaces (AVI)*, pp. 110–119. ACM Press, 2000.
- [13] J. Blaas, C. Botha, and F. Post. Extensions of parallel coordinates for interactive exploration of large multi-timepoint data sets. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1436–1451, 2008. doi: 10.1109/TVCG.2008.131
- [14] L. A. Bruckner. On chernoff faces. In *Graphical representation of multivariate data*, pp. 93–121. Elsevier, 1978.
- [15] C. Collins, G. Penn, and S. Carpendale. Bubble Sets: revealing set relations with isocontours over existing visualizations. *IEEE Trans. Vis. Comput. Graph.*, 15(6):1009–1016, 2009. doi: 10.1109/TVCG.2009.122
- [16] P. C. Dodwell and G. K. Humphrey. A functional theory of the mclough effect. *Psychological Review*, 97(1):78–89, 1990.
- [17] D. Dorling, A. Barford, and M. Newman. WORLDMAPPER: the world as you've never seen it before. *IEEE transactions on visualization and computer graphics*, 12(5):757–764, 2006.
- [18] D. A. Ellsworth, C. E. Henze, and B. C. Nelson. Interactive visualization of high-dimensional petascale ocean data. In *Proc. of IEEE Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 36–44. IEEE, 2017.
- [19] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. LineUp: visual analysis of multi-attribute rankings. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2277–2286, 2013. doi: 10.1109/TVCG.2013.173
- [20] N. Gurran and P. Phibbs. When tourists move in: How should urban planners respond to Airbnb? *Journal of The American Planning Association*, 83(1):80–92, 2017.
- [21] X. He, Y. Tao, Q. Wang, and H. Lin. Multivariate spatial data visualization: a survey. *J. of Visualization*, 22(5):897–912, 2019. doi: 10.1007/s12650-019-00584-3
- [22] C. G. Healey and J. T. Enns. Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE transactions on visualization and computer graphics*, 5(2):145–167, 1999.
- [23] C. G. Healey and J. T. Enns. Attention and visual memory in visualization and computer graphics. *IEEE Trans. Vis. Comput. Graph.*, 18(7):1170–1188, 2012. doi: 10.1109/TVCG.2011.127
- [24] J. Heinrich and D. Weiskopf. State of the art of parallel coordinates. In M. Shert and L. Szirmay-Kalos, eds., *Proc. of European Association for Computer Graphics (Eurographics)*, pp. 95–116. Eurographics Association, 2013.
- [25] A. Inselberg and B. Dimsdale. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In A. E. Kaufman, ed., *Proc. of IEEE Visualization Conference, San Francisco, CA, USA, October 23–26, 1990*, pp. 361–378. IEEE, 1990. doi: 10.1109/VISUAL.1990.146402
- [26] B. Jacobsen, M. Wallinger, S. G. Kobourov, and M. Nöllenburg. Metrossets: Visualizing sets as metro maps. *IEEE Trans. Vis. Comput. Graph.*, 27(2):1257–1267, 2021.
- [27] R. Kosara, F. Bendix, and H. Hauser. Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE Trans. Vis. Comput. Graph.*, 12(4):558–568, 2006. doi: 10.1109/TVCG.2006.76
- [28] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister. Upset: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.*, 20(12):1983–1992, 2014. doi: 10.1109/TVCG.2014.2346248
- [29] F. Liu and R. W. Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE transactions on pattern analysis and machine intelligence*, 18(7):722–733, 1996.
- [30] D. Mondal, M. Mondal, C. K. Roy, K. A. Schneider, Y. Li, and S. Wang. Clone-world: A visual analytic system for large scale software clones. *Vis. Informatics*, 3(1):18–26, 2019. doi: 10.1016/j.visinf.2019.03.003
- [31] T. Munzner. *Visualization Analysis and Design*. A.K. Peters visualization series. A K Peters, 2014.
- [32] J. C. Roberts. State of the art: Coordinated multiple views in exploratory visualization. In *Proc. of International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*, pp. 61–71, 2007. doi: 10.1109/CMV.2007.20
- [33] S. Sadri, M. Pan, Y. Wada, N. Vergopolan, J. Sheffield, J. S. Famiglietti, Y. Kerr, and E. Wood. A global near-real-time soil moisture index monitor for food security using integrated SMOS and SMAP. *Remote Sensing of Environment*, 246(7):111864, 2020.
- [34] M. Sarkar, S. S. Snibbe, O. J. Tversky, and S. P. Reiss. Stretching the rubber sheet: A metaphor for viewing large layouts on small screens. In S. E. Hudson, R. Pausch, B. T. V. Zanden, and J. D. Foley, eds., *Proc. of ACM Symposium on User Interface Software and Technology (UIST), Atlanta, GA, USA*, pp. 81–91. ACM, November 3–5, 1993. doi: 10.1145/168642.168650
- [35] G. Sharma and R. Bala. *Digital color imaging handbook*. CRC press, 2017.
- [36] H. H. Shenas and V. Interrante. Compositing color with texture for multi-variate visualization, 2005.
- [37] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, man, and cybernetics*, 8(6):460–473, 1978.
- [38] C. Ware. *Information Visualization, Second Edition: Perception for Design (Interactive Technologies)*. 04 2004.
- [39] K. Wu and S. Zhang. Visualizing 2D scalar fields with hierarchical topology, 2015.
- [40] G. M. H. Zahran, D. Mondal, and C. Gutwin. Contour line stylization to visualize multivariate information. In *Proc. of Graphics Interface Conference (GI)*, pp. 198 – 207, 2021.