Visualization of Cluster Structure and Separation in Multivariate Mixed Data: A Case Study of Diversity Faultlines in Work Teams

Tuan Pham1,∗, Ronald Metoyer1, Katerina Bezrukova2, Chester Spell3

Abstract
In organizational management, researchers and managers study separations or faultlines that occur in diverse teams when members form subgroups based on the alignment of multiple demographic characteristics. The team faultline concept is operationalized using multivariate cluster analysis—analysts use faultline measures to identify subgroups/clusters in a team and to quantify how subgroups/clusters are separated. Unfortunately, these measures have limited capacity to enable users to observe and explore faultlines and subgroup structure across the examined attributes efficiently. We address this problem and make three contributions. First, we propose a visual representation for communicating faultline information that is based on multiple linked, stacked histograms in an axis-parallel layout. Second, we evaluate the effectiveness of the proposed technique in a controlled user study, comparing it to the two other common multivariate representations of clusters: parallel coordinates and scatter plot matrices. While we chose faultline-related tasks based on the requirements by domain experts in organizational management, the study findings can be generalized to representations and tasks involving distributions of clusters of multivariate objects in mixed-type data. Finally, inspired by geological faultlines, we propose several visual enhancements to stacked histograms to further facilitate the task of identifying “cracks” within work teams.

Keywords:
clustering, cluster separation, functional diversity, evaluation, faultlines, stacked histograms, user study

1. Introduction

Effective management of work teams is widely regarded as critical to the success of organizations. Therefore, leveraging the benefits of teamwork while reducing negative outcomes associated with groups has been a central focus of organizational research [1, 2, 3, 4]. For example, researchers study how the demographic diversity of team members such as age, gender, ethnicity, and functional background affects outcomes such as performance and productivity as well as group processes such as collaboration and conflict. They investigate diversity not only as a distribution along one employee attribute, for instance, group ethnic diversity, but also as a complex composition of multiple attributes that results in diversity faultlines [1]. For instance, faultlines may split a diverse project team into two subgroups: one of two senior male software engineers and the other of two junior female QA testers.

A common approach to understanding faultlines within a team relies on faultline metrics [5, 2, 3], which measure the extent to which the given team is divided into relatively homogeneous subgroups across the attributes of interest, and tabular data of subgroup structure (see Table 1 for an example).

From an analysis point of view, the goal is not necessarily to identify the objects that cluster together, but to identify how attribute space is divided up into the clustered subgroups. Unfortunately, as the number of attributes and team members to be examined both increase, table-based assessment of faultlines and subgroup structure becomes difficult, time-consuming, and tedious. To our knowledge, very little work has been done to develop visual representations that reveal faultlines across multiple attributes. In fact, this lack of tools is considered a challenge in management research that hinders the development of the faultline theory to a more applicable and useful level [3, 4].

We envision that the analysis of faultlines would be complemented by a visual analytics approach that leverages faultline metrics with appropriate representation and interaction techniques. Specifically, such a visual interface would allow researchers to explore faultlines and structure of subgroups within teams quickly and iteratively. Managers or human resources departments could use data visualization to inspect team dynamics based on faultlines and potentially reassign members in hopes of improving performance. Such visualizations could also prove useful in understanding the dynamics of online volunteer teams, for instance, open source software development teams [6, 7].

In addition to team faultlines, the problem of visualizing cluster structure and separation in multivariate mixed data represents itself in other application domains. For instance, ecologists and microbiologists recognize functional diversity as the variety of roles played by different species or their equivalents.

1Email: pham@eecs.oregonstate.edu (or tuanpham@live.com)
2School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon 97331 USA
3Department of Psychology, Santa Clara University, Santa Clara, California 95053 USA
4Rutgers School of Business–Camden, Rutgers University, Camden, New Jersey 08102 USA

Preprint submitted to Computers & Graphics

October 7, 2013
based on their composition of multiple functional traits such as rooting depth and maximum growth rate of plants \[8, 9, 10\]. Technically, composition of these traits can be used to cluster different species (or their equivalents) present in a unit of study into different functional groups and to derive, for example, the functional diversity (FD) metric \[8, 9, 10\]. In summary, clusters may represent functional groups in an ecological unit of study (e.g., ecological communities) or subgroups in an organizational unit of study (e.g., work teams); clusters may also represent different units of study under comparison.

In this paper, we formalize and generalize the faultlines visualization problem as visual analysis of cluster structure and separation in multivariate mixed data. In doing so, we provide three contributions. First, we propose a representation that aims to reveal faultlines and subgroup structure of diverse teams across multiple attributes. The proposed representation, HIST, is based on multiple linked, stacked histograms in an axis-parallel layout \[11, 12\], as depicted in Figure 1. To our knowledge, while these techniques separately are well-known, as a whole, their application to representing clusters of multivariate objects in general and team diversity faultlines in particular is novel and it is a first attempt to explore the design space for the problem. Moreover, the novelty of HIST is in the emphasis on attribute visibility \[13\], such as the distribution of clusters in attribute space, as opposed to object visibility \[14, 15\] when representing clusters.

Second, we contribute results of a controlled user study to compare HIST to the parallel coordinate plot (PCP) \[11, 12\] and the scatter plot matrix (SPLOM) \[16\], the two other commonly used techniques for representing clusters of multivariate objects, as identified in previous work by Holten and van Wijk \[14\]. With respect to user performance, our results show that (1) users can judge faultlines using HIST as or more accurately than when using the other two methods and (2) HIST performance holds consistent across task questions and data set sizes. Furthermore, the findings can be generalized to representations and tasks involving distributions of clusters of multivariate objects in mixed-type data, extending the previous work \[14\]. The generalization is reflected in our choice of task questions that are relevant to both faultlines and general cluster representations and in the application of HIST to other application domains such as functional diversity in ecology.

Finally, we incorporate computational analysis into HIST to assist users in detecting faultlines or subgroup separation. Specifically, inspired by the physical form of geological faults, we propose novel visual enhancements as connecting dashed lines across attribute axes to represent “cracks” within a team, as depicted in Figure 2. In our algorithm, we cluster attribute values by subgroups using Bertin Classification Criterion \[17\] and we introduce a metric, Total Separation Criterion, to automatically detect attributes with separable subgroups.

### 2. Diversity Faultlines Background and Design Requirements

#### 2.1. Diversity Faultlines Concept

Faultlines are described as hypothetical dividing lines that may split a team into relatively homogeneous subgroups based on one or more attributes \[1\]. Measuring faultlines of a team is adopted from multivariate clustering—that is, the measure assigns team members into subgroups (or clusters) according to their similarity across the attributes of interest, for instance, demographic characteristics. Clusters (or subgroups) have maximum internal homogeneity or between-cluster heterogeneity.

Team data represent team members characterized by multiple demographic attributes of varying types, including numeric, ordinal, and nominal. As an example, consider two teams as shown in Table 1: Teams 1 and 2 consist of five and seven members, respectively. We computed team faultlines along the three characteristics of AGE, ETHNICITY, and EDUCATION (degree) using a widely accepted measure proposed by Thatcher et al. \[5\]. For each team, the measure identifies the subgroups (Subgroup column) corresponding to the strongest group partitioning following the formula:

\[
Fau_g = \sqrt{\frac{\sum_{j=1}^{p} \sum_{k=1}^{n} n^g_k (\bar{x}_{jk} - \bar{x}_j)^2}{\sum_{j=1}^{p} \sum_{k=1}^{n} \sum_{i=1}^{n^g_k} (x_{ijk} - \bar{x}_j)^2}}}, \quad g = 1, 2, \ldots, S. \tag{1}
\]

where \(p\) is the number of members of interest, \(n_g\) is the number of subgroups in the partition \(g\), \(n^g_k\) is the number of members in subgroup \(k\) of partition \(g\), \(\bar{x}_{jk}\) is the mean value of attribute \(j\) in subgroup \(k\), \(\bar{x}_j\) is the overall mean value of attribute \(j\), and \(x_{ijk}\) is the value of attribute \(j\) of member \(i\) in subgroup \(k\). Simply put, the measure iterates through all possible partitions (splits) of the team into subgroups and finds the largest ratio of the between subgroup sum of squares to the total sum of squares. Since \(Fau_g\) takes numeric values, each categorical attribute must be recoded into a series of dummy variables and rescaled across the attributes \[5\]. For example, a five year difference in AGE is equivalent to a difference in ETHNICITY and a difference in EDUCATION based on a given data sample.

The variable faultline strength \(Fau\), which always takes a value between zero and one, is the maximum over all \(\{Fau_g\}_{g=1}^{S}\). The larger the faultline strength value, the stronger the separation between subgroups or equivalently, the more attributes in which the subgroups are separable. The concept is inspired by geological faults whose strength increases with the number of

<table>
<thead>
<tr>
<th>Team</th>
<th>AGE</th>
<th>ETHNICITY</th>
<th>EDU</th>
<th>Subgroup</th>
<th>Fau</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>T</td>
<td>E</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>T</td>
<td>E</td>
<td></td>
<td>0.56</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>T</td>
<td>E</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>Y</td>
<td>A</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>52</td>
<td>Y</td>
<td>A</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>W</td>
<td>E</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>W</td>
<td>A</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>U</td>
<td>B</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>X</td>
<td>B</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>Z</td>
<td>D</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>Z</td>
<td>C</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>Z</td>
<td>B</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1: Synthetic data of the two work teams. Faultline measure \[5\] clusters each of the two teams into subgroups (Subgroup) and identifies the team faultline strength (Fau).
layers it cuts through [1]. Since Fauf is based on a brute-force search, it is suited only for small teams. Thatcher et al. [3] and Meyer and Glenz [18] present thorough surveys of existing faultline measures.

2.2. Design Requirements

While Table 1 describes each of the team members of the two teams in detail, the table does not clearly show where the separation (or “cracks”) occur in a team. That is, within an attribute, it is unobvious how the attribute space is potentially occupied by different subgroups, which is precisely the problem that we address with our visual representation. Here we discuss design requirements as validated by our two collaborators, who are experts in management research and also co-authors on this paper. These requirements aim to capture the experts’ information needs when they study team faultlines data (e.g., [19]). Moreover, these requirements are empirically associated with team outcomes in the faultlines literature, as we cite in the list of requirements below. Specifically, a faultline representation of a given team should allow users to explore efficiently:

- **R1.** Faultline value (e.g., faultline strength Fauf). Such numeric quantification of a faultline can be used to compare different teams quickly or to predict the effects of faultlines on outcome processes [2, 3].
- **R2.** Faultlines themselves, or where do the “cracks” occur in the team? A “crack” or total separation occurs within an attribute when members of different subgroups fall into different subsets of values in the attribute space.
- **R3.** The inner structure of subgroups in the team including the number of subgroups, evenness of subgroups, and multivariate distribution of members across the examined attributes [20]. These important constructs are associated with distribution of power, resources, and abilities in the team [1, 4].

In addition, the representation should scale well to the number of members in a team and number of attributes of interest. Management researchers have typically studied small teams of up to 16 members that may potentially split into up to seven subgroups, depending on team size and the number of attributes [21, 4], yet they are also interested in teams of larger sizes, for instance, online volunteer groups [6, 7].

2.3. Conventional Cluster Analysis vs. Faultlines Analysis

Finally, while conventional multivariate cluster analysis usually concerns object visibility and separation in attribute space of quantitative attributes [15], we note that faultlines analysis emphasizes distribution or alignment of objects across multiple attributes of varying types or, in other words, the role of attributes in structure and separation of clusters. Furthermore, a faultlines visualization requires a faultline measure or a clustering algorithm as an external data pre-processing step to pre-assign team members to subgroups, as opposed to letting analysts identify potential subgroups or implicit clusters from representations of raw data [14].

3. Related Work

Design and evaluation of our proposed technique was informed by related work on visual representations and user studies of cluster representations, which we discuss here.

3.1. Representing Clusters

Here we review a subset of existing representation techniques that are potentially applicable to multivariate cluster analysis and team faultlines. More general surveys of visual representations can be found in [22, 23].

Scatter plots are probably the most common technique to represent clusters of objects [14]. However, without additional encoding, possible data overlap/occlusion may lead to ambiguous interpretation of the abundance of objects, especially among categorical attributes. The histogram, on the other hand, takes advantage of data overlap to show the distribution of objects over a single attribute. Our proposed technique, which is based on histograms, aims to convey object distribution instead of object visibility. As noted, these techniques display only one or two attributes of interest.

The dimensionality problem may be solved by using multiples. For example, the scatter plot matrix (SPLOM) [16] extends scatter plots to represent clusters of multivariate objects, although multiple pairwise projections of the data attributes require more screen space and potentially cognitive load placed on the user. On the other hand, multiple histograms could be useful for representing distribution of multiple attributes in an axis-parallel layout [24]. Furthermore, histograms have been proven effective in communicating diversity information in separate attributes in previous work [25, 26]. Our proposed representation of diversity faultlines is in fact multiple histograms augmented with histogram stacking and color encoding.

The parallel coordinates plot (PCP) [11, 12] is another common approach to representing clusters of high-dimensional objects [14]. Similar to SPLOM, PCP may suffer from occlusions caused by data overlap as the number of objects increases and many categorical attributes exist, as in the case of demographic data. Several variants of PCP such as Parallel Sets [27] and Diversity Map [26] overcome this limitation by providing information on the distribution of values for each attribute. However, it is not clear how multiple clusters are embedded into these techniques.

Star coordinates [28] may be suited to representing the overall structure of a set of objects over multiple attributes. Additional encoding such as colors may be used to reveal explicit clusters in the data. Unfortunately, the mapping between a data point and its location in star coordinates is not one-to-one. Consequently, several different data points may end up in the same location if they have equal vector sums.

Among stacked displays [22, 29], the mosaic plot [30] could be used for showing subgroup structure since subgroups are stacked within a team. While in theory, the stacking process may be repeated multiple times, in practice, space constraints limit the number of attributes as well as the number of possible values in an attribute. Therefore, mosaic plots can be useful only when the number of attributes is relatively small.
our proposed histogram-based technique, we apply the stacking process to histogram bars only once.

Finally, there are hybrid approaches that integrate multiple representations in a single view or in multiple coordinated views. In the former group, the most relevant technique is DICON [31], a treemap- and icon-based technique designed to visualize structure of clusters. Unfortunately, the technique supports only quantitative attributes, which is not sufficient for team faultlines data that are usually multivariate with varying attribute types (e.g., numeric, nominal, and ordinal) as described in Section 2. In the latter group of multiple views, VisBricks [32] clusters an inhomogeneous data set into different subsets and visualizes them using different representation techniques augmented with coordination interaction features. Our proposed representation technique could be potentially incorporated into VisBricks as a building block.

3.2. Evaluating Cluster Representations

The closest exemplar to our user study is that of Holten and van Wijk [14]. They evaluated cluster identification performance of nine PCP variants, two of which are the standard PCP and a variant with embedded scatterplots (SP). Nevertheless, unlike our scenario involving explicit clusters in demographic data, their study used simulated quantitative data with no pre-computation of clusters. The most interesting finding from their study is that despite the apparently valid improvements of the PCP variants, scatterplots are more effective than PCPs with respect to PCP-based cluster identification tasks. Furthermore, participants favored SP as the least difficult variation. Following this result, the authors called for further evaluation of techniques that explicitly highlight pre-computed clusters, for example, with unique colors.

4. Visualization Design

4.1. Design Considerations and Prototype

A histogram is well suited to showing the diversity or distribution of objects within an attribute (requirement R3). According to Mackinlay [33], position and length are ranked highly for encoding nominal and numeric values such as variety of attribute values and abundance of objects, respectively. In addition, previous work suggests that the axis-parallel layout [11, 12] of multiple distributions is capable of conveying a holistic object distribution over multiple attributes [24, 26]. However, the previous work does not consider how distributions of multiple subgroups align over multiple attributes. Since subgroups are nested within a team, to maintain bar length encoding, a natural solution to encoding subgroups is to stack bars within each bin (Figures 1 and 2). We then use distinct color hues on a white background to differentiate stacked subgroups. Our choice of qualitative colors provided by ColorBrewer [34] meets the requirement of encoding up to seven subgroups. On another note, the length of each bar is scaled according to $l(x) = |x|/|x_{MAX}|$, where $|x|$ denotes the number of objects in bin $x$, and $x_{MAX}$ is the bin with the most objects for the attribute in question. We also discretized numeric attributes into bins based on their rescaled factors (Equation 1).

Following this design, a total separation or “crack” occurs at a nominal attribute when distinct subgroups (or distinct colors) occupy distinct positions along the vertical axis (requirement R2). Total separation at a numeric or ordinal attribute further requires that these distinct positions—including ones without objects (zero-length bars)—are contiguous, for instance, AGE and MLB_TENURE histograms in Figure 2.

The HIST representation communicates the overall degree of separation of subgroups in a given team (i.e., faultline strength) as the combined separation of all demographic attributes under investigation (requirement R1). In the limit of perfectly strong faultlines, where different subgroups occupy different subsets
of attribute values across all the attributes, all the bars of the histograms will have solid colors, as depicted in Figures 1(a) and 2. On the contrary, a team with very weak faultlines will produce a visualization with most of the bars stacked with at least two colors like the AGE histogram in Figure 1(b). Moreover, while the chosen Fa measure (Equation 1) [5] does not consider how far apart the subgroups are, especially on quantitative attributes (i.e., faultline distance [2]), we note that stacked histograms of quantitative axes are able to reveal the potential gaps or distances between subgroups. For instance, the AGE histogram in Figure 1(a) shows a big “generation gap” between the two subgroups.

4.2. Informal Evaluation and Motivation for a Formal Study

A close collaboration between management and visualization researchers was critical for the design of HIST. The domain experts helped validate the design requirements and evaluate the design iterations and prototypes. We adopted an iterative user participatory design approach [35] in which the management researchers participated in every step of the design process. First, we used the real-world data collected by the management researchers [19] to identify their information needs. The example data were valuable for helping the visualization researchers understand the problem and facilitate the second step involving iterative discussions of the data and visualization design requirements. After ensuring all the requirements were captured, the visualization team sketched several mockup prototypes using hypothetical team data and gathered feedback from the management researchers. As the design gains maturity, the subsequent steps involved design implementation, testing and exploration with real-world data sets, and analysis of tasks and questions for the user study.

Thus far, we have applied the prototypes to two real-world data sets: Major League Baseball (MLB) teams (Figure 2) and an empirical faultlines study [19] (Figure 3). The domain experts found the representation helpful in inspecting subgroup structure of different teams and in developing a sense of where the separations are likely to occur following their configuration of the faultline measures.

While the qualitative results from our informal evaluation are encouraging, they have limitations. First, the evaluation is observational [36] and lacks controlled visualization techniques (control groups) as well as various data sets with controllable characteristics serving as ground truth answers. Second, our two management researcher collaborators represent only a small set of potential users of the visualization. The proposed faultlines visualization, HIST, could potentially support a wide range of target users: (1) management researchers who study faultlines and subgroups theories [4], (2) human resources departments who manage current employees and recruit new employees [37], and (3) managers and officials from many areas concerning work teams such as education, sports, and entertainment to name a few. Third, thus far, the management researchers limited the use of the faultline visualizations to data exploration only, accompanied by further statistical analysis. The design of HIST targets both data exploration such as data
analysis and communication such as charts in a publication or training. Finally, while HIST is designed based on the requirements of faultlines and subgroup structures in work teams, it can be potentially utilized to communicate distributions of clusters of multivariate objects in mixed-type data, for example, to compare structures of functional groups in ecological and microbiological data [8, 9, 10, 38], as we mentioned in the Introduction section.

To overcome these limitations and make the evaluation results generalizable, in the next section, we extend our evaluation with a controlled user study designed to understand the effectiveness of a visual representation in a broader context of communicating separations and distributions of clusters/subgroups of multivariate objects in mixed-type data.

5. User Study Design and Implementation

In this section, we describe the design and implementation of a formal user study intended to evaluate the effectiveness of HIST at communicating cluster separation and structure in the specific context of team faultlines. In particular, we compare HIST to PCP [11, 12] and SPLOM [16], the two common techniques for representing clusters of multivariate objects. In fact, a previous study has shown that the standard PCP and a PCP variant with embedded scatterplots are the most effective among variants of PCP for cluster identification tasks [14]. Figure 4 depicts examples of the three techniques.

5.1. Task Design and Implementation

The task design includes three important components: (1) a set of task-oriented questions, (2) a procedure for generating synthetic team data, and (3) design of the three visualization techniques under comparison.

User Study Task Questions. The study contains six types of questions intended to assess the capability of a particular visual representation in conveying different aspects of team faultlines (requirements R1 - R3). Note that in accordance with the previous cluster identification study [14], we design the tasks to be relevant to both faultlines and general cluster representations of mixed-type data and not tied to users with specialized knowledge of demographics. Therefore, for each of the question types, we also provide the equivalent generic evaluation question in parentheses.

Q1: How many subgroups are there in the given team? (generic form: How many clusters are there in the data set?) (possible answers: 1 to 7) This question type is designed to determine if a representation technique supports users in identifying the number of subgroups/clusters in a team/data set (requirement R3). This type is equivalent to the only cluster identification task in the previous study [14].

Q2a/b: Among the existing subgroups in the given team, which one is the biggest/smallest? (generic form: Among the existing clusters in the data set, which one is the biggest/smallest?) (possible answers: Subgroup 1 to 7). These two types are intended to measure the user’s ability to determine evenness of subgroups or equivalently, isolate subgroups/clusters that contain most and least members/objects using a representation (requirement R3).

Q3: In which attributes are the subgroups totally separated? (generic form: In which attributes are the clusters totally separated?) (possible answers: the attributes under investigation). The goal of this question type is to test if a representation technique supports users in isolating the attributes that totally separate subgroups/clusters and result in faultlines (or “cracks”) within a team (requirement R2).

Q4: To what extent are the subgroups separated across all attributes? (generic form: To what extent are the clusters separated across all attributes?) (possible answers: Very Weak, Somewhat Weak, Medium, Somewhat Strong, Very Strong). This question type is intended to gauge how well a user can interpret and assign a faultline level to a team using a visual representation (requirement R1). Within the scope of this study, the faultline level of a team is determined by the number of attributes in which the subgroups are totally separated. While this assessment does not consider attributes with partial separation of subgroups as the way the Fau measure (Equation 1) quantifies separation of subgroups, it makes answering this task question more straightforward to participants.

Q5: Between two different teams, which team has stronger separation of subgroups? (generic form: Between two different data sets, which one has stronger separation of clusters?) (possible answers: Team A or Team B). This last question type is intended to determine if a representation technique is discriminative enough to allow a user to compare the faultline levels of two teams depicted in two visualizations of the same technique (requirement R1).

In our user study, each of the question types was asked multiple times on different teams/data sets. We identified the best answers to the questions based on the distribution of members across subgroups and the attributes in which subgroups are separable. These constructs are achieved using our synthetic data generation procedure, which is described next.

Synthetic Team Data Generation. For the study, it is difficult to find real data sets that can serve as ground truth stimuli for the six types of questions. Therefore, we create work teams formed from automatically generated data sets. Technically, our method generates pre-clustered teams over a manually defined set of mixed-type demographic attributes, where team size, number of subgroups, evenness of subgroups, and separation of subgroups are controlled. The aim was to simulate teams with realistic distributions of members while controlling the faultlines and subgroup structure.

In our setting, we have one variable X for each attribute, and we hand-specify the categorical values or range of values for X. To generate a team, we specify its input parameters including the number of subgroups k, subgroup sizes $[n_i]_{i=1}^k$, and the set of attributes in which the subgroups are totally separated $[X_s]$. Note that $n = \sum n_i$ denotes the size of the entire team. For each $X_s$, we randomly partition its attribute space into k distinct subsets of values following a multinomial distribution and we draw randomly $n_i$ samples from each subset for each subgroup i. This guarantees that the subgroups are totally
Figure 4: Example team of size 18 visualized using (a) HIST, (b) PCP, and (c) SPLOM. Distinct colors are used to differentiate the three subgroups: subgroup 1, subgroup 2, and subgroup 3. While subgroup 3 is the biggest, subgroup 1 is the smallest. The three subgroups are totally separated along ETHNICITY, EDUCATION, and EXPERIENCE because different subgroups occupy different subsets of values along these attributes. The three subgroups overlap in GENDER and AGE because there exist values of these attributes shared by different subgroups. The faultline level is MEDIUM considering that the subgroups are totally separated in three out of five attributes.

For the rest of the attributes \(X_n\), we model the distribution over its possible values either as uniform or skewed distribution and we draw randomly \(n\) samples from each of these distributions. We choose these specific distributions based on the realistic distributions of the team demographics widely accepted in management literature: uniform distribution corresponds to diversity as variety and skewed (or relatively homogeneous) distribution corresponds to diversity as disparity [20]. For example, while both genders may be uniformly represented in some teams (e.g., student body), either male or female gender may be dominant in other teams (e.g., organizational groups). Once the samples are created for each of the attributes, we use the \(j^{th}\) sample for each attribute as the corresponding attribute value of the \(j^{th}\) member in the generated team. Finally, since the team is already clustered into subgroups, we simply use the Fau formula (quation 1) to calculate the faultline strength value for the team.

In our generated teams, team members or objects are char-
acterized by the following five independent demographic attributes. We chose these attributes because they are the most commonly used in faultline literature [3].

- GENDER: F or M
- AGE: 20-60, discretized for HIST by steps of 5 corresponding to a pre-defined rescale factor (Equation 1)
- ETHNICITY: T, U, V, W, X, Y, or Z
- EDUCATION (degree): A, B, C, D, or E
- EXPERIENCE (level): 0-9

While we believed the study participants would be familiar with these attributes, we used single-letter labels as values of categorical attributes, e.g., T, U, V, . . . for ETHNICITY, to prevent participants from associating their own knowledge of demographics, e.g., ethnic differences, into their answers.

We generated teams whose sizes range from four to 50 members and number of subgroups range from two to seven. The teams with at most 16 members were considered small teams. The teams with more than 16 and less than 50 members are considered large teams to simulate other workgroup settings such as online volunteer groups [6, 7].

Visual Representations. Figure 4 presents examples of stimulus materials of the three techniques under comparison. The design of HIST (without visual enhancements) was described in Section 4. In our design of PCP and SPLOM, we also use distinct color hues to differentiate subgroups. To prevent total occlusion due to data overlap, both PCP polylines and SPLOM dots are drawn at a constant opacity of 40% and 60%, respectively. The opacity encoding matches our PCP implementation with that of Holten and Van Wijk [14]. Furthermore, we employed a jittering technique [16] to alleviate data overlap issues in SPLOM.

The resolution of each image produced by the three techniques was 630 × 430 pixels. Each visualization image was accompanied by a subgroup color legend of 80 × 270 pixels. We chose these resolutions to ensure that visualization images would fit into a standard 1024 × 768 pixel screen without requiring any scrolling—the usable screen space for a web page is approximately 960 × 600 pixels.

5.2. Experiment Design and Implementation

Participants. Participants were recruited from Amazon’s Mechanical Turk (mTurk), a popular crowdsourcing Internet marketplace which has been shown to be a viable platform for graphical representation experiments [39]. The marketplace allows requesters to post jobs, also called Human Intelligence Tasks or HITs, for a large pool of users (or turkers) to consider and complete. Since mTurk is a world-wide marketplace, we targeted our participants specifically to those registered in the US with normal vision, at least 95% “approval” rating, and at least 100 tasks approved. After passing the color blindness qualification test hosted on the mTurk website, each participant visited our external study website, read an explanation of the research study (in lieu of a signed consent form), and was randomly assigned to a visualization technique. The qualification test, which is based on the Ishihara Color Test [40], was to detect and exclude interested individuals with color blindness.

In total, 57 participants completed the study or 19 for each visualization technique. They represented a diverse range of majors/occupations, gender, and ages (Figure 5). Most of them were unfamiliar with the field of InfoVis. In addition to the 57 participants, we excluded 10 participants who stopped at the beginning or in the middle of the study. These withdrawn participants are evenly distributed across the three techniques: 4 for HIST, 3 for PCP, and 3 for SPLOM.

Experiment Design and Procedure. We followed a randomized between subjects study design where the primary factor consisted of three levels (HIST, PCP, SPLOM). Each of the techniques was randomly assigned to each of the participants. We used a common collection of synthetic team data sets for each of the three visualization techniques.

A participant first completed a short tutorial that explained the technique. The tutorial included several baseline visualization examples of very strong, weak, and medium faultline levels. The participant then answered six task questions of each of the types described earlier: three for smaller teams and three for larger teams. During a question, the participant could access a visualization example with annotations highlighting various aspects of faultlines. Note that the questions of one type are the same, but each one is asked about visualizations of different data sets. The ordering of question types was randomized across participants, but all questions of the same type were asked as a block. The ordering of questions in each type was also randomized to avoid ordering effects, such as primacy and recency effects, among participants. In total, the number of task questions was 36.

Following the data collection approach in previous work...
In fact, there was convincing evidence for an effect. As Figure 6 indicates, participants answered Q1 questions consistent across task questions and data set sizes. Our results show that (1) across the tasks, users can effectively identify evenness of subgroups using the other two methods and (2) HIST performance holds across the tasks and the mean of response times across the questions of interest. These results confirm our initial hypothesis that HIST is more effective in supporting users in determining attributes of visualization on both error distance, F(2, 54) = 9.809, p = 0 and response time, F(2, 54) = 10.87, p = 0. Tukey’s HSD multiple comparison tests showed statistically significant differences between HIST and PCP as well as between HIST and SPLOM in terms of error distance (HIST−PCP = 0.001) and response time (HIST−SPLOM = 0.002). The results for error distance held consistent when small and large teams were analyzed separately. The results for Q2a were similar to Q2a’s, with participants tending to identify the smallest subgroup more accurately with HIST. With respect to error distance, Tukey’s HSD tests revealed convincing evidence for the difference in the two pairs of techniques (HIST−PCP = 0; HIST−SPLOM = 0). Interestingly, when we analyzed error distance data separately over small and large teams, the results held true for large teams only. With small teams, while we found a statistically significant difference between HIST and PCP (HIST−PCP = 0.014), there was no such evidence when comparing HIST and SPLOM (HIST−SPLOM = 0.890).

The results confirm our hypothesis that users would make better judgments about subgroup evenness with HIST than with SPLOM or PCP. Again, PCP is the least favorable choice for this task perhaps due to both occlusion caused by data overlap and visual clutter caused by large data sets. As the results suggest, data overlap also hurts SPLOM’s performance, especially when the task involved identifying the biggest subgroup in large teams. In contrast, participants using HIST produced consistent answers for both smallest and biggest subgroups and independent of the data set size.

Results for Q3. In which attributes are the subgroups totally separated? The results also favored HIST (Figure 6). We found statistically significant effects of visualization on both error distance, F(2, 54) = 17.58, p = 0 and response time, F(2, 54) = 12.15, p = 0. Tukey’s HSD tests yielded significant differences between HIST and PCP as well as between HIST and SPLOM on both error distance (HIST−PCP = 0; HIST−SPLOM = 0.001) and response time (HIST−PCP = 0.013; HIST−SPLOM = 0). The results held true when we analyzed error distances for small and large teams separately.

These results confirm our initial hypothesis that HIST is the most effective in supporting users in determining attributes in which subgroups/clusters are totally separated, followed by SPLOM and PCP. This finding is important considering that to the best of our knowledge, no previous work has explored the
Figure 6: Boxplots of mean of error distances (a) and of response times (b) for each question type as a function of visualization technique (HIST, PCP, and SPLOM).

6.1. Subjective Evaluation

After answering the task questions, participants also completed a short questionnaire requesting their thoughts on the visualization technique and their study experience. The questionnaire consisted of 10 Likert-style statements and three open-ended questions, which we adopted from [42, 25], as well as four NASA TLX questions [43] (see Appendix A).

We first discuss the results for the Likert-style statements and NASA TLX questions. Figure 7 presents the responses to each of the Likert-style statements from the samples of HIST, PCP, and SPLOM. Figure 8 presents the responses to each of the four NASA TLX questions from the samples of HIST, PCP, and SPLOM.

Figure 7: Boxplot of responses to each of 10 Likert-style statements as a function of visualization method (HIST, PCP, and SPLOM). The participants were asked to indicate their level of agreement on a scale of 1 (strongly disagree) to 5 (strongly agree).

Figure 8: Boxplot of responses to each of four NASA TLX questions as a function of visualization method (HIST, PCP, and SPLOM). The participants were asked to indicate the level on a scale of 1 (very low) to 10 (very high).
Participants (out of 19) praised HIST for its extensive analysis of the three open-ended questions. Overall, nine participants and subgroups (seven participants). However, four participants expressed concern about transparency of polylines, which are difficult to discern especially when they are of similar colors, e.g., red and orange. Seven participants also mentioned that the charts become extremely overwhelming for large data sets. Commenting on SPLOM, three participants (out of 19) liked the technique for its familiarity and ease of understanding. However, similar to PCP, five participants disliked the similar colors among dots. Additionally, ten participants requested bigger charts or the zoom-in ability. This confirmed our initial assessment that without interaction techniques [44], the matrix form space requirement of SPLOM is a limitation.

7. Faultlines Visualization Enhancement

To further facilitate the faultlines identification tasks, we incorporate computational analysis into HIST. The visual enhancement is inspired by the analogy between team faultlines and the physical layered form of geological faultlines, as first introduced by Lau and Murnighan [1], in a sense that team members’ multiple demographic attributes resemble multiple layers of the earth’s crust. We augment the representation with connecting dashed lines to indicate the holistic boundaries of existing separation among the subgroups across the attributes of interest or “layers” (Figures 2 and 9). To our knowledge, this visual enhancement is novel considering that while measures exist to detect separable clusters of quantitative data in 2D scatterplots [45, 15], measures and enhancements for mixed type data in stacked histograms are non-existent.

Technically, the augmentation requires three main computational procedures: (1) reordering attribute values in space, (2) identifying attributes with total subgroup separation, and (3) drawing the lines.

7.1. Reordering of Attribute Values

The first step is to reorder values within nominal attributes to reveal meaningful boundaries among subgroups along the corresponding axes. For each attribute \( X \), we first construct the corresponding contingency table (or matrix), \( A \), by subgroups. Second, we reorder attribute values or matrix rows by optimizing the Bertin Classification Criterion (BCC), as illustrated in Figure 10. The criterion, which is proposed by Pilhöfer et al. [17] and related to Kendall’s \( \tau \) [46], is an implementation of Bertin’s idea that reordering of data would improve the understandability of graphical displays [47]. The goal is to minimize

\[
BCC(X) = \sum_{i \neq j, j \neq f} A_{ij} A_{jf}
\]

where \( A_{ij} \) denotes the entry value at row \( i \) and column \( j \) and similarly, \( A_{ij} \) the entry value at row \( i' \) and column \( j' \). Note that optimization of \( BCC \) does not indicate whether total separation of subgroups/clusters occurs within an attribute. Also note that since we want to preserve the stacking order of subgroups, that is, subgroup 1–red followed by subgroup 2–blue and subgroup 3–green as in Figure 9, this first step optimizes \( BCC \) by re-arranging attribute values only, instead of both subgroups and attribute values.
7.2. Total Separation Criterion

The second step is to determine if total separation of subgroups/clusters occurs within an attribute $X$. Technically, if $X$ is a nominal attribute, total separation occurs when each row (attribute value) of the matrix is fully contained in exactly one column (subgroup). In other words, different subgroups share no common attribute values, or

$$R(X) = \sum_{i \neq j, x \neq y} A_{ij}A_{xy} = 0 \quad (3)$$

If $X$ is a numeric or ordinal attribute, total separation of subgroups further requires that rows fully contained in one specific column must be contiguous, or $BCC(X) = 0$, assuming the ordering of subgroups (or matrix columns) are optimized to achieve the pseudo-diagonal form of the matrix $[17]$. Combining the two requirements, total separation of subgroups occurs within an attribute $X$ when

$$TSC(X) = R(X) + \min(BCC(X)) = 0 \quad (4)$$

We refer to $TSC$ as Total Separation Criterion. Its values are also used to reorder attribute axes in ascending order from left to right (Figure 9) before executing the faultlines drawing algorithm. Note that for the purpose of computing $TSC$, this step simply calculates $BCC$ with different permutations of subgroups (or matrix columns), as opposed to actual re-arrangement of attribute values as in the first step.

7.3. Faultlines Drawing Algorithm

For each of the attributes with total separation of subgroups, since its values are already in optimal ordering after the first two steps, our algorithm simply traverses the values and marks the boundary between two adjacent subgroups. The traversal also wraps around the values to include the boundary between the two subgroups occupying the top and bottom values along the attribute axis. Finally, we draw a dashed polyline along the boundaries of the two specific subgroups across the attributes with total separation of subgroups. Note that values of nominal attributes without objects (zero-length bars) can be selectively excluded to adjoin boundaries among subgroups. We also apply a jittering technique to alleviate the possible overlap of vertical line segments (Figure 9). Our informal test indicates that realtime computation of the lines is reasonably fast on a typical desktop PC.

8. Discussion and Future Work

In this paper, we propose, design, and evaluate visualization solutions to a new and worthwhile domain-specific problem concerning separation and structure of multivariate clusters instantiated in the context of diversity faultlines in work teams. Like most studies, ours has limitations that we discuss here along with suggested directions for future work as well as implications of our work for other application domains that concern cluster structure and separation.

8.1. Study Design Issues

First, our study evaluated static visualizations only to first understand the merits and shortcomings of HIST, SPLOM, and PCP as standalone representations. Since the faultline concept is still new to end-users, e.g., managers, and no visualization solution exists, we must begin by understanding representation approaches that are linked to generic clustering. This decision was also made to keep the study implementation feasible in the online setting of mTurk. Future work will address the interactive capabilities of HIST. For example, interaction features can potentially allow users to configure their faultline requirements, such as faultline measures, attributes of interest, and rescale factors for each of the examined attributes.

Second, while we collected response time, we did not set a time limit for each question considering that the online setting of the study may be associated with more interruptions than in a lab setting. The online setting could be a factor that caused several unexpected outliers as shown in Figure 6(b). Nevertheless, these outliers were counter-balanced among the three visualizations techniques and we applied a rank transformation [41] to the data before performing statistical tests.

Third, faultlines visualization enhancement, such as reordering of attribute values and drawing of connecting dashed lines, also requires formal evaluation. Early feedback from our management researcher collaborators were highly positive—they praised the enhancement for its simplicity and usefulness. However, an interesting point was suggested regarding reordering of attribute values not only in nominal attributes—as currently implemented—but also in ordinal and discretized quantitative attributes. The aim would be to make the separation of subgroups along the dashed lines more clear-cut (i.e., no crossing lines) but at the expense of losing the information on the possible distance/gaps among subgroups in ordinal and quantitative attributes. A follow-up user study of such trade-off in design choices with target users such as managers would be a potential direction for future work.

8.2. Limitations of HIST

Multiple histograms also have limitations. First, the technique requires a discretization of quantitative attributes. Second, since the technique treats each attribute independently, it provides limited insight into the correlation between attributes, at least with the static representation. On the other hand, PCP is well-suited to showing the correlation between two neighboring attributes. To enable correlational analysis in HIST, we envision that PCP polylines can be selectively overlaid to allow the user
to inspect the relationship among attributes as well as individual objects. Alternatively, it would be informative to consider approaches that decouple the primary faultlines/subgroups view from a relationship view where correlations are shown, for example, in a scatter plot matrix.

With regard to scalability, HIST is scalable to the number of visualized objects. Nevertheless, like many other multivariate representation techniques, screen space is a limiting factor when the number of dimensions increases. As a remedy, the HIST design places histograms vertically side-by-side—as opposed to horizontally—to allow more attributes to fit in a wide-screen display as well as to facilitate the placement and reading of labels from left to right. Across the three examined techniques, scaling with the number of clusters or subgroups would also be limited by the number of distinct qualitative color hues perceivable to the human eye. Harrower and Brewer provides detailed guidelines for qualitative color schemes in their ColorBrewer paper and tool [34]. Besides color hues, future work would investigate other encodings (e.g., textures or patterns) or combination of encodings to improve scalability in terms of the number of clusters.

On a related note, implementing an interactive faultlines visualization would require efficient faultline measures as an external data clustering step. Nevertheless, to our knowledge, there are currently no well-established measures that would be scalable to large teams with multiple subgroups [3]. We suspect that modern cluster algorithms from the field of data mining such as Affinity Propagation [48] deserve further investigation for the faultline measurement challenge. These algorithms may prompt revision of HIST and other cluster representations.

8.3. Implications of Faultlines and HIST in Other Domains

To further show that the faultlines concept in management corresponds well to functional diversity in ecology, as mentioned in the Introduction section, we also apply the multiple linked stacked histograms representation (HIST) to visualizing the two groups/clusters of common moths and rare moths from the moth data set [49] (Figure 11). To some extent, the two groups are equivalent to functional groups in functional diversity. The results are encouraging. The visualization provides insights into the separation between the two groups with respect to species, genus, and family as well as food plant: common moths are mostly conifer-feeders and rare moths are mostly hardwood, herb, and grass-feeders (Figure 11). Readers interested in the ecological viewpoint are encouraged to refer to Petchey and Gaston [9, 10] and Ramette [50] for in-depth reviews of cluster analysis techniques for ecological and microbial diversity data, respectively.
Finally, we are also intrigued by the possibility of investigating faultlines in online collaborations and other social contexts such as political science. There have been studies of the effects of group diversity on productivity as well as member withdrawal behaviors among Wikipedia projects [7], however, the effects of attributes are studied one at a time. In future work, it will be informative to revisit the problem and investigate the effects of multiple attributes simultaneously with diversity faultlines as the primary measure. In addition, political science, which studies demographic diversity and how diversity relates to voting patterns and election results, opens another area that may benefit from faultline-based visualization.

9. Conclusion

We present the first study exploring the design space for graphical representation of team faultlines, a fundamental construct in management that shares many characteristics with clustering in computation. In doing so, we contribute (1) the novel application and refinement of existing stacked histograms technique to the faultlines visualization in particular and visual analysis of cluster structure and separation in multivariate mixed data in general, (2) a rigorous evaluation of the effectiveness of the proposed technique, (3) additional visual enhancements and metrics to further facilitate the faultlines identification tasks. To visualization researchers, the findings from our study suggest the need for revisiting cluster representations in general and investigating techniques for the important problem of faultlines in particular. To management researchers, our proposed visualization provides a useful means to conceptualize visually the output of faultlines measures, a requirement which is extremely difficult to achieve with a table-based assessment. We also hope the visualization will help bring the benefits of studying faultlines to more end-users such as managers or human resources departments.

Acknowledgments

The authors wish to thank Rob Hess, Crystal Ju, and Thuan Duong-Ba for valuable feedback on our user study. We would also like to thank the three anonymous reviewers for their constructive comments and suggestions.

Appendix A. Subjective Evaluation Questionnaire

After answering the task questions, participants also completed a short questionnaire requesting their thoughts on the visualization technique and their study experience. The questionnaire consisted of 10 Likert-style statements, four NASA TLX questions, and three open-ended questions:

- **L1.** I was able to identify the number of subgroups in a team using the chart.
- **L2.** I was able to identify the biggest/smallest subgroup in a team using the chart.

**References**


