"Show Me the Cracks in Our Teams": Visual Representations of Demographic Diversity Faultlines

Tuan Pham Ronald Metoyer *
School of EECS
Oregon State University

Katerina Bezrukova †
Department of Psychology
Santa Clara University

Chester Spell ‡
School of Business
Rutgers University

Abstract

We address the problem of visualizing demographic faultlines, a fundamental construct developed to understand dynamics among members in diverse workgroups. We propose a visual representation for this purpose that is based on multiple linked, stacked histograms in a parallel axis layout. In collaboration with management researchers, we evaluate our technique qualitatively using both synthetic data and real-world data of an empirical faultlines study.

1 Introduction

Teams or workgroups are core units that contribute to the success of organizations. Therefore, leveraging the benefits of teamwork while reducing negative outcomes associated with groups has been a central focus in organizational management [6]. For example, researchers are interested in understanding how the demographic diversity of workgroup members (e.g., age, gender, ethnicity, education) affects performance and outcome processes (e.g., productivity, creativity, collaboration). Researchers study the group charateristics not only within a single-attribute perspective (e.g., group ethnic diversity) but also as a complex composition of multiple demographic attributes that results in *faultlines*. For instance, faultlines may split a diverse project team into the two subgroups of two old male software engineers and two young female QA testers.

A common and traditional analysis approach to understanding faultlines and their effects follows three main steps [1]: (1) formulate initial hypotheses regarding faultlines then (2) collect demographic data of workgroups and other outcome factors and finally (3) use faultline metrics, such as [1] to measure the extent to which a given team is divided into relatively homogeneous subgroups, and run statistical tests to quantify the relationships between faultline measures and outcome variables. This confirmatory analysis approach relies on assumptions that initial hypotheses are comprehensive and statistical models and measures are well-defined. When these assumptions are not satisfied, the approach loses its utility due to the lack of support for direct exploration of the data. Furthermore, the use of many existing faultline measures [6], which assume different statistical constraints on the data, contributes to the fracture of the current approach. To our understanding, visualizations play a small role in the current process–static histograms may be used to show distributions of team members in separate attributes; static scatterplots or line charts may be used to present the relationships between faultline values and outcome variables, two variables at a time as in [1]. Little work has been done to develop representations that emphasize faultlines across multiple attributes.

We envision that a visual analysis tool that leverages faultline metrics with appropriate representation and interaction techniques would be a useful complement to the current approach (see Fig. 1). In particular, such a tool would allow researchers to explore their diversity data iteratively, inspect faultlines and subgroup structure of different workgroups, and test their effects on outcome factors quickly. Managers or human resources departments could also use the tool to inspect the dynamics of their workgroups based on faultlines and potentially re-arrange team members interactively to predict behavior and to improve team performance.

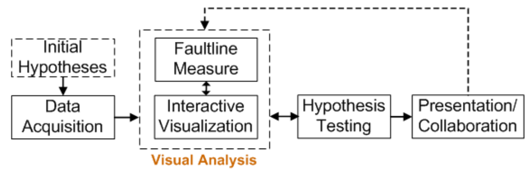

Figure 1: Proposed visualization-driven faultline analysis approach.

As a first step to achieve this vision, we propose a representation that aims to reveal demographic faultlines and subgroup structure in diverse workgroups. The representation is based on previous work to visualize diversity which utilizes multiple linked histograms arranged in a parallel coordinates layout [5, 3]. We augment this approach with histogram stacking and color encoding. While these techniques themselves are well-known, to our knowledge, their application to visualizing diversity faultlines is novel and the first to explore the design space for the problem. Furthermore, to design our representation, we have collaborated with management researchers, who helped validate its effectiveness.

2 Background

Measuring Group Faultlines. In many ways, measuring faultlines of a workgroup is equivalent to clustering analysis–that is, the measure classifies members of a group into subgroups (or clusters) according to their similarity across the attributes of interest (e.g. demographics). Clusters (or subgroups) have maximum internal homogeneity and maximum between-cluster heterogeneity.

Consider synthetic data of the two teams of six members each, as shown in Table 1. The visualizations presented in this poster computed group faultlines along three characteristics (age, gender, and ethnicity) with the measure proposed by Bezrukova *et al.* [1]. For each team, the measure does two things: first, it identifies the two subgroups (*Subgroup* column) and second, it quantifies the degree of alignment of attributes within the resulting subgroups (i.e., internal homogeneity) and the degree of difference between subgroups (i.e., between-cluster heterogeneity) with a faultline value (*Fau* column). The larger the faultline value, the stronger the split between

TeamID	Age	Gender	Ethnicity	*Subgroup*	*Fau*
1	44	f	A	*1*	*1.04*
1	18	m	B	*2*	
1	40	f	A	*1*	
1	33	f	D	*2*	
1	33	m	C	*2*	
1	50	f	B	*1*	
2	22	f	A	*1*	*2.17*
2	23	f	A	*1*	
2	39	m	B	*2*	
2	42	m	B	*2*	
2	57	m	C	*2*	
2	51	m	C	*2*	

Table 1: Synthetic data of the two teams of six members each. Faultline measure [1] clusters each of two teams into two subgroups (*Subgroup*) and identifies faultline values (*Fau*) for each of the teams.

*e-mail: [pham|metoyer]@eecs.oregonstate.edu
†e-mail: ybezrukova@scu.edu
‡e-mail: cspell@camden.rutgers.edu

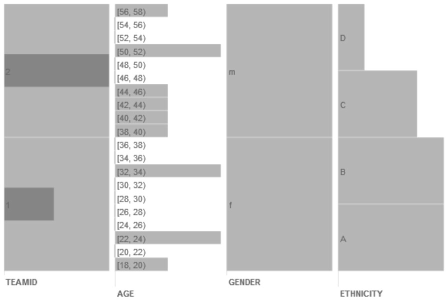

Figure 2: Representation of the entire synthetic workgroups data set (Table 1). Clicking on each of the TeamID's opens Figure 3. The visualization can be accessed at http://purl.oclc.org/faultlines/synthetic1.

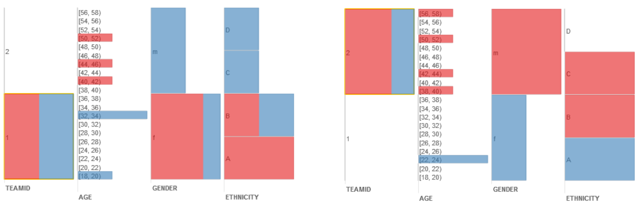

Figure 3: Representation of the subgroup structure of Team 1 (left view, *Fau* = 1.04) and Team 2 (right view, *Fau* = 2.17)

subgroups. Note that Table 1 does not clearly show where the splits (or "cracks") occur in a team and this is precisely what we address with our visual representation.

Representing Clusters. Thus far, besides Table 1, static scatterplots are a common technique to represent objects within clusters [2]. The technique is effective for showing structure of clusters over a small number of attributes. However, scatterplots may lead to ambiguous communication of abundance of objects due to occlusions caused by data overlap, especially with categorical attributes. Our proposed representation, which is based on histograms, takes a different approach of focusing on the distribution of objects in attribute space instead of focusing on object visibility. A survey of other multivariate representations of clusters can be found in [2].

3 Visualization Design

Design Requirements. Based on the requirements of management researchers, a faultlines visual representation should allow users to explore efficiently: (1) the faultline value (*Fau*) of a given team, (2) the inner structure (or distribution) of subgroups and the possible splits over the attributes of interest of a given team. In addition, the representation should scale well to the number of teams and the number of members in a team. Management researchers have been studied teams of up to 16 members that may potentially split up to 7 subgroups, depending on team size and number of attributes, yet they are also interested in teams of larger sizes (e.g., online volunteer groups).

Design Considerations. Previous work on diversity visualization [5] has shown multiple linked histograms are well suited to showing the distribution of objects over multiple attributes. Position and length are ranked highly for encoding nominal and quantitative values such as variety of attribute values and abundance of objects, respectively [4]. However, the previous work did not consider how distributions of multiple subgroups align over multiple attributes. Since subgroups are nested within a group, to maintain bar length encoding, a natural solution to encoding subgroups is to stack bars within each bin (Fig. 3). Then we use distinct color hues to differentiate stacked subgroups. Following this design, a split occurs at a particular attribute when distinct subgroups (or distinct colors) occupy distinct positions along the vertical axis. In addition, we arrang attribute axes in parallel [3] to facilitate a holistic overview of the subgroup structure and the possible splits over multiple attributes of interest. Note that while the faultline measure we chose [1] limits the number of subgroups to two, our choice of qualitative colors meets the requirement of encoding up to 7 (or more) subgroups. On another note, the length of each bin 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 quantitative attributes into bins.

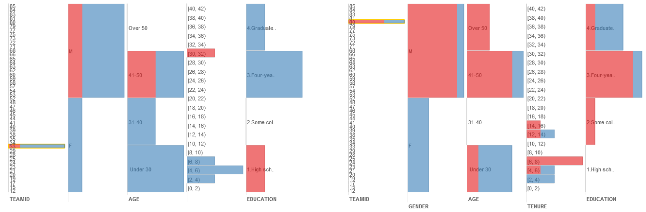

Figure 4: Representation of the subgroup structure of a team with strong faultlines (left view, Team 33) and a team with weak faultlines (right view, Team 80) from the faultlines study data set [1].

Visualization Prototype. Fig. 2 visualizes the entire synthetic data set (Table 1) with multiple linked histograms in a parallel axis layout. We overlaid additional bars on the TeamID axis to encode faultline values of each of the teams. This representation provides an overview how members of all teams are distributed across the attributes of interest and which teams have strong or weak faultlines. Clicking on each of the TeamID's opens Fig. 3, which shows subgroup structure of the two teams by stacking colored histograms of subgroups at each attribute axis. Red represents the first subgroup and blue represents the second one. While the two subgroups of Team 1 (left view, $Fau = 1.04$) are split only on the Age attribute (column 2), those of Team 2 (right view, $Fau = 2.17$) are split on both Age, Gender, and Ethnicity (columns 2, 3, and 4).

4 Evaluation and Discussions

We also applied the representation to a real world data set of an empirical faultlines study [1]. The data contains multiple teams with faultlines measured along four characteristics: gender, age, company tenure, and education (columns 2, 3, 4, 5 in Fig. 4). While we included Fig. 4 to demonstrate subgroup structure of only two teams with strong and weak faultlines, we encourage you to visit the visualization of the entire data set at http://purl.oclc.org/faultlines/study1. Management researchers found this view helpful in inspecting subgroup structure of different workgroups and in developing a sense of where the splits are likely to occur following their configuration of the faultline measures, prior to testing the faultlines effects (Fig. 1).

In future work, we plan to subject our method to a rigorous formal evaluation in which we compare our method to existing methods, such as scatterplot matrix and parallel coordinates [3, 2] through user studies. We also intend to develop this representation into a team building tool that enables users to build and re-arrange team members interactively based on faultlines.

References

[1] K. Bezrukova, C. Spell, and J. Perry. Violent splits or healthy divides? coping with injustice through faultlines. *Personnel Psychology*, 63(3):719–751, 2010.

[2] D. Holten and J. Van Wijk. Evaluation of cluster identification performance for different pcp variants. *Computer Graphics Forum*, 2010.

[3] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proc. IEEE Vis*, 1990.

[4] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2), 1986.

[5] T. Pham, R. Hess, C. Ju, E. Zhang, and R. Metoyer. Visualization of diversity in large multivariate data sets. *IEEE TVCG*, 2010.

[6] S. Thatcher and P. Patel. Group faultlines: A review, integration, and guide to future research. *Journal of Management*, 2011.