Environ Earth Sci
DOI 10.1007/s12665-014-3365-8

THEMATIC ISSUE

Toward exploratory analysis of diversity unified across fields of study: an information visualization approach

Tuan Pham · Julia Jones · Ronald Metoyer · Frederick Colwell

Received: 31 October 2013 / Accepted: 16 May 2014

Abstract The study of the diversity of multivariate objects shares common characteristics and goals across disciplines, including ecology and organizational management. Nevertheless, subject-matter experts have adopted somewhat separate diversity concepts and analysis techniques, limiting the potential for sharing and comparing across disciplines. Moreover, while large and complex diversity data may benefit from exploratory data analysis, most of the existing techniques emphasize confirmatory analysis based on statistical metrics and models. This work aims to bridge these gaps. First, by cross comparing the analyses of species diversity, microbial diversity, and workgroup diversity, we introduce a framework of diversity concerns aligned across the three areas. The alignment framework is validated and refined by feedback from subject-matter experts. Then, guided by the framework and theoretical information visualization and visual analytics principles (as distinguished from scientific visualization), we propose a unified taxonomy of common analytical tasks for exploration of diversity.

Keywords Information visualization · Species diversity · Microbial diversity · Workgroup diversity · Multivariate data · Visual Analytics

T. Pham (✉) · R. Metoyer
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon, USA
e-mail: pham@eecs.oregonstate.edu; tuanpham@live.com

J. Jones · F. Colwell
College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA

Introduction

Understanding diversity *patterns* and their causes and consequences (*processes*) is one of the greatest challenges in ecology, both at the scales of species such as plants and animals and of microorganisms, (e.g., Gunderson 2000; Magurran 2003; Ogunseitan 2005; Fuhrman 2009). This problem is critical because diversity is an important factor for the assessment of complex systems: changes in biodiversity may influence the stability and functioning of the ecosystem, (e.g., McCann 2000; Ives and Carpenter 2007). Although the problem is shared by other disciplines, ecologists might not be fully aware of the improvements potentially gained from understanding diversity studies in other arenas. For instance, researchers and managers of human organizations are concerned with diversity of work teams, (e.g., Lau and Murnighan 1998, Harrison and Klein 2007, and Bezrukova et al. 2009).

A common approach to understanding diversity patterns and processes is *hypothesis-driven* or *confirmatory* analysis that relies on rigorous statistical metrics and tests of data observations (Magurran 2003; Gotelli and Ellison 2004; Harrison and Klein 2007; Thatcher and Patel 2012). These techniques may work well when hypotheses are falsifiable and testable with reasonable metrics and tests. Otherwise, the utility of the current approach diminishes quickly when the number of diversity attributes under investigation is large, multiple subsets of data are involved, and/or hypotheses are not pre-established. Still, indices of diversity have greatly dominated over more direct exploration of diversity in studies of ecology and human organizations. In addition, discipline-specific metrics may preclude the understanding of how diversity functions and how it could be characterized similarly across disciplines.

Decades ago, Whittaker (1965), Sanders (1968), and Hurlbert (1971) suggested that in addition to diversity indices, ecologists should gauge diversity patterns by direct observation of data. Following this advice, visual representations of data such as histograms and rank-abundance plots (Whittaker 1965) have been employed to communicate species variety and abundance. Nevertheless, these techniques supported a limited number of variables, with no interaction, and thus limited exploration capacities—perhaps due to a lack of computational interfaces and tools at that time. Recently, experts who study human organizations have also suggested that configurations of work team structure are important and have direct consequences on team outcome processes (Carton and Cummings 2012). Yet no tools exist to enable direct investigation of team structure, besides text- or table-based assessment of data.

Recently, visual analytics, "the science of analytical reasoning facilitated by interactive visual interfaces" (Thomas and Cook 2006), offers a new, and powerful aid to the analytical reasoning of diversity patterns and processes in complex data. By leveraging the human visual system, visual analytics—a subfield of data visualization—provides a visual gateway to the data, complementing existing diversity metrics and allowing users to explore data directly and iteratively prior to further statistical analysis (Fig. 1). As distinguished from confirmatory analysis which is centered around hypothesis testing, data exploration facilitates the generation of hypotheses and insights into the data (Tukey 1977; Andrienko and Andrienko 2006).

The visualization community has shown considerable interest in interactive visualization tools for exploring diversity in ecology and its subfield—microbial ecology. Notably, there are tools designed to facilitate understanding of (1) patterns of species distributions in separate attributes, e.g., the EcoDATE tool (Pham et al. 2013), (2) structures of microbial populations, e.g., the MicrobiVis tool (Fernstad et al. 2011), and (3) taxonomic classification and structure, e.g., the TaxonTree tool (Lee et al. 2004). Unfortunately, these tools serve specific subsets of information needs that are somewhat separated and not transferrable from one to another. To our understanding, very little work has focused on abstracting diversity analyses from various fields to *unified analytical tasks* that target all facets of diversity in multivariate data sets. By analytical task, we mean one or a series of actions carried out by the target users on the data to fulfill an information need. Analytical tasks serve as prerequisites for designing visual-analysis tools that in turn support those tasks (Fig. 2).

This paper identifies and answers the following two research questions:

RQ1: How is analysis of diversity conceptualized and aligned across the multiple fields that study it? More

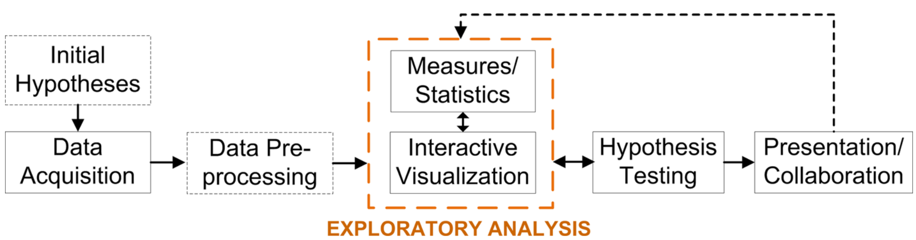

Fig. 1 Proposed visual-analysis process of exploring diversity data. Each *rectangle* indicates an analysis stage and each *arrow* indicates a path the analyst can take to navigate the stages. This work focuses on exploratory analysis tasks (the *orange rectangle*), as distinguished from data pre-processing or hypothesis testing tasks. Image redrawn from Pham et al. (2013)

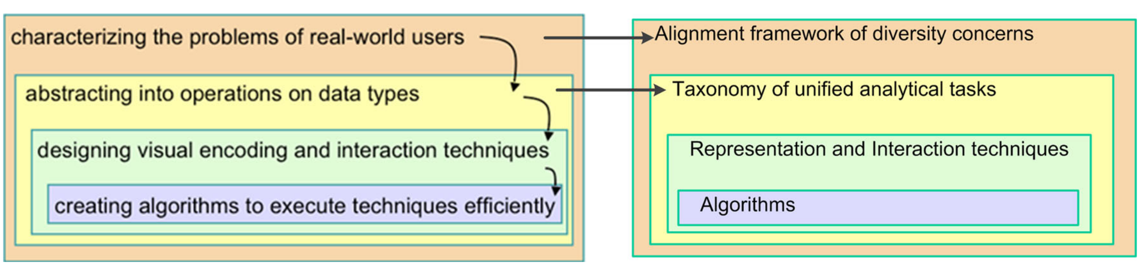

Fig. 2 A model of visualization creation with four nested layers introduced by Munzner (2009) (*left*) and its application in the context of diversity analysis (*right*). This paper emphasizes the two outer layers: (1) characterize the problem in terms of diversity concerns and information needs ("the framework", see "Alignment of diversity concerns") and (2) abstract the concerns into a list of common analytical tasks ("the taxonomy", see "A unified task taxonomy for exploratory diversity analysis") that can be accomplished with visual-analysis tools

Table 1 Alignment of diversity concerns across the analyses of species diversity (ecology), microbial diversity (microbial ecology and microbiology), and workgroup diversity (organizational management) summarized using terminology that is common to or distinct between the disciplines

	Species diversity	Microbial/genomic diversity	Workgroup diversity	Data behavior characterization
Typical unit of study	Community (α-diversity)	Microbe sample (α-diversity)	Work team	N/A
Typical Unit of Observation	Individual of known species or biomass	Operational taxonomic units with abundance (classified from microbe sample)	Individual person	N/A
Diversity components concerning separate attributes	Variety and abundance	Variety and abundance	Variety	Distributions
	Niche separation	–	Separation	Metrics
	Dominance/rarity	Dominance/rarity	Disparity	
Diversity components concerning interactions among attributes	Functional diversity	Functional diversity	Faultines/subgroups	Distributions Clusters Metrics
	Taxonomic diversity	Taxonomic diversity	–	Distributions Clusters Hierarchies Metrics
Diversity in space and time	β-diversity or turnover; γ-diversity	β-diversity or turnover	Between-unit diversity; Macro-faultlines	Spatial & Temporal characterization Metrics
Diversity as responder (cause of diversity)	Landscape patterns (climate, disturbance, land use)	Environmental patterns or biological patterns (human body)	Organizational factors (e.g., culture, recruitment)	Correlations/ Regressions Metrics
Diversity as driver or moderator (effect of diversity)	Ecosystem functions and processes	Eco or human system functions and processes	Workgroup functions and outcomes	

Table cells marked with "–" indicate missing concerns that may not yet be studied in the corresponding fields. The last column suggests how the data behavior for each of the concerns (if applicable) should be characterized

specifically, what are the fundamental scientific questions and hypotheses of interest regarding the diversity of multivariate objects?

RQ2: Given that analysis of diversity can be aligned across the fields (RQ1), which common analytical tasks are particularly useful in exploring diversity data?

In answering these questions, we draw upon our lessons from designing diversity visualizations (Pham et al. 2010, 2011, 2013, 2014) and provide two contributions: (1) an alignment framework of diversity concerns (RQ1)—the orange outermost layer in Fig. 2 and (2) a classification/ taxonomy of analytical tasks for exploratory analysis of diversity (RQ2)—the yellow layer in Fig. 2. Specifically, to answer RQ1, we review, cross compare, and align diversity concerns across the three areas of *species diversity* (ecology), *microbial diversity* (ecology/ microbiology), and *workgroup diversity* (organizational management). By *concerns*, we mean elements of diversity that can be conceptualized in a manner that transcends the three areas and the type of question being asked. The aim of the alignment framework is to set up a shared understanding between subject-matter experts and visualization researchers in terms of common diversity-related vocabulary and design considerations.

We also illustrate these concerns with several examples of commonly used visualization techniques. This work emphasizes techniques that apply to datasets where the objects of concern are described by abstract attributes that do not necessarily have a natural mapping to 2D or 3D space. Thus, this work falls under the field of Information Visualization (InfoVis) (Spence 2007) as opposed to Scientific Visualization (SciVis) that tends to deal with objects with physical properties such as surface location or density, that map naturally to 3D space. While the two areas increasingly overlap under the umbrella of Data Visualization (Weiskopf et al. 2006), SciVis techniques have been adopted by many related visualization work for environmental science, including visualization of geotechnical data (Orhan and Tosun 2010), of hydrology data and models (Rink et al. 2012; Velasco et al. 2013).

To address RQ2, we then translate these concerns into analytical tasks that are well defined by existing generic task taxonomies in visual analytics (e.g., Amar et al. 2005; Andrienko and Andrienko 2006). Simply put, while the diversity concerns are the vocabulary of subject-matter experts that represent their information needs and transcend disciplinary boundaries, the analytical tasks are the vocabulary of computer science, or more specifically, of visual analytics that represent user requirements that can be met by design of visual-analysis tools.

Our results aim to benefit various users. Subject-matter experts can cross compare diversity concerns and scientific findings as well as adopt analytical tasks and visualization techniques. Further, visualization designers and researchers have common vocabulary and abstractions for designing and evaluating different diversity visual-analysis tools. Finally, we are aware that the proposed framework and taxonomy are by no means comprehensive considering the complexity of ecological and human systems and their interactions. Therefore, we expect this work will stimulate further discussions regarding validation and improvement to both the framework and the taxonomy.

Alignment of diversity concerns

To answer RQ1, we propose a framework for aligning diversity concerns ("the framework") across the analyses of species diversity in ecology, microbial/genomic diversity in microbial ecology, and workgroup diversity in organizational management. By framework, we mean a set of thoughts, theories, and approaches that are accepted by subject-matter experts as the guiding principles for characterizing the problem. The concerns of interest include (1) characteristics of diversity data (see "Data characteristics"), (2) description of diversity patterns (see "Diversity patterns"), and (3) hypotheses regarding the causes and consequences of diversity (processes) (see "Diversity processes"). The framework is summarized in Table 1.

Data characteristics

Ecologists typically make a distinction between two types of phenomena concerning diversity: (1) the description of diversity (*diversity patterns*) and (2) the causes and consequences of diversity (*diversity processes*) (Magurran 2003). To understand these phenomena, a common approach is to undertake scientific studies. Specifically, experts collect data and make inferences about the underlying phenomena based on *data behaviors* (or data patterns). Data behavior is defined as a set of inherent features specific to a (sub)set of data observations considered *as a whole* as opposed to individual observations (Andrienko and Andrienko 2006). For instance, a data behavior may manifest itself as notions of distributions, clusters, or trends.

Diversity data are samples of independent *observations* collected from the population of interest within one or multiple *units of study* (Table 1, Rows 1 and 2). In workgroup diversity, a work team represents a typical unit of study while an individual person represents a unit of observation (or measurement) (Harrison and Klein 2007; Thatcher and Patel 2012). Comparatively, in species diversity, a typical unit of observation is an individual of a known species such as animals and plants collected in a community or assemblage (Magurran 2003). A typical unit of study of microbial community diversity is a biological sample (i.e., biological specimen) that contains multiple so-called Operational Taxonomic Units (OTUs), which are a close approximation to microbial species (as opposed to plant or animal species) with corresponding abundances (Ogunseitan 2005; Fuhrman 2009). The identification of OTUs is performed by extracting DNA from cells and then sequencing DNA from the biological sample (Fuhrman 2009).

Each unit of observation may be characterized by multiple mix-typed and, in some cases, hierarchical characteristics (*attributes*) necessary for gauging diversity of the corresponding unit of study and its role in the examined ecological or human system. For instance, a team member may be characterized by multiple demographic and non-demographic attributes; an individual of a known species, whether macrobiotic or microbiotic, may be described by multiple known characteristics, e.g., size, food type, and physiology, and hierarchical levels of Linnaean taxonomy, e.g., family, genus, and species. In addition, observations can be collected in space and time (*independent variables*) and associated with system process factors, e.g., team performance or ecosystem functions. In essence, diversity data sets are mix-typed, multivariate, and in many cases, hierarchical, spatiotemporal, and large (up to thousands of records/observations).

Diversity patterns

Diversity patterns are an overarching concept that includes various and related components adopted by the three areas of interest but usually under slightly different terms, especially between species/microbial diversity and workgroup diversity. The components can be loosely classified based on the ideas that (1) diversity is *attribute-specific*—that is, attributes are not treated as equal and (2) one or multiple diversity attributes can be investigated either *separately* (i.e., one by one) or *simultaneously* (Lau and Murnighan 1998; Magurran 2003; Harrison and Klein 2007). These two important ideas are captured in Table 1, Rows 3 and 4 and demonstrated in this section.

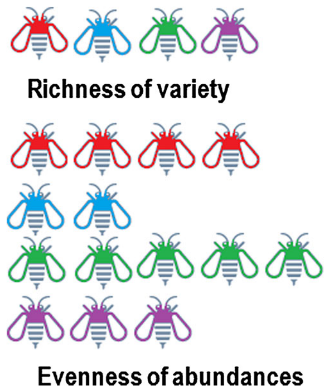

Fig. 3 Illustration of species richness and evenness. Each icon represents an individual of a known species (e.g., insects). Species richness refers to the number of different species represented in a unit of study and species evenness concerns the degree to which the respective species abundances are similar to one another, e.g., a highly even distribution has equal numbers of individuals of all represented species

Diversity patterns concerning separate attributes

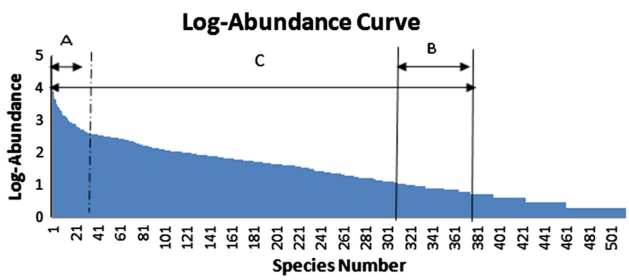

Fig. 4 Rank abundance curve (with logarithmic scale) showing the evenness of moth species in the moth dataset (Miller 2005). The technique, which is limited to a single attribute, is a variation of the histogram in which species are ranked from most to least abundant and then plotted along the *x*-axis. '*A*' shows the common moths, '*B*' shows the rare moths, and '*C*' shows the common through rare moths. Note that '*B*' excludes extremely rare moths because they do not provide enough information to identify the diversity and abundance of the respective moths. Image taken from Pham et al. (2011)

This section describes the common components of diversity concerning separate attributes, demonstrates how they are aligned across the three areas, and characterizes and gives examples of visualizations that depict data behaviors (Table 1, Row 3). For example, consider the investigation of biodiversity at species level only. Species diversity (or α-diversity) is "the variety and abundance of species in a defined unit of study", as defined by Magurran (2003). This definition emphasizes the two main components and corresponding metrics of *richness of variety* and *evenness of abundance* of species (Fig. 3). Similarly but at the genomic level, microbial community diversity also concerns variety and abundance of microorganisms in a community. With respect to data behaviors, in addition to diversity metrics, richness of variety and evenness of abundance are typically characterized by *the shapes of distribution*, as depicted by a rank-abundance curve in Fig. 4.

Similar to species diversity, a widely accepted definition of workgroup diversity in separate attributes is also centered on the generalization of *diversity as distributions*. The definition is described as "the distribution of differences among the members of a unit with respect to a common attribute, *X*, such as tenure, ethnicity, conscientiousness, task attitude, or pay" (Harrison and Klein 2007).

In addition, workgroup diversity is explicitly *attribute-specific*. Depending on the attributes under investigation, the experts conceptualize diversity not only as *variety* but also as *separation* and *disparity*, as introduced by Harrison and Klein (2007). Variety represents differences in kind or category, e.g., different skill sets, and reflects information in the unit. Separation represents differences in position or opinion and is considered a horizontal difference between members of a unit. For instance, different cultural values of members represent team separation. Disparity represents differences in concentration of valued social assets or resources and is considered a vertical difference between members of a unit. For example, difference in pay among members may create disparity in a team. Disparity thus reflects differences in possession.

While these diversity types have different names conceptually, from an analysis point of view, their patterns differ only in the *shapes of the distribution* of interest for minimum, moderate, and maximum diversity (Fig. 5). These shapes of distributions are in turn empirically associated with different outcomes for the examined unit of study (Harrison and Klein 2007).

Interestingly, ecologists also discuss *species dominance* and *niche separation*, which correspond well to disparity and separation in management if these components are considered in separate attributes. Species dominance refers to the degree to which a species is more numerous than others are or makes up (or possesses) more of the biomass, thus representing a vertical difference in makeup (as in disparity) (Begon et al. 2005; Gobet et al. 2010, 2011). Niche separation is the process of naturally partitioning competing species into different patterns of resource use or different niches so that they do not out-compete each other (Lawlor 1980). For instance, food type of animals could be considered as a separation attribute: carnivore (meat eater) and herbivore (plant eater) may represent two extreme ends of the food type spectrum; for microbes, it might be autotrophic metabolic lifestyle compared to a heterotrophic metabolic lifestyle.

In all, we argue that when diversity is considered in separate attributes, the concept of species diversity matches well with that of workgroup diversity in which team members equate to individuals of species (or their equivalents such as OTUs). These components are centered on

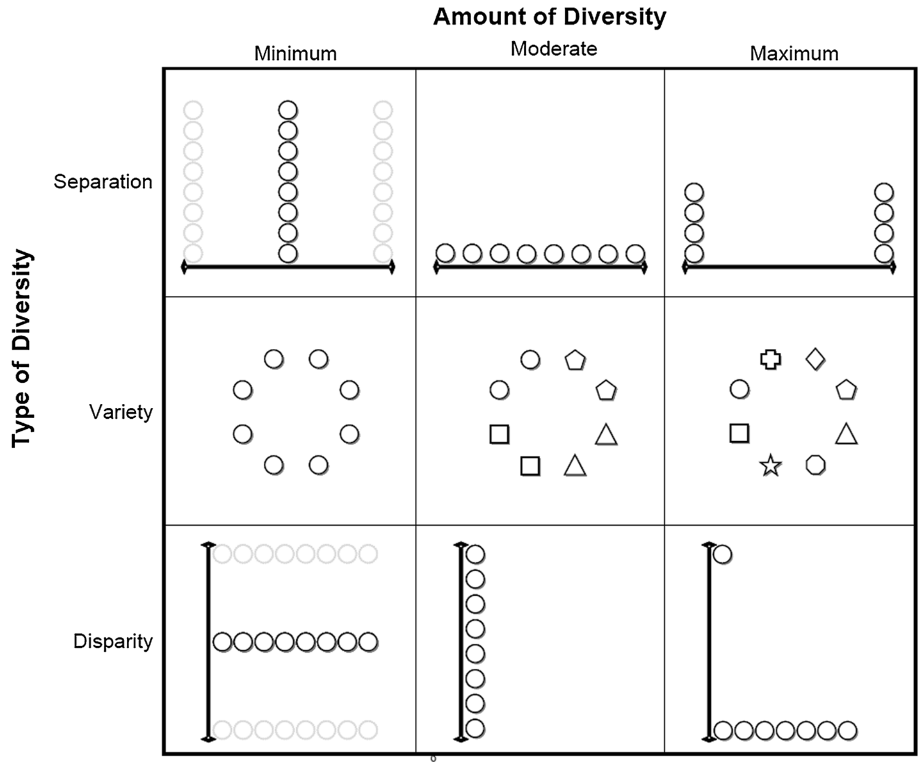

Fig. 5 Illustration of the three types of diversity within work teams and the corresponding shapes of distributions for the three levels of diversity: minimum, moderate, and maximum. Each of the icons represents a team member. Examples of distribution shapes include uniform distribution depicting maximum variety, bimodal distribution—maximum separation, and skewed distribution—maximum disparity (third column from the *left*). Image reused with permission from Harrison and Klein (2007). © 2007, Academy of Management

the generalization of *diversity as distributions*. Furthermore, it is important that the analysts choose the correct conceptualization, e.g. type of diversity, and apply the appropriate data characterization, e.g. statistical metrics or shapes of distribution. To summarize, we propose the following consideration for characterizing data behavior of diversity patterns in separate attributes:

> *Data behavior characterization–Consideration 1. From an analysis point of view, when diversity patterns are considered in separate attributes, depending on the types of diversity under consideration, e.g., variety, separation, and disparity, the corresponding data behaviors are typically characterized by the shapes of distributions of observations in separate attributes, in addition to summary statistics such as diversity metrics. If time and space are involved, the data behavior should also consider how the distribution patterns and summary statistics vary over time and space.*

To demonstrate how this consideration may benefit design of interactive visualization techniques, consider Fig. 6, which depicts the multiple histogram representation of the moth diversity and abundance data set supported by the EcoDATE tool (Pham et al. 2013). Consideration 1 emphasizes the characterization of data behavior as shapes of distributions in separate attributes. According to information visualization design principles (Mackinlay 1986), a histogram is well suited to showing the distribution of objects within an attribute. Further, placing histograms vertically side-by-side in parallel (Inselberg 2009) aims to convey a holistic object distribution over multiple attributes. Finally, the characterization of distributions (Consideration 1) is further assisted by interaction features. For instance, users can sort bins within a histogram by abundances to form the rank-abundance curve (e.g., green histogram LEP_NAME); annotate histograms with different colors to distinguish attributes of different diversity types (i.e., variety, separation, and disparity); subset data by time (COLLECT_YEAR) or space (TRAP_ID) to see how distribution patterns vary over time and space.

Diversity patterns concerning interactions among multiple attributes

Diversity definitions that look at the diversity of each attribute separately have a limitation. They do not take into account the *interaction among attributes*. Consider an example of two teams of employees that have four members each in Table 2. While it is obvious that Team 2 is divided into more subgroups, the current definition concludes that both teams are at the same level of overall diversity with respect to gender and age—that is, in each of the two teams, members are uniformly distributed in both gender and age. To address this limitation, here we discuss

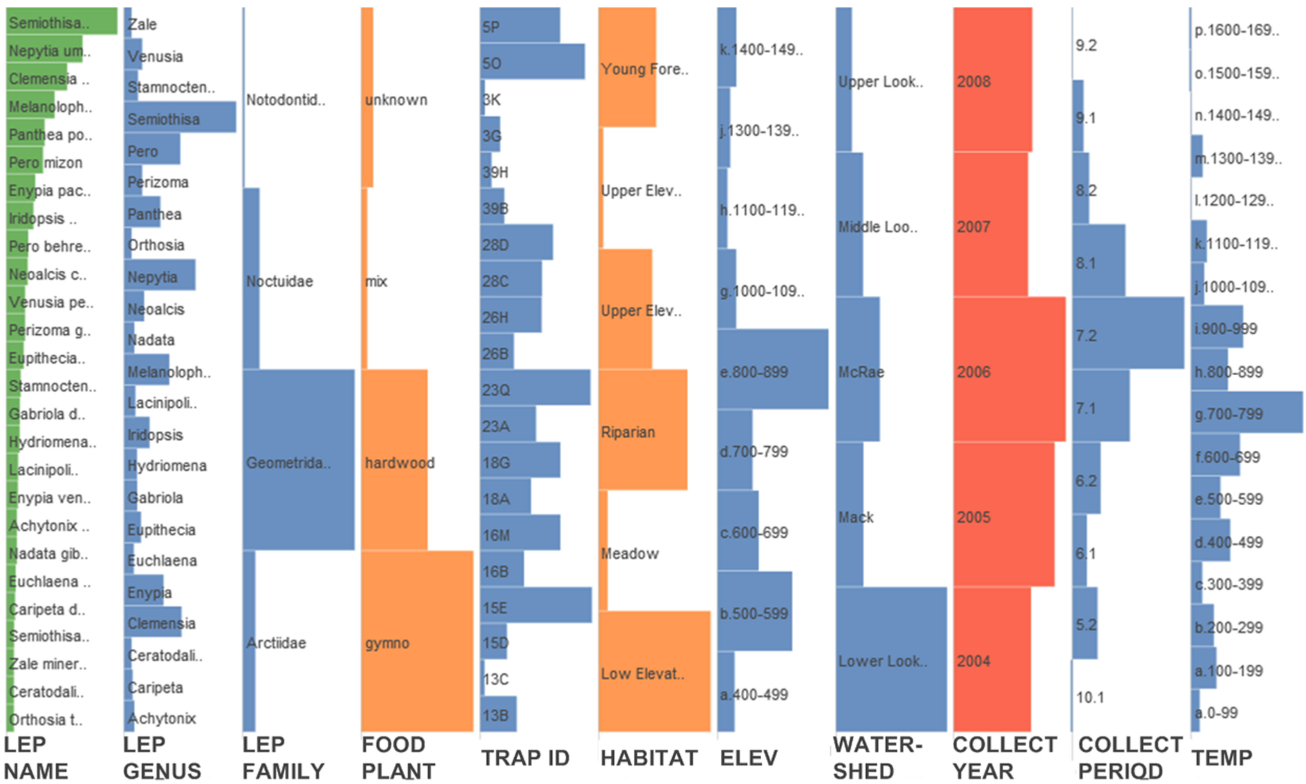

Fig. 6 The multiple histogram representation of common moths. The visualized attributes from *left* to *right* are LEP NAME (moth scientific name including genus and species), LEP GENUS, LEP FAMILY, FOOD PLANT, TRAP ID, HABITAT, ELEVATION, WATERSHED, COLLECT YEAR, COLLECT PERIOD, TEMPERATURE. Note that LEP is short for *Lepidoptera* (moth). In each of the histograms, the bars are pointing to the *right* (in contrast to the familiar upward-pointing display). The structure of the moth data set is described in Pham et al. (2011). The interactive version of the visualization is available at http://purl.oclc.org/ecodate/commonmoth

Table 2 Employee Diversity Example

Team 1	
Female, over 50	Male, under 50
Female, over 50	Male, under 50
Team 2	
Female, over 50	Male, over 50
Female, under 50	Male, under 50

Each of the two teams has four members

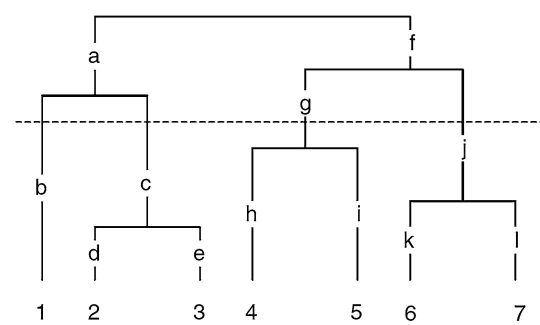

Fig. 7 A dendrogram representation demonstrating how seven species 1–7 are assigned to four functional groups based on hierarchical clustering of the species across multiple functional traits. The four functional groups include {1}, {2, 3}, {4, 5}, {6, 7}. The *dashed line* indicates an arbitrary stopping condition for the clustering process. Image reused with permission from Petchey and Gaston (2002). Copyright © 2002, John Wiley and Sons

diversity patterns that consider interactions among multiple attributes. In this regard, we also find parallel components across the three areas (Table 1, Row 4). This section describes and demonstrates how the common diversity components can be aligned and gives examples of visualizations that depict the corresponding data behaviors of interests.

Functional diversity is recognized by ecologists and microbiologists as different roles or functions played by species (or their equivalents) in communities and ecosystems. These roles can be determined based on *the composition of multiple functional traits* such as rooting depth and maximum growth rate of plants (Petchey and Gaston 2002). Technically, the ideas are (1) to cluster different species (or their equivalents) present in a unit of study into different *functional groups* based on composition of these traits, (2) to derive, for example, the functional diversity

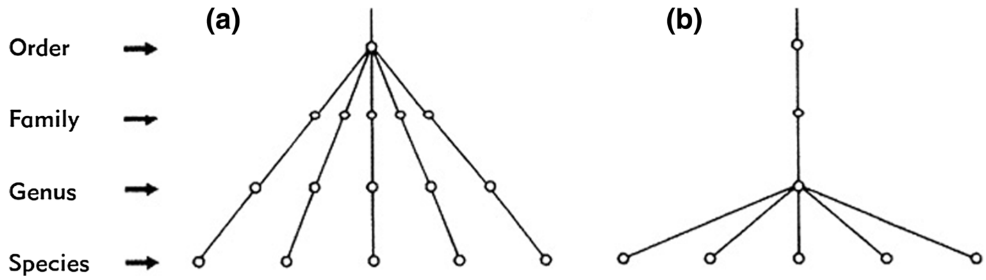

Fig. 8 A node-link diagram (tree) representation of two hypothetical units of study (e.g., assemblages) with the same level of species richness (i.e., five species represented) but different levels of taxonomic diversity when higher taxa such as genus and family are considered; unit of study (**a**) is more diverse than unit of study (**b**). Image reused with permission from Magurran (2003). Copyright © 2003, John Wiley and Sons

(FD) metric (Petchey and Gaston 2002) (Fig. 7), and (3) finally, to quantify and predict the associations between the functional diversity metric and other system processes. Ramette (2007) provides an in-depth review of cluster analysis techniques for microbial diversity data.

Furthermore, species and OTUs are inherently *hierarchical*—that is, species are grouped into taxa. Therefore, multiple traits or attributes under investigation might be extended to taxonomic organization such as *species*, *genus*, and *family*, resulting in *taxonomic diversity* (diversity across taxa) and corresponding metrics such as taxonomic distinctness (Warwick and Clarke 1998). Figure 8 illustrates an example of two hypothetical units of study whose diversity levels are determined by not only species level but also as composition of higher taxa. From an analysis perspective, the *hierarchy* of different species present is the primary data behavior of interest for taxonomic diversity.

It is important to note that functional diversity and taxonomic diversity also concern richness of variety and evenness of abundances within or between clusters, e.g., functional groups (Petchey and Gaston 2006), making the generalization of *diversity as distributions* still applicable. As an example, while the dendrogram alone in Fig. 7 does not consider evenness of observations in each of the four functional groups, a heatmap is commonly used along with a dendrogram to communicate evenness of abundances as demonstrated in Fig. 9.

Interestingly, in parallel with functional diversity in ecology, the *diversity faultlines* concept in organizational management, which is also derived from *multivariate clustering*, concerns *subgroups* or *clusters* formed in a work team based on alignment (or composition) of multiple demographic or non-demographic characteristics of members, as first introduced by Lau and Murnighan (1998). Figure 10 depicts an example of how the faultlines concept is applied to a work team. Just as ecologists studying functional diversity, management experts are also interested in (1) structure of subgroups with respect to the number of subgroups, evenness of subgroups, and subgroup variety and abundance; and (2) faultlines or attributes in which subgroups are separable or far apart from each other (Bezrukova et al. 2009; Carton and Cummings 2012; Pham et al. 2014). The goal is not necessarily to identify objects that cluster together but to determine how attribute space is divided or shared across the attributes of interest by clustered subgroups. Surveys of various faultline concerns and metrics can be found in Thatcher and Patel (2012) and Meyer and Glenz (2013).

In all, we argue that the concept of faultlines in organizational management could be matched with that of functional diversity in ecology from an analysis perspective. Both are derived from *multivariate cluster analysis*. Therefore, appropriate operationalizations of the concepts in terms of diversity metrics or data behaviors of interest could potentially be exchangeable. We summarize a consideration for characterizing data behavior of diversity patterns concerning interactions among multiple attributes as follows:

> *Data behavior characterization—Consideration 2. From an analysis point of view, when diversity patterns involve interaction among multiple attributes simultaneously, the corresponding data behaviors are typically characterized by the shapes of distributions of observations that are grouped into clusters across multiple attributes, in addition to summary statistics such as diversity metrics. The term cluster may refer a functional group of species in an ecological unit of study, e.g., communities, or a subgroup of people in an organizational unit of study, e.g., work teams; clusters may also represent different units of study under comparison. In addition, in some cases, the data behavior of interest is the hierarchical relationships if the patterns of interest concern taxonomic organization, e.g., taxonomic diversity, or hierarchical clustering. If time and space are involved, the data behavior should also consider how these distributions, clusters, and/or hierarchies as well as corresponding summary statistics vary over time and space.*

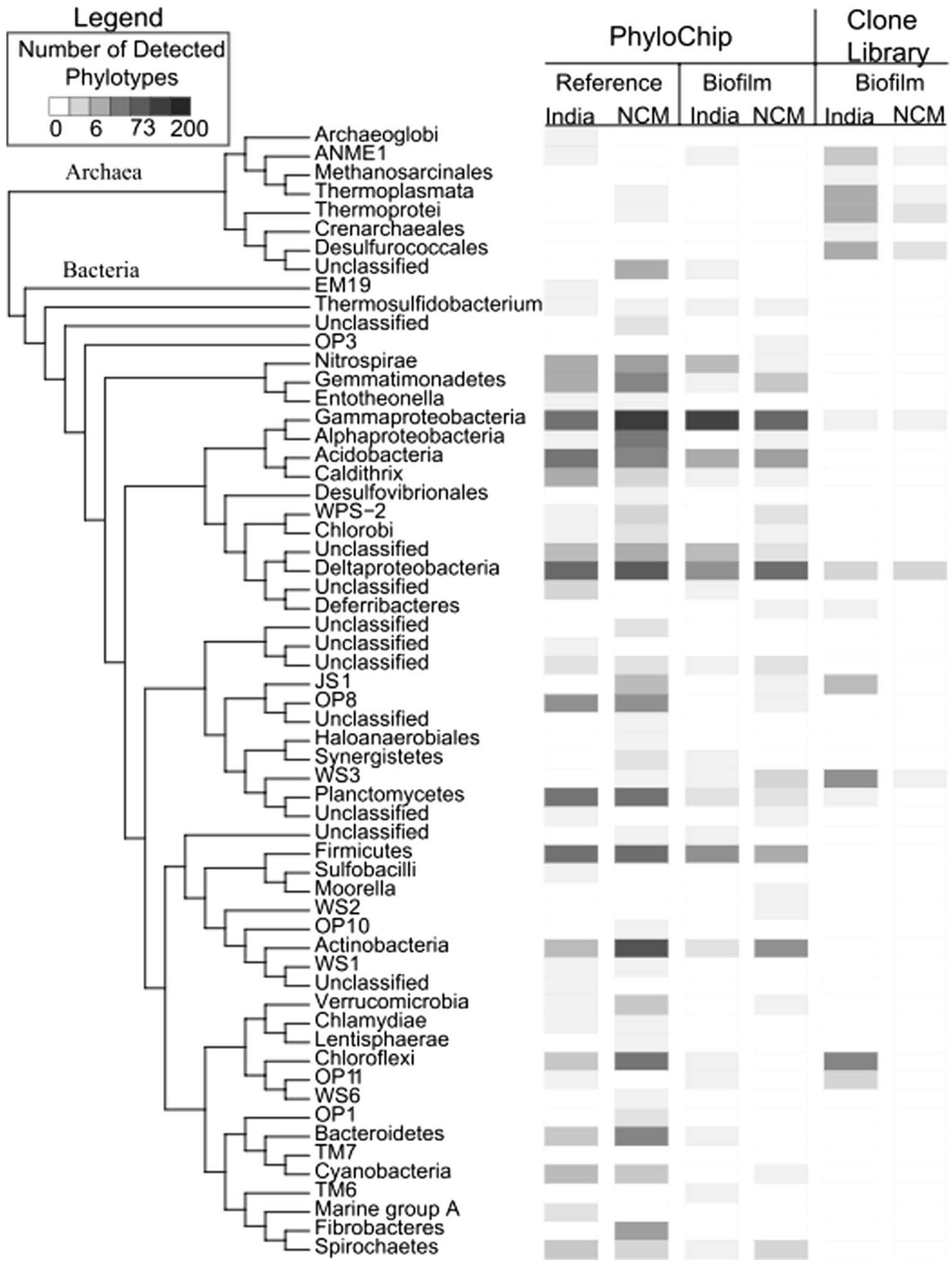

Fig. 9 A hybrid representation of dendrogram and heatmap used to depict the taxonomic diversity of archaeal and bacteria phyla along with corresponding abundances detected in several samples of a microbial diversity study. The term *phylotype* refers to an OTU that has been detected in a sample but for which there may be no microbe cultured. Image reused with permission from Briggs et al. (2011). Copyright © 2011, American Society for Microbiology

Figures 11 and 12 demonstrate a multiple linked stacked histogram design (HIST) of diversity faultlines in work teams that follow closely Consideration 2 (Pham et al. 2014). First, to depict distributions of observations across attributes, the design reuses multiple histograms (Fig. 6). Then, to help discern whether distributions of different subgroups overlap or separate along an attribute space, bars for each of the subgroups are stacked within each bin (Fig. 12). Finally, distinct color hues on a white background are used to differentiate stacked subgroups. Following this design, the holistic structure of each of the subgroups is conveyed across attributes. In addition, a total separation of subgroups at a nominal attribute is indicated by distinct subgroups (or distinct colors) occupying distinct positions along the vertical axis. At a numeric or ordinal attribute, total separation further demands that these

Player	AGE	COUNTRY	RACE	MLB TENURE	*Sub-group*	*Fau Metric*
A	29	USA	CAUCASIAN	8	*1*	*1.96 (Very Strong)*
B	33	USA	CAUCASIAN	14		
C	27	USA	AFRICAN-AMERICAN	8		
D	24	DOMINICAN/CARIBBEAN	FOREIGN	3	*2*	
E	22	DOMINICAN/CARIBBEAN	FOREIGN	2		

Fig. 10 An example of how a faultline metric (Bezrukova et al. 2009) is used to cluster a group of starting pitchers of an MLB team into two subgroups (subgroup 1 and subgroup 2) based on the similarity of group members across the attributes of interest: AGE, COUNTRY (of origin), RACE, and MLB TENURE (in years). The table does not clearly show how the subgroups (or clusters) are separable or far apart across the attributes under investigation. Figure 11 depicts a multivariate visualization technique that addresses this issue. Data courtesy of Katerina Bezrukova and Chester Spell

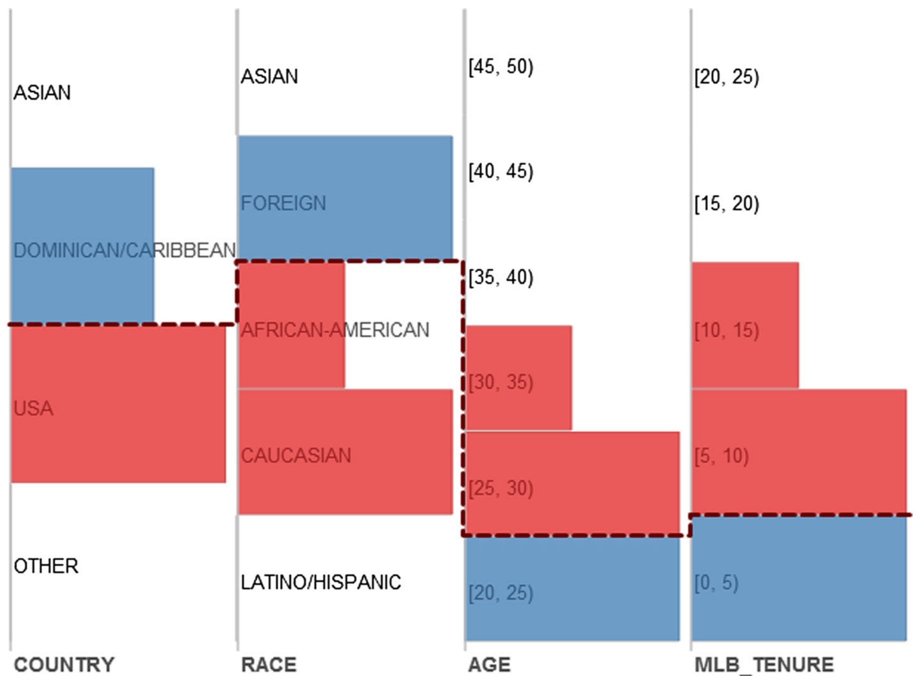

Fig. 11 Multiple linked stacked histograms (HIST) of the group of starting pitchers in Fig. 10. In all four attributes of COUNTRY, RACE, AGE, and MLB TENURE, the two subgroups (subgroup 1 and subgroup 2) are totally separated. In addition, the design automatically computes and draws connecting dashed lines to reveal the holistic separation between the two subgroups. Image reused with permission from Pham et al. (2014). Copyright © 2014, Elsevier

distinct positions—including ones without objects (zero-length bars)—are adjacent. The visual representation in Fig. 11 makes it obvious that the two subgroups formed in a group of baseball players are totally separated in all four attributes under investigation; the team visualized in Fig. 12 represents a less extreme example of faultline separation in a work team: the three subgroups are separated along several attributes but the members overlap in other attributes.

To further show that analyses of team faultlines and ecological functional diversity can be aligned, we also visualize the two groups of common moths and rare moths from the moth data set using HIST (Fig. 13). To some extent, the two groups of moths represent two functional groups in their respective communities. The visualization helps reveal the possible separation between the two groups with respect to species, genus, and family as well as food plant: while common moths are mostly conifer-feeders (i.e., *gymno*), rare moths are mostly hardwood and herb feeders (Fig. 13).

Diversity processes

Thus far, we have focused on diversity patterns; however these patterns are causally associated with other phenomena in the system under investigation. This section describes the component of diversity as cause or consequence, demonstrates how these are aligned across the

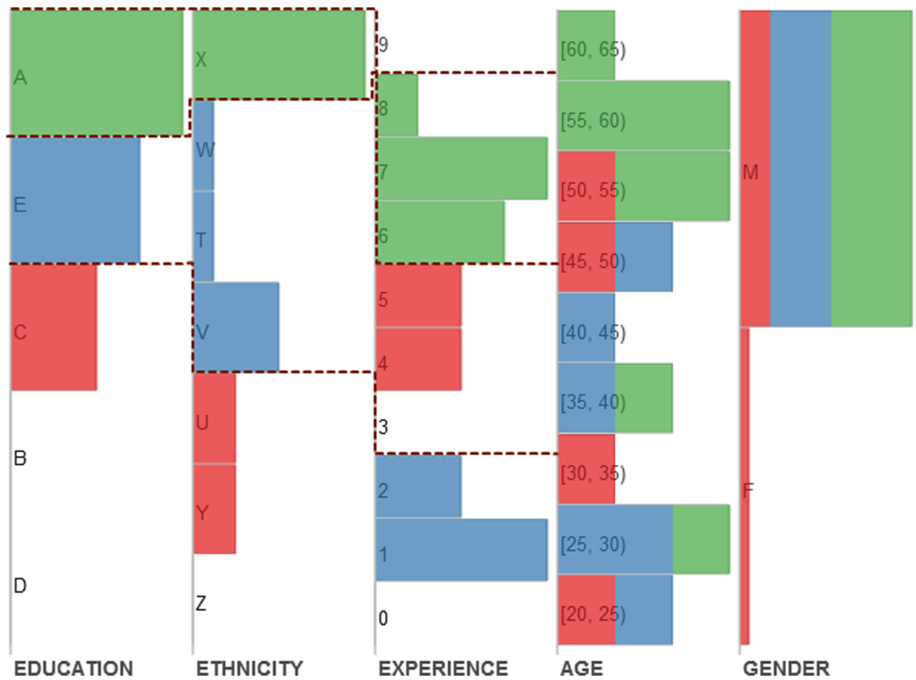

Fig. 12 Multiple linked stacked histograms (HIST) of an example team of 18 members clustered into three subgroups: subgroup 1 (smallest), subgroup 2, and subgroup 3 (biggest). Along ETHNICITY, EDUCATION, and EXPERIENCE, the three subgroups occupy different subsets of values; therefore, the subgroups are totally separated. Along GENDER and AGE, the three subgroups overlap. Image reused with permission from Pham et al. (2014). Copyright © 2014, Elsevier

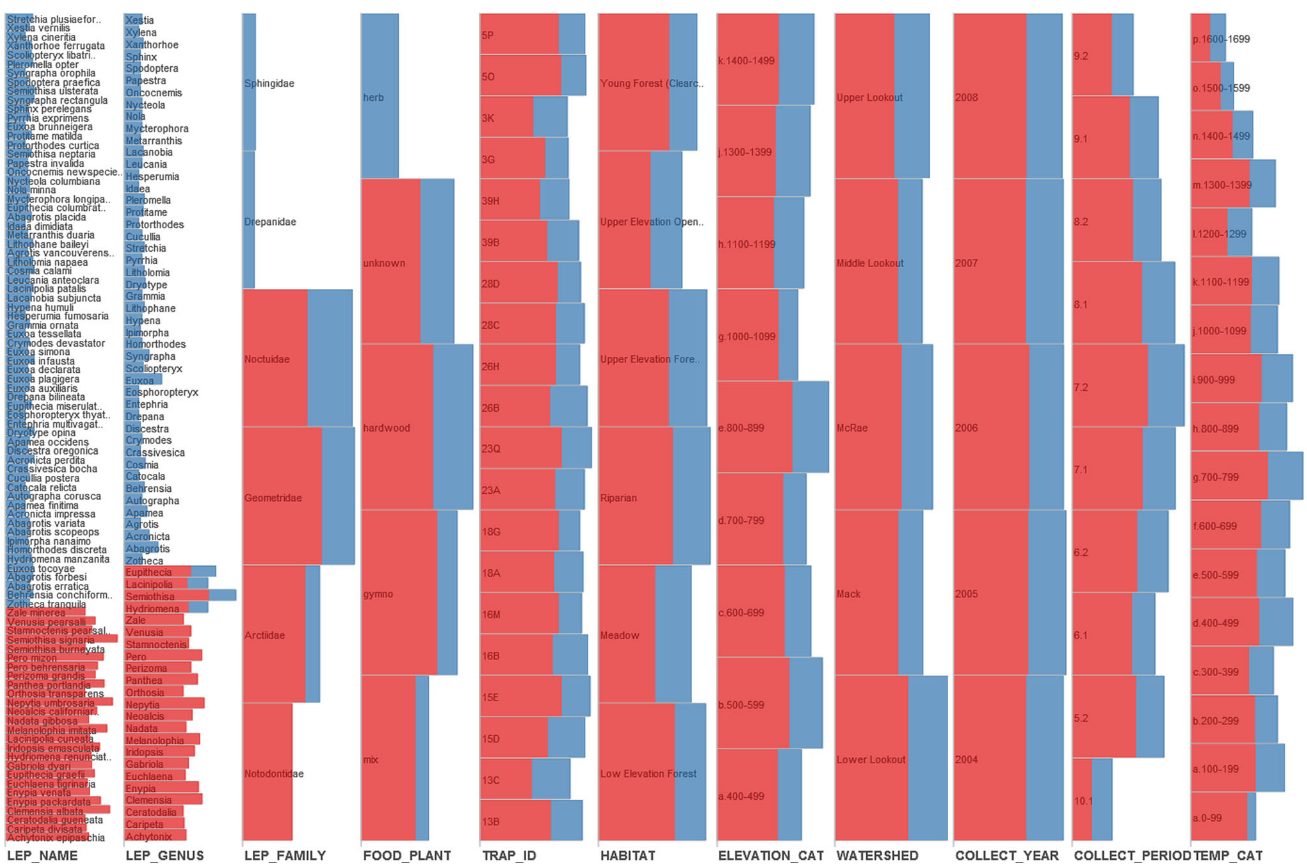

Fig. 13 Multiple linked stacked histograms (HIST) of two groups (or clusters) of common moths and rare moths. Bar lengths are scaled according the logarithm with base 10 because the common moths are significantly more abundant than the rare moths. The view helps ecologists hypothesize that the two groups may be functionally separated in terms of species, genus, and family as well as food plant—attribute axes 1–4 from *left* to *right*. However, the two groups overlap in the other attributes. The structure of the moth data set is described in Pham et al. (2011). Image reused with permission from Pham et al. (2014). Copyright © 2014, Elsevier

three areas, and characterizes and gives examples of visualizations that depict data behaviors (Table 1, Rows 6 and 7).

Across the three examined areas, we can find parallels in the roles of diversity as *responder* (cause), *driver* (effect), or *moderator* (effect). For instance, ecologists refer to positive effects of diversity such as sustainability and resilience in an ecological system (e.g., Gunderson 2000; Begon et al. 2005) while organizational management experts seek innovation and flexibility, just to name a few (e.g., Harrison and Klein 2007; Knippenberg et al. 2004; Meyer et al. 2011). The causes of diversity in ecology are related to climate, disturbance, and land use while in organizational management they are organizational factors such as culture or recruitment.

During the data exploration, there are two main approaches to making sense of diversity processes. First, the analyst may be able to make inferences about the diversity processes from direct observation of diversity patterns considering that the causal links are well understood (Andrienko and Andrienko 2006; Fernstad et al. 2011; Pham et al. 2011). For example, ecologists found that the richness of species tends to be higher in lower latitudes than in higher latitudes (Hillebrand 2004). Second, based on information needs of users and availability of process data (e.g., environmental factors or performance), visual analysis tools may support users in examining the associations between observed diversity patterns and system processes directly via *correlation and regression analyses*, before further statistical analysis. Note that regression and correlation indicate only how or to what extent data variables are associated with each other. To make conclusions about the causal relationships, analysists may need to involve their domain knowledge. We introduce another consideration for characterizing data behavior of diversity processes as follows:

> *Data behavior characterization—Consideration 3. From an analysis point of view, the data behaviors of diversity processes are typically characterized by how diversity patterns and system processes are correlated, if the observed diversity patterns are investigated as a driver or responder; or by how diversity patterns moderate correlations between system processes, if the observed diversity patterns are investigated as a moderator. If time and space are involved, the data behavior should also consider how these correlations/regressions vary over time and space. Note that following Consideration 1 and 2, diversity patterns and system processes may be characterized by corresponding data behaviors, e.g., distributions, clusters, hierarchies, or summary statistics, e.g., diversity metrics.*

To demonstrate the relevance of this consideration to designing visual representations, we present two examples from ecology and organizational management. According to information visualization guidelines, scatter plots and line charts are effective for communicating relationships between two variables (Seo and Shneiderman 2005). In Fig. 14, scatter plots are used to demonstrate possible *correlations* between measures of species richness, functional diversity, and ecosystem processes, e.g., retention of nitrogen, total aboveground biomass. In Fig. 15, a line chart is used to depict the role of diversity faultlines as "moderator". These two static examples, which are taken from research papers, serve the primary purpose of *explaining* the correlations and regressions found from hypothesis testing. Nevertheless, the techniques, if equipped with appropriate interaction features such as highlight/select and filter/subset (Heer and Shneiderman 2012), would be still applicable for enabling *exploration* of the correlations. On a related note, in both examples, the examined diversity patterns and system processes are quantified by summary statistics, as opposed to more descriptive data behaviors such as distributions or clusters, which is demonstrated in Pham et al. (2011).

Summary of the diversity concerns

In answering the first research question RQ1, we synthesize a variety of diversity concerns that represent information needs aligned across the three areas into a framework. There are two key points. First, exploratory analysis of diversity patterns aims to reveal the descriptive structure of multivariate objects of interest, e.g., species individuals, team members, in units of study of interest, e.g., communities, work teams. Such structure may manifest itself in the observed data as distributions, clusters, and/or hierarchies (Considerations 1 and 2). Second, exploration of diversity processes concerns the existence of the causal relationships between the diversity patterns and system processes. Such relationships are typically characterized by correlations and regressions among values of corresponding data variables (Consideration 3).

Moreover, conceptualization of diversity varies with different compositions of diversity attributes. Based on research questions of interest and data collected, it becomes very important that experts choose correct diversity concerns and apply the appropriate operationalization such as statistical metrics or visual representations of data behaviors such as those in Table 1, Column 5. The process could be iterative and exploratory (Fig. 1). We accompany each of the diversity concerns with selected examples of appropriate visualizations. By discussing these examples, we wish to emphasize how visualization design

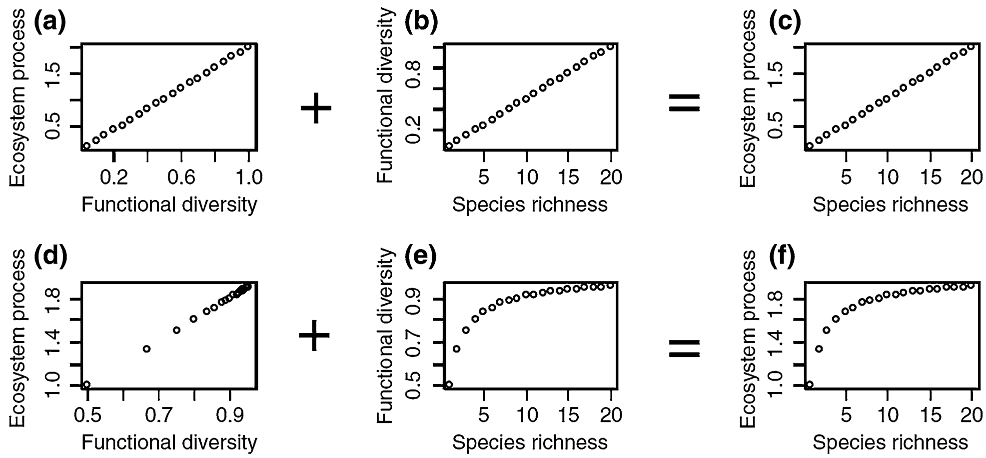

Fig. 14 An example of scatter plots used to illustrate possible relationships between species richness, functional diversity metric, and ecosystem processes. Data points may represent unique units of study or a unit of study repeatedly measured over time. The two rows (**a**, **b**, **c**) and (**d**, **e**, **f**) depict how the relationship between ecosystem process and species richness can be determined by a combination of the relationships between ecosystem process and functional diversity and between functional diversity and species richness. Image reused with permission from Petchey and Gaston (2006). Copyright © 2006, John Wiley and Sons

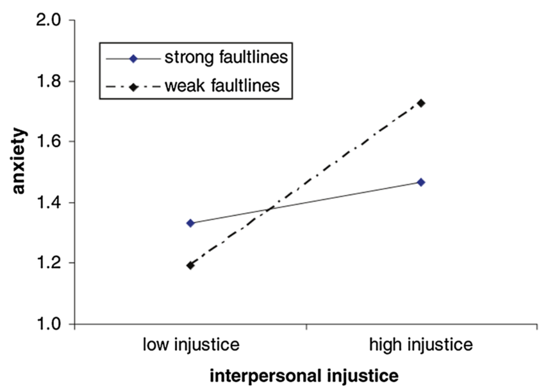

Fig. 15 An example of a line chart used to depict the role of diversity faultline as "moderator": psychological distress of team members was positively related to their perceived injustice in the team (the *dashed line*); strong group faultlines weakened that positive relationship (the *solid line*). Image reused with permission from Bezrukova et al. (2010). Copyright © 2010 Wiley Periodicals, Inc

should be guided by the information needs of users that can be abstracted into corresponding data behaviors.

Motivations for the taxonomy of analytical tasks. The alignment framework is only the first layer in the nested model of visualization creation process (Fig. 2). The framework does not yet illuminate possible *analysis processes* of these aligned diversity concerns. For instance, investigation of diversity patterns typically precedes that of diversity processes. In other scenarios, ecologists may wish to experiment with different combinations of functional traits and different clustering algorithms when investigating functional diversity; management researchers may wish to conceptualize the 'age' attribute as variety in one case (age comes with experience) and as separation in other cases (age represents generation gaps). These analytical tasks and processes represent user requirements that can be met by design of visual representation and interaction techniques (the third layer in Fig. 2).

Moreover, the framework, which is expressed in the vocabulary of the domains, e.g., richness, evenness, functional diversity, faultlines, must be translated into the vocabulary of visual analytics, e.g., characterize distribution, clusters, etc. (Fig. 2). The aim is to establish a shared understanding between subject-matter experts and visualization researchers.

Next, to answer research question RQ2, we review existing generic taxonomies of analytical tasks (see "Assessment of existing generic taxonomies of analytical tasks") and introduce a specific task taxonomy for diversity analysis unified across ecology and organizational management (see "A unified task taxonomy for exploratory diversity analysis").

Assessment of existing generic taxonomies of analytical tasks

Design of our taxonomy of analytical tasks was informed by existing generic task taxonomies in the fields of information visualization and visual analytics. In this section, we assess applicability of a subset of relevant taxonomies to diversity analysis. More thorough reviews of existing task taxonomies can be found in Amar et al. (2005) and Andrienko and Andrienko (2006).

To guide the design of information visualization tools, Shneiderman (1996) proposed the now well-known visual information seeking mantra "overview first, zoom and filter, then details on demand" followed by a classification of corresponding analytical tasks. The mantra and tasks are potentially useful to guide analysis strategies. Nevertheless, the tasks are somewhat driven by the tool capabilities, e.g., support of zoom and filter features, and there are no explicit mappings between the tasks and specific information needs in the context of diversity studies, e.g., what is the purpose of overview or filter?

Following a different approach based on user analytical activities when using visualization tools, Amar et al. (2005) introduced a taxonomy of ten *low-level* tasks (top 10 rows of Table 3). Applied to diversity analysis, these tasks, while not necessarily comprehensive, are relevant as building blocks since they aim to capture fundamental analytical operations, e.g., filter/subset, sort, characterize distribution. Nevertheless, to be more useful, the low-level operations need to be coupled with specific high-level information needs, e.g., which components of diversity require the tasks of characterizing distributions, hierarchies, and/or clusters?

Our proposed taxonomy is further motivated by the work of Andrienko and Andrienko (2006), who introduced a classification of tasks strongly based on information needs of analysts and tied closely to spatiotemporal data. In their framework, a task is defined as a *query* to find the unknown information (*target*) corresponding to the specified or known information (one or more *constraints*). The target represents data behaviors of interest such as distributions, clusters, and/or correlations fulfilled by one or many constraints such as population, space, and time. The classification, whose general outline is illustrated in Fig. 16, makes a distinction between the two classes of task: *elementary tasks*—which concern individual elements of data, e.g., what is the height of a given tree measured in a given date?, and *synoptic tasks*—which involves data behaviors in a set or subset of data *as a whole*, e.g., what is the shape of distribution of moth species caught in a given date and location? Synoptic tasks are further classified into *descriptive* (e.g., characterize distributions) and *connectional* (e.g., characterize correlation) tasks. Since the classification presents high-level analytical tasks, it can potentially serve as a generic framework for building a task taxonomy for field-specific needs like diversity analysis.

A unified task taxonomy for exploratory diversity analysis

While the generic task taxonomies do not necessarily consider or readily support specific tasks in diversity analysis, they serve as a framework and building blocks for our proposed unified task taxonomy (RQ2). In fact, our taxonomy offers an application, combination, and extension of the taxonomy of data-centric queries by Andrienko and Andrienko (2006) and the analytic low-level operations by Amar et al. (2005) in the context of a specific analysis. Figure 17 outlines our proposed taxonomy.

The taxonomy can be viewed at three levels of organization (or abstraction), representing the reasoning process of transforming information needs into knowledge and insights via analytical tasks. An information need starts in an abstract form of synopsis (*Generic Level*), then is realized with specific queries on diversity patterns and processes in the analyst's mind (*Data-Centric Level*), and finally can be achieved with low-level operations on appropriate analysis tools (*Analytic Low Level*). The following subsections describe each of the three levels.

Generic-level tasks

At the generic and also highest level, the taxonomy considers only synoptic tasks (Fig. 17, top orange level) as opposed to both elementary and synoptic tasks as in Andrienko and Andrienko's framework (Andrienko and Andrienko 2006) (Fig. 16). We made that decision based on the understanding that diversity patterns and processes concern behaviors of (sub)sets of observations *as a whole* as opposed to individual data elements (Table 1). While specific individual observations, for instance, rare or extreme observations, may be of interest to researchers, these observations are usually assessed in relation to other (sub)sets of observations and are still considered as a whole. Also, the value of a visualization typically lies in its capacity to uncover patterns or behaviors in data as a whole. Investigation of individual observations may be better served by raw tables coupled with database queries.

Data-centric queries

Decomposed from synoptic tasks, data-centric queries (Fig. 17, middle green level) encompass specific information needs regarding *building, detecting, or comparing diversity patterns and processes* as presented in the alignment framework (Table 1). To some extent, patterns and processes match the descriptive and connectional tasks in Andrienko and Andrienko's framework (Andrienko and Andrienko 2006) (Fig. 16). At this level, we also adopt their definition of task as *query*, which consists of two parts: one target (unknown information) and one or many constraints (known information).

Diversity patterns. These queries aim to gain knowledge into diversity patterns. The main objective is to characterize data behaviors (targets) as distributions, hierarchies, clusters, or summary statistics, following the three

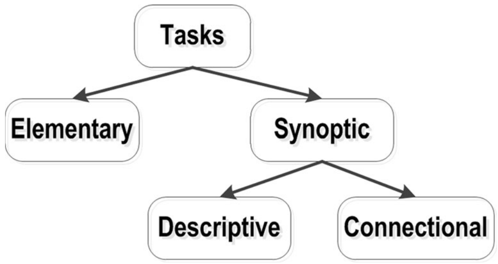

Fig. 16 General outline of the classification of analytical tasks proposed by Andrienko and Andrienko (2006). Image redrawn from Andrienko and Andrienko (2006)

considerations. The primary constraint is *population*, which is represented by collected samples of independent observations characterized by multiple attributes. In addition, data samples could be collected in the context of *space* and *time*, two additional secondary constraints. For example, information needs regarding functional diversity in ecology as well as faultlines in organizational management may involve building, detecting, or comparing distributions of clusters of data observations (targets) collected from a specific population—and in some cases—in space and time (constraints) (Consideration 2, see "Diversity patterns concerning interactions among multiple attributes").

Diversity processes. These queries examine the scientifically meaningful relationships between diversity patterns and system processes. Diversity patterns can play multiple roles in such causal relationships: diversity as driver, as responder, and/or as moderator (Table 1). These roles are characterized by the correlation between data behaviors of diversity patterns/metrics and of system processes. As an example, in ecology, as the name suggests, functional diversity, which is often characterized by statistical metrics or distribution of functional groups, is directly correlated with various ecosystem processes (Petchey and Gaston 2002). During exploratory analysis, the queries on diversity patterns usually serve as prerequisites for understanding diversity processes. This kind of "workflow" is represented as horizontal dashed arrows in the task taxonomy (Fig. 17).

Table 3 Ten low-level analytical tasks by Amar et al. (2005) followed by the three additional tasks of *Characterize Hierarchy*, *Annotate*, and *Fit Models/Metrics*. The tasks are described in the context of diversity analysis

Task	Description	Example
Retrieve value	Given a set of observations, find attributes of those observations	What is the tenure of a given player in a given baseball team (Fig. 10)?
Filter/subset	Given some concrete conditions on attribute values, find observations satisfying those conditions	What are the moth observations collected in HJA Forest in 2008?
Compute derived value/ Metric	Given a set of observations, compute an aggregate numeric representation of those observation	What is the faultline level of a given team (Fig. 10)?
Find extremum	Find data observations possessing an extreme value of an attribute over its range within the data set	What is the moth species with highest abundance (Fig. 4)?
Sort	Given a set of observations, rank them according to some ordinal metric	Sort the moth species observations by abundances (Fig. 4)
Determine range	Given a set of observations and an attribute of interest, find the span of values within the set	What is the age range of members in a given team (Fig. 11)?
Characterize distribution	Given a set of observations and an attribute of interest, describe the distribution of that attributes values over the set	What is the tenure distribution of members in given team (Fig. 11)?
Find anomalies	Identify any anomalies within a given set of observations with respect to a given relationship or expectation, e.g., statistical outliers	Are there any rare moth species (Fig. 4)?
Characterize clusters	Given a set of observations and multiple attributes of interest, find clusters of similar attribute values	Are there functional groups of trees with similar traits (Fig. 7)?
Correlate	Given a set of observations and two attributes, determine useful relationships between the values of those attributes	Is there a correlation between species richness, functional diversity, and ecosystem processes (Fig. 14)?
Characterize hierarchy	Given a set of data observations and hierarchy-based attributes, describe the hierarchical classification of the set over the attributes	What is the hierarchy of species in a unit of study (Fig. 8)?
Annotate	Note or distinguish among attributes or observations based on their common or user-defined characteristics	Annotate 'age' attribute as variety or as separation
Fit models/ metrics	Given a set of observations, fit a statistical or computational model to those observations—usually in the forms of visual indicators such as lines or colors	Fit a specific distribution curve to the data (i.e., dash line on the data histogram)

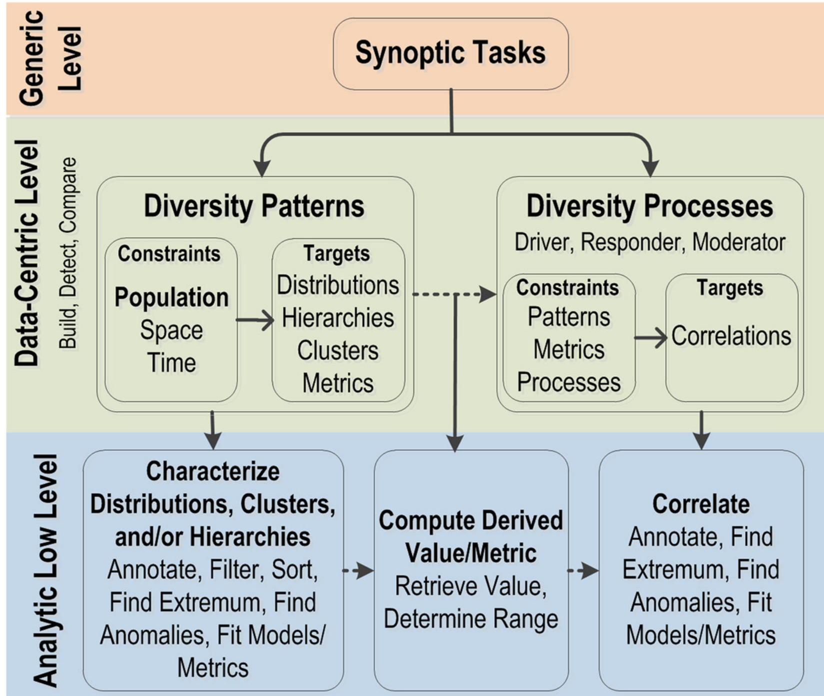

Fig. 17 Proposed task taxonomy for exploratory analysis of diversity organized at three levels of abstractions: (1) Generic Tasks, (2) Data-centric Queries, and (3) Low-level Analytical Operations. Vertical solid arrows represent how the tasks in an upper level can be mapped to one or many tasks in a lower level. Horizontal dashed arrows suggest the workflow between tasks within the same level

Low-level analytical operations

Data-centric queries in the analyst's mind are finally realized with low-level operations to be fulfilled by visual-analysis tools (Fig. 17, bottom blue level). All Amar et al.'s operations (Amar et al. 2005) described in Table 3 are relevant to diversity analysis. Note that secondary operations can be combined to accomplish a primary operation, which is denoted as bold texts in Fig. 17. For instance, characterizing distribution of a data subset may require filtering data first, and then sorting the data.

The ten original operations Amar et al. (2005) are not comprehensive. Guided by the alignment framework of diversity concerns, we introduce three additional low-level operations: *Characterize Hierarchy*, *Annotate*, and *Fit Models/Metrics* (Table 3). Hierarchy characterization is required when users inspect taxonomic diversity of species (or their equivalents) (Consideration 2, see "Diversity patterns concerning interactions among multiple attributes"). Annotation is useful when analysts wish to note or distinguish among attributes or observations based on their common or user-defined characteristics (Consideration 1, see "Diversity patterns concerning separate attributes"). For example, management researchers may wish to annotate the 'age' attribute as *variety* in one case and as *separation* in other cases. Fitting Models/Metrics represents a scenario in which analysts, given a set of observations, may want to fit a statistical or computational model to those observations—usually in the forms of visual indicators. For example, they may want to fit a straight line to a set of observations in a scatter plot to represent linear correlation (Fig. 15); they may want to see some visual indicator to represent separation among clusters of observations (Figs. 11, 12).

Summary of the task taxonomy

Guided by the alignment framework and existing generic task taxonomies, our proposed taxonomy aims to capture all possible queries and operations in the process of exploring diversity data. The reasoning process of the analyst may start with high-level queries on scientific phenomena such as "what are the functional diversity patterns?", followed by low-level analytical operations such as "characterize clusters of the observed data". The data behavior characterization in turn enables the analyst to understand and make inferences about the underlying scientific phenomena. Understanding the reasoning process as well as the specific queries and operations on diversity data is a critical requirement for the design of visual-analysis tools.

Discussion

This work presents the first cross-disciplinary synthesis study targeting exploratory analysis of diversity. Our study provides two contributions: (1) understanding of the diversity concerns aligned across the analyses of macrobiotic species, microbial taxa, and workgroup diversity (RQ1) and (2) a unified taxonomy of analytical tasks guiding the design of visual-analysis tools to address these concerns

(RQ2). Here we extend our discussion on (1) validation and refinement of the alignment framework with subject-matter experts, (2) limitations and future work, and (3) implications for diversity studies and design of visualizations.

Formative evaluation of the alignment framework with subject-matter experts

Feedback from experts is critical to ensure the alignment framework fulfills its intended purpose of characterizing the diversity analysis problem (RQ1). Our formative evaluation of the framework consists of two phases: (1) a pilot phase with our two domain expert collaborators and (2) a survey study with other external experts. To stimulate our discussion, we adopt the following feedback criteria (Ahn et al. 2013) (Table 4): comprehensiveness, ease of use, precision, usefulness, discoverability, and alignability of the framework.

Table 4 Criteria and corresponding questions for validation and refinement of the alignment framework of diversity concerns

Feedback Criterion	Question
Comprehensiveness	Are any concerns missing from the framework?
Ease of use	Is the framework easy to understand?
Precision	Does the framework describe precisely the concerns and the corresponding data behaviors?
Usefulness	Can the framework be used by experts to organize and cross compare their studies?
Discoverability	Does the framework help experts discover new concerns they had not thought of?
Alignability	Would the experts think concerns could be aligned across the three fields of interest?

The criteria are adopted from Ahn et al. (2013)

Pilot feedback. In developing the alignment framework, we have set up multiple face-to-face and email discussions between two visualization researchers and two domain expert collaborators, who co-author this paper: one ecologist and one microbiologist/microbial ecologist. The aims are to understand their information needs and to collect feedback on early thoughts on the framework before a full survey study.

Our ecologist (Jones) was instrumental in helping validate the analysis of species diversity as well as refine the overall framework vocabulary and presentation. Discussing the framework's comprehensiveness, she pointed out niche separation and dominance in ecology as potentially parallel concepts to separation and disparity in organizational management, respectively. With regards to the usefulness criterion, she requested compelling examples to demonstrate the operationalizations of diversity concepts as well as their alignment across the three areas. We responded with examples of visualization and introduced the three considerations for data behavior characterization. Interestingly, after seeing how multiple stacked histograms are used to communicate subgroups and faultlines in organizational management (Fig. 11), the ecologist immediately requested the same chart for comparison of the structure of different groups of moths from the moth data set (Fig. 13). We take that request as a positive sign that the framework helped the ecologist discover new diversity concerns she had not thought of, such as separation between clusters of observations or functional groups.

The discussions with our microbiologist/microbial ecologist (Colwell) suggested that between microbial diversity and macrobiotic species diversity, while there are some parallels in analysis approach, there also exist distinctions in information needs and characteristics of diversity data. Specifically, microbial diversity analysis emphasizes exploration of genomic information to identify previously unknown microorganisms and ultimately, to understand their functionality. The classification is usually performed in the data pre-processing stage (Fig. 1), using DNA extracting and sequencing (Sogin et al. 2006; Fuhrman 2009). Microbial genomic information is often rich in terms of representative OTUs but can be limited in terms of the number of biological samples and the number of attributes (e.g., spatial and temporal) because in many conditions such as subsurface environments, sampling remains a challenge. However, surface microbial communities such as in soils, waters, humans, and animals can be sampled much more frequently as in the MicrobiVis example (Fernstad et al. 2011). In addition, the costs of genomic analyses have decreased dramatically, making it possible to analyze more samples. On the other hand, species diversity deals with already known species and their well-understood characteristics, e.g., taxonomic classification, food types, habitats, so its analysis emphasis is really on the diversity patterns of multiple observations and their causes and consequences, providing that ecologists have access to larger number of observations and other environmental factors.

In all, our microbiologist assessed that the alignment framework was useful for cross-comparison of diversity studies. It helped him discover new diversity concerns such as diversity faultlines and corresponding techniques such as multiple stacked histograms. It is also encouraging to hear his comment that the future of microbiology would benefit from a similar species diversity analysis, and essentially from the alignment framework, providing that microorganisms are well classified and more data replicates are available. He also recommended related work on microbial diversity that we reference in this work.

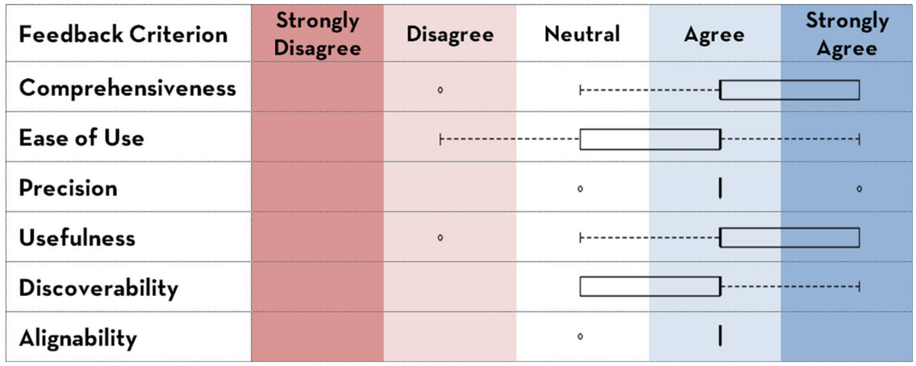

Fig. 18 *Boxplot* of responses from nine domain-experts to each of the six Likert-style feedback criteria/statements (Table 4). The experts were asked to indicate their level of agreement on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree)

Survey study with other experts. After the pilot phase, we further evaluated the framework in a survey study involving nine domain experts whose expertise was in species diversity (four), both species and microbial diversity (one), and workgroup diversity (four). All of them, who authored published research work cited in this paper, volunteered to participate in the survey in response to our emails soliciting their feedback. They answered the survey after reading a technical report presenting the framework.

The evaluation survey consisted of six Likert-style statements (Table 4), in which the experts were asked to indicate their level of agreement on a scale of one (Strongly Disagree) to five (Strongly Agree), and two open-ended questions: (1) if you disagree with any of the above statements, please explain your reason and (2) please comment on any aspects concerning the framework or the technical report.

The survey results were encouraging (Fig. 18). Most of the experts strongly agreed or agreed on the framework's comprehensiveness (seven out of nine), ease of use (six), precision (seven), usefulness (seven), discoverability (six), and alignability (seven). Several experts expressed their enthusiasm for the work, especially its novelty, necessity, and timeliness: "I really like how you bring together three so different disciplines in the first cross-disciplinary synthesis study about diversity";. "It's [the framework] looking great! I'm so happy that you're tackling this challenge—it's sorely needed"; and "The subject is also very timely".

Nevertheless, some experts also pointed out several limitations of the work. With respect to comprehensiveness of the framework, one mentioned the lack of diversity components concerning data acquisition and pre-processing, which we elaborate in the next subsection. Commenting on the role of visual exploration in diversity analysis, one expressed concern about the issue with post-hoc analysis—that is, the use of visualization to look for patterns that were not specified a priori. We respond to that comment that visual exploration, which is only part of a larger analysis process (Fig. 1), may prompt further statistical tests (that take into account post hoc analysis) or graphical inference tests (Wickham et al. 2010), additional data collection, and experiments. We also argue that traditional statistical tests may not be able to uncover unexpected data behaviors such as shapes of distribution, outliers, or separation of clusters of interest to diversity analysis. Complementing statistics, visualizations are particularly effective for those tasks.

In addition to critical comments, the experts also offered suggestions for improvement. One pointed out related domains that may share common diversity analysis such as community detection among social networks and additional analysis techniques such as Bayesian approaches for workgroup data. Several of them suggested other related work as well as minor changes for the terms used in Table 1, such as correlation vs. regression and taxonomy vs. ontology vs. typology. We considered them carefully, followed up with the corresponding experts via email if necessary, and incorporated them into the framework.

Finally, in addition to the interaction with domain experts, this manuscript draws upon our lessons from designing and evaluating diversity visualizations (Pham et al. 2010, 2011, 2013, 2014). Our previous publications describe and discuss our in-depth and long-term interdisciplinary collaboration processes in more detail.

Limitations and future work

This work emphasizes the exploration stage of the analysis process (i.e., hypothesis generation), following data acquisition and pre-processing stages and preceding further hypothesis testing, as illustrated in Fig. 1. Other stages may involve additional diversity concerns and corresponding analytical tasks. For example, microbe samples could be pre-classified into OTUs at different taxonomic levels using the Ribosomal Database Project (RDP) (Cole et al. 2009) and the process could benefit from dedicated analytical tasks such as dimensionality reduction using principal component analysis (PCA) and Nonmetric multidimensional scaling (NMDS) (Ramette 2007). In another example, data acquisition (or sampling) plays a critical role because it affects diversity patterns and processes. Species richness, for instance, tends to increase when the number of samples increases (Magurran 2003). The dependence of species richness on sample size can be revealed by dedicated analytical tasks such as constructing and comparing species accumulation curves or rarefaction curves (Magurran 2003). Extending this work beyond the exploration stage deserves deeper investigation in future work.

To keep our proposed taxonomy concise, we excluded analytical tasks necessary for collaborative exploration. For instance, analysts may wish to keep track of their findings and share their findings with other users (Pham et al. 2013). These tasks are generic and relevant to almost all scientific analysis workflows (Heer and Agrawala 2008).

Our literature review examines only three areas, ecology, microbiology, and organizational management because understanding diversity patterns and processes are fundamental problems in these areas (Gunderson 2000; Magurran 2003; Ogunseitan 2005; Fuhrman 2009; Lau and Murnighan 1998; Harrison and Klein 2007; Bezrukova et al. 2009). Additionally, the three areas cover the diversity of multivariate objects at *three encompassing scales*: microbial taxa (ecology/microbiology), species (ecology), and human beings (organizational management). Finally, although adopting somewhat separate vocabularies, interestingly, the three areas share many common characteristics and analysis goals as synthesized throughout this paper.

However, diversity represents itself in many other fields. For example, chemists consider the similarity/diversity of molecular models in exploring the multitude of designs generated by simulation (Izrailev and Agrafiotis 2004); scholars study language diversity in order to understand societies (Nettle 1998). All of these fields are advancing and new findings and analysis techniques may prompt revision of the framework and the taxonomy. Alternatively, we may have to create new ones for specific fields.

Implications for diversity studies

The alignment framework aims to support experts in adopting new diversity concerns within their own field of expertise or across fields. In addition to the examples presented in "Alignment of diversity concerns", we discuss several other usage scenarios here.

The first scenario demonstrates how the three types of diversity as variety, separation, and disparity in separate attributes could be extended to interaction among multiple attributes (Table 1, Rows 3 and 4). In fact, depending on the types of attribute under investigation, experts studying workgroups already conceptualize variety-based, separation-based, and disparity-based faultlines and subgroups, as introduced by Carton and Cummings (2012). For example, composition of disparity attributes such as pay, rank, and decision power may form disparity-based faultlines and subgroups in a team (Carton and Cummings 2012). The same conceptualization might be applied to functional diversity in ecology, depending on the types of examined functional traits. For example, composition of resource-based functional traits for plants such as nutrient consumption, tree density, body size could create disparity-based functional groups in the examined unit of study.

The second usage scenario extends our discussion on the alignability between diversity faultlines in organizational management and functional diversity in ecology. Across the two areas, it would be informative to cross compare statistical metrics (Petchey and Gaston 2002, 2006; Thatcher and Patel 2012) and visual representations. For example, while the faultline metric used in the baseball data (Fig. 10) does not involve a hierarchy of clusters (Bezrukova et al. 2009), hierarchical clustering algorithms such as the FD metric in ecology (Petchey and Gaston 2002) could potentially be adopted and vice versa. Other modern cluster algorithms from the field of data mining such as Affinity Propagation (Frey and Dueck 2007) could be potentially utilized. Further, configuration of attribute weighting is another unique feature of diversity faultlines potentially applicable to ecological functional diversity. For example, management researchers studying the impact of faultlines in workgroups may ask how many years of age difference between team members should be considered as equally important as a difference in gender or ethnicity (Thatcher et al. 2003). Ecologists studying functional diversity may adopt similar configuration of the relative importance of functional traits (Petchey and Gaston 2006), depending on the corresponding system processes of interest.

The third usage scenario discusses the missing component of taxonomic diversity in workgroup diversity (Table 1). To our understanding, experts studying workgroups have not yet examined hierarchical classification of attributes. That missing link may suggest a potential research direction. For example, functional expertise of team members is potentially hierarchical (e.g., ecology and microbiology majors are closely related since they are classified under life sciences) and the hierarchical information can be taken into account during investigation of faultlines and subgroups.

Implications for the design of visualization

Use of the alignment framework and the task taxonomy also has implications for the design of visualizations. Specifically, it provides visualization designers and researchers with a common vocabulary and considerations for designing and evaluating different visual-analysis tools

Table 5 Data behaviors of interest to diversity analysis and corresponding typical visual representations

Data Behavior	Examples of Diversity Concern	Data Characteristics	Typical visual representations (with example citations)
Distributions	Variety and abundance in separate attributes	Univariate	Boxplot (Tukey 1977)
			Histogram (Magurran 2003)
			Stacked Bar Chart (Caporaso et al. 2010)
			Rank-abundance Curve (Whittaker 1965; Magurran 2003)
			Cumulative Frequency Curve (Magurran 2003)
		Multivariate	2D Scatter plot and its variants (2D Heatmap, Fluctuation Diagram)
			Multiples of univariate representations, e.g., Boxplot (Tukey 1977), Histogram, Scatter plot matrix (Cleveland and McGill 1984), Diversity Map (Pham et al. 2011)
Distributions + clusters	Functional diversity; Subgroups/ Faultlines	Bivariate	Scatter Plot (Sedlmair et al. 2012)
			Mosaic Plot (Hartigan and Kleiner 1981)
		Multivariate	Multiple Stacked Histograms (Pham et al. 2014)
			Scatter Plot Matrix (Pham et al. 2014)
Distributions + hierarchies	Taxonomic Diversity (Richness+Evenness)	Multivariate	Treemap (Shneiderman 1992; Horn et al. 2009)
			Sunburst, Icicle (Stasko and Zhang 2000)
Hierarchies	Taxonomic diversity (Richness)	Multivariate	Node-link diagram and its variants, e.g., Tree (Lee et al. 2004), Dendrogram (Briggs et al. 2011)
Correlations	Processes of diversity	Bivariate	Scatter plot or line chart (Petchey and Gaston 2006; Bezrukova et al. 2010)
		Multivariate	Scatter plot matrix (Cleveland and McGill 1984)
			Parallel coordinates (Inselberg 2009; Fernstad et al. 2011)
			Parallel sets (Kosara et al. 2006)

targeting diversity data. We expect a set of base visualization techniques and tools for illuminating various components of diversity and providing new ways of looking at data across fields.

Typical visual representations. Following the three considerations and examples of visualization presented in the Alignment Framework Section, Table 5 is a useful list of typical visual representations that are well-suited to communicating the data behaviors of interest concerning diversity. The techniques, which by no means represent an exhaustive list, are suggested based on the understanding of their pros and cons from the field of information visualization (Pham et al. 2010, 2011, 2014). This classification would serve as a useful reference for visualization designers targeting specific diversity concerns. A thorough survey of various existing visualization techniques for general purpose can be found in Keim (2002).

This tabulation (Table 5) could be extended to include techniques targeting diversity in space and time. Recall that the three considerations suggest that if time and space are involved, the techniques should support users to explore how the data behaviors of interest (e.g., summary statistics, distributions, clusters, and/or hierarchies) vary over time and space. To communicate spatial distributions or clusters in univariate data, a geographical map with an additional encoding (e.g., a heat map) is widely used. However, visualizing data behaviors of multivariate data on a map remains a challenge. Potential solutions include overlaying other representations on a geographical map or alternatively, presenting geographical maps and other representations in separate windows connected by interactions (Andrienko and Andrienko 2006). On the other hand, to convey how the data behaviors of interest vary over time, one possible solution is to employ multiple snapshots of visual representations—for example, multiple histograms—one for each time point. Alternatively, animation of visualization states over time may potentially be useful. Andrienko and Andrienko (2006) present a thorough investigation of exploratory analysis of spatial and temporal data in their book.

Assessment of existing visual-analysis tools. In addition to guiding the invention of future visualizations, the three data behavior characterization considerations (see "Diversity patterns and Diversity processes") could be used to assess existing techniques and tools. For example, consider the MicrobiVis tool (Fernstad et al. 2011), which employed parallel coordinate plot (PCP)—among other techniques—to convey the separation between two groups of microbial samples across multiple OTUs (Fig. 19). PCP is well suited to make the data observations visible as well as to convey the correlation between two neighboring attribute axes (Inselberg 2009) (Table 5). However, we argue that the choice of PCP does not support Consideration 2—PCP is

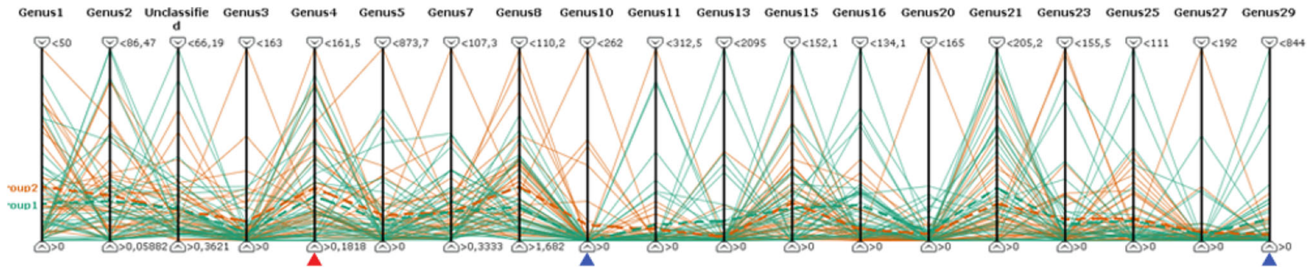

Fig. 19 Samples from two oral bacterial populations visualized using Parallel Coordinate Plot (PCP) supported by the MicrobiVis tool (Fernstad et al. 2011). *Vertical axes* represent a set of Genus OTUs of interest to the analyst; each of the polylines represents a sample that intersects each genus axis at the value corresponding to the abundance of the genus detected in the sample. *Distinct colors* are used to differentiate the two populations: group 1 and group 2. *Red* and *blue arrows* indicate some interesting genera identified by the analyst. For example, the first *blue arrow* from the *left marks* Genus 4 in which group 1 and group 2 are separated with respect to abundance. Image reused with permission from Fernstad et al. (2011). Copyright © 2011 IEEE

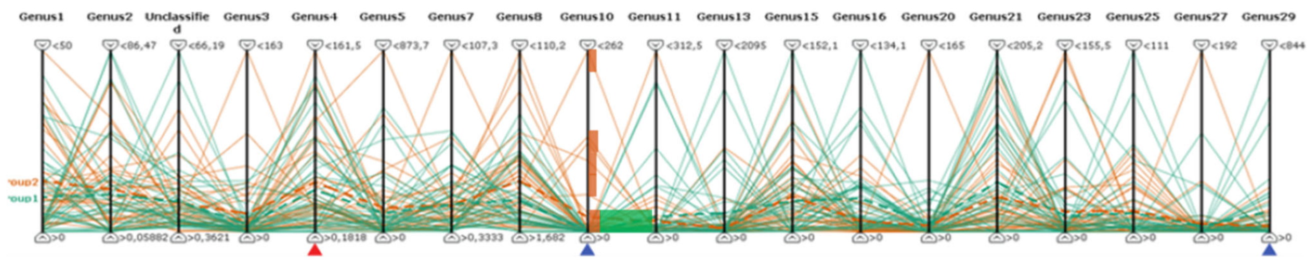

Fig. 20 An alternative design to the PCP in Fig. 19 in which stacked histograms are selectively overlaid along the axes to convey the distribution of clusters as well as separation among clusters across the attributes of interest. For demonstration purpose, Genus 4 axis—marked by the first blue arrow from the left—is re-drawn with stacked histograms. We argue that the stacked histograms make separation between the two groups of microbe samples within Genus 4 stand out: while all samples of group 1 contain low abundance of Genus 4, most of the samples of group 2 contain higher abundance of Genus 4. We adapt the Fig. from Fernstad et al. (2011). Original Figure © 2011 IEEE

not effective in supporting users in comparing the distributions and separation of clusters across multiple attributes (Pham et al. 2014). Fundamentally, a more effective design should start with synotic tasks in mind, as opposed to elementary tasks. Figure 20 presents an alternative design in which stacked histograms are selectively overlaid along the axes to convey the distribution of clusters as well as separation among clusters across the attributes of interest.

Conclusions

Ecologists are increasingly concerned about changes in diversity patterns of species communities and how they influence ecosystem functioning and stability. However, ecologists may not be aware of statistical and visual analysis techniques in other fields, such as organizational management, that may help improve their own understanding. Reciprocally, understanding concerns and analysis techniques of diversity in ecosystems may widen the perspectives of researchers who study diversity in human organizations. Aiming to connect that missing link, this interdisciplinary work abstracts diversity concerns across the analyses of species diversity, microbial diversity, and workgroup diversity in an alignment framework and offers an operationalization of these concerns in terms of data behaviors of interest and common analytical tasks. Subject-matter experts and tool designers may take advantage of this work to find a common ground for the diversity analysis problem. We expect this work will help guide the evaluation and refinement of existing visualization techniques as well as the invention of future ones. We also anticipate further discussions regarding validation and amendment to both the alignment framework and the unified task taxonomy.

Acknowledgments The authors wish to thank the subject-matter experts who participated in the survey study and provided valuable feedback on the manuscript. Support for this research was provided by National Science Foundation funding to the H.J. Andrews Long-term Ecological Research program (NSF 0823380) and ongoing U.S. Forest Service support to the H.J. Andrews Experimental Forest.

References

Ahn J, Plaisant C, Shneiderman B (2013) A task taxonomy for network evolution analysis. IEEE Trans Visual Comput Graph 20(3):365–376

Amar R, Eagan J, Stasko J (2005) Low-level components of analytic activity in information visualization. In: IEEE symposium on information visualization. IEEE, pp 111–117
Andrienko N, Andrienko G (2006) Exploratory analysis of spatial and temporal data. Springer, Berlin
Begon M, Townsend CR, Harper JL (2005) Ecology: from individuals to ecosystems. Wiley-Blackwell, New York
Bezrukova K, Jehn KA, Zanutto EL, Thatcher SM (2009) Do workgroup faultlines help or hurt? A moderated model of faultlines, team identification, and group performance. Organ Sci 20(1):35–50
Bezrukova K, Spell CS, Perry JL (2010) Violent splits or healthy divides? Coping with injustice through faultlines. Pers Psychol 63(3):719–751
Briggs B, Pohlman J, Torres M, Riedel M, Brodie E, Colwell F (2011) Macroscopic biofilms in fracture-dominated sediment that anaerobically oxidize methane. Appl Environ Microbiol 77(19):6780–6787
Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Meth 7(5):335–336
Carton AM, Cummings JN (2012) A theory of subgroups in work teams. Acad Manag Rev 37(3):441–470
Cleveland WS, McGill R (1984) Graphical perception: theory, experimentation, and application to the development of graphical methods. J Amer Stat Assoc 79(387):531–554
Cole J, Wang Q, Cardenas E, Fish J, Chai B, Farris R, Kulam-Syed-Mohideen A, McGarrell D, Marsh T, Garrity G et al (2009) The ribosomal database project: improved alignments and new tools for rRNA analysis. Nucl Acids Res 37(suppl 1):D141–D145
Fernstad SJ, Johansson J, Adams S, Shaw J, Taylor D (2011) Visual exploration of microbial populations. In: IEEE symposium on biological data visualization (BioVis). IEEE, pp 127–134
Frey BJ, Dueck D (2007) Clustering by passing messages between data points. Science 315(5814):972–976
Fuhrman JA (2009) Microbial community structure and its functional implications. Nature 459(7244):193–199
Gobet A, Böer SI, Huse SM, van Beusekom JE, Quince C, Sogin ML, Boetius A, Ramette A (2011) Diversity and dynamics of rare and of resident bacterial populations in coastal sands. ISME J 6(3):542–553
Gobet A, Quince C, Ramette A (2010) Multivariate cutoff level analysis (multicola) of large community data sets. Nucl Acids Res 38(15):e155
Gotelli N, Ellison A (2004) A primer of ecological statistics. Sinauer Associates, Sunderland
Gunderson LH (2000), Ecological resilience-in theory and application. Ann Rev Ecol Systemat pp 425–439
Harrison DA, Klein KJ (2007) What's the difference? Diversity constructs as separation, variety, or disparity in organizations. Acad Manag Rev 32(4):1199–1228
Hartigan JA, Kleiner B (1981) Mosaics for contingency tables. Proc Interf 1981:268–273
Heer J, Agrawala M (2008) Design considerations for collaborative visual analytics. Inf Visual 7(1):49–62
Heer J, Shneiderman B (2012) Interactive dynamics for visual analysis. ACM Queue 10(2):30
Hillebrand H (2004) On the generality of the latitudinal diversity gradient. Amer Nat 163(2):192–211
Horn MS, Tobiasz M, Shen C (2009) Visualizing biodiversity with voronoi treemaps. In: Sixth international symposium on Voronoi diagrams (ISVD'09). IEEE, pp 265–270
Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. Ecology 52(4):577–586
Inselberg A (2009) Parallel coordinates: visual multidimensional geometry and its applications. Springer, Berlin
Ives AR, Carpenter SR (2007) Stability and diversity of ecosystems. Science 317(5834):58–62
Izrailev S, Agrafiotis DK (2004) A method for quantifying and visualizing the diversity of QSAR models. J Mol Graph Model 22(4):275–284
Keim DA (2002) Information visualization and visual data mining. IEEE Trans Visual Comput Graph 8(1):1–8
Kosara R, Bendix F, Hauser H (2006) Parallel sets: interactive exploration and visual analysis of categorical data. IEEE Trans Visual Comput Graph 12(4):558–568
Lau DC, Murnighan JK (1998) Demographic diversity and faultlines: the compositional dynamics of organizational groups. Acad Manag Rev 23(2):325–340
Lawlor LR (1980) Overlap, similarity, and competition coefficients. Ecology, pp 245–251
Lee B, Parr CS, Campbell D, Bederson BB (2004) How users interact with biodiversity information using taxontree. In: Proceedings of the working conference on Advanced visual interfaces. ACM, New York, pp 320–327
Mackinlay J (1986) Automating the design of graphical presentations of relational information. ACM Trans Graph (TOG) 5(2):110–141
Magurran AE (2003) Measuring biological diversity. Wiley-Blackwell, New York
McCann KS (2000) The diversity–stability debate. Nature 405(6783):228–233
Meyer B, Glenz A (2013) Team faultline measures a computational comparison and a new approach to multiple subgroups. Organ Res Meth 16(3):393–424
Meyer B, Shemla M, Schermuly CC (2011) Social category salience moderates the effect of diversity faultlines on information elaboration. Small Group Res 42(3):257–282
Miller J (2005) Spatial and temporal distribution and abundance of moths in the Andrews experimental forest. http://andrewsforest.oregonstate.edu/data/abstract.cfm?dbcode=SA015
Munzner T (2009) A nested model for visualization design and validation. IEEE Trans Visual Comput Graph 15(6):921–928
Nettle D (1998) Explaining global patterns of language diversity. J Anthropol Archaeol 17(4):354–374
Ogunseitan O (2005) Microbial diversity: form and function in prokaryotes. Blackwell Publishing, UK
Orhan A, Tosun H (2010) Visualization of geotechnical data by means of geographic information system: a case study in eskisehir city (nw turkey). Environ Earth Sci 61(3):455–465
Petchey OL, Gaston KJ (2002) Functional diversity (FD), species richness and community composition. Ecol Lett 5(3):402–411
Petchey OL, Gaston KJ (2006) Functional diversity: back to basics and looking forward. Ecol Lett 9(6):741–758
Pham T, Hess R, Ju C, Zhang E, Metoyer R (2010) Visualization of diversity in large multivariate data sets. IEEE Trans Visual Comput Graph 16(6):1053–1062
Pham T, Highland S, Metoyer R, Henshaw D, Miller J, Jones J (2011) Interactive visualization of spatial and temporal patterns of diversity and abundance in ecological data. In: Proceedings of Environmental Information Management. Publisher of University of California, California, pp 104–110
Pham T, Jones J, Metoyer R, Swanson F, Pabst R (2013) Interactive visual analysis promotes exploration of long-term ecological data. Ecosphere 4(9). doi:10.1890/ES13-00121.1
Pham T, Metoyer R, Bezrukova K, Spell C (2014) Visualization of cluster structure and separation in multivariate mixed data: A case study of diversity faultlines in work teams. Comput Graph 38(2014):117–130

Rammette A (2007) Multivariate analyses in microbial ecology. FEMS Microbiol Ecol 62(2):142–160
Rink K, Kalbacher T, Kolditz O (2012) Visual data exploration for hydrological analysis. Environ Earth Sci 65(5):1395–1403
Sanders HL (1968) Marine benthic diversity: a comparative study. American Naturalist, USA
Sedlmair M, Tatu A, Munzner T, Tory M (2012) A taxonomy of visual cluster separation factors. Comput Graph Forum 31(3pt4):1335–1344
Seo J, Shneiderman B (2005) A rank-by-feature framework for interactive exploration of multidimensional data. Info Visual 4(2):96–113
Shneiderman B (1992) Tree visualization with tree-maps: 2D space-filling approach. ACM Trans Graph (TOG) 11(1):92–99
Shneiderman B (1996) The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings of Symposium on Visual Languages. IEEE, pp 336–343
Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". In: Proceedings of the National Academy of Sciences 103(32): 12,115–12,120
Spence R (2007) Information visualization: design for interaction. Prentice Hall, New Jersey
Stasko J, Zhang E (2000) Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In: IEEE symposium on information visualization (InfoVis 2000). IEEE, pp 57–65
Thatcher SM, Jehn KA, Zanutto E (2003) Cracks in diversity research: the effects of diversity faultlines on conflict and performance. Group Dec Negot 12(3):217–241
Thatcher SM, Patel PC (2012) Group faultlines a review, integration, and guide to future research. J Manag 38(4):969–1009
Thomas JJ, Cook KA (2006) A visual analytics agenda. IEEE Comput Graph Appl 26(1):10–13
Tukey JW (1977) Exploratory data analysis. Addison-Wesley, Boston
Van Knippenberg D, De Dreu CK, Homan AC et al (2004) Work group diversity and group performance: an integrative model and research agenda. J Appl Psychol 89(6):1008–1022
Velasco V, Gogu R, Vázquez-Suñè E, Garriga A, Ramos E, Riera J, Alcaraz M (2013) The use of gis-based 3D geological tools to improve hydrogeological models of sedimentary media in an urban environment. Environ Earth Sci 68(8):2145–2162
Warwick R, Clarke K (1998) Taxonomic distinctness and environmental assessment. J Appl Ecol 35(4):532–543
Weiskopf D, Ma KL, van Wijk JJ, Kosara R, Hauser H (2006) Scivis, infovis-bridging the community divide? In: Proceedings of the IEEE Visualization Conference. IEEE
Whittaker RH (1965) Dominance and diversity in land plant communities: numerical relations of species express the importance of competition in community function and evolution. Science 147(3655):250
Wickham H, Cook D, Hofmann H, Buja A (2010) Graphical inference for infovis. IEEE Trans Visual Comput Graph 16(6):973–979