Data Visualization (DSC 530/CIS 602-02)

Focus+Context

Dr. David Koop
Overview

Reducing Items and Attributes

- Filter
  - Items
    - [Diagram]
  - Attributes
    - [Diagram]

- Aggregate
  - Items
    - [Diagram]
  - Attributes
    - [Diagram]

Reduce

- Filter
  - [Diagram]

- Aggregate
  - [Diagram]

- Embed
  - [Diagram]

[Munzner (ill. Maguire), 2014]
Restaurant locations are derived from the New York City Department of Health and Mental Hygiene database. Due to the limitations of the Health Department’s database, some restaurants could not be placed.

By JEREMY WHITE
Scented Widgets

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hue</td>
<td>Varies the hue of the widget (or of a visualization embedded in it)</td>
<td><img src="OptionA" alt="Example" /> <img src="OptionB" alt="Example" /></td>
</tr>
<tr>
<td>Saturation</td>
<td>Varies the saturation of the widget (or of a visualization embedded in it)</td>
<td><img src="OptionA" alt="Example" /> <img src="OptionB" alt="Example" /></td>
</tr>
<tr>
<td>Opacity</td>
<td>Varies the saturation of the widget (or of a visualization embedded in it)</td>
<td><img src="OptionA" alt="Example" /> <img src="OptionB" alt="Example" /></td>
</tr>
<tr>
<td>Text</td>
<td>Inserts one or more small text figures into the widget</td>
<td>![Example](2 OptionA) ![Example](10 OptionB)</td>
</tr>
<tr>
<td>Icon</td>
<td>Inserts one or more small icons into the widget</td>
<td><img src="OptionA" alt="Example" /> <img src="OptionB" alt="Example" /></td>
</tr>
<tr>
<td>Bar Chart</td>
<td>Inserts one or more small bar chart visualizations into the widget</td>
<td><img src="OptionA" alt="Example" /> <img src="OptionB" alt="Example" /></td>
</tr>
<tr>
<td>Line Chart</td>
<td>Inserts one or more small line chart visualizations into the widget</td>
<td><img src="OptionA" alt="Example" /> <img src="OptionB" alt="Example" /></td>
</tr>
</tbody>
</table>

[Willett et al., 2007]
Attribute Filtering on Star Plots

[Yang et al., 2003]
Aggregation: Histograms

- Very similar to bar charts
- Often shown without space between (continuity)
- Choice of number of bins
  - Important!
  - Viewers may infer different trends based on the layout

[Munzner (ill. Maguire), 2014]
Two-Dimensional Binning
Spatial Aggregation

In cartography, changing the boundaries of the regions used to analyze data can yield dramatically different results.
Spatial Aggregation

[Images illustrating spatial aggregation and the modifiable areal unit problem]

[Penn State, GEOG 486]
Spatial Aggregation

[Penn State, GEOG 486]
Gerrymandering

- Massachusetts
  - Gov. Elbridge Gerry
- South Essex County
- Looks like a salamander

[E. Tisdale, Boston Gazette, 1812]
Modifiable Areal Unit Problem

• How you draw boundaries impacts the type of aggregation you get
• Similar to bins in histograms
• Gerrymandering

[Wonkblog, Washington Post, Adapted from S. Nass]
Boxplots

• Show **distribution**

• Single value (e.g. mean, max, min, quartiles) doesn't convey everything

• Created by John Tukey who grew up in New Bedford!

• Show **spread** and **skew** of data

• Best for **unimodal** data

• Variations like vase plot for multimodal data

• Aggregation here involves many different marks
Four Distributions, Same Boxplot...

Normal

Bimodal

Peaked

Skewed

Box plot

[C. Choonpradub and D. McNeil, 2005]
Exam 2

- **Wednesday, April 5, 3:30pm-4:45pm**
- Similar Format to Exam 1: Multiple Choice + Free Response
- Cumulative but emphasis on the topics covered since Exam 1
  - Trees
  - Geospatial Data
  - Color & Colormaps
  - Interaction
  - Multiple Views
  - Filtering & Aggregation
- Reading: Chs. 9.5, 8.1-8.3, 10, 6, 11, 12, 13
- Research papers discussed in class on the website
Project Designs

• Due April 7 April 14
• Feedback soon
• Start thinking about ways to evolve the sketch or go in new directions based on more recent class discussions
• 3 Designs ≠ 3 Views
  - Want to see how your designs evolve, not three basic encodings
• Some inspiration:
  - University of Washington
  - Stanford University
  - University of Miami (more journalism-focused)
Overview

Reducing Items and Attributes

- **Filter**
  - Items
  - Attributes

- **Aggregate**
  - Items
  - Attributes

[Reduce]

[Filter]

[Aggregate]

[Embed]

[Munzner (ill. Maguire), 2014]
Attribute Aggregation

• Remember reducing attributes—use statistics: either one variable matches another or doesn't change!
• We can also use similar criteria for aggregating attributes
• Cluster similar attributes together
  - How?
K-Means
K-Means Issues

Shape

Number of Clusters

[D. Robinson, 2015]
Dimensionality Reduction

• Attribute Aggregation: Use fewer attributes (dimensions) to represent items
• Combine attributes in a way that is more instructive than examining each individual attribute
• Example: Understanding the language in a collection of books
  - Count the occurrence of each non-common word in each book
  - Huge set of features (attributes), want to represent each with an aggregate feature (e.g. high use of "cowboy", lower use of "city") that allows clustering (e.g. "western")
  - Don't want to have to manually determine such rules
• Techniques: Principle Component Analysis, Multidimensional Scaling family of techniques
Principle Component Analysis (PCA)
Principal Component Analysis (PCA) is a technique used to emphasize variation and bring out strong patterns in a dataset. It's often used to make data easy to explore and visualize.

First, consider a dataset in only two dimensions, like (height, weight). This dataset can be plotted as points in a plane. But if we want to tease out variation, PCA finds a new coordinate system in which every point has a new (x,y) value. The axes don't actually mean anything physical; they're combinations of height and weight called "principal components" that are chosen to give one axes lots of variation.

Drag the points around in the following visualization to see PC coordinate system adjusts.

Here's the plot of the data along the first principal component. Already we can see something is different about Northern Ireland.

Now, see the first and second principal components, we see Northern Ireland a major outlier. Once we go back and look at the data in the table, this makes sense: the Northern Irish eat way more grams of fresh potatoes and way fewer of fresh fruits, cheese, fish and alcoholic drinks. It's a good sign that structure we've visualized reflects a big fact of real-world geography: Northern Ireland is the only of the four countries not on the island of Great Britain. (If you're confused about the differences among England, the UK and Great Britain, see: this video.)

For more explanations, visit the Explained Visually project homepage. Or subscribe to our mailing list.

Non-linear Dimensionality Reduction

Original data space $\mathcal{X}$

Component space $\mathcal{Z}$

$\Phi_{gen} : \mathcal{Z} \rightarrow \mathcal{X}$

$\Phi_{extr} : \mathcal{X} \rightarrow \mathcal{Z}$

[M. Scholz, CC-BY-SA 2.0]
Dimensionality Reduction in Visualization

Fig. 8. MDS layouts showing visual quality, time, and stress for the Glimmer, Hybrid, and PivotMDS algorithms. The data set name, the number of nodes (N), and the number of dimensions (D) appear above each column. Time in seconds appears at the bottom left of each entry, with normalized stress on the bottom right.

[D. Koop, DSC 530, Spring 2017] [Glimmer, Ingram et al., 2009]
Tasks in Understanding High-Dim. Data

**Task 1**
- **In:** HD data
- **Out:** 2D data
- **What?** In High-dimensional data
- **Why?** Produce, Derive

**Task 2**
- **In:** 2D data
- **Out:** Scatterplot, Clusters & points
- **What?** In 2D data
- **Why?** Discover, Explore
- **How?** Encode, Navigate

**Task 3**
- **In:** Scatterplot, Clusters & points
- **Out:** Labels for clusters
- **What?** In Scatterplot
- **Why?** Produce
- **Why?** Annotate

[Munzner (ill. Maguire), 2014]
Probing Projections

Abstract

—We introduce a set of integrated interaction techniques to interpret and interrogate dimensionality-reduced data. Projection techniques generally aim to make a high-dimensional information space visible in form of a planar layout. However, the meaning of the resulting data projections can be hard to grasp. It is seldom clear why elements are placed far apart or close together and the inevitable approximation errors of any projection technique are not exposed to the viewer. Previous research on dimensionality reduction focuses on the efficient generation of data projections, interactive customisation of the model, and comparison of different projection techniques. There has been only little research on how the visualization resulting from data projection is interacted with. We contribute the concept of probing as an integrated approach to interpreting the meaning and quality of visualizations and propose a set of interactive methods to examine dimensionality-reduced data as well as the projection itself. The methods let viewers see approximation errors, question the positioning of elements, compare them to each other, and visualize the influence of data dimensions on the projection space. We created a web-based system implementing these methods, and report on findings from an evaluation with data analysts using the prototype to examine multidimensional datasets.

1INTRODUCTION

A primary goal of information visualization is to find patterns and relationships in multivariate datasets. Many visualization techniques have been developed towards this goal such as multiple coordinated views [2], parallel coordinates [14], scatterplot matrices [28], and dimensionality reductions such as multidimensional scaling (MDS) [5]. Dimensionality reductions are a particular class of techniques that synthesise high-dimensional data spaces onto projection spaces with much fewer dimensions, typically the two dimensions of the plane. While most visualization techniques juxtapose the different data dimensions as matrices or columns, dimensionality reductions integrate them into a planar canvas. The projection results in a so-called spatialisation (i.e., embedding) of data elements that approximately represents similarity as proximity and in turn dissimilarity as distance. Considering that the human perceptional system comprises a well-developed capacity for spatial reasoning, the assumption is that spatialisation would be a more natural way [31] to analyse high-dimensional datasets since groupings, separations, and other patterns among data elements become immediately discernible.

However, there are two major caveats linked with dimensionality reduction: first, it can be challenging to interpret the positions of projected elements, and second, the errors that occur with any projection technique are not exposed to the viewer.

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication xx Aug. 2015; date of current version 25 Oct. 2015.

For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

[J. Stahnke et al., 2015]
Probing Projection Goals

• Examining the Projection
• Exploring the Data
• Design Goals:
  - Show and correct approximation errors
  - Allow for multi-level comparisons
  - Spatial orientation
  - Consistent design
• Allow grouping of samples
  - Selections
  - Classes
  - Clusters

[J. Stahnke et al., 2015]
4.3 Comparing elements

Elements can be analysed by viewing their values and comparing them to the rest of the dataset. The means are used instead. The graphical representation also takes into account, showing a density plot instead of a bar. As shown in Figure 4, points in the projection are displayed as bar charts, with density plots of the distribution, providing a quick visual way of locating the relevant elements.

4.4 Analysing dimensions

Analysing a single sample is done by hovering the mouse pointer over a dot on the projection. The values for the corresponding sample can be compared to other samples by selecting it and hovering over other samples. A tooltip will appear and visualise the differences. The means are used instead. The graphical representation also takes into account, showing a density plot instead of a bar. As shown in Figure 4, points in the projection are displayed as bar charts, with density plots of the distribution, providing a quick visual way of locating the relevant elements.

Tooltips with statistics

[Image 41x71 to 980x626]
Comparing Two Groups

**South America** 3 samples

**Northern Europe** 9 samples

<table>
<thead>
<tr>
<th>Dimension</th>
<th>South America</th>
<th>Northern Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational attainment</td>
<td>50</td>
<td>77</td>
</tr>
<tr>
<td>Employees working very...</td>
<td>18</td>
<td>6.2</td>
</tr>
<tr>
<td>Life expectancy</td>
<td>75</td>
<td>81</td>
</tr>
<tr>
<td>Life satisfaction</td>
<td>7.1</td>
<td>7.4</td>
</tr>
<tr>
<td>Self-reported health</td>
<td>65</td>
<td>77</td>
</tr>
<tr>
<td>Student skills</td>
<td>420</td>
<td>500</td>
</tr>
<tr>
<td>Time devoted to leisure...</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Years in education</td>
<td>16</td>
<td>19</td>
</tr>
</tbody>
</table>

---

[J. Stahnke et al., 2015]
Heatmap from Dimension Hover

---

The dimension itself to see how elements compare to other elements, to account for clustering of the data, and the distribution of values in matter in this regard: the spatial distribution of values in the projection when analysing a dimensionality-reduced projection. Two things it is important to be able to quickly reference original dimensions and groups work together, making it possible to compare a sample to density plots for each of them. The methods for comparing samples in which the sample deviates most. An individual sample can also be compared to other samples by selecting it and hovering over other points in the projection. Some additional information, such as the distribution, providing a quick visual way of locating the relevant selection, cluster, or class is displayed with a thumbnail of its spatial range. The values for the corresponding sample are indicated in the list of dimensions. Additionally, a tooltip appears, over a dot on the projection. The values for the corresponding sample can be compared to other samples by selecting it and hovering over other samples. A tooltip will appear and visualise the differences.

Elements can be analysed by viewing their values and comparing them to the dataset in general, or to other selections in particular. Even though it is useful to look at individual dimensions, it is important to know how those values correspond with the positioning of the dots. (For some techniques, such as PCA, projected dimensions. It would then be possible to display this as a dynamic heatmaps in the projection and density plots in the sidebar. Projections created with most dimensionality-reduction techniques, such as MDS, have no meaningful axes, complicating spatial orientation because dimensional values are distributed nonlinearly. Yet, in principle, interpolating values for the projection space can also be used to answer a more theoretical question: what values would a fictive sample have to have to be in a specific cluster? We used a heatmap to try to answer this question.

Analysing groups works similarly. When selecting a group of elements, they are indicated in the list of dimensions, and on the projection itself as thumbnails. All heatmaps are also shown as a heatmap in the projection on the left. The heatmap is a grid of cells each representing the value for a certain dimension at its position, with higher values being darker. Brightness is used to avoid confusion with the group colours. This allows to visually assess the value distribution for a given dimension, and for comparing the selection to the dataset. A tooltip comparison is also available, showing a density plot instead of a bar. As shown in Figure 2, a tooltip displays the sample’s absolute values, standard deviations, and graphical representations for each dimension. The graphical representation also takes into account, showing a density plot instead of a bar. As shown in Figure 3, hovering over a dimension in the sidebar displays its distribution as a heatmap in the projection on the left.

Analysing a single sample is done by hovering the mouse pointer over a dot on the projection. The values for the corresponding sample can be compared to other samples by selecting it and hovering over other dots on the projection. One solution is to use a glyph plot, with the dots themselves being used to represent an additional dimension, for example by varying their size according to the values. This technique is available in the prototype and can be used to visualise a dimension spatially. Where the original data has multiple clusters, the contribution of each original dimension can be mapped to the positioning of the dots. (For some techniques, such as PCA, one can try to answer this question: what values would a fictive sample have to have to be in a specific cluster? We used a heatmap to try to answer this question.)

4.4 Analysing dimensions

Analysing groups works similarly. When selecting a group of elements, they are indicated in the list of dimensions, and on the projection itself as thumbnails. All heatmaps are also shown as a heatmap in the projection on the left. The heatmap is a grid of cells each representing the value for a certain dimension at its position, with higher values being darker. Brightness is used to avoid confusion with the group colours. This allows to visually assess the value distribution for a given dimension, and for comparing the selection to the dataset. A tooltip comparison is also available, showing a density plot instead of a bar. As shown in Figure 2, a tooltip displays the sample’s absolute values, standard deviations, and graphical representations for each dimension. The graphical representation also takes into account, showing a density plot instead of a bar. As shown in Figure 3, hovering over a dimension in the sidebar displays its distribution as a heatmap in the projection on the left.

Analysing a single sample is done by hovering the mouse pointer over a dot on the projection. The values for the corresponding sample can be compared to other samples by selecting it and hovering over other dots on the projection. One solution is to use a glyph plot, with the dots themselves being used to represent an additional dimension, for example by varying their size according to the values. This technique is available in the prototype and can be used to visualise a dimension spatially. Where the original data has multiple clusters, the contribution of each original dimension can be mapped to the positioning of the dots. (For some techniques, such as PCA, one can try to answer this question: what values would a fictive sample have to have to be in a specific cluster? We used a heatmap to try to answer this question.)

4.4.1 Heatmaps

The heatmap is a grid of cells each representing the value for a certain dimension at its position, with higher values being darker. Brightness is used to avoid confusion with the group colours. This allows to visually assess the value distribution for a given dimension, and for comparing the selection to the dataset. A tooltip comparison is also available, showing a density plot instead of a bar. As shown in Figure 2, a tooltip displays the sample’s absolute values, standard deviations, and graphical representations for each dimension. The graphical representation also takes into account, showing a density plot instead of a bar. As shown in Figure 3, hovering over a dimension in the sidebar displays its distribution as a heatmap in the projection on the left.

Analysing a single sample is done by hovering the mouse pointer over a dot on the projection. The values for the corresponding sample can be compared to other samples by selecting it and hovering over other dots on the projection. One solution is to use a glyph plot, with the dots themselves being used to represent an additional dimension, for example by varying their size according to the values. This technique is available in the prototype and can be used to visualise a dimension spatially. Where the original data has multiple clusters, the contribution of each original dimension can be mapped to the positioning of the dots. (For some techniques, such as PCA, one can try to answer this question: what values would a fictive sample have to have to be in a specific cluster? We used a heatmap to try to answer this question.)

4.3 Comparing elements

Comparing the selection to the dataset. A tooltip comparison is also available, showing a density plot instead of a bar. As shown in Figure 2, a tooltip displays the sample’s absolute values, standard deviations, and graphical representations for each dimension. The graphical representation also takes into account, showing a density plot instead of a bar. As shown in Figure 3, hovering over a dimension in the sidebar displays its distribution as a heatmap in the projection on the left.

Analysing a single sample is done by hovering the mouse pointer over a dot on the projection. The values for the corresponding sample can be compared to other samples by selecting it and hovering over other dots on the projection. One solution is to use a glyph plot, with the dots themselves being used to represent an additional dimension, for example by varying their size according to the values. This technique is available in the prototype and can be used to visualise a dimension spatially. Where the original data has multiple clusters, the contribution of each original dimension can be mapped to the positioning of the dots. (For some techniques, such as PCA, one can try to answer this question: what values would a fictive sample have to have to be in a specific cluster? We used a heatmap to try to answer this question.)
Fig. 5. Comparing an individual sample with a class from the well-known Iris flower data set. In addition to the distribution of dimensions of the class and the value of the sample, the visualization also indicates sample-centric distortions using grey and white halos.

The grid is constructed by dividing up the projection space into a number of cells, with their size based on the mean of distances for the closest three points from each point. A minimum value is used to prevent the grid from getting too small to be useful, and calculations taking too long.

The values for each grid cell are calculated by averaging the values of the points in the cell, or, if there are none, the three closest points for the cell, weighting them according to their distances from the cell. This ensures smooth transitions over large gaps in the projection space, while being responsive to abrupt changes at the same time.

4.4.2 Density plots

While the heatmaps show how the values are spatially distributed in the projection space, kernel density plots in the list of dimensions show their value distributions. In the prototype, currently the plots are generated roughly equivalent to R's bw.nrd0 function which uses Silverman's 'rule of thumb' [24], p. 48. Percentiles are indicated on the density plots to support the visual assessment. Used together with brushing and linking, it is possible to assess how a sample, or a group of samples, relates to the whole dataset.

Markers or sub-plots for selected elements are shown on the density plots in the list of dimensions, providing dynamic highlights of samples being examined (see Figure 5, lower right). Additional markers display the dimensional values mouse position in the projection space, based on the calculations done for the heatmap, making it possible to gradually track value progressions for multiple dimensions.

4.5 Examining projection errors

Besides exploring the distribution of samples and dimensions, the visualization environment allows for the integrated examination of projection errors by providing per-sample halos, distance corrections, and a dendrogram.

4.5.1 Error halos

The cumulative distance error for each point is displayed as a halo around the dot, with the radius corresponding to the relative amount of error and the value indicating the direction of the error (see Figure 6).

This is intended to help the user visually understand the quality of the projection and find potentially unreliable spots. Hovering over a dot shows the errors in relation to the hovered point, to check if the distances between certain points are correct or just projection artefacts, and learn which points should be closer together or further apart.

Halos were chosen because their visual properties are a good match for the properties of the error they represent. The error does not belong to the sample, but to the projection, and as such should be displayed by the projection. A halo is clearly connected to the dot, but also part of the projection. Size was chosen as a very intuitive metric to display the amount of error, with points with a large error standing out easily.

The brightness of the halo displays the direction of the error. If the other points are farther away in the projection than in the high-dimensional space, the halo is white; if they are too close in the projection, the halo is dark. White was chosen for points that should be nearer because it stands out more, and while using the prototype ourselves, we often ended up looking for 'similar' samples.

Size and brightness were chosen over colour or shape, as using colour would have clashed with the colouring of the dots, and different shapes were not as glanceable as changes in brightness.

4.5.2 Distance correction

After examining the approximation errors, the viewer might decide that the errors of a certain point warrant more attention. They can then visualize them by selecting to view the high-dimensional distances between the selected point and all others. This removes all projection errors when it comes to distance, but for the selected sample only. The new, corrected positions for each point are calculated by taking the vector between the selected and the other point and multiplying it by the distance error ratio between the two. The angle between them is kept as is. As a result, the other point moves directly towards or away from the specified point, correcting the distance.

[O. Stahnke et al., 2015]
Fig. 6. Halos represent the cumulative error for the respective samples. White indicates that a majority of samples is more similar than indicated by their distance to the given sample; grey indicates the opposite. The paths travelled by the points are shown as lines, leading from the points' original positions in the projection to the new, corrected positions (see Figure 8). This connects them to their original positions in the projection, and displays the size of the distance error at the same time. Resembling the brightness encoding of the halos, the brightness of the lines indicates whether they've moved closer or farther away.

A problem with this solution is that it introduces new distortions in the spatial relationship between all other points. Only the distances directly between the selected point and the other points are reliable, whereas all the other distances are distorted, and the new positioning might lead to wrong assumptions about potential clusterings. To mitigate this problem, the correction paths are shown. Another solution would be to recompute the projection while preserving the distances from and to the selected point and being more generous with distance errors among the remaining points. This would somewhat reduce the introduced distortions. However, in a recomputed projection, the positions of the points might change significantly, most likely leading to completely different positions for all points, possibly confusing the observer even if an animation is used.

Fig. 7. Dendrograms mapped onto the projection. Left: projection with low projection error. Right: high projection error.

4.5.3 Dendrogram

In addition to the visualization of errors and corrections, a dendrogram can visualize the samples with regard to their position in the clustering hierarchy. Such a dendrogram (using the same agglomerative algorithm as the clusters) overlaid onto the projection may also help to visualise high-dimensional distances on the projection space. It graphically emphasises clusters by connecting close dots through dense lines. Interestingly, the dendrogram is a surprisingly good indicator of goodness of fit: if many thick, long lines intersect, it is likely that the projection is of low quality.

Example: OECD Countries

To illustrate the functionality of the interface we visualize the dataset of OECD countries in the prototype (see Figure 9). The dataset contains 8 dimensions for 36 countries. First, the viewer is drawn to the projection and notices Turkey that seems to be a clear outlier, far away from all other countries. To explore why this is, the viewer can examine this sample by hovering over it. A tooltip relating Turkey to the rest of the dataset appears, showing that it deviates strongly from the mean in nearly every dimension. This indicates the positioning as outlier is probably correct.

To test this assumption and build up trust in the visualization, the viewer selects ‘correct distances’, showing the high-dimensional distances between Turkey and the other countries. This reveals that Turkey should be even farther apart from several of the other countries. Having confirmed that Turkey is an outlier in this dataset, the viewer uses the built-in clustering to get a sense of how the countries are grouped. Playing around with the number of clusters, they notice that there seem to be seven clusters roughly corresponding to the geographical and geopolitical placement of the countries.

Taking a closer look at the positioning of the clustered countries, they realise that the arrangement seems to roughly correspond to geographic directions: Northern and Southern countries are roughly distributed along the vertical axes, East and West along the horizontal. To find out if or how this correlates with the dimensions, the viewer first compares the different clusters. Here the differences along the dimensions are very much pronounced. Interestingly though, life expectancy is lower in Latin America than Asia, while the self-reported health is higher for the former than the latter.

After a few more comparisons between the clusters, the viewer becomes interested in the dimension life satisfaction and turns towards the heatmaps. They notice that the values for life satisfaction and self-reported health seem to be higher in the Western countries, whereas the value for employees working very long hours seems to be especially high in the countries of the far East and the South.


White: higher levels of similarity
Gray: lower levels of similarity

[J. Stahnke et al., 2015]
User Study & Results

• Types of Questions:
  - How would you try to characterize the type X?
  - In what way are X and Y different in their properties?
  - Are the projections of X and Y correct or do they deviate? How do you interpret this?
  - Can you discover which parts of the cluster combinations are A, B, and C?

• Discussion:
  - Learnability: need more effective mechanisms for grasping the concepts behind dimensionality reduction
  - Manipulation: What happens with results?
  - Large data: What about text corpora?

[J. Stahnke et al., 2015]
Focus+Context

• Show everything at once but compress regions that are not the current focus
  - User shouldn't lose sight of the overall picture
  - May involve some aggregation in non-focused regions
  - "Nonliteral navigation" like semantic zooming

• Elision

• Superimposition: more directly tied than with layers

• Distortion
Focus+Content Overview

- Embed
  - Elide Data
    - [Diagram of Elide Data]
  - Superimpose Layer
    - [Diagram of Superimpose Layer]
  - Distort Geometry
    - [Diagram of Distort Geometry]

Reduce

- Filter
  - [Diagram of Filter]
- Aggregate
  - [Diagram of Aggregate]
- Embed
  - [Diagram of Embed]

[Munzner (ill. Maguire), 2014]
Elision

• There are a number of examples of elision including in text, DOI Trees, …

• Includes both filtering and aggregation but goal is to give overall view of the data

• In visualization, usually correlated with focus regions
Degree of Interest Function

- DOI = I(x) - D(x,y)
  - I: interest function
  - D: distance (semantic or spatial)
  - x: location of item
  - y: current focus point (could be more than one)
- Interactive: y changes
Elision: DOITrees

- Example: 600,000 node tree
  - Multiple foci (from search results or via user selection)
  - Distance computed topologically (levels, not geometric)

[Heer and Card, 2004]
Superimposition

• Different from layers because this is restricted to a particular region
  - For Focus+Context, superimposition is **not global**
  - More like overloading

• Lens may occlude the layer below
Superimposition with Interactive Lenses

(a) Alteration

(b) Suppression

[ChronoLenses and Sampling Lens in Tominski et al., 2014]
Superimposition with Interactive

(c) Enrichment

[Extended Lens in Tominski et al., 2014]
It can be difficult to observe micro and macro features simultaneously with complex graphs. If you zoom in for detail, the graph is too big to view in its entirety. If you zoom out to see the overall structure, small details are lost.

Focus + context techniques allow interactive exploration of an area.

[M. Bostock, http://bost.ocks.org/mike/fisheye/]
Distortion Choices

- How many focus regions?
  - One
  - Multiple
- Shape of the focus?
  - Radial
  - Rectangular
  - Other
- Extent of the focus
  - Constrained similar to magic lenses
  - Entire view changes
- Type of interaction:
  - Geometric, moveable lenses, rubber sheet
Overplotting

[M. Bostock, http://bost.ocks.org/mike/fisheye/]
Cartesian Distortion

[M. Bostock, http://bost.ocks.org/mike/fisheye/]
Cartesian Distortion

[M. Bostock, http://bost.ocks.org/mike/fisheye/]
Figure 3. LiveRAC shows a full day of system management time-series data using a reorderable matrix of area-aware charts. Over 4000 devices are shown in rows, with 11 columns representing groups of monitored parameters. (a): The user has sorted by the maximum value in the CPU column. The first several dozen rows have been stretched to show sparklines for the devices, with the top 13 enlarged enough to display text labels. The time period of business hours has been selected, showing the increase in the In pkts parameter for many devices. (b): The top three rows have been further enlarged to show fully detailed charts in the CPU column and partially detailed ones in Swap and two other columns. The time marker (vertical black line on each chart) indicates the start of anomalous activity in several of spire's parameters. Below the labeled rows, we see many blocks at the lowest semantic zoom level, and further below we see a compressed region of highly saturated blocks that aggregate information from many charts.

Principle: multiple views are most effective when coordinated through explicit linking.

The principle of linked views [15] is that explicit coordination between views enhances their value. In LiveRAC, as the user moves the cursor within a chart, the same point in time is marked in all charts with a vertical line. Similarly, selecting a time segment in one chart shows a mark in all of them. This technique allows direct comparison between parameter values at the same time on different charts. In addition, people can easily correlate times between large charts with detailed axis labels, and smaller, more concise charts.

Assertion: showing several levels of detail simultaneously provides useful high information density in context. Several technique choices are based on this assertion. First, LiveRAC uses stretch and squish navigation, where expanding one or many regions compresses the rest of the view [11, 17]. The accompanying video shows the look and feel of this navigation technique. The stretching and squishing operates on rectangular regions, so expanding a single chart also magnifies the entire row for the device it represents, and the entire column for the parameters that it shows. The edges of the display are fixed so that all cells remain within the visible area, as opposed to conventional zooming where some regions are pushed off-screen. There are rapid navigation shortcuts to zoom a single cell, a column, an aggregated group of devices, the results of a search, or to zoom out to an overview. Users can also directly drag grid lines or resize freely drawn on-screen rectangles. Navigation shortcuts can also be created for any arbitrary grouping, whose cells do not need to be contiguous. This interaction mechanism affords multiple focus regions, supporting multiple levels of detail.

Second, charts in LiveRAC dynamically adapt to show visual representations adapted in each cell to the available screen space. This technique, called semantic zooming [13], allows a hierarchy of representations for a group of device-parameter time-series. In Figure 3, the largest charts have multiple overlaid curves and detailed axis and legend labels. Smaller charts show fewer curves and less labeling, and at smaller sizes only one curve is shown as a sparkline [24]. On each curve, the maximum value over the displayed time period is indicated with a red dot, the minimum with a blue dot, and the current value with a green one. All representation levels color code the background rectangle according to dynamically changeable thresholds of the minimum, maximum, or average values of the parameters within the current time window. The smallest view is a simple block, where this color coding is the only information shown.

Third, aggregation techniques achieve visual scalability by ensuring dense regions show meaningful visual representations. Given our target scale of dozens of parameters and thousands of devices, the size of the matrix could easily surpass 100,000 cells. Stretch and squish navigation allows users to quickly create a mosaic with cells of many different sizes. [McLachlan et al., 2008]
Distortion Concerns

- Distance and length judgments are **harder**
  - Example: Mac OS X Dock with Magnification
  - Spatial position of items changes as the focus changes

- Node-link diagrams not an issue… why?

- Users have to be made aware of distortion
  - Back to scatterplot with distortion example
  - Lenses or shading give clues to users

- **Object constancy**: understanding when two views/frames show the same object
  - What happens under distortion?
  - 3D Perspective is distortion… but we are well-trained for that

- Think about **what** is being shown (filtering) and method (fisheye)