Multidimensional Visualization

Álvaro Figueira, PhD.
MSc in Data Science - Data Visualization

[dcc] DEPARTAMENTO DE CIÊNCIA DE COMPUTADORES
FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO

FC FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO
U.PORTO

Introduction

- It is common to classify multidimensional visualization techniques into
  - **Point-based** techniques
  - **Line-based** techniques
  - **Region-based** techniques

The techniques discussed in this module can only be used for lists and data tables/frames that do not have spatial attribute(s)

Point-based Techniques

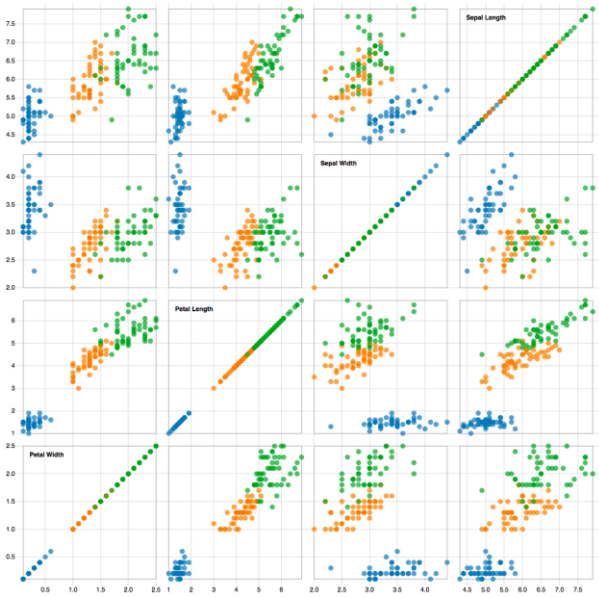
- Point-based techniques are projections of m-dimensional instances into a p-dimensional visual space (p={1,2,3})
  - And then, a glyph is associated to each point

Scatterplots and Scatterplot Matrices

- Scatterplots are the most common visual representations, but as dimensionality increases new strategies have to be sought
  - **Dimension selection**: manually or using some algorithm
  - **Dimensionality reduction**: using techniques such as PCA
  - **Incorporate dimensions**: map other dimensions to graphical elements
  - **Multiple displays**: showing several displays side-by-side

Scatterplots and Scatterplot Matrices

Scatterplot matrices are grids of scatterplots that shows all combinations of the n-dimensions (n^2)



https://bl.ocks.org/mbostock/4063663

226

226

Scatterplots and Scatterplot Matrices

- Papers to read
  - T. N. Dang and L. Wilkinson, "ScagExplorer: Exploring Scatterplots by Their Scagnostics," 2014 IEEE Pacific Visualization Symposium, Yokohama, 2014, pp. 73-80.
  - N. Elmqvist, P. Dragicevic and J. D. Fekete, "Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation," in IEEE Transactions on Visualization and Computer Graphics, vol. 14, no. 6, pp. 1539-1148, Nov.-Dec. 2008. (https://www.youtube.com/watch?v=E1birsp9iYk)

227

227

Multidimensional Projections

- Note that the goal of projection methods is to keep, as much as possible, the relationships of the *m*-dimensional space into the *p*-dimensional space of the visualization.
  - For instance, similarity relationships, neighborhood, etc.
- The result is a set of points (glyphs) on the plane
  - Close points indicate related instances, distant points indicate non-related objects

228

228

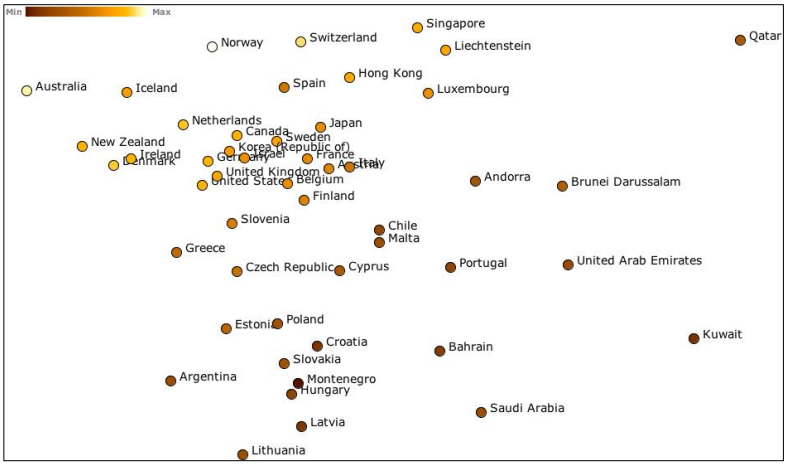
Multidimensional Projections

- Example: Human Development Index of 2006

Variables:

- Life expectancy at birth
- Expected years of schooling
- Mean years of schooling
- Gross national income (GNI) per capita

Can you identify the axis?

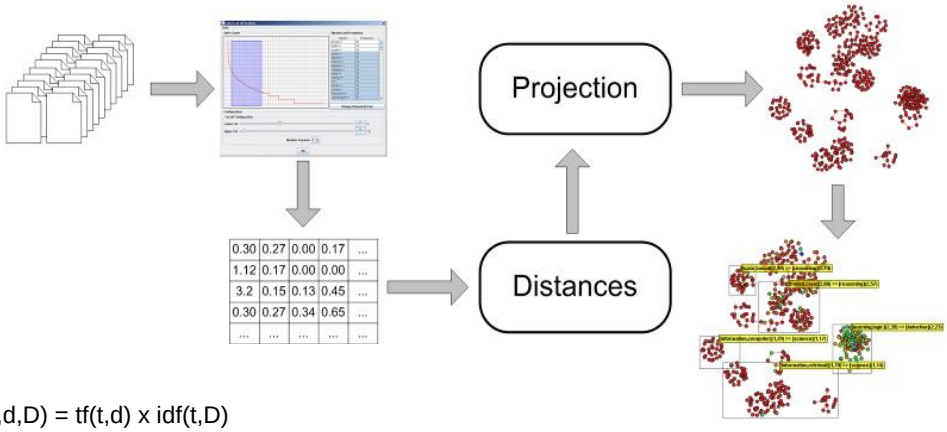


230

230

Multidimensional Projection for documents

- Example of Bag-of-words representation of document collections



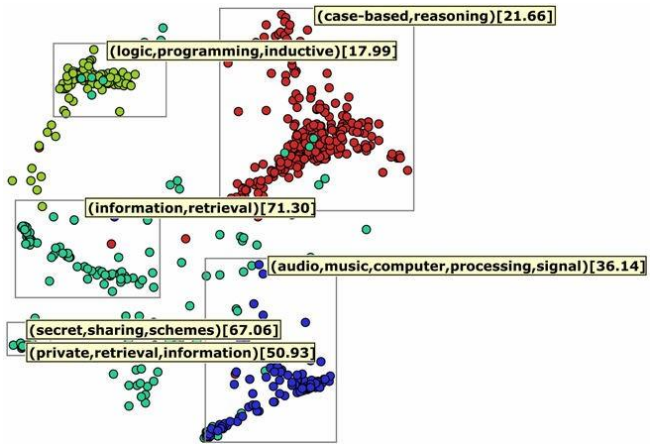
TF-IDF(t,d,D) = tf(t,d) x idf(t,D)
= (#t in d) x log (|D| / |(d in D: t in d)|)

231

231

Multidimensional Projection

- Example: collection of scientific papers



232

232

Point-based Techniques

- **Radviz** is another example of a point-based technique
  - For a *m*-dimensional dataset, *m* anchors are created and distributed over a circumference

- The position A_j of the j^{th} anchor ($j=[0,m-1]$) is calculated as

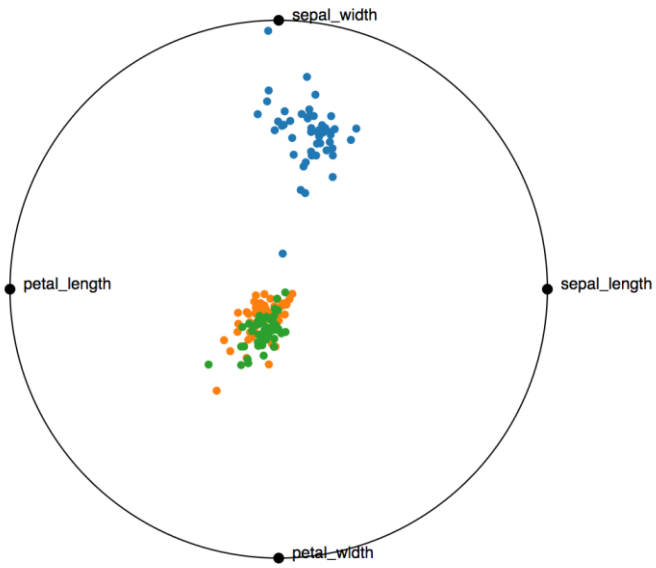
$$A_j^x = r * cos(j * 360/m) + cx$$
$$A_j^y = r * sin(j * 360/m) + cy$$

  - where $r=1$ is the circumference radius and $c=(cx,cy)$ its center

237

237

RadViz



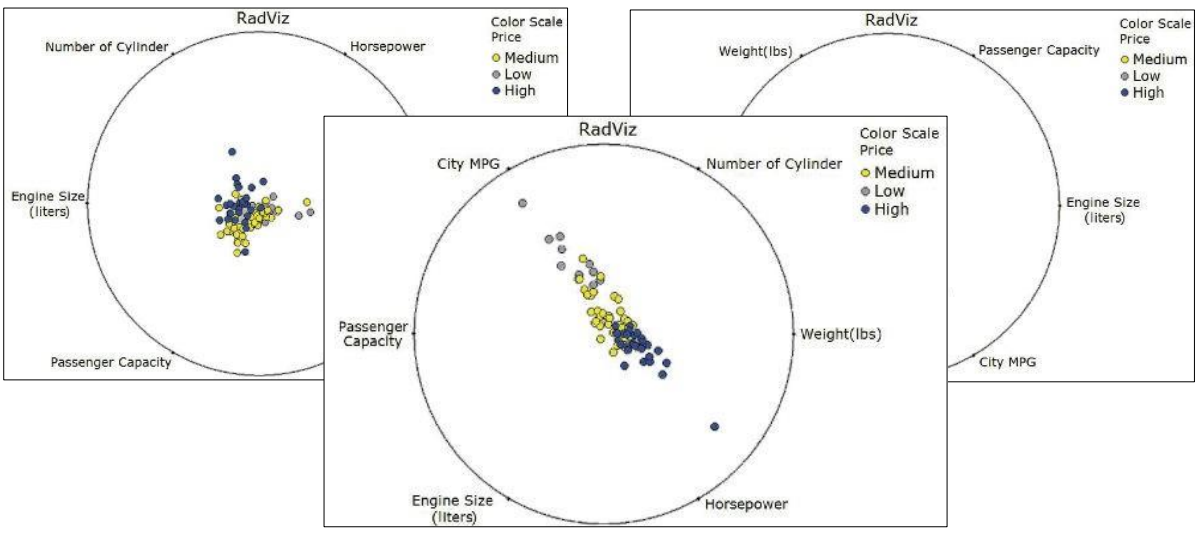
239

239

RadViz

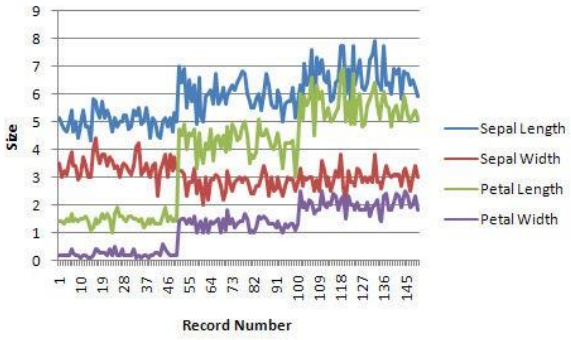
- Observe that the order of the anchors dictates the final result
  - Interaction can help users
  - There are techniques that maximize the points' spread

RadViz



Line-based Techniques

- On line-based techniques, the data patterns are represented through line crossing, curvatures, etc.



242

242

Line-based Techniques

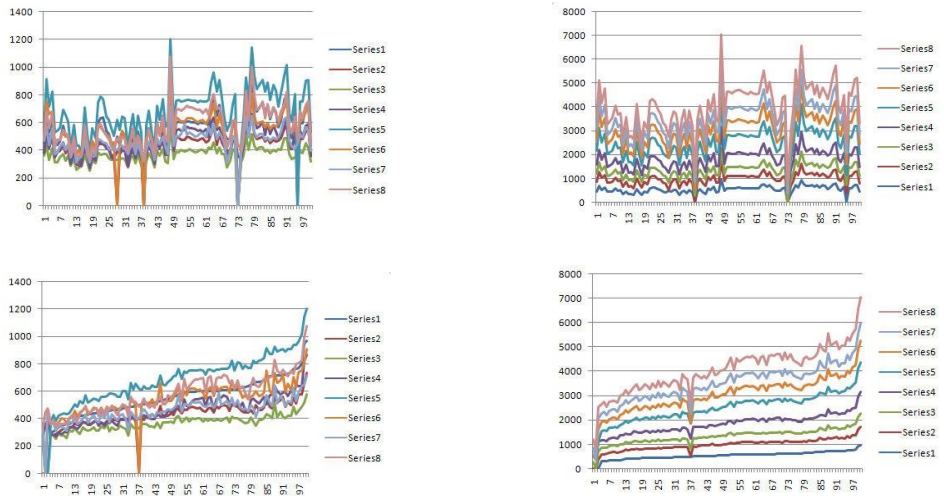
- Superimposition should not be used on datasets with many dimensions due to occlusion problems

- Potential solutions
  - Stack up the lines considering the previous dimension (difficult to evaluate the real value)
  - Order the instances considering one dimension

243

243

Line-based Techniques



Example: Professor salaries in 100 universities.

244

Line-based Techniques

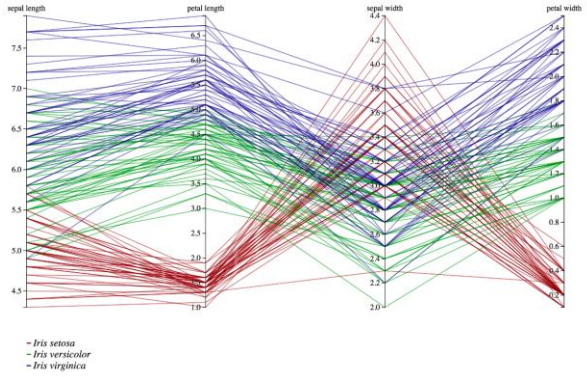
- These last techniques can only be used if the units of the dimensions are related in all series

- Solutions
  - Multiple coordinated stacked graphs
  - Multiple vertical axes can be used for different dimensions

245

Parallel Coordinates

- On the parallel coordinates, the axes are parallel (non orthogonal) and the data instances are represented as polylines that cross the axes on positions proportional to the value on the dimension



http://mbostock.github.io/d3/talk/20111116/iris-parallel.html

246

246

Parallel Coordinates

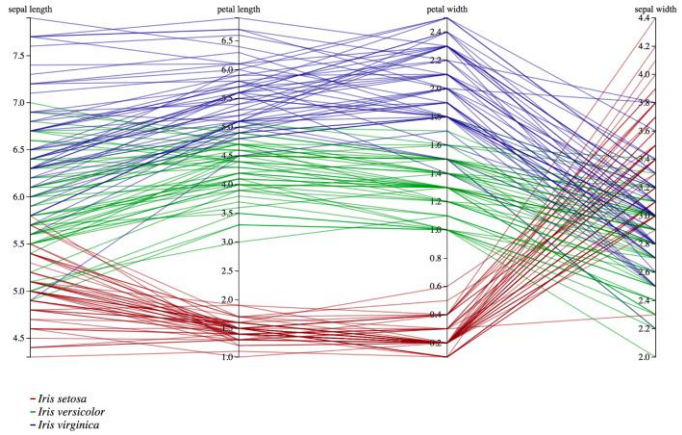
- With this visual representation it is possible to
  - Locate groups of similar polylines
  - Locate crossing points
  - Locate different polylines

247

247

Parallel Coordinates

- However, correlating non-consecutive dimensions is difficult
  - Interaction can be used to reorder the dimensions



248

248

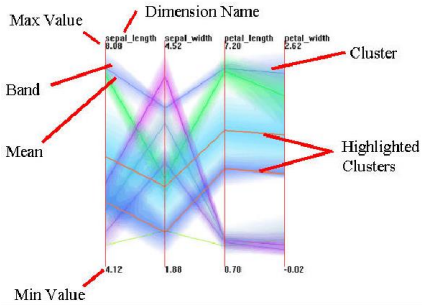
Parallel Coordinates

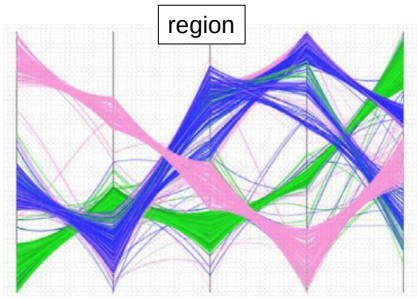
- There are many variants of the original parallel coordinates technique
  - Hierarchical parallel coordinates show groups, not individual data instances
  - Semi-transparency can be used to show groups on large databases
  - Grouping, re-ordering, and different spacing based on correlation
  - Using histograms on the axes can help on interpreting the data distribution
  - Curves can be used on the crossing points to improve the axes continuity

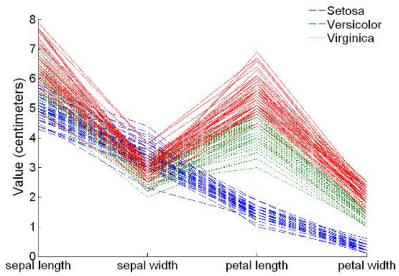
249

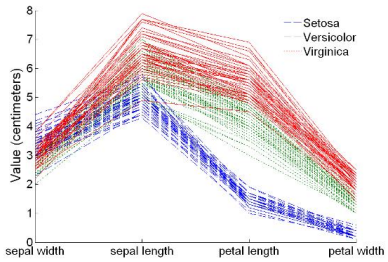
249

Parallel Coordinates









Reordering the vertical axes

250

250

Parallel Coordinates

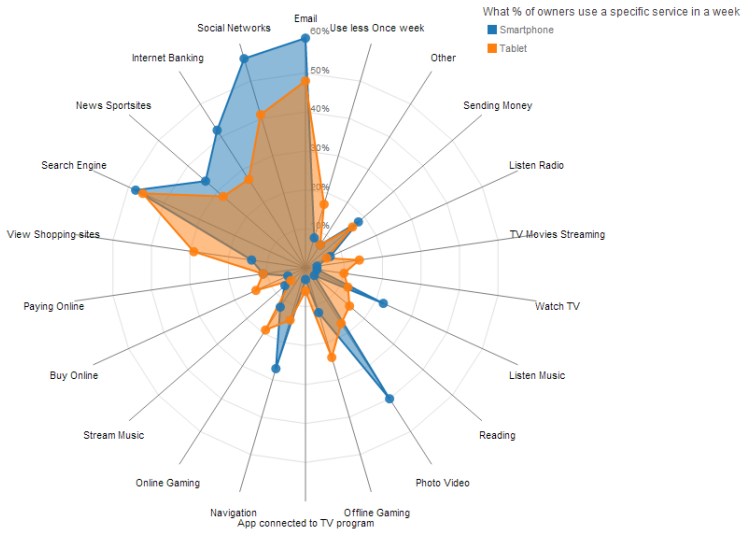
- For automatic axes reordering read
  - A.O. Artero, M.C.F.D. Oliveira, H. Levkowitz, **Enhanced high dimensional data visualization through dimension reduction and attribute arrangement**, in: Proceedings of the conference on Information Visualization, IEEE Computer Society, 2006, pp. 707–712.
  - Liang Fu Lu, Mao Lin Huang, Jinson Zhang, **Two axes re-ordering methods in parallel coordinates plots**, In Journal of Visual Languages & Computing, Volume 33, Pages 3-12, 2016.

251

251

Radial Axes Techniques

- For each technique with parallel axes there is a technique with radial axes

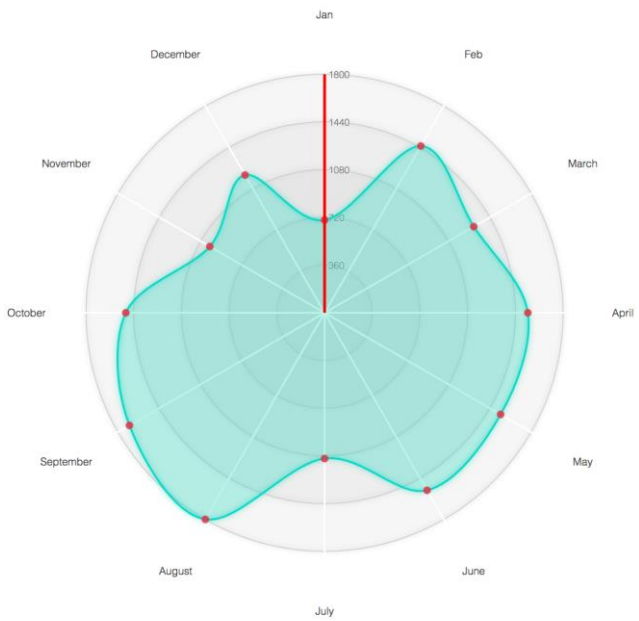


http://bl.ocks.org/nbremer/6506614

252

252

Radial Axes Techniques



The circular line graph is one example indicated for cyclic events

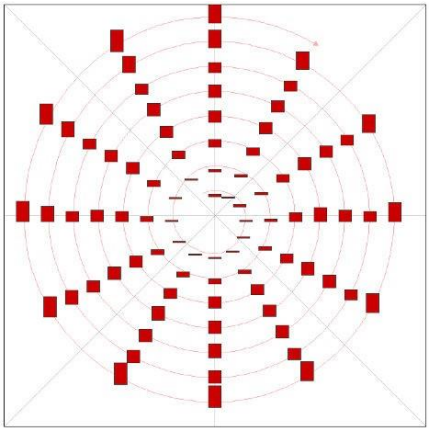
http://bl.ocks.org/tezzutezzu/c9d8706587e8f5b5d72084b083b502f8

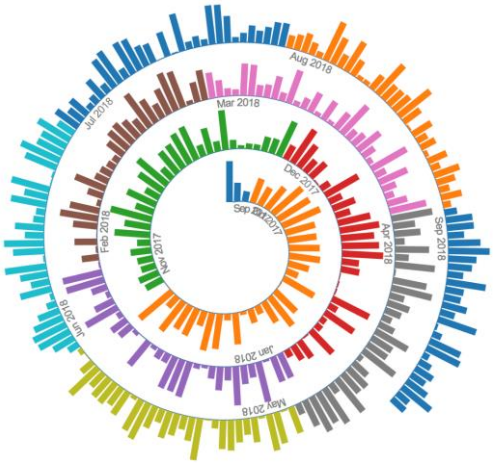
253

253

Radial Axes Techniques

- For layouts with more than one cycle, concentric circles or spirals can be used





https://bl.ocks.org/arpitnarechania/027e163073864ef2ac4ceb5c2c0bf616

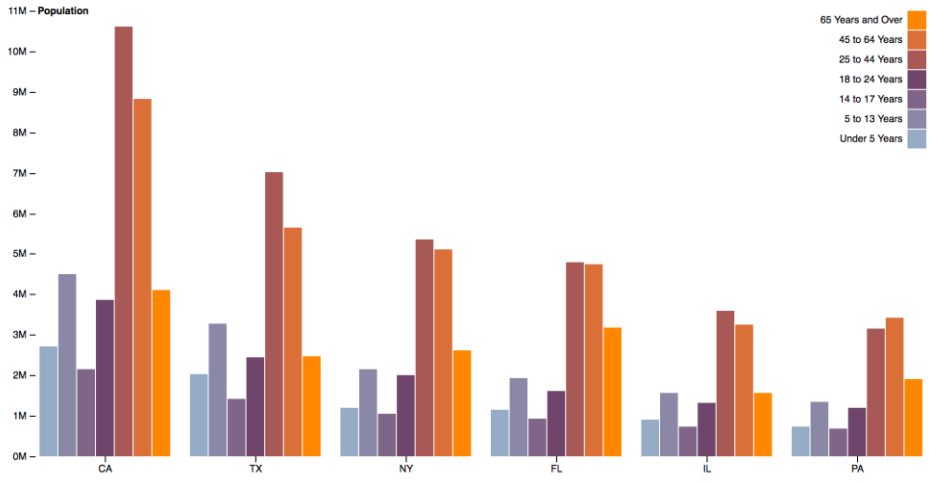
Region-based Techniques

- Region-based techniques present data values using filled polygons varying the size, color, shape, and other visual attributes
  - Although it present several cognitive problems, many techniques have been developed for:
    - Pie graphs
    - Bar graphs

Bar Graphs and Histograms

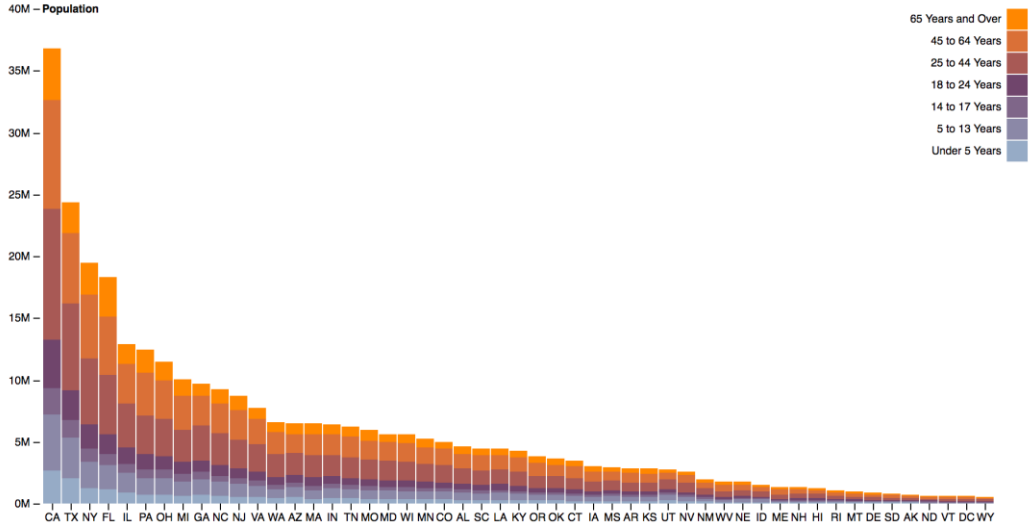
- One of the most common visual representation are the bar graphs
  - Stacked Bar Charts
  - Grouped Bar Charts
  - 3D Histogram

Grouped Bar Chart



https://bl.ocks.org/mbostock/3887051

Stacked Bar Chart



https://bl.ocks.org/mbostock/3886208

258

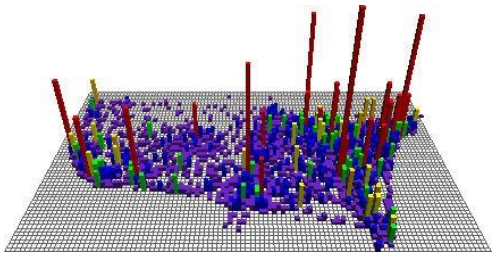
Bar Graphs and Histograms

- If the task is to understand the data distribution, then histograms can be used.
- It is simple for nominal values or few integers – the same number of bars and distinct values

259

Bar Graphs and Histograms

- The 3D version of bar graphs is called Cityscape
  - It is also called 3D histogram
  - Often used for georeferenced data

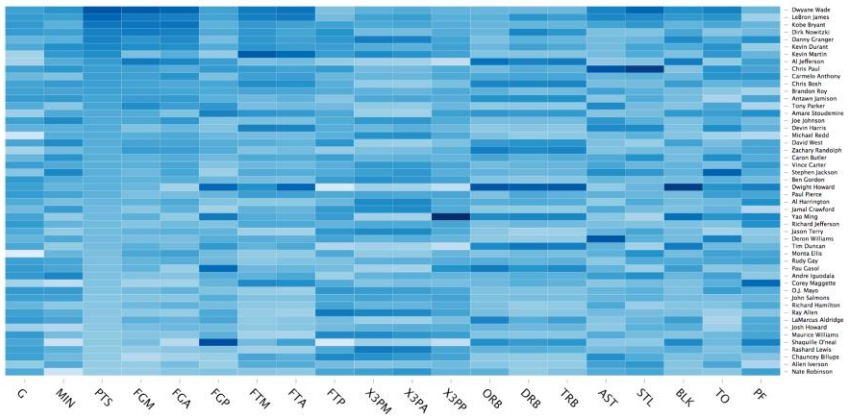


Tabular Displays

- It is easy to generate visual representations if the data is organized in a table

Tabular Displays

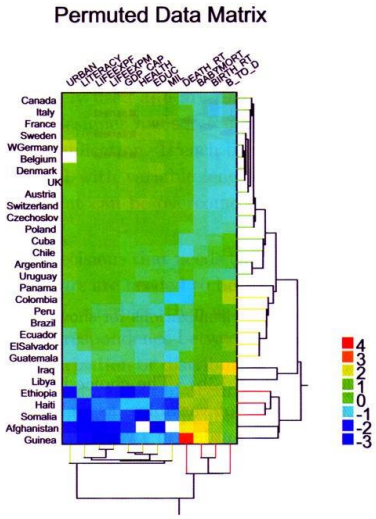
- Heatmaps map data values into rectangles filled with colors given by a color scale (colormap)



Tabular Displays

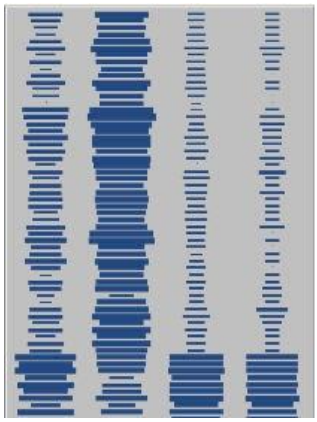
- Permutation and reordering can be used to improve the visual representation, changing the lines and columns positions

Tabular Displays



Tabular Displays

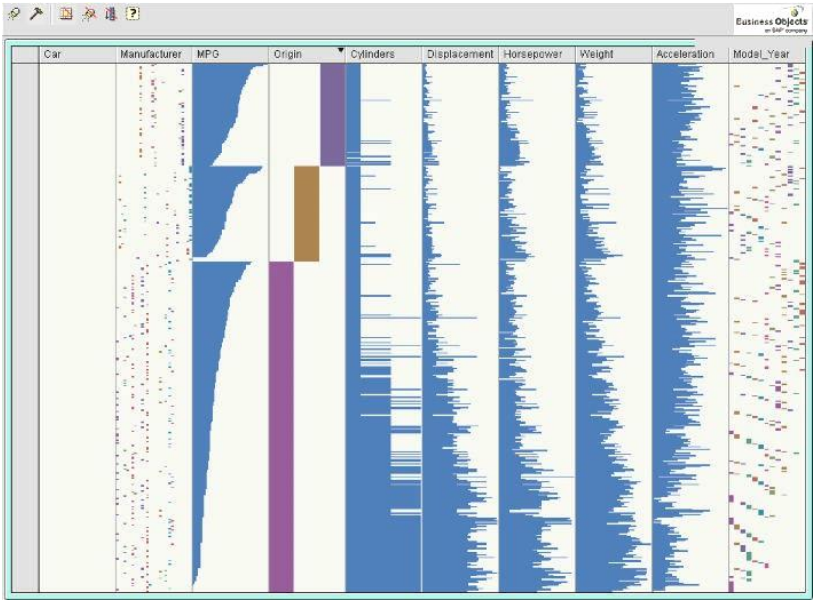
- Survey Plots varies the cell size instead of coloring
  - To avoid problems related to adjacent colors



Tabular Displays

- TableLens combines these ideas and add mechanisms to visualize all the table while providing a detailed view

(sorted by manufacturer and then by MPG)



266

266

Dense Pixel Displays

- Dense pixel display techniques map each data value to individual pixels, creating a filled polygon to represent each data dimension
  - Employ all visual space
  - Each value defines a pixel color

267

267

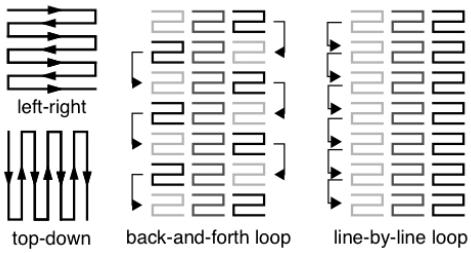
Dense Pixel Displays

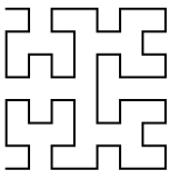
- The most basic approach create an image per dimension
- The elements are drawn so that close values are place together on the display
- An image can be created traversing the display from left-to-right (right-to-left)
- A spiral traverse can also be used
- Space-filling recursive curves can also de used (close elements in a list stay close on the display)

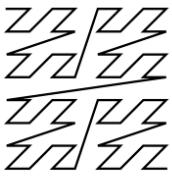
268

268

Dense Pixel Displays





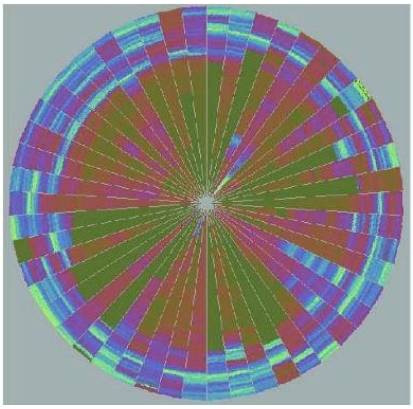


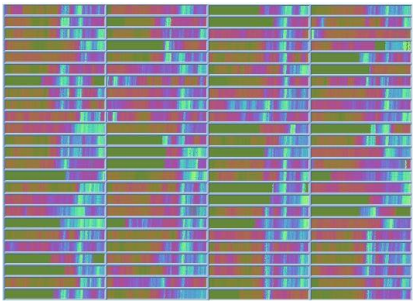
269

269

Dense Pixel Displays

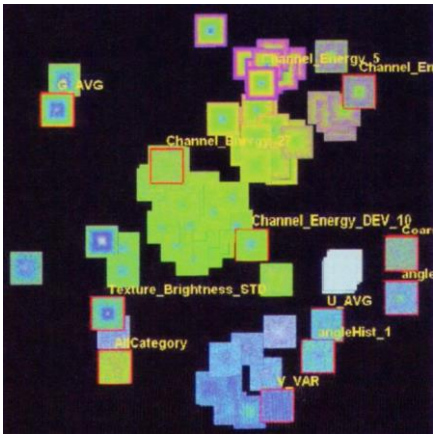
- The images of each dimension can be placed on the screen using different strategies: circle segments or grid





Dense Pixel Displays

- A projection method can also be used to place similar instances close together



Dense Pixel Displays

- Elements order (whenever possible) can help on the identification of interesting patterns
  - Order based on one dimension help to reveal clusters in that dimension
  - Order based on the distance to a vector can help to reveal clusters on several dimensions

Combining Techniques

- Some techniques combine the features of two or more classes of techniques
  - Glyphs creation

Glyphs and Icons

- Different types of mappings can be used when creating glyphs
  - One-to-one: each data attribute is mapped to a distinct graphical attribute
  - One-to-many: redundant mapping is used to improve precision
  - Many-to-one: different attributes are mapped to a single graphical attribute using the space, orientation or other transformation to segregate them

Give examples!

274

274

Glyphs and Icons

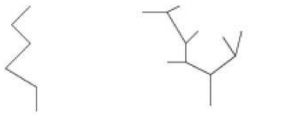
- Some example of glyphs are
  - Profile – height and color of bars
  - Stars – size of rays emanating from the center
  - Anderson/metroglyphs – size of rays
  - Stick figures – size, angle, color
  - Trees – size, thickness, branch angles, etc.
  - Autoglyph – color
  - Boxes – height, width, depth, etc.
  - Hedgehogs – arrows in a vector field, varying thickness and orientation
  - Faces – size and position of eyes, nose, mouth, etc.
  - Arrows – size, length, color, etc.

See next slide

275

275

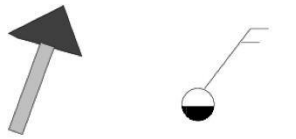
Glyphs and Icons



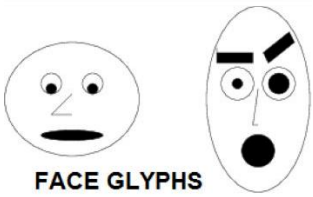
STICKS AND TREES

STARS AND METROGLYPHS

PROFILE GLYPHS



ARROWS/WEATHERVANES



FACE GLYPHS

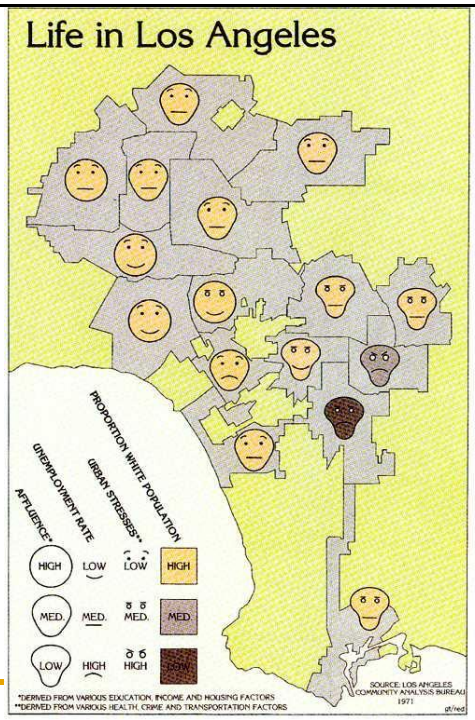
AUTOGLYPH/BOX GLYPH

All convey information regarding several dimensions

276

276

Glyphs and Icons



277

277

Glyphs and Icons

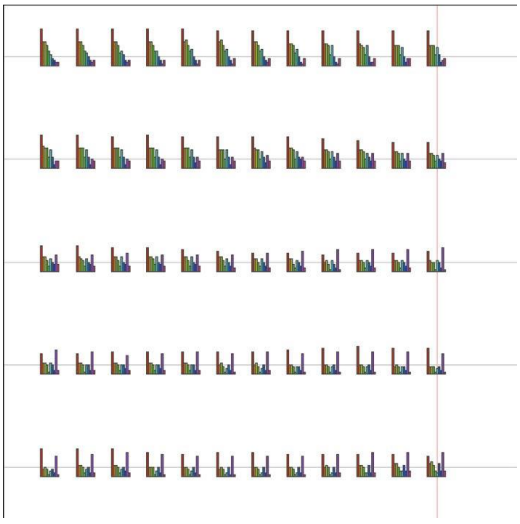
- There are three different strategies to position the glyphs
  - Uniform: glyphs are scaled and positioned with equal spacing between them to fill the entire screen - avoid overlaps
  - Data-driven: data values are used to determine the position of glyphs - can be positioned using scatterplots or multi-dimensional projections
  - Structure-driven: If there is any implicit structure, such as cyclic or hierarchical, it can be used to position the glyphs

See next slides

280

280

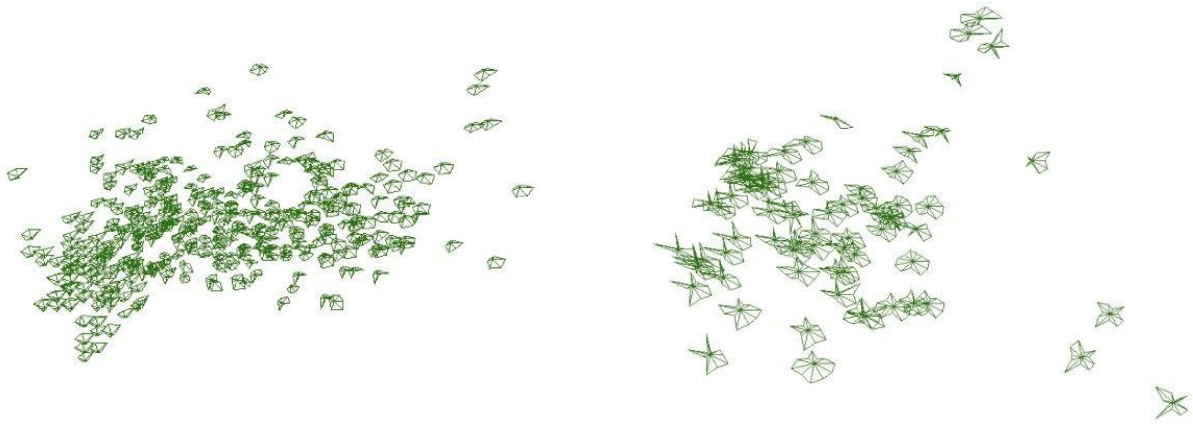
Glyphs and Icons (uniform)



281

281

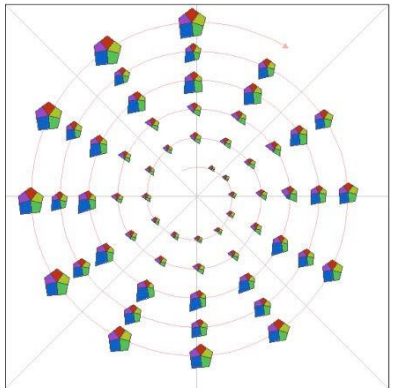
Glyphs and Icons (data-driven)

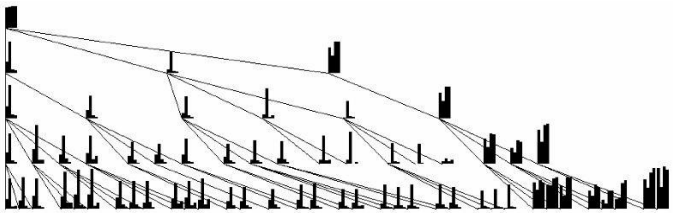


282

282

Glyphs and Icons (structure-driven)





283

283

Reference

- Ward, M., Grinstein, G. G., Keim, D. **Interactive data visualization foundations, techniques, and applications**. Natick, Mass., A K Peters, 2010.

284