Ph. D. Thesis

Curran Kelleher

1   The Ideal Data Management Infrastructure

2   The Ideal Visual Analytics Environment

- Tables generated from a data cube navigator

  - Time slider is part of the navigation environment
  - Tree navigator for hierarchical dimensions

- Missing values are handled correctly

  - A set of special values represent missing data
  - NULL values in the database treated as missing data

- One global selection

  - All tools modify and reflect the global selection
  - All tools support the following selection interactions

    * Persistent selection shapes
      · Rectangle, Circle, Lassoo
      · Shapes are movable, resizable, deletable, storable
    * Set operations union and difference

- One global filter

  - Filter created by creating a subset from selection

- One global color map

- Color map defined as a set of (value,color) pairs
- The color map is editable
- Color map presets can be managed (CRUD) by priveleged users

- One global probing specification

  - A set of columns to include
  - A text template specification for each chosen column

- The following visualization tools are available

  - Scatterplot
    * 1 dimension
      · Records are points
    * 5 measures
      · Point X
      · Point Y
      · Point Size
      · Point Glyph
      · Point Color
  - Table, Table Lens
    * 1 dimension
      · Records are rows
    * n measures
      · Each measure is a column
      · Values shown as text
  - Heatmap
    * 1 dimension
      · Records are rows
    * n measures
      · Each measure is a column
      · Values shown as color
  - Parallel Coordinates

* 1 dimension
    · Records are polylines
  * n measures
    · Each measure is a vertical axis
- Choropleth Map (colored polygons)
  * 1 dimension (must be space)
    · Records are geographic regions
  * 1 measure
    · Region Color
- Pie Chart
  * 1 dimension
    · Records are pie slices
  * 2 measures
    · Pie slice size
    · Pie slice color
- Bar Chart
  * 1 dimension
    · Records are bars
  * 1 measure
    · Bar Height
    · Bar Width
    · Bar Color
- Tree Map (and other tree views)
  * 1 dimension (hierarchical)
    · Records are tree nodes
  * 2 measures
    · Node size
    · Node color
- Timeseries Line Chart
  * 2 dimensions (one must be time)
    · Records of an arbitrary dimension are polylines

- · Records of the time dimension are X locations
  - * 1 measure
    - · Polyline point Y Location
- – Timeseries Stacked Area Chart
  - * 2 dimensions (one must be time)
    - · Records of an arbitrary dimension are polylines
    - · Records of the time dimension are X locations
  - * 1 measure
    - · Area size in the Y direction
- – Histogram (Barchart of density)
  - * 1 dimension
    - · Records are X points
  - * 1 measure
    - · X point location
- – Correlation Matrix (Heatmap of correlation)
  - * 1 dimension
    - · specifies record type (records not directly visualized)
  - * n measures
    - · Each measure is a row and a column
    - · Color of each measure pair represents correlation

- Integrated R scripts

  - – R scripts expose
    - * A unique name
    - * A description of what it does
    - * Who created it
    - * When it was created
    - * Input table structure
    - * Output table structure
  - – R scripts create derived data sets
  - – Visualization tools handle data sets derived from R scripts
    - * Scatterplot handles regression line derived dataset
    - * Histogram handles distribution fitting derived dataset

4

3 Mathematorium

4 Universal Visualization Platform

5 JyVis

6 Weave

7 Quadstream

8 Universal Data Cube Ontology

9 Universal Data Cube Server

Urls:

- `http://datacubes.info/dimensions/time`

- `http://datacubes.info/dimensions/space`

- `http://datacubes.info/dimensions/industry`

- `http://datacubes.info/dimensions/ownership`

- `http://datacubes.info/meausures/population`

- `http://datacubes.info/meausures/employment`

- `http://datacubes.info/meausures/average_annual_income`

- `http://datacubes.info/meausures/total_wages`

- `http://datacubes.info/meausures/num_establishments`

- RDF Description of the Time dimension:
  `/dimensions/time`

- listing of all years
  `/dimensions/time/years`

- RDF Description of the year 1990

  `/dimensions/time/years/`9901

- listing of all quarters in year 1990

  `/dimensions/time/years/`990/quarters1

- RDF Description of Quarter 1 of the year 1990

  `/dimensions/time/years/`990/quarters/q11

- listing of all months in Q1 of year 1990

  `/dimensions/time/years/`990/quarters/q1/months1

- RDF Description of month January 1990

  `/dimensions/time/years/`990/quarters/q1/months/january1

- listing of all quarters

  `/dimensions/time/quarters/`

- listing of all countries

  `/dimensions/space/countries`

- RDF Description of the country USA

  `/dimensions/space/countries/USA`

- listing of all US States

  `/dimensions/space/countries/USA/states`

- RDF Description of US state massachusetts

  `/dimensions/space/countries/USA/states/massachusetts`

- listing of all MA counties

  `/dimensions/space/countries/USA/states/massachusetts/counties`

  ————————-query time————————

- listing of all US counties

  `/dimensions/space/countries/USA/counties`

- listing of all months in year 1990

  `/dimensions/time/years/`990`/months/`1

Test citation [1]

References

[1] Sabine Timpf and Andrew U. Frank. A multi-scale data structure for car-
    tographic objects. In *17th International Cartographic Conference ICC'95,
    at*, pages 1381–1386, 1995.