Effortless Data Exploration with **zenvisage**: An Expressive and Interactive Visual Analytics System



Tarique Siddiqui, Albert Kim, John Lee, Karrie Karahalios, Aditya Parameswaran

zenvisage.github.io



Motivation

Everyone doing exploratory data analysis uses some combination of the following workflow:

- 1. Load dataset into an interactive viz tool like Excel or Tableau
- 2. Select visualization to be generated
- 3. See if the visualization satisfies desired "insights" or "visual property"
- 4. If yes, stop; if not, back to step 2

With LARGE datasets and LARGE # of attributes, this is a tedious and time consuming process, all for a single visual property.

Motivation

This is a **real** problem!

- Advertising Data Analysis: (our collaborators at Turn Inc.)
 - a. Finding keywords with similar click-through rates
- **Genomic Data Analysis:** (our collaborators at the NIH center at Illinois)
 - a. Finding pairs of genes that can visually explain the difference between clinical outcomes
- **Environmental Data Analysis:** (our collaborators at the Great Lakes initiative)
 - a. Finding sensors (on buoys) that are behaving anomalously
- **Engineering Data Analysis:** (our collaborators at CMU)
 - a. Finding solvents with desired behaviours ("hockey stick" shape for a certain property)
- ..

Common theme: There are multiple settings where finding the "right" visualization that reveals the desired insight can take *hours or days*!

Enter zenvisage



Zenvisage = zen + envisage (to "effortlessly" visualize)

A visual data exploration system for "fast-forwarding to desired insights"

We've been building the system for the last 2 years; being developed in collaboration with the 4 collaborator groups

A significant generalization of the previous system SeeDB

Zenvisage: our design goals



Expressive: Specify desired insights using a declarative "data exploration" language for experts

Interactive: For non-experts, support simple interaction primitives to support effortless data exploration

Scalable: Must be able to traverse through a large space of visualizations and recommend interesting ones instantly

Expressiveness via ZQL

We've developed a data exploration language called ZQL (Zenvisage Query Language) enabling users to specify the desired visual insights

Using a small number of ZQL lines (often < 2 lines), users can specify desired trends, patterns, insights from visualizations

ZQL draws from QBE (Query By Example)/SQL + ggplot/polaris algebra

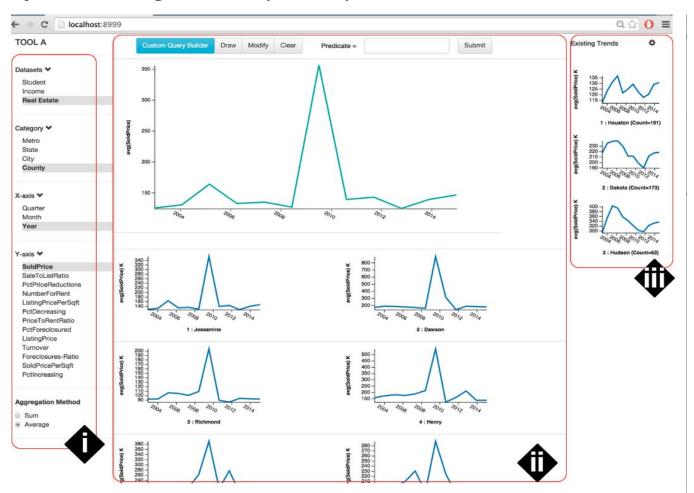
We've formally developed a visual exploration algebra, and shown that ZQL is visual exploration complete with respect to that algebra ⇒ ZQL has *nice*, *formal semantics*!

Expressiveness via ZQL

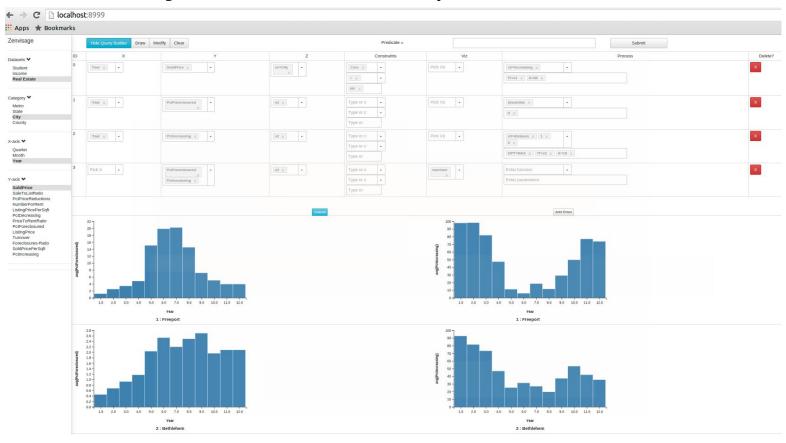
All within two-three lines:

- Find x and y attributes on which chairs and desks differ the most
- Find products whose sales over years and the profit over years trends are most dissimilar
- From among products that are similar to staplers on sales over time,
 find typical trends on profits over time
- Find products whose sales over time has an increasing trend while profit over time has a decreasing trend

Interactivity: The "Drag-and-Drop" Perspective



Interactivity: The ZQL Perspective



What else?

- Expressiveness & Interactivity -- briefly covered
- Scalability:
 - Automated guery translation and execution for ZQL
 - Proposed General Query Optimization techniques, similar to MQO

Evaluation

- Performance of optimization techniques
- Usability and effectiveness via a user study (12 people w/ varying experience with data exploration and programming); findings:
 - **Fast and accurate:** up to 100% faster than baseline, 30% more accurate results
 - **Easy:** Even non-programmers could learn a subset of ZQL and use it within a short period
 - Better: Unanimously, everyone preferred zenvisage over the baseline and wanted to incorporate it into their current data analysis workflow

What else?

Project webpage: http://zenvisage.github.io/

Technical Report: http://data-people.cs.illinois.edu/zenvisage.pdf
20+ pages (!!!) paper, circa April 2016: LOTS MORE HERE!!!
under review at VLDB, also up on ArXiV

To be open-sourced soon (in a few weeks)!!

Email me: tsiddiq2@illinois.edu