You can’t always sketch what you want: Understanding Sensemaking in Visual Query Systems

Doris Jung-Lin Lee, John Lee, Tarique Siddiqui, Jaewoo Kim, Karrie Karahalios, Aditya Parameswaran
Long-term (1880-2017) global warming compared to short-term temperature trends

Dow Jones Industrial Average (1927 to 1932)

10/29: Day of the Stock Market Crash

MY OVERALL HEALTH

THE DAY I REALIZED I COULD COOK BACON WHENEVER I WANTED.
Long-term (1880-2017) global warming compared to short-term temperature trends

15-YEAR TRENDS
°F per decade


average global temperature (°F)


Dow Jones Industrial Average (1927 to 1932)

10/29: Day of the Stock Market Crash

MY OVERALL HEALTH

THE DAY I REALIZED I COULD COOK BACON WHenever I wanted.
Long-term (1880-2017) global warming compared to short-term temperature trends

Dow Jones Industrial Average (1927 to 1932)

MY OVERALL HEALTH

THE DAY I REALIZED I COULD COOK BACON WHENEVER I WANTED.
Astronomy Use Case

Luminosity vs. Time

- Massive star: ~10% error, ~50 days
- White dwarf: error < 0.5%
Visual Query Systems (VQSs)

- TimeSearcher [Hochheiser & Shneiderman 2004]
- QuerySketch [Wattenberg 2001]
- QueryLines [Ryall et al. 2005]
- SoftSelect [Holz & Feiner 2009]
- Google Correlate [Mohebbi 2011]
- TimeSketch [Eichmann & Zgraggen 2015]
- SketchQuery [Correll & Gleicher 2016]
- Zenvisage [Siddiqui et al 2017]
- Qetch [Mannino & Abouzied 2018]
Visual Query Systems (VQSs)

Most Similar

TimeSearcher [Hochheiser & Shneiderman 2004]
QuerySketch [Wattenberg 2001]
QueryLines [Ryall et al. 2005]
SoftSelect [Holz & Feiner 2009]
Google Correlate [Mohebbi 2011]

Least Similar

TimeSketch [Eichmann & Zgraggen 2015]
SketchQuery [Correll & Gleicher 2016]
Zenvisage [Siddiqui et al. 2017]
Qetch [Mannino & Abouzied 2018]
Outline

1) Methodology

   Motivation: Challenge of VQS adoption
   Our Approach: User-Centered Design Process

2) Study Findings

   Taxonomy
   Usage observation and analysis
   Design guidelines
Outline

1) Methodology

Motivation: Challenge of VQS adoption
Our Approach: User-Centered Design Process

2) Study Findings

Taxonomy
Usage observation and analysis
Design guidelines
Munzner, *A Nested Model for Visualization Design and Validation* (TVCG 2009)
Munzner, *A Nested Model for Visualization Design and Validation* (TVCG 2009)
1) Methodology

Motivation: Challenge of VQS adoption

Our Approach: User-Centered Design Process

2) Study Findings

Taxonomy
Usage observation and analysis
Design guidelines
Our Approach

Phase I
Need-Finding

Interviews & Contextual Inquiry

Phase II
Prototyping

Participatory Design

Phase III
Evaluation

Grounded Evaluation

Understanding Users, Datasets, Challenges

Outcome

Zenvisage++

VQS Sensemaking Model

Zenvisage++

Top-down

Bottom-up

Context Creation
Phase I
Need-Finding

Interviews & Contextual Inquiry

Our Approach

Understanding Users, Datasets, Challenges

Phase II
Prototyping

Participatory Design

Grounded Evaluation

Phase III
Evaluation

VQS Sensemaking Model

Zenvisage++
Phase I
Need-Finding

Interviews & Contextual Inquiry

Our Approach
Understanding Users, Datasets, Challenges

Phase II
Prototyping

Participatory Design

Grounded Evaluation

Phase III
Evaluation

VQS Sensemaking Model

Top-down
Bottom-up
Context Creation
Phase II Prototyping

Participatory Design

Our Approach

Interviews & Contextual Inquiry

Grounded Evaluation

Understanding Users, Datasets, Challenges

Zenvisage++

VQS Sensemaking Model
Phase II
Prototyping

Participatory Design

Zenvisage++

VQS Sensemaking Model

Our Approach

Interviews & Contextual Inquiry

Grounded Evaluation

Phase I
Need-Finding

Phase III
Evaluation

Understanding Users, Datasets, Challenges
Phase III Evaluation

Grounded Evaluation

Our Approach

Phase I: Need-Finding
- Interviews & Contextual Inquiry

Phase II: Prototyping
- Participatory Design

Phase III: Evaluation

VQS Sensemaking Model

Top-down
Bottom-up
Context Creation

Outcome

Understanding Users, Datasets, Challenges

Zenvisage++
Phase III Evaluation

Grounded Evaluation

VQS Sensemaking Model

Top-down

Bottom-up

Context Creation

Phase I
Need-Finding

Interviews & Contextual Inquiry

Our Approach

Phase II
Prototyping

Participatory Design

Outcome

Understanding Users, Datasets, Challenges

Zenvisage++
What are the characteristic differences between stars that harbor planets versus ones that don’t?

Which battery components have sustainable levels of energy-efficiency and are safe and cheap to manufacture in production?

How does a treatment affect the expression of a gene in a breast cancer cell-line?
Outline

1) Methodology

Motivation: Challenge of VQS adoption
Our Approach: User-Centered Design Process

2) Study Findings

Taxonomy
Usage observation and analysis
Design guidelines
Taxonomy of Sensemaking Process and Key Capabilities in VQSs

Top-down
- Pattern Specification
  - What is the shape of the pattern query?
- Match Specification
  - How should the pattern query be matched with other visualizations?

Bottom-up
- Result Querying
  - What other vizzes “look similar” to the selected pattern?
- Recommendation
  - What are the key patterns in this dataset?

Context Creation
- View Specification
  - What data to visualize and how should it be displayed?
- Slice-and-Dice
  - How does navigating to another data subset change the query result?
Top-down

Pattern Specification

What is the shape of the pattern query?

Match Specification

How should the pattern query be matched with other visualizations?
What is the shape of the pattern query?

How should the pattern query be matched with other visualizations?
Result Querying

What other vizzes “look similar” to the selected pattern?

Recommendation

What are the key patterns in this dataset?

Representative Patterns

234 like this
145 like this
What other vizzes “look similar” to the selected pattern?

What are the key patterns in this dataset?
Context Creation

View Specification

Slice-and-Dice

What data to visualize and how should it be displayed?

is_star = True
X: Time
Y: Expression
Z: Gene

is_star = False
X: Time
Y: Expression
Z: Gene

How does navigating to another data subset change the query result?

Dataset: real_estate
Pattern: 
Category: 
X-axis: 
Y-axis: 
soldpricependsoft
Error Attribute: none

Expression Time Gene
Time Expression

Results

city: Laguna Hills (0.468)
city: Saint Augustine (0.451)
city: Sacramento (0.427)
city: Rohnert Park (0.4025)
city: Lomav (0.401)
city: Sandy Springs (0.374)
Context Creation

View Specification

Slice-and-Dice

What data to visualize and how should it be displayed?

How does navigating to another data subset change the query result?

- is_star = True
- is_star = False

X: Time
Y: Expression
Z: Gene

Expression
Gene
Time
1) Methodology

Motivation: Challenge of VQS adoption
Our Approach: User-Centered Design Process

2) Study Findings

Taxonomy
Usage observation and analysis
Design guidelines
Bottom-up Data-Driven Inquiries via Recommendations
Bottom-up Data-Driven Inquiries via Recommendations

Expression Level vs. Time

Results

- **gene: Capns1** (0.78)
  - Value by timestep
- **gene: Derl13** (0.63)
  - Value by timestep
- **gene: Zifp704** (0.570)
  - Value by timestep
- **gene: HsfBst1** (0.557)
  - Value by timestep
- **gene: Tead1** (0.53)
  - Value by timestep
- **gene: Dct** (0.499)
  - Value by timestep

Representative patterns

- **Sgfk1** (24 more like this)
  - Value by timestep
- **Zifp43** (22 more like this)
  - Value by timestep

Outliers

- **Etl4** (38 more like this)
  - Value by timestep
- **Socs3**
  - Value by timestep
- **Neat1**
  - Value by timestep
Narrow Search Space via Context Creation

Results

- pdf flux by objid_67000
  - objid_cycle_band: 371420660:0 [G] (0.41)
  - objid_cycle_band: 370221280:0 [I] (0.27)
- pdf flux by objid_67000
  - objid_cycle_band: 374104387:3 [Z] (0.225)
  - objid_cycle_band: 374113780:0 [I] (0.205)
- pdf flux by objid_67000
  - objid_cycle_band: 374105867:0 [I] (0.177)
  - objid_cycle_band: 367499786:0 [I] (0.149)

Representative patterns

- Luminosity: ~10%
- Time: ~50 days
Narrow Search Space via Context Creation

Results

- pdf_flux by mj57000
  - objid_cycle_band: 371420660:0 [G] (0.41)
  - objid_cycle_band: 370221280:0 [I] (0.27)

- pdf_flux by mj57000
  - objid_cycle_band: 374104367:3 [Z] (0.225)
  - objid_cycle_band: 374113708:0 [I] (0.205)

- pdf_flux by mj57000
  - objid_cycle_band: 374105867:0 [I] (0.177)
  - objid_cycle_band: 367499786:0 [I] (0.149)

Representative patterns

- Luminosity:
  - ~10%
  - ~50 days

Options:
- Consider x-range
- Show original sketch
- Show scatterplot
- Show Bar Chart
- Reverse y-axis

Input Equation:
- e.g. y=x^2
  - add

Cluster Size:
- 3

Data Smoothing:
- None

Number of Results:
- 50

Smoothing Constant:
- 0.5

Similarity Cutoff:
- 0

Filter Constraint:
- e.g. m(jo>20)
  - Submit

Similarity:
- Euclidean Distance
- Segmentation
- DTW
- MVIP

Aggregate:
- Sum
- Average
- None
**Finding:** You can't always sketch what you want!

All sensemaking processes are **essential**.
How do participants transition between different sensemaking processes?
How do participants transition between different sensemaking processes?
Each sensemaking process fulfills a **central goal** in participants’ analysis, while other sensemaking processes are used as **support**.
Each sensemaking process fulfills a **central goal** in participants’ analysis, while other sensemaking processes are used as **support**.
Each sensemaking process fulfills a **central goal** in participants’ analysis, while other sensemaking processes are used as **support**.
Each sensemaking process fulfills a central goal in participants’ analysis, while other sensemaking processes are used as support.
Each sensemaking process fulfills a **central goal** in participants’ analysis, while other sensemaking processes are used as **support**.
Each sensemaking process fulfills a **central goal** in participants’ analysis, while other sensemaking processes are used as *support*. 

---

**Created classes**

<table>
<thead>
<tr>
<th>Ranges [li, bp]</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10.0 &lt; li &lt;= 37.23</td>
<td>7938</td>
</tr>
<tr>
<td>280.0 &lt; bp &lt;= 1232.69</td>
<td></td>
</tr>
<tr>
<td>-17.79 &lt; li &lt;= -10.0</td>
<td>19</td>
</tr>
<tr>
<td>280.0 &lt; bp &lt;= 1232.69</td>
<td></td>
</tr>
<tr>
<td>-10.0 &lt; li &lt;= 37.23</td>
<td>40</td>
</tr>
<tr>
<td>102.24 &lt; bp &lt;= 280.0</td>
<td></td>
</tr>
</tbody>
</table>
1) Methodology

Motivation: Challenges to VQS adoption
Our Approach: User-Centered Design Process

2) Study Findings

Taxonomy
Usage observation and analysis
Design guidelines
Finding: You can't always sketch what you want!
Finding: You can't always sketch what you want!
Finding: You can't always sketch what you want!

Need to "close the loop" between all sensemaking processes
Methodology: Beyond the lab

User-centered design crucial for tool adoption
Methodology: Beyond the lab

User-centered design crucial for tool adoption
Methodology: Beyond the lab

User-centered design crucial for tool adoption
Finding: Closing the loop

Towards integrative VQS
You can’t always sketch what you want: Understanding Sensemaking in Visual Query Systems

Doris Jung-Lin Lee, John Lee, Tarique Siddiqui, Jaewoo Kim, Karrie Karahalios, Aditya Parameswaran

Contact:

@dorisjlee
dorislee@berkeley.edu