Visualization By Example

Chenglong Wang\textsuperscript{1}, Yu Feng\textsuperscript{2}, Alvin Cheung\textsuperscript{3}, Ras Bodik\textsuperscript{1}, Isil Dillig\textsuperscript{4}

\textsuperscript{1}University of Washington
\textsuperscript{2}University of California, Santa Barbara
\textsuperscript{3}University of California, Berkeley
\textsuperscript{4}University of Texas, Austin

Read the paper!
Visualizations

1. How to formalize visualizations?
2. How to synthesize visualizations to bring it to the masses?

Product price in different region  Net cash flow in a year  Survey result

Performance rating distribution for each department  Housing price in different region
Formalizing Visualizations

“Transformation of the symbolic into the geometric”
[McCormick et al. 1987]

A set of geometric objects

Table

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
</tr>
</tbody>
</table>

Visual Program
ggplot2, Matplotlib, Excel, Vega-Lite
Visualization in Practice

Visual program alone is often insufficient.

\[ T \xrightarrow{\phi_V} V \]

Table \quad A set of geometric objects

<table>
<thead>
<tr>
<th>X</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

\( \phi_V \) expects a certain shape of the input table

Expects 3 columns that map to bar.x, bar.height, bar.color

bar(x=1, h=1, color=A)
bar(x=1, h=4, color=B)
bar(x=1, h=3, color=C)
bar(x=2, h=2, color=A)

......
Visualization in Practice

\[ T \xrightarrow{\phi_T} T' \xrightarrow{\phi_V} V \]

\[ \text{Data Adapter} \]

\[ \text{Visual Program} \]

\[ \text{Gather (1) turn A,B,C into the Key column (2) move values in A,B,C columns into the Val column} \]

\[ \text{Data adapter prepares the input table to match the shape expected by } \phi_V \]
Visualization Challenges

1. Users need to master both data prep libraries and visualization libraries.

2. Reshaping and aggregation of data requires deep data transformation insight. [Feng et al. 2018]

3. Change of visualization designs requires frequent change of data adapters.
Visualization Challenges [Gatto 2015]

1. Diverse data shapes in practice

2. Limited Software Knowledge

3. Limited knowledge about visualization concepts

Visualization Synthesis
(1) handle diverse data shapes
(2) expressive $\phi_T$ and $\phi_V$
(3) user specification requires little visualization concepts

How to specify visualization?

How would you explain intent of this visualization?

Partial visualization
A subset of geometric objects of the final visualization
Visualization by Example

\[ T \xrightarrow{\phi_T} T' \xrightarrow{\phi_V} V \]
Visualization by Example

\[ T \xrightarrow{\phi_T} T' \xrightarrow{\phi_V} V_{\text{partial}} \subseteq V \]
Visualization by Example

Given $T$, $V_{\text{partial}}$, synthesize $\phi_T$, $\phi_V$, such that $\phi_V(\phi_T(T)) \supseteq V_{\text{partial}}$
1. Compositional Synthesis:
\[\phi_T \in \mathcal{L}_T \text{ and } \phi_V \in \mathcal{L}_V\]

2. Potentially large input table
\[\text{e.g. } 3000 \times 10\]

3. Weak Specification:
“\(\subseteq\)” instead of “="

**Requirement:**
\[\phi_V(\phi_T(T)) \supseteq V_{\text{partial}}\]
Visualization synthesis

Requirement:
\[ \phi_V(\phi_T(T)) \supseteq V_{\text{partial}} \]
Visualization synthesis

Step 1: decompile visualization
s.t., $\phi_V(T_{\text{sketch}}) = V_{\text{partial}}$

Requirement:
$\phi_V(\phi_T(T)) \supseteq V_{\text{partial}}$
Visualization synthesis

Step 2: Synthesize data adapter
s.t., $T_{sketch} \subseteq \phi_T(T)$

Key: push the containment requirement from visualization to data adapter.

Requirement:
$\phi_V(\phi_T(T)) \supseteq V_{partial}$
Step 1: Decompile Visualization

\[ T \xrightarrow{\subseteq} T' \xrightarrow{\phi_V} V \]

Requirement:
\[ \phi_V(T_{sketch}) = V_{partial} \]

Key: formalize visualization as mappings (and leave the challenges for tables)

What mapping generates \( V_{partial} \)?

\[ \phi_V \]

What data generates \( V_{partial} \)?

\[ \phi_V^{-1} \]

\( T_{sketch} \)

bar\( (x=1, h=1, \text{color}=A) \)
bar\( (x=1, h=4, \text{color}=B) \)

Color:
- a
- b

<table>
<thead>
<tr>
<th>x</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Other alternatives …
Step 2: Data Adapter Synthesis

\[ \phi_T(T) \supseteq T_{sketch} \]

\begin{tabular}{|c|c|c|}
\hline
C1 & C2 & C3 \\ \hline
1 & A & 1 \\ 1 & B & 4 \\ \hline
\end{tabular}

bar(x=1, h=1, color=A) 
bar(x=1, h=4, color=B)

Requirement: \( \phi_T(T) \supseteq T_{sketch} \)
Step 2: Data Adapter Synthesis

Requirement: $\phi_T(T) \supseteq T_{sketch}$

Holes “☐” are uninstantiated parameters of the partial program.
Step 2: Data Adapter Synthesis

For any predicate, we have

\[ \phi_T(T) \subseteq T_{\text{sketch}} \]

Requirement: \( \phi_T(T) \supseteq T_{\text{sketch}} \)

Forward reasoning
Given \( T \) and partial \( \phi_T \), what’s the property of the output \( \phi_T(T) \)?

Contradicts:

[Wang PLDI17, Feng PLDI18]
Step 2: Data Adapter Synthesis

Requirement: \( \phi_T(T) \supseteq T_{sketch} \)

\[ T_{sketch} = \]

Thus, \( T_{sketch} \subseteq T_{in} \)

Contradicts:

Backward reasoning:
Given property \( \phi_T(T) \supseteq T_{sketch} \) and partial \( \phi_T \),
what's the property of \( T \)?
Step 2: Data Adapter Synthesis

Requirement: $\phi_T(T) \supseteq T_{\text{sketch}}$

Contribution: **Bidirectional reasoning**
Inductively defined for all operators in $\mathcal{L}_T$

Start

- gather($T$, key=$\Box$, val=$\Box$)
- filter($T$, $\Box$)
- spread($T$, id=$\Box$, key=$\Box$, val=$\Box$)

......

- gather($T$, key=A, val=$\Box$)
- gather($T$, key=A, val=$\Box$)
- gather($T$, key=C, val=$\Box$)
- gather($T$, key=C, val=$\Box$)
- gather($T$, key=$\Box$, val=$\Box$)

......

- gather($T$, key=X, val=[A])
- gather($T$, key=X, val=[A])
- gather($T$, key=X, val=[A])
- gather($T$, key=X, val=[A])
- gather($T$, key=X, val=[A])
- gather($T$, key=X, val=[A])
- gather($T$, key=X, val=[A, B])
- gather($T$, key=X, val=[A, B])
- gather($T$, key=X, val=[A, B, C])

$T \xrightarrow{\phi_T} T' \xrightarrow{\phi_V} V_{\text{partial}} \subseteq V$
Visualization by Example

\[ \phi_V(\phi_T(\textbf{T})) \supseteq V_{\text{partial}} \]

Potentially multiple \((\phi_T, \phi_V)\) pairs can satisfy the specification.

\(T\)

\[
\begin{array}{cccc}
X & A & B & C \\
1 & 1 & 4 & 3 \\
2 & 2 & 3 & 2 \\
3 & 5 & 2 & 1 \\
4 & 3 & 6 & 1 \\
\end{array}
\]

\(\phi_T\)

gather(\(T\), \(id=X\), \(key=[A,B,C]\))

\(\phi_V\)

\[
\begin{array}{cccc}
X & Key & Val \\
1 & A & 1 \\
1 & B & 4 \\
1 & C & 3 \\
2 & A & 2 \\
\end{array}
\]

\[
X \rightarrow \text{bar.x}, \quad \text{Val} \rightarrow \text{bar.height}, \quad \text{Key} \rightarrow \text{bar.color}
\]

\(V_{\text{partial}}\)

\[
\text{bar(x=1, h=1, color=A)} \quad \text{bar(x=1, h=4, color=B)} \quad \ldots
\]

\(\subseteq\)

\[
\text{bar(x=1, h=1, color=A)} \quad \text{bar(x=1, h=4, color=B)} \quad \text{bar(x=2, h=3, color=C)} \quad \text{bar(x=2, h=2, color=A)} \quad \ldots
\]
Experiment: Viser

**Question 1 (Performance):**
Can fast can Viser solve practical visualization problems?

**Question 2 (Usability):**
How many geometric objects does the user need to demonstrate?

- **Data Adapter** (R tidyverse library)
  - filter, join, gather, spread, mutate, unite, separate, select
- **Visual Program** (Vega-Lite)
  - Line, Bar, Scatter, Area
  - Stacked charts, Faceted chart, Layered chart
- **83 benchmarks from**
  - Stack Overflow
  - Excel/R tutorials
- **Evaluation:**
  - 600s timeout
  - Partial visualization sampled from full visualization

*Input table size*
Ranges from **4x3** to **3686x9**, average size **100x10**

*Program Complexity*
1-4 statements,
On average 20 decisions to make for each program
Partial visualization size = 4
(i.e., 4 random geometric objects from the full visualization)

Solves 70 out of 83 benchmarks

26 benchmarks with in 1 seconds
49 benchmarks with in 10 seconds
**Performance Experiment**

Partial visualization size = 4

**Baseline Viser-M**
a variation of Viser without bidirectional pruning
[Feng PLDI18]

Solves 17 more benchmarks

On average 7X faster

Unsolved benchmarks
Large input table & complex transformations

Look ahead
Multi-modal synthesis
Usability Experiment

Can solve a lot of benchmarks with a small number of examples.

Increasing size of $V_{\text{partial}}$, makes expected solutions rank higher.

Difficult tasks often involve custom filtering.

Look ahead
Negative examples, Interactive refinement.
Visualizations

Survey result

Net cash flow in a year

Housing price in different region

Housing price in different region

Product price in different region
Visualization by Example

Given $T$, $V_{\text{partial}}$, synthesize $\phi_T, \phi_V$, such that $\phi_V(\phi_T(T)) \supseteq V_{\text{partial}}$

Contribution 1
Define and formalize the visualization synthesis problem

Contribution 2
Compositional synthesis of visualizations

Contribution 3
Bidirectional analysis for pruning partial programs

More questions?
Talk to: Chenglong Wang
clwang@cs.washington.edu