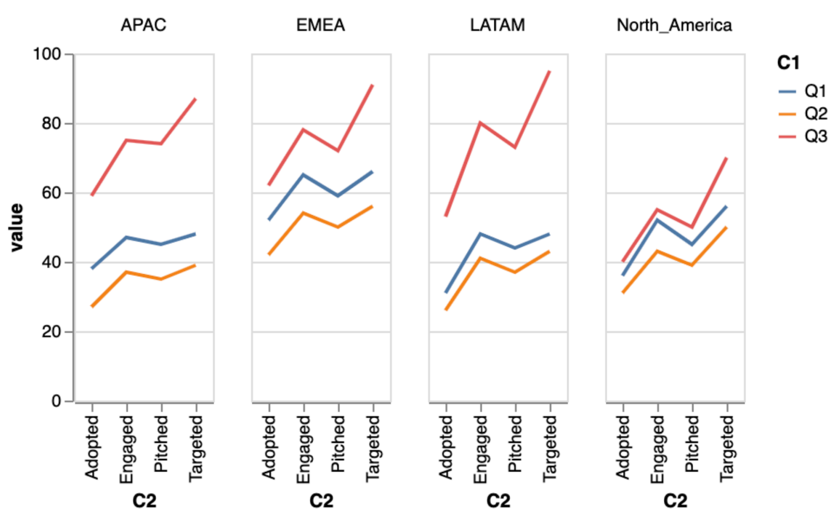
Visualization By Example

Chenglong Wang[1], Yu Feng[2], Alvin Cheung[3], Ras Bodik[1], Isil Dillig[4]

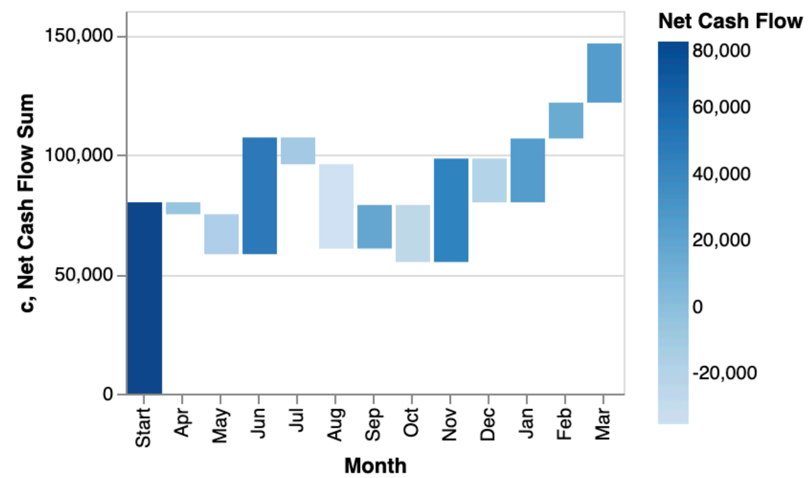
[1]University of Washington
[2]University of California, Santa Barbara
[3]University of California, Berkeley
[4]University of Texas, Austin

Read the paper!

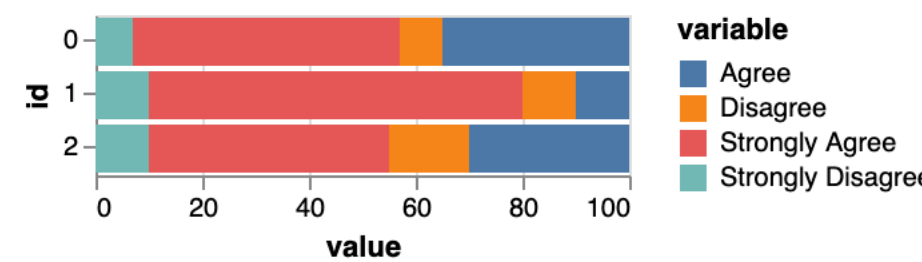
Visualizations



Product price in different region

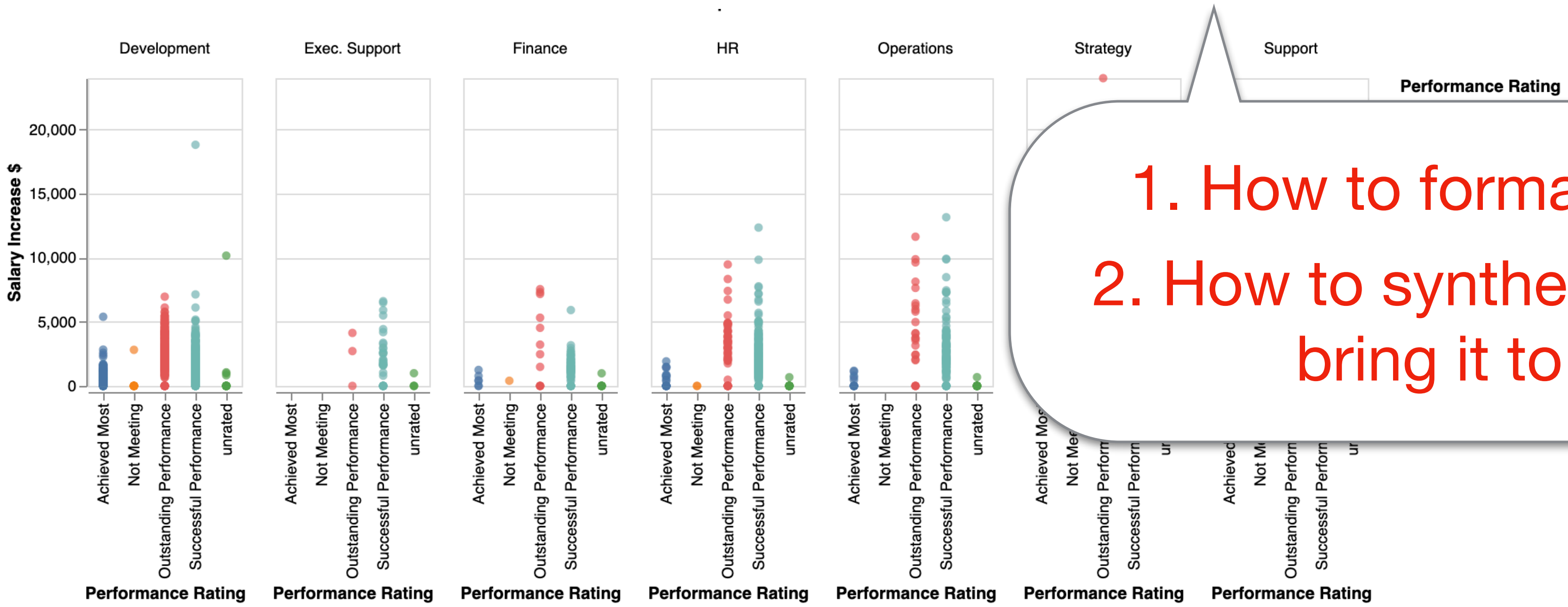


Net cash flow in a year

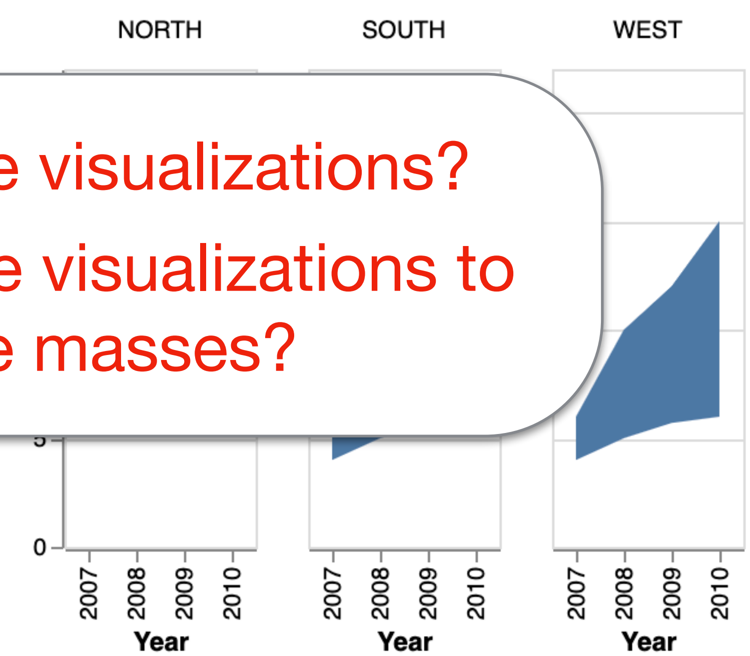


Survey result

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



Performance rating distribution for each department



Housing price in different region

Formalizing Visualizations

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

Visual Program
ggplot2, Matplotlib, Excel, Vega-Lite

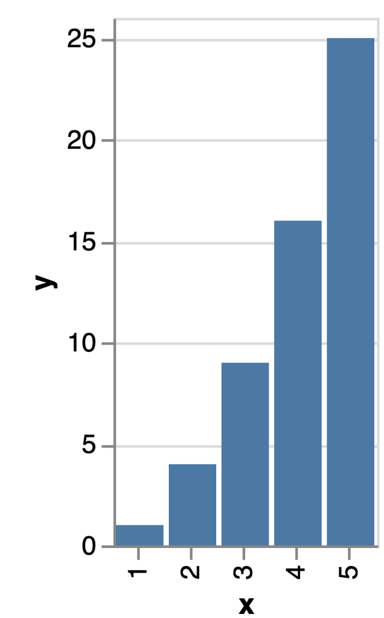
$$T \xrightarrow{\phi_V} V$$

Table ***A set of geometric objects***

A	B
1	1
2	4
3	9
4	16
5	25

A → *bar.x*
B → *bar.height*

bar(x=1, h=1)
bar(x=2, h=4)
bar(x=3, h=9)
bar(x=4, h=16)
bar(x=5, h=25)



Visualization in Practice

Visual program alone is often insufficient.

$$T \xrightarrow{\phi_V} V$$

Visual Program

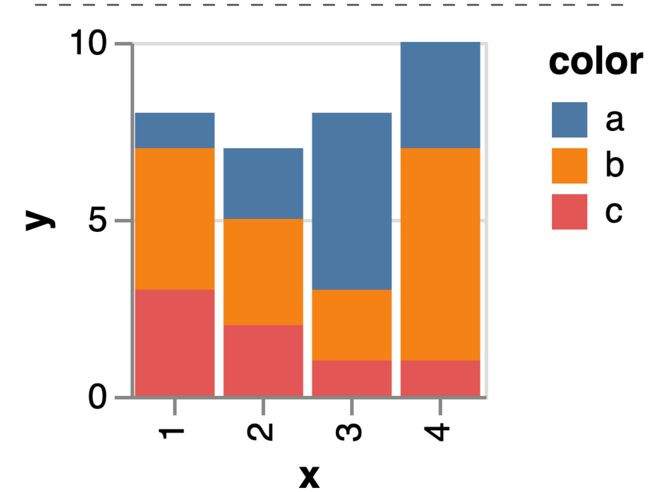
Table

A set of geometric objects

X	A	B	C
1	1	4	3
2	2	3	2
3	5	2	1
4	3	6	1

?

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)
……



ϕ_V expects a certain shape of the input table

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

Visualization in Practice

Data Adapter

Visual Program

$$T \xrightarrow{\phi_T} T' \xrightarrow{\phi_V} V$$

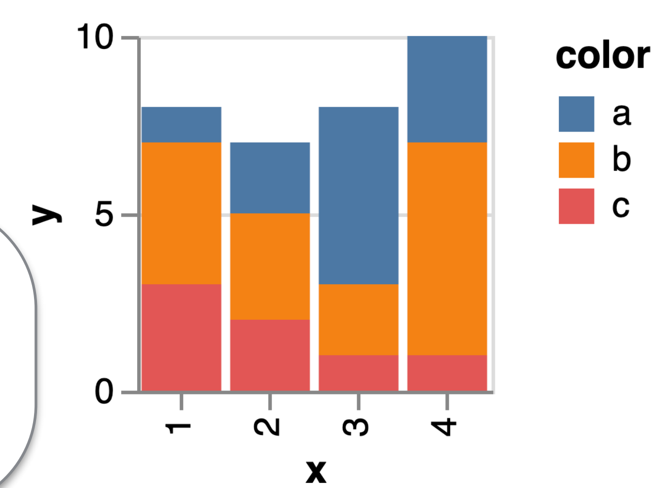
X	A	B	C
1	1	4	3
2	2	3	2
3	5	2	1
4	3	6	1

gather(
T,
X,
[A,B,C])

X	Key	Val
1	A	1
1	B	4
1	C	3
2	A	2
…	…	…

X → bar.x
Val → bar.height
Key → 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)
……



Gather
(1) turn **A,B,C** into the **Key** column
(2) move values in **A,B,C** columns into the **Val** column

Data adapter prepares the input table to match the shape expected by ϕ_V

Visualization Challenges

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

Python, R, …

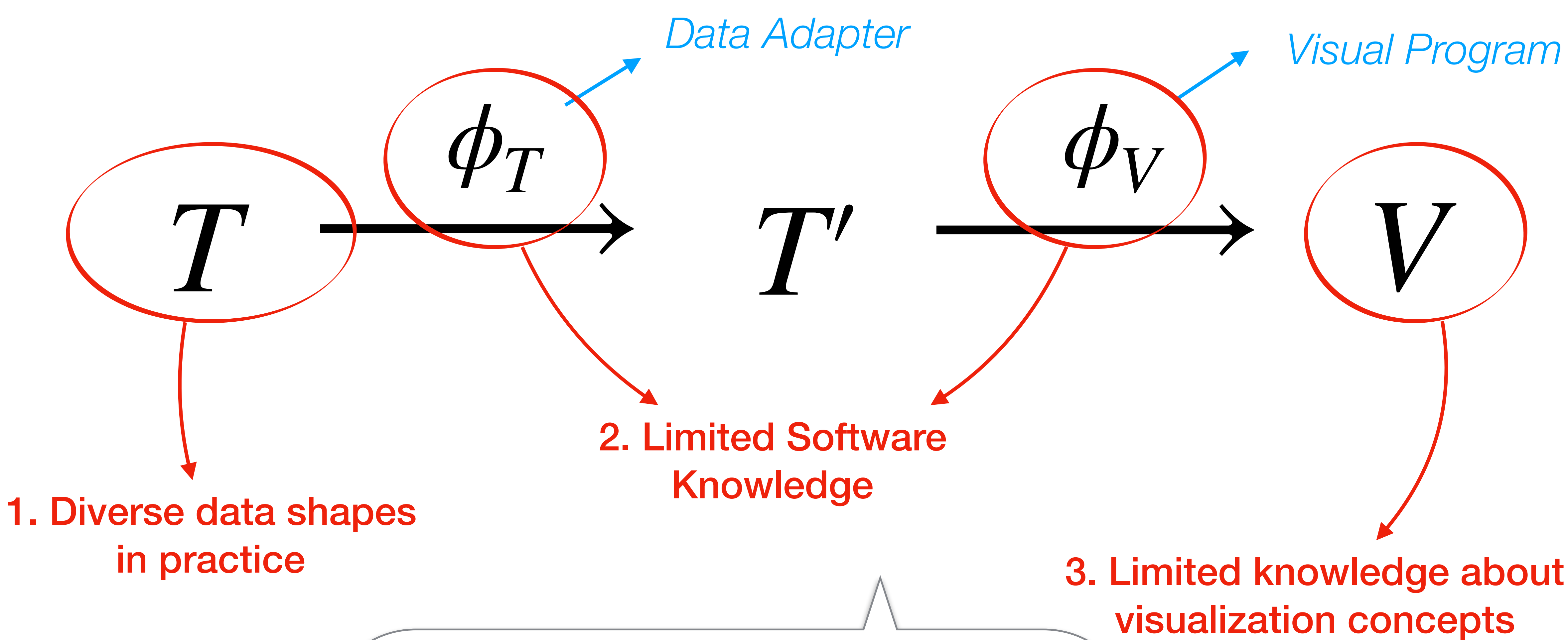
ggplot2, Vega-Lite, matplotlib …

$$T \xrightarrow{\phi_T} T' \xrightarrow{\phi_V} V$$

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]



Visualization Synthesis

(1) handle diverse data shapes

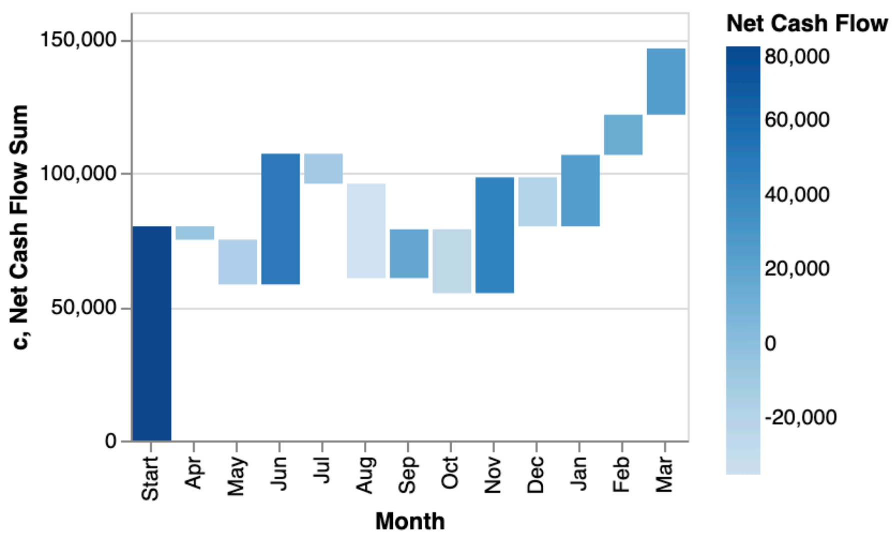
(2) expressive ϕ_T and ϕ_V

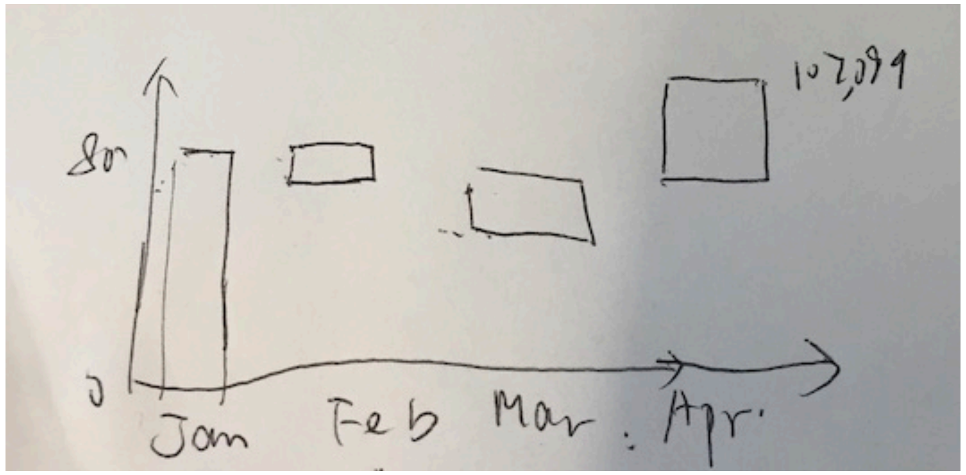
(3) user specification requires little visualization concepts

Gatto, Malu AC. "Making research useful: Current challenges and good practices in data visualization." (2015).

How to specify visualization?

How would you explain intent of this visualization?

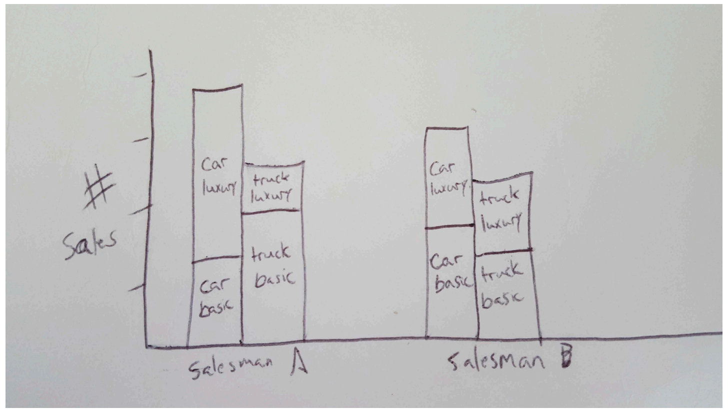


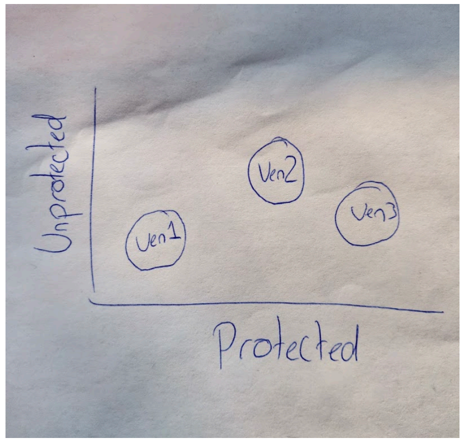


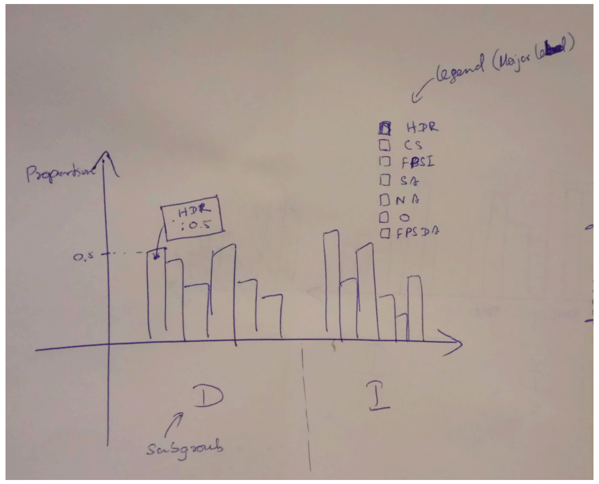
Partial visualization

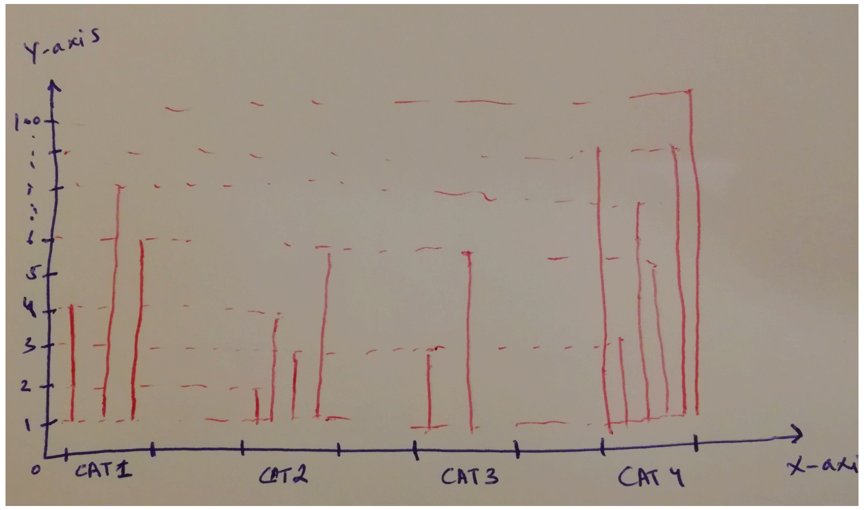
A subset of geometric objects of the final visualization

stackoverflow









Visualization by Example

$$T \xrightarrow{\phi_T} T' \xrightarrow{\phi_V} V$$

Visualization by Example

Input

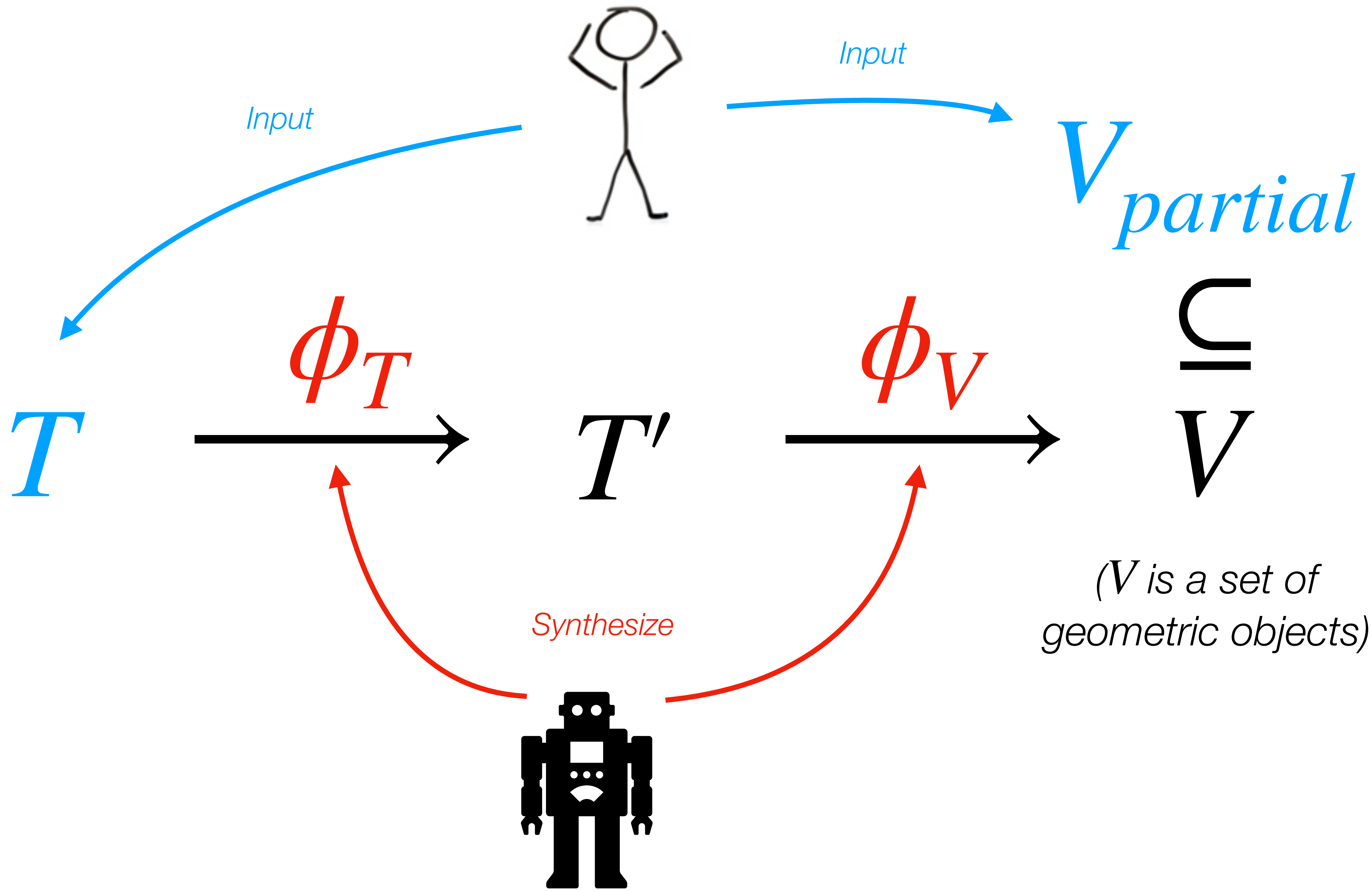
Input

$$V_{partial}$$

$$\subseteq$$

$$T \xrightarrow{\phi_T} T' \xrightarrow{\phi_V} V$$

Visualization by Example



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

Visualization by Example

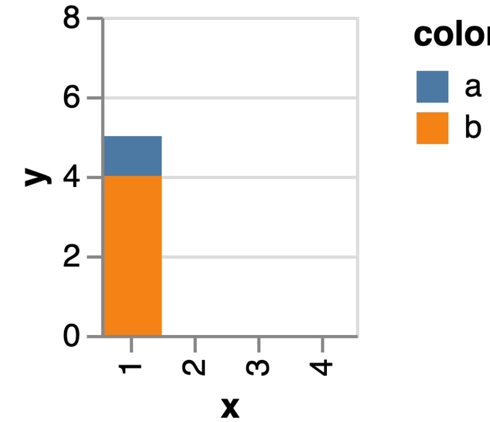
3. Weak Specification:

“⊆” instead of “="

2. Potentially large input table

e.g. 3000 x 10

$V_{partial}$



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

$\subseteq$

T

X	A	B	C
1	1	4	3
2	2	3	2
3	5	2	1
4	3	6	1

ϕ_T

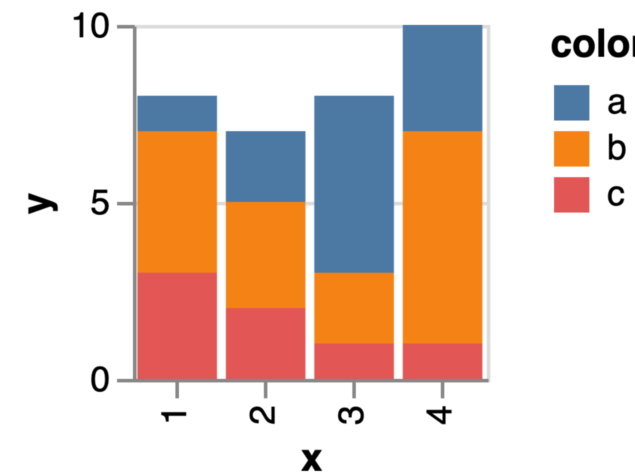
gather(
T,
id=X,
key=[A,B,C])

X	Key	Val
1	A	1
1	B	4
1	C	3
2	A	2
…	…	…

ϕ_V

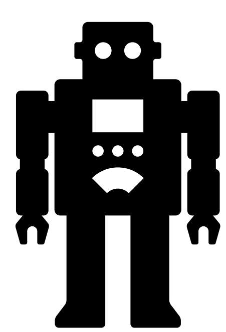
X → bar.x
Val → bar.height
Key → bar.color

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)
……



1. Compositional Synthesis:

$\phi_T \in \mathscr{L}_T$ and $\phi_V \in \mathscr{L}_V$



Requirement:

$\phi_V(\phi_T(\,T\,)) \supseteq V_{partial}$

Visualization synthesis

$V_{partial}$

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

$\subseteq$

T

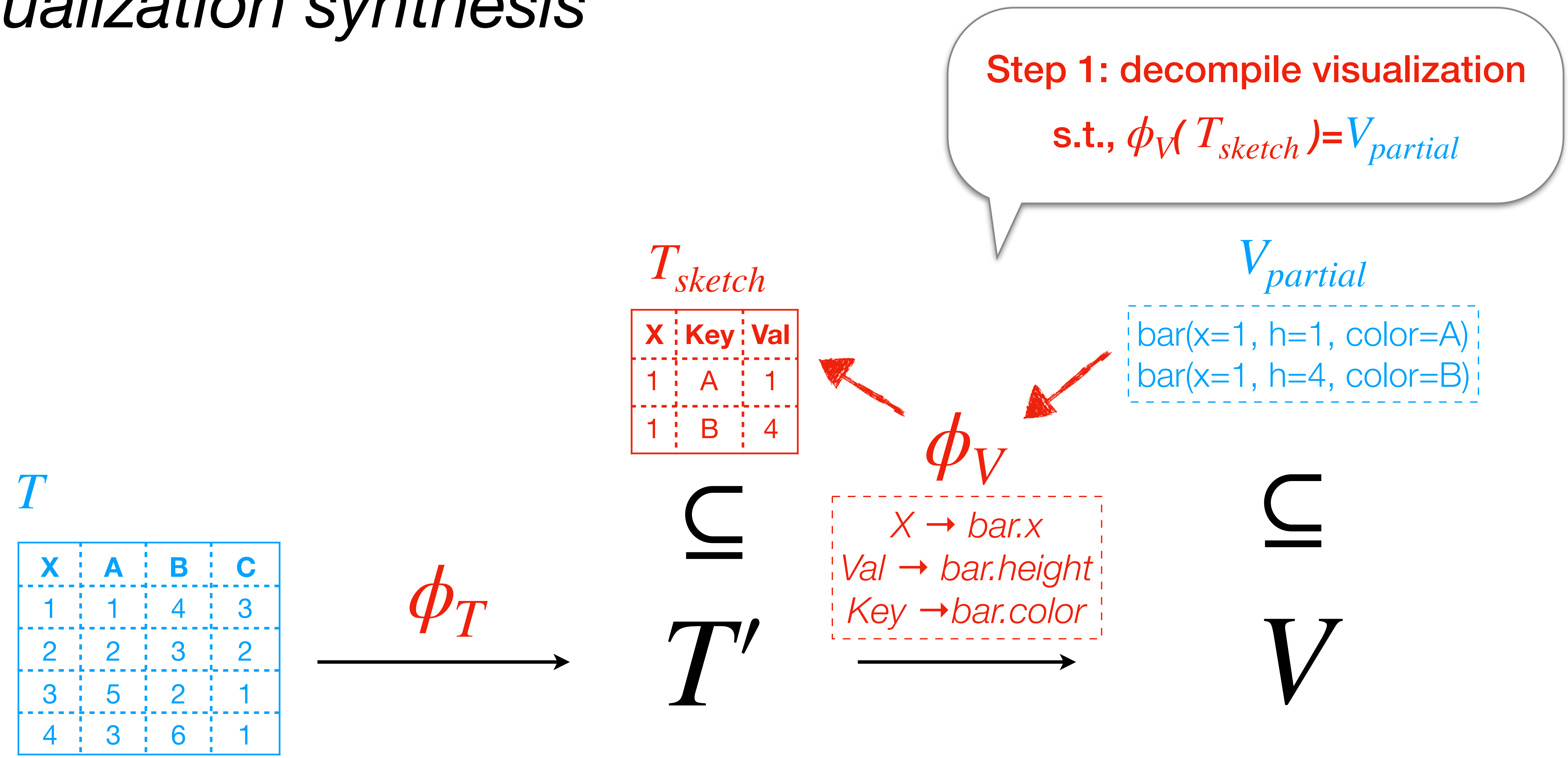
X	A	B	C
1	1	4	3
2	2	3	2
3	5	2	1
4	3	6	1

$$\xrightarrow{\phi_T} T' \xrightarrow{\phi_V} V$$

Requirement:

$\phi_V(\phi_T(\,T\,)) \supseteq V_{partial}$

Visualization synthesis



T_{sketch}:

X	Key	Val
1	A	1
1	B	4

T:

X	A	B	C
1	1	4	3
2	2	3	2
3	5	2	1
4	3	6	1

Requirement:

$$\phi_V(\phi_T(T)) \supseteq V_{partial}$$

Visualization synthesis

Step 1: decompile visualization

s.t., $\phi_V(T_{sketch}) = V_{partial}$

Step 2: Synthesize data adapter

s.t., $T_{sketch} \subseteq \phi_T(T)$

T

X	A	B	C
1	1	4	3
2	2	3	2
3	5	2	1
4	3	6	1

ϕ_T

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

T_{sketch}

X	Key	Val
1	A	1
1	B	4

$\subseteq$

T'

ϕ_V

X → bar.x
Val → bar.height
Key → bar.color

$V_{partial}$

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

$\subseteq$

V

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

Requirement:

$\phi_V(\phi_T(T)) \supseteq V_{partial}$

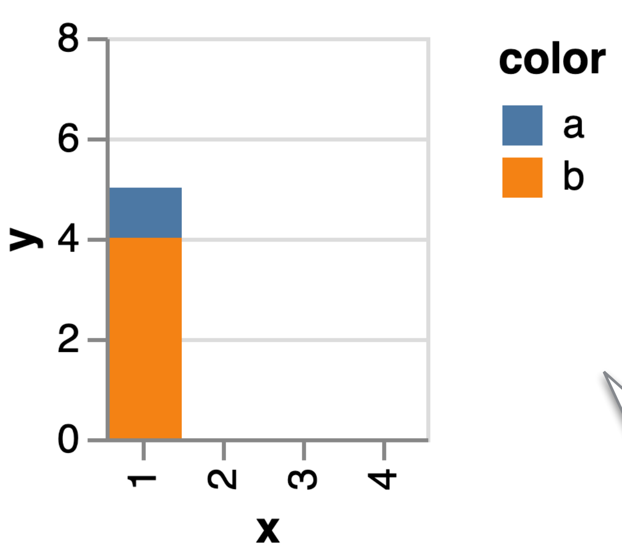
Step 1: Decompile Visualization

$$
\begin{array}{ccccc}
 & & T_{sketch} & \xleftarrow{\phi_V} & V_{partial} \\
 & & \subseteq & & \subseteq \\
T & \longrightarrow & T' & \xrightarrow{\phi_V} & V
\end{array}
$$

Requirement:

$\phi_V(\,T_{sketch}\,) = V_{partial}$

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



What mapping generates $V_{partial}$?

ϕ_V C1 → bar.x
C2 → bar.height
C3 → bar.color

What data generates $V_{partial}$?

ϕ_V^{-1}

T_{sketch}

C1	C2	C3
1	A	1
1	B	4

(other alternatives …)

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

Step 2: Data Adapter Synthesis

C1	C2	C3
1	A	1
1	B	4

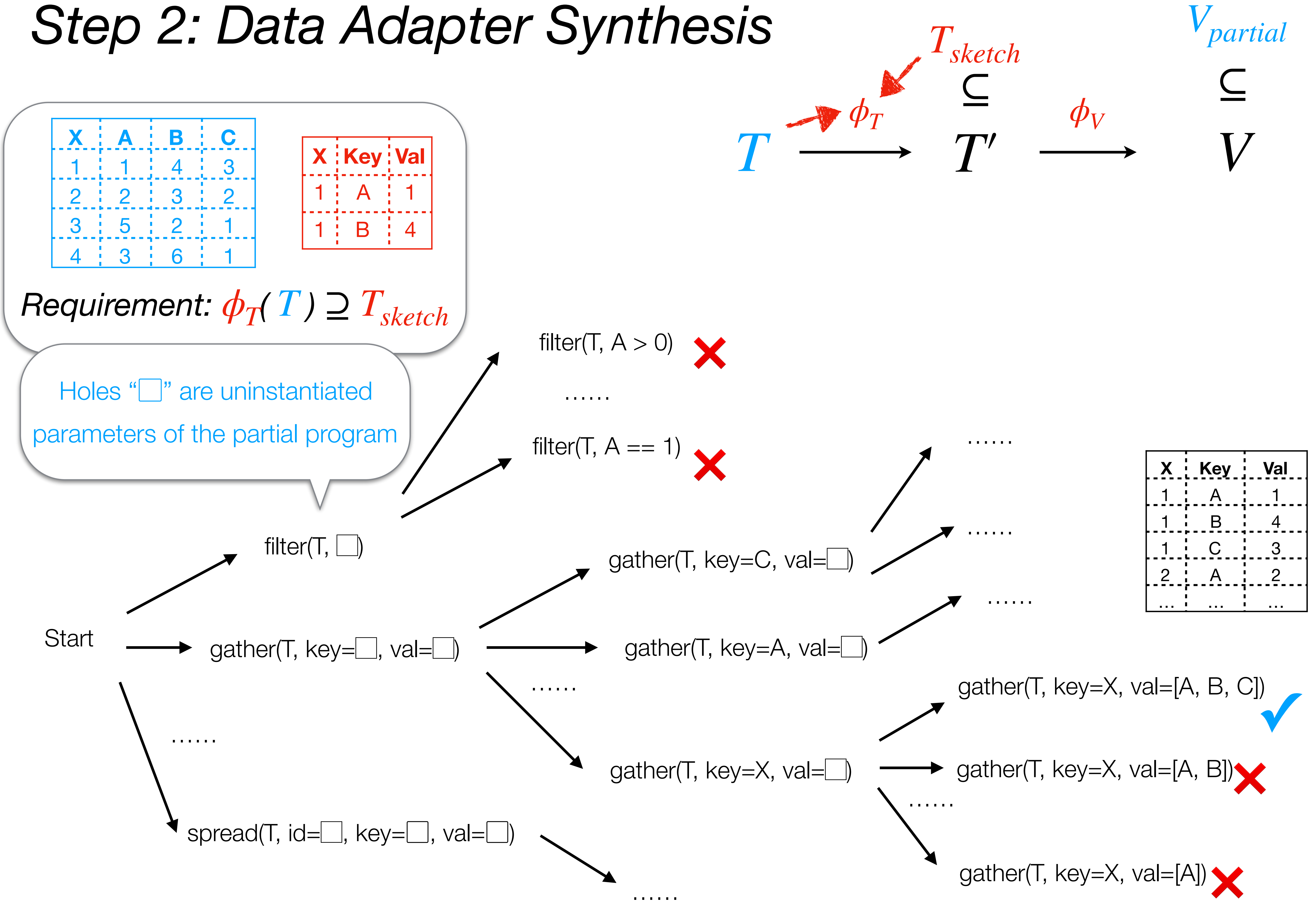
bar(x=1, h=1, color=A)
bar(x=1, h=4, color=B)

$$
\begin{array}{ccccc}
 & & T_{sketch} & & V_{partial} \\
 & & \subseteq & & \subseteq \\
T & \xrightarrow{\phi_T} & T' & \xrightarrow{\phi_V} & V
\end{array}
$$

C1 → bar.x
C2 → bar.height
C3 → bar.color

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

Step 2: Data Adapter Synthesis



X	A	B	C
1	1	4	3
2	2	3	2
3	5	2	1
4	3	6	1

X	Key	Val
1	A	1
1	B	4

X	Key	Val
1	A	1
1	B	4
1	C	3
2	A	2
...	...	...

[Wang PLDI17, Feng PLDI17, Feng PLDI18]

Step 2: Data Adapter Synthesis

$$\begin{array}{ccccc} & & T_{sketch} & & V_{partial} \\ & & \subseteq & & \subseteq \\ T & \xrightarrow{\phi_T} & T' & \xrightarrow{\phi_V} & V \end{array}$$

X	A	B	C
1	1	4	3
2	2	3	2
3	5	2	1
4	3	6	1

X	Key	Val
1	A	1
1	B	4

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

For any predicate, we have

$\phi_T(T) \subseteq$

X	A	B	C
1	1	4	2
2	2	3	2
3	5	2	1
4	3	6	1

Contradicts:

1	A	1
1	B	4

$\not\subseteq$

X	A	B	C
1	1	4	2
2	2	3	2
3	5	2	1
4	3	6	1

Start

filter(T, □) ✗

gather(T, key=□, val=□)

……

spread(T, id=□, key=□, val=□)

gather(T, key=C, val=□)

gather(T, key=X, val=[A])

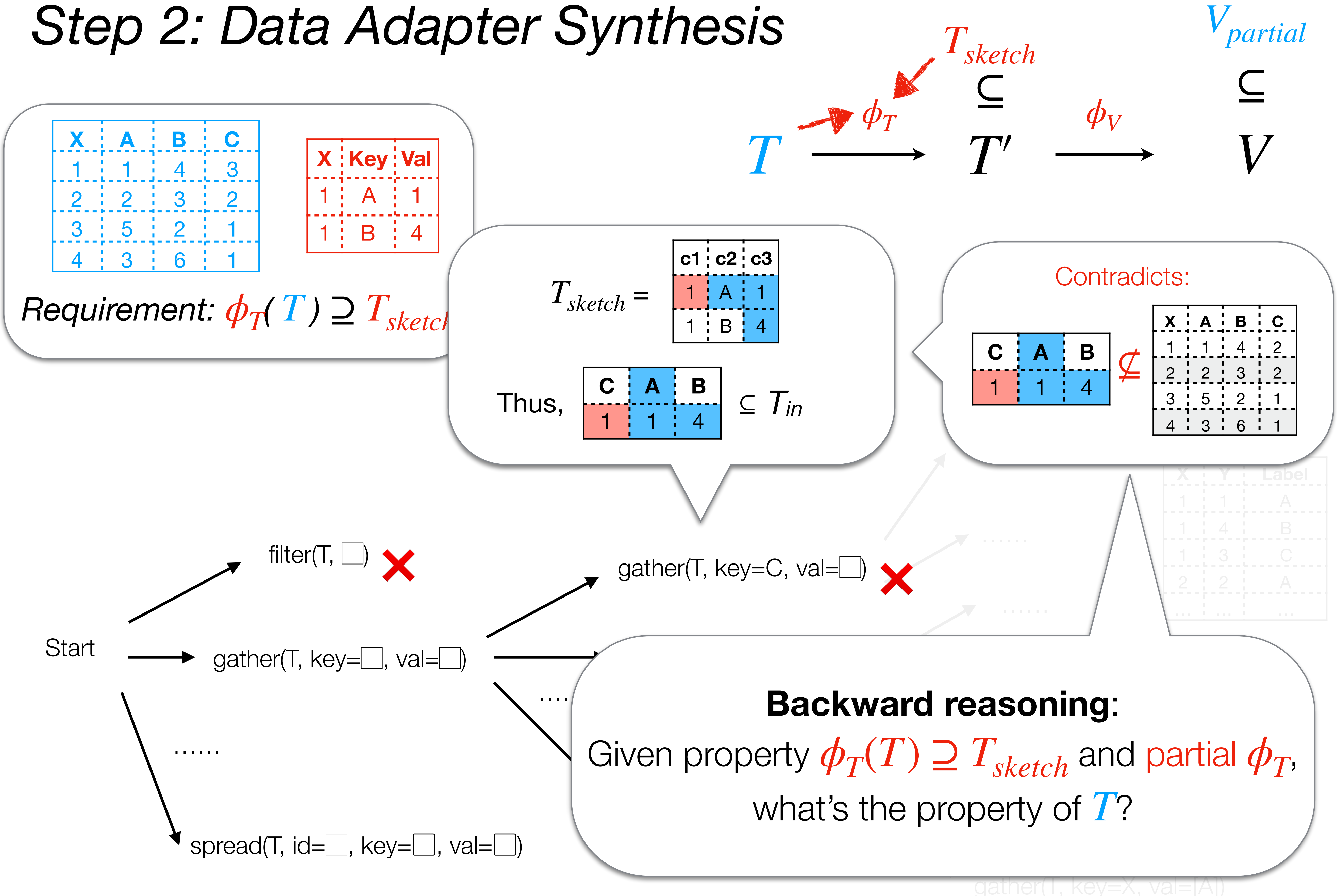
Forward reasoning

Given T and partial ϕ_T,

what's the property of the output $\phi_T(T)$?

[Wang PLDI17, Feng PLDI18]

Step 2: Data Adapter Synthesis



Step 2: Data Adapter Synthesis

$$T \xrightarrow{\phi_T} T' \xrightarrow{\phi_V} V$$

$$T_{sketch} \subseteq T' \qquad V_{partial} \subseteq V$$

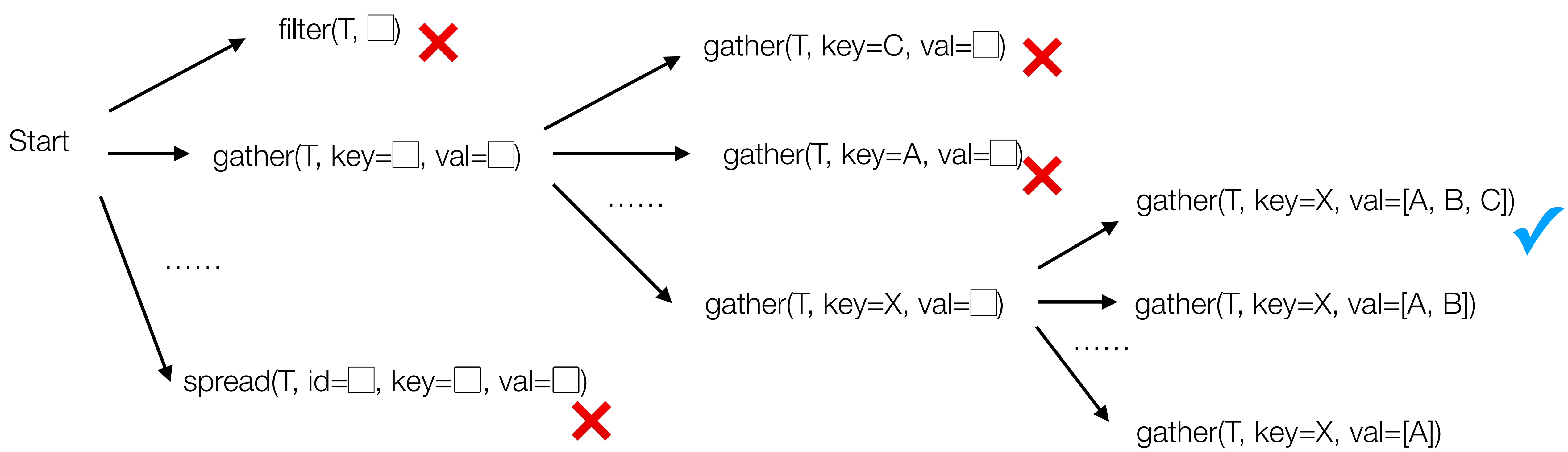
X	A	B	C
1	1	4	3
2	2	3	2
3	5	2	1
4	3	6	1

X	Key	Val
1	A	1
1	B	4

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

Contribution: **Bidirectional reasoning**

Inductively defined for all operators in $\mathscr{L}_T$



Visualization by Example

$$\phi_V(\phi_T(\,T\,)) \supseteq V_{partial}$$

Potentially multiple (ϕ_T, ϕ_V) pairs can satisfy the specification.

T

X	A	B	C
1	1	4	3
2	2	3	2
3	5	2	1
4	3	6	1

ϕ_T

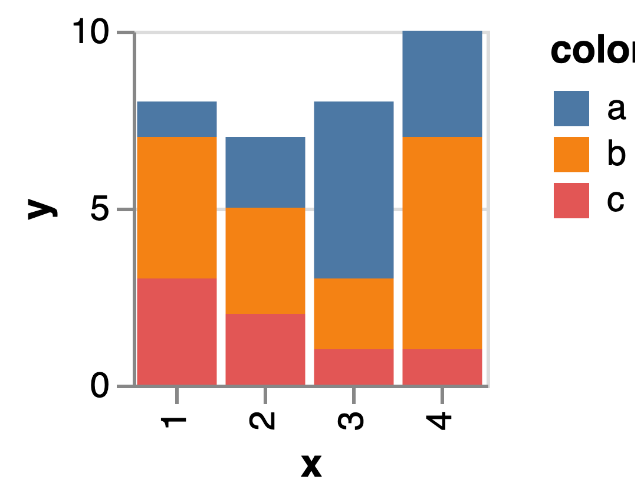
gather(
T,
id=X,
key=[A,B,C])

X	Key	Val
1	A	1
1	B	4
1	C	3
2	A	2
…	…	…

ϕ_V

X → bar.x
Val → bar.height
Key → bar.color

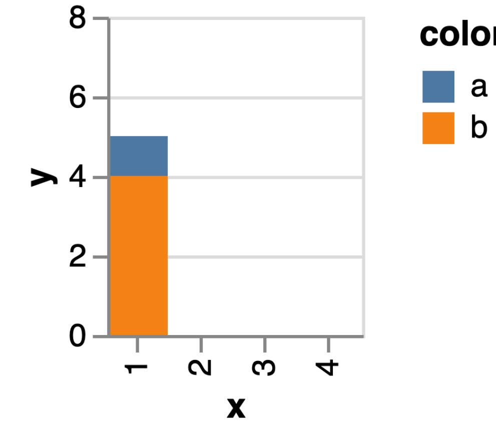
bar(x=1, h=1, color=A)
bar(x=1, h=4, color=B)
bar(x=1, h=3, color=C)
bar(x=2, h=2, color=A)
……



$\supseteq$

$V_{partial}$

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



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 t orials
- **Evalu**
  - 60
  - Pa
    fro

<u>*Input table size*</u>

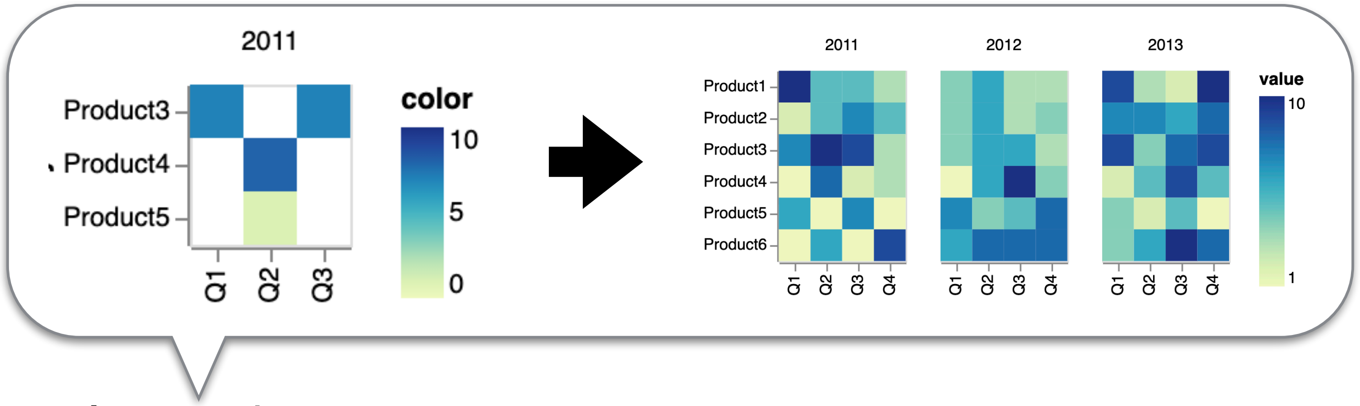
Ranges from **4x3** to **3686x9**, average size **100x10**

<u>*Program Complexity*</u>

1-4 statements,

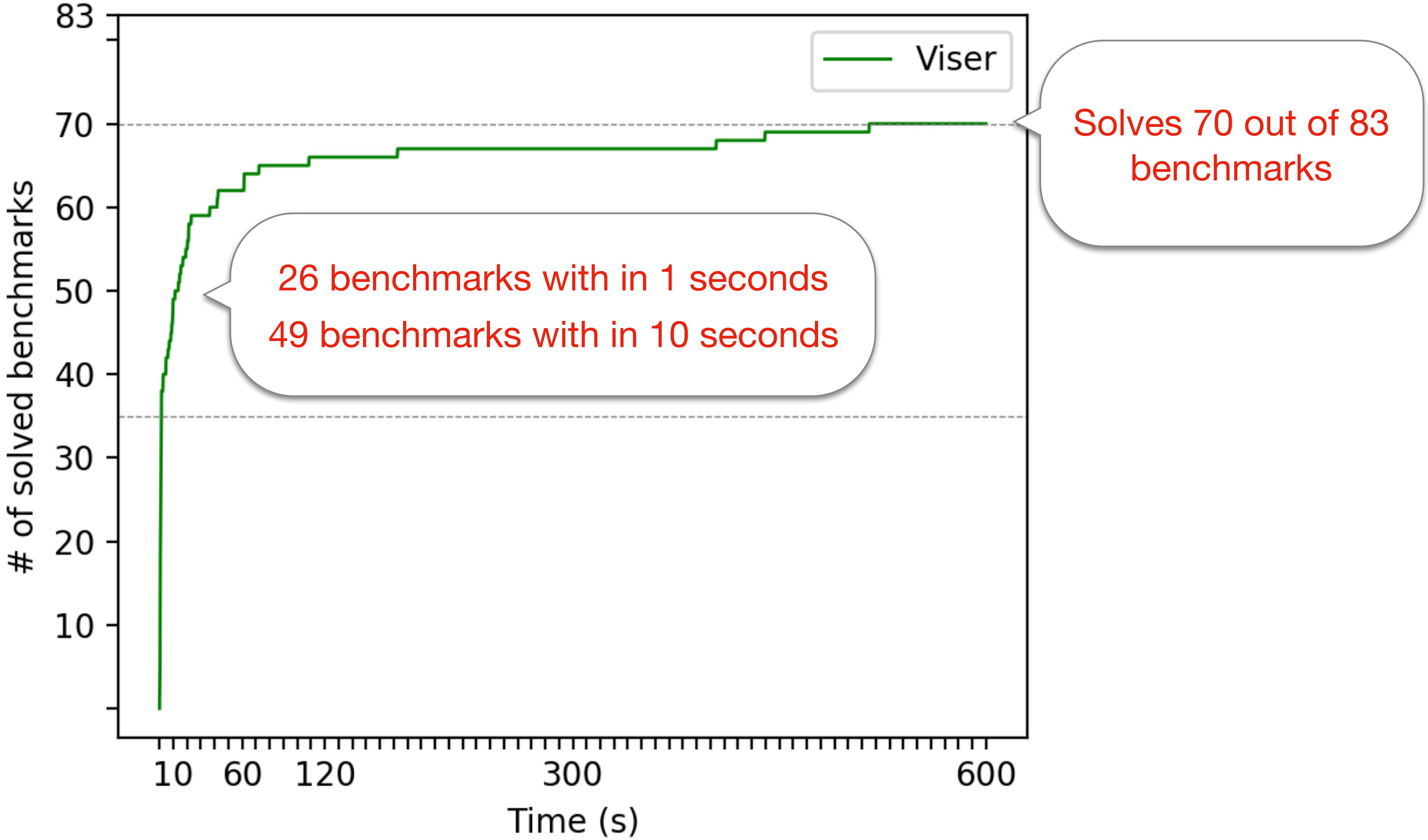
On average 20 decisions to make for each program

Experiment

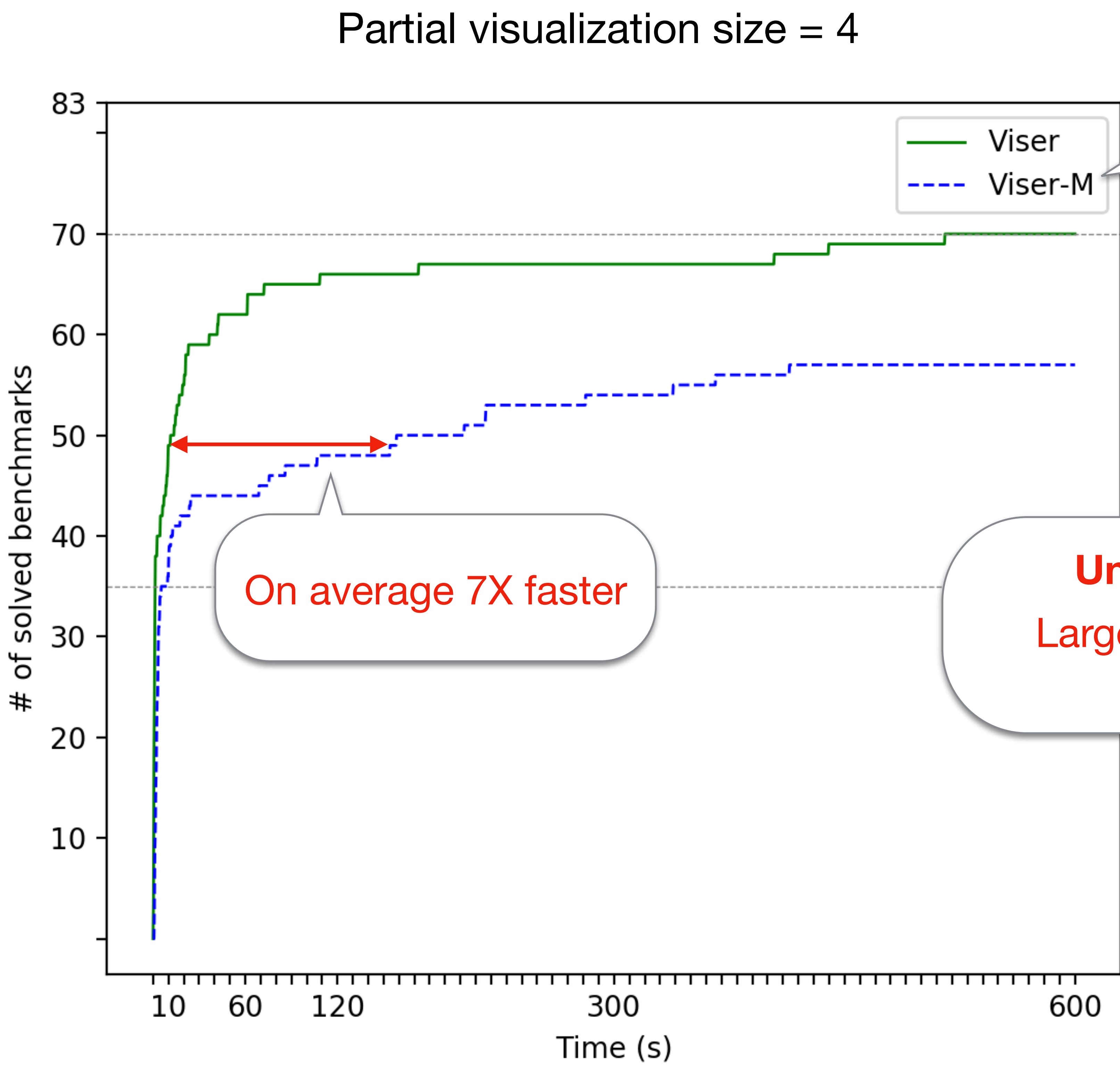


Partial visualization size = 4

(i.e., 4 random geometric objects from the full visualization)



Performance Experiment



Baseline Viser-M

a variation of Viser without bidirectional pruning

[Feng PLDI18]

Solves 17 more benchmarks

On average 7X faster

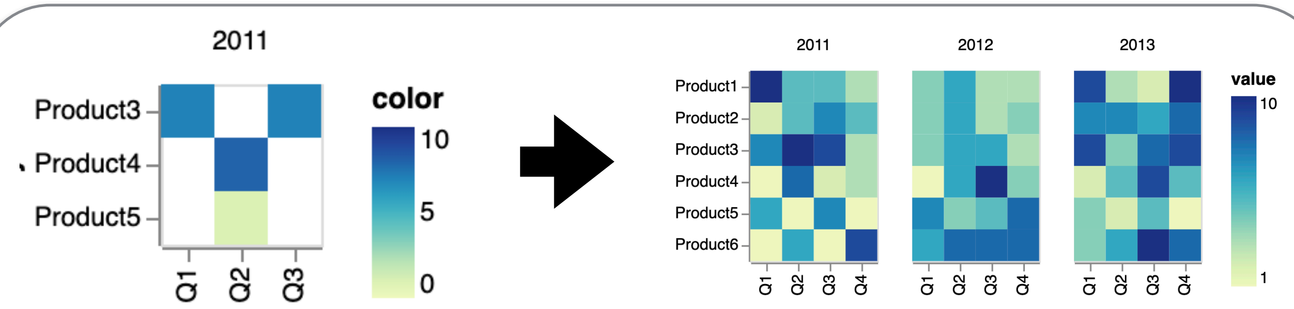
Unsolved benchmarks

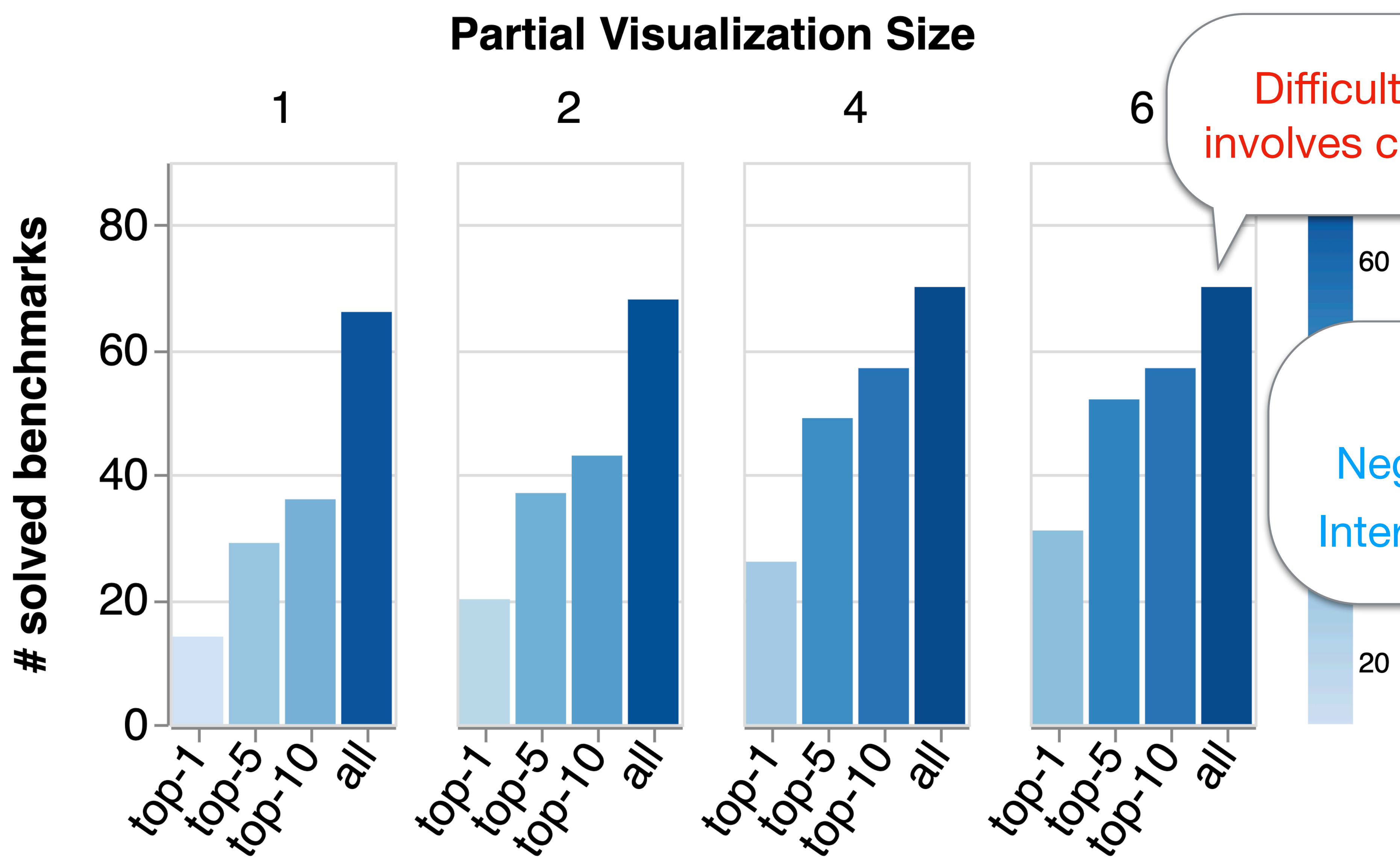
Large input table & complex transformations

Look ahead

Multi-modal synthesis

Usability Experiment





Difficult tasks often involves custom filtering

Look ahead

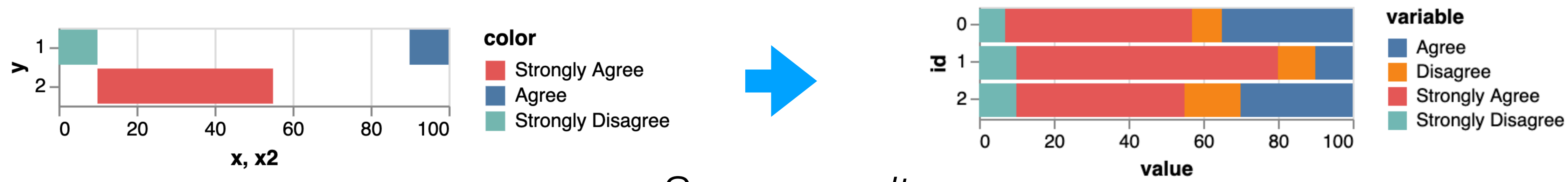
Negative examples,

Interactive refinement

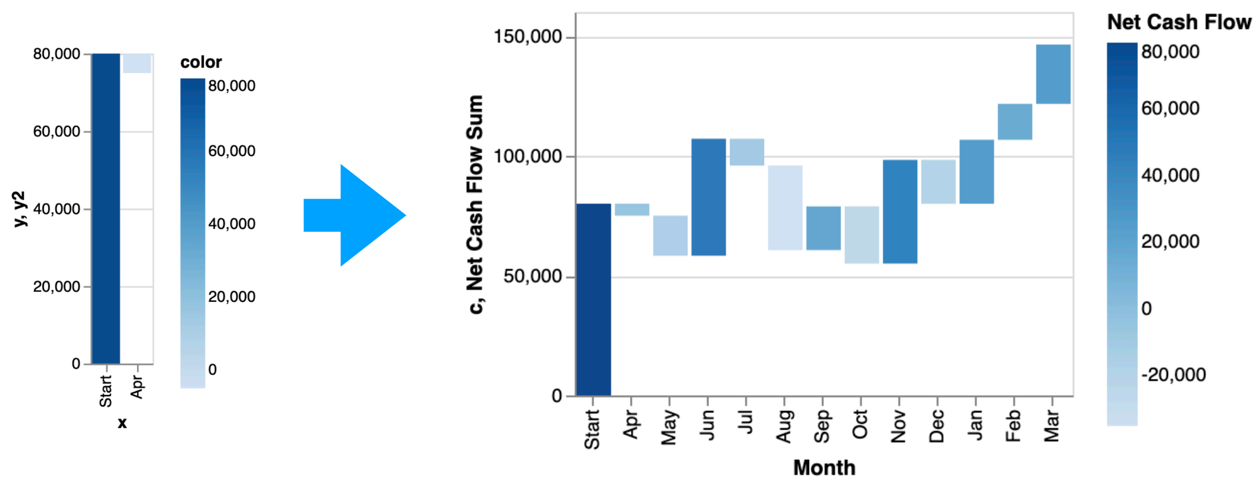
Can solve a lot benchmarks with a small number of examples.

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

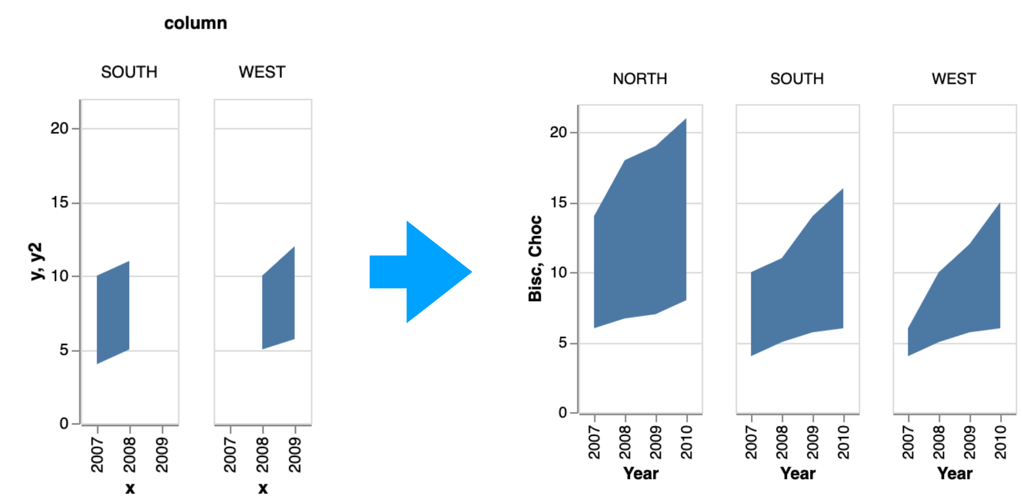
Visualizations



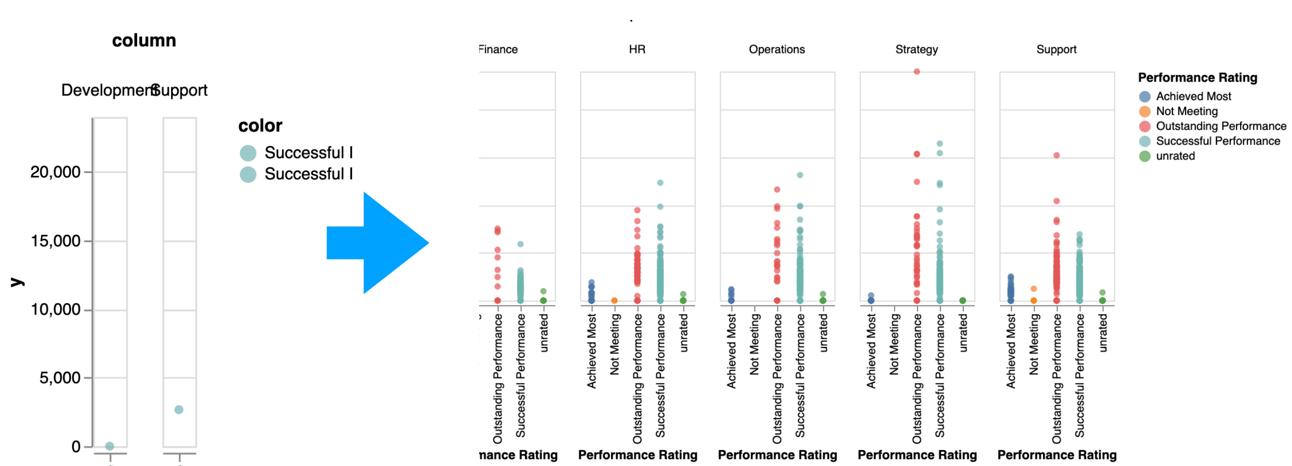
Survey result



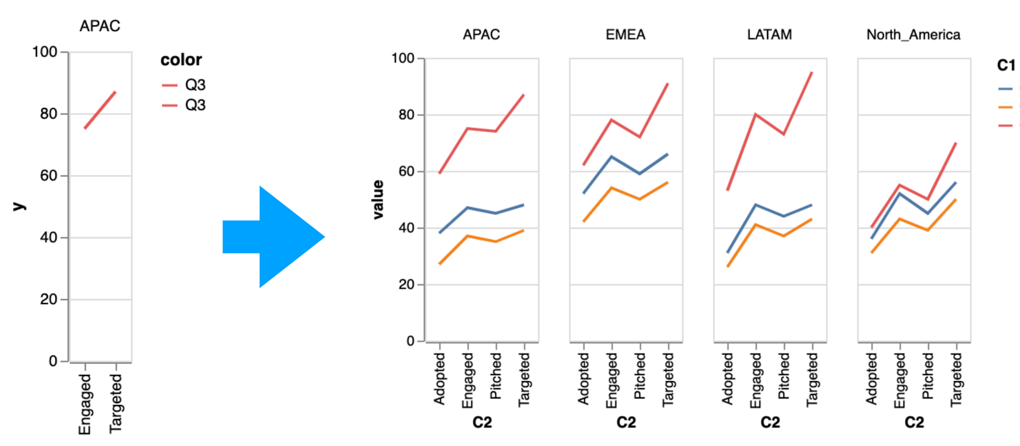
Net cash flow in a year



Housing price in different region



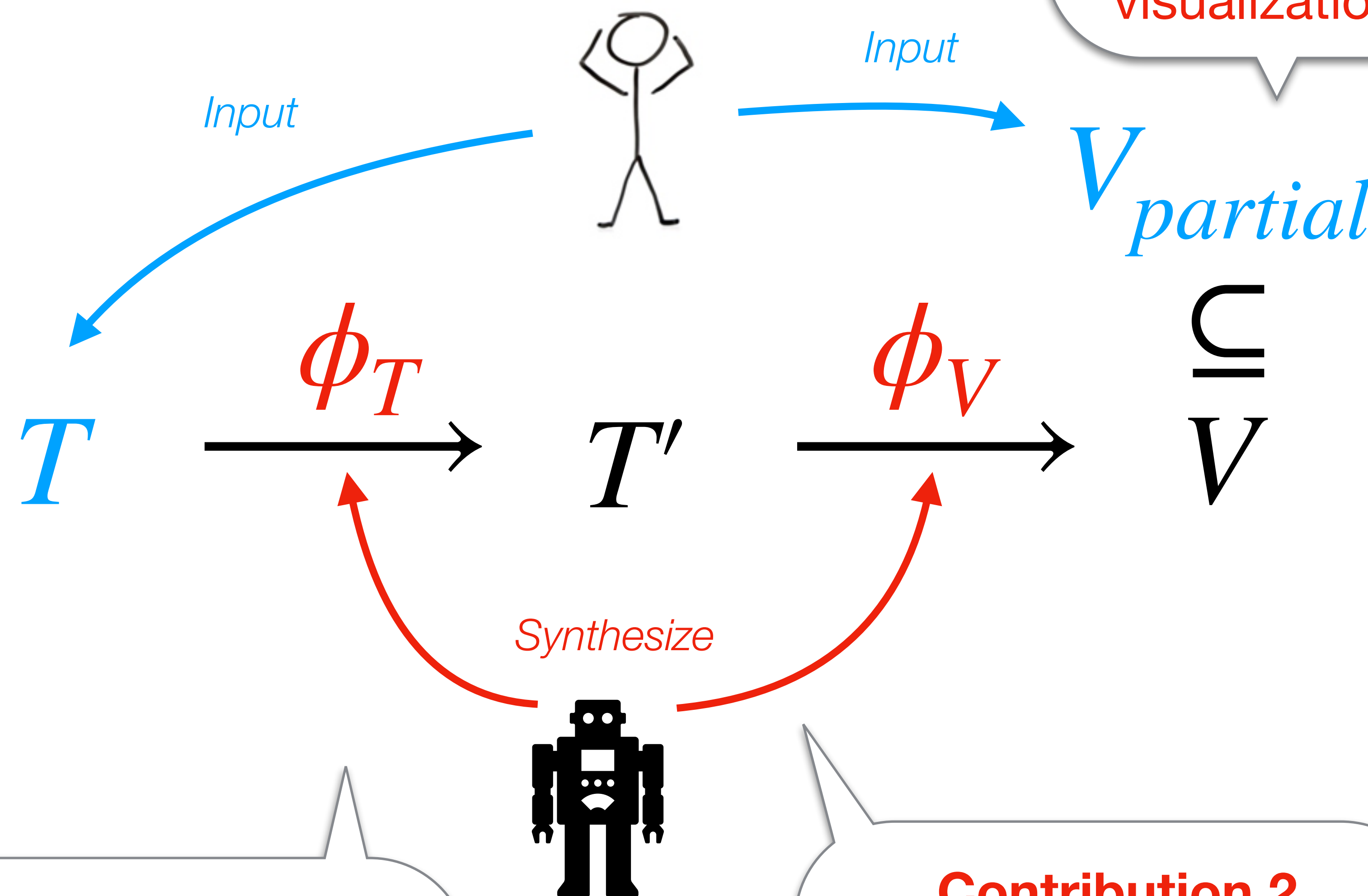
Housing price in different region



Product price in different region

Visualization by Example

Contribution 1
Define and formalize the visualization synthesis problem



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

Contribution 3
Bidirectional analysis for pruning partial programs

Contribution 2
Compositional synthesis of visualizations

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