ACM TIST, 2022
Supplemental material

Steering-by-Example for Progressive Visual Analytics

Marco Angelini,[a] Marius Hogräfer,[b] Giuseppe Santucci,[a] Hans-Jörg Schulz[b]

a SAPIENZA UNIVERSITÀ DI ROMA
b AARHUS UNIVERSITY

Overview

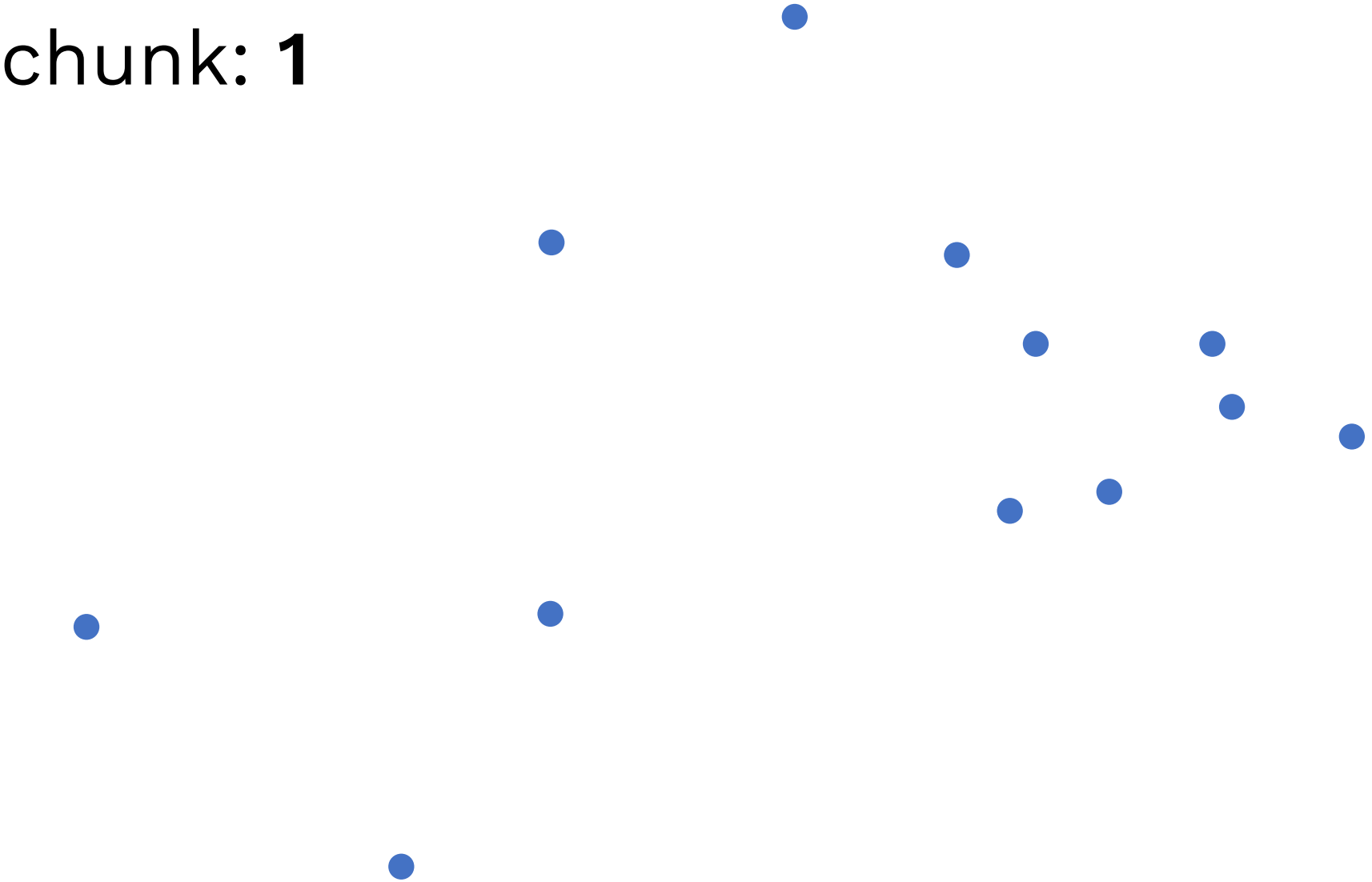
1) Introducing Steering-by-Example

2) ProSteer's Interface

3) Demo

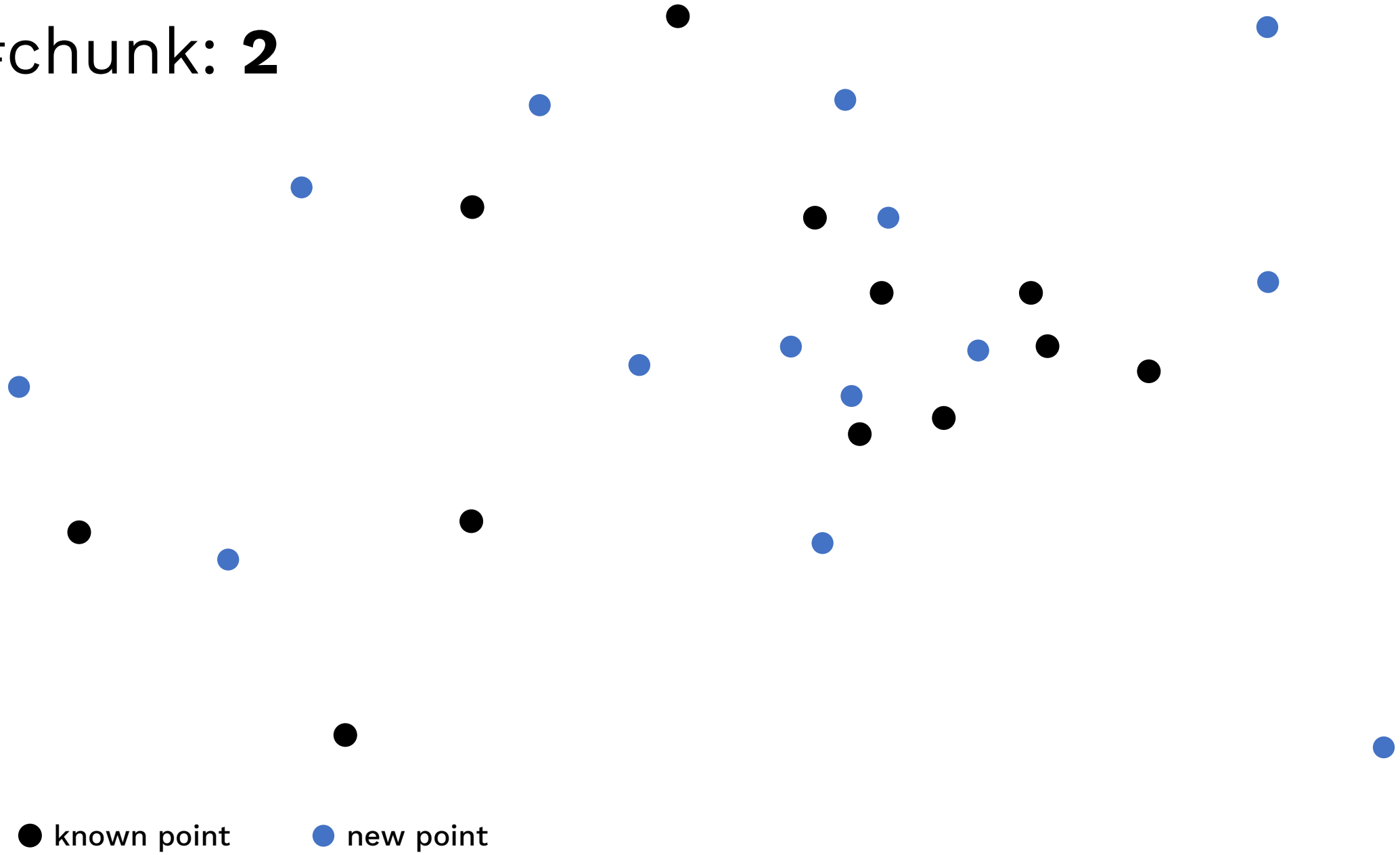
1.1 Motivation

#chunk: 1

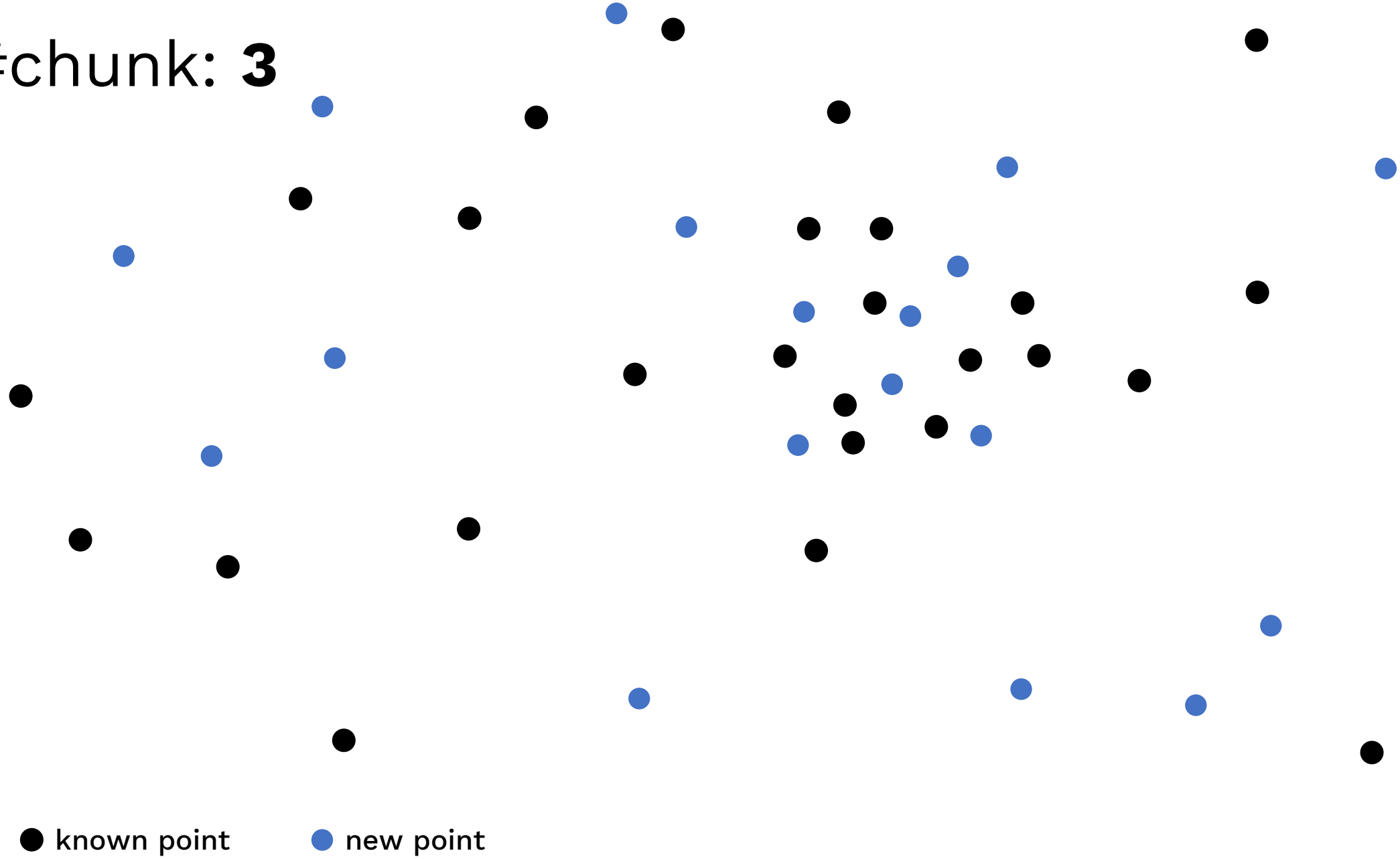


new point

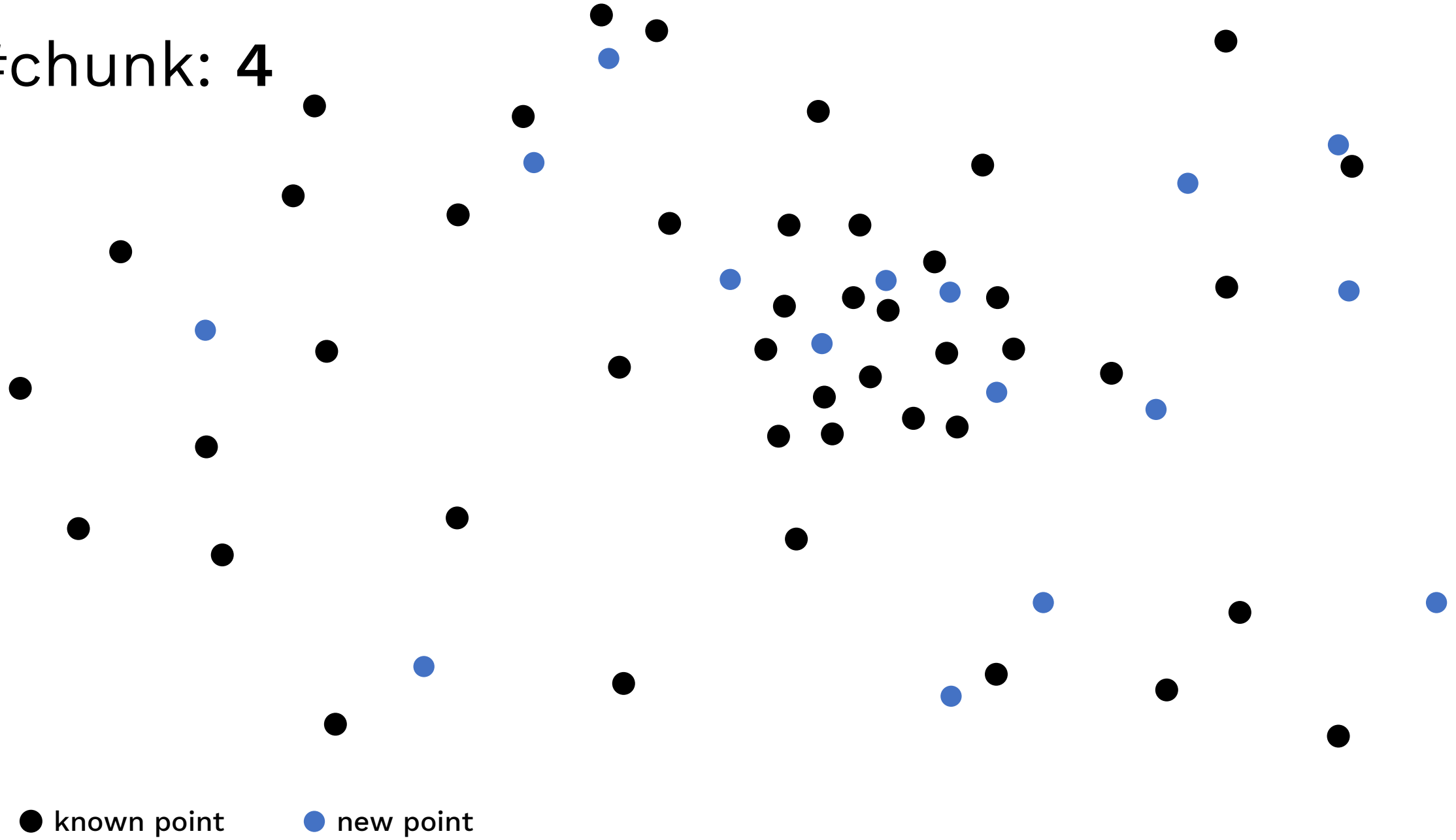
#chunk: **2**



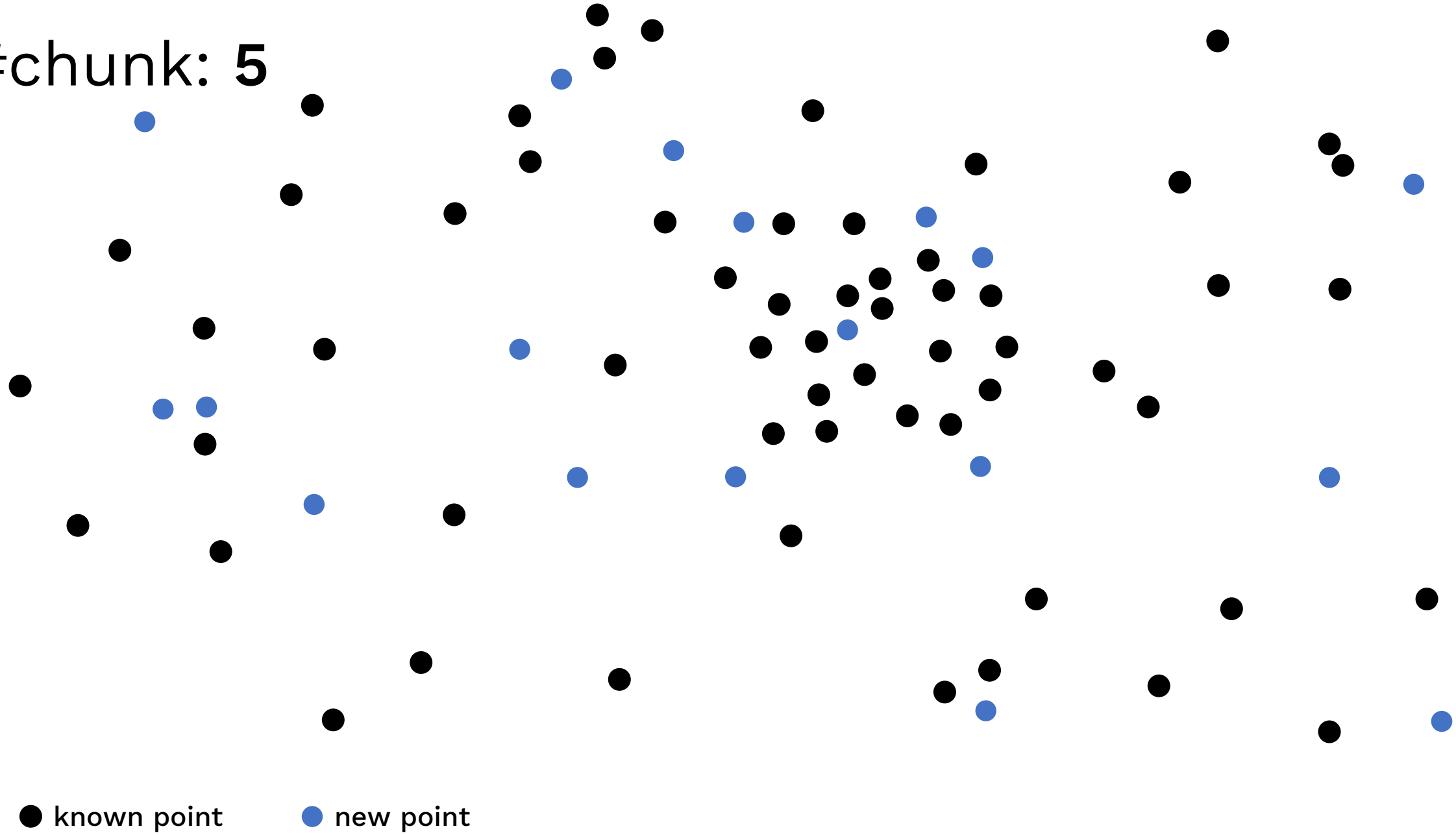
#chunk: **3**



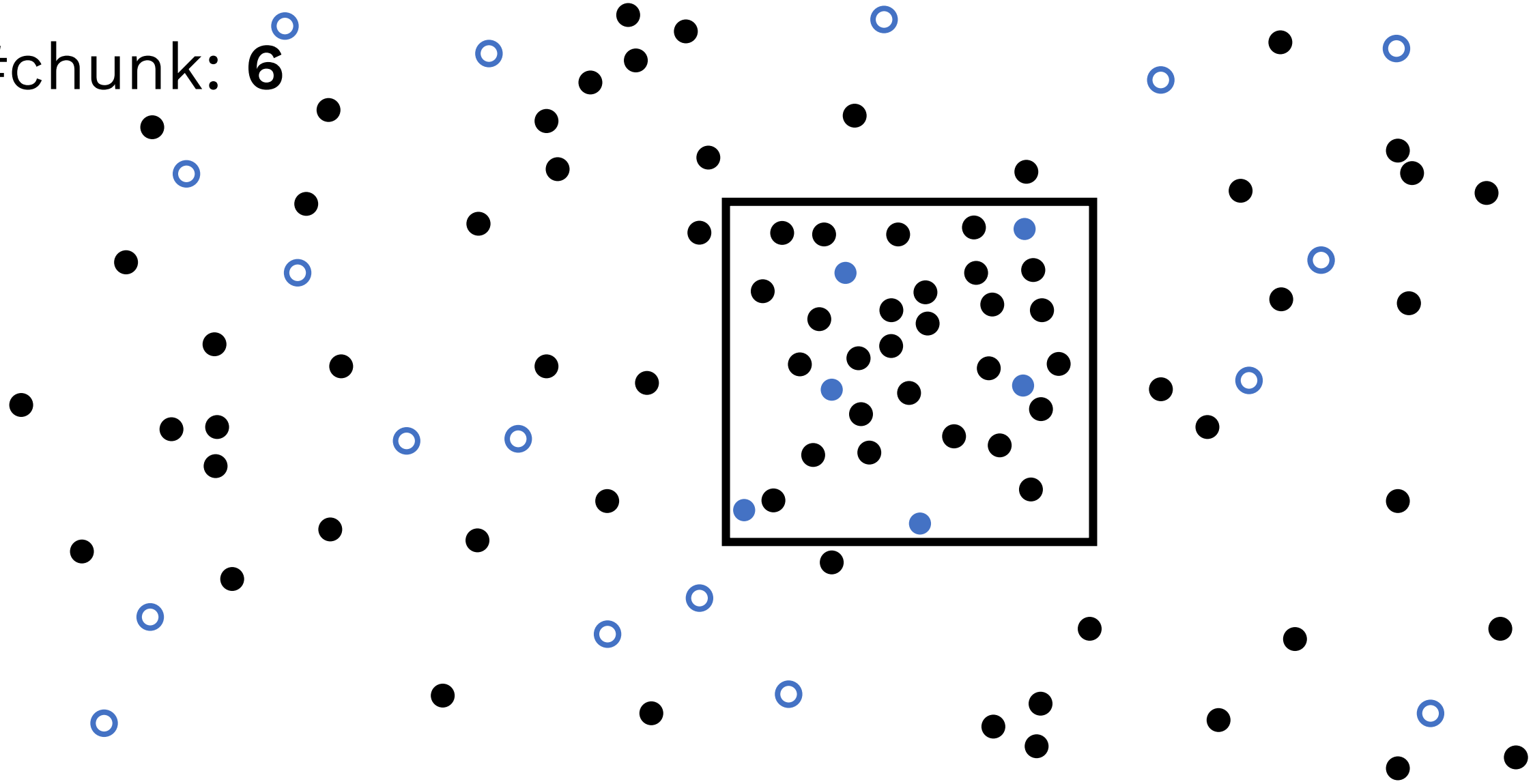
#chunk: **4**



#chunk: 5

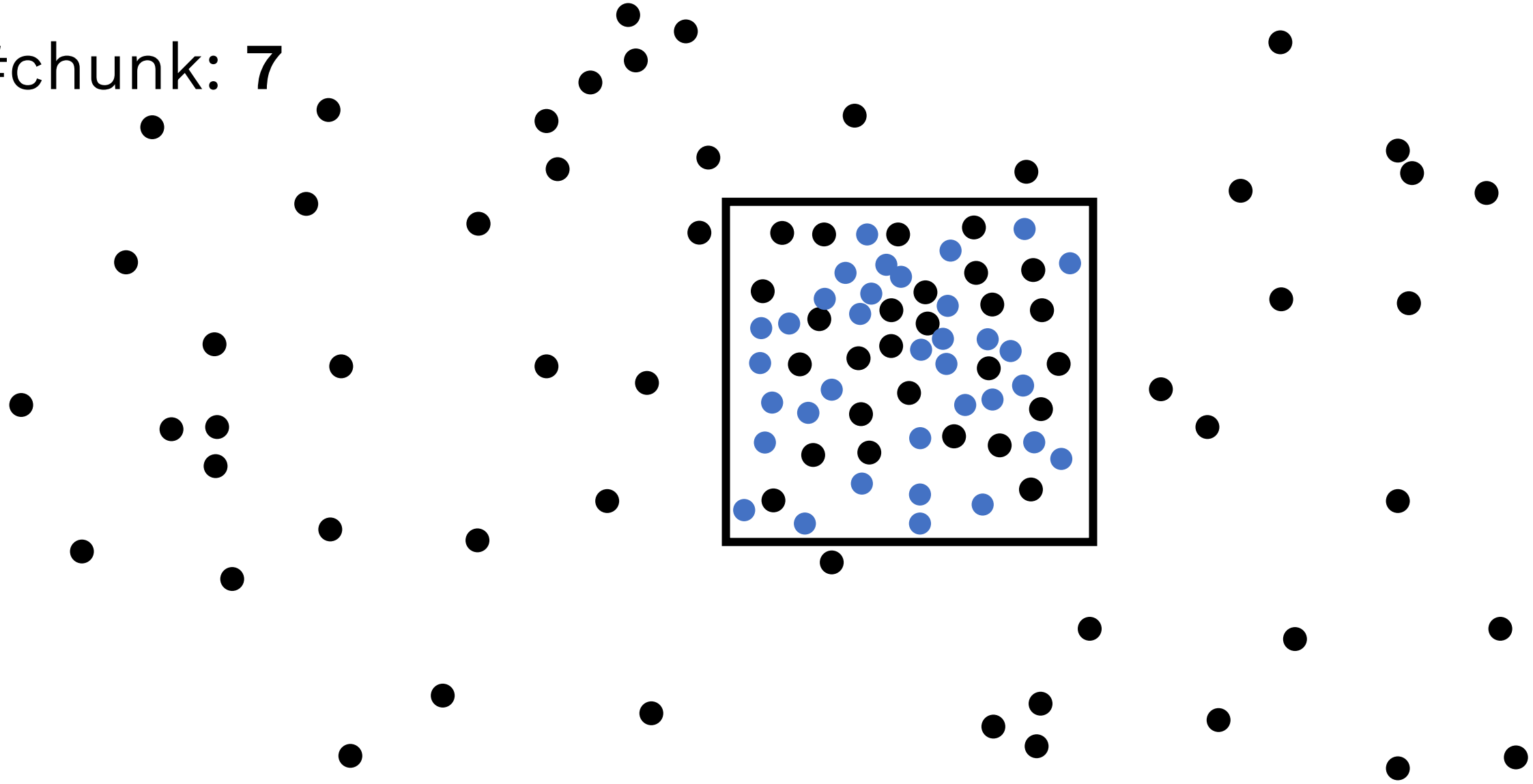


#chunk: 6

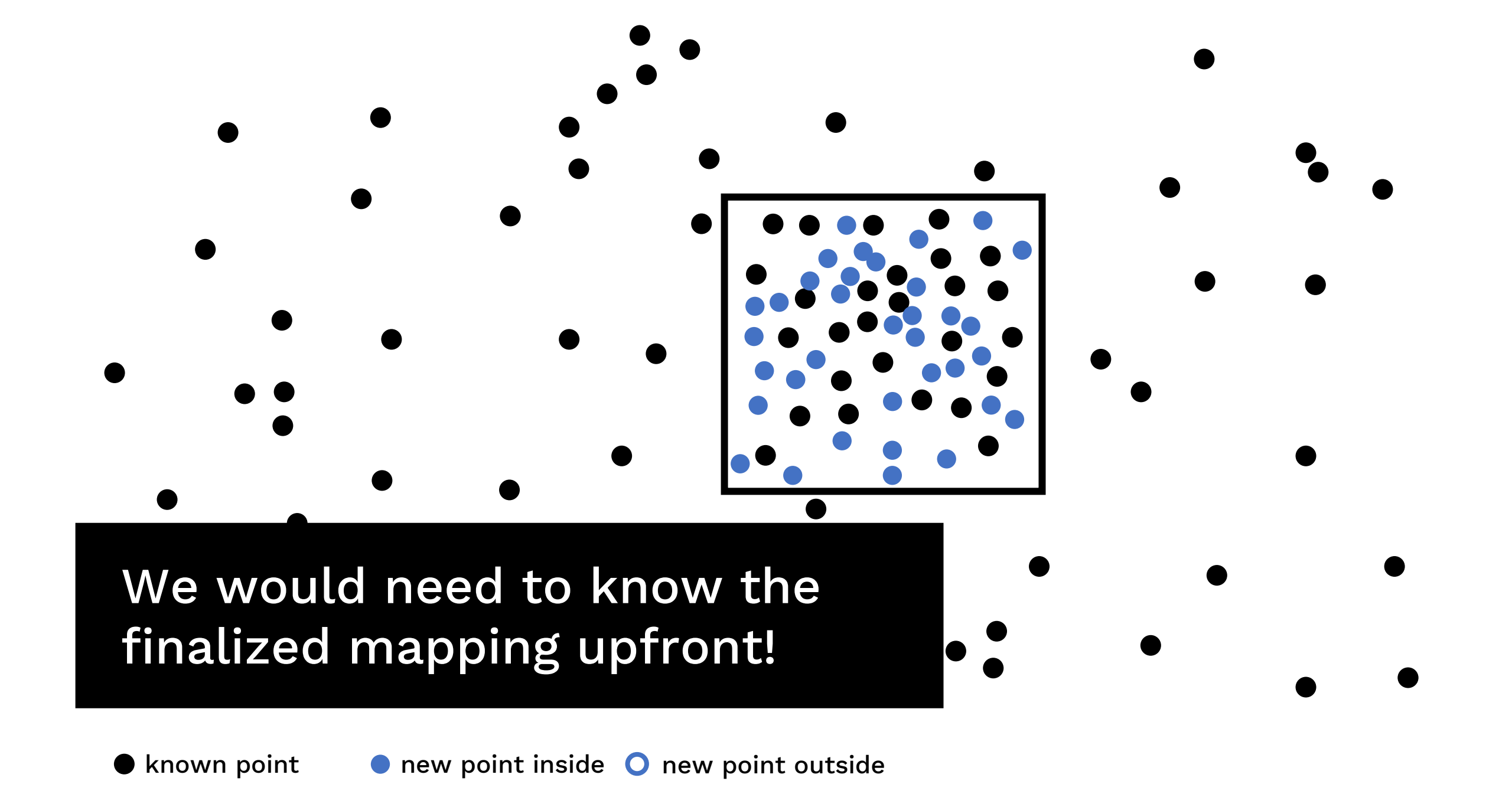


known point new point inside new point outside

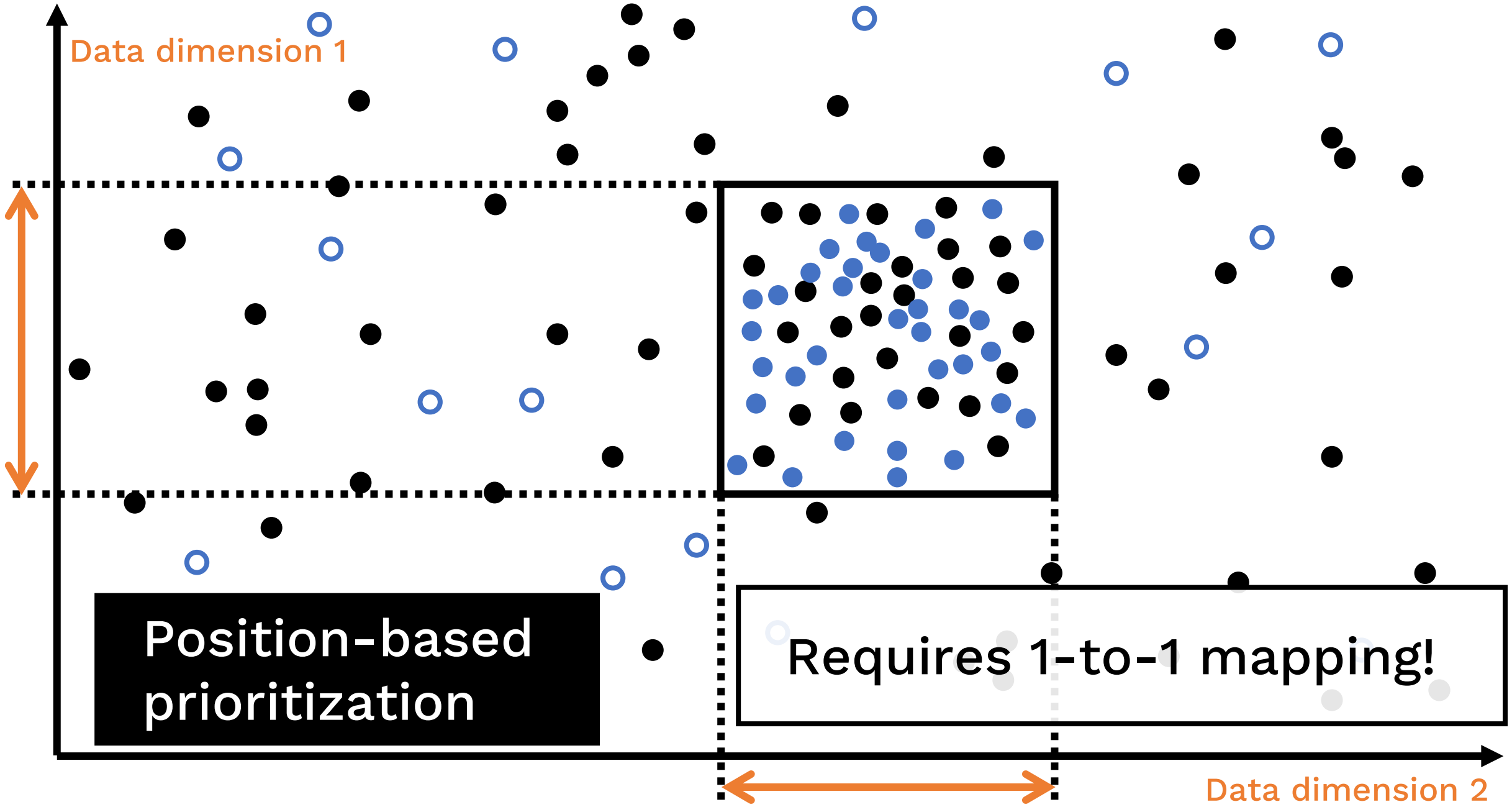
#chunk: 7



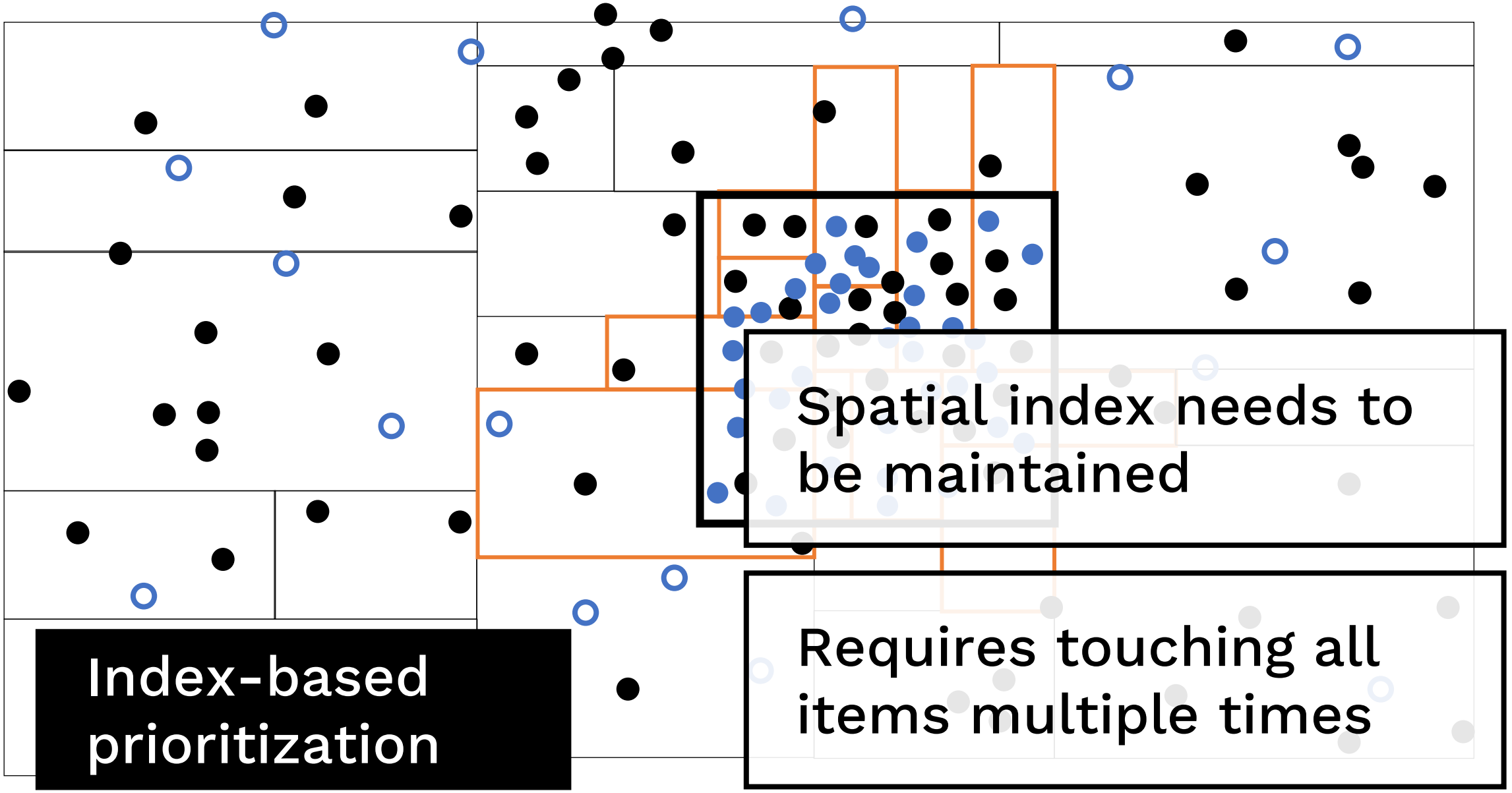
known point new point inside new point outside


We would need to know the finalized mapping upfront!
known point
new point inside
new point outside

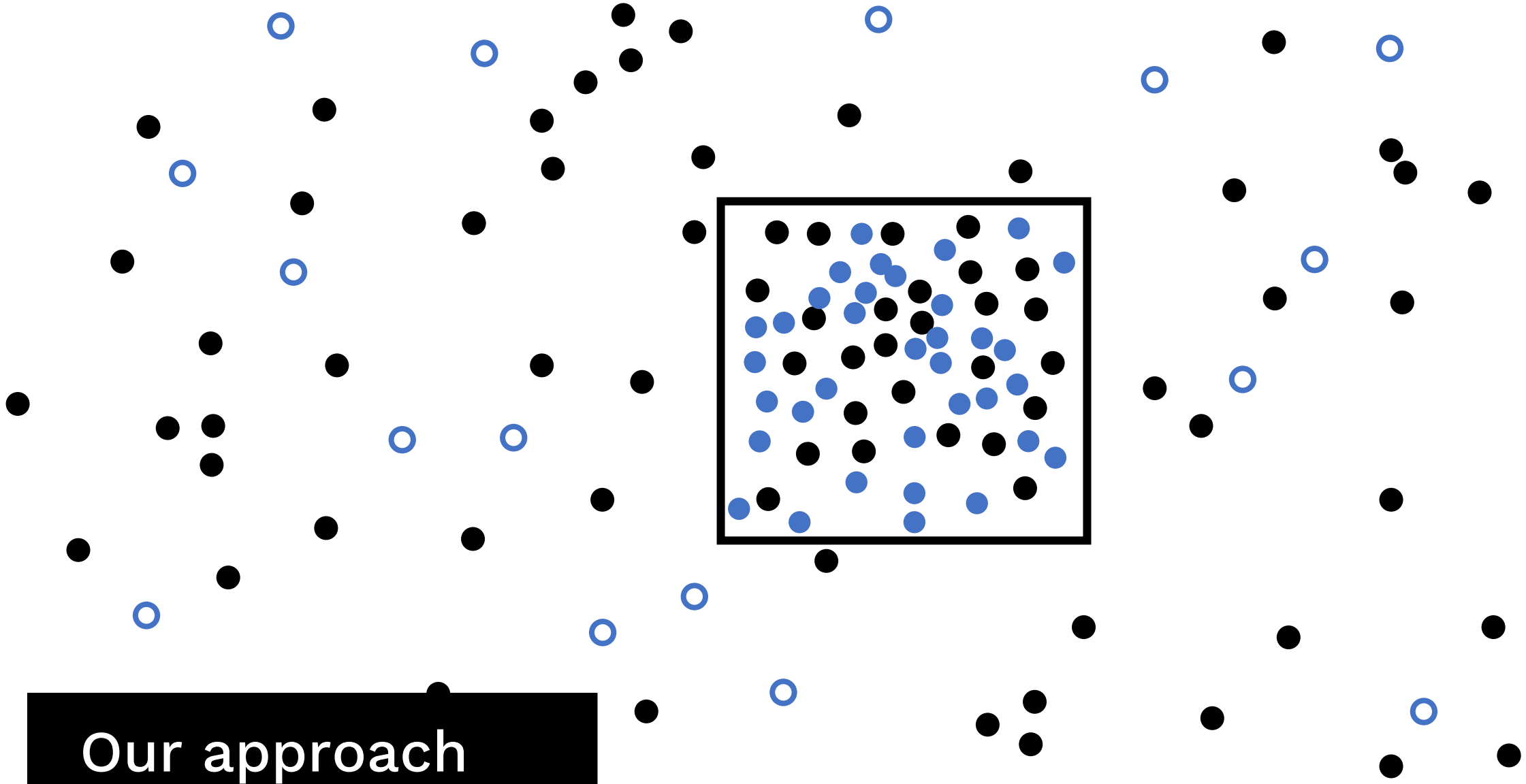
1.2 Related Work



[Cui et al., 2019]



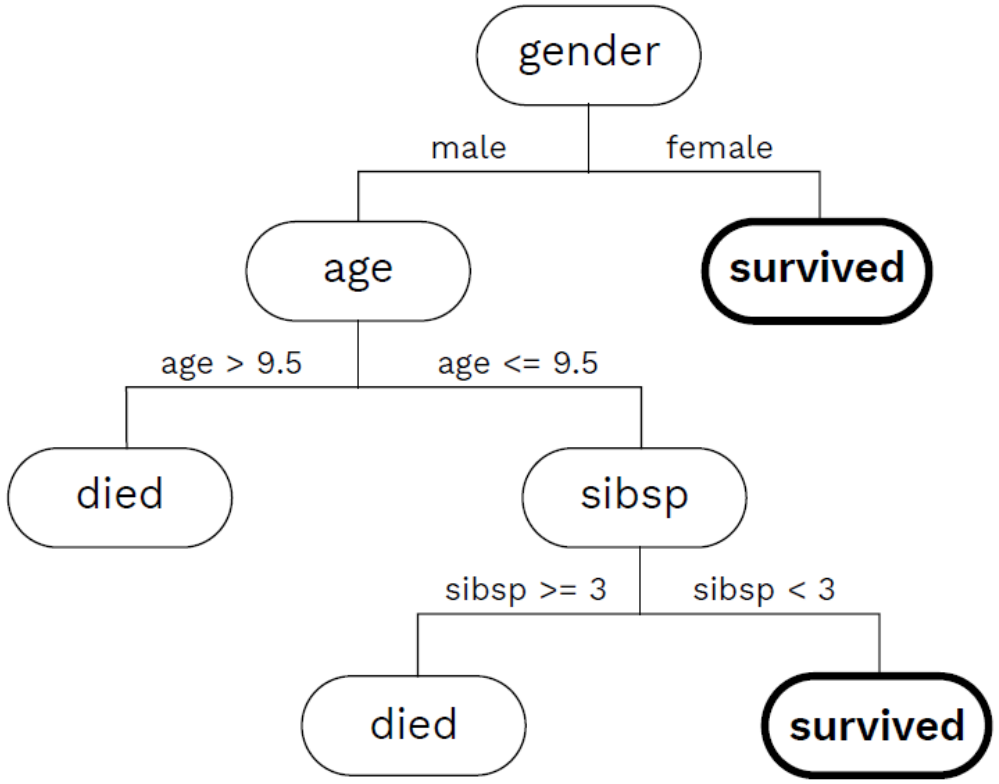
[Williams and Munzner, 2008]



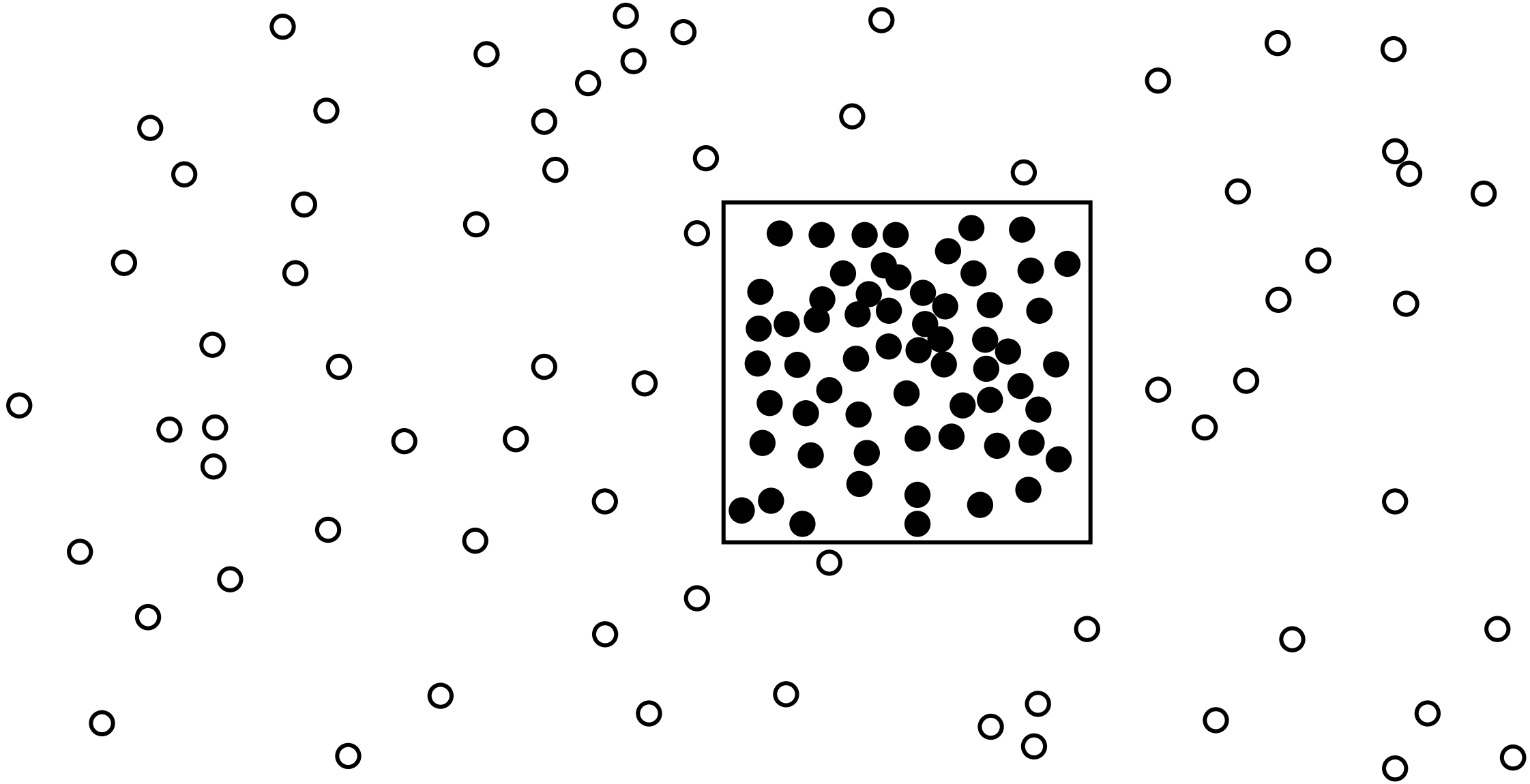
Our approach

1.3 Our Approach

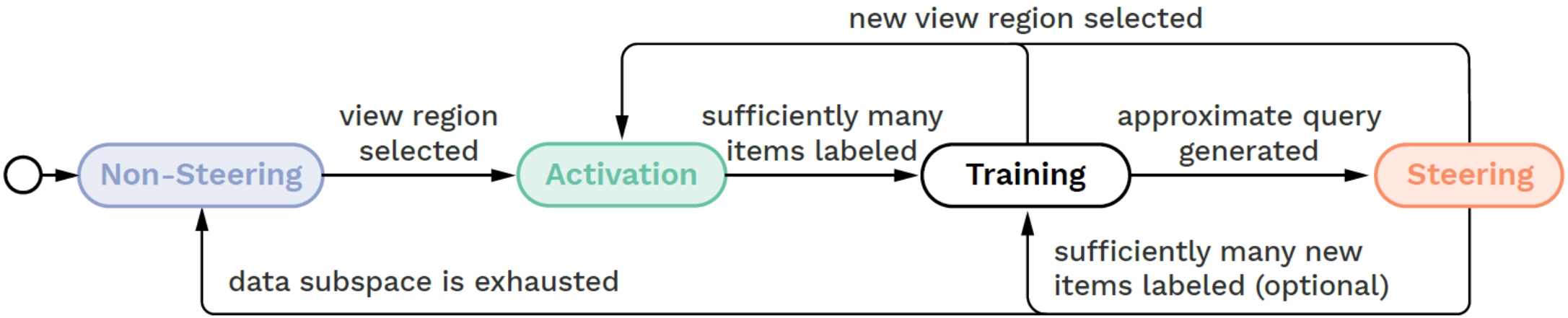
Our Approach



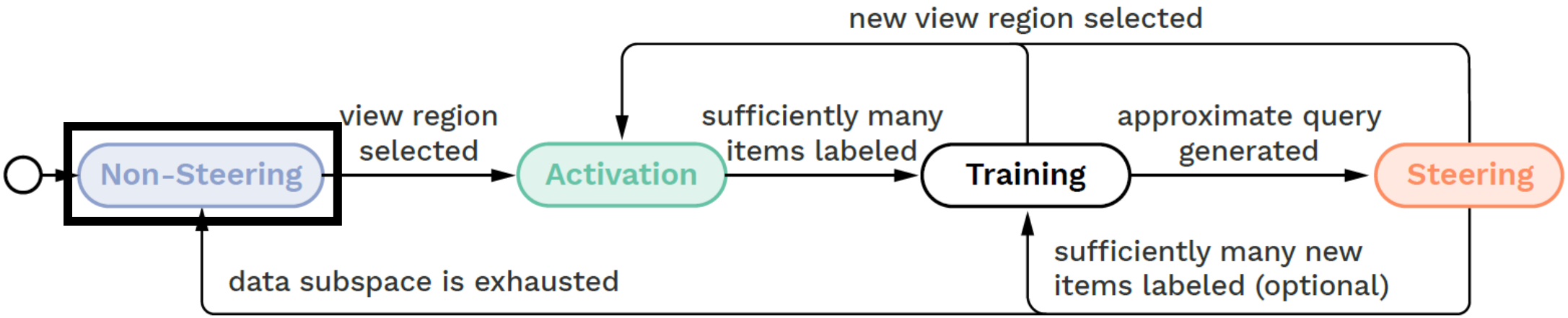

```
SELECT * FROM table

WHERE gender="female"
  OR (gender="male"
        AND age <= 9.5
        AND sibsp < 3);
```


Point inside
point outside

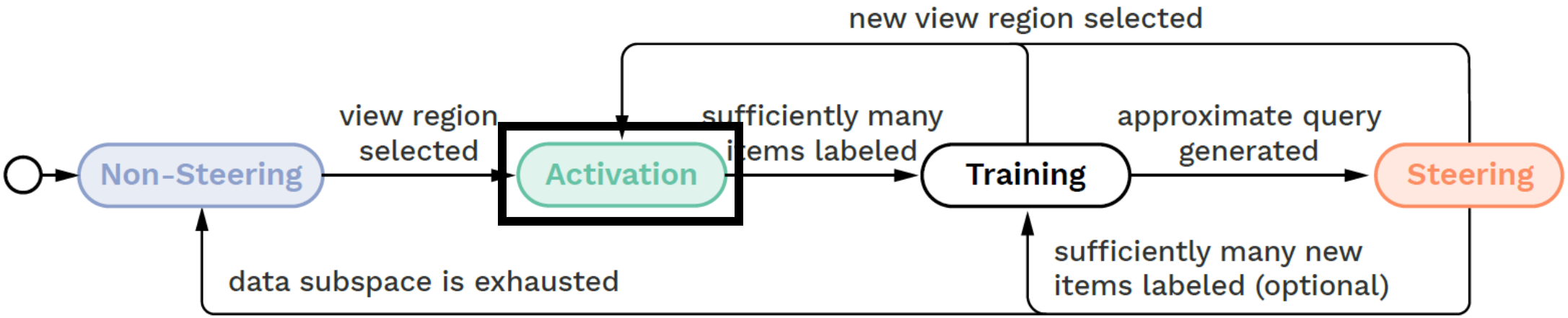
Phases


new view region selected
view region selected
sufficiently many items labeled
approximate query generated
Non-Steering
Activation
Training
Steering
data subspace is exhausted
sufficiently many new items labeled (optional)

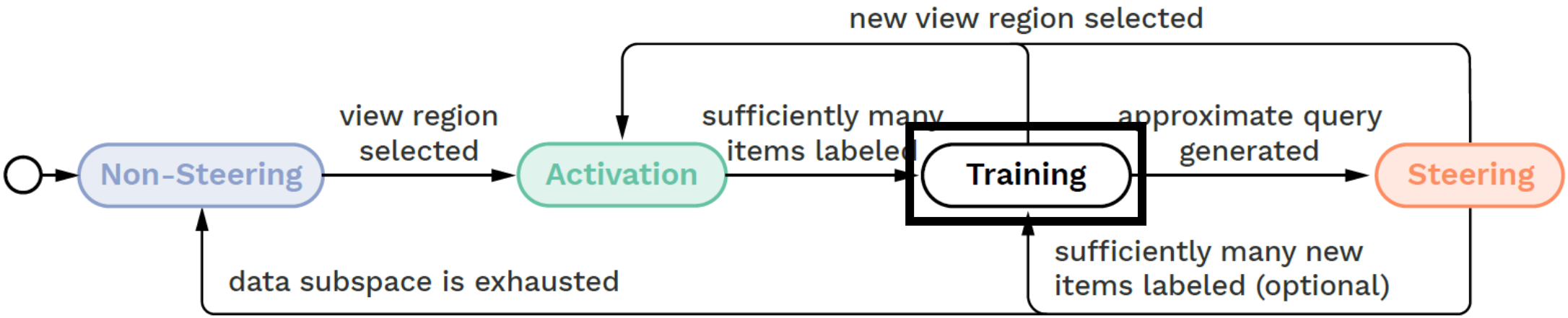
Phases


new view region selected
view region selected
sufficiently many items labeled
approximate query generated
Non-Steering
Activation
Training
Steering
data subspace is exhausted
sufficiently many new items labeled (optional)

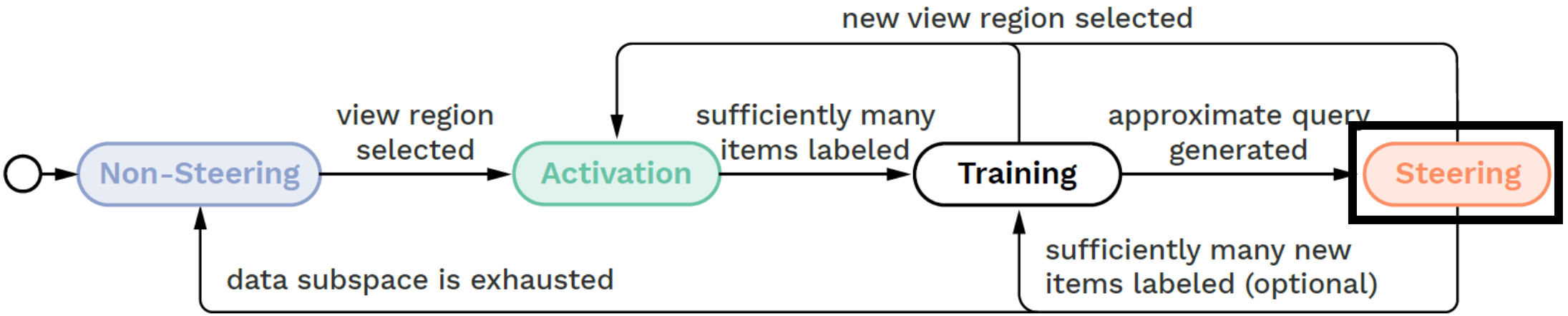
Phases


new view region selected
view region selected
sufficiently many items labeled
approximate query generated
Non-Steering
Activation
Training
Steering
data subspace is exhausted
sufficiently many new items labeled (optional)

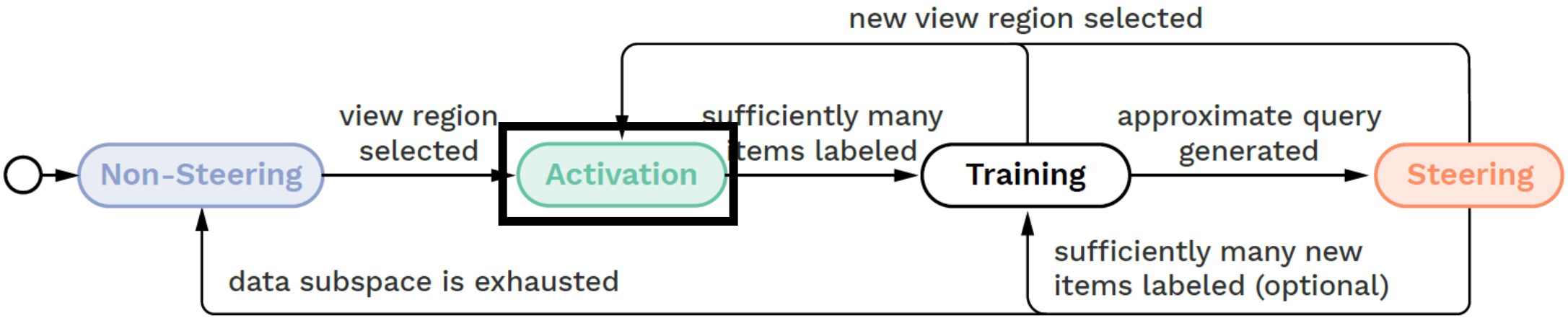
Phases


new view region selected
view region selected
sufficiently many items labeled
approximate query generated
Non-Steering
Activation
Training
Steering
data subspace is exhausted
sufficiently many new items labeled (optional)

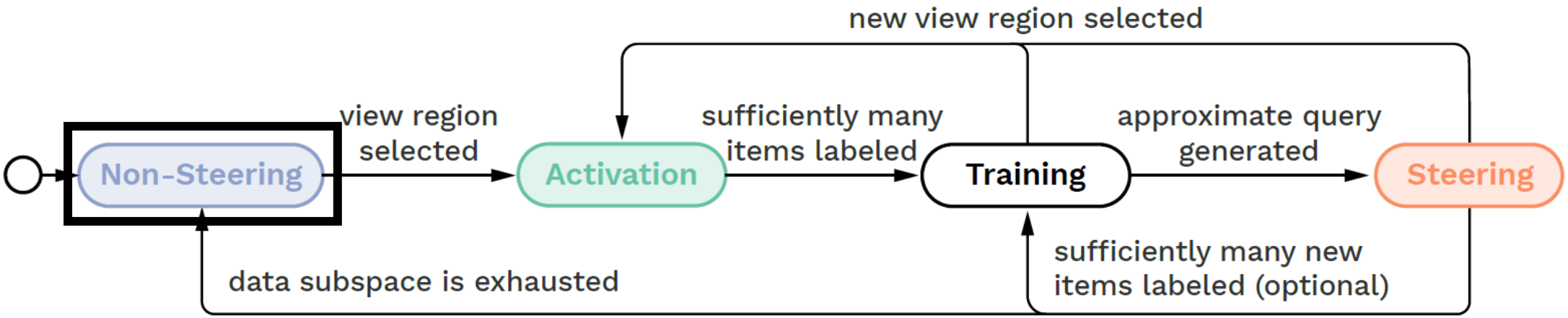
Phases


new view region selected
view region selected
sufficiently many items labeled
approximate query generated
Non-Steering
Activation
Training
Steering
data subspace is exhausted
sufficiently many new items labeled (optional)

Phases


new view region selected
view region selected
sufficiently many items labeled
approximate query generated
Non-Steering
Activation
Training
Steering
data subspace is exhausted
sufficiently many new items labeled (optional)

Phases

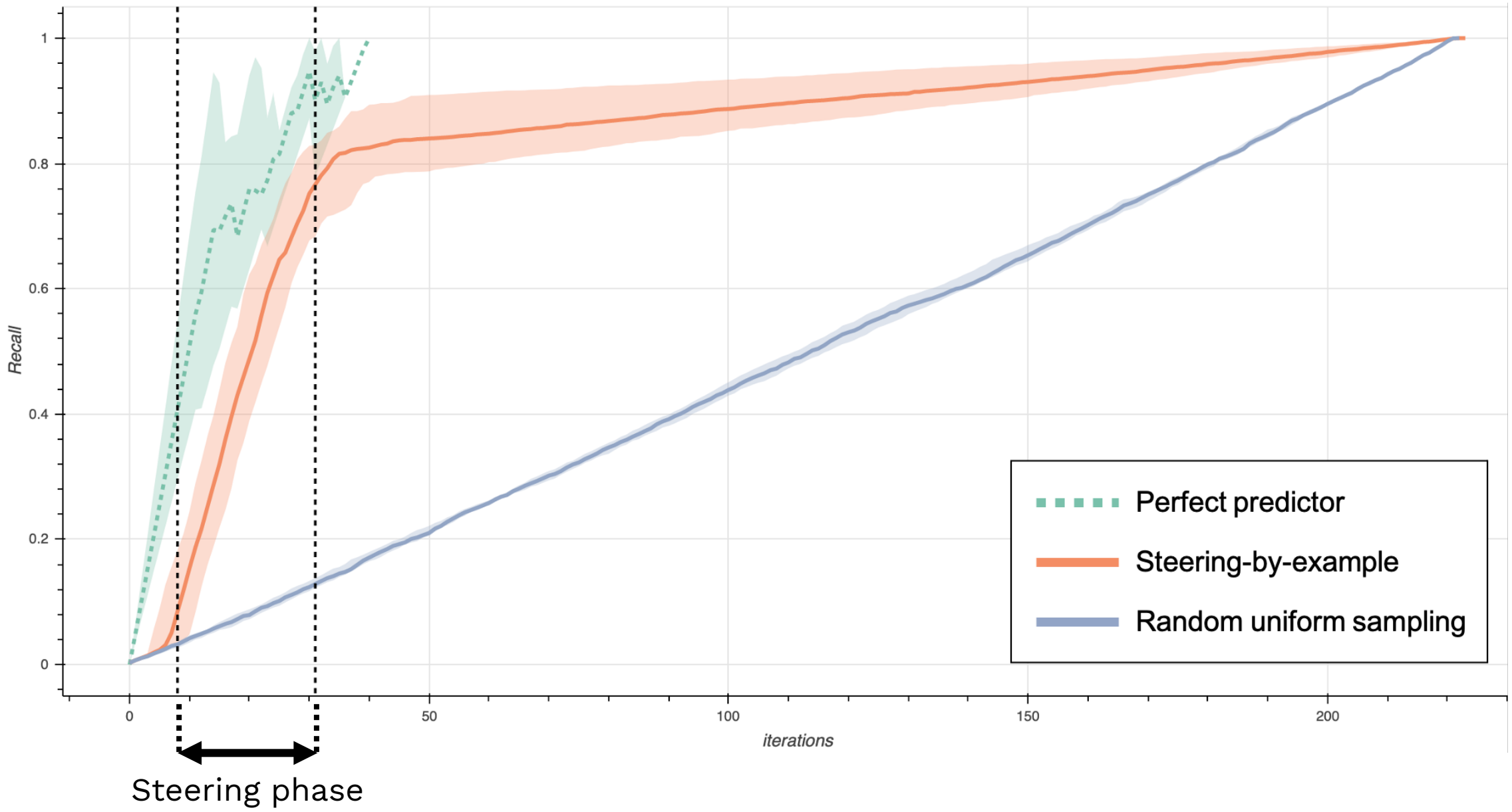

new view region selected
view region selected
sufficiently many items labeled
approximate query generated
Non-Steering
Activation
Training
Steering
data subspace is exhausted
sufficiently many new items labeled (optional)

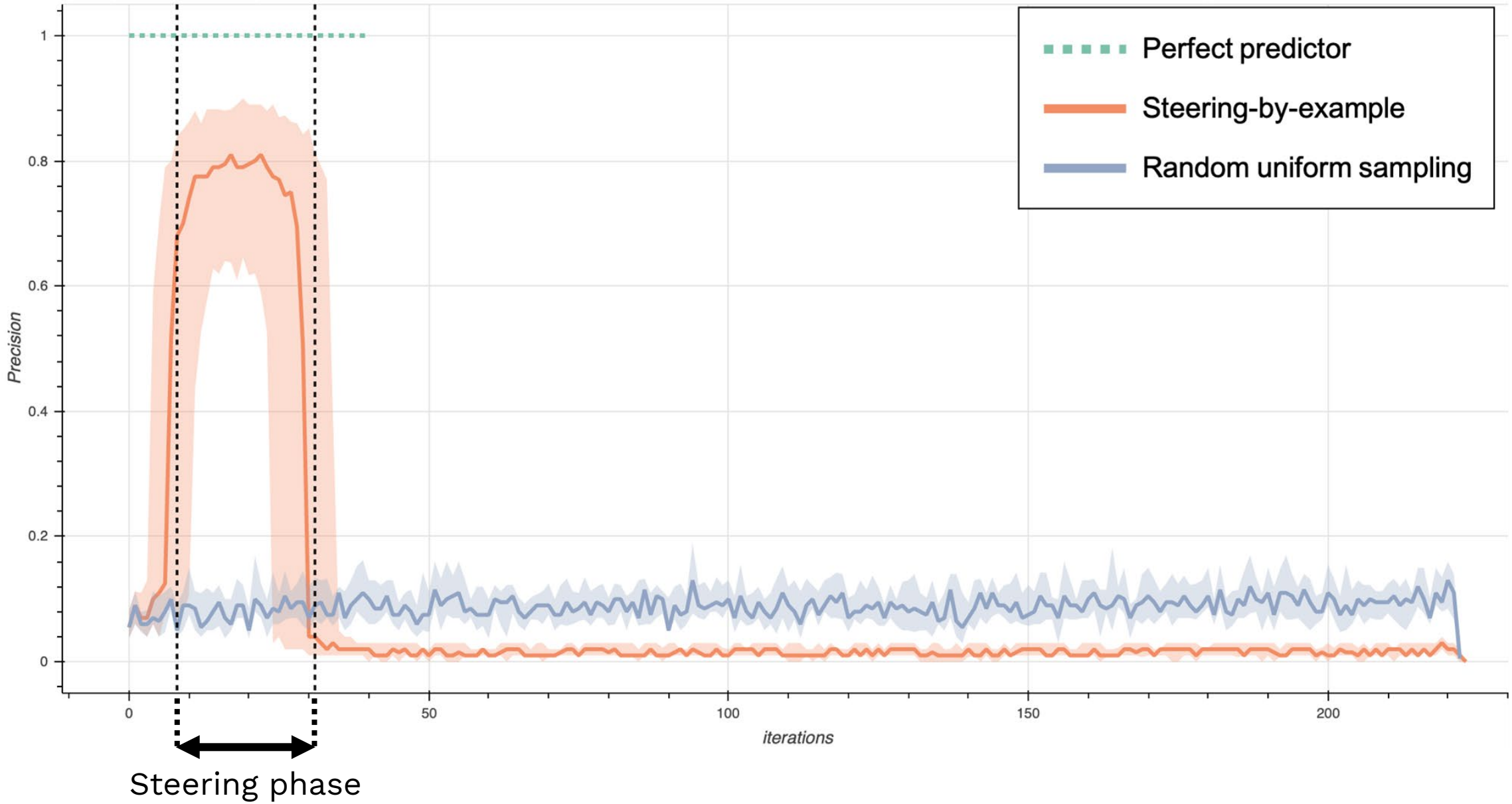
1.4 Benchmarks

Setup

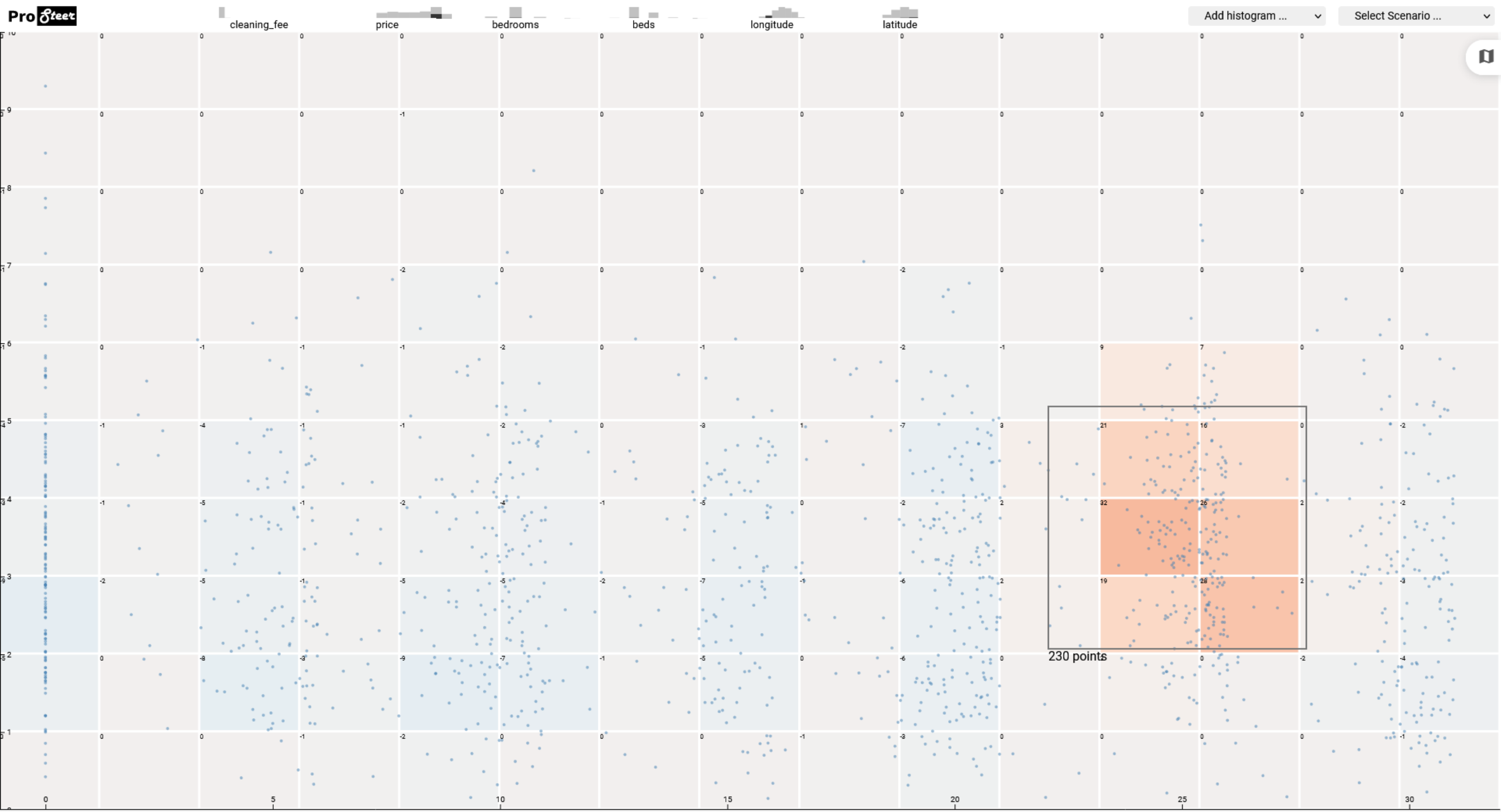
- Tested approaches:
  - steering-by-example
  - random uniform sampling (baseline, lower-bound)
  - hypothetical "perfect predictor" (upper bound)

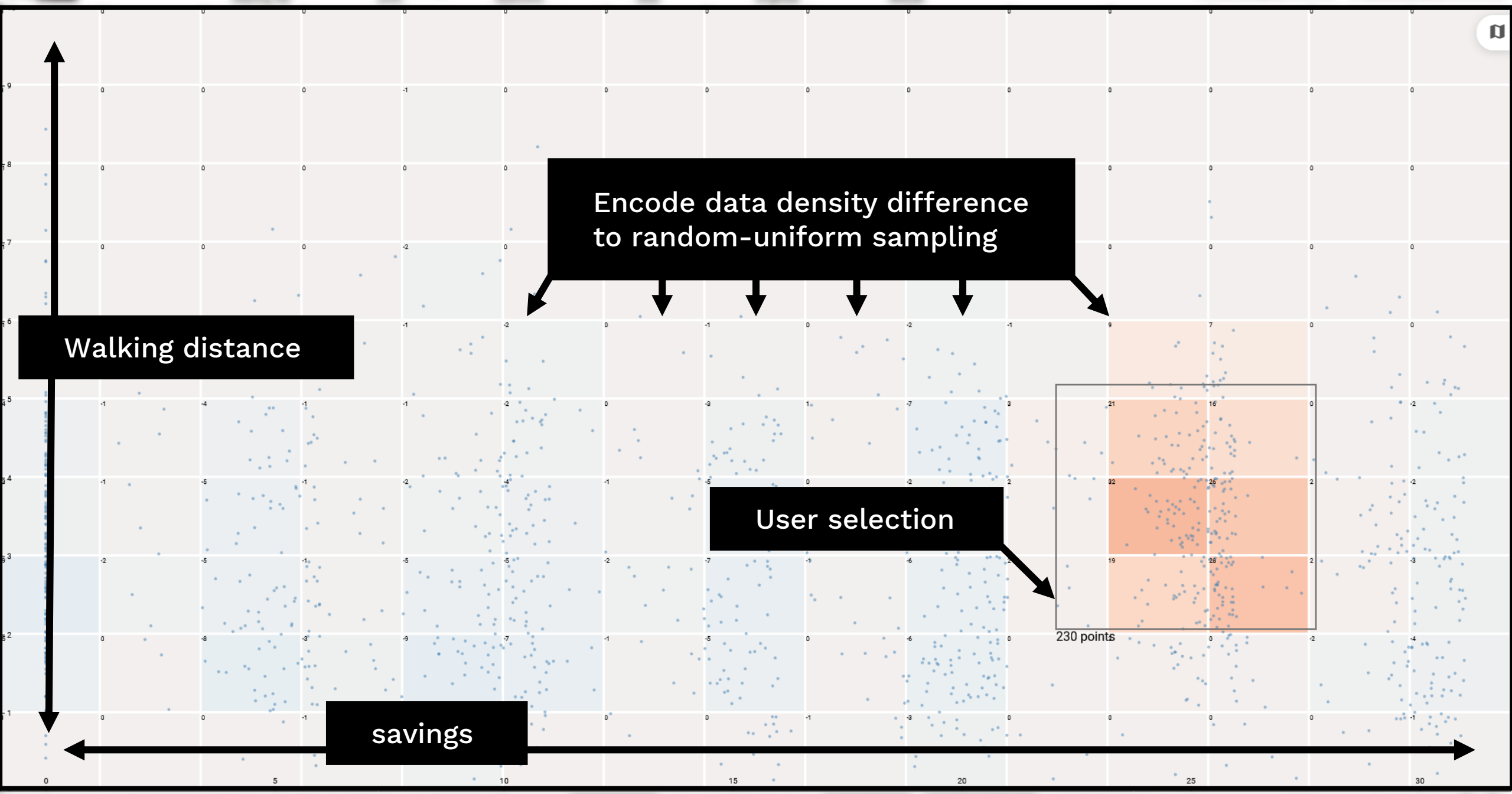
- Test cases:
  - 3 chunk sizes {50, 100, 150} items per iteration
  - 6 thresholds {10, 20, 40, 60, 80, 100} before tree is trained
  - 60 selections with different cardinality {high, medium, low}

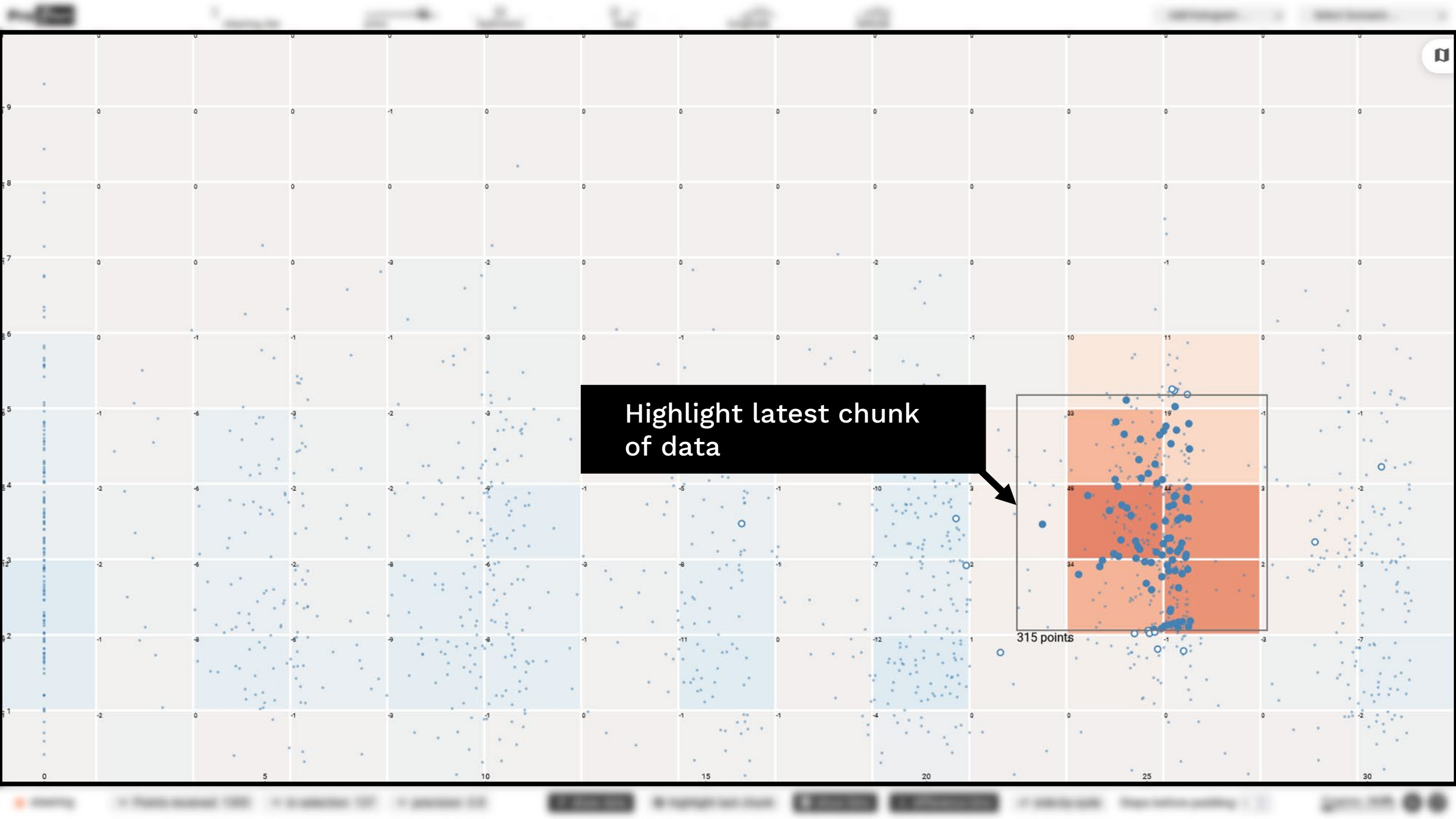

Recall
iterations
0
0.2
0.4
0.6
0.8
1
0
50
100
150
200
Perfect predictor
Steering-by-example
Random uniform sampling
Steering phase

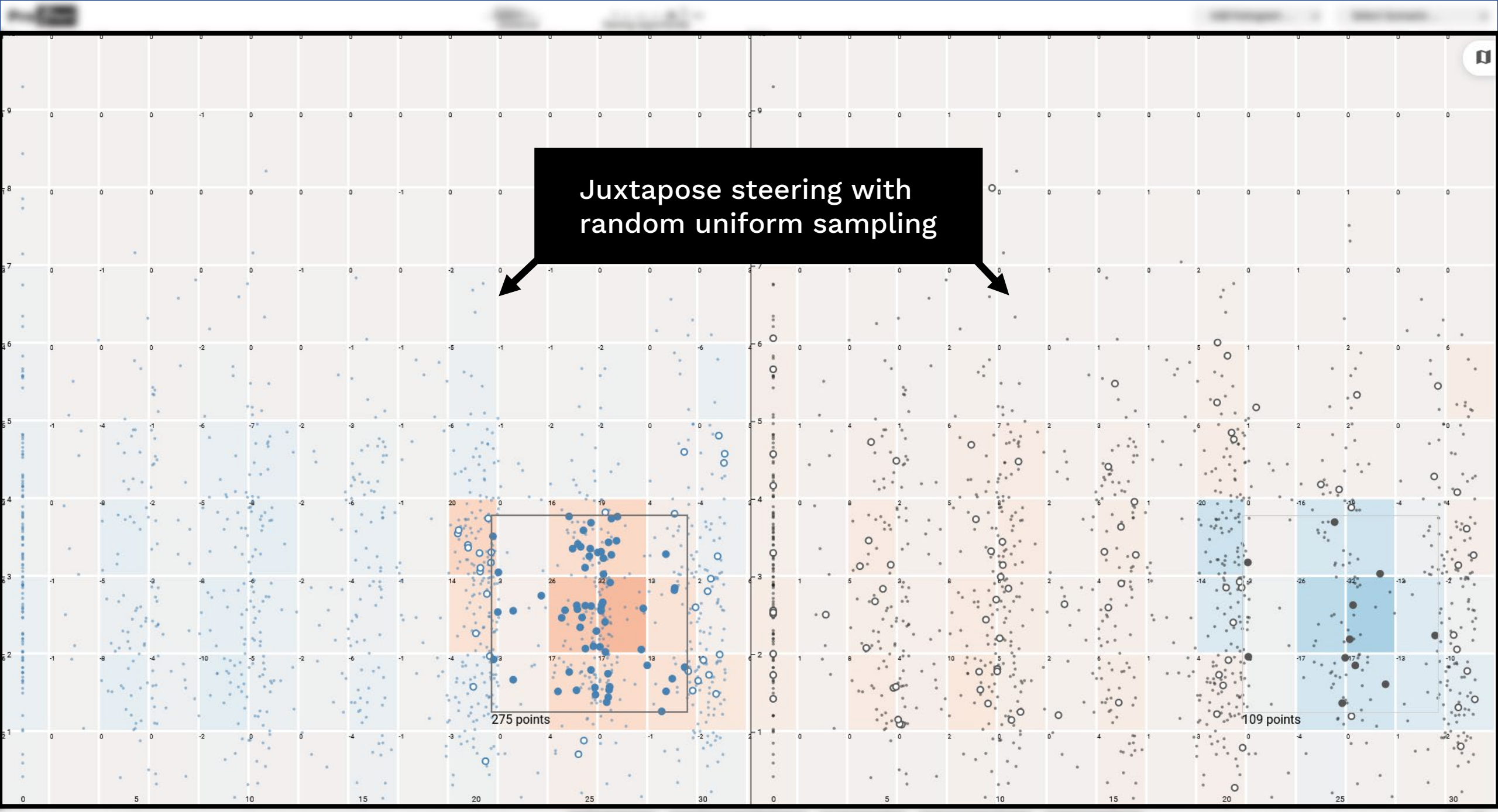

Perfect predictor
Steering-by-example
Random uniform sampling
Precision
iterations
1
0.8
0.6
0.4
0.2
0
0
50
100
150
200
Steering phase

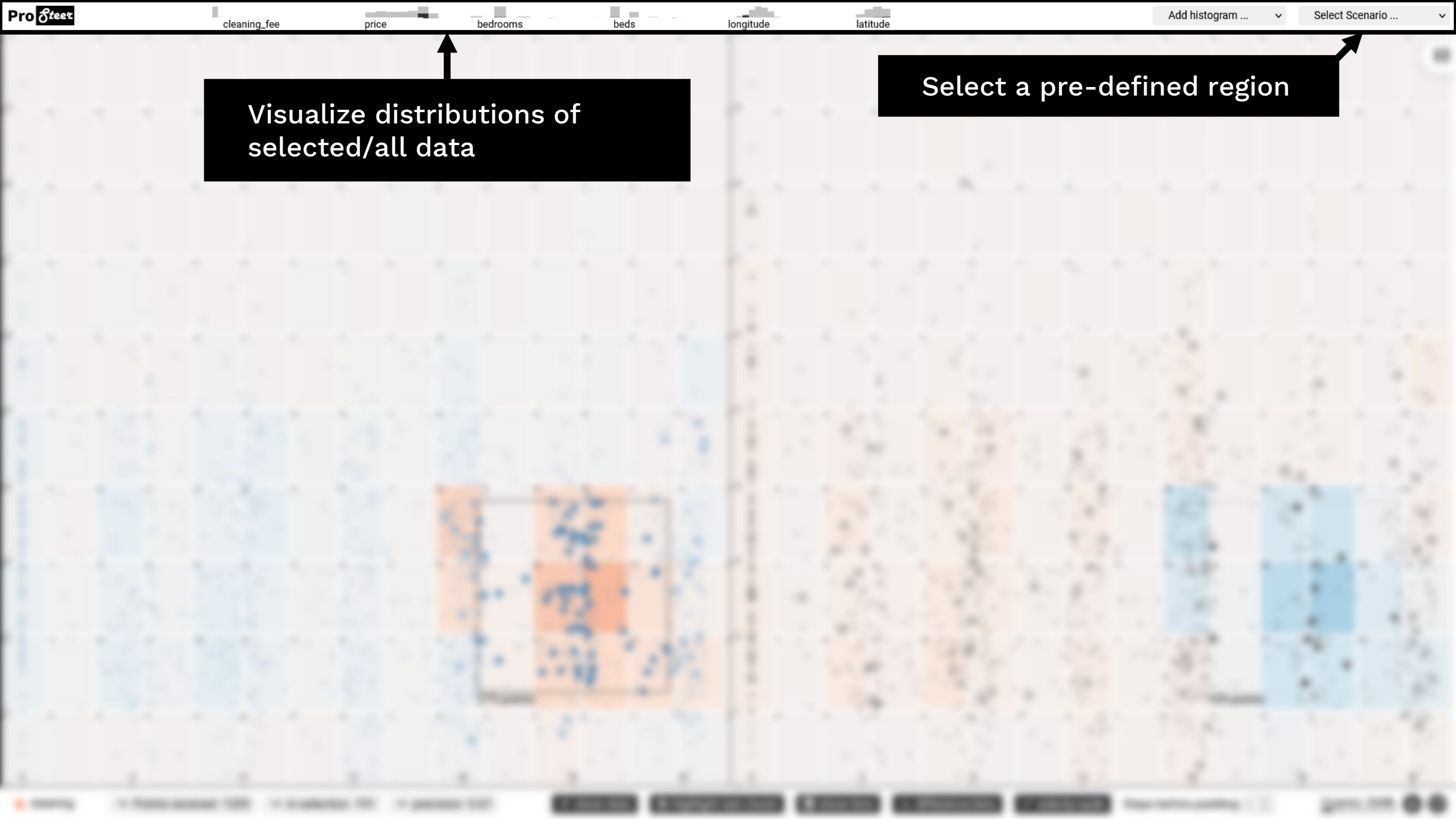
2 ProSteer's Interface

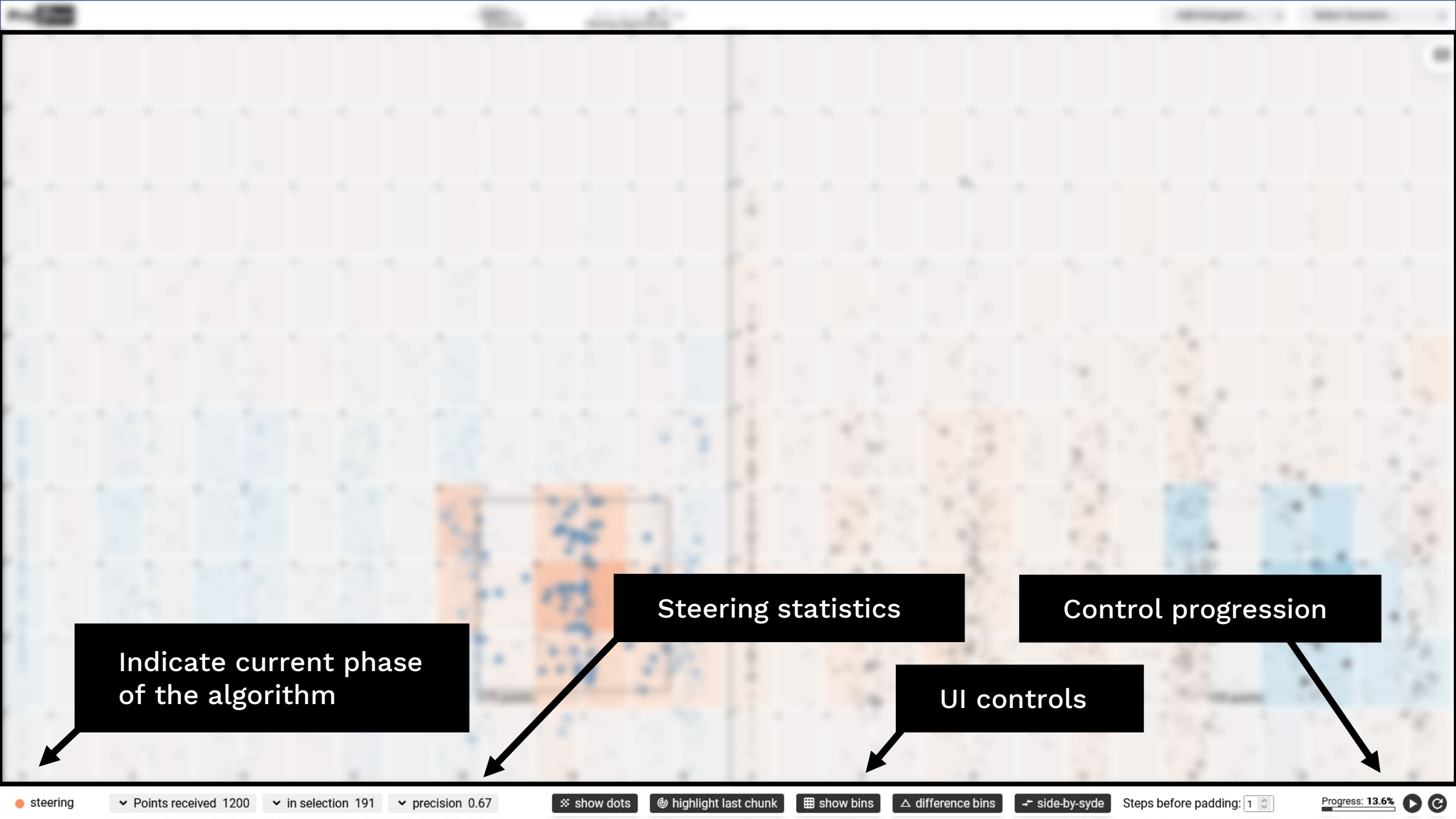

ProSteer
cleaning_fee
price
bedrooms
beds
longitude
latitude
Add histogram ...
Select Scenario ...
230 points
steering
Points received 1300
in selection 137
precision 0.8
show dots
highlight last chunk
show bins
difference bins
side-by-syde
Steps before padding: 1
Progress: 14.8%


Encode data density difference
to random-uniform sampling
Walking distance
User selection
230 points
savings

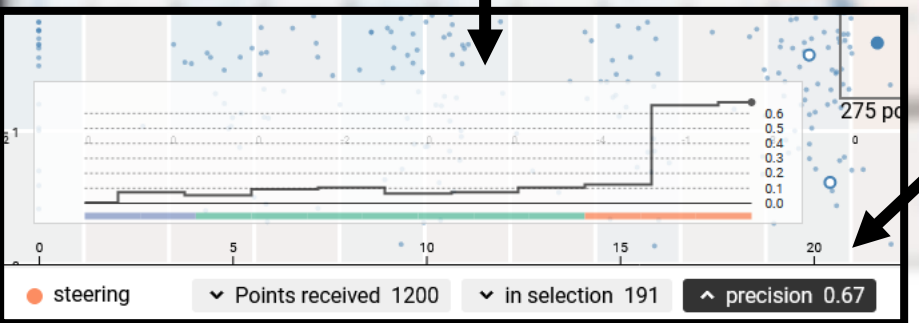

Highlight latest chunk of data
315 points


Juxtapose steering with
random uniform sampling
275 points
109 points


cleaning_fee
price
bedrooms
beds
longitude
latitude
Add histogram ...
Select Scenario ...
Visualize distributions of selected/all data
Select a pre-defined region


Indicate current phase of the algorithm
Steering statistics
UI controls
Control progression
steering
Points received 1200
in selection 191
precision 0.67
show dots
highlight last chunk
show bins
difference bins
side-by-syde
Steps before padding: 1
Progress: 13.6%

Statistic over time



Steering statistics

3 Demo