Visual Analysis of **High-Dimensional Event Sequence Data** via **Dynamic Hierarchical Aggregation**

David Gotz, Jonathan Zhang, Wenyuan Wang, Joshua Shrestha, David Borland

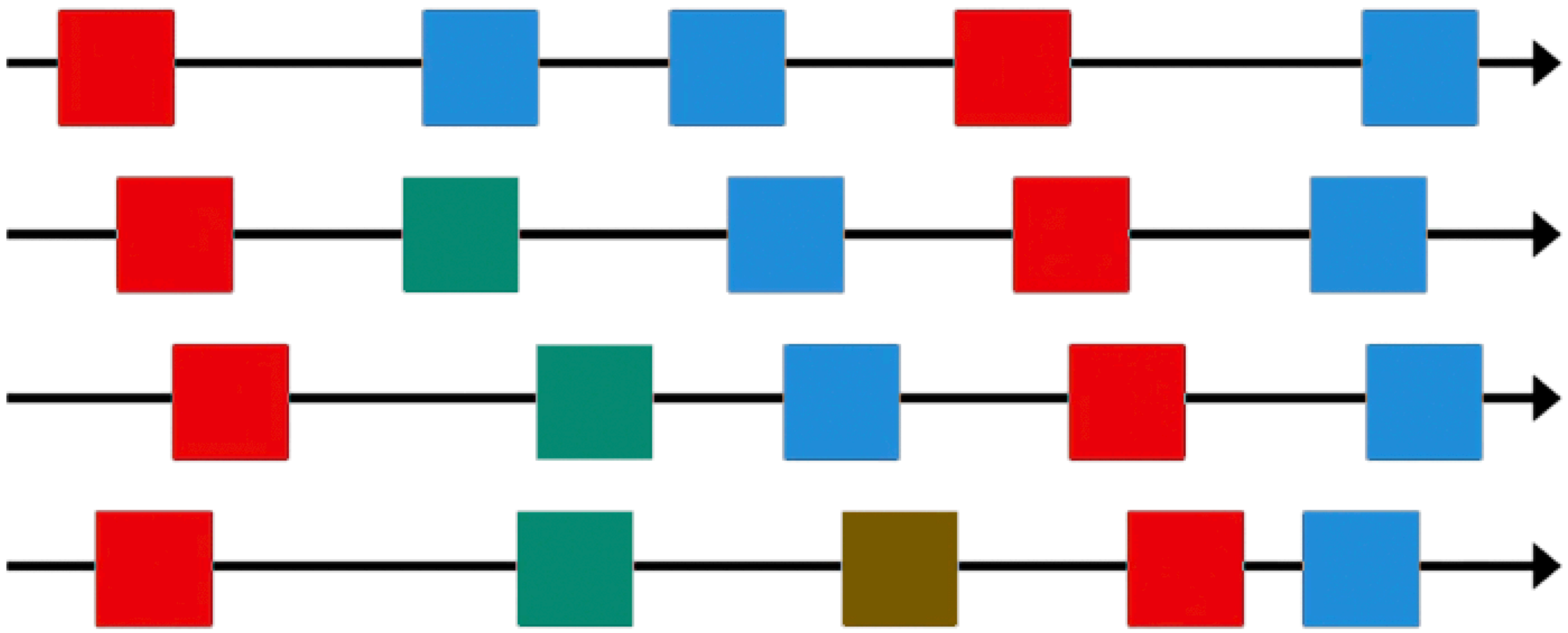
University of North Carolina at Chapel Hill

IEEE Transactions on Visualization and Computer Graphics, 2019

CPSC 547 | Kevin Chow

Event Sequences

- Time-ordered lists of discrete events
- Analyze to discover **patterns** or **rare event paths**
- But… real-world datasets are large and complex:
  - Volume and length of event sequences
  - High-dimensional event data



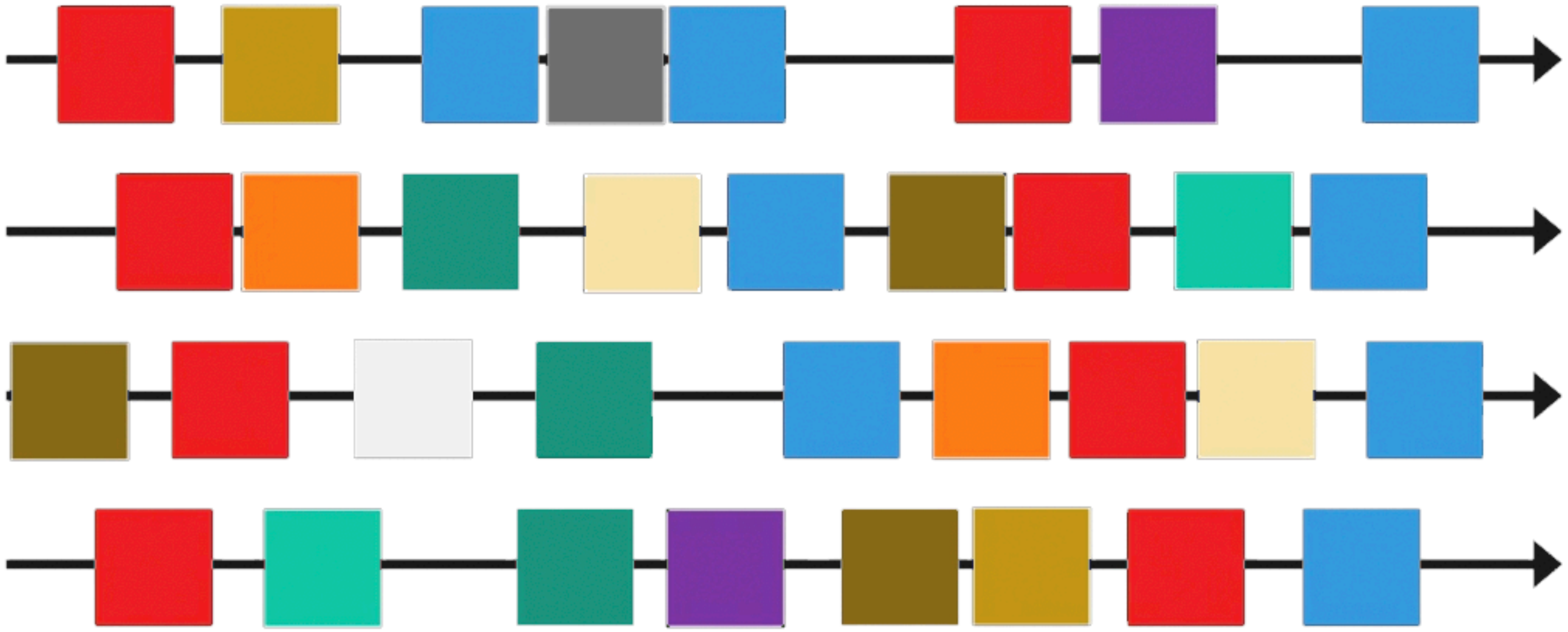
Volume and length of event sequences

Volume and length of event sequences

Aggregate sequences

Volume and length of event sequences

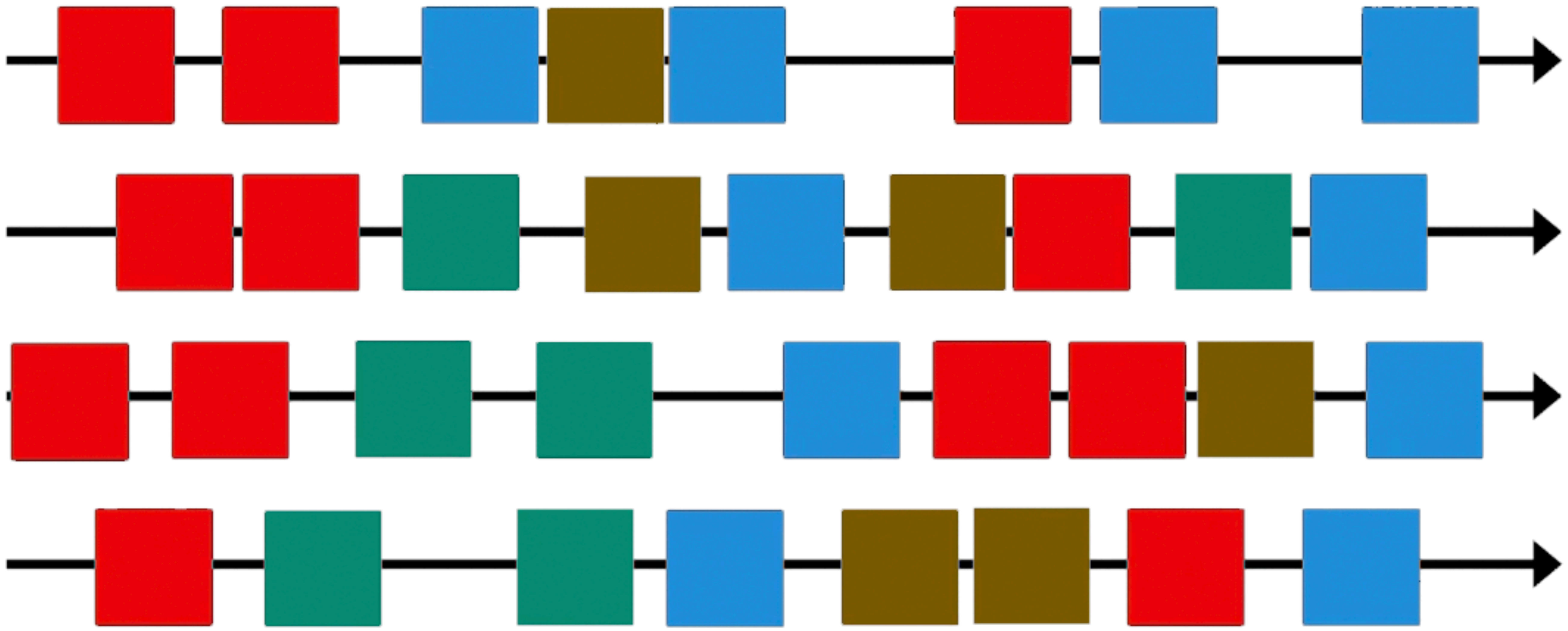
Aggregate sequences



High-dimensional event data

Volume and length of event sequences

Aggregate sequences

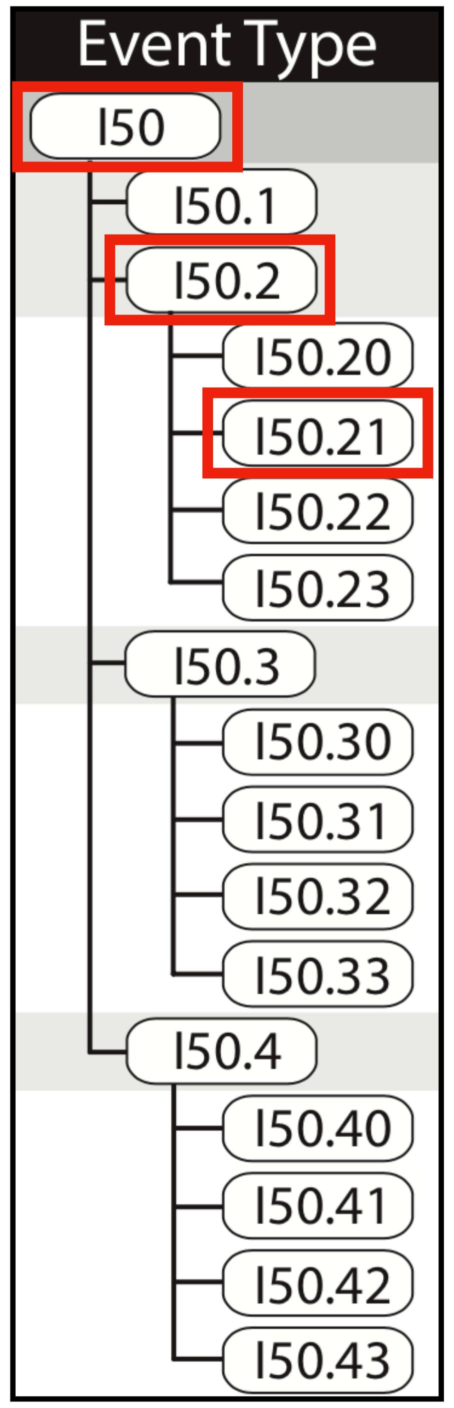


High-dimensional event data

Group events

Grouping Events

- Typically, events are grouped in a **pre-processing** step
- Requires **foreknowledge** and **expertise** about events



Event type hierarchy

ICD-10 Coding System

I50: Heart Failure

I50.2: Systolic Heart Failure

I50.21: Acute Systolic Heart Failure

……

Grouping Events

- Can't change event groups interactively
  - May want multiple groupings — **different levels of detail**
  - An ideal grouping may not exist — **data- and task-dependent**

Cadence

Visual Analysis for Medical Event Sequences

Cadence

Visual Analysis for Medical Event Sequences

Cadence

Visual Analysis for Medical Event Sequences

Cadence

Visual Analysis for Medical Event Sequences

Dynamic Hierarchical Aggregation

Dynamic Hierarchical Aggregation

1. Determining an **optimal** and **adjustable** level of grouping events based on an informativeness score

Dynamic Hierarchical Aggregation

1. Determining an **optimal** and **adjustable** level of grouping events based on an informativeness score
2. Supporting navigation of the **event type hierarchy** with a scatter-plus-focus visualization

Dynamic Hierarchical Aggregation

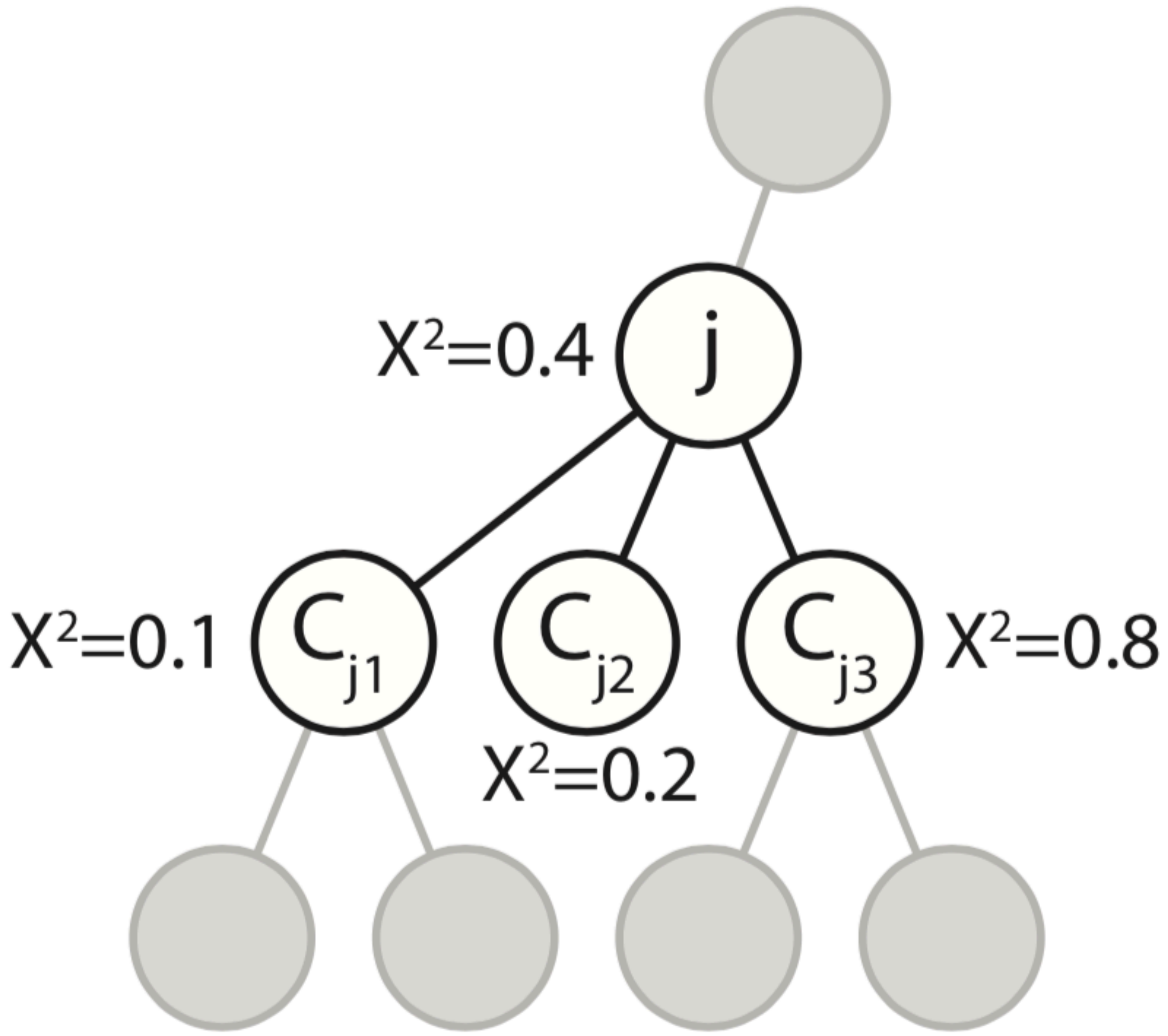
1. Determining an **optimal** and **adjustable** level of grouping events based on an **informativeness score**
2. Supporting navigation of the **event type hierarchy** with a **scatter-plus-focus** visualization
3. **Scenting** to enable discovery of interesting event types

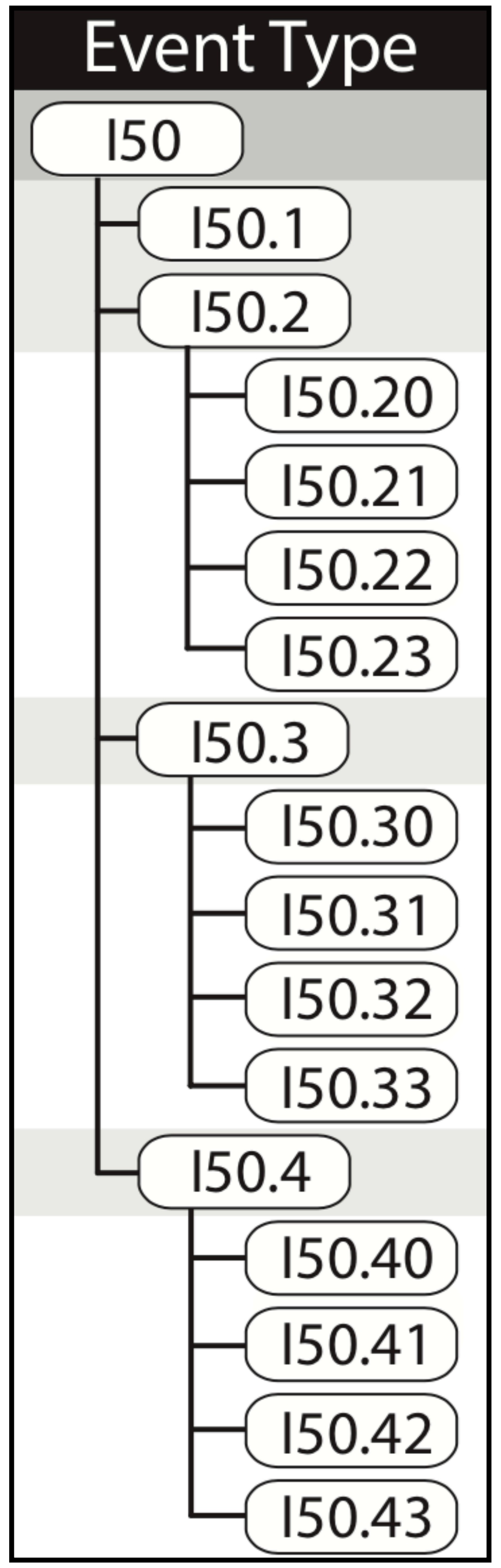
Informativeness Score

- Computed for each event type j in the event type hierarchy
- Measures the **strength of the association** between an **event type** and the **outcome**
    - If this patient had outcome v, did they also experience event type j?
- Based on the **chi-square test statistic** X_j^2

Algorithm: **Optimal Grouping Level**

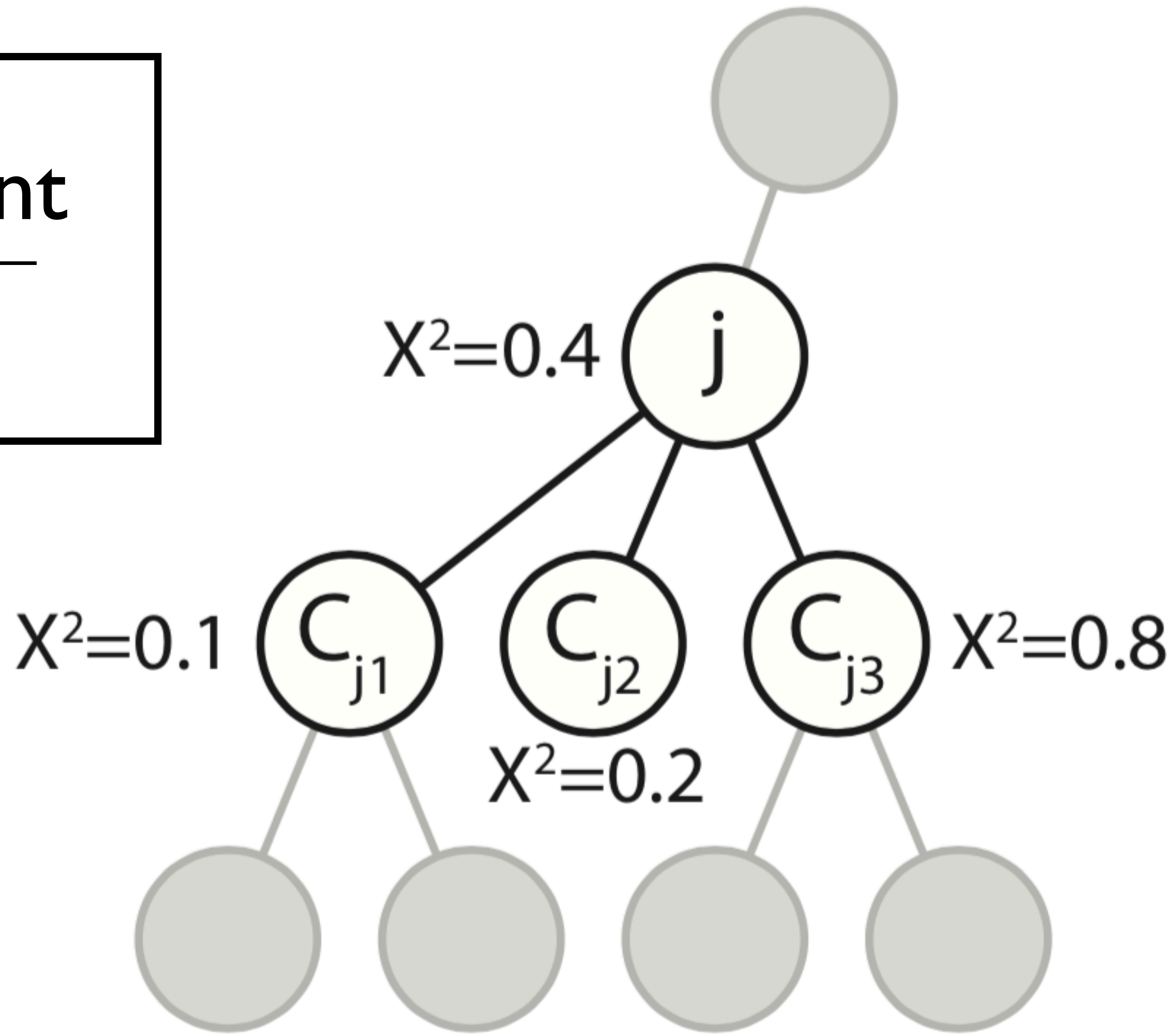
- <u>Goal</u>: Determine the **most informative cut** through the event type hierarchy
- Recursively traverse event type hierarchy
- **Compare informativeness score of parent with each child**





Algorithm: **Optimal Grouping Level**

$$R_j = \frac{\text{\# of children more informative than parent}}{\text{total \# of children}}$$

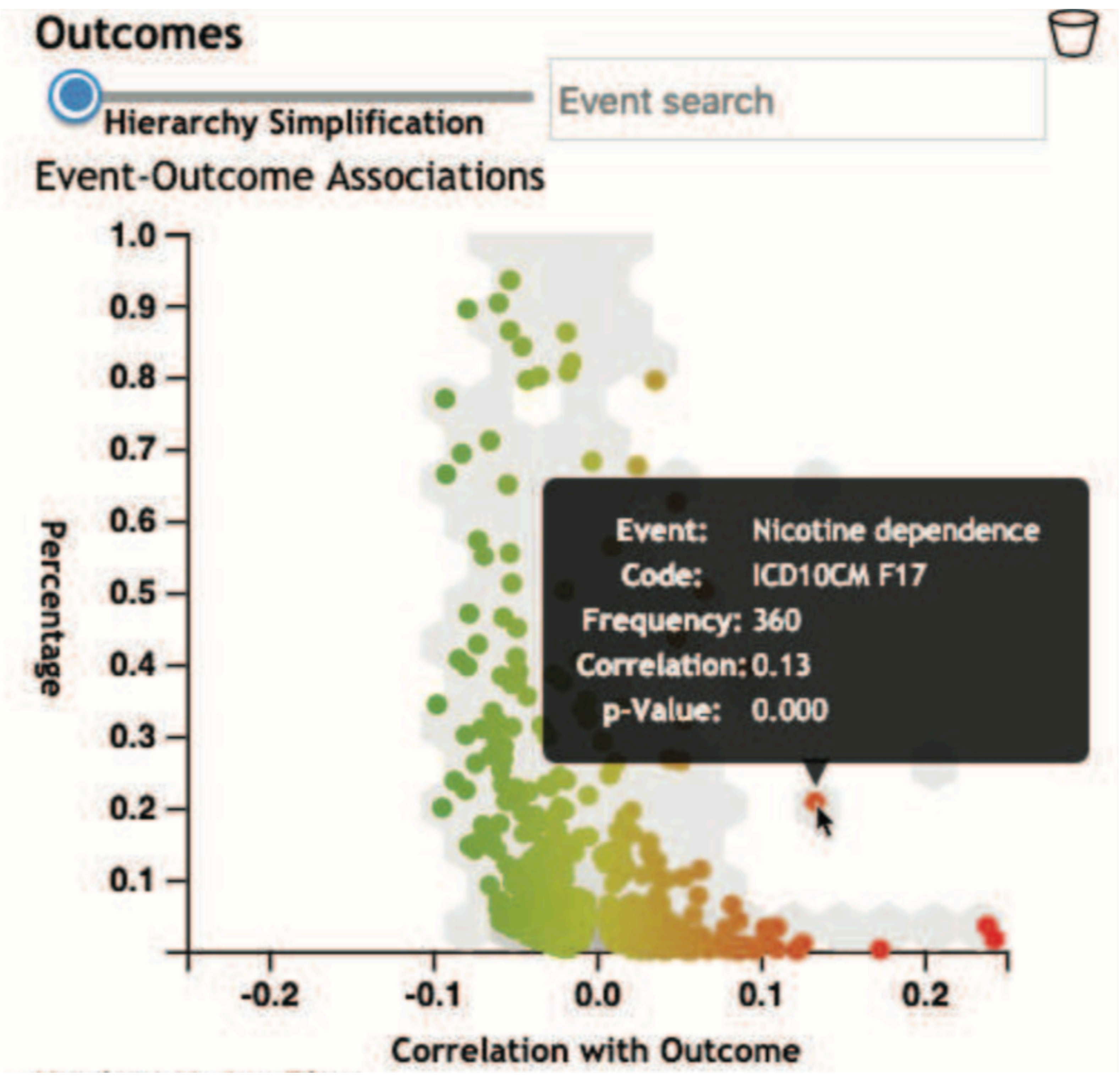


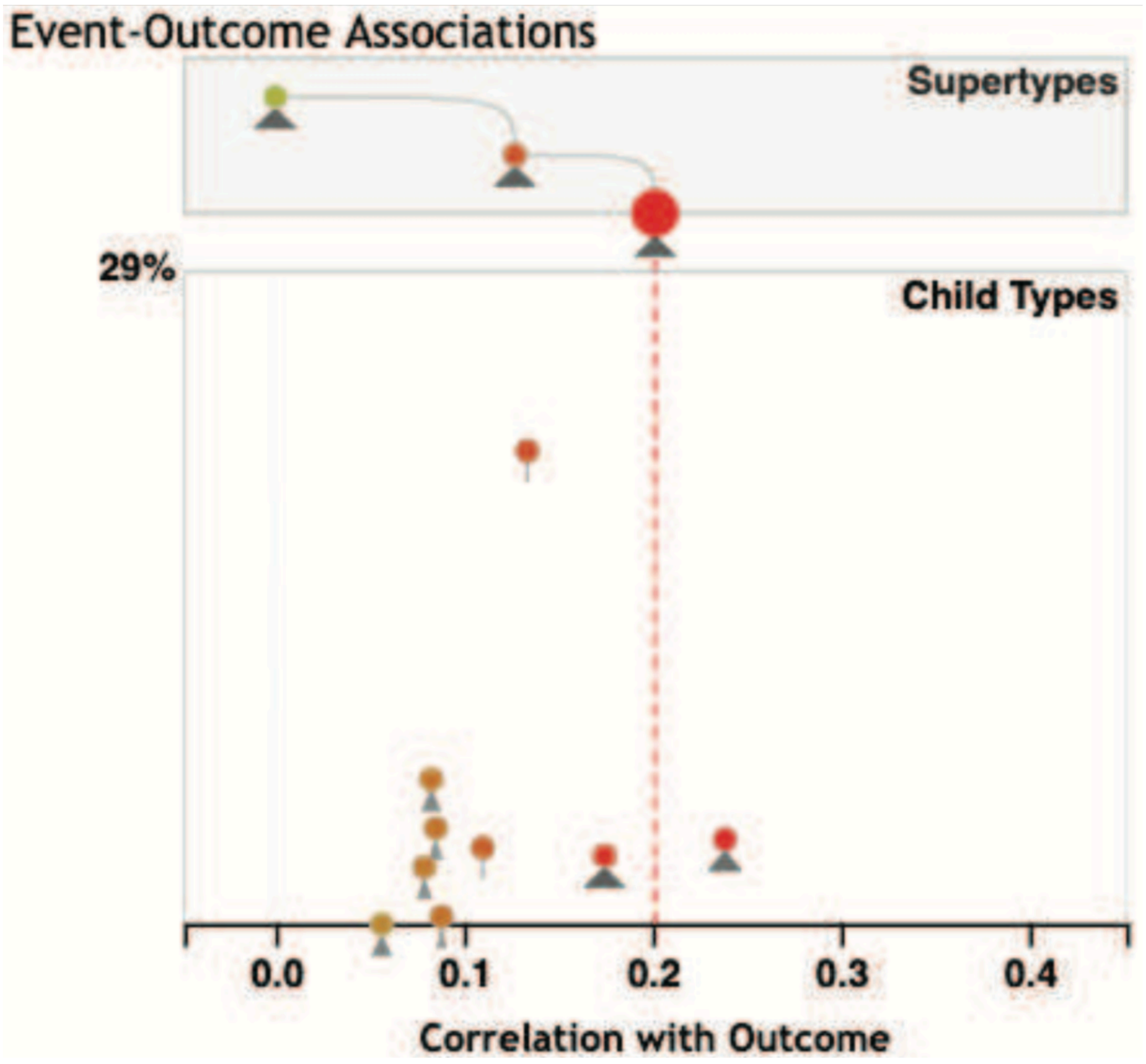
Add j to cut if: (else, recurse)

1. No more children (leaf)
2. $R_j \leq R$ where $0 \leq R \leq 1$

R controls level of aggregation (larger = more aggregation)

Scatter-plus-Focus



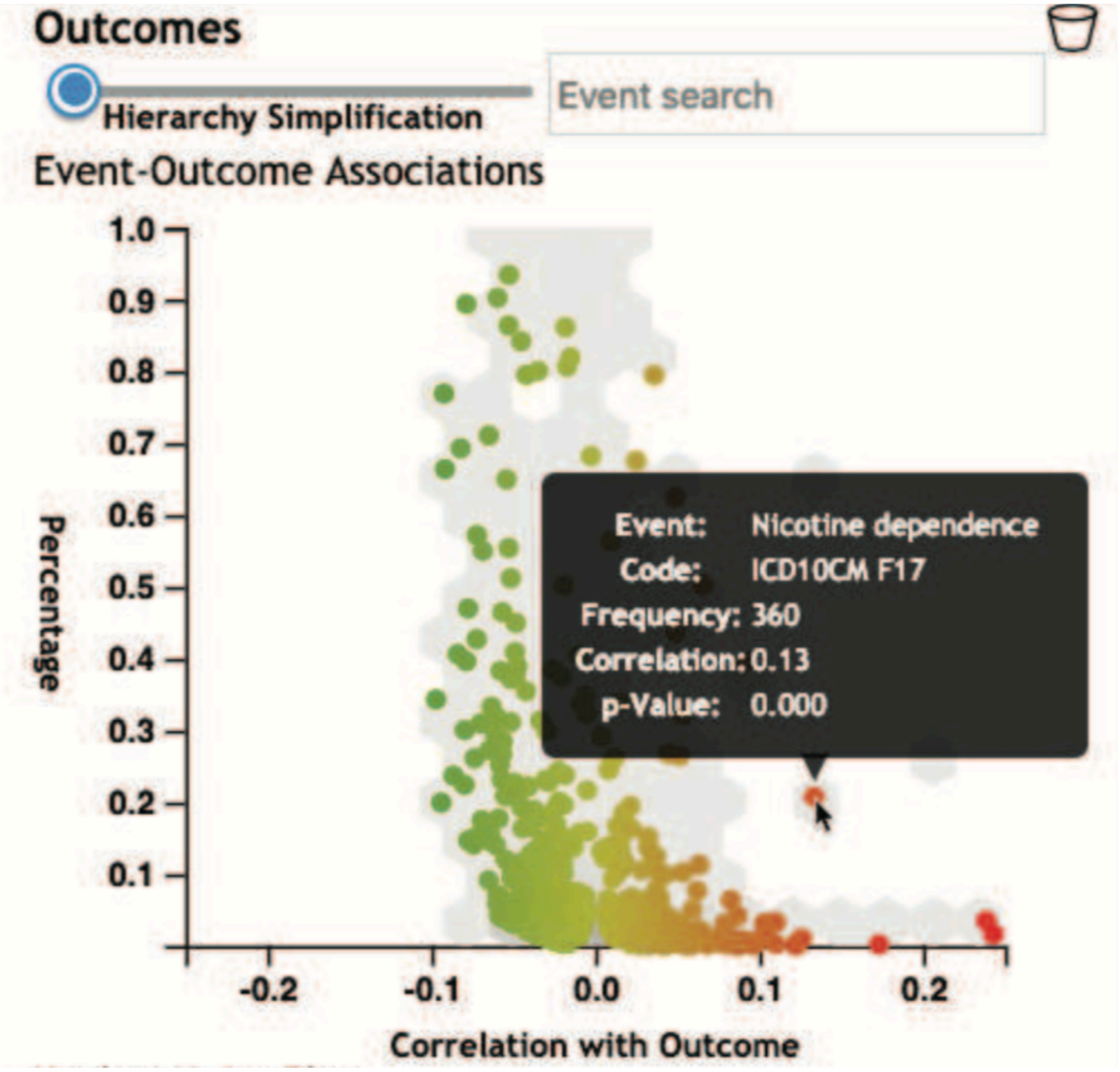


Scatter plot

Focused dual-view

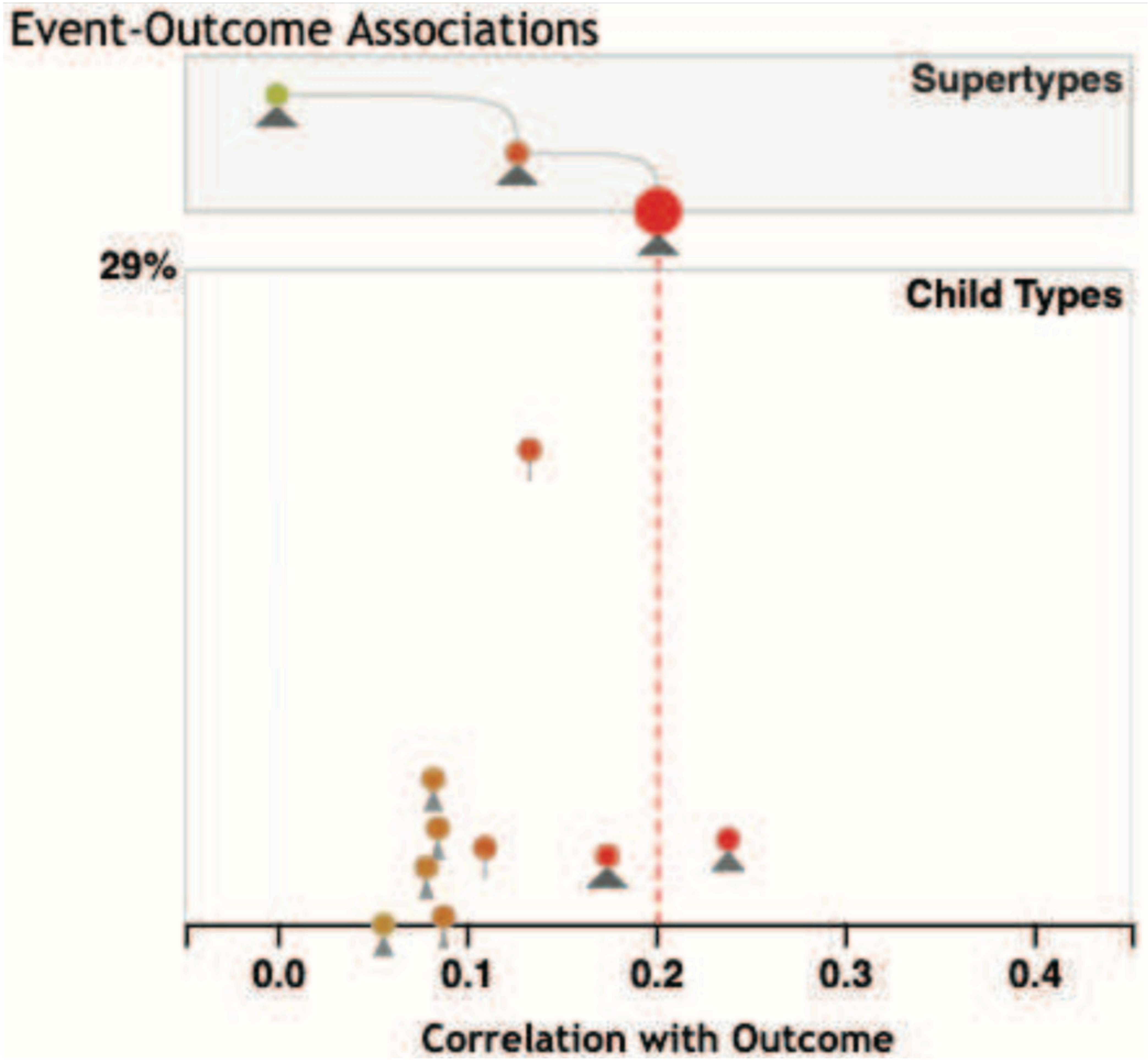
Scatter-plus-Focus

- Challenges of **overplotting!**
- Grey hexes hint at **density** of all possible event types
- Marks are only event types part of **informative cut**
  - Control R with slider



Scatter-plus-Focus

- Focuses on **hierarchy of selected event type**
- X-axis is centred on correlation
- Y-axis: determined by **optimization-based layout algorithm**



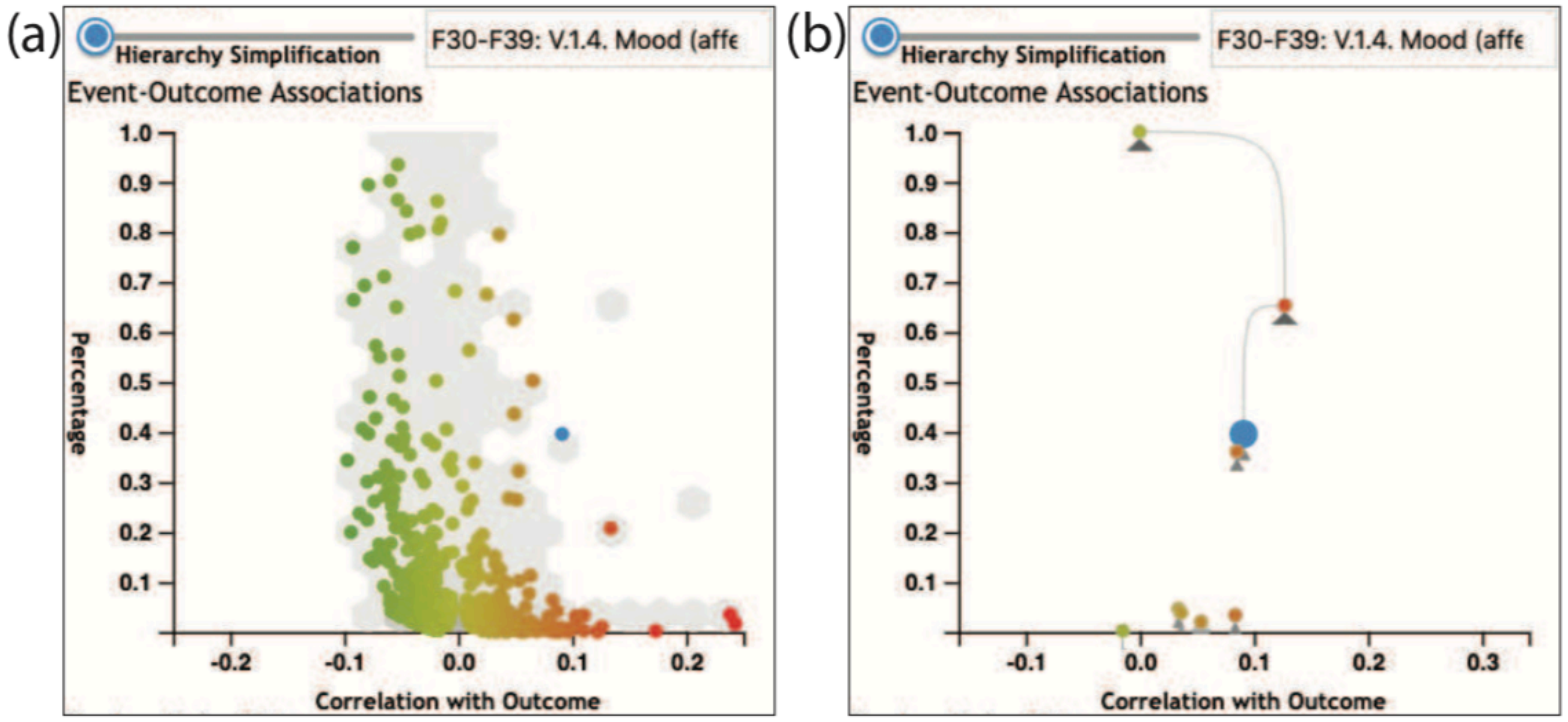
Algorithm: **Optimize Layout**

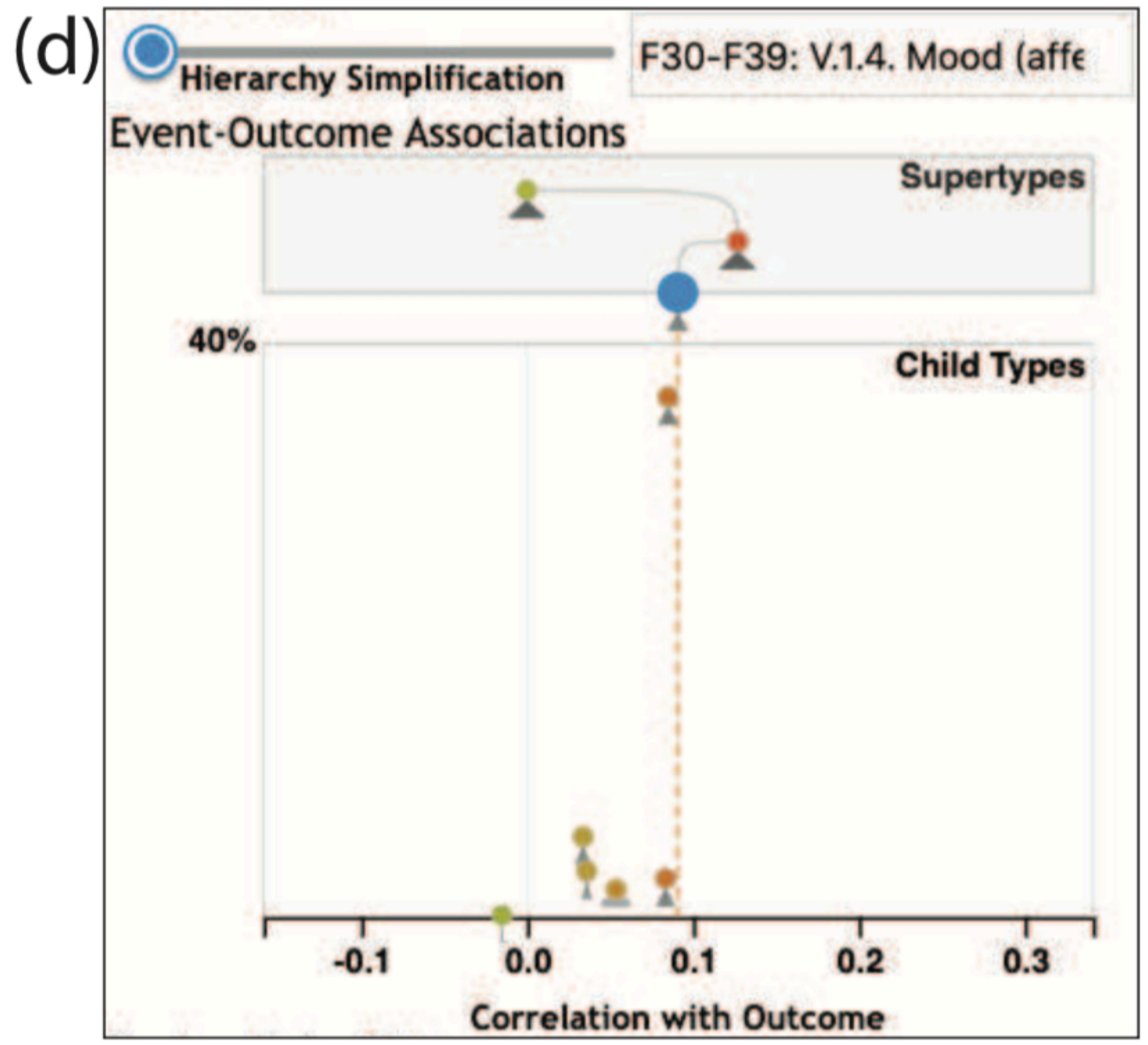
- Cost function that balances **two layout priorities:**
  - Y-positions should be **close to original in scatter view**
  - Marks should **not** overlap
- Two constraints:
  - Optimized y-positions must be within y-axis scale
  - Original y-position **order of marks** must be preserved

Algorithm: **Optimize Layout**

No changes to y-positions

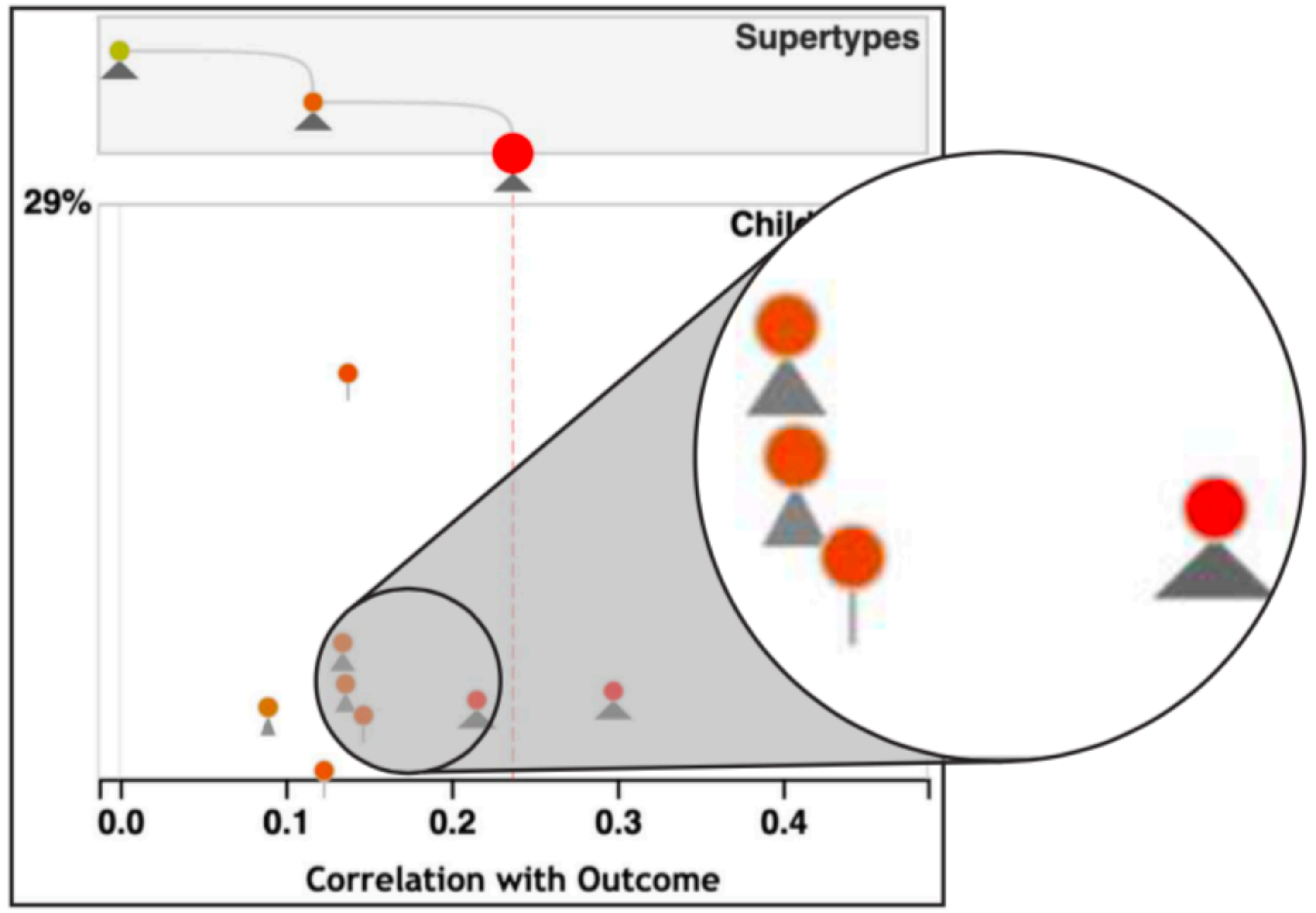
With algorithm

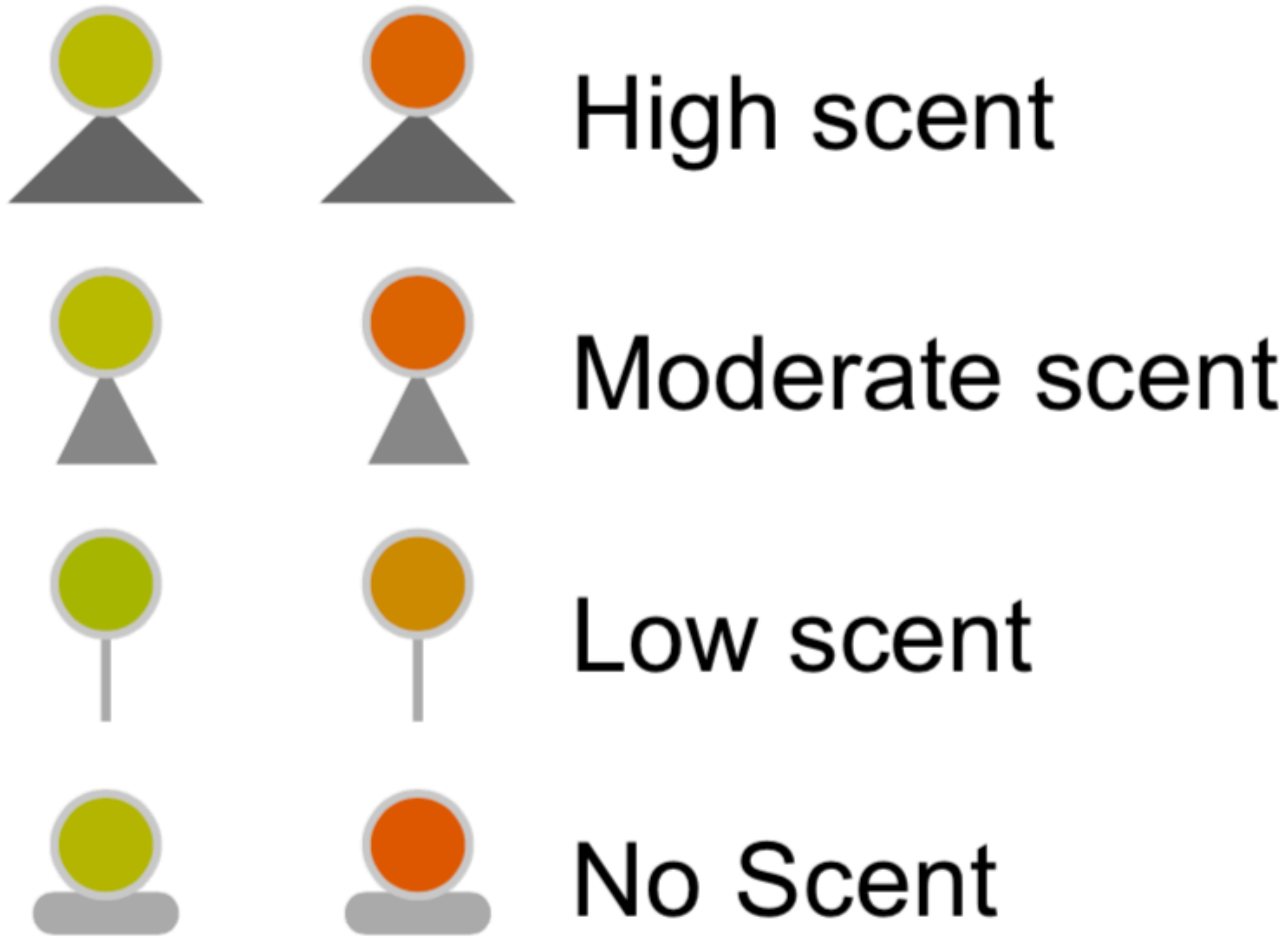




Scenting

- Shows up when exploring type hierarchy in focused view
- Scent value: **range of correlations to outcome** in children
- Size of glyph indicates magnitude of scent value





Evaluation

- **3 medical experts:** health researchers with data analysis experience
- Hands-on demonstration and semi-structured interviews
- Results from **thematic analysis:**
  - Training is required
  - Automated selection of aggregation level useful
  - Navigating through event type hierarchy was intuitive

What-Why-How Analysis

What: Data

- Tree (event type hierarchy)
- Table (patient data)

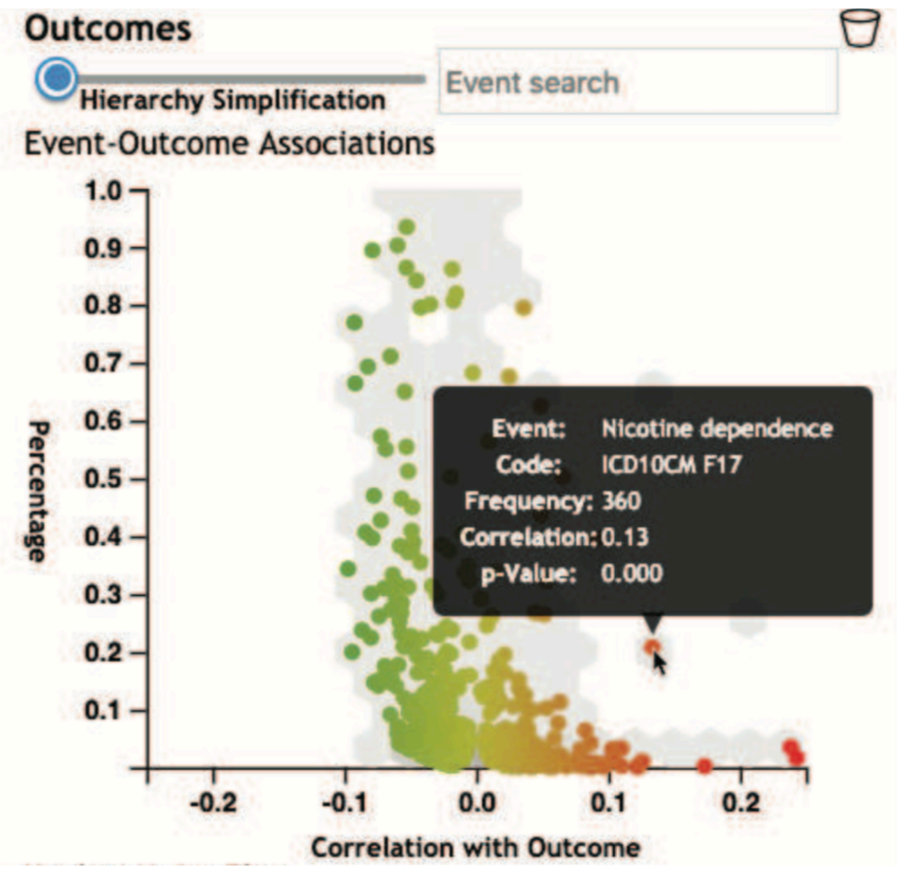
What: Derived

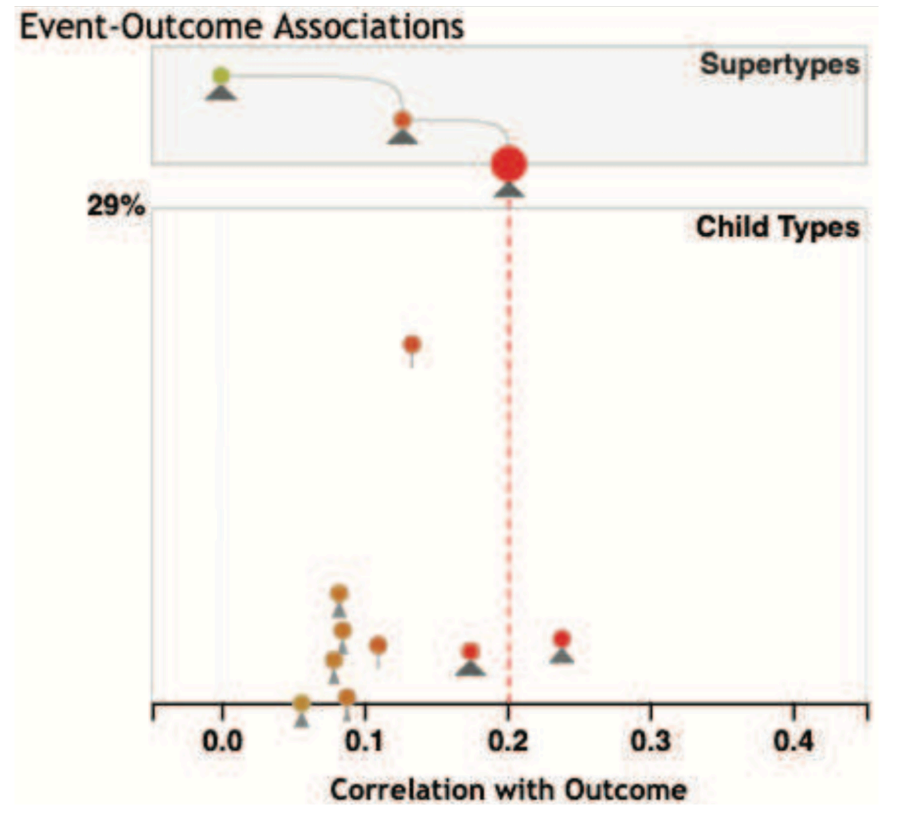
- Optimal event grouping
- Informativeness score, scent value, optimized y-positions

Why

- Discover and produce (event type groupings)

Scale: 5,000 patients, 700,000 events, 10,000 unique event types





What-Why-How Analysis

How: Encode

- Scatterplots
- Color (outcome correlation)

How: Reduce

- Item aggregation (grouping event types)
- Scenting (picking event type)

How: Change

- Select (mark in scatter)

How: Facet

- Overview+detail view (scatter-plus-focus)
- Layering (grey hexes in background)

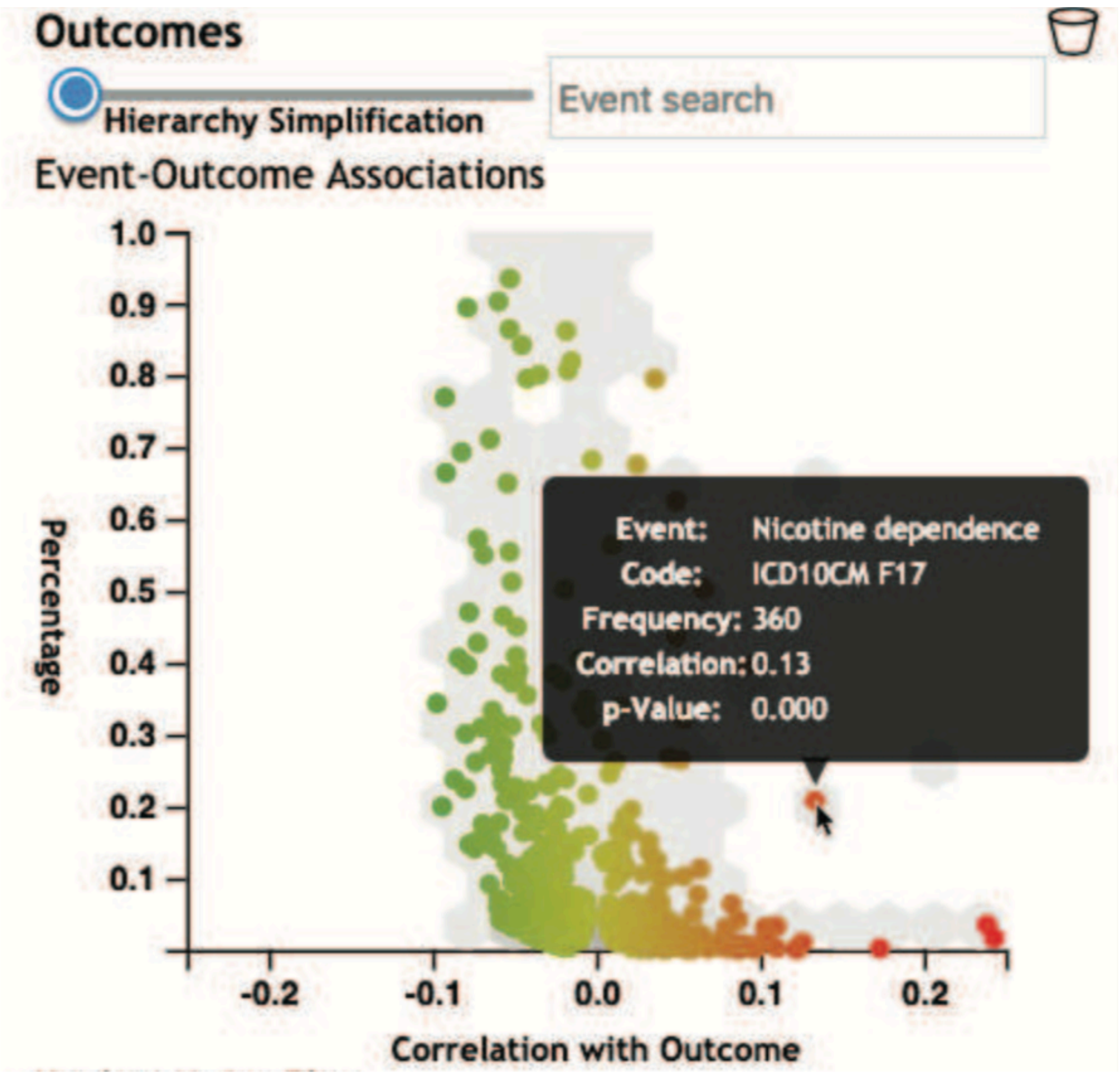
Critique

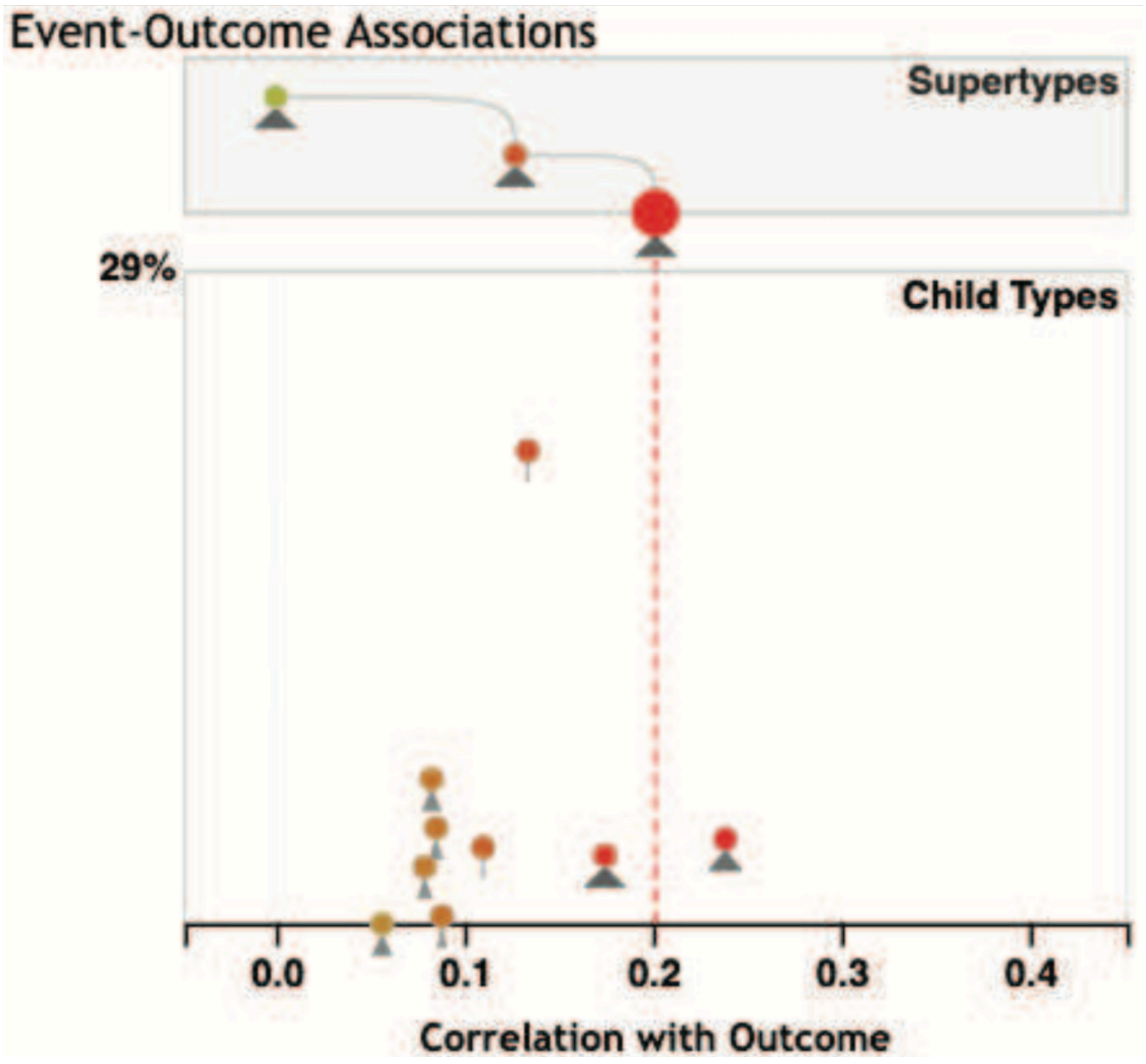
- Strengths
  - Intuitive, simple algorithms
  - Dealt with challenges of occlusion and distortion
  - Switching between views and parameter control reduces load
  - Generalizable to contexts other than health

Critique

- Weaknesses/Limitations
  - Automated approach to aggregation may hide better custom groupings
  - Adding event type groups can be tedious
  - Reliance on tree-based event type hierarchy

Thank You!





Visual Analysis of **High-Dimensional Event Sequence Data** via **Dynamic Hierarchical Aggregation**