

# INFO 290T

## Human-Centered Data Management

Datatone: Managing Ambiguity in NLI for  
Data Visualization



# Announcement: Intermediate Report

- Due **10/31**: 10% of your grade
- Report of up to 5 (two col) pages: references don't count towards the page limit
- Follow the template of any conf. from which we've studied papers
- Introduction – cover:
  - What is the problem,
  - Why is it interesting/important
  - Why hasn't it been done before/what are the limitations of related work (briefly)
  - What is the approach, and what have you accomplished so far
- Related Work – can go right after intro or towards the end
  - Here, we don't just expect a small number of refs, but a thorough exploration of the space
  - Hard to give a rule of thumb for # of references, but <10-15 is usually too few – lots of related work in pretty much any area you will look at
  - Don't just do it to check a box: actually find closest related papers and carefully contrast to your approach



# Announcement: Intermediate Report (contd.)

- Body of the Paper
  - Talk about your approach.
  - Set up the problem more formally;
  - Perhaps describe any preliminaries/background;
  - Your methodology/architecture/design;
  - What have you learned/built/analyzed so far – preliminary findings and evaluation
- Plan for the last month
  - Talk about how you're planning to wrap up the project in the last month
  - If you haven't started thinking about evaluation yet: How will you evaluate? Have this be as detailed as you can:
    - What are the datasets & metrics?
    - Who are you testing with, on what workload?
    - What are the comparison points?
  - Are there any anticipated roadblocks? How are you planning to avoid them?



# Aside: modalities we've considered for vis(ish)...

- GUI/menu-selection, e.g., Polaris/Tableau, Falcon, SeeDB, Profiler
- Direct manipulation, e.g., Excel, Sigma Worksheet
- Sketch, e.g., Qetch
- Gestures (for SQL), e.g., GestureQuery
  
- Still pretty hard, despite all that!



# Democratizing Data Visualization

- BI Tools like Tableau and PowerBI (+Excel) are great at supporting data vis via a GUI ...
- But still hard for many users to map “question” ↔ “interactions” to generate a vis that answers that question
  - Also a cold-start problem
  - Anecdote: ~1% of journalists would be able to operate a GUI tool to get their questions answered, rest send their questions to a data journalist
- *Can we support a NL interface for data visualization?*



# Prior Work (in 2012)

- Prior to GenAI wave
- IBM Watson Analytics & PowerBI
  - User enters NL query
  - System suggests related (canned) queries (Watson) OR autocompletes to them (PowerBI)
  - But: doesn't work beyond the small set of canned queries
- NL-to-SQL work
  - Doesn't quite work for vis aspects



# Main Issue with NL approaches: Ambiguity

- NL system interpretation is often wrong.
  - Side note: gotten much better with LLMs, but mistakes still abound due to ambiguity

State or city?

“Show revenue for NY and Washington DC in 2012”

Could be multiple attributes corresponding to revenue

Is it NY for all years and DC for 2012, or both NY and DC for 2012?

Visual ambiguity:

- Two line charts for each city?
- One line chart with two lines?
- A stacked bar chart (of city) vs. time?
- Two bar charts?
- A single bar chart of aggregate revenue?



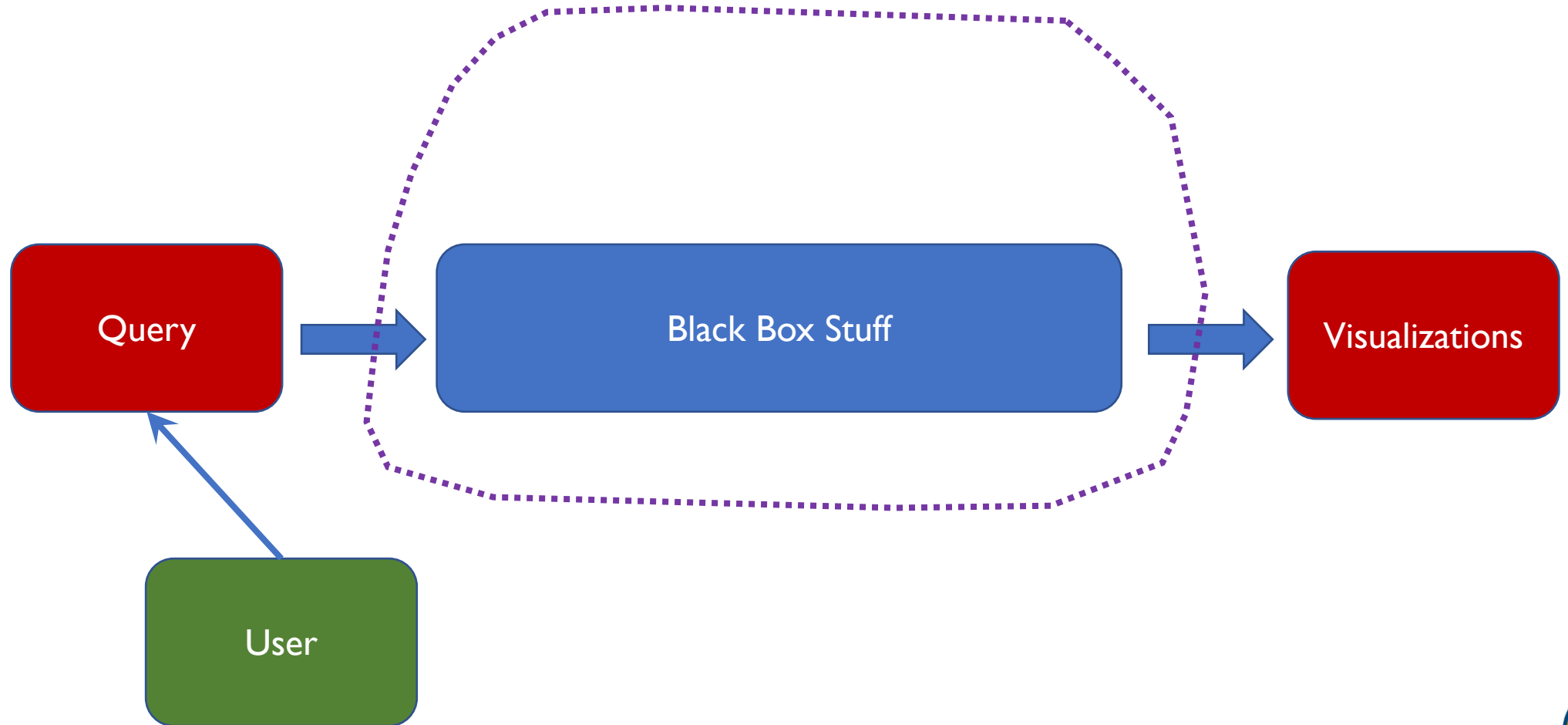
# Main Issue with NL approaches: Ambiguity

- As a user, the only option is to change the query (+ cross your fingers)
  - If you don't get it right the first time, try, try again??
  - How often do you keep trying??
- *Key Question: Can we allow users to tweak system interpretations of human utterances?*

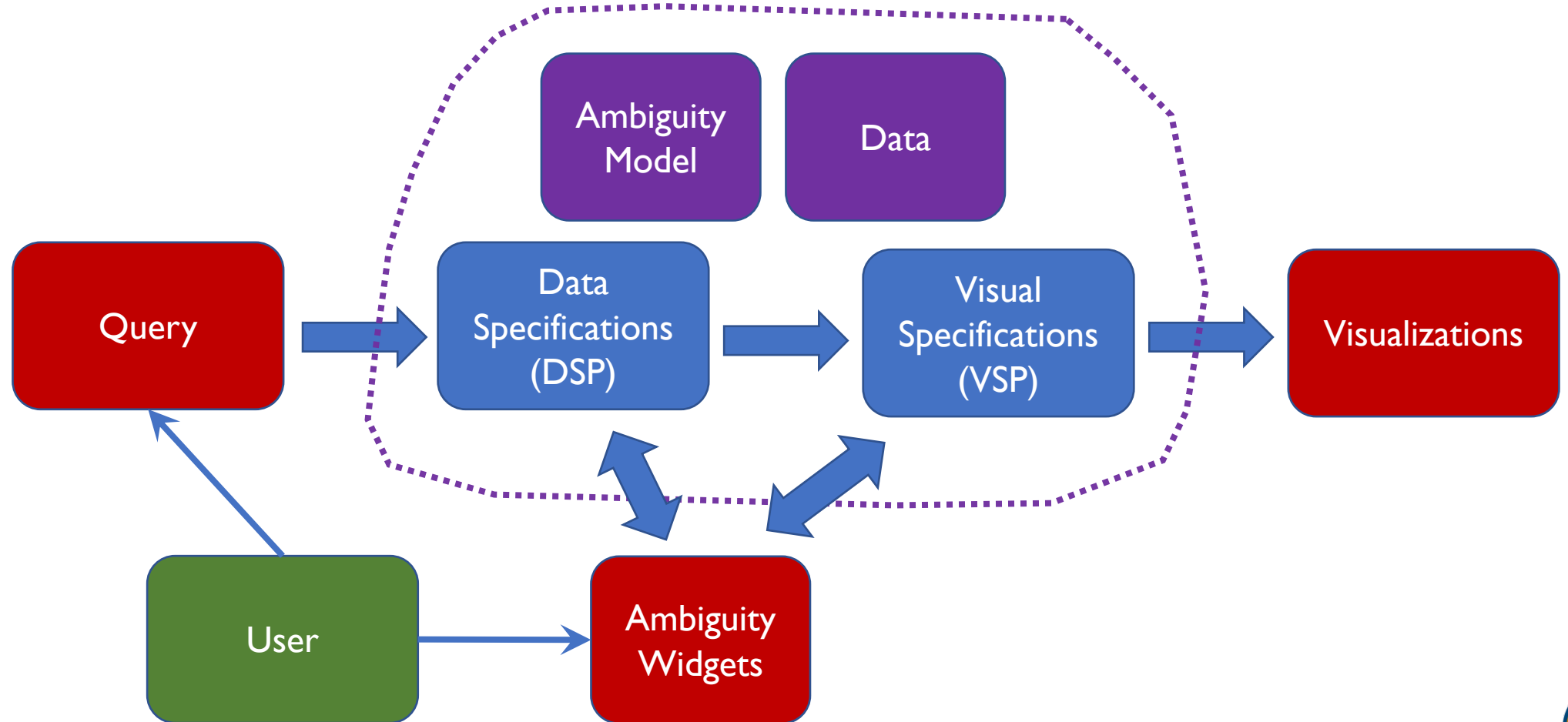




# Prior Conceptual Framework



# Conceptual Framework



# Demo

- <https://youtu.be/yjJ3k8fCGVo?t=219&si=4bXUenKL7gB2dj-W>



# Search Query Bar

Olympic Athletes

show me medals for hockey and skating by country

Click to Speak

Submit

## Data Overview:

**Athlete:** Michael Phelps, Natalie Coughlin, Sun Yang, ...  
**Age:** 15-61  
**Country:** United States, Australia, Russia, ...  
**Year:** 2000-2012  
**Sport:** Swimming, Diving, Cycling, ...  
**Gold Medal:** 0-8  
**Silver Medal:** 0-3  
**Bronze Medal:** 0-3  
**Total Medal:** 1-8



show me medal for hockey and skating by country



Attribute Selector



## Data Ambiguity Widgets

**Dimensions**

Country | Sport

Country

**Chart Templates**



**Color**

Color by Country

Color by Sport

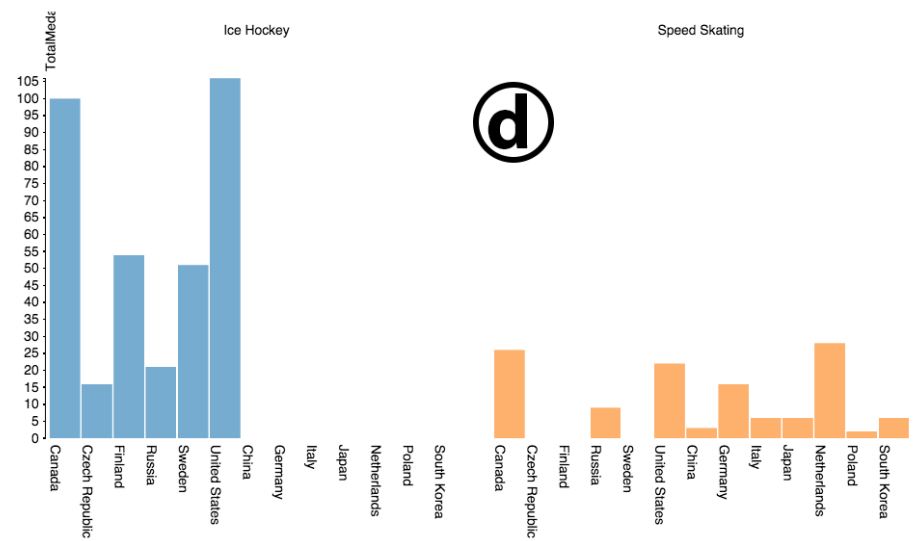
Single Color

**Group Order**

Group by Sport then by Country

Group by Country then by Sport

Sum of TotalMedals (Sport: Speed Skating and Ice Hockey) by Sport, Country



## Sample Queries:

[Empty box for sample queries]

## Visual Ambiguity Widgets



# Approach

- NL Query → Data Specification
- Data Specification → Visual Specification
- Ambiguity Resolution throughout
  
- But first, tokenization and similarity matching to a lexicon + parsing



# Tokenization and Similarity Matching

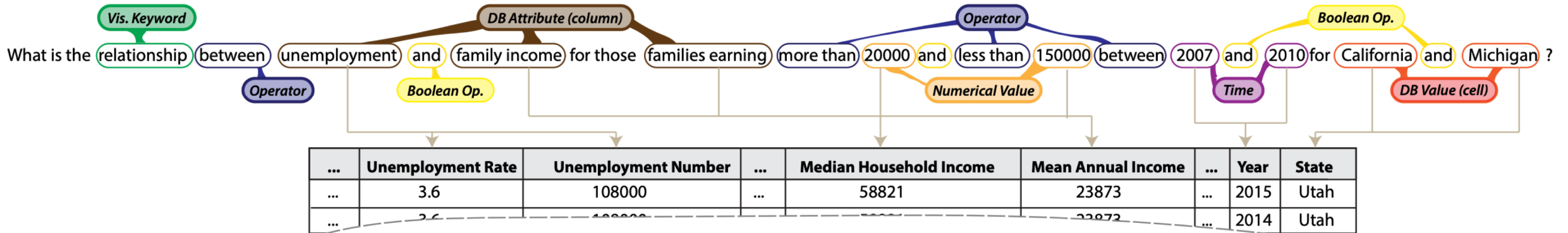
E.g., *show me medals of hockey and skating by country*

Each n-gram matched to one of the following:

1. Attribute names
2. Attribute values
3. Numerical values (Q: why is this different from 2?)
4. Time expressions
5. Data operators and functions, e.g., greater than, equal, sum, ...
6. Visualization key phrases, e.g., trend, relationship, bar chart, ...
7. Conjunction and disjunction terms, e.g., and, between, ...
8. Direct manipulation actions, e.g., add, color

Uses a threshold of match (0.8) – leads to multiple possible match candidates for each token/n-gram





# Next: Relationship Identification

- Stanford CoreNLP Parser used to do parsing
- Parsing yields:
  - relations between data operators and values → a condition
  - attribute and conditions → a filter
  - aggregation phrases and attribute → an aggregate
  - Etc.
- Sidenote: some of this is not too sophisticated & based on heuristics, e.g., how far away is the ordering keyword from the attribute





# NL → Data Specification

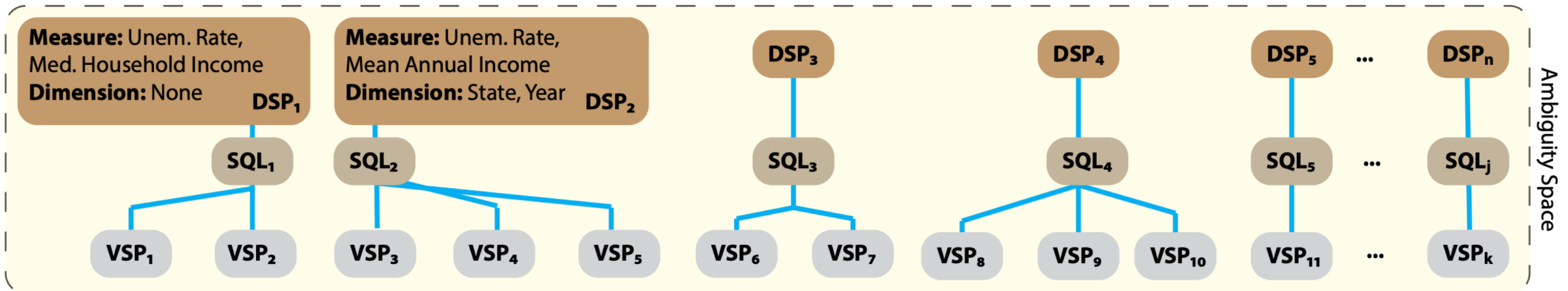
Given **filters** & **aggregates**, plus:  
other unconstrained **attributes** + **order** (plus **attributes** “close by”)

```
SELECT {Aggregates}, {Dimension Attributes} FROM Table  
WHERE {Implicit Filters}  
GROUP BY {Dimension Attributes}  
HAVING {Explicit Filters}  
ORDER BY {Order Attributes}
```

Drop all DSPs with empty results



# Data Specification → Visual Specification



Many Visual Specs per DSP / SQL query

# Visual Specifications (VSP)

- Builds on GoG (recall: alternative to Polaris approach)
- Each VSP specifies:
  - Graphic type
  - X,Y axis
  - Additional Encoding (e.g., Color)
  - Additional Faceting (to get small multiples)
- Visualizations: grouped/stacked bar chart, single/multi line chart, scatterplots, scatter plot matrix, histogram





# Ambiguity Resolution

## S4. Vis type

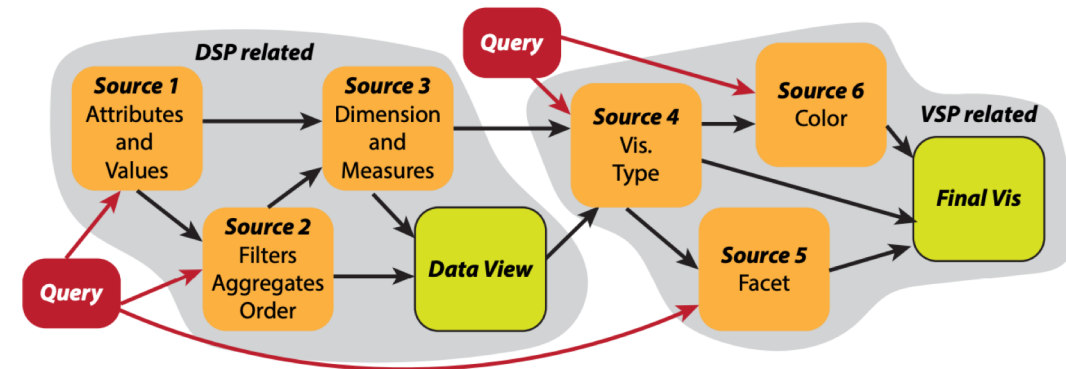
- Based on keywords, infer if the goal is comparison, correlation, distribution analysis and trends → pick vis accordingly

## S5. Faceting

- Datatone uses order of attributes
- Y by X by Z

## S6. Color

- Generally prefer vis with fewer colors



# Ambiguity Resolution History

- History of ambiguity resolution is preserved and tracked
- Used as soft constraints for subsequent interactions
- “Forgets” after a while – weighted combination



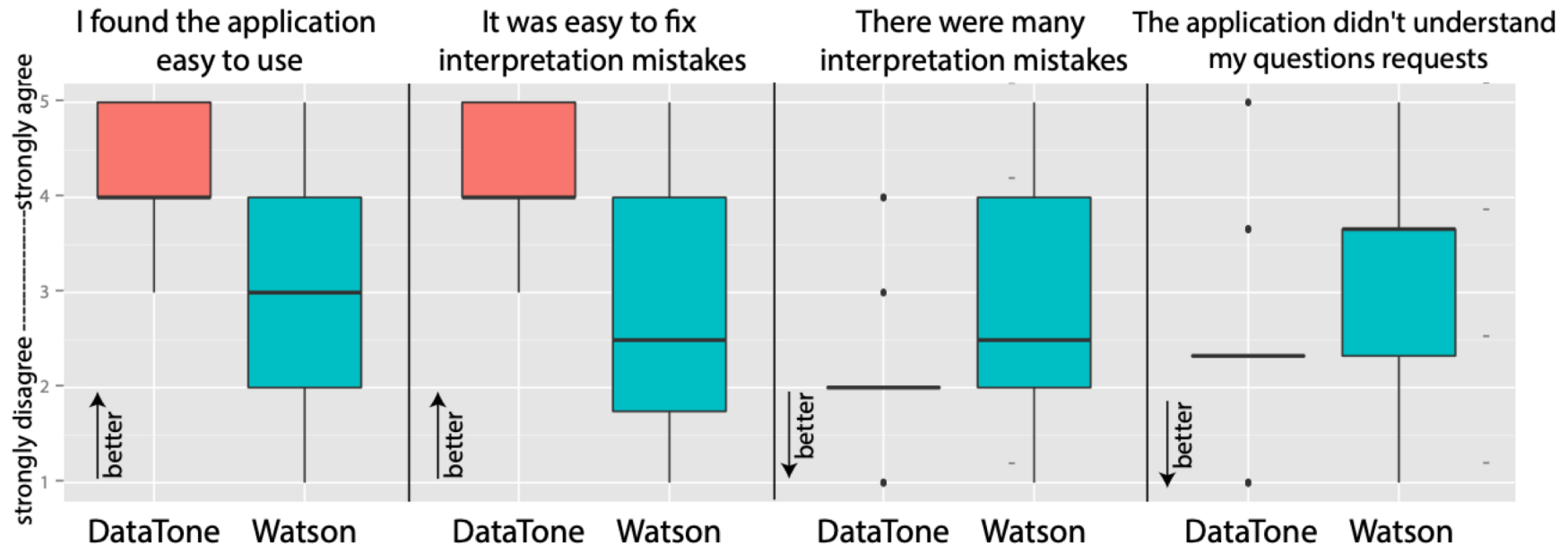
# Evaluation: Jeopardy Style

- Challenge with traditional eval: if you give a NL prompt, people would simply use the same prompt as is to the system
- Instead, eval used “facts” to be proved or disproved
  - E.g., *North Dakota has the fewest number of people without jobs*
- Comparing against Watson Analytics (randomized order)
- 16 participants, 10 facts



# Evaluation Results

More facts with datatone (5.56 vs 2.38) + more correct facts



Easier to use and more flexible (“*It seemed much easier to figure out how to get the visualizations I wanted. It was also much easier to figure out how to fix errors*”)

Widgets helped (“*Very natural interface, and I wasn’t worried about being syntactically accurate since it was easy to correct mistakes.*”)





# Limitations

- No probabilistic approach: heuristics & rules
- Single table
- No memory/context beyond one query at a time
- Limited types of ambiguity widgets (e.g., could resolve at vis itself)



# Takeaways

- Ambiguity resolution is the way to fix human and system issues in specifying and interpreting NL
- A mixed initiative, carefully designed approach aiming to fix data and vis ambiguities can go a long way in helping make NL vis interfaces usable



# Thoughts?

- What did you think of this paper:
  - Interface/Approach?
  - Writing?
  - Evaluation?



# Discussion Q: Apply to LLMs!

- What, out of this paper, would apply to LLMs?



# Discussion Q: Contrast with Wrangler

- How does the mixed Initiative approach here contrast with wrangler? What ways is it better or worse?



# Discussion Q: Conversation

- How would you extend this approach to conversation?



# Discussion Q: Limitations

- What are limitations in the system design? Ways it can be more robust, more useful?



# Issues

- Column names need to be expressed accurately
- Conversation

