

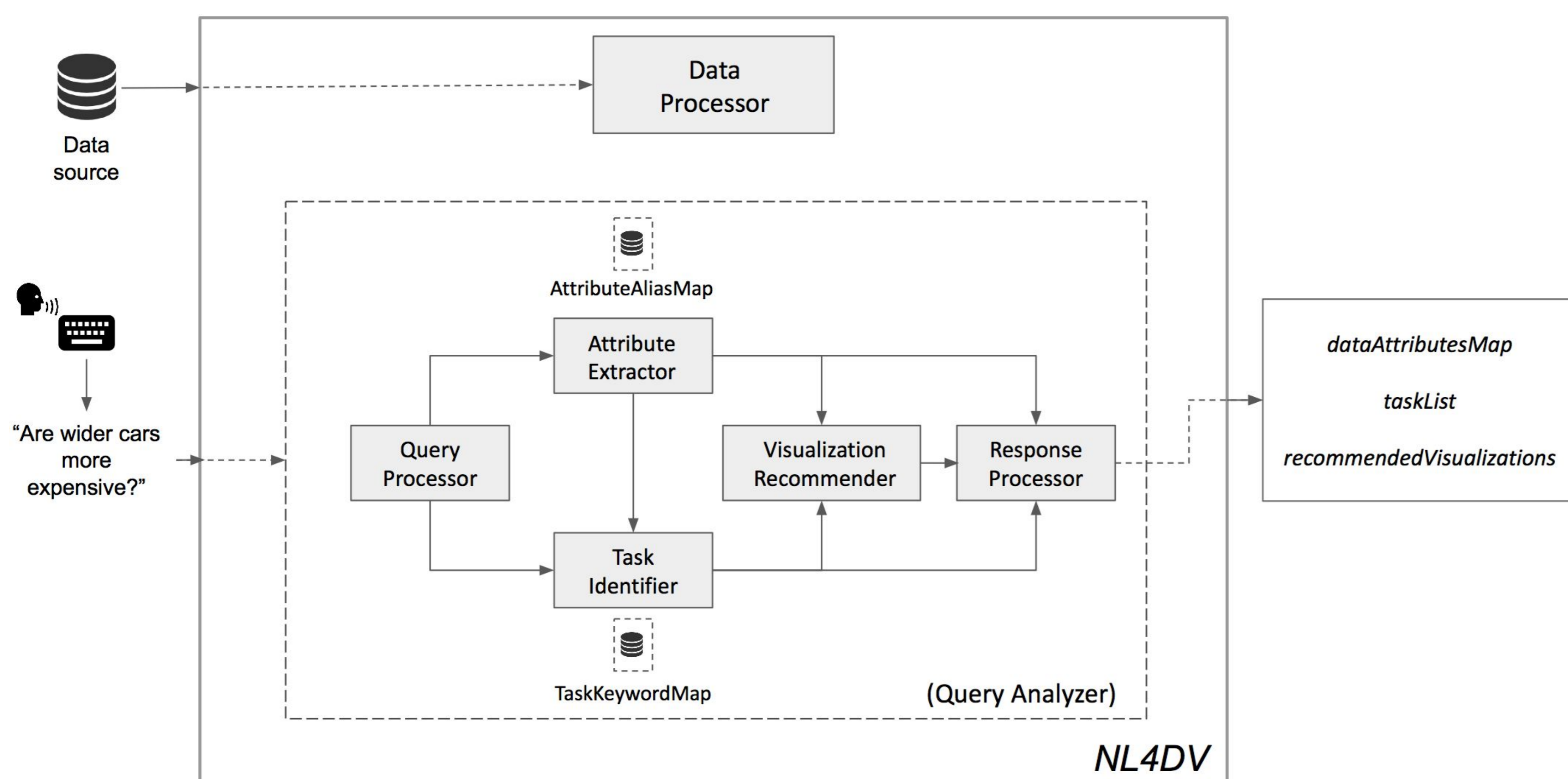
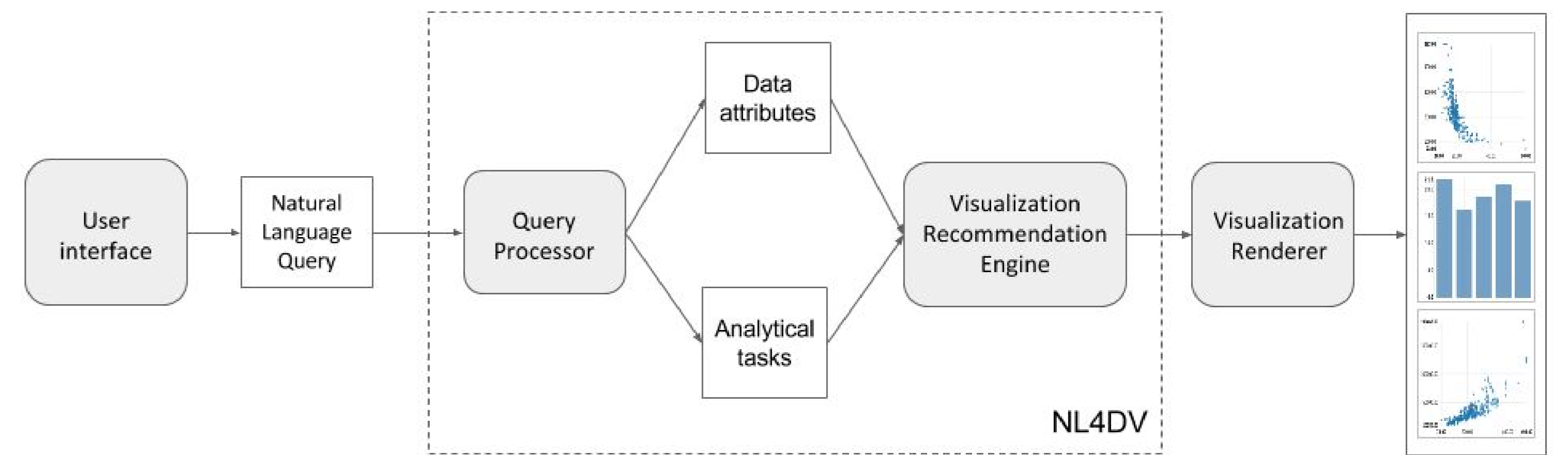
# NL4DV: Toolkit for Natural Language Driven Data Visualization

Arjun Srinivasan      John Stasko

School of Interactive Computing, Georgia Tech

Developing natural language interfaces for visualization systems is a challenging task and requires system developers to spend time and effort on implementing Natural Language Processing (NLP) components necessary to convert natural language queries into visualizations.

NL4DV provides high-level functions developers can use to create natural language-driven data visualization systems. In addition to extracting data attributes and analytical tasks from a query, NL4DV also suggests a ranked list of visualizations that are most relevant to a given query.



## Components:

- **DataProcessor.** Processes the input data source to extract details about individual attributes (data type, domain, related attributes, etc.)
- **QueryProcessor.** Processes the input query to generate POS-tags and n-grams.
- **AttributeExtractor.** Identifies the attributes mentioned in the query and maps these attributes to the identifying substrings in the query.
- **TaskIdentifier.** Identifies analytical tasks a user is trying to perform using a query.
- **VisualizationRecommender.** Uses the data attributes and tasks identified to generate a ranked list of relevant visualizations for the input query.
- **ResponseProcessor.** Merges the output from the other modules to generate the response object.

NL4DV is currently being used by graduate students. Students have used the toolkit to build natural language driven visualization systems from scratch. Additionally, some students have used the toolkit to add an optional natural language query interface to existing systems.

Ongoing work includes. . .

- Experiments to evaluate the toolkit's performance based on its responses.
- Using advanced dependency parsing techniques to better map tasks to attributes in a given query.
- Using Artificial Intelligence Markup Language (AIML) to integrate a conversational agent into the toolkit.

## Usage

### Step 1: Import NL4DV.

```
import nl4dv
```

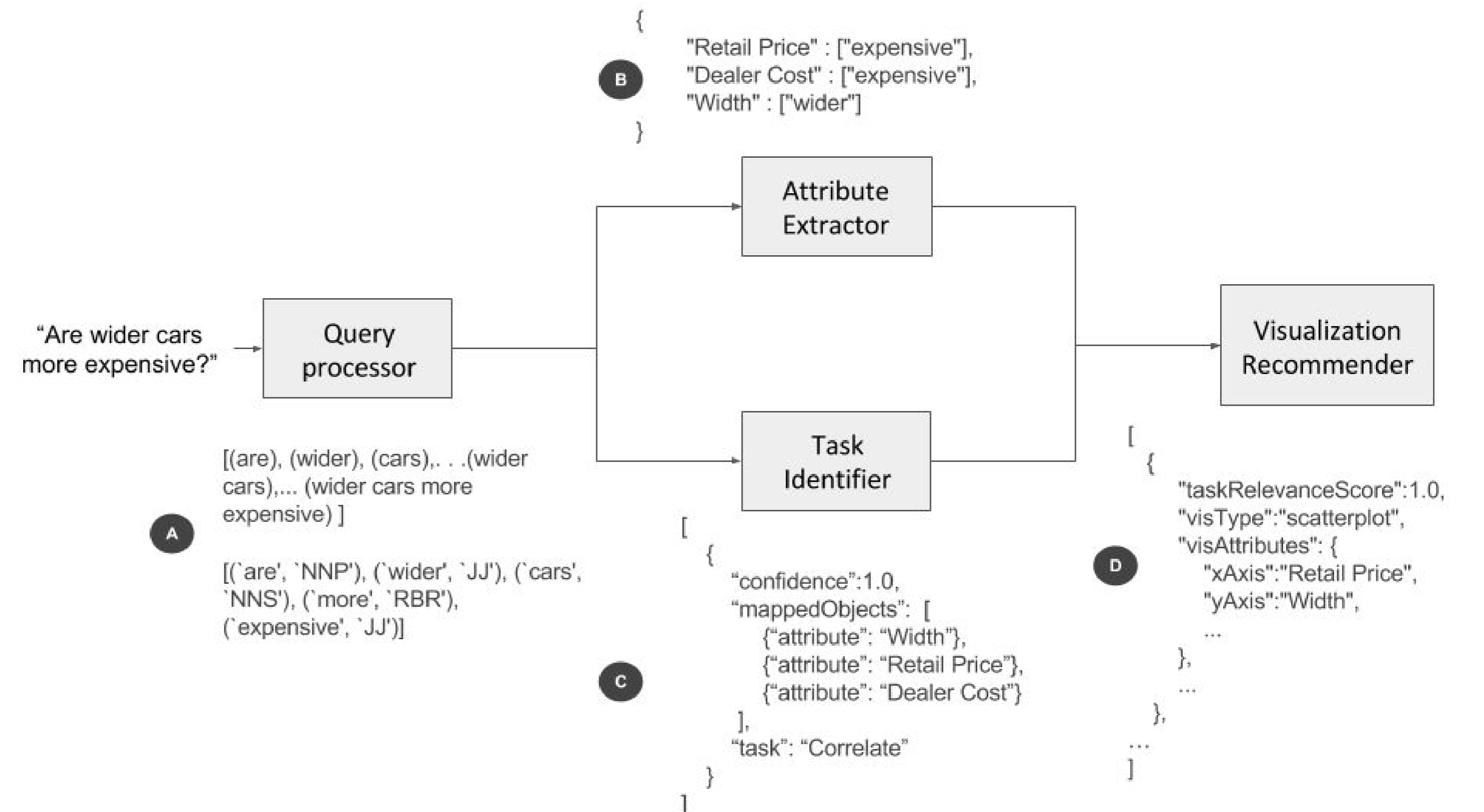
### Step 2: Initialize data source.

```
nl4dv.data(dataFile)
```

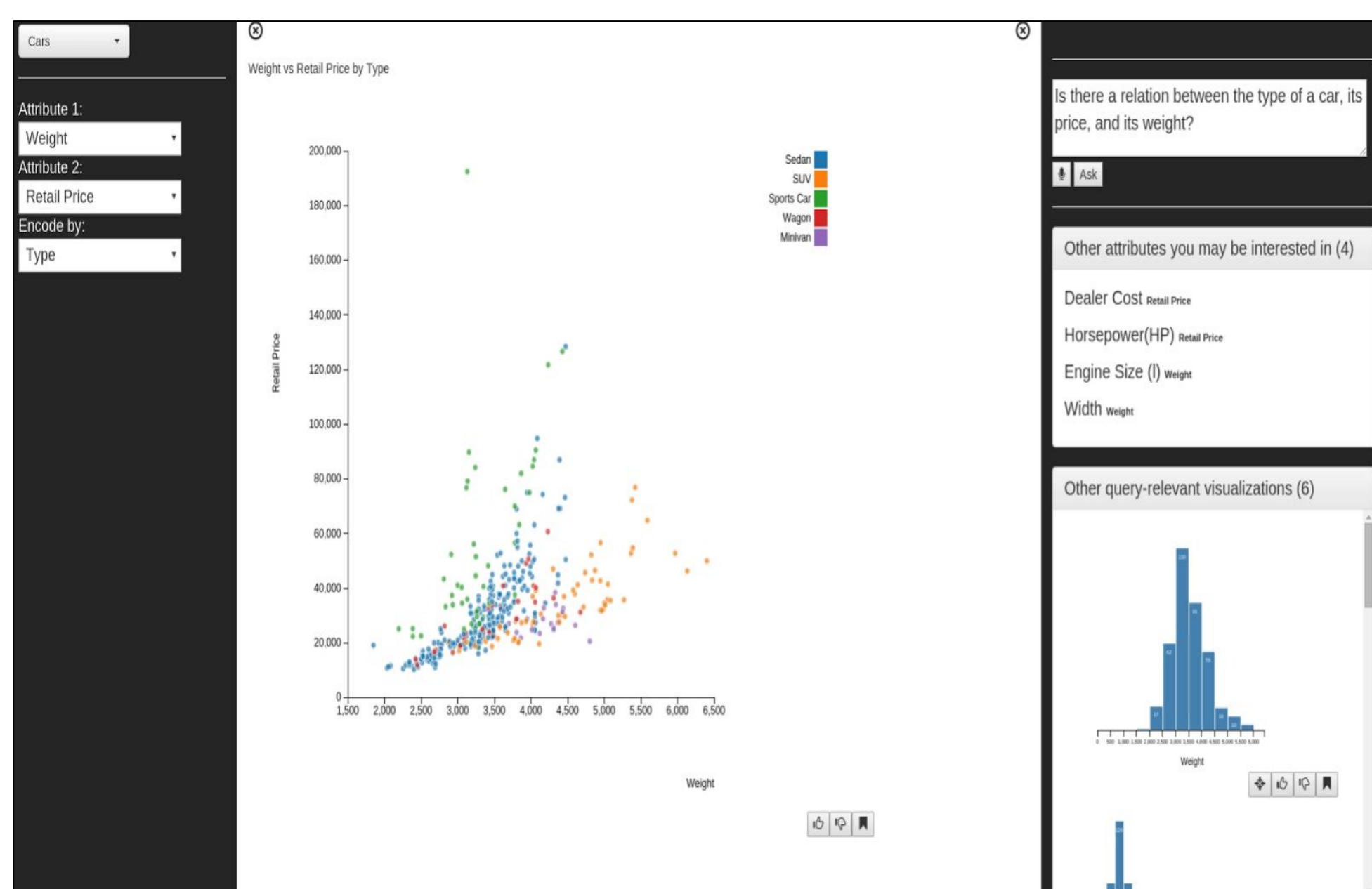
### Step 3: Pass the query as a string.

```
nl4dv.analyzeQuery(query)
```

All responses are in the form of structured JSON objects. These can be parsed easily to update the user interface or render visualizations.

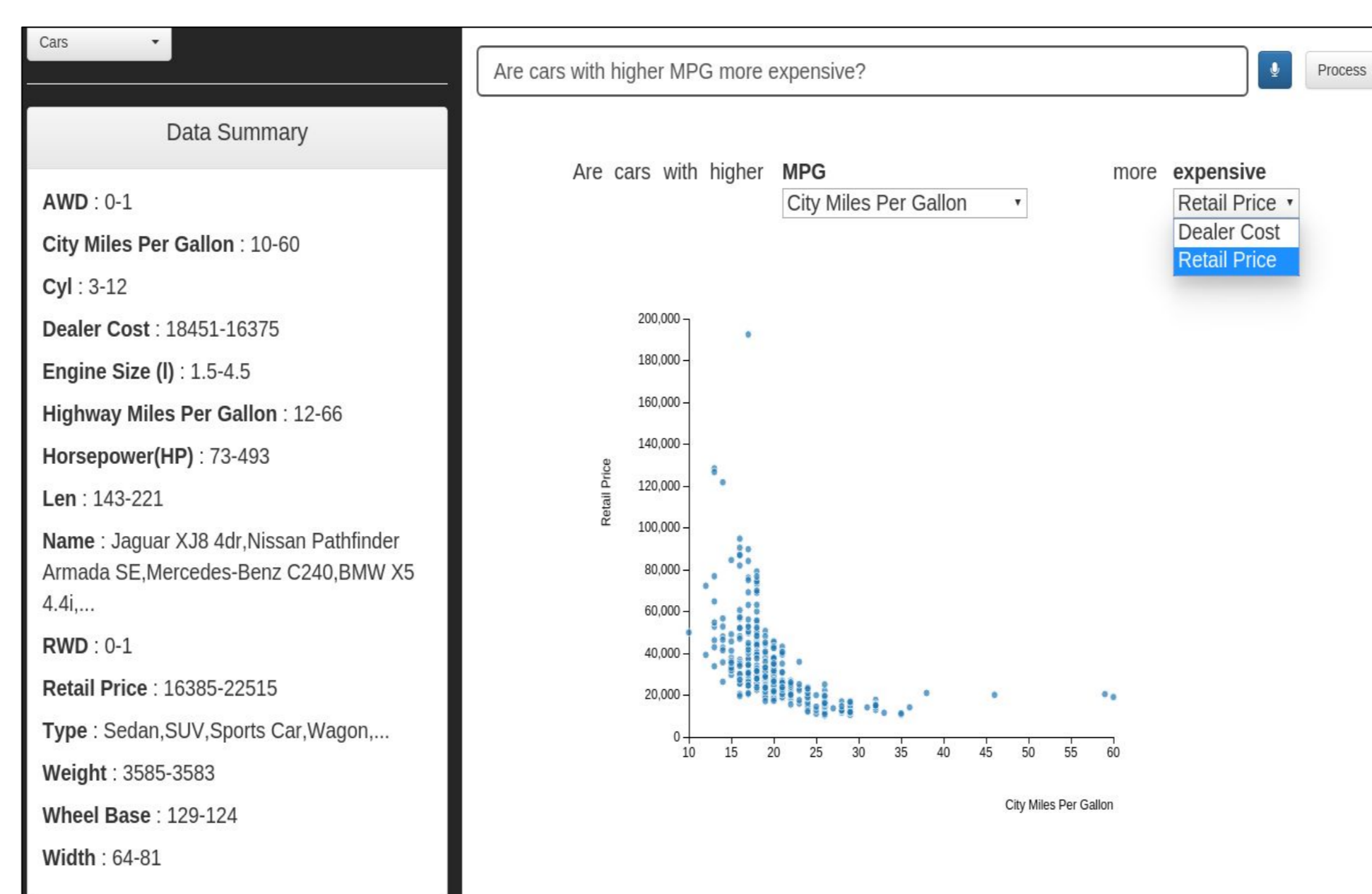


## Sample applications built using NL4DV



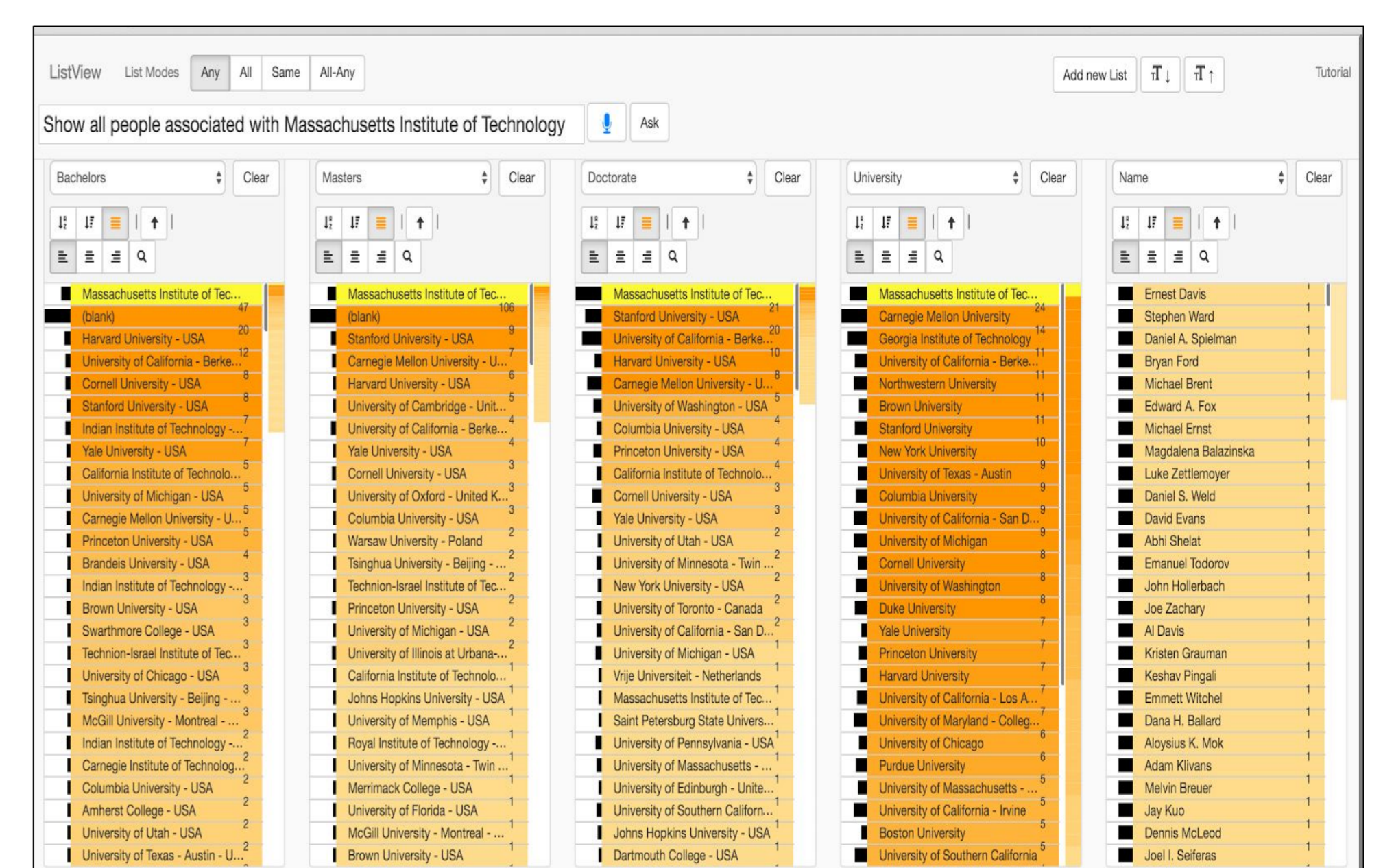
SermoViz

SermoViz is an exploratory data analysis tool which allows users to ask questions to explore a dataset. It uses NL4DV to process the questions, suggest potentially interesting attributes, and select and order visualizations.



SpeechVis

SpeechVis uses NL4DV to present a dataset summary and lets users generate visualizations using natural language. Similar to DataTone by Gao et al, it also highlights ambiguity in a question with the help of dropdowns.



ListView

ListView is using NL4DV to test whether providing an optional natural language query interface can help users in performing tasks faster and more easily compared to the existing graphical user interface.

NL4DV will soon be available as open-source software for developers to use for building and experimenting with natural language interfaces for visualization.