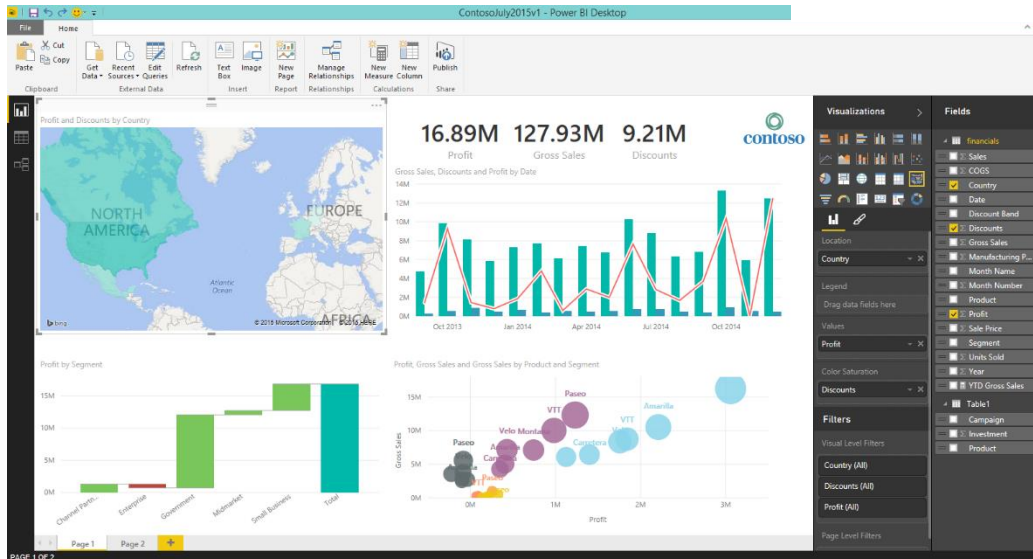
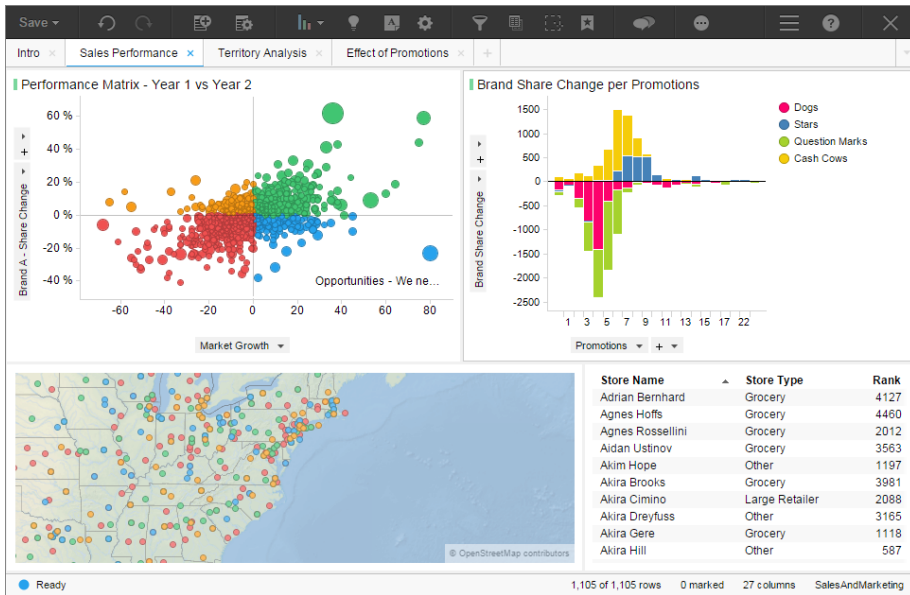


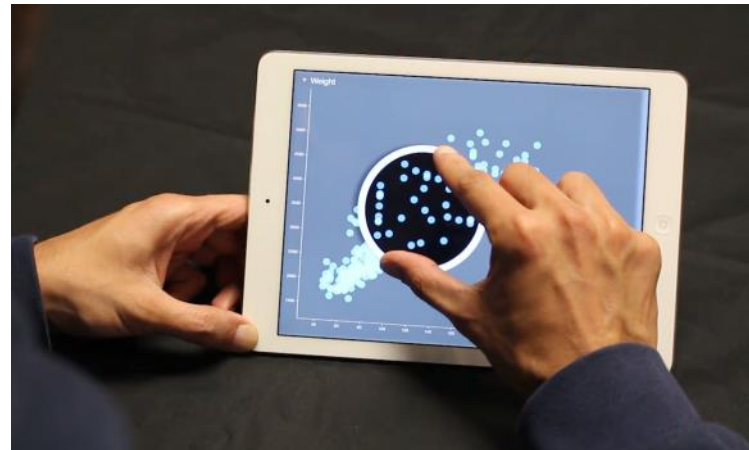
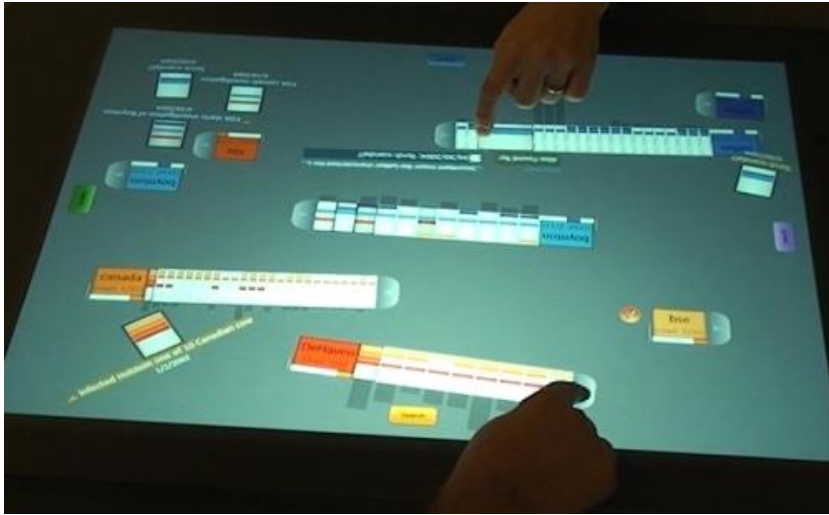
Natural Language Interfaces for Data Analysis with Visualization: Considering What Has and Could Be Asked

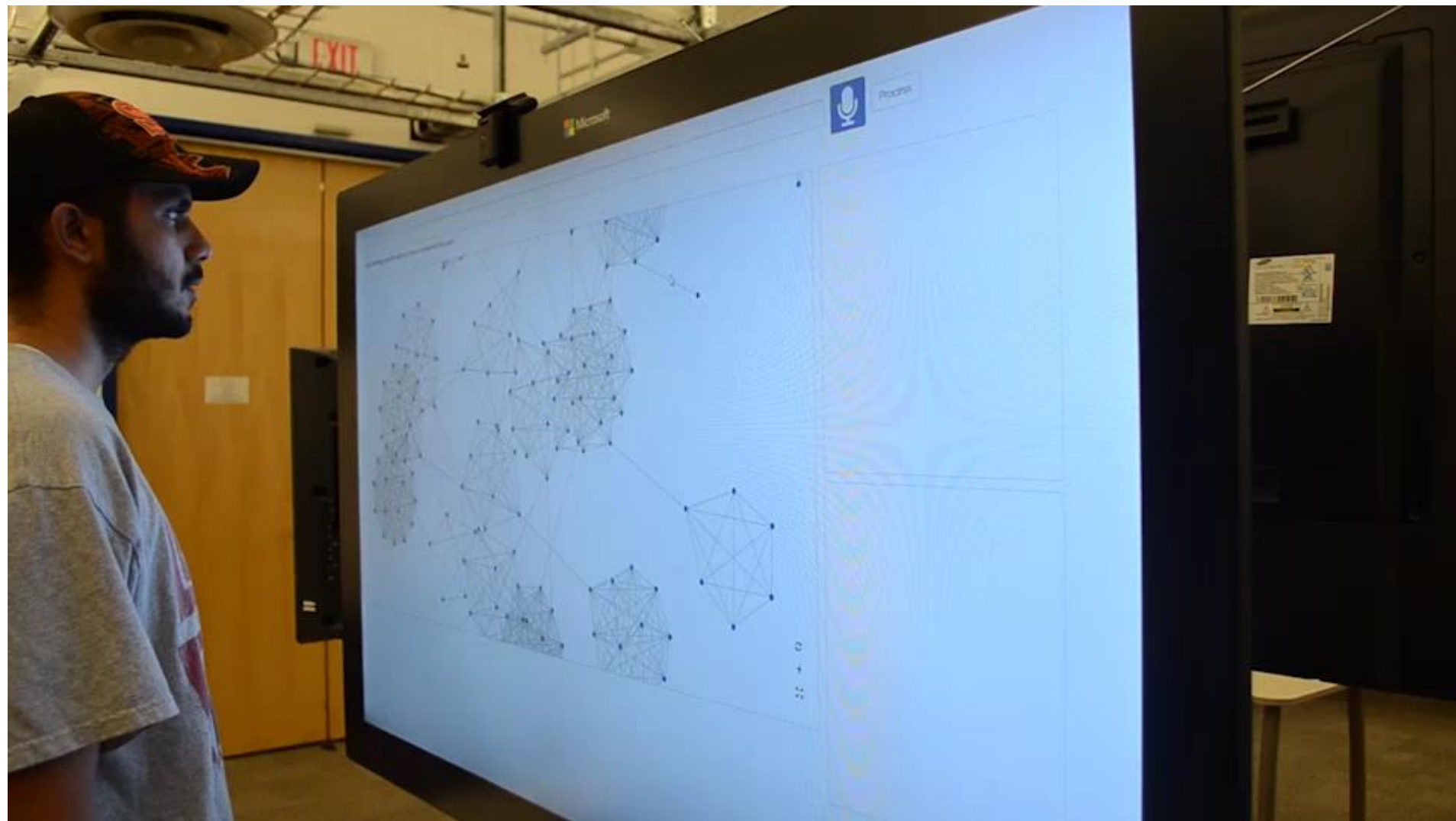
Arjun Srinivasan & John Stasko

School of Interactive Computing
Georgia Institute of Technology







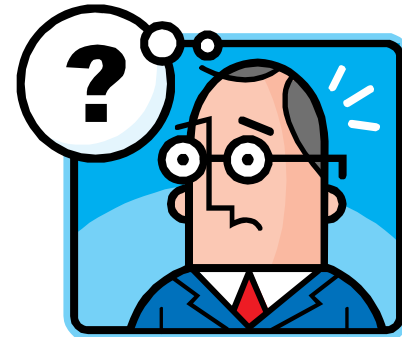


Natural Language Interfaces to Visualization

Help people express what they want to accomplish?

Help remove the need for complicated controls and widgets?

Make interaction more fluid and natural?



Natural Language Interfaces to Visualization

Existing work:

Unnamed	Cox et al	Intl Jrnl Speech Technology '01
Articulate	Sun et al	Intl Symp Smart Graphics '10
DataTone	Gao et al	UIST '15
Eviza	Setlur et al	UIST '16
Articulate2	Kumar et al	Meeting of SIG on Discourse & Dialog '16

Our Analytical Study

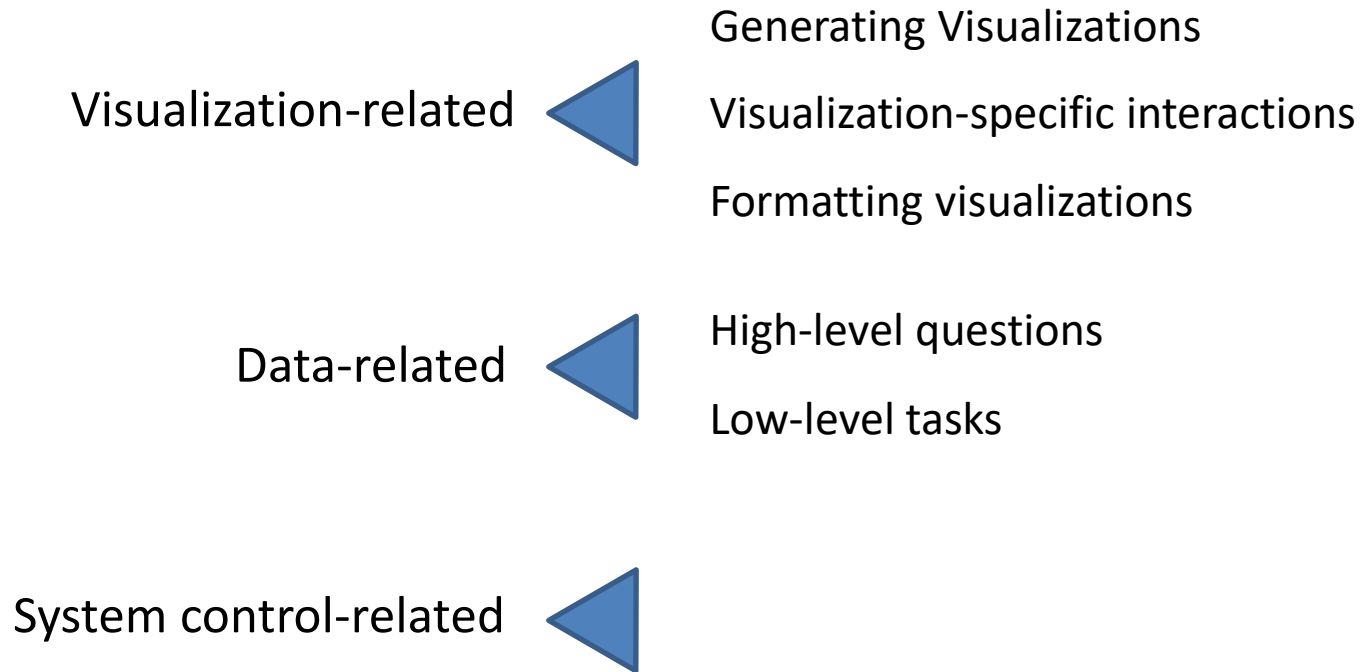
Goals:

1. Characterize what people can (and will) say to a visualization system
2. Situate the five existing research projects in that context

Methodology

1. Take example utterances from existing project papers
2. Consider existing visualization task taxonomies
3. Employ an affinity diagramming approach

Descriptive Task/Utterance Framework



Visualization-related

Add a visualization or modify an existing one

Generating Visualizations

- Add a scatterplot of unemployment versus GDP.
- Show me medals for hockey and skating by country.
- Display a map of crimes in the northern half of the city.

Examples

Visualization-specific interactions

- Sort by average unemployment.
- Select the largest bar in the histogram.
- Show only eastern Europe.
- Zoom in to Spain.

Formatting visualizations

- Change coloring of European countries to blue.
- Show a y-axis label.



Data-related

Asking a question about the data being visualized

High-level questions

Is there a seasonal trend for bike usage?

What is the best time to produce decaffeinated coffee?

Which of these stocks should I buy?

Low-level analytical operations

What is the range of MPG?

Which country had the most outbreaks?

How many transactions occurred before 9am?

System control-related

Controlling the operations of the system and user interface

Move this window to the top right corner.

Can you close the line graph for me?

Help, what can I ask now?

Go back a step.

What other chart can I look at?

Challenges in NLI for Visualization

Input ambiguity

"Show if expensive cars have a higher mileage"

"Show medals for hockey and skating by country"

Output uncertainty

Deciding which visualizations to show or change

What to do when proper reaction is not clear

among others...

Research Opportunities

1. Multimodal interaction in Post-WIMP settings

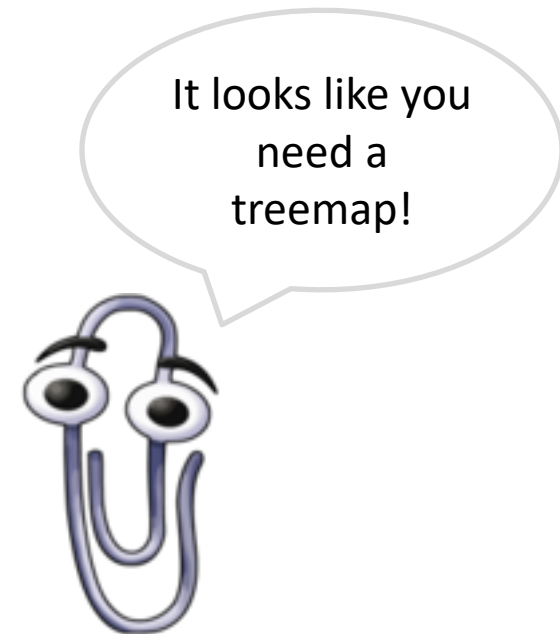
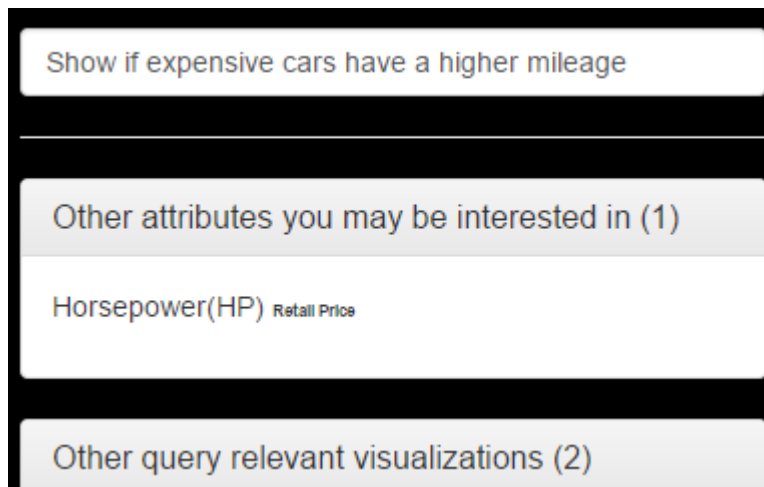
Can combining NLI with touch and gesture provide an even better experience?



Research Opportunities

2. Proactive system behavior

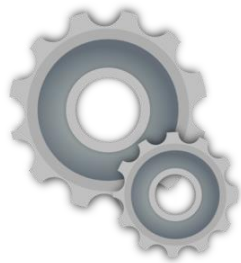
Ranging from the use of context and past queries to influence system behavior all the way up to intelligent virtual assistants



Research Opportunities

3. Mixed-initiative interaction

What can a system provide to allow users to correct and personalize its behavior, and how can it learn from user reactions and interactions?



Are cars with higher **MPG**
City Miles Per Gallon

more **expensive**
Retail Price
Dealer Cost
Retail Price

- division
- level of education
- training level
- work experience
- salary
- sales



Research Opportunities

4. Instruction and feedback

How does the system let the user know what is possible and how does it provide helpful feedback about why something happened or what went wrong?

The figure shows four stages of the Eviza auto-complete interface:

- 1**: The input field is empty. A dropdown menu shows suggestions: `find [where] at [when]`, `show [where] at [when]`, `highlight [where] at [when]`, `find the correlation for [where] at [when]`, `show the correlation for [where] at [when]`, `find [attribute] between [number] and [number]`, `show [attribute] between [number] and [number]`, `highlight the correlation for [where] at [when]`, `highlight [attribute] between [number] and [number]`, and `reset`.
- 2**: The input field contains `show the co`. A dropdown menu shows `highlight the correlation for [where] at [when]`.
- 3**: The input field contains `show the correlation for|`. A dropdown menu shows country names: `Chad`, `Cuba`, `Fiji`, `Iraq`, `Mali`, `Oman`, `Peru`, `Togo`, `Aruba`, and `Benin`.
- 4**: The input field contains `show the correlation for India and China between|`. A dropdown menu shows `complete` and `1800 and 2009`.

Eviza auto-complete

It looks like you're trying to spell "Jamboree". It doesn't have a "U" in it. What are you, British?

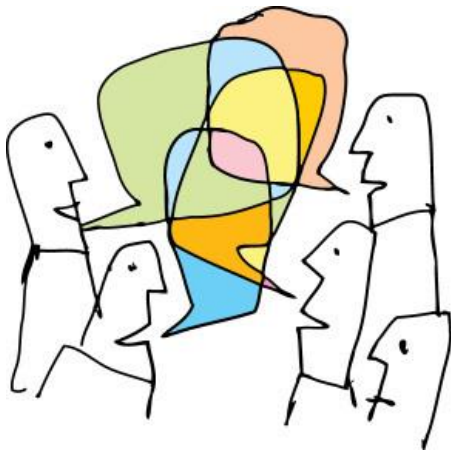
- Leave the spelling to me, idiot.
- That "u" was left in sarcasm.



Research Opportunities

5. Studying the nature of interaction

Understanding how people ask questions may assist NLI design, parsing techniques, architectural & algorithmic decisions, and multimodal integration



Explicit	Find Ronaldo. – Show Pepe’s connections. – Show connections between Pogba and Bale. – Show the shortest path from Evra to Kroos. – Color by position. – Size nodes by betweenness centrality. – What is the clustering coefficient of this network. – Only show German forwards. – Clear all filters. – Resize graph to fit the screen. – Add a filter widget for country. – Change value of the age slider to show players over the age of 30. – Change red nodes to blue.
Follow-up & Contextual	Are any of these players right footed. – Filter by this player’s club. – Show connections of these players. – Do any of these players play for the same club and national team. – Show the different countries players come from. – Ronaldo and Rooney. – Color nodes by country > Now club > How about position?
High-level	How are France and Italy connected. – Players from which countries tend to play more with clubs in the same country. – Which clubs have more left footed players. – Which countries have highest number of common players. – Modify the network layout to focus on England players. – Which three nodes have highest betweenness centralities. – Modify layout to show least edge crossings. – Find clusters.

System Analyses

		Cox et al.	Articulate	DataTone	Eviza	Articulate2
Visualization-related	Generating Visualizations	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
	Visualization specific interactions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	Formatting visualizations				*	
Data-related	Low-level analytical operations		<input type="checkbox"/>		<input checked="" type="checkbox"/>	*
	High-level questions				*	
System control-related		<input type="checkbox"/>				<input type="checkbox"/>

	Cox et al.	Articulate	DataTone	Eviza	Articulate2
Post-WIMP + Multimodal					*
Proactive vs Reactive	Reactive	Reactive	Minimally proactive	Moderately proactive	Reactive*
Instruction & Feedback	<input type="checkbox"/>		<input type="checkbox"/>	<input checked="" type="checkbox"/>	*
Mixed-initiative Interaction			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

(Details in paper)

Conclusion

Natural language interfaces to visualization an exciting new area

Relatively little existing work

Challenges

- Ambiguity

- Multimodal integration

- System capability and feedback

- ...

Many future research opportunities

Acknowledgments

- Supported by NSF IIS-1320537



- Supported by gifts from Adobe, Google, Microsoft, and Nokia

Natural Language Interfaces for Data Analysis with Visualization: Considering What Has and Could Be Asked

Arjun Srinivasan & John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu

