Articulate2: Toward a Conversational Interface for Visual Data Exploration

1. Electronic Visualization Laboratory

University of Illinois at Chicago www.evl.uic.edu

University of Illinois at Chicago

2. Natural Language Processing **Laboratory (nlp)**

3. Laboratory for Advanced Visualization and Applications (lava)

> University of Hawaii at Manoa lava.manoa.hawaii.edu

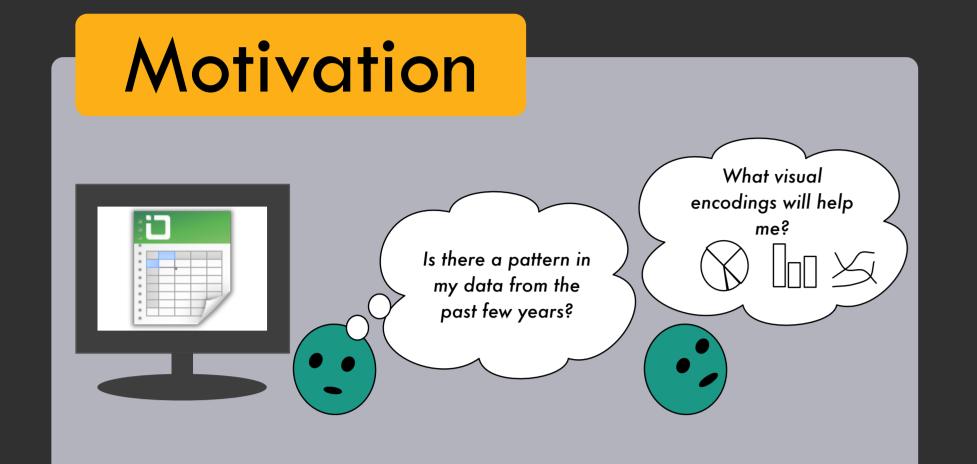
Jillian Aurisanoevi* Abhinav Kumarnıp Alberto Gonzaleziava, Jason Leighiava Barbara DiEugenionip Andrew Johnsonevi

*jauris2@uic.edu

voice

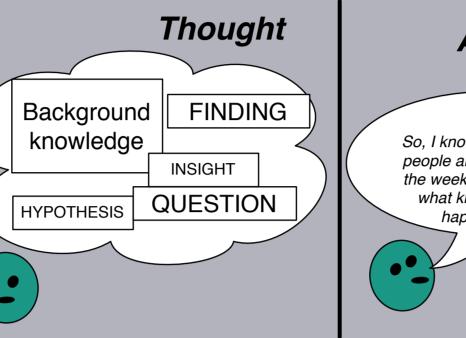
interface

www.nlp.cs.uic.edu



InfoVis novices' struggle with visualization construction. Even with the aid of visualization software, such users may face challenges when translating their questions into appropriate visual encodings, or interactively refining the representation to achieve a desired result.

Conversational interface



Response Action the weekends. I wonder > Does this answer your

A 'conversational interface' which maintains a dialog with the user through natural language and gestures, could allow users to engage in repeated cycles of visualization generation and modification, asking questions directly through speech.

Articulate 2 multi-window web-based environment (sage2) system communication console

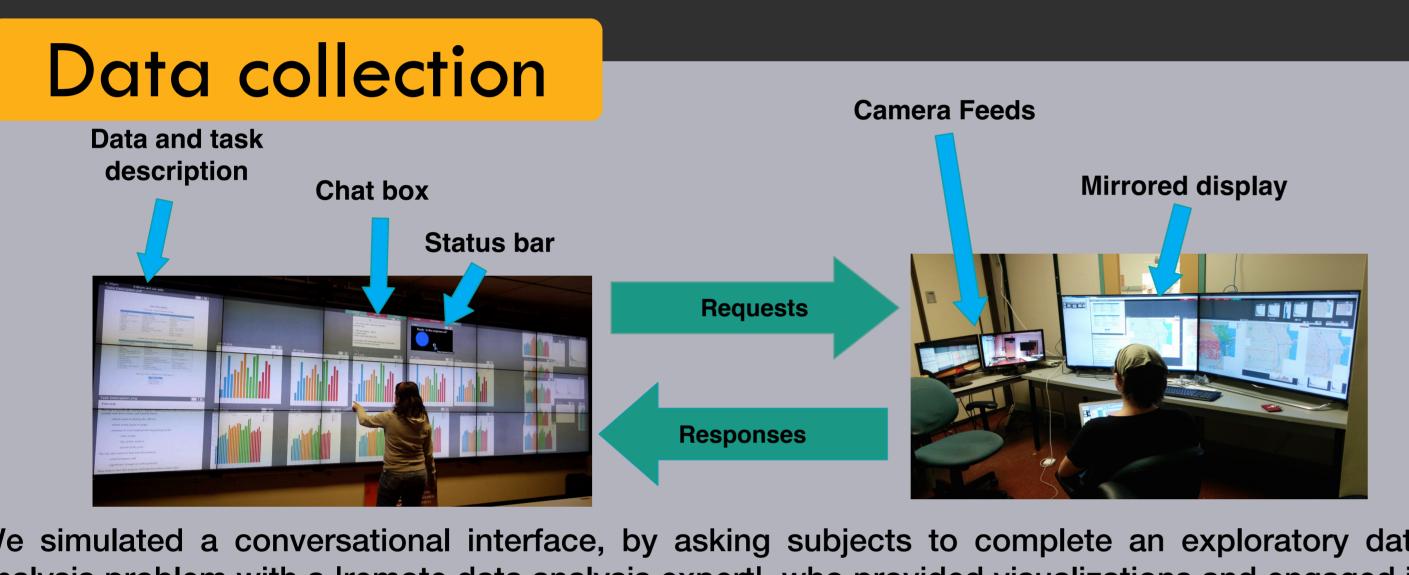
In this poster we present a prototype conversational visual data analysis system. Our prototype was developed from a corpus consisting in 15subjects engaging in exploratory data visualization with a simulated conversational interface. It features 1) speech to visualization pipeline, 2) classification system to divide utterances into major types, 3) history manager and knowledge-base.

During what time

is the crime rate

maximum, during

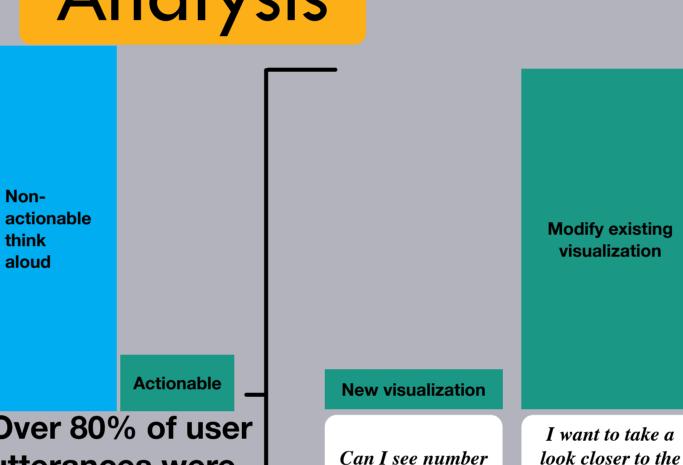
the day or the



We simulated a conversational interface, by asking subjects to complete an exploratory data analysis problem with a 'remote data analysis expert', who provided visualizations and engaged in a dialog through a chat box.

Analysis

question?



visualizations

Over 80% of user utterances were 'non-actionable' think-aloud

Can I see number metro right here, a of crimes by day of little bit eastward of the week? Greektown.

Users most often refined existing visualizations, or asked for new visualizations

graphs as I won't

be needing it

anymore

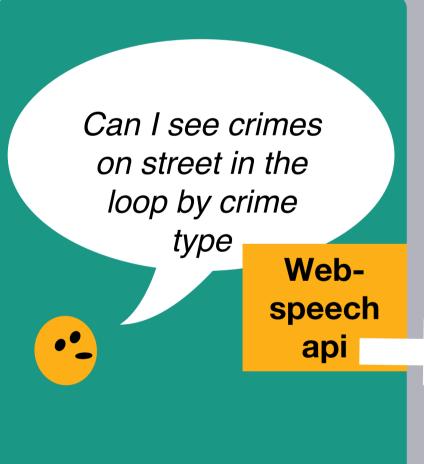
using an existing visualization as a template.

clarification **Express preference** If you want you Okay, so is this can close these

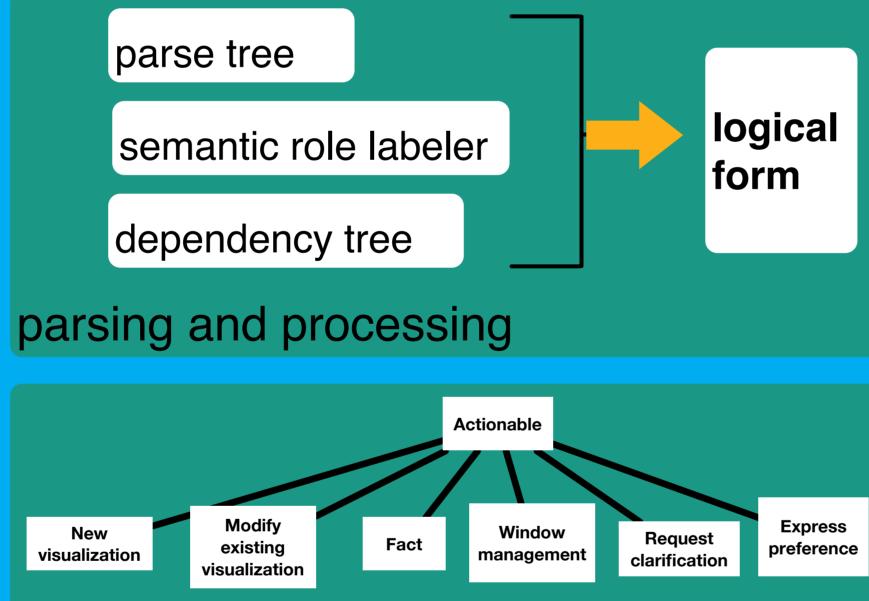
a particular year?

The first graph is a better way to statistics from all 5 visualize rather years? Or is this for than these four separately

Architecture



Natural language to vis pipeline



question type classifier Month of the Year Time of Day e.g. Loop, River North, iversity of Illinois at Chicago, Near W knowledge base

1. process

history manager

SELECT count(*) as TOTAL_CRIME, crimetype FROM chicagocrime WHERE 'neighborhood'='loop') AND (`location type`='street') GROUP BY crimetype {"id":0, "horizontalAxis":"NON_UNIT","horizontal

GroupAxis":"crime type","verticalAxis":"TOTAL_CRIME","plo tType":"BAR", ", "dataQueryResult": ["January": 1748....

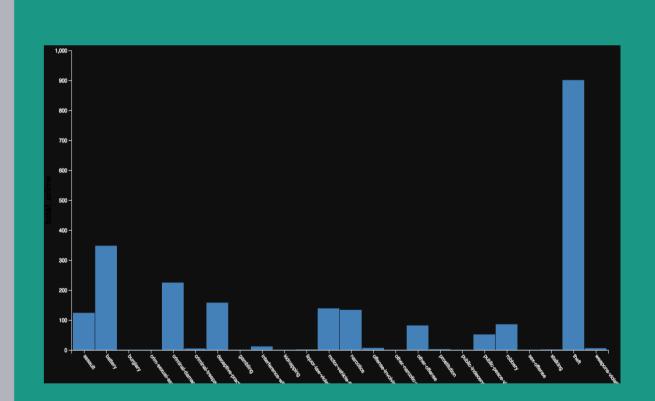
sql and visualization specification generation



specification

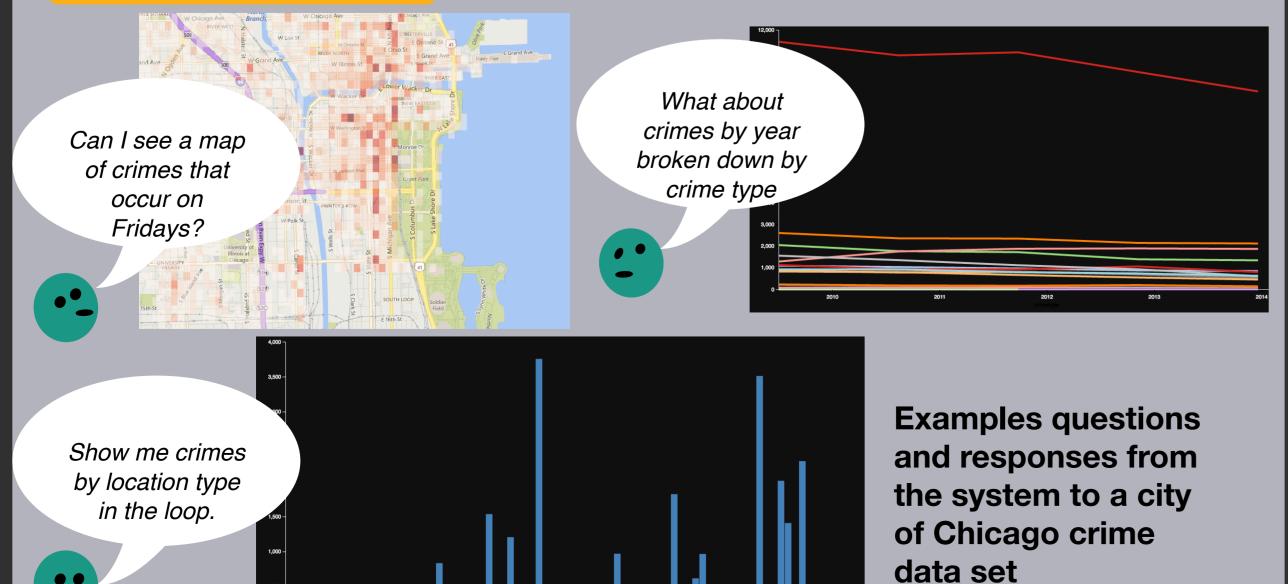
2. specify

visualization executer:



Visualizations were executed using vega.js, which leverages principals from grammar of graphics. Results are displayed in a multi-windowed, web-based environment.

Examples



Question classification accuracy

Classifying questions into major types (new visualization, modify visualization, fact, window management, request clarification, express preference)

Classifier Model	Accuracy
Support Vector Machines	87.66
Random Forest	85.6
Multinomial Naïve Bayes	85.6
Naïve Bayes	74.28

Future work

Future work with focus on:

- Resolving references to on-screen visualizations and objects through speech and gesture
- "Can I see this plot <points> but with data for the Loop?"
- Enriched visualization history manager, to provide context to new questions
- Responses to user questions about points of fact and user expressions of preference
- Expanded visualization modification operation

This work is supported by NSF awards 1445751 and 1445796.