

# Enhancing Natural Language-Based Data Exploration with Analysis Pipeline Illustration

Yi Guo, Nan Cao, Xiaoyu Qi, Haoyang Li, Danqing Shi, Jing Zhang, Qing Chen, Daniel Weiskopf

Intelligent Big Data Visualization Lab, Tongji University, Shanghai, China Renmin University of China, Beijing, China  
University of Stuttgart, Stuttgart, Germany Finnish Center for Artificial Intelligence, Helsinki, Finland

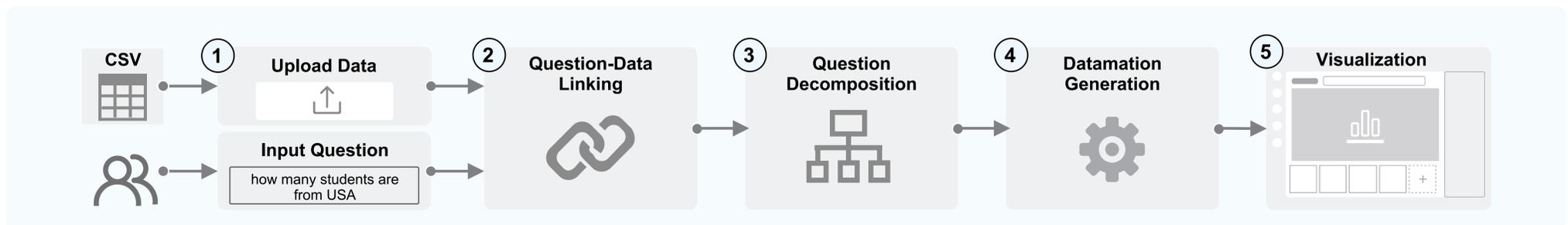


## Introduction

Existing natural language interfaces (NLIs) for data exploration focus on providing answers to questions, with few offering explanations or presentations of the data analysis pipeline used to uncover the answer. Such presentations are important in exploratory data analysis (EDA), improving interpretability, reliability, and aiding users in understanding the analysis process and gaining insights.

We introduce Urania, an NLI that visualizes data analysis pipelines for answering input questions. It utilizes a novel language model to generate an analysis pipeline that solves a user's question, and then presents the pipeline and the answer as a datamation, with animated operations and data changes.

## Workflow

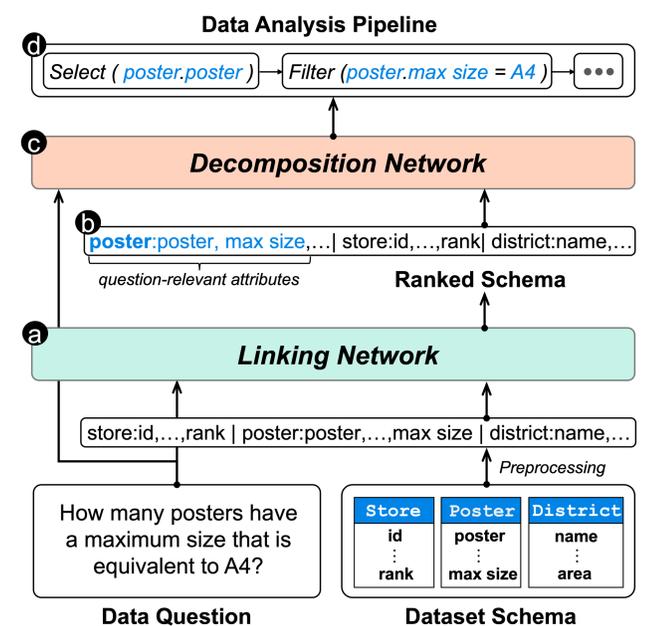


## Urania

<https://github.com/idvxlabs/urania>

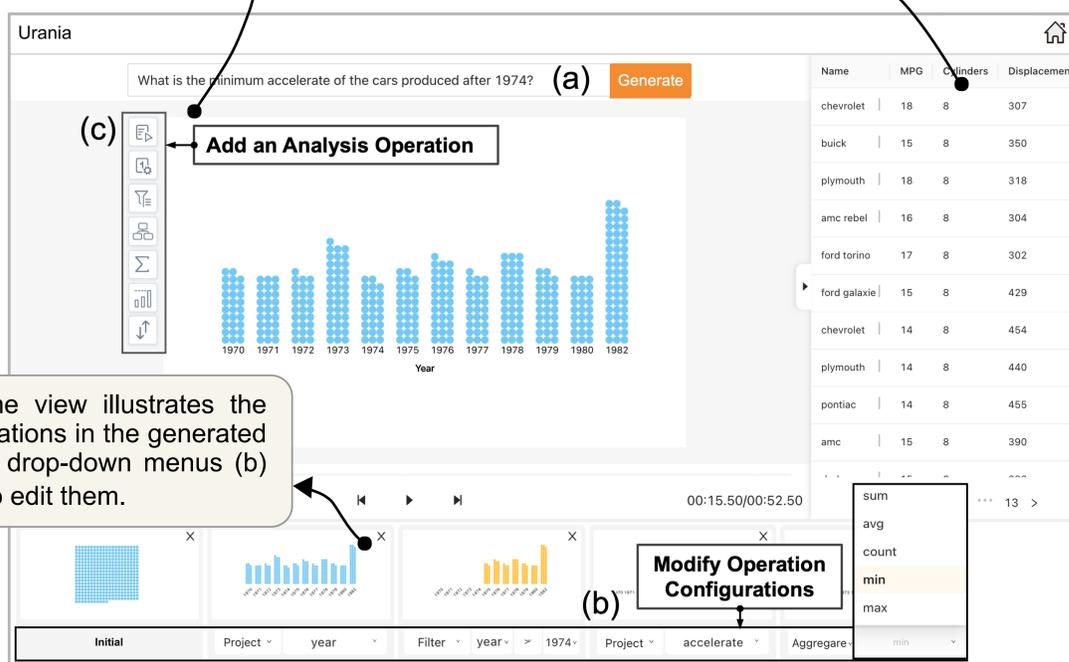
## Algorithm

- We introduce a data-aware question decomposition algorithm that is able to incorporate the data-question links in the question decomposition process to generate more accurate data analysis pipelines.



The datamation view visualizes the data analysis pipeline as a datamation. Users can use shortcut buttons (c) to add new operators to data analysis pipelines.

The data view is designed to illustrate the uploaded data, users can preview the data and enter a question of interest into the input box (a).



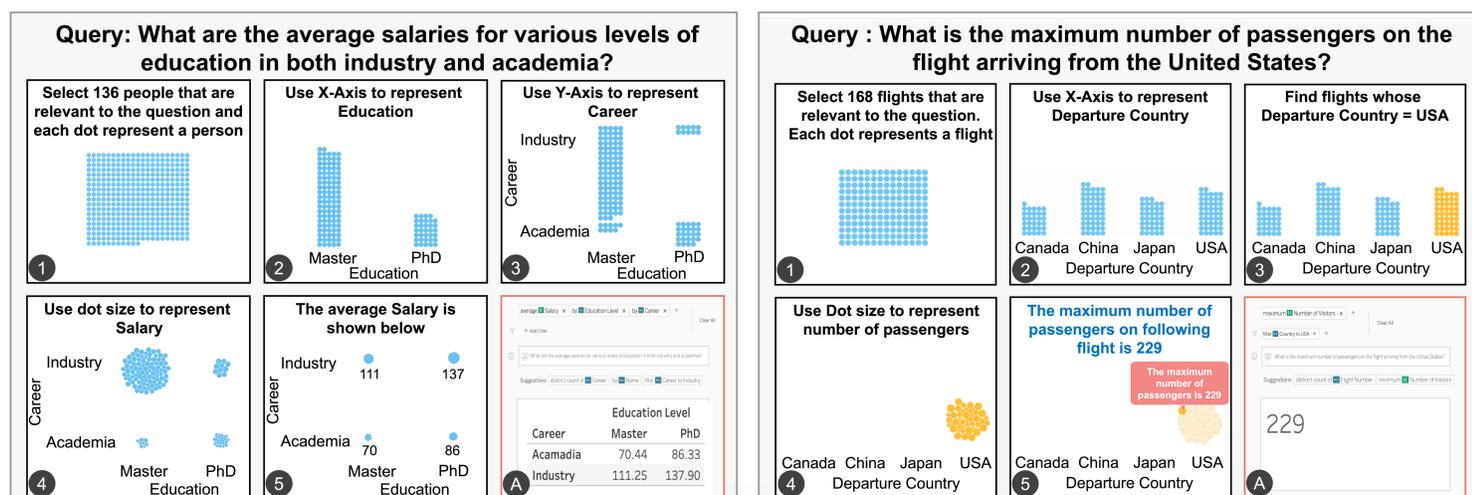
The key-frame view illustrates the analysis operations in the generated pipeline. The drop-down menus (b) allow users to edit them.

## Evaluation

- We conducted interviews with three domain experts. They are asked to explore a dataset with two NLI systems, Urania and Tableau Ask Data.

Experts agree that:

- Urania offers an intuitive and reliable way to perform EDA, with verification and protection against false discoveries.
- Its interactive features streamline exploration, and personalized sessions make users feel in control.
- Urania strikes a balance between manual and automatic exploration.



Example of Urania (1-5) and Ask Data (A) responded to experts' questions