Towards Natural Interactions with Data

Haotian Li
Joint work with Nithin Chalapathi, Alvin Cheung, Aditya Parameswaran

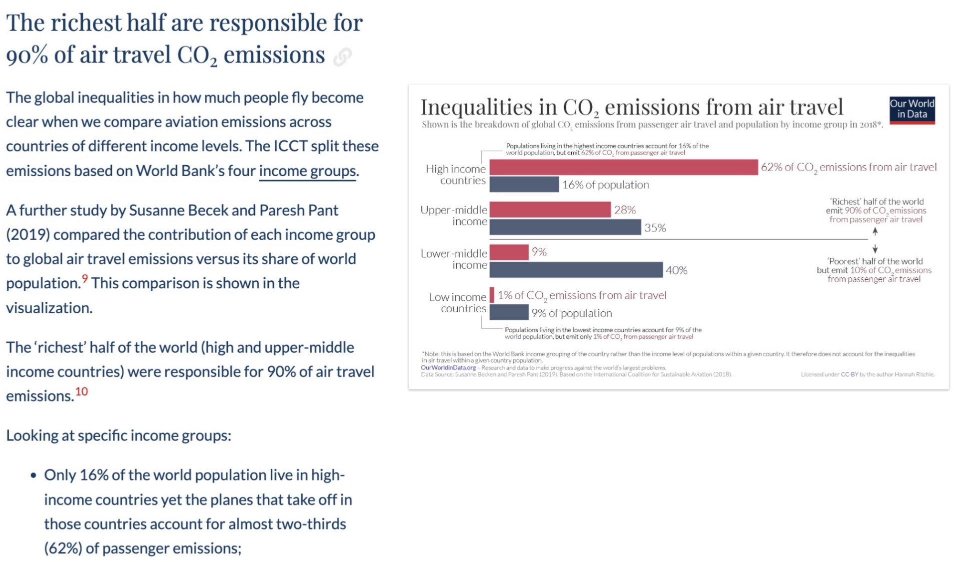
EPIC Retreat
2023.4.17

Berkeley
UNIVERSITY OF CALIFORNIA

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Interacting with Data

- More and more users with non-computer science background have been involved in data work as well, such as journalists and business analysts.



The richest half are responsible for 90% of air travel CO_2 emissions

The global inequalities in how much people fly become clear when we compare aviation emissions across countries of different income levels. The ICCT split these emissions based on World Bank's four income groups.

A further study by Susanne Becek and Paresh Pant (2019) compared the contribution of each income group to global air travel emissions versus its share of world population.[9] This comparison is shown in the visualization.

The 'richest' half of the world (high and upper-middle income countries) were responsible for 90% of air travel emissions.[10]

Looking at specific income groups:

- Only 16% of the world population live in high-income countries yet the planes that take off in those countries account for almost two-thirds (62%) of passenger emissions;

Data article by Our World in Data[1]

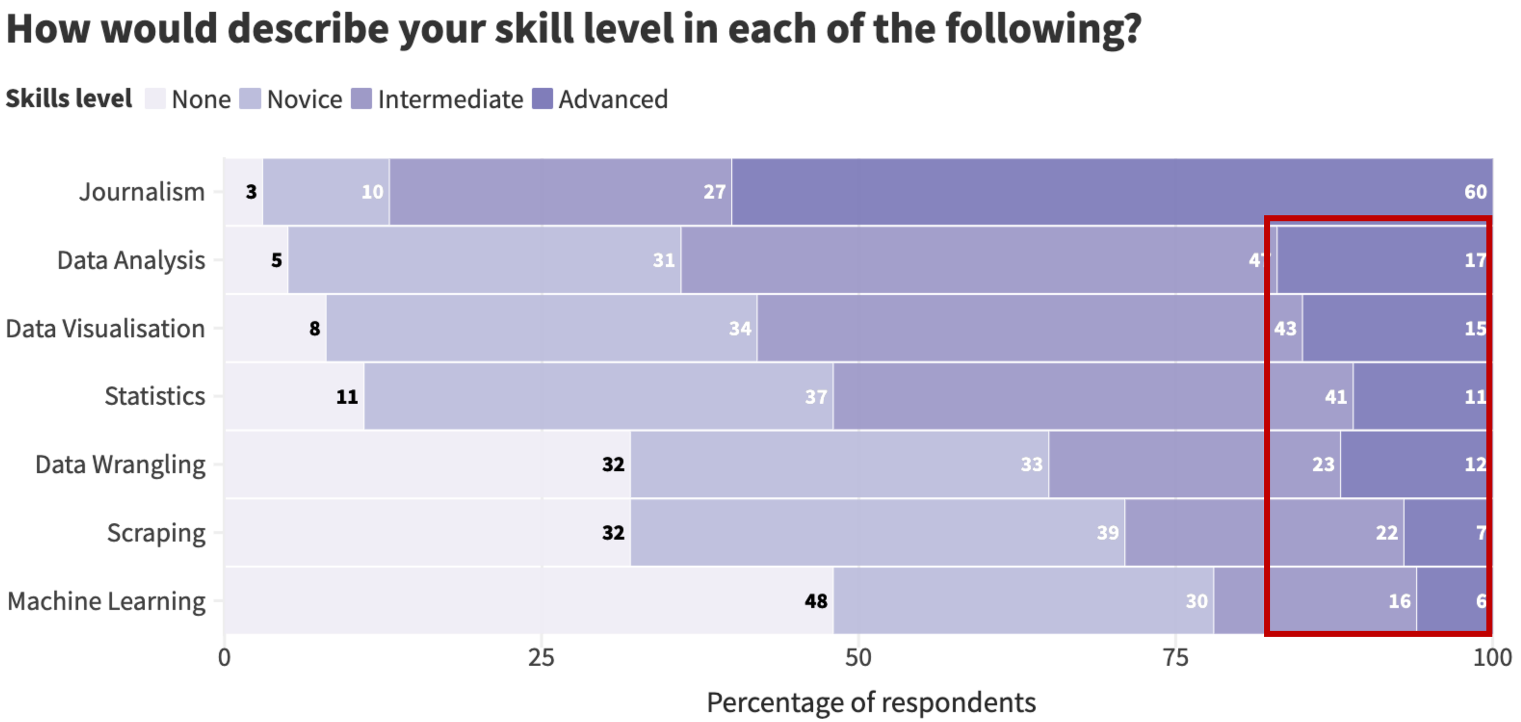
Data video by Vox[2]

[1] https://ourworldindata.org/transport#the-richest-half-are-responsible-for-90-of-air-travel-co2-emissions
[2] https://youtu.be/sv0dQfRRrEQ

Interacting with Data Can be Challenging!

- How do they feel about data analysis?

- Let's take journalists as an example and have a close look at a survey[3] with more than 1,700 data journalists.



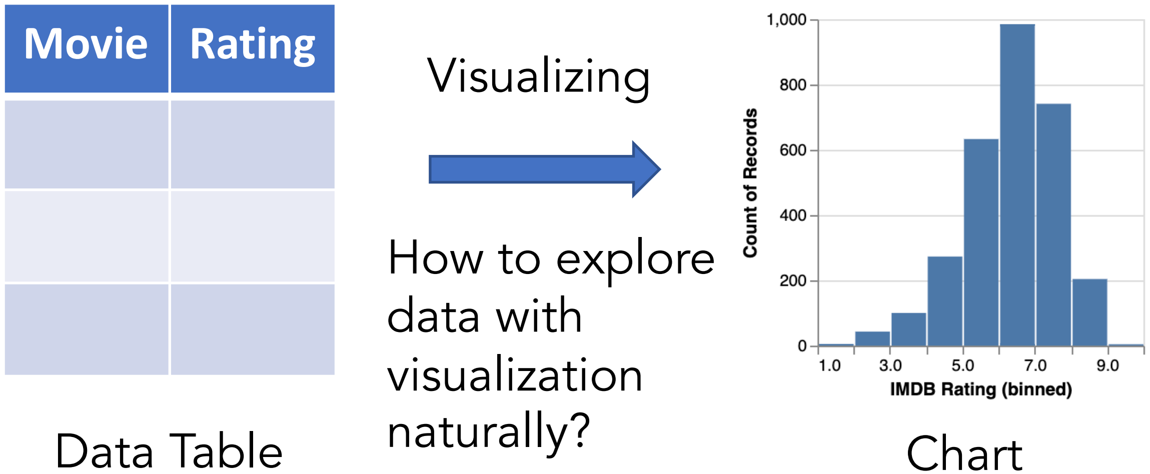
- According to the results, data analysis can be challenging to them.

[3] https://datajournalism.com/survey/2022/skills-and-tools/

Towards Natural Interactions with Data

- To make data analysis more accessible to these data novices, one promising direction is to equip them with <u>natural approaches of interaction with data</u>.

- In this talk, we would like to share our study about exploratory visual analysis towards this direction.

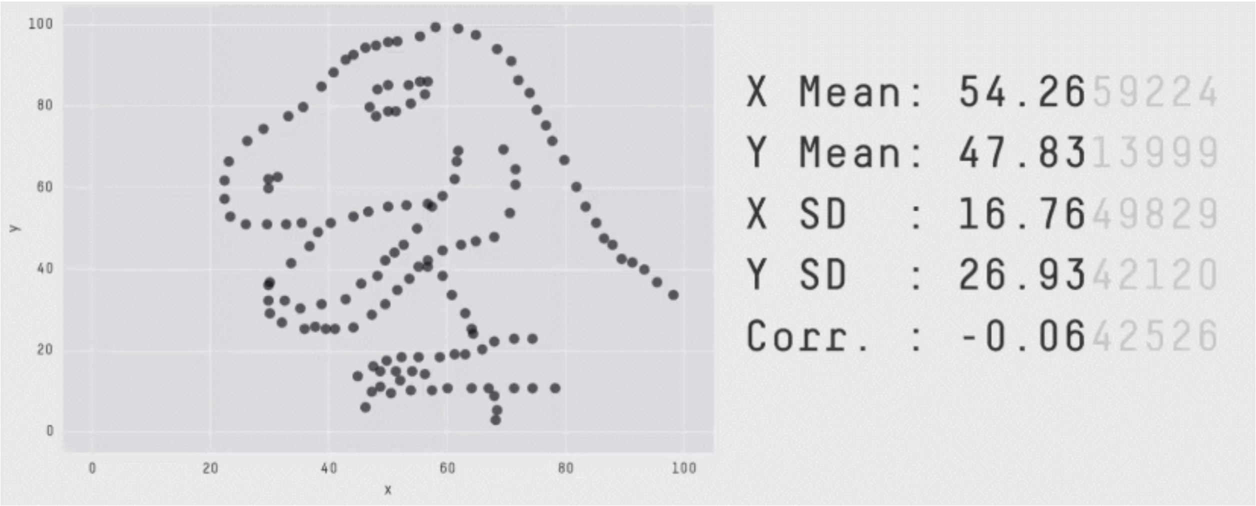


Data Table

Chart

Exploratory Visual Analysis

- Visualizations present data graphically with details.
- Compared to statistical methods, they better facilitates exploring structured data.

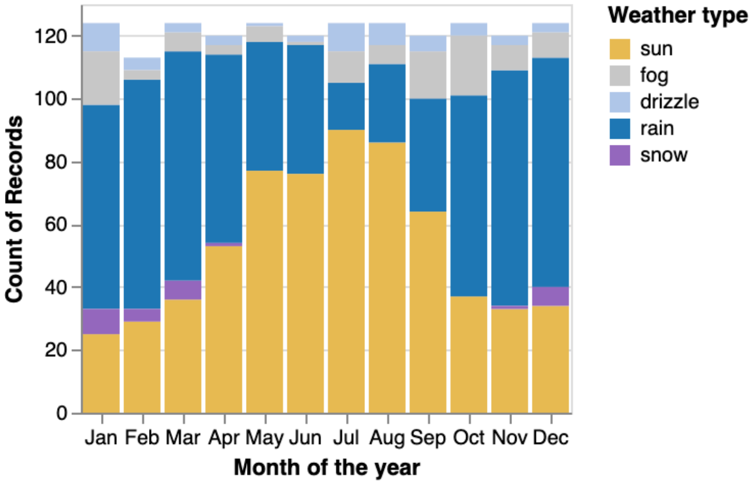


4

[4] Matejka, Justin, and George Fitzmaurice. "Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing." CHI '17. 2017.

Exploratory Visual Analysis

- However, it can be challenging to author visualizations from scratch.



Stacked Bar Chart

```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "data": {"url": "data/seattle-weather.csv"},
  "mark": "bar",
  "encoding": {
    "x": {
      "timeUnit": "month",
      "field": "date",
      "type": "ordinal",
      "title": "Month of the year"
    },
    "y": {
      "aggregate": "count",
      "type": "quantitative"
    },
    "color": {
      "field": "weather",
      "type": "nominal",
      "scale": {
        "domain": ["sun", "fog", "drizzle", "rain", "snow"],
        "range": ["#e7ba52", "#c7c7c7", "#aec7e8", "#1f77b4", "#9467bd"]
      },
      "title": "Weather type"
    }
  }
}
```

Vega-Lite[5]

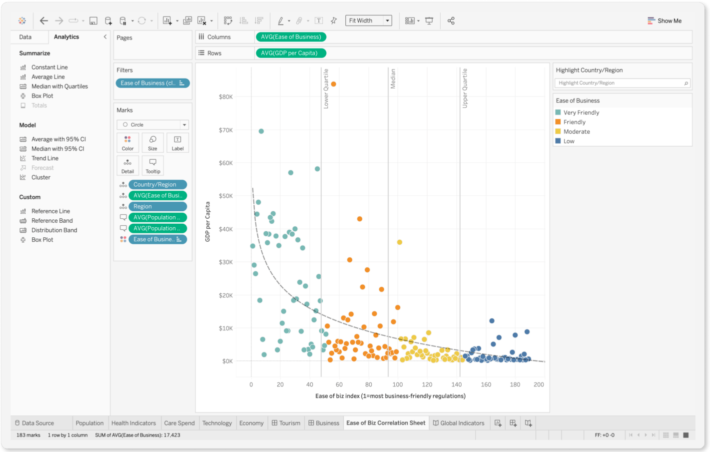
```
fig, ax = plt.subplots()
bottom = np.zeros(3)

for boolean, weight_count in weight_counts.items():
    p = ax.bar(species, weight_count, width, label=boolean, bottom=bottom)
    bottom += weight_count

ax.set_title("Number of penguins with above average body mass")
ax.legend(loc="upper right")

plt.show()
```

Matplotlib[6]

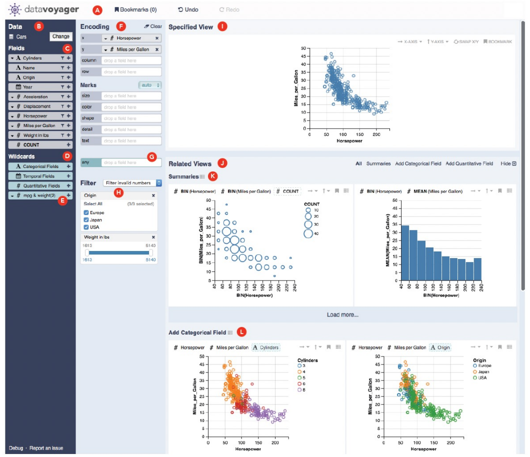
[5] https://vega.github.io/vega-lite/examples/stacked_bar_weather.html
[6] https://matplotlib.org/stable/gallery/lines_bars_and_markers/bar_stacked.html#sphx-glr-gallery-lines-bars-and-markers-bar-stacked-py

Exploratory Visual Analysis

- Multiple methods have been proposed to address the challenge, such as
  - leveraging interactive graphical user interfaces (GUI);
  - recommending visualizations based on users' intent.



Tableau[7]



Voyager[8]

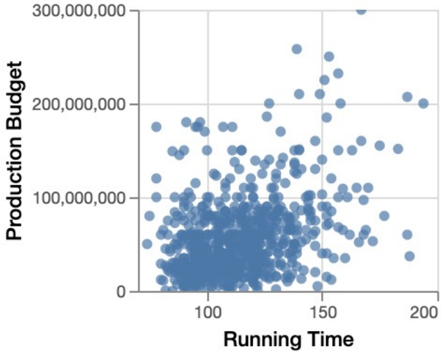
- According to the survey with journalists, the proportion of using Tableau for data analysis or visualization is the same as using Python (24%).
  - Such GUI tools may still not be convenient enough for journalists.

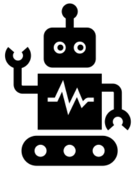
[7] https://www.tableau.com/products/desktop
[8] Wongsuphasawat, Kanit, et al. "Voyager 2: Augmenting visual analysis with partial view specifications." CHI' 17. 2017.

Natural Language Interface

- A more natural way is to introduce natural language-based interaction to facilitate exploratory visual analysis.
- Users can directly ask questions about data in natural language and receive visualizations as responses.

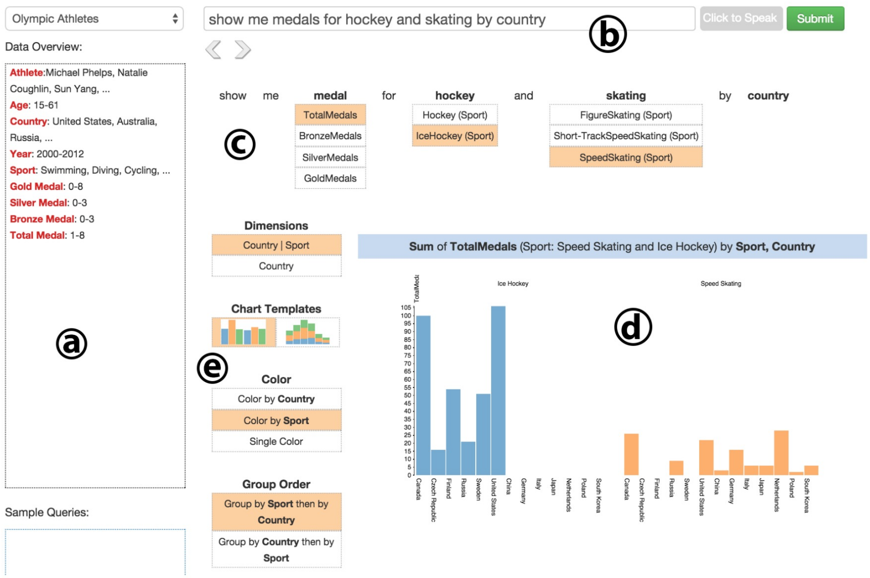
What is the relationship between production budget and running time?





Natural Language Interface

- Some existing work include:
  - Methods supporting one-shot questions without considering query history



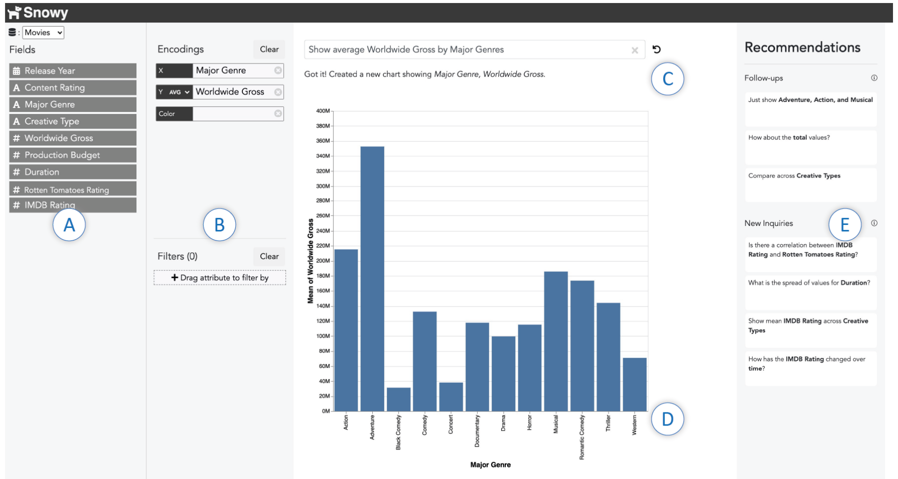
DataTone[9]

Sevi[10]

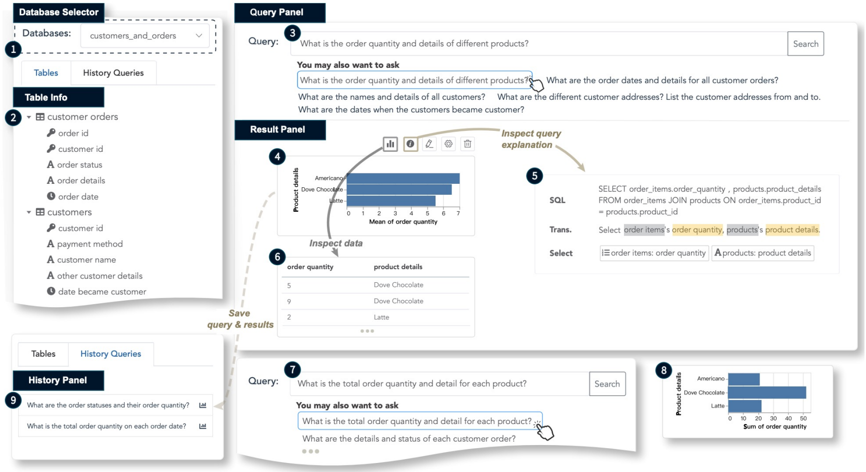
[9] Gao, Tong, et al. "Datatone: Managing ambiguity in natural language interfaces for data visualization." UIST '15. 2015.
[10] Tang, Jiawei, et al. "Sevi: Speech-to-visualization through neural machine translation." SIGMOD '22. 2022.

Natural Language Interface

- Some existing work include:
  - Methods supporting one-shot questions without considering query history
  - Systems recommending follow-up queries



Snowy[11]



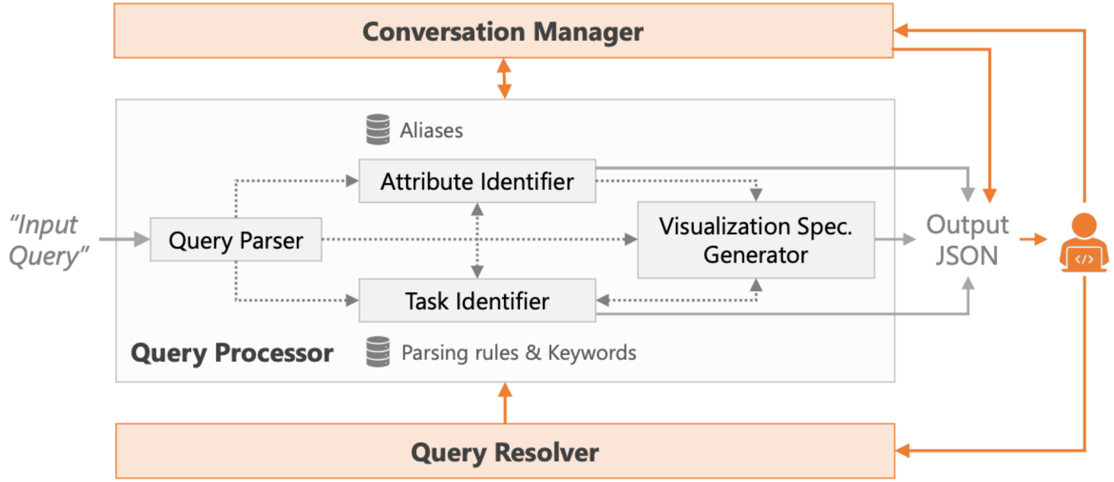
QRec-NLI[12]

- They do not support multi-turn conversation-like data exploration, which might not suit the iterative nature well.

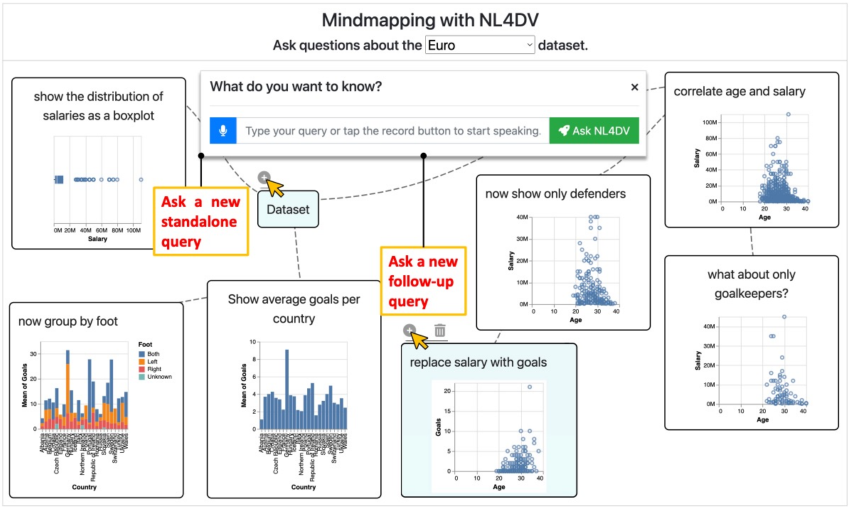
[11] Srinivasan, Arjun, and Vidya Setlur. "Snowy: Recommending utterances for conversational visual analysis." UIST '21. 2021.
[12] Wang, Xingbo, et al. "Interactive data analysis with next-step natural language query recommendation." arXiv (2022).

Challenges

- The only work considering query history is NL4DV,[13] a heuristics-based python package.



NL4DV Workflow



NL4DV Application

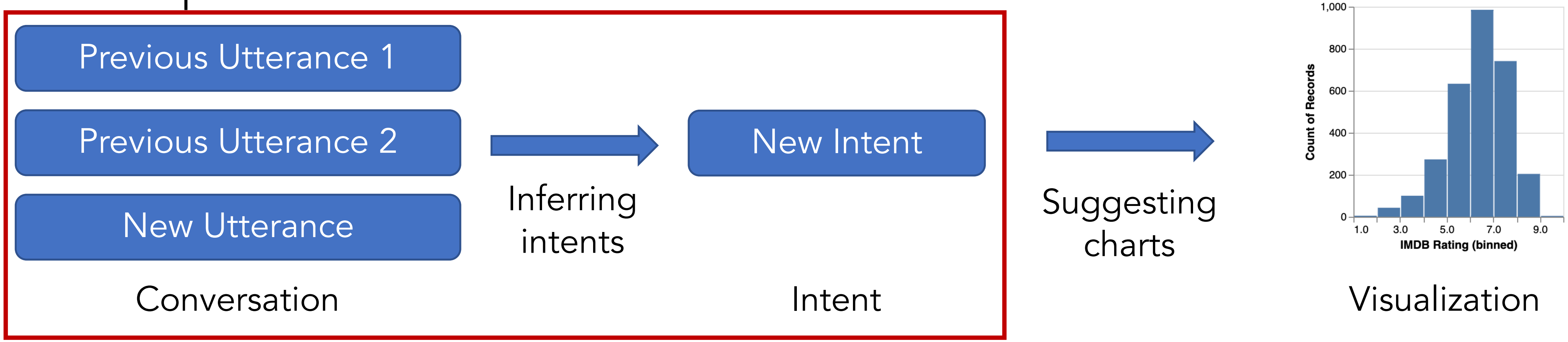
- However, its heuristic-based method limits its flexibility and performance,
  - e.g., it has a limited number of mapping between words and operators.

[13] Narechania, Arpit, Arjun Srinivasan, and John Stasko. "NL4DV: A toolkit for generating analytic specifications for data visualization from natural language queries." IEEE TVCG 27.2 (2020): 369-379.

Inferring visualization intents from conversations

- Informed by the recent advances of large language models (LLMs), we explore enhancing conversation-like NL-based data exploration with them.

- We consider suggesting visualizations from natural language queries as a two-step workflow.



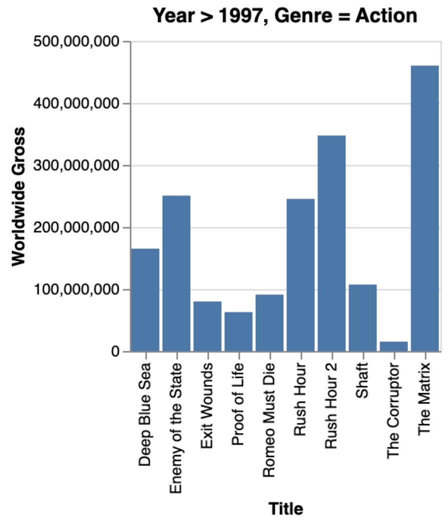
- We focus on the part of <u>understanding users' intent development</u> according to their sequence of questions.

Problem Formulation

- We consider users' intent as two parts following previous research[14]:
  - <u>**Visualized attributes**</u>, e.g., Gross and Title;
  - <u>**Data filters**</u>, e.g., Year > 1997, Genre = Action.

- With these two parts, visualizations can be recommended according to users' focused data attributes and the data subset.

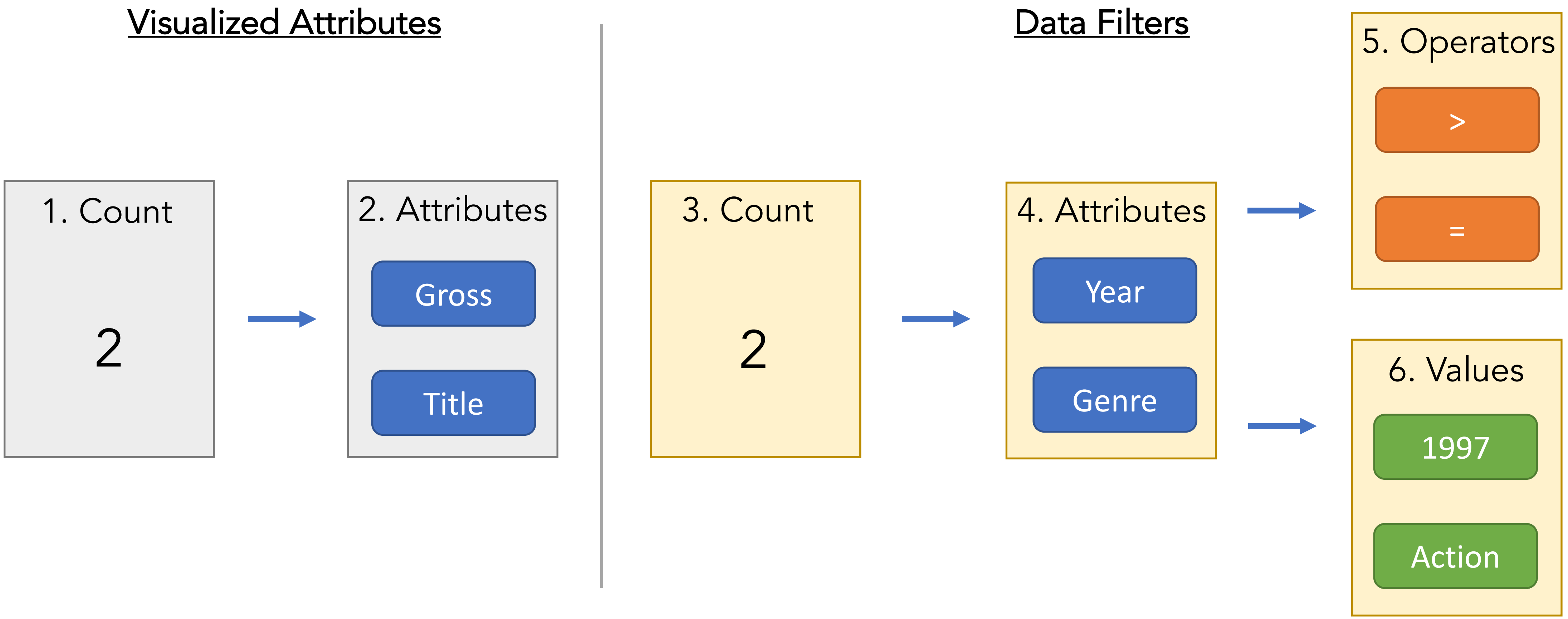
Title	Gross	Year	Genre
Titanic	1842879955	1997	Drama
Rush Hour	245300000	1998	Action
The Matrix	460279930	1999	Action



[14] Lee, Doris Jung-Lin, et al. "Lux: always-on visualization recommendations for exploratory dataframe workflows." Proceedings of the VLDB Endowment 15.3 (2021): 727-738.

Proposed Framework

- To predict the intent, we propose a framework where the whole task can be broken into two groups of six sub-tasks :



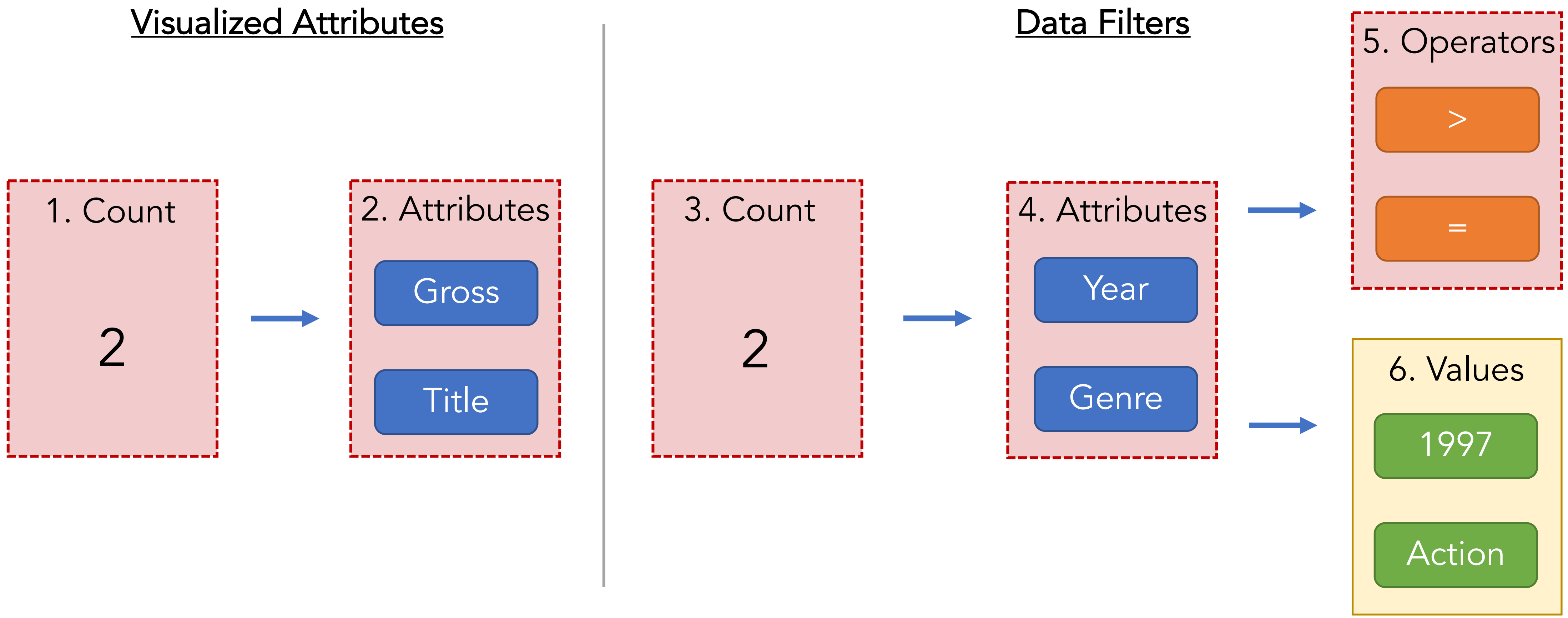
Proposed Framework

- Compared to formulating it as a translation task, the benefit of dividing intent prediction into sub-tasks is to eliminate the learning of structure of visualization intent.
- Therefore, we can
  - Make sure the output is always valid
    - For example, Picard[15], a state-of-the-art NL2SQL model, has 2% of SQL queries that cannot be executed.
  - Simplify the task to improve the performance

[15] Scholak, Torsten, Nathan Schucher, and Dzmitry Bahdanau. "PICARD: Parsing incrementally for constrained auto-regressive decoding from language models." EMNLP '21. 2021.

Task Overview

- Due to the features of different tasks, we have different setup of them.
- Among all six tasks, five are conducted with LLMs.



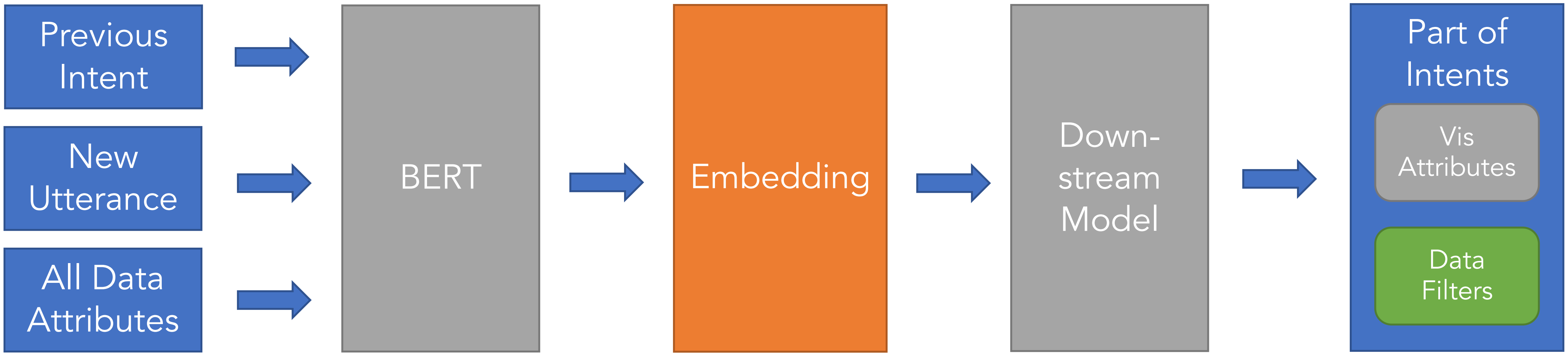
LLM Task Overview

- We fine-tune pre-trained BERTs[16] (110 million parameters base version) with a variant of CoSQL[17] (Conversation to SQL) dataset.
- We extract the attributes selected in SQL queries as visualized attributes and the predicates as data filters.
- Example:
  - Utterance: What are the populations of every country in Africa?
  - SQL: SELECT Name, Population FROM Country WHERE Continent = "Africa"
  - Intent:
    - Attribute: Name, Population
    - Filter: Continent = "Africa"

[16] Devlin, Jacob, et al. "BERT: Pre-training of deep bidirectional transformers for language understanding." NAACL '19. 2019.
[17] Yu, Tao, et al. "CoSQL: A conversational text-to-sql challenge towards cross-domain natural language interfaces to databases." EMNLP '19. 2019.

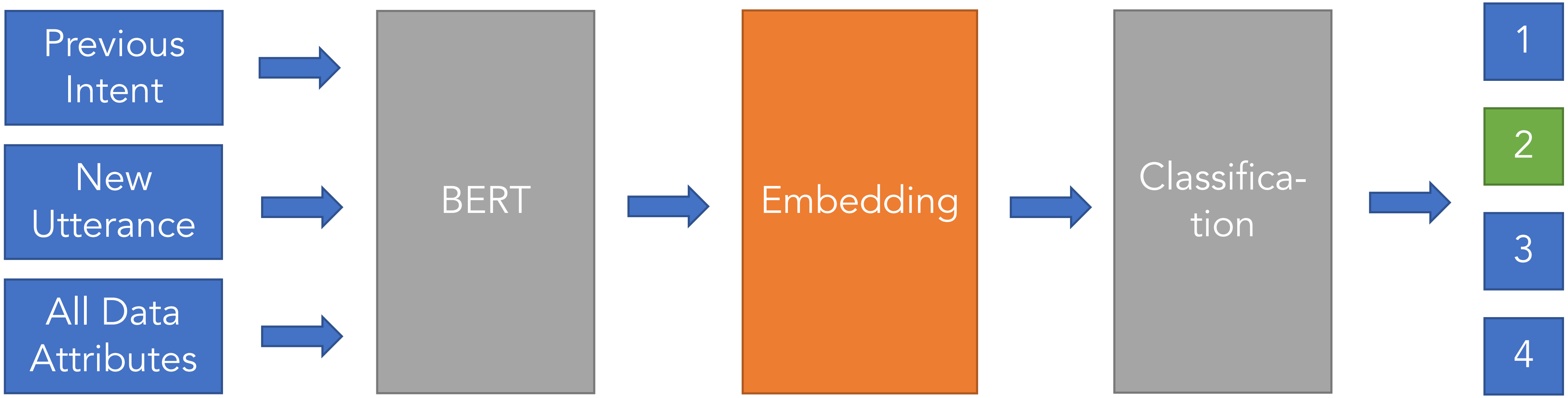
LLM Task Overview

- Structure overview:



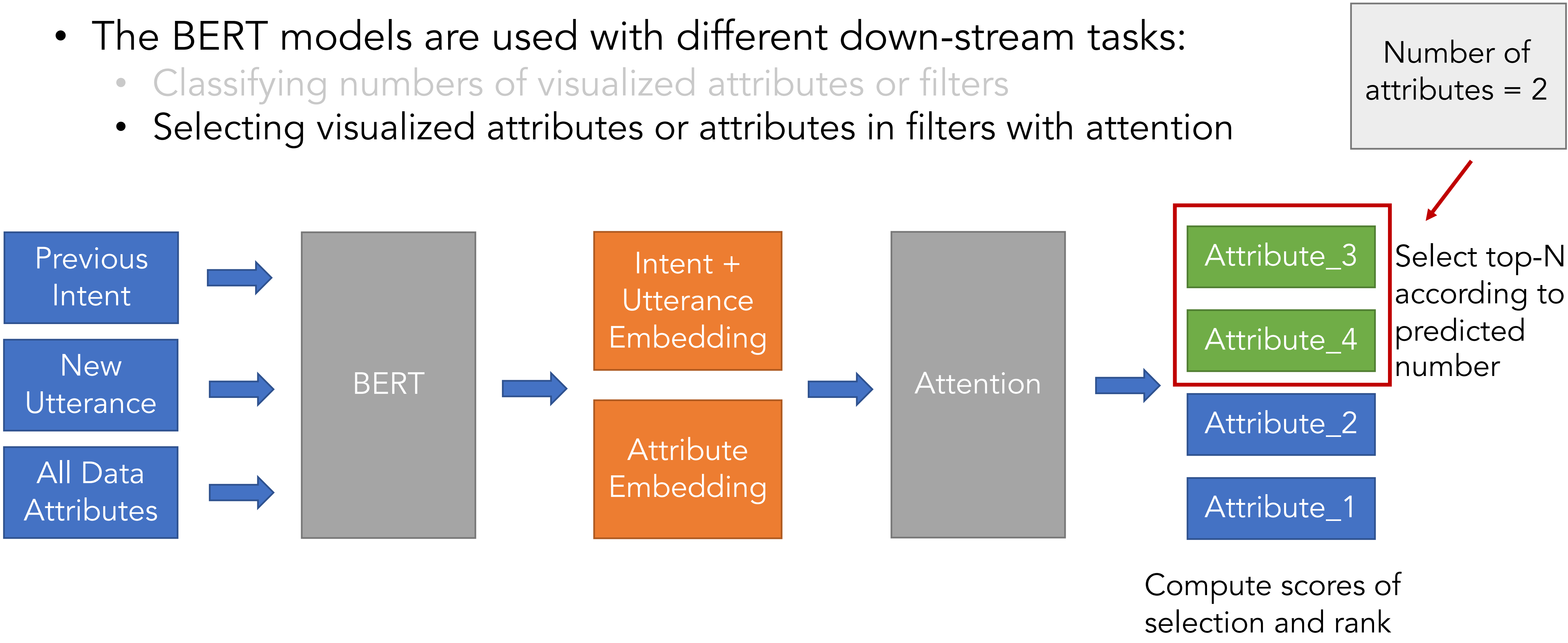
Number of Visualized Attributes

- The BERT models are used with different down-stream tasks:
  - Classifying numbers of visualized attributes or filters



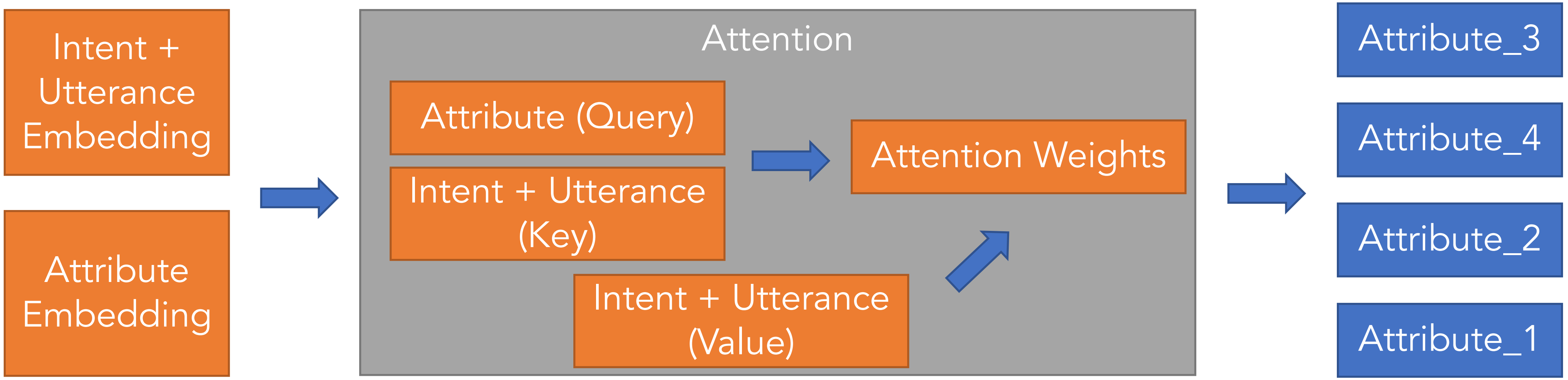
Visualized Attributes

- The BERT models are used with different down-stream tasks:
  - Classifying numbers of visualized attributes or filters
  - Selecting visualized attributes or attributes in filters with attention



Attention Mechanism

- In our models, we leverage attention mechanisms[18] since it demonstrates great power in dealing with natural language problems.
- It can match users' natural language utterances with the semantically meaningful attribute names.

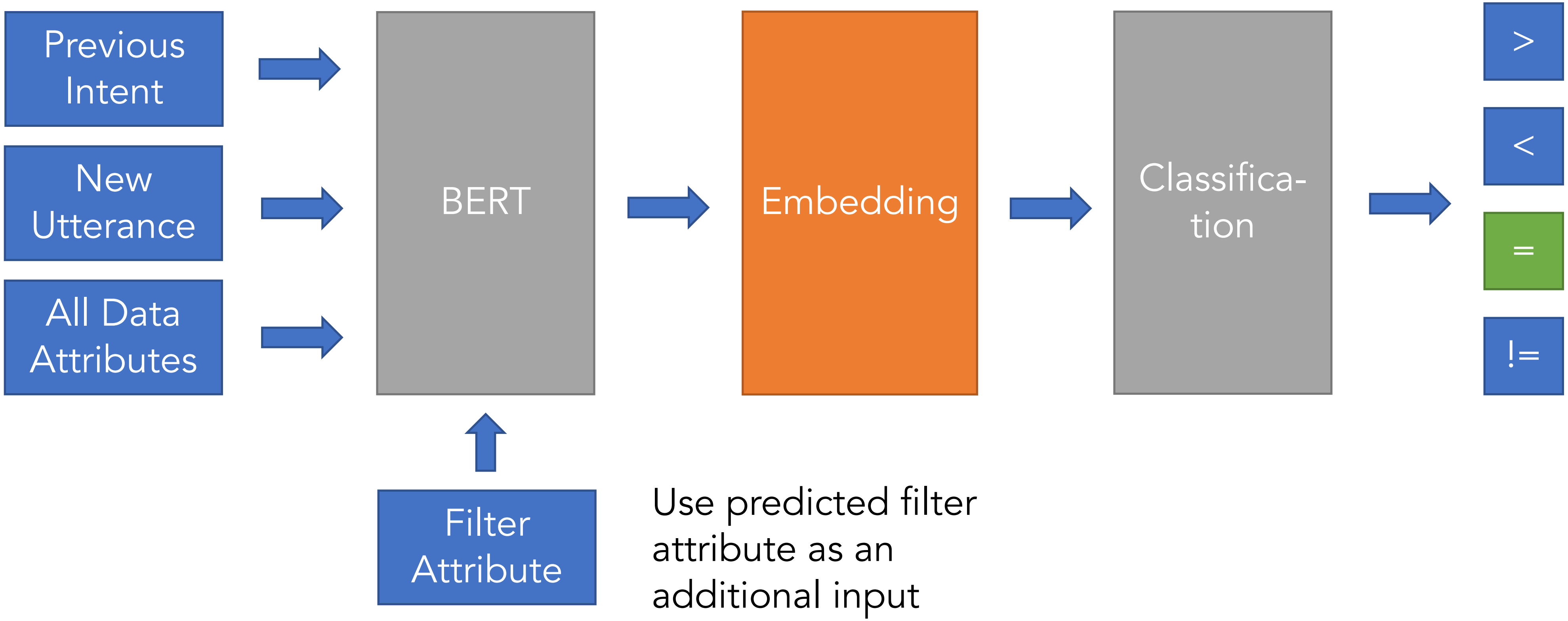


Receive scores of selection and rank

[18] Vaswani, Ashish, et al. "Attention is all you need." NeurIPS' 17. 2017

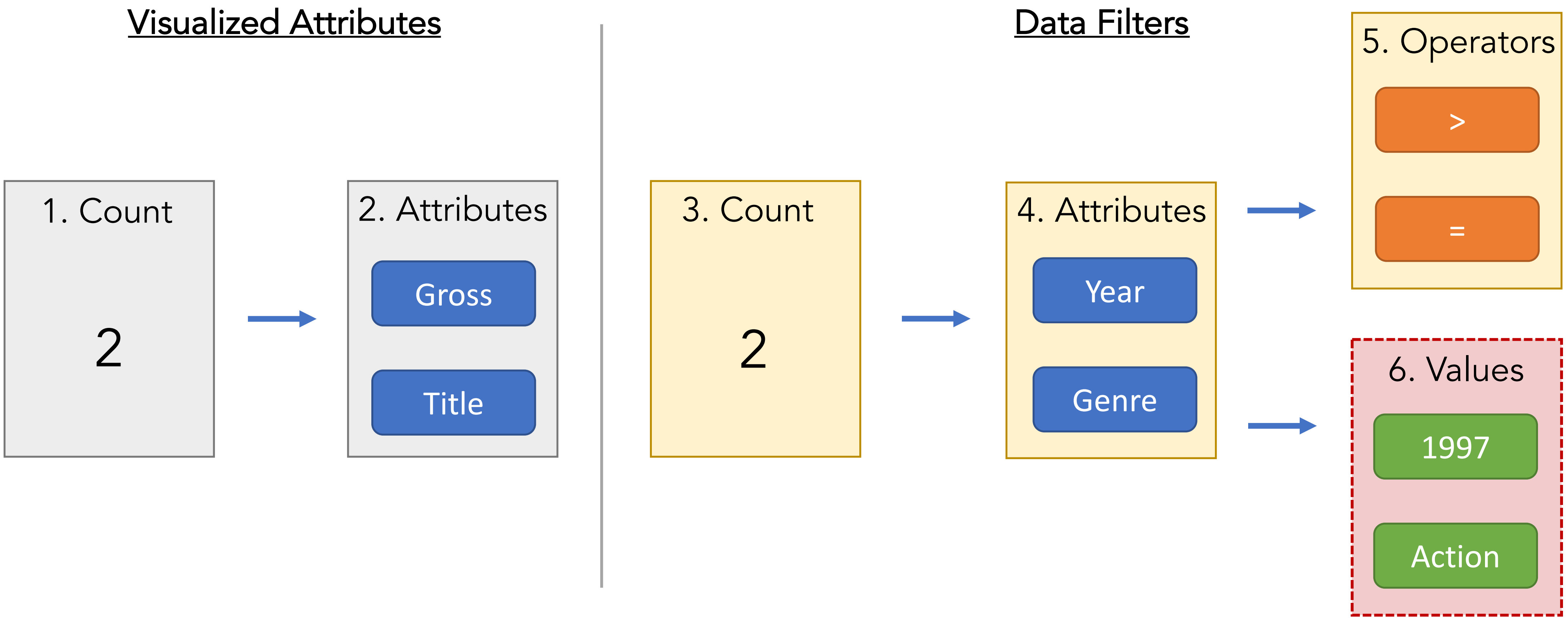
Operators in Filters

- The BERT models are used with different down-stream tasks:
  - Classifying numbers of visualized attributes or filters
  - Selecting visualized attributes or attributes in filters with attention
  - Classifying operators in filters



Task Overview

- Due to the features of different tasks, we have different setup of them.
- Our approach identifies values in filters with a heuristic-based method.

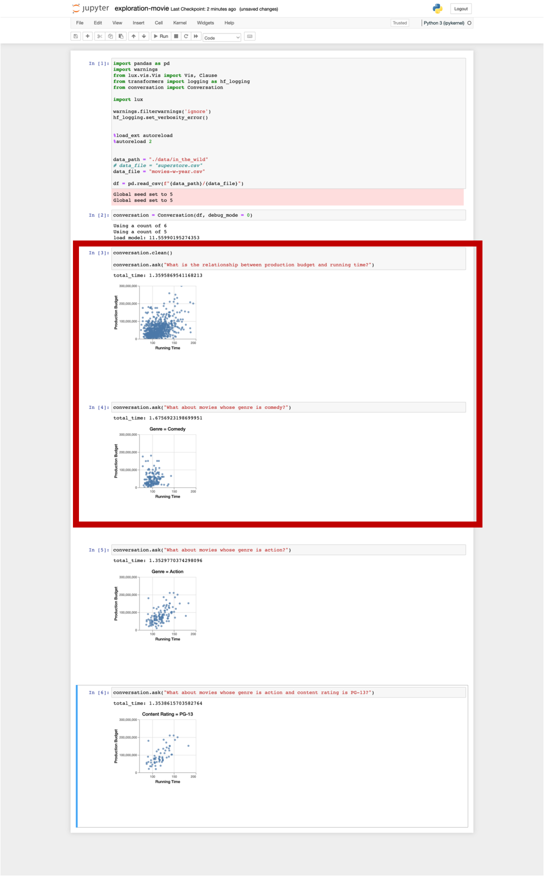


Demo Case

- We implement a demo of our approach as a notebook plugin with Lux as the visualization recommendation engine.
- Let's see a demo case about the movie dataset.

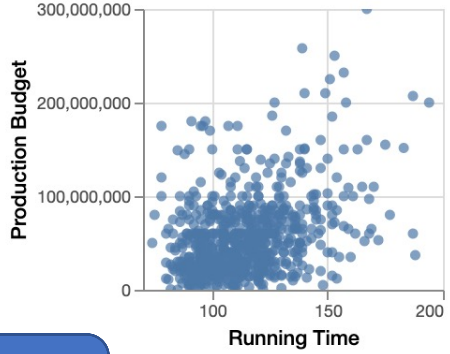
Title	Worldwide Gross	Production Budget	Running Time	Year	Genre	IMDB Rating
Titanic	1842879955	200000000	194	1997	Drama	7.4
Rush Hour	245300000	35000000	98	1998	Action	6.8
The Matrix	460279930	65000000	136	1999	Action	8.7

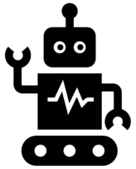
Demo Case - 1



Intent

What is the relationship between production budget and running time?

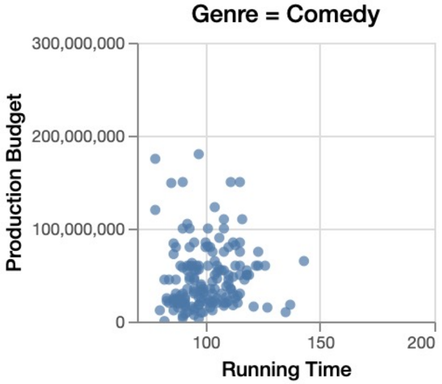


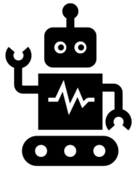


Production Budget

Running Time

What about movies whose genre is comedy?



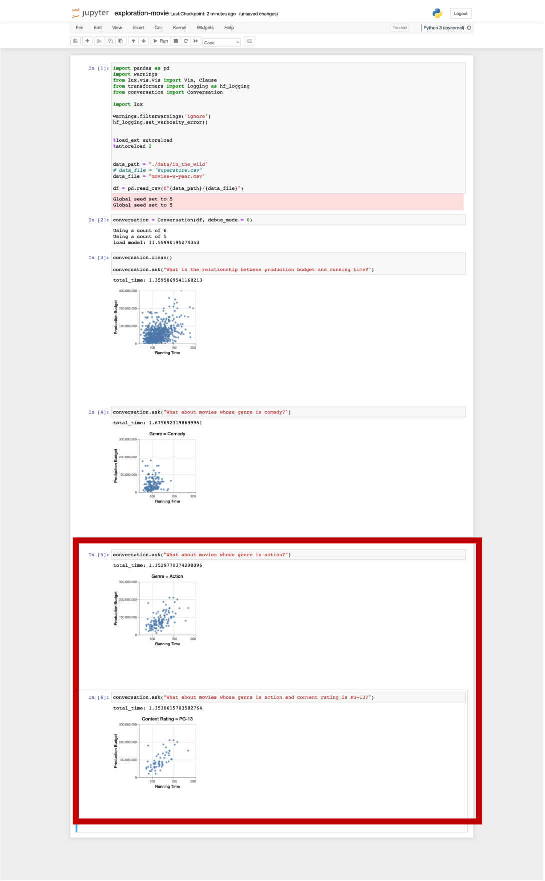


Production Budget

Running Time

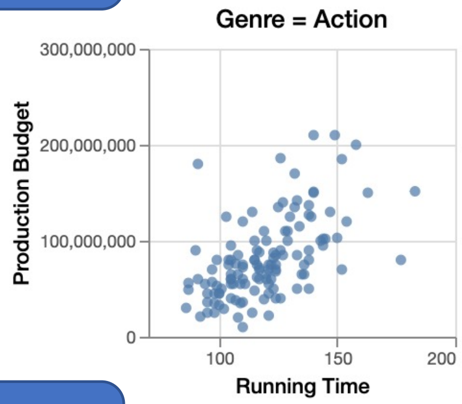
Genre = Comedy

Demo Case - 2



Intent

What about movies whose genre is action?

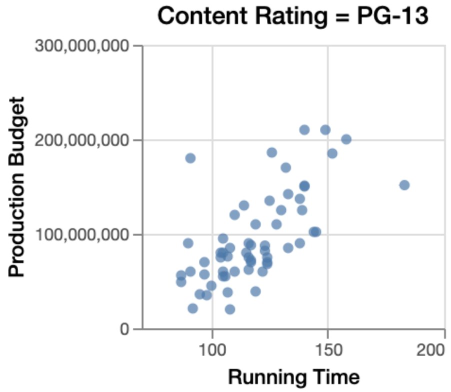


Production Budget

Running Time

Genre = Action

What about movies whose genre is action and content rating is PG-13?



Production Budget

Running Time

Genre = Action

Content Rating = PG-13

Quantitative Evaluation - Setup

- Since there is no previous models that follow the same problem formulation as ours, our preliminary evaluation compares our models with two baselines:
  - Picard: a state-of-the-art NL2SQL model with T5[19] (3-billion and 250-million versions) trained on CoSQL;
    - (As a reference, our models have ~550 million parameters in total.)
    - The intent is extracted from their generated SQL queries.
  - NL4DV: a heuristics-based which supports recommend visualizations according to conversations.
    - The intent is extracted from its recommended visualizations.
- All models are evaluated with 20% of the CoSQL dataset using the metric of accuracy. The accuracy delineate the number of completely correct intent.

[19] Raffel, Colin, et al. "Exploring the limits of transfer learning with a unified text-to-text transformer." JMLR 21.1 (2020): 5485-5551.

Quantitative Evaluation - Overall Accuracy

Method	Test Accuracy
Ours	50.6112
Picard-3B	48.4108
Picard-Base	33.4963
NL4DV	10.4218

- It is clear that our model both outperforms Picard and NL4DV.
- Notably, our model achieves the performance using a smaller number of parameters (550million) than the best baseline model, Picard-3B (3 billion).

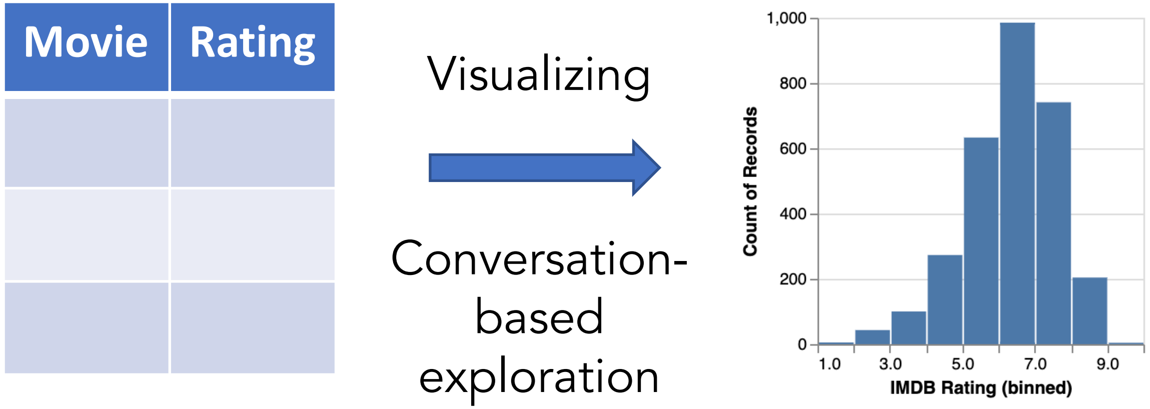
Quantitative Evaluation - Breakdown Accuracy

	Vis Attribute Accuracy	Filter Accuracy
Ours	65.7702	77.9951
Picard-3B	72.8606	61.8582
Picard-Base	53.0562	52.8117
NL4DV	20.3474	33.4988

- Our approach outperforms all other methods in filter accuracy.
- It also outperforms most of the other approaches in attribute accuracy.

Conclusion & Discussion

- In our research, we explore a new approach of understanding users' intent of visualizations from their conversation to facilitate a natural language-based interaction with data.



Data Table Chart

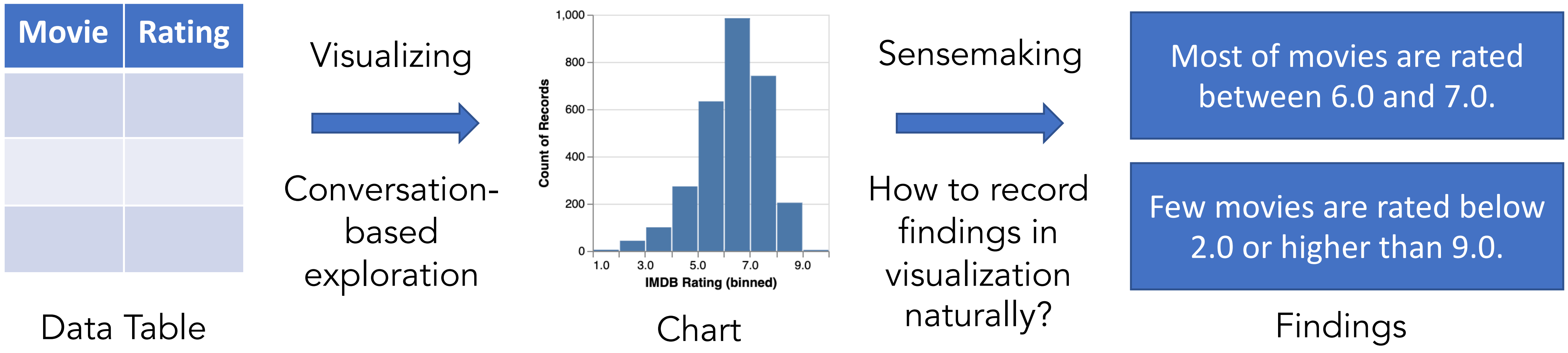
GPT?

- GPT families[20] (GPT-3, GPT-4, and ChatGPT) have demonstrated their power in dealing with conversations.
- Our method presents a general framework of how we can address the challenge of inferring visualization intent from conversations through a divide-and-conquer strategy.
- As our evaluation shows, smaller models can achieve better performance than a large model.
- It is easy to change BERT models to GPT models as backbone in our approach and we may expect a better performance.

[20] https://platform.openai.com/docs/models

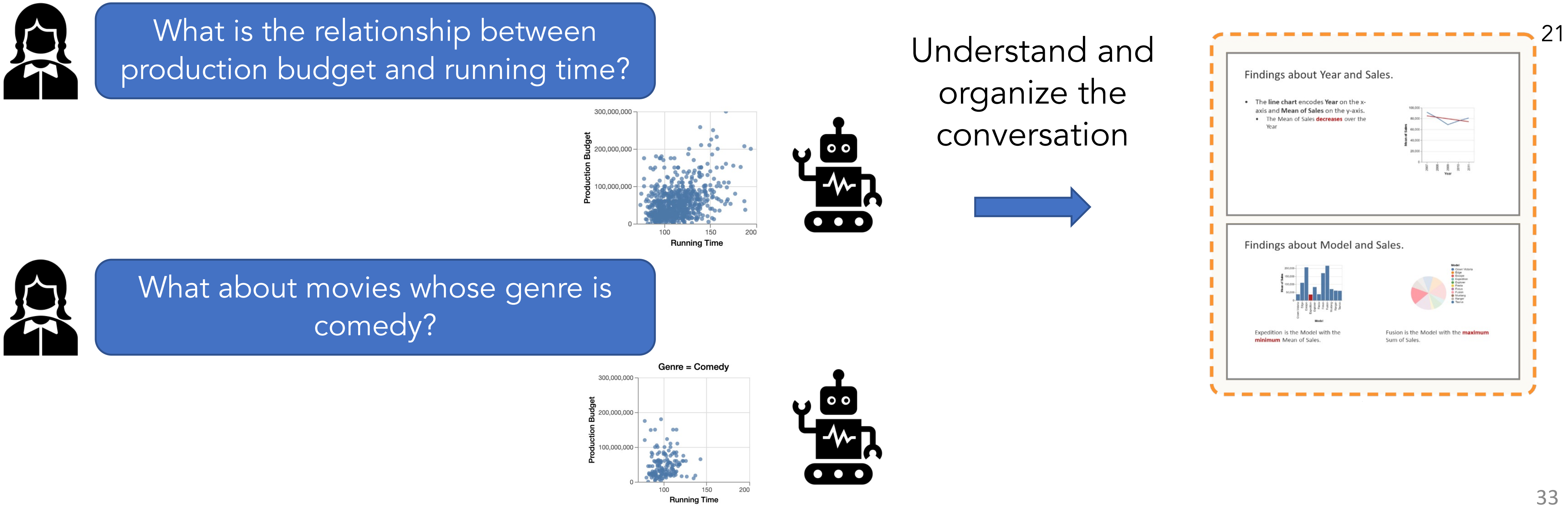
Future Work

- In the future, we plan to further extend our work from multiple perspectives, such as:
  - Introducing natural interactions to other steps in exploratory analysis



Future Work

- In the future, we plan to further extend our work from multiple perspectives, such as:
  - Introducing natural interactions to other steps in exploratory analysis
  - Leveraging conversations to bridge exploring and communicating data



[21] Li, Haotian, et al. "Notable: On-the-fly assistant for data storytelling in computational notebooks." CHI '23. 2023.

Thank you!

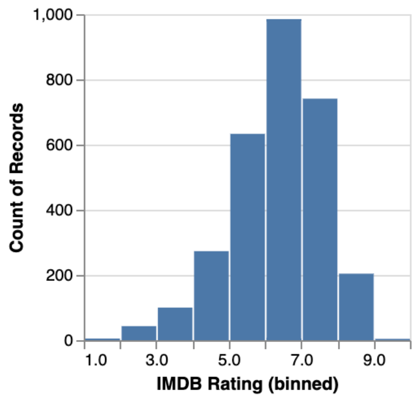
Towards Natural Interactions with Data

Movie	Rating

Data Table

Visualizing

Conversation-based exploration



Chart

Homepage: https://haotian-li.com
Email: haotian.li@berkeley.edu