VIS 2021

KG4Vis: A Knowledge Graph-Based Approach for Visualization Recommendation

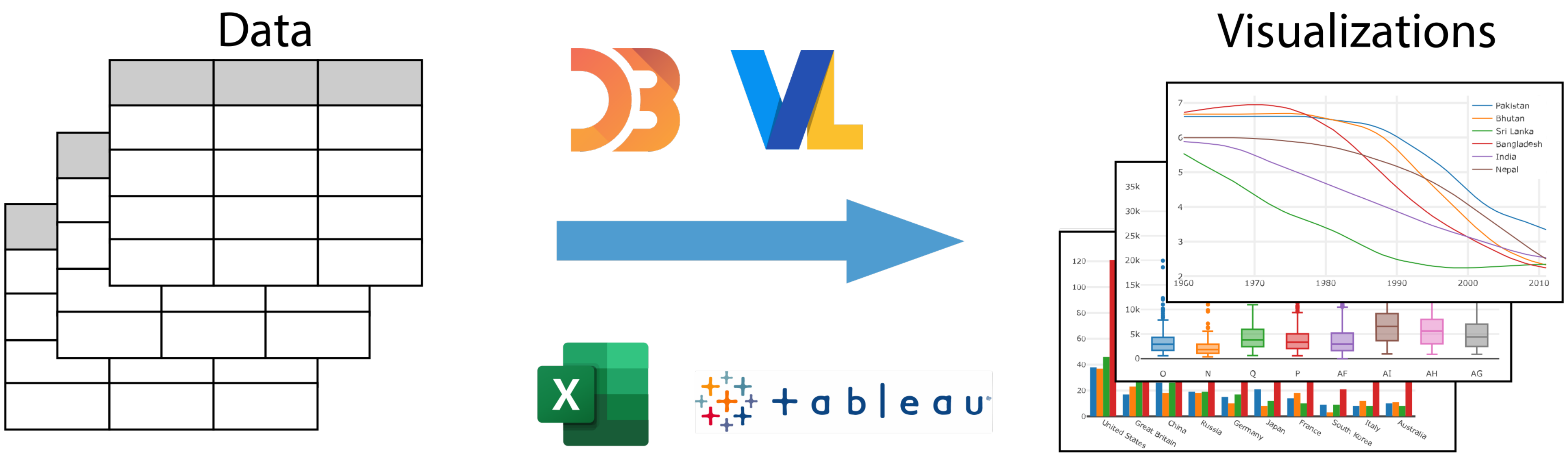
    

Haotian Li[1,2] Yong Wang[2] Songheng Zhang[2] Yangqiu Song[1] Huamin Qu[1]

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

SINGAPORE MANAGEMENT UNIVERSITY

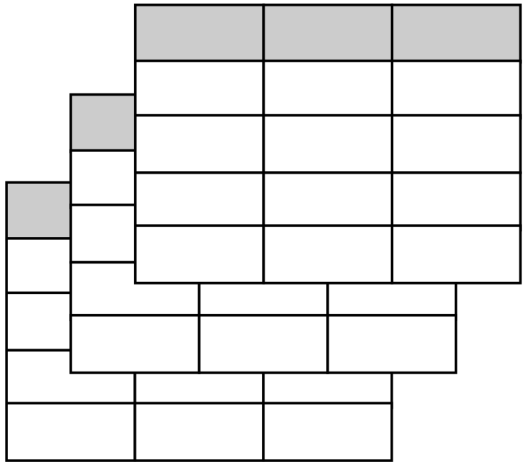
Motivation



The barrier of creating effective visualizations is high.

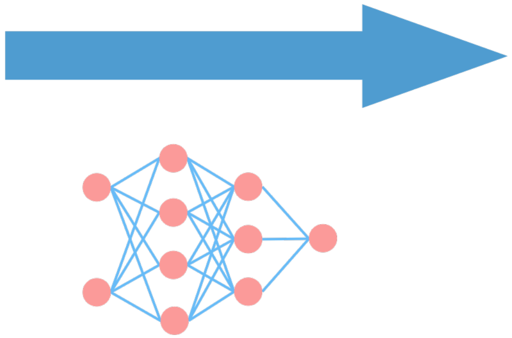
Visualization Recommendation

Data



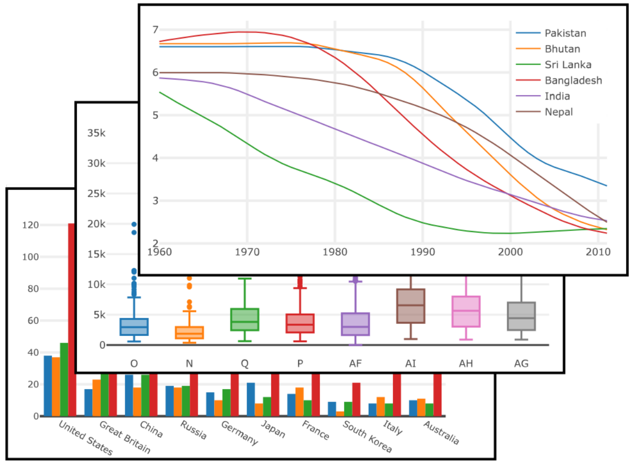
1. Rule-based

IF this THEN that



2. ML-based

Visualizations

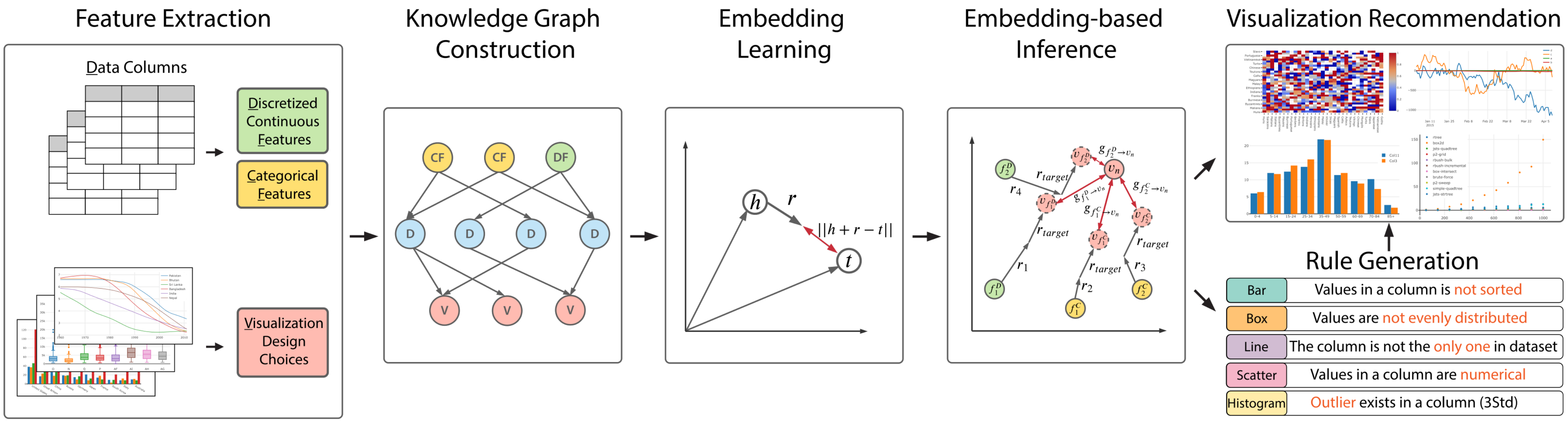


K. Z. Hu, M. A. Bakker, S. Li, T. Kraska, and C. A. Hidalgo. Vizml: A machine learning approach to visualization recommendation. CHI, 2019.

Research Question

Can we achieve visualization recommendation that requires no manual specifications of rules and guarantees good explainability?

KG4Vis



Our knowledge graph (KG)-based visualization recommendation approach is **data-driven** and **explainable**.

Introduction

Method

Feature Extraction

KG Construction

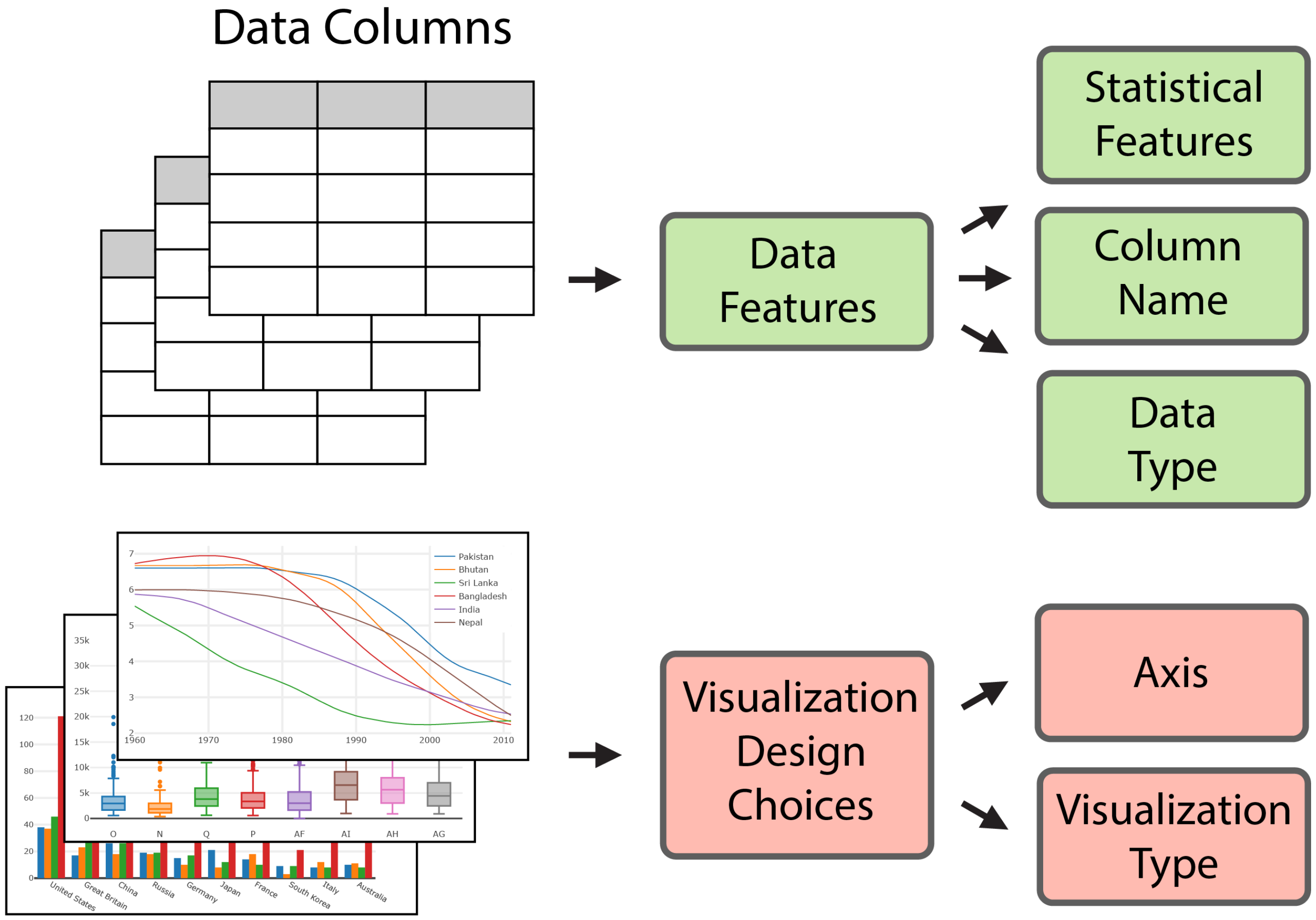
Embedding Learning

Inference

Evaluation

Discussion

Feature Extraction



K. Z. Hu, M. A. Bakker, S. Li, T. Kraska, and C. A. Hidalgo. Vizml: A machine learning approach to visualization recommendation. CHI, 2019.

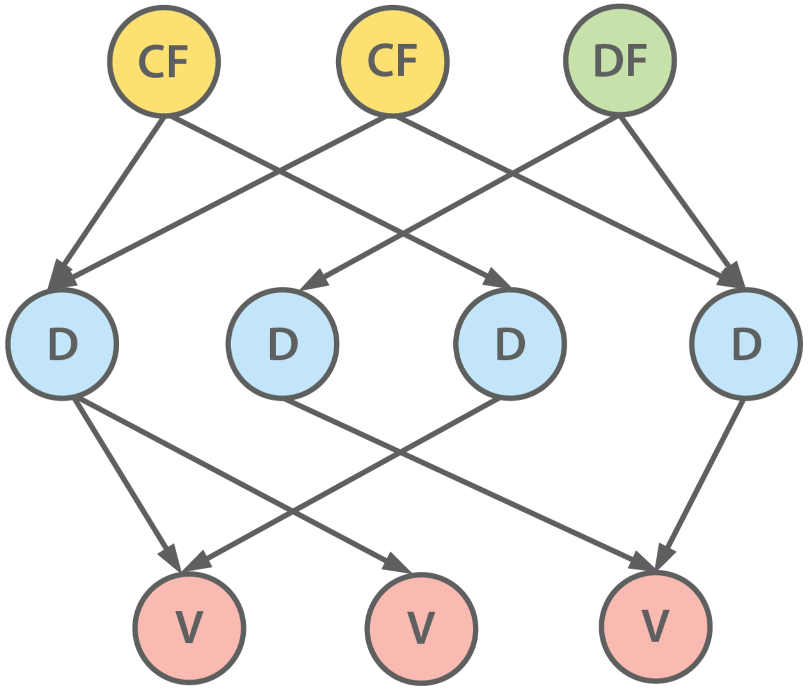
KG Construction

Entities:

- Discretized continuous features
  - Use each interval after discretization as an entity
  - We employ a discretization algorithm based on minimum description length principle (MDLP)
- Categorical features
- Data columns
- Visual designs

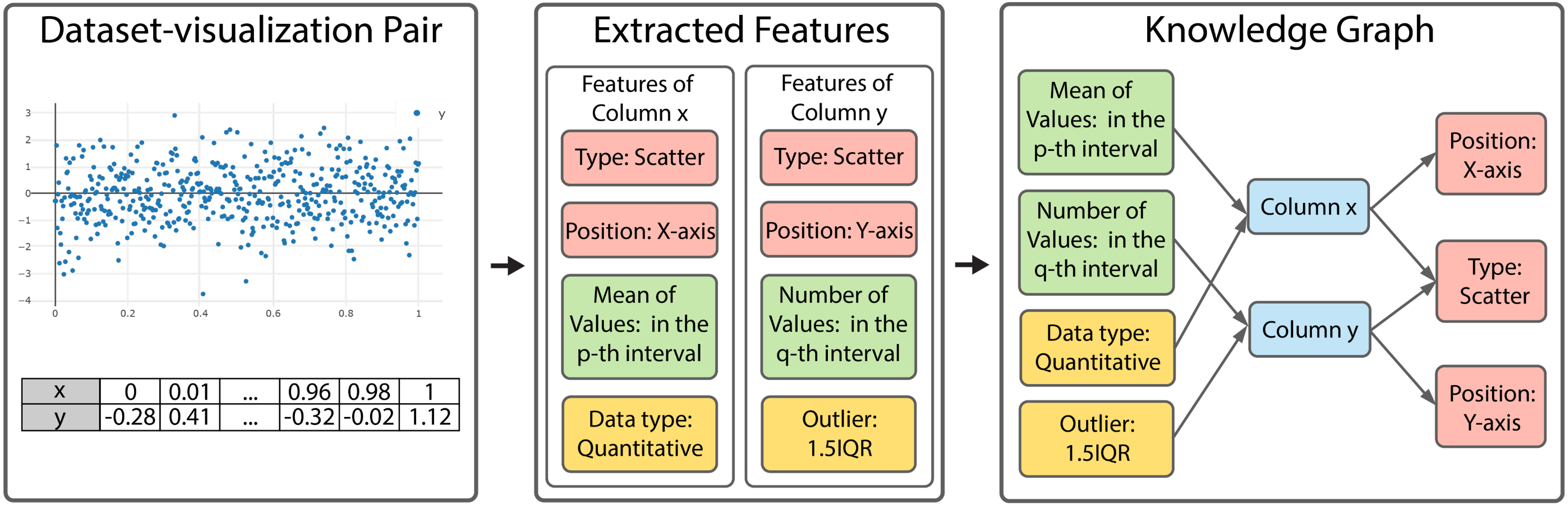
Relations:

- Defined based on entity types



U. M. Fayyad and K. B. Irani. Multi-interval discretization of continuous valued attributes for classification learning. IJCAI, 1993.

A Example KG


Dataset-visualization Pair
y
3
2
1
0
−1
−2
−3
−4
0
0.2
0.4
0.6
0.8
1
x | 0 | 0.01 | ... | 0.96 | 0.98 | 1
y | -0.28 | 0.41 | ... | -0.32 | -0.02 | 1.12
Extracted Features
Features of Column x
Type: Scatter
Position: X-axis
Mean of Values: in the p-th interval
Data type: Quantitative
Features of Column y
Type: Scatter
Position: Y-axis
Number of Values: in the q-th interval
Outlier: 1.5IQR
Knowledge Graph
Mean of Values: in the p-th interval
Number of Values: in the q-th interval
Data type: Quantitative
Outlier: 1.5IQR
Column x
Column y
Position: X-axis
Type: Scatter
Position: Y-axis

Embedding Learning

Triplet (denotes a edge in KG):

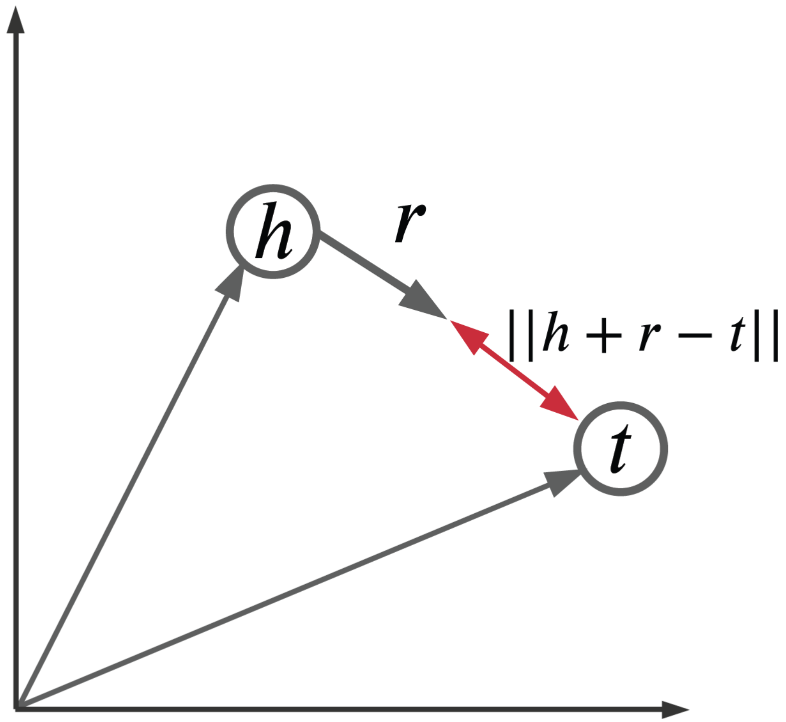
$$(head\,entity, relation, tail\,entity) \text{ or } (h, r, t)$$

TransE assumption:

$$\mathbf{h} + \mathbf{r} \approx \mathbf{t}$$

TransE scoring function (measures the possibility of a triplet):

$$g(h, r, t) = -||\mathbf{h} + \mathbf{r} - \mathbf{t}||_{1/2}$$



A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. NeurIPS, 2013.
Z. Sun, Z. Deng, J. Nie, and J. Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. ICLR, 2019.

Inference

Rule structure:

a data feature $\rightarrow$ ***a visual design choice*** or $f_i \rightarrow v_n$

Inference steps:

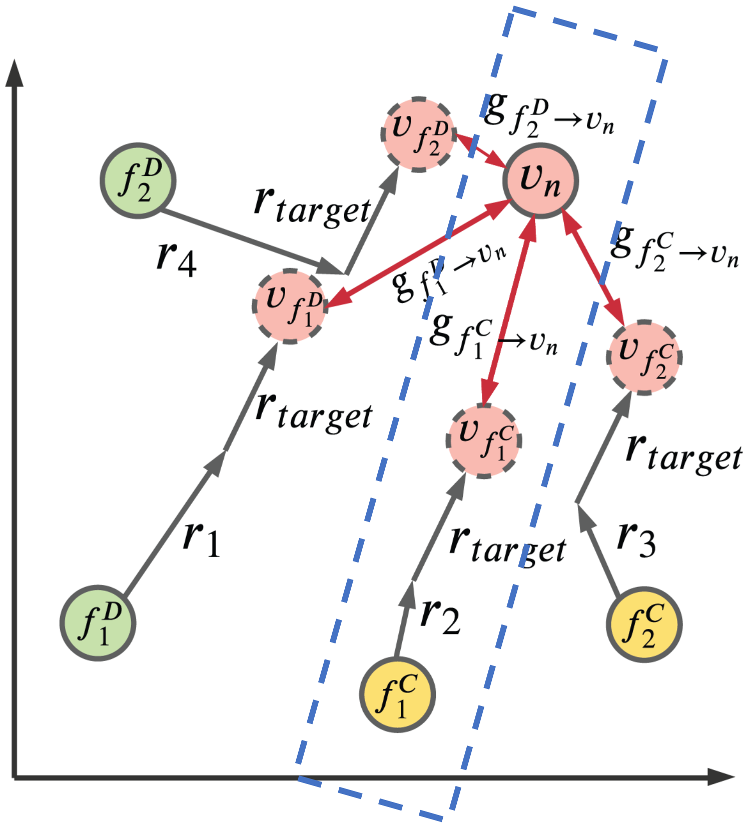
1. Compute rule score (possibility of the rule):

$$g_{f_i \rightarrow v_n} = -||\mathbf{f}_i + \mathbf{r}_j + \mathbf{r}_{target} - \mathbf{v}_n||$$

2. Aggregate all suitable rules' scores of a data column:

$$g(d_{new}, r_{target}, v_n) = \frac{1}{|F_{new}|} \sum_{f_i \in F_{new}} g_{f_i \rightarrow v_n}$$

3. Recommend the design with the highest score



Introduction

Method

Evaluation

Quantitative Evaluation

User Study

Case Study

Discussion

Quantitative Evaluation

Methods in comparison:

- TransE-adv (used in KG4Vis): TransE with self-adversarial negative sampling
- TransE: original TransE
- RotatE: $g(h,r,t) = -\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|_{1/2}$

Metrics:

- *MR*: mean rank of the correct design choices
- *Hits@2*: proportion of correct design choices ranked in the top two recommendations

	Axis	Visualization Type	
	Accuracy	MR	Hits@2
TransE-adv	**0.7350**	**1.9567**	**0.7489**
TransE	0.7214	1.9718	0.7445
RotatE	0.7193	1.9608	0.7458

TransE-adv outperforms others.

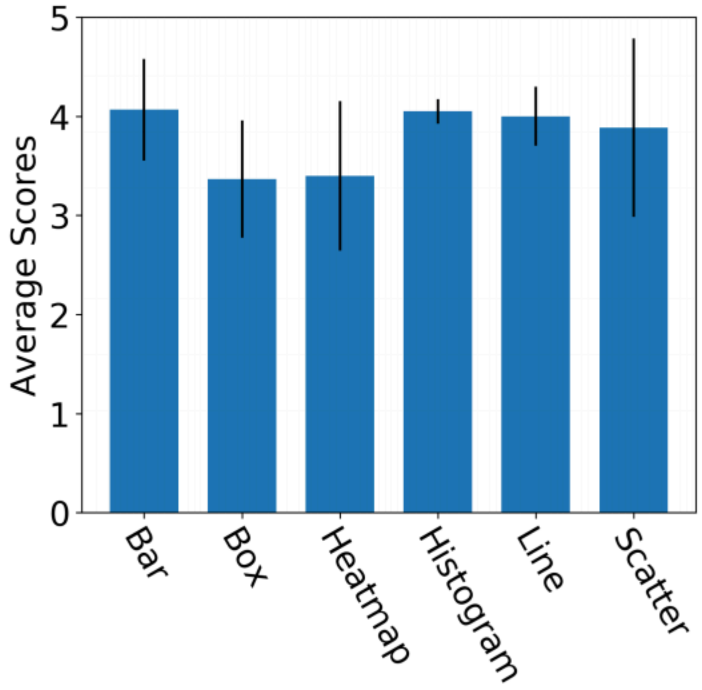
A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. NeurIPS, 2013.
Z. Sun, Z. Deng, J. Nie, and J. Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. ICLR, 2019.

Expert Interview

Results:

- Most of the rules are of high quality, but some features need to be further improved.

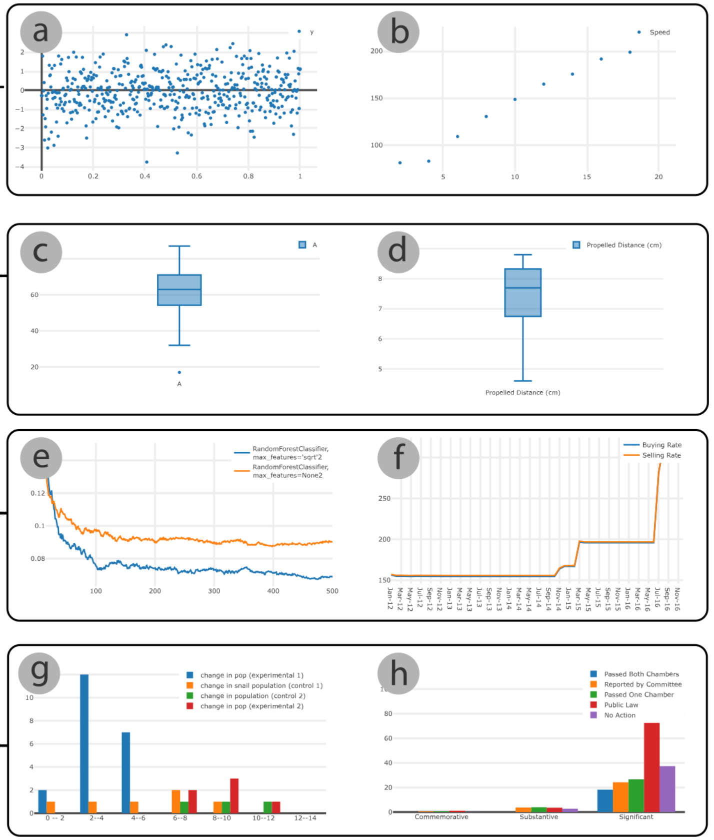
- The recommended visualizations are correct. Users' analytical tasks should be further taken into consideration.



Average scores of recommended visualizations

Case Study

Scatter	Values in a column are numerical
Scatter	Outlier exists in a column (1.5IQR)
Box	Values are not evenly distributed
Box	The entropy of values is large
Line	Values are monotonic
Line	The data type of a column is decimal
Line	The column is not the only one in dataset
Bar	All values in a column are unique
Bar	Values in a column is not sorted

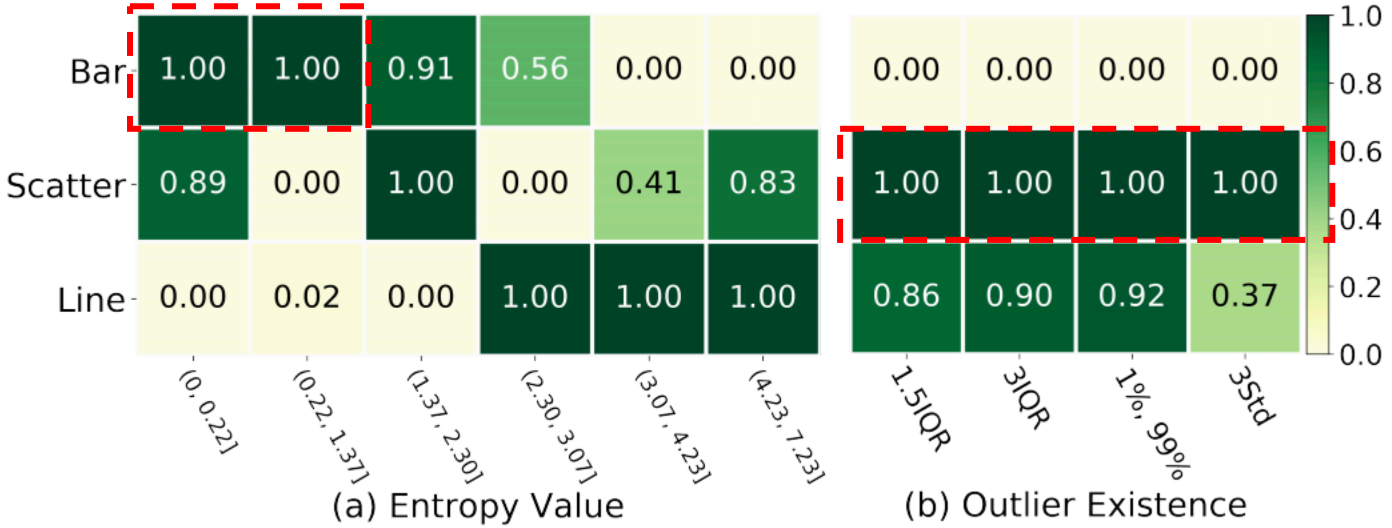


Our methods can learn commonly used explicit and implicit visual design rules.

Comparison with Empirical Studies

Ours align well with empirical rules:

- Bar charts are suitable for identifying clusters
- Scatter plots should be used to find anomalies



B. Saket, A. Endert, and C. Demiralp. Task-based effectiveness of basic visualizations. TVCG, 2019.

Introduction

Method

Evaluation

Discussion

Lessons

Pros and Cons

Future Work

Lessons

1. Knowledge graph for visualizations
    - Entity construction: discretize continuous features
    - Embedding learning: facilitate inference and rule generation

2. Explainability of rules
    - Straightforward features in conditions
    - Number of conditions

Pros and Cons

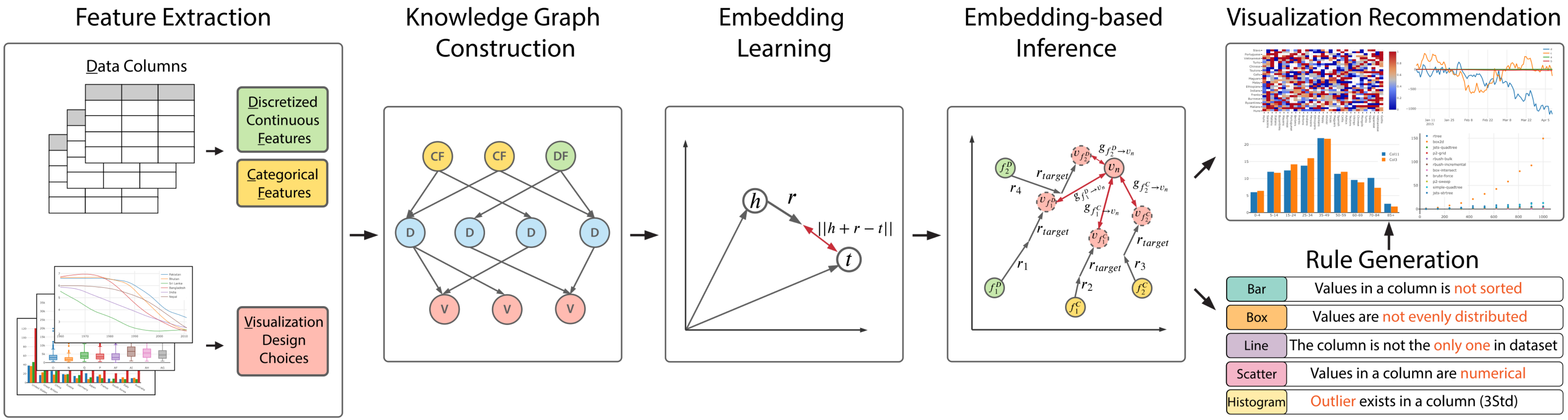
1. Compared with rule-based methods
   - Have better extendability and require less human effort
   - Rely on the quality of corpus

2. Compared with ML-based methods
   - Improve the explainability
   - Have potential performance drop

Future Work

- Introduce more visual designs such as color usage
- Consider analytical tasks and cross-column data features
- Extend to more visualization types including infographics

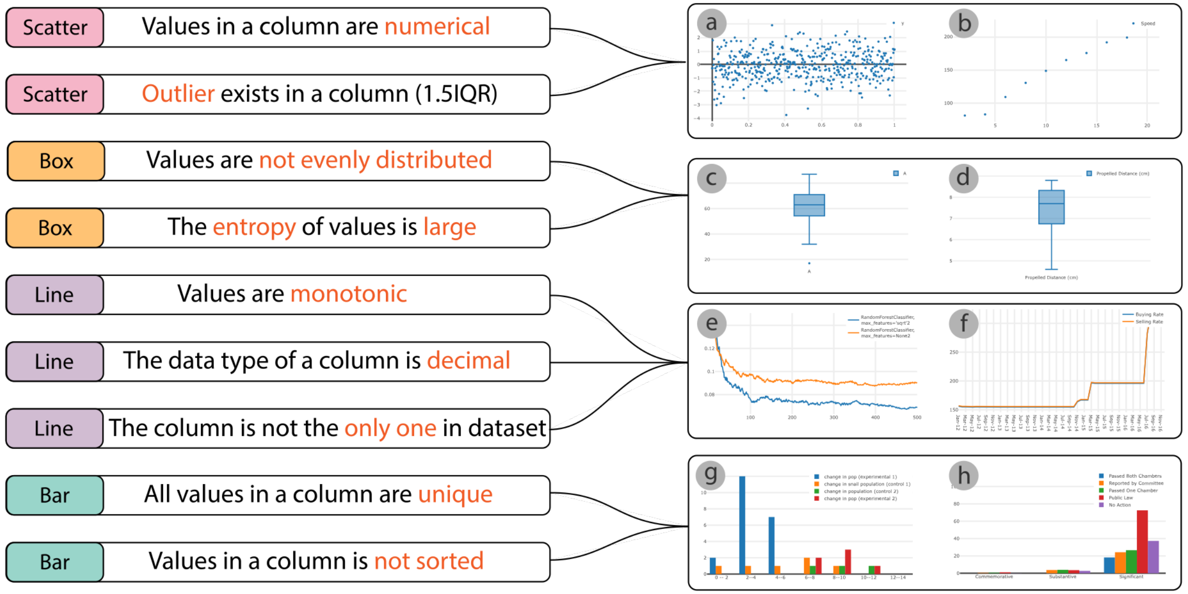
Take-home Message



- Knowledge graph provides an intuitive way to model the relationship between data and visualizations.
- Representing entities and relations with embeddings facilitates the further inference and the rule generation.
- Many factors affects the explainability of visualization rules, such as the complexity of features and the number of conditions.

VIS 2021

KG4Vis: A Knowledge Graph-Based Approach for Visualization Recommendation



https://kg4vis.github.io/

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

SINGAPORE MANAGEMENT UNIVERSITY