PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Recommender System to Support Chart Constructions with Statistical Data

Taissa Abdalla Filgueiras de Sousa
tsousa@inf.puc-rio.br

Simone Diniz Junqueira Barbosa
simone@inf.puc-rio.br

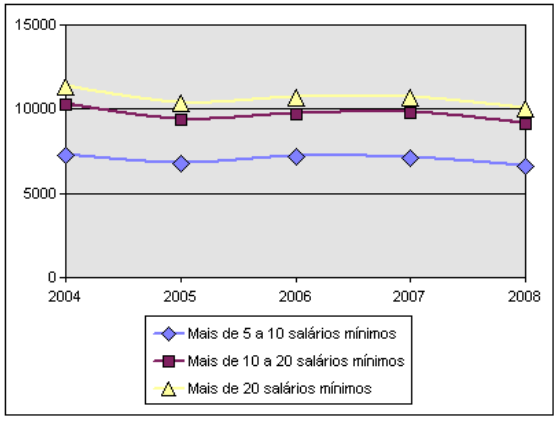
Problem

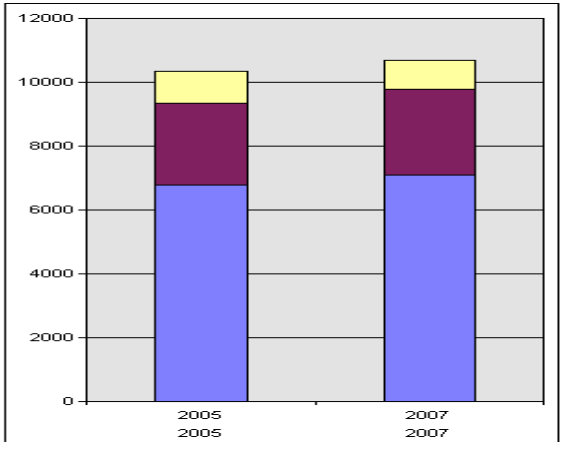
Difficulty in the construction of efficient charts

“the basic problem of chart construction is the selection of representation.” (Bertin,1918)

“... only few have skills to design effective graphic presentations of information” (Mackinlay, 2007)

– Example: Was there an increase in the total number of people with income higher than 5 monthly minimum wages between 2005 and 2007?





Research question

How can we support novice users to create efficient visualizations with statistical data?

Users:

- Students and professionals not related to statistics, journalism or data analysis.

Efficient visualizations:

- Those that can answer some specific questions in a single instant of perception

Related Work

1. Rules of graphic system
2. Techniques for data visualization
3. Research of Visualization tools
4. Evaluation with users using different visualization tools

Requirements

1. Generate efficient, clear and accurate charts
2. Motivate analysis
3. Allow many types of construction and math operations. Ex: calculate average, sum and difference
4. Develop precise meanings of view
5. Provide visual feedback, automatic visualizations and default values.
6. **Provide an interactive help feature**

Preliminary studies

- Visualization ontology that interrelates user questions, data features and efficient visualizations
- Techniques for recommender systems

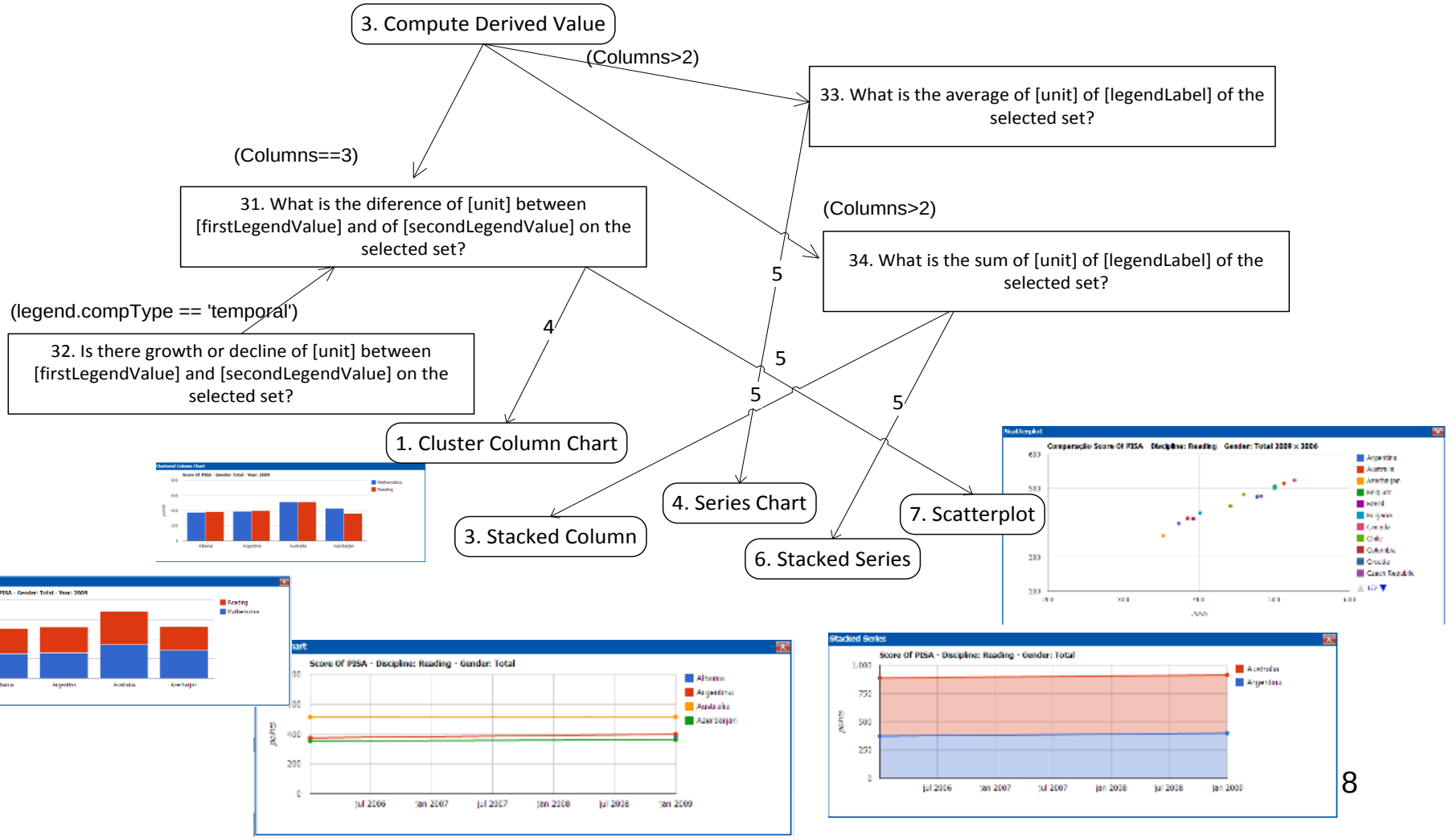
The ViSC ontology

Example of questions and tasks

- Where is there more people in the range of 14 years of education? (find extreme)
- What was the ranking of places in the range of 14 years of education? (sort)
- What was the PISA average score in the selected countries (calculate derived values)
- What was Canada's PISA score in math in 2003? (retrieve value)

The ViSC ontology

Exemple of task class



The ViSC recommender system

– Knowledge-based Recommender System

Background data	Input data	Process
Feature of **items**. Knowledge of how the items meet users' needs.	Description of needs or user interests.	Infer a correspondence between an item and a user need.

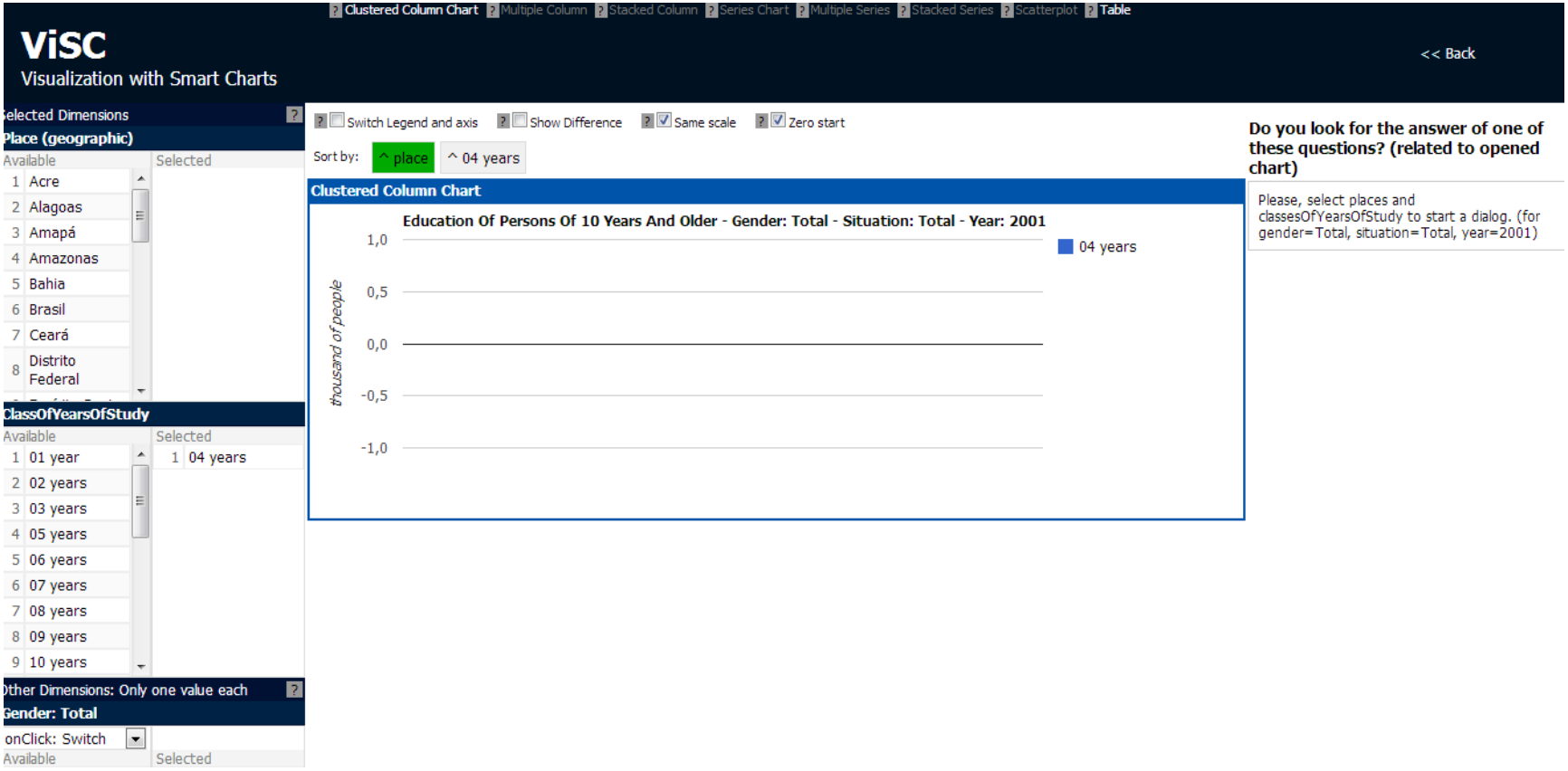
- Items: Efficient visualizations
- User need: Answer his question

The interface

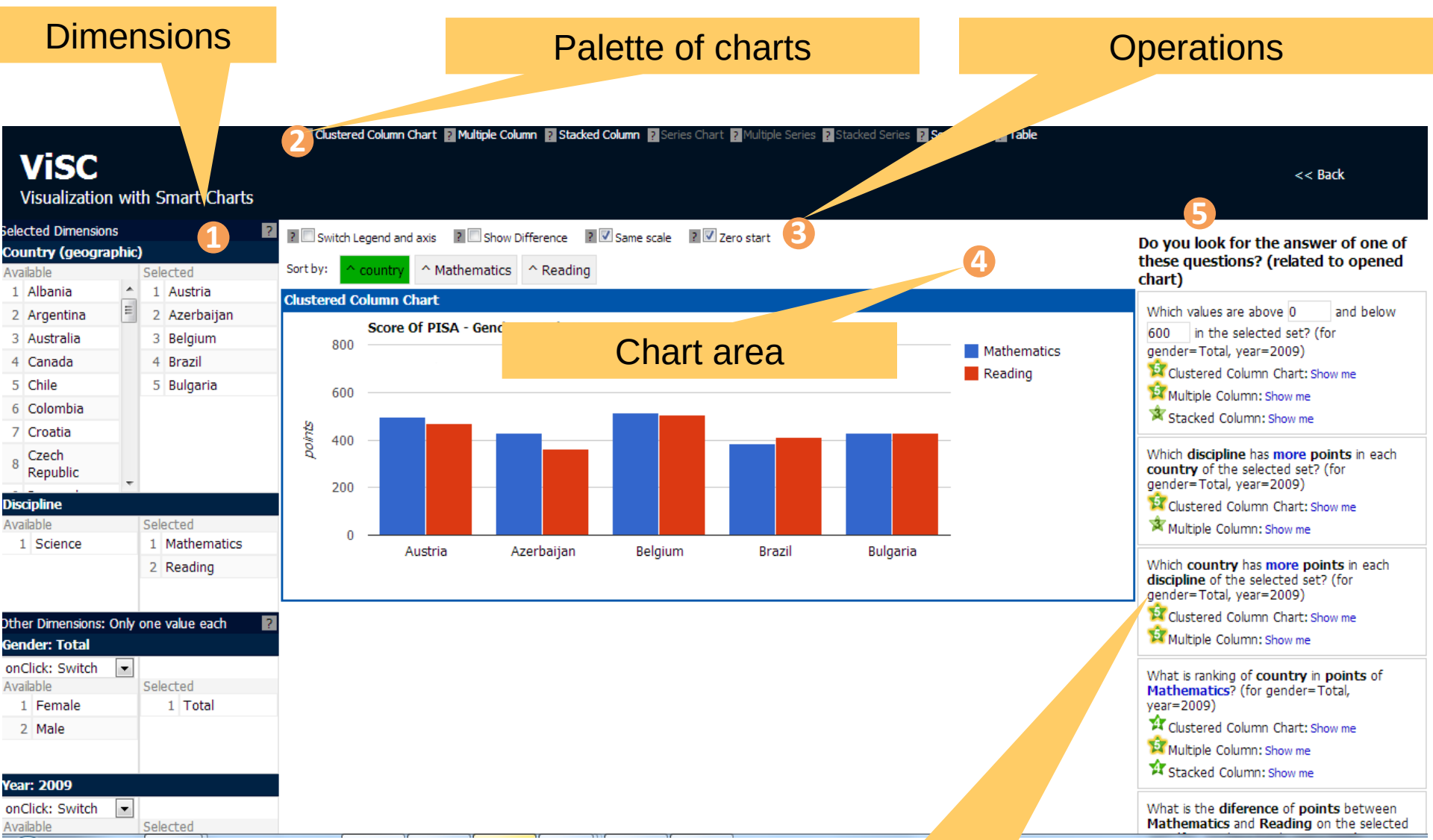
Selection of theme and 2 dimensions

Available visualizations

The interface



The ViSC tool



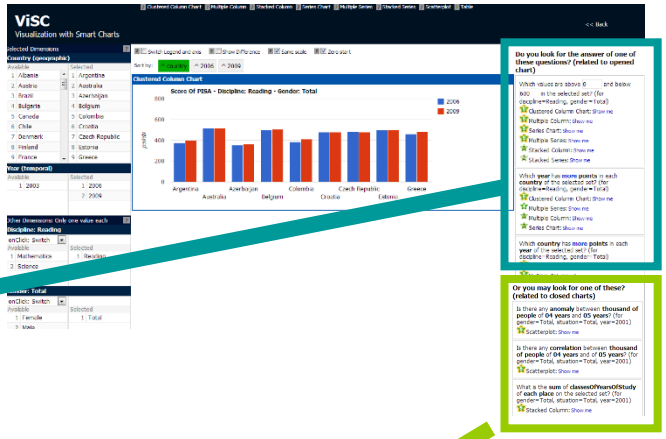
Interaction with questions

Do you look for the answer of one of these questions? (related to opened chart)

Which values are above 0 and below 700 in the selected set? (for gender=Total, situation=Total, year=2001)
- Clustered Column Chart: Show me
- Multiple Column: Show me
- Stacked Column: Show me

Which **classOfYearsOfStudy** has **more thousand of people** in each **place** of the selected set? (for gender=Total, situation=Total, year=2001)
- Clustered Column Chart: Show me
- Multiple Column: Show me

Which **place** has **more thousand of people** in each **classOfYearsOfStudy** of the selected set? (for gender=Total, situation=Total, year=2001)
- Clustered Column Chart: Show me
- Multiple Column: Show me



Or you may look for one of these? (related to closed charts)

Is there any **anomaly** between **thousand of people** of **04 years** and **05 years**? (for gender=Total, situation=Total, year=2001)
- Scatterplot: Show me

Is there any **correlation** between **thousand of people** of **04 years** and of **05 years**? (for gender=Total, situation=Total, year=2001)
- Scatterplot: Show me

What is the **sum** of **classesOfYearsOfStudy** of **each place** on the selected set? (for gender=Total, situation=Total, year=2001)
- Stacked Column: Show me

Expected contributions

- Develop interactive solution for visualization construction by novice users;

Secondary contributions

- Indirect evaluation of the visualization ontology
- Motivation potential in learning by analysis

Evaluation

How do the related questions influence the task performance and the generated visualizations?

Methods

- Semiotic Inspection Method - **SIM**
- Retrospective Communicability Evaluation - **RCE** (Retrospective Think Aloud + Tagging from Communicability Evaluation Method)

Preparation

- User profile: 6 undergraduate or master's degrees students from exact science areas at PUC
- Selected tools: ViSC and Tableau

Users	**Task 1**	**Task 2**
Odd (group 1)	ViSC	Tableau Public
Even (group 2)	Tableau Public	ViSC

Results

Well understood parts of the **ViSC** metamensage

- Include values and select the form of visualization.
- Visualization recommendations based on questions.
- *You only need to find the question and analyse one or more recomended visualizations.*

Results

Well understood parts of the **Tableau** metamensage

- *You can do math operations*
- *You can change visualization preferences*
- *Shows when charts are active or inactive*
- *Transform data in accordance to the selected visualization*

Results

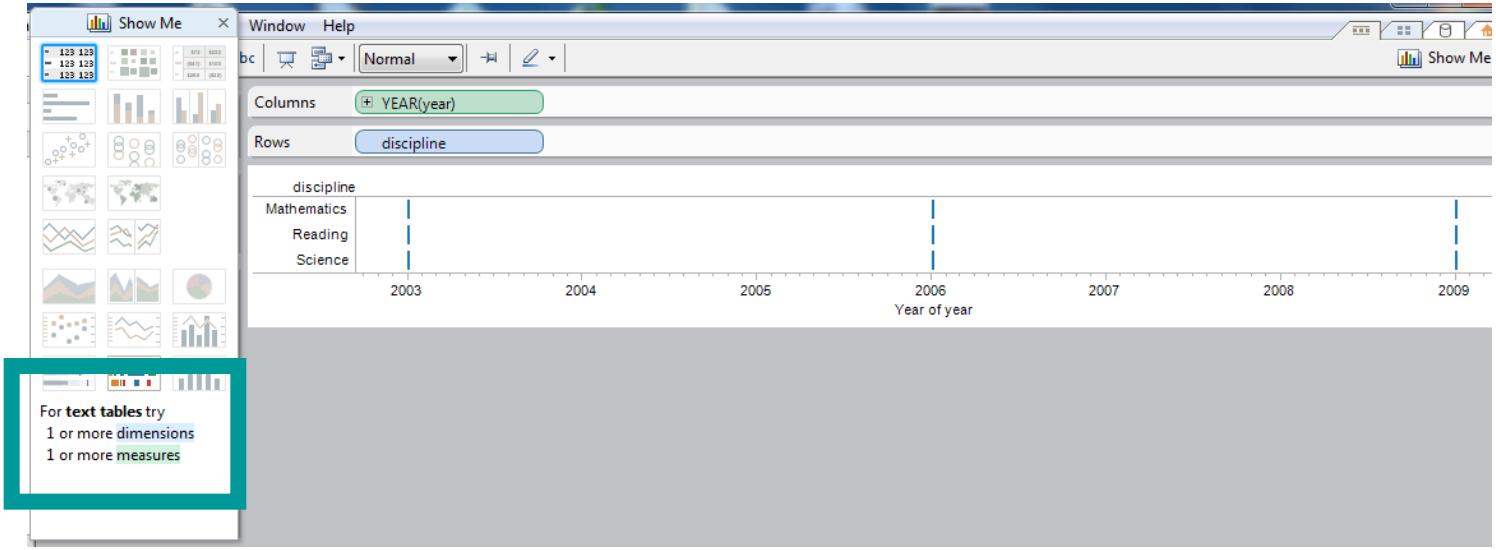
Not understood parts of **ViSC** metamensage

- Recommendations were classified by score

Results

Not understood parts of **Tableau** metamensage

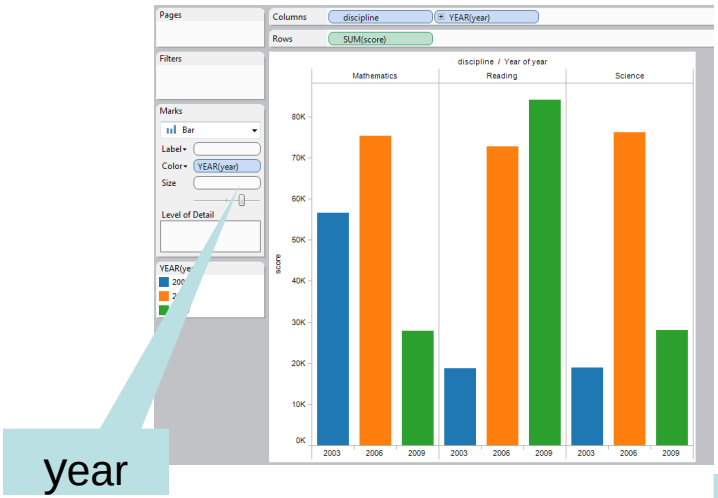
- Explains why each visualization is active or inactive

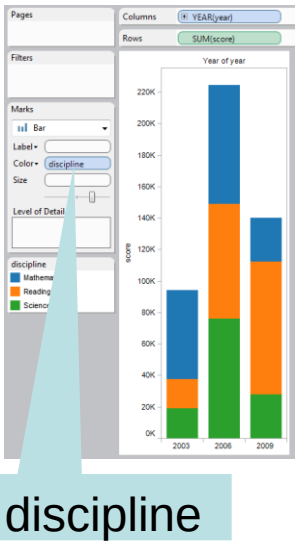


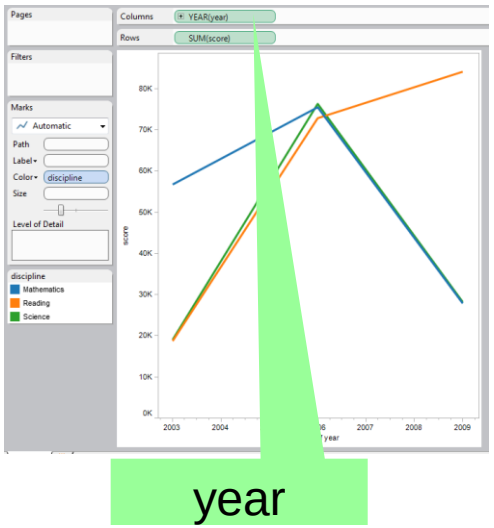
Results

Not understood parts of **Tableau** metamensage

- Interaction and chart changes can change variable features.







Results

“(...) I decided to see if there was any questions that could help me. And I found!" (U03)

“ (...) I was looking for something better to improve this chart or to put all bars together in a single color. (...) I found exactly what I had done. It was already there." (U05)

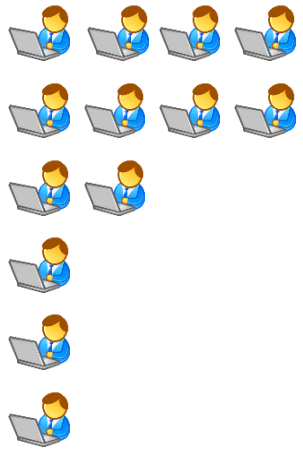
“(...) It might have the question I want to answer (...) I selected the questions and then I changed to “sum” and I found the correct chart. (U07)

“I looked at straight to the questions seeking for something to help me. I clicked on this question (...) but the chart (...) was really bad..” (U04)

ViSC

About the questions

- The questions had an important influence on the results.
- Users understood how they were generated
- Score was not observed
- A user did not read the questions
- A user did not use the questions
- The questions helped to find problems in the ontology



HCI problems

Tableau

About the window Show me

- Thumbnails helped user to select charts 
- Explanation were not observed 

Problems in undestanding some concepts

- Dimension X measure 
- General view 

Evaluation

How do the related questions influence the task performance and the generated visualizations?

- *Helped*
  - *Quick answers*
  - *New visualizations*
  - *Check with previous answer*
- *Did not influence*
  - *Not used*
- *Misled*
  - *Generate inefficient chart*

Conclusions

Goal

- Develop a solution to support novice users in chart construction with statistical data

Solution

- ViSC with recommender system through common questions

Evaluation

- Questions were efficient solutions to support chart construction by novice users

Contributions

- interactive solution to visualization construction by novice users; **achieved**
- Indirect evaluation of the visualization ontology: **improvement is required**
- Potential in learning by analysis motivation: **new evaluation is recommended**

Future work

- Expansion of evaluation
  - New group of users
  - Classification of the ontology
  - Evaluate learning potential
- Correction and extension of the ontology
- Hybrid recommender systems

Thank you!