



# 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

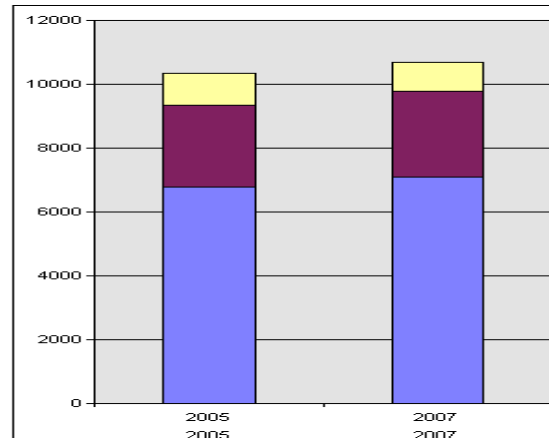
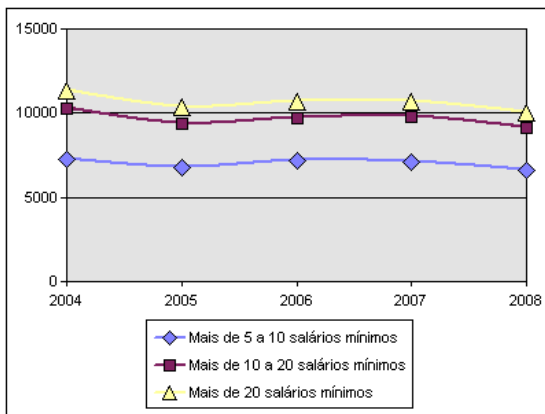


## Difficulty in the construction of efficient charts

“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?





*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



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



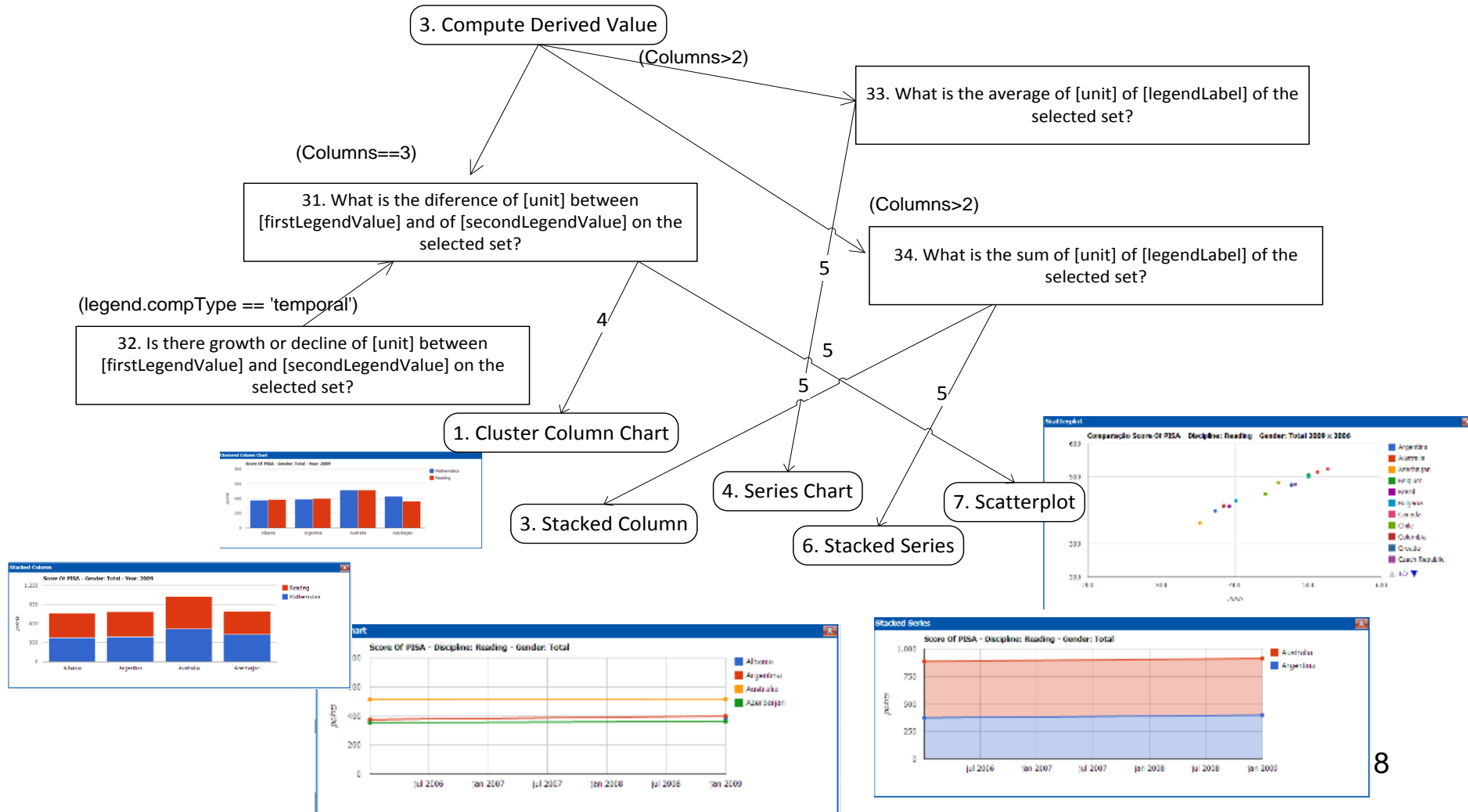
## The ViSC ontology

### **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 <b>items</b> . 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

### ViSC

Visualization with Smart Charts

ViSC is a visualization tool that provides charts through a smart way. You just need to select the theme and the two dimensions you may want to compare. Thus, through picking a question you want to answer, you will have efficient visualizations.

Theme:

Education of persons of 10 years and o

Do you want a visualization to compare these variables?

In horizontal Axis:

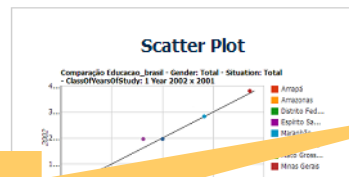
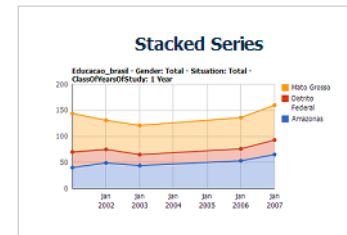
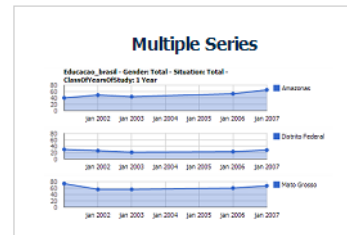
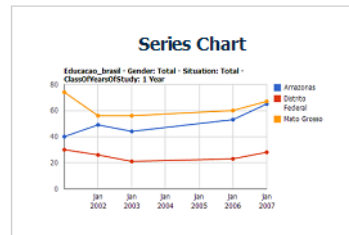
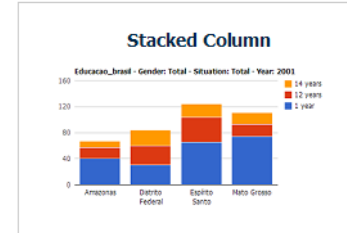
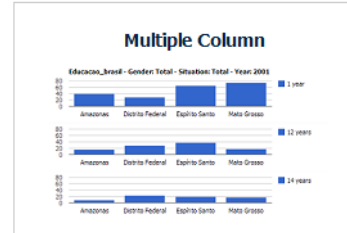
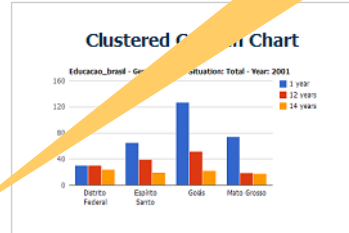
place (28)

In Legend:

classOfYearsOfStudy (19)

YES! Start ViSC now!

### Available visualization



**Table**

place	10 years	11 years	12 years
Alagoas	41	173	10
Amapá	21	87	3
Amazonas	78	325	17
Bahia	275	1211	49
Brasil	4621	19239	1756
Ceará	169	581	48
Distrito Federal	77	242	30

Available visualizations



## The interface

Clustered Column Chart | Multiple Column | Stacked Column | Series Chart | Multiple Series | Stacked Series | Scatterplot | Table

**ViSC**  
Visualization with Smart Charts

<< Back

Selected Dimensions

**Place (geographic)**

Available Selected

- 1 Acre
- 2 Alagoas
- 3 Amapá
- 4 Amazonas
- 5 Bahia
- 6 Brasil
- 7 Ceará
- 8 Distrito Federal

**ClassOfYearsOfStudy**

Available Selected

- 1 01 year
- 2 02 years
- 3 03 years
- 4 05 years
- 5 06 years
- 6 07 years
- 7 08 years
- 8 09 years
- 9 10 years

Other Dimensions: Only one value each

**Gender: Total**

onClick: Switch

Available Selected

Switch Legend and axis Show Difference Same scale Zero start

Sort by: place 04 years

**Clustered Column Chart**

Education Of Persons Of 10 Years And Older - Gender: Total - Situation: Total - Year: 2001

thousand of people

1,0

0,5

0,0

-0,5

-1,0

04 years

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

Please, select places and classesOfYearsOfStudy to start a dialog. (for gender=Total, situation=Total, year=2001)



# The ViSC tool

Dimensions

Palette of charts

Operations

**ViSC**  
Visualization with Smart Charts

Selected Dimensions  
**Country (geographic)**

Available	Selected
1 Albania	1 Austria
2 Argentina	2 Azerbaijan
3 Australia	3 Belgium
4 Canada	4 Brazil
5 Chile	5 Bulgaria
6 Colombia	
7 Croatia	
8 Czech Republic	

**Discipline**

Available	Selected
1 Science	1 Mathematics
	2 Reading

Other Dimensions: Only one value each  
**Gender: Total**

onClick: Switch	Selected
1 Female	1 Total
2 Male	

**Year: 2009**

onClick: Switch	Selected

Clustered Column Chart ? Multiple Column ? Stacked Column ? Series Chart ? Multiple Series ? Stacked Series ? Table

Switch Legend and axis ? Show Difference ? Same scale ? Zero start

Sort by: country ^ Mathematics ^ Reading

**Chart area**

Score Of PISA - Gender

Country	Mathematics	Reading
Austria	500	480
Azerbaijan	430	380
Belgium	520	510
Brazil	400	420
Bulgaria	430	430

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

Which values are above 0 and below 600 in the selected set? (for gender=Total, year=2009)  
[Clustered Column Chart: Show me](#)  
[Multiple Column: Show me](#)  
[Stacked Column: Show me](#)

Which discipline has more points in each country of the selected set? (for gender=Total, year=2009)  
[Clustered Column Chart: Show me](#)  
[Multiple Column: Show me](#)

Which country has more points in each discipline of the selected set? (for gender=Total, year=2009)  
[Clustered Column Chart: Show me](#)  
[Multiple Column: Show me](#)

What is ranking of country in points of Mathematics? (for gender=Total, year=2009)  
[Clustered Column Chart: Show me](#)  
[Multiple Column: Show me](#)  
[Stacked Column: Show me](#)

What is the difference of points between Mathematics and Reading on the selected

Recommendations



## 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)

- Clustering 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)

- Clustering 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)

- Clustering 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)

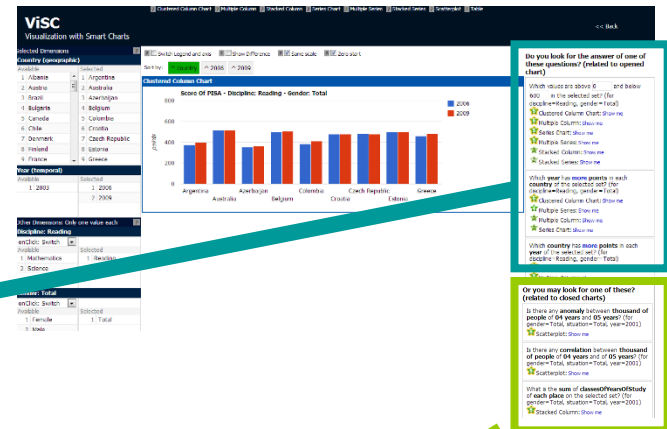
- Scatterplot: Show me

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



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

<b>Users</b>	<b>Task 1</b>	<b>Task 2</b>
Odd (group 1)	ViSC	Tableau Public
Even (group 2)	Tableau Public	ViSC



## Results

Well understood parts of the **ViSC** metamessage

- 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 recommended visualizations.*





## Results

Well understood parts of the **Tableau** metamodel

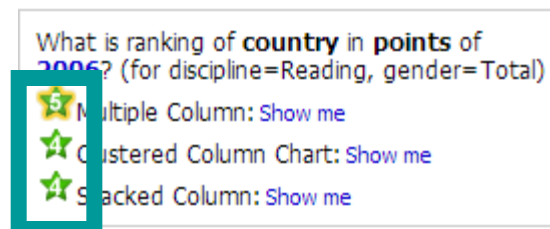
- *You can do math operations*
- *You can change visualization preferences*
- *Shows when charts are active or inactive*
- *Transform data in accordance to the selected visualization*



## Results

**Not** understood parts of **ViSC** metamessage

- Recommendations were classified by score

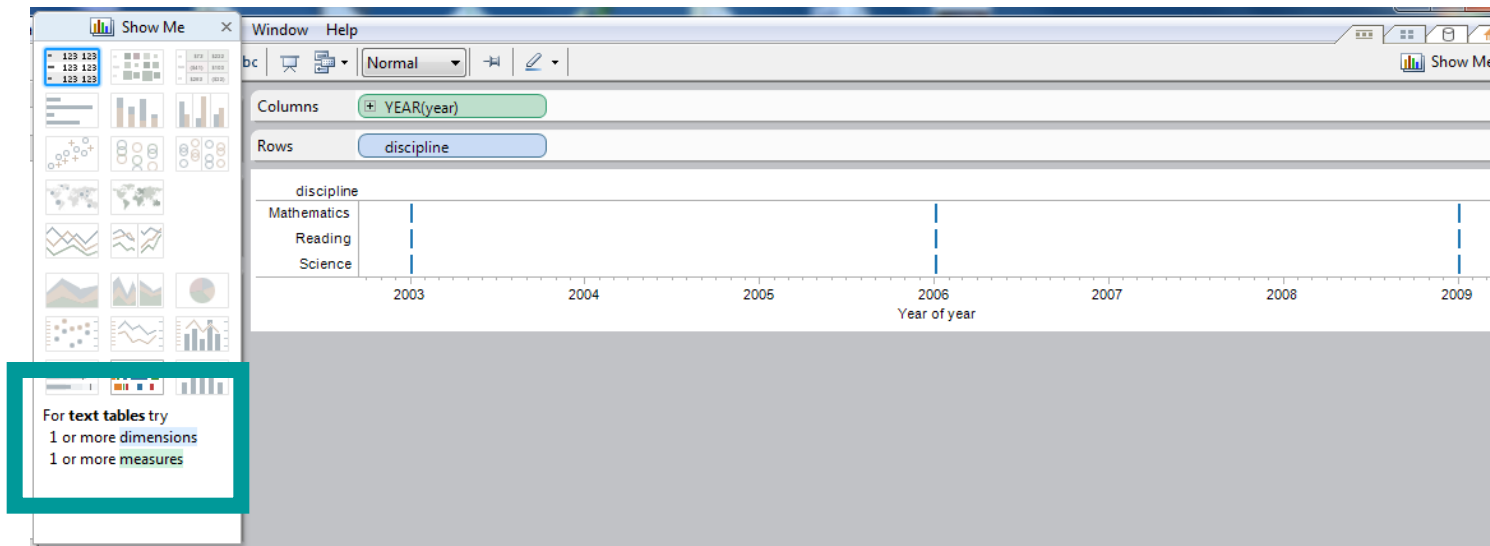




## Results

**Not** understood parts of **Tableau** metamessage

- Explains why each visualization is active or inactive

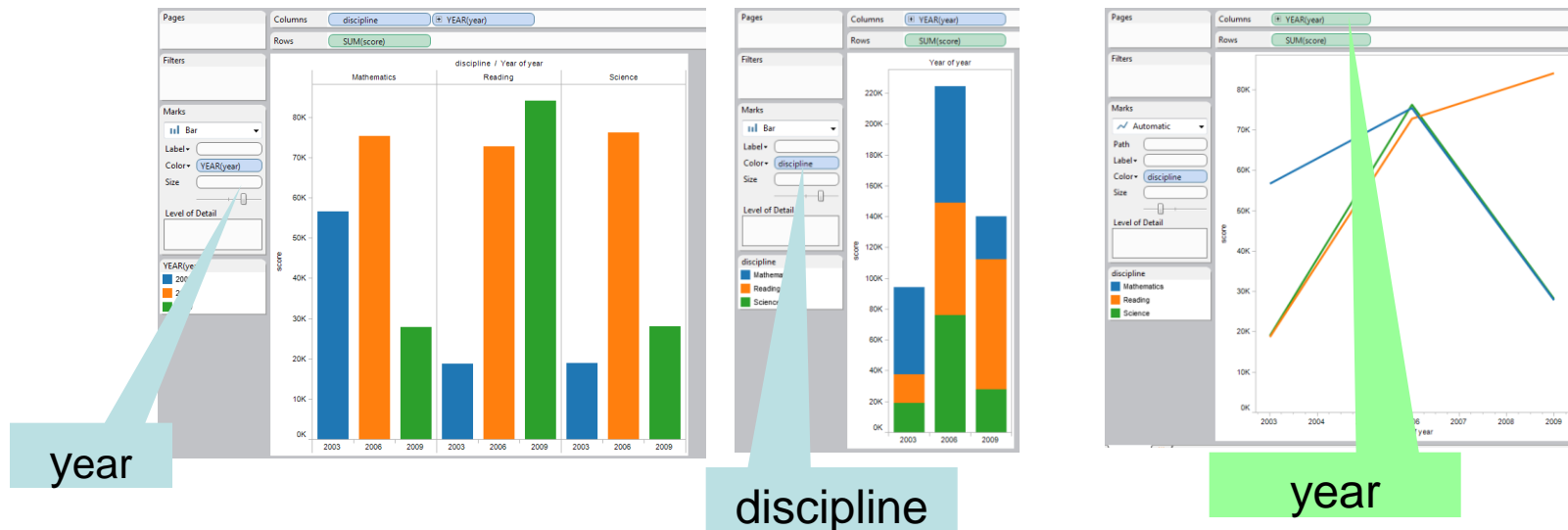




## Results

**Not** understood parts of **Tableau** metamessage

- 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)*

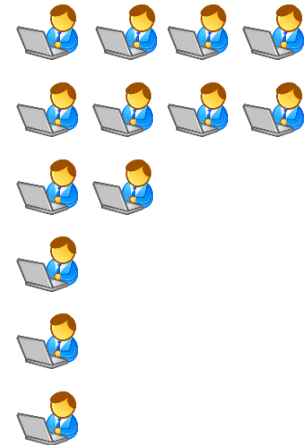
*“I looked at straight to the questions seeking for something to help me. I clicked on this question (...) but the chart (...) was really bad..” (U04)*



## 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 
- Explanations were not observed    

### Problems in understanding some concepts

- Dimension X measure    
- General view    



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*





## 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!**