

QualityFlow

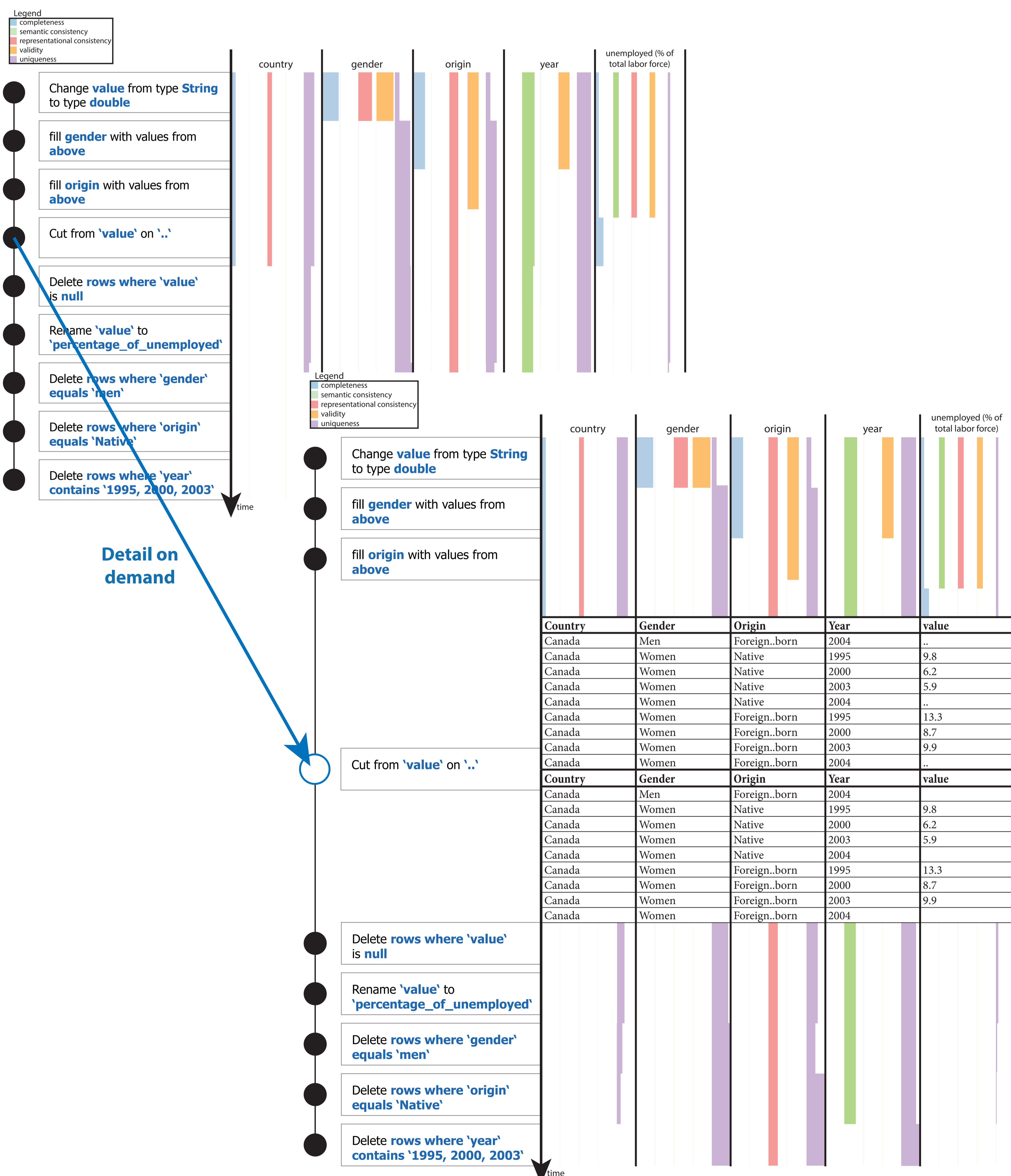
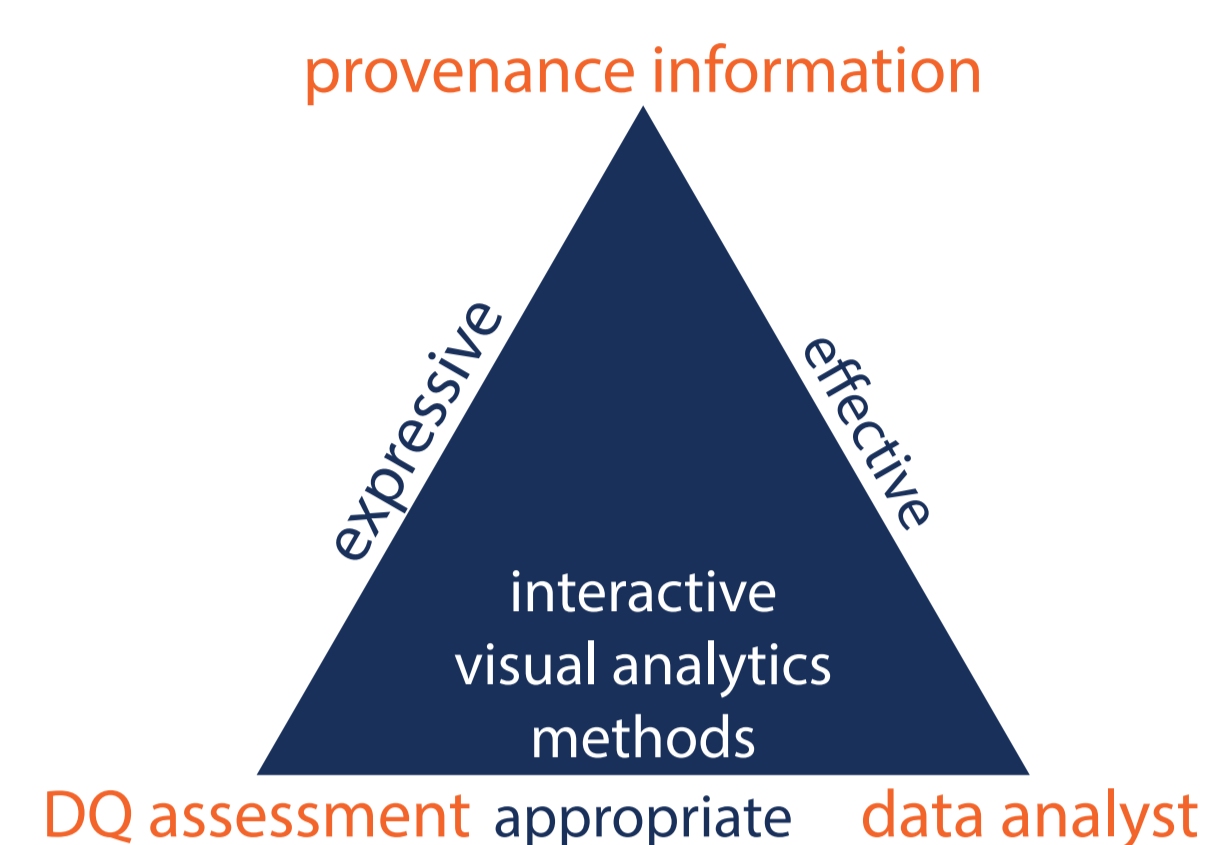
Provenance Generation from Data Quality

Christian Bors, Theresia Gschwandtner, Silvia Miksch

'What has been done to a data set?'

Data Quality assessment widely disregards means for providing overview of already performed actions on a given data set.

Re-tracing collaborators' activity often leaves the analyst with scattered evidence and requires him/her to draw conclusions on the thought process behind transformations performed on the data set.



Tasks

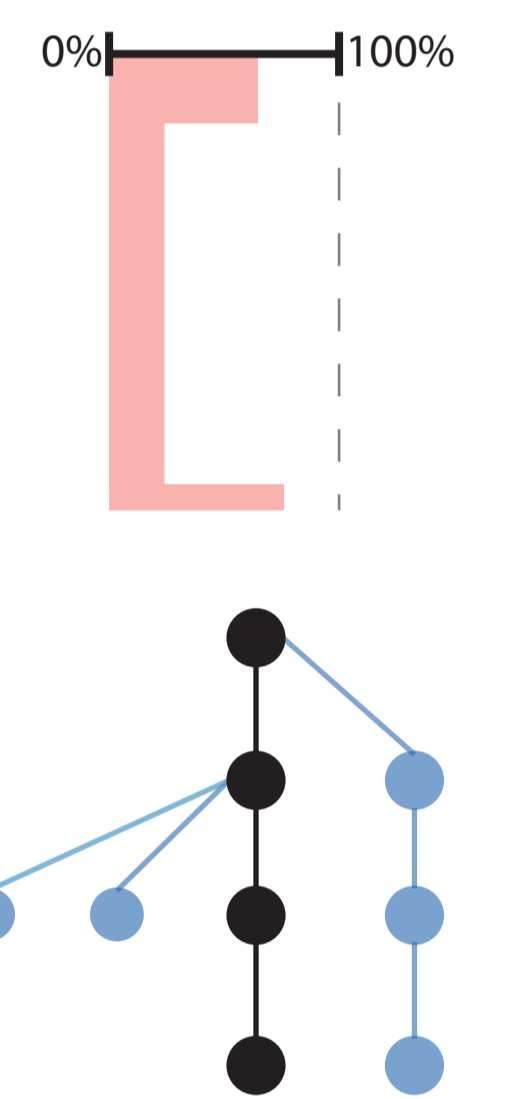
- Understanding the impact of operations on the data's quality and structural changes
- Assisting the sensemaking process
- Helping data analysts in coming back to a priory wrangled or transformed data set
- Understanding the rationale behind their own or others' actions
- Notifying collaborators of idiosyncratic or interesting occurrences of particular data sets
- Determining errors in users' operations due to an increase of quality issues in the data

Generate Provenance

- 1 Information from:
 - Data pre-processing
 - Data Quality metrics
- 2 Combine provenance to:
 - Support investigation of the data's history and quality
- 3 Support collaboration:
 - Visual transformation history
 - Means for annotation

Design Rationales

- Multiple Data Quality metrics for each data column
- Metric width equals to amount of quality issues (in percent)
- Graph of transformation actions
- Detail view: Combine raw tabular data view with Data Quality metrics
- Structural transformation indicator



Contributions

- Guide attention to comprehensive information about transform actions executed on the data
- Users can notice conspicuous and confirm valid actions
- Convey information between collaborators, through annotations alongside respective data

Future Work

- Create means to customize and specify metrics
- Address scalability issues for large tabular data
- Log users' activity during exploration and trace their learning process



Vienna University of Technology, Faculty for Informatics, Information Engineering Group



Laura Bassi Centre of Expertise
Centre for Visual Analytics
Science & Technology