The Turnip: Visualizing Statistical Data Cleaning

The Approach

Data Viewer + Data Visualization + Statistical analyzer + Outlier Detector

The Results

I Present The Turnip:

The Motivation

Current tools complicated
Big, made for the enterprise
Not very visual
Time consuming
Require advanced knowledge of:
Database management
Statistical theories

Future Work

Multiple Columns in visualizations
Cleaning up the UI
Effects
New Statistical Models
New Data Domains

Data cleaning usually not visual
Database tools not visual
Hard to apply statistical theories to databases
SQL syntax complex for large calculation

Author: Ragnar Skulason