Visual Data Science in Practice

Johanna Schmidt

VRVis Forschungs GmbH
Vienna, Austria
VRVis – Research Center for Virtual Reality and Visualization

Founded in 2000

Located in Ares Tower, 1220 Vienna

Staff approx. 70 researchers

**Mission:** application-oriented research in Visual Computing for linking science to industry

**Research focus:** Computer graphics, Scientific Visualization, Visual Analytics

Close cooperation with national and international universities and companies
The Visual Analytics Group at VRVis – ... and Projects

Energy sector

Industrial manufacturing

Automotive engineering

Healthcare planning

... and others ...
Focus of this Talk

Energy sector

Industrial manufacturing

Automotive engineering

Healthcare planning

... and others ...
Visual Analytics

... „combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets”

[D. Keim, J. Kohlhammer, G. Ellis, F. Mansmann. "Mastering the information age: Solving problems with visual analytics". VisMaster, 2010.]

Our focus: high-dimensional and time-dependent data
Recurring Questions ...

Is the data quality sufficient for correct interpretation?
Does the data reveal unexpected trends and relationships?
Does the data enable to build (predictive) models?
How to use the data for efficient decision making?
How to effectively communicate insights and decisions?
... Resulting (High-Level) Tasks

Data Quality Assessment and Cleansing
   Is the data quality sufficient for correct interpretation?

Exploratory Data Analysis
   Does the data reveal unexpected trends and relationships?

Model Building and Validation
   Does the data enable to build (predictive) models?

Multi-Criteria Decision Making
   How to use the data for efficient decision making?

Reporting
   How to effectively communicate insights and decisions?
Data Science Lifecycle

01 BUSINESS UNDERSTANDING
Ask relevant questions and define objectives for the problem that needs to be tackled.

02 DATA MINING
Gather and scrape the data necessary for the project.

03 DATA CLEANING
Fix the inconsistencies within the data and handle the missing values.

04 DATA EXPLORATION
Form hypotheses about your defined problem by visually analyzing the data.

05 FEATURE ENGINEERING
Select important features and construct more meaningful organizing the raw data that you have.

06 PREDICTIVE MODELING
Train machine learning models, evaluate their performance, and use them to make predictions.

07 DATA VISUALIZATION
Communicate the findings with key stakeholders using plans and interactive visualizations.

DATA SCIENCE LIFECYCLE
sudop.co
Data Science Lifecycle

01 BUSINESS UNDERSTANDING
Ask relevant questions and define objectives for the problem that needs to be tackled.

02 DATA MINING
Gather and scope the data necessary for the project.

03 DATA CLEANING
Fix the inconsistencies within the data and handle the missing values.

04 DATA EXPLORATION
Form hypotheses about your defined problem by visually analyzing the data.

05 FEATURE ENGINEERING
Select important features and construct more meaningful conclusions using the raw data that you have.

06 PREDICTIVE MODELING
Train machine learning models, evaluate their performance, and use them to make predictions.

07 DATA VISUALIZATION
Communicate the findings with key stakeholders using plots and interactive visualizations.

DATA SCIENCE LIFECYCLE
sadrep.co
Our approach: Visualization Dashboards

- Based on C++ Visual Analytics Platform “Visplore” (12+ years)
- Multiple coordinated views
- Linking and brushing
- High performance for large data
- Task-tailored entry points
- Integrated in Data Science Environments (Python, Matlab,..)
Energy Example 1: Data Quality Assessment of Time Series Data

Goal: Overview of data quality problems for > 100 time series

Energy Example 2: Exploratory Analysis of Time Series Data

Goal: Detection of structural anomalies and relationships
Energy Example 3: Feature Selection for Regression

Goal: Regression modeling of wind power production

Energy Example 4: Detection of Structural Breaks

Goal: Explain why automated models show bad performance
Energy Example 5: Model Selection

Goal: Identify best decision trees under trade-offs
Industry Example: Analysis of Product Quality Indicators

Goal: Detect and explain quality problems in manufacturing
Engineering Example 1: Exploratory Analysis of Simulation ensembles

Goal: Qualitative validation of complex simulation results

Engineering Example 2: Multi-Objective Optimization

Goal: Identify engine settings meeting conflicting objectives

Integration in Data Science Environments

Goal: Enable data science workflows based on familiar tools

- Rapid exploration
- Computational flexibility
- Interactive selection
- Libraries, modules
- Visual model tuning
- Familiarity
- Visual Debugging
- Reproducability
Our Software Platform – Visplore

Visualization system for exploratory data analysis

- Extensible software framework (C++, Python, OpenGL, Gtk+)
- Initially designed as expert tool
- Current focus as platform for creating task-specific dashboards
Some Lessons Learned

In general positive attitude towards visualization

- But: for most people, visualization = Excel and static reports
- Visual analysis far less known and still a bit “suspicious”
- Visual literacy in general quite limited
Some Lessons Learned

The unknown may be perceived as evil rather than good

- Be careful with unfamiliar visualization techniques
- Be even more careful with complex visualization techniques
- Real users have very limited time for familiarization
- Domain conventions often beat perceptual superiority
Some Lessons Learned

Acceptance requires integration in user work flow

- Straightforward data import and result export
- Real analysts require access to Matlab / Python / R
- Most users prefer dashboards rather than build-it-yourself kits
- Automation and guidance is appreciated
Interested ... ?

Internship / diploma thesis

- Interesting topics to address real tasks of real users on real data
- Deepen know-how in C++, software engineering, visualization, ...
- Tight collaboration with Visplore team
- Payment
Thank you for your attention!

https://www.vrvis.at/research/visual-analytics-group/
Johanna.schmidt@vrvis.at