Visual Sensemaking of Big Data
Data Science Lecture, Tuesday September 20, 2016

dr.ir. Stef van den Elzen
Visualization Architect
stef.van.den.elzen@synerscope.com
@StefvandenElzen
Introduction – Stef

2011  **ir. degree** – *Visual Analytics*  
*TU/e* Computer Science

2011-2015  **dr. degree** – Interactive Visualization of Dynamic Multivariate Networks  
*TU/e* + *SynerScope B.V.*

Prof.dr.ir. **Jack** van Wijk
Introduction

Dr. Stef van den Eerenbeeck - 2011-2015 - TU/e

Dr. Jack van Wijk - 2015-present - SynerScope B.V.

Prof. dr. ir. Jack van Wijk
Introduction – Stef

2011  
**ir. degree** – *Visual Analytics*  
TU/e Computer Science

2011-2015  
**dr. degree** – Interactive Visualization of Dynamic Multivariate Networks  
TU/e + SynerScope B.V.

2015-present  
**Visualization Architect**  
SynerScope B.V.

Prof.dr.ir. **Jack** van Wijk
Introduction – SynerScope

- TU/e visualization group spin-off, 2011
- PhD – Visualization of graphs and trees for Software Visualization

dr.ir. Danny Holten
Introduction – SynerScope

Apply PhD techniques to other data?

Of course!

dr.ir. Danny Holten

Jan-Kees Buenen, MBA
Introduction – SynerScope

• TU/e visualization group spin-off, 2011
• 20 people

From raw complex data → clear visual order for fast sensemaking

Main application areas:
• Insurance
• Cyber security
• Critical infrastructure
Visual Sensemaking of Big Data
Big Data
Big Data

- Not just about numbers
- Mix of structured and unstructured data
- How to store? Hadoop?
Big Data

- Just storage.
- ‘Data lake’ soon becomes data swamp
Dark Data

“The information assets organizations collect, process and store during regular business activities, but generally fail to use for other purposes.”

– Gartner
Big Data

Main challenges:

- Discovery
  - How to extract interesting data from data lake?

- Analysis
  - How to analyze this data?
Main challenges:

**Discovery**

- How to extract interesting data from data lake?

**Analysis**

- How to analyze this data?
Discovery
Discovery

Challenges:

• Data silos
• No overview
• Lack of structure
• Lack of metadata

No mechanism to effectively search, explore, and select for further analysis.
Discovery – Legato

Self-service data navigation solution on top of Apache Hadoop.

Provides *business users* the capabilities to quickly navigate, search, link and improve *structured* and *unstructured* data.
Discovery – Legato

• Gain an overview of your data
Discovery – Legato

- Gain an overview of your data
- Search your data
Discovery – Legato

• Gain an overview of your data
• Search your data
• Link data from multiple sources together
Discovery – Legato

• Gain an overview of your data
• Search your data
• Link data from multiple sources together
• Improve data using external data sources
• Automate & standardize data processing
• Govern your data. Not everyone needs all data
• Collaborate with others
Discovery – Legato

- Gain an overview of your data
- Search your data
- Link data from multiple sources together
- Improve data using external data sources
- Automate and standardize data processing
- Govern your data. Not everyone needs all data
- Collaborate
- Harness power of Apache Hadoop & Spark
Discovery – Legato

• Gain an overview of your data
• Search your data
• Link data from multiple sources together
• Improve data
• Automate & standardize data processing
• Govern your data. Not everyone needs all data
• Collaborate
• Harness power of Apache Hadoop & Spark
• Export to Marcato, Notebook and third-party software
Discovery – Legato

- Gain an overview of your data
- Search your data
- Link data from multiple sources together
- Improve data using external data sources
- Automate & standardize data processing
- Govern your data. Not everyone needs all data
- Collaborate with others
- Harness power of Apache Hadoop & Spark
- Export to Marcato, Notebook and third-party software
SynerScope Legato

Demo
Sensemaking
Analysis

• 80-90% is unstructured

• Lack of infrastructure and tools for analysis

• **Business users** unable to ask questions but need help from **data scientists** or from **IT**

• Analysis only happens in isolation – per data type
Analysis

Towards a one-interface technology
Analysis

Towards a one-interface technology
Analysis – Sensemaking

• Traditional data analysis

• Models
• Machine Learning
• Data Mining
• Automated detection
• Predictive analytics
• Deep-learning

• **No** direct *Sense making*
• Used as *end result*
• Part of *complex tool-chain*
• Hard to interpret
Analysis – Sensemaking

Domain Expert
Formulate questions

Data scientist
Configuration – trial & error

Input
Output

Interpret output
Visual sensemaking

• Combination of visualization, automated methods & human interaction.
Why visualization

- Exploit human perception and cognition.
- Reduce search time
- Enhance recognition of patterns
- Perceptually monitor large number of events
- Provide an interactive medium that enables exploration
Why visualization

Ideal if one needs “insight” and does not know what to look for (exactly)
Visualization – dashboards

- No real visual data exploration
- Current situation only
- Single values
- No explanation
- No exploration of additional variables
Visualizations should

• Fit the task at hand
• Interaction should play a key-role
• Not need any explanation
• Be scalable
• Guide the user in the exploration
• Avoid unnecessary visual elements
• Not rely on human memory
• Have an automatic support mechanism for explanation and presentation
• Be consistent (visual variables & interaction metaphors)
• Provide a context
Visualization – All dimensions
Visual Analytics

“The science of analytical reasoning facilitated by interactive visual interfaces.” [Thomas and Cook, 2005]
Visual Analytics – Multi-level exploration

• Result features → visual exploration

• Combine with external data to create context
  • Structured: numbers, temporal, network, geospatial
  • Unstructured: text, images, speech

• Explore model parameters
  • Influence on outcome
  • Comparison of different models
  • Improve model performance
Visual sensemaking

- Effortless machine assisted visual sensemaking
- New class of users
Visual analytics

Demo
Visual analytics

Demo
Data Landscape

SynerScope Appliance
- Legato
- Marcato

SynerScope Front-End
- Legato UI
- Marcato UI

Data

Control

Scalable DBs
GPU Processing

Users / Analysts
Practical solutions – how it is done

• Backend leverages **existing frameworks** (theano, tensorflow, caffe, torch) through **apache Spark** as an orchestrator.

• Use **GRID** technology to provide low-latency **OpenGL** rendering of the UI.

• Use interactive **data linkage** and **user interaction** to set training goals, closing the learning loop for discovery.

• Currently moving to cloud (fra.me, amazon)
Take aways

• **Combine** the strengths of human decision making, visualization and GPU-driven **machine learning**

• **Bridging the gap** between **data-scientist** and **business user**
Take aways – Visual sensemaking

• **Generic** solutions

• Simple **interaction** / querying mechanisms

• Deal with large **volume** of data with
  • **visual analytics**, augmentation

• Deal with large **variety** of data
  • **Interlinked views**

• System needs to support **speed of thinking**
Thank you!

Visual Sensemaking of Big Data

www.synerscope.com

(Yes, we are hiring)

dr.ir. Stef van den Elzen
Visualization Architect
stef.van.den.elzen@synerscope.com

@StefvandenElzen

(Yes, we are hiring)