FACULTY
OF INFORMATICS
Masaryk University

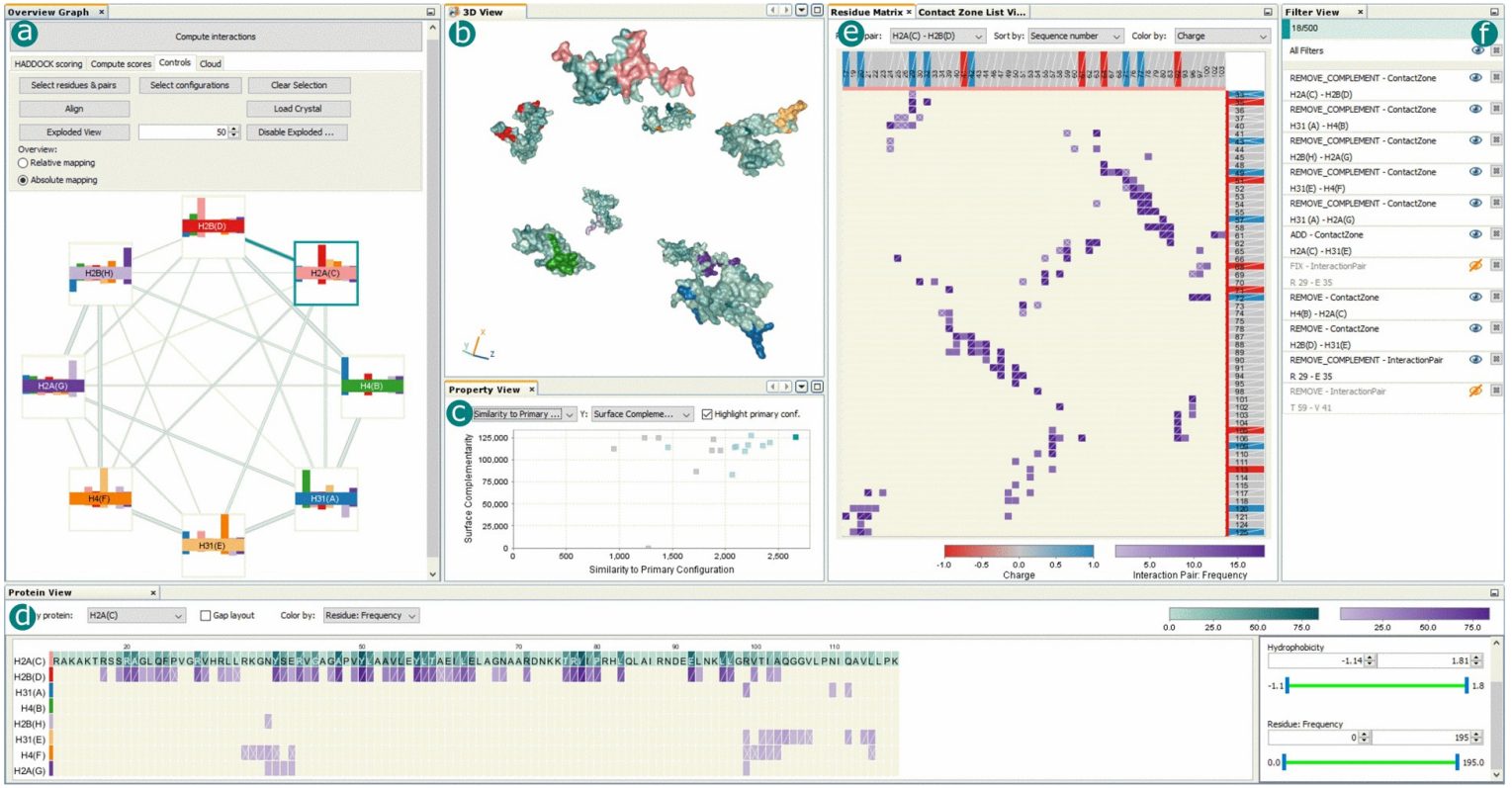
Data Visualization – Principles and Challenges

Radek Ošlejšek

Data Visualization

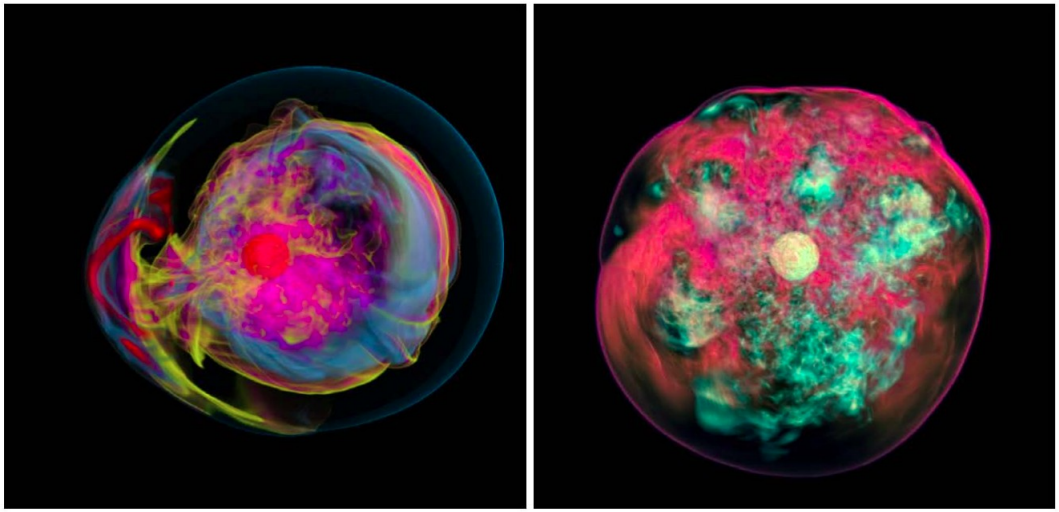
- Avoiding confusion: **IT IS NOT** about the design of GUI, e.g., where to place info window, what color to choose (although these UX aspects are part of any good graphical tool).
- Three main fields in visualization
  - Scientific visualization (SciVis)
  - Information visualization (InfoVis)
  - Visual analytics (VA)

[Furmanová et al.: Multiscale Visual Drilldown for the Analysis of Large Ensembles of Multi-Body Protein Complexes, 2019]



Scientific visualization (SciVis)

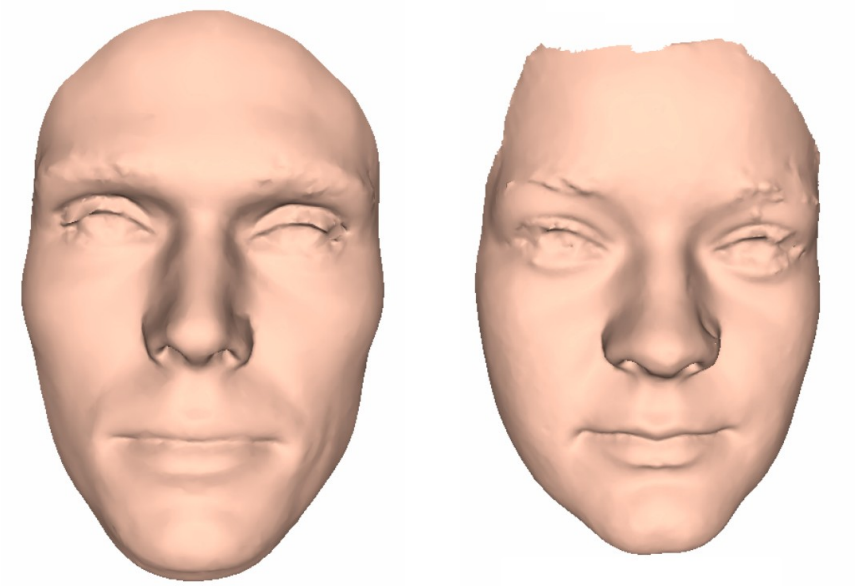
- **Goal:** To graphically illustrate scientific phenomena
  - Techniques to create a **renderable geometric model of the data**
  - Explanatory visualizations: Graphics representation is used for understanding and interpretation
  - Studied phenomenon is usually modeled by measurements
- Interdisciplinary: medical, molecular, flow, volumetric … visualization



A visual approach to illustrate the complex relationships within a Supernova
[Ma et al.: Scientific discovery through advanced visualization, 2005]

SciVis: Face Morpohology

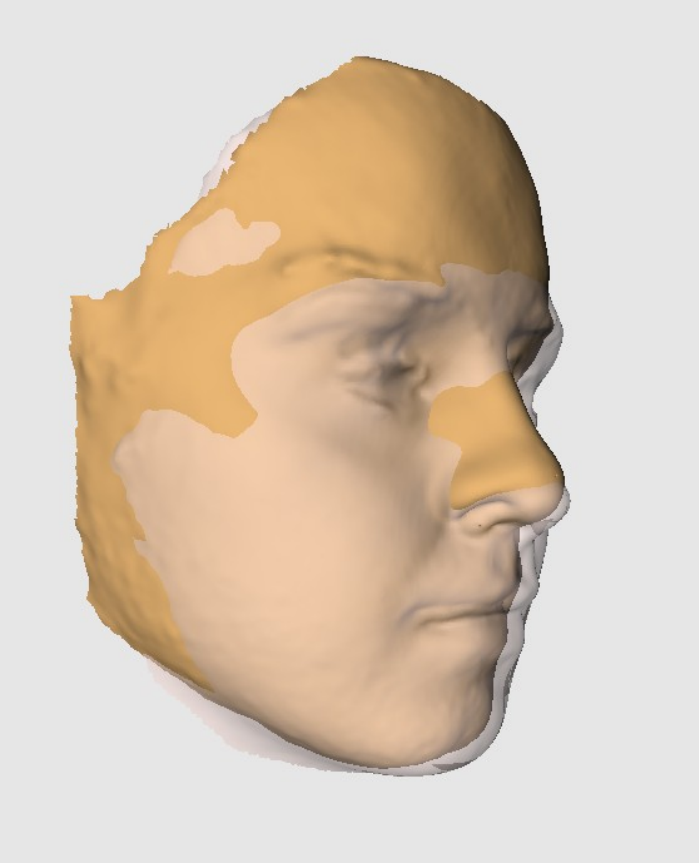
- **Research activities:**
  - Similarity of Human Faces in to cooperation with doc. Petra Urbanová, Department of Anthropology, Faculty of Science, MU.
  - FIDENTIS Analyst 2: https://www.radek-oslejsek.cz/it/fidentis-analyst-2/
    - For anthropologists, police, (plastic) surgeons
- **Exercise:** Given two 3D scans of human faces, what visualization techniques we can use to compare their (dis)similarity?



[Ferková, Z.: Comparison and Analysis of Multiple 3D Shapes, 2015]

SciVis: Similarity of Human Faces

- Approach: Overlapping with **transparency**



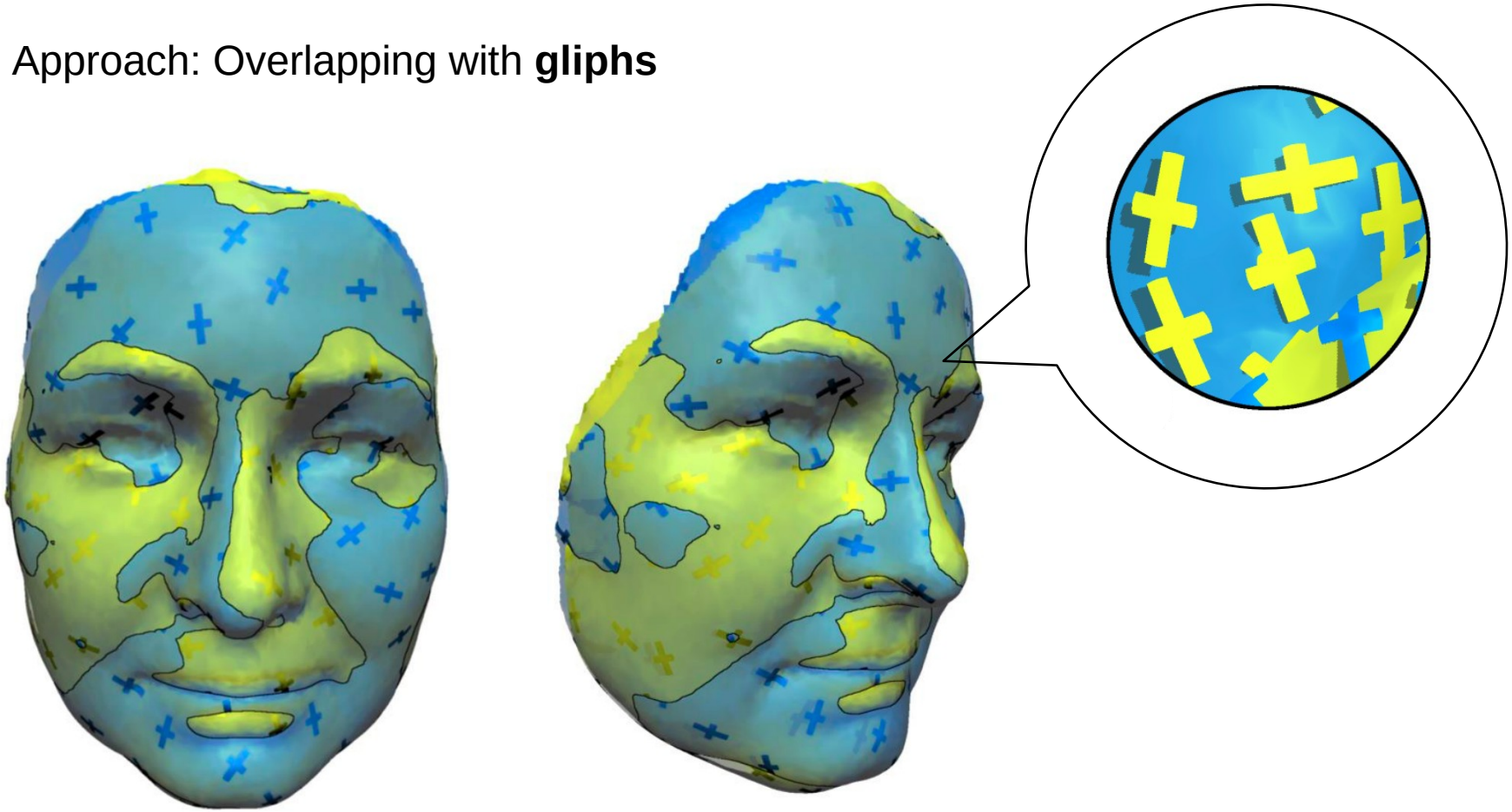
SciVis: Similarity of Human Faces

- Approach: Overlapping with **contours**

[Furmanová. K.: Visualization techniques for 3D facial comparison, 2015]

SciVis: Similarity of Human Faces

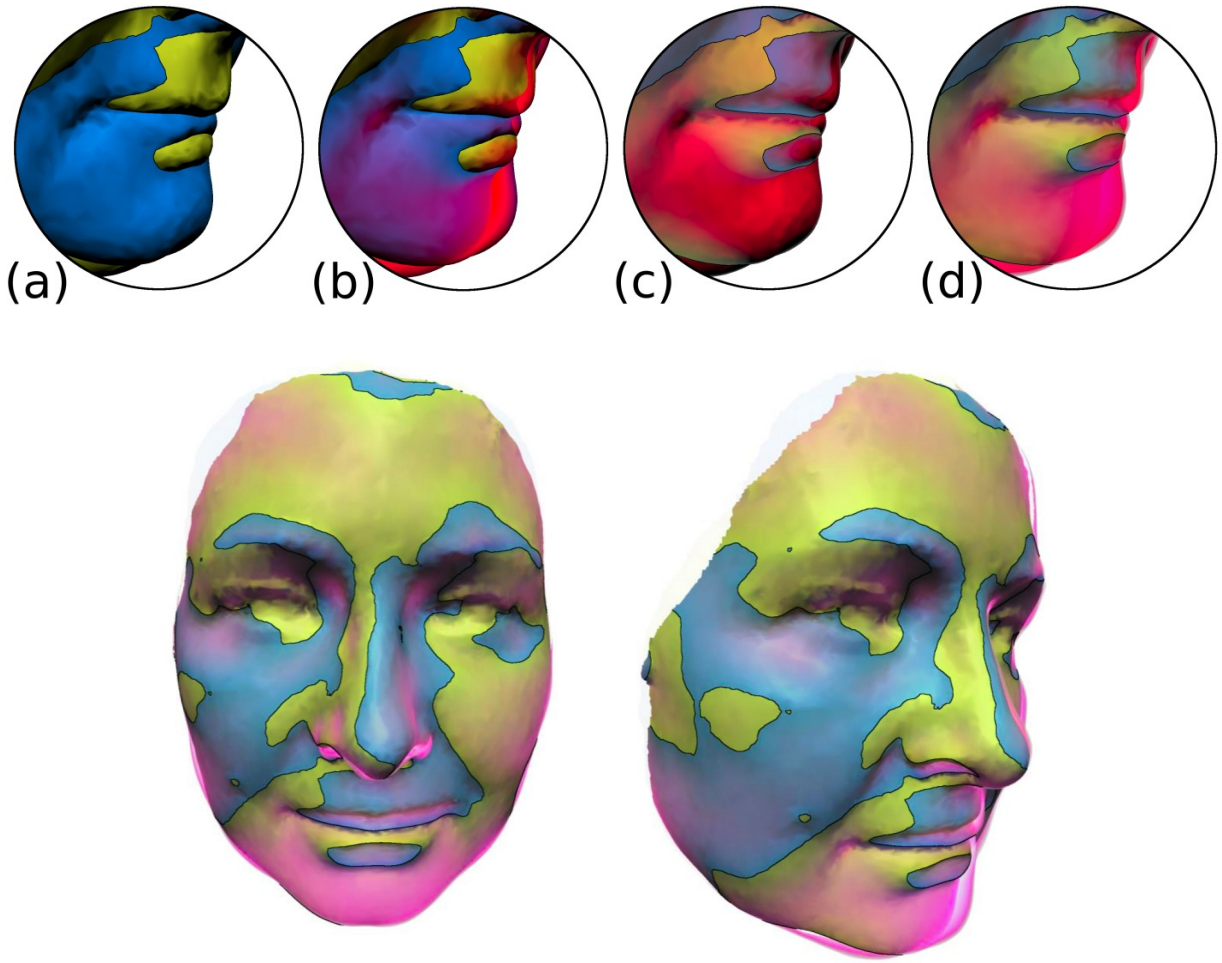
- Approach: Overlapping with **gliphs**



[Furmanová. K.: Visualization techniques for 3D facial comparison, 2015]

SciVis: Similarity of Human Faces

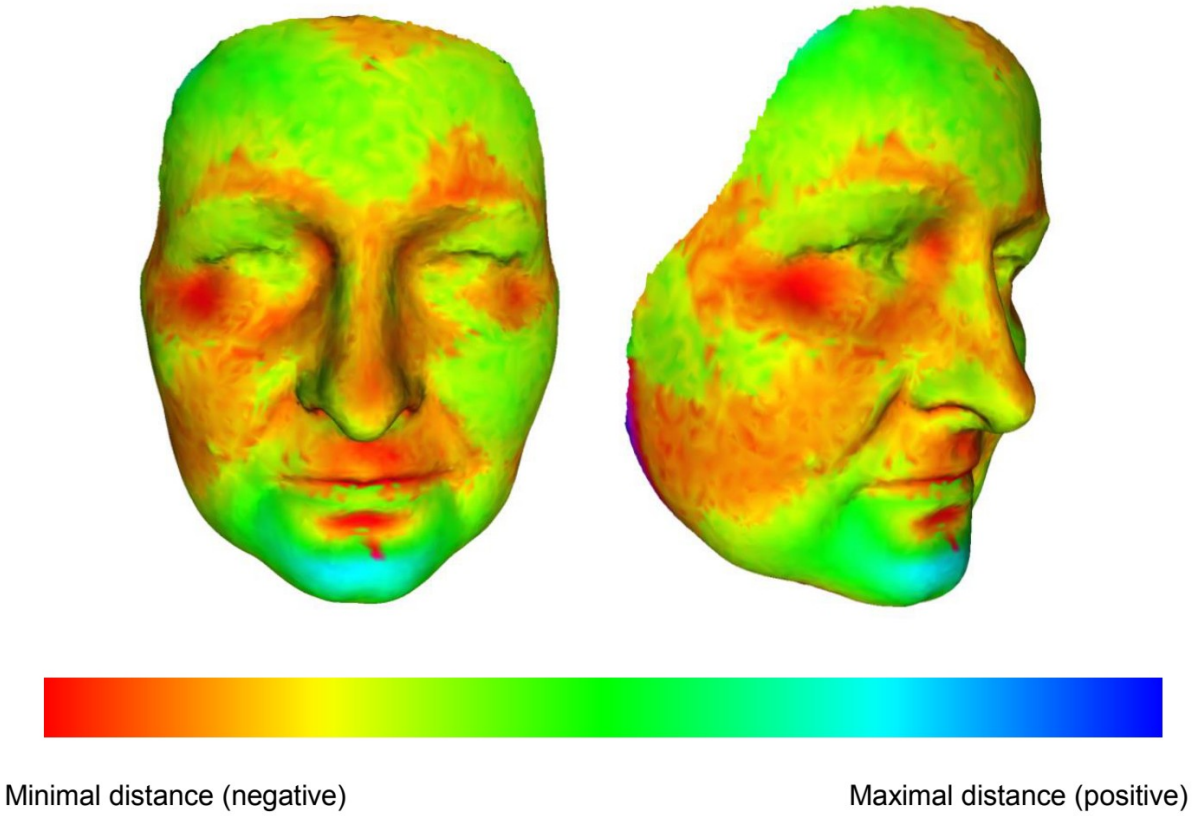
- Approach: Overlapping with **volume highlighting** (by fog)



[Furmanová. K.: Visualization techniques for 3D facial comparison, 2015]

SciVis: Similarity of Human Faces

- Approach: Distance **heatmap** of Hausdorff distance

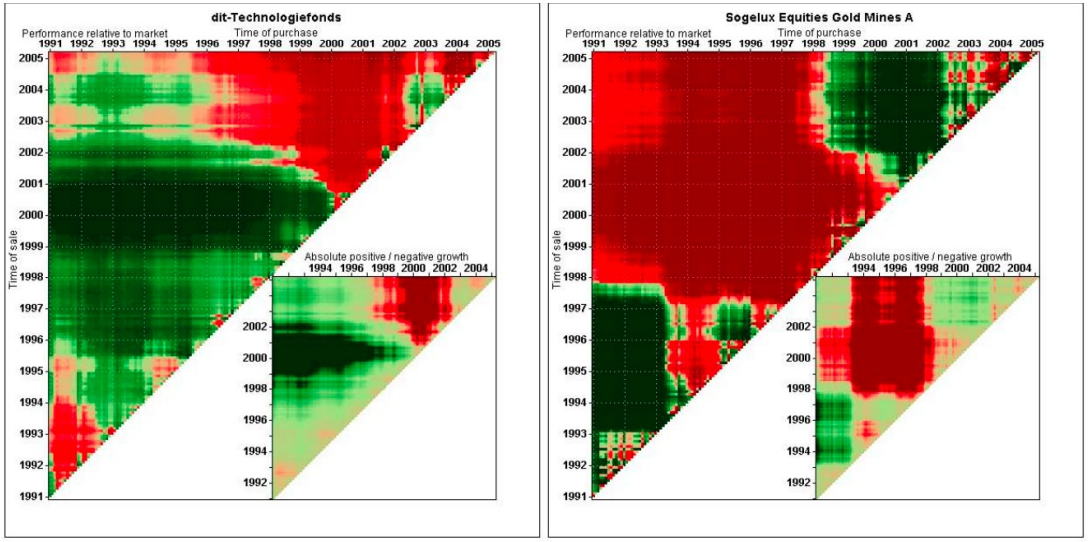


[Furmanová. K.: Visualization techniques for 3D facial comparison, 2015]

FACULTY OF INFORMATICS
Masaryk University

Information visualization (InfoVis)

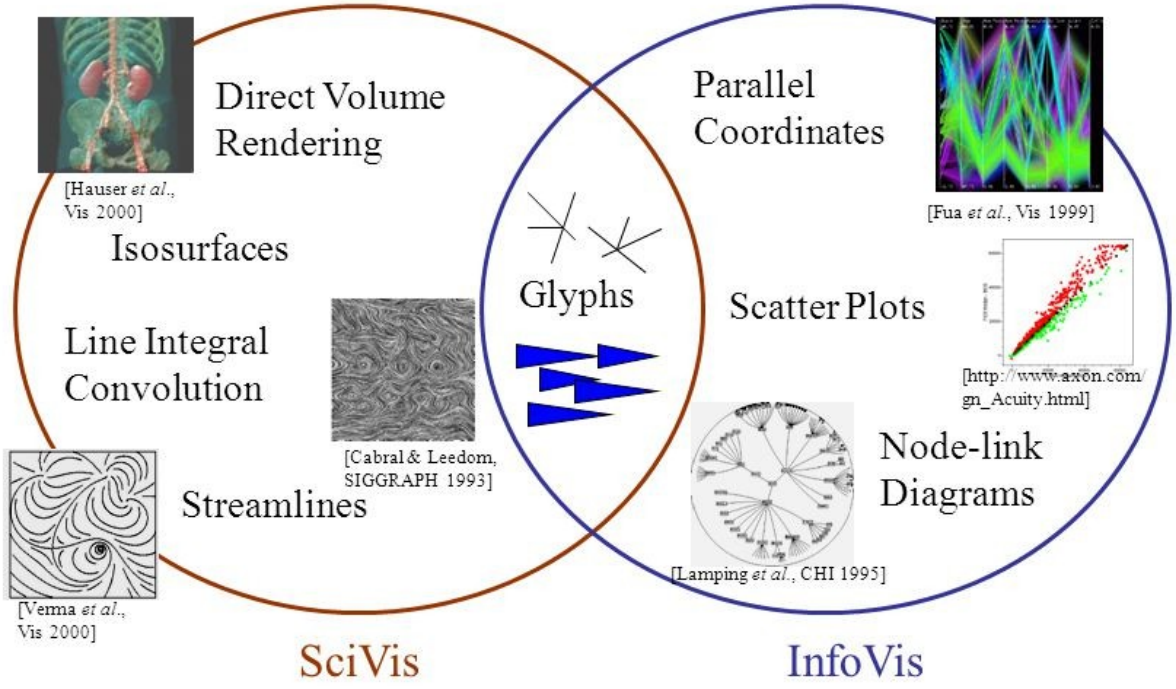
- **Goal:** To provide (interactive) visual representations of abstract data to reinforce human cognition
  - Main focus on representing data in **easily understandable way, supported by intuitive interaction**
  - Transformation of data and phenomenons to complementary, often 2D, views
  - Distribution of the data in space or time, relationships between data values, etc.
- The most common uses:
  - Presentation, **exploratory analysis**, **confirmation analysis**



[Keim et al.: FinDEx: A spectral visualization system for analyzing financial time series data., 2006]

SciVis vs. InfoVis

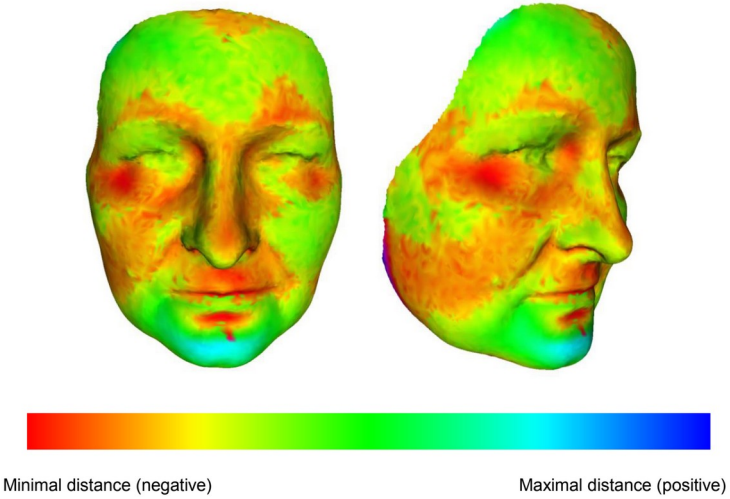
- Explanatory (SciVis) vs. exploratory or confirmation analysis (InfoVis)
- Direct visualization of the phenomena (SciVis) vs. indirect views of phenomena aspects (e.g., the distribution of the data in space or time)
- It is SciVis when the spatial representation is given, and InfoVis when the spatial representation is chosen

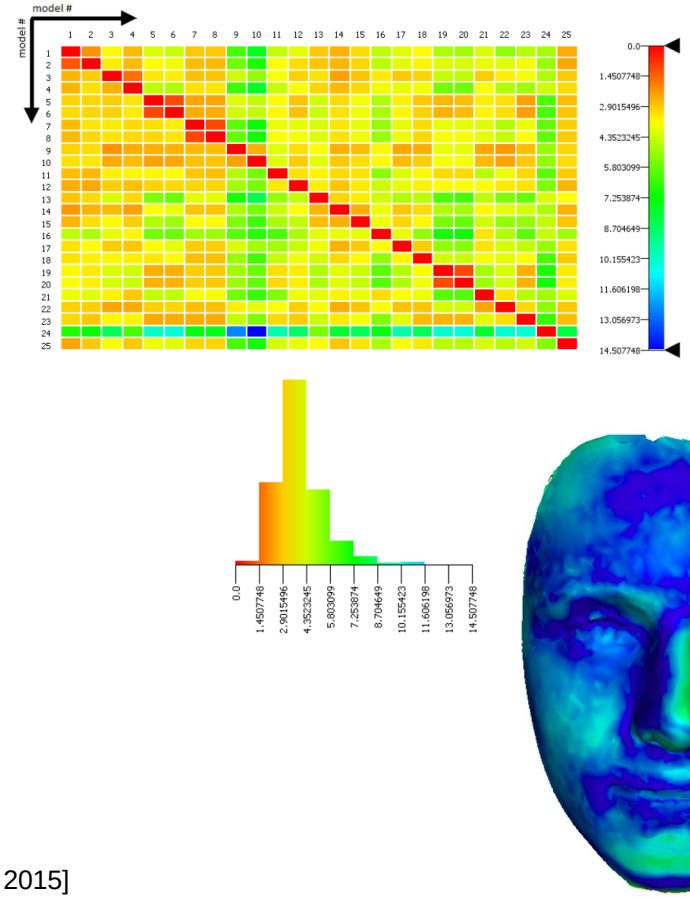


[Hamminger, B.: Information Visualization, https://slideplayer.com/slide/4644055/]

SciVis vs. InfoVis – Face Morphology Example

- SciVis: visualize distances in a 3D heatmap
- InfoVis: View of data distribution (histogram), filtering (e.g., the noise in the eyebrow areas), mutual similarity of N faces, etc.

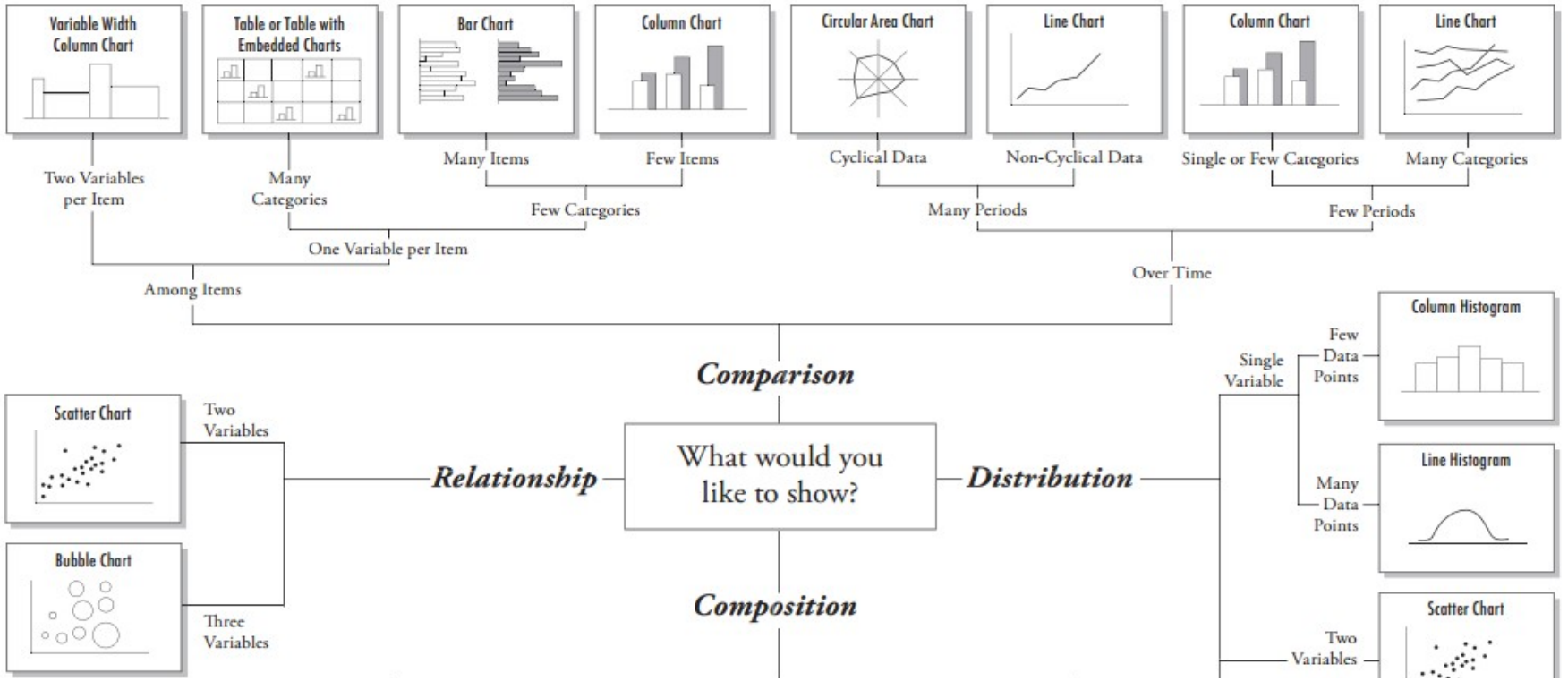




[Furmanová. K.: Visualization techniques for 3D facial comparison, 2015]

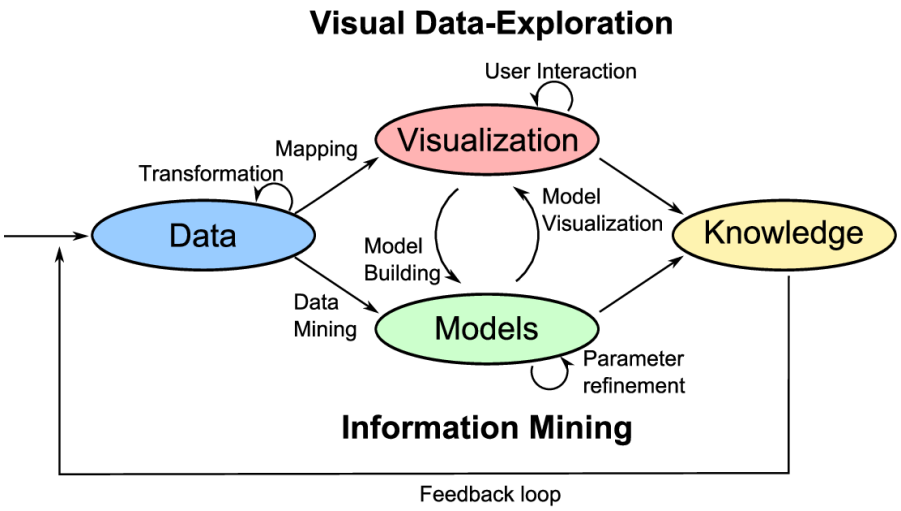
InfoVis Techniques

- Many visualization techniques has been designed for concrete data characteristics
  - Spatial data, geospatial data, multivariate data, trees and networks, …
- Many visualization techniques has been designed for concrete goals
  - Data distribution, comparison, composition, etc.
- Still, the design of concrete solution is challenging
  - www.datavizproject.com

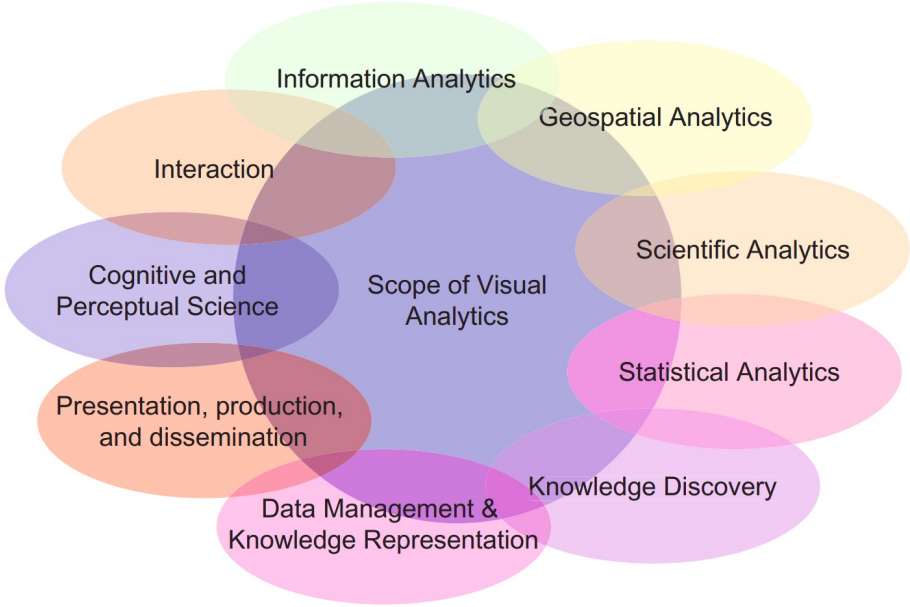


Visual Analytics (VA)

- **Goal:** Analytical **reasoning** supported by interactive visual interfaces
  - Hypotheses-based (what-if) analyses
  - Revealing relationships hidden in the data
  - Building the knowledge (step by step) from the data
- Analytical goals of domain experts (users) drive the VA solution



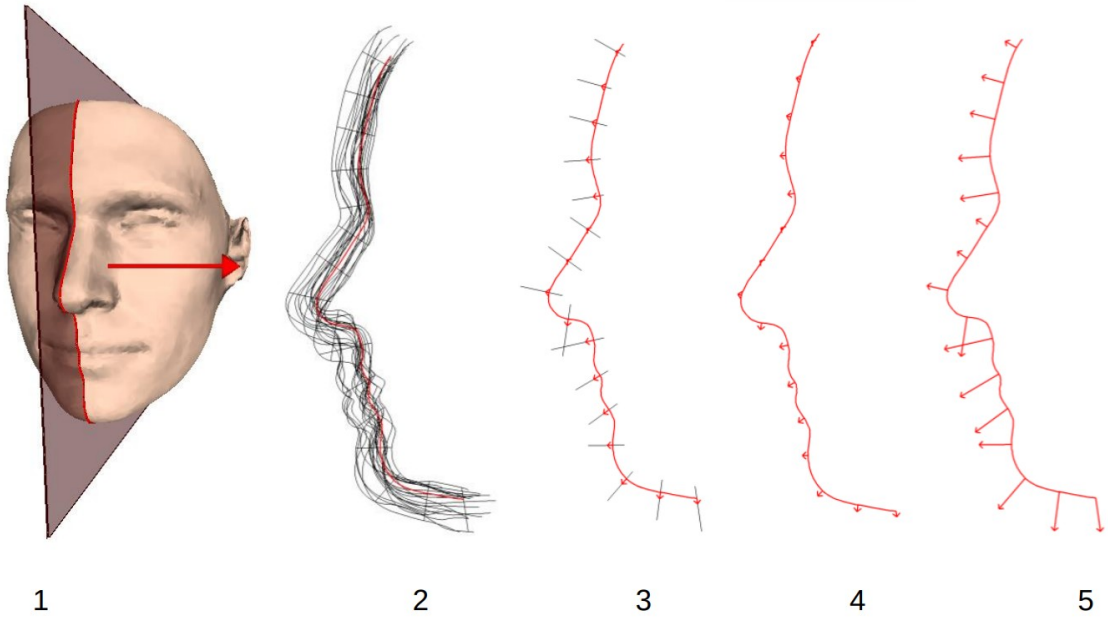
[Keim et al.: Visual Analytics, 2009]



[Keim et al.: Visual Analytics: Scope and Challenges, 2008]

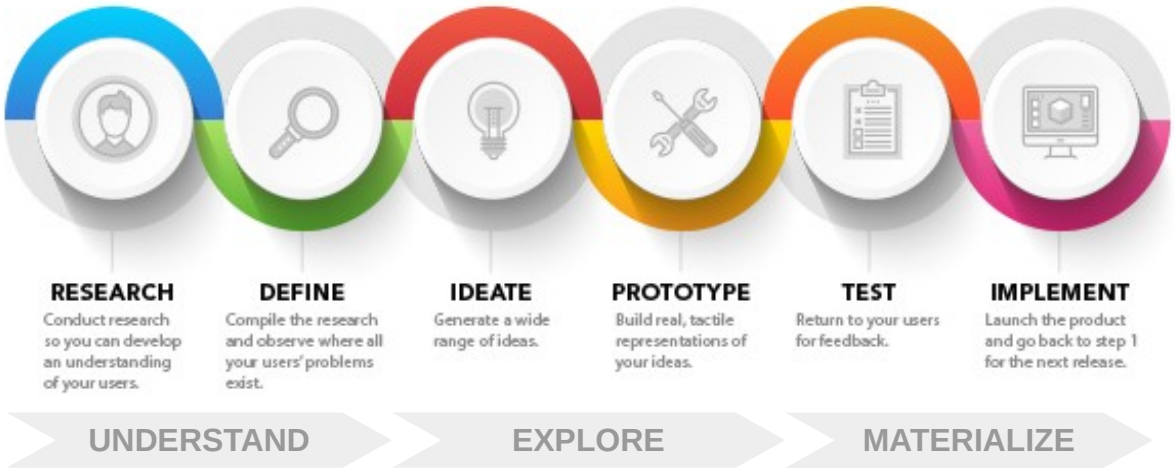
VA – Face Morphology Example

- High-level analytical goals (examples):
  - Face symmetry: Modeling for plastic surgery
  - Morphing: Estimation how a face will change when a child grows up
  - Identification: Searching for a similar faces in a database, e.g., by police
  - Classification: Identification of clusters in a set of cases, e.g., by race, gender, age, etc.



VA methodology

- The development of a really useful VA tool is challenging. It is necessary to follow many rules and best practices to achieve good results and to prove usability
  - Tight cooperation with *domain experts* for both requirements analysis and usability evaluation
  - Using iterative design methodologies, e.g., *user-centered design* (it isn't an ad-hoc process)
  - Formal *evaluation* of results, e.g., quantitative and qualitative methods of measuring user experience
- The development process can be considered a special discipline of software engineering



FACULTY
OF INFORMATICS
Masaryk University

Visual Analytics for Practical Cybersecurity Training

FACULTY OF INFORMATICS
Masaryk University

Problem statement

No tangible output (like a code in programming courses)

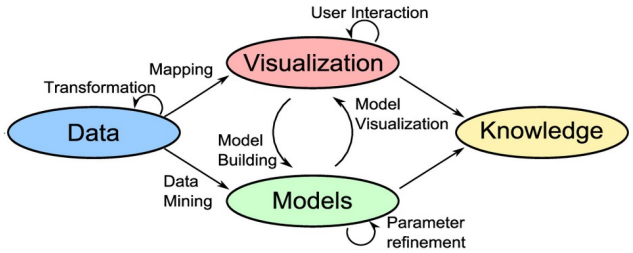
- ***Tutors*** have no idea, what trainees do, whether they are stacked in some task, etc.
- ***Trainees*** don't know whether what they did wrong, or whether there was a faster solution to the tasks.
- ***Training designers*** don't know whether the game was too easy or difficult.

- **Research Goal:** To research and develop data analysis tools providing insight into educational aspects of cybersecurity training and enable comparison, assessment, and continuous improvement.

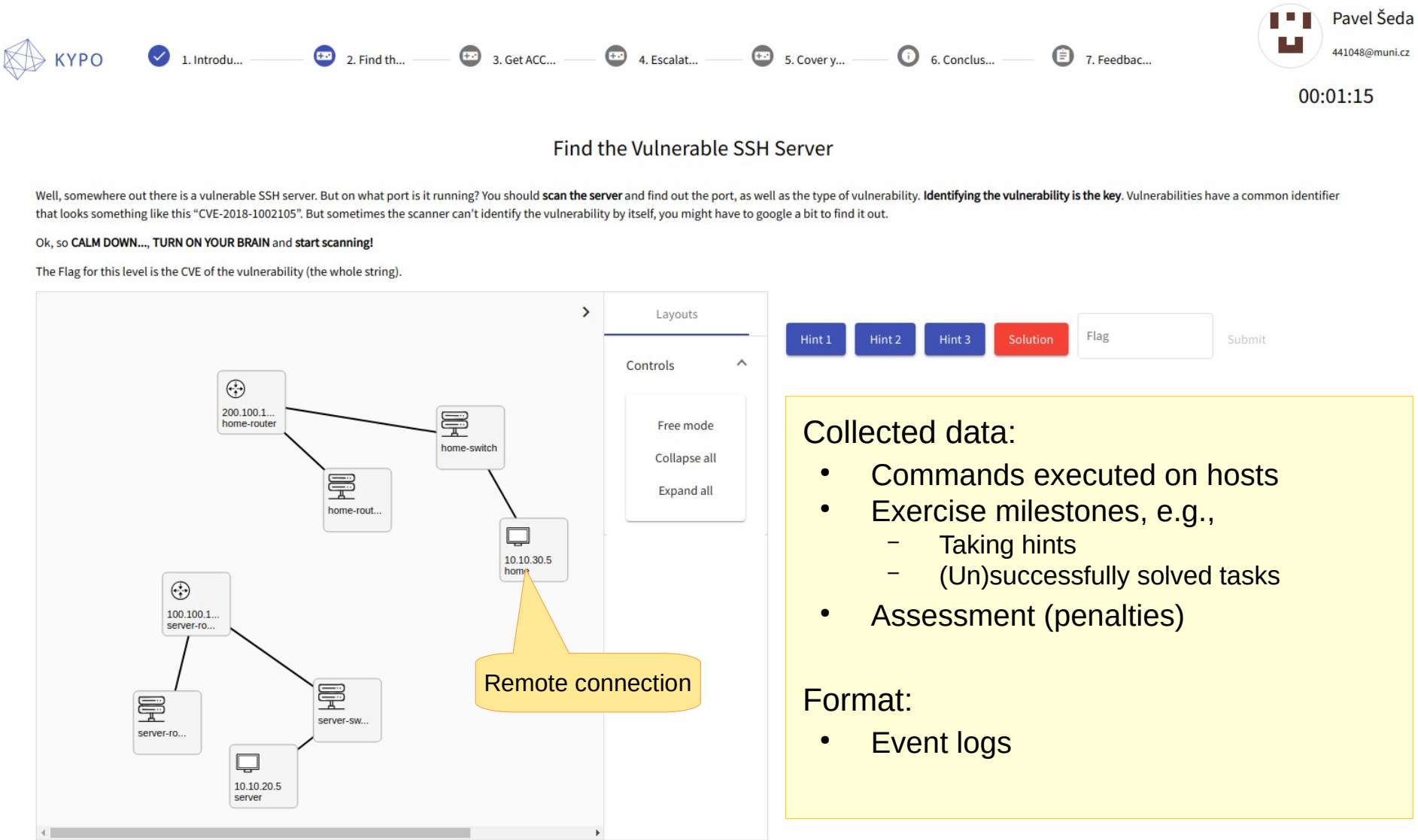
VA for Practical Cybersecurity Training: Data

raw training logs

```
User1;2.08.2020 10:31:43;use webmin_backdoor
User1;2.08.2020 10:32:44;set RHOST
User1;2.08.2020 10:33:19;set LHOST
User1;2.08.2020 10:34:27;set SSL
User1;2.08.2020 10:34:35;set TARGET
User2;2.08.2020 10:32:17;use webmin_backdoor
User2;2.08.2020 10:32:43;exploit
User2;2.08.2020 10:44:33;set RPORT
User2;2.08.2020 10:45:21;exploit
User2;2.08.2020 10:56:02;set LHOST
User2;2.08.2020 10:56:20;set SSL
User2;2.08.2020 10:58:35;set TARGET
...
```



KYPO Cyber Range Platform



Collected data:

- Commands executed on hosts
- Exercise milestones, e.g.,
  - Taking hints
  - (Un)successfully solved tasks
- Assessment (penalties)

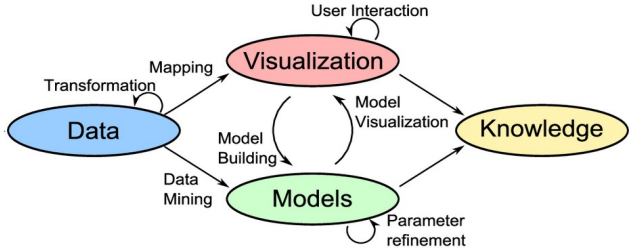
Format:

- Event logs

VA for Practical Cybersecurity Training: Knowledge

raw training logs

```
User1;2.08.2020 10:31:43;use webmin_backdoor
User1;2.08.2020 10:32:44;set RHOST
User1;2.08.2020 10:33:19;set LHOST
User1;2.08.2020 10:34:27;set SSL
User1;2.08.2020 10:34:35;set TARGET
User2;2.08.2020 10:32:17;use webmin_backdoor
User2;2.08.2020 10:32:43;exploit
User2;2.08.2020 10:44:33;set RPORT
User2;2.08.2020 10:45:21;exploit
User2;2.08.2020 10:56:02;set LHOST
User2;2.08.2020 10:56:20;set SSL
User2;2.08.2020 10:58:35;set TARGET
...
```



insight of trainees V_1	insight of organizing participants V_2
personal feedback V_3	quality of training exercise V_4
behavior analysis V_5	infrastructure analysis V_6

FACULTY OF INFORMATICS
Masaryk University

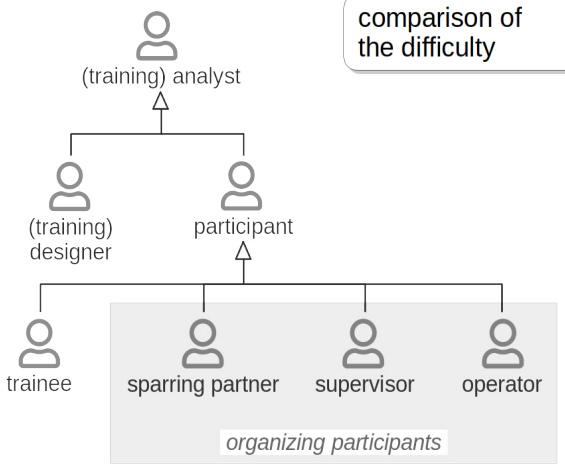
VA for Practical Cybersecurity Training: Knowledge

visual situational awareness

- insight of trainees V_1
  - trainee
    - awareness of the state of network environment V_{1A}
    - awareness of cybersecurity posture V_{1B}
- insight of organizing participants V_2
  - sparring partner
    - training progression V_{2A}
  - supervisor
    - training management V_{2B}
  - operator
    - infrastructure management V_{2C}

visual data analytics

- personal feedback V_3
  - trainee
    - personal reflections on trainees V_{3A}
  - supervisor
    - impact of supervision V_{3B}
- quality of training exercise V_4
  - designer
    - correctness of a training definition V_{4A}
  - designer
    - difficulty of a training definition V_{4B}
  - designer
    - comparison of the difficulty V_{4C}
- behavior analysis V_5
  - analyst
    - successful strategies V_{5A}
  - analyst
    - cooperation patterns V_{5B}
- infrastructure analysis V_6
  - operator & designer
    - performance analysis V_{6A}
  - operator & designer
    - reliability analysis V_{6B}

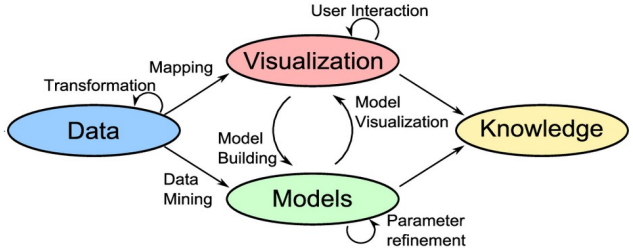


OŠLEJŠEK, Radek, Vít RUSŇÁK, Karolína DOČKALOVÁ BURSKÁ, Valdemar ŠVÁBENSKÝ, Jan VYKOPAL and Jakub ČEGAN. **Conceptual Model of Visual Analytics for Hands-on Cybersecurity Training.** In *IEEE Transactions on Visualization and Computer Graphics*, 2021.

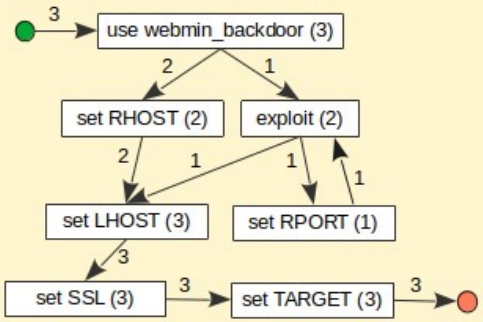
VA for Practical Cybersecurity Training: Models

raw training logs

```
User1;2.08.2020 10:31:43;use webmin_backdoor
User1;2.08.2020 10:32:44;set RHOST
User1;2.08.2020 10:33:19;set LHOST
User1;2.08.2020 10:34:27;set SSL
User1;2.08.2020 10:34:35;set TARGET
User2;2.08.2020 10:32:17;use webmin_backdoor
User2;2.08.2020 10:32:43;exploit
User2;2.08.2020 10:44:33;set RPORT
User2;2.08.2020 10:45:21;exploit
User2;2.08.2020 10:56:02;set LHOST
User2;2.08.2020 10:56:20;set SSL
User2;2.08.2020 10:58:35;set TARGET
...
```



process model

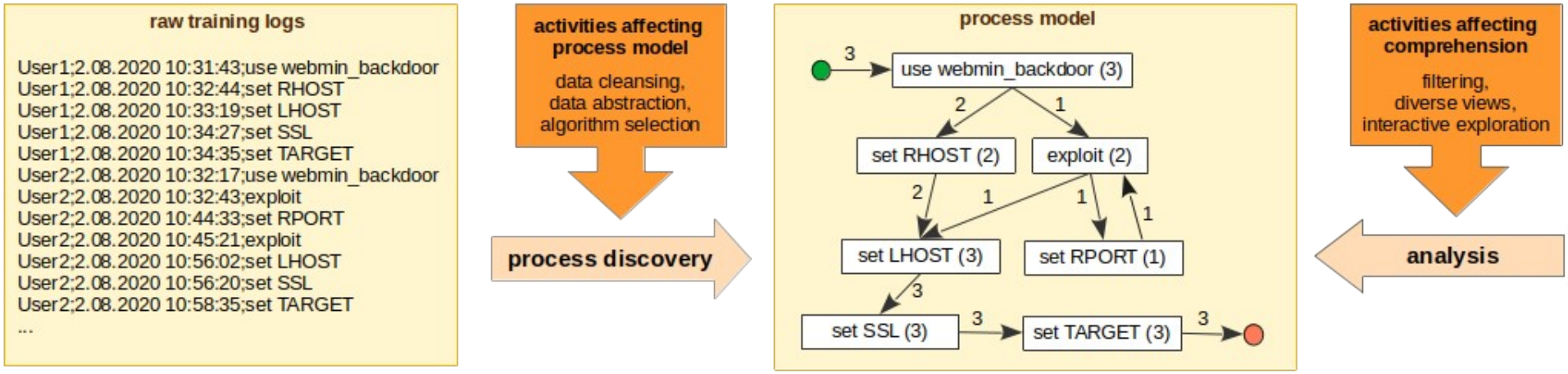


- insight of trainees V_1
- insight of organizing participants V_2
- personal feedback V_3
- quality of training exercise V_4
- behavior analysis V_5
- infrastructure analysis V_6

FACULTY OF INFORMATICS
Masaryk University

VA for Practical Cybersecurity Training: Models

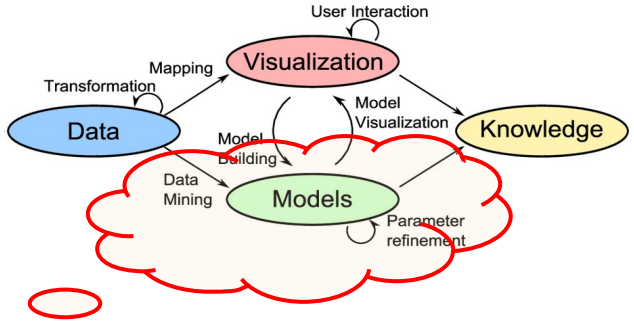
- Statistical models
- Process mining models
  - Challenge: Data pre-processing and mapping affect obtained graphs
  - Challenge: The selection of process discovery algorithm affects obtained graphs
  - Challenge: Problem with the scalability of obtained graphs

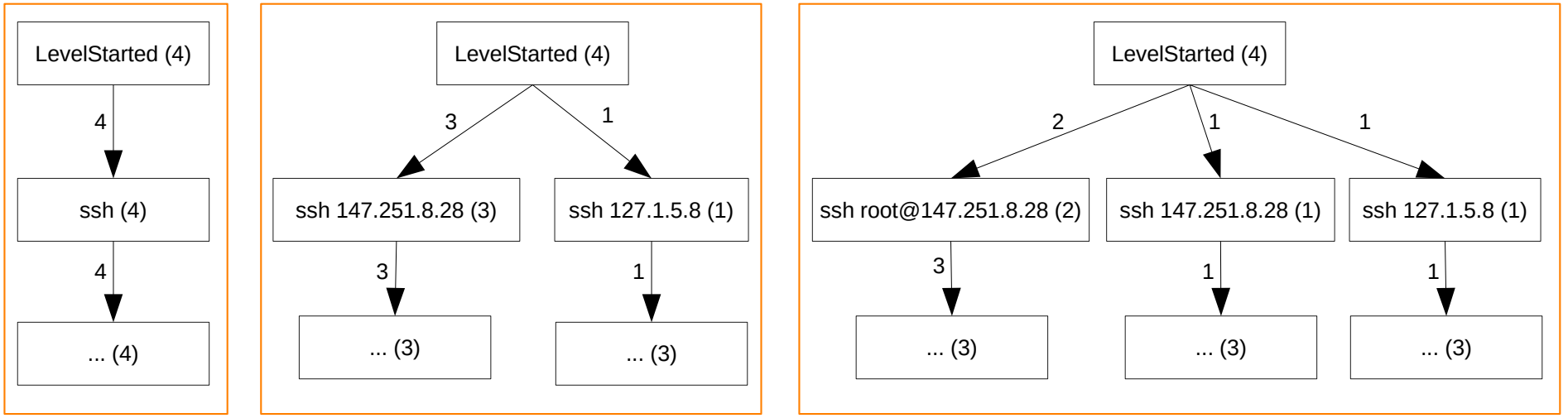


FACULTY OF INFORMATICS
Masaryk University

Data aggregation and filtering

- What is the same or sufficiently similar commands?
  - User 1: `ssh root@147.251.8.28`
  - User 2: `ssh 147.251.8.28`
  - User 3: `ssh -4 root@147.251.8.28`
  - User 4: `ssh 127.1.5.8`

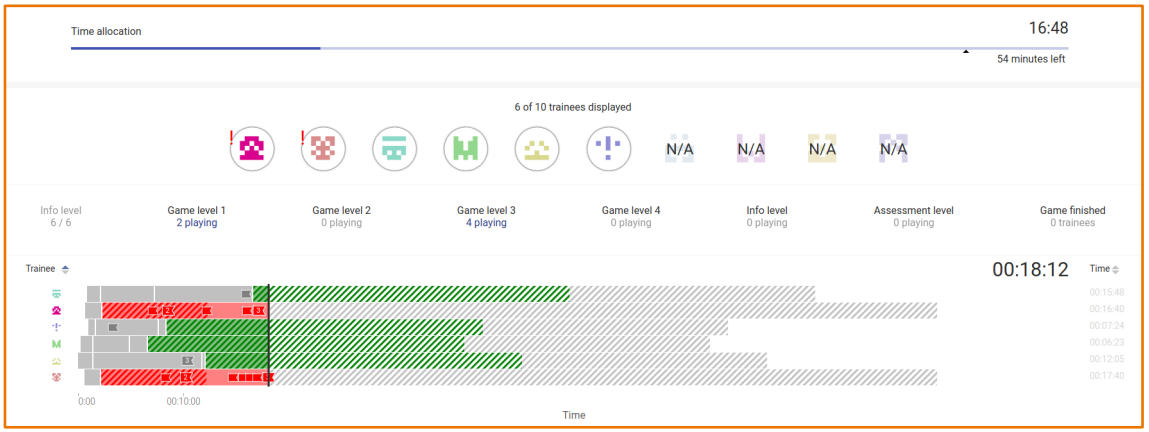


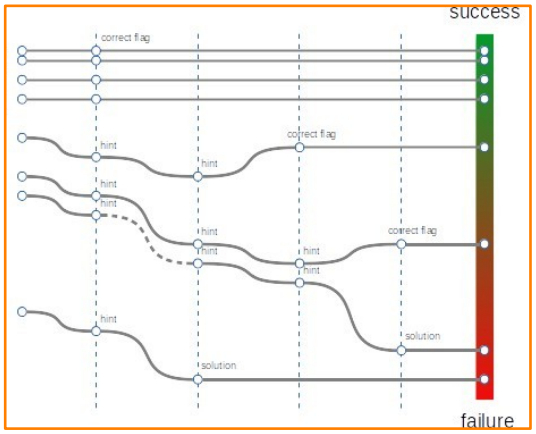


MACÁK, Martin, Radek OŠLEJŠEK and Barbora Bühnová. **Process Mining Analysis of Puzzle-Based Cybersecurity Training.** Innovation and Technology in Computer Science Education (ITiCSE'22), to appear, 2022.

FACULTY OF INFORMATICS
Masaryk University

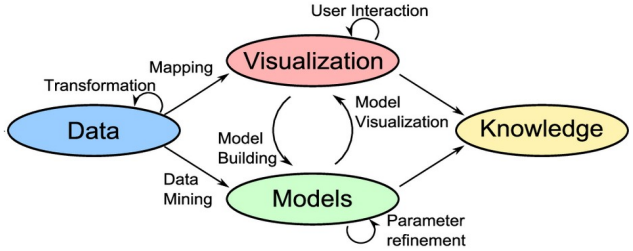
VA for Practical Cybersecurity Training: Visualizations



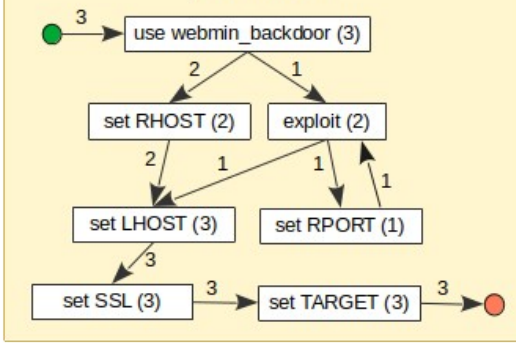


raw training logs

```
User1;2.08.2020 10:31:43;use webmin_backdoor
User1;2.08.2020 10:32:44;set RHOST
User1;2.08.2020 10:33:19;set LHOST
User1;2.08.2020 10:34:27;set SSL
User1;2.08.2020 10:34:35;set TARGET
User2;2.08.2020 10:32:17;use webmin_backdoor
User2;2.08.2020 10:32:43;exploit
User2;2.08.2020 10:44:33;set RPORT
User2;2.08.2020 10:45:21;exploit
User2;2.08.2020 10:56:02;set LHOST
User2;2.08.2020 10:56:20;set SSL
User2;2.08.2020 10:58:35;set TARGET
...
```



process model



- insight of trainees V_1
- insight of organizing participants V_2
- personal feedback V_3
- quality of training exercise V_4
- behavior analysis V_5
- infrastructure analysis V_6

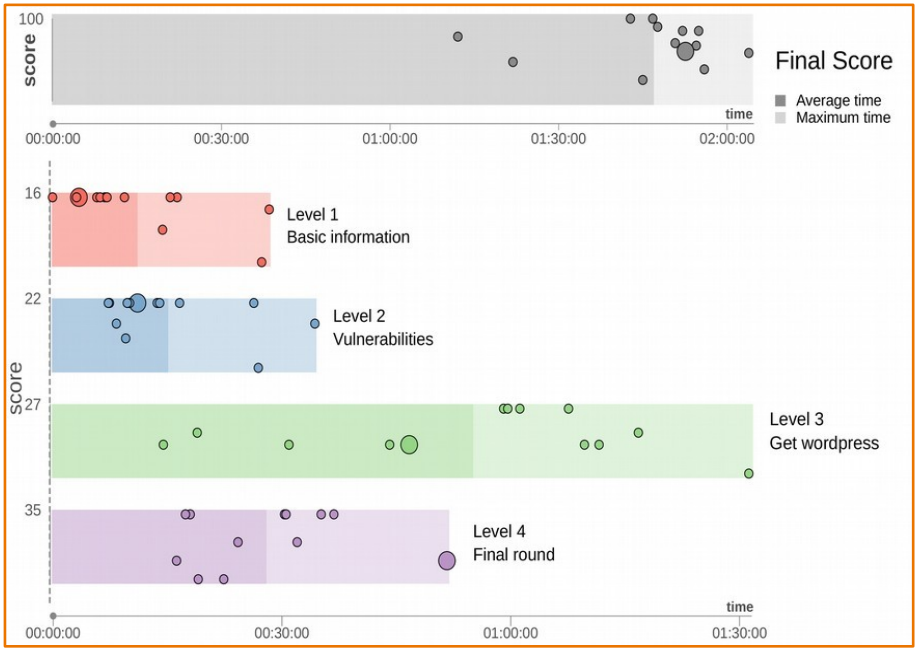
FACULTY OF INFORMATICS
Masaryk University

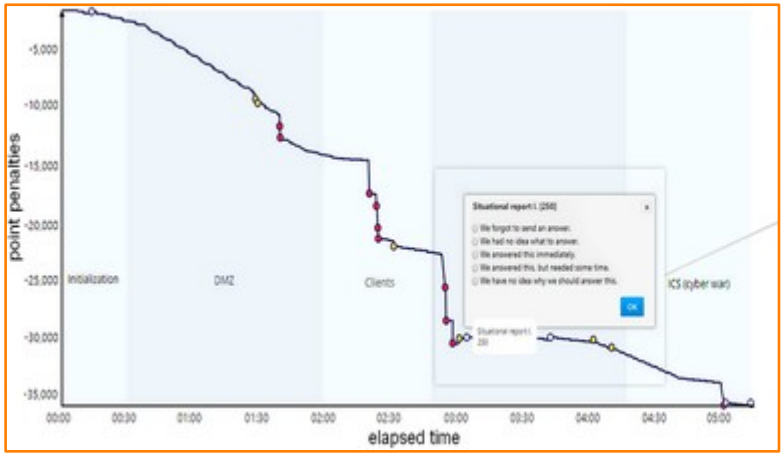
Data-Based Vis: Personalized feedback to trainees

Goal: Learning from own mistakes

- What did I do wrong in selected tasks?
- Where I lost most points and why?
- ...

OŠLEJŠEK, Radek, Vít RUSŇÁK, Karolína BURSKÁ, Valdemar ŠVÁBENSKÝ a Jan VYKOPAL. **Visual Feedback for Players of Multi-Level Capture the Flag Games: Field Usability Study.** In *IEEE Symposium on Visualization for Cyber Security (VizSec'19)*





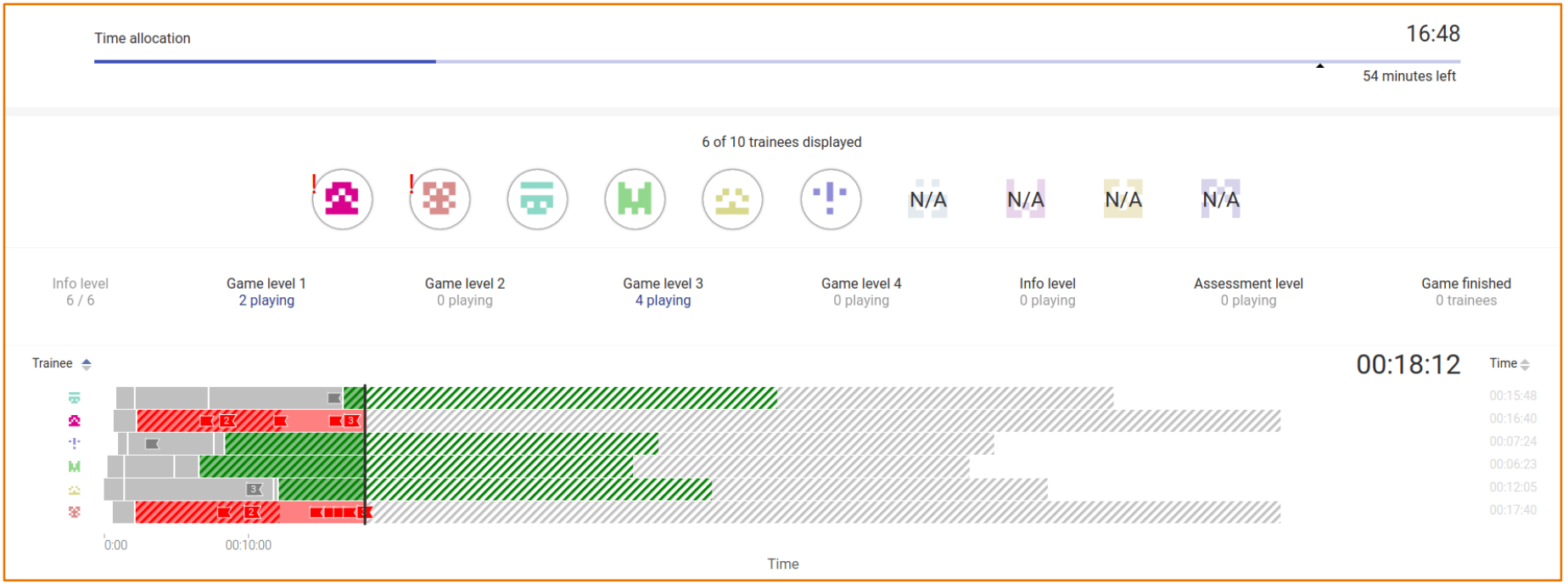
VYKOPAL, Jan, Radek OŠLEJŠEK, Karolína BURSKÁ and Kristína ZÁKOPČANOVÁ. **Timely Feedback in Unstructured Cybersecurity Exercises.** In *ACM Technical Symposium on Computer Science Education (SIGCSE'18)*

FACULTY OF INFORMATICS
Masaryk University

Data-Based Vis: Insight for organizing participants

Goal: Situational awareness and timely intervention

- Which trainees are in trouble? Why?
- Is the training session on schedule, or is there some delay?
- ...



DOČKALOVÁ BURSKÁ Karolína, Vít RUSŇÁK and Radek OŠLEJŠEK.
Enhancing Situational Awareness for Tutors of Cybersecurity Capture the Flag Games.
In *International Conference Information Visualization (iV'21)*.

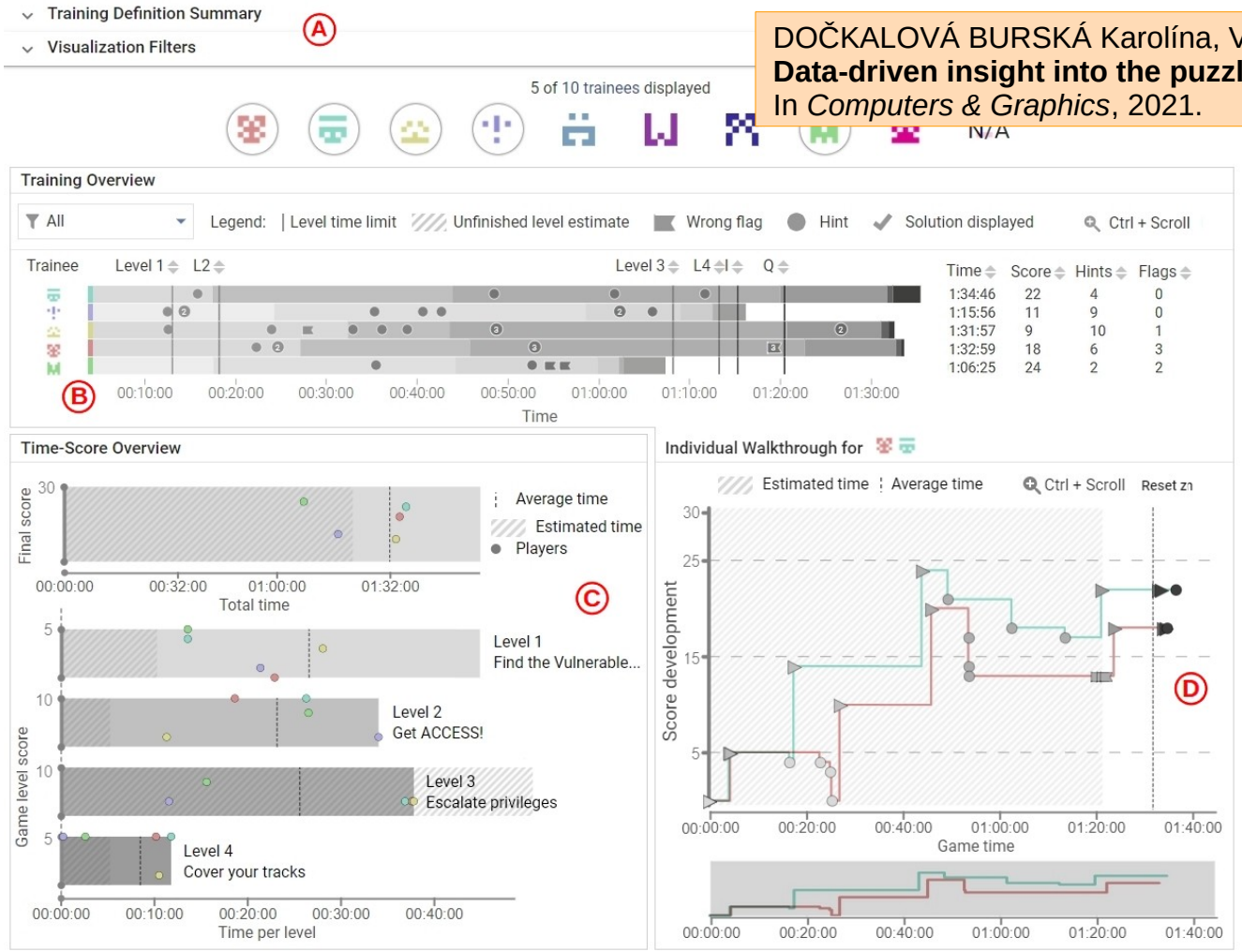
FACULTY OF INFORMATICS
Masaryk University

Data-Based Vis: Post-training analysis

Goal: Improve the impact of learning

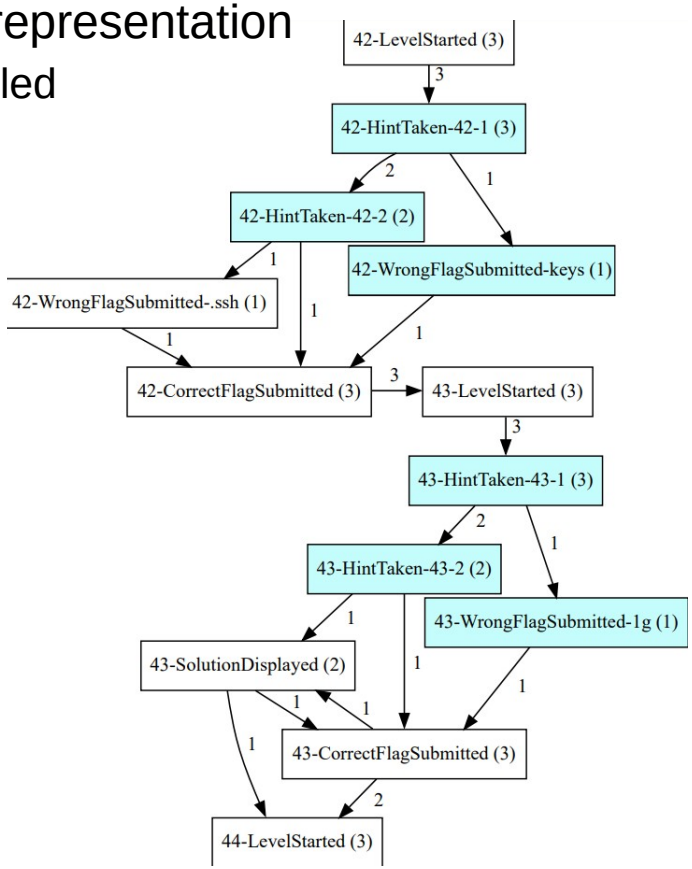
- Was training too easy or difficult?
- What are the sources of losing motivation and giving up the training?
- Are there some flows in the scenario, requirements, etc.?

DOČKALOVÁ BURSKÁ Karolína, Vít RUSŇÁK and Radek OŠLEJŠEK. **Data-driven insight into the puzzle-based cybersecurity training.** In *Computers & Graphics*, 2021.



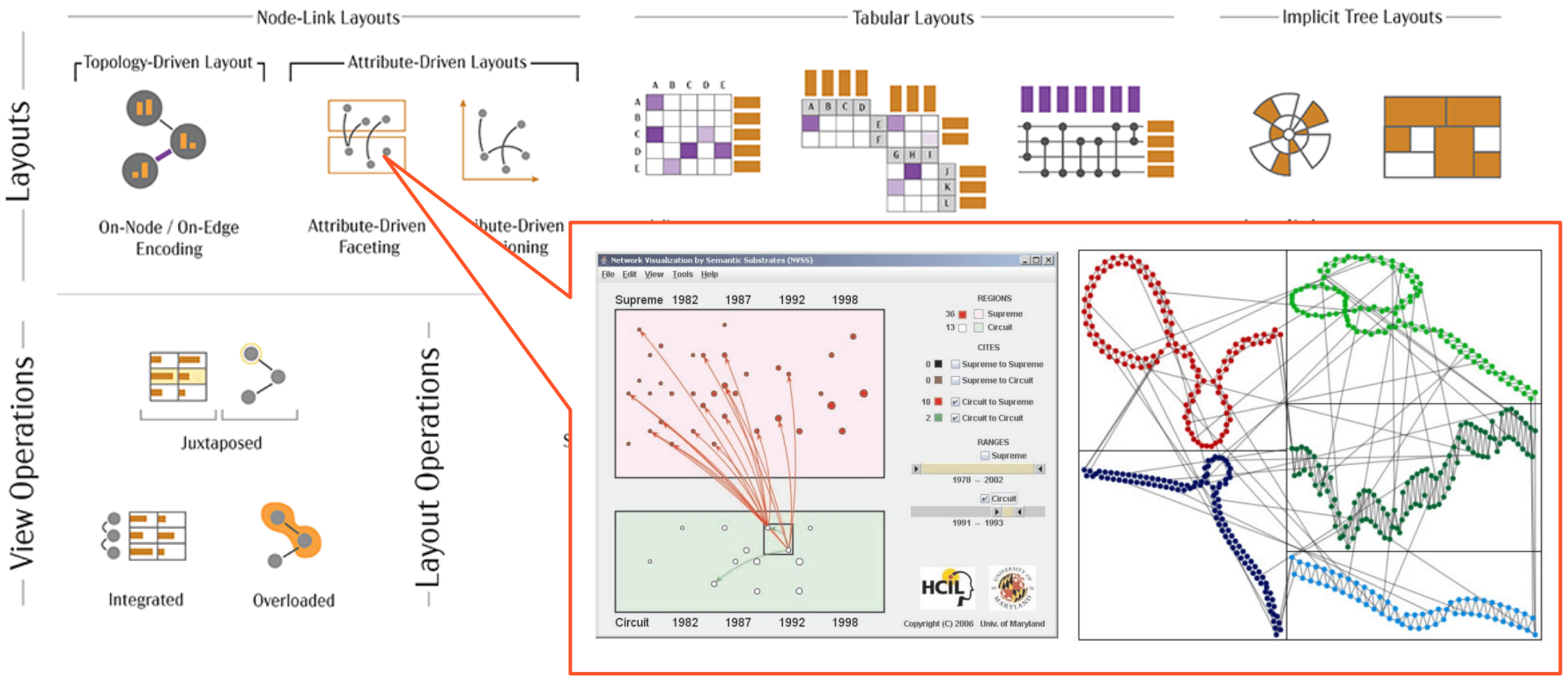
Model-Based Vis: Process Graphs

- The previous examples demonstrate visualization of data
- Visualization of process models brings challenges in comprehensibility and scalability of models.
- Idea: Provide alternate view to a traditional graph representation
  - From the VA perspective, process graphs are so-called **multivariate networks**



Multivariate Networks

- No silver bullet solution available.
- A combination of carefully selected, adapted, and mutually connected visualizations and interactions is required.
- A suitable approach depends on the properties of the multivariate network

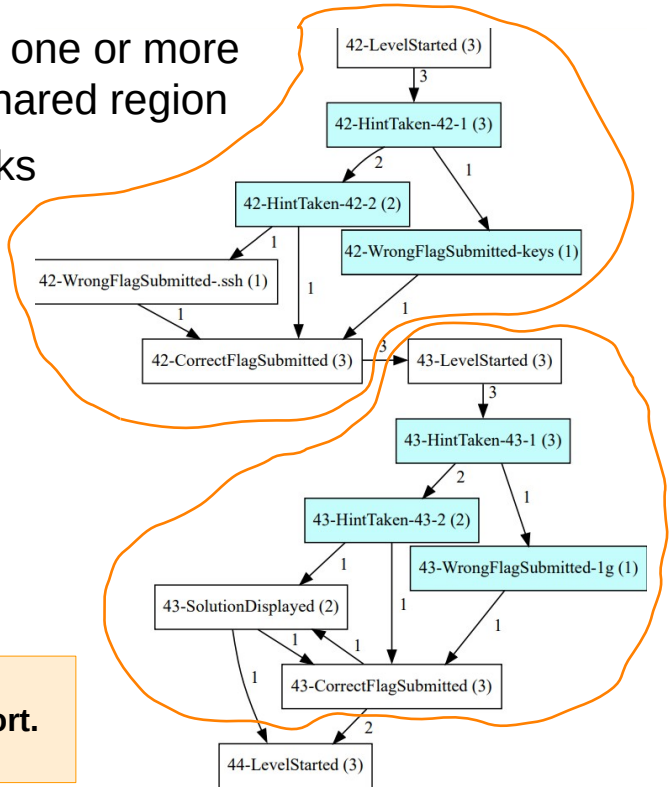


[Nobre, C. et al.: The state of the art in visualizing multivariate networks. In Computer Graphics Forum, 2019]

FACULTY OF INFORMATICS
Masaryk University

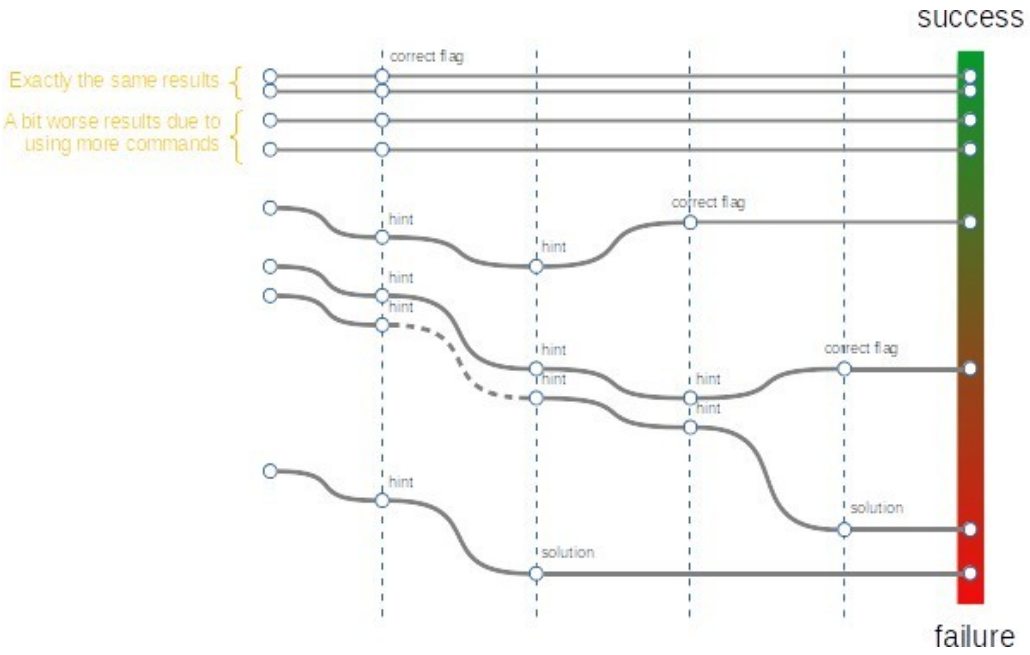
Process Models as Multivariate Networks

- Properties of our process models:
  - Medium (less than 1.000 nodes) or large size (more than 1.000 nodes)
  - K-partite
  - Heterogeneous nodes with few attributes and homogeneous edges.
- Suitable visualization tactics:
  - Attribute-driven faceting: groups nodes according to one or more attributes and places the elements of a group in a shared region
  - Quilts: A tabular layout optimized for layered networks
  - Integrated or juxtaposed view operations.
    - Integrated: the topology and the attribute visualizations are laid out with the other view in mind.
    - Juxtaposed: separates the topology visualization from the attribute visualization into two or more views.



MACÁK, Martin, Radek OŠLEJŠEK and Barbora Bühnová.
Applying Process Discovery to Cybersecurity Training: An Experience Report.
Cyber Range Technologies and Applications (CACOE'22), under review.

Infrastructure analysis



Highlight possible flows:

- ☐ Unclear task description
  - *Taking hint immediately (without trying any command)*
- ☐ The loss of motivation to finish the task or the lack of time
  - *Taking hints or solution without trying any command*
- ☐ Useful hints, but possibly too much instructive
  - *Taking a hint is followed by the correct flag*
- ☐ Useless hints
  - *Taking a hint is followed by taking another hint, solution, or incorrect flag*

Thank you for your attention!