

Visual Analytics

Yvo Keller, FS2023

Introduction

This report will cover my learnings as part of the visual analytics course, based on the ADNI dataset (Alzheimer's Disease Neuroimaging Initiative, 2023). I am familiar with it from the challenge "Understanding Brain Health" which I completed during HS22. I will focus on providing visual analytics tools built using GlueViz on this rich dataset consisting of patient data with biomarkers, clinical tests and diagnosis over many years.

My GitHub repository with all the code can be found here: <https://github.com/kelleryvo/van>

Table of Content

1	LO1: ALGORITHMS	1
1.1	What is Visual Analytics?	1
1.2	Computational vs. Visual Support	1
1.3	Spatialization and Grouping	2
1.4	Linked visualizations	2
1.5	Iterative manner of computational methods	2
2	LO2: COGNITION AND COMPUTATION	3
2.1	Adding computation	3
2.2	Cognition	4
2.3	Results and takeaways	5
3	LO3: TOOLS	6
3.1	GlueViz	6
3.2	Tableau	6
3.3	PowerBI	7
3.4	Side-by-side comparison	7
4	LO4: EVALUATION	9
4.1	An overview of the usability testing methods	9
4.2	Performing the usability test	9
4.3	Conclusion and suggestions for improvement	10
5	BIBLIOGRAPHY	11
Yvo Keller	Visual Analytics	1

6	APPENDIX	13
6.1	My visual analytics solution	13
6.2	Cognitive walk-through	14

1 LO1: Algorithms

1.1 What is Visual Analytics?

Visual analytics systems combine machine learning or other analytic techniques with interactive data visualization to promote sensemaking and analytical reasoning. This combination allows for the interpretation of large, complex datasets. (Endert, 2023). In the context of visual analytics, computational techniques are commonly used to enable effective visual presentation of voluminous and/or complex information to human analysts. This becomes especially relevant when there is no meaningful way to create a visualization showing every data point in detail. The main goal is to enable an overall view of the whole bulk of information, achieved through a high degree of abstraction and omission of details. The details, when needed, can be obtained by means of interaction. (Andrienko et al., 2020)

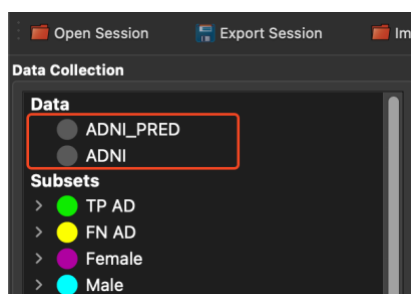


Figure 1 The imported datasets in GlueViz

In line with the principles of visual analytics, my goal was to create an overview of the ADNI dataset, specifically to interpret the CSF biomarkers ABETA, TAU, and PTAU. Having an overview of the data, especially understanding relations between different features, becomes increasingly complex the more features and more dimensions a dataset has. This is where visual analytics becomes helpful. As several introductory videos looked promising and it was recommended in the course, I decided to go with the open-source software GlueViz. I created a new session, imported the relevant datasets (Figure 1) and set up several visualizations based on the Histogram and 2D-Scatter Plot Data Viewers.

1.2 Computational vs. Visual Support

Visual analytics combines interactive visualizations with computational techniques for data processing and analysis. This combination has **two aspects** (Andrienko et al., 2020):

- **Computational support** to visual analysis: This involves using the outcomes of computations to provide input to human cognition, which are then visually represented.
- **Visual support** to the application of computational methods: This includes visual exploration of data properties to prepare data for computations, evaluation of computation outcomes, and comparison of results from different runs of computational techniques.

In GlueViz, visual exploring of data properties becomes possible by dragging-and-dropping an imported dataset into the work area, then selecting the type of plot to use. I set this up specifically for the properties *Gender*, *ADNI Phase* and *Year of Birth* and *Diagnosis* (Figure 2). These plots will provide visual support to evaluating the results of the computation classifying a patient's diagnosis, which I plan to add in LO2.

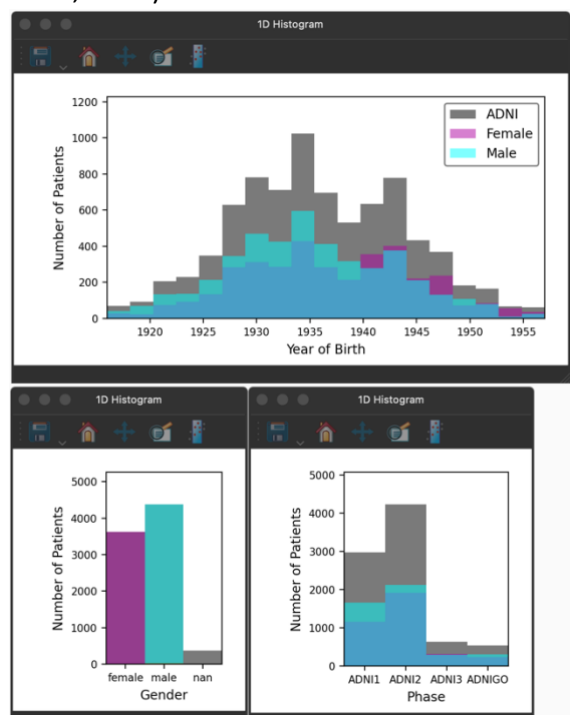


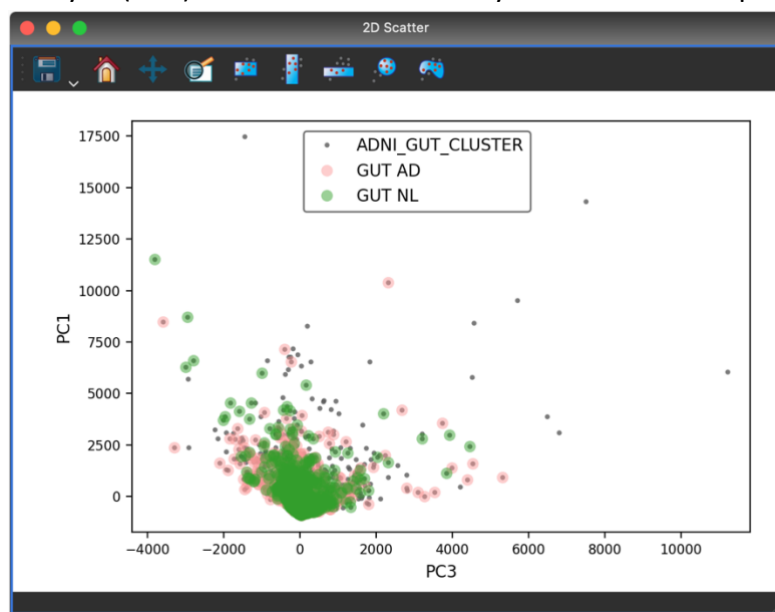
Figure 2 ADNI Histogram Plots in GlueViz

1.3 Spatialization and Grouping

A major purpose for using computational methods in visualization is to enable an overview of voluminous and complex data. Andrienko mentions that the **two general approaches** to achieving this are spatialization and grouping.

- **Spatialization** is achieved through data embedding techniques, which position data items in an artificial two or three-dimensional space according to their similarity or relatedness. This spatial arrangement can be visually represented in a two- or three-dimensional plot that can be interactively viewed from different perspectives.
- **Grouping** is achieved using clustering algorithms, which organize data items into groups based on their similarity. The resulting groups can be treated as units and are characterized by statistical summaries of the characteristics of their elements.

Grouping can be helpful in my scenario with ADNI data to identify groups of similar patients according to a given set of features based on an unsupervised learning algorithm like Principal Component Analysis (PCA). PCA is a dimensionality reduction technique to derive low-dimensional set of features



from a much larger set while still preserving as much variance as possible. Thus, it allows to show high-dimensional data in a low dimension. When we color the datapoints based on a feature like the patient's diagnosis, it can be viewed as a clustering algorithm (Li, 2019). The ADNI dataset has data on over 120 different gut metabolites for each patient (Alzheimer's Disease Neuroimaging Initiative, 2023). I ran PCA on the gut metabolites data and plotted the result as a new 2D-Scatter Data Viewer in GlueViz (Figure 3). It becomes apparent that patients in the AD group tend to be outliers in the spectrum of the first two principal components.

Figure 3 Grouping high-dimensional data in GlueViz using PCA for clustering.

1.4 Linked visualizations

Visualizations that are linked to each other and adjust their appearance based on user selection are another key part of allowing visual analytics. They can be linked by means of color propagation and highlighting of corresponding items in other plots. (Andrienko et al., 2020)

This is exactly what GlueViz allows with its concept of subsets. I can use a lasso tool to select some specific outliers in a 2D-Scatter plot and will then see the selected data points pop up in all other plots, with them being highlighted in the subsets color (Figure 4).

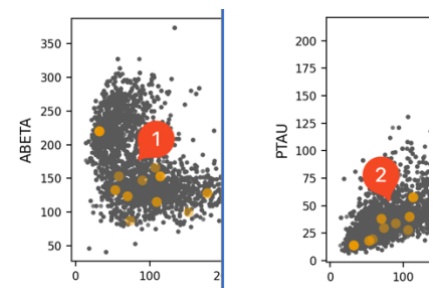


Figure 4 Linked visualization

1.5 Iterative manner of computational methods

Computational methods often work iteratively, allowing for inspection of the process and intermediate results, with visualization playing a key role in this inspection. This involves investigating data properties before applying methods, inspecting data usage and processing during, and evaluating results after (Andrienko et al., 2020). My GlueViz implementation supports all these three phases: showing (1) the raw ADNI data (2), the training data for the computational method like a Decision Tree, and (3) the results of the computation as a confusion matrix. I will add (3) as part of implementing computation in LO2.

2 LO2: Cognition and computation

2.1 Adding computation

The goal of this learning objective is to understand the human factors involved in visual analytics and put them into relation with my visual analytics solution. Key to any visual analytics solution is computation. In LO1, I put together a selection of plots that might pose information relevant to understanding the relationship between Alzheimer's disease and CSF biomarkers. Now, I added a machine learning model to the mix. I trained a decision tree to classify a patient within the three diagnosis groups of Normal (NL), Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD). I imported the result of this classification as a new dataset to GlueViz and linked it to the extensive ADNI dataset via the datapoints UID using Glue's "Link Data" feature. This allowed me to add a confusion matrix plot, which now allows the user to create a subset on it and thereby highlighting, as an example, all cases of False Negative MCI classification on each linked visualization.

2.1.1 Cognitive load theory

When combining artificial intelligence with human intelligence, it becomes important to optimize the software tool in consideration of the human factors to broaden the intersection between human and machine intelligence (Greitzer et al., 2011). One important concept is Cognitive Load theory. We differentiate between three types of cognitive load ("Cognitive Load," 2023):

- **Extraneous Cognitive Load:** This is generated by the way in which information is presented to learners and is under the control of instructional designers. It refers to the cognitive resources required to process information that is not directly relevant to the learning task.
- **Intrinsic Cognitive Load:** This is the inherent difficulty associated with a specific instructional topic. It is related to the complexity of the content being learned and is generally considered to be immutable.
- **Germane Cognitive Load:** This refers to the cognitive resources used to process and understand the learning task itself ("belonging to the matter"). It is the work put into creating a permanent store of knowledge, or a schema.

In the context of my visual analytics tool in GlueViz, these concepts are crucial. It must be designed in a way that minimizes extraneous cognitive load, i.e., it should present information in an organized manner to avoid overloading the user's working memory with unnecessary information. This can be achieved by only showing plots relevant to the goal of the analysis. At the same time, it should support the user in dealing with the intrinsic cognitive load, i.e., understanding the complex data and analytics results. Important for this are well-described plots. Finally, it should facilitate the germane cognitive load, i.e., help the user in creating a mental model or schema of the data and the analytics process. This is partially made possible by the understanding of the computational algorithm's decisions a user can gain when creating subsets and using the linked visualization feature of GlueViz.

2.1.2 Magical number 7

Another key concept is Miller's law / "**Magical Number 7**", which is the argument that a human can only hold $7 \pm$ items in short-term-memory ("The Magical Number Seven, Plus or Minus Two," 2022). In terms of visual analytics, this law suggests to not overwhelm to user with too much information. A good balance between providing much of the relevant information to gain insights and not showing too much as to overwhelm the user needs to be found.

2.1.3 Limitations

Next to its merits, visual analytics also has its limitations. It can help provide deeper insight, but also poses a risk of over-simplification (e.g., when clustering high-dimensional data) or leading to incorrect conclusions or misinterpretation if plots are not read correctly. While it's a great way to get a handle on

large data sets and understand the computations that run on them, it can't provide a simple solution for visualizing millions of data points.

2.2 Cognition

For writing the following introspection ("Introspection," 2023) and conducting an objective cognitive walk-through (Nielsen Norman Group, 2023d) on my visual analytics solution based on Nielsen Norman Groups proposed steps, I put myself in the position of the persona of Walker. He works within a research group focused on finding an effective cure for Alzheimer's disease (AD). He possesses a medical degree, and therefore his strengths and contribution are not in the field for machine learning, but rather the anatomic and medical understanding of AD. Visual analytics helps him get an intuition for the computation (the Decision Tree ML model) using his cognitive abilities.

2.2.1 Walker's Introspection

«Before I start performing visual analytics, I find myself consider the scenario at hand. Anne, the Data Scientist in my team, has trained a decision tree model on ADNI data, specifically the cerebrospinal fluid biomarkers ABETA, TAU and PTAU, to classify Normal (NL), Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD). ABETA, TAU, and PTAU are particularly relevant as they are key biomarkers in the cerebrospinal fluid (CSF) that have been associated with Alzheimer's disease, making them crucial in the diagnosis and understanding of the disease's progression (Blennow et al., 2001). The model achieved an impressive precision of 96%. However, precision is not the only metric that matters. I am particularly interested in the cases that were misclassified as Alzheimer's disease. These false positives represent an opportunity for learning. They are the outliers, the exceptions that defy the model's understanding of the data. By studying these cases, I can gain insights into the limitations of the model and identify potential areas for improvement. While being optimistic, I am also aware that this task is not straightforward as the data is complex and multidimensional, making it difficult to recognize patterns and relationships.»

2.2.2 Walker's Cognitive Walk-Through

The full cognitive walk-through can be found in the appendix (My visual analytics solution). My visual analytics solution consists of a GlueViz session with two tabs. Tab 1 focuses on the ADNI CSF biomarker data, while Tab 2 is centered around gut metabolites utilizing a clustering technique to allow analytics on this high-dimensional dataset.



Figure 9 Tab 1: Visual Analytics on ADNI CSF data

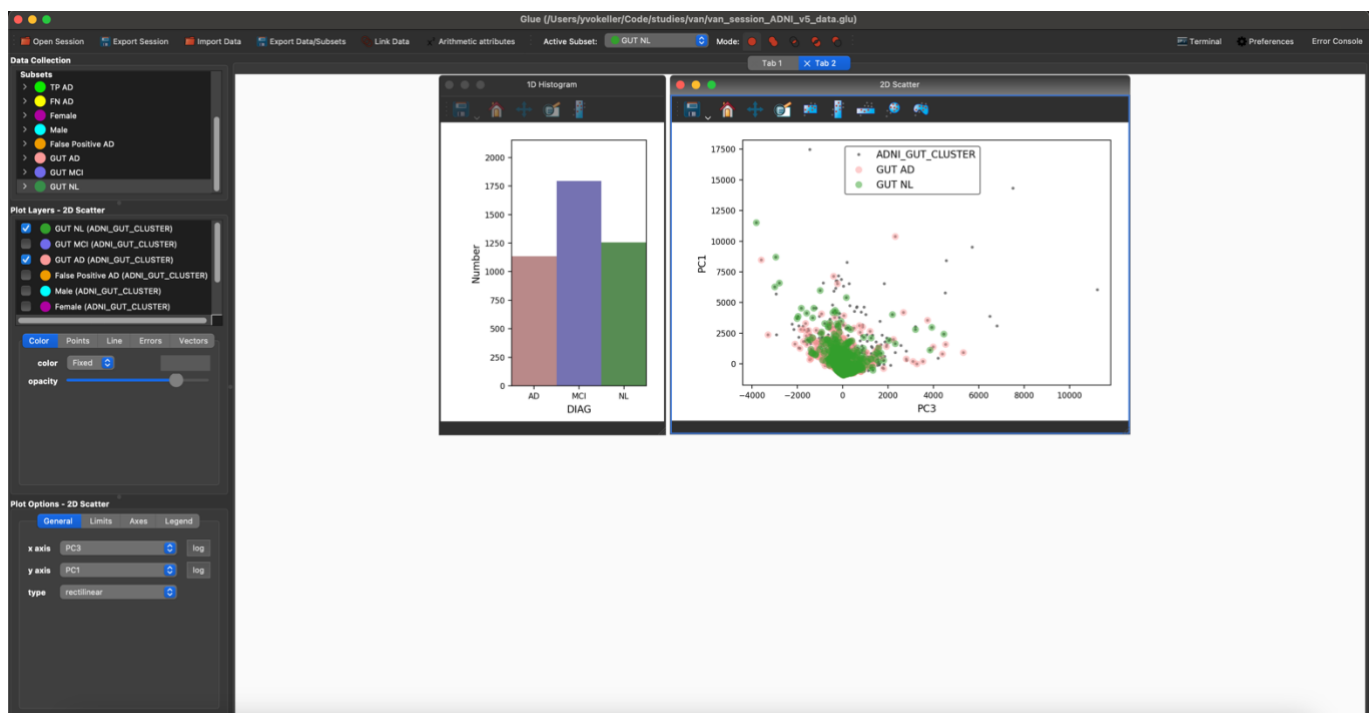


Figure 10 Tab 2: Visual Analytics on ADNI Gut Metabolite data

Cognitive walk-through). While I only did one iteration of a cognitive walkthrough, in practice this is an iterative process (Nielsen Norman Group, 2023d). «My first step is to open the prepared session, containing the imported datasets, their linking configuration and predefined relevant plots (Figure 5).»

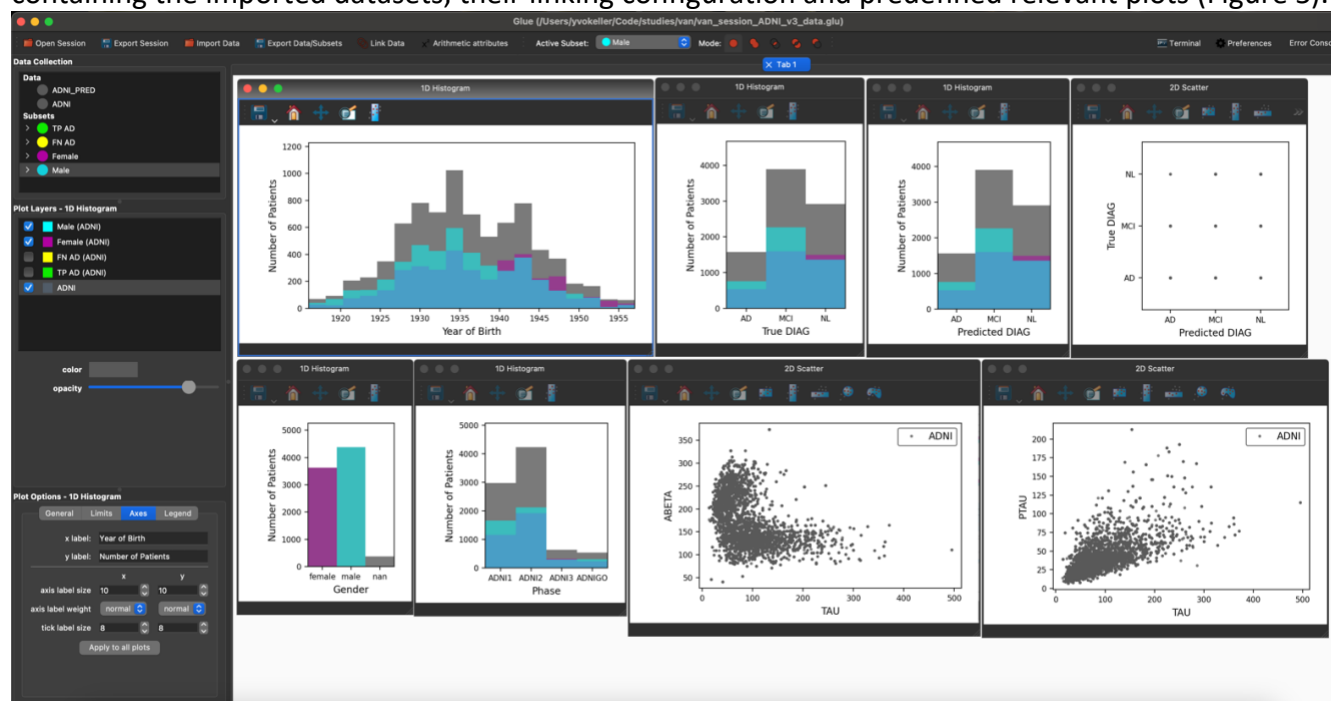


Figure 5 The visual analytics session for ADNI CSF data in GlueViz

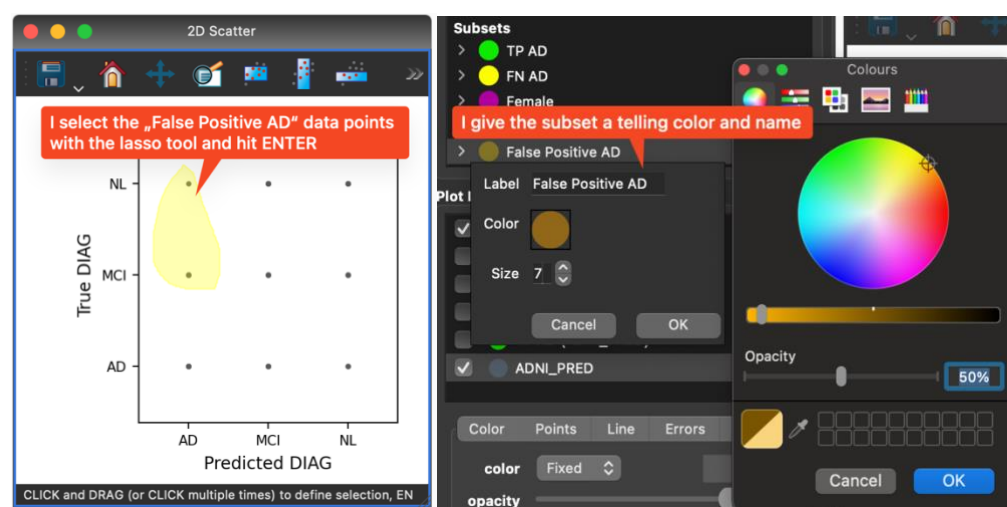


Figure 7 Creating a new subset in GlueViz Figure 6 Configuring label and color of a subset in GlueViz

data points across multiple visualizations. Then, I assign a distinct color to this subset (Figure 6) to visually highlight these data points in any plot. Now, the scatter plots of the ABETA, TAU, and PTAU biomarkers provide a visual representation of the distribution and relations of these biomarkers against the computed diagnosis. The color-coded data points stand out, allowing me to quickly identify all the falsely classified AD cases (Figure 8). As I examine these plots, I start to form hypotheses about why these data points were falsely classified as AD. Perhaps these cases have biomarker levels that are unusually high for their true classification, causing the model to misclassify them.»



Figure 8 The GlueViz visual analytics session, with a subset of False Positive AD classifications selected.

2.3 Results and takeaways

The cognitive walk-through showed that generally the interface of GlueViz in combination with my plots works well for getting insights on this data. However, there are multiple areas of improvement I noticed: (1) the number of plots shown on one screen is on the upper limit with eight. (2) changing the color of a subset should be more user friendly. (3) while the 2D-scatter plot provides a good starting point, it would be a lot easier to gain insight with 3D plots and additional statistical visualizations, for example a grouped boxplot showing the distribution of the seemingly important feature TAU for one or more subsets and the whole dataset. To address (1), I would reassess whether all plots are needed, and if yes, utilize the tabs feature of GlueViz to show less on the same tab. (2) and (3) are both things that could be addressed, but need software development effort. Adding such a statistical group plot feature would mean developing a whole new GlueViz plugin.

3 LO3: Tools

To get to know the technology behind other possible visual analytics platforms and be able to make a fair comparison, I researched GlueViz and the two other most popular solutions Microsoft PowerBI and Tableau on eight criteria that I identified to be relevant for a visual analytics use case like mine.

3.1 GlueViz

CRITERIA	GLUEVIZ
EASE OF USE	GlueViz is a Python library designed for interactive data visualization. It is built with "data-hacking" workflows in mind and can be run locally (GlueViz, 2023c). Everything is based around a session, in which you import relevant datasets, create links between them and then add different plots to the workspace. A more advanced way of interacting with Glue is via its Python library. (GlueViz, 2023a)
DATA HANDLING CAPABILITIES	It allows for easy linking of data over a primary key and supports the creation of subsets of data. However, it does not provide a way to refresh the data when the data schema changes, requiring re-importing of data and recreation of all the plots. It only supports a very limited amount of data sources, most of which rely on a separate plugin (GlueViz, 2023d).
VISUALIZATION CAPABILITIES	It supports the creation of scatter plots, histograms, and images (2D and 3D) of data. However, it has a limited range of visualization types, and some, like 3D scatter plots, do not work properly (anymore?) (GlueViz, 2023b).
INTERACTIVITY	It is focused on the brushing and linking paradigm, where selections in any graph propagate to all others. This allows for a high degree of interactivity in the visualizations.
INTEGRATION WITH MACHINE LEARNING	It does not have built-in machine learning capabilities. The easiest way to implement computation is using Python a machine learning library like Scikit-Learn and saving the resulting predictions as part of the data you are importing.
REAL-TIME ANALYSIS	GlueViz does not offer real-time data analysis as of now.
COST	GlueViz is an open-source tool, so it is free to use.
SUPPORT AND DOCUMENTATION	GlueViz has a documentation page, but it lacks detailed documentation. It does not appear to be actively maintained, which raises concerns about the availability of support.

3.2 Tableau

CRITERIA	TABLEAU
EASE OF USE	Tableau offers an intuitive drag-and-drop interface that allows users to uncover hidden insights quickly, even when offline. It is designed to be user-friendly and is suitable for users of all skill levels. It is available on mobile in addition to web and desktop, although that version has some limitations. (Biswal, 2023)
DATA HANDLING CAPABILITIES	Tableau provides robust data handling capabilities. It allows users to connect to data on-premises or in the cloud, and then combine and clean the data with clicks, not code. There are many connectors available out-of-the-box (Tableau, 2023a).
VISUALIZATION CAPABILITIES	Tableau supports a wide range of visualization types and allows users to build and iterate on visualizations with a drag-and-drop experience and dynamic previews. It offers more advanced charting options than other solutions like PowerBI (Biswal, 2023).
INTERACTIVITY	Tableau is designed for real-time data exploration with live visual analytics (Tableau, 2023b). It provides a high degree of interactivity and interactive visualization filter options (Tableau, 2023b).
INTEGRATION WITH MACHINE LEARNING	Tableau integrates AI and machine learning into its platform to democratize what-if scenario planning, guided model building, AI-powered predictions, and other data science techniques. It also allows for the integration and visualization of live results from R, Python, Einstein Discovery, MATLAB, and other extensions. (Tableau, 2023d)
REAL-TIME ANALYSIS	Tableau supports real-time data exploration with visual analytics. Still, real-time is usually limited to around 10-15 minutes delay (Tableau, 2023e).
COST	Tableau is a commercial product with various pricing plans. It's plans start at 70\$ per user for the "Tableau Creator" subscription. (Tableau, 2023c)

SUPPORT AND DOCUMENTATION

Tableau provides detailed documentation and has a strong community of users who can provide support. Still, I found their documentation to not be as straightforward to understand as for example the one of PowerBI.

3.3 PowerBI

CRITERIA	POWERBI
EASE OF USE	Power BI is a data analytics platform originally built for enterprise business intelligence by Microsoft. The user interface is not as intuitive and takes some time to learn, but it allows users to connect to and visualize data from a variety of sources. It is available in a mobile, a web and a desktop version. (Biswal, 2023)
DATA HANDLING CAPABILITIES	Power BI provides robust data handling capabilities. It allows users to connect to hundreds of on-premises and cloud data sources such as Dynamics 365, Azure SQL Database, Salesforce, Excel, and SharePoint (davidiseminger, 2023b). However, it is more tightly integrated to and focused on Microsoft products than other solutions like Tableau (Biswal, 2023).
VISUALIZATION CAPABILITIES	Power BI supports over 23 different visualization types, from a basic scatter plot to tree maps. (mihart, 2023)
INTERACTIVITY	Power BI provides a high degree of interactivity. It's "visual interactions" feature allows for linked visualizations, for example selecting a region on a map cross-filters the data in the other visualizations. (maggiesMSFT, 2022)
INTEGRATION WITH MACHINE LEARNING	Power BI integrates AI and machine learning into its platform to democratize what-if scenario planning, guided model building, AI-powered predictions, and other data science techniques. It also allows for the integration and visualization of live results from R, Python, Einstein Discovery, MATLAB, and other extensions (otarb, 2023).
REAL-TIME ANALYSIS	Power BI supports real-time data exploration with live visual analytics. It provides the tools needed for real-time insights, from data preparation to dashboard updates. (davidiseminger, 2023a)
COST	Power BI is a closed-source commercial product with a variety of pricing plans, starting from 9.80 CHF per user/month. Reports within the desktop app can be created for free (PowerBI, 2023).
SUPPORT AND DOCUMENTATION	Power BI is a successful product, therefore it is actively maintained and provides detailed documentation and has a strong community of users.


3.4 Side-by-side comparison

CRITERIA	GLUEVIZ	TABLEAU	POWERBI
EASE OF USE	3	5	5
Data Handling Capabilities	2	5	5
Visualization Capabilities	2	5	5
Interactivity	5	5	5
Integration with Machine Learning	2	5	5
Real-time Analysis	1	4	5
Cost	5	2	3

Support and Documentation	2	5	5	To quantitatively compare the tools against each other, I assigned 0 to 5 points in each of the 8 categories
Total (of 40)	22	36	38	

based on the above evaluation. PowerBI is the clear winner, with Tableau close behind, and quite a bit behind, GlueViz. What do these results mean in practice, and for my visual analytics scenario with ADNI?

I had quite many usability issues with GlueViz. Just to name some: All plot configurations (like labels) are not persisted within the session export. Changing the data schema means you need to build the session with all plots & subsets from scratch to get updated data. I was also really missing a working 3D-Scatter plot. Some built-in grouping capability to calculate statistics like mean, std would also have been helpful. What worked well and what I would use it again for, is the whole concept of linked visualizations where subsets are propagated to all other plots of the same data.

At this stage of my knowledge on visual analytics and available tools, I would try to implement the same solution with PowerBI. I am quite sure that the  experience would be better, connected visualizations (cross-filters) on a whole report are available too, and additionally I could possibly go as far as integrating with Azure Machine Learning to provide an option to directly train a new model version with adjusted hyperparameters from the dashboard. This would address the iterative manner of computational methods.

4 LO4: Evaluation

There are many methods one can consider for performing usability testing to evaluate a visual analytics solution. To find out what is the right one for my case, I started by researching some of them to get a rough understanding and orientation in the space.

4.1 An overview of the usability testing methods

One option is to perform a heuristic evaluation for example using **Nielsen's Heuristics**. These include principles like "visibility of system status", e.g., showing the user where he currently is on a map, or "match between system and the real world", e.g., avoiding technical terms but instead speaking the user's natural language and providing information in logical order (Nielsen Norman Group, 2023c).

Performing a **cognitive walkthrough** is another option. This is what I already did as part of LO2. Key idea here was to simulate a new user's cognitive processes, thoughts and path of exploration when learning to use the tool (Nielsen, 1994).

Another method I found is the **Think-Aloud Protocol**. Users are asked to say out loud what they are thinking as they move through the user interface, which can help identify what they find confusing or difficult ("Think Aloud Protocol," 2023).

User Testing, on the other hand, involves observing users as they use the system, typically while they are performing specific tasks. This can provide insights into how users interact with the system, what they find intuitive, and where they encounter difficulties (Rubin, 2008).

Another option would be **Eye Tracking**. This method involves tracking where the user's gaze lands within the interface. This can provide insights into what users find most attention-grabbing and how they navigate the interface (Bojko, 2013).

Alternatively, I could write a **Survey** or conduct an **Interview** to gather user feedback on various aspects of the system, such as ease of use, satisfaction, and perceived usefulness and any other improvement suggestions the users might have.

More on the technical side and especially good to gain quantitative results, **Analytics** and **Log Files** can provide data on how users interact with a system, including what features they use, how long they spend on certain tasks, and where they encounter errors. For a visual analytics tool built on the web, there are tools like Google Analytics that could be set up and would then provide daily stats on interactions of the users within the tool. This is an interesting option but is surely not the right approach for performing a user test in the moment and within a python environment like GlueViz.

4.2 Performing the usability test

I chose to perform the usability test of my visual analytics solution with the **Think-Aloud Protocol**, as this method allows me to interact with the user during the test and learn from their thoughts on the tool.

The **process** can be summarized as follows ("Think Aloud Protocol," 2023):

- Design the study and write the guide.
- Recruit the participants.
- Conduct think-aloud protocol.
- Analyze the findings and summarize insights.

During the test, the **moderator** should adhere to the following **guidelines** for the best results (Nielsen Norman Group, 2023a):

- Explain the process and what the participant awaits beforehand.
- Encourage continuous verbalization during the test.
- Stay neutral to not influence the participants actions.
- Take notes on key observations and issues.

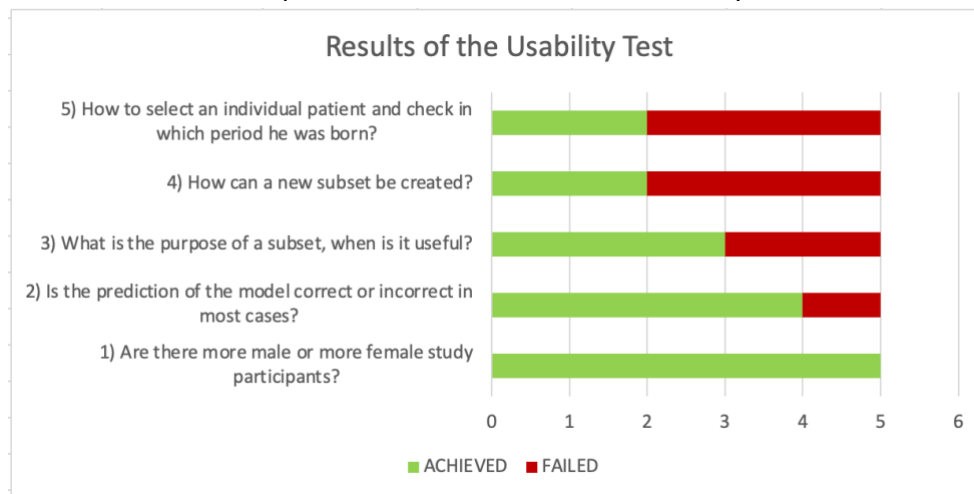
4.2.1 Study Design

I will conduct my study with five participants, as per Nielsen the 6th participant would usually mostly repeat what previous testers already commented on. At this point, in practice I would implement changes according to the results, and then perform another round of usability testing (Nielsen Norman Group, 2023b, p. 5). Within the scope of this module and the restricted timeframe, I will perform only one iteration.

As I have limited resources for this project, I recruited participants within my family and friends. Again, in practice I would make sure the cohort exists of people that will actually work with the visual analytics solution I am building and are familiar with the domain knowledge behind the data being analyzed. I will give each user the same set of prepared questions. To be able to answer them, they will need to understand the user interface of my GlueViz solution and interact with the control elements of the app.

4.2.2 Study Results

I visualized the tasks posed to the user and whether they were able to fulfill them. For task (1), users



had no problem spotting the correct chart and reading out the information. (2) required some more understanding of how such classification works, but most users were able to answer it. (3) the purpose of a subset was not clear for 2 users. They did not understand the relation subsets have to the original dataset. (4)

only two users were able to create a new subset themselves, as this requires interaction with the toolbar and selecting the correct tool on a plot. (5) needed the users to zoom in as much as that a single patient datapoint became visible on the y-axis of histogram, which only 2 managed to do. This showed to be not intuitive at all. Independently of the specific question, users really struggled with the initial understanding of the user interface. For somebody not familiar with the data and software, there are many different controls, and the UI consists of 5 different sections (Global Toolbar, Data Collection, Plot Layers, Plot Options, and the Working Area). There is quite a learning curve between the user seeing the interface the first time and being able to confidently interact with it. This could be addressed with a Product Tour guiding new users through the UI (Whatfix, 2023). The fact that Glue mixes colors of different subsets when they overlay each other (e.g., in the Histogram) also showed to be really confusing to the interpretation. Finally, the interpretation possibilities with GlueViz are limited, as there is no way to click a single datapoint and view the data behind it.

4.3 Conclusion and suggestions for improvement

Reviewing the results not only of LO4 but my overall work, it becomes clear to me that the visual analytics solution as it is implemented now works for what it is intended. Still, there remains much potential to improve upon. Unfortunately, to implement most of the improvements, I would need to rewrite much of the GlueViz interface. This would mean lots of development effort and is not worth pursuing, as GlueViz seems to not really be actively maintained anymore. Rather, in the future I would opt for a different solution like Microsoft PowerBI, which should be better up to the task, as I assessed in LO3: Tools. This still won't allow for all kinds of changes, so if one has the resources to support it, a custom development that allows for full customization and the overall best user experience could be considered.

5 Bibliography

- Alzheimer's Disease Neuroimaging Initiative. (2023, June 13). *ADNI | Alzheimer's Disease Neuroimaging Initiative*. <https://adni.loni.usc.edu/>
- Andrienko, N., Andrienko, G., Fuchs, G., Slingsby, A., Turkay, C., & Wrobel, S. (2020). Computational Techniques in Visual Analytics. In N. Andrienko, G. Andrienko, G. Fuchs, A. Slingsby, C. Turkay, & S. Wrobel, *Visual Analytics for Data Scientists* (pp. 89–147). Springer International Publishing. https://doi.org/10.1007/978-3-030-56146-8_4
- Biswal, A. (2023, July 6). *Microsoft Power BI Vs Tableau: Here are the Difference You Should Know | Simplilearn*. Simplilearn.Com. <https://www.simplilearn.com/tutorials/power-bi-tutorial/power-bi-vs-tableau>
- Blennow, K., Vanmechelen, E., & Hampel, H. (2001). CSF total tau, A β 42 and phosphorylated tau protein as biomarkers for Alzheimer's disease. *Molecular Neurobiology*, 24(1), 87. <https://doi.org/10.1385/MN:24:1-3:087>
- Bojko, A. (2013). *Eye Tracking the User Experience: A Practical Guide to Research*. Cognitive load. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Cognitive_load&oldid=1157571238
- davidiseminger. (2023a, January 12). *Real-time streaming in Power BI - Power BI*. <https://learn.microsoft.com/en-us/power-bi/connect-data/service-real-time-streaming>
- davidiseminger. (2023b, May 23). *Data sources in Power BI Desktop—Power BI*. <https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-data-sources>
- Endert, A. (2023, June 15). *The State of the Art in Integrating Machine Learning into Visual Analytics*. Ar5iv. <https://ar5iv.labs.arxiv.org/html/1802.07954>
- GlueViz. (2023a, February 6). *Starting Glue from Python*. https://docs.glueviz.org/en/stable/python_guide/glue_from_python.html
- GlueViz. (2023b, May 14). *Getting started*. https://docs.glueviz.org/en/v0.7.x/getting_started/index.html#plotting-data
- GlueViz. (2023c, May 18). *Installing and running glue*. <https://docs.glueviz.org/en/stable/installation/installation.html>
- GlueViz. (2023d, May 27). *Frequently Asked Questions*. <https://docs.glueviz.org/en/stable/faq.html>
- Greitzer, F. L., Noonan, C. F., & Franklin, L. (2011). *Cognitive Foundations for Visual Analytics* (PNNL-20207, 1013936; p. PNNL-20207, 1013936). <https://doi.org/10.2172/1013936>
- Introspection. (2023). In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=Introspection&oldid=1159907487>
- Li, L. (2019, May 27). *Principal Component Analysis for Dimensionality Reduction*. Medium. <https://towardsdatascience.com/principal-component-analysis-for-dimensionality-reduction-115a3d157bad>
- maggiesMSFT. (2022, November 23). *Change how visuals interact in a report—Power BI*. <https://learn.microsoft.com/en-us/power-bi/create-reports/service-reports-visual-interactions>
- mihart. (2023, March 17). *Visualization types in Power BI - Power BI*. <https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-types-for-reports-and-q-and-a>
- Nielsen, J. (1994). *Usability Engineering*. Morgan Kaufmann.
- Nielsen Norman Group. (2023a, June 15). *Checklist for Moderating a Usability Test*. Nielsen Norman Group. <https://www.nngroup.com/articles/usability-checklist/>
- Nielsen Norman Group. (2023b, June 16). *Why You Only Need to Test with 5 Users*. Nielsen Norman Group. <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>
- Nielsen Norman Group. (2023c, December 6). *10 Usability Heuristics for User Interface Design*. Nielsen Norman Group. <https://www.nngroup.com/articles/ten-usability-heuristics/>
- Nielsen Norman Group. (2023d, December 6). *How to Conduct a Cognitive Walkthrough Workshop*. Nielsen Norman Group. <https://www.nngroup.com/articles/cognitive-walkthrough-workshop/>

otarb. (2023, January 18). *Create Power BI visuals using Python in Power BI Desktop—Power BI*.
<https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-python-visuals>

PowerBI. (2023, May 6). *Preis- und Produktvergleich*. <https://powerbi.microsoft.com/de-ch/pricing/>

Rubin, J. (2008). *Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests*.

Tableau. (2023a, June 13). *Supported Connectors*. https://help.tableau.com/current/pro/desktop/en-us/exampleconnections_overview.htm

Tableau. (2023b, June 15). *Filter Data from Your Views*. Filter Data from Your Views.
<https://help.tableau.com/current/pro/desktop/en-us/filtering.htm>

Tableau. (2023c, June 15). *Pricing for data people*. Tableau. <https://www.tableau.com/pricing/teams-orgs>

Tableau. (2023d, December 6). *AI + Analytics*. Tableau. <https://www.tableau.com/solutions/ai-analytics>

Tableau. (2023e, December 6). *Realtime Streaming of Data through Tableau Dashboard*. Realtime Streaming of Data. <https://community.tableau.com/s/question/0D58b00009gVHgeCAG/realtime-streaming-of-data-through-tableau-dashboard>

The Magical Number Seven, Plus or Minus Two. (2022). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=The_Magical_Number_Seven,_Plus_or_Minus_Two&oldid=1127506856

Think aloud protocol. (2023). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Think_aloud_protocol&oldid=1151214818

Whatfix. (2023, June 16). *Ultimate Guide to Creating Product Tours in 2023*.
<https://website.whatfix.com/product-tour/>

6 Appendix

6.1 My visual analytics solution

My visual analytics solution consists of a GlueViz session with two tabs. Tab 1 focuses on the ADNI CSF biomarker data, while Tab 2 is centered around gut metabolites utilizing a clustering technique to allow analytics on this high-dimensional dataset.



Figure 9 Tab 1: Visual Analytics on ADNI CSF data

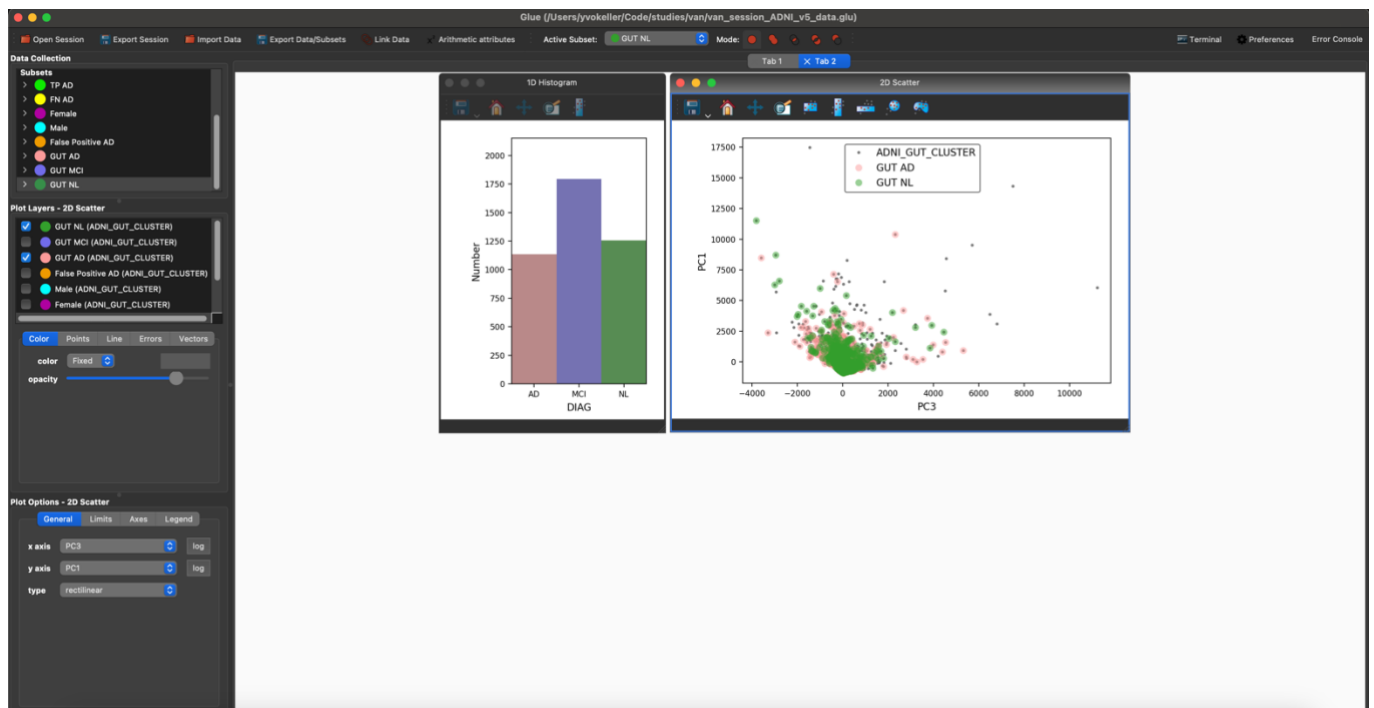


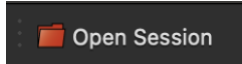
Figure 10 Tab 2: Visual Analytics on ADNI Gut Metabolite data

6.2 Cognitive walk-through

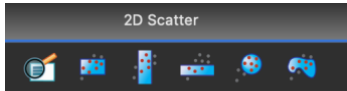
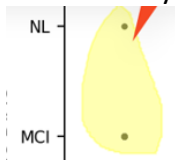
User's Goal

As Walker, a medical researcher focused on Alzheimer's disease (AD), my goal is to understand the characteristics of the data points that have been falsely classified as AD by a Decision Tree Model.

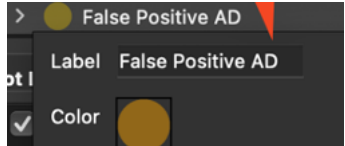

Task 1: Open the prepared session in GlueViz, which contains the imported datasets, their linking configuration, and predefined relevant plots.

Questions	Result
Will the user try to achieve the right result?	Yes
Will the user notice that the correct action is available?	Yes, the button «Open Session» is clearly visible on the GlueViz UI 
Will the user associate the correct action with the effect he is trying to achieve?	Yes, the buttons action is clearly labeled
If the correct action is performed, will the user see that progress is being made toward the solution of the task?	Yes, once the session is opened, the datasets and plots will be visible.

Task 2: Open the prepared session in GlueViz, which contains the imported datasets, their linking configuration, and predefined relevant plots.

Questions	Result
Will the user try to achieve the right result?	Yes, he knows he needs to create a subset of falsely classified AD cases.
Will the user notice that the correct action is available?	Yes, the different subset selection tools like lasso are clearly visible and available in GlueViz. 
Will the user associate the correct action with the effect he is trying to achieve?	Yes, he understands that using the lasso tool will allow him to create the subset.
If the correct action is performed, will the user see that progress is being made toward the solution of the task?	Yes, once the subset is created, it will be highlighted in the confusion matrix, and during its creation the currently selected area is always shown. 

Task 3: Assign a distinct color to this subset (Figure 6).

Questions	Result
Will the user try to achieve the right result?	Yes, he knows he needs to assign a distinct color to the subset to easily identify it.
Will the user notice that the correct action is available?	Somewhat unintuitive, he must double-click the subset for these settings to appear. 
Will the user associate the correct action with the effect he is trying to achieve?	Yes, he understands that assigning a color will make the subset easily identifiable in the plots.
If the correct action is performed, will the user see that progress is being made toward the solution of the task?	Yes, once the color is assigned, the subset will stand out in the plots. 

Task 4: Examine the scatter plots of the ABETA, TAU, and PTAU biomarkers to identify the falsely classified AD cases (Figure 8).

Questions	Result
Will the user try to achieve the right result?	Yes, he knows he needs to examine the scatter plots to understand the characteristics of the falsely classified AD cases.
Will the user notice that the correct action is available?	Yes, the scatter plots are clearly visible in the session.
Will the user associate the correct action with the effect he is trying to achieve?	Yes, he understands that examining the scatter plots will provide insights into the characteristics of the falsely classified AD cases.
If the correct action is performed, will the user see that progress is being made toward the solution of the task?	Yes, by examining the scatter plots, he can start to form hypotheses about why these data points were falsely classified as AD. However, some statistical charts like a boxplot be helpful, to allow for statistically based judgement and not just with the naked eye.