Intelligence

■ □ □

Goal: Identify the important decision points in a Visual Analytics system and assign their priorities

Deliverable: An inventory of decision points and their priorities

1. What is the <u>context of use</u> for the Visual Analytics system?

Outline the context in which your system is going to be used. Who are the users? What are their goals? What are the available resources and environments that limit their options? Talk to your users for this analysis, but you can also discuss internally with your team for this step.

Users	Goals and Tasks	Resources	Environment
Epidemiologists who have domain knowledge about the diseases under study and knowledge about different VA methods	Explore and identify different types of patients for the same disease. Data cleaning, processing, and finding patterns.	Limited manpower and computing power.	Visual interface More possible risks, as medical contexts could influence patients' lives

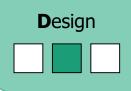## 2. What are the <u>decision points</u> in the analysis?

In the contexts of use you just outlined, brainstorm about decision points where users need to **make a decision between multiple alternatives**. Walk through the users' analytical process in each of the following components and consider where they need to choose between multiple options to move the analysis forward.

	Data	Algorithm	Visualization	Reasoning
Which	Which dimensions of the data to include?	Which dimension reduction and clustering algorithm to use for processing the data?	Which primary visual encoding to use to find the clusters?	Which type of clusters/patterns to discover -- e.g., to find different mechanisms that cause the same disease, or
How	How to clean up messy entries and make placeholders for the missing data?	How to parameterize the algorithms? e.g. number of clusters	How to specify the visualization to bring out the patterns more clearly?	How to combine different data dimensions and algorithms according to the reasoning -- e.g. how to combine dimension

3. <u>How much support</u> do users need in each decision point?

Reflect on if support (guidance) is needed for each decision point, and how much support it needs?
Rate the decision points with your preferred method and then combine them in the final assessment.

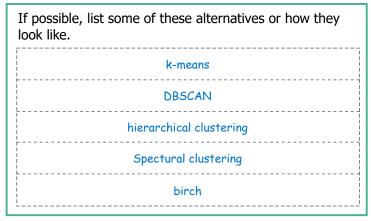| Decision Points	Probability of a "wrong" decision	Impact of getting it "wrong"	Number of potential alternatives	Final Assessment
which data features/dimensions	medium	high	medium	2 x 3 x 2 = 12
which clustering algorithm	**high**	**high**	**medium**	**3 x 3 x 2 = 18**
order of the algorithms	high	medium	low	3 x 2 x 1 = 6
which primary encoding	medium	medium	low	2 x 2 x 1 = 4
…	…	…	…	…

Goal: Design an evaluation model for the options/alternatives

Deliverable: Criteria as well as the model for the evaluation that generates guidance

1. Choose a decision point, and consider <u>what are the alternatives</u> to choose from?

Estimate the possible number of alternatives that should be considered by the guidance generation process. Think about where these alternatives come from and if they could or should be evaluated.

Scikit-learn has a handful of clustering algorithms -- around 10 were listed in their documentation for comparisons

If possible, list some of these alternatives or how they look like.

k-means

DBSCAN

hierarchical clustering

Spectural clustering

birch

2. How to produce the <u>criteria</u> for evaluating the alternatives?

Think about when users are choosing between these alternatives - **what criteria** their decisions are based on? And **on what basis** should these criteria be produced?

Based on:	Full results of each alternative	Partial samples of each alternative	Abstract features of each alternative	Human-rating of each alternative
Produced Criteria:	Full results might be too computationally expensive	Silhouette Coefficient Davies-Bouldin Index Calinski-Harabasz Index Rand index and mutual information based score can be used when the ground truth is available	Computation time	Subjective quality

3. How to <u>combine the criteria</u> into an evaluation model?

Summarize the **criteria from above**, choose the relevant one(s), and combine them together in a model through conditions, weighted sum, or both. Think about which criteria should be used in which model, how user should change these criteria and conditions/weights.

Which criteria to use? ⟶ What is the degree of guidance? ⟶ What is the level of user control?

Consider and list which criteria are clear and relevant.

Should the guidance be provided in orienting, directing, or prescribing degree?

Should the user change the criteria and the conditions/weights? And how should the changes be made?

Silhouette Coefficient
Davies-Bouldin Index
Calinski-Harabasz Index
Computation time

Mostly directing
Possibly some orienting

Users are experts, so they might want to change the criteria weights to their liking

<table>
<tr><td>

Choice

</td><td>

Goal: Presenting the alternatives to users to guide their decision

Deliverable: Design of the specification and presentation of the alternatives
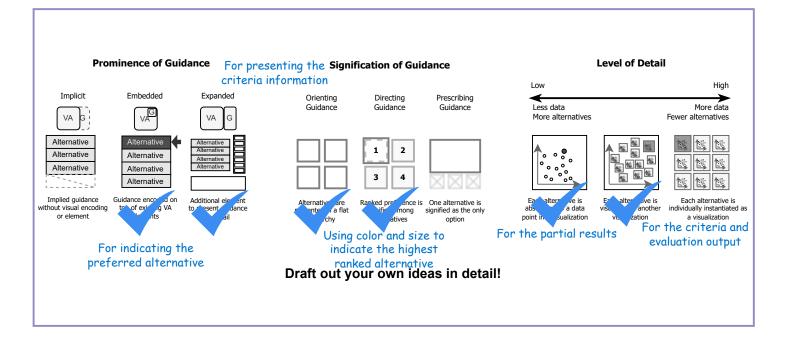
</td></tr>
</table>

1. What <u>information/data</u> about each alternative is relevant for users' choice?

Think about the **evaluation, criteria, and full/partial results of each alternative** that you have produced during the Design stage -- Which one(s) of them are relevant for helping users make their choice?

Data about the alternatives:	Evaluation output	Criteria in the evaluation	Full/partial results to produce the criteria
Data content:	Ranking of the alterntaives	Scores in the criteria: Silhouette Coefficient Davies-Bouldin Index Calinski-Harabasz Index Computation time	Partial results of each alternative are also available as labels for each patient

2. How should the information about the alternatives be <u>presented</u>?

Reflect back on the data and structure about the alternatives, brainstorm about how they should be presented. For example - think about **which encodings** work for the data, **what detail** each alternative should be shown in, and how to **signify the guidance**. Following are some examples as a thinking tool.

Prominence of Guidance

For presenting the criteria information — Signification of Guidance

Level of Detail

Implicit — Implied guidance without visual encoding or element

Embedded — Guidance encoded on top of existing VA elements

Expanded — Additional element to present guidance detail

For indicating the preferred alternative

Orienting Guidance — Alternatives are presented in a flat hierarchy

Directing Guidance — Ranked preference is signified among alternatives

Prescribing Guidance — One alternative is signified as the only option

Using color and size to indicate the highest ranked alternative

Low — Less data More alternatives — High — More data Fewer alternatives

Each alternative is absorbed a data point in visualization — For the partial results

Each alternative is visualized another visualization

Each alternative is individually instantiated as a visualization — For the criteria and evaluation output

Draft out your own ideas in detail!

3. How to adapt the guidance to <u>user feedback</u>?

How should different components -- **presence, generation, and presentation**, of the provided guidance adapt to user feedback? Should the adaptation be **implicitly inferred** from user interactions or **directly controlled** by users?

	Guidance Presence	Guidance Generation	Guidance Presentation
Implicitly Inferred	Possibly turn off the guidance when users do not follow the suggestions	Users' preferences among the alternatives can be inferred	Similar to presence -- the details can be cut down when users do not interact with them
Directly Controlled	User should be able to turn off some of the guidance	The criteria and weights can be changed	Details of the evaluation criteria can be hidden when not needed