In our dataset search homework, you may have experienced the challenges in dataset search - despite the many dataset repositories available, it is so difficult for novice users to find a dataset of interest. Now, it is your turn to change it!

Team: please form teams of 3 students to conduct this project.

Input to this project: the search logs recorded by the whole class; jason files containing meta data of more than 3.3k datasets from Dataverse. They are stored in files/dataset\_search\_project.

"Tasks": The search logs reflect the experiences of the students in dataset search. You can learn user requirements from the search logs and they are the input to you for requirement analysis.

"Data": You also need to analyze the jason files. They decide the visualizations you are able to build. You can follow the http links in the Jason files to access the original webpages of the datasets. They will help you understand the descriptors in the Jason file. You help users search datasets better by making use of the information in the Jason files.

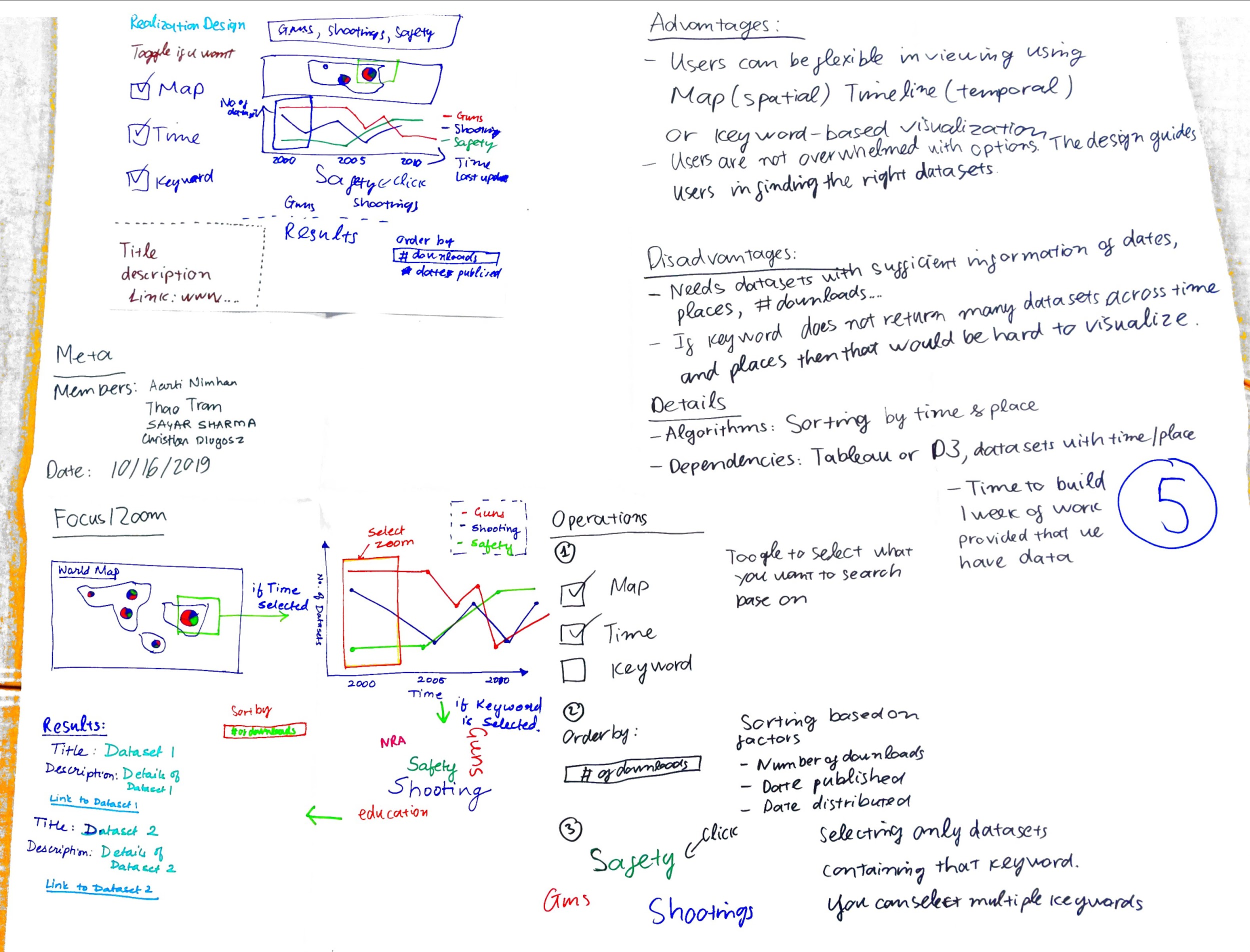
Output of this project:

1. Design of a visualization system that can provide better dataset search experience to users. Both tasks and data need to be considered while you are designing. Following the five sheet design methodology (see the file about it in the dataset search project folder) in the design process.

2. A working prototype that Implements one or three views in your design (at least one view is required if you implement your visualizations with D3 or other programming languages; at least three views are required if you generate your visualizations with Tableau).

3. Presentations: You are required to submit and present your designs in the class of 10/16 (we will work on teams to generate the designs in the class of 10/09), and represent a living demo of your implementation with real data (the jason files I provided to you. The bigcollection.zip should be used. The smallcollection.zip is for your testing and debugging) on 12/11.

Grading metrics: Design: 8 points; Implementation: 9 points. I will evaluate your results by their usefulness, novelty, usability, and implementation quality if you code by yourself.



1. Sheet 1: Geographical map including zoom and pan (Heavy task) -Aarti
2. Sheet 2: Timeline including zoom and pan (Heavy task) Sagar
3. Sheet 3: Word clouds with Filtering - Christian
4. Dashboard: Bring them all together : Listing results based on filters of: Map, Time, Keyword. Order by #of downloads… (Heavy task) Thao
5. Make a presentation in Powerpoint for in-class presentation - Christian

Use the small collection.zip to get the data to visualize