

Visualization and Visual Data Analysis

Nicole Cherches, Alexander Gelb, Benjamin Neckam, Axinya Tokareva

December 10, 2017

1 M3 Hi-Fi Prototyping

1.1 Task

Since the GAIA satellite is observing million of stars and scientists only focusing on a small subset of the data, we wanted to provide a tool which gives information about the whole dataset. This should include finding patterns, correlations, outliers or maybe some other useful information.

1.2 Current visualization design

Pictures of program here

1.3 M2 use cases iteration

No idea what is meant

1.4 Changes

After the feedback of M2 and a meeting with Mr. Möller we had to rethink our approach. We focused too much on details and requirements our customers mentioned and lost focus of providing a big picture of the data.

Between M2 and M3 we were building a program for a very specific task and not for exploring the data. Mr. Möller indicated us that we were on a wrong track and gave us some input to adapt it to the needs of the "Vis" class. One of the biggest changes is that we will not provide a 3D representation of the stars itself because this would be more a "Computer graphics problem" and not a "Visualization-problem". Therefore we decided to offer a scatterplot, scatterplotmatrix, barplot and a correlogram at the moment.

With these type of plots the user should be able to explore the data and gain interesting information about it.

Since the dataset is huge with many columns, we got the hint to use "Principal Component Analysis" to see patterns in the data and reduce columns.

1.5 Major challenges and problems

The biggest challenge of all is to handle the amount of data. Not only the dataset consists of almost 2 million stars, a single star also has 58 features (columns). At the moment it is challenging to filter out data and create a useful and meaningful plot out of it. Furthermore plotting so many information has a bad performance and is very time consuming in D3.