# Visualization and Visual Data Analysis

Nicole Cherches (01506832), Alexander Gelb (01268620),
Benjamin Neckam (01301917), Axinya Tokareva (01368965)

December 13, 2017

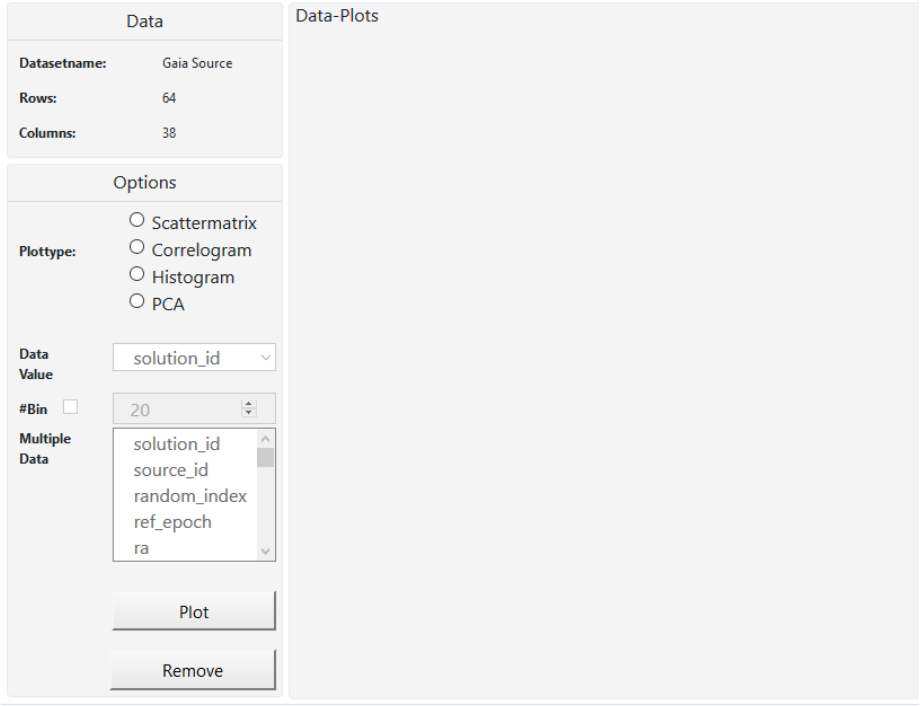
## M3 Hi-Fi Prototyping

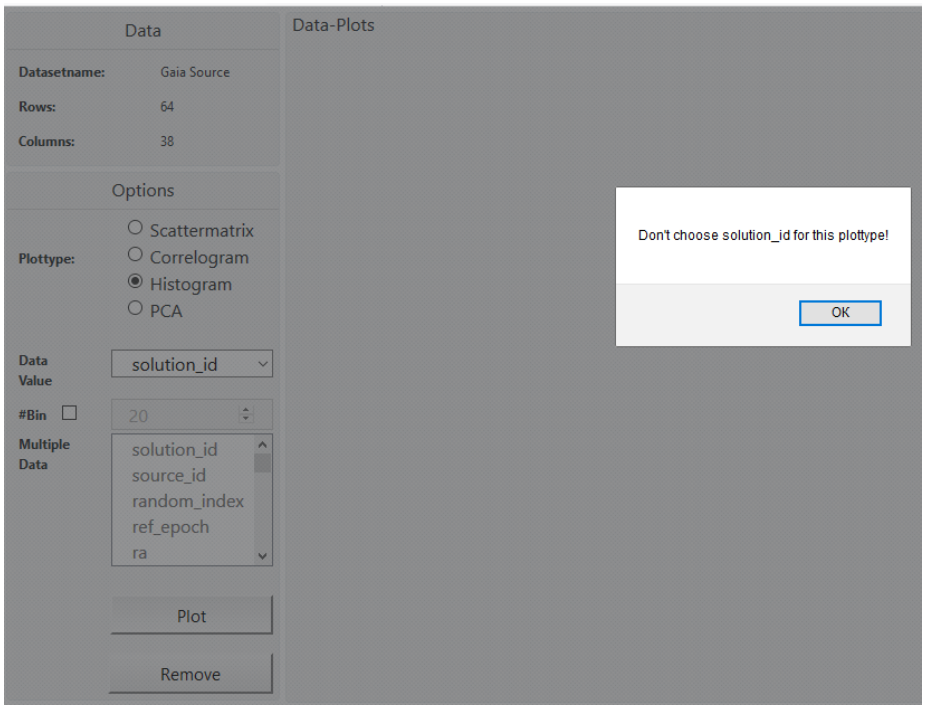
## 1 Task

Since the GAIA satelite is oberving millions of stars and scientists are only focusing on a small subset of the data, we wanted to provide a tool which gives information about the whole dataset. We didn't really know much about the data, so we couldn't tell anything about the data itself, also it was too complicated for us to understand. So we focused on exploring the data by finding patters, correlations, outliers or maybe some other useful information.

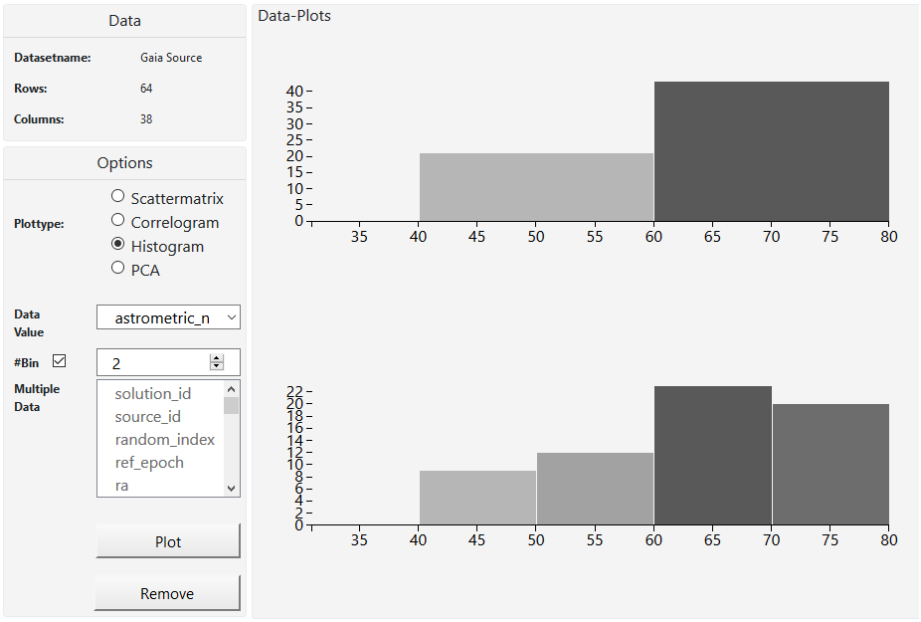
## 2 Current visualization design

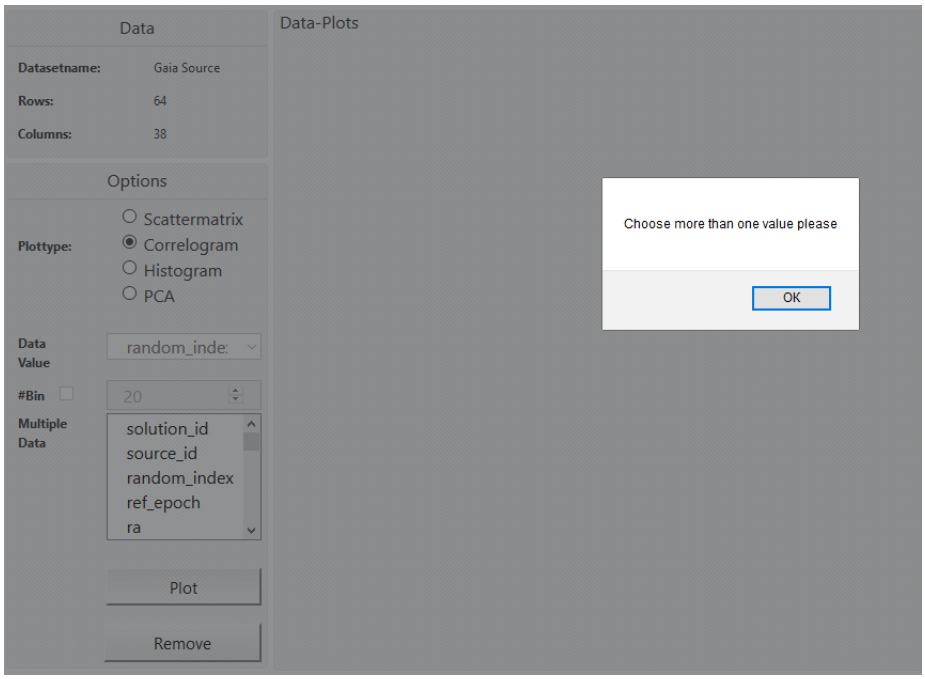
TODO another images

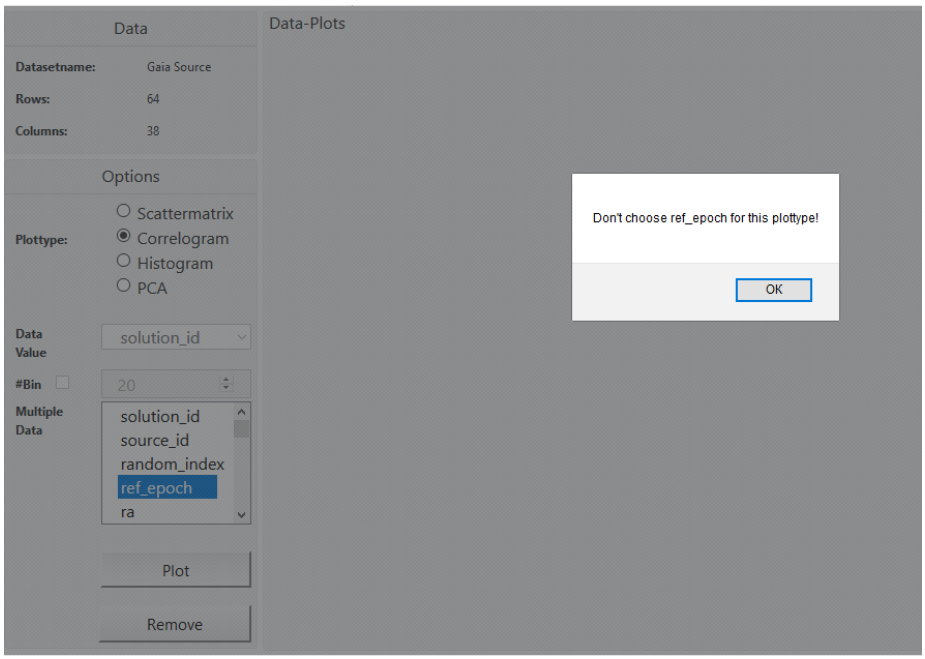

Data
Datasetname:
Gaia Source
Rows:
64
Columns:
38
Options
Plottype:
Scattermatrix
Correlogram
Histogram
PCA
Data Value
solution_id
#Bin
20
Multiple Data
solution_id
source_id
random_index
ref_epoch
ra
Plot
Remove
Data-Plots

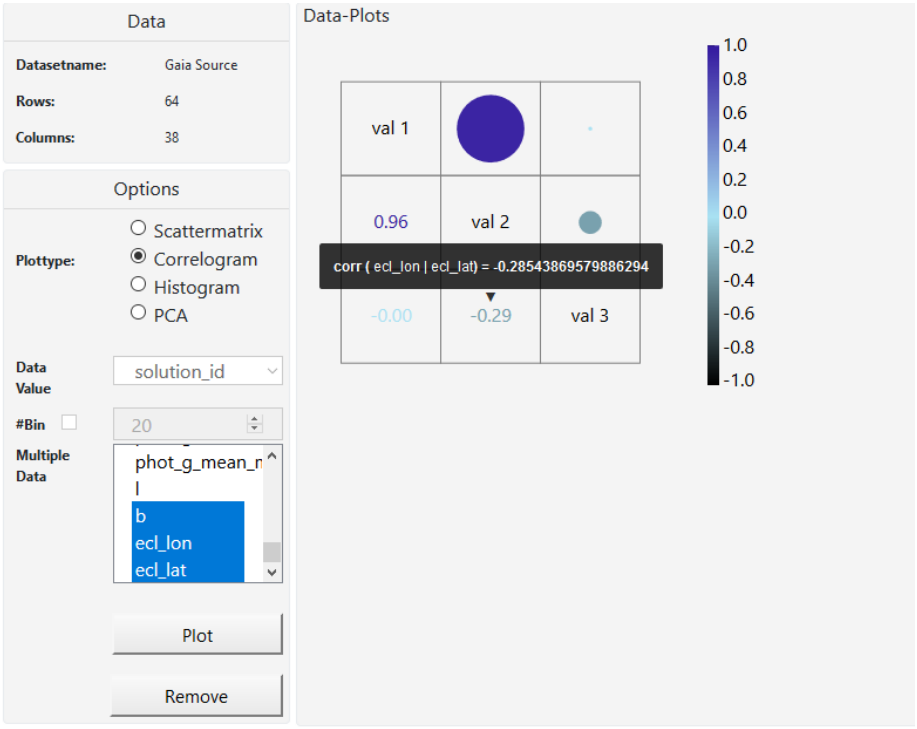

Data
Datasetname: Gaia Source
Rows: 64
Columns: 38
Options
Plottype:
Scattermatrix
Correlogram
Histogram
PCA
Data Value
solution_id
#Bin
20
Multiple Data
solution_id
source_id
random_index
ref_epoch
ra
Plot
Remove
Data-Plots
Don't choose solution_id for this plottype!
OK

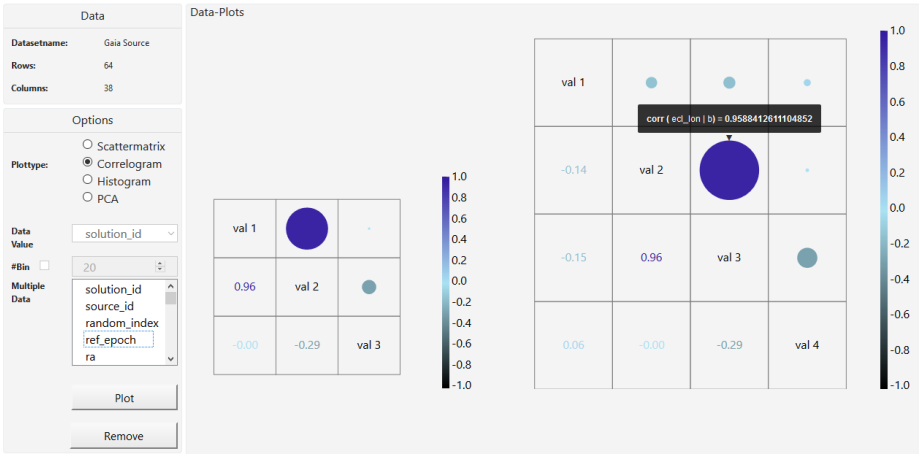
Data
Datasetname: Gaia Source
Rows: 64
Columns: 38
Options
Plottype:
Scattermatrix
Correlogram
Histogram
PCA
Data Value
matched_obs
#Bin
20
Multiple Data
solution_id
source_id
random_index
ref_epoch
ra
Plot
Remove
Data-Plots
matched_observations: 29
25
20
15
10
5
0
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0
10.5
11.0


Data
Datasetname: Gaia Source
Rows: 64
Columns: 38
Options
Plottype:
Scattermatrix
Correlogram
Histogram
PCA
Data Value
astrometric_n
#Bin
2
Multiple Data
solution_id
source_id
random_index
ref_epoch
ra
Plot
Remove
Data-Plots
40
35
30
25
20
15
10
5
0
35 40 45 50 55 60 65 70 75 80
22
20
18
16
14
12
10
8
6
4
2
0
35 40 45 50 55 60 65 70 75 80


Data
Datasetname: Gaia Source
Rows: 64
Columns: 38
Options
Plottype:
Scattermatrix
Correlogram
Histogram
PCA
Data Value
random_inde:
#Bin
20
Multiple Data
solution_id
source_id
random_index
ref_epoch
ra
Plot
Remove
Data-Plots
Choose more than one value please
OK


Data
Datasetname:
Gaia Source
Rows:
64
Columns:
38
Options
Plottype:
Scattermatrix
Correlogram
Histogram
PCA
Data Value
solution_id
#Bin
20
Multiple Data
solution_id
source_id
random_index
ref_epoch
ra
Plot
Remove
Data-Plots
Don't choose ref_epoch for this plottype!
OK


Data
Datasetname: Gaia Source
Rows: 64
Columns: 38
Options
Plottype: Scattermatrix
Correlogram
Histogram
PCA
Data Value solution_id
#Bin 20
Multiple Data
phot_g_mean_r
l
b
ecl_lon
ecl_lat
Plot
Remove
Data-Plots
val 1
0.96
val 2
corr ( ecl_lon | ecl_lat) = -0.2854386957886294
-0.00
-0.29
val 3
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0


Data
Datasetname: Gaia Source
Rows: 64
Columns: 38
Options
Plottype: Scattermatrix
Correlogram
Histogram
PCA
Data Value solution_id
#Bin 20
Multiple Data
solution_id
source_id
random_index
ref_epoch
ra
Plot
Remove
Data-Plots
val 1
0.96
val 2
-0.00
-0.29
val 3
corr ( ecl_lon | b) = 0.9588412611104852
-0.14
-0.15
0.06
val 4
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0

# 3 M2 use cases iteration

In M2 we described our use case scenario. We wanted to create a visualization tool for Joao Alves, who is an astrophysics professor at the university of Vienna. The main goal to achieve was a tool that gives astronomers a bigger picture of the universe because they only look at small areas most of the time. So we mainly focused on that and started with making a scatterplot. After that we decided to create a scatterplot matrix and histograms which shows how a subset of the data works together. A whole matrix with scatterplots helps us to look at many columns from our set at the same time, better than just one scatterplot. We made a correlogram which gives us a look on how the data in our subset or in the whole dataset correlates. What we didn't do, is analyzing clusters of stars or locating them. Our task is to help the user to find as many correlations in the data as possible (with the correlogram) and then study this data in a more detailed way (with histograms, scatterplots and PCA). Also we didn't have the time to create interaction between the plots, so the user is only able to view one plot at a time. Also the user cannot save. We will try to develop this further in M4.

# 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 hug with many columns, we got the hint to use "Principal Component Analysis" to see patterns in the data and reduce columns.

# 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.

# 6 VIS techniques

- correlation over multiple var
- filter
- tooltips
- zoom
- brushing and linking

# 7 The work distribution

TODO

- Nicole Cherches —
- Alexander Gelb —
- Benjamin Neckam —
- Axinya Tokareva —

# 8 References


1. Our website: http://wwwlab.cs.univie.ac.at/~a1368965/vis17/
2. https://medienportal.univie.ac.at/uniview/professuren/cv/artikel/univ-prof-dr-joao-alves/
3. vis lectures
4. http://gaia.ari.uni-heidelberg.de/tap/tables
5. http://sci.esa.int/gaia/
6. http://sci.esa.int/gaia/58275-data-release-1/
7. https://en.wikipedia.org/wiki/
8. http://www.mattboldt.com/demos/typed-js/
9. another links in README