Exploratory analysis

March 18, 2021

[91]:
```
%reload_ext autoreload
%autoreload 2
default_figsize=(14,12)
```

[92]:
```
import datasets
import numpy as np
import pandas as pd
import seaborn as sn
import matplotlib.pyplot as plt
import matplotlib
matplotlib.rcParams['figure.figsize'] = (14, 12)

dataset_name = "mcswain"
dataset_module = datasets.datasets_by_name_all[dataset_name]
x,y,metadata = dataset_module.load(dropna=True,verbose=True)
y = datasets.map_y_em(y,dataset_name)

# generate dataframe with both x and y
xy = pd.concat([x,y],axis=1)
xy.describe()
```

```
Warning loading data from McSwain2005-2009_VPHAS-2MASS.csv:
Dropped 2313 rows with missing values.
Rows (original):   5455
Rows (after drop): 3142
```

[92]:

	umag	gmag	rmag	imag	Hamag
count	3142.000000	3142.000000	3142.000000	3142.000000	3142.000000
mean	16.304806	15.668695	14.651512	14.076811	14.398883
std	1.572447	1.225343	1.047182	1.035455	1.041201
min	12.260000	12.530000	11.990000	11.450000	11.690000
25%	15.150000	14.722500	13.842500	13.260000	13.580000
50%	16.445000	15.740000	14.700000	14.100000	14.435000
75%	17.337500	16.650000	15.510000	14.900000	15.250000
max	20.840000	19.050000	17.500000	17.090000	17.170000

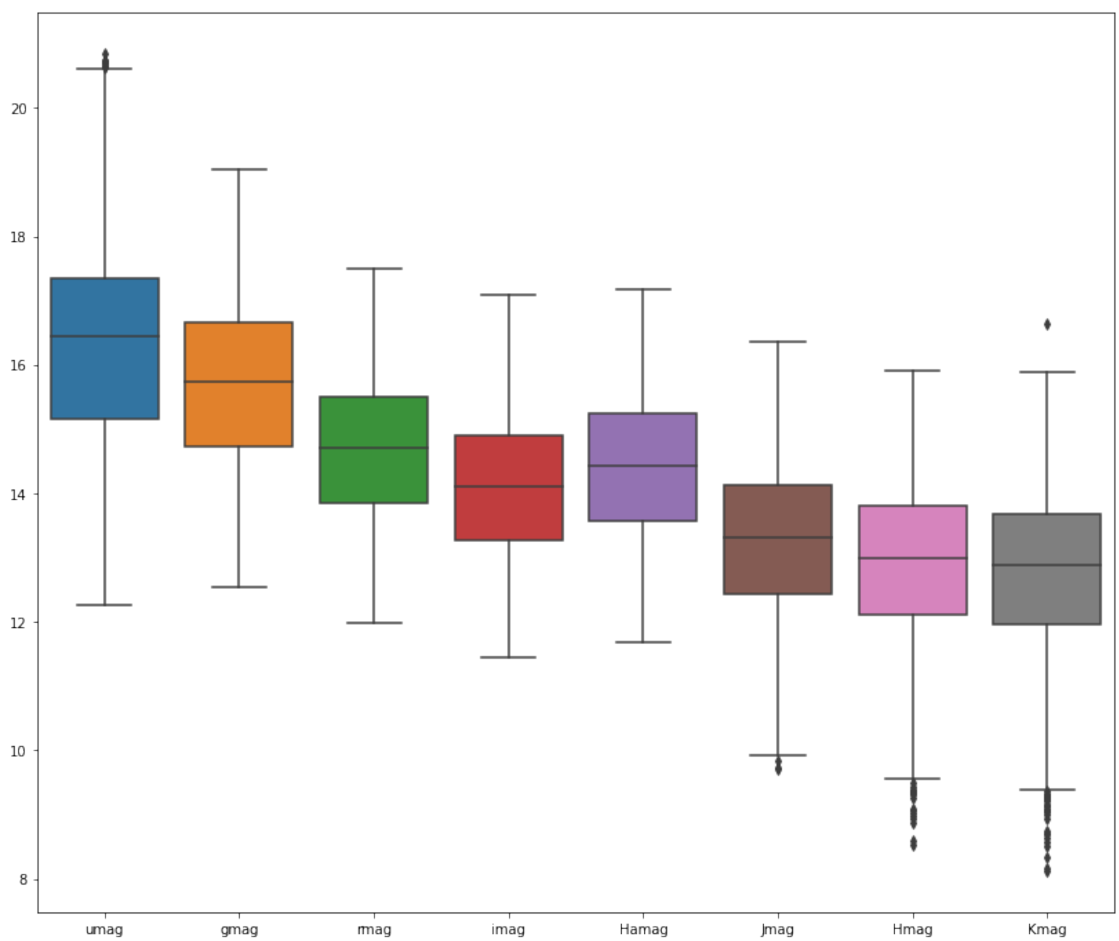
	Jmag	Hmag	Kmag	em
count	3142.000000	3142.000000	3142.000000	3142.000000

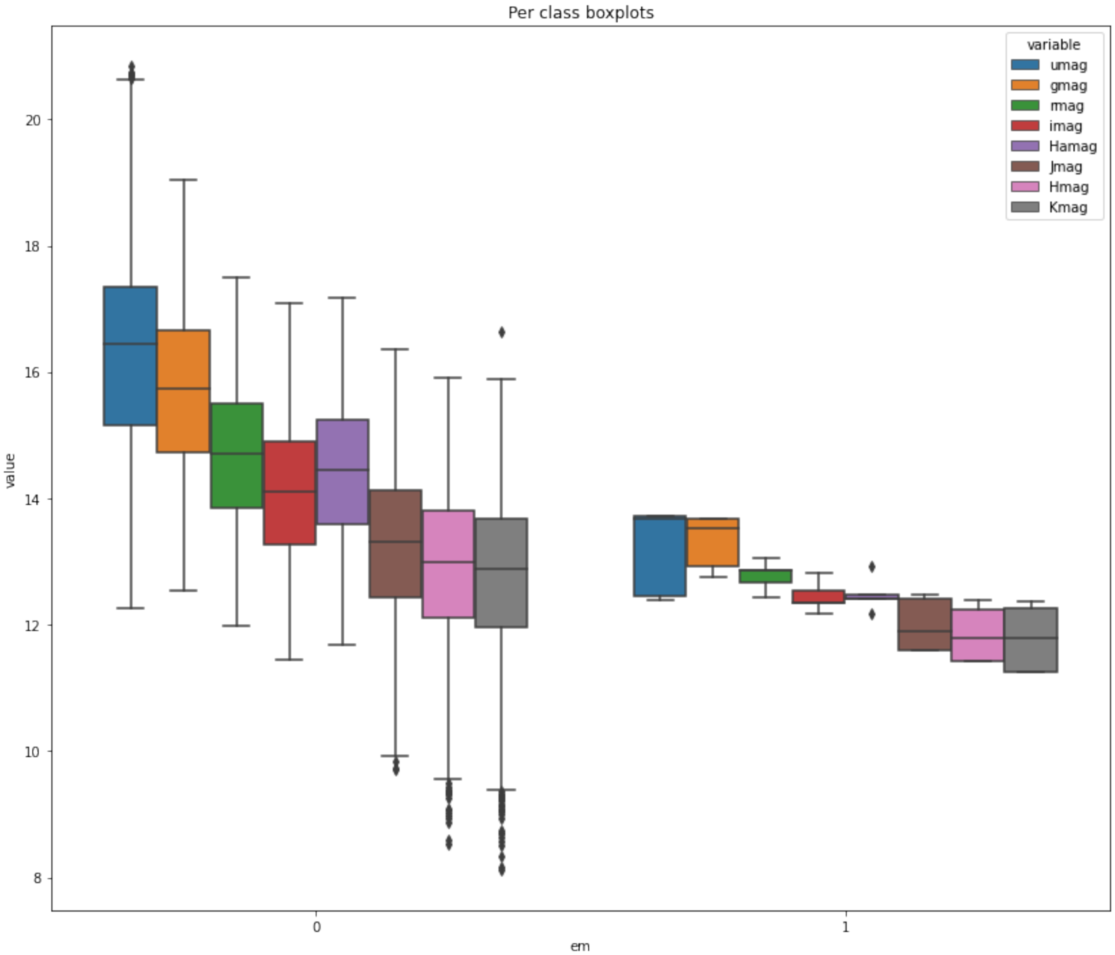
```
mean     13.248104   12.876838   12.739313    0.001591
std       1.128721    1.221364    1.259059    0.039866
min       9.700000    8.521000    8.118000    0.000000
25%      12.434000   12.101000   11.957000    0.000000
50%      13.306000   13.000000   12.876000    0.000000
75%      14.126750   13.801000   13.682000    0.000000
max      16.368000   15.912000   16.631000    1.000000
```

1 Variable visualization

[93]:
```python
sn.boxplot(data=x)

plt.figure()
xy_long = pd.melt(xy, id_vars='em')
sn.boxplot(x='em', y='value', hue='variable', data=xy_long)
plt.title("Per class boxplots")
```

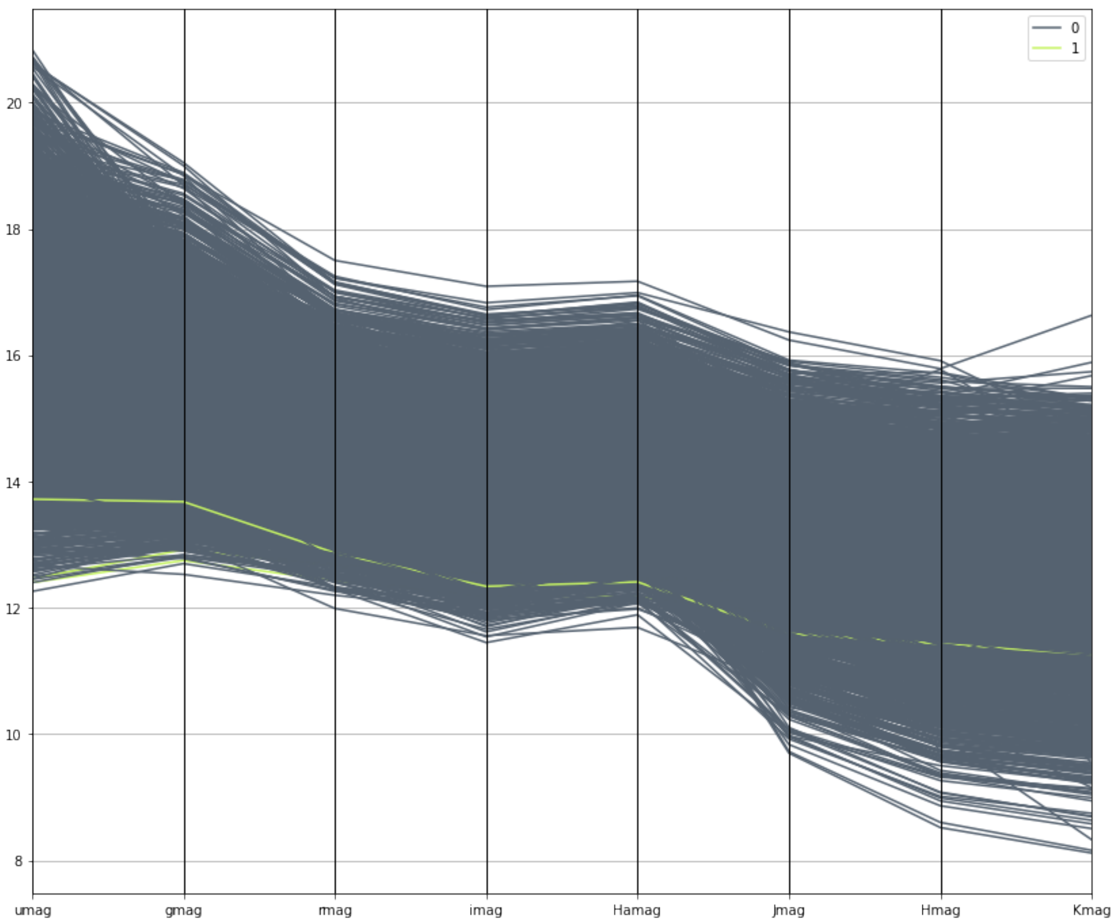
[93]: Text(0.5, 1.0, 'Per class boxplots')





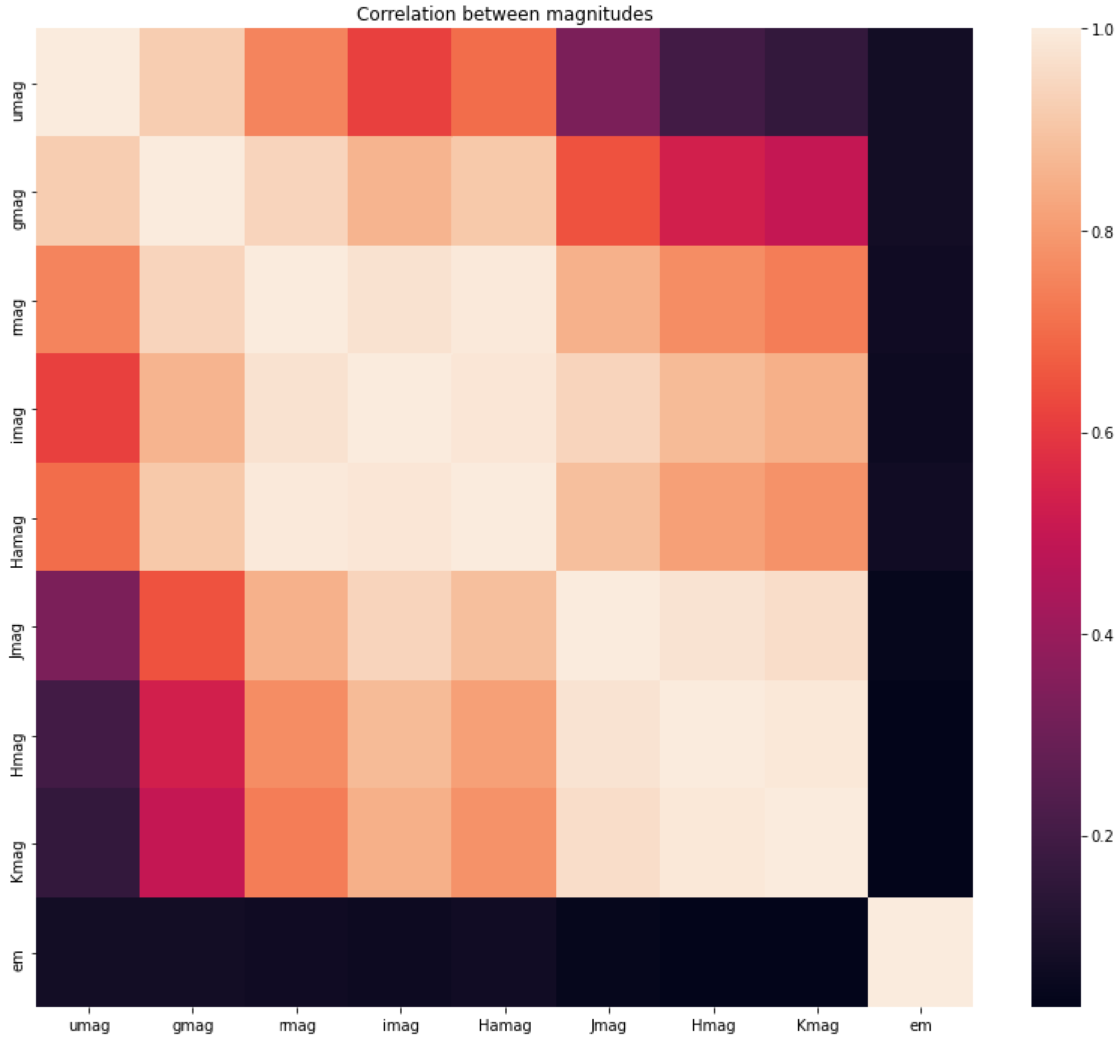

```
[94]: pd.plotting.parallel_coordinates(xy,"em",color=('#556270','#C7F464'))
```

```
[94]: <AxesSubplot:>
```

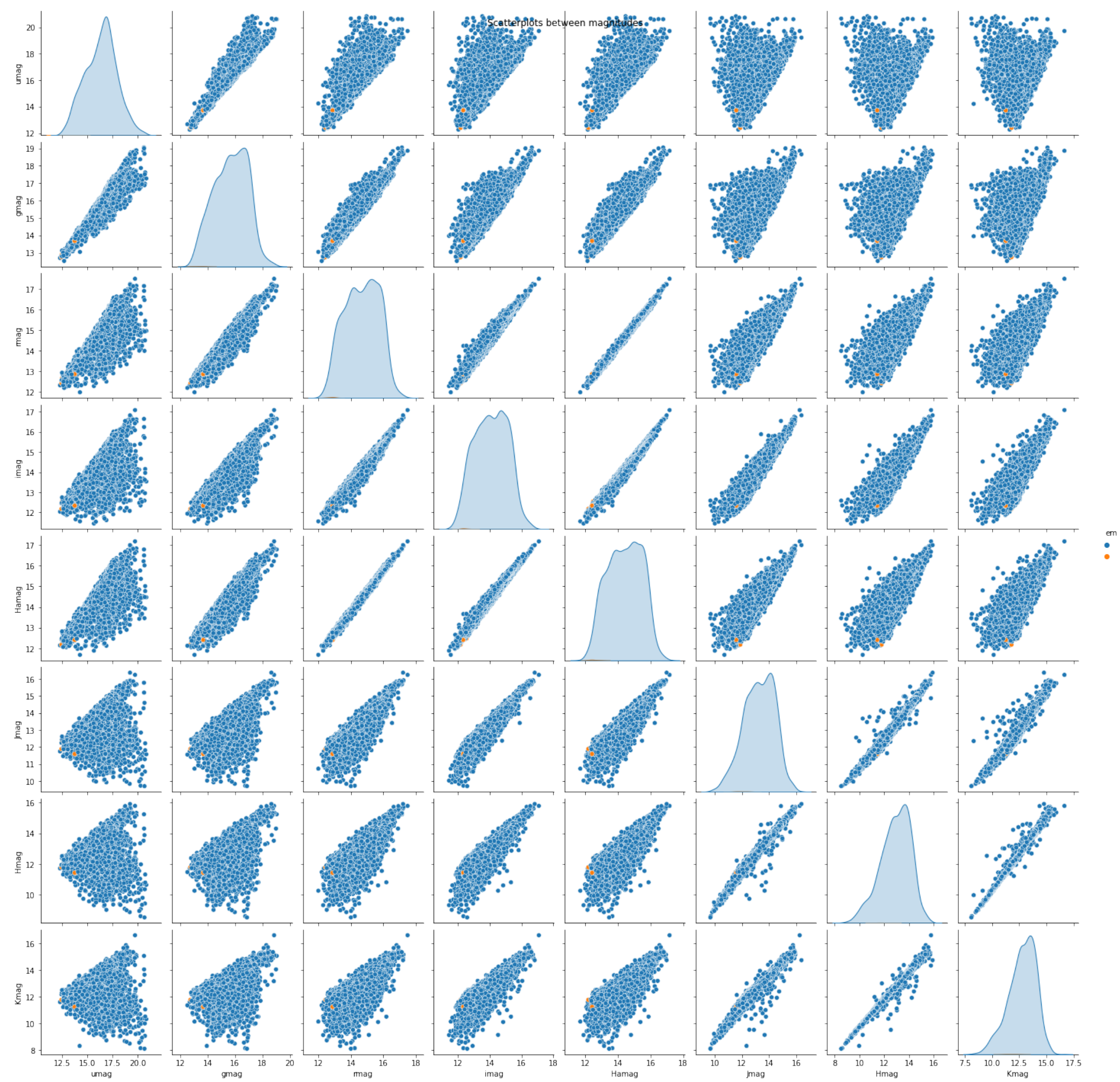


[95]:

```
sn.heatmap(xy.corr().abs())
plt.title("Correlation between magnitudes")
plt.show()

sn.pairplot(xy,hue="em")
plt.suptitle("Scatterplots between magnitudes")
# axes=pd.plotting.scatter_matrix(x,c=y["em"],alpha=0.
↪9,grid=False,figsize=(14,12))
```



[95]: Text(0.5, 0.98, 'Scatterplots between magnitudes')



2 Outlier detection via confidence interval

[96]:
```python
from scipy import stats
m = len(x.columns) # number of columns = number of hypothesis
confidence= 0.98
adjusted_confidence = 1- (1-confidence)/m  # bonferroni-adjusted confidence
max_zscore = stats.norm.ppf(adjusted_confidence)
print(f"Confidence  (desired): {confidence}")
print(f"Confidence (adjusted): {adjusted_confidence}")
print(f"Z-score    (adjusted): {max_zscore}")

indices = (np.abs(stats.zscore(x-x.mean())) > max_zscore).any(axis=1)
outliers_x = x[indices]
```

```
if dataset_name != "all_em":
    outliers_metadata = metadata[indices]
    outliers_x = outliers_x.
 ↪merge(outliers_metadata,left_index=True,right_index=True)
outliers_x
```

```
Confidence  (desired): 0.98
Confidence (adjusted): 0.9975
Z-score    (adjusted): 2.807033768343811
```

[96]:

	umag	gmag	rmag	imag	Hamag	Jmag	Hmag	Kmag	Rad	\
240	19.72	18.86	17.50	17.09	17.17	16.239	15.786	16.631	NaN	
1272	16.87	14.44	12.51	11.62	12.13	9.924	9.015	8.747	NaN	
1345	18.19	15.56	13.42	12.18	12.96	10.290	9.387	9.042	NaN	
1462	19.19	15.99	13.72	12.08	13.14	9.836	8.866	8.505	NaN	
1533	20.84	17.27	14.97	13.54	14.47	11.570	10.644	10.342	NaN	
1539	20.73	17.59	15.46	14.16	15.01	12.216	11.319	11.070	NaN	
1570	16.64	14.45	12.64	11.76	12.22	10.061	9.265	8.993	NaN	
1836	20.73	17.62	15.27	13.97	14.77	11.809	10.792	10.536	NaN	
1967	20.71	16.97	14.51	12.96	13.92	10.553	9.426	9.080	NaN	
1994	18.69	15.60	13.52	12.09	13.01	10.053	9.082	8.692	NaN	
2163	20.27	16.75	14.24	12.35	13.66	9.725	8.603	8.161	NaN	
2415	15.63	13.97	12.29	11.45	11.89	10.025	9.318	9.145	NaN	
2437	17.55	14.91	12.91	11.91	12.47	10.235	9.362	9.122	NaN	
2463	16.69	14.61	12.67	11.70	12.25	10.084	9.333	9.075	NaN	
2483	14.21	13.95	13.34	12.93	13.15	12.348	9.971	8.329	NaN	
2661	18.26	15.52	13.27	12.22	12.79	10.289	9.340	9.070	NaN	
2678	15.84	14.97	13.10	11.84	12.66	9.967	9.502	9.238	NaN	
2683	18.88	15.78	13.35	12.23	12.90	9.957	8.990	8.636	NaN	
2699	16.36	16.05	15.11	14.53	14.80	13.671	10.260	9.155	NaN	
2718	20.66	16.88	14.00	12.59	13.46	9.700	8.521	8.118	NaN	
3015	20.29	16.89	14.10	12.64	13.60	10.106	9.070	8.703	NaN	
3028	20.14	16.66	14.01	12.48	13.43	9.984	8.943	8.580	NaN	
3102	20.42	16.94	14.25	12.54	13.67	10.398	9.341	8.945	NaN	

	_DEJ2000	…	Cluster	vsini	e_Hmag	e_rmag	e_Rad	e_y-Ha	\
240	-25.483167	…	Haffner 16	NaN	0.126	0.05	NaN	0.115	
1272	-57.660417	…	IC 2581	NaN	0.022	0.00	NaN	0.073	
1345	-5.765472	…	Basel 1	NaN	0.022	0.00	NaN	0.064	
1462	-5.966361	…	Basel 1	NaN	0.042	0.00	NaN	0.065	
1533	-5.886528	…	Basel 1	NaN	0.026	0.01	NaN	0.088	
1539	-5.854250	…	Basel 1	NaN	0.023	0.01	NaN	0.092	
1570	-58.287500	…	NGC 3293	NaN	0.023	0.00	NaN	0.064	
1836	-60.696194	…	Trumpler 20	NaN	0.021	0.00	NaN	0.101	
1967	-60.611472	…	Trumpler 20	NaN	0.022	0.00	NaN	0.075	
1994	-60.631167	…	Trumpler 20	NaN	0.025	0.01	NaN	0.067	
2163	-60.574639	…	Trumpler 20	NaN	0.018	0.01	NaN	0.072	

```
2415 -47.466806  …     NGC 6200   NaN   0.024   0.00   NaN   0.050
2437 -47.544444  …     NGC 6200   NaN   0.024   0.00   NaN   0.061
2463 -47.478972  …     NGC 6200   NaN   0.026   0.00   NaN   0.058
2483 -46.933944  …     NGC 6204   NaN     NaN   0.01   NaN   0.102
2661 -44.725972  …     NGC 6249   NaN   0.023   0.00   NaN   0.057
2678 -33.431944  …  Trumpler 27   NaN   0.024   0.00   NaN   0.092
2683 -32.467667  …  Trumpler 28   NaN   0.029   0.00   NaN   0.081
2699 -32.420083  …  Trumpler 28   NaN     NaN   0.00   NaN   0.091
2718 -32.443972  …  Trumpler 28   NaN   0.034   0.00   NaN   0.098
3015  -8.509889  …  Trumpler 34   NaN   0.023   0.00   NaN   0.047
3028  -8.389750  …  Trumpler 34   NaN   0.023   0.00   NaN   0.078
3102  -5.930500  …      Basel 1   NaN   0.022   0.01   NaN   0.084

      e_Kmag    _RAJ2000  SB?  e_gmag
240      NaN  117.481417  NaN    0.01
1272   0.021  157.007667  NaN    0.01
1345   0.024  282.029125  NaN    0.00
1462   0.021  282.104792  NaN    0.00
1533   0.025  282.153042  NaN    0.01
1539   0.023  282.156125  NaN    0.01
1570   0.019  158.856375  NaN    0.00
1836   0.023  189.831292  NaN    0.02
1967   0.020  189.910875  NaN    0.01
1994   0.018  189.925042  NaN    0.01
2163   0.029  190.020042  NaN    0.01
2415   0.025  250.958125  NaN    0.00
2437   0.023  251.046333  NaN    0.00
2463   0.023  251.131750  NaN    0.01
2483     NaN  251.693333  NaN    0.01
2661   0.020  254.576000  NaN    0.00
2678   0.023  264.084292  NaN    0.00
2683   0.026  264.121042  NaN    0.00
2699     NaN  264.186458  NaN    0.01
2718   0.031  264.235167  NaN    0.01
3015   0.020  280.074042  NaN    0.00
3028   0.023  280.084833  NaN    0.01
3102   0.021  281.981750  NaN    0.01

[23 rows x 43 columns]
```

3 Outlier detection via IQR

[97]:
```
iqr_factor=1.5
q25,q75=x.quantile(0.25),x.quantile(0.75)
iqr=q75-q25
min_values = q25-iqr_factor*iqr
```

```
max_values = q75+iqr_factor*iqr
# ou
indices = (np.logical_or(x<min_values,x>max_values)).any(axis=1)
outliers_x = x[indices]
if dataset_name != "all_em":
    outliers_metadata = metadata[indices]
    outliers_x = outliers_x.
 ↪merge(outliers_metadata,left_index=True,right_index=True)
outliers_x
```

[97]:

	umag	gmag	rmag	imag	Hamag	Jmag	Hmag	Kmag	Rad	\
165	19.95	16.73	14.29	12.98	13.77	10.709	9.637	9.278	NaN	
235	18.90	16.00	13.75	12.54	13.20	10.597	9.569	9.236	NaN	
240	19.72	18.86	17.50	17.09	17.17	16.239	15.786	16.631	NaN	
256	20.66	18.69	16.47	15.72	16.01	14.595	13.883	13.670	NaN	
470	20.62	19.00	16.91	16.25	16.49	15.197	14.309	14.071	NaN	
906	20.11	16.77	14.16	12.97	13.82	10.665	9.660	9.270	NaN	
1272	16.87	14.44	12.51	11.62	12.13	9.924	9.015	8.747	NaN	
1345	18.19	15.56	13.42	12.18	12.96	10.290	9.387	9.042	NaN	
1395	19.78	16.67	14.23	12.84	13.77	10.572	9.561	9.217	NaN	
1405	19.27	16.12	13.86	12.60	13.40	10.574	9.658	9.369	NaN	
1462	19.19	15.99	13.72	12.08	13.14	9.836	8.866	8.505	NaN	
1533	20.84	17.27	14.97	13.54	14.47	11.570	10.644	10.342	NaN	
1539	20.73	17.59	15.46	14.16	15.01	12.216	11.319	11.070	NaN	
1570	16.64	14.45	12.64	11.76	12.22	10.061	9.265	8.993	NaN	
1577	15.15	14.38	12.88	12.04	12.51	10.323	9.570	9.301	NaN	
1836	20.73	17.62	15.27	13.97	14.77	11.809	10.792	10.536	NaN	
1967	20.71	16.97	14.51	12.96	13.92	10.553	9.426	9.080	NaN	
1994	18.69	15.60	13.52	12.09	13.01	10.053	9.082	8.692	NaN	
2163	20.27	16.75	14.24	12.35	13.66	9.725	8.603	8.161	NaN	
2415	15.63	13.97	12.29	11.45	11.89	10.025	9.318	9.145	NaN	
2437	17.55	14.91	12.91	11.91	12.47	10.235	9.362	9.122	NaN	
2463	16.69	14.61	12.67	11.70	12.25	10.084	9.333	9.075	NaN	
2483	14.21	13.95	13.34	12.93	13.15	12.348	9.971	8.329	NaN	
2661	18.26	15.52	13.27	12.22	12.79	10.289	9.340	9.070	NaN	
2674	16.20	14.57	12.89	11.99	12.51	10.388	9.647	9.247	NaN	
2678	15.84	14.97	13.10	11.84	12.66	9.967	9.502	9.238	NaN	
2683	18.88	15.78	13.35	12.23	12.90	9.957	8.990	8.636	NaN	
2699	16.36	16.05	15.11	14.53	14.80	13.671	10.260	9.155	NaN	
2713	16.57	14.70	12.85	12.04	12.47	10.359	9.581	9.304	NaN	
2718	20.66	16.88	14.00	12.59	13.46	9.700	8.521	8.118	NaN	
2750	18.60	15.82	13.67	12.50	13.21	10.497	9.607	9.329	NaN	
3015	20.29	16.89	14.10	12.64	13.60	10.106	9.070	8.703	NaN	
3027	20.69	17.51	14.94	13.62	14.48	11.284	10.356	10.017	NaN	
3028	20.14	16.66	14.01	12.48	13.43	9.984	8.943	8.580	NaN	
3102	20.42	16.94	14.25	12.54	13.67	10.398	9.341	8.945	NaN	

	_DEJ2000	…	Cluster	vsini	e_Hmag	e_rmag	e_Rad	e_y-Ha	\
165	-20.711139	…	NGC 2421	NaN	0.022	0.00	NaN	0.069	
235	-20.649944	…	NGC 2421	NaN	0.024	0.00	NaN	0.064	
240	-25.483167	…	Haffner 16	NaN	0.126	0.05	NaN	0.115	
256	-25.517722	…	Haffner 16	NaN	0.022	0.01	NaN	0.089	
470	-25.544778	…	Haffner 16	NaN	NaN	0.01	NaN	0.099	
906	-53.920361	…	Ruprecht 79	NaN	0.022	0.00	NaN	0.077	
1272	-57.660417	…	IC 2581	NaN	0.022	0.00	NaN	0.073	
1345	-5.765472	…	Basel 1	NaN	0.022	0.00	NaN	0.064	
1395	-5.913583	…	Basel 1	NaN	0.027	0.01	NaN	0.073	
1405	-5.772222	…	Basel 1	NaN	0.021	0.00	NaN	0.071	
1462	-5.966361	…	Basel 1	NaN	0.042	0.00	NaN	0.065	
1533	-5.886528	…	Basel 1	NaN	0.026	0.01	NaN	0.088	
1539	-5.854250	…	Basel 1	NaN	0.023	0.01	NaN	0.092	
1570	-58.287500	…	NGC 3293	NaN	0.023	0.00	NaN	0.064	
1577	-58.192194	…	NGC 3293	NaN	0.028	0.00	NaN	0.057	
1836	-60.696194	…	Trumpler 20	NaN	0.021	0.00	NaN	0.101	
1967	-60.611472	…	Trumpler 20	NaN	0.022	0.00	NaN	0.075	
1994	-60.631167	…	Trumpler 20	NaN	0.025	0.01	NaN	0.067	
2163	-60.574639	…	Trumpler 20	NaN	0.018	0.01	NaN	0.072	
2415	-47.466806	…	NGC 6200	NaN	0.024	0.00	NaN	0.050	
2437	-47.544444	…	NGC 6200	NaN	0.024	0.00	NaN	0.061	
2463	-47.478972	…	NGC 6200	NaN	0.026	0.00	NaN	0.058	
2483	-46.933944	…	NGC 6204	NaN	NaN	0.01	NaN	0.102	
2661	-44.725972	…	NGC 6249	NaN	0.023	0.00	NaN	0.057	
2674	-35.635611	…	Bochum 13	NaN	0.030	0.00	NaN	0.093	
2678	-33.431944	…	Trumpler 27	NaN	0.024	0.00	NaN	0.092	
2683	-32.467667	…	Trumpler 28	NaN	0.029	0.00	NaN	0.081	
2699	-32.420083	…	Trumpler 28	NaN	NaN	0.00	NaN	0.091	
2713	-32.481056	…	Trumpler 28	NaN	0.026	0.00	NaN	0.060	
2718	-32.443972	…	Trumpler 28	NaN	0.034	0.00	NaN	0.098	
2750	-32.552806	…	Trumpler 28	NaN	0.025	0.00	NaN	0.068	
3015	-8.509889	…	Trumpler 34	NaN	0.023	0.00	NaN	0.047	
3027	-8.526639	…	Trumpler 34	NaN	0.024	0.00	NaN	0.065	
3028	-8.389750	…	Trumpler 34	NaN	0.023	0.00	NaN	0.078	
3102	-5.930500	…	Basel 1	NaN	0.022	0.01	NaN	0.084	

	e_Kmag	_RAJ2000	SB?	e_gmag
165	0.021	114.083458	NaN	0.00
235	0.019	114.173500	NaN	0.00
240	NaN	117.481417	NaN	0.01
256	0.043	117.495292	NaN	0.01
470	NaN	117.722875	NaN	0.01
906	0.023	145.126708	NaN	0.01
1272	0.021	157.007667	NaN	0.01
1345	0.024	282.029125	NaN	0.00
1395	0.021	282.059333	NaN	0.01

```
1405   0.019  282.067375  NaN    0.00
1462   0.021  282.104792  NaN    0.00
1533   0.025  282.153042  NaN    0.01
1539   0.023  282.156125  NaN    0.01
1570   0.019  158.856375  NaN    0.00
1577   0.026  159.009542  NaN    0.00
1836   0.023  189.831292  NaN    0.02
1967   0.020  189.910875  NaN    0.01
1994   0.018  189.925042  NaN    0.01
2163   0.029  190.020042  NaN    0.01
2415   0.025  250.958125  NaN    0.00
2437   0.023  251.046333  NaN    0.00
2463   0.023  251.131750  NaN    0.01
2483     NaN  251.693333  NaN    0.01
2661   0.020  254.576000  NaN    0.00
2674   0.027  259.347000  NaN    0.00
2678   0.023  264.084292  NaN    0.00
2683   0.026  264.121042  NaN    0.00
2699     NaN  264.186458  NaN    0.01
2713   0.023  264.221958  NaN    0.00
2718   0.031  264.235167  NaN    0.01
2750   0.023  264.319042  NaN    0.00
3015   0.020  280.074042  NaN    0.00
3027   0.023  280.084708  NaN    0.01
3028   0.023  280.084833  NaN    0.01
3102   0.021  281.981750  NaN    0.01

[35 rows x 43 columns]
```

4 Analysis of q-features (q_3) (all magnitudes)

[98]:
```
x_np=x.to_numpy()
import qfeatures
coefficients = dataset_module.coefficients
systems = dataset_module.systems
coefficients_np = np.array([coefficients[k] for k in x.columns])
systems = [systems[k] for k in x.columns]
q=qfeatures.calculate(x_np,coefficients_np,x.columns,systems,combination_size=3)
m = q.magnitudes

q_df = pd.DataFrame(m, columns = q.column_names)
q_df.describe()
```

[98]:
```
       umag_gmag_rmag  umag_gmag_imag  umag_gmag_Hamag  umag_gmag_Jmag  \
count     3142.000000     3142.000000      3142.000000     3142.000000
mean         0.156141       -0.378599        -0.010663       -3.028395
```

std	0.477268	0.333868	0.437317	0.940956
min	-0.759654	-1.321813	-0.903692	-7.115069
25%	-0.126526	-0.615351	-0.282897	-3.564993
50%	0.016104	-0.469503	-0.131121	-2.951368
75%	0.231807	-0.232851	0.094217	-2.415524
max	2.579221	1.192398	2.186495	-0.466389

	umag_gmag_Hmag	umag_gmag_Kmag	umag_rmag_imag	umag_rmag_Hamag	\
count	3142.000000	3142.000000	3142.000000	3142.000000	
mean	-5.979377	-9.798617	0.954243	1.407748	
std	2.229224	3.864487	0.827491	0.977587	
min	-16.027196	-27.431046	-0.711111	-0.596636	
25%	-7.162250	-11.795907	0.425526	0.813995	
50%	-5.624293	-9.041667	0.664269	1.106449	
75%	-4.540071	-7.278962	1.343070	1.737780	
max	-0.952413	-1.686405	4.944912	6.135140	

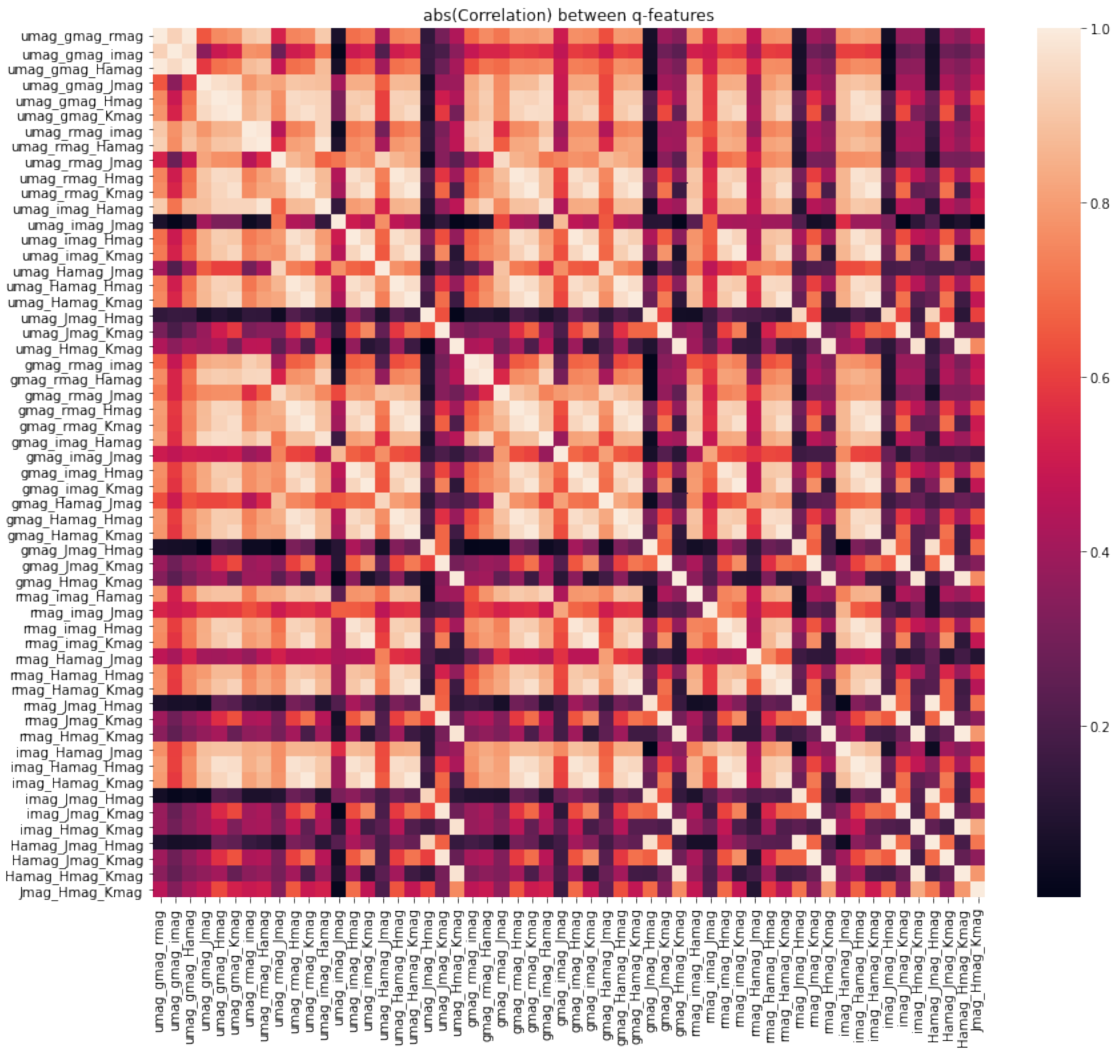
	umag_rmag_Jmag	umag_rmag_Hmag	…	imag_Hamag_Jmag	imag_Hamag_Hmag	\
count	3142.000000	3142.000000	…	3142.000000	3142.000000	
mean	-2.400995	-6.371318	…	0.365199	1.100709	
std	0.870262	2.649514	…	0.184691	0.539457	
min	-7.689333	-21.841478	…	-0.028903	0.046913	
25%	-2.914667	-7.681826	…	0.241431	0.755772	
50%	-2.415333	-5.949609	…	0.319889	0.973728	
75%	-1.835278	-4.581130	…	0.463094	1.322918	
max	0.383778	-0.032348	…	1.375556	4.200283	

	imag_Hamag_Kmag	imag_Jmag_Hmag	imag_Jmag_Kmag	imag_Hmag_Kmag	\
count	3142.000000	3142.000000	3142.000000	3142.000000	
mean	2.010004	0.029678	-0.817382	0.638187	
std	0.976299	0.406704	0.884542	0.772481	
min	0.031895	-6.482065	-13.751588	-11.218902	
25%	1.366315	-0.095897	-1.125529	0.350629	
50%	1.780310	0.037130	-0.677735	0.625464	
75%	2.436038	0.174163	-0.364721	0.986067	
max	7.662516	4.936783	4.142765	7.507065	

	Hamag_Jmag_Hmag	Hamag_Jmag_Kmag	Hamag_Hmag_Kmag	Jmag_Hmag_Kmag
count	3142.000000	3142.000000	3142.000000	3142.000000
mean	0.004697	-1.210278	0.767005	0.254415
std	0.581073	1.279983	1.001007	0.264408
min	-9.400609	-19.827601	-15.024588	-2.622131
25%	-0.192011	-1.668518	0.427647	0.119691
50%	0.027500	-0.991654	0.778657	0.231033
75%	0.236761	-0.543185	1.206074	0.380542
max	6.374304	5.383026	9.521608	2.715843

```
[8 rows x 56 columns]
```

[99]:
```python
sn.heatmap(q_df.corr().abs())
plt.title("abs(Correlation) between q-features")
plt.show()
```



5 Analysis of q-features (q_3) (calculated by system)

[100]:
```python
x_np=x.to_numpy()
import qfeatures
coefficients = dataset_module.coefficients
systems = dataset_module.systems
coefficients_np = np.array([coefficients[k] for k in x.columns])
systems = [systems[k] for k in x.columns]
```

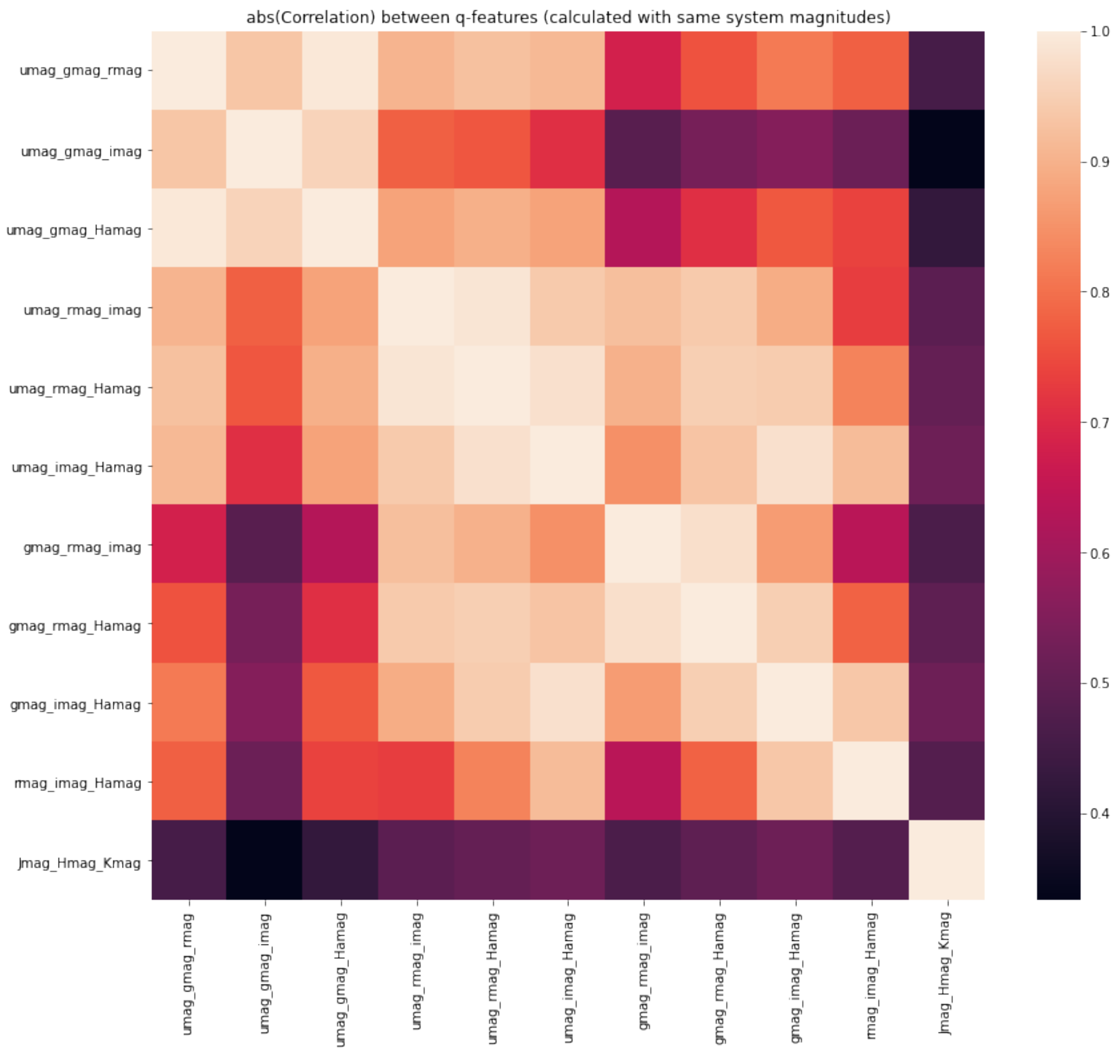
```
q= qfeatures.calculate(x_np,coefficients_np,x.
 ↪columns,systems,combination_size=3,by_system=True)

m = q.magnitudes

q_df = pd.DataFrame(m, columns = q.column_names)
q_df.describe()
```

[100]:

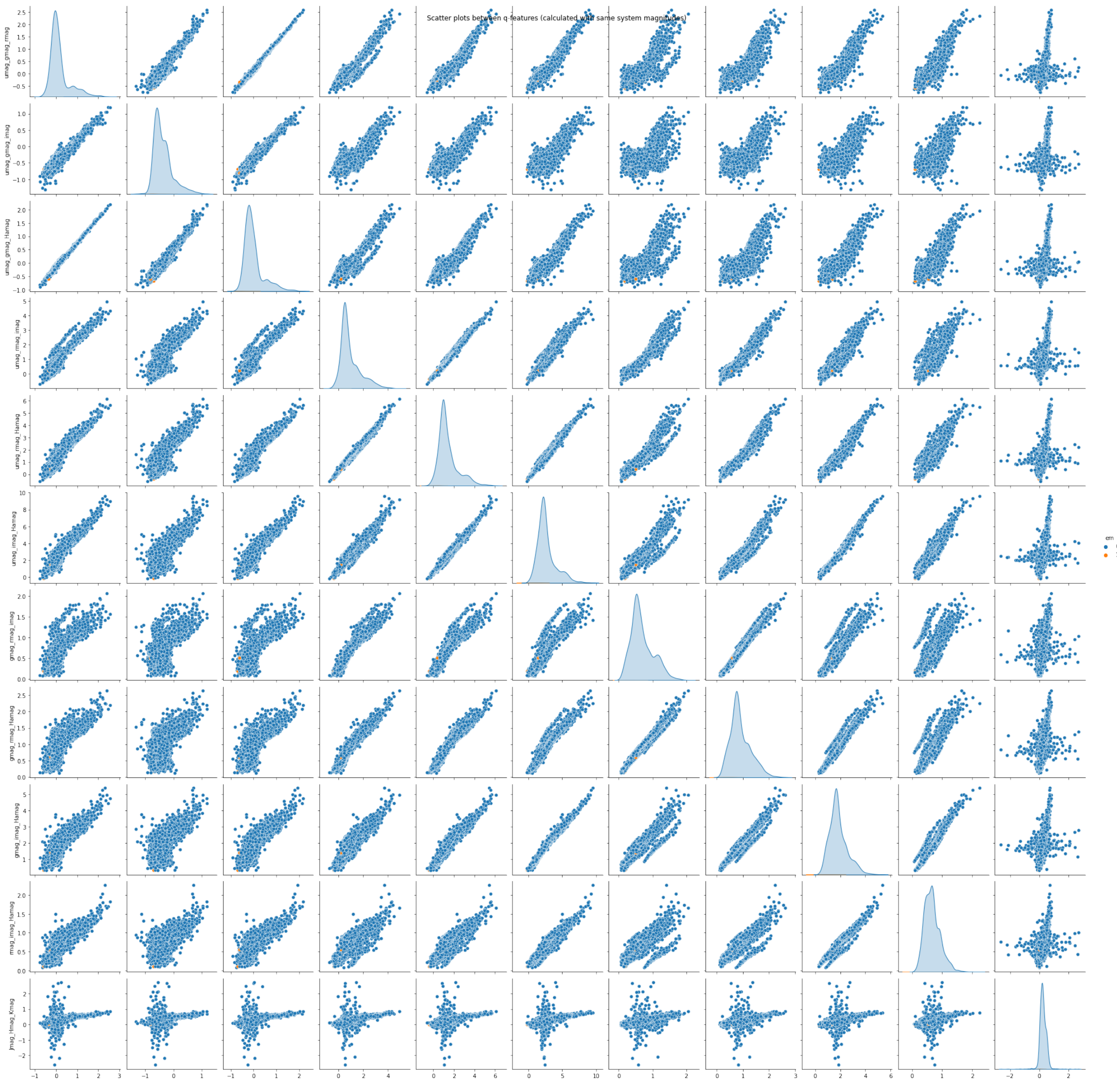
```
       umag_gmag_rmag  umag_gmag_imag  umag_gmag_Hamag  umag_rmag_imag  \
count     3142.000000     3142.000000      3142.000000     3142.000000
mean         0.156141       -0.378599        -0.010663        0.954243
std          0.477268        0.333868         0.437317        0.827491
min         -0.759654       -1.321813        -0.903692       -0.711111
25%         -0.126526       -0.615351        -0.282897        0.425526
50%          0.016104       -0.469503        -0.131121        0.664269
75%          0.231807       -0.232851         0.094217        1.343070
max          2.579221        1.192398         2.186495        4.944912

       umag_rmag_Hamag  umag_imag_Hamag  gmag_rmag_imag  gmag_rmag_Hamag  \
count      3142.000000      3142.000000     3142.000000      3142.000000
mean          1.407748         2.631337        0.684462         0.900313
std           0.977587         1.403822        0.334951         0.397890
min          -0.596636        -0.202336        0.068421         0.117617
25%           0.813995         1.770654        0.461711         0.643785
50%           1.106449         2.298271        0.602632         0.812547
75%           1.737780         3.101028        0.885789         1.145245
max           6.135140         9.560561        2.063684         2.630187

       gmag_imag_Hamag  rmag_imag_Hamag  Jmag_Hmag_Kmag
count      3142.000000      3142.000000     3142.000000
mean          1.831181         0.665001        0.254415
std           0.723852         0.265588        0.264408
min           0.310000         0.073178       -2.622131
25%           1.364100         0.474112        0.119691
50%           1.712173         0.636916        0.231033
75%           2.188703         0.817757        0.380542
max           5.373318         2.257290        2.715843
```

[101]:

```
sn.heatmap(q_df.corr().abs())
plt.title("abs(Correlation) between q-features (calculated with same system␣
 ↪magnitudes)")
plt.show()
q_dfy=pd.concat([q_df,y],axis=1)
sn.pairplot(q_dfy,hue="em")
plt.suptitle("Scatter plots between q-features (calculated with same system␣
 ↪magnitudes)")
```




```
[101]: Text(0.5, 0.98, 'Scatter plots between q-features (calculated with same system
magnitudes)')
```


Scatter plots between q-features (calculated with same system magnitudes)
em
0
1
umag_gmag_rmag
umag_gmag_imag
umag_gmag_Hamag
umag_rmag_imag
umag_rmag_Hamag
umag_imag_Hamag
gmag_rmag_imag
gmag_rmag_Hamag
gmag_imag_Hamag
rmag_imag_Hamag
Jmag_Hmag_Kmag