main

April 25, 2023

[ ]:
```
pip install pandas numpy  seaborn missingno matplotlib scikit-learn joblib
```

[1]:
```
import pandas as pd
import json
import os
import missingno as ms
import seaborn as sns
import matplotlib.pyplot as plt
import sklearn as skl
```

[2]:
```
dataCsvPath = os.path.join("..","results.csv")
data_df = pd.read_csv(dataCsvPath)
src_paths = list(data_df["src_path"])

print("Number of rows:", data_df.shape[0])
print("Number of columns:", data_df.shape[1])
print("Column names:", list(data_df.columns))
```

```
Number of rows: 4427
Number of columns: 22
Column names: ['model_name', 'src_path', 'conv_path', 'src_ext', 'is_parsed',
'is_sys_design', 'sys_name', 'no_components', 'no_connectors',
'no_hardware_comp', 'understandability', 'no_size', 'no_data_comp',
'no_software_comp', 'no_sys_comp', 'coupling', 'cohesion', 'complexity',
'graph_density', 'avg_shortest_path', 'avg_deg_cent', 'doc_files']
```

[3]:
```
data_df.head()
```

[3]:

	model_name	src_path
0	NaN	
1	isolette_heat_source	
2	isolette_operator_interface	
3	isolette_integration_1	
4	isolette	
0		/mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
1		/mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…

```
2  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
3  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
4  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…

                                          conv_path src_ext  is_parsed  \
0  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…    aadl      False
1  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…    aadl       True
2                                              NaN    aadl      False
3                                              NaN    aadl      False
4  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…    aadl       True

   is_sys_design                          sys_name  no_components  \
0          False                               NaN            NaN
1           True  Heat_Source_with_devices_Instance            8.0
2          False                               NaN            NaN
3          False                               NaN            NaN
4           True   Temperature_Sensor_impl_Instance            2.0

   no_connectors  no_hardware_comp  …  no_data_comp  no_software_comp  \
0            NaN               NaN  …           NaN               NaN
1            7.0               4.0  …           0.0               3.0
2            NaN               NaN  …           NaN               NaN
3            NaN               NaN  …           NaN               NaN
4            0.0               0.0  …           0.0               0.0

   no_sys_comp  coupling  cohesion  complexity  graph_density  \
0          NaN       NaN       NaN         NaN            NaN
1          1.0  3.166667  0.333333        11.0            1.0
2          NaN       NaN       NaN         NaN            NaN
3          NaN       NaN       NaN         NaN            NaN
4          2.0  0.000000  0.000000         0.0            0.0

   avg_shortest_path  avg_deg_cent  \
0                NaN           NaN
1           0.517857         0.125
2                NaN           NaN
3                NaN           NaN
4           0.000000         0.000

                                          doc_files
0  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
1  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
2  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
3  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
4  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…

[5 rows x 22 columns]
```

0.1 Cleaning the data, filtering the non parsed models and removing some columns

[4]:
```
data_df =data_df[(data_df["is_parsed"] == True) & (data_df["is_sys_design"] ==␣
 ↪True)]
df = data_df.loc[:, ~data_df.columns.isin(['src_path',␣
 ↪'conv_path',"src_ext","doc_files"])]
print("Number of rows:", data_df.shape[0])
print("Number of columns:", data_df.shape[1])
df.head()
```

```
Number of rows: 3368
Number of columns: 22
```

[4]:

	model_name	is_parsed	is_sys_design	\
1	isolette_heat_source	True	True	
4	isolette	True	True	
8	isolette_heat_source	True	True	
9	isolette_operator_interface	True	True	
13	isolette_temperature_sensor	True	True	

	sys_name	no_components	no_connectors	\
1	Heat_Source_with_devices_Instance	8.0	7.0	
4	Temperature_Sensor_impl_Instance	2.0	0.0	
8	Heat_Source_impl_Instance	6.0	0.0	
9	Operator_Interface_impl_Instance	2.0	0.0	
13	Temperature_Sensor_impl_Instance	2.0	0.0	

	no_hardware_comp	understandability	no_size	no_data_comp	\
1	4.0	0.125	15.0	0.0	
4	0.0	0.000	2.0	0.0	
8	4.0	0.000	6.0	0.0	
9	0.0	0.000	2.0	0.0	
13	0.0	0.000	2.0	0.0	

	no_software_comp	no_sys_comp	coupling	cohesion	complexity	\
1	3.0	1.0	3.166667	0.333333	11.0	
4	0.0	2.0	0.000000	0.000000	0.0	
8	0.0	2.0	1.000000	0.000000	2.0	
9	0.0	2.0	0.000000	0.000000	0.0	
13	0.0	2.0	0.000000	0.000000	0.0	

	graph_density	avg_shortest_path	avg_deg_cent
1	1.0	0.517857	0.125
4	0.0	0.000000	0.000
8	0.0	0.000000	0.000
9	0.0	0.000000	0.000
13	0.0	0.000000	0.000

0.2 Extracting the String Data and Creating a Data Set for that

[5]:
```
text_models_data = data_df.loc[:,data_df.columns.
 ↪isin(["model_name","graph_str_rep","doc_files"])]
text_models_data = text_models_data.drop_duplicates()
text_models_data.to_csv("data_text.csv",index=True)
text_models_data
```

[5]:
```
                          model_name  \
1               isolette_heat_source
4                           isolette
9         isolette_operator_interface
13        isolette_temperature_sensor
17    JustSubprogramGroup_process_fg
…                                 …
4420            heterogenous_systems
4421                     libmathtest
4422                           model
4423                         systems
4424                ardupilot_system

                                                 doc_files
1     /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
4     /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
9     /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
13    /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
17    /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
…                                                   …
4420  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
4421  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
4422  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
4423  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…
4424  /mnt/DATA/00-GSSI/00-WORK/EXAMPLE_ROOT_DIRECTO…

[1238 rows x 2 columns]
```

0.3 Exploratory analysis

[6]:
```
df_num = df.loc[:, ~df.columns.isin(["is_parsed","is_sys_design"])]
df_num.head()
```

[6]:
```
                    model_name                          sys_name  \
1          isolette_heat_source  Heat_Source_with_devices_Instance
4                      isolette   Temperature_Sensor_impl_Instance
8          isolette_heat_source          Heat_Source_impl_Instance
9   isolette_operator_interface   Operator_Interface_impl_Instance
13  isolette_temperature_sensor   Temperature_Sensor_impl_Instance
```

```
    no_components  no_connectors  no_hardware_comp  understandability  \
1             8.0            7.0               4.0              0.125
4             2.0            0.0               0.0              0.000
8             6.0            0.0               4.0              0.000
9             2.0            0.0               0.0              0.000
13            2.0            0.0               0.0              0.000

    no_size  no_data_comp  no_software_comp  no_sys_comp  coupling  cohesion  \
1      15.0           0.0               3.0          1.0  3.166667  0.333333
4       2.0           0.0               0.0          2.0  0.000000  0.000000
8       6.0           0.0               0.0          2.0  1.000000  0.000000
9       2.0           0.0               0.0          2.0  0.000000  0.000000
13      2.0           0.0               0.0          2.0  0.000000  0.000000

    complexity  graph_density  avg_shortest_path  avg_deg_cent
1         11.0            1.0           0.517857         0.125
4          0.0            0.0           0.000000         0.000
8          2.0            0.0           0.000000         0.000
9          0.0            0.0           0.000000         0.000
13         0.0            0.0           0.000000         0.000
```

[9]:
```
df_num.isnull().sum()
```

[9]:
```
model_name           0
sys_name             0
no_components        0
no_connectors        0
no_hardware_comp     0
understandability    0
no_size              0
no_data_comp         0
no_software_comp     0
no_sys_comp          0
coupling             0
cohesion             0
complexity           0
graph_density        0
avg_shortest_path    0
avg_deg_cent         0
dtype: int64
```

0.3.1 Filtering by num of componnents >= 3

[10]:
```
df_num = df_num[(df_num["no_components"] >= 3)]
df_num
```

[10]:

	model_name	sys_name	\
1	isolette_heat_source	Heat_Source_with_devices_Instance	
8	isolette_heat_source	Heat_Source_impl_Instance	
17	JustSubprogramGroup_process_fg	p1_impl_Instance	
18	JustSubprogramGroup_process_fg	Root_impl_Instance	
21	sc3	src3_i_Instance	
…	…	…	
4422	model	main_i_Instance	
4423	systems	mysystem_impl_Instance	
4424	ardupilot_system	ardupilot_i_Instance	
4425	main	main_i_Instance	
4426	case_study_osal	osal_i_Instance	

	no_components	no_connectors	no_hardware_comp	understandability	\
1	8.0	7.0	4.0	0.125000	
8	6.0	0.0	4.0	0.000000	
17	4.0	2.0	0.0	0.166667	
18	5.0	2.0	0.0	0.100000	
21	3.0	1.0	0.0	0.166667	
…	…	…	…	…	
4422	36.0	0.0	15.0	0.000000	
4423	29.0	0.0	15.0	0.000000	
4424	15.0	0.0	6.0	0.000000	
4425	37.0	0.0	17.0	0.000000	
4426	9.0	0.0	4.0	0.000000	

	no_size	no_data_comp	no_software_comp	no_sys_comp	coupling	\
1	15.0	0.0	3.0	1.0	3.166667	
8	6.0	0.0	0.0	2.0	1.000000	
17	6.0	0.0	3.0	1.0	1.000000	
18	7.0	0.0	3.0	2.0	1.000000	
21	4.0	0.0	0.0	3.0	0.000000	
…	…	…	…	…	…	
4422	36.0	0.0	20.0	1.0	1.142857	
4423	29.0	0.0	13.0	1.0	1.000000	
4424	15.0	0.0	8.0	1.0	0.000000	
4425	37.0	0.0	19.0	1.0	3.933333	
4426	9.0	0.0	4.0	1.0	0.000000	

	cohesion	complexity	graph_density	avg_shortest_path	avg_deg_cent
1	0.333333	11.0	1.000000	0.517857	0.125000
8	0.000000	2.0	0.000000	0.000000	0.000000
17	0.666667	0.0	0.666667	0.666667	0.333333
18	0.333333	0.0	0.500000	0.400000	0.200000
21	1.000000	0.0	0.500000	0.666667	0.333333
…	…	…	…	…	…
4422	0.000000	24.0	0.000000	0.000000	0.000000

```
4423  0.000000         8.0       0.000000           0.000000      0.000000
4424  0.000000         0.0       0.000000           0.000000      0.000000
4425  0.000000        31.0       0.000000           0.000000      0.000000
4426  0.000000         0.0       0.000000           0.000000      0.000000

[2476 rows x 16 columns]
```

0.3.2 Droping duplicates with build-in

[11]:
```
df_num.drop_duplicates(inplace=True)
df_num
```

[11]:
```
                          model_name  \
1               isolette_heat_source
8               isolette_heat_source
17    JustSubprogramGroup_process_fg
18    JustSubprogramGroup_process_fg
21                               sc3
…                                …
4338                             APS
4339                             APS
4341                             APS
4347                 Mobile5GNetwork
4360                            syst

                                    sys_name  no_components  no_connectors  \
1            Heat_Source_with_devices_Instance            8.0            7.0
8                    Heat_Source_impl_Instance            6.0            0.0
17                           p1_impl_Instance            4.0            2.0
18                          Root_impl_Instance            5.0            2.0
21                             src3_i_Instance            3.0            1.0
…                                          …              …              …
4338              COLAVandSSITAW_impl_Instance           12.0            6.0
4339      COLAVandSSITAW_Backup_impl_Instance           12.0            8.0
4341  EmergencyButtonTransmitter_impl_Instance            3.0            2.0
4347            APSCommunication_impl_Instance            3.0            8.0
4360      SecuritySystem_with_devices_Instance           13.0           12.0

      no_hardware_comp  understandability  no_size  no_data_comp  \
1                  4.0           0.125000     15.0           0.0
8                  4.0           0.000000      6.0           0.0
17                 0.0           0.166667      6.0           0.0
18                 0.0           0.100000      7.0           0.0
21                 0.0           0.166667      4.0           0.0
…                  …                  …        …             …
4338               0.0           0.045455     18.0           0.0
4339               0.0           0.060606     20.0           0.0
```

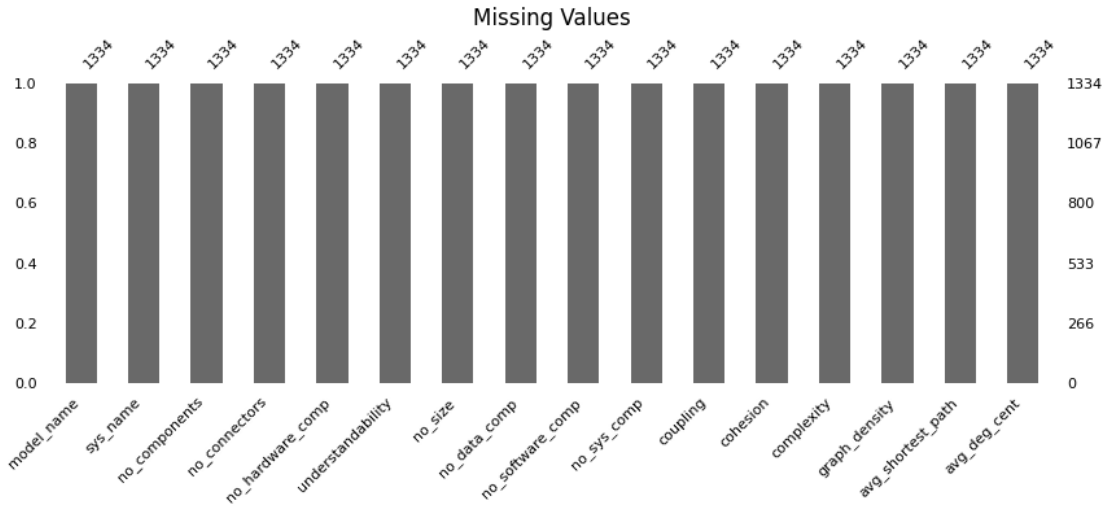
```
4341                1.0           0.333333      5.0           0.0
4347                0.0           1.333333     11.0           0.0
4360                7.0           0.076923     25.0           0.0

      no_software_comp  no_sys_comp  coupling  cohesion  complexity  \
1                  3.0          1.0  3.166667  0.333333        11.0
8                  0.0          2.0  1.000000  0.000000         2.0
17                 3.0          1.0  1.000000  0.666667         0.0
18                 3.0          2.0  1.000000  0.333333         0.0
21                 0.0          3.0  0.000000  1.000000         0.0
…                  …            …        …         …           …
4338              11.0          1.0  5.000000  0.109091        30.0
4339              11.0          1.0  5.000000  0.145455        30.0
4341               0.0          2.0  0.500000  2.000000         1.0
4347               0.0          3.0  1.000000  8.000000         8.0
4360               5.0          1.0  6.766667  0.181818        16.0

      graph_density  avg_shortest_path  avg_deg_cent
1          1.000000           0.517857      0.125000
8          0.000000           0.000000      0.000000
17         0.666667           0.666667      0.333333
18         0.500000           0.400000      0.200000
21         0.500000           0.666667      0.333333
…             …                  …             …
4338       0.545455           0.037879      0.022727
4339       0.727273           0.030303      0.030303
4341       1.000000           0.666667      0.333333
4347       4.000000           0.333333      0.333333
4360       1.000000           0.692308      0.102564

[1334 rows x 16 columns]
```

0.4 Visualization, metrics and statistics over the data

[12]:
```
df_data = df_num
```

[13]:
```
# Missing values
plt.title("Missing Values", fontsize=12)
ms.bar(df_data,fontsize=8,figsize=(10,3))
```

[13]:
```
<AxesSubplot: title={'center': 'Missing Values'}>
```

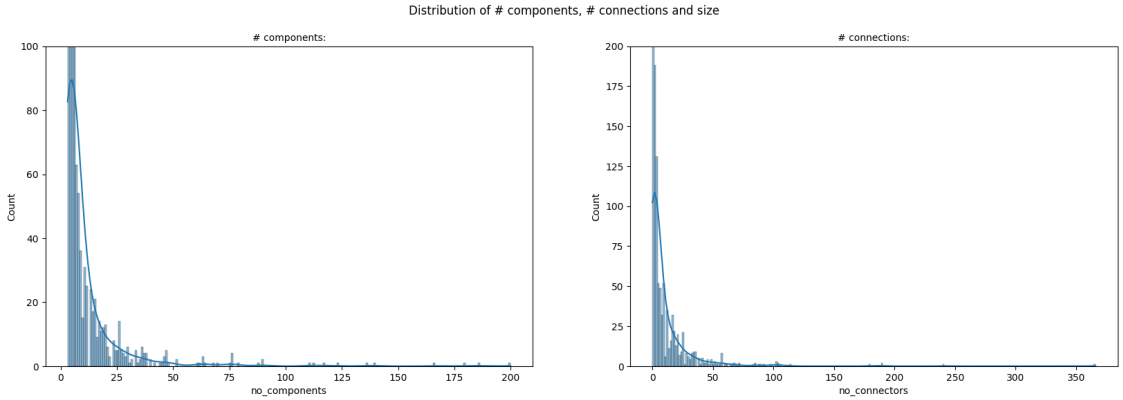


0.5 Distribution for numerical metrics

0.5.1 Distributions for amount of component, connection and size of every model

[14]:
```
fig, axes = plt.subplots(1, 2, figsize=(20,6))
fig.suptitle('Distribution of # components, # connections and size',fontsize=12)
axes[0].set_title('# components:',fontsize=10)
axes[1].set_title('# connections:',fontsize=10)
# axes[2].set_title('# size:',fontsize=10)
sns.histplot(ax=axes[0],data=df_data['no_components'],kde=True);
sns.histplot(ax=axes[1],data=df_data['no_connectors'],kde=True);
axes[0].set_ylim([0, 100])
axes[1].set_ylim([0, 200])
# sns.histplot(ax=axes[2],data=df_num['size'],kde=True);
```

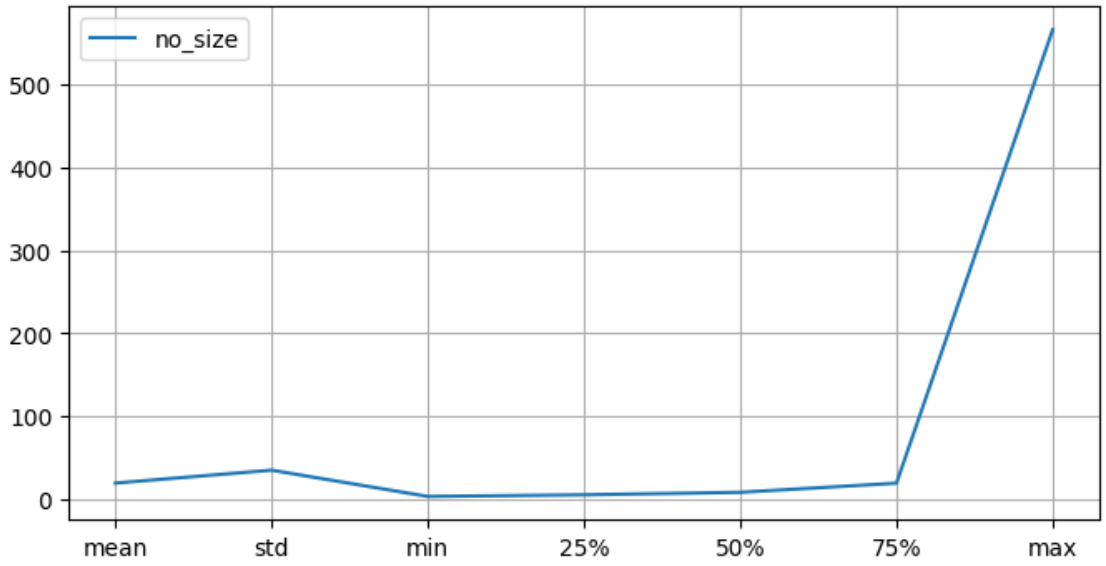
[14]: (0.0, 200.0)



0.5.2 Description for the size parametter

[15]:
```
plt.figure(figsize=(8,4))
plt.grid()
sns.lineplot(data=df_data[["no_size"]].describe().drop("count",axis=0))
```

[15]: <AxesSubplot: >



[16]:
```
df_data[["no_size"]].describe()
```

[16]:

	no_size
count	1334.000000
mean	19.159670
std	34.716892
min	3.000000
25%	5.000000
50%	8.000000
75%	19.000000
max	566.000000

[17]:
```
small_size = df_data[(df_num["no_size"] >= 1) & (df_data["no_size"] <= 13) ]
(x,_) = small_size.shape
print("percent: ", ((x) / df_data.shape[0])*100)
small_size
```

```
percent:  67.31634182908546
```

[17]:

	model_name	sys_name	no_components	no_connectors
8	isolette_heat_source	Heat_Source_impl_Instance	6.0	0.0
17	JustSubprogramGroup_process_fg	p1_impl_Instance	4.0	2.0
18	JustSubprogramGroup_process_fg	Root_impl_Instance	5.0	2.0
21	sc3	src3_i_Instance	3.0	1.0
24	sc3	whole_i_Instance	6.0	1.0
…	…	…	…	…
4334	APS	GNSS_IMU_Main_impl_Instance	3.0	0.0
4335	APS	GNSS_IMU_Backup_impl_Instance	3.0	0.0
4337	APS	IS3MS_impl_Instance	3.0	6.0
4341	APS	EmergencyButtonTransmitter_impl_Instance	3.0	2.0
4347	Mobile5GNetwork	APSCommunication_impl_Instance	3.0	8.0

	no_hardware_comp	understandability	no_size	no_data_comp
8	4.0	0.000000	6.0	0.0
17	0.0	0.166667	6.0	0.0
18	0.0	0.100000	7.0	0.0
21	0.0	0.166667	4.0	0.0
24	0.0	0.033333	7.0	0.0
…	…	…	…	…
4334	0.0	0.000000	3.0	0.0
4335	0.0	0.000000	3.0	0.0
4337	0.0	1.000000	9.0	0.0
4341	1.0	0.333333	5.0	0.0
4347	0.0	1.333333	11.0	0.0

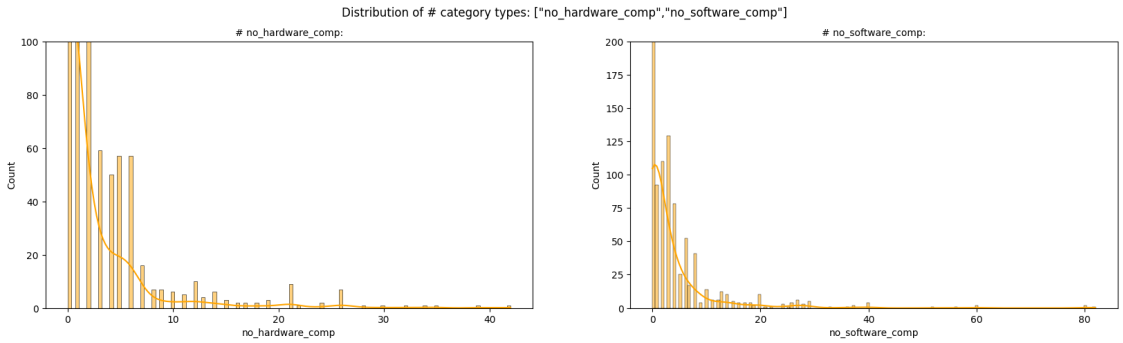
	no_software_comp	no_sys_comp	coupling	cohesion	complexity
8	0.0	2.0	1.0	0.000000	2.0
17	3.0	1.0	1.0	0.666667	0.0
18	3.0	2.0	1.0	0.333333	0.0
21	0.0	3.0	0.0	1.000000	0.0
24	0.0	6.0	2.0	0.100000	0.0
…	…	…	…	…	…
4334	2.0	1.0	1.0	0.000000	2.0

```
4335                2.0          1.0       1.0  0.000000         2.0
4337                2.0          1.0       1.0  6.000000        10.0
4341                0.0          2.0       0.5  2.000000         1.0
4347                0.0          3.0       1.0  8.000000         8.0

      graph_density  avg_shortest_path  avg_deg_cent
8          0.000000           0.000000      0.000000
17         0.666667           0.666667      0.333333
18         0.500000           0.400000      0.200000
21         0.500000           0.666667      0.333333
24         0.200000           0.666667      0.133333
...             ...                ...           ...
4334       0.000000           0.000000      0.000000
4335       0.000000           0.000000      0.000000
4337       3.000000           0.833333      0.500000
4341       1.000000           0.666667      0.333333
4347       4.000000           0.333333      0.333333

[898 rows x 16 columns]
```

0.5.3 Distributions for amount of category of the components

[18]:
```
fig, axes = plt.subplots(1, 2, figsize=(20,5))
fig.suptitle('Distribution of # category types:␣
 ↪["no_hardware_comp","no_software_comp"]',fontsize=12)
axes[0].set_title('# no_hardware_comp:',fontsize=10)
axes[1].set_title('# no_software_comp:',fontsize=10)
sns.
 ↪histplot(ax=axes[0],data=df_data['no_hardware_comp'],kde=True,color='orange');
 ↪
sns.
 ↪histplot(ax=axes[1],data=df_data['no_software_comp'],kde=True,color='orange');
 ↪
axes[0].set_ylim([0, 100])
axes[1].set_ylim([0, 200])
```

[18]:
```
(0.0, 200.0)
```



0.5.4 Description for no_hardware_comp and no_software_comp

[19]:
```
df_data[["no_hardware_comp"]].describe()
```

[19]:
```
       no_hardware_comp
count       1334.000000
mean           2.182909
std            4.550572
min            0.000000
25%            0.000000
50%            0.000000
75%            2.000000
max           42.000000
```

[20]:
```
df_data[["no_software_comp"]].describe()
```

[20]:
```
       no_software_comp
count       1334.000000
mean           3.320840
std            7.339617
min            0.000000
25%            0.000000
50%            0.000000
75%            3.000000
max           82.000000
```

[21]:
```
df_data[df_data["no_software_comp"]>= 5]
```

[21]:
```
                model_name                                sys_name  \
82     SimpleControlSystem                    SCS_tier2_Instance
83     SimpleControlSystem                  SCS_dualtier2_Instance
251              issue2056                   Example_impl_Instance
281      integration::main                      main_impl_Instance
294   DigitalControlSystem                DCS_singletier2_Instance
…                      …                                       …
4320                 APS                ASC_Backup_impl_Instance
4336                 APS  NetworkAndSystemManagement_impl_Instance
4338                 APS              COLAVandSSITAW_impl_Instance
4339                 APS       COLAVandSSITAW_Backup_impl_Instance
4360                syst      SecuritySystem_with_devices_Instance

      no_components  no_connectors  no_hardware_comp  understandability  \
82             16.0           10.0               6.0           0.041667
83             17.0           11.0               7.0           0.040441
251             8.0            2.0               1.0           0.035714
```

```
281            47.0           53.0               4.0           0.024514
294            12.0           11.0               3.0           0.083333
…               …              …                  …              …
4320           30.0           26.0               0.0           0.029885
4336           14.0           12.0               0.0           0.065934
4338           12.0            6.0               0.0           0.045455
4339           12.0            8.0               0.0           0.060606
4360           13.0           12.0               7.0           0.076923

      no_size  no_data_comp  no_software_comp  no_sys_comp   coupling  \
82       26.0           0.0               6.0          4.0  10.100000
83       28.0           0.0               6.0          4.0  11.100000
251      10.0           0.0               6.0          1.0   2.000000
281     100.0           0.0               8.0         35.0  14.788095
294      23.0           0.0               6.0          3.0   7.166667
…         …             …                 …            …        …
4320     56.0           0.0              26.0          4.0   9.500000
4336     26.0           0.0              13.0          1.0   2.000000
4338     18.0           0.0              11.0          1.0   5.000000
4339     20.0           0.0              11.0          1.0   5.000000
4360     25.0           0.0               5.0          1.0   6.766667

      cohesion  complexity  graph_density  avg_shortest_path  avg_deg_cent
82    0.095238        22.0       0.666667           1.545833      0.087500
83    0.091667        22.0       0.687500           1.363971      0.077206
251   0.095238         4.0       0.285714           0.357143      0.107143
281   0.051208       235.0       1.152174           0.388067      0.024977
294   0.200000        12.0       1.000000           0.939394      0.136364
…          …           …              …                  …             …
4320  0.064039        69.0       0.896552           0.079310      0.024138
4336  0.153846        12.0       0.923077           0.137363      0.038462
4338  0.109091        30.0       0.545455           0.037879      0.022727
4339  0.145455        30.0       0.727273           0.030303      0.030303
4360  0.181818        16.0       1.000000           0.692308      0.102564

[255 rows x 16 columns]
```

[22]:
```
df_data[df_data["no_hardware_comp"] >= 5]
```

[22]:
```
                                 model_name  \
81                      SimpleControlSystem
82                      SimpleControlSystem
83                      SimpleControlSystem
106   AircraftSafetyExample::AOADiscrepancy
107   AircraftSafetyExample::AOADiscrepancy
…                                       …
4329                                   APS
```

```
4330                                    APS
4331                                    APS
4333                                    APS
4360                                   syst

                                                sys_name  no_components  \
81                                   SCS_tier1_Instance           12.0
82                                   SCS_tier2_Instance           16.0
83                                SCS_dualtier2_Instance           17.0
106                            ac_OneSensorSpec_Instance            9.0
107               ac_OSSPermanentDiscrepancyFail_Instance            9.0
…                                                    …            …
4329                       NavigationSystem_impl_Instance           26.0
4330        NavigationSystem_impl_withRedundancy_Instance           27.0
4331       NavigationSystem_impl_withRedundancy2_Instance           33.0
4333  NavigationSystem_Backup_impl_withRedundancy_In…           13.0
4360                 SecuritySystem_with_devices_Instance           13.0

      no_connectors  no_hardware_comp  understandability  no_size  \
81              9.0               6.0           0.068182     21.0
82             10.0               6.0           0.041667     26.0
83             11.0               7.0           0.040441     28.0
106            10.0               6.0           0.138889     19.0
107            10.0               6.0           0.138889     19.0
…               …                  …                  …        …
4329           52.0              21.0           0.080000     78.0
4330           54.0              21.0           0.076923     81.0
4331           58.0              21.0           0.054924     91.0
4333           16.0               8.0           0.102564     29.0
4360           12.0               7.0           0.076923     25.0

      no_data_comp  no_software_comp  no_sys_comp   coupling  cohesion  \
81             0.0               2.0          4.0   8.100000  0.163636
82             0.0               6.0          4.0  10.100000  0.095238
83             0.0               6.0          4.0  11.100000  0.091667
106            0.0               0.0          1.0   2.433333  0.357143
107            0.0               0.0          1.0   2.433333  0.357143
…               …                 …            …         …         …
4329           0.0               0.0          5.0  12.666667  0.173333
4330           0.0               0.0          6.0  13.166667  0.166154
4331           0.0               4.0          8.0  16.333333  0.116935
4333           0.0               0.0          5.0   6.166667  0.242424
4360           0.0               5.0          1.0   6.766667  0.181818

      complexity  graph_density  avg_shortest_path  avg_deg_cent
81          18.0       0.818182           0.901515      0.106061
82          22.0       0.666667           1.545833      0.087500
```

```
83          22.0       0.687500           1.363971      0.077206
106         14.0       1.250000           0.555556      0.138889
107         14.0       1.250000           0.555556      0.138889
…            …           …                  …             …
4329       318.0       2.080000           2.283077      0.070769
4330       382.0       2.076923           1.072650      0.066952
4331       395.0       1.812500           0.953598      0.044508
4333       170.0       1.333333           0.282051      0.076923
4360        16.0       1.000000           0.692308      0.102564

[213 rows x 16 columns]
```

[23]:
```
df_data[["no_data_comp"]].describe()
```

[23]:
```
       no_data_comp
count   1334.000000
mean       1.184408
std        9.354430
min        0.000000
25%        0.000000
50%        0.000000
75%        0.000000
max      170.000000
```

[24]:
```
df_data[df_data["no_data_comp"] >= 5].head(10)
```

[24]:
```
              model_name                                        sys_name  \
30     data_port_to_data                              top_impl_Instance
231  CasePositionControl                           SMS_buffered_Instance
234  CasePositionControl  SMS_Operational_Environment_buffered_Instance
463       access_to_data                              top_impl_Instance
611     DeclarativeTests         Sub_to_provides_comp_outgoing_Instance
612     DeclarativeTests         Sub_to_provides_comp_incoming_Instance
613     DeclarativeTests            Sub_to_provides_comp_bidir_Instance
614     DeclarativeTests           Sub_to_provides_comp_bidir2_Instance
615     DeclarativeTests      Sub_FG_to_provides_comp_outgoing_Instance
616     DeclarativeTests      Sub_FG_to_provides_comp_incoming_Instance

     no_components  no_connectors  no_hardware_comp  understandability  \
30            11.0          16.0               0.0           0.145455
231           12.0          13.0               2.0           0.098485
234           19.0          21.0               6.0           0.061404
463           11.0          16.0               0.0           0.145455
611            6.0           9.0               0.0           0.300000
612            6.0           9.0               0.0           0.300000
613            6.0          18.0               0.0           0.600000
614            6.0          18.0               0.0           0.600000
```

```
615            6.0            9.0               0.0           0.300000
616            6.0            9.0               0.0           0.300000

     no_size  no_data_comp  no_software_comp  no_sys_comp  coupling  cohesion  \
30      27.0           8.0               2.0          1.0  0.500000  0.355556
231     25.0           6.0               3.0          1.0  2.238095  0.236364
234     40.0           6.0               3.0          4.0  5.654762  0.137255
463     27.0           8.0               2.0          1.0  1.000000  0.355556
611     15.0           5.0               0.0          1.0  0.000000  0.900000
612     15.0           5.0               0.0          1.0  0.000000  0.900000
613     24.0           5.0               0.0          1.0  0.000000  1.800000
614     24.0           5.0               0.0          1.0  0.000000  1.800000
615     15.0           5.0               0.0          1.0  0.000000  0.900000
616     15.0           5.0               0.0          1.0  0.000000  0.900000

     complexity  graph_density  avg_shortest_path  avg_deg_cent
30          1.0       1.600000           1.163636      0.145455
231        22.0       1.181818           0.143939      0.053030
234        29.0       1.166667           0.836257      0.064327
463         0.0       1.600000           1.163636      0.145455
611         0.0       1.800000           0.166667      0.166667
612         0.0       1.800000           0.000000      0.000000
613         0.0       3.600000           0.166667      0.166667
614         0.0       3.600000           0.166667      0.166667
615         0.0       1.800000           0.166667      0.166667
616         0.0       1.800000           0.000000      0.000000
```

[25]:
```
df_data[df_data["no_sys_comp"] >= 5]
```

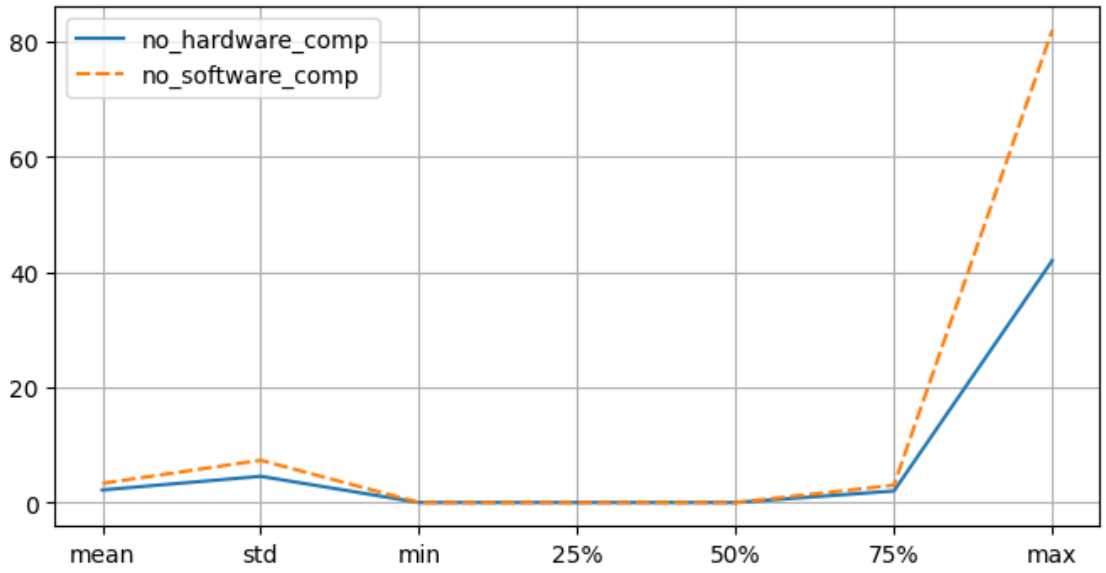
[25]:
```
                    model_name  \
24                         sc3
218                    DualFGS
222   findSubcomponentInstance
281          integration::main
283               OptimizeTree
…                            …
4318                       APS
4329                       APS
4330                       APS
4331                       APS
4333                       APS

                                      sys_name  no_components  \
24                               whole_i_Instance            6.0
218                             FGS_impl_Instance            7.0
222                           toplevel_i_Instance            7.0
281                            main_impl_Instance           47.0
```

```
283                                  Top_impl_Instance            7.0
…                                                  …            …
4318                                 ASC_impl_Instance           61.0
4329                    NavigationSystem_impl_Instance           26.0
4330     NavigationSystem_impl_withRedundancy_Instance           27.0
4331    NavigationSystem_impl_withRedundancy2_Instance           33.0
4333  NavigationSystem_Backup_impl_withRedundancy_In…           13.0

      no_connectors  no_hardware_comp  understandability  no_size  \
24              1.0               0.0           0.033333      7.0
218             7.0               1.0           0.166667     14.0
222             0.0               0.0           0.000000      7.0
281            53.0               4.0           0.024514    100.0
283             0.0               0.0           0.000000      7.0
…             …                 …                  …      …
4318           50.0               0.0           0.013661    111.0
4329           52.0              21.0           0.080000     78.0
4330           54.0              21.0           0.076923     81.0
4331           58.0              21.0           0.054924     91.0
4333           16.0               8.0           0.102564     29.0

      no_data_comp  no_software_comp  no_sys_comp   coupling  cohesion  \
24             0.0               0.0          6.0   2.000000  0.100000
218            0.0               0.0          6.0   2.666667  0.466667
222            0.0               0.0          7.0   0.000000  0.000000
281            0.0               8.0         35.0  14.788095  0.051208
283            0.0               0.0          7.0   0.000000  0.000000
…            …                 …            …        …       …
4318           0.0              52.0          9.0  20.000000  0.028249
4329           0.0               0.0          5.0  12.666667  0.173333
4330           0.0               0.0          6.0  13.166667  0.166154
4331           0.0               4.0          8.0  16.333333  0.116935
4333           0.0               0.0          5.0   6.166667  0.242424

      complexity  graph_density  avg_shortest_path  avg_deg_cent
24           0.0       0.200000           0.666667      0.133333
218          6.0       1.166667           0.333333      0.119048
222          0.0       0.000000           0.000000      0.000000
281        235.0       1.152174           0.388067      0.024977
283          0.0       0.000000           0.000000      0.000000
…          …            …                 …            …
4318       218.0       0.833333           0.250000      0.013115
4329       318.0       2.080000           2.283077      0.070769
4330       382.0       2.076923           1.072650      0.066952
4331       395.0       1.812500           0.953598      0.044508
4333       170.0       1.333333           0.282051      0.076923
```

```
[148 rows x 16 columns]
```

[26]:
```
plt.figure(figsize=(8,4))
plt.grid()
sns.lineplot(data=df_data[["no_hardware_comp","no_software_comp"]].describe().
 ↪drop("count",axis=0))
```

[26]:
```
<AxesSubplot: >
```

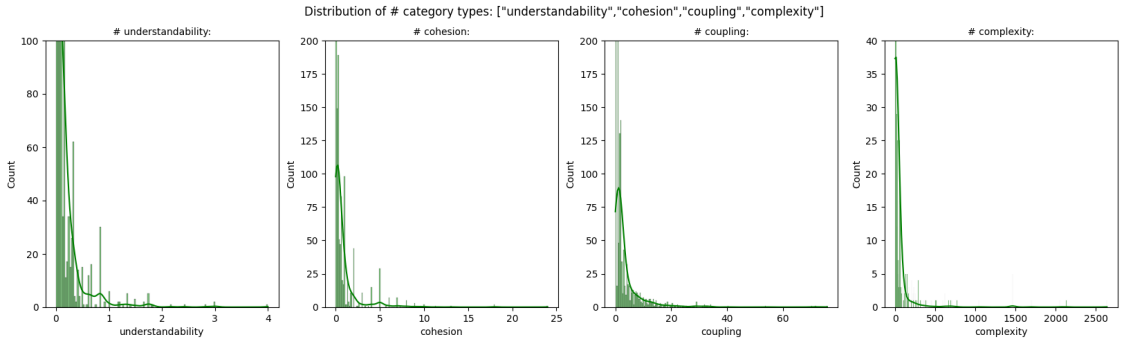


0.5.5 Distributions for understandability, cohesion, coupling

[27]:
```
fig, axes = plt.subplots(1, 4, figsize=(20,5))
fig.suptitle('Distribution of # category types:␣
 ↪["understandability","cohesion","coupling","complexity"]',fontsize=12)
axes[0].set_title('# understandability:',fontsize=10)
axes[1].set_title('# cohesion:',fontsize=10)
axes[2].set_title('# coupling:',fontsize=10)
axes[3].set_title('# complexity:',fontsize=10)
sns.
 ↪histplot(ax=axes[0],data=df_data['understandability'],kde=True,color='green');
 ↪
sns.histplot(ax=axes[1],data=df_data['cohesion'],kde=True,color='green');
sns.histplot(ax=axes[2],data=df_data['coupling'],kde=True,color='green');
sns.histplot(ax=axes[3],data=df_data['complexity'],kde=True,color='green');
axes[0].set_ylim([0, 100])
axes[1].set_ylim([0, 200])
axes[2].set_ylim([0, 200])
```

```
axes[3].set_ylim([0, 40])
```

[27]: (0.0, 40.0)



0.5.6 Description for understandability and cohesion

[28]:
```
df_data["coupling"].describe()
```

[28]:
```
count    1334.000000
mean        3.290963
std         6.130500
min         0.000000
25%         1.000000
50%         1.438462
75%         3.000000
max        75.961328
Name: coupling, dtype: float64
```

[29]:
```
df_data[(df_data["complexity"] >=6) & (df_data["complexity"] <= 8) ].head(10)
```

[29]:

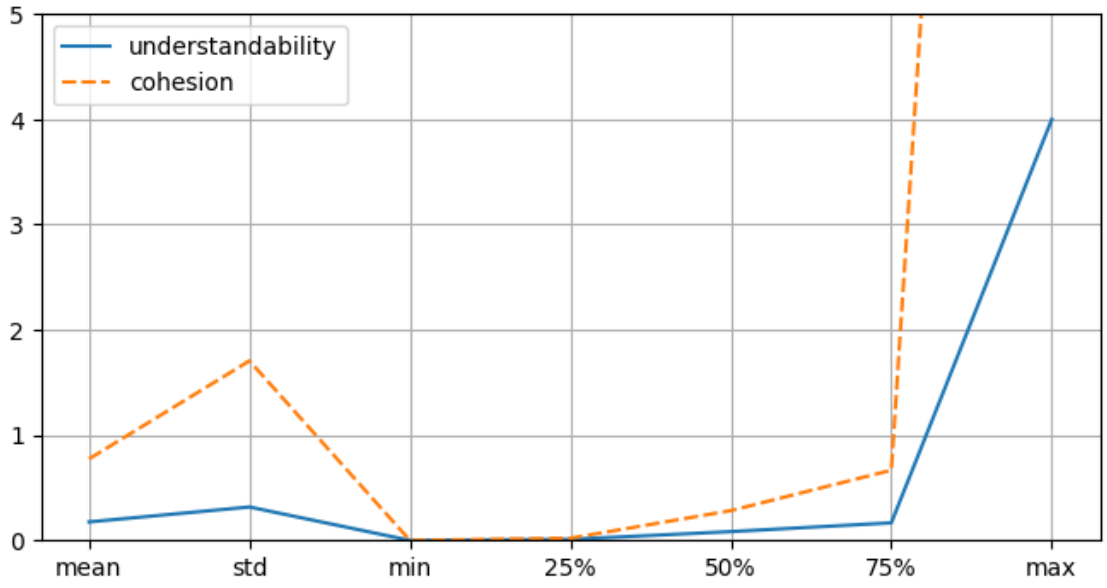
	model_name	sys_name	\
218	DualFGS	FGS_impl_Instance	
293	DigitalControlSystem	DCS_singletier1_Instance	
412	Refinement	Example_Low_Instance	
414	data_to_data_port	top_impl_Instance	
491	FlightSystem	FlightSystem_tier1parts_Instance	
492	FlightSystem	FlightSystem_tier1_Instance	
571	SubprogramWithSubprogram	Root_impl_Instance	
767	nestedcomposite	main_nestedstate_Instance	
969	GPSSystem	GPS_parts_TwoSensor_Instance	
970	GPSSystem	GPS_basic_Instance	

	no_components	no_connectors	no_hardware_comp	understandability	\
218	7.0	7.0	1.0	0.166667	

```
293            8.0           10.0              3.0           0.178571
412            3.0            1.0              0.0           0.166667
414           11.0           16.0              0.0           0.145455
491            5.0            0.0              1.0           0.000000
492            5.0            7.0              1.0           0.350000
571            9.0            3.0              0.0           0.041667
767           10.0            3.0              6.0           0.033333
969            6.0            0.0              5.0           0.000000
970            7.0           15.0              5.0           0.357143

     no_size  no_data_comp  no_software_comp  no_sys_comp  coupling  cohesion  \
218     14.0           0.0               0.0          6.0  2.666667  0.466667
293     18.0           0.0               2.0          3.0  5.166667  0.476190
412      4.0           0.0               0.0          3.0  1.000000  1.000000
414     27.0           1.0               9.0          1.0  4.000000  0.355556
491      5.0           0.0               0.0          4.0  1.916667  0.000000
492     12.0           0.0               0.0          4.0  1.916667  1.166667
571     12.0           0.0               8.0          1.0  3.500000  0.107143
767     13.0           0.0               3.0          1.0  2.500000  0.083333
969      6.0           0.0               0.0          1.0  3.500000  0.000000
970     22.0           0.0               1.0          1.0  4.166667  1.000000

     complexity  graph_density  avg_shortest_path  avg_deg_cent
218         6.0       1.166667           0.333333      0.119048
293         8.0       1.428571           0.428571      0.196429
412         8.0       0.500000           0.166667      0.166667
414         8.0       1.600000           1.163636      0.145455
491         6.0       0.000000           0.000000      0.000000
492         6.0       1.750000           0.700000      0.250000
571         7.0       0.375000           0.097222      0.041667
767         7.0       0.333333           0.166667      0.055556
969         6.0       0.000000           0.000000      0.000000
970         8.0       2.500000           1.095238      0.309524
```

[30]:
```
plt.figure(figsize=(8,4))
plt.grid()
sns.lineplot(data=df_data[["understandability","cohesion"]].describe().
 ↪drop("count",axis=0))
plt.ylim([0, 5])
```

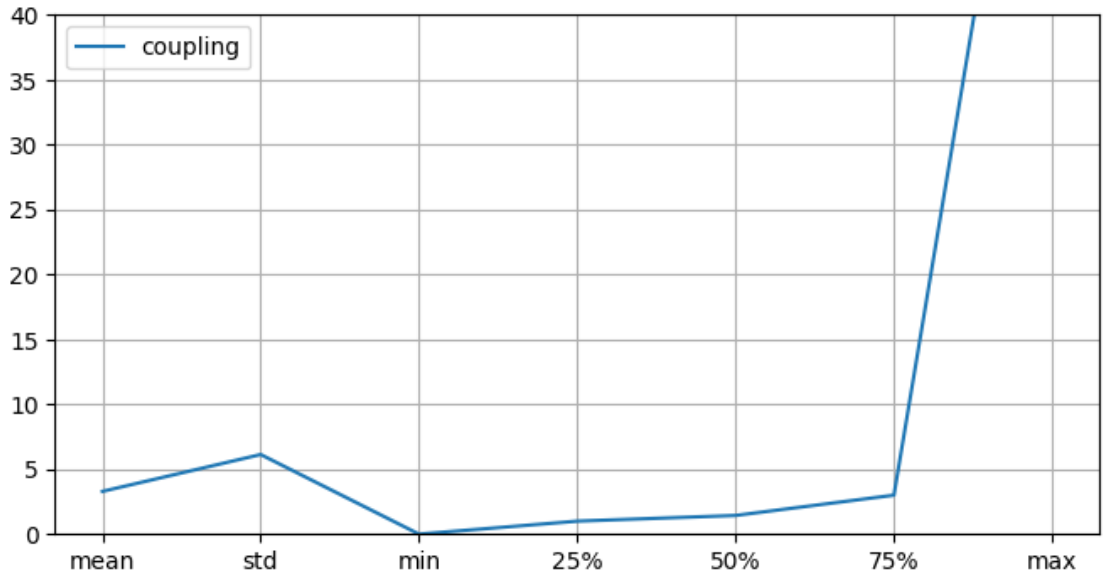
[30]:
```
(0.0, 5.0)
```



0.5.7 Description for coupling

[31]:
```
plt.figure(figsize=(8,4))
plt.grid()
sns.lineplot(data=df_data[["coupling"]].describe().drop("count",axis=0))
plt.ylim([0, 40])
```

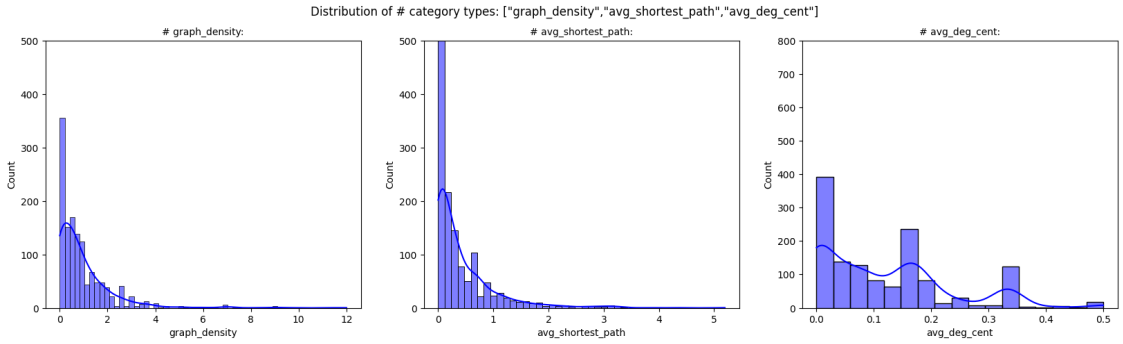
[31]: (0.0, 40.0)



0.5.8 Distributions for graph metrics graph_density, avg_shortest_path, avg_clust_coeff, avg_deg_cent

[32]:
```
fig, axes = plt.subplots(1, 3, figsize=(20,5))
fig.suptitle('Distribution of # category types:␣
 ↪["graph_density","avg_shortest_path","avg_deg_cent"]',fontsize=12)
axes[0].set_title('# graph_density:',fontsize=10)
axes[1].set_title('# avg_shortest_path:',fontsize=10)
axes[2].set_title('# avg_deg_cent:',fontsize=10)
sns.histplot(ax=axes[0],data=df_data['graph_density'],kde=True,color='blue');
sns.
 ↪histplot(ax=axes[1],data=df_data['avg_shortest_path'],kde=True,color='blue');
sns.histplot(ax=axes[2],data=df_data['avg_deg_cent'],kde=True,color='blue');
axes[0].set_ylim([0, 500])
axes[1].set_ylim([0, 500])
axes[2].set_ylim([0, 800])
```

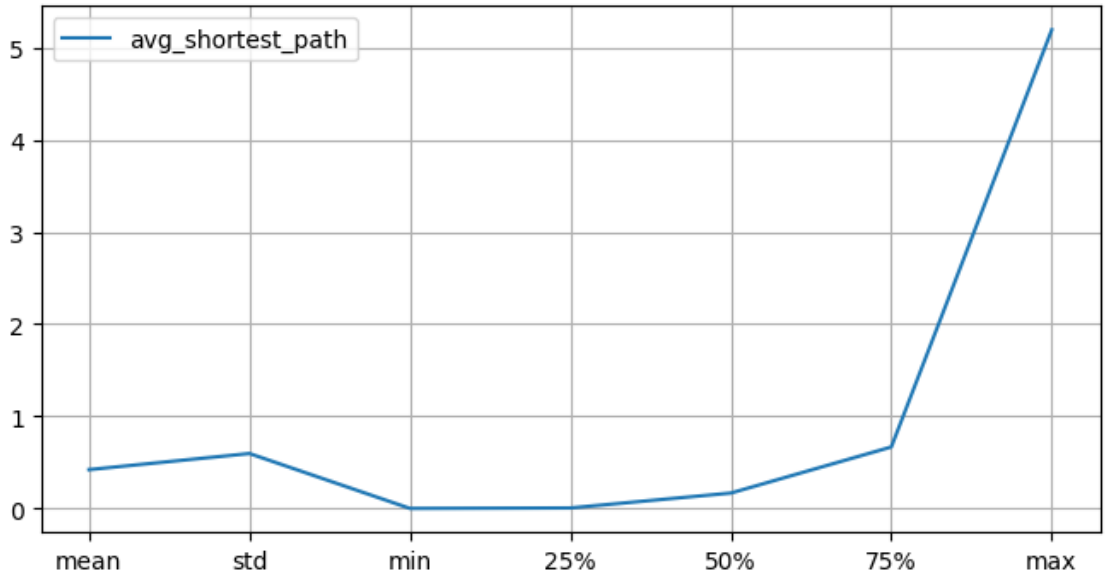
[32]:
```
(0.0, 800.0)
```



0.5.9 Description for avg_shortest_path, graph_density, avg_deg_cent, avg_clust_coeff

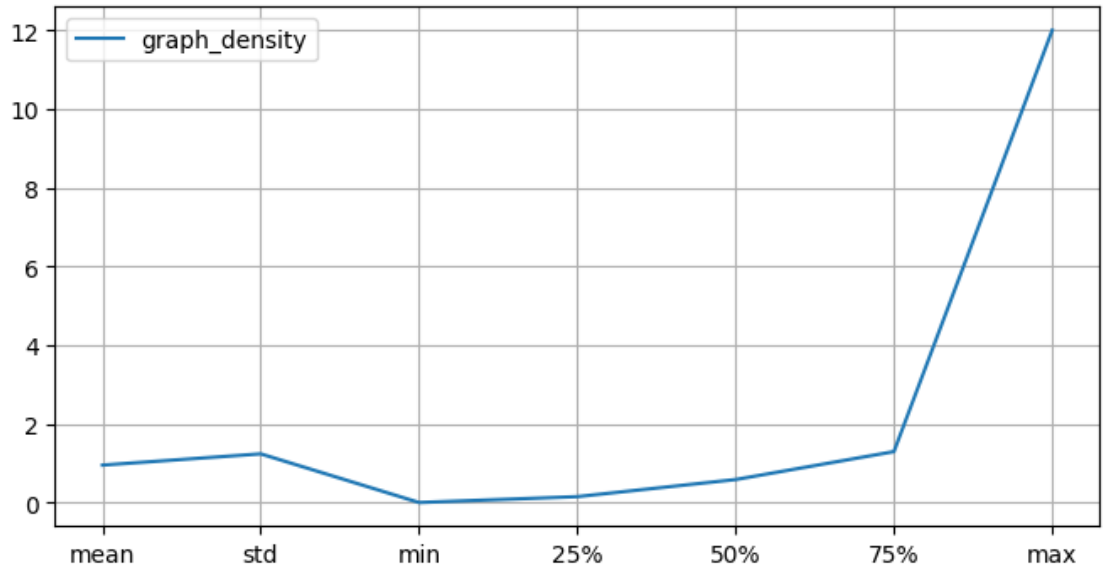
[33]:
```
plt.figure(figsize=(8,4))
plt.grid()
sns.lineplot(data=df_data[["avg_shortest_path"]].describe().
 ↪drop("count",axis=0))
```

[33]:
```
<AxesSubplot: >
```



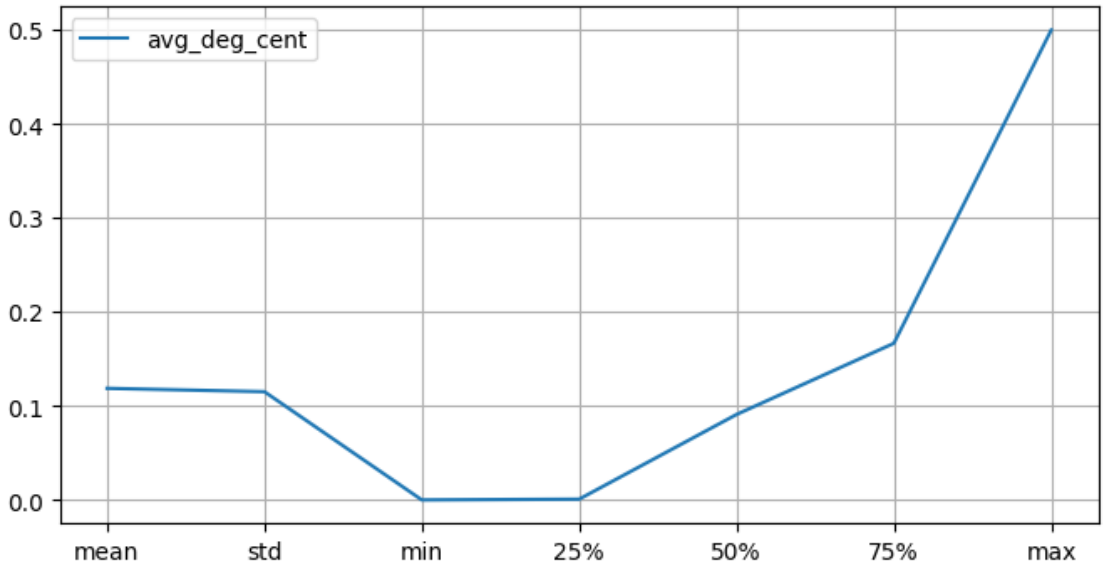
[34]:
```
plt.figure(figsize=(8,4))
plt.grid()
sns.lineplot(data=df_data[["graph_density"]].describe().drop("count",axis=0))
```

[34]: <AxesSubplot: >



[35]:
```
plt.figure(figsize=(8,4))
plt.grid()
sns.lineplot(data=df_data[["avg_deg_cent"]].describe().drop("count",axis=0))
```

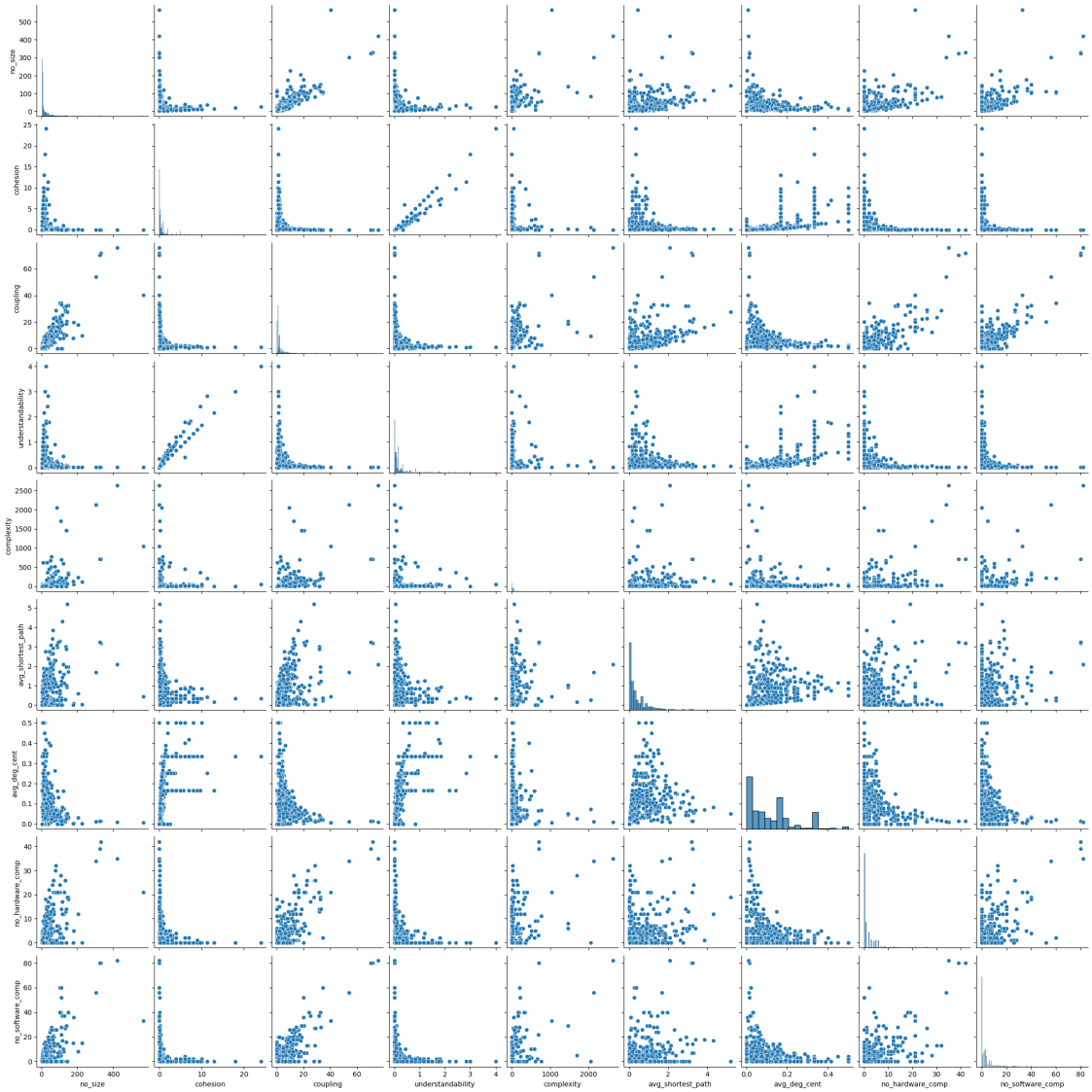
[35]: <AxesSubplot: >



0.6 Trying to get reelation between variables

[36]:
```
sns.
↪pairplot(df_data[["no_size","cohesion","coupling","understandability","complexity","avg_sho
                   "no_software_comp"]])
```

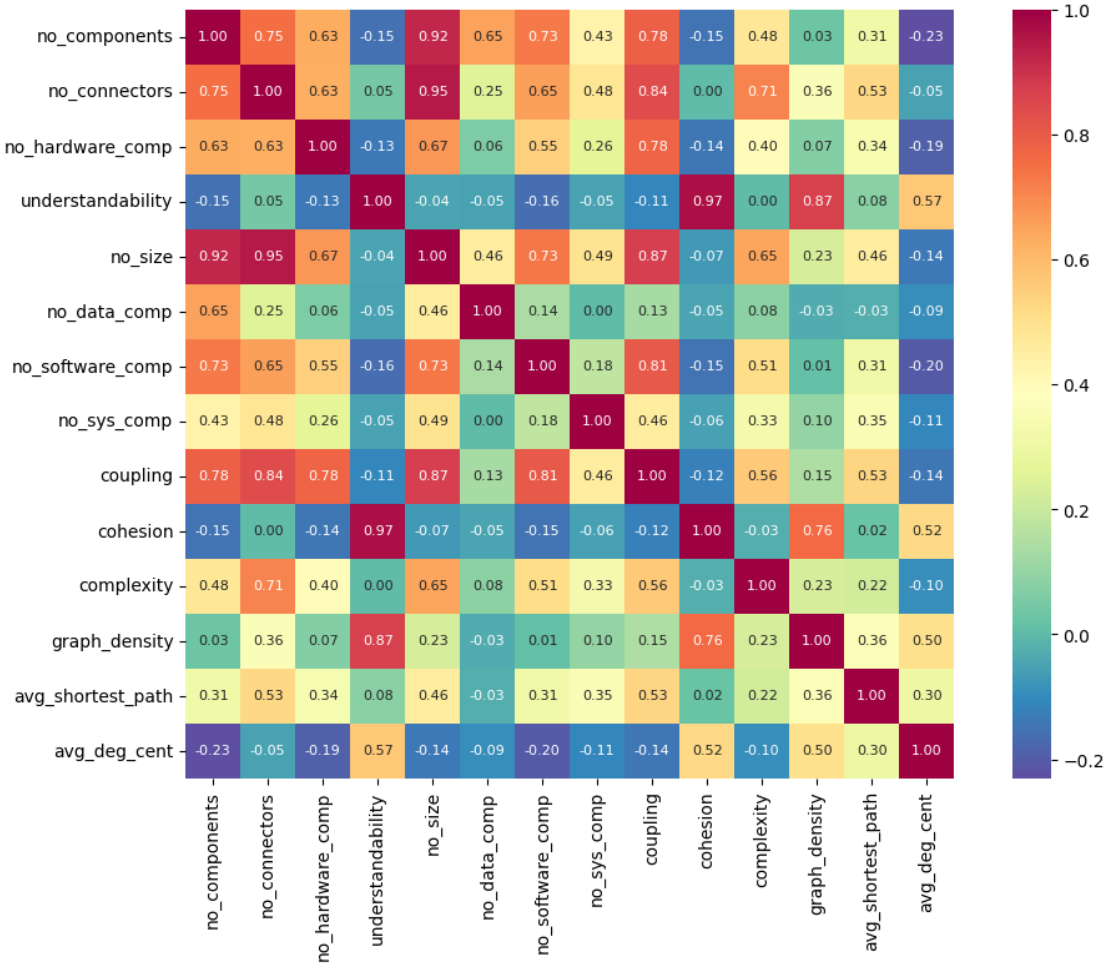
[36]: <seaborn.axisgrid.PairGrid at 0x7fdce1f93160>



0.6.1 Correlation Matrix

[37]:
```
plt.figure(figsize=(12,8))
df_num_corr =  df_data.loc[:, ~df_data.columns.isin(['model_name', "sys_name"])]
corrmat = df_num_corr.corr()
hm = sns.heatmap(corrmat,
                 cbar=True,
                 annot=True,
                 square=True,
                 fmt='.2f',
                 annot_kws={'size': 8},
                 yticklabels=df_num_corr.columns,
                 xticklabels=df_num_corr.columns,
```

```
                    cmap="Spectral_r")
plt.show()
```



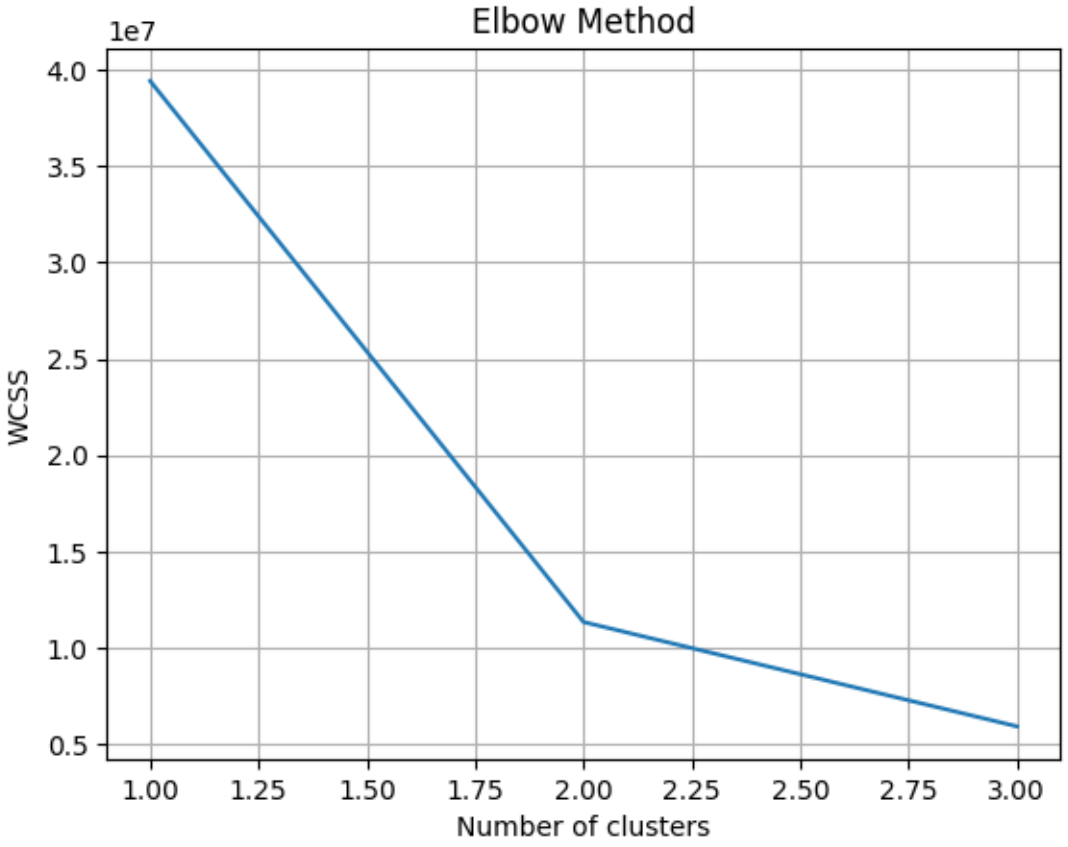
0.7 Clustering

[38]:
```
from numpy import unique
from numpy import where
from sklearn.datasets import make_classification
from sklearn.cluster import KMeans
```

[39]:
```
X =  df_data.loc[:, ~df_data.columns.isin(['model_name', "sys_name"])].values
```

[40]:
```
wcss = []
no_clusters = 4
for i in range(1, no_clusters):
```

```
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10,␣
 ↪random_state=0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)

# Plot the WCSS versus the number of clusters
plt.plot(range(1, no_clusters), wcss)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.grid()
plt.show()
```



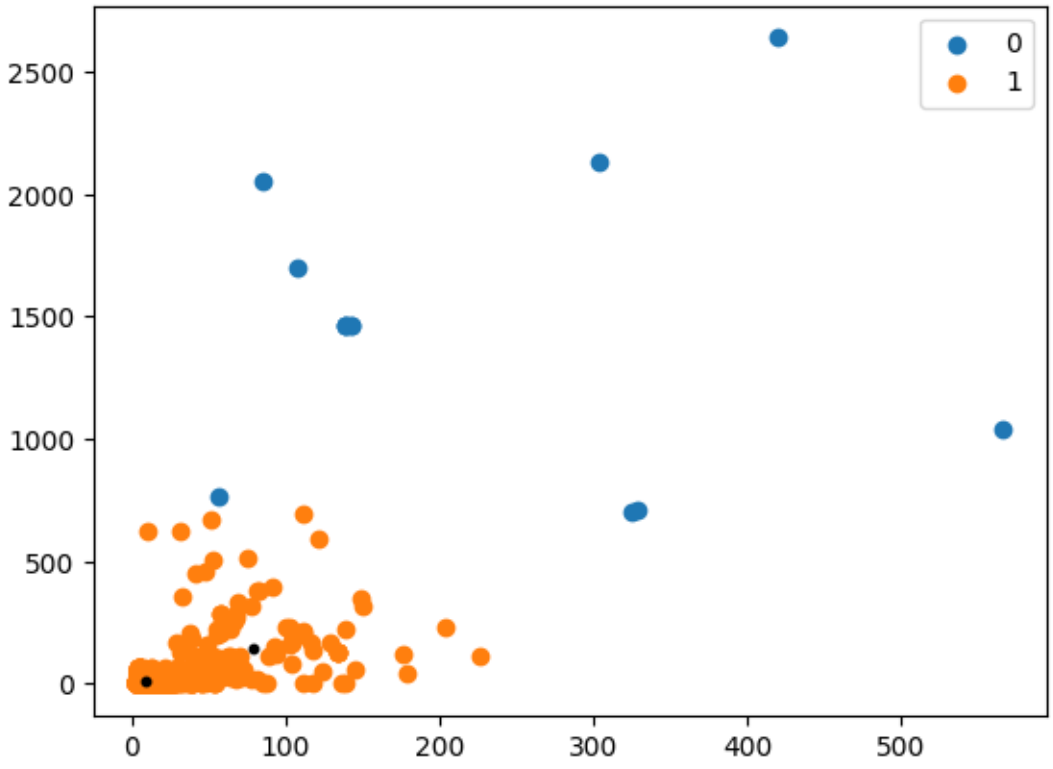
[41]:
```
model = KMeans(n_clusters=2,n_init="auto")
model.fit(X)
```

[41]:
```
KMeans(n_clusters=2, n_init='auto')
```

[42]:
```
# assign a cluster to each example
labels = model.predict(X)
clusters = unique(labels)
centroids = model.cluster_centers_
print(labels)
print(clusters)
```

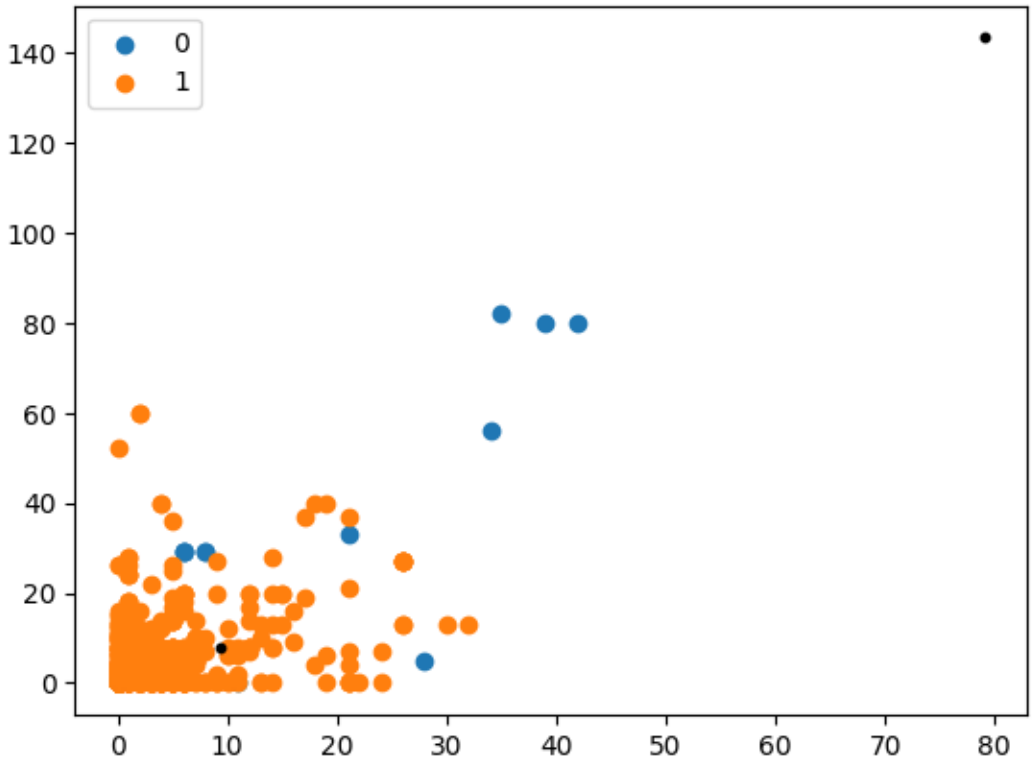
```
[1 1 1 … 1 1 1]
[0 1]
```

[43]:
```
for cluster in clusters:
    plt.scatter(df_data[labels == cluster]["no_size"],df_data[labels ==␣
 ↪cluster]["complexity"], label = cluster)
plt.scatter(centroids[:,0] , centroids[:,1] , s = 10, color = 'k')
plt.legend()
plt.show()
```

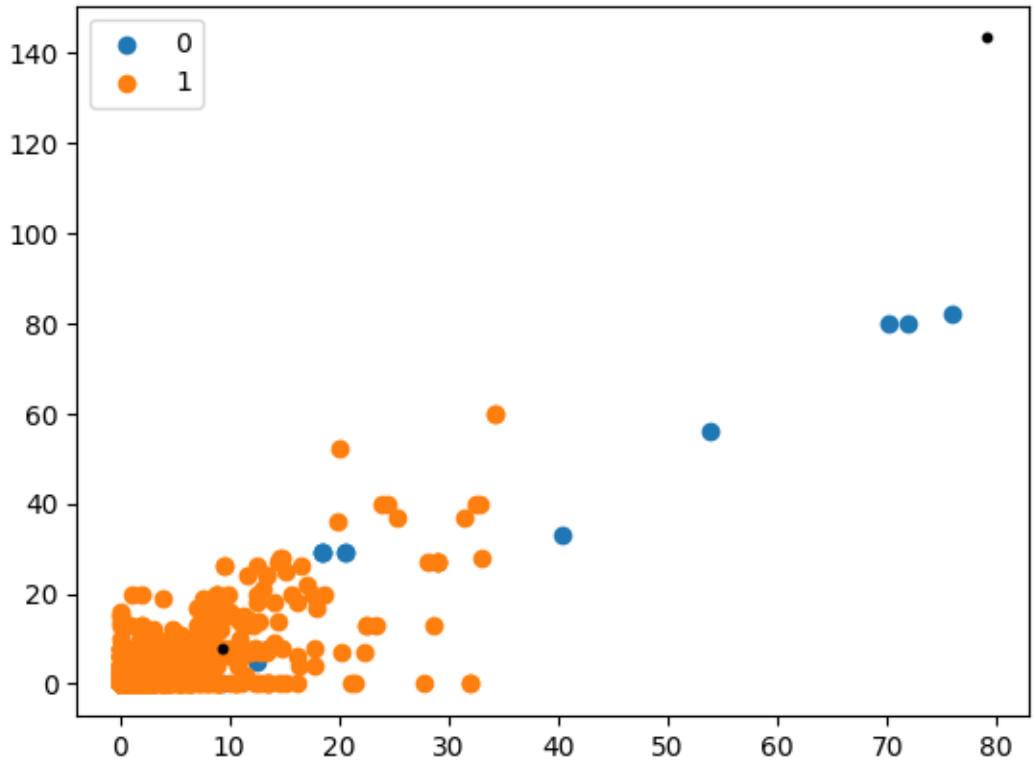


[44]:
```
for cluster in clusters:
    plt.scatter(df_data[labels == cluster]["no_hardware_comp"],df_data[labels␣
 ↪== cluster]["no_software_comp"], label = cluster)
plt.scatter(centroids[:,0] , centroids[:,1] , s = 10, color = 'k')
plt.legend()
```

```
plt.show()
```



[45]:
```
for cluster in clusters:
    plt.scatter(df_data[labels == cluster]["coupling"],df_data[labels ==␣
 ↪cluster]["no_software_comp"], label = cluster)
plt.scatter(centroids[:,0] , centroids[:,1] , s = 10, color = 'k')
plt.legend()
plt.show()
```



[46]:
```
df_data[labels == 0]
```

[46]:

	model_name	sys_name	\
1306	Boards::Ardupilot	Ardupilot_impl_Instance	
1361	Main_Ardupilot	Ardupilot_Map_impl_Instance	
1423	RAP	RAP_GENERIC_IMPL_Instance	
1424	RAP	RAP_LEON_RTEMS_Instance	
1425	RAP	RAP_LEON_ORK_Instance	
1426	RAP	RAP_ERC32_ORK_Instance	
1427	RAP	RAP_Native_Instance	
1598	complete	complete_impl_Instance	
1605	systems	ARS_impl_Instance	
2890	UAV	UAV_Impl_Instance	
2929	FCS	FCS_Impl_Instance	
4304	Overall	APSCommunicationArchitecture_impl_Instance	
4316	APS	APSCommunication_impl_Instance	

	no_components	no_connectors	no_hardware_comp	understandability	\
1306	12.0	44.0	11.0	0.333333	
1361	35.0	72.0	28.0	0.060504	
1423	36.0	102.0	6.0	0.080952	
1424	36.0	102.0	6.0	0.080952	

```
1425          38.0          104.0              8.0          0.073969
1426          38.0          104.0              8.0          0.073969
1427          36.0          102.0              6.0          0.080952
1598         140.0          189.0             42.0          0.009712
1605         136.0          189.0             39.0          0.010294
2890         200.0          366.0             21.0          0.009196
2929          17.0           68.0              0.0          0.250000
4304         180.0          240.0             35.0          0.007449
4316         123.0          180.0             34.0          0.011995

      no_size  no_data_comp  no_software_comp  no_sys_comp   coupling  \
1306     56.0           0.0               0.0          1.0   2.807857
1361    107.0           0.0               5.0          2.0  12.505714
1423    138.0           0.0              29.0          1.0  18.509280
1424    138.0           0.0              29.0          1.0  18.509280
1425    142.0           0.0              29.0          1.0  20.509280
1426    142.0           0.0              29.0          1.0  20.509280
1427    138.0           0.0              29.0          1.0  18.509280
1598    329.0           5.0              80.0         13.0  71.897223
1605    325.0           5.0              80.0         12.0  70.230557
2890    566.0         141.0              33.0          5.0  40.317128
2929     85.0           0.0               0.0         17.0   9.362183
4304    420.0           0.0              82.0         63.0  75.961328
4316    303.0           0.0              56.0         33.0  53.888889

      cohesion  complexity  graph_density  avg_shortest_path  avg_deg_cent
1306  0.800000       769.0       4.000000           0.204545      0.068182
1361  0.128342      1698.0       2.117647           0.160504      0.025210
1423  0.171429      1463.0       2.914286           1.034921      0.050000
1424  0.171429      1463.0       2.914286           1.034921      0.050000
1425  0.156156      1463.0       2.810811           0.928876      0.046230
1426  0.156156      1463.0       2.810811           0.928876      0.046230
1427  0.171429      1463.0       2.914286           1.034921      0.050000
1598  0.019706       707.0       1.359712           3.184943      0.013669
1605  0.020896       705.0       1.400000           3.237255      0.014270
2890  0.018578      1041.0       1.839196           0.431583      0.007588
2929  0.566667      2052.0       4.250000           0.250000      0.073529
4304  0.015065      2639.0       1.340782           2.085723      0.008411
4316  0.024387      2131.0       1.475410           1.685259      0.011995
```

[47]:
```
df_data[labels == 1]
```

[47]:
```
                          model_name  \
1               isolette_heat_source
8               isolette_heat_source
17    JustSubprogramGroup_process_fg
18    JustSubprogramGroup_process_fg
```

```
21                              sc3
…                               …
4338                            APS
4339                            APS
4341                            APS
4347                Mobile5GNetwork
4360                           syst

                                    sys_name  no_components  no_connectors  \
1           Heat_Source_with_devices_Instance            8.0            7.0
8                   Heat_Source_impl_Instance            6.0            0.0
17                         p1_impl_Instance            4.0            2.0
18                        Root_impl_Instance            5.0            2.0
21                           src3_i_Instance            3.0            1.0
…                                          …            …              …
4338             COLAVandSSITAW_impl_Instance           12.0            6.0
4339      COLAVandSSITAW_Backup_impl_Instance           12.0            8.0
4341  EmergencyButtonTransmitter_impl_Instance            3.0            2.0
4347           APSCommunication_impl_Instance            3.0            8.0
4360     SecuritySystem_with_devices_Instance           13.0           12.0

      no_hardware_comp  understandability  no_size  no_data_comp  \
1                  4.0           0.125000     15.0           0.0
8                  4.0           0.000000      6.0           0.0
17                 0.0           0.166667      6.0           0.0
18                 0.0           0.100000      7.0           0.0
21                 0.0           0.166667      4.0           0.0
…                  …                  …        …             …
4338               0.0           0.045455     18.0           0.0
4339               0.0           0.060606     20.0           0.0
4341               1.0           0.333333      5.0           0.0
4347               0.0           1.333333     11.0           0.0
4360               7.0           0.076923     25.0           0.0

      no_software_comp  no_sys_comp  coupling  cohesion  complexity  \
1                  3.0          1.0  3.166667  0.333333        11.0
8                  0.0          2.0  1.000000  0.000000         2.0
17                 3.0          1.0  1.000000  0.666667         0.0
18                 3.0          2.0  1.000000  0.333333         0.0
21                 0.0          3.0  0.000000  1.000000         0.0
…                  …            …         …         …            …
4338              11.0          1.0  5.000000  0.109091        30.0
4339              11.0          1.0  5.000000  0.145455        30.0
4341               0.0          2.0  0.500000  2.000000         1.0
4347               0.0          3.0  1.000000  8.000000         8.0
4360               5.0          1.0  6.766667  0.181818        16.0
```

```
      graph_density  avg_shortest_path  avg_deg_cent
1          1.000000           0.517857      0.125000
8          0.000000           0.000000      0.000000
17         0.666667           0.666667      0.333333
18         0.500000           0.400000      0.200000
21         0.500000           0.666667      0.333333
…             …                  …             …
4338       0.545455           0.037879      0.022727
4339       0.727273           0.030303      0.030303
4341       1.000000           0.666667      0.333333
4347       4.000000           0.333333      0.333333
4360       1.000000           0.692308      0.102564

[1321 rows x 16 columns]
```

[48]:
```
df_data[labels == 2]
```

[48]:
```
Empty DataFrame
Columns: [model_name, sys_name, no_components, no_connectors, no_hardware_comp,
understandability, no_size, no_data_comp, no_software_comp, no_sys_comp,
coupling, cohesion, complexity, graph_density, avg_shortest_path, avg_deg_cent]
Index: []
```

[49]:
```
df_data[labels == 3]
```

[49]:
```
Empty DataFrame
Columns: [model_name, sys_name, no_components, no_connectors, no_hardware_comp,
understandability, no_size, no_data_comp, no_software_comp, no_sys_comp,
coupling, cohesion, complexity, graph_density, avg_shortest_path, avg_deg_cent]
Index: []
```

[ ]:
```

```