

Visualizations

Video overview

- Visualization tools to...
 - **Explore individual features**
 - Histograms
 - Plots
 - Statistics
 - **Explore feature relations**
 - Scatter plots
 - Correlation plots
 - Plot (index vs feature statistics)
 - And more

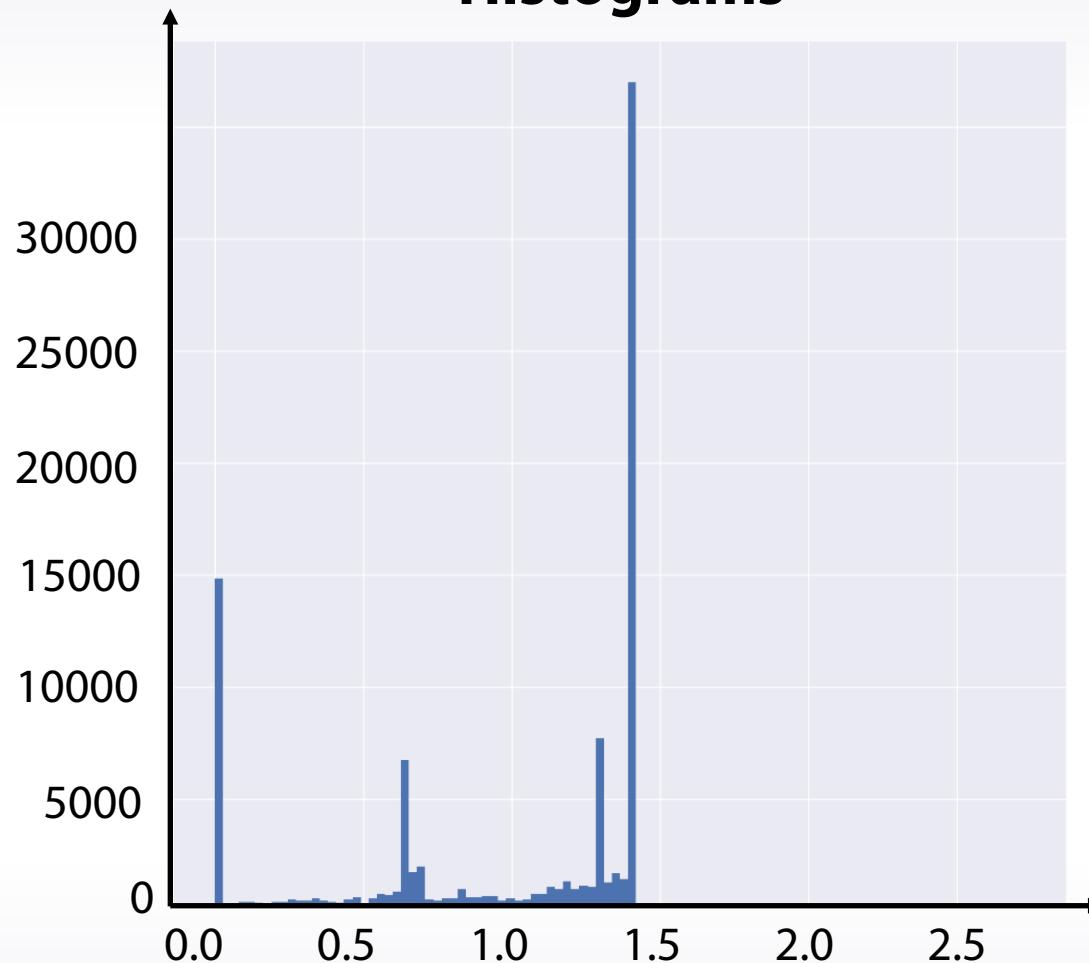
Visualizations

EDA is an art!

And visualizations are our art tools

Art tools

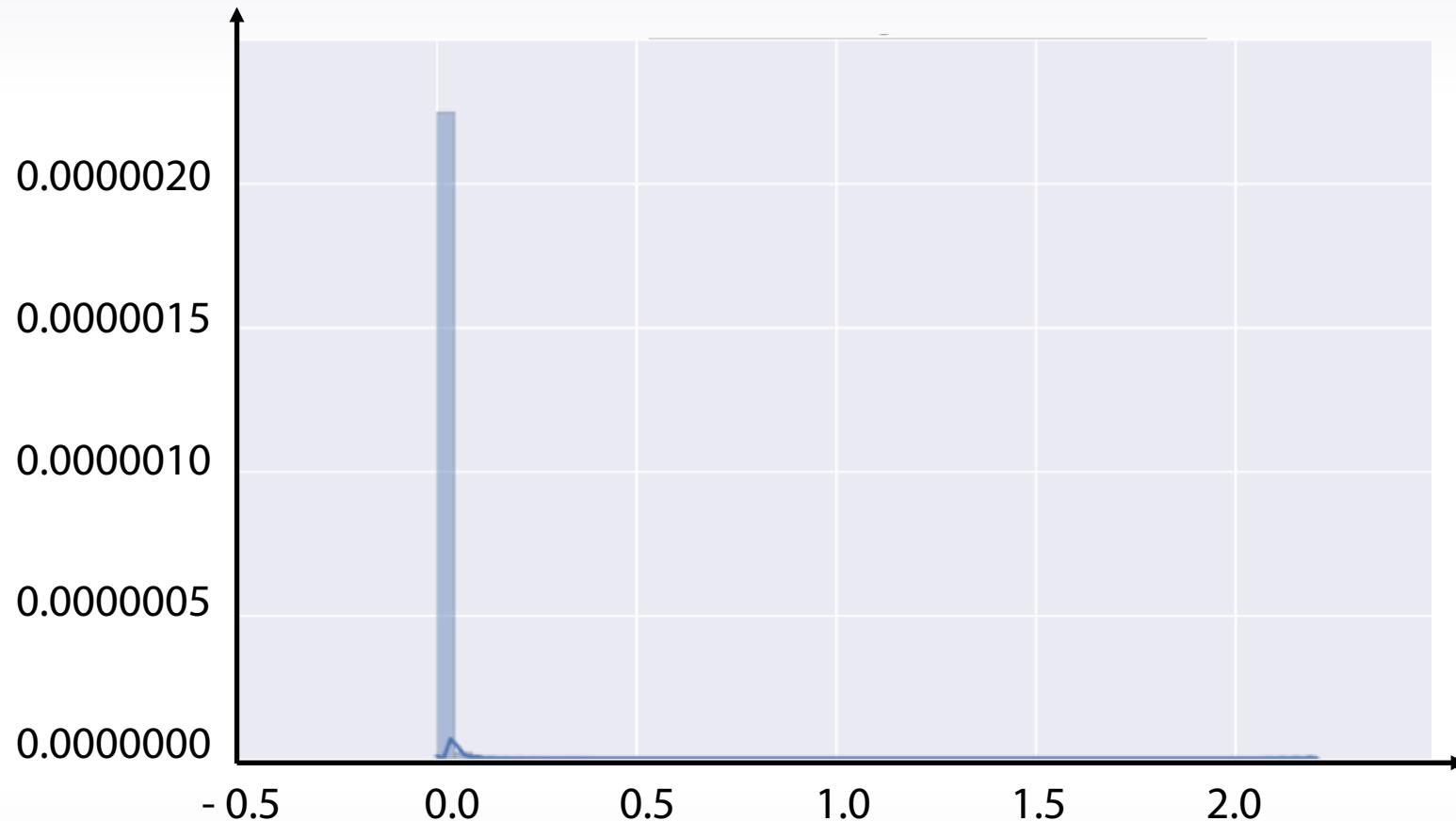
Histograms



```
| plt.hist(x)
```

Art tools

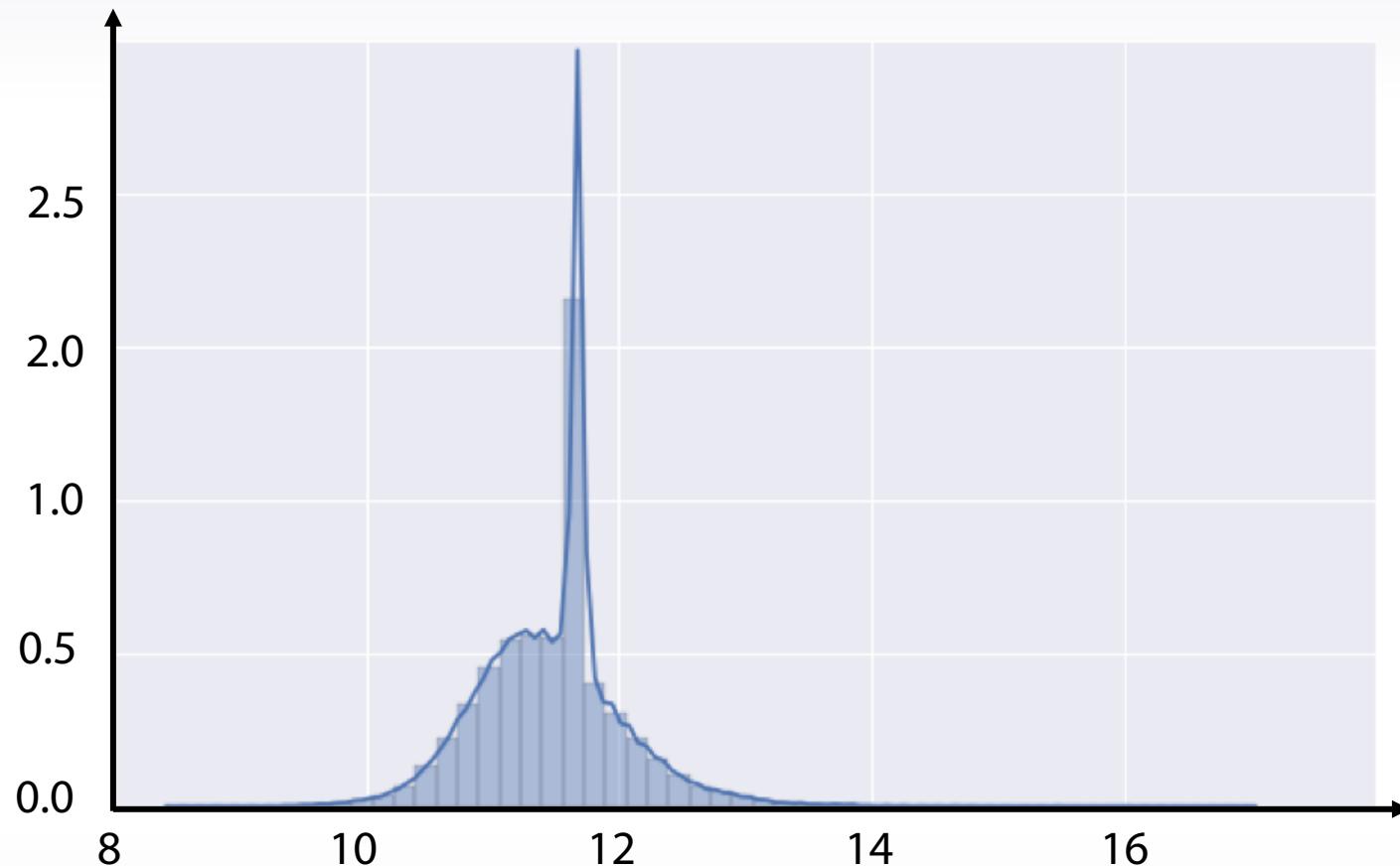
Histograms



```
| plt.hist(x)
```

Art tools

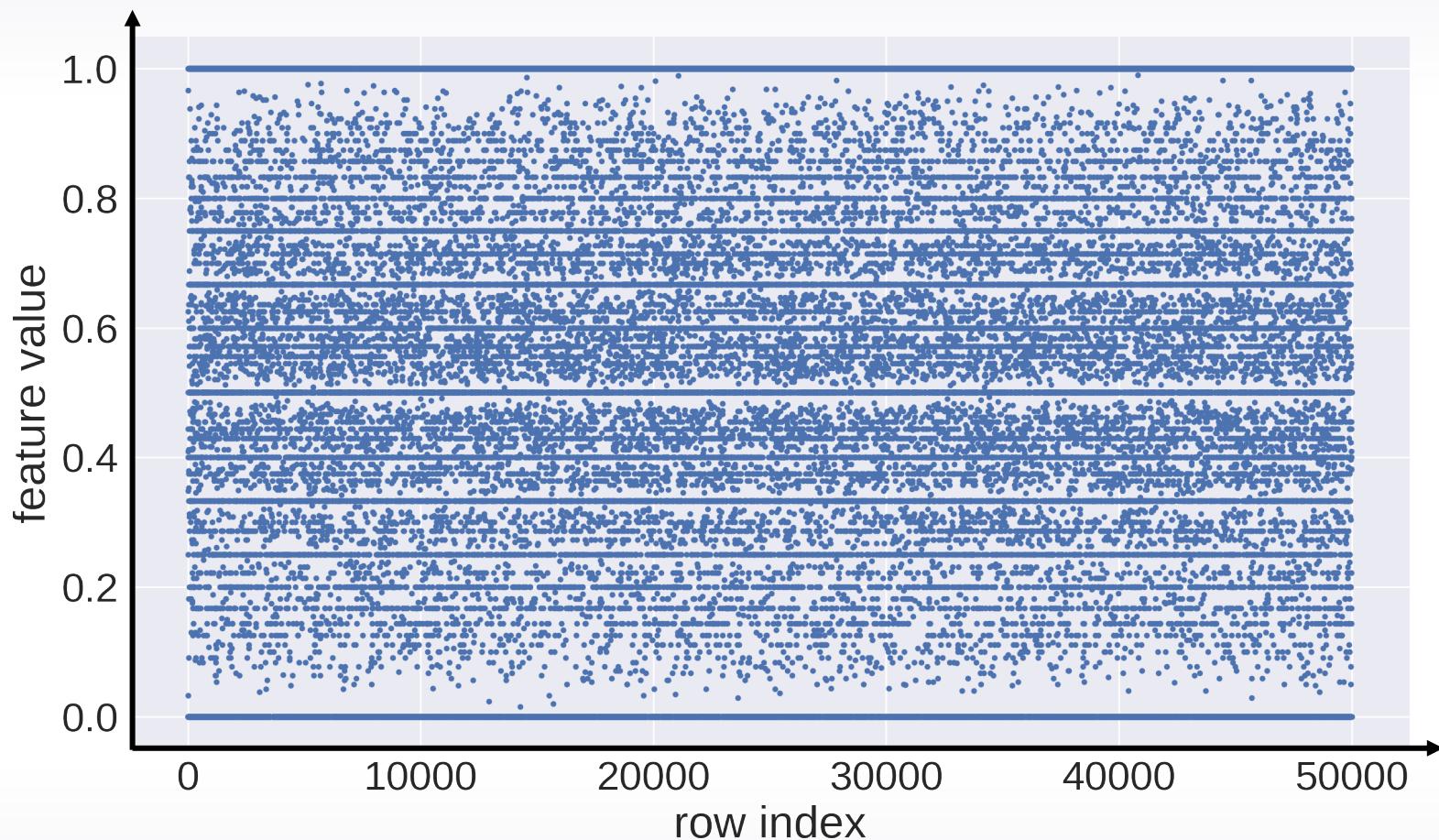
Histograms



```
| plt.hist(x)
```

Art tools

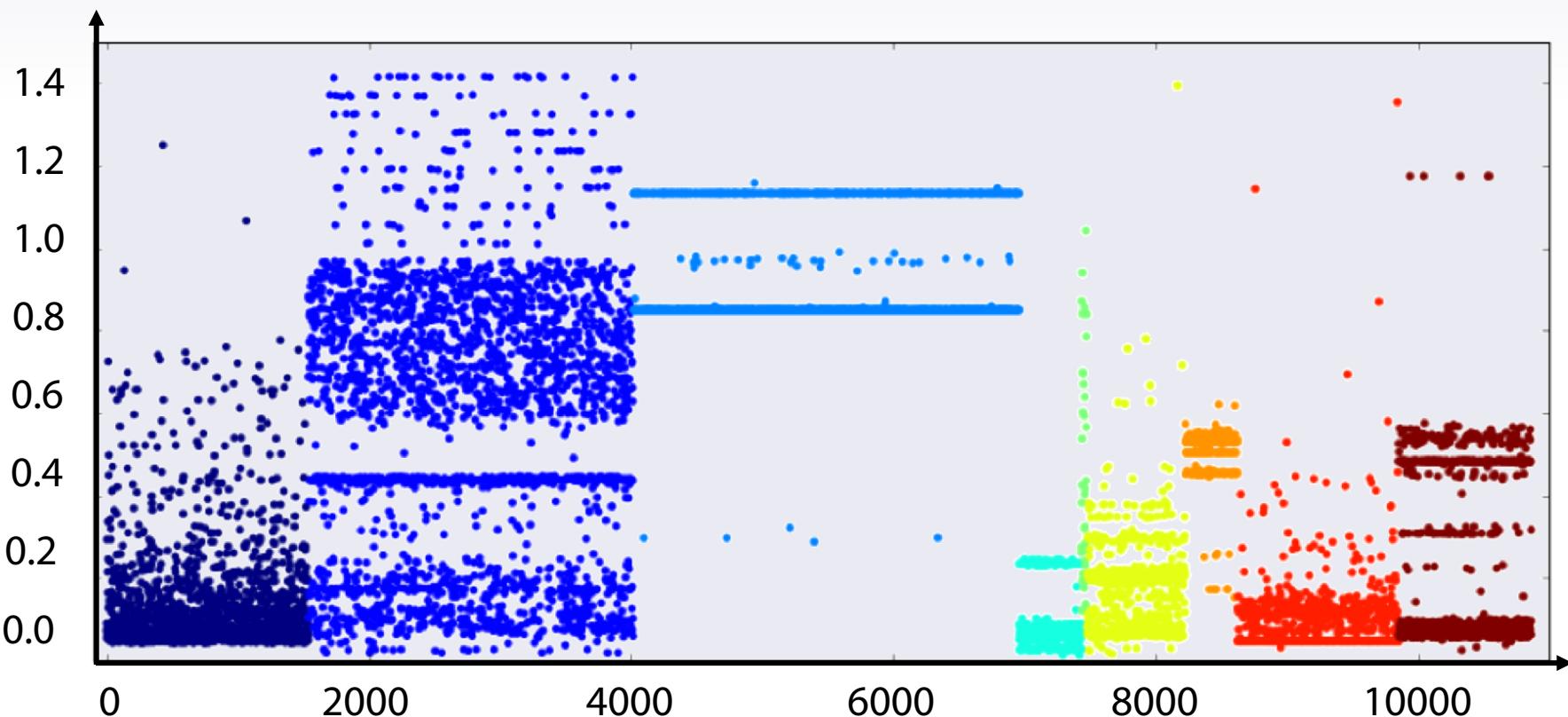
Plot (index versus value)



```
| plt.plot(x, '.' )
```

Art tools

Plot (index versus value)



```
| plt.scatter(range(len(x)), x, c=y)
```

Art tools

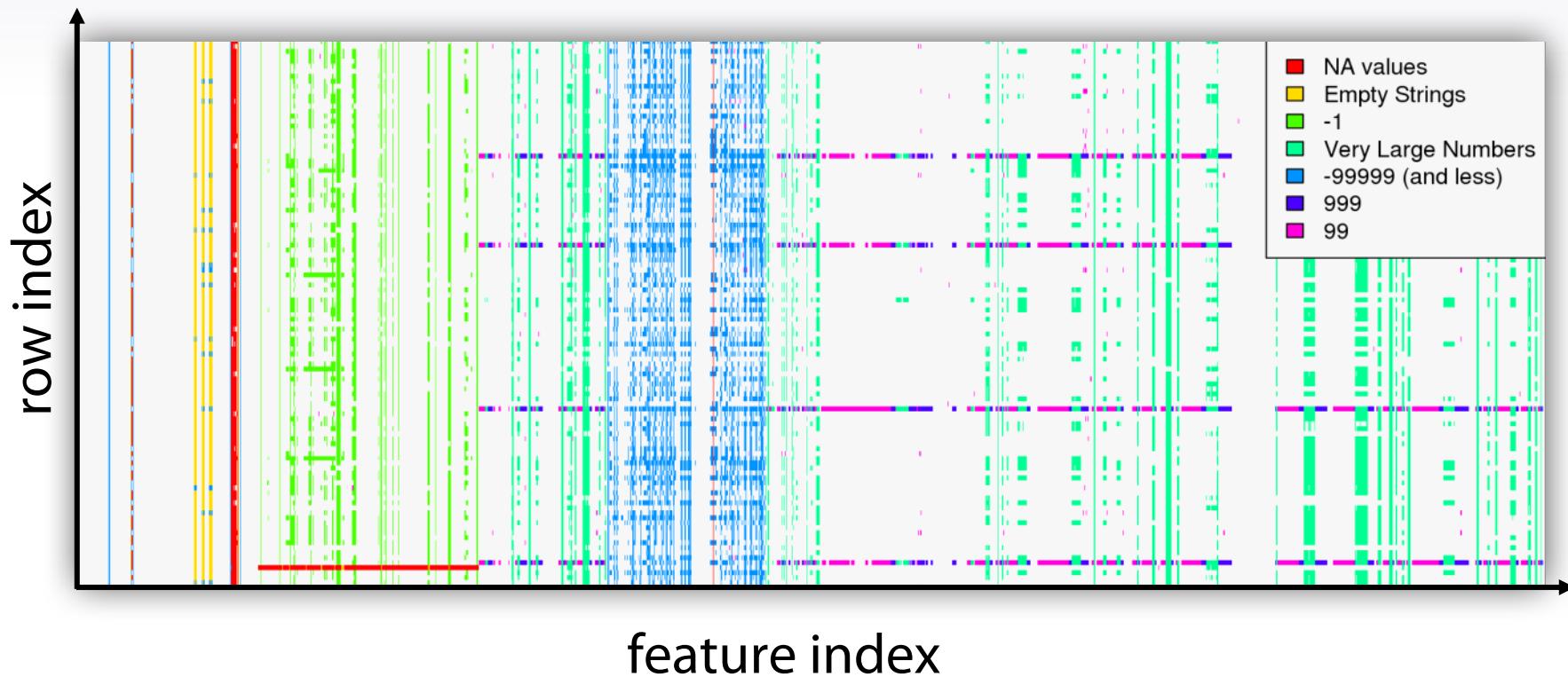
Feature statistics

	x6	x7	x8	x13
count	50000.00000	50000.00000	48793.00000	45512.00000
mean	0.99296	0.975860	-0.000252	4428.915253
std	0.08361	0.153485	1.023282	10943.884658
min	0.00000	0.000000	-85.252444	-99.000000
25%	1.00000	1.000000	-0.255490	0.000000
50%	1.00000	1.000000	0.221047	1817.000000
75%	1.00000	1.000000	0.567620	5582.000000
max	1.00000	1.000000	3.426844	776759.000000

```
df.describe()
x.mean()
x.var()
```

Art tools

Other tools



```
x.value_counts()  
x.isnull()
```

Tools for individual features exploration

Histograms:

| plt.hist(x)

Plot (index versus value):

| plt.plot(x, '.')

Statistics:

| df.describe()

| x.mean()

| x.var()

Other tools:

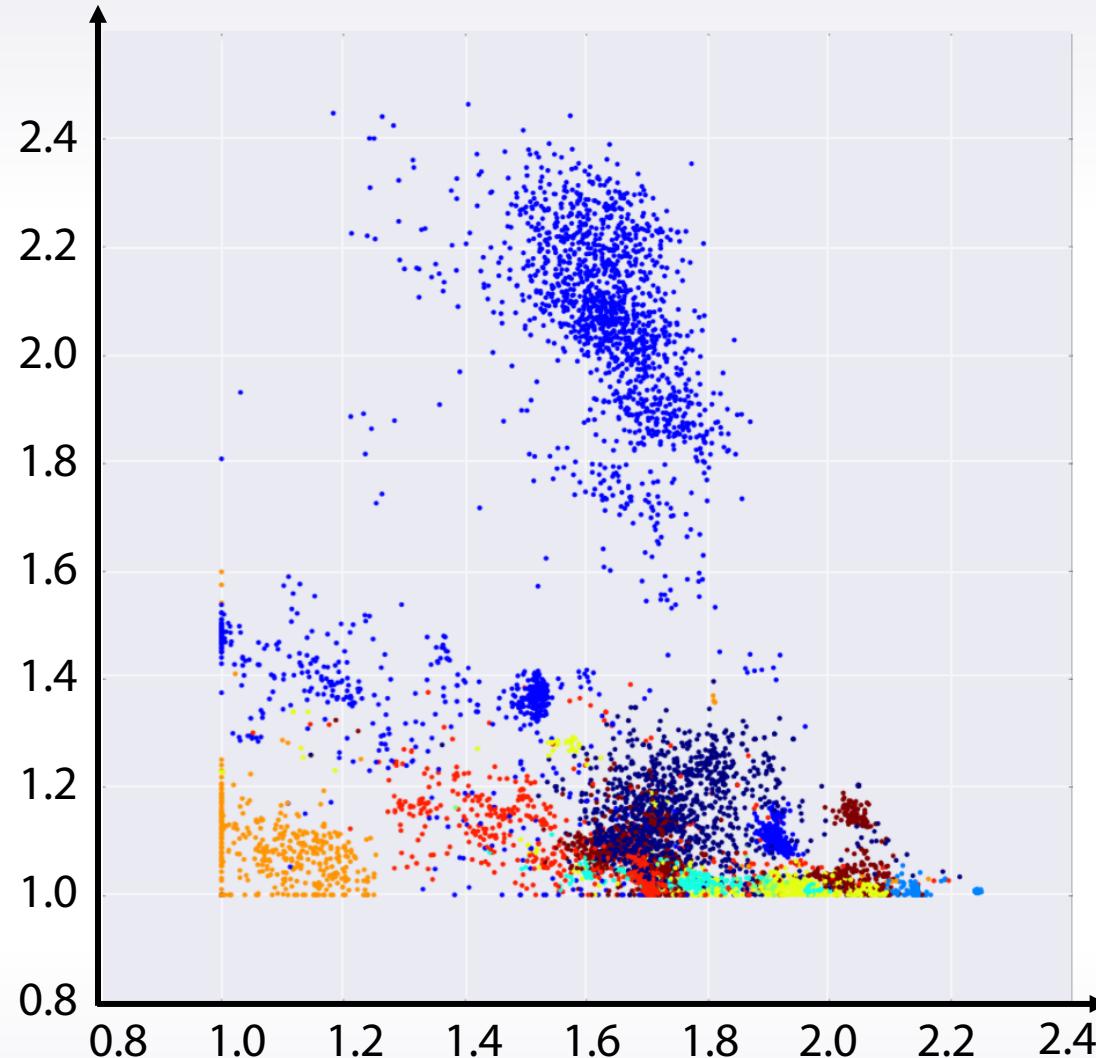
| x.value_counts()

| x.isnull()

Next in this video

- Visualization tools to...
 - **Explore individual features**
 - Histograms
 - Plots
 - Statistics
 - **Explore feature relations**
 - Scatter plots
 - Correlation matrices
 - ...

Exploring feature relations



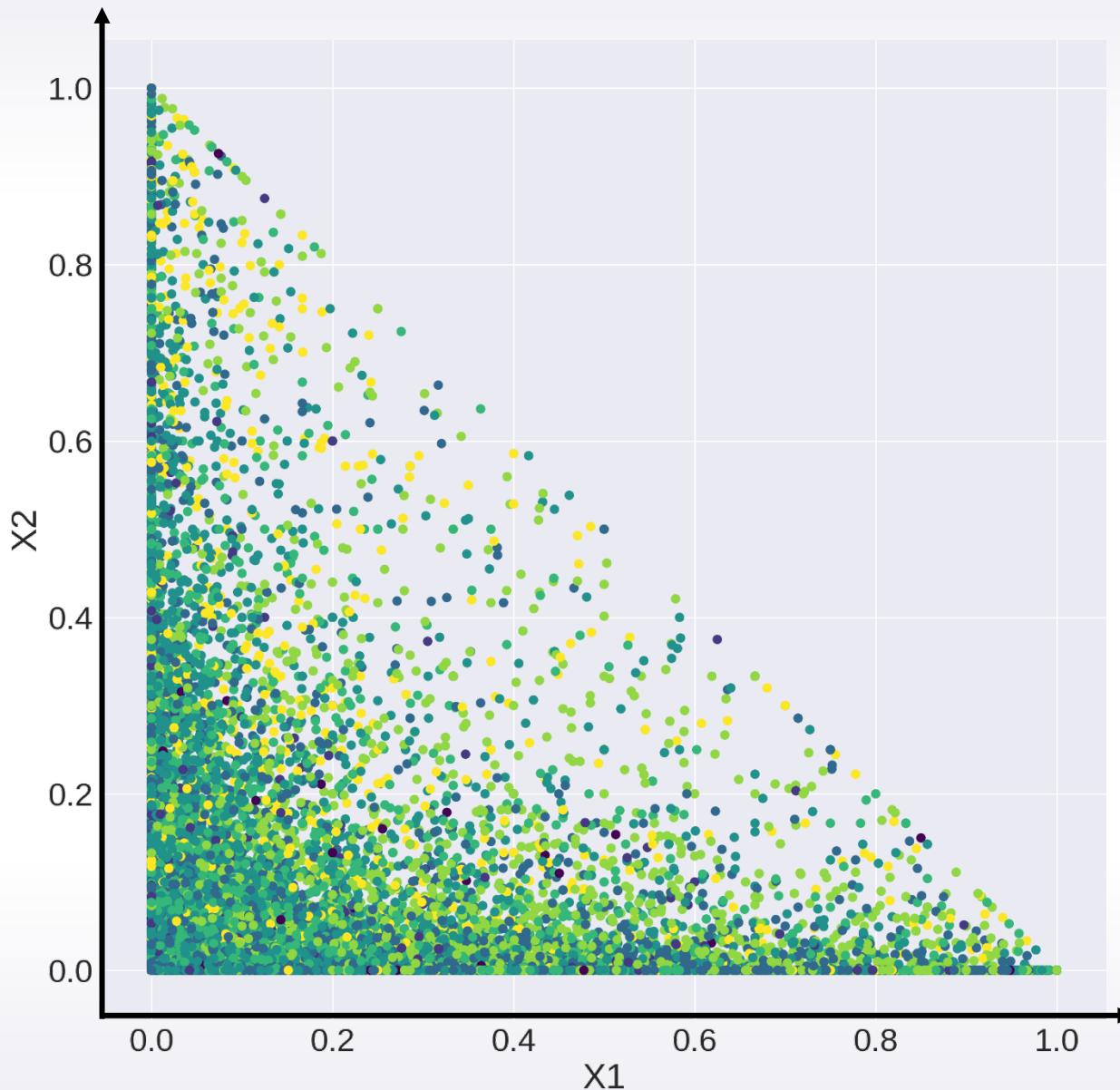
```
| plt.scatter(x1, x2)
```

Exploring feature relations

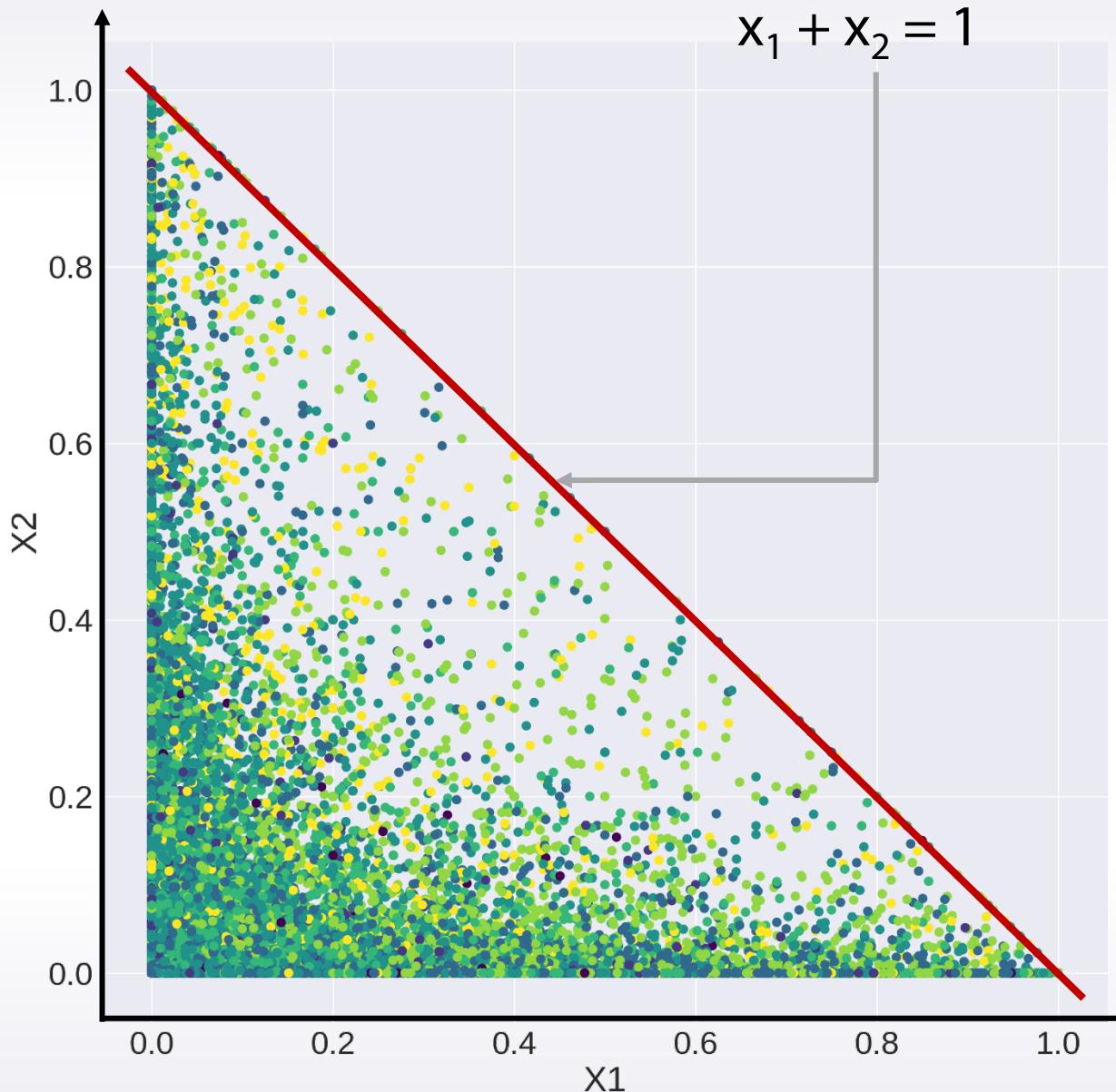


```
| plt.scatter(x1, x2)
```

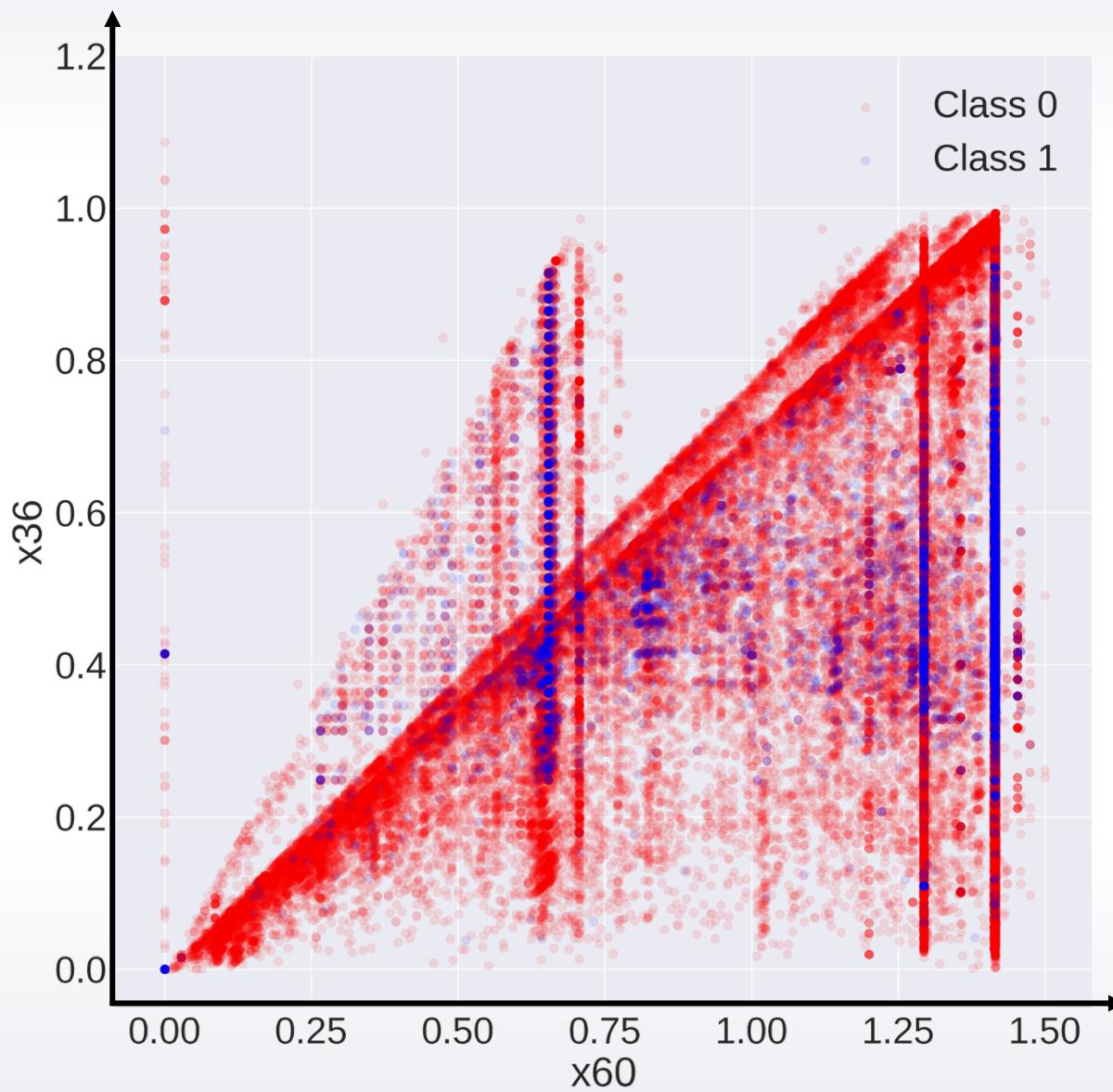
Exploring feature relations



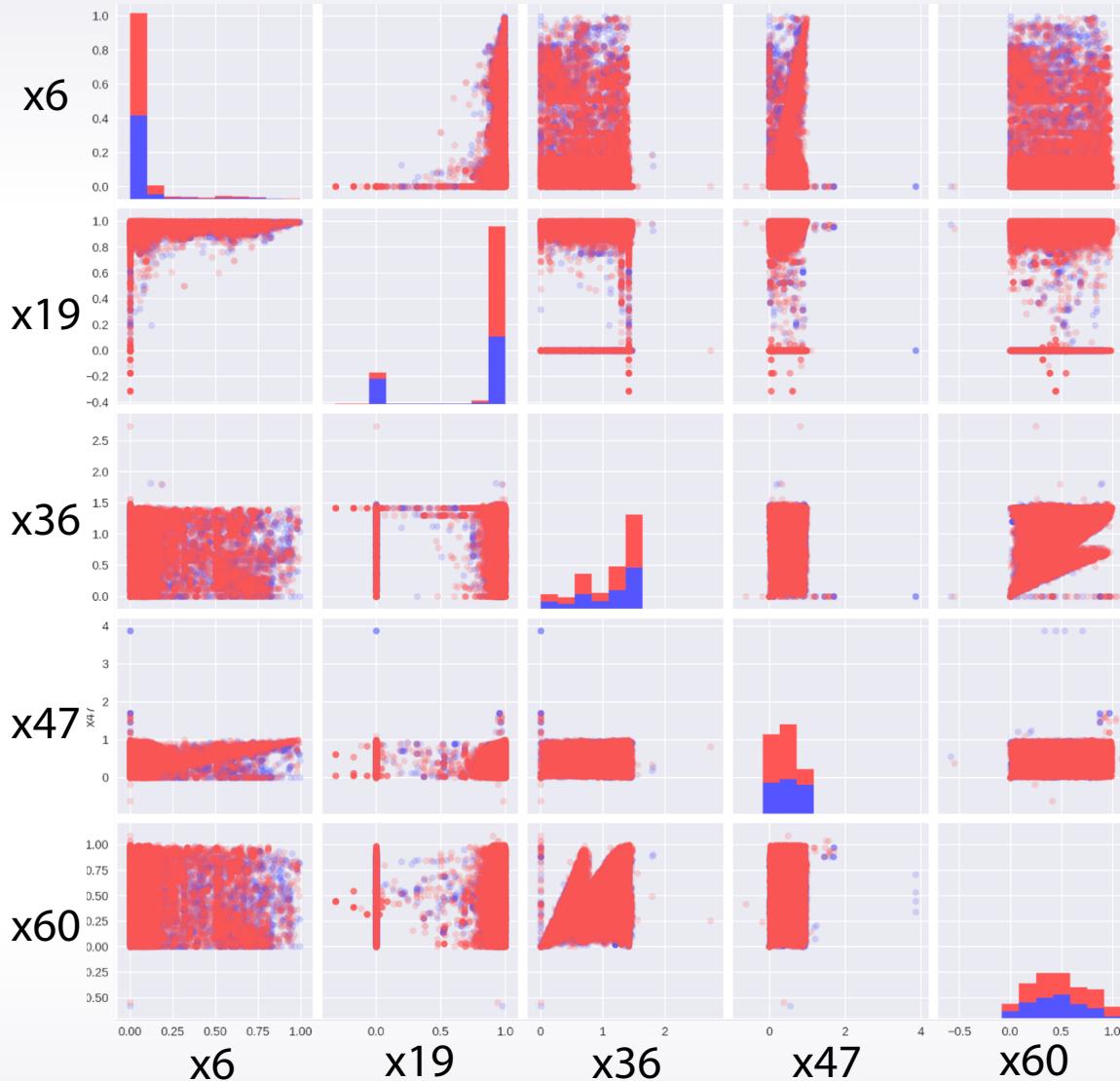
Exploring feature relations



Exploring feature relations

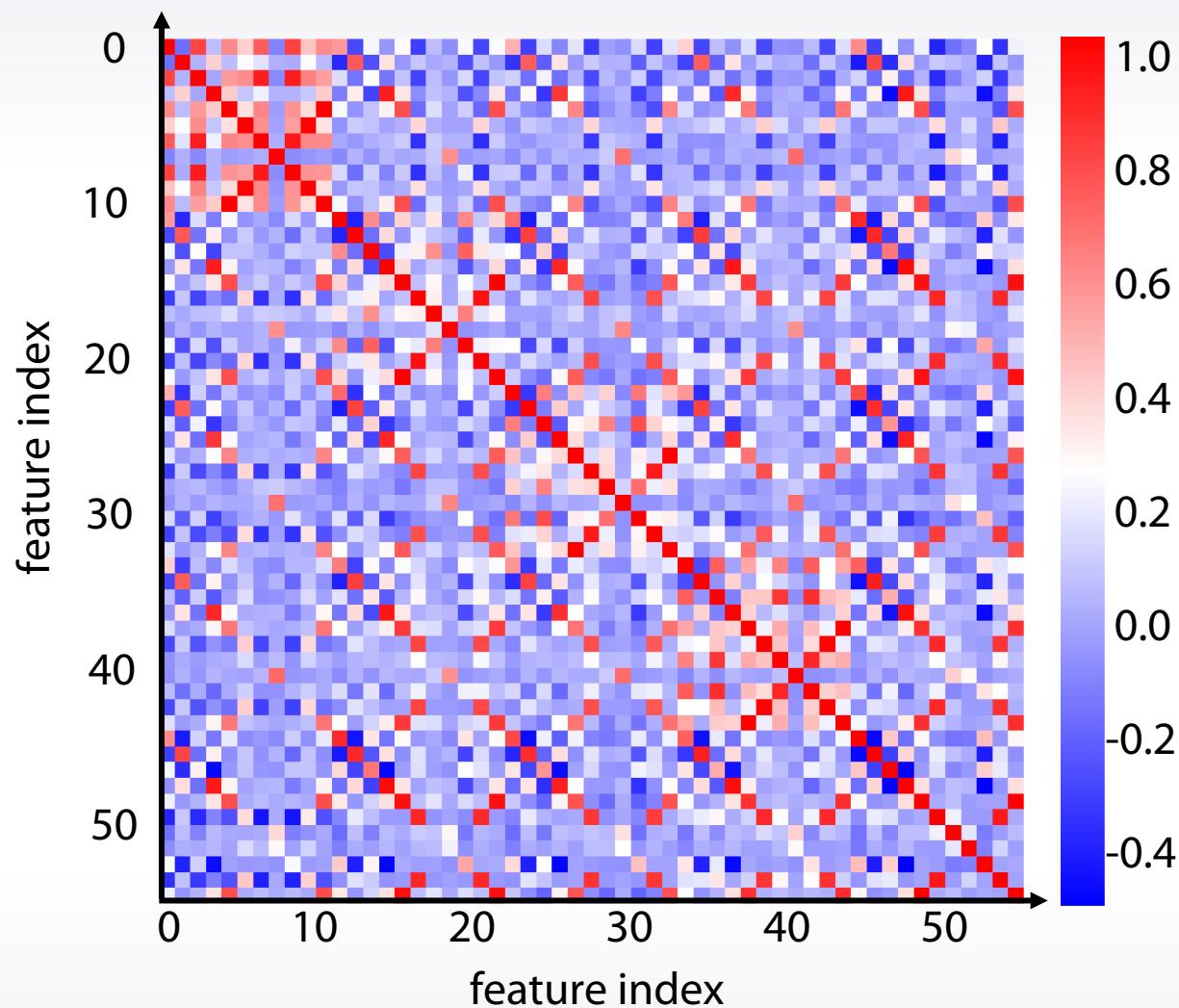


Exploring individual features: pairs



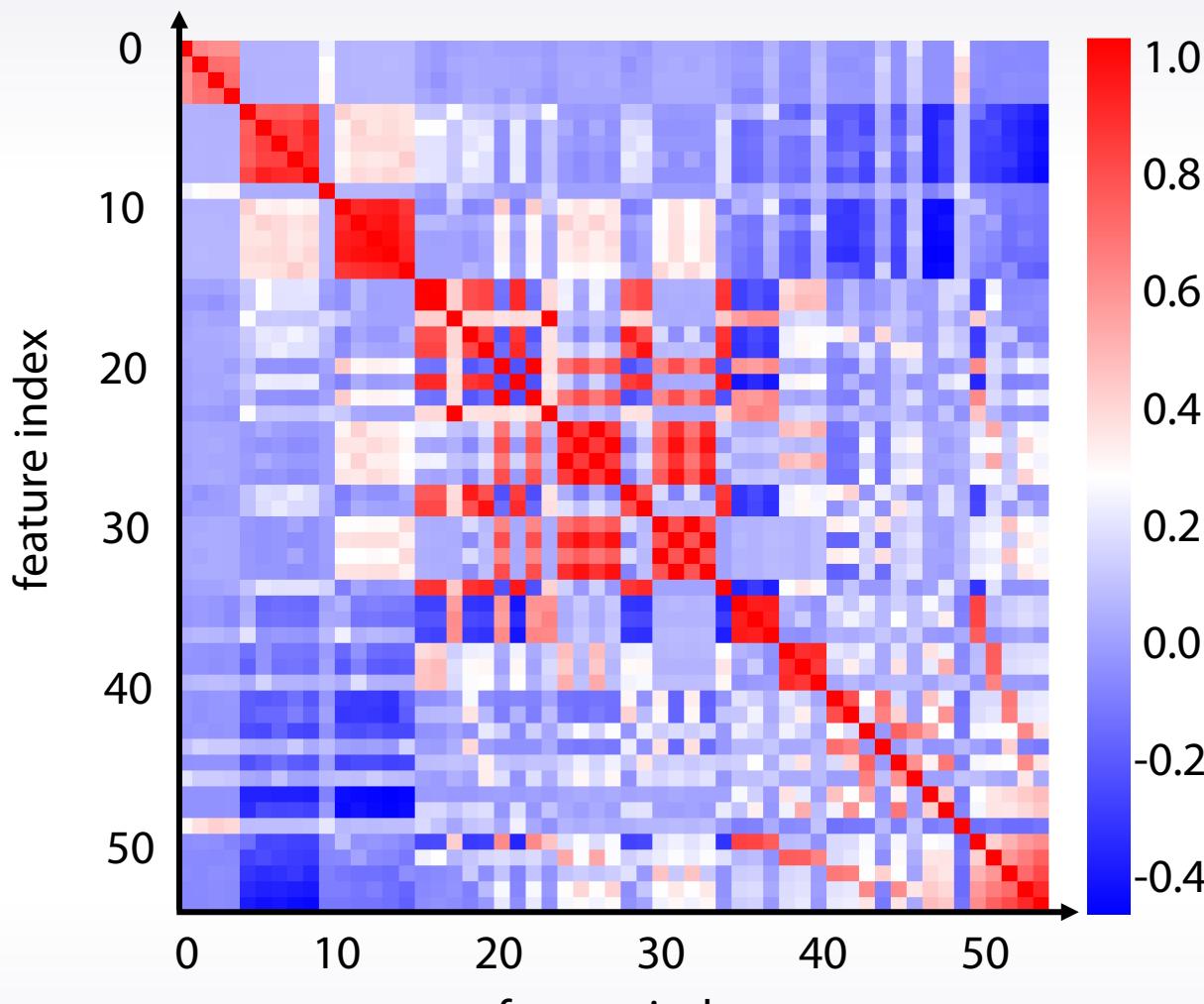
```
| pd.scatter_matrix(df)
```

Exploring individual features: pairs



```
| df.corr(), plt.matshow( ... )
```

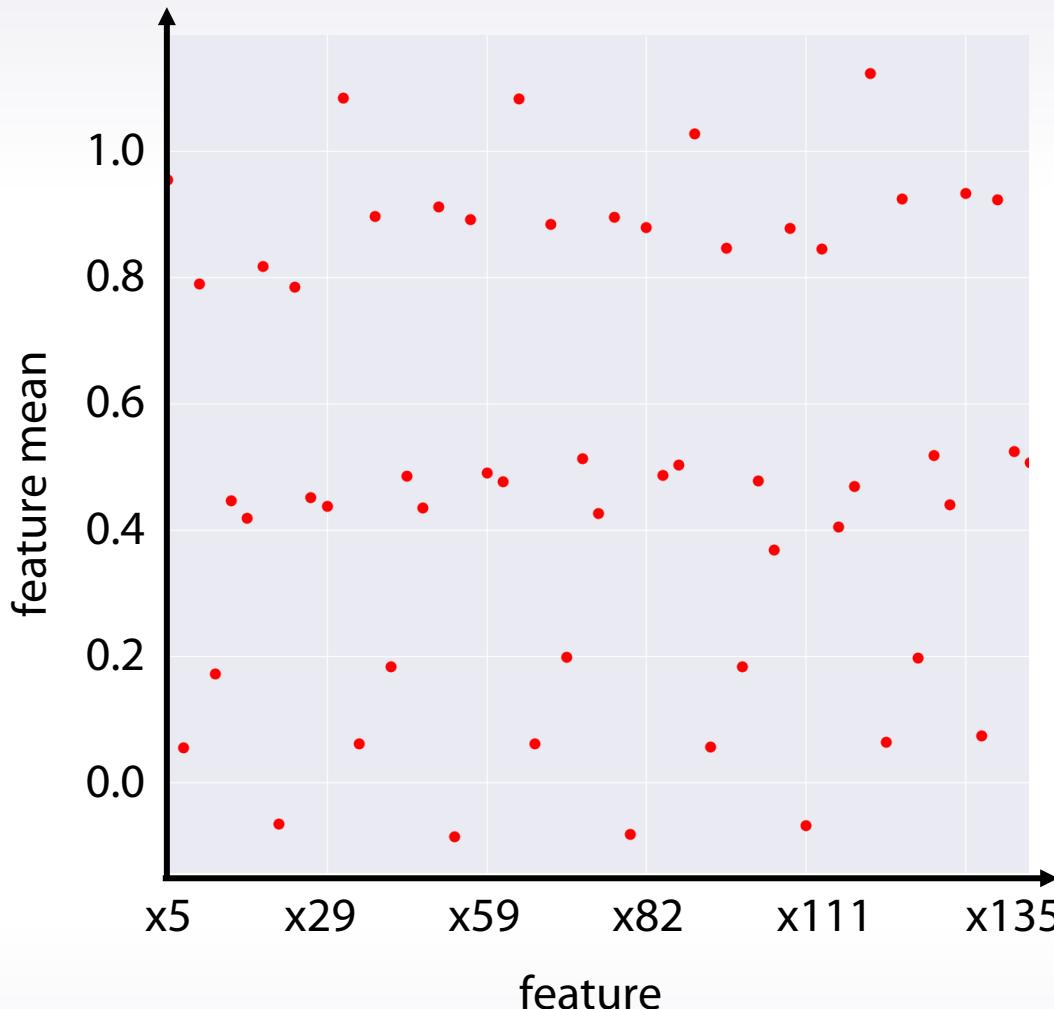
Exploring individual features: pairs/groups



Tools:

| `df.corr()`, `plt.matshow(...)`

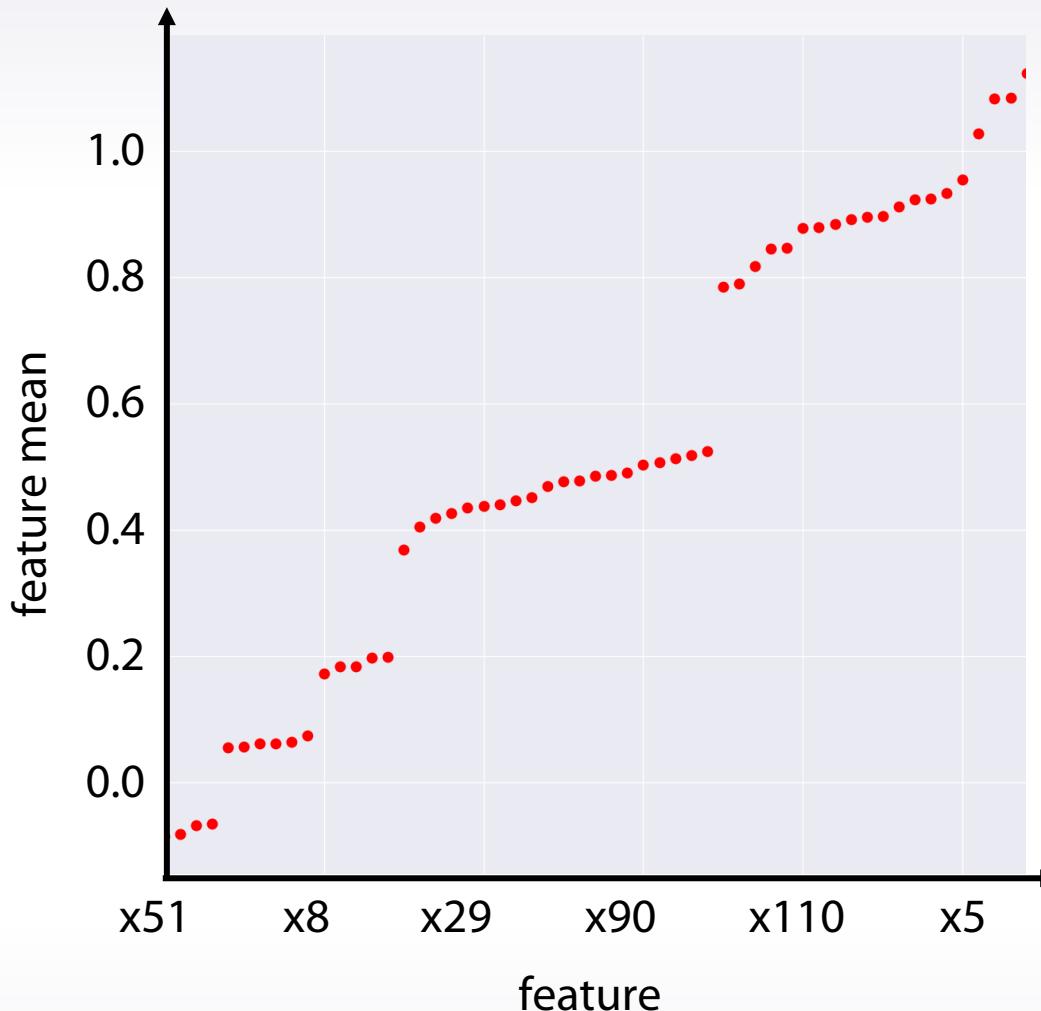
Exploring individual features: groups



Tools:

```
df.mean().plot(style='.')
```

Exploring individual features: groups



Tools:

```
df.mean().sort_values().plot(style='.')
```

Exploring individual features

Tools:

```
plt.scatter(x1, x2)
pd.scatter_matrix(df)
df.corr(), plt.matshow( ... )
df.mean().sort_values().plot(style='.' )
```

Conclusion

- Explore individual features
 - Histogram
 - Plot (index vs value)
 - Statistics
- Explore feature relations
 - **Pairs**
 - Scatter plot, scatter matrix
 - Corrplot
 - **Groups**
 - Corrplot + clustering
 - Plot (index vs feature statistics)