

Aula 5 – Mineração de Dados Weka

Profa. Elaine Faria
UFU - 2021

O que é o Weka?

- Ferramenta para Aprendizado de Máquina desenvolvida em Java
- Criado por pesquisadores da Universidade de Waikato – Nova Zelândia
- Implementa inúmeras funcionalidades
 - Classificadores
 - Métodos de agrupamento
 - Técnicas de pré-processamento

Weka

- Operações
 - Usando interface gráfica
 - Usando linha de comando
 - JAVA API
- Para download, acesse:
<http://www.cs.waikato.ac.nz/ml/weka/>
 - Você pode fazer o download da versão executável e dos códigos fontes

Weka

- Formato dos dados
 - Usa arquivos textos para descrever os dados
 - Trabalha com uma grande variedade de tipos de arquivos, incluindo .arff
 - É possível importar dados
 - CSV, por exemplo
 - Dados podem ser lidos de uma URL ou de uma base SQL

Formato dos dados

- .arff

```
@RELATION iris
```

```
@ATTRIBUTE sepallength NUMERIC
```

```
@ATTRIBUTE sepalwidth NUMERIC
```

```
@ATTRIBUTE petallength NUMERIC
```

```
@ATTRIBUTE petalwidth NUMERIC
```

```
@ATTRIBUTE class {Iris-setosa,Iris-versicolor, Iris-virginica}
```

```
@DATA
```

```
5.1,3.5,1.4,0.2,Iris-setosa
```

```
4.9,3.0,1.4,0.2,Iris-setosa
```

```
4.7,3.2,1.3,0.2,Iris-setosa
```

Formato dos dados

- **.arff**
 - @attribute attrName {numeric, string, <nominal>, date}
 - Numeric: número
 - Nominal: conjunto finito de strings
 - String: concatenação de caracteres
 - Date: data – formato yyyy-MM-dd

Formato dos dados

- .arff

```
@relation heart-disease-simplified
```

```
@attribute age numeric
```

```
@attribute sex { female, male}
```

```
@attribute chest_pain_type { typ_angina, asympt, non_anginal,  
    atyp_angina}
```

```
@attribute cholesterol numeric
```

```
@attribute exercise_induced_angina { no, yes}
```

```
@attribute class { present, not_present}
```

```
@data
```

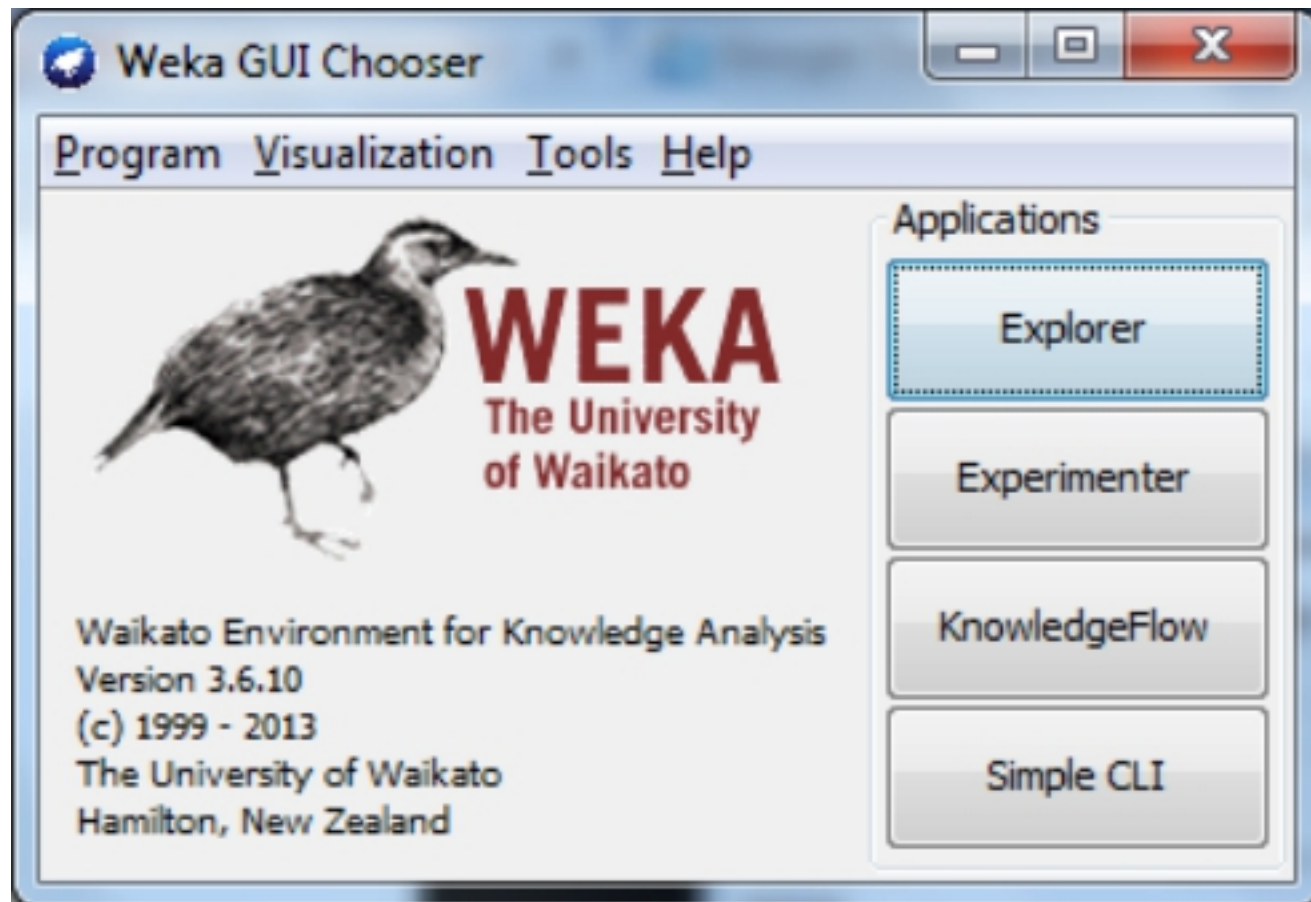
```
63,male,typ_angina,233,no,not_present
```

```
67,male,asympt,286,yes,present
```

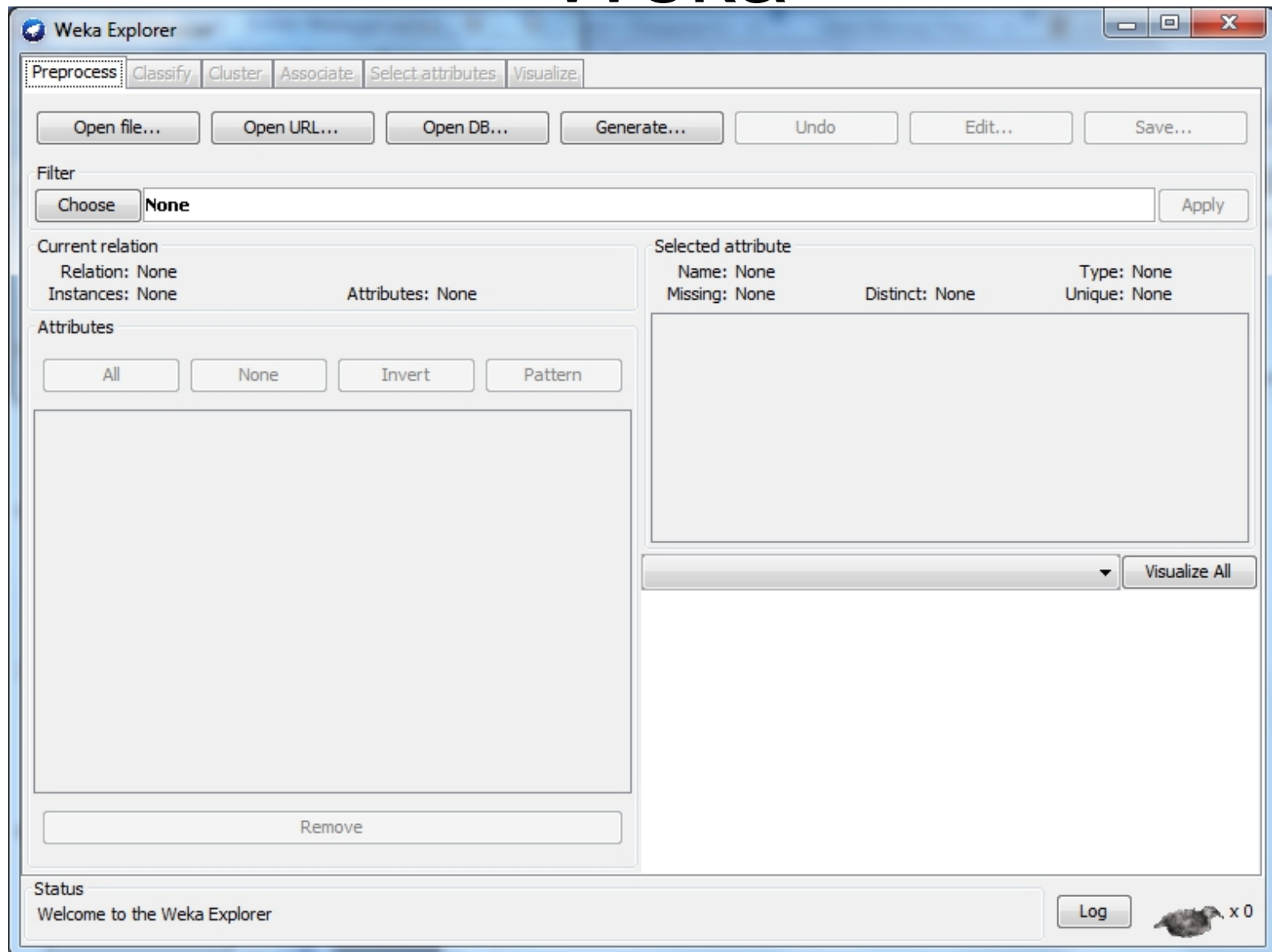
```
67,male,asympt,229,yes,present
```

```
38,female,non_anginal,?,no,not_present
```

Weka



Weka



Pré-processamento

- É chamado no Weka de filtro (*filters*)
- Weka contém filtros
 - Discretização, normalização, reamostragem, seleção de atributos, etc.

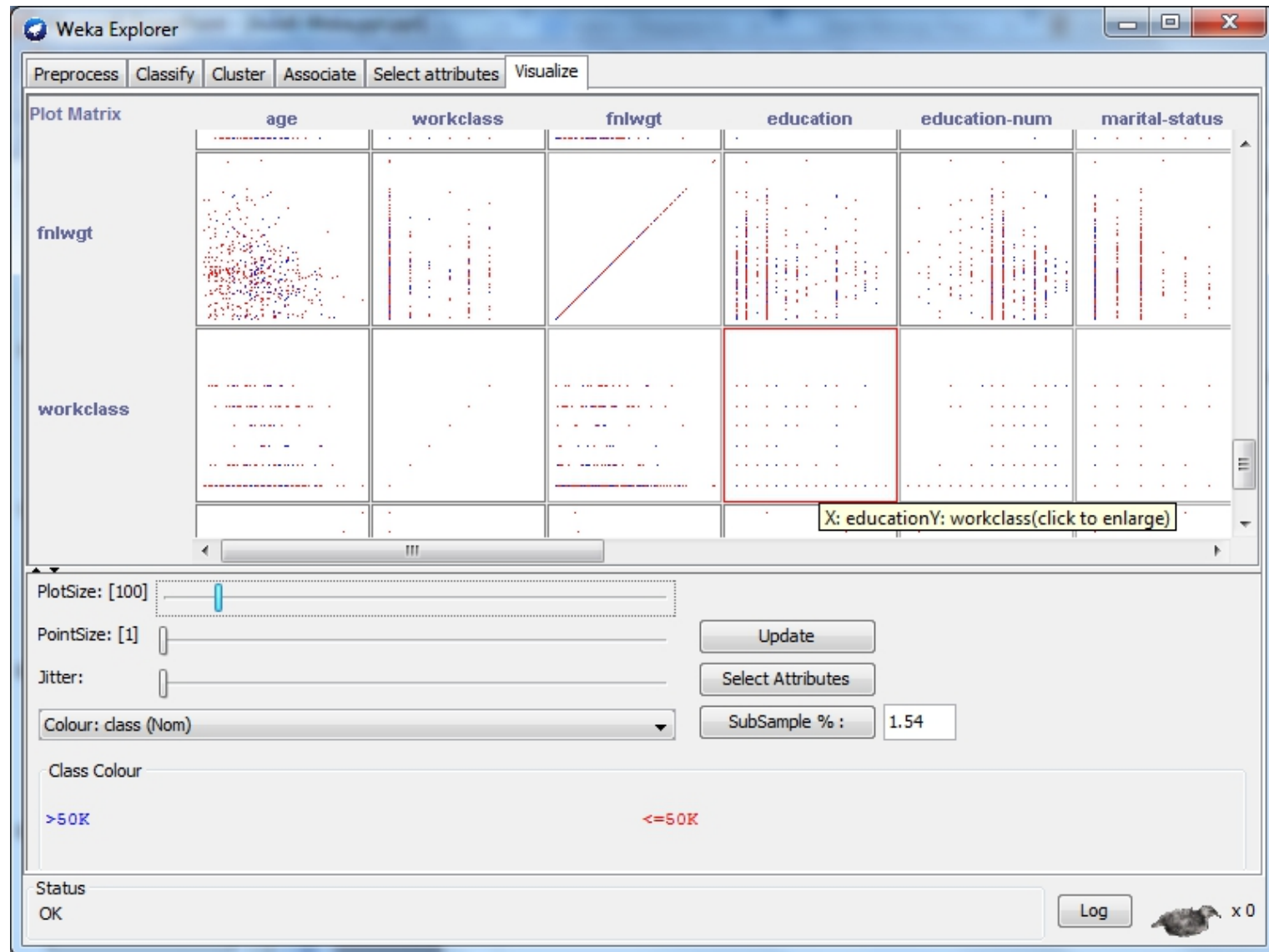
Vídeos sobre o Weka

- <https://weka.waikato.ac.nz/explorer>

Pré-processamento

- *Remove (attribute)*
- *RemoveWithValues (instance)*
- *Discretization (attribute)*
- *NominalToBinary (attribute)*
- *Normalize (attribute)*
- *ReplaceMissingValues*
- *ReplaceWithMissingValue*

Visualize



Referências

- www.cs.columbia.edu/~kathy/.../weka-tutorial.ppt
- www.cs.ubbcluj.ro/~gabis/ml/.../WekaTutorial.ppt
- www.cs.waikato.ac.nz/ml/weka/