National Institute Of Technology – Calicut

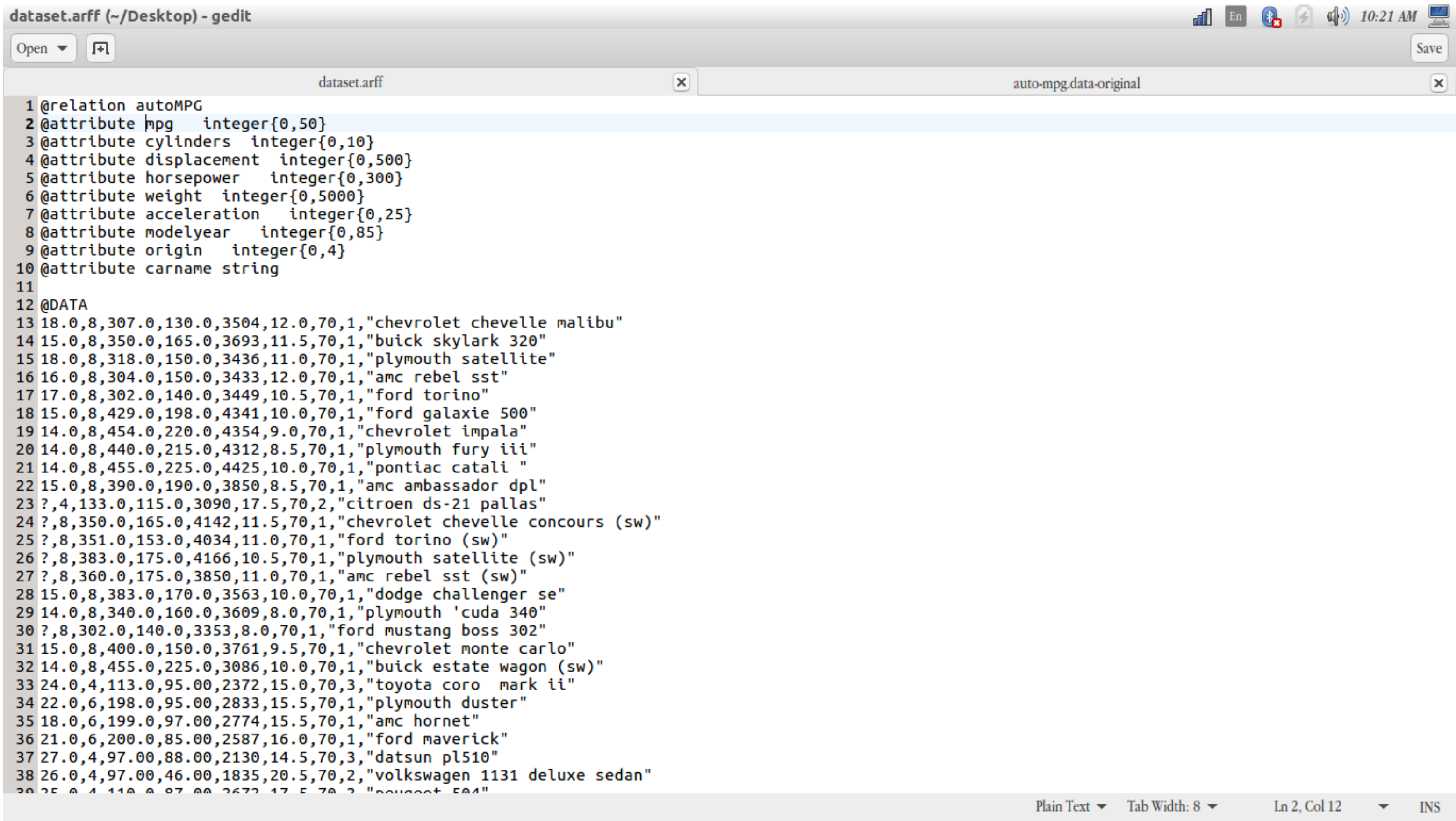
Data Mining

(Data pre-processing assignment)

Que 02.
(AUTO MPG data set)
Data Pre-processing Using Weka

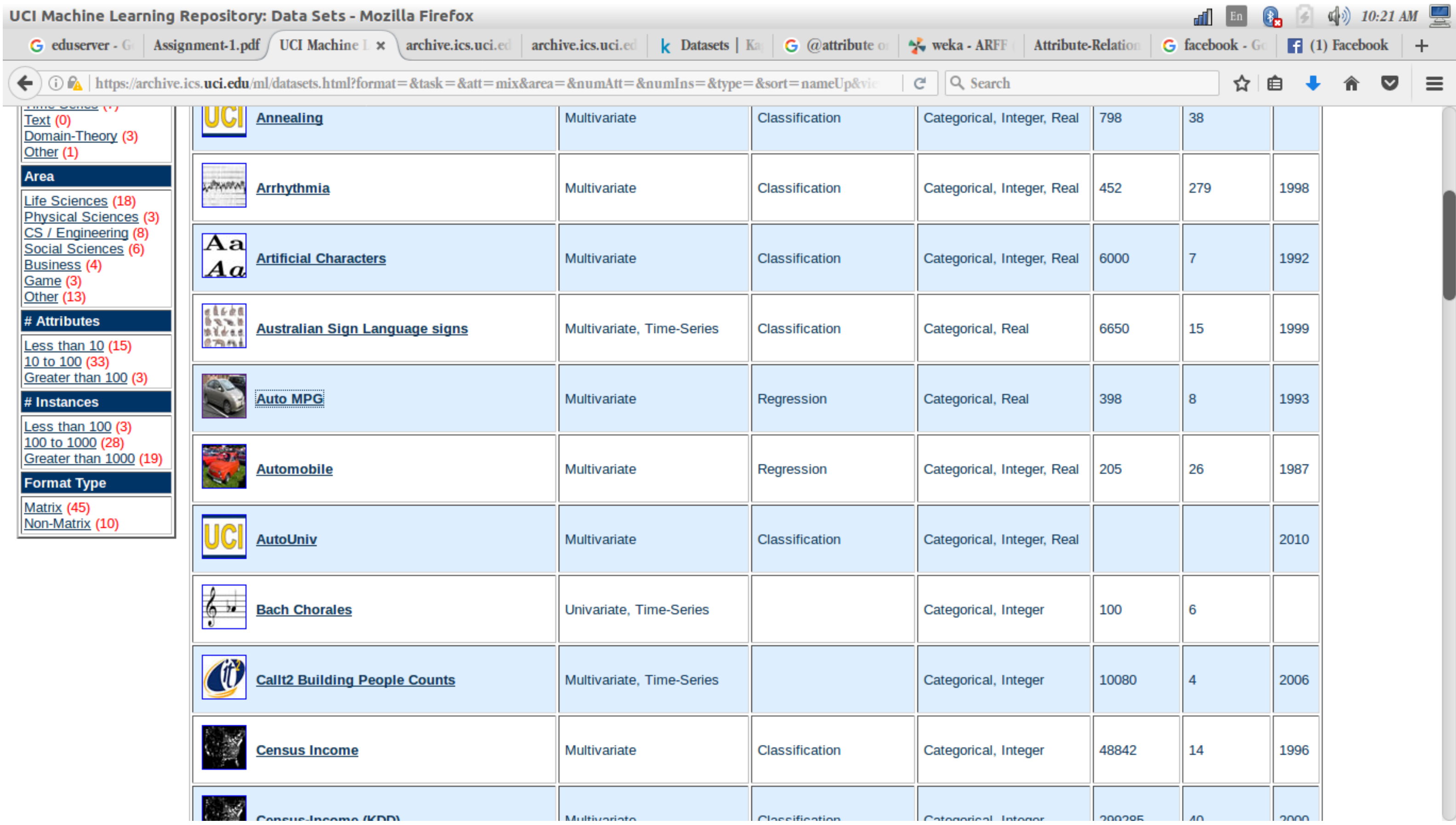
T.G. Deshan K. Sumanathilaka
B150413CS
Computer Science and Engineering
(B.Tech)

Selected Data Set ....



```
@relation autoMPG
@attribute mpg   integer{0,50}
@attribute cylinders  integer{0,10}
@attribute displacement  integer{0,500}
@attribute horsepower   integer{0,300}
@attribute weight  integer{0,5000}
@attribute acceleration   integer{0,25}
@attribute modelyear   integer{0,85}
@attribute origin   integer{0,4}
@attribute carname string

@DATA
18.0,8,307.0,130.0,3504,12.0,70,1,"chevrolet chevelle malibu"
15.0,8,350.0,165.0,3693,11.5,70,1,"buick skylark 320"
18.0,8,318.0,150.0,3436,11.0,70,1,"plymouth satellite"
16.0,8,304.0,150.0,3433,12.0,70,1,"amc rebel sst"
17.0,8,302.0,140.0,3449,10.5,70,1,"ford torino"
15.0,8,429.0,198.0,4341,10.0,70,1,"ford galaxie 500"
14.0,8,454.0,220.0,4354,9.0,70,1,"chevrolet impala"
14.0,8,440.0,215.0,4312,8.5,70,1,"plymouth fury iii"
14.0,8,455.0,225.0,4425,10.0,70,1,"pontiac catali "
15.0,8,390.0,190.0,3850,8.5,70,1,"amc ambassador dpl"
?,4,133.0,115.0,3090,17.5,70,2,"citroen ds-21 pallas"
?,8,350.0,165.0,4142,11.5,70,1,"chevrolet chevelle concours (sw)"
?,8,351.0,153.0,4034,11.0,70,1,"ford torino (sw)"
?,8,383.0,175.0,4166,10.5,70,1,"plymouth satellite (sw)"
?,8,360.0,175.0,3850,11.0,70,1,"amc rebel sst (sw)"
15.0,8,383.0,170.0,3563,10.0,70,1,"dodge challenger se"
14.0,8,340.0,160.0,3609,8.0,70,1,"plymouth 'cuda 340"
?,8,302.0,140.0,3353,8.0,70,1,"ford mustang boss 302"
15.0,8,400.0,150.0,3761,9.5,70,1,"chevrolet monte carlo"
14.0,8,455.0,225.0,3086,10.0,70,1,"buick estate wagon (sw)"
24.0,4,113.0,95.00,2372,15.0,70,3,"toyota coro  mark ii"
22.0,6,198.0,95.00,2833,15.5,70,1,"plymouth duster"
18.0,6,199.0,97.00,2774,15.5,70,1,"amc hornet"
21.0,6,200.0,85.00,2587,16.0,70,1,"ford maverick"
27.0,4,97.00,88.00,2130,14.5,70,3,"datsun pl510"
26.0,4,97.00,46.00,1835,20.5,70,2,"volkswagen 1131 deluxe sedan"
25.0,4,110.0,87.00,2672,17.5,70,2,"peugeot 504"
```


UCI Machine Learning Repository: Data Sets - Mozilla Firefox
10:21 AM
eduserver - G
Assignment-1.pdf
UCI Machine L
archive.ics.uci.ed
archive.ics.uci.ed
Datasets | Ka
@attribute o
weka - ARFF
Attribute-Relation
facebook - G
(1) Facebook
https://archive.ics.uci.edu/ml/datasets.html?format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&vie
Search

Text (0)
Domain-Theory (3)
Other (1)

Area

Life Sciences (18)
Physical Sciences (3)
CS / Engineering (8)
Social Sciences (6)
Business (4)
Game (3)
Other (13)

# Attributes

Less than 10 (15)
10 to 100 (33)
Greater than 100 (3)

# Instances

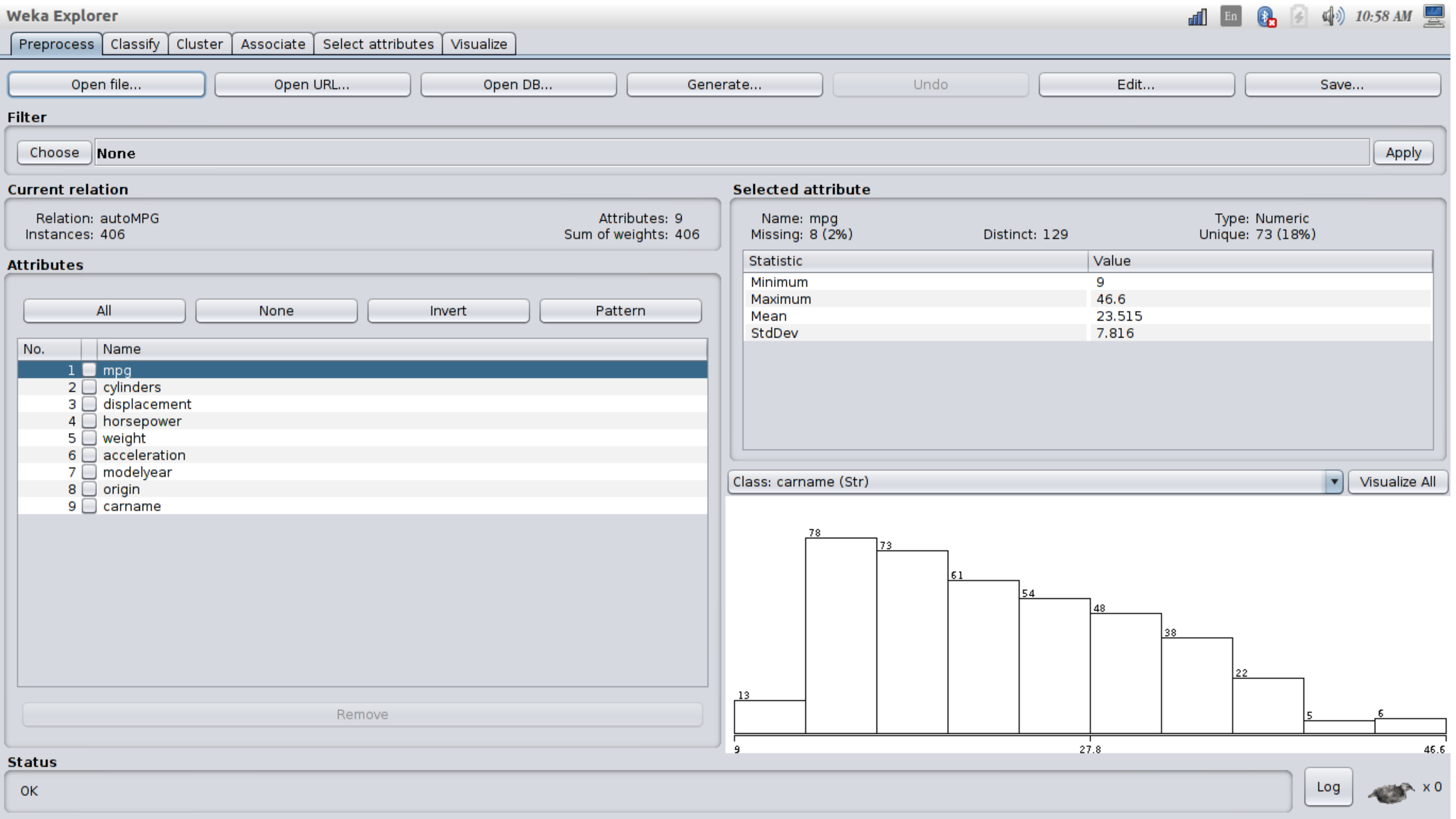
Less than 100 (3)
100 to 1000 (28)
Greater than 1000 (19)

Format Type

Matrix (45)
Non-Matrix (10)

Annealing	Multivariate	Classification	Categorical, Integer, Real	798	38	
Arrhythmia	Multivariate	Classification	Categorical, Integer, Real	452	279	1998
Artificial Characters	Multivariate	Classification	Categorical, Integer, Real	6000	7	1992
Australian Sign Language signs	Multivariate, Time-Series	Classification	Categorical, Real	6650	15	1999
Auto MPG	Multivariate	Regression	Categorical, Real	398	8	1993
Automobile	Multivariate	Regression	Categorical, Integer, Real	205	26	1987
AutoUniv	Multivariate	Classification	Categorical, Integer, Real			2010
Bach Chorales	Univariate, Time-Series		Categorical, Integer	100	6	
CalIt2 Building People Counts	Multivariate, Time-Series		Categorical, Integer	10080	4	2006
Census Income	Multivariate	Classification	Categorical, Integer	48842	14	1996
Census-Income (KDD)	Multivariate	Classification	Categorical, Integer	299285	40	2000

Step 1 :Loading Data Set to Weka



Filter

None

Current relation

Relation: autoMPG
Instances: 406
Attributes: 9
Sum of weights: 406

Attributes

No.	Name
1	□ mpg
2	□ cylinders
3	□ displacement
4	□ horsepower
5	□ weight
6	□ acceleration
7	□ modelyear
8	□ origin
9	□ carname

Selected attribute

Name: mpg
Missing: 8 (2%)
Distinct: 129
Type: Numeric
Unique: 73 (18%)

Statistic	Value
Minimum	9
Maximum	46.6
Mean	23.515
StdDev	7.816

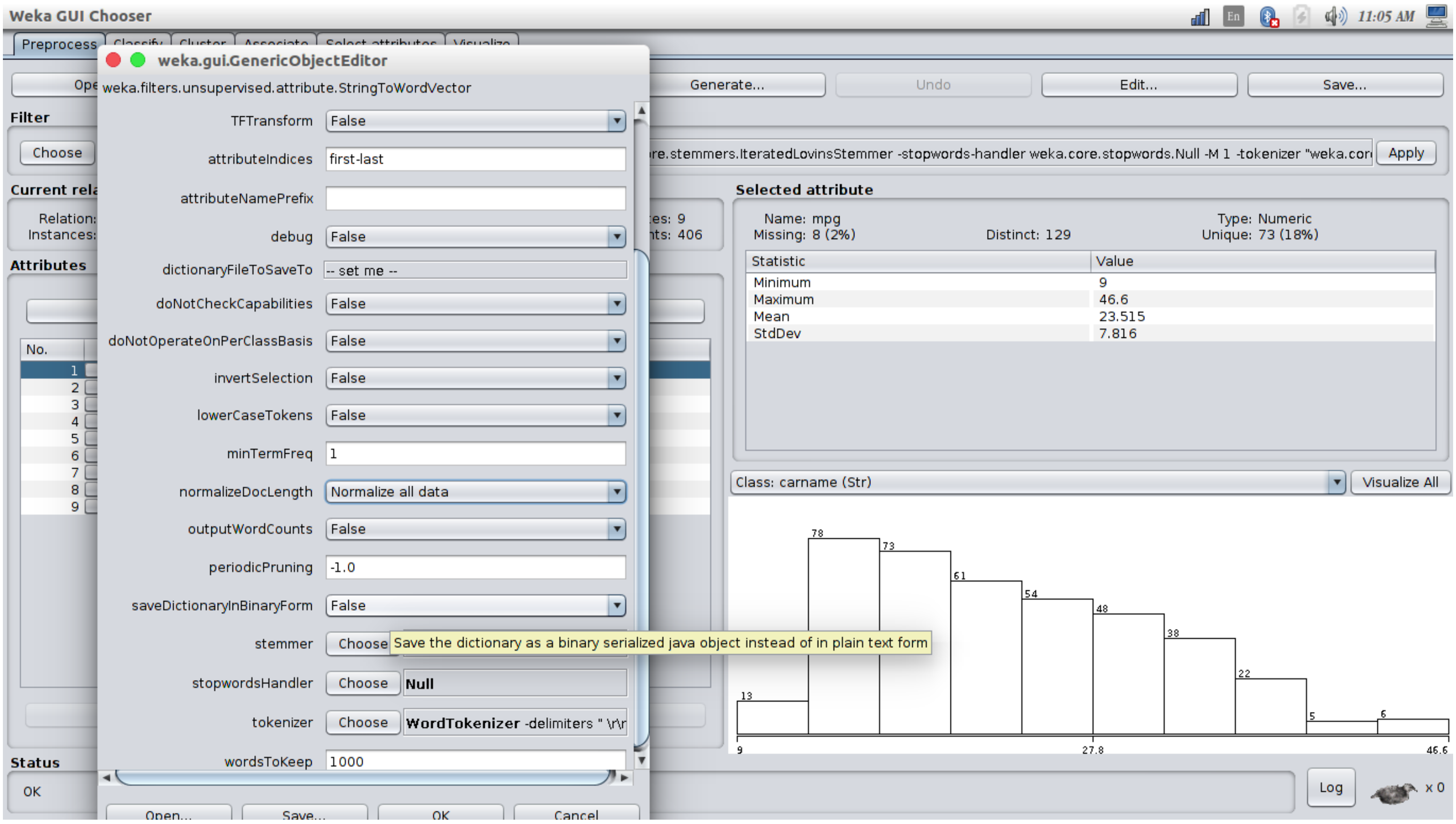
Class: carname (Str)

Status

OK

Step 2

a)Cleaning up inconsistent spelling of terms



weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.StringToWordVector

TFTransform	False
attributeIndices	first-last
attributeNamePrefix	
debug	False
dictionaryFileToSaveTo	-- set me --
doNotCheckCapabilities	False
doNotOperateOnPerClassBasis	False
invertSelection	False
lowerCaseTokens	False
minTermFreq	1
normalizeDocLength	Normalize all data
outputWordCounts	False
periodicPruning	-1.0
saveDictionaryInBinaryForm	False
stemmer	Choose
stopwordsHandler	Choose Null
tokenizer	Choose WordTokenizer -delimiters " \r\r
wordsToKeep	1000

Save the dictionary as a binary serialized java object instead of in plain text form

Open... Save... OK Cancel

re.stemmers.IteratedLovinsStemmer -stopwords-handler weka.core.stopwords.Null -M 1 -tokenizer "weka.cor

es: 9
ts: 406

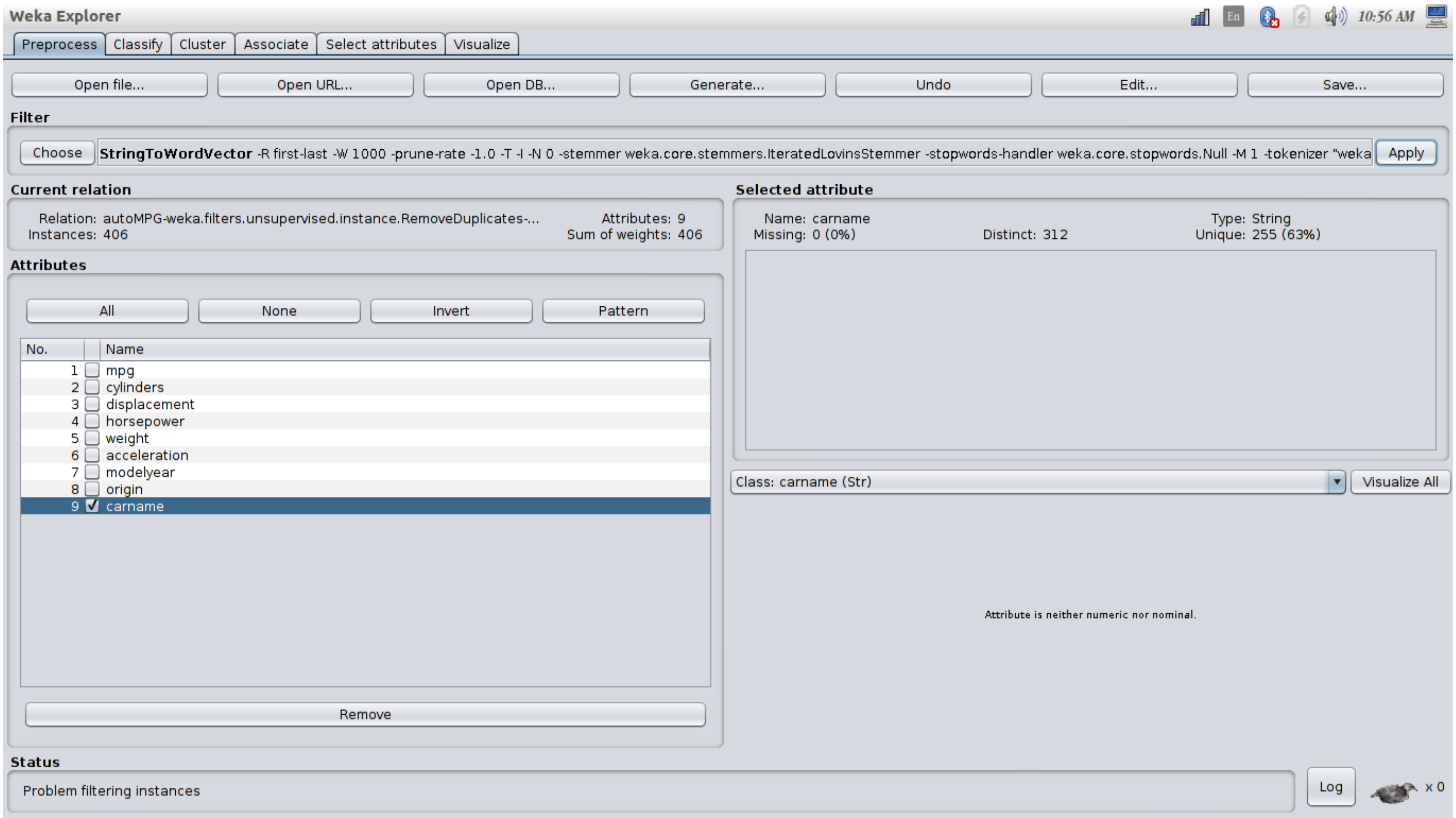
Selected attribute

Name: mpg
Missing: 8 (2%)
Distinct: 129
Type: Numeric
Unique: 73 (18%)

Statistic	Value
Minimum	9
Maximum	46.6
Mean	23.515
StdDev	7.816

Class: carname (Str)

Visualize All



Filter

StringToWordVector -R first-last -W 1000 -prune-rate -1.0 -T -I -N 0 -stemmer weka.core.stemmers.IteratedLovinsStemmer -stopwords-handler weka.core.stopwords.Null -M 1 -tokenizer "weka

Current relation

Relation: autoMPG-weka.filters.unsupervised.instance.RemoveDuplicates-...
Instances: 406
Attributes: 9
Sum of weights: 406

Attributes

No.		Name
1	☐	mpg
2	☐	cylinders
3	☐	displacement
4	☐	horsepower
5	☐	weight
6	☐	acceleration
7	☐	modelyear
8	☐	origin
9	☑	carname

Selected attribute

Name: carname
Missing: 0 (0%)
Distinct: 312
Type: String
Unique: 255 (63%)

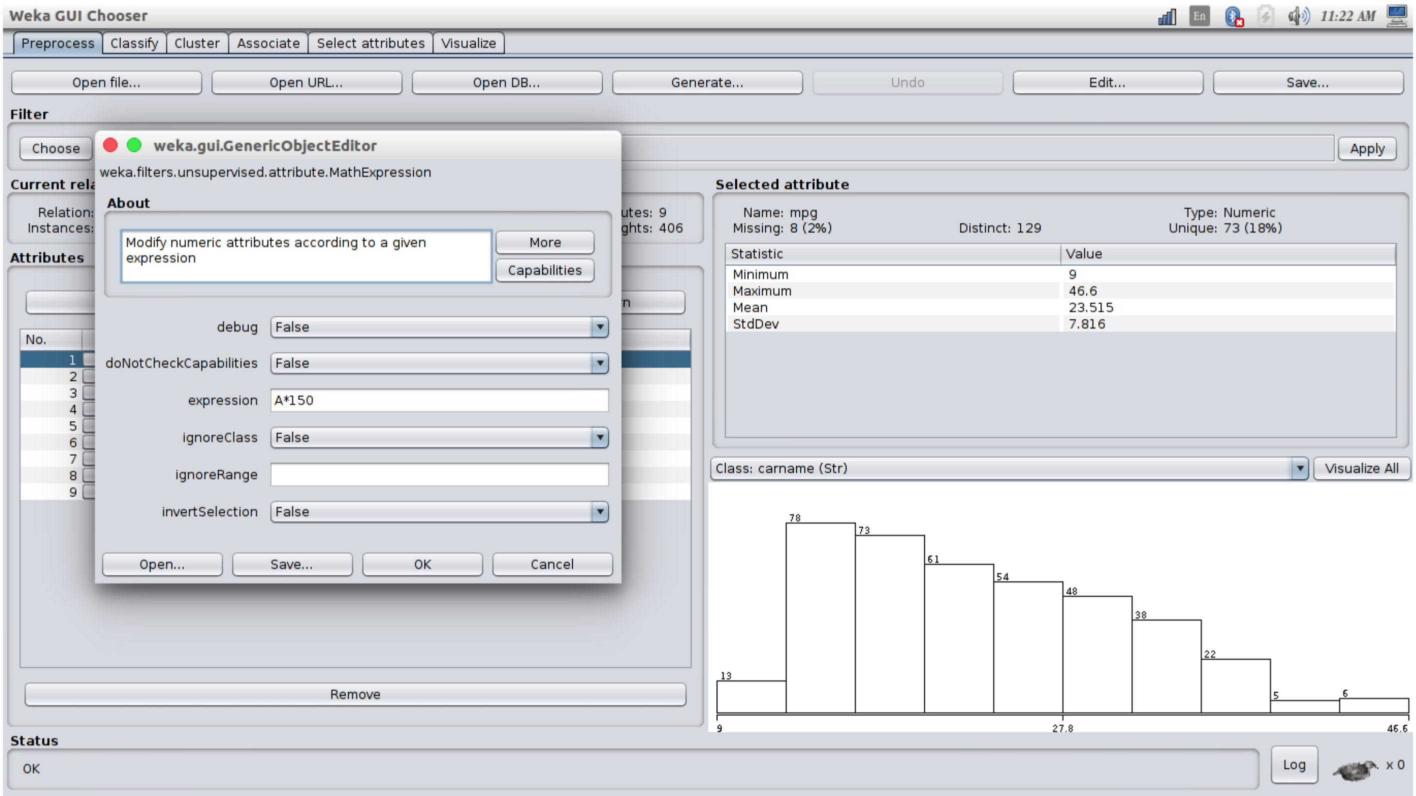
Class: carname (Str)

Attribute is neither numeric nor nominal.

Status

Problem filtering instances

b)Converting values that are text descriptions of numeric values to actual numeric values which are usable for analysis.



weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.MathExpression

About

Modify numeric attributes according to a given expression

debug: False

doNotCheckCapabilities: False

expression: A*150

ignoreClass: False

ignoreRange:

invertSelection: False

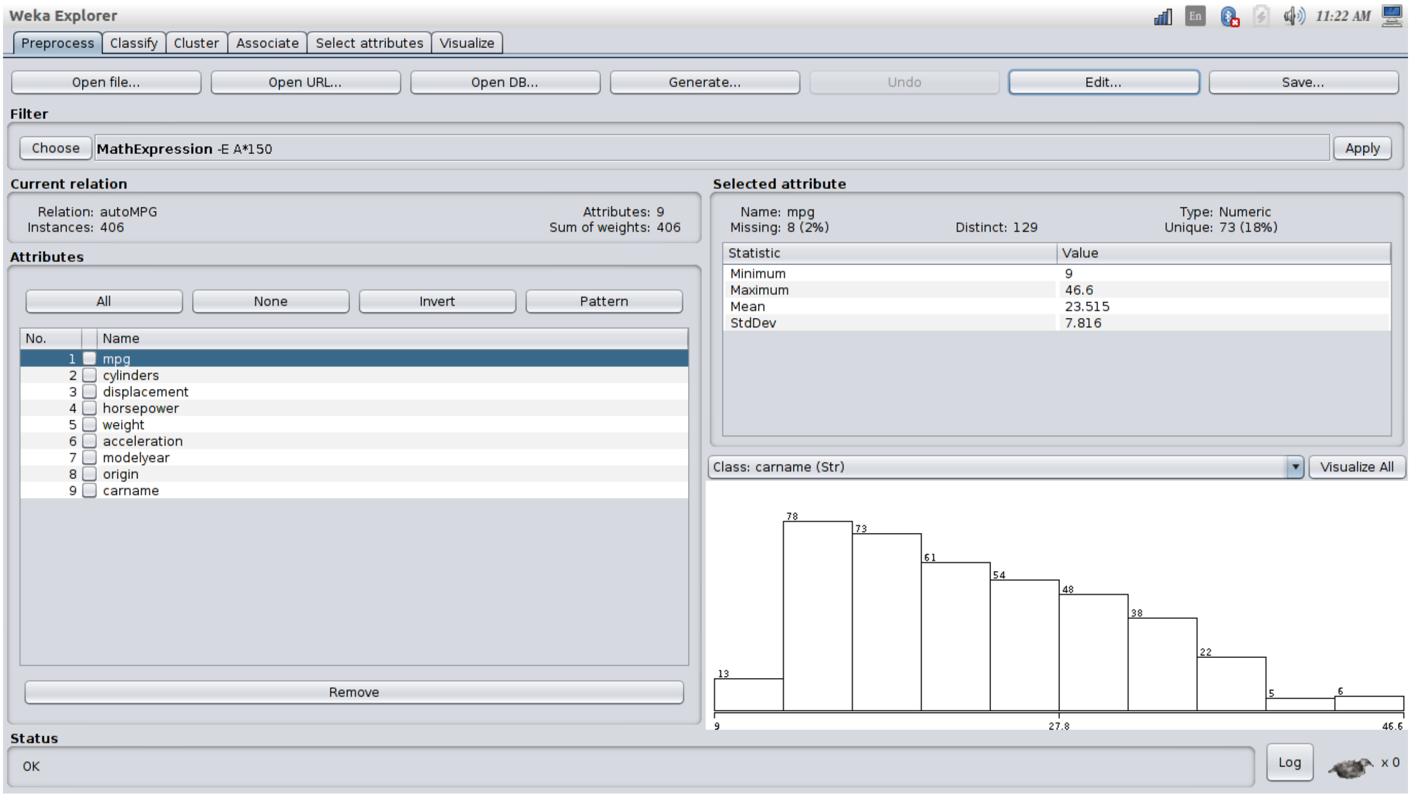
Selected attribute

Name: mpg
Missing: 8 (2%)
Distinct: 129
Type: Numeric
Unique: 73 (18%)

Statistic	Value
Minimum	9
Maximum	46.6
Mean	23.515
StdDev	7.816

Class: carname (Str)

OK



MathExpression -E A*150

Current relation

Relation: autoMPG
Instances: 406
Attributes: 9
Sum of weights: 406

Attributes

No.	Name
1	mpg
2	cylinders
3	displacement
4	horsepower
5	weight
6	acceleration
7	modelyear
8	origin
9	carname

Selected attribute

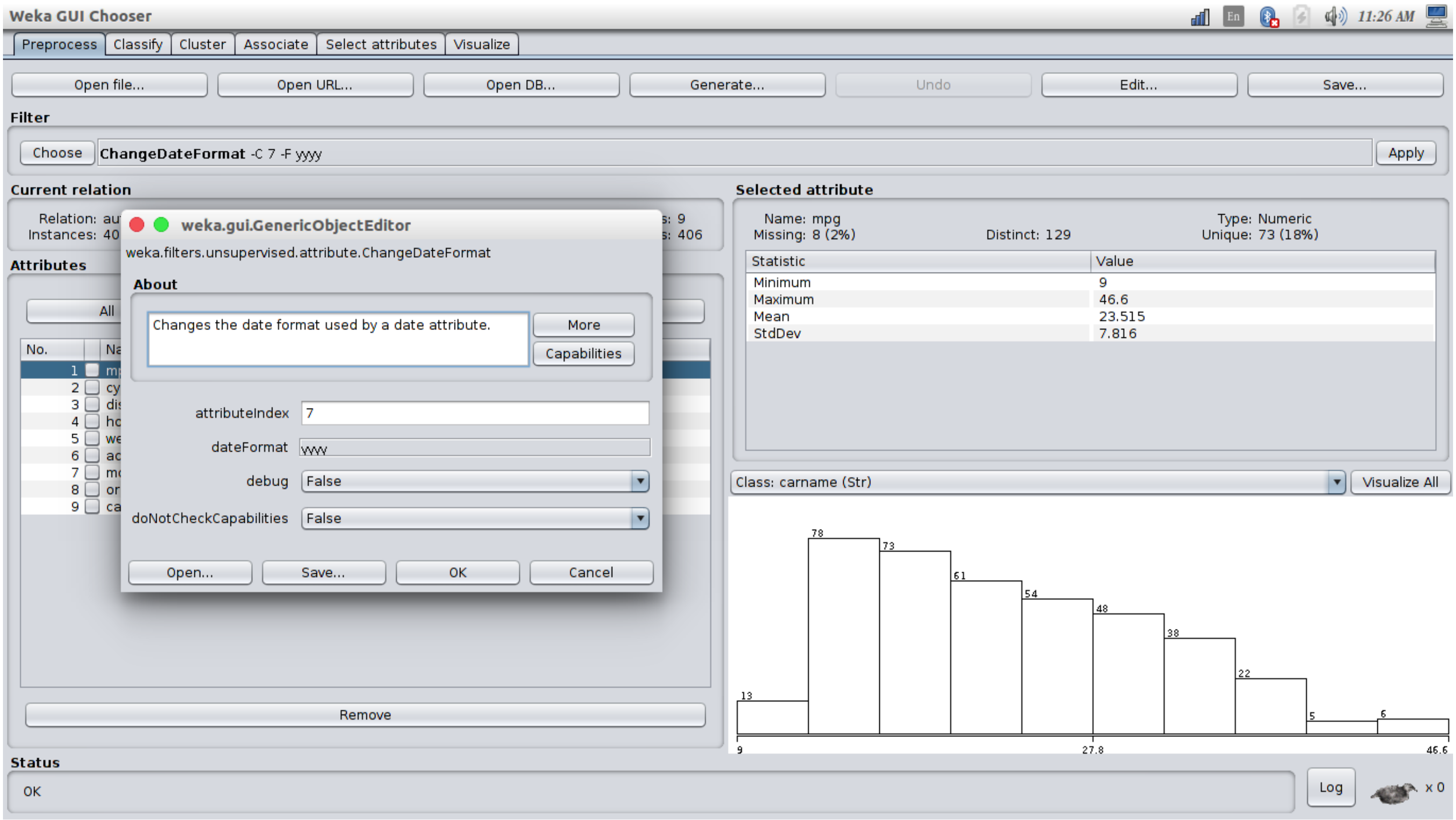
Name: mpg
Missing: 8 (2%)
Distinct: 129
Type: Numeric
Unique: 73 (18%)

Statistic	Value
Minimum	9
Maximum	46.6
Mean	23.515
StdDev	7.816

Class: carname (Str)

OK

c)Extracting and cleaning values for dates



Filter: **ChangeDateFormat** -C 7 -F yyyy

Current relation

Relation: au… Instances: 40…

Attributes

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.ChangeDateFormat

About

Changes the date format used by a date attribute.

attributeIndex: 7

dateFormat: yyyy

debug: False

doNotCheckCapabilities: False

Selected attribute

Name: mpg Type: Numeric

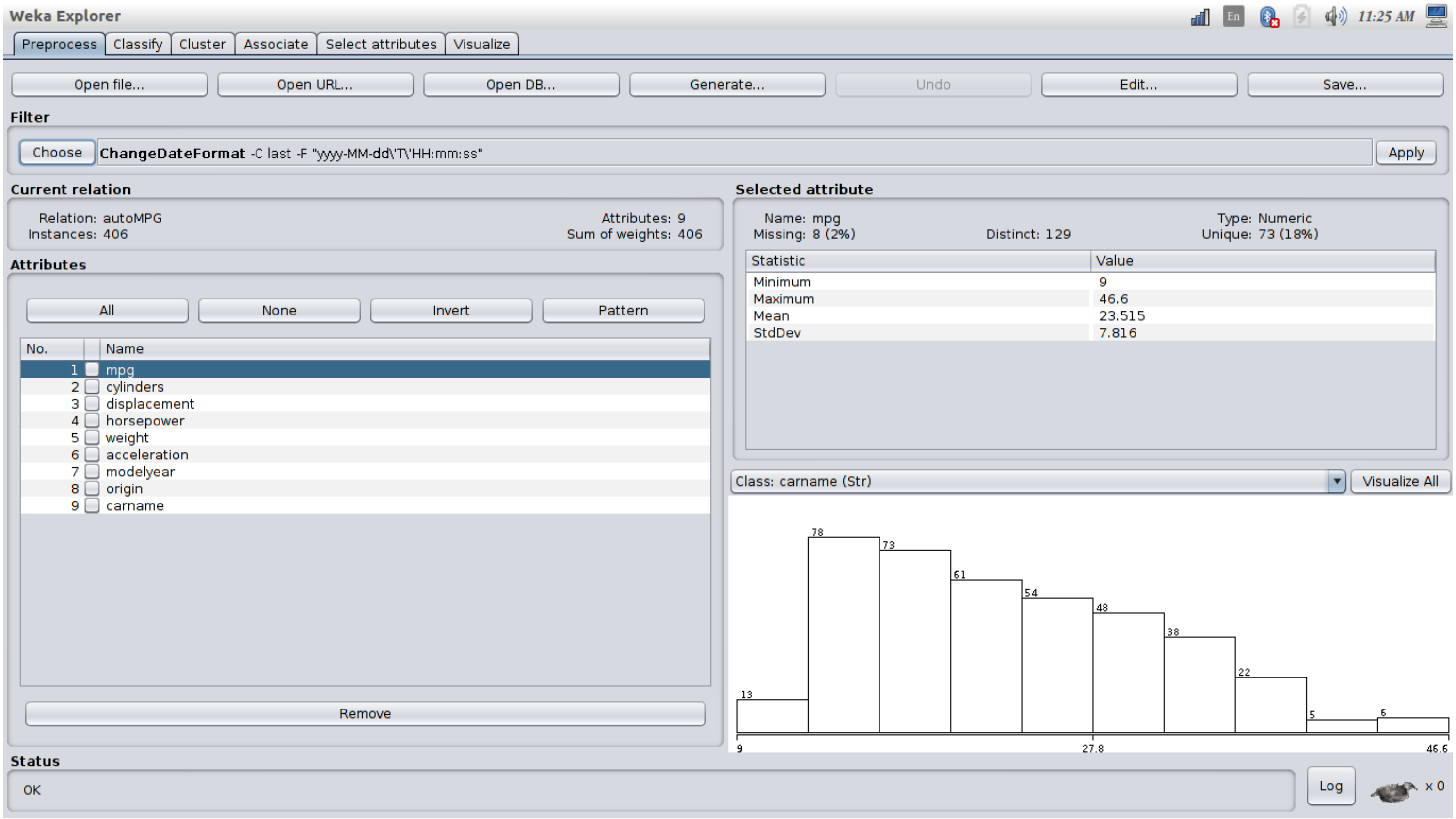
Missing: 8 (2%) Distinct: 129 Unique: 73 (18%)

Statistic	Value
Minimum	9
Maximum	46.6
Mean	23.515
StdDev	7.816

Class: carname (Str)

Status

OK



Filter: **ChangeDateFormat** -C last -F "yyyy-MM-dd\'T\'HH:mm:ss"

Current relation

Relation: autoMPG Attributes: 9

Instances: 406 Sum of weights: 406

Attributes

No.	Name
1	mpg
2	cylinders
3	displacement
4	horsepower
5	weight
6	acceleration
7	modelyear
8	origin
9	carname

Selected attribute

Name: mpg Type: Numeric

Missing: 8 (2%) Distinct: 129 Unique: 73 (18%)

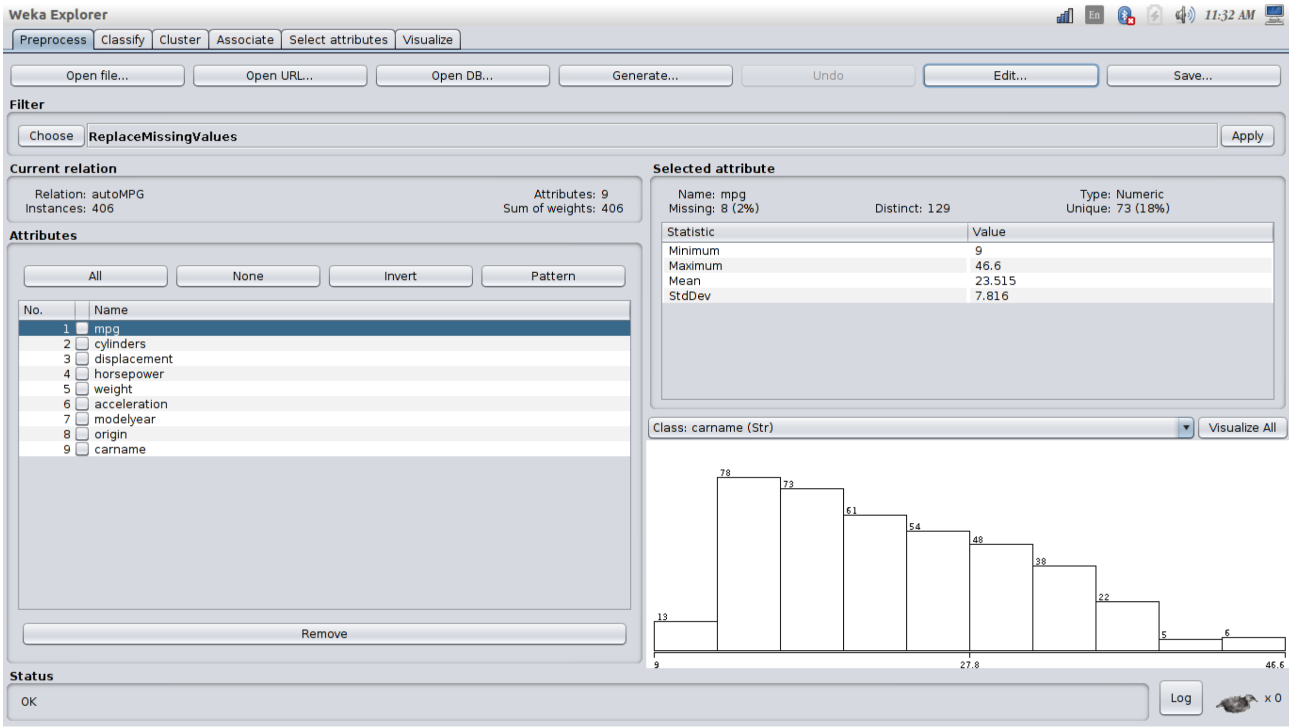
Statistic	Value
Minimum	9
Maximum	46.6
Mean	23.515
StdDev	7.816

Class: carname (Str)

Status

OK

d)Fill in the missing values by the various options



Filter: ReplaceMissingValues

Current relation

Relation: autoMPG   Attributes: 9
Instances: 406   Sum of weights: 406

Selected attribute

Name: mpg   Type: Numeric
Missing: 8 (2%)   Distinct: 129   Unique: 73 (18%)

Statistic	Value
Minimum	9
Maximum	46.6
Mean	23.515
StdDev	7.816

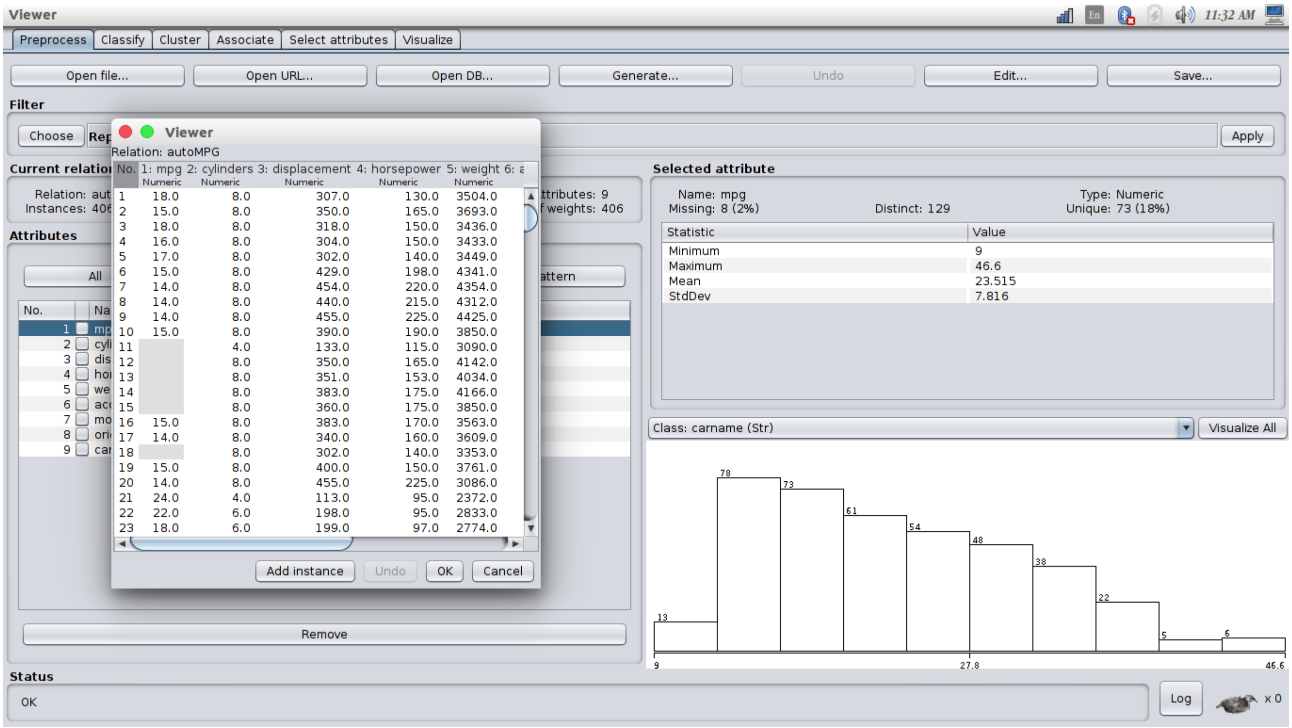
Attributes

No.	Name
1	mpg
2	cylinders
3	displacement
4	horsepower
5	weight
6	acceleration
7	modelyear
8	origin
9	carname

Class: carname (Str)

Status

OK



Relation: autoMPG

No.	1: mpg Numeric	2: cylinders Numeric	3: displacement Numeric	4: horsepower Numeric	5: weight Numeric
1	18.0	8.0	307.0	130.0	3504.0
2	15.0	8.0	350.0	165.0	3693.0
3	18.0	8.0	318.0	150.0	3436.0
4	16.0	8.0	304.0	150.0	3433.0
5	17.0	8.0	302.0	140.0	3449.0
6	15.0	8.0	429.0	198.0	4341.0
7	14.0	8.0	454.0	220.0	4354.0
8	14.0	8.0	440.0	215.0	4312.0
9	14.0	8.0	455.0	225.0	4425.0
10	15.0	8.0	390.0	190.0	3850.0
11		4.0	133.0	115.0	3090.0
12		8.0	350.0	165.0	4142.0
13		8.0	351.0	153.0	4034.0
14		8.0	383.0	175.0	4166.0
15		8.0	360.0	175.0	3850.0
16	15.0	8.0	383.0	170.0	3563.0
17	14.0	8.0	340.0	160.0	3609.0
18		8.0	302.0	140.0	3353.0
19	15.0	8.0	400.0	150.0	3761.0
20	14.0	8.0	455.0	225.0	3086.0
21	24.0	4.0	113.0	95.0	2372.0
22	22.0	6.0	198.0	95.0	2833.0
23	18.0	6.0	199.0	97.0	2774.0

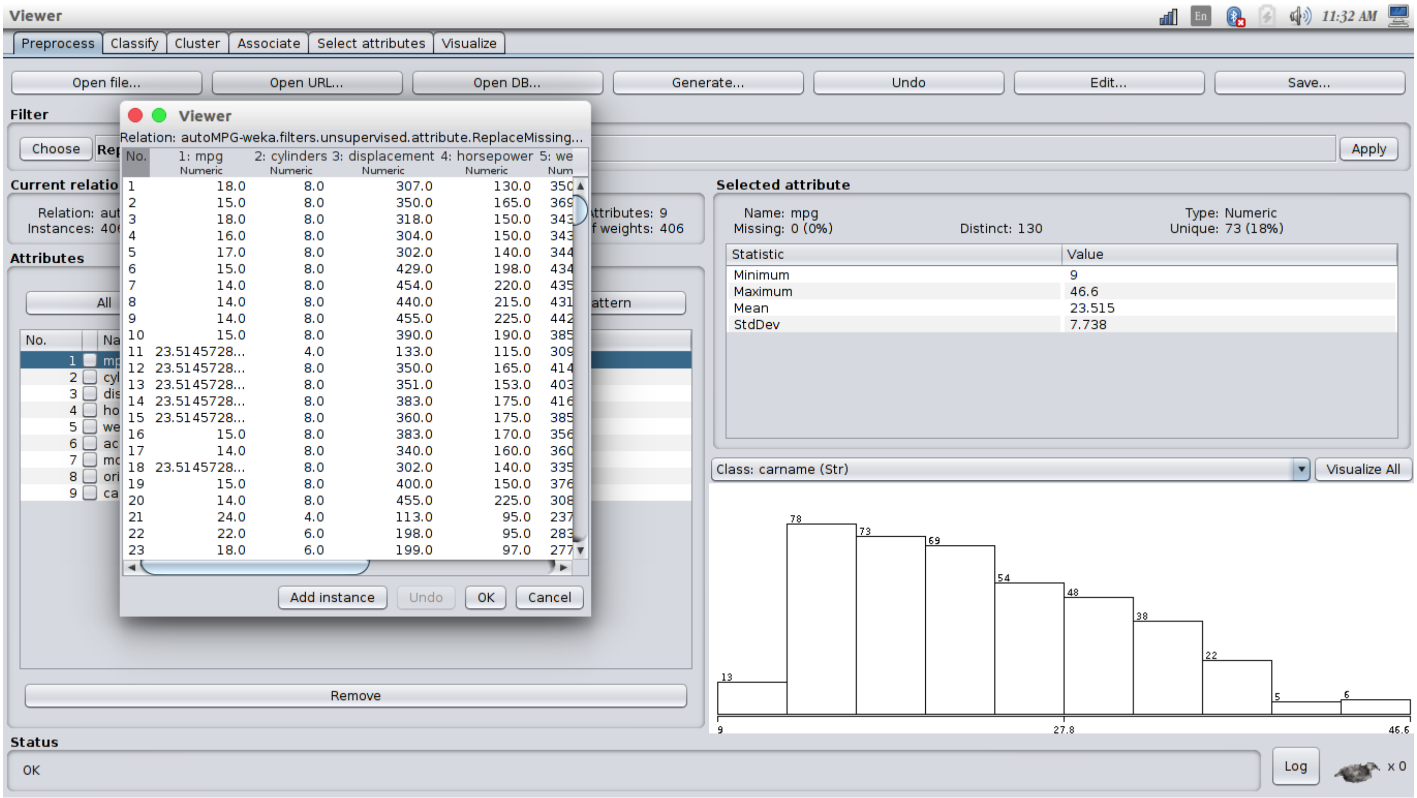
Name: mpg   Type: Numeric
Missing: 8 (2%)   Distinct: 129   Unique: 73 (18%)

Statistic	Value
Minimum	9
Maximum	46.6
Mean	23.515
StdDev	7.816

Class: carname (Str)

Status

OK



Relation: autoMPG-weka.filters.unsupervised.attribute.ReplaceMissing...

No.	1: mpg Numeric	2: cylinders Numeric	3: displacement Numeric	4: horsepower Numeric	5: we Num
1	18.0	8.0	307.0	130.0	350
2	15.0	8.0	350.0	165.0	369
3	18.0	8.0	318.0	150.0	343
4	16.0	8.0	304.0	150.0	343
5	17.0	8.0	302.0	140.0	344
6	15.0	8.0	429.0	198.0	434
7	14.0	8.0	454.0	220.0	435
8	14.0	8.0	440.0	215.0	431
9	14.0	8.0	455.0	225.0	442
10	15.0	8.0	390.0	190.0	385
11	23.5145728...	4.0	133.0	115.0	309
12	23.5145728...	8.0	350.0	165.0	414
13	23.5145728...	8.0	351.0	153.0	403
14	23.5145728...	8.0	383.0	175.0	416
15	23.5145728...	8.0	360.0	175.0	385
16	15.0	8.0	383.0	170.0	356
17	14.0	8.0	340.0	160.0	360
18	23.5145728...	8.0	302.0	140.0	335
19	15.0	8.0	400.0	150.0	376
20	14.0	8.0	455.0	225.0	308
21	24.0	4.0	113.0	95.0	237
22	22.0	6.0	198.0	95.0	283
23	18.0	6.0	199.0	97.0	277

Relation: aut
Instances: 40

ttributes: 9
f weights: 406

No.	Na
1	mp
2	cyl
3	dis
4	ho
5	we
6	ac
7	mc
8	ori
9	ca

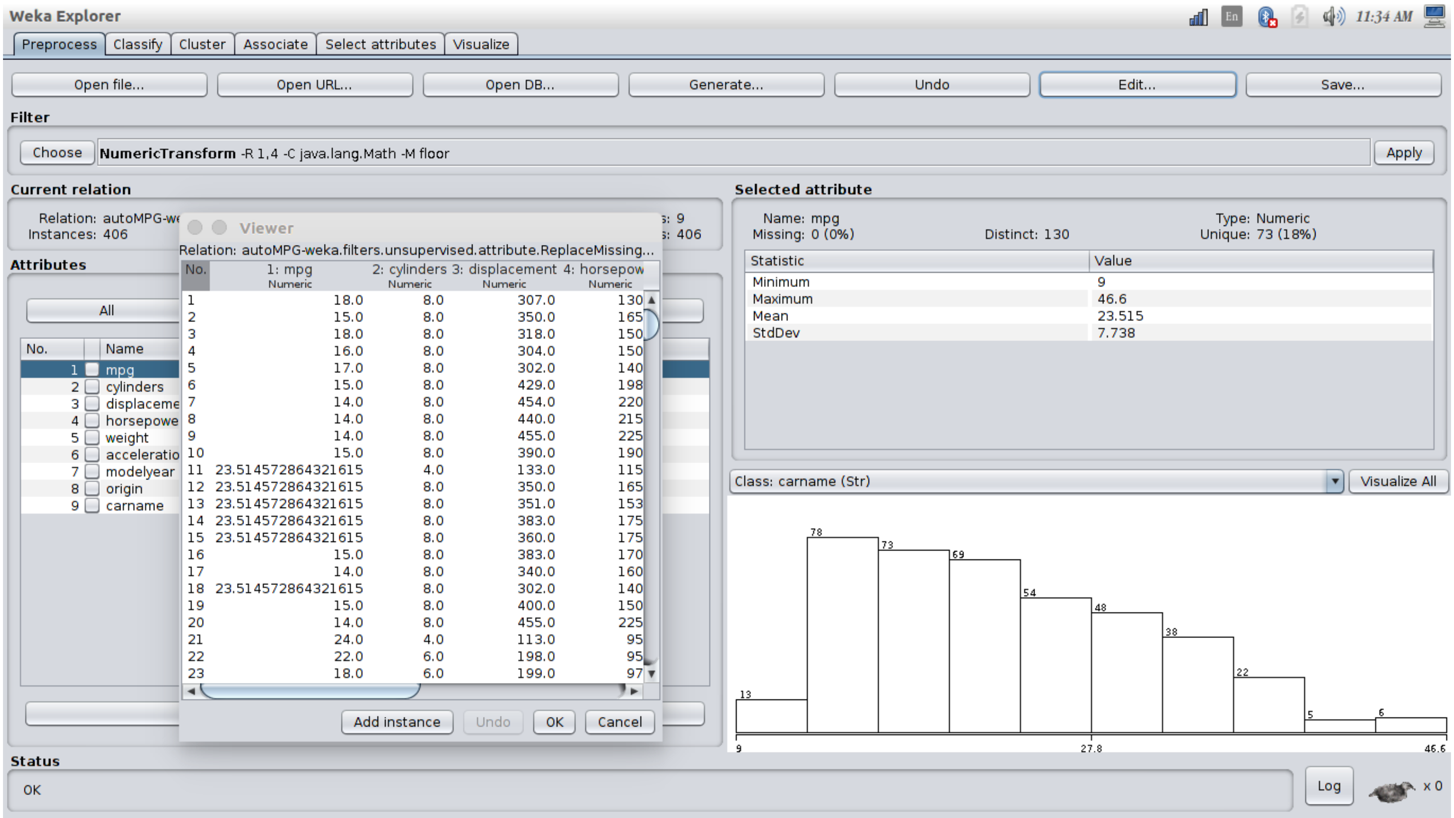
Selected attribute

Name: mpg
Missing: 0 (0%)
Distinct: 130
Type: Numeric
Unique: 73 (18%)

Statistic	Value
Minimum	9
Maximum	46.6
Mean	23.515
StdDev	7.738

Class: carname (Str)

OK



NumericTransform -R 1,4 -C java.lang.Math -M floor

Relation: autoMPG-we
Instances: 406

s: 9
s: 406

Relation: autoMPG-weka.filters.unsupervised.attribute.ReplaceMissing...

No.	1: mpg Numeric	2: cylinders Numeric	3: displacement Numeric	4: horsepow Numeric
1	18.0	8.0	307.0	130
2	15.0	8.0	350.0	165
3	18.0	8.0	318.0	150
4	16.0	8.0	304.0	150
5	17.0	8.0	302.0	140
6	15.0	8.0	429.0	198
7	14.0	8.0	454.0	220
8	14.0	8.0	440.0	215
9	14.0	8.0	455.0	225
10	15.0	8.0	390.0	190
11	23.514572864321615	4.0	133.0	115
12	23.514572864321615	8.0	350.0	165
13	23.514572864321615	8.0	351.0	153
14	23.514572864321615	8.0	383.0	175
15	23.514572864321615	8.0	360.0	175
16	15.0	8.0	383.0	170
17	14.0	8.0	340.0	160
18	23.514572864321615	8.0	302.0	140
19	15.0	8.0	400.0	150
20	14.0	8.0	455.0	225
21	24.0	4.0	113.0	95
22	22.0	6.0	198.0	95
23	18.0	6.0	199.0	97

No.	Name
1	mpg
2	cylinders
3	displaceme
4	horsepowe
5	weight
6	acceleratio
7	modelyear
8	origin
9	carname

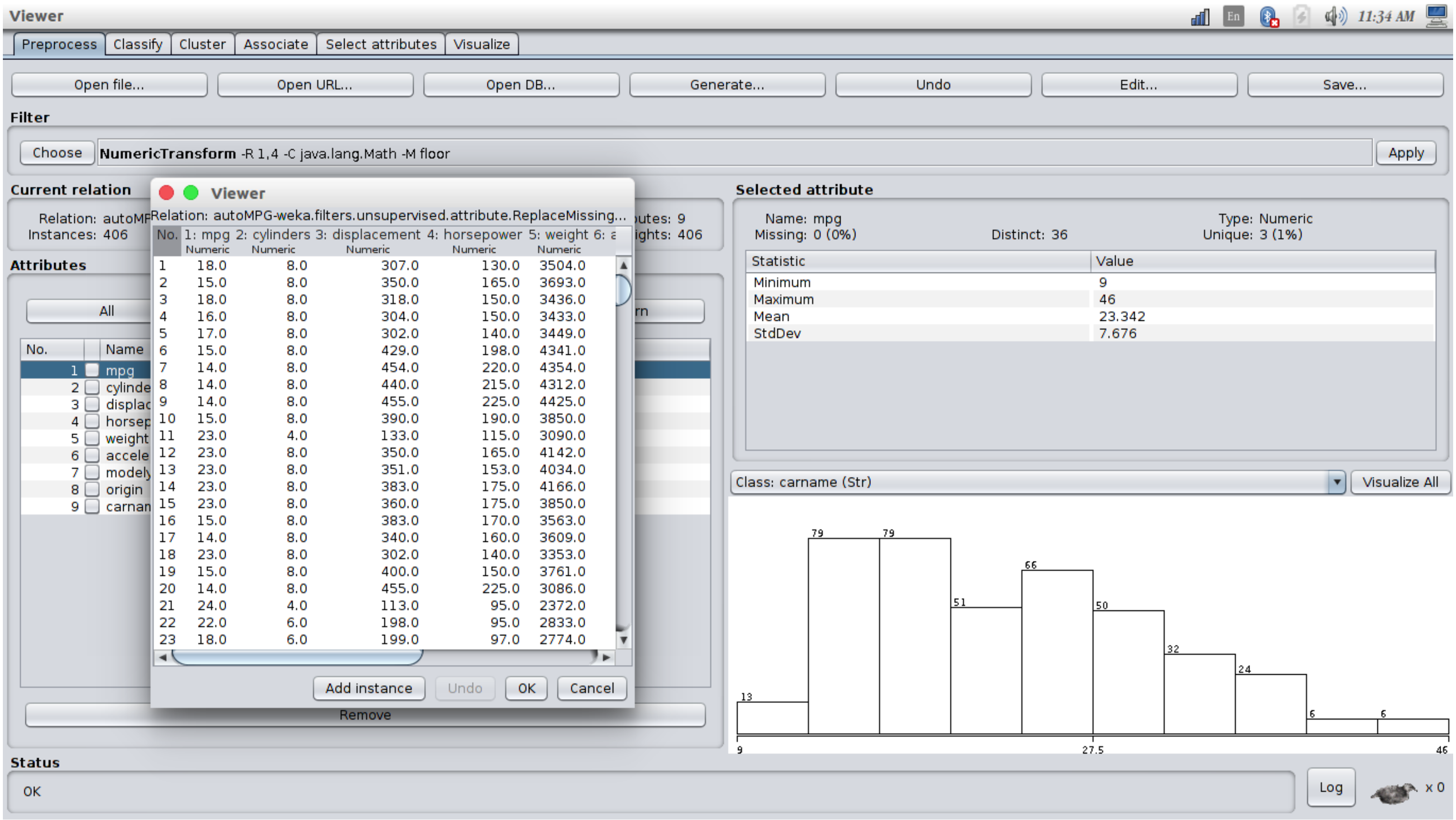
Selected attribute

Name: mpg
Missing: 0 (0%)
Distinct: 130
Type: Numeric
Unique: 73 (18%)

Statistic	Value
Minimum	9
Maximum	46.6
Mean	23.515
StdDev	7.738

Class: carname (Str)

OK



NumericTransform -R 1,4 -C java.lang.Math -M floor

Relation: autoMP
Instances: 406
...utes: 9
...ights: 406

No.		Name
1	□	mpg
2	□	cylinde
3	□	displac
4	□	horsep
5	□	weight
6	□	accele
7	□	modely
8	□	origin
9	□	carnan

Viewer

Relation: autoMPG-weka.filters.unsupervised.attribute.ReplaceMissing...

No.	1: mpg Numeric	2: cylinders Numeric	3: displacement Numeric	4: horsepower Numeric	5: weight Numeric	6: a
1	18.0	8.0	307.0	130.0	3504.0	
2	15.0	8.0	350.0	165.0	3693.0	
3	18.0	8.0	318.0	150.0	3436.0	
4	16.0	8.0	304.0	150.0	3433.0	
5	17.0	8.0	302.0	140.0	3449.0	
6	15.0	8.0	429.0	198.0	4341.0	
7	14.0	8.0	454.0	220.0	4354.0	
8	14.0	8.0	440.0	215.0	4312.0	
9	14.0	8.0	455.0	225.0	4425.0	
10	15.0	8.0	390.0	190.0	3850.0	
11	23.0	4.0	133.0	115.0	3090.0	
12	23.0	8.0	350.0	165.0	4142.0	
13	23.0	8.0	351.0	153.0	4034.0	
14	23.0	8.0	383.0	175.0	4166.0	
15	23.0	8.0	360.0	175.0	3850.0	
16	15.0	8.0	383.0	170.0	3563.0	
17	14.0	8.0	340.0	160.0	3609.0	
18	23.0	8.0	302.0	140.0	3353.0	
19	15.0	8.0	400.0	150.0	3761.0	
20	14.0	8.0	455.0	225.0	3086.0	
21	24.0	4.0	113.0	95.0	2372.0	
22	22.0	6.0	198.0	95.0	2833.0	
23	18.0	6.0	199.0	97.0	2774.0	

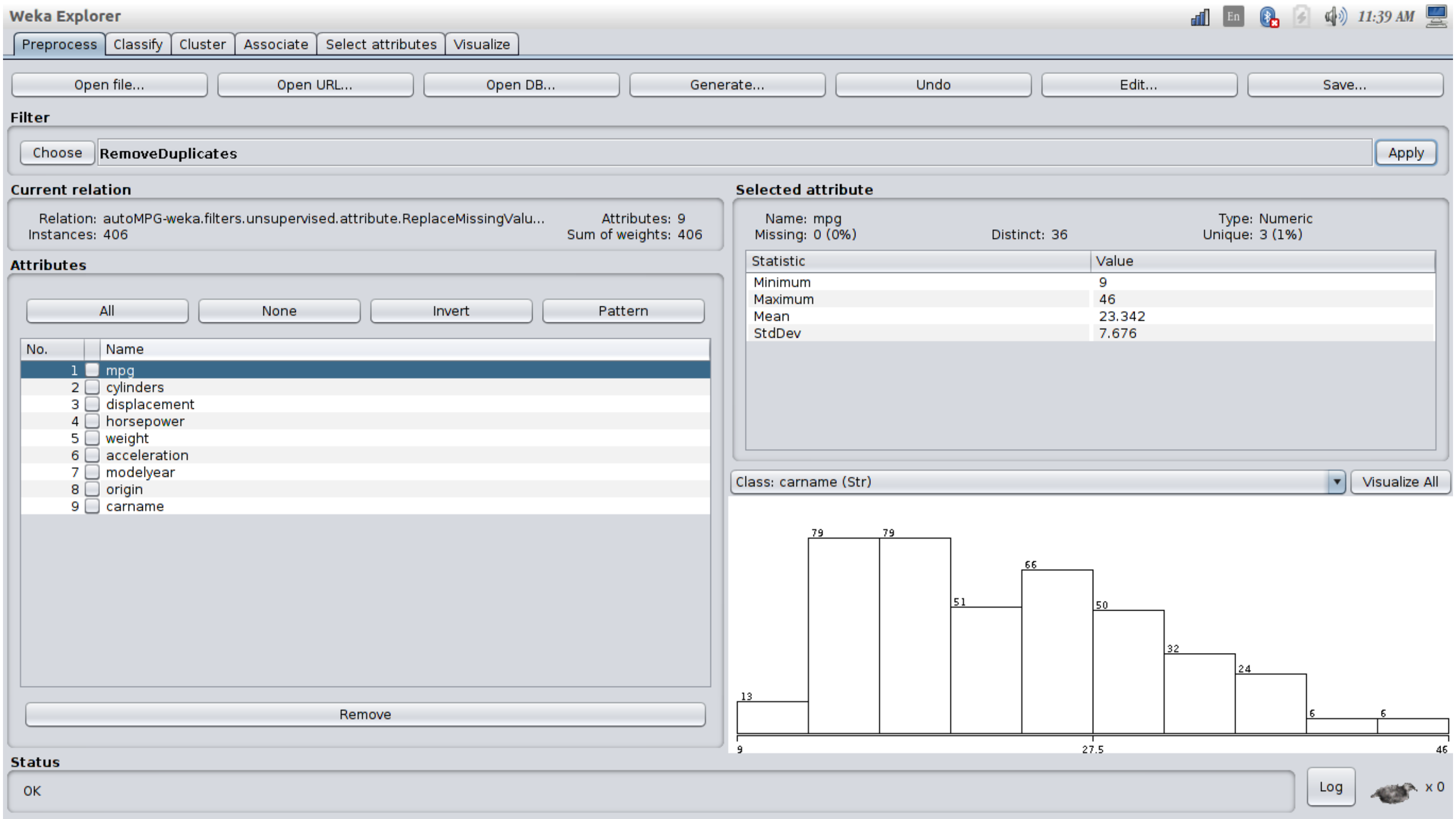
Name: mpg
Type: Numeric
Missing: 0 (0%)
Distinct: 36
Unique: 3 (1%)

Statistic	Value
Minimum	9
Maximum	46
Mean	23.342
StdDev	7.676

Class: carname (Str)

OK

e) Removing duplicate rows



Filter

RemoveDuplicates

Current relation

Relation: autoMPG-weka.filters.unsupervised.attribute.ReplaceMissingValu... Attributes: 9
Instances: 406 Sum of weights: 406

Selected attribute

Name: mpg Type: Numeric
Missing: 0 (0%) Distinct: 36 Unique: 3 (1%)

Statistic	Value
Minimum	9
Maximum	46
Mean	23.342
StdDev	7.676

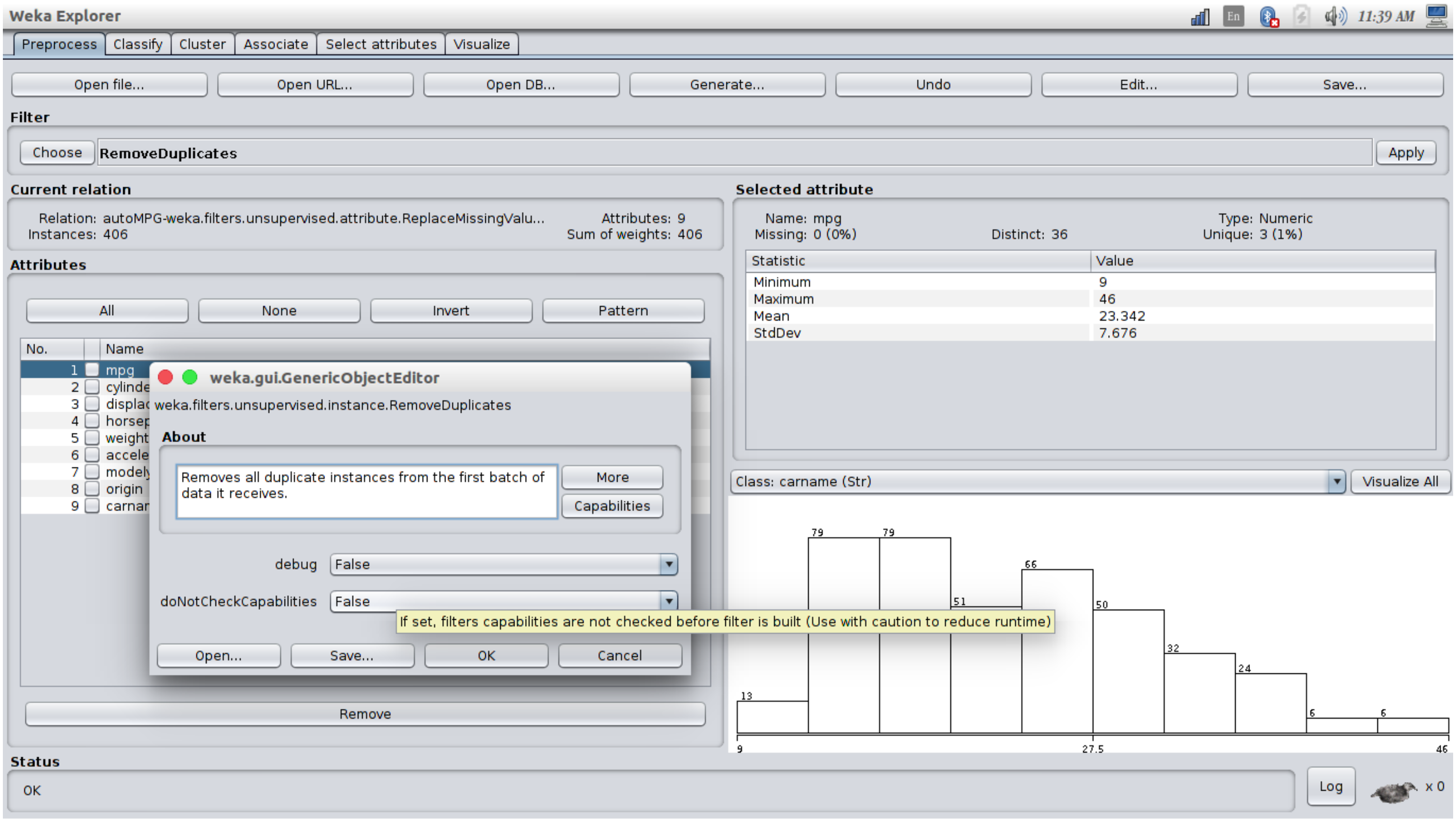
Attributes

No.	Name
1	☐ mpg
2	☐ cylinders
3	☐ displacement
4	☐ horsepower
5	☐ weight
6	☐ acceleration
7	☐ modelyear
8	☐ origin
9	☐ carname

Class: carname (Str)

Status

OK



Filter

RemoveDuplicates

Current relation

Relation: autoMPG-weka.filters.unsupervised.attribute.ReplaceMissingValu... Attributes: 9
Instances: 406 Sum of weights: 406

Selected attribute

Name: mpg Type: Numeric
Missing: 0 (0%) Distinct: 36 Unique: 3 (1%)

Statistic	Value
Minimum	9
Maximum	46
Mean	23.342
StdDev	7.676

Attributes

No.	Name
1	☐ mpg
2	☐ cylinde
3	☐ displac
4	☐ horsep
5	☐ weight
6	☐ accele
7	☐ modely
8	☐ origin
9	☐ carnar

weka.gui.GenericObjectEditor

weka.filters.unsupervised.instance.RemoveDuplicates

About

Removes all duplicate instances from the first batch of data it receives.

debug: False

doNotCheckCapabilities: False

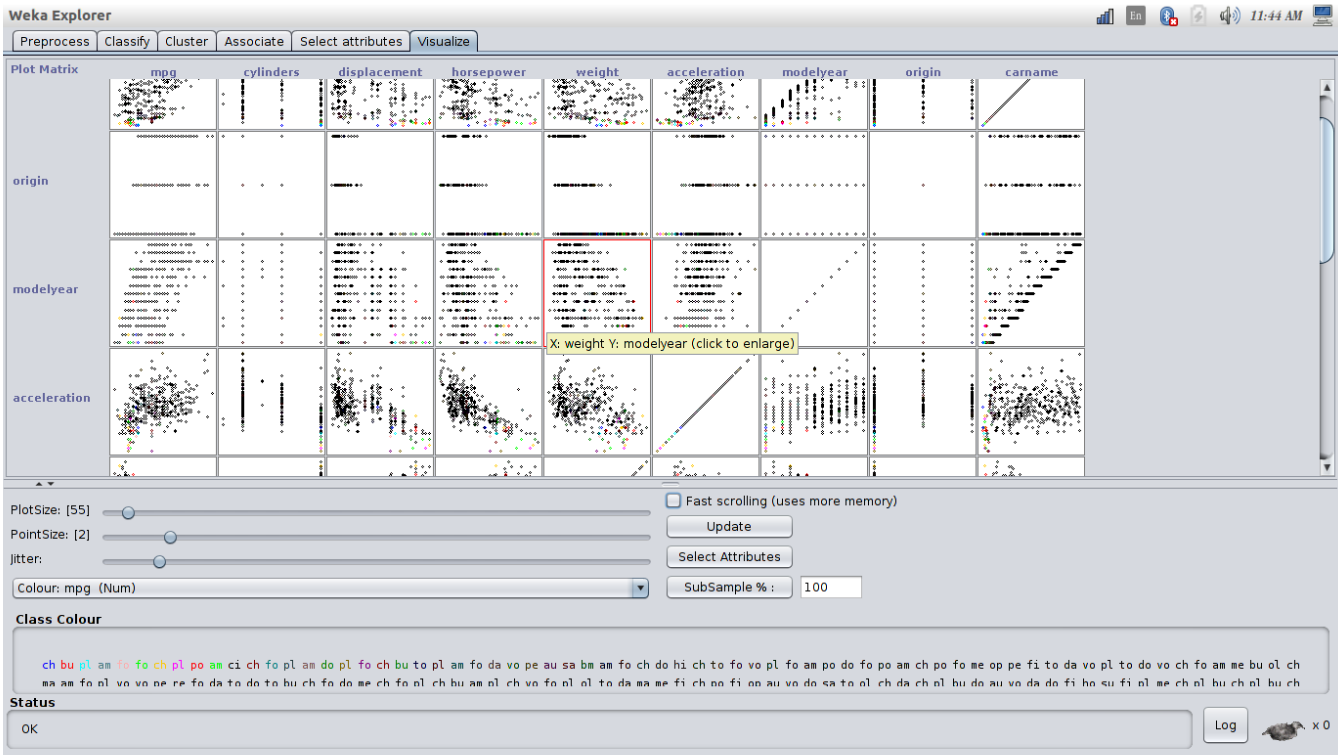
If set, filters capabilities are not checked before filter is built (Use with caution to reduce runtime)

Class: carname (Str)

Status

OK

f)Using a scatter plot to visualize relationships between values in different columns



PlotSize: [55]

PointSize: [2]

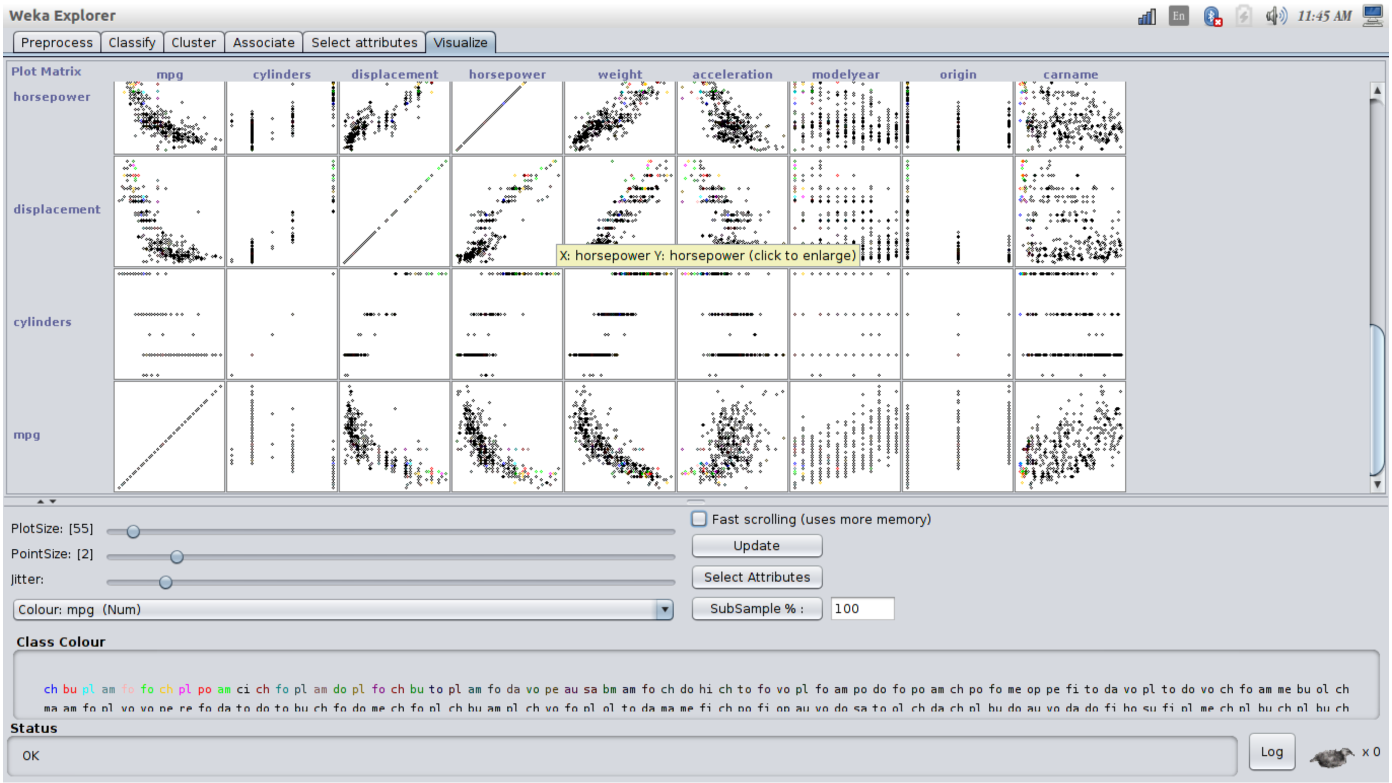
Jitter:

Colour: mpg (Num)

SubSample % : 100

ch bu pl am fo ch pl po am ci ch fo pl am do pl fo ch bu to pl am fo da vo pe au sa bm am fo ch do hi ch to fo vo pl fo am po do fo po am ch po fo me op pe fi to da vo pl to do vo ch fo am me bu ol ch
ma am fo pl vo vo pe re fo da to do to bu ch fo do me ch fo pl ch bu am pl ch vo fo pl ol to da ma me fi ch po fi op au vo do sa to ol ch da ch pl bu do au vo da do fi ho su fi pl me ch pl bu ch pl bu ch

Status: OK



PlotSize: [55]

PointSize: [2]

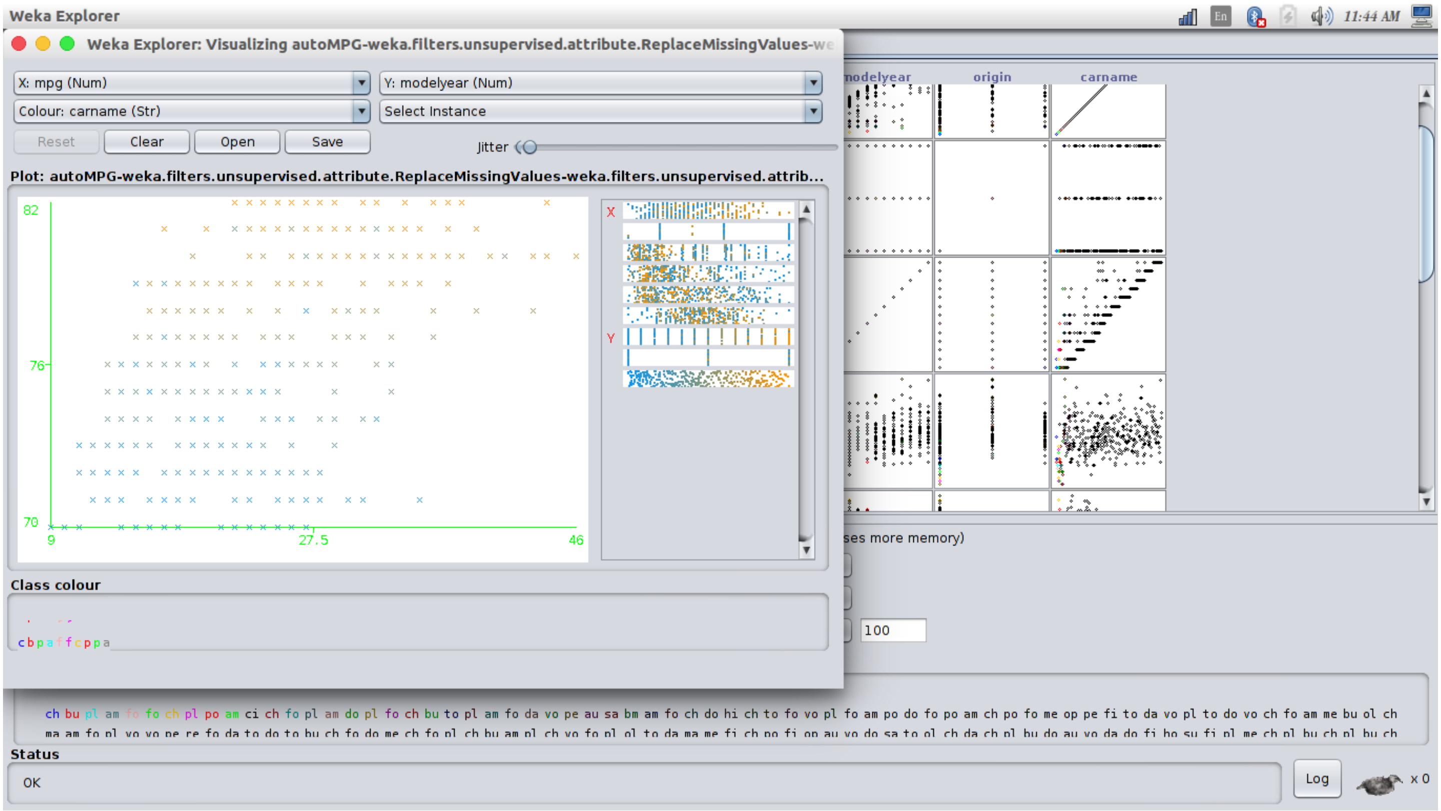
Jitter:

Colour: mpg (Num)

SubSample % : 100

ch bu pl am fo ch pl po am ci ch fo pl am do pl fo ch bu to pl am fo da vo pe au sa bm am fo ch do hi ch to fo vo pl fo am po do fo po am ch po fo me op pe fi to da vo pl to do vo ch fo am me bu ol ch
ma am fo pl vo vo pe re fo da to do to bu ch fo do me ch fo pl ch bu am pl ch vo fo pl ol to da ma me fi ch po fi op au vo do sa to ol ch da ch pl bu do au vo da do fi ho su fi pl me ch pl bu ch pl bu ch

Status: OK



Weka Explorer

Weka Explorer: Visualizing autoMPG-weka.filters.unsupervised.attribute.ReplaceMissingValues-we

X: mpg (Num)

Y: modelyear (Num)

Colour: carname (Str)

Select Instance

Plot: autoMPG-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attrib...

Class colour

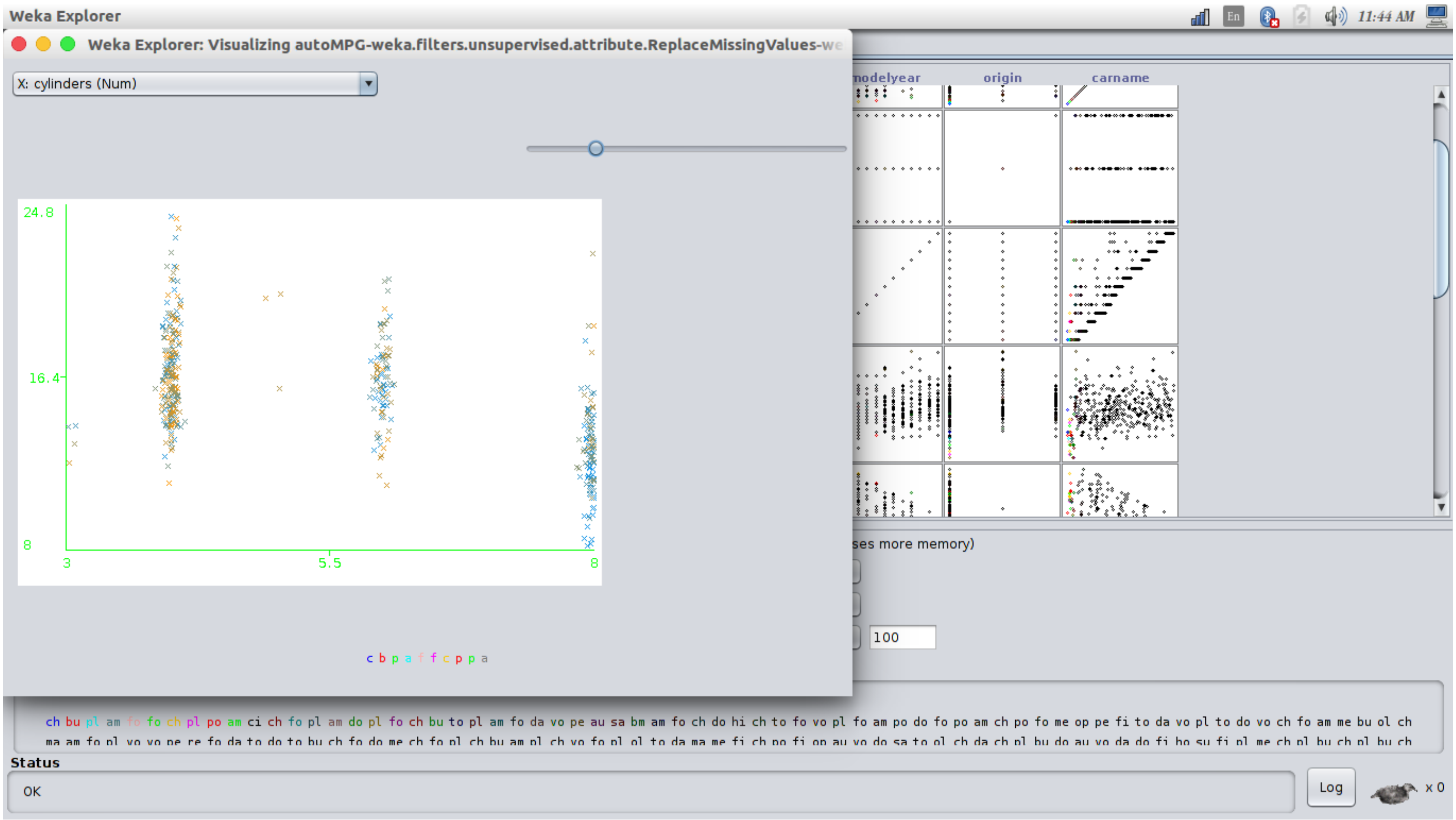
c b p a f f c p p a

ch bu pl am fo fo ch pl po am ci ch fo pl am do pl fo ch bu to pl am fo da vo pe au sa bm am fo ch do hi ch to fo vo pl fo am po do fo po am ch po fo me op pe fi to da vo pl to do vo ch fo am me bu ol ch

ma am fo pl vo vo pe re fo da to do to bu ch fo do me ch fo pl ch bu am pl ch vo fo pl ol to da ma me fi ch po fi op au vo do sa to ol ch da ch pl bu do au vo da do fi ho su fi pl me ch pl bu ch pl bu ch

Status

OK



Weka Explorer

Weka Explorer: Visualizing autoMPG-weka.filters.unsupervised.attribute.ReplaceMissingValues-we

X: cylinders (Num)

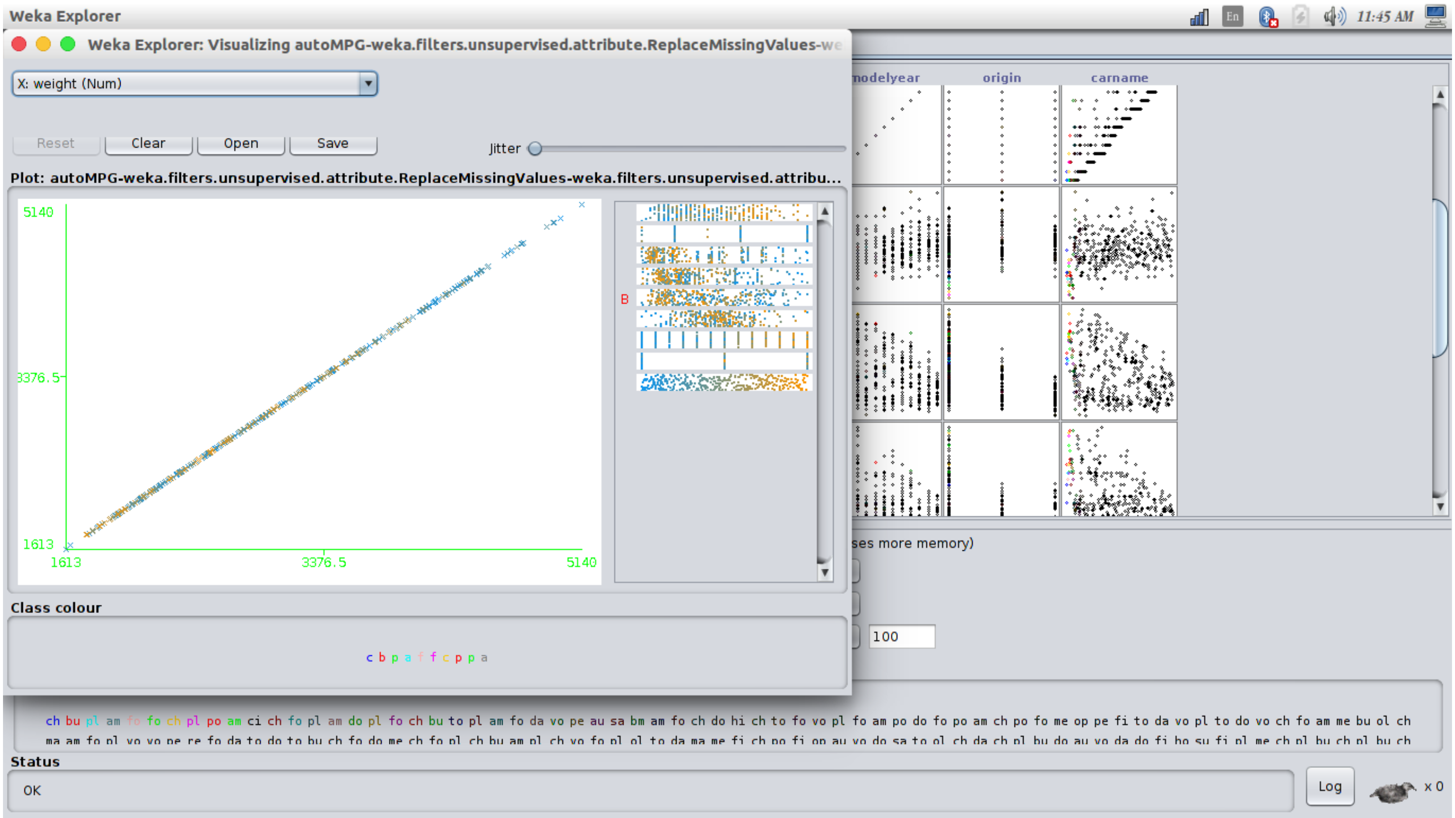
c b p a f f c p p a

ch bu pl am fo fo ch pl po am ci ch fo pl am do pl fo ch bu to pl am fo da vo pe au sa bm am fo ch do hi ch to fo vo pl fo am po do fo po am ch po fo me op pe fi to da vo pl to do vo ch fo am me bu ol ch

ma am fo pl vo vo pe re fo da to do to bu ch fo do me ch fo pl ch bu am pl ch vo fo pl ol to da ma me fi ch po fi op au vo do sa to ol ch da ch pl bu do au vo da do fi ho su fi pl me ch pl bu ch pl bu ch

Status

OK



Weka Explorer

Weka Explorer: Visualizing autoMPG-weka.filters.unsupervised.attribute.ReplaceMissingValues-we

X: weight (Num)

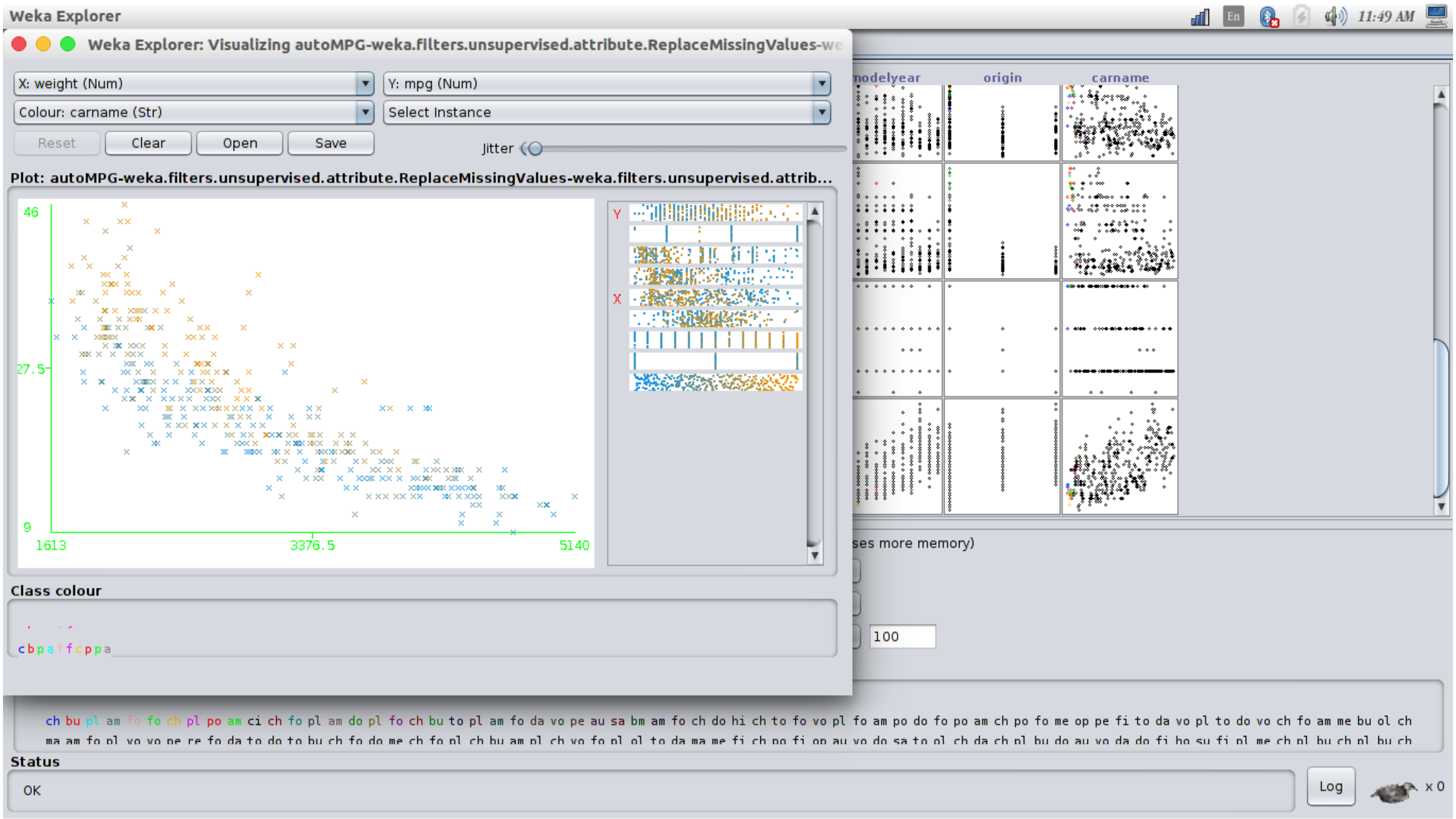
Plot: autoMPG-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribu...

Status

OK

11:45 AM
Reset
Clear
Open
Save
Jitter
5140
3376.5
1613
1613
3376.5
5140
B
Class colour
c b p a f f c p p a
nodelyear
origin
carname
ses more memory)
100
ch bu pl am fo ch pl po am ci ch fo pl am do pl fo ch bu to pl am fo da vo pe au sa bm am fo ch do hi ch to fo vo pl fo am po do fo po am ch po fo me op pe fi to da vo pl to do vo ch fo am me bu ol ch
ma am fo pl vo vo pe re fo da to do to bu ch fo do me ch fo pl ch bu am pl ch vo fo pl pl to da ma me fi ch po fi op au vo do sa to ol ch da ch pl bu do au vo da do fi ho su fi pl me ch pl bu ch pl bu ch
Log
x 0

(using same attribute as x-axis and y-axis)



Weka Explorer

Weka Explorer: Visualizing autoMPG-weka.filters.unsupervised.attribute.ReplaceMissingValues-we

X: weight (Num)

Y: mpg (Num)

Colour: carname (Str)

Select Instance

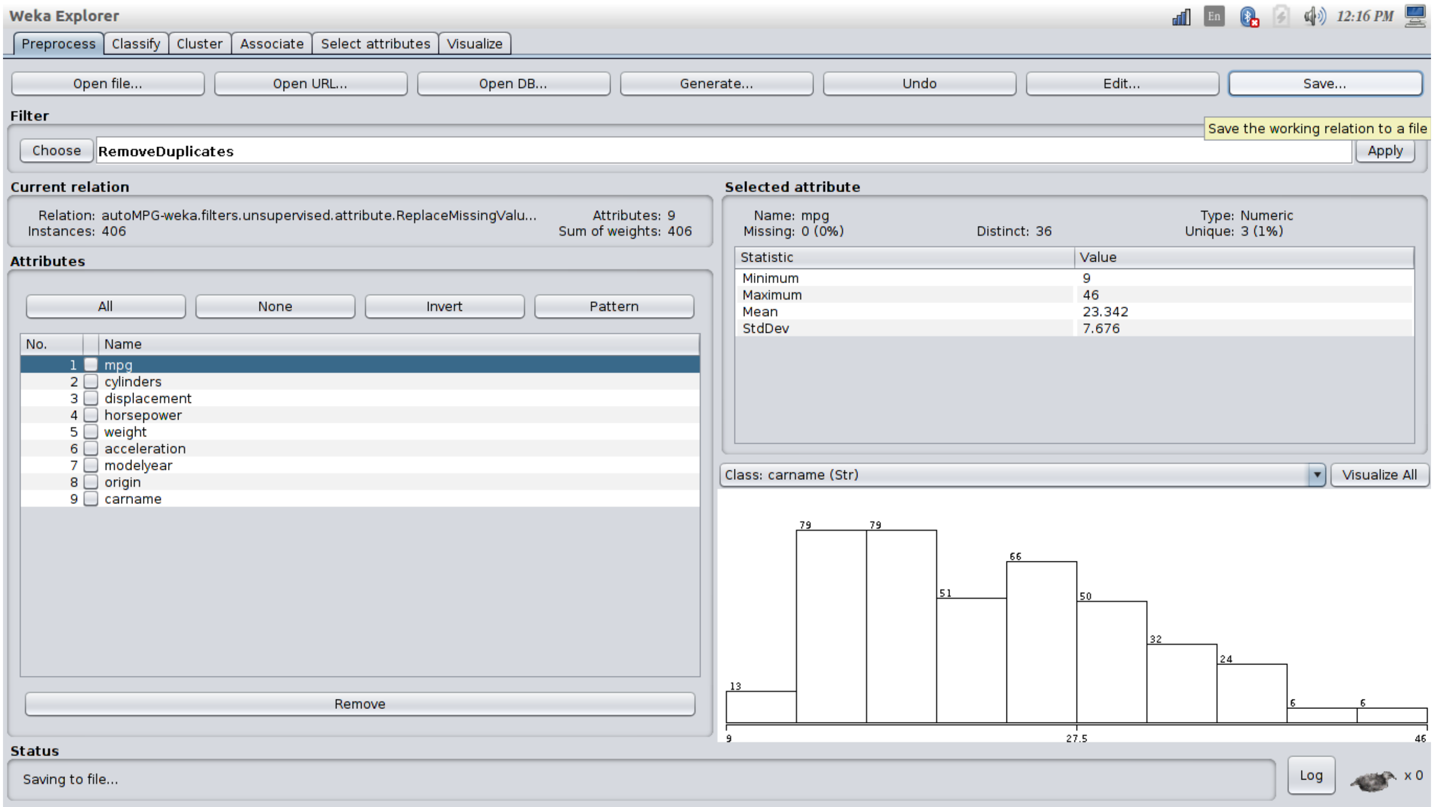
Plot: autoMPG-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attrib...

Status

OK

11:49 AM
Reset
Clear
Open
Save
Jitter
46
27.5
9
1613
3376.5
5140
Y
X
Class colour
c b p a f f c p p a
nodelyear
origin
carname
ses more memory)
100
ch bu pl am fo ch pl po am ci ch fo pl am do pl fo ch bu to pl am fo da vo pe au sa bm am fo ch do hi ch to fo vo pl fo am po do fo po am ch po fo me op pe fi to da vo pl to do vo ch fo am me bu ol ch
ma am fo pl vo vo pe re fo da to do to bu ch fo do me ch fo pl ch bu am pl ch vo fo pl pl to da ma me fi ch po fi op au vo do sa to ol ch da ch pl bu do au vo da do fi ho su fi pl me ch pl bu ch pl bu ch
Log
x 0

g)Exporting cleaned data to Excel



Filter

RemoveDuplicates

Current relation

Relation: autoMPG-weka.filters.unsupervised.attribute.ReplaceMissingValu...   Attributes: 9
Instances: 406   Sum of weights: 406

Selected attribute

Name: mpg   Type: Numeric
Missing: 0 (0%)   Distinct: 36   Unique: 3 (1%)

Statistic	Value
Minimum	9
Maximum	46
Mean	23.342
StdDev	7.676

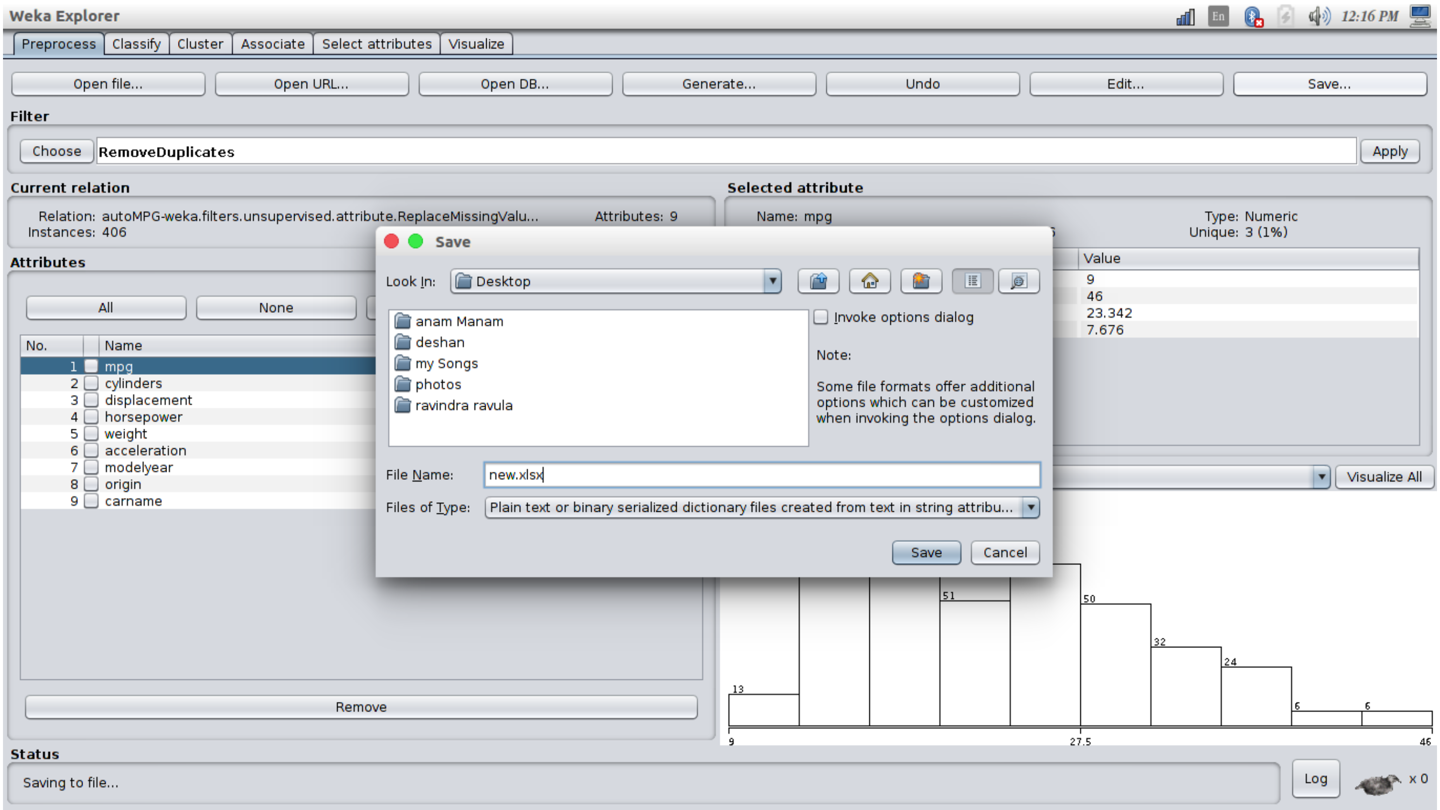
Attributes

No.		Name
1	□	mpg
2	□	cylinders
3	□	displacement
4	□	horsepower
5	□	weight
6	□	acceleration
7	□	modelyear
8	□	origin
9	□	carname

Class: carname (Str)

Status

Saving to file...



Filter

RemoveDuplicates

Current relation

Relation: autoMPG-weka.filters.unsupervised.attribute.ReplaceMissingValu...   Attributes: 9
Instances: 406

Selected attribute

Name: mpg   Type: Numeric
Unique: 3 (1%)

Value
9
46
23.342
7.676

Attributes

No.		Name
1	□	mpg
2	□	cylinders
3	□	displacement
4	□	horsepower
5	□	weight
6	□	acceleration
7	□	modelyear
8	□	origin
9	□	carname

Save

Look In: Desktop

- anam Manam
- deshan
- my Songs
- photos
- ravindra ravula

□ Invoke options dialog

Note:

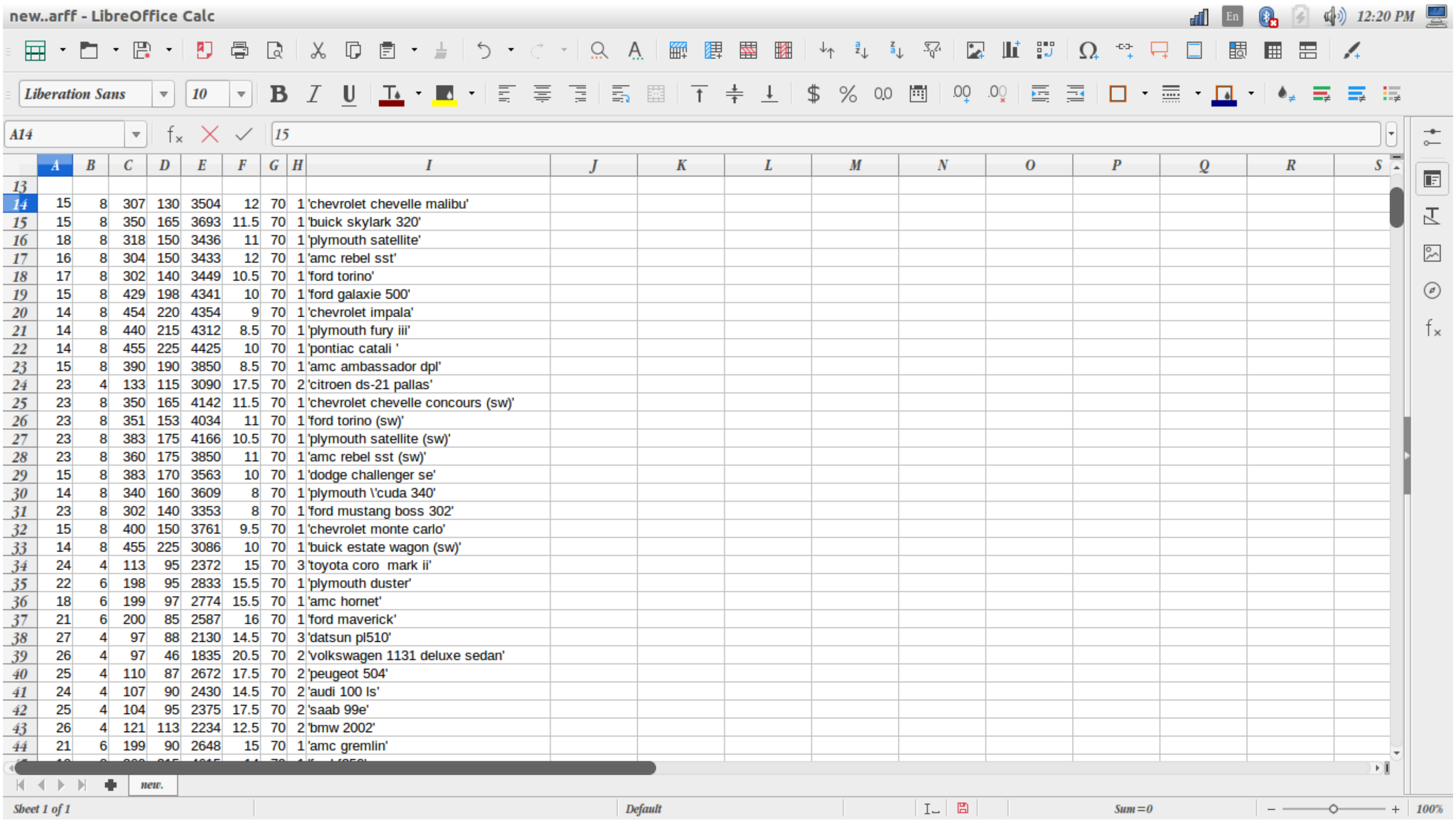
Some file formats offer additional options which can be customized when invoking the options dialog.

File Name: new.xlsx

Files of Type: Plain text or binary serialized dictionary files created from text in string attribu...

Status

Saving to file...



	A	B	C	D	E	F	G	H	I
13									
14	15	8	307	130	3504	12	70	1	'chevrolet chevelle malibu'
15	15	8	350	165	3693	11.5	70	1	'buick skylark 320'
16	18	8	318	150	3436	11	70	1	'plymouth satellite'
17	16	8	304	150	3433	12	70	1	'amc rebel sst'
18	17	8	302	140	3449	10.5	70	1	'ford torino'
19	15	8	429	198	4341	10	70	1	'ford galaxie 500'
20	14	8	454	220	4354	9	70	1	'chevrolet impala'
21	14	8	440	215	4312	8.5	70	1	'plymouth fury iii'
22	14	8	455	225	4425	10	70	1	'pontiac catali '
23	15	8	390	190	3850	8.5	70	1	'amc ambassador dpl'
24	23	4	133	115	3090	17.5	70	2	'citroen ds-21 pallas'
25	23	8	350	165	4142	11.5	70	1	'chevrolet chevelle concours (sw)'
26	23	8	351	153	4034	11	70	1	'ford torino (sw)'
27	23	8	383	175	4166	10.5	70	1	'plymouth satellite (sw)'
28	23	8	360	175	3850	11	70	1	'amc rebel sst (sw)'
29	15	8	383	170	3563	10	70	1	'dodge challenger se'
30	14	8	340	160	3609	8	70	1	'plymouth \'cuda 340'
31	23	8	302	140	3353	8	70	1	'ford mustang boss 302'
32	15	8	400	150	3761	9.5	70	1	'chevrolet monte carlo'
33	14	8	455	225	3086	10	70	1	'buick estate wagon (sw)'
34	24	4	113	95	2372	15	70	3	'toyota coro mark ii'
35	22	6	198	95	2833	15.5	70	1	'plymouth duster'
36	18	6	199	97	2774	15.5	70	1	'amc hornet'
37	21	6	200	85	2587	16	70	1	'ford maverick'
38	27	4	97	88	2130	14.5	70	3	'datsun pl510'
39	26	4	97	46	1835	20.5	70	2	'volkswagen 1131 deluxe sedan'
40	25	4	110	87	2672	17.5	70	2	'peugeot 504'
41	24	4	107	90	2430	14.5	70	2	'audi 100 ls'
42	25	4	104	95	2375	17.5	70	2	'saab 99e'
43	26	4	121	113	2234	12.5	70	2	'bmw 2002'
44	21	6	199	90	2648	15	70	1	'amc gremlin'

Differences of Weka And Open Refine

1) visualization of data in weka is much more clear compare to refine and easy. The plot matrix of weka will give a good visualization and easy to understand the properties of data set.

2)weka supports histogram visualized method.but not in refine.

3)Operations like Removing Duplicates, removing inconsistent spelling of terms can be easily perform in weka.

4)Tuple format data consideration can perform in weka while column format data classification can perform in open refine.

5)Clustering and merging can perform easily in open refine compared to Weka.

6) Replacing missing Values can be perform only in weka while removing data tuples with   missing value can be perform in both open refine and weka.

7) No inbuild, direct function to remove duplicate rows in open refine.but in weka .