Knowledge Discovery and Data Mining Lab-01

Name : Kunal Sanjay Patil
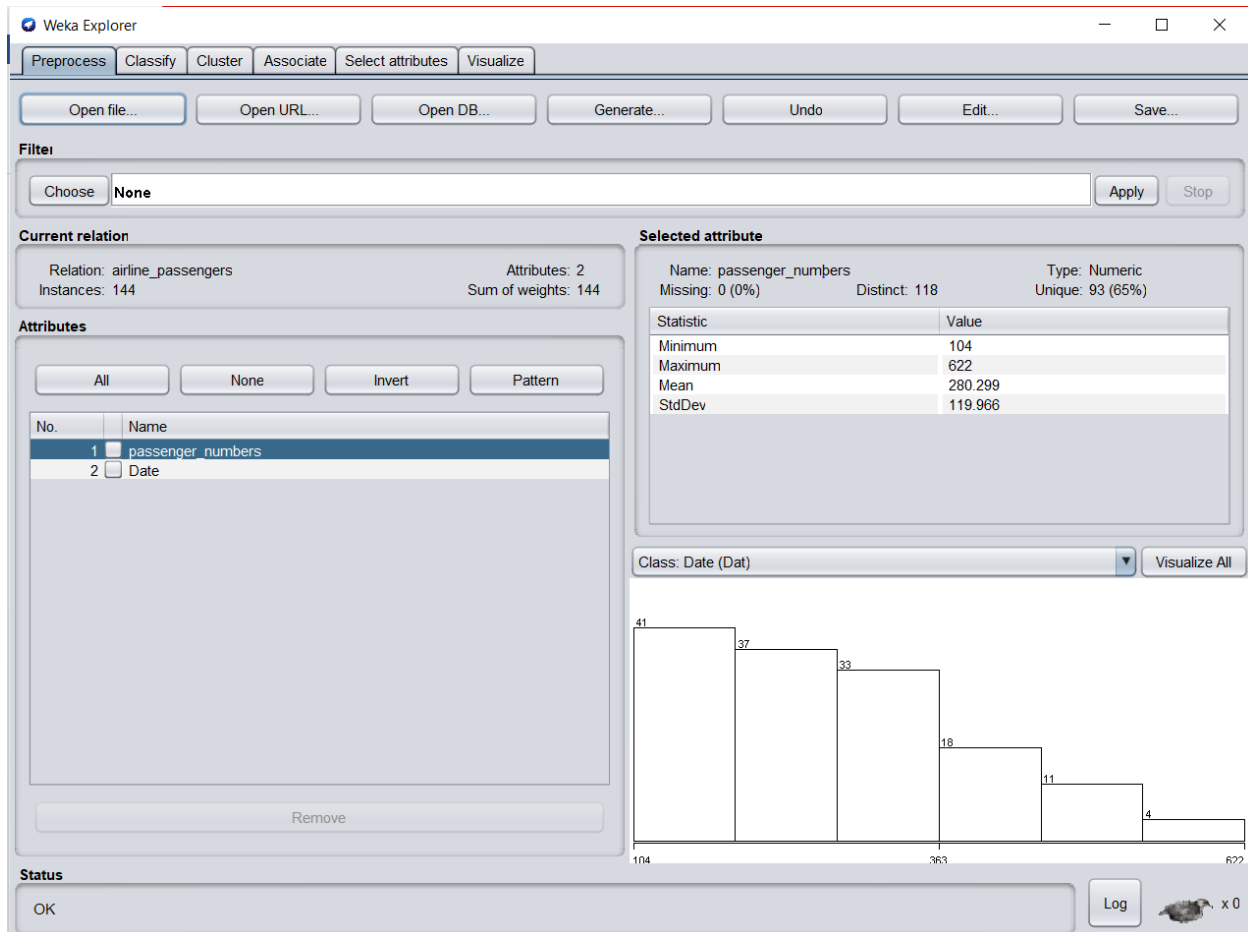PRN : 20190802025

AIM:

1. airline.arff

The airline dataset consists;

Number of instances: 144

Attribute information:

Passenger_numbers:

Type: Numeric, Missing values: 0, Distinct: 118, Unique: 93

Maximum value: 622, Minimum value: 104, Mean: 280.299, Standard Deviation: 119.966

Date:

Type: Date, Missing values: 0, Distinct: 144, Unique: 144

Maximum: -286695000000, Minimum: -662707800000, Mean: -474769800000, Standard Deviation: 109697660835.156

2. Breast-cancer.arff

The breast-cancer dataset consists;

Number of instances: 286

Attribute Information:

1) age:

Type: Nominal, Missing values: 0, Distinct: 6, Unique: 1

Label	Count	Weight
10-19	0	0.0
20-29	1	1.0
30-39	36	36.0
40-49	90	90.0
50-59	96	96.0
60-69	57	57.0
70-79	6	6.0
80-89	0	0.0
90-99	0	0.0

2) menopause:

Type: Nominal, Missing: 0, Distinct: 3, Unique: 0

Label	Count	Weight
lt40	7	7.0
ge40	129	129.0
premeno	150	150.0

3) tumor-size:
Type: Nominal, Missing: 0, Distinct: 11, Unique: 0

Label	Count	Weight
0-4	8	8.0
5-9	4	4.0
10-14	28	28.0
15-19	30	30.0
20-24	50	50.0
25-29	54	54.0
30-34	60	60.0
35-39	19	19.0
40-44	22	22.0
45-49	3	3.0
50-54	8	8.0
55-59	0	0.0

4) inv-nodes:
Type: Nominal, Missing: 0, Distinct: 7, Unique: 1

Label	Count	Weight
0-2	213	213.0
3-5	36	36.0
6-8	17	17.0
9-11	10	10.0
12-14	3	3.0
15-17	6	6.0
18-20	0	0.0
21-23	0	0.0
24-26	1	1.0
27-29	0	0.0
30-32	0	0.0
33-35	0	0.0
36-39	0	0.0

5) node-caps:
Type: Nominal, Missing 8, Distinct: 2, Unique: 0

Label	Count	Weight
yes	56	56.0
no	222	222.0

6) deg-malig:
Type: Nominal, Missing:0, Distinct: 3, Unique: 0

Label	Count	Weight
1	71	71.0
2	130	130.0
3	85	85.0

7) breast:

Type: Nominal, Missing: 0, Distinct: 2, Unique: 0

Label	Count	Weight
left	152	152.0
right	134	134.0

8) breast-quad:

Type: Nominal, Missing: 1, Distinct: 5, Unique: 0

Label	Count	Weight
left_up	97	97.0
left_low	110	110.0
right_up	33	33.0
right_low	24	24.0
central	21	21.0

9) irradiat:

Type: Nominal, Missing:0, Distinct: 2, Unique: 0

Label	Count	Weight
yes	68	68.0
no	218	218.0

10) Class:

Type: Nominal, Missing: 0, Distinct: 2, Unique: 0

Label	Count	Weight
no-recurrence...	201	201.0
recurrence-ev...	85	85.0

3. Contact-lenses.arff
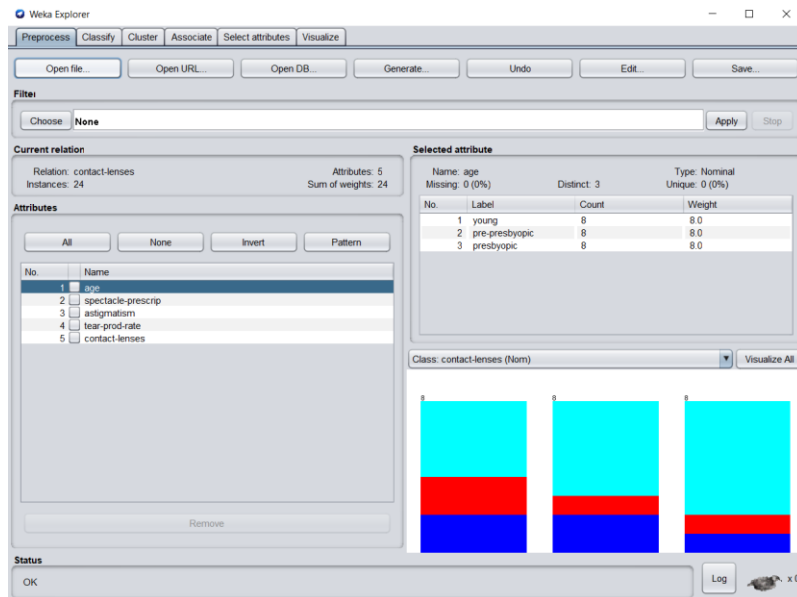
The contact-lenses dataset consists:

No. of instances: 24

No. of attributes: 5

Class objects: age, spectacle-prescrip, astigmatism, tear-prod-rate, contact-lenses

Attribute Information:

1) age:

Type: Nominal, Missing: 0, Distinct: 3, Unique: 0

Label	Count
young	8
pre-presbyopic	8
presbyopic	8

2) spectacle-prescrip:

Type: Nominal, Missing: 0, Distinct: 2, Unique: 0

Label	Count
myope	12
hypermetrope	12

3) astigmatism:

Type: Nominal, Missing: 0, Distinct: 2, Unique: 0

Label	Count
no	12
yes	12

4) tear-prod-rate:

Type: Nominal, Missing: 0, Distinct: 2, Unique: 0

Label	Count
reduced	12
normal	12

5) contact-lenses:
Type: Nominal, Missing: 0, Distinct: 3, Unique: 0

Label	Count
soft	5
hard	4
none	15

4. Cpu.arff
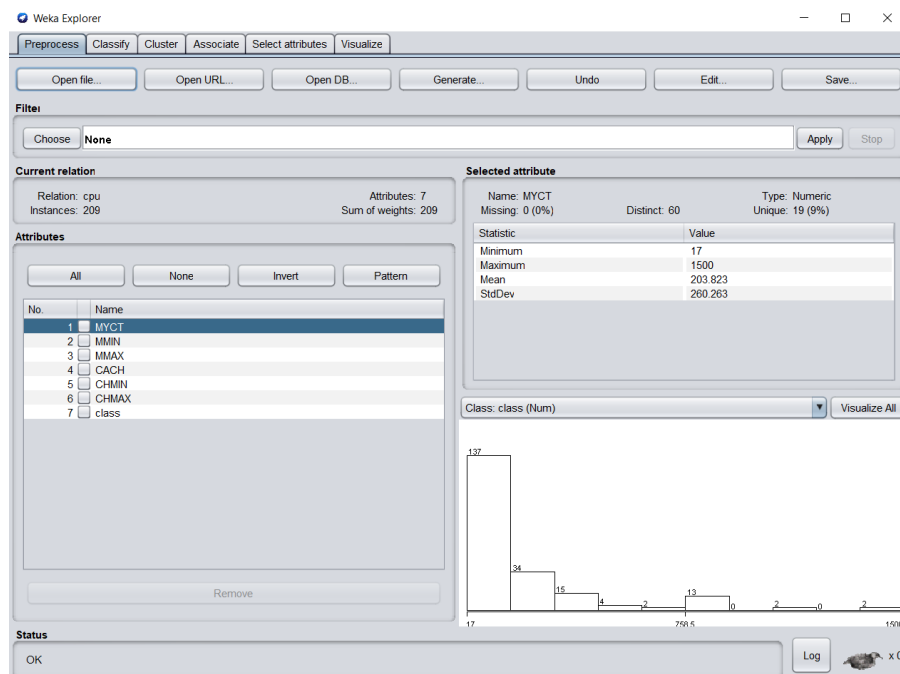
The cpu dataset consists:
No. of instances: 209
No. of attributes: 7
Class objects: MYCT, MMIN, MMAX, CACH, CHMIN, CHMAX, class
Data type: Numeric
Missing Values: 0
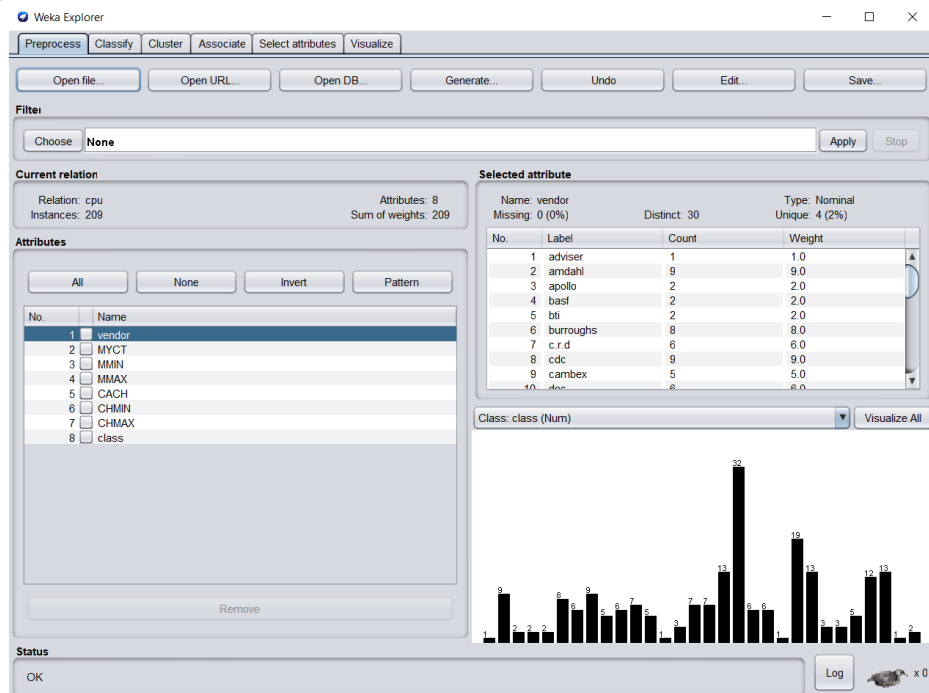Attribute information:
1) MYCT:
Distinct: 60, Unique: 19
Minimum value: 17, Maximum value: 1500, Mean: 203.823, Standard Deviation: 260.263

2) MMIN:

Distinct: 25, Unique: 11
Minimum value: 64, Maximum value: 32000, Mean: 2867.981, Standard Deviation: 3878.743

3) MMAX:
Distinct: 23, Unique: 6,
Minimum value: 64, Maximum value: 64000, Mean: 11796.153, Standard Deviation: 11726.564

4) CACH:
Distinct: 22, Unique: 4
Minimum: 0, Maximum: 256, Mean: 25.206, Standard Deviation: 40.629

5) CHMIN:
Minimum: 0, Maximum: 52, Mean: 4.699, Standard Deviation: 6.816

6) CHMAX:
Minimum: 0, Maximum: 176, Mean: 18.268, Standard Deviation: 25.997

7) Class
Minimum: 6, Maximum: 1150, Mean: 105.622, Standard Deviation: 160.831

5. Cpu,with.vendor.arff

No. of instances: 209
No. of attributes: 8
Missing Values: 0

Attribute information:

1) Vendor:
   Type: Nominal, Missing: 0, Distinct: 30, Unique: 4
   Labels: adviser, amdahl, apollo, basf, bti, burroughs, c.r.d, cdc, cambex, dec, dg, formation, four-phase, gould, hp, harris, honeywell, ibm, ipl, magnuson, microdata, nas, ncr, nixdorf, perkin-elmer, prime, siemens, sperry, sratus, wang

2) MYCT:
   Distinct: 60, Type: Numeric, Unique: 19
   Minimum value: 17, Maximum value: 1500, Mean: 203.823, Standard Deviation: 260.263

3) MMIN:
   Type: Numeric, Distinct: 25, Unique: 11
   Minimum: 64, Maximum: 32000, Mean: 2867.981, Standard Deviation: 38783743

4) MMAX:
   Type: Numeric, Distinct: 23, Unique: 6
   Minimum value: 64, Maximum value: 64000, Mean: 11796.153, Standard Deviation: 11726.564

8) CACH:
   Distinct: 22, Unique: 4
   Minimum: 0, Maximum: 256, Mean: 25.206, Standard Deviation: 40.629

9) CHMIN:
   Minimum: 0, Maximum: 52, Mean: 4.699, Standard Deviation: 6.816
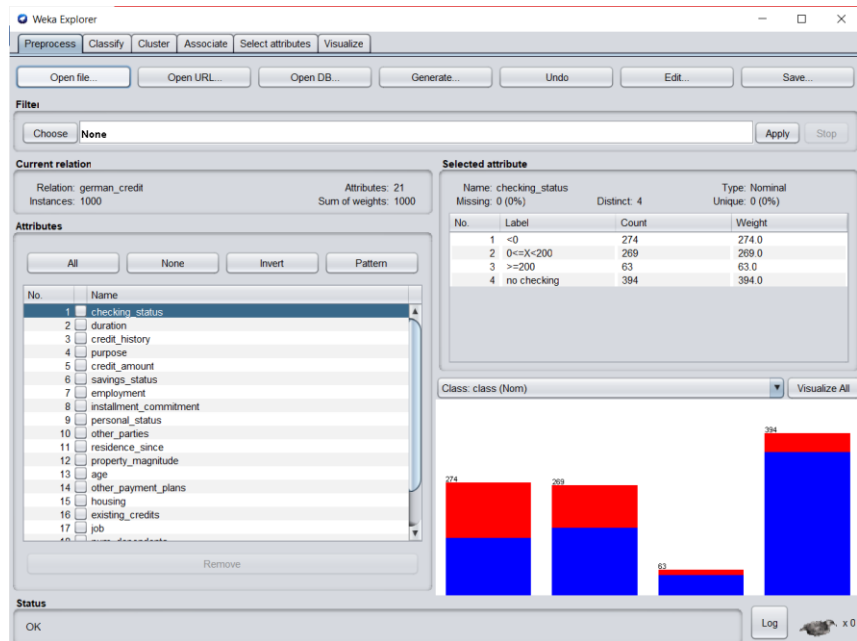
10) CHMAX:
    Minimum: 0, Maximum: 176, Mean: 18.268, Standard Deviation: 25.997

11) Class
    Minimum: 6, Maximum: 1150, Mean: 105.622, Standard Deviation: 160.831

6. Credit-g.arff

No. of instances: 1000
No. of attributes: 21
Missing values: 0
Attribute information:

1) Checking_status:

   Type: Nominal, Distinct:4, Unique: 0

Label	Count
<0	274
0<=X<200	269
>=200	63
no checking	394

2) Duration:

   Type: Numeric, Distinct: 33, Unique: 5

   Minimum: 4, Maximum: 72, Mean: 20.903, Standard Deviation: 12.059

3) Credit_history:

   Type: Nominal, Distinct: 5, Unique: 0

Label	Count
no credits/all p...	40
all paid	49
existing paid	530
delayed previo...	88
critical/other e...	293

4) Purpose:

   Type: nominal, Distinct: 10, Unique: 0

Label	Count
new car	234
used car	103
furniture/equip...	181
radio/tv	280
domestic appli...	12
repairs	22
education	50
vacation	0
retraining	9
business	97
other	12

5) Credit_amount:
   Type: Numeric, Distinct: 921, Unique: 847

Statistic	Value
Minimum	250
Maximum	18424
Mean	3271.258
StdDev	2822.737

6) Savings_status:
   Type: Nominal, Distinct: 5, Unique: 0

Label	Count
<100	603
100<=X<500	103
500<=X<1000	63
>=1000	48
no known savi...	183

7) Employment:
   Type: Nominal, Distinct: 5, Unique: 0

Label	Count
unemployed	62
<1	172
1<=X<4	339
4<=X<7	174
>=7	253

8) Installment_commitment:
   Type: Numeric, Distinct: 4, Unique: 0
   Minimum: 1, Maximum, 4, Mean: 2.973, Standard Deviation: 1.119

9) Personal_status:
   Type: Nominal, Distinct: 4, Unique: 0

Label	Count
male div/sep	50
female div/dep...	310
male single	548
male mar/wid	92
female single	0

10) Other_parties:

Type: Nominal, Distinct: 3, unique: 0

Label	Count
none	907
co applicant	41
guarantor	52

11) Residence_since:

Type: Numeric, Distinct: 4

Statistic	Value
Minimum	1
Maximum	4
Mean	2.845
StdDev	1.104

12) Property_magnitude:

Type: Nominal, Distinct: 4, Unique: 0

Label	Count
real estate	282
life insurance	232
car	332
no known pro...	154

13) Age:

Type: Numeric, Distinct: 53, Unique:1

Statistic	Value
Minimum	19
Maximum	75
Mean	35.546
StdDev	11.375

14) Other_payment_plans:

Type: Nominal, Distinct: 3, Unique: 0

Label	Count
bank	139
stores	47
none	814

15) Housing:

Type: Nominal, Distinct: 3

Label	Count
rent	179
own	713
for free	108

16) Existing_credits:

Type: Numeric, Distinct: 4

Statistic	Value
Minimum	1
Maximum	4
Mean	1.407
StdDev	0.578

17) Job:

Type: Nominal, Distinct: 4, Unique: 0

Label	Count
unemp/unskill...	22
unskilled resid...	200
skilled	630
high qualif/self ...	148

18) Num_dependents:

Type: Numeric, Distinct: 2, Unique: 0

Statistic	Value
Minimum	1
Maximum	2
Mean	1.155
StdDev	0.362

19) Own_telephone:

Type: Nominal, Distinct: 4

Label	Count
none	596
yes	404

20) Foreign_worker:

Type: Nominal, Distinct: 2

Label	Count
yes	963
no	37

21) Class:

Type: Nominal, Distinct: 2

Label	Count
good	700
bad	300

7. diabetes.arff
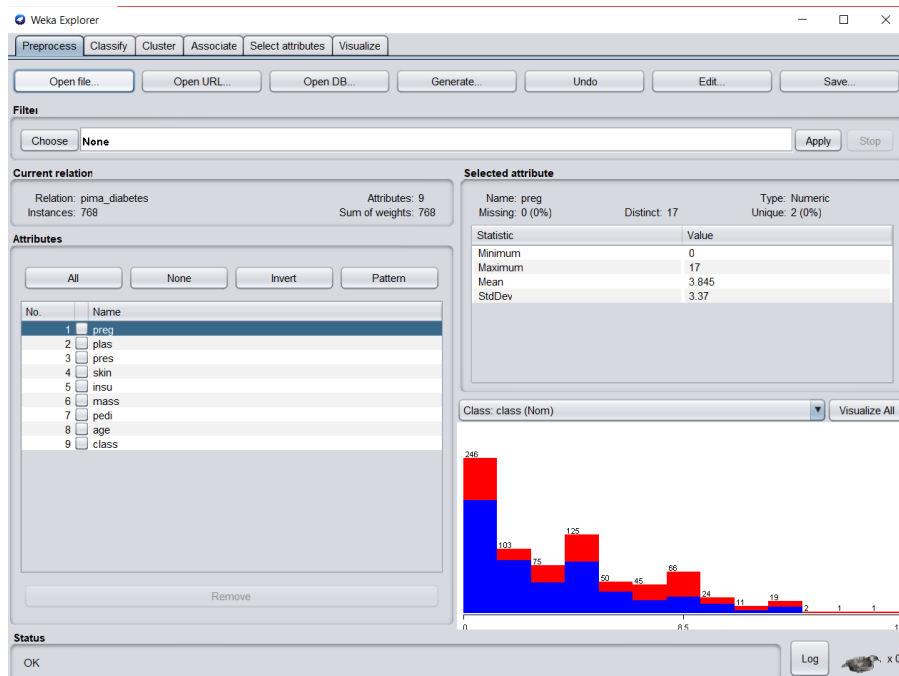
No. of instances: 768

No. of attributes: 9

Attribute information:

1) preg:

Type: Numeric, Distinct: 17, Unique: 2

Statistic	Value
Minimum	0
Maximum	17
Mean	3.845
StdDev	3.37

2) plas:

Type: Numeric, Distinct: 136, Unique: 19

Statistic	Value
Minimum	0
Maximum	199
Mean	120.895
StdDev	31.973

3) pres:

Type: Numeric, Distinct: 47, Unique: 8

Statistic	Value
Minimum	0
Maximum	122
Mean	69.105
StdDev	19.356

4) skin:

Type: Numeric, Distinct: 51, Unique: 5

Statistic	Value
Minimum	0
Maximum	99
Mean	20.536
StdDev	15.952

5) insu:

Type: Numeric, Distinct: 186, Unique: 93

Statistic	Value
Minimum	0
Maximum	846
Mean	79.799
StdDev	115.244

6) mass:

Type: Numeric, Distinct: 248, Unique: 76

Statistic	Value
Minimum	0
Maximum	67.1
Mean	31.993
StdDev	7.884

7) pedi:

Type: Numeric, Distinct: 517, Unique: 346

Statistic	Value
Minimum	0.078
Maximum	2.42
Mean	0.472
StdDev	0.331

8) age:

Type: Numeric, Distinct: 52, Unique: 5

Statistic	Value
Minimum	21
Maximum	81
Mean	33.241
StdDev	11.76

9) class:

Type: Nominal, Distinct: 2, Unique: 0

Label	Count
tested_negative	500
tested_positive	268

8. Glasses.arff
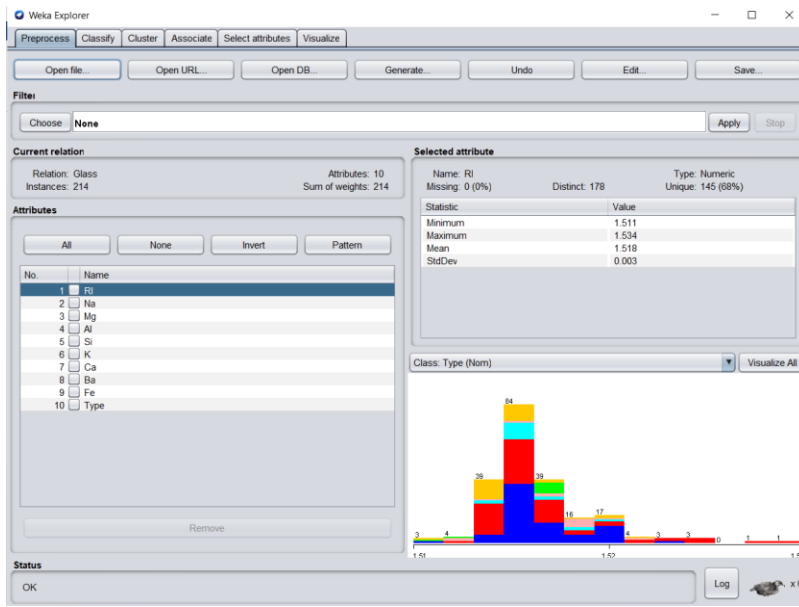
No. of instances: 214
No. of attributes: 10
Missing values: 0
Attribute information:

1) Rl:

Type: Numeric, Distinct: 178, Unique: 145

Statistic	Value
Minimum	1.511
Maximum	1.534
Mean	1.518
StdDev	0.003

2) Na:

Type: Numeric, Distinct: 142, Unique: 97

Statistic	Value
Minimum	10.73
Maximum	17.38
Mean	13.408
StdDev	0.817

3) Mg:

Type: Numeric, Distinct: 94, Unique: 62

Statistic	Value
Minimum	0
Maximum	4.49
Mean	2.685
StdDev	1.442

4) Al:

Type: Numeric, Distinct: 118, Unique: 64

Statistic	Value
Minimum	0.29
Maximum	3.5
Mean	1.445
StdDev	0.499

5) Si:

Type: Numeric, Distinct: 133, Unique: 78

Statistic	Value
Minimum	69.81
Maximum	75.41
Mean	72.651
StdDev	0.775

6) K:

Type: Numeric, Distinct: 65, Unique: 30

Statistic	Value
Minimum	0
Maximum	6.21
Mean	0.497
StdDev	0.652

7) Ca:

Type: Numeric, Distinct: 143, Unique: 95

Statistic	Value
Minimum	5.43
Maximum	16.19
Mean	8.957
StdDev	1.423

8) Ba:

Type: Numeric, Distinct: 34, Unique: 28

Statistic	Value
Minimum	0
Maximum	3.15
Mean	0.175
StdDev	0.497

9) Fe:

Type: Numeric, Distinct: 32, Unique: 17

Statistic	Value
Minimum	0
Maximum	0.51
Mean	0.057
StdDev	0.097

10) Type:

Type: Nominal, Distinct: 6, Unique: 0

Label	Count
build wind float	70
build wind non...	76
vehic wind float	17
vehic wind no...	0
containers	13
tableware	9
headlamps	29

9. Hypothyroid.arff
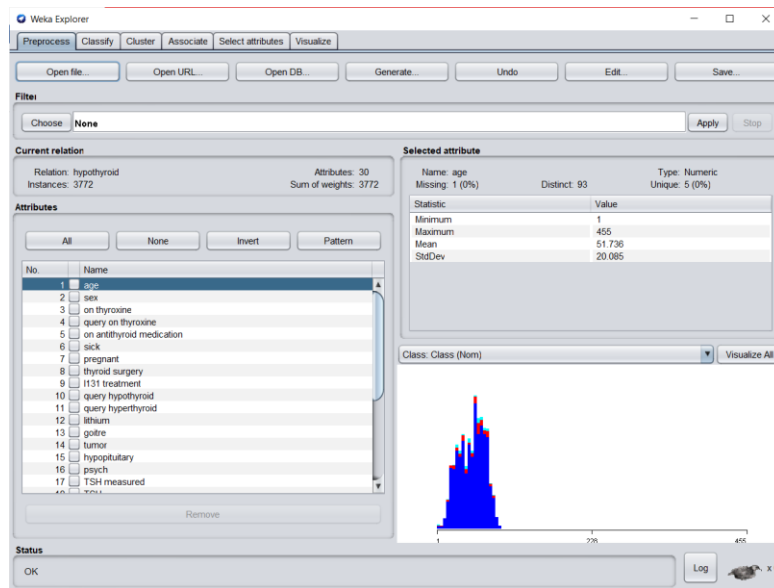
No. of instances: 3772
No. of attributes: 30
Attribute information:
'age' integer
'sex' { F, M}
'on thyroxine' { f, t}
'query on thyroxine' { f, t}
'on antithyroid medication' { f, t}
'sick' { f, t}
'pregnant' { f, t}
'thyroid surgery' { f, t}
'I131 treatment' { f, t}
'query hypothyroid' { f, t}
'query hyperthyroid' { f, t}
'lithium' { f, t}
'goitre' { f, t}
'tumor' { f, t}
'hypopituitary' { f, t}

'psych' { f, t}
'TSH measured' { t, f}
'TSH' real
'T3 measured' { t, f}
'T3' real
'TT4 measured' { t, f}
'TT4' real
'T4U measured' { t, f}
'T4U' real
'FTI measured' { t, f}
'FTI' real
'TBG measured' { f}
'TBG' real
'referral source' { SVHC, other, SVI, STMW, SVHD}
'Class' { negative, compensated_hypothyroid, primary_hypothyroid, secondary_hypothyroid}

10. Ionosphere.arff
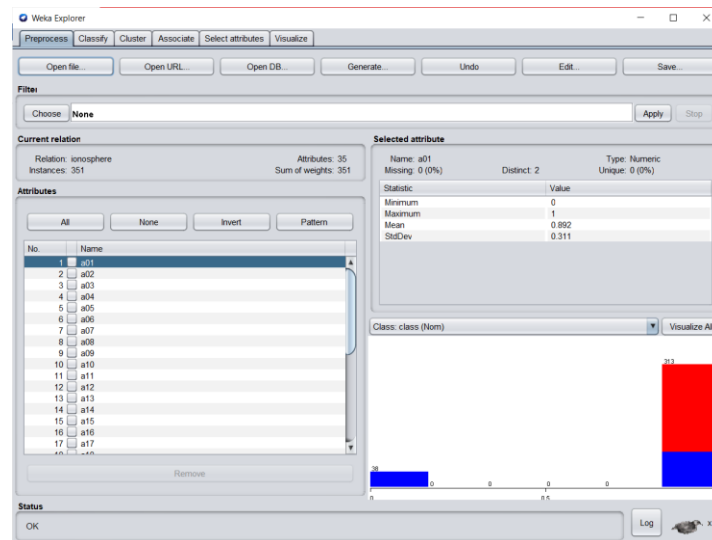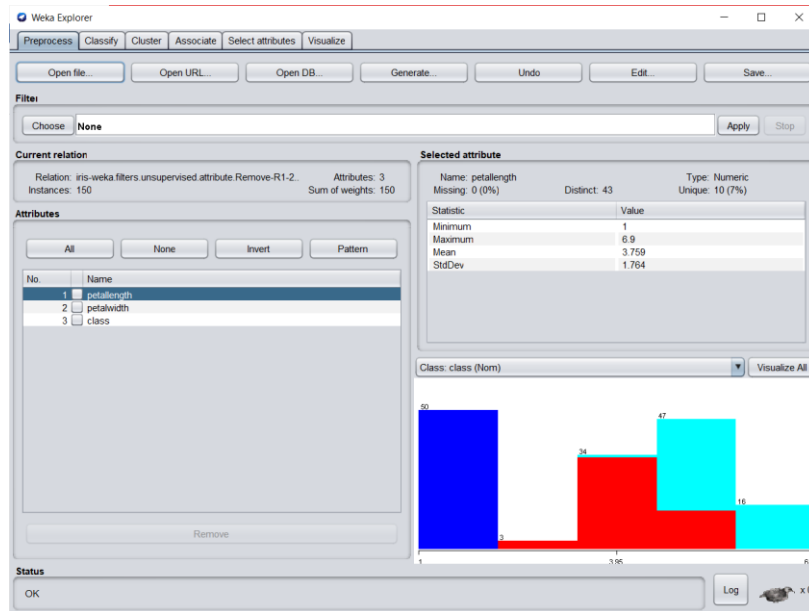

No. of instances: 351
No. of attributes: 35
Attribute information:
Type: Numeric (a01-a34), Nominal (class)
a01, a02, a03, a04, a05, a06, a07, a08, a09,a10, a11, a12, a13, a14, a15, a16, a17, a18, a19, 20, a22, a23, a24, a25, a26, a27, a28, a29, a30, a31, a32, a33, a34
class {b, g}

11. Iris.2D.arff

No. of instances: 150

No. of attributes: 3

Class objects: petallength, petalwidth, class

Missing values: 0

Attribute information:

1) Petallength:

Type: Numeric, Distinct: 43, Unique: 10

Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

2) Petalwidth:

Type: Numeric, Distinct: 22, Unique: 2

Statistic	Value
Minimum	0.1
Maximum	2.5
Mean	1.199
StdDev	0.763

3) Class:

Type: Nominal, Distinct: 3, Unique: 0

Label	Count
Iris-setosa	50
Iris-versicolor	50
Iris-virginica	50

12. Iris.arff
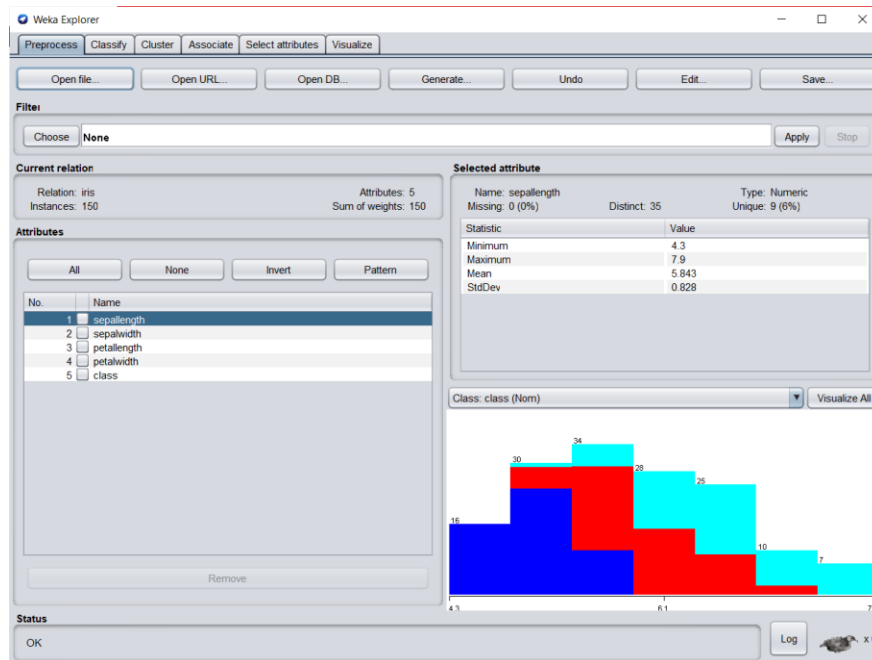
No. of instances: 150
No. of attributes: 5
Missing values: 0
Attribute Information:
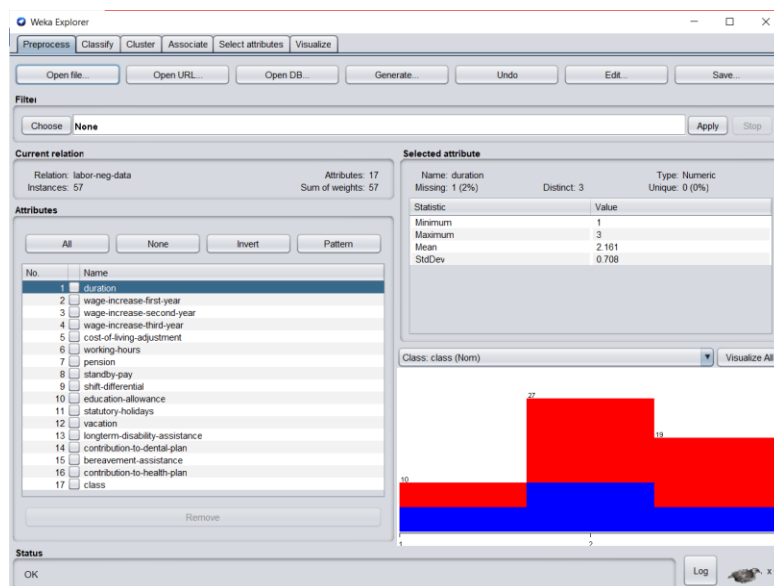1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
Iris Setosa, Iris Versicolour, Iris Virginica
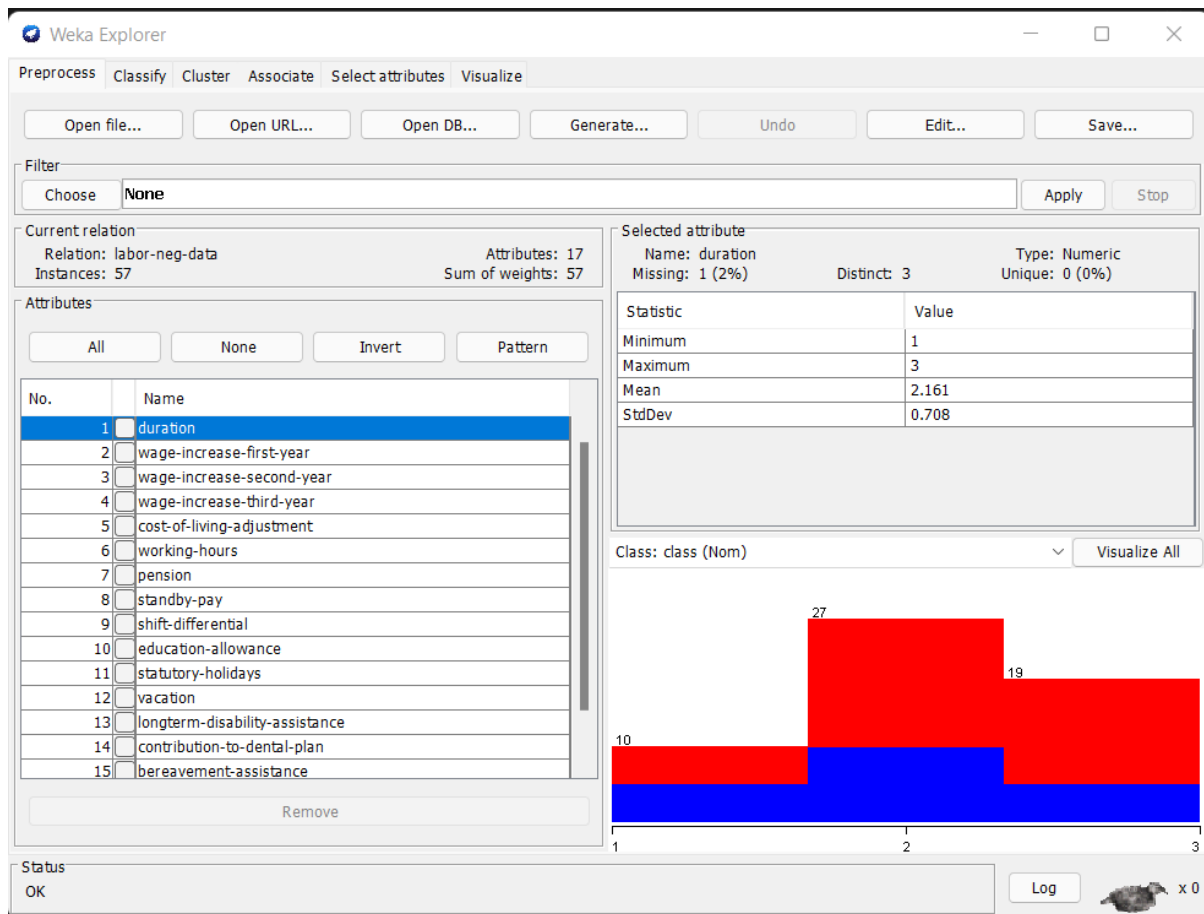
13. Labor.arff

No. of instances: 57
No. of attributes: 17
Missing values: 0
Attribute Information:
1. dur: duration of agreement   [1..7]
2 wage1.wage : wage increase in first year of contract [2.0 .. 7.0]
3 wage2.wage : wage increase in second year of contract [2.0 .. 7.0]
4 wage3.wage : wage increase in third year of contract [2.0 .. 7.0]
5 cola : cost of living allowance [none, tcf, tc]
6 hours.hrs : number of working hours during week [35 .. 40]
7 pension : employer contributions to pension plan [none, ret_allw, empl_contr]
8 stby_pay : standby pay [2 .. 25]
9 shift_diff : shift differencial : supplement for work on II and III shift [1 .. 25]
10 educ_allw.boolean : education allowance [true false]
11 holidays : number of statutory [9 .. 15]
12 vacation : number of paid vacation days [ba, avg, gnr]
13 lngtrm_disabil.boolean : employer's help during employee longterm disability [true , false]
14 dntl_ins : employers contribution towards the dental plan [none, half, full]
15 bereavement.boolean : employer's financial contribution towards the costs of bereavement [true , false]
16 empl_hplan : employer's contribution towards the health plan [none, half, full]

15) Laboror

Number of Instances: 57
Number of Attributes: 16
==
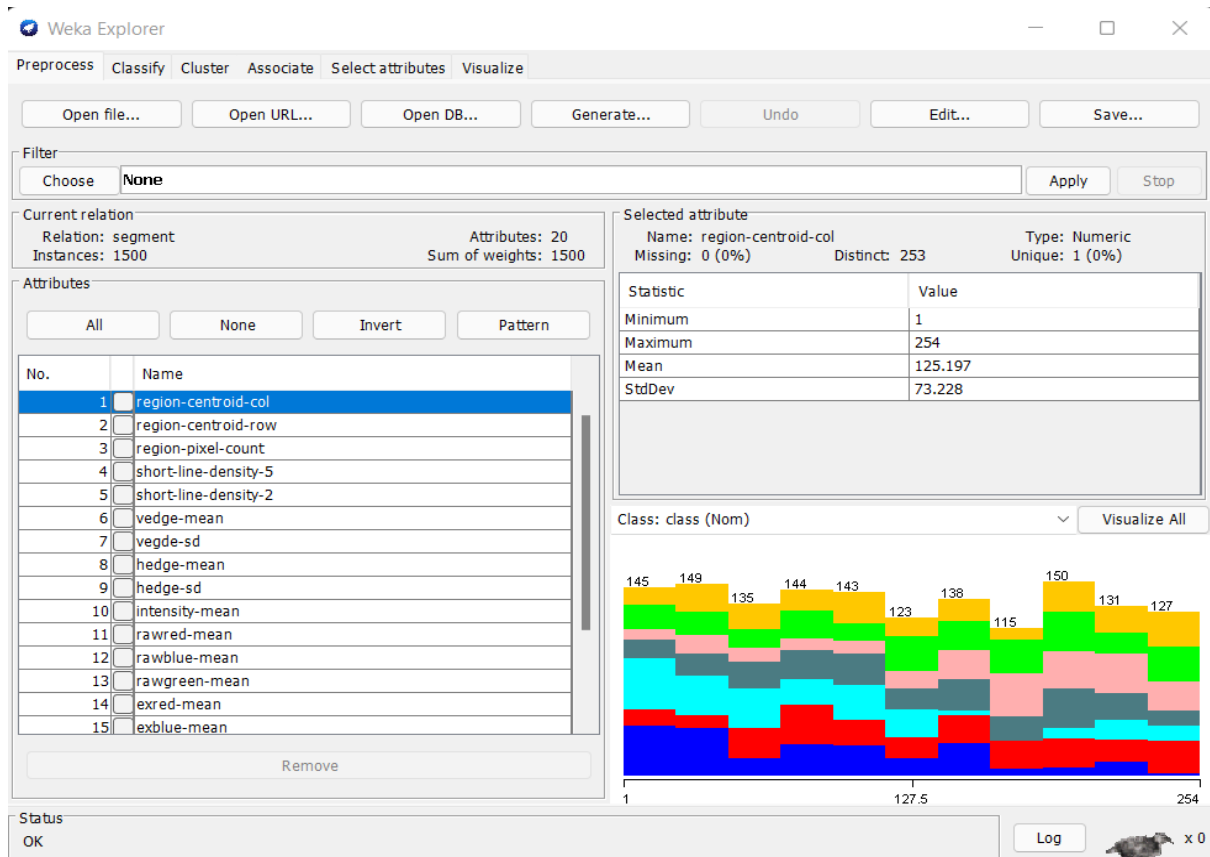
7. Attribute Information:
  1. dur: duration of agreement
     [1..7]
  2 wage1.wage : wage increase in first year of contract
     [2.0 .. 7.0]
  3 wage2.wage : wage increase in second year of contract
     [2.0 .. 7.0]
  4 wage3.wage : wage increase in third year of contract
     [2.0 .. 7.0]
  5 cola : cost of living allowance
     [none, tcf, tc]
  6 hours.hrs : number of working hours during week
     [35 .. 40]
  7 pension : employer contributions to pension plan
     [none, ret_allw, empl_contr]

8 stby_pay : standby pay
    [2 .. 25]
9 shift_diff : shift differencial : supplement for work on II and III shift
    [1 .. 25]
10 educ_allw.boolean : education allowance
    [true false]
11 holidays : number of statutory holidays
    [9 .. 15]
12 vacation : number of paid vacation days
    [ba, avg, gnr]
13 lngtrm_disabil.boolean :
    employer's help during employee longterm disabil
    ity [true , false]
14 dntl_ins : employers contribution towards the dental plan
    [none, half, full]
15 bereavement.boolean : employer's financial contribution towards the
    covering the costs of bereavement
    [true , false]
16 empl_hplan : employer's contribution towards the health plan
    [none, half, full]

16) segment-challenge

7 Attribute Information:

1   region-centroid-col: the column of the center pixel of the region.

2   region-centroid-row: the row of the center pixel of the region.

    region-pixel-count: the number of pixels in a region = 9.

    short-line-density-5: the results of a line extractoin algorithm that
    counts how many lines of length 5 (any orientation) with
    low contrast, less than or equal to 5, go through the region.

    short-line-density-2: same as short-line-density-5 but counts lines of
    high contrast, greater than 5.

    vedge-mean: measure the contrast of horizontally
    adjacent pixels in the region. There are 6, the mean and
    standard deviation are given. This attribute is used as
    a vertical edge detector.

7   vegde-sd: (see 6)

    hedge-mean: measures the contrast of vertically adjacent
    pixels. Used for horizontal line detection.

    hedge-sd: (see 8).

1   intensity-mean: the average over the region of (R + G + B)/3

11   rawred-mean: the average over the region of the R value.

12   rawblue-mean: the average over the region of the B value.

1   rawgreen-mean: the average over the region of the G value.

1   exred-mean: measure the excess red: (2R - (G + B))

1   exblue-mean: measure the excess blue: (2B - (G + R))

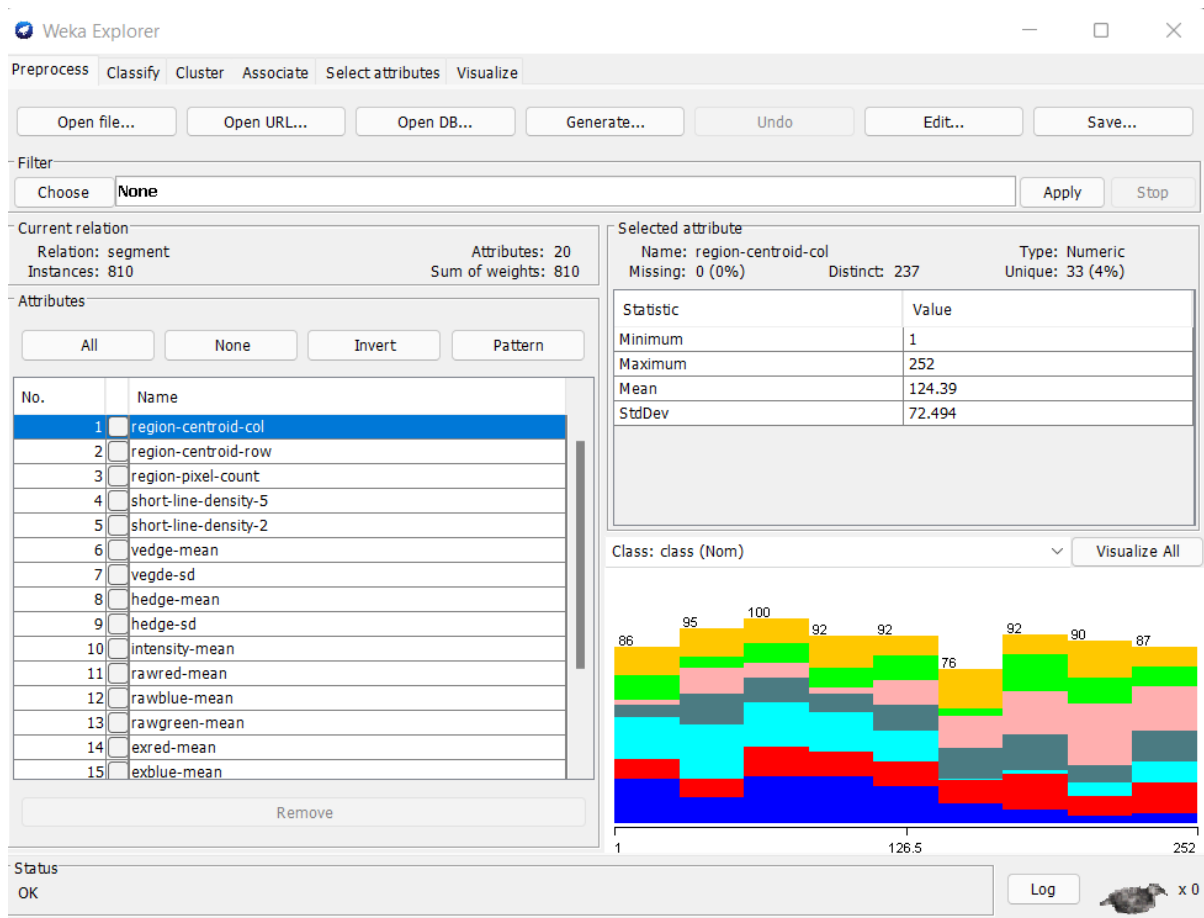1   exgreen-mean: measure the excess green: (2G - (R + B))

17   value-mean: 3-d nonlinear transformation
    of RGB. (Algorithm can be found in Foley and VanDam, Fundamentals
    of Interactive Computer Graphics)

1   saturatoin-mean: (see 17)

1   hue-mean: (see 17)

17) segment-test

Number of Instances: Training data: 210 Test data: 2100

Number of Attributes: 19 continuous attributes

==

Attribute Information:

1. region-centroid-col: the column of the center pixel of the region.

2. region-centroid-row: the row of the center pixel of the region.

3. region-pixel-count: the number of pixels in a region = 9.

4. short-line-density-5: the results of a line extractoin algorithm that

   counts how many lines of length 5 (any orientation) with

   low contrast, less than or equal to 5, go through the region.

5. short-line-density-2: same as short-line-density-5 but counts lines

   of high contrast, greater than 5.

6. vedge-mean: measure the contrast of horizontally

   adjacent pixels in the region. There are 6, the mean and

   standard deviation are given. This attribute is used as

   a vertical edge detector.

7. vegde-sd: (see 6)

8. hedge-mean: measures the contrast of vertically adjacent

   pixels. Used for horizontal line detection.

9. hedge-sd: (see 8).

10. intensity-mean: the average over the region of (R + G + B)/3

11. rawred-mean: the average over the region of the R value.

12. rawblue-mean: the average over the region of the B value.

13. rawgreen-mean: the average over the region of the G value.

14. exred-mean: measure the excess red: (2R - (G + B))

15. exblue-mean: measure the excess blue: (2B - (G + R))

16. exgreen-mean: measure the excess green: (2G - (R + B))

17. value-mean: 3-d nonlinear transformation

    of RGB. (Algorithm can be found in Foley and VanDam, Fundamentals

    of Interactive Computer Graphics)

18. saturatoin-mean: (see 17)

19. hue-mean: (see 17)

18) soybean
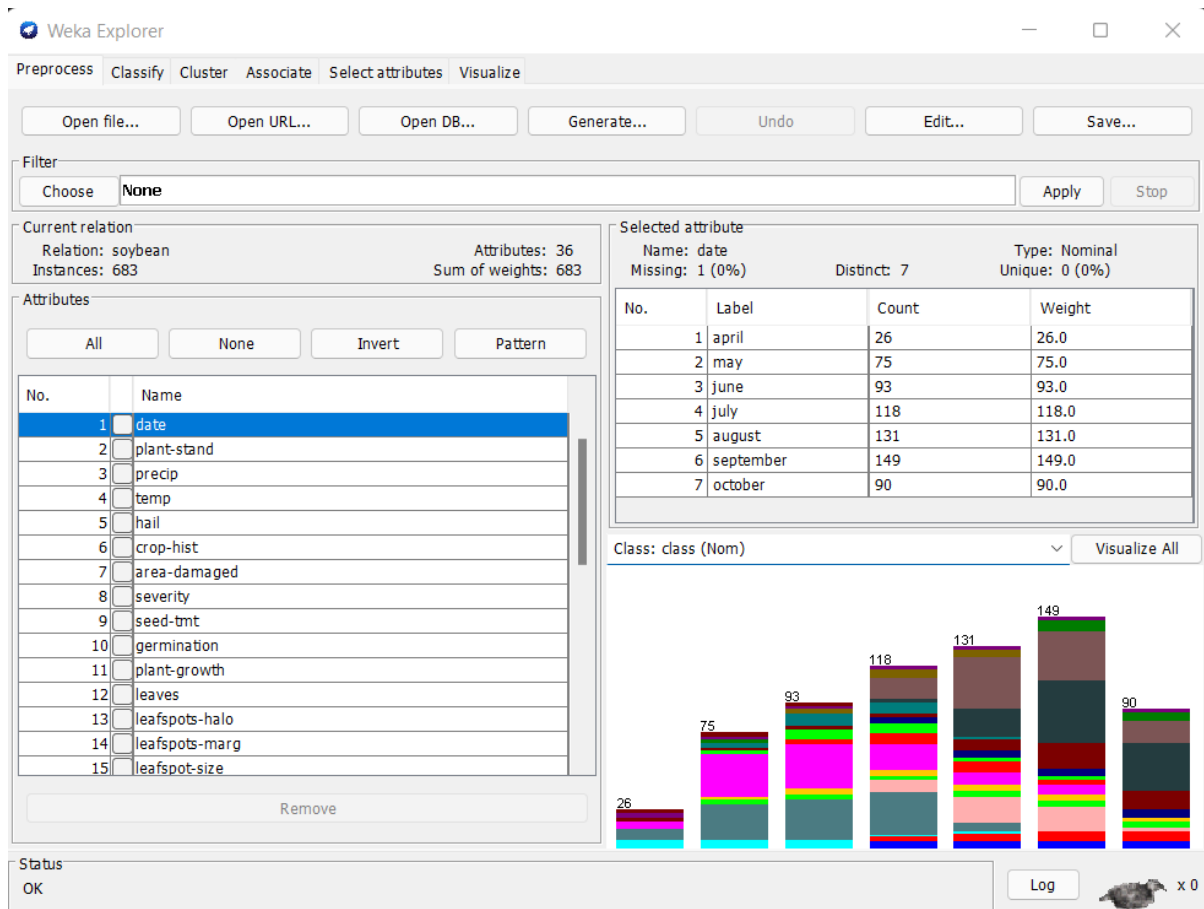
Number of Instances: 683

Number of Attributes: 35

==

|	
date	{april,may,june,july,august,september,october}
plant-stand	{normal,lt-normal}
precip	{lt-norm,norm,gt-norm}
temp	{lt-norm,norm,gt-norm}
hail	{yes,no}
crop-hist	{diff-lst-year,same-lst-yr,same-lst-two-yrs, same-lst-sev-yrs}

area-damaged	{scattered,low-areas,upper-areas,whole-field}
severity	{minor,pot-severe,severe}
seed-tmt	{none,fungicide,other}
germination	{90-100,80-89,lt-80}
plant-growth	{norm,abnorm}
leaves	{norm,abnorm}
leafspots-halo	{absent,yellow-halos,no-yellow-halos}
leafspots-marg	{w-s-marg,no-w-s-marg,dna}
leafspot-size	{lt-1/8,gt-1/8,dna}
leaf-shread	{absent,present}
leaf-malf	{absent,present}
leaf-mild	{absent,upper-surf,lower-surf}
stem	{norm,abnorm}
lodging	{yes,no}
stem-cankers	{absent,below-soil,above-soil,above-sec-nde}
canker-lesion	{dna,brown,dk-brown-blk,tan}
fruiting-bodies	{absent,present}
external-decay	{absent,firm-and-dry,watery}
mycelium	{absent,present}
int-discolor	{none,brown,black}
sclerotia	{absent,present}
fruit-pods	{norm,diseased,few-present,dna}
fruit-spots	{absent,colored,brown-w/blk-specks,distort,dna}
seed	{norm,abnorm}
mold-growth	{absent,present}
seed-discolor	{absent,present}
seed-size	{norm,lt-norm}
shriveling	{absent,present}
roots	{norm,rotted,galls-cysts}

class	{diaporthe-stem-canker, charcoal-rot, rhizoctonia-root-rot, phytophthora-rot, brown-stem-rot, powdery-mildew, downy-mildew, brown-spot, bacterial-blight, bacterial-pustule,

purple-seed-stain, anthracnose, phyllosticta-leaf-spot, alternarialeaf-spot, frog-eye-leaf-spot, diaporthe-pod-&-stem-blight, cyst-nematode, 2-4-d-injury, herbicide-injury}

19) unbalanced
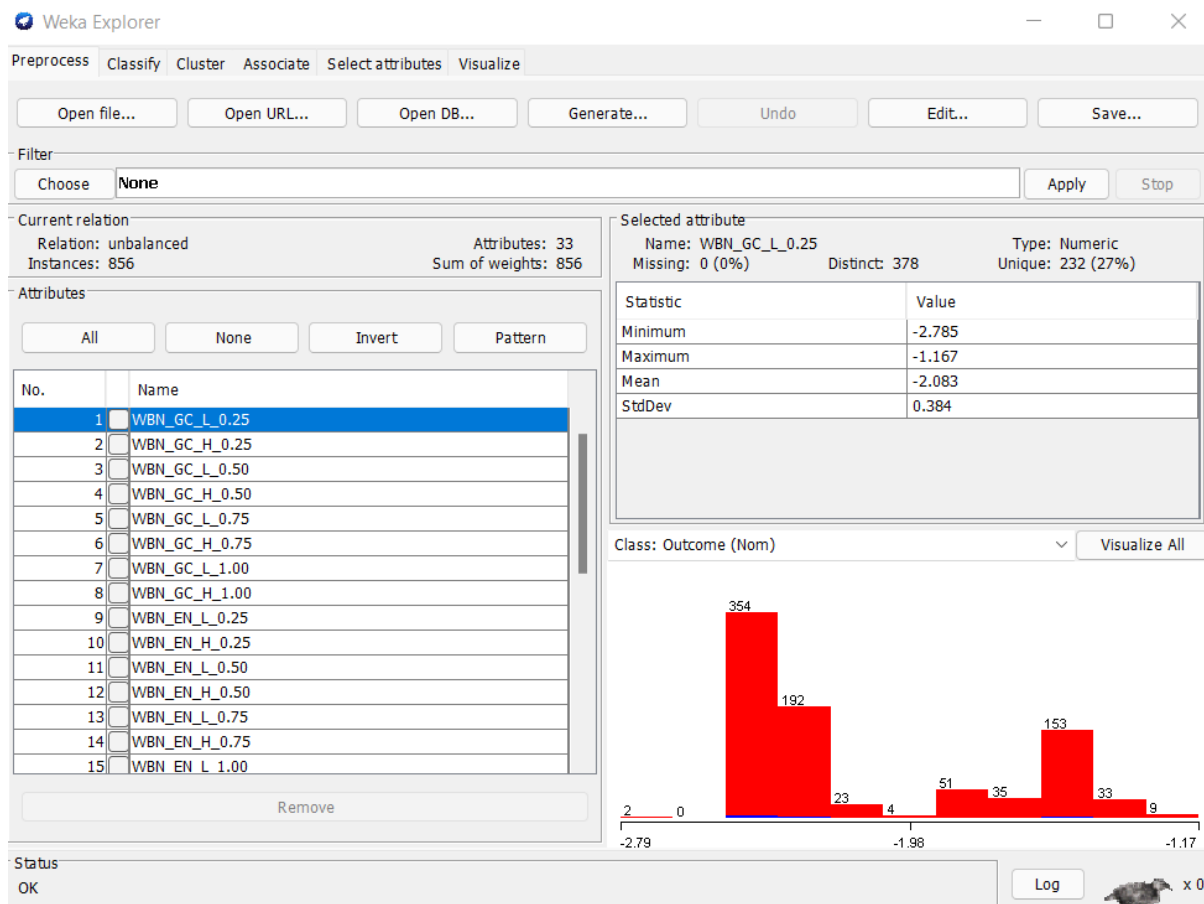
WBN_GC_L_0.25 numeric

WBN_GC_H_0.25 numeric

WBN_GC_L_0.50 numeric

WBN_GC_H_0.50 numeric

WBN_GC_L_0.75 numeric

WBN_GC_H_0.75 numeric

WBN_GC_L_1.00 numeric

WBN_GC_H_1.00 numeric

WBN_EN_L_0.25 numeric

WBN_EN_H_0.25 numeric

WBN_EN_L_0.50 numeric

WBN_EN_H_0.50 numeric

WBN_EN_L_0.75 numeric

WBN_EN_H_0.75 numeric

WBN_EN_L_1.00 numeric

WBN_EN_H_1.00 numeric

WBN_LP_L_0.25 numeric

WBN_LP_H_0.25 numeric

WBN_LP_L_0.50 numeric

WBN_LP_H_0.50 numeric

WBN_LP_L_0.75 numeric

WBN_LP_H_0.75 numeric

WBN_LP_L_1.00 numeric

WBN_LP_H_1.00 numeric

XLogP numeric

PSA numeric

NumRot numeric

NumHBA numeric

NumHBD numeric

MW numeric

BBB numeric

BadGroup numeric

Outcome {Active,Inactive}

20) vote

Selected attribute

Name: handicapped-infants   Type: Nominal
Missing: 12 (3%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	236	236.0
2	y	187	187.0

Selected attribute

Name: water-project-cost-sharing   Type: Nominal
Missing: 48 (11%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	192	192.0
2	y	195	195.0

Selected attribute

Name: adoption-of-the-budget-resolution   Type: Nominal
Missing: 11 (3%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	171	171.0
2	y	253	253.0

Selected attribute

Name: physician-fee-freeze   Type: Nominal
Missing: 11 (3%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	247	247.0
2	y	177	177.0

Selected attribute

Name: el-salvador-aid   Type: Nominal
Missing: 15 (3%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	208	208.0
2	y	212	212.0

Selected attribute

Name: religious-groups-in-schools   Type: Nominal
Missing: 11 (3%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	152	152.0
2	y	272	272.0

Selected attribute

Name: anti-satellite-test-ban   Type: Nominal
Missing: 14 (3%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	182	182.0
2	y	239	239.0

Selected attribute

Name: aid-to-nicaraguan-contras   Type: Nominal
Missing: 15 (3%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	178	178.0
2	y	242	242.0

Selected attribute

Name: mx-missile   Type: Nominal
Missing: 22 (5%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	206	206.0
2	y	207	207.0

Selected attribute

Name: immigration   Type: Nominal
Missing: 7 (2%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	212	212.0
2	y	216	216.0

Selected attribute

Name: synfuels-corporation-cutback   Type: Nominal
Missing: 21 (5%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	264	264.0
2	y	150	150.0

Selected attribute

Name: education-spending   Type: Nominal
Missing: 31 (7%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	233	233.0
2	y	171	171.0

Selected attribute

Name: superfund-right-to-sue   Type: Nominal
Missing: 25 (6%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	201	201.0
2	y	209	209.0

Selected attribute

Name: crime   Type: Nominal
Missing: 17 (4%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	170	170.0
2	y	248	248.0

Selected attribute

Name: duty-free-exports   Type: Nominal
Missing: 28 (6%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	233	233.0
2	y	174	174.0

Selected attribute

Name: export-administration-act-south-africa   Type: Nominal
Missing: 104 (24%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	n	62	62.0
2	y	269	269.0

Selected attribute

Name: Class   Type: Nominal
Missing: 0 (0%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	democrat	267	267.0
2	republican	168	168.0

21) weather.nominal

Selected attribute

Name: outlook
Missing: 0 (0%)   Distinct: 3   Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

Selected attribute

Name: temperature
Missing: 0 (0%)   Distinct: 3   Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	hot	4	4.0
2	mild	6	6.0
3	cool	4	4.0

Selected attribute

Name: humidity
Missing: 0 (0%)   Distinct: 2   Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	high	7	7.0
2	normal	7	7.0

Selected attribute

Name: windy
Missing: 0 (0%)   Distinct: 2   Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	TRUE	6	6.0
2	FALSE	8	8.0

Selected attribute

Name: play
Missing: 0 (0%)   Distinct: 2   Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	yes	9	9.0
2	no	5	5.0

22) weather.numeric

Selected attribute

Name: outlook   Type: Nominal
Missing: 0 (0%)   Distinct: 3   Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

Selected attribute

Name: temperature   Type: Numeric
Missing: 0 (0%)   Distinct: 12   Unique: 10 (71%)

Statistic	Value
Minimum	64
Maximum	85
Mean	73.571
StdDev	6.572

Selected attribute

Name: humidity   Type: Numeric
Missing: 0 (0%)   Distinct: 10   Unique: 7 (50%)

Statistic	Value
Minimum	65
Maximum	96
Mean	81.643
StdDev	10.285

Selected attribute

Name: windy   Type: Nominal
Missing: 0 (0%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	TRUE	6	6.0
2	FALSE	8	8.0

Selected attribute

Name: play   Type: Nominal
Missing: 0 (0%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	yes	9	9.0
2	no	5	5.0