

```

1  local help=[
2  LOOK.LUA: landscape analysis
3  (c) 2022 Tim Menzies, tim@ieee.org, BSD-2 license
4  *I think the highest and lowest points are the important ones.
5  Anything else is just... in between." ~Jim Morrison
6
7  INSTALL: requires: lua 5.4+
8             download: lib.lua, look.lua, looked.lua
9             test      : lua egs.lua -h
10
11  USAGE: lua looked.lua [OPTIONS]
12
13             defaults
14             -----
15  --also      -a size of rest=best*also = 4
16  --p         -p distance coefficient   = 2
17  --far       -f far                    = .95
18  --Some      -S sample size           = 256
19  --seed      -s random number seed    = 10019
20  --min       -m min size pass1        = .5
21  --Min       -M min size pass2        = 10
22
23  --file      -f csv file with data    = ../etc/data/auto93.csv
24  --help      -h show help              = false
25  --verbose   -v verbose mode          = false
26  --go        -g start up action        = nothing]]
27
28 local _ = require"hb"
29 local big,is,push,shuffle,tothing = _.big, _.is, _.push, _.shuffle, _.tothing
30 local the={}
31 help:gsub("[^-]|([%s]+)|(%n)%s([%s]+)",function(k,x) the[k]=_tothing(x)end)
32 -----
33 local ROWS="ROW"
34 function ROW.new(i,of,cells) i.cells, i.of, i.evaluated = cells,of,0 end
35 function ROW.better(i,j, n,s1,s2,v1,v2)
36     n,s1,s2 = 0,0,0
37     for _,_ in pairs(i.of.ys) do n = n + 1 end
38     for c,w in pairs(i.of.ys) do
39         v1,v2 = i.of.norm(c, r1[c]), i.of.norm(c, r2[c])
40         s1 = s1 - 2.7183^(w * (v1 - v2) / n)
41         s2 = s2 - 2.7183^(w * (v2 - v1) / n) end
42     return s1/n < s2/n end
43
44 function ROW.dist(i,j, d,n,dist1)
45     function dist1(c,v1,v2)
46         if v1=="?" and v2=="?" then return 0 end
47         if not i.of.nums[c]
48             then return v1==v2 and 0 or 1
49             else if v1=="?" then v2=i.of.norm(c,v2); v1= v2<.5 and 1 or 0
50                   elseif v2=="?" then v1=i.of.norm(c,v1); v2= v1<.5 and 1 or 0
51                   else v1,v2 = i.of.norm(c,v1), i.of.norm(c,v2) end
52             return math.abs(v1-v2) end
53     end
54     d,n = 0,0
55     for c,_ in pairs(i.xs) do n,d = n+1, d + (dist1(c,i[c], j[c]))^the.p end
56     return (d/n)^(1/the.p) end
57 -----
58 local ROWS="ROWS"
59 local function num(s) return s:find"^[A-Z]" end
60 local function goal(s) return s:find"[4-5]" end
61 local function wght(s) return s:find"-$" and -1 or 1 end
62
63 function ROWS.new(i,src)
64     i.rows, i.nums, i.xs, i.ys, i.names = {}, {}, {}, {}, nil
65     if type(src)=="table" then for _,r in pairs(src) do i:add(r) end
66     else for r in csv(src) do i:add(r) end end end
67
68 function ROWS.add(i,t)
69     if i.names
70     then r = t.cells and r or ROW(i,t); i:update(r.cells); push(i.rows, r)
71     else i:header(r) end end
72
73 function ROWS.header(i,r)
74     i.names = r
75     for c,s in pairs(r) do if num(s) then i.nums[c]={lo=big,hi=-big} end end
76     for c,s in pairs(r) do if goal(s) then i.ys[c]=wght(s) else i.xs[c]=c end end end
77
78 function ROWS.update(i,t, v)
79     for c,num in pairs(i.nums) do
80         v = t[c]
81         if v == "?" then num.lo = math.min(v, num.lo)
82                     num.hi = math.max(v, num.hi) end end end
83
84 function ROWS.norm(i,c,v, lo,hi)
85     lo,hi = i.nums[c].lo, i.nums[c].hi
86     return (v=="?" and v) or ((hi-lo) < 1E-9 and 0) or (v-lo)/(hi-lo) end
87
88 function ROWS.around(i,r1,t, fun)
89     function fun(r2) return (dist=dist(r1,r2), row=r2) end
90     return sort(map(t or i.rows,fun),lt"dist") end
91
92 function ROWS.far(i,r1,t, tmp)
93     tmp= i:around(r1,t)
94     return tmp[#tmp*the.far//1].row end
95
96 function ROWS.look(i, w,sample,ra,rest)
97     w = shuffle(i.rows)
98     sample = many(w, the.Some)
99     ra = i:far(any(sample), sample)
100     rest = {}
101     for _,stop in pairs(({(#w)^the.min, the.Min)) do
102         while #w > stop do
103             local rb = far(ra, sample)
104             if rb:better(ra) then ra,rb = rb,ra end
105             ra.evaluated, rb.evaluated = true,true
106             local c = ra:dist(rb)
107             for _,rc in pairs(w) do rc.x=(rc:dist(ra)^2 +c^2- rc:dist(rb)^2)/(2*c) end
108             local best = {}
109             for n,rc in pairs(sort(w,lt"x")) do push(n<=#w/2 and best or rest,rc) end
110             if #best==#w then break else w=best end
111             sample = many(w,the.Some) end end
112     return ra,w,many(rest, #w*the.also) end
113
114 return (ROWS=ROWS, ROW=ROW, help=help, the=the)
115
116 -- i=self
117 -- j=other
118 -- k=loop counter
119 -- v= cell value
120 -- c = column index

```