

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210

```



```

local _ = require"cli"
local any = _,any
local cat,cli,coerce,copy,csv = _cat, _cli, _coerce, _copy, _csv
local lines,many,obj,push,rogues = _lines, _many, _obj,_push, _rogues
local words = _words

local rand = math.random

local Cols, Data, Num, Sym = obj"Cols", obj"Data", obj"Num", obj"Sym"
local the={ratios=256, bins=8, seed=10019, some=512}

-- Num -----
function Num:new(at,txt)
  txt = txt or ""
  return {n=0,at=at or 0, txt=txt, ready=false, has={},
    hi=-math.huge, lo= math.huge, w=txt:find"$" and -1 or 1} end

function Num:add(x)
  if x ~= "?" then
    local pos
    self.n = self.n + 1
    self.lo = math.min(x, self.lo)
    self.hi = math.max(x, self.hi)
    if #self.has < the.ratios then pos = 1 + (#self.has)
    elseif rand() < the.ratios/self.n then pos = rand(#self.has) end
    if pos then self.ready=false
    if self.has[pos]=x end end end

function Num:discretize(x)
  local b = (self.hi - self.lo)/the.bins
  return self.hi==self.lo and 1 or math.floor(x/b+.5)*b end

function Num:dist(x,y)
  if x=="?" and y=="?" then return 1 end
  if x=="?" then y=self:norm(y); x=y<.5 and 1 or 0
  elseif y=="?" then x=self:norm(x); y=x<.5 and 1 or 0
  else x,y = self:norm(x), self:norm(y) end
  return math.abs(x-y) end

function Num:holds()
  if not self.ready then table.sort(self.has); self.ready=true end
  return self.has end

function Num:norm(num)
  return self.hi - self.lo < 1E-9 and 0 or (num-self.lo)/(self.hi-self.lo) end

-- Sym -----
function Sym:new(at,txt)
  return {n=0,at=at or 0, txt=txt or "", ready=false, has={}} end

function Sym:add(x)
  if x ~= "?" then
    self.n = self.n + 1
    self.has[x] = 1+(self.has[x] or 0) end end

function Sym:discretize(x) return x end
function Sym:dist(x,y)
  return (x=="?" or y=="?") and 1 or x==y and 0 or 1 end

-- Row -----
function Row:new(cells) return {cells=cells, cooked=copy(cells)} end

-- Cols -----
function Cols:new(names)
  self.names, self.x, self.y, self.all= names, {}, {}, {}
  for at,txt in pairs(names) do
    local what = txt:find"[A-Z]" and Num or Sym
    local col = push(self.all, what(at,txt))
    if not txt:find"$" then
      push(txt:find"[+-]$" and self.y or self.x, col) end end end

-- Data -----
local function rows(src)
  if type(src) == "table" then return src else
  local u={} ; csv(src, function(t) push(u,t) end); return u end end

function Data:new(rows)
  self.rows, self.cols = {},{}
  for i,row in pairs(rows) do
    if i==1
    then self.cols = Cols(row)
    else push(self.rows, Row(row))
    for cols in pairs(self.cols.x, self.cols.y) do
      for col in pairs(cols) do col:add(row[col.at]) end end end end
  for cols in pairs(self.cols.x, self.cols.y) do
    for _,row in pairs(self.rows) do
      row.cooked[col.at] = col:discretize(row.cells[col.at]) end end end

function Data:dist(row1,row2)
function Data:around(row1, rows)
  return sort(map(rows, function(row2) return {row=row2,d = row1-row2} end),--#
    lt"d") end
function Data:far(XXX) end

function Data:half(rows, above, all)
  local all = all or self.rows
  local some = many(all, the.some)
  local left = above or far(any(some), some)
  -- (defmethod half ((i rows) (optional all above)
  --   "Split rows in two by their distance to two remove points."
  --   (let* ((all (or (all ? i .has)))
  --         (some (many all (! my some)))
  --         (left (or above (far (any some) some)))
  --         (right (far left some))
  --         (c- (dists left right))
  --         (n 0) lefts rights)
  --     (labels ((project (row)
  --       (let ((a (dists row left))

```