```lua
local b4={}; for k,v in pairs(_ENV) do b4[k]=v end -- LUA trivia. Ignore.
local help=[[
CSV : summarized csv file
(c) 2022 Tim Menzies <timm@ieee.org> BSD-2 license

USAGE: lua seen.lua [OPTIONS]

OPTIONS:
 -e  --eg        start-up example                      = nothing
 -d  --dump      on test failure, exit with stack dump = false
 -f  --file      file with csv data                    = ../data/auto93.csv
 -h  --help      show help                             = false
 -n  --nums      number of nums to keep                = 512
 -s  --seed      random number seed                    = 10019
 -S  --seperator feild seperator                       = ,]]

-- Function argument conventions:
-- 1. two blanks denote optionas, four blanls denote locals:
-- 2. prefix n,s,is,fun denotes number,string,bool,function;
-- 3. suffix s means list of thing (so names is list of strings)
-- 4. c is a column index (usually)

-- ## Misc routines
-- ### Handle Settings
local the,coerce,cli
-- Parse `the` config settings from `help`.
function coerce(s,     fun)
  function fun(s1)
    if s1=="true"  then return true end
    if s1=="false" then return false end
    return s1 end
  return math.tointeger(s) or tonumber(s) or coerce1(s:match"^%s*(.-)%s*$") end

-- Create a `the` variables
the={}
help:gsub("\n [-][%S]+[%s]+[-][-]([%S]+)[^\n]+= ([%S]+)",
          function(k,x) the[k]=coerce(x) end)

-- Update settings from values on command-line flags. Booleans need no values
-- (we just flip the defeaults).
function cli(t)
  for slot,v in pairs(t) do
    v = tostring(v)
    for n,x in ipairs(arg) do
      if x=="-"..(slot:sub(1,1)) or x=="--"..slot then
        v = v=="false" and "true" or v=="true" and "false" or arg[n+1] end end
    t[slot] = coerce(v) end
  if t.help then os.exit(print("\n"..help.."\n")) end
  return t end

-- ### Linting code
local rogues
-- Find rogue locals.
function rogues()
  for k,v in pairs(_ENV) do if not b4[k] then print("?",k,type(v)) end end end

-- ### lists
local copy,per,push,csv
-- deepcopy
function copy(t,    u)
  if type(t) ~= "table" then return t end
  u={}; for k,v in pairs(t) do u[k] = copy(v) end
  return setmetatable(u,getmetatable(t))  end

-- Return the `p`-th thing from the sorted list `t`.
function per(t,p)
  p=math.floor(((p or .5)*#t)+.5); return t[math.max(1,math.min(#t,p))] end

-- Add to `t`, return `x`.
function push(t,x) t[1+#t]=x; return x end

-- ## Call `fun` on each row. Row cells are divided in `the.seperator`.
function csv(fname,fun,     sep,src,s,t)
  sep = "([^" .. the.seperator .. "]+)"
  src = io.input(fname)
  while true do
    s = io.read()
    if not s then return io.close(src) else
      t={}
      for s1 in s:gmatch(sep) do t[1+#t] = coerce(s1) end
      fun(t) end end end

-- ### Strings
local o,oo
-- `o` is a telescopt and `oo` are some binoculars we use to exam stucts.
-- `o`:  generates a string from a nested table.
function o(t,   show,u)
  if type(t) ~=  "table" then return tostring(t) end
  function show(k,v)
    if not tostring(k):find"^_"  then
      v = o(v)
      return #t==0 and string.format(":%s %s",k,v) or tostring(v) end end
  u={}; for k,v in pairs(t) do u[1+#u] = show(k,v) end
  if #t==0 then table.sort(u) end
  return "{"..table.concat(u," ").."}" end

-- `oo`: prints the string from `o`.
function oo(t) print(o(t)) return t end

-- ### OO
local ako,obj
-- obj("Thing") enables a constructor Thing:new() ... and a pretty-printer
-- for Things.
ako=setmetatable
function obj(name,    t)
  t={}
  t.__index,t.__tostring = t, function(x) return name .. o(x) end
  return ako(t,{__call=function(k,...)
                       x=ako({},k); return ako(x.new(t,...) or t,x) end}) end
```



```lua
-- ------------------------------------
-- ## Objects
Cols,Data,Num,Rows,Sym=obj"Cols", obj"Data", obj"Num", obj"Rows", obj"Sym"
-- `Data` is a holder of `rows` and their sumamries (in `cols`).
function Data:new() return { cols= nil,  -- summaries of data
                             rows= {}    -- kept data
                           } end

-- `Columns` Holds of summaries of columns.
-- Columns are created once, then may appear in  multiple slots.
function Cols:new() return {
  names={},  -- all column names
  all={},    -- all the columns (including the skipped ones)
  klass=nil, -- the single dependent klass column (if it exists)
  x={},      -- independent columns (that are not skipped)
  y={}       -- depedent columns (that are not skipped)
  } end

-- `Sym`s summarize a stream of symbols.
function Sym:new(c,s)
  return {n=0,          -- items seen
          at=c or 0,    -- column position
          name=s or "", -- column name
          _has={}       -- kept data
         } end

-- `Num` ummarizes a stream of numbers.
function Num(c,s)
  return {n=0,at=c or 0, name=s or "", _has={}, -- as per Sym
          lo= math.huge,   -- lowest seen
          hi= -math.huge,  -- highest seen
          isSorted=true,    -- no updates since last sort of data
          w = ((s or ""):find"-$" and -1 or 1)
         } end

-- `Row` holds one record
function Row(t) return {cells=t,           -- one record
                        cooked=copy(t), -- used if we discretize data
                        isEvaled=false    -- true if y-values evaluated.
                       } end
-- ------------------------------------
-- ## Columns
-- Add one thing to `col`. For Num, keep at most `nums` items.
function Sym:add(v)
  if v~="?" then
   self.n=self.n+1; self._has[v] = 1 + (self._has[v] or 0) end end

-- Reservoir sampler. Keep at most `the.nums` numbers
-- (and if we run out of room, delete something old, at random).,
function Num:add(col,v,     pos)
  if v~="?" then
    self.n=self.n+1
    self.lo = math.min(v, self.lo)
    self.hi = math.max(v, self.hi)
    if      #self._has < the.nums           then pos = 1 + (#self._has)
    elseif math.random() < the.nums/col.n then pos = math.random(#self._has) end
    if pos then self.isSorted = false
                self._has[pos] = tonumber(v) end end end

--- Add a `row` to `data`. Calls `add()` to  updatie the `cols` with new values.
function Data:add(xs)
  local row= push(data.rows, xs.cells and xs or Row(xs)) -- ensure xs is a Row
  for _,todo in pairs{data.cols.x, data.cols.y} do
    for _,col in pairs(todo) do
      col:add(row.cells[col.at]) end end end

function Cols:new(names,     col)
  self.names = namess
  for c,s in pairs(names) do
    local col = push(self.all, -- Numerics start with Uppercase.
                     (s:find"^[A-Z]*" and Num or Sym)(c,s))
    if not s:find":$" then -- some columns are skipped
      push(s:find"[!+-]" and self.y or self.x, col) -- some cols are goal cols
      if s:find"!$" then self.klass=col end end end end

-- else read rows from a `src`  table. When reading, use row1 to define columns.
-- Generate rows from some `src`.  If `src` is a string, read rows from file;
function records(src,       data,head,body)

  function body(t) -- treat first row differently (defines the columns)
    if data.cols then record(data,t) else data.cols=head(t) end
  end ----------
  data =  Data()
  if type(src)=="string" then csv(src, body) else
    for _,t in pairs(src or {}) do body(t) end end
  return data end

-- ### Query
-- Return kept numbers, sorted.
local function nums(num)
  if not num.isSorted then table.sort(num._has); num.isSorted=true end
  return num._has end

-- Diversity (standard deviation for Nums, entropy for Syms)
local function div(col)
  if  col.isNum then local a=nums(col); return (per(a,.9)-per(a,.1))/2.58 else
    local function fun(p) return p*math.log(p,2) end
    local e=0
    for _,n in pairs(col._has) do if n>0 then e=e-fun(n/col.n) end end
    return e end end

-- Central tendancy (median for Nums, mode for Syms)
local function mid(col)
  if col.isNum then return per(nums(col),.5) else
    local most,mode = -1
    for k,v in pairs(col._has) do if v>most then mode,most=k,v end end
    return mode end end

-- Diversity (standard deviation for Nums, entropy for Syms)
local function div(col)
  if  col.isNum then local a=nums(col); return (per(a,.9)-per(a,.1))/2.58 else
    local function fun(p) return p*math.log(p,2) end
    local e=0
    for _,n in pairs(col._has) do if n>0 then e=e-fun(n/col.n) end end
    return e end end

-- For `showCols` (default=`data.cols.x`) in `data`, report `fun` (default=`mid`).
local function stats(data,  showCols,fun,    t)
  showCols, fun = showCols or data.cols.y, fun or mid
  t={}; for _,col in pairs(showCols) do t[col.name]=fun(col) end; return t end
```

```
230
231 -- ---------------------------------
232 -- ## Test Engine
233 local eg, fails = {},0
234
235 -- [1] reset random number seed before running something.
236 -- [2] Cache the detaults settings, and [3] restore them after the test
237 -- [4] Print error messages or stack dumps as required.
238 -- Return true if this all went well.
239 local function runs(k,     old,status,out,msg)
240   if not eg[k] then return end
241   math.randomseed(the.seed) -- reset seed [1]
242   old={}; for k,v in pairs(the) do old[k]=v end --  [2]
243   if the.dump then
244     status,out = true, eg[k]()
245   else
246     status,out = pcall(eg[k])  -- pcall means we do not crash and dump on errror
247   end
248   for k,v in pairs(old) do the[k]=v end -- restore old settings [3]
249   msg = status and ((out==true and "PASS") or "FAIL") or "CRASH" -- [4]
250   print("!!!!!!", msg, k, status)
251   return out or err end
252
253 -- ---------------------------------
254 -- ## Tests
255 -- Test that the test  happes when something crashes?
256 function eg.BAD() print(eg.dont.have.this.field) end
257
258 -- Sort all test names.
259 function eg.LIST(   t)
260   t={}; for k,_ in pairs(eg) do t[1+#t]=k end; table.sort(t); return t end
261
262   -- List test names.
263 function eg.LS()
264   print("\nExamples lua csv -e ...")
265   for _,k in pairs(eg.LIST()) do print(string.format("\t%s",k)) end
266   return true end
267
268 -- Run all tests
269 function eg.ALL()
270   for _,k in pairs(eg.LIST()) do
271     if k ~= "ALL" then
272       print"\n-----------------------------------"
273       if not runs(k) then fails=fails+ 1 end end end
274   return true end
275
276 -- Settings come from big string top of "sam.lua"
277 -- (maybe updated from comamnd line)
278 function eg.the() oo(the); return true end
279
280 -- The middle and diversity of a set of symbols is called "mode"
281 -- and "entropy" (and the latter is zero when all the symbols
282 -- are the same).
283 function eg.sym(  sym,entropy,mode)
284   sym= adds(Sym(), {"a","a","a","a","b","b","c"})
285   mode, entropy = mid(sym), div(sym)
286   entropy = (1000*entropy)//1/1000
287   oo({mid=mode, div=entropy})
288   return mode=="a" and 1.37 <= entropy and entropy <=1.38 end
289
290 -- The middle and diversity of a set of numbers is called "median"
291 -- and "standard deviation" (and the latter is zero when all the nums
292 -- are the same).
293 function eg.num(  num)
294   num=Num()
295   for i=1,100 do add(num,i) end
296   local med,ent = mid(num), div(num)
297   print(mid(num) ,div(num))
298   return 50<= med and med<= 52 and 30.5 <ent and ent <32 end
299
300 -- Nums store only a sample of the numbers added to it (and that storage
301 -- is done such that the kept numbers span the range of inputs).
302 function eg.bignum(  num)
303   num=Num()
304   the.nums = 32
305   for i=1,1000 do add(num,i) end
306   oo(nums(num))
307   return 32==#num._has; end
308
309 -- Show we can read csv files.
310 function eg.csv()
311   local n=0
312   csv("../data/auto93.csv",function(row)
313     n=n+1; if n> 10 then return else oo(row) end end); return true end
314
315 -- Print some stats on columns.
316 function eg.stats()
317   oo(stats(records("../data/auto93.csv"))); return true end
318
319 -- ---------------------------------
320 the = cli(the)
321 runs(the.eg)
322 rogues()
323 os.exit(fails)
```