Ana Cristina Fazza

Doutoranda em Genética e Biologia Evolutiva

Laboratório de Genética e Evolução Molecular de Aves

Instituto de Biociências, USP.

Exercícios

exec

Proposta de projeto

PLANO A

A idéia inicial é criar uma função para colorir redes de haplótipos.

Um rede de haplótipos é uma representação gráfica da diversidade genética. Por meio dessa análise podem ser feitas diversas inferências como estruturação geográfica das populações, ocorrência de expansão demográfica, haplótipos ancentrais e derivados.

Diversos programas calculam redes de haplótipos utilizando diferentes algoritmos. O programa que eu utilizo para construir as redes de haplótipos é um programa gratuito. Esse programa possui a possibilidade de inserir informações para facilitar a coloração. Porém, para usar a função de colorir a rede tem que comprar um programa (associado ao “gratuito”) de US$ 135 cuja licença dura apenas 1 ano! Ou seja, quem faz essa análise sem querer gastar esse valor anual tem que colorir os haplótipos um a um!

Se inicialmente eu opto por separar as populações por munícipio de coleta o que eu faço é o seguinte:

- uma tabela com os haplótipos e quais as amostras que possuem cada haplótipo

- associo com as amostras por localidade de coleta

- faço as divisões no gráfico de pizza de cada haplótipo manualmente e então resta “apenas” colorir um por vez

Se depois você acha que ficaria mais informativo se dividisse por região geográfica, você refaz tudo novamente, clicando com o botão direito sobre cada haplótipo, redividindo e recolorindo...

Por exemplo:

Haplotipo	Amostra da espécie	Localidade	Rede 1	Região Geográfica	Rede 2
C	Individuo 128	Serra do Chuchu	amarelo	Norte	Azul
A	Individuo 129	Serra do Chuchu	amarelo	Norte	Azul
C	Individuo 200	Fim do mundo	preto	Sul	Vermelho
B	Individuo 230	Fim do mundo	preto	Sul	Vermelho
D	Individuo 256	Fim do mundo	preto	Sul	Vermelho
A	Individuo 375	Vale Encantado	vermelho	Norte	Azul
A	Individuo 567	Vale Encantado	vermelho	Norte	Azul

Bom...pode parecer tranquilo...mas quando tem muitas amostras/localidades é bem cansativo verificar uma por vez e colorir. Como isso:

E nem é uma amostragem tão grande... Além do que tem sempre que conferir várias vezes se não fez nada errado.

É isso!

PLANO B

Fazer uma tabela frequencia e notas dos alunos. Não é nada muito original, mas eu vou precisar fazer esse cálculo pelo excel em breve. Então, se conseguir fazer uma rotina no R será ótimo! A idéia é que o arquivo de entrada contenha dados como nome, data das aulas contendo a informação se o aluno estava presente ou ausente para calcular sua frequencia geral no curso, notas de trabalhos e provas. A estrutura básica do arquivo de entrada será mais ou menos assim:

Nome	Aula 05.03.12	Aula 12.03.12	Aula 19.03.12	Aula 26.03.12	Trab 1	Trab 2	Trab 3	Prova 1	Prova 2
Joãozinho	Presente	Presente	Ausente	Presente	9.0	5.0	8.5	7.0	5.0

e eu quero obter um arquivo de saída com os dados:

Nome	Frequencia	Nota final

onde nota final será a aplicação de uma fórmula somando os trabalhos e provas.

Bem mais simples que o Plano A, mas espero fazer os dois pelo menos para facilitar a minha vida ;)

Comentários

Sua explicação da rede é boa, colocou até o tipo de dado e fica fácil visualizar a tarefa da função. Entretanto, como nunca fiz nenhuma figura parecida, fica difícil avaliar a dificuldade que irá enfrentar. Acho que o desafio é grande para a primeira proposta e a segunda mais tranquila... Rode o código abaixo para ter uma idéia das possibilidades, se conseguir estudar o código e entende-lo fica meio caminho andado para a proposta A (nao se assuste com o jeito que o codigo cria os objetos de dados,

é uma forma que não apresetamos em aula!).

```
set.seed( 1 ) # fixa a semente aleatória
install.packages("network") ## instala pacotes necessários
install.packages("sna")
install.packages("Hmisc")
## sobe os pacotes instalados (mesmo que library)
require(network) # network
require(sna) # plot.network
require(Hmisc) # largest.empty

# Cria os dados a partir de uma linha de texto
## aqui poderia ser a tabela que postou!

con <- textConnection("name,age,jobyrs,inc,saved,int,chld,addryrs
Adam,21,1,46,110,3.9,2.62,16.46
Ben,32,6,82,110,3.9,2.875,17.02
Chris,22,2,38,93,3.85,2.32,18.61
David,46,16,158,110,3.08,3.215,19.44
Elvin,58,4,110,175,3.15,3.44,17.02
Fred,18,1,25,105,2.76,3.46,20.22
Greg,23,1,60,245,3.21,3.57,15.84
Henry,44,12,46,62,3.69,3.19,20
Ivan,32,2,40,95,3.92,3.15,22.9
James,29,5,67,123,3.92,3.44,18.3
Kevin,27,10,67,123,3.92,3.44,18.9
Luke,46,18,75,180,3.07,4.07,17.4")
nodelist <-read.csv(con)
close( con )

### Aqui são os dados das conexoes entre individuos ou no seu caso dos
haplótipos
con <- textConnection("from,to
Adam,Ben
Adam,Chris
Chris,David
Chris,Elvin
David,Fred
Fred,Greg
Greg,Henry
Henry,James
Henry,Ivan
James,Kevin
Ivan,Kevin
Kevin,Luke")

edgelist <- read.csv(con) ## lendo o texto anterior e colocando em um data
frame
close( con )
str(edgelist)
nodes <- levels(as.factor(nodelist[[1]]))
```

```
# Uma matrix que ira representar as conexoes da rede, preenchida com zeros

m <- matrix(data = 0, nrow=length(nodes), ncol=length(nodes))
rownames(m) <- colnames(m) <- nodes

## aqui uma função para colocar 1 na matrix m onde há conexao entre os
elementos
linque1<-function(x){ m[x[[1]], x[[2]]] <<- 1}

## aplicando a função para preencher o m usando os dados de edgelist
apply(edgelist, 1, linque1 )

## veja como ficou a matrix m e compare com  a edgelist:
m
edgelist
### preparando o grafico
graph <- network(m, matrix.type="adjacency")

# primeiro plot as linhas de conexoes entre os individuos (haplotipos)

par(xpd=TRUE)
## aqui pega os x y do objeto montado anteriormente e plot as ligacoes
xy <- plot(graph, vertex.cex=5, vertex.col="white", vertex.border=0)

# aqui procura um espaço vazio para plotar uma legenda (pizza)
kl <- largest.empty(xy[,1], xy[,2], 2, 2)
# aqui plot os dados
stars(nodelist[-1], labels=nodelist[[1]], locations=xy, draw.segments=TRUE,
      key.loc=c(kl$x, kl$y), add=TRUE)
```

— *Alexandre Adalardo de Oliveira 2012/04/03 18:07*

TRABALHO FINAL - PLANO B

Como o Ale disse nos comentários, o Plano A realmente foi um desafio muito grande para a minha primeira função, então eu não consegui...ainda! (Porque se até o Pink ficou inteligente em um episódio, eu ainda tenho a esperança de conseguir, em alguma noite, finalmente fazer esse Plano A!)

Segue o código para o Plano B.

Código da função

```
#funcao para gerar frequencia e medias finais dos alunos
notafinal <- function(x, n.aulas, n.notas, grafico=TRUE)
{
```

```
  alunos <- x[,1]
  frequencia <- x[,2:n.aulas+1]=="Presente"
  freq <- rep (0, dim(x)[1])
  for (i in 1:dim(frequencia) [1])
  {
    freq[i] <- paste(round(sum(frequencia[i,])/dim(frequencia)[2]*100, 2),
"%")
  }
  notas <- x[,(n.aulas +2):(n.aulas + n.notas +1)]
  media <- rep (0, dim(x)[1])
  for(k in 1:dim(notas)[1]) #calcula média com peso 1
  {
    media[k] <- sum(notas[k,1:n.notas])/n.notas
  }
  aprovacao <- rep (0, dim(x)[1])
  for(i in 1:length(media))
  {
    if(media[i] >= 5)
    {
      aprovacao[i] = "Aprovado"
    }
    else
    {
      aprovacao[i] = "Reprovado"
    }
  }
  if (grafico==TRUE)
  {
  x11()
  plot(alunos, media, ylab="Médias", main="Rendimento da turma", pch=19,
las=3)
  abline(h=mean(media), col="blue")
  abline(h=6, col="red")
  legend(7,4,legend=c("Média da turma", "Média para aprovação"),pch=c("_",
"_"),bty="n", col=c("blue", "red"))
  }
  final <- data.frame(alunos, freq, media, aprovacao)
  return(final)
}
```

Página de ajuda

```
notafinal                package:unknown                R Documentation

Nota final e frequencia dos alunos

Descrição:

     Função para calcular a frequencia dos alunos nas aulas e media final
calculada pela média aritmética simples dos trabalhos e provas. E mostra o
```

```
rendimento da turma, com a média de notas dos alunos e a média necessária
para a aprovação.

Uso:

notafinal(x, n.aulas, n.notas, ...)

Argumentos:

x:  tabela com nome dos alunos na primeira coluna, seguida pela informação
    "Presente" ou "Ausente" para cada aula, e notas de exercicios e provas.
n.aulas: valor para o número de aulas.
n.notas: valor para o número de avaliações.
grafico: argumento lógico, por padrão TRUE, retorna o rendimento da turma.
     Se FALSE, não retorna gráfico.


Autora:

     Ana Cristina Fazza


Exemplos:

##exemplo para rodar a funcao
baixar o arquivo "exemplo.csv"
planilha <- read.table("exemplo.csv", header=T, sep=";") #nao usar
row.names=1
notafinal(planilha, 10, 7)

##exemplo para rodar a funcao sem baixar o arquivo
alunos <- c("Homer", "Marge", "Bart", "Lisa", "Maggie", "Fred", "Wilma",
"Pedrita", "Barney", "Betty", "Bambam")
aula <- matrix(sample(factor(c("Presente", "Ausente"))), ncol=10,
nrow=length(alunos))
notas <- matrix(sample(round(rnorm(11, 7.5, 2), 2)), ncol=7,
nrow=length(alunos))
planilha <- data.frame(alunos, aula, notas)

notafinal(planilha, 10, 7)
```

Arquivos

Código da funçãoArquivo de ajuda Exemplo

From:
http://ecor.ib.usp.br/ - **ecoR**

Permanent link:
http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2013:alunos:trabalho_final:crisfazza:start

Last update: **2020/08/12 06:04**