

GGEBILOT VÀ NGÔN NGỮ R

Nguyễn Đình Hiền

Khoa Công nghệ Thông tin, Trường Đại học Nông nghiệp Hà Nội

Email: dinhhien@fpt.vn

Ngày gửi bài: 19.12.2013

Ngày chấp nhận: 05.01.2014

TÓM TẮT

Biplot là kiểu đồ thị vẽ đồng thời nhân tố hàng và nhân tố cột trên cùng một hệ tọa độ do đó có thể dùng để nghiên cứu một cách trực quan mối quan hệ hàng - cột của một bảng 2 chiều. Biplot dùng để xem xét tương tác kiểu gen - môi trường được gọi là GGEBiplot. Có thể dùng các chương trình máy tính như SPSS, MiniTab, Irristat để tạo ra Biplot nhưng muốn có hình ảnh đẹp, dễ dùng và giúp trả lời một số vấn đề hay gặp trong nghiên cứu tương tác kiểu gen - môi trường thì nên dùng các chương trình chuyên dụng như GGEBiplot của Weikai Yan. Ở đây chúng tôi giới thiệu một số đoạn chương trình trong ngôn ngữ R có thể giúp tạo ra GGEBiplot.

Từ khóa: Biplot, GGEBiplot, R.

GGEBiplot and R Language

ABSTRACT

Biplot is a scatter plot that graphically displays both the row factors and the column factors of a two way table. When a Biplot is used to examine genotype - environment interaction we have a GGEBiplot. Some procedures in R are given to generate GGEBiplot, visualize results and answer to some questions in analyzing GxE interaction.

Keywords: Graphically display, R, GGEBiplot.

1. Giới thiệu ngôn ngữ R

Năm 1996, trong một bài báo về tính toán thống kê, hai nhà thống kê học Ross Ihaka và Robert Gentleman thuộc Trường đại học Auckland, New Zealand phác họa một ngôn ngữ mới cho phân tích thống kê mà họ đặt tên là R. Sáng kiến này được rất nhiều nhà thống kê học trên thế giới tán thành và tham gia vào việc phát triển R.

Cho đến nay càng ngày càng có nhiều nhà thống kê học, toán học, nghiên cứu trong mọi lĩnh vực đã chuyển sang sử dụng R để phân tích dữ liệu khoa học. Trên toàn cầu đã có một mạng lưới hàng triệu người sử dụng R.

R là một phần mềm sử dụng cho phân tích thống kê và vẽ biểu đồ. Thật ra, về bản chất, R là ngôn ngữ máy tính đa năng, có thể sử dụng cho nhiều mục tiêu khác nhau, từ tính toán đơn giản, toán học giải trí, tính toán ma trận (matrix), đến các phân tích thống kê phức tạp. Vì là một ngôn ngữ, cho nên người ta có thể sử dụng R để phát triển thành các phần mềm chuyên môn cho một vấn đề tính toán cá biệt.

2- Cài đặt R

Để cài đặt R trong máy tính của mình phải truy nhập vào website "Comprehensive R Archive Network" (CRAN) sau đây: <http://cran.R-project.org>., sau đó chọn Cran mirrors. Thí dụ:

<http://cran.csiro.au/>

<http://cran.ms.unimelb.edu.au/>

CSIRO

University of Melbourne

Dựa vào vào phiên bản và hệ điều hành để chọn tài liệu cần tải về.

Tại các website này có thể tìm thấy rất nhiều tài liệu chỉ dẫn cách sử dụng R, đủ trình độ, từ đơn giản đến phức tạp.

Khi đã tải R xuống phải cài đặt vào máy tính. Để làm việc này cần nhấn chuột vào tài liệu trên và làm theo hướng dẫn cách cài đặt trên màn hình.

R cung cấp một “ngôn ngữ” máy tính và một số chức năng để làm các phân tích căn bản và đơn giản. Nếu muốn làm những phân tích phức tạp hơn cần phải tải về máy tính một số package khác. Package là một phần mềm nhỏ được các nhà thống kê phát triển để giải quyết một vấn đề cụ thể, và có thể chạy trong hệ thống R. Chẳng hạn như để phân tích hồi qui tuyến tính, R có chức năng lm để sử dụng cho mục đích này, nhưng để làm các phân tích sâu hơn và phức tạp hơn cần đến các package như **lme4**. Các package này cần tải về và cài đặt. Địa chỉ các package vẫn là: [http://cran.r-](http://cran.r-project.org)

[project.org](http://cran.r-project.org), bấm vào phần Packages” để tìm, kèm các trang web để tải về. Một số package thường dùng trong các phân tích thống kê là:

Các package này có thể cài đặt trực tuyến bằng cách chọn **Install packages** trong phần **packages** của R. Nếu package đã được tải xuống máy tính việc cài đặt có thể nhanh hơn bằng cách chọn **Install package(s) from local zip file** cũng trong phần **packages**.

3. Ggebiplot trong ngôn ngữ R

Nghiên cứu tính ổn định là vấn đề được đề cập nhiều khi nghiên cứu mối quan hệ giữa giống và môi trường. Có nhiều định nghĩa về ổn định với nhiều chỉ tiêu khác nhau và phần lớn các chỉ tiêu đó đòi hỏi khối lượng tính toán khá lớn và phức tạp. Có thể dùng các chương trình máy tính chuyên về tính ổn định như Ondinh (1) hay có mục chuyên về tính ổn định như Irristat (2). Cũng có thể kết hợp tính một phần bằng các chương trình máy tính như Excel, MiniTab, SPSS, SAS . . . sau đó tính tay.

Tên package	Chức năng
lattice	Dùng để vẽ đồ thị và làm cho đồ thị đẹp hơn trong nghiên cứu thống kê nhiều chiều
agricolae	Phân tích thống kê một số mô hình thí nghiệm trong Nông nghiệp (Statistical Procedures for agricultural Research)
agridat	Các tệp số liệu dùng trong nghiên cứu Nông nghiệp
Design	Một số mô hình thiết kế thí nghiệm trong nghiên cứu khoa học
Epi	Phân tích thống kê trong dịch tễ học
epitools	Một package khác chuyên cho các phân tích thống kê trong dịch tễ học
Foreign	Nhập dữ liệu từ các phần mềm như SPSS, Stata, SAS, ...
lme4	Mô hình tuyến tính hỗn hợp (Linear mixed effects models)
pspearman	Tương quan thứ tự Spearman (Spearman's rank correlation test)
BiplotGUI	Phân tích các bảng hai chiều qua đồ thị
GGEbiplotGUI	Phân tích bảng đồ thị mối quan hệ giữa giống và môi trường
bpca	Phân tích thành phần chính

survival	Mô hình Cox (Cox's proportional hazard model)
Zelig	Phân tích thống kê trong xã hội học
Genetics	Phân tích số liệu trong di truyền học
BMA	Bayesian Model Average

Muốn hình dung cụ thể mối quan hệ giống và môi trường các nhà khoa học thường dùng phương pháp vẽ đồng thời giống và môi trường lên cùng một mặt phẳng với 2 trục tọa độ là hai thành phần chính của ma trận tương quan giữa giống và môi trường. Để có hình ảnh vừa đẹp vừa phản ánh trung thực mối quan hệ giữa giống và môi trường phải lựa chọn tỷ lệ xích của hai trục cho phù hợp. Tất cả các vấn đề này được đề cập trong các chương trình Biplot.

Trong ngôn ngữ R có các Packages như BiplotGUI, GGEBiplot GUI, bPCA, gplot2 có đề cập đến Biplot.

Sau đây là một số đoạn chương trình và kết quả trong R

Dùng GGEBiplotGUI

```
> library(GGEBiplotGUI)
> data(Ontario)
# tệp số liệu có sẵn trong GGEBiplotGUI
gồm 18 loại lúa mì trồng ở 9 địa điểm
> Ontario
```

> GGEBiplot(Ontario)

Ta thu được đồ thị Biplot có cả giống và địa điểm.

Trên cơ sở Biplot này có thể tìm được các đồ thị về:hc

Trồng giống nào ở đâu thì tốt (Which won where/what **Hình 1c**)

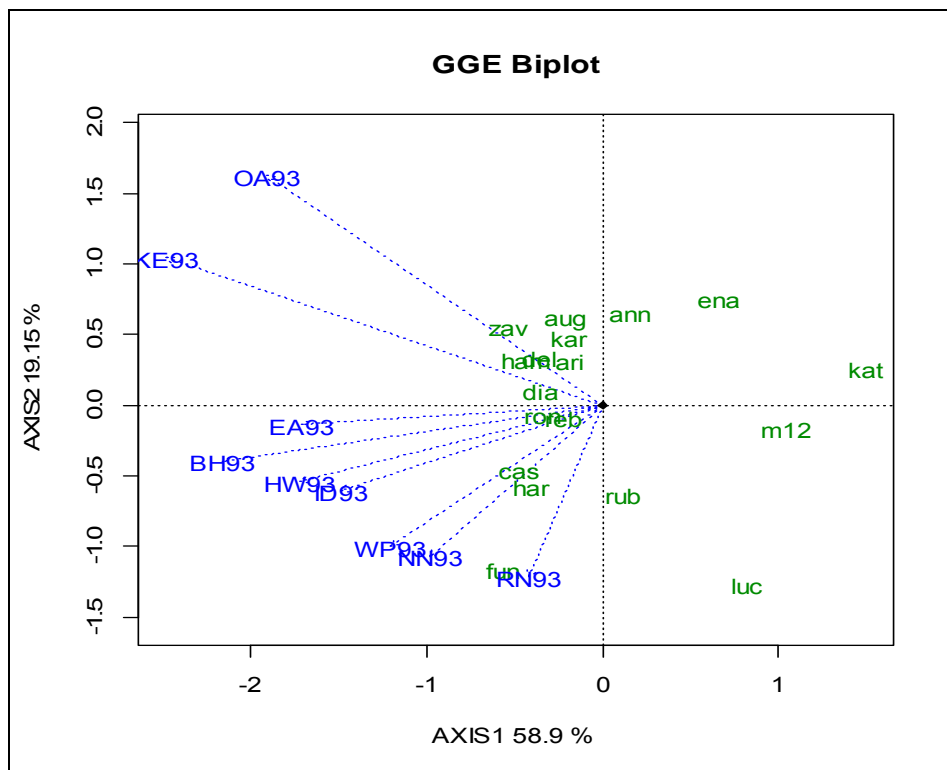
Xem xét một giống (Examine a Genotype **Hình 1d**)

Xem xét một địa điểm (Examine an Environment **Hình 1e**)

Tính ổn định (Mean vs stability **hình 1f**)

So sánh 2 giống (Compare two genotypes)

Sắp xếp thứ tự các giống dựa vào địa điểm lý tưởng (Rank Genotype/ Environment with reference to the “ideal Environment” **hình 1j**) hoặc sắp xếp thứ tự các địa điểm dựa vào giống lý tưởng (Rank Environment with reference to the “ideal Genotype”)

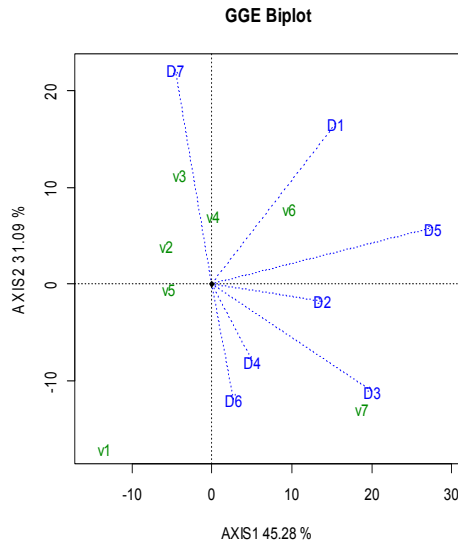


Hình 1a

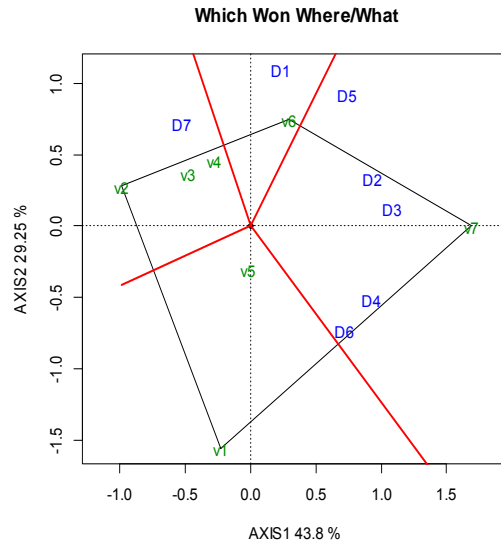
Có thể dùng thí dụ về 7 giống lúa trồng tại 7 địa điểm (tệp ondingh.txt) để minh họa.

```
>library(GGEBiplotGUI)
>
test<-
read.table("C:/nnr/ondingh.txt",header=TRUE)
> test
  D1 D2 D3 D4 D5 D6 D7
v1 27.5 35.7 46.4 53.7 33.3 64.9 43.3
```

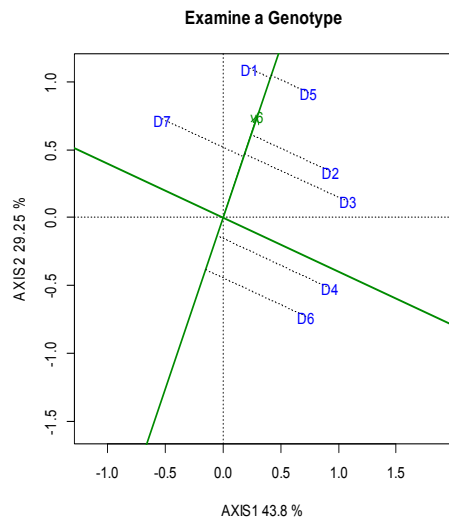
```
v2 35.7 37.5 46.2 40.8 51.9 45.6 57.5
v3 46.4 46.2 38.7 49.1 50.4 55.6 69.4
v4 53.7 40.8 49.1 51.2 49.4 48.1 57.5
v5 33.3 51.9 50.4 49.4 42.5 63.1 68.9
v6 54.9 45.6 55.6 48.1 63.1 60.0 63.1
v7 43.3 57.5 69.4 57.5 68.9 63.1 43.7
> GGEBiplot(test)
```



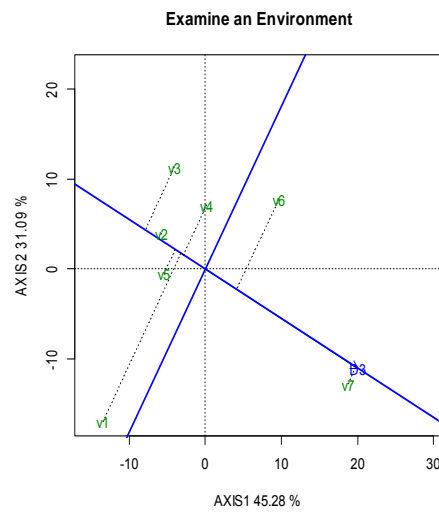
Hình 1b (tương tự hình 1a)



Hình 1c



Hình 1d



Hình 1e

Dùng BiplotGUI

```
> library(BiplotGUI)
> data(Countries)
> Biplots(Countries)
```

Chúng ta được một số hình chiếu trên hai trục thành phần chính.

Dùng bpca

```
> library(bpca)
> test<-
read.table("C:/nnr/ondinhg.txt",header=TRUE)
> plot(bpca(test))
```

Chúng ta được **hình 1b**

```
> plot(bpca(test), type='eo', obj.id=6,
obj.cex=1)
```

Chúng ta được **hình 1d**

```
> plot(bpca(test), type='ev', var..id=36,
var.cex=1)
```

Chúng ta được **hình 1e**

```
> plot(bpca(test),type='ww')
```

Chúng ta được **hình 1c**

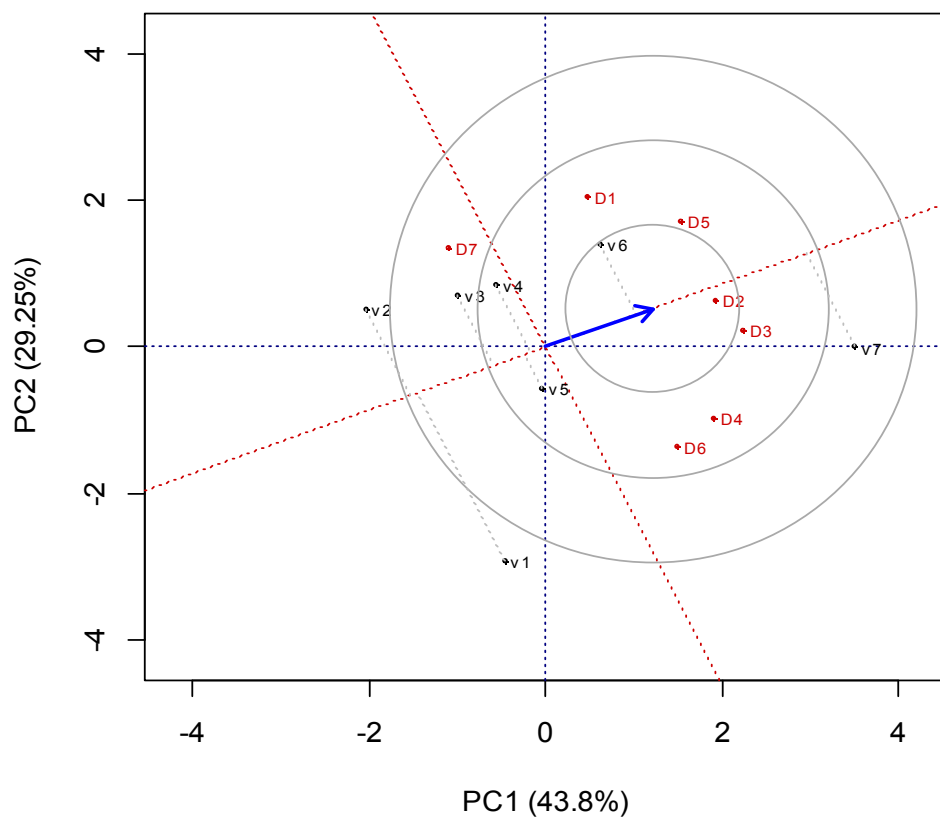
```
> plot(bpca(test),type='ms')
```

Chúng ta được hình vẽ về tính ổn định

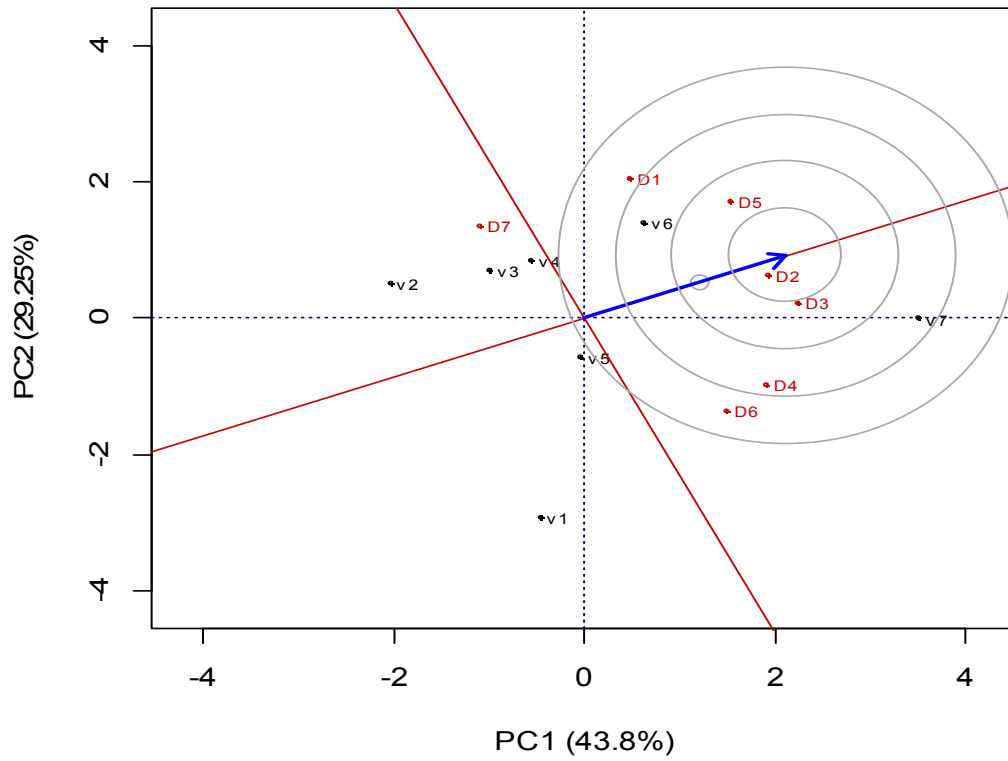
hình 1f

```
> plot(bpca(test),type='ms')
```

Chúng ta được hình vẽ sắp xếp các genotype căn cứ vào Ideal environment **hình 1j**



Hình 1f



Hình 1j

TÀI LIỆU THAM KHẢO

Chương trình onđinh của Nguyễn Đình Hiền Bộ môn
công nghệ phần mềm Đại học Nông nghiệp Hà nội.
Phần mềm thống kê Irristat của viện lúa quốc tế IRRI.

Weikai Yan. Software Ggebiplot.(Bản Beta Ggebiplot
trên mạng tại địa chỉ w.w.w.ggebiplot.com).