

Shine Mosquito, shine !!!!

STAT 585 Final Project

Ignacio Alvarez

Natalia da Silva

April 24, 2014

1 Introduction

We want to learn and build a shiny app as a final project for Stat585, using **shiny** package we can create web application for data analysis that allow the final user make the data exploration by themselves.

Iowa State University Medical Entomology Laboratory and Iowa Department of Public Health and the University Hygienic Lab (Iowa City, IA), have been monitored on an annual basis mosquito populations and mosquito-borne disease in the State of Iowa since 1969. “The primary goal of this project was to integrate light trap data from these efforts into a centralized back-end database and interactive website that is available through the internet at <http://iowa-mosquito.ent.iastate.edu>.” (this web page is down from April 7-th).

In the web page from Entomology Laboratory you can find weekly mosquito abundance compared with data from previous years. Additional interactive capabilities facilitate analyses of the data based on mosquito species, distribution, or a time frame of interest. All data can be viewed in graphical and tabular format and can be downloaded.

Objective: The objective of this project is design a web application using shiny package to help entomologist to visualize, summarize and analyze the data produced by the surveillance program. The information presented in the Entomology Laboratory web today allows see information by specie along all the years but we can never see more than one specie or more than one site at the same time. Then the information is there but the comparisons are difficult. We will include different visualization tools in the shiny app to make simpler the comparison at different levels (individual, group and community).

2 Application structure and design

Here we describe how we design the web app, what are the questions we want to address and how this is done in the app. The app have 5 tabset panels each of them try are focused in some analysis level, we describe how to use the shiny app based on examples.

We use part of the dataset, the subset that entomologists are more confident with and they think are the most relevant for the analysis. We use yearly mosquito counts for 8 sites (cfall, green valley, edale, ewing, gildea, union, wloo, wlp) and 20 years of information. However complete data has more sites and years on weekly basis but the information with this desegregation is not comparable.

The shiny app is designed to answer questions at three analysis level:

1. *Within specie:* the objective is to analyse the individual behavior of one specie across years and for different sites. We want to answer: The specie behavior have changed across the years? Are some

differences for each specie between sites? How many years the proportion of each specie is above the mean proportion of the specie across sites?

2. *Between genotype*: the objective is to analyse more than one genotype at the same time. We want to answer: which genotype are more relevant in each site along all the years? Are there some years where some genotype is more relevant in one site?
3. *Community level*: we want to see the community dynamic, for that we analyse different indexes to describe this dynamic.

2.1 Tabset 1: Specie-site

In this tabset the user can find a **selectInput** with all the mosquito species and a **checkbox** with all sites. The idea is to analyse one species at time across all years and compare the behavior in the different sites. The information presented after selection will be a plot with the proportion of the mosquito specie in a specific year in a specific site. In each plot also there are a reference red line, this line represents the mean proportion for each species across sites and is useful to see how far is the behavior of the specie in the specific site respect to its mean behavior.

Additionally in the same tab set we can see a numeric summary of results like the proportion of the specie across all the years in the same site, the maximum proportion of that specie in a site and finally the number of years that in a specific site the specie is above its mean.

Choices on shiny: Species (only one) and multiple Sites. As an example, we show here the proportion of

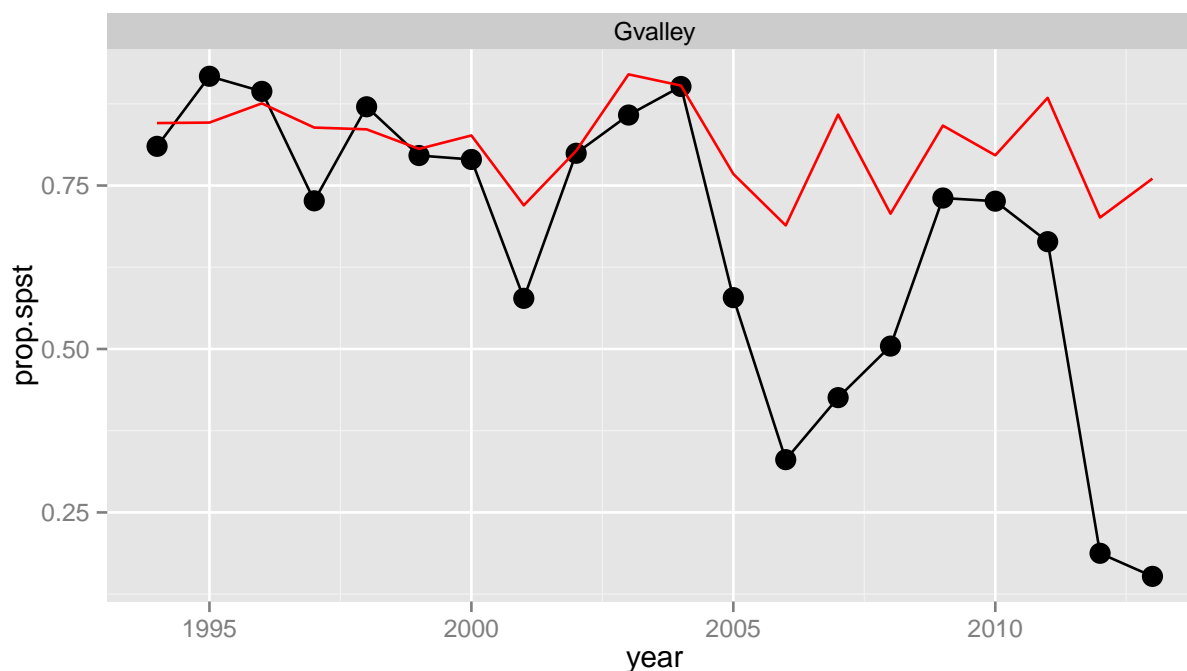


Figure 1: Example plot for tabset 1

Aedes.Vexan on Green Valley over time, the red line represents the average proportion of this species on each year. So on average across all years and sites, this species represents around 75% of the mosquito counts. However on Green Valley the Aedes.Vexan proportion is most of the time below the mean, What happen in Green Valley ? The Vexans don't like it anymore !!!!

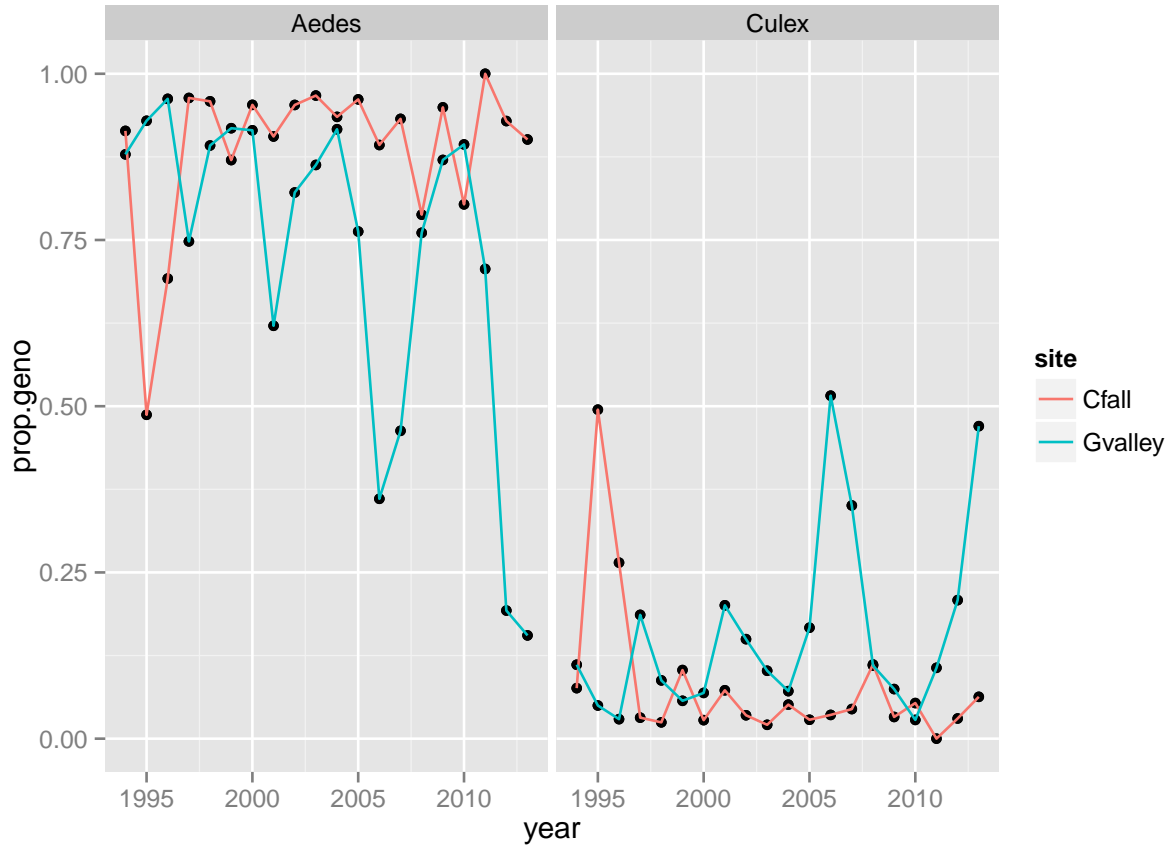


Figure 2: Example plot for tabset 2

2.2 Tabset 2: Genotype-site

In this tabset the user will find two checkboxes, one to select the genotype (Aedes, Anopheles, Coquilletidia, Culex, Culiseta, Orthopodomyia, Psorophora and Uranotaenia) and the other one to select the sites. In this case the user can select multiple sites and multiple genotypes. Here the idea is compare the genotypes in all the sites. After selection the user will see in each panel the proportion of a genotype in each year for a site or more than one. The plot is faceted by genotype and each additional selected site will be added as a new line in the specific genotype plot.

The example shows the two most abundant genotypes in iowa mosquitos, Aedes and culex.

2.3 Tabset 3: Location

In this tabset the user can select multiple species, after selection the user will see a plot with the Iowa map with the distribution of that specie in each site, the color indicate the abundance of the specific specie in the place. The idea in this tabset is see *where* is more frequent to see a particular specie. The current dataset has only 8 sites and some species are not abundant a complete white map indicate the specie in the specific site does not like that place. If the color in the site is red then that specie is more abundant in that specific site. We can see the spatial distribution of the specie across Iowa

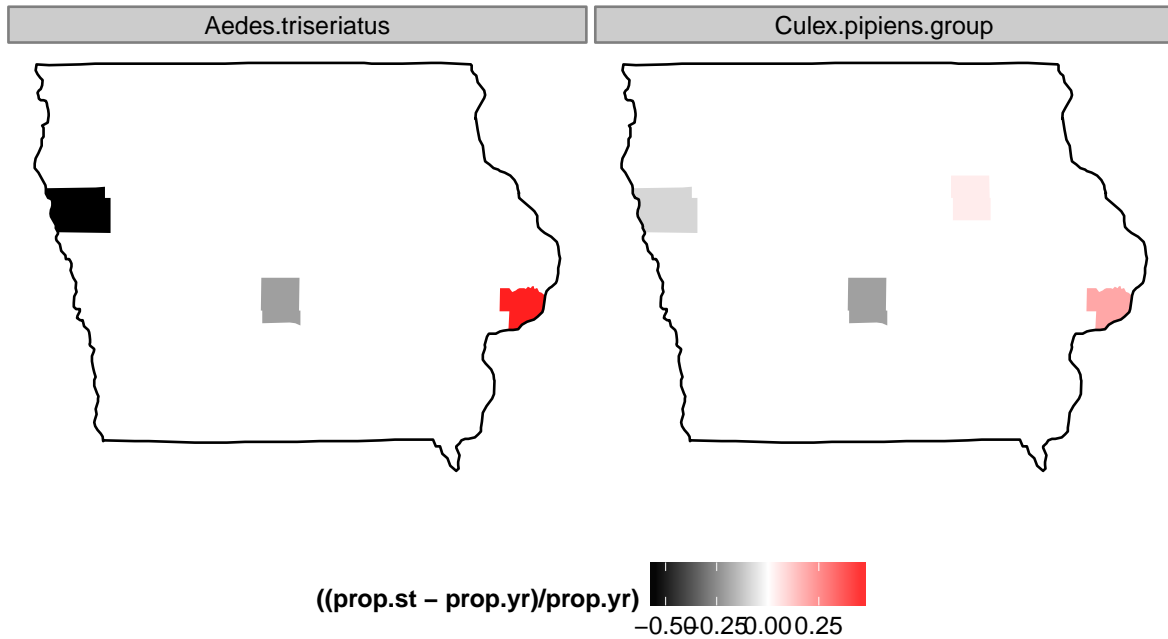


Figure 3: Example plot for tabset 3

2.4 Tabset 4: Rare communities

Besides describing each community we may be interesting in comparing them, in particular we want to know if all communities are alike or there are some site-year observations that are different from the rest. One important feature for the entomologist is to study the community characteristics, there are several indices reflecting different aspects of the mosquito community composition like abundance, simpson index, etc that will be useful for a community level comparison.

Working with species proportion within one site and year (instead of total counts) we compute an "average community" which simply taking average of each species proportion. Then for each observations we compute the distance respect to that average community.

In this tabset the user will find two select inputs with variables to describe the mosquito community composition and a scrol bar to select a specific quantile. After selection the user will see the density of the distance to the mean community, the red line is the quantile. In left panel each point is a site-year and the red ones are the extreme communities, the extreme values are colored based on the selected quantile (scrol bar) and in the axis we can select differenc variables to explain the community composition.

We plot a two panel figure, on the right is the estimated density of the distances to the mid point community. The vertical red line represents the 90 quantile (as defect).

On the left panel there is a scatter plot of two community variables and we colored the points corresponding to communities further from the center community, this is with distance bigger than its 90% quantile.

On the shiny it is possible to choose the quantile that determine which communities are consider extreme and also the variables plot in the right panel. In this example the selected quantile is 90% and the variables are Precipitation (horizontal) and Dominance index (vertical). Basically we can see all extreme communities shows big values in Dominance but are not extreme on precipitation.

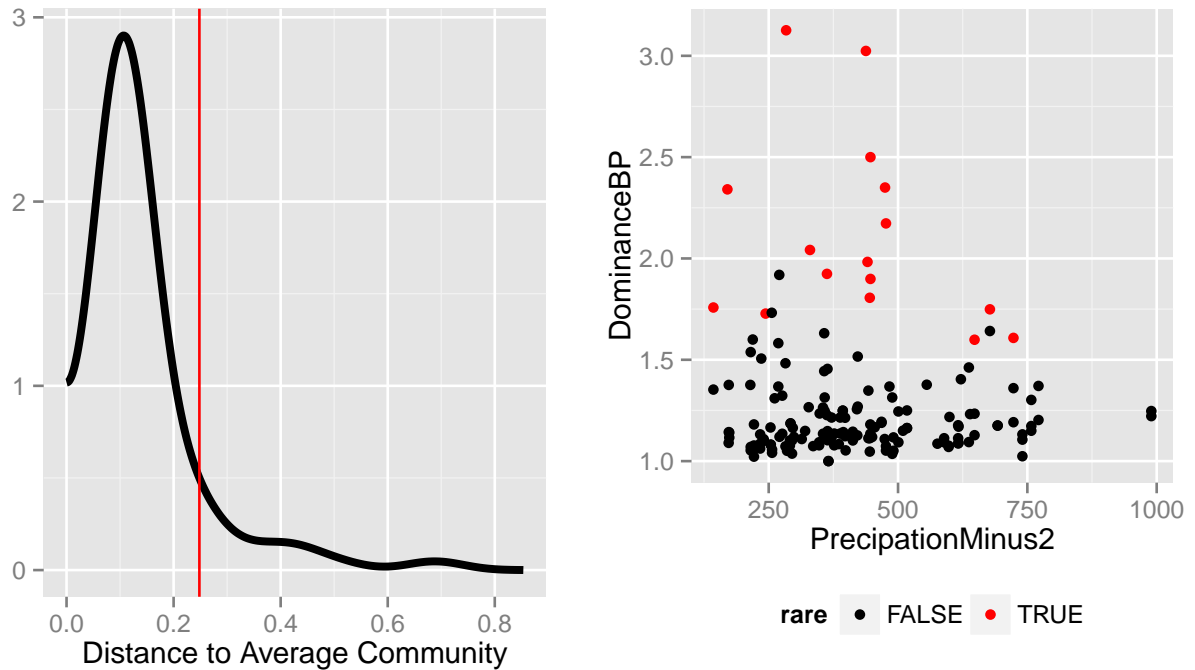


Figure 4: Example plot for tabset 4

2.5 Tabset MDS

The last tabset presents also information at community level, this time we compute a distance matrix with every distances among each pair of communities. We use Multidimensional Scaling (MDS) to visualize the distance matrix, the distance computation and the MDS are done using `vegan` package with `vegdist` and `metaMDS` functions.

The user will see a scatterplot with the MDS results and going over any point with the mouse it will display some information to identify the specific community, in particular, information about site, year and distance from the center will be displayed.

Additionally the user has the option to select a specific site and a specific year and will see these values colored and with different shape in the plot representation.

2.6 Using the app

We finish the web app description with a little instruction on how to run the application and get ready to analyze the mosquito surveillance data. You just need to follow these steps :

1. go to https://github.com/nachalca/mosquito_shiny
2. copy the folder `shiny_msq` into your local machine
3. Open R and change the working directory where `shiny_msq` folder is saved
4. run `library(shiny)`
5. run `runApp('shiny_msq')`

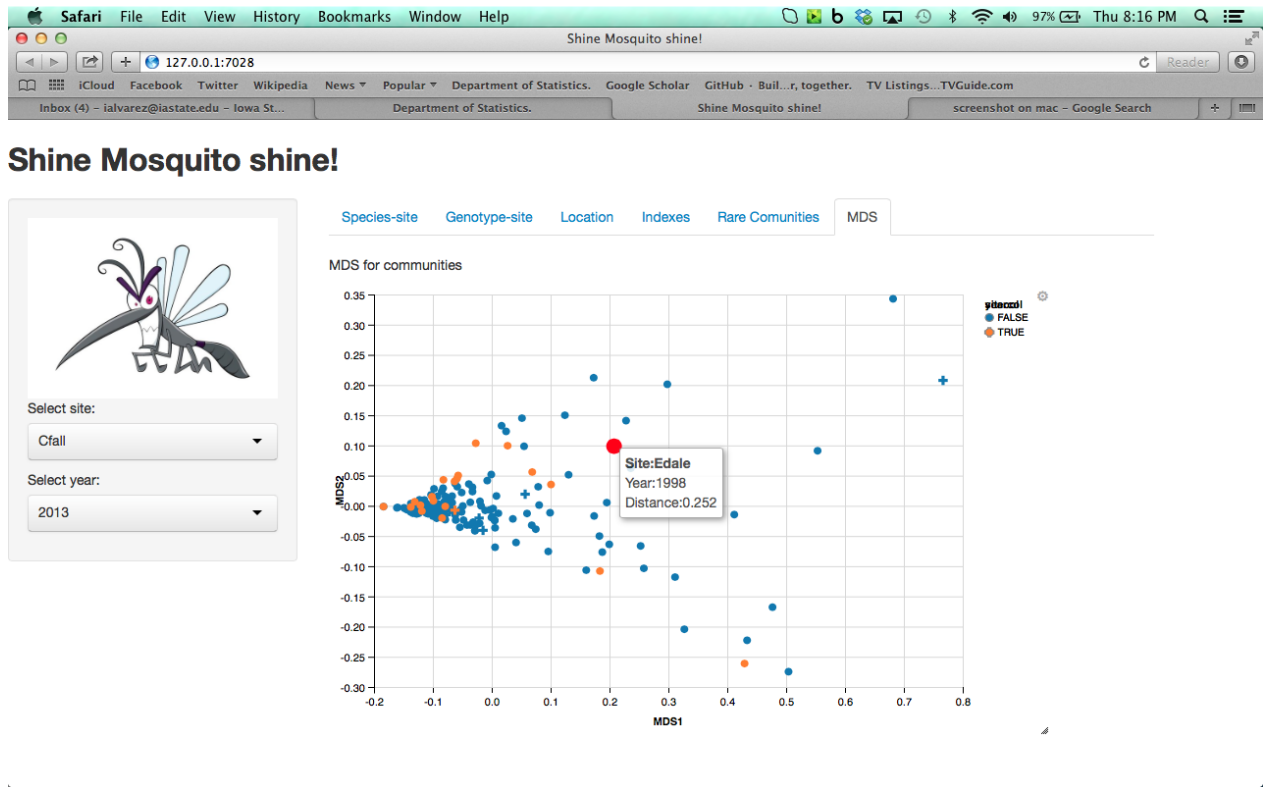


Figure 5: Example plot for tabset 5

3 Computing

There are 3 R files needed to run this app: `ui.R`, `server.R` and `datasources.R`. The last one is a code read and transform the data set used by the shiny app. We describe the main characteristics of `ui.R` and `server.R` files in this section.

3.1 UI

The ui have two principal aspects, the `sidebarPanel` and `mainPanel`, we have 5 tabsets in the web app and we design it to have different shiny options on each tabset. Then the `sidebarPanel` consist in a series of 5 `conditionalPanel` one per tabset, for example the tabset 4 options are implemented by the following code

```
conditionalPanel(condition="input.conditionedPanels==4",
  img(src="Mosquito_da_Dengue_by_Lukemaciel.png", height = 100, width = 300),
  sliderInput("q", "Quantile",
    min=0, max=100, value=90, step=1),

  selectInput(inputId = "index.X",
    label = "Select x axis variable",
    choices = lab.index, selected='PrecipitationExact'),

  selectInput(inputId = "index.Y",
    label = "Select y axis variable",
    choices = lab.index, selected='DegreeDayExact')
```

```
)
```

this create 3 options, the first one for selecting a quantile to act as a limit of what is consider as an extreme community, the other two allows the selection of the x and y variables for the scatterplot.

The `mainPanel` basically set up a tabpanel for each tabset we want to organize the output, the code is the following

```
mainPanel(
  tabsetPanel(
    tabPanel("Species-site",p(cap1), plotOutput("plot1"),tableOutput("tab1"),p(cap1.1),value=1),

    tabPanel("Genotype-site",p(cap6), plotOutput("plot6"),value=6),

    tabPanel("Location", plotOutput("plot2"),value=2),

    tabPanel("Indexes", plotOutput("plot3"),value=3),

    tabPanel("Rare Communities", p(cap4),plotOutput("plot4"),value=4),

    tabPanel("MDS",p(cap5), ggvis_output("my_plot"),value=5),

    id="conditionedPanels")

  ) # close mainPanel
```

3.2 Server

We have organized the server file in two main parts, a first part where we create reactive data set for each plot we want to include, and after using each reactive data set we create a reactive plot or table to include in the tabset.

As an example, here is the dataset and the plot to create the last tabset,

```
# data for plot 5
d5 <- reactive({
  data.frame(mds2, sitecol = as.factor(mds2$site == input$site5), yearcol = as.factor(mds2$year ==
    input$year5))
})

# function for display information with the tooltip
showSite <- function(x) {
  if (is.null(x))
    return(NULL)
  xx <- as.numeric(x)
  xx <- round(xx, 4)
  ss <- mds2[round(mds2$MDS1, 4) == xx[1] & round(mds2$MDS2, 4) == xx[2],
    ]
  paste0("<b>", "Site:", ss$site, "</b><br>", "Year:", ss$year, "<br>", "Distance:",
    round(ss$dist, 3))
}
```

```

# ggvis plot
gv <- reactive({
  # check<-input_select(unique((mds2fsite)))
  p <- ggvis(d5(), props(~MDS1, ~MDS2, shape = ~yearcol, fill = ~sitecol,
    `:=`(fill.hover, "red"), `:=`(size.hover, 200)))
  p + layer_point() + tooltip(showSite)
})

# set up the output plot
output$controls <- renderControls(gv)
observe_ggvis(gv, "my_plot", session)
output$plot6 <- reactivePlot(function() {
  print(ggplot(data = d6(), aes(x = year, y = prop.geno)) + geom_point(size = 2) +
    geom_line(aes(color = site)) + facet_wrap(facets = ~geno) + scale_x_continuous("Year") +
    scale_y_continuous("Proportion of geno"))
})

```