Student Dryad Projects – Bioinformatics 490S Duke University
Three projects from the Fall 2012 Bioinformatics course. Work shown with student permission.

Student learning objectives of this assignment:
Students will be able to interpret the results of primary literature.
Students will collect and analyze publically available datasets.
Students will analyze data and evaluate claims made by authors.
Students will create their own graphs and figures from the data in R and reinterpret the results.

Phage mutation project:
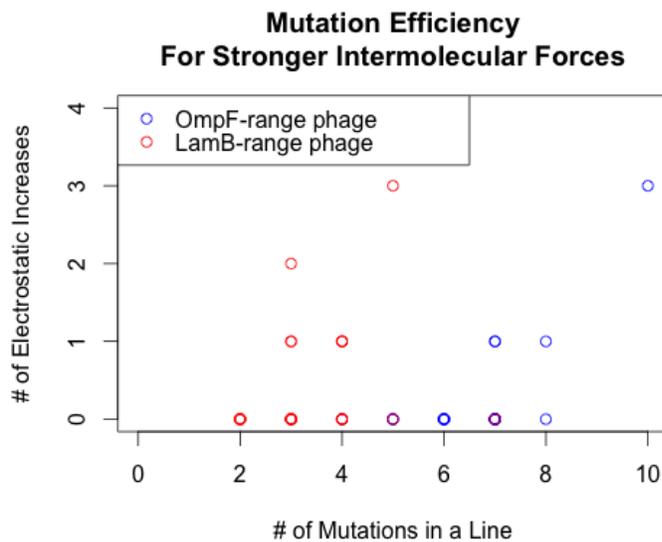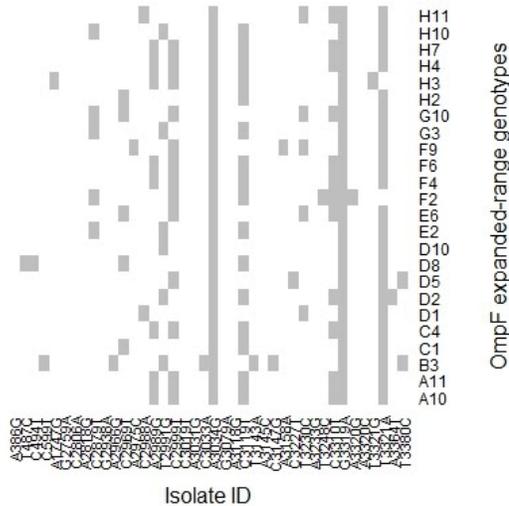
R code *(These students were docked points because their code lacked comments)*

```
setwd("~/_Duke/2012 Fall/Biology490S/Dryad Project/DRYAD FILES")
# read in names that will eventually be the x-labels
genes = read.table("JSNPs_names.txt", header=FALSE)
genes = t(genes)
lambnames = read.table("lamb_names.txt", header=FALSE)
lambnames = t(lambnames)
dataRead = read.table("lamb_only.txt", header=FALSE)
dataMatrix = t(data.matrix(dataRead))

my.colors = c("white", "gray")
plot<-heatmap(t(dataMatrix), Rowv=NA, Colv=NA, col=my.colors,
labRow=lambnames, labCol=genes, scale="row",
margins=c(5,5), revC=FALSE, xlab="Isolate ID",
ylab="LamB ancestral-range genotypes")

pomut <- c(7, 7, 10, 5, 7, 5, 7, 6, 7, 4, 6, 7, 8, 7, 7, 7, 6, 8, 5, 7, 7, 7, 6, 6)
plmut <- c(2, 4, 3, 2, 3, 3, 3, 3, 5, 3, 3, 4, 5, 2, 3, 4, 4, 2, 3, 4, 7, 4, 2, 4)
pobond <- c(0, 0, 3, 0 ,0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0)
plbond <- c(0, 0, 0, 0, 0, 2, 0, 0, 3, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0)

plot(pomut, pobond, xlab = "# of Mutations in a Line", col="blue", ylab = "# of Electrostatic Increases",
xlim=c(0,10), ylim=c(0,4))
points(plmut, plbond, col="red")
characters = c("OmpF-range phage", "LamB-range phage")
my.colors = c("blue","red")
legend("topleft", characters, col=my.colors, pch=1)
title(main="Mutation Efficiency\nFor Stronger Intermolecular Forces")
```
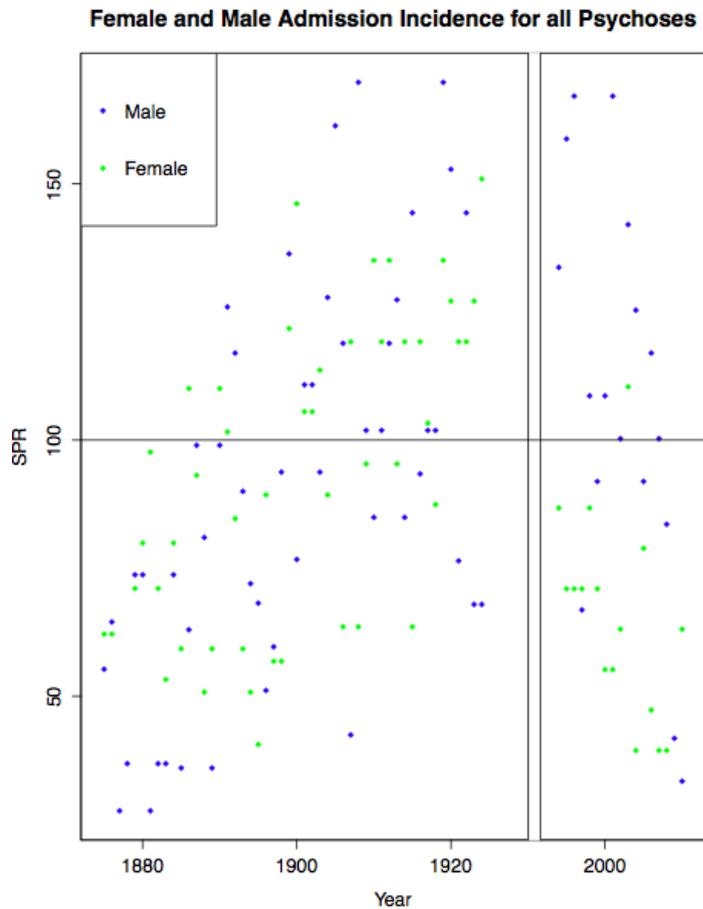
Human schizophrenia project:

**Female and Male Admission Incidence for all Psychoses**

R code

```
library(gplots)
#Allows us to use tools from installed package known as gplots. This was used to create a confidence
interval graph.
library(plotrix)
# Allows us to use tools from installed package known as plotrix. This was used to create a gap in our
graph on the x-axis.
malecomplete = read.table("cimen.txt",header=T,sep="\t")
# Imports table of total male admission incidence for all psychoses in the text file. This table was defined
as "malecomplete".
femalecomp = read.table("femalecomplete.txt",header=T,sep="\t")
# Imports table of total female admission incidence for all psychoses in the text file. This table was
defined as "femalecomp".
gap.plot(x=malecomplete$yearADM,y=malecomplete$SPR,gap.axis="x",gap=c(1930,1990), pch=18,
col= 'blue', ylab = "SPR", xlab = "Year")
```

# This creates a plot and a gap between the years of 1930 to 1990 in the x-axis of the entire graph. The x-values are the years, and y-values are the SPR (standard psychosis reate) for men.
title(main= 'Female and Male Admission Incidence for all Psychoses')
# This creates the title of the graph
abline(100,0)
# sets a line at a confidence of 100
gap.plot(x=femalecomp$yearADM,y=femalecomp$SPR,gap.axis="x",gap=c(1930,1990),ylim = c(0,250), pch=18, col='green', add= TRUE)
# This adds to existing plot. The x- values are the years, and y-values are the SPR (standard psychosis reate) for women.
legend("topleft", legend = c('Male','Female'), col = c('blue','green'),pch=18)
# This creates a legend on the top left corner with blue dots = male and green dots = female.

Drosophila recombination project:
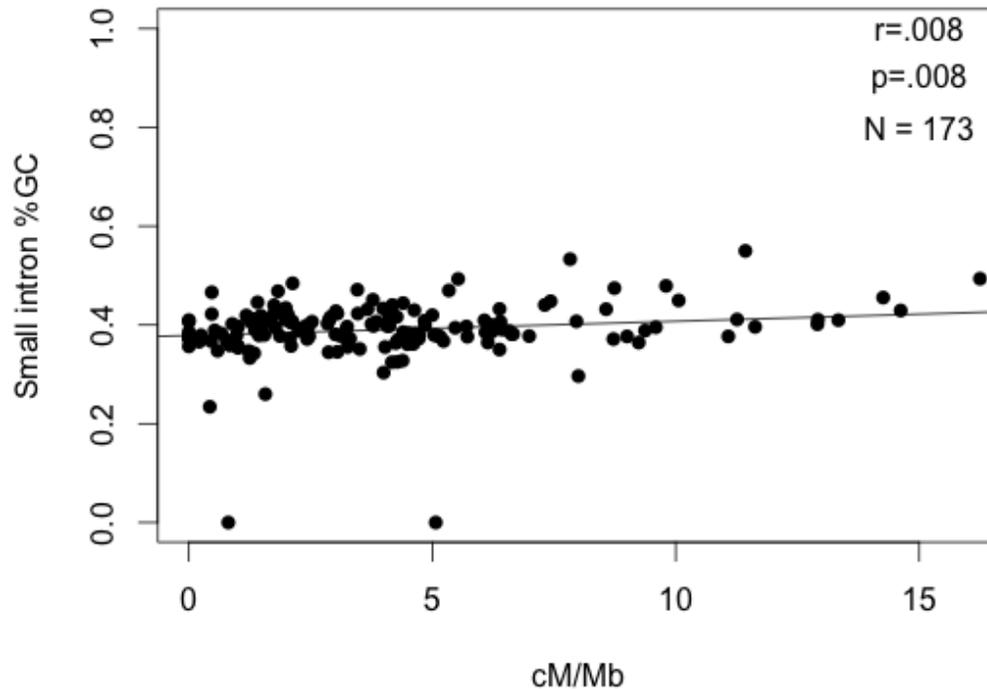
R code

```
#imports data and reads it as a table- sets header to top column, and says that data is comma delineated
data <- read.table("noor.csv", header=T,sep=",")
#creates plot based on the rate column and data column. sets x and y axis regions and labels both axes.
sets pch to 16.
plot( data$rate, data$GC, xlim=c(0,16), ylim=c(0,1), xlab= "cM/Mb", ylab = "Small intron %GC", pch
=16)
#creates best-fit line
abline(fit <- lm(data$GC ~ data$rate))
#utilizes summary for p-value and r^2
summary(lm(data$GC ~ data$rate))

#legend for r, p, and n values
text(15, 1, "r=.008")
text(15, 0.9, "p=.008")
text(15, 0.8, "N = 173")
```