

Aventuras aplicando R

Edith Seier PhD

Profesora Emérita UNMSM

Profesora Emérita East Tennessee St. Univ.

<http://faculty.etsu.edu/seier>

seier@etsu.edu

edithseier@Gmail.com

Por que programar y por que R?

Por que programar?

- Para poder repetir análisis con ligeras variaciones en el proceso de investigación
- Para poder hacer análisis nuevos

Por que R?

He usado SAS, Matlab, Gauss y son muy buenos pero son paquetes comerciales mientras que R es gratis, para mi el hecho que cualquier persona con acceso a una computadora y la internet pueda usarlo es importante

Esto ha hecho posible que se cree toda una comunidad de usuarios que crean paquetes que comparten con otros su trabajo y que haya un crecimiento explosivo de paquetes con fines específicos.

R es flexible, tiene excelentes facilidades graficas y muchas funciones estan implementadas (Claro también podría ser Phyton por las mismas razones)

- Analizando datos utilizo también varios de los paquetes estadísticos tradicionales y en consultoría busco algo que sea asequible a la otra persona.
- R es ideal cuando nos enfrentamos a situaciones de análisis no rutinarias pues podemos diseñar nuestra propia herramienta o buscar un paquete que realice ese análisis.
- Veremos algunas de las experiencias que he tenido con R motivadas, la mayoría, por aplicaciones reales para las cuales usar R convenía más que un paquete estadístico comercial
- R se considera junto con Python uno de los principales lenguajes en Ciencias de Datos. Cada uno de ustedes tiene sus propias experiencias utilizando R. Estas son solo algunas de mis 'aventuras' usando R.

Utilizo R aprox. desde el 2003 para:

- Enseñanza
 - 1) dentro de algunos cursos de Estadística
 - 2) el proyecto Symbiosis (2006-2009) y el libro que salio de el
 - 3) el curso de Computación Estadística (+ SAS)
 - 4) en un libro introductorio en castellano que estoy escribiendo ←4
- Mi propia investigación en Estadística y colaboraciones
 - 1) simulaciones
 - 2) gráficos para ilustrar o contestar preguntas investigación ←1a
 - 2) implementar cosas que definimos (Polyplot) ←1b
- Colaboraciones con investigadores de otras especialidades, 2 ejemplos:
 - 1) uso del paquete TraMineR ←2
 - 2) código para periodogramas móviles ←3

Gráficos

R tiene grandes capacidades graficas y paquetes especializados como ggplot2, spatstat etc. Pero también podemos trabajar con el R básico.

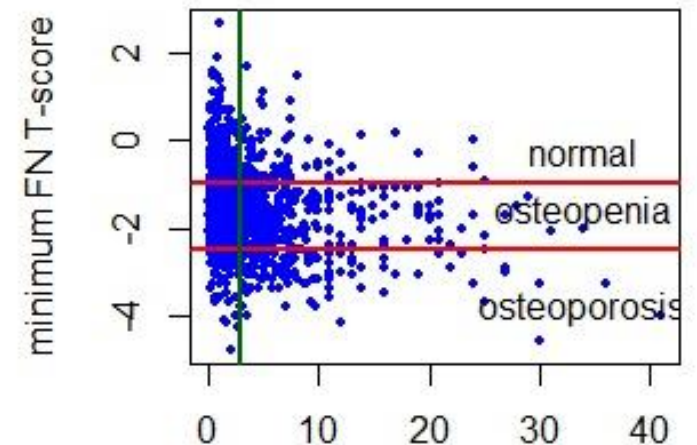
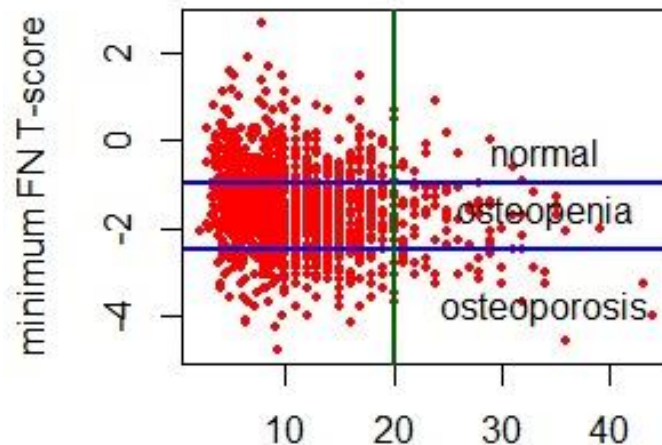
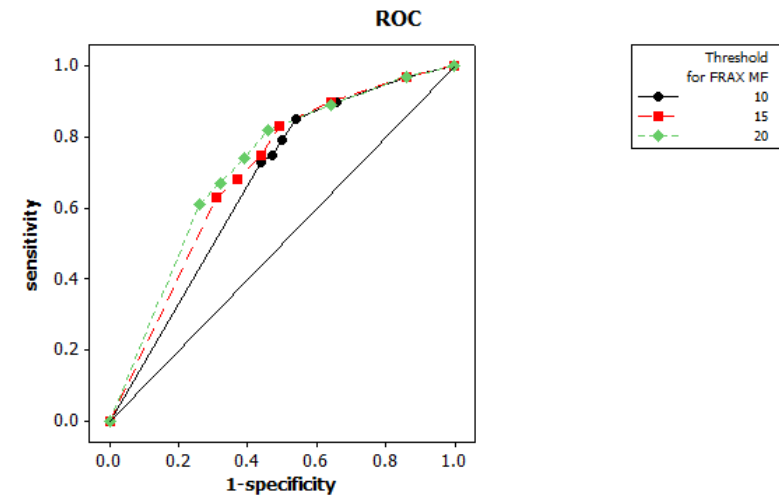
Tres ejemplos:

- Preparando un grafico estándar simple para una publicación nos permite añadir las anotaciones que lo hagan mas informativo
- Gráficos especiales implementados por paquetes en R, ejemplos del análisis espacial
- Creando un grafico nuevo (ejemplo el Polyplot) y escribiendo una función para que otros lo usen

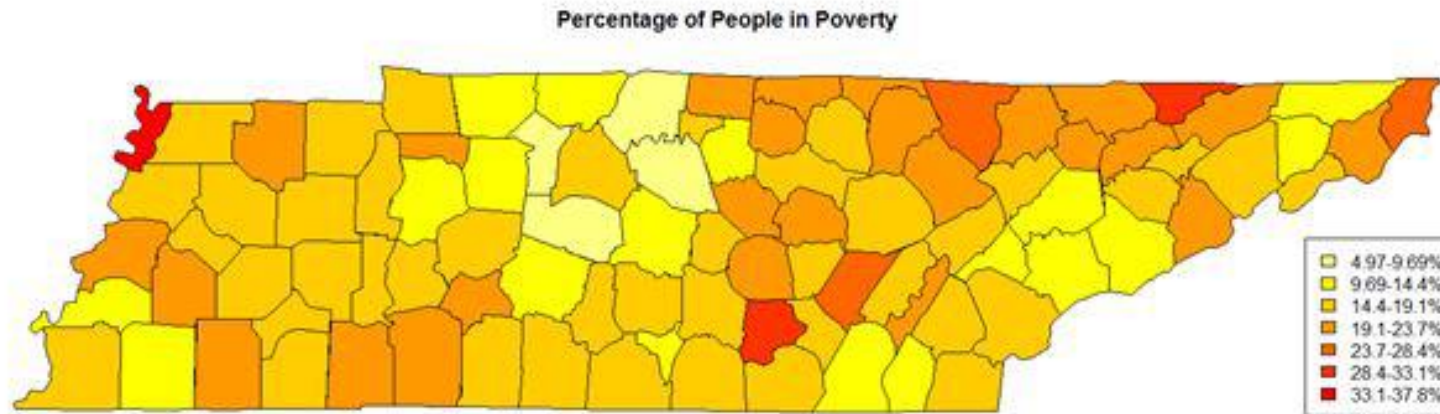
Preparando un grafico para publicación

- Escribimos código en R para hacer el grafico y borrando/aumentando líneas de código vemos facilmente diferentes versiones del grafico hasta que todos los colaboradores queden contentos

Hamdy, R.C., Seier, E., Whalen, K., Clark, W.A., Hicks, K. & Pigeon, T.B. (2018) **"FRAX calculated without BMD does not correctly identify Caucasian men with densitometric evidence of osteoporosis."** Osteoporosis International. Published online Feb 3, 2018



Mapas coroplético: % de personas bajo el nivel de pobreza en Tennessee por condado



```
poverty<-scan("poverty.txt")
hist(poverty)
pov.ints<-cut(poverty, breaks=7)
colors<-rev(heat.colors(7))
map("county", "Tennessee", fill=T, plot=T, col=colors[pov.ints])
title("Percentage of People in Poverty")
leg.txt <- c("4.97-9.69%", "9.69-14.4%", "14.4-19.1%", "19.1-23.7%",
"23.7-28.4%", legend("bottomright", leg.txt, horiz = FALSE, fill = colors)
```

Los datos están a nivel de cada distrito u otra división política, los colores se asignan de acuerdo a una escala utilizando una de las paletas en R

Trabajo hecho con Allison Foster, alumna de pregrado para su proyecto final sobre las disparidades en esperanza de vida entre condados en Tennessee y correlación espacial

Se utiliza el paquete **maps**. Se podría también hacer el programa comercial ArcGIS

puntos y su densidad (para Proyecto CRAWL)

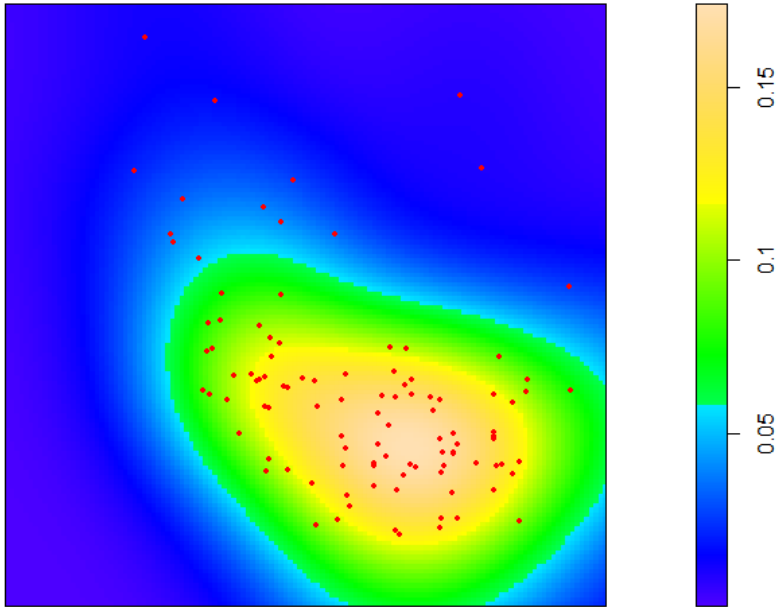
con el paquete **spatstat**

```
bees2pm<-read.table('OD1-coor-2-4pm.txt')
```

```
X<-bees2pm[,1] ; Y<-bees2pm[,2]
```

```
bees2<-ppp(X,Y,c(50,100),c(0,50)) ## definimos marco ;  
plot(density(bees2))
```

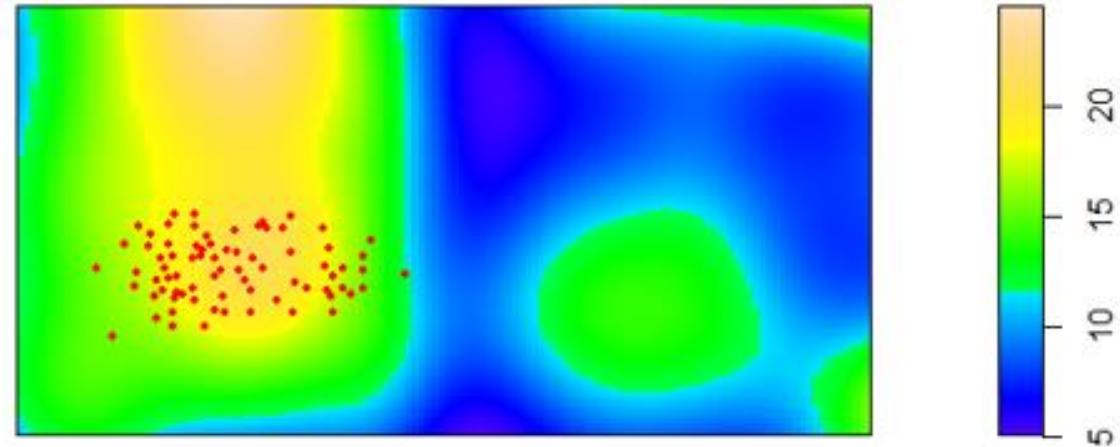
```
points(X,Y,pch=19,col='red',cex=0.5)
```



La densidad de un conjunto de puntos y los puntos de otros conjuntos de datos

Graficamos la densidad de todas las abejas y los puntos indican la localización de las danzas. Que hayan abejas es condición necesaria pero no suficiente, prefieren estar cerca a la puerta. No en todo sitio que hay alta concentración de abejas las exploradoras bailan.

density of all bees and dance locations



El Polyplot

- Cuando se define una nueva herramienta es conveniente incluir el código en R para calcularla para que otros puedan usarla. Cuando son varias herramientas se construye un ‘paquete’.
- En el 2011 publicamos un artículo en el American Statistician Vol. 65, No. 4, 258-261: Seier&Bonett **A polyplot for visualizing location, spread, skewness and kurtosis.**
- Igual que el “boxplot” representa gráficamente estadísticas descriptivas, pero tiene una mayor cantidad de estadísticas de posición y dispersión, que permite percibir también la kurtosis de la distribución al comparar visualmente varias estadísticas de dispersión. El polyplot incluye la **media**, **mediana**, **los cuartiles**, **extremos**, **rango inter-cuartílico**, **desviación media con respecto a la mediana**, **desviación standard**, y **rango**.
- La intención inicial no fue hacer una variante del boxplot pero en investigación las cosas a veces siguen un camino inesperado.
- La función polyplot <http://faculty.etsu.edu/seier/Rpolyplot.txt>. Primero se copia la función en la ventana de la sesión de R, luego bastará con aplicarla a un conjunto de datos de una variable.

La función polyplot

```
polyplot<- function(x) {
  n<-length(x);n2=floor(n/2);if(n%%2==0) n2m1=n2+1 else n2m1=n2+2;
  sx<-sort(x);n2m1=n2+1;first<-sx[1:n2];second<-sx[n2m1:n];
  m1<-mean(first);m2<-mean(second);xmean<-mean(x);difm<-(m2-m1)/2;xmed<-median(x);
  mi<-min(x);ma<-max(x);s<-sd(x);delta<-(sum(abs(x-xmed)))/n;
  cinco<-summary(x);
  q1<-cinco[2];q3<-cinco[5];
  ##par(mfcol=c(2,1))
  layout(matrix(c(1,2)),heights=c(3,2), widths=c(1,1))
  hist1<-hist(x,plot=
  TRUE, main= 'Histogram ',col= 'beige');nh=length(hist1$breaks);
  xa=hist1$breaks[1];xb=hist1$breaks[nh];par(mar=c(2, 4,2, 4) + 0.1);
  mm=max(abs(xa),abs(xb));fskew<-(sum((x-xmean)^3));
  plot(x,x,'n',xlab=' ',ylab=' ',ylim=c(-0.7,3.6),xlim=c(xa-0.1*mm,xb+0.1*mm),
  axes=FALSE, main = 'Polyplot');
  axis(1,at=hist1$breaks);
  if (fskew>=0) {segments(m1,2,xmean,2, lty='dashed');
  }else{ segments(m1,2,xmean,2, lwd=2)};
  if(fskew>=0) {segments(xmean,2, m2 ,2, lty='solid',lwd=2);
  }else {segments(xmean,2, m2 ,2, lty='dashed')};
  segments(m2,2,xmean+s,1, lty='dotted');segments(m1,2,xmean-s,1, lty='dotted');
  segments(xmean-s,1, mi,0,lty='dotted');segments(xmean+s,1, ma,0,lty='dotted');
  segments(q3,3,m2,2, lty='dotted');segments(q1,3,m1,2, lty='dotted');
  if (fskew>=0) {segments(xmean,1,xmean+s,1, lwd=2)
  }else{ segments(xmean,1,xmean+s,1, lty='dashed')};
  if (fskew>=0) {segments(xmean-s,1,xmean,1, lty='dashed')
  }else{ segments(xmean-s,1,xmean,1, lwd=2)};
  segments(mi,0,ma,0,, lwd=2);points(xmean,2,pch=19,col='black');
  points(xmean,1,pch=19,col='black');points((mi+ma)/2,0,pch=3,col='black');
  points(m1,2,pch=19, col='black');points(m2,2,pch=19, col='black');
  points(xmed,3.1,pch=25,col='black',bg='black');
  points(q1,3.1,pch=25,col='black',bg='black');
  points(q3,3.1, pch=25,col='black',bg='black');points(mi,0,pch= 15 );
  points(ma,0,pch= 15 );segments(xmed,3,xmean,2, lty='dotted');
  segments(q1,3,q3,3, lwd=2); rmi=round(mi,4);rma=round(ma,4);
```

```
## PRINTING STATISTICS
iqr=q3-q1;riqr=round(iqr,2);rdelta=round(delta,2);
rs=round(s,2);R=ma-mi;rR=round(R,2);
rq1=round(q1,2);rq3=round(q3,2);rxmed=round(xmed,2);
fence2<-q3+(1.5*iqr)
fence1<-q1-(1.5*iqr)
text(fence1,0, '(')
  text(fence2,0,')')
rxmean=round(xmean,2);
mtext('MAD',side=4,line=2,at=2,cex=0.8,las=1,adj=0);
mtext('IQR',side=4,line=2,at=3.5, cex=0.8, las=1,adj=0);
mtext('IQR/2',side=4,line=2,at=2.8, cex=0.8, las=1,adj=0);
mtext('SD',side=4,line=2,at=1.2, cex=0.8,las=1,adj=0);
mtext('R/2',side=4,line=2,at=0.4, cex=0.8,las=1,adj=0);

mtext('R',side=4,line=2,at=-0.4, cex=0.8,las=1,adj=0);
mtext(rdelta,side=4,line=1,at=2,cex=0.8,las=1, adj=1);
mtext(riqr,side=4,line=1,at=3.5, cex=0.8, las=1, adj=1);
mtext(riqr/2,side=4,line=1,at=2.8, cex=0.8, las=1, adj=1);
mtext(rs,side=4,line=1,at=1.2, cex=0.8,las=1, adj=1);
mtext(rR,side=4,line=1,at=-0.4, cex=0.8,las=1, adj=1);
mtext(round(rR/2,2),side=4,line=1,at=0.4, cex=0.8,las=1, adj=1);
mtext('Q1      ',side=2,line=2,at=3.5,cex=0.8,las=1,adj=1);
mtext('Q2      ',side=2,line=2,at=3,cex=0.8,las=1,adj=1);
mtext('Q3      ',side=2,line=2,at=2.5,cex=0.8,las=1,adj=1);
mtext('mean   ',side=2,line=2,at=1.5,cex=0.8,las=1,adj=1);
mtext('n       ',side=2,line=2,at=0.5,cex=0.8,las=1,adj=1);
mtext(rq1,side=2,line=1.5, at=3.5,cex=0.8,las=1,adj=0);
mtext(rxmed,side=2,line=1.5, at=3,cex=0.8,las=1,adj=0);
mtext(rq3,side=2,line=1.5, at=2.5,cex=0.8,las=1,adj=0);
mtext(rxmean,side=2,line=1.5, at=1.5,cex=0.8,las=1,adj=0);
mtext(n,side=2,line=1.5,at=0.5,cex=0.8,las=1,adj=0)
}
```

Ejemplo 1 : pulso de 210 estudiantes
<http://faculty.etsu.edu/seier/pulserate.txt>

Ejemplo 2: mil valores
 simulados con $U(0,1)$

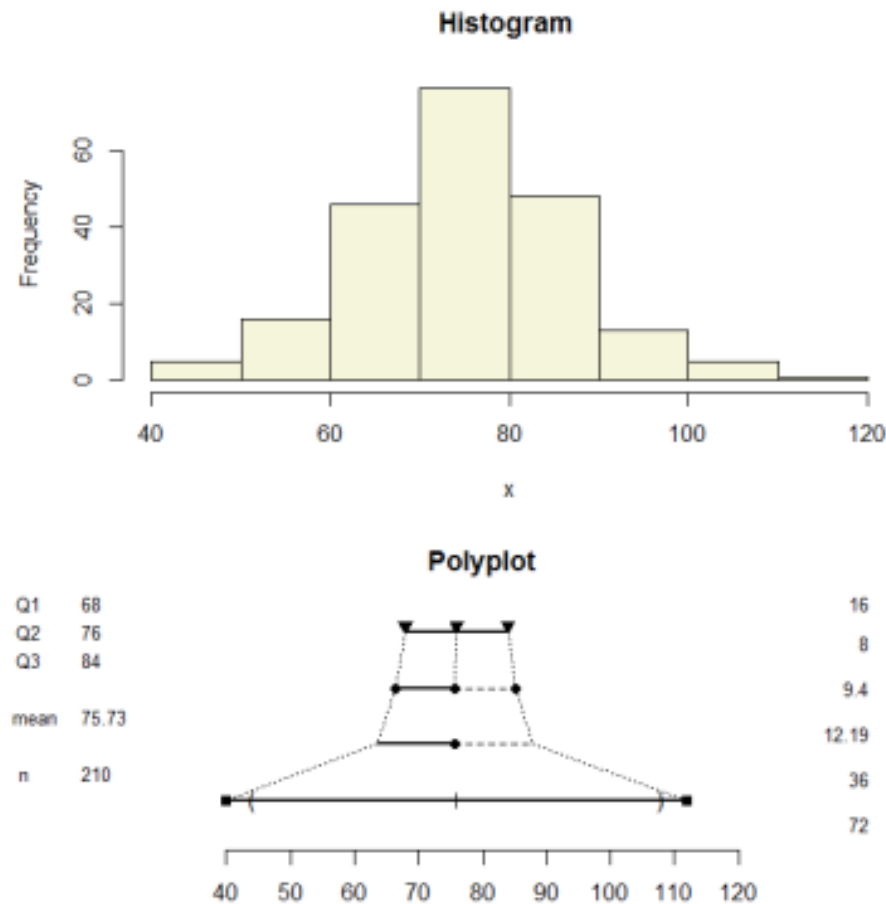
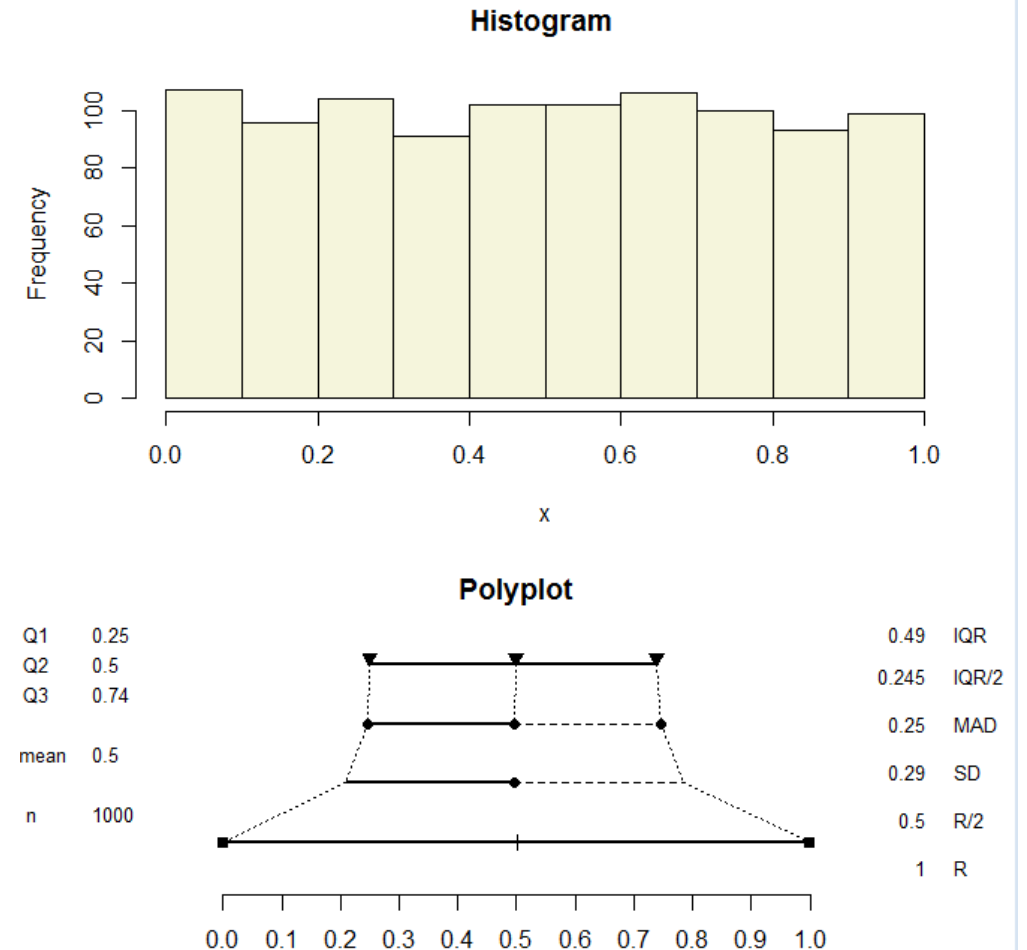


Figura 1 . El Polyplot para el pulso de 210 estudiantes.



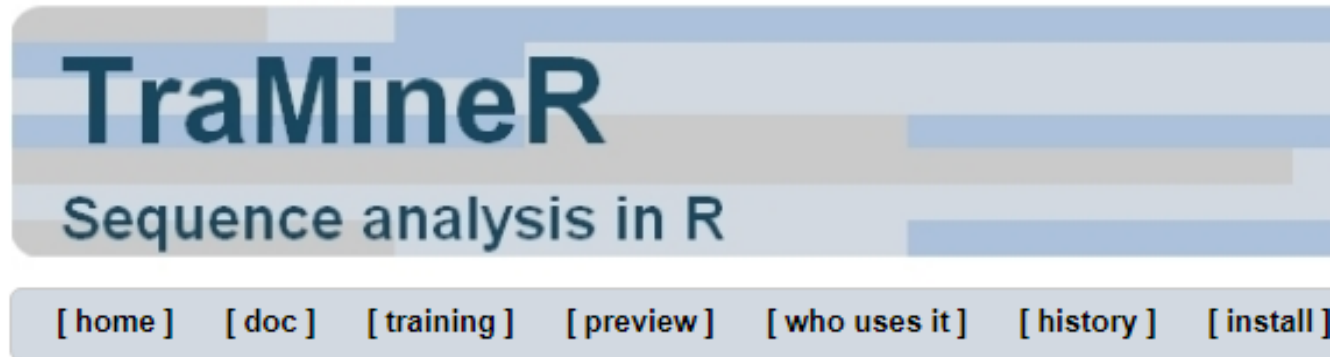
Uso del paquete TraMineR

- Por que lo usamos?
 - Para que sirve el TraMineR?
 - Una particularidad del TraMineR
 - Que hicimos con los resultados del TraMineR?
-
- Este fue un trabajo conjunto con un alumno de pre-grado Adam Chase quien luego fue a estudiar un master en Business Analytics
- El trabajo fue parte del proyecto CRAWL y también el proyecto final del bachillerato de Adam.

Por que usamos TraMineR?

- Un biólogo con el que he colaborado un largo tiempo, Karl Joplin, quien estudia el comportamiento de un tipo de *mosca Sarcophaga crasipalpis* tenia datos de secuencias de comportamientos y el programa T-pattern que quería usar era caro y me pregunto si habría algo equivalente en R para analizar esos datos
- El objetivo era ver si las secuencias de comportamientos difieren dependiendo en el sexo del animal y en el hecho de estar solo o en un par en el mismo ambiente así como la edad (en días) de la mosca.
- Buscamos entonces un paquete en R que analizara secuencias de las categorías de una variable categórica. Puede que actualmente existan otros paquetes pero en ese momento (2014) encontramos el TraMineR

TraMineR <http://traminer.unige.ch/>



Nosotros utilizamos una versión antigua de TraMineR pero hay una versión inclusive de Enero 2020 que puede realizar mas tareas.

TraMineR: a toolbox for exploring sequence data

TraMineR is a [R](#)-package for mining, describing and visualizing sequences of states or events, and more generally discrete sequence data. Its primary aim is the analysis of biographical longitudinal data in the social sciences, such as data describing careers or family trajectories. However, most of its features also apply to many other kinds of categorical sequence data. They include:

TraMineR fue escrito por expertos en demografía y sociología de la Universidad de Ginebra para estudiar secuencias de variables categóricas en las ciencias sociales

Gabadinho, A., Ritschard, G., Müller, N.S. & Studer, M. (2011), Analyzing and visualizing state sequences in R with TraMineR, *Journal of Statistical Software*. Vol. 40(4), pp. 1-37.

Que es una secuencia para una variable categorica?

- Para una variable cuantitativa una secuencia de observaciones produce una 'serie de tiempo', una secuencia de valores numéricos
- Para una variable categórica con categorías A,B,C,D,E una secuencia puede por ejemplo verse así AACDDDDAE
- Lo que se estudia en una secuencia de este tipo es su complejidad, por ejemplo, es muy diferente una secuencia AAAAAA que una CBADE
- Hay diferentes estadísticas o métricas para cuantificar la complejidad

El alfabeto y el tipo de datos: Estados y duraciones

Behavior	Abbreviation
Approach	App
Avoid	Avo
Bobbing	Bob
Grooming	Gro
Hold	Hol
Immobilized	Imm
Jump	Jum
Low Intensity Agression	Low/Lia
Lunge	Lun
Mating	Mat
Retreat	Ret
Standing	Sta
Stilt	Sti
Turn-toward	Tur
Upside-down	Ups
Walking	Wal
Wrestle	Wre

App - Avo - Bob - Gro - Hol

(App, 32) - (Avo, 56) - (Bob, 2) - (Gro, 41) - (Hol, 24)

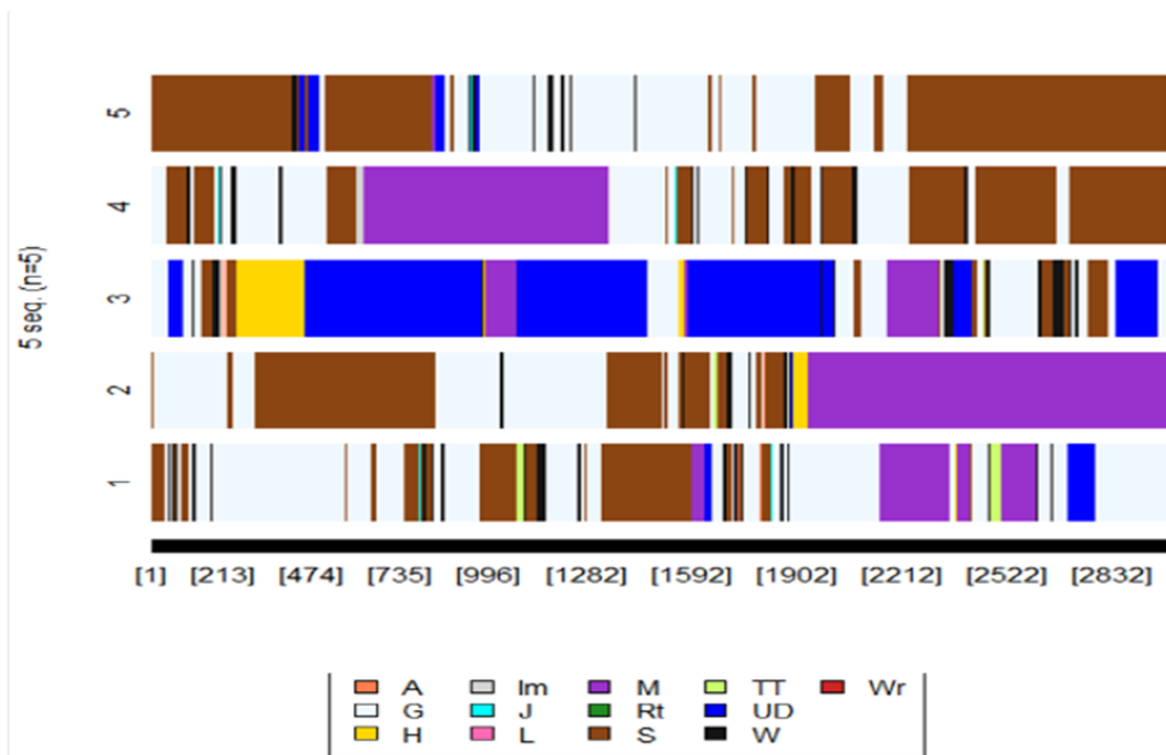


Grafico para 5 individuos

Estadísticas para cuantificar la complejidad

- 1) Numero de transiciones o cambios de estado,4 en AACDDDDAE
- 2) Turbulencia, toma en cuenta las duraciones de cada sub-secuencia

A la turbulencia no le importa si todos los posibles estados ocurren o no en una secuencia.

De estas 4 secuencias la primera es la mas 'turbulenta' y la ultima es la menos turbulenta porque no hay dos subsecuencias con el mismo estado

$$a = (\text{Sta}, 48) - (\text{Jum}, 47) - (\text{Wal}, 31) - (\text{Ups}, 18) \quad b = (\text{Sta}, 35) - (\text{Jum}, 29) - (\text{Sta}, 24) - (\text{Wal}, 30)$$
$$c = (\text{Sta}, 8) - (\text{Jum}, 36) - (\text{Wal}, 44) - (\text{Sta}, 12)$$
$$d = (\text{Sta}, 85)$$

- 3) Entropia (Ley de Shannon).- La entropía es máxima cuando cada estado aparece el mismo numero de veces o el mismo numero de tiempos. La entropía es mínima (0) cuando el individuo permanece siempre en el mismo estado. Valores altos indican balance entre estados, cerca al 0 indican no balance
- 4) Indice de complejidad El minimo valor es 0 y sucede cuando solo un estado sucede AAAAAAAAAAAAAA

El máximo valor es 1 y se logra cuando:

- i). Cada estado del alfabeto ocurre
- ii). El tiempo para cada estado es el mismo
- lii) El numero de transiciones es el maximo possible.

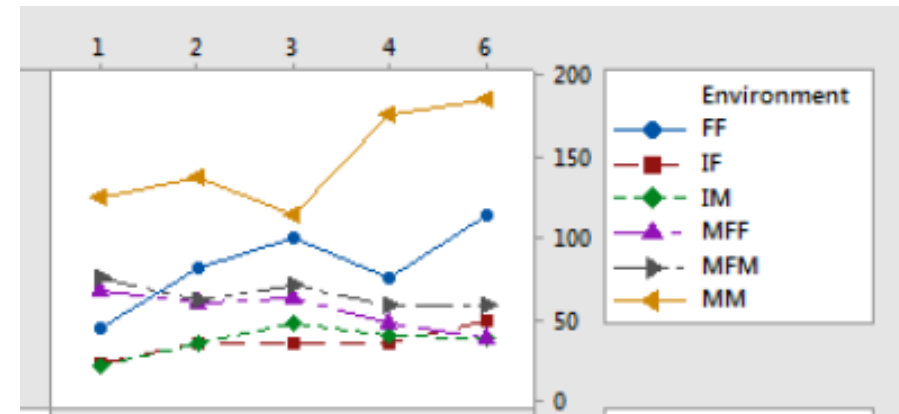
El análisis

- Se comparo la complejidad de las secuencias obtenidas bajo diferentes condiciones
- Los grupos a comparar se definieron por sexo y condición (solo o en pares) IM IF MF FF MM
- También se considero el factor dia porque en el caso de las moscas los comportamientos pueden estar asociados al grado de desarrollo de los individuos (alcanzan madurez sexual al tercer dia)
- Se uso un modelo ANOVA (dos criterios con interacción) para ver si había diferencia en complejidad entre los grupos

General Linear Model: Transitions versus Environment, Day

Analysis of Variance

Source	DF	Seq SS	Adj SS	Adj MS	F-Value	P-Value
Environment	5	753733	747648	149530	59.04	0.000
Day	4	43097	18226	4556	1.80	0.128
Environment*Day	20	114588	114588	5729	2.26	0.002
Error	402	1018202	1018202	2533		
Total	431	1929620				



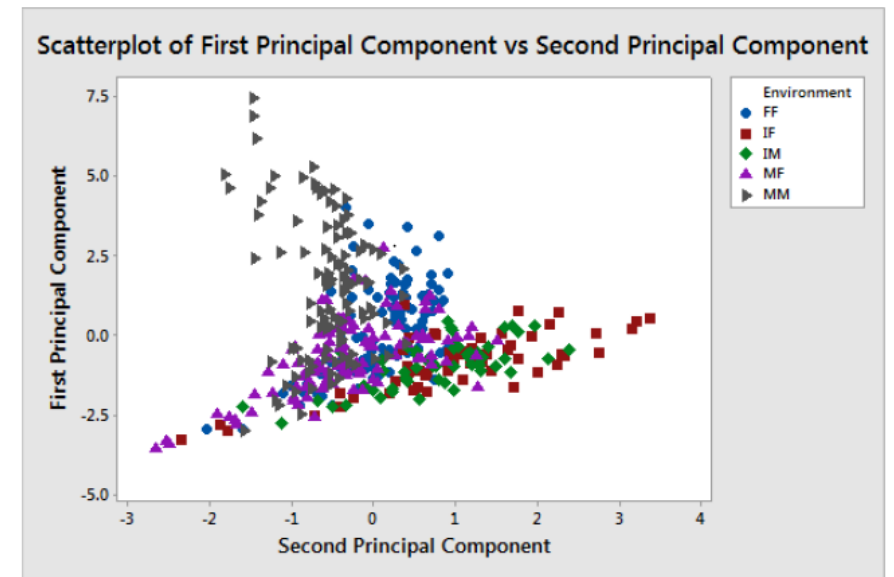
Reduciendo la dimensión de indicadores

- También aplicamos componentes principales para ver si podíamos trabajar con 2 estadísticas de complejidad en lugar de 4. Fijándonos en los coeficientes vemos que Z1 es casi equivalente a un promedio ponderado de los 4 y Z2 es un contraste entre entropía y la combinación de turbulencia y numero de transiciones

$$Z_1 = (0.549)(\text{NumberofTransitions}) + (0.287)(\text{Entropy}) \\ + (0.549)(\text{Turbulence}) + (0.562)(\text{Co})$$

$$Z_2 = (-0.259)(\text{Numberof Transitions}) + (0.930)(\text{Entropy}) \\ + (-0.258)(\text{Turbulence}) + (0.030)(\text{ComplexityIndex})$$

Eigenvalue	3.1202	0.8579
Proportion	0.780	0.214
Cumulative	0.780	0.995



Calculando indicadores de complejidad con R

- El calculo en si de los indicadores es directo con TraMineR una vez que los datos son puestos en forma apropiada

```
xtrans<-seqtransn(flyseqwdr)
```

```
xent<-seqient(flyseqwdr)
```

```
xtur<-seqST(flyseqwdr)
```

```
xcim<-seqici(flyseqwdr)
```

- El mayor trabajo esta en la preparación de los datos, el alfabeto, colores a usar. El formato de los datos algo interesante en el uso de TraMineR que vale la pena comentar por inusual

Usualmente en análisis de datos, los datos pueden venir en forma desordenada y el problema es ponerlos en columnas o formato 'tidy' para analizarlos.

En este caso teníamos los datos en columnas (comportamiento y duración) pero TraMineR pide (o al menos pedía en el 2014) los datos en un formato extraño.

`(EM,4)-(TR,2)-(EM,64)`

Felizmente con R podemos escribir código que transforme el formato en que se ingresan los datos

```
f1wd<-read.table('M-MFd4-1.txt')
behav<-f1wd[,1]
duratr<-round(f1wd[,2])
f1bdr<-paste(behav,duratr,sep=",")
f1bdpr<-paste("(",f1bdr,sep="")
f1bdppr<-paste(f1bdpr,")",sep="")
fly1dr<-noquote(f1bdppr)
fly1wdr<-paste(fly1dr,collapse="-")
```

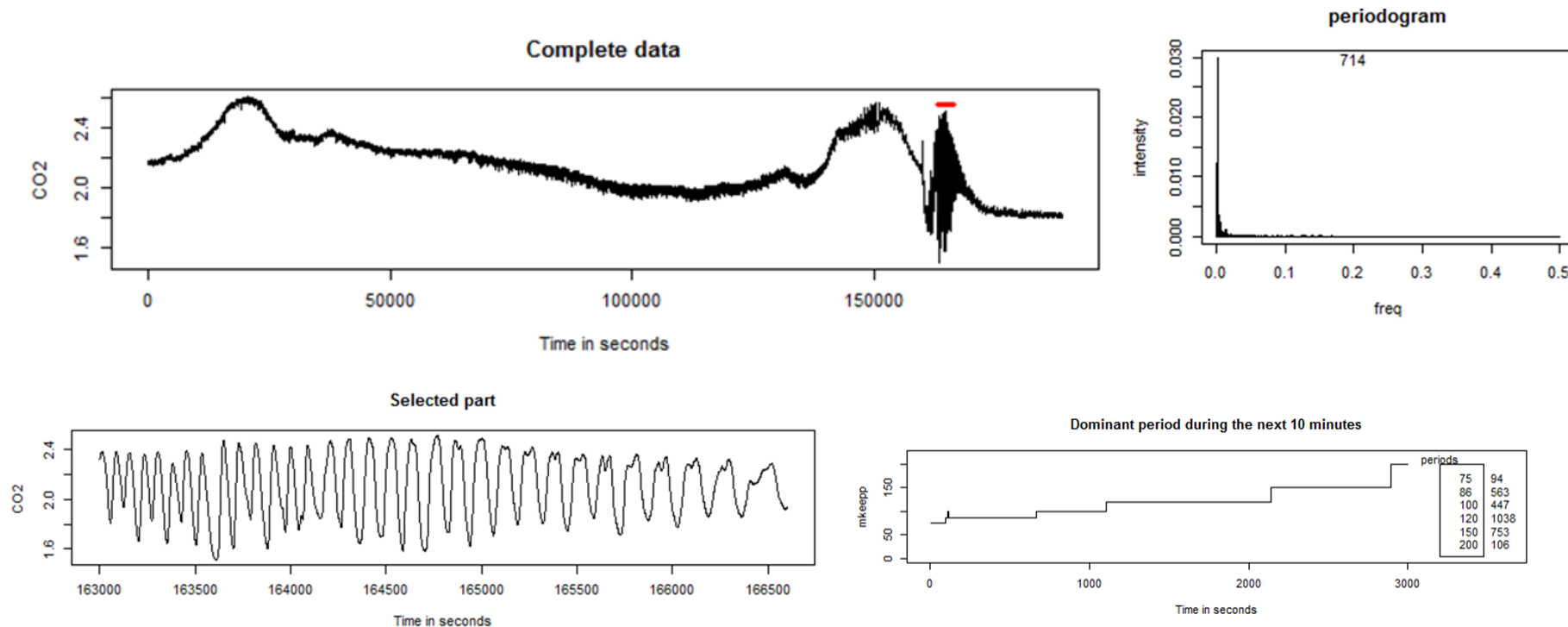
Bibliografía para los interesados en el análisis de secuencias para una variable categórica

- [1] C. H. Elzinga. Complexity of Categorical Time Series. *Sociological Methods Research*, 38:463–481, 2010.
- [2] C. H. Elzinga and A. C. Liefbroer. De-standardization of Family-Life Trajectories of Young Adults: A Cross-National Comparison Using Sequence Analysis. *Springer Science+Business Media*, 23:231–234, 2007.

A. Gabadinho, G. Ritschard, N. Muller, and M. Studer. Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, 40:1–37, 2011.

Cuando alguien quiere repetir un análisis muchas veces mejor le escribimos una función. El caso del periodograma móvil

K. Joplin tenía series de tiempo (nivel de CO2 en la respiración de una larva) muy cambiantes a lo largo del tiempo y quería poder escoger a voluntad una porción y analizarla, escribimos una función que producía los siguientes gráficos. El único argumento ('input') de la función eran el tiempo de inicio y termino del intervalo a analizar. El periodograma móvil cubría intervalos de 10 minutos (el CO2 se media cada segundo). La función tenía como 90 líneas de código pero para aplicarla, Karl solo tenía que reemplazar los argumentos por números en `partperio(x,froma,tob,howmany,myz)`. Preparamos también versiones adicionales para porciones no estacionarias en que se ajusta primero un modelo de regresión y se trabaja con los residuos



R desde el comienzo

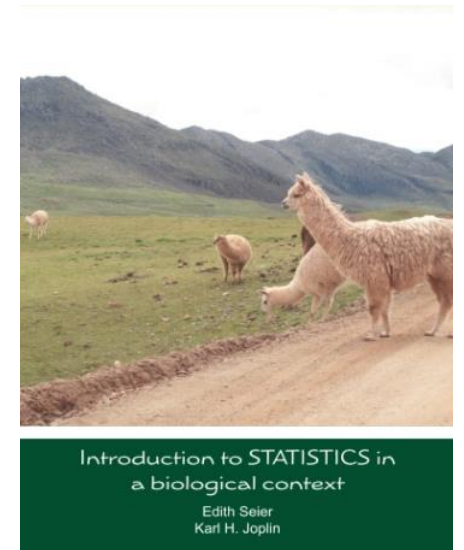
La experiencia del proyecto Symbiosis-

El libro del 2011 en ingles Introduction to statistics in a biological context

- el proyecto del libro para el Perú. 'Introducción al mundo de la estadística'. La idea es de introducir al lector a la estadística pero dando una visión de cosas que vienen después y teniendo en cuenta las discusiones actuales sobre la practica y enseñanza de la estadística

Algunos capítulos se pueden utilizar en forma independiente para minicursos sobre un tema específico (descriptiva, probabilidad, inferencia etc.)

Si alguien quisiera utilizar el libro para uno o varios cursos o talleres podría usar el software de su preferencia para la practica PERO se incluyen los comandos en R para el que desee aprovechar la oportunidad de aprenderlo. Igual que con un idioma extranjero, mientras más estamos expuestos a el lo aprendemos mejor.



Idea general con respecto a R en el libro

Instrucciones básicas sobre el lenguaje en el apéndice y se incluyen comandos para los temas que aprendemos en el momento y el código (script) que se uso para obtener determinado análisis o figura.

Se tratan de utilizar datos del Peru o datos propios de trabajos en los que he participado

Se utiliza principalmente R básico y paquetes solo cuando indispensable (por ejemplo ggplot2 cuando se habla de visualización)

Sobre el libro en general: No tiene un fin comercial sino ofrece una visión algo personal de la estadística al que recién se inicia

He recibido mucho consejo de amigas en particular Ysela Agüero y Emma Cambillo sobre el contenido estadístico del libro.

Todavía me falta terminarlo, a ver si me apuro ...😊

Capítulos del libro en castellano

1. Que es la estadística
2. Diseño de experimentos
3. Planeamiento y ejecución de encuestas
4. Explorando datos de variables cuantitativas
5. Estadísticas y gráficos para variables categóricas
6. Graficando con respecto a tiempo y espacio
7. Exploración y visualización de datos
8. Trabajando con archivos de datos
9. Indicadores
10. Modelos de regresión
11. Introducción a la probabilidad
12. Introducción a la inferencia

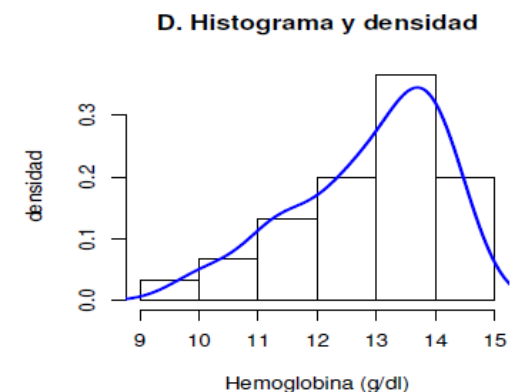
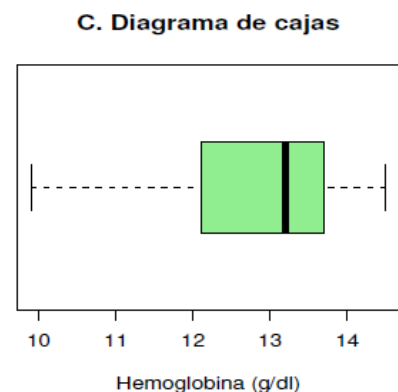
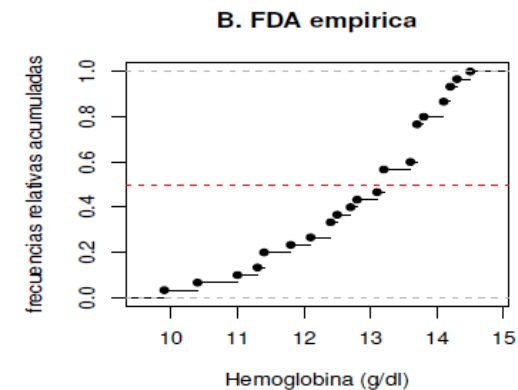
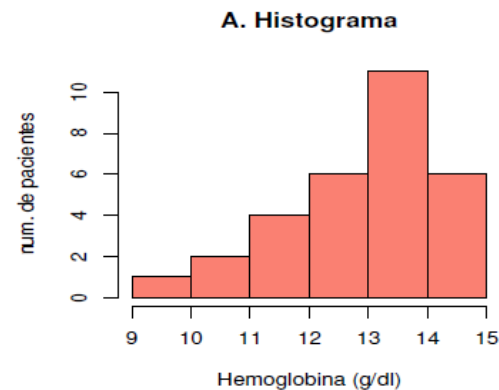
En el libro incluiremos comandos en R para calcular las estadísticas que en el se incluyen.

Ahora veremos algunos ejemplos de figuras cuyo correspondiente código en R se incluye en el libro. El lector puede o no utilizar R pero si desea aprovechar la oportunidad para aprender R puede utilizar esos códigos como referencia.

El plan es preparar una pagina web para el libro en que pongamos esos scripts para ser usados.

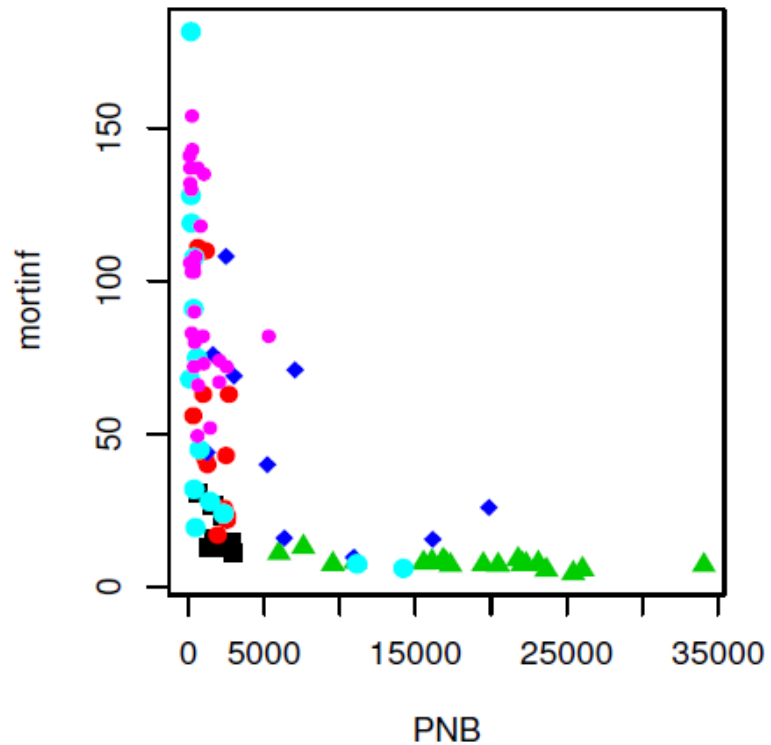
Para el libro en ingles pusimos algunos comandos en R en <https://faculty.etsu.edu/seier/IntroStatsBioBook.htm> pero ahora lo pensamos hacer de una forma mas completa.

En el capítulo 4

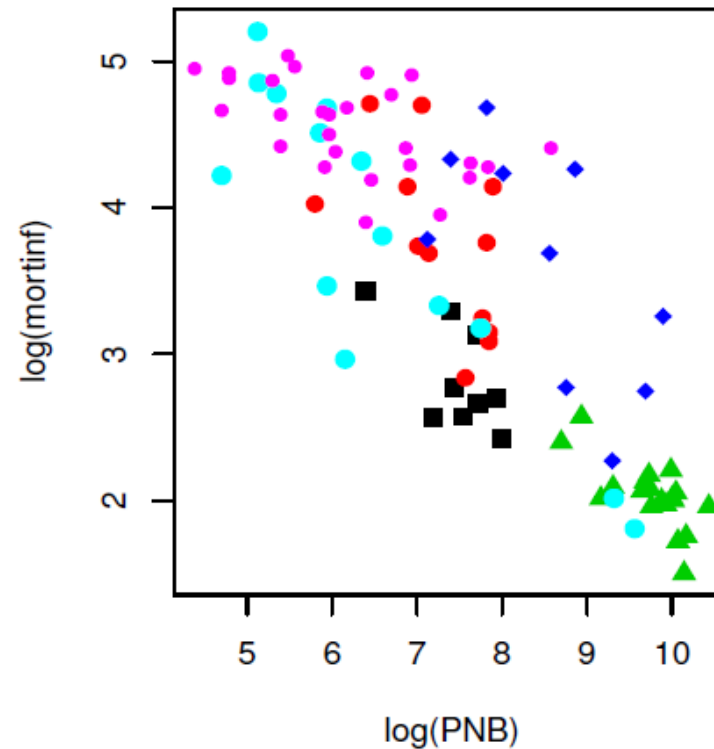


```
hbbmi<-read.table("http://faculty.etsu.edu/seier/data/HBBMI30.txt",skip=1,
  col.names=c("Hb","BMI"))
attach(hbbmi)      ## conecta los nombres de variables con los datos
par(mfrow=c(2,2))  ## indica preparar figura con 4 graficos
hist(Hb,main='A. Histograma',xlab='Hemoglobina (g/dl)',
  ylab='numero de pacientes',col='salmon')
plot(ecdf(Hb),main='B. FDA empirica',ylab='frecuencias relativas acumuladas',
  xlab='Hemoglobina (g/dl)')
abline(h=0.5,col='red',lty=2) ## linea horizontal en y=0.5
boxplot(Hb, main='C. Diagrama de cajas', col='lightgreen', horizontal=TRUE,
  lab='Hemoglobina (g/dl)')
## para graficar la densidad
hist(Hb, freq=F, breaks = 6,main="D. Histograma y densidad",xlab='Hemoglobina (g/dl)',
  ylab='densidad')
lines(density(Hb), lwd=2, col="blue") ## agregamos la densidad al histograma
summary(Hb)      ## calcula estadisticas basicas
```

Mortalidad infantil y PNB

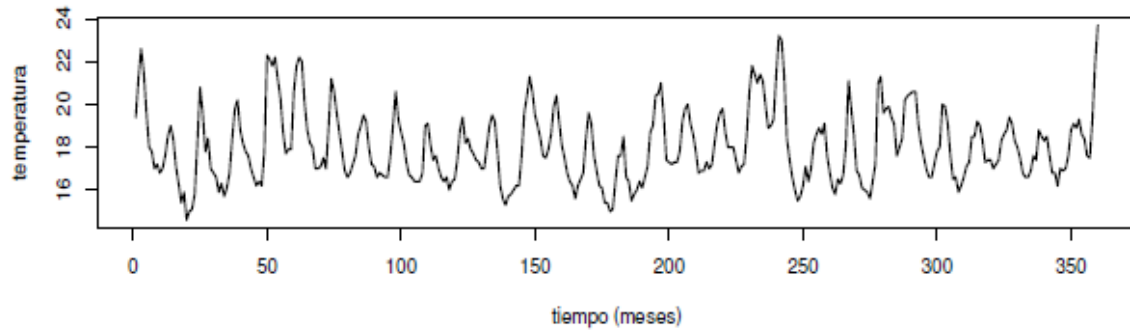


Graficando los logaritmos

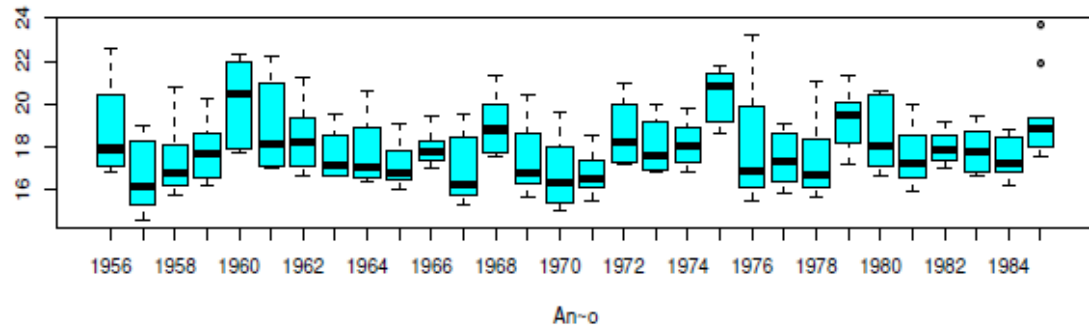


```
plot(PNB,mortinf,pch=grupo+14,cex=1.2,col=grupo,
main='Mortalidad infantil y PNB')
plot(log(PNB),log(mortinf),pch=grupo+14,cex=1.2,col=grupo,
main='Graficando los logaritmos')
```

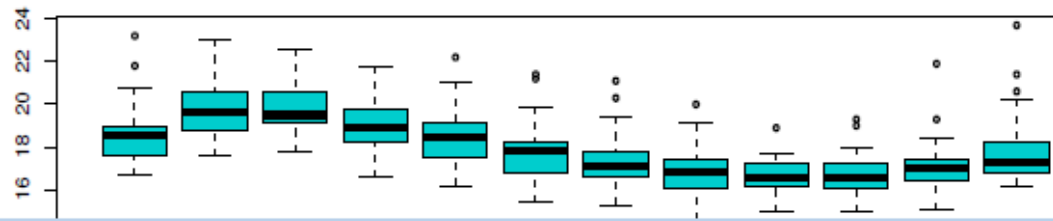
A. Temperatura mensual promedio (TMP) del mar del Callao 1956-1985



B. TMP por año



C. TMP para cada uno de los 12 meses

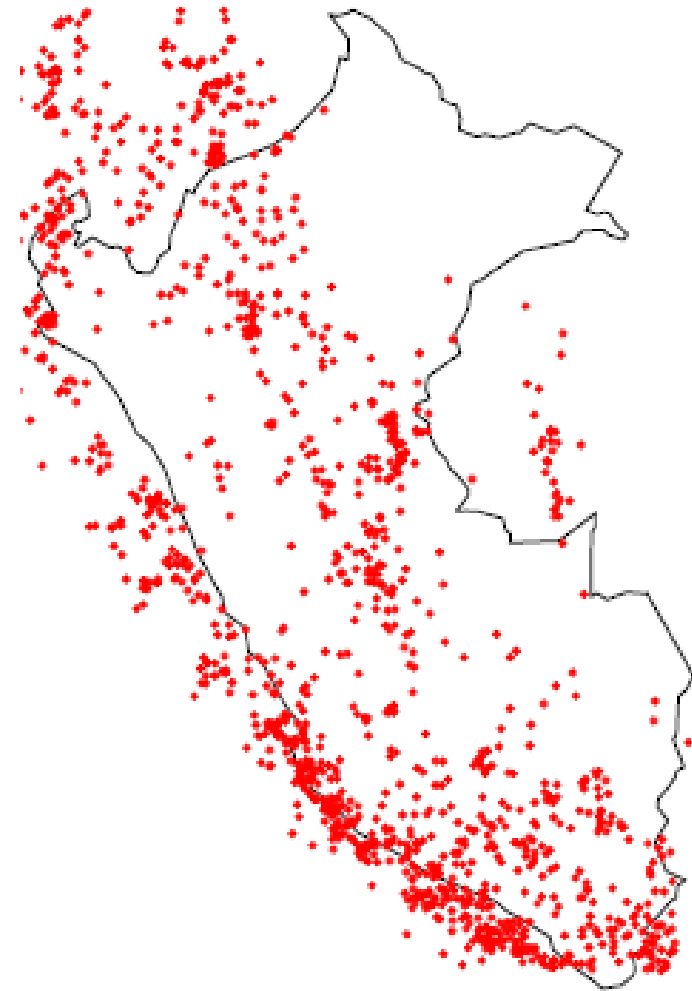


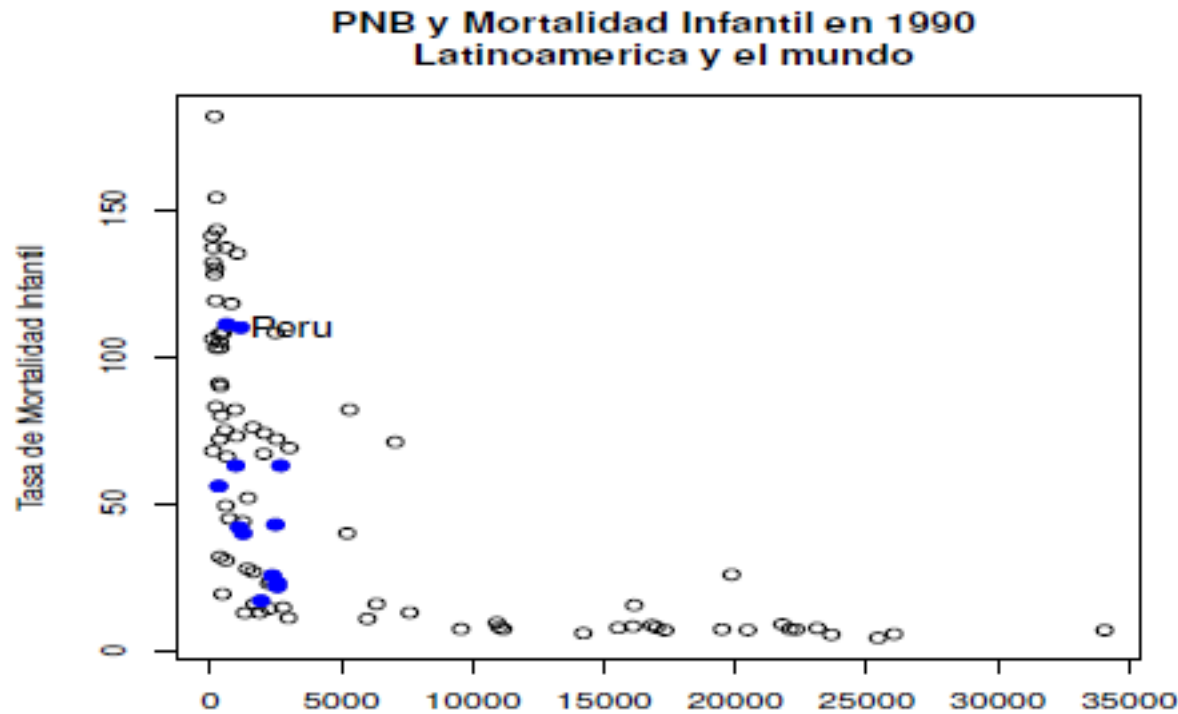
En el capítulo 6

```
## script tempemar
martemp<-read.table('temperaturamar.txt',header=TRUE)
attach(martemp) ## asigna los nombres de las variables a cada columna de datos
par(mfcol=c(3,1)) ## prepara figura con 3 filas y una columna
ts.plot(temperatura,xlab='tiempo (meses)',main='A. Temperatura mensual promedio
(TMP) del mar del Callao 1956-1985')
boxplot(temperatura~A,main='B. TMP por año', xlab='Año', col='cyan')
boxplot(temperatura~M,main='C. TMP para cada uno de los 12 meses',
xlab='Meses (1=Enero,... 12=Diciembre)',col='cyan3')}
```

Cap 6

```
## sismos magnitud >=5 en el Peru 1974-2012
## instalar primero el paquete maps
library(maps) ## se carga el paquete
par(mfcol=c(1,1)) ## un solo grafico en la figura
map("world","Peru") ## mapa del Peru
sismos<-read.table('sismosperu5.TXT',header=TRUE)
attach(sismos) ## conecta nombres de variables con datos
points(longitud,latitud,pch=19,cex=0.5,col='red') ## ag
```





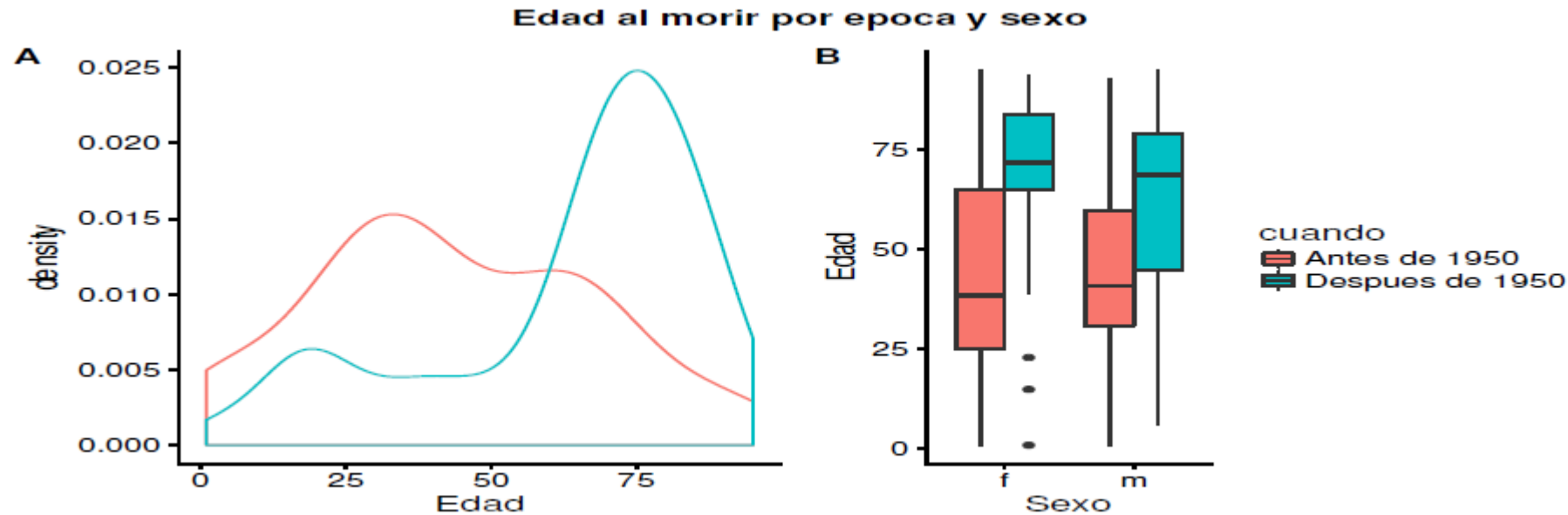
En el capitulo 7

Identificando algunos individuos

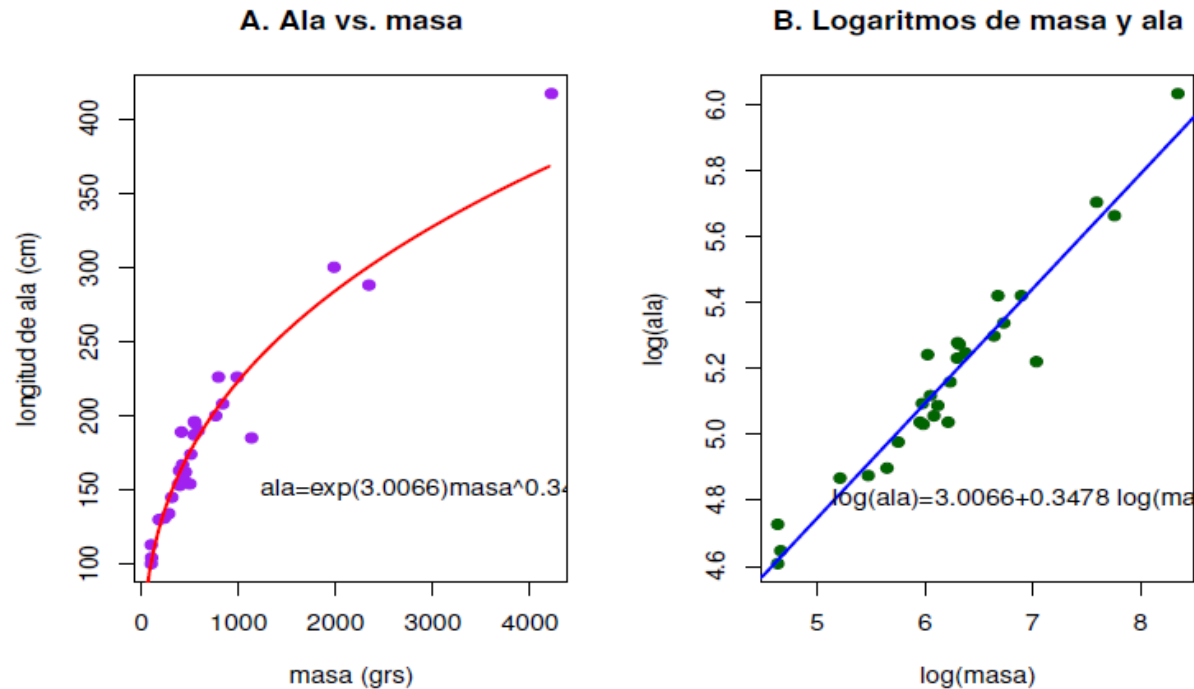
```
##script Fig7-infmortLA.R
idemeco90<-read.table("https://ww2.amstat.org/publications/jse/datasets/poverty.dat.txt"
  col.names=c("tnata","tmort","mortinf","EVhombres","Evmujeres","PNB","grupo","pais"), r
attach(idemeco90)
head(idemeco90)
plot(PNB,mortinf,pch=19,main="PNB y Mortalidad Infantil en 1990 \n Latinoamerica y el mur
indLA<-which(grupo==2) ## identificar a los de Latinoamerica
points(PNB[indLA],mortinf[indLA],col='blue',pch=19)
text(PNB[20],mortinf[20],"Peru",adj=-0.1,cex=1.2)
```


Usando ggplot2

- especificación de los datos (indicar con que conjuntos de datos se va a trabajar)
- mapeo estético o identificación de las variables a ser proyectadas en el gráfico y en que roles
- objetos geométricos (lineas, puntos, rectángulos) a utilizarse.

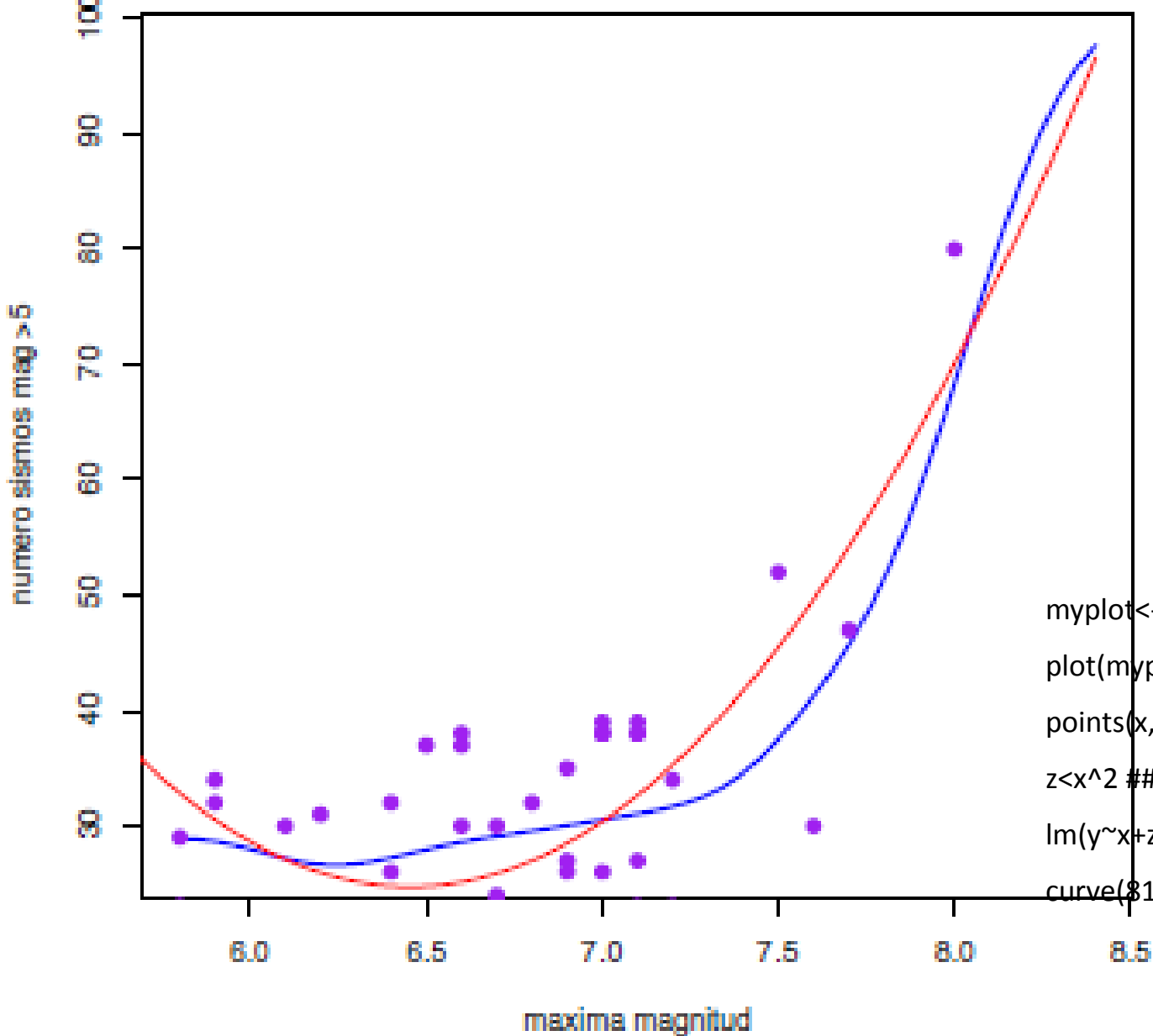


Cap 10 Regresion



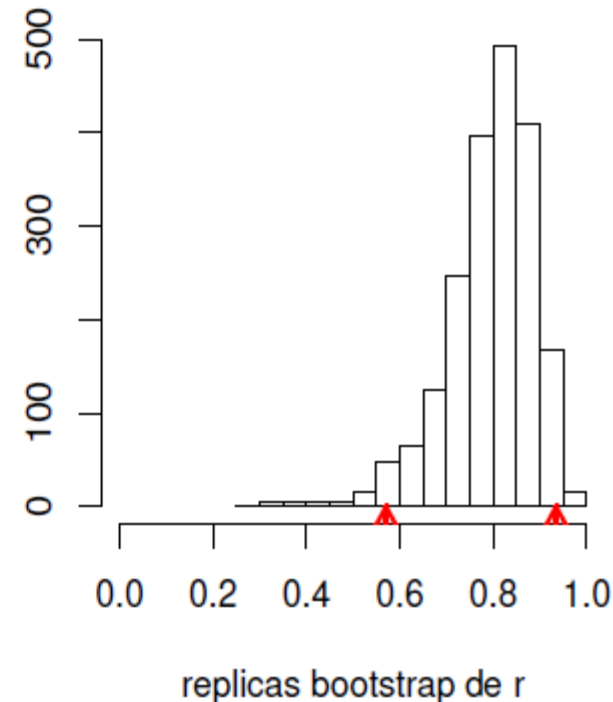
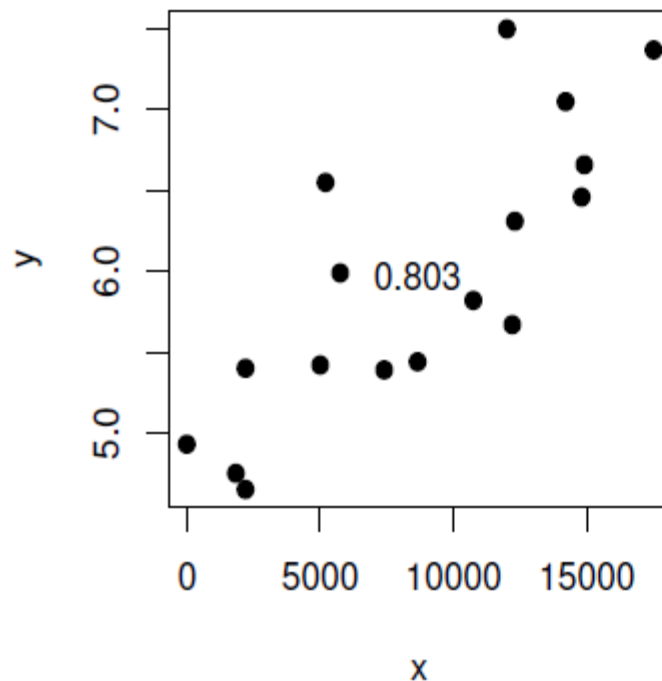
```
par(mfcol=c(1,2))
plot(masa,ala,pch=19,col='purple',main='A. Ala vs.
masa',xlab='masa (grs)',
ylab='longitud de ala (cm)')
curve(20.21854*(x^0.3478),0,4200,lwd=2,col='red',add=TRUE)
text(3000,150,'ala=exp(3.0066)masa^0.3478')
plot(log(masa),log(ala),pch=19,col='darkgreen',main='B.
Logaritmos de masa y ala')
abline(lm(log(ala)~log(masa)),col='blue',lwd=2)
text(7,4.8,'log(ala)=3.0066+0.3478 log(masa)')
lm(log(ala)~log(masa))
```

Figura 10.15: Modelo de potencia o alométrico para 28 especies de Phasiniidae

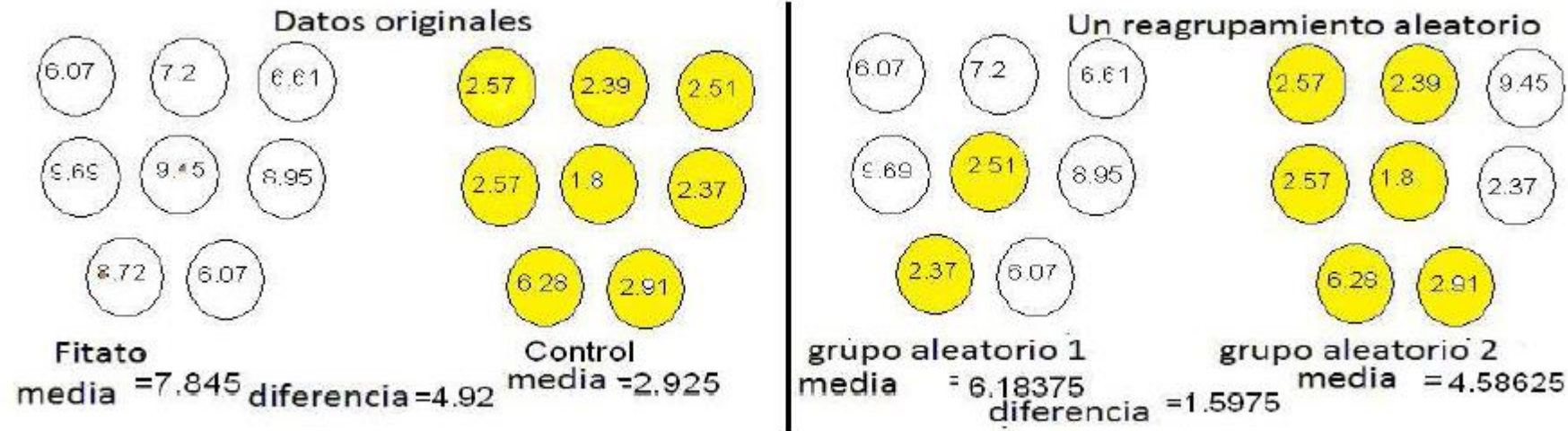


Capitulo 12 Introduccion a la inferencia

- Incluye métodos tradicionales (t-Student) y también de bootstrap y test de permutación. Este es el intervalo bootstrap para la correlacion
- El código esta en <https://faculty.etsu.edu/seier/RcommCh3.txt>



Por ahora el código en R esta en <https://faculty.etsu.edu/seier/IntroStatsBioBook.htm>



Test de permutación o aleatorización
para comparar dos grupos

