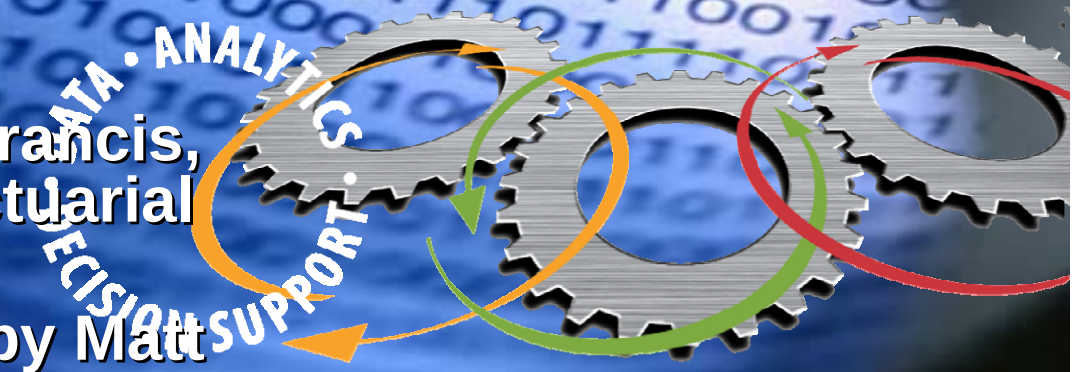
2008 PREDICTIVE MODELING SEMINAR

# Software and Utilities: Free or Inexpensive – R utilities/add-ons

Presentation by Louise Francis, Francis Analytics and Actuarial Data Mining, Inc.

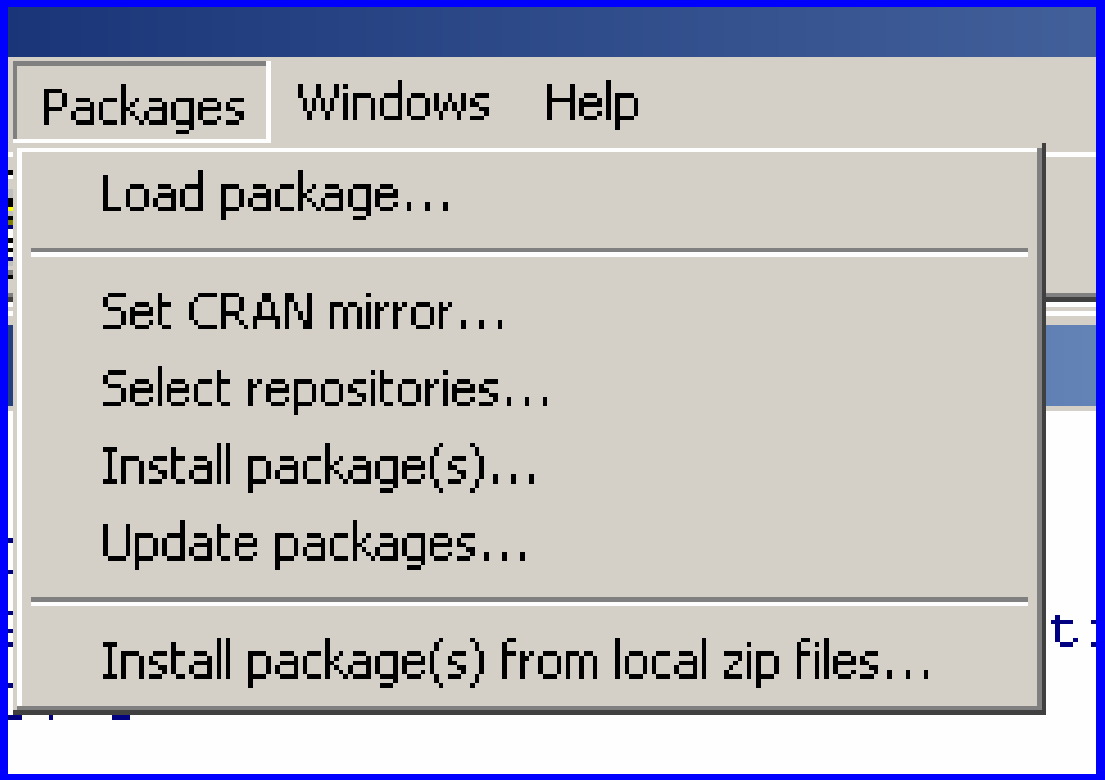
Many slides contributed by Matt Flynn, ISO

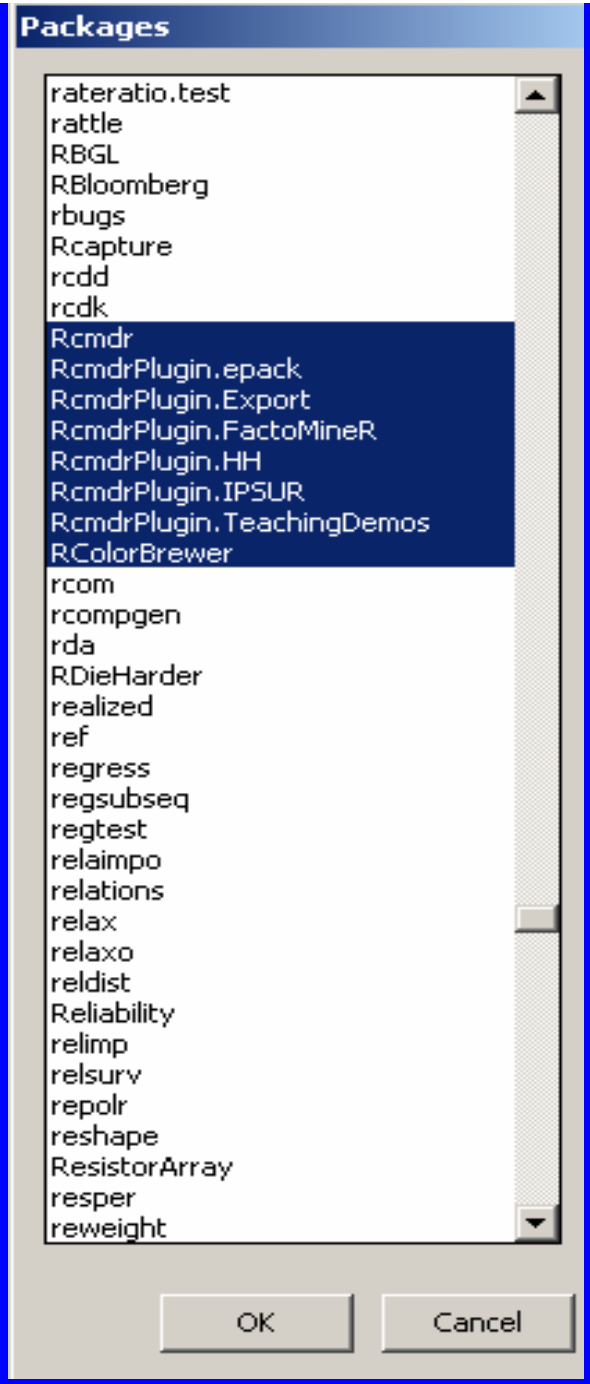


# Some Useful R utilities/add-ons

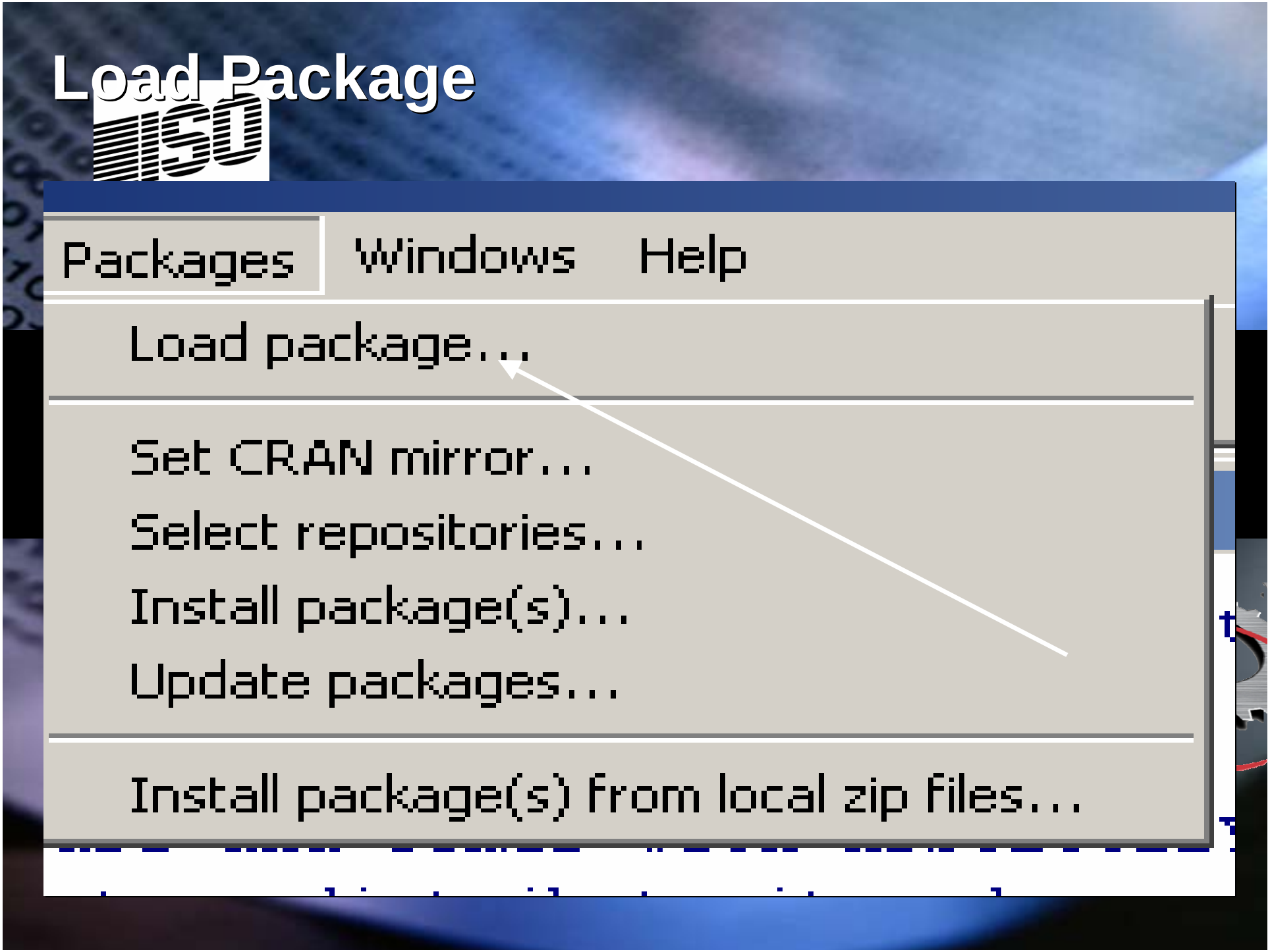
- Calling R code from SAS
- R Commander – Rcmdr
- actuar
- rpanel
- iplots
- RExcel
- JGR – “Jaguar”

# Install Package from CRAN Mirror

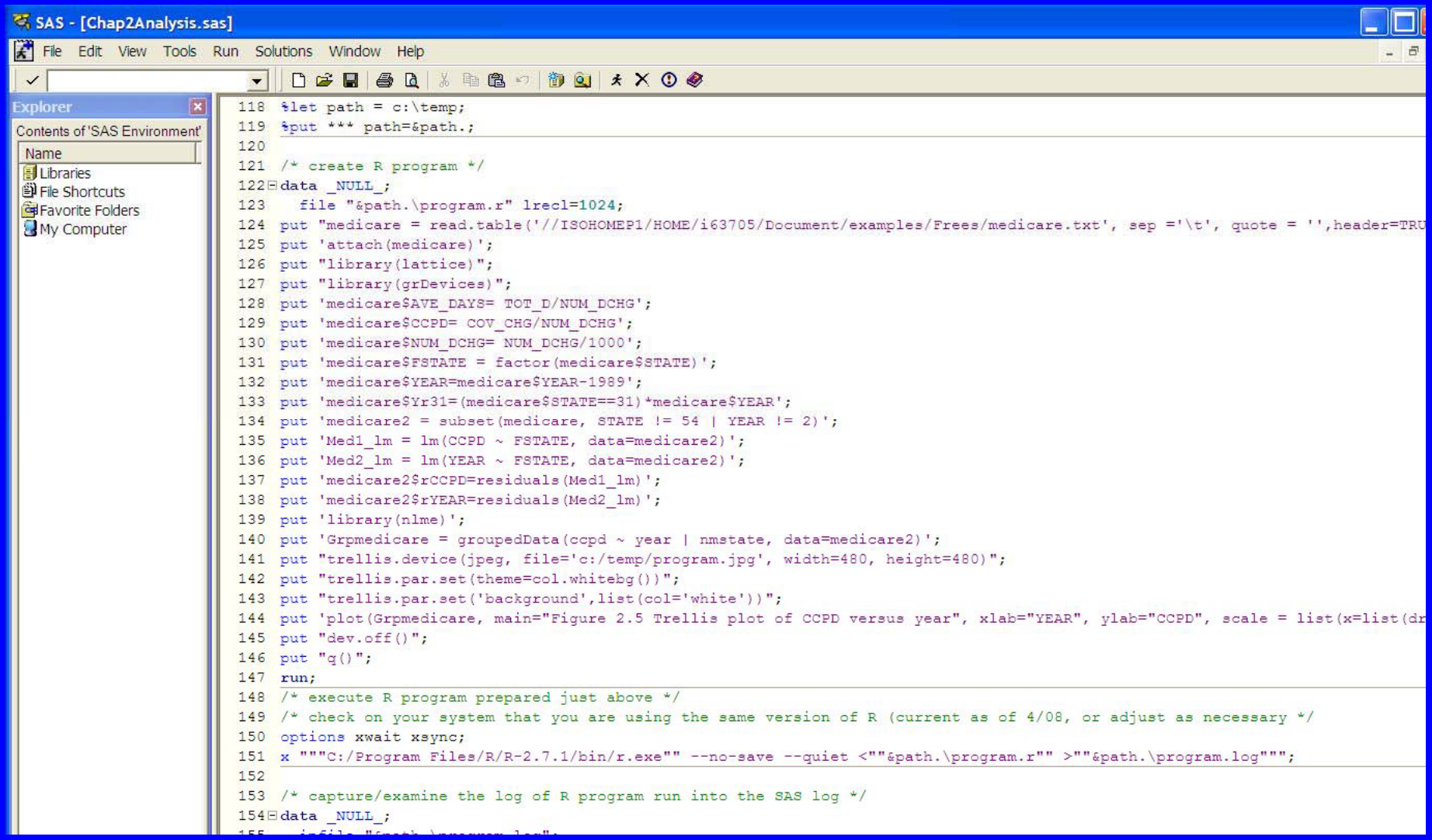



Packages
rateratio.test
rattle
RBGL
RBloomberg
rbugs
Rcapture
rcdd
rcdk
Rcmdr
RcmdrPlugin.epack
RcmdrPlugin.Export
RcmdrPlugin.FactoMineR
RcmdrPlugin.HH
RcmdrPlugin.IPSUR
RcmdrPlugin.TeachingDemos
RColorBrewer
rcom
rcompgen
rda
RDieHarder
realized
ref
regress
regsubseq
regtest
relaimpo
relations
relax
relaxo
reldist
Reliability
relimp
relsurv
repolr
reshape
ResistorArray
resper
reweight
OK
Cancel

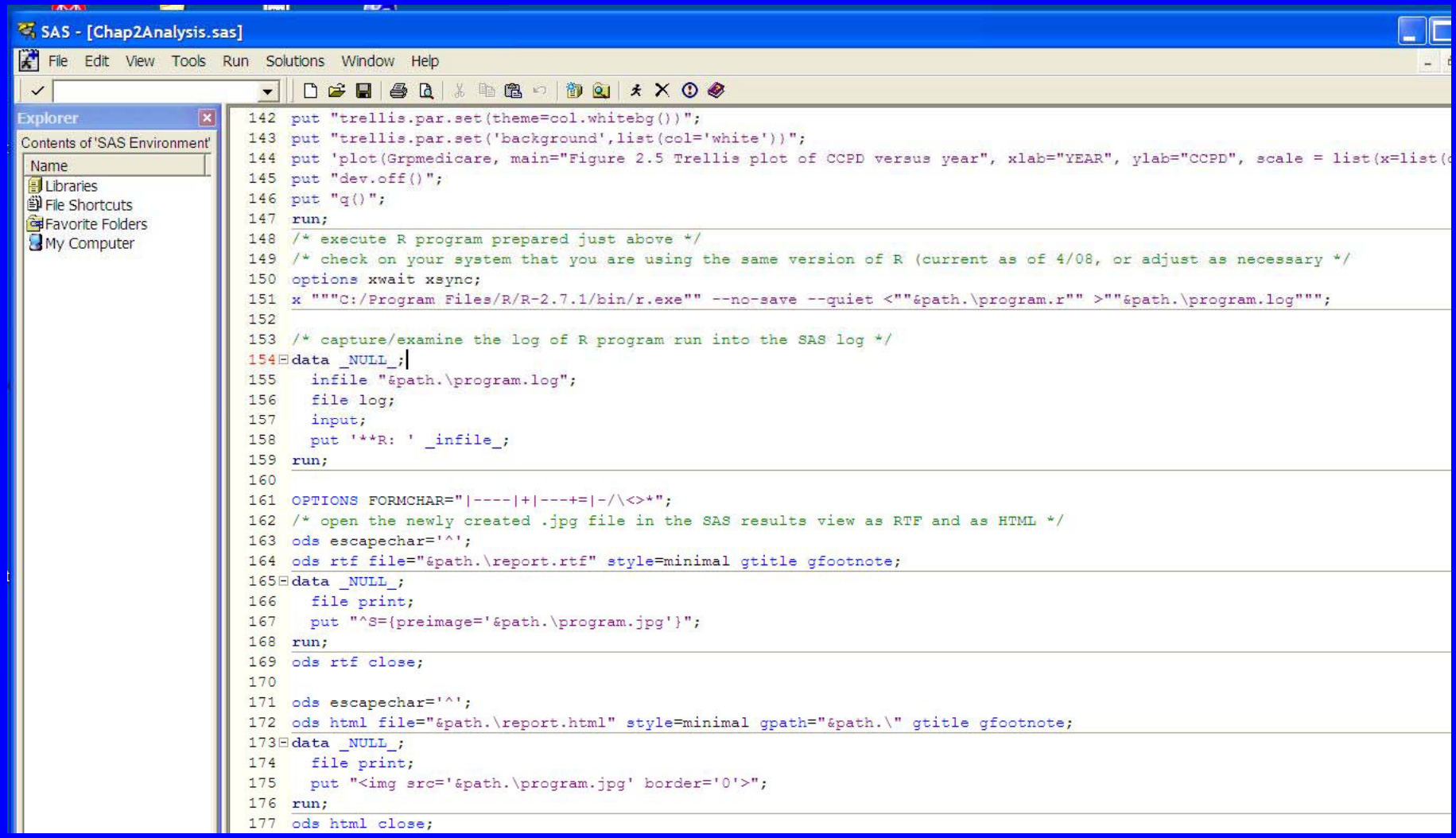
# Load Package



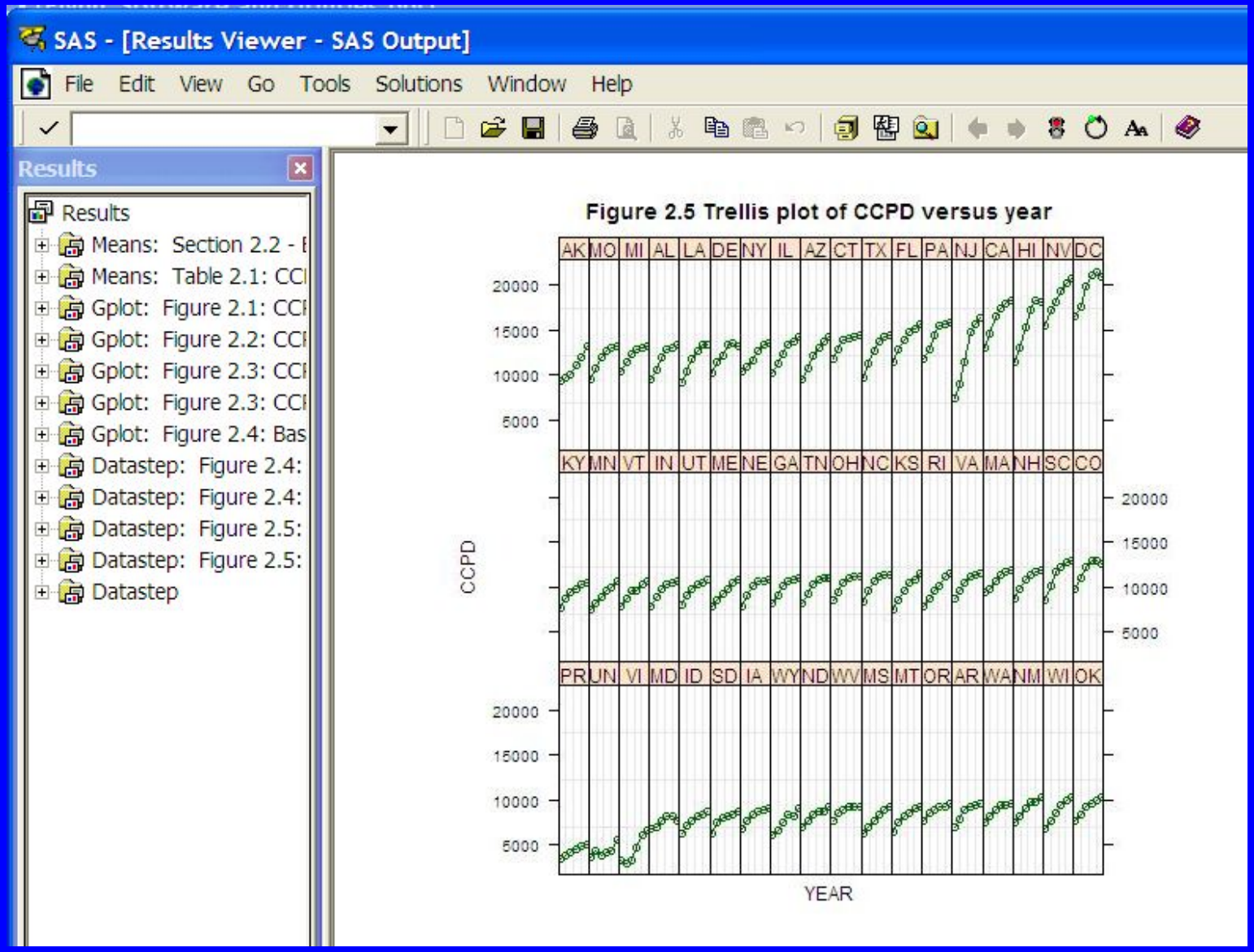
# Calling R from SAS



# Calling R from SAS



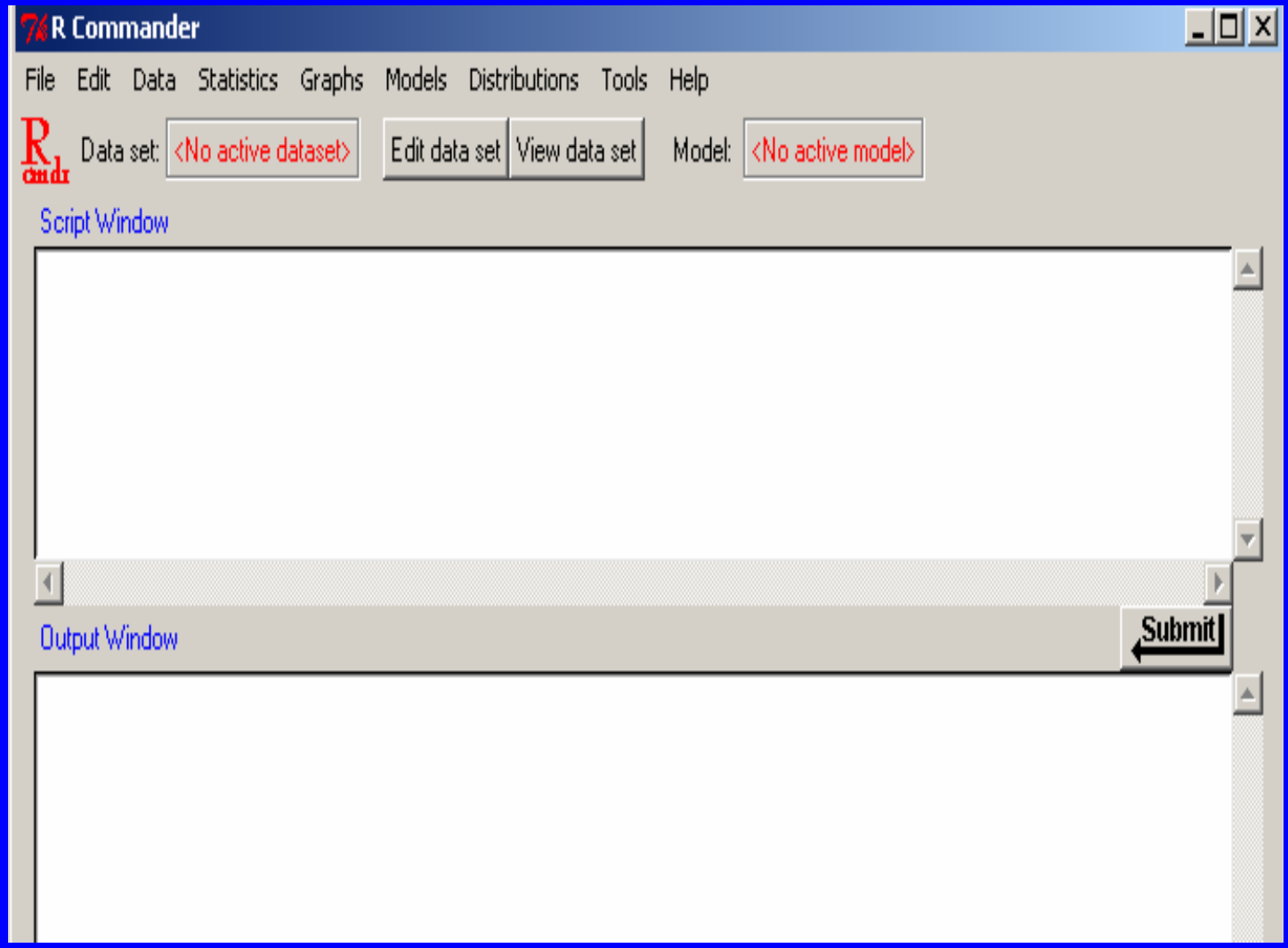
# Calling R from SAS



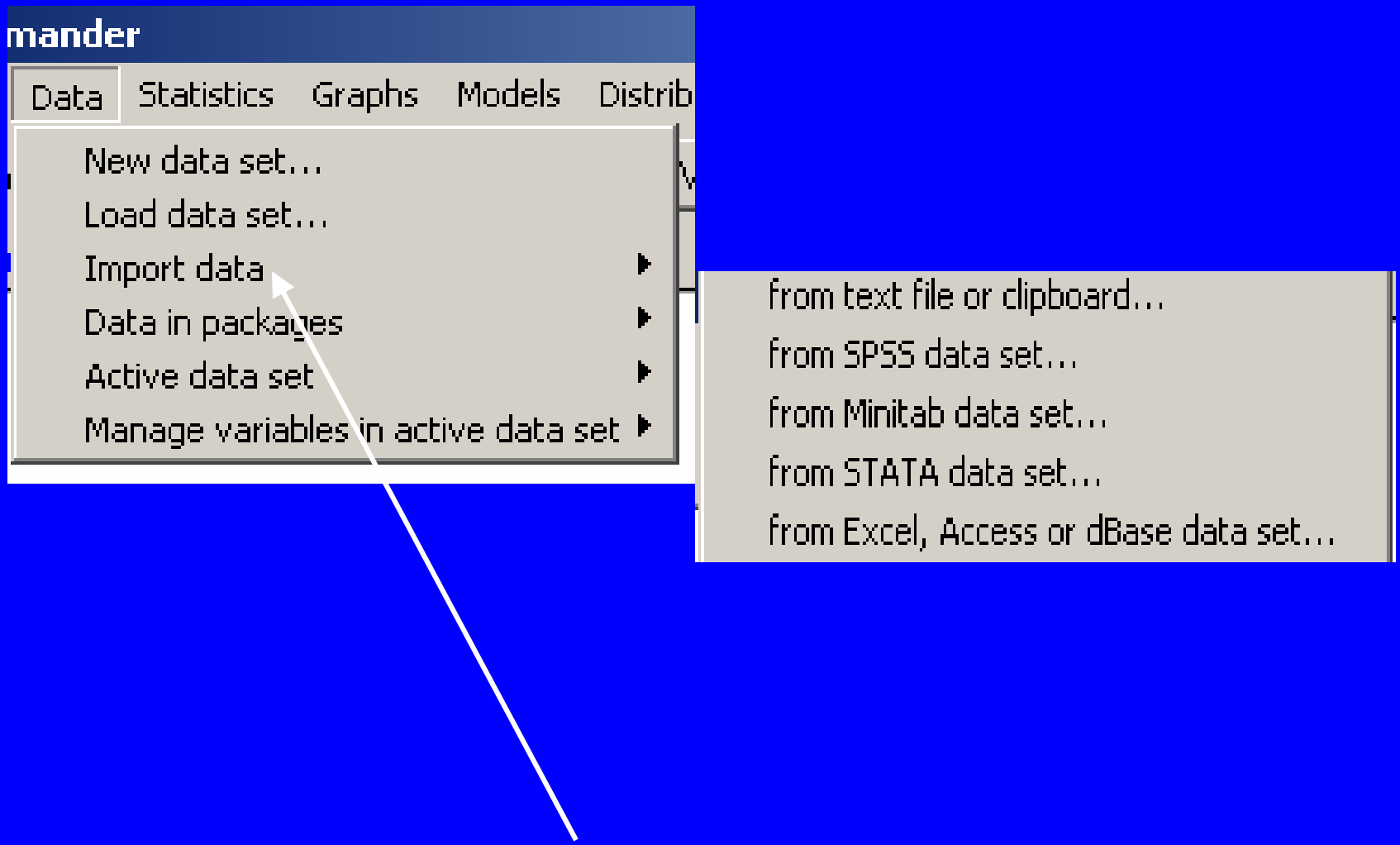
# Rcmdr utility

- A GIU Inter face for R
- Developed by John Fox
- Makes basic data importing and statistical procedures available by menu

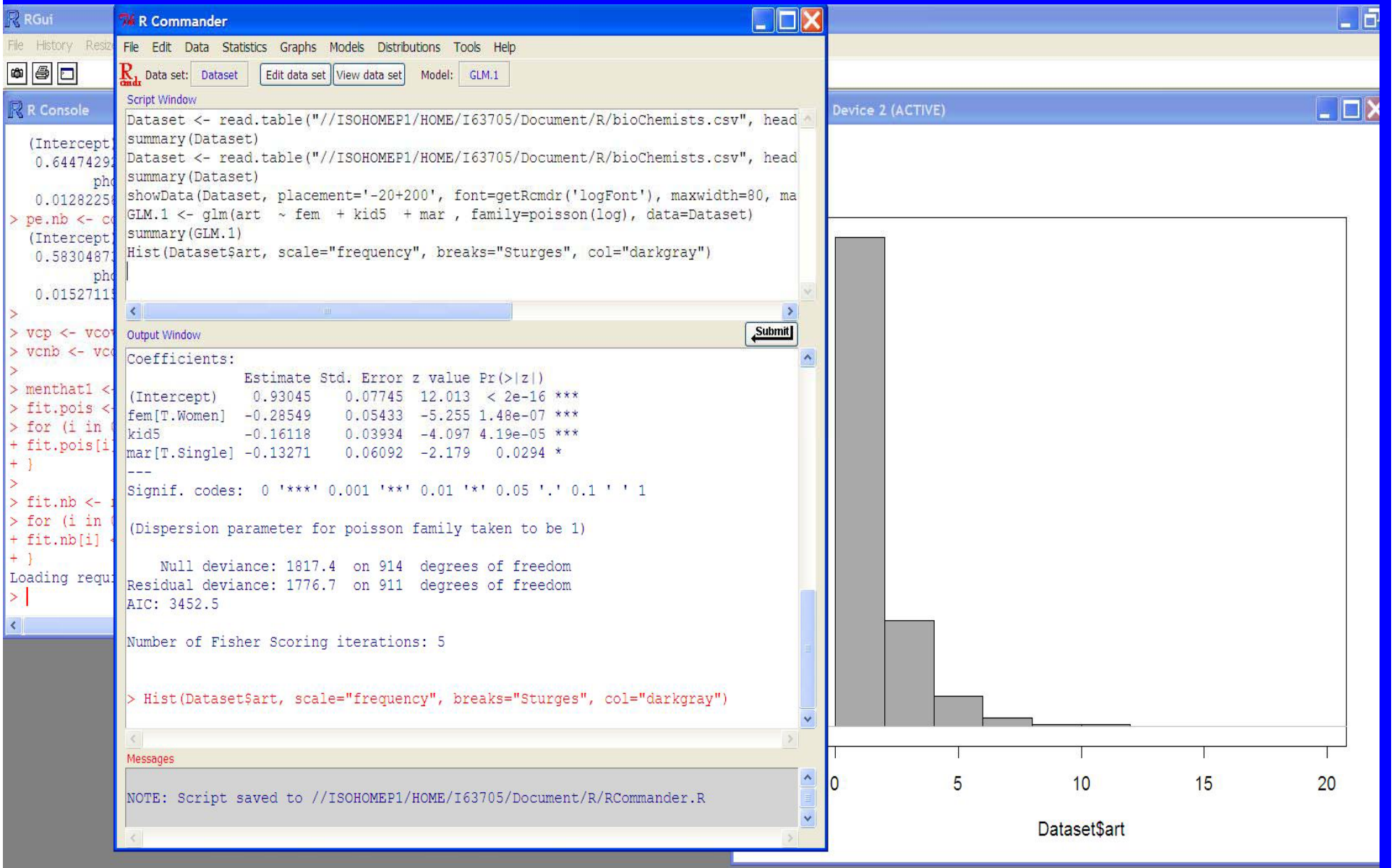
# Rcmdr GUI


R Commander
File Edit Data Statistics Graphs Models Distributions Tools Help
Data set: <No active dataset>
Edit data set
View data set
Model: <No active model>
Script Window
Output Window
Submit

# Import Data from Other Software (Excel, Access, SPSS) from Rcmdr



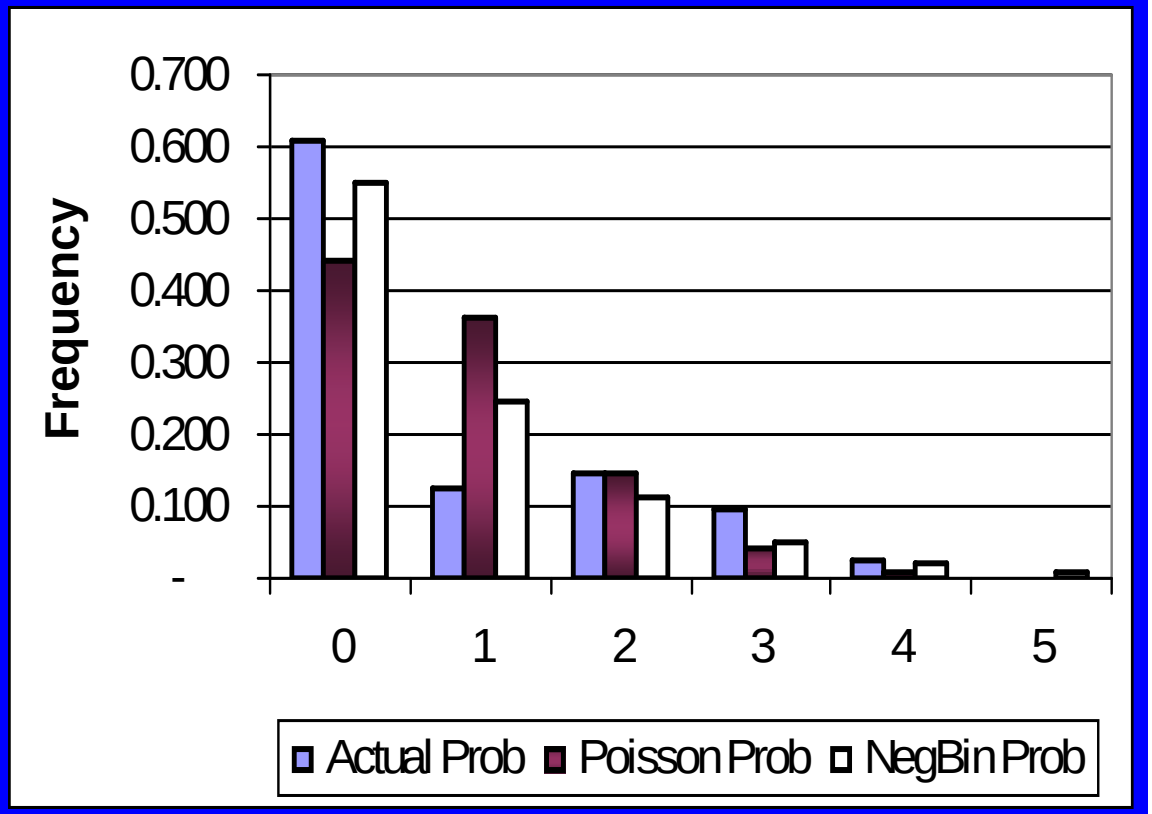
# Rcmdr


RGui
File History Resize
R Console
(Intercept)
0.64474292
pho
0.01282258
> pe.nb <- co
(Intercept)
0.58304871
pho
0.01527113
>
> vcp <- vco
> vcnb <- vc
>
> menthat1 <-
> fit.pois <-
> for (i in
+ fit.pois[i
+ }
>
> fit.nb <-
> for (i in
+ fit.nb[i]
+ }
Loading requi
>
R Commander
File Edit Data Statistics Graphs Models Distributions Tools Help
Data set: Dataset
Edit data set
View data set
Model: GLM.1
Script Window
Dataset <- read.table("//ISOHOMEP1/HOME/I63705/Document/R/bioChemists.csv", head
summary(Dataset)
Dataset <- read.table("//ISOHOMEP1/HOME/I63705/Document/R/bioChemists.csv", head
summary(Dataset)
showData(Dataset, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80, ma
GLM.1 <- glm(art ~ fem + kid5 + mar , family=poisson(log), data=Dataset)
summary(GLM.1)
Hist(Dataset$art, scale="frequency", breaks="Sturges", col="darkgray")
Output Window
Submit
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.93045 0.07745 12.013 < 2e-16 ***
fem[T.Women] -0.28549 0.05433 -5.255 1.48e-07 ***
kid5 -0.16118 0.03934 -4.097 4.19e-05 ***
mar[T.Single] -0.13271 0.06092 -2.179 0.0294 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for poisson family taken to be 1)
Null deviance: 1817.4 on 914 degrees of freedom
Residual deviance: 1776.7 on 911 degrees of freedom
AIC: 3452.5
Number of Fisher Scoring iterations: 5
> Hist(Dataset$art, scale="frequency", breaks="Sturges", col="darkgray")
Messages
NOTE: Script saved to //ISOHOMEP1/HOME/I63705/Document/R/RCommander.R
Device 2 (ACTIVE)
0
5
10
15
20
Dataset$art

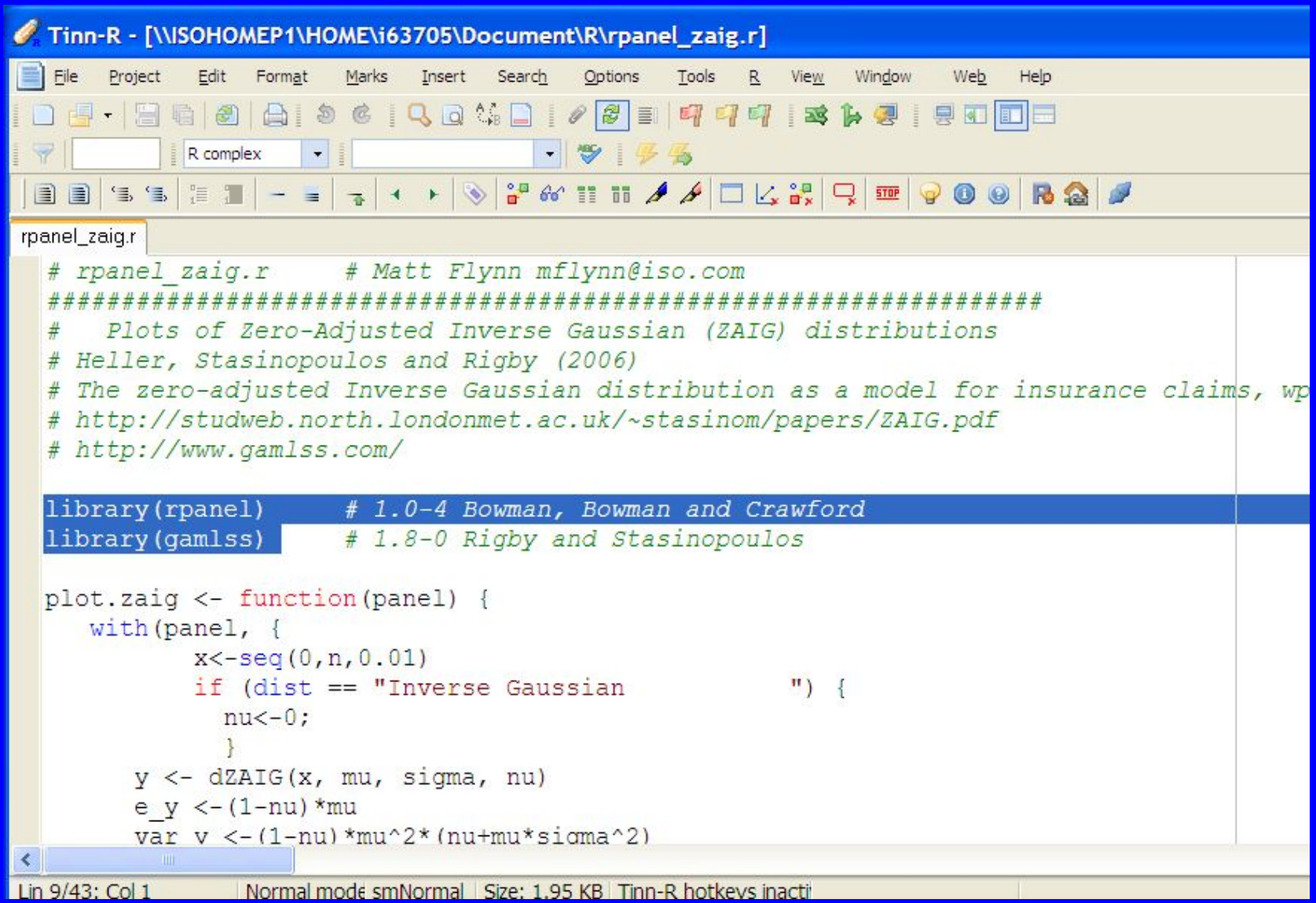
# Alternative Distributions: ZIP (zero inflated Poisson)

$$\varphi + (1-\varphi)e^{-\lambda} \quad x=0$$

$$(1-\varphi)\frac{\lambda^{x}}{x!}e^{-\lambda} \quad x>0$$



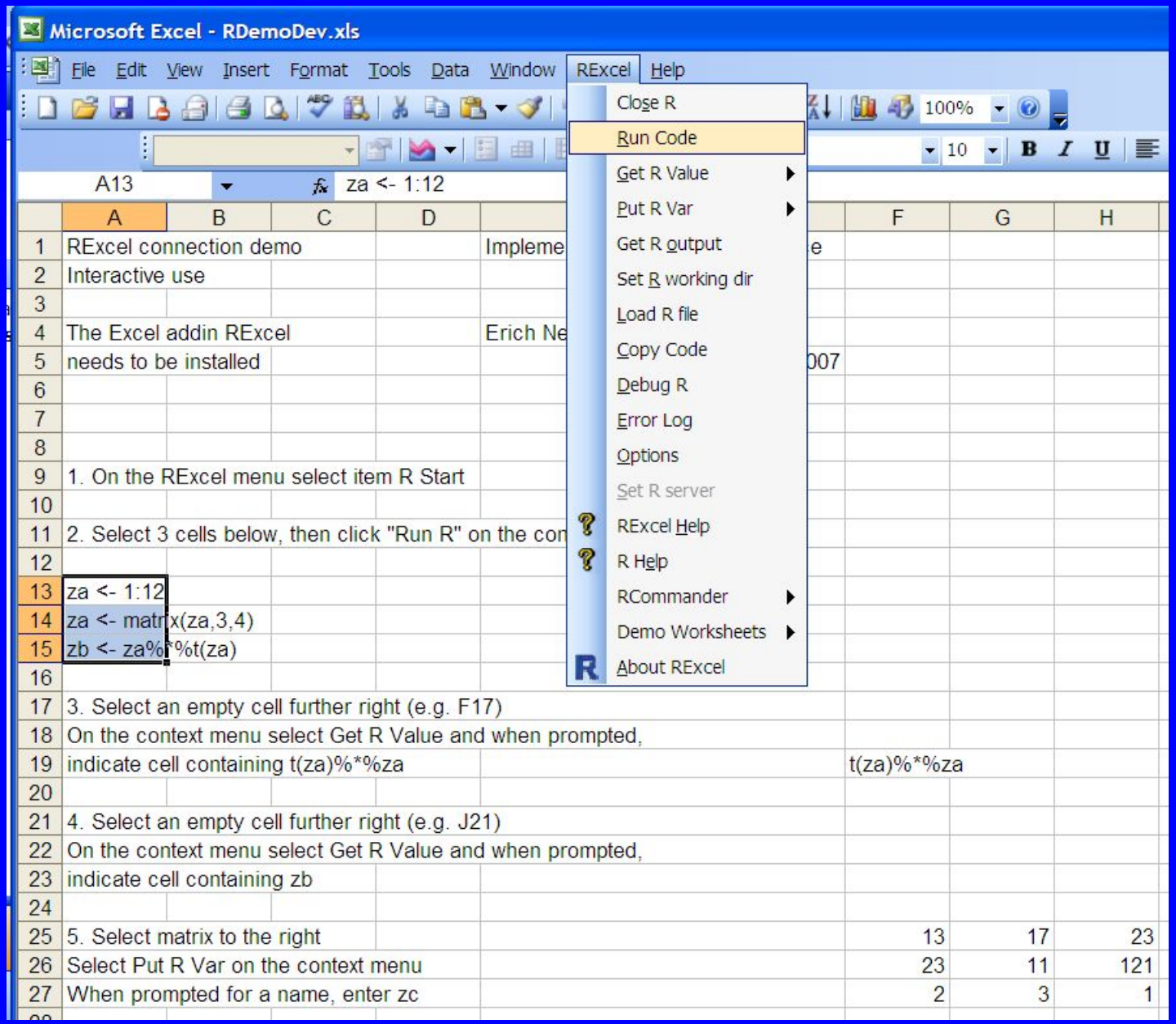
# Tinn-R: An R Editor



# RExcel

- An Add-in for Excel
- Lets you run R from within Excel
- You must first run an execute file from R to install it

# RExcel


Microsoft Excel - RDemoDev.xls
File Edit View Insert Format Tools Data Window RExcel Help
Close R
Run Code
Get R Value
Put R Var
Get R output
Set R working dir
Load R file
Copy Code
Debug R
Error Log
Options
Set R server
RExcel Help
R Help
RCommander
Demo Worksheets
About RExcel
A13
za <- 1:12
RExcel connection demo
Interactive use
The Excel addin RExcel
needs to be installed
1. On the RExcel menu select item R Start
2. Select 3 cells below, then click "Run R" on the con
za <- 1:12
za <- matrix(za,3,4)
zb <- za%*%t(za)
3. Select an empty cell further right (e.g. F17)
On the context menu select Get R Value and when prompted,
indicate cell containing t(za)%*%za
t(za)%*%za
4. Select an empty cell further right (e.g. J21)
On the context menu select Get R Value and when prompted,
indicate cell containing zb
5. Select matrix to the right
Select Put R Var on the context menu
When prompted for a name, enter zc
13 17 23
23 11 121
2 3 1

# RPanel & ZIGP

```
# rpanel_zigp.r     # Matt Flynn mflynn@iso.com
#################################################################
#   Plots of Zero-Inflated Generalized Poisson (ZIGP) distributions
# Czado, Claudia and Aleksey Min
# Zero-inflated generalized Poisson regression models:
# Asymptotic theory and applications
# Munich University of Technology WP 2005
# http://www-m4.ma.tum.de/Papers/Min/Czado-Min.pdf

library(rpanel)     # 1.0-4 Bowman, Bowman and Crawford
library(ZIGP)       # 1.2 Vinzenz Erhardt

plot.zigp <- function(panel) {
  with(panel, {
    if (dist == "Poisson") {
      phi<-1; omega<-0;
    } else
      if (dist == "Generalized Poisson    ") {
        omega<-0;
      } else
        if (dist == "ZIP") {
          phi<-1;
        }
    probs <- dzigp(0:n, mu, phi, omega)
    plot(0:n, probs, type='n', ylim=c(0,0.5), xlab="x", ylab = "Probability")
    segments(0:n, rep(0, n+1), 0:n, probs, col="blue", lwd=3)
    abline(h=0, col="grey")
    mutext <- as.character(round(mu, 3))
    phitext <- as.character(round(phi, 3))
    omegatext <- as.character(round(omega, 3))
    title(paste("Dist:", dist, ", mu =", mutext, ", phi =", phitext, ", omega =", omegatext))
    text(10,0.4,expression(P(Y==y, mu,phi,omega) == (1['x=0'])*(omega + (1 - omega)*exp(frac(-mu, phi))) + "     "))
    text(10,0.35, expression((1[x>0])*(1 - omega)*frac(mu*(mu + (phi - 1)*x)^(x - 1),"x!")*phi^(-x)*
```

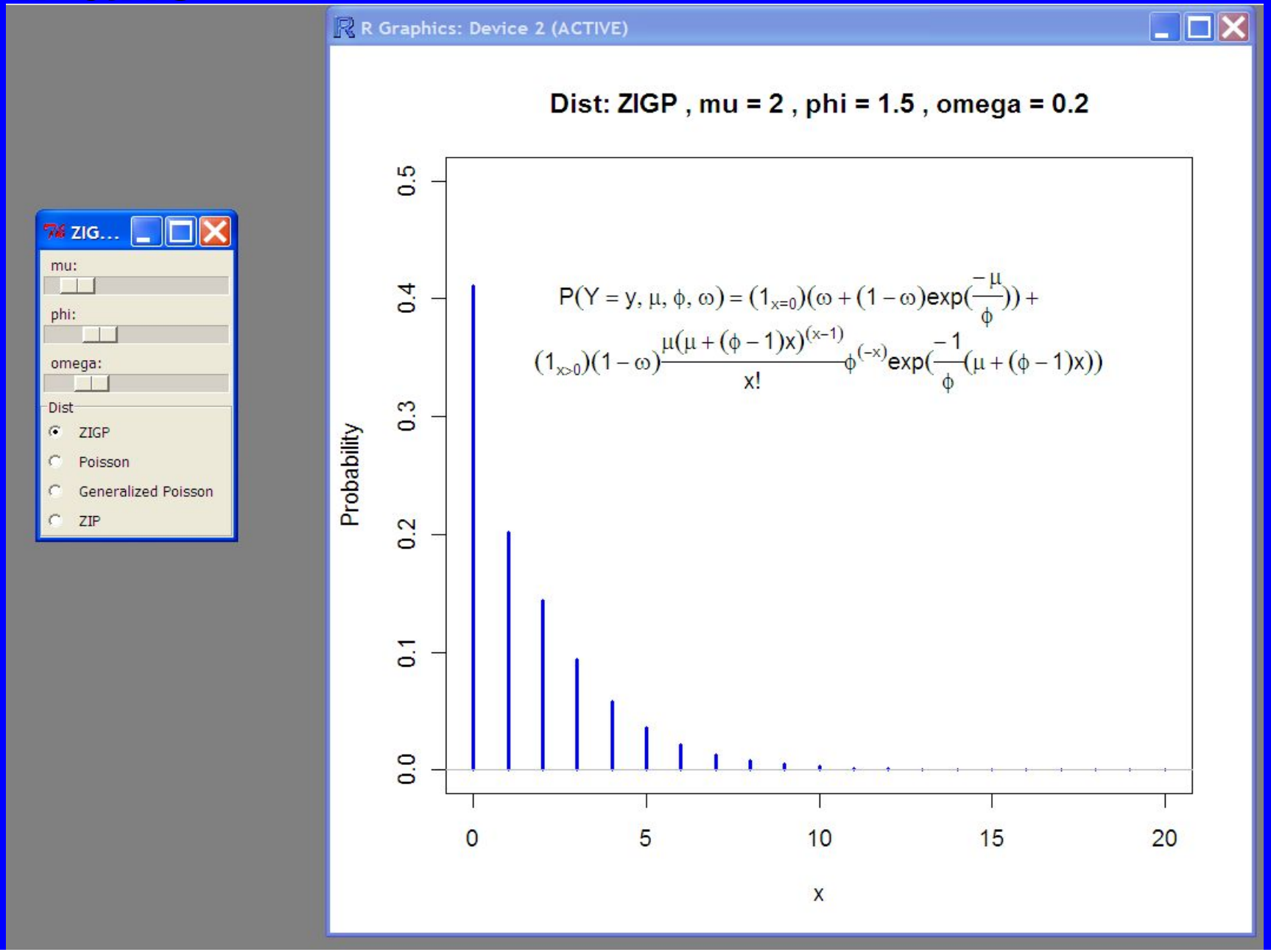
# actuar

- **Package of actuarial functions**
  - Loss distributions
  - Risk theory
  - Some kinds of simulation
  - Credibility theory

# RPanel & ZIGP

```
  with(panel, {
    if (dist == "Poisson") {
      phi<-1; omega<-0;
    } else
    if (dist == "Generalized Poisson   ") {
      omega<-0;
    } else
    if (dist == "ZIP") {
      phi<-1;
    }
    probs <- dzigp(0:n, mu, phi, omega)
    plot(0:n, probs, type='n', ylim=c(0,0.5), xlab="x", ylab = "Probability")
    segments(0:n, rep(0, n+1), 0:n, probs, col="blue", lwd=3)
    abline(h=0, col="grey")
    mutext <- as.character(round(mu, 3))
    phitext <- as.character(round(phi, 3))
    omegatext <- as.character(round(omega, 3))
    title(paste("Dist:", dist, ", mu =", mutext, ", phi =", phitext, ", omega =", omegatext))
    text(10,0.4,expression(P(Y==y, mu,phi,omega) == (1[x=0])*(omega + (1 - omega)*exp(frac(-mu, phi))) + "     "))
    text(10,0.35, expression((1[x>0])*(1 - omega)*frac(mu*(mu + (phi - 1)*x)^(x - 1),"x!")*phi^(-x)*
                             exp(frac(-1,phi)*(mu + (phi - 1)*x))))
  })
  panel
  }
panel <- rp.control("ZIGP Parameters", dist="ZIGP", n=20, mu=2, phi=1.5, omega=0.2)
rp.slider(panel, mu, 0.01, 20,
      title = "mu:", action = plot.zigp)
rp.slider(panel, phi, 1, 3,
      title = "phi:", action = plot.zigp)
rp.slider(panel, omega, 0, 1,
      title = "omega:", action = plot.zigp)
rp.radiogroup(panel, dist, c('ZIGP', 'Poisson', 'Generalized Poisson   ', 'ZIP'), action=plot.zigp, title = 'Dist')
```

# RPanel


ZIG...
mu:
phi:
omega:
Dist
ZIGP
Poisson
Generalized Poisson
ZIP
R Graphics: Device 2 (ACTIVE)
Dist: ZIGP , mu = 2 , phi = 1.5 , omega = 0.2
$P(Y = y, \mu, \phi, \omega) = (1_{x=0})(\omega + (1-\omega)\exp(\frac{-\mu}{\phi})) + (1_{x>0})(1-\omega)\frac{\mu(\mu+(\phi-1)x)^{(x-1)}}{x!}\phi^{(-x)}\exp(\frac{-1}{\phi}(\mu+(\phi-1)x))$
Probability
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
x

# iPlots: Interactive Plots

# JGR – “Jaguar”

JGR Features Screenshots Download JGR Installation
JGR on Linux FAQ

## JGR - Java GUI for R

**JGR** (speak 'Jaguar') is a universal and unified Graphical User Interface for R (it actually abbreviates **J**ava **G**ui for **R**). JGR was introduced at the **useR!** meeting in **2004** and there is an introductory article in the **Statistical Computing and Graphics Newsletter Vol 16 nr 2** p9-12

**What's new?**

**2008/07/24** JGR 1.6-2 and new launchers 1.61 released
- please download the new launchers

**2008/06/18** JGR 1.6-1 released
- minor fix in editor and object-browser

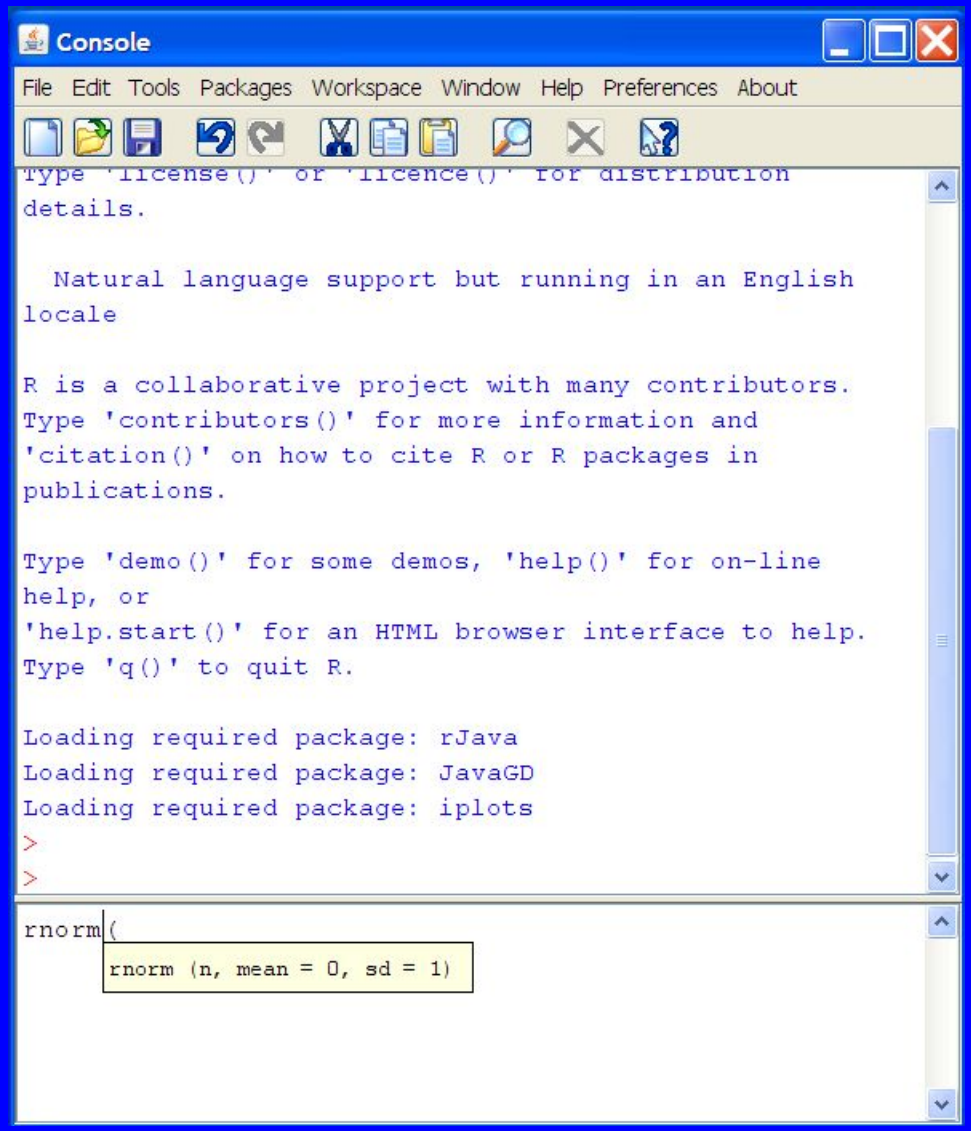
**2008/06/15** JGR 1.6 and new launchers released
- please delete your .JGRprefs file in your home directory before and download the new launchers

**2008/06/15** Launch of new JGR webpage

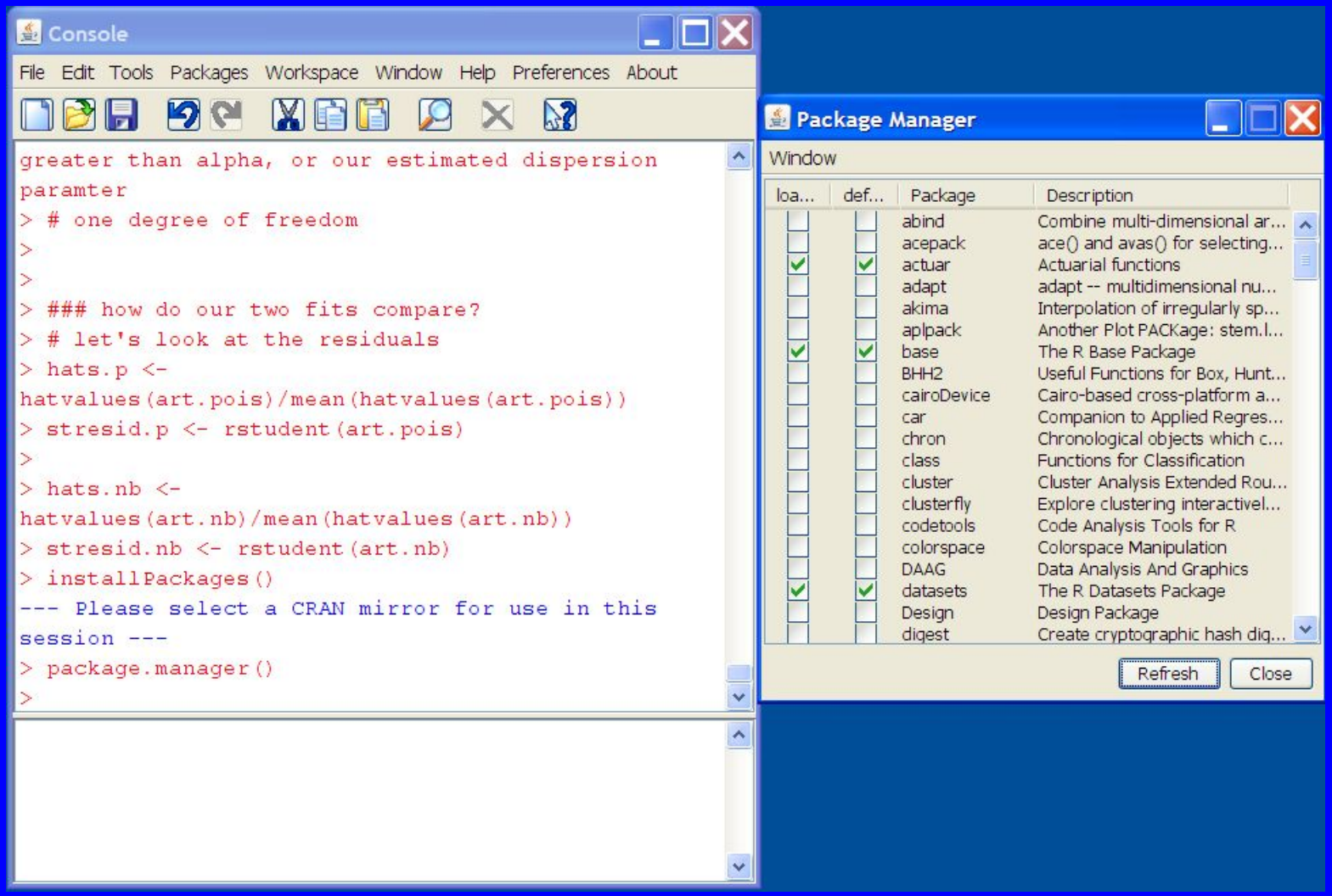
JGR's friends at RoSuDa.org

Made on a Mac

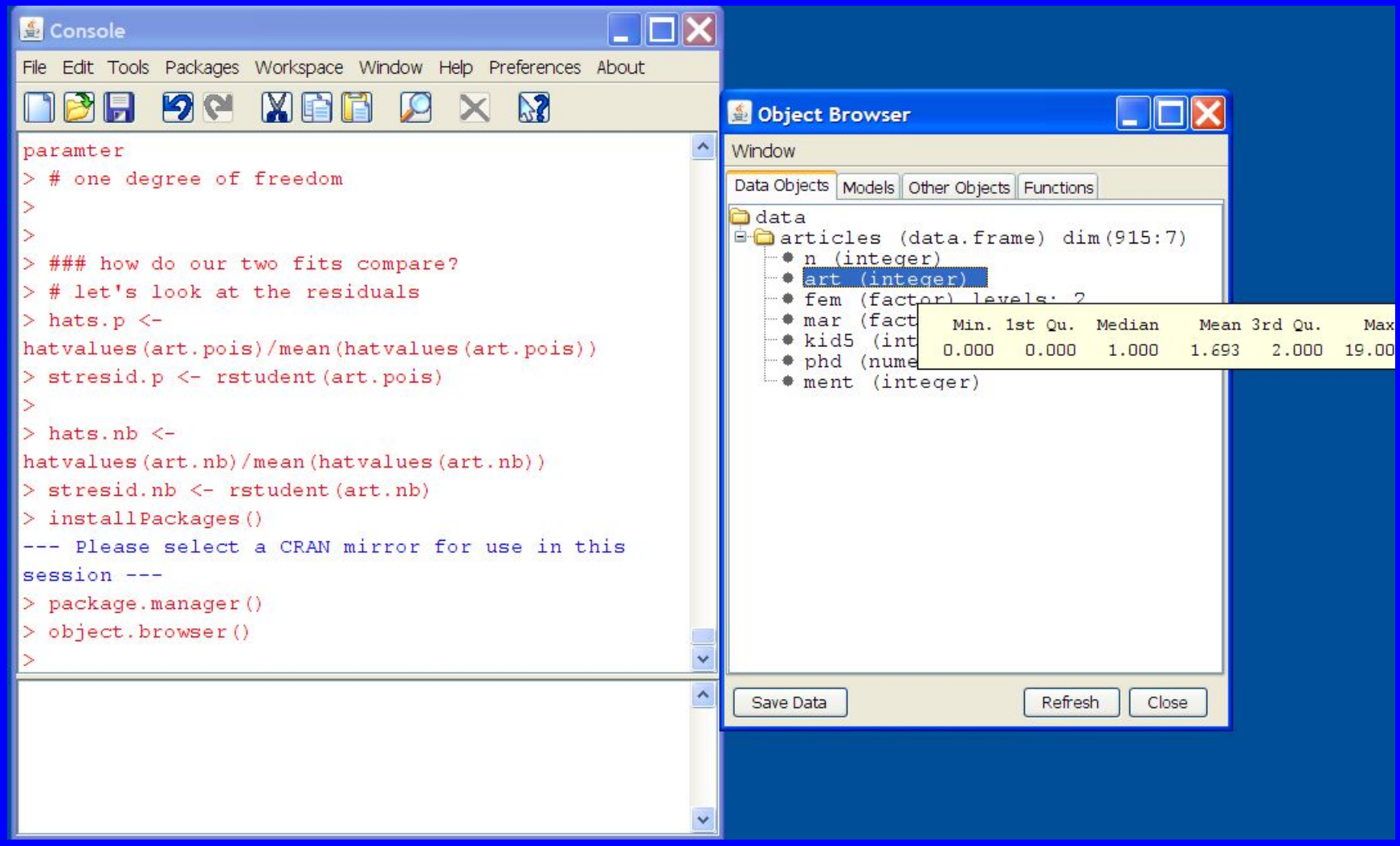
# JGR – “Jaguar”


Console
File Edit Tools Packages Workspace Window Help Preferences About
Type 'license()' or 'licence()' for distribution
details.
Natural language support but running in an English
locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in
publications.
Type 'demo()' for some demos, 'help()' for on-line
help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Loading required package: rJava
Loading required package: JavaGD
Loading required package: iplots
>
>
rnorm(
rnorm (n, mean = 0, sd = 1)

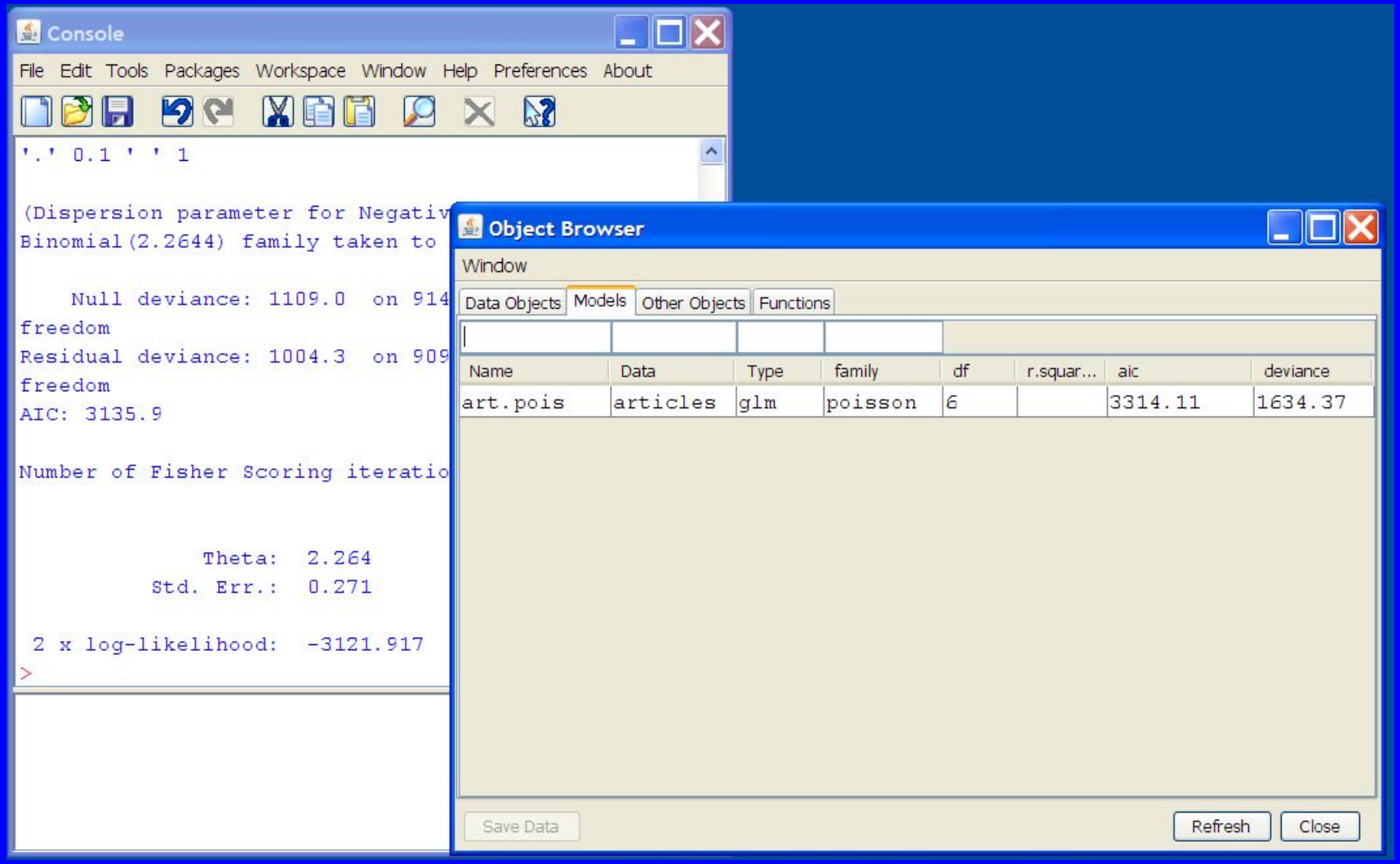
# JGR – “Jaguar”


Console
File Edit Tools Packages Workspace Window Help Preferences About
greater than alpha, or our estimated dispersion paramter
> # one degree of freedom
>
>
> ### how do our two fits compare?
> # let's look at the residuals
> hats.p <- hatvalues(art.pois)/mean(hatvalues(art.pois))
> stresid.p <- rstudent(art.pois)
>
> hats.nb <- hatvalues(art.nb)/mean(hatvalues(art.nb))
> stresid.nb <- rstudent(art.nb)
> installPackages()
--- Please select a CRAN mirror for use in this session ---
> package.manager()
>
Package Manager
Window
loa... def... Package Description
abind Combine multi-dimensional ar...
acepack ace() and avas() for selecting...
actuar Actuarial functions
adapt adapt -- multidimensional nu...
akima Interpolation of irregularly sp...
aplpack Another Plot PACKage: stem.l...
base The R Base Package
BHH2 Useful Functions for Box, Hunt...
cairoDevice Cairo-based cross-platform a...
car Companion to Applied Regres...
chron Chronological objects which c...
class Functions for Classification
cluster Cluster Analysis Extended Rou...
clusterfly Explore clustering interactivel...
codetools Code Analysis Tools for R
colorspace Colorspace Manipulation
DAAG Data Analysis And Graphics
datasets The R Datasets Package
Design Design Package
digest Create cryptographic hash dig...
Refresh
Close

# JGR – “Jaguar”


Console
File Edit Tools Packages Workspace Window Help Preferences About
paramter
> # one degree of freedom
>
>
> ### how do our two fits compare?
> # let's look at the residuals
> hats.p <-
hatvalues(art.pois)/mean(hatvalues(art.pois))
> stresid.p <- rstudent(art.pois)
>
> hats.nb <-
hatvalues(art.nb)/mean(hatvalues(art.nb))
> stresid.nb <- rstudent(art.nb)
> installPackages()
--- Please select a CRAN mirror for use in this
session ---
> package.manager()
> object.browser()
>
Object Browser
Window
Data Objects
Models
Other Objects
Functions
data
articles (data.frame) dim(915:7)
n (integer)
art (integer)
fem (factor) levels: 2
mar (fact
kid5 (int
phd (nume
ment (integer)
Min. 1st Qu. Median Mean 3rd Qu. Max
0.000 0.000 1.000 1.693 2.000 19.00
Save Data
Refresh
Close

# JGR – “Jaguar”


Console
File Edit Tools Packages Workspace Window Help Preferences About
'.' 0.1 ' ' 1
(Dispersion parameter for Negativ
Binomial(2.2644) family taken to
Null deviance: 1109.0 on 914
freedom
Residual deviance: 1004.3 on 909
freedom
AIC: 3135.9
Number of Fisher Scoring iteratio
Theta: 2.264
Std. Err.: 0.271
2 x log-likelihood: -3121.917
>
Object Browser
Window
Data Objects
Models
Other Objects
Functions
Name
Data
Type
family
df
r.squar...
aic
deviance
art.pois
articles
glm
poisson
6
3314.11
1634.37
Save Data
Refresh
Close

# Helpful websites

The mothership

http://cran.r-project.org/

R FAQ

http://cran.r-project.org/doc/FAQ/R-FAQ.html

An introduction to R

http://cran.r-project.org/doc/manuals/R-intro.pdf

R News

http://cran.r-project.org/doc/Rnews/Rnews_2008-1.pdf

Kickstarting R

http://cran.r-project.org/doc/contrib/Lemon-kickstart/index.html

**A Brief Guide to R for Beginners in Econometrics**

http://people.su.se/~ma/R_intro/

R Wiki

http://wiki.r-project.org/rwiki/doku.php

# Contact info:

- **-Matt Flynn**
  - Office: (860) 570-0015
  - Mobile: (203) 435-7362
  - Fax: (201) 748-1900
  - Email: mflynn@iso.com
- **Louise Francis**
  - 215-923-1567
  - Louise_francis@msn.com