



# Actuarial Modeling in

CAS Webinar  
August 19, 2010

Jim Guszczka, FCAS, MAAA

# Aganda

---

---

HistoRy

---

Installing R on Your PC

---

Warm-up

---

Example 1: Fitting Loss Distributions

---

Example 2: GLM

---

Example 3a: Chain-Ladder Loss Reserving

---

Example 3b: GLM Loss Reserving

---

Example 3c: Longitudinal Loss Reserving

---

Example 4: Geospatial Modeling in R

---

# R Background



# R Background

---

**R is an open-source, object-oriented statistical programming language**

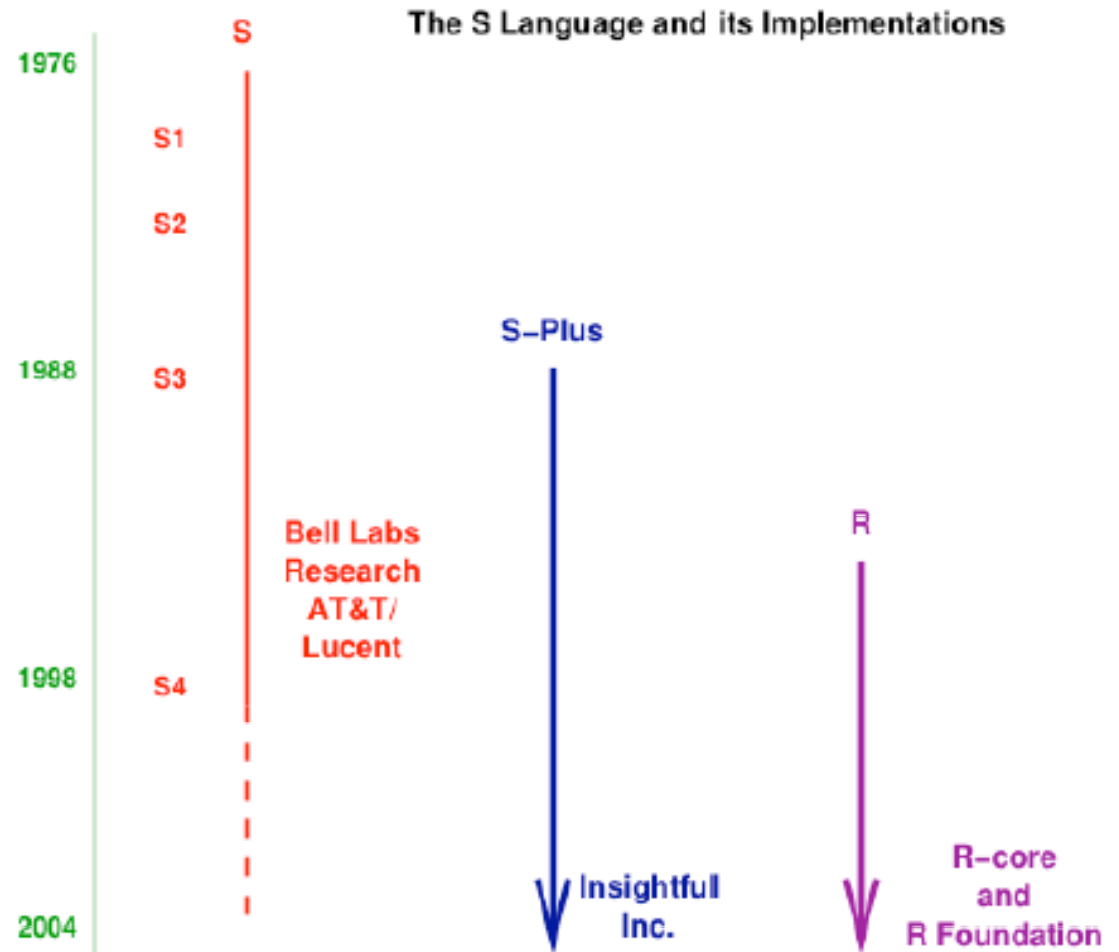
- History:
  - R is based on the S statistical programming language developed by John Chambers at Bell labs in the 1980's
  - The commercial package S-plus is based on the S language
  - R is an open-source implementation of the S language
  - Developed by Robert Gentleman and Ross Ihaka in New Zealand
- Features:
  - R is a high-level, object-oriented programming environment
  - R has advanced graphical capabilities
  - Statisticians around the world contribute add-on packages... therefore:

“The great beauty of R is that you can modify it to do all sorts of things,” said Hal Varian, chief economist at Google. “And you have a lot of prepackaged stuff that’s already available, so you’re standing on the shoulders of giants.”

# R Evolution

---

- S is the original language
- S-plus is a commercial implementation of S
- R is an open-source implementation of S
- R is very similar to, but not identical with, other implementations of S



# Facets of R

---

- In a recent article John Chambers discussed 6 “Facets of R”
  1. *An interface* to computational procedures of many kinds
  2. *Interactive*, hands-on in real time
  3. *Functional* in its model of programming
  4. *Object-oriented*, “everything is an object”
  5. *Modular*, built from standardized pieces
  6. *Collaborative*, a world-wide, open-source effort
- **Interactive interface: Chambers was influenced by APL**
  - One of the rare interactive scientific computing environments
  - Gives user ability to express novel computations
  - Heavy emphasis on matrices and arrays
  - But: unlike R, APL had no interface to procedures
- In the days before spreadsheets, APL was very popular in the actuarial community

“Facets of R”, John M. Chambers, *The R Journal* Vol. 1/1, May 2009

# Modular and Collaborative: A Network ExteRnality

- Hal Varian's "giant" has grown at an exponential rate.
- The open-source nature of R has encouraged top researchers from around the world to contribute new, often highly advanced, packages.
- Result: a powerful "network effect".
  - The value of a product increases as more people use it.
- R has become something like the Wikipedia of the statistics world.

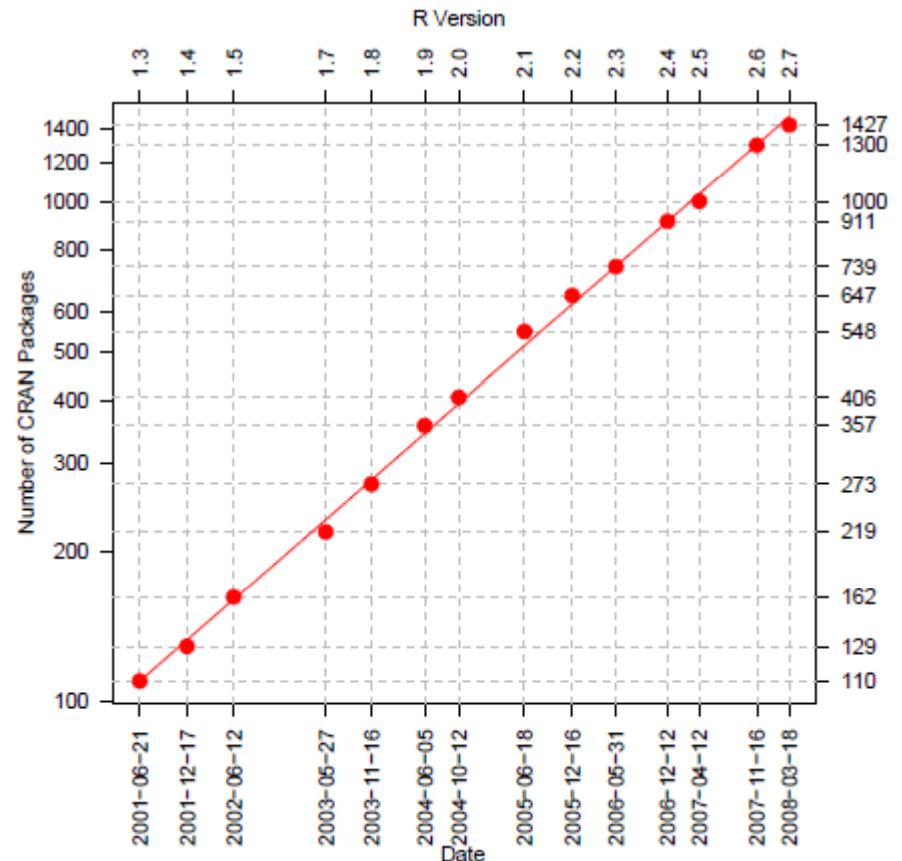


Figure 1: The number of R packages on CRAN has grown exponentially since R version 1.3 in 2001. Source of Data: <https://svn.r-project.org/R/branches/>.

# Growing interest in R

---

- August 2006



## Limited Attendance Seminar on Predictive Modeling

August 15-16, 2006  
Chicago, Illinois

**NEW!** The Limited Attendance Seminar on Predictive Modeling is now full. No additional registrations will be taken at this time. Thank you for your interest.

### PURPOSE

This two-day limited attendance seminar is designed to impart an understanding of predictive modeling applications in insurance. Examples of personal lines ratemaking and commercial lines underwriting applications will be covered in detail using simulated data. The seminar format will be a mixture of lecture, demonstration, and hands-on exercises. Pre-assignments will be given to the participants to familiarize them with some of the concepts to be covered. Datasets that will be used during the seminar will also be provided.

The seminar will cover both theoretical and practical aspects of insurance predictive modeling. Topics to be covered will include:

- Data scrubbing and manipulation
- Modeling methodologies
- Model validation
- Interpreting and analyzing model results

The statistical software that will be used during the seminar is the widely used, freely available shareware package "R". Prior to the seminar, students will be provided with instructions on how to install and use R. Participants are expected to bring their own laptop to the seminar with the R software installed.

The instructors are James Guszczka, Ph.D., FCAS, and Jun Yan, Ph.D. Both instructors are members of Deloitte Consulting LLP's Advanced Quantitative Services group.



# Growing interest in R

---

- November 2006

<http://www.casact.org/newsletter/index.cfm?fa=viewart&id=5311>

CLICK HERE TO DOWNLOAD A .PDF VERSION OF THIS NEWSLETTER.

**Actuarial Review**

RETURN TO MAIN PAGE

**The R Programming Language—My “Go To” Computational Software**  
Glenn Meyers

My involvement in a number of predictive modeling projects in the past few years has given me the opportunity to work with professional statisticians. These statisticians introduced me to something that I believe will be useful to many actuaries.

The R programming language is a software environment for statistical computing and graphics. R is widely used for statistical software development and data analysis. R’s source code is free and available at the Web Site [www.r-project.org](http://www.r-project.org) where precompiled binary versions are provided for Microsoft Windows, Mac OS X, and other UNIX-like operating systems.

R is the result of a collaborative effort with contributions from all over the world. R was initially written by Robert Gentleman and Ross Ihaka—also known as “R & R” of the Statistics Department of the University of Auckland. Since mid-1997 there has been a core group with access to write the actual source code for R.

R supports a wide variety of statistical and numerical techniques. R is also highly extensible through the use of packages, which are user-submitted libraries for specific functions or areas of study. A core set of packages are included with the installation of R, with over 700 more available at the [Comprehensive R Archive Network](http://www.cran.r-project.org) (CRAN) as of 2006.

# Growing interest in R

---

- November 2008 – CAS Annual Meeting, Seattle

## **C-17: LOSS RESERVING WITH R**

**Tuesday, November 18, 10:00 a.m. – 11:30 a.m.**

R is a free, open-source (GPL-licensed) software environment that has become very popular in academic, scientific, and financial communities for statistical modeling and problem solving. CAS members may be familiar with the application of R to predictive modeling. This session will show how R can also be used for reserving. Markus Gesmann wrote the R ChainLadder package which carries out some of the basic deterministic and stochastic reserving methods familiar to casualty actuaries. Vincent Goulet wrote the R package actuar that provides additional R functionality in loss distribution modeling, credibility theory, and risk and ruin theory. Vincent will begin this session with a brief introduction to the R language and actuar. Dan Murphy will show how to use R with Excel via the add-in RExcel. Markus will then give a live demonstration of the capabilities of his ChainLadder package. The session will focus on R as a tool rather than on advanced actuarial techniques. Attendees can expect to leave the session somewhat more at ease with the notion that actuarial reserving methods and models need not be relegated to the realm of the spreadsheet.

**Moderator:**

Simon Lilley, Senior Actuarial Associate, SAFECO Insurance Companies

**Panelists:**

Markus Gesmann, Chief Analyst, Lloyd's of London

Vincent Goulet, Associate Professor, Université Laval

Daniel Murphy, Trinostics

# Growing interest in R

---

- January 2009

[http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?\\_r=1&pagewanted=print](http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?_r=1&pagewanted=print)

**The New York Times**

This copy is for your personal, noncommercial use only. You can order presentation-ready copies for distribution to your colleagues, clients or customers [here](#) or use the "Reprints" tool that appears next to any article. Visit [www.nytreprints.com](http://www.nytreprints.com) for samples and additional information. [Order a reprint of this article now.](#)

PRINTER-FRIENDLY FORMAT  
SPONSORED BY

**(500) DAYS OF  
SUMMER**

January 7, 2009

## Data Analysts Captivated by R's Power

By [ASHLEE VANCE](#)

To some people R is just the 18th letter of the alphabet. To others, it's the rating on racy movies, a measure of an attic's insulation or what pirates in movies say.

R is also the name of a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca partly because data mining has entered a golden age, whether being used to set ad prices, find new drugs more quickly or fine-tune financial models. Companies as diverse as [Google](#), [Pfizer](#), [Merck](#), [Bank of America](#), the InterContinental Hotels Group and Shell use it.

But R has also quickly found a following because statisticians, engineers and scientists without computer programming skills find it easy to use.

"R is really important to the point that it's hard to overvalue it," said Daryl Pregibon, a research scientist at Google, which uses the software widely. "It allows statisticians to do very intricate and complicated analyses without knowing the blood and guts of computing systems."

# Growing interest in R

- April 2009

[http://www.actuaries.org.uk/media\\_centre/news\\_stories/2009/april/r\\_you\\_ready](http://www.actuaries.org.uk/media_centre/news_stories/2009/april/r_you_ready)

- Interest in the UK actuarial community

The screenshot shows a web browser window with the URL [http://www.actuaries.org.uk/media\\_centre/news\\_stories/2009/april/r\\_you\\_ready](http://www.actuaries.org.uk/media_centre/news_stories/2009/april/r_you_ready). The page header includes the logo for 'The Actuarial Profession' and the tagline 'making financial sense of the future'. A navigation menu contains links for 'The Profession', 'Media centre', 'Jobs', 'Events', and 'Publications'. A secondary menu lists 'Careers', 'Students', 'Members', 'Practice areas', 'Regulation', and 'Knowledge services'. A sidebar on the left has a 'Home' button and a menu with 'Media centre' and 'R you ready?'. The main content area features the article title 'R you ready for something useful?' with a breadcrumb trail: 'Actuaries | Media centre | News Stories | 2009 | April | R you ready for something useful?'. The article text states: 'At a previous GIRO, the R Toolkit Working Party (all R experts) did some great work in introducing the UK actuarial profession to R. This is an amazing language and environment which every actuary should have as part of his or her toolkit. Now the 2009 GIRO R Working Party has set up a beginners R workstream for believe it or not beginners. This workstream will help to: point you in the right direction for useful learning material for beginners; get you onto useful mailing lists; ask you to choose a work-related problem and solve it using R; put you in touch with an expert who will encourage you through any early teething problems, help you choose your work-related problem, and so on; and run an initial one-day workshop at the Institute to get you up to speed. To participate in this workstream, or to express an interest in a one-day workshop or just to get further information, contact Neil Hilary at Staple Inn and he will get in touch. If you are an expert you can guess why we need you. If you are happy to mentor a beginner through the early stages please let Neil know at the above address. In addition if you have suggestions for other more specific (expert) workstreams please let us know.'

# On to Bigger Things?

- A company that aspires to be R's "Redhat"
  - Enterprise versions of R



Customer Login

Products & Services

Downloads

Why Revolution R?

Support

News & Events

About Us

**scale**

Big Data Analytics with  
**Revolution R Enterprise**

[Learn More Now](#)

**QUICKLINKS:**

- [Which R is Right for Me?](#)
- [Getting Started with Revolution R Enterprise](#)
- [Solutions for Finance & Life Sciences](#)
- [Free On-Demand Webinars](#)

[Buy Now](#)



### Changing the Face of Analytics

Watch CEO Norman H. Nie interviewed on Fox Business Network

### LATEST NEWS & EVENTS

[Revolution Analytics Brings 'Big Data' Analysis to R](#)

[August 25th Big Data Webinar](#)

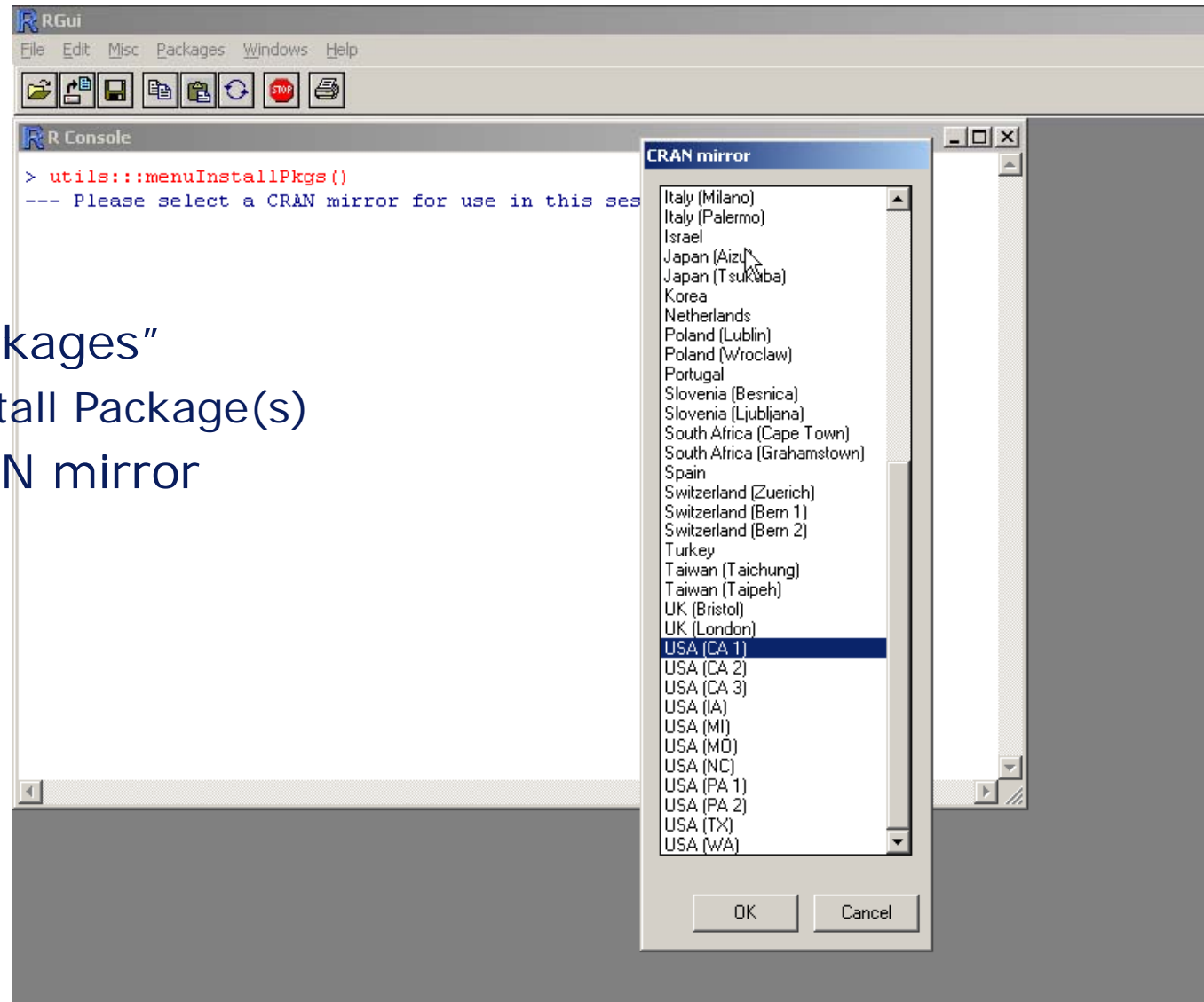
# Installing R

---

- Go to <http://cran.r-project.org/>
  - Or just type "R" into Google and click "I feel lucky"
- Click on "Download CRAN" on the left of the screen
- Click on one of the USA CRAN mirror sites
- Click on "Windows (95 and later)"
- Click on "base"
- Right-click on R-2.11.1-win32.exe
  - "Save target as" into any directory
- After you've downloaded this setup program, double-click on it and follow the instructions

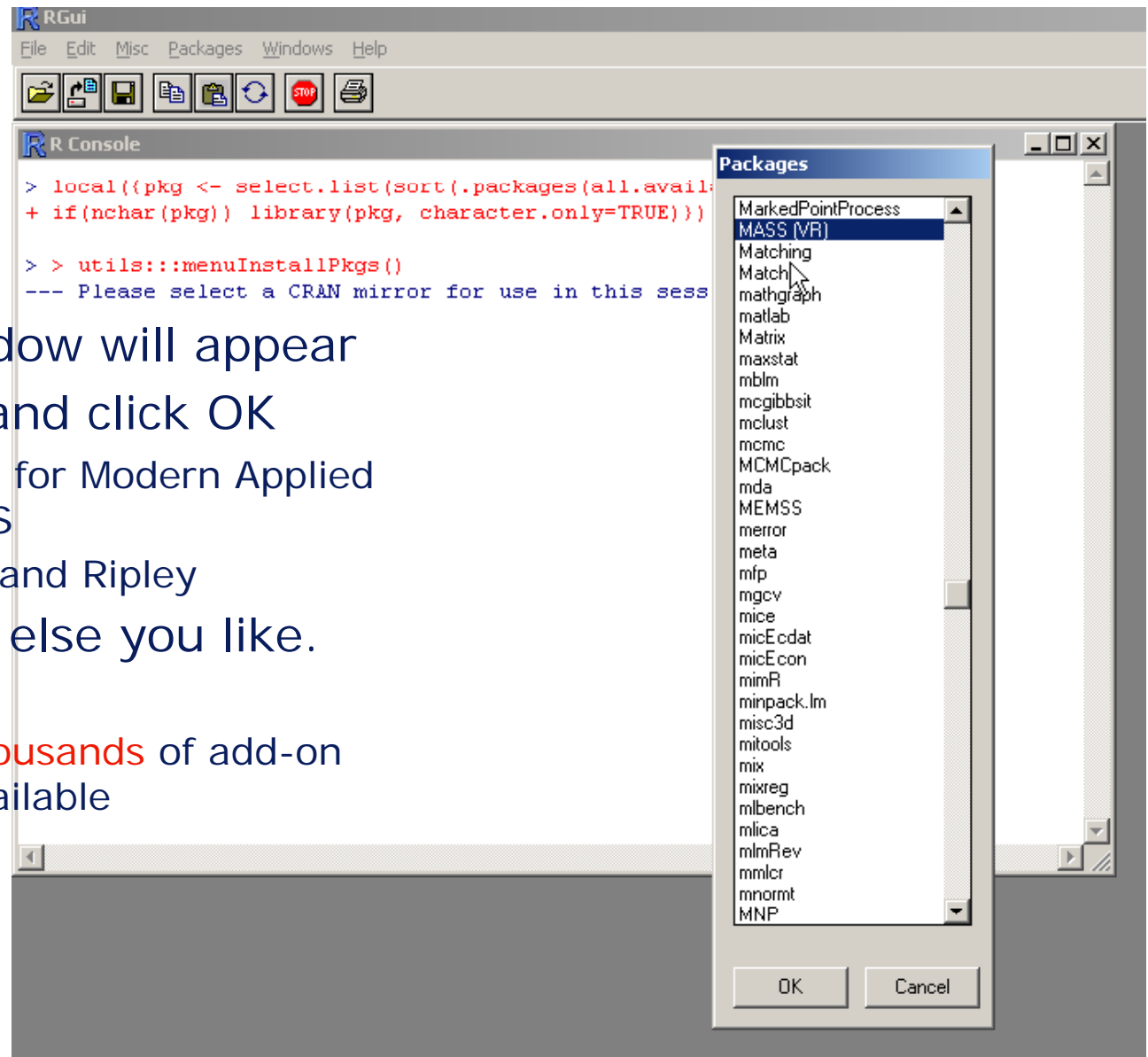
# Add-on Packages

- Click on “Packages”
  - Select “Install Package(s)”
- Select a CRAN mirror



# Add-on Packages

- “Packages” window will appear
- Select “MASS” and click OK
  - MASS stands for Modern Applied Statistics in S
  - By Venables and Ripley
- ... add anything else you like.
  - It's all free
  - There are **thousands** of add-on packages available





# R Warm-Up



R as a Calculator

Assignments

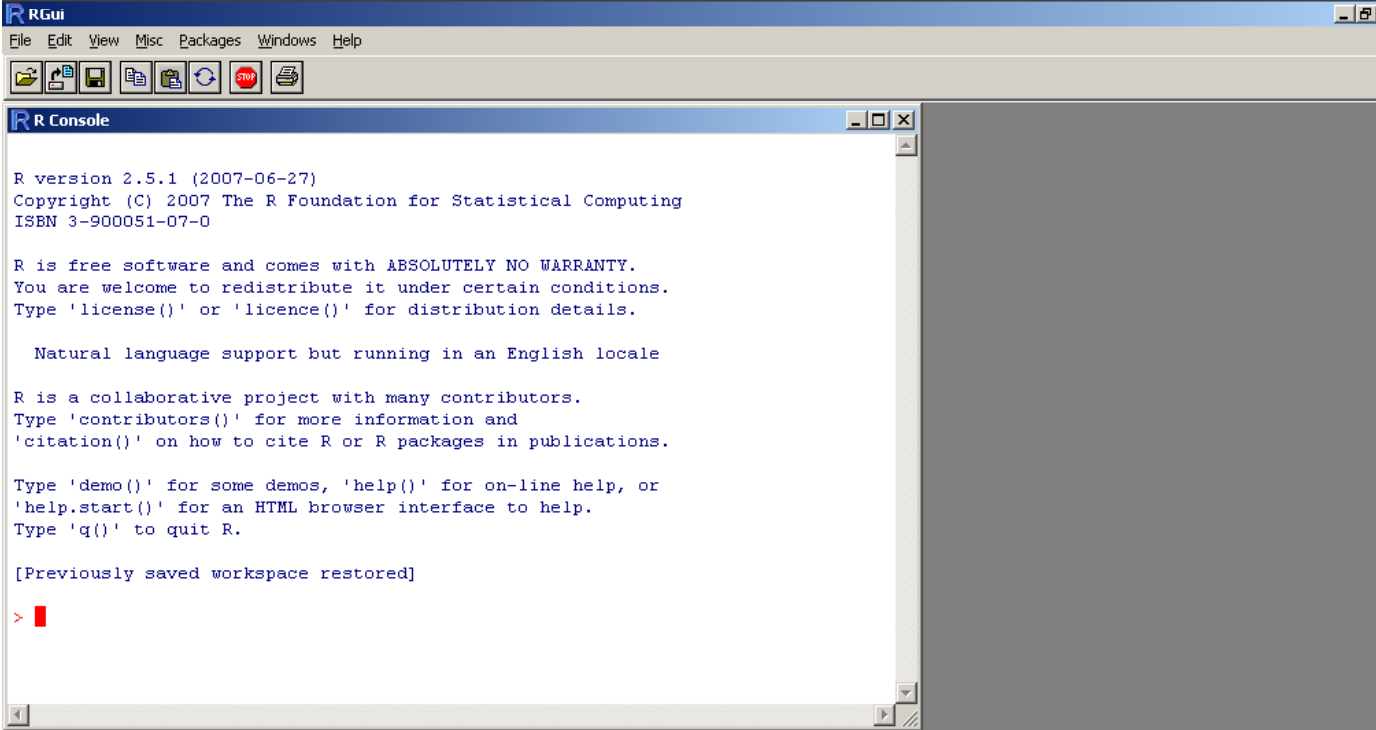
Vectors, Matrices, Data Frames

Getting Help

# Getting Started with R

---

- Double-click on the “R” icon to start the program
- You will see the following screen:



The screenshot shows the RGui application window. The title bar reads "RGui". The menu bar includes "File", "Edit", "View", "Misc", "Packages", "Windows", and "Help". Below the menu bar is a toolbar with icons for file operations and execution. The main window contains an "R Console" pane. The console displays the following text:

```
R version 2.5.1 (2007-06-27)
Copyright (C) 2007 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
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.

[Previously saved workspace restored]

> █
```

Below the console window, a dark grey box contains the following text:

**Note: you can always click ctrl-I to clear the screen**

# Getting Started with R

---

- This screen gives you the “command line”.
  - Type commands at the red “>”
- You can use R as a calculator.
  - Type “2+3” at the command line and hit enter
  - Similarly “2-3”, “2\*3”, “2/3”, “2^3” (or “2\*\*3”)
- R handles vectors very naturally.
  - Type “c(1,2,3,4,5)” at the command line and hit enter
  - “c” stands for “concatenate”
  - This is how to create a vector of numbers
  - Alternately:
    - Type “1:5”
    - Type “seq(1,5)”
    - What is “seq”? Don’t ask me... go to the online help
    - “help(seq)” or better, “?seq”

# Getting Started with R

---

- An aside: test to see whether your additional libraries were successfully added.
  - Type "library(MASS)"
    - Note: R is case-sensitive!
  - Type "library(statmod)"
  - If there are no error messages you're ok
  - Type "?tweedie" to see documentation of Tweedie GLM
  - Type "?truehist" for a description of a graphical histogram package from MASS.

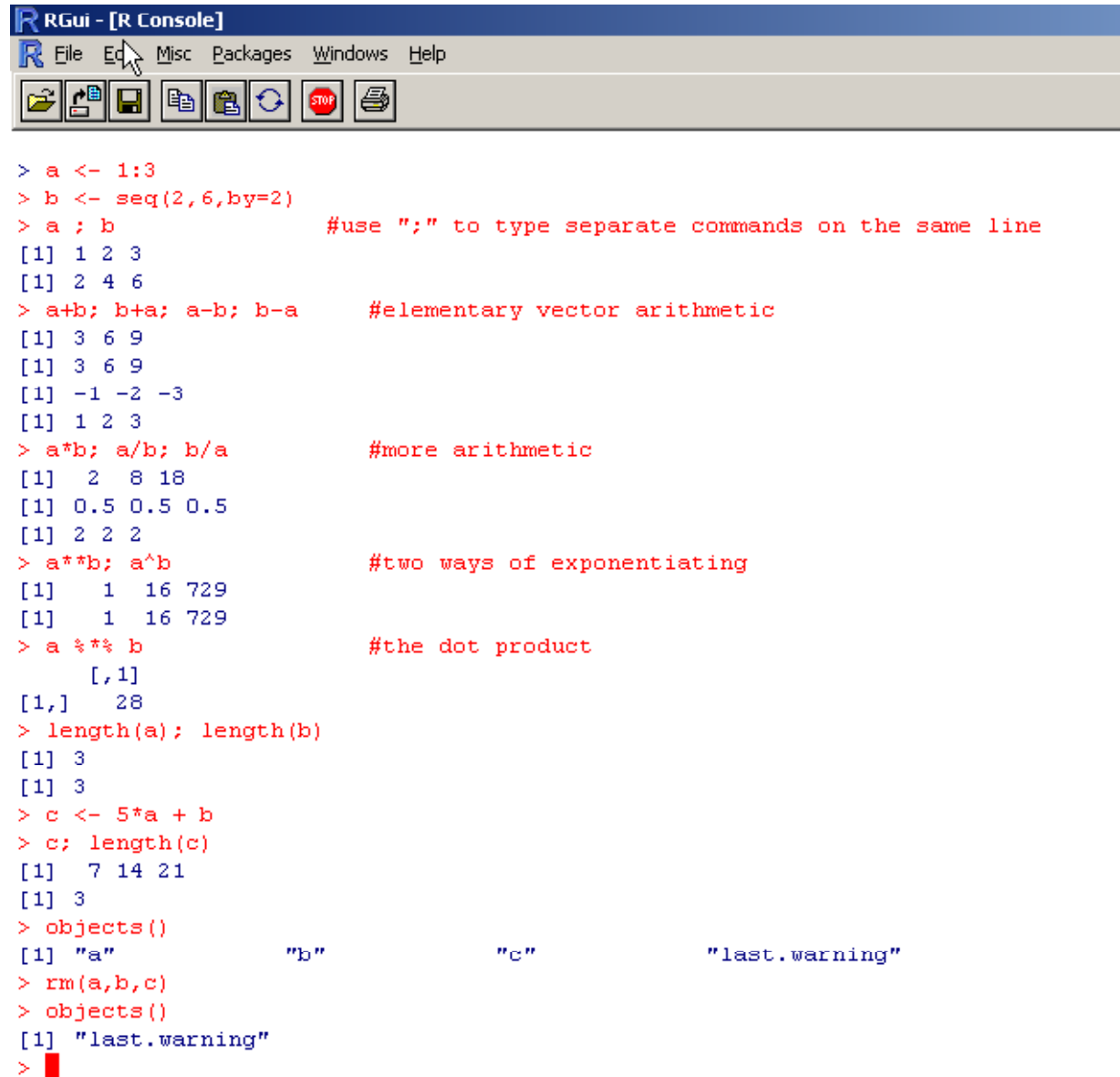
# Assignments

---

- Suppose you want to set the variable `x` to equal 5
- Type `"x <- 5"`
  - Combine the less than sign `"<"` and the minus sign `"-"`
  - You could also type `"5->x"` or `"x=5"`
- In words: `"x gets 5"`
- Now type `"x"` at the command line
- Now type `"objects()"`
  - `x` has been saved as an R object
- Now type `"rm(x)"`
  - To remove the object `x` if we're done with it
- Now type `"objects()"` again
  - The object `x` is gone

# Working with Vectors

- R handles vectors very naturally
- Type these commands into your R session to gain comfort.
  - Feel free to play around

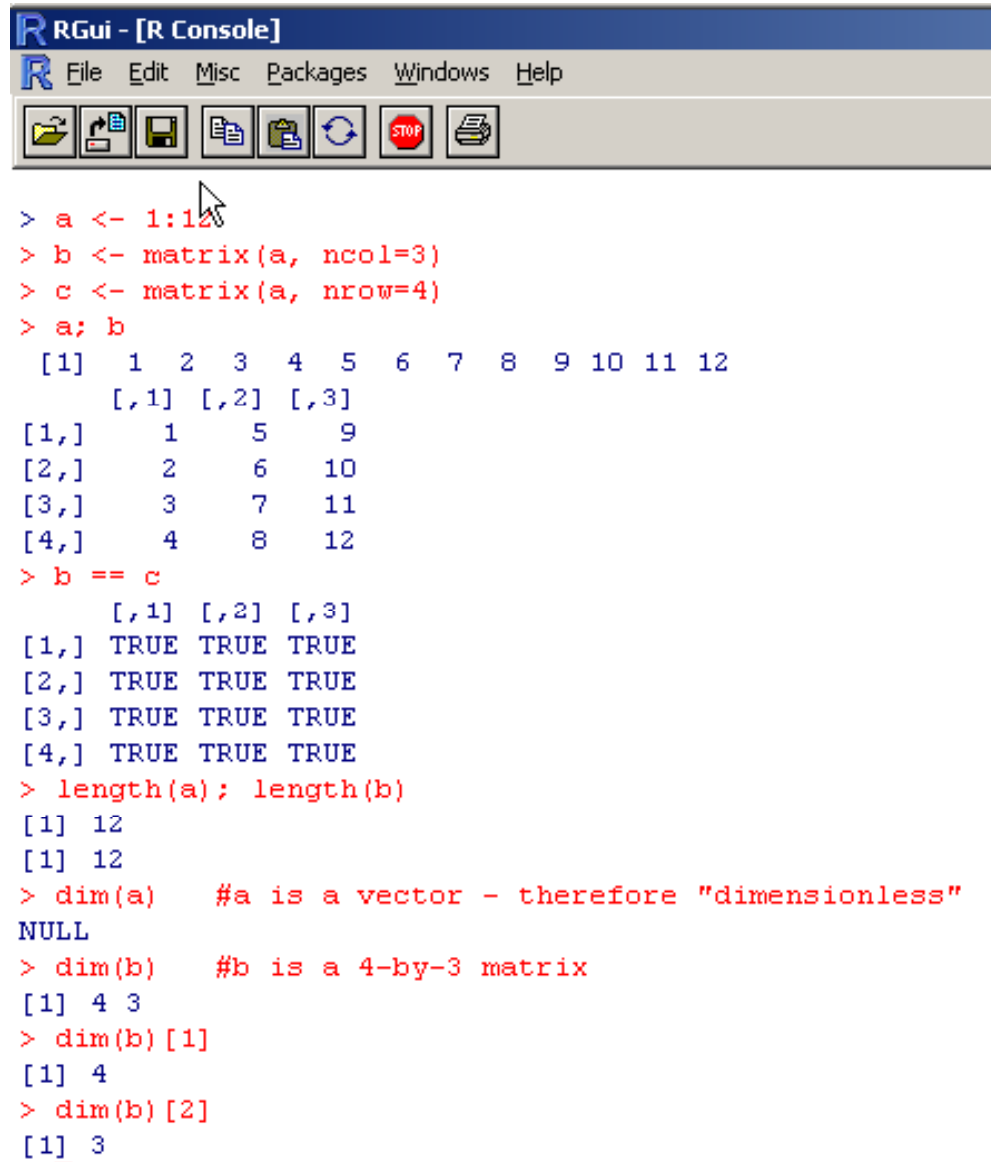


```
RGui - [R Console]
File Edit Misc Packages Windows Help

> a <- 1:3
> b <- seq(2,6,by=2)
> a ; b #use ";" to type separate commands on the same line
[1] 1 2 3
[1] 2 4 6
> a+b; b+a; a-b; b-a #elementary vector arithmetic
[1] 3 6 9
[1] 3 6 9
[1] -1 -2 -3
[1] 1 2 3
> a*b; a/b; b/a #more arithmetic
[1] 2 8 18
[1] 0.5 0.5 0.5
[1] 2 2 2
> a**b; a^b #two ways of exponentiating
[1] 1 16 729
[1] 1 16 729
> a %*% b #the dot product
 [,1]
[1,] 28
> length(a); length(b)
[1] 3
[1] 3
> c <- 5*a + b
> c; length(c)
[1] 7 14 21
[1] 3
> objects()
[1] "a" "b" "c" "last.warning"
> rm(a,b,c)
> objects()
[1] "last.warning"
>
```

# Working with Matrices

- This screen illustrates how to create a matrix from a vector
- R treats vectors as “dimensionless”
  - So `dim(a)=NULL`
- Note that “`==`” is a logical test
  - “`!=`” is a similar test for “not equals”
  - Try typing “`b!=c`”



```
RGui - [R Console]
File Edit Misc Packages Windows Help

> a <- 1:12
> b <- matrix(a, ncol=3)
> c <- matrix(a, nrow=4)
> a; b
 [1] 1 2 3 4 5 6 7 8 9 10 11 12
      [,1] [,2] [,3]
[1,]  1   5   9
[2,]  2   6  10
[3,]  3   7  11
[4,]  4   8  12
> b == c
      [,1] [,2] [,3]
[1,] TRUE TRUE TRUE
[2,] TRUE TRUE TRUE
[3,] TRUE TRUE TRUE
[4,] TRUE TRUE TRUE
> length(a); length(b)
 [1] 12
 [1] 12
> dim(a) #a is a vector - therefore "dimensionless"
NULL
> dim(b) #b is a 4-by-3 matrix
 [1] 4 3
> dim(b)[1]
 [1] 4
> dim(b)[2]
 [1] 3
```

# The R Mindset

---

- R is designed to handle matrices naturally.
- The bracket notation “mat[row, column]” allows you to access any element of a matrix.
- Notice what happens when you leave the row or column entry blank.
- Try an analogous exercise with a vector.
  - `vec <- 5:15`
  - `vec[5], vec[6:8]`
  - `vec[c(6,2,8)]`



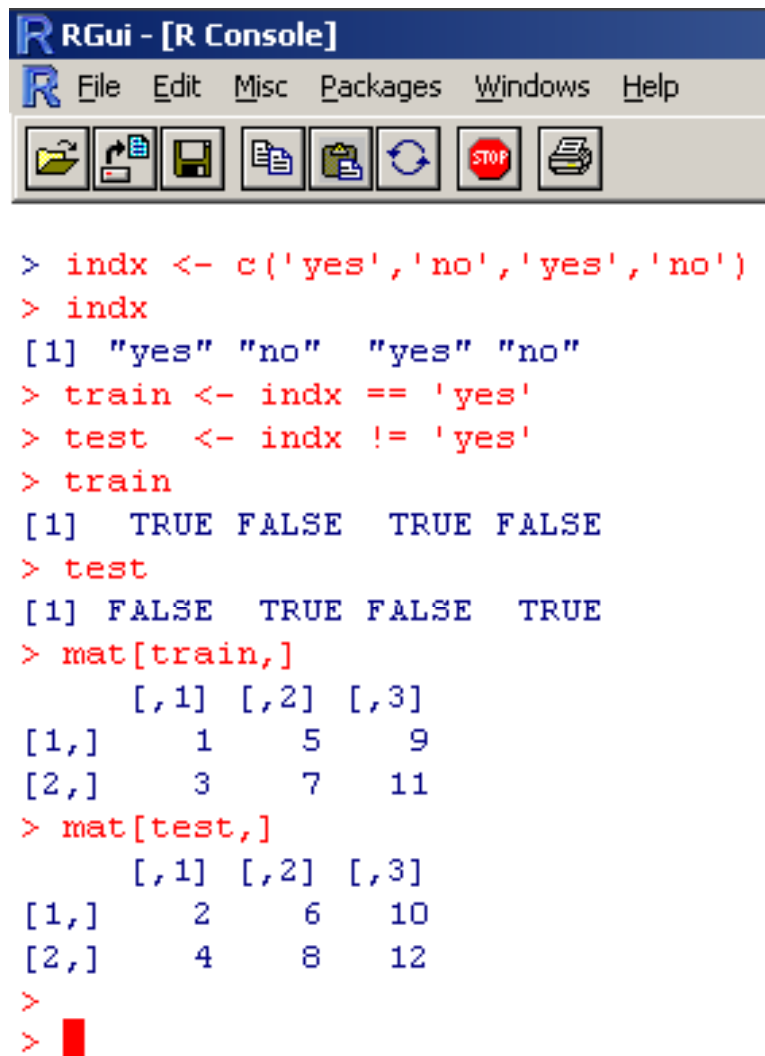
```
> mat <- matrix(1:12,ncol=3)
> mat
      [,1] [,2] [,3]
[1,]    1    5    9
[2,]    2    6   10
[3,]    3    7   11
[4,]    4    8   12
> mat[2,3]
[1] 10
> mat[3,2]
[1] 7
> mat[2,]           #returns the 2nd row
[1] 2 6 10
> mat[,2]           #returns the 2nd column
[1] 5 6 7 8
```



# The R Mindset

---

- We can get fancier by creating an index.
  - Let's use an index to divide the matrix into disjoint sets of rows.
- Think about how this trick can be used in predictive modeling projects.
  - We divide a dataset either by a random number or some other dimension.



```
RGui - [R Console]
File Edit Misc Packages Windows Help

> indx <- c('yes','no','yes','no')
> indx
[1] "yes" "no"  "yes" "no"
> train <- indx == 'yes'
> test  <- indx != 'yes'
> train
[1] TRUE FALSE TRUE FALSE
> test
[1] FALSE TRUE FALSE TRUE
> mat[train,]
      [,1] [,2] [,3]
[1,]    1    5    9
[2,]    3    7   11
> mat[test,]
      [,1] [,2] [,3]
[1,]    2    6   10
[2,]    4    8   12
>
>
```

# Example 1

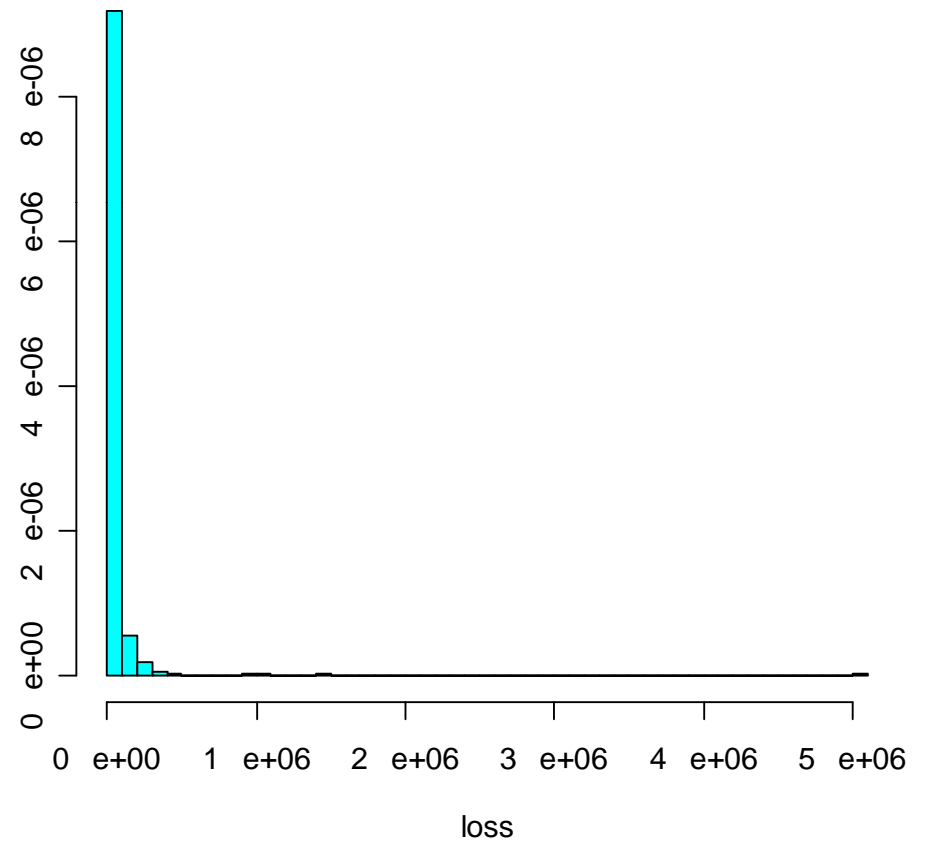


Estimating a non-trivial loss distribution

# Example 1: Fitting a Non-trivial Loss Distribution

---

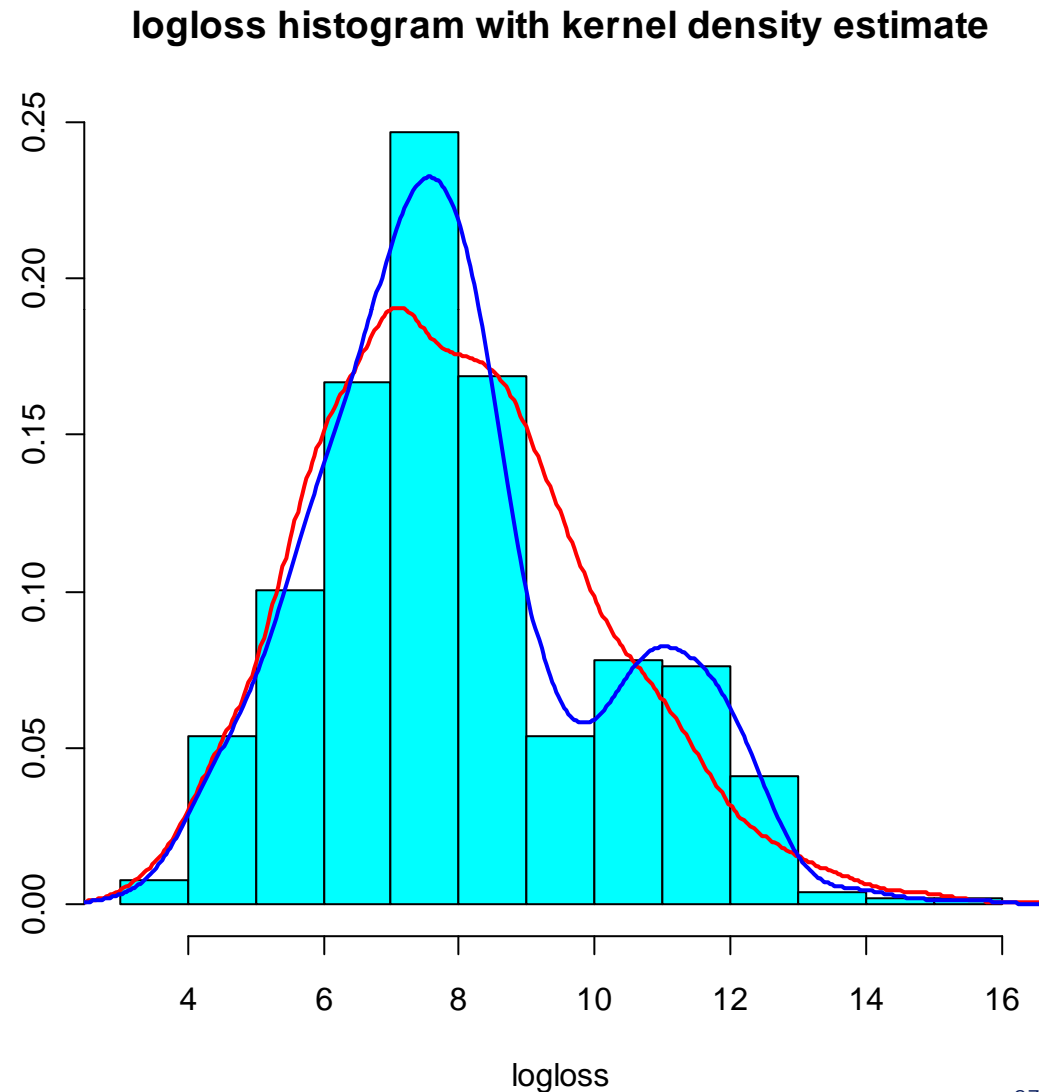
- Here is a size-of-loss histogram for 539 claims
- Let's estimate the true distribution that generated these claims.



# Graphical Data Exploration

---

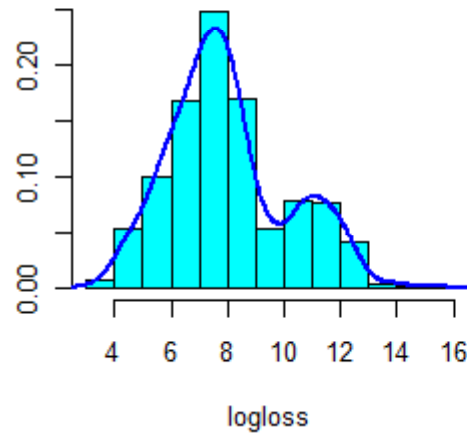
- The first attempt at data visualization didn't help much.
- But knowing how to use R effectively enables one to quickly gain insight.



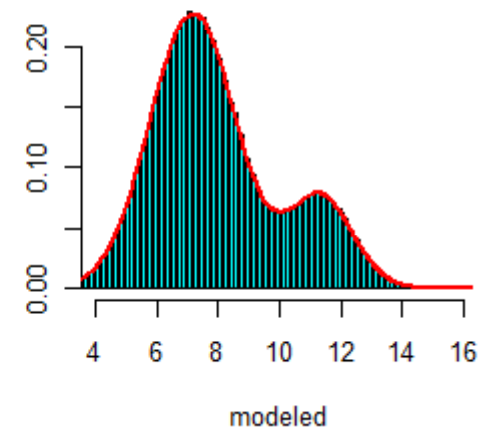
# Result of Fitting a Mixture of Lognormals

- Arriving at this solution involved:
  - Graphical data exploration
  - Writing a mixture-of-lognormals likelihood function
  - Using an optimizer to maximize the likelihood function
  - Producing the diagnostic plots to the right
- This combination of programming flexibility and graphical features illustrates R's strengths.

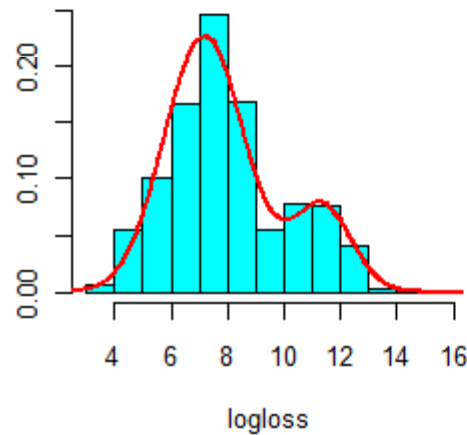
actual data w/ kernel estimate



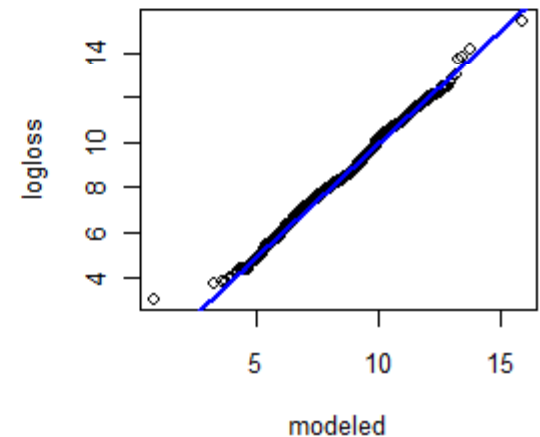
modeled loss distribution



actual data w/ modeled dist



QQ plot

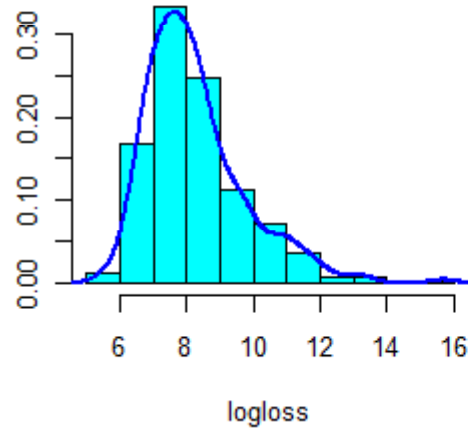


# Coda: a couple of years later...

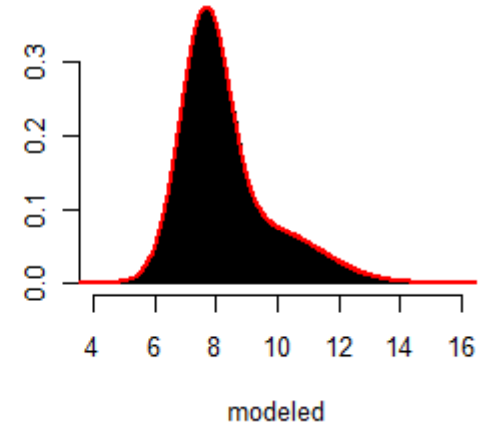
---

- I recycled precisely the same code to fit the data from an old COTOR challenge.
- Worked like a charm

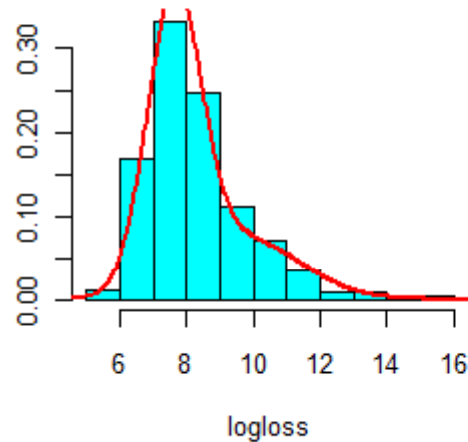
actual data w/ kernel estimate



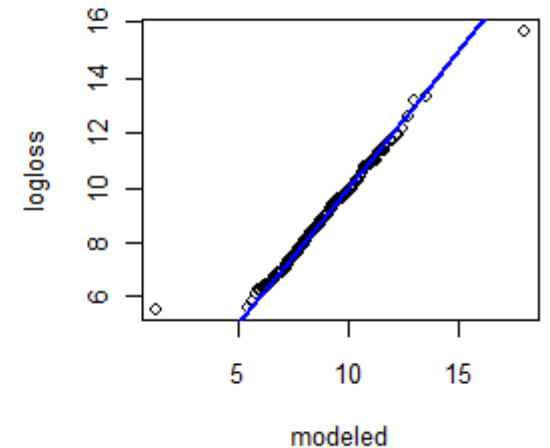
modeled loss distribution



actual data w/ modeled dist



QQ plot



# Example 2



Simple Generalized Linear Modeling Example

The dataset analyzed in this case study is discussed in the textbook:

*Generalized Linear Models for Insurance Data* by Piet de Jong and Gillian Z. Heller  
(Cambridge University Press)

## Example 2: Diabetes Deaths

---

- This dataset contains the number of deaths due to Diabetes in New South Wales, Australia in 2002.
  - From the mortality database of the Australian Institute of Health and Welfare.
- Note that this is *grouped data*.
  - There are only 16 data points, but each data point corresponds to thousands of people.
  - Because each data point corresponds to a different amount of “exposure”, it will be critical to use either a weight or an offset here.

```
> diab
  gender  age ageMid   pop deaths
1  Male  <25    20 1141100     3
2  Male 25-34    30  485571     0
3  Male 35-44    40  504312    12
4  Male 45-54    50  447315    25
5  Male 55-64    60  330902    61
6  Male 65-74    70  226403   130
7  Male 75-84    80  130527   192
8  Male  85+    90   29785   102
9 Female  <25    20 1086408     2
10 Female 25-34    30  489948     1
11 Female 35-44    40  504030     3
12 Female 45-54    50  445763    11
13 Female 55-64    60  323669    30
14 Female 65-74    70  241488    63
15 Female 75-84    80  179686   174
16 Female  85+    90   67203   159
```

Target variable: deaths

Offset variable: log(population)

Predictive vars: age, gender



# Candidate Models for Diabetes Deaths Data

---

- Let's fit a sequence of models and use the Akaike Information Criterion [AIC] to select the most appropriate one.
- Illustrate features of glm() function along the way.

```
> d0 <- glm(deaths ~ gender + age
+ , offset=log(pop), family=poisson, data=diab)
>
> d1 <- glm(deaths ~ I(gender=="Female") + C(age, base=4)
+ , offset=log(pop), family=poisson, data=diab)
>
> summary(predict(d0) - predict(d1))#test equivalence
      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
-2.487e-14 -8.882e-16  4.441e-16 -2.387e-15  1.776e-15  2.665e-15
>
> d2 <- glm(deaths ~ I(gender=="Female")
+ + ageMid + I((ageMid - mean(ageMid))^2)
+ , offset=log(pop), family=poisson, data=diab)
>
> ###add age:gender interaction
> d3 <- update(d2, .~. + ageMid : I(gender=="Female") )
>
> ###select best model
> AIC(d1); AIC(d2); AIC(d3) #the interaction is significant
[1] 104.4905
[1] 101.4789
[1] 98.97633
>
```

## Selected Model for Diabetes Deaths Data

---

- It is easy to inspect the selected model's various coefficients and statistics.

```
> summary(d3)

Call:
glm(formula = deaths ~ I(gender == "Female") + ageMid + I((ageMid -
  mean(ageMid))^2) + I(gender == "Female"):ageMid, family = poisson,
  data = diab, offset = log(pop))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3098  -0.4498  -0.1755   0.6008   1.6254

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.524e+01  3.298e-01 -46.203 < 2e-16 ***
I(gender == "Female")TRUE -1.418e+00  4.325e-01  -3.279  0.00104 **
ageMid            1.115e-01  5.123e-03  21.761 < 2e-16 ***
I((ageMid - mean(ageMid))^2) -3.501e-04  1.453e-04  -2.409  0.01601 *
I(gender == "Female")TRUE:ageMid 1.157e-02  5.532e-03   2.091  0.03655 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

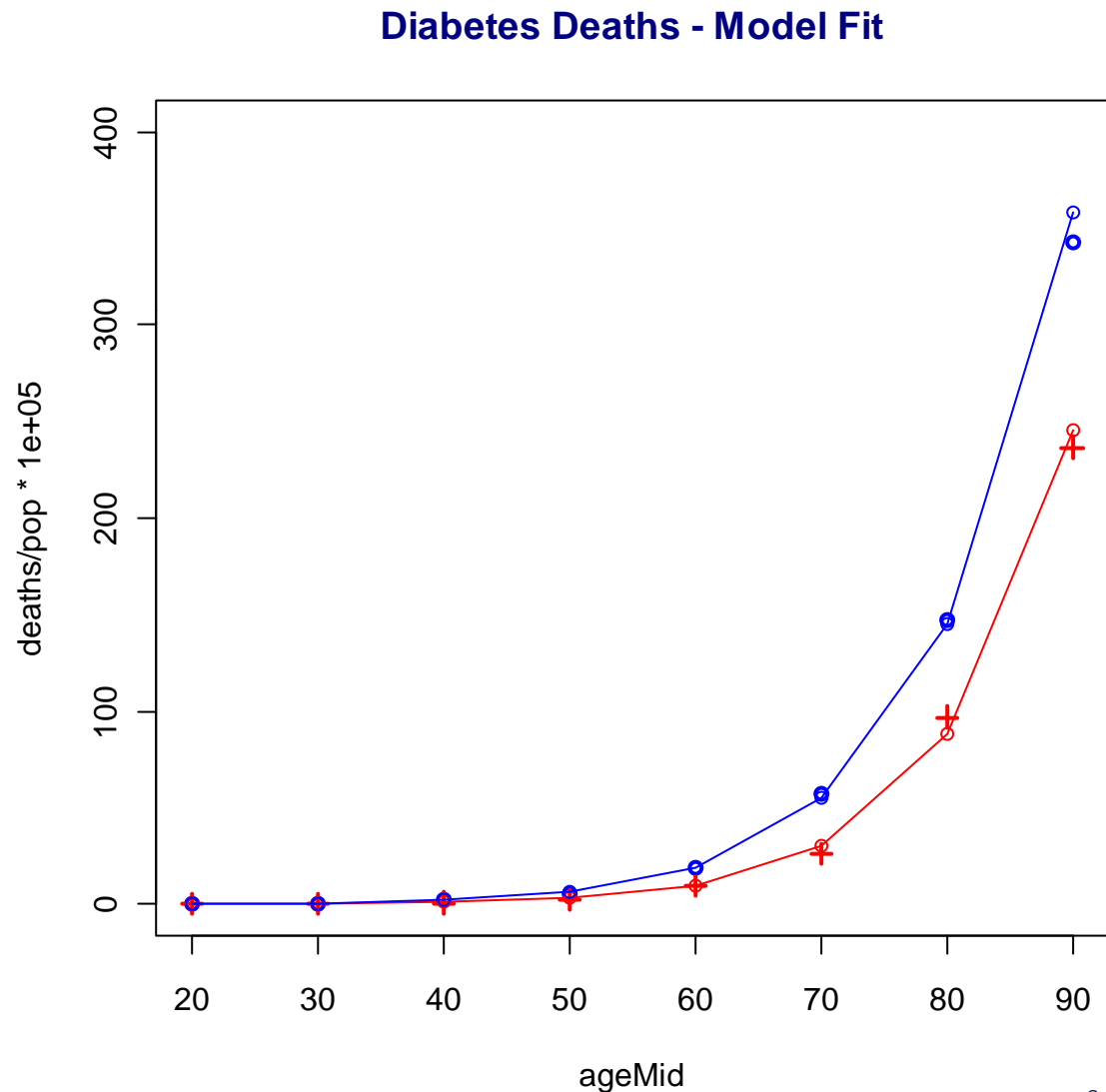
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 3306.383  on 15  degrees of freedom
Residual deviance:  13.375  on 11  degrees of freedom
AIC: 98.976
```

# Diabetes Deaths – Visual Inspection of Model Fit

---

- A great virtue of R is that it enables one to visually explore data and also create custom graphics to evaluate model fit.
- AIC is fine as far as it goes, but this picture is perhaps more informative.



# Example 3



Chain Ladder Loss Reserving

Over-Dispersed Poisson Loss Reserving

Hierarchical Growth Curve Loss Reserving Model

## Example 3: Three Approaches to Loss Reserving

- A garden variety loss triangle (Dave Clark, 2003 *Forum Paper*):

Cumulative Losses in 1000's													
AY	12	24	36	48	60	72	84	96	108	120	reported	est ult	reserve
1991	358	1,125	1,735	2,183	2,746	3,320	3,466	3,606	3,834	3,901	3,901	3,901	0
1992	352	1,236	2,170	3,353	3,799	4,120	4,648	4,914	5,339		5,339	5,434	95
1993	291	1,292	2,219	3,235	3,986	4,133	4,629	4,909			4,909	5,379	470
1994	311	1,419	2,195	3,757	4,030	4,382	4,588				4,588	5,298	710
1995	443	1,136	2,128	2,898	3,403	3,873					3,873	4,858	985
1996	396	1,333	2,181	2,986	3,692						3,692	5,111	1,419
1997	441	1,288	2,420	3,483							3,483	5,672	2,189
1998	359	1,421	2,864								2,864	6,787	3,922
1999	377	1,363									1,363	5,644	4,281
2000	344										344	4,971	4,627
chain link	3.491	1.747	1.455	1.176	1.104	1.086	1.054	1.077	1.018	1.000	34,358	53,055	18,697
chain ldf	14.451	4.140	2.369	1.628	1.384	1.254	1.155	1.096	1.018	1.000			
growth curve	6.9%	24.2%	42.2%	61.4%	72.2%	79.7%	86.6%	91.3%	98.3%	100.0%			

- We will estimate outstanding losses using three methods:
  - Replicate the above chain-ladder spreadsheet calculation in R
  - Use the Over-dispersed Poisson GLM model (not advocated!... Just illustrated)
  - Longitudinal data analysis: Hierarchical Growth Curve Model

# Loss Reserving 101: Chain Ladder Calculation

This is just home-made chain ladder R code... see the actuar package for a fully developed version

```

> ddev <- unique(dat$dev)
> ddev <- ddev[order(ddev)]
> nn <- length(ddev)
> llink <- NULL
> for (ii in 2:nn){
+ numer <- dat[dat$dev==ddev[ii], "cum"]
+ denom <- dat[dat$dev==ddev[ii-1], "cum"]
+ denom <- denom[1:length(numer)]
+ temp <- sum(numer) / sum(denom)
+ llink <- c(llink, temp)
+ }
> llink
[1] 3.490607 1.747333 1.455050 1.176119 1.103824 1.086269 1.053874 1.076555 1.017725
> ttail <- 1
> llink <- c(llink, ttail)
> ldf <- cumprod(llink[nn:1])[nn:1]
> ldf
[1] 14.451008 4.139971 2.369309 1.628335 1.384499 1.254276 1.154664 1.095637 1.017725 1.000000
>
> cumsum(10:1)
[1] 10 19 27 34 40 45 49 52 54 55
> now <- dat[cumsum(nn:1), "cum"]
> ult <- now[nn:1] * ldf
> sum(ult)
[1] 53055.22
> sum(now)
[1] 34358.09
> reserve <- sum(ult) - sum(now)
> reserve
[1] 18697.13

```

Cumulative Losses in 1000's													
AY	12	24	36	48	60	72	84	96	108	120	reported	est ult	reserve
1991	358	1,125	1,735	2,183	2,746	3,320	3,466	3,606	3,834	3,901	3,901	3,901	0
1992	352	1,236	2,170	3,353	3,799	4,120	4,648	4,914	5,339		5,339	5,434	95
1993	291	1,292	2,219	3,235	3,986	4,133	4,629	4,909			4,909	5,379	470
1994	311	1,419	2,195	3,757	4,030	4,382	4,588				4,588	5,298	710
1995	443	1,136	2,128	2,898	3,403	3,873					3,873	4,858	985
1996	396	1,333	2,181	2,986	3,692						3,692	5,111	1,419
1997	441	1,288	2,420	3,483							3,483	5,672	2,189
1998	359	1,421	2,864								2,864	6,787	3,922
1999	377	1,363									1,363	5,644	4,281
2000	344										344	4,971	4,627
chain link	→ 3.491	1.747	1.455	1.176	1.104	1.086	1.054	1.077	1.018	1.000	34,358	53,055	18,697
chain ldf	14.451	4.140	2.369	1.628	1.384	1.254	1.155	1.096	1.018	1.000			
growth curve	6.9%	24.2%	42.2%	61.4%	72.2%	79.7%	86.6%	91.3%	98.3%	100.0%			

# Loss Reserving 201: England-Verral GLM Chain Ladder

Cumulative Losses in 1000's													
AY	12	24	36	48	60	72	84	96	108	120	reported	est ult	reserve
1991	358	1,125	1,735	2,183	2,746	3,320	3,466	3,606	3,834	3,901	3,901	3,901	0
1992	352	1,236	2,170	3,353	3,799	4,120	4,648	4,914	5,339		5,339	5,434	95
1993	291	1,292	2,219	3,235	3,986	4,133	4,629	4,909			4,909	5,379	470
1994	311	1,419	2,195	3,757	4,030	4,382	4,588				4,588	5,298	710
1995	443	1,136	2,128	2,898	3,403	3,873					3,873	4,858	985
1996	396	1,333	2,181	2,986	3,692						3,692	5,111	1,419
1997	441	1,288	2,420	3,483							3,483	5,672	2,189
1998	359	1,421	2,864								2,864	6,787	3,922
1999	377	1,363									1,363	5,644	4,281
2000	344										344	4,971	4,627
chain link	3.491	1.747	1.455	1.176	1.104	1.086	1.054	1.077	1.018	1.000	34,358	53,055	18,697
chain ldf	14.451	4.140	2.369	1.628	1.384	1.254	1.155	1.096	1.018	1.000			
growth curve	6.9%	24.2%	42.2%	61.4%	72.2%	79.7%	86.6%	91.3%	98.3%	100.0%			

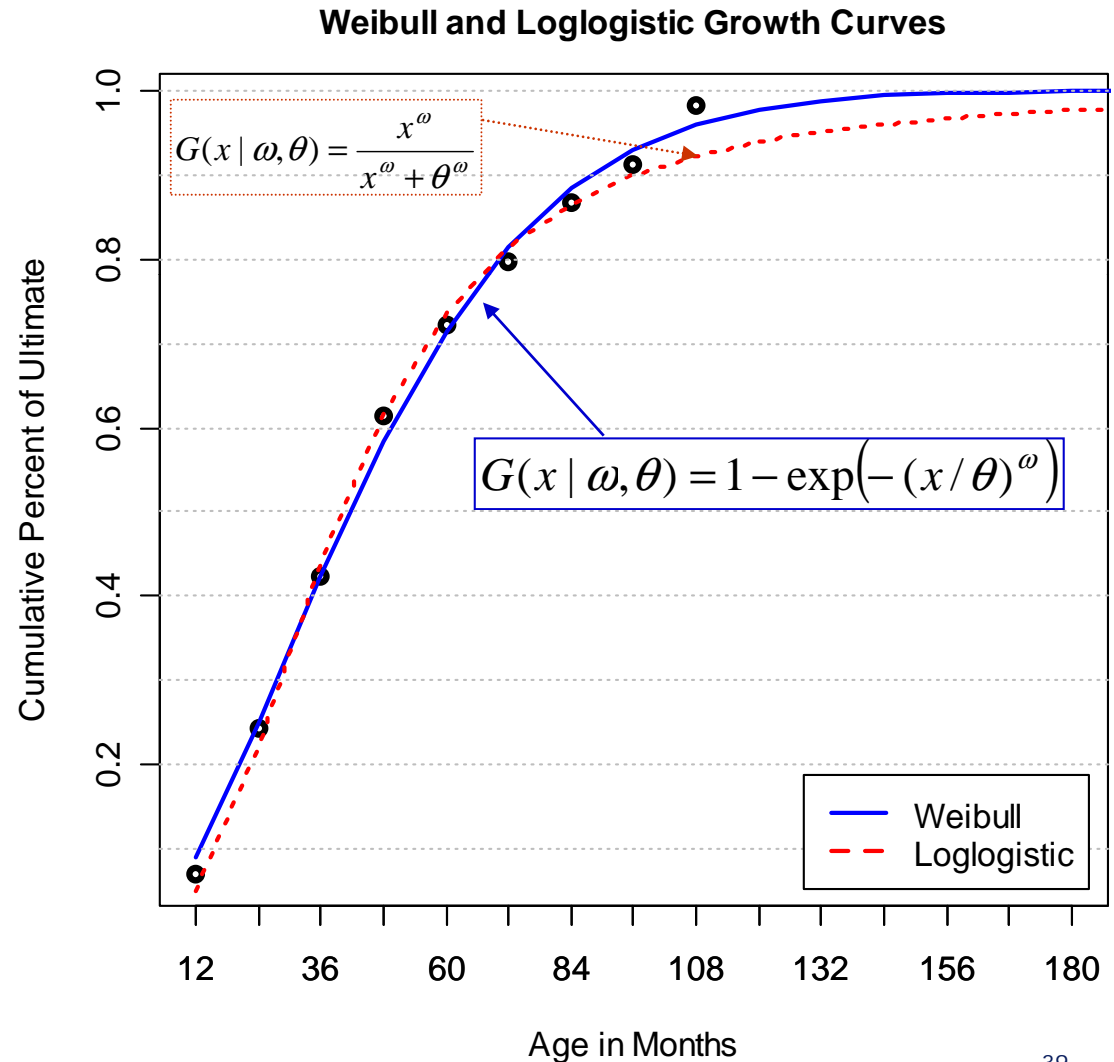
```

> ###this GLM model is equivalent to the chain ladder
> ###we are using incremental (not cumulative) losses
> g1 <- glm(inc ~ factor(AY) + factor(dev) - 1, family="quasipoisson", data=dat)
>
> ###create a dataset for the "lower triangle"
> AY <- rep(1992:2000, times=1:9)
> dev <- NULL
> for (ii in 0:8) dev <- c(dev, ddev[10:(10-ii)])
> future <- data.frame(AY, dev)
>
> eta <- predict(g1, future)
> yhat <- exp(eta)
> sum(yhat)#calculate outstanding losses (lower triangle)
[1] 18697.13
> reserve #confirm same answer as chain-ladder
[1] 18697.13

```

# Something Completely Different: Growth Curves

- Let's build a **non-linear** model of the loss triangle.
  - Are GLMs natural models for loss triangles?
- Uses growth curve to model the loss development process
  - 2-parameter curves
  - $\theta$  = scale
  - $\omega$  = shape
- Roughly speaking, we fit these curves to the LDFs and add random effects to  $\theta$  and/or  $\omega$  to allow the curves to vary by year.





# Hierarchical Growth Curve Model

---

$$\begin{aligned} CumLoss_{AY,dev} &= ULT_{AY} \left[ 1 - \exp\left(- (dev / \theta)^\omega\right) \right] + \varepsilon_{AY,dev} \\ ULT_{AY} &\sim N(\mu_{ULT}, \sigma_{ULT}^2) \\ Var(\varepsilon_{AY,dev}) &= \sigma^2 \hat{C}L_{AY,dev} \end{aligned}$$

- **Cumulative losses @ dev = (Ult losses) \* (modeled growth)**
- We must estimate the parameters:  $\{\mu_{ULT}; \omega; \theta; \sigma_{ULT}; \sigma\}$
- Random effects are added to ultimate loss ( $ULT$ ) parameter.
  - Analogous to a “random intercept” linear model.
- Random scale ( $\theta$ ) and shape ( $\omega$ ) effect can also be tested.
  - Do not seem to be significant.
  - Use NLME function in R. (Non-Linear Mixed Effects)

See “Hierarchical Growth Curve Models for Loss Reserving”, James Guszczka, *CAS Forum* (Fall 2008)

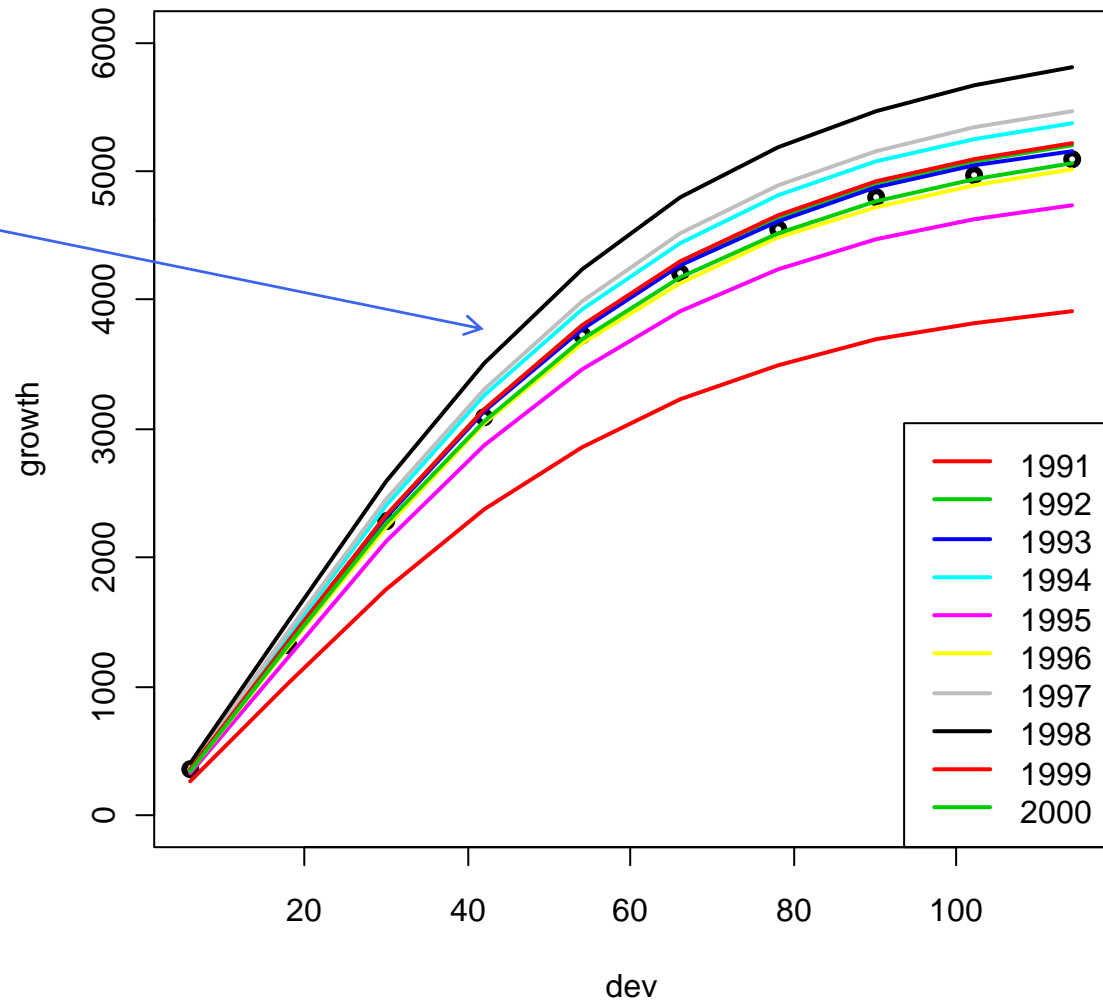
<http://www.casact.org/pubs/forum/08fforum/7Guszczka.pdf>

Copyright © 2010 Deloitte Development LLC. All rights reserved.

# Modeled Loss Development Curves

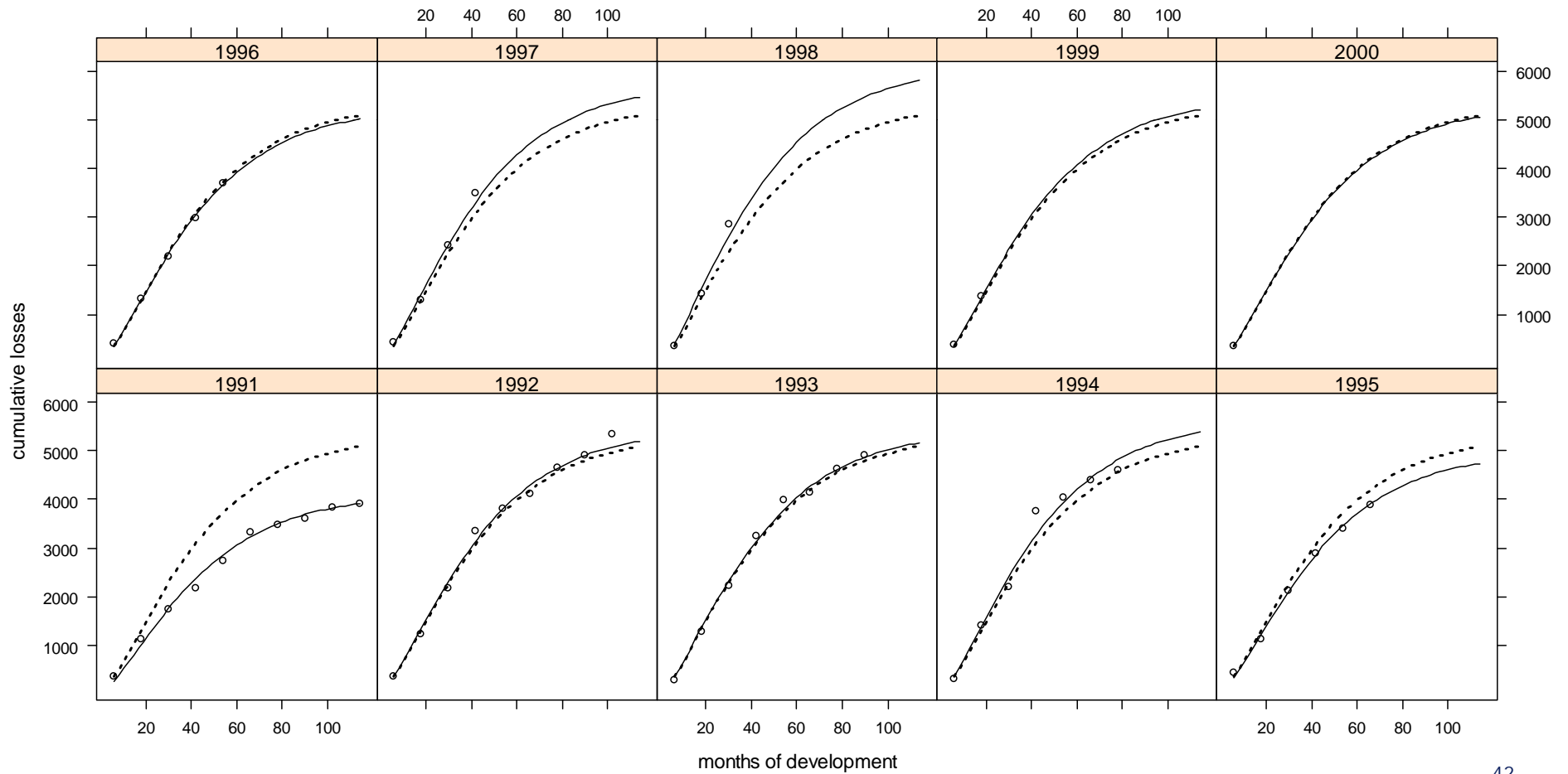
- An advantage of the hierarchical modeling framework.
- **We can allow loss development curves to vary randomly by accident year.**
- A parsimonious modeling approach that achieves good fit.
- Incorporates the insights of credibility theory.

Loss Development Patterns by AY



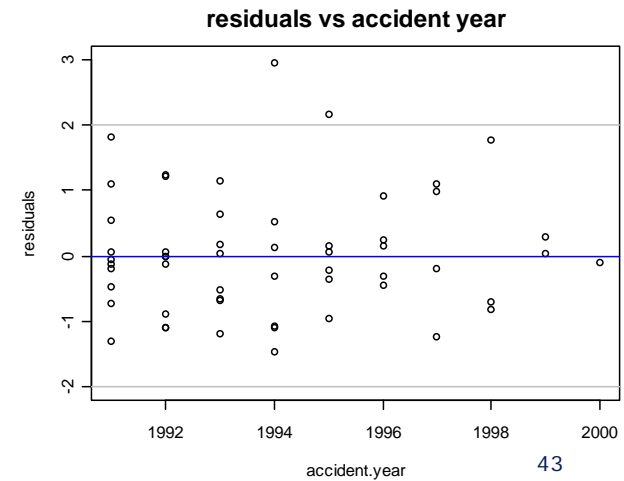
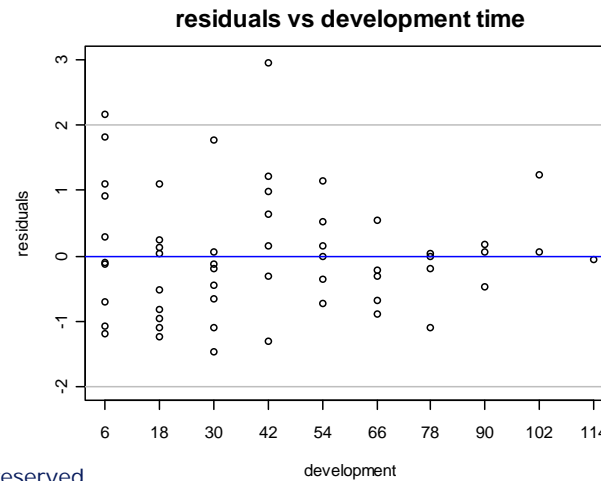
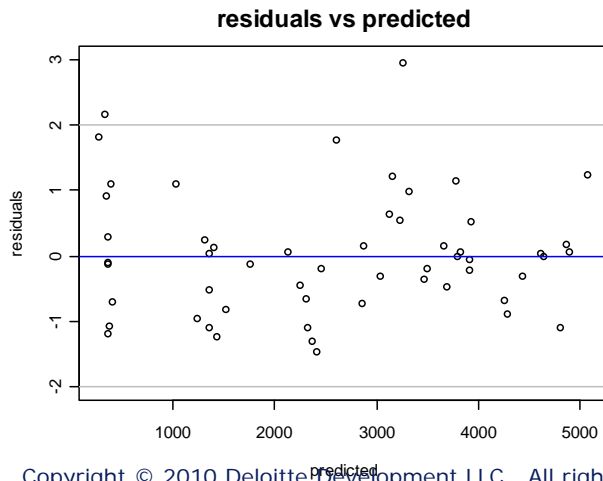
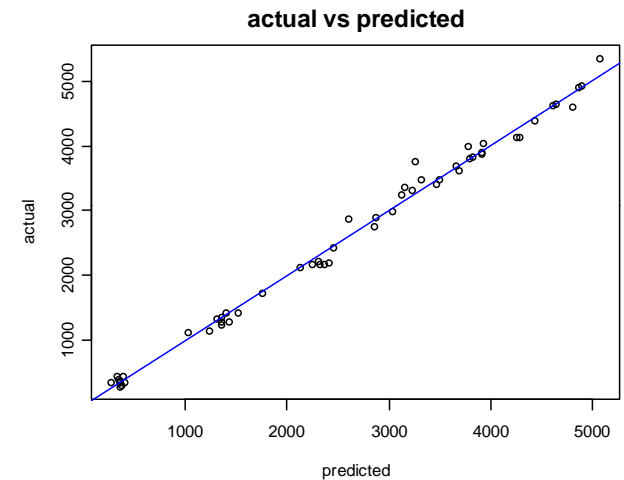
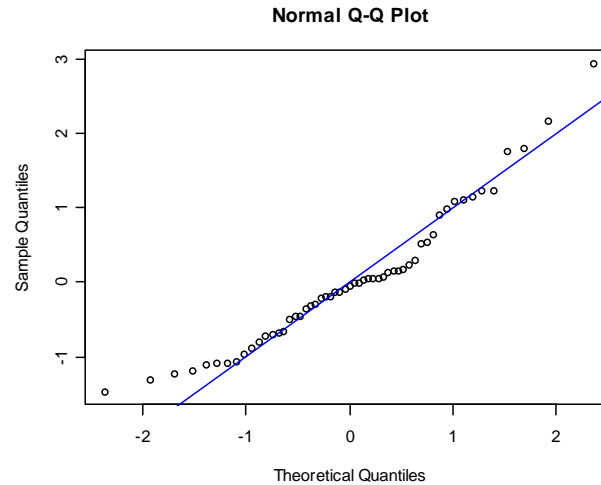
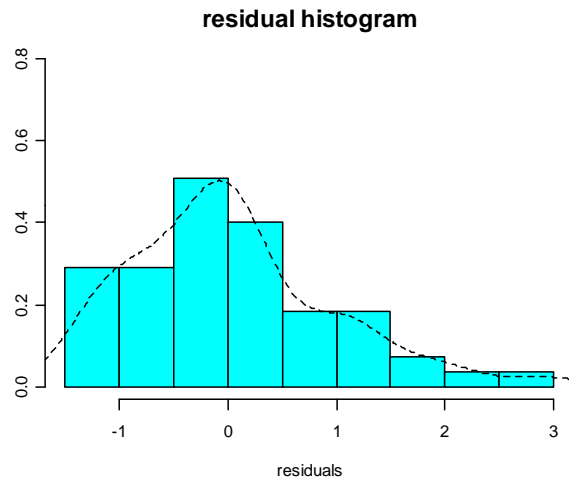
# Compare Actual to Modeled Loss Triangle

Weibull Growth Curve Model - Baseline



# Residual Analysis of Hierarchical Model

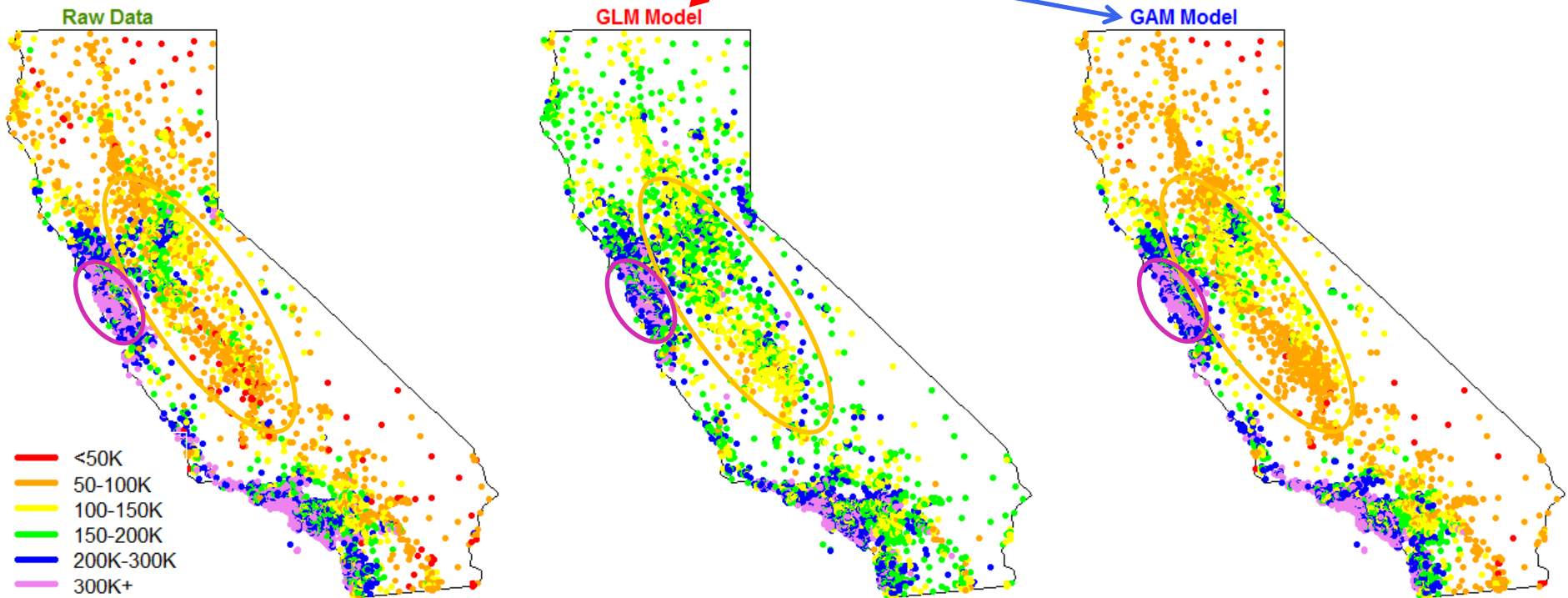
(Just try doing this in Excel!)



# GAM fit to California Home Price Data

A GAM model (added to a 3-factor GLM) does a credible job with this data. R's combined modeling and data visualization power was helpful here.

$$\log(\text{VALUE}) = \alpha + \beta_1 \text{INCOME} + \beta_2 \text{AGE} + \beta_3 \text{ROOMS} + f(\text{lat}, \text{long})$$



Further improvements could result from superimposing one or more local GAM models built for specific metropolitan areas.