

# Use of STATA in Pediatric Research - An Indian Perspective



# Who is a Pediatrician ?



Dr. Bhavneet Bharti, PGIMER- Chandigarh

# Background

---

- Research is an important part of curriculum of pediatric medicine
- Research Project is necessary for Postgraduates
- In order to fulfill their MD/DM/MCH requirements thesis mandatory



# Pediatrics-Endless queries?

- Which needle causes less pain in infants undergoing vaccination?



- Which drug is better for the treatment of Pediatric HIV, sepsis and many other diseases



# Statistics and Pediatric Research

- For answering these queries -Statistics plays increasingly important role
- It is not possible, for example, to have a new drug treatment approved for use without solid, statistical evidence to support claims of efficacy and safety

# Statistics and research

- Many new statistical methods have been developed with particular relevance for medical researchers
- these methods can be applied routinely using statistical software packages



# Statistical softwares

- Statistical knowledge of most physicians may be best described as "limited"





# Available Statistical Packages

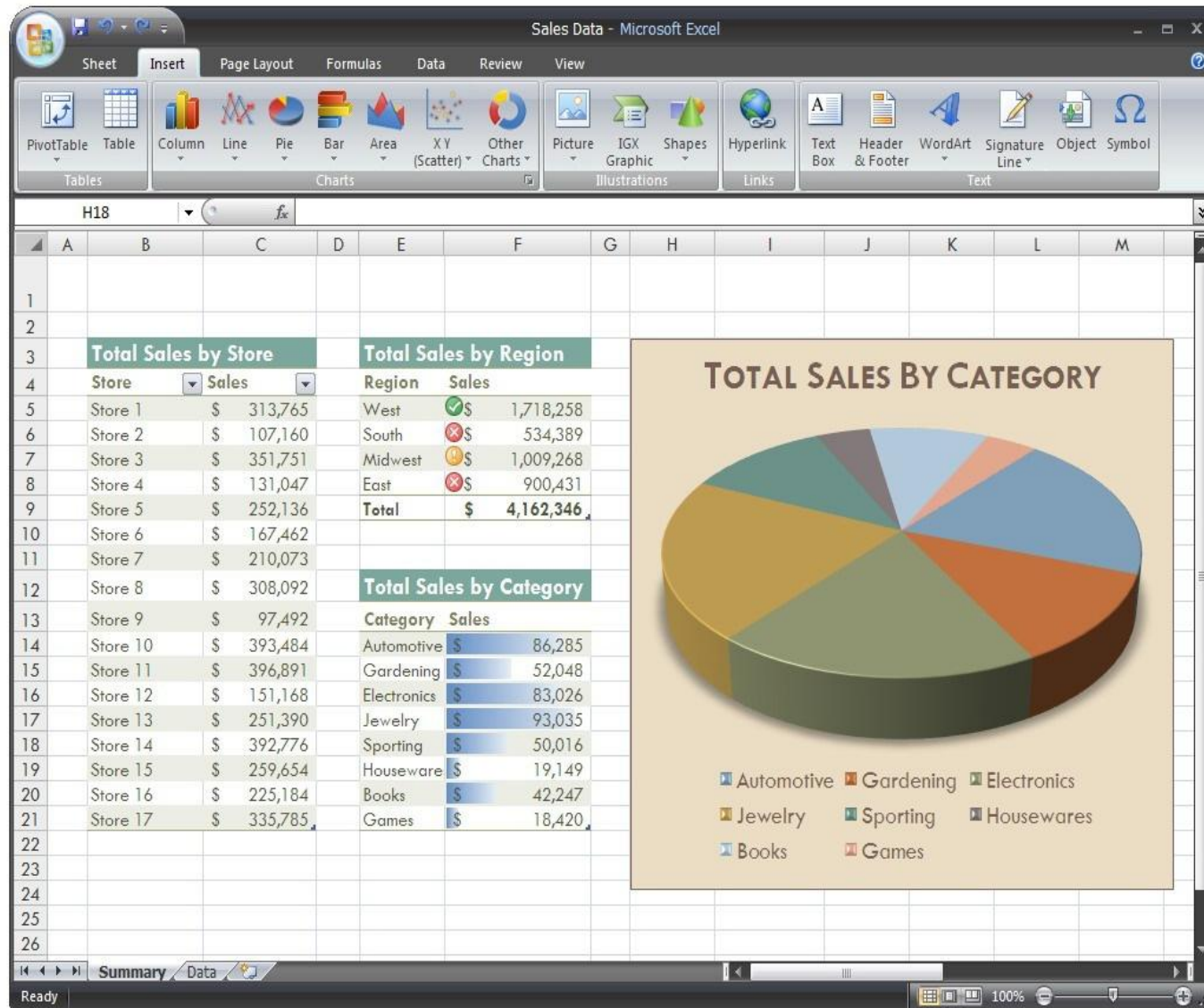
## Proprietary

- Excel
- SPSS
- STATA
- MINITAB
- SAS
- Comprehensive  
metanalysis

## Free Software

- EpiInfo
- R
- Revman
- LibreOffice Calc
- PSPP

# Microsoft Excel



# Microsoft Excel

## **COST**

- Individual License for Microsoft Office Professional \$350
- Volume Discounts available for large organizations and universities
- Free Starter Version available on new PCs

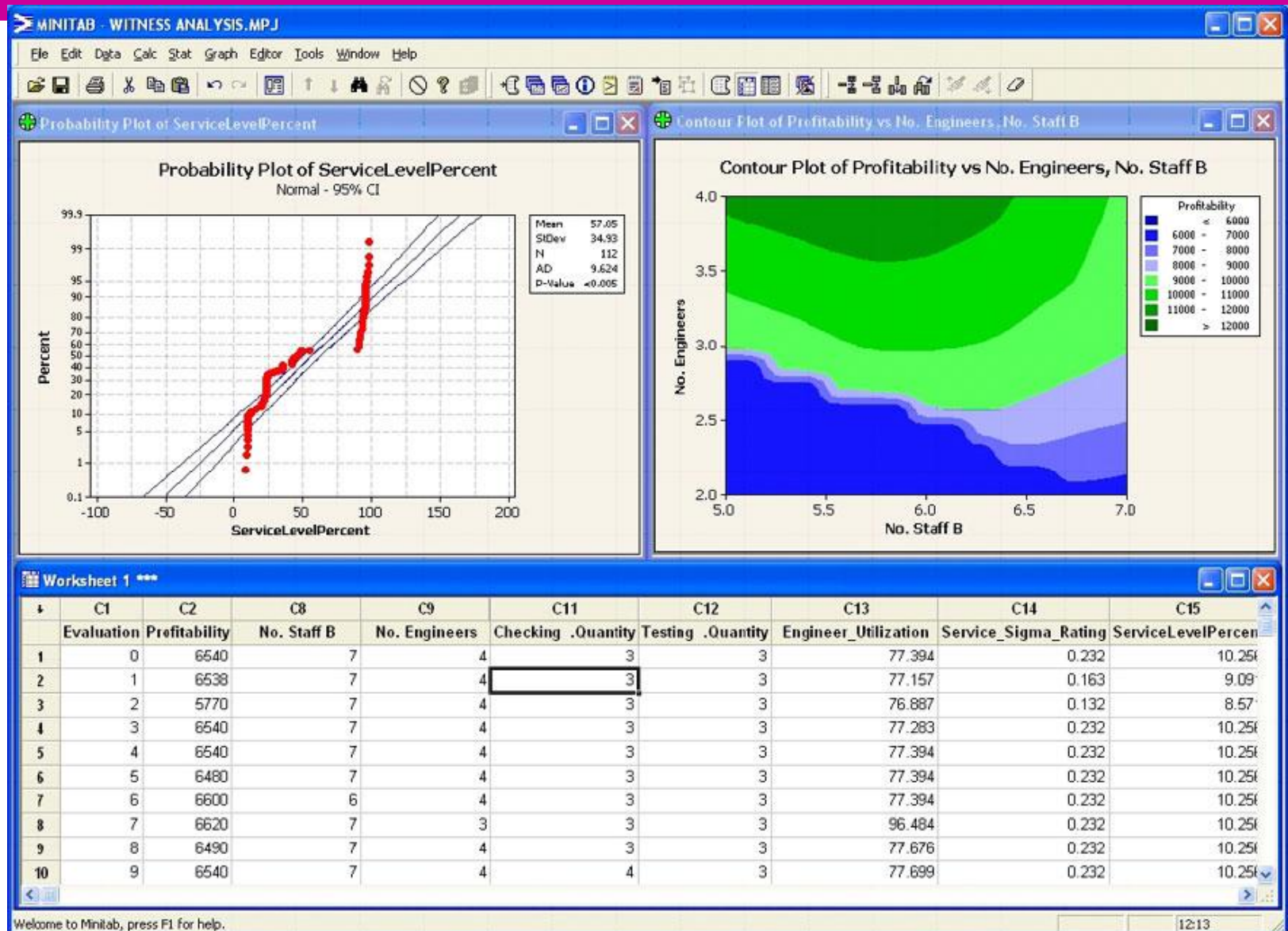
## **PRO**

- Nearly ubiquitous and is often pre-installed on new computers
- User friendly
- Very good for basic descriptive statistics, charts and plots

## **CON**

- Costs money
- Not sufficient for anything beyond the most basic statistical analysis

# Minitab



# Minitab

## **COST**

- \$1,395.00 per single user license

## **CON**

- *Costs Money*
- Not suitable for very complicated statistical computation and analysis
- Not often used in academic research

## **PRO**

- Easy to learn and use
- Often taught in schools in introductory statistics courses
- Widely used in engineering for process improvement

# SPSS

\*stroke\_survival.sav [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 42 of 42 Variables

	gender	active	obesity	diabetes	bp
1	Female	Yes	No	No	Hypotension
2	Male	Yes	Yes	No	Hypertension
3	Female	Yes	Yes	Yes	Hypertension
4	Male	Yes	No	No	Normal
5			No	No	Hypertension
6			Yes	No	Normal
7			No	Yes	Normal
8			No	No	Normal
9			No	No	Normal
10			No	No	Hypertension
11			Yes	No	Normal
12			No	No	Normal
13			No	No	Hypotension
14			Yes	No	Normal
15			Yes	Yes	Normal
16			Yes	No	Hypertension
17			No	No	Hypertension
18	Male	Yes	No	No	Hypertension
19	Female	No	No	No	Normal
20	Male	Yes	No	No	Hypotension
21	Female	Yes	Yes	No	Normal

Regression

- Automatic Linear Modeling...
- Linear...**
- Curve Estimation...
- Partial Least Squares...
- Binary Logistic...
- Multinomial Logistic...
- Ordinal...
- Probit...
- Nonlinear...
- Weight Estimation...
- 2-Stage Least Squares...
- Optimal Scaling (CATREG)...

Data View Variable View

Linear... IBM SPSS Statistics Processor is ready

# SPSS

## **COST**

- From \$1000 to \$12000 per license depending on license type.

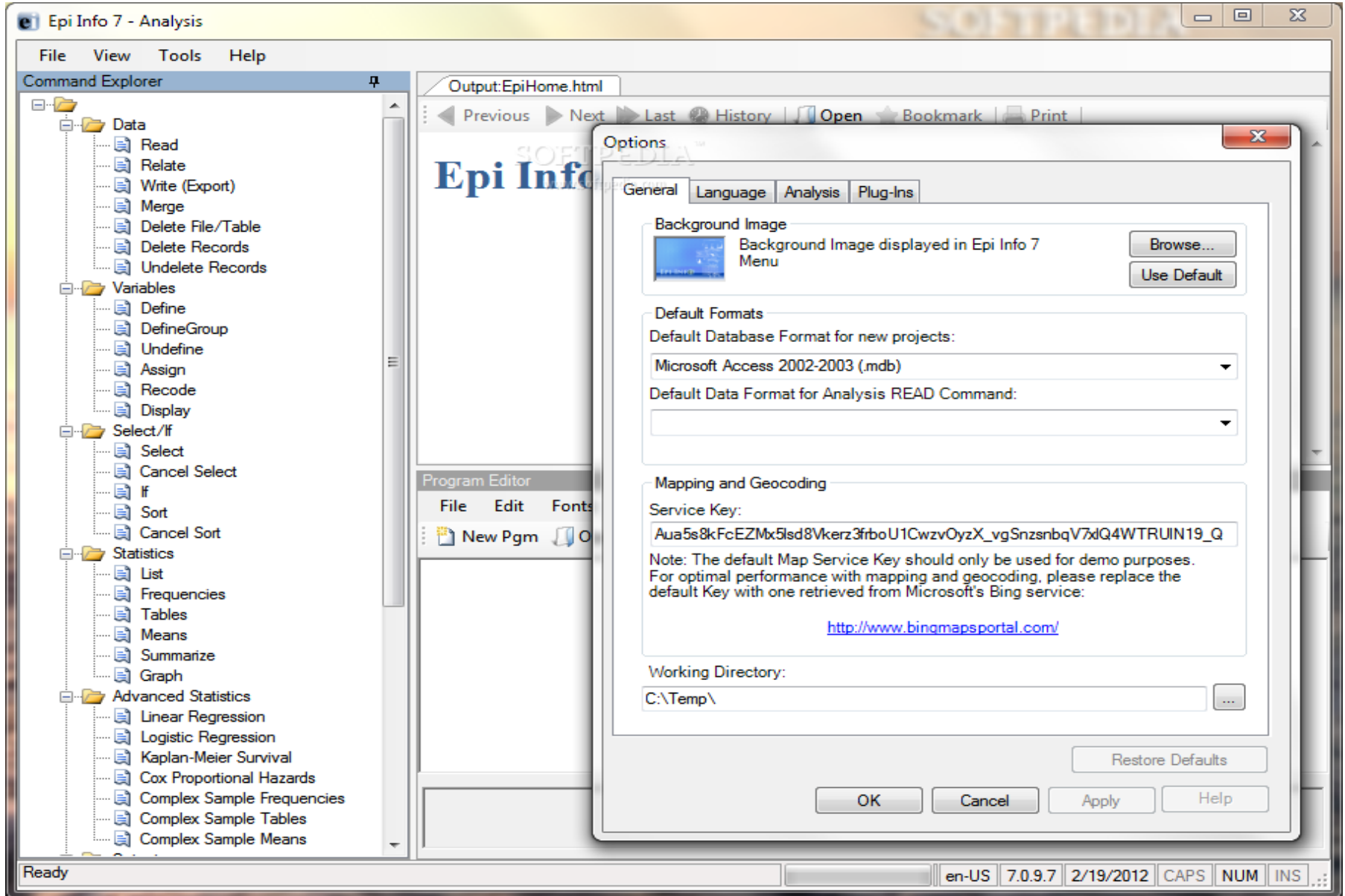
## **CON**

- Very expensive
- Not adequate for modeling and cutting edge statistical analysis

## **PRO**

- Easy to learn and use
- More powerful than Minitab
- One of the most widely used statistical packages in academia and industry
- Has a command line interface in addition to menu driven user interface
- One of the most powerful statistical packages that is also easy to use.

# EpiInfo





# EpiInfo

## PRO

- Consists of multiple modules to accomplish various tasks beyond just statistical analysis.
- ability to rapidly develop a questionnaire
- customize the data entry process
- quickly enter data into that questionnaire
- analyze the data

## COST

- Free

## CON

- Not a dedicated statistical package
- Not as powerful as commercial alternative for performing advanced analysis and modeling

# SAS

SAS

File Edit View Go Tools Solutions Window Help

Log - (Untitled)

```

2308 goptions reset=all;
2309 goptions hsize=5 in vsize=4 in ;
2310 ods html file="fig4_short.html" nogtitle nogfootnote opt
2310! ;
2311 goptions noimageprint;
2312 title "2008 Year to Date Weekly Report";
2313 proc tabulate data=yr2008 noseps ;
2314   var volnew high low close;
2315   table date =', (high='Weekly High' low='Weekly Low'
2315! volnew='Volume(100,000)')
2316     * mean=' ' * f=comma15. / rts=15;
2317   class date;
2318 run;
2319 title;
2320 proc gchart data=work.sectors;
2321   pie Sector / sumvar=Percentage descending detail=Issuer de
2322     value=none other=5 otherlabel='Combined'
2323     noheading legend html=htmlvar name='figure_
2324 run;
2324!   quit;
2325 ods html close;
  
```

Results Viewer - file://C:\SAS\temp\fig4\_short.html

### 2008 Year to Date Weekly Report

	Weekly High	Weekly Low	Weekly Close	Volume(100,000)
04JAN08	13,365	12,789	12,800	10,789
11JAN08	12,931	12,502	12,606	15,895
18JAN08	12,795	12,022	12,099	20,082
25JAN08	12,487	11,635	12,207	18,246

Editor - Untitled1 \*

```

goptions reset=all;
goptions hsize=5 in vsize=4 in ;
ods html file="fig4_short.html" nogtitle nogfootnote optio
goptions noimageprint;
title "2008 Year to Date Weekly Report";
proc tabulate data=yr2008 noseps ;
  var volnew high low close;
  table date =', (high='Weekly High' low='Weekly Low' cl
    * mean=' ' * f=comma15. / rts=15;
  class date;
run;
title;
proc gchart data=work.sectors;
  pie Sector / sumvar=Percentage descending detail=Issuer de
    value=none other=5 otherlabel='Combined'
  
```

Sector

- Consumer Discretionary
- Energy
- Health Care
- Information Technology
- Consumer Staples
- Financials
- Industrials
- Materials
- Combined

Output - (Untitle... | Log - (Untitled) | Editor - Untitled... | Explorer | Results Viewe...

# SAS

## **COST**

- Complicated pricing model
- \$8,500 first year license fee

## **CON**

- Very very expensive
- Not user friendly
- Steep learning curve
- Relatively poor graphics capabilities

## **PRO**

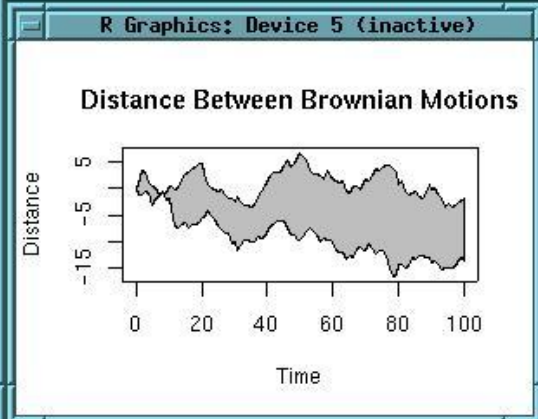
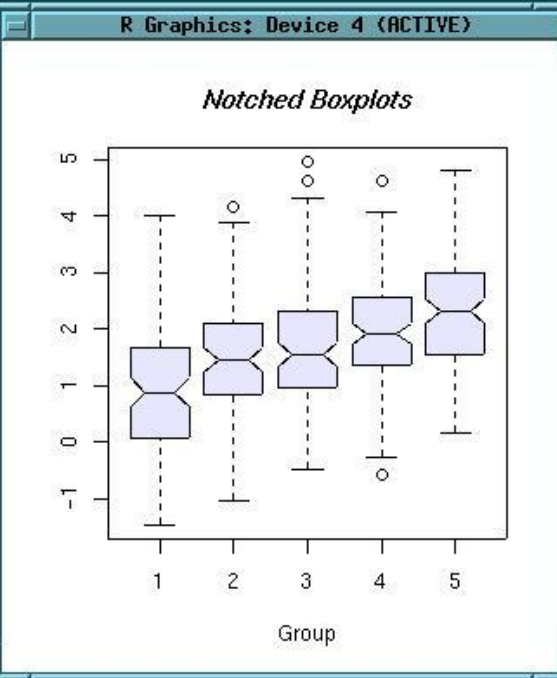
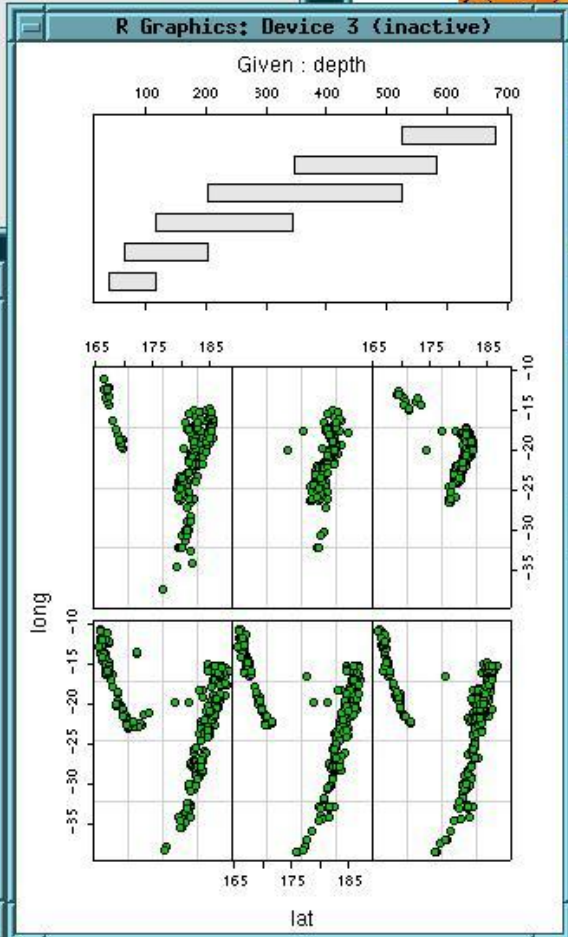
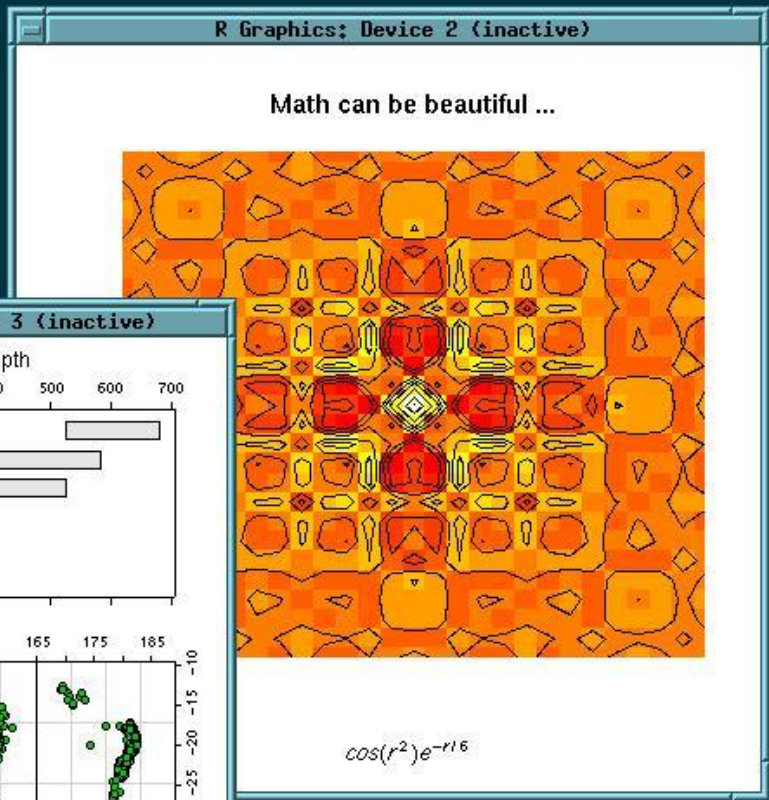
- Widely accepted as the leader in statistical analysis and modeling
- Widely used in the industry and academia
- Very flexible and very powerful.

```
leisch@galadriel:~/work/tnp
R> n <- 5
R> g <- gl(n, 100, n*100)
R> x <- rnorm(n*100) + sqrt(codes(g))
R> boxplot(split(x,g), col="lavender", notch=TRUE)
R> title(main="Notched Boxplots", xlab="Group", font.main=4, font.lab=1)
R>
R> ctl <- c(4.17,5.58,5.18,6.11,4.50,4.61,5.17,4.53,5.33,5.14)
R> trt <- c(4.81,4.17,4.41,3.59,5.87,3.83,6.03,4.89,4.32,4.69)
R> group <- gl(2,10,20,labels=c("Ctl","Trt"))
R> weight <- c(ctl,trt)
R> anova(lm.D9 <- lm(weight~group))

Analysis of Variance Table
Response: weight

  group    Df Sum Sq Mean Sq    F Pr(>F)
  Residual 18  8.7293   0.4850

R>
R>
```



# R

## PRO

- Widely used and accepted in industry and academia
- Very powerful and flexible
- Very large user base
- Lots of books and manuals
- Several User Interface Shells available

## COST

- Free / Open Source

## CON

- Not user friendly
- Requires steep learning curve

# R

- R is free but more difficult for those who are not into the world of math and pure statistics.
- There is R Commander package that can ease your comprehension
- SAS is more visual than R that makes its use simpler for those who are not familiar with programming languages.

# STATA is one thing which has united us today



# STATA

---

- It is a general-purpose statistical software package used by people of different backgrounds and professional disciplines
- Most Stata users are non-physicians



# STATA

---

- Although Stata has simple commands, these may be difficult for non-programmers to use.
- Generally, physicians are familiar with "clicking on" rather than writing commands

# STATA

- 
- No matter which book you choose or which course you attend....
  - Since Stata is used by people in many fields, most training programs offered are geared toward programmers and non-physicians

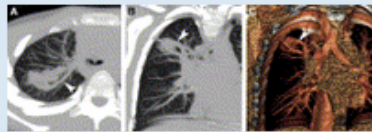
# Use of STATA in Pediatric research



**12 -24%**

# Pediatric Journals ( US vs Indian)


### Featured Image



Chest MSCT in patient 2 showing a large PAVM in the posterior segment of right upper lobe (*arrowhead*). A, Axial maximum intensity projection image. B, Coronal maximum intensity projection image. C, 3-dimensional volume rendering.

[Click here to read more](#)

### Current Issue | July 2013, Vol. 183, No. 1

 Full Table of Contents


### Featured Articles

Thanks to our Valued Reviewers—2012

Exposure to Bath Salts and Synthetic Tetrahydrocannabinol from 2009 to 2012 in the United States Wood KE

Epidemiology of Kawasaki Disease: Prevalence from National Database and Future Trends Projection by System Dynamics Modeling Huang S-K, Lin M-T, Chen H-C, et al

Growth and Body Composition of Uninfected Children Exposed to Human Immunodeficiency Virus: Comparison with a Contemporary Cohort and United States National Standards Neri D, Somarriba GA, Schaefer NN, et al

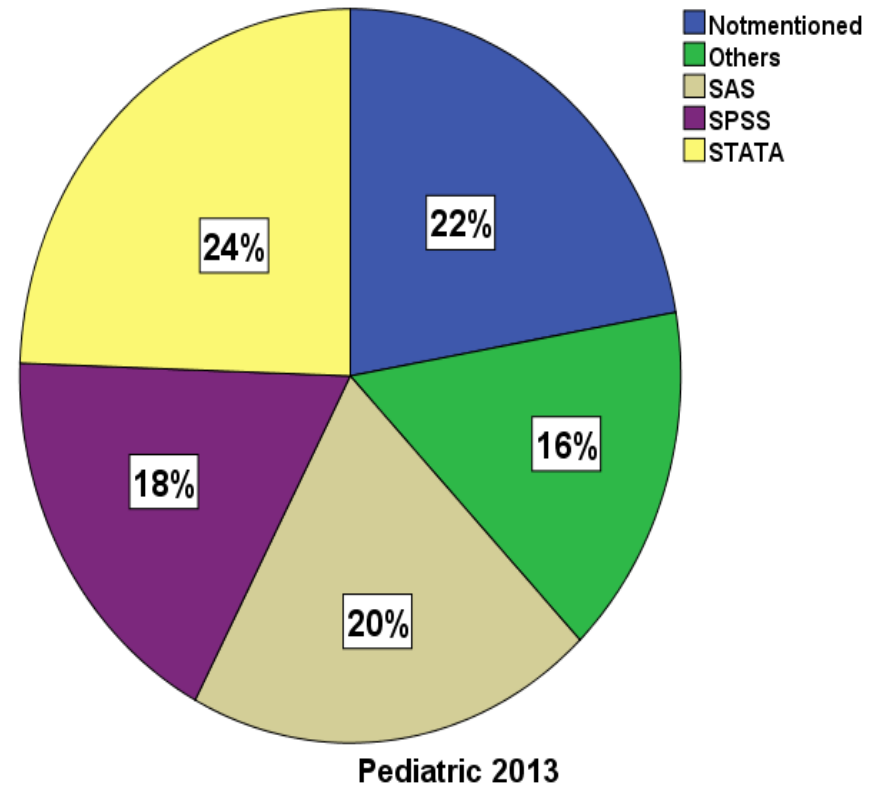
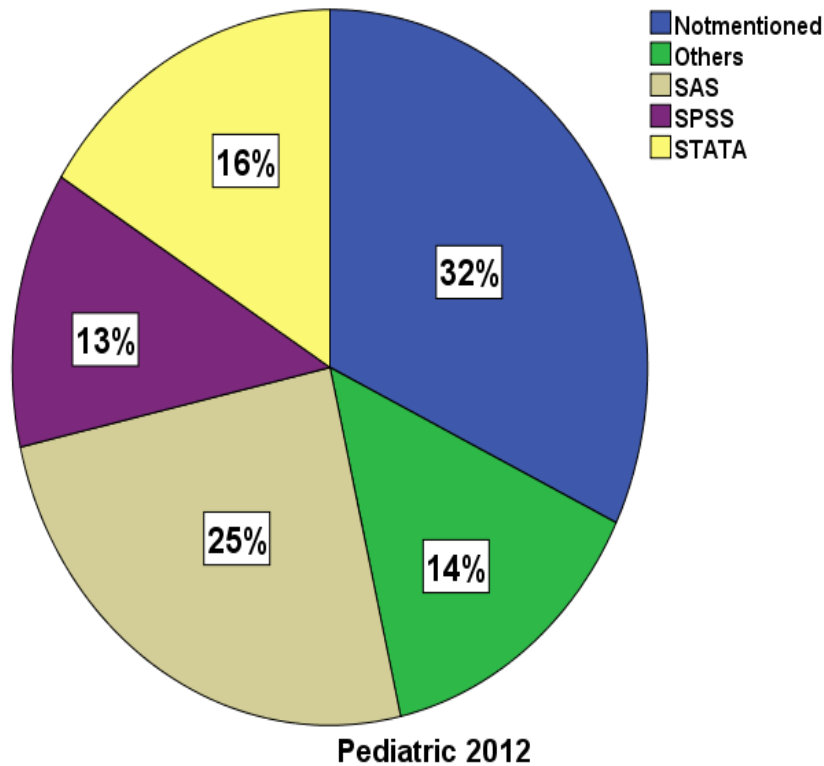
Access this journal on [ScienceDirect](#) 

 [Print or Share This Page](#)

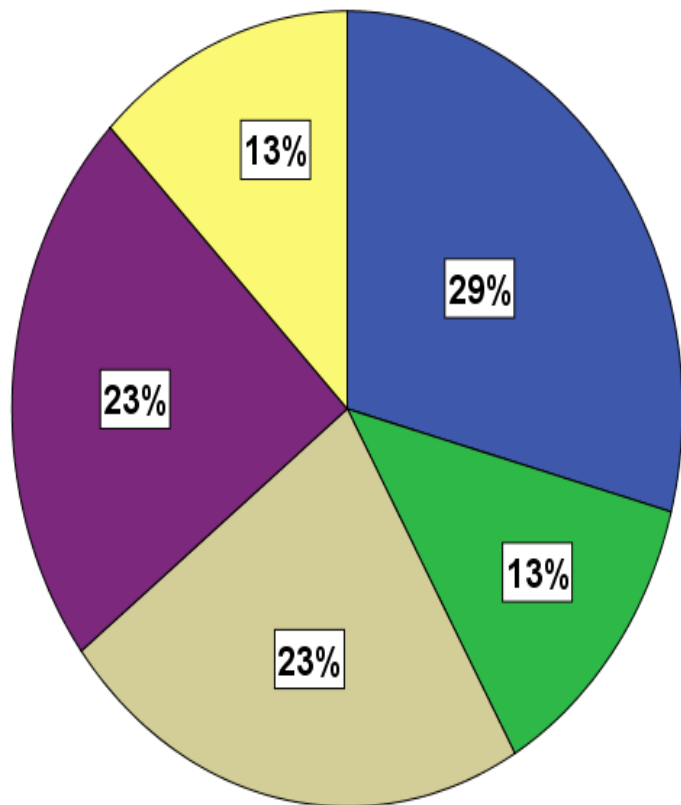
### Insights and Images Quiz



# Use of STATA and other soft wares in Pediatrics research

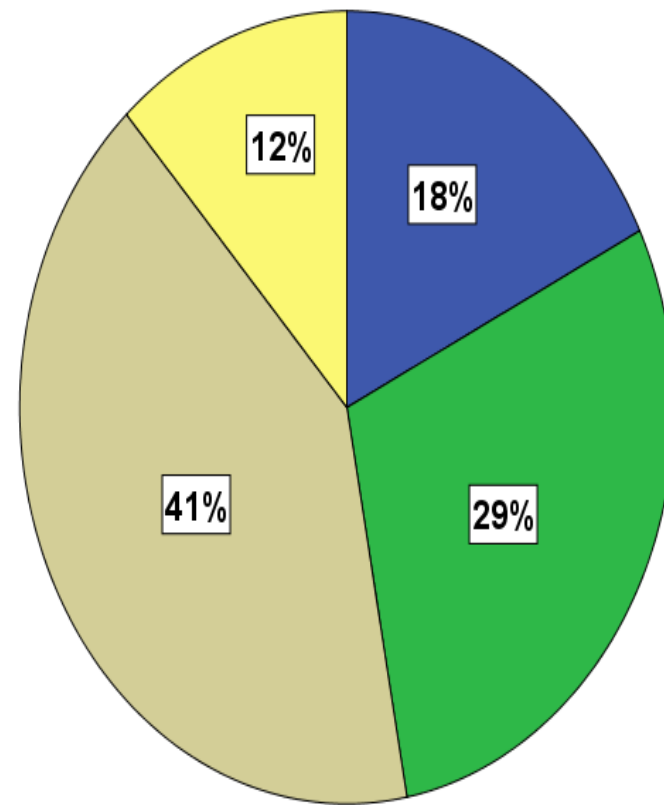


# Use of STATA and other softwares in Pediatric research



J of Pediatrics 2013

Notmentioned  
Others  
SAS  
SPSS  
STATA



Ind Pediatrics 2013

Notmentioned  
Others  
SPSS  
STATA

# SPSS -13700 hits



SPSS in Pediatric research



Scholar

About 13,700 results (0.05 sec)

My Citations

Any time

Since 2013

Since 2012

Since 2009

Custom range...

[Attitudes of canadian researchers toward the return to participants of incidental and targeted genomic findings obtained in a \*\*pediatric research\*\* setting](#)

CV Fernandez, C Strahlendorf, D Avar... - *Genetics in ...*, 2013 - nature.com

... Data were analyzed using **SPSS** version 14 (IBM, Somers, NY). ... CPCGC, Canadian **Pediatric** Cancer Genome Consortium; FORGE, Finding of Rare Genes Canada Consortium. ... results that should be offered to participants as indicated by Canadian **Pediatrics** Cancer Genome ...

[Cited by 2](#) [Related articles](#) [All 4 versions](#) [Cite](#)

Sort by relevance

Sort by date

[\[HTML\] Effect of child health status on parents' allowing children to participate in \*\*pediatric research\*\*](#)

[biomedcentral.com \[HTML\]](#)

J Vanhelst, L Hardy, D Bert, S Duhem... - *BMC medical ...*, 2013 - biomedcentral.com

... using the Statistical Package for the Social Sciences, version 11.5 for Windows (**SPSS** Inc., Chicago ... time in discussions with investigators before enrolling their child in a **pediatric** clinical **research** ... Ramet J, Van den Anker J: The European Academy of **pediatrics** (EAP/CESP) and ...

[Related articles](#) [All 13 versions](#) [Cite](#) [More](#)

include patents

include citations

# STATA-3780 hits



STATA in Pediatric research



Scholar

About 3,780 results (0.03 sec)

My Citations



Any time

Since 2013

Since 2012

Since 2009

Custom range...

[Medical complexity and \*\*pediatric\*\* emergency department and inpatient utilization](#)

[pediatricsdigest.mobi \[HTML\]](#)

L O'Mahony, DS O'Mahony, TD Simon, J Neff... - **Pediatrics**, 2013 - Am Acad **Pediatrics**

... Data analysis was completed by using **Stata** 10SE (**Stata** Corp, College Station, TX) and R statistical software programs (Vienna ... the National Institute of Neurologic Disorders and Stroke, the Child Health Corporation of America via the **Pediatric Research** in Inpatient ... **Pediatrics** . ...

[Cited by 1](#) [Related articles](#) [All 6 versions](#) [Cite](#)

Sort by relevance

Sort by date

[Umbilical choline and related methylamines betaine and dimethylglycine in relation to birth weight](#)

M Hogeveen, M den Heijer, BA Semmekrot... - **Pediatric** ..., 2013 - nature.com

... Journal name: **Pediatric Research** (2013) Volume: 73, Pages: 783-787 DOI: doi:10.1038/pr.2013.54 ... retardation, placental dysfunction or low birth weight should be involved in such **research**. ... LBW was defined as standardized birth weight  $\leq 2,500$  g. **STATA** for Windows version ...

[Related articles](#) [All 4 versions](#) [Cite](#)

include patents

include citations

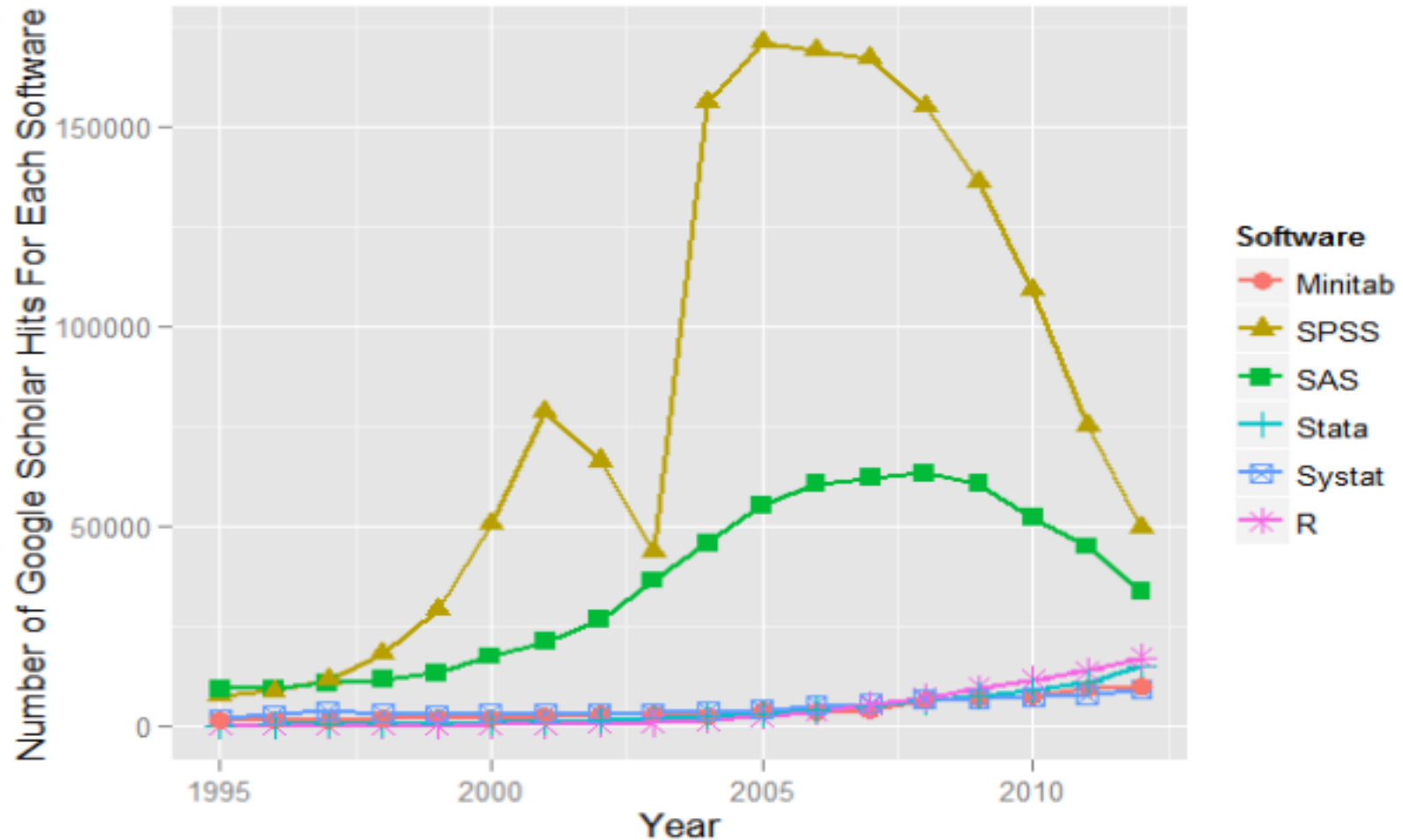
[Racial Differences in Antibiotic Prescribing by Primary Care Pediatricians](#)

[pediatricsdigest.mobi \[HTML\]](#)

JS Gerber, PA Prasad, AR Localio, AG Fiks... - ..., 2013 - pediatricsdigest.mobi



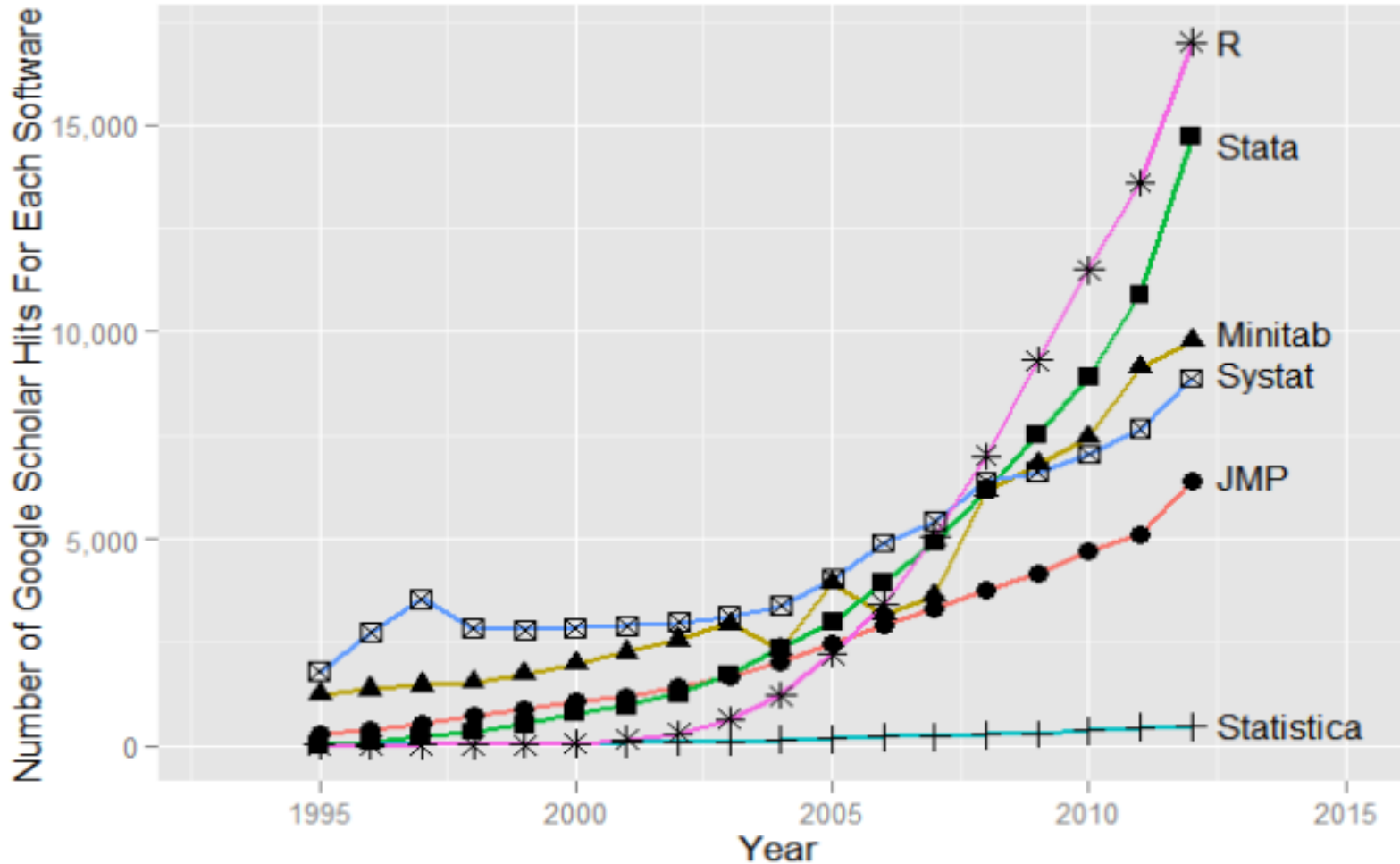
# No of Google scholar hits for various softwares



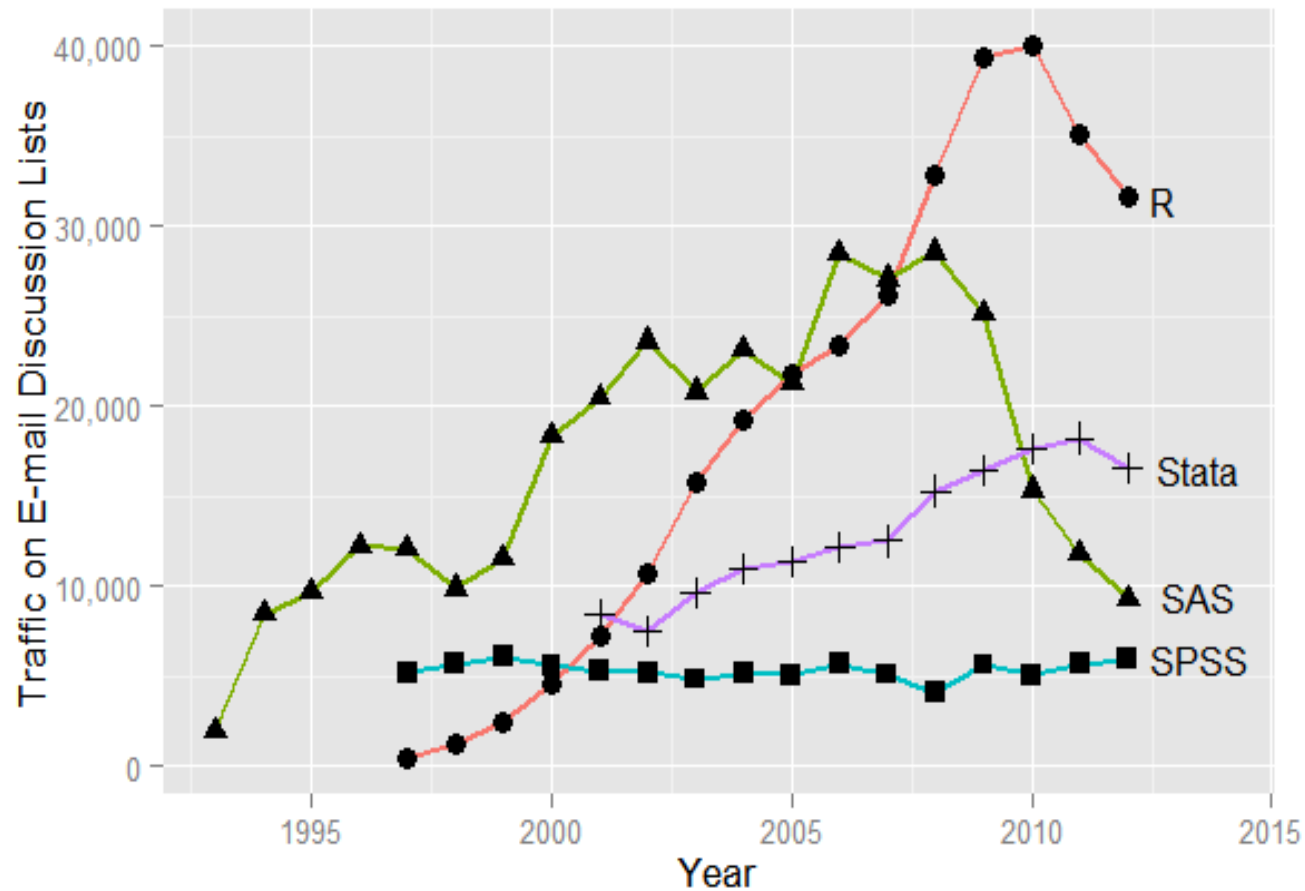
# Popularity of analytic softwares

	<b>R</b>	<b>SAS</b>	<b>SPSS</b>	<b>Stata</b>
1995	7	9120	7310	24
1996	4	9130	8560	92
1997	9	10600	11400	214
1998	16	11400	17900	333
1999	25	13100	29000	512
2000	51	17300	50500	785
2001	155	20900	78300	969
2002	286	26400	66200	1260
2003	639	36300	43500	1720
2004	1220	45700	56000	2350
2005	2210	55100	71000	2980
2006	3420	60400	69000	3940
2007	5070	61900	167000	4900
2008	7000	63100	155000	6150
2009	9320	60400	136000	7530
2010	11500	52000	109000	8890
2011	13600	44800	74900	10900
2012	17000	33500	49400	14700

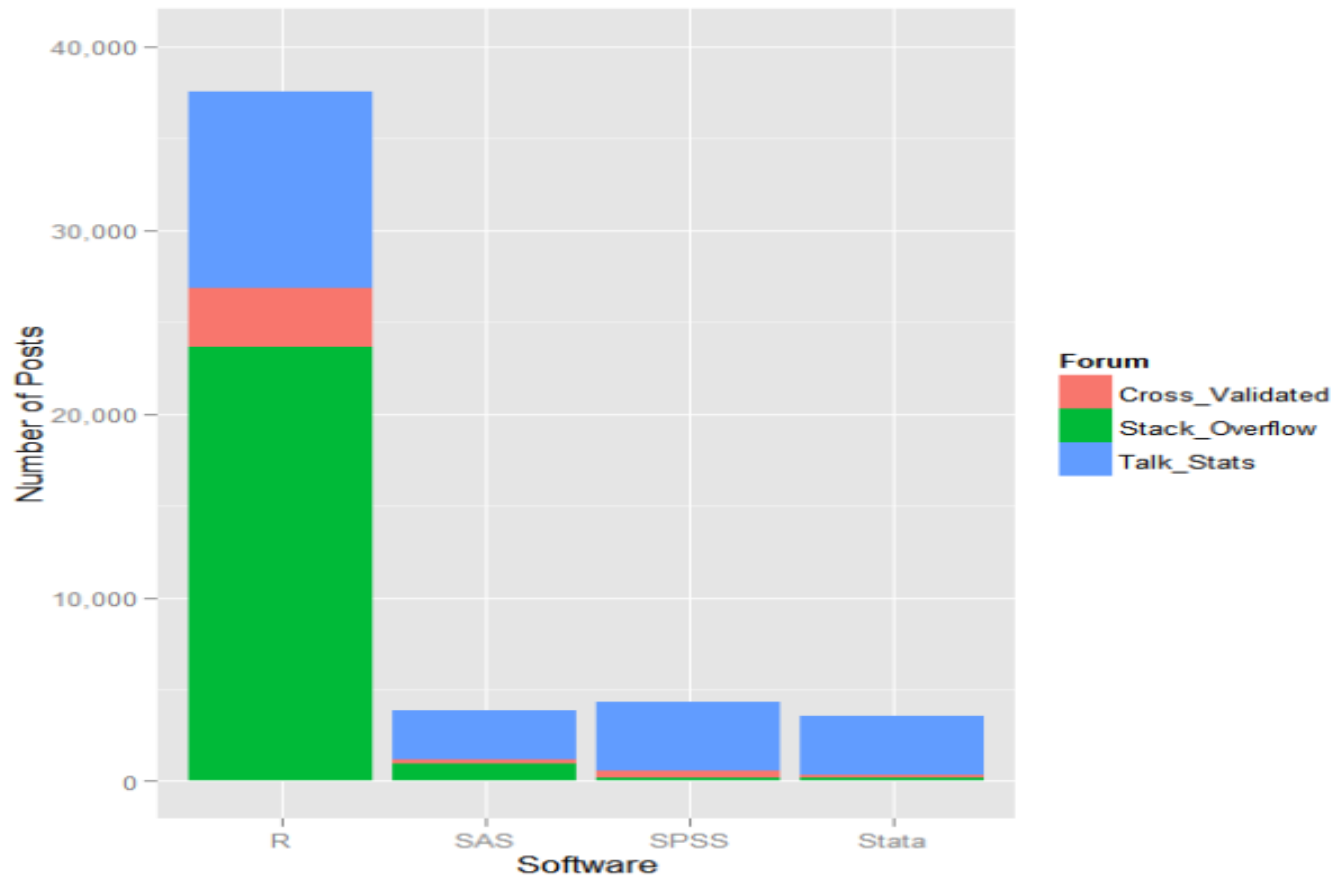
# Forecasting-Muenchen



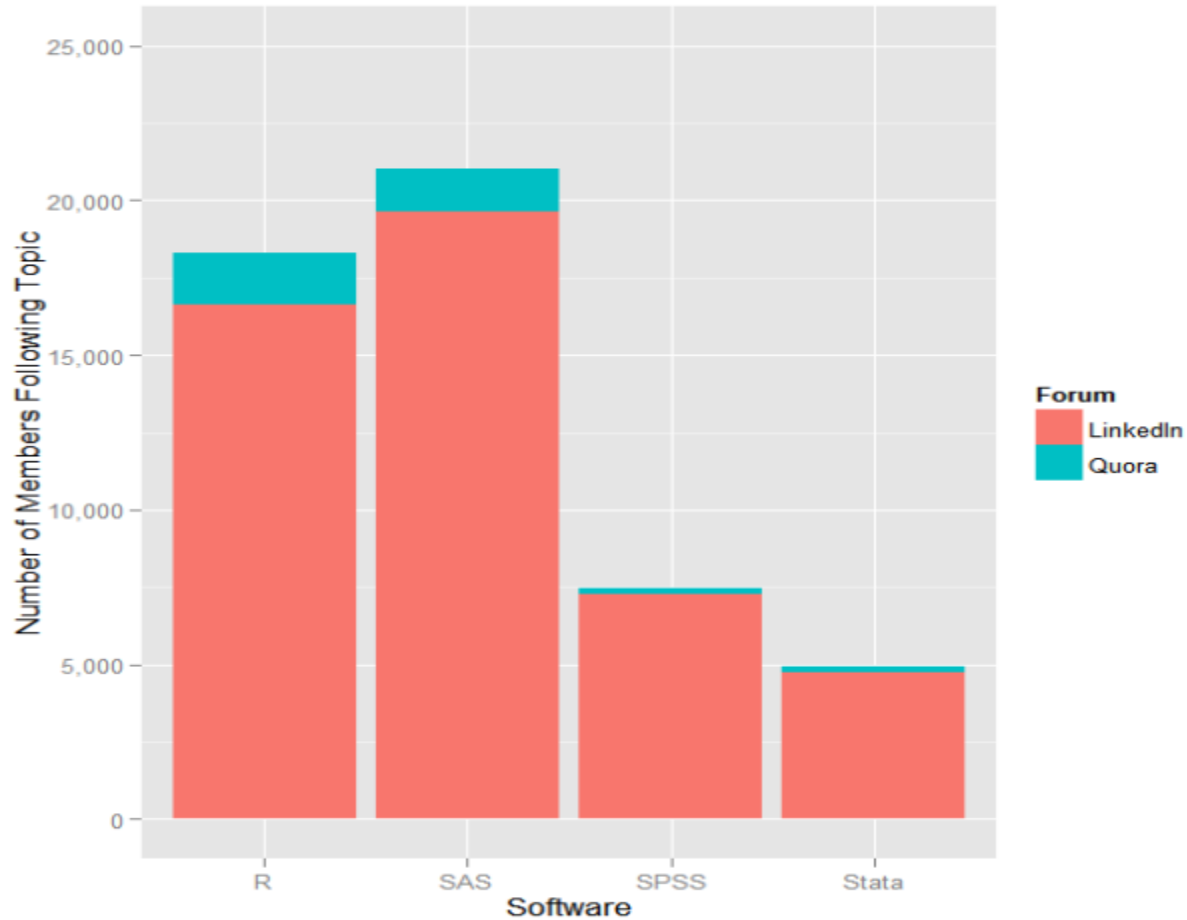
# Traffic on Email discussion lists



# Number of posts



# Number of people registered in the main discussion group for each software-2013

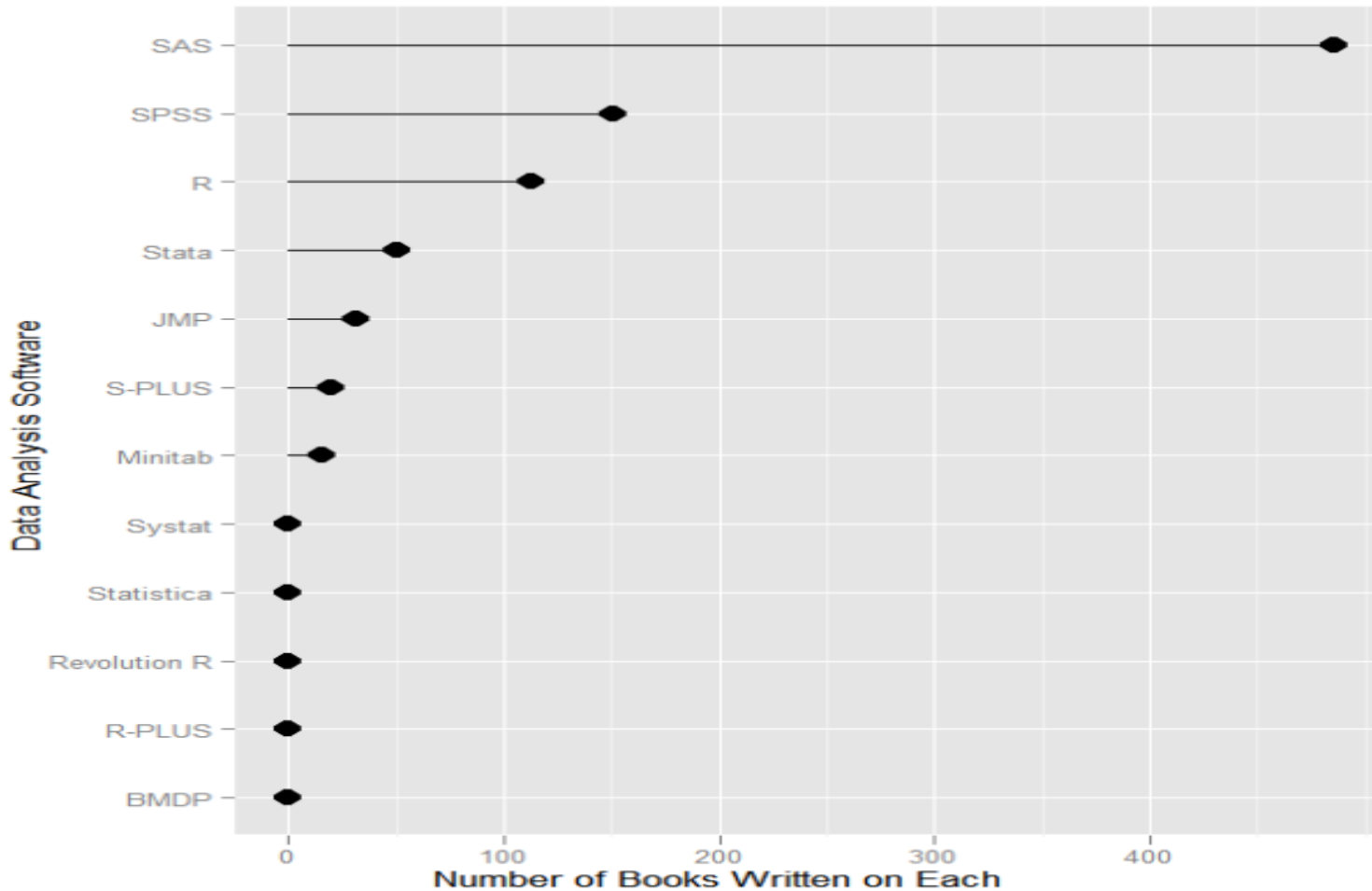


# Comparison using no of Blogs

---

Software	Number of Blogs
R	452
SAS	40
Stata	8
Others	0-3

# Comparison no of books written





# Health services research

Dembe et al. *BMC Health Services Research* 2011, **11**:252  
<http://www.biomedcentral.com/1472-6963/11/252>



## RESEARCH ARTICLE

## Open Access

# Statistical software applications used in health services research: analysis of published studies in the U.S

Allard E Dembe<sup>1\*</sup>, Jamie S Partridge<sup>2</sup> and Laurel C Geist<sup>3</sup>

### Abstract

**Background:** This study aims to identify the statistical software applications most commonly employed for data analysis in health services research (HSR) studies in the U.S. The study also examines the extent to which information describing the specific analytical software utilized is provided in published articles reporting on HSR studies.

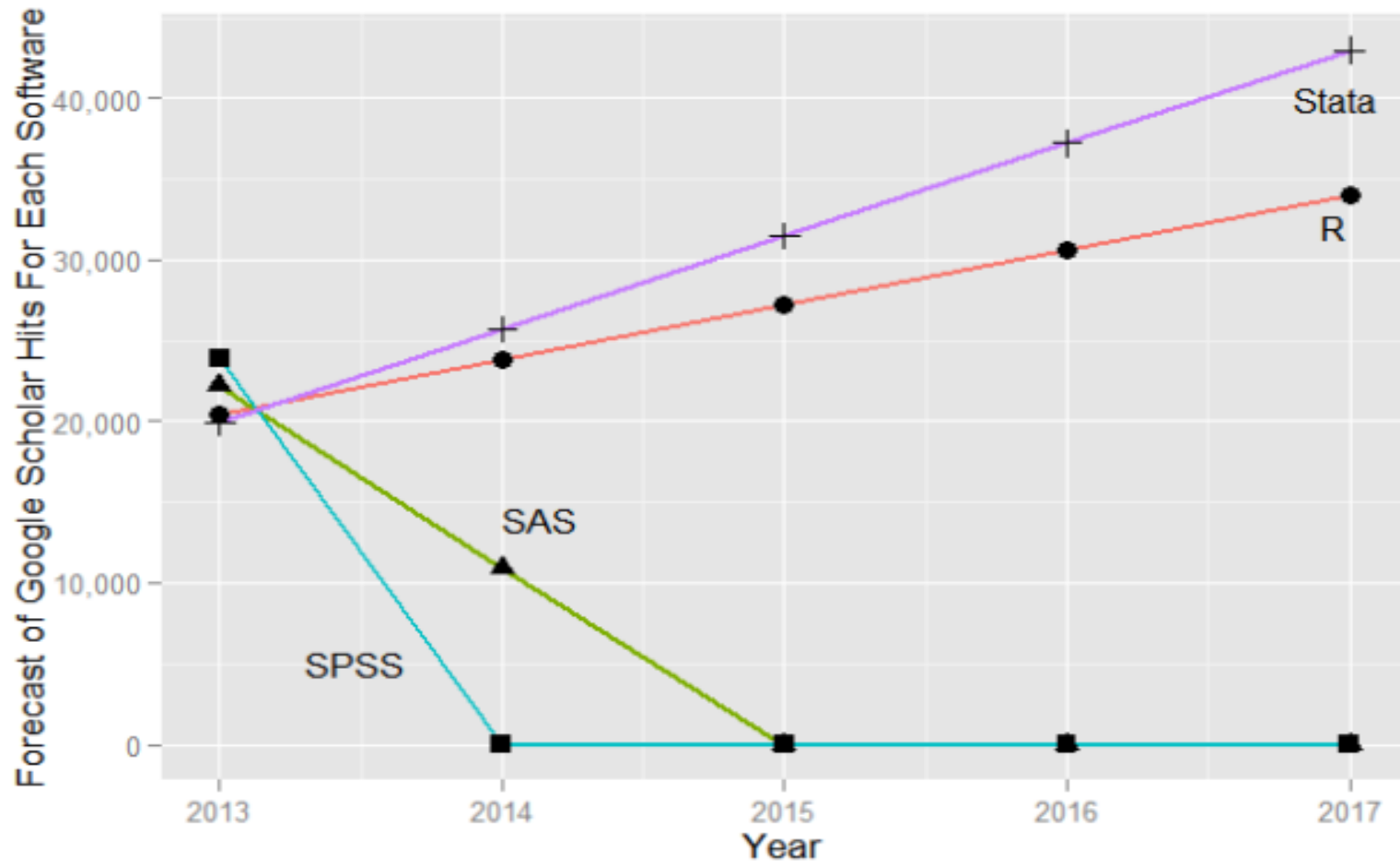
**Methods:** Data were extracted from a sample of 1,139 articles (including 877 original research articles) published between 2007 and 2009 in three U.S. HSR journals, that were considered to be representative of the field based upon a set of selection criteria. Descriptive analyses were conducted to categorize patterns in statistical software usage in those articles. The data were stratified by calendar year to detect trends in software use over time.

**Results:** Only 61.0% of original research articles in prominent U.S. HSR journals identified the particular type of statistical software application used for data analysis. Stata and SAS were overwhelmingly the most commonly used software applications employed (in 46.0% and 42.6% of articles respectively). However, SAS use grew considerably during the study period compared to other applications. Stratification of the data revealed that the type of statistical software used varied considerably by whether authors were from the U.S. or from other countries.

**Conclusions:** The findings highlight a need for HSR investigators to identify more consistently the specific analytical software used in their studies. Knowing that information can be important, because different software packages might produce varying results, owing to differences in the software's underlying estimation methods.

**Keywords:** Statistical software, data analysis, SAS, Stata

# Forecast



# Growth of R-Why?

- The continued rapid growth in add-on packages
- The attraction of R's powerful language
- The near monopoly R has on the latest analytic methods
- Its free price
- The freedom to teach with real-world examples from outside organizations, which is forbidden to academics by SAS and SPSS licenses (IBM is loosening up on this a bit)

# What will slow R's growth

- is its lack of a graphical user interface that:
  - Is powerful
  - Is easy to use
  - Provides direct cut/paste access to journal style output in word processor format
  - Is standard, i.e. widely accepted as *The One to Use*
  - Is open source

# STATA-Forecast

- Although Stata is currently the fastest growing package,
- it's growth will slow in 2013 and level off by 2015 at around 23,000 articles, leaving it in fourth place.

# STATA-forecast

- The main cause of this will be inertia of users of the established leaders, SPSS and SAS, as well as the competition from all the other packages, most notably R.
- R and Stata share many strengths and with R being free, there is doubt Stata will be able to beat R in the long run

# Why the difference?

- Learning to use a data analysis tool well takes significant effort, so people tend to continue using the tool they learned in college for much of their careers.
- As a result, the software used by professors and their students is likely to predict what the next generation of analysts will use for years to come

# How to decide?

- Does it run natively on your computer?
- Does the software provide all the methods you use? If not, how extensible is it?
- Does that extensibility use its own language, or an external one (e.g. Python, R, SQL) that is commonly accessible from many packages?
- Does it fully support the style (programming vs. point-and-click) that you like?



# How to decide?

- Are its visualization options (e.g. static vs. interactive) adequate for your problems?
- Does it provide output the form you prefer (e.g. cut & paste vs. LaTeX integration)?
- Does it handle large enough data sets?
- Do your colleagues use it so you can easily share data and programs?
- Can you afford it?

# How to decide doctors?

- Keep it simple
- Utilize pull-down menus (rather than commands)
  - No discussion on commands
  - Provision of cut and paste
- Scientific methods to disseminate information

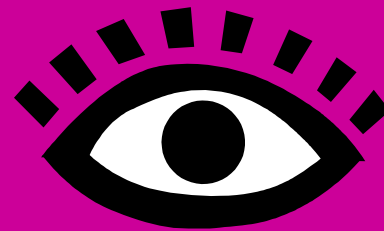


There are five primary learning styles:

1 visual (picture),  
2 visual (text)

3 auditory  
4 verbal

5 kinesthetic



# Types of Learners

- Visual learners
- Auditory learners
- Kinesthetic learners



# Dale's Cone of Experience

PEOPLE GENERALLY REMEMBER:

PEOPLE ARE ABLE TO:

10% of what they read

Read

• Define

• Describe

20% of what they hear

Hear

• List

• Explain

30% of what they see

View Images

Watch Videos

• Demonstrate

50% of what they hear  
and see

Attend Exhibit/Sites

Watch a Demonstration

• Apply

• Practice

70% of what they  
say and write

Participate in Hands-on Workshop

Design Collaborative Lesson

• Analyze

• Design

• Create

• Evaluate

90% of what they  
say, discuss,  
and do

Simulate or Model Lesson or Experience

Design/Perform a Presentation – Do the “Real Thing”

THANK YOU

