



# INTEGRAÇÃO DA ESTATÍSTICA ESPACIAL EM AMBIENTES GIS

**Monografia de Qualificação em  
Computação Aplicada**

Aluna: Karla Donato Fook

INPE/2005



## Roteiro

- Introdução
- Estatística Espacial
- Estatística Espacial e GIS
- Integração R e TerraLib via aRT
- Considerações Finais





## Introdução

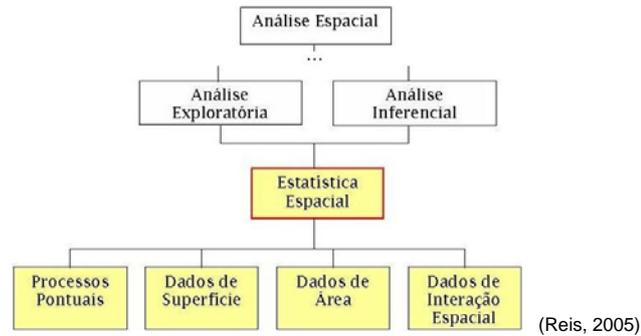
- Ferramentas de estatísticas espaciais não contemplam componentes providos por um GIS
  - Banco de dados espacial robusto
  - Modelos espaciais
  - Algoritmos de visualização
  
- Usuários de GIS encontram dificuldades para fazer análises mais sofisticadas
  
- Motivação
  - A integração das duas tecnologias une a capacidade de gerenciamento de dados dos GIS às técnicas de análise da Estatística Espacial



## Estatística Espacial



# Estatística Espacial

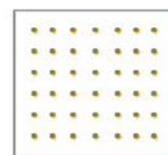
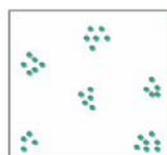
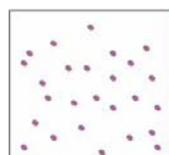


- Conjunto de técnicas utilizadas na análise exploratória e análise inferencial, onde a localização dos dados é relevante
- Inserida em um contexto mais amplo, a análise espacial de dados
- Subdividida em áreas de análise conforme o tipo de dado tratado



## Análise de Dados Pontuais ou Eventos

- Estuda a distribuição espacial de fenômenos expressos através de ocorrências identificadas como pontos localizados no espaço
- Detecção de Padrões Espaciais
- Exemplos
  - A localização de crimes
  - Ocorrência de doenças





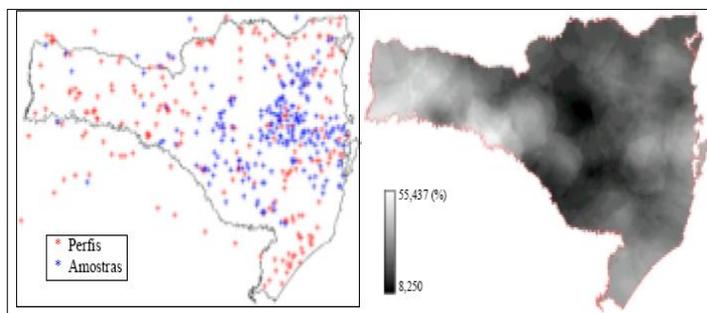
## Análise de Dados de Superfície

- Permite que dados pontuais sejam interpolados gerando uma superfície
- Detecção de padrões nos valores dos atributos
- Modelos utilizados
  - Determinísticos de efeitos locais
  - Determinísticos de efeitos globais
  - Modelos estatísticos de efeitos locais e globais
- Exemplos
  - Medidas geológicas de depósitos minerais
  - Salinidade e permeabilidade do solo
  - Medição de chuva



## Análise de Dados de Superfície

- a) Distribuição de perfis e amostras de solo em Santa Catarina
- b) Distribuição contínua estimada para a variável saturação por bases



(a)

(b)

(Druck et al., 2004)



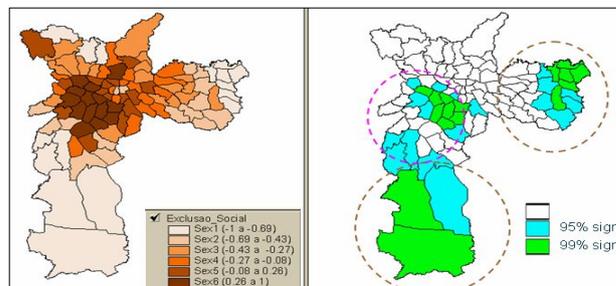
## Análise de Dados de Área

- Aplicada quando não se dispõe da localização exata dos eventos e sim um valor por área
- Detecção e possíveis explicações para o padrão nos valores dos atributos
- Utiliza
  - Análise exploratória
  - Modelos de regressão espacial
  - Modelos inferenciais



## Análise de Dados de Área

- Exemplos
  - Censos
  - Estatísticas de saúde
  - Distribuição espacial do índice de exclusão/inclusão social de S. Paulo
    - a) Tipo de dado de área
    - b) Agregamentos de exclusão e inclusão social



(a)

(b)

Sposati *apud* (Druck et al., 2004)



## Análise de Dados Espaço-Temporais

- Lida com modelos preditivos que possuem representação espaço-temporal
- Aplicada para avaliar padrões que variam no tempo
- Técnicas de análise capturam relações dinâmicas e estabelecem conexões entre fenômenos através de modelos multi-escala
- Desafio na Geoinformação



## Análise de Dados Espaço-Temporais

- Exemplos
  - Exame do impacto da dinâmica temporal no comportamento de um modelo urbano (Liu and Anderson, 2004);
  - Avaliação do padrão de procriação de pássaros (Unwin, 1996);
  - Visualização e análise exploratória de dados epidemiológicos (AvRuskin et al., 2004);
    - Estudo da relação entre exposição de arsênio e câncer de bexiga em Michigan
  - Descoberta de agrupamento espaço-temporal (Assunção et al., 2003);
    - Incidência de crime em Belo Horizonte, durante 1995-2001 coletados pela Polícia Militar de Minas Gerais



## Estatística Espacial e GIS

---



## Estatística Espacial e GIS

- Integração fortalece as áreas e aumenta a eficácia individual e comum de ambas
- A tecnologia GIS oferece um pronto acesso a rotinas complexas e sofisticadas da estatística espacial para uma grande comunidade de usuários
- Estratégias de integração
  - Completa (*full integration*)
  - Acoplamento fraco (*loose coupling*)
  - Acoplamento forte (*close coupling*)

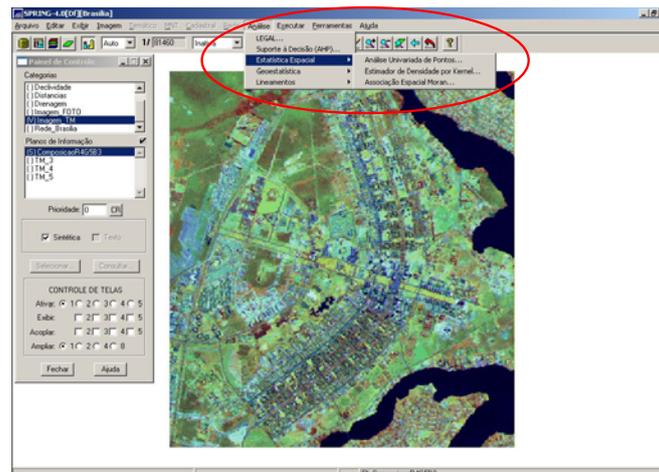


## Integração Completa

- As ferramentas de análise estatística são incorporadas diretamente no software GIS
- Utilizada em ferramentas como
  - Geoestatística
  - Índices de autocorrelação espacial
- Desvantagens
  - Aumento da complexidade do GIS para usuários comuns
  - Alto investimento de programação e manutenção de software
  - Duplicação da funcionalidade já disponível em outros sistemas



## Integração Completa



Geoestatística no SPRING



## Integração via Acoplamento

- Os softwares GIS e de análise estatística funcionam separadamente e compartilham os dados a serem analisados
- Permitem que o software de análise de dados espaciais evolua independentemente do software GIS
  - Incorpora técnicas que não são, a princípio, vantajosas de ser integradas completamente ao GIS do ponto de vista comercial

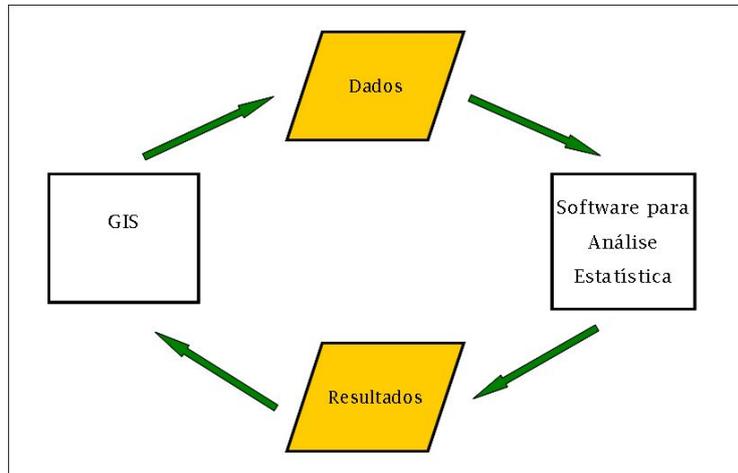


## Acoplamento Fraco

- Os dados são exportados do GIS para dentro de um software de análise estatística de dados espaciais
- Os resultados da análise são exportados para o GIS para que sejam visualizados
- Vantagem
  - Implementação computacionalmente simples
- Desvantagem
  - *Overhead* nos processos de exportação e importação dos dados e/ou resultados



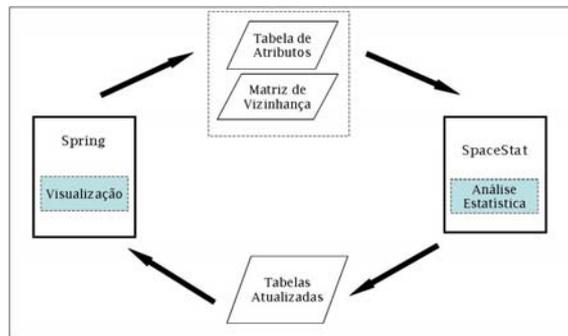
## Acoplamento Fraco



## Acoplamento Fraco

### ■ Exemplos

- ❑ GIS ArcInfo e o R
- ❑ GIS Spring e o SpaceStat



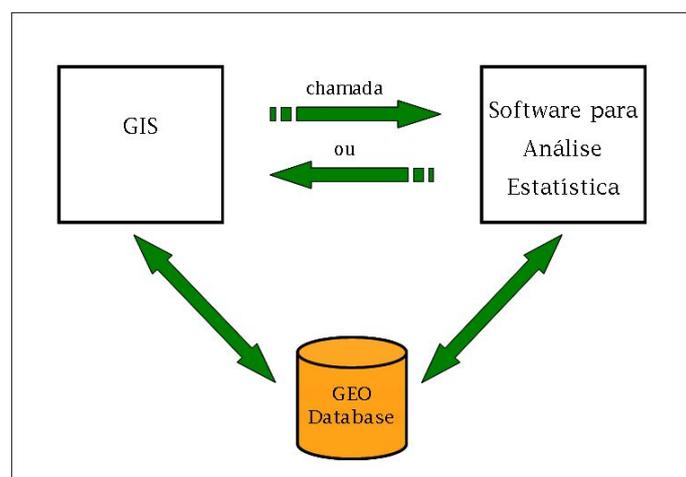


## Acoplamento Forte

- GIS e o software de análise compartilham dados
- Não há perda de estruturas mais altas
  - Topologia
  - Identidade do objeto
  - Metadados
  - Relacionamentos



## Acoplamento Forte





## Acoplamento Forte

### ■ Exemplos

- GIS ArcView (ESRI) e S-PLUS (MathSoft)
- GIS IDRISI e o software geoestatístico Gstat
- GIS GRASS e R
  - Executa o R de dentro do ambiente GRASS através do pacote "GRASS"
  - O R herda as configurações de todas as variáveis de ambiente necessárias aos programas GRASS
  - Suporta
    - Transferência binária direta
    - Transferência através de arquivos ASCII temporários



## Acoplamento Forte/TerraLib e Open Modeller

### ■ Open Modeller

- Um dos produtos do projeto speciesLink
- Ambiente computacional para o desenvolvimento de modelos e cenários
- Usado para predição de distribuição de espécies

### ■ Projeto TerraModeller

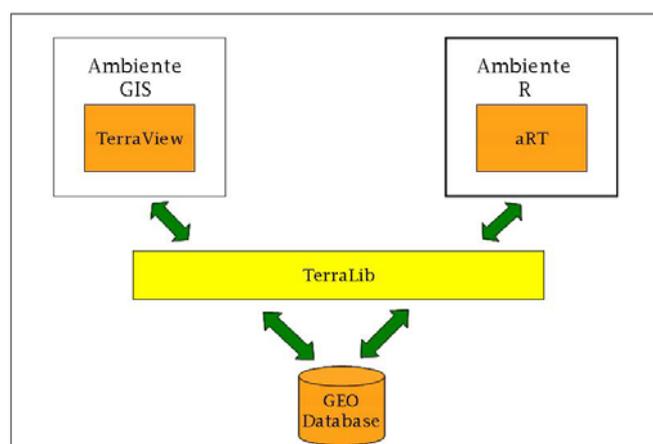
- Integrar o OpenModeller e GIS para assegurar o acesso a grandes volumes de dados ambientais
  - Dados: clima, vegetação, topografia, solos e uso da terra
- Combinar as funcionalidades da TerraLib com o ambiente OpenModeller



## Integração R e TerraLib via aRT



## Integração R e TerraLib



Mecanismo de *acoplamento forte*



## Integração R e TerraLib

- Provê ao R acesso a Banco de Dados TerraLib
- Os resultados gerados no R são incorporados ao Banco de Dados de forma direta e transparente
  - Análises são realizadas no R através de pacotes como *gstat*, *geoR* ou *R2WinBUGS*
- Os resultados são visualizados no TerraView
- A integração é realizada através da API aRT



## aRT: API R-TerraLib

- Pacote do R desenvolvido pela UFPR, inicialmente para a plataforma Linux
- É carregado no ambiente R
- Propicia ao usuário do R acesso
  - Bancos de Dados Geográficos
  - Operações implementadas na TerraLib



## aRT: API R-TerraLib

### ■ Classes para manipular os dados e funções da TerraLib

#### □ aRT

- Permite conexão com DBMS para a realização de funções de administração de banco de dados

#### □ aRTdb

- Provê a criação e o acesso a um novo banco de dados

#### □ aRTlayer

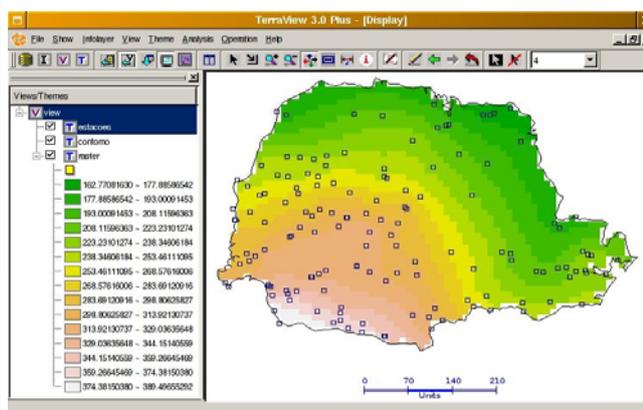
- Permite a manipulação de *layers* em aRT

#### □ aRTtheme

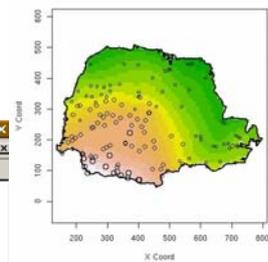
- Propicia a visualização de *themes* no TerraView



## Integração R e TerraLib



Visualização do Resultado no TerraView



Resultado R

(Neto et al., 2005)

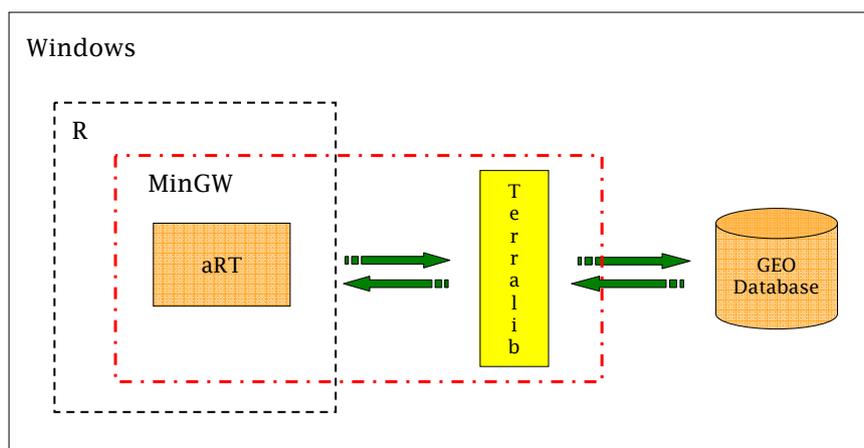


## Porte aRT para a Plataforma Windows

- Realizado de forma conjunta por equipes do INPE e da UFPR
- Migra a integração R e TerraLib, e conseqüentemente R e TerraView para o Windows
- O aRT foi compilado sob um ambiente que simula, de forma mínima, o ambiente operacional do Linux



## Porte aRT para a Plataforma Windows





## Considerações Finais

---



## Considerações Finais

- A integração da Estatística Espacial em ambientes GIS
  - Pode tornar tanto GIS quanto as estatísticas espaciais mais robustos
  - Torna a tomada de decisões mais efetiva
  - Unifica a capacidade de gerenciamento de dados dos GIS com as técnicas de análise da estatística espacial
  - Eleva o grau de complexidade das análises
  - Melhora a qualidade dos resultados



## Considerações Finais

### ■ Perspectivas

- Desafios emergentes como a abordagem espaço-tempo podem propiciar novas necessidades e alternativas de integração
- Surgimento de demanda para a integração entre outros tipos de softwares
- Forma de integração a ser selecionada pelo usuário de acordo com sua necessidade e requisitos

### ■ Entretanto ...

- O processo de interligação entre as tecnologias deve ser direcionado a cada classe de usuário de forma mais sistematizada



## Referências



## Referências

- ANSELIN, L., 1992, SpaceStat tutorial: a workbook for using SpaceStat in the analysis of spatial data, Santa Barbara, NCGIA (National Center for Geographic Information and Analysis).
- ANSELIN, L., 1998. Exploratory Spatial Data Analysis in a Geocomputational Environment. In: LONGLEY, P.; BROOKS, S.; MCDONNELL, R.; MACMILLAN, B., eds., **Geocomputation: A Primer**: New York, John Wiley & Sons.
- ANSELIN, L. **GeoDA**. Spatial Analysis Laboratory. University of Illinois, Urbana-Champaign, 2005. <https://geoda.uiuc.edu/default.php>.
- ASSUNÇÃO, R. Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia. São Carlos, SP: UFScar, 2001.
- ASSUNÇÃO, R. M.; TAVARES, A. I.; KULLDORFF, M. An early warning system for space-time cluster detection. In: V Workshop Brasileiro em Geoinformática - GeoInfo2003. Campos do Jordão - São Paulo, 2003.
- ANSELIN, L.; BAO, S., 1997. Exploratory spatial data analysis linking SpaceStat and ArcView. In: FISCHER, M.; GETIS, A., eds., **Recent developments in spatial analysis**: Berlin, Springer-Verlag, p. 35-59.
- ANSELIN, L.; SYABRI, I.; KHO, Y. **GeoDa: An Introduction to Spatial Data Analysis, Geographical Analysis (forthcoming)**. 2005, <https://geoda.uiuc.edu/documentation.php#pubs>.
- BAILEY, T.; GATRELL, A. **Interactive Spatial Data Analysis**. London: Longman Scientific and Technical, 1995.
- BAO, S.; ANSELIN, L.; MARTIN, D.; STRALBERG, D. Seamless integration of spatial statistics and GIS: The S-PLUS for ArcView and the S+Grassland Links. **Journal of Geographical Systems**, v. 2, n.3, 2000.
- BIVAND, R. Using the R statistical data analysis language on GRASS 5.0 GIS data base files. **Computers and Geosciences**, v. 26, p. 1043-1052, 2000.



## Referências

- BIVAND, R. R and geographical information systems, especially GRASS. In: 2nd International Workshop on Distributed Statistical Computing - DSC. Vienna, Austria, 2001. p.
- BIVAND, R., 2005, Interfacing GRASS 6 and R: Status and development directions, GRASS-News, Germany, p. 11-16.
- BIVAND, R.; NETELER, M. Open Source geocomputation: using the R data analysis language integrated with GRASS GIS and PostgreSQL data base systems. In: The Fifth International Conference on GeoComputation. Kent, UK, 2000. p.
- BURROUGH, P., 1998. Dynamic Modelling and Geocomputation. In: LONGLEY, P.; BROOKS, S.; MCDONNELL, R.; MACMILLAN, B., eds., **Geocomputation: A Primer**: New York, John Wiley.
- CÂMARA, G.; MONTEIRO, A. M.; DAVIS, C. **Introdução à Ciência da Geoinformação**. São José dos Campos: INPE, 2003a.
- CÂMARA, G.; MONTEIRO, A. M.; MENEZES, J. S. Representações Computacionais do Espaço: Fundamentos Epistemológicos da Ciência da Geoinformação. *Revista Geografia (UNESP)*, v. 28, n.1, p. 83-96, 2003b
- CÂMARA, G.; SOUZA, R.; PEDROSA, B.; VINHAS, L.; MONTEIRO, A. M.; PAIVA, J.; CARVALHO, M. T.; GATTASS, M. TerraLib: Technology in Support of GIS Innovation. In: II Brazilian Symposium on Geoinformatics, GeoInfo2000. São Paulo, 2000. p.
- CASANOVA, M.; CAMARA, G.; DAVIS, C.; VINHAS, L.; QUEIROZ, G., eds., 2005, Bancos de Dados Geográficos (Spatial Databases): Curitiba, Editora MundoGEO.
- CRIA; FAPESP. **openModeller: Static Spatial Distribution Modelling Tool**. CRIA, 2005. <http://openmodeller.cria.org.br/>.
- CRIA; INPE; USP, 2005, Project Proposal: A framework for species distribution modeling, INPE, p. 20.
- DINIZ, A., 2000, Estatística Espacial, UFMG - Universidade Federal de Minas Gerais, p. 15.



## Referências

- DRUCK, S.; CARVALHO, M. S.; CÂMARA, G.; MONTEIRO, A. M. V. **Análise Espacial de Dados Geográficos**. Brasília: EMBRAPA (ISBN 85-7383-260-6), 2004.
- FISCHER, M. M.; SCHOLTEN, H. J.; UNWIN, D., 1996. Geographic Information Systems, spatial data analysis and spatial modelling: an introduction. In: FISCHER, M. M.; SCHOLTEN, H. J.; UNWIN, D., eds., **Spatial analysis perspectives in GIS**, Taylor and Francis.
- GOODCHILD, M. F. The current status of GIS and spatial analysis. **Journal of Geographical Systems**, v. 2, p. 5-10, 2000.
- GOODCHILD, M. F.; HAINING, R. P. GIS and spatial data analysis: covering perspectives. **Papers in Regional Science**, v. 83, p. 363-385, 2004.
- GRUNSKY, E. C. R: a data analysis and statistical programming environment—an emerging tool for the geosciences. **Computers & Geosciences**, v. 28, p. 1219-1222, 2002.
- KRIVORUCHKO, K. Using Spatial Statistics In GIS. In: International Congress on Modelling and Simulation. Townsville, Australia, 2003. p.
- MARBLE, D. Some thoughts on the integration of spatial analysis and Geographic Information Systems. **Geographical Systems**, v. 2, p. 31-35, 2000.
- NETO, P. R. D. A.; CARRERO, M. A.; JR., P. J. R., 2005, aRT: API R-TerraLib, Paraná, Brasil, Universidade Federal do Paraná, p. 19.
- PEBESMA, E. J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, v. 30, p. 683-691, 2004.
- PEBESMA, E. J.; WESSELING, C. G. Gstat: a program for geostatistical modelling, prediction and simulation. **Computers & Geosciences**, v. 24, n.1, p. 17-31, 1998.
- REIS, I. A., 2005, O estado da arte da integração entre Sistemas de Informação Geográfica e Modelos Inferenciais Bayesianos, São José dos Campos, INPE, p. 27.
- RIBEIRO-JR., P. J.; CHRISTENSEN, O. F.; DIGGLE, P. J. geoR and geoRglm: Software for Model-Based Geostatistics. In: 3rd International Workshop on Distributed Statistical Computing (DSC 2003). Vienna, Austria, 2003. p.



## Questões ?