

MAD-CB: Introdução à Matéria

James Hunter - Laboratório de Retrovirologia, EPM/UNIFESP

7 de fevereiro de 2017

Manuseio e Análise dos Dados

- Como organizar e preparar os dados para a análise
- Uso de Excel como organizador
- Tipos de análises e modelos que podemos construir
 - ▶ Regressão
 - ▶ Análise de variância (ANOVA)
 - ▶ Machine Learning (supervised & unsupervised)
 - ▶ Análises especializadas para estudos biológicos
- Preparando estudos para a publicação
- Assunto primária das aulas

- Um derivativo open-source de linguagem “S”
 - ▶ Desenvolvido pelo antigo Bell Labs
- Primeiro versão: 1999
- Agora, Versão 3.3.2
- Comunidade ativa de desenvolvedores e programadores
- >10.000 pacotes/modulos publicados pela comunidade
- Ferramenta mais popular para modelos de estatística na academia hoje
- Sem custo/GRATIS/FREE

R É Uma Linguagem Não Um Software Completo

- Modelo diferente dos softwares tradicionais
 - ▶ SPSS
 - ▶ Graphpad Prism
 - ▶ SAS
 - ▶ Statistica
 - ▶ STATA

Problemas com os Tradicionais

- Sistemas de menus e caixas escondem o código que os sistemas escrevem

Problemas com os Tradicionais

- Sistemas de menus e caixas escondem o código que os sistemas escrevem
- Difícil saber se você pode reproduzir exatamente a seqüência de comandos

Problemas com os Tradicionais

- Sistemas de menus e caixas escondem o código que os sistemas escrevem
- Difícil saber se você pode reproduzir exatamente a seqüência de comandos
- Custo frequentemente absurdo

- Sistemas de menus e caixas escondem o código que os sistemas escrevem
- Difícil saber se você pode reproduzir exatamente a seqüência de comandos
- Custo frequentemente absurdo
- “*You can only do what the buttons say you can do.*” — Sacha Epskamp, U de Amsterdã¹

¹Baker, Monya, “Code Alert”, **Nature**, Vol 541, 26/1/2017, p. 563 - 565.

Problemas com os Tradicionais

- Sistemas de menus e caixas escondem o código que os sistemas escrevem
- Difícil saber se você pode reproduzir exatamente a seqüência de comandos
- Custo frequentemente absurdo
- *“You can only do what the buttons say you can do.”* — Sacha Epskamp, U de Amsterdã¹
- *“Você pode fazer somente o que os programadores dos botões acham que o botão faz.”* — Jim Hunter, UNIFESP

¹Baker, Monya, “Code Alert”, **Nature**, Vol 541, 26/1/2017, p. 563 - 565.

Com R, Você Tem Controle de Resultados

- Com um pouco treinamento, pode depender que os resultados seriam confiáveis
- Você escreve uma serie de instruções (código) para dirigir as operações do programa
- Você controle, outros podem facilmente auditar
 - ▶ Chave para “reproducible research”
- Curva de aprendizagem
 - ▶ Inicial – suave
 - ▶ Avançado – mais inclinada

Quase Toda a Matéria do Curso Está Sendo Preparada com R

- Estes slides
- Handouts
- Programas
- Graficos
- Com ajuda de IDE para R, RStudio

Projeto Final da Matéria

- Divisão da turma em até 10 grupos
 - ▶ De 1 até 3 pessoas
- Escolha de uma pesquisa para analisar
 - ▶ Pode ser um projeto de tese ou do laboratório
 - ▶ Nós podemos ajudar você com a escolha
- Grupo vai fazer uma apresentação sobre a análise quantitativa dos dados relacionados a tema do projeto
 - ▶ Começando ao final do março
 - ▶ Ordem seria sorteada na aula
- Na aula final, grupo vai submeter um relatório resumindo a análise que fez
 - ▶ Técnicas e ferramentas usadas
 - ▶ Resultados
 - ▶ Quais técnicas funcionaram bem, quais não tão bem
 - ▶ No estilo de uma publicação formal no jornal científico

- Vai ter
- **MAS** são diagnósticos, não avaliações formais
- Provas como experiência de aprendizagem

- Têm em Khan Academy
- Vou criar conjuntos de problemas usando um pacote de R – `swirl`
- Faz parte de nota final

- Jim Hunter estará disponível quintas de 13.30h até 16.00h
 - ▶ Ed. de Pesquisa 2, Pedro de Toledo, 669, 6º Andar fundos
- Para reforçar o que estão aprendendo na aula, na Khan Academy, etc.
- Tirar dúvidas sobre o conteúdo e a administração da matéria
- Não precisa marcar antecipadamente; **Just show up**
- Email do Jim: jhunter@unifesp.br
- Cel do Jim (para Whatsapp): 11-9-5327-5656

- Crawley, **Statistics: An Introduction Using R**, (Wiley)

- Crawley, **Statistics: An Introduction Using R**, (Wiley)
- Diez, Barr & Cetinkaya-Rundel, **OpenIntro Statistics**, (<http://openintro.org>)

- Crawley, **Statistics: An Introduction Using R**, (Wiley)
- Diez, Barr & Cetinkaya-Rundel, **OpenIntro Statistics**, (<http://openintro.org>)
- Navarro, **Learning statistics with R: A tutorial for psychology students and other beginners**, (<http://learningstatisticswithr.com>)

- Crawley, **Statistics: An Introduction Using R**, (Wiley)
- Diez, Barr & Cetinkaya-Rundel, **OpenIntro Statistics**, (<http://openintro.org>)
- Navarro, **Learning statistics with R: A tutorial for psychology students and other beginners**, (<http://learningstatisticswithr.com>)
- Nolan & Speed, **Stat Labs: Mathematical Statistics through Applications**, Springer

- Crawley, **Statistics: An Introduction Using R**, (Wiley)
- Diez, Barr & Cetinkaya-Rundel, **OpenIntro Statistics**, (<http://openintro.org>)
- Navarro, **Learning statistics with R: A tutorial for psychology students and other beginners**, (<http://learningstatisticswithr.com>)
- Nolan & Speed, **Stat Labs: Mathematical Statistics through Applications**, Springer
- Vickers, **What is a P-Value Anyway?**, Addison-Wesley

- Kabacoff, **R in Action: Data analysis and graphics with R, 2e** (Manning)

- Kabacoff, **R in Action: Data analysis and graphics with R**, 2e (Manning)
- Peng, **R Programming for Data Science** (Leanpub)

- Kabacoff, **R in Action: Data analysis and graphics with R**, 2e (Manning)
- Peng, **R Programming for Data Science** (Leanpub)
- Peng, Kross & Anderson, **Mastering Software Development in R** (Leanpub)

- Kabacoff, **R in Action: Data analysis and graphics with R**, 2e (Manning)
- Peng, **R Programming for Data Science** (Leanpub)
- Peng, Kross & Anderson, **Mastering Software Development in R** (Leanpub)
- Phillips, **YaRrr!: The Pirate's Guide to R** (<http://www.thepiratesguidetor.com>)

- Kabacoff, **R in Action: Data analysis and graphics with R**, 2e (Manning)
- Peng, **R Programming for Data Science** (Leanpub)
- Peng, Kross & Anderson, **Mastering Software Development in R** (Leanpub)
- Phillips, **YaRrr!: The Pirate's Guide to R** (<http://www.thepiratesguidetor.com>)
- Wickham & Grolemund, **R for Data Science**, (<http://r4ds.had.co.nz> or Riley)

- Kabacoff, **R in Action: Data analysis and graphics with R**, 2e (Manning)
- Peng, **R Programming for Data Science** (Leanpub)
- Peng, Kross & Anderson, **Mastering Software Development in R** (Leanpub)
- Phillips, **YaRrr!: The Pirate's Guide to R** (<http://www.thepiratesguidetor.com>)
- Wickham & Grolemund, **R for Data Science**, (<http://r4ds.had.co.nz> or Riley)
- Zumel & Mount, **Practical Data Science with R** (Manning)

- Hansen, **Bioconductor: An Introduction to Core Technologies** (Leanpub)
- Irizarry & Love, **Data Analysis for the Life Sciences** (Leanpub)

Livros de Leanpub: <https://leanpub.com>

5 Livros que Deve Ler Porque São Bons

- Leonard Mlodinow, **O Andar do Bêbado**

5 Livros que Deve Ler Porque São Bons

- Leonard Mlodinow, **O Andar do Bêbado**
- David Salsburg, **Uma Senhora Toma Chá**

5 Livros que Deve Ler Porque São Bons

- Leonard Mlodinow, **O Andar do Bêbado**
- David Salsburg, **Uma Senhora Toma Chá**
- Ian Stewart, **17 Equações que Mudaram o Mundo**

5 Livros que Deve Ler Porque São Bons

- Leonard Mlodinow, **O Andar do Bêbado**
- David Salsburg, **Uma Senhora Toma Chá**
- Ian Stewart, **17 Equações que Mudaram o Mundo**
- Peter L. Bernstein, **Desafiando os Deuses: A História do Risco**

5 Livros que Deve Ler Porque São Bons

- Leonard Mlodinow, **O Andar do Bêbado**
- David Salsburg, **Uma Senhora Toma Chá**
- Ian Stewart, **17 Equações que Mudaram o Mundo**
- Peter L. Bernstein, **Desafiando os Deuses: A História do Risco**
- Randall Munroe, **What If?: Serious Scientific Answers to Absurd Hypothetical Questions**

- Precisa contar para mim o que vocês querem aprender

- Precisa contar para mim o que vocês querem aprender
- Precisa me avisar se estou indo rápido demais ou devagar demais

- Precisa contar para mim o que vocês querem aprender
- Precisa me avisar se estou indo rápido demais ou devagar demais
- Vou tentar deixar cálculo e álgebra linear fora (mas, não é sempre possível)

- Precisa contar para mim o que vocês querem aprender
- Precisa me avisar se estou indo rápido demais ou devagar demais
- Vou tentar deixar cálculo e álgebra linear fora (mas, não é sempre possível)
- Vou usar dados de nossa área quando for possível – se tiver alguns datasets interessantes, me avise

- Pergunte bastante; participação ajuda todos
- Se você fica com uma pergunta, pode garantir que outra pessoa tem a mesma pergunta
- Não existem perguntas burras

*There are naive questions, tedious questions, ill-phrased questions, questions put after inadequate self-criticism. But every question is a cry to understand the world. **There is no such thing as a dumb question.***

- Carl Sagan, **The Demon-Haunted World: Science as a Candle in the Dark**, p. 303.