





R e *Data Science*: como entender o que dizem os dados com o **R**?

IMIL na Sala de Aula - IE/UFRJ

Vítor Wilher

analisemacro.com.br

9 de Novembro de 2018

O plano de voo para hoje

- Sobre o Autor
- O mundo dos dados
 - Getting and Cleaning data
 - Exploratory Data Analysis
 - Modeling
 - Communication
- O mundo do R
- Alguns problemas práticos
- Data Science e os Economistas
- Contato

2/21

Vítor Wilh

Sobre o Autor

O mundo dos

Getting and Clear data

Exploratory Data Analysis

Modeling Communication

O mundo do F

Alguns problemas

e os Economistas

Contato

Sobre o Autor

Vítor Wilher é Bacharel e Mestre em Economia, pela Universidade Federal Fluminense, com especialização em Data Science pela Johns Hopkins University. Sócio-Fundador da Análise Macro, empresa especializada em treinamento e consultoria em data science. É também Conselheiro do Instituto Millenium.

Maiores informações, visite www.analisemacro.com.br

Vítor Wilh

Sobre o Auto

O mundo dos dados

Getting and Clean data

Exploratory Dat Analysis

Modeling Communication

O mundo do F

Alguns problemas práticos

e os Economistas

Contato

O mundo dos dados

O avanço da informática e das telecomunicações possibilitou o armazenamento e a distribuição de conjuntos de dados cada vez mais complexos. Lidar com essas bases de dados exigiu a sistematização de diversas técnicas de coleta, tratamento, análise e apresentação de dados.

Vítor Wilh

Sobre o Autor

O mundo dos dados

data
Exploratory Data
Analysis
Modeling
Communication

O mundo do F

Alguns problemas práticos

e os Economistas

Contat

O mundo dos dados

Essa sistematização de técnicas deu origem ao que hoje chamamos de **data science**, cujo objetivo principal é extrair informações úteis de conjuntos de dados aparentemente confusos.

Aplicações interessantes:

- Identificar mensagens indesejáveis em um e-mail (spam);
- Segmentação do comportamento de consumidores para propagandas direcionadas;
- Redução de fraudes em transações de cartão de crédito;
- Predição de eleições;
- Otimização do uso de energia em casas ou prédios;
- etc, etc, etc...

Vítor Wilh

Sobre o Autor

O mundo dos dados

Exploratory Data Analysis Modeling

O mundo do F

Alguns problemas

Data Science e os Economistas

Contato

O mundo dos dados

De modo a responder esse tipo de pergunta, é necessário cumprir aquelas quatro etapas da ciência de dados.

As quatro operações:

- É preciso coletar os dados;
- Dados brutos precisam ser tratados;
- Uma vez disponíveis, os dados precisam ser analisados de forma a extrair informações relevantes e/ou responder determinados questionamentos;
- Com as respostas em mãos, é preciso apresentar os resultados.

Vitor Wilh

Sobre o Autor

O mundo dos dados

Getting and Clear

Exploratory Data Analysis

Modeling Communication

O mundo do l

Alguns problemas práticos

e os Fronomistas

Contato

O mundo dos dados

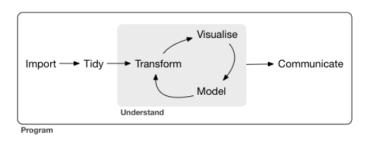


Figura: Fonte: R for Data Science.

Cada uma dessas etapas exige conhecimentos específicos, de modo a lidar com diferentes formatos de dados, bem como responder questões distintas.

Vítor Wilh

Sobre o Auto

O mundo dos dados

Getting and Cleaning

Exploratory Da Analysis

Modeling Communication

O mundo do F

Alguns

Data Science os

Economista:

Contato

Getting and Cleaning data

Dados podem estar dispostos em diferentes formatos:

- Excel;
- XML;
- JSON;
- txt;
- HTML;
- MySQL;
- Formatos proprietários (Weka, Stata, Minitab, Octave, SPSS, SAS, etc).

Vítor Will

Sobre o Auto

O mundo dos dados

Getting and Cleaning data

Analysis

Modeling

Communication

O mundo do l

problemas práticos

e os Economistas

Contat

Getting and Cleaning data

Dados precisam ser tratados:

- Limpeza de dados;
- Tratamento de missing values;
- Construção de números índices;
- Deflacionar valores correntes;
- Obtenção de taxas de crescimento, a partir de comparações mensais, interanuais, acumuladas em 12 meses, etc;
- Tratando tendências;
- Dessazonalização;
- Obtendo subconjuntos (subsetting) relevantes;
- Classificando dados de acordo com algum critério;
- Transformando dados de acordo com alguma operação.

Vítor Wilh

Sobre o Auto

O mundo do: dados

Getting and Clea

Exploratory Data Analysis Modeling

Communication

O mundo do F

problemas práticos

Data Scienc e os Economistas

Contato

Exploratory Data Analysis

Dados precisam ser visualizados:

- Gráficos simples;
- Gráficos de correlação;
- Clustering;

data

Exploratory Data

Modeling

O mundo do R

A.

práticos

Data Scienc

e os Economistas

Contate

Modeling

Dados podem ser relacionados uns aos outros.

- Modelos ARIMA;
- Regressão linear;
- Árvores de regressão;
- Neural Network;
- Support Vector Machine;
- Naive Bayes;
- etc, etc, etc.

Vítor Wilh

Sobre o Auto

O mundo dos

Getting and Clea

Exploratory Da Analysis

Communication

O mundo do R

Alguns problemas

Data Scien e os –

Economista

Contato

Communication

Os resultados precisam ser comunicados através de *documentos reprodutíveis*, que unam **código** e **texto**.

Vítor Wilh

Sobre o Autor

O mundo do: dados

data
Exploratory Data
Analysis
Modeling

O mundo do R

praticos Data Scienci e os

Contato

O mundo do R

Era necessário construir uma plataforma que unisse todas essas etapas. O $\bf R$ é uma das melhores soluções atualmente disponíveis, dados os seguintes motivos:

- A existência de uma comunidade grande e bastante entusiasmada, que compartilha conhecimento todo o tempo;
- o R é gratuito, open source, de modo que você não precisa comprar licenças de software para instalá-lo;
- Tem inúmeras bibliotecas (pacotes) em estatística, machine learning, visualização, importação e tratamento de dados;
- Possui uma linguagem estabelecida para data analysis;
- Ferramentas poderosas para comunicação dos resultados da sua pesquisa, seja em forma de um website ou em pdf.

Vítor Wilh

Sobre o Auto

O mundo dos

Getting and Cleadata

Exploratory Data Analysis Modeling

O mundo do R

Alguns problema práticos

Data Science e os Economistas

Contato

O mundo do R

Ao aprender **R**, você conseguirá integrar as etapas de coleta, tratamento, análise e apresentação de dados em um único ambiente. Você vai esquecer ter de abrir o excel, algum pacote estatístico, depois o power point ou o word, depois um compilador de pdf para gerar seu relatório. Todas essas etapas serão feitas em um único ambiente. E essa talvez seja a grande motivação para você entrar de cabeça nesse mundo.

Vítor Wilh

Sobre o Auto

O mundo dos

data
Exploratory Data
Analysis
Modeling

O mundo do R

problemas práticos

e os Economistas

Contato

O mundo do R

- Baixe o R em http://cran-r.c3sl.ufpr.br/;
- Baixe o RStudio em https://www.rstudio.com/products/rstudio/download/;
- Baixe o MikTex se você for usuário de Windows em http://miktex.org/download;
- Baixe o MacTex se você for usuário de Mac em http://www.tug.org/mactex/.

litor Will

Sobre o Autor

O mundo dos dados

Getting and Cle

data Exploratory Data Analysis

O mundo do R

o manao ao r

problem

Data Scien e os

Economista

Contato

O mundo do R

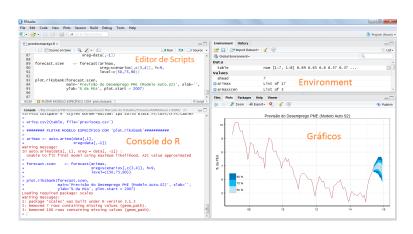


Figura: Ambiente do RStudio.

Vítor Will

Sobre o Auto

O mundo dos dados

data Exploratory Data Analysis

Modeling Communication

Alguns

práticos Data Science e os

Contato

Alguns problemas práticos

- Existe relação entre desemprego e procura pela Uber?
- ② Os desembolsos do BNDES fizeram aumentar a taxa de investimento da economia brasileira?
- Omo podemos explicar a inflação de alimentos?
- O Banco Central brasileiro reage a choques cambiais?
- Qual o efeito do aumento de volatilidade no mercado sobre a taxa de câmbio?

Vítor Will

Sobre o Auto

O mundo dos dados Getting and Cleani data Exploratory Data

Modeling Communication

O mundo do F

problemas práticos Data Scienc

Alguns

e os Economistas

Contat

Alguns problemas práticos

De modo a analisar essas questões, alguns problemas imediatos surgem:

- Onde estão os dados?
- Qual proxy utilizar para representar, por exemplo, o interesse pela Uber?
- Omo tratar os dados brutos obtidos das fontes primárias?
- Qual a estrutura dos dados?
- Uma vez que as questões anteriores estejam resolvidas, qual o melhor modelo para analisar a relação entre as variáveis?

Vítor Will

Sobre o Auto

O mundo dos

data

Analysis

Modeling

Communication

O mundo do F

Data Science

e os Economistas

Contato

Data Science e os Economistas

Os economistas estão acostumados com **modelagem** de variáveis econômicas, de modo que podem e devem se beneficiar das técnicas de *data science*, que envolvem coleta, tratamento, análise e apresentação de dados.

Em um mundo onde se discute a perda de empregos para robôs, os economistas devem se preparar para temas como *Big Data*, *Machine Learning*, *Statistical Learning*, *Predictive Modeling*, etc...

Vítor Wilhe

Sobre o Auto

O mundo dos dados

Getting and Clean

Exploratory Data

Modeling

Communication

O manao do n

Alguns

Data Science e os Economistas

Contato

Data Science e os Economistas

Para terminar, uma provocação. Enquanto vocês são obrigados a ler Marx, no MIT...

Computer Science, Economics, and Data Science (Course 6-14) Computer Science, Economics, and Data Science Bachelor of Science in Computer Science, Economics, and Data Science General Institute Requirements (GIRs) The General Institute Requirements include a Communication Requirement that is integrated into both the HASS Requirement and the requirements of each major; see details below. Summary of Subject Requirements Subjects Science Requirement Humanities, Arts, and Social Sciences (HASS) Requirement (between one and three subjects can be from the Departmental Program); at least two of these subjects must be designated as communication-intensive (CI-H) to fulfill the Communication Requirement. Restricted Electives in Science and Technology (REST) Requirement I can be satisfied by 6.042[J] and 18.06 in the Departmental Program]

Figura: Uma nova graduação no MIT.

ítor Wilhe

Sobre o Auto

O mundo dos dados

detting and Cit

Exploratory Data Analysis

Modeling Communication

O mundo do F

Data Scien

Economistas

Contato

Slides estão disponíveis no repositório da Análise Macro no Github: https://github.com/analisemacro/degustacao.

Visite:

www.analisemacro.com.br

