

Introdução

1 Objetivos

2 Introdução

- Econometria e Teoria Econômica
- Econometria e Teoria Econômica
- Modelo Econômico
- Modelo Econométrico
- A Estrutura dos Dados Econômicos
 - Dados de Corte Transversal
 - Série Temporal
 - Dados em Pannel
- Ciência de Dados
 - Definição
 - Áreas de Conhecimento
 - O Processo de Ciência de Dados
- Big Data
 - Dados Gerados por Minuto
 - Velocidade de Crescimento dos Dados

Ao final desta aula os alunos deverão ser capazes de:

- Conhecer alguns campos de atuação da Econometria e sua relação com a teoria econômica;
- Diferir sobre Modelos Econômicos e Modelos Econométricos;
- Entender o papel dos dados na definição do método econométrico a ser utilizado;
- Entender conceitos de Ciência de Dados e sua relação com a Econometria;
- Conhecer a plataforma DataCamp que pode ser usada para aprendizagem da linguagem R.

- Por meio da Econometria é possível avaliar empiricamente a teoria econômica:
 - Explicar fatos passados;
 - Testar teorias e hipóteses;
 - Prever resultados de políticas ou eventos futuros;
 - Estimar relações entre variáveis econômicas.
- Isso é viável porque, em geral, **há relações de equilíbrio de longo prazo** entre variáveis econômicas.

- O economista pode fazer uso de diversos campos da econometria de acordo com o **fundamento econômico**:
 - Modelos de Regressão: Regressão Linear Múltipla, Regressão Quantílica, Regressões Penalizadas (Ridge, LASSO, ...)
 - Modelos de séries temporais (ARIMA, ARCH/GARCH, VAR, VEC, MIDAS, ...)
 - Modelagem não-paramétrica and semi-paramétrica (Regressão Linear Local, Splines, LOESS, Wavelets, Modelos Aditivos, ...)
 - Microeconometria (dados em painel, ...)
 - Macroeconometria (DSGE, DSGE-VAR, ...)
 - Classificação (Análise Discriminante, ...)
 - Clustering (K-Means, ...)
 - Redução da Dimensionalidade (PCA, Análise Fatorial, ...)
 - Machine Learning e Inteligência Artificial (Redes Neurais, Support Vector Machines, Boosting, Random Forests, Ensemble, ...)

Quais os efeitos do **treinamento** sobre a **produtividade** do trabalhador?

$$saláριοh = f(educ, \quad exper, \quad treina) \quad (1)$$

onde:

- *salário_h* é o salário-hora;
- *educ* representa os anos de educação formal;
- *exper* refere-se aos de experiência no mercado de trabalho;
- *treina* corresponde as semanas ocupadas em treinamento.

Hipótese: Os trabalhadores são pagos de acordo com sua produtividade.

Um modelo econométrico para o exemplo anterior

$$\text{salário}_i = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{exper}_i + \beta_3 \text{treina}_i + u_i \quad (2)$$

em que o termo u contém fatores observados ou não observados, mas que podem influenciar a produtividade, tais como:

- aptidão inata;
- qualidade da educação;
- formação da família.

Objetivo: Testar hipótese sobre o parâmetro β_3 . Exemplo: ele é diferente de zero?

Dados de Corte Transversal

Amostra de indivíduos, empresas, estados, países em um determinado ponto no tempo. Ignoramos o tempo. Tipo de dado mais usado em **Estatística Econômica**.

Conjunto de Dados de Corte Transversal sobre Salários e outras Características Individuais

<i>nobsi</i>	<i>salárioh</i>	<i>educ</i>	<i>exper</i>	<i>feminino</i>	<i>casado</i>
1	3,10	11	2	1	0
2	3,24	12	22	1	1
3	3,00	11	2	0	0
4	6,00	8	44	0	1
5	5,30	12	7	0	1
.
.
.
525	11,56	16	5	0	1
526	3,50	14	5	1	0

Série Temporal

Dados de séries de tempo consistem de observações sobre uma variável ou muitas variáveis ao longo do tempo (preços de ações, IPCA, PIB, Câmbio, Taxa de Juros, ...). Dados mais usados em **Econometria**. Frequência: intradiária, diária, semanal, mensal, trimestral, anual, etc.

Salário Mínimo, Desemprego e Dados Relacionados para Porto Rico

<i>nobsa</i>	<i>ano</i>	<i>minmed</i>	<i>cobmed</i>	<i>desemp</i>	<i>pnb</i>
1	1950	0,20	20,1	15,4	878,7
2	1951	0,21	20,7	16,0	925,0
3	1952	0,23	22,6	14,8	1.015,9
.
.
.
37	1986	3,35	58,1	18,9	4.281,6
38	1987	3,35	58,2	16,8	4.496,7

Dados em Painel

Um conjunto de dados em painel (ou dados longitudinais) consiste em uma série de tempo para cada variável do corte transversal. Este tipo de dado costuma ser enfocado em cursos de **Microeconometria**. Temos as mesmas variáveis do corte transversal ao longo do tempo.

Conjunto de Dados de Painel sobre Estatísticas de Crime nas Cidades para Dois Anos

<i>nobs</i>	<i>cidade</i>	<i>ano</i>	<i>homicds</i>	<i>população</i>	<i>desemp</i>	<i>polícia</i>
1	1	1986	5	350.000	8,7	440
2	1	1990	8	359.200	7,2	471
3	2	1986	2	64.300	5,4	75
4	2	1990	1	65.100	5,5	75
.
.
.
297	149	1986	10	260.700	9,6	286
298	149	1990	6	245.000	9,8	334
299	150	1986	25	543.000	4,3	520
300	150	1990	32	546.200	5,2	493

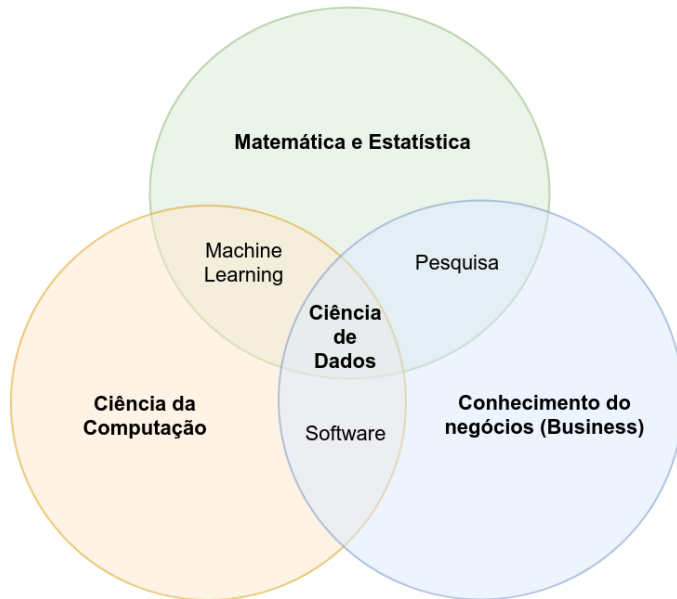
Definição

- Com a migração para a internet, produzimos um fluxo constante e exaustivo de informação digital;
- Estima-se que **90% dos dados armazenados no mundo** foram produzidos apenas nos **últimos dois anos**;
- Informação tornou-se a moeda mais poderosa e exige que o mercado saiba interpretá-la a seu favor (para gerar valor);
- Logo, a profissão de Cientista de Dados se torna crucial;
- Áreas de atuação: varejo, saúde, finanças, telecomunicações, segurança, transporte, economia, recursos humanos, ...

Definição 1: Ciência de Dados é o termo usado para o processo de extrair *insights* de dados que são coletados de várias fontes (estruturadas e não estruturadas)

Definição 2: Consiste de um conjunto de especialidades, tais como estatística, matemática, programação, computação e *business*

Áreas de Conhecimento

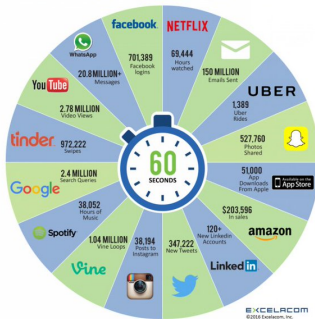


O Processo de Ciência de Dados

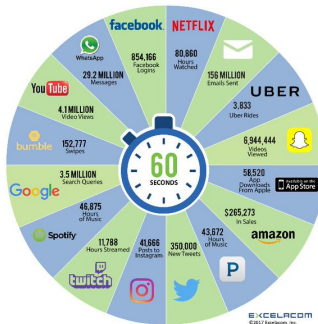
- 1 Identificar o problema da área de negócio;
- 2 Compreender o problema (entidades e atributos);
- 3 Coletar conjuntos de dados que representem as entidades;
- 4 Limpar e transformar os dados;
- 5 Compreender os relacionamentos entre os dados;
- 6 Criar modelos estatísticos e matemáticos que representem os relacionamentos;
- 7 Utilizar os modelos para fazer previsões;
- 8 Entregar valor e resultado.

Dados Gerados por Minuto

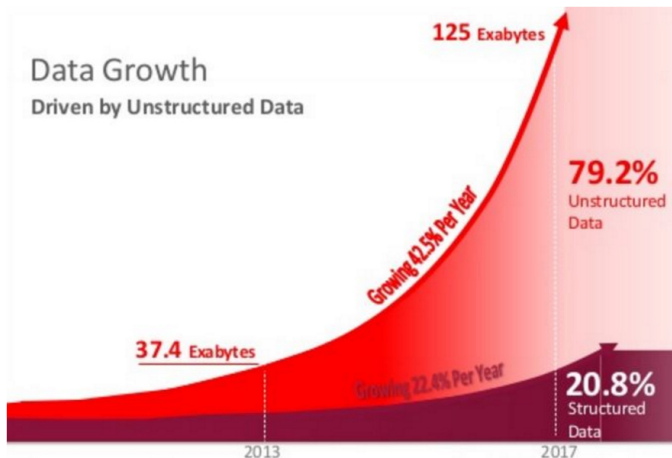
2016 What happens in an INTERNET MINUTE?



2017 What happens in an INTERNET MINUTE?



Velocidade de Crescimento dos Dados



Source: ISD - 2014, *Structured Data vs. Unstructured Data: The Balance of Power Continues to Shift*

Tipos de Dados

Structured Data



0.103	0.176	0.387	0.300	0.379
0.333	0.384	0.564	0.587	0.857
0.421	0.309	0.654	0.729	0.228
0.266	0.750	1.056	0.936	0.911
0.225	0.326	0.643	0.337	0.721
0.187	0.586	0.529	0.340	0.829
0.153	0.485	0.560	0.428	0.628

Unstructured Data



Economia dos Dados

The Economist: publicação de 2017 que diz que ter dados atualmente é o mesmo que ter petróleo há 100 anos e que essa nova onda está criando o que os autores chamam de **Economia dos Dados**



Harvard Business Review: publicação de 2012 que abordava o perfil do cientista de dados

Data Scientist: *The Sexiest Job of the 21st Century*

**Meet the people who
can coax treasure out of
messy, unstructured data.**

by Thomas H. Davenport
and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't swiping out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

70 Harvard Business Review October 2012