

Introdução à Ciência de Dados como Prática na Pesquisa Acadêmica

Seção 01 - Conceitos Básicos de Ciências de Dados

Jadson Pessoa

Professor do DECON | Membro GAPE



Plano de Trabalho

Objetivos

O que é Ciência de Dados?

Linguagens de Programação em Ciência de Dados

Workflow em Ciência de Dados

R, RStudio e Repositórios

Objetivos

- Apresentar qual o entendimento que temos hoje do termo “**Ciência de Dados**” ou, como ficou mais conhecido, **Data Science** (DS), em inglês.

O que é Ciência de Dados?

Data science [...] allow you to turn raw data into understanding, insight, and knowledge. (Wickham and Grolmund [2017])

Data science combines multiple fields, including statistics, scientific methods, artificial intelligence (AI), and data analysis, to extract value from data Oracle.

Data science combines the scientific method, math and statistics, specialized programming, advanced analytics, AI, and even storytelling to uncover and explain the business insights buried in data IBM.

Áreas multidisciplinares

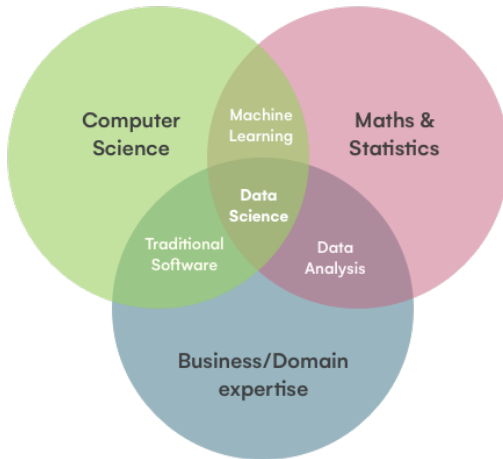


Figure 1: Data Science

Big Data

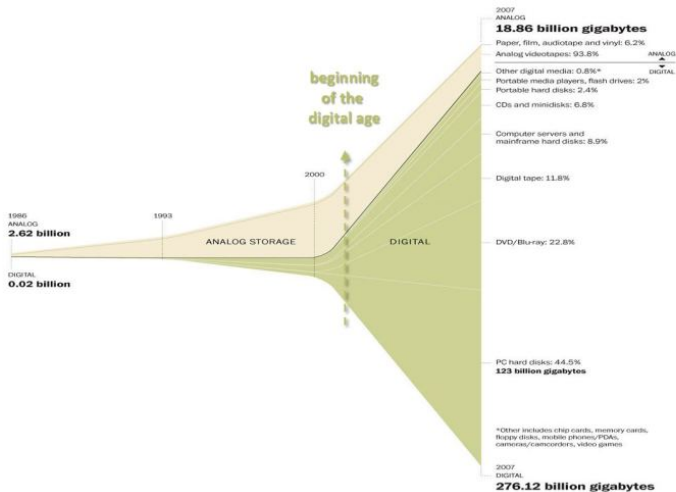


Figure 2: Big Data (Hilbert and López [2011])

Mundo Conectado

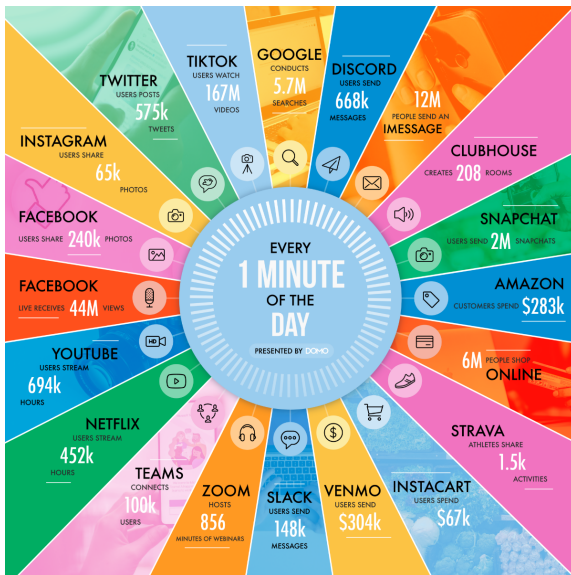


Figure 3: Big Data (DOMO [2020])

Linguagens de Programação em Ciência de Dados

- Dados e Escalabilidade;
- Em um projeto de DS temos a utilização de linguagem, ou melhor, *linguagens* de programação;
- Qual é melhor linguagem de programação?
- Programação orientada à problema.
- R

Não iremos trabalhar com:

- Big Data;
- Python, Julia e outras linguagens;
- Dados não regulares (imagens, sons, textos, etc.).

Workflow em Ciência de Dados

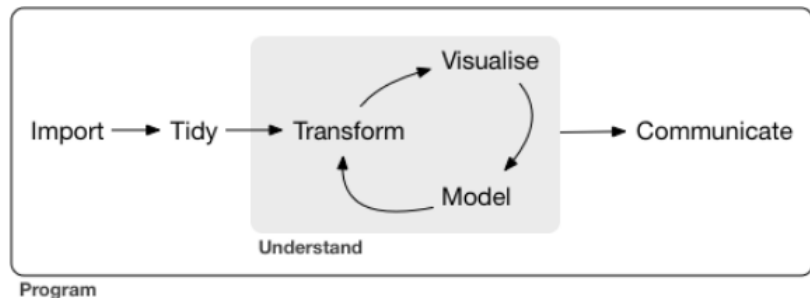


Figure 4: O processo de compreensão dos dados ([Wickham and Grolemund \[2017\]](#))

Sistematização - Projeto em DS:

- Os verbos do DS:

1. Coletar;
2. Tratar;
3. Visualizar;
4. Modelar;
5. Comunicar.

R, RStudio e Repositórios

- Baixar a linguagem R - [CRAN](#).
- Baixar um "Ambiente de Desenvolvimento (IDE)" - [RStudio](#)
- Repositórios - [Github](#)
- Dúvidas: [Stack Overflow](#)

Referências I

- DOMO. Data Never Sleeps 9.0, 2020. URL <https://www.domo.com/learn/infographic/data-never-sleeps-9>.
- Martin Hilbert and Priscila López. The world's technological capacity to store, communicate, and compute information. *science*, 332(6025):60–65, 2011.
- H. Wickham and G. Grolemund. *R for Data Science*. O'Reilly Media, 2017.