

# Data Science: muito além do ‘hype’

Daniel Ikenaga - Data Scientist and CTO at Jupyter

“Os dados estão se tornando a nova matéria-prima do negócio”.

**Craig Mundie, Senior Advisor to the CEO at Microsoft.**

# ...2012 Global Market Capitalization Leaderboard = Tech = 20% of Top 20 Companies...40% of Top 5

| Rank  | Company                 | Region      | Industry Segment   | 5/31/2012 Value (\$B) | 2011 Revenue (\$B) |
|-------|-------------------------|-------------|--------------------|-----------------------|--------------------|
| 1     | Apple                   | USA         | Tech – Hardware    | \$540                 | \$128              |
| 2     | Exxon Mobil             | USA         | Financial Services | 368                   | 434                |
| 3     | PetroChina              | China       | Energy             | 267                   | 318                |
| 4     | Microsoft               | USA         | Tech – Software    | 245                   | 72                 |
| 5     | ICBC                    | China       | Financial Services | 227                   | 70                 |
| 6     | Wal-Mart                | USA         | Retail             | 224                   | 447                |
| 7     | IBM                     | USA         | Tech – Hardware    | 223                   | 107                |
| 8     | China Mobile            | China       | Telecom            | 203                   | 84                 |
| 9     | General Electric        | USA         | Industrial         | 202                   | 143                |
| 10    | AT&T                    | USA         | Telecom            | 200                   | 127                |
| 11    | Royal Dutch Shell       | Netherlands | Energy             | 197                   | 470                |
| 12    | Berkshire Hathaway      | USA         | Financial Services | 196                   | 141                |
| 13    | Chevron                 | USA         | Energy             | 194                   | 236                |
| 14    | Google / Alphabet       | USA         | Tech – Internet    | 189                   | 38                 |
| 15    | Nestlé                  | Switzerland | Food / Beverages   | 180                   | 90                 |
| 16    | China Construction Bank | China       | Financial Services | 173                   | 58                 |
| 17    | Johnson & Johnson       | USA         | Healthcare         | 171                   | 65                 |
| 18    | Procter & Gamble        | USA         | Consumer Goods     | 171                   | 84                 |
| 19    | Wells Fargo             | USA         | Financial Services | 170                   | 73                 |
| 20    | BHP Billiton            | Australia   | Metals / Mining    | 170                   | 75                 |
| Total |                         |             |                    | \$4,512               | \$3,257            |

# 2017 Global Market Capitalization Leaderboard =

## Tech = 40% of Top 20 Companies...100% of Top 5...

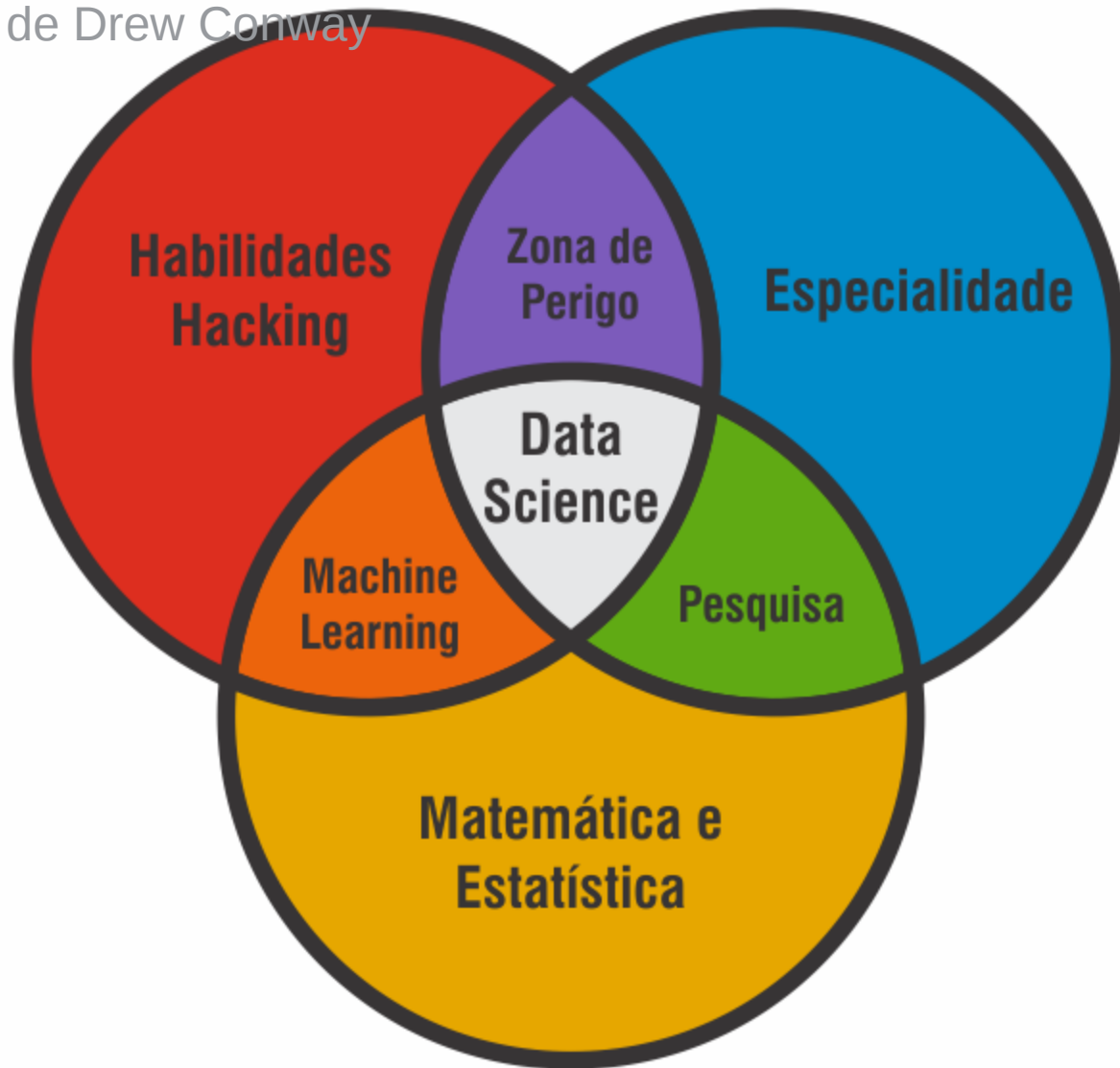
| Rank  | Company             | Region      | Industry Segment   | Current Market Value (\$B) | 2016 Revenue (\$B) |
|-------|---------------------|-------------|--------------------|----------------------------|--------------------|
| 1     | Apple               | USA         | Tech – Hardware    | \$801                      | \$218              |
| 2     | Google / Alphabet   | USA         | Tech – Internet    | 680                        | 90                 |
| 3     | Microsoft           | USA         | Tech – Software    | 540                        | 86                 |
| 4     | Amazon              | USA         | Tech – Internet    | 476                        | 136                |
| 5     | Facebook            | USA         | Tech – Internet    | 441                        | 28                 |
| 6     | Berkshire Hathaway  | USA         | Financial Services | 409                        | 215                |
| 7     | Exxon Mobil         | USA         | Energy             | 346                        | 198                |
| 8     | Johnson & Johnson   | USA         | Healthcare         | 342                        | 72                 |
| 9     | Tencent             | China       | Tech – Internet    | 335                        | 22                 |
| 10    | Alibaba             | China       | Tech – Internet    | 314                        | 21                 |
| 11    | JP Morgan Chase     | USA         | Financial Services | 303                        | 90                 |
| 12    | ICBC                | China       | Financial Services | 264                        | 85                 |
| 13    | Nestlé              | Switzerland | Food / Beverages   | 263                        | 88                 |
| 14    | Wells Fargo         | USA         | Financial Services | 262                        | 85                 |
| 15    | Samsung Electronics | Korea       | Tech – Hardware    | 259                        | 168                |
| 16    | General Electric    | USA         | Industrial         | 238                        | 120                |
| 17    | Wal-Mart            | USA         | Retail             | 237                        | 486                |
| 18    | AT&T                | USA         | Telecom            | 234                        | 164                |
| 19    | Roche               | Switzerland | Healthcare         | 233                        | 51                 |
| 20    | Bank of America     | USA         | Financial Services | 231                        | 80                 |
| Total |                     |             |                    | \$7,207                    | \$2,497            |

# Data Scientist

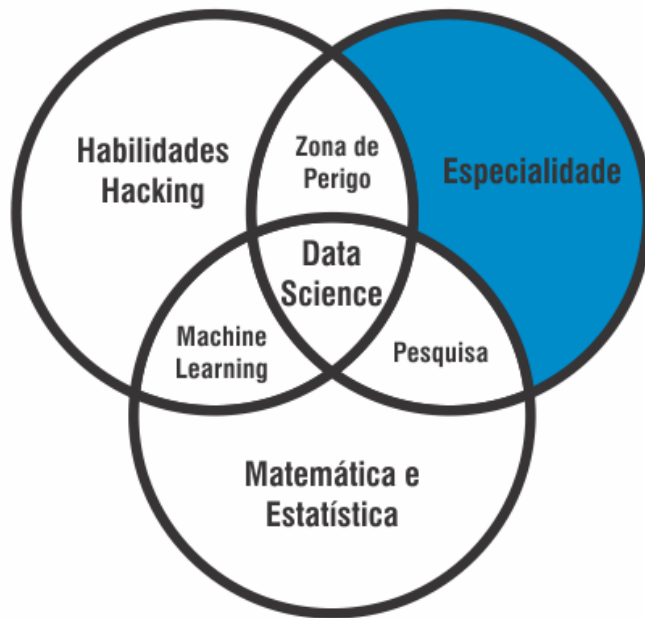
100% Data Science

# Data Science: Conjunto de habilidades

Modificado de Drew Conway

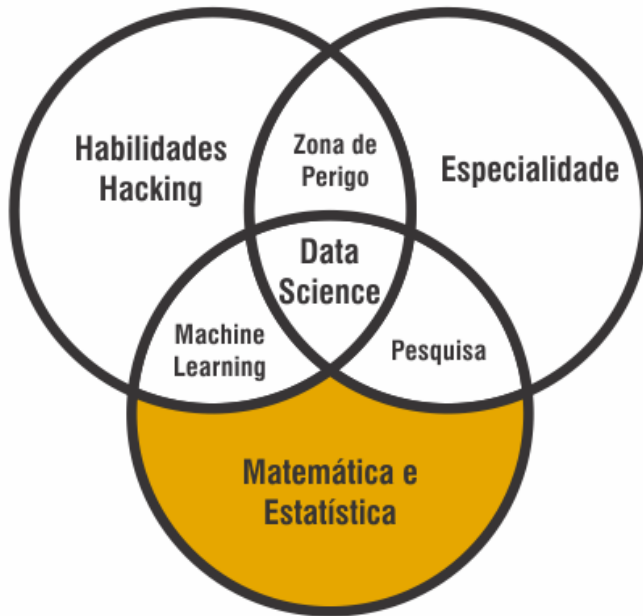


# Especialidade



1. Formação
2. Conhecimento específico
3. Motivação de trabalho

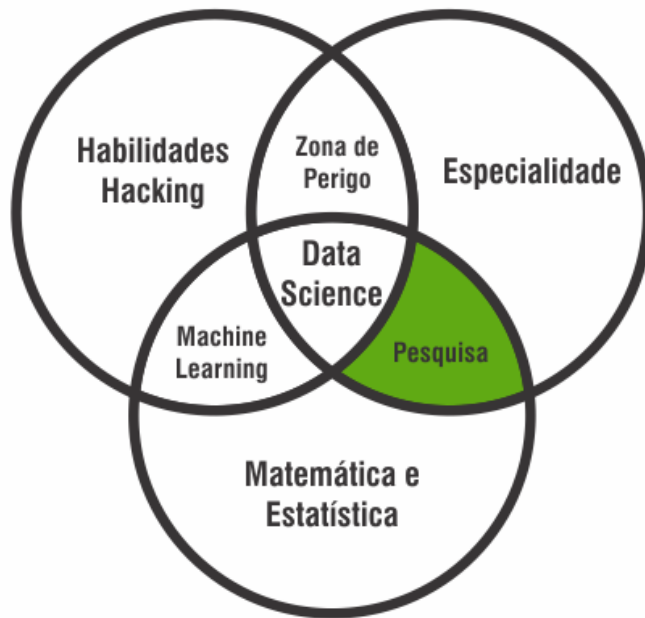
# Matemática e Estatística



1. Álgebra linear
2. Probabilidade
3. Inferência
4. Regressão linear
5. Multivariada
6. Estatística computacional

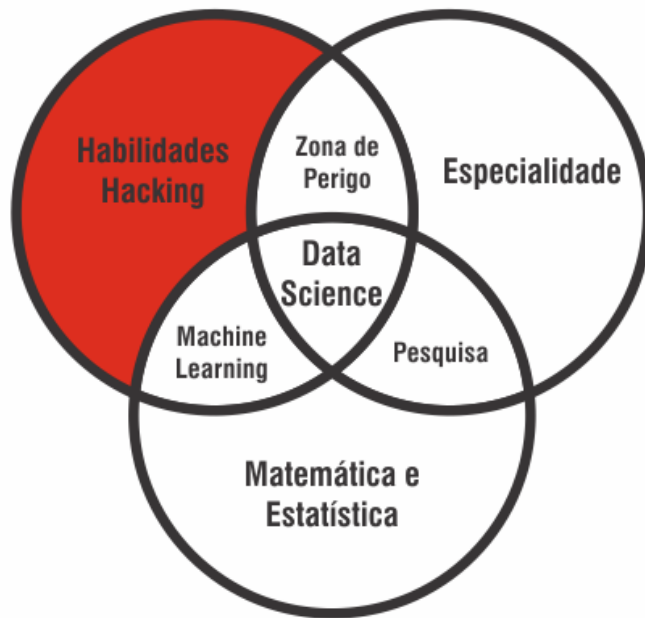


# Pesquisa



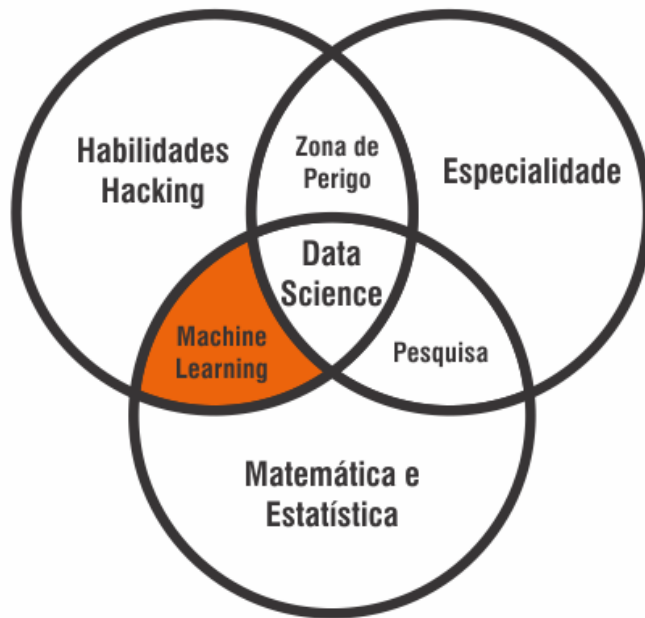
## 1. Método científico

# Habilidades hacking



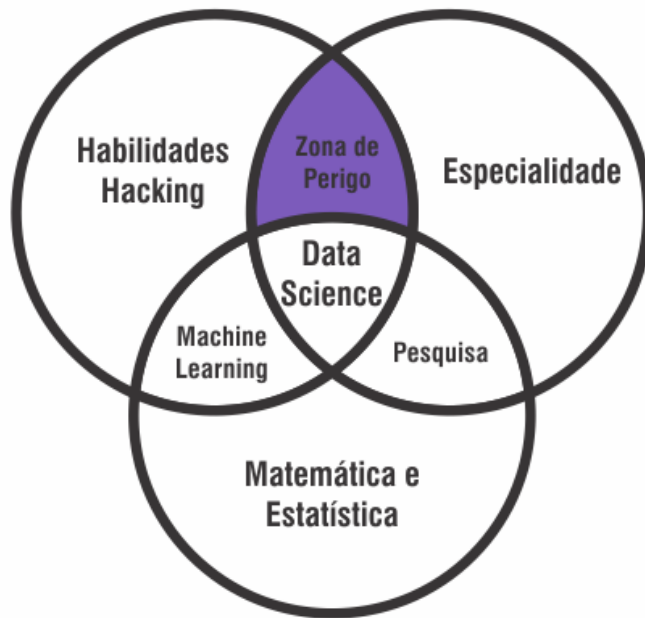
1. Compreensão das tecnologias
2. Armazenamento de dados
3. Desenvolvimento de software

# Machine Learning

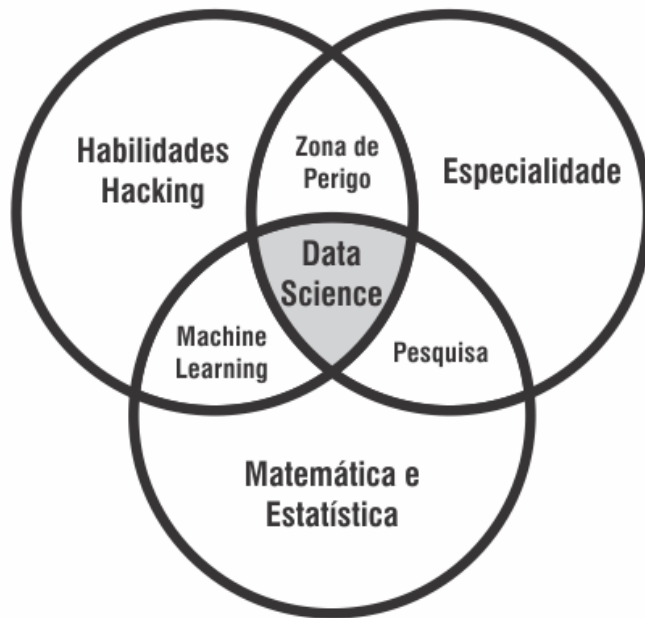


1. Regressão linear/logística
2. Gradiente descendente
3. Regularização
4. Classificação
5. Redução de dimensionalidade
6. Support vector machine
7. Árvores de decisão

# Zona de Perigo



# Data Science



1. Data driven
2. Analisar dados, gerar insights e criar novas e rentáveis soluções em produtos

# Linguagens de programação e marcação

1. R
2. Python
3. Julia
4. Markdown
5. LaTeX

# Fonte de dados

1. SQL
2. NoSQL
3. Graph
4. API
5. Scrapping

# Ciclos de projeto em Data Science

1. Compreensão do assunto
2. Definir as hipóteses
3. Coleta e tratamento de dados
4. Análise exploratória
5. Desenvolvimento de modelos e testes
6. Avaliação e testes de predição
7. Reporte, apresentação e visualização
8. Desenvolvimento da aplicação



# Obrigado!