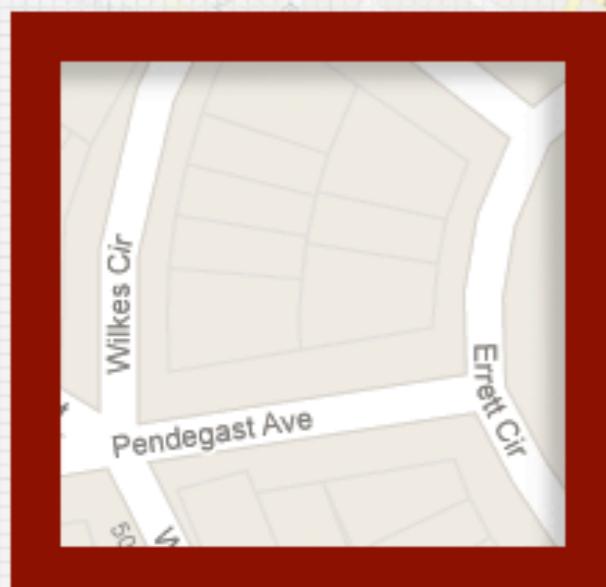




# Data Science, Machine Learning and Big Data

Fabrício J. Barth

---



# PredPol®

## Predict Crime in Real Time.®

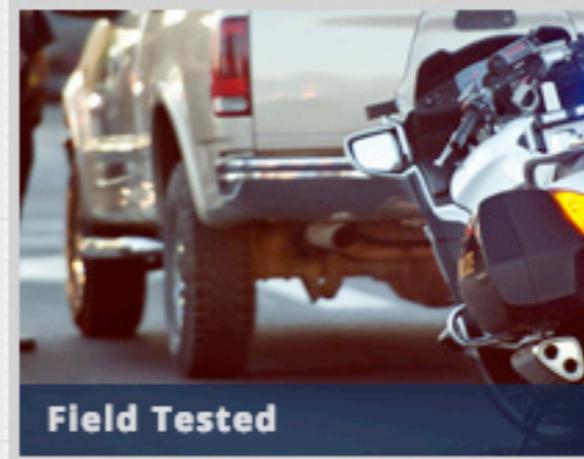
PredPol provides targeted, real-time crime prediction designed for and successfully tested by officers in the field.

■ SEE PREDICTIVE POLICING IN ACTION. LOS ANGELES, CALIFORNIA

[LEARN MORE](#)



Preventative Tactics



Field Tested



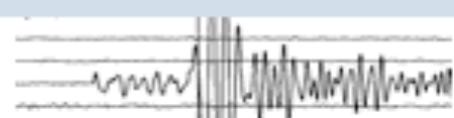
Easy to Use



In The News

Entrada: 13 milhões de registros históricos sobre crimes em LA.  
Saída: determinar quando um crime irá acontecer.

## Earthquake Hazards Program

[Home](#)[About Us](#)[Contact Us](#)[EARTHQUAKES](#)[HAZARDS](#)[LEARN](#)[PREPARE](#)[MONITORING](#)[RESEARCH](#)

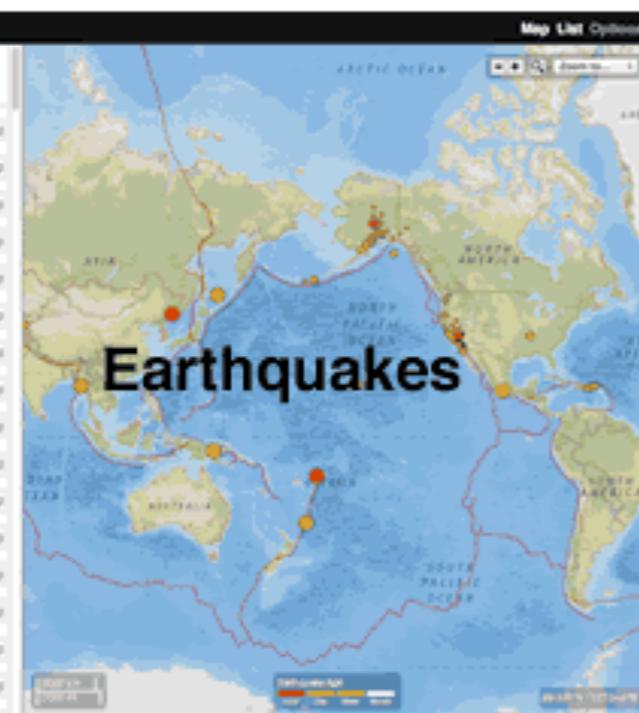
The USGS Earthquake Hazards Program is part of the [National Earthquake Hazards Reduction Program \(NEHRP\)](#), established by Congress in 1977. We monitor and report earthquakes, assess earthquake impacts and hazards, and research the causes and effects of earthquake.

### Latest Earthquakes



1 Day, All Magnitudes

Last Updated: 2013-08-20 12:55:44 (180s ago) [refresh]

[Download: CSV | XML | JSON | KML](#)

### Significant Earthquakes

Past 30 Days

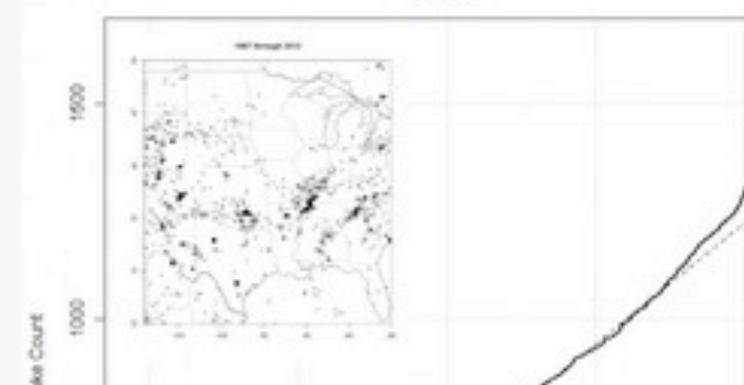
- |            |   |                         |              |
|------------|---|-------------------------|--------------|
| <b>6.6</b> | <a href="#">98km WSW of Mutis, Colombia</a>       | 2013-08-13 15:43:15 UTC | 12.0 km deep |
| <b>4.9</b> | <a href="#">9km SSW of Volcano, Hawaii</a>        | 2013-08-11 15:54:05 UTC | 31.8 km deep |
| <b>5.9</b> | <a href="#">13km E of Chabu, China</a>            | 2013-07-21 23:45:56 UTC | 9.8 km deep  |
| <b>6.5</b> | <a href="#">46km ESE of Blenheim, New Zealand</a> | 2013-07-21 05:09:31 UTC | 14.0 km deep |

[Significant Earthquake Archive](#)

### Featured Items



### [Man-Made Earthquakes](#)

Cumulative Number of Earthquakes  
 $M \geq 3$ 

View recent events or search for past earthquakes. Optimized for mobile and desktop.

[Real-time Feeds & Notifications](#)

Get real-time earthquake notifications sent to you using a number of popular mediums: Feeds, Email, Twitter, etc...

**Entrada: rede de sensores que cobre todo o mundo**  
**Saída: determinar quando um terremoto irá acontecer**

Shop by  
Department ▾

Search

Books ▾

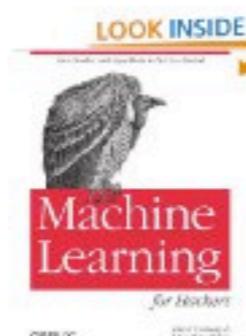
Go

Hello, **Fabrício**  
Your Account ▾

Your Amazon.com Your Browsing History Recommended For You Amazon Betterizer Improve Your Recommendations Your Profile Learn More

[Your Amazon.com](#) > [Recommended for You](#) > **Books****Just For Today**[Browse Recommended](#)**Recommendations**  
**Books**[Arts & Photography](#)[Audible Audiobooks](#)[Bargain Books](#)[Biographies & Memoirs](#)[Books on CD](#)[Business & Investing](#)[Calendars](#)[Children's Books](#)[Christian Books & Bibles](#)[Comics & Graphic Novels](#)[Computers & Technology](#)[Cookbooks, Food & Wine](#)[Crafts, Hobbies & Home](#)[Education & Reference](#)[Gay & Lesbian](#)[Health, Fitness & Dieting](#)[History](#)These recommendations are based on [items you own](#) and more.view: [All](#) | [New Releases](#) | [Coming Soon](#)

1.

**Machine Learning for Hackers**

by Drew Conway (February 22, 2012)

Average Customer Review: ★★★★☆ (21)

In Stock

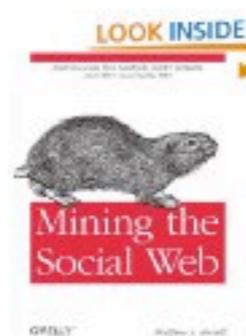
**List Price:** \$39.99**Price:** \$26.48[74 used & new from \\$21.90](#)

Add to Cart

Add to Wish List

 I own it Not interested Rate this itemRecommended because you added **Data Mining with R** to your Wishlist and more ([Fix this](#))

2.

**Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Networks**

by Matthew A. Russell (February 8, 2011)

Average Customer Review: ★★★★☆ (23)

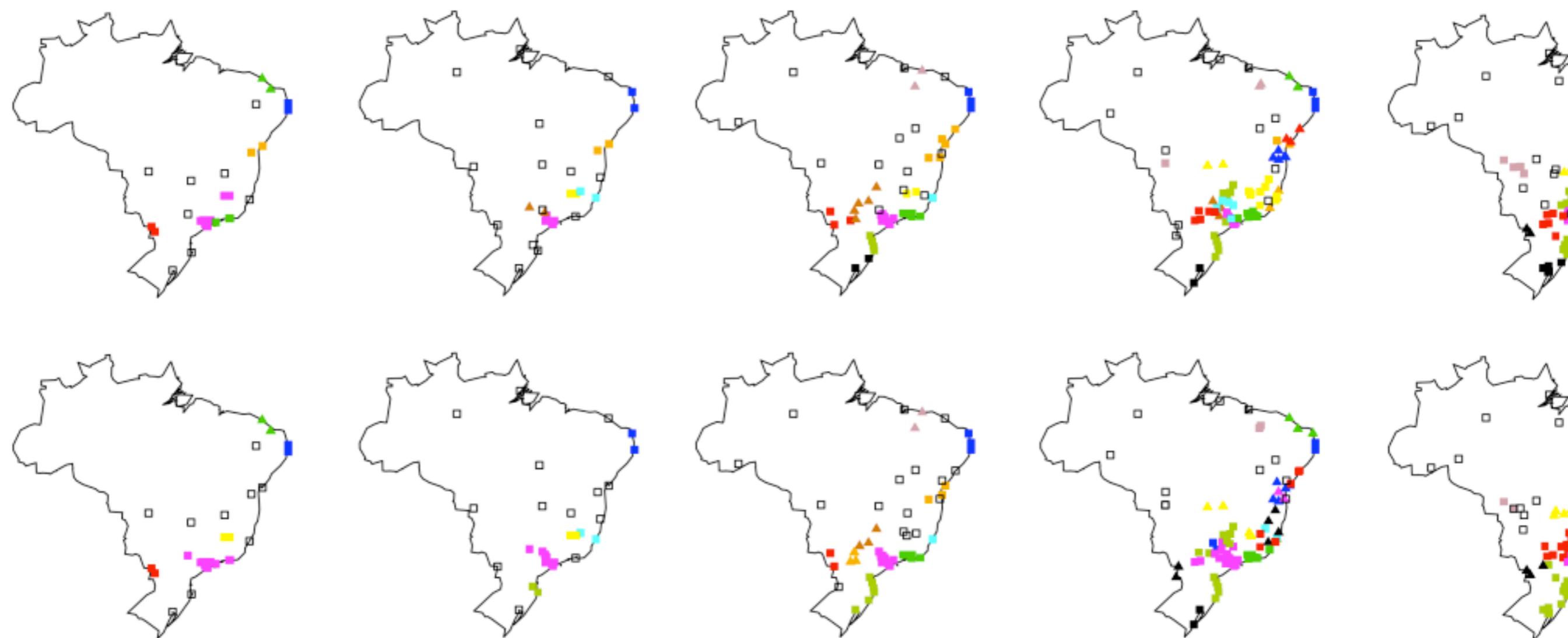
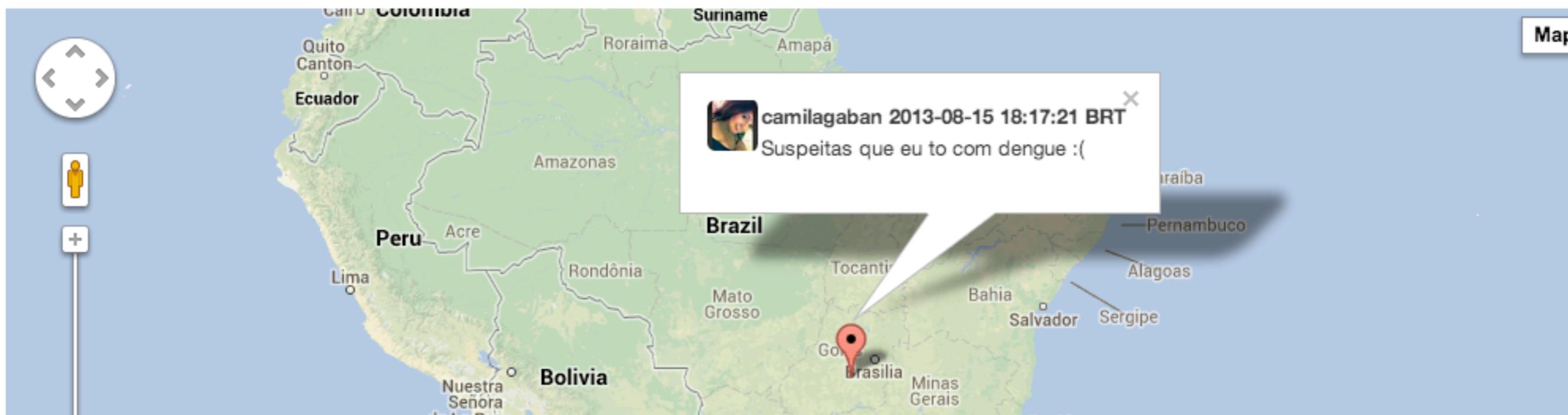
In Stock

**List Price:** \$39.99**Price:** \$26.48[84 used & new from \\$20.00](#)

Add to Cart

Add to Wish List

 I own it Not interested Rate this itemRecommended because you added **Data Mining with R** to your Wishlist and more ([Fix this](#))**Entrada: milhões de itens e milhões de usuários****Saída: recomendar itens com acurácia alta para os usuários (clientes)**





# Yelp Recruiting Competition

Finished

Wednesday, March 27, 2013

Jobs • 352 teams

Sunday, June 30, 2013

Dashboard

Home

Data

Make a submission

Information

Description

Evaluation



Competition Details » Get the Data » Make a submission

How many "useful" votes will a Yelp review receive? Show off your skills to land an interview for a position on a Yelp data mining team!

Português ▾

NETFLIX

1  
MÊS  
GRÁTIS

Assista a filmes e séries de TV quando quiser, onde quiser. Apenas R\$16,90 ao mês.

Comece a utilizar seu mês grátis

Arquivo Editar Exibir Histórico FAVORITOS Ferramentas Ajuda

http://www.opsys.com.br/index.php

Control panel - Web... Meus Mapas - Mind... Registro.br - Autent... Yahoo! Search BOS... Yahoo! Developer N... Technology Rev...

Diigo related:UOti-1rv3XgJ:schola

Bookmark Highlight Send Read Later Unread Recent Add

opSys! Mineração de Opinião e...

### Meus Sets

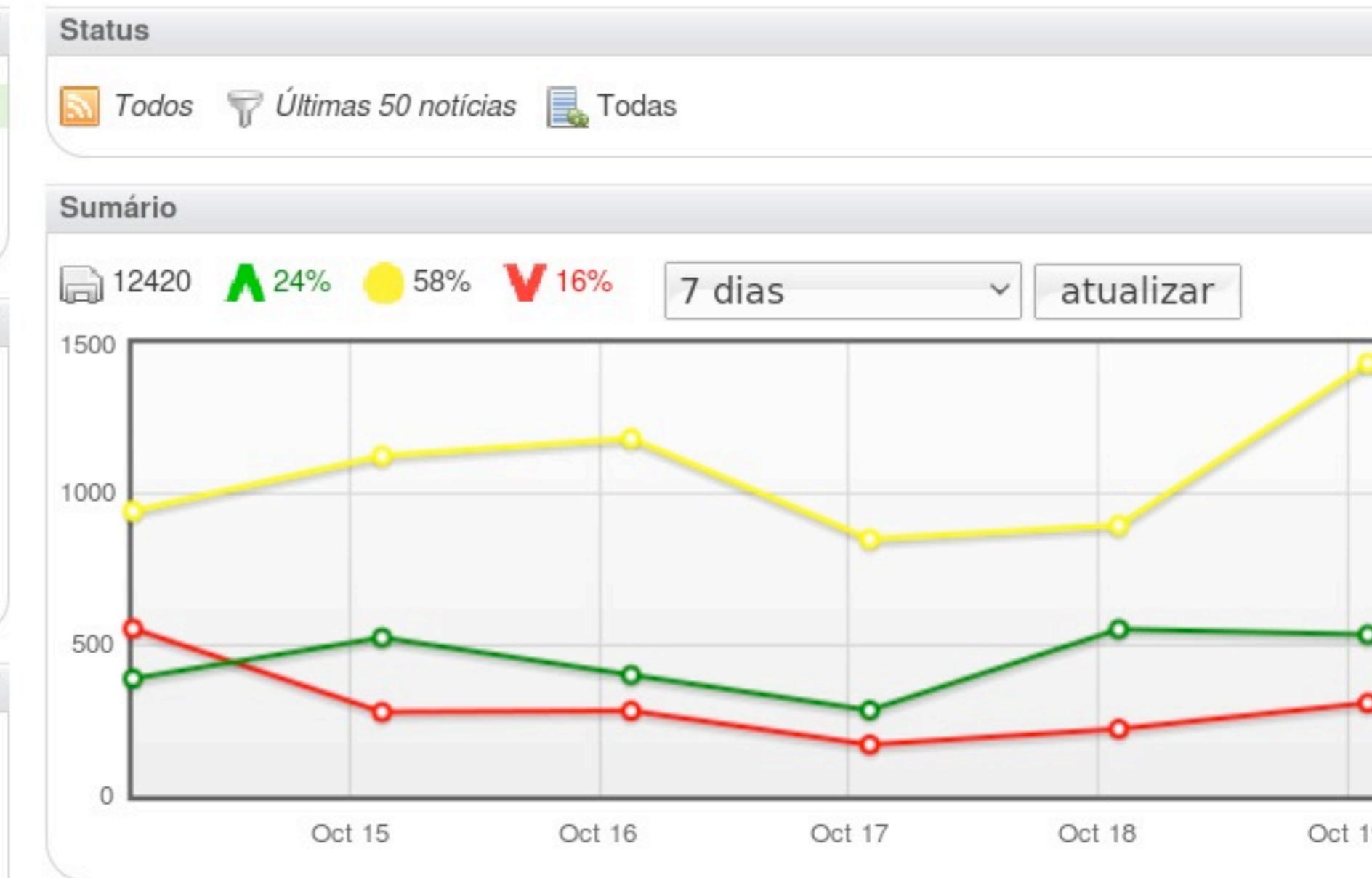
- Análise de Investimentos
- Eleições 2010
- Criar Novo

### Empresas / Entidades

- BRADESCO
- GERDAU
- PETROBRAS
- USIMINAS
- VALE R DOCE

### Feeds

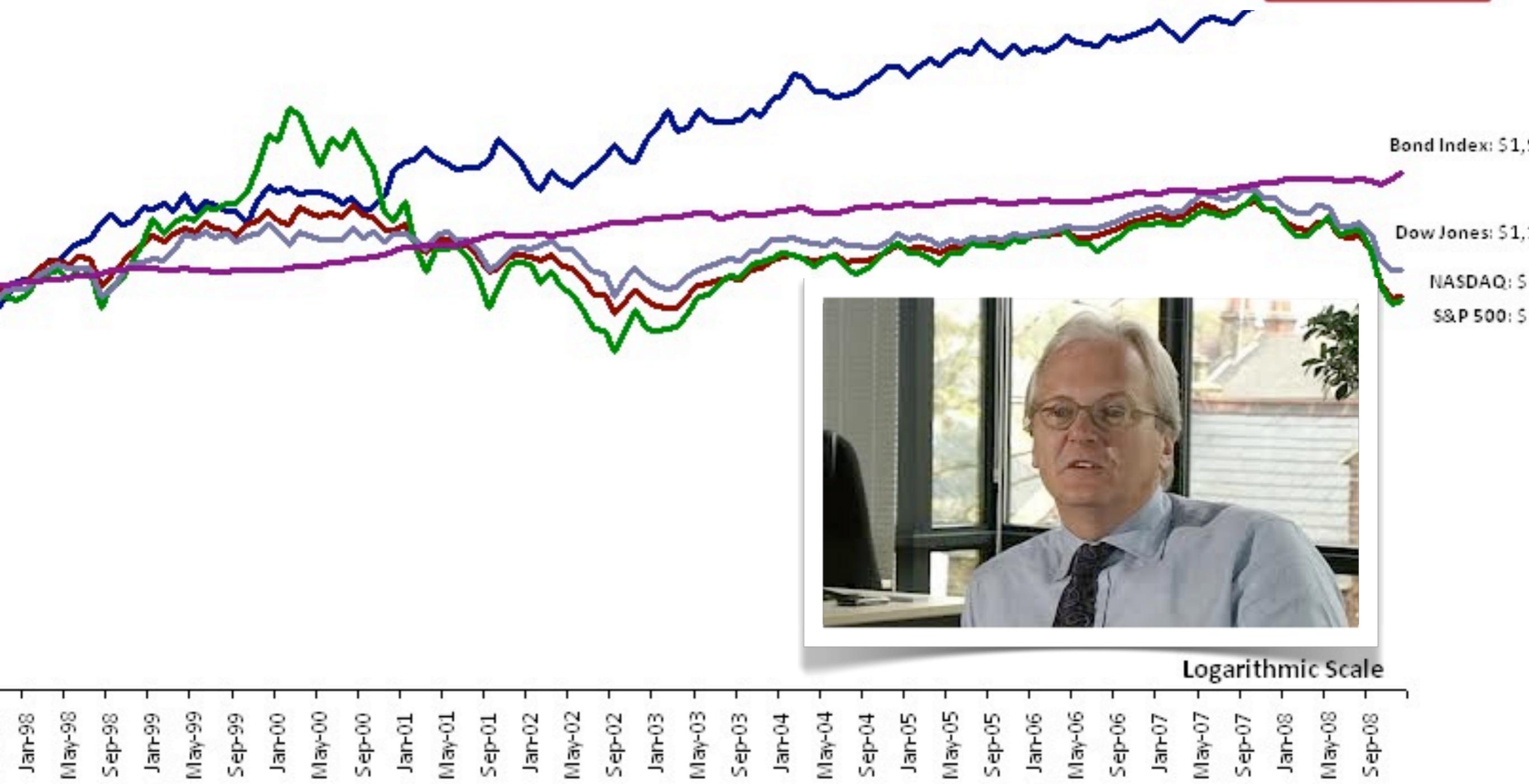
- Abril.com - Economia
- BBCBrasil.com | Tópicos | Economia
- Direto do Pregão
- estadao.com.br - Últimas notícias
- Folha Online - Dinheiro - Principal
- G1 Economia e Negócios
- Google News - Detalhe



### Notícias filtradas

Data/Hora	Titulo
-----------	--------

# Criar rôbos que compram e vendem ações



# O que estes projetos têm em comum?

---

- \* **Manipulam grandes volumes de informação**
- \* Outros exemplos de grandes volumes de informação:
  - \* A380: Heathrow - JFK: 640 TBs de log
  - \* Twitter: 12+ TBs of tweet every day
  - \* Facebook: 25+ TBs of log data every day

# O que estes projetos têm em comum?

---

- \* A origem dos dados é muito variada.



Mobile Sensors



Social Media



Video Surveillance

Video Rendering



READING METERS  
EVERY 15 MINS.  
IS 3,000X MORE  
DATA INTENSIVE



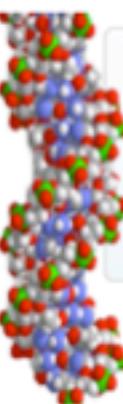
Smart Grids



Geophysical  
Exploration



Medical Imaging



Gene Sequencing

COST TO SEQUENCE  
ONE GENOME  
HAS FALLEN FROM  
\$100M IN 2001  
TO \$10K IN 2011

---

O que estes  
projetos têm  
em comum?





**Queremos modelos predictivos**

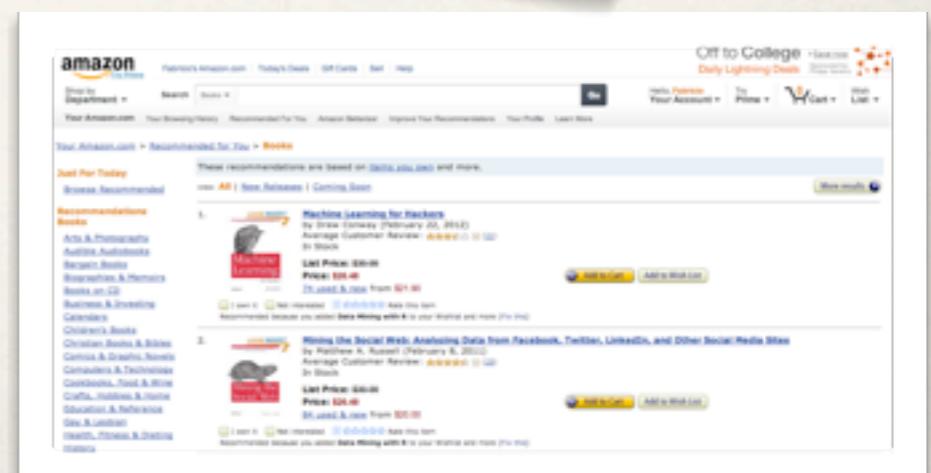
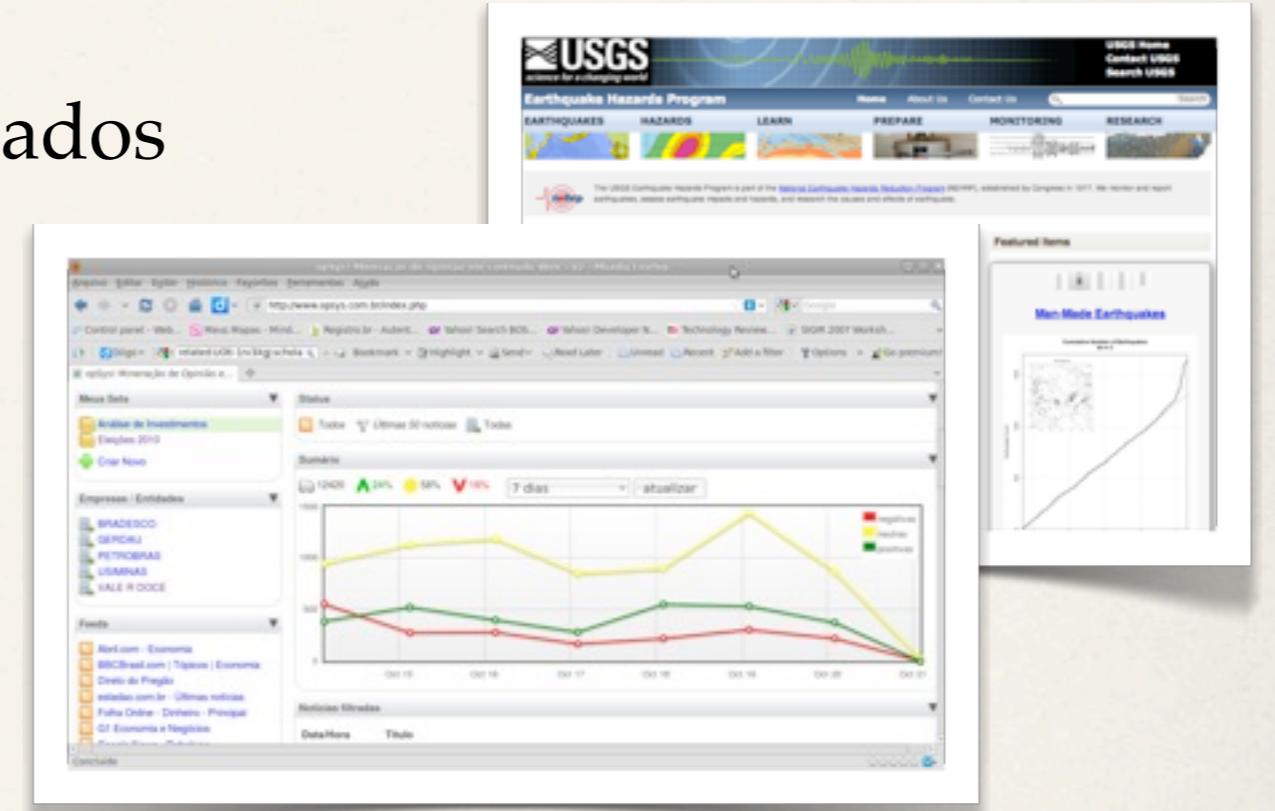
# Big Data refers to:

---

- \* All data that comes at high **Volume**
- \* All data that comes at high **Velocity**
- \* All data that comes from a **Variety of Sources** (structured + unstructured data)
- \* **Future**

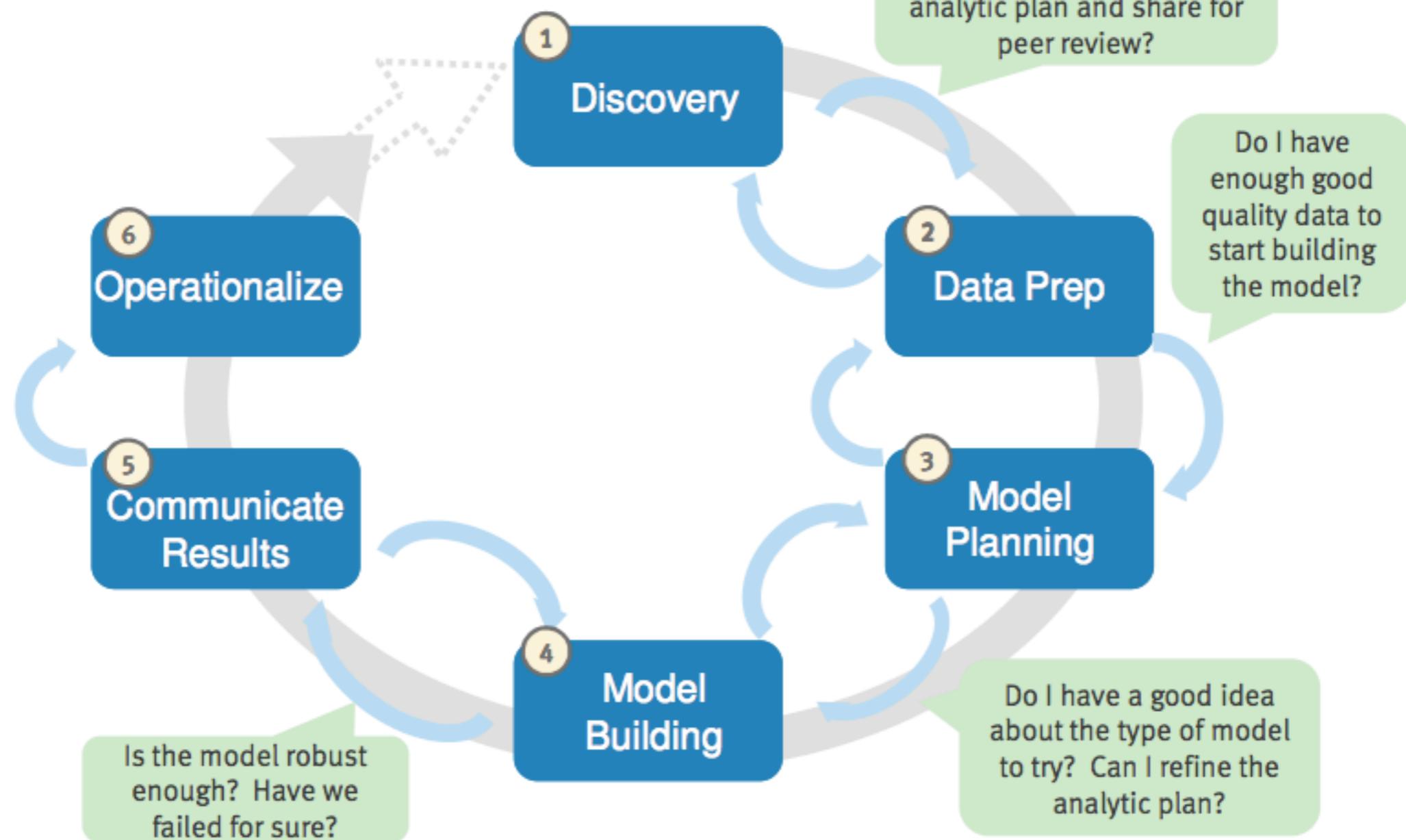
# Fluxo de transformação dos dados

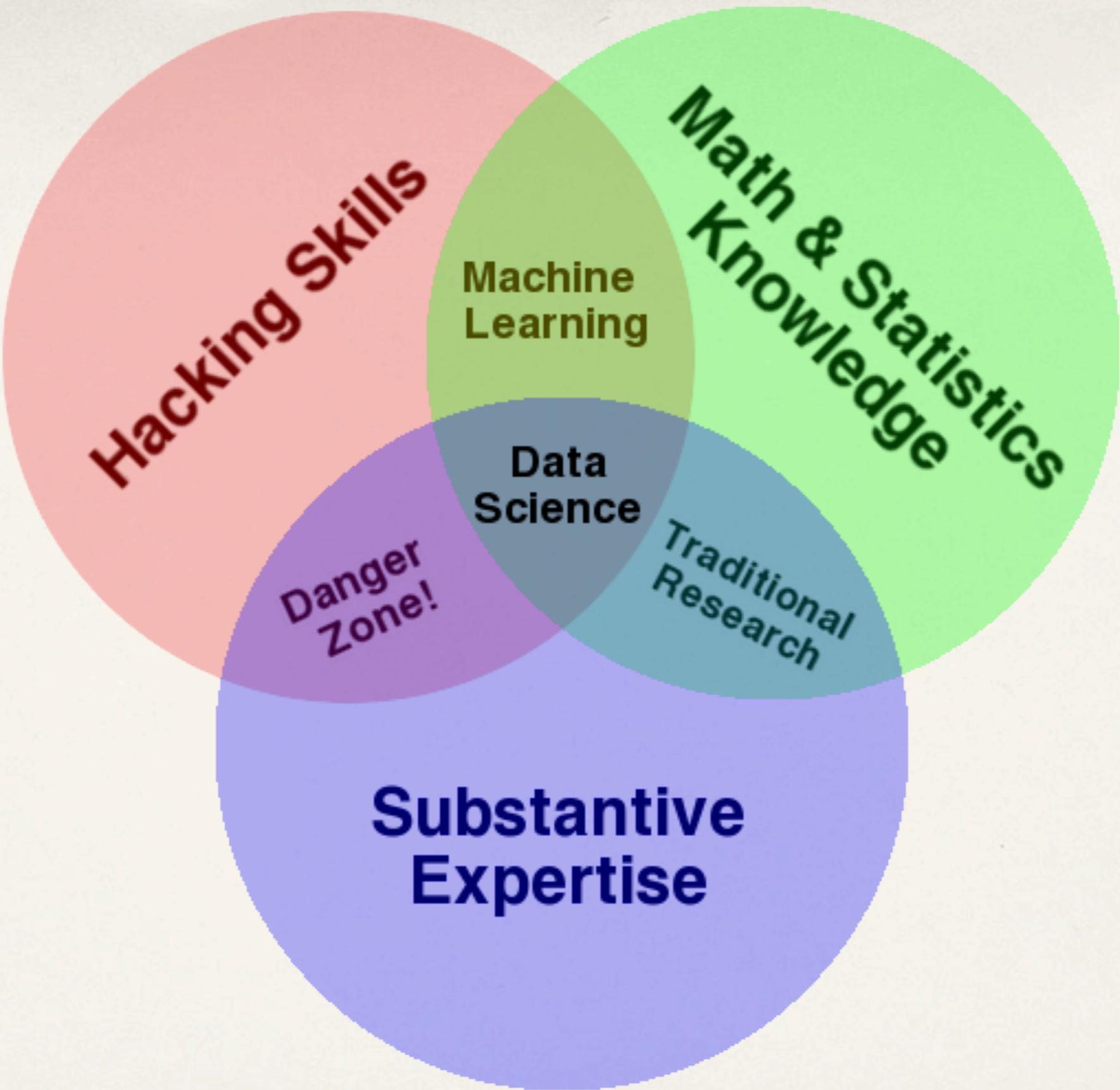
- **Coleta** e pré-processamento dos dados
- **Aplicação** do modelo
- **Apresentação** dos resultados



# Como o modelo é desenvolvido?

## Data Analytics Lifecycle





---

Habilidades de um cientista de dados