

# Análise e Sistematização de Big Data

Leonardo Sangali Barone

April 20, 2017

# Análise e Sistematização de Big Data

O objetivo do curso, tal como construímos ao longo dos 4 dias, foi fazer um breve introdução ao R com foco e manipulação de dados e apresentar diversos tópicos relacionados ao termo pouco preciso, porém popular, “Big Data”.

# Big Data é um termo impreciso

Uma das grandes dificuldades do curso é lidar com o fato de que há muita coisa sob o guarda-chuva de Big Data. Quando falamos de Big Data, estamos falando alternadamente de:

- ▶ Ciência de Dados (roupagem nova para estatística)
- ▶ Armazenamento de dados (que é um dos problemas da computação desde o princípio)
- ▶ Captura de dados (que é um tema novo)
- ▶ Aprendizado de Máquina (que é um uso novo para ferramentas estatísticas velhas)

## O plano inicial do curso

Se tivéssemos que mudar o nome do curso, o mais adequado seria “R para análise de dados e Big Data”. O plano inicial do curso, antes de conhecer vocês, era trabalhar:

- ▶ Uma introdução ao R e à gramática *dplyr*
- ▶ Webscrapping
- ▶ Aprendizado de máquina

# O curso como termina

Conhecendo vocês e ampliando o material do curso, terminamos com os seguintes tópicos

- ▶ Uma introdução ao R e à gramática *dplyr* (ampliado)
- ▶ Webscrapping
- ▶ RevoScaleR e Microsoft R Client
- ▶ Aprendizado de máquina
- ▶ Dados volumosos e integração R e SGBD
- ▶ Aprendizado de máquina com Spark

# O prometido não entregue

Faltou entregarmos um tópico menos importante (tanto como solução quanto para vocês):

- ▶ Bibliotecas básicas do R para dados volumosos

B