



# Reddit viewer



Equipe:  
Ana França  
Édipo Andersen  
Rodrigo Farias

# Análises - O que queremos descobrir?

- Semânticas

- Que categorias são mais discutidas?
- Como é a distribuição de tópicos por categoria?
- Com que frequência um usuário se liga a vários subfóruns?
- Como os interesses dos usuários estão relacionados?
- Quais são as palavras chave em determinadas categorias?

} Treemap

} Force Directed Layout

- Temporais

- Como as palavras chave variam ao longo do tempo?
- Como a popularidade das categorias varia ao longo do tempo?
- Em que momentos as pessoas escrevem mais postagens?
- Como varia a intensidade das discussões nas categorias ao longo do tempo?

Word Cloud + Treemap  
+ StackedGraph

} StackedGraph  
+ Histograms

# Estrutura Técnica

---

# Partes Principais

---

- Ferramental
  - Funções básicas genéricas Ex:
  - GetKeyWords(text) - remove stopwords em @text
  - CountKeyWords(text, keywords) - retorna o mapeamento keyword->#aparições do texto
- Pré-Processamento
  - Funções para iterar em toda a base de dados, extraíndo informações de mais alto nível Ex:
  - CountCommonAuthors(subredditA, subredditB) - retorna a contagem de autores que publicam no subredditA e no subredditB
  - CalculateSubredditPopularity(subreddit) - varre os dados contando as threads únicas
- Visualização (on-the-fly)
  - Funções e classes relacionadas com a visualização, tendo como base os dados pré-processados contendo apenas as informações relevantes para cada visualização Ex:
  - TreeMap - Precisa apenas das informações de popularidade e keywords de cada subreddit

# Visualizações

---

# Tree Map

---

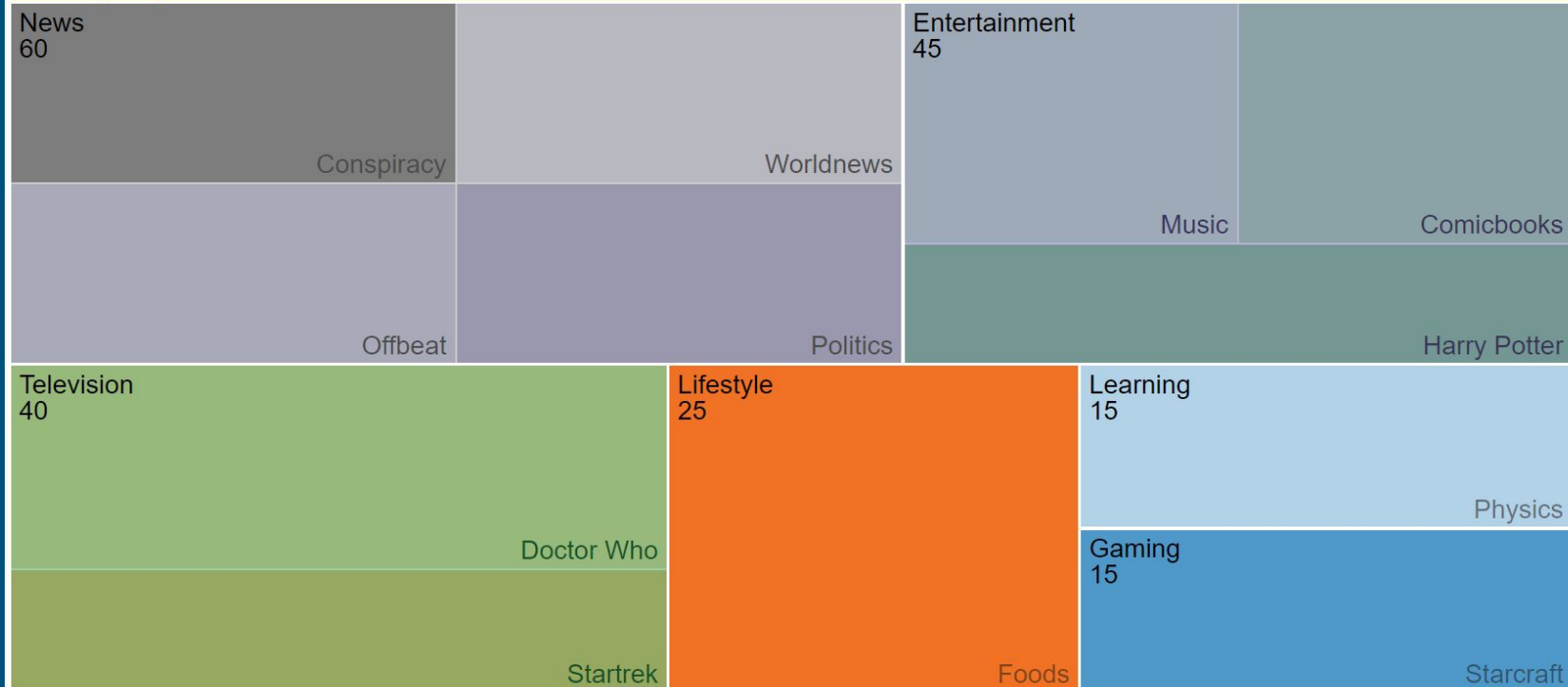
- Dois níveis de hierarquia: Meta e Subreddit
- Área significa a quantidade de tópicos
- Aplicar zoom
- Atributos necessários:

Meta	Subreddit	Text
------	-----------	------

# Tree Map

## Reddit Population

### World (200)



# Tree Map

## Reddit Population

World (200) / Learning (15)

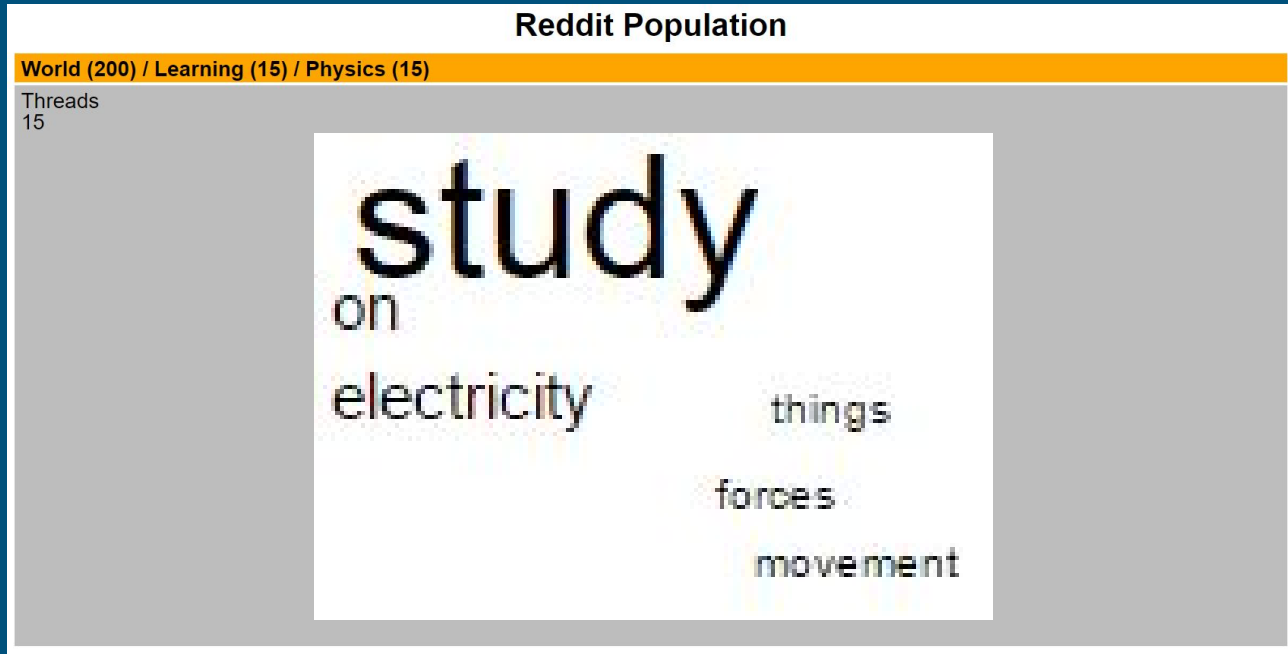
Physics  
15

Threads



# Word Cloud

---



# Histograma de Tempo

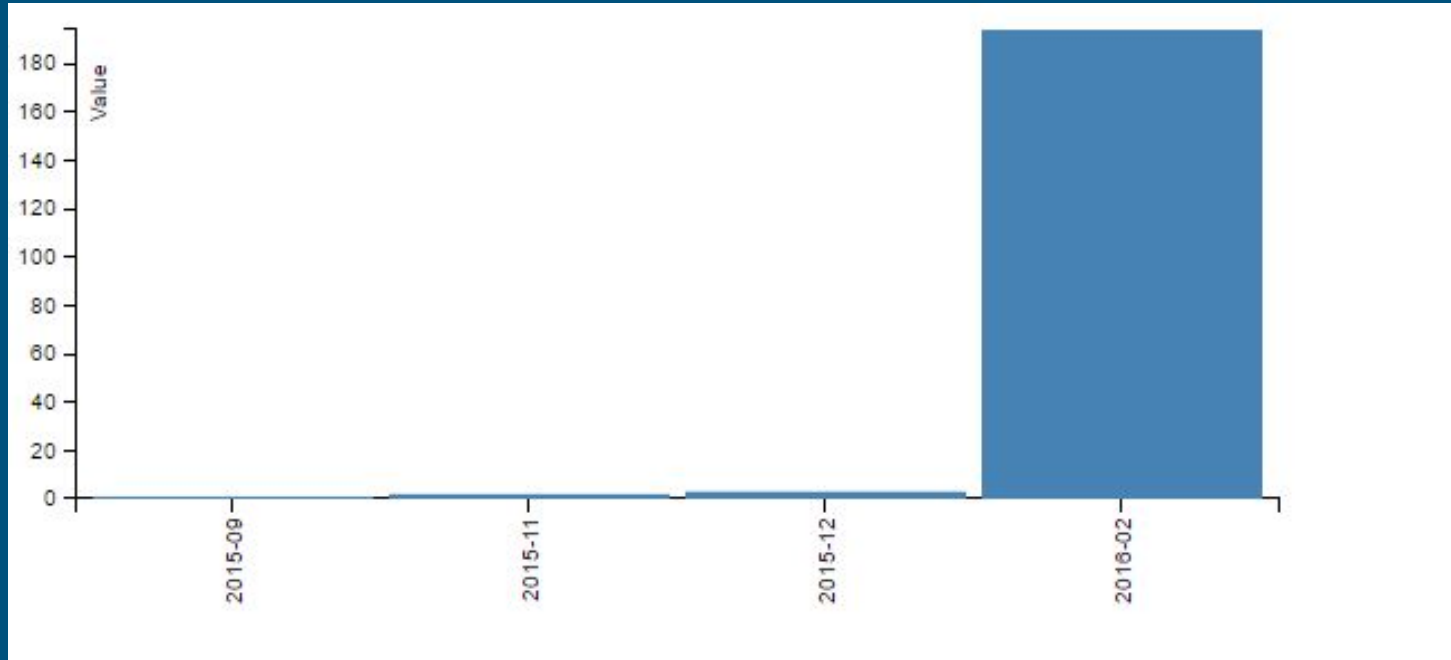
---

- Horário das postagens distribuídos
- Dividir em 3 Hora do dia/Dia da semana/Mês do Ano
- Paleta de cores configuráveis.
- Filtro para diferentes comunidades
- Atributos necessários:

Time	Subreddit
------	-----------

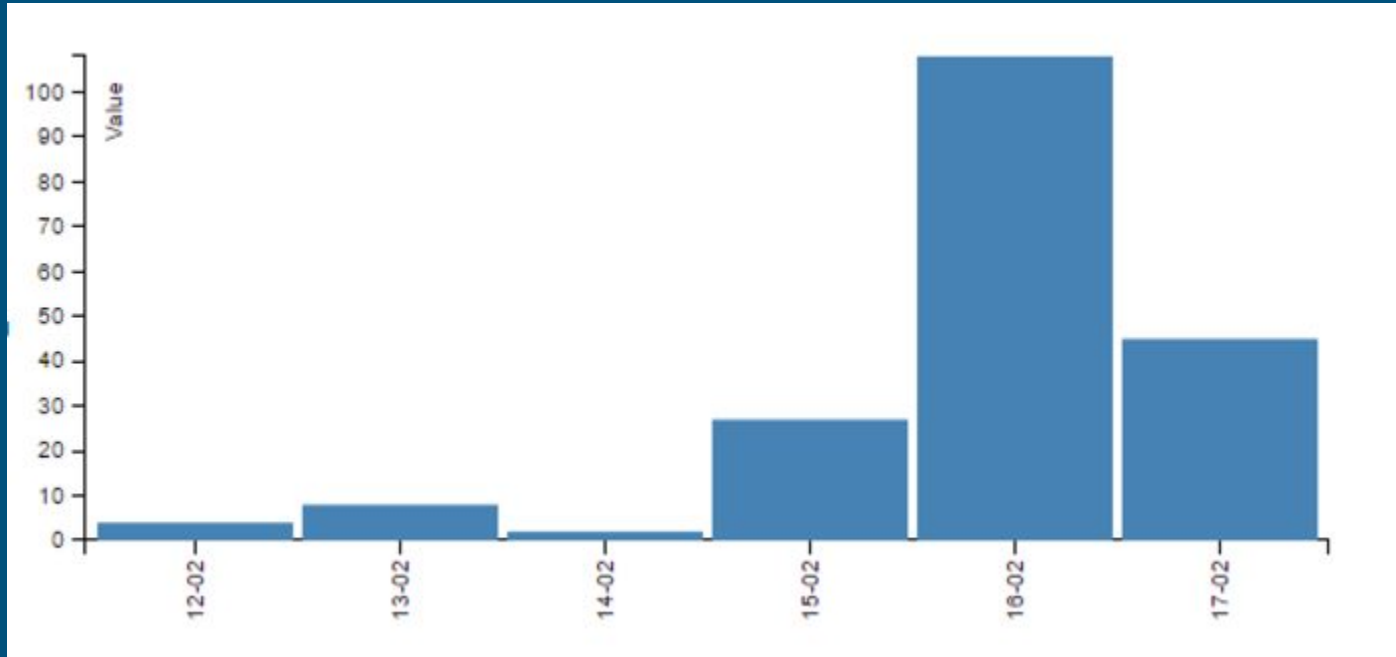
- Data mínima: 23/05/2014
- Data máxima: 17/02/2016

# Histograma de Tempo(ano/mês)



# Histograma de Tempo(dia/mês)

---



# Force Directed Layout

- Cada nó representa um *subreddit* (51) e os links representam as relações entre os subreddits
- Atributos necessários:

Subreddit	Author
-----------	--------



# Tiara/Streamgraph

- 7 faixas, 1 para cada Meta.
- Atributos necessários:

Text	Meta	Time	Ups
------	------	------	-----

Referências:

Básico, Interativo, Nuvem de palavras