

# Análise de dados: sites de divulgação de Ciência

Prof. Caio Azevedo

# Introdução

- O interesse da pesquisadora é avaliar diversas características de sites (considerados) como de divulgação de Ciência.
- Objetivo da parte quantitativa do projeto: compreender quais são os fatores responsáveis pela popularidade (calculada como número de visualizações por dia) de vídeos de ciência do YouTube (modelo de regressão).
- Interesse: sites vídeos produzidos pelo projeto ScienceVlogs Brasil (<http://scienceblogs.com.br/sciencevlogs/>).
- Atualmente, 38 canais fazem parte deste grupo; neste trabalho, por razões relacionadas ao conceito de divulgação científica usado, ela considerou apenas 33 canais.

# Variáveis de interesse

- Utilizou-se a categorização de Borghol et al (2012) para a criação de dois tipos de variáveis: aquelas relacionadas ao conteúdo (content-related) e as não-relacionadas ao conteúdo (content-agnostic).
- Quanto às relacionadas ao conteúdo, decidiu-se pelas seguintes: TEMA DO VÍDEO, TAMANHO DO VÍDEO, FORMATO DO VÍDEO, RECURSOS DE EDIÇÃO.
- Quanto às variáveis não relacionadas ao conteúdo, utilizou-se as seguintes: PRODUTIVIDADE DO CANAL, IDADE DO CANAL, LIKES, COMENTÁRIOS.

# Variáveis relacionadas ao conteúdo

- TEMA DO VÍDEO, de acordo com a classificação de oito áreas de pesquisa proposta pela Fapesp (Ciências Exatas e da Terra; Ciências Biológicas; Engenharias; Ciências da Saúde; Ciências Agrárias; Ciências Sociais Aplicadas; Ciências Humanas; Linguística, Letras e Artes). A cada vídeo, foi atribuído apenas um tema. Acrescentamos ainda a área Interdisciplinar, na qual classificamos vídeos muito híbridos para serem colocados em apenas uma categoria.
- TAMANHO DO VÍDEO, em minutos.

# Variáveis relacionadas ao conteúdo

- **FORMATO DO VÍDEO**, que pode ser um em oito. A cada um dos formatos atribuímos um número, que usamos a título de etiqueta para facilitar a referência. Aqui, eles aparecem entre parênteses. São eles vlog (1,) entrevista (2), documentário ou reportagem (3), hangout (conversa em grupo transmitida via internet 4), animação (5), conversa em grupo ao vivo (6), vídeo comentado (7) e palestra (8).
- A escolha por estes formatos foi feita após uma pré-análise dos vídeos presentes no ScienceVlogs Brasil, feita muito antes do início da coleta de dados.

# Variáveis relacionadas ao conteúdo

- RECURSOS DE EDIÇÃO, selecionados com inspiração no trabalho de Morcillo et al (2016). Consideramos cada um dos nove recursos como um ponto, e conforme os recursos foram aparecendo no vídeo, somamos um ponto para ele. Cada vídeo, portanto, poderia ter entre zero e dez pontos nesta variável. São os recursos:
  - Uso de trilha sonora ou efeitos sonoros;
  - uso de imagens, textos ou gráficos no vídeo (legendas não contam);
  - uso de excertos de outros vídeos;
  - exibição de logo ou vinheta no vídeo (consideramos apenas os logos em evidência, não os discretos, no canto do vídeo);
  - uso de filtros;

# Variáveis relacionadas ao conteúdo

- RECURSOS DE EDIÇÃO (cont.):
  - uso da técnica fast-forward (vídeo rodado a um ritmo mais acelerado);
  - uso da técnica jump cut (cortes que removem parte de uma mesma tomada, gerando transições bruscas);
  - uso da técnica quatro-a-quadro (stop motion);
  - uso de chroma key (green screen)

# Variáveis não relacionadas ao conteúdo

- **PRODUTIVIDADE DO CANAL**, calculada como o número de vídeos no canal dividido pelo número de meses de existência do canal, resultando em um
- índice de vídeos por mês. Consideramos a data inicial de inauguração do canal a partir de sua primeira publicação.
- **IDADE DO CANAL**, calculada em número de meses de existência do canal (modo de cálculo explicado no item acima).
- **LIKES** dados para cada vídeo (resposta).
- **COMENTÁRIOS** feitos em cada vídeo.



# Análise/planejamento amostral

- A avaliação de algumas características do canal foi feito a partir da pesquisadores, que assistia à tais vídeos na íntegra.
- Considere que o interesse, neste momento, seja apenas de estimar alguma quantidade relativa à variável “likes’. Por exemplo, o número médio de likes por dias em que o vídeo está no ar.
- Como proceder?
- Unidade elementar?
- O que os canais representam?

# Análise/planejamento amostral

- O que os canais representam?
- Nenhuma das (co)variáveis é conhecida em nível populacional.
- De acordo com a pesquisadora, os vídeos, dentro de cada canal, tendem a ser mais parecidos do que vídeos de diferentes canais.
- Plano amostral?

# Referências

- Borghol, Y., Ardon, S., Carlsson, N., Eager, D., Mahanti, A. (2013) The Untold Story of the Clones: Content-agnostic Factors that Impact YouTube Video Popularity, eprint arXiv:1311.6526, <https://arxiv.org/abs/1311.6526>.
- Covington, P., Adams, J., Sargin, E. (2016) Deep Neural Networks for YouTube Recommendations, Proceedings of the 10th ACM Conference on Recommender Systems, , <https://ai.google/research/pubs/pub45530>.

# Referências

- Morcillo, J. M., Czurda, K., Trotha, R.-von (2016). Typologies of the popular science web video, *Journal of Science Communication*, 15, 4, 1-32.
- Szabo, G., Huberma, E. A. (2010) Predicting the Popularity of Online Content, *Communications of the ACM*, 53, 8, 80-88.
- Welbourne, D. J., Grant, W. J. (2015) Science communication on YouTube: Factors that affect channel and video popularity, *Public Understanding of Science* 25, 6, 706-718.