

Universidade Estadual Paulista "Júlio de Mesquita Filho"

Trabalho final de Ciência de dados - Grupo 3

Leonardo Oliveira
Kawe Antônio dos Santos Marcelino
Pedro Henrique Nunes Barros
Rafael Nunes Caseiro
Vitor de Souza Cruzeiro

Bauru
2018

Introdução

Conforme o passar dos últimos anos, a quantidade de dados aumentou de forma gigantesca. Qualquer dispositivo eletroeletrônico, capaz de ter conexão com a internet, gera ou pode gerar dados. Porém, os dados gerados, por si sós, não têm uma finalidade correta caso não sejam formatados de forma a se tornarem informação, com a qual é possível identificar problemas ou ter uma visão melhorada da situação que gerou aqueles dados. Os dados não estruturados alcançam escalas empresariais. Os dados estão em abundância em todos os lugares. Projeta-se que, em 2020, cada pessoa gerará cerca 1.7MB de dados por segundo. Colocando isto numa escala global e em um determinado tempo, essa quantidade de dados mostra-se difícil de ser manejada. Então surgiu a ciência de dados, que é uma área interdisciplinar voltada para o estudo e a análise de dados estruturados ou não. A ciência de dados visa a extração de conhecimento (*insights*) para tomadas de decisão; compreende tudo relacionado à preparação, limpeza e análise de dados (slides do professor).

Sobre o trabalho

Este relatório foi realizado como parte do trabalho final da disciplina de Ciências de Dados, do segundo termo de 2018. Ele visa demonstrar um conjunto de técnicas exploratórias da linguagem R, aprendidas durante o curso, assim como a elaboração de grafos e documentos através do software R-Studio, e o próprio relatório, em conjunto com uma apresentação de slides, produzidos pelo software R-Markdown. Ele contempla a pesquisa realizada pelos alunos da disciplina, aplicada ao corpo discente da Universidade Estadual Paulista “Julio de Mesquita Filho” - UNESP - do campus de Bauru – SP, como parte da pesquisa em progresso de autoria do professor João Pedro Albino, com participação de 61 alunos, com o intuito de responder a pergunta:

*“Quais são as influências das mídias sociais sobre discentes da instituição de nível superior
UNESP-Bauru-SP?”*

Análise exploratória de dados

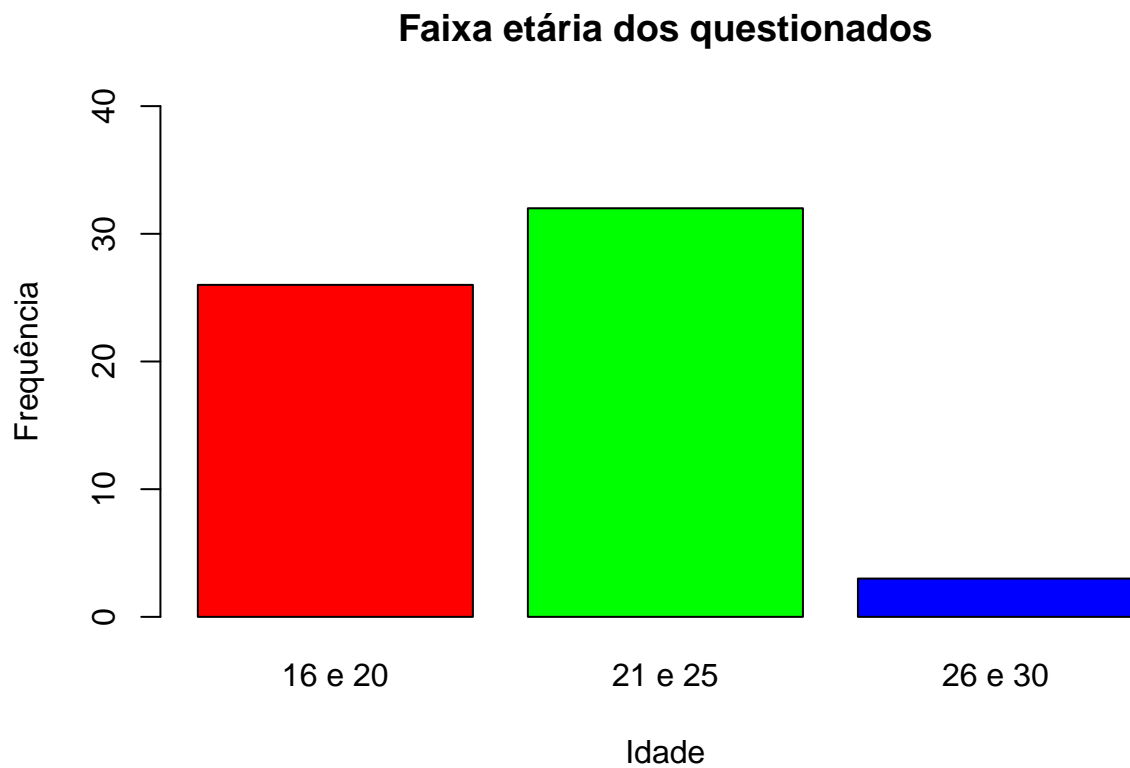
Utilizando a planilha *tidy* fornecida pelo professor, foram analisados todos os dados das 15 questões apresentadas nos formulários coletados. A análise foi dividida em primeira (análises isoladas das questões) e segunda grau (cruzamentos de dados entre questões para gerar informações mais complexas).

Análise de primeira ordem

A partir dos dados coletados pela pesquisa foi gerado um conjunto de gráficos exploratórios de primeira ordem. Este conjunto apresenta os resultados de forma simples e direta, servindo como embasamento inicial da análise, facilitando a proposta de perguntas mais interessantes sobre o conjunto de dados. Ainda assim, a análise destes gráficos já aponta algumas características interessantes do corpo de pesquisa, permitindo a observação de aspectos prevalentes e possivelmente inesperados. Os gráficos gerados foram os seguintes:

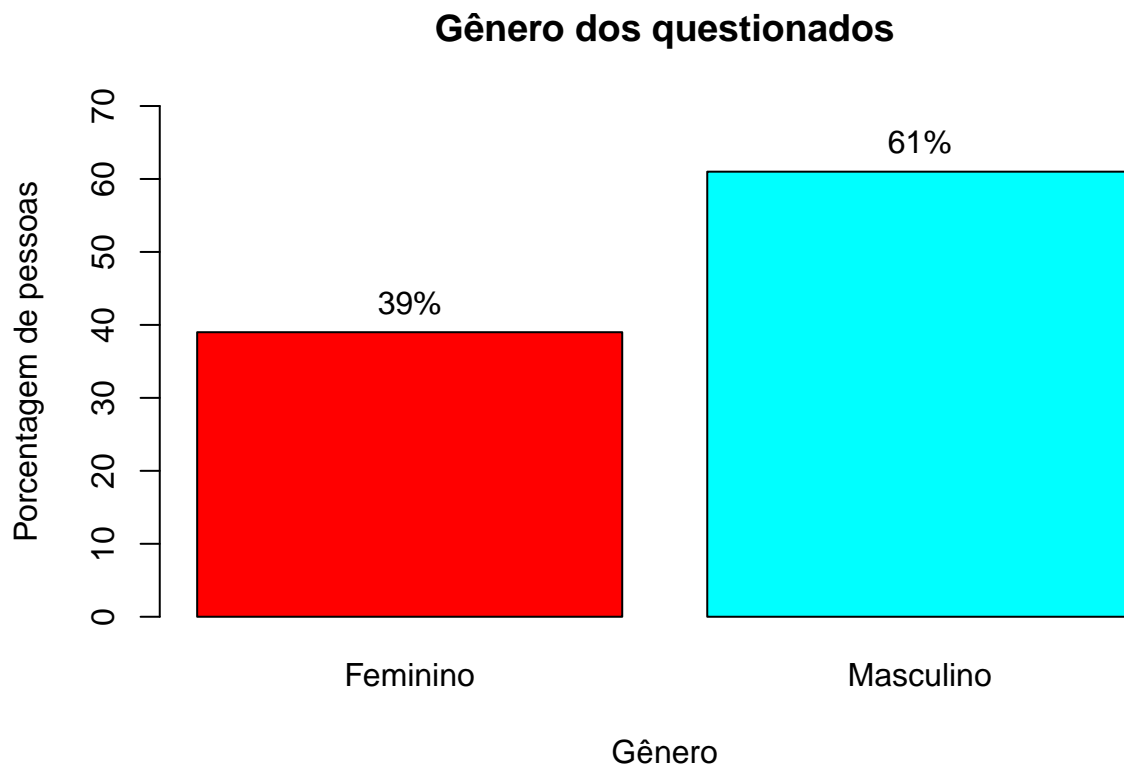
Faixa etária

O gráfico a seguir apresenta a distribuição de faixa etária dos entrevistados. Percebe-se que a maioria dos entrevistados tem entre 21 e 25 anos:



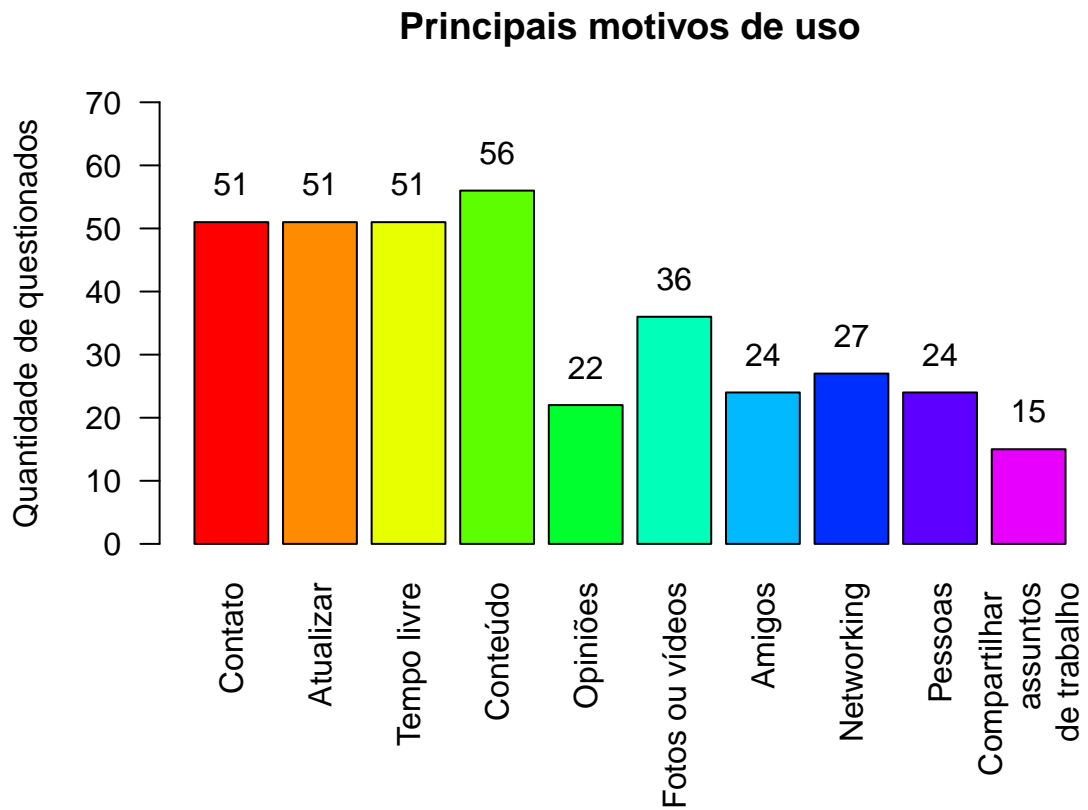
Sexo

Uma vez que a maioria dos respondentes eram do curso de Ciência da Computação ou de Sistemas de Informação da Unesp de Bauru, é normal observar um número maior de homens, como mostra o gráfico abaixo:



Principais motivos de uso das redes sociais

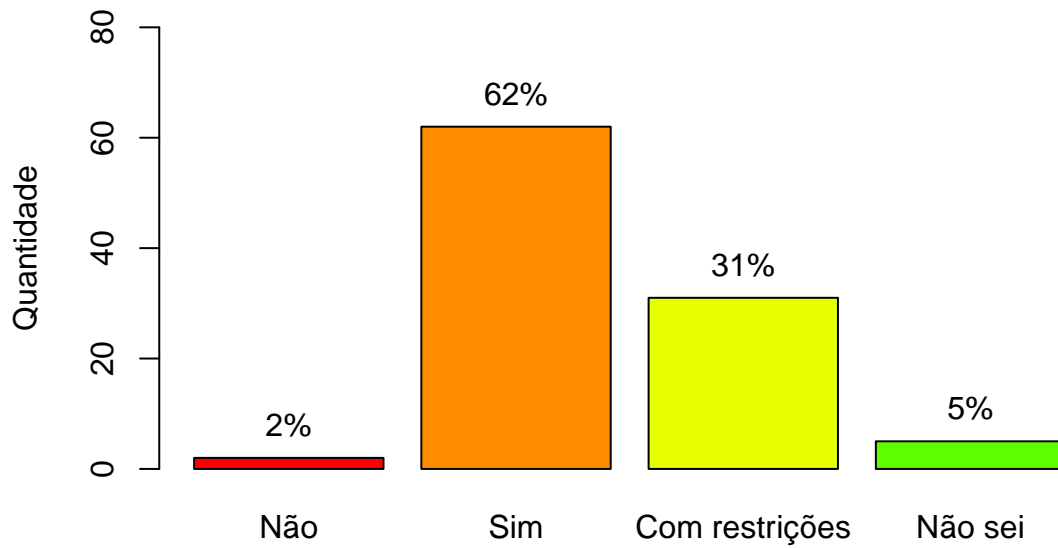
Através da análise deste dado, notou-se uma diversificação no que tange os motivos pelos quais os usuários utilizam as redes sociais.



Uso de mídias sociais por professores

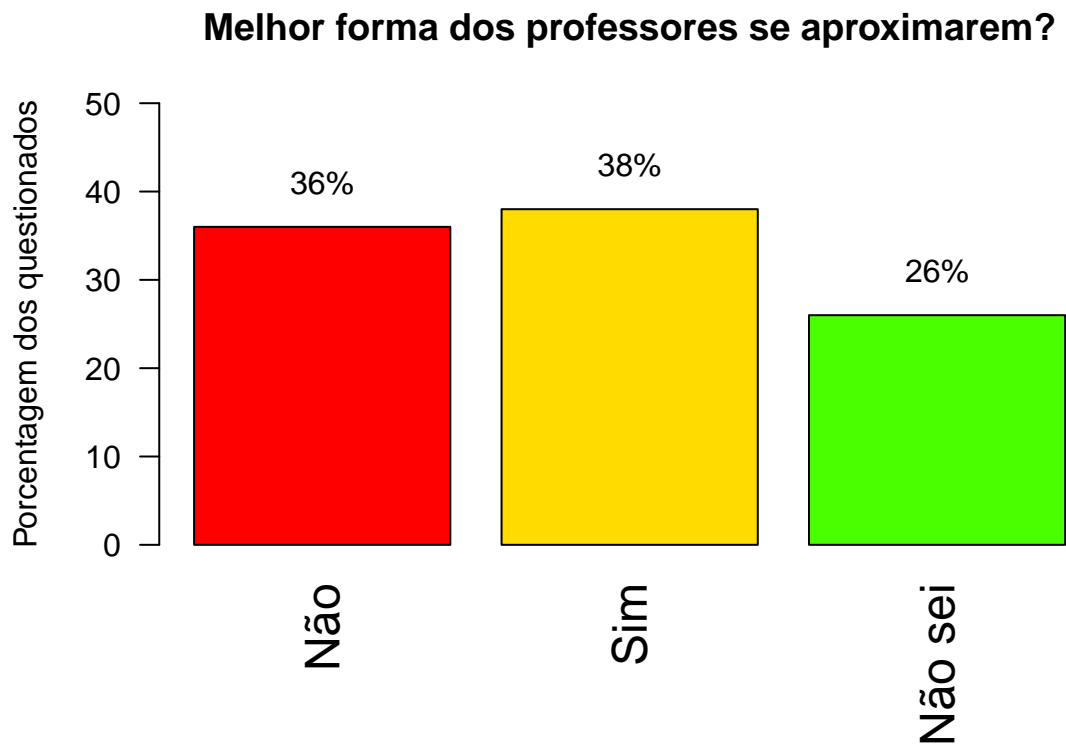
Pelo menos 93% dos entrevistados concorda que as mídias sociais podem ser utilizadas por professores, mesmo que com ressalvas. Uma população muito pequena discorda da ideia e não representa um conjunto significativo para futuras análises.

Mídia social pode ser usada pelos professores?



Mídias sociais como a melhor forma de aproximação entre professor e aluno

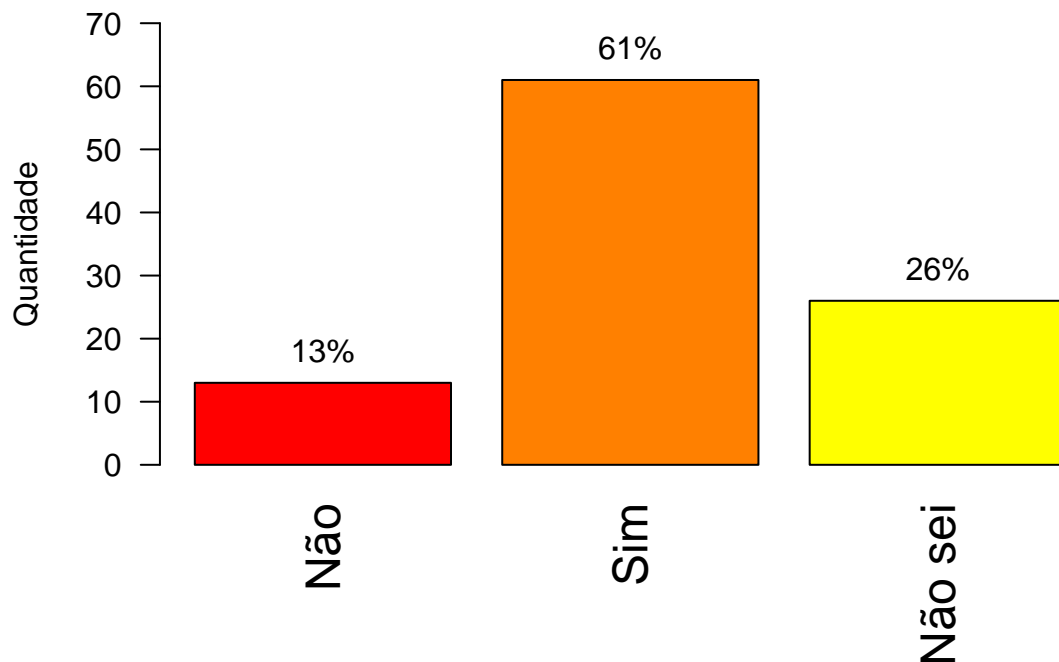
O gráfico apresenta as respostas à pergunta *"Você acredita que a mídia social é a melhor forma dos professores se aproximarem de seus alunos?"*. Notou-se que, apesar de a maioria dos respondentes concordarem com o uso das mídias sociais na educação, ainda há muita dúvida quanto a sua real eficiência, uma vez que poucos responderam "Sim" à pergunta e muitos ficaram indecisos ou discordaram.



Mídias sociais com potencial de melhora no desempenho dos alunos

Ainda seguindo a tendência da questão anterior (que mostra que as mídias sociais ainda são um pouco “obscuras” para uso na educação), o índice de abstenção ficou próximo de um quarto dos entrevistados, mas os que concordam com a afirmação de que estes mecanismos podem melhorar o desempenho dos alunos permaneceu alto.

Melhores resultados com a integração das redes sociais



Análise de segunda ordem

Enquanto que a projeção individual de resultados gera gráficos mais simples e diretos, ideais para uma primeira análise dos resultados, uma exploração mais profunda exige a elaboração de gráficos de segunda ordem, isto é, a projeção cruzada entre variáveis. Esse tipo de análise expõe correlações entre dados que dificilmente seriam percebidas de outra forma. São através destes gráficos que os aspectos mais interessantes e surpreendentes da pesquisa realizada expõem-se. Portanto, foram selecionadas diversas variáveis do corpo de dados, e, a partir destas, foi projetado um conjunto de gráficos de segunda ordem. Ao analisar estes gráficos pode-se notar que alguns não parecem transmitir nenhuma correlação relevante, com os dados cruzados se relacionando de forma dispersa. Estes gráficos não foram inclusos neste relatório. Por outro lado, as demais projeções puderam ser destacadas em dois grupos: aquelas que indicam correlação entre dados de forma a fortalecer seus resultados e outras que se relacionam de forma contraditória aos próprios resultados. Este segundo grupo em particular despertou grande interesse, pois pode ser usado para apontar aspectos da pesquisa que podem não ter sido bem interpretados pelos entrevistados, e que, portanto, possam, talvez, ser melhor elaborados. Seguem os resultados obtidos:

Resultados

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque ut augue pretium, luctus erat vitae, sollicitudin eros. Aliquam erat volutpat. Nulla facilisi. Vivamus ipsum ipsum, finibus vel venenatis non, tincidunt sit amet turpis. Aliquam aliquam purus vitae eros aliquet ullamcorper. Fusce imperdiet ut nisl tempus iaculis. Nunc vehicula, erat id vestibulum tincidunt, diam risus ullamcorper odio, quis tincidunt sapien libero sed nibh. Vivamus in velit diam. Ut nulla erat, eleifend et luctus nec, convallis id ligula. Aliquam gravida quis odio at facilisis. Suspendisse semper venenatis est, quis convallis nunc tincidunt et. Nam egestas sapien non nulla luctus iaculis. Donec vel ipsum id metus consequat aliquam. Interdum et malesuada fames ac ante ipsum primis in faucibus.