

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 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:

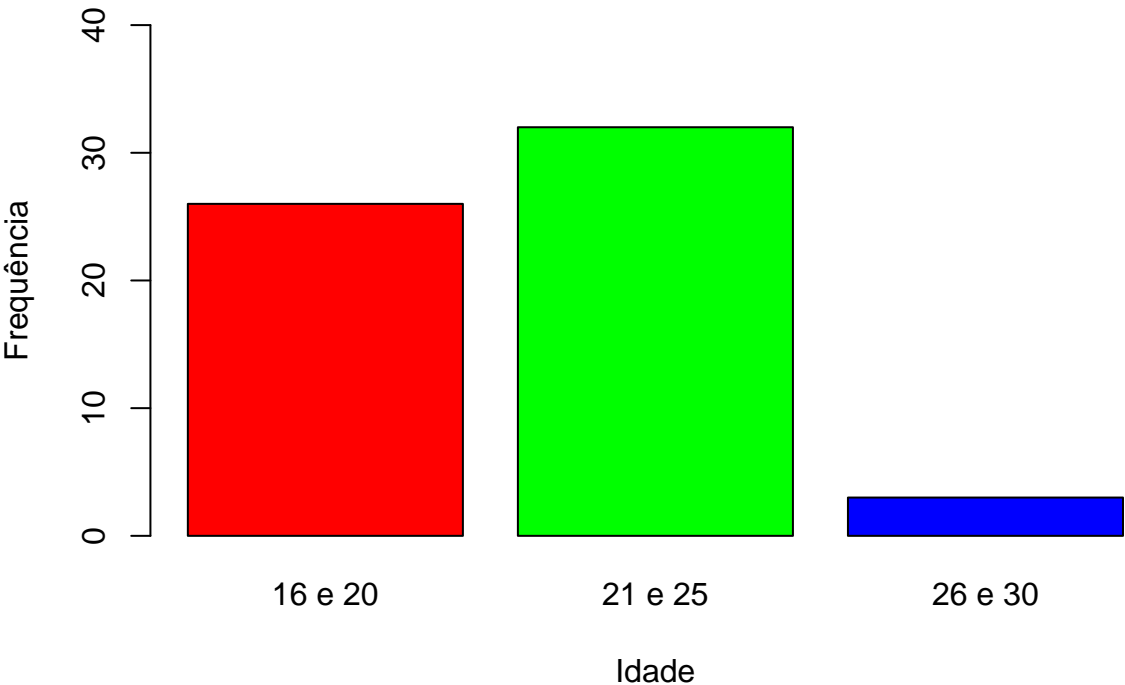
Faixa etária

O gráfico a seguir apresenta a distribuição de faixa etária dos entrevistados:

```
## Loading required package: readxl
```

```
## [1] "1. Entre 16 e 20 anos\r\n2. Entre 21 e 25 anos\r\n3. Entre 26 e 30 anos\r\n4. Entre 30 e 35 anos\r\n5. Entre 36 e 40 anos\r\n6. Entre 41 e 45 anos\r\n7. Entre 46 e 50 anos\r\n8. Entre 51 e 55 anos\r\n9. Entre 56 e 60 anos\r\n10. Entre 61 e 65 anos\r\n11. Entre 66 e 70 anos\r\n12. Entre 71 e 75 anos\r\n13. Entre 76 e 80 anos\r\n14. Entre 81 e 85 anos\r\n15. Entre 86 e 90 anos\r\n16. Entre 91 e 95 anos\r\n17. Entre 96 e 100 anos\r\n18. Não sabe responder"
```

Faixa etária dos questionados



Os dados

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.

Os gráficos

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.

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.