

RELATÓRIO – TITANIC

NANODEGREE FUNDAMENTOS DE DATA SCIENCE II – REINALDO LEPSCH NETO – versão 3

1. LINK PARA O WORKBOOK DO TABLEAU

Versão 1:

https://public.tableau.com/profile/reinaldo.lepsch.neto#!/vizhome/NDFDS2-projeto2-titanic-rlepsch-PRONTO/HISTORYv_1?publish=yes

Versão 2:

https://public.tableau.com/profile/reinaldo.lepsch.neto#!/vizhome/NDFDS2-projeto2-titanic-rlepsch-PRONTO-v_2/HISTORYv_2?publish=yes

Versão 3:

https://public.tableau.com/profile/reinaldo.lepsch.neto#!/vizhome/NDFDS2-projeto2-titanic-rlepsch-PRONTO-v_3/HISTORYv_3?publish=yes

2. RESUMO

O presente trabalho tem como objetivo apresentar, de modo visual, estatísticas básicas sobre os passageiros do HMS TITANIC, que naufragou no ano de 1912. Reunindo informações essenciais coletadas após a tragédia, busca construir relações entre os dados de passageiros sobreviventes e dessa forma apoiar a análise do ocorrido. Análises semelhantes, realizadas nas décadas que se seguiram ao acidente, permitiram grande melhora nas condições de segurança na navegação.

Com base em diversas análises feitas em épocas mais recentes (após a descoberta dos destroços na década de 80 e o lançamento da versão de maior sucesso no cinema, em 1997), realizei meu estudo procurando focar na principal explicação para o grande número de mortos no naufrágio: foi dada prioridade para o salvamento de passageiros de maior poder aquisitivo (primeira classe), que estavam em menor quantidade. Nas demais classes, foi dada prioridade ao salvamento de mulheres e crianças. Essas premissas foram buscadas nas visualizações apresentadas.

3. DESIGN

Procurei criar visualizações, inicialmente, de dados isolados (sexo, idade, classe, origem) e em seguida relacionando pares de variáveis e sua relação com a sobrevivência ou não. O gráfico de dispersão foi aquele em que mais informação pôde ser incluída numa única visualização. Procurei sempre que possível associar o gráfico construído com a proporção de sobreviventes, e assim analisar as variáveis de maior impacto.

4. COMENTÁRIOS

➤ Enviados por Luiza Bissoli em 11/04/2018.

- Acho que uns filtros interativos para as duas primeiras tabelas seriam ótimos para explorar os dados; **FEITO. As duas primeiras tabelas (sobreviventes e mortos) foram unificadas em "passageiros" onde é possível escolher pelo filtro interativo a visualização: sobreviventes / mortos / todos.**
- Muito interessante o recorte feito no gráfico de dispersão, pois alguns passageiros se comportavam como outliers e prejudicavam uma visualização melhor dos dados. **Acrescentei filtros interativos para os gráficos de dispersão e destaquei a área onde havia maior concentração.**
- Talvez seja interessante descartar os dados com idades inválidas (como você disse, acima de 120 anos), porque isso tornou o histograma pouco informativo, já que a distribuição que realmente poderia informar algo pegaram um intervalo grande de idade. Em outras palavras, 2 barras até 100 anos não diz

muita coisa. **FEITO. Criei um filtro interativo para esse histograma, acrescentei a faixa de idade (campo que eu criei para ter idade como dimensão) e diminui a largura do compartimento (bin).**

- Por fim, acho que os gráficos estão bacanas, o que está faltando é uma narrativa aos seus achados (dá uma olhada na rubrica). **FEITO. Melhorei a seção RESUMO acima, e enriqueci as explicações nas legendas da história.**

➤ Enviados por rodns01 em 12/04/2018.

- Quando eu abri vi várias abas no topo e fiquei confuso, acho que pode retirar. Você precisasempre salvar na primeira aba e enviar para o site, para quando abrimos a visualização abrir na primeira aba. Eu abri e apareceu no final, fiquei confuso. **FEITO na publicação.**
- Um dos requisitos do projeto é ter várias conclusões, vi pouquíssimo texto no trabalho. **FEITO conforme feedback anterior. Acrescentei diversas conclusões na história.**
- Um outro requisito é ter interatividade, não encontrei. **FEITO conforme feedback anterior. Acrescentei vários filtros interativos.**
- É legal ter uma aba introdução onde você fala sobre o que irá mostrar e uma aba conclusão. **FEITO.**
- Na aba 4 os rotulos do eixo estão em inglês. **FEITO. Alterado.**
- Não consegui concluir nada com os gráficos na aba 4(painel de correlações), acho que ficou muito confuso, poluído visualmente. **FEITO. Mantive o painel, mas inclui filtros interativos, além da legenda na história. Espero que tenha ficado menos confuso...**
- A lista de mortos achei curiosa, uma inovação. **Obrigado, inclui dois gráficos ao final, espero que fique legal.**
O resto está legal, abraço!

➤ Enviado por REVISÃO #1 em 12/04/2018 – o detalhamento abaixo é o texto dos itens a serem revisados:

Visualização é explanatória

A visualização centraliza em uma descoberta específica clara.

As visualizações da história devem ser do tipo explanatórias, quando queremos mostrar um achado específico que encontramos em nossos dados. Perceba que temos que criar uma narrativa com foco em uma descoberta específica e clara. Escolha um achado da análise e utilize diferentes gráficos para detalhar o achado específico que quer mostrar a partir de diversos ângulos.

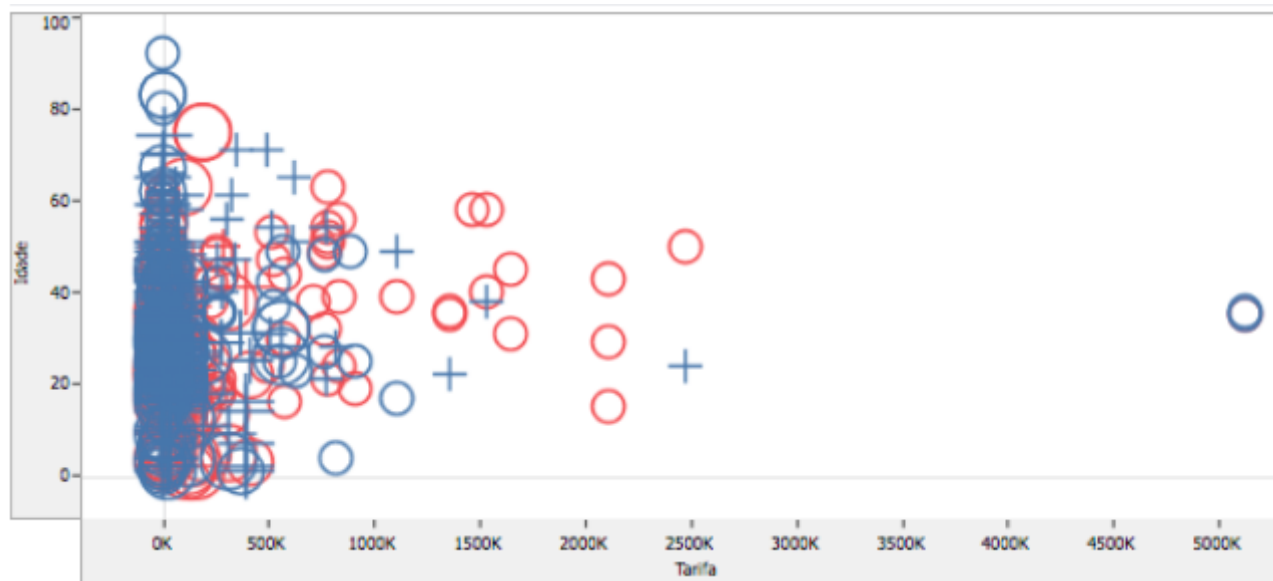
Os títulos de histórias aparecem sempre na exibição e são uma forma prática de centralizar o objetivo da história. Por favor, altere o título da história para que indique qual descoberta você quer mostrar. Títulos em formato de perguntas cativam mais a atenção do leitor. Um exemplo de pergunta que pode ser utilizada para indicar o achado da análise é: "Mulheres e crianças tiveram prioridade no desembarque?". **FEITO**.

Utilize as legendas dos painéis do Tableau para guiar o leitor pela história destacando pontos de interesse e tendências respaldados pelos gráficos apresentados. Utilize texto explicativos dentro dos painéis descrever seus achados. **FEITO. As legendas já continham bastante textos explicativos, eu procurei melhorá-los onde cabia.**

Veja aqui um [exemplo](#) do Tableau.

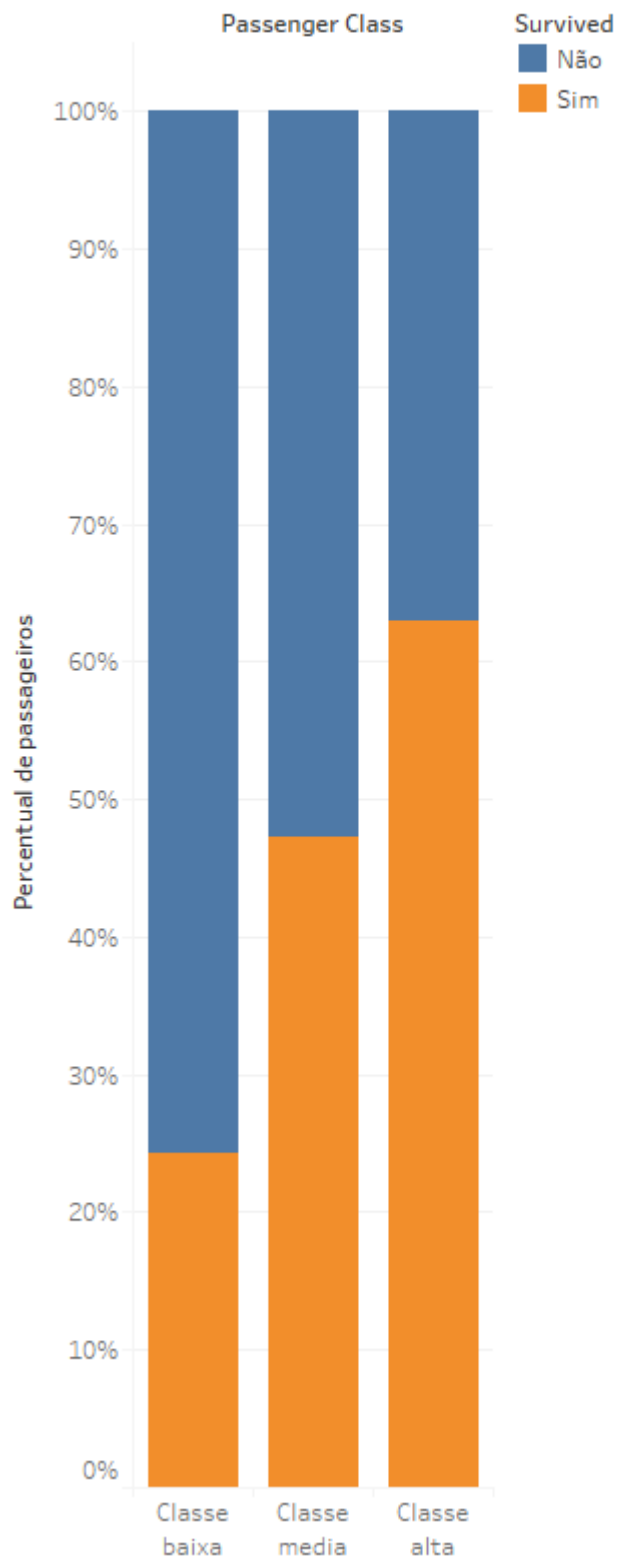
A descoberta é claramente comunicada. Escolhas de design favorecem a comunicação entre o leitor e a visualização.

Os gráficos devem ser claros para facilitar a interpretação dos dados pelos leitores. Evite gráficos com muitas variáveis simultâneas que deixam a interpretação bem confusa. Por exemplo, o gráfico a seguir é muito confuso e de interpretação extremamente complicada por envolver diversas variáveis simultâneas:



Sugiro que comece com gráficos simples, mostrando a diferença na taxa de sobrevivência com base em variáveis únicas. Veja como no gráfico a seguir é bem simples identificar a classe com menor taxa de sobreviventes:

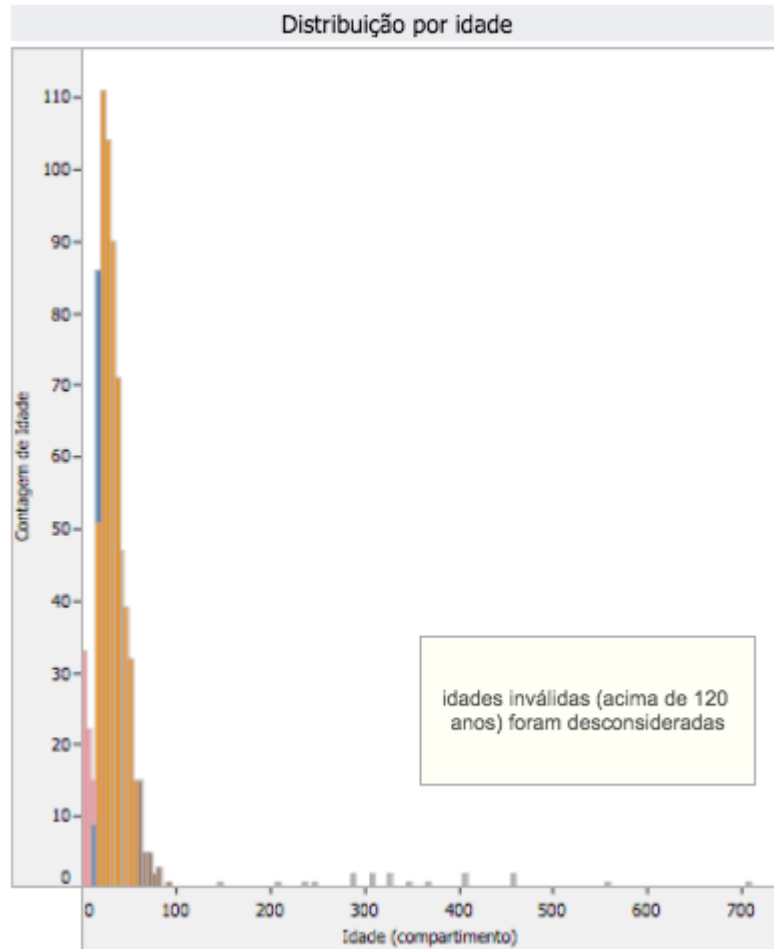
Taxa de sobrevivência por classe do passageiro



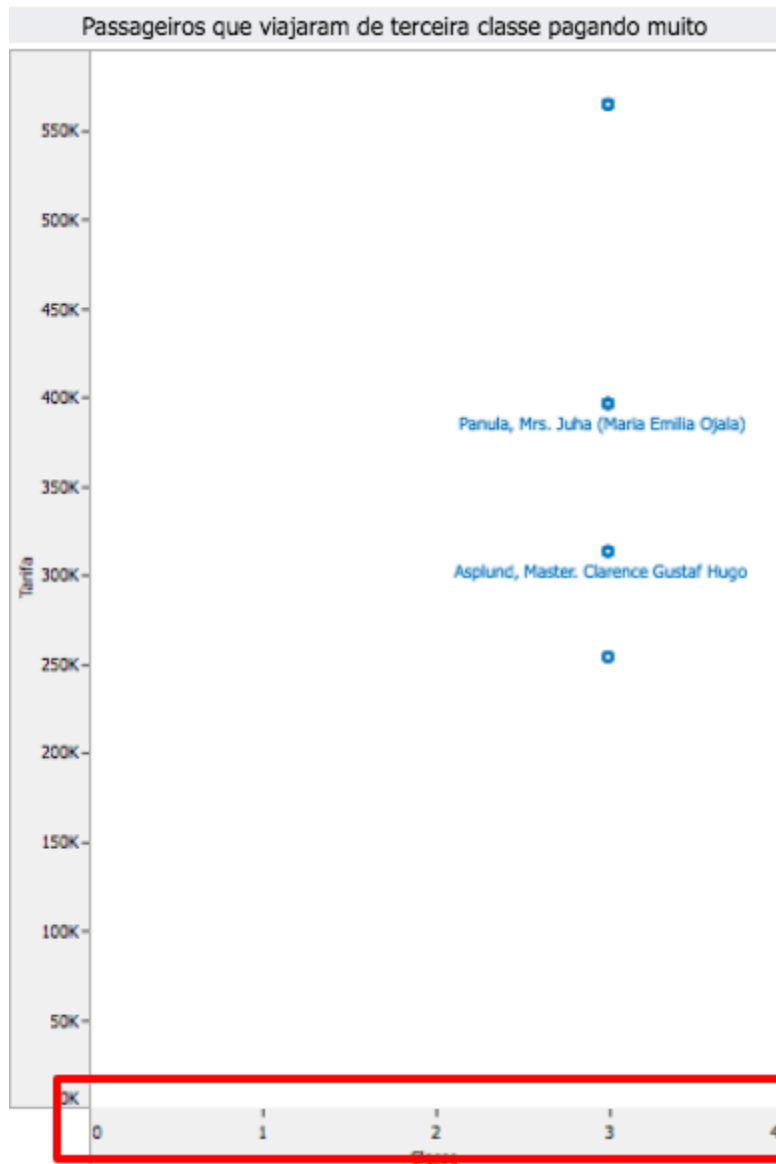
Esse tipo de gráfico é de fácil compreensão pelo leitor e comunica o achado de forma clara (classe baixa teve mais mortos). **FEITO. Exclui os gráficos mais**

confusos e poluídos da história (mas continuam no workbook). Mas não consegui deixar as barras de percentuais com a mesma altura.

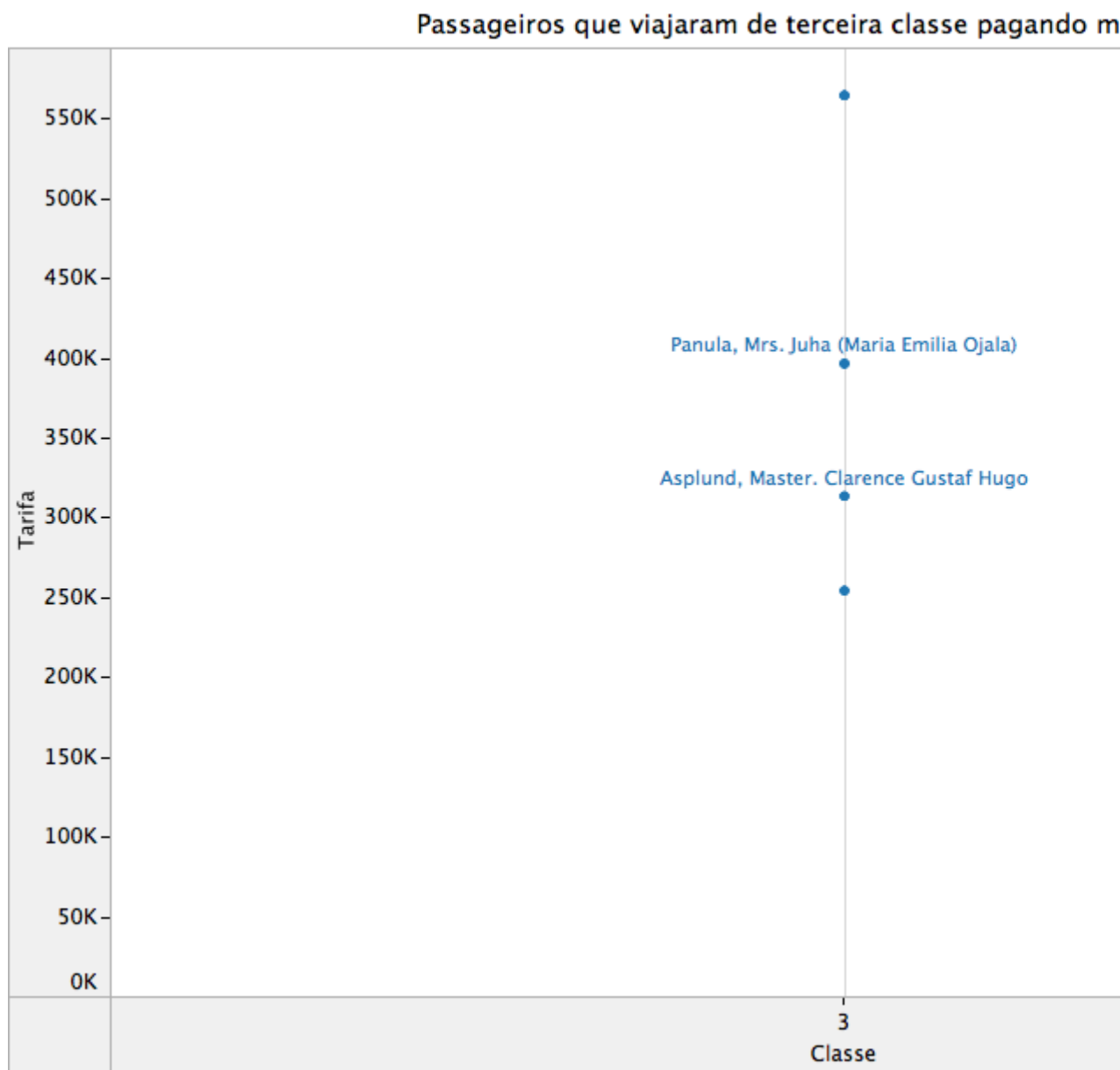
Os histogramas não apresentam dados corretos. No dataset do Titanic disponibilizado não tem passageiros com mais de 100 anos. Verifique se tratou corretamente os números decimais e corrija os gráficos. **FEITO. Foi um erro ao converter o arquivo CSV para XLSX.**



Variáveis categóricas estão sendo tratadas como variáveis contínuas. Veja por exemplo no gráfico a seguir:



Perceba que estão sendo mostrados valores de Pclass que nem existem (como 0 e 4). Converta a variável para DImensão e o gráfico ficará correto:



FEITO. Na verdade eu preferi mudar a visualização, para o formato mapa de árvore, onde o tamanho do retângulo mostra o valor da tarifa paga.

5. RECURSOS

Consultei diversos exemplos de visualizações no github e na nuvem do Tableau (Tableau public), além da própria documentação do produto. Iniciei meu trabalho diretamente no Tableau public mas devido a erros da própria ferramenta tive que instalar a versão Desktop 10.5 (versão de avaliação) e refazer tudo lá.

Os dados sobre casos pontuais de passageiros foram consultados na Wikipedia.