

< [Return to "Data Analyst Nanodegree" in the classroom](#)

Explore and Summarize Data

REVISÃO

REVISÃO DE CÓDIGO

HISTORY

Meets Specifications

Seu trabalho ficou sensacional!

Continue assim para manter sua excelente trajetória.



Boa sorte em seus próximos projetos!

Se quiser me adicionar no [LinkedIn \(Rafael Buck\)](#) fique à vontade.

Funcionalidade do código



Todo o código é funcional (e.g. Nenhum erro ocorre durante a execução e o documento RMD não está protegido de ser usado no knit.)

Perfeito! Código executa sem erros. Parabéns 🤗

Sugestão: você poderia definir knitr para tornar os tamanhos de plotagem uniformes, com largura de página e centralizado. Você consegue fazer isso com este código:

```
knitr::opts_chunk$set(fig.width=9,fig.height=5,fig.path='Figs/',
  fig.align='center',tidy=TRUE,
  echo=FALSE,warning=FALSE,message=FALSE)
```

Você também pode tornar seu relatório mais fácil de navegar adicionando `toc_float: true` ao seu YAML, desta forma:

```
---
title: "White Wine Quality Analysis"
output:
  html_document:
    toc: true
    toc_depth: 2
    toc_float: true
    theme: united
---
```

Introduction
Univariate Plots Section
Distribution charts of variables
Univariate Analysis
Bivariate Plots Section
Bivariate Analysis
Multivariate Plots Section
Multivariate Analysis
Final Plots and Summary
Summary
Reflection
Sources:

White Wine Quality Analysis by David Filip

Introduction

This tidy data set contains 4,898 white wines with 11 variables on quantifying the chemical properties of each wine. At least 3 wine experts rated the quality of each wine, providing a rating between 0 (very bad) and 10 (very excellent). (Source: Dataset option description)

In this analysis we will look at what chemical properties are found with highly rated wines.

Univariate Plots Section

Data Properties

Displaying basic properties of the data.

```
## 'data.frame': 4898 obs. of 13 variables:
## $ X : int 1 2 3 4 5 6 7 8 9 10 ...
## $ fixed.acidity : num 7 6.3 8.1 7.2 7.2 8.1 6.2 7 6.3 8.1 ...
## $ volatile.acidity : num 0.27 0.3 0.28 0.23 0.23 0.28 0.32 0.27 0.3 0.22 ...
## $ citric.acid : num 0.36 0.34 0.4 0.32 0.32 0.4 0.16 0.36 0.34 0.43 ...
## $ residual.sugar : num 20.7 1.6 6.9 8.5 8.5 6.9 7 20.7 1.6 1.5 ...
## $ chlorides : num 0.045 0.049 0.05 0.058 0.058 0.05 0.045 0.045 0.049 0.044 ...
## $ free.sulfur.dioxide : num 45 14 30 47 47 30 30 45 14 28 ...
## $ total.sulfur.dioxide : num 170 132 97 186 186 97 136 170 132 129 ...
## $ density : num 1.001 0.994 0.995 0.996 0.996 ...
## $ pH : num 3 3.3 3.26 3.19 3.19 3.26 3.18 3 3.3 3.22 ...
## $ sulphates : num 0.45 0.49 0.44 0.4 0.4 0.44 0.47 0.45 0.49 0.45 ...
## $ alcohol : num 8.8 9.5 10.1 9.9 9.9 10.1 9.6 8.8 9.5 11 ...
## $ quality : int 6 6 6 6 6 6 6 6 6 ...
```

All columns are of the data type num, except for quality and x that are of the type int.

Statistical Summary

Displaying a statistical summary of the data.

```
## X fixed.acidity volatile.acidity citric.acid
## Min. : 1 Min. : 3.800 Min. : 0.0800 Min. : 0.0000
```



O projeto quase nunca usa código repetido onde uma função seria mais apropriada. O código referencia variáveis por nome ao invés de usar constantes e números de colunas.

Código bem estruturado. Ficou top!

Sugestão: segue [aqui uma referência de vários cheat sheets](#) de como utilizar funções e outras coisas em R e no RStudio. É sempre legal ter uma dessas para consulta rápida e para manter o código sempre bem estruturado 😊

Leitura do código



O código usa técnicas de formatação de forma consistente e efetiva para melhorar a legibilidade do código. Todas as linhas possuem menos de 80 caracteres.

O código está bem estruturado.

Sugestão: De uma olhada nas melhores práticas de estilo de R [nesse link](#). Tem alguns trechos onde podem seu código podem ser aprimorados. Use isso em seus próximos projetos que seu código ficará top 😊



Todo o código complexo está adequadamente explicado com comentários. Está sempre claro o que o código faz, como e por que algumas decisões não usuais de código foram feitas.

O código está bastante legível, especialmente pelo uso de textos explicativos nos resultados.

Sugestão: Sempre que possível, procure comentar mais seu código. Assim, sua análise ficará muito boa do ponto de vista da leitura de um terceiro. Isso enriquecerá muito seu trabalho 😊 Aqui uma [referência bem legal sobre o padrão de código R do Google](#)



Sintaxe Markdown é usada no arquivo RMD para melhorar a legibilidade do arquivo.

Usou adequadamente a sintaxe Markdown no arquivo RMD. Show!

Sugestão: segue [aqui uma referência bem legal dos principais formatos da sintaxe markdown em R](#). Utilizar ao máximo essa funcionalidade deixa seu trabalho sempre muito profissional e muito bom para ser publicado. Recomendo sempre explorar bastante isso 😊

Qualidade das análises



O projeto utiliza de forma adequada gráficos univariados, bivariados e multivariados para explorar maior parte dos relacionamentos de um conjunto de dados.

Excelente! O projeto utiliza de forma adequada gráficos univariados, bivariados e multivariados.



Questões e descobertas são dispostas entre blocos de código R de forma regular, fazendo com que se torne claro o que o aluno estava pensando durante a análise.

Curti o fluxo e a evolução da análise.

Sugestão: desenvolva mais as explicações de cada gráfico após a geração/plot de cada um deles, como se fosse um artigo. Isso deixará sua análise rica e muito profissional 😊



Explicações são fornecidas com os gráficos durante a análise. Os gráficos formam um fluxo lógico. Comentários após os gráficos refletem de forma adequada os conteúdos apresentados nos gráficos.

Novamente, curti bastante o fluxo e a evolução da análise, com muitos gráficos interessantes.

Sugestão: aqui também sugiro que desenvolva mais as explicações de cada gráfico após a geração/plot de cada um deles, como se fosse um artigo. Isso deixará sua análise rica e muito profissional 😊



O projeto contém ao menos 10 visualizações. As visualizações são variadas e mostram múltiplas comparações e tendências. Estatísticas relevantes (e.g. média, mediana, intervalos de confiança, correlações) são calculadas durante a análise quando uma inferência é feita sobre os dados.

A análise contém pelo menos 10 visualização (por sinal, muito interessantes) 😊



Visualizações feitas durante o projeto apresentam os dados de forma a permitir interpretação instantânea. A escolha do tipo do gráfico, variáveis e parâmetros visuais (e.g. largura das partições, color, eixos) é apropriada.

Top!

Gráficos finais e resumo



O projeto inclui uma seção de gráficos finais e sumário contendo três gráficos e um comentário. Todos os gráficos nesta seção refletem o que foi explorado durante o corpo da análise.

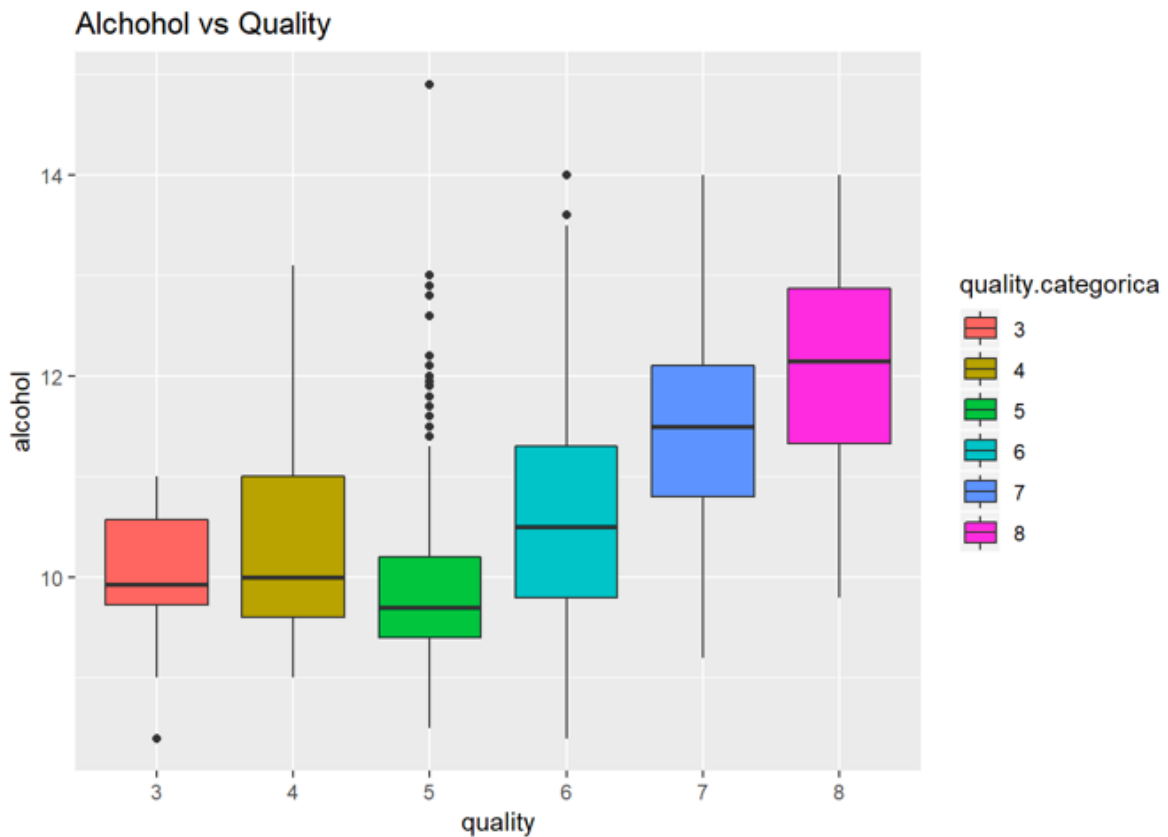
Excelente a seção de conclusão e seleção dos três gráficos finais. Ficou bem legal!



Os gráficos foram bem escolhidos e atendem ao menos dois critérios, cada. Os gráficos são variados e revelam interessantes tendências e relacionamentos.

Requisitos cumpridos corretamente! Os gráficos foram bem escolhidos e revelam tendências e relacionamentos interessantes, como este (o que achei muito interessante):

Segundo Gráfico



Todos os gráficos possuem variáveis selecionadas apropriadamente e foram apresentados de uma forma que representa bem os dados/informações (i.e. as descobertas do Gráfico Final 1 não dependem dos resultados do Gráfico Final 2).

Os três gráficos atendem aos requisitos da Udacity. Ficou show!



O raciocínio e descobertas de cada gráfico são explicados e o texto sobre cada gráfico é descritivo o suficiente para ser auto-suficiente. Comentários refletem os conteúdos dos gráficos a que são associados.

Excelente trabalho! Espero que tenha se divertido 😊

Sugestão: a principal ideia nesse trabalho é envolver o leitor com um storytelling de dados. [Aqui um referência legal do próprio blog da Udacity que fala sobre o assunto - Data Storytelling](#). Seguem também posts de duas ferramentas bem legais, e bastante utilizadas pela indústria, para montar storytelling sobre dados, [aqui um post do Tableau](#) e [aqui um post do Microsoft PowerBI](#).



Todos os gráficos estão rotulados adequadamente (rótulos dos eixos, títulos dos gráficos e unidades dos eixos) e podem ser lidos e facilmente interpretados. Gráficos foram apresentados em um tamanho adequado.

Perfeito, todos os gráficos estão rotulados adequadamente (rótulos dos eixos e títulos dos gráficos) e podem ser lidos e facilmente interpretados. Parabéns!

Sugestão: com relação à unidade, por exemplo, você poderia usar a seguinte sintaxe nos ggplot

```
... x = expression(Density~(g/cm^3)), ...
```

Reflexão



O projeto inclui uma seção de referência que discute a análise realizada.



A seção reflete como a análise foi conduzida e reporta as dificuldades e sucessos da análise. Esta seção fornece ao menos uma idéia ou questão para trabalhos futuros. A seção explica decisões importantes da análise e como estas decisões afetaram a análise.

Conclusões excelentes. Ficou top! 😊

 [BAIXAR PROJETO](#)

RETORNAR