

Teste Técnico DTI – Estatístico@

Processo Seletivo

Autor: Marina Alves Amorim

Introdução



Este arquivo contém análises estatísticas referente a uma base de dados que avalia a Utilização da Wikipedia como Ferramenta de Ensino.



A base de dados "wiki4HE", é fruto de uma pesquisa enviada a professores de duas universidades espanholas entre 2012 e 2013: Universitat Oberta da Catalunya (UOC) e Universitat Pompeu Fabra (UPF). Ela contém 913 respostas (linhas) e 53 atributos (colunas).



A pesquisa foi organizada em duas partes. A primeira parte teve como objetivo coletar dados demográficos como: sexo, idade, área de especialização, doutorado, anos de experiência em ensino universitário, nível acadêmico e associação registrada na Wikipédia.

A segunda parte da pesquisa teve como objetivo reunir informações sobre os diferentes aspectos da Wikipedia no que diz respeito ao ensino superior e às opiniões dos professores. Essas perguntas tiveram que ser respondidas através de uma escala Likert de 5 pontos. Essa escala se refere ao nível de concordância ou discordância com uma afirmação (1 = "Discordo totalmente" e 5 = "Concordo totalmente").

Informações técnicas



O download da base de dados pode ser feita através do link:
<http://archive.ics.uci.edu/ml/machine-learning-databases/00334/>



Mais informações sobre a base de dados estão disponíveis no link:
<http://archive.ics.uci.edu/ml/datasets/wiki4HE>.



A linguagem de programação utilizada nas análises será o R (<https://www.r-project.org/>). O R é um ambiente de software livre para modelagens e gráficos estatísticos.



Todas as análises contidas neste documentos vão estar disponíveis no meu github. Link :
<https://github.com/MarinaAmorim/Analise-dos-dados-wiki4HE>

Objetivos

Exibir estatísticas básicas sobre os professores participantes, como a proporção de homens e mulheres, por meio de gráficos e tabelas.

Etapa 1

Analisar respostas de diferentes grupos de usuários para os itens da categoria Perceived Enjoyment: ENJ1 e ENJ2.

Etapa 2

Realizar outras análises que sejam interessantes e que apresentem informações relevantes sobre os dados.

Etapa 3

Análise descritiva



A base de dados apresenta informações de 953 professores e 53 perguntas respondidas.

Para análise exploratória, vamos utilizar as variáveis: Sexo, Idade, Anos de experiência, qual universidade o professor pertence, a área de ensino, se ele possui PhD, a posição dele no departamento e se é usuário da Wikipédia.

Também fizemos algumas descritivas para as variáveis gerais que estão na escala likert.



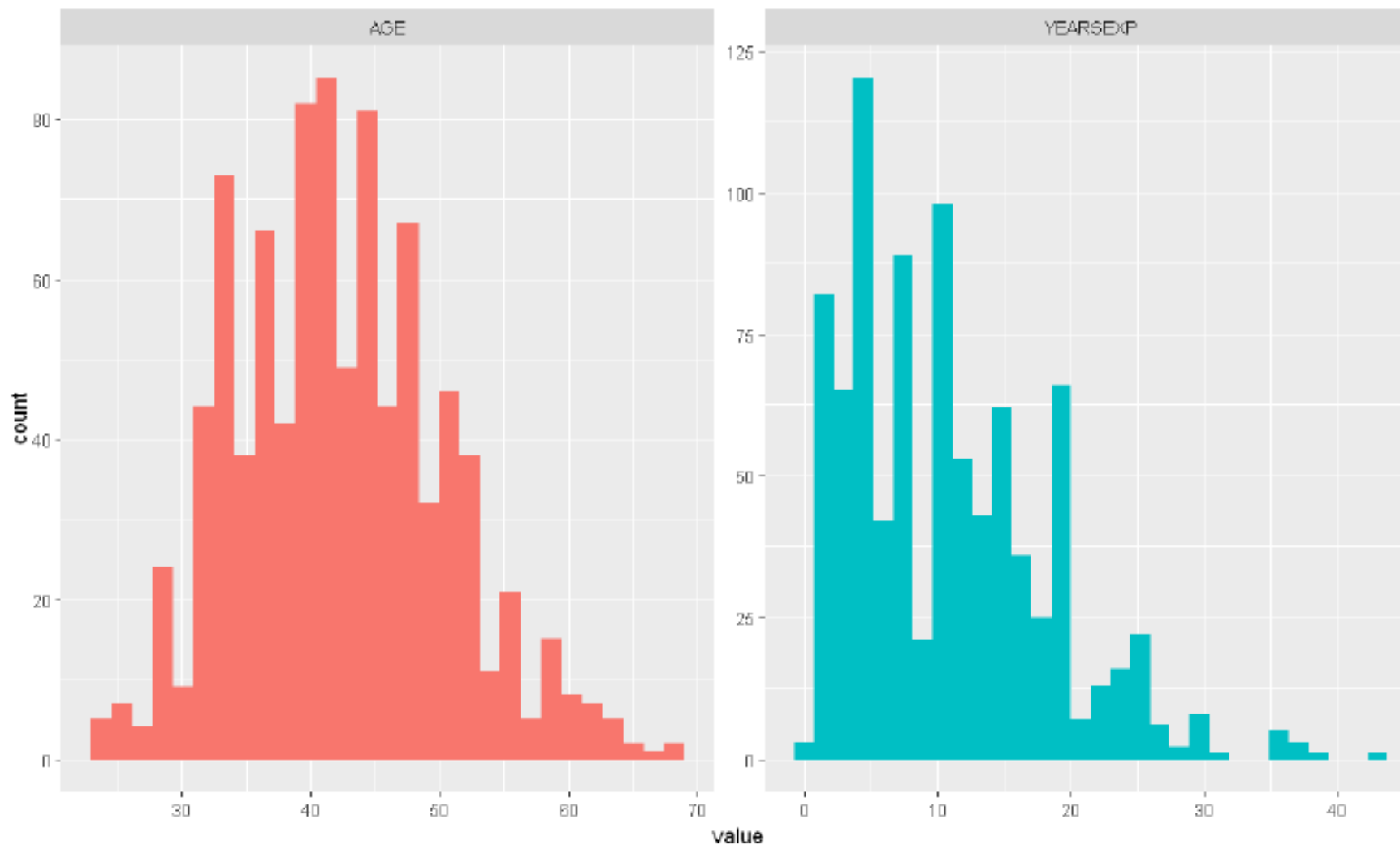
Para análise exploratória vamos utilizar gráficos e tabelas.



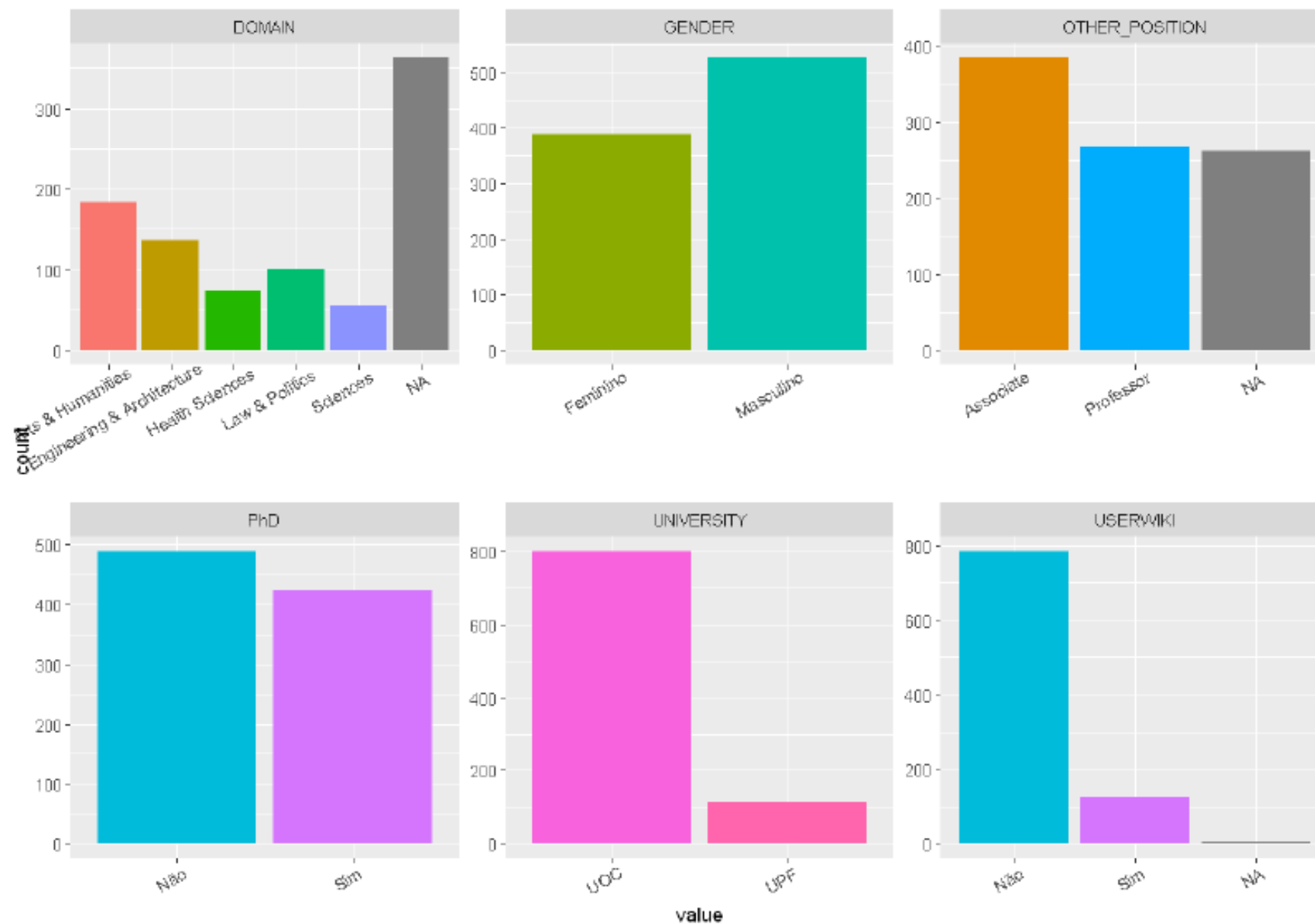
O relatório completo com todas as tabelas e com o código pode ser encontrado no meu github:

https://github.com/MarinaAmorim/Analise-dos-dados-wiki4HE/blob/master/Teste_DTI.ipynb

Análise descritiva



Análise descritiva



Análise descritiva



Etapa 2



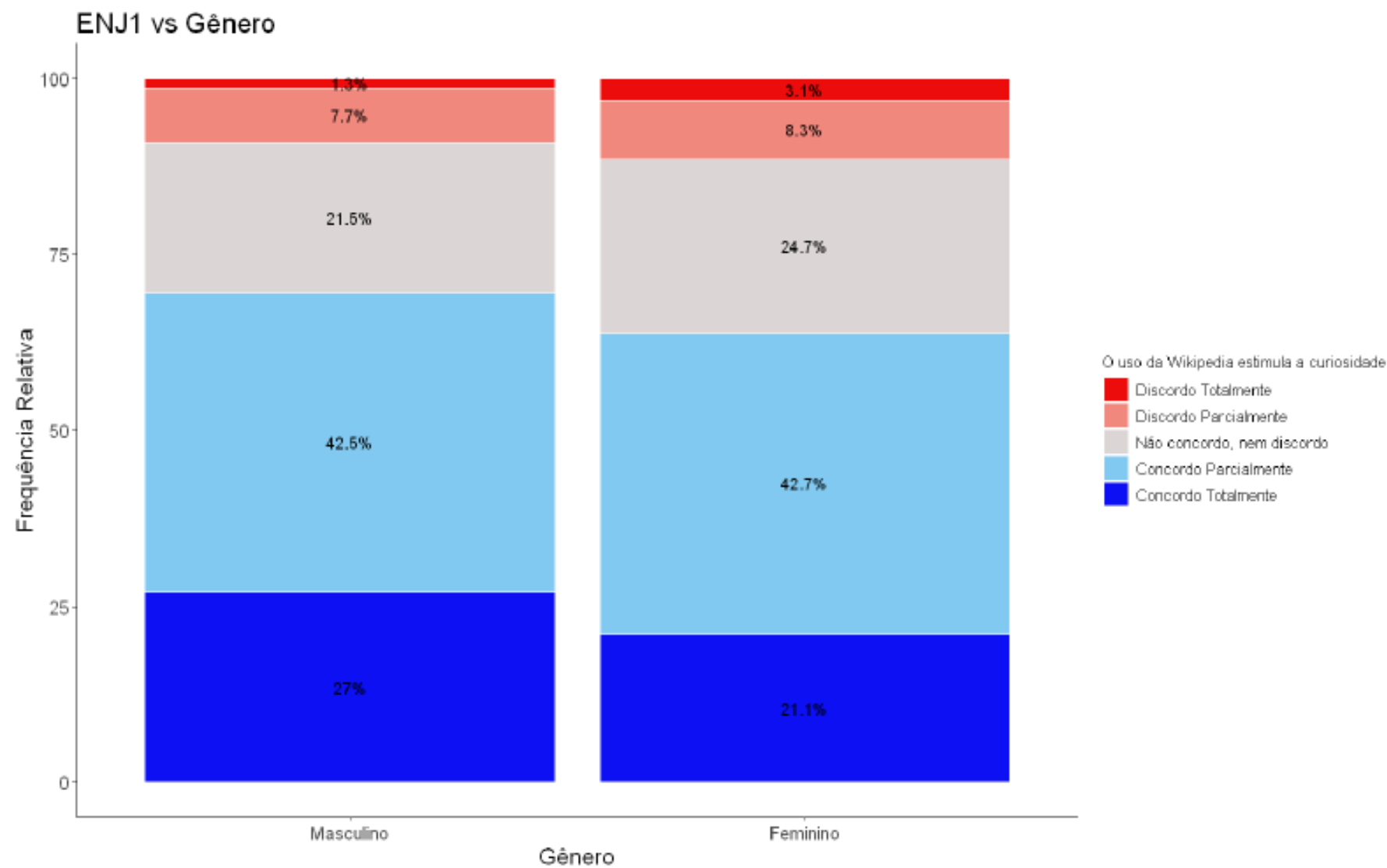
Nesta parte do trabalho, vamos analisar as respostas de diferentes grupos de usuários para os itens da categoria "Perceived Enjoyment": ENJ1 e ENJ2.

- ENJ1: O uso da Wikipedia estimula a curiosidade
- ENJ2: O uso da Wikipedia é divertido

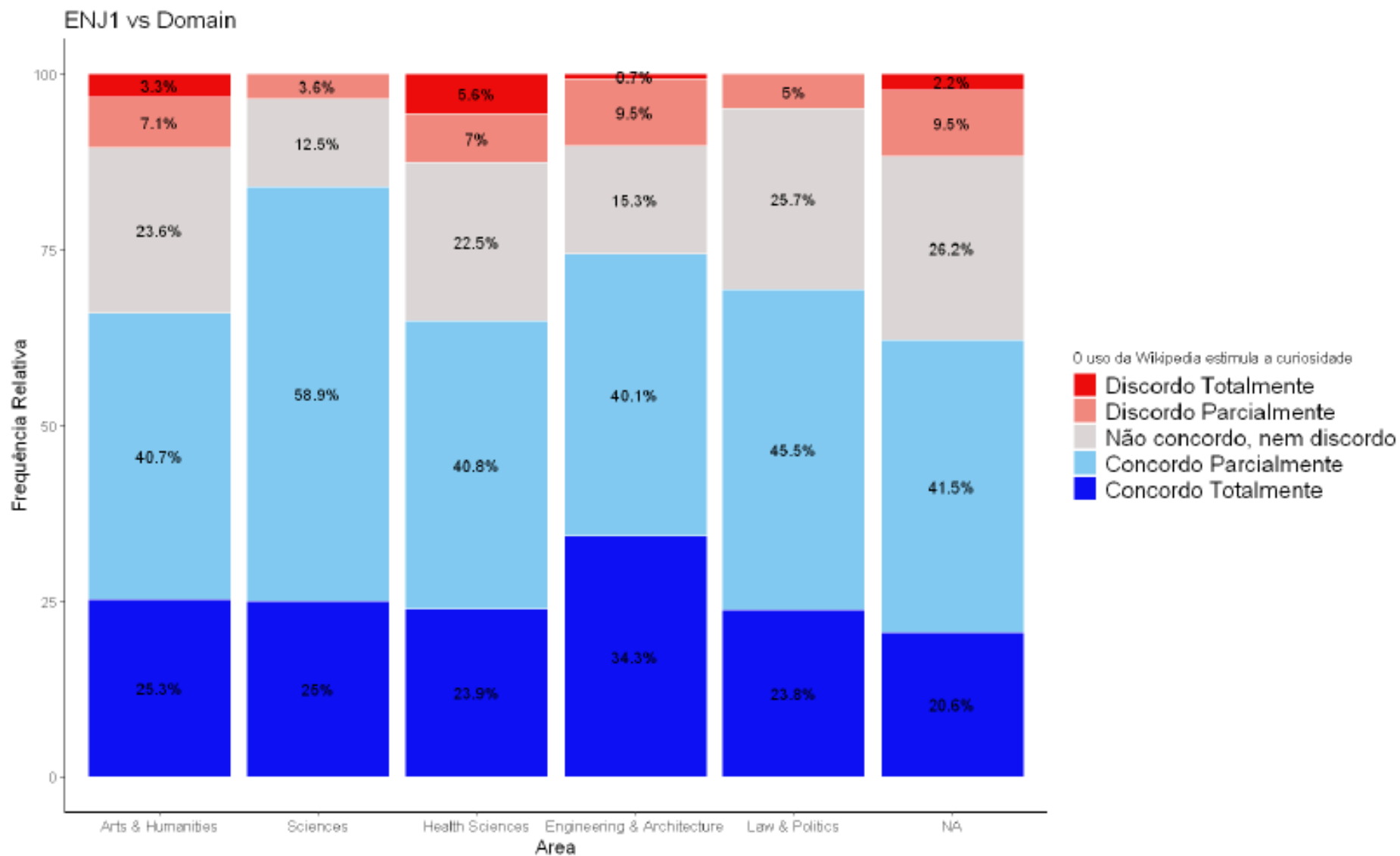


Vamos utilizar as variáveis que já estão sendo observadas no decorrer do trabalho para comparar com essas duas variáveis, são elas: sexo, área, PhD, universidade e se utiliza a Wikipédia.

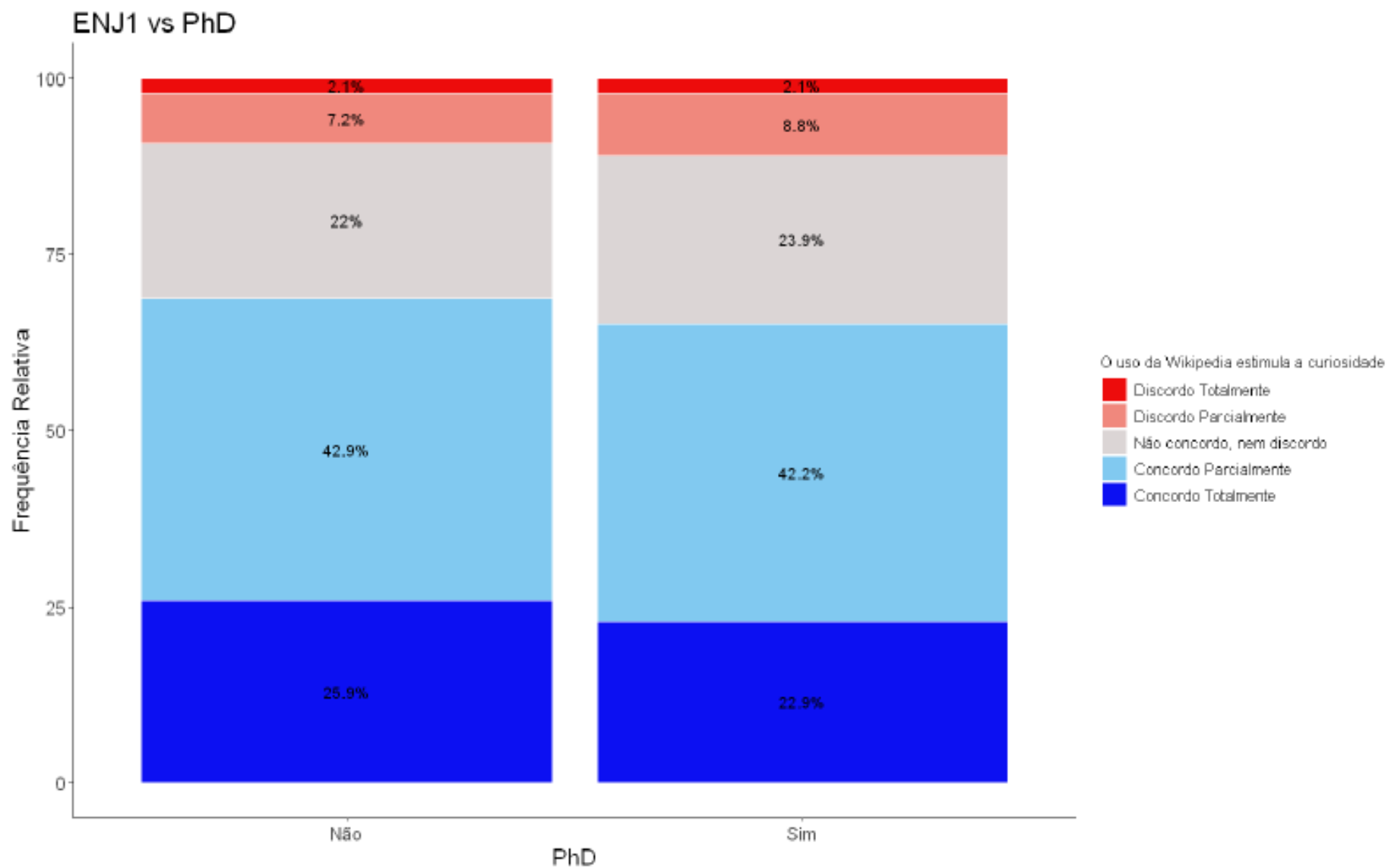
O uso da Wikipédia estimula a criatividade?



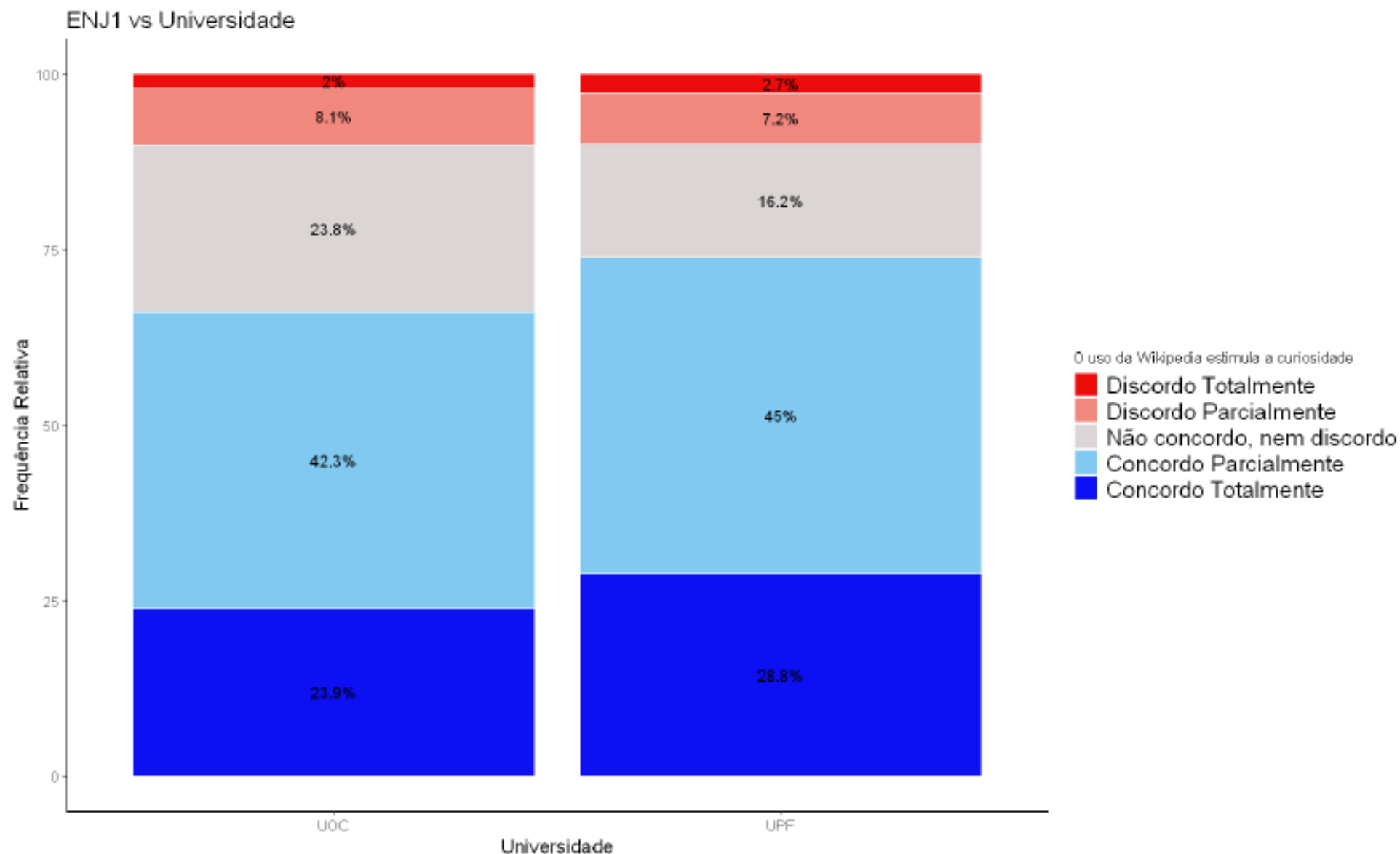
O uso da Wikipédia estimula a criatividade?



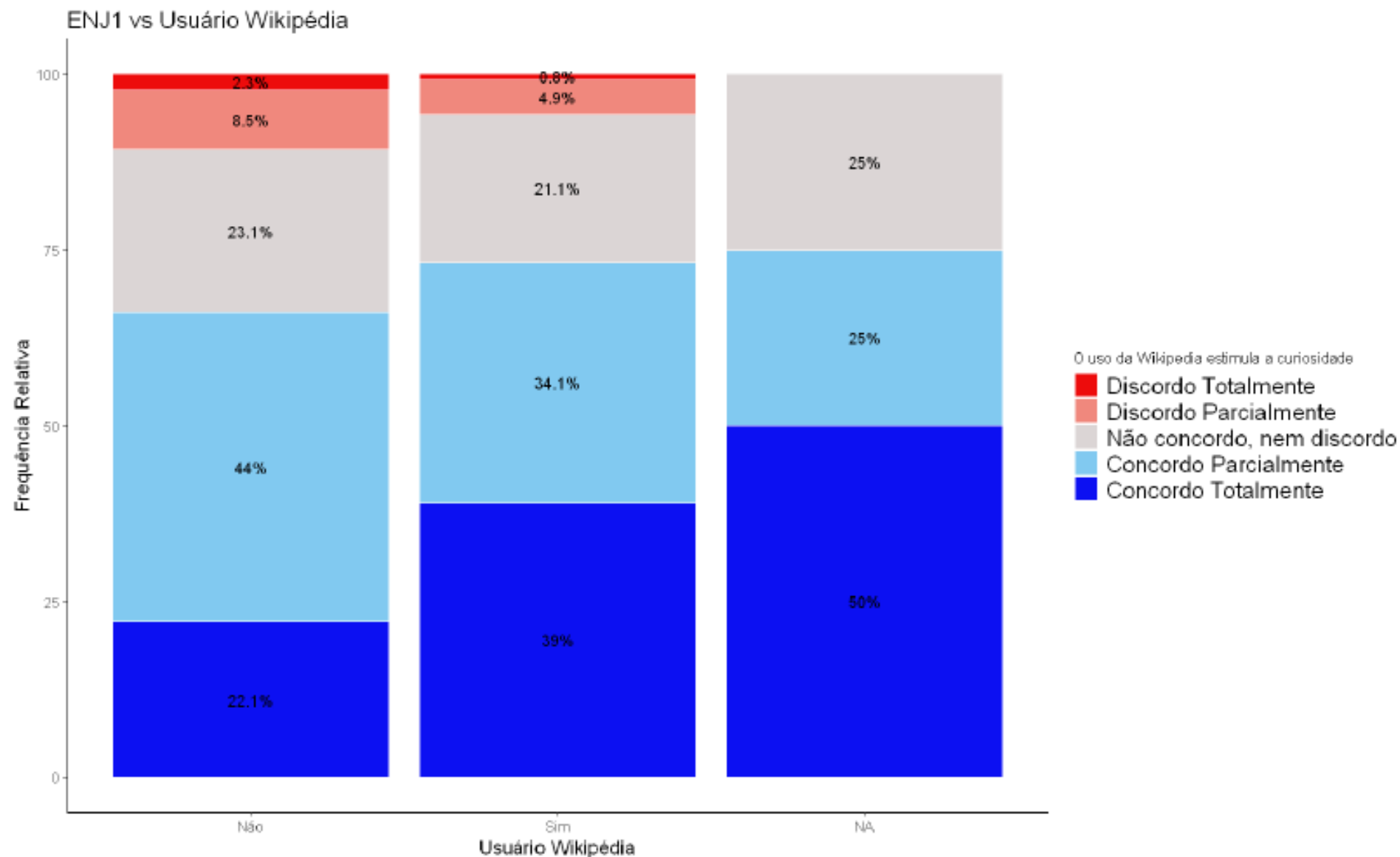
O uso da Wikipédia estimula a criatividade?



O uso da Wikipédia estimula a criatividade?



O uso da Wikipédia estimula a criatividade?



Análises

O uso da Wikipédia estimula a criatividade?



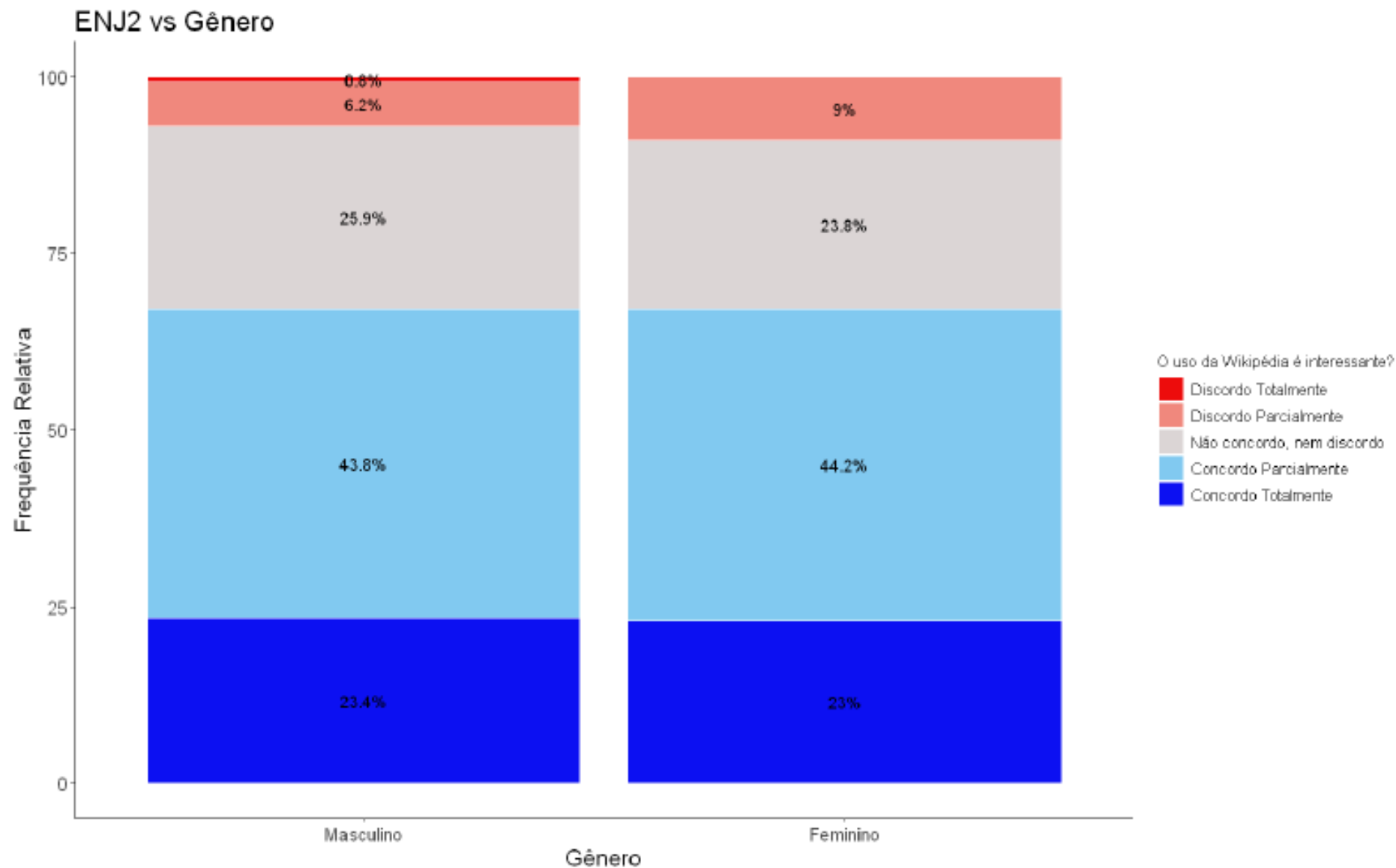
Quando analisamos a variável ENJ1 (O uso da Wikipédia estimula a curiosidade) comparada com sexo, área, PhD, universidade e se utiliza a Wikipédia, os percentuais são muito parecidos entre as categorias. Apenas a variável que leva em consideração o fato do professor utilizar ou não a Wikipédia que mostra uma leve diferença nas categorias, mostrando que indivíduos que usam a ferramenta tende a concordar com o fato que ela estimula a criatividade.

O uso da Wikipédia é interessante?

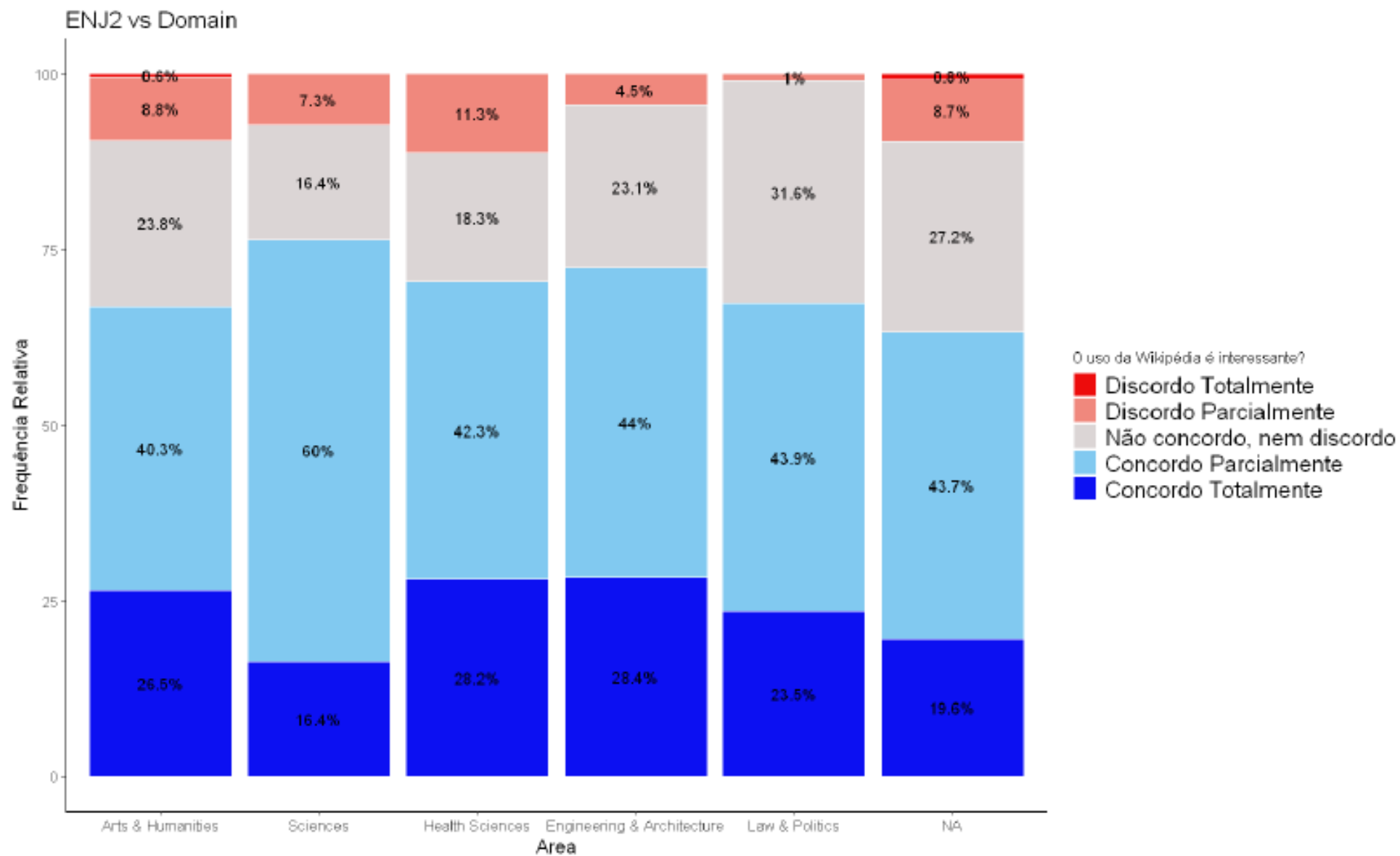


Já para a variável ENJ2 (O uso da Wikipedia é interessante) comparada com sexo, área, PhD, universidade e se utiliza a Wikipédia, os percentuais são muito parecidos entre as categorias, os resultados são bem parecidos com os anteriores. Apenas a variável que leva em consideração o fato do professor utilizar ou não a Wikipédia que mostra uma leve diferença nas categorias, mostrando que indivíduos que usam a ferramenta tende a concordar com o fato que ela é interessante.

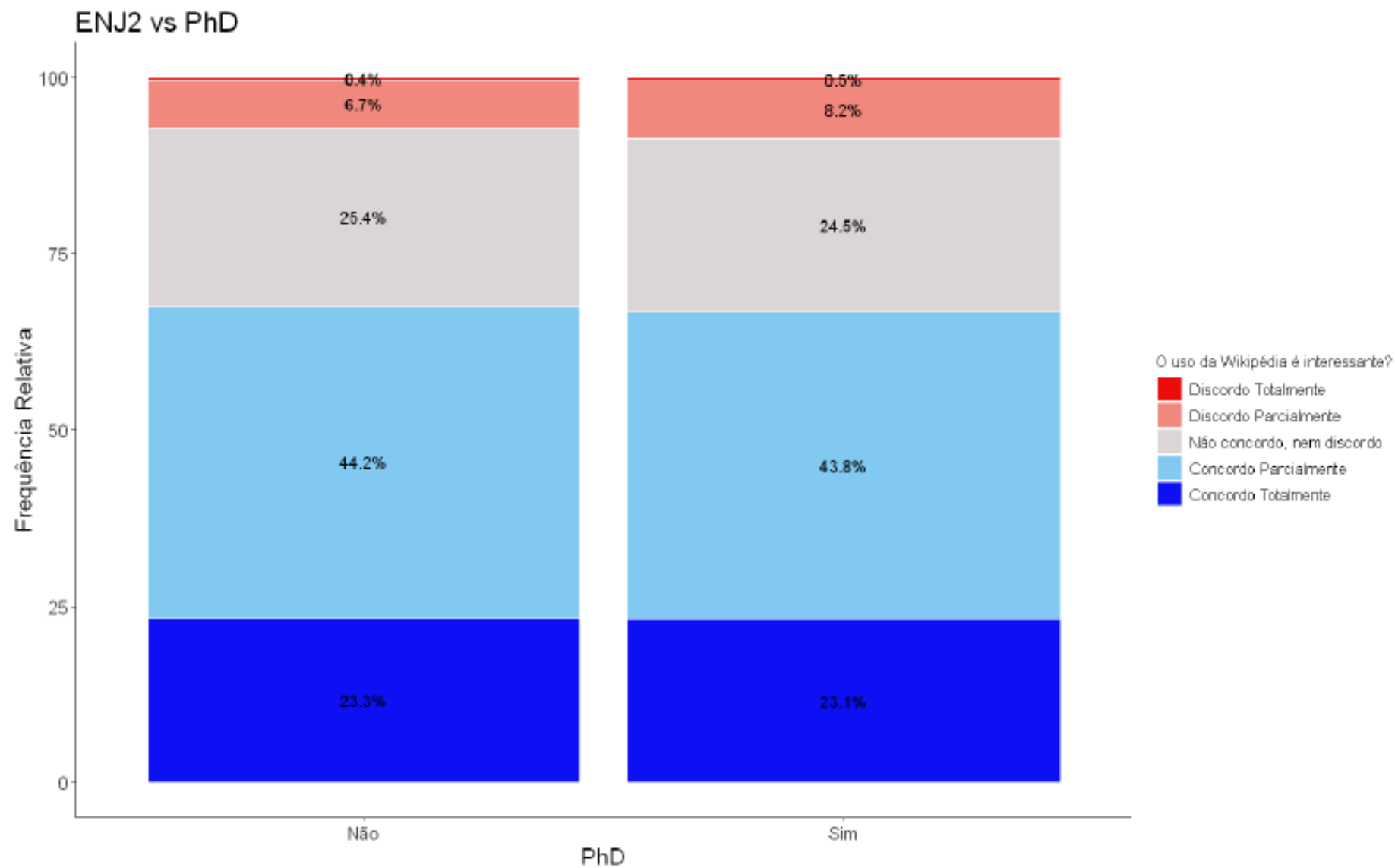
O uso da Wikipédia é interessante?



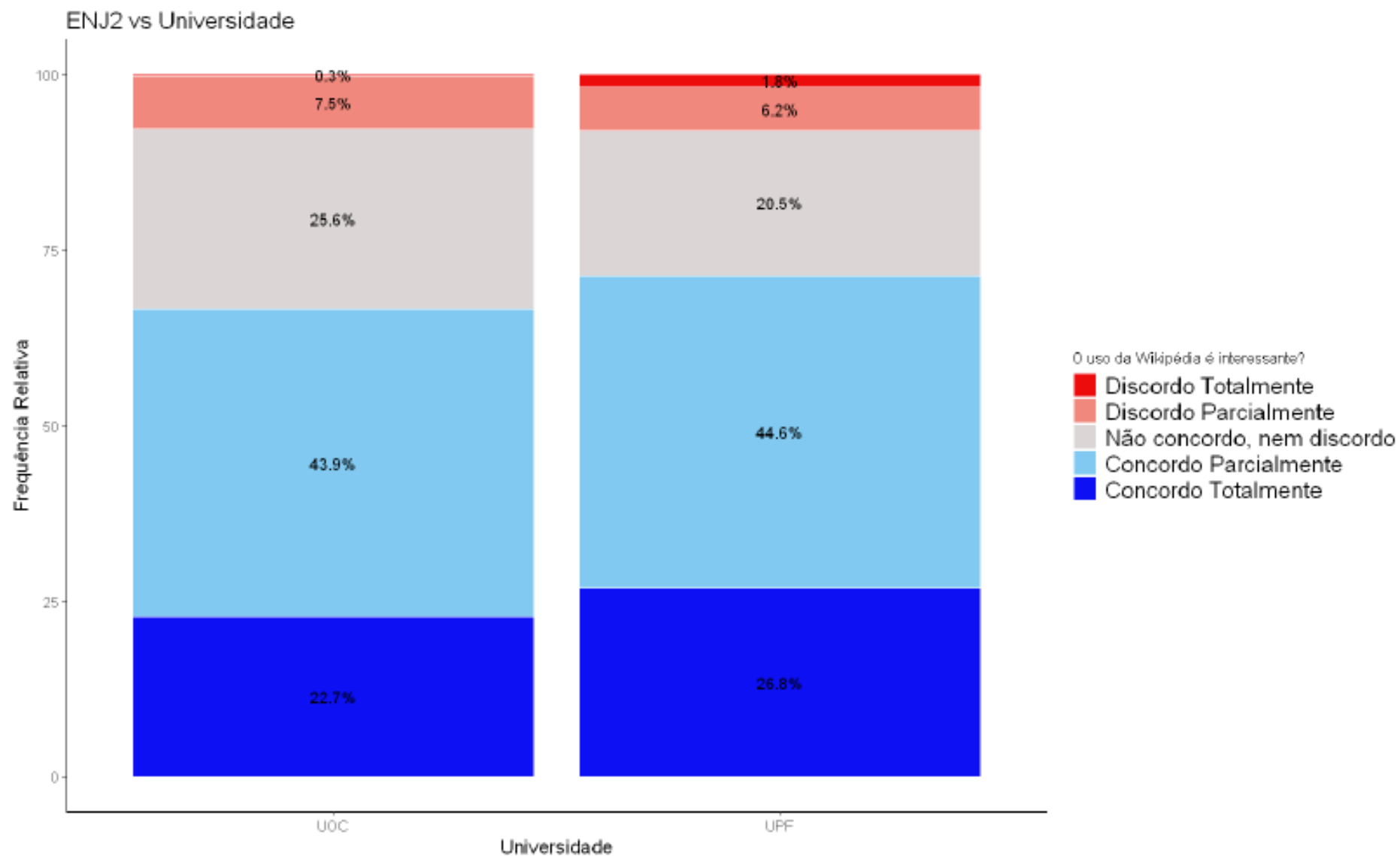
O uso da Wikipédia é interessante?



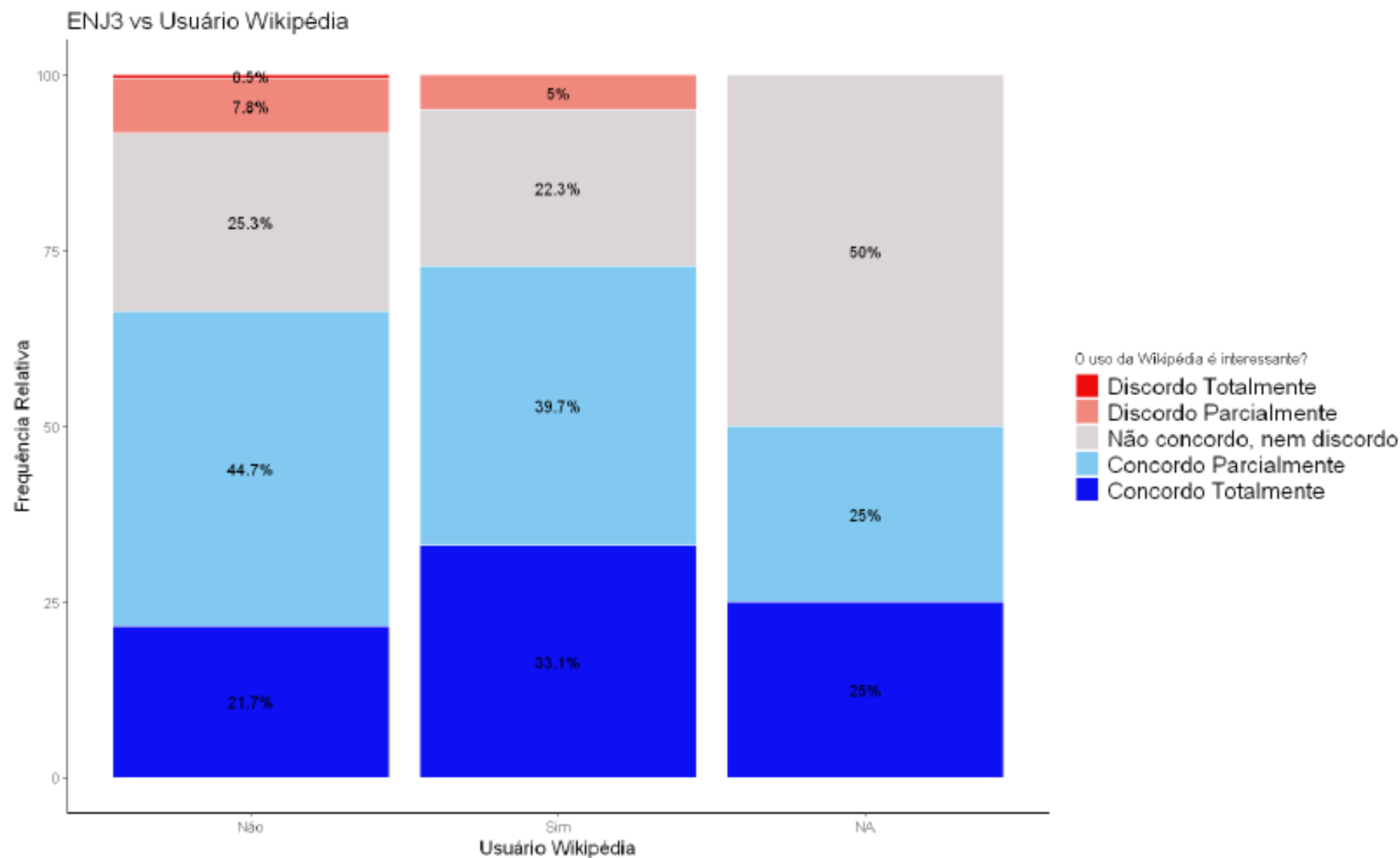
O uso da Wikipédia é interessante?



O uso da Wikipédia é interessante?



O uso da Wikipédia é interessante?



Etapa 3



Em nossa base de dados temos a variável **USE3** (recomendo que meus alunos usem a Wikipédia), está variável está na escala likert, ou seja, de 1 a 5.

Vamos criar um indicador para transformar a variável em binária. Para isso, vamos definir que professores que marcaram para a questão os valores 4 ou 5 recomendariam o uso e os que marcaram 1,2 e 3 não recomendariam. Essa regra de decisão é arbitrária e poderia ser feita de outra forma se preferir. Agora que a variável de recomendação do professor a Wikipédia é uma variável binária, podemos utilizar regressão logística.



A regressão logística é uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, binária, a partir de uma série de variáveis explicativas.



Como variável explicativa vamos utilizar as questões que já discutimos neste relatório, são elas: Idade, Gênero, Universidade, Anos de Experiência, Área, Se tem ou não PhD e se utiliza a Wikipédia.

Etapa 3



Com a regressão logística podemos identificar quais variáveis são importantes para explicar o fato de os professores recomendarem ou não o uso da Wikipédia e com isso, conseguimos traçar os perfis dos professores que recomendam.



Após ajustar o modelo, temos um teste de hipótese para cada variável explicativa, esse teste nos informa quais variáveis são estatisticamente significativas para explicar a recomendação da Wikipédia para os alunos.



Concluimos que, apenas as variáveis Gênero, Domain, Universidade e o fato de usar ou não a Wikipédia ajuda a explicar o fato do professor recomendar ou não o uso para os alunos.

Análises



Para encontrar todos os cálculos realizados nesta apresentação acessar o relatório completo do GitHub.

Essa base de dados tem muito potencial e é ideal para fazer análise de componentes principais, aplicar várias técnicas de agrupamento para agrupar os professores de acordo com o perfil de recomendação dos mesmos ou outros atributos. Eu não utilizei nenhuma dessas técnicas pois, elas dependem dos dados estarem completos, ou seja não ter informações faltantes. Nesta base de dados temos muitos casos no qual isso acontece, em 50 das 53 variáveis observamos informações faltantes.



Existem várias técnicas de imputações de dados para corrigir o problema. Essas técnicas são utilizadas depois de estudar cada tipo de variável e forma como a mesma foi coletada. Afim, de não viciar os dados e devido ao curto período de tempo optei por não utilizar a imputação dos dados, o que limita o leque de análises possíveis.

Obrigada !



Email : marinaaamorim@hotmail.com

Tel: (31) 99455-7615