

House of Excellence

Consultores Responsáveis:

Estatiano 1

Estatiano 2

...

Estatiano n

Requerente:

ESTAT

Brasília, 12 de outubro de 2024.



Sumário

	Página
1 Mudanças que estou ciente:	3
1.1 Ajeitar a planilha, estava usando apenas uma aba do banco de dados, agora vou juntar todas em uma só e transformar em apenas uma com as mesmas variáveis.	3
1.2 Não sei se todas as questões de fontes e cores estão corretas, gostaria de saber se tem algo que estou fazendo errado e que o pacote da estat não cobre e preciso fazer manualmente.	3
1.3 Suponde que o banco de dados estivesse correto, essa seria uma solução razoável para a questão 1?	3
2 Introdução	4
3 Referencial Teórico	5
4 Análises	6
4.1 Top 5 países com maior número de mulheres medalhistas	6
4.2 2. Valor IMC por esporte, estes sendo, ginástica, futebol, judô, atletismo e badminton (20/10)	6
4.3 3. Top 3 medalhistas gerais por quantidade de cada tipo de medalha (27/10)	7
5 Conclusões	8

1 Mudanças que estou ciente:

- 1.1 Ajeitar a planilha, estava usando apenas uma aba do banco de dados, agora vou juntar todas em uma só e transformar em apenas uma com as mesmas variáveis.**
- 1.2 Não sei se todas as questões de fontes e cores estão corretas, gostaria de saber se tem algo que estou fazendo errado e que o pacote da estat não cobre e preciso fazer manualmente.**
- 1.3 Suponde que o banco de dados estivesse correto, essa seria uma solução razoável para a questão 1?**

2 Introdução

O seguinte projeto tem como objetivo trazer uma análise descritiva e exploratória de dados fornecidos pela “House of Excellence”, visando otimizar o desempenho de seus atletas.

O banco de dados conta com 9 variáveis e 67.474 observações, sendo as mais importantes (tirando as informações básicas de cada atleta, como nome, peso e altura); “Sport”: Esporte praticado pelo atleta, “Event”: Nome do evento esportivo que o atleta participou, “Medal”: medalha adquirida pelo atleta na competição.

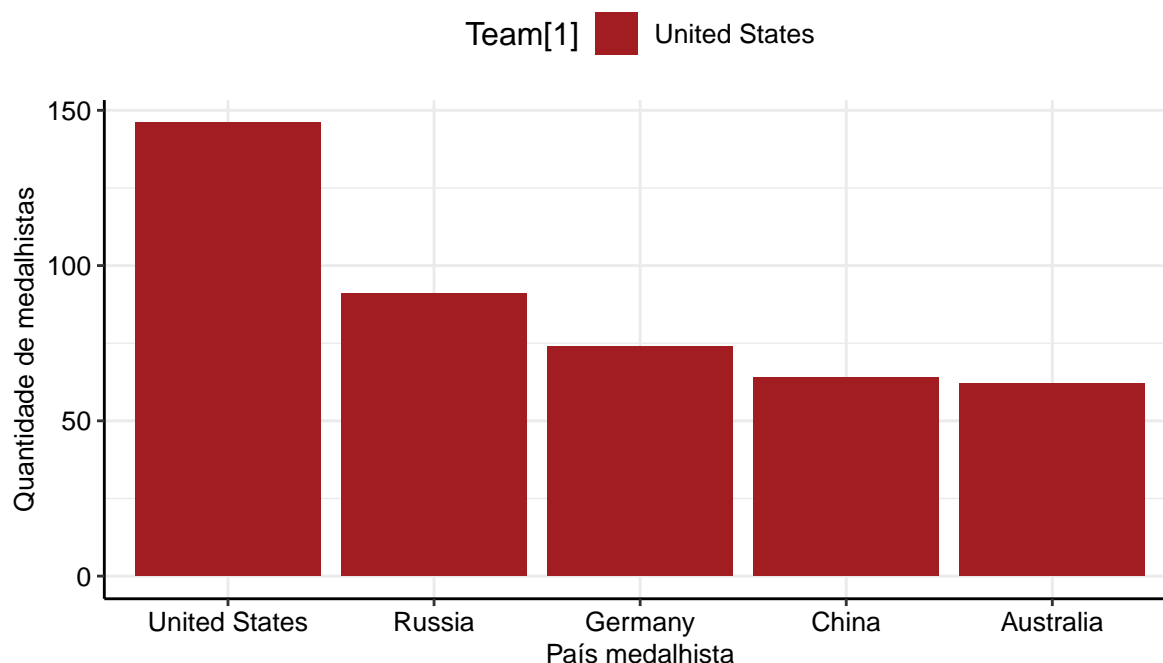
Todas as análises foram feitas utilizando o software RStudio

3 Referencial Teórico

?

4 Análises

4.1 Top 5 países com maior número de mulheres medalhistas



4.2 2. Valor IMC por esporte, estes sendo, ginástica, futebol, judô, atletismo e badminton (20/10)

Quantidade de “na’s” nas variáveis de altura e peso respectivamente

```
[1] 1013
```

```
[1] 1013
```

Como a quantidade de “na’s” é muito pequena, os “na’s” serão considerados um evento aleatório e as linhas serão removidas.

```
{r}
```

```
tabela1 <- imc_calculado %>% group_by(Sport) %>% summarise(media = round(mean(imc), 2), variancia = round(var(imc), 2), mediana = round(median(imc), 2))
```

```
tabela1 %>% ggplot(aes(x = "Sport", y = "media")) + geom_col() + theme_estat()
```

```
{r}
```

```
imc_calculado %>% ggplot(aes(x = reorder(Sport, imc, FUN = median, decreasing = T), y = imc, fill = Sport[1])) + geom_boxplot() + theme_estat() + labs(y = "IMC", x = "Esporte", fill = "Esporte")
```

4.3 3. Top 3 medalhistas gerais por quantidade de cada tipo de medalha (27/10)

```
{r} df <- df %>% mutate(Medal_value = case_when( Medal == "Gold" ~ 3, Medal ==  
"Silver" ~ 2, Medal == "Bronze" ~ 1 ))  
{r}  
top_medalhistas <- df %>% group_by(Names) %>% summarise(n = n(), ) %>% ar-  
range(desc(n)) %>% head()  
  
{r}  
  
top_nomes <- top_medalhistas %>%  
  filter(n == 4) %>%  
  select(Names) %>%  
  as_vector()  
  
terceiro <- df %>%  
  filter(Names %in% top_nomes) %>%  
  group_by(Names) %>%  
  summarise(Medal_value = sum(Medal_value)) %>%  
  arrange(desc(Medal_value)) %>%  
  head(1) %>%  
  select(Names)
```

5 Conclusões

?