---

title: "House of Excellence"

output-file: titulo do projeto

---

```{r}

library(kableExtra)

library(DT)

library(knitr)

library(pacman)

library(tidyverse)

library(ggplot2)

library(dplyr)

library(readxl)

library(plotly)

athenas = read\_excel("OlimpiadasPS.xlsx",

sheet = "Athina")

rj = read\_excel("OlimpiadasPS.xlsx",

sheet = "Rio de Janeiro")

londres = read\_excel("OlimpiadasPS.xlsx",

sheet = "London")

sydney = read\_excel("OlimpiadasPS.xlsx",

sheet = "Sydney")

pequim = read\_excel("OlimpiadasPS.xlsx",

sheet = "Beijing")

# Organizando o banco de dados

colnames(sydney) = c("Nome", "Sexo", "Idade", "Altura", "Peso", "Pais", "Esporte", "Evento", "Medalha")

colnames(rj) = c("Nome", "Sexo", "Idade", "Altura", "Peso", "Pais", "Esporte", "Evento", "Medalha")

colnames(londres) = c("Nome", "Sexo", "Idade", "Altura", "Peso", "Pais", "Esporte", "Evento", "Medalha")

colnames(pequim) = c("Nome", "Sexo", "Idade", "Altura", "Peso", "Pais", "Esporte", "Evento", "Medalha")

colnames(athenas) = c("Nome", "Sexo", "Idade", "Altura", "Peso", "Pais", "Esporte", "Evento", "Medalha")

primeiro = rbind(athenas,rj,londres,sydney,pequim)

olimpiadasG = subset(primeiro, Medalha== "Gold")

olimpiadasS = subset(primeiro, Medalha== "Silver")

olimpiadasB = subset(primeiro, Medalha== "Bronze")

olimpiadas = rbind(olimpiadasG,olimpiadasS,olimpiadasB)

estat\_colors <- c(

"#A11D21", "#003366", "#CC9900",

"#663333", "#FF6600", "#CC9966",

"#999966", "#006606", "#008091",

"#041835", "#666666" )

theme\_estat <- function(...) {

theme <- ggplot2::theme\_bw() +

ggplot2::theme(

axis.title.y = ggplot2::element\_text(colour = "black", size = 12),

axis.title.x = ggplot2::element\_text(colour = "black", size = 12),

axis.text = ggplot2::element\_text(colour = "black", size = 9.5)

,

panel.border = ggplot2::element\_blank(),

axis.line = ggplot2::element\_line(colour = "black"),

legend.position = "top",

...

)

return(

list(

theme,

scale\_fill\_manual(values = estat\_colors),

scale\_colour\_manual(values = estat\_colors)

)

)

}

```

```{r}

medalW = subset(olimpiadas, Sexo== "F")

medalW$Pais[medalW$Pais == "United States"] = "Estados Unidos"

medalW$Pais[medalW$Pais == "Germany"] = "Alemanha"

tabelamedalhasgeral= table(medalW$Pais)

top5medalhas = sort(tabelamedalhasgeral, decreasing = TRUE)

top5= data.frame(Team = names(head(top5medalhas, 5)),

Medalhas = as.numeric(head(top5medalhas, 5)))

ggplot(top5, aes(x = reorder(Team, -Medalhas), y = Medalhas)) +

geom\_bar(stat = "identity", fill = "#A11D21") +

geom\_text(aes(label = paste0(Medalhas, " (", round(Medalhas / sum(Medalhas) \* 100, 1), "%)")),

vjust = -0.5, size = 2.5) +

labs(title = "Top 5 países com mais medalhas entre as mulheres",

x = "Países", y = "Número de Medalhas") +

theme\_estat()

head( top5medalhas, 5)

```

```{r}

olimpiadas$Peso <- as.numeric(olimpiadas$Peso)

olimpiadas$pesokg = olimpiadas$Peso \* 0.453592

olimpiadas$alturam = olimpiadas$Altura / 100

olimpiadas$imc <- olimpiadas$pesokg / (olimpiadas$alturam^2)

judo = subset(olimpiadas, Esporte == "Judo")

badminton = subset(olimpiadas, Esporte == "Badminton")

futebol = subset(olimpiadas, Esporte == "Football")

ginastica = subset(olimpiadas, Esporte == "Gymnastics")

atletismo = subset(olimpiadas, Esporte == "Athletics")

espotes=rbind(judo,badminton,futebol,ginastica,atletismo)

espotes$Esporte[espotes$Esporte == "Football"] = "Futebol"

espotes$Esporte[espotes$Esporte == "Gymnastics"] = "Ginastica"

espotes$Esporte[espotes$Esporte == "Athletics"] = "Atletismo"

ggplot(espotes) +

aes(x = reorder(Esporte, imc, FUN = median), y = imc) +

geom\_boxplot(aes(fill = "Mediana"), width = 0.5) +

stat\_summary(

aes(shape = "Média"), fun = "mean", geom = "point", size = 3, fill = "white"

) +

labs(x = "Esporte", y = "IMC", title = "IMC POR ESPORTE") +

scale\_shape\_manual(values = c(23)) +

scale\_fill\_manual(values = c("Mediana" = "#A11D21")) +

guides(fill = "none", shape = "none") +

theme\_estat()

```

```{r, echo=FALSE, results='asis'}

cat("\\begin{quadro}[H]

\\caption{Medidas resumo dos esportes}

\\centering

\\begin{tabular}{| l |

S[table-format = 2.2]

S[table-format = 1.2]

S[table-format = 2.2]

S[table-format = 2.2]

S[table-format = 2.2]

|}

\\toprule

\\textbf{Estatística} & \\textbf{Atletismo} & \\textbf{Badminton} & \\textbf{Futebol} & \\textbf{Ginastica} & \\textbf{Judo} \\\\

\\midrule

Média & 22.30 & 22.21 & 22.51 & 20.68 & 25.70 \\\\

Desvio Padrão & 3.86 & 1.50 & 1.73 & 2.38 & 5.12 \\\\

Variância & 14.92 & 2.26 & 2.99 & 5.67 & 26.23 \\\\

Mínimo & 15.82 & 18.94 & 16.73 & 15.16 & 18.52 \\\\

1º Quartil & 20.03 & 21.22 & 21.34 & 18.61 & 22.06 \\\\

Mediana & 21.45 & 22.28 & 22.49 & 21.09 & 24.68 \\\\

3º Quartil & 23.67 & 23.21 & 23.71 & 22.48 & 27.70 \\\\

Máximo & 44.38 & 26.73 & 29.07 & 26.45 & 56.50 \\\\

\\bottomrule

\\end{tabular}

\\label{quad:quadro\_resumo1}

\\end{quadro}

")

```

```{r}

olimpiadas$Medalha[olimpiadas$Medalha == "Gold"] = "Ouro"

olimpiadas$Medalha[olimpiadas$Medalha == "Silver"] = "Prata"

ouro = subset(olimpiadas, Medalha == "Ouro")

prata = subset(olimpiadas, Medalha == "Prata")

bronze = subset(olimpiadas, Medalha == "Bronze")

medalhaalltime = rbind(ouro,prata,bronze)

michaelphelps = subset(medalhaalltime,Nome == "Michael Fred Phelps, II")

natalie = subset(medalhaalltime, Nome == "Natalie Anne Coughlin (-Hall)")

ryan = subset(medalhaalltime,Nome == "Ryan Steven Lochte")

top3= rbind(michaelphelps, natalie, ryan)

```

```{r}

library(dplyr)

library(ggplot2)

library(forcats)

library(stringr)

# Adaptando o código para o novo banco de dados e variáveis

trans\_drv <- top3 %>%

mutate(Nome = case\_when(

str\_detect(Nome, "auto") ~ "auto",

str\_detect(Nome, "manual") ~ "manual",

TRUE ~ Nome

)) %>%

group\_by(Nome, Medalha) %>%

summarise(freq = n(), .groups = "drop") %>%

mutate(

freq\_relativa = round(freq / sum(freq) \* 100, 1)

)

porcentagens <- str\_c(trans\_drv$freq\_relativa, "%") %>% str\_replace("\\.", ",")

legendas <- str\_squish(str\_c(trans\_drv$freq, " (", porcentagens, ")"))

ggplot(trans\_drv) +

aes(

x = fct\_reorder(Nome, freq, .desc = TRUE),

y = freq,

fill = Medalha,

label = legendas

) +

geom\_col(position = position\_dodge2(preserve = "single", padding = 0)) +

geom\_text(

position = position\_dodge(width = 0.9),

vjust = -0.5, hjust = 0.5,

size = 3

) +

labs(x = "Nome do Atleta", y = "Quantidade de Medalhas") +

theme\_estat() +

ggtitle("Quantidade de Medalhas por Atleta e Tipo de Medalha")

```

```{r}

ggplot(olimpiadas, aes(x = pesokg, y = Altura)) +

geom\_point(color = "#A11D21") +

labs(

title = "Altura x Peso",

x = "Peso",

y = "Altura",

color = "Legenda"

) +

theme\_estat()

correlacao <- cor(olimpiadas$pesokg, olimpiadas$Altura, use = "complete.obs")

```