~~~
---
title: "BF"
author: "Atuti"
date: "2025-12-11"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

## importing and exploring data
## 3000 participants
## There are missing valus/blanks on partcipant ids

```{r}
BF <- read_csv("fitness data capstone.csv")
view(BF)
 BF %>%
   distinct(participant_id) %>%
   count()
 BF %>%
   distinct(participant_id) %>%
   drop_na()
  BF
   colnames(BF)
   ## table

BF%>%
  select(participant_id,age,gender,hydration_level,stress_level,daily_steps, fitness_level)%>%
  distinct(participant_id, age, gender, fitness_level, stress_level) %>%
table()


BF%>%
  distinct(participant_id, age, gender,calories_burned, fitness_level, stress_level) %>%
  view()

 ## removing duplicates data
BF_single <- BF %>%
  group_by(participant_id, activity_type) %>%
  slice_max(calories_burned, n = 1) %>%
  ungroup()

view(BF_single)

summary(BF_single)

## visuals

BF_single %>%
mean(calories_burned) %>%
  ggplot()+ geom_bar(fill=activity_type)

avg_calories <- BF_single %>%
  drop_na(calories_burned) %>%
  group_by(activity_type) %>%
  summarise(avg_calories = mean(calories_burned, na.rm = TRUE)) %>%
  ungroup()
view(avg_calories)

avg_calories %>%
ggplot(avg_calories, aes(x=fct_reorder(activity_type, avg_calories, .desc = FALSE),, y=
avg_calories, fill= activity_type))+ geom_col() + labs( title = "Acitivity type vs calories
~~~

```
burned", x = "Activity", y ="Calories")

avg_steps <- BF_single %>%
  drop_na(daily_steps) %>%
  group_by(stress_level) %>%
  summarise(avg_steps = mean(daily_steps, na.rm = TRUE)) %>%
  ungroup
view(avg_steps)
ggplot(avg_steps, aes(x=stress_level, y=avg_steps, fill= stress_level,)) +geom_col()+labs(title =
"Steps vs stress_level", x = "Stress", y= "Steps")

## Hours_sleep vs stress

avg_sleep <- BF_single %>%
  drop_na(hours_sleep) %>%
  group_by(stress_level) %>%
  summarise(avg_sleep = mean(hours_sleep,na.rm =TRUE)) %>%
  ungroup()
 view(avg_sleep)

ggplot(avg_sleep, aes(x=stress_level, y= avg_sleep) + geom_col()+ labs(title = "Sleep's impact on
stress levels", x = "stress")

fitness<- BF_single %>%
  drop_na(fitness_level) %>%
  group_by(health_condition) %>%
  summarise(fitness = mean(fitness_level)) %>%
  ungroup()
view(fitness)
   ggplot(fitness, aes(x= health_condition, y = fitness, fill= fitness))+geom_col(width =1) +
coord_polar(theta="y") + labs(title = "Fitness per health condition") + theme_void()

   ##donut chart
    ggplot(fitness, aes(x= 2, y = fitness, fill= health_condition))+geom_col(width =1) +
coord_polar(theta="y") + xlim(0.5, 2.5)+ labs(title = "Fitness per health condition") +
theme_void()

##hydration level vs stress

  BF_hy <-BF_single %>%
  select(participant_id, hydration_level,stress_level) %>%
  group_by(participant_id) %>%
  arrange(desc(hydration_level), stress_level) %>%
  slice(1) %>%
  ungroup()
  view(BF_hy)

  ## grouping average hydration per stress

  BF_hy_clean <- BF_hy %>%
  drop_na(hydration_level) %>%
  group_by(stress_level) %>%
  filter(
    hydration_level >= (quantile(hydration_level, 0.25) - 1.5 * IQR(hydration_level)) &
    hydration_level <= (quantile(hydration_level, 0.75) + 1.5 * IQR(hydration_level))
  ) %>%
  ungroup()
  view(BF_hy_clean)


  hydration <- BF_hy_clean %>%
    drop_na(hydration_level) %>%
    group_by(stress_level) %>%
    summarise(hydration = mean(hydration_level)) %>%
    ungroup()
```

~~~
  view(hydration)
   ggplot(hydration, aes(x=stress_level, y= hydration))+ geom_line()+ geom_point()+labs(title =
"Hydration vs Stress level", x = "Stress", y = "Mean hydration")

   ggplot(BF_hy_clean, aes(x=stress_level, y= hydration_level))+ geom_smooth()+
     labs(title= "Hydration vs Stress on entire population", x ="stress", y = "Hydration")
 ##smokers
   ggplot(BF_single, aes(x=smoking_status, fill= health_condition)) + geom_bar()+ labs(title =
"Smoking status")+ facet_wrap(~gender)
   ## age distrubution

  age <- BF %>%
  select(participant_id, age, gender) %>%
    group_by(partcipant_id) %>%
    slice_sample(n=1) %>%
  ungroup()

   view(age)


```{r pressure, echo=FALSE}
plot(pressure)
```
~~~