Final Project

Due Wednesday, December 3, 11:59 PM

Danielle, Beckett, and Alexia

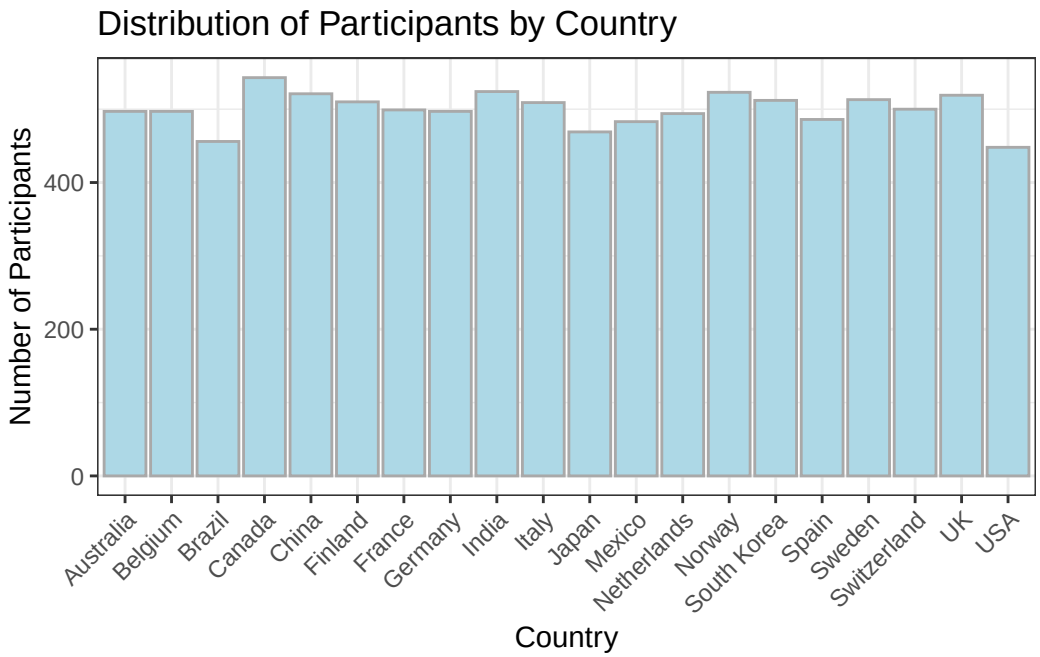
Use of AI tools

I did not use AI in the completion of this assignment.

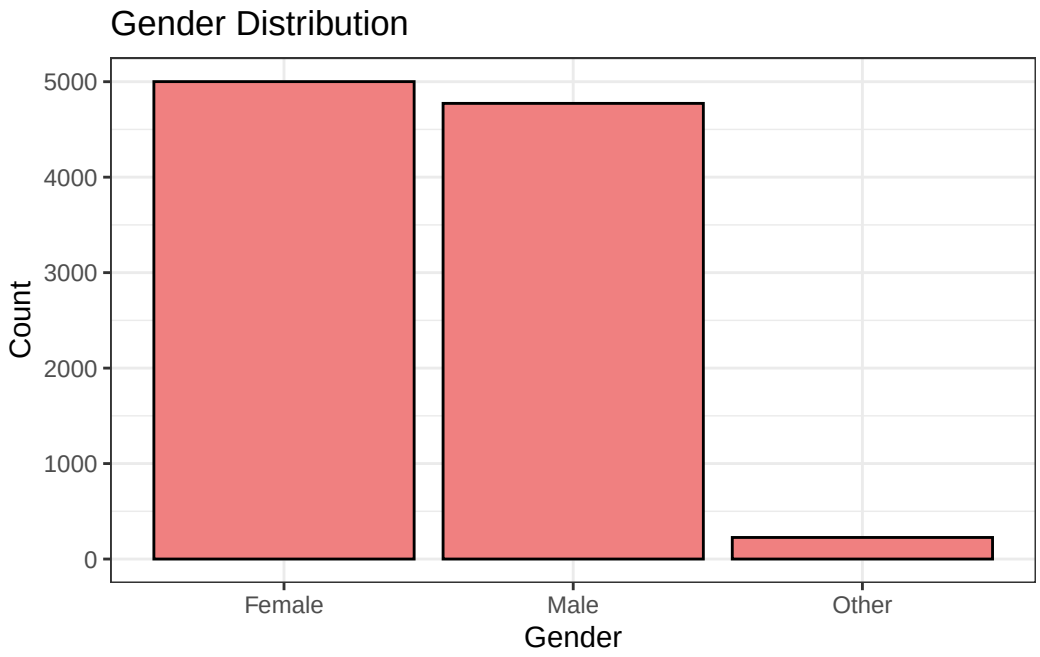
Loading Libraries/Dataset

Visualizations of Patients' Demographic

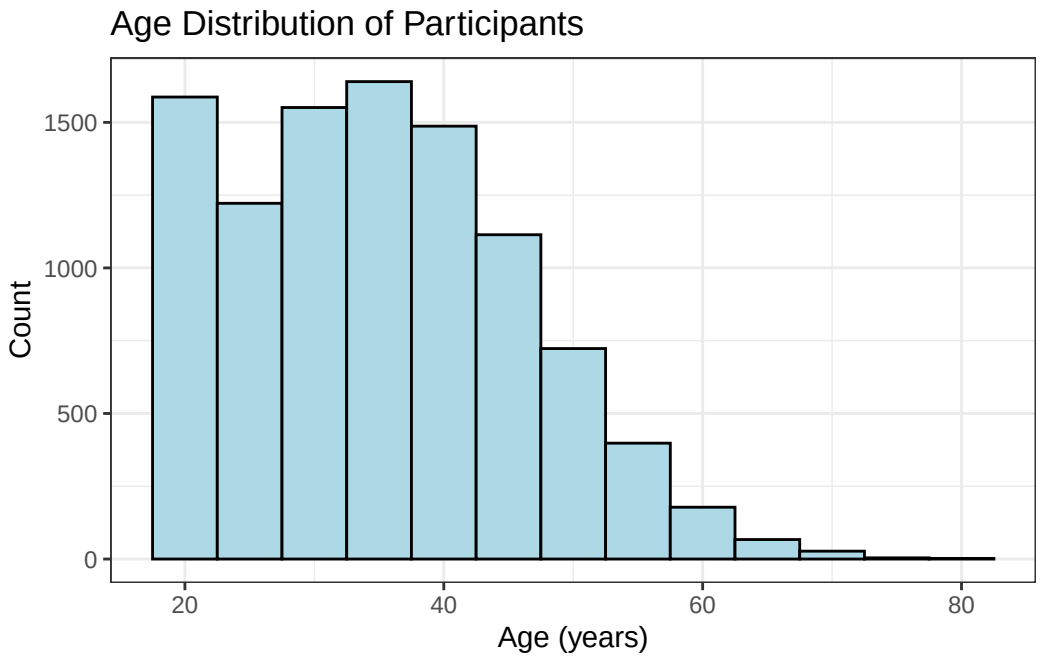
```
# m <- lm(Sleep_Hours ~ ., data = data)

# Distribution of participants by country
ggplot(data, aes(x = Country)) +
  geom_bar(color = "darkgrey", fill = "lightblue") +
  labs(
    x = "Country",
    y = "Number of Participants",
    title = "Distribution of Participants by Country"
  ) +
  theme_bw() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# gender distribution
ggplot(data, aes(x = Gender)) +
  geom_bar(fill = "lightcoral", color = "black") +
  labs(title = "Gender Distribution", y = "Count", x = "Gender") +
  theme_bw()
```

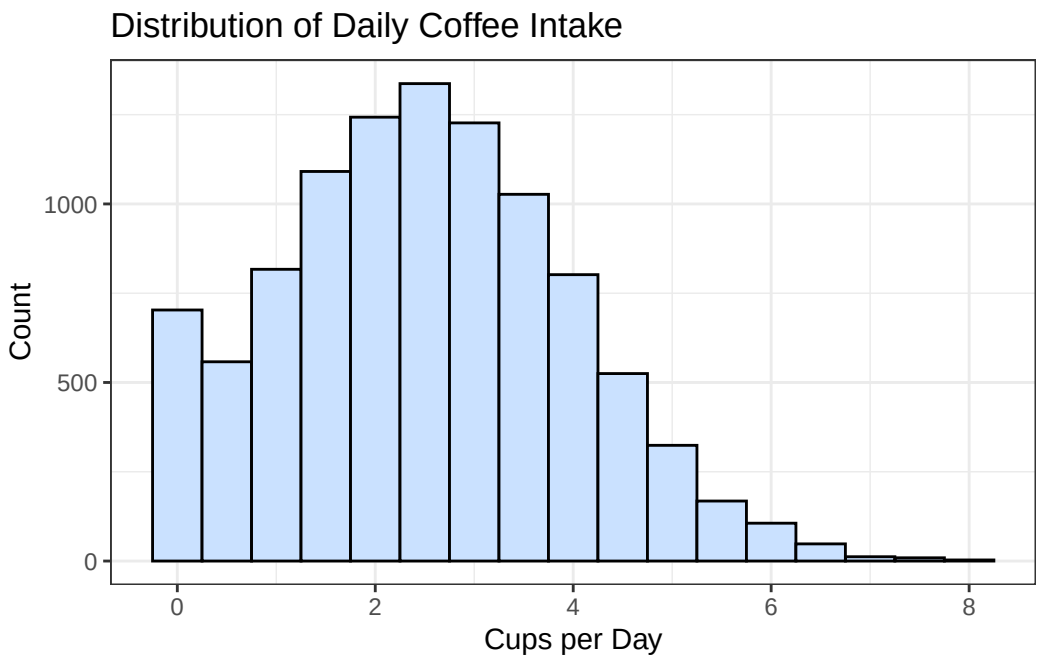



```
# age distribution
ggplot(data, aes(x = Age)) +
  geom_histogram(binwidth = 5, fill = "lightblue", color = "black") +
  labs(title = "Age Distribution of Participants", x = "Age (years)", y = "Count") +
  theme_bw()
```



Visualizations of Coffee Consumption

```r
# Coffee Intake Distribution
ggplot(data, aes(x = Coffee_Intake)) +
  geom_histogram(binwidth = 0.5, fill = "lightsteelblue1", color = "black") +
  labs(title = "Distribution of Daily Coffee Intake", x = "Cups per Day", y = "Count") +
  theme_bw()
```



Models

```
#BMI, Heart rate, sleep quality, and stress levels
m2 <- glm(high_caffeine ~ BMI + factor(Sleep_Quality) + Heart_Rate + factor(Stress_Level),
          data = data,
          family = "binomial")
summary(m2)
```

```
Call:
glm(formula = high_caffeine ~ BMI + factor(Sleep_Quality) + Heart_Rate +
    factor(Stress_Level), family = "binomial", data = data)

Coefficients: (2 not defined because of singularities)
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                -3.3092014  0.3108830 -10.645  < 2e-16 ***
BMI                        -0.0002889  0.0078145  -0.037  0.97051
factor(Sleep_Quality)Fair   0.9813082  0.1265191   7.756 8.75e-15 ***
factor(Sleep_Quality)Good   0.6120603  0.1186116   5.160 2.47e-07 ***
factor(Sleep_Quality)Poor   1.3318303  0.1364594   9.760  < 2e-16 ***
Heart_Rate                  0.0091048  0.0031016   2.935  0.00333 **
```

```
factor(Stress_Level)Low            NA         NA      NA       NA
factor(Stress_Level)Medium         NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7519.8  on 9999  degrees of freedom
Residual deviance: 7380.3  on 9994  degrees of freedom
AIC: 7392.3

Number of Fisher Scoring iterations: 5
```

```r
table(data$Stress_Level, data$high_caffeine)
```

```
           0    1
  High    761  200
  Low    6264  725
  Medium 1729  321
```