

poster3.R

17ada

2024-01-15

```
chooseCRANmirror(ind=1)
install.packages("vcd")
```

```
## Installing package into 'C:/Users/17ada/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'vcd' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\17ada\AppData\Local\Temp\RtmpMJW7wJ\downloaded_packages
```

```
install.packages("forcats")
```

```
## Installing package into 'C:/Users/17ada/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'forcats' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\17ada\AppData\Local\Temp\RtmpMJW7wJ\downloaded_packages
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```
library(forcats)
```

```
## Warning: package 'forcats' was built under R version 4.3.2
```

```
library(vcd)
```

```
## Warning: package 'vcd' was built under R version 4.3.2
```

```
## Loading required package: grid
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.2
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##     filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##     intersect, setdiff, setequal, union
```

```
file_path = "C:\\\\Users\\\\17ada\\\\Desktop\\\\R\\\\asrData.csv"

org_data = read.csv(file_path)
data1 = org_data

new_col_names = c(
  "Timestamp" = "SurveyTimestamp",
  "Email Address" = "Email",
  "Your age" = "Age",
  "Gender" = "Gender",
  "Your academic major or field of study?" = "Major",
  "What is your current year of study?" = "YearOfStudy",
  "What is your current living situation?" = "LivingSituation",
  "Do you have any chronic medical condition?" = "ChronicCondition",
  "What is your current employment status?" = "EmploymentStatus",
  "Do you have any specific goals related to your lifestyle (e.g., weight management, stress reduction, etc.)?" = "LifestyleGoals",
  "How many hours of sleep do you typically get per night on an average?" = "HoursOfSleep",
  "Please select the relevant option [How often do you engage in physical activity or exercise per week?]" = "PhysicalActivityFrequency",
  "Please select the relevant option [How often do you go to dine out or order food from outside?]" = "DiningOutFrequency",
  "Please select the relevant option [How often do you smoke or indulge in tobacco products?]" = "SmokingFrequency",
  "Please select the relevant option [How frequently do you consume alcohol or engage in binge drinking?]" = "AlcoholFrequency",
  "How much time do you spend on screens per day(TV, computer, smartphone?)" = "ScreenTime",
  "On a scale of 1 to 5, choose the relevant answer. [I am influenced by peer pressure or social norms when making my lifestyle choices?]" = "PeerInfluence",
  "On a scale of 1 to 5, choose the relevant answer. [Factors, such as my budget or financial situation, affect my lifestyle choices?]" = "FinancialInfluence",
  "On a scale of 1 to 5, choose the relevant answer. [I feel that academic demands and stress impact my ability to maintain a healthy lifestyle?]" = "AcademicStress",
  "On a scale of 1 to 5, choose the relevant answer. [I have received support or guidance related to my lifestyle choices (e.g., from health services, counseling, or friends)?]" = "LifestyleSupport",
  "On a scale of 1 to 5, choose the relevant answer. [I am aware of the impact of my lifestyle choices on my academic performance, and my behavior?]" = "AwarenessImpact",
  "Which of the following strategies do you use to balance academic pressures and a healthy lifestyle?" = "BalancingStrategy"
)

dim(data1)
```

```
## [1] 143 22
```

```
str(data1)
```

```

## 'data.frame': 143 obs. of 22 variables:
## $ Timestamp
: chr "11/2/2023 18:36:08" "11/9/2023 19:06:40" "11/10/2023 10:09:09" "11/10/2023 18:31:52"
...
## $ Email.Address
: chr "ravi7896@gmail.com" "raj7733@gmail.com" "roshap0199@gmail.com" "mushabbar.ahmed@mca.
christuniversity.in" ...
## $ Your.age
: chr "21-25" "21-25" "25-35" "18-21" ...
## $ Gender
: chr "Male" "Female" "Male" "Male" ...
## $ Your.academic.major.or.field.of.study.
: chr "UG" "PG" "UG" "PG" ...
## $ What.is.your.current.year.of.study.
: chr "1" "1" "4" "1" ...
## $ What.is.your.current.living.situation.
: chr "With family" "PG" "With family" "With family" ...
## $ Do.you.have.any.chronic.medical.condition.
: chr "No" "No" "Yes" "No" ...
## $ What.is.your.current.employment.status.
: chr "Unemployed" "Unemployed" "Part Time Job/ Internship" "Prefer not to say" ...
## $ Do.you.have.any.specfic.goals.related.to.your.lifestyle..e.g...weight.management..stre
ss.reduction..etc... : chr
"No" "No" "Yes" "Yes" ...
## $ How.many.hours.of.sleep.do.you.typically.get.per.night.on.an.average.
: chr "6-8" "6-8" "More than 8" "6-8" ...
## $ Please.select.the.relevant.option..How.often.do.you.engage.in.physical.activity.or.exer
cise.per.week.. : chr
"Less than 3 days per week" "Less than 3 days per week" "More than 5 days a week" "3 to 5 day
s per week" ...
## $ Please.select.the.relevant.option..How.often.do.you.go.to.dine.out.or.order.food.from.o
utside.. : chr
"3 to 5 days per week" "3 to 5 days per week" "More than 5 days a week" "3 to 5 days per wee
k" ...
## $ Please.select.the.relevant.option..How.often.do.you.smoke.or.indulge.in.tobacco.product
s.. : chr
"Never" "Never" "More than 5 days a week" "Never" ...
## $ Please.select.the.relevant.option..How.frequently.do.you.consume.alcohol.or.engage.in.b
inge.drinking.. : chr
"Less than 3 days per week" "Less than 3 days per week" "More than 5 days a week" "Never" ...
## $ How.much.time.do.you.spend.on.screens.per.day.TV..computer..smartphone..
: chr "3 to 6 hours" "3 to 6 hours" "More than 9 hours" "3 to 6 hours" ...
## $ On.a.scale.of.1.to.5..choose.the.relevant.answer...I.am.influenced.by.peer.pressure.or.
social.norms.when.making.my..lifestyle.choices.. : chr
"Neutral" "Neutral" "Strongly Disagree" "Neutral" ...
## $ On.a.scale.of.1.to.5..choose.the.relevant.answer...Factors..such.as.my.budget.or.financ
ial.situation..affect.my.lifestyle.choices.. : chr
"Agree" "Agree" "Neutral" "Neutral" ...
## $ On.a.scale.of.1.to.5..choose.the.relevant.answer...I.feel.that.academic.demands.and.str
ess.impacts.my.ability.to.maintain.a.healthy.lifestyle.. : chr
"Agree" "Agree" "Agree" "Agree" ...
## $ On.a.scale.of.1.to.5..choose.the.relevant.answer...I.have..received.support.or.guidanc
e.related.to.my.lifestyle.choices..e.g...from.health.services..counseling..or.friends...: chr
"Neutral" "Neutral" "Agree" "Agree" ...
## $ On.a.scale.of.1.to.5..choose.the.relevant.answer...I.am..aware.of.the.impact.of.my.life

```

```
style.choices.on.my.academic.performance..and.my.behavior.. : chr
"Agree" "Agree" "Agree" "Strongly Agree" ...
## $ Which.of.the.following.strategies.do.you.use.to.balance.academic.pressures.and.a.health
y.lifestyle. : chr
"Regular exercise" "Setting specific academic goals" "Setting specific academic goals" "Regul
ar exercise and Healthy eating habits" ...
```

```
summary(data1)
```

```
##   Timestamp      Email.Address      Your.age      Gender
## Length:143      Length:143      Length:143      Length:143
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
## Your.academic.major.or.field.of.study. What.is.your.current.year.of.study.
## Length:143                  Length:143
## Class :character          Class :character
## Mode  :character          Mode  :character
## What.is.your.current.living.situation.
## Length:143
## Class :character
## Mode  :character
## Do.you.have.any.chronic.medical.condition.
## Length:143
## Class :character
## Mode  :character
## What.is.your.current.employment.status.
## Length:143
## Class :character
## Mode  :character
## Do.you.have.any.specfic.goals.related.to.your.lifestyle..e.g...weight.management..stres
s.reduction..etc...
## Length:143
## Class :character
## Mode  :character
## How.many.hours.of.sleep.do.you.typically.get.per.night.on.an.average.
## Length:143
## Class :character
## Mode  :character
## Please.select.the.relevant.option..How.often.do.you.engage.in.physical.activity.or.exerci
se.per.week..
## Length:143
## Class :character
## Mode  :character
## Please.select.the.relevant.option..How.often.do.you.go.to.dine.out.or.order.food.from.out
side..
## Length:143
## Class :character
## Mode  :character
## Please.select.the.relevant.option..How.often.do.you.smoke.or.indulge.in.tobacco.product
s..
## Length:143
## Class :character
## Mode  :character
## Please.select.the.relevant.option..How.frequently.do.you.consume.alcohol.or.engage.in.bri
ge.drinking..
## Length:143
## Class :character
## Mode  :character
## How.much.time.do.you.spend.on.screens.per.day.TV..computer..smartphone..
## Length:143
## Class :character
## Mode  :character
## On.a.scale.of.1.to.5..choose.the.relevant.answer...I.am.influenced.by.peer.pressure.or.so
cial.norms.when.making.my..lifestyle.choices..
```

```
## Length:143
## Class :character
## Mode  :character
## On.a.scale.of.1.to.5..choose.the.relevant.answer...Factors..such.as.my.budget.or.financia
l.situation..affect.my.lifestyle.choices..
## Length:143
## Class :character
## Mode  :character
## On.a.scale.of.1.to.5..choose.the.relevant.answer...I.feel.that.academic.demands.and.stres
s.impacts.my.ability.to.maintain.a.healthy.lifestyle..
## Length:143
## Class :character
## Mode  :character
## On.a.scale.of.1.to.5..choose.the.relevant.answer...I.have..received.support.or.guidance.r
elated.to.my.lifestyle.choices..e.g...from.health.services..counseling..or.friends...
## Length:143
## Class :character
## Mode  :character
## On.a.scale.of.1.to.5..choose.the.relevant.answer...I.am..aware.of.the.impact.of.my.lifest
yle.choices.on.my.academic.performance..and.my.behavior..
## Length:143
## Class :character
## Mode  :character
## Which.of.the.following.strategies.do.you.use.to.balance.academic.pressures.and.a.healthy.
lifestyle.
## Length:143
## Class :character
## Mode  :character
```

```
head(data1)
```

```

##           Timestamp          Email.Address      Your.age
## 1 11/2/2023 18:36:08      ravi7896@gmail.com    21-25
## 2 11/9/2023 19:06:40      raj7733@gmail.com    21-25
## 3 11/10/2023 10:09:09     roshanp0199@gmail.com 25-35
## 4 11/10/2023 18:31:52 mushabbar.ahmed@mca.christuniversity.in 18-21
## 5 11/11/2023 20:05:31      shaizk77@gmail.com   25 and Above
## 6 11/11/2023 22:19:12      paulankita562@gmail.com 21-25
##   Gender Your.academic.major.or.field.of.study.
## 1   Male                 UG
## 2 Female                PG
## 3   Male                 UG
## 4   Male                 PG
## 5   Male                 UG
## 6 Female                PG
##   What.is.your.current.year.of.study. What.is.your.current.living.situation.
## 1                   1             With family
## 2                   1                  PG
## 3                   4             With family
## 4                   1             With family
## 5            Other             With family
## 6                   1             With family
##   Do.you.have.any.chronic.medical.condition.
## 1                   No
## 2                   No
## 3                   Yes
## 4                   No
## 5                   No
## 6                   No
##   What.is.your.current.employment.status.
## 1           Unemployed
## 2           Unemployed
## 3 Part Time Job/ Internship
## 4 Prefer not to say
## 5 Part Time Job/ Internship
## 6           Unemployed
##   Do.you.have.any.specific.goals.related.to.your.lifestyle..e.g...weight.management..stres
## s.reduction..etc...
## 1
No
## 2
No
## 3
Yes
## 4
Yes
## 5
Yes
## 6
Yes
##   How.many.hours.of.sleep.do.you.typically.get.per.night.on.an.average.
## 1                   6-8
## 2                   6-8
## 3                   More than 8
## 4                   6-8
## 5                   6-8

```

6-8

```

## 6
## Please.select.the.relevant.option..How.often.do.you.engage.in.physical.activity.or.exerc
ise.per.week..
## 1                                         Less than 3
days per week
## 2                                         Less than 3
days per week
## 3                                         More than
5 days a week
## 4                                         3 to 5
days per week
## 5                                         3 to 5
days per week
## 6                                         Less than 3
days per week
## Please.select.the.relevant.option..How.often.do.you.go.to.dine.out.or.order.food.from.ou
tside..
## 1                                         3 to 5 days p
er week
## 2                                         3 to 5 days p
er week
## 3                                         More than 5 days
a week
## 4                                         3 to 5 days p
er week
## 5                                         Less than 3 days p
er week
## 6                                         Less than 3 days p
er week
## Please.select.the.relevant.option..How.often.do.you.smoke.or.indulge.in.tobacco.product
s..
## 1                                         Nev
er
## 2                                         Nev
er
## 3                                         More than 5 days a we
ek
## 4                                         Nev
er
## 5                                         Nev
er
## 6                                         Nev
er
## Please.select.the.relevant.option..How.frequently.do.you.consume.alcohol.or.engage.in.bi
nge.drinking..
## 1                                         Less than 3
days per week
## 2                                         Less than 3
days per week
## 3                                         More than
5 days a week
## 4
Never
## 5
Never
## 6

```

```
Never
## How.much.time.do.you.spend.on.screens.per.day.TV..computer..smartphone..
## 1 3 to 6 hours
## 2 3 to 6 hours
## 3 More than 9 hours
## 4 3 to 6 hours
## 5 6 to 9 hours
## 6 6 to 9 hours
## On.a.scale.of.1.to.5..choose.the.relevant.answer...I.am.influenced.by.peer.pressure.or.social.norms.when.making.my..lifestyle.choices..
## 1
Neutral
## 2
Neutral
## 3
Strongly Disagree
## 4
Neutral
## 5
Neutral
## 6
Agree
## On.a.scale.of.1.to.5..choose.the.relevant.answer...Factors..such.as.my.budget.or.financial.situation..affect.my.lifestyle.choices..
## 1
Agree
## 2
Agree
## 3
Neutral
## 4
Neutral
## 5
Strongly Agree
## 6
Neutral
## On.a.scale.of.1.to.5..choose.the.relevant.answer...I.feel.that.academic.demands.and.stress.impacts.my.ability.to.maintain.a.healthy.lifestyle..
## 1
Agree
## 2
Agree
## 3
Agree
## 4
Agree
## 5
Agree
## 6
Agree
## On.a.scale.of.1.to.5..choose.the.relevant.answer...I.have..received.support.or.guidance.related.to.my.lifestyle.choices..e.g...from.health.services..counseling..or.friends...
## 1
Neutral
## 2
Neutral
```

```

## 3
Agree
## 4
Agree
## 5
Strongly Agree
## 6
Strongly Agree
## On.a.scale.of.1.to.5..choose.the.relevant.answer...I.am..aware.of.the.impact.of.my.lifes
tyle.choices.on.my.academic.performance..and.my.behavior..
## 1
Agree
## 2
Agree
## 3
Agree
## 4
Strongly Agree
## 5
Strongly Agree
## 6
Strongly Agree
## Which.of.the.following.strategies.do.you.use.to.balance.academic.pressures.and.a.health
y.lifestyle.
## 1                               Regul
ar exercise
## 2                               Setting specific aca
demic goals
## 3                               Setting specific aca
demic goals
## 4                               Regular exercise and Healthy ea
ting habits
## 5                               Regular exercise and Healthy ea
ting habits
## 6                               Time management
techniques

```

```

colnames(data1) = new_col_names
data1$Major = as.factor(data1$Major)
data1$Gender = as.factor(data1$Gender)
data1$Age = as.factor(data1$Age)
data1$LivingSituation = as.factor(data1$LivingSituation)
data1$ChronicCondition = as.factor(data1$ChronicCondition)
data1$EmploymentStatus = as.factor(data1$EmploymentStatus)
data1$PhysicalActivityFrequency = as.factor(data1$PhysicalActivityFrequency)
data1$HoursOfSleep = as.factor(data1$HoursOfSleep)
data1$AcademicStress = as.factor(data1$AcademicStress)
data1$AlcoholFrequency = as.factor(data1$AlcoholFrequency)
data1$EmploymentStatus = as.factor(data1$EmploymentStatus)

sleep_age_table = table(data1$Age, data1$HoursOfSleep)
data1$AlcoholFrequency

```

```
## [1] Less than 3 days per week Less than 3 days per week
## [3] More than 5 days a week Never
## [5] Never Never
## [7] Never Never
## [9] Less than 3 days per week Never
## [11] Never Never
## [13] Never Never
## [15] Less than 3 days per week Never
## [17] Never Never
## [19] Never Less than 3 days per week
## [21] Less than 3 days per week Never
## [23] Never Never
## [25] Never Never
## [27] Never Never
## [29] Never Never
## [31] Less than 3 days per week Never
## [33] Less than 3 days per week Less than 3 days per week
## [35] Less than 3 days per week Never
## [37] Never Never
## [39] Never Never
## [41] Never Never
## [43] Never Never
## [45] Never Never
## [47] More than 5 days a week Never
## [49] Never Never
## [51] 3 to 5 days per week Never
## [53] More than 5 days a week 3 to 5 days per week
## [55] Never Never
## [57] Never Never
## [59] Never Never
## [61] Never Never
## [63] Never Never
## [65] Never Never
## [67] Less than 3 days per week Never
## [69] Less than 3 days per week Never
## [71] Never Never
## [73] Never 3 to 5 days per week
## [75] Never Never
## [77] Never Never
## [79] Never Never
## [81] Never Never
## [83] Never Never
## [85] Less than 3 days per week Never
## [87] Never Never
## [89] 3 to 5 days per week Less than 3 days per week
## [91] Never Never
## [93] Never Never
## [95] Never Never
## [97] Less than 3 days per week Never
## [99] Less than 3 days per week Never
## [101] Never Never
## [103] Never Never
## [105] Less than 3 days per week Never
## [107] Never Never
## [109] Never Never
```

```

## [111] Never           Never
## [113] Never           Never
## [115] Never           Never
## [117] Never           Never
## [119] Never           Never
## [121] Never           Never
## [123] 3 to 5 days per week Less than 3 days per week
## [125] Less than 3 days per week More than 5 days a week
## [127] Never           Never
## [129] Never           Less than 3 days per week
## [131] Never           Never
## [133] Never           Never
## [135] Never           Less than 3 days per week
## [137] Never           3 to 5 days per week
## [139] More than 5 days a week Never
## [141] Less than 3 days per week More than 5 days a week
## [143] Never           Never
## 4 Levels: 3 to 5 days per week ... Never

```

```

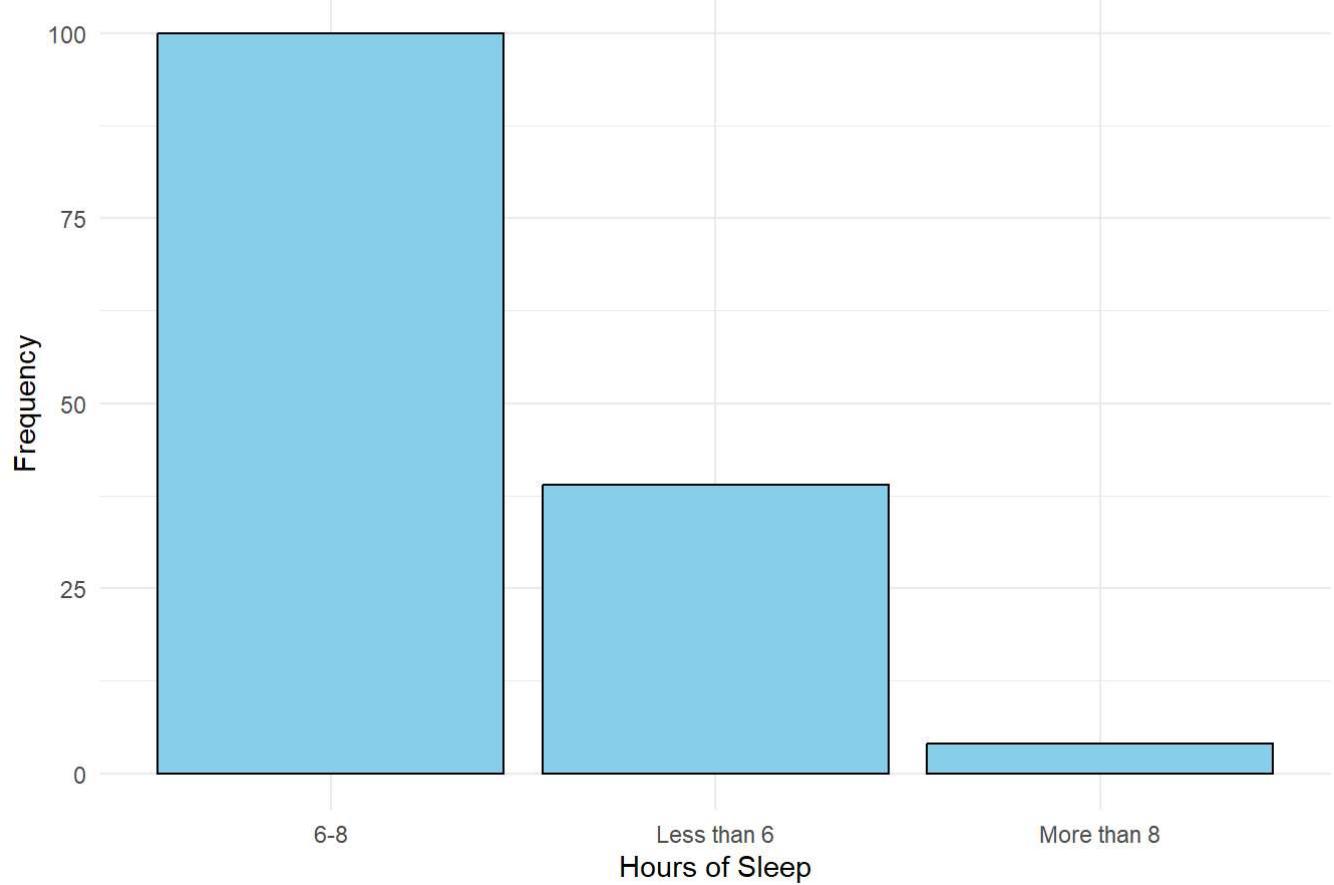
# Dropping timestamp and email
data = data1[, -c(1, 2)]

custom_colors <- c("skyblue", "lightgreen", "coral", "purple")

# Histogram for HoursOfSleep
# Bar plot for Distribution of Hours of Sleep with custom colors
ggplot(data1, aes(x = HoursOfSleep)) +
  geom_bar(fill = custom_colors[1], color = "black") + # Use the first custom color
  labs(title = "Distribution of Hours of Sleep", x = "Hours of Sleep", y = "Frequency") +
  theme_minimal()

```

Distribution of Hours of Sleep

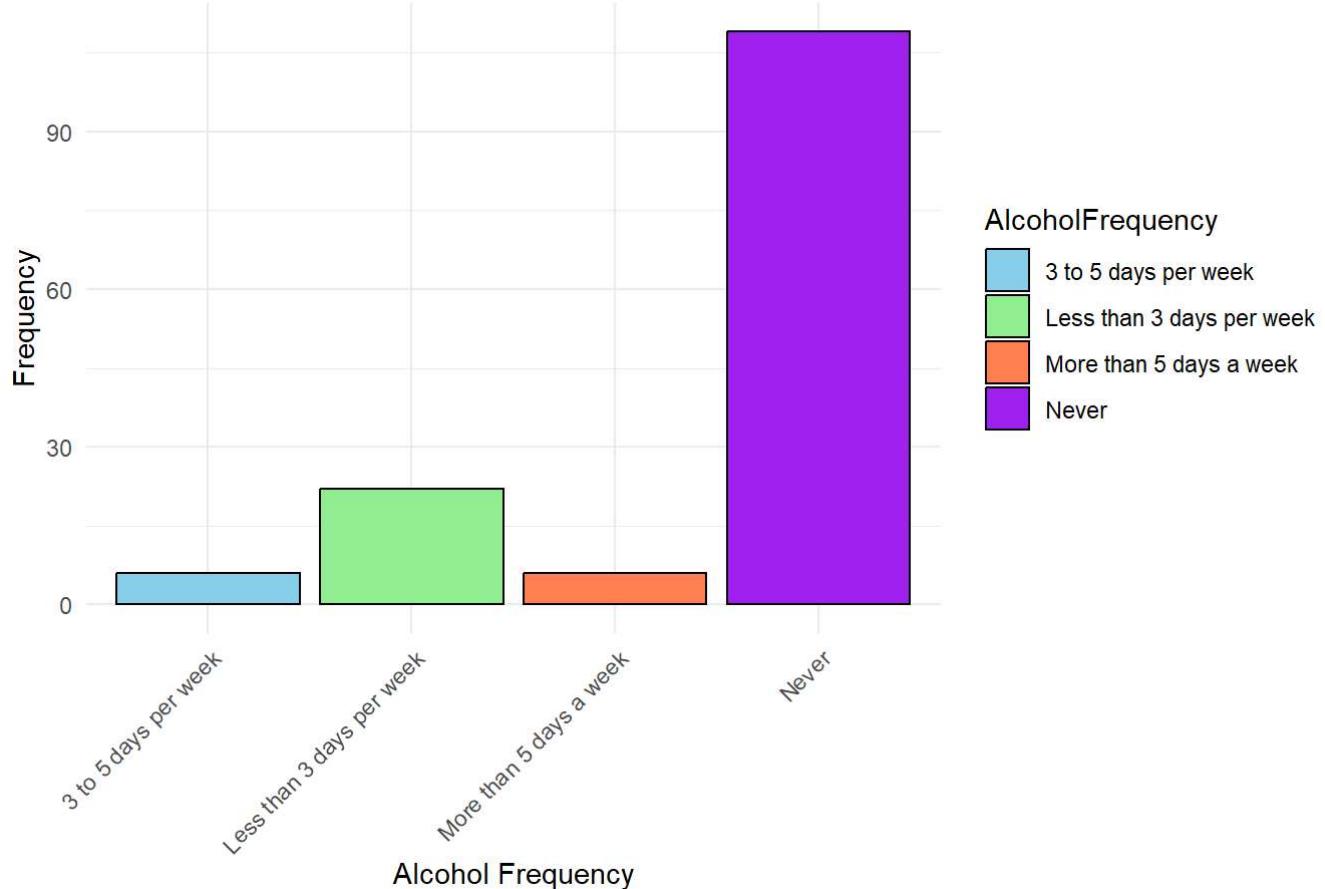


#findings

There are majority of students having atleast 6-8 hours of sleep

```
# Histogram for Distribution of Alcohol Frequency with custom colors
# Histogram for Distribution of Alcohol Frequency with custom colors and rotated Labels
ggplot(data1, aes(x = AlcoholFrequency, fill = AlcoholFrequency)) +
  geom_bar(color = "black") +
  labs(title = "Distribution of Alcohol Frequency", x = "Alcohol Frequency", y = "Frequency") +
  scale_fill_manual(values = custom_colors) + # Use custom colors
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Distribution of Alcohol Frequency



#The bar chart shows the distribution of alcohol frequency among a sample of people. The bar for "Never" is the tallest, followed by "Less than 3 days per week", "3 to 5 days per week", and "More than 5 days a week". This suggests that most people in this sample drink alcohol occasionally or never, and only a small number of people drink alcohol frequently.

#Here are some specific findings from the bar chart:

The most common alcohol frequency is "Never", with 51% of people in the sample reporting this.

#The second most common alcohol frequency is "Less than 3 days per week", with 38% of people in the sample reporting this.

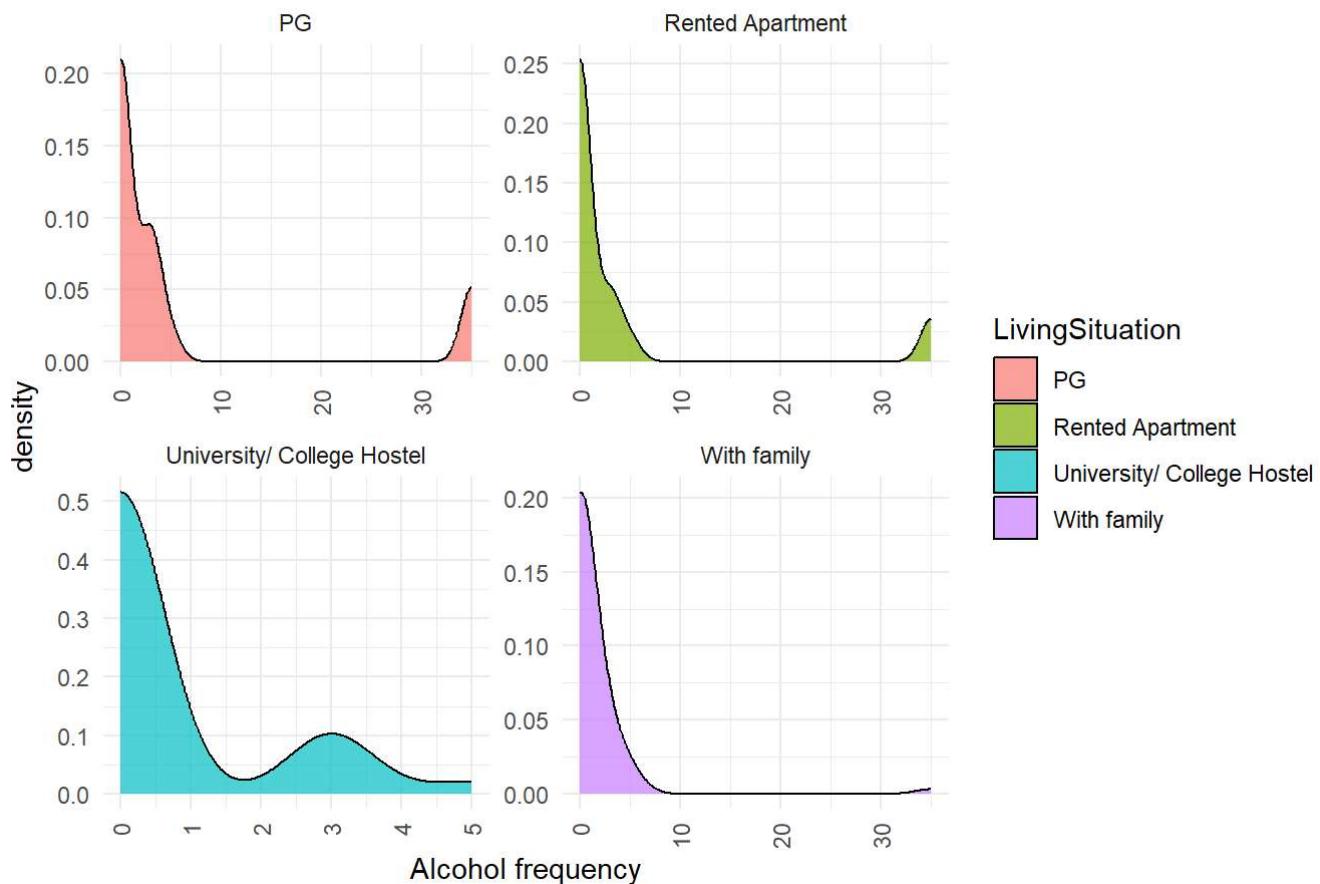
#The third most common alcohol frequency is "3 to 5 days per week", with 11% of people in the sample reporting this.

#The least common alcohol frequency is "More than 5 days per week", with only 0.02% of people in the sample reporting this.

```
data$AlcoholFrequencyNumeric <- as.numeric(gsub("[^0-9-]+", "", data$AlcoholFrequency))
data$AlcoholFrequencyNumeric[data$AlcoholFrequency == "Never"] <- 0 # Treat "Never" as 0
```

```
library(ggplot2)
ggplot(data, aes(x = AlcoholFrequencyNumeric, fill = LivingSituation)) +
  geom_density(alpha = 0.7) +
  labs(title = "Density Plot of AlcoholFrequency by LivingSituation",
       x = "Alcohol frequency") +
  facet_wrap(~ LivingSituation, scales = "free") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```

Density Plot of AlcoholFrequency by LivingSituation



#The graph shows the distribution of alcohol frequency for different Living situations. The blue line represents people who live in a PG, the green line represents people who live in a University/College Hostel, and the red line represents people who live in a Rented Apartment.

#Findings:

People who live in a PG drink more alcohol than people who live in a University/College Hostel or a Rented Apartment.

#There is a bimodal distribution of alcohol frequency among people who live in a PG, with one peak at around 20 and another peak at around 30.

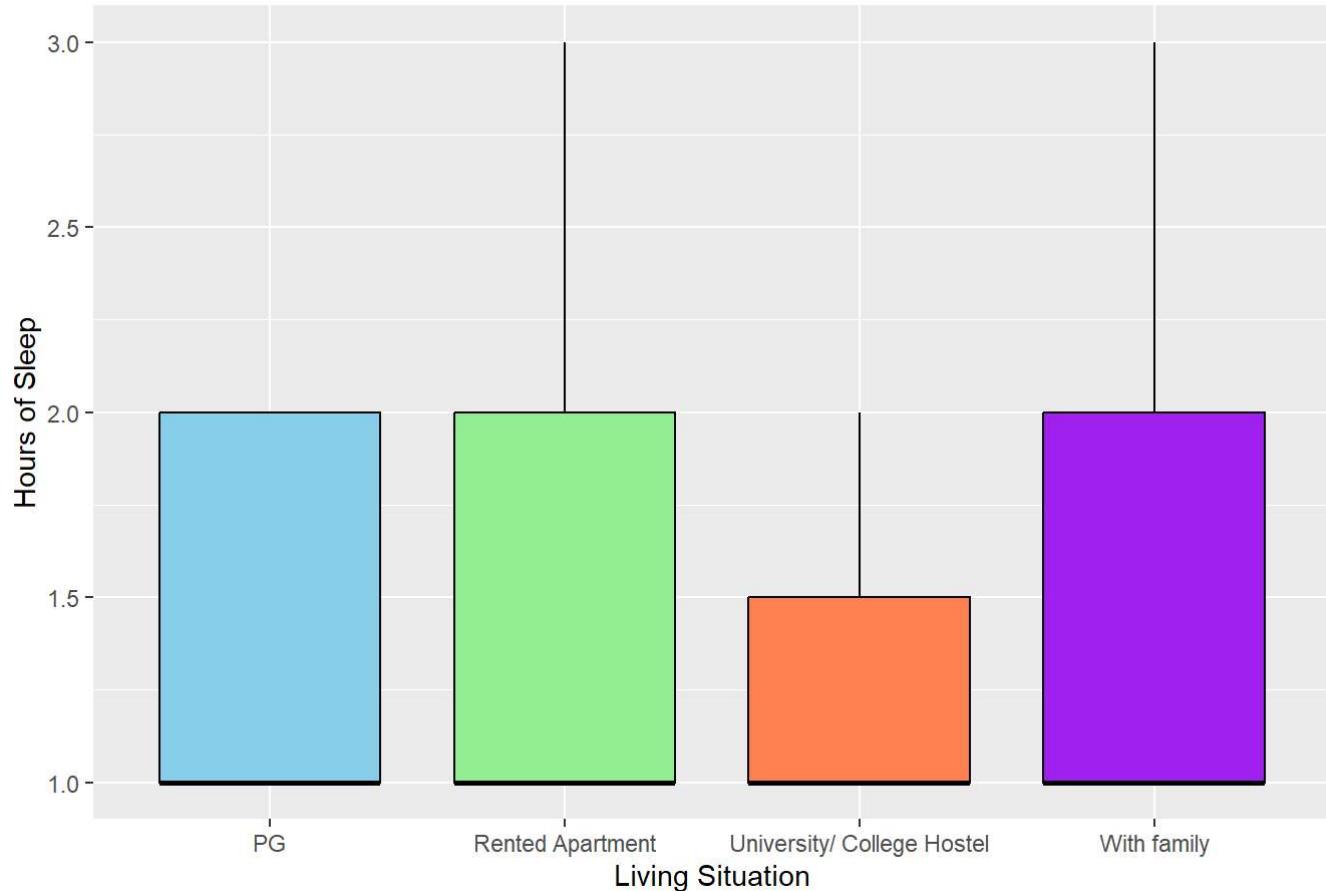
#There is a unimodal distribution of alcohol frequency among people who live in a University/College Hostel, with a peak at around 20.

#There is a unimodal distribution of alcohol frequency among people who live in a Rented Apartment, with a peak at around 10.

Box plot for HoursOfSleep across LivingSituation

```
ggplot(data1, aes(x = LivingSituation, y = as.numeric(HoursOfSleep))) +
  geom_boxplot(fill = custom_colors, color = "black") +
  labs(title = "Box plot of Hours of Sleep by Living Situation", x = "Living Situation", y =
  "Hours of Sleep")
```

Box plot of Hours of Sleep by Living Situation



```
# Subset data for the T-test
t_test_data <- data1[data1$LivingSituation %in% c("Hostel", "Apartment"), ]

# Extract numeric values from the ranges
data1$HoursOfSleepNumeric <- as.numeric(gsub("[^0-9-]+", "", as.character(data1$HoursOfSleep)))
```

```
## Warning: NAs introduced by coercion
```

```

# Take the midpoint for ranges
data1$HoursOfSleepNumeric <- ifelse(grepl("-", data1$HoursOfSleep),
                                      (as.numeric(strsplit(as.character(data1$HoursOfSleep), "-"))[[1]])[1] +
                                          as.numeric(strsplit(as.character(data1$HoursOfSleep), "-"))[[1]][2]) / 2,
                                      data1$HoursOfSleepNumeric)

# hypotheses 1:t-test(2 variables (independent))
#Null Hypothesis ( $H_0$ ): There is no significant difference in the mean hours of sleep
#between students living with family and students
#living in a PG (Paying Guest) accommodation.

#Alternative Hypothesis ( $H_1$ ): There is a significant difference in the mean hours of sleep between students living with family and students living in a PG accommodation.
# Set seed for reproducibility
set.seed(123)

```

```

# Create a random sample of 29 observations
random_sample <- data1[sample(nrow(data1), 29), ]

# Extract numeric values from the ranges
random_sample$HoursOfSleepNumeric <- as.numeric(gsub("[^0-9-]+", "", random_sample$HoursOfSleep))

```

```
## Warning: NAs introduced by coercion
```

```

# Take the midpoint for ranges
random_sample$HoursOfSleepNumeric <- ifelse(grepl("-", random_sample$HoursOfSleep),
                                              (as.numeric(strsplit(as.character(random_sample$HoursOfSleep), "-"))[[1]])[1] +
                                                  as.numeric(strsplit(as.character(random_sample$HoursOfSleep), "-"))[[1]][2]) / 2,
                                              random_sample$HoursOfSleepNumeric)

# Check the structure of the new variable
str(random_sample$HoursOfSleepNumeric)

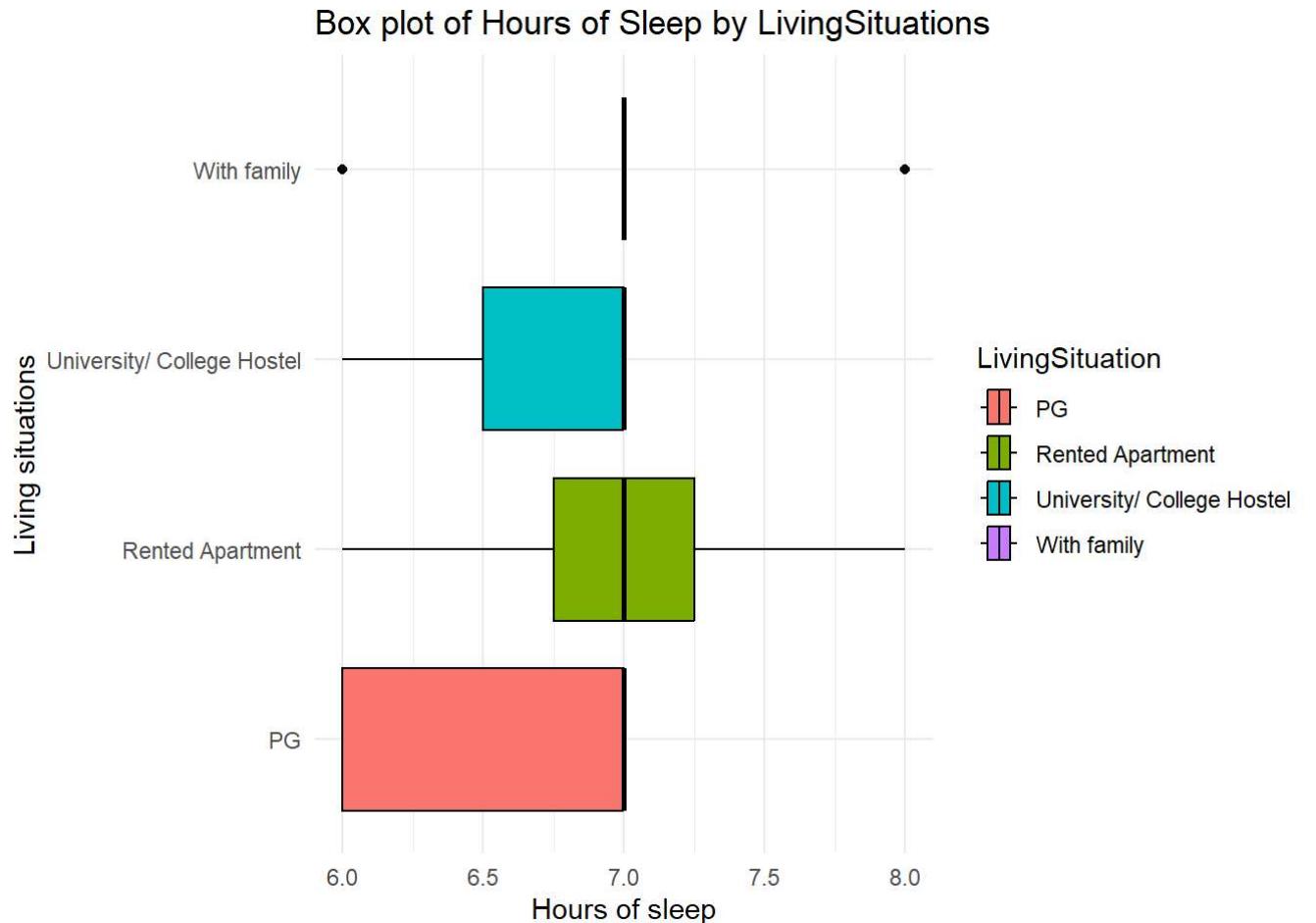
```

```
## num [1:29] 7 7 7 6 6 7 7 8 7 8 ...
```

```
random_sample$HoursOfSleepNumeric
```

```
## [1] 7 7 7 6 6 7 7 8 7 8 7 6 7 7 7 7 6 6 7 8 6 6 7 7 7 7 6
```

```
# Box plot for HoursOfSleep across and living situations
ggplot(random_sample, aes(x =HoursOfSleepNumeric , y = LivingSituation, fill = LivingSituation)) +
  geom_boxplot(color = "black") +
  labs(title = "Box plot of Hours of Sleep by LivingSituations",
       x = "Hours of sleep",
       y = "Living situations") +
  theme_minimal()
```



#The box plot shows the distribution of hours of sleep for different living situations. The blue boxplot represents people who live in a PG, the green boxplot represents people who live in a University/College Hostel, the red boxplot represents people who live in a Rented Apartment, and the purple boxplot represents people who live with family.

#The median number of hours of sleep is highest for people who live with family, followed by people who live in a Rented Apartment, University/College Hostel, and then PG. The interquartile range is also widest for people who live with family, followed by people who live in a Rented Apartment, University/College Hostel, and then PG. This suggests that there is more variability in the number of hours of sleep for people who live with family than for people who live in other living situations.

```
# Subset data for the t-test
t_test_data_random <- random_sample %>% filter(LivingSituation %in% c("With family", "PG"))

# Conduct the t-test
t_test_result_random <- t.test(HoursOfSleepNumeric ~ LivingSituation, data = t_test_data_random)

t_test_result_random
```

```
##  
## Welch Two Sample t-test  
##  
## data: HoursOfSleepNumeric by LivingSituation  
## t = -1.0677, df = 8.5279, p-value = 0.3149  
## alternative hypothesis: true difference in means between group PG and group With family is  
not equal to 0  
## 95 percent confidence interval:  
## -1.0133933 0.3672394  
## sample estimates:  
## mean in group PG mean in group With family  
## 6.600000 6.923077
```

```
summary(t_test_result_random)
```

```
##             Length Class  Mode  
## statistic      1   -none- numeric  
## parameter     1   -none- numeric  
## p.value        1   -none- numeric  
## conf.int       2   -none- numeric  
## estimate       2   -none- numeric  
## null.value     1   -none- numeric  
## stderr         1   -none- numeric  
## alternative    1   -none- character  
## method         1   -none- character  
## data.name      1   -none- character
```

```
#descriptive statistics  
install.packages("e1071")
```

```
## Installing package into 'C:/Users/17ada/AppData/Local/R/win-library/4.3'  
## (as 'lib' is unspecified)
```

```
## package 'e1071' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'e1071'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying  
## C:\Users\17ada\AppData\Local\R\win-library\4.3\00LOCK\e1071\libs\x64\e1071.dll  
## to C:\Users\17ada\AppData\Local\R\win-library\4.3\e1071\libs\x64\e1071.dll:  
## Permission denied
```

```
## Warning: restored 'e1071'
```

```
##  
## The downloaded binary packages are in  
## C:\Users\17ada\AppData\Local\Temp\RtmpMJW7wJ\downloaded_packages
```

```
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.3.2
```

```
# Calculate skewness and kurtosis for Hours of Sleep
skewness_hours <- skewness(t_test_data_random$HoursOfSleepNumeric)
kurtosis_hours <- kurtosis(t_test_data_random$HoursOfSleepNumeric)
skewness_hours
```

```
## [1] 0.07832666
```

```
kurtosis_hours
```

```
## [1] -0.6724195
```

#Skewness: The skewness value for the Hours of Sleep is approximately 0.0783. Since the skewness is close to zero, it indicates that the distribution is approximately symmetric.

#Kurtosis: The kurtosis value for the Hours of Sleep is approximately -0.6724. A negative kurtosis suggests that the distribution has lighter tails than a normal distribution.

Skewness of Hours of Sleep in the T-test analysis: 0.0783 (approximately normal distribution).

#Kurtosis of Hours of Sleep in the T-test analysis: -0.6724 (platykurtic distribution).

#Analysis:

The p-value of 0.3149 is greater than the common significance level of 0.05.

#Fail to reject the null hypothesis (H_0) as there is not enough evidence to suggest a significant difference in means between the PG and With Family groups in terms of hours of sleep.

#The 95% confidence interval includes 0, further supporting the lack of a significant difference.

#Conclusion:

Based on the Welch Two Sample t-test, we do not find sufficient evidence to conclude that there is a significant difference in the mean hours of sleep between individuals living with their family (With Family) and individuals in a PG group.

hypotheses 2: One-way ANOVA

#Null Hypothesis (H_0): The mean alcohol frequency is the same across different living situations (PG, Rented Apartment, University/College Hostel, With Family).

#Alternative Hypothesis (H_1): There is a significant difference in the mean alcohol frequency across different living situations (PG, Rented Apartment, University/College Hostel, With Family).

```
data$LivingSituation <- as.factor(data$LivingSituation)
```

```
str(data$LivingSituation)
```

```
## Factor w/ 4 levels "PG","Rented Apartment",..: 4 1 4 4 4 4 4 1 4 ...
```

```
summary(data$LivingSituation)
```

```
## PG Rented Apartment
## 21 20
## University/ College Hostel With family
## 31 71
```

```
# Convert AlcoholFrequency to a numeric variable (assuming it represents a scale)
data$AlcoholFrequencyNumeric <- as.numeric(gsub("[^0-9-]+", "", data$AlcoholFrequency))
data$AlcoholFrequencyNumeric[data$AlcoholFrequency == "Never"] <- 0 # Treat "Never" as 0
```

```
# Check the updated structure and summary
str(data$AlcoholFrequencyNumeric)
```

```
## num [1:143] 3 3 5 0 0 0 0 0 3 0 ...
```

```
summary(data$AlcoholFrequencyNumeric)
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.00 0.00 0.00 2.14 0.00 35.00
```

```
# Perform One-Way ANOVA
```

```
one_way_anova_result <- aov(AlcoholFrequencyNumeric ~ LivingSituation, data = data)
one_way_anova_result
```

```
## Call:
## aov(formula = AlcoholFrequencyNumeric ~ LivingSituation, data = data)
##
## Terms:
## LivingSituation Residuals
## Sum of Squares 538.461 6504.742
## Deg. of Freedom 3 139
##
## Residual standard error: 6.840812
## Estimated effects may be unbalanced
```

```
# Display ANOVA results
summary(one_way_anova_result)
```

```
## Df Sum Sq Mean Sq F value Pr(>F)
## LivingSituation 3 538 179.5 3.835 0.0112 *
## Residuals 139 6505 46.8
## ---
## Signif. codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#descriptive statistics
# Calculate skewness and kurtosis for Alcohol Frequency
skewness_alcohol <- skewness(data$AlcoholFrequencyNumeric)
kurtosis_alcohol <- kurtosis(data$AlcoholFrequencyNumeric)

# Print the results
cat("Skewness for Alcohol Frequency:", skewness_alcohol, "\n")
```

Skewness for Alcohol Frequency: 4.243584

```
cat("Kurtosis for Alcohol Frequency:", kurtosis_alcohol, "\n")
```

Kurtosis for Alcohol Frequency: 16.89283

#Skewness for Alcohol Frequency: 4.2436

#Interpretation: The skewness value of 4.2436 indicates a highly positively skewed distribution. The majority of respondents may have lower alcohol frequency, with a few having significantly higher values, leading to the skewness towards the right.

#Kurtosis for Alcohol Frequency: 16.8928

#Interpretation: The kurtosis value of 16.8928 suggests a distribution with heavy tails and a higher peak than a normal distribution. This high positive kurtosis indicates a leptokurtic distribution, emphasizing the presence of outliers or extreme values in the data.

#Analysis:

The mean alcohol frequency across different living situations (PG, Rented Apartment, University/College Hostel, With Family) is statistically different, as indicated by the p-value of 0.0112.

#The F-statistic value of 3.835 suggests that there is a significant variation in alcohol frequency means among the living situations.

#The effect of living situation on alcohol frequency is significant, supporting the rejection of the null hypothesis.

The results of the One-Way ANOVA suggest that there is a significant difference in alcohol frequency means among individuals living in different situations. Further post-hoc tests or pairwise comparisons may be conducted to identify specific differences between living situations.

#Conclusion:

#T-Test Hypothesis:

The results of the Welch Two Sample t-test do not provide sufficient evidence to conclude a significant difference in mean hours of sleep between individuals living with family and those in PG accommodation (p-value = 0.3149).

#One-Way ANOVA Hypothesis:

The One-Way ANOVA suggests a significant difference in alcohol frequency means among individuals living in different situations (p-value = 0.0112), supporting the rejection of the null hypothesis. Further post-hoc tests may explore specific differences between living situations.