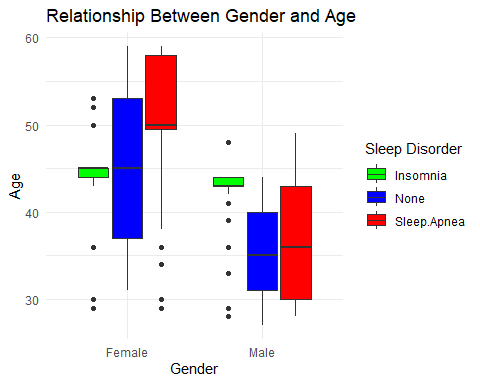
Phân tích

library(tidyverse)   
library(dplyr)  
library(ggplot2)  
library(GGally)  
library(caret)  
library(skimr)  
  
library(scales)  
library(hrbrthemes)  
library(viridis)

## Load dữ liệu   
preprocessed\_data <- read.csv("Sleep\_health\_and\_lifestyle\_dataset.csv")  
  
## Loại bỏ các cột không sử dụng khi training  
data <- preprocessed\_data[, c(2:13)]   
  
## Transform dứ liệu  
data$Blood.Pressure <- sub("/", " ", data$Blood.Pressure)  
data <- separate(data, Blood.Pressure, into = c("Systolic", "Diastolic"), sep = " ")  
data$Systolic <- as.numeric(data$Systolic)  
data$Diastolic <- as.numeric(data$Diastolic)  
  
data$BMI.Category<- gsub("Normal Weight", "Normal", data$BMI.Category)  
  
data$BMI.Category<- gsub("Obese", "Obesity", data$BMI.Category)  
  
data$Sleep.Disorder[data$Sleep.Disorder == "Sleep Apnea"] = "Sleep.Apnea"  
data <- data %>% transform(Sleep.Disorder = as.factor(Sleep.Disorder))  
  
data$Occupation <- as.factor(data$Occupation)  
  
data$Gender <- as.factor(data$Gender)  
  
data$BMI.Category <- as.factor(data$BMI.Category)

### The relationship between gender and age affect insomnia and sleep apnea?

ggplot(data, aes(x = Gender, y = Age, fill = Sleep.Disorder)) +  
 geom\_boxplot() +  
 labs(title = "Relationship Between Gender and Age",  
 x = "Gender",  
 y = "Age",  
 fill = "Sleep Disorder") +  
 scale\_fill\_manual(values = c("None" = "blue", "Sleep.Apnea" = "red", "Insomnia" = "green")) +  
 theme\_minimal()



t\_test\_insomnia <- t.test(Age ~ Gender, data = subset(data, Sleep.Disorder == "Insomnia"))  
t\_test\_sleep\_apnea <- t.test(Age ~ Gender, data = subset(data, Sleep.Disorder == "Sleep.Apnea"))  
# In kết quả kiểm định t-test  
print(t\_test\_insomnia)

##   
## Welch Two Sample t-test  
##   
## data: Age by Gender  
## t = 2.4786, df = 65.3, p-value = 0.01578  
## alternative hypothesis: true difference in means between group Female and group Male is not equal to 0  
## 95 percent confidence interval:  
## 0.5200259 4.8322776  
## sample estimates:  
## mean in group Female mean in group Male   
## 44.94444 42.26829

print(t\_test\_sleep\_apnea)

##   
## Welch Two Sample t-test  
##   
## data: Age by Gender  
## t = 5.7109, df = 13.007, p-value = 7.147e-05  
## alternative hypothesis: true difference in means between group Female and group Male is not equal to 0  
## 95 percent confidence interval:  
## 9.064489 20.094262  
## sample estimates:  
## mean in group Female mean in group Male   
## 51.76119 37.18182

print(t\_test\_sleep\_apnea)

##   
## Welch Two Sample t-test  
##   
## data: Age by Gender  
## t = 5.7109, df = 13.007, p-value = 7.147e-05  
## alternative hypothesis: true difference in means between group Female and group Male is not equal to 0  
## 95 percent confidence interval:  
## 9.064489 20.094262  
## sample estimates:  
## mean in group Female mean in group Male   
## 51.76119 37.18182

### Mối quan hệ giữa nghề nghiệp và thời lượng giấc ngủ ảnh hưởng đến chất lượng giấc ngủ như thế nào?

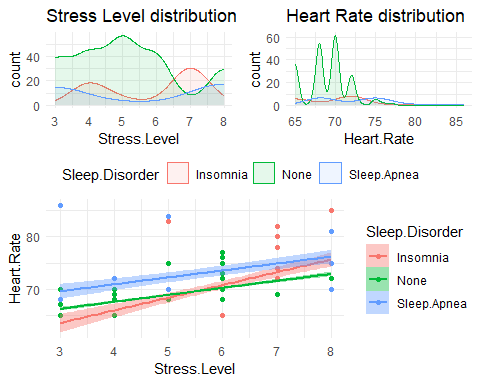
# Tạo một data frame mới với các biến cần thiết  
data\_analysis <- data %>%  
 select(Occupation, Sleep.Duration, Quality.of.Sleep)  
# Thực hiện phân tích ANOVA  
anova\_result <- aov(Quality.of.Sleep ~ Occupation \* Sleep.Duration, data = data\_analysis)  
# Hiển thị kết quả phân tích ANOVA  
summary(anova\_result)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Occupation 10 241.91 24.19 195.50 <2e-16 \*\*\*  
## Sleep.Duration 1 223.35 223.35 1805.01 <2e-16 \*\*\*  
## Occupation:Sleep.Duration 8 25.33 3.17 25.59 <2e-16 \*\*\*  
## Residuals 354 43.80 0.12   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Mối quan hệ Stress levels và Heart Rate

plot1 <- data %>% ggplot(aes(x = Stress.Level, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("Stress Level distribution") +  
 theme\_minimal()  
  
plot2<- data %>% ggplot(aes(x = Heart.Rate, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("Heart Rate distribution") +  
 theme\_minimal()  
  
plot3 <- data %>% ggplot(aes(y = Heart.Rate, x = Stress.Level, color = Sleep.Disorder)) +  
 geom\_point() +  
 geom\_smooth(method=lm , aes(fill= Sleep.Disorder), se=TRUE) +  
 theme\_minimal()  
library(ggpubr)  
ggarrange(ggarrange(plot1, plot2, ncol = 2, nrow = 1, common.legend = TRUE,  
legend = "bottom"), plot3, ncol = 1, nrow = 2 )

## `geom\_smooth()` using formula = 'y ~ x'

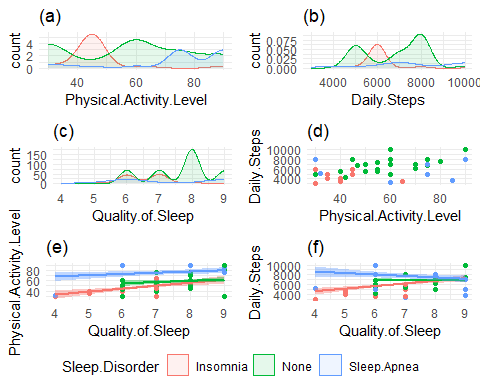


### Mối quan hệ giữa physical activity,daily steps ảnh hưởng đến Quality of Sleep và Sleep Disorder

Physic\_plot <-data %>% ggplot(aes(x = Physical.Activity.Level, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("(a)") +  
 theme\_minimal()  
  
DStep\_plot <- data %>% ggplot(aes(x = Daily.Steps, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("(b)") +  
 theme\_minimal()  
  
QSleep\_plot <- data %>% ggplot(aes(x = Quality.of.Sleep, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("(c)") +  
 theme\_minimal()  
  
  
Physic\_DStep\_plot <- data %>% ggplot(aes(x = Physical.Activity.Level, y = Daily.Steps, color = Sleep.Disorder)) +  
 geom\_point() +  
 ggtitle("(d)") +  
 theme\_minimal()  
  
Physic\_QSleep\_plot <-data %>% ggplot(aes(x = Quality.of.Sleep, y = Physical.Activity.Level, color = Sleep.Disorder)) +  
 geom\_point() +  
 geom\_smooth(method=lm , aes(fill= Sleep.Disorder), se=TRUE) +  
 ggtitle("(e)") +  
 theme\_minimal()  
  
DStep\_QSleep\_plot <- data %>% ggplot(aes(x = Quality.of.Sleep, y = Daily.Steps, color = Sleep.Disorder)) +  
 geom\_point() +  
 ggtitle("(f)")+  
 geom\_smooth(method=lm , aes(fill= Sleep.Disorder), se=TRUE) +  
 theme\_minimal()

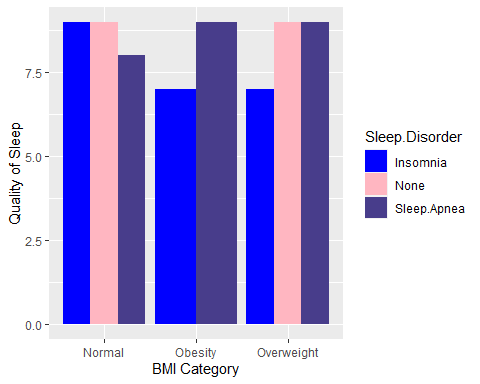
library(ggpubr)  
ggarrange(  
 Physic\_plot, DStep\_plot, QSleep\_plot, Physic\_DStep\_plot, Physic\_QSleep\_plot, DStep\_QSleep\_plot,  
 ncol = 2, nrow = 3,  
 common.legend = TRUE, legend = "bottom"  
)

## `geom\_smooth()` using formula = 'y ~ x'  
## `geom\_smooth()` using formula = 'y ~ x'



### Mối quan hệ giữa BMI categories và sleep quality

ggplot(data, aes(x = BMI.Category, y = Quality.of.Sleep, fill=Sleep.Disorder)) +  
 geom\_bar(stat = "identity", position = "dodge") +  
 scale\_fill\_manual(values = c("None" = "lightpink", "Sleep.Apnea" = "darkslateblue","Insomnia"="blue")) +  
 labs(x = "BMI Category", y = "Quality of Sleep")



### Mối quan hệ blood pressure và heart rate ảnh hưởng đến insomnia và sleep apnea

ggplot(data, aes(x =data$Heart.Rate, y =data$Systolic , color= data$Sleep.Disorder)) +   
 geom\_point() +  
 labs(x = "Heart Rate", y = "Systolic") +   
 geom\_smooth(method = 'loess',formula = 'y ~ x',mapping = aes(x =data$Heart.Rate, y =data$Systolic , color= data$Sleep.Disorder))

## Warning: Use of `data$Heart.Rate` is discouraged.  
## ℹ Use `Heart.Rate` instead.

## Warning: Use of `data$Systolic` is discouraged.  
## ℹ Use `Systolic` instead.

## Warning: Use of `data$Sleep.Disorder` is discouraged.  
## ℹ Use `Sleep.Disorder` instead.

## Warning: Use of `data$Heart.Rate` is discouraged.  
## ℹ Use `Heart.Rate` instead.

## Warning: Use of `data$Systolic` is discouraged.  
## ℹ Use `Systolic` instead.

## Warning: Use of `data$Sleep.Disorder` is discouraged.  
## ℹ Use `Sleep.Disorder` instead.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : pseudoinverse used at 70

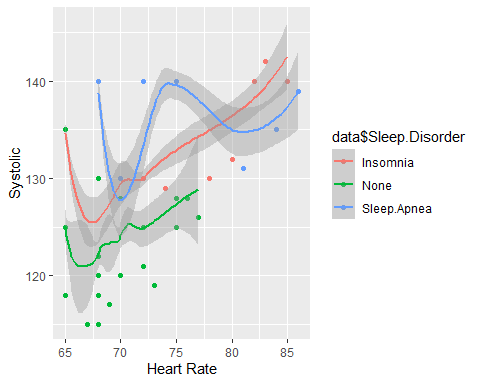
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : neighborhood radius 2

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : reciprocal condition number 1.5705e-16

## Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x  
## else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at  
## 70

## Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x  
## else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius 2

## Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x  
## else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition  
## number 1.5705e-16



ggplot(data, aes(x =data$Heart.Rate, y =data$Diastolic , color= data$Sleep.Disorder)) +   
 geom\_point() +   
 labs(x = "Heart Rate", y = "Diastolic") +  
 geom\_smooth(method = 'loess',formula = 'y ~ x',mapping = aes(x =data$Heart.Rate, y =data$Diastolic , color= data$Sleep.Disorder))

## Warning: Use of `data$Heart.Rate` is discouraged.  
## ℹ Use `Heart.Rate` instead.

## Warning: Use of `data$Diastolic` is discouraged.  
## ℹ Use `Diastolic` instead.

## Warning: Use of `data$Sleep.Disorder` is discouraged.  
## ℹ Use `Sleep.Disorder` instead.

## Warning: Use of `data$Heart.Rate` is discouraged.  
## ℹ Use `Heart.Rate` instead.

## Warning: Use of `data$Diastolic` is discouraged.  
## ℹ Use `Diastolic` instead.

## Warning: Use of `data$Sleep.Disorder` is discouraged.  
## ℹ Use `Sleep.Disorder` instead.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : pseudoinverse used at 70

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : neighborhood radius 2

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : reciprocal condition number 1.5705e-16

## Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x  
## else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at  
## 70

## Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x  
## else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius 2

## Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x  
## else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition  
## number 1.5705e-16

