Plot

## Đọc dữ liệu

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

data <- read.csv("Sleep\_health\_and\_lifestyle\_dataset.csv")  
  
data <- data[, c(2:13)]

## Tiền xử lý dữ liệu

* Blood Pressure

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)

* BMI Caterogy

data$BMI.Category<- gsub("Normal Weight", "Normal", data$BMI.Category)   
data$BMI.Category<- gsub("Obese", "Obesity", data$BMI.Category)

* Outcome

data$Sleep.Disorder[data$Sleep.Disorder == "Sleep Apnea"] = "Sleep.Apnea"   
data <- data %>% transform(Sleep.Disorder = as.factor(Sleep.Disorder))

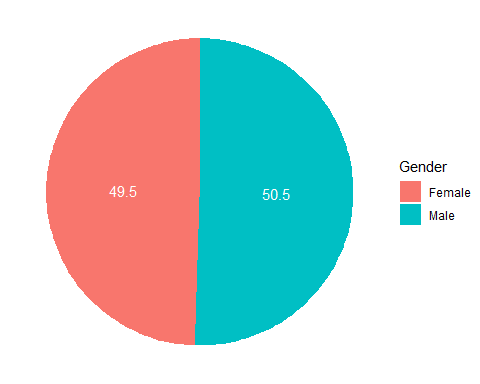
* factor

data$Gender <- as.factor(data$Gender)   
data$Occupation <- as.factor(data$Occupation)   
data$BMI.Category <- as.factor(data$BMI.Category)

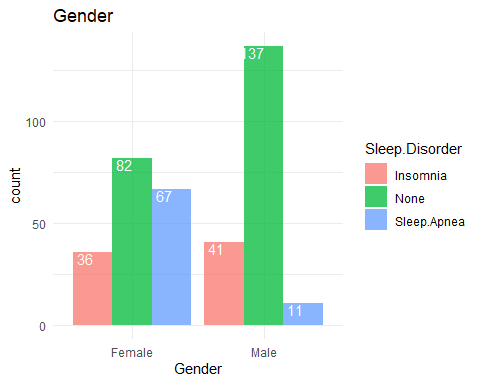
## Trực quan hóa dữ liệu

### Gender

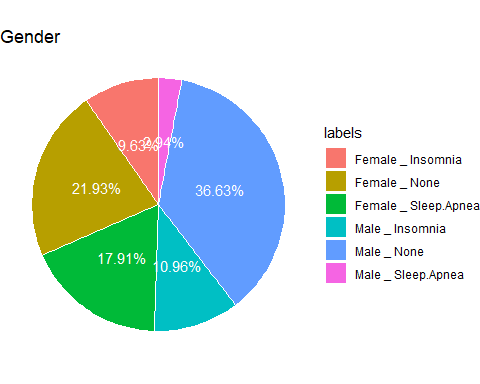
data %>% count(Gender) %>% mutate(labels = round(100 \* n/ sum(n), 1)) %>%  
 ggplot(aes(x = "", y = labels, fill = Gender, group = Gender))+  
 geom\_bar(stat = "identity")+  
 coord\_polar(theta = "y", start=0) +  
 geom\_text(aes(label = labels), position = position\_stack(vjust = 0.5), color = "white") +  
 theme\_void()



data %>% count(Gender, Sleep.Disorder) %>% group\_by(Gender) %>%  
 ggplot(aes(x = Gender, y = n, fill = Sleep.Disorder)) +  
 geom\_bar(stat = "identity", position = 'dodge', alpha = 0.75) +  
 ggtitle("Gender") +  
 geom\_text(aes(label = n), position = position\_dodge(0.9), color = "white", vjust = 1, hjust = 1) +  
 labs(y = 'count') +  
 theme\_minimal()

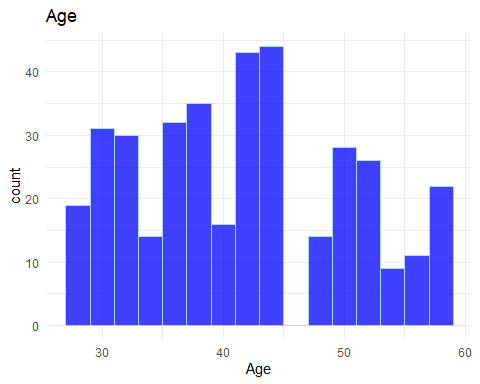


data %>% select(Gender, Sleep.Disorder) %>%   
 count(Gender, Sleep.Disorder) %>% arrange(Gender) %>%  
 mutate(values = round(100\*n/sum(n), 2)) %>% mutate(labels = paste(Gender, '\_', Sleep.Disorder)) %>%  
 ggplot(aes(x = "", y = values, fill = labels, group = labels)) +  
 geom\_col(width=1, color="white") +  
 coord\_polar(theta = "y", start=0) +  
 ggtitle("Gender") +  
 theme\_void() +  
 geom\_text(aes(label = paste0(values, "%")), position = position\_stack(vjust = 0.5), color = "white") +  
 theme(  
 panel.spacing = unit(0.1, "lines"),  
 axis.ticks.x=element\_blank()  
 )

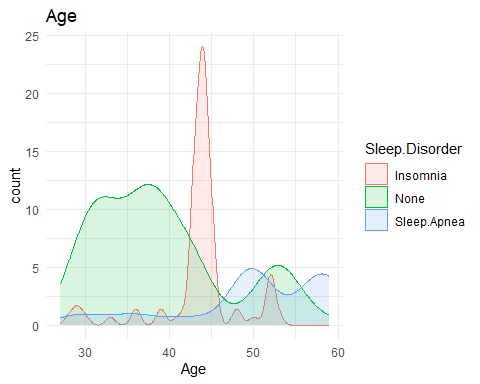


### Age

data %>% ggplot(aes(x = Age)) +  
 geom\_histogram(binwidth = 2, color = "lightblue", fill = "blue", alpha = 0.75) +  
 # geom\_density(aes(y= 2 \* after\_stat(count)),color = "blue", alpha = 0.15) +  
 ggtitle("Age") +  
 ylab("count")+  
 theme\_minimal()

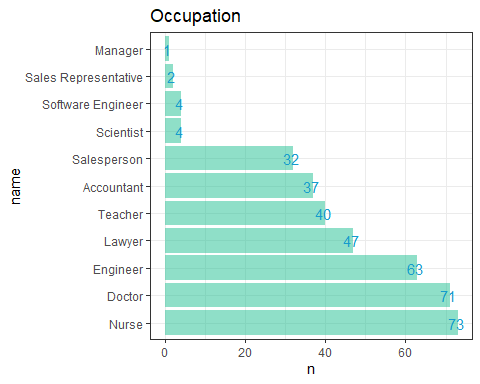


data %>% ggplot(aes(x = Age, fill = Sleep.Disorder, color = Sleep.Disorder,)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.15) +  
 ggtitle("Age") +  
 theme\_minimal()



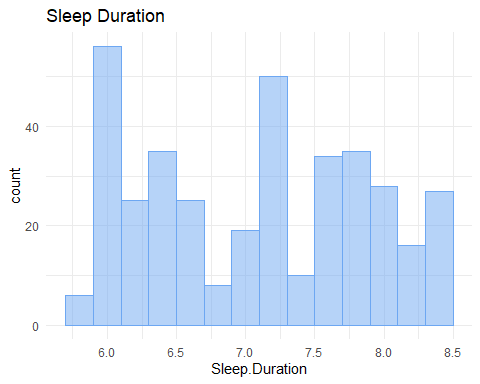
### **Occupation**

data %>% count(Occupation) %>% arrange(n) %>% mutate(name = fct\_reorder(Occupation, desc(n))) %>%  
 ggplot(aes(x = name, y = n)) +  
 geom\_bar(stat = "identity",fill="#1FBF92", alpha=0.5) +  
 coord\_flip() +  
 ggtitle("Occupation") +  
 geom\_text(aes(label = n), nudge\_y = -0.5, color = "#169CCC") +  
 theme\_bw()

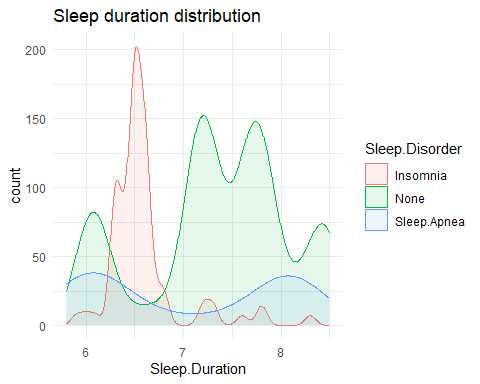


### **Sleep Duration**

ggplot(data = data, aes(x = Sleep.Duration)) +  
 geom\_histogram(color = "#6DA7F2", fill = "#6DA7F2", alpha = 0.5, binwidth = 0.2) +  
 ggtitle("Sleep Duration") +  
 theme\_minimal()

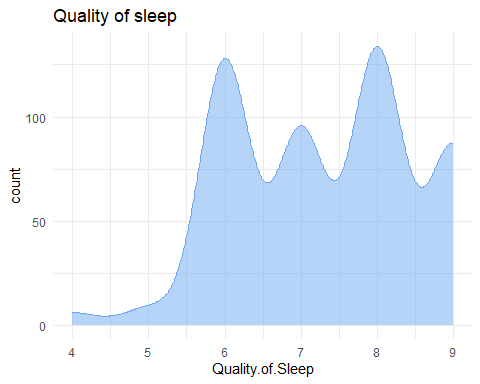


data %>% ggplot(aes(x = Sleep.Duration, fill = Sleep.Disorder, color = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("Sleep duration distribution") +  
 theme\_minimal()

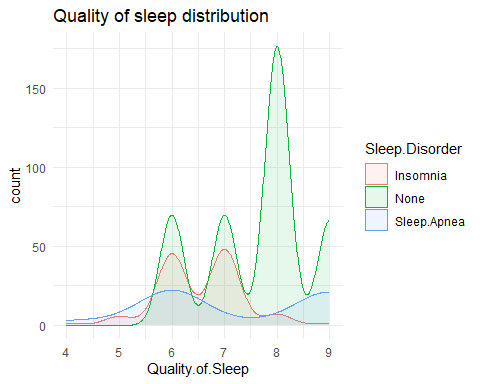


### **Quality of sleep**

data %>% ggplot(aes(x = Quality.of.Sleep)) +  
 geom\_density(aes(y= after\_stat(count)), color = "#6DA7F2", fill = "#6DA7F2", alpha = 0.5) +  
 ggtitle("Quality of sleep") +  
 theme\_minimal()

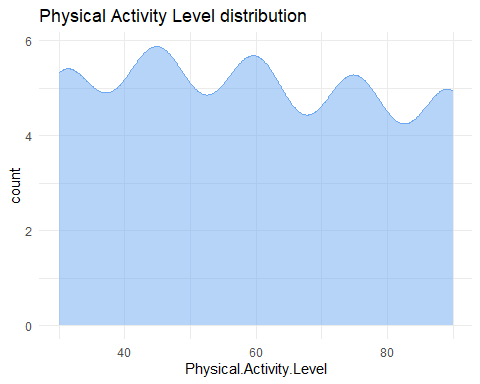


data %>% ggplot(aes(x = Quality.of.Sleep, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("Quality of sleep distribution") +  
 theme\_minimal()

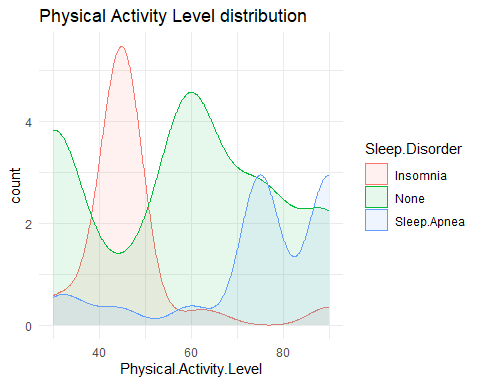


### Physical Activity level Distribution

data %>% ggplot(aes(x = Physical.Activity.Level)) +  
 geom\_density(aes(y = after\_stat(count)),color = "#6DA7F2", fill = "#6DA7F2", alpha = 0.5, binwidth = 5) +  
 ggtitle("Physical Activity Level distribution") +  
 theme\_minimal()

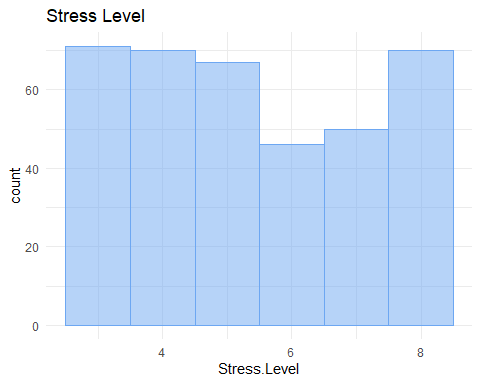


data %>% ggplot(aes(x = Physical.Activity.Level, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("Physical Activity Level distribution") +  
 theme\_minimal()

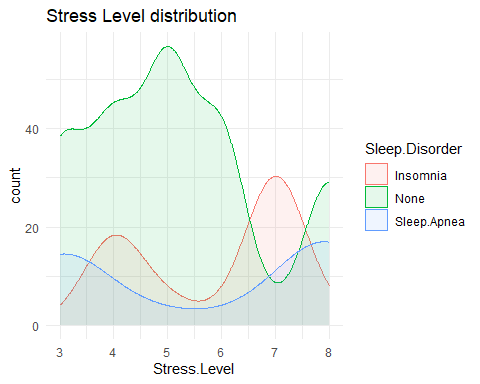


### Stress level

data %>% ggplot(aes(x = Stress.Level)) +  
 geom\_histogram(fill = "#6DA7F2", color="#6DA7F2", alpha = 0.5, binwidth = 1) +  
 ggtitle("Stress Level") +  
 theme\_minimal()

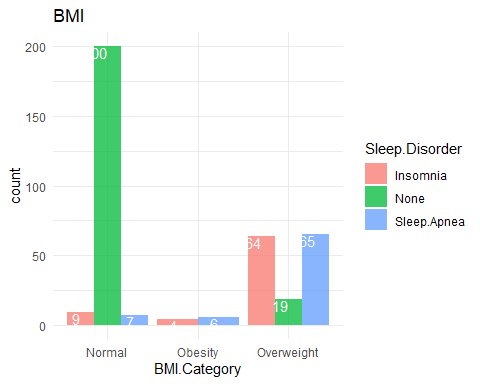


data %>% ggplot(aes(x = Stress.Level, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("Stress Level distribution") +  
 theme\_minimal()



### BMI

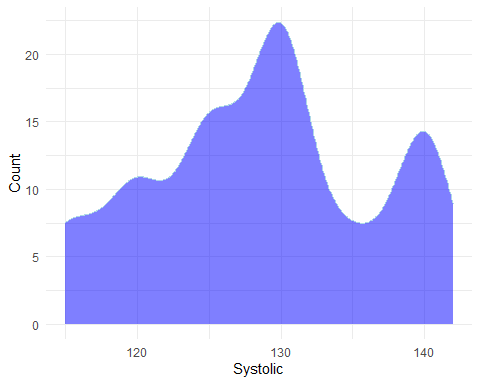
data %>% count(BMI.Category, Sleep.Disorder) %>% group\_by(BMI.Category) %>%  
 ggplot(aes(x = BMI.Category, y = n, fill = Sleep.Disorder)) +  
 geom\_bar(stat = "identity", position = 'dodge', alpha = 0.75) +  
 ggtitle("BMI") +  
 geom\_text(aes(label = n), position = position\_dodge(0.9), color = "white", vjust = 1, hjust = 1) +  
 labs(y = 'count') +  
 theme\_minimal()



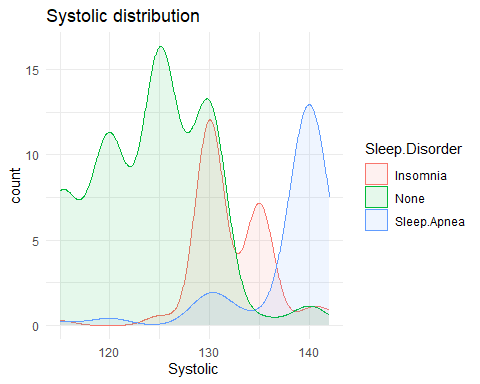
### Systolic

ggplot(data, aes(x=data$Systolic)) +   
 geom\_density(aes(y = after\_stat(count)),color="lightblue", fill = "blue", alpha = 0.5)+  
 labs(x = "Systolic", y = "Count") +  
 theme\_minimal()

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



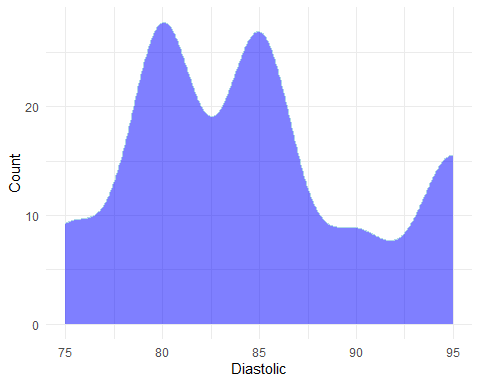
data %>% ggplot(aes(x = Systolic, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("Systolic distribution") +  
 theme\_minimal()



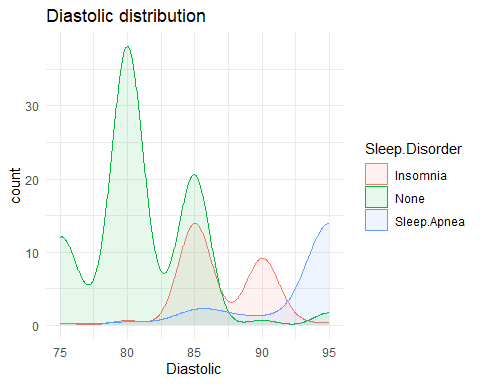
### Diatolic

ggplot(data, aes(x=data$Diastolic)) +   
 geom\_density(aes(y = after\_stat(count)),color="lightblue", fill = "blue", alpha = 0.5)+  
 labs(x = "Diastolic", y = "Count") +  
 theme\_minimal()

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

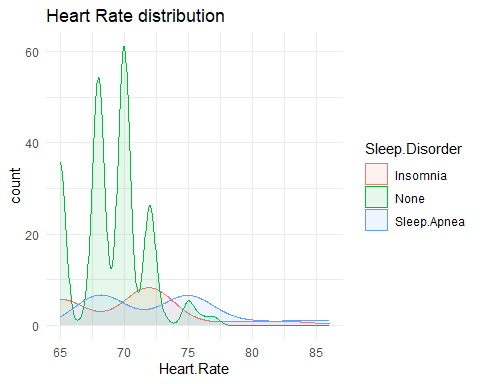


data %>% ggplot(aes(x = Diastolic, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("Diastolic distribution") +  
 theme\_minimal()



### Heart rate

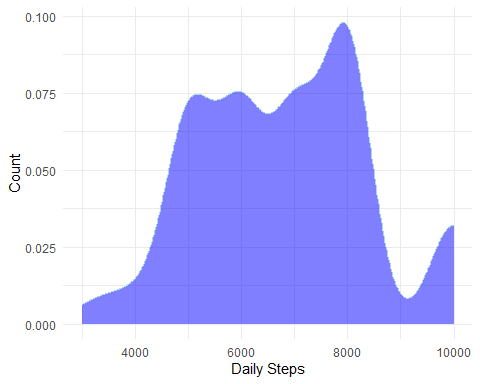
data %>% ggplot(aes(x = Heart.Rate, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("Heart Rate distribution") +  
 theme\_minimal()



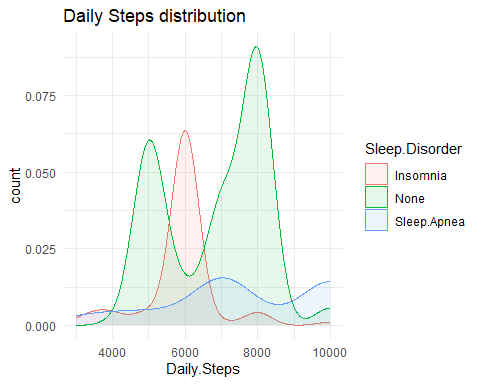
### Daily Steps

ggplot(data, aes(x=data$Daily.Steps)) +   
 geom\_density(aes(y = after\_stat(count)),color="lightblue", fill = "blue", alpha = 0.5)+  
 labs(x = "Daily Steps", y = "Count") +  
 theme\_minimal()

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



data %>% ggplot(aes(x = Daily.Steps, color = Sleep.Disorder, fill = Sleep.Disorder)) +  
 geom\_density(aes(y= after\_stat(count)), alpha = 0.1) +  
 ggtitle("Daily Steps distribution") +  
 theme\_minimal()



### Sleep Disorder

data %>% select(Sleep.Disorder) %>%   
 count(Sleep.Disorder) %>% arrange(n) %>%   
 mutate(labels = scales::percent(n/sum(n))) %>%  
 ggplot(aes(x = "", y = labels, fill = Sleep.Disorder)) +  
 geom\_bar(stat="identity", width=1, color="white") +  
 coord\_polar("y", start=0) +  
 ggtitle("Sleep Disorder") +  
 theme\_void() +  
 geom\_text(aes(label = labels), position = position\_stack(vjust = 0.5), color = "white")

