732A98-Visualization Lab 1

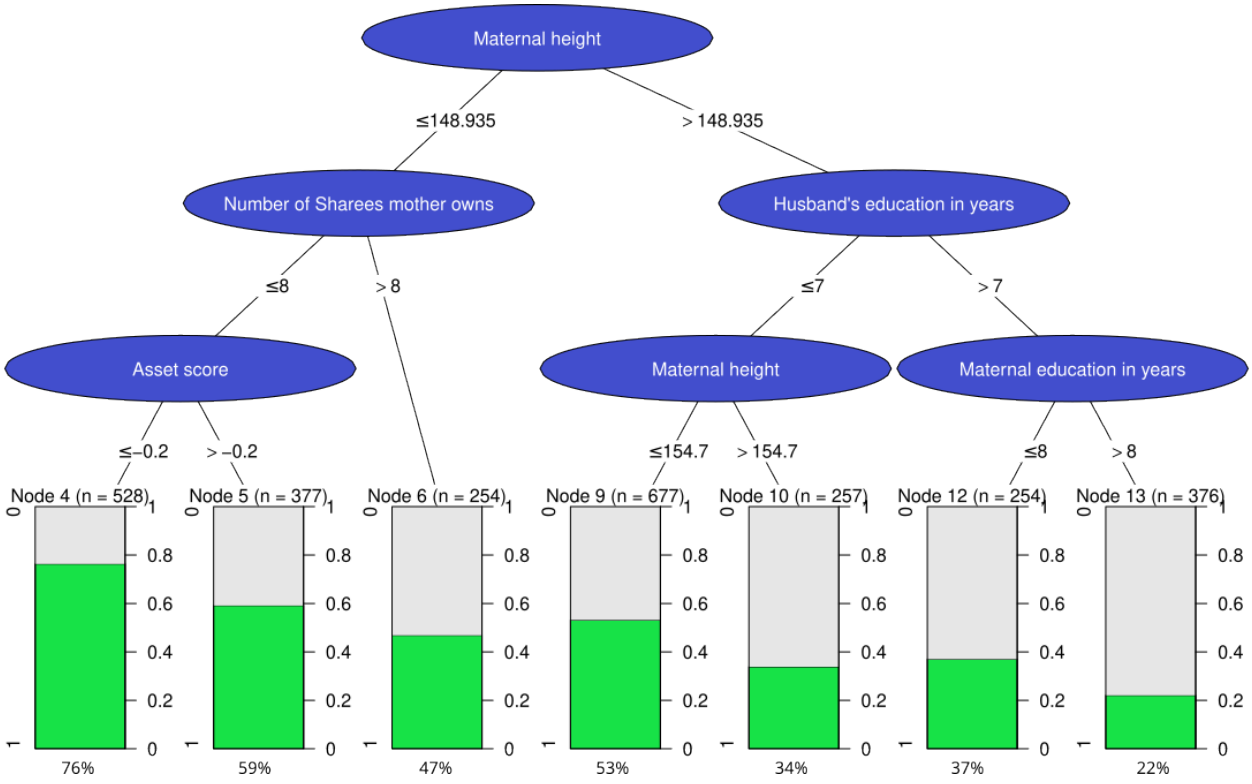
Qinyuan Qi(qinqi464)

2024-08-30

Assignment 1

The modified graph is as follows. Plot generated by tree package saved to an SVG file, then edited by Inkscape. To be included in Rmd, it is converted to a png file. The result is as follows.

```
knitr::include_graphics("tree.png")
```



Assignment 2

2.1 Read file

```
############################  Code For Assignment 2.1 ############################
# read data
# since there have different space count between columns, we use  read.table and use fill=TRUE
data <- read.table("SENIC.txt", header = FALSE, fill = TRUE)

# Assign column names
```

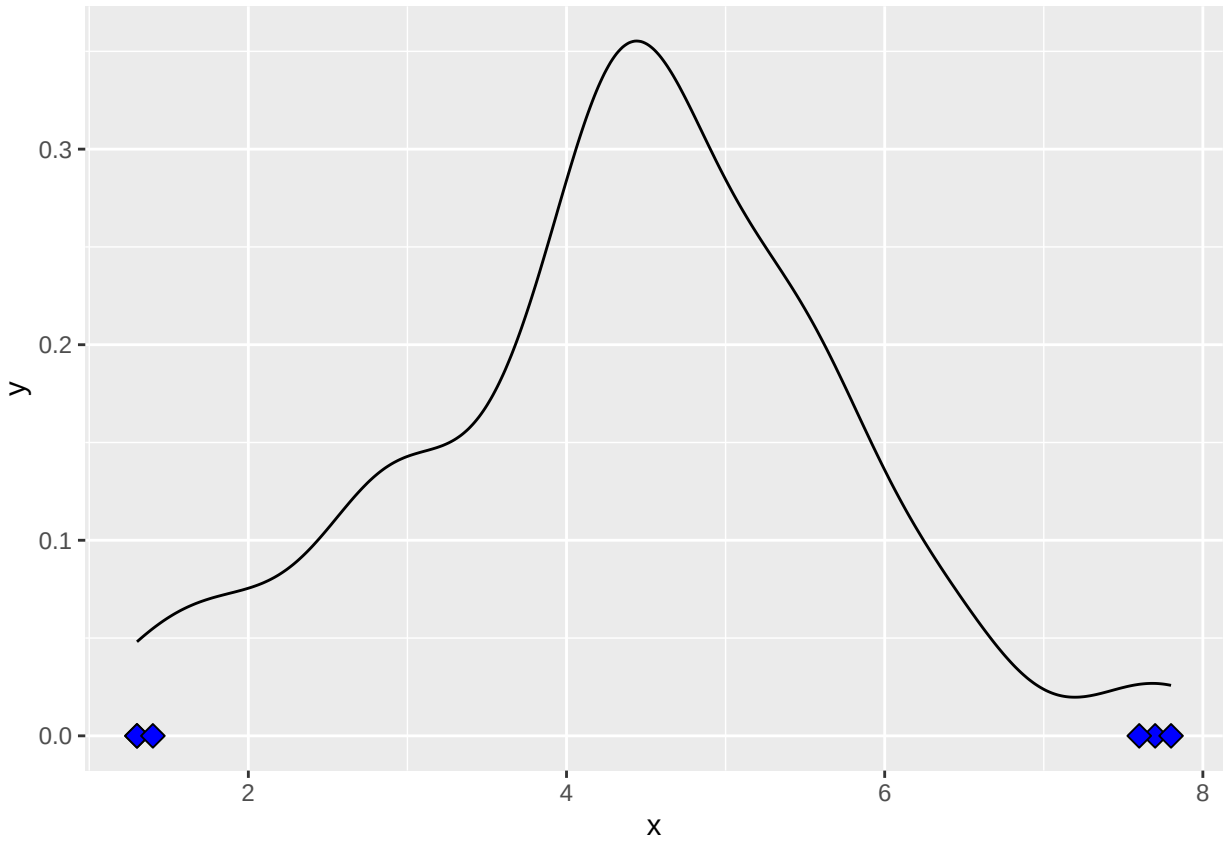
```r
colnames(data) <- c("ID", "X1", "X2", "X3", "X4", "X5", "X6", "X7", "X8", "X9", "X10", "X11")

# Convert columns to appropriate types
data$ID <- as.integer(data$ID)
data$X1 <- as.numeric(data$X1)
data$X2 <- as.numeric(data$X2)
data$X3 <- as.numeric(data$X3)
data$X4 <- as.numeric(data$X4)
data$X5 <- as.numeric(data$X5)
data$X6 <- as.numeric(data$X6)
data$X7 <- as.integer(data$X7)
data$X8 <- as.integer(data$X8)
data$X9 <- as.integer(data$X9)
data$X10 <- as.integer(data$X10)
data$X11 <- as.numeric(data$X11)
```

2.2 Calculate outlier observations

```r
############################  Code For Assignment 2.2 ############################
calc_outlier_observations <- function(data_column) {

  # Compute the first and third quartiles
  Q1 <- quantile(data_column, 0.25, na.rm = TRUE)
  Q3 <- quantile(data_column, 0.75, na.rm = TRUE)

  outlie_indices <- which(data_column > Q3+1.5 * (Q3-Q1) | data_column < Q1-1.5 * (Q3-Q1))
  return(outlie_indices)
}
```

2.3 Create a density plot of Infection risk in which outliers are plotted as a diamond symbol.

```r
############################  Code For Assignment 2.3 ############################
library(ggplot2)
# get the Infection Risk outliers
x3_outliers <- calc_outlier_observations(data$X3)

df <- data.frame(x = data$X3)
df_outliers  <- data.frame(x = data$X3[x3_outliers])

graph2.3 <- ggplot(data = df, aes(x=x)) +
  geom_density(alpha = 0.5) +
  geom_point(data=df_outliers,aes(x=x,y=0),shape = 23, size = 3, fill = "blue")

graph2.3
```




```
## [1] "Outliers are:"
```

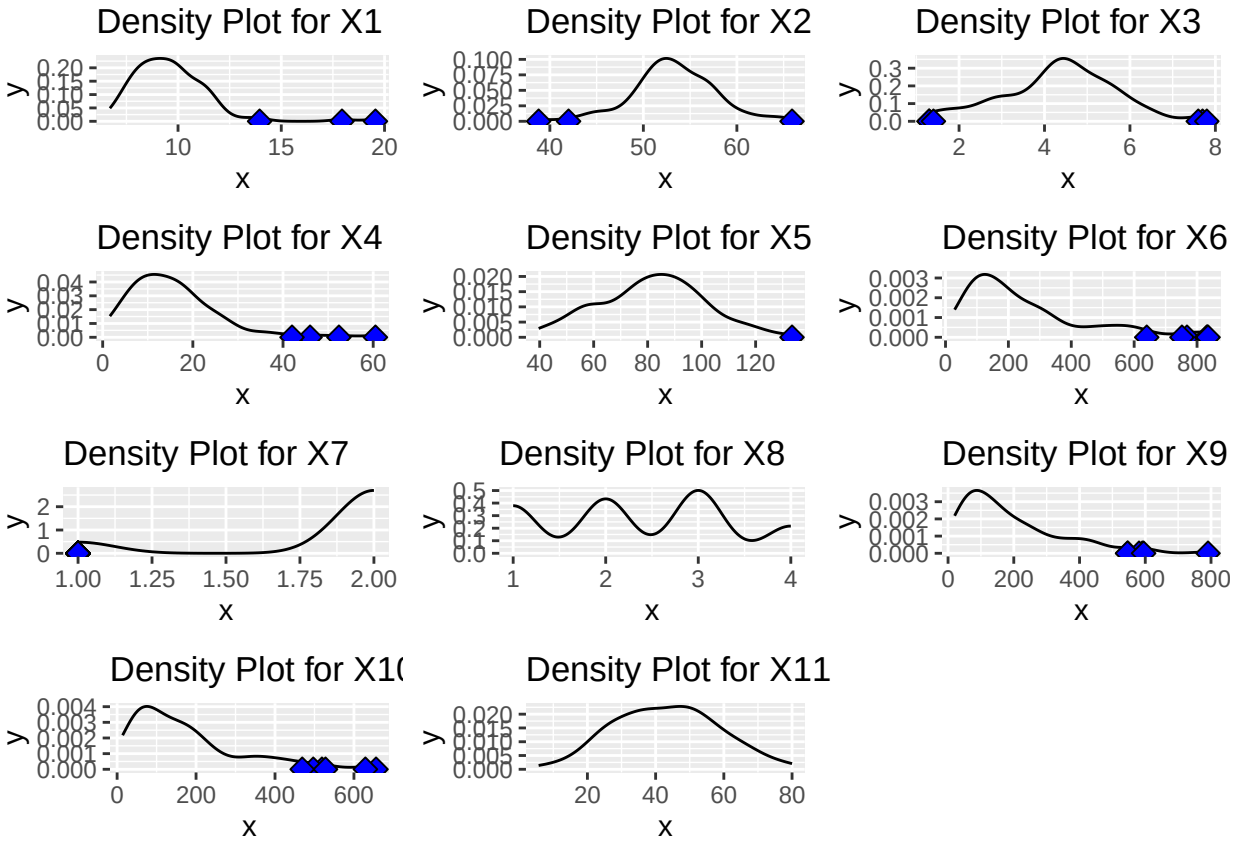
```
## [1] 7.7 1.3 7.6 7.8 1.3 1.4
```

According to the graph and output above, we can see that the Infection risk follows some bell-shaped distribution, approximately a Gaussian distribution. Outlier data are located in the 2 tails of this distribution.

2.4 Produce graphs of the same kind as in step 3 but for all other quantitative variables in the data

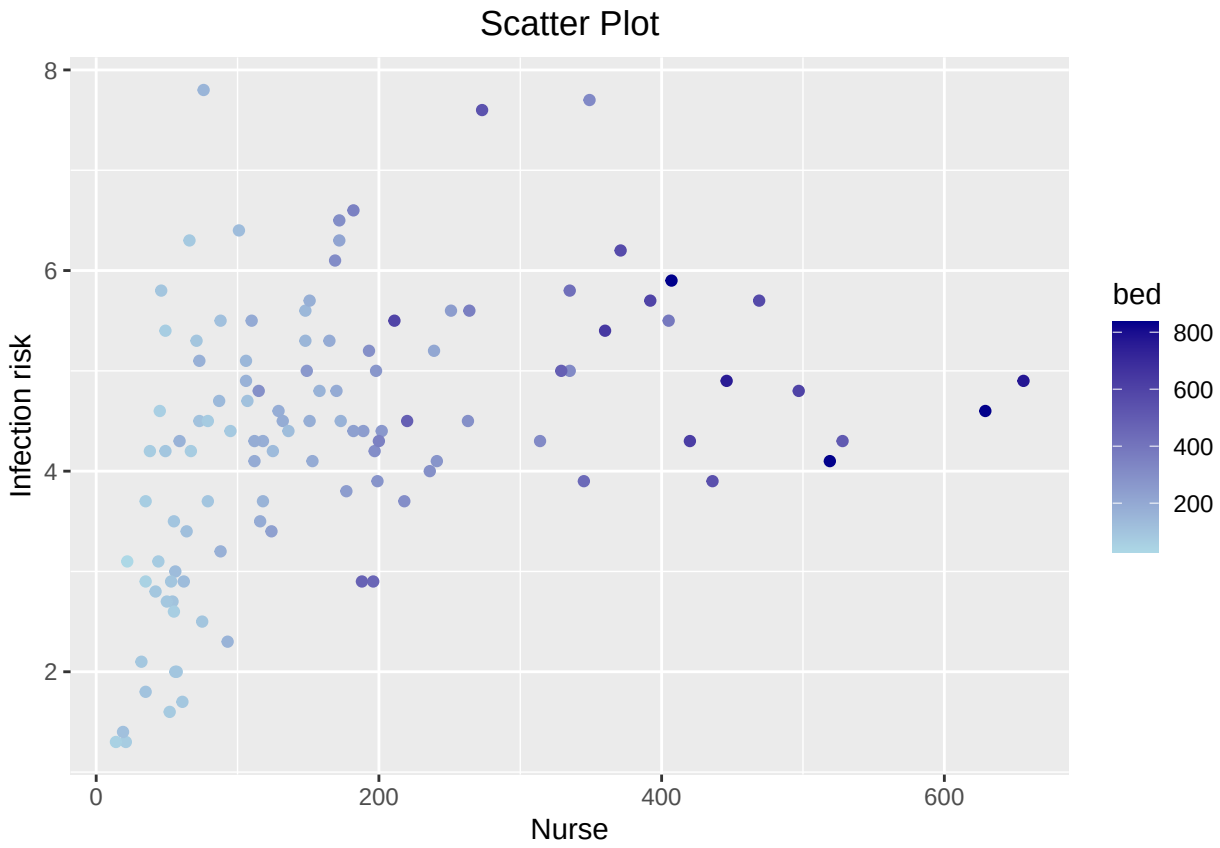
```
############################  Code For Assignment 2.4 ############################
library(gridExtra)

variables <- paste0("X", 1:11)

create_plot <- function(var_name) {
  # Calculate outliers
  outliers <- calc_outlier_observations(data[[var_name]])

  # Create data frames
  df <- data.frame(x = data[[var_name]])
  df_outliers <- data.frame(x = data[[var_name]][outliers])

  # Create the plot
  plot <- ggplot(data = df, aes(x = x)) +
    geom_density(alpha = 0.5) +
    geom_point(data = df_outliers, aes(x = x, y = 0), shape = 23, size = 3, fill = "blue") +
    ggtitle(paste("Density Plot for", var_name))
```

```
  return(plot)
}

# Create a list of plots
plot_list <- lapply(variables, create_plot)

grid_layout <- arrangeGrob(grobs = plot_list, ncol = 3)

# Display the combined plot
grid.arrange(grid_layout)
```



2.5 Create a ggplot2 scatter plot showing the dependence of Infection risk on the Number of Nurses

```
############################  Code For Assignment 2.5 ############################

df <- data.frame(x = data$X10,y = data$X3, bed=data$X6)
graph2.5 <- ggplot(data = df) +
  geom_point(aes(x=x,y=y,color=bed)) +
  scale_color_gradient(low = "lightblue", high = "darkblue") +
  labs(title = 'Scatter Plot', x = 'Nurse', y = 'Infection risk') +
  theme(
    plot.title = element_text(hjust = 0.5)  # Center the title
  )
graph2.5
```



2.6 Convert graph from step 3 to Plotly with ggplotly function

```
############################  Code For Assignment 2.6 ############################
library(plotly)
```

```
##
## Attaching package: 'plotly'
## The following object is masked from 'package:ggplot2':
##
##     last_plot
## The following object is masked from 'package:stats':
##
##     filter
## The following object is masked from 'package:graphics':
##
##     layout
```

```
#ggplotly(graph2.3)
```

2.7 Use pipeline operator to make a histogram of Infection risk, use diamond symbol when plot outliers

```
############################  Code For Assignment 2.7 ############################
```

2.8 Shiny App to do same thing as step 2.4

```
##########################  Code For Assignment 2.8 ############################
```