STA413 Data Analysis Assignment
Group2(TEAM ALPHA) Report

Team members and contribution

Amoo Barakat	Simulated Data In R
Edet Mercy	Computed Descritives statistics In R
Dada Israel	Drew Insights from Data
Olusayo Moses	Visualized the Data
Oluyomi Stephen	Prepared the Report

This report contains the analysis and insight derived from simulated data (patients-data) for a fictional hospital.

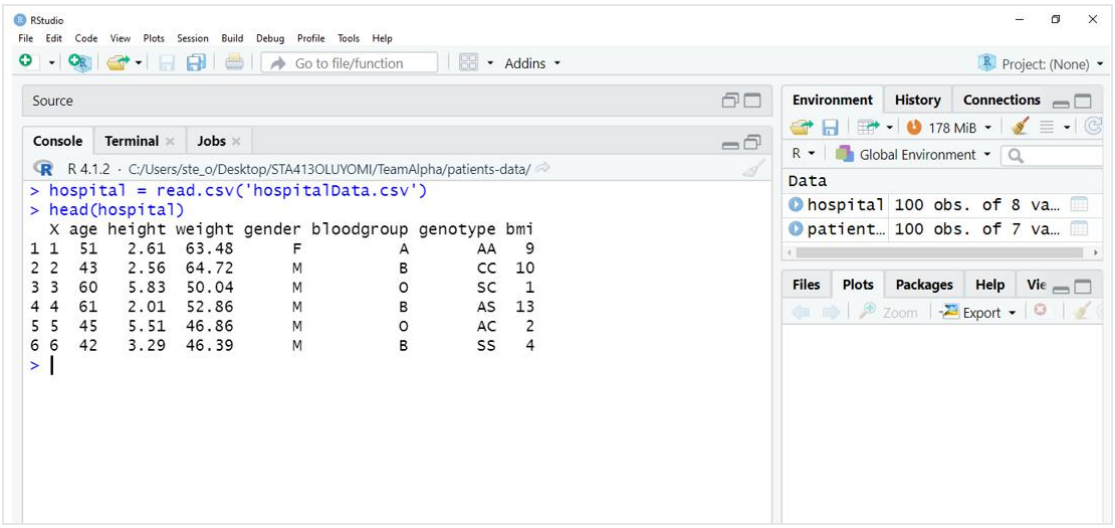
Appropriate graphs/plots are used to gain insight, examine the distribution of the variables and communicate our findings about the data overall.

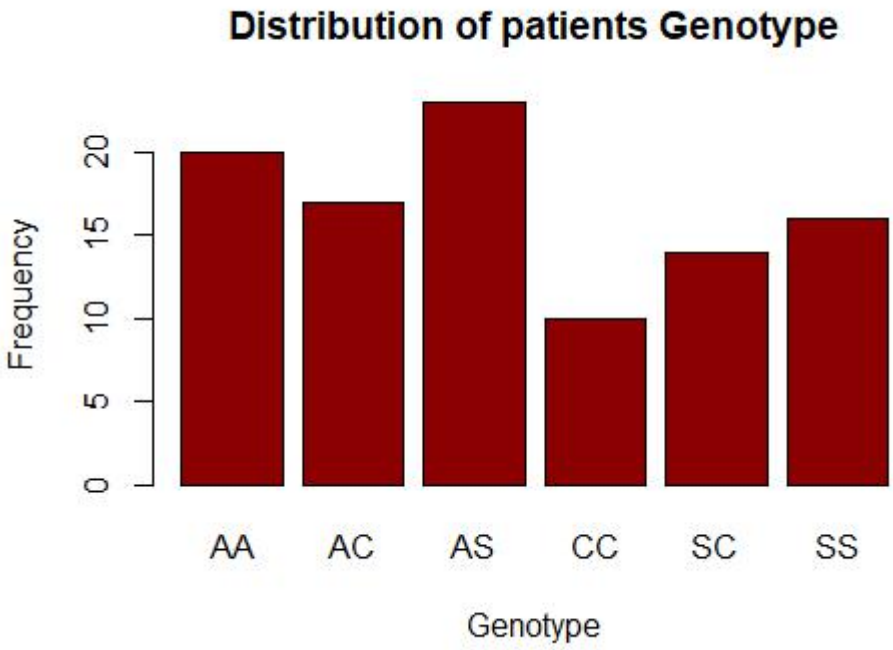
A full version of the analysis consisting of the plots, visualizations used, source code including this report is available on github.

NOTE: The souce code is written in R.

Data: patients-data
Format: CSV
Rows: 100
Variables: 7

Source code: https://github.com/stephen-lakes/STA413-data-analysis/tree/master/patients-data



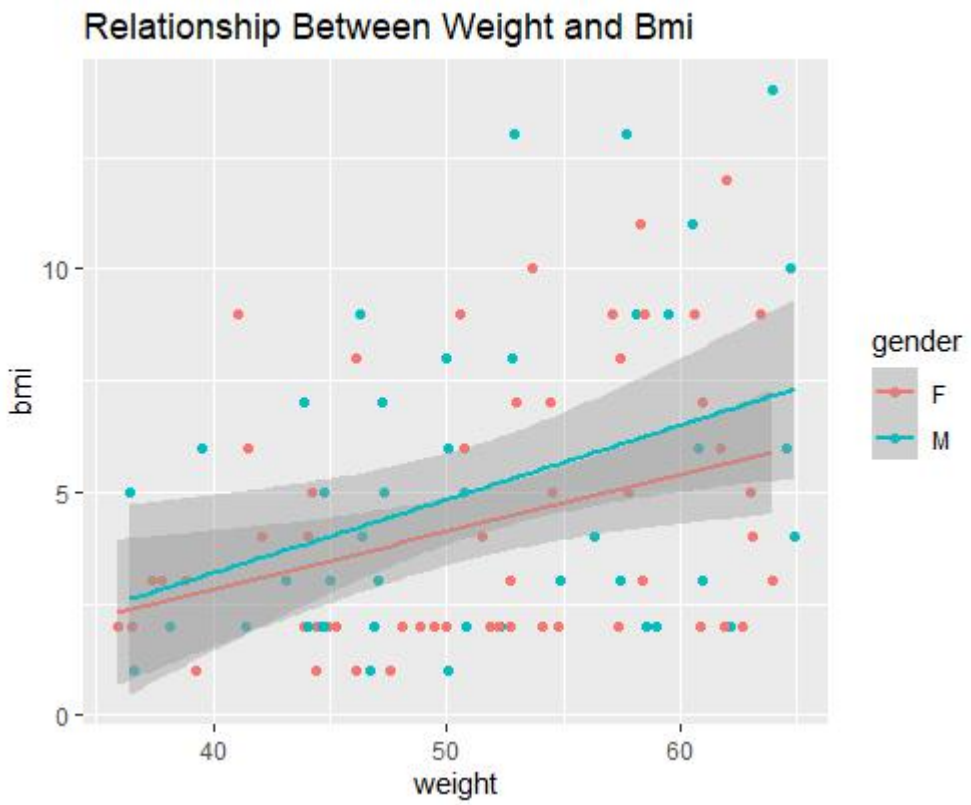


Observations:

- Most patients have an AS genotype.
- Few patients have a genotype of type CC.

Conclusion:

- Patients who use this hospital have type AS genotype.

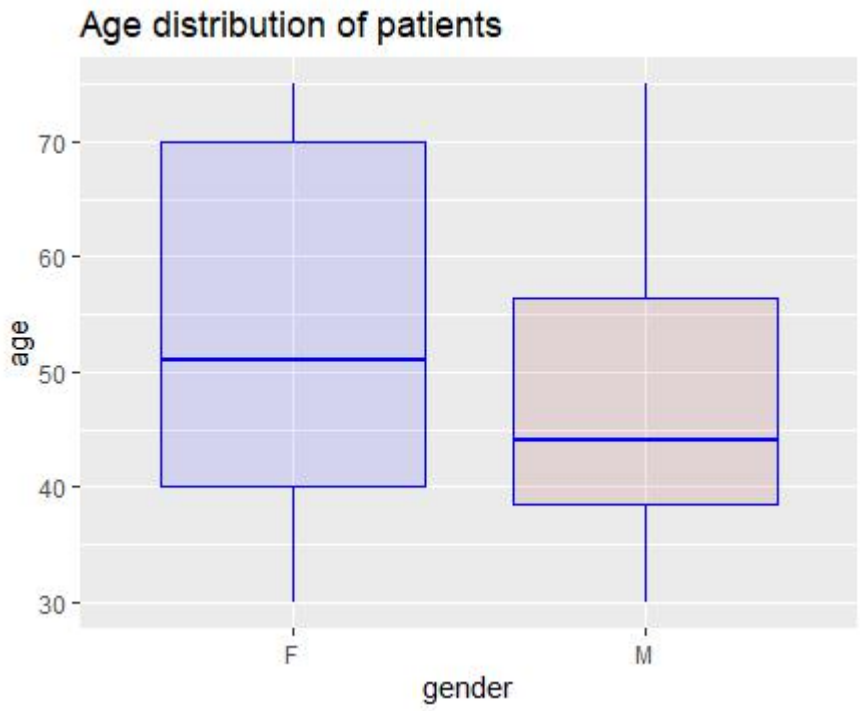


Observations:

- There is a positive relationship between weight and Bmi.
- The positive relationship between weight and Bmi in Males is much stronger than that of the Females.

Conclusion:

- Males Bmi tend to go up as they gain weight compared to Females.

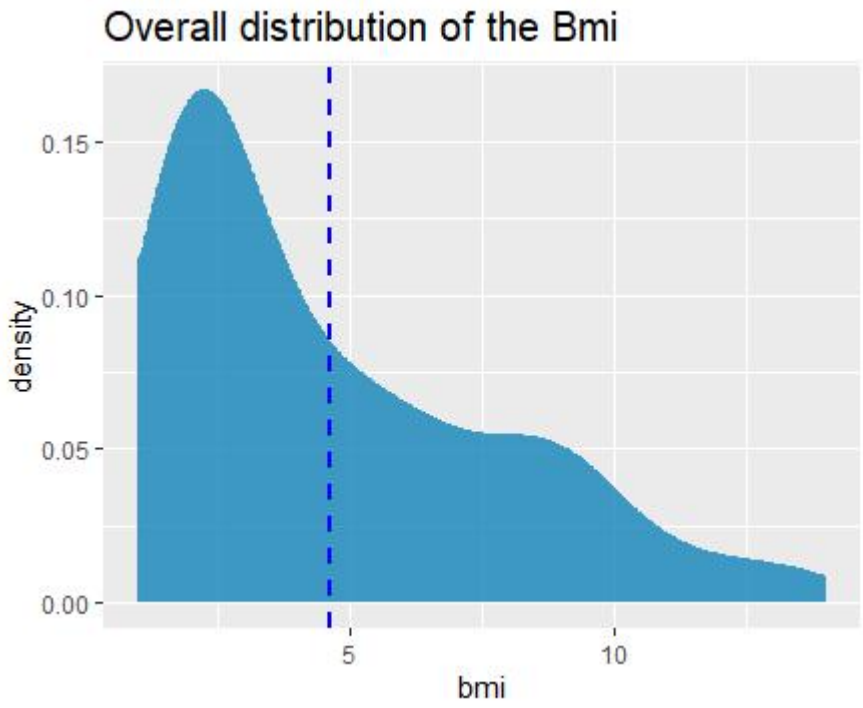


Observations:

- No patient is less than 30 years of age for both categories.
- There are more Male patients than Female patients in the hospital.
- Atleast 75% of the Male patients are 40 years or older.
- Atleast 75% of the Female patients are 40 years or older.

Conclusion:

- Males Bmi tend to go up as they gain weight compared to Females.

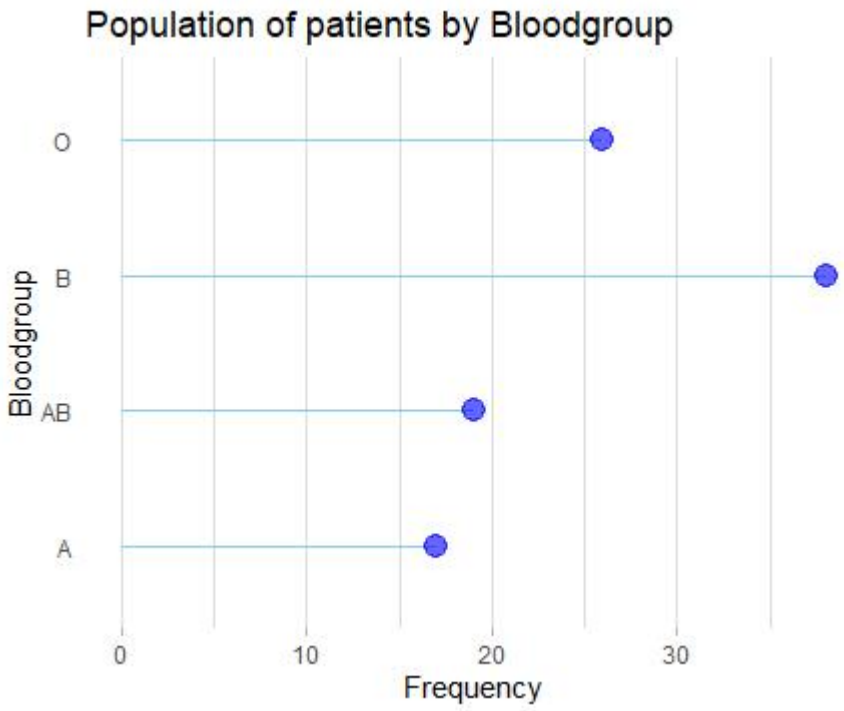


Observations:

- The overall Bmi of patients follows a chi square distribution.
- It's positive skewed with an average bmi of 4.61
- Overall average bmi of the patients is less than 5.

Conclusion:

- Males Bmi tend to go up as they gain weight compared to Females.



Observations:

- Most patients have a blood group of type B.
- Fewer of the patients have blood group of type A

Appendix

```
#  Libraries
library(ggplot2)
library(dplyr)

#  Load  hospitalData
hospital = read.csv('hospitalData.csv')

# UNIVARIATE QUANTITATIVE VARIABLE ( DENSITY PLOT )

ggplot(data = hospital, aes(x=bmi)) +
  geom_density(fill="#1185ba", color="#e9ecef", alpha=0.8) +
  geom_vline(aes(xintercept=mean(bmi)), color="blue", linetype="dashed", size=1) +
  ggtitle("Overall distribution of the Bmi") +
  theme(plot.title = element_text(size=15))

# SUMMARY STATISTICS
summary(hospital$bmi)
Min.    1st Qu.   Median    Mean   3rd Qu.    Max.
1.00      2.00       3.00       4.61      7.00       14.00


# BIVARIATE QUANITATIVE VARIABLE ( SCATTER PLOT )

ggplot(data = hospital, aes(x = weight, y = bmi, color = gender)) +
  geom_point() +
  geom_smooth(method="lm", se=TRUE, fullrange=FALSE, level=0.95) +
  ggtitle("Relationship Between Weight and Bmi")


# UNIVARIATE QUALITATIVE VARIABLE ( LOLLIPOP CHART, PIE CHART, BAR CHART )

# BLOOD GROUP
group = aggregate(hospital,by=list(hospital$bloodgroup), FUN=length)
data = data.frame(
  x= group$Group.1,
  y= group$X)

ggplot(data, aes(x=x, y=y)) +
  geom_segment( aes(x=x, xend=x, y=0, yend=y), color="skyblue") +
  geom_point( color="blue", size=4, alpha=0.6) +
  theme_light() +
  coord_flip() +
  theme(
    panel.grid.major.y = element_blank(),
    panel.border = element_blank(),
```

```
    axis.ticks.y = element_blank()
  ) +

  xlab("Bloodgroup") +
  ylab("Frequency") +
  ggtitle("Population of patients by Bloodgroup")

# Gender
data <- data %>%
  arrange(desc(Gender)) %>%
  mutate(prop = Freq / sum(data$Freq) *100) %>%
  mutate(ypos = cumsum(prop)- 0.5*prop )


ggplot(data,aes(x="", y=Freq, fill=Gender)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y") +
  theme_void() +
  geom_text(aes(y = ypos, label = Gender), color = "white", size=6)

# GENOTYPE
barplot(data$Freq,
        main = "Distribution of patients Genotype",
        xlab = "Genotype",
        ylab = "Frequency",
        names.arg = data$Genotype,
        col = "darkred",
        horiz = FALSE
)


# 2 QUALITATIVE 1 QUANITATIVE MULTIVARIATE VARIABLE (BOX PLOT)

ggplot(data = hospital, aes(x=gender, y=age)) +
  geom_boxplot(
    # custom boxes
    color="blue",
    fill= c("blue", "darkred"),
    alpha=0.1,

    # Notch?
    notch=FALSE,
    notchwidth = 0.5,

    # custom outliers
    outlier.colour="red",
    outlier.fill="red",
    outlier.size=3
  ) +
  ggtitle("Age distribution of patients")
```