Community Health Evaluation Data Analysis

Comprehensive Statistical Analysis and Modeling

Community Health and Opportunity Group

Introduction

Objective: Analyze community health evaluation data to understand:

- Patient demographics and service utilization
- Biomechanical measurements and their relationships
- Factors affecting patient satisfaction and quality of life
- Predictive modeling for health outcomes

Dataset: 347 participants with 12 variables including demographics, service types, biomechanical measures, and outcomes.

Data Import and Preprocessing

```r
# Load required libraries
library(tidyverse)
library(knitr)
library(kableExtra)
library(corrplot)
library(caret)
library(randomForest)
library(ggplot2)
library(plotly)
library(GGally)

# Import data
data <- readRDS("C:/Users/Hp/Desktop/ff-comm-health/data/Rdata.rds")
```

```r
# Display structure
str(data)
```

```
'data.frame':	347 obs. of  12 variables:
 $ Participant.ID              : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Age                         : int  56 69 46 32 60 25 38 56 36 40 ...
 $ Gender                      : chr  "F" "M" "M" "F" ...
 $ SES                         : int  4 1 4 1 3 1 3 1 3 1 ...
 $ Service.Type                : chr  "Rehab" "Preventive" "Rehab" "Consultation" ...
 $ Visit.Frequency             : chr  "Weekly" "Yearly" "Yearly" "Weekly" ...
 $ Step.Frequency..steps.min. : int  85 80 81 66 73 74 66 68 67 82 ...
 $ Stride.Length..m.           : num  0.54 0.7 0.57 0.78 0.84 0.9 0.6 0.58 0.55 0.82 ...
 $ Joint.Angle....             : num  18 13.1 29.9 28.5 20.8 ...
 $ EMG.Activity                : chr  "Low" "Moderate" "Moderate" "Moderate" ...
 $ Patient.Satisfaction..1.10.: int  1 8 4 9 5 3 9 9 4 4 ...
 $ Quality.of.Life.Score       : int  57 94 66 66 98 82 81 57 68 87 ...
```

Data Cleaning

```r
# Convert categorical variables to factors
data <- data %>%
  mutate(
    Gender = as.factor(Gender),
    SES = as.factor(SES),
    Service.Type = as.factor(Service.Type),
    Visit.Frequency = as.factor(Visit.Frequency),
    EMG.Activity = as.factor(EMG.Activity)
  )

# Check for missing values
missing_summary <- data %>%
  summarise(across(everything(), ~sum(is.na(.)))) %>%
  pivot_longer(everything(), names_to = "Variable", values_to = "Missing_Count")

kable(missing_summary, caption = "Missing Values Summary") %>%
  kable_styling(bootstrap_options = c("striped", "hover"))
```

Table 1: Missing Values Summary

Variable	Missing_Count
Participant.ID	0
Age	0
Gender	0
SES	0
Service.Type	0
Visit.Frequency	0
Step.Frequency..steps.min.	0
Stride.Length..m.	0
Joint.Angle....	0
EMG.Activity	0
Patient.Satisfaction..1.10.	0
Quality.of.Life.Score	0

Descriptive Statistics

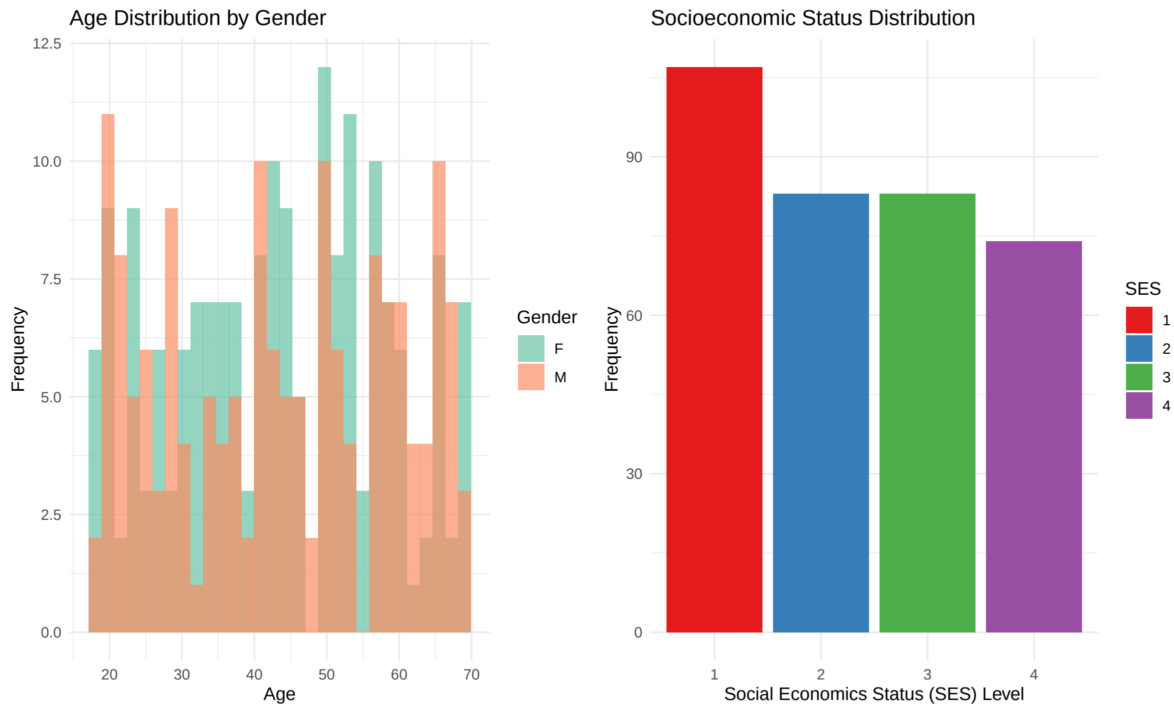
```
# Summary statistics for continuous variables
continuous_vars <- data %>%
  select(Age, Step.Frequency..steps.min., Stride.Length..m.,
         Joint.Angle...., Patient.Satisfaction..1.10., Quality.of.Life.Score)

summary_stats <- data.frame(
  Variable = names(continuous_vars),
  Mean = sapply(continuous_vars, mean, na.rm = TRUE),
  SD = sapply(continuous_vars, sd, na.rm = TRUE),
  Min = sapply(continuous_vars, min, na.rm = TRUE),
  Max = sapply(continuous_vars, max, na.rm = TRUE),
  Median = sapply(continuous_vars, median, na.rm = TRUE)
)

kable(summary_stats, digits = 2,
      caption = "Descriptive Statistics for Continuous Variables") %>%
  kable_styling(bootstrap_options = c("striped", "hover"))
```

Table 2: Descriptive Statistics for Continuous Variables

	Variable	Mean	SD	Min	Max	Median
Age	Age	43.37	15.18	18.00	69.00	43.00
Step.Frequency..steps.min.	Step.Frequency..steps.min.	80.12	11.26	60.00	99.00	81.00
Stride.Length..m.	Stride.Length..m.	0.75	0.14	0.50	1.00	0.76
Joint.Angle....	Joint.Angle....	20.06	5.81	10.06	29.97	20.19
Patient.Satisfaction..1.10.	Patient.Satisfaction..1.10.	5.21	2.83	1.00	10.00	5.00
Quality.of.Life.Score	Quality.of.Life.Score	74.20	13.95	50.00	99.00	74.00

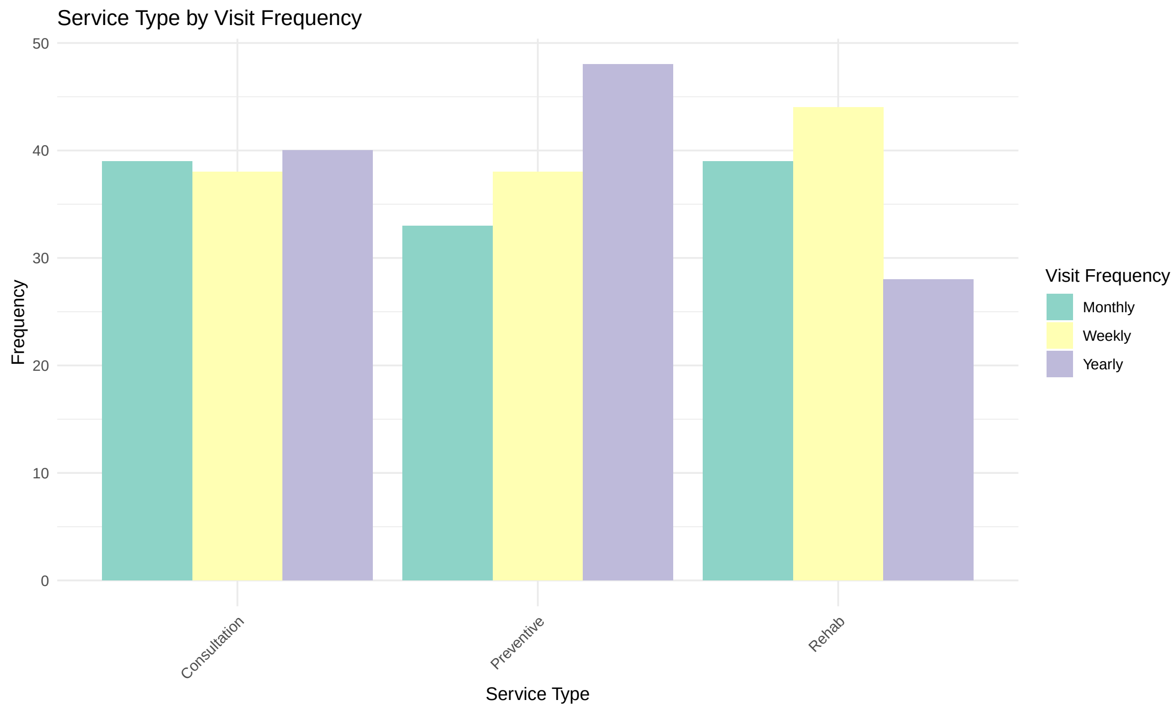
Demographic Distribution

```
# Age distribution by gender
p1 <- ggplot(data, aes(x = Age, fill = Gender)) +
  geom_histogram(bins = 30, alpha = 0.7, position = "identity") +
  theme_minimal() +
  labs(title = "Age Distribution by Gender", x = "Age", y = "Frequency") +
  scale_fill_brewer(palette = "Set2")

# SES distribution
p2 <- ggplot(data, aes(x = SES, fill = SES)) +
  geom_bar() +
  theme_minimal() +
  labs(title = "Socioeconomic Status Distribution", x = "Social Economics Status (SES) Level"
  scale_fill_brewer(palette = "Set1")

gridExtra::grid.arrange(p1, p2, ncol = 2)
```



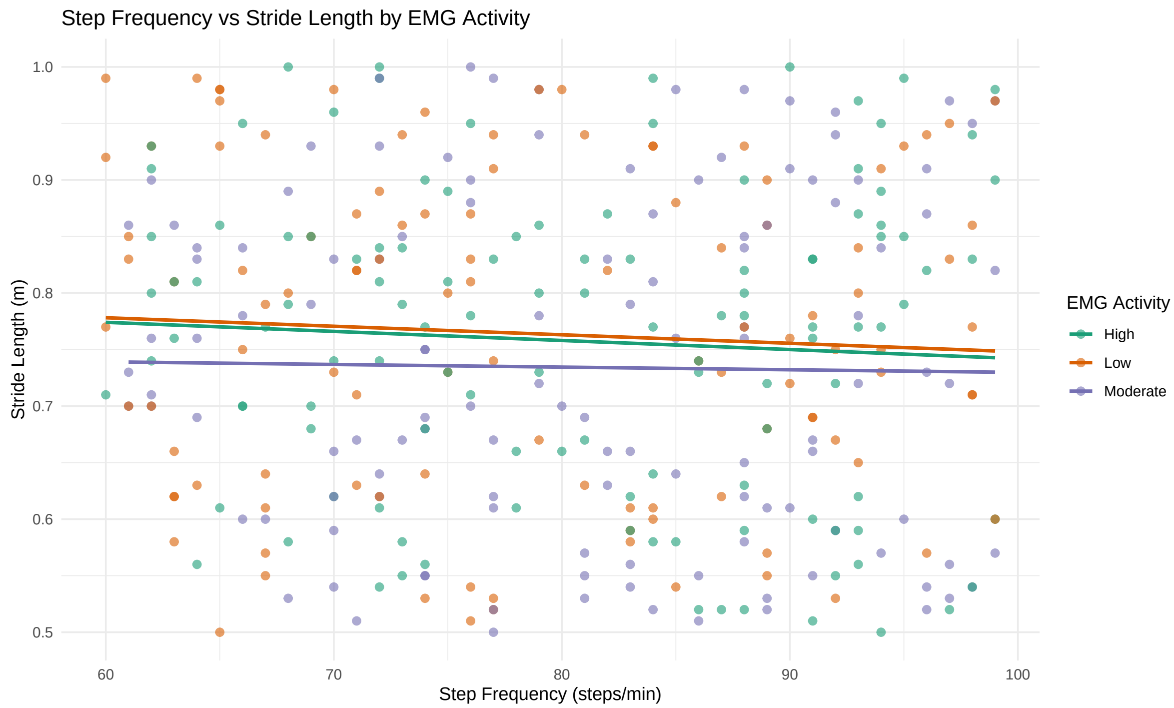
Service Utilization

```r
# Service type by visit frequency
service_visit <- data %>%
  group_by(Service.Type, Visit.Frequency) %>%
  summarise(Count = n(), .groups = "drop")

ggplot(service_visit, aes(x = Service.Type, y = Count, fill = Visit.Frequency)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  labs(title = "Service Type by Visit Frequency",
       x = "Service Type", y = "Frequency", fill = "Visit Frequency") +
  scale_fill_brewer(palette = "Set3") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



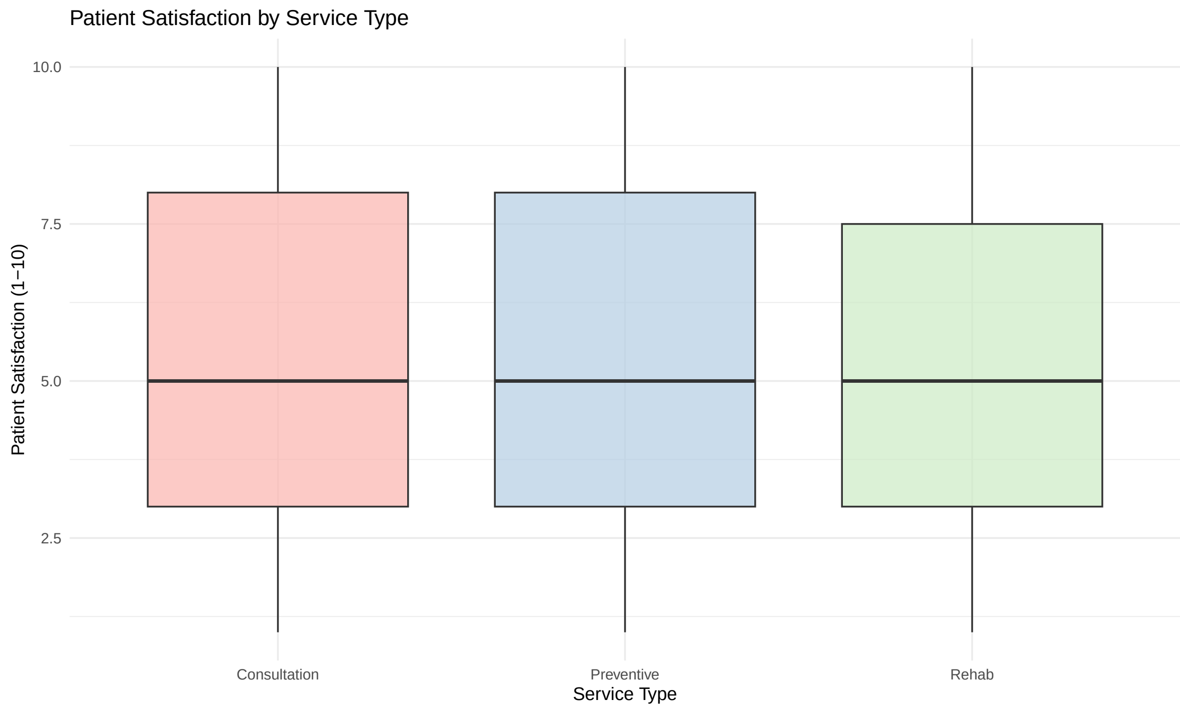
Biomechanical Measures Overview

```
# Step frequency vs stride length
ggplot(data, aes(x = Step.Frequency..steps.min., y = Stride.Length..m.,
                 color = EMG.Activity)) +
  geom_point(alpha = 0.6, size = 2) +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal() +
  labs(title = "Step Frequency vs Stride Length by EMG Activity",
       x = "Step Frequency (steps/min)", y = "Stride Length (m)",
       color = "EMG Activity") +
  scale_color_brewer(palette = "Dark2")
```



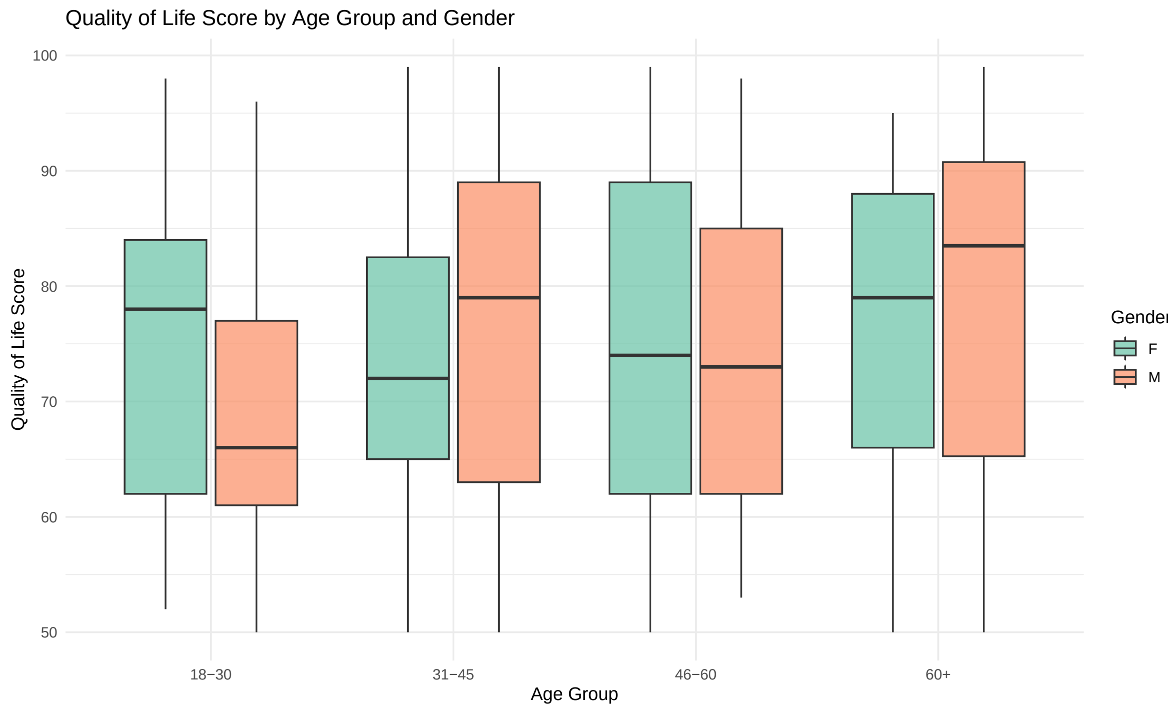
Patient Satisfaction Analysis

```
# Satisfaction by service type
ggplot(data, aes(x = Service.Type, y = Patient.Satisfaction..1.10.,
                 fill = Service.Type)) +
  geom_boxplot(alpha = 0.7) +
  theme_minimal() +
  labs(title = "Patient Satisfaction by Service Type",
       x = "Service Type", y = "Patient Satisfaction (1-10)") +
  scale_fill_brewer(palette = "Pastel1") +
  theme(legend.position = "none")
```



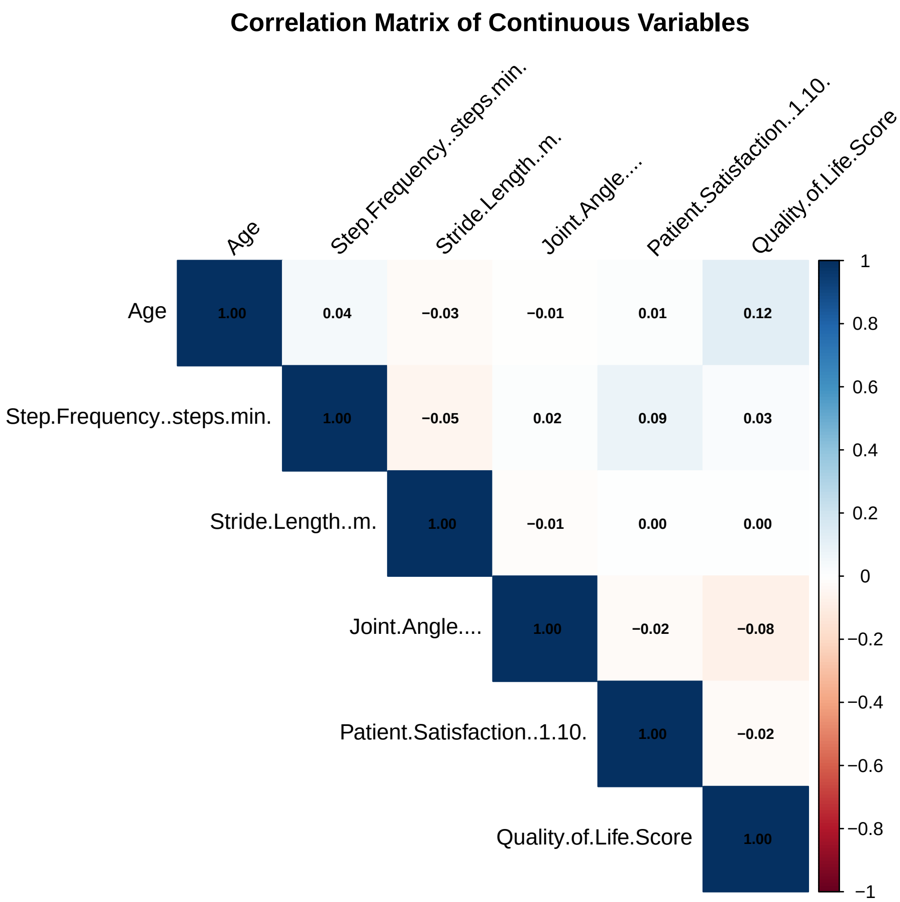
Quality of Life Analysis

```
# QoL by age groups and gender
data_age_groups <- data %>%
  mutate(Age_Group = cut(Age, breaks = c(0, 30, 45, 60, 100),
                         labels = c("18-30", "31-45", "46-60", "60+")))

ggplot(data_age_groups, aes(x = Age_Group, y = Quality.of.Life.Score,
                            fill = Gender)) +
  geom_boxplot(alpha = 0.7) +
  theme_minimal() +
  labs(title = "Quality of Life Score by Age Group and Gender",
       x = "Age Group", y = "Quality of Life Score") +
  scale_fill_brewer(palette = "Set2")
```



Correlation Analysis

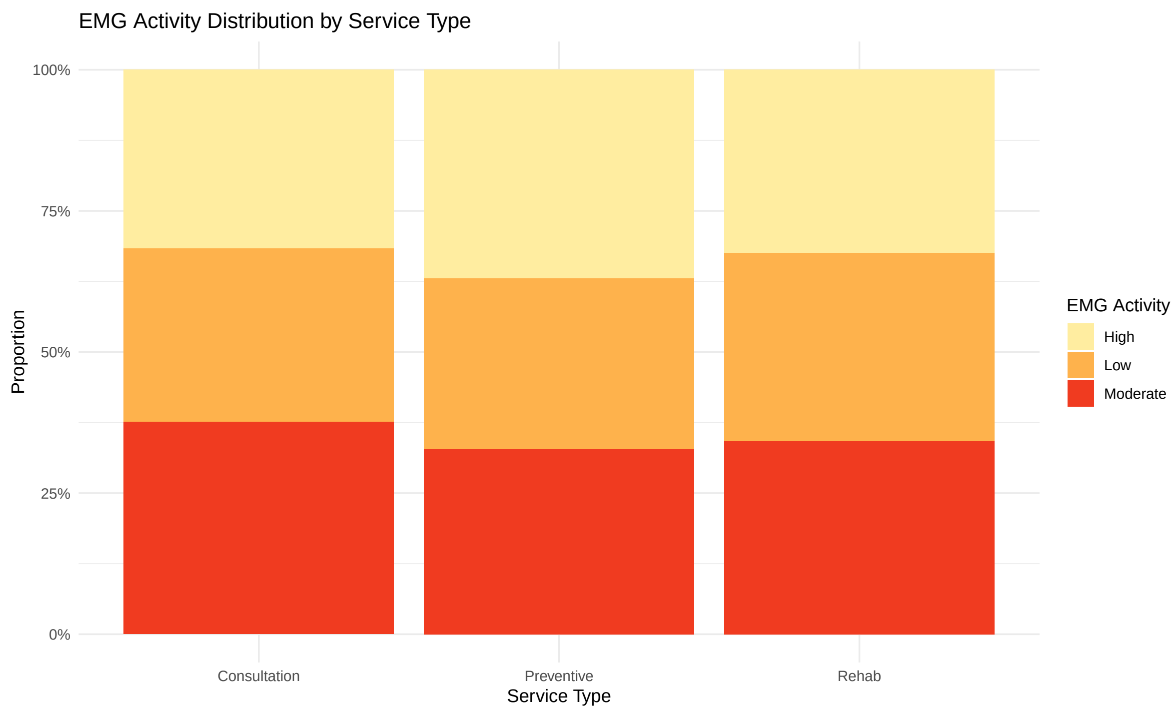
```
# Select numeric variables for correlation
numeric_data <- data %>%
  select(Age, Step.Frequency..steps.min., Stride.Length..m.,
         Joint.Angle...., Patient.Satisfaction..1.10.,
         Quality.of.Life.Score)

# Compute correlation matrix
cor_matrix <- cor(numeric_data, use = "complete.obs")

# Plot correlation matrix
corrplot(cor_matrix, method = "color", type = "upper",
         tl.col = "black", tl.srt = 45,
         addCoef.col = "black", number.cex = 0.7,
         title = "Correlation Matrix of Continuous Variables",
         mar = c(0,0,2,0))
```



EMG Activity Patterns

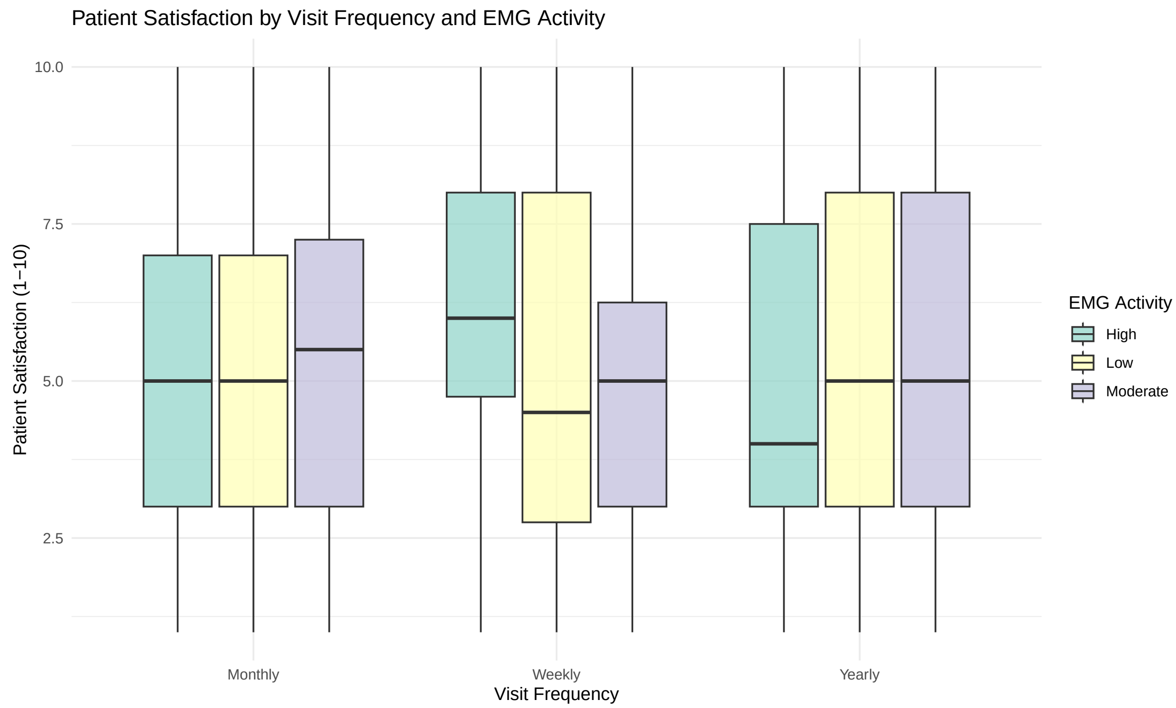
```r
# EMG activity distribution across service types
emg_service <- data %>%
  group_by(Service.Type, EMG.Activity) %>%
  summarise(Count = n(), .groups = "drop") %>%
  group_by(Service.Type) %>%
  mutate(Percentage = Count / sum(Count) * 100)

ggplot(emg_service, aes(x = Service.Type, y = Percentage, fill = EMG.Activity)) +
  geom_bar(stat = "identity", position = "fill") +
  theme_minimal() +
```

```r
  labs(title = "EMG Activity Distribution by Service Type",
       x = "Service Type", y = "Proportion", fill = "EMG Activity") +
  scale_y_continuous(labels = scales::percent) +
  scale_fill_brewer(palette = "YlOrRd")
```



Key Insights: Satisfaction Factors

```r
# Satisfaction by EMG activity and visit frequency
ggplot(data, aes(x = Visit.Frequency, y = Patient.Satisfaction..1.10.,
                 fill = EMG.Activity)) +
  geom_boxplot(alpha = 0.7) +
  theme_minimal() +
  labs(title = "Patient Satisfaction by Visit Frequency and EMG Activity",
       x = "Visit Frequency", y = "Patient Satisfaction (1-10)",
       fill = "EMG Activity") +
  scale_fill_brewer(palette = "Set3")
```



Predictive Modeling: Data Preparation

```
# Prepare data for modeling
set.seed(123)

# Create training and testing sets
trainIndex <- createDataPartition(data$Quality.of.Life.Score,
                                  p = 0.8, list = FALSE)
train_data <- data[trainIndex, ]
test_data <- data[-trainIndex, ]

# Display split
cat("Training set size:", nrow(train_data), "\n")
```

```
Training set size: 279
```

```r
cat("Testing set size:", nrow(test_data), "\n")
```

```
Testing set size: 68
```

Linear Regression Model

```r
# Build linear regression model for Quality of Life
lm_model <- lm(Quality.of.Life.Score ~ Age + Gender + SES + Service.Type +
                 Visit.Frequency + Step.Frequency..steps.min. +
                 Stride.Length..m. + Joint.Angle.... + EMG.Activity +
                 Patient.Satisfaction..1.10.,
               data = train_data)

# Model summary
summary_lm <- summary(lm_model)
cat("R-squared:", round(summary_lm$r.squared, 4), "\n")
```

```
R-squared: 0.0436
```

```r
cat("Adjusted R-squared:", round(summary_lm$adj.r.squared, 4), "\n")
```

```
Adjusted R-squared: -0.011
```

```r
cat("RMSE:", round(sqrt(mean(lm_model$residuals^2)), 4), "\n")
```

```
RMSE: 13.6999
```

Model Coefficients