Appendix- R-CODE

```r
library(ggplot2)
library(dplyr)
library(knitr)
data = read.csv("CardioGoodFitness.csv")
kable(head(data, 5), caption = "First 5 Rows of CardioGoodFitness Data")
```

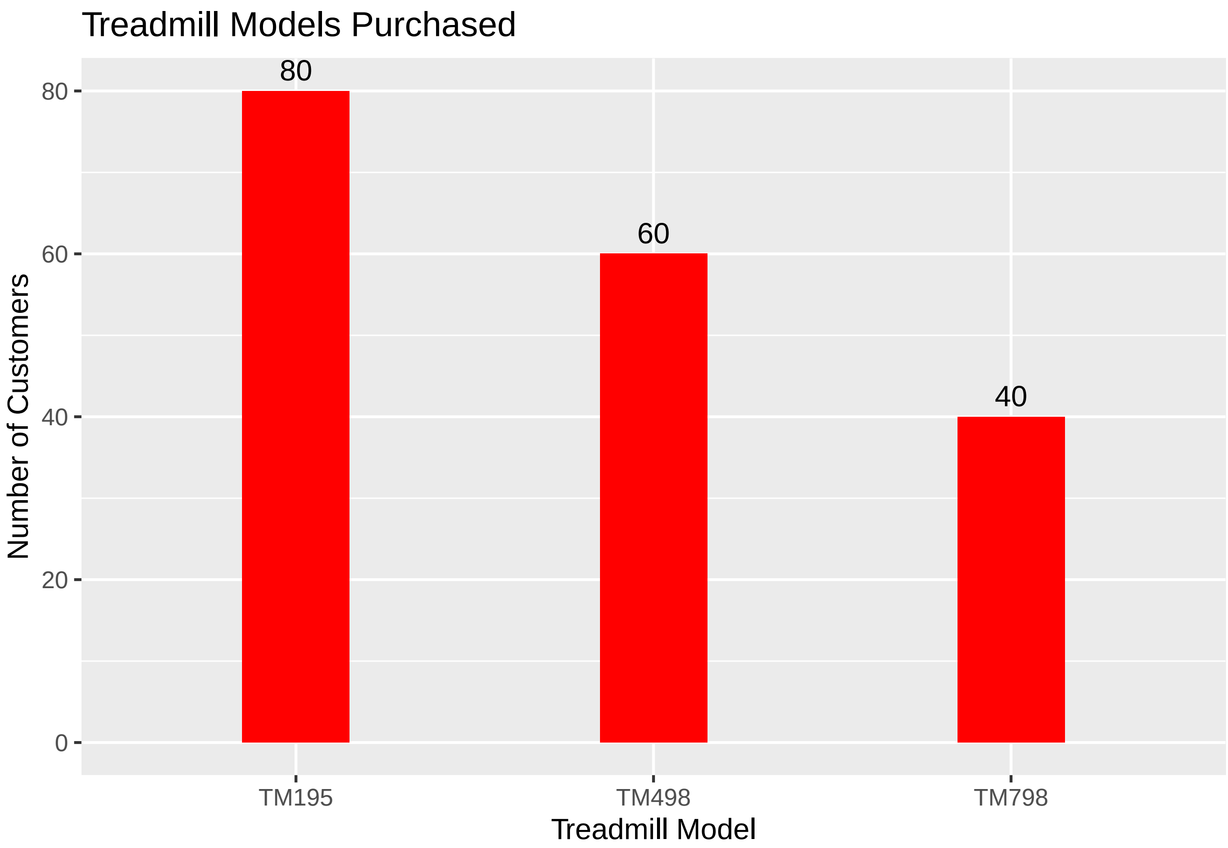
Table 1: First 5 Rows of CardioGoodFitness Data

Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
TM195	18	Male	14	Single	3	4	29562	112
TM195	19	Male	15	Single	2	3	31836	75
TM195	19	Female	14	Partnered	4	3	30699	66
TM195	19	Male	12	Single	3	3	32973	85
TM195	20	Male	13	Partnered	4	2	35247	47

Univariate Analysis

1. Product

```r
ggplot(data, aes(x = Product)) +
  geom_bar(fill = "red", width = 0.3) +
  geom_text(stat = 'count', aes(label = ..count..), vjust = -0.5) +
  labs(title = "Treadmill Models Purchased",
       x = "Treadmill Model",
       y = "Number of Customers")
```

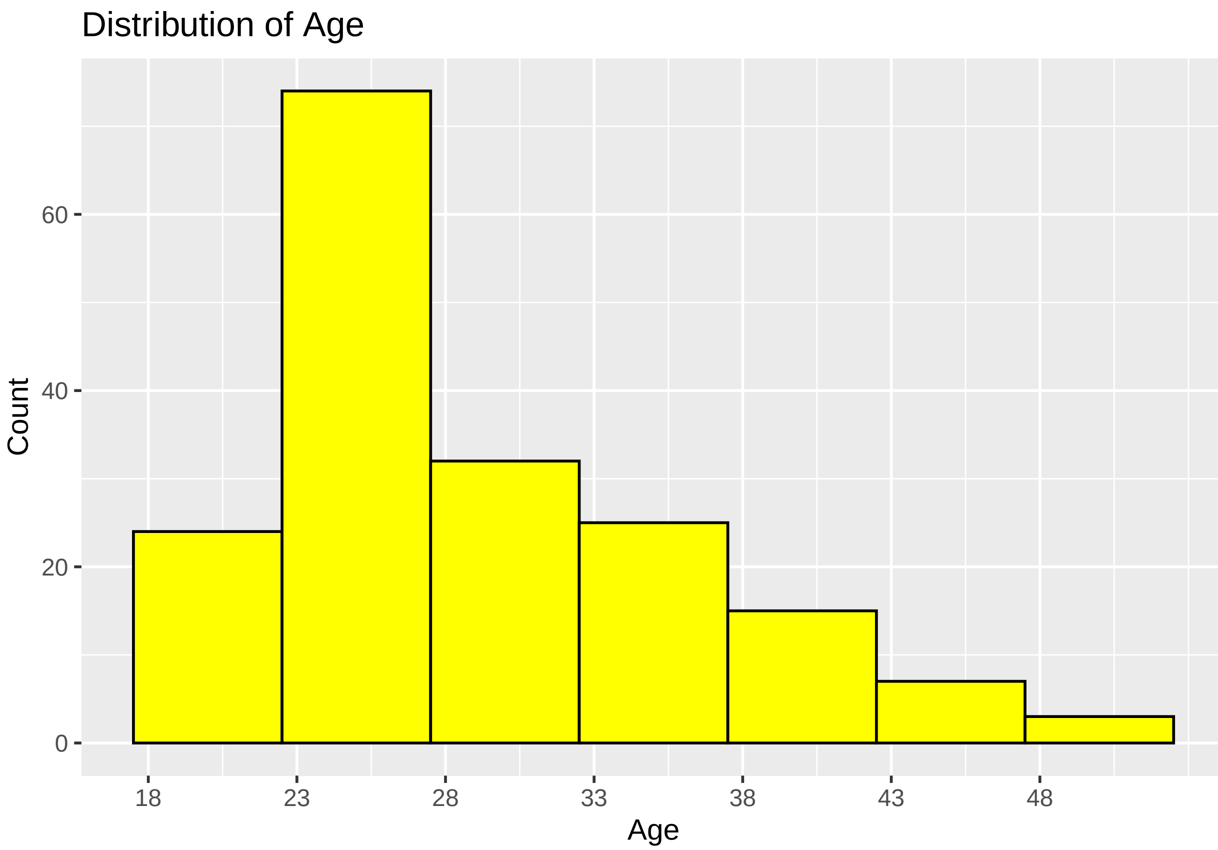


2. Gender

```
ggplot(data, aes(x = Gender)) +
  geom_bar(fill = "Purple", width = 0.2) +
  labs(title = "Treadmill Purchases by Gender", x = "Gender", y = "Number of Customers")
```

3. Age

```r
ggplot(data, aes(x = Age)) +
  geom_histogram(binwidth = 5, fill = "yellow", color = "black") +
  labs(title = "Distribution of Age", x = "Age", y = "Count") +
  scale_x_continuous(breaks = seq(min(data$Age), max(data$Age), by = 5))
```



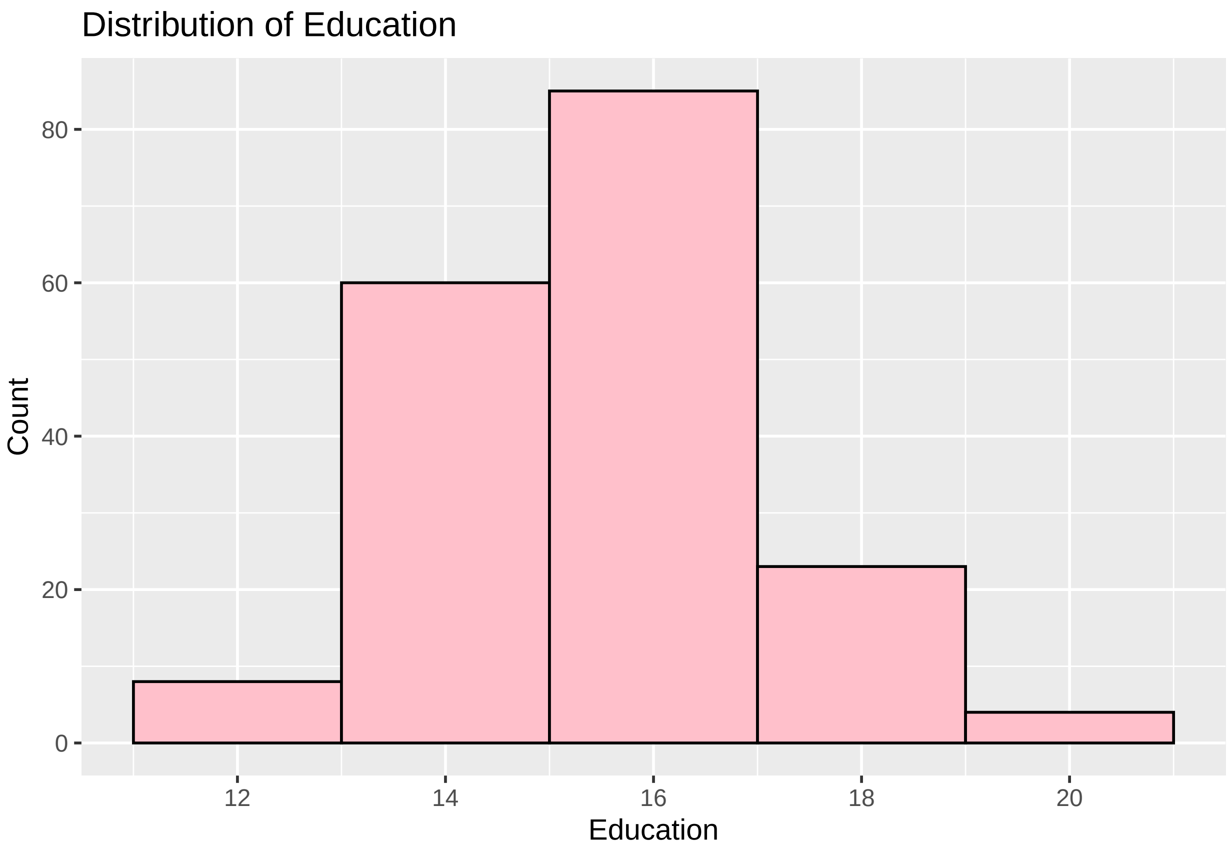

```
age_summary <- summary(data$Age)
age_summary_df <- data.frame(Statistic = names(age_summary), Value = as.numeric(age_summary))
kable(age_summary_df, caption = "Age Summary")
```

Table 2: Age Summary

Statistic	Value
Min.	18.00000
1st Qu.	24.00000
Median	26.00000
Mean	28.78889
3rd Qu.	33.00000
Max.	50.00000

4. Education

```
ggplot(data, aes(x = Education)) +
  geom_histogram(binwidth = 2, fill = "pink", color = "black") +
  labs(title = "Distribution of Education", x = "Education", y = "Count") +
  scale_x_continuous(breaks = seq(min(data$Education), max(data$Education), by = 2))
```



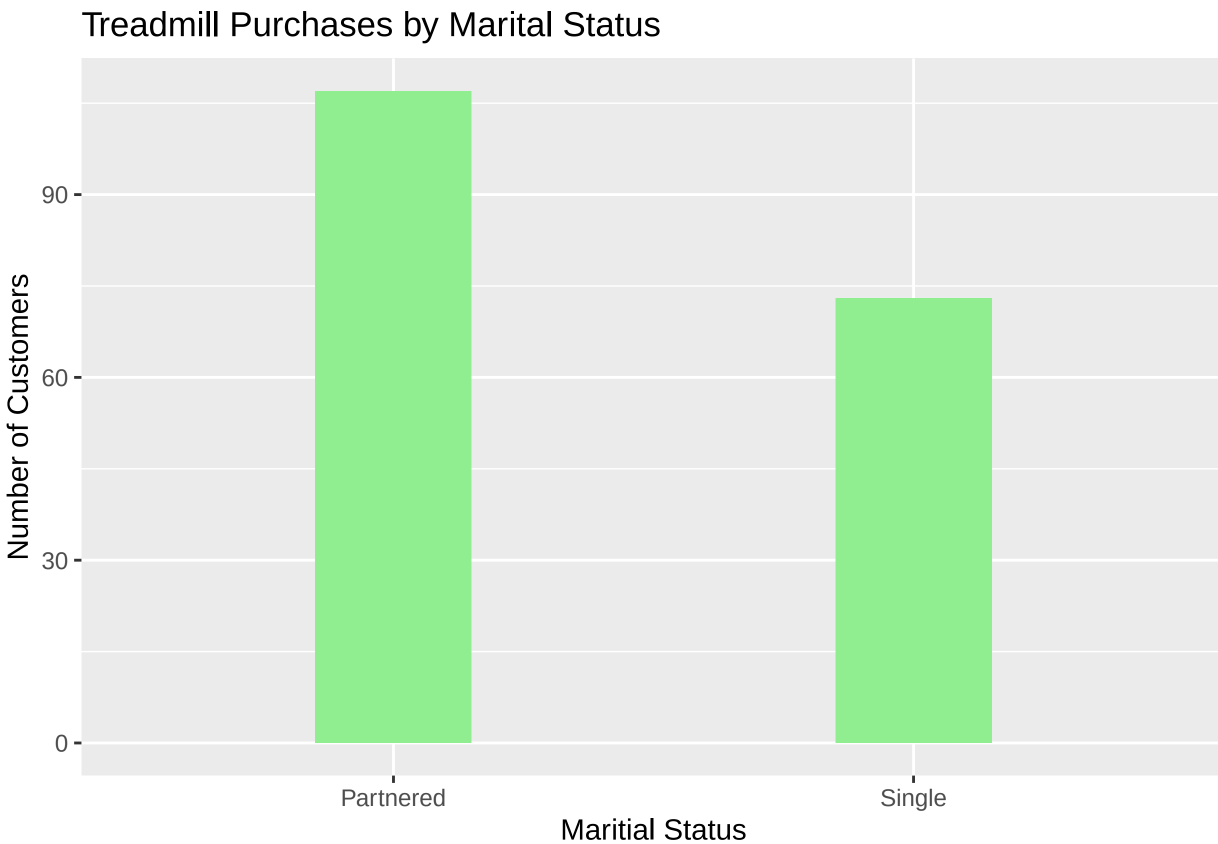

```
education_summary <- summary(data$Education)
education_summary_df <- data.frame(Statistic = names(education_summary), Value = as.numeric(education_su
kable(education_summary_df, caption = "Education Summary")
```

Table 3: Education Summary

Statistic	Value
Min.	12.00000
1st Qu.	14.00000
Median	16.00000
Mean	15.57222
3rd Qu.	16.00000
Max.	21.00000

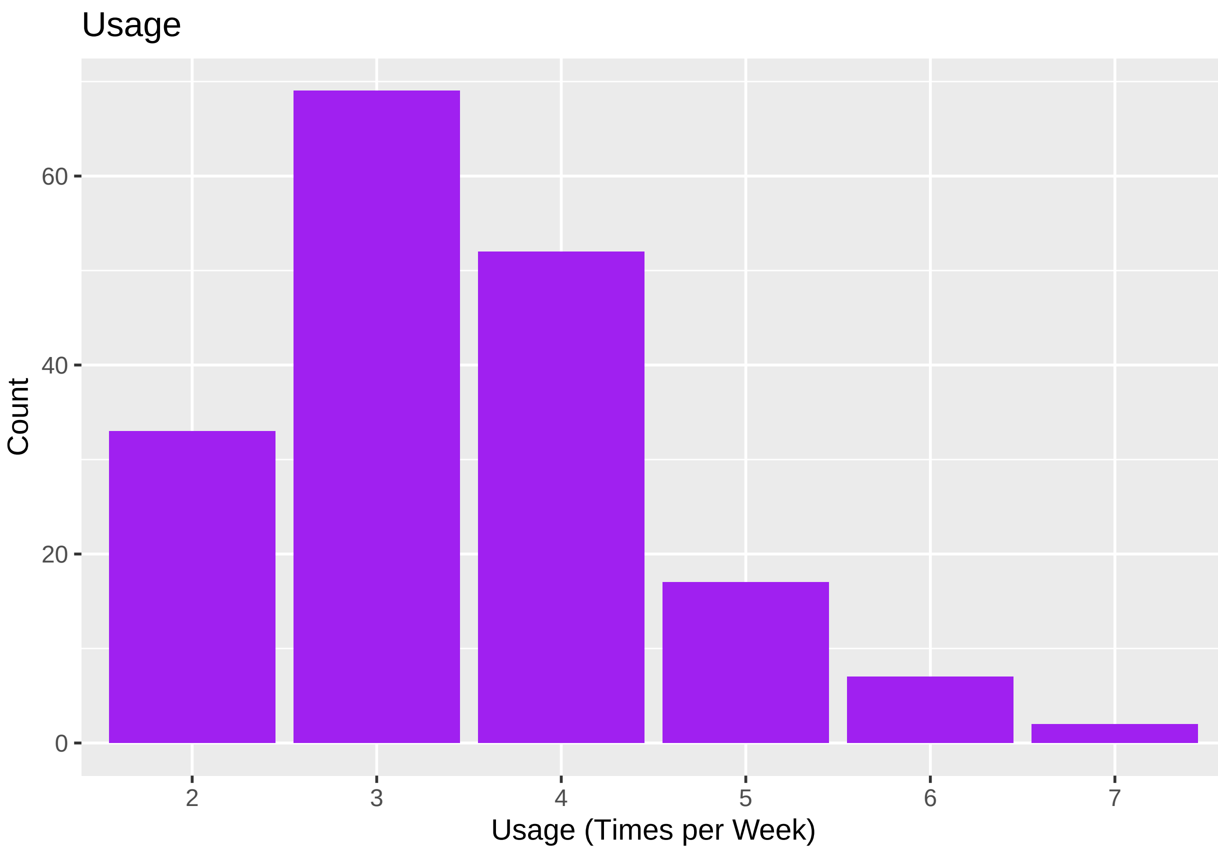
5. Martial Status

```
ggplot(data, aes(x = MaritalStatus)) +
  geom_bar(fill = "LightGreen", width = 0.3) +
  labs(title = "Treadmill Purchases by Marital Status", x = "Maritial Status", y = "Number of Customers
```



6. Usage

```
ggplot(data, aes(x = as.factor(Usage))) +
  geom_bar(fill = "purple") +
  labs(title = "Usage", x = "Usage (Times per Week)", y = "Count")
```



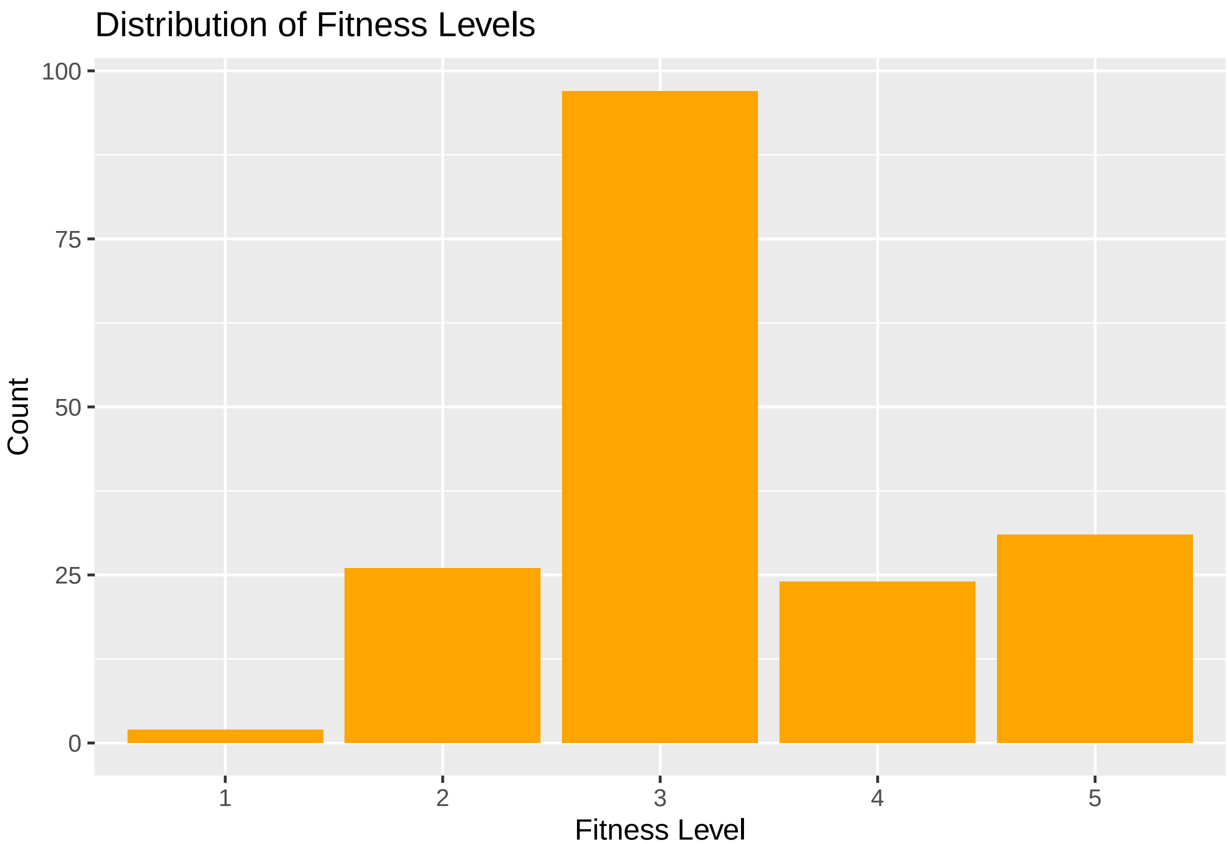

```
usage_summary = summary(data$Usage)
usage_summary_df = data.frame(Statistic = names(usage_summary), Value = as.numeric(usage_summary))
kable(usage_summary_df, caption = "Usage Summary")
```

Table 4: Usage Summary

Statistic	Value
Min.	2.000000
1st Qu.	3.000000
Median	3.000000
Mean	3.455556
3rd Qu.	4.000000
Max.	7.000000

7. Fitness

```
ggplot(data, aes(x = as.factor(Fitness))) +
  geom_bar(fill = "orange") +
  labs(title = "Distribution of Fitness Levels", x = "Fitness Level", y = "Count")
```




```
fitness_summary <- summary(data$Fitness)
fitness_summary_df <- data.frame(Statistic = names(fitness_summary), Value = as.numeric(fitness_summary
kable(fitness_summary_df, caption = "Fitness Summary")
```

Table 5: Fitness Summary

Statistic	Value
Min.	1.000000
1st Qu.	3.000000
Median	3.000000
Mean	3.311111
3rd Qu.	4.000000
Max.	5.000000

:

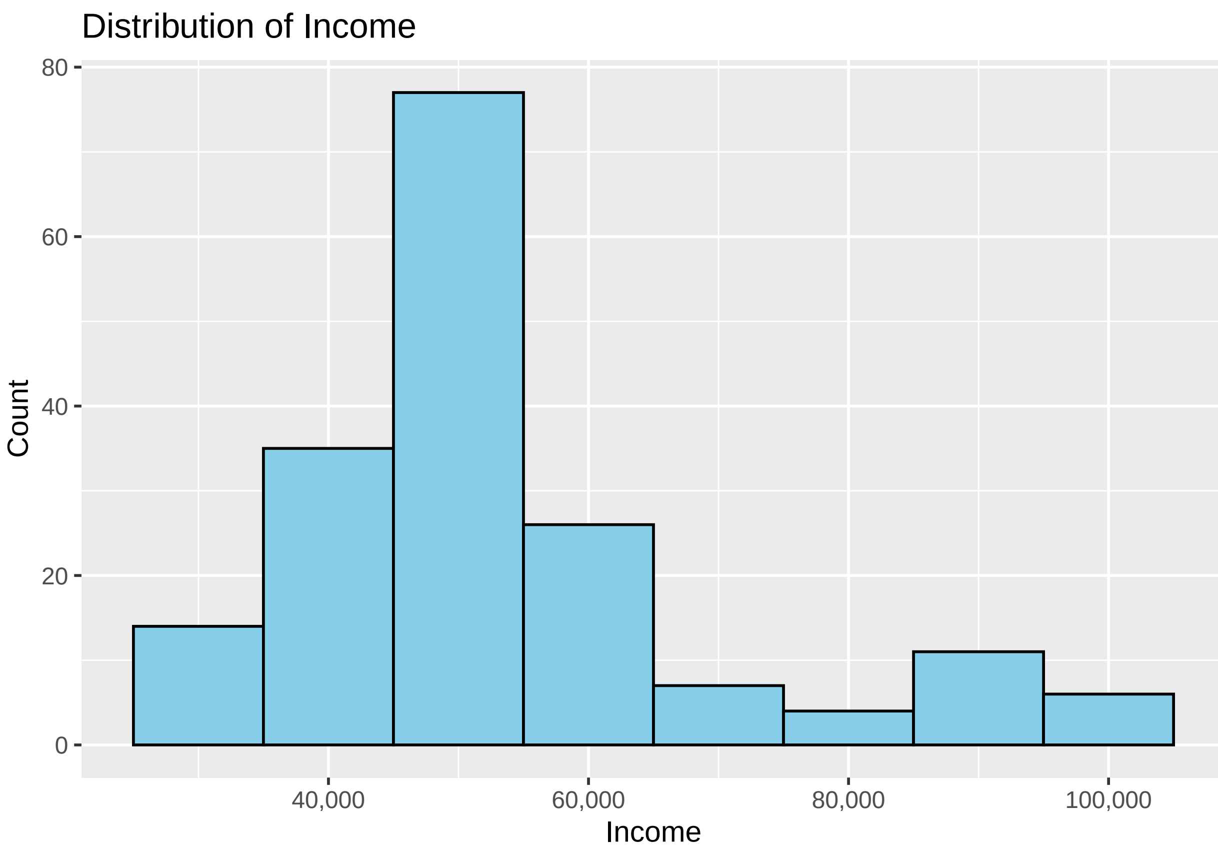
```
self_rate = data %>% summarise(Avg_Rating = mean(Fitness))
kable(self_rate, caption = "Average self-rated fitness of customers buying treadmill")
```

Table 6: Average self-rated fitness of customers buying treadmill

Avg_Rating
3.311111

8. Income

```r
ggplot(data, aes(x = Income)) +
  geom_histogram(binwidth = 10000, fill = "skyblue", color = "black") +
  labs(title = "Distribution of Income", x = "Income", y = "Count") +
  scale_x_continuous(labels = scales::comma)
```




```r
income_summary <- summary(data$Income)
income_summary_df <- data.frame(Statistic = names(income_summary), Value = as.numeric(income_summary))
kable(income_summary_df, caption = "Income Summary")
```

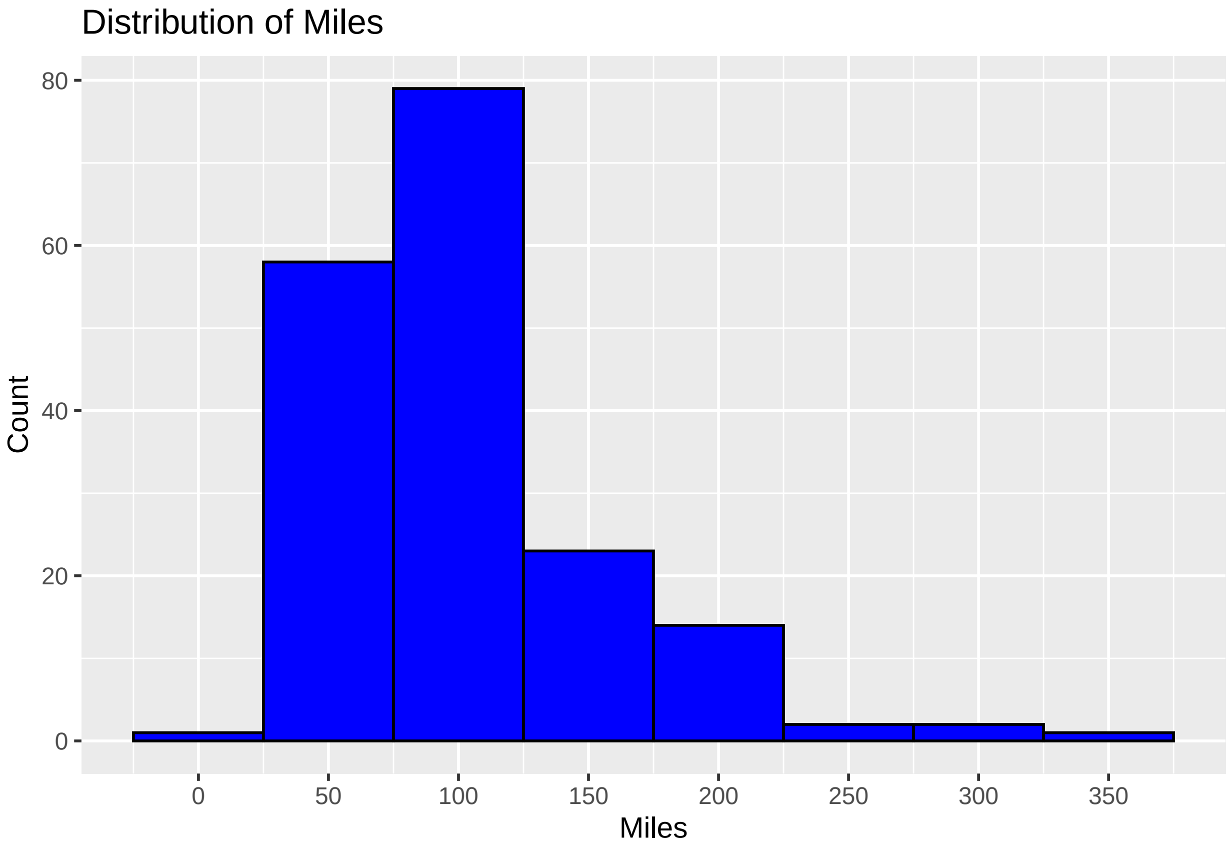
Table 7: Income Summary

Statistic	Value
Min.	29562.00
1st Qu.	44058.75
Median	50596.50
Mean	53719.58
3rd Qu.	58668.00
Max.	104581.00

9. Miles

```r
ggplot(data, aes(x = Miles)) +
  geom_histogram(binwidth = 50, fill = "blue", color = "black") +
  labs(title = "Distribution of Miles", x = "Miles", y = "Count") +
```

```
  scale_x_continuous(breaks = seq(0, max(data$Miles), by = 50))  # Customizing axis labels
```




```
miles_summary <- summary(data$Miles)
miles_summary_df <- data.frame(Statistic = names(miles_summary), Value = as.numeric(miles_summary))
kable(miles_summary_df, caption = "Miles Summary")
```

Table 8: Miles Summary

Statistic	Value
Min.	21.0000
1st Qu.	66.0000
Median	94.0000
Mean	103.1944
3rd Qu.	114.7500
Max.	360.0000

Table 9: Average Days and Miles Expected on Treadmill

Avg_Usage	Avg_Miles
3.455556	103.1944

```
averages = data %>% summarize(Avg_Income = mean(Income), Avg_Age = mean(Age), Avg_Education = mean(Educa
kable(averages, caption = "Averages of Treadmill Buyers")
```

Table 10: Averages of Treadmill Buyers

Avg_Income	Avg_Age	Avg_Education
53719.58	28.78889	15.57222

Bivariate Analysis

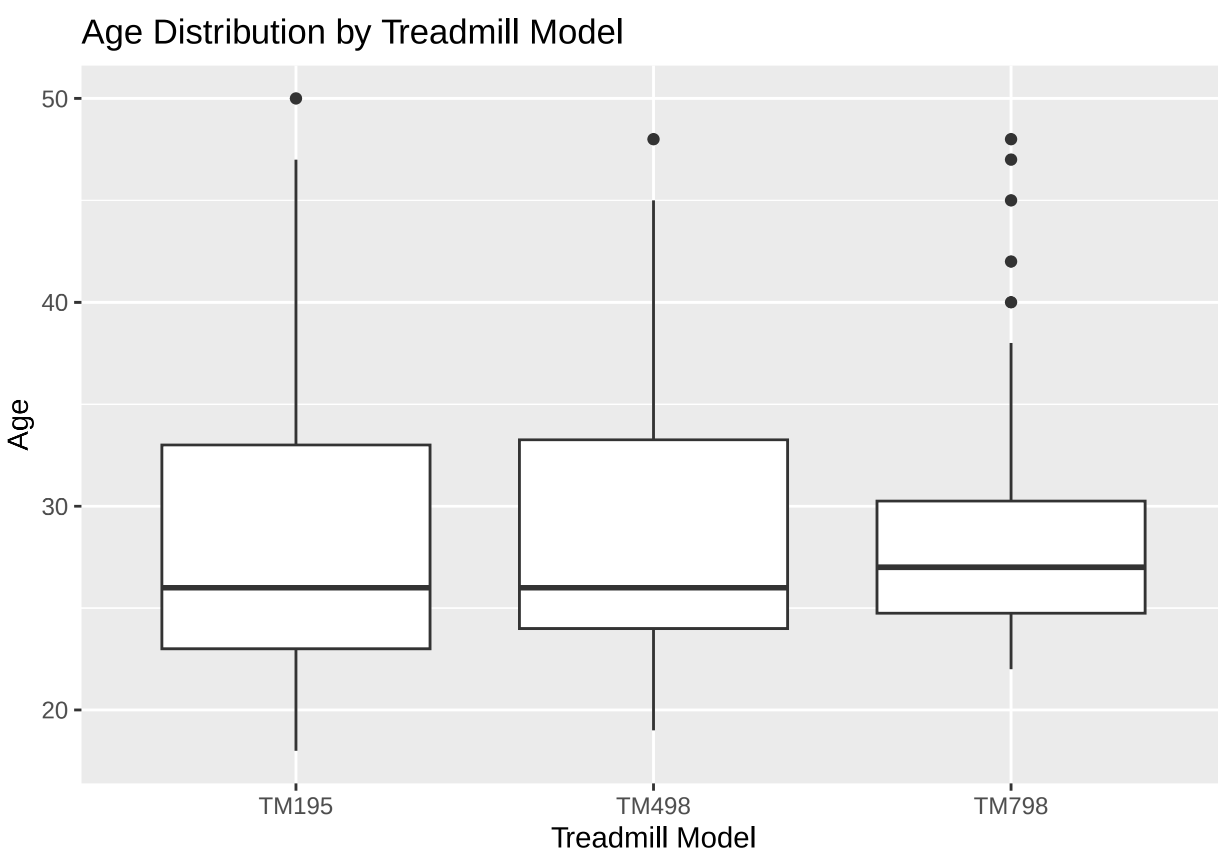
1. Age vs Model

```
anova_age_model = aov(Age ~ Product, data = data)
anova_result = summary(anova_age_model)
kable(anova_result[[1]], caption = "ANOVA Table: Age vs Product")
```

Table 11: ANOVA Table: Age vs Product

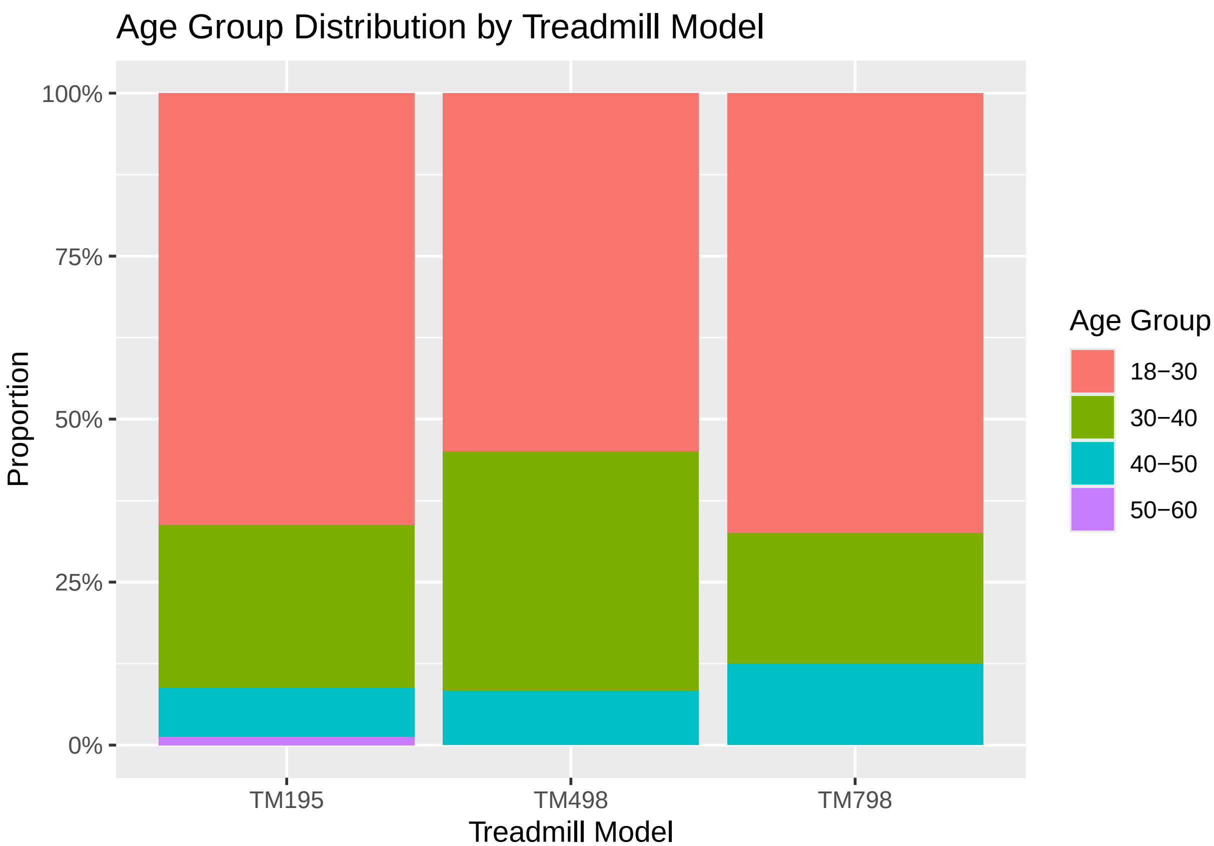
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Product	2	9.177778	4.588889	0.0942179	0.9101301
Residuals	177	8620.800000	48.705085	NA	NA

```
ggplot(data, aes(x = Product, y = Age)) +
  geom_boxplot() +
  labs(title = "Age Distribution by Treadmill Model", x = "Treadmill Model", y = "Age")
```




```
library(scales)
data$AgeGroup <- cut(data$Age, breaks = c(18, 30, 40, 50, 60),
```

```
                     labels = c("18-30", "30-40", "40-50", "50-60"), right = FALSE)
ggplot(data, aes(x = Product, fill = AgeGroup)) +
  geom_bar(position = "fill") +
  labs(title = "Age Group Distribution by Treadmill Model", x = "Treadmill Model",
       y = "Proportion",
       fill = "Age Group") +
  scale_y_continuous(labels = percent_format())
```



2. Income vs Model

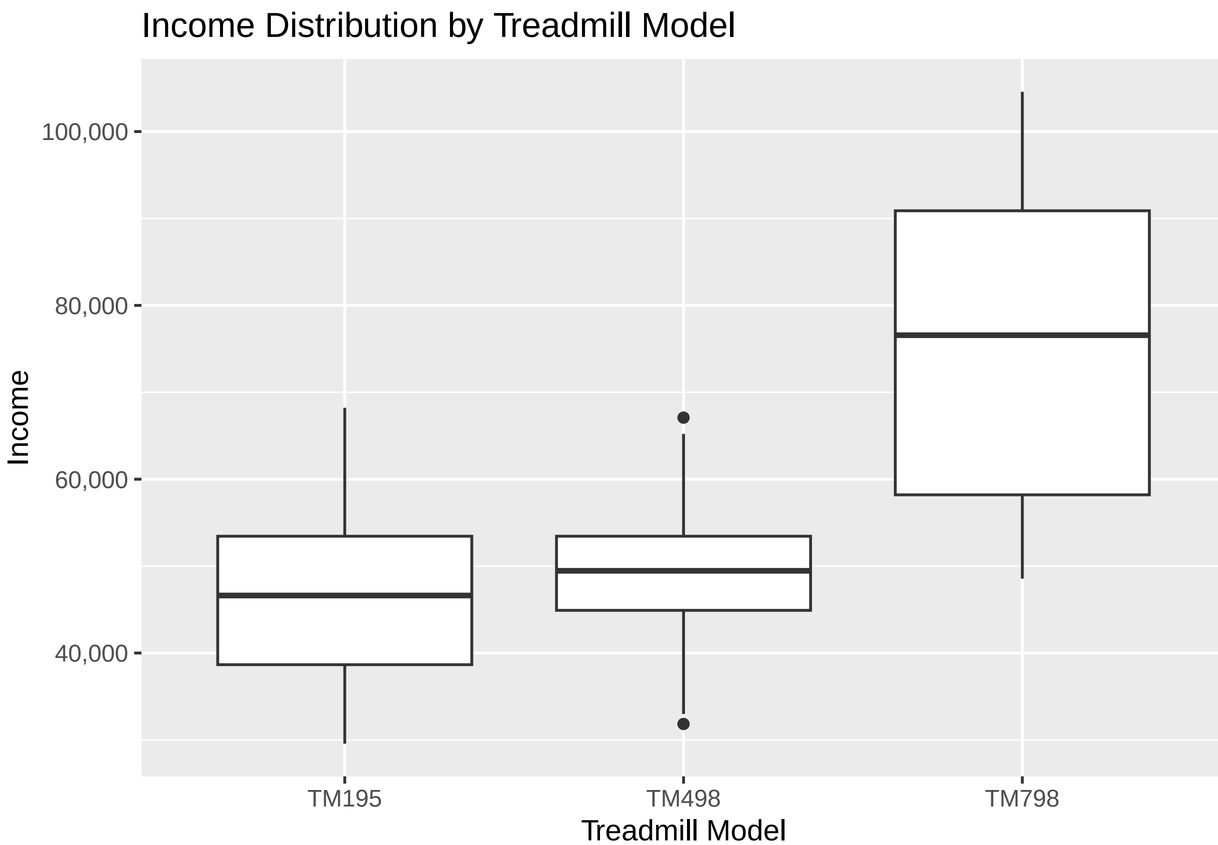
```
library(knitr)
anova_result = aov(Income ~ Product, data = data)
anova_summary = summary(anova_result)
anova_df = as.data.frame(anova_summary[[1]])
kable(anova_df, format = "markdown", caption = "ANOVA Table: Income ~ Product")
```

Table 12: ANOVA Table: Income ~ Product

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Product	2	24490250199	12245125099	89.25904	0
Residuals	177	24281991523	137186393	NA	NA

```
ggplot(data, aes(x = Product, y = Income)) +
  geom_boxplot() +
```

```r
  labs(title = "Income Distribution by Treadmill Model", x = "Treadmill Model", y = "Income") +
  scale_y_continuous(labels = scales::comma_format())
```



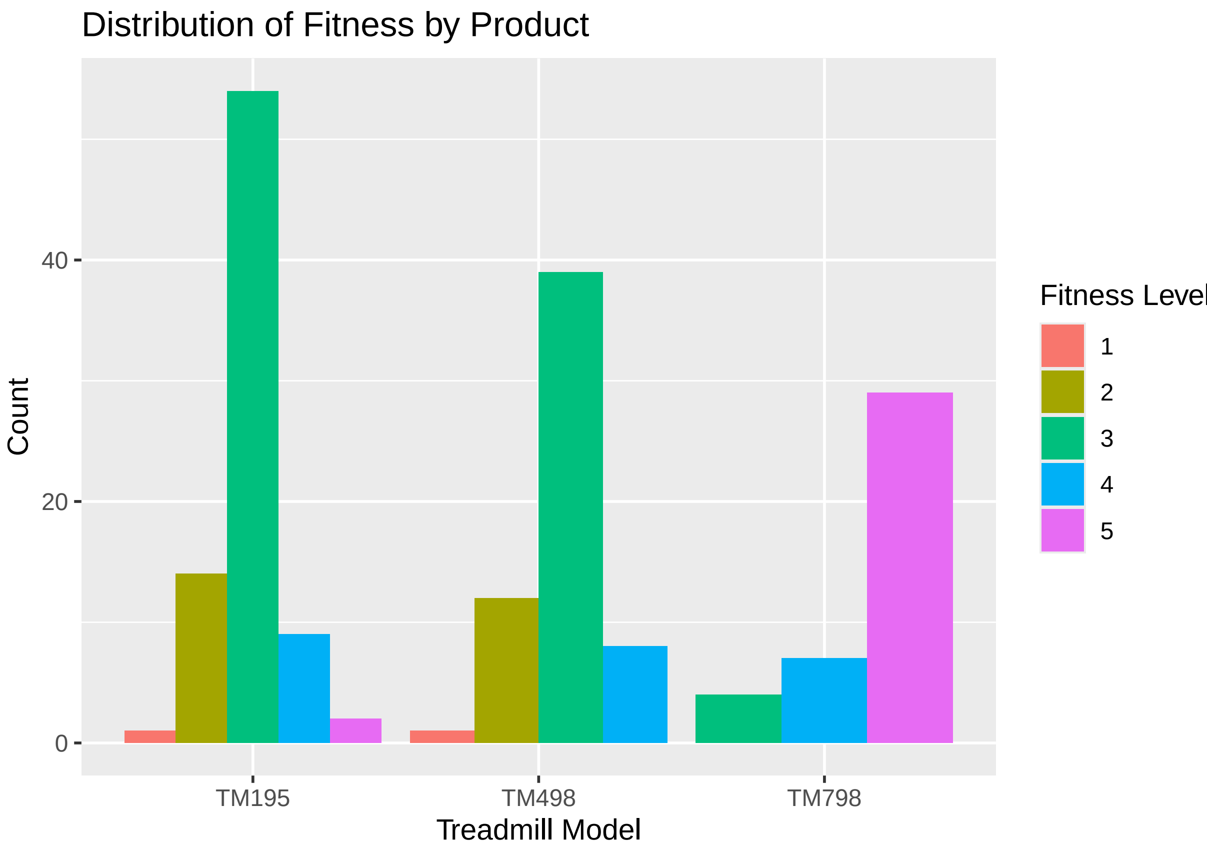
3. Fitness vs Model

```r
table_fitness_model = table(data$Fitness, data$Product)
chisq_test = chisq.test(table_fitness_model)
chi_result_df = data.frame(Statistic = chisq_test$statistic, DF = chisq_test$parameter, P_value = chisq_
kable(chi_result_df, caption = "Chi-Square Test Results: Fitness vs Product")
```

Table 13: Chi-Square Test Results: Fitness vs Product

	Statistic	DF	P_value
X-squared	118.7768	8	0

```r
ggplot(data, aes(x = Product, fill = factor(Fitness))) +
  geom_bar(position = "dodge") +
  labs(title = "Distribution of Fitness by Product", x = "Treadmill Model",y = "Count", fill = "Fitness
```



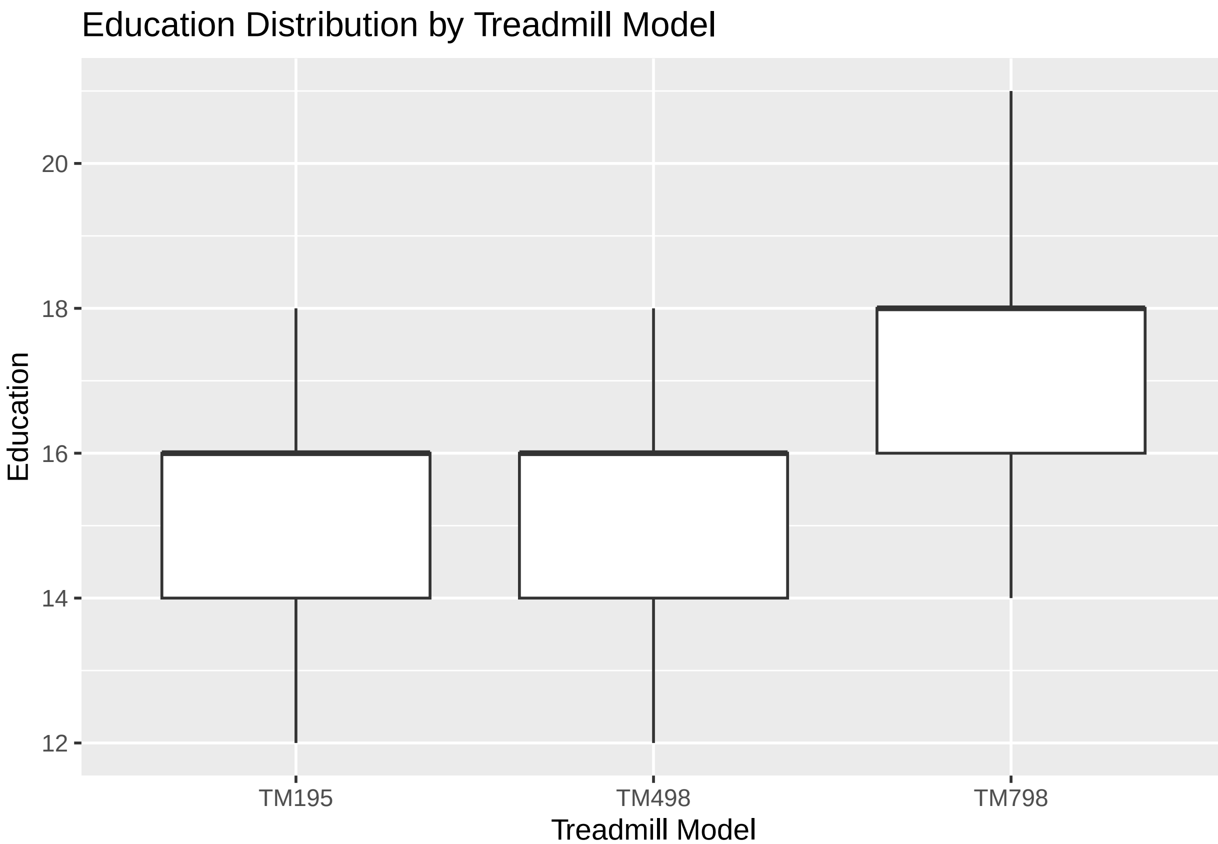
4. Education vs Model

```
anova_edu_model = aov(Education ~ Product, data = data)
anova_result = summary(anova_edu_model)
kable(anova_result[[1]], caption = "ANOVA Table: Education vs Product")
```

Table 14: ANOVA Table: Education vs Product

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Product	2	158.2153	79.107639	45.19038	0
Residuals	177	309.8458	1.750541	NA	NA

```
ggplot(data, aes(x = Product, y = Education)) +
  geom_boxplot() +
  labs(title = "Education Distribution by Treadmill Model", x = "Treadmill Model", y = "Education")
```



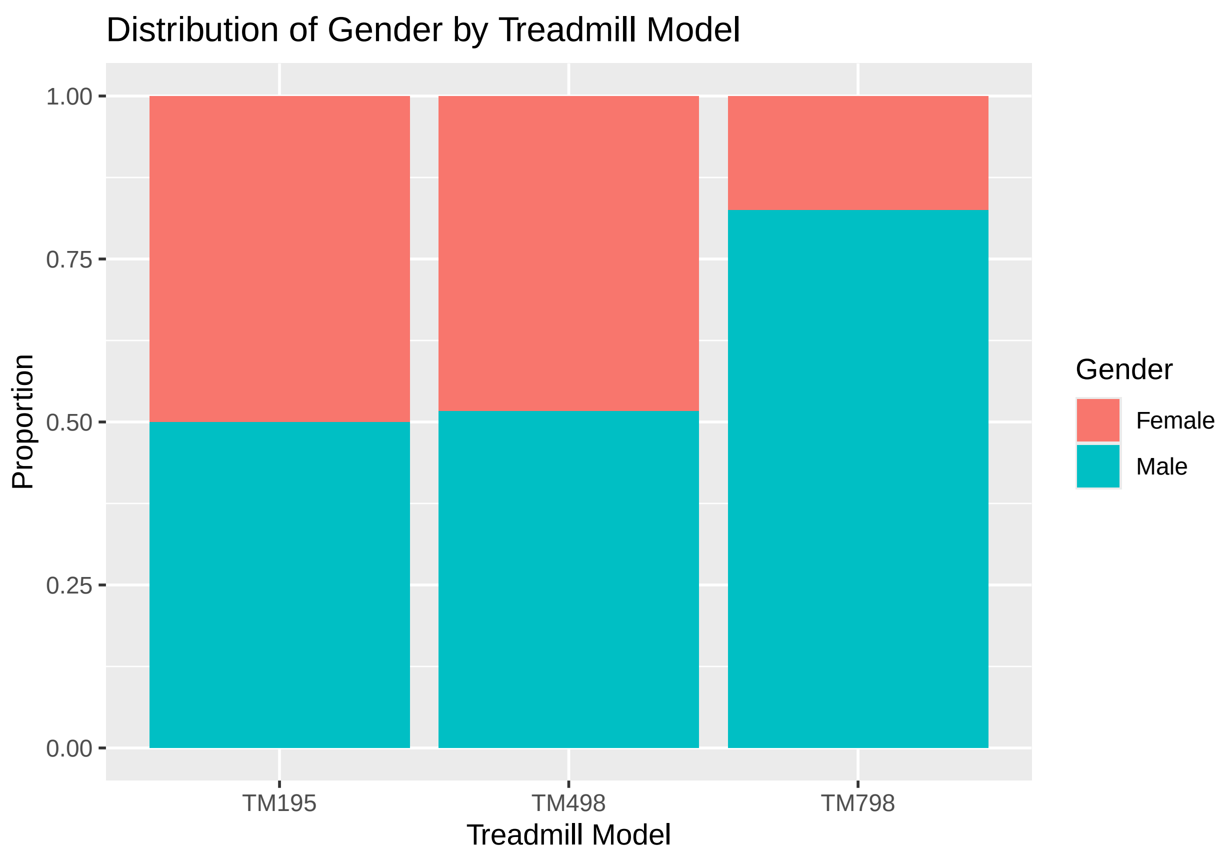
5. Gender vs Model

```
table_gender_model = table(data$Gender, data$Product)
chi_square_gender_model = chisq.test(table_gender_model)
chi_square_gender_df = data.frame(Statistic = chi_square_gender_model$statistic, DF = chi_square_gender_
kable(chi_square_gender_df, caption = "Chi-Square Test: Gender vs Treadmill Model")
```

Table 15: Chi-Square Test: Gender vs Treadmill Model

	Statistic	DF	P_value
X-squared	12.92384	2	0.0015618

```
ggplot(data, aes(x = Product, fill = Gender)) +
  geom_bar(position = "fill") +
  labs(title = "Distribution of Gender by Treadmill Model", y = "Proportion", x = "Treadmill Model")
```



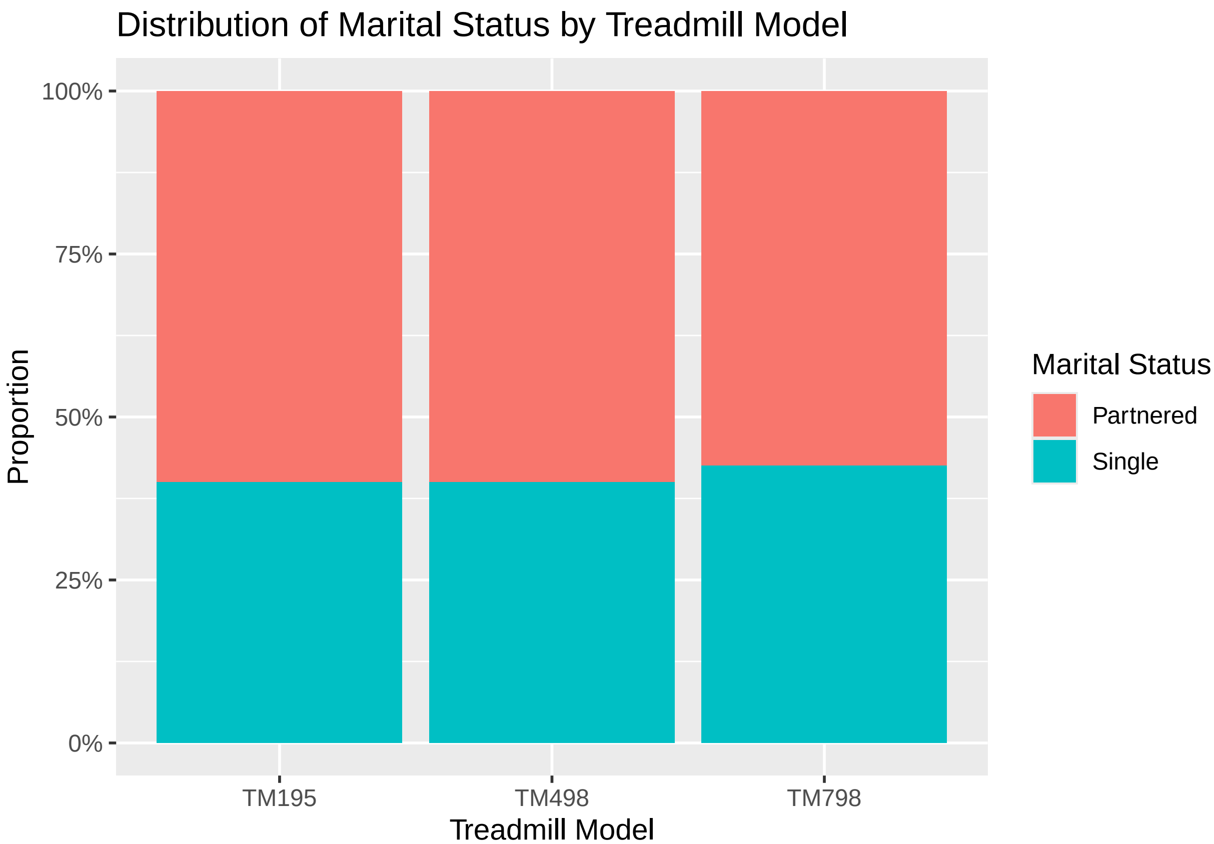
6. Gender vs Model

```
table_marital_model = table(data$MaritalStatus, data$Product)
chi_square_marital_model = chisq.test(table_marital_model)
chi_square_martial_df = data.frame(Statistic = chi_square_marital_model$statistic, DF = chi_square_mari
kable(chi_square_martial_df, caption = "Chi-Square Test: Martial Status vs Treadmill Model")
```

Table 16: Chi-Square Test: Martial Status vs Treadmill Model

	Statistic	DF	P_value
X-squared	0.0806555	2	0.9604746

```
ggplot(data, aes(x = Product, fill = MaritalStatus)) +
  geom_bar(position = "fill") +
  labs(title = "Distribution of Marital Status by Treadmill Model",
       x = "Treadmill Model",
       y = "Proportion",
       fill = "Marital Status") +
  scale_y_continuous(labels = scales::percent)
```



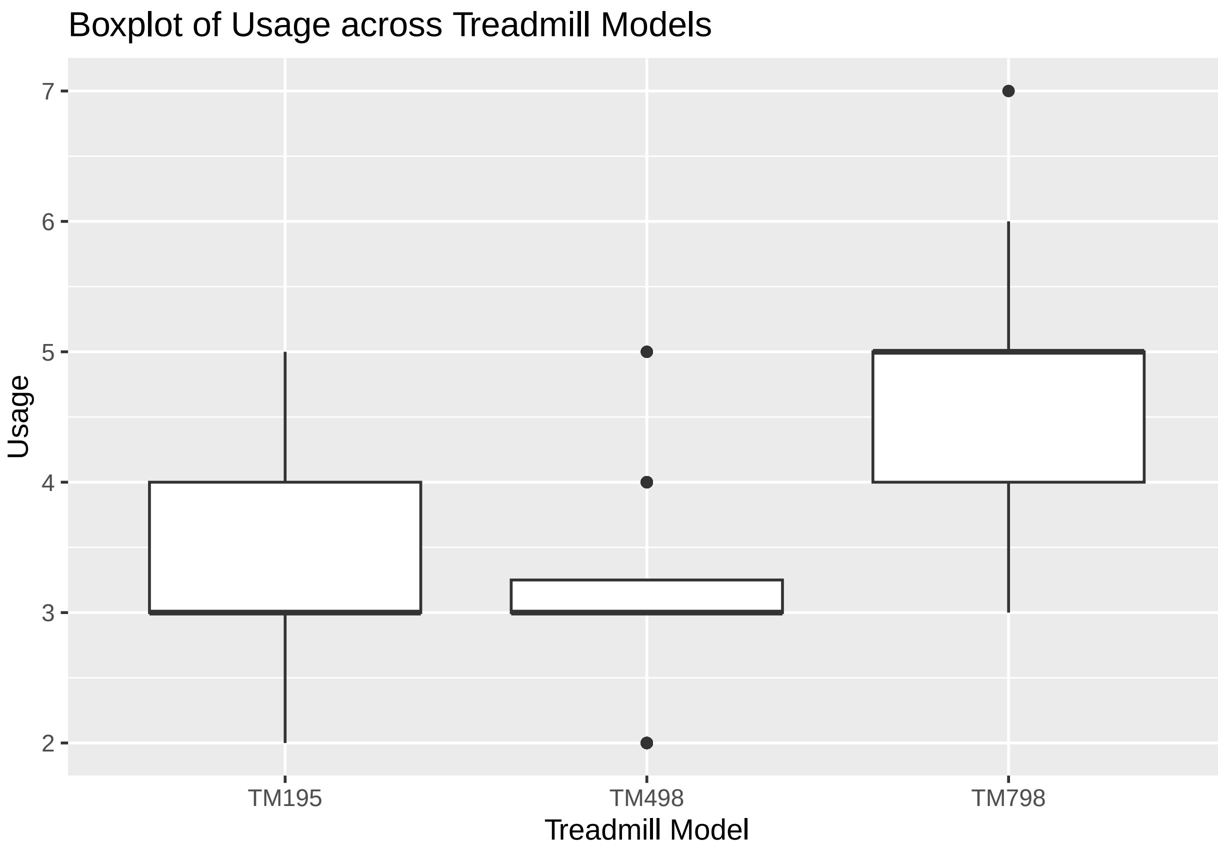
7. Usage vs Model

```
anova_usage_product = aov(Usage ~ Product, data = data)
result_usage = summary(anova_usage_product)
kable(result_usage[[1]], caption = "ANOVA Table: Usage vs Product")
```

Table 17: ANOVA Table: Usage vs Product

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Product	2	89.54861	44.7743056	65.44447	0
Residuals	177	121.09583	0.6841573	NA	NA

```
ggplot(data, aes(x = Product, y = Usage)) +
  geom_boxplot() +
  labs(title = "Boxplot of Usage across Treadmill Models", x = "Treadmill Model", y = "Usage")
```



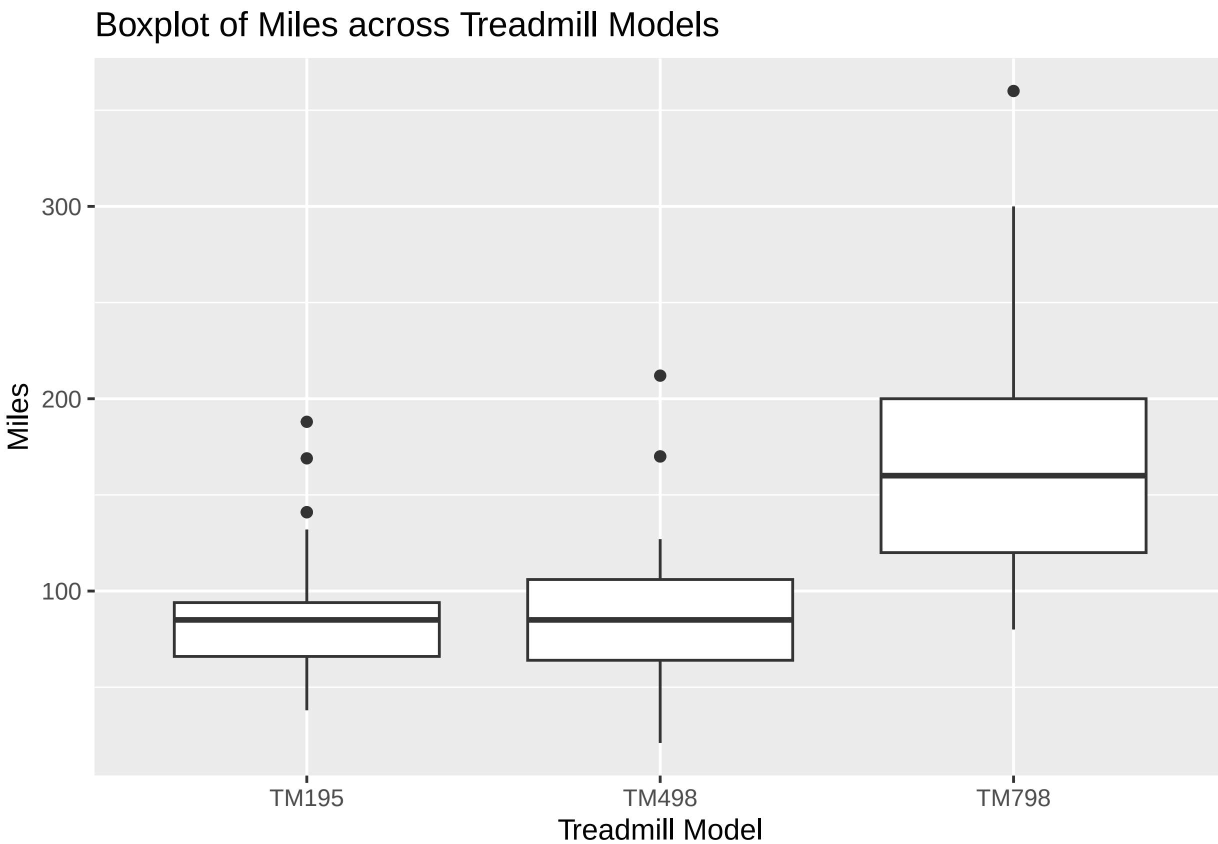
8. Miles vs Product

```
anova_miles_product = aov(Miles ~ Product, data = data)
result_miles = summary(anova_miles_product)
kable(result_miles[[1]], caption = "Miles vs Product")
```

Table 18: Miles vs Product

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Product	2	209625.5	104812.737	68.2418	0
Residuals	177	271854.7	1535.902	NA	NA

```
ggplot(data, aes(x = Product, y = Miles)) +
  geom_boxplot() +
  labs(title = "Boxplot of Miles across Treadmill Models", x = "Treadmill Model", y = "Miles")
```



Multivariate Analysis

```
ggplot(data, aes(x = Age, y = Income, color = Product)) +
  geom_point() +
  labs(title = "Income vs Age by Treadmill Model", x = "Age", y = "Income") +
  scale_y_continuous(labels = label_comma())
```


```r
multi_reg_model <- lm(Income ~ Age + Fitness + Education + Product, data = data)
summary(multi_reg_model)
```

```
##
## Call:
## lm(formula = Income ~ Age + Fitness + Education + Product, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -24019  -4166   -599   4004  25969
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -6262.16    7690.21  -0.814  0.41658
## Age             1089.11      93.97  11.590  < 2e-16 ***
## Fitness           31.45     948.25   0.033  0.97358
## Education       1429.29     494.83   2.888  0.00436 **
## ProductTM498    2063.25    1405.64   1.468  0.14395
## ProductTM798   25102.74    2537.39   9.893  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8219 on 174 degrees of freedom
## Multiple R-squared:  0.759,  Adjusted R-squared:  0.752
## F-statistic: 109.6 on 5 and 174 DF,  p-value: < 2.2e-16
```

```r
two_way_anova <- aov( Age * Fitness ~  Product, data = data)
summary(two_way_anova)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## Product       2  81600   40800   41.16 2.09e-15 ***
## Residuals   177 175449     991                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```