```r
# Load necessary packages
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.2     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.2     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
come errors
```

```r
# Read the dataset
telco <- read.csv("telco_churn.csv")
```

```r
# Print the first six rows of the telco dataset
head(telco)
```

```
##   customerID gender SeniorCitizen Partner Dependents tenure PhoneService
## 1 7590-VHVEG Female             0     Yes         No      1           No
## 2 5575-GNVDE   Male             0      No         No     34          Yes
## 3 3668-QPYBK   Male             0      No         No      2          Yes
## 4 7795-CFOCW   Male             0      No         No     45           No
## 5 9237-HQITU Female             0      No         No      2          Yes
## 6 9305-CDSKC Female             0      No         No      8          Yes
##      MultipleLines InternetService OnlineSecurity OnlineBackup DeviceProtection
## 1 No phone service             DSL             No          Yes               No
## 2               No             DSL            Yes           No              Yes
## 3               No             DSL            Yes          Yes               No
## 4 No phone service             DSL            Yes           No              Yes
## 5               No     Fiber optic             No           No               No
## 6              Yes     Fiber optic             No           No              Yes
##   TechSupport StreamingTV StreamingMovies       Contract PaperlessBilling
## 1          No          No              No Month-to-month              Yes
## 2          No          No              No       One year               No
## 3          No          No              No Month-to-month              Yes
## 4         Yes          No              No       One year               No
## 5          No          No              No Month-to-month              Yes
## 6          No         Yes             Yes Month-to-month              Yes
##               PaymentMethod MonthlyCharges TotalCharges Churn
## 1          Electronic check          29.85        29.85    No
## 2              Mailed check          56.95      1889.50    No
## 3              Mailed check          53.85       108.15   Yes
## 4 Bank transfer (automatic)          42.30      1840.75    No
## 5          Electronic check          70.70       151.65   Yes
## 6          Electronic check          99.65       820.50   Yes
```

```
# Display summary statistics for the telco dataset, including the mean, median, standard deviation, and range for each variable.
summary(telco)
```

```
##   customerID          gender          SeniorCitizen      Partner
##  Length:7043        Length:7043        Min.   :0.0000   Length:7043
##  Class :character   Class :character   1st Qu.:0.0000   Class :character
##  Mode  :character   Mode  :character   Median :0.0000   Mode  :character
##                                        Mean   :0.1621
##                                        3rd Qu.:0.0000
##                                        Max.   :1.0000
##
##   Dependents            tenure      PhoneService       MultipleLines
##  Length:7043        Min.   : 0.00   Length:7043        Length:7043
##  Class :character   1st Qu.: 9.00   Class :character   Class :character
##  Mode  :character   Median :29.00   Mode  :character   Mode  :character
##                     Mean   :32.37
##                     3rd Qu.:55.00
##                     Max.   :72.00
##
##  InternetService    OnlineSecurity     OnlineBackup       DeviceProtection
##  Length:7043        Length:7043        Length:7043        Length:7043
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  TechSupport        StreamingTV        StreamingMovies      Contract
##  Length:7043        Length:7043        Length:7043        Length:7043
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  PaperlessBilling   PaymentMethod      MonthlyCharges    TotalCharges
##  Length:7043        Length:7043        Min.   : 18.25   Min.   :  18.8
##  Class :character   Class :character   1st Qu.: 35.50   1st Qu.: 401.4
##  Mode  :character   Mode  :character   Median : 70.35   Median :1397.5
##                                        Mean   : 64.76   Mean   :2283.3
##                                        3rd Qu.: 89.85   3rd Qu.:3794.7
##                                        Max.   :118.75   Max.   :8684.8
##                                                         NA's   :11
##    Churn
##  Length:7043
##  Class :character
##  Mode  :character
##
##
##
##
```

```
# Print the structure of the telco dataset, including the names of the variables, the number of rows and columns, the data type of each variable, and the first few values of each variable.
str(telco)
```

```
## 'data.frame':    7043 obs. of  21 variables:
##  $ customerID      : chr  "7590-VHVEG" "5575-GNVDE" "3668-QPYBK" "7795-CFOCW" ...
##  $ gender          : chr  "Female" "Male" "Male" "Male" ...
##  $ SeniorCitizen   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Partner         : chr  "Yes" "No" "No" "No" ...
##  $ Dependents      : chr  "No" "No" "No" "No" ...
##  $ tenure          : int  1 34 2 45 2 8 22 10 28 62 ...
##  $ PhoneService    : chr  "No" "Yes" "Yes" "No" ...
##  $ MultipleLines   : chr  "No phone service" "No" "No" "No phone service" ...
##  $ InternetService : chr  "DSL" "DSL" "DSL" "DSL" ...
##  $ OnlineSecurity  : chr  "No" "Yes" "Yes" "Yes" ...
##  $ OnlineBackup    : chr  "Yes" "No" "Yes" "No" ...
##  $ DeviceProtection: chr  "No" "Yes" "No" "Yes" ...
##  $ TechSupport     : chr  "No" "No" "No" "Yes" ...
##  $ StreamingTV     : chr  "No" "No" "No" "No" ...
##  $ StreamingMovies : chr  "No" "No" "No" "No" ...
##  $ Contract        : chr  "Month-to-month" "One year" "Month-to-month" "One year" ...
##  $ PaperlessBilling: chr  "Yes" "No" "Yes" "No" ...
##  $ PaymentMethod   : chr  "Electronic check" "Mailed check" "Mailed check" "Bank transfer
(automatic)" ...
##  $ MonthlyCharges  : num  29.9 57 53.9 42.3 70.7 ...
##  $ TotalCharges    : num  29.9 1889.5 108.2 1840.8 151.7 ...
##  $ Churn           : chr  "No" "No" "Yes" "No" ...
```

```
# Create a table of the Churn variable in the telco dataset, showing the number of observations in each category.
table(telco$Churn)
```

```
##
##   No  Yes
## 5174 1869
```

```
# Handle missing values
telco <- telco %>% na.omit()  # Remove rows with missing values
```

```
# Convert categorical variables
telco$Churn <- as.factor(telco$Churn)
```

```
# Perform feature engineering
telco <- telco %>% mutate(TotalCharges = ifelse(TotalCharges == " ", 0, as.numeric(TotalCharges)))
```

```
library(caret) #load the caret library
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
# Split the dataset
set.seed(123)  # For reproducibility
train_index <- caret::createDataPartition(telco$Churn, p = 0.8, list = FALSE)
train_data <- telco[train_index, ]
test_data <- telco[-train_index, ]
```

```
# Select an algorithm
library(randomForest)
```

```
## randomForest 4.7-1.1
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

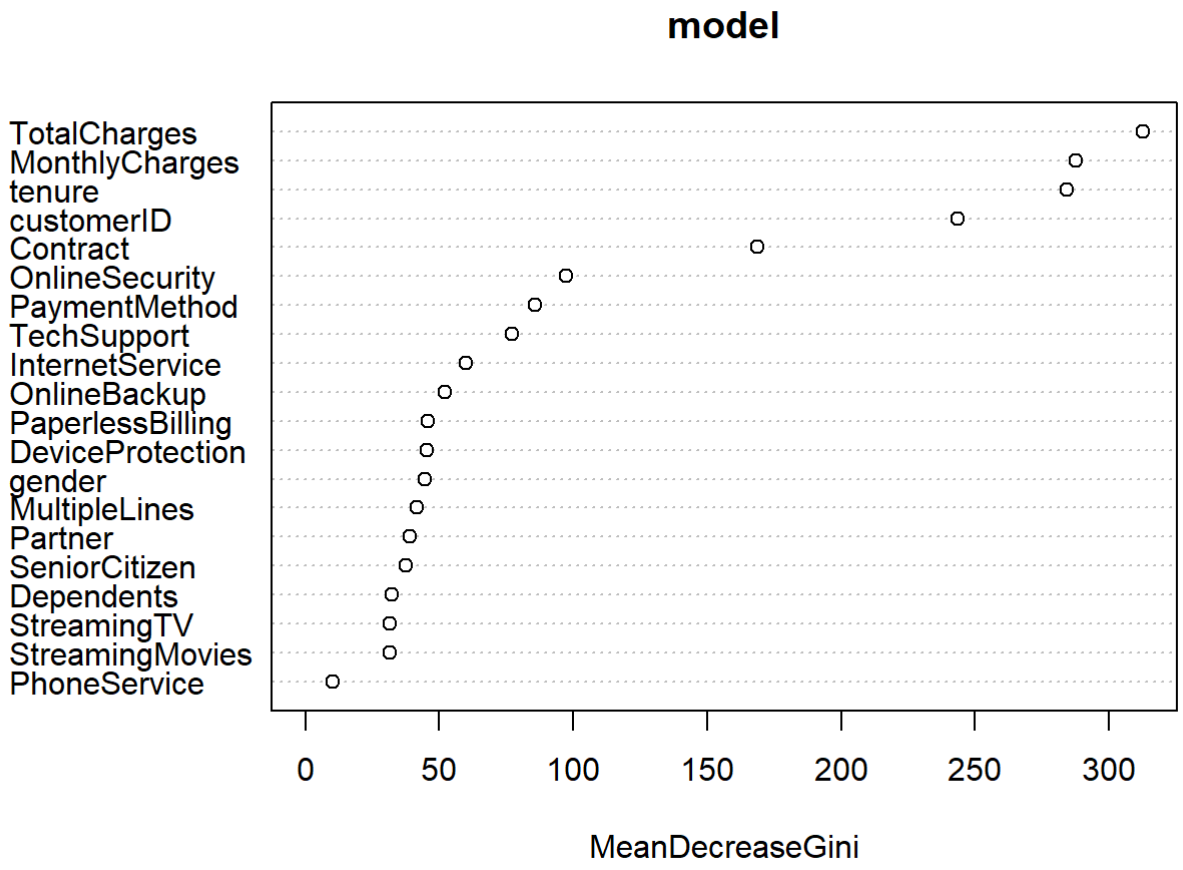
```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
# Train the model
model <- randomForest(Churn ~ ., data = train_data)
```

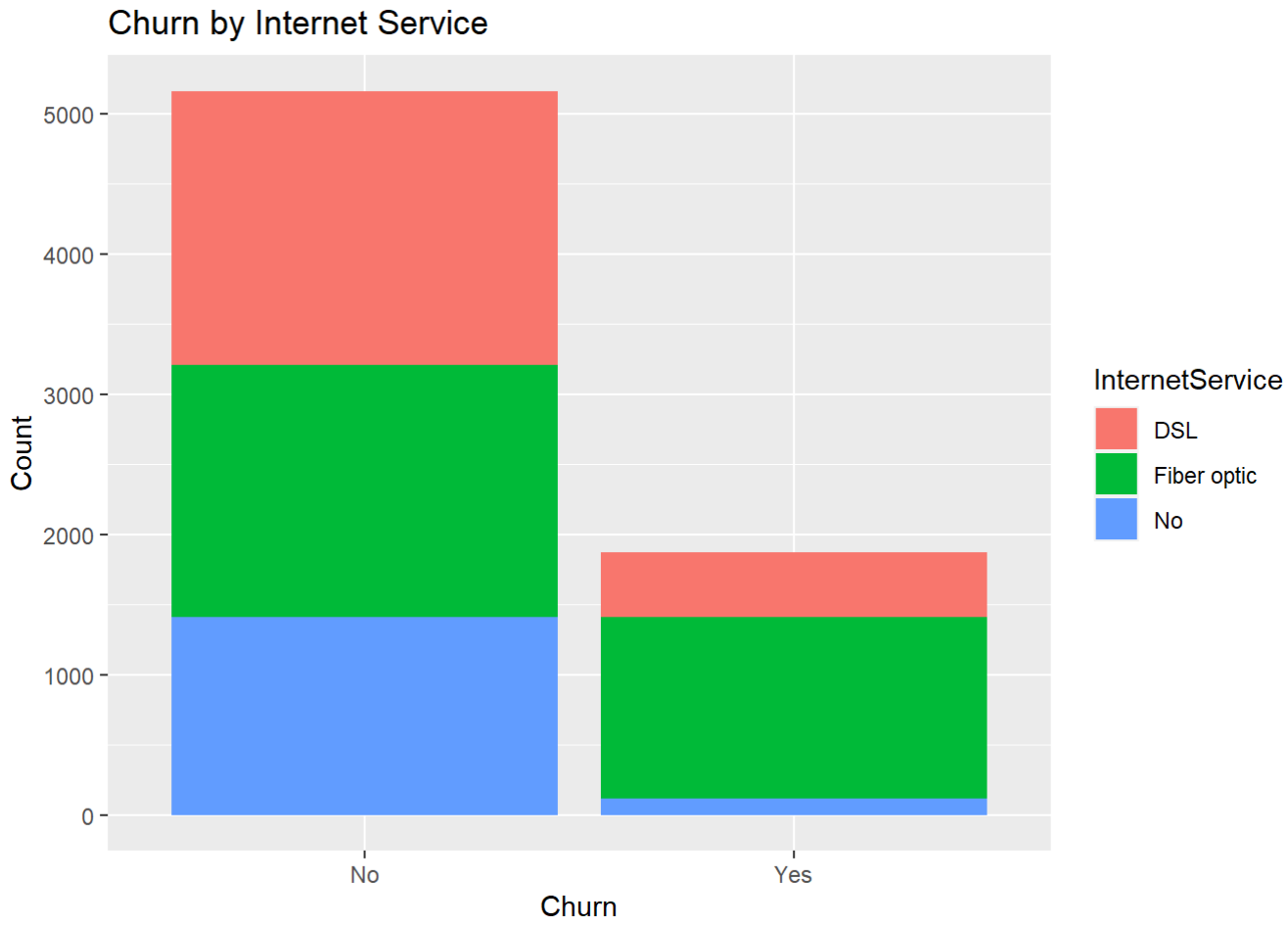
```
# Evaluate the model
predictions <- predict(model, newdata = test_data)
confusionMatrix(predictions, test_data$Churn)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No Yes
##        No  928 180
##        Yes 104 193
##
##                Accuracy : 0.7979
##                  95% CI : (0.7759, 0.8186)
##     No Information Rate : 0.7345
##     P-Value [Acc > NIR] : 1.935e-08
##
##                   Kappa : 0.4456
##
##  Mcnemar's Test P-Value : 8.570e-06
##
##             Sensitivity : 0.8992
##             Specificity : 0.5174
##          Pos Pred Value : 0.8375
##          Neg Pred Value : 0.6498
##              Prevalence : 0.7345
##          Detection Rate : 0.6605
##    Detection Prevalence : 0.7886
##       Balanced Accuracy : 0.7083
##
##        'Positive' Class : No
##
```

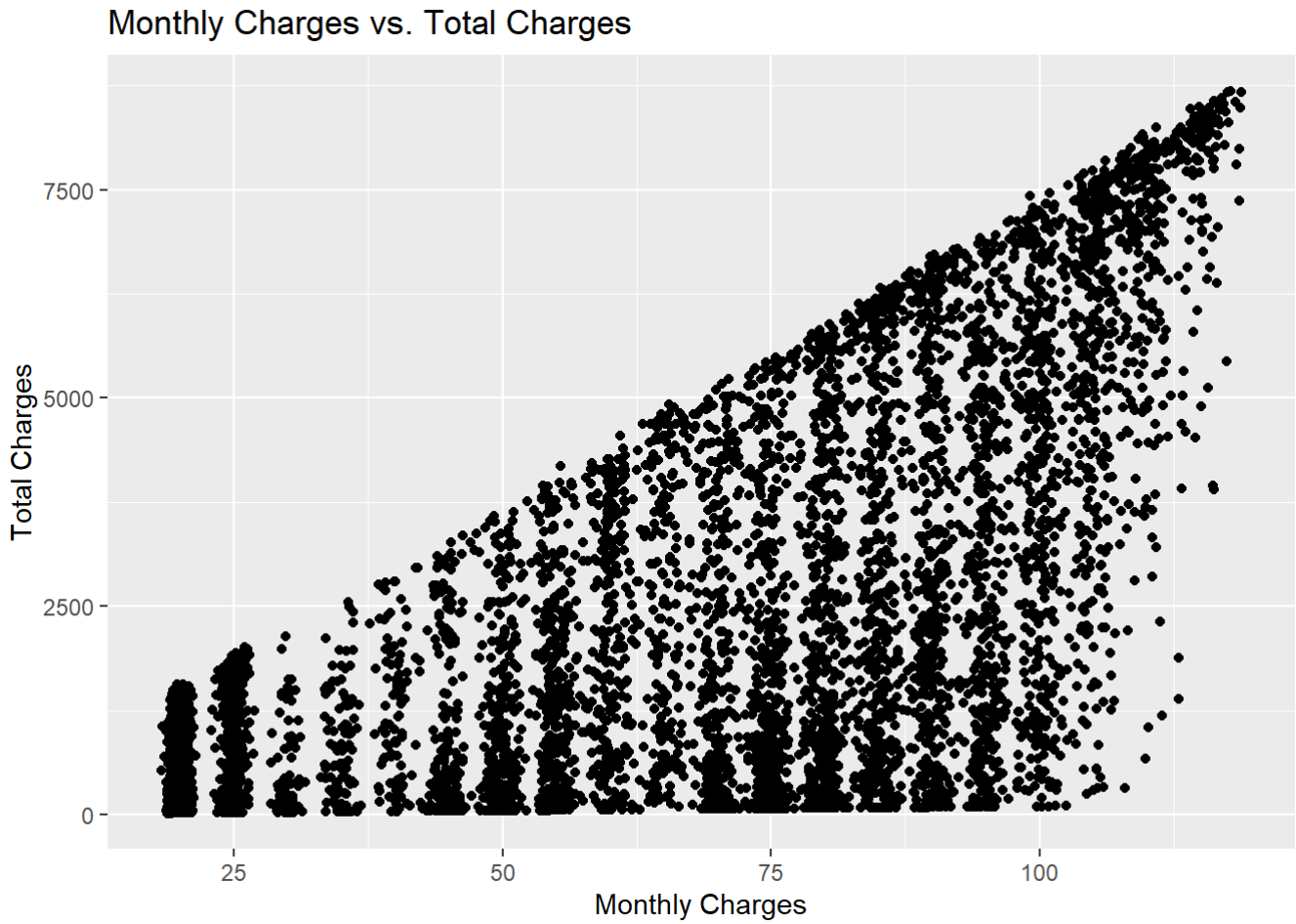
```
# Analyze the predictions and interpret the results
varImpPlot(model)
```



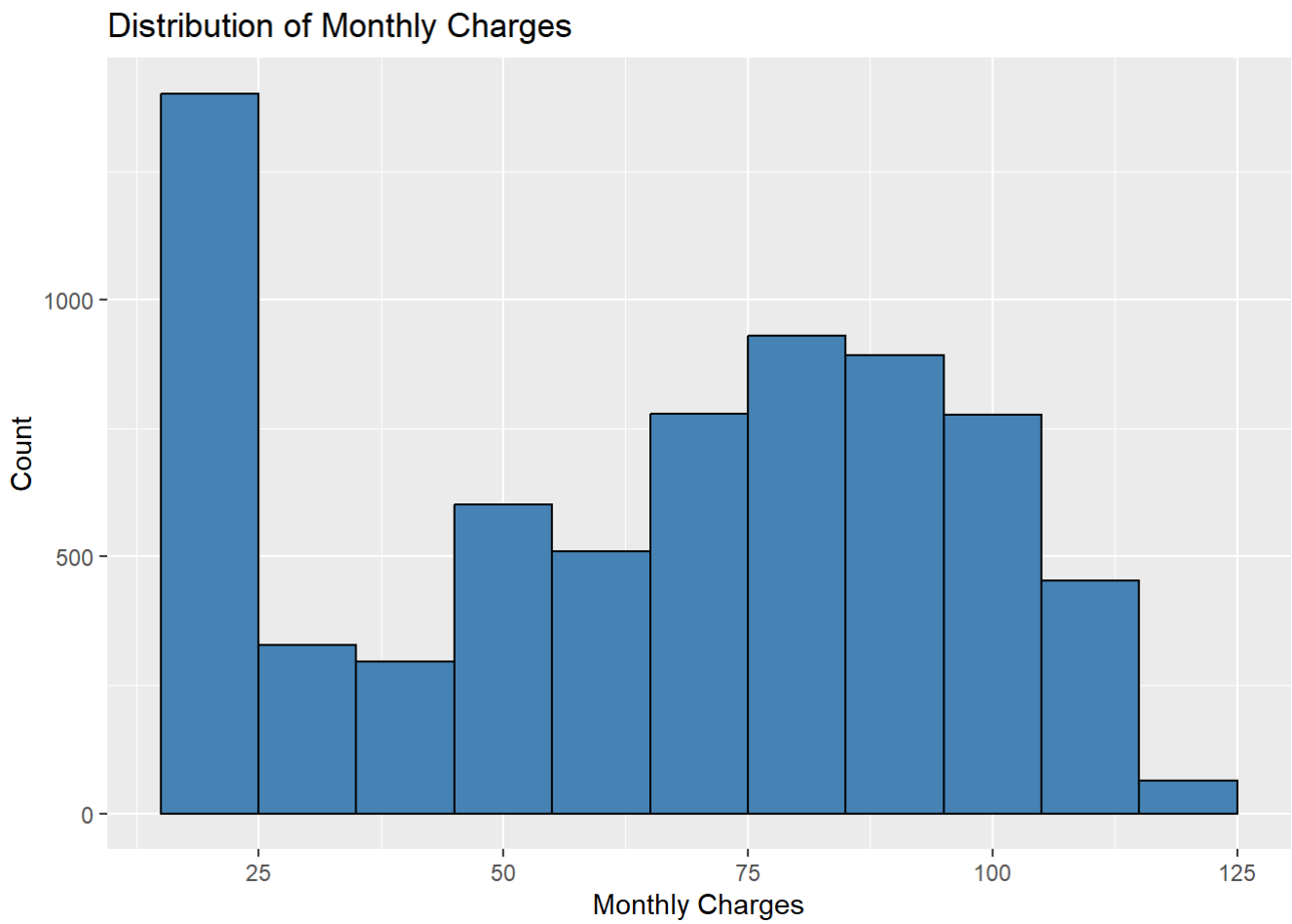

```
# Summarize and visualize the results
library(ggplot2)
ggplot(telco, aes(x = Churn, fill = InternetService)) +
  geom_bar() +
  labs(title = "Churn by Internet Service", x = "Churn", y = "Count")
```



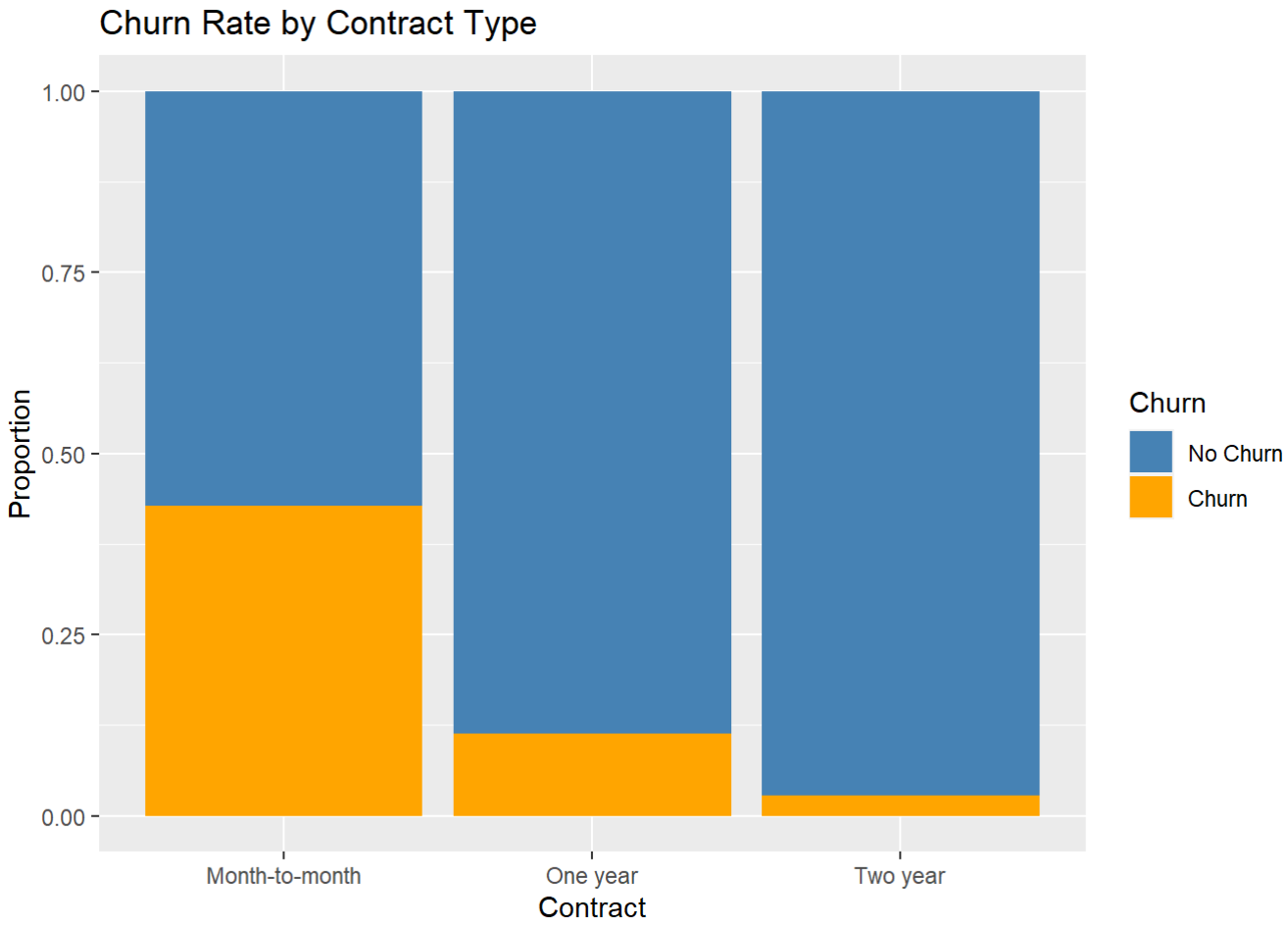

```
# Scatter plot of Monthly Charges vs. Total Charges
ggplot(data = telco, aes(x = MonthlyCharges, y = TotalCharges)) +
  geom_point() +
  labs(title = "Monthly Charges vs. Total Charges", x = "Monthly Charges", y = "Total Charge
s")
```



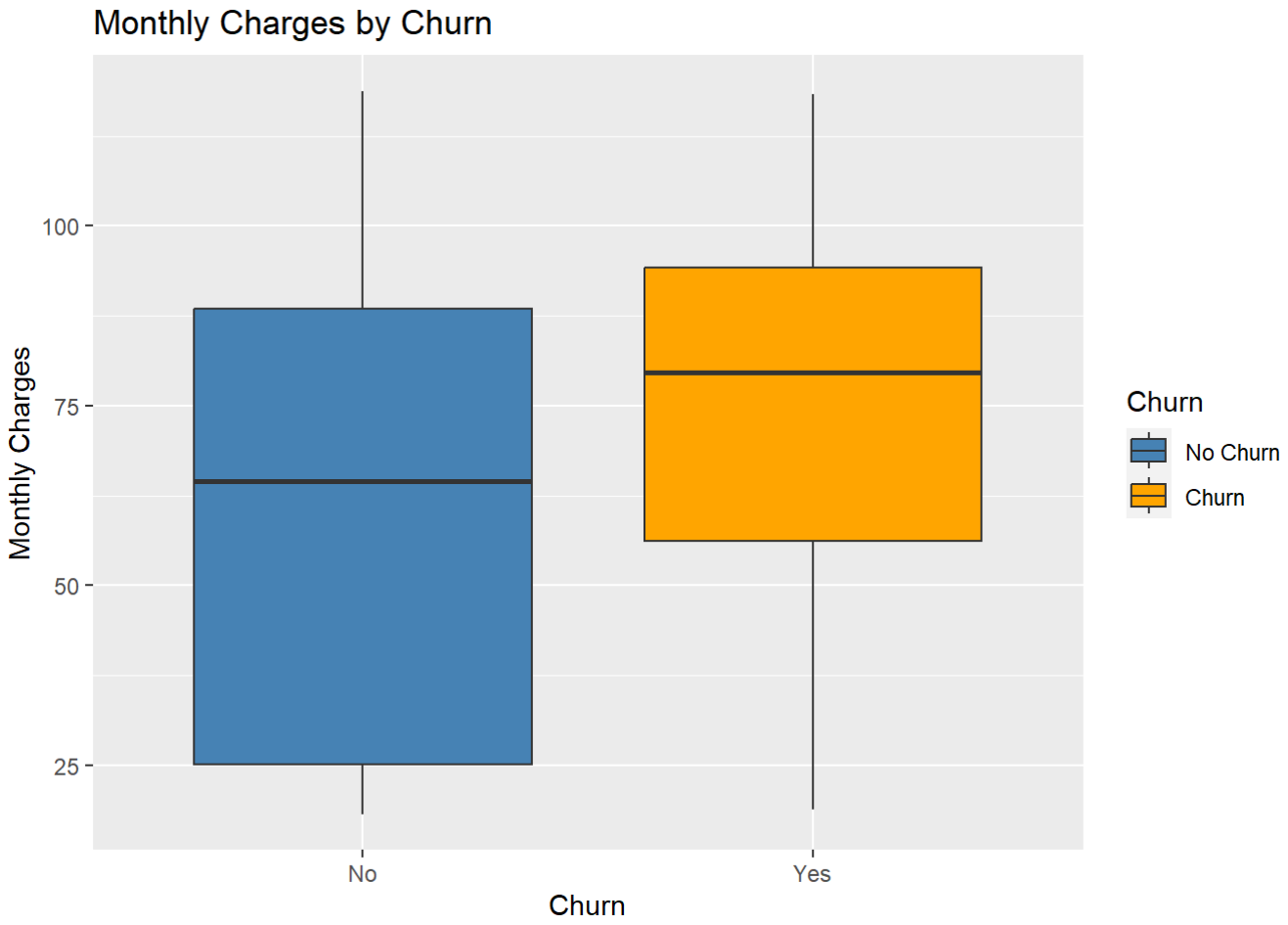

```
# Histogram of Monthly Charges
ggplot(data = telco, aes(x = MonthlyCharges)) +
  geom_histogram(binwidth = 10, fill = "steelblue", color = "black") +
  labs(title = "Distribution of Monthly Charges", x = "Monthly Charges", y = "Count")
```




```
# Bar plot of Churn Rate by Contract Type
ggplot(data = telco, aes(x = Contract, fill = Churn)) +
  geom_bar(position = "fill") +
  labs(title = "Churn Rate by Contract Type", x = "Contract", y = "Proportion") +
  scale_fill_manual(values = c("steelblue", "orange"), labels = c("No Churn", "Churn"))
```



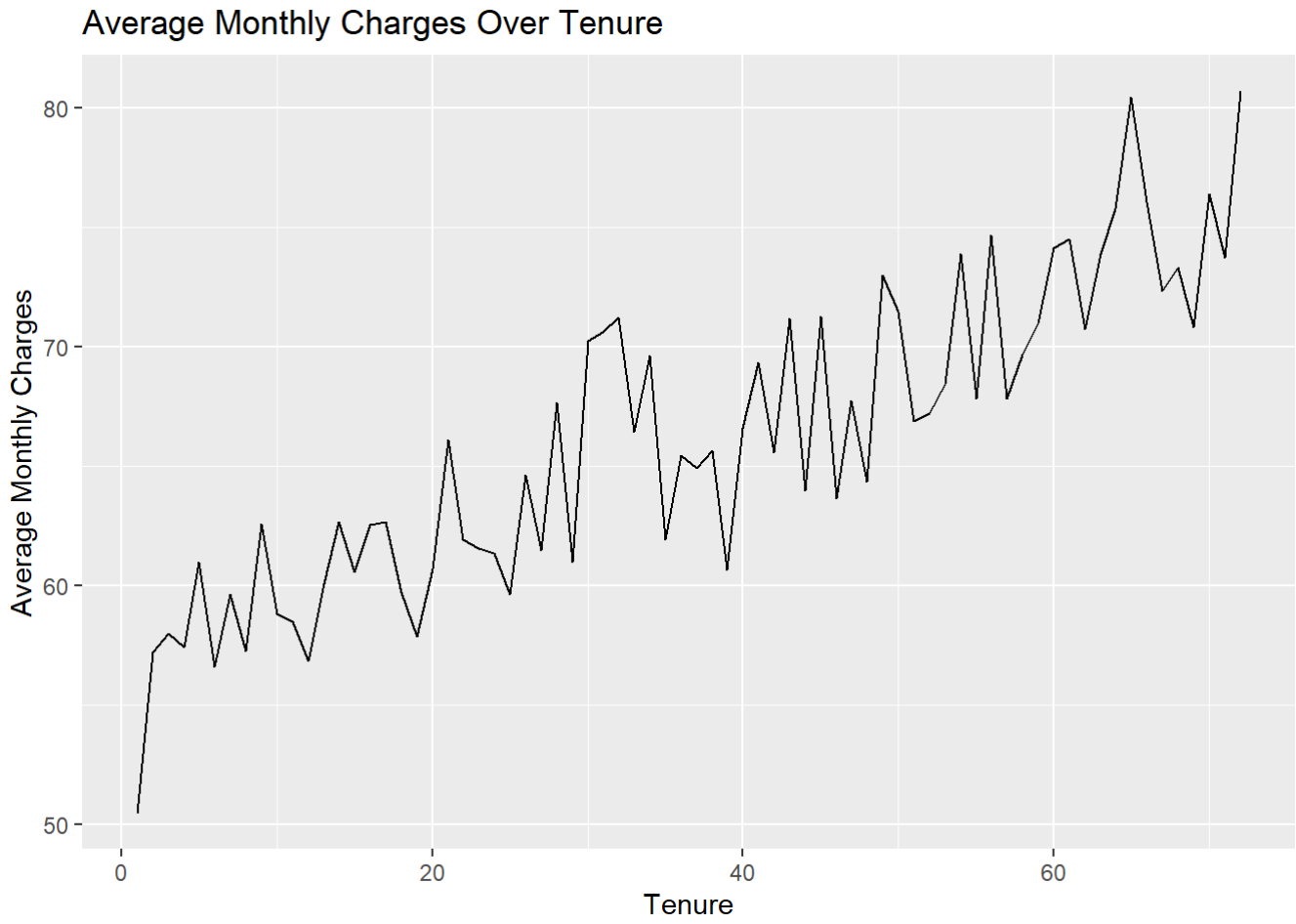

```
# Box plot of Monthly Charges by Churn
ggplot(data = telco, aes(x = Churn, y = MonthlyCharges, fill = Churn)) +
  geom_boxplot() +
  labs(title = "Monthly Charges by Churn", x = "Churn", y = "Monthly Charges") +
  scale_fill_manual(values = c("steelblue", "orange"), labels = c("No Churn", "Churn"))
```




```
# Line plot of Average Monthly Charges Over Tenure
library(dplyr)

telco_summary <- telco %>%
  group_by(tenure) %>%
  summarise(avg_charges = mean(MonthlyCharges))

ggplot(data = telco_summary, aes(x = tenure, y = avg_charges)) +
  geom_line() +
  labs(title = "Average Monthly Charges Over Tenure", x = "Tenure", y = "Average Monthly Char
ges")
```


Average Monthly Charges Over Tenure
Average Monthly Charges
80
70
60
50
0
20
40
60
Tenure