```
R version 4.4.3 (2025-02-28 ucrt) -- "Trophy Case"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> > data <- read.csv ("C:/Users/sray8/OneDrive/Desktop/MIS581/hrdata.csv")
Error: unexpected '>' in ">"
> > head(data)
Error: unexpected '>' in ">"
>
> data <- read.csv ("C:/Users/sray8/OneDrive/Desktop/MIS581/hrdata.csv")
> data$Attrition <- factor(data$Attrition, levels = c("No", "Yes"))
>  model_rq1 <- glm(Attrition ~ Age + Gender + YearsAtCompany + OverTime + MonthlyIncome,
+                  data = data, family = binomial)
> summary(model_rq1)

Call:
glm(formula = Attrition ~ Age + Gender + YearsAtCompany + OverTime +
    MonthlyIncome, family = binomial, data = data)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -4.416e-01  3.342e-01  -1.321 0.186427
Age            -3.389e-02  9.885e-03  -3.428 0.000608 ***
GenderMale      2.091e-01  1.561e-01   1.339 0.180576
YearsAtCompany -4.000e-02  1.841e-02  -2.173 0.029816 *
OverTimeYes     1.421e+00  1.525e-01   9.319  < 2e-16 ***
MonthlyIncome  -7.494e-05  2.606e-05  -2.876 0.004031 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1298.6  on 1469  degrees of freedom
Residual deviance: 1146.9  on 1464  degrees of freedom
AIC: 1158.9

Number of Fisher Scoring iterations: 5

> # Load the ggplot2 library
> library(ggplot2)
>
> # Create a bar chart showing attrition count by OverTime
> ggplot(data, aes(x = OverTime, fill = Attrition)) +
+   geom_bar(position = "fill") +  # Shows proportion within each OverTime group
+   labs(
+     title = "Attrition Rate by Overtime Status",
+     x = "Overtime",
+     y = "Proportion of Employees",
+     fill = "Attrition"
+   ) +
+   scale_y_continuous(labels = scales::percent) +
+   theme_minimal()
> ggplot(data, aes(x = OverTime, fill = Attrition)) +
+   geom_bar(position = "fill", width = 0.5) +  # Set width to a smaller value (default is 0.9)
+   labs(
```

```
+     title = "Attrition Rate by Overtime Status",
+     x = "Overtime",
+     y = "Proportion of Employees",
+     fill = "Attrition"
+   ) +
+   scale_y_continuous(labels = scales::percent) +
+   theme_minimal()
> ggplot(data, aes(x = OverTime, fill = Attrition)) +
+   geom_bar(position = "fill", width = 0.5) +  # Set width to a smaller value (default is 0.9)
+   labs(
+     title = "Attrition Rate by Overtime Status",
+     x = "Overtime",
+     y = "Proportion of Employees",
+     fill = "Attrition"
+   ) +
+   scale_y_continuous(labels = scales::percent) +
+   theme_minimal()
> library(ggplot2)
> library(dplyr)

Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’:

    filter, lag

The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union

>
> # Create a summary data frame: attrition rate by JobRole
> attrition_summary <- data %>%
+   group_by(JobRole, Attrition) %>%
+   summarise(Count = n()) %>%
+   mutate(Percent = Count / sum(Count))
`summarise()` has grouped output by 'JobRole'. You can override using the `.groups` argument.
>
> # Plot: grouped bar chart
> ggplot(attrition_summary, aes(x = JobRole, y = Percent, fill = Attrition)) +
+   geom_bar(stat = "identity", position = "dodge") +
+   labs(title = "Attrition Rate by Job Role",
+        x = "Job Role",
+        y = "Percentage",
+        fill = "Attrition Status") +
+   scale_y_continuous(labels = scales::percent) +
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1))
> library(ggplot2)
>
> ggplot(data, aes(x = JobRole, fill = Attrition)) +
+   geom_bar(position = "fill") +
+   labs(title = "Attrition Rate by Job Role", x = "Job Role", y = "Proportion") +
+   scale_fill_manual(values = c("No" = "gray", "Yes" = "purple")) +
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1))
> ggplot(attrition_summary, aes(x = JobRole, y = Percent, fill = Attrition)) +
+ +   geom_bar(stat = "identity", position = "dodge") +
+ +   labs(title = "Attrition Rate by Job Role",
+ +        x = "Job Role",
Error: unexpected '=' in:
"+   labs(title = "Attrition Rate by Job Role",
+        x ="
> +        y = "Percentage",
Error: unexpected ',' in "+        y = "Percentage","
> +        fill = "Attrition Status") +
Error: unexpected ')' in "+        fill = "Attrition Status")"
> +   scale_y_continuous(labels = scales::percent) +
+ +   theme_minimal() +
+ +   theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
Error in `+.gg`:
! Cannot use `+` with a single argument.
ⓘ Did you accidentally put `+` on a new line?
Run `rlang::last_trace()` to see where the error occurred.
> library(ggplot2)
> library(dplyr)
>
> # Create a summary data frame: attrition rate by JobRole
> attrition_summary <- data %>%
+   group_by(JobRole, Attrition) %>%
+   summarise(Count = n()) %>%
+   mutate(Percent = Count / sum(Count))
`summarise()` has grouped output by 'JobRole'. You can override using the `.groups` argument.
>
> # Plot: grouped bar chart
> ggplot(attrition_summary, aes(x = JobRole, y = Percent, fill = Attrition)) +
+   geom_bar(stat = "identity", position = "dodge") +
+   labs(title = "Attrition Rate by Job Role",
+        x = "Job Role",
+        y = "Percentage",
+        fill = "Attrition Status") +
+   scale_y_continuous(labels = scales::percent) +
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1))
> library(ggplot2)
> library(dplyr)
>
> # Create summary table of attrition rates by job satisfaction
> satisfaction_summary <- data %>%
+   group_by(JobSatisfaction) %>%
+   summarise(
+     Count = n(),
+     Attrition_Rate = mean(Attrition == "Yes")
+   )
>
> # Plot
> ggplot(satisfaction_summary, aes(x = factor(JobSatisfaction), y = Attrition_Rate)) +
+   geom_bar(stat = "identity", fill = "purple") +
+   labs(title = "Attrition Rate by Job Satisfaction",
+        x = "Job Satisfaction (1 = Low, 4 = High)",
+        y = "Attrition Rate") +
+   theme_minimal()
> library(ggplot2)
> library(dplyr)
>
> # Create summary table of attrition rates by job satisfaction
> satisfaction_summary <- data %>%
+   group_by(JobSatisfaction) %>%
+   summarise(
+     Count = n(),
+     Attrition_Rate = mean(Attrition == "Yes")
+   )
>
> # Plot
> ggplot(satisfaction_summary, aes(x = factor(JobSatisfaction), y = Attrition_Rate)) +
+   geom_bar(stat = "identity") +
+   labs(title = "Attrition Rate by Job Satisfaction",
+        x = "Job Satisfaction (1 = Low, 4 = High)",
+        y = "Attrition Rate") +
+   theme_minimal()
> library(ggplot2)
> library(dplyr)
>
> # Create summarized data
> satisfaction_summary <- data %>%
+   group_by(JobSatisfaction, Attrition) %>%
+   summarise(Count = n(), .groups = "drop")
>
> # Plot with default ggplot2 fill colors
> ggplot(satisfaction_summary, aes(x = factor(JobSatisfaction), y = Count, fill = Attrition)) +
```

```
+   geom_bar(stat = "identity", position = "fill") +
+   labs(title = "Attrition Proportion by Job Satisfaction",
+        x = "Job Satisfaction (1 = Low, 4 = High)",
+        y = "Proportion") +
+   theme_minimal()
> library(pROC)
Type 'citation("pROC")' for a citation.

Attaching package: ‘pROC’

The following objects are masked from ‘package:stats’:

    cov, smooth, var

>
> # Predict probabilities
> test_probs <- predict(model_rq4, newdata = test_clean, type = "response")
Error: object 'model_rq4' not found
>
> # Generate ROC object and plot
> roc_obj <- roc(test_clean$Attrition, test_probs)
Error: object 'test_clean' not found
>
> # Plot
> plot(roc_obj, col = "purple", lwd = 2, main = "ROC Curve for Attrition Prediction")
Error: object 'roc_obj' not found
> abline(a = 0, b = 1, lty = 2, col = "gray")
Error in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...) :
  plot.new has not been called yet
>
```