AIT664_Group_Project

Neha Reddy Yenugu

2024-02-29

```r
# Load the readr package
library(readr)

# Specify the file path of your CSV file
file_path <- "C:\\Users\\nredd\\OneDrive\\Documents\\3rd sem\\AIT664\\MetroPT3(AirCompressor).csv"

# Read the CSV file
data <- read_csv(file_path)
```

```
## New names:
## Rows: 1516948 Columns: 17
## ── Column specification
## ─────────────────────────────────────────────── Delimiter: "," dbl
## (16): ...1, TP2, TP3, H1, DV_pressure, Reservoirs, Oil_temperature, Mot... dttm
## (1): timestamp
## ℹ Use `spec()` to retrieve the full column specification for this data. ℹ
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## • `` -> `...1`
```

```r
# Verify the data
head(data)
```

```
## # A tibble: 6 × 17
##    ...1 timestamp               TP2   TP3    H1 DV_pressure Reservoirs
##   <dbl> <dttm>                <dbl> <dbl> <dbl>       <dbl>      <dbl>
## 1     0 2020-02-01 00:00:00 -0.0120  9.36  9.34     -0.0240       9.36
## 2    10 2020-02-01 00:00:10 -0.0140  9.35  9.33     -0.0220       9.35
## 3    20 2020-02-01 00:00:19 -0.0120  9.34  9.32     -0.0220       9.34
## 4    30 2020-02-01 00:00:29 -0.0120  9.33  9.31     -0.0220       9.33
## 5    40 2020-02-01 00:00:39 -0.0120  9.32  9.30     -0.0220       9.32
## 6    50 2020-02-01 00:00:49 -0.0120  9.31  9.29     -0.0240       9.31
## # ℹ 10 more variables: Oil_temperature <dbl>, Motor_current <dbl>, COMP <dbl>,
## #   DV_eletric <dbl>, Towers <dbl>, MPG <dbl>, LPS <dbl>,
## #   Pressure_switch <dbl>, Oil_level <dbl>, Caudal_impulses <dbl>
```

```r
# Load necessary libraries
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
```

```
library(ggplot2)
library(lubridate) # for working with dates
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(corrplot) # for correlation plots
```

```
## corrplot 0.92 loaded
```

```
# Convert timestamp to datetime format
data$timestamp <- as_datetime(data$timestamp)
```

```
summary(data)
```

```
##        ...1            timestamp                        TP2
##  Min.   :       0   Min.   :2020-02-01 00:00:00.00   Min.   :-0.032
##  1st Qu.: 3792368   1st Qu.:2020-03-23 05:05:04.50   1st Qu.:-0.014
##  Median : 7584735   Median :2020-05-17 08:07:06.00   Median :-0.012
##  Mean   : 7584735   Mean   :2020-05-16 22:58:36.63   Mean   : 1.368
##  3rd Qu.:11377102   3rd Qu.:2020-07-10 03:07:27.50   3rd Qu.:-0.010
##  Max.   :15169470   Max.   :2020-09-01 03:59:50.00   Max.   :10.676
##       TP3              H1          DV_pressure       Reservoirs
##  Min.   : 0.730   Min.   :-0.036   Min.   :-0.03200   Min.   : 0.712
##  1st Qu.: 8.492   1st Qu.: 8.254   1st Qu.:-0.02200   1st Qu.: 8.494
##  Median : 8.960   Median : 8.784   Median :-0.02000   Median : 8.960
##  Mean   : 8.985   Mean   : 7.568   Mean   : 0.05596   Mean   : 8.985
##  3rd Qu.: 9.492   3rd Qu.: 9.374   3rd Qu.:-0.01800   3rd Qu.: 9.492
##  Max.   :10.302   Max.   :10.288   Max.   : 9.84400   Max.   :10.300
##  Oil_temperature Motor_current      COMP          DV_eletric
##  Min.   :15.40   Min.   :0.020   Min.   :0.000   Min.   :0.0000
##  1st Qu.:57.77   1st Qu.:0.040   1st Qu.:1.000   1st Qu.:0.0000
##  Median :62.70   Median :0.045   Median :1.000   Median :0.0000
##  Mean   :62.64   Mean   :2.050   Mean   :0.837   Mean   :0.1606
##  3rd Qu.:67.25   3rd Qu.:3.808   3rd Qu.:1.000   3rd Qu.:0.0000
##  Max.   :89.05   Max.   :9.295   Max.   :1.000   Max.   :1.0000
##      Towers           MPG              LPS         Pressure_switch
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.:1.0000   1st Qu.:0.00000   1st Qu.:1.0000
##  Median :1.0000   Median :1.0000   Median :0.00000   Median :1.0000
##  Mean   :0.9198   Mean   :0.8327   Mean   :0.00342   Mean   :0.9914
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
##   Oil_level      Caudal_impulses
##  Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.:1.0000
##  Median :1.0000   Median :1.0000
##  Mean   :0.9042   Mean   :0.9371
##  3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000
```

```
# Time Series Analysis
# Example: Plot oil temperature over time
ggplot(data, aes(x = timestamp, y = Oil_temperature)) +
  geom_line() +
  labs(title = "Oil Temperature Over Time",
       x = "Timestamp",
       y = "Oil Temperature")
```

## Oil Temperature Over Time
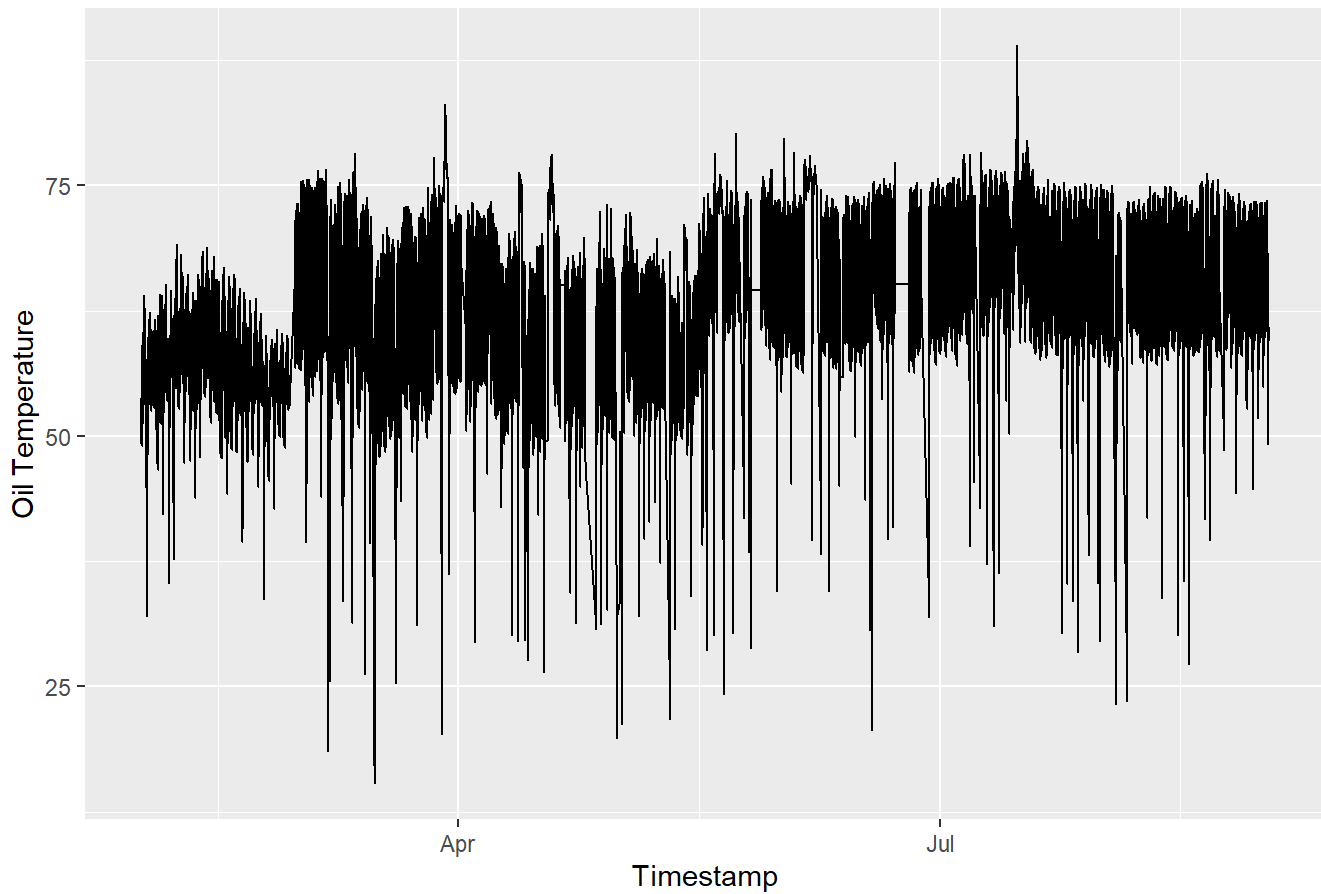


```
ggplot(data, aes(x = timestamp, y = DV_pressure)) +
  geom_line() +
  labs(title = "DV Pressure Over Time",
       x = "Timestamp",
       y = "DV Pressure")
```
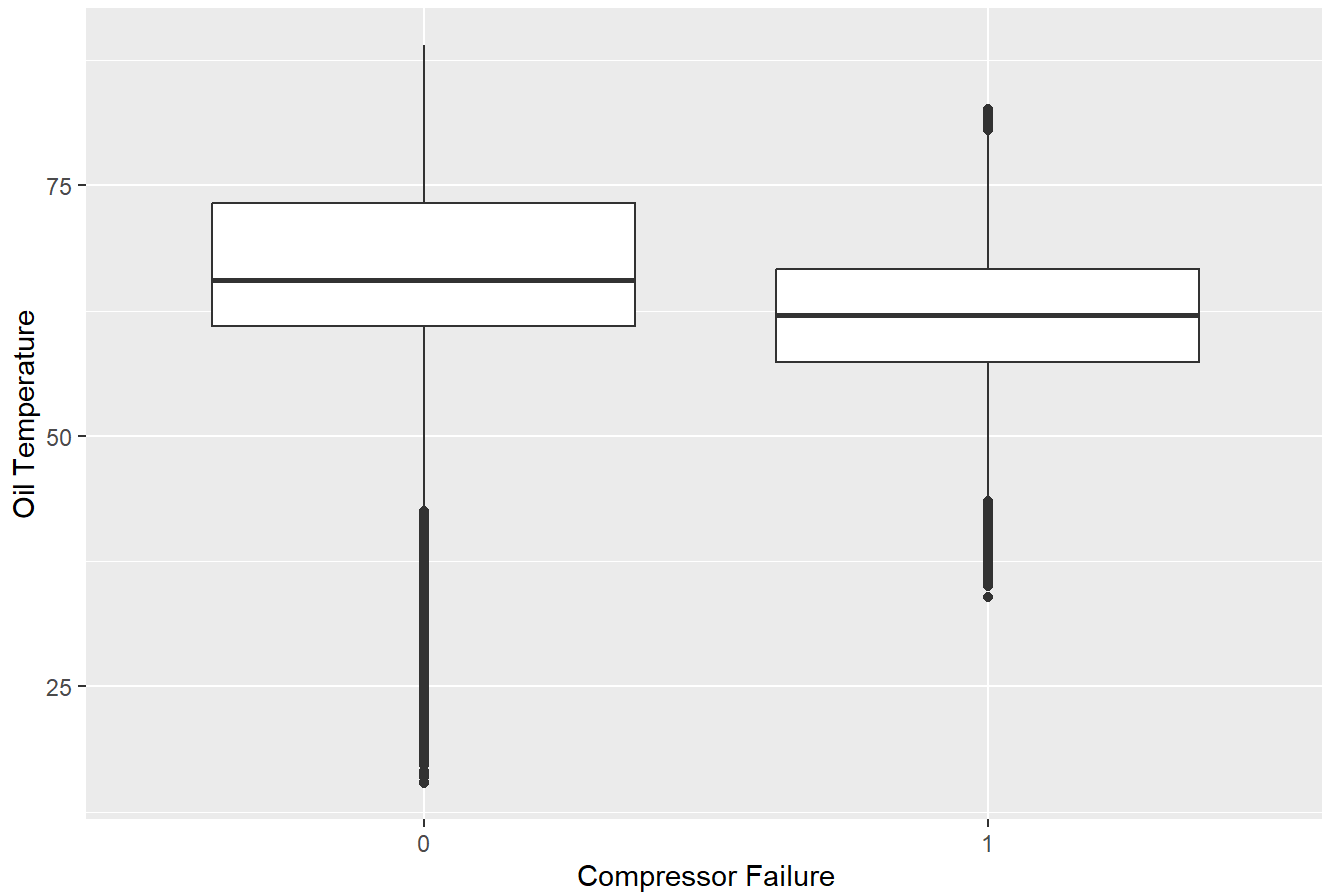
## DV Pressure Over Time
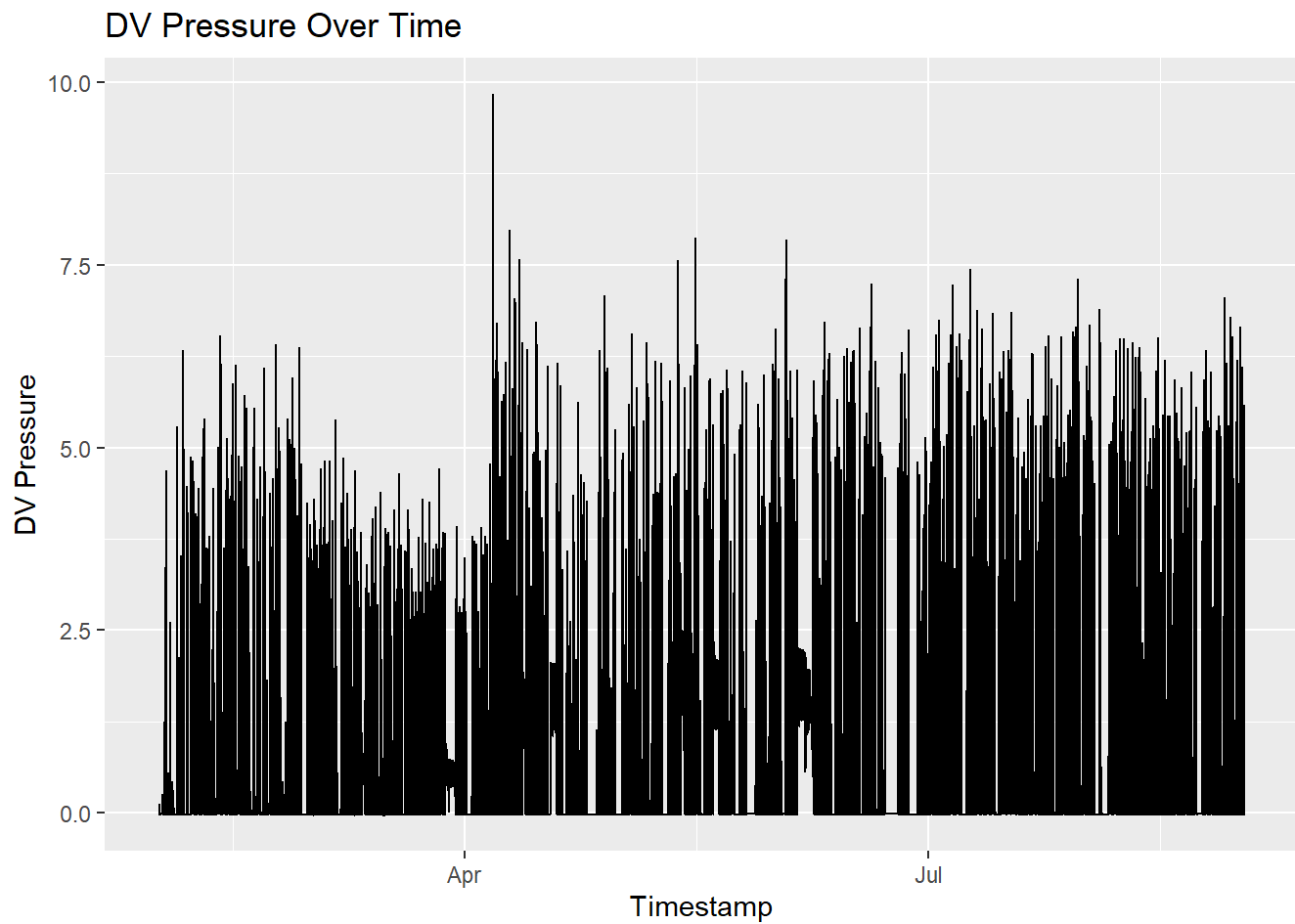


```
# Example: Scatter plot of motor current vs. oil temperature
ggplot(data, aes(x = Motor_current, y = Oil_temperature)) +
  geom_point() +
  labs(title = "Motor Current vs. Oil Temperature",
       x = "Motor Current",
       y = "Oil Temperature")
```

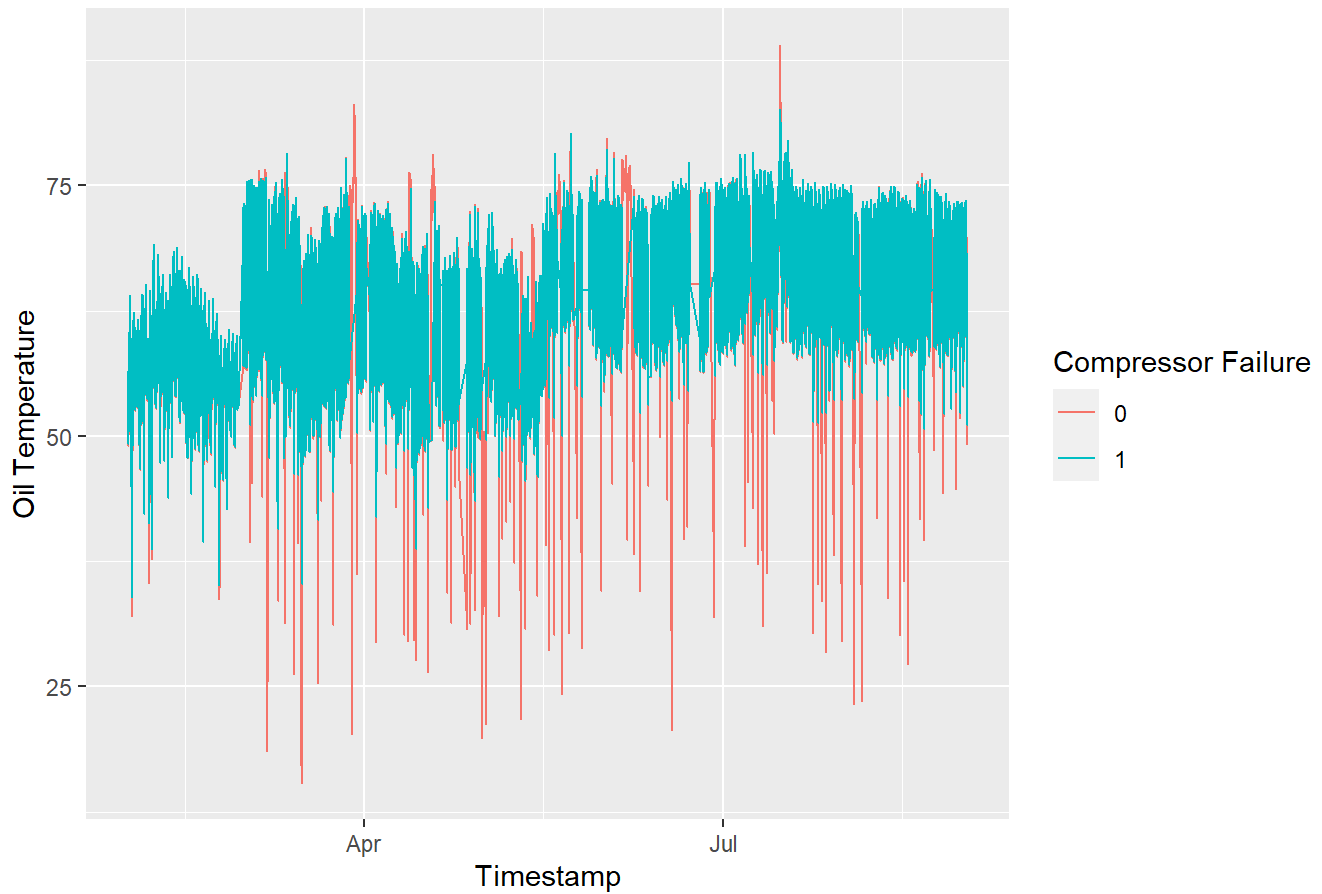## Motor Current vs. Oil Temperature
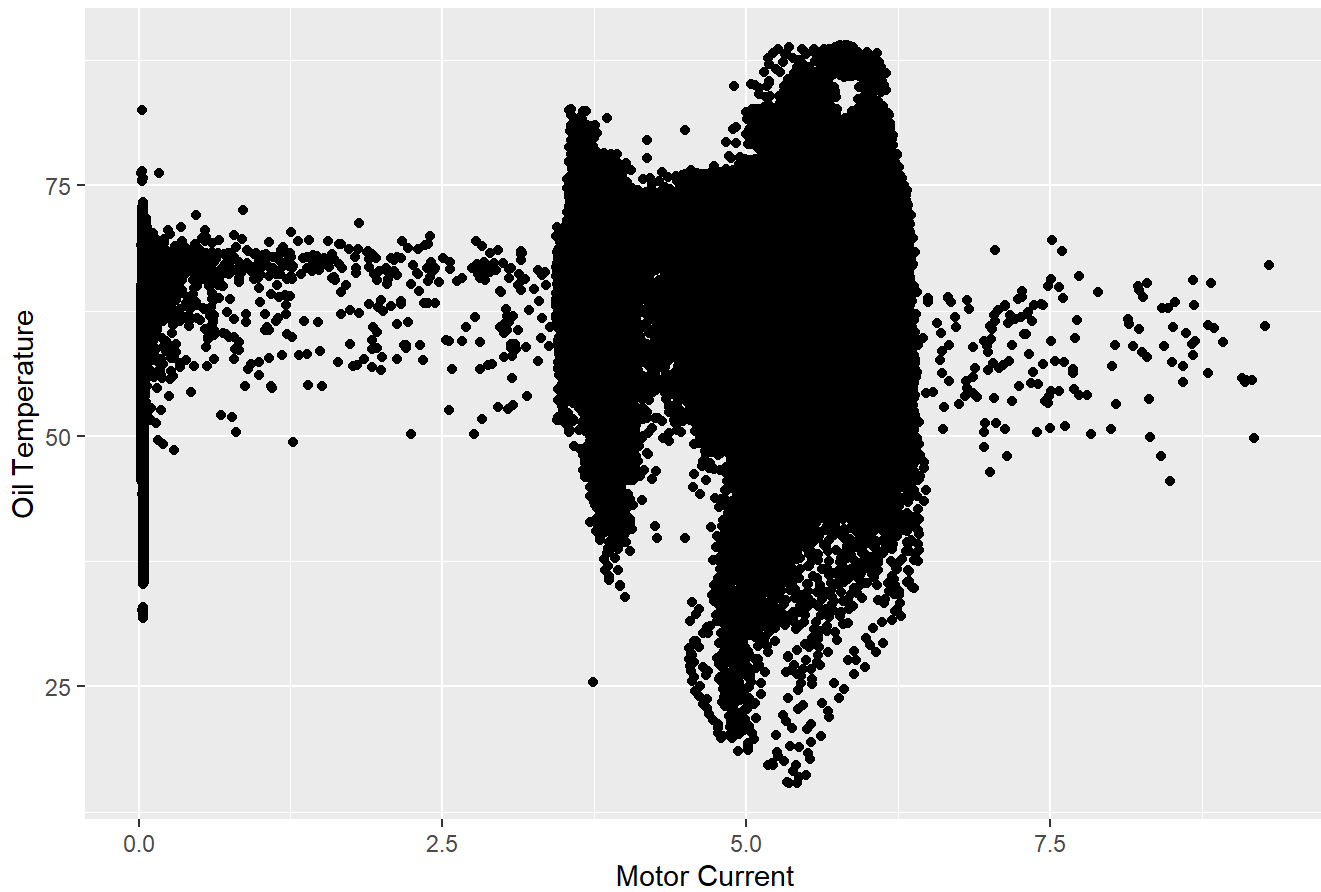

```
# Example: Box plot of oil temperature for failed and non-failed compressors
ggplot(data, aes(x = factor(COMP), y = Oil_temperature)) +
  geom_boxplot() +
  labs(title = "Box Plot of Oil Temperature for Failed and Non-Failed Compressors",
       x = "Compressor Failure",
       y = "Oil Temperature")
```

Box Plot of Oil Temperature for Failed and Non-Failed Compressors


```
# Time Series Plot
ggplot(data, aes(x = timestamp, y = Oil_temperature, color = factor(COMP))) +
  geom_line() +
  labs(title = "Oil Temperature Over Time by Compressor Failure",
       x = "Timestamp",
       y = "Oil Temperature",
       color = "Compressor Failure")
```

Oil Temperature Over Time by Compressor Failure


```
# Conditional Box Plots
ggplot(data, aes(x = factor(COMP), y = data$DV_pressure, fill = factor(COMP))) +
  geom_boxplot() +
  facet_wrap(~COMP) +
  labs(title = "Conditional Box Plot of DV pressure for Failed and Non-Failed Compressors",
       x = "Compressor Failure",
       y = "Oil Temperature",
       fill = "Compressor Failure")
```

Conditional Box Plot of DV pressure for Failed and Non-Failed Compressors
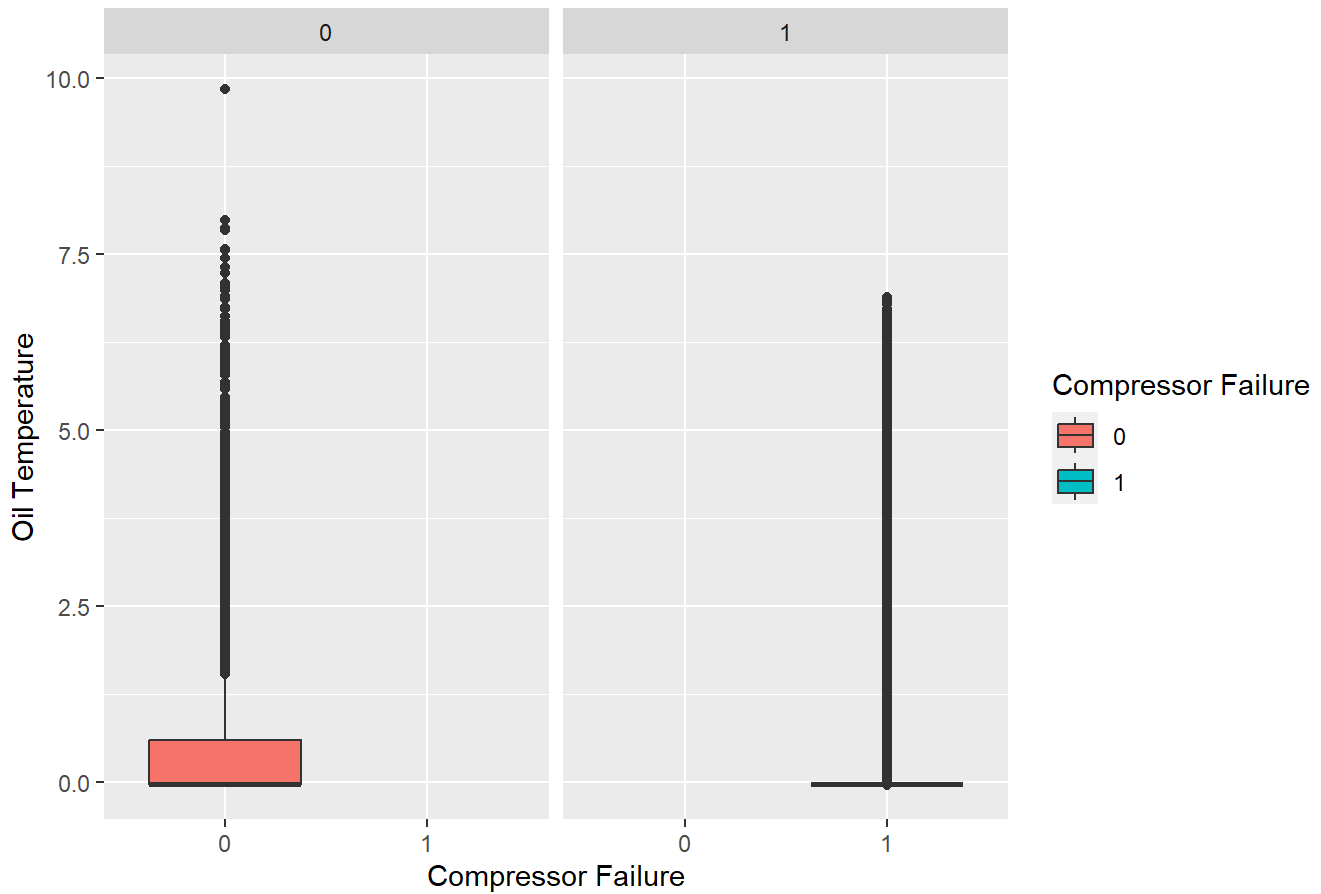

```
# Visualizing the distribution of compressor failures
ggplot(data, aes(x = COMP)) +
  geom_bar() +
  labs(title = "Distribution of Compressor Failures",
       x = "Compressor Failure",
       y = "Count")
```

Distribution of Compressor Failures