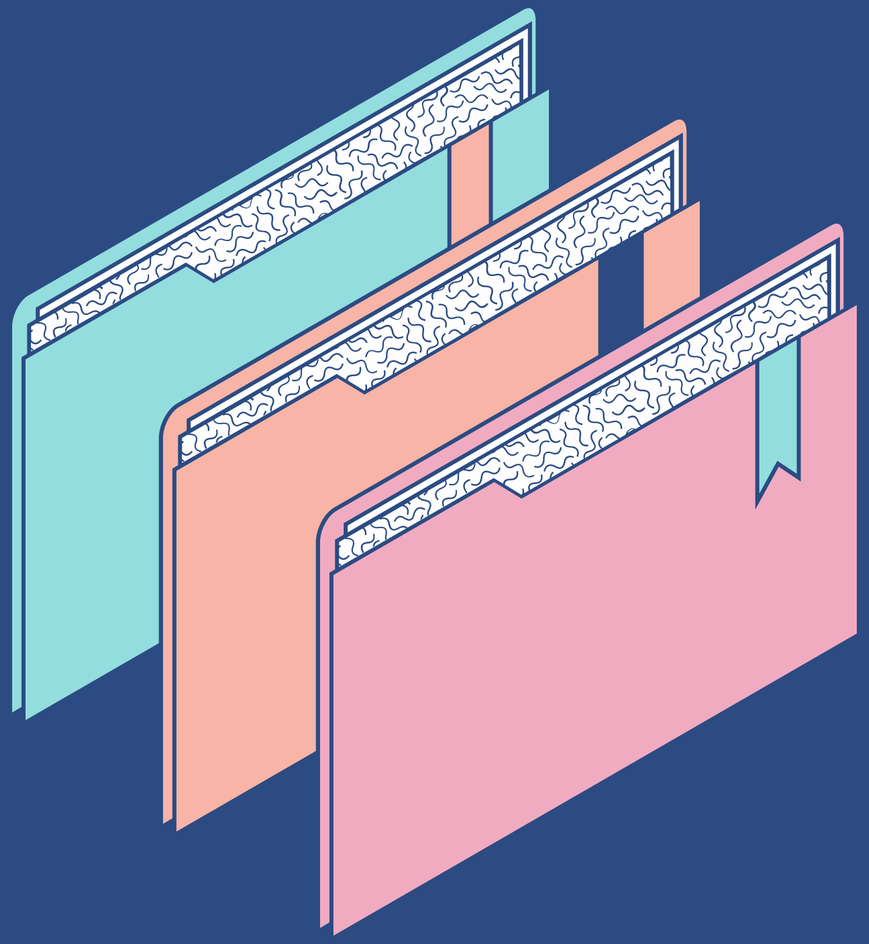
PHM08 Prognostics Data Challenge

Joud AlFarra S20106605

Leena Alam S20106371

Lamar AlSobaihi S20106553



Content

Introduction & Problem Statement

- Predicting aircraft component life by using data to predict remaining operational cycles until failure is the problem statement addressed by the PHM 2008 Challenge.
- Our project aims to predict the remaining useful life of turbofan engines using machine learning.
- Our performance will be evaluated using the mean absolute error metric.

Data

- We will use the dataset provided by the PHM 2008 Challenge, which includes sensor measurements and failure times of 218 turbofan engines.
- The datasets are:
  - train.txt
  - test.txt
  - final_test.txt
- Each row is a snapshot of data taken during a single operational cycle.
- The columns correspond to:
  - unit number
  - time, in cycles
  - operational setting 1-3
  - sensor measurement 1-21

Imported Libraries

```
LIBRARY(TIDYVERSE)
LIBRARY(CARET)
LIBRARY(GGPLOT2)
LIBRARY(RANDOMFOREST)
```

Loading the Data

- train_data <- read.table("train.txt", header = TRUE)
- test_data <- read.table("test.txt", header = TRUE)
- final_test_data <- read.table("final_test.txt", header = TRUE)

Data	
final_test_data	55156 obs. of 26 variables
test_data	29820 obs. of 26 variables
train_data	45918 obs. of 26 variables

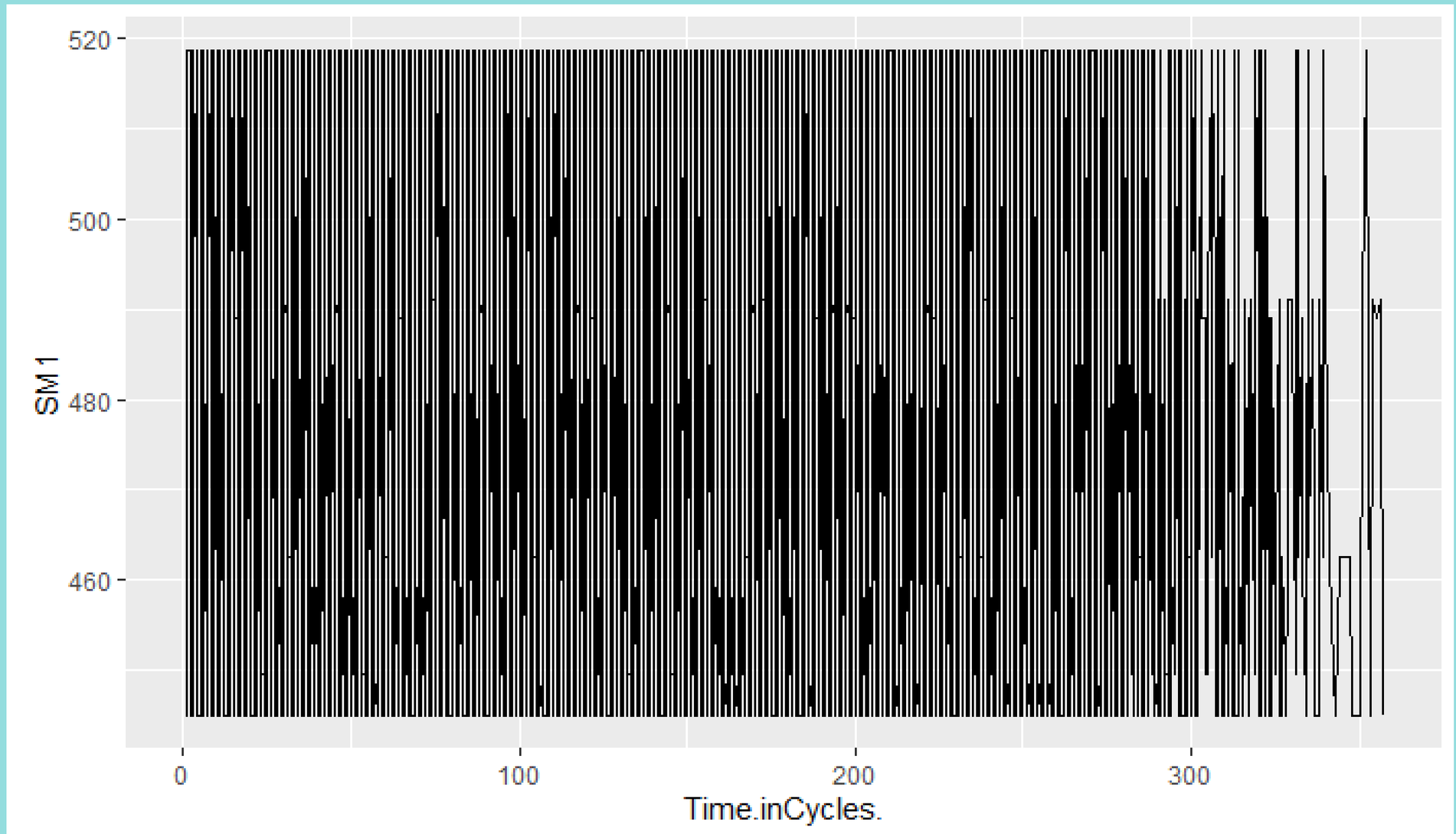
Cleaning the Data

Checked for missing values using the sum(is.na()) function. If there are any missing values, they were removed.

- sum(is.na(train_data))

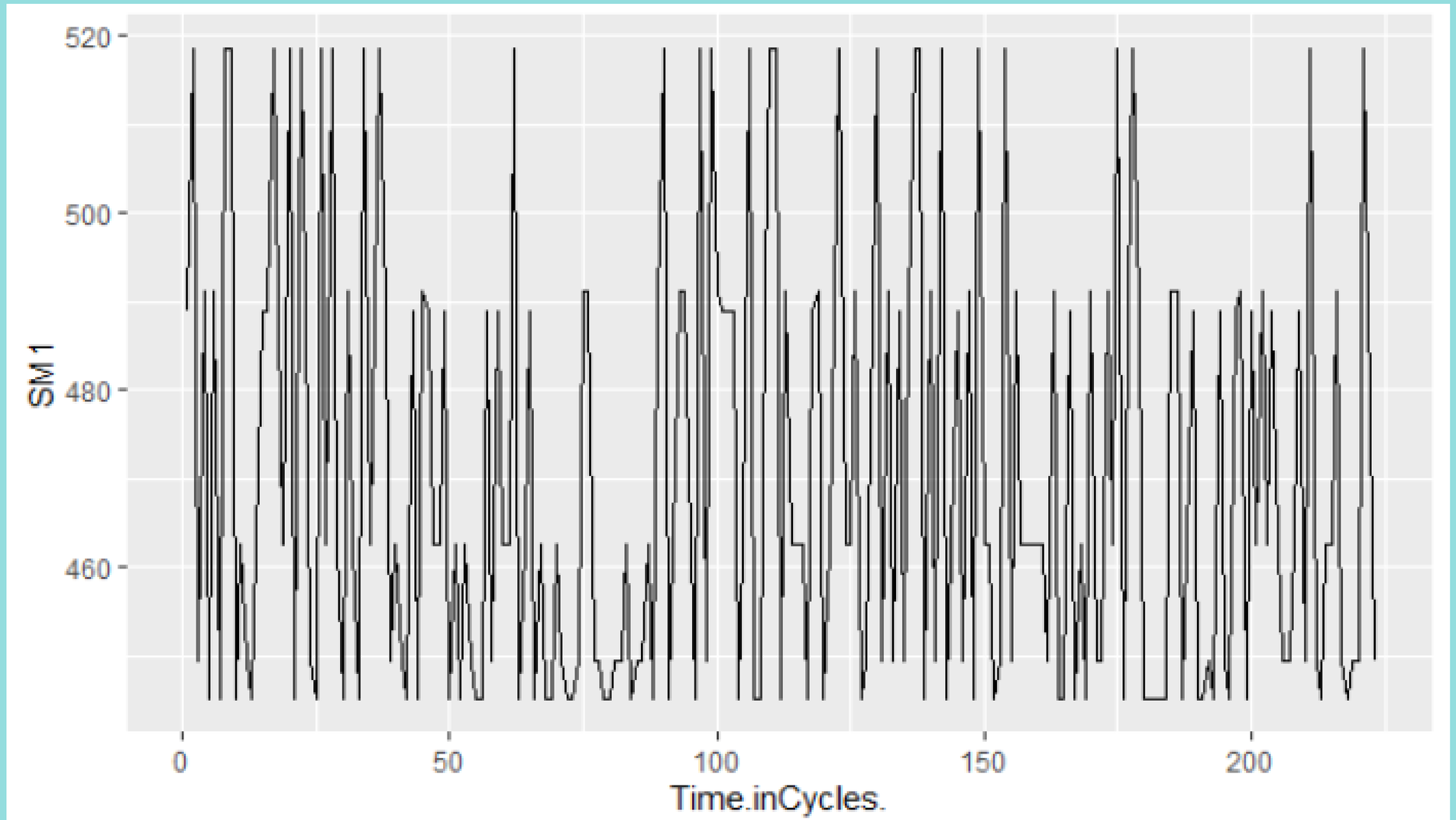
Visualization of Data

Sensor Measurement 1 vs Time Cycles



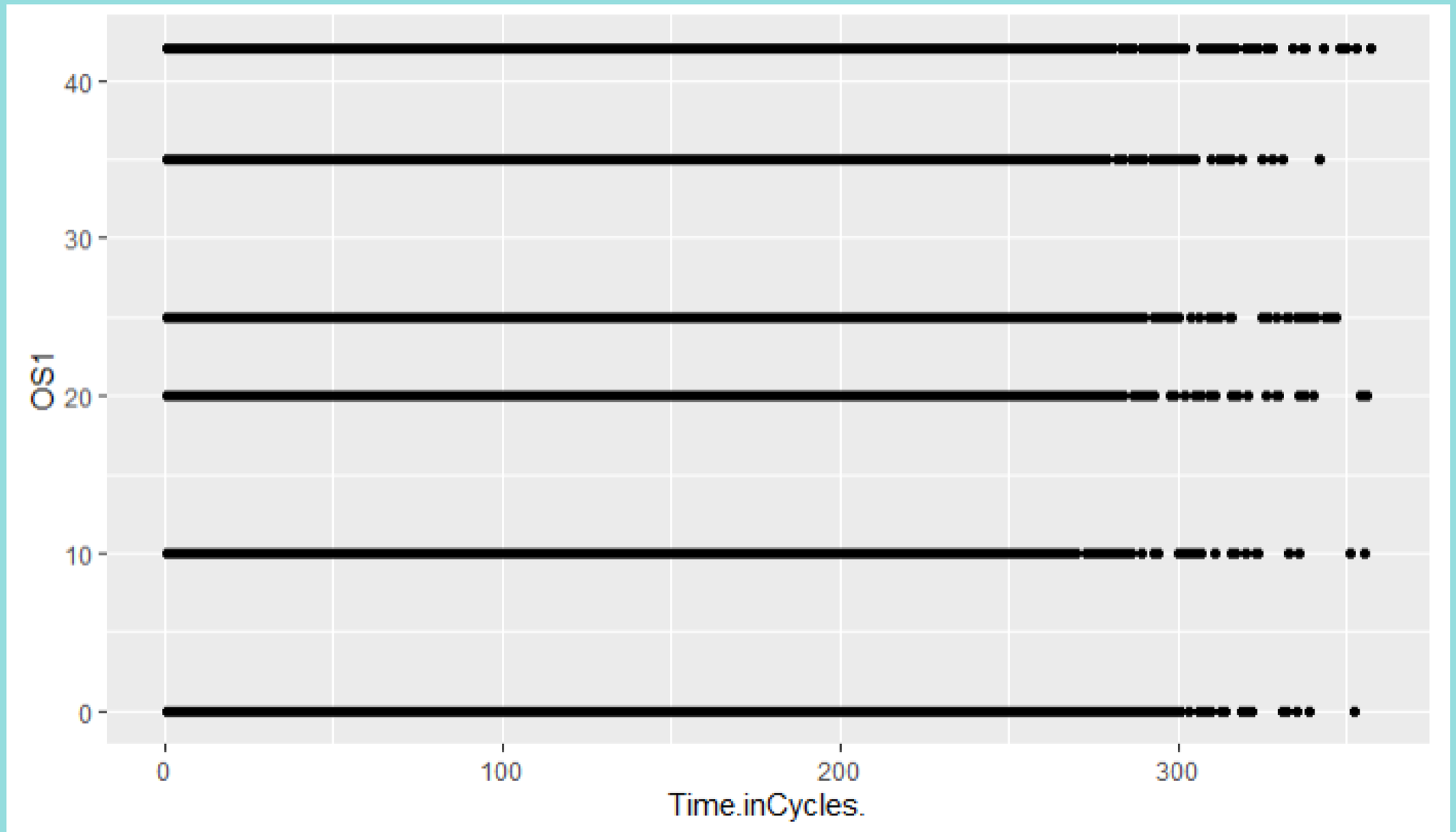
Visualization of Data

Sensor Measurement 1 vs Time Cycles for Engine 1



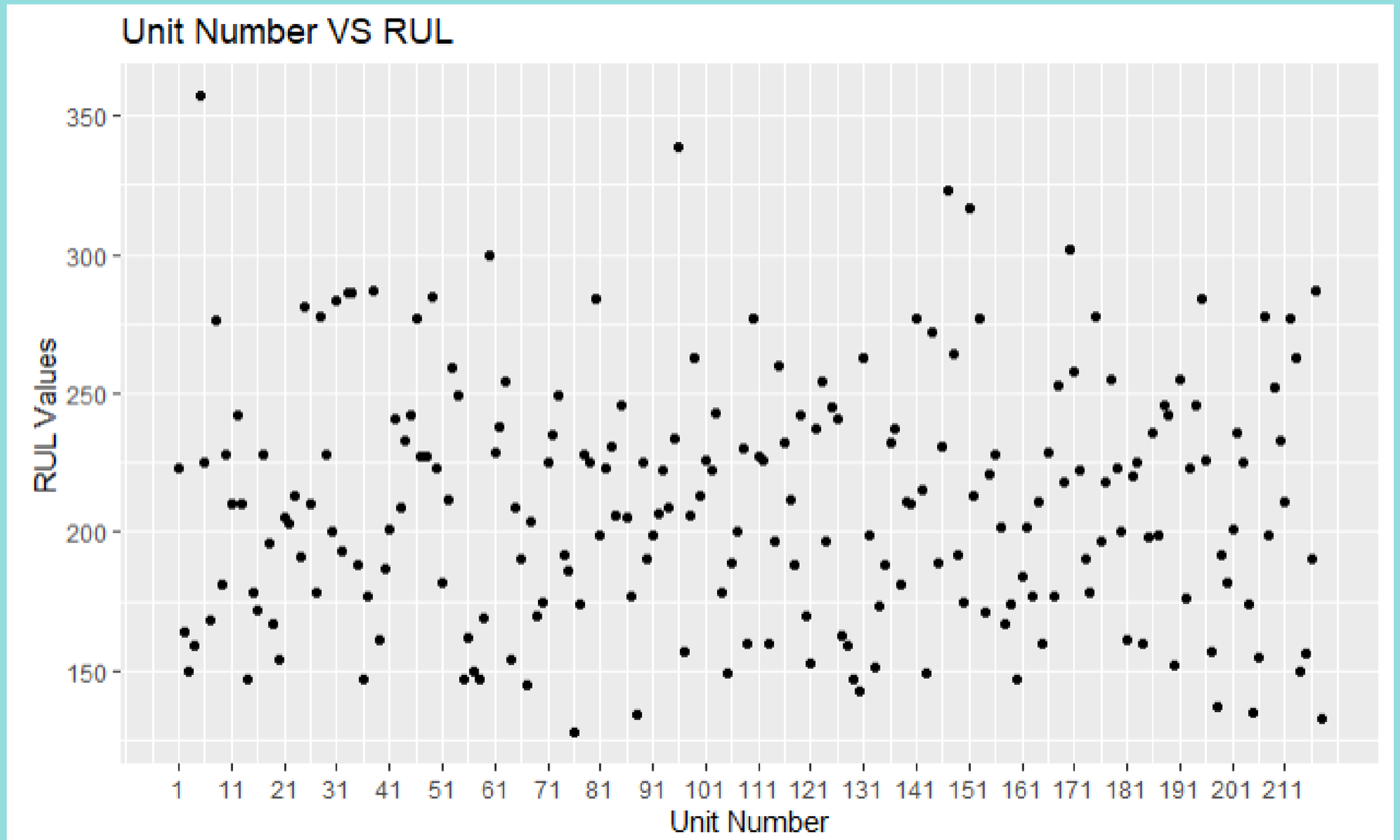
Visualization of Data

Operational Setting 1 vs Time Cycles for All Engines



Calculating the RUL (Remaining Useful Life) for the train_data

Calculated the RUL for each engine and visualized it as a scatterplot



Training

Training the data using the Random Forest Machine Learning Algorithm

model <- randomForest(RUL ~ ., data = train_data)

```
```{r}
set.seed(123)
# Train a random forest model
library(randomForest)
model <- randomForest(RUL ~ ., data = train_data)
```
```

Prediction

Trying to predict the RULs of the test_data

```
# Make predictions on the testing data
predictions <- predict(model, newdata = test_data)
```

	UnitNumber	predicted_RUL
1	1	198.5743
2	2	197.8591
3	3	193.1419
4	4	177.8723
5	5	188.6867
6	6	197.3574
7	7	206.4155
8	8	195.1573
9	9	185.6197

Evaluation

After prediction of the RULs, we calculated the actual values and evaluated the accuracy of the prediction.

```
# Evaluate the model performance
MAE <- mean(abs(predictions - test_data$predicted_RUL))
```

MAE	46.546851332746

Conclusion

- Our model predicts the remaining useful life of engines.
- The Random Forest method was used.
- We cleaned the data, trained our model, predicted, and evaluated our results.
- Key machine learning concepts were demonstrated.
- The code provides a framework for predictive model development.