ISYE 6414B Project

Yuanting Fan XXXX

2024-11-30

Contents

Clean the data: handle missing values

(1) Data collection: 29 variables (monthly or quarterly frequency), 157 data points in total

- Time data: Month, Date(the last day of each month of each year)
- LDV (light duty vehicle) sales unit: NEV car sales, Gasoline car sales
- Macro economy data GDP, Durable consumption, Non durable consumption,Residential fixed investment, Nonresidential fixed investment, unemployment rate, federal effective rate,M1,M2 (we remove the M3 variable due to the data incompleteness),Per capita disposable income
- Demographically data:Population,Percentage of employees who are middle-aged (25-55),Percentage of employees with a bachelor's degree or higher education
- Industry specific data:Tesla model S price, Gasoline price,Electric retail price
- Number of government new energy vehicles policies: these figures represent the incentive from the US government
- Covid 19 period: categorical data "1" for from Jan 2020 - May 2023 (the end of the federal COVID-19 PHE declaration) , otherwise "0"

(2) handling missing values

- transform quarterly data into monthly data by linear interpolation

- fill the missing values of Tesla model S price by using the previous valid price (It also makes sense because only when the company decides to change their price, the price floats)

(3) we do not scale data as it will stop doing boxcox transformation

```
data <- read.csv("D:/GT Master/1. Academic/Semester 1/ISYE6414/6414 Project/us market da
                      sep=",")

data$Year <- NULL

data$GDP[which(is.na(data$GDP))] <- approx(data$Date, data$GDP, xout = data$Date[which(

data$Durable.goods.consumption[which(is.na(data$Durable.goods.consumption))] <- approx(d

data$Nondurable.goods.consumption[which(is.na(data$Nondurable.goods.consumption))] <- ap

data$Residential.fixed.Investment[which(is.na(data$Residential.fixed.Investment))] <- ap

data$Nonresidential.fixed.Investment[which(is.na(data$Nonresidential.fixed.Investment))]

data$Population[which(is.na(data$Population))] <- approx(data$Date, data$Population, xou

data$Per.capita.disposable.income[which(is.na(data$Per.capita.disposable.income))] <- ap

library(tidyr)
data <- data %>% fill(Tesla.model.S.price, .direction = "down")
data <- data %>% fill(Electric.retail.price, .direction = "down")
data <- data %>% fill(Population, .direction = "down")
data <- data %>% fill(Per.capita.disposable.income, .direction = "down")
```

Exploratory Data Analysis

```
install.packages("plotly")
```

```
##   'plotly'   MD5
##
##
##  C:\Users\yuanting\AppData\Local\Temp\RtmpWUhATB\downloaded_packages
```

```
library(plotly)
```

```
## Warning:   'plotly' R 4.4.2
##      ggplot2
##
```

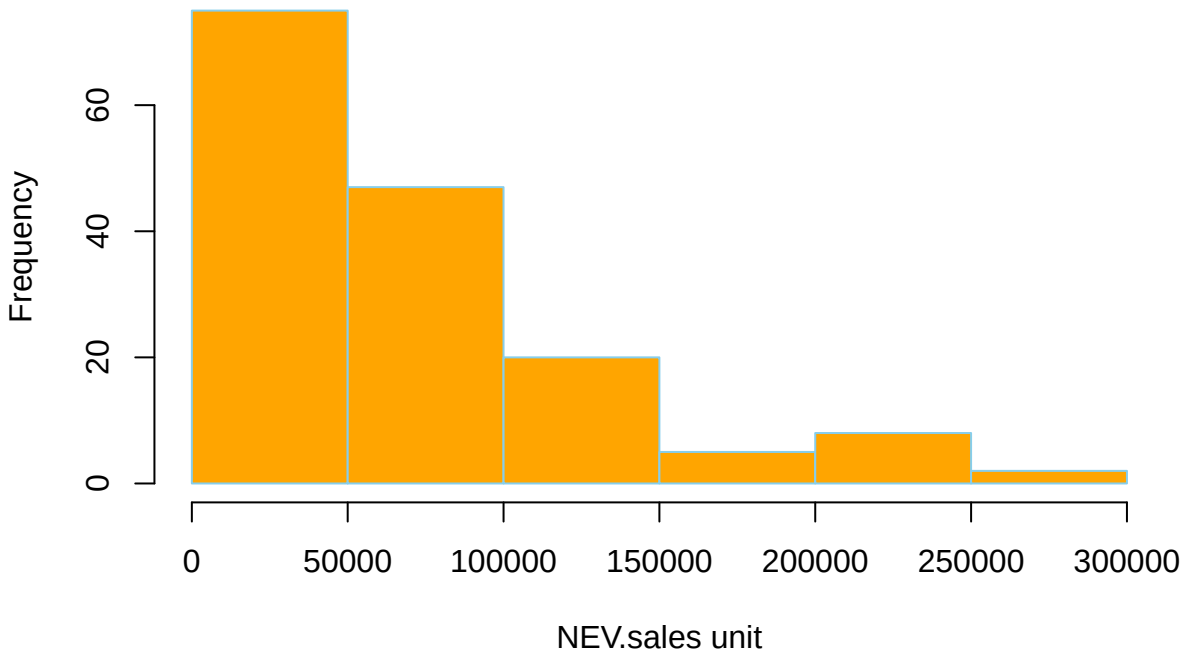
```
##     'plotly'
```

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

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
hist(data$Total.NEV.Sales,main="",xlab="NEV.sales unit",border = "skyblue",col="orange")
```



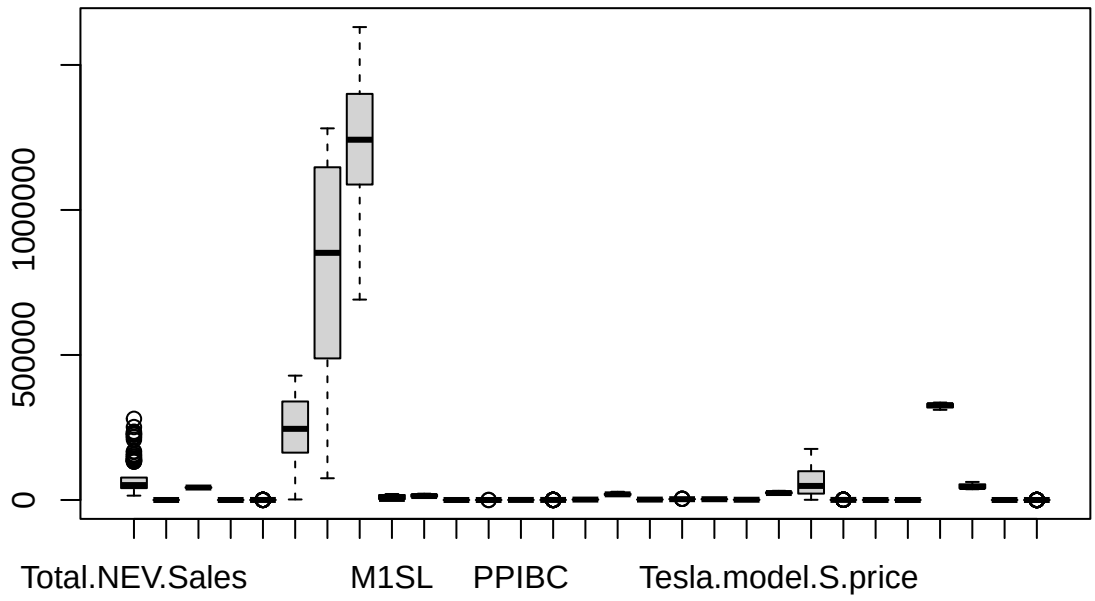

```
data$formatedate <- as.Date(data$Date, origin = "1970-01-01")
data$formatedate <- format(data$formatedate, "%Y%m")

plot_ly(data = data, x = ~formatedate) %>%
  add_lines(y = ~data$Total.NEV.Sales, name = "NEV.Sales", line = list(color = "blue")) 
  add_lines(y = ~data$Gasoline.LDV.sales, name = "Gasoline.sales", line = list(color = '
  layout(
    title = "LDV sales unit in US",
```

```
    yaxis = list(title = "NEV.Sales", side = "left"),
    yaxis2 = list(title = "Gasoline.sales", side = "right", overlaying = "y"),
    xaxis = list(title = "Date")
  )
```

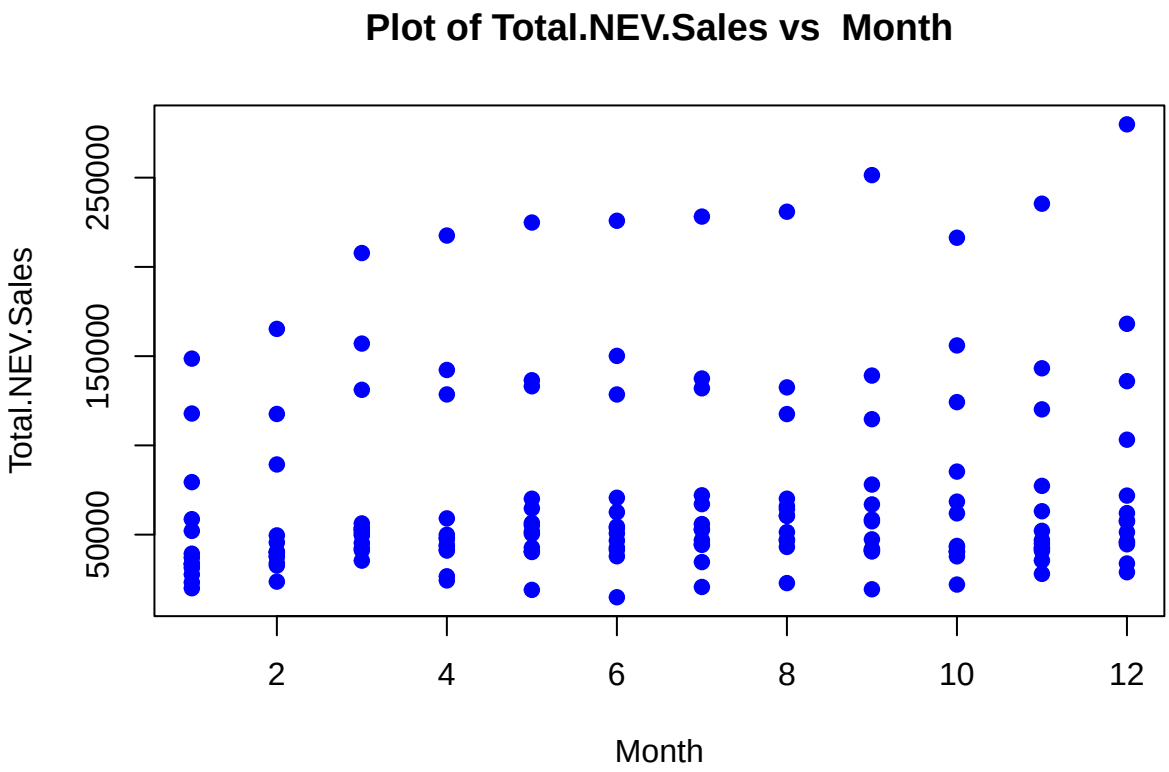
```
plot_ly(data = data, x = ~formatedate) %>%
  add_lines(y = ~log(data$Total.NEV.Sales), name = "log NEV.Sales", line = list(color =
  layout(
    title = "log NEV.Sales unit in US",
    xaxis = list(title = "Date")
  )
```

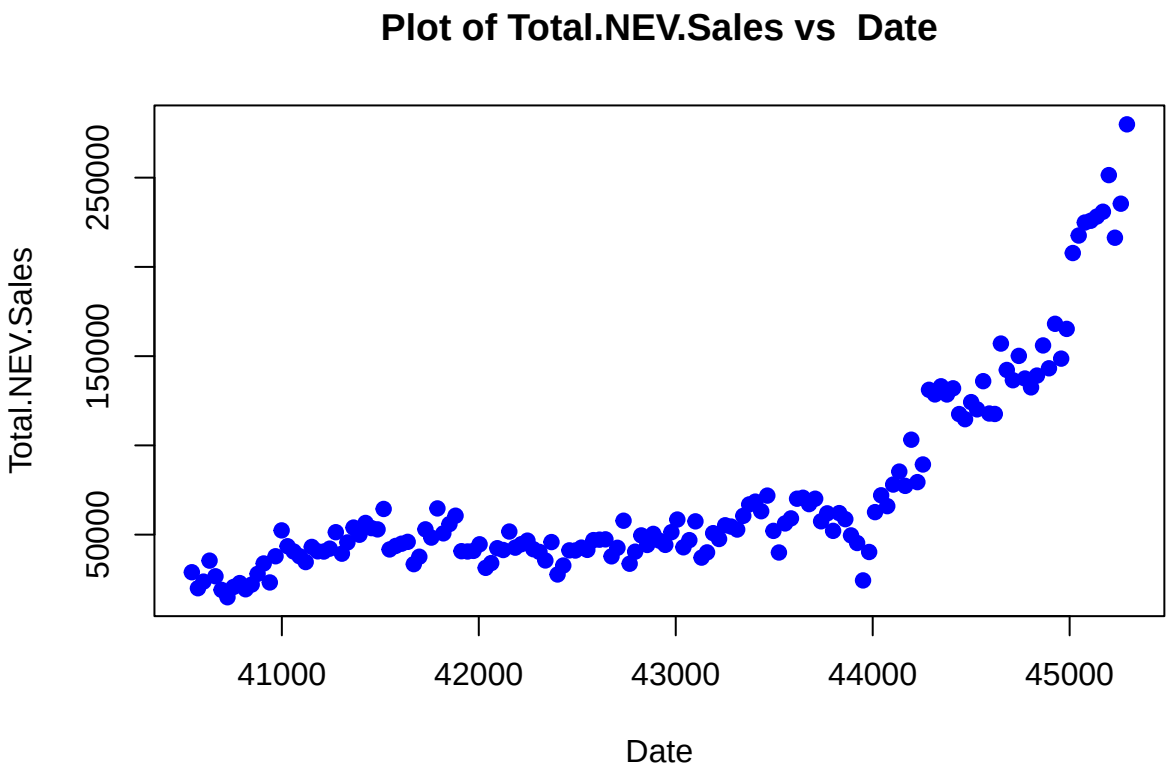
```
data$formatedate <- NULL
```

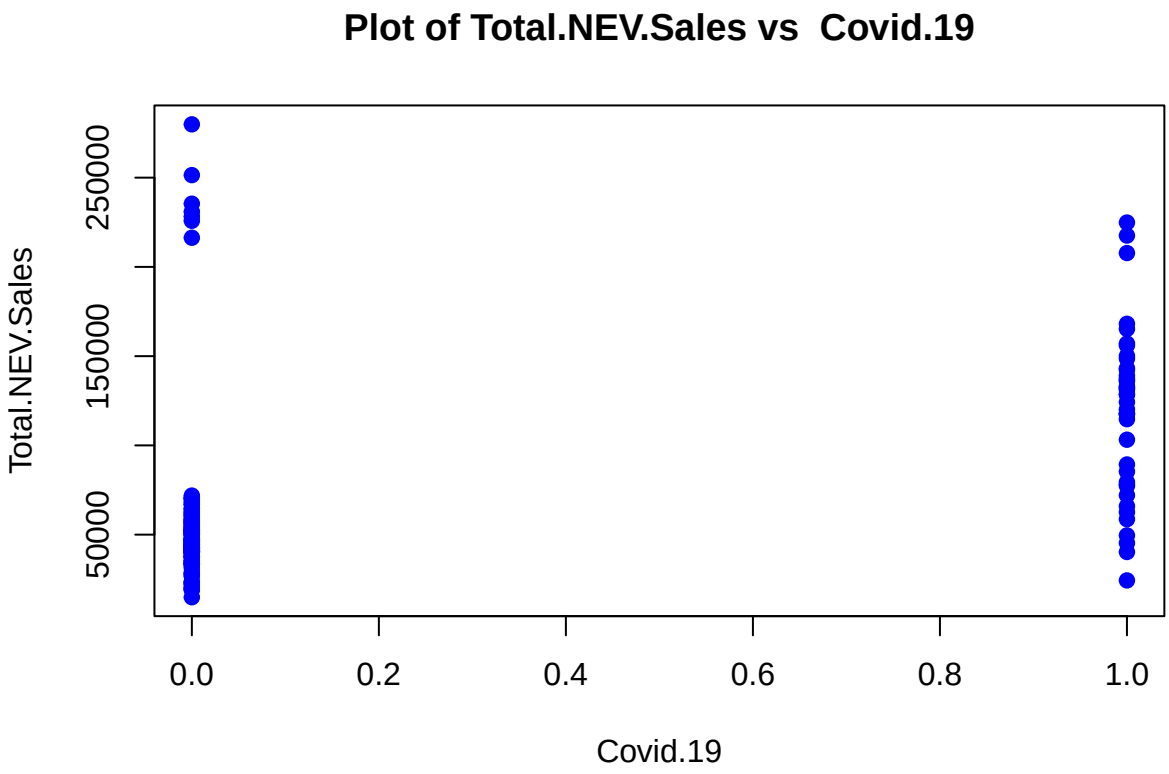
```
boxplot(data)
```

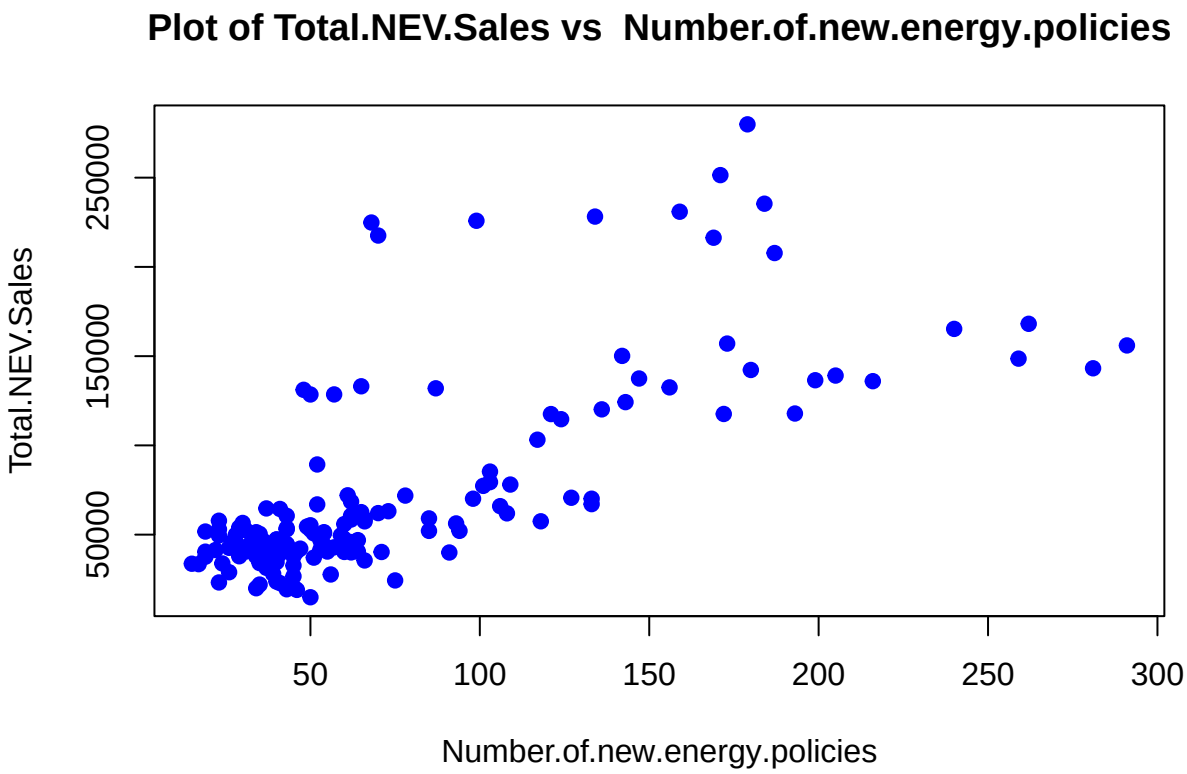


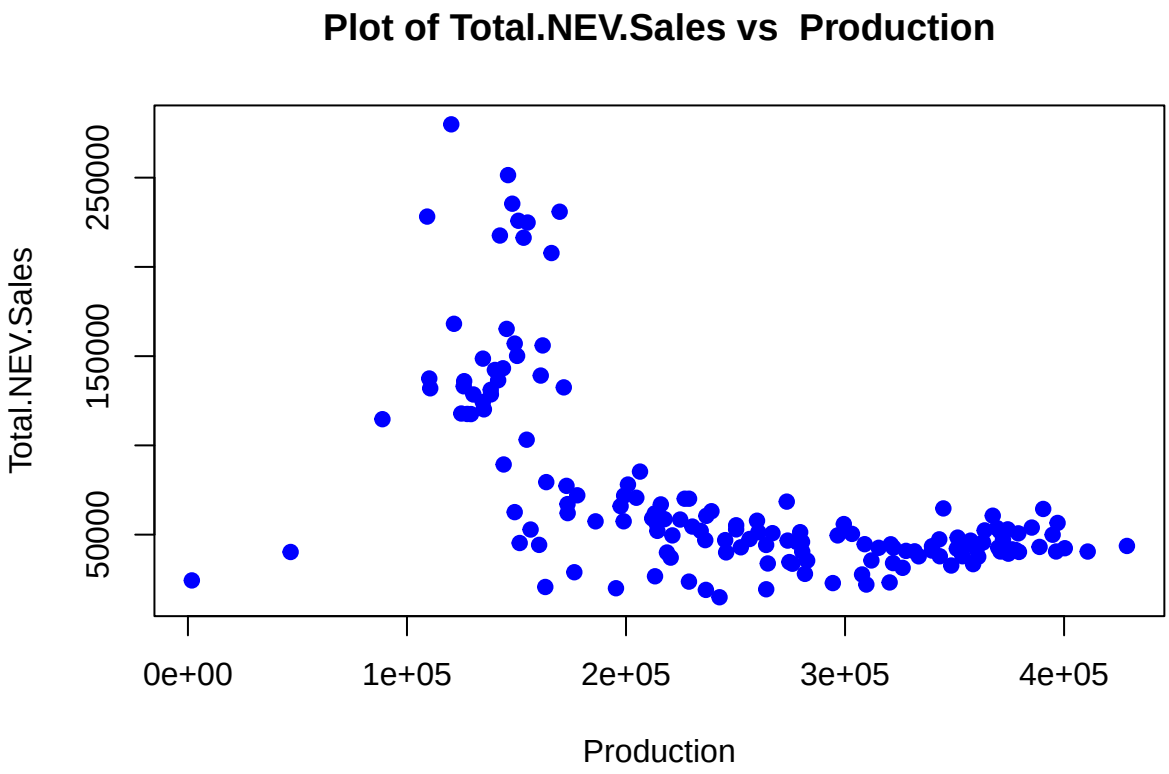

```
for (i in 2:29) {  # Looping through the other 10 variables
  plot(data[[i]], data$Total.NEV.Sales,
       main = paste("Plot of Total.NEV.Sales vs ", colnames(data)[i]),
       xlab = colnames(data)[i],ylab = "Total.NEV.Sales",
       col = "blue", pch = 19)
}
```

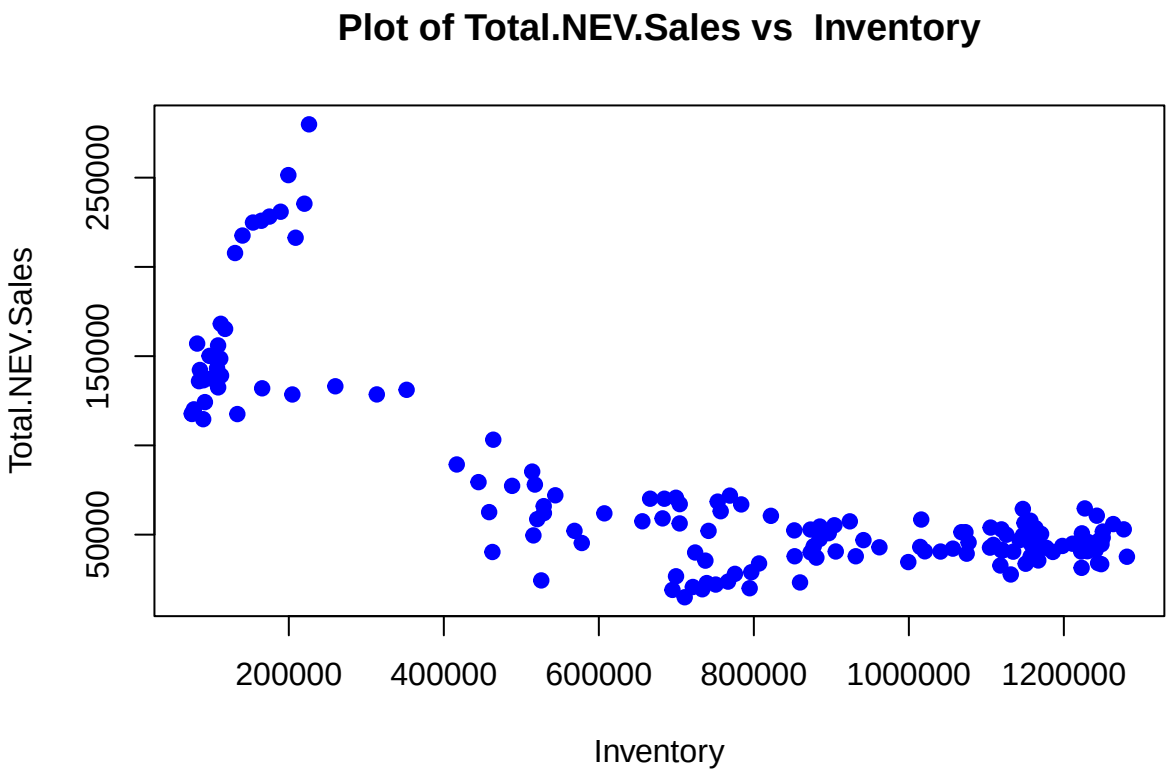

Plot of Total.NEV.Sales vs Month
Total.NEV.Sales
50000
150000
250000
2
4
6
8
10
12
Month

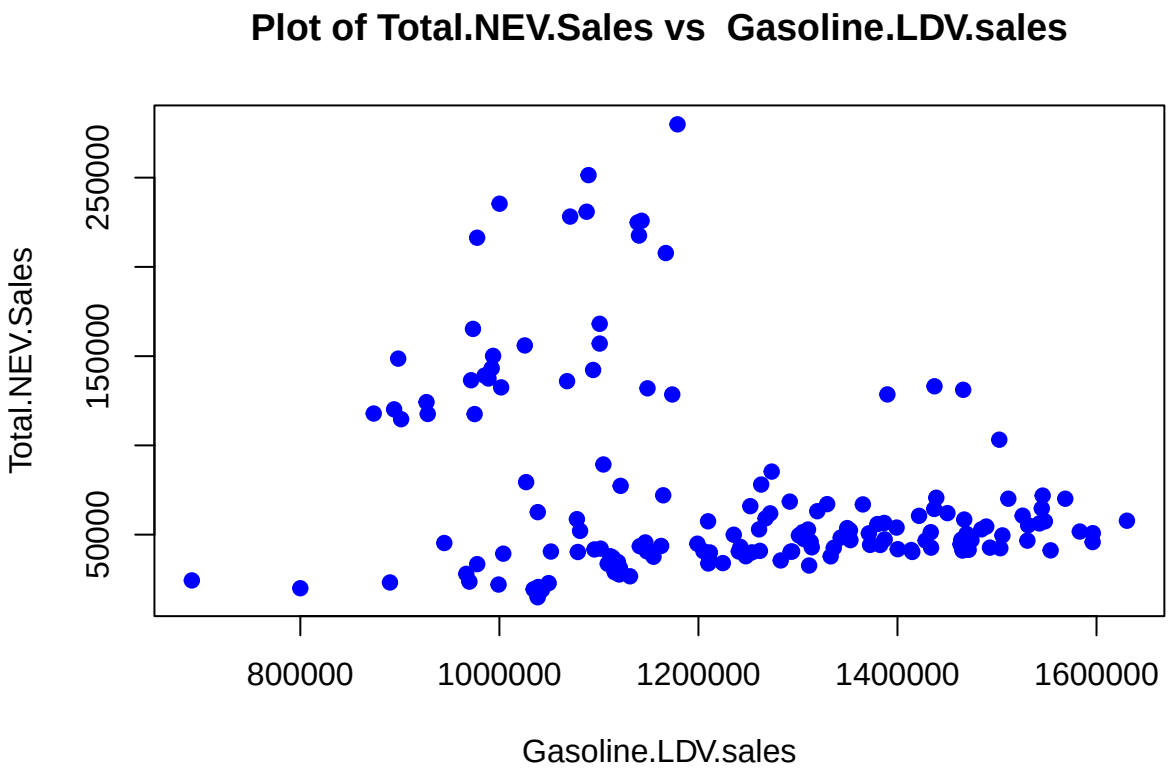

Plot of Total.NEV.Sales vs Date
Total.NEV.Sales
50000
150000
250000
41000
42000
43000
44000
45000
Date

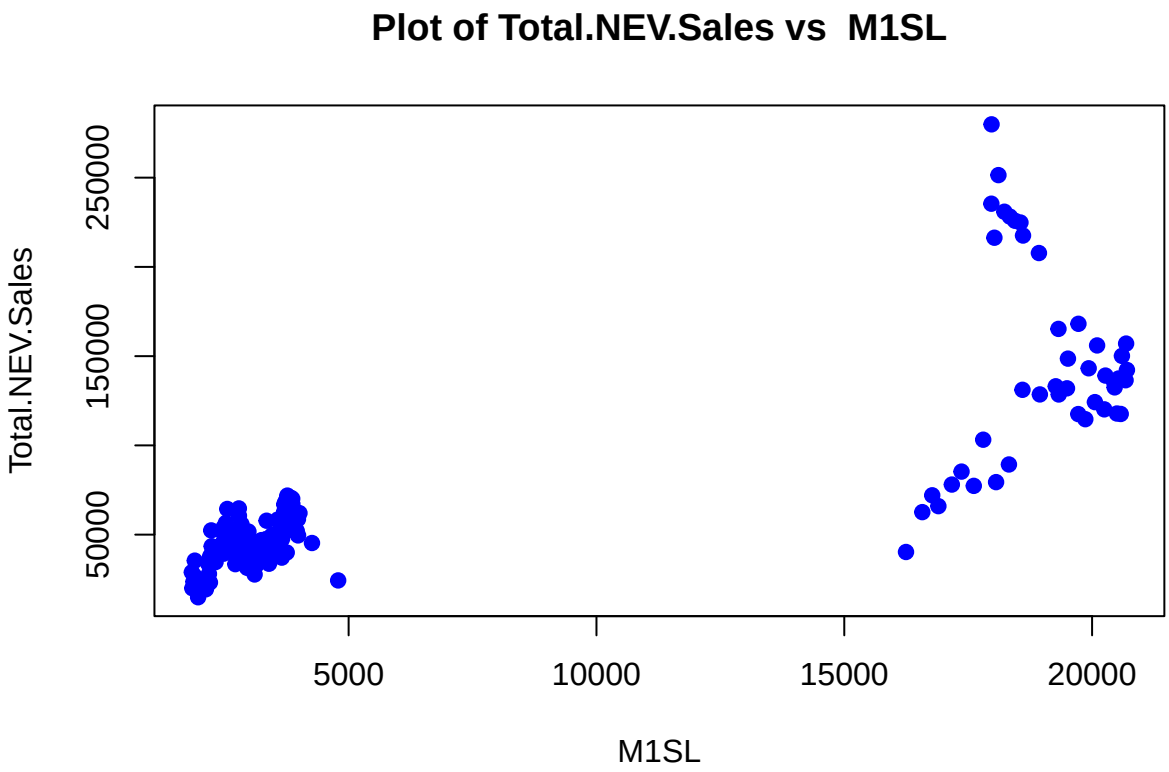

Plot of Total.NEV.Sales vs Covid.19
Total.NEV.Sales
50000
150000
250000
0.0
0.2
0.4
0.6
0.8
1.0
Covid.19

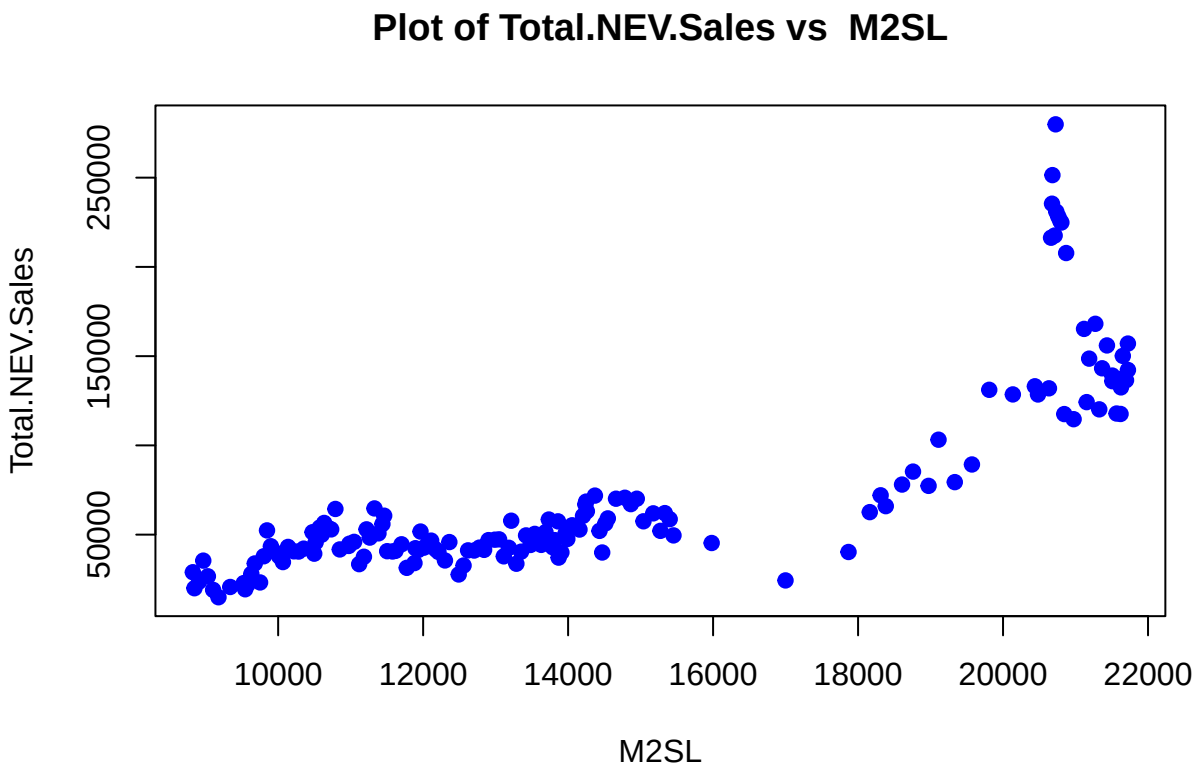

Plot of Total.NEV.Sales vs Number.of.new.energy.policies
Total.NEV.Sales
50000
150000
250000
50
100
150
200
250
300
Number.of.new.energy.policies

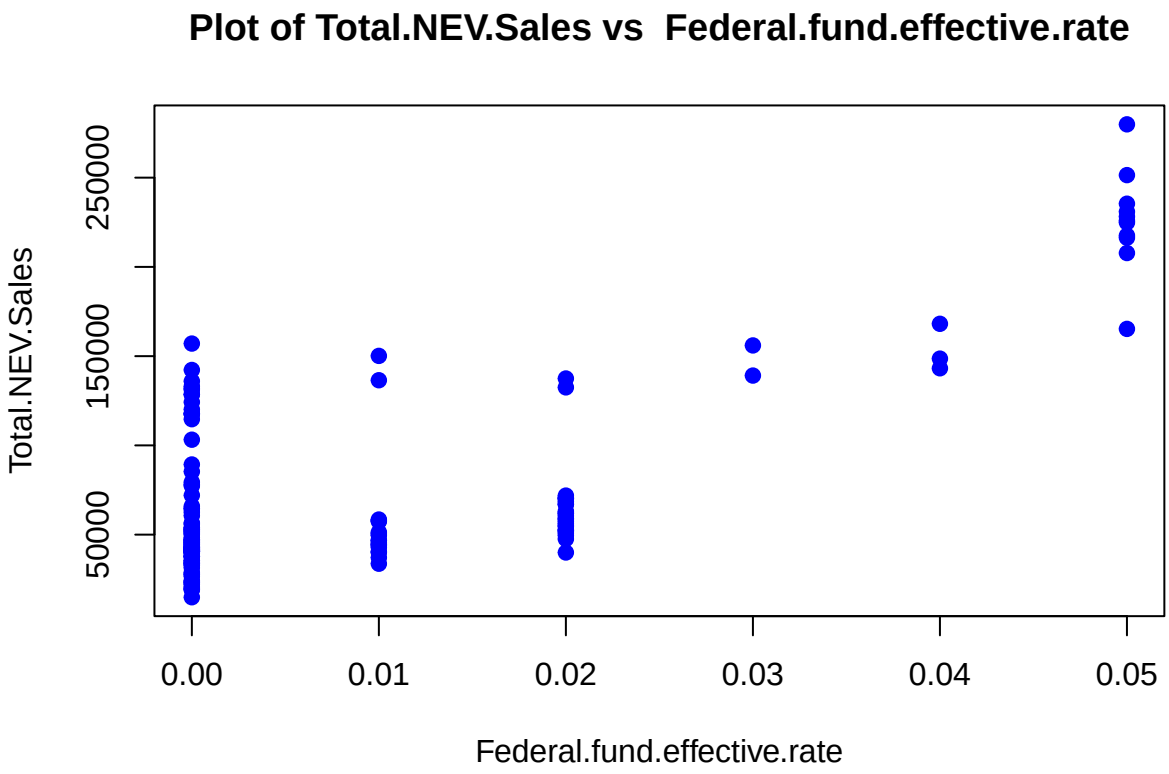

Plot of Total.NEV.Sales vs Production
Total.NEV.Sales
50000
150000
250000
0e+00
1e+05
2e+05
3e+05
4e+05
Production

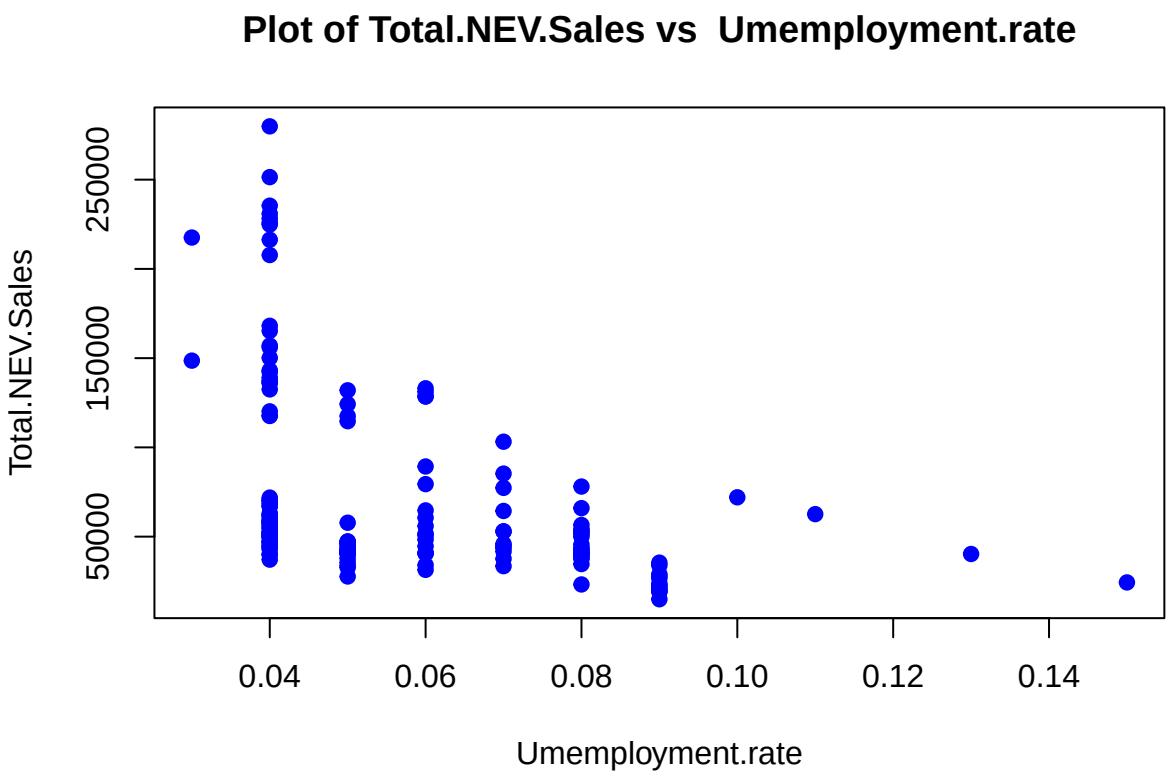

Plot of Total.NEV.Sales vs Inventory
Total.NEV.Sales
50000
150000
250000
200000
400000
600000
800000
1000000
1200000
Inventory

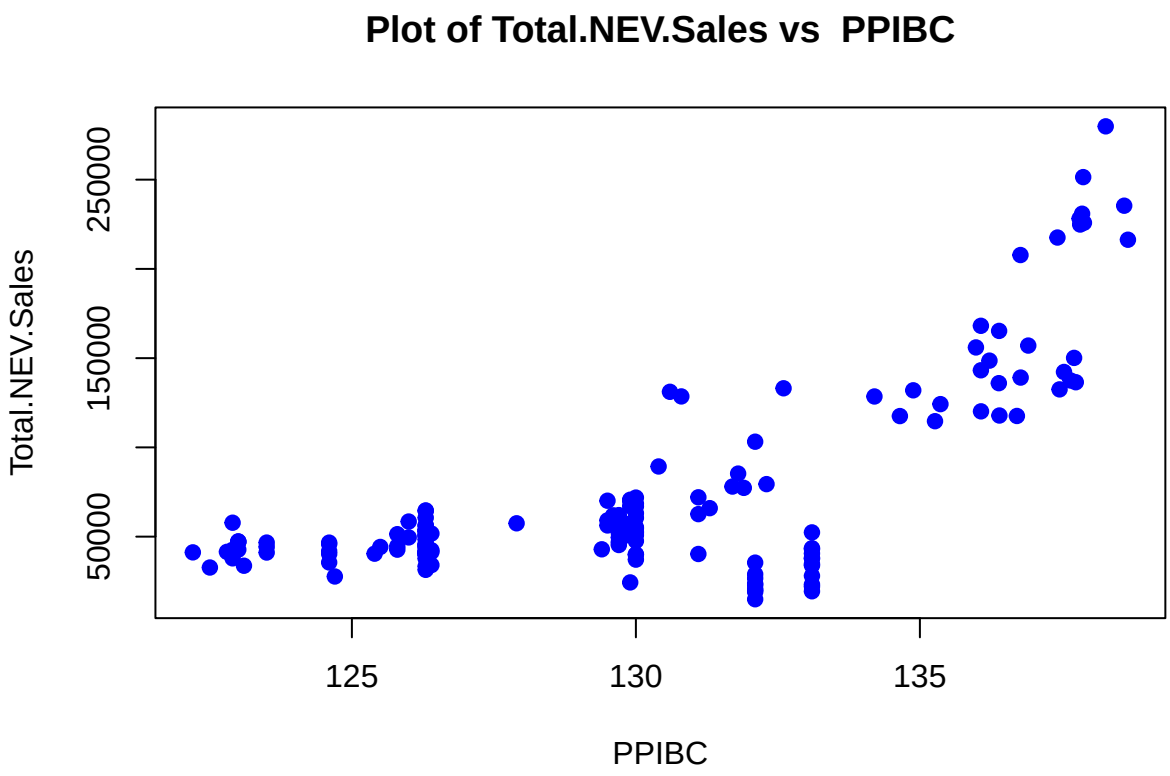

Plot of Total.NEV.Sales vs Gasoline.LDV.sales
Total.NEV.Sales
50000
150000
250000
800000
1000000
1200000
1400000
1600000
Gasoline.LDV.sales

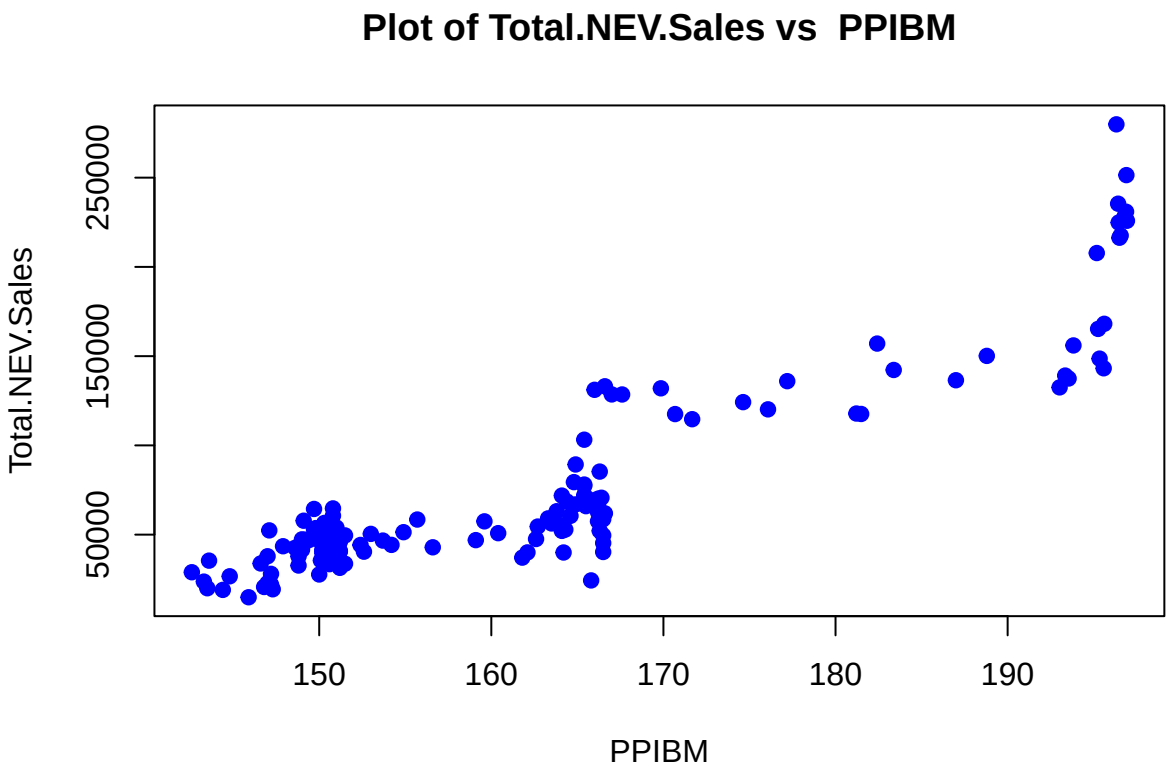

Plot of Total.NEV.Sales vs M1SL
Total.NEV.Sales
50000
150000
250000
5000
10000
15000
20000
M1SL

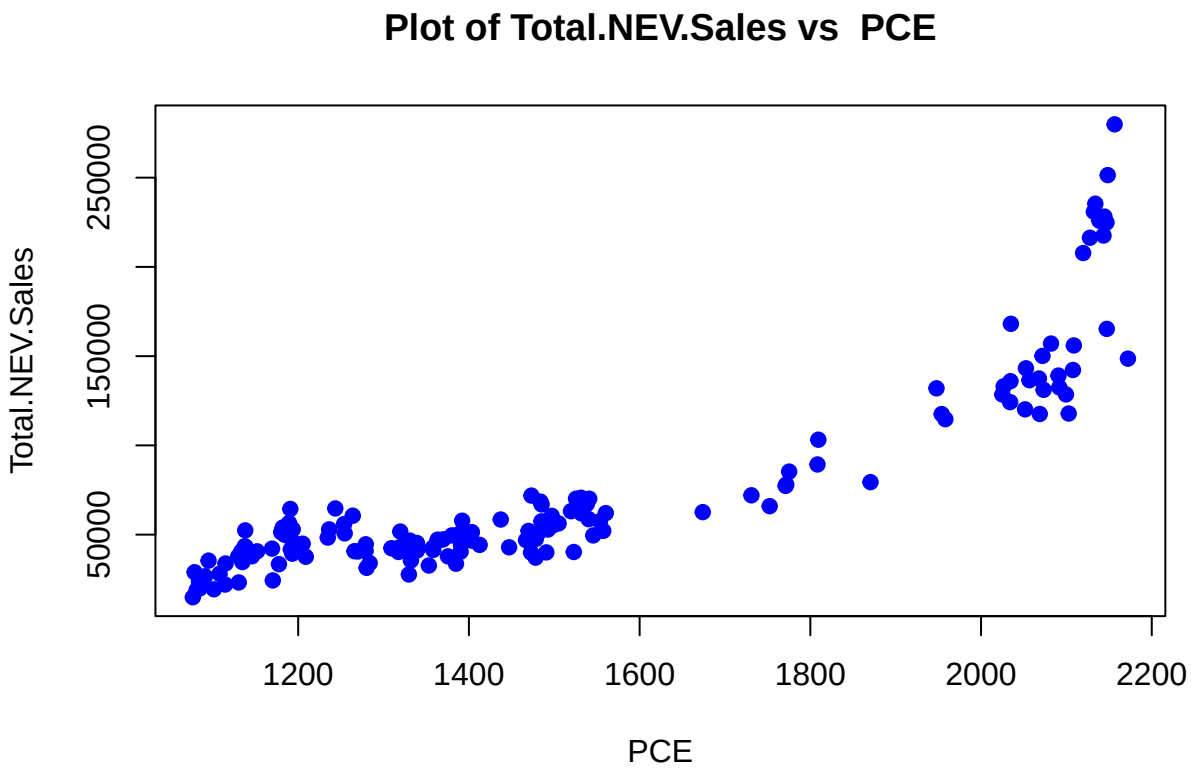

Plot of Total.NEV.Sales vs M2SL
Total.NEV.Sales
50000
150000
250000
10000
12000
14000
16000
18000
20000
22000
M2SL

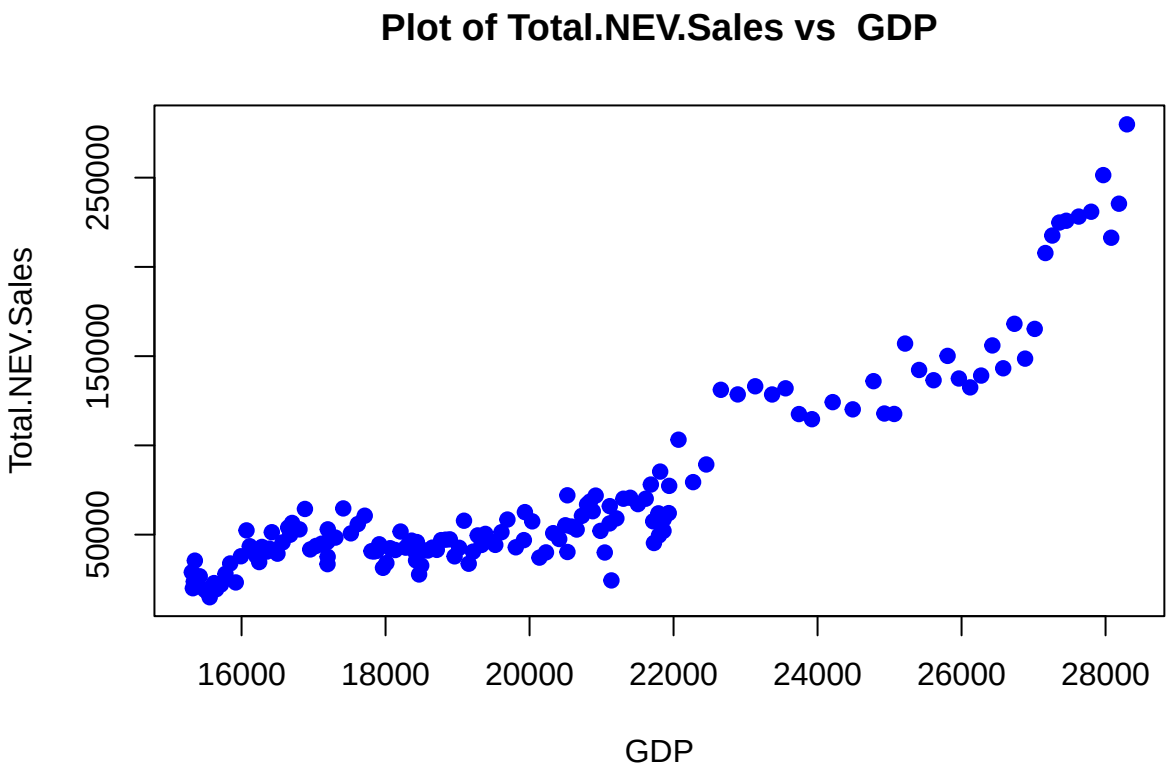

Plot of Total.NEV.Sales vs Federal.fund.effective.rate
Total.NEV.Sales
50000
150000
250000
0.00
0.01
0.02
0.03
0.04
0.05
Federal.fund.effective.rate

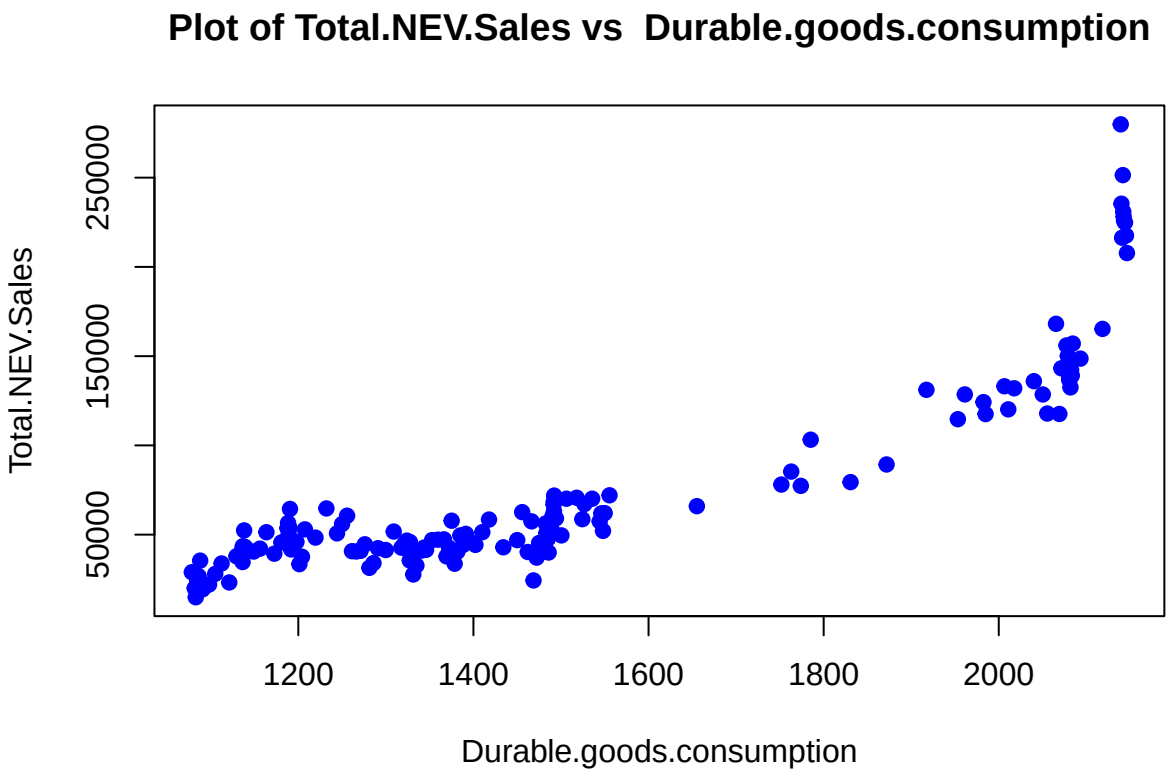

Plot of Total.NEV.Sales vs Umemployment.rate
Total.NEV.Sales
50000
150000
250000
0.04
0.06
0.08
0.10
0.12
0.14
Umemployment.rate

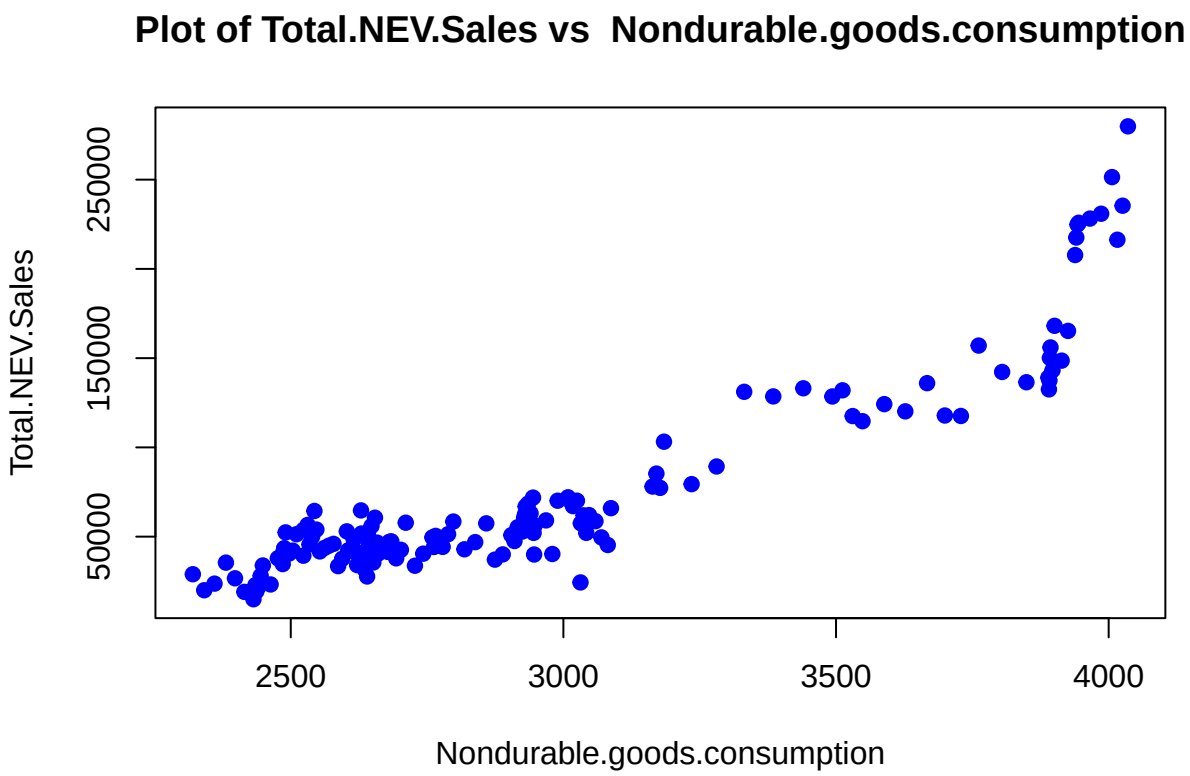

Plot of Total.NEV.Sales vs PPIBC
Total.NEV.Sales
50000
150000
250000
125
130
135
PPIBC

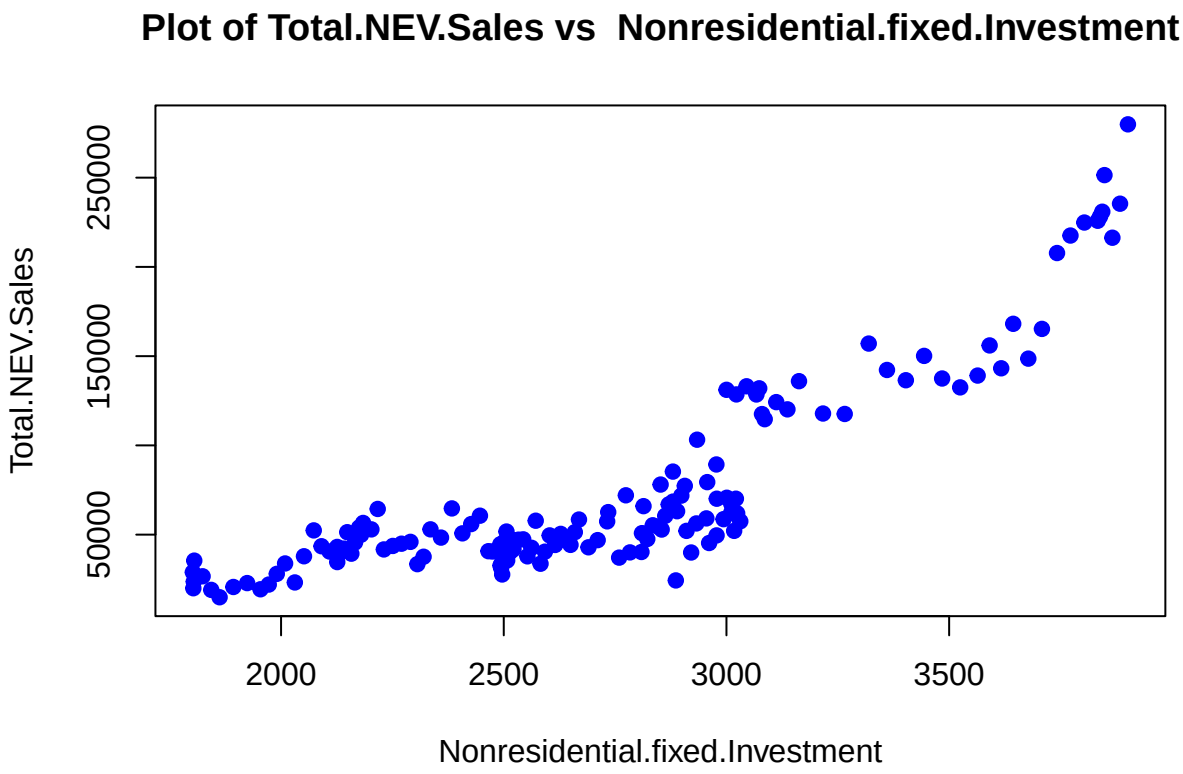

Plot of Total.NEV.Sales vs PPIBM
Total.NEV.Sales
50000
150000
250000
150
160
170
180
190
PPIBM

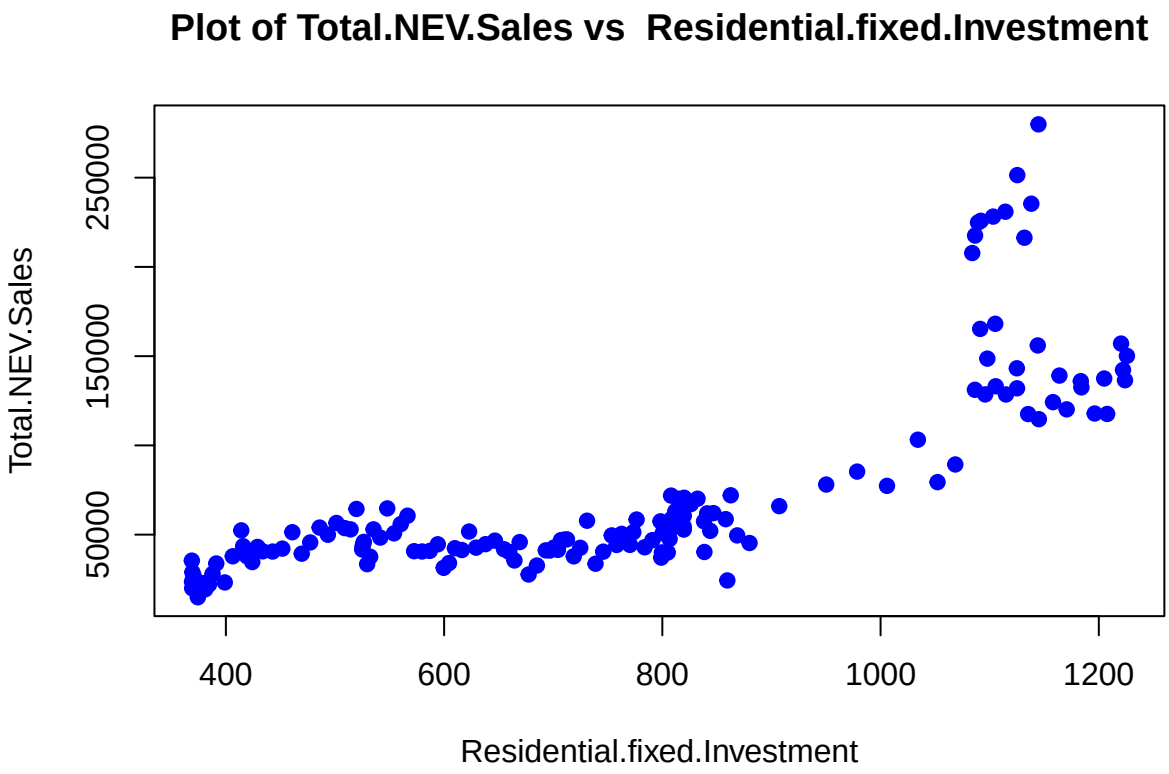

Plot of Total.NEV.Sales vs PCE
Total.NEV.Sales
50000
150000
250000
1200
1400
1600
1800
2000
2200
PCE

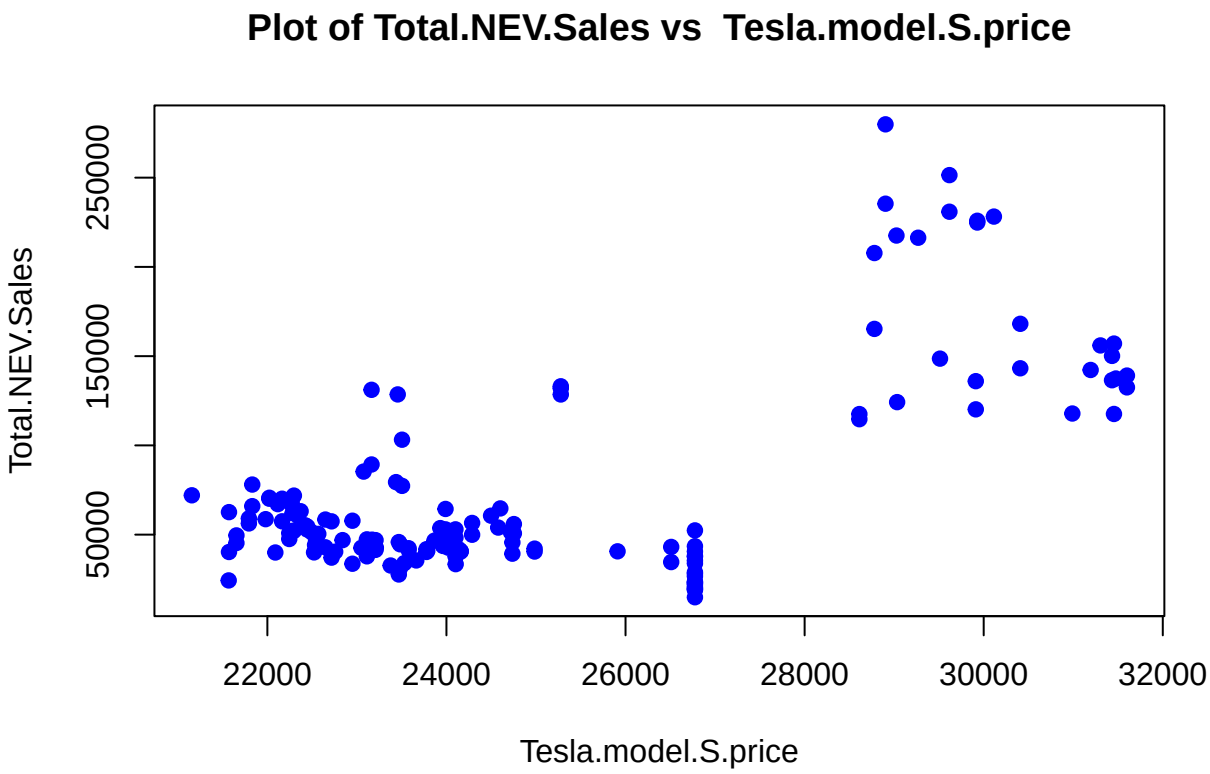

Plot of Total.NEV.Sales vs GDP
Total.NEV.Sales
50000
150000
250000
16000
18000
20000
22000
24000
26000
28000
GDP

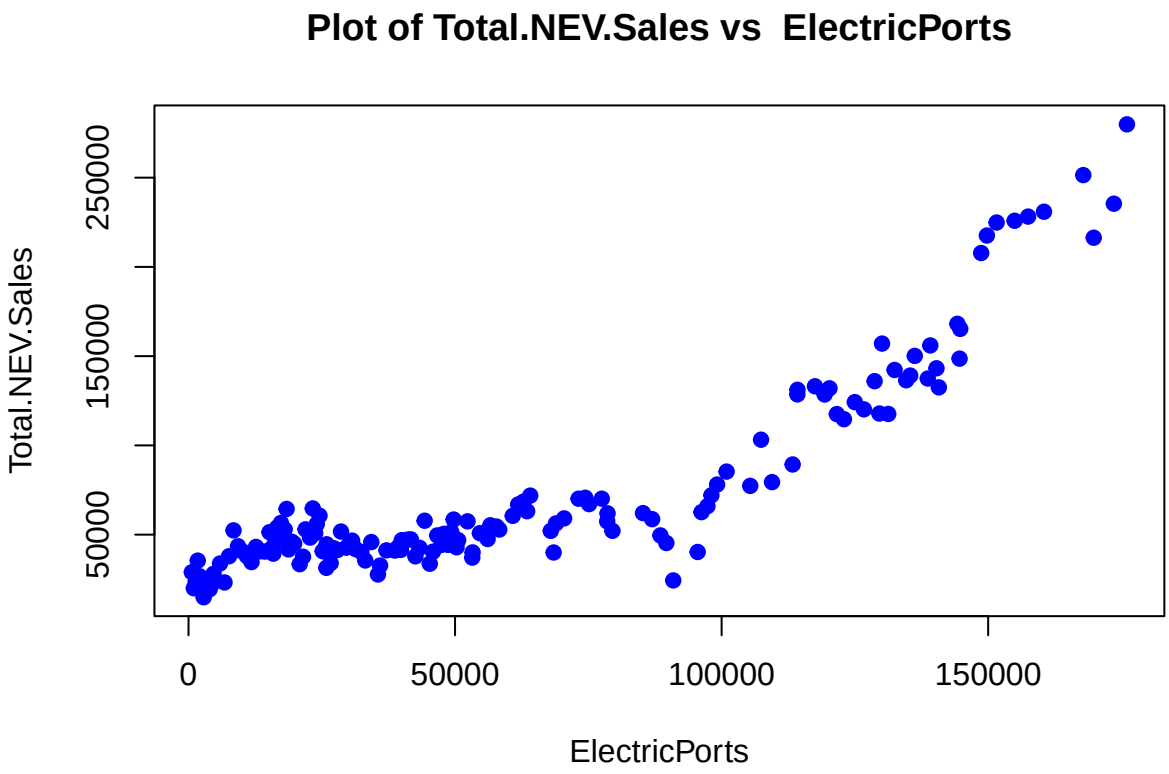

Plot of Total.NEV.Sales vs Durable.goods.consumption
Total.NEV.Sales
50000
150000
250000
1200
1400
1600
1800
2000
Durable.goods.consumption

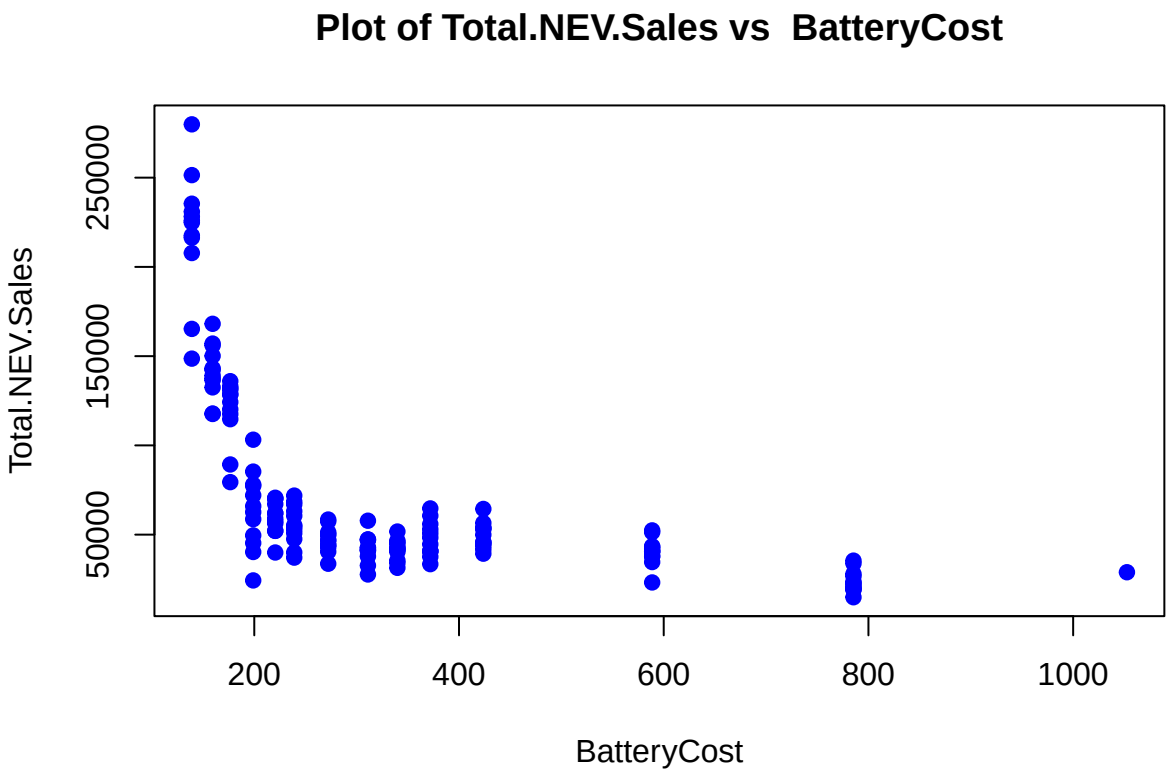

Plot of Total.NEV.Sales vs Nondurable.goods.consumption
Total.NEV.Sales
50000
150000
250000
2500
3000
3500
4000
Nondurable.goods.consumption

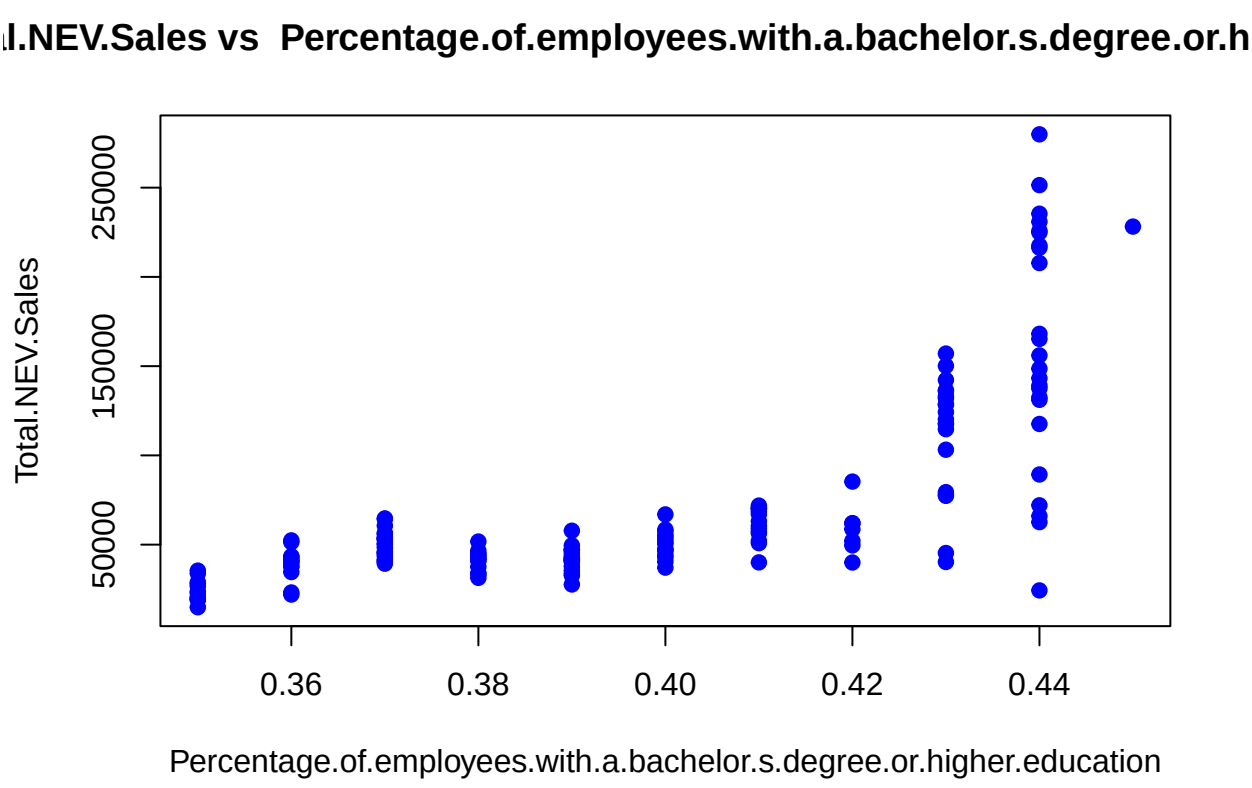

Plot of Total.NEV.Sales vs Nonresidential.fixed.Investment
Total.NEV.Sales
50000
150000
250000
2000
2500
3000
3500
Nonresidential.fixed.Investment

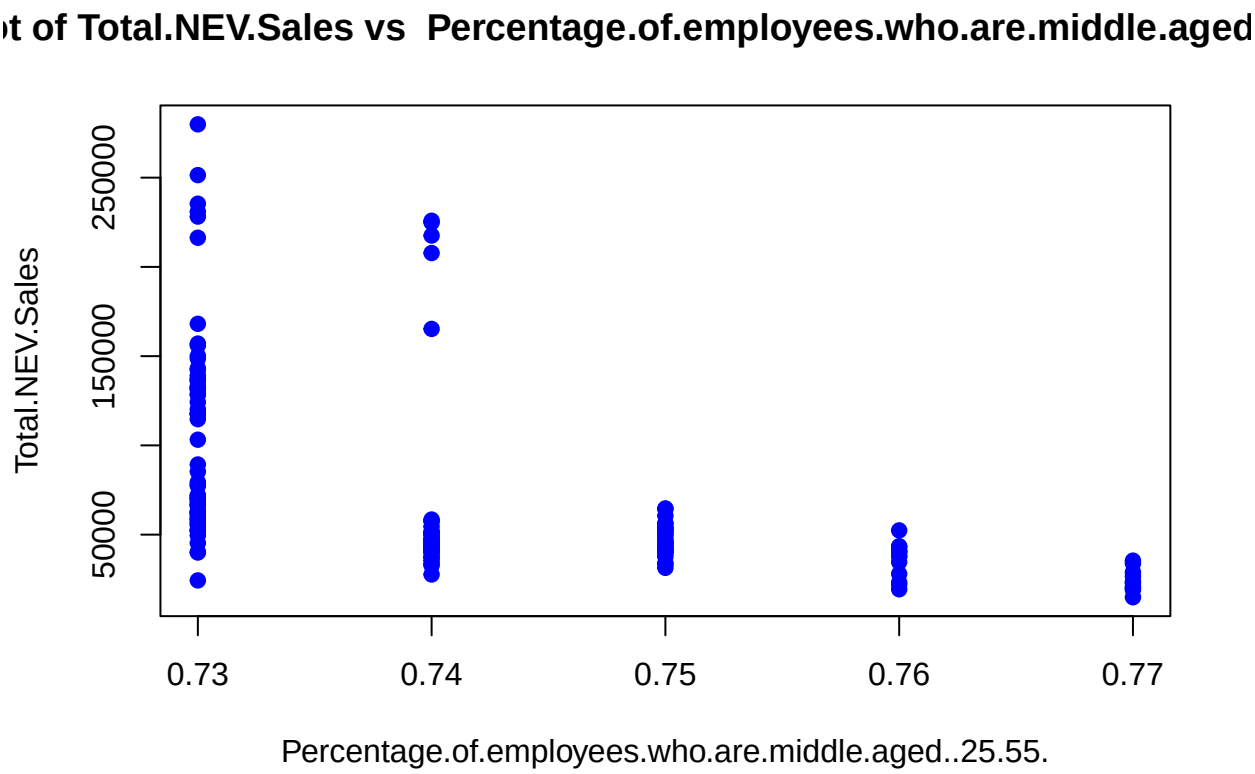

Plot of Total.NEV.Sales vs Residential.fixed.Investment
Total.NEV.Sales
50000
150000
250000
400
600
800
1000
1200
Residential.fixed.Investment

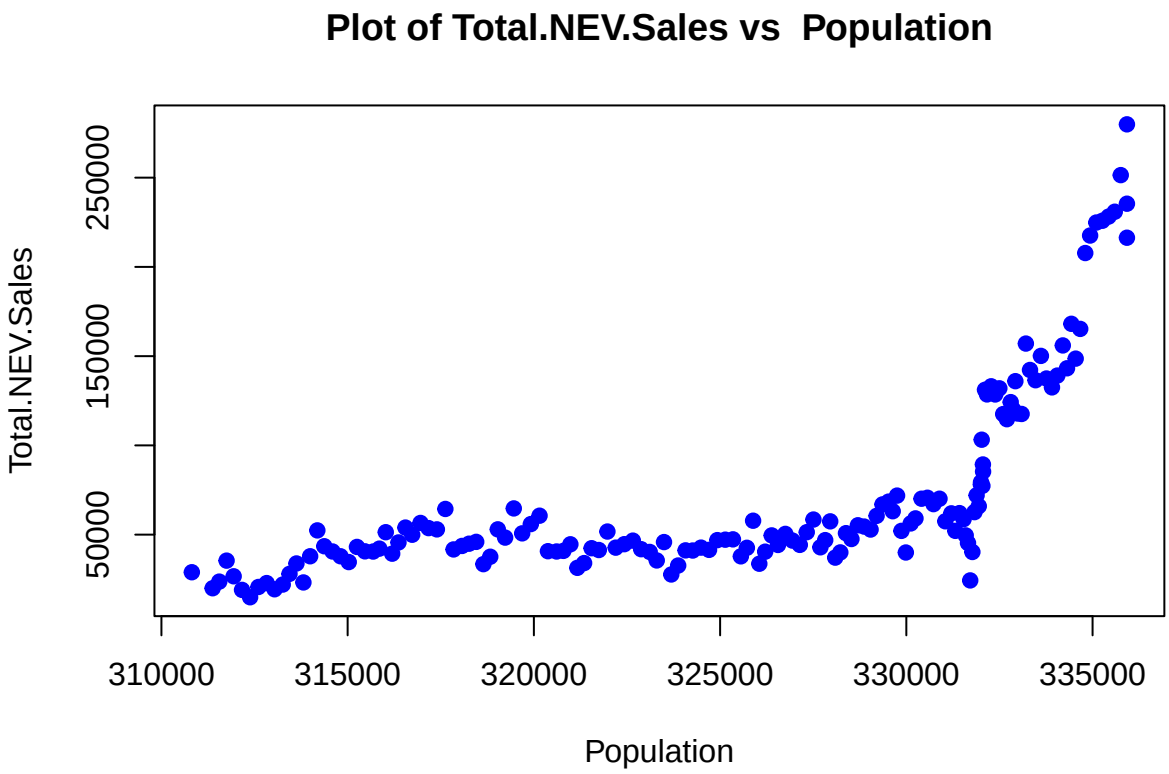

Plot of Total.NEV.Sales vs Tesla.model.S.price
Total.NEV.Sales
50000
150000
250000
22000
24000
26000
28000
30000
32000
Tesla.model.S.price

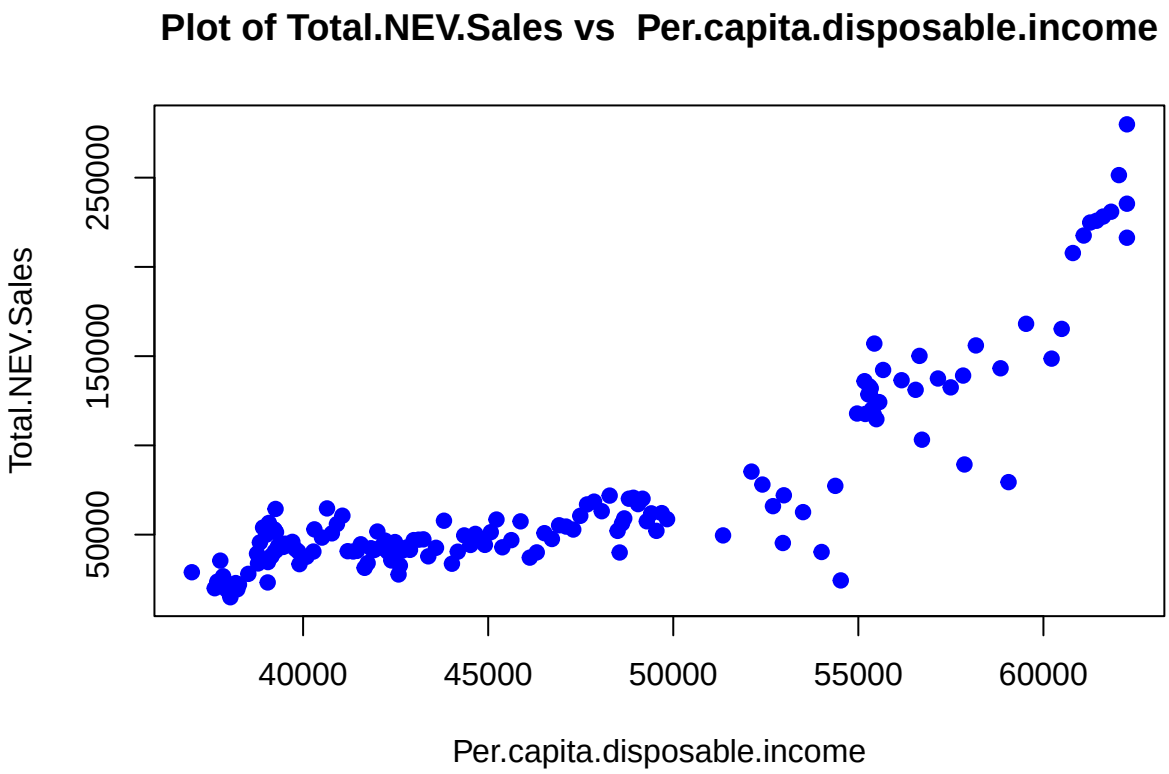

Plot of Total.NEV.Sales vs ElectricPorts
Total.NEV.Sales
50000
150000
250000
0
50000
100000
150000
ElectricPorts

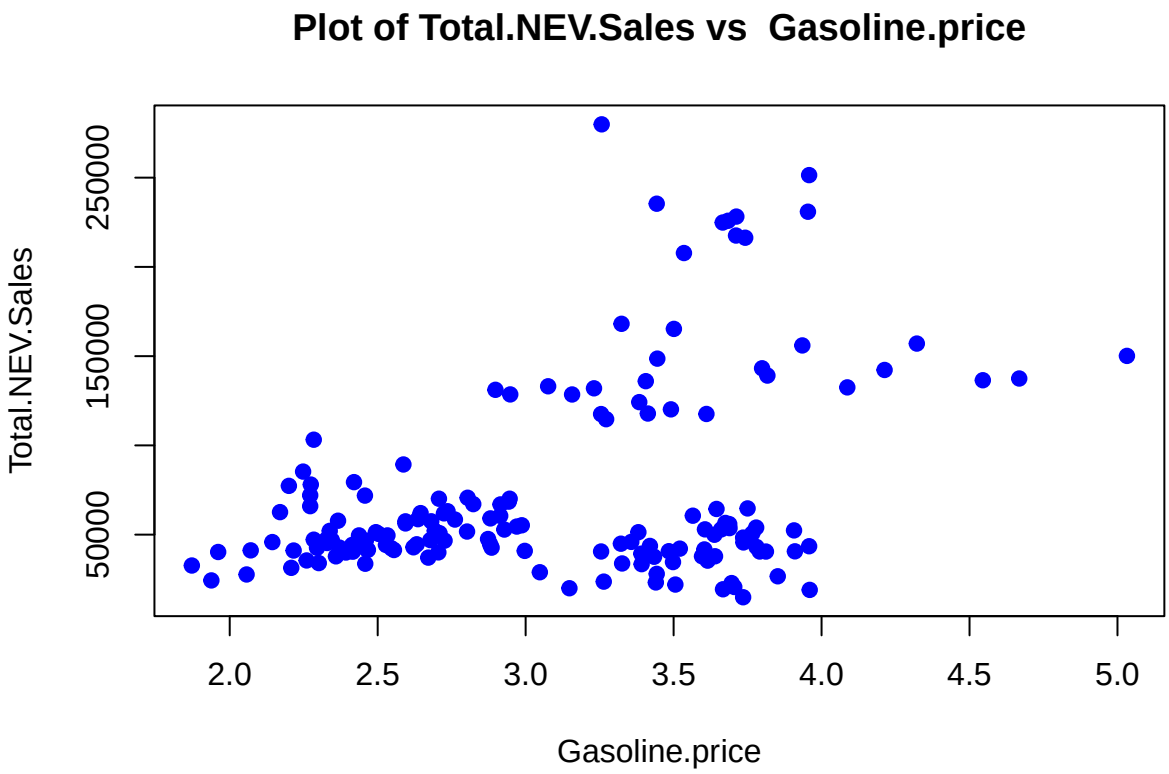

Plot of Total.NEV.Sales vs BatteryCost
Total.NEV.Sales
50000
150000
250000
200
400
600
800
1000
BatteryCost

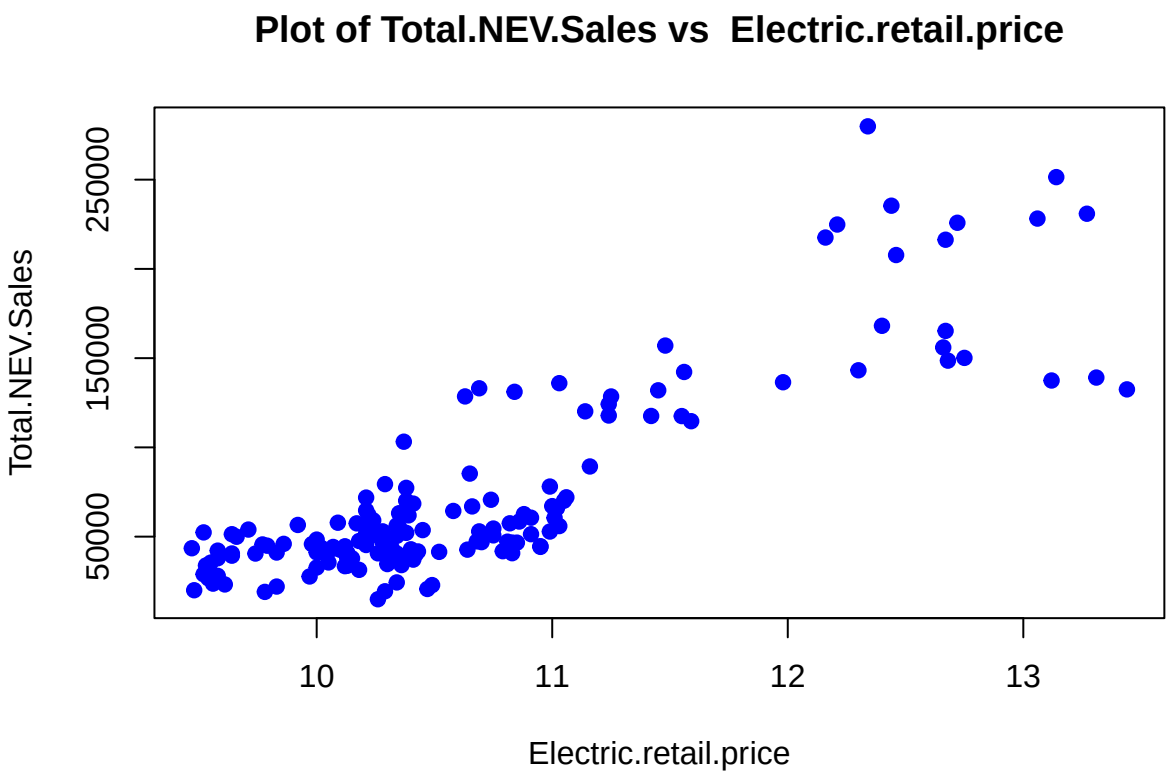

l.NEV.Sales vs Percentage.of.employees.with.a.bachelor.s.degree.or.h
Total.NEV.Sales
50000
150000
250000
0.36
0.38
0.40
0.42
0.44
Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education


t of Total.NEV.Sales vs Percentage.of.employees.who.are.middle.aged
Total.NEV.Sales
50000
150000
250000
0.73
0.74
0.75
0.76
0.77
Percentage.of.employees.who.are.middle.aged..25.55.


Plot of Total.NEV.Sales vs Population
Total.NEV.Sales
50000
150000
250000
310000
315000
320000
325000
330000
335000
Population


Plot of Total.NEV.Sales vs Per.capita.disposable.income
Total.NEV.Sales
50000
150000
250000
40000
45000
50000
55000
60000
Per.capita.disposable.income


Plot of Total.NEV.Sales vs Gasoline.price
Total.NEV.Sales
50000
150000
250000
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Gasoline.price




```
# cor_matrix <- cor(data[5:22])
# print(cor_matrix)
```

Trend and Seasonlity

```
trend <- lm(data$Total.NEV.Sales~data$Date+as.factor(data$Month))
summary(trend)
```

```
##
## Call:
## lm(formula = data$Total.NEV.Sales ~ data$Date + as.factor(data$Month))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -79738 -25424  -1119  18530 122044
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.262e+06  8.664e+04 -14.565   <2e-16 ***
## data$Date                 3.077e+01  2.013e+00  15.284   <2e-16 ***
## as.factor(data$Month)2    3.076e+03  1.362e+04   0.226   0.8217
```

```
## as.factor(data$Month)3   1.923e+04  1.362e+04   1.412   0.1601
## as.factor(data$Month)4   1.356e+04  1.362e+04   0.996   0.3211
## as.factor(data$Month)5   1.809e+04  1.362e+04   1.328   0.1862
## as.factor(data$Month)6   1.675e+04  1.362e+04   1.229   0.2209
## as.factor(data$Month)7   1.651e+04  1.362e+04   1.212   0.2276
## as.factor(data$Month)8   1.741e+04  1.363e+04   1.278   0.2034
## as.factor(data$Month)9   1.526e+04  1.363e+04   1.120   0.2647
## as.factor(data$Month)10  1.295e+04  1.363e+04   0.950   0.3437
## as.factor(data$Month)11  1.143e+04  1.363e+04   0.838   0.4032
## as.factor(data$Month)12  2.607e+04  1.338e+04   1.949   0.0533 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34720 on 144 degrees of freedom
## Multiple R-squared:  0.6273, Adjusted R-squared:  0.5963
## F-statistic:  20.2 on 12 and 144 DF,  p-value: < 2.2e-16
```

Outlier Detection

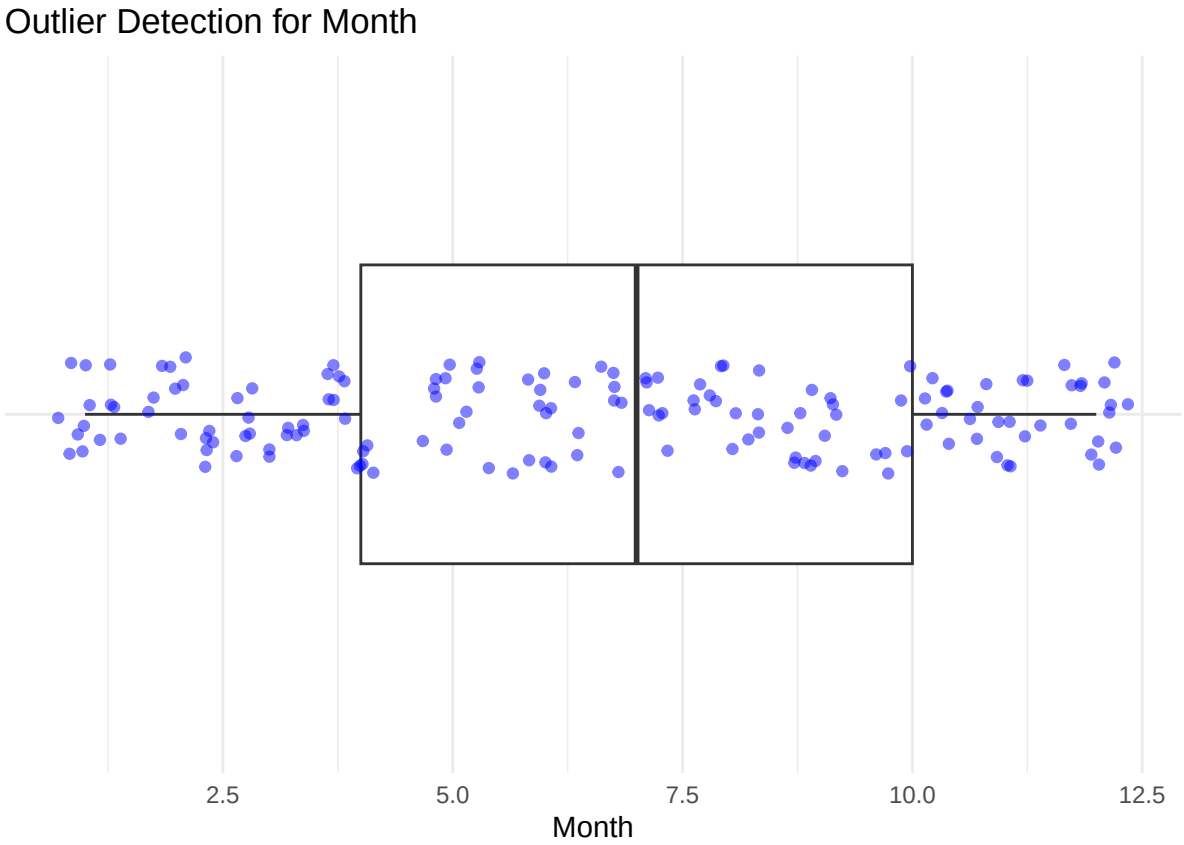
```r
detect_multiple_outliers <- function(data, start_col = 2, end_col = 29) {
  # Store original options
  original_options <- options()

  # Set options for better display
  options(tibble.width = Inf)
  options(tibble.print_max = Inf)
  options(width = 150)

  all_results <- list()

  for (i in start_col:end_col) {
    if (!is.numeric(data[[i]])) {
      cat("Skipping column", names(data)[i], "as it's not numeric\n")
      next
    }

    x <- data[[i]]
    var_name <- names(data)[i]

    results <- list()

    # Z-score method
    z_scores <- scale(x)
    results$zscore_indices <- which(abs(z_scores) > 3)
```

```r
    results$zscore_values <- x[results$zscore_indices]

    # IQR method
    Q1 <- quantile(x, 0.25, na.rm = TRUE)
    Q3 <- quantile(x, 0.75, na.rm = TRUE)
    IQR <- Q3 - Q1
    lower_bound <- Q1 - 1.5 * IQR
    upper_bound <- Q3 + 1.5 * IQR
    results$iqr_indices <- which(x < lower_bound | x > upper_bound)
    results$iqr_values <- x[results$iqr_indices]
    results$iqr_bounds <- c(lower_bound = lower_bound, upper_bound = upper_bound)

    # Modified Z-score method
    median_x <- median(x, na.rm = TRUE)
    mad_x <- mad(x, na.rm = TRUE)
    modified_z_scores <- 0.6745 * (x - median_x) / mad_x
    results$modified_zscore_indices <- which(abs(modified_z_scores) > 3.5)
    results$modified_zscore_values <- x[results$modified_zscore_indices]

    # Create visualization
    df_plot <- data.frame(
      value = x,
      y = 1
    )

    p <- ggplot2::ggplot(df_plot, ggplot2::aes(y = factor(y), x = value)) +
      ggplot2::geom_boxplot(outlier.color = "red", width = 0.5) +
      ggplot2::geom_jitter(height = 0.1, alpha = 0.5, color = "blue") +
      ggplot2::labs(title = paste("Outlier Detection for", var_name),
                    x = var_name,
                    y = "") +
      ggplot2::theme_minimal() +
      ggplot2::theme(axis.text.y = ggplot2::element_blank(),
                     axis.ticks.y = ggplot2::element_blank())

    results$plot <- p

    # Store results for this variable
    all_results[[var_name]] <- results

    # Print section header
    cat("\n================================================================\n")
    cat("=== Variable:", var_name, "===\n")
    cat("================================================================\n")
```

```
    # Function to print outlier values
    print_outliers <- function(values, method_name) {
      if (length(values) > 0) {
        cat("\n", method_name, "\n")
        cat("Number of outliers:", length(values), "\n")
        cat("Outlier values:", paste(round(values, 3), collapse = ", "), "\n")
      }
    }

    # Print results for each method
    cat("\nOutlier Detection Results:\n")

    # Z-score outliers
    print_outliers(results$zscore_values, "Z-score method (|z| > 3)")

    # IQR outliers
    cat("\nIQR method\n")
    cat("IQR bounds: Lower =", round(lower_bound, 3), ", Upper =", round(upper_bound, 3)
    print_outliers(results$iqr_values, "IQR outliers")

    # Modified Z-score outliers
    print_outliers(results$modified_zscore_values, "Modified Z-score method (|modified z

    # Print summary statistics
    cat("\nSummary Statistics:\n")
    summary_stats <- list(
      mean = mean(x, na.rm = TRUE),
      median = median(x, na.rm = TRUE),
      sd = sd(x, na.rm = TRUE),
      Q1 = Q1,
      Q3 = Q3
    )
    print(lapply(summary_stats, round, 3))

    # Display plot
    print(results$plot)
    cat("\n----------------------------------------------------------------\n")
  }

  options(original_options)

  return(all_results)
}
```

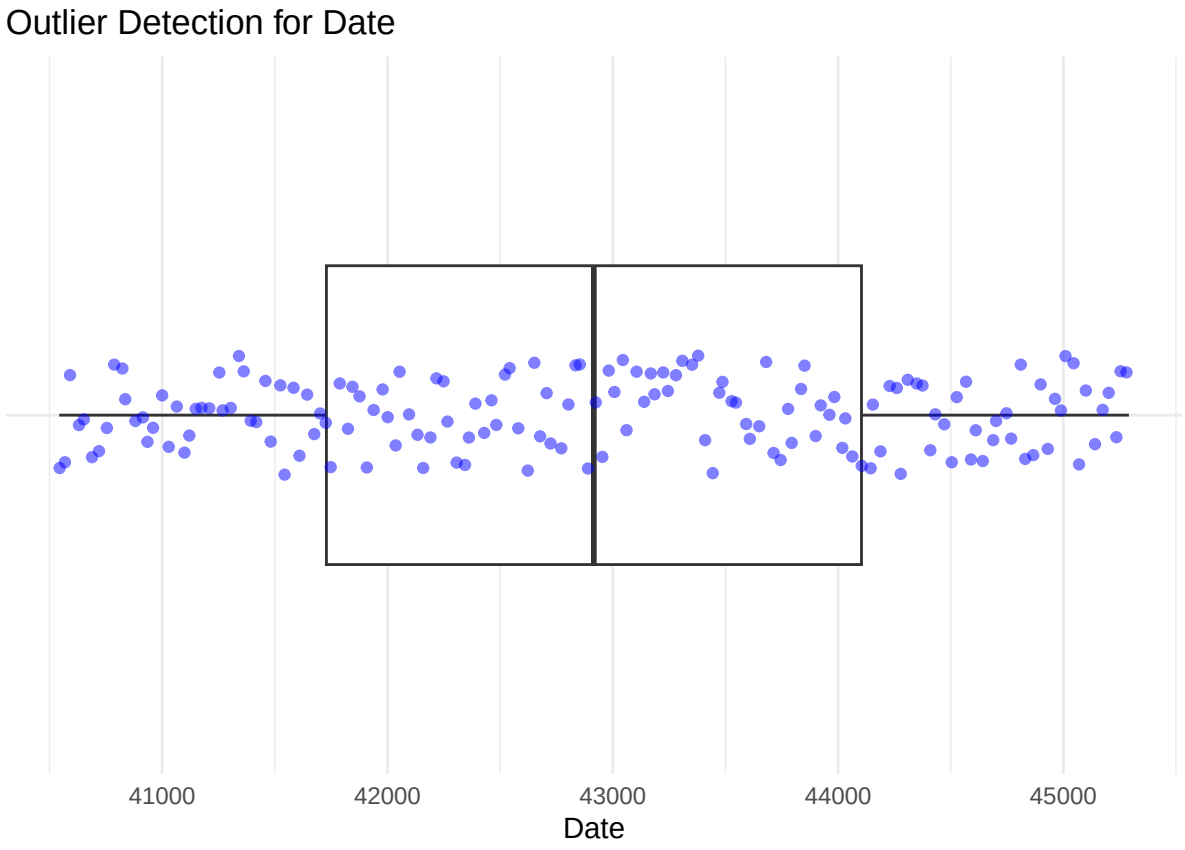
```r
results <- detect_multiple_outliers(data)
```

```
##
## ================================================================
## === Variable: Month ===
## ================================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = -5 , Upper = 19
##
## Summary Statistics:
## $mean
## [1] 6.535
##
## $median
## [1] 7
##
## $sd
## [1] 3.48
##
## $Q1
## 25%
##   4
##
## $Q3
## 75%
##  10
```



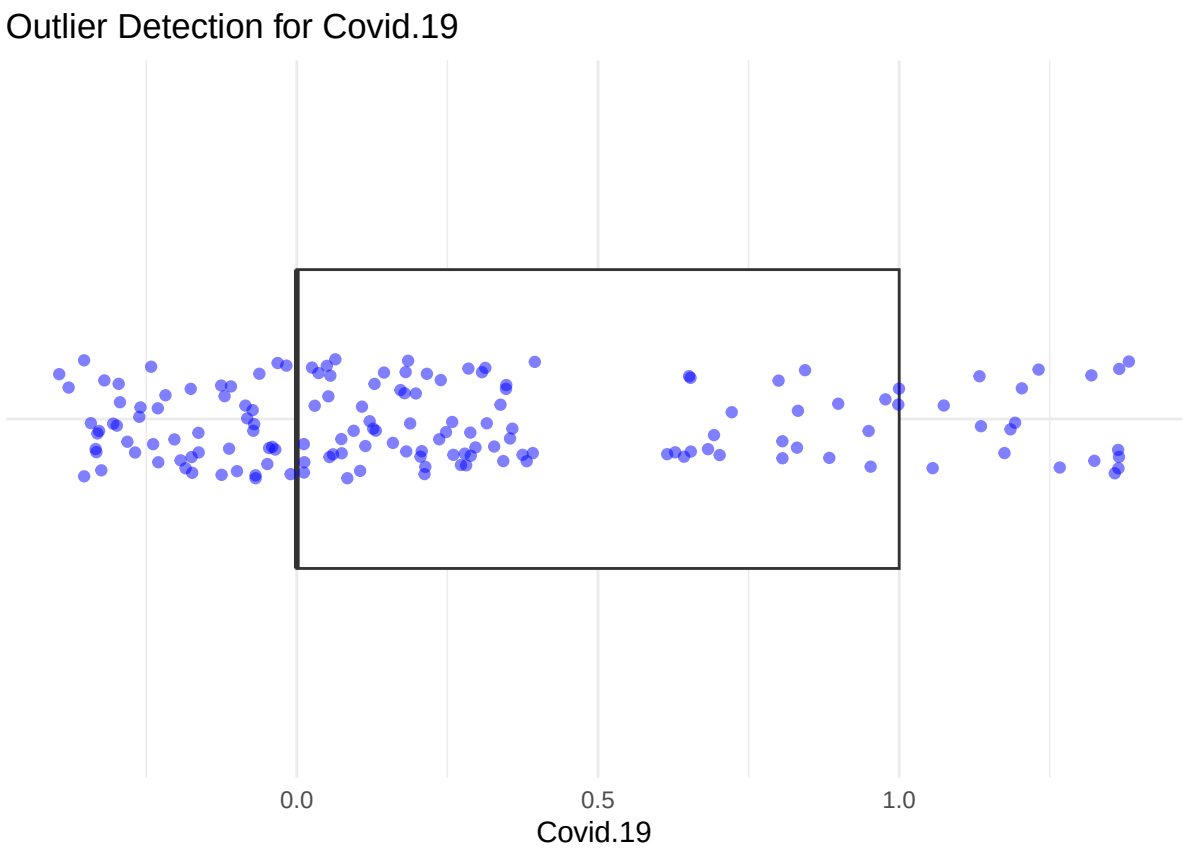

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Date ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 38166.5 , Upper = 47666.5
##
## Summary Statistics:
## $mean
## [1] 42916.53
##
## $median
## [1] 42916
##
## $sd
## [1] 1383.858
```

```
##
## $Q1
##   25%
## 41729
##
## $Q3
##   75%
## 44104
```



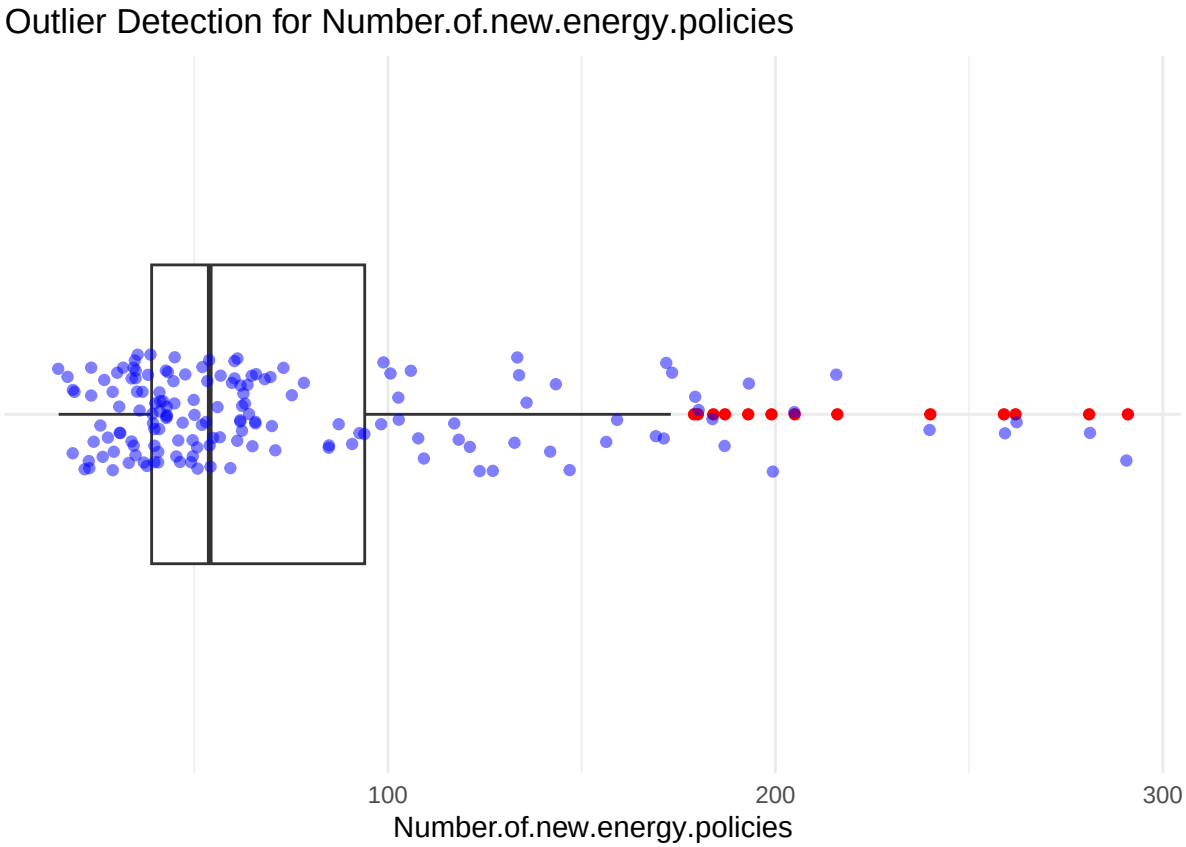

```
##
## ----------------------------------------------------------------
##
## ================================================================
## === Variable: Covid.19 ===
## ================================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = -1.5 , Upper = 2.5
##
##  Modified Z-score method (|modified z| > 3.5)
```

```
## Number of outliers: 41
## Outlier values: 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
##
## Summary Statistics:
## $mean
## [1] 0.261
##
## $median
## [1] 0
##
## $sd
## [1] 0.441
##
## $Q1
## 25%
##   0
##
## $Q3
## 75%
##   1
```



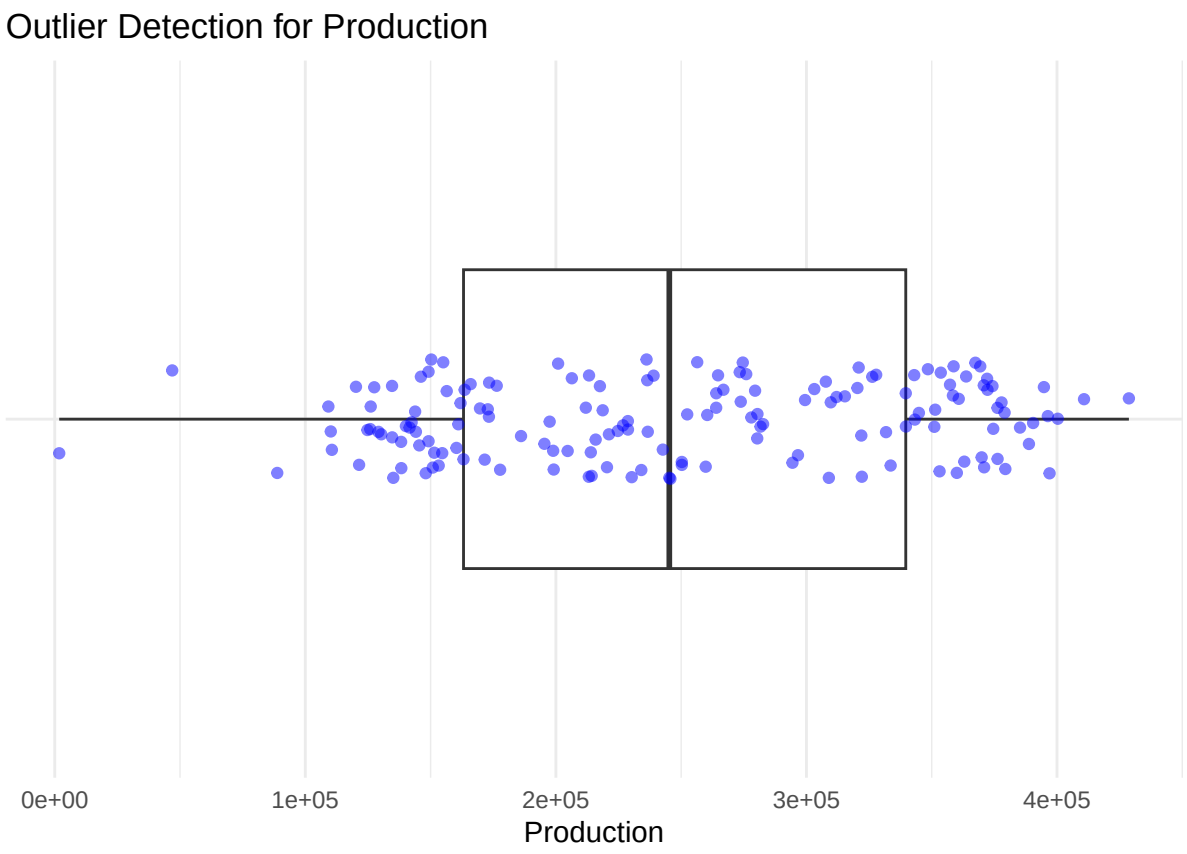

```
##
```

```
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Number.of.new.energy.policies ===
## ============================================================
##
## Outlier Detection Results:
##
##  Z-score method (|z| > 3)
## Number of outliers: 4
## Outlier values: 291, 281, 262, 259
##
## IQR method
## IQR bounds: Lower = -43.5 , Upper = 176.5
##
##  IQR outliers
## Number of outliers: 13
## Outlier values: 216, 193, 180, 199, 205, 291, 281, 262, 259, 240, 187, 184, 179
##
##  Modified Z-score method (|modified z| > 3.5)
## Number of outliers: 7
## Outlier values: 216, 205, 291, 281, 262, 259, 240
##
## Summary Statistics:
## $mean
## [1] 75.478
##
## $median
## [1] 54
##
## $sd
## [1] 57.753
##
## $Q1
## 25%
##  39
##
## $Q3
## 75%
##  94
```



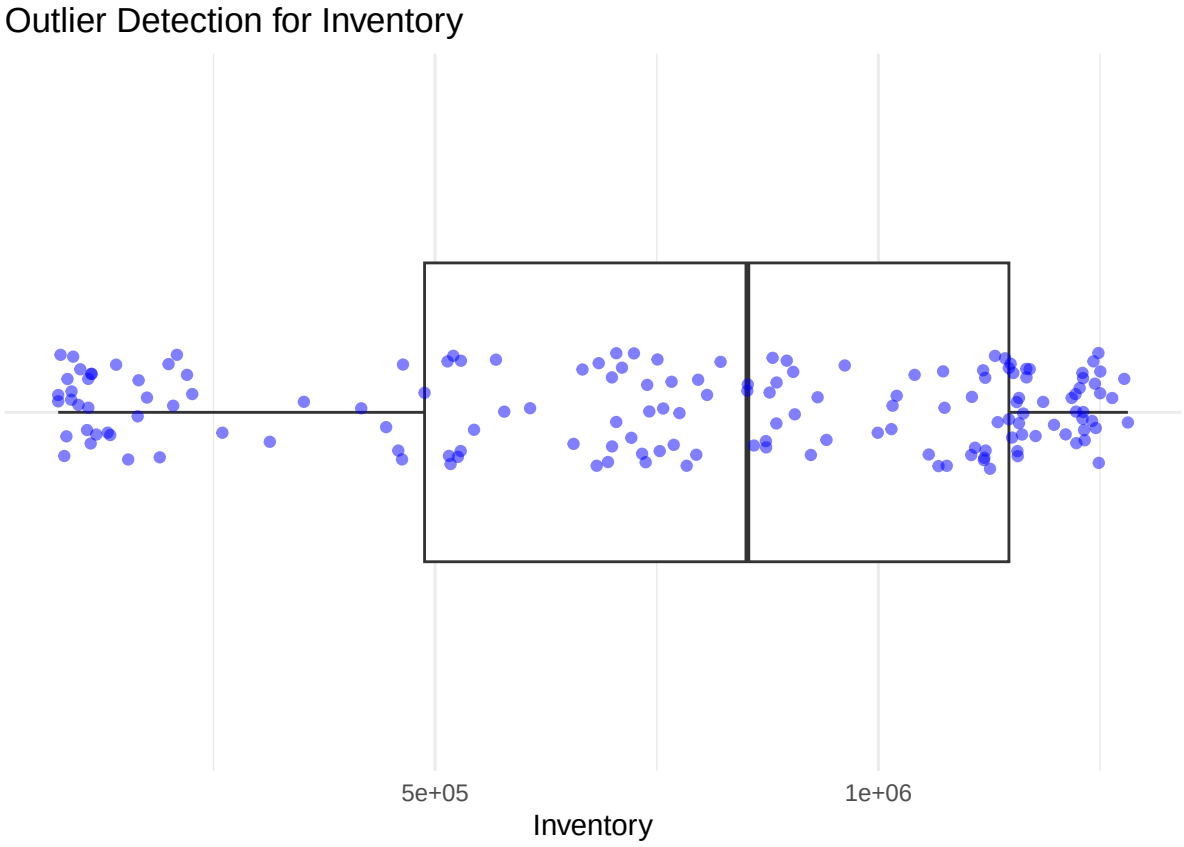

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Production ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = -101620 , Upper = 604412
##
## Summary Statistics:
## $mean
## [1] 249727.2
##
## $median
## [1] 245279
##
## $sd
## [1] 93137.35
```

```
##
## $Q1
##     25%
## 163142
##
## $Q3
##     75%
## 339650
```



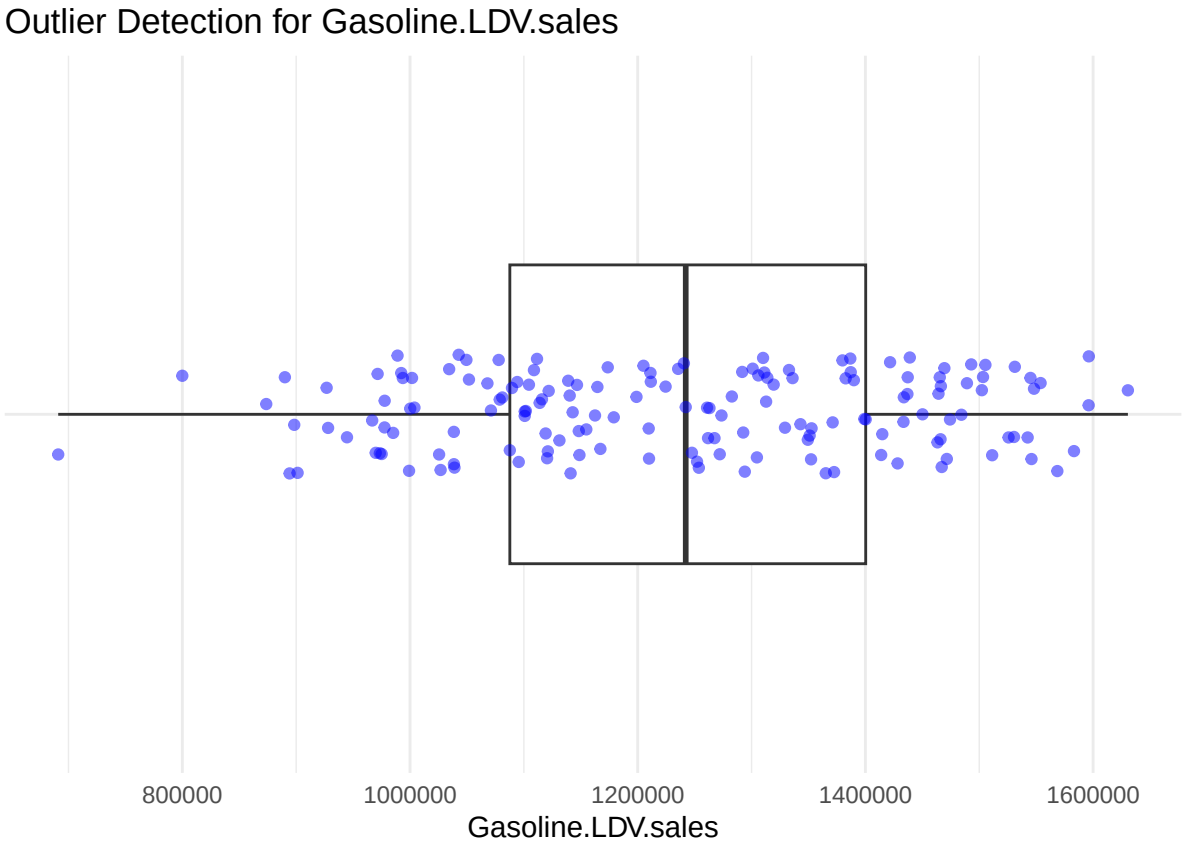

```
##
## ----------------------------------------------------------------
##
## ================================================================
## === Variable: Inventory ===
## ================================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = -500591 , Upper = 2135921
##
## Summary Statistics:
```

```
## $mean
## [1] 771654.7
##
## $median
## [1] 852200
##
## $sd
## [1] 400402.5
##
## $Q1
##    25%
## 488101
##
## $Q3
##     75%
## 1147229
```



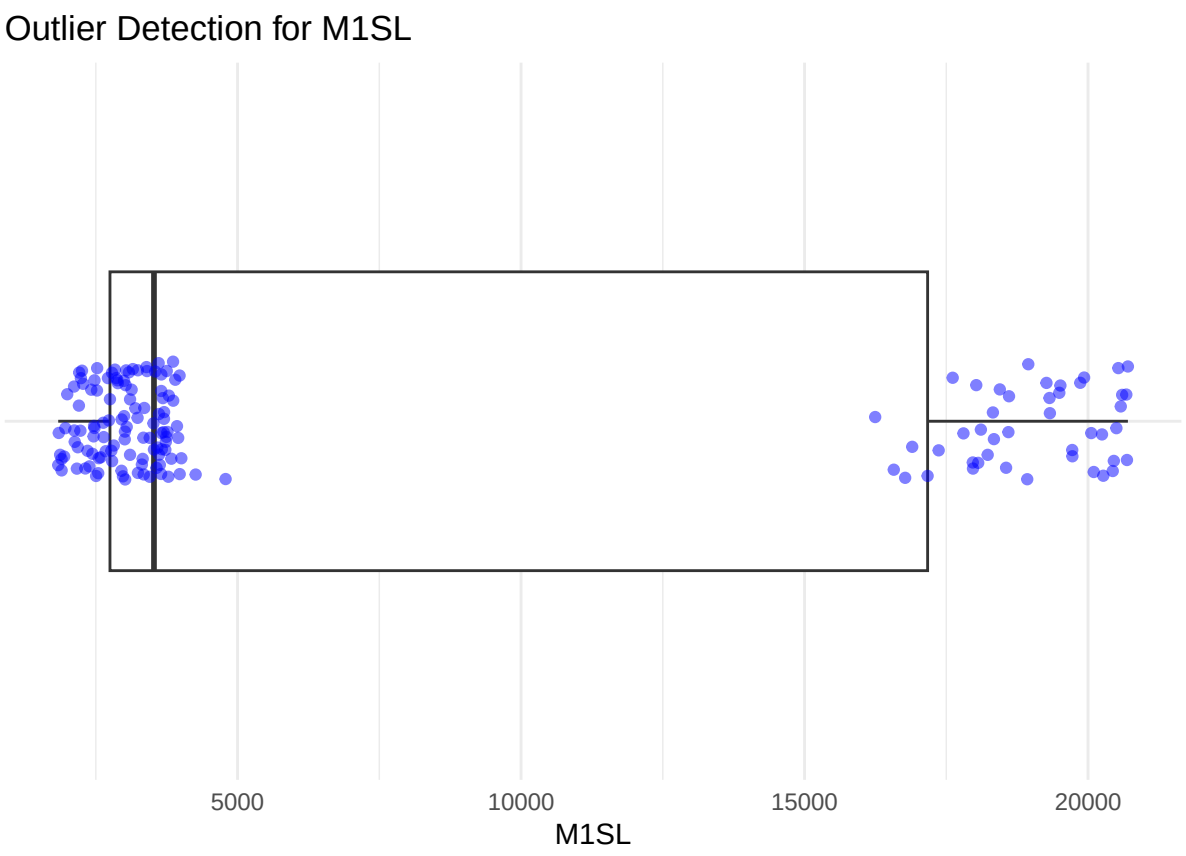

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Gasoline.LDV.sales ===
```

```
## ================================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 618790 , Upper = 1869150
##
## Summary Statistics:
## $mean
## [1] 1236707
##
## $median
## [1] 1242179
##
## $sd
## [1] 200399.2
##
## $Q1
##     25%
## 1087675
##
## $Q3
##     75%
## 1400265
```



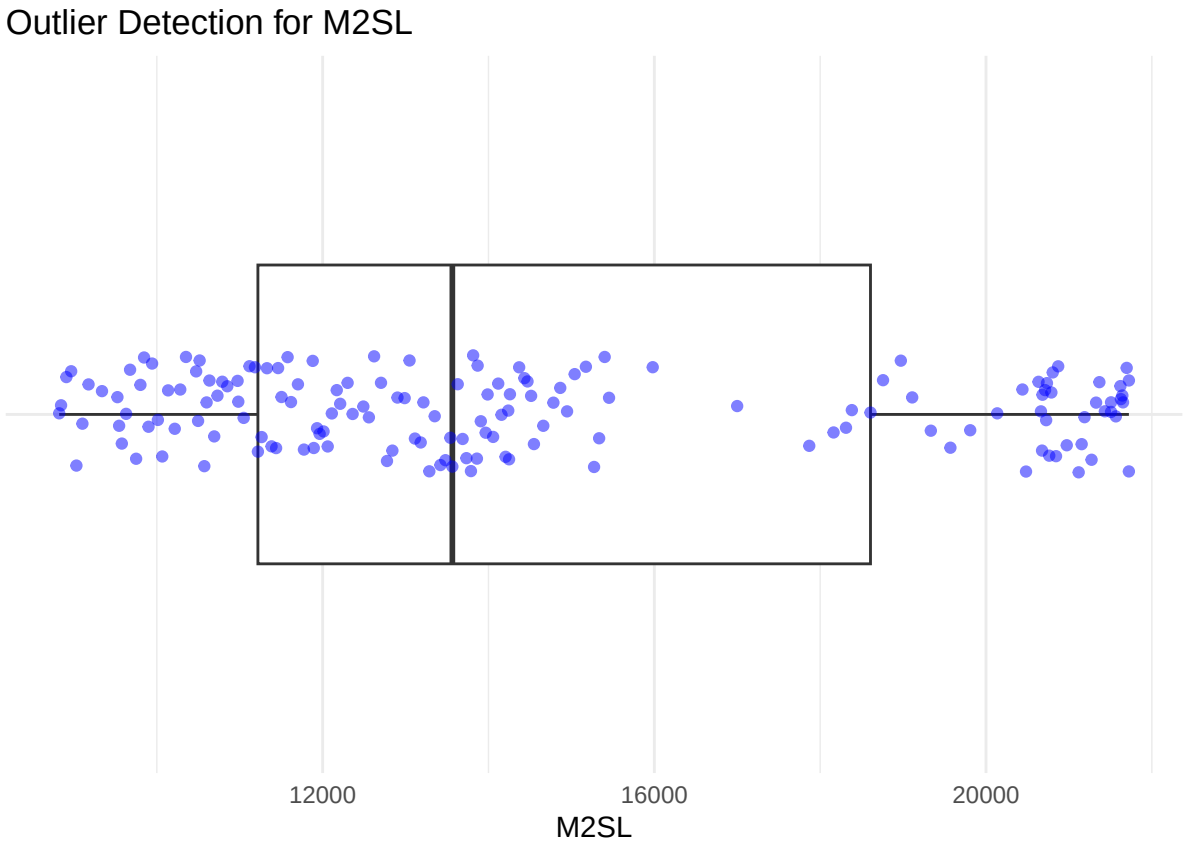

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: M1SL ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = -18882.5 , Upper = 38802.3
##
##  Modified Z-score method (|modified z| > 3.5)
## Number of outliers: 44
## Outlier values: 16245.5, 16574.1, 16774.5, 16898.8, 17170.5, 17365.7, 17612.9, 17803,
##
## Summary Statistics:
## $mean
## [1] 7506.645
##
## $median
```

```
## [1] 3525.3
##
## $sd
## [1] 7240.829
##
## $Q1
##    25%
## 2749.3
##
## $Q3
##     75%
## 17170.5
```



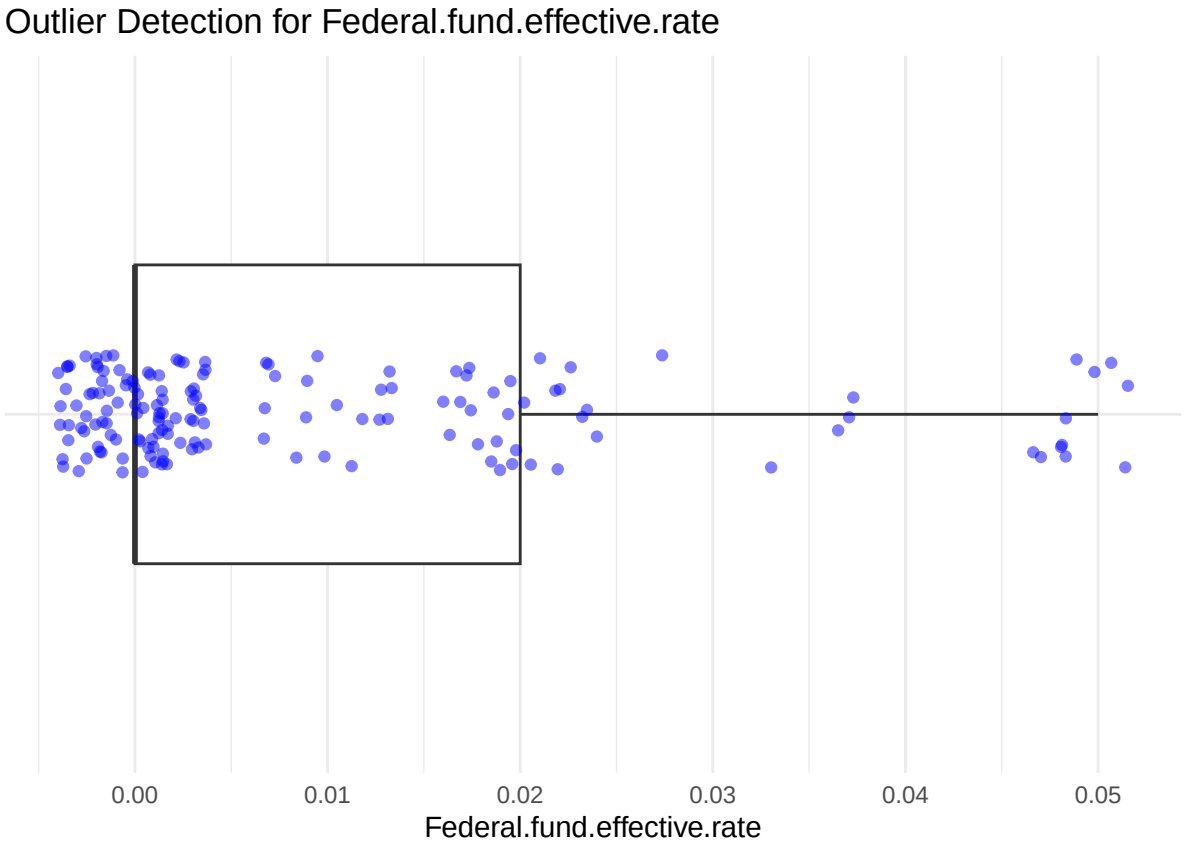

```
##
## ----------------------------------------------------------------
##
## ================================================================
## === Variable: M2SL ===
## ================================================================
##
## Outlier Detection Results:
##
```

```
## IQR method
## IQR bounds: Lower = 137.55 , Upper = 29688.35
##
## Summary Statistics:
## $mean
## [1] 14541.98
##
## $median
## [1] 13564.1
##
## $sd
## [1] 4129.859
##
## $Q1
##     25%
## 11219.1
##
## $Q3
##     75%
## 18606.8
```



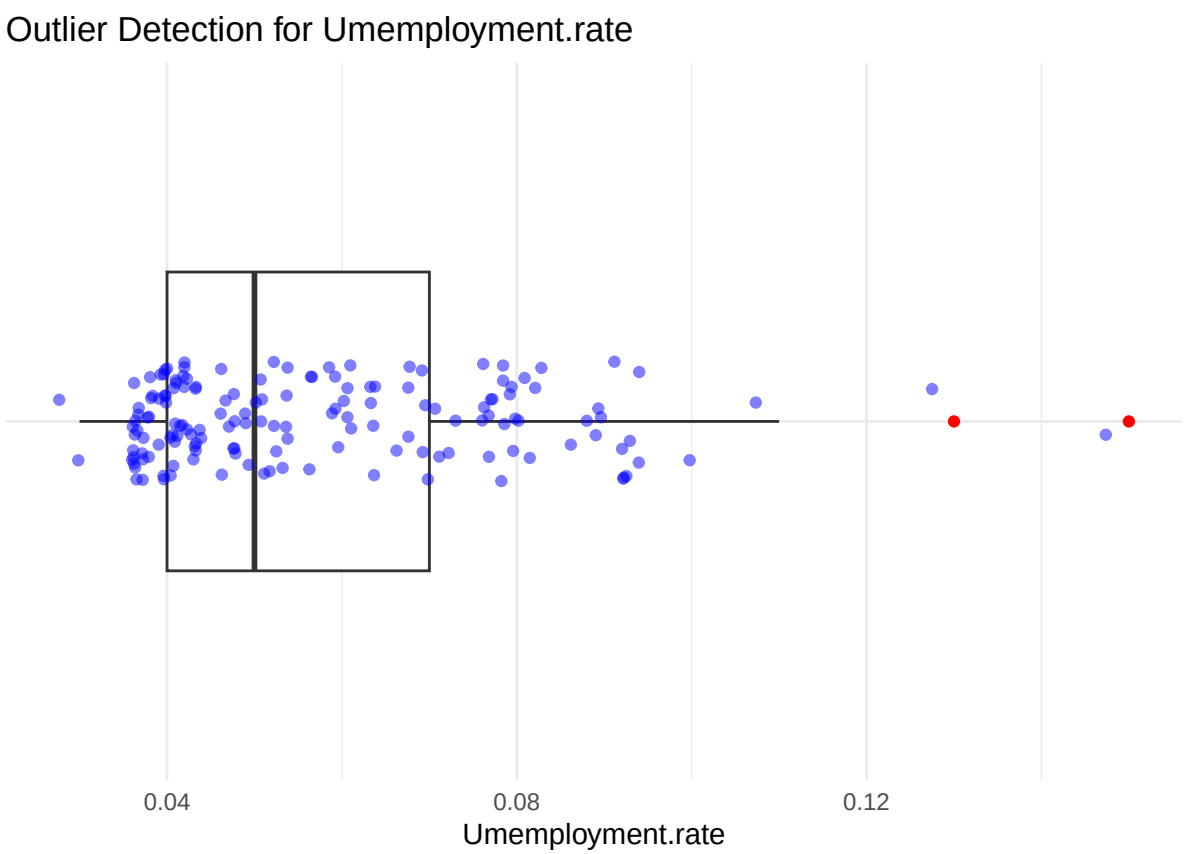

```
##
```

```
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Federal.fund.effective.rate ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = -0.03 , Upper = 0.05
##
##  Modified Z-score method (|modified z| > 3.5)
## Number of outliers: 60
## Outlier values: 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.0
##
## Summary Statistics:
## $mean
## [1] 0.009
##
## $median
## [1] 0
##
## $sd
## [1] 0.015
##
## $Q1
## 25%
##   0
##
## $Q3
##  75%
## 0.02
```



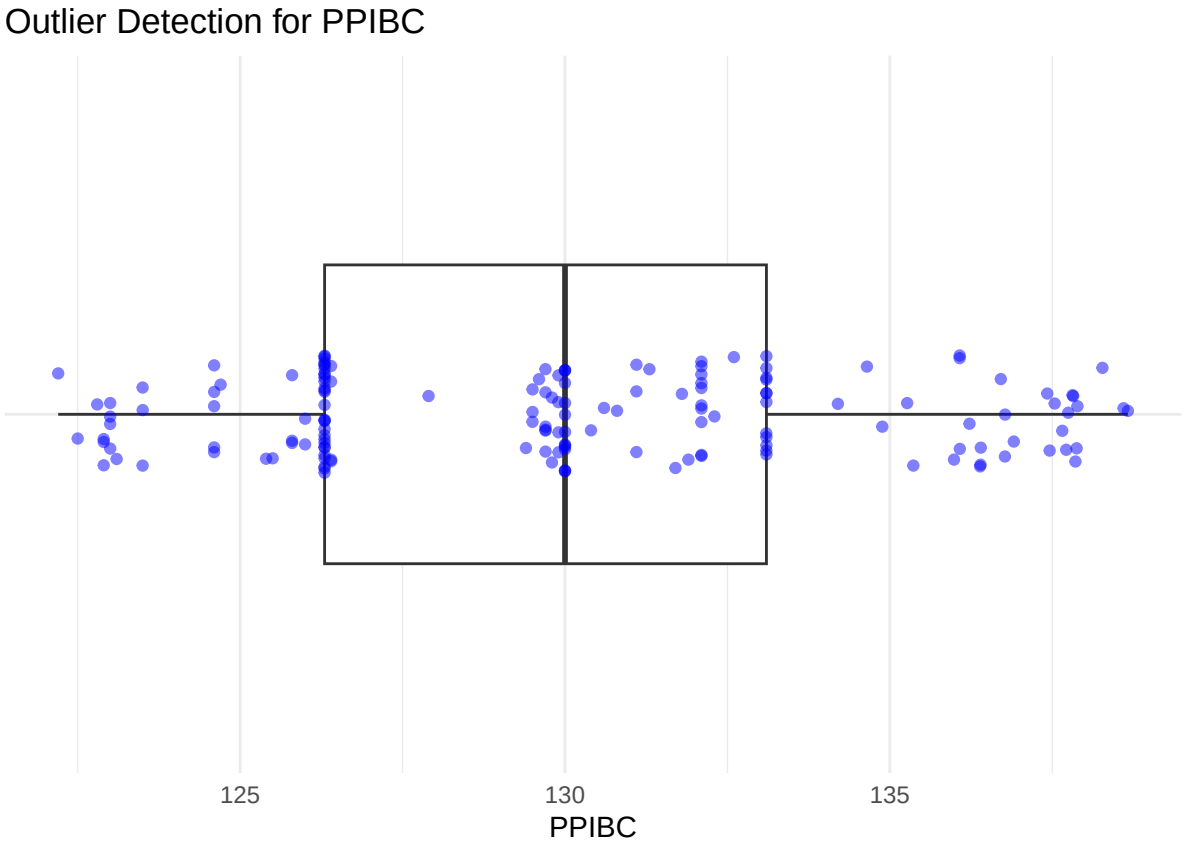

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Umemployment.rate ===
## ============================================================
##
## Outlier Detection Results:
##
##  Z-score method (|z| > 3)
## Number of outliers: 2
## Outlier values: 0.15, 0.13
##
## IQR method
## IQR bounds: Lower = -0.005 , Upper = 0.115
##
##  IQR outliers
## Number of outliers: 2
## Outlier values: 0.15, 0.13
##
##  Modified Z-score method (|modified z| > 3.5)
```

```
## Number of outliers: 2
## Outlier values: 0.15, 0.13
##
## Summary Statistics:
## $mean
## [1] 0.058
##
## $median
## [1] 0.05
##
## $sd
## [1] 0.021
##
## $Q1
##  25%
## 0.04
##
## $Q3
##  75%
## 0.07
```



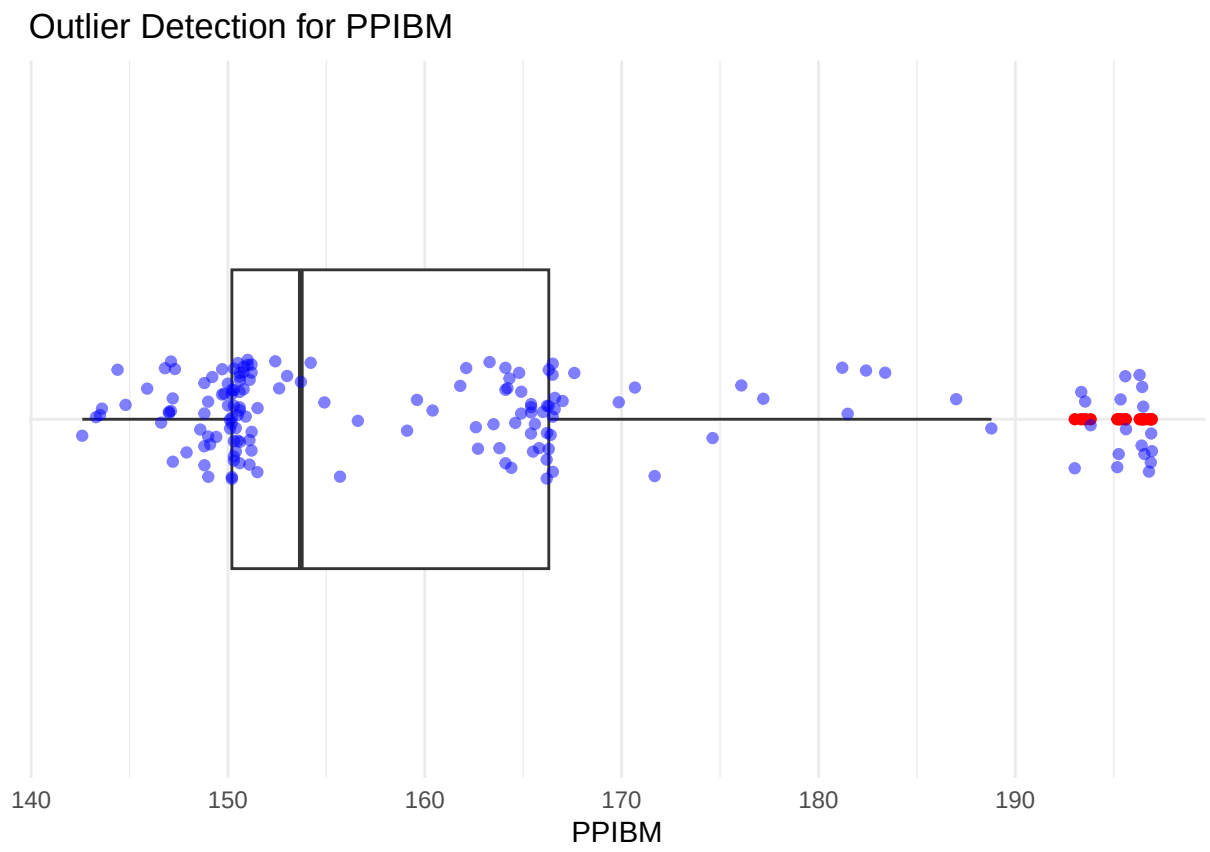

```
##
```

```
## ------------------------------------------------------------
##
## ============================================================
## === Variable: PPIBC ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 116.1 , Upper = 143.3
##
## Summary Statistics:
## $mean
## [1] 129.983
##
## $median
## [1] 130
##
## $sd
## [1] 4.47
##
## $Q1
##   25%
## 126.3
##
## $Q3
##   75%
## 133.1
```



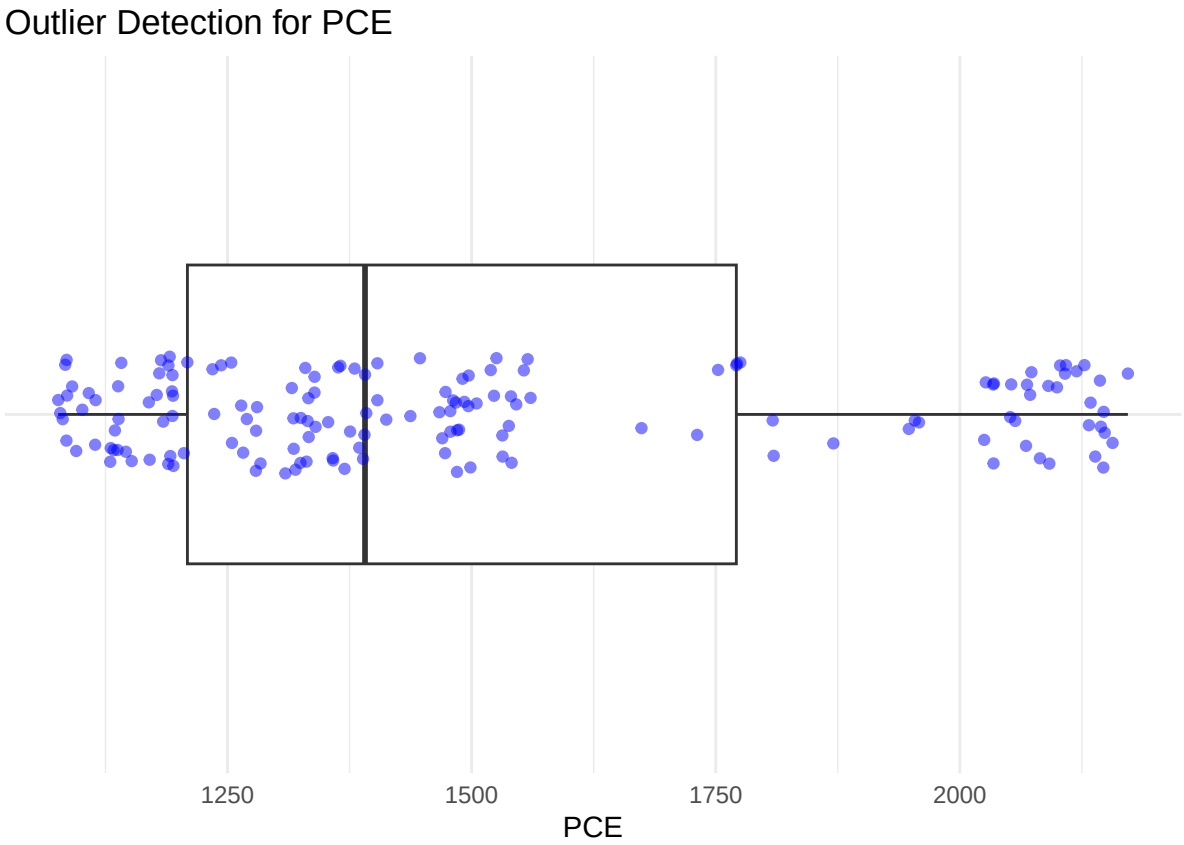

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: PPIBM ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 126.05 , Upper = 190.45
##
##  IQR outliers
## Number of outliers: 18
## Outlier values: 193.543, 193.021, 193.345, 193.825, 195.581, 195.614, 195.345, 195.25
##
## Summary Statistics:
## $mean
## [1] 161.423
##
## $median
```

```
## [1] 153.7
##
## $sd
## [1] 15.528
##
## $Q1
##   25%
## 150.2
##
## $Q3
##   75%
## 166.3
```



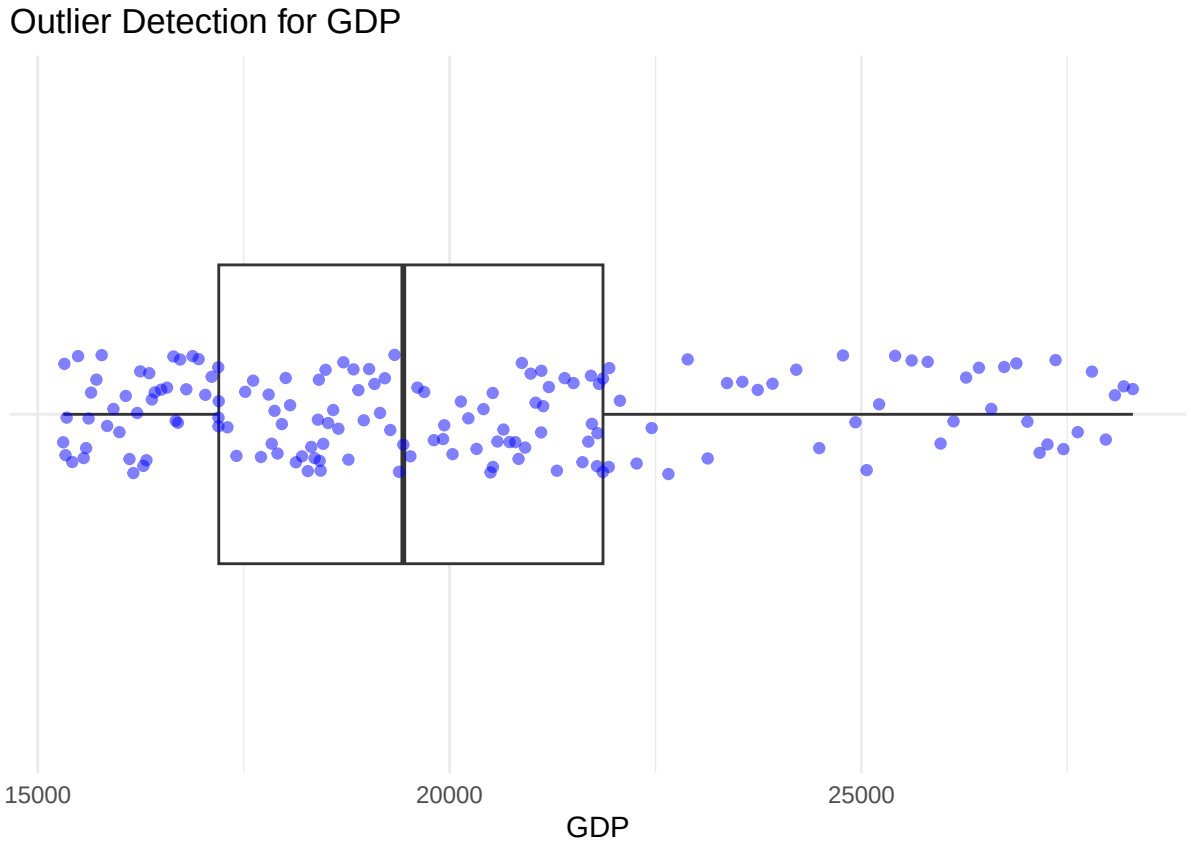

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: PCE ===
## ============================================================
##
## Outlier Detection Results:
##
```

```
## IQR method
## IQR bounds: Lower = 365.6 , Upper = 2614.4
##
## Summary Statistics:
## $mean
## [1] 1505.364
##
## $median
## [1] 1390.8
##
## $sd
## [1] 348.248
##
## $Q1
##    25%
## 1208.9
##
## $Q3
##    75%
## 1771.1
```



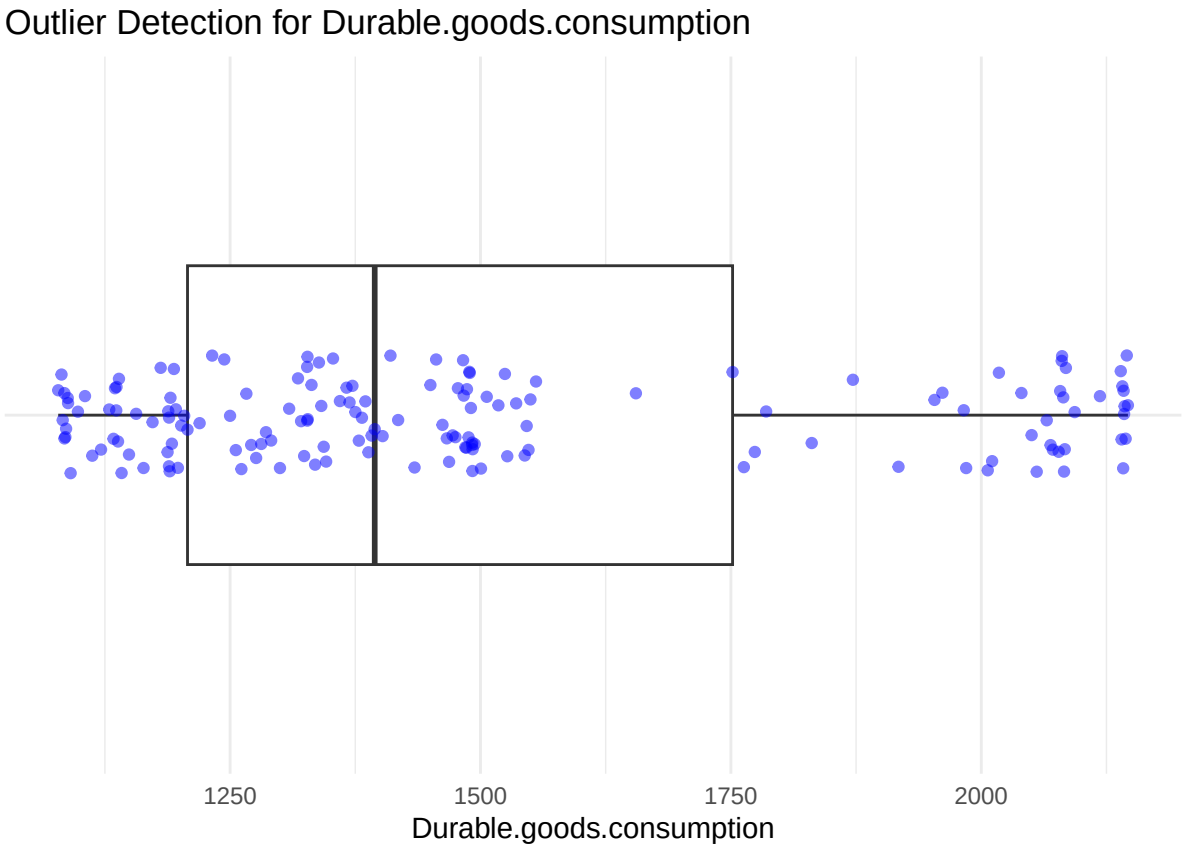

```
##
```

```
## ------------------------------------------------------------
##
## ============================================================
## === Variable: GDP ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 10199.46 , Upper = 28861.44
##
## Summary Statistics:
## $mean
## [1] 20194.55
##
## $median
## [1] 19438.6
##
## $sd
## [1] 3578.868
##
## $Q1
##     25%
## 17197.7
##
## $Q3
##      75%
## 21863.19
```



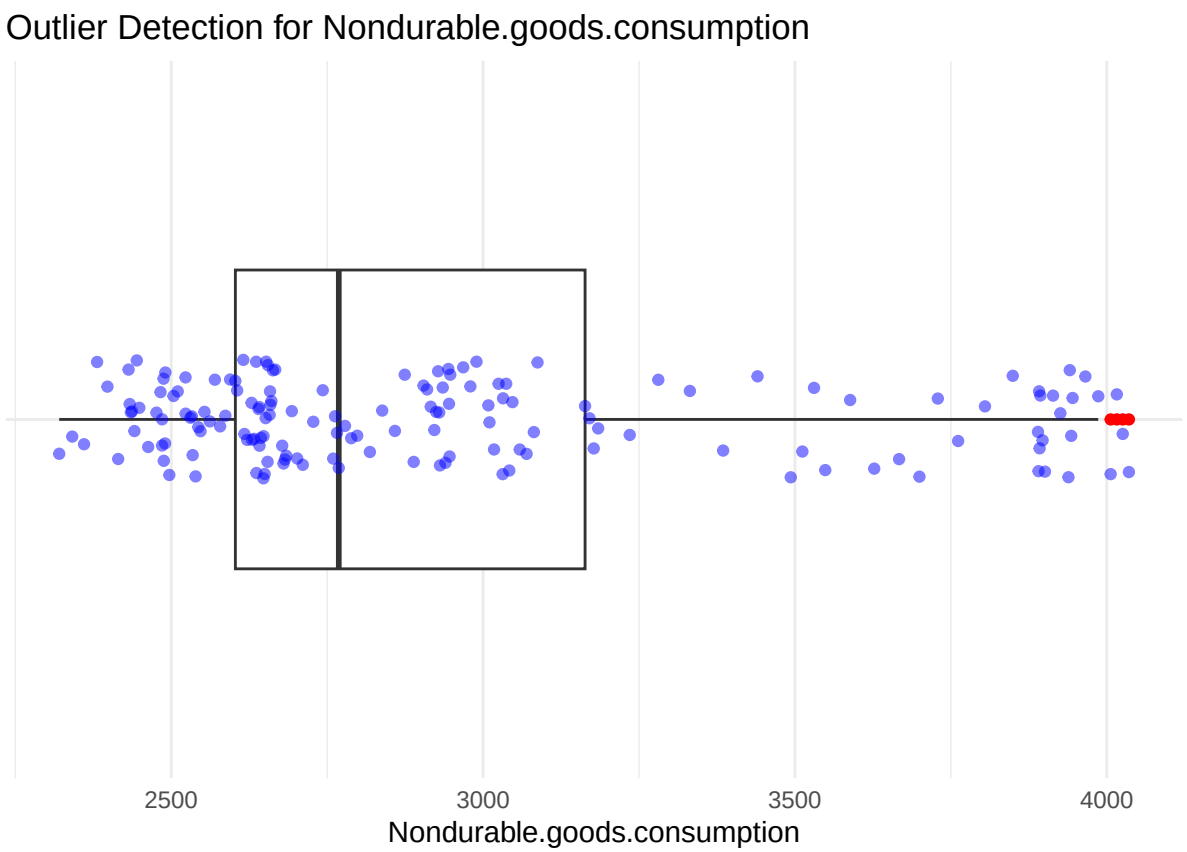

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Durable.goods.consumption ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 391.2 , Upper = 2568
##
## Summary Statistics:
## $mean
## [1] 1498.626
##
## $median
## [1] 1394.4
##
## $sd
## [1] 343.381
```

```
##
## $Q1
##    25%
## 1207.5
##
## $Q3
##    75%
## 1751.7
```



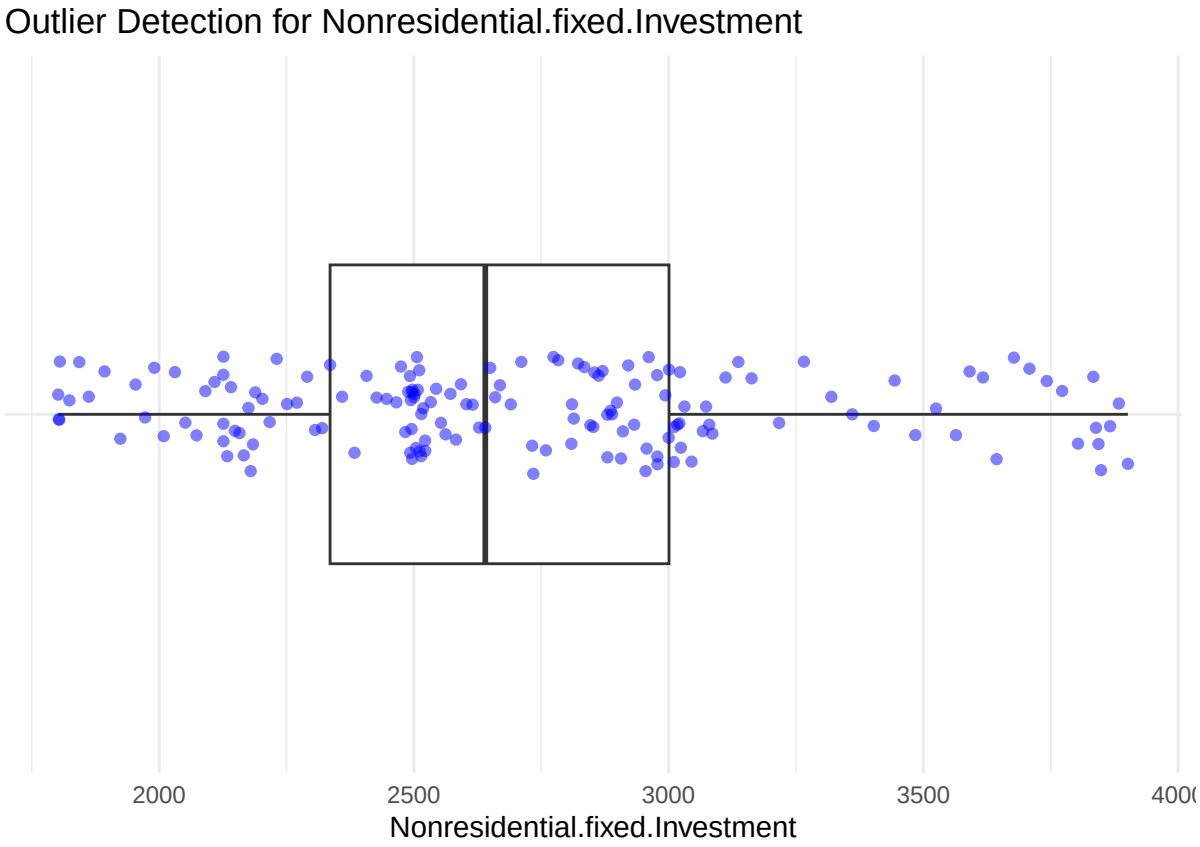

```
##
## ----------------------------------------------------------------
##
## ================================================================
## === Variable: Nondurable.goods.consumption ===
## ================================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 1761.75 , Upper = 4004.55
##
##  IQR outliers
```

```
## Number of outliers: 4
## Outlier values: 4006.2, 4016.039, 4025.561, 4035.4
##
## Summary Statistics:
## $mean
## [1] 2953.967
##
## $median
## [1] 2768.6
##
## $sd
## [1] 493.195
##
## $Q1
##    25%
## 2602.8
##
## $Q3
##    75%
## 3163.5
```



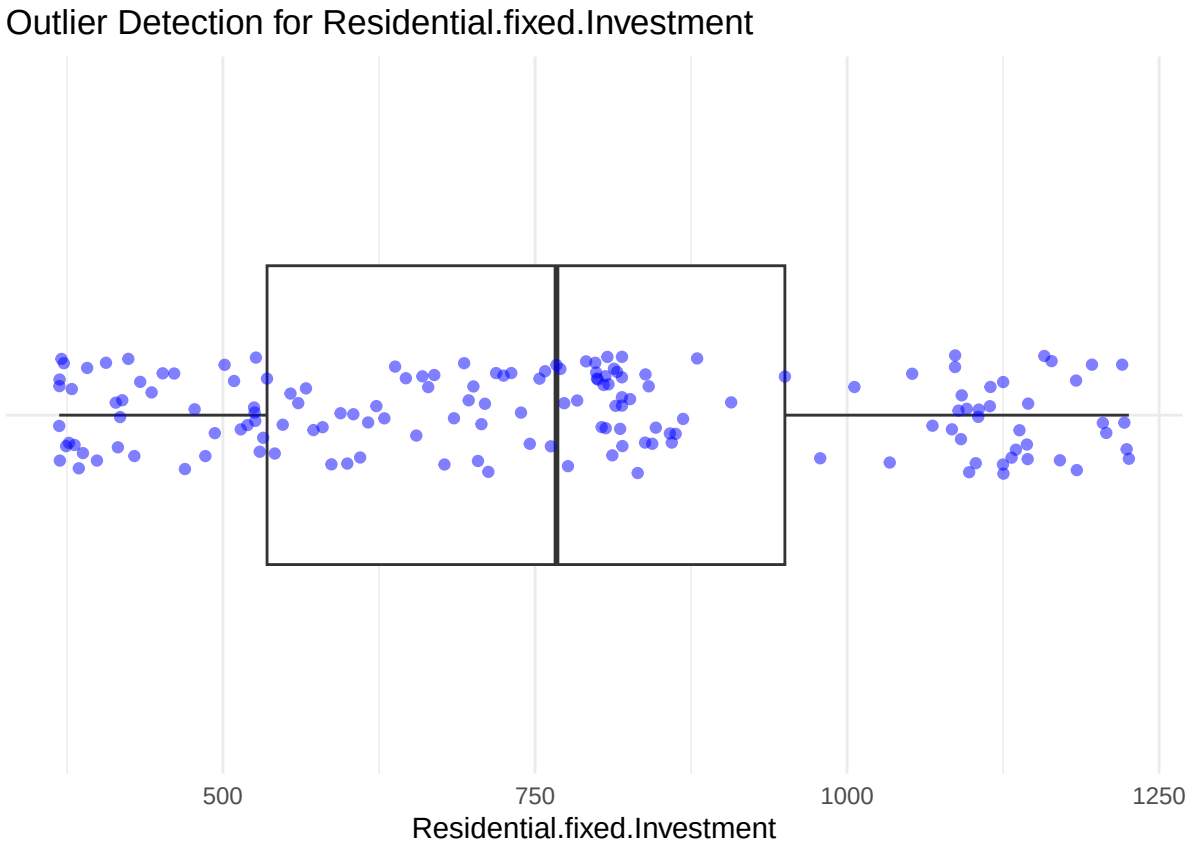

```
##
```

```
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Nonresidential.fixed.Investment ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 1337.55 , Upper = 3998.75
##
## Summary Statistics:
## $mean
## [1] 2711.07
##
## $median
## [1] 2640.2
##
## $sd
## [1] 524.978
##
## $Q1
##    25%
## 2335.5
##
## $Q3
##    75%
## 3000.8
```



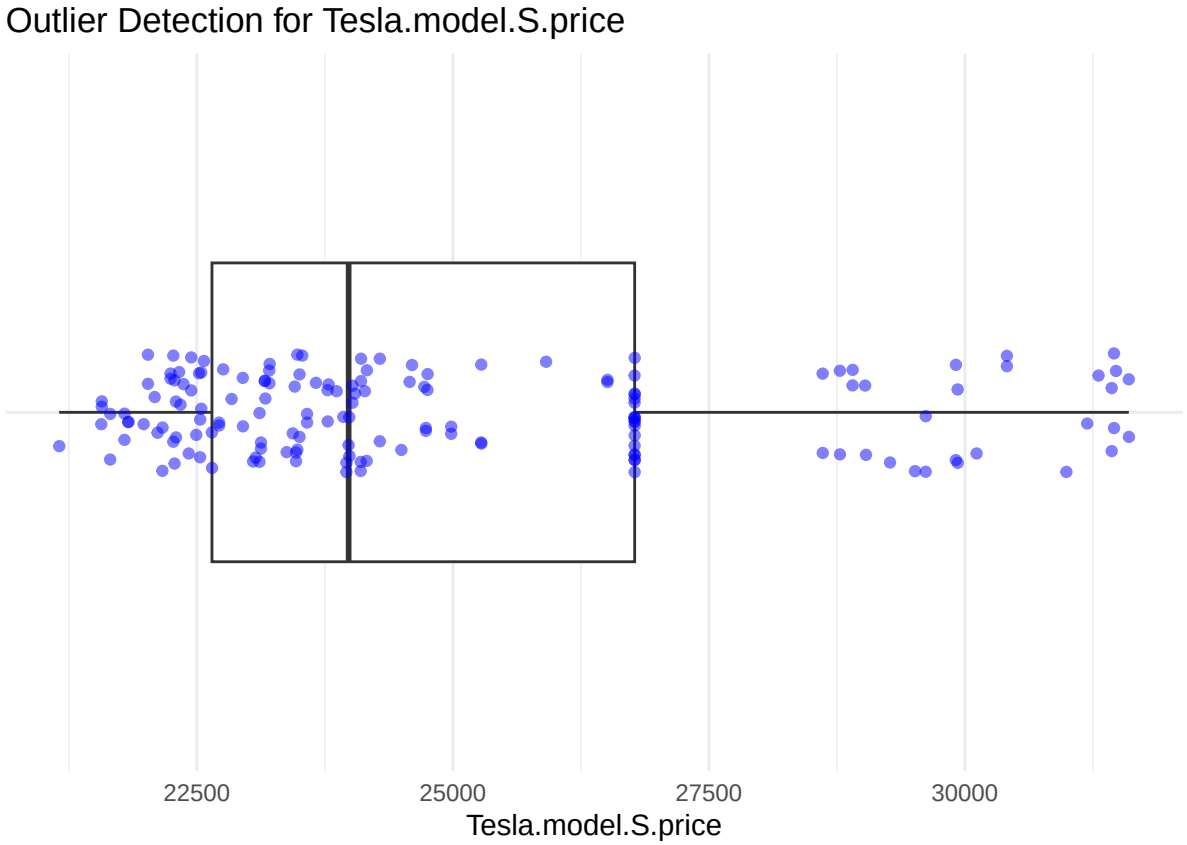

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Residential.fixed.Investment ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = -87.05 , Upper = 1572.55
##
## Summary Statistics:
## $mean
## [1] 761.288
##
## $median
## [1] 767.2
##
## $sd
## [1] 257.924
```

```
##
## $Q1
##   25%
## 535.3
##
## $Q3
##   75%
## 950.2
```



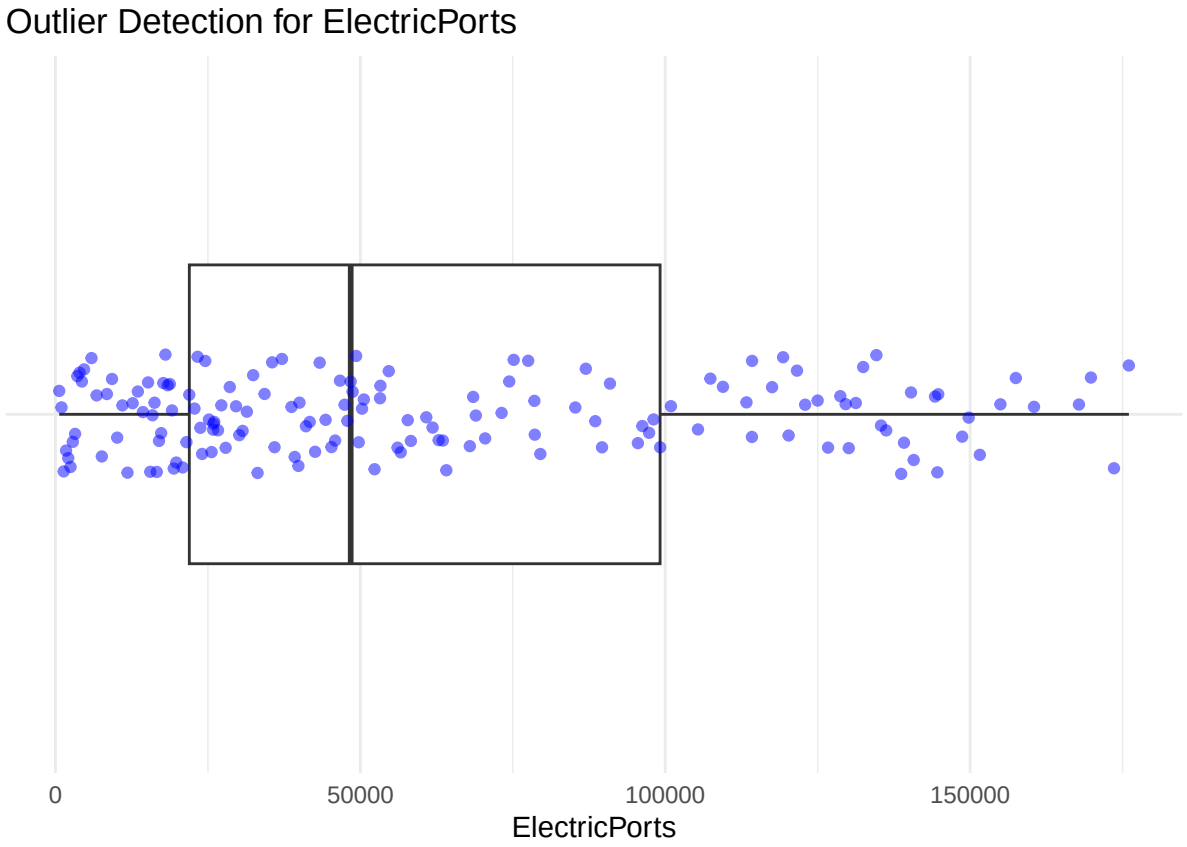

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Tesla.model.S.price ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 16456.83 , Upper = 32966.32
##
## Summary Statistics:
```

```
## $mean
## [1] 24964.04
##
## $median
## [1] 23982.49
##
## $sd
## [1] 2889.943
##
## $Q1
##      25%
## 22647.89
##
## $Q3
##      75%
## 26775.26
```



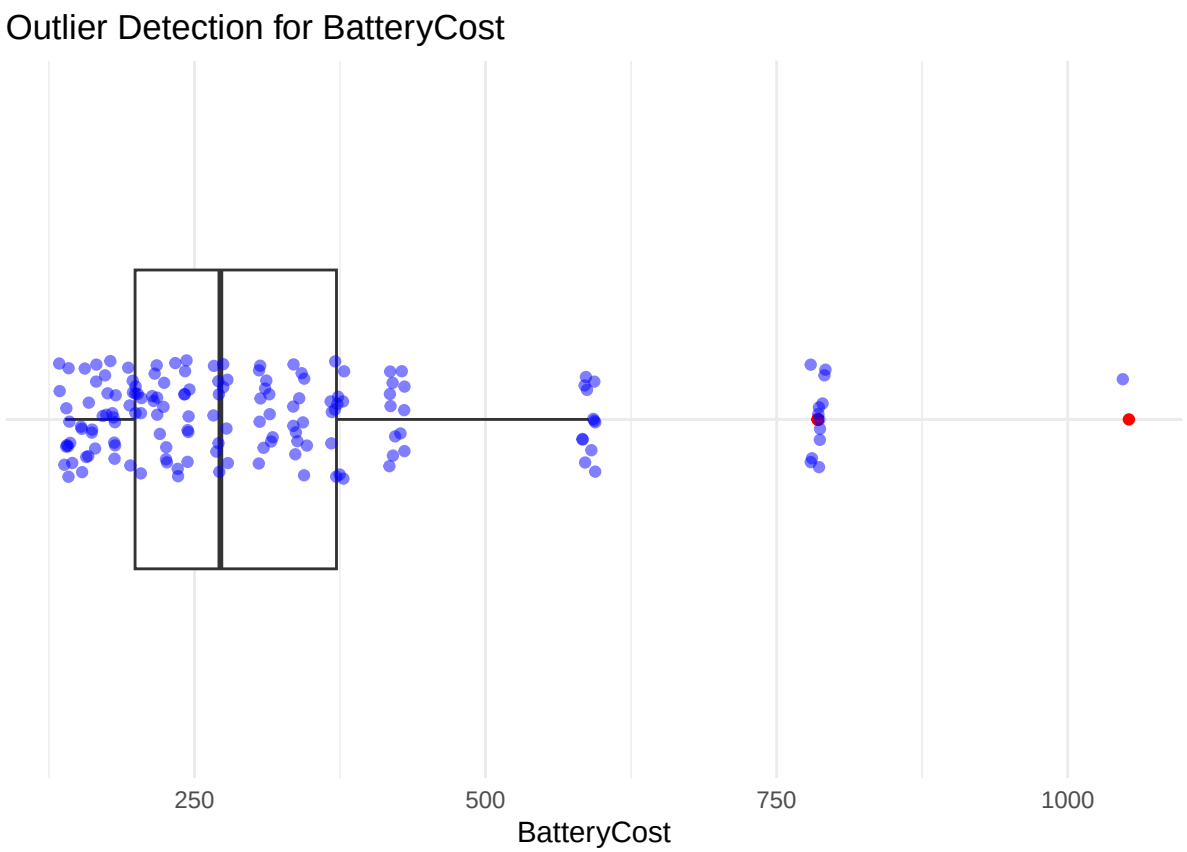

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: ElectricPorts ===
```

```
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = -93844.5 , Upper = 214951.5
##
## Summary Statistics:
## $mean
## [1] 62525.83
##
## $median
## [1] 48416
##
## $sd
## [1] 48945.74
##
## $Q1
##   25%
## 21954
##
## $Q3
##   75%
## 99153
```



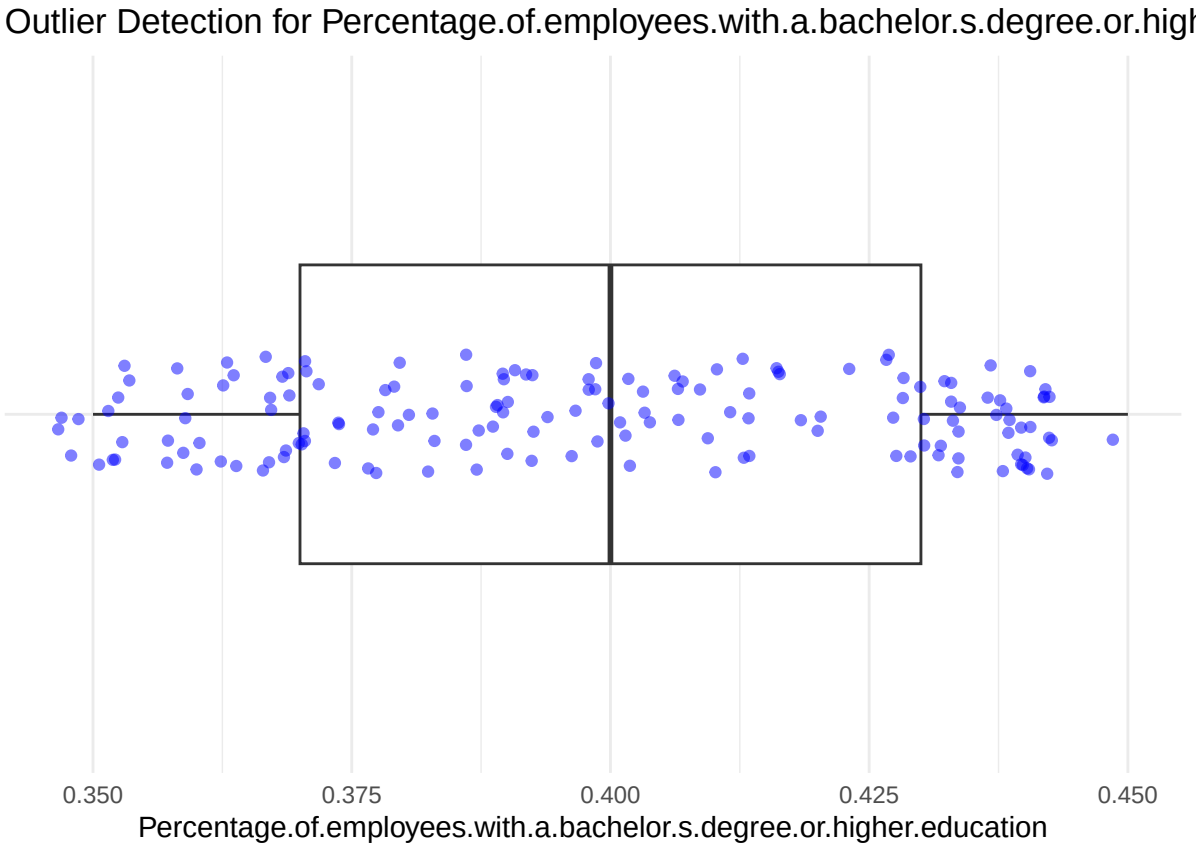

```
##
## ----------------------------------------------------------------
##
## ================================================================
## === Variable: BatteryCost ===
## ================================================================
##
## Outlier Detection Results:
##
##  Z-score method (|z| > 3)
## Number of outliers: 1
## Outlier values: 1052.559
##
## IQR method
## IQR bounds: Lower = -60.541 , Upper = 631.478
##
##  IQR outliers
## Number of outliers: 13
## Outlier values: 1052.559, 785.396, 785.396, 785.396, 785.396, 785.396, 785.396, 785.3
##
##  Modified Z-score method (|modified z| > 3.5)
```

```
## Number of outliers: 1
## Outlier values: 1052.559
##
## Summary Statistics:
## $mean
## [1] 329.734
##
## $median
## [1] 272.233
##
## $sd
## [1] 187.858
##
## $Q1
##     25%
## 198.966
##
## $Q3
##     75%
## 371.971
```



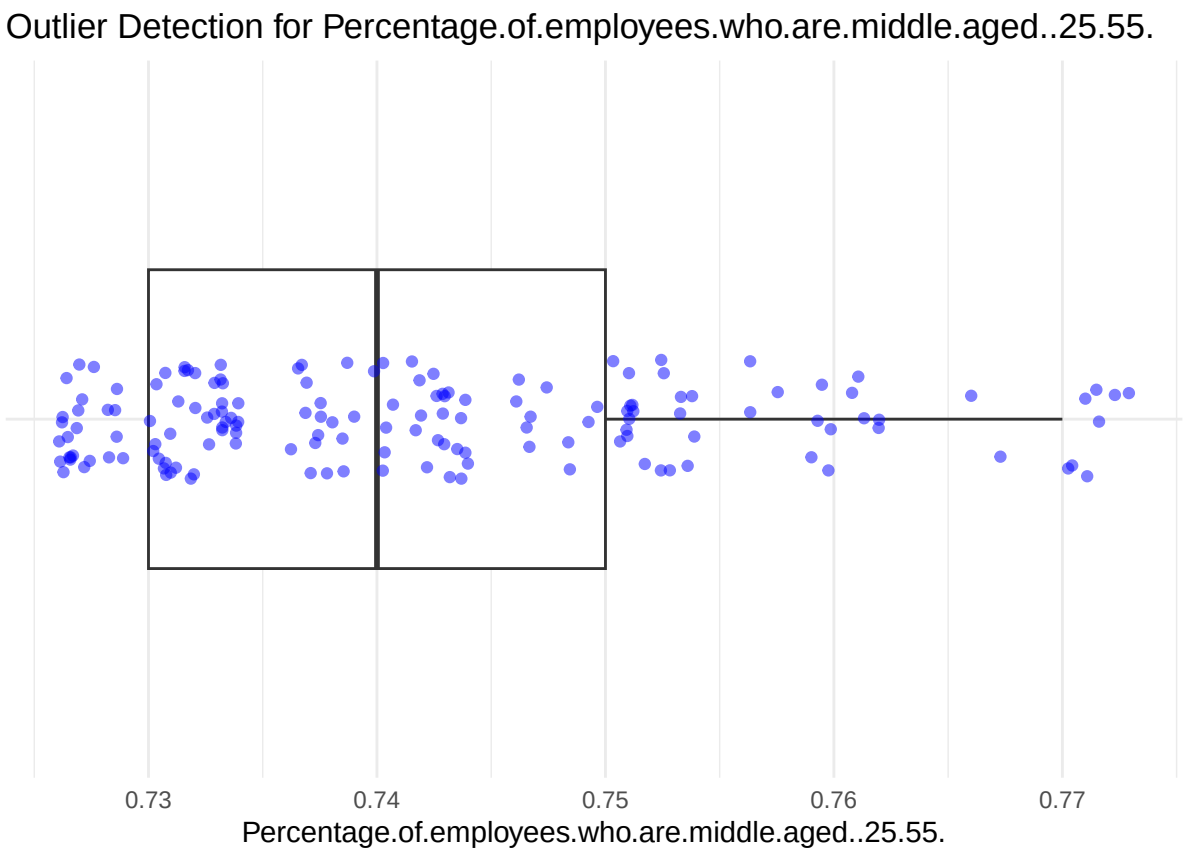

```
##
```

```
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education ==
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 0.28 , Upper = 0.52
##
## Summary Statistics:
## $mean
## [1] 0.398
##
## $median
## [1] 0.4
##
## $sd
## [1] 0.03
##
## $Q1
##  25%
## 0.37
##
## $Q3
##  75%
## 0.43
```



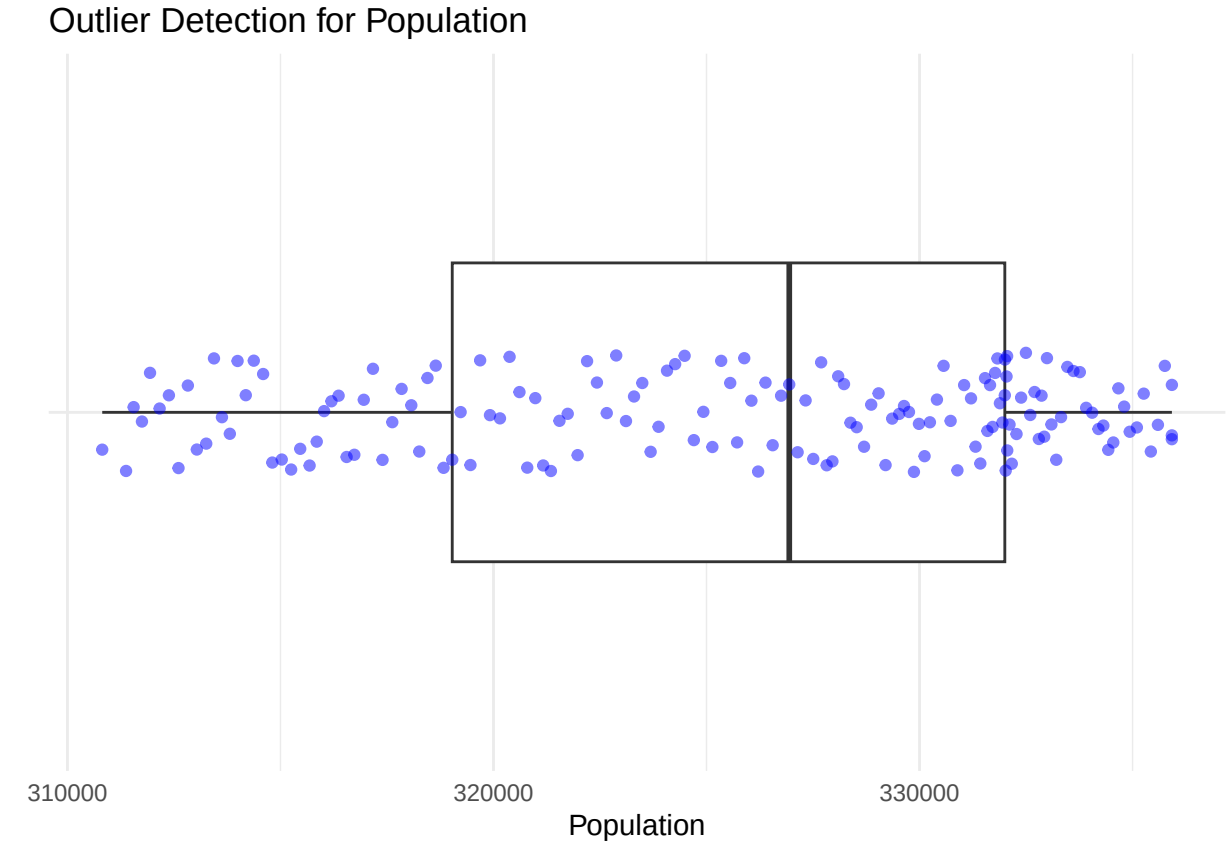

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Percentage.of.employees.who.are.middle.aged..25.55. ===
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 0.7 , Upper = 0.78
##
## Summary Statistics:
## $mean
## [1] 0.742
##
## $median
## [1] 0.74
##
## $sd
## [1] 0.012
```

```
##
## $Q1
##  25%
## 0.73
##
## $Q3
##  75%
## 0.75
```



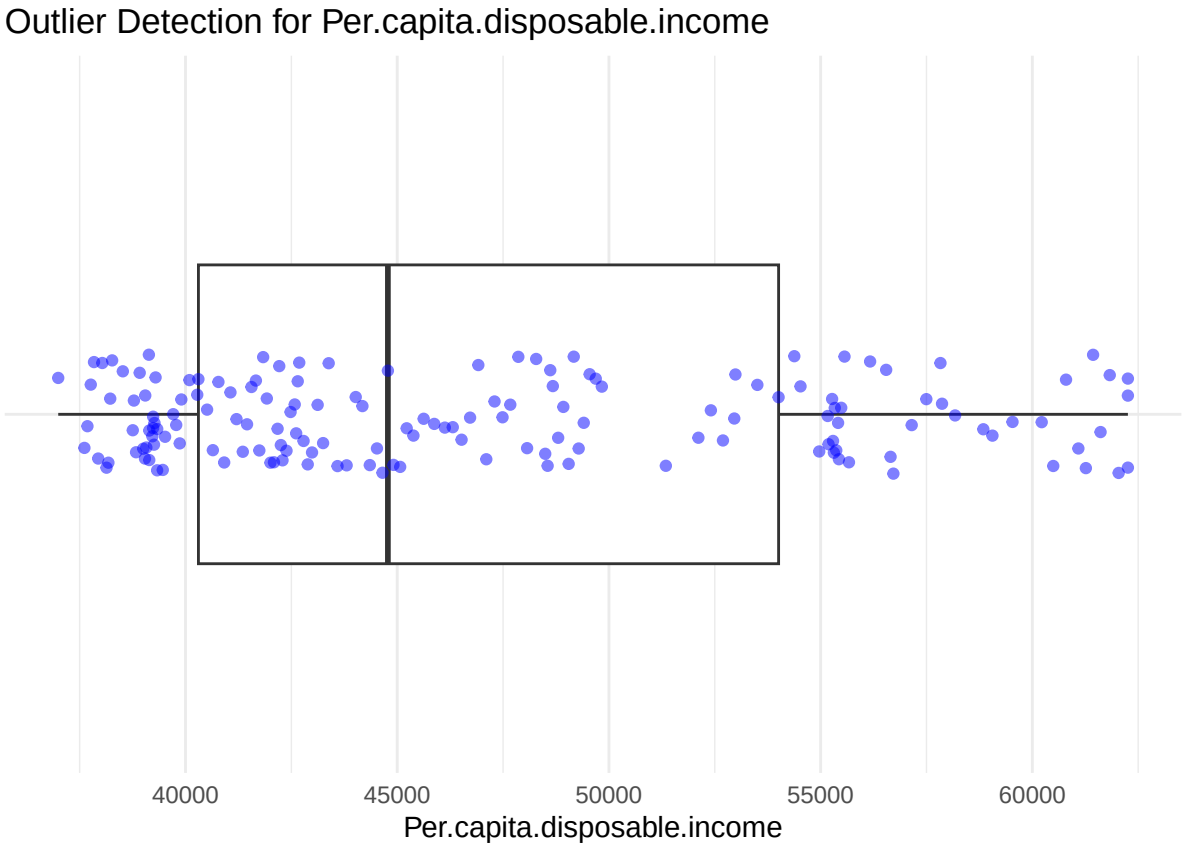

```
##
## ----------------------------------------------------------------
##
## ================================================================
## === Variable: Population ===
## ================================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 299576.1 , Upper = 351457
##
## Summary Statistics:
```

```
## $mean
## [1] 325413
##
## $median
## [1] 326942.5
##
## $sd
## [1] 7372.479
##
## $Q1
##      25%
## 319031.5
##
## $Q3
##      75%
## 332001.7
```



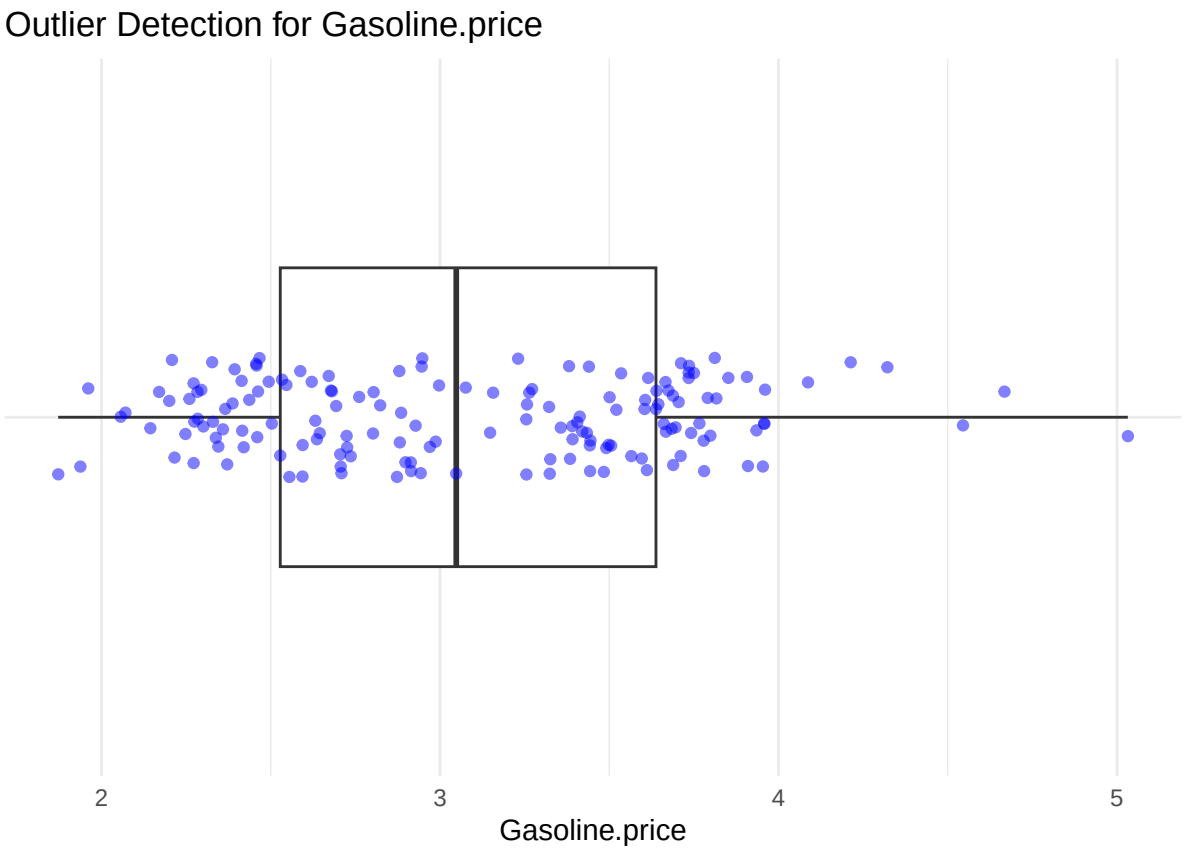

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Per.capita.disposable.income ===
```

```
## ============================================================
##
## Outlier Detection Results:
##
## IQR method
## IQR bounds: Lower = 19751.03 , Upper = 74561.13
##
## Summary Statistics:
## $mean
## [1] 46949.09
##
## $median
## [1] 44779.71
##
## $sd
## [1] 7481.783
##
## $Q1
##      25%
## 40304.82
##
## $Q3
##      75%
## 54007.35
```



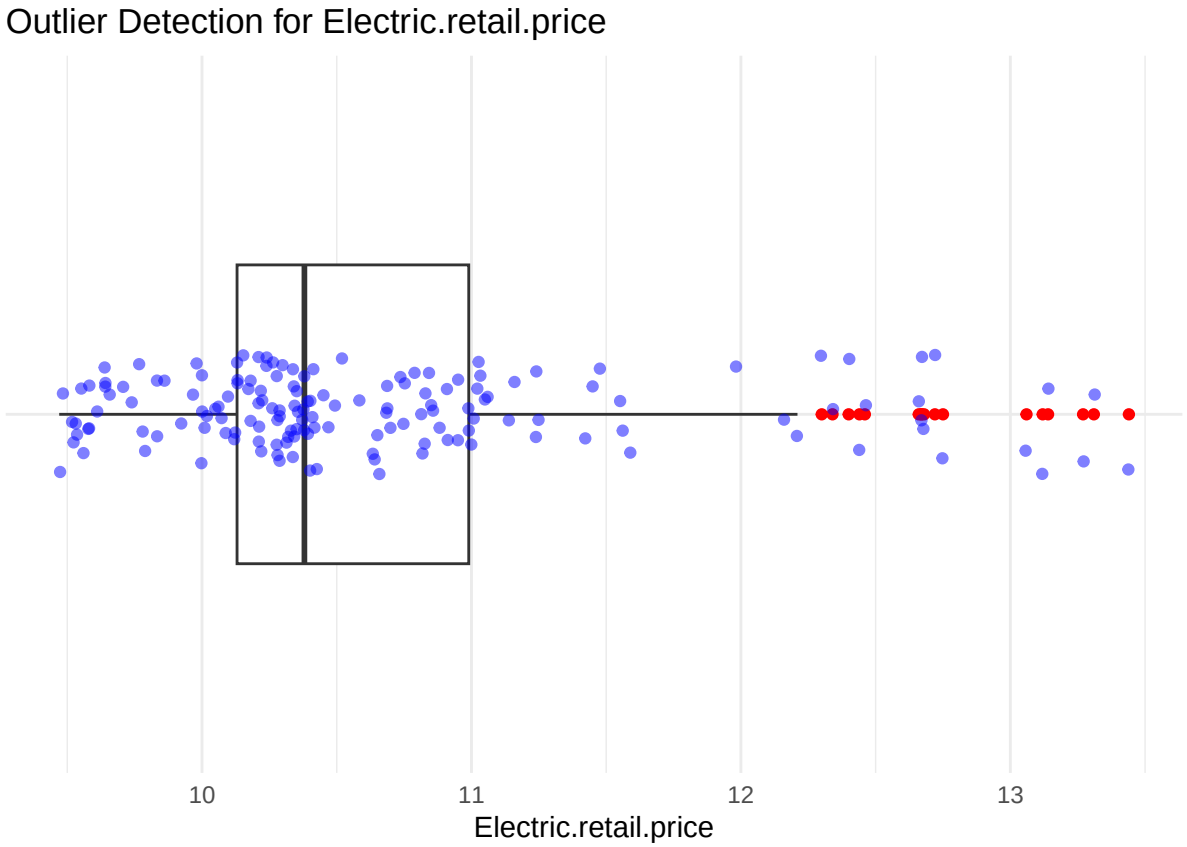

```
##
## ------------------------------------------------------------
##
## ============================================================
## === Variable: Gasoline.price ===
## ============================================================
##
## Outlier Detection Results:
##
##  Z-score method (|z| > 3)
## Number of outliers: 1
## Outlier values: 5.032
##
## IQR method
## IQR bounds: Lower = 0.863 , Upper = 5.303
##
## Summary Statistics:
## $mean
## [1] 3.087
##
## $median
```

```
## [1] 3.048
##
## $sd
## [1] 0.637
##
## $Q1
##   25%
## 2.528
##
## $Q3
##   75%
## 3.638
```




```
##
## ----------------------------------------------------------------
##
## ================================================================
## === Variable: Electric.retail.price ===
## ================================================================
##
## Outlier Detection Results:
##
```

```
##  Z-score method (|z| > 3)
## Number of outliers: 1
## Outlier values: 13.44
##
## IQR method
## IQR bounds: Lower = 8.84 , Upper = 12.28
##
##  IQR outliers
## Number of outliers: 17
## Outlier values: 12.75, 13.12, 13.44, 13.31, 12.66, 12.3, 12.4, 12.68, 12.67, 12.46, 1
##
## Summary Statistics:
## $mean
## [1] 10.673
##
## $median
## [1] 10.38
##
## $sd
## [1] 0.918
##
## $Q1
##   25%
## 10.13
##
## $Q3
##   75%
## 10.99
```




```
##
## ------------------------------------------------------------
```

Stepwise MLR

```
install.packages("leaps")
```

```
##   'leaps'   MD5
## Warning:      'leaps'
## Warning in file.copy(savedcopy, lib, recursive = TRUE):
##  D:\Software\R-4.4.1\library\00LOCK\leaps\libs\x64\leaps.dll D:\Software\R-4.4.1\libr
## denied
## Warning:   'leaps'
##
##
##  C:\Users\yuanting\AppData\Local\Temp\RtmpWUhATB\downloaded_packages
```

```
library("leaps")
```

```
## Warning:   'leaps' R 4.4.2
```

```r
full_model <- lm(data$Total.NEV.Sales~., data = data)
stepwise_model <- step(full_model, direction = "both", trace = 0)
summary(stepwise_model)
```

```
##
## Call:
## lm(formula = data$Total.NEV.Sales ~ Month + Date + Covid.19 +
##     Number.of.new.energy.policies + Inventory + Gasoline.LDV.sales +
##     M2SL + Umemployment.rate + PCE + Durable.goods.consumption +
##     Nondurable.goods.consumption + Nonresidential.fixed.Investment +
##     Residential.fixed.Investment + Tesla.model.S.price + ElectricPorts +
##     BatteryCost + Percentage.of.employees.with.a.bachelor.s.degree.or.higher.educatio
##     Population, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23665.1  -3936.7    146.7   3616.7  22207.9
##
## Coefficients:
##                                                                    Estimate
## (Intercept)                                                      -2.260e+05
## Month                                                             7.202e+02
## Date                                                              1.657e+02
## Covid.19                                                         -1.253e+04
## Number.of.new.energy.policies                                    -1.265e+02
## Inventory                                                         3.758e-02
## Gasoline.LDV.sales                                                3.695e-02
## M2SL                                                             -1.769e+01
## Umemployment.rate                                                 1.598e+05
## PCE                                                               4.052e+01
## Durable.goods.consumption                                         9.803e+01
## Nondurable.goods.consumption                                      6.942e+01
## Nonresidential.fixed.Investment                                  -7.402e+01
## Residential.fixed.Investment                                     -1.212e+02
## Tesla.model.S.price                                               2.816e+00
## ElectricPorts                                                     9.869e-01
## BatteryCost                                                      -6.182e+01
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education -3.573e+05
## Population                                                       -2.065e+01
##                                                                  Std. Error
## (Intercept)                                                       6.211e+05
## Month                                                             2.193e+02
## Date                                                              5.764e+01
## Covid.19                                                          4.596e+03
```

```
## Number.of.new.energy.policies                                    2.333e+01
## Inventory                                                        8.642e-03
## Gasoline.LDV.sales                                               5.178e-03
## M2SL                                                             4.651e+00
## Umemployment.rate                                                1.024e+05
## PCE                                                              1.881e+01
## Durable.goods.consumption                                        3.106e+01
## Nondurable.goods.consumption                                     2.902e+01
## Nonresidential.fixed.Investment                                  2.564e+01
## Residential.fixed.Investment                                     4.047e+01
## Tesla.model.S.price                                              1.150e+00
## ElectricPorts                                                    3.611e-01
## BatteryCost                                                      1.802e+01
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education 1.670e+05
## Population                                                       6.372e+00
##                                                                 t value
## (Intercept)                                                      -0.364
## Month                                                             3.283
## Date                                                              2.874
## Covid.19                                                         -2.727
## Number.of.new.energy.policies                                    -5.423
## Inventory                                                         4.349
## Gasoline.LDV.sales                                                7.135
## M2SL                                                             -3.803
## Umemployment.rate                                                 1.560
## PCE                                                               2.154
## Durable.goods.consumption                                         3.156
## Nondurable.goods.consumption                                      2.393
## Nonresidential.fixed.Investment                                  -2.886
## Residential.fixed.Investment                                     -2.994
## Tesla.model.S.price                                               2.449
## ElectricPorts                                                     2.733
## BatteryCost                                                      -3.432
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education -2.140
## Population                                                       -3.241
##                                                                 Pr(>|t|)
## (Intercept)                                                     0.716547
## Month                                                           0.001300
## Date                                                            0.004689
## Covid.19                                                        0.007226
## Number.of.new.energy.policies                                   2.55e-07
## Inventory                                                       2.64e-05
## Gasoline.LDV.sales                                              4.98e-11
## M2SL                                                            0.000214
## Umemployment.rate                                               0.121025
```

```
## PCE                                                                    0.032941
## Durable.goods.consumption                                              0.001961
## Nondurable.goods.consumption                                           0.018078
## Nonresidential.fixed.Investment                                        0.004525
## Residential.fixed.Investment                                           0.003263
## Tesla.model.S.price                                                    0.015569
## ElectricPorts                                                          0.007104
## BatteryCost                                                            0.000792
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education 0.034152
## Population                                                             0.001495
##
## (Intercept)
## Month                                                                  **
## Date                                                                   **
## Covid.19                                                               **
## Number.of.new.energy.policies                                          ***
## Inventory                                                              ***
## Gasoline.LDV.sales                                                     ***
## M2SL                                                                   ***
## Umemployment.rate
## PCE                                                                    *
## Durable.goods.consumption                                              **
## Nondurable.goods.consumption                                           *
## Nonresidential.fixed.Investment                                        **
## Residential.fixed.Investment                                           **
## Tesla.model.S.price                                                    *
## ElectricPorts                                                          **
## BatteryCost                                                            ***
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education   *
## Population                                                             **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7405 on 138 degrees of freedom
## Multiple R-squared:  0.9838, Adjusted R-squared:  0.9816
## F-statistic: 464.3 on 18 and 138 DF,  p-value: < 2.2e-16
```

```
library(car)
```

```
##      carData
```

```
vif(stepwise_model)
```

```
##                                                                    Month
##                                                                1.657330
##                                                                     Date
```

```
##                                                                 18101.439629
##                                                                     Covid.19
##                                                                    11.668300
##                                                  Number.of.new.energy.policies
##                                                                     5.166106
##                                                                    Inventory
##                                                                    34.059699
##                                                               Gasoline.LDV.sales
##                                                                     3.063568
##                                                                          M2SL
##                                                                  1049.702132
##                                                               Umemployment.rate
##                                                                     12.613786
##                                                                          PCE
##                                                                  122.058974
##                                                     Durable.goods.consumption
##                                                                    323.540754
##                                                   Nondurable.goods.consumption
##                                                                    582.539717
##                                                Nonresidential.fixed.Investment
##                                                                    515.587131
##                                                   Residential.fixed.Investment
##                                                                    309.945416
##                                                            Tesla.model.S.price
##                                                                     31.408739
##                                                                  ElectricPorts
##                                                                     888.792974
##                                                                   BatteryCost
##                                                                     32.582692
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education
##                                                                     70.251108
##                                                                    Population
##                                                                  6278.272065
```

```
# there is problem with multi colinearity, we should not use all of these predictors
```

```
data_stepwise <- subset(data, select = -c(M2SL,GDP,Durable.goods.consumption,Nondurable.
data_stepwise0<-data_stepwise
full_model2 <- lm(Total.NEV.Sales~., data = data_stepwise)
stepwise_model2 <- step(full_model2, direction = "both", trace = 0)
summary(stepwise_model2)
```

```
##
## Call:
## lm(formula = Total.NEV.Sales ~ Date + Covid.19 + Number.of.new.energy.policies +
##     Production + Gasoline.LDV.sales + M1SL + Federal.fund.effective.rate +
```

```
##     Umemployment.rate + PPIBC + Tesla.model.S.price + Gasoline.price +
##     Electric.retail.price, data = data_stepwise)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -37728  -5242   -588   5908  62345
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -6.345e+05  2.016e+05  -3.148  0.00200 **
## Date                           1.634e+01  3.988e+00   4.097 6.97e-05 ***
## Covid.19                      -3.250e+04  5.516e+03  -5.892 2.59e-08 ***
## Number.of.new.energy.policies -1.605e+02  3.586e+01  -4.475 1.54e-05 ***
## Production                     3.810e-02  2.370e-02   1.608  0.11011
## Gasoline.LDV.sales             3.391e-02  8.076e-03   4.199 4.67e-05 ***
## M1SL                           5.750e+00  6.435e-01   8.936 1.76e-15 ***
## Federal.fund.effective.rate    1.627e+06  1.532e+05  10.615  < 2e-16 ***
## Umemployment.rate              3.008e+05  1.023e+05   2.942  0.00380 **
## PPIBC                         -1.553e+03  6.110e+02  -2.542  0.01208 *
## Tesla.model.S.price            5.712e+00  1.105e+00   5.168 7.77e-07 ***
## Gasoline.price                 1.262e+04  3.191e+03   3.955  0.00012 ***
## Electric.retail.price         -7.435e+03  2.963e+03  -2.509  0.01320 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12910 on 144 degrees of freedom
## Multiple R-squared:  0.9485, Adjusted R-squared:  0.9442
## F-statistic:   221 on 12 and 144 DF,  p-value: < 2.2e-16
```

```r
library(car)
vif(stepwise_model2)
```

```
##                           Date                      Covid.19
##                      28.509174                      5.532017
## Number.of.new.energy.policies                     Production
##                       4.016289                      4.561375
##             Gasoline.LDV.sales                           M1SL
##                       2.451947                     20.328002
##    Federal.fund.effective.rate             Umemployment.rate
##                       4.689438                      4.137900
##                           PPIBC           Tesla.model.S.price
##                       6.983916                      9.551348
##                  Gasoline.price         Electric.retail.price
##                       3.862100                      6.924861
```

Autocorrelation—adding lagged values

```
library(lmtest)
```

```
##      zoo
```

```
## Warning:   'zoo' R 4.4.2
```

```
##
##    'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
dwtest(stepwise_model2)
```

```
##
##  Durbin-Watson test
##
## data:  stepwise_model2
## DW = 0.92266, p-value = 1.541e-14
## alternative hypothesis: true autocorrelation is greater than 0
```

```
library(dplyr)
```

```
## Warning:   'dplyr' R 4.4.2
```

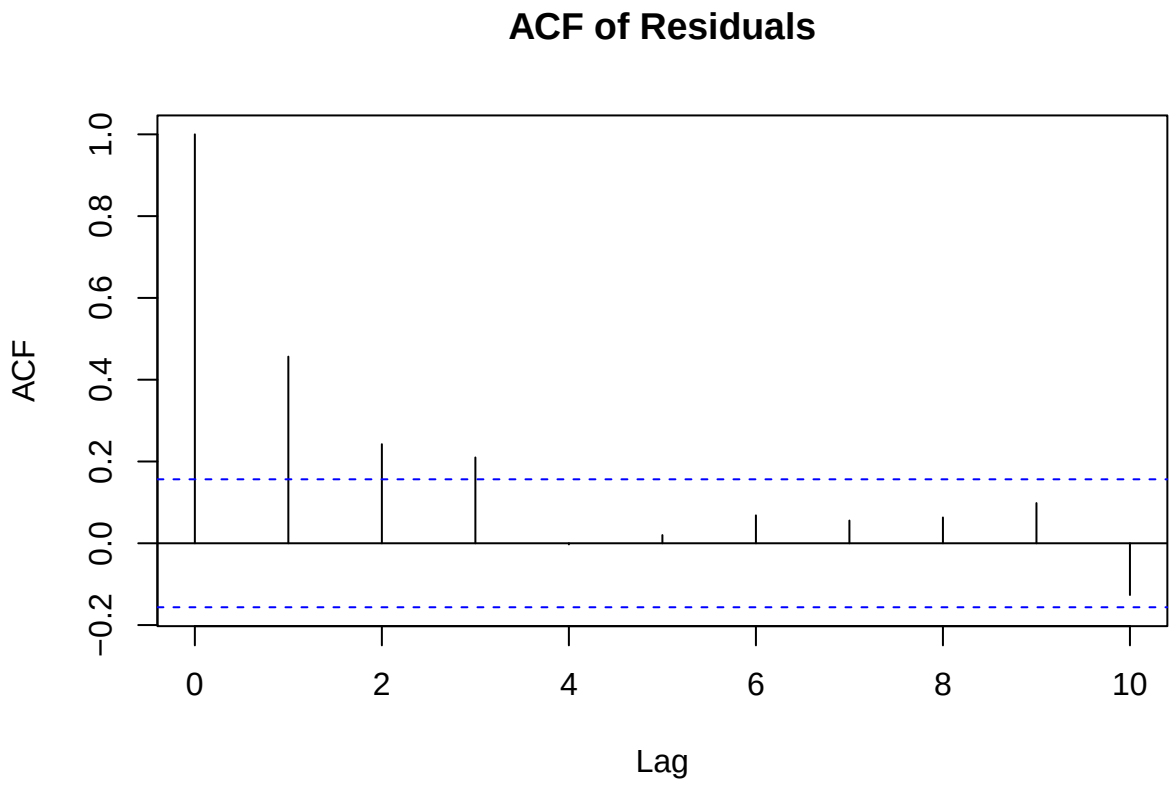
```
##
##    'dplyr'
```

```
## The following object is masked from 'package:car':
##
##     recode
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

```
residuals_lagged <- residuals(stepwise_model2)
acf(residuals_lagged, main = "ACF of Residuals", lag.max = 10)
```




```
# From the ACF of Residuals, the residual autocorrelation at lag orders 1, 7 and 10 is
data_stepwise <- data_stepwise %>%
  mutate(lagged_value1 = lag(data_stepwise$Total.NEV.Sales, n = 1),lagged_value2 = lag(d

data_stepwise <- subset(data_stepwise, select = -c(PPIBM,Date))


# test if lagged variables solve the autocorrelation problem
lagged_model <-  lm(data_stepwise$Total.NEV.Sales~., data = data_stepwise)
dwtest(lagged_model)
```

```
##
##  Durbin-Watson test
##
## data:  lagged_model
## DW = 1.8647, p-value = 0.033
## alternative hypothesis: true autocorrelation is greater than 0
```

```
summary(lagged_model)
```

```
##
## Call:
```

```
## lm(formula = data_stepwise$Total.NEV.Sales ~ ., data = data_stepwise)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24667.1  -3802.9    107.8   3869.5  28580.9
##
## Coefficients:
##                                                   Estimate Std. Error
## (Intercept)                                     -2.247e+04  1.606e+05
## Month                                            5.767e+01  2.266e+02
## Covid.19                                         3.153e+03  4.305e+03
## Number.of.new.energy.policies                   -4.584e+01  2.495e+01
## Production                                       9.140e-04  1.693e-02
## Inventory                                        2.104e-03  8.814e-03
## Gasoline.LDV.sales                               4.820e-02  5.220e-03
## M1SL                                             2.272e+00  4.520e-01
## Federal.fund.effective.rate                      7.529e+05  1.161e+05
## Umemployment.rate                                1.444e+05  6.064e+04
## PPIBC                                           -3.979e+02  4.507e+02
## Tesla.model.S.price                              1.264e+00  7.506e-01
## BatteryCost                                      9.801e+00  1.860e+01
## Percentage.of.employees.who.are.middle.aged..25.55.  2.290e+04  2.257e+05
## Gasoline.price                                   4.228e+03  2.004e+03
## Electric.retail.price                           -6.482e+03  1.872e+03
## lagged_value1                                    3.965e-01  6.398e-02
## lagged_value2                                    8.182e-02  6.849e-02
## lagged_value3                                    3.786e-01  6.157e-02
##                                                 t value Pr(>|t|)
## (Intercept)                                      -0.140 0.888974
## Month                                             0.255 0.799488
## Covid.19                                          0.732 0.465159
## Number.of.new.energy.policies                    -1.837 0.068433 .
## Production                                        0.054 0.957037
## Inventory                                         0.239 0.811668
## Gasoline.LDV.sales                                9.233 4.88e-16 ***
## M1SL                                              5.026 1.56e-06 ***
## Federal.fund.effective.rate                       6.485 1.55e-09 ***
## Umemployment.rate                                 2.381 0.018677 *
## PPIBC                                            -0.883 0.378872
## Tesla.model.S.price                               1.684 0.094463 .
## BatteryCost                                       0.527 0.599018
## Percentage.of.employees.who.are.middle.aged..25.55.  0.101 0.919359
## Gasoline.price                                    2.110 0.036715 *
## Electric.retail.price                            -3.462 0.000717 ***
## lagged_value1                                     6.197 6.52e-09 ***
```

```
## lagged_value2                                          1.195 0.234324
## lagged_value3                                          6.148 8.29e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7767 on 135 degrees of freedom
##   (   3    )
## Multiple R-squared:  0.9822, Adjusted R-squared:  0.9799
## F-statistic: 414.9 on 18 and 135 DF,  p-value: < 2.2e-16
```

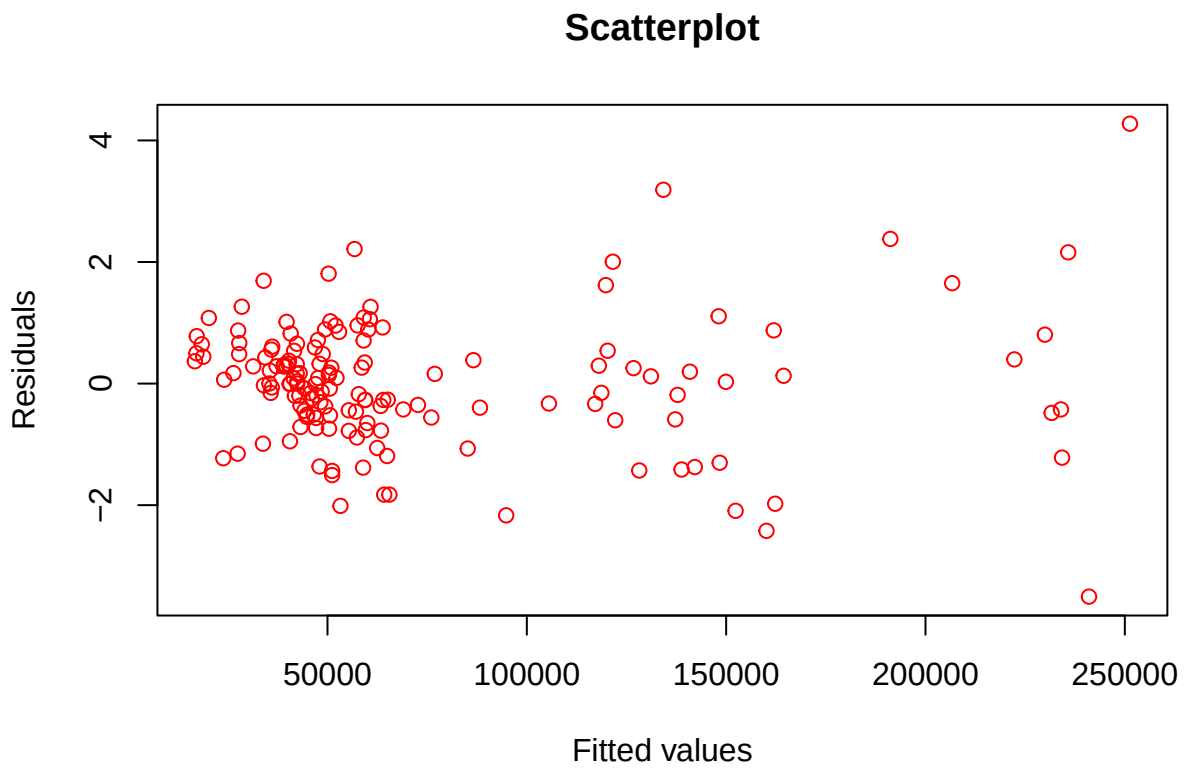
```
library(car)
vif(lagged_model)
```

```
##                                              Month
##                                           1.538686
##                                           Covid.19
##                                           9.244310
##                         Number.of.new.energy.policies
##                                           5.314122
##                                         Production
##                                           6.390757
##                                          Inventory
##                                          32.206621
##                                  Gasoline.LDV.sales
##                                           2.703344
##                                               M1SL
##                                          27.374849
##                            Federal.fund.effective.rate
##                                           7.380239
##                                    Umemployment.rate
##                                           3.822012
##                                              PPIBC
##                                          10.448824
##                                 Tesla.model.S.price
##                                          12.076718
##                                        BatteryCost
##                                          26.087605
## Percentage.of.employees.who.are.middle.aged..25.55.
##                                          17.365337
##                                       Gasoline.price
##                                           4.205110
##                                  Electric.retail.price
##                                           7.400783
##                                        lagged_value1
##                                          28.389926
```

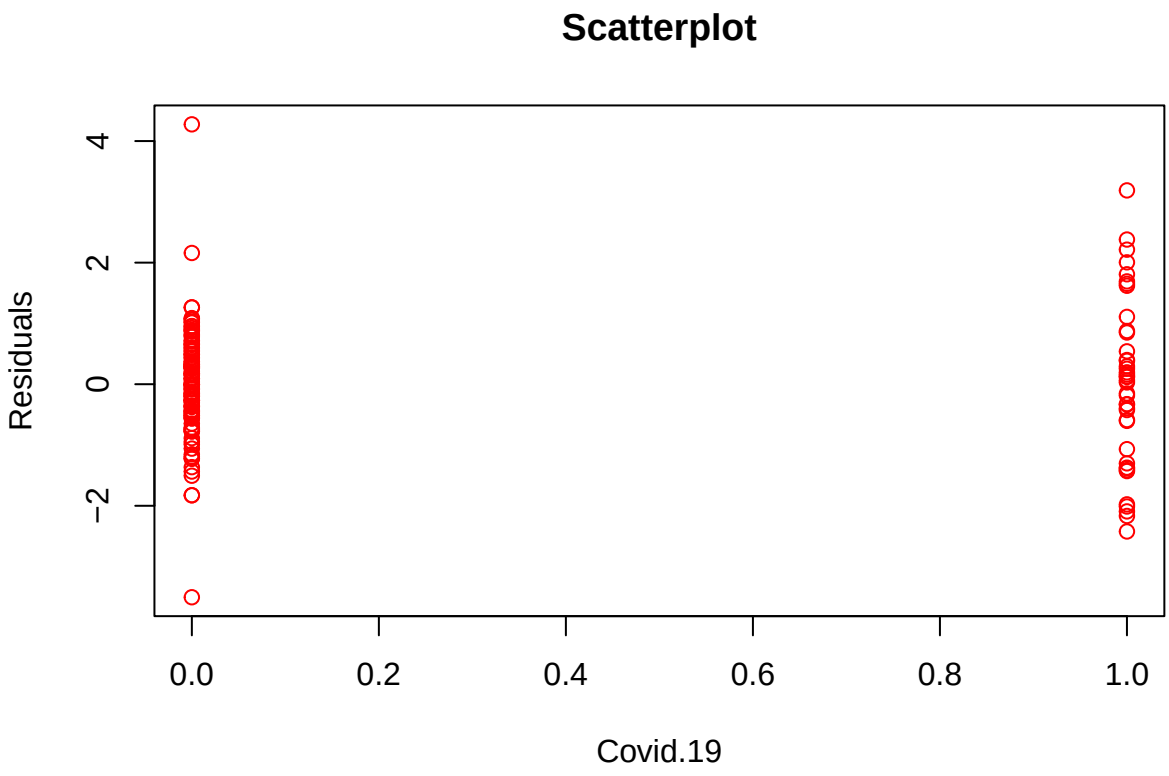
```
##                                     lagged_value2
##                                         30.645384
##                                     lagged_value3
##                                         23.536047
```

Goodness of fit

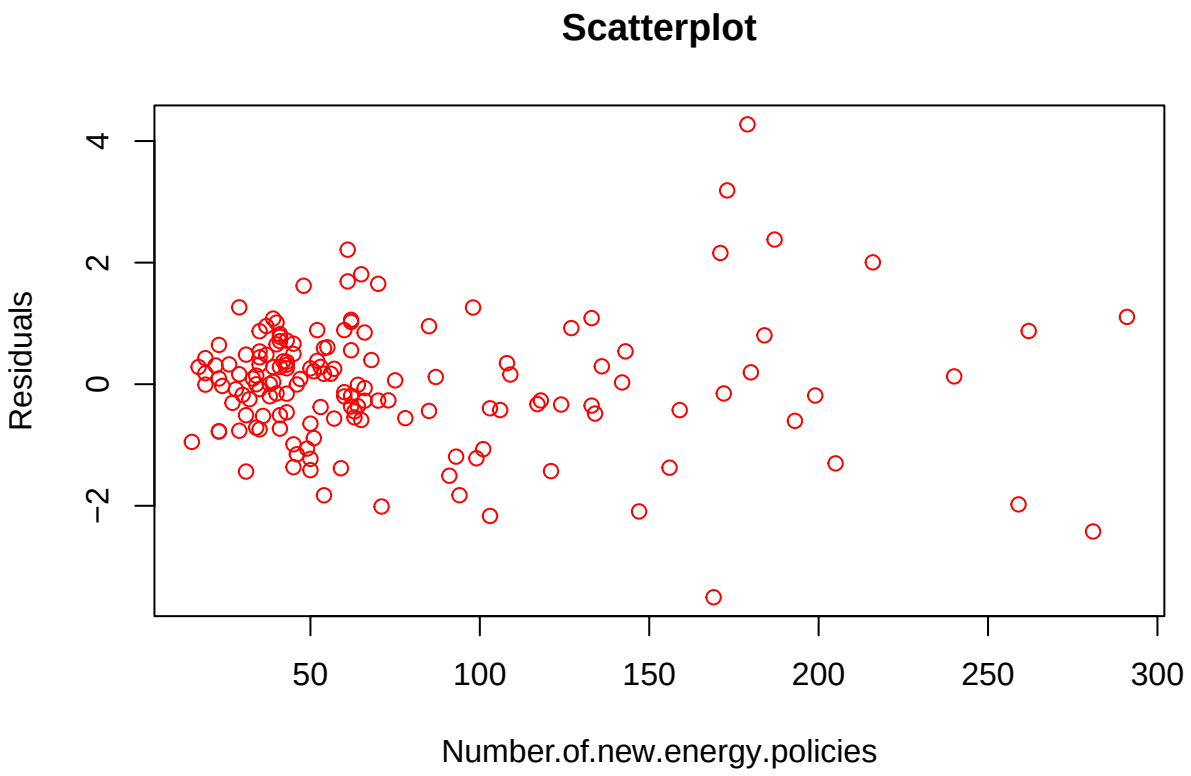
```
# Constant Variance Assumption: hold
resids=rstandard(lagged_model)
fits=lagged_model$fitted
plot(fits,resids,xlab="Fitted values",ylab="Residuals",main="Scatterplot",col="red")
```



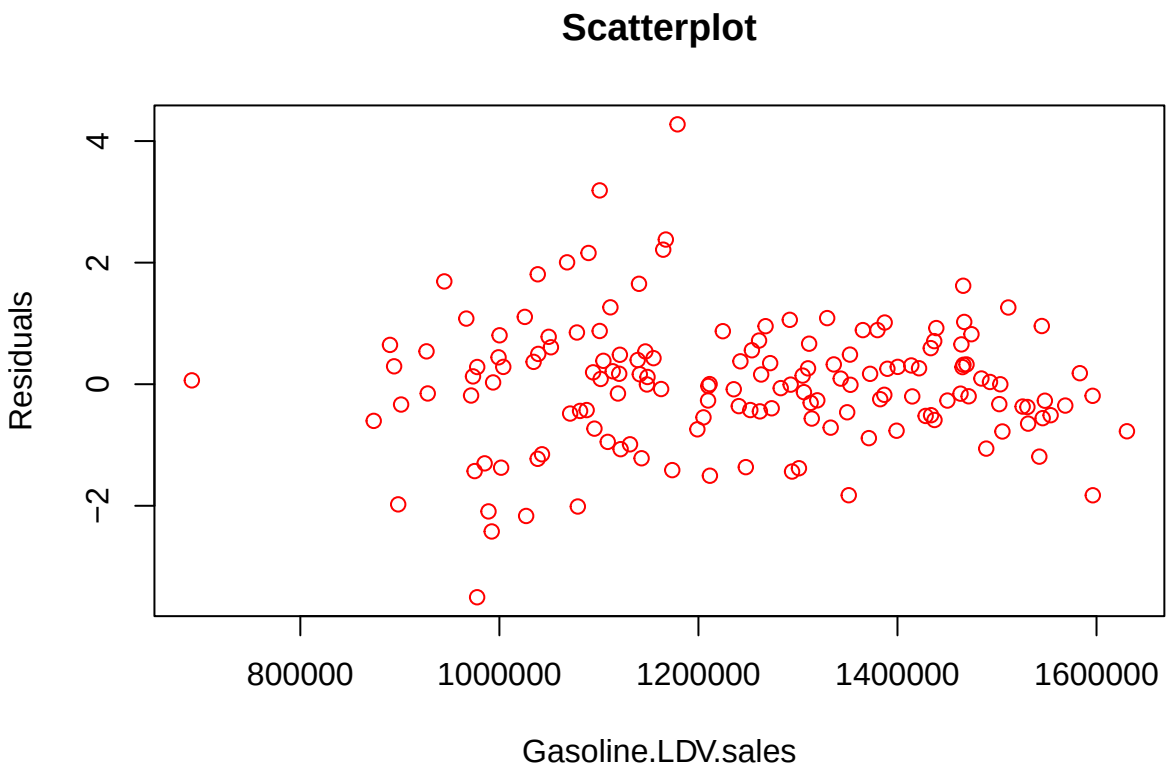

```
# Linearity Assumption: hold
# following codes should be checked according to the variable selection results in cas
plot(data$Covid.19[4:nrow(data)],resids,xlab="Covid.19",ylab="Residuals",main="Scatterpl
```



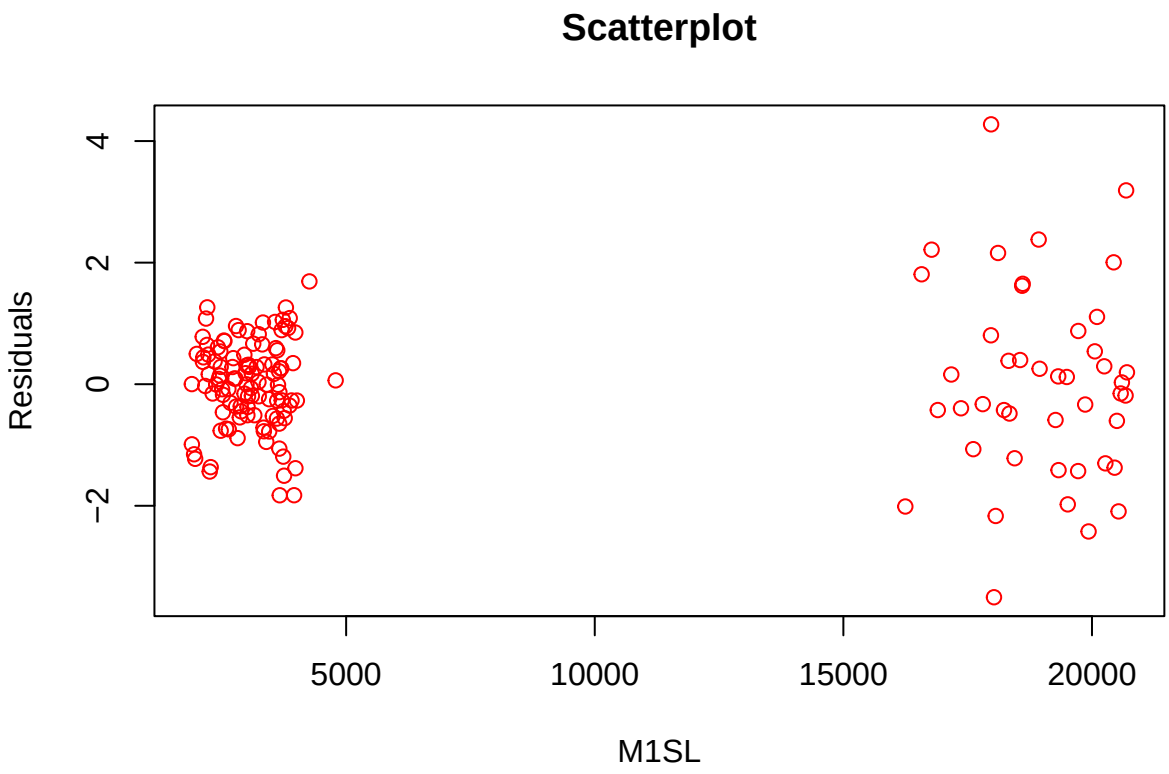

```
plot(data$Number.of.new.energy.policies[4:nrow(data)],resids,xlab="Number.of.new.energy.
```



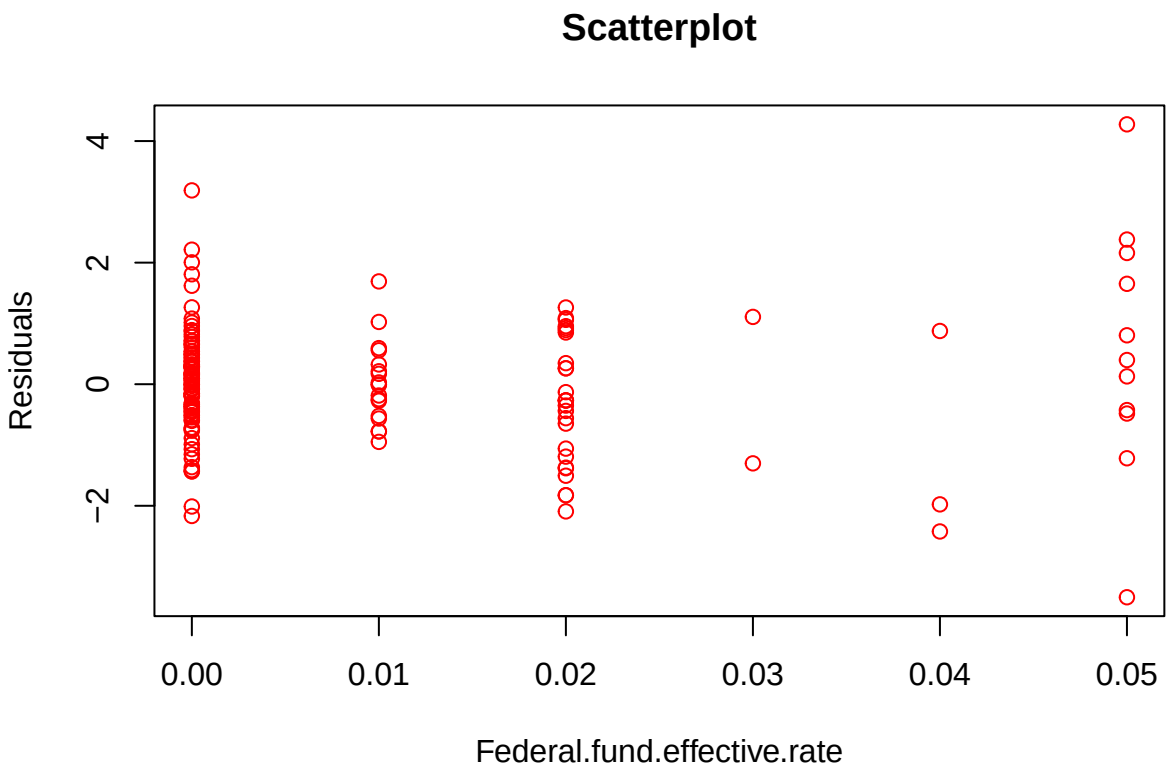

```
plot(data$Gasoline.LDV.sales[4:nrow(data)],resids,xlab="Gasoline.LDV.sales",ylab="Residu
```



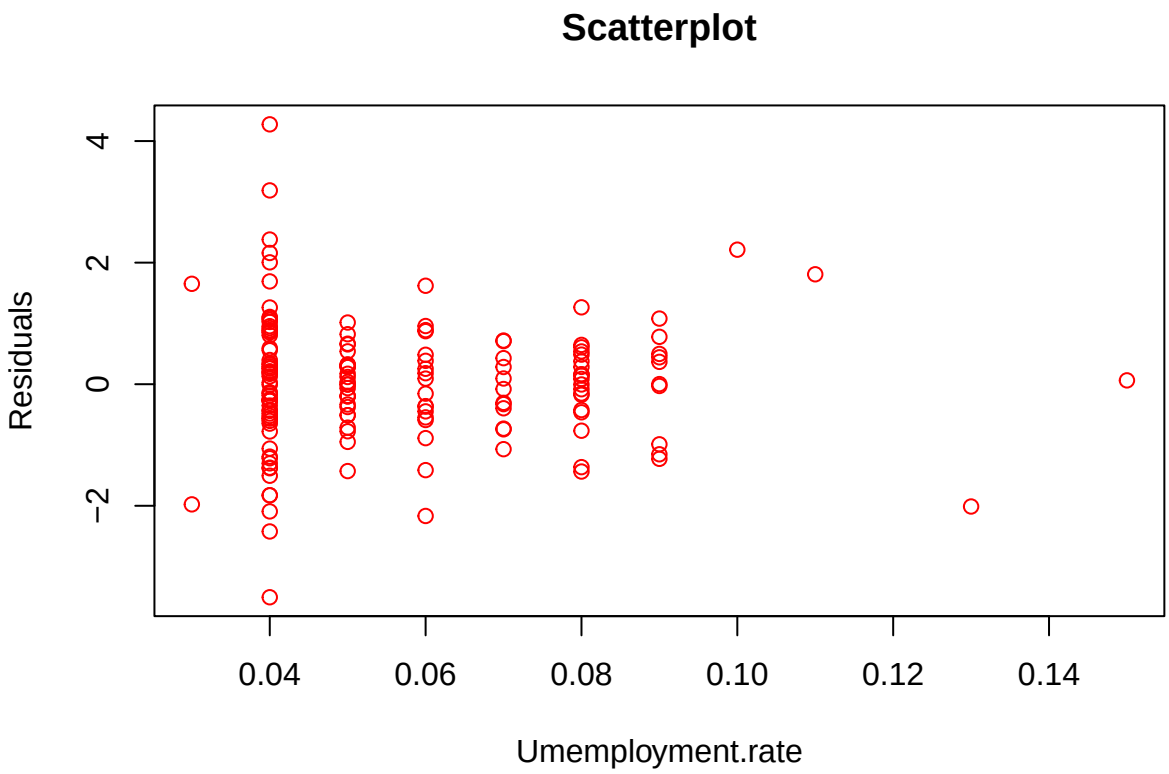

```
plot(data$M1SL[4:nrow(data)],resids,xlab="M1SL",ylab="Residuals",main="Scatterplot",col=
```



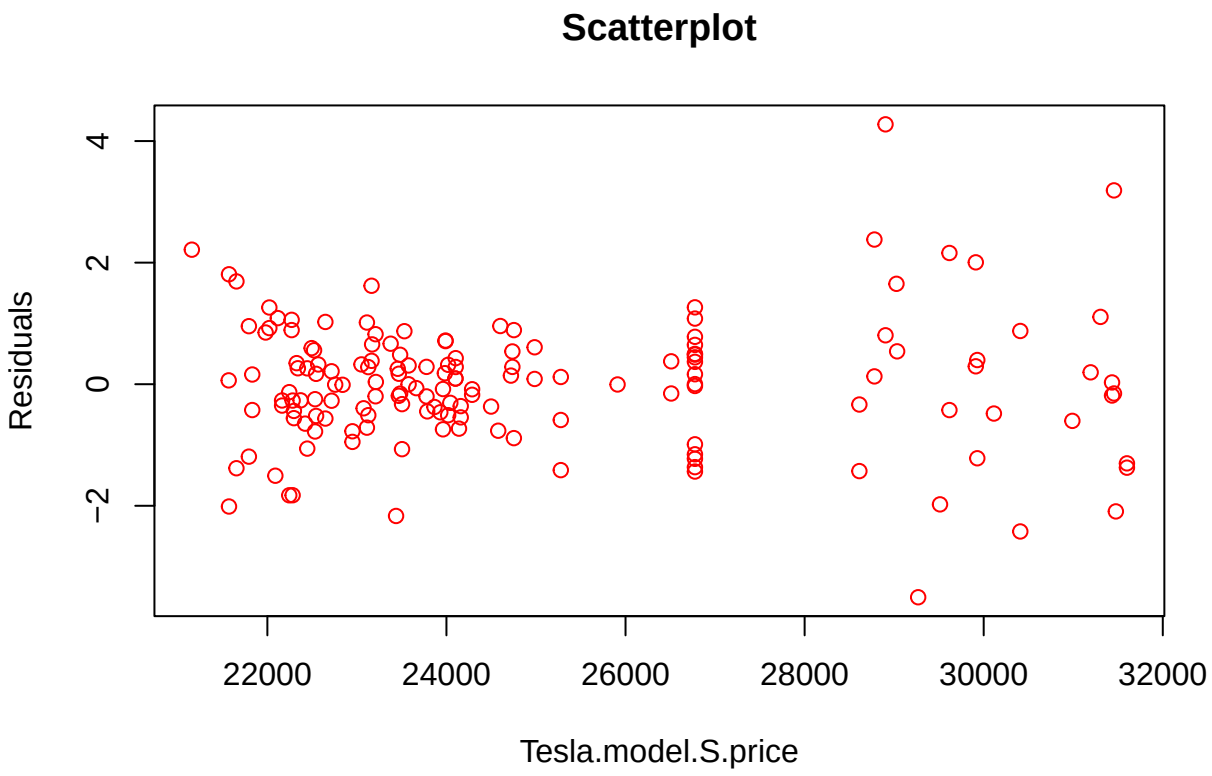

```
plot(data$Federal.fund.effective.rate[4:nrow(data)],resids,xlab="Federal.fund.effective.
```



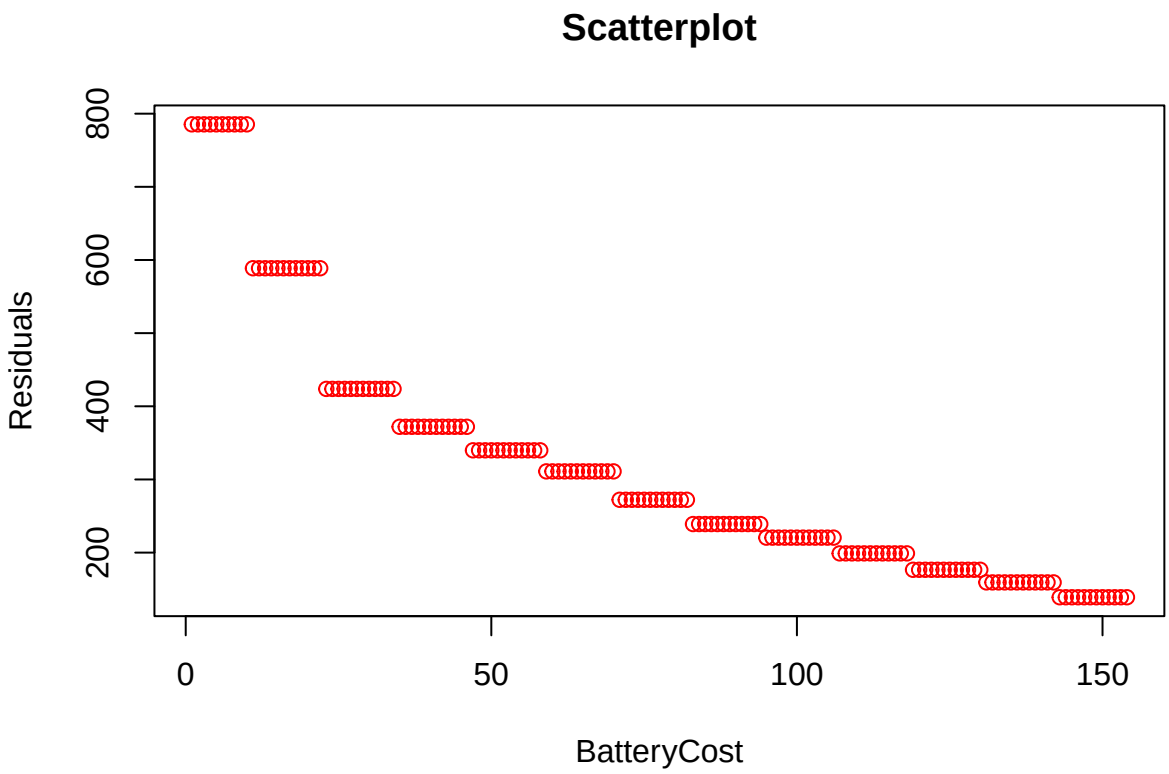

```
plot(data$Umemployment.rate[4:nrow(data)],resids,xlab="Umemployment.rate",ylab="Residual
```



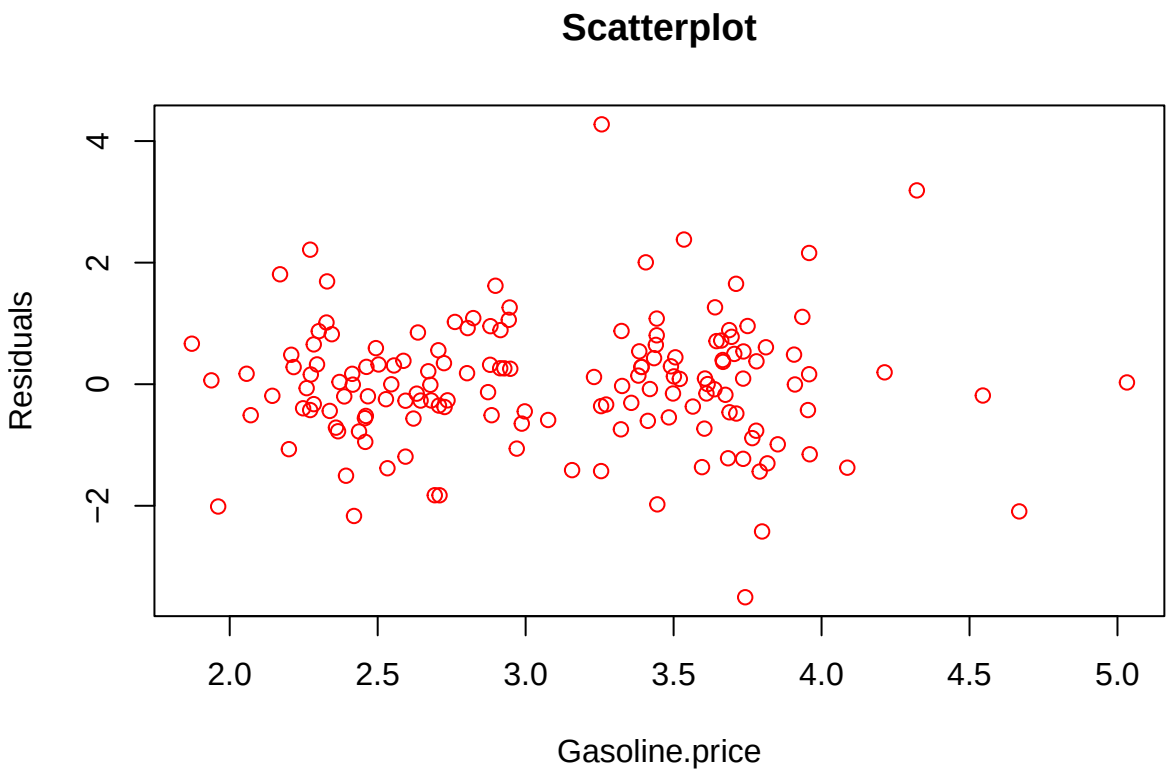

```
plot(data$Tesla.model.S.price[4:nrow(data)],resids,xlab="Tesla.model.S.price",ylab="Resi
```



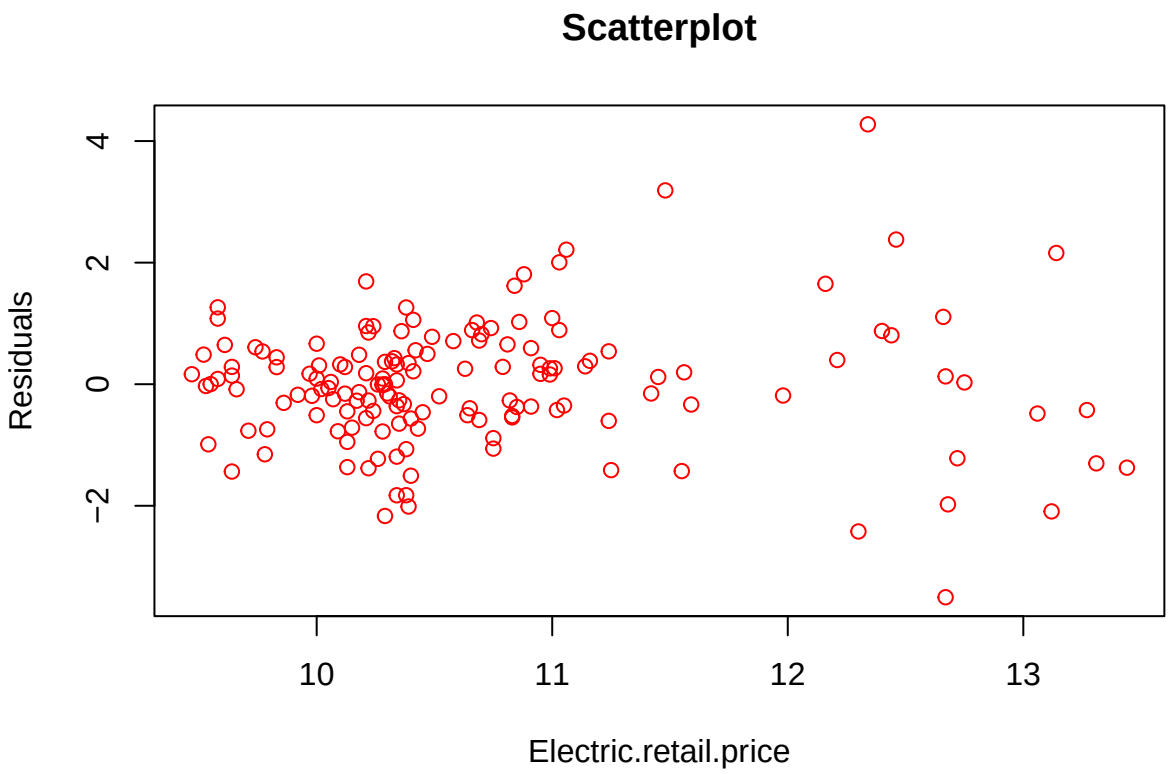

```
plot(data$BatteryCost[4:nrow(data)],xlab="BatteryCost",ylab="Residuals",main="Scatterplo
```



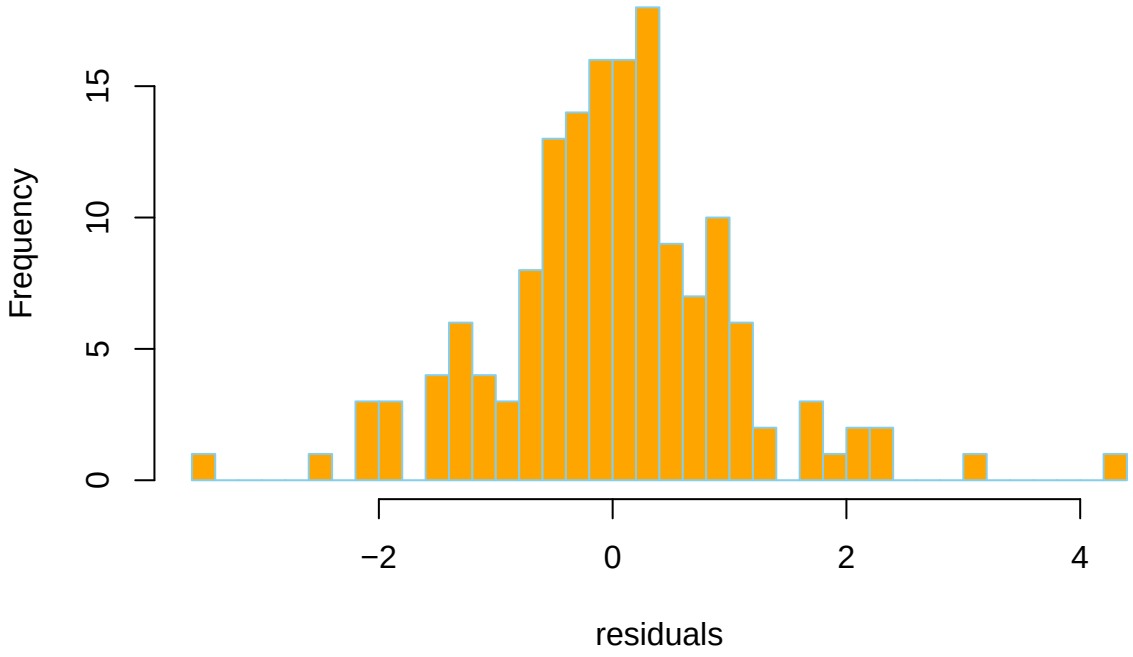

```
plot(data$Gasoline.price[4:nrow(data)],resids,xlab="Gasoline.price",ylab="Residuals",mai
```



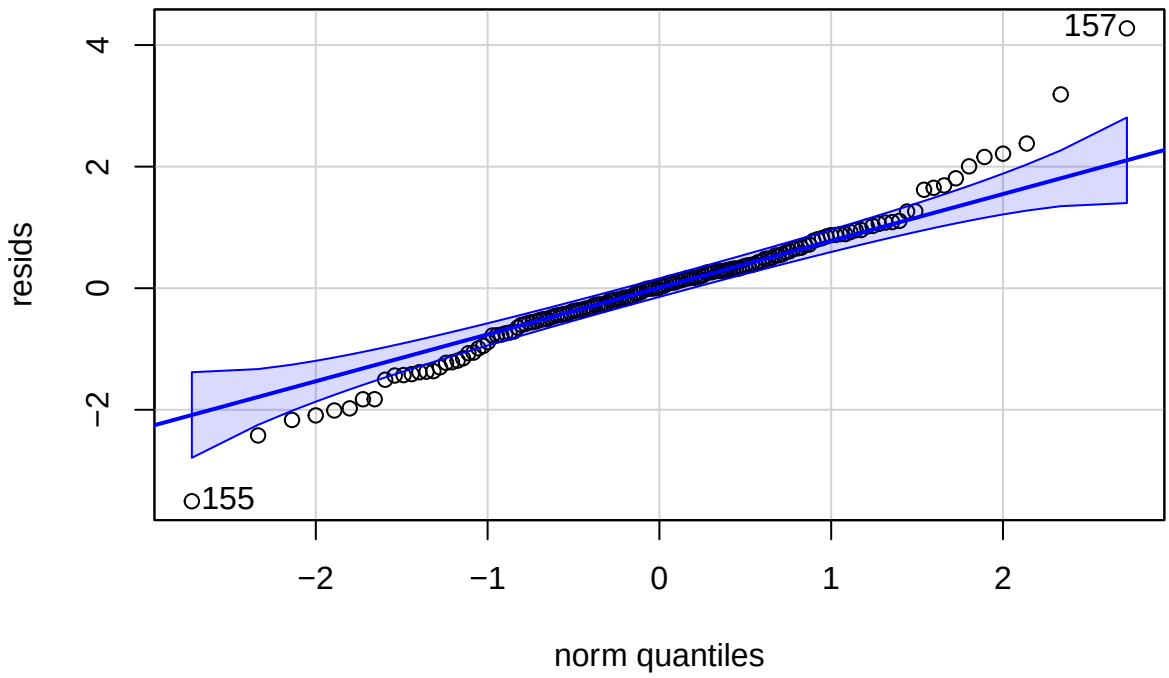

```
plot(data$Electric.retail.price[4:nrow(data)],resids,xlab="Electric.retail.price",ylab='
```




```
# Normality Assumption: hold
hist(resids,breaks = 30,main="",xlab="residuals",border = "skyblue",col="orange")
```




```
qqPlot(resids)
```




```
## 157 155
## 154 152
```

Goodness of fit: Some Outliers

```
cook=cooks.distance(lagged_model)
outlier_table <- data.frame(Standardized_Residuals=resids,Cooks_Distance = cook)
print(outlier_table)
```

```
##     Standardized_Residuals Cooks_Distance
## 4              0.001730881   2.078387e-08
## 5             -0.988794916   8.098317e-03
## 6             -1.152720975   9.388277e-03
## 7             -1.227472874   1.082723e-02
## 8              0.498604927   2.554984e-03
## 9              0.779782974   4.863038e-03
## 10             0.367627509   1.119733e-03
## 11             0.442222028   1.563929e-03
## 12             1.079963505   9.834924e-03
## 13            -0.029014095   6.186683e-06
## 14             0.646659129   2.591857e-03
## 15             1.264789428   8.031495e-03
```

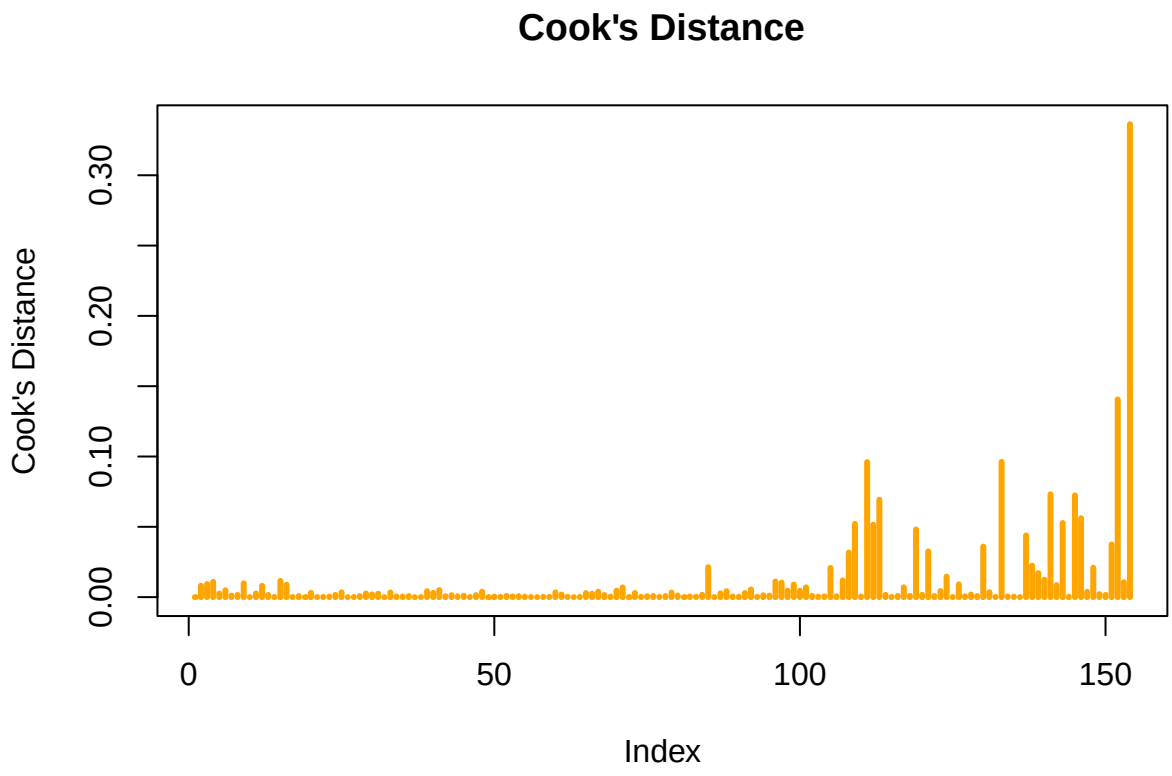
```
## 16            0.486908040   1.554683e-03
## 17            0.164212787   1.452780e-04
## 18           -1.436425682   1.152518e-02
## 19           -1.364028509   8.850997e-03
## 20           -0.152742905   1.161564e-04
## 21            0.376688155   7.668700e-04
## 22           -0.004370129   8.398940e-08
## 23            0.607671067   3.031759e-03
## 24            0.086397823   6.033840e-05
## 25            0.141823658   1.889679e-04
## 26            0.284957596   4.309222e-04
## 27            0.538863856   1.545594e-03
## 28           -0.764833835   3.368316e-03
## 29           -0.082806564   2.860297e-05
## 30           -0.174052199   1.237197e-04
## 31           -0.461695011   7.671713e-04
## 32            0.718436603   2.589570e-03
## 33            0.708920466   1.901912e-03
## 34           -0.730868790   2.416304e-03
## 35           -0.079493133   4.231583e-05
## 36           -0.742264829   3.307188e-03
## 37           -0.306557416   5.220389e-04
## 38            0.282885903   4.308713e-04
## 39            0.428900293   7.825739e-04
## 40            0.093843481   3.738577e-05
## 41            0.092311530   3.522389e-05
## 42            0.957261300   4.267924e-03
## 43           -0.886726820   3.029302e-03
## 44            0.890141036   5.028655e-03
## 45           -0.367328026   6.896800e-04
## 46           -0.545624435   1.407357e-03
## 47           -0.360386554   6.246327e-04
## 48           -0.446840526   1.015872e-03
## 49           -0.155125003   9.398998e-05
## 50            0.484438916   1.387810e-03
## 51            0.872473485   3.838758e-03
## 52           -0.001354384   1.123750e-08
## 53            0.308352260   4.196920e-04
## 54            0.181449116   1.928938e-04
## 55           -0.509499071   9.100296e-04
## 56            0.319434635   5.302491e-04
## 57           -0.374979504   6.827613e-04
## 58            0.285181395   3.170995e-04
## 59           -0.200785072   1.336213e-04
## 60           -0.064390725   1.342047e-05
```

```
## 61          -0.190575072   1.823678e-04
## 62           0.172297886   1.670319e-04
## 63           0.666545256   3.358143e-03
## 64          -0.509857222   1.845323e-03
## 65           0.281769356   3.016884e-04
## 66           0.037033392   4.893632e-06
## 67          -0.198610615   1.828594e-04
## 68           0.822549219   2.887697e-03
## 69           0.655250711   2.420904e-03
## 70           1.014777807   3.821256e-03
## 71          -0.714439962   1.573952e-03
## 72           0.325632911   4.041038e-04
## 73          -0.774041586   4.630368e-03
## 74          -0.948832227   6.907983e-03
## 75          -0.006682400   2.289971e-07
## 76          -0.777664601   2.908098e-03
## 77          -0.246253707   2.227533e-04
## 78           0.324726042   6.517159e-04
## 79          -0.521932153   8.180493e-04
## 80           0.170626077   2.383742e-04
## 81           0.593661818   7.594255e-04
## 82           1.024380769   3.237978e-03
## 83          -0.564791622   1.224030e-03
## 84          -0.009656212   4.042819e-07
## 85          -0.271617973   5.118950e-04
## 86           0.212883482   2.716805e-04
## 87           0.558551420   1.703413e-03
## 88          -1.827223384   2.129067e-02
## 89          -0.130457740   8.453493e-05
## 90          -0.648279888   2.693851e-03
## 91          -1.059419054   4.235291e-03
## 92           0.261056297   4.891802e-04
## 93           0.262423367   3.123203e-04
## 94           0.890054683   2.920306e-03
## 95           1.058998818   5.432728e-03
## 96          -0.265153025   2.335382e-04
## 97          -0.558360992   1.342547e-03
## 98          -0.439355042   1.066517e-03
## 99          -1.505484460   1.109104e-02
## 100         -1.191594441   1.029366e-02
## 101          0.954571097   4.656965e-03
## 102          1.262311792   8.918235e-03
## 103          0.923623075   4.527929e-03
## 104          1.087547626   6.931653e-03
## 105         -0.351197818   9.102641e-04
```

```
## 106          -0.267442607   3.541310e-04
## 107           0.346615790   5.918072e-04
## 108          -1.826001853   2.077991e-02
## 109          -0.268598586   6.425773e-04
## 110           0.850753126   1.191653e-02
## 111          -1.382078525   3.168898e-02
## 112           1.691401460   5.211264e-02
## 113           0.063124638   4.706428e-04
## 114          -2.011986010   9.602353e-02
## 115           1.808319140   5.151643e-02
## 116           2.213475582   6.924829e-02
## 117          -0.424873329   1.650048e-03
## 118           0.159754295   1.997854e-04
## 119          -0.395996383   8.932357e-04
## 120          -1.068671880   6.988187e-03
## 121          -0.327605476   9.401444e-04
## 122          -2.167391254   4.806715e-02
## 123           0.384696828   1.657604e-03
## 124           1.619764738   3.250021e-02
## 125           0.253985673   7.794773e-04
## 126          -0.588788879   4.320743e-03
## 127          -1.413304928   1.459515e-02
## 128           0.118963014   1.158233e-04
## 129          -1.428957434   9.058827e-03
## 130          -0.334052172   6.400536e-04
## 131           0.540737403   1.830320e-03
## 132           0.294574737   7.321892e-04
## 133           2.004860513   3.590477e-02
## 134          -0.602924137   3.399034e-03
## 135          -0.152270562   2.017406e-04
## 136           3.188569181   9.619568e-02
## 137           0.195182398   4.784648e-04
## 138          -0.185713020   3.504129e-04
## 139           0.029054829   1.051316e-05
## 140          -2.092431920   4.384706e-02
## 141          -1.373131728   2.238077e-02
## 142          -1.302037155   1.707928e-02
## 143           1.107292372   1.240538e-02
## 144          -2.422540248   7.312728e-02
## 145           0.875715966   8.503295e-03
## 146          -1.976694108   5.272610e-02
## 147           0.128723708   2.985956e-04
## 148           2.379898536   7.235181e-02
## 149           1.651648506   5.606962e-02
## 150           0.397926013   3.759211e-03
```

```
## 151           -1.218306004   2.091852e-02
## 152           -0.482418688   2.111637e-03
## 153           -0.425270620   1.572610e-03
## 154            2.158928322   3.740619e-02
## 155           -3.504525836   1.406168e-01
## 156            0.804657134   1.061166e-02
## 157            4.275133388   3.362954e-01
```

```
plot(cook,type="h",lwd=3,col="orange",ylab="Cook's Distance",main="Cook's Distance")
```




```
condition <- (abs(resids) > 2) & (cook > 4 / 157)
count_rows <- sum(condition)
count_rows/157
```

```
## [1] 0.07006369
```

BoxCox MLR

```
# Box-Cox transformation
library(MASS)
```

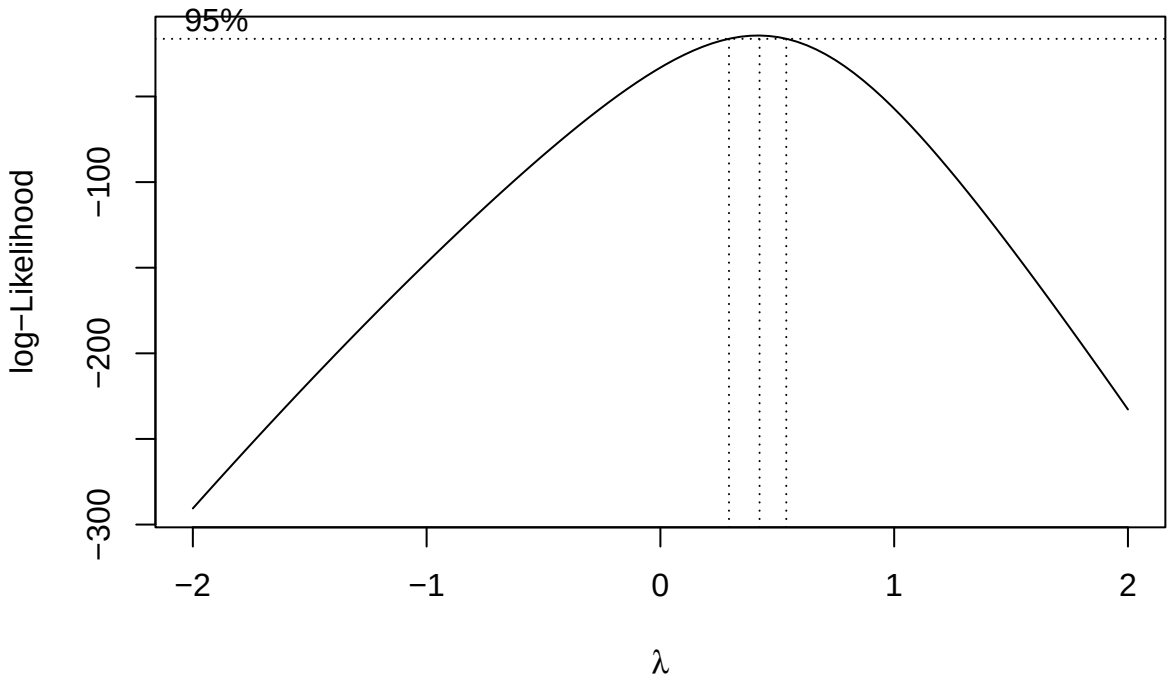
```
##
```

```
##     'MASS'
```

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

```
## The following object is masked from 'package:plotly':
##
##     select
```

```
boxcox_result <- boxcox(full_model, plotit = TRUE)
```




```
best_lambda <- boxcox_result$x[which.max(boxcox_result$y)]
print(best_lambda)
```

```
## [1] 0.4242424
```

```
# If lambda = 0, use log transformation, otherwise use (y^lambda - 1) / lambda
model_boxcox <- lm(((Total.NEV.Sales^best_lambda) - 1) / best_lambda ~ ., data = data)
summary(model_boxcox)
```

```
##
## Call:
## lm(formula = ((Total.NEV.Sales^best_lambda) - 1)/best_lambda ~
```

```
##     ., data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.1193  -4.9063  -0.0071   5.9883  25.3090
##
## Coefficients:
##                                                                          Estimate
## (Intercept)                                                             1.298e+03
## Month                                                                   1.531e+00
## Date                                                                    2.772e-01
## Covid.19                                                               -6.494e+00
## Number.of.new.energy.policies                                          -6.342e-02
## Production                                                             -9.753e-06
## Inventory                                                               6.016e-05
## Gasoline.LDV.sales                                                      6.175e-05
## M1SL                                                                   -3.752e-03
## M2SL                                                                   -1.266e-03
## Federal.fund.effective.rate                                             9.869e+02
## Umemployment.rate                                                      -1.455e+02
## PPIBC                                                                   4.446e-01
## PPIBM                                                                   1.397e-01
## PCE                                                                     4.624e-02
## GDP                                                                    -2.351e-02
## Durable.goods.consumption                                               2.024e-01
## Nondurable.goods.consumption                                           -6.756e-02
## Nonresidential.fixed.Investment                                        -4.324e-02
## Residential.fixed.Investment                                           -1.340e-01
## Tesla.model.S.price                                                     5.022e-03
## ElectricPorts                                                           1.536e-03
## BatteryCost                                                            -1.192e-01
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education -4.502e+02
## Percentage.of.employees.who.are.middle.aged..25.55.                    -3.308e+02
## Population                                                             -3.780e-02
## Per.capita.disposable.income                                           -8.256e-04
## Gasoline.price                                                          1.602e+01
## Electric.retail.price                                                  -4.115e+00
##                                                                        Std. Error
## (Intercept)                                                             1.272e+03
## Month                                                                   3.270e-01
## Date                                                                    1.019e-01
## Covid.19                                                                7.125e+00
## Number.of.new.energy.policies                                           3.699e-02
## Production                                                              2.634e-05
## Inventory                                                               1.812e-05
```

```
## Gasoline.LDV.sales                                                    7.729e-06
## M1SL                                                                  1.907e-03
## M2SL                                                                  1.045e-02
## Federal.fund.effective.rate                                           3.762e+02
## Umemployment.rate                                                     1.545e+02
## PPIBC                                                                 7.196e-01
## PPIBM                                                                 7.528e-01
## PCE                                                                   3.277e-02
## GDP                                                                   1.650e-02
## Durable.goods.consumption                                             6.194e-02
## Nondurable.goods.consumption                                          8.059e-02
## Nonresidential.fixed.Investment                                       4.954e-02
## Residential.fixed.Investment                                          8.153e-02
## Tesla.model.S.price                                                   2.269e-03
## ElectricPorts                                                         6.151e-04
## BatteryCost                                                           3.249e-02
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education  2.641e+02
## Percentage.of.employees.who.are.middle.aged..25.55.                   3.347e+02
## Population                                                            1.073e-02
## Per.capita.disposable.income                                          1.634e-03
## Gasoline.price                                                        5.386e+00
## Electric.retail.price                                                 2.776e+00
##                                                                      t value
## (Intercept)                                                            1.020
## Month                                                                  4.682
## Date                                                                   2.720
## Covid.19                                                              -0.911
## Number.of.new.energy.policies                                         -1.714
## Production                                                            -0.370
## Inventory                                                              3.320
## Gasoline.LDV.sales                                                     7.990
## M1SL                                                                  -1.967
## M2SL                                                                  -0.121
## Federal.fund.effective.rate                                            2.623
## Umemployment.rate                                                     -0.942
## PPIBC                                                                  0.618
## PPIBM                                                                  0.186
## PCE                                                                    1.411
## GDP                                                                   -1.424
## Durable.goods.consumption                                              3.268
## Nondurable.goods.consumption                                          -0.838
## Nonresidential.fixed.Investment                                       -0.873
## Residential.fixed.Investment                                          -1.644
## Tesla.model.S.price                                                    2.214
## ElectricPorts                                                          2.497
```

```
## BatteryCost                                                          -3.670
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education  -1.705
## Percentage.of.employees.who.are.middle.aged..25.55.                   -0.988
## Population                                                           -3.522
## Per.capita.disposable.income                                         -0.505
## Gasoline.price                                                        2.974
## Electric.retail.price                                                -1.482
##                                                                     Pr(>|t|)
## (Intercept)                                                         0.309459
## Month                                                               7.14e-06
## Date                                                                0.007435
## Covid.19                                                            0.363789
## Number.of.new.energy.policies                                       0.088880
## Production                                                          0.711829
## Inventory                                                           0.001172
## Gasoline.LDV.sales                                                  6.82e-13
## M1SL                                                                0.051320
## M2SL                                                                0.903708
## Federal.fund.effective.rate                                         0.009761
## Umemployment.rate                                                   0.348104
## PPIBC                                                               0.537778
## PPIBM                                                               0.853059
## PCE                                                                 0.160723
## GDP                                                                 0.156780
## Durable.goods.consumption                                           0.001391
## Nondurable.goods.consumption                                        0.403425
## Nonresidential.fixed.Investment                                     0.384383
## Residential.fixed.Investment                                        0.102630
## Tesla.model.S.price                                                 0.028631
## ElectricPorts                                                       0.013786
## BatteryCost                                                         0.000355
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education 0.090668
## Percentage.of.employees.who.are.middle.aged..25.55.                  0.324973
## Population                                                          0.000595
## Per.capita.disposable.income                                        0.614190
## Gasoline.price                                                      0.003514
## Electric.retail.price                                               0.140758
##
## (Intercept)
## Month                                                               ***
## Date                                                                **
## Covid.19
## Number.of.new.energy.policies                                       .
## Production
## Inventory                                                           **
```

```
## Gasoline.LDV.sales                                                    ***
## M1SL                                                                  .
## M2SL
## Federal.fund.effective.rate                                           **
## Umemployment.rate
## PPIBC
## PPIBM
## PCE
## GDP
## Durable.goods.consumption                                             **
## Nondurable.goods.consumption
## Nonresidential.fixed.Investment
## Residential.fixed.Investment
## Tesla.model.S.price                                                   *
## ElectricPorts                                                         *
## BatteryCost                                                           ***
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education .
## Percentage.of.employees.who.are.middle.aged..25.55.
## Population                                                            ***
## Per.capita.disposable.income
## Gasoline.price                                                        **
## Electric.retail.price
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.26 on 128 degrees of freedom
## Multiple R-squared:  0.984,  Adjusted R-squared:  0.9805
## F-statistic: 280.9 on 28 and 128 DF,  p-value: < 2.2e-16
```

```
library(car)
vif(model_boxcox)
```

```
##                                                      Month
##                                                   1.920368
##                                                       Date
##                                                29501.153805
##                                                   Covid.19
##                                                  14.618061
##                                 Number.of.new.energy.policies
##                                                   6.767884
##                                                 Production
##                                                   8.927369
##                                                  Inventory
##                                                  78.038416
##                                          Gasoline.LDV.sales
```

```
##                                                                 3.557467
##                                                                     M1SL
##                                                                282.814081
##                                                                     M2SL
##                                                               2759.810876
##                                                 Federal.fund.effective.rate
##                                                                44.756856
##                                                       Umemployment.rate
##                                                                14.968166
##                                                                     PPIBC
##                                                                15.339518
##                                                                     PPIBM
##                                                               202.623852
##                                                                      PCE
##                                                               193.148413
##                                                                      GDP
##                                                              5172.768085
##                                                   Durable.goods.consumption
##                                                                670.782082
##                                                 Nondurable.goods.consumption
##                                                               2342.557766
##                                              Nonresidential.fixed.Investment
##                                                               1002.806743
##                                                 Residential.fixed.Investment
##                                                                655.749898
##                                                         Tesla.model.S.price
##                                                                  63.749847
##                                                               ElectricPorts
##                                                                1344.055530
##                                                                BatteryCost
##                                                                  55.226187
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education
##                                                                91.605819
##                  Percentage.of.employees.who.are.middle.aged..25.55.
##                                                                24.533040
##                                                                Population
##                                                               9287.203160
##                                              Per.capita.disposable.income
##                                                                221.534172
##                                                             Gasoline.price
##                                                                  17.426933
##                                                    Electric.retail.price
##                                                                  9.631857
```

```r
data_boxcox <- subset(data, select = -c(Date,M1SL,M2SL,GDP,Nondurable.goods.consumption,
data_boxcox0<-data_boxcox
model_boxcox2 <- lm(((Total.NEV.Sales^best_lambda) - 1) / best_lambda ~ ., data = data_b
summary(model_boxcox2)
```

```
##
## Call:
## lm(formula = ((Total.NEV.Sales^best_lambda) - 1)/best_lambda ~
##     ., data = data_boxcox)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.628  -8.070   0.667   7.233  61.779
##
## Coefficients:
##                                                  Estimate Std. Error
## (Intercept)                                    -2.115e+02  3.036e+02
## Month                                           1.577e+00  4.301e-01
## Covid.19                                       -7.225e+00  6.970e+00
## Number.of.new.energy.policies                  -1.184e-01  4.739e-02
## Production                                      4.312e-05  3.262e-05
## Inventory                                       1.688e-05  1.815e-05
## Gasoline.LDV.sales                              4.868e-05  1.051e-05
## Federal.fund.effective.rate                     1.424e+03  1.796e+02
## Umemployment.rate                               7.793e+02  1.116e+02
## PPIBC                                          -5.618e-01  8.978e-01
## PCE                                             2.052e-01  1.865e-02
## Tesla.model.S.price                             5.312e-03  1.384e-03
## BatteryCost                                    -7.419e-02  3.127e-02
## Percentage.of.employees.who.are.middle.aged..25.55.  2.466e+01  4.317e+02
## Gasoline.price                                  2.206e+01  3.880e+00
## Electric.retail.price                          -9.746e+00  3.667e+00
##                                                t value Pr(>|t|)
## (Intercept)                                     -0.697 0.487176
## Month                                            3.666 0.000348 ***
## Covid.19                                        -1.037 0.301642
## Number.of.new.energy.policies                   -2.498 0.013639 *
## Production                                       1.322 0.188309
## Inventory                                        0.930 0.354127
## Gasoline.LDV.sales                               4.630 8.22e-06 ***
## Federal.fund.effective.rate                      7.925 6.25e-13 ***
## Umemployment.rate                                6.983 1.05e-10 ***
## PPIBC                                           -0.626 0.532499
## PCE                                             11.003  < 2e-16 ***
```

```
## Tesla.model.S.price                                    3.839 0.000186 ***
## BatteryCost                                           -2.373 0.018997 *
## Percentage.of.employees.who.are.middle.aged..25.55.    0.057 0.954524
## Gasoline.price                                         5.685 7.24e-08 ***
## Electric.retail.price                                 -2.657 0.008783 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.72 on 141 degrees of freedom
## Multiple R-squared:  0.9586, Adjusted R-squared:  0.9542
## F-statistic: 217.5 on 15 and 141 DF,  p-value: < 2.2e-16
```

```
Date <- data$Date
model_date_included <- lm(((Total.NEV.Sales^best_lambda) - 1) / best_lambda ~ . + I(Date
anova_result <- anova(model_boxcox2,model_date_included)
print(anova_result)
```

```
## Analysis of Variance Table
##
## Model 1: ((Total.NEV.Sales^best_lambda) - 1)/best_lambda ~ Month + Covid.19 +
##     Number.of.new.energy.policies + Production + Inventory +
##     Gasoline.LDV.sales + Federal.fund.effective.rate + Umemployment.rate +
##     PPIBC + PCE + Tesla.model.S.price + BatteryCost + Percentage.of.employees.who.are
##     Gasoline.price + Electric.retail.price
## Model 2: ((Total.NEV.Sales^best_lambda) - 1)/best_lambda ~ Month + Covid.19 +
##     Number.of.new.energy.policies + Production + Inventory +
##     Gasoline.LDV.sales + Federal.fund.effective.rate + Umemployment.rate +
##     PPIBC + PCE + Tesla.model.S.price + BatteryCost + Percentage.of.employees.who.are
##     Gasoline.price + Electric.retail.price + I(Date)
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    141 34847
## 2    140 34837  1    9.2838 0.0373 0.8471
```

```
## Reject the null hypothesis, which means we shouldn't keep date in the boxcox transf
```

```
library(car)
vif(model_boxcox2)
```

```
##                                              Month
##                                           1.413901
##                                           Covid.19
##                                           5.954011
##                          Number.of.new.energy.policies
##                                           4.727917
##                                         Production
##                                           5.825524
```

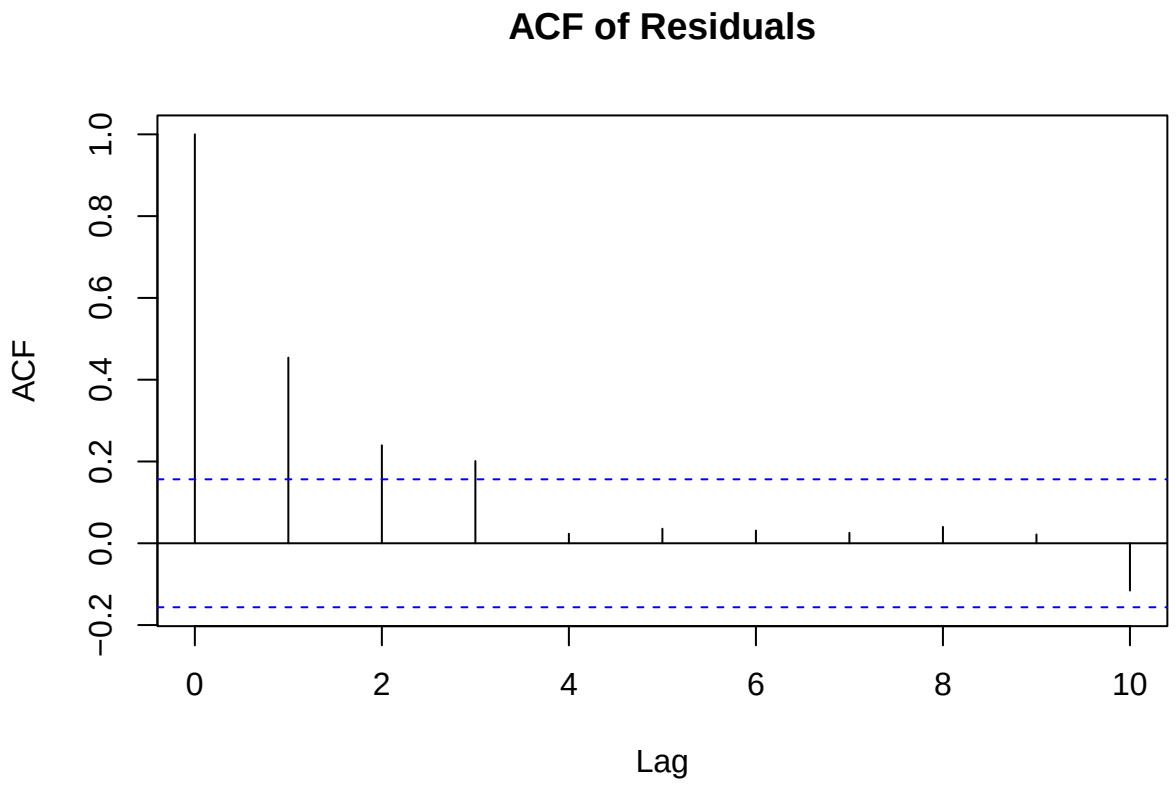
```
##                                         Inventory
##                                         33.351361
##                                 Gasoline.LDV.sales
##                                          2.802633
##                           Federal.fund.effective.rate
##                                          4.345362
##                                 Umemployment.rate
##                                          3.322561
##                                             PPIBC
##                                         10.165497
##                                               PCE
##                                         26.632445
##                                 Tesla.model.S.price
##                                         10.094932
##                                       BatteryCost
##                                         21.775594
## Percentage.of.employees.who.are.middle.aged..25.55.
##                                         17.370260
##                                   Gasoline.price
##                                          3.849206
##                                Electric.retail.price
##                                          7.153222
```

Autocorrelation

```
library(lmtest)
dwtest(model_boxcox2)
```

```
##
##  Durbin-Watson test
##
## data:  model_boxcox2
## DW = 1.0071, p-value = 2.163e-13
## alternative hypothesis: true autocorrelation is greater than 0
```

```
residuals_lagged <- residuals(model_boxcox2)
acf(residuals_lagged, main = "ACF of Residuals", lag.max = 10)
```




```
# From the ACF of Residuals, the residual autocorrelation at lag orders 1, 2 and 3 is
library(dplyr)

data_boxcox <- data_boxcox %>%
  mutate(
    lagged_value1 = ((lag(Total.NEV.Sales, n = 1))^best_lambda - 1) / best_lambda,
    lagged_value2 = ((lag(Total.NEV.Sales, n = 2))^best_lambda - 1) / best_lambda,
    lagged_value3 = ((lag(Total.NEV.Sales, n = 3))^best_lambda - 1) / best_lambda
  )
# test if lagged variables solve the autocorrelation problem
lagged_model2 <-  lm(((Total.NEV.Sales^best_lambda) - 1) / best_lambda ~ ., data = data_
dwtest(lagged_model2)
```

```
##
##  Durbin-Watson test
##
## data:  lagged_model2
## DW = 1.8235, p-value = 0.01811
## alternative hypothesis: true autocorrelation is greater than 0
```

```r
summary(lagged_model2)
```

```
##
## Call:
## lm(formula = ((Total.NEV.Sales^best_lambda) - 1)/best_lambda ~
##     ., data = data_boxcox)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.368  -6.616   0.174   6.915  33.049
##
## Coefficients:
##                                                   Estimate Std. Error
## (Intercept)                                     -3.651e+01  2.255e+02
## Month                                            5.022e-01  3.306e-01
## Covid.19                                         7.873e+00  5.324e+00
## Number.of.new.energy.policies                   -5.218e-02  3.433e-02
## Production                                       2.938e-05  2.433e-05
## Inventory                                        1.436e-05  1.375e-05
## Gasoline.LDV.sales                               7.343e-05  7.841e-06
## Federal.fund.effective.rate                      8.083e+02  1.389e+02
## Umemployment.rate                                5.483e+02  8.258e+01
## PPIBC                                            1.285e-01  6.419e-01
## PCE                                              1.038e-01  1.637e-02
## Tesla.model.S.price                              2.477e-03  1.026e-03
## BatteryCost                                      1.405e-02  2.869e-02
## Percentage.of.employees.who.are.middle.aged..25.55. -2.839e+02  3.260e+02
## Gasoline.price                                   1.093e+01  3.005e+00
## Electric.retail.price                           -8.083e+00  2.646e+00
## lagged_value1                                    3.494e-01  5.892e-02
## lagged_value2                                    7.277e-02  5.991e-02
## lagged_value3                                    2.369e-01  5.316e-02
##                                                 t value Pr(>|t|)
## (Intercept)                                      -0.162 0.871612
## Month                                             1.519 0.131028
## Covid.19                                          1.479 0.141530
## Number.of.new.energy.policies                    -1.520 0.130872
## Production                                        1.208 0.229268
## Inventory                                         1.044 0.298196
## Gasoline.LDV.sales                                9.365 2.29e-16 ***
## Federal.fund.effective.rate                       5.821 4.05e-08 ***
## Umemployment.rate                                 6.640 7.03e-10 ***
## PPIBC                                             0.200 0.841607
## PCE                                               6.342 3.17e-09 ***
```

```
## Tesla.model.S.price                                    2.414 0.017113 *
## BatteryCost                                            0.490 0.625175
## Percentage.of.employees.who.are.middle.aged..25.55.   -0.871 0.385327
## Gasoline.price                                         3.639 0.000389 ***
## Electric.retail.price                                 -3.055 0.002710 **
## lagged_value1                                          5.929 2.41e-08 ***
## lagged_value2                                          1.215 0.226641
## lagged_value3                                          4.457 1.73e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.15 on 135 degrees of freedom
##   (   3    )
## Multiple R-squared:  0.9794, Adjusted R-squared:  0.9767
## F-statistic: 357.4 on 18 and 135 DF,  p-value: < 2.2e-16
```

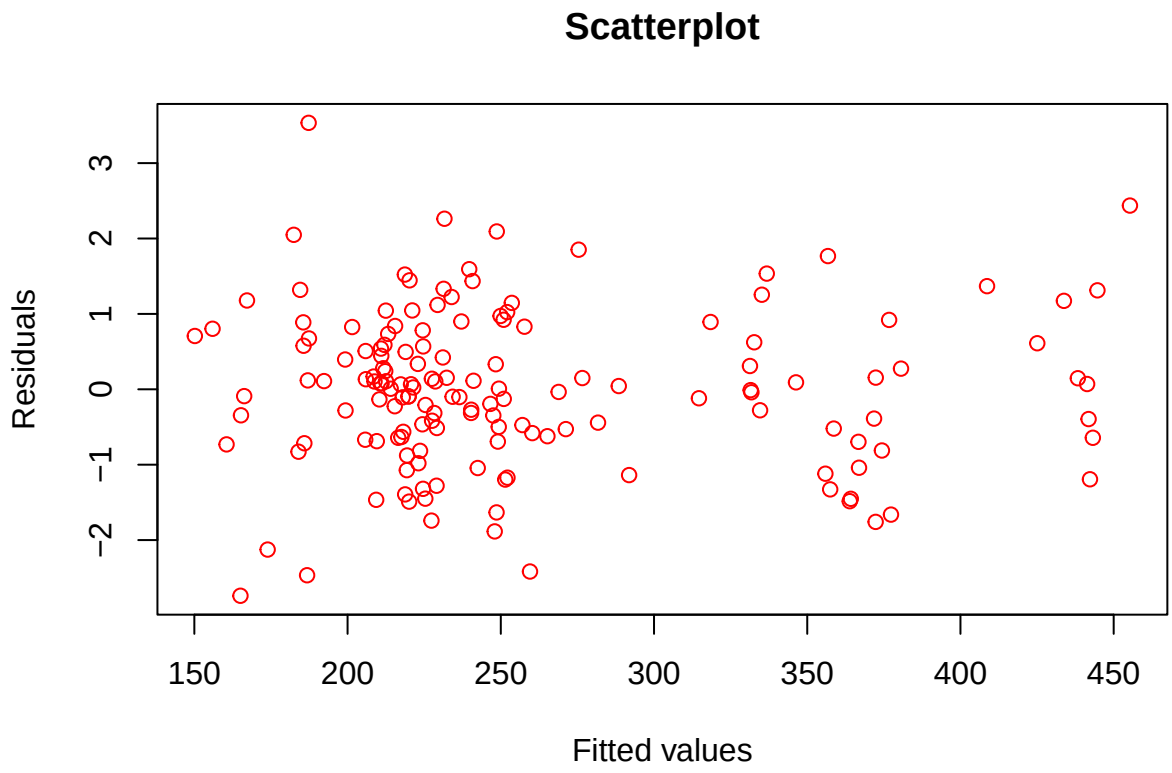
```
library(car)
vif(lagged_model2)
```

```
##                                              Month
##                                           1.588687
##                                            Covid.19
##                                           6.857932
##                          Number.of.new.energy.policies
##                                           4.879246
##                                         Production
##                                           6.398033
##                                          Inventory
##                                           38.037019
##                                    Gasoline.LDV.sales
##                                           2.958084
##                             Federal.fund.effective.rate
##                                           5.119534
##                                    Umemployment.rate
##                                           3.438718
##                                              PPIBC
##                                           10.281360
##                                                PCE
##                                           39.598421
##                                  Tesla.model.S.price
##                                           10.946846
##                                         BatteryCost
##                                           30.114201
## Percentage.of.employees.who.are.middle.aged..25.55.
##                                           17.564326
```

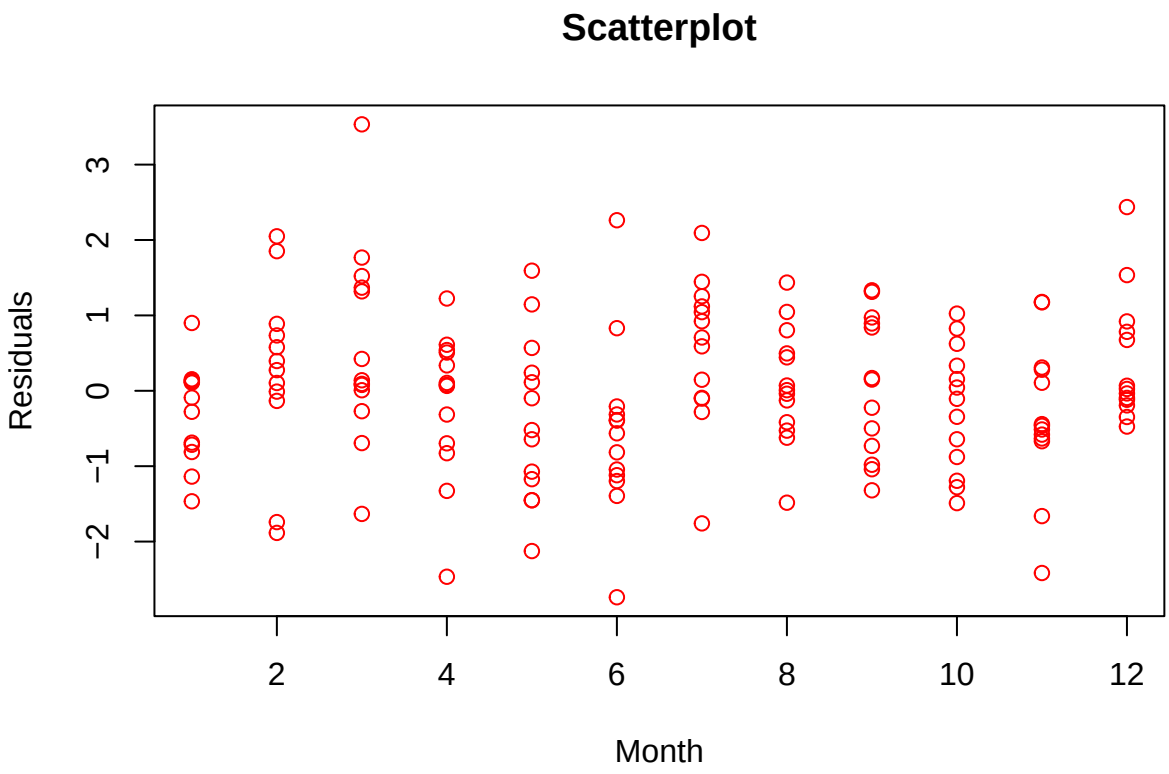
```
##                                  Gasoline.price
##                                        4.585222
##                             Electric.retail.price
##                                        7.169106
##                                   lagged_value1
##                                       21.643336
##                                   lagged_value2
##                                       21.612604
##                                   lagged_value3
##                                       16.424033
```

Goodness of fit

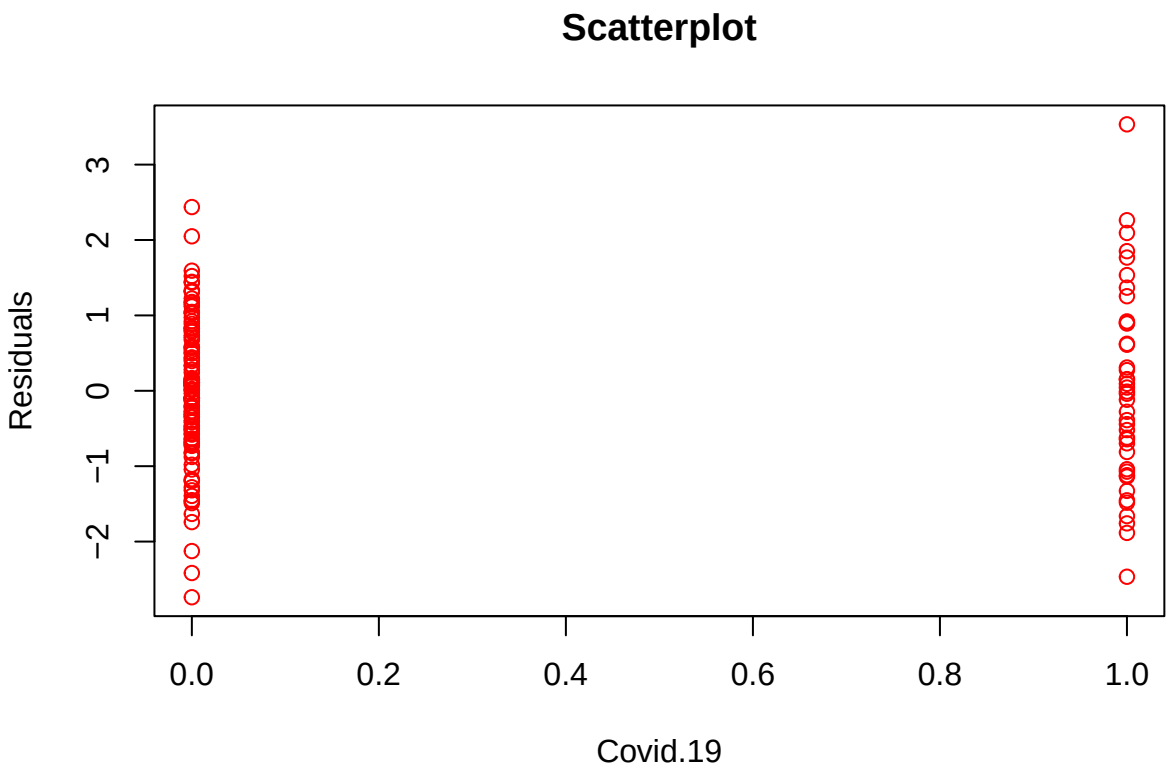
```
# Constant Variance Assumption: hold
resids=rstandard(lagged_model2)
fits=lagged_model2$fitted
plot(fits,resids,xlab="Fitted values",ylab="Residuals",main="Scatterplot",col="red")
```



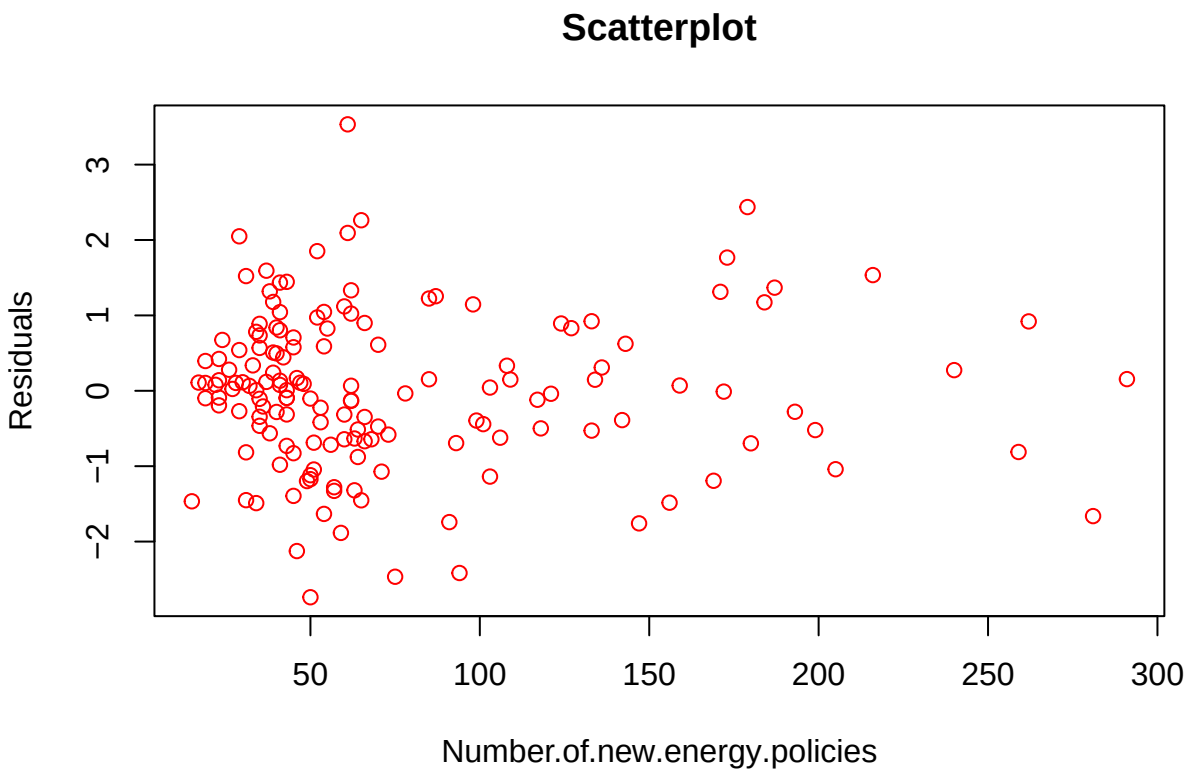

```
# Linearity Assumption: hold
# following codes should be checked according to the variable selection results in cas
plot(data$Month[4:nrow(data)],resids,xlab="Month",ylab="Residuals",main="Scatterplot",co
```



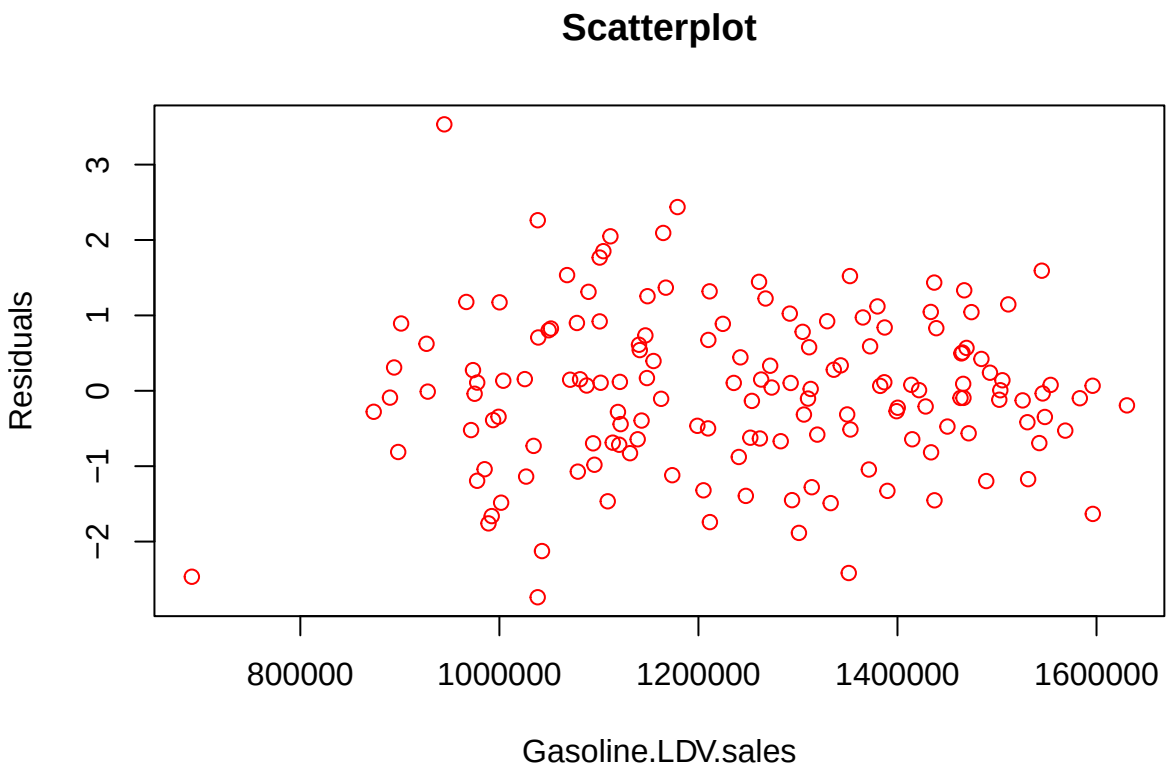

```
plot(data$Covid.19[4:nrow(data)],resids,xlab="Covid.19",ylab="Residuals",main="Scatterpl
```



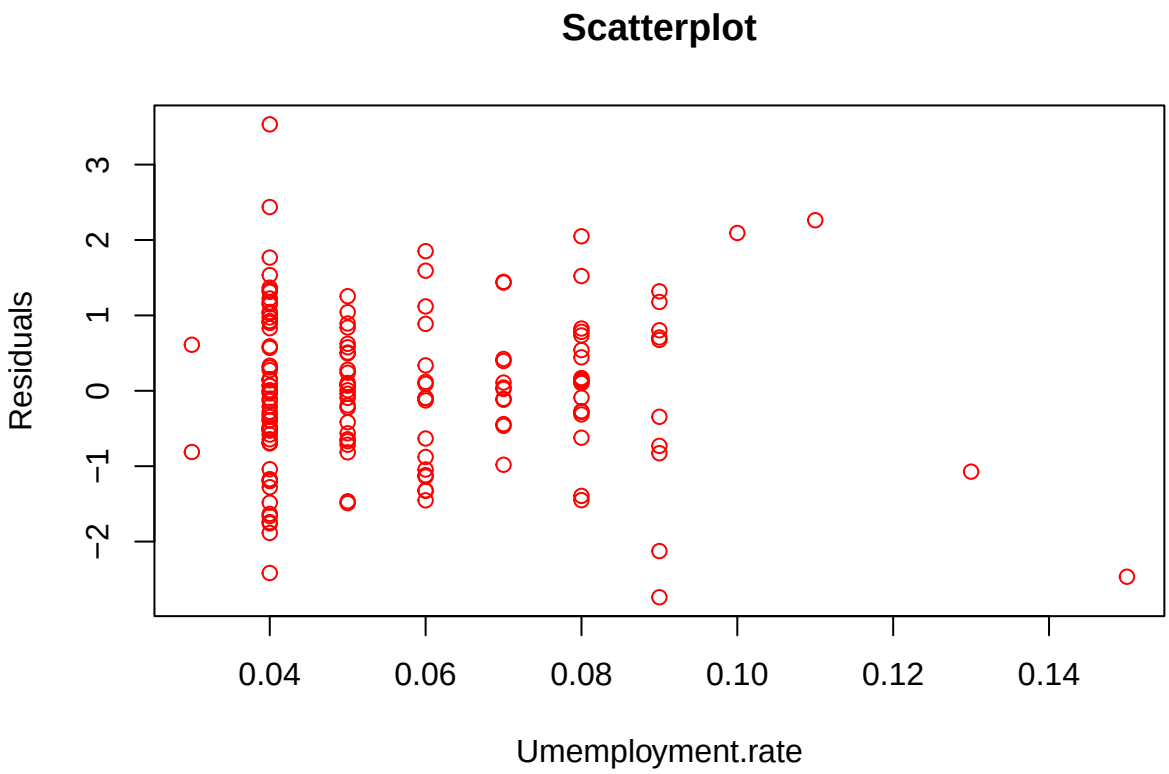

```
plot(data$Number.of.new.energy.policies[4:nrow(data)],resids,xlab="Number.of.new.energy.
```



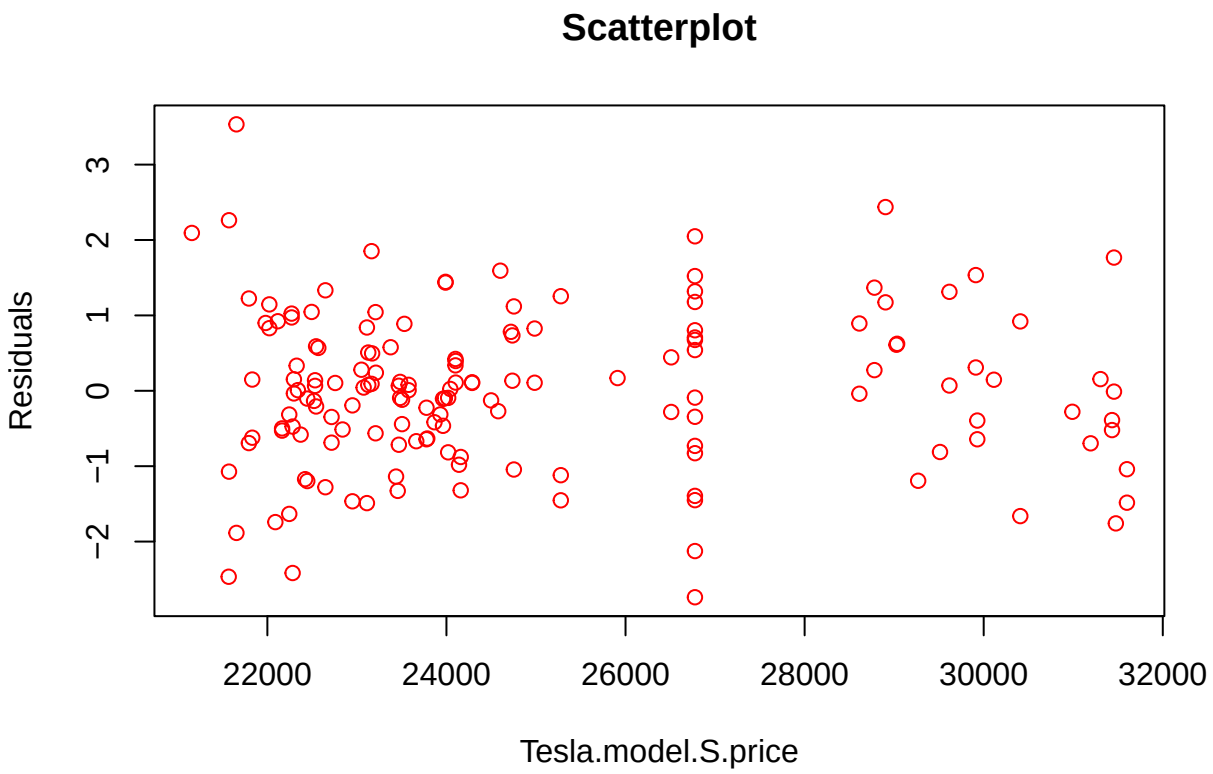

```
plot(data$Gasoline.LDV.sales[4:nrow(data)],resids,xlab="Gasoline.LDV.sales",ylab="Residu
```




```
plot(data$Umemployment.rate[4:nrow(data)],resids,xlab="Umemployment.rate",ylab="Residual
```



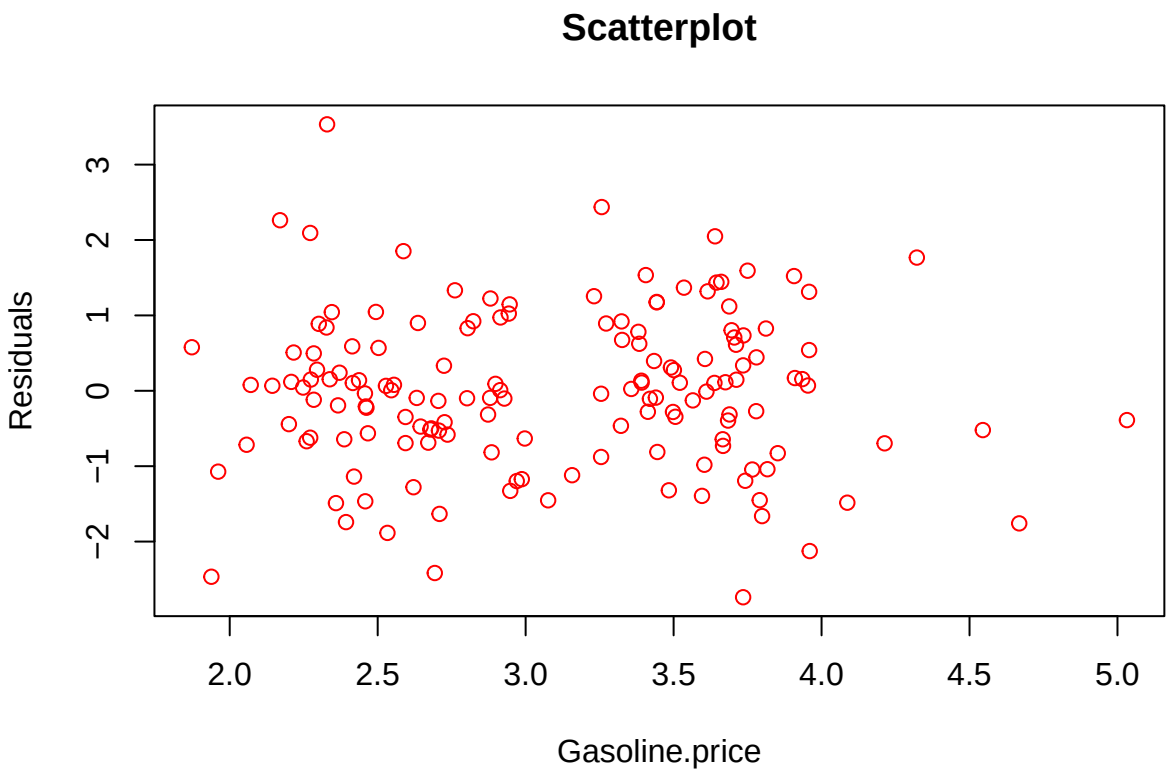

```
plot(data$Tesla.model.S.price[4:nrow(data)],resids,xlab="Tesla.model.S.price",ylab="Resi
```



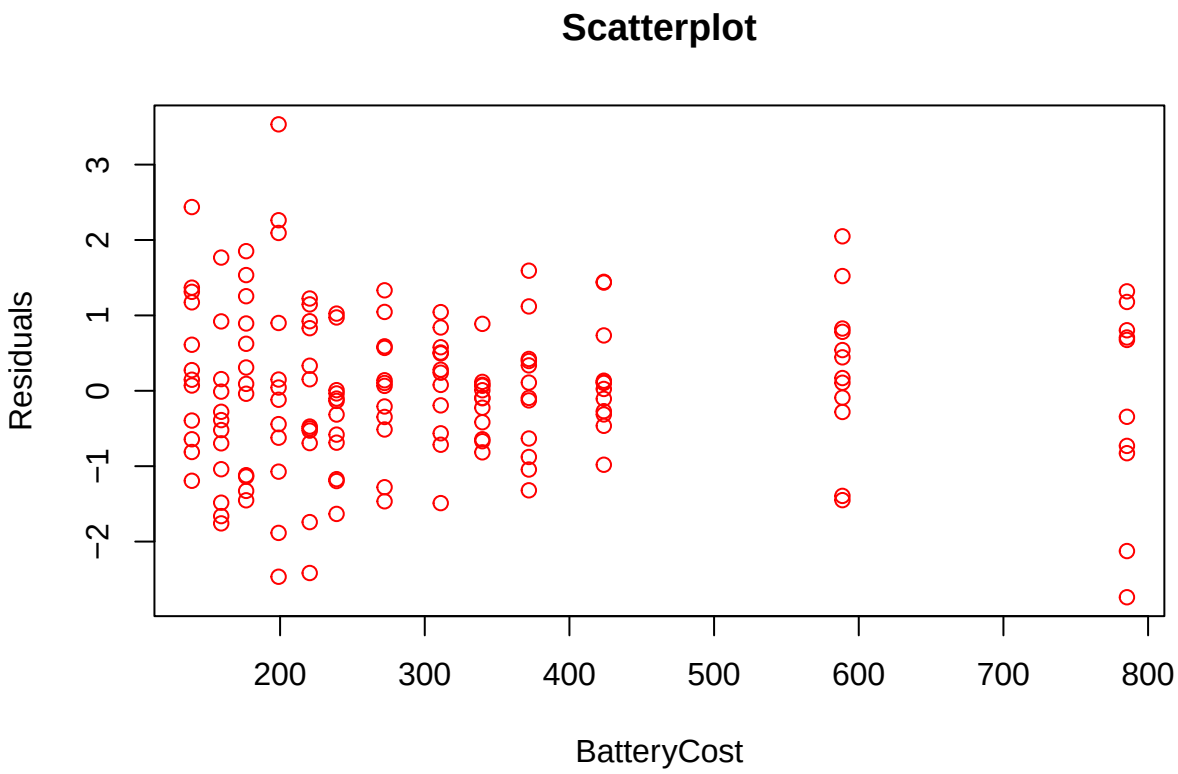

```
plot(data$Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education[4:nrow(da
```


```
plot(data$Percentage.of.employees.who.are.middle.aged..25.55.[4:nrow(data)],resids,xlab=
```

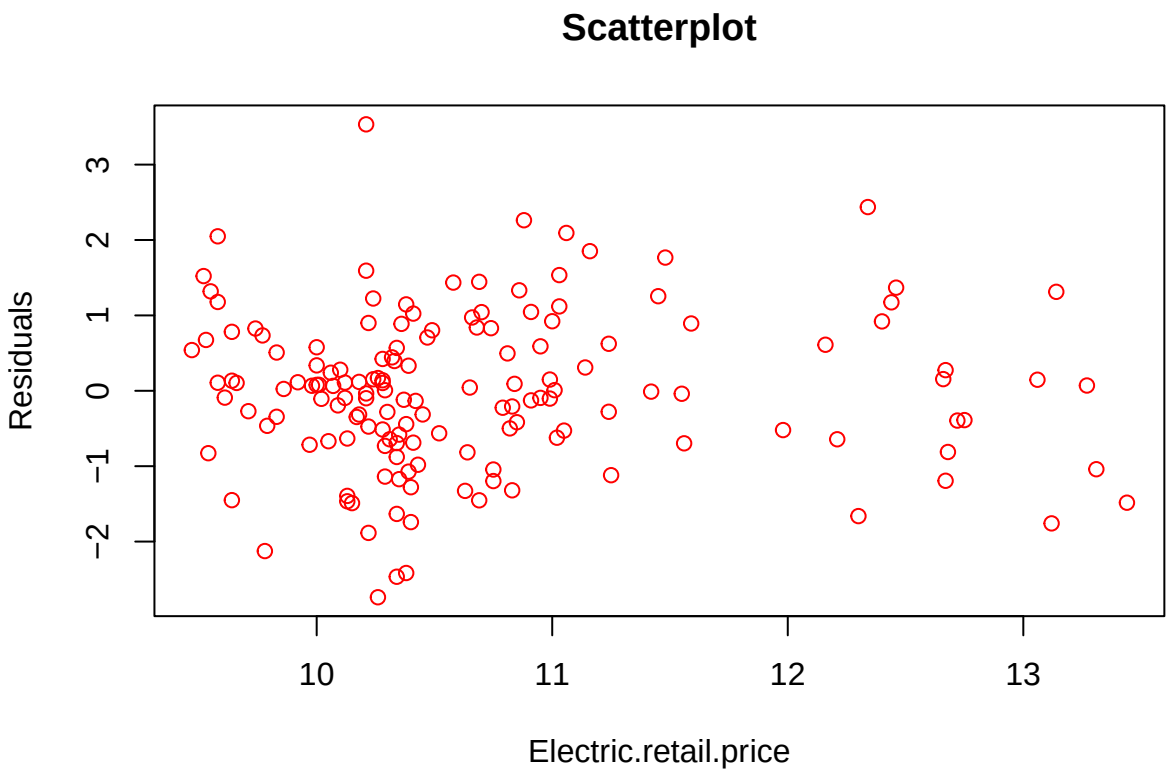

```
plot(data$Gasoline.price[4:nrow(data)],resids,xlab="Gasoline.price",ylab="Residuals",mai
```



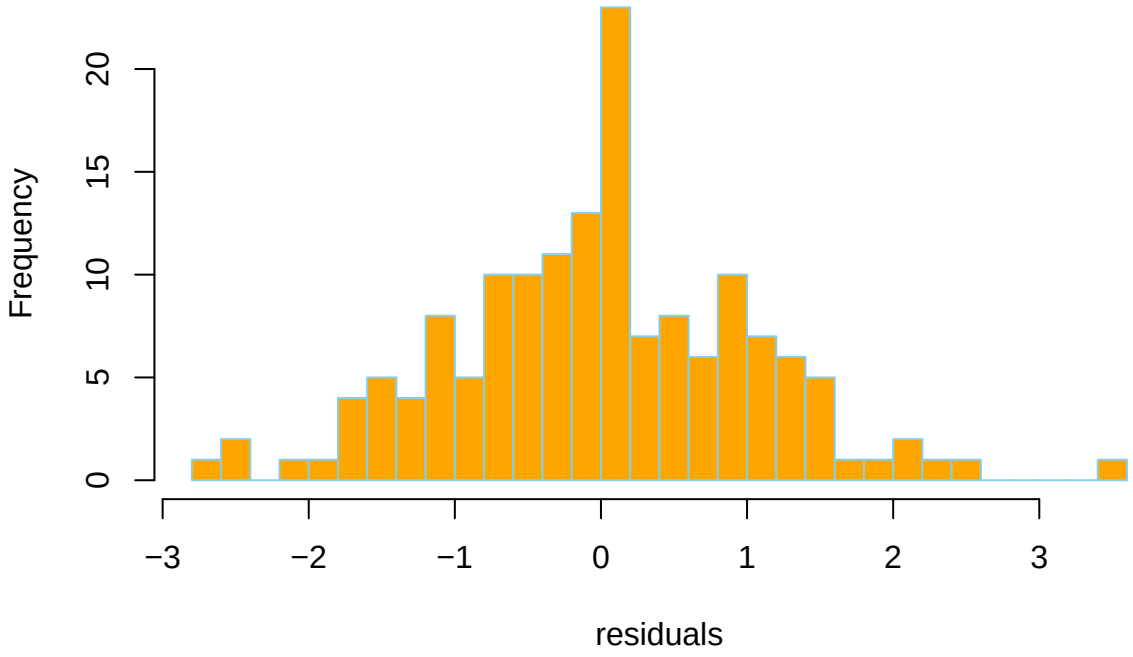

```
plot(data$BatteryCost[4:nrow(data)],resids,xlab="BatteryCost",ylab="Residuals",main="Sca
```



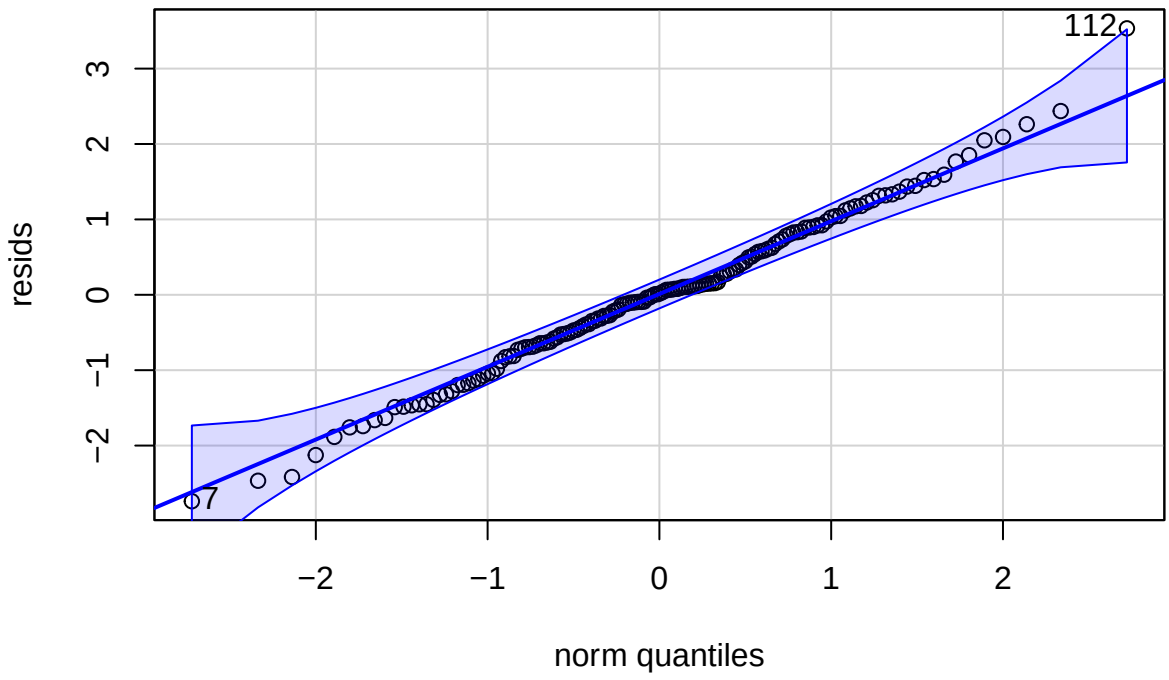

```
plot(data$Electric.retail.price[4:nrow(data)],resids,xlab="Electric.retail.price",ylab='
```




```
# Normality Assumption: hold
hist(resids,breaks = 30,main="",xlab="residuals",border = "skyblue",col="orange")
```




```
qqPlot(resids)
```




```
## 112   7
## 109   4
```

```
# Goodness of fit: Outliers
cook=cooks.distance(lagged_model2)
outlier_table <- data.frame(Standardized_Residuals=resids,Cooks_Distance = cook)
print(outlier_table)
```

```
##     Standardized_Residuals Cooks_Distance
## 4              1.318873255   1.280524e-02
## 5             -0.827890540   6.613645e-03
## 6             -2.125730720   3.656977e-02
## 7             -2.738400964   6.268763e-02
## 8              0.707047540   5.704628e-03
## 9              0.802825531   5.882041e-03
## 10            -0.730981282   4.797001e-03
## 11            -0.344917609   1.065168e-03
## 12             1.178076709   1.293981e-02
## 13             0.674567870   3.573641e-03
## 14            -0.090510637   5.464039e-05
## 15             2.049052137   2.564434e-02
## 16             1.521580522   1.935937e-02
```

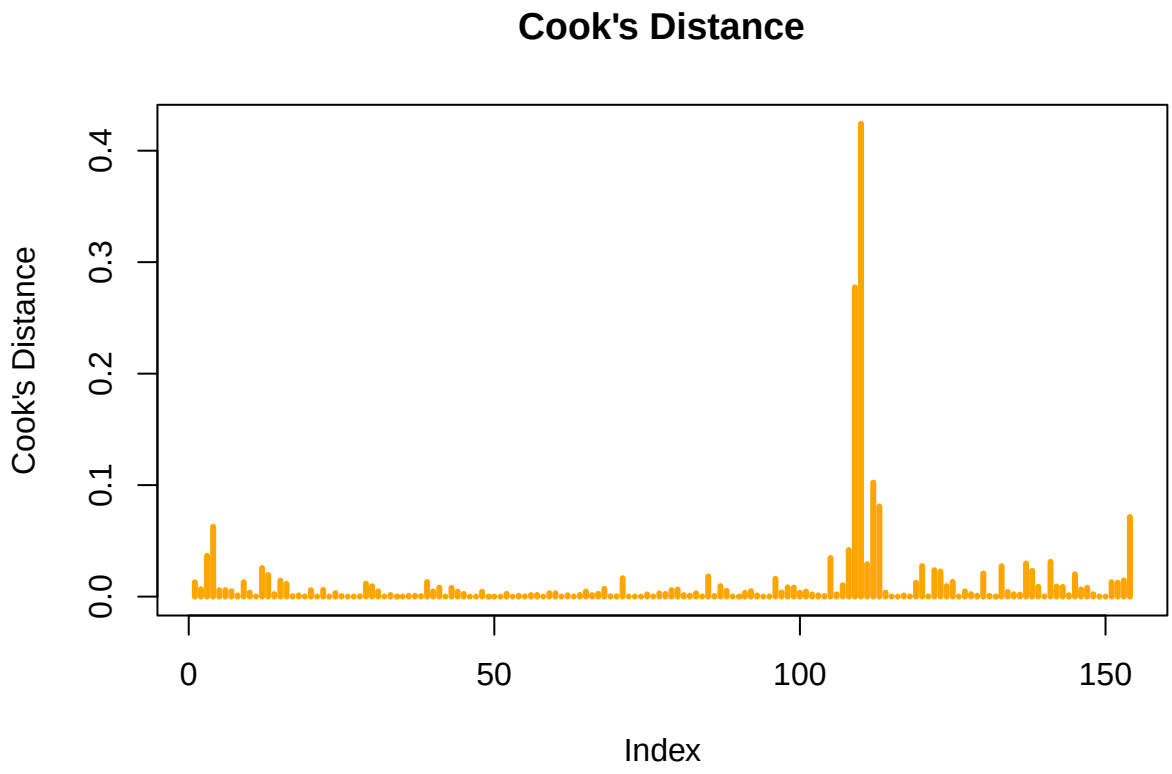
```
## 17            0.540073688   2.187315e-03
## 18           -1.450820421   1.429204e-02
## 19           -1.393934077   1.139757e-02
## 20           -0.280963009   4.198874e-04
## 21            0.443392547   1.049613e-03
## 22            0.168855118   1.273577e-04
## 23            0.825713704   5.733842e-03
## 24            0.106837854   1.011397e-04
## 25            0.781920741   6.157900e-03
## 26            0.133677082   9.643278e-05
## 27            0.734303521   3.161590e-03
## 28           -0.270697863   4.411891e-04
## 29            0.104255817   4.687857e-05
## 30            0.113039515   5.471776e-05
## 31           -0.313811568   3.862688e-04
## 32            1.445009979   1.161541e-02
## 33            1.435198526   9.352260e-03
## 34           -0.980953162   4.715081e-03
## 35           -0.106409150   8.566746e-05
## 36           -0.464613953   1.444501e-03
## 37            0.024257879   3.182997e-06
## 38            0.108693893   5.904110e-05
## 39            0.394953290   7.888445e-04
## 40            0.422175862   8.606240e-04
## 41            0.336846688   5.046400e-04
## 42            1.593103245   1.309515e-02
## 43           -1.044371661   4.556266e-03
## 44            1.118923448   8.067907e-03
## 45           -0.127320047   8.514419e-05
## 46           -1.320061095   7.884359e-03
## 47           -0.877874346   4.378189e-03
## 48           -0.632982709   2.393664e-03
## 49           -0.094537670   3.467461e-05
## 50            0.117959742   8.467226e-05
## 51            0.887700583   4.400974e-03
## 52            0.007259284   3.540186e-07
## 53            0.079296563   2.891347e-05
## 54           -0.098870942   5.821837e-05
## 55           -0.816349335   2.439691e-03
## 56           -0.094145997   4.654888e-05
## 57           -0.417304604   8.083407e-04
## 58           -0.224681766   1.892588e-04
## 59           -0.643140318   1.360337e-03
## 60           -0.668946816   1.497192e-03
## 61            0.066861228   2.215132e-05
```

```
## 62          -0.715762760   3.089234e-03
## 63           0.577381956   3.042694e-03
## 64           0.077977248   4.943646e-05
## 65           0.507401028   1.178790e-03
## 66           0.239793355   2.376969e-04
## 67          -0.563821882   1.490762e-03
## 68           1.043353158   4.612496e-03
## 69           0.495640722   1.355906e-03
## 70           0.839455327   2.536661e-03
## 71          -1.489953664   7.003614e-03
## 72           0.279262318   3.261494e-04
## 73          -0.193567000   2.851817e-04
## 74          -1.466031291   1.648252e-02
## 75           0.103424092   7.091027e-05
## 76           0.139492996   1.012007e-04
## 77           0.065122473   1.520491e-05
## 78           0.568709097   2.009168e-03
## 79          -0.208247959   1.273364e-04
## 80           0.589597765   2.832634e-03
## 81           1.045652740   2.410041e-03
## 82           1.332704211   5.767936e-03
## 83          -1.279169407   6.340623e-03
## 84          -0.513171955   1.355326e-03
## 85          -0.347622926   8.775396e-04
## 86          -0.688206296   2.927720e-03
## 87          -0.134304439   1.310334e-04
## 88          -1.633064883   1.807726e-02
## 89          -0.314950507   5.583732e-04
## 90          -1.172356942   9.408994e-03
## 91          -1.197387140   5.276728e-03
## 92          -0.103443147   7.536748e-05
## 93           0.008459442   3.054405e-07
## 94           0.972604497   3.329854e-03
## 95           1.024021900   4.869051e-03
## 96          -0.581413377   1.063181e-03
## 97          -0.035111532   6.327061e-06
## 98           0.152585582   1.533293e-04
## 99          -1.741830816   1.602726e-02
## 100         -0.693674035   3.746607e-03
## 101          1.224155839   8.280634e-03
## 102          1.145959792   8.043919e-03
## 103          0.830280726   3.478806e-03
## 104          0.922277586   4.557723e-03
## 105         -0.528282603   1.983495e-03
## 106         -0.498666306   1.172616e-03
```

```
## 107           0.332389481   5.442026e-04
## 108          -2.417044179   3.461226e-02
## 109          -0.474455499   1.989403e-03
## 110           0.899243006   1.013667e-02
## 111          -1.884981976   4.175172e-02
## 112           3.534029462   2.773797e-01
## 113          -2.466681075   4.241906e-01
## 114          -1.072506740   2.907432e-02
## 115           2.262323241   1.021906e-01
## 116           2.093488178   8.085797e-02
## 117          -0.622331239   3.629181e-03
## 118           0.149146592   1.718984e-04
## 119           0.043102895   9.674828e-06
## 120          -0.442247952   1.052066e-03
## 121          -0.119126285   1.197681e-04
## 122          -1.138027651   1.245745e-02
## 123           1.851962915   2.752749e-02
## 124           0.091684245   1.377569e-04
## 125          -1.327453922   2.372098e-02
## 126          -1.452742409   2.253422e-02
## 127          -1.119491159   9.407800e-03
## 128           1.254370091   1.323698e-02
## 129          -0.039463528   8.531127e-06
## 130           0.892548002   4.737300e-03
## 131           0.623503563   2.203396e-03
## 132           0.308947978   7.819442e-04
## 133           1.534667418   2.073486e-02
## 134          -0.278618061   6.947758e-04
## 135          -0.011167843   9.712139e-07
## 136           1.767010137   2.731544e-02
## 137          -0.696797969   4.159599e-03
## 138          -0.521068291   2.043547e-03
## 139          -0.388976383   1.741098e-03
## 140          -1.758862555   2.969062e-02
## 141          -1.484434329   2.304015e-02
## 142          -1.040456164   8.758900e-03
## 143           0.154739010   2.446211e-04
## 144          -1.662147844   3.119934e-02
## 145           0.920414392   9.020116e-03
## 146          -0.811901783   8.656039e-03
## 147           0.273928903   1.230224e-03
## 148           1.367997370   2.004612e-02
## 149           0.610676998   6.414160e-03
## 150          -0.643076364   8.013115e-03
## 151          -0.394821078   2.171496e-03
```

```
## 152            0.146795830   1.953388e-04
## 153            0.069766670   4.050768e-05
## 154            1.312530849   1.287868e-02
## 155           -1.193691728   1.246608e-02
## 156            1.173036367   1.451519e-02
## 157            2.436967434   7.143228e-02
```

```
plot(cook,type="h",lwd=3,col="orange",ylab="Cook's Distance",main="Cook's Distance")
```




```
condition <- (abs(resids) > 2) & (cook > 4 / 157)
count_rows <- sum(condition)
count_rows/157
```

```
## [1] 0.05732484
```

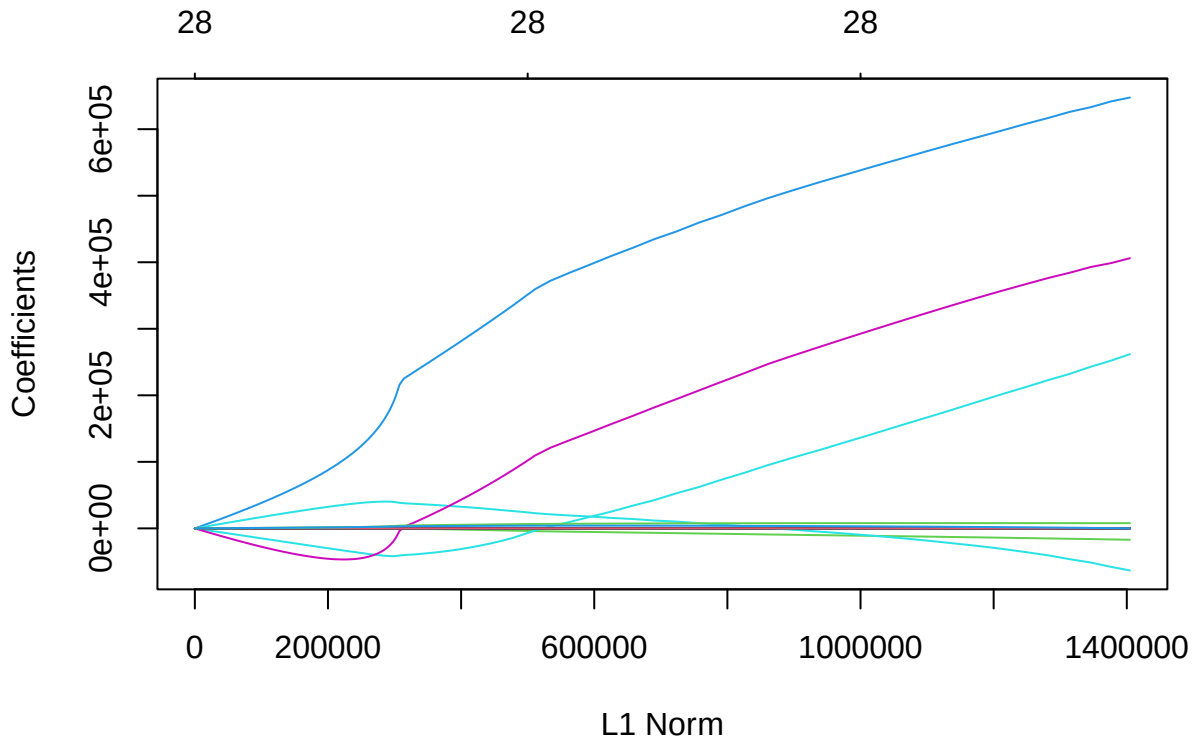
```
# Conclusion: no need to transform the predicting variables
```

Ridge MLR (considering the dataset has high multicoliearity)

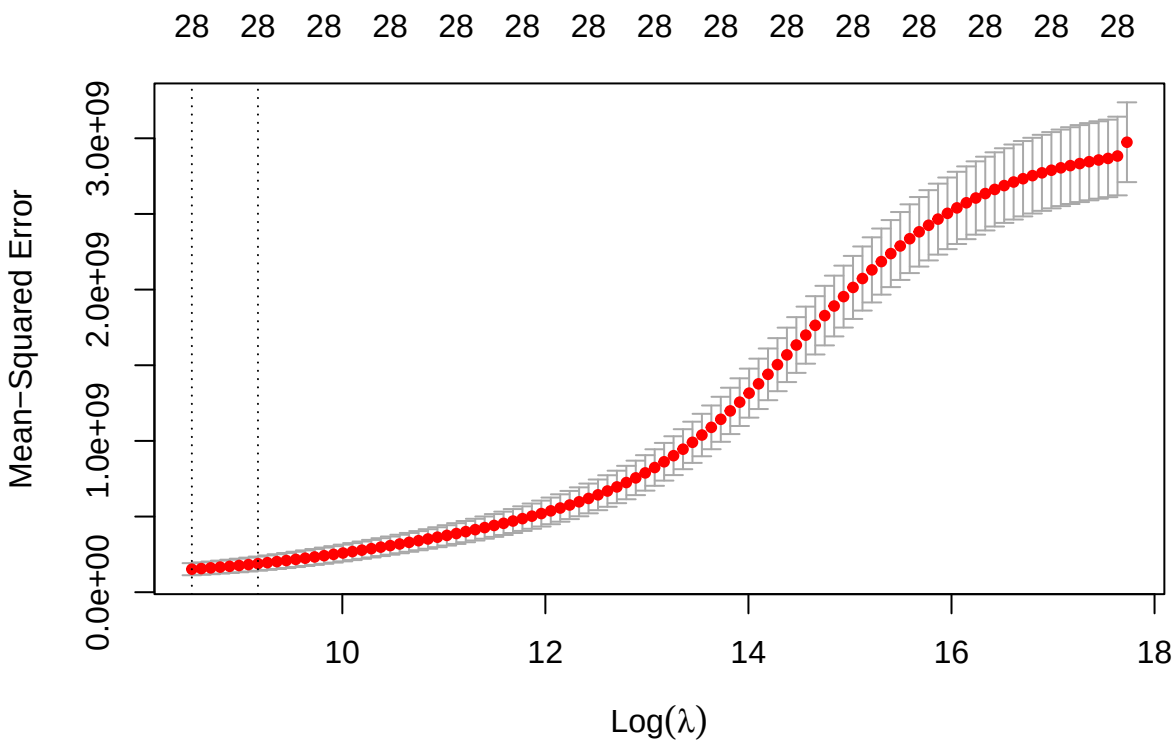
```
X <- as.matrix(data[, -which(names(data) == "Total.NEV.Sales")])
y <- data$Total.NEV.Sales
library(glmnet)
```

```
##      Matrix
##
##     'Matrix'
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
## Loaded glmnet 4.1-8
```

```r
ridge_model <- glmnet(X, y, alpha = 0)
plot(ridge_model )
```




```r
cv_ridge_model  <- cv.glmnet(X, y, alpha = 0)
plot(cv_ridge_model )
```




```r
best_lambda2 <- cv_ridge_model$lambda.min
print(best_lambda2)
```

```
## [1] 5003.271
```

```r
final_ridge_model <- glmnet(X, y, alpha = 0, lambda = best_lambda2)
print(final_ridge_model$beta)
```

```
## 28 x 1 sparse Matrix of class "dgCMatrix"
##                                                s0
## Month                                5.622378e+02
## Date                                -8.760754e-02
## Covid.19                            -1.707478e+04
## Number.of.new.energy.policies       -1.088210e+02
## Production                           4.564174e-02
## Inventory                            4.621963e-03
## Gasoline.LDV.sales                   2.751609e-02
## M1SL                                 1.263975e+00
## M2SL                                 5.417634e-01
## Federal.fund.effective.rate          6.405528e+05
## Umemployment.rate                    2.701036e+05
## PPIBC                               -1.103099e+02
```

```
## PPIBM                                                                  2.166538e+02
## PCE                                                                    2.478844e+01
## GDP                                                                    1.960865e+00
## Durable.goods.consumption                                              2.536275e+01
## Nondurable.goods.consumption                                           1.631283e+01
## Nonresidential.fixed.Investment                                        1.062938e+01
## Residential.fixed.Investment                                          -3.438602e+00
## Tesla.model.S.price                                                    2.411760e+00
## ElectricPorts                                                          2.122293e-01
## BatteryCost                                                           -2.308608e+00
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education -5.744401e+04
## Percentage.of.employees.who.are.middle.aged..25.55.                    4.043770e+05
## Population                                                            -6.745629e-01
## Per.capita.disposable.income                                           9.417733e-01
## Gasoline.price                                                         7.844055e+03
## Electric.retail.price                                                 -1.098814e+01
```

Poisson: not suitable (fail the goodness-of-fit test)

```
# Poisson regression is not suitable for this data set
poisson_model<- glm(Total.NEV.Sales~., family=poisson, data=data)
summary(poisson_model)
```

```
##
## Call:
## glm(formula = Total.NEV.Sales ~ ., family = poisson, data = data)
##
## Coefficients:
##                                                                        Estimate
## (Intercept)                                                           1.826e+01
## Month                                                                 1.461e-02
## Date                                                                  2.785e-03
## Covid.19                                                             -1.597e-02
## Number.of.new.energy.policies                                        -9.604e-05
## Production                                                           -9.801e-08
## Inventory                                                             4.324e-07
## Gasoline.LDV.sales                                                    5.841e-07
## M1SL                                                                 -3.417e-05
## M2SL                                                                  2.034e-05
## Federal.fund.effective.rate                                           8.011e+00
## Umemployment.rate                                                    -3.328e+00
## PPIBC                                                                 7.212e-03
## PPIBM                                                                -7.119e-04
## PCE                                                                   4.180e-04
```

```
## GDP                                                                    -1.826e-04
## Durable.goods.consumption                                               1.610e-03
## Nondurable.goods.consumption                                           -7.930e-04
## Nonresidential.fixed.Investment                                        -6.890e-04
## Residential.fixed.Investment                                           -1.671e-03
## Tesla.model.S.price                                                     3.786e-05
## ElectricPorts                                                           8.340e-06
## BatteryCost                                                            -1.642e-03
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education -4.855e+00
## Percentage.of.employees.who.are.middle.aged..25.55.                    -4.713e+00
## Population                                                             -3.633e-04
## Per.capita.disposable.income                                           -2.683e-06
## Gasoline.price                                                          1.920e-01
## Electric.retail.price                                                  -2.330e-02
##                                                                        Std. Error
## (Intercept)                                                             4.784e-01
## Month                                                                   1.210e-04
## Date                                                                    3.674e-05
## Covid.19                                                                2.282e-03
## Number.of.new.energy.policies                                           1.145e-05
## Production                                                              1.109e-08
## Inventory                                                               7.678e-09
## Gasoline.LDV.sales                                                      3.133e-09
## M1SL                                                                    8.053e-07
## M2SL                                                                    3.996e-06
## Federal.fund.effective.rate                                             1.381e-01
## Umemployment.rate                                                       6.389e-02
## PPIBC                                                                   3.134e-04
## PPIBM                                                                   2.873e-04
## PCE                                                                     1.053e-05
## GDP                                                                     5.858e-06
## Durable.goods.consumption                                               2.199e-05
## Nondurable.goods.consumption                                            3.298e-05
## Nonresidential.fixed.Investment                                         1.925e-05
## Residential.fixed.Investment                                            2.722e-05
## Tesla.model.S.price                                                     7.865e-07
## ElectricPorts                                                           2.163e-07
## BatteryCost                                                             1.460e-05
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education    9.943e-02
## Percentage.of.employees.who.are.middle.aged..25.55.                     1.291e-01
## Population                                                              3.918e-06
## Per.capita.disposable.income                                            6.154e-07
## Gasoline.price                                                          1.955e-03
## Electric.retail.price                                                   1.030e-03
##                                                                         z value
```

```
## (Intercept)                                                          38.160
## Month                                                               120.767
## Date                                                                 75.795
## Covid.19                                                             -6.997
## Number.of.new.energy.policies                                        -8.391
## Production                                                           -8.835
## Inventory                                                            56.313
## Gasoline.LDV.sales                                                  186.473
## M1SL                                                                -42.430
## M2SL                                                                  5.091
## Federal.fund.effective.rate                                          57.993
## Umemployment.rate                                                   -52.087
## PPIBC                                                                23.015
## PPIBM                                                                -2.478
## PCE                                                                  39.698
## GDP                                                                 -31.173
## Durable.goods.consumption                                            73.204
## Nondurable.goods.consumption                                        -24.047
## Nonresidential.fixed.Investment                                     -35.801
## Residential.fixed.Investment                                        -61.394
## Tesla.model.S.price                                                  48.144
## ElectricPorts                                                        38.562
## BatteryCost                                                        -112.495
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education -48.827
## Percentage.of.employees.who.are.middle.aged..25.55.                  -36.509
## Population                                                          -92.727
## Per.capita.disposable.income                                         -4.359
## Gasoline.price                                                       98.197
## Electric.retail.price                                               -22.625
##                                                                    Pr(>|z|)
## (Intercept)                                                         < 2e-16
## Month                                                               < 2e-16
## Date                                                                < 2e-16
## Covid.19                                                           2.61e-12
## Number.of.new.energy.policies                                       < 2e-16
## Production                                                          < 2e-16
## Inventory                                                           < 2e-16
## Gasoline.LDV.sales                                                  < 2e-16
## M1SL                                                                < 2e-16
## M2SL                                                               3.56e-07
## Federal.fund.effective.rate                                         < 2e-16
## Umemployment.rate                                                   < 2e-16
## PPIBC                                                               < 2e-16
## PPIBM                                                                0.0132
## PCE                                                                 < 2e-16
```

```
## GDP                                                                  < 2e-16
## Durable.goods.consumption                                            < 2e-16
## Nondurable.goods.consumption                                         < 2e-16
## Nonresidential.fixed.Investment                                      < 2e-16
## Residential.fixed.Investment                                         < 2e-16
## Tesla.model.S.price                                                  < 2e-16
## ElectricPorts                                                        < 2e-16
## BatteryCost                                                          < 2e-16
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education  < 2e-16
## Percentage.of.employees.who.are.middle.aged..25.55.                  < 2e-16
## Population                                                           < 2e-16
## Per.capita.disposable.income                                        1.31e-05
## Gasoline.price                                                       < 2e-16
## Electric.retail.price                                                < 2e-16
##
## (Intercept)                                                         ***
## Month                                                               ***
## Date                                                                ***
## Covid.19                                                            ***
## Number.of.new.energy.policies                                       ***
## Production                                                          ***
## Inventory                                                           ***
## Gasoline.LDV.sales                                                  ***
## M1SL                                                                ***
## M2SL                                                                ***
## Federal.fund.effective.rate                                         ***
## Umemployment.rate                                                   ***
## PPIBC                                                               ***
## PPIBM                                                               *
## PCE                                                                 ***
## GDP                                                                 ***
## Durable.goods.consumption                                           ***
## Nondurable.goods.consumption                                        ***
## Nonresidential.fixed.Investment                                     ***
## Residential.fixed.Investment                                        ***
## Tesla.model.S.price                                                 ***
## ElectricPorts                                                       ***
## BatteryCost                                                         ***
## Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education ***
## Percentage.of.employees.who.are.middle.aged..25.55.                 ***
## Population                                                          ***
## Per.capita.disposable.income                                        ***
## Gasoline.price                                                      ***
## Electric.retail.price                                               ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 5123903  on 156  degrees of freedom
## Residual deviance:   72056  on 128  degrees of freedom
## AIC: 74128
##
## Number of Fisher Scoring iterations: 4
```

Overall significance

```
1-pchisq(poisson_model$null.deviance-poisson_model$deviance,38)
```

```
## [1] 0
```

Goodness-of-fit

```
c(deviance(poisson_model),1-pchisq(deviance(poisson_model),118))
```

```
## [1] 72056.26     0.00
```

```
pearres=residuals(poisson_model,type="pearson")
pearson.tvalue=sum(pearres^2)
c(pearson.tvalue,1-pchisq(pearson.tvalue,118))
```

```
## [1] 71256.07     0.00
```

```
# conlusion: not a good fit
```

Model Comparision - Prediction accuracy

```
install.packages("Metrics")
```

```
## Warning in download.file(url, destfile, method, mode = "wb", ...): URL
## 'https://cran.rstudio.com/bin/windows/contrib/4.4/Metrics_0.1.4.zip': status
## was 'SSL connect error'
```

```
## Error in download.file(url, destfile, method, mode = "wb", ...) :
##     URL'https://cran.rstudio.com/bin/windows/contrib/4.4/Metrics_0.1.4.zip'
```

```
## Warning in download.packages(pkgs, destdir = tmpd, available = available, :
##    'Metrics'
```

```
library(Metrics)  # For MAE, MAPE, RMSE
```

```
## Warning:   'Metrics' R 4.4.2
```

```r
library(glmnet)   # For ridge regression
library(MASS)     # For stepwise regression
library(caret)    # For cross-validation
```

```
##      lattice
```

```
##
##    'caret'
```

```
## The following objects are masked from 'package:Metrics':
##
##     precision, recall
```

```r
set.seed(123)

calc_metrics <- function(predictions, actual) {
  # Mean Squared Prediction Error (MSPE)
  MSPE <- mean((predictions - actual)^2)

  # Mean Absolute Error (MAE)
  MAE <- mean(abs(predictions - actual))

  # Mean Absolute Percentage Error (MAPE)
  MAPE <- mean(abs((predictions - actual) / actual)) * 100

  # Precision Measure (PM)
  PM <- sum((predictions - actual)^2) / sum((actual - mean(actual))^2)

  # R-squared
  SS_res <- sum((predictions - actual)^2)  # Residual sum of squares
  SS_tot <- sum((actual - mean(actual))^2)  # Total sum of squares
  R_squared <- 1 - (SS_res / SS_tot)

  # Residuals
  residuals <- actual - predictions
  mean_residual <- mean(residuals)  # Mean of residuals
  residual_sum_of_squares <- sum(residuals^2)  # Sum of squared residuals

  # Return all metrics as a named vector
  return(c(MSPE = MSPE, MAE = MAE, MAPE = MAPE, PM = PM,
           R_squared = R_squared, Mean_Residual = mean_residual, Residual_Sum_of_Squares
}

# Cross-validation function
cv_model <- function(model, X, Y, k = 10) {
  folds <- createFolds(Y, k = k, list = TRUE, returnTrain = FALSE)
```

```
  metrics <- matrix(NA, nrow = k, ncol = 7)
  colnames(metrics) <- c("MSPE", "MAE", "MAPE", "PM", "R-squared", "Mean Residual", "Res


  for (i in 1:k) {
    test_index <- folds[[i]]
    train_index <- setdiff(1:nrow(X), test_index)

    X_train <- X[train_index, , drop = FALSE]
    Y_train <- Y[train_index]
    X_test <- X[test_index, , drop = FALSE]
    Y_test <- Y[test_index]

    # Train model on training data
    if (inherits(model, "glmnet")) {
      # Ridge regression
      pred <- predict(model, newx = X_test)
      pred <- pred[, 1]
    } else if (inherits(model, "lm")) {
      # For linear models like stepwise and boxcox
      pred <- predict(model, newdata = data.frame(X_test))
    }

    # Calculate performance metrics
    metrics[i, ] <- calc_metrics(pred, Y_test)
  }

  # Return the mean of each metric over all folds
  colMeans(metrics)
}

stepwise_metrics_multicolinearity_corrected <- cv_model(stepwise_model2,data_stepwise0[,

data_cv_step <-data_stepwise[4:nrow(data_stepwise),]
stepwise_metrics_multicolinearity_autocorrelation_corrected <- cv_model(lagged_model,dat

boxcox_metrics_multicolinearity_corrected<-cv_model(model_boxcox2,data_boxcox0[, -1], (c

data_cv_boxcox <- data_boxcox[4:nrow(data_boxcox),]
boxcox_metrics_multicolinearity_autocorrelation_corrected <- cv_model(lagged_model2,data


ridge_metrics <- cv_model(final_ridge_model,as.matrix(data[, -1]),as.matrix(data$Total.
```

```r
cat("Stepwise multicolinearity corrected Model Metrics:", "\n")
```

```
## Stepwise multicolinearity corrected Model Metrics:
```

```r
print(stepwise_metrics_multicolinearity_corrected )
```

```
##                     MSPE                      MAE                     MAPE
##            1.529660e+08            8.586711e+03            1.491908e+01
##                       PM               R-squared           Mean Residual
##            7.987972e-02            9.201203e-01           -1.477414e+01
## Residual Sum of Squares
##            2.399444e+09
```

```r
cat("Stepwise multicolinearity & autocorrelation corrected Model Metrics:", "\n")
```

```
## Stepwise multicolinearity & autocorrelation corrected Model Metrics:
```

```r
print(stepwise_metrics_multicolinearity_autocorrelation_corrected)
```

```
##                     MSPE                      MAE                     MAPE
##            5.281100e+07            5.343426e+03            8.998258e+00
##                       PM               R-squared           Mean Residual
##            2.339006e-02            9.766099e-01           -1.935032e+01
## Residual Sum of Squares
##            8.143157e+08
```

```r
cat("Box-Cox multicolinearity corrected Model Metrics:", "\n")
```

```
## Box-Cox multicolinearity corrected Model Metrics:
```

```r
print(boxcox_metrics_multicolinearity_corrected)
```

```
##                     MSPE                      MAE                     MAPE
##            2.218659e+02            1.091726e+01            4.475114e+00
##                       PM               R-squared           Mean Residual
##            4.908420e-02            9.509158e-01            5.622677e-02
## Residual Sum of Squares
##            3.484675e+03
```

```r
cat("Box-Cox multicolinearity & autocorrelation corrected Model Metrics:", "\n")
```

```
## Box-Cox multicolinearity & autocorrelation corrected Model Metrics:
```

```r
print(boxcox_metrics_multicolinearity_autocorrelation_corrected)
```

```
##                     MSPE                      MAE                     MAPE
##            1.082560e+02            8.089410e+00            3.355103e+00
##                       PM               R-squared           Mean Residual
##            2.205188e-02            9.779481e-01            4.811473e-02
```

```
## Residual Sum of Squares
##            1.678848e+03
```

```
cat("Ridge Regression Model Metrics:", "\n")
```

```
## Ridge Regression Model Metrics:
```

```
print(ridge_metrics)
```

```
##                     MSPE                      MAE                     MAPE
##            1.281676e+08             7.844014e+03             1.313336e+01
##                       PM                R-squared           Mean Residual
##            4.450664e-02             9.554934e-01            -1.444706e+01
## Residual Sum of Squares
##            2.021927e+09
```