week.5..p

sahrash fatima Lab

2024-10-02

```
# Load the dataset from the specified path
dataset <- read.csv("D:/iupui/2rd sem/R- stats/parkinsons_disease_data.csv")

# View the first few rows of the dataset to verify it loaded correctly
head(dataset)
```

```
##   PatientID Age Gender Ethnicity EducationLevel      BMI Smoking
## 1      3058  85      0         3              1 19.61988       0
## 2      3059  75      0         0              2 16.24734       1
## 3      3060  70      1         0              0 15.36824       0
## 4      3061  52      0         0              0 15.45456       0
## 5      3062  87      0         0              1 18.61604       0
## 6      3063  68      1         2              1 39.42331       1
##   AlcoholConsumption PhysicalActivity DietQuality SleepQuality
## 1           5.108241        1.3806599    3.893969     9.283194
## 2           6.027648        8.4098041    8.513428     5.602470
## 3           2.242135        0.2132746    6.498805     9.929824
## 4           5.997788        1.3750452    6.715033     4.196189
## 5           9.775243        1.1886071    4.657572     9.363925
## 6          13.596889        7.7967040    7.070239     7.737549
##   FamilyHistoryParkinsons TraumaticBrainInjury Hypertension Diabetes
Depression
## 1                       0                    0            0        0
0
## 2                       0                    0            0        0
0
## 3                       0                    0            0        1
0
## 4                       0                    0            0        0
0
## 5                       0                    0            0        0
0
## 6                       0                    0            0        0
0
##   Stroke SystolicBP DiastolicBP CholesterolTotal CholesterolLDL
CholesterolHDL
## 1      0        129          60         222.8423      148.12562
37.86778
## 2      0        163          76         210.5011      153.75646
77.22812
## 3      0        113          93         287.3880      118.70260
85.58830
## 4      0        146          78         280.3395      136.29919
```

```
51.86963
## 5      0        115          94          284.0142      108.44945
25.06942
## 6      0        151          90          290.1331       91.75022
54.48892
##   CholesterolTriglycerides      UPDRS      MoCA FunctionalAssessment
Tremor
## 1                 337.3071   6.458713 29.181289             1.572427
1
## 2                 264.6355  37.306703 12.332639             4.787551
0
## 3                 395.6626  67.838170 29.927783             2.130686
1
## 4                 362.1897  52.964696 21.304268             3.391288
1
## 5                 149.9566  21.804880  8.336364             3.200969
0
## 6                 253.7973 101.912536 27.370580             6.824779
0
##   Rigidity Bradykinesia PosturalInstability SpeechProblems SleepDisorders
## 1        0            0                   0              0              0
## 2        1            0                   1              0              1
## 3        0            0                   0              1              0
## 4        1            1                   0              0              0
## 5        0            0                   1              0              1
## 6        0            0                   0              0              0
##   Constipation Diagnosis DoctorInCharge
## 1            0         0    DrXXXConfid
## 2            0         1    DrXXXConfid
## 3            1         1    DrXXXConfid
## 4            1         1    DrXXXConfid
## 5            0         0    DrXXXConfid
## 6            0         0    DrXXXConfid
```

```
# Check the structure of the dataset to ensure variables are correctly loaded
str(dataset)
```

```
## 'data.frame':    2105 obs. of  35 variables:
##  $ PatientID                : int  3058 3059 3060 3061 3062 3063 3064 3065
3066 3067 ...
##  $ Age                      : int  85 75 70 52 87 68 78 70 80 71 ...
##  $ Gender                   : int  0 0 1 0 0 1 1 1 0 0 ...
##  $ Ethnicity                : int  3 0 0 0 0 2 0 0 2 3 ...
##  $ EducationLevel           : int  1 2 0 0 1 1 0 0 1 2 ...
##  $ BMI                      : num  19.6 16.2 15.4 15.5 18.6 ...
##  $ Smoking                  : int  0 1 0 0 0 1 1 1 1 1 ...
##  $ AlcoholConsumption       : num  5.11 6.03 2.24 6 9.78 ...
##  $ PhysicalActivity         : num  1.381 8.41 0.213 1.375 1.189 ...
##  $ DietQuality              : num  3.89 8.51 6.5 6.72 4.66 ...
##  $ SleepQuality             : num  9.28 5.6 9.93 4.2 9.36 ...
##  $ FamilyHistoryParkinsons : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ TraumaticBrainInjury    : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ Hypertension            : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ Diabetes                : int  0 0 1 0 0 0 0 1 1 0 ...
##  $ Depression              : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ Stroke                  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ SystolicBP              : int  129 163 113 146 115 151 122 129 133 169
...
##  $ DiastolicBP             : int  60 76 93 78 94 90 60 99 113 105 ...
##  $ CholesterolTotal        : num  223 211 287 280 284 ...
##  $ CholesterolLDL          : num  148 154 119 136 108 ...
##  $ CholesterolHDL          : num  37.9 77.2 85.6 51.9 25.1 ...
##  $ CholesterolTriglycerides: num  337 265 396 362 150 ...
##  $ UPDRS                   : num  6.46 37.31 67.84 52.96 21.8 ...
##  $ MoCA                    : num  29.18 12.33 29.93 21.3 8.34 ...
##  $ FunctionalAssessment    : num  1.57 4.79 2.13 3.39 3.2 ...
##  $ Tremor                  : int  1 0 1 1 0 0 1 1 0 0 ...
##  $ Rigidity                : int  0 1 0 1 0 0 0 0 0 0 ...
##  $ Bradykinesia            : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ PosturalInstability     : int  0 1 0 0 1 0 0 1 0 0 ...
##  $ SpeechProblems          : int  0 0 1 0 0 0 1 0 0 0 ...
##  $ SleepDisorders          : int  0 1 0 0 1 0 0 0 0 1 ...
##  $ Constipation            : int  0 0 1 1 0 0 0 1 0 0 ...
##  $ Diagnosis               : int  0 1 1 1 0 0 0 1 1 0 ...
##  $ DoctorInCharge          : chr  "DrXXXConfid" "DrXXXConfid"
"DrXXXConfid" "DrXXXConfid" ...
```

```
# Install and load required packages
if (!require(ggplot2)) install.packages("ggplot2", dependencies=TRUE)
```

```
## Loading required package: ggplot2
```

```
if (!require(psych)) install.packages("psych", dependencies=TRUE)
```

```
## Loading required package: psych
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
if (!require(corrplot)) install.packages("corrplot", dependencies=TRUE)
```

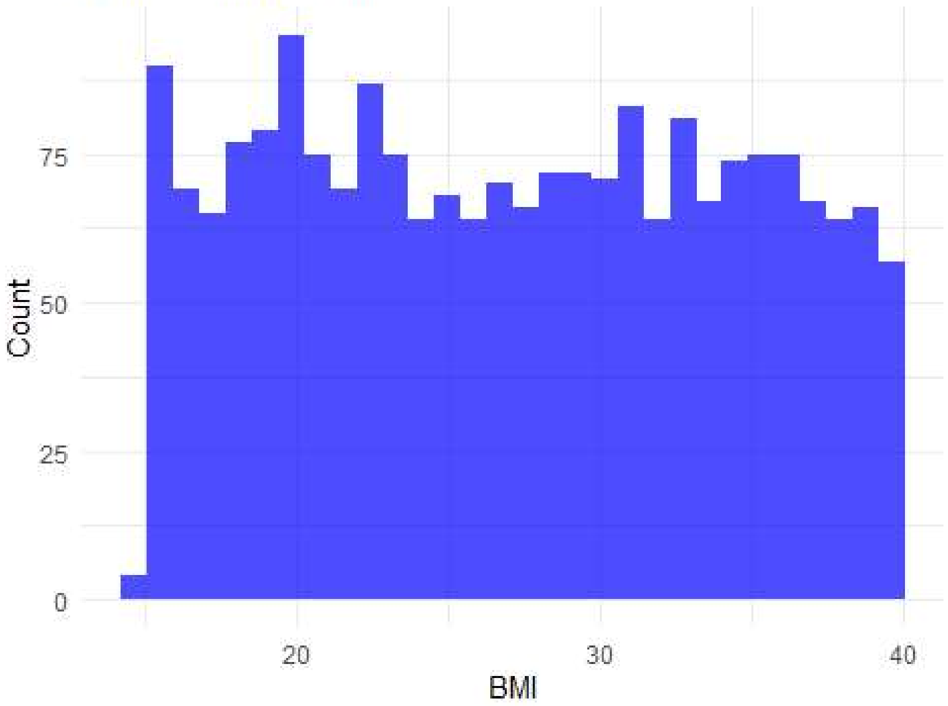
```
## Loading required package: corrplot
```

```
## corrplot 0.94 loaded
```

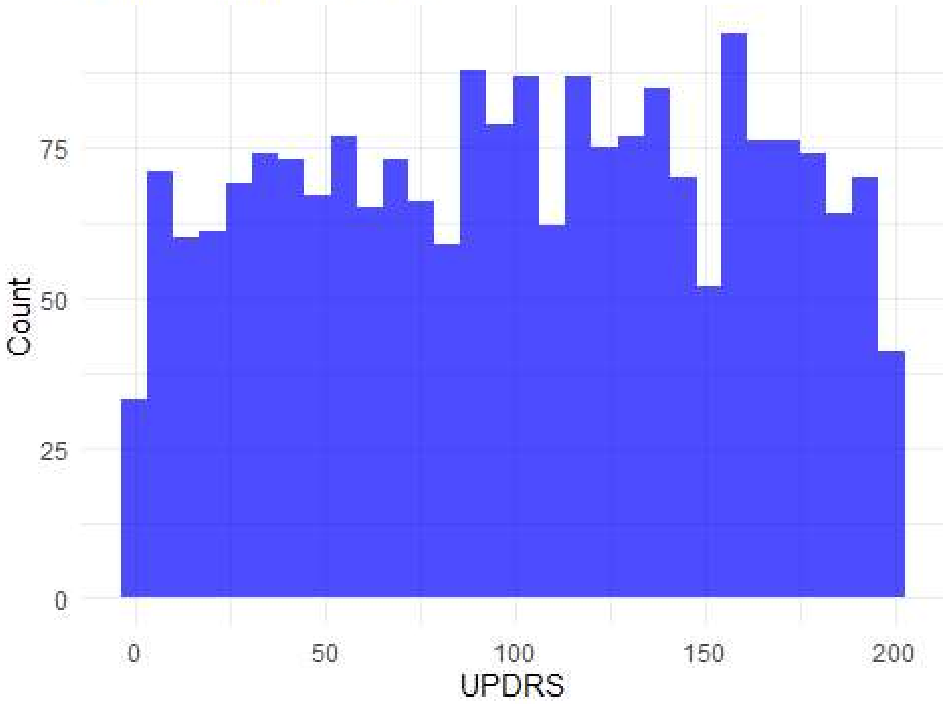
```
library(ggplot2)
library(psych)
library(corrplot)
```

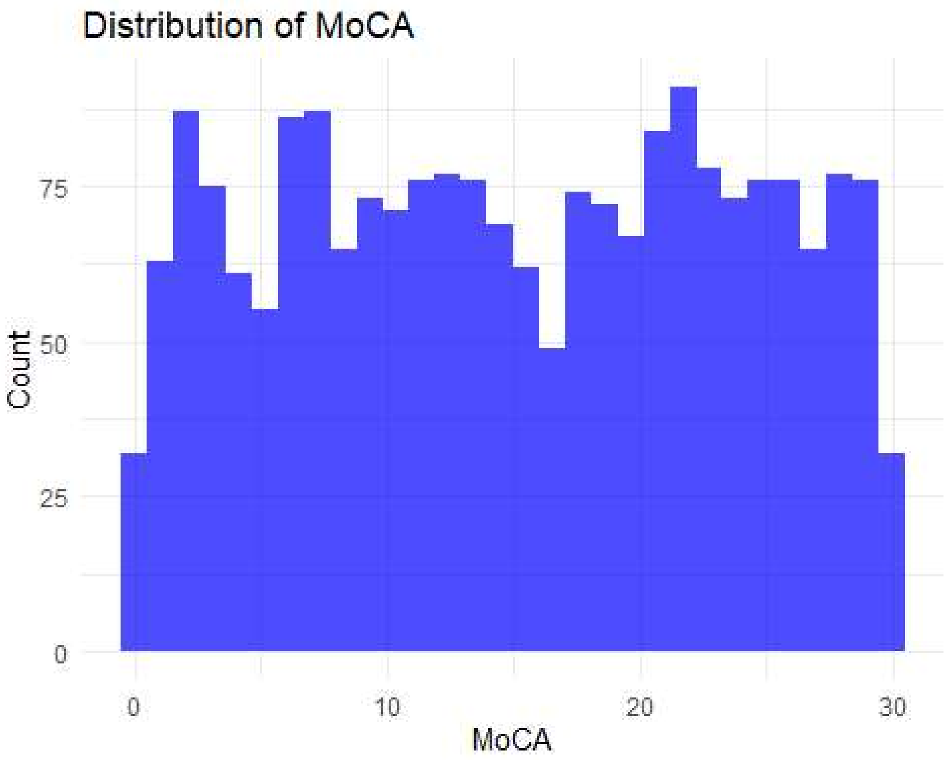
```r
# Plot distribution of variables and check for skewness
plot_distribution <- function(variable_name, dataset) {
  tryCatch({
    ggplot(dataset, aes(x = !!sym(variable_name))) +
      geom_histogram(bins = 30, fill = "blue", alpha = 0.7) +
      labs(title = paste("Distribution of", variable_name), x =
variable_name, y = "Count") +
      theme_minimal()
  }, error = function(e) {
    return(paste("Error in plotting", variable_name, ":", e$message))
  })
}

# Variables to plot (based on dataset)
variables <- c("BMI", "UPDRS", "MoCA", "PhysicalActivity", "SystolicBP")

# Loop through and plot distributions
for (var in variables) {
  print(plot_distribution(var, dataset))
}
```

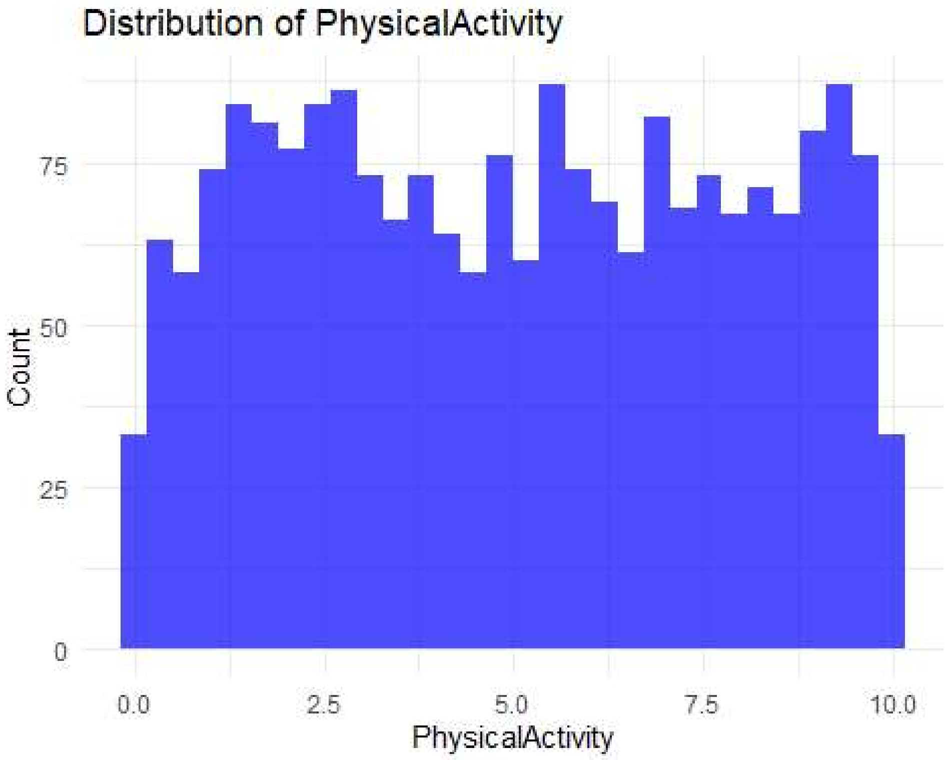
Distribution of BMI

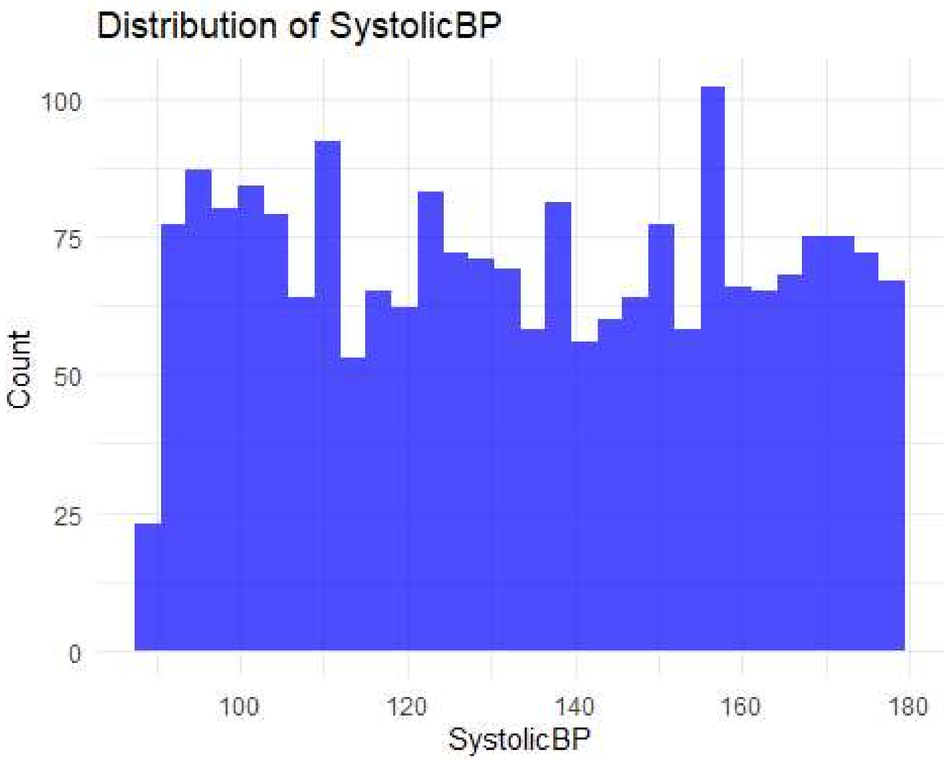


Distribution of UPDRS

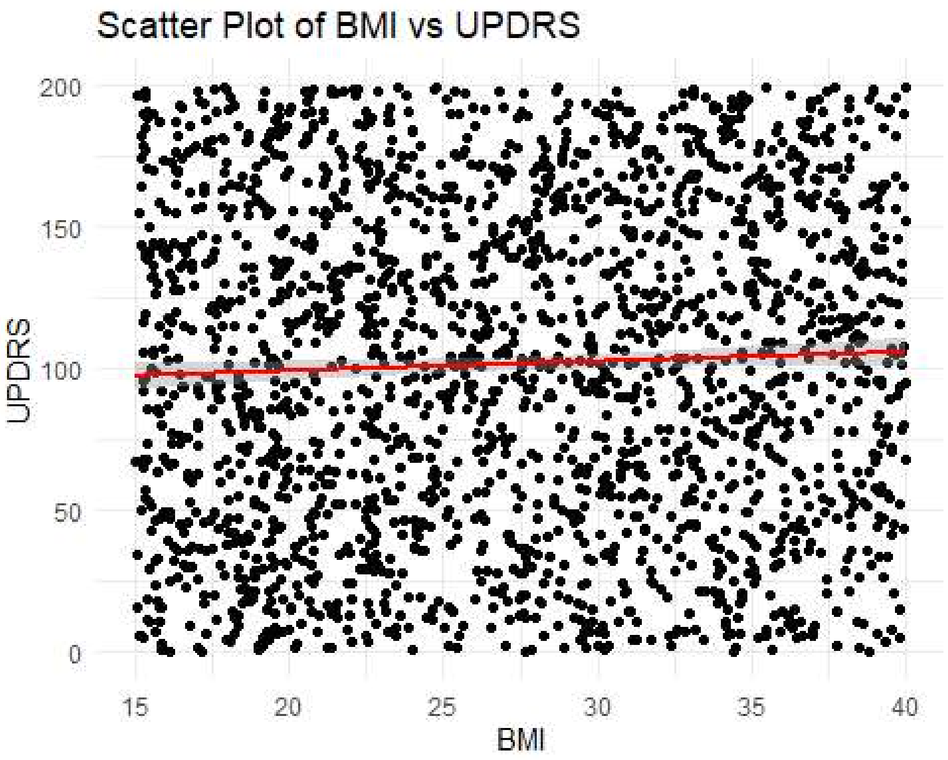



Distribution of MoCA
Count
0
25
50
75
0
10
20
30
MoCA

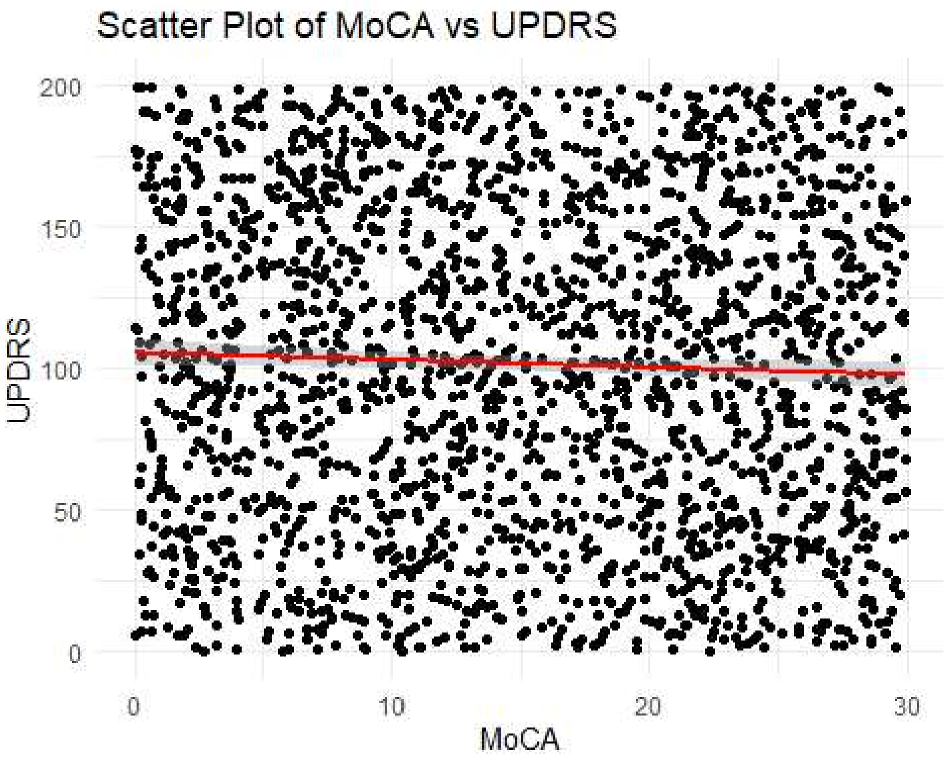

Distribution of PhysicalActivity
Count
0
25
50
75
0.0
2.5
5.0
7.5
10.0
PhysicalActivity



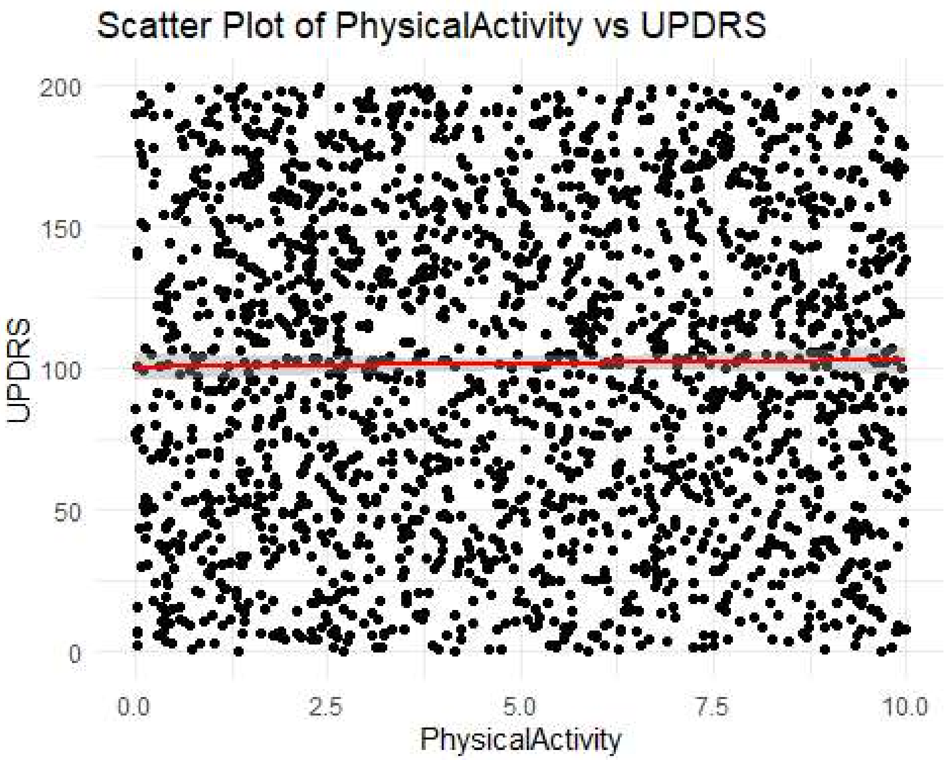
```
# Scatter plot function with regression line
scatter_plot_with_regression <- function(predictor, outcome, dataset) {
  tryCatch({
    ggplot(dataset, aes(x = !!sym(predictor), y = !!sym(outcome))) +
      geom_point() +
      geom_smooth(method = "lm", color = "red") +
      labs(title = paste("Scatter Plot of", predictor, "vs", outcome),
           x = predictor, y = outcome) +
      theme_minimal()
  }, error = function(e) {
    return(paste("Error in plotting", predictor, "vs", outcome, ":",
e$message))
  })
}

# Variables to plot against UPDRS (adjust if needed)
predictors <- c("BMI", "MoCA", "PhysicalActivity", "SystolicBP")
outcome <- "UPDRS"

# Loop through and plot
for (var in predictors) {
  print(scatter_plot_with_regression(var, outcome, dataset))
}

## `geom_smooth()` using formula = 'y ~ x'
```



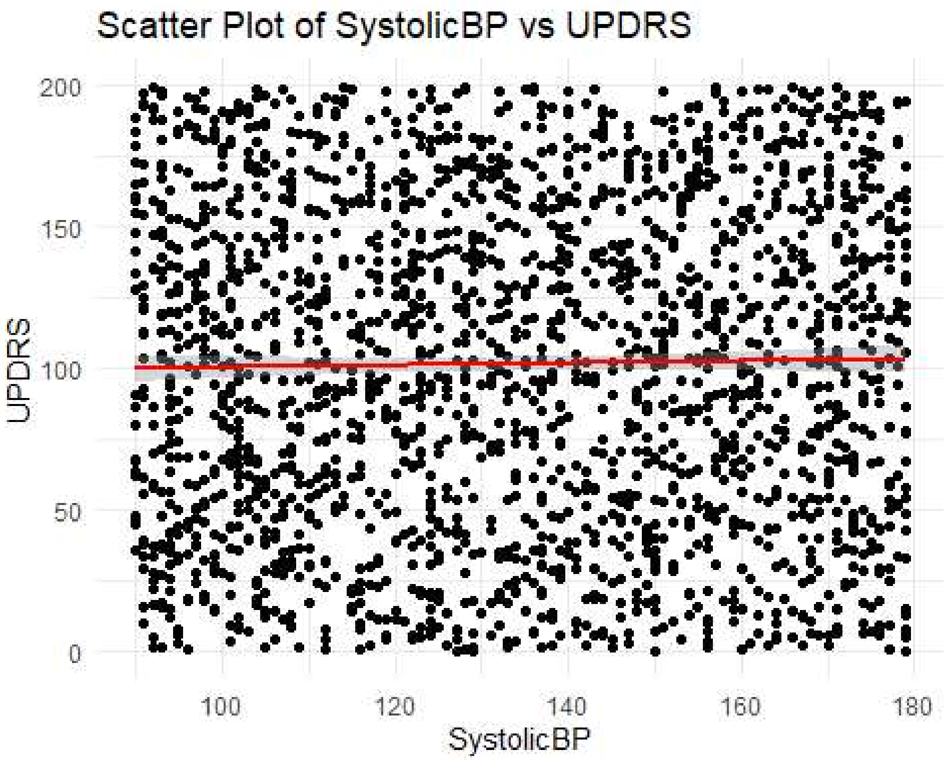
```
## `geom_smooth()` using formula = 'y ~ x'
```



```
## `geom_smooth()` using formula = 'y ~ x'
```



```
## `geom_smooth()` using formula = 'y ~ x'
```



``` Correlations:
Because the regression lines in all of the scatter plots are almost flat, it appears that none of the variables (BMI, SystolicBP, PhysicalActivity, MoCA, and Physical Activity)
```

have a significant linear association with UPDRS. Distributions: Each variable’s data is approximately uniform, with no significant skewness or strong trends in any direction, according to the histograms for the distributions.