

EDA on Human Activity recognition

2023-09-08

Load dataset

```
d513 <- read.csv("C:\\Users\\windows 10 pro\\Documents\\har70\\har70plus\\513.csv")
library(readr)
# Display the first few rows of the dataset
head(d513)
```

```
## timestamp back_x back_y back_z thigh_x thigh_y thigh_z label
## 1 44337.47 -0.968262 -0.109375 0.012451 -0.964600 0.162842 -0.213135 standing
## 2 44337.47 -0.973145 -0.109375 0.012451 -0.964600 0.160645 -0.217041 standing
## 3 44337.47 -0.987793 -0.109375 0.012451 -0.966553 0.174805 -0.216309 standing
## 4 44337.47 -0.987549 -0.109375 0.012451 -0.949707 0.178711 -0.218506 standing
## 5 44337.47 -1.002686 -0.109375 0.012939 -0.961426 0.179932 -0.203125 standing
## 6 44337.47 -1.001953 -0.109375 0.028564 -0.964844 0.165039 -0.200195 standing
```

```
# Get summary statistics
summary(d513)
```

```
## timestamp back_x back_y back_z
## Min. :44337 Min. :-1.5686 Min. :-0.913330 Min. :-0.76123
## 1st Qu.:44337 1st Qu.: -0.9993 1st Qu.: -0.102051 1st Qu.: -0.14990
## Median :44337 Median : -0.9812 Median : -0.041504 Median : -0.06909
## Mean :44337 Mean : -0.8779 Mean : -0.086929 Mean : 0.03086
## 3rd Qu.:44337 3rd Qu.: -0.9346 3rd Qu.: 0.002686 3rd Qu.: 0.05640
## Max. :44338 Max. : 0.1455 Max. : 0.429932 Max. : 1.03101
## thigh_x thigh_y thigh_z label
## Min. :-3.3787 Min. :-1.669922 Min. :-2.41675 Length:123599
## 1st Qu.: -0.9849 1st Qu.: -0.008545 1st Qu.: -0.83032 Class :character
## Median : -0.9358 Median : 0.085449 Median : -0.23145 Mode :character
## Mean : -0.6614 Mean : 0.130380 Mean : -0.40402
## 3rd Qu.: -0.1309 3rd Qu.: 0.167725 3rd Qu.: -0.08032
## Max. : 0.4199 Max. : 1.781250 Max. : 0.90088
```

```
# Display the structure of the dataset
str(d513)
```

```
## 'data.frame': 123599 obs. of 8 variables:
## $ timestamp: num 44337 44337 44337 44337 44337 ...
## $ back_x : num -0.968 -0.973 -0.988 -0.988 -1.003 ...
## $ back_y : num -0.109 -0.109 -0.109 -0.109 -0.109 ...
## $ back_z : num 0.0125 0.0125 0.0125 0.0125 0.0129 ...
## $ thigh_x : num -0.965 -0.965 -0.967 -0.95 -0.961 ...
## $ thigh_y : num 0.163 0.161 0.175 0.179 0.18 ...
## $ thigh_z : num -0.213 -0.217 -0.216 -0.219 -0.203 ...
## $ label : chr "standing" "standing" "standing" "standing" ...
```

```
# Check for missing values
any(is.na(d513))
```

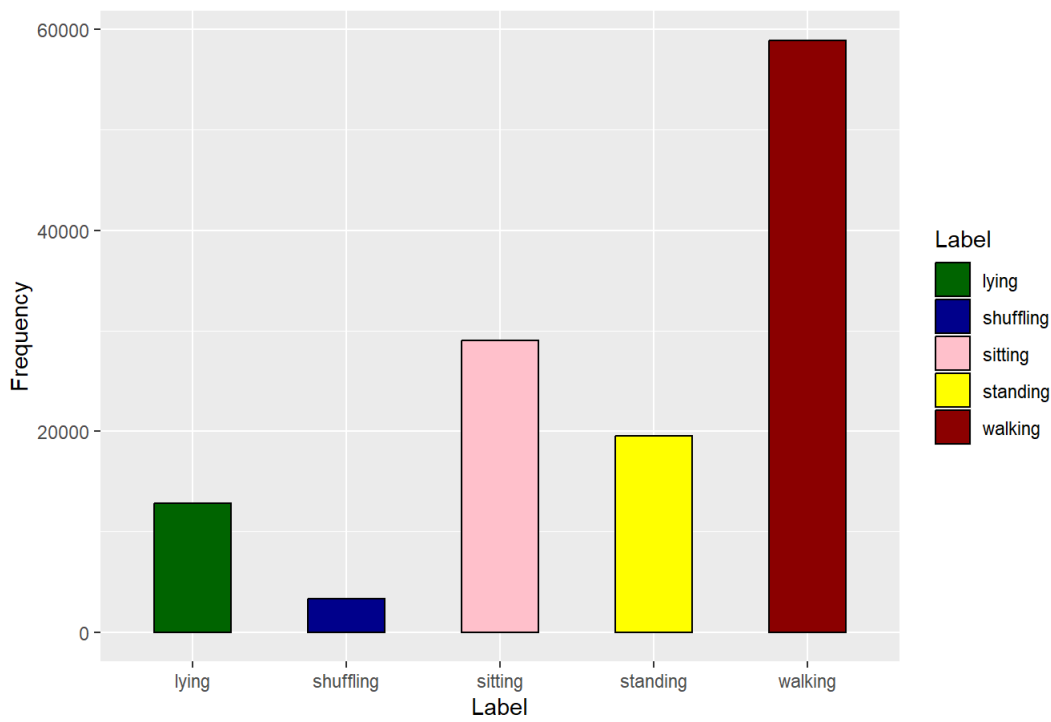
```
## [1] FALSE
```

Data Visualization

```
library(ggplot2)
# Create the histogram
ggplot(d513, aes(x = label, fill = label)) +
  geom_histogram(stat = "count", width = 0.5, color = "black") +
  labs(title = "Histogram of label count", x = "Label", y = "Frequency", fill = "Label") +
  scale_fill_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

```
## Warning in geom_histogram(stat = "count", width = 0.5, color = "black"):
## Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```

Histogram of label count



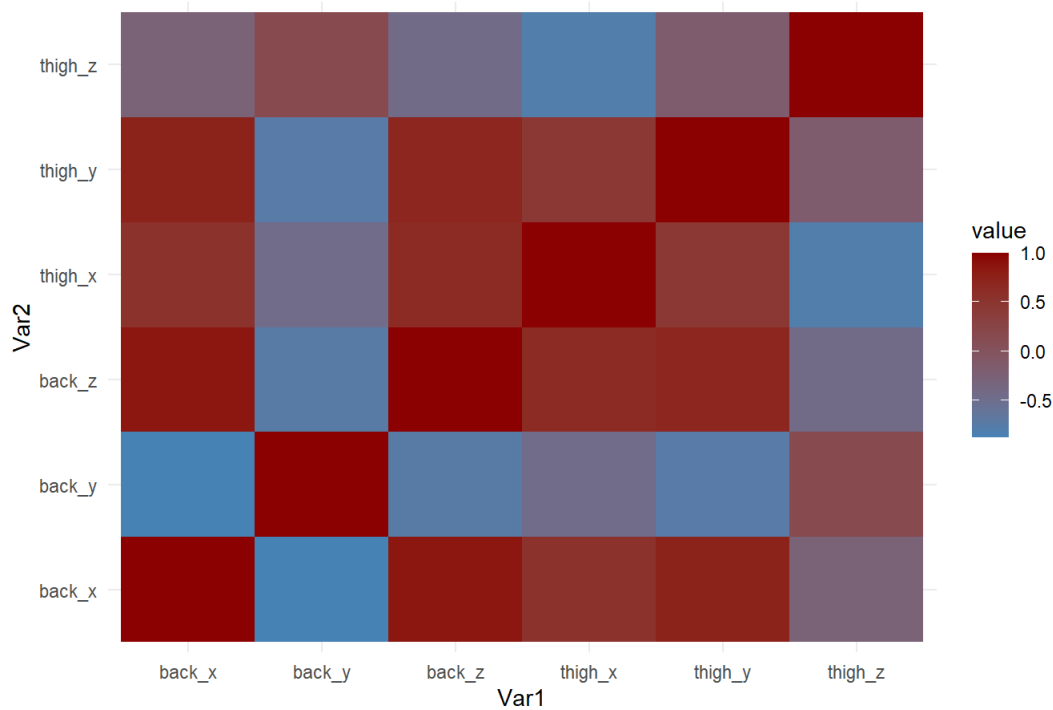
Correlation table

```
# create correlation matrix of (rounded to 2 decimal places)
round(cor(d513[c('back_x','back_y','back_z','thigh_x','thigh_y','thigh_z')]), 2)
```

```
##      back_x back_y back_z thigh_x thigh_y thigh_z
## back_x   1.00 -0.87  0.86   0.55   0.74  -0.26
## back_y  -0.87  1.00 -0.74  -0.46  -0.74   0.17
## back_z   0.86 -0.74  1.00   0.65   0.70  -0.43
## thigh_x  0.55 -0.46  0.65   1.00   0.48  -0.79
## thigh_y  0.74 -0.74  0.70   0.48   1.00  -0.16
## thigh_z -0.26  0.17 -0.43  -0.79  -0.16   1.00
```

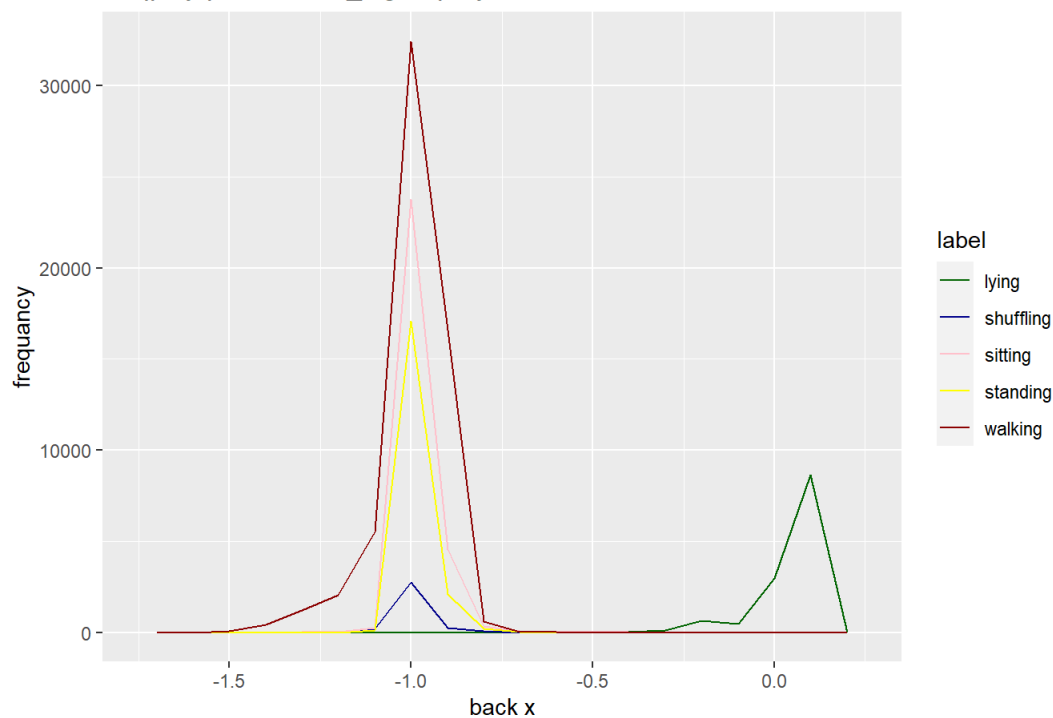
```
# Calculate the correlation matrix
cor_matrix <- cor(d513[, 2:7])
# Create a basic correlation heatmap
ggplot(data = reshape2::melt(cor_matrix)) +
  geom_tile(aes(Var1, Var2, fill = value)) +
  scale_fill_gradient(low = "steelblue", high = "darkred") +
  labs(title = "Correlation Heatmap") +
  theme_minimal()
```

Correlation Heatmap



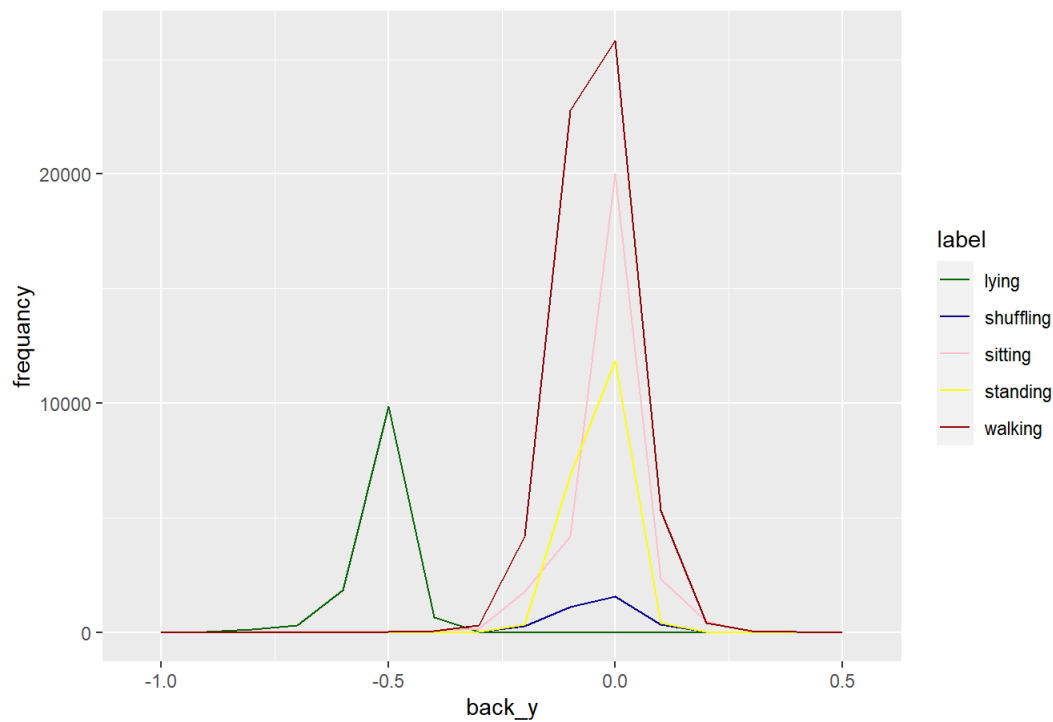
```
ggplot(d513, mapping = aes(x = back_x)) +
  geom_freqpoly(mapping = aes(colour = label), binwidth = 0.1) +
  labs(title = "freqpoly plot of back_x group by label",
       x = "back x",
       y = "frequency",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

freqpoly plot of back_x group by label



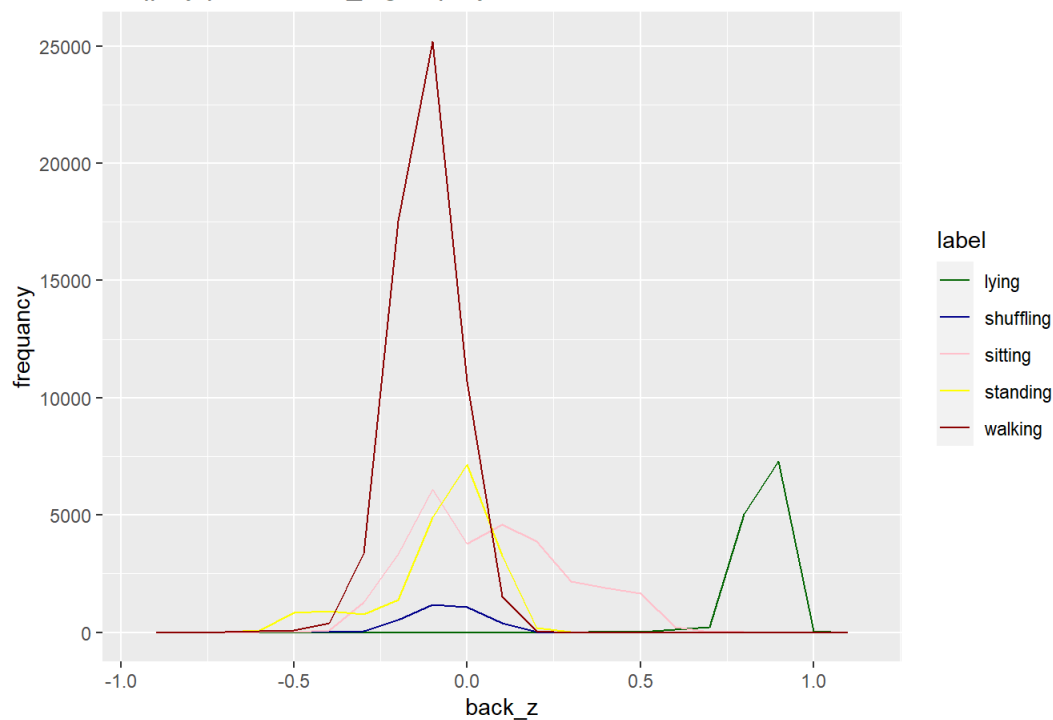
```
ggplot(d513, mapping = aes(x = back_y)) +
  geom_freqpoly(mapping = aes(colour = label), binwidth = 0.1) +
  labs(title = "freqpoly plot of back_y group by label",
       x = "back_y",
       y = "frequency",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

freqpoly plot of back_y group by label



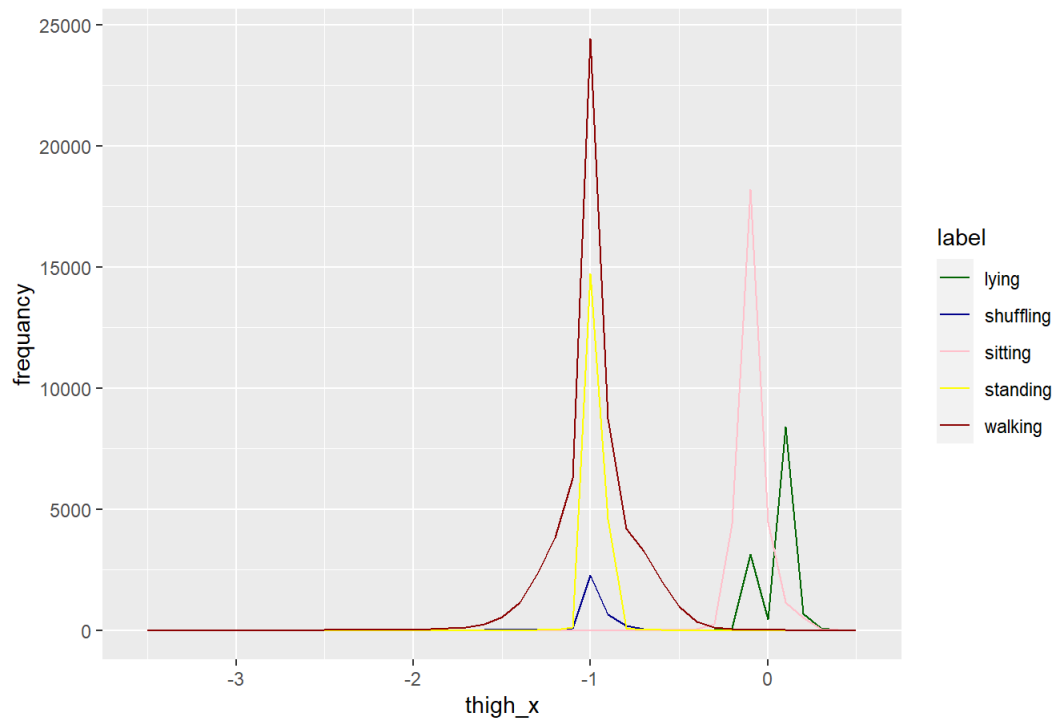
```
ggplot(d513, mapping = aes(x = back_z)) +
  geom_freqpoly(mapping = aes(colour = label), binwidth = 0.1) +
  labs(title = "freqpoly plot of back_z group by label",
       x = "back_z",
       y = "frequency",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

freqpoly plot of back_z group by label



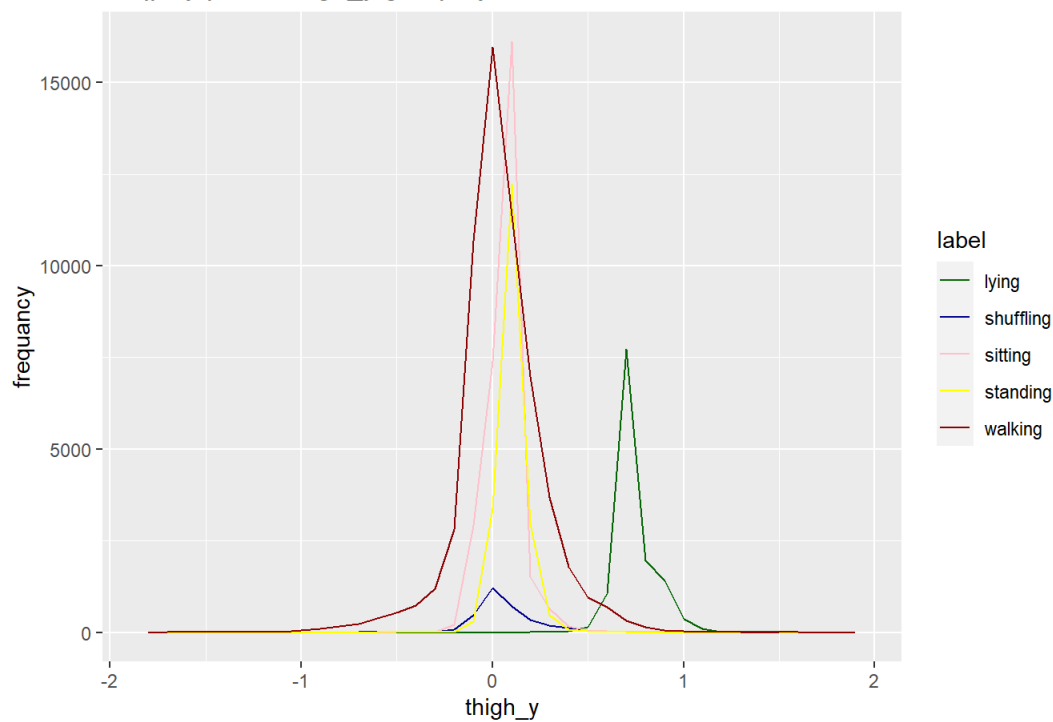
```
ggplot(d513, mapping = aes(x = thigh_x)) +
  geom_freqpoly(mapping = aes(colour = label), binwidth = 0.1) +
  labs(title = "freqpoly plot of thigh_x group by label",
       x = "thigh_x",
       y = "frequency",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

freqpoly plot of thigh_x group by label



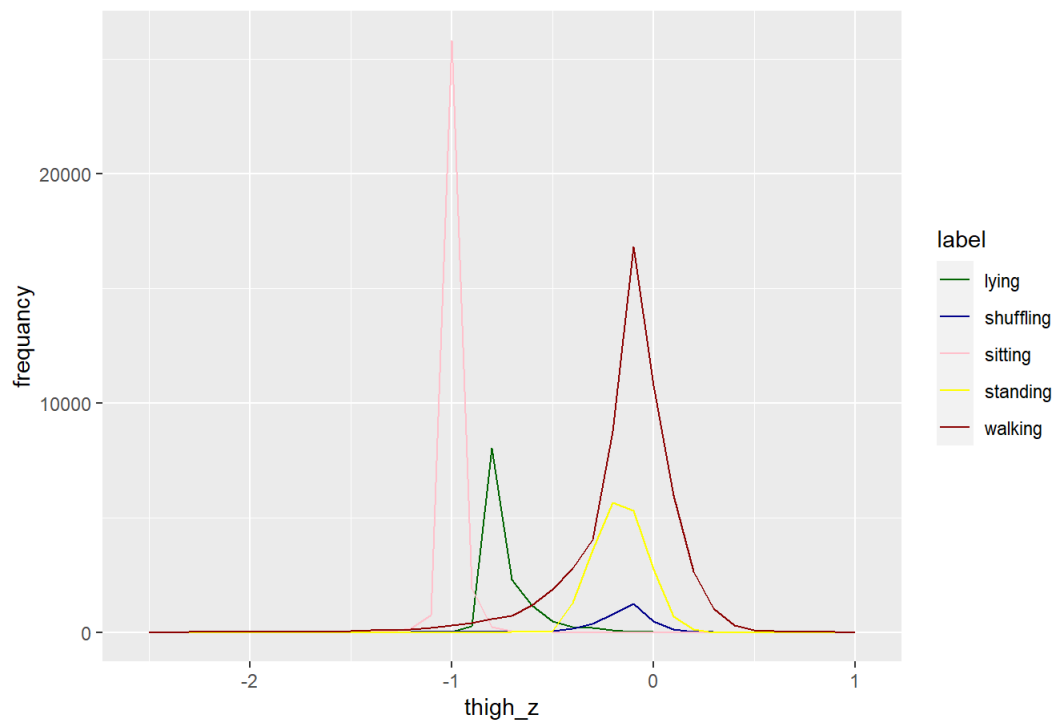
```
ggplot(d513, mapping = aes(x = thigh_y)) +
  geom_freqpoly(mapping = aes(colour = label), binwidth = 0.1) +
  labs(title = "freqpoly plot of thigh_y group by label",
       x = "thigh_y",
       y = "frequency",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

freqpoly plot of thigh_y group by label



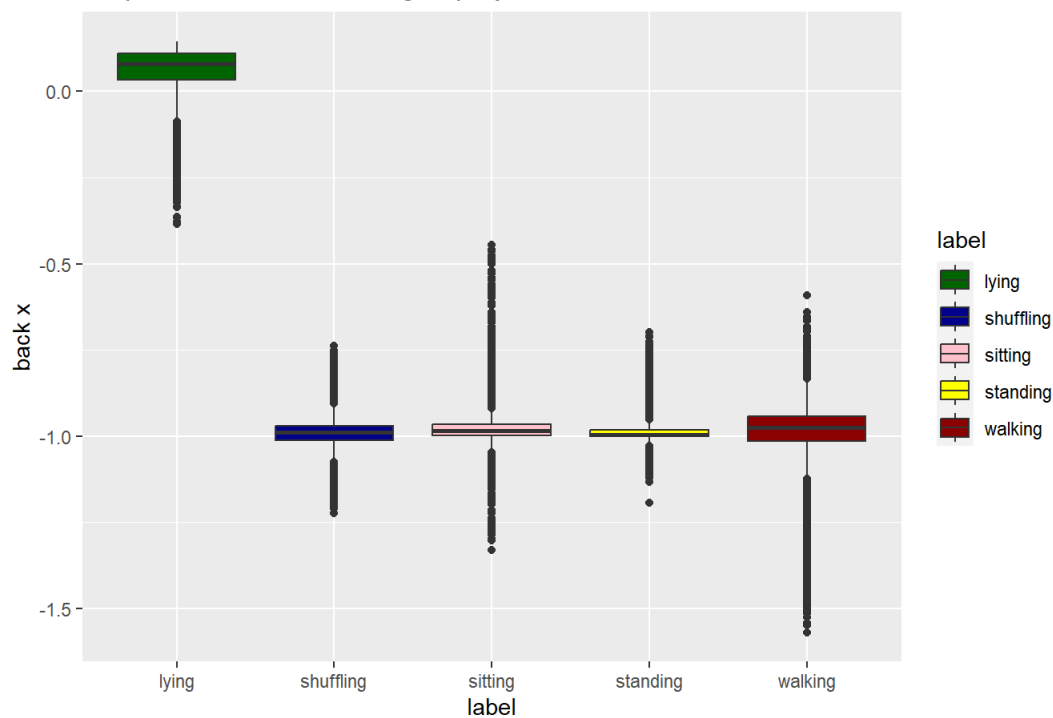
```
ggplot(d513, mapping = aes(x = thigh_z)) +
  geom_freqpoly(mapping = aes(colour = label), binwidth = 0.1) +
  labs(title = "freqpoly plot of thigh_z group by label",
       x = "thigh_z",
       y = "frequency",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

freqpoly plot of thigh_z group by label



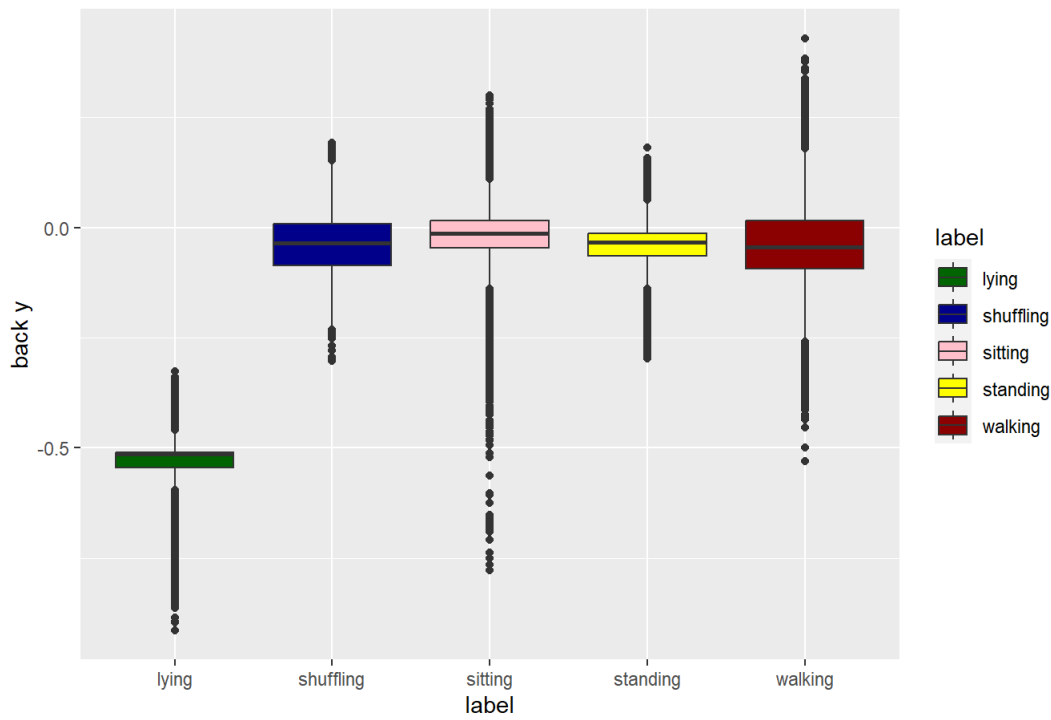
```
ggplot(d513, aes(x = label, y= back_x, fill = label)) +
  geom_boxplot() +
  labs(title = "boxplott of label vs back x group by label",
       x = "label",
       y = "back x",
       color = "label") +
  scale_fill_manual(values = c("darkgreen","darkblue","pink","yellow","darkred"))
```

boxplott of label vs back x group by label



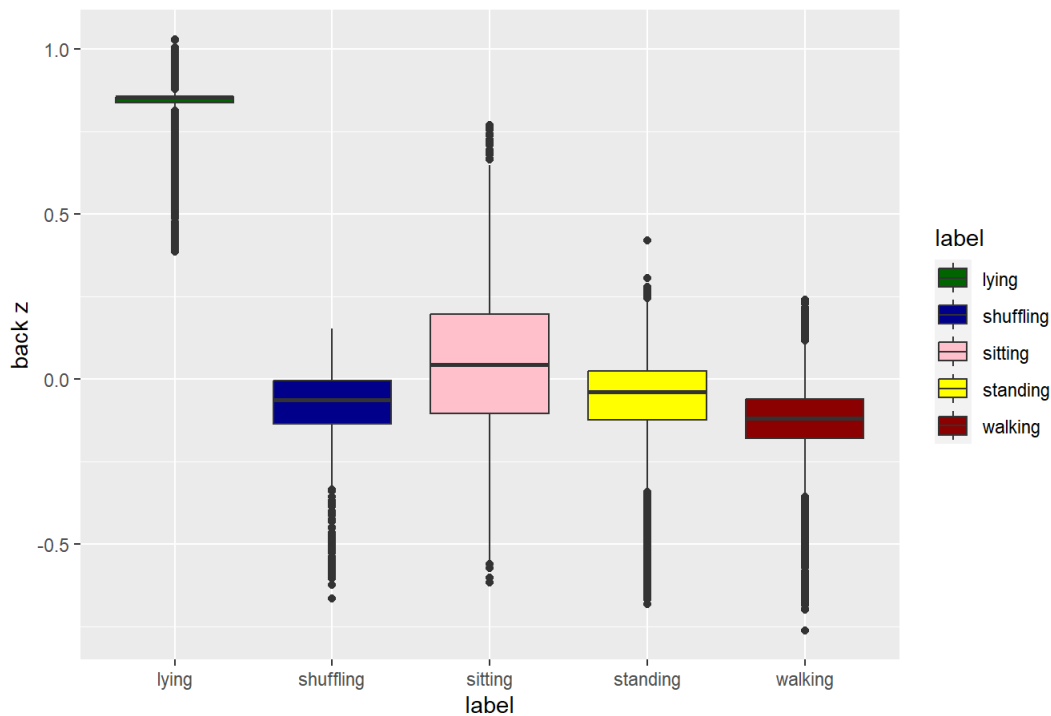
```
ggplot(d513, aes(x = label, y= back_y, fill = label)) +
  geom_boxplot() +
  labs(title = "boxplott of label vs back y group by label",
       x = "label",
       y = "back y",
       color = "label") +
  scale_fill_manual(values = c("darkgreen","darkblue","pink","yellow","darkred"))
```

boxplott of label vs back y group by label



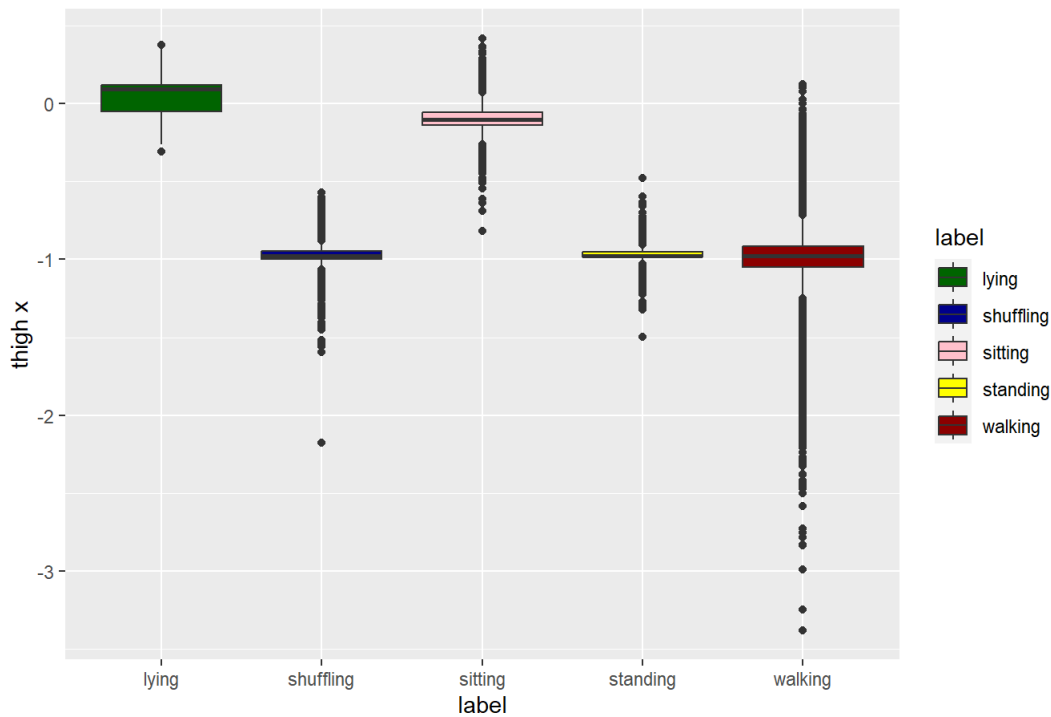
```
ggplot(d513, aes(x = label, y= back_z, fill = label)) +
  geom_boxplot() +
  labs(title = "boxplott of label vs back z group by label",
       x = "label",
       y = "back z",
       color = "label") +
  scale_fill_manual(values = c("darkgreen","darkblue","pink","yellow","darkred"))
```

boxplott of label vs back z group by label



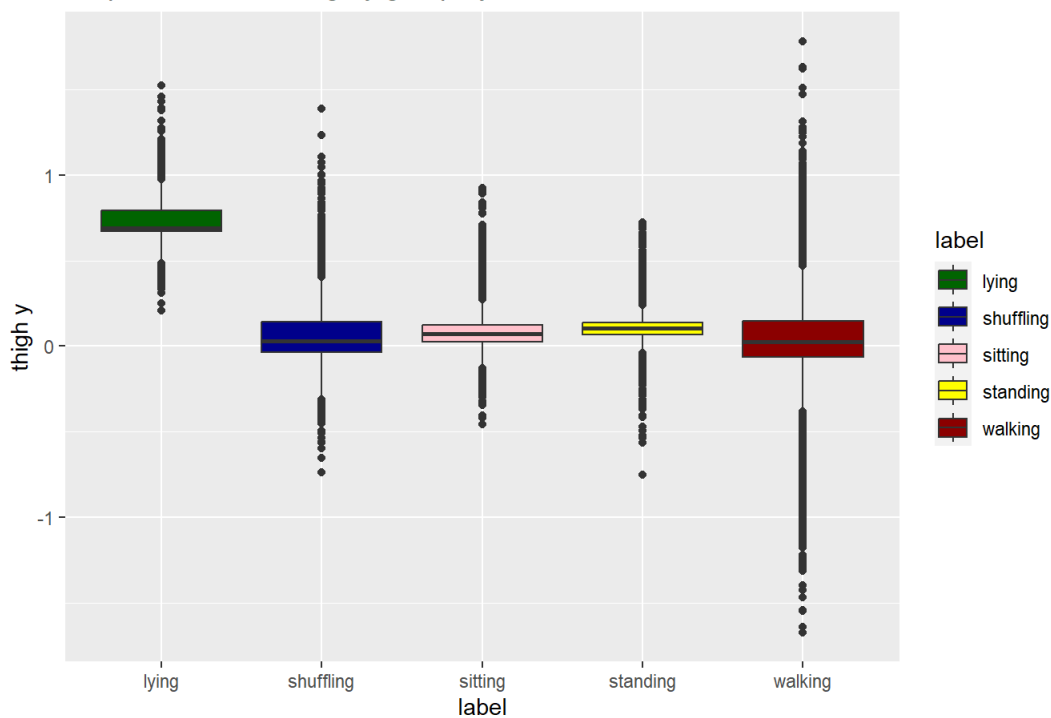
```
ggplot(d513, aes(x = label, y= thigh_x, fill = label)) +
  geom_boxplot() +
  labs(title = "boxplott of label vs thigh x group by label",
       x = "label",
       y = "thigh x",
       color = "label") +
  scale_fill_manual(values = c("darkgreen","darkblue","pink","yellow","darkred"))
```

boxplott of label vs thigh x group by label



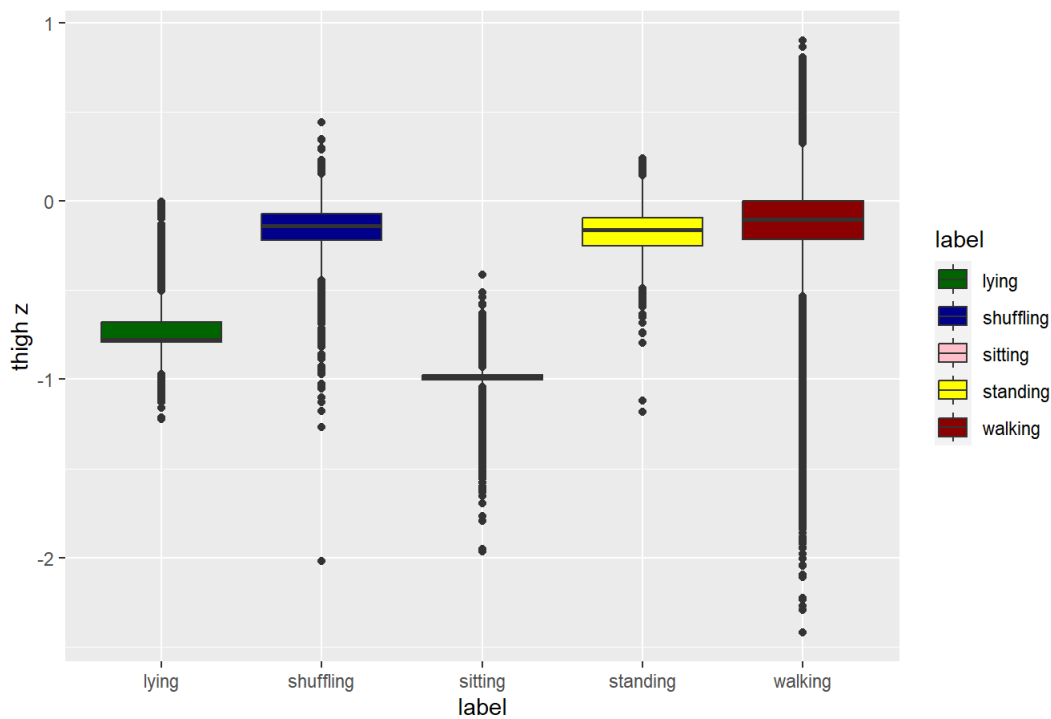
```
ggplot(d513, aes(x = label, y= thigh_y, fill = label)) +
  geom_boxplot() +
  labs(title = "boxplott of label vs thigh y group by label",
    x = "label",
    y = "thigh y",
    color = "label") +
  scale_fill_manual(values = c("darkgreen","darkblue","pink","yellow","darkred"))
```

boxplott of label vs thigh y group by label



```
ggplot(d513, aes(x = label, y= thigh_z, fill = label)) +
  geom_boxplot() +
  labs(title = "boxplott of label vs thigh z group by label",
    x = "label",
    y = "thigh z",
    color = "label") +
  scale_fill_manual(values = c("darkgreen","darkblue","pink","yellow","darkred"))
```

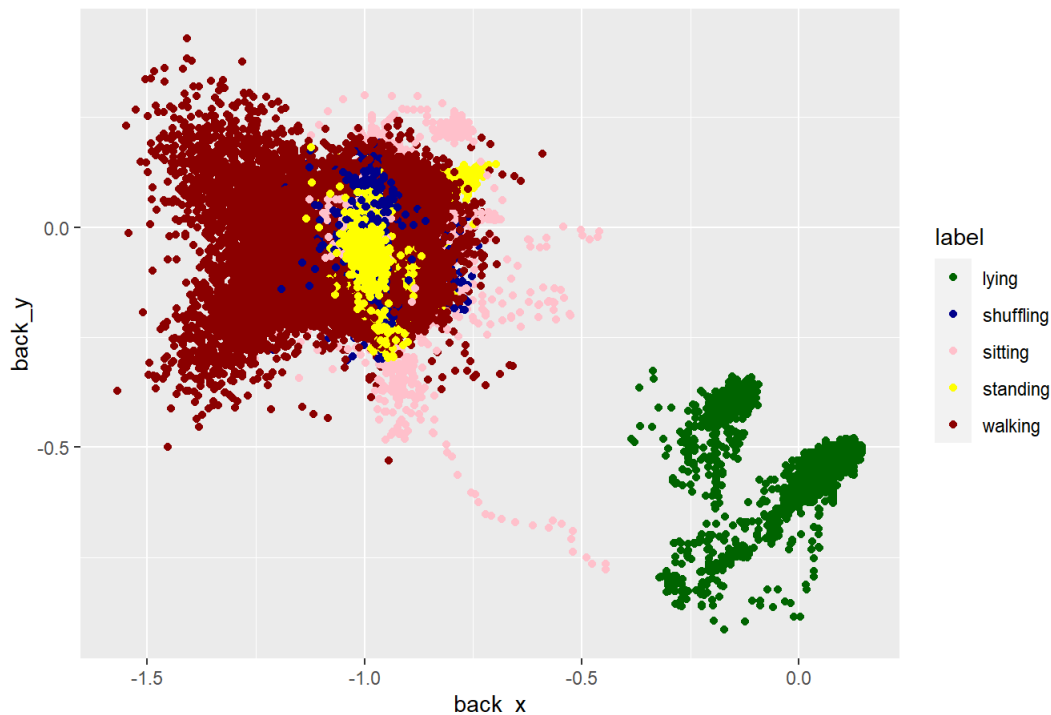

boxplott of label vs thigh z group by label



For more understanding of the correlation table, i will make use of scatter plot to view the column with values more than 0.5 from the correlation table.

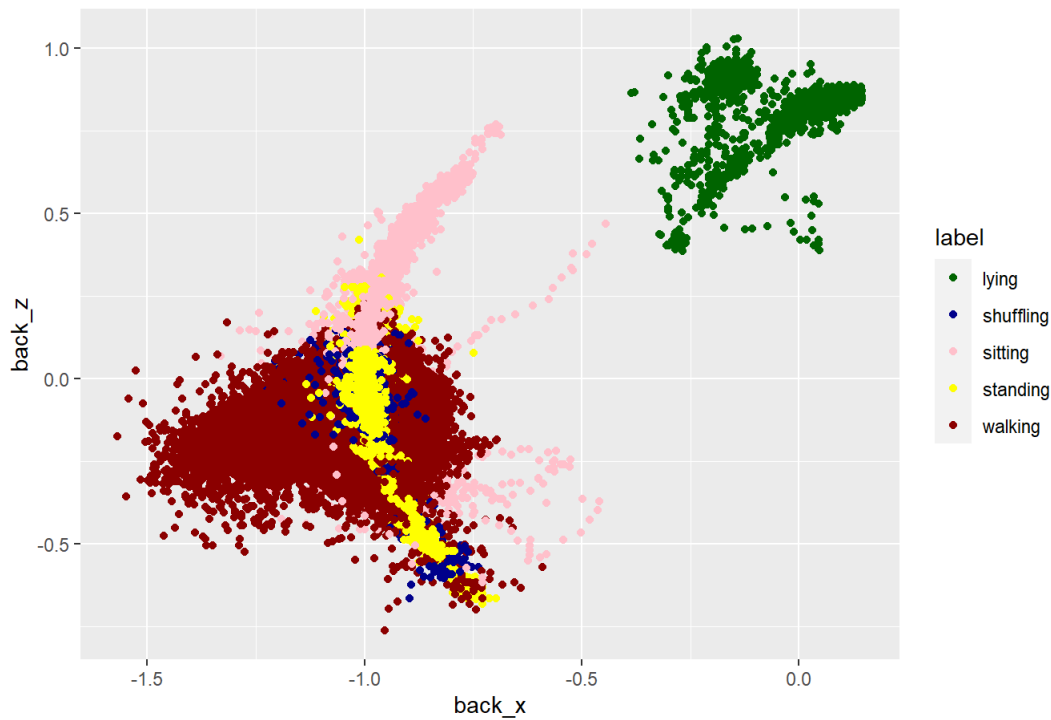
```
# From correlation table (Back x vs back y) = -0.87
ggplot(d513, aes(x= back_x, y = back_y, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of back_x vs back_y group by label",
    x = "back_x",
    y = "back_y",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of back_x vs back_y group by label



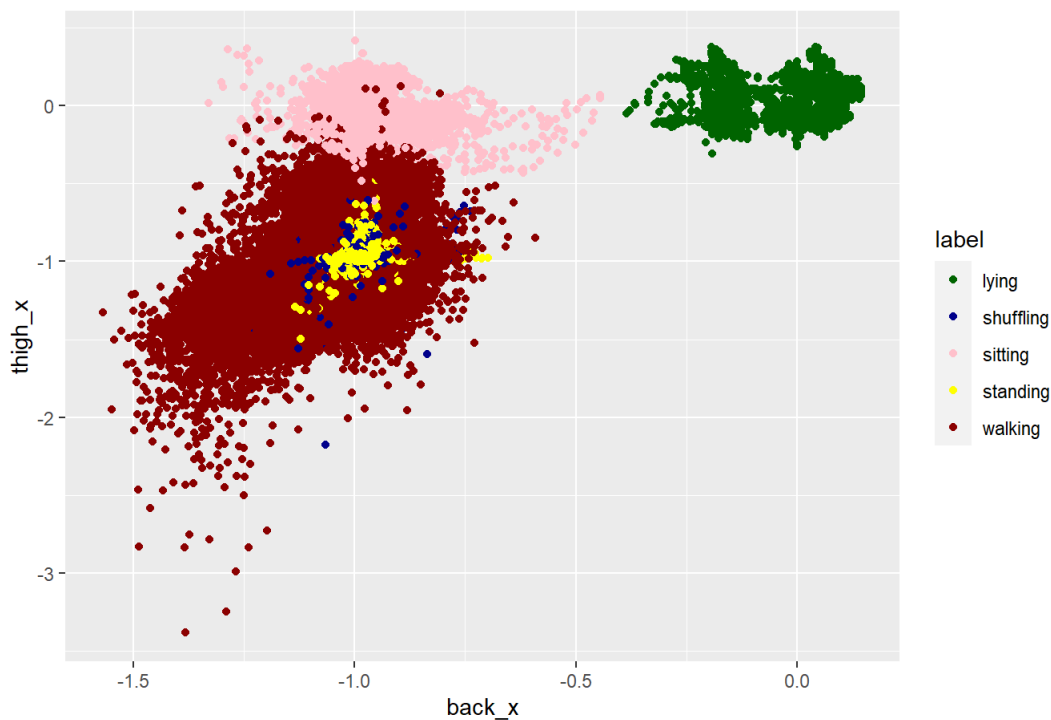
```
# From correlation table (Back x vs back z) = 0.86
ggplot(d513, aes(x= back_x, y = back_z, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of back_x vs back_z group by label",
    x = "back_x",
    y = "back_z",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of back_x vs back_z group by label



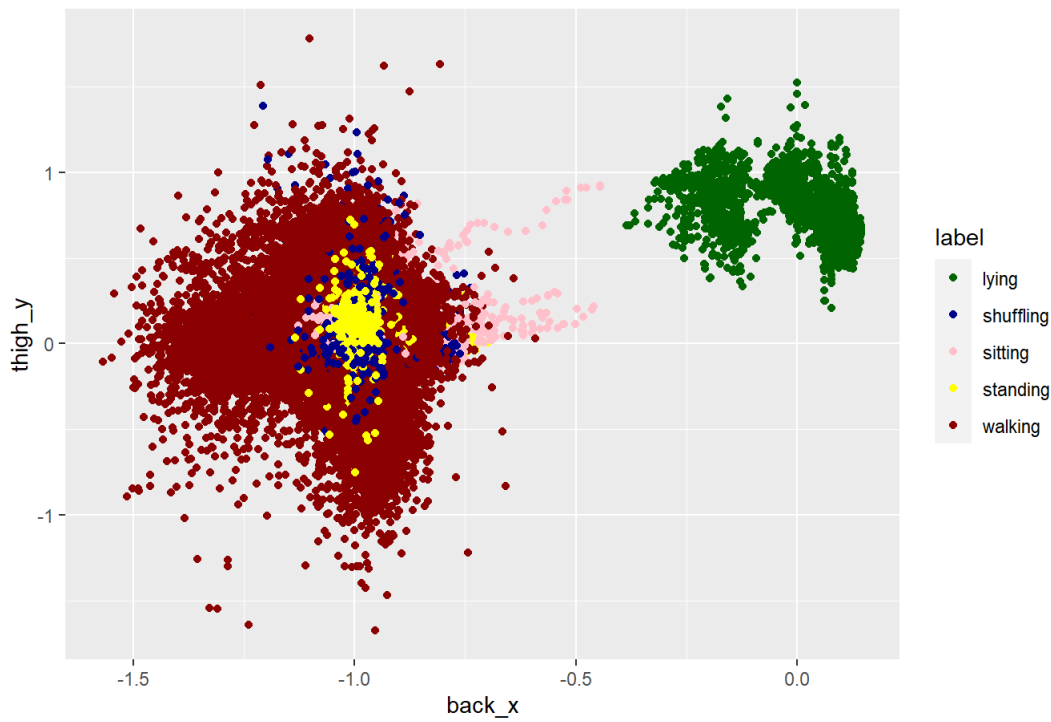
```
# From correlation table (Back x vs thigh x) = 0.55
ggplot(d513, aes(x= back_x, y = thigh_x, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of back_x vs thigh_x group by label",
    x = "back_x",
    y = "thigh_x",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of back_x vs thigh_x group by label



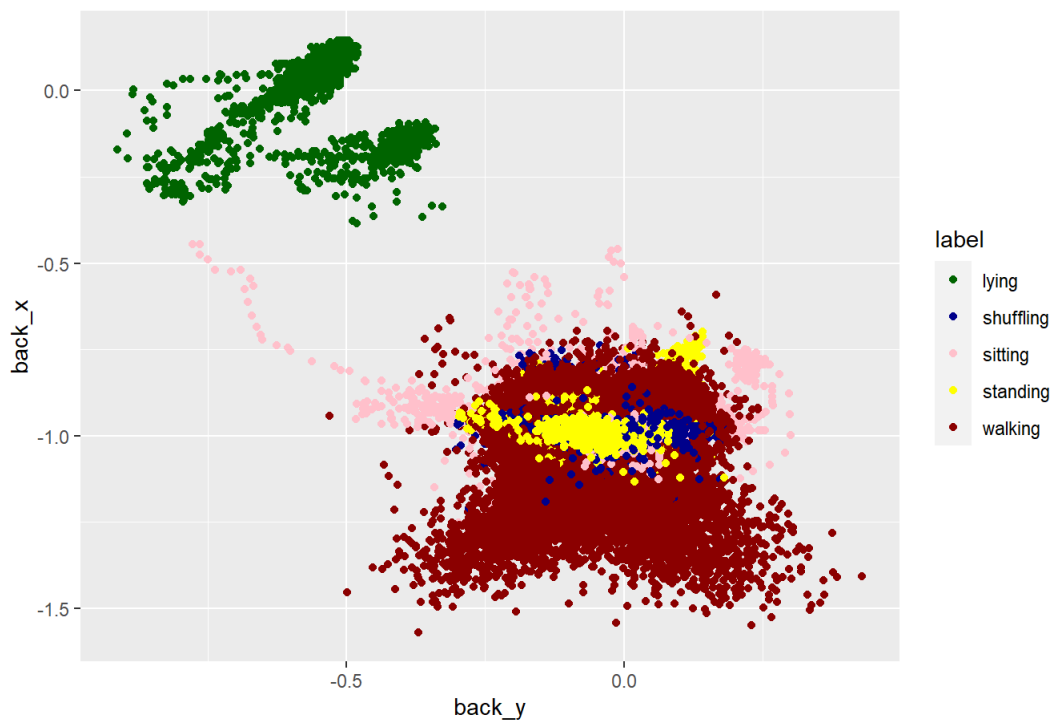
```
# From correlation table (Back x vs thigh y) = 0.74
ggplot(d513, aes(x= back_x, y = thigh_y, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of back_x vs thigh_y group by label",
    x = "back_x",
    y = "thigh_y",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of back_x vs thigh_y group by label



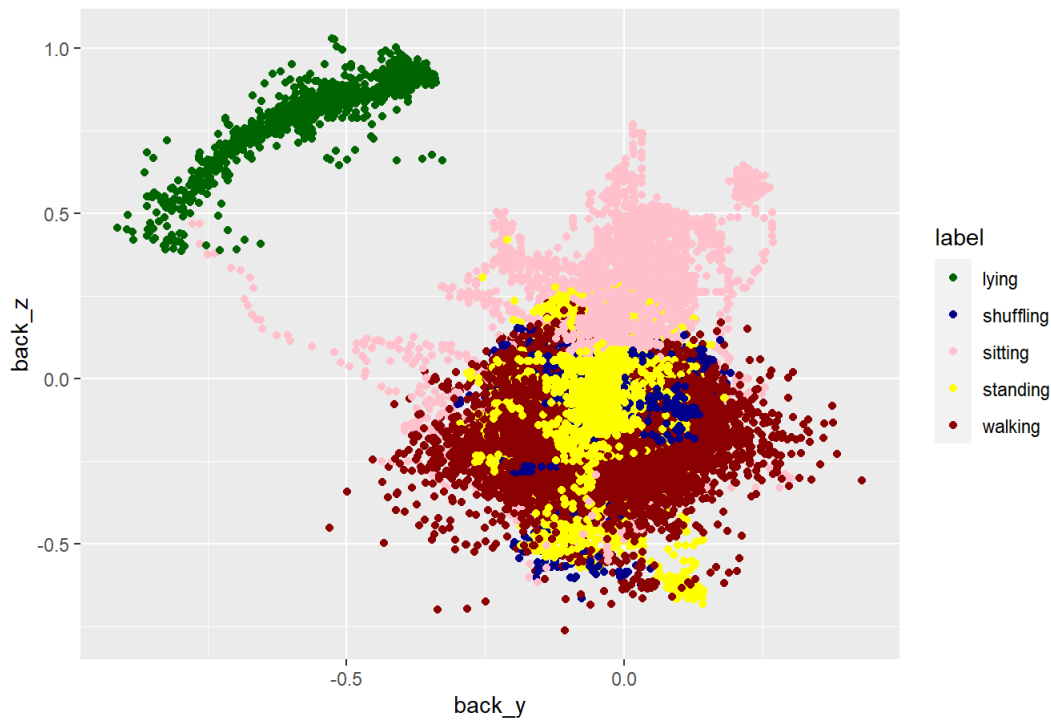
```
# From correlation table (Back y vs back x) = -0.87
ggplot(d513, aes(x= back_y, y = back_x, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of back_y vs back_x group by label",
    x = "back_y",
    y = "back_x",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of back_y vs back_x group by label



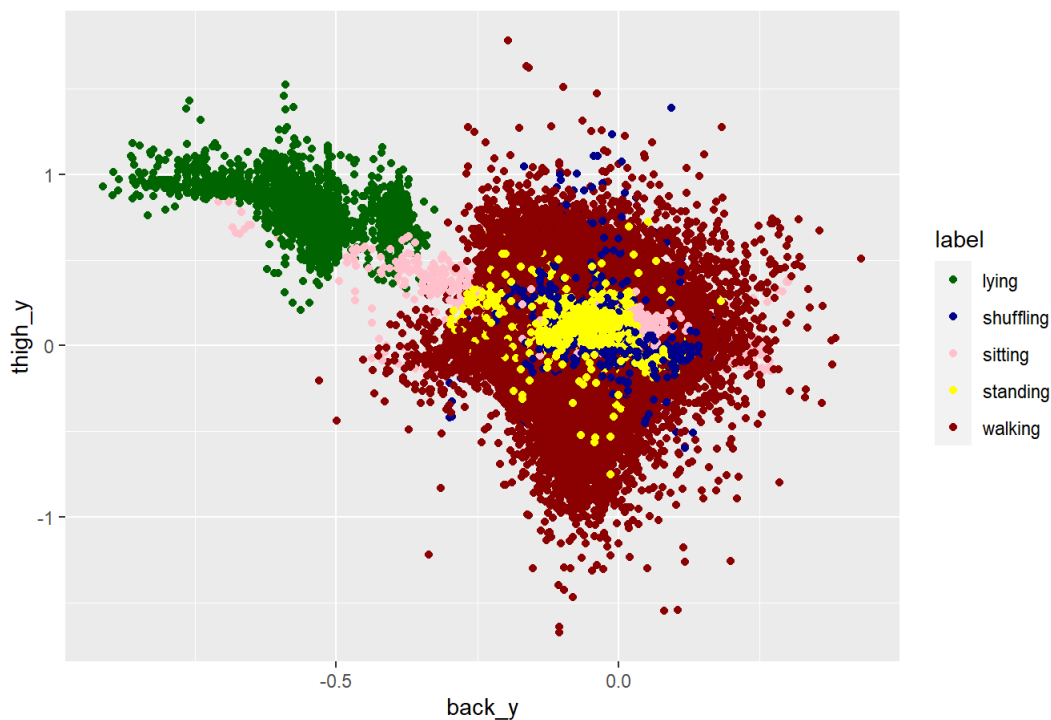
```
# From correlation table (Back y vs back z) = -0.74
ggplot(d513, aes(x= back_y, y = back_z, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of back_y vs back_z group by label",
    x = "back_y",
    y = "back_z",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of back_y vs back_z group by label



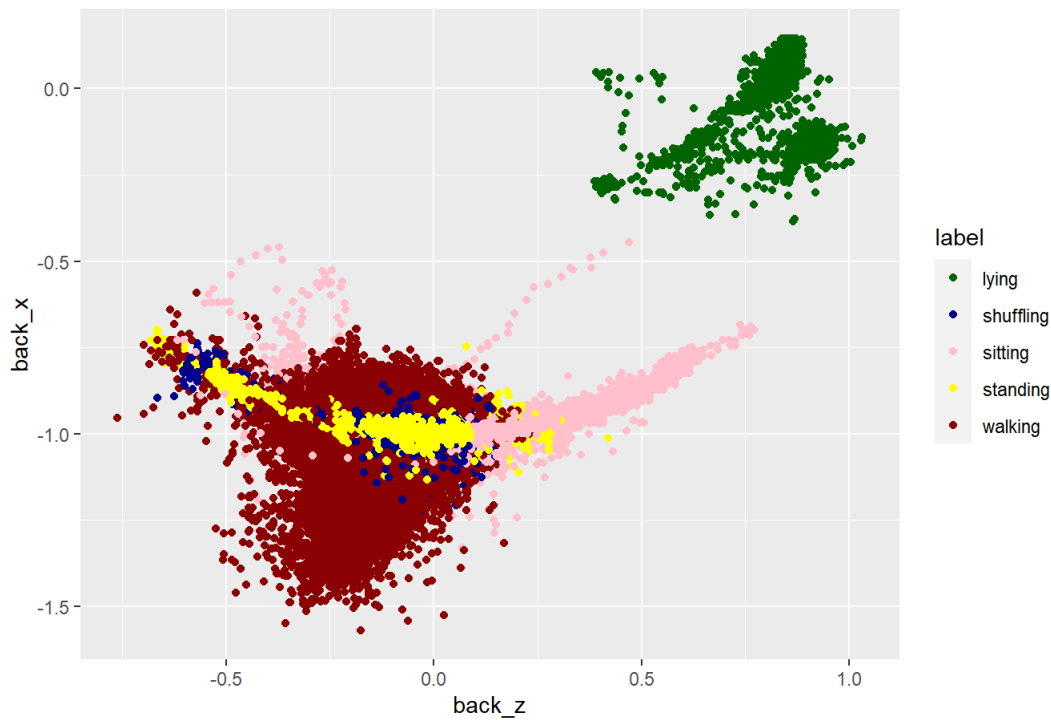
```
# From correlation table (Back y vs thigh y) = -0.74
ggplot(d513, aes(x= back_y, y = thigh_y, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of back_y vs thigh_y group by label",
       x = "back_y",
       y = "thigh_y",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of back_y vs thigh_y group by label



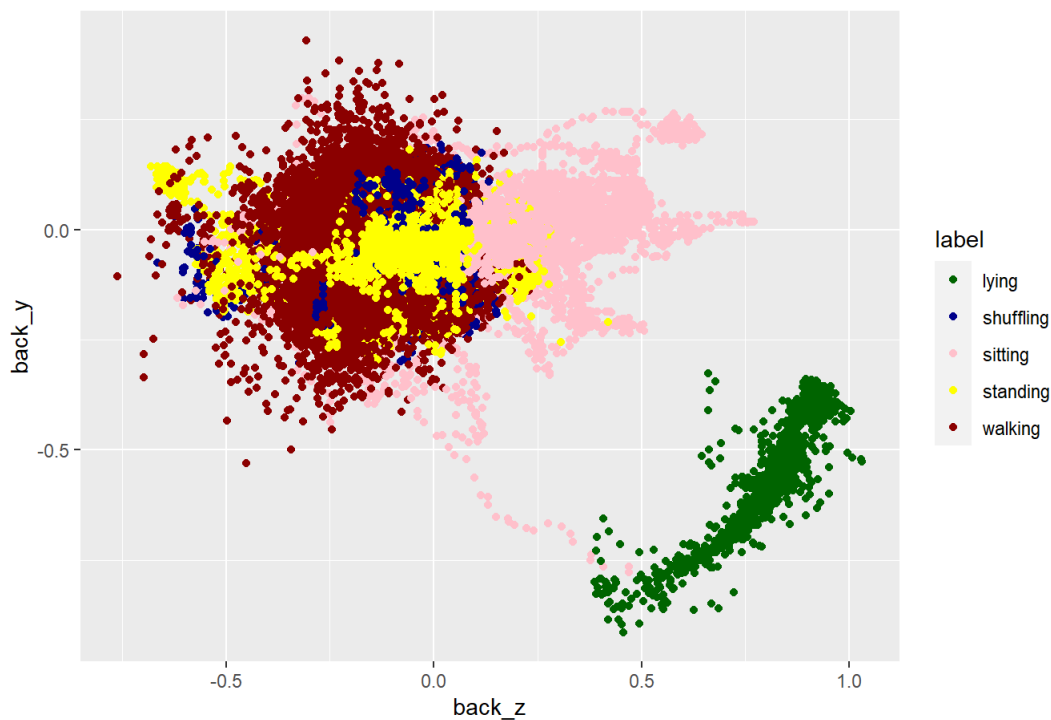
```
# From correlation table (Back z vs back x) = 0.86
ggplot(d513, aes(x= back_z, y = back_x, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of back_z vs back_x group by label",
       x = "back_z",
       y = "back_x",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of back_z vs back_x group by label



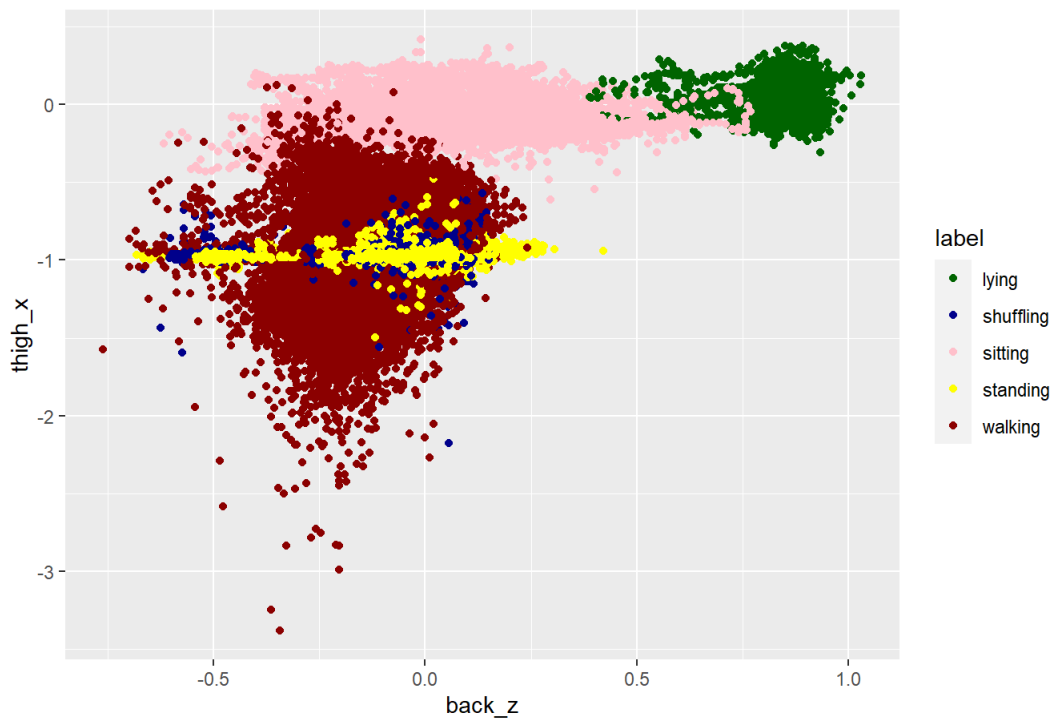
```
# From correlation table (Back z vs back y) = 0.74
ggplot(d513, aes(x= back_z, y = back_y, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of back_z vs back_y group by label",
       x = "back_z",
       y = "back_y",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of back_z vs back_y group by label



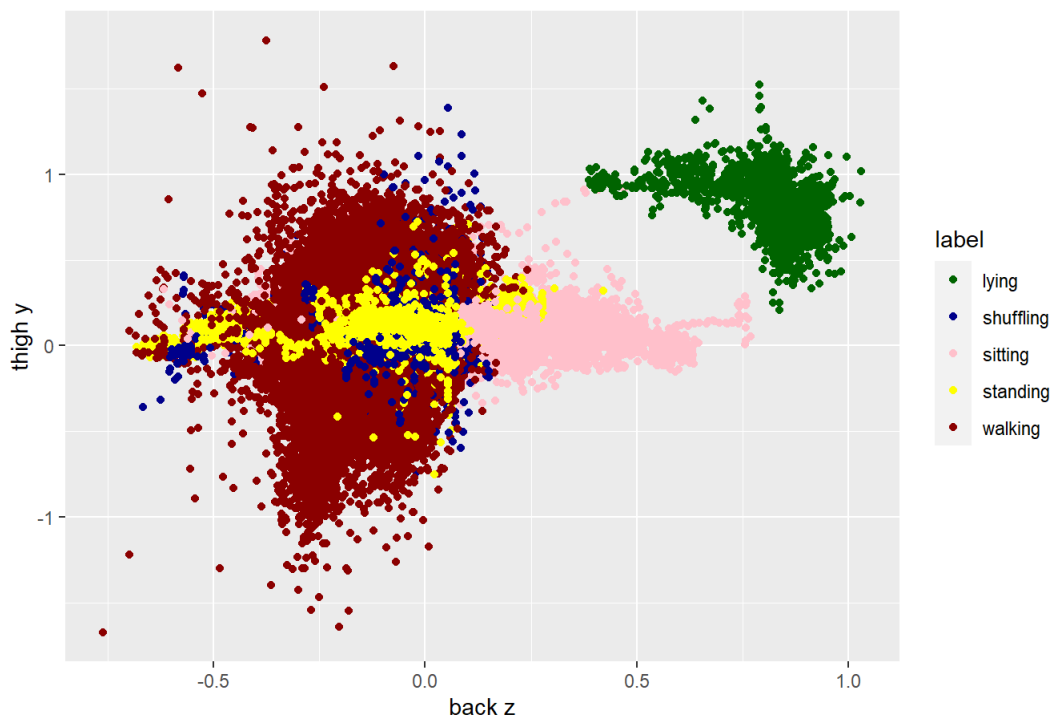
```
# From correlation table (Back z vs thigh x) = 0.65
ggplot(d513, aes(x= back_z, y = thigh_x, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of back_z vs thigh_x group by label",
       x = "back_z",
       y = "thigh_x",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of back_z vs thigh_x group by label



```
# From correlation table (Back z vs thigh y) = 0.70
ggplot(d513, aes(x= back_z, y = thigh_y, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of back z vs thigh y group by label",
    x = "back z",
    y = "thigh y",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of back z vs thigh y group by label



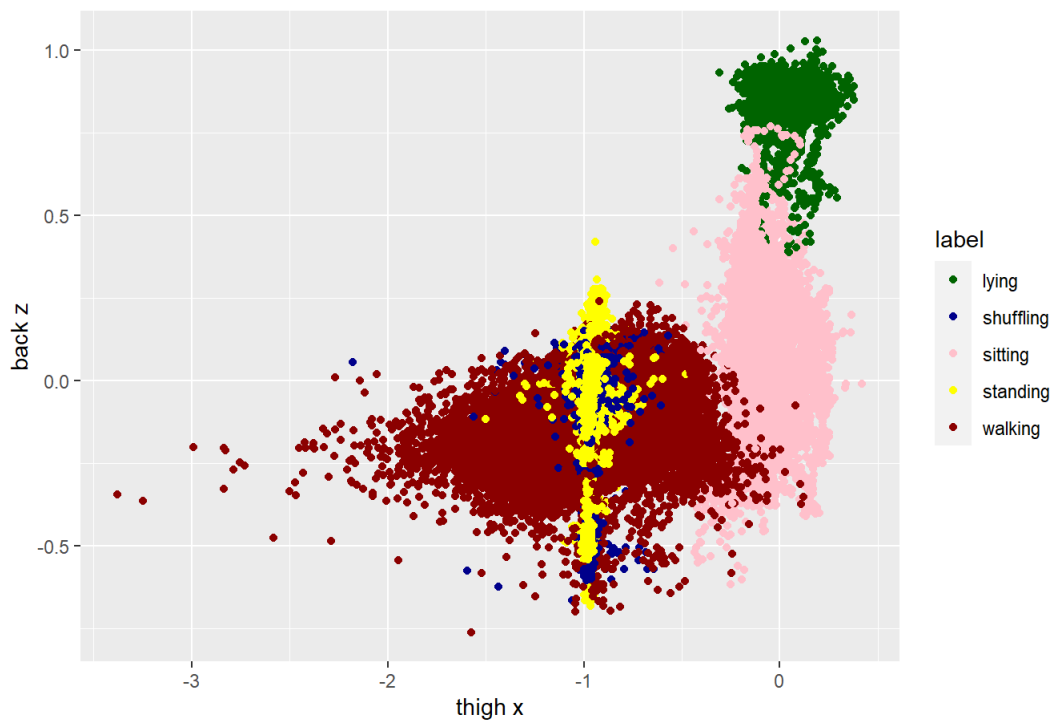
```
# From correlation table (thigh x vs back x) = 0.55
ggplot(d513, aes(x= thigh_x, y = back_x, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of thigh x vs back x group by label",
    x = "thigh x",
    y = "back x",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of thigh x vs back x group by label



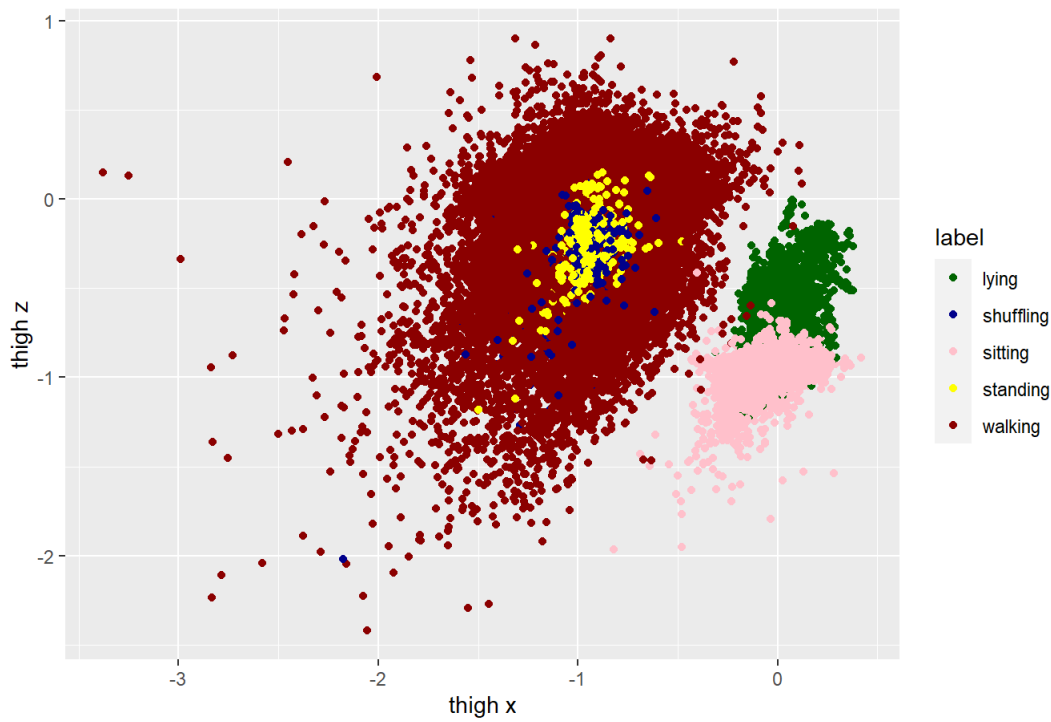
```
# From correlation table (thigh x vs back z) = 0.65
ggplot(d513, aes(x= thigh_x, y = back_z, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of thigh x vs back z group by label",
    x = "thigh x",
    y = "back z",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of thigh x vs back z group by label



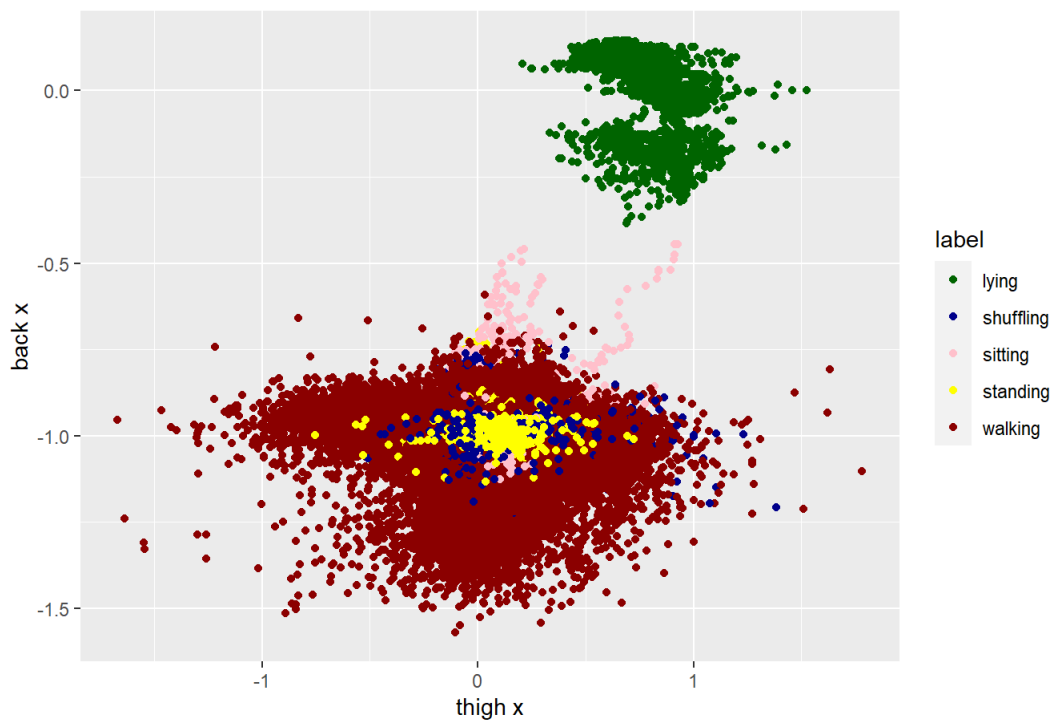
```
# From correlation table (thigh x vs thigh z) = -0.79
ggplot(d513, aes(x= thigh_x, y = thigh_z, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of thigh x vs thigh z group by label",
    x = "thigh x",
    y = "thigh z",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of thigh x vs thigh z group by label



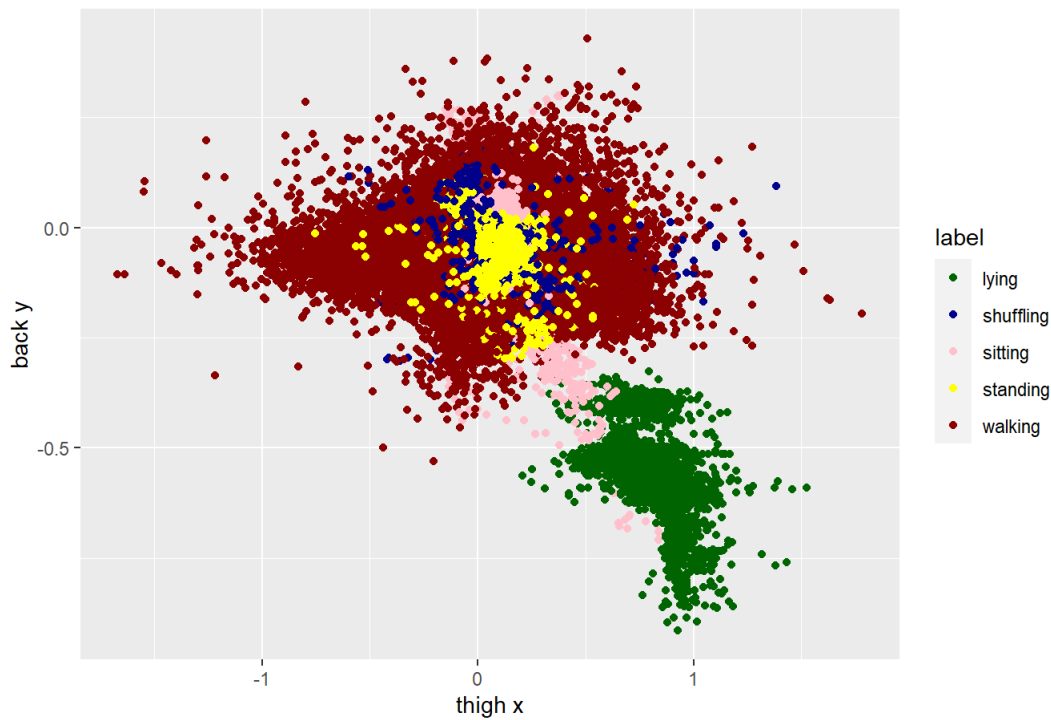
```
# From correlation table (thigh y vs back x) = 0.74
ggplot(d513, aes(x= thigh_y, y = back_x, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of thigh y vs back x group by label",
    x = "thigh x",
    y = "back x",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of thigh y vs back x group by label



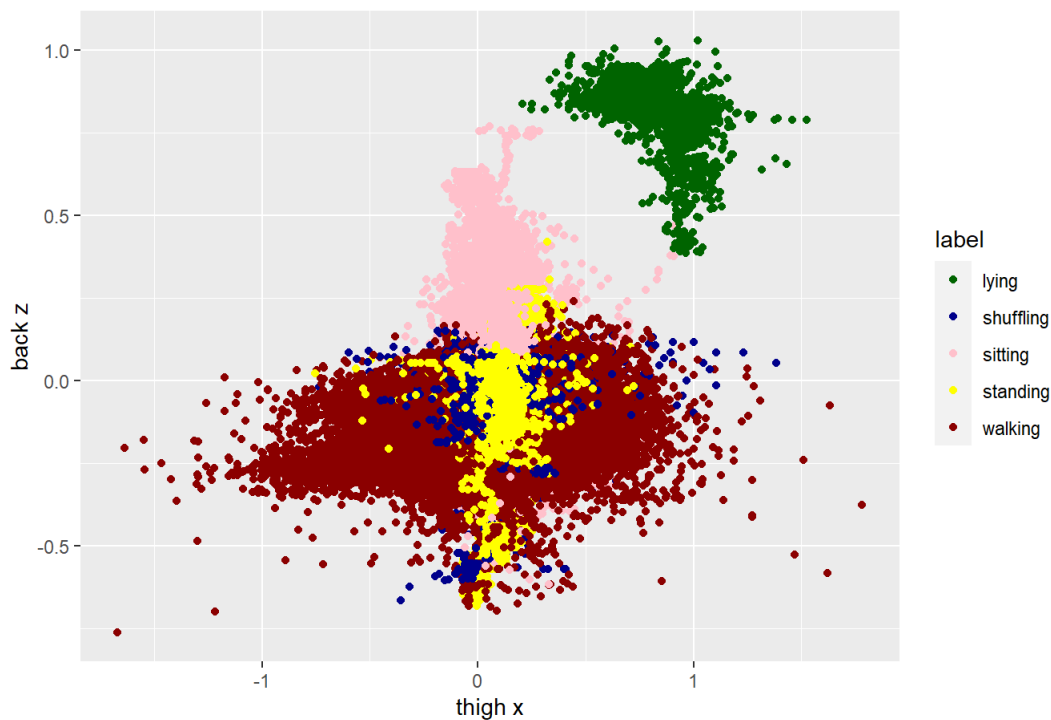
```
# From correlation table (thigh y vs back y) = 0.74
ggplot(d513, aes(x= thigh_y, y = back_y, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of thigh y vs back y group by label",
    x = "thigh x",
    y = "back y",
    color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```


Scatter plot of thigh y vs back y group by label



```
# From correlation table (thigh y vs back z) = 0.70
ggplot(d513, aes(x= thigh_y, y = back_z, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of thigh y vs back z group by label",
       x = "thigh x",
       y = "back z",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of thigh y vs back z group by label



```
# From correlation table (thigh z vs thigh x) = -0.79
ggplot(d513, aes(x= thigh_z, y = thigh_x, color = label)) +
  geom_point() +
  labs(title = "Scatter plot of thigh z vs thigh x group by label",
       x = "thigh z",
       y = "thigh x",
       color = "label") +
  scale_color_manual(values = c("darkgreen", "darkblue", "pink", "yellow", "darkred"))
```

Scatter plot of thigh z vs thigh x group by label

