DSE Project - using csv file

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.4.2
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 4.4.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.2
```

```
##
## Attaching package: 'dplyr'
```

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

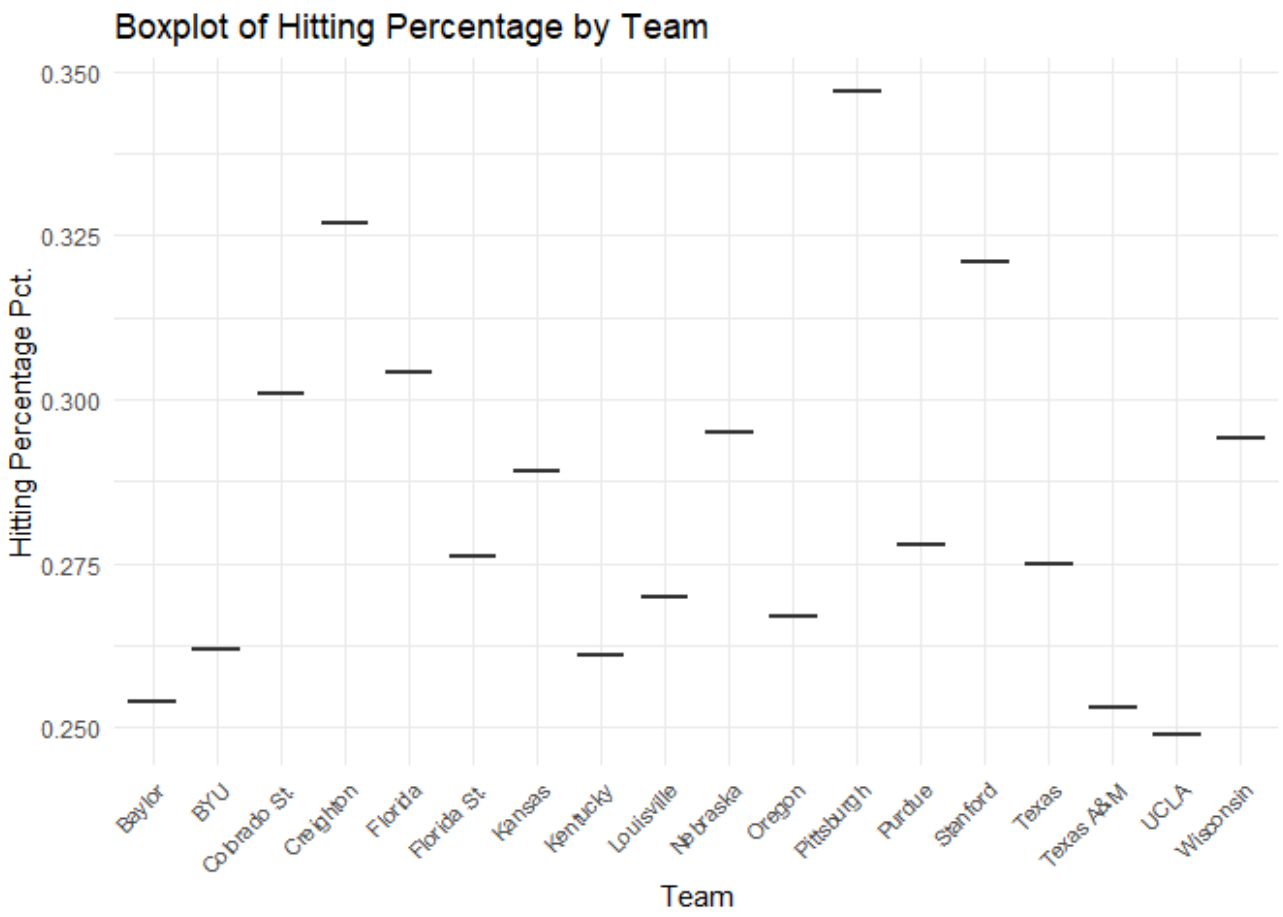
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data <- read_excel("team_stats_combined.xlsx")
```

```
head(data)
```

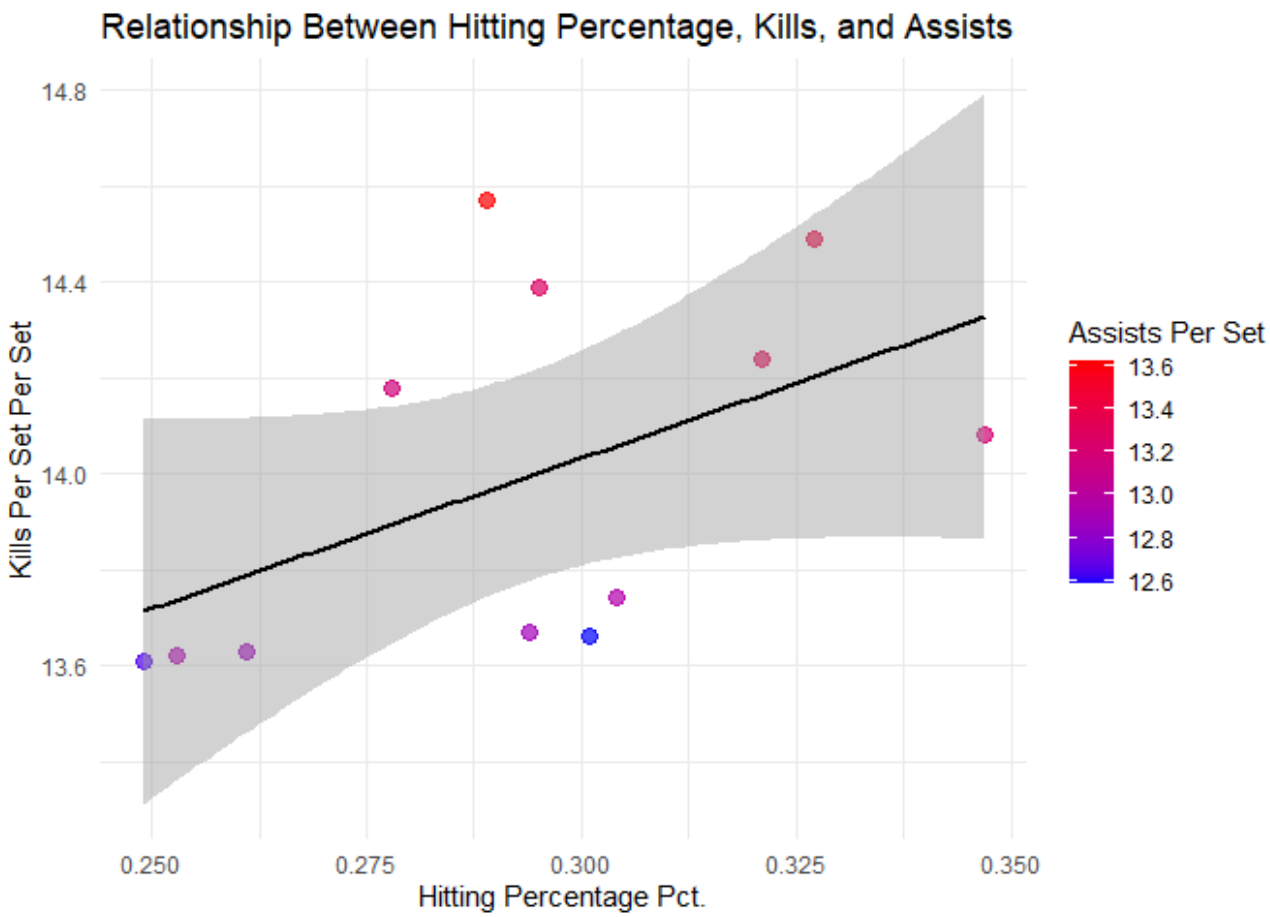
```
## # A tibble: 6 × 34
##   Team              `Aces Per Set Rank` `Aces Per Set S` `Aces Per Set Aces`
##   <chr>             <chr>                         <dbl>                <dbl>
## 1 Florida A&M       5                               111                  234
## 2 Pittsburgh        -                                90                  186
## 3 Florida           11                               96                  187
## 4 Creighton         14                               96                  184
## 5 Kentucky          23                               92                  173
## 6 George Washington 37                              115                  205
## # ℹ 30 more variables: `Aces Per Set Per Set` <dbl>,
## #   `Assists Per Set Rank` <dbl>, `Assists Per Set S` <dbl>,
## #   `Assists Per Set Assists` <dbl>, `Assists Per Set Per Set` <dbl>,
## #   `Blocks Per Set Rank` <chr>, `Blocks Per Set S` <dbl>,
## #   `Blocks Per Set Block Solos` <dbl>, `Blocks Per Set Block Assists` <dbl>,
## #   `Blocks Per Set Per Set` <dbl>, `Digs Per Set Rank` <dbl>,
## #   `Digs Per Set S` <dbl>, `Digs Per Set Digs` <dbl>, …
```

```
# Remove rows with NA (NaN) values in 'Hitting Percentage Pct.' or 'Team' columns
data_clean <- data %>%
  filter(!is.na(`Hitting Percentage Pct.`) & !is.na(Team))

# Create the boxplot
ggplot(data_clean, aes(x = Team, y = `Hitting Percentage Pct.`)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)) + # Adjust angl
  labs(title = "Boxplot of Hitting Percentage by Team",
       x = "Team",
       y = "Hitting Percentage Pct.") +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 8, angle = 45, hjust = 1, vjust = 1))  # Sm
```




```
# Clean data: Remove NA values for 'Hitting Percentage Pct.', 'Kills Per Set Per Set'
data_clean <- data %>%
  filter(!is.na(`Hitting Percentage Pct.`) &
         !is.na(`Kills Per Set Per Set`) &
         !is.na(`Assists Per Set Per Set`))

# Create a scatter plot with color encoding for 'Assists Per Set Per Set'
ggplot(data_clean, aes(x = `Hitting Percentage Pct.`,
                       y = `Kills Per Set Per Set`,
                       color = `Assists Per Set Per Set`)) +
  geom_point(alpha = 0.7, size = 3) + # Scatter plot points
  geom_smooth(method = "lm", color = "black", se = TRUE) + # Overall regression line
  labs(title = "Relationship Between Hitting Percentage, Kills, and Assists",
       x = "Hitting Percentage Pct.",
       y = "Kills Per Set Per Set",
       color = "Assists Per Set") +
  scale_color_gradient(low = "blue", high = "red") + # Color gradient for assists
  theme_minimal()
```

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


Relationship Between Hitting Percentage, Kills, and Assists
Kills Per Set Per Set
14.8
14.4
14.0
13.6
0.250
0.275
0.300
0.325
0.350
Hitting Percentage Pct.
Assists Per Set
13.6
13.4
13.2
13.0
12.8
12.6