

ECS111-Preprocessing

Samarth Sridhara

2025-05-13

```
# Load data  
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
library(tidyr)  
library(dplyr)
```

```
##  
## 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
```

```
library(readr)
```

```
# Step 1: Load the data
```

```
df <- read_csv("/Users/samarthsridhara/Downloads/valorant_players_processedMay12,2025.csv")
```

```
## Rows: 553 Columns: 22
```

```
## -- Column specification -----  
## Delimiter: ","  
## chr (10): puuid, user, tag, deaths_per_game, kills_per_game, assists_per_gam...  
## dbl (12): hs_percent, body_percent, leg_percent, s_damage_per_round, s_kd_ra...  
##  
## i Use 'spec()' to retrieve the full column specification for this data.  
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
# Step 2: Coerce target columns to numeric using dplyr::across()
```

```
cols_to_numeric <- c("hs_percent", "body_percent", "leg_percent", "s_kd_ratio",  
                     "s_win_percent", "s_kast_percent", "s_damage_per_round",  
                     "s_acs", "s_kills_per_round", "deaths_per_game",
```

```

      "kills_per_game", "assists_per_game", "first_bloods_per_game",
      "flawless_rounds_per_game", "aces_per_game")

df <- df %>%
  mutate(across(all_of(cols_to_numeric), as.numeric))

```

```

## Warning: There were 6 warnings in 'mutate()'.
## The first warning was:
## i In argument: 'across(all_of(cols_to_numeric), as.numeric)'.
## Caused by warning:
## ! NAs introduced by coercion
## i Run 'dplyr::last_dplyr_warnings()' to see the 5 remaining warnings.

```

```

# Step 3: Pivot longer for faceting
df_long <- pivot_longer(df, cols = all_of(cols_to_numeric),
  names_to = "statistic", values_to = "value")

# Step 4: Plot with facet_wrap
ggplot(df_long, aes(x = smurf_label, y = value, fill = smurf_label)) +
  geom_boxplot() +
  facet_wrap(~ statistic, scales = "free_y") +
  theme_minimal() +
  labs(title = "Distribution of Player Statistics by Smurf Category",
    x = "Smurf Label", y = "Value") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

```

## Warning: Removed 30 rows containing non-finite outside the scale range
## ('stat_boxplot()').

```

Figure 1 displays 12 box plots comparing various game statistics across three player types: most likely smurf (red), normal player (green), and suspicious (blue). The x-axis represents the Smurf Label, and the y-axis represents the Value. The statistics are arranged in a 3x4 grid. The legend indicates that red represents 'most likely smurf', green represents 'normal player', and blue represents 'suspicious'.

| Statistic | most likely smurf | normal player | suspicious |
|-------------------|-------------------|---------------|------------|
| ices_per_game | ~0.08 | ~0.02 | ~0.04 |
| sists_per_game | ~5.5 | ~5.5 | ~5.5 |
| body_percent | ~65 | ~65 | ~65 |
| maths_per_game | ~14.5 | ~15.5 | ~15.5 |
| bloods_per_game | ~4.5 | ~2.5 | ~3.5 |
| s_rounds_per_game | ~0.8 | ~0.7 | ~0.7 |
| hs_percent | ~50 | ~50 | ~50 |
| kills_per_game | ~35 | ~25 | ~30 |
| leg_percent | ~5.5 | ~6.5 | ~5.5 |
| s_acs | ~250 | ~200 | ~210 |
| amage_per_ro | ~170 | ~140 | ~150 |
| s_kast_percent | ~75 | ~65 | ~70 |
| s_kd_ratio | ~1.2 | ~1.0 | ~1.1 |
| kills_per_round | ~0.9 | ~0.6 | ~0.8 |
| s_win_percent | ~60 | ~50 | ~60 |

```
## Rows: 553 Columns: 22
## -- Column specification -----
## Delimiter: ","
## chr (10): puuid, user, tag, deaths_per_game, kills_per_game, assists_per_gam...
## dbl (12): hs_percent, body_percent, leg_percent, s_damage_per_round, s_kd_ra...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
df <- df %>%
```

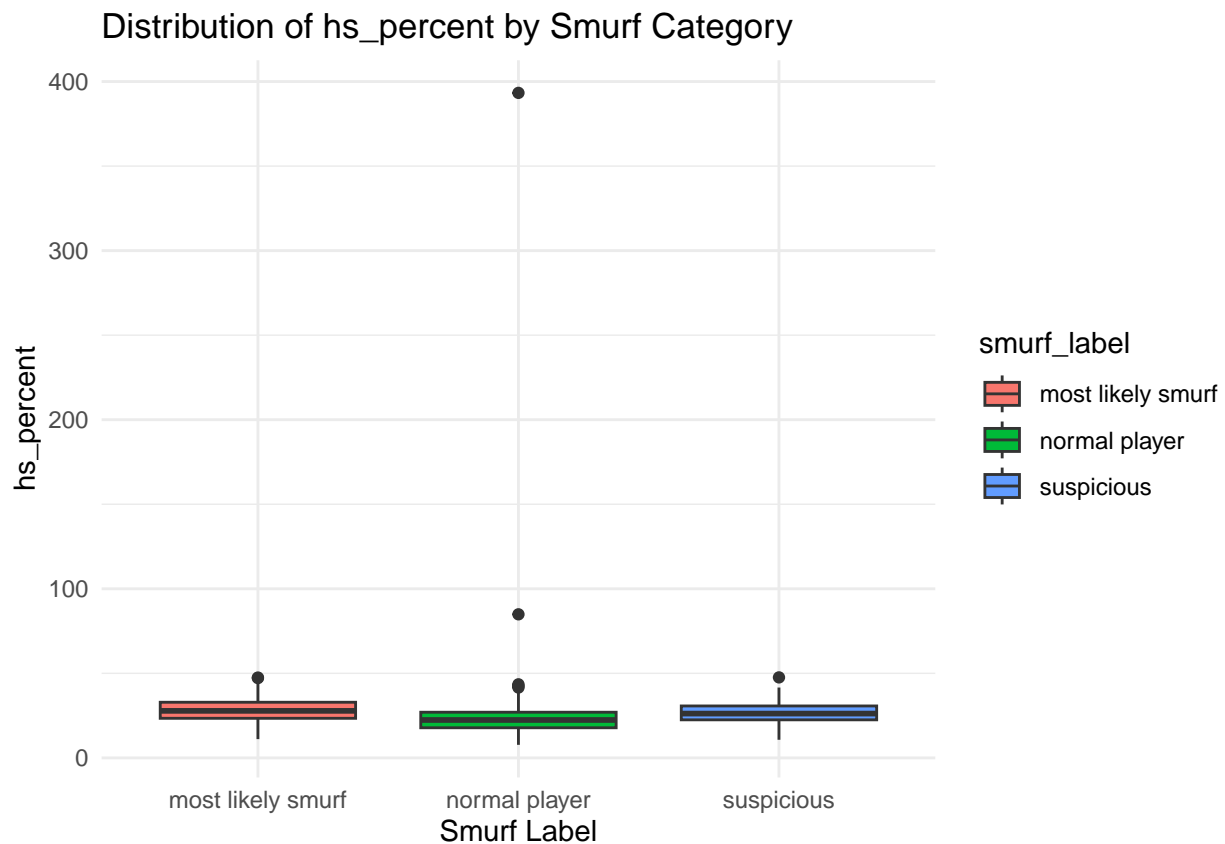
```
mutate(across(all_of(cols_to_numeric), as.numeric)) %>%
pivot_longer(cols = all_of(cols_to_numeric), names_to = "statistic", values_to = "value")
```

```
## Warning: There were 6 warnings in 'mutate()'.
## The first warning was:
## i In argument: 'across(all_of(cols_to_numeric), as.numeric)'.
## Caused by warning:
## ! NAs introduced by coercion
## i Run 'dplyr::last_dplyr_warnings()' to see the 5 remaining warnings.
```

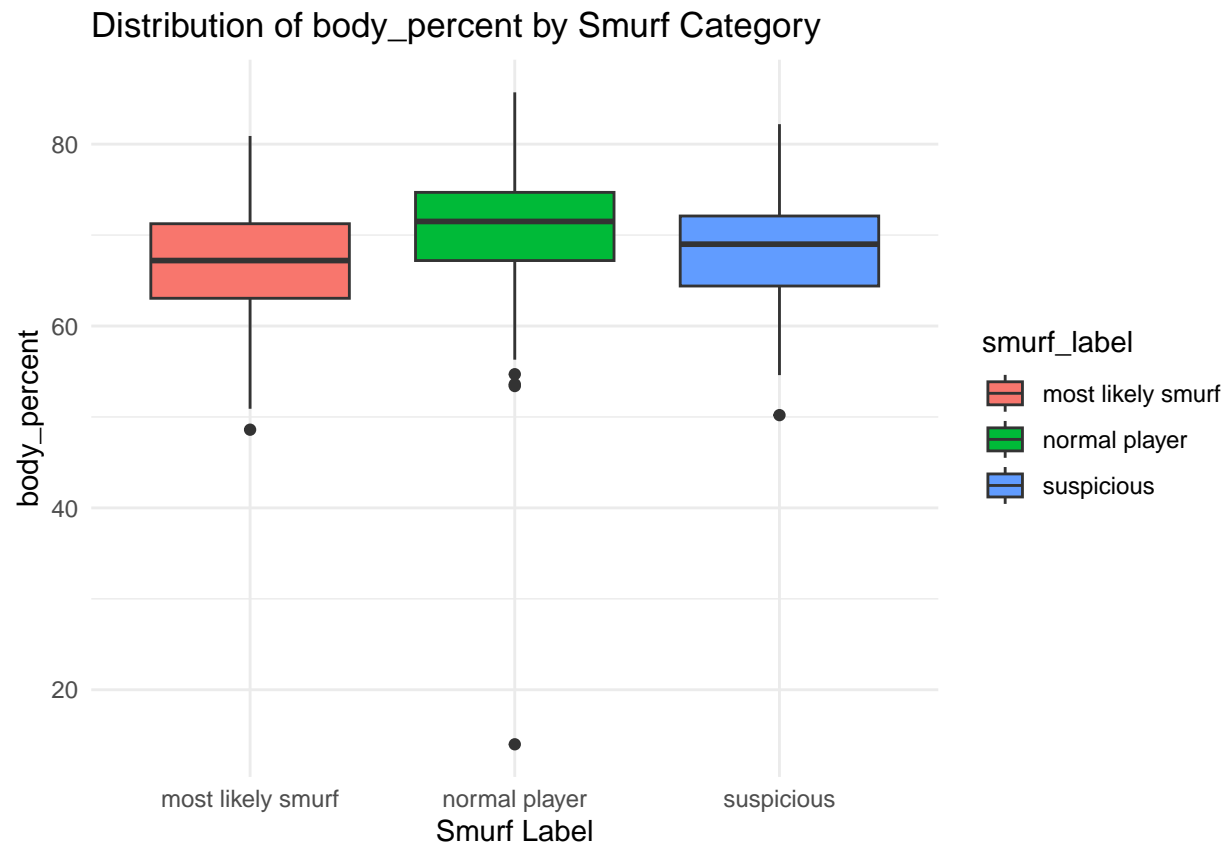
```
# Create and print each plot separately
```

```
unique_stats <- unique(df$statistic)
```

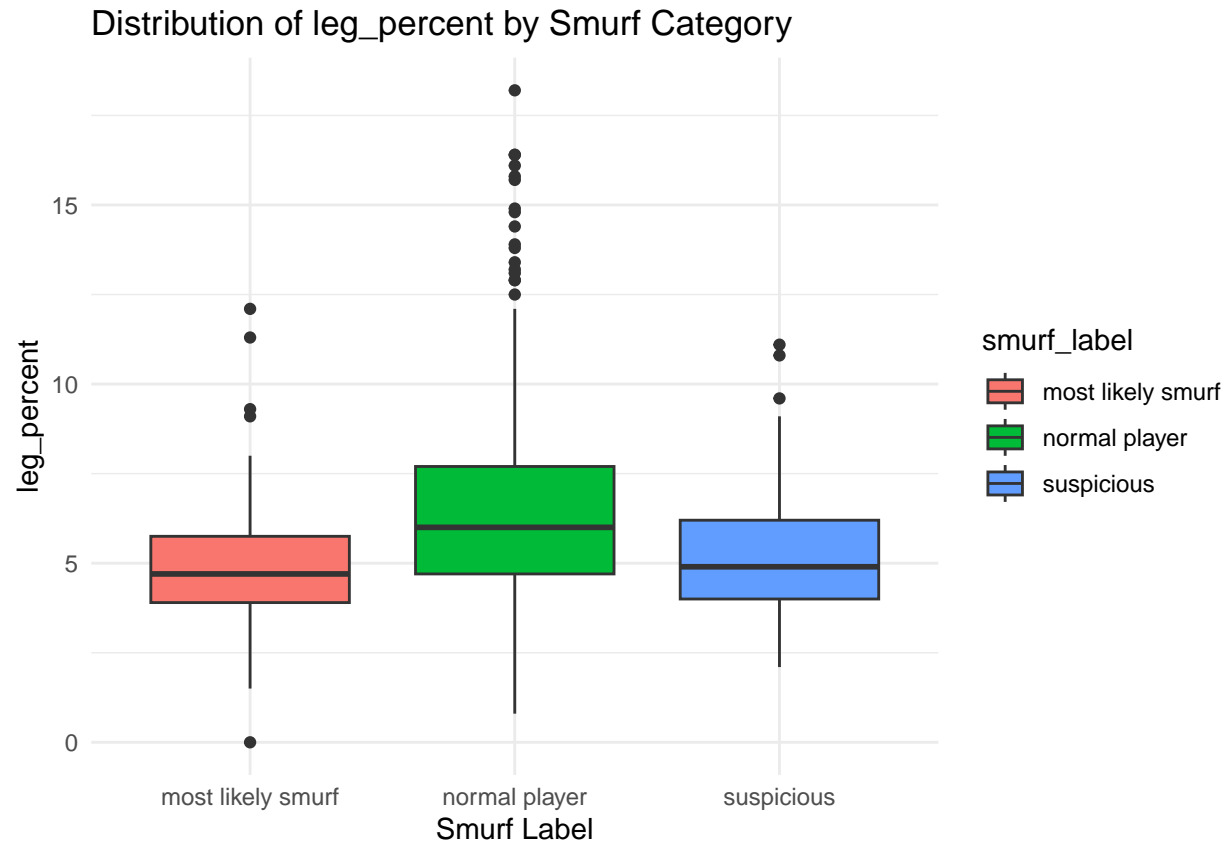
```
for (stat in unique_stats) {
  p <- ggplot(filter(df, statistic == stat), aes(x = smurf_label, y = value, fill = smurf_label)) +
    geom_boxplot() +
    theme_minimal() +
    labs(title = paste("Distribution of", stat, "by Smurf Category"),
         x = "Smurf Label", y = stat)
  print(p)
  readline(prompt = "Press [enter] to continue to next plot")
}
```



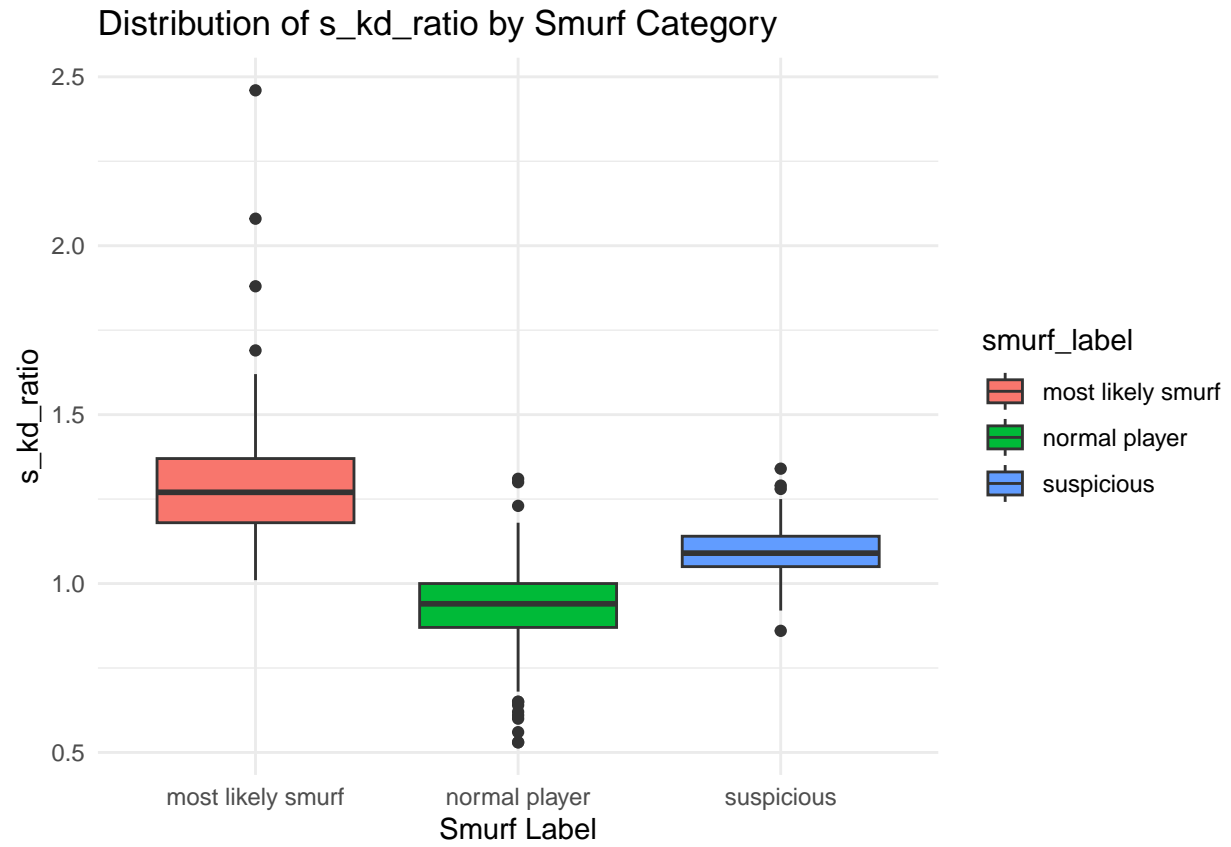
```
## Press [enter] to continue to next plot
```



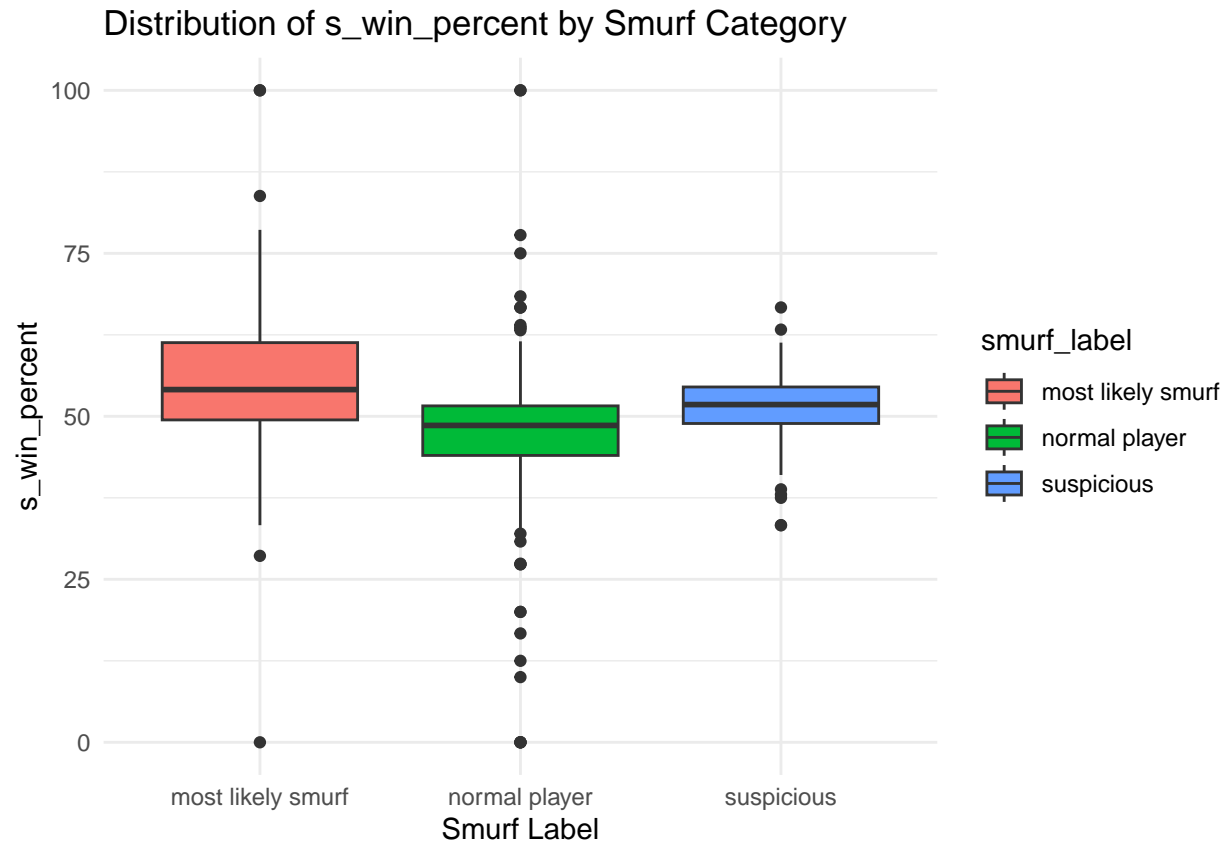
Press [enter] to continue to next plot



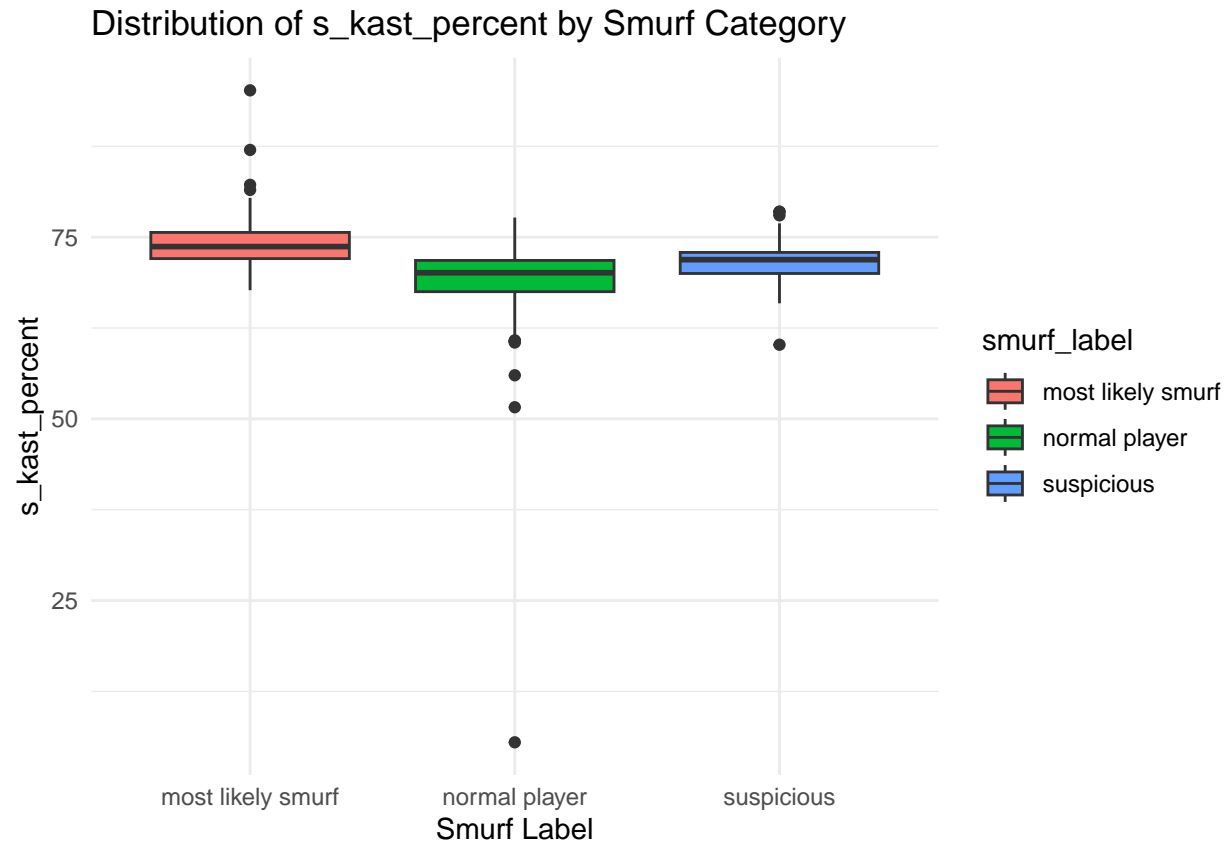
Press [enter] to continue to next plot



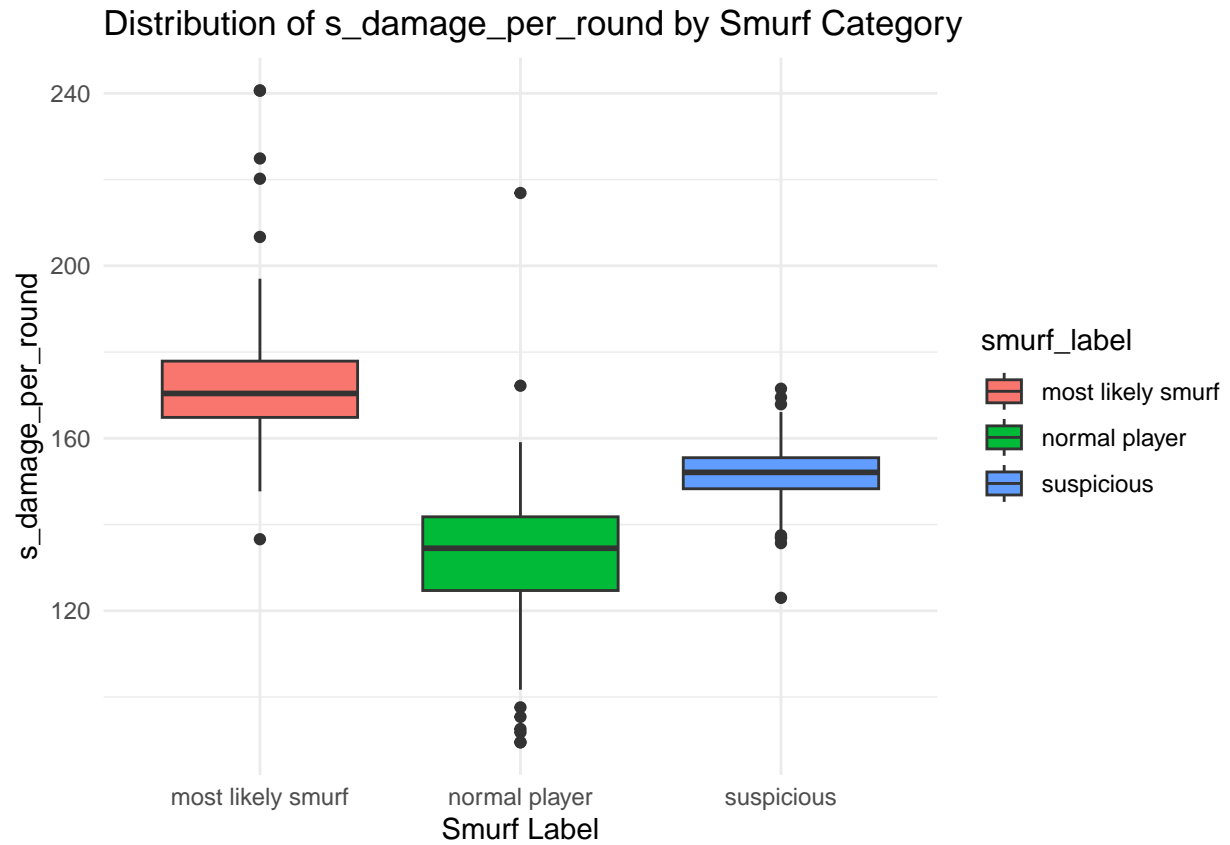
Press [enter] to continue to next plot



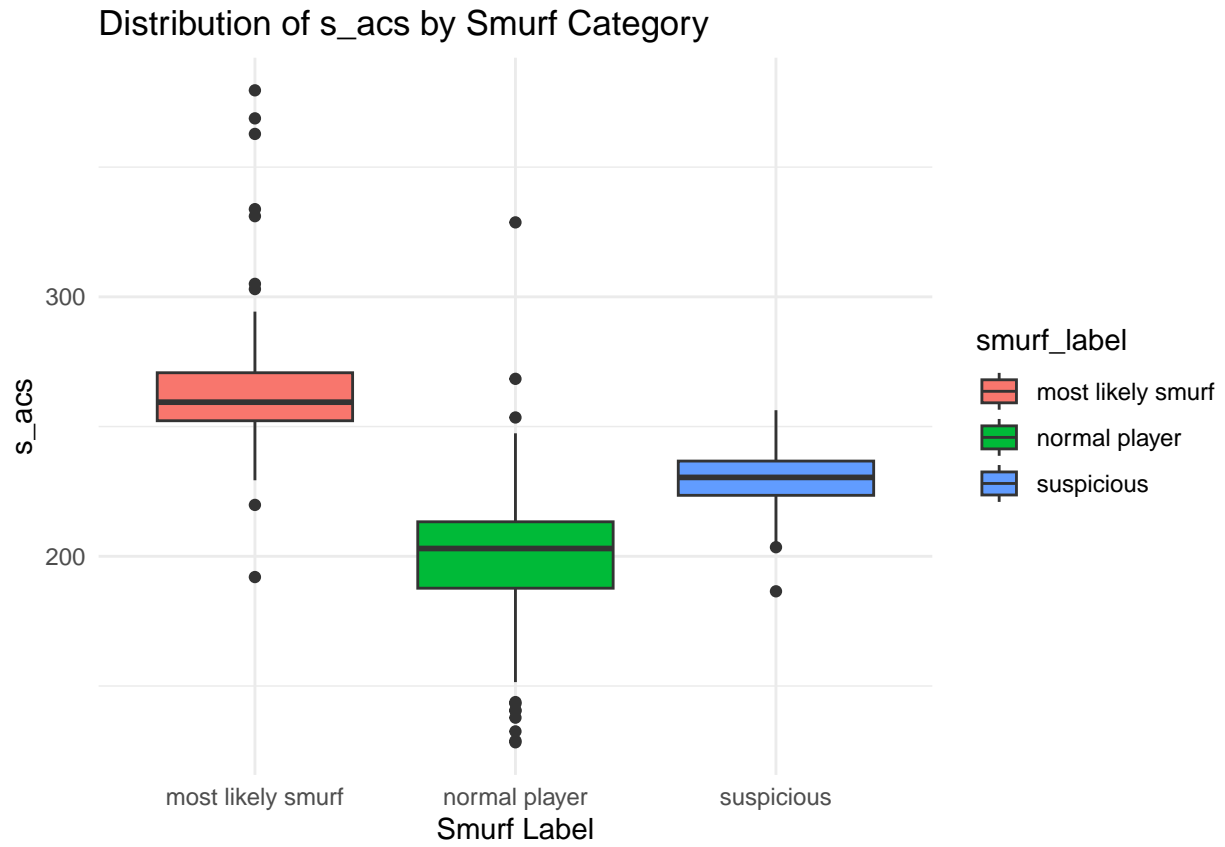
Press [enter] to continue to next plot



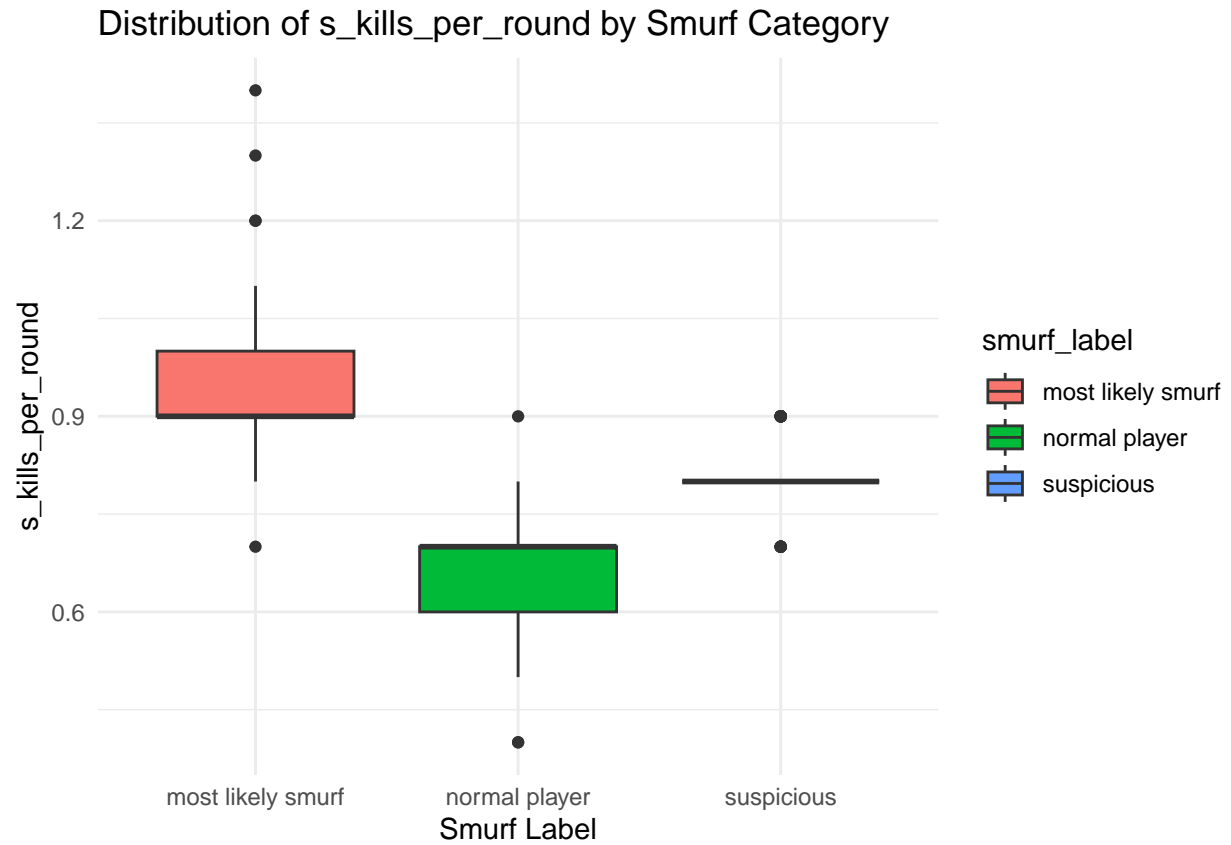
Press [enter] to continue to next plot



Press [enter] to continue to next plot



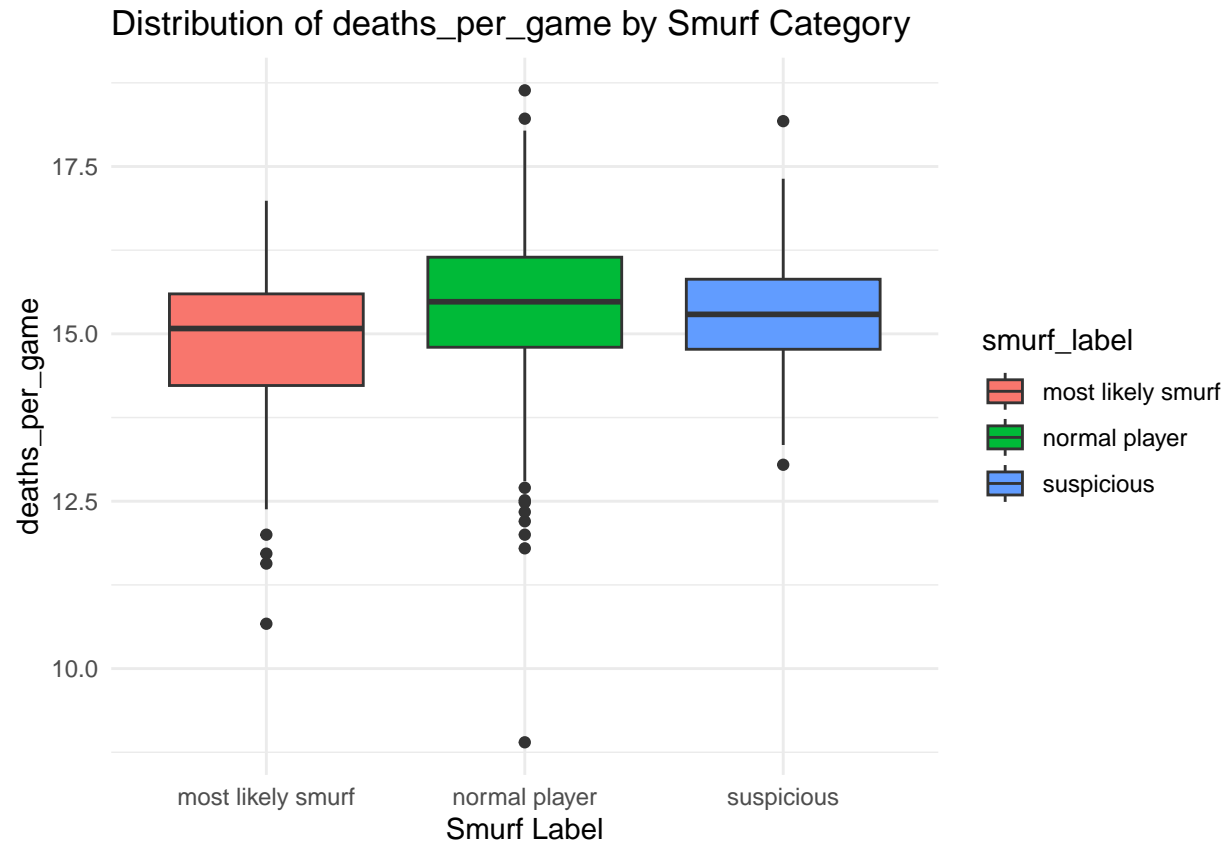
Press [enter] to continue to next plot



```
## Press [enter] to continue to next plot
```

```
## Warning: Removed 5 rows containing non-finite outside the scale range
```

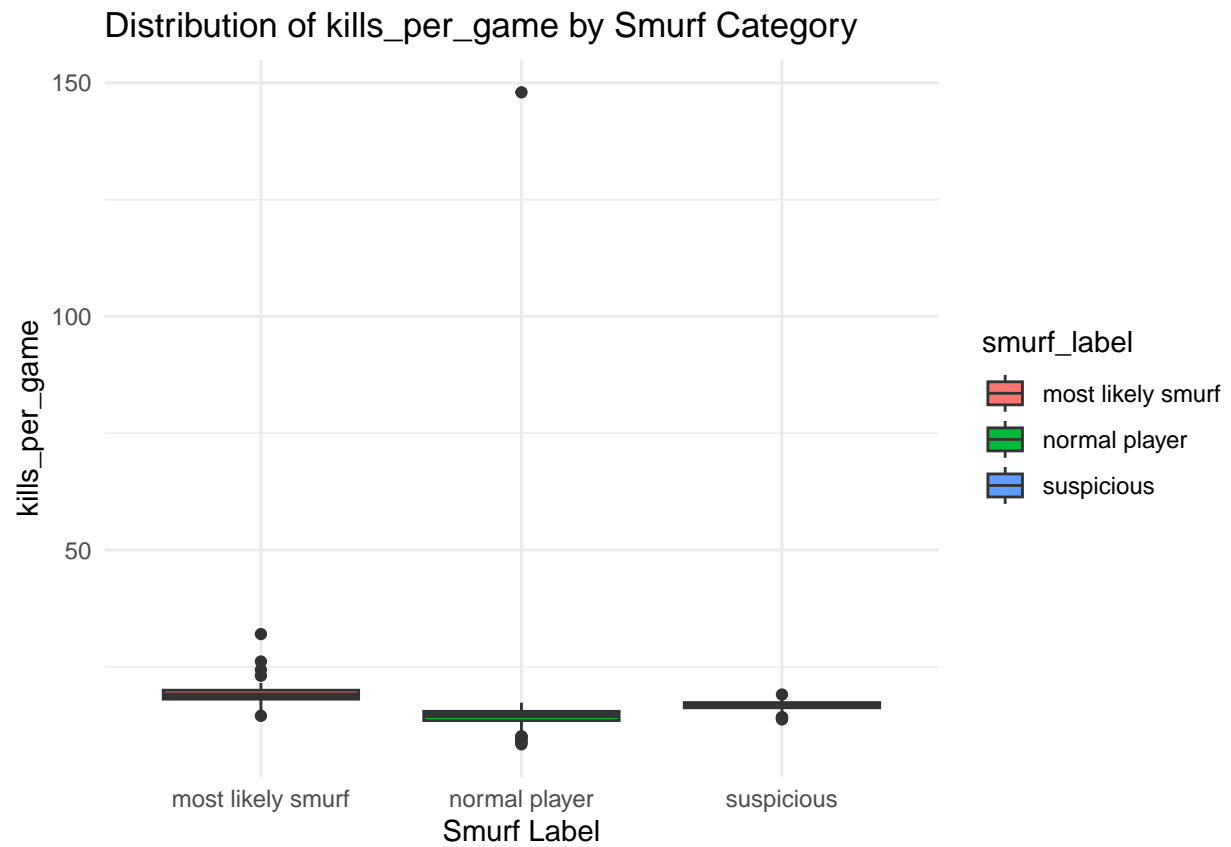
```
## ('stat_boxplot()').
```



```
## Press [enter] to continue to next plot
```

```
## Warning: Removed 5 rows containing non-finite outside the scale range
```

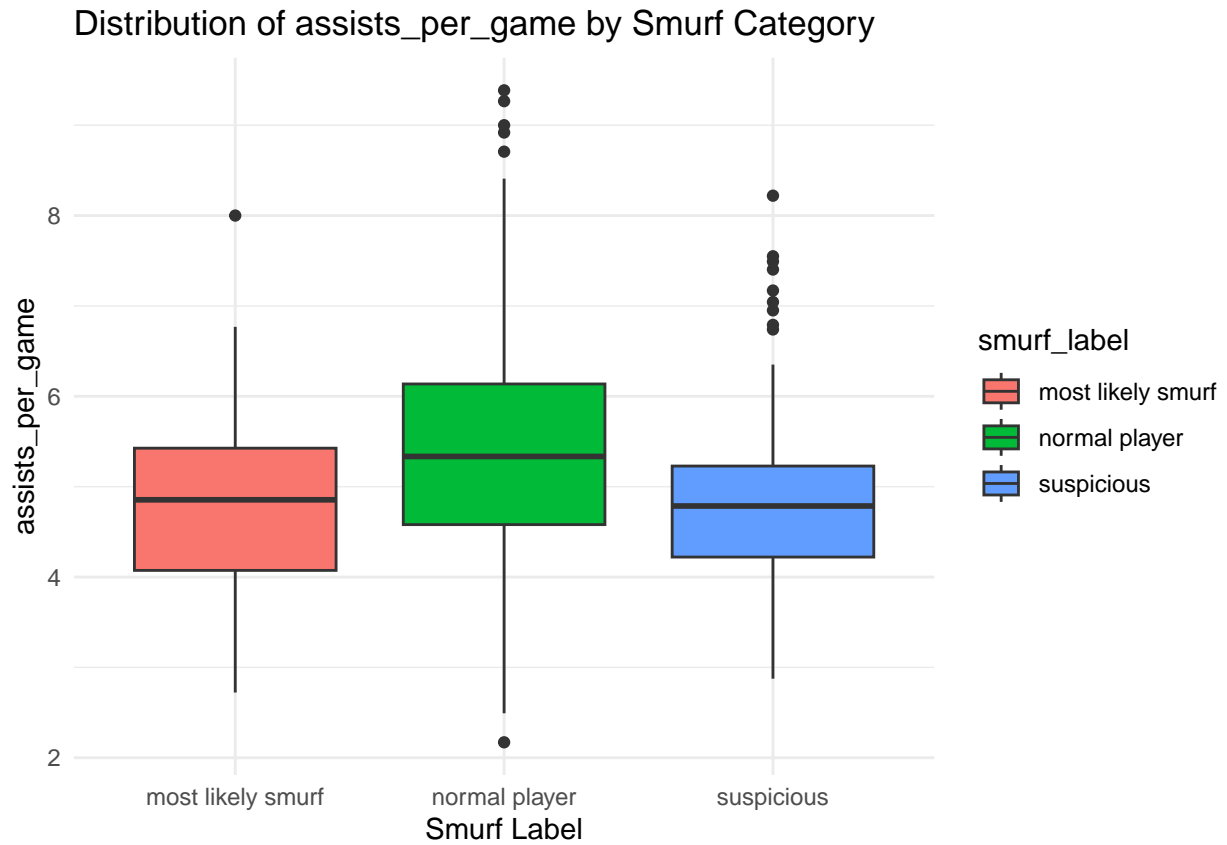
```
## ('stat_boxplot()').
```



```
## Press [enter] to continue to next plot
```

```
## Warning: Removed 5 rows containing non-finite outside the scale range
```

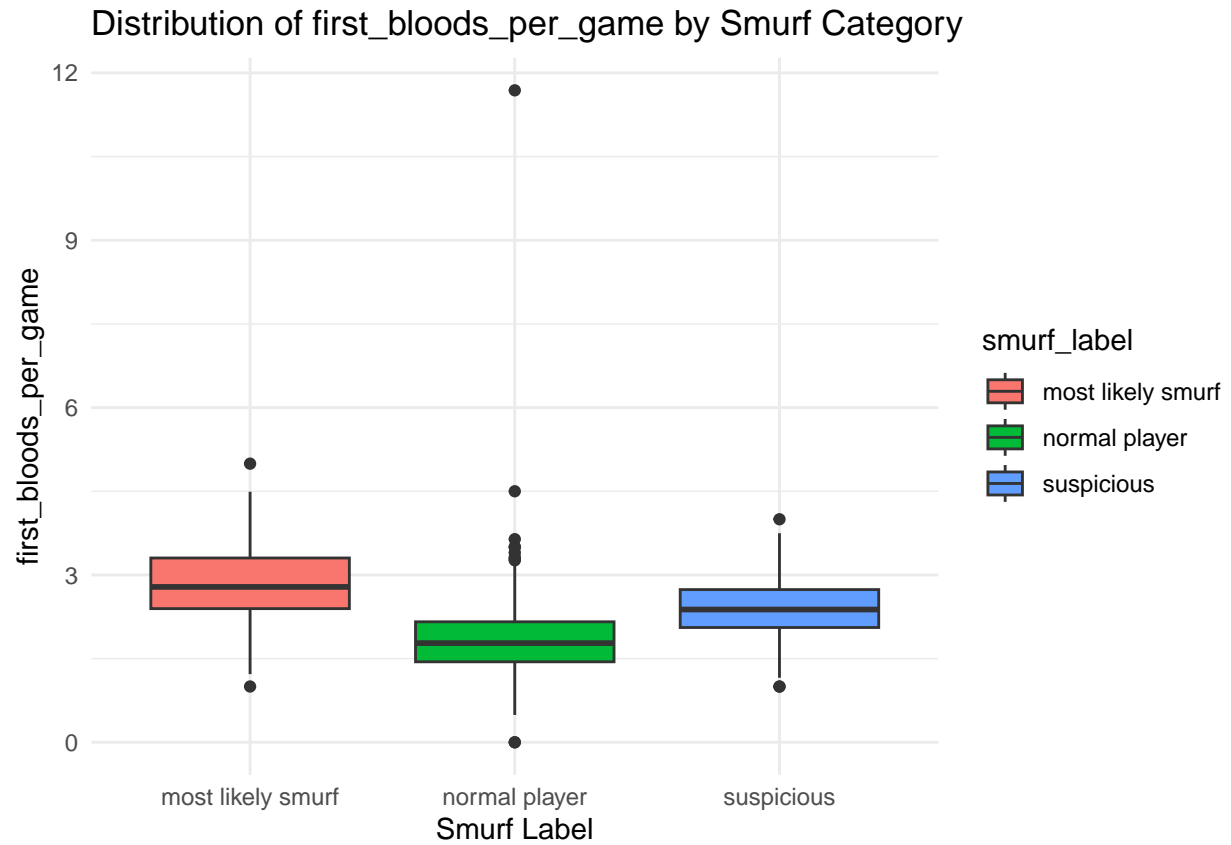
```
## ('stat_boxplot()').
```



```
## Press [enter] to continue to next plot
```

```
## Warning: Removed 5 rows containing non-finite outside the scale range
```

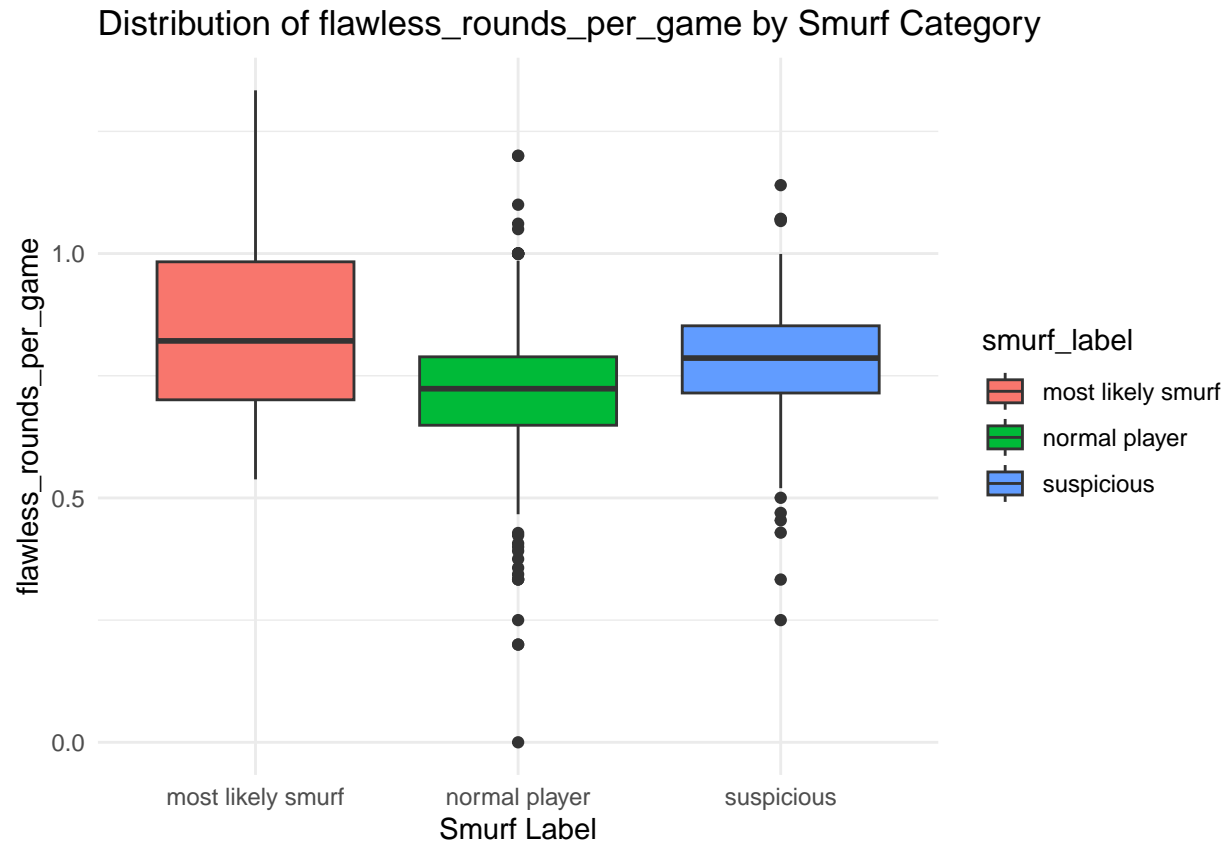
```
## ('stat_boxplot()').
```



```
## Press [enter] to continue to next plot
```

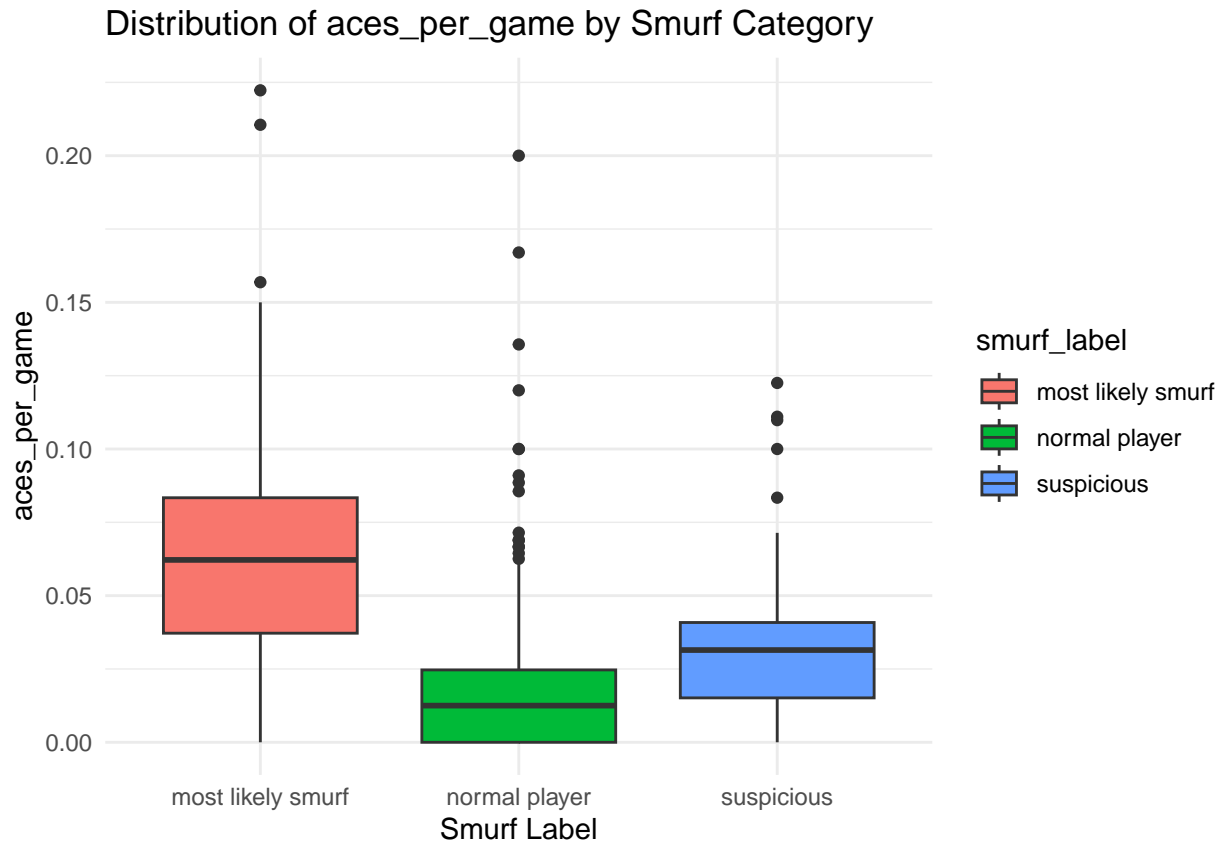
```
## Warning: Removed 5 rows containing non-finite outside the scale range
```

```
## ('stat_boxplot()').
```

```
## Press [enter] to continue to next plot
```

```
## Warning: Removed 5 rows containing non-finite outside the scale range  
## ('stat_boxplot()').
```



Press [enter] to continue to next plot