project preprocessing

Takshsheel Goswami

2024-04-23

Including Plots

You can also embed plots, for example:

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.4
## v forcats   1.0.0      v stringr   1.5.0
## v ggplot2   3.5.1      v tibble    3.2.1
## v lubridate 1.9.2      v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
salary_data = read.csv("salary_Proj.csv")
nflverse_data = read.csv("nflverse.csv")

names(salary_data)[1] <- "player_display_name"
total <- merge(salary_data,nflverse_data,by="player_display_name")
```

```r
total = subset(total, select = -c(player_name,player_id,X,Position,season_type,headshot_url,recent_team

library(dplyr)
total <- total %>%
  group_by(player_display_name,position,position_group) %>%
  summarise_all(mean)
```
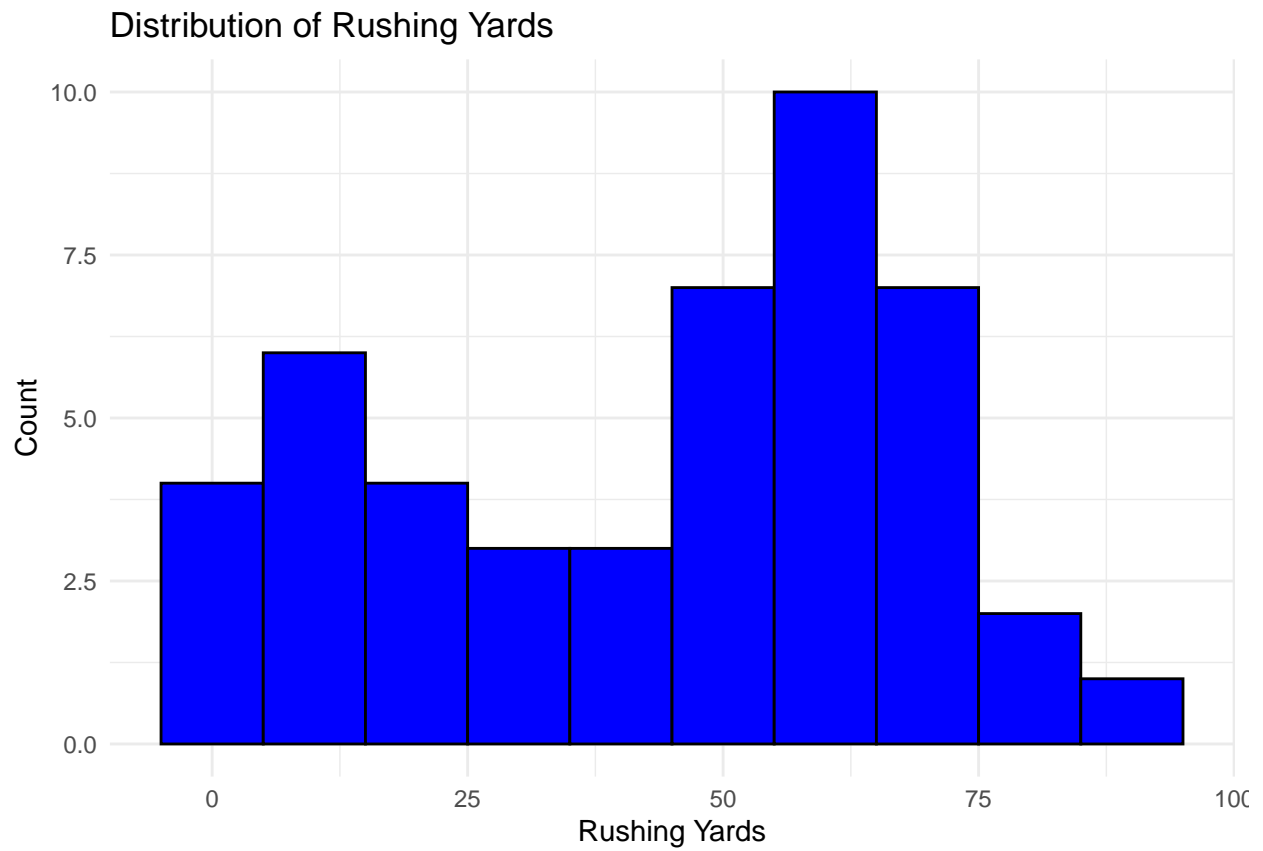
```r
data = read.csv("preprocessed_combined_mean.csv")
```
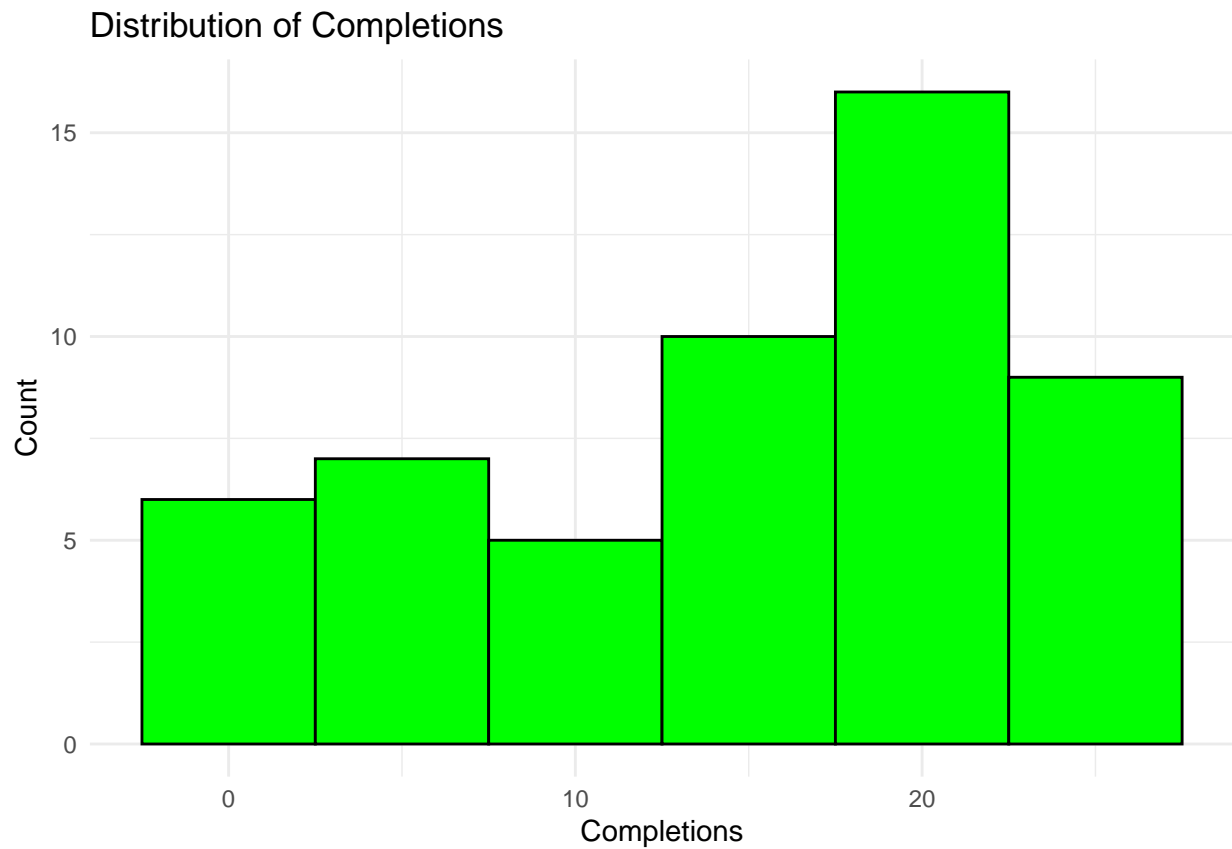
```r
library(ggplot2)

data_rb <- data[data$position == "RB", ]
data_qb <- data[data$position == "QB", ]

ggplot(data_rb, aes(x=rushing_yards)) +
  geom_histogram(binwidth=10, fill="blue", color="black") +
  labs(title="Distribution of Rushing Yards",
       x="Rushing Yards",
       y="Count") +
  theme_minimal()
```

Distribution of Rushing Yards


```r
ggplot(data_qb, aes(x=completions)) +
  geom_histogram(binwidth=5, fill="green", color="black") +
  labs(title="Distribution of Completions",
       x="Completions",
       y="Count") +
  theme_minimal()
```

Distribution of Completions


```r
ggplot(data_qb, aes(x=passing_yards)) +
  geom_histogram(binwidth=15, fill="red", color="black") +
  labs(title="Distribution of Passing Yards",
       x="Passing Yards",
       y="Count") +
  theme_minimal()
```

Distribution of Passing Yards