

Offense vs Defense: An Analysis in the NBA's Play Style

Team 22

2025-11-07

Abstract

Dunk or get dunked on — that's the name of the game when it comes to the NBA. You have twenty-four seconds to sink a shot or get one put down on your hoop, and currently the rules have seemed to strongly favor offenses and penalize physical defenses. Does that mean defense is less useful now, or does defense still play just as crucial a part and simply look different from what we expect? This paper seeks to investigate what efficient defenses and offenses look like by identifying key metrics that aggregate to describe efficient play, and ultimately analyze what wins games more, scoring points or allowing points. To conduct this analysis, we will develop models describing offensive and defensive efficiency and apply nested F-tests across the models and an aggregate to identify which contributes more to winning games. (Insert findings later)

Introduction

The NBA's rulebook since 2004 and ongoing regularly makes updates that drive the game to at least seem both extremely offense-driven and defense-second, with the latter seemingly being a mere afterthought if not outright not integrated into defensive philosophies. This perception of the game may have a strong sense of truth, but there is a very viable possibility that it only describes half the game and leaves sorely underdeveloped defensive outlook and strategy. To bring the best out of players' technical abilities, strengthen and develop the NBA-seeking talent pool with both the players and aspiring coaches, and help the NBA tangibly understand how to fight now-rampant criticisms of the NBA 'going soft', it is crucial to better understand the current defensive structure of the game and how to best develop it. By analyzing which key metrics contribute to defensive or offensive efficiency and developing models to understand how much each throughput affects winning outcomes, we can create this better understanding of the flipside to the offensive-minded game style. In order to best conduct this analysis, we will break the analysis into two phases.

For the first phase of analysis, we will first categorize the variables into defensive and offensive categories. Once we class the variables, we will test for intercategory collinearity and develop a full model that selects one variable per collinear relationship to avoid redundancy. Once the full model is developed, we will employ techniques like backwards elimination to develop statistically significant offensive and defensive models that explain a team's win within a game accurately.

For the second phase of analysis, we will then create an aggregate model from the two initial models that is computed as a composite. Using all three models the, we will conduct nested F-tests to determine whether the offensive model or defensive model provides the stronger signal within the composite.

Data

In order to conduct this analysis on NBA gameplay statistics, we are using regular season data (original data) compiled over a time range from 2010-2024 (Korolyk, 2024). This dataset is compiled by Vasili Korolyk and is publicly available at <https://github.com/NocturneBear/NBA-Data-2010-2024> for academic use under its MIT License. The original data contains over 33,000 observations and documents 57 variables

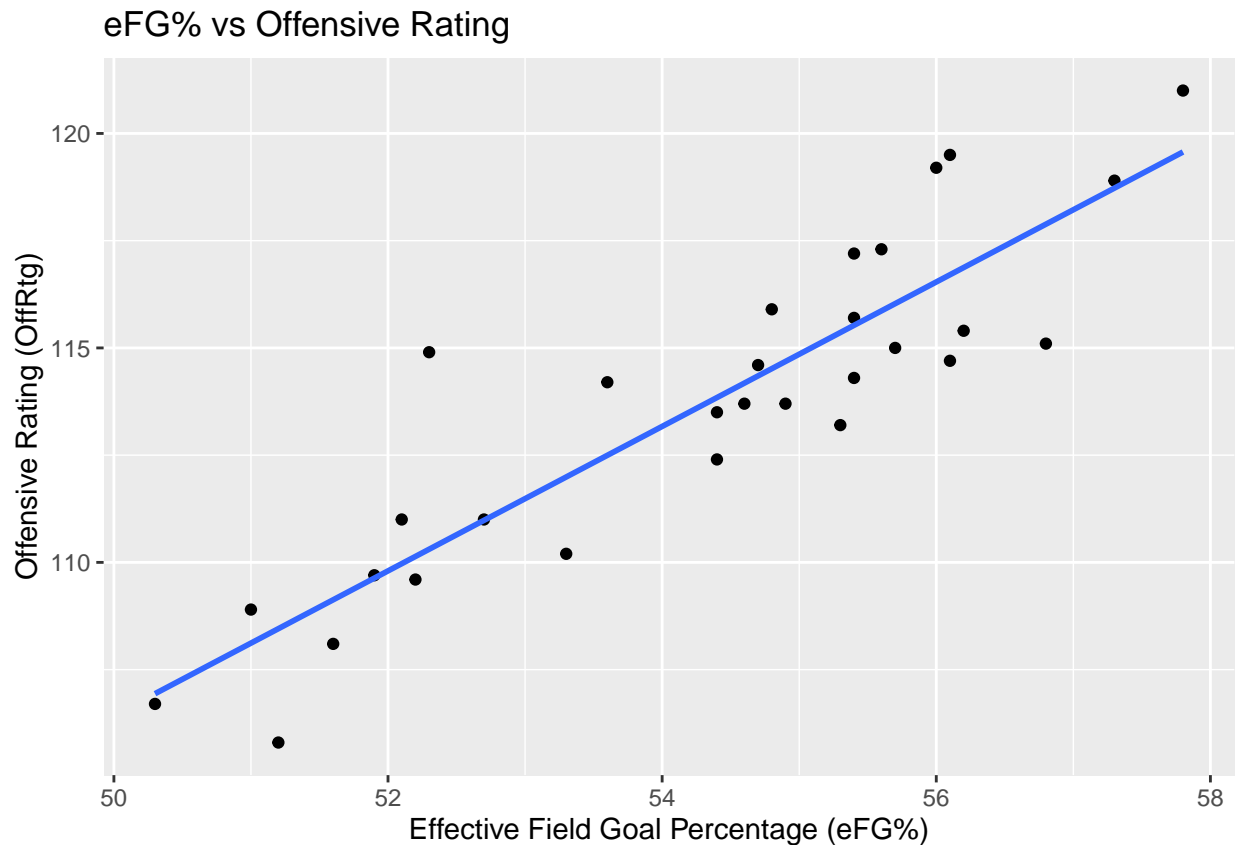
over each entry. Certain variables from the raw data are not useful for our research scope and analysis, so we initially pared down the variable size down to only the relevant ones required for both clean nomenclature and analysis. One of the key utility variables was the 'AVAILABLE_FLAG', which indicated whether the data for that row was usable. As a result, we utilized it to remove any potential entries with empty data points, only removing it after running the previous job.

EDA & Data Visualization

Offensive Efficiency

```
# eFG% vs Offensive Rating
ggplot(df, aes(x = `eFG%`, y = OffRtg)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "eFG% vs Offensive Rating",
       x = "Effective Field Goal Percentage (eFG%)",
       y = "Offensive Rating (OffRtg)")
```

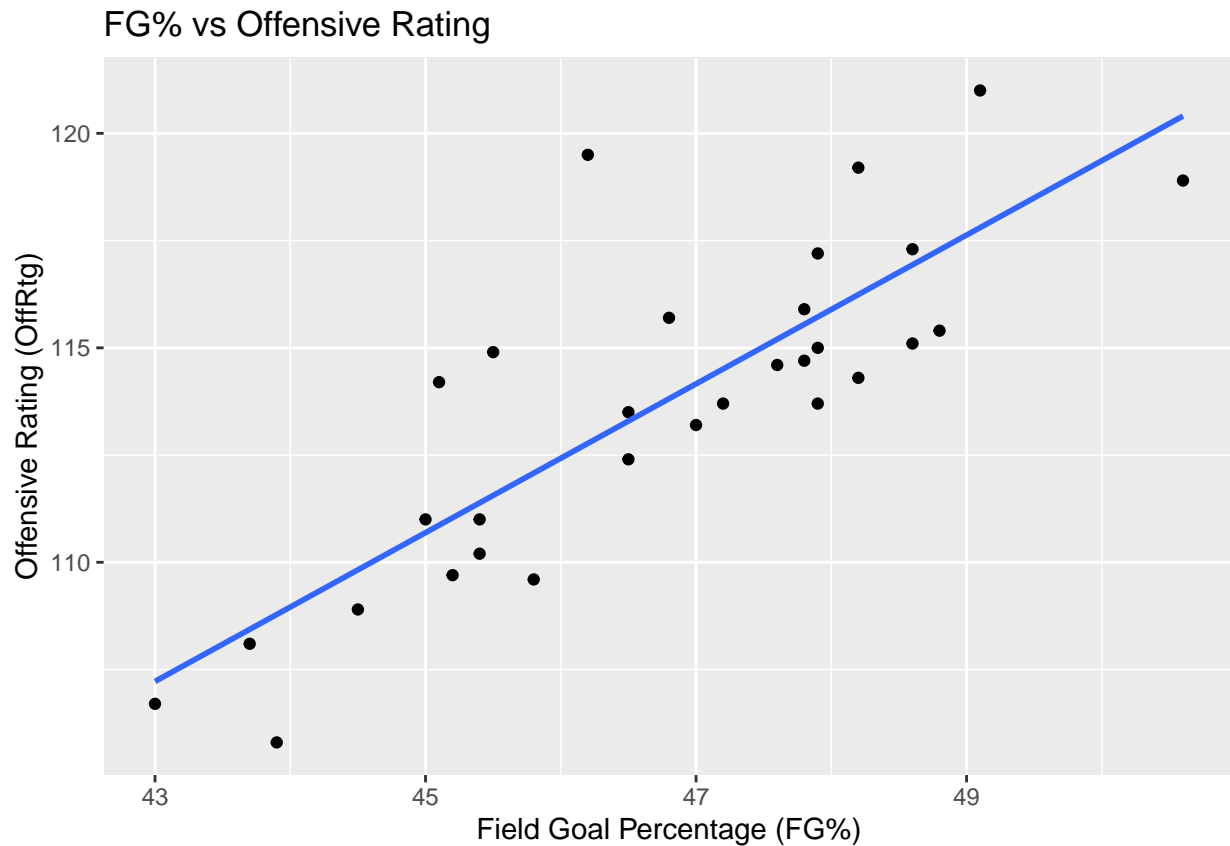
```
## 'geom_smooth()' using formula = 'y ~ x'
```



```
# FG% vs Offensive Rating
ggplot(df, aes(x = `FG%`, y = OffRtg)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
```

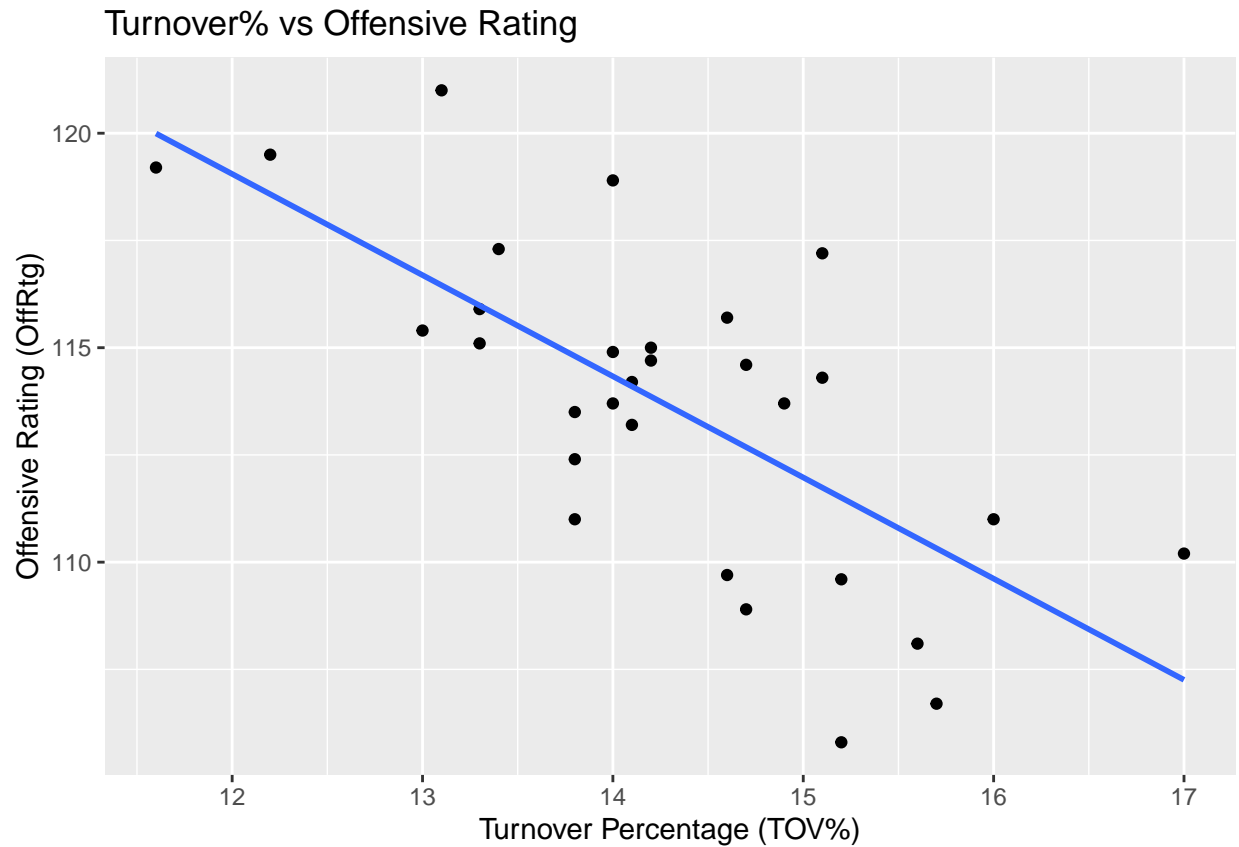
```
labs(title = "FG% vs Offensive Rating",
     x = "Field Goal Percentage (FG%)",
     y = "Offensive Rating (OffRtg)")
```

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



```
# Turnover% vs Offensive Rating
ggplot(df, aes(x = `TOV%`, y = OffRtg)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Turnover% vs Offensive Rating",
       x = "Turnover Percentage (TOV%)",
       y = "Offensive Rating (OffRtg)")
```

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

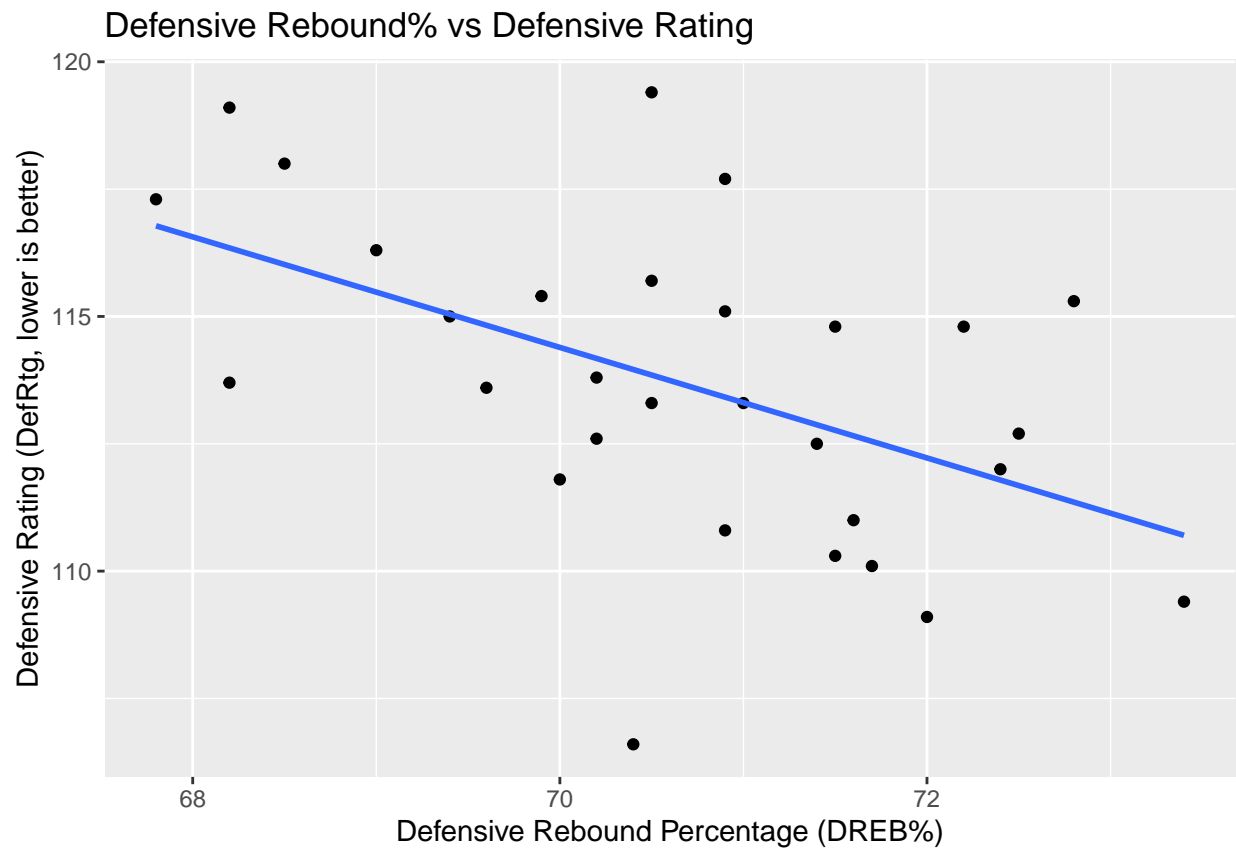


Observation: Teams with better shooting efficiency (higher eFG% and FG%) generally have higher offensive efficiency. Teams that commit fewer turnovers also tend to score more effectively.

Defensive Efficiency

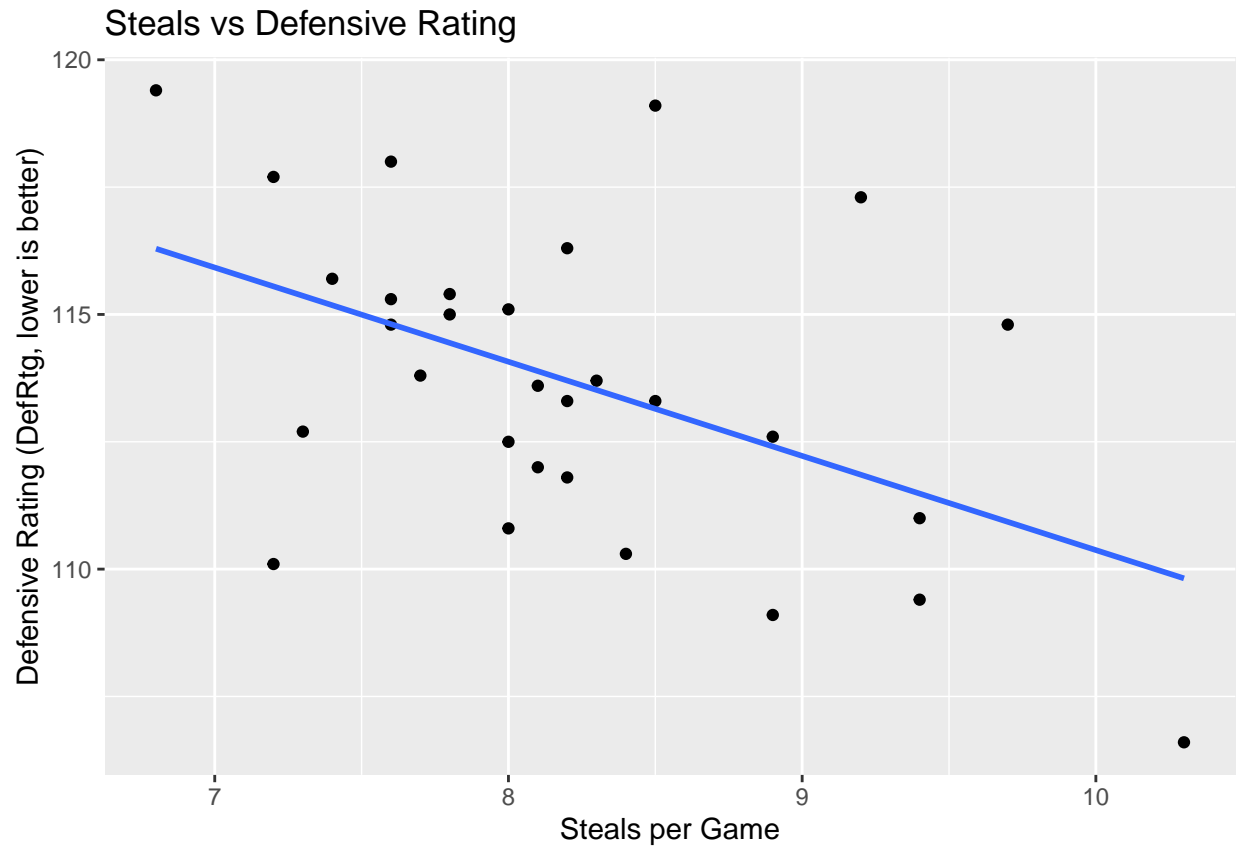
```
# Defensive Rebound% vs Defensive Rating
ggplot(df, aes(x = `DREB%`, y = DefRtg)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Defensive Rebound% vs Defensive Rating",
       x = "Defensive Rebound Percentage (DREB%)",
       y = "Defensive Rating (DefRtg, lower is better)")
```

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



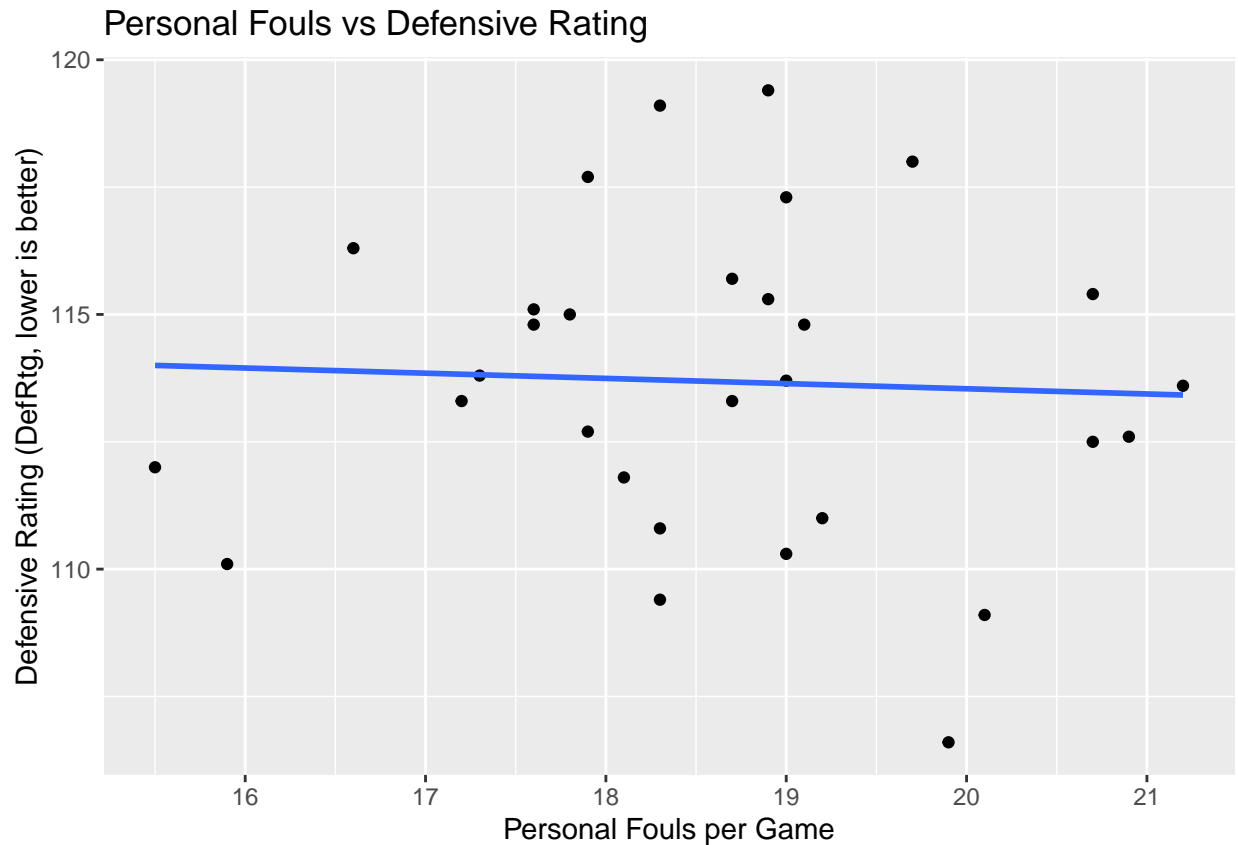
```
# Steals vs Defensive Rating
ggplot(df, aes(x = STL, y = DefRtg)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Steals vs Defensive Rating",
       x = "Steals per Game",
       y = "Defensive Rating (DefRtg, lower is better)")
```

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



```
# Personal Fouls vs Defensive Rating
ggplot(df, aes(x = PF, y = DefRtg)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Personal Fouls vs Defensive Rating",
       x = "Personal Fouls per Game",
       y = "Defensive Rating (DefRtg, lower is better)")
```

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



Observation: Teams that rebound well on defense and generate steals generally allow fewer points. Personal fouls do not show a clear direct relationship with defensive efficiency.

Plan for Further Analysis

In the next stage of the project, we will study how offensive and defensive efficiency relate to winning more games. Specifically, we will:

Identify whether offense or defense contributes more to win percentage. Basic data is not a good measure of the importance of offense and defense. More analysis and calculations are needed.

Create a four-quadrant “Offense vs Defense” chart to compare team play styles.

Focus on the top 10 teams in the league to determine common efficiency patterns.

This will allow us to better answer the question: Does offense or defense win more games in the NBA?

```
df_top10 %>%
  select(Team, `WIN%`, OffRtg, off_rank, DefRtg, def_rank)
```

```
## # A tibble: 10 x 6
##   Team                'WIN%' OffRtg off_rank DefRtg def_rank
##   <fct>              <dbl>  <dbl>   <int>  <dbl>   <int>
## 1 Cleveland Cavaliers 0.78    121     1    112.     8
## 2 Boston Celtics      0.744   120.     2    110.     4
## 3 Oklahoma City Thunder 0.829   119.     3    107.     1
## 4 Denver Nuggets      0.61    119.     4    115.    19
## 5 New York Knicks     0.622   117.     5    113.    13
## 6 Minnesota Timberwolves 0.598   116.     8    111.     6
```

## 7 Indiana Pacers	0.61	115.	9	113.	13
## 8 Los Angeles Lakers	0.61	115	11	114.	16
## 9 Houston Rockets	0.634	115.	12	110.	5
## 10 LA Clippers	0.61	114.	15	109.	3