FP_Rymer/Ben

Rymer Popovich, Ben Tartaglia

Stat 184 Final Project

Import Necessary Packages:

```
# We will be using Google's R style during this project
#| echo: TRUE
#| include: FALSE

library(tidyverse)
```

```
Warning: package 'tidyverse' was built under R version 4.4.2
```

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

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

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
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 becor
```

```
library(janitor)
```

```
Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```
library(ggrepel)
```

```
Warning: package 'ggrepel' was built under R version 4.4.2
```

```
library(ggplot2)
library(dplyr)
library(fmsb)
```

```
Warning: package 'fmsb' was built under R version 4.4.2
```

```
library(gt)
```

```
Warning: package 'gt' was built under R version 4.4.2
```

Reading the Data:

```
nba_data <- read_csv("NBA_2024_per_game(03-01-2024).csv")
```

Research Questions:

All of the following are opinion based questions purely from statistics.

1. Who do we believe would be MVP (Most Valuable Player) in the final stretch of the NBA season (March 1 -) ?
2. Who do we believe would be SMOY (Six Man of the Year) in the final stretch of the NBA season (March 1-) ?
3. Who do we believe would be DPOY (Defensive Player of the Year) in the final stretch of the NBA season (March 1-) ?

Provenance of the Data:

This dataset was pulled from NBA Dataset and it consists of the comprehensive statistics for NBA players during the 2023-2024 regular season. The data was derived from Basketball Reference which is the most reliable source of NBA data which dates all the way back to 1946. The cases are made up of players and teams. The reason for this is due to some players moving teams, therefore having to record statistics from the same player but on two different teams, so two cases are needed.

FAIR and CARE Principles:

F indable: This dataset is publicly viewable on Kaggle and the name (NBA 2024 Per Game) contains common search keywords to make it easily findable.

A ccessible: Since this data is stored in a CSV file, it is easily accessible.

I nteroperable: The dataset uses commonly found basketball statistics that align with other sports (Points, Assists, Field Goals, etc.)

R eusable: The dataset contains a license for use: Attribution 4.0 International (CC BY 4.0)

C ollective Benefit: This data can both benefit the players, by showing them what aspects of their game they need to work on, and general public for sports betting purposes and to understand what type of player each guy is.

A uthority to Control: This data is controlled by the NBA but are shared publicly.

R esponsibility: Researchers and analysts should be transparent about their methods and findings so that they can ensure responsibility.

E thics: This data isn't necessarily sensitive data, so the ethics doesn't really apply to this dataset.

Focused Attributes:

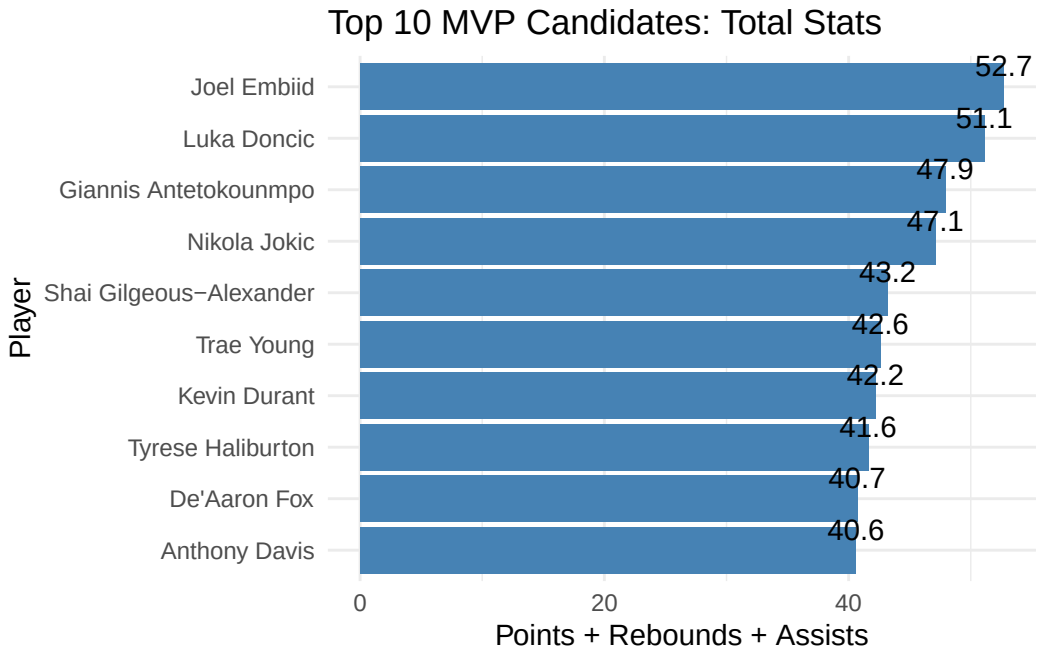
Regarding the three awards that we will be distributing, we will be focusing on a number of attributes from the data. All of which are already in the dataset. They will consist of the following:

- Player, PTS (Points), AST (Assists), TRB (Total Rebounds), STL (Steals), BLK (Blocks), TOV (Turnovers), MP (Minutes Played), GS (Games Started), Pos (Position)

Data Visualizations:

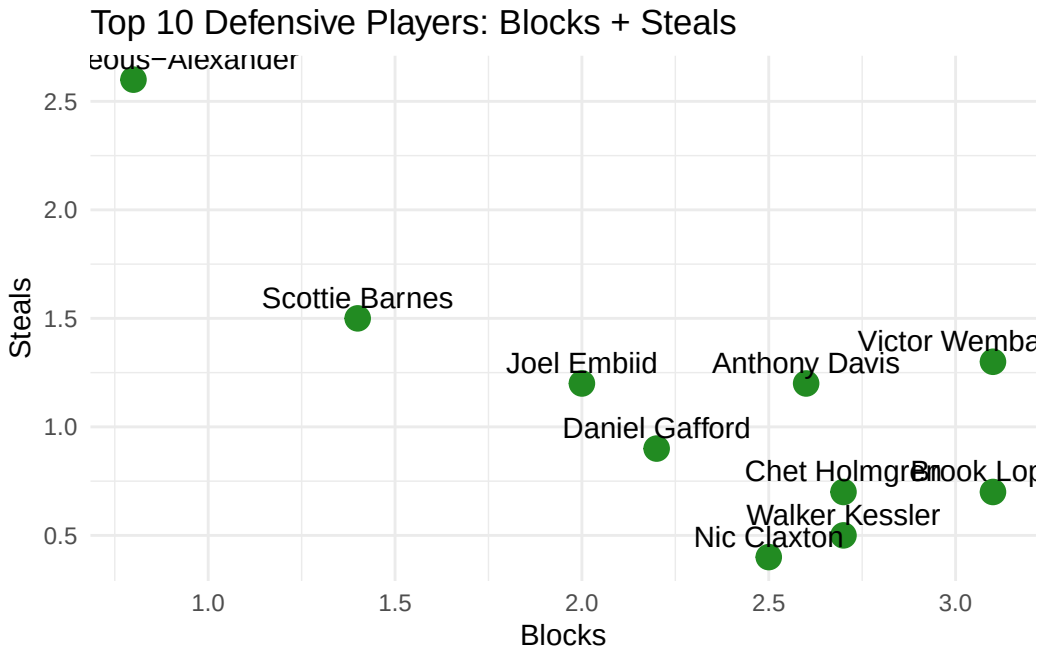
(Figure 1) MVP Candidates:

This Bar Graph is showing the top ten players with the most combined points, assists, and rebounds per game. One of the top 3 or 4 candidates will end up winning the MVP award.



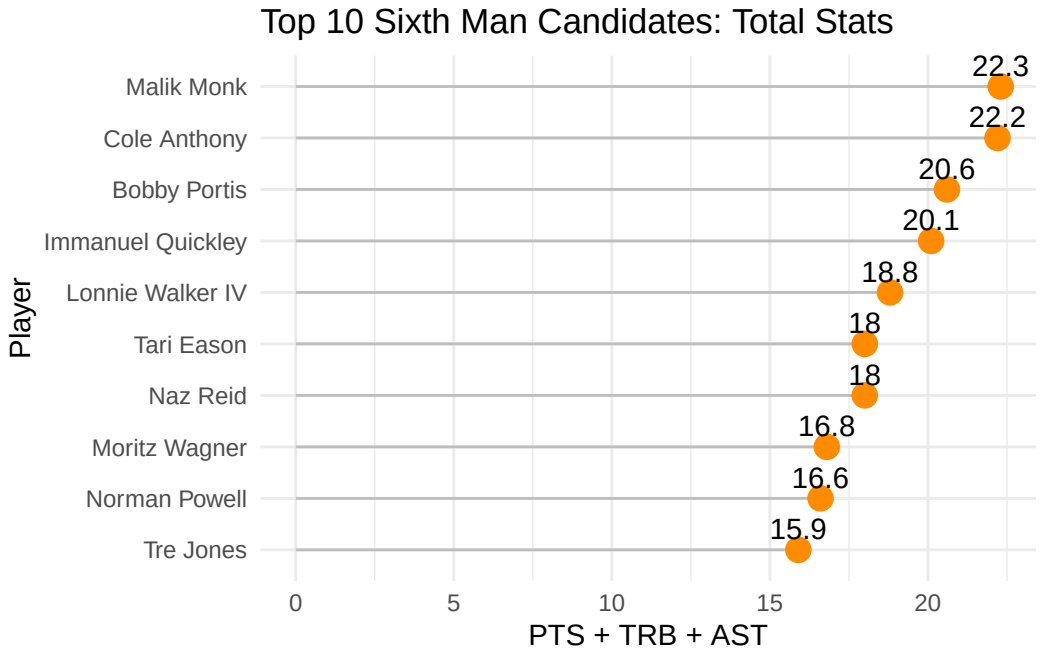
(Figure 2) DPOY Candidates:

This scatter plot is showing a comparison between total blocks and steals. It is only displaying the top ten players with the most combined blocks and steals.



(Figure 3) SMOY Candidates:

This lollipop chart is showing the top ten non-starter players that have the most combined points, rebounds, and assists per game. One of the top 3 or 4 players will end up winning the SMOY award.



Top 10 Team Performances

Team	Total Points	Avg Minutes Played	Total Rebounds	Total Assists
BRK	177.5	21.2	74.3	41.4
MIA	172.7	22.0	65.6	40.1
CHO	172.1	20.9	65.2	40.5
MEM	166.6	19.9	61.6	41.6
NOP	163.5	19.7	62.5	36.1
POR	163.5	19.4	67.0	34.5
DAL	162.4	18.0	61.6	36.2
TOR	161.5	18.9	66.9	39.4
LAC	160.0	18.1	64.6	37.1
PHO	158.2	20.7	58.8	35.4

(Figure 4) Team Performance:

This Table consists of the top ten team performances. This includes total average points, rebounds, and assists per game. This chart can be used to help us when considering who wins the awards.

NOTICE: There is some issue with LaTeX where it can't recognize this text before the code due to float values so it is stuck down here instead of above the graph. (Played around with it for hours and nothing fixed it)

Code Appendix:

```
#Figure 1
mvp_data <- nba_data %>%
  select(Player, PTS, AST, TRB) %>%
  mutate(Total_Offensive_Stats = PTS + AST + TRB) %>%
  arrange(desc(Total_Offensive_Stats)) %>%
  head(10)

ggplot(mvp_data, aes(x = reorder(Player, Total_Offensive_Stats), y = Total_Offensive_Stats))
  geom_bar(stat = "identity", fill = "steelblue") +
  geom_text(aes(label = round(Total_Offensive_Stats, 1)), vjust = -0.5) +
  labs(title = "Top 10 MVP Candidates: Total Stats",
       x = "Player",
       y = "Points + Rebounds + Assists") +
  theme_minimal() +
  coord_flip()


#Figure 2
defensive_data <- nba_data %>%
  select(Player, BLK, STL) %>%
  mutate(Total_Defensive_Stats = BLK + STL) %>%
  arrange(desc(Total_Defensive_Stats)) %>%
  head(10)

ggplot(defensive_data, aes(x = BLK, y = STL, label = Player)) +
  geom_point(size = 4, color = "forestgreen") +
  geom_text(vjust = -0.5, hjust = 0.5) +
  labs(title = "Top 10 Defensive Players: Blocks + Steals",
       x = "Blocks",
       y = "Steals") +
  theme_minimal()


#Figure 3
sixth_man_data <- nba_data %>%
  filter(GS == 0) %>%
  mutate(Total_Offensive_Stats = PTS + AST + TRB) %>%
  select(Player, Total_Offensive_Stats, G) %>%
  arrange(desc(Total_Offensive_Stats)) %>%
  head(10)
```

```r
ggplot(sixth_man_data, aes(x = reorder(Player, Total_Offensive_Stats), y = Total_Offensive_St
  geom_segment(aes(x = Player, xend = Player, y = 0, yend = Total_Offensive_Stats), color = "
  geom_point(size = 4, color = "darkorange") +
  geom_text(aes(label = round(Total_Offensive_Stats, 1)), vjust = -0.5) +
  labs(title = "Top 10 Sixth Man Candidates: Total Stats",
       x = "Player",
       y = "PTS + TRB + AST") +
  theme_minimal() +
  coord_flip()


#Figure 4
team_performance <- nba_data %>%
  group_by(Tm) %>%
  summarise(
    Total_Points = sum(PTS, na.rm = TRUE),
    Average_Minutes = mean(MP, na.rm = TRUE),
    Total_Rebounds = sum(TRB, na.rm = TRUE),
    Total_Assists = sum(AST, na.rm = TRUE)
  ) %>%
  arrange(desc(Total_Points)) %>%
  head(10)

team_performance %>%
  gt() %>%
  tab_header(
    title = "Top 10 Team Performances",
  ) %>%
  cols_label(
    Tm = "Team",
    Total_Points = "Total Points",
    Average_Minutes = "Avg Minutes Played",
    Total_Rebounds = "Total Rebounds",
    Total_Assists = "Total Assists"
  ) %>%
  fmt_number(
    columns = c(Total_Points, Average_Minutes, Total_Rebounds, Total_Assists),
    decimals = 1
  ) %>%
  data_color(
    columns = Total_Points,
    fn = scales::col_numeric(
```

```
      palette = "Blues",
      domain = NULL
    )
  ) %>%
  tab_options(
    table.font.size = 12,
    heading.align = "center",
    column_labels.font.weight = "bold"
  )
```