NBA_GMS_2010_2024_Project

2024-12-25

```r
library(tidyverse)
```

```
## -- 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   3.5.0     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 become errors
```

```r
library(ggplot2)
library(dplyr)
library(stringr)
library(readr)
```

Introduction

Background and Objective

This project examines the career trajectories and success metrics of NBA general managers (GMs) from 2010 to 2024. With aspirations of becoming a GM, I developed a dataset to analyze the credentials, career paths, and success levels of individuals who have held these positions. The dataset includes variables such as education, professional playing experience, and prior roles. I also created a General Manager Success Index (GMSI) to quantify performance, though it remains a work in progress. GMSI is a weighted average including the following columns and weight: Tenure (10%), Regular season winning percentage (15%), Playoff winning percentage (25%), Playoff appearances (10%), and a special equation for Championships (35%). This is a project I hope to continue to revise in the future as my skills as a data scientist improve. The primary goal of this project is to identify trends and patterns among NBA GMs that can inform aspiring professionals about the qualifications and pathways to success. Additionally, the study evaluates the relationship between certain variables and measurable success, as quantified by the GMSI.

Data Overview

Dataset Source and Description

The dataset was self-created using information from Basketball Reference, RealGM, and Wikipedia. The dataset includes 79 rows, each representing a unique GM from the 2009-2010 season onward. In order to keep the data tidy, GMs that had multiple tenures within this time period were only included once, that being their most recent tenure. Key variables include:

- **GM name and team**
- **Tenure (in years)**
- **Educational background** (College, Graduate School)
- **Playing experience** (NCAA or Professional Basketball)
- **Career history** (Job Histories)
- **Team performance metrics** (regular season and playoff win percentages, playoff appearances, championships)

I wanted to make note of the asterisks in the data set. Brad Stevens, Leon Rose, and Trajan Langdon are not general managers of their respective teams, but are the presidents of basketball operations. The Celtics, Knicks, and Pistons do not have active general managers. With that being said, I placed all three of them into my dataset. The president of basketball operations and general manager differ; the President of Basketball Operations (PoBO) has a macro role while the GM has a micro role in terms of roster construction and the team itself. PoBO sets the big-picture strategy that the general manager puts into place on a day-to-day basis.

Some values, particularly degree information or job history, were unavailable online and were marked as "Unknown." For neatness, front office roles that GMs worked in the past were standardized into the following abbreviations:

- , NBA = position was held for an NBA team
- AGM = Assitant General Manager
- AS = Assitant Scout
- Agent = Sports Agent
- ADoS = Assitant Director of Scouting
- BO = Basketball Operations
- BOA = Basketball Operations Analyst
- BOa = Basketball Operations Assitant
- BOi = Basketball Operations Intern
- DoBO = Director of Basketball Operations
- DoPP = Director of Player Personel
- DoPD = Director of Player Personel
- EVP = Executive Vice President
- EVPoBO = Executive Vice President of Basketball Operations
- FOi = Front Office Intern
- , GL = position was held for a G League Team
- HC = Head Coach
- HS = Head Scout
- SVPoBO = Senior Vice President of Basketball Operations
- TC = Team Counsel
- SC = Scouting Coordinator
- SMoBO = Senior Manager of Basketball Operations
- VPoBO = Vice President of Basketball Operations
- BOM = Basketball Operations Manager
- DoCS = Director of College Scouting
- DoS = Director of Scouting
- DoTO = Director of Team Operations
- GC = General Counsel
- PDD = Player Devlopment Director
- PDA = Player Development Assistant
- PPM = Player Personel Manager

- VPoPP = Vice President of Player Personel
- VPoS = Vice President of Scouting
- VC = Video Coordinator

Methodology

The project answers six key statistical questions by employing descriptive statistics and visualizations. Data cleaning and transformations were performed using R, and analysis was conducted to uncover patterns in GM tenure, education, work history, and success. The first data cleaning can be found below. To calculate GMSI, I will filter out the GMs who were hired before the 2024-2025 season, as this data set stopped recording data at the conclusion of the 2023-2024 season.

```
NBA_GMS_Temp_Final <- read_csv("C:/Users/lucas/Downloads/PERSONAL_PROJECTS/NBA_GMS_Temp_Final")
```

```
## Rows: 79 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (20): Name, Team, Current, College, Degree_Name, Pro, NCAA, Job_History,...
## dbl  (1): Tenure
##
## 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.
```

```
NBA_GMS_Temp_Final
```

```
## # A tibble: 79 x 21
##    Name         Team  Current Tenure College Degree_Name Pro   NCAA  Job_History
##    <chr>        <chr> <chr>    <dbl> <chr>   <chr>       <chr> <chr> <chr>
##  1 Elton Brand  76ers Y            6 Duke    Unknown     Y     Y     GM, GL
##  2 Jon Horst    Bucks Y            7 Roches~ Sports Man~ N     Y     DoBO, NBA
##  3 Marc Eversl~ Bulls Y            4 Urbana  Business A~ N     N     SVPoPP, NBA
##  4 Mike Gansey  Cavs  Y            3 WVU     Criminology Y     Y     GM, GL
##  5 Brad Steven~ Celt~ Y            3 DePauw  Economics   N     Y     HC, NBA
##  6 Trent Redden Clip~ Y            1 SMU     Accounting  N     Y     AGM, NBA
##  7 Zach Kleiman Griz~ Y            5 USC     Psychology  N     N     TC, NBA
##  8 Landry Fiel~ Hawks Y            2 Stanfo~ Communicat~ Y     Y     AGM, NBA
##  9 Andy Elisbu~ Heat  Y           11 St Tho~ Sports Adm~ N     N     VPoBO, NBA
## 10 Jeff Peters~ Horn~ Y            0 Arkans~ Marketing   N     Y     AGM, NBA
## # i 69 more rows
## # i 12 more variables: Job_History_II <chr>, Job_History_III <chr>,
## #   Job_History_IV <chr>, Job_History_V <chr>, Job_History_VI <chr>,
## #   Job_History_VII <chr>, Promoted <chr>, Grad_School <chr>,
## #   Regular_Season_Winning_Percentage <chr>, Playoff_Winning_Percentage <chr>,
## #   Playoff_Appearances <chr>, Championships <chr>
```

```
# Rather than manually editing the data set in sheets and reloading the csv file, I am going to make my
NBA_GMs_Cleaned <- NBA_GMS_Temp_Final |>
  mutate(
    Regular_Season_Winning_Percentage = na_if(Regular_Season_Winning_Percentage, 'Unknown'),
    Playoff_Winning_Percentage = na_if(Playoff_Winning_Percentage, 'Unknown'),
    Playoff_Appearances = na_if(Playoff_Appearances, 'Unknown'),
```

```r
    Championships = na_if(Championships, 'Unknown')
  )

# Now I am going to convert it to numeric.
NBA_GMs_Numeric <- NBA_GMs_Cleaned |>
  mutate(
    Regular_Season_Winning_Percentage = as.numeric(Regular_Season_Winning_Percentage),
    Playoff_Winning_Percentage = as.numeric(Playoff_Winning_Percentage),
    Playoff_Appearances = as.numeric(Playoff_Appearances),
    Championships = as.numeric(Championships)
  )

# Now I am going to filter out the NA values
NBA_GMs_Numeric <- NBA_GMs_Numeric |>
  filter(!is.na(Regular_Season_Winning_Percentage) &
         !is.na(Playoff_Winning_Percentage) &
         !is.na(Playoff_Appearances) &
         !is.na(Championships))

# Since my data is all squared away, I am going to create a weighted average variable called the Genera
NBA_GMs_GMSI <- NBA_GMs_Numeric |>
  mutate(
    GMSI = ((Tenure / 24) * 0.1) +
           (Regular_Season_Winning_Percentage * 0.15) +
           (Playoff_Winning_Percentage * 0.25) +
           ((Playoff_Appearances / Tenure) * 0.1) +
           (((Championships / Tenure + 1) / 0.28) * 0.35) -
           (((Tenure / 24) * (1 - Championships)) * 0.05)

)
# I now have a metric that measures success.
NBA_GMs_GMSI |> arrange(desc(GMSI)) |> relocate(GMSI)
```

```
## # A tibble: 73 x 22
##     GMSI Name   Team  Current Tenure College Degree_Name Pro   NCAA  Job_History
##    <dbl> <chr>  <chr> <chr>    <dbl> <chr>   <chr>       <chr> <chr> <chr>
##  1  31.1 Bob M~ Warr~ N           11 UCLA    Business    Y     Y     AGM, NBA
##  2  29.4 R.C. ~ Spurs N           17 Friends Management  N     Y     AGM, NBA
##  3  29.0 Brad ~ Celt~ Y            3 DePauw  Economics   N     Y     HC, NBA
##  4  27.0 David~ Cavs  N            4 ASU     Poli Sci    N     N     VPoBO, NBA
##  5  26.2 Pat R~ Heat  N           24 Kentuc~ Unknown     Y     Y     HC, NBA
##  6  25.7 Calvi~ Nugg~ Y            5 Penn S~ Unknown     Y     Y     AGM, NBA
##  7  25.6 Bobby~ Rapt~ Y            7 UCSB    Economics   N     N     AGM, NBA
##  8  25.2 Jon H~ Bucks Y            7 Roches~ Sports Man~ N     Y     DoBO, NBA
##  9  24.9 Geoff~ Kings N           19 Prince~ English     Y     Y     VPoBO, NBA
## 10  24.7 Andy ~ Heat  Y           11 St Tho~ Sports Adm~ N     N     VPoBO, NBA
## # i 63 more rows
## # i 12 more variables: Job_History_II <chr>, Job_History_III <chr>,
## #   Job_History_IV <chr>, Job_History_V <chr>, Job_History_VI <chr>,
## #   Job_History_VII <chr>, Promoted <chr>, Grad_School <chr>,
## #   Regular_Season_Winning_Percentage <dbl>, Playoff_Winning_Percentage <dbl>,
## #   Playoff_Appearances <dbl>, Championships <dbl>
```

Question 1 [What is the average tenure of a GM since 2010?]

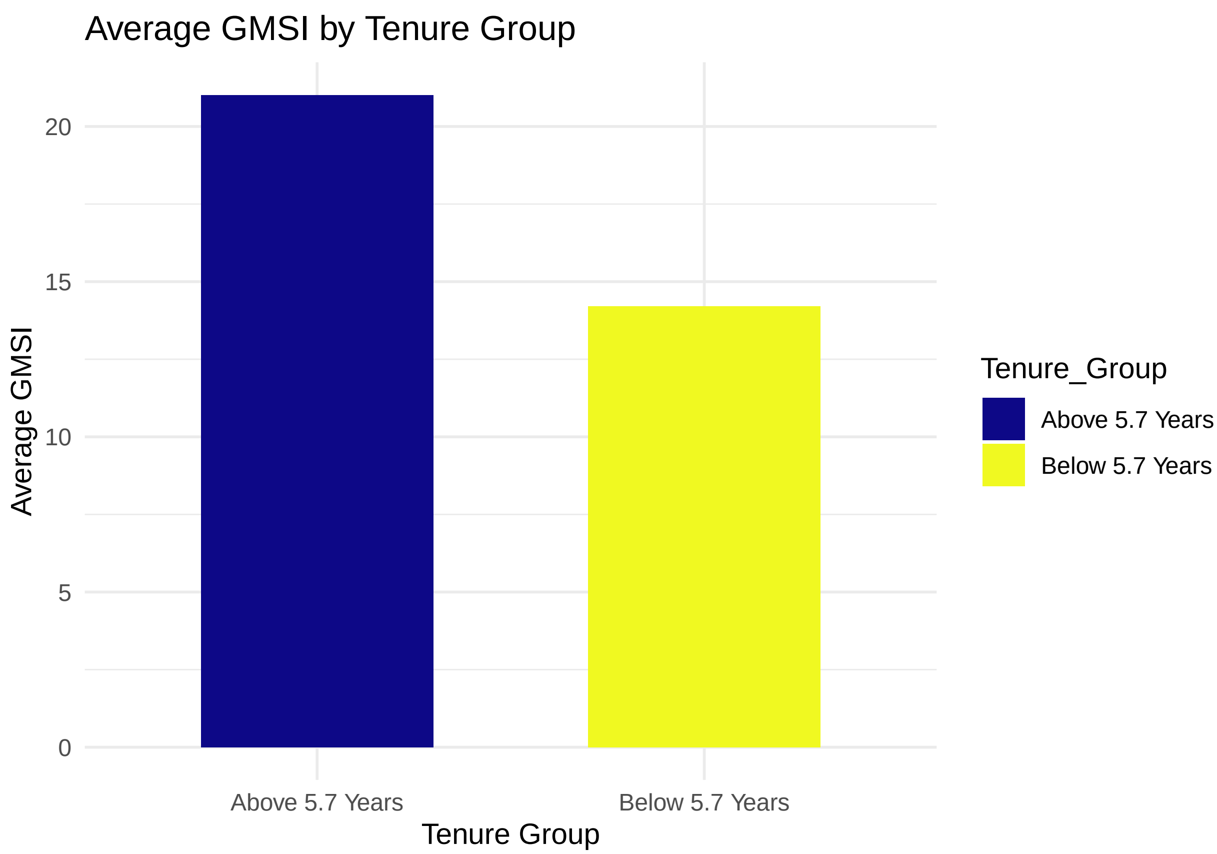
I will answer this first question by finding the mean of the tenure column.

```r
# I simply calculated the mean using the mean function.
summary(NBA_GMS_Temp_Final$Tenure)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   3.000   5.000   5.709   7.000  24.000
```

The average tenure of an NBA GM is 5.7 years. This finding aligns with the long-term nature of roster construction and the time required for strategic decisions to manifest in team success. However, 5.7 years can be seen as a double-edged sword. While it allows GMs to implement their vision, it may not always be sufficient to overcome challenges such as injuries, roster missteps, or rebuilding cycles. On the other hand, a tenure exceeding this average often reflects stability and alignment between the GM and ownership, which are critical for sustained success. Therefore, comparing GMSI between GMs who exceed this average tenure and GMs who fail to meet this average tenure would be telling.

```r
# Now I am going to compare GMSI between those who exceed 5.4 years and those who do not by calculating
GMSI_summary <- NBA_GMs_GMSI |>
  mutate(Tenure_Group = ifelse(Tenure > 5.7, "Above 5.7 Years", "Below 5.7 Years")) |>
  group_by(Tenure_Group) |>
  summarise(Average_GMSI = mean(GMSI))

# Now I will create the graph
GMSI_summary |> ggplot(aes(x = Tenure_Group, y = Average_GMSI, fill = Tenure_Group)) +
  geom_col(width = 0.6) +
  labs(
    title = "Average GMSI by Tenure Group",
    x = "Tenure Group",
    y = "Average GMSI"
  ) +
  scale_fill_viridis_d(option = "C") +
  theme_minimal()
```



As you can see, GMs who's tenure exceeds 5.7 years have much better GMSI, on average, than those who do not exceed 5.7 years. The difference appears to be about 5 points on the graph.

Question 2 [Which teams have had the most amount of GMs since 2010?]

A high frequency of changing GMs will likely signal poor team success.

```
# Here I create a frequency table to give me the mode of the categorical variable.
created_table <- table(NBA_GMS_Temp_Final$Team)
# Then I found the mode of the table.
max_freq <- max(created_table)
max_freq
```

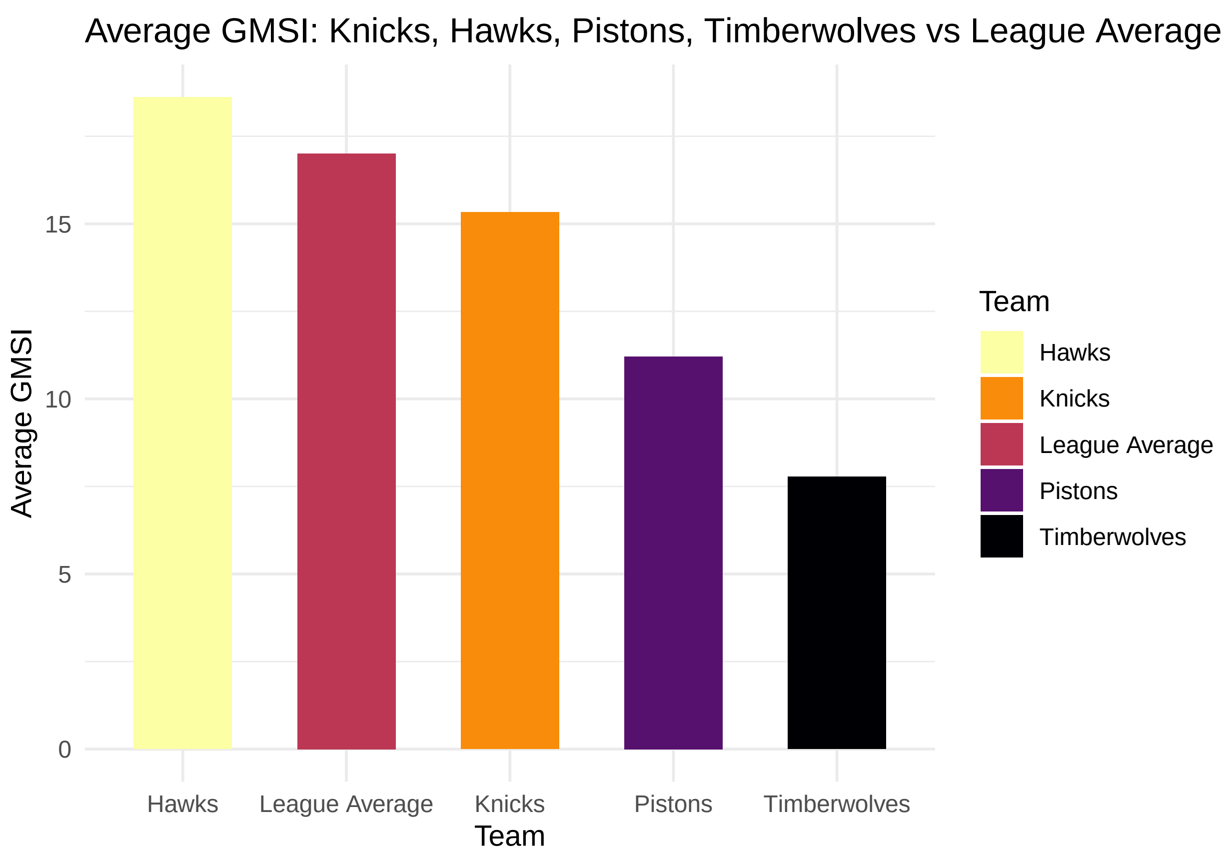
```
## [1] 5
```

```
# max_freq outputs a value of 5, meaning that the mode of the table is 5. I then used this information
most_frequent_team <- names(created_table[created_table == max_freq])
# This is me printing the value.
most_frequent_team
```

```
## [1] "Hawks"        "Knicks"       "Pistons"      "Timberwolves"
```

The Hawks, Knicks, Pistons, and Timberwolves have each employed five GMs, indicating instability within their front offices. Frequent turnover often correlates with poor team performance and a lack of consistent vision. Let's compare GMSI of these teams to the league average.

```r
# Calculate average GMSI for each of the four teams and league average
team_analysis <- NBA_GMs_GMSI |>
  filter(Team %in% c("Knicks", "Hawks", "Pistons", "Timberwolves")) |>
  group_by(Team) |>
  summarise(Average_GMSI = mean(GMSI)) |>
  bind_rows(
    NBA_GMs_GMSI |>
      summarise(Team = "League Average", Average_GMSI = mean(GMSI))
  )

# Create the graph
team_analysis |> ggplot(aes(x = reorder(Team, -Average_GMSI), y = Average_GMSI, fill = Team)) +
  geom_col(width = 0.6) +
  labs(
    title = "Average GMSI: Knicks, Hawks, Pistons, Timberwolves vs League Average",
    x = "Team",
    y = "Average GMSI"
  ) +
  scale_fill_viridis_d(option = "B", direction = -1) +
  theme_minimal()
```

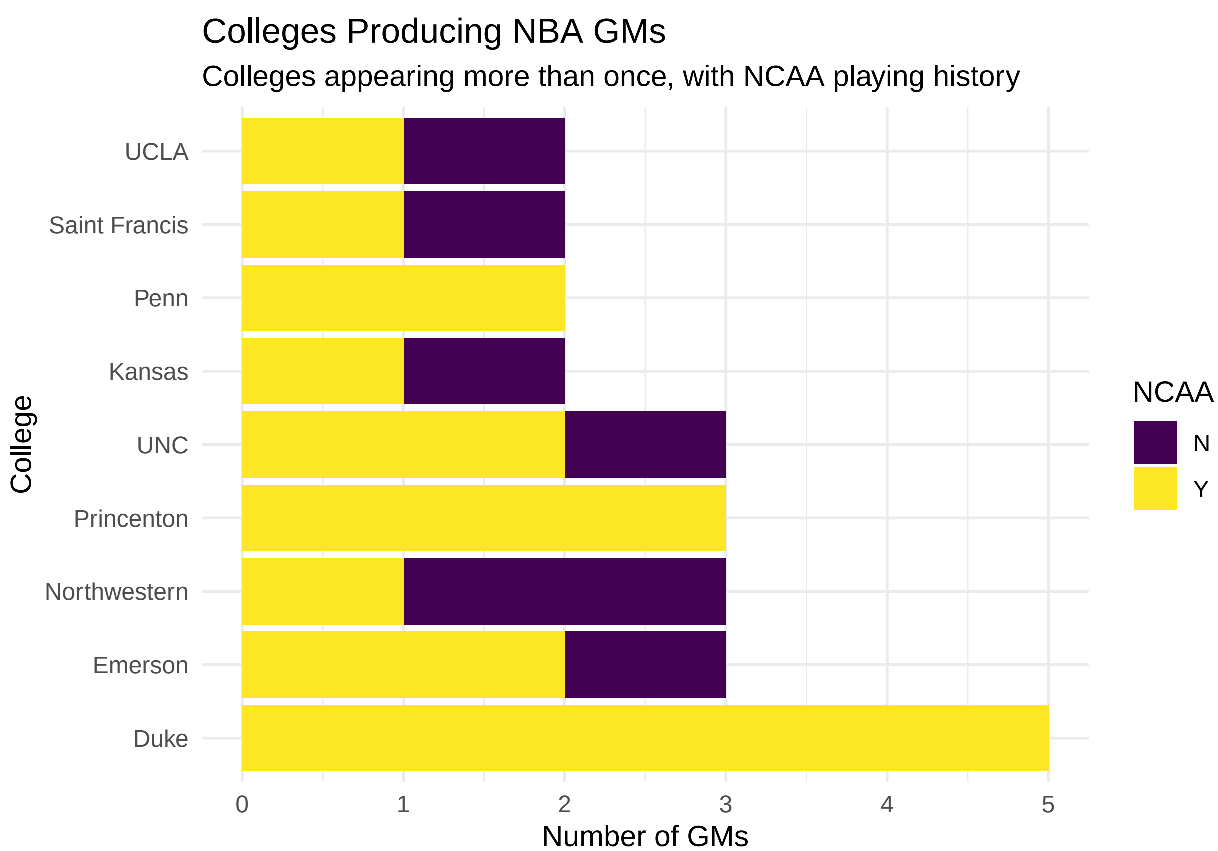


Surprisingly, the Hawks exceed the league average despite having had 5 GMs over this time period. While they haven't ranked atop of the NBA, excluding the 2014-2015 season, they have consistently made the playoffs across all of their GMs. However, the Knicks, Pistons, and Timberwolves have all had terrible losing stretches across this time period, which explains their low GMSI. The Timberwolves being at the bottom of the list makes sense since the Western conference has been much stronger during this time period than the Eastern conference, making it more difficult for them to make the playoffs compared to the Knicks and Pistons. This is important because of the weight that making the playoffs has on the GMSI metric.

Question 3 [Where did GMs go to college?]

Now I'm interested in the colleges where these GMs went. I'll make a graph displaying their colleges to see if there's any signs of a potential pattern. To avoid clutter and make the data more relevant to observing patterns, I am only going to display colleges that appear more than once.

```
# First I need to create a new data set that has a filtered "Colleges" column that only has colleges th
filtered_colleges_NBA_GMS <- NBA_GMS_Temp_Final |> add_count(College) |> filter(n > 1)
# Now I am going to create a bar graph to display this data. I want to fill by the NCAA variable to see
filtered_colleges_NBA_GMS |>
  ggplot(aes(x = reorder(College, -n), fill = NCAA)) +
  geom_bar() +
  coord_flip() +
  labs(
    title = "Colleges Producing NBA GMs",
    subtitle = "Colleges appearing more than once, with NCAA playing history",
    x = "College",
    y = "Number of GMs"
  ) +
  scale_fill_viridis_d(option = "D") +
  theme_minimal()
```



```r
# I wanted to see which college had the most amount of GMs who didn't play college basketball, which I
filtered_colleges_NBA_GMS |> count(NCAA)
```

```
## # A tibble: 2 x 2
##   NCAA      n
##   <chr> <int>
## 1 N         7
## 2 Y        18
```

```r
NBA_GMS_Temp_Final |> filter(College == "Northwestern") |> filter(NCAA == "N")
```

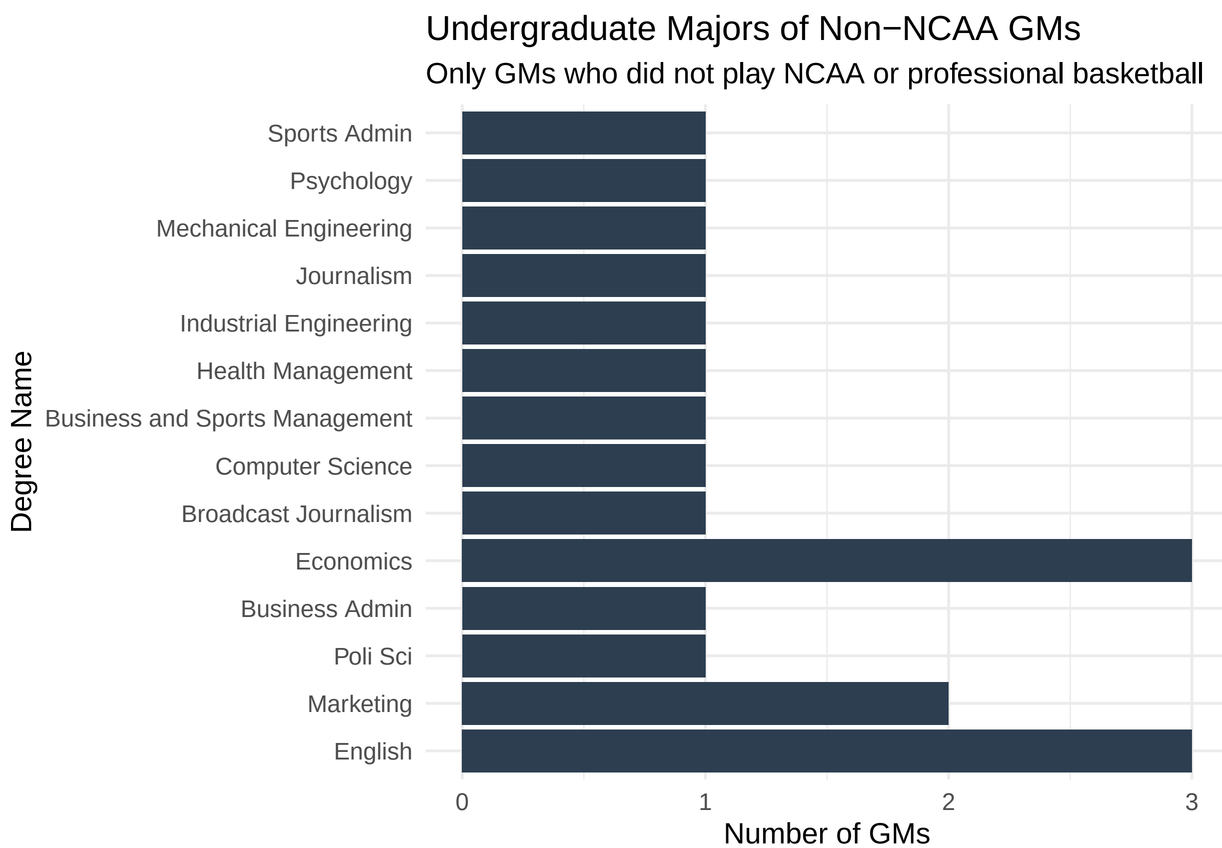
```
## # A tibble: 2 x 21
##   Name         Team   Current Tenure College Degree_Name Pro   NCAA  Job_History
##   <chr>        <chr>  <chr>    <dbl> <chr>   <chr>       <chr> <chr> <chr>
## 1 Justin Zanik Jazz   Y            6 Northw~ Economics   N     N     AGM, NBA
## 2 Daryl Morey  Rocke~ N           13 Northw~ Computer S~ N     N     DoBO, NBA
## # i 12 more variables: Job_History_II <chr>, Job_History_III <chr>,
## #   Job_History_IV <chr>, Job_History_V <chr>, Job_History_VI <chr>,
## #   Job_History_VII <chr>, Promoted <chr>, Grad_School <chr>,
## #   Regular_Season_Winning_Percentage <chr>, Playoff_Winning_Percentage <chr>,
## #   Playoff_Appearances <chr>, Championships <chr>
```

Analysis of educational backgrounds reveals that certain schools such as Duke, Princeton, UNC, and Northwestern appear frequently in this dataset. This is most likely attributed to these institutions' combination of strong academic reputations and successful basketball programs, which can provide both connections and credibility in the sports industry. In general, playing college basketball helps connect individuals with the NBA network, which is reflected by the 18 out of 25 GMs on this list who played in college. Northwestern stands out as the only school producing multiple GMs who did not play basketball, those two being Justin Zanik and Daryl Morey.

Since I do not play college basketball, I am interested in the GMs who did not play in college, whom I will explore below.

Question 4 [What did GMs who did not play college basketball major in?]

```r
# First I will make a new filtered data set so that I can easily create a graph.
NBA_GMS_filtered_by_degree <- NBA_GMS_Temp_Final |> add_count(Degree_Name) |> filter(NCAA == "N") |> fil
# Now I will create a bar graph
NBA_GMS_filtered_by_degree |>
  ggplot(aes(x = reorder(Degree_Name, -n))) +
  geom_bar(fill = "#2C3E50") +
  coord_flip() +
  labs(
    title = "Undergraduate Majors of Non-NCAA GMs",
    subtitle = "Only GMs who did not play NCAA or professional basketball",
    x = "Degree Name",
    y = "Number of GMs"
  ) +
  theme_minimal()
```



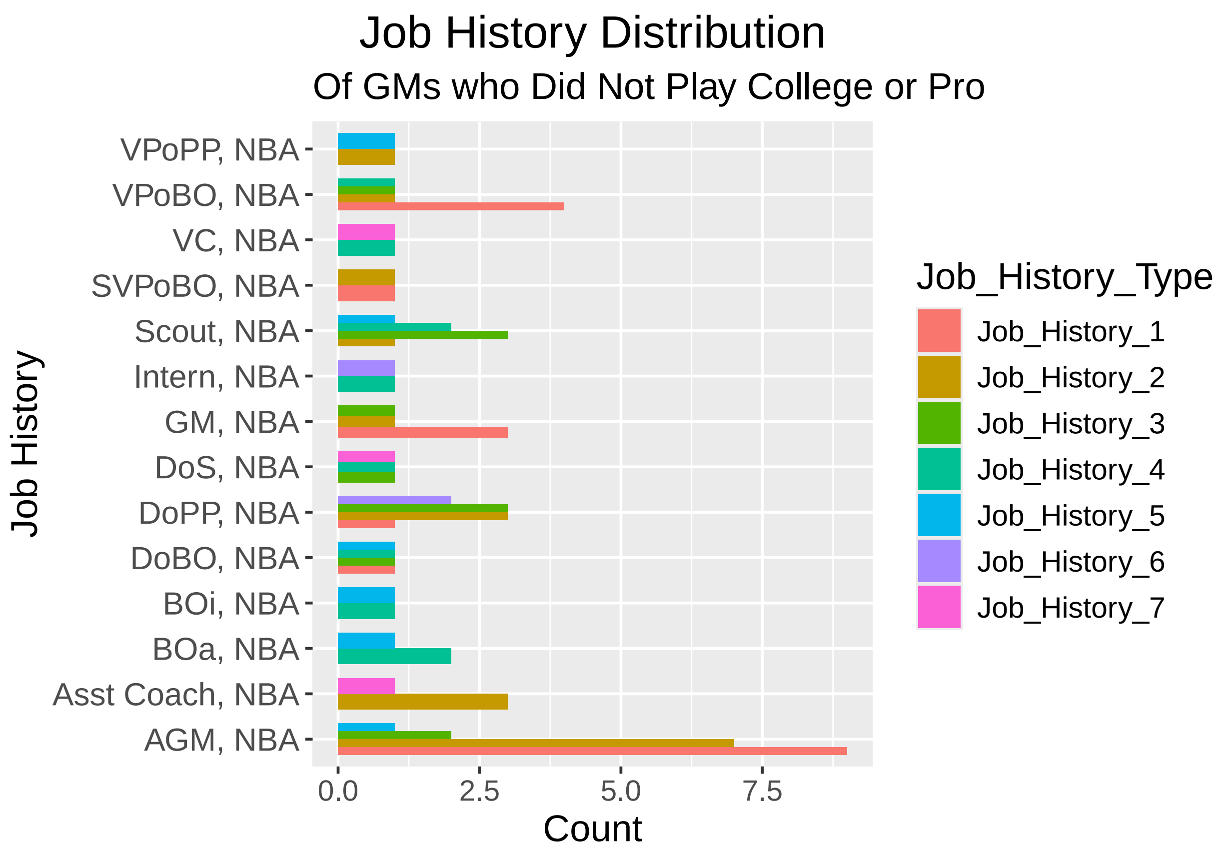
Among GMs who lacked collegiate or professional basketball experience, the most common undergraduate majors were Marketing, English, and Economics. These findings suggest that while technical degrees may not dominate, skills in communication, strategy, and analysis are valued. Communication skills enable GMs to effectively negotiate contracts, manage team dynamics, and collaborate with stakeholders. Strategic thinking is crucial for long-term planning, including drafting, trades, and cap management. Analytical skills are indispensable for interpreting performance metrics and identifying undervalued talent, making these competencies essential to navigating the multifaceted challenges of a GM role. With that being said, the small sample size makes it difficult to make any conclusive statements about college degrees increasing someone's chance of becoming a GM, or a good one for that matter.

Question 5 [Where did GMs work prior to their hiring?]

As a result, I am now interested in the work experience of these GMs who did not play pro or college basketball. I will display that below.

```
# I will once again make an modified data set to create my graph.
NBA_GMS_work_experience <- NBA_GMS_Temp_Final |> add_count(Job_History, Job_History_II, Job_History_III

# Now I will make a column that includes all of the job history values. To do this, I am going to have
job_history_long <- NBA_GMS_work_experience |>
  select(Name,
         Job_History_1 = Job_History,
         Job_History_2 = Job_History_II,
         Job_History_3 = Job_History_III,
         Job_History_4 = Job_History_IV,
         Job_History_5 = Job_History_V,
```

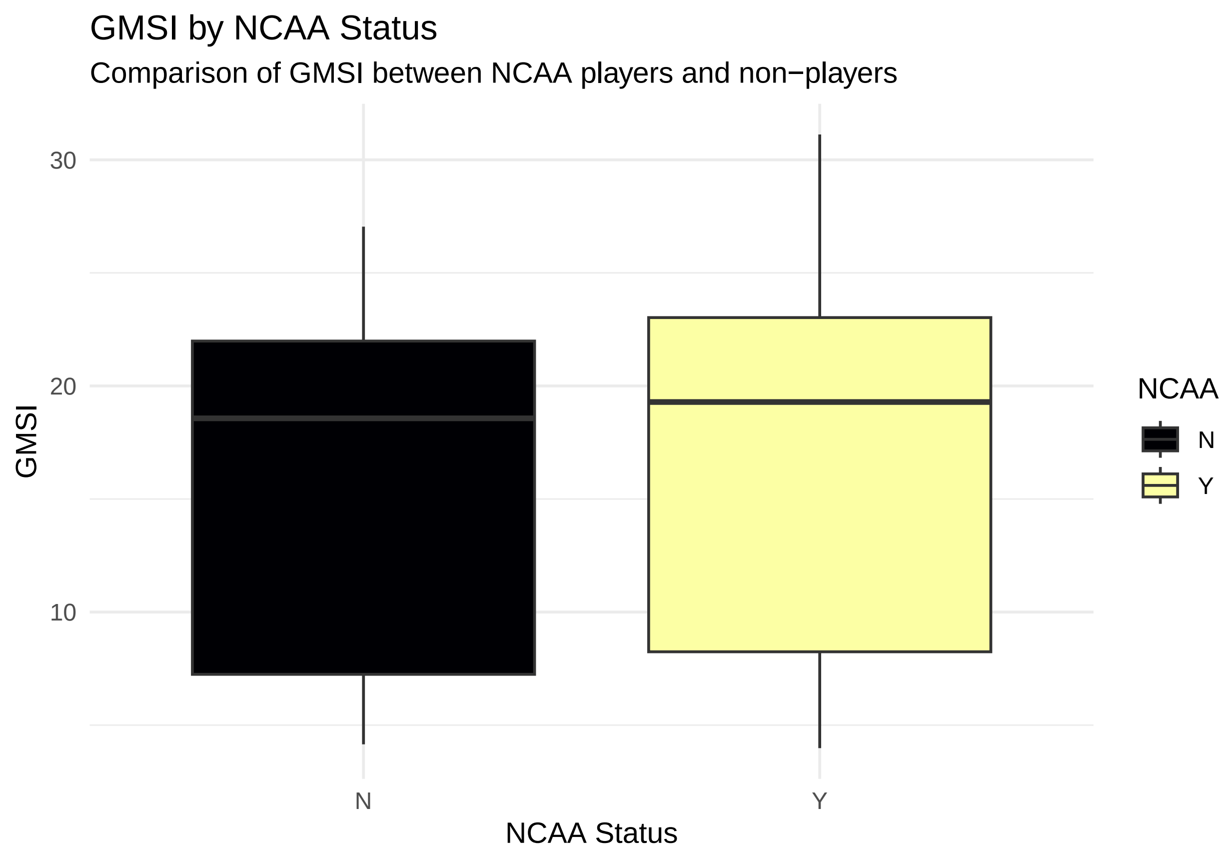
```
        Job_History_6 = Job_History_VI,
        Job_History_7 = Job_History_VII) |>
  pivot_longer(cols = starts_with('Job_History'),
               names_to = 'Job_History_Type',
               values_to = 'Job_History') |>
  filter(Job_History != 'Unknown')

# Count occurrences of each job history. I am doing this because in the code above, I realized that the
job_history_counts <- job_history_long |>
  group_by(Job_History) |>
  summarise(Count = n()) |>
  filter(Count > 1)

# Filtered the original data set to include only job histories that occur more than once
filtered_job_history_long <- job_history_long |>
  filter(Job_History %in% job_history_counts$Job_History)

# Now I will create a graph displaying their job histories along with when in their respective careers
filtered_job_history_long |>
  ggplot(aes(x = Job_History, fill = Job_History_Type)) +
  geom_bar(position = 'dodge', width = 0.7) +
  coord_flip() +
  theme(text = element_text(size = 14),
        axis.text.y = element_text(size = 12),
        plot.title = element_text(hjust = 0.5)) +
  labs(title = 'Job History Distribution', subtitle = 'Of GMs who Did Not Play College or Pro', x = 'Jol
```



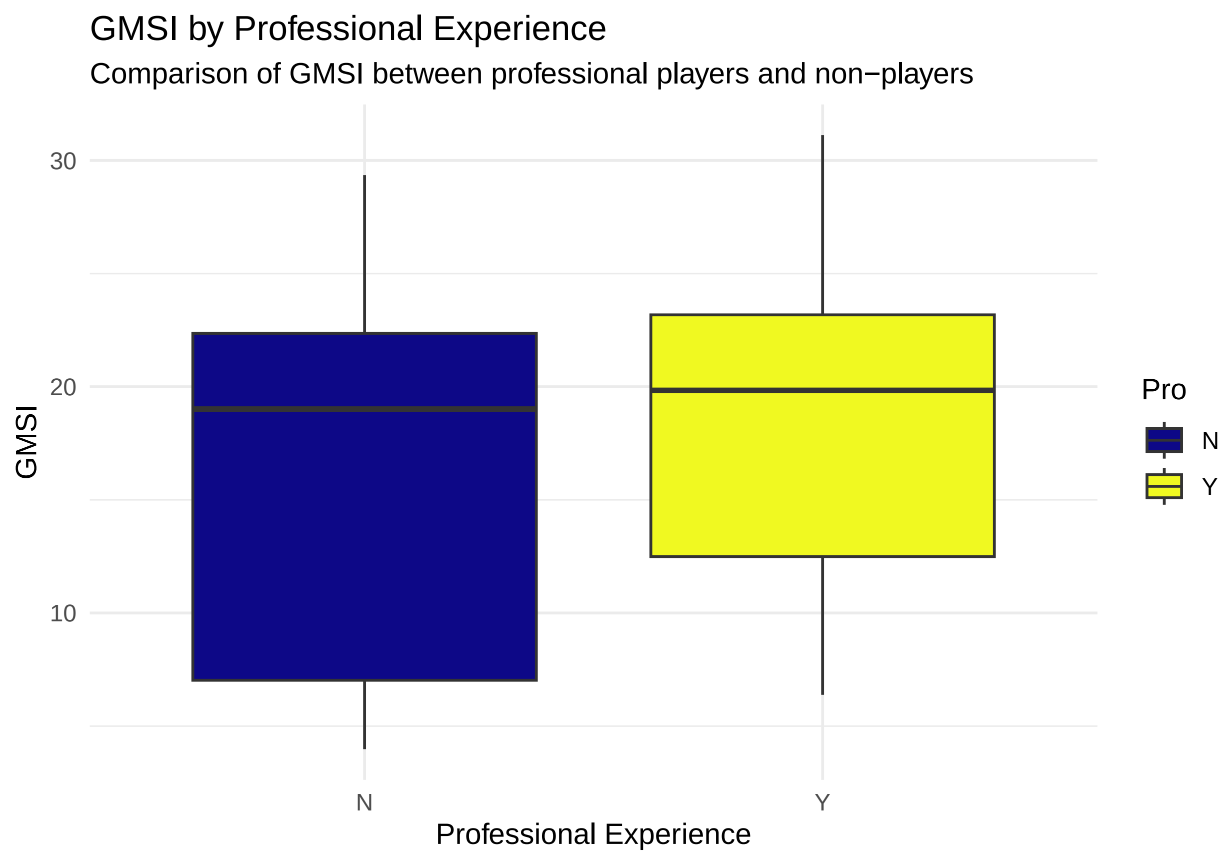
The job with the most experience prior to becoming a GM was Assistant GM of an NBA team. This makes sense since being an assistant to your future position would give you the utmost amount of relevant experience. Another notable job with relevant application is Vice President of Basketball Operations. Oftentimes, a GM is either promoted to President of Basketball Operations or holds the title simultaneously to being GM to give them maximum control over the team. I'd argue that being the VPoBO is one of the best jobs to prepare yourself for being a GM since you've seen your position from your boss' perspective. Being a scout is also a common path taken towards working your way up the front office hierarchy as GMs must succeed in evaluating talent in the draft and free agency. Director of Player Personnel also fits this description as they are in charge of overseeing all scouting and player acquisition. These roles provide critical experience in player evaluation and team operations, all of which are foundational to GM responsibilities.

Question 6 [What variables contribute to success in terms of GMSI?]

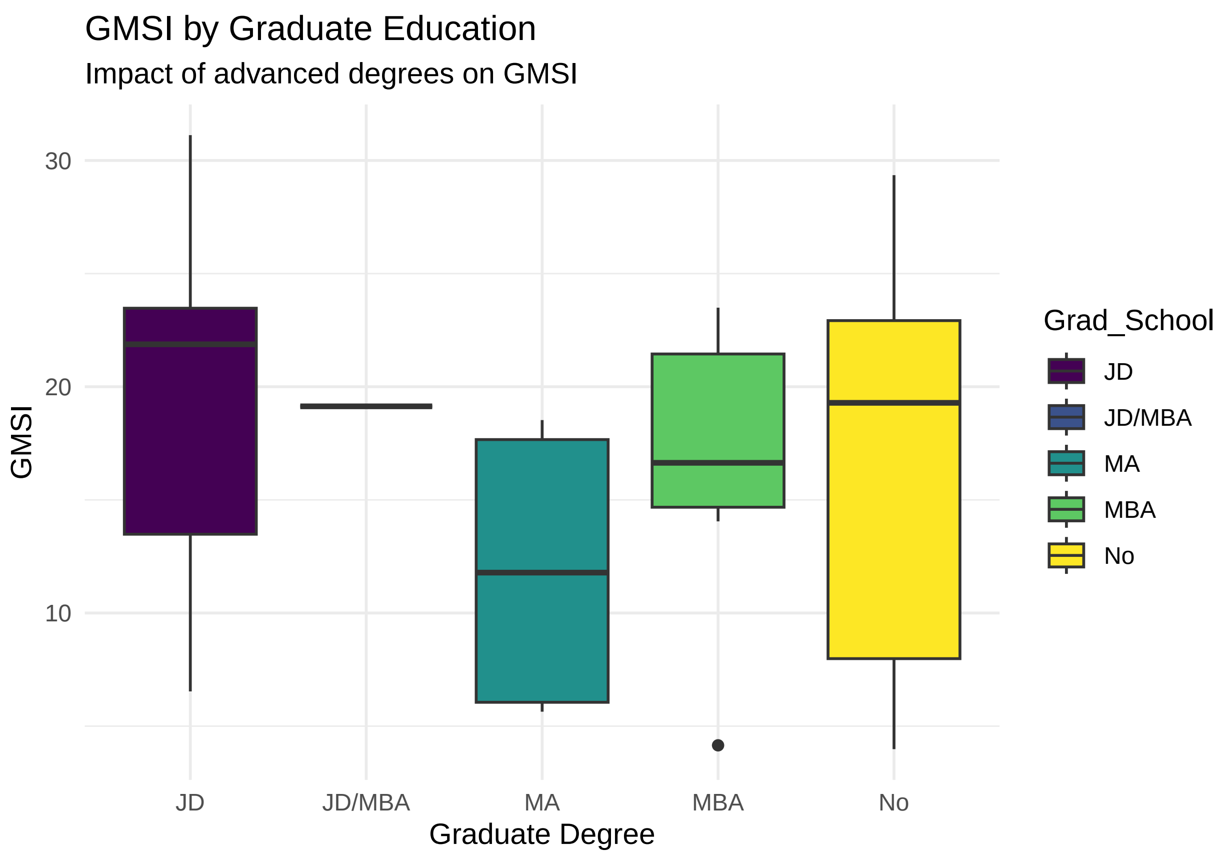
```
# I will now visualize the relationship between GMSI and various categorical variables. First up is GMS
NBA_GMs_GMSI |>
  ggplot(aes(x = NCAA, y = GMSI, fill = NCAA)) +
  geom_boxplot() +
  labs(
    title = "GMSI by NCAA Status",
    subtitle = "Comparison of GMSI between NCAA players and non-players",
    x = "NCAA Status",
    y = "GMSI"
  ) +
  scale_fill_viridis_d(option = "B") +
  theme_minimal()
```



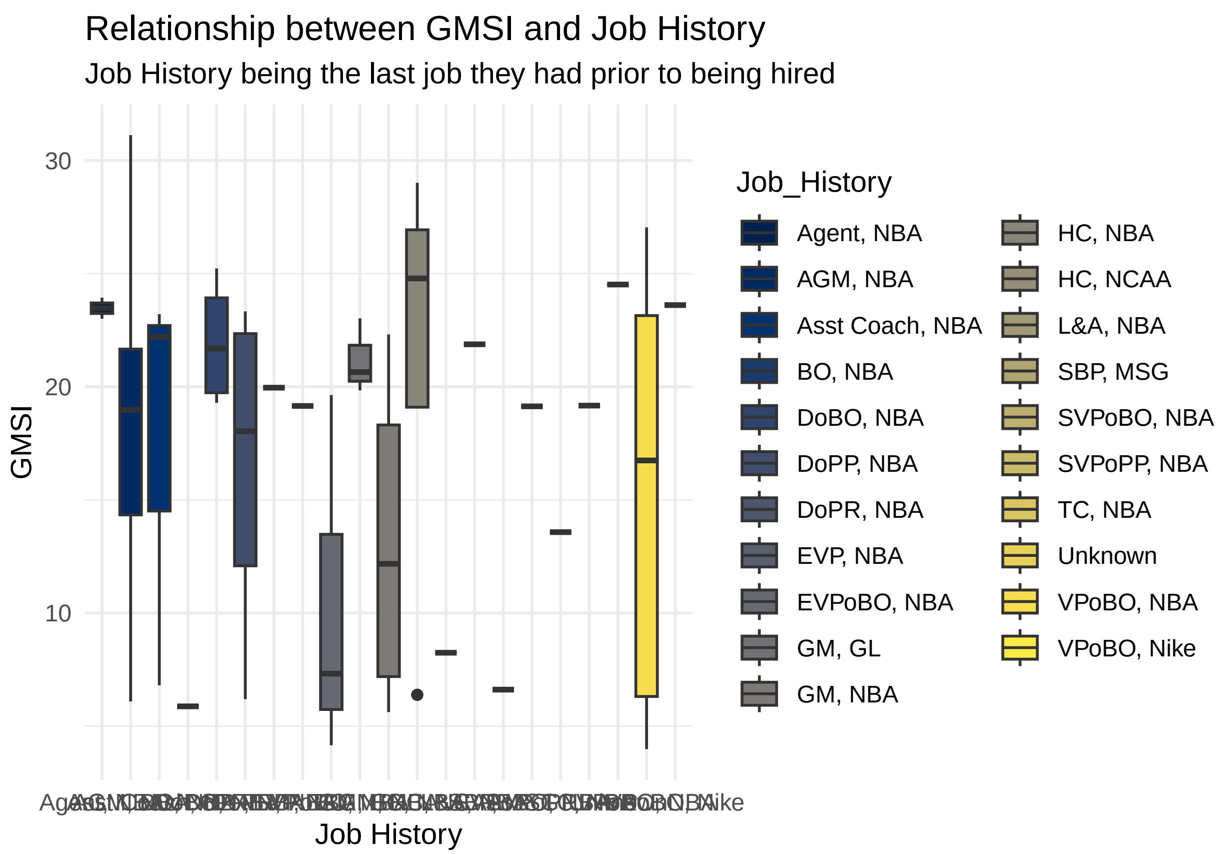

```
# This is GMSI and Pro Status
NBA_GMs_GMSI |>
  ggplot(aes(x = Pro, y = GMSI, fill = Pro)) +
  geom_boxplot() +
  labs(
    title = "GMSI by Professional Experience",
    subtitle = "Comparison of GMSI between professional players and non-players",
    x = "Professional Experience",
    y = "GMSI"
  ) +
  scale_fill_viridis_d(option = "C") +
  theme_minimal()
```



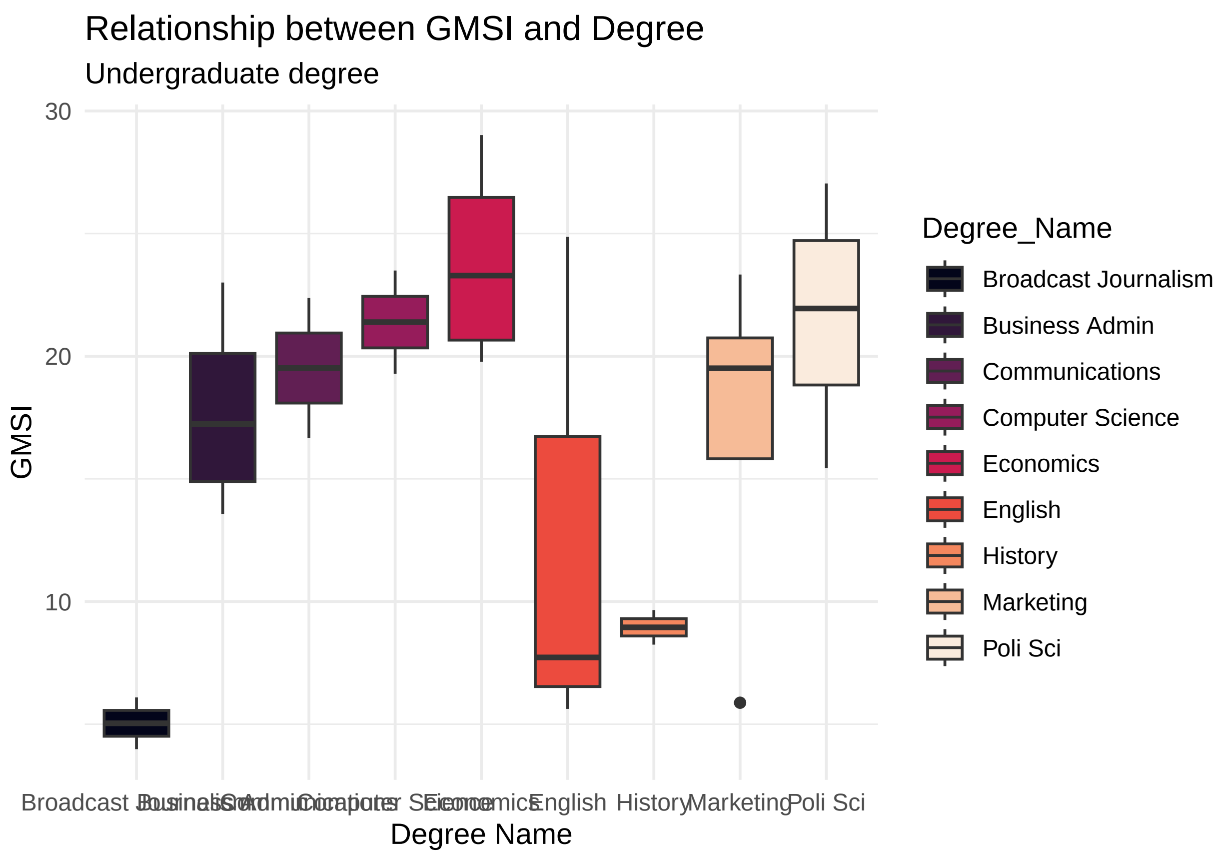

```r
# This is GMSI and whether or not they went to graduate school
NBA_GMs_GMSI |>
  ggplot(aes(x = Grad_School, y = GMSI, fill = Grad_School)) +
  geom_boxplot() +
  labs(
    title = "GMSI by Graduate Education",
    subtitle = "Impact of advanced degrees on GMSI",
    x = "Graduate Degree",
    y = "GMSI"
  ) +
  scale_fill_viridis_d(option = "D") +
  theme_minimal()
```




```r
# This is GMSI and the most recent job the GMs had prior to being hired
NBA_GMs_GMSI |>
  ggplot(aes(x = Job_History, y = GMSI, fill = Job_History)) +
  geom_boxplot() +
  labs(
    title = "Relationship between GMSI and Job History",
    subtitle = "Job History being the last job they had prior to being hired",
    x = "Job History",
    y = "GMSI"
  ) +
  scale_fill_viridis_d(option = "E") +
  theme_minimal()
```




```r
# Now I am going to filter out the data set again so I can analyze college degrees
GMSI_filtered_by_degree <- NBA_GMs_GMSI |> add_count(Degree_Name) |> filter(Degree_Name != "Unknown") |

GMSI_filtered_by_degree <- NBA_GMs_GMSI |>
  filter(Degree_Name %in% GMSI_filtered_by_degree$Degree_Name)

# This is the visualization of GMSI and college degrees
GMSI_filtered_by_degree |>
  ggplot(aes(x = Degree_Name, y = GMSI, fill = Degree_Name)) +
  geom_boxplot() +
  labs(
    title = "Relationship between GMSI and Degree",
    subtitle = "Undergraduate degree",
    x = "Degree Name",
    y = "GMSI"
  ) +
  scale_fill_viridis_d(option = "F") +
  theme_minimal()
```



GMSI vs. NCAA Status:

GMs who played college basketball have slightly higher GMSI scores than non-NCAA GMs. While one could argue that the difference is sizable, the reality is that this small difference is likely attributed to the high frequency of GMs who played in college.

GMSI vs. Professional Basketball Experience:

A similar trend emerges for professional playing experience, with slight advantages in GMSI for former professional players. Playing experience likely offers insights into team dynamics and player evaluation, and could potentially give GMs an edge in the locker room. Reliability amongst management and players can easily make a difference in morale and trust. However, I think the large difference in IQR between the two groups reflects that GMs with professional playing experience perform better than those without.

GMSI vs. Graduate Education:

GMs with JDs demonstrate high GMSI variability. Most of them have work experience in sports agency or the legal field of sports, likely making them great negotiators and contract makers. Graduate education may enhance strategic thinking and negotiation skills, though it is not a consistent predictor of success, as seen by the varying level of success across different graduate degrees.

GMSI vs. Last Job Held:

Assistant GM, DoBO, and scouting roles appear most correlated with higher GMSI scores. These positions directly involve decision-making and talent evaluation, integral to GM success. Agents have also seen a plentiful amount of success, with GMs such as Bob Myers and Rob Pelinka winning championships. However, their share is extremely small so it's difficult to make conclusive statements about them. The Head Coach category is skewed by the success of both Pat Riley and Brad Stevens. EVPoBOs have had the least amount of success, reflected by their extremely low median value and IQR.

GMSI vs. Undergraduate Degree:

Degrees in Economics and Political Science correspond to higher GMSI scores, reinforcing the importance of analytical and managerial skills in front-office roles. However, their success could be reflected by how common these degrees are within this dataset.

Conclusion

This project highlights the diverse pathways and qualifications of NBA GMs while providing insights into the variables influencing their success. Key findings include:

- Playing experience, both collegiate and especially professional, can be advantageous but are not mandatory for success.
- Certain educational backgrounds, such as Economics or Political Science, and especially graduate degrees such as JDs or MBAs, are associated with higher success metrics.
- Prior roles like Assistant GM and Director of Basketball Operations are strong precursors to GM appointments and success, while sports agent seem to be on the rise in terms of success.
- The GMSI serves as a useful, albeit incomplete, metric for evaluating GM performance. Future iterations of this project could refine the index by incorporating additional variables, such as draft outcomes or free-agent signings. Ultimately, this study underscores the multifaceted skill set required to excel as an NBA GM, combining analytical acumen, strategic foresight, and effective leadership.

Future

I plan on quantifying draft success in the near future. This is something that no one that I know of has done yet, but is something that is necessary. Doug Drinen, the founder of Pro Football Reference, created a metric, Approximate Value (AV), that places a number on the value of a player a given season. AV is a great metric for measuring draft success since it takes into account amount of starts, pro bowl appearances, and overall impact. I would love to be able to create a statistic for the NBA so that I could analyze the average AV of players that GMs drafted. Draft success is the most obvious thing missing from the dataset, so I plan on fixing that once my skills as a data scientist improve.