Blue Lock Player Analysis

Lillian Lin

Summer 2025

```
##
## 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
```

1 R Coding Final - A Blue Lock Analysis

1.1 Lillian Lin

1.1.1 GA Tech, Summer 2025

1.1.1.1 Abstract

This project analyzes football performance statistics from the manga *Blue Lock*, using player attributes such as Speed, Defense, Pass, Dribble, Shoot, and Offense.

The **central research questions** are:

1. Which players demonstrate the highest individual skill scores, and are they consistent top performers overall?
2. How closely does a player's average skill score align with their overall Player Score?
3. Do the characters show tendencies towards specific roles (ex. offense-heavy or well-rounded)?
4. What does the overall distribution of skills look like across all characters - are specific attributes more emphasized than other ones?

The data was compiled from published *Blue Lock* character stat charts found in the manga that was produced onto a website by a fan. Though fictional, these stats mirror real-world performance indicators.

While direct academic literature on fictional sports data is limited, this project is inspired by broader sports analytics works and performance modeling. For example:

- **Bunker & Thabtah (2019)** explored predictive modeling in football, focusing on performance indicators.

- **Miller et al. (2014)** emphasized the growing role of analytics in decision-making in professional sports teams.

Data Source: https://blue-lock-marcelones.vercel.app/

GitHub Profile: lillianlin01
Repository: `econ4001_fin`

1.1.2 Table of Contents

1. Background
2. The Data
    - Cleaning Goals
    - What worked from my Midterm plans
    - What did not work from the Midterm and why
3. Research Questions
4. Analysis Goals & Visualizations
    - Tables
    - Graphs
5. Beyond our R
6. Next Steps

1.1.3 Background

This project is motivated by both personal interest and academic development. As a fan of *Blue Lock*, a manga that emphasizes performance metrics and psychological attributes of football players, I wanted to explore how data science can be applied even in unconventional data sets like fictional sports statistics. At the same time, this analysis will be an opportunity to improve my skills in data cleaning, transformation, and visualization using R. The *Blue Lock* manga, written by Kaneshiro Muneyuki and illustrated by Nomura Yusuke, quantifies character skills in areas such as Speed, Dribbling, and Shooting. These stats can be analyzed like real performance data.

Through this project, I hope to connect entertainment and analysis, asking real data-driven question of fictional athletes, and to present my findings through visualizations and statistical analysis.

1.1.4 The Data

My data set for this project consists of performance statistics from *Blue Lock,* a sports manga series centered on elite football training. Each character is evaluated on six attributes - Speed, Defense, Pass, Dribble, Shoot, and Offense – along with an overall Total score. These ratings were extracted from official character charts published in the manga. It was imported into R using the read_csv() function from the readr package, with further manipulation done using dplyr and tidyr.

Here are the summary statistics of all variables before the cleaning.

```r
summary(stats)
```

```
##      Name               Speed          Defense           Pass      
##  Length:31          Min.   : 0.00   Min.   : 0.00   Min.   : 0.00  
##  Class :character   1st Qu.: 0.00   1st Qu.: 0.00   1st Qu.: 0.00  
##  Mode  :character   Median :78.00   Median :71.00   Median :77.00  
##                     Mean   :53.74   Mean   :50.94   Mean   :52.65  
##                     3rd Qu.:83.50   3rd Qu.:76.50   3rd Qu.:84.00  
##                     Max.   :98.00   Max.   :99.00   Max.   :97.00  
##     Dribble          Shoot         Offense          Total      
##  Min.   : 0.00   Min.   : 0.0   Min.   : 0.00   Min.   : 0.00  
##  1st Qu.: 0.00   1st Qu.: 0.0   1st Qu.: 0.00   1st Qu.: 0.00  
##  Median :76.00   Median :79.0   Median :82.00   Median :82.00  
##  Mean   :53.48   Mean   :54.1   Mean   :56.19   Mean   :56.35  
##  3rd Qu.:86.00   3rd Qu.:84.5   3rd Qu.:89.50   3rd Qu.:89.00  
##  Max.   :95.00   Max.   :98.0   Max.   :96.00   Max.   :98.00
```

1.1.4.1 Cleaning Goals

My cleaning process focused on ensuring the data set accurately reflected the player statistics shown in the manga while making the structure ready for analysis. Below is the data set and the summary table after cleaning and creating as well as a breakdown of what worked and what didn't.

```r
skills = c('Speed', 'Defense', 'Pass', 'Dribble', 'Shoot', 'Offense')

stats_clean = stats %>% filter(rowSums(select(., all_of(skills)), na.rm = TRUE) > 0)

stats_clean = stats_clean %>% rename(Player_Score = Total)

stats_clean = stats_clean %>% mutate(Average_Score = rowMeans(select(., all_of(skills)), na.rm = TRUE))

stats_clean
```

```
## # A tibble: 20 x 9
##    Name     Speed Defense  Pass Dribble Shoot Offense Player_Score Average_Score
##    <chr>    <dbl>   <dbl> <dbl>   <dbl> <dbl>   <dbl>        <dbl>         <dbl>
##  1 Isagi Y~    77      75    78      70    82      94           88          79.3
##  2 Bachira~    84      68    88      95    82      86           86          83.8
##  3 Kunigam~    85      75    65      71    95      94           91          80.8
##  4 Chigiri~    98      68    80      93    88      95           90          87  
##  5 Gagamar~    87      93    88      79    80      84           87          85.2
##  6 Raichi ~    77      90    81      79    82      82           81          81.8
##  7 Shoei B~    80      68    61      88    96      95           92          81.3
##  8 Ikki Ni~    78      84    88      76    67      82           82          79.2
##  9 Nagi Se~    82      75    85      86    92      91           92          85.2
```

```
## 10 Mikage ~     81       86     89       83     82       86              85              84.5
## 11 Aryu Jy~     74       90     68       69     79       76              80              76
## 12 Yukimiy~     89       76     77       91     86       88              84              84.5
## 13 Kurona ~     95       68     83       84     72       86              83              81.3
## 14 Kiyora ~     82       76     80       83     88       81              80              81.7
## 15 Aiku Ol~     81       93     83       76     77       72              85              80.3
## 16 Sendo S~     82       73     76       77     83       84              81              79.2
## 17 Michael~     91       74     81       86     98       96              98              87.7
## 18 Alexis ~     76       71     97       90     78       94              93              84.3
## 19 Don Lor~     83       99     87       93     82       84              96              88
## 20 Charles~     84       77     97       89     88       92              93              87.8
```

```r
summary(stats_clean)
```

```
##      Name               Speed          Defense           Pass      
##  Length:20          Min.   :74.0   Min.   :68.00   Min.   :61.00  
##  Class :character   1st Qu.:79.5   1st Qu.:72.50   1st Qu.:77.75  
##  Mode  :character   Median :82.0   Median :75.50   Median :82.00  
##                     Mean   :83.3   Mean   :78.95   Mean   :81.60  
##                     3rd Qu.:85.5   3rd Qu.:87.00   3rd Qu.:88.00  
##                     Max.   :98.0   Max.   :99.00   Max.   :97.00  
##     Dribble          Shoot          Offense      Player_Score   Average_Score
##  Min.   :69.00   Min.   :67.00   Min.   :72.0   Min.   :80.00   Min.   :76.00  
##  1st Qu.:76.75   1st Qu.:79.75   1st Qu.:83.5   1st Qu.:82.75   1st Qu.:80.71  
##  Median :83.50   Median :82.00   Median :86.0   Median :86.50   Median :82.83  
##  Mean   :82.90   Mean   :83.85   Mean   :87.1   Mean   :87.35   Mean   :82.95  
##  3rd Qu.:89.25   3rd Qu.:88.00   3rd Qu.:94.0   3rd Qu.:92.00   3rd Qu.:85.17  
##  Max.   :95.00   Max.   :98.00   Max.   :96.0   Max.   :98.00   Max.   :88.00
```

1.1.4.2 What Worked from the Midterm Plans

- **Removing irrelevant rows**: I was able to filter out any rows where all individual skill values were 0 or missing using a combination of rowSums() and filter() from the dplyr package. These rows did not reflect actual performance and would have skewed later analysis.
- **Renaming columns:** The 'Total' column was renamed to 'Player_Score' using the rename() function to better reflect its purpose. The original name was ambiguous and could be misinterpreted as a sum or average.
- **Creating the 'Average_Score' column:** I used mutate() and rowMeans() to generate a new column that represents each player's average score across the six performance attributes. This helped standardize performance and provided a way to compare with the Player Score.

1.1.4.3 What Didn't Worked from the Midterm Plans

- **Dependent column logic:** I originally was going to compute the derived columns earlier in the cleaning stage. However, I learned that creating the Average_Score column before filtering out the zero-value rows led to averages that were not accurate. I had to restructure the order in which I did things, making sure I filter before making new conclusions.

1.1.5 Research Questions

The purpose of this project is to investigate patterns and insights in the performance statistics of Blue Lock characters using a structured data analysis approach. Below are the guiding research questions that inform the visualizations and statistical summaries presented later in the report:

1. **Which players demonstrate the highest individual skill scores, and are they consistent top performers overall?**

By identifying the top scorers for each skill, I aim to assess whether certain players dominate specific attributes or are more well-rounded across the board.

2. **How closely does a player's average skill score align with their overall Player Score?**

The Player_Score column reflects the manga's overall evaluation of a character. Comparing this to the calculated Average_Score allows us to see whether external factors — such as in-game leadership, ego, or narrative importance — influence their total score beyond skill.

3. **Do the characters show tendencies towards specific roles (ex. offense-heavy or well-rounded)?**

Given that *Blue Lock* emphasizes offensive power and striker mentality, this question explores whether the character stats reflect a bias toward attributes like Shooting and Offense or whether there's a balance across all six skills.

4. **What does the overall distribution of skills look like across all characters - are specific attributes more emphasized than other ones?**

This question aims to identify the spread and central tendencies of each attribute. It also helps determine whether some skills (e.g., Pass or Defense) are underemphasized across the cast, or whether there's a consistent baseline for most players.

1.1.6 Analysis Goals & Visualizations

After cleaning the data, my goal is to explore relationships and trends in the performance of *Blue Lock* characters through both summary statistics and visualizations.

1.1.6.1 Tables To support the following visualizations, I've included a summary table of all skill-related variables.

This table was also included in the *Cleaning Goals* section, but it is shown again here for easier reference alongside the graphs.

```
summary(stats_clean)
```

```
##      Name               Speed         Defense          Pass      
##  Length:20          Min.   :74.0   Min.   :68.00   Min.   :61.00  
##  Class :character   1st Qu.:79.5   1st Qu.:72.50   1st Qu.:77.75  
##  Mode  :character   Median :82.0   Median :75.50   Median :82.00  
##                     Mean   :83.3   Mean   :78.95   Mean   :81.60  
```

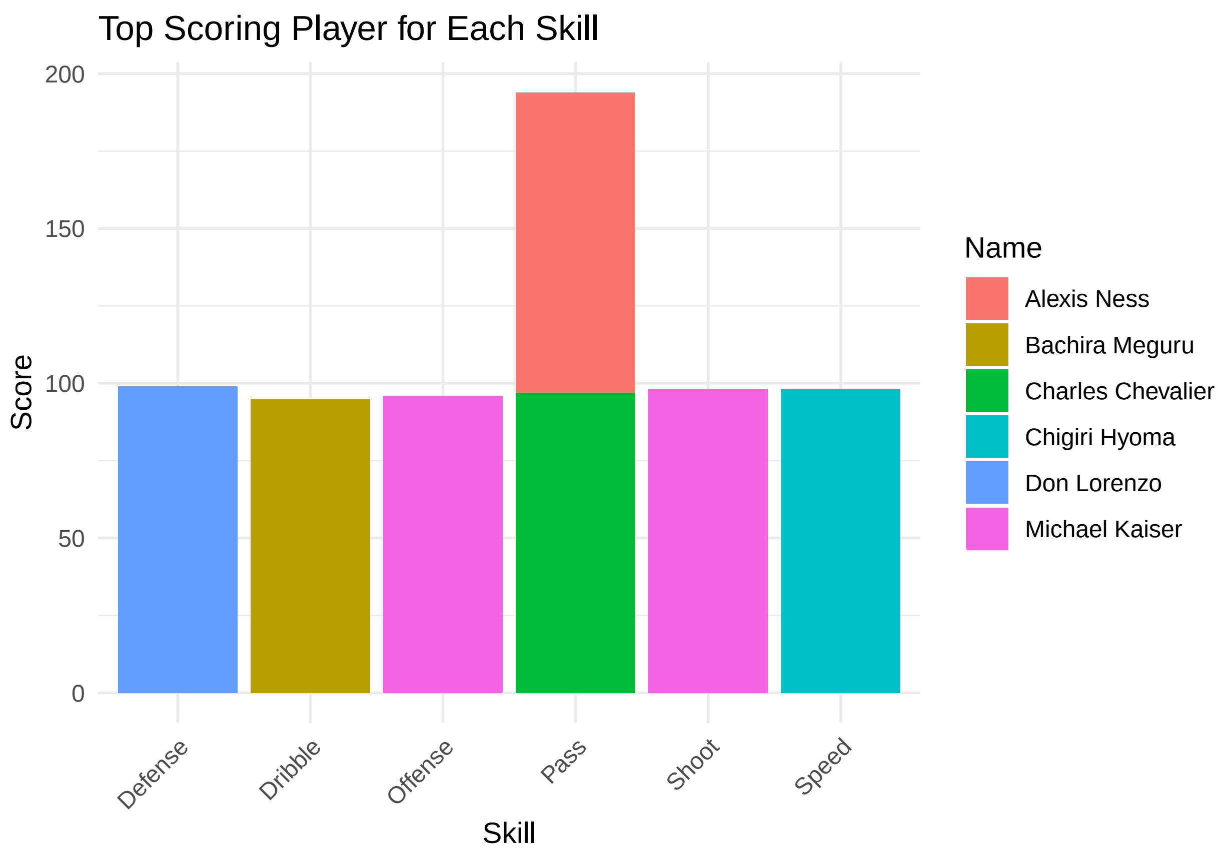
```
##                       3rd Qu.:85.5   3rd Qu.:87.00   3rd Qu.:88.00
##                       Max.   :98.0   Max.   :99.00   Max.   :97.00
##     Dribble          Shoot          Offense      Player_Score   Average_Score
##  Min.   :69.00   Min.   :67.00   Min.   :72.0   Min.   :80.00   Min.   :76.00
##  1st Qu.:76.75   1st Qu.:79.75   1st Qu.:83.5   1st Qu.:82.75   1st Qu.:80.71
##  Median :83.50   Median :82.00   Median :86.0   Median :86.50   Median :82.83
##  Mean   :82.90   Mean   :83.85   Mean   :87.1   Mean   :87.35   Mean   :82.95
##  3rd Qu.:89.25   3rd Qu.:88.00   3rd Qu.:94.0   3rd Qu.:92.00   3rd Qu.:85.17
##  Max.   :95.00   Max.   :98.00   Max.   :96.0   Max.   :98.00   Max.   :88.00
```

This table shows the distribution (min, median, mean, max) of each attribute across all players, providing key context for interpreting outliers and identifying general trends in the data.

1.1.6.2 Graphs Each of the following graphs supports one of the research questions mentioned prior.

1. **Bar Plot: Highest Individual Stat Holders** This bar chart shows which player holds the highest score for each of the six skill attributes.

```r
top_stats <- stats_clean %>%
  pivot_longer(cols = c(Speed, Defense, Pass, Dribble, Shoot, Offense), names_to = "Skill", values_to =
  group_by(Skill) %>%
  slice_max(order_by = Score, n = 1)

ggplot(top_stats, aes(x = Skill, y = Score, fill = Name)) +
  geom_col() +
  labs(title = "Top Scoring Player for Each Skill", x = "Skill", y = "Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```r
ggsave("top_player_plot.png", width = 8, height = 6, dpi = 300)
```

This chart identifies the top-performing character for each skill. The results show that no single character dominates across all attributes, suggesting specialization rather than all-around performance. This supports the idea that some characters are designed to excel in specific roles (e.g., a shooter vs. a defender), rather than being universally strong across the board.

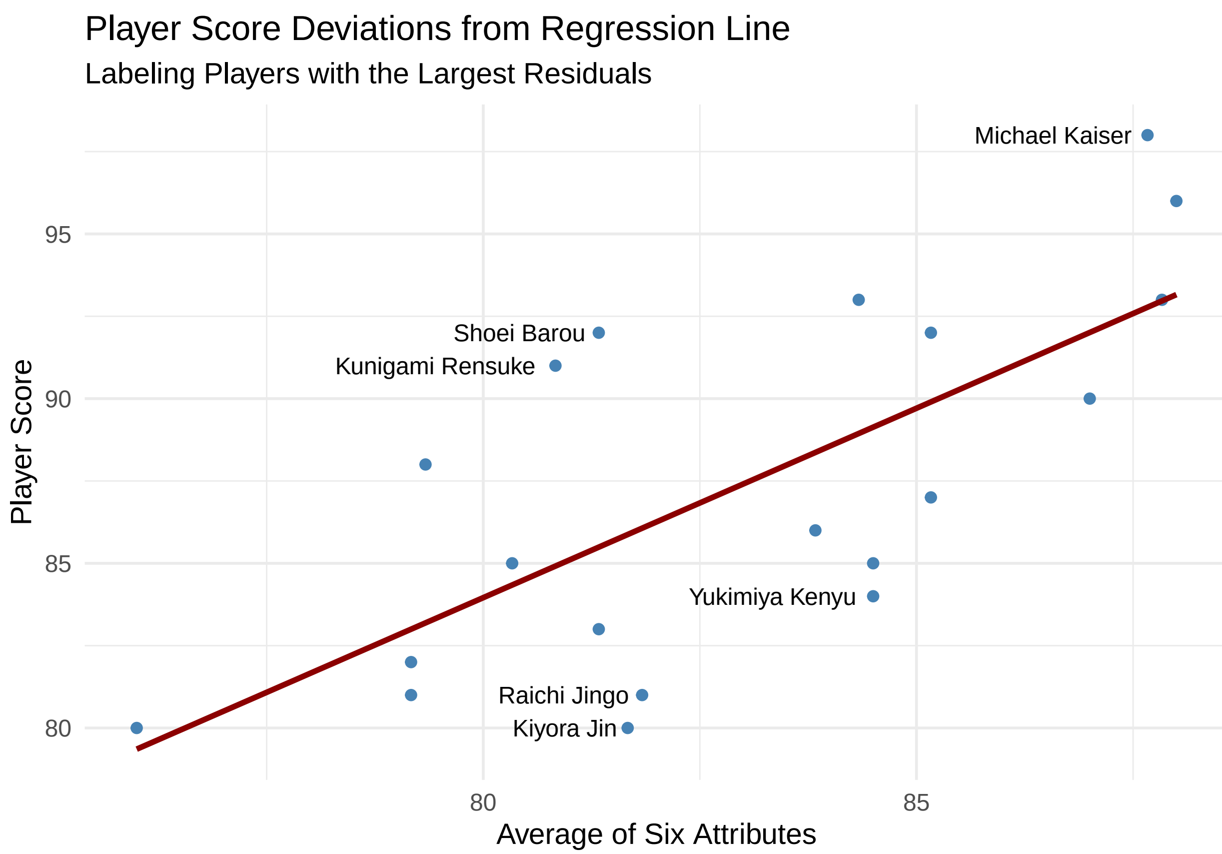
2. **Scatter Plot: Average Score vs. Player Score** This plot visualizes the relationship between a player's calculated Average_Score and their overall Player_Score, highlighting the top players with the largest gaps between their average score and player score.

```r
model <- lm(Player_Score ~ Average_Score, data = stats_clean)

stats_clean <- stats_clean %>%
  mutate(
    predicted_score = predict(model),
    residual = Player_Score - predicted_score,
    abs_residual = abs(residual)
  )

stats_clean %>%
  arrange(desc(abs_residual)) %>%
  select(Name, Average_Score, Player_Score, predicted_score, residual, abs_residual) %>%
  head(10)
```

```
## # A tibble: 10 x 6
##    Name          Average_Score Player_Score predicted_score residual abs_residual
##    <chr>                 <dbl>        <dbl>           <dbl>    <dbl>        <dbl>
##  1 Shoei Barou            81.3           92            85.5     6.51         6.51
##  2 Kunigami Re~           80.8           91            84.9     6.08         6.08
##  3 Kiyora Jin             81.7           80            85.9    -5.87         5.87
##  4 Michael Kai~           87.7           98            92.8     5.23         5.23
##  5 Yukimiya Ke~           84.5           84            89.1    -5.13         5.13
##  6 Raichi Jingo           81.8           81            86.1    -5.07         5.07
##  7 Isagi Yoichi           79.3           88            83.2     4.81         4.81
##  8 Mikage Reo             84.5           85            89.1    -4.13         4.13
##  9 Alexis Ness            84.3           93            88.9     4.06         4.06
## 10 Gagamaru Gin           85.2           87            89.9    -2.90         2.90
```

```r
ggplot(stats_clean, aes(x = Average_Score, y = Player_Score)) +
  geom_point(color = "steelblue") +
  geom_smooth(method = "lm", se = FALSE, color = "darkred") +
  geom_text(
    data = filter(stats_clean, abs_residual > 5),
    aes(label = Name),
    size = 3, hjust = 1.1, vjust = 0.5
  ) +
  labs(
    title = "Player Score Deviations from Regression Line",
    subtitle = "Labeling Players with the Largest Residuals",
    x = "Average of Six Attributes",
    y = "Player Score"
  ) +
  theme_minimal()
```

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




```
ggsave("player_score_deviation.png", width = 8, height = 6, dpi = 300)
```

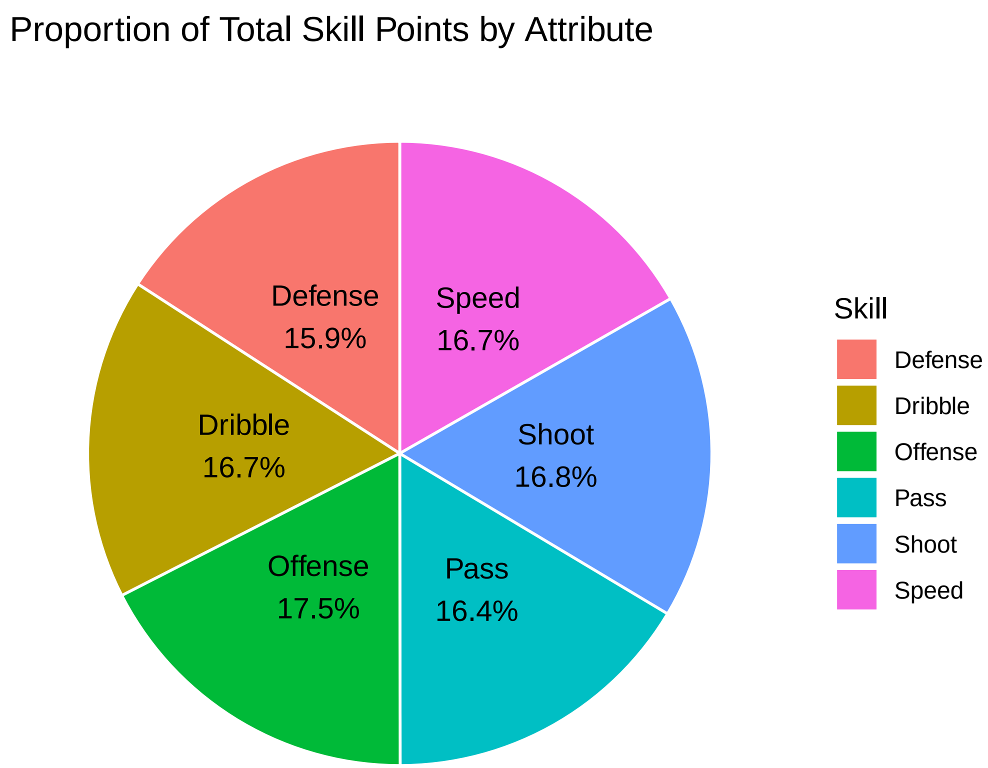
```
## `geom_smooth()` using formula = 'y ~ x'
```

There is a general positive trend between average skill and total player score, but some players appear as outliers, having a higher or lower Player Score than their average would suggest. This discrepancy may reflect the influence of narrative or intangible qualities (such as "ego" or leadership potential) on the overall Player Score — elements emphasized heavily in *Blue Lock*.

3. **Pie Chart: Skill Totals Across All Players** This chart shows the aggregate skill values for each attribute across all the players.

```
skill_totals <- stats_clean %>%
  summarise(across(c(Speed, Defense, Pass, Dribble, Shoot, Offense), \(x) sum(x, na.rm = TRUE))) %>%
  pivot_longer(cols = everything(), names_to = "Skill", values_to = "Total") %>%
  mutate(Percent = Total / sum(Total) * 100,
         Label = paste0(Skill, "\n", round(Percent, 1), "%"))

ggplot(skill_totals, aes(x = "", y = Total, fill = Skill)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y") +
  geom_text(aes(label = Label), position = position_stack(vjust = 0.5)) +
  labs(title = "Proportion of Total Skill Points by Attribute") +
  theme_minimal() +
```

```
  theme(axis.title = element_blank(),
        axis.text = element_blank(),
        panel.grid = element_blank())
```

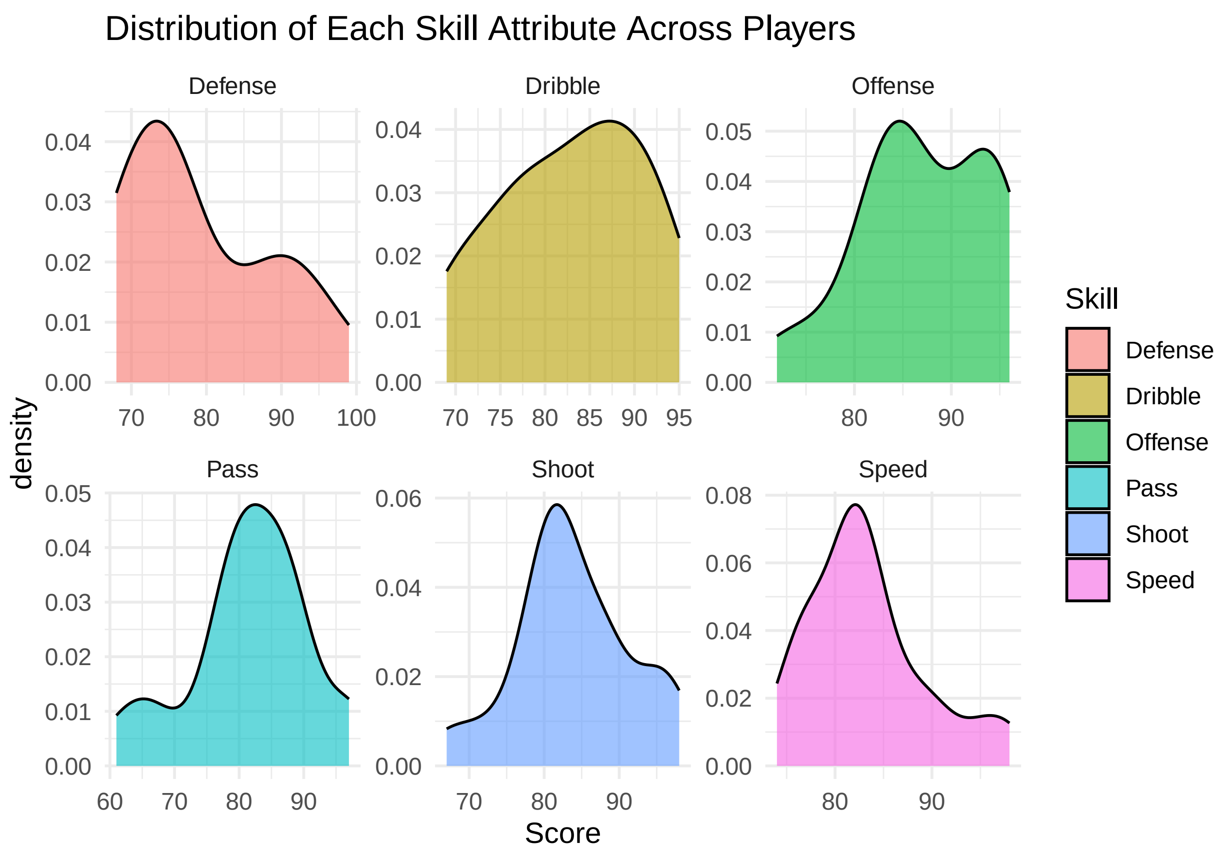



```
ggsave("total_skill_pie.png", width = 8, height = 6, dpi = 300)
```

The pie chart reveals that the skill attributes are fairly evenly distributed across players, indicating a generally balanced skill set in the cast. However, Offense stands out slightly, accounting for approximately 17.5% of the total skill points, reflecting the *Blue Lock* manga's emphasis on offensive capabilities and striker-focused play.

4. **Density Plots: Distribution of Each Attribute** This series of density plots helps visualize how each attribute is distributed across players.

```
stats_clean %>%
  pivot_longer(cols = c(Speed, Defense, Pass, Dribble, Shoot, Offense), names_to = "Skill", values_to =
  ggplot(aes(x = Score, fill = Skill)) +
  geom_density(alpha = 0.6) +
  facet_wrap(~ Skill, scales = "free") +
  labs(title = "Distribution of Each Skill Attribute Across Players") +
  theme_minimal()
```




```r
ggsave("each_skill_distribution.png", width = 8, height = 6, dpi = 300)
```

The distribution plots reveal that some attributes — like Shoot and Speed — are more tightly clustered, while others — like Dribble and Defense — have broader spreads with higher outliers. This supports the earlier insight that the data set favors attacking capabilities, with a few characters being extremely specialized or exceptional in those areas.

1.1.7 Beyond our R

With some background and the help functions in R, I used functions beyond the scope of our class. While most of my analysis builds on the core functions introduced in class, I also incorporated a few additional functions to improve data visualization and manipulation. These include:

- **pivot_longer()** – to reshape wide-format data into long format for faceted plots and group-wise summaries.
- **slice_max()** – to extract the top-performing player(s) for each skill.
- **geom_col()** – for creating bar plots where height is directly mapped to a variable (used for skill leaders).
- **summarise() + across()** – used with anonymous functions ((x)) for calculating totals across multiple columns.
- **geom_bar() + coord_polar()** – to create pie charts from bar data.

- **geom_density() + facet_wrap()** – to visualize the distribution of each skill across all players in separate plots.

These additions helped me produce cleaner, more understandable visualizations that aligned with my research questions.

1.1.8 Next Steps

This project successfully cleaned and explored the *Blue Lock* player data set, created multiple visualizations to support my research questions, and identified several interesting trends — such as the relatively balanced skill distribution and the slight bias toward offense. I was able to successfully fulfill my Midterm goals, but considering to expand this research, I was thinking about how I could continue.

For future work, here are the next steps I would consider:

- **Expand the Data set:** Include more characters or updated stats as the manga progresses. New chapters may introduce revised ratings or new players, which could provide richer analysis and allow for time-series comparisons.
- **Add Narrative-Based Metrics**: Introduce qualitative or subjective attributes such as "ego," "leadership," or "flow state" — core themes in *Blue Lock* — to explore how intangible factors align (or not) with player stats.
- **Compare Subgroups:** Group players by team, arc, or rank (e.g., top 10, bottom 10) to compare skill distributions across narrative roles or tournament stages.