

Six Nations Fantasy Rugby Visualisation and Analysis



Background

The six nations is an annual international rugby union competition between **England, France, Ireland, Italy, Scotland** and **Wales**. There are five rounds to ensure that all teams play each other. Fantasy rugby is an interactive game that fans can play alongside the competition. Players pick their own team from all the squads before each round kicks off and earn points based on the performance of their squad in the match.

Players can pick a maximum of 18 people for their team which includes 15 starting players and 3 substitutions. All the positions in the team must be filled as normal and a maximum of 4 players can be picked from each nation. They also must pick a starting player as **Captain**, which doubles that individual's points, and a bench player as **Supersub**, which triples that person's points.

The players score points during the match for things like tries, conversions, meters carried, tackles made etc.

Data Origins

The data sets were obtained from Fantasy Six Nations Rugby (<https://fantasy.sixnationsrugby.com>) and contains the statistics gathered from the 2023 Six Nations Championship. The data is organised per player per team. Most people fill in the highest scoring positions, such as back row and back 3, with the top performing nations players (at the moment- Ireland). It's therefore important to be able to compare the top performing players from the different nations since you can only have a maximum of 4 players per nation on your fantasy squad.

Shown below is the raw data set for Ireland which shows the categories collected.

```
#Upload raw data sets
Ireland <- read.csv(here("data", "raw_data", "Ireland.csv"))

#Print head of data frame
if (knitr::is_html_output()) {
  head(Ireland)          #If html output then print header with df_print: paged applied in YAML header
} else {
  head(Ireland[,1:5], 5) #If not html output (eg pdf) print header with first 5 columns
}
```

```
##           X Number.of.matches Man.of.the.match Tackles Dominant.tackles
## 1      J. Lowe                5                NA        13                5
```

## 2	M. Hansen	5	2	19	2
## 3	H. Keenan	5	1	15	2
## 4	J. Ryan	5	NA	59	6
## 5	D. Sheehan	4	1	18	4

Project Aim

It is important to know the rankings in teams in terms of fantasy points as well as their actual performance in real life in order to pick the best squad possible. This project will visualise the top 3 scoring players in terms of fantasy points from each country. This will enable players to decide which countries to prioritise in their 18 available positions and the variety in how the top scoring players earn their points.

Project Organisation

The /data folder contains the processed data frames as .csv files as well as a sub folder /raw_data which contains the original raw data sets. The /plots folder contains the final visualisation saved as a .png with fixed dimensions. There is also a /font and /logo folder which contains the official branding font and logo for the Guinness Six Nations.

A codebook has also been included which describes the different statistics and data frames used within this project.

Data Preparation

Loading Packages

The first step in data preparation was to import the packages used in this project.

```
#Import packages
library(tidyverse)
library(here)
library(showtext)
```

Summarising and combining data

The raw data consisted of six separate csv. files, one per nation. For easier data manipulation these data frames were combined into one data frame.

Since the original data sets contained images of the players coloured shirts to indicate their country, a separate column was added to show this instead.

```
#Create a list of the data sets
list_nations <- c("England", "France", "Ireland", "Italy", "Scotland", "Wales")

#Loop through each data set
for (nation in list_nations) {
  dataset <- get(nation)
  country <- gsub("[[:digit:]]", "", nation)
  dataset$Country <- country
  assign(nation, dataset)
}
```

#Select each data frame from the list
#Extract country from data set name
#Add new column for country
#Assign changes back to original list

The next step was to filter the data frames to only contain the top 3 scoring players.

```
#List players to keep from each country
players_to_keep <- list(
  England = c("J. Willis", "L. Ludlam", "M. Malins"),
  France = c("T. Ramos", "D. Penaud", "T. Flament"),
  Ireland = c("J. Lowe", "M. Hansen", "H. Keenan"),
  Italy = c("A. Capuozzo", "T. Allan", "D. Fischetti"),
  Scotland = c("F. Russell", "D. van der Merwe", "H. Jones"),
  Wales = c("R. Dyer", "J. Tipuric", "L. Williams")
)

#Create empty list to store filtered data sets
filtered_datasets <- list()

#Loop through each nation and filter the players then save in the new list
for(nation in list_nations) {
  dataset <- get(nation) #Retrieve each data frames from the list
  filtered_nation <- dataset %>% #Assign filtered version to new data frames
  filter(X %in% players_to_keep[[nation]]) #Apply filter
  filtered_datasets[[nation]] <- filtered_nation #Assign filtered data frames to new list
}
```

This list was then combined into one data frame.

```
#Combine list into one data frame
combined_nations <- bind_rows(filtered_datasets)
```

Columns were then renamed for ease of manipulation and readability. This was then saved as a new .csv file.

```
#Rename columns
combined_nations <- combined_nations %>%
  rename(PlayerName= "X", N= Number.of.matches, MOTM= Man.of.the.match, TM=Tackles,
    DM=Dominant.tackles, LB=Line.breaks, MC=Meters.carried, FT=X50.22,
    LS=Line.out.steals, BS=Breakdown.steal, T=Try, TA=Assist, C=Conversion, P=Penalty,
    DG=Drop.goal, YC=Yellow.card, RC=Red.card, AP=Average.points)

#Save new data set in folder /data within directory
combined_nations_data <- paste(here("data"), "2023_combined_nations.csv", sep = "/") #Name .csv file
write.csv(combined_nations, file= combined_nations_data) #Save .csv file
```

Shown below are the top 5 rows of the combined data frame.

```
head(combined_nations)
```

##	PlayerName	N	MOTM	TM	DM	LB	MC	FT	LS	BS	T	TA	C	P	DG	YC	RC	AP	Country
## 1	J. Willis	4	NA	54	5	1	76	NA	NA	3	1	NA	NA	NA	NA	1	NA	38.4	England
## 2	L. Ludlam	5	NA	60	7	1	193	NA	NA	2	NA	1	NA	NA	NA	NA	NA	38.3	England
## 3	M. Malins	4	NA	31	2	2	231	NA	NA	NA	2	NA	NA	NA	NA	NA	NA	37.8	England
## 4	T. Ramos	5	1	6	2	5	502	NA	NA	NA	3	2	18	10	1	NA	NA	70.9	France
## 5	D. Penaud	5	NA	21	2	7	397	NA	NA	NA	5	1	NA	NA	NA	NA	NA	55.5	France
## 6	T. Flament	5	NA	77	4	3	165	NA	1	NA	3	NA	NA	NA	NA	NA	NA	48.6	France

Manipulation

Finally, the data had to be manipulated to represent the points scored in fantasy rugby, not just the raw value of each category. For example, only 0.2 points are scored per meter carried, whereas 15 points are scored per try scored. Each statistic was multiplied by the number of points scored in fantasy rugby to reflect this. These scores were then divided by the number of games each player played in order to show their average score per game, reflective of the whole competition.

```
#Multiply statistics by the points they earn in fantasy rugby
fantasy_data <- combined_nations %>%
  mutate(MOTM=MOTM*15, DM=DM*7, LB=LB*7, MC=MC*0.2, FT=FT*10, LS=LS*7, BS=BS*7, T=T*15,
         TA=TA*7, C=C*2, P=P*5, DG=DG*7, YC=YC*-5, RC=RC*-10)

#Divide the statistics by the number of games from each player
fantasy_data[, 3:18] <- fantasy_data[, 3:18] / fantasy_data[, 2] #Divide values in columns 3-18 by va
```

Statistics that aren't used to calculate fantasy points were excluded. This new data set was then saved as a separate .csv file.

```
#Remove columns for average points and number of games played
fantasy_data <- fantasy_data[, -c(which(names(fantasy_data)=="AP"),
                                which(names(fantasy_data)=="N"))]

#Save new data set
final_fantasy_data <- paste(here("data"), "2023_fantasy_data.csv", sep = "/") #Name .csv file
write.csv(fantasy_data, file= final_fantasy_data) #Save .csv file
```

Shown below are the first 5 rows of the final fantasy data frame.

```
options(digits = 3) #Only show max 2 decimal places

head(fantasy_data)
```

```
## PlayerName MOTM TM DM LB MC FT LS BS T TA C P DG YC
## 1 J. Willis NA 13.50 8.75 1.75 3.80 NA NA 5.25 3.75 NA NA NA NA -1.25
## 2 L. Ludlam NA 12.00 9.80 1.40 7.72 NA NA 2.80 NA 1.4 NA NA NA NA
## 3 M. Malins NA 7.75 3.50 3.50 11.55 NA NA NA 7.50 NA NA NA NA NA
## 4 T. Ramos 3 1.20 2.80 7.00 20.08 NA NA NA 9.00 2.8 7.2 10 1.4 NA
## 5 D. Penaud NA 4.20 2.80 9.80 15.88 NA NA NA 15.00 1.4 NA NA NA NA
## 6 T. Flament NA 15.40 5.60 4.20 6.60 NA 1.4 NA 9.00 NA NA NA NA NA
## RC Country
## 1 NA England
## 2 NA England
## 3 NA England
## 4 NA France
## 5 NA France
## 6 NA France
```

Visualisaion

Six Nations Colours

Colours for the bar graph were chosen from the countries official team colours found on Encycolorpedia.

```
#ffffff for England (white)
#3f48cc for France
#00845c for Ireland
#8f8140 for Italy
#483d8b for Scotland
#ff0000 for Wales
```

Custom Text

The official Guinness Six Nations font was then installed and used in the visualisation using the `showtext` package.

```
#Upload custom font
font_add("Gotham_book",
         regular= here("font", "Gotham_Book_Font.otf"))

#Load new font
showtext_auto()
```

Organising Data

First the data frame was converted from wide to long format to easily work with `ggplot()`. Then, variables were grouped to summarise different ways players could score points. They were grouped into:

- “Cards”: red cards and yellow cards,
- “Try/Posts”: any in game point scoring i.e. tries, conversions and drop goals,
- “Tackles”: tackles made and dominant tackles,
- “Meters”: meters carried,
- “Other”: everything else e.g. line breaks, man of the match award, line out steals etc.

The player names were then reorganised along the x axis to be grouped by country instead of the alphabetical default.

```
#STEP 1: TRANSFORM TO LONG/TIDY FORMAT
fantasy_long <- fantasy_data %>%
  pivot_longer(cols= -c(PlayerName, Country),      #Keep PlayerName and Country columns
               names_to = "Statistic",             #Name the title of statistics
               values_to="Points")                 #Name the value of statistics

#STEP 2: GROUP VARIABLES
fantasy_long$Group <-
  ifelse(fantasy_long$Statistic %in% c("T", "C", "DG"), "Try/Posts",      #Insert new column "Group"
         ifelse(fantasy_long$Statistic %in% c("YC", "RC"), "Cards",      #Seperate statistics into t
         ifelse(fantasy_long$Statistic %in% c("TM", "DT"), "Tackles",
         ifelse(fantasy_long$Statistic %in% c("MC"), "Meters",
               "Other")))) #Any other statistics saved

#STEP 3: REORDER PLAYERS
fantasy_long$PlayerName <-
  factor(fantasy_long$PlayerName, levels = unique(fantasy_long$PlayerName)) #Convert PlayerName to a f
```

Plot Graph

```
#STEP 4: VISUALISATION
#Plot graph
myplot <- ggplot(fantasy_long, aes(x= PlayerName, y=Points))+ #ggplot with "PlayerName" on x axis and
  geom_bar(stat="identity", aes(fill=Country))+ #Make bar graph and fill columns based on "Country" v
  scale_fill_manual(values=c("England"="white", "France"="#3f48cc", "Ireland"="#00845c", "Italy"="#8f81b1"))
  geom_bar(stat="identity", fill="black", aes(alpha=factor(Group, levels=c("Cards", "Other", "Meters", "Tackles", "Try/Posts"))))
  scale_alpha_manual(values=c("Cards"=0.3, "Try/Posts"=1, "Tackles"=0.4, "Meters"=0.2, "Other"=0), #S
                    name="Groups", #Name legend title
                    breaks=c("Cards", "Try/Posts", "Tackles", "Meters", "Other"))+ #Reorder "Group" var
xlab("PLAYER NAME")+ #Name x axis
ylab("POINTS")+ #Name y axis
labs(title="2023 SIX NATIONS CHAMPIONSHIP STATISTICS", subtitle="TOP 3 PLAYERS PER COUNTRY")+ #Add ti
theme(text = element_text(family = "Gotham_book"), #Apply custom font
      plot.title=element_text(size=40),
      plot.subtitle=element_text(size=30),
      axis.text.x=element_text(size=15, angle=45, vjust=0.5, hjust=0.5, margin=margin(t=-0.4, unit="1.5cm")),
      axis.text.y=element_text(size=18),
      axis.title.x=element_text(margin=margin(t=10), size=25), #Move x axis title so doesn't overlap
      axis.title.y=element_text(size=25),
      legend.key.size=unit(1, "lines"),
      legend.title=element_text(size=30),
      legend.text=element_text(size=20),
      legend.margin=margin(-5,0,-2,0))
```

Save and Print Graph

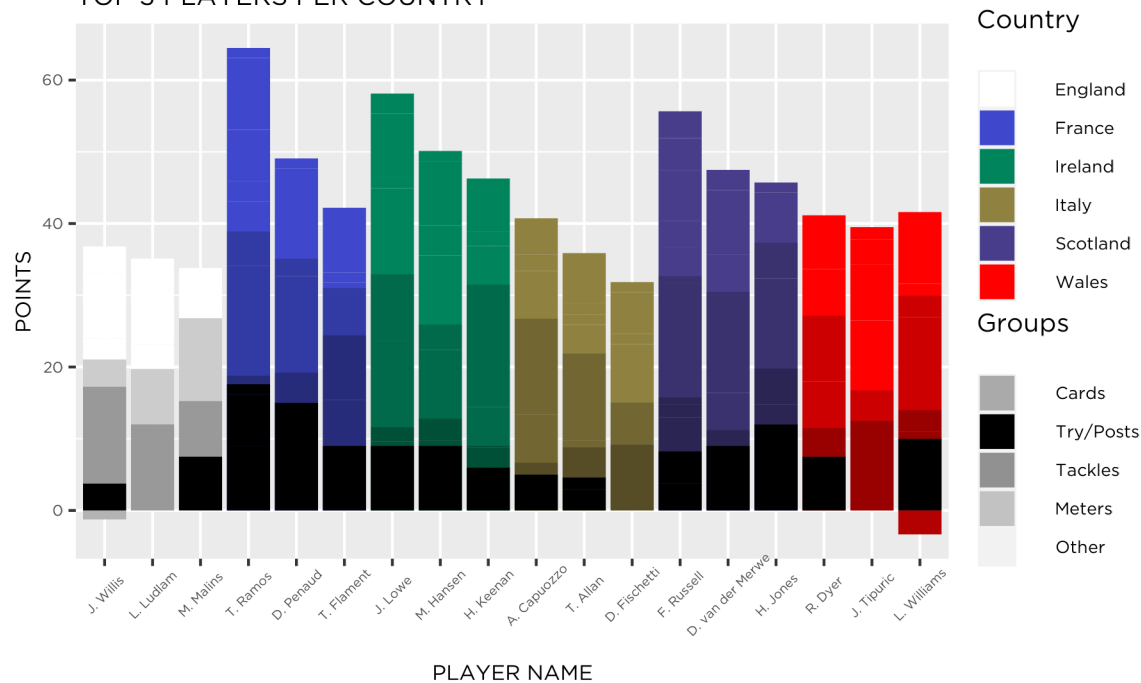
The graph was then saved to the /plots folder within the directory. It was saved with fixed size so it won't change dimensions when printed in a different format.

```
#Save plot as .png and specify fixed dimensions
ggsave(here("plots", "myplot.png"), width=6, height=4)

#Upload plot .png and insert graphic
final_plot <- here("plots", "myplot.png")
knitr::include_graphics(final_plot)
```

2023 SIX NATIONS CHAMPIONSHIP STATISTICS

TOP 3 PLAYERS PER COUNTRY



Design Choices

Due to the different categories of statistics, a stacked bar graph was the best way to visualise these groups. This automatically placed the negative groups (ie red and yellow cards) underneath the x axis instead of being deducted from the total which occurred when the statistics weren't stacked. It was also possible to place the "Try/Posts" category closest to the x axis which enables viewers to also compare in-game points scored by the players.

Each group was set to a different transparency to highlight the differences in these groups. The "Try/Posts" group was designed to be fully black since this is the group people are usually most concerned with when picking fantasy players.

Interpretation

There are a few interesting things to note from this graph. First is that there isn't a strong correlation between players who scored highly in "Try/Posts" category and players scoring most fantasy points overall. This is best reflected by J. Tipuric for Wales who scored no in-game points but still managed to score better than all 3 of the top players for England. Secondly, whilst Ireland may have won the Six Nations Championship and achieved a Grand Slam, they didn't have the top scoring player. As well as this, England beat both Italy and Wales in the competition, but they scored the lowest in fantasy points. This is useful to know when picking a fantasy squad since competition performance is not necessarily reflected by fantasy points.

Conclusion

I had a lot of fun making this project felt like I really got to grips with data wrangling and making for loops. I significantly increased my confidence in data analysis and visualisation, even up to the point of helping other

people with queries about r. I have never coded before this project but found the process really enjoyable and satisfying. I would keep showing my friends and family updated pictures of my lines of code and graphs, even if I knew they had no idea what it meant.

Follow-ups

It is important to recognise that as well as the above categories, players also score points depending on which team won during the match. Teams get 8 points per winning player and 4 points per drawing player. Since this required separating each players statistics into the 5 rounds where they acquired them and then calculating the players that won or even played that roun, this felt too difficult for this visualisation.

After finishing this graph I think it might also be interesting to compare the position of the players since there is a clear correlation between position and points. For example, people always prioritise back row players in the forwards and back 3 players in the backs. Comparing all the flankers that played in the Six Nations might also be helpful in terms of player selection for fantasy squads.

References

- *Fantasy Six Nations* - Original raw data sets from <https://fantasy.sixnationsrugby.com>
- *Encyclopedic* - Official hex codes for branding colours for international rugby teams <https://encyclopedia.com/teams/rugby-union/international>
- *Official Championship Logo and Brand Guidelines* - <https://d2cx26qpfwuhvu.cloudfront.net/sixnatlg/wp-content/uploads/2019/08/16113233/Guinness-Six-Nations-Championship-Guidelines.pdf>