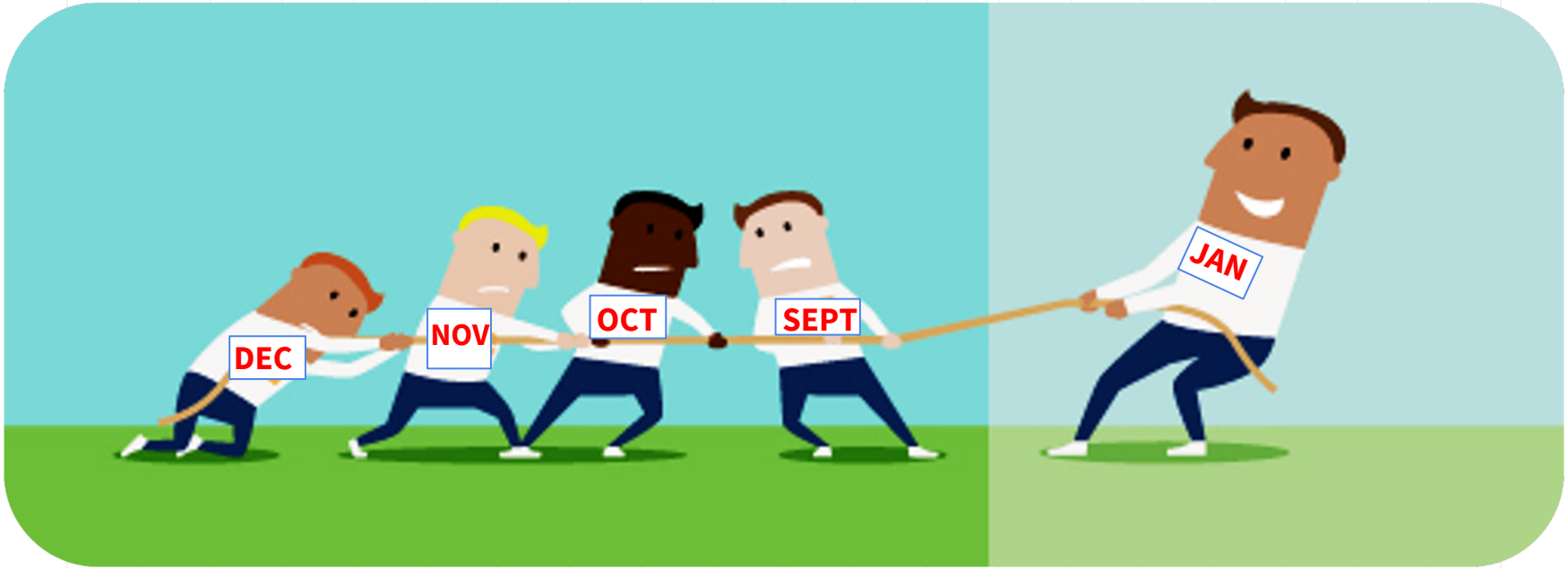



# THE IMPACT OF THE RELATIVE AGE EFFECT IN ELITE SOCCER

Roma Gandhi | Joanna Yang | Bryan Zhang

*Wharton Moneyball Academy 2019, Sports Analytics Student Research Journal*

# What is the Relative Age Effect?



January is over and February has begun. What may seem like a banal piece of information is anything but. It is alarming news for German boys who will be born this year: their chances of becoming a top-level football player have just been dramatically reduced.

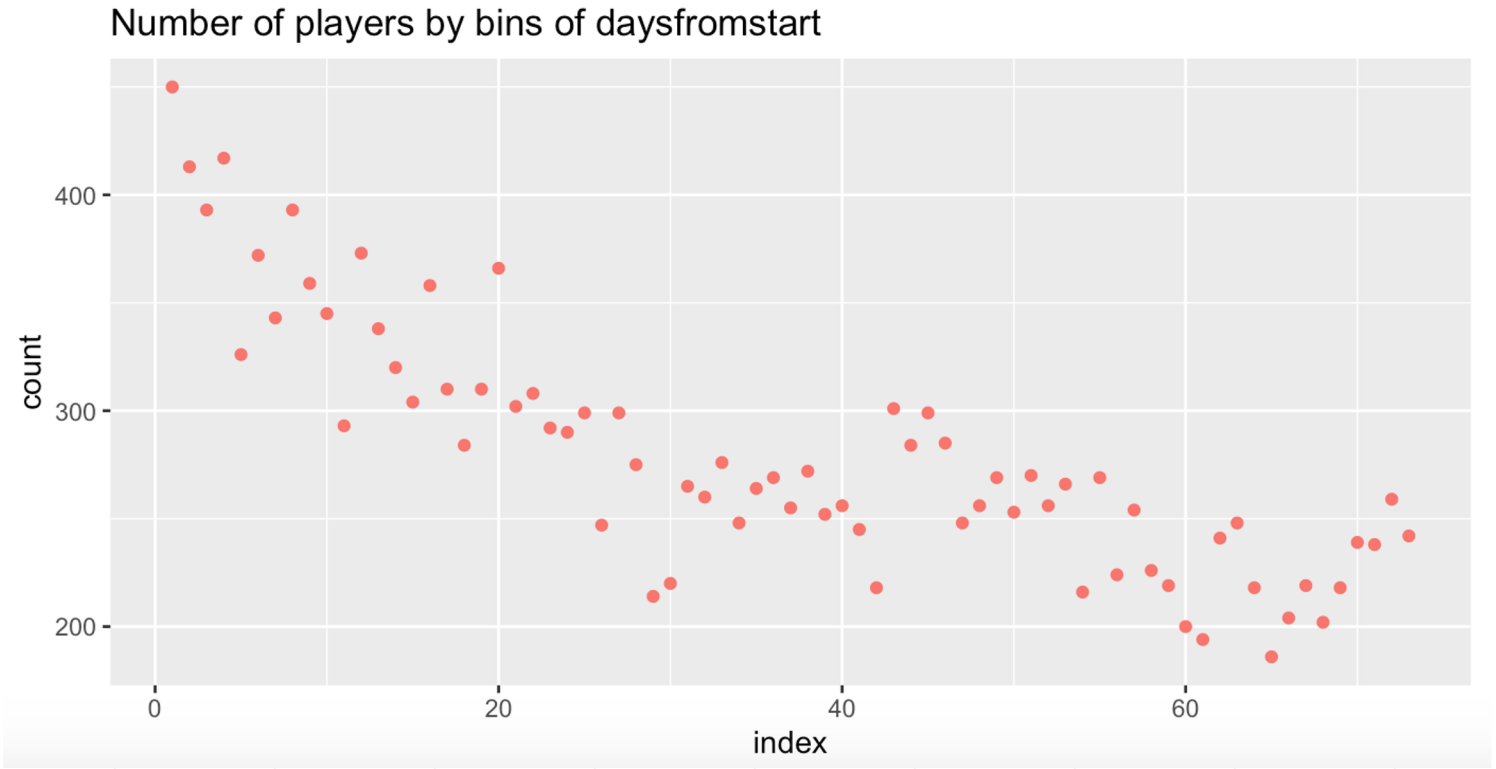
(ESPN)

Players with a greater relative age are more likely to be identified as "talented" because of the likely physical advantages they have over their "younger" peers. Some options for reducing the relative age effect are offered.

(Werner F Helsen, Jan van Winckel & A Mark Williams 2005)

Children born early in their school year enjoy a considerable advantage over those born later in the school year, simply by being up to 12 months older than later-borns in the same class. They do comparatively well in terms of academic attainment [1, 2], sport [3, 4], and in their emotional and social life [5, 6]. Such Relative Age Effects (RAEs) tend to persist longer than the natural ironing out of maturational advantages suggests they should. One possibility is psychological:

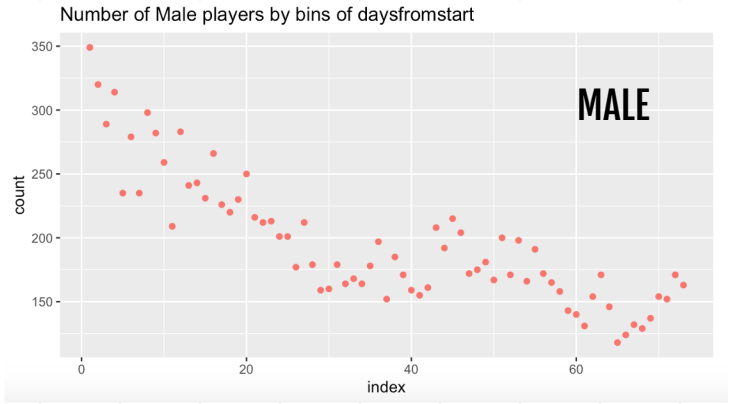
(Doyle, Bottomley 2018)



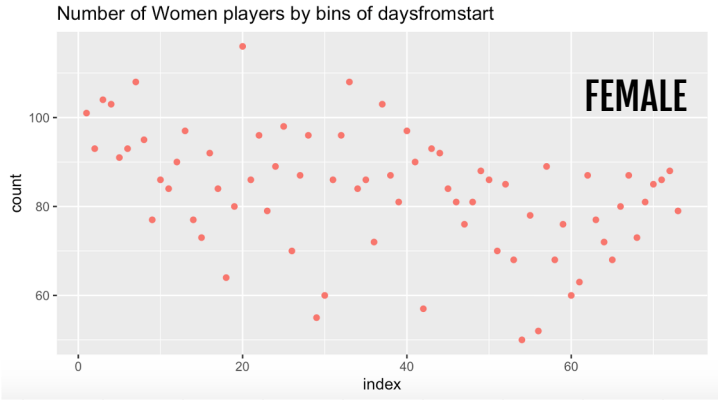

Correlation (R-value): -0.829

***Index*** *= number of bins (each bin = 5 days from start of season)*

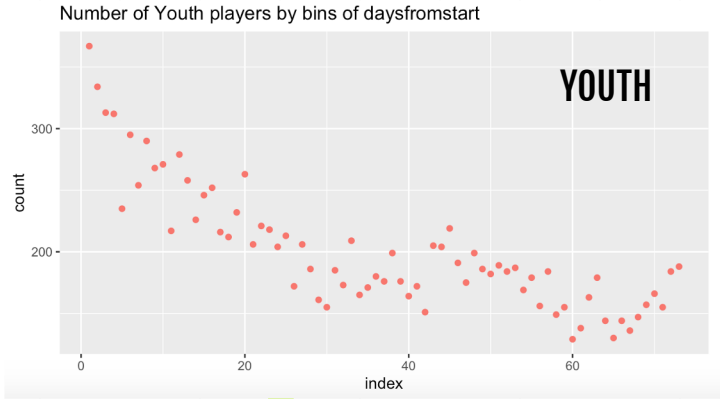
***count*** *= number of players within each bin*



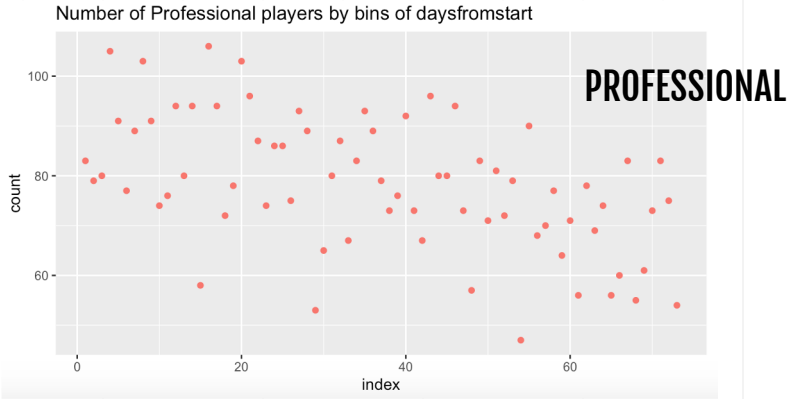

**Correlation (R-value): -0.836**




**Correlation (R-value): -0.416**

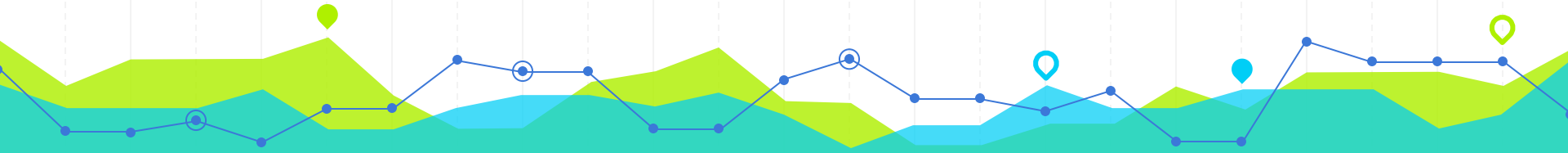



**Correlation (R-value): -0.815**




**Correlation (R-value): -0.509**

# How does the relative age effect impact elite soccer players internationally?



# Data Collection 1

# WORLD CUP DATA

## INCLUDED DATA

**Number**

- Total of 16,996 players

**Teams**

- Women and Men
- Youth (U-17, U-20, etc) and Professional

**Countries**

- 68 countries, with 887 players from the United States

## EXCLUDED DATA

- Players born on February 29th
- Countries that have played less than three times
- Repeated players
- Countries that no longer exist

## MUTATED DATA

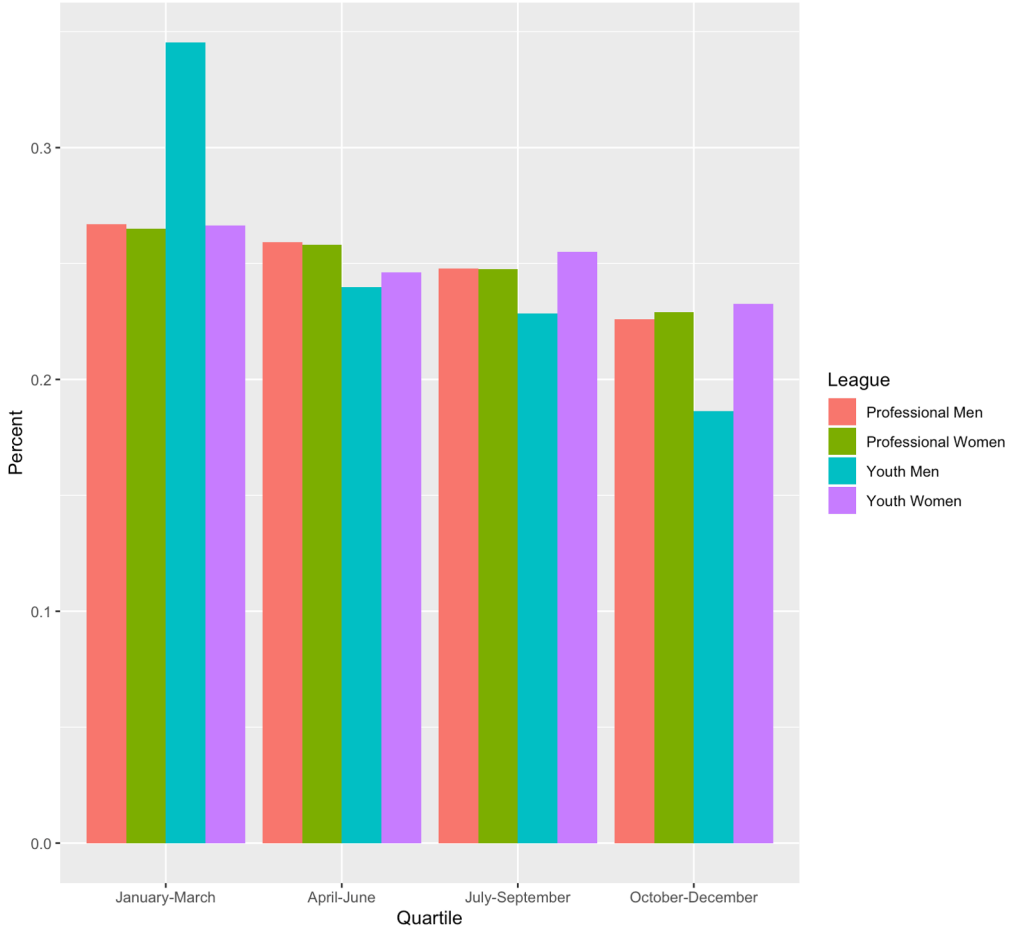
**Start of Season Dates**

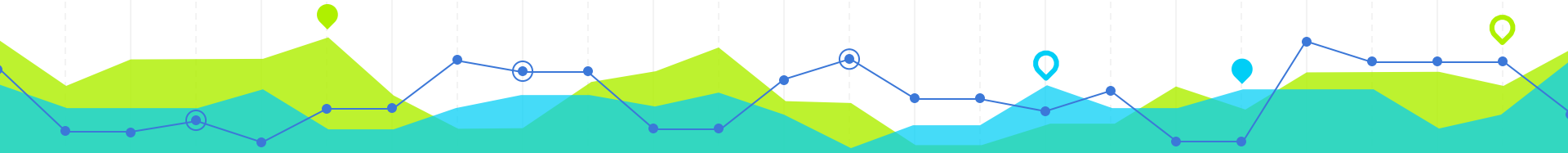
- International Standard: January 1st.

**Exceptions:**

- England: September 1st
- Japan: April 1st
- United States/Canada: August 1st

# Bar Graph of the Percentage of World Cup Players per Yearly Quartile





# Relative Age Effect On Gender 2

# NOW, WE MUST CONDUCT A TWO SAMPLE T-TEST FOR GENDER VS DAYS FROM START OF SEASON

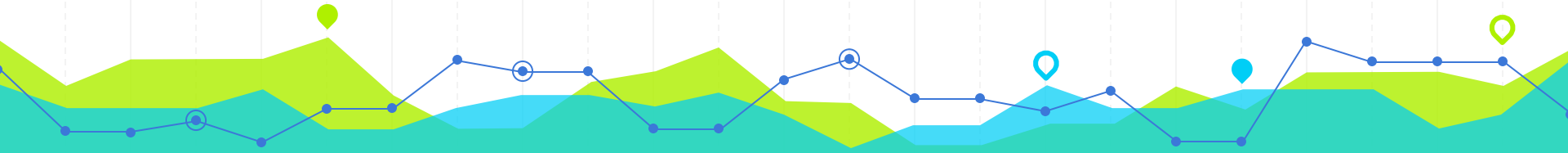
Null Hypothesis ($H_O$) = The difference of days from start of season between genders is 0.

Research Hypothesis ($H_R$) = The difference of days from start of season between genders is not 0.

$H_O = \mu diff = 0$ vs $H_R : \mu diff \neq 0$

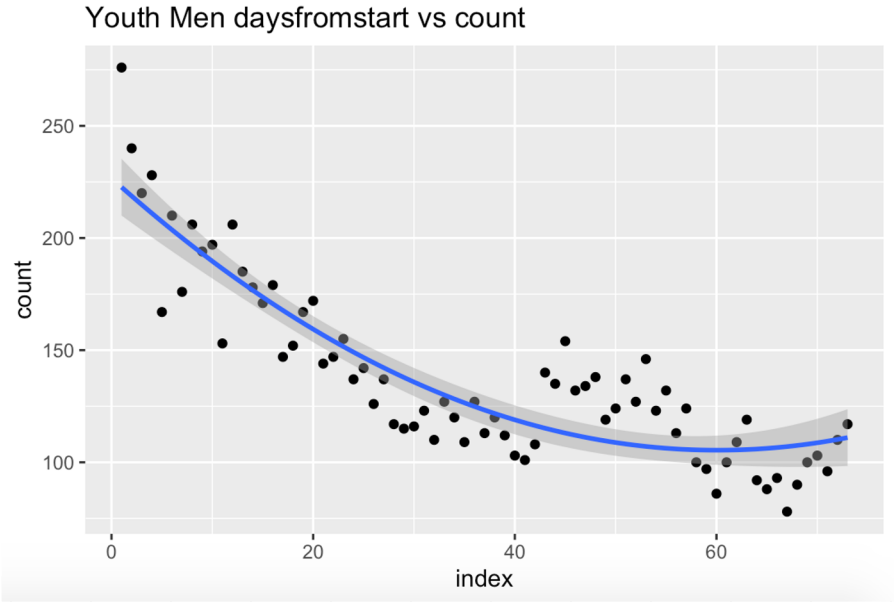
p =2.1*e^-15

The small p-value supports that we have enough data to reject the null, showing that the days from start of season is different in men and women.



# Relative Age Effect on Level of Soccer Played 3

# YOUTH MEN




**Polynomial Regression Model**

Adjusted R-squared:
- 0.7843

P-value:
- <2.2e-16

Residual standard error:
- 18.57 on 70 degrees of freedom

Correlation (R-value):
- -0.828

***Index*** *= number of bins (each bin = 5 days from start of season)*

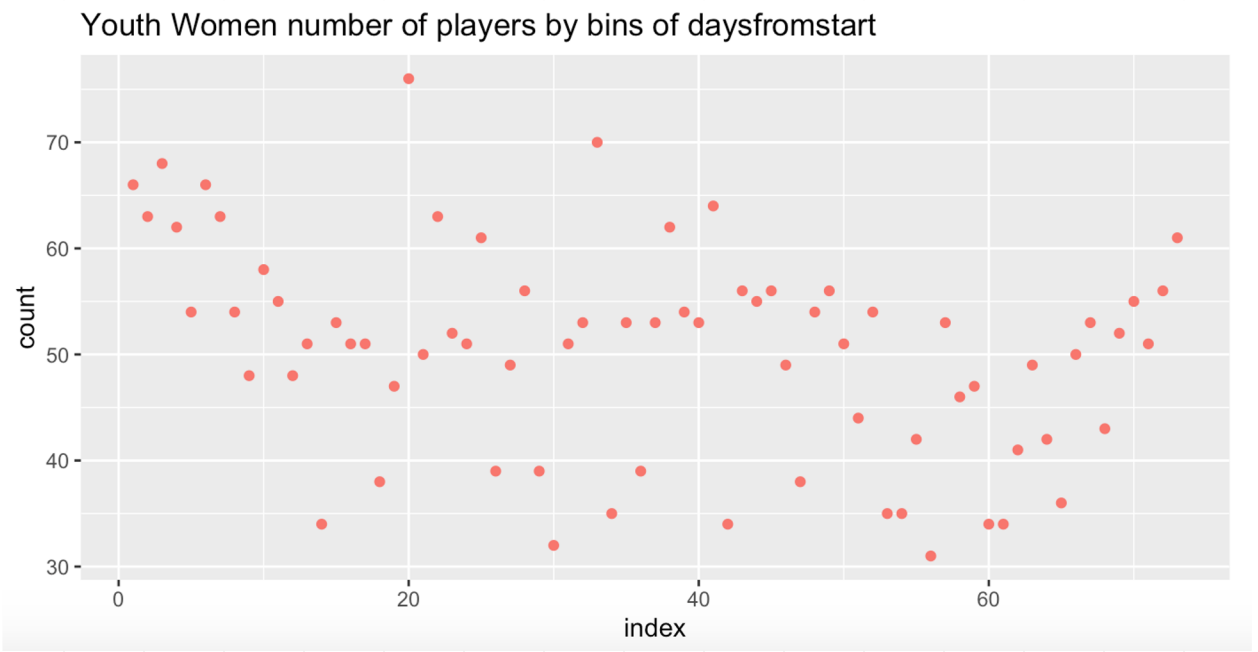
***count*** *= number of players within each bin*

# YOUTH WOMEN

***Index*** *= number of bins (each bin = 5 days from start of season)*

***count*** *= number of players within each bin*

**Correlation (R-value): -0.349**

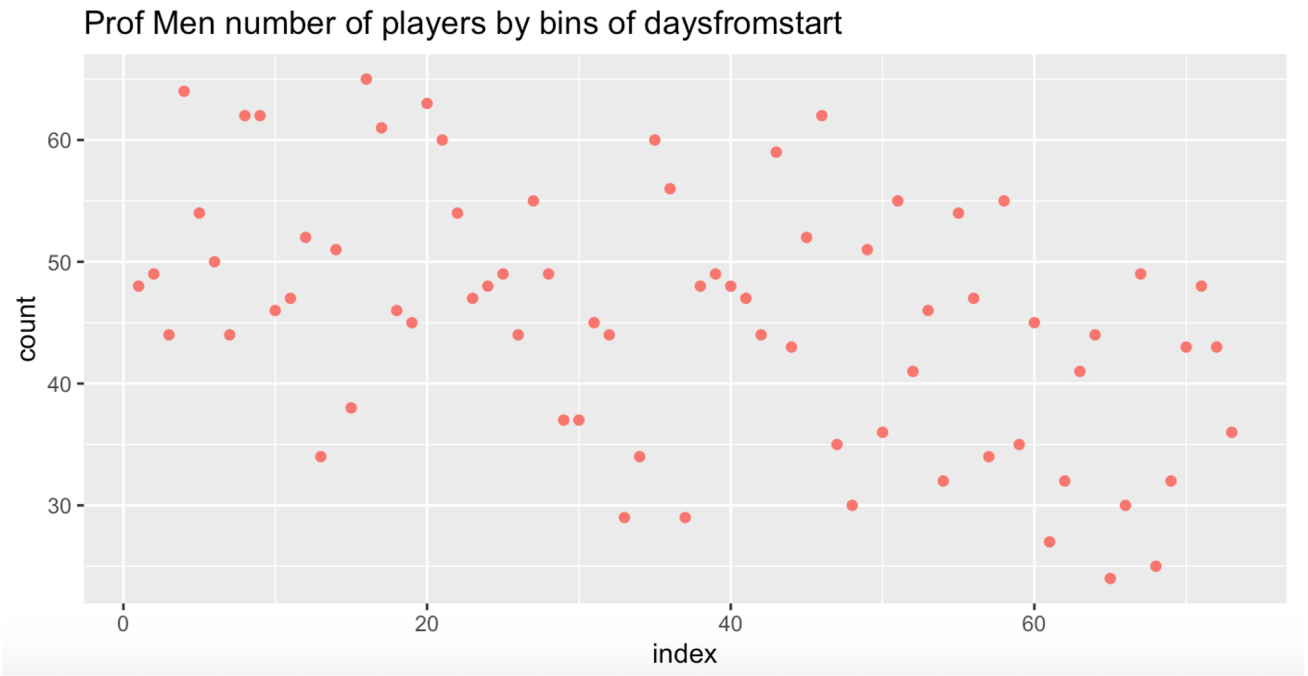


# PROFESSIONAL MEN

***Index*** *= number of bins (each bin = 5 days from start of season)*

***count*** *= number of players within each bin*

**Correlation (R-value): -0.454**

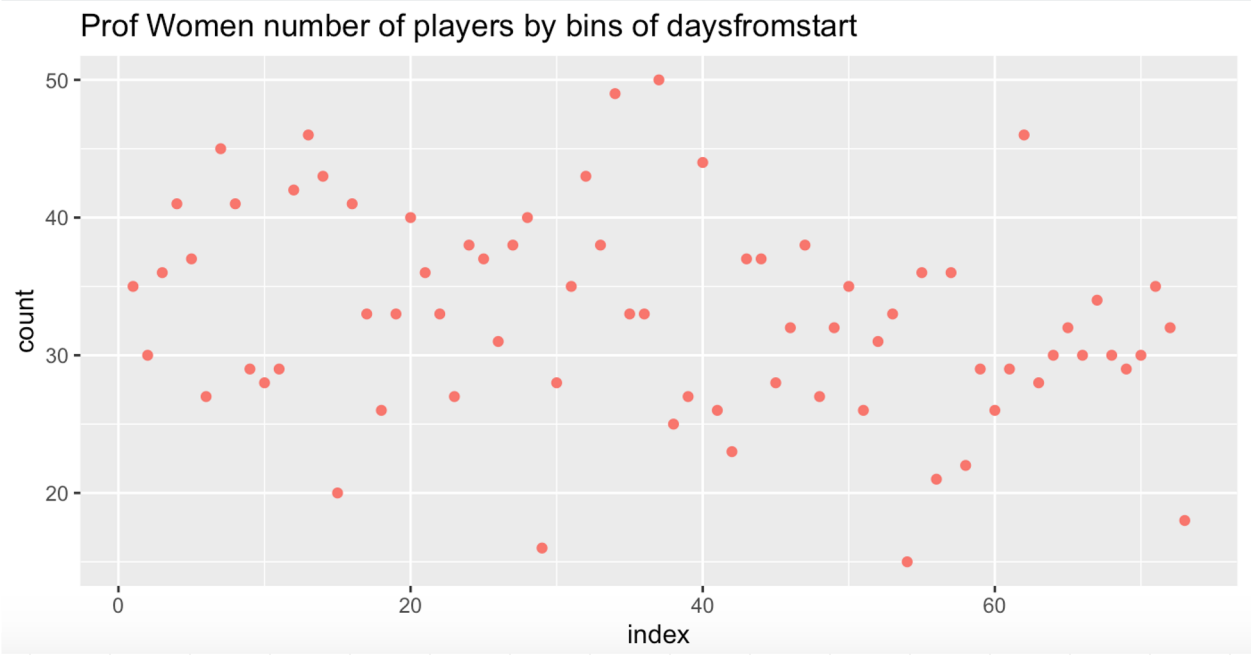


# PROFESSIONAL WOMEN

***Index*** *= number of bins (each bin = 5 days from start of season)*

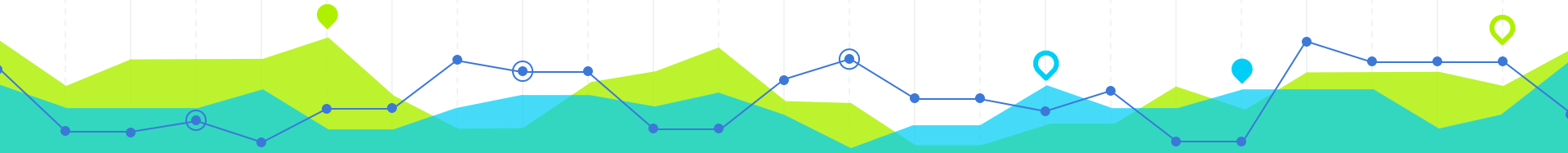
***count*** *= number of players within each bin*

**Correlation (R-value): -0.289**

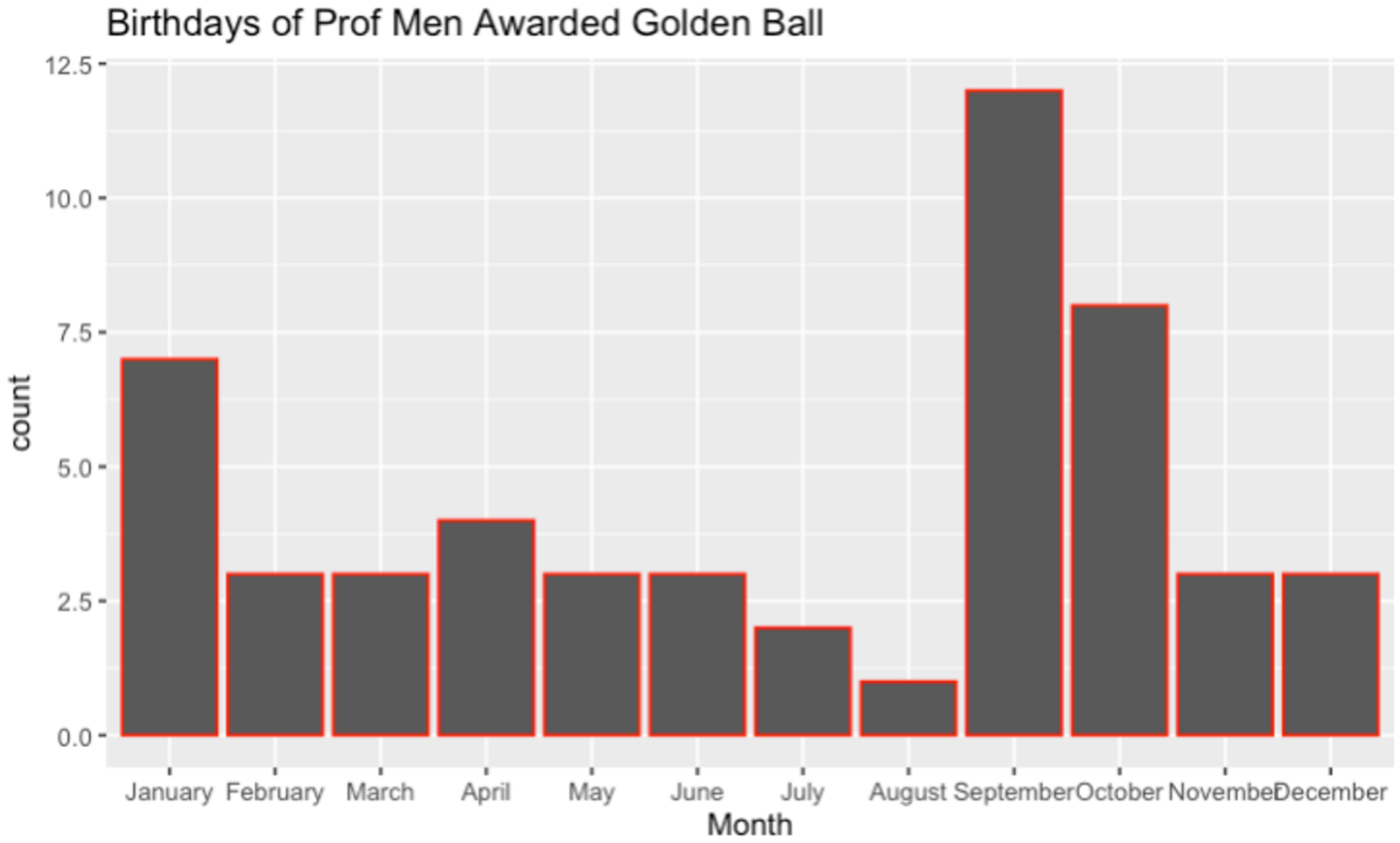


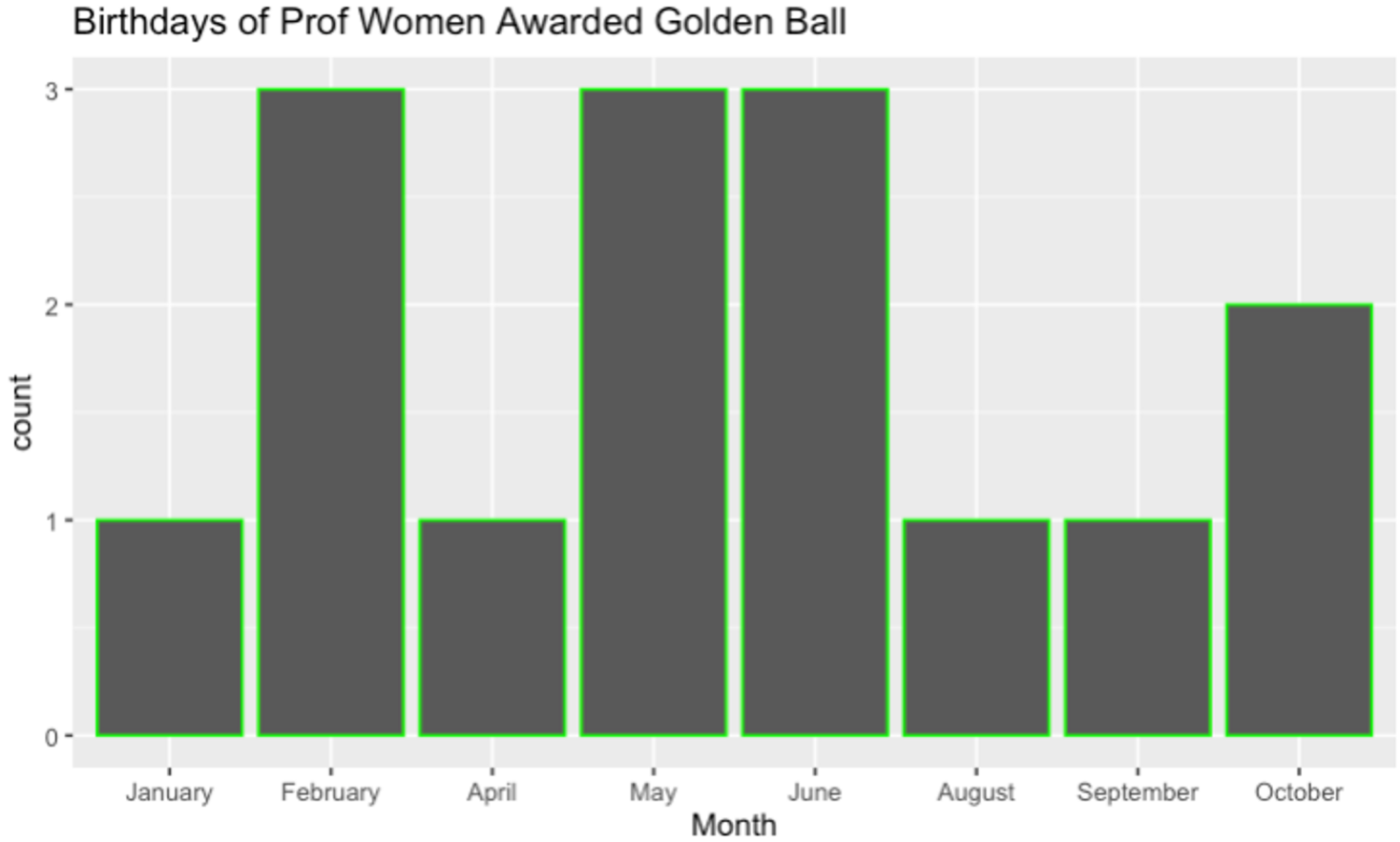
# FINDINGS

| | MEN | WOMEN |
|---|---|---|
| YOUTH | The highest correlation is drawn from this population, due to later pubescent years in comparison to women, and thus earlier born players experienced larger advantages. | The correlation in this population is low, due to the earlier pubescent years of women and thus there is an absence of RAE in ages 17 and up. |
| PROF | The correlation in this population is higher than both populations of women, but lower than youth men. | The lowest correlation of the study is in this population, which follows the hypothesis that RAE diminishes after players undergo puberty. |



# Golden Ball
## Award


Birthdays of Prof Men Awarded Golden Ball
12.5
10.0
7.5
5.0
2.5
0.0
count
January
February
March
April
May
June
July
August
September
October
November
December
Month


Birthdays of Prof Women Awarded Golden Ball
count
3
2
1
0
January
February
April
May
June
August
September
October
Month

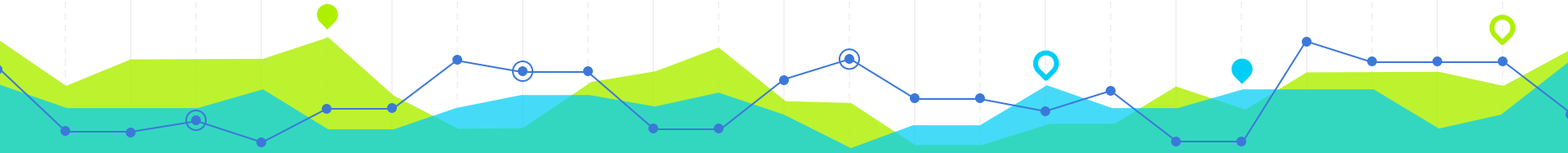
# FINDINGS

MEN

WOMEN

PROF

Both men and women had random birthdays without a visually obvious trend that follows the relative age effect, supporting the hypothesis that the relative age effect fades in professional soccer, due to the end of pubescent years.

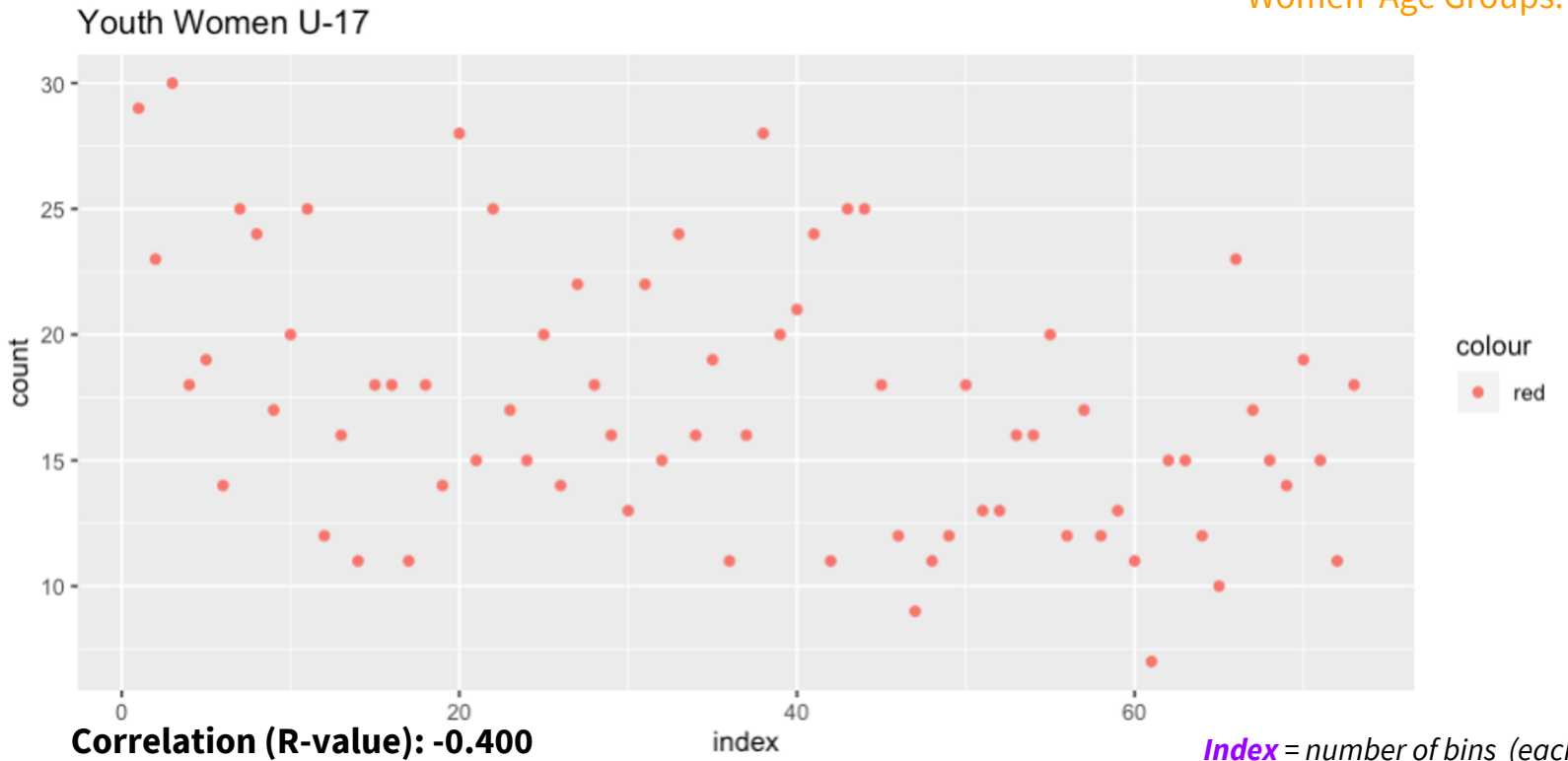


# Relative Age Effect Across Age Groups

# 4

# YOUTH WOMEN: U-17

Correlation of All Youth Women Age Groups: -0.349

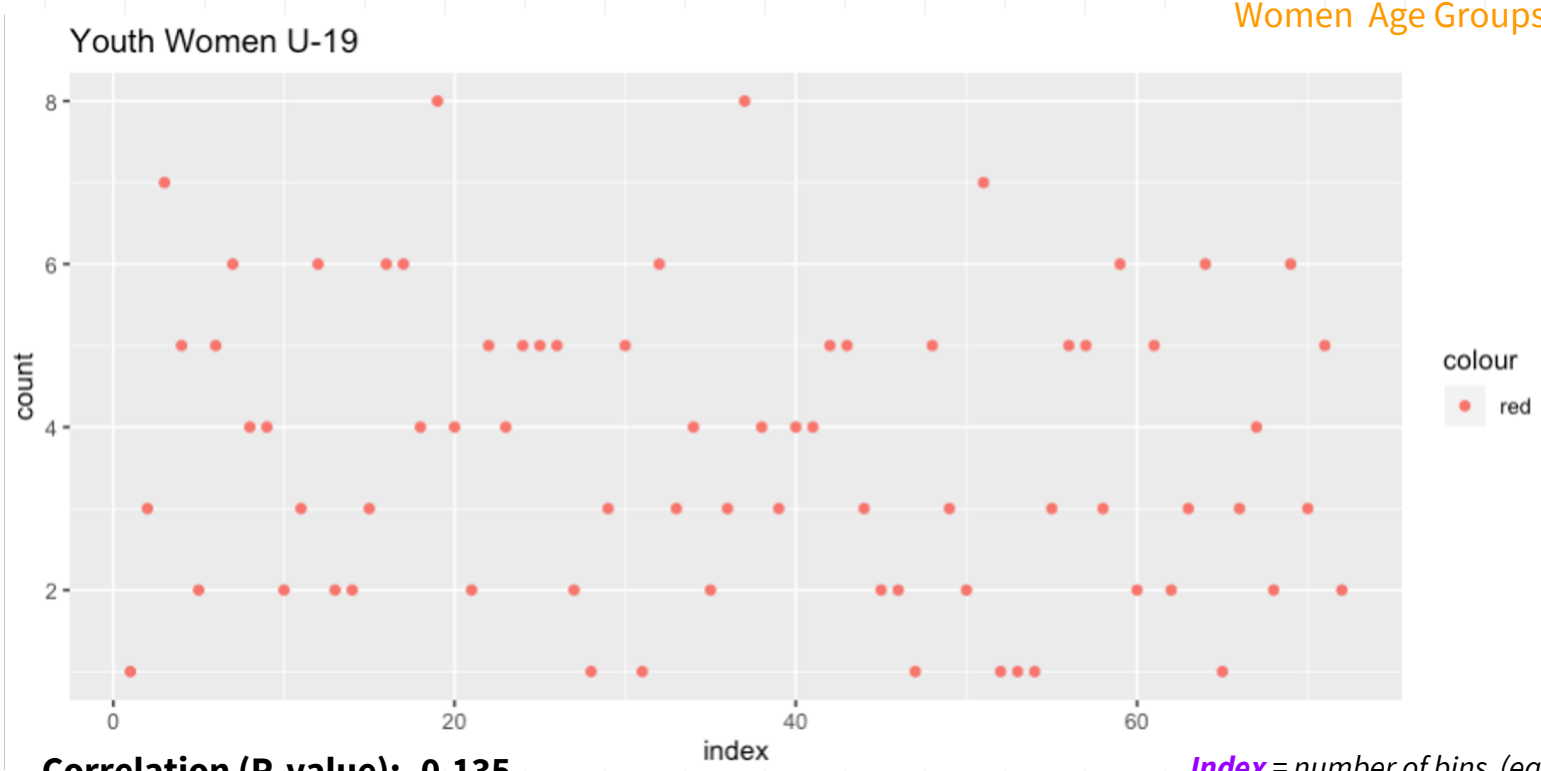



**Correlation (R-value): -0.400**

***Index*** *= number of bins (each bin = 5 days from start of season)*

# YOUTH WOMEN: U-19

Correlation of All Youth Women Age Groups: -0.349

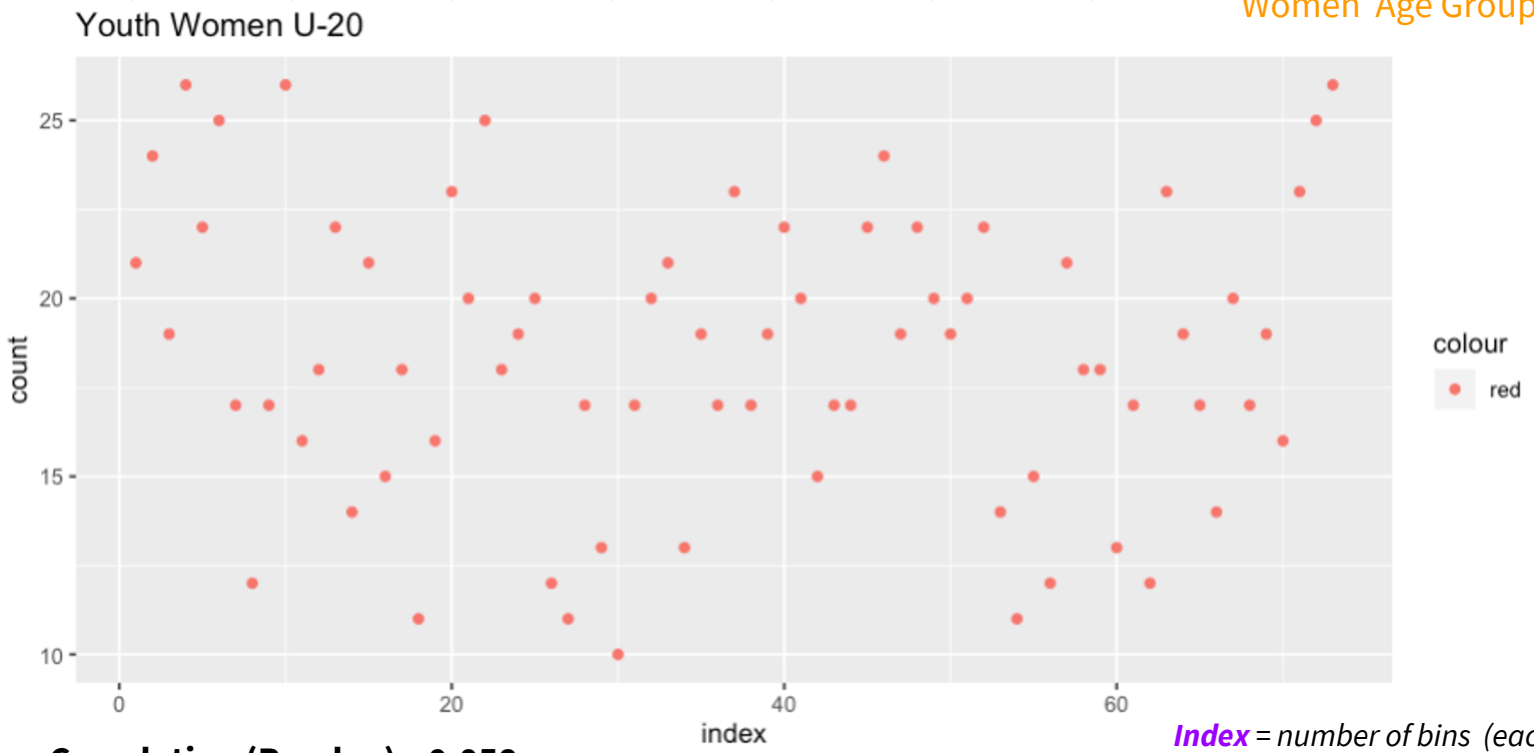



**Correlation (R-value): -0.135**

***Index*** *= number of bins (each bin = 5 days from start of season)*

# YOUTH WOMEN: U-20

Correlation of All Youth Women Age Groups: -0.349

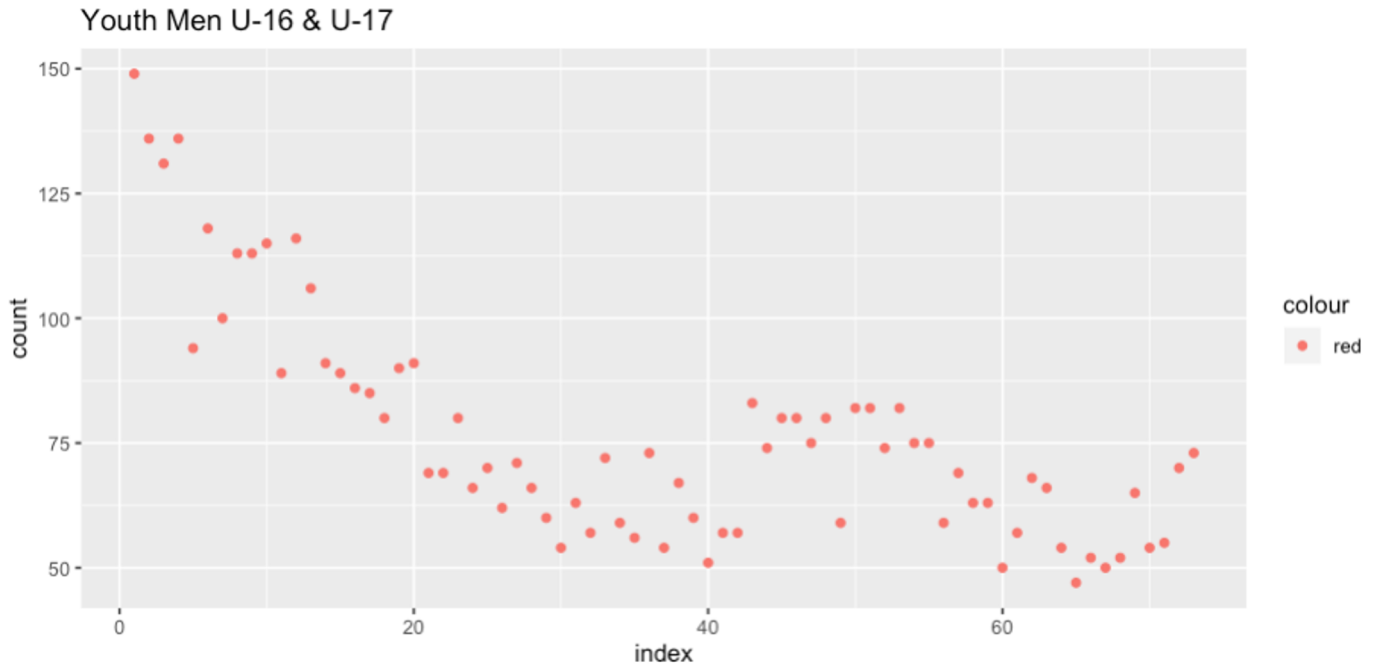



**Correlation (R-value): -0.058**

***Index*** *= number of bins (each bin = 5 days from start of season)*

# YOUTH MEN: U-16/U-17

Correlation of All Youth Men Age Groups: -0.828

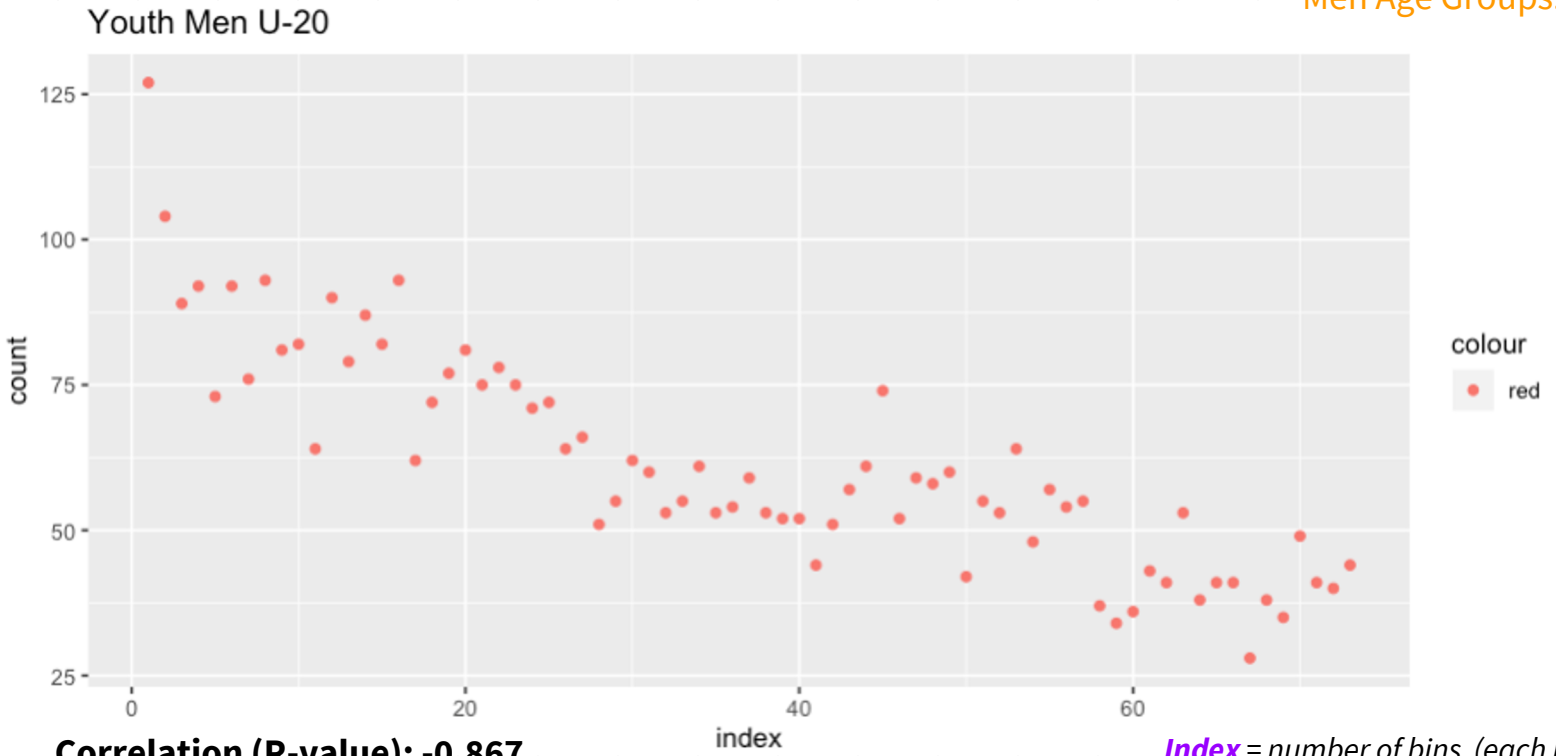



**Correlation (R-value): -0.723**

***Index*** *= number of bins (each bin = 5 days from start of season)*

# YOUTH MEN: U-20

Correlation of All Youth Men Age Groups: -0.828




**Correlation (R-value): -0.867**

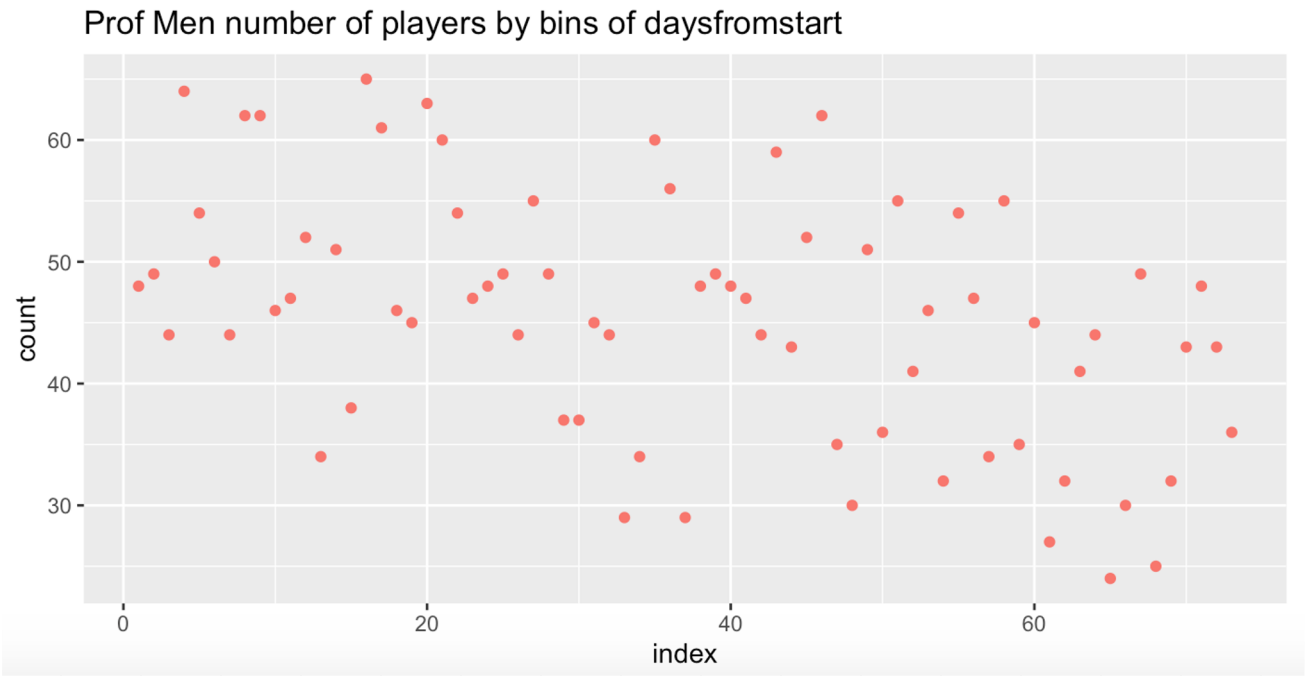
***Index*** *= number of bins (each bin = 5 days from start of season)*

# PROFESSIONAL MEN

***Index*** *= number of bins (each bin = 5 days from start of season)*

***count*** *= number of players within each bin*

**Correlation (R-value): -0.454**



# FINDINGS

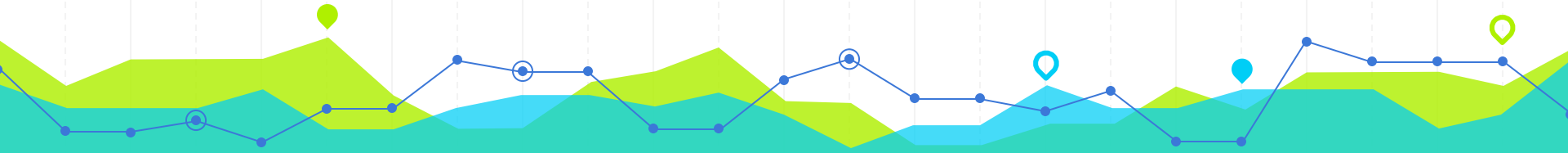
## MEN

## WOMEN

**YOUTH**

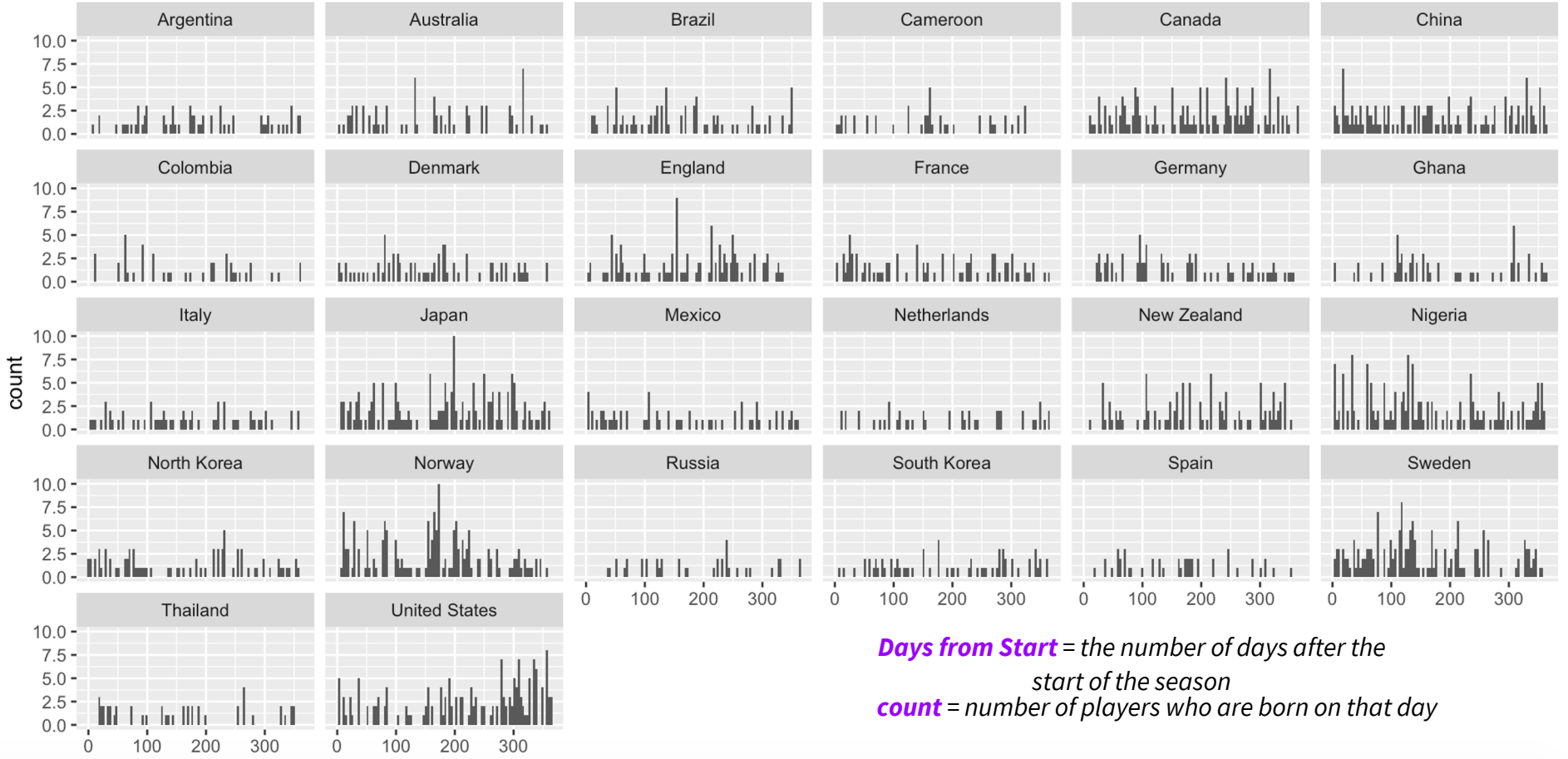
An interesting find within the male data is that rather than the RAE diminsing from U-17 to U-20, it increases with a strong correlation. However, looking at professional men data, the RAE fades, which supports our conclusion.

As female players grew older, a lower correlation emerged, thus proving that the RAE fades each year after pubescent years.



# Relative Age Effect by Country 5

# PROFESSIONAL WOMEN FACETED BY COUNTRY

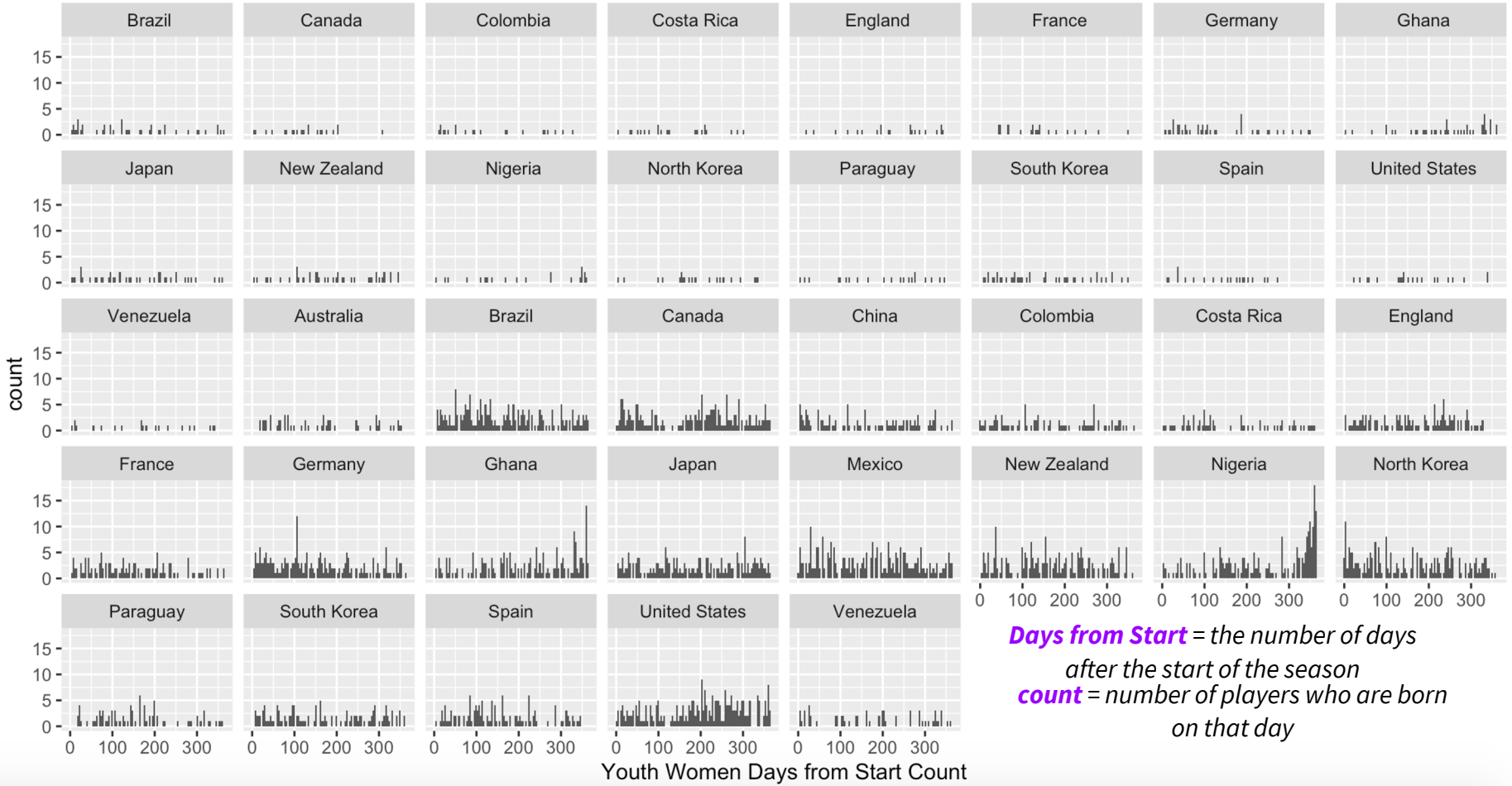



***Days from Start*** *= the number of days after the start of the season*

***count*** *= number of players who are born on that day*

Prof Women Days from Start Count

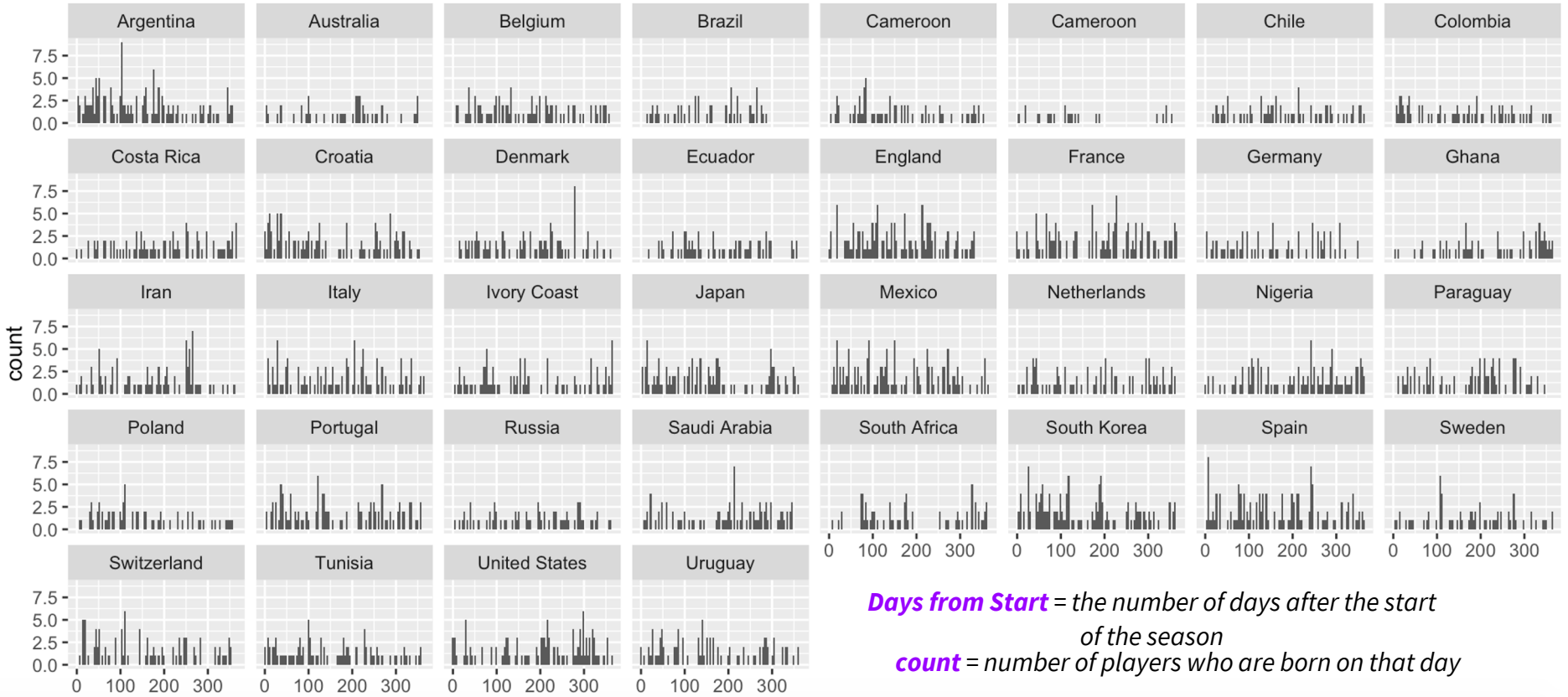
# YOUTH WOMEN FACETED BY COUNTRY




***Days from Start*** *= the number of days after the start of the season*

***count*** *= number of players who are born on that day*

# PROFESSIONAL MEN FACETED BY COUNTRY



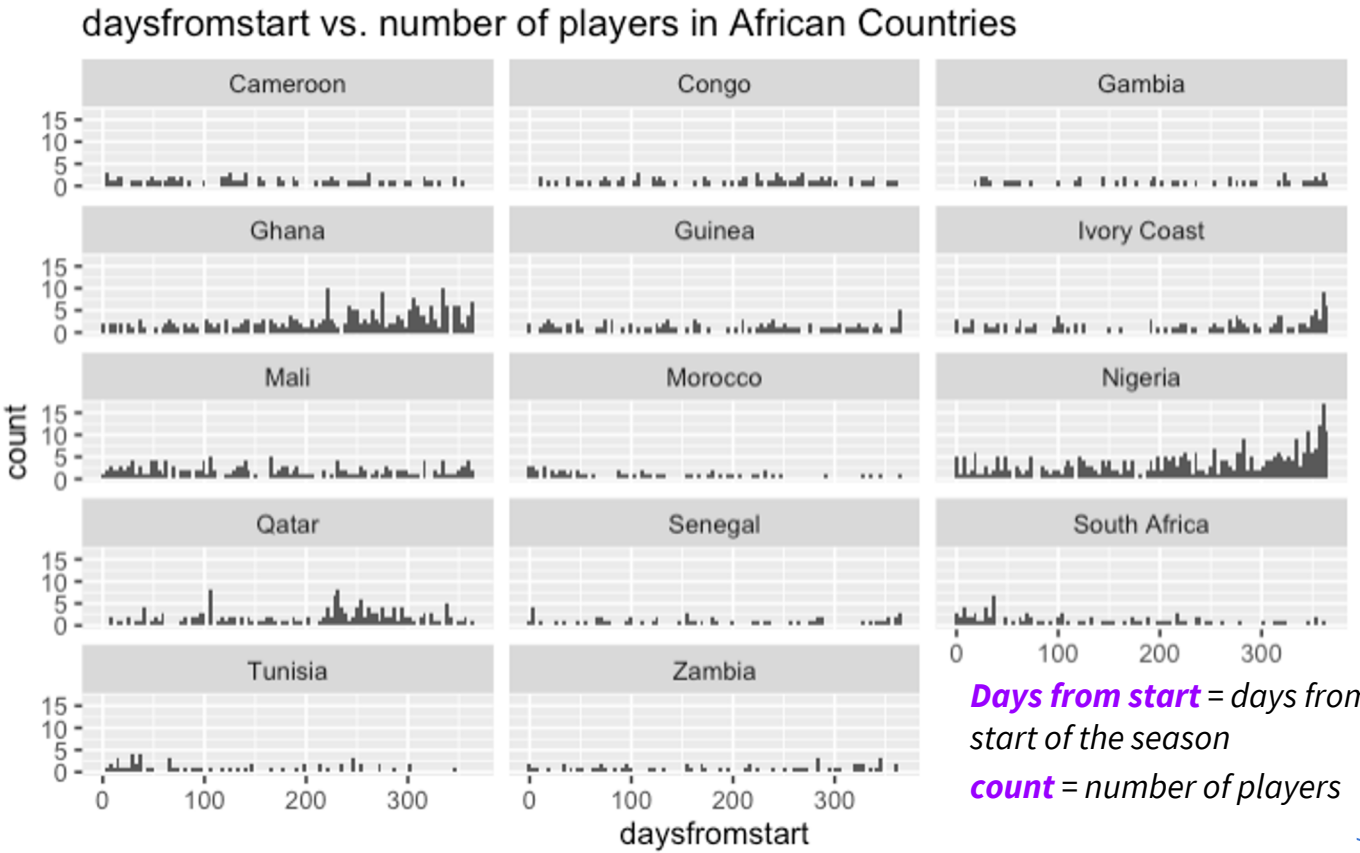

***Days from Start*** = *the number of days after the start of the season*

***count*** = *number of players who are born on that day*

# YOUTH MEN FACETED BY COUNTRY

# YOUTH MEN FACETED BY COUNTRY

# AFRICAN COUNTRIES ONLY




***Days from start*** *= days from start of the season*

***count*** *= number of players*

# CONCLUSION

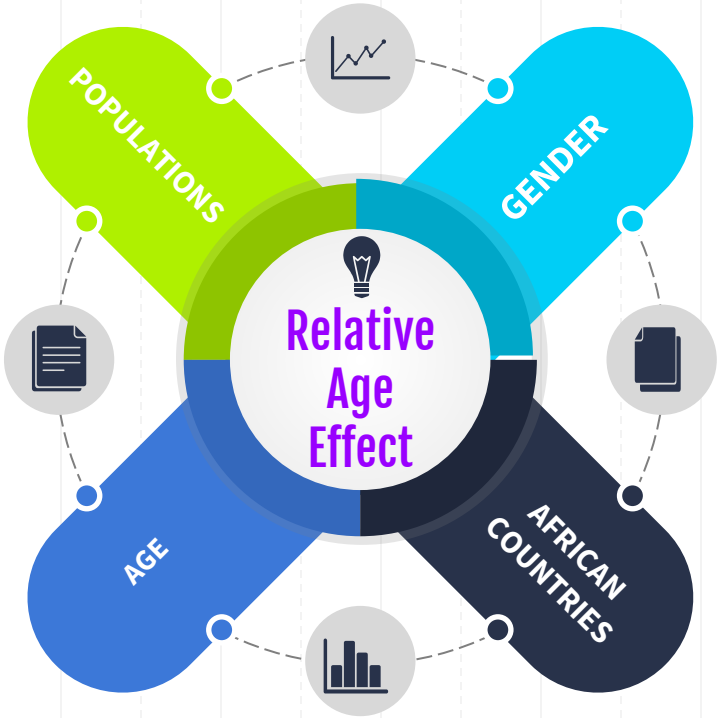
Across professional and youth, men and women

## Population

The youth men population has the strongest correlation and proves to be the only population that exhibits a relative age effect at a high degree.

## Age

Age impacts the RAE, supporting that as time goes on, the effect of RAE diminishes due to the end of puberty and a level playing field.




## Gender

Men have, on average, higher correlations that support existence of a RAE. The absence of RAE in youth women could be due to earlier maturation years, and thus earlier years with exhibited RAE.

## African Countries

African countries are an anomaly in our data, with an inverse RAE that spikes up the farther away from the season start.