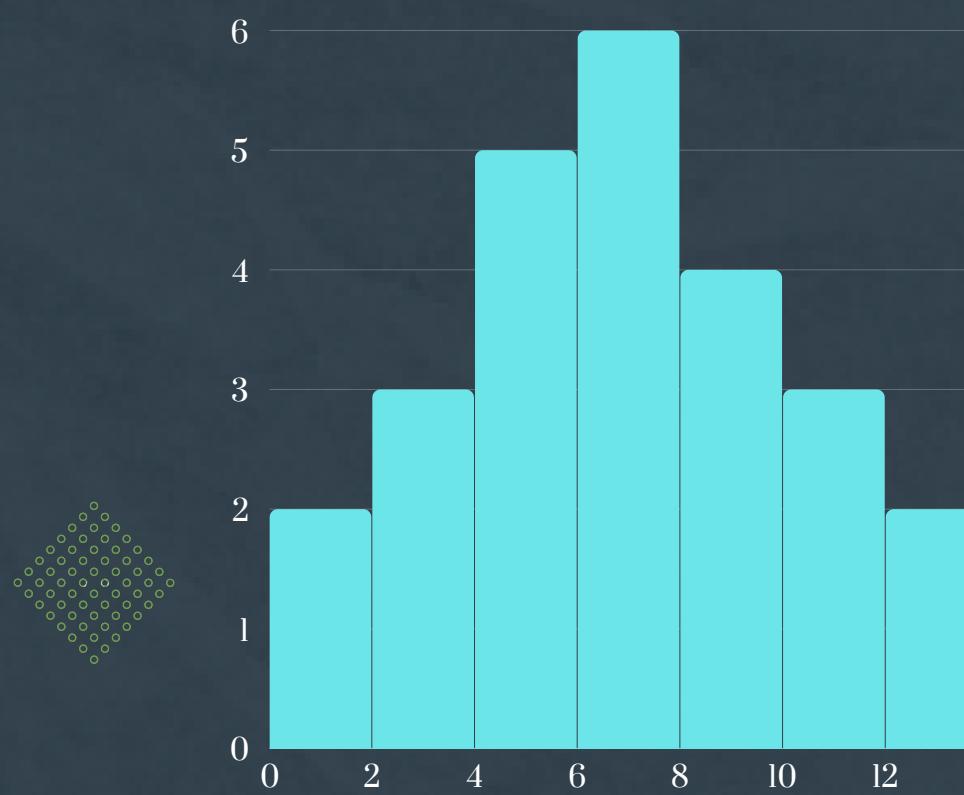


# TEAM C - PROJECT 1

DO HUU DAT  
NGO HOANG DUNG  
NGUYEN DUC LOI



# DATASET INTRODUCTION

- Dataset sourced from Statistics Canada, NHL team list endpoint, and NHL API
- Investigation focuses on professional ice hockey demographics
- We aim to elucidate whether similar patterns persist among Canadian-born NHL players.





# KEY QUESTION



Gladwell's hypothesis posits that individuals born in these early months hold a statistical advantage in sports, a concept often referred to as the "relative age effect." Through rigorous analysis of player birthdates in the National Hockey League (NHL), we endeavor to ascertain whether Gladwell's claims hold true within the context of professional ice hockey.

The second question at the forefront of our investigation pertains to the utilization of local talent by NHL teams. By scrutinizing team rosters and player demographics, we seek to elucidate whether NHL teams exhibit a propensity towards selecting players from their immediate geographic regions or whether talent acquisition transcends local boundaries.





# INTRODUCTION AND APPROACH

## QUESTION 1: IS THERE ANY BIAS REGARDING THE BIRTH CITY?

- WE ENCOUNTERED DISCUSSIONS SURROUNDING MALCOLM GLADWELL'S HYPOTHESIS, WHICH SUGGESTS THAT INDIVIDUALS BORN IN THE FIRST THREE MONTHS OF THE YEAR MAY POSSESS A STATISTICAL ADVANTAGE IN SPORTS DUE TO DEVELOPMENTAL FACTORS
- FIRSTLY, WE WILL ASSESS WHETHER THERE IS A PREFERENCE FOR PLAYERS FROM CERTAIN COUNTRIES ACROSS THE LEAGUE.
- WE WILL ALSO LOOK INTO THE PRESENCE AND PERFORMANCE OF INTERNATIONAL SCOUTING NETWORKS TO SEE IF THERE ARE DISPARITIES IN HOW PLAYERS FROM DIFFERENT REGIONS ARE SCOUTED AND VALUED
- SECONDLY, WE WILL EXPLORE WHETHER INDIVIDUAL TEAMS SHOW A BIAS TOWARDS LOCAL PLAYERS, SPECIFICALLY THOSE BORN IN THE SAME CITY OR REGION AS THE TEAM



# ANALYSIS

## QUESTION 1: IS THERE ANY BIAS REGARDING THE BIRTH CITY?

UTILIZED GGPLOT2 AND SF LIBRARIES TO VISUALIZE NHL PLAYERS' BIRTH COUNTRIES ON A WORLD MAP.

EMPLOYED DPLYR PACKAGE FOR DATA MANIPULATION.

LOADED MEDIUM-SCALE WORLD MAP AS A SIMPLE FEATURES OBJECT.

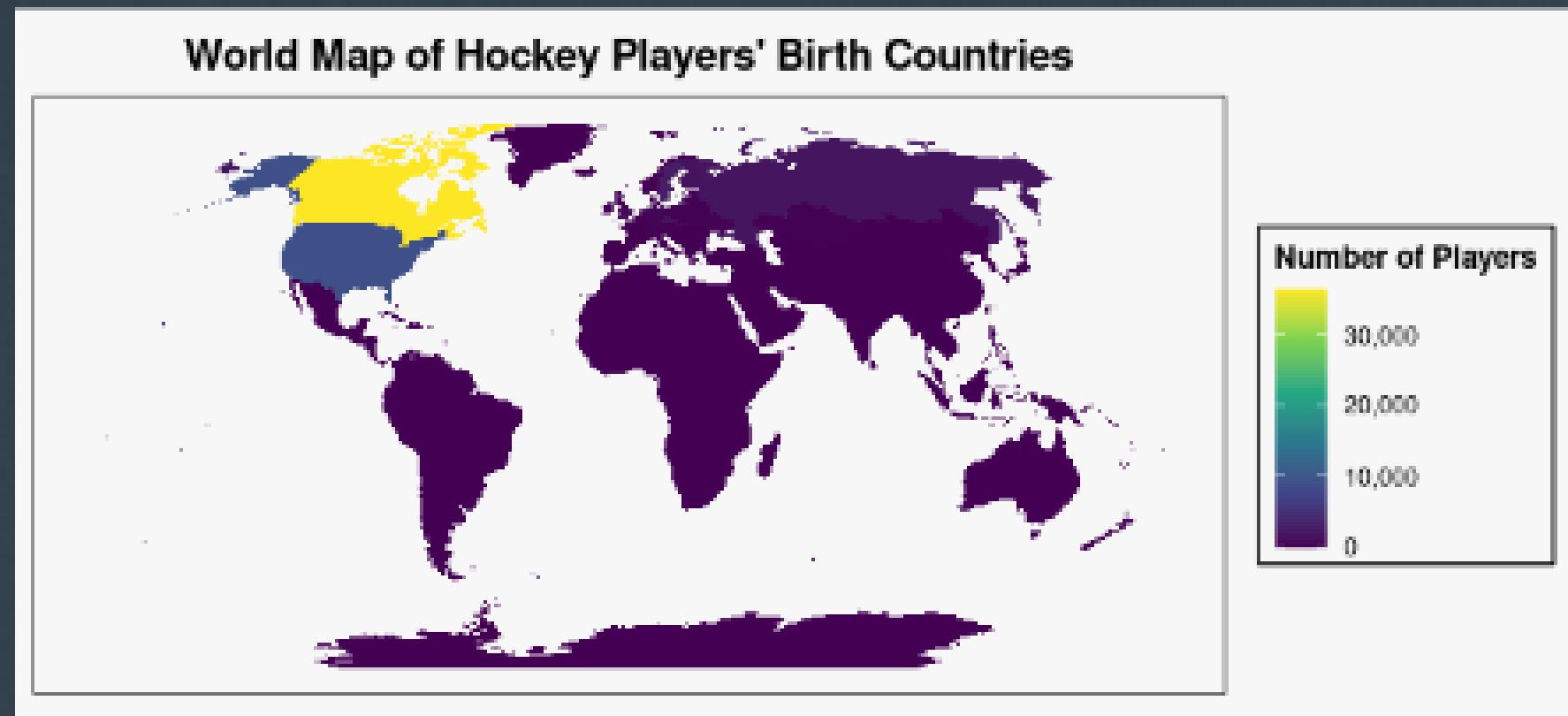
GROUPED NHL ROSTERS DATASET BY BIRTH COUNTRY TO COUNT PLAYERS.

MERGED PLAYER DATA WITH WORLD MAP DATA, SETTING ZERO COUNTS TO NA FOR VISUAL CLARITY.

CHOSE VIRIDIS COLOR SCALE FOR AESTHETIC CLARITY AND COMMA-SEPARATED LABELS.

IMPLEMENTED ROBINSON PROJECTION FOR ENHANCED VISUAL APPEAL.

MADE AESTHETIC ADJUSTMENTS TO IMPROVE PLOT READABILITY AND PRESENTATION.





# ANALYSIS

## QUESTION 1: IS THERE ANY BIAS REGARDING THE BIRTH CITY?

LOAD TEAM DATA FROM "NHL.CSV" INTO DATA FRAME TEAM\_DATA.

MERGE TEAM\_DATA WITH NHL\_ROSTERS BASED ON MATCHING 'TEAM\_CODE' AND 'ABBREVIATION'.

USE HEAD FUNCTION TO CHECK INITIAL ROWS FOR CORRECTNESS.

ADD NEW COLUMN SAME\_CITY TO MERGED DATA TO INDICATE PLAYER'S BIRTH CITY MATCHING TEAM'S CITY USING MUTATE FUNCTION.

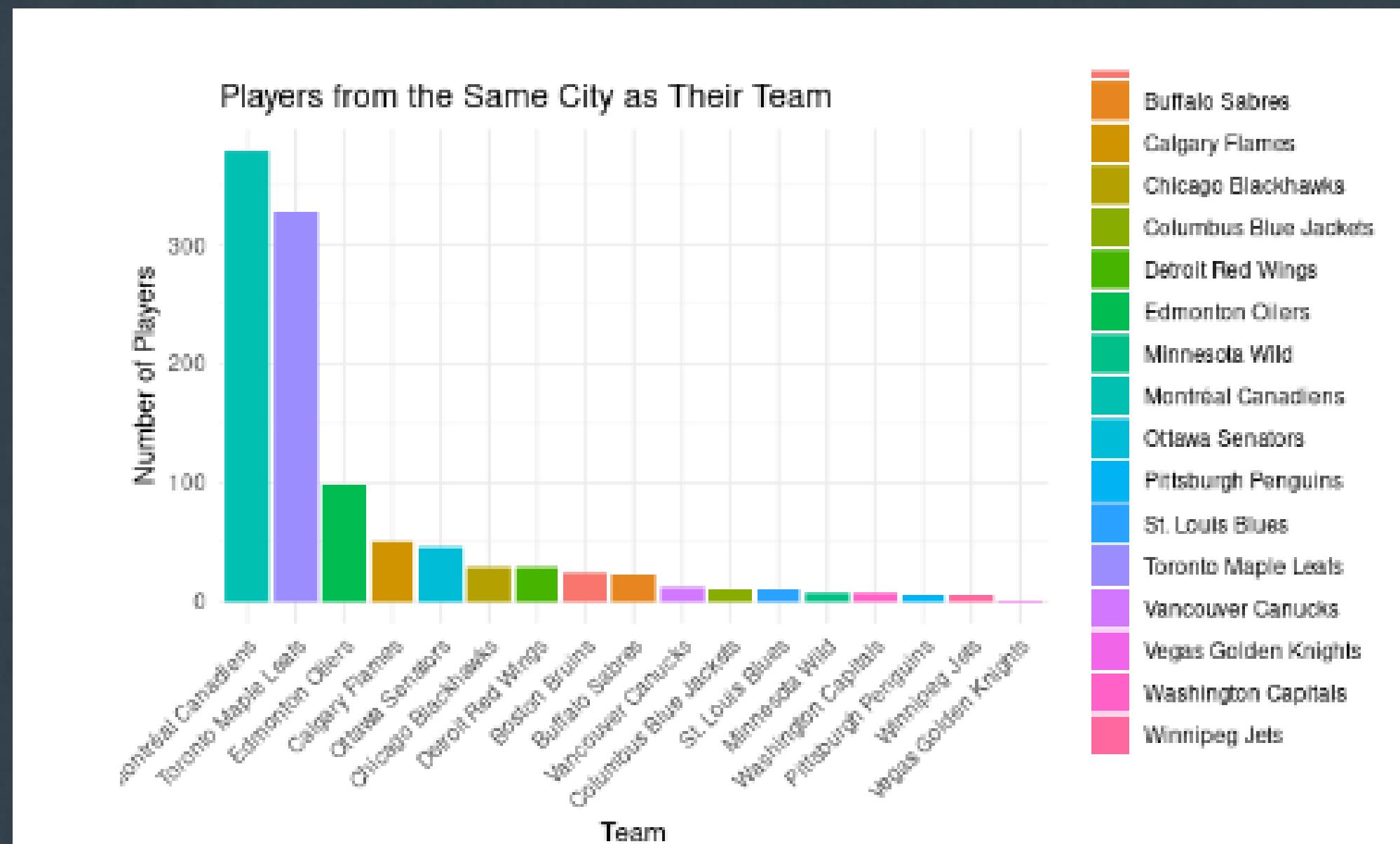
SUMMARIZE DATA TO COUNT PLAYERS PER TEAM FROM THE SAME CITY AS THEIR TEAM.

FILTER OUT TEAMS WITH NO PLAYERS FROM THEIR CITY.

CREATE BAR PLOT USING GGPLOT2 TO VISUALIZE NUMBER OF PLAYERS FROM SAME CITY AS THEIR TEAM.

ARRANGE TEAMS BY NUMBER OF SUCH PLAYERS AND APPLY MINIMAL THEME STYLING.

SET APPROPRIATE LABELS FOR AXES AND PLOT TITLE, AND ROTATE X-AXIS LABELS FOR BETTER VISIBILITY.



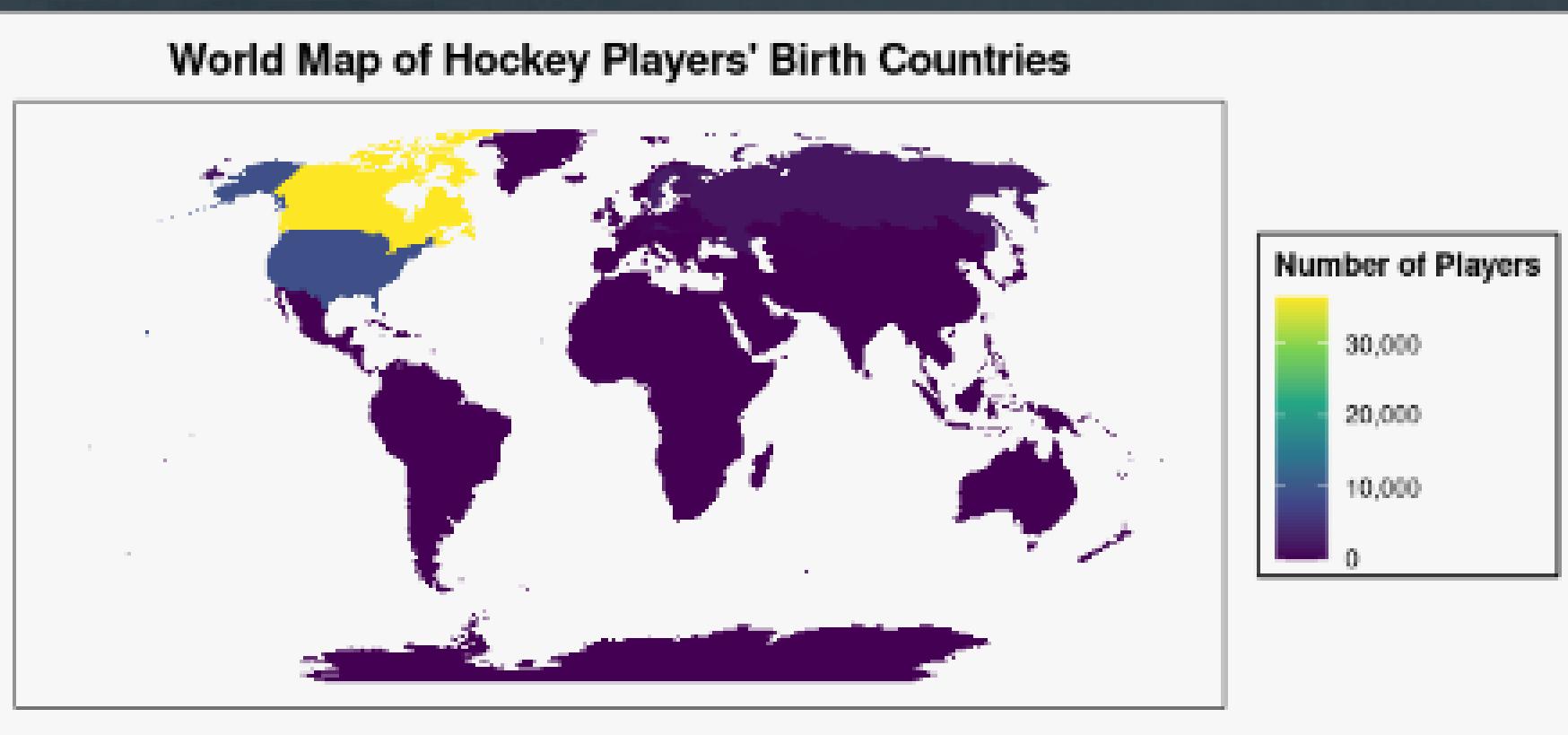
# DISCUSSION

## QUESTION 1: IS THERE ANY BIAS REGARDING THE BIRTH CITY?

FIRST FIGURE: WORLD MAP DISPLAYING BIRTH COUNTRIES OF HOCKEY PLAYERS.

DARKER HUES DOMINATE NORTH AMERICA AND NORTHERN EUROPE, INDICATING HIGH PLAYER DENSITY.

SUGGESTS A POTENTIAL GEOGRAPHICAL BIAS DUE TO CULTURAL ROOTS, ICE AVAILABILITY, AND STRUCTURED TRAINING PROGRAMS IN COLDER CLIMATES.



# DISCUSSION

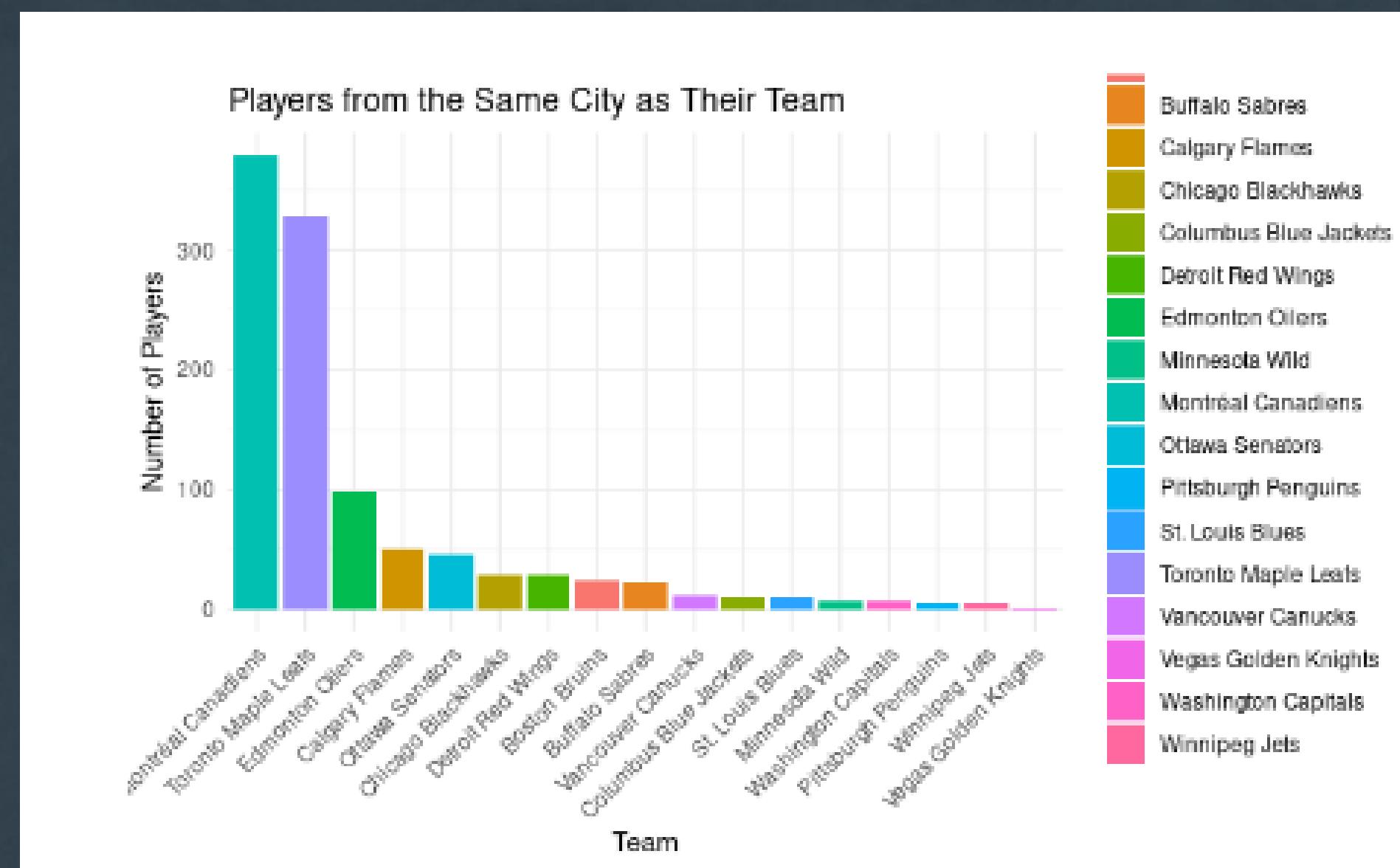


## QUESTION 1: IS THERE ANY BIAS REGARDING THE BIRTH CITY?

SECOND FIGURE: BAR CHART SHOWING LOCAL RECRUITMENT BIASES IN NHL TEAMS.

SOME TEAMS HAVE A SIGNIFICANT NUMBER OF PLAYERS BORN IN THEIR CITY, IMPLYING A BIAS TOWARDS HOMEGROWN TALENT.

REFLECTS MARKETING STRATEGIES AND INVESTMENTS IN LOCAL SPORTS PROGRAMS TO NURTURE TALENT.





# INTRODUCTION AND APPROACH

## QUESTION 2: WHETHER MALCOLM GLADWELL'S ASSERTION IS STILL VALID?

- WE ENCOUNTERED DISCUSSIONS SURROUNDING MALCOLM GLADWELL'S HYPOTHESIS, WHICH SUGGESTS THAT INDIVIDUALS BORN IN THE FIRST THREE MONTHS OF THE YEAR MAY POSSESS A STATISTICAL ADVANTAGE IN SPORTS DUE TO DEVELOPMENTAL FACTORS.
- FIRSTLY, WE WILL EMPLOY A LINE CHART TO ILLUSTRATE THE OVERALL TREND OF BIRTHDATE DISTRIBUTIONS AMONG NHL PLAYERS OVER THE YEARS.
- SECONDLY, TO ASSESS THE VALIDITY OF MALCOLM GLADWELL'S ASSERTION AND ITS IMPACT ON INDIVIDUAL NHL TEAMS, WE WILL UTILIZE A BAR CHART. THIS CHART WILL DISPLAY THE PROPORTION OF HOW MANY SEASONS DID THEY HAVE DOMINANT NUMBER OF FIRST-QUARTER-BORN PLAYERS.
- SECONDLY, WE WILL EXPLORE WHETHER INDIVIDUAL TEAMS SHOW A BIAS TOWARDS LOCAL PLAYERS, SPECIFICALLY THOSE BORN IN THE SAME CITY OR REGION AS THE TEAM

# ANALYSIS

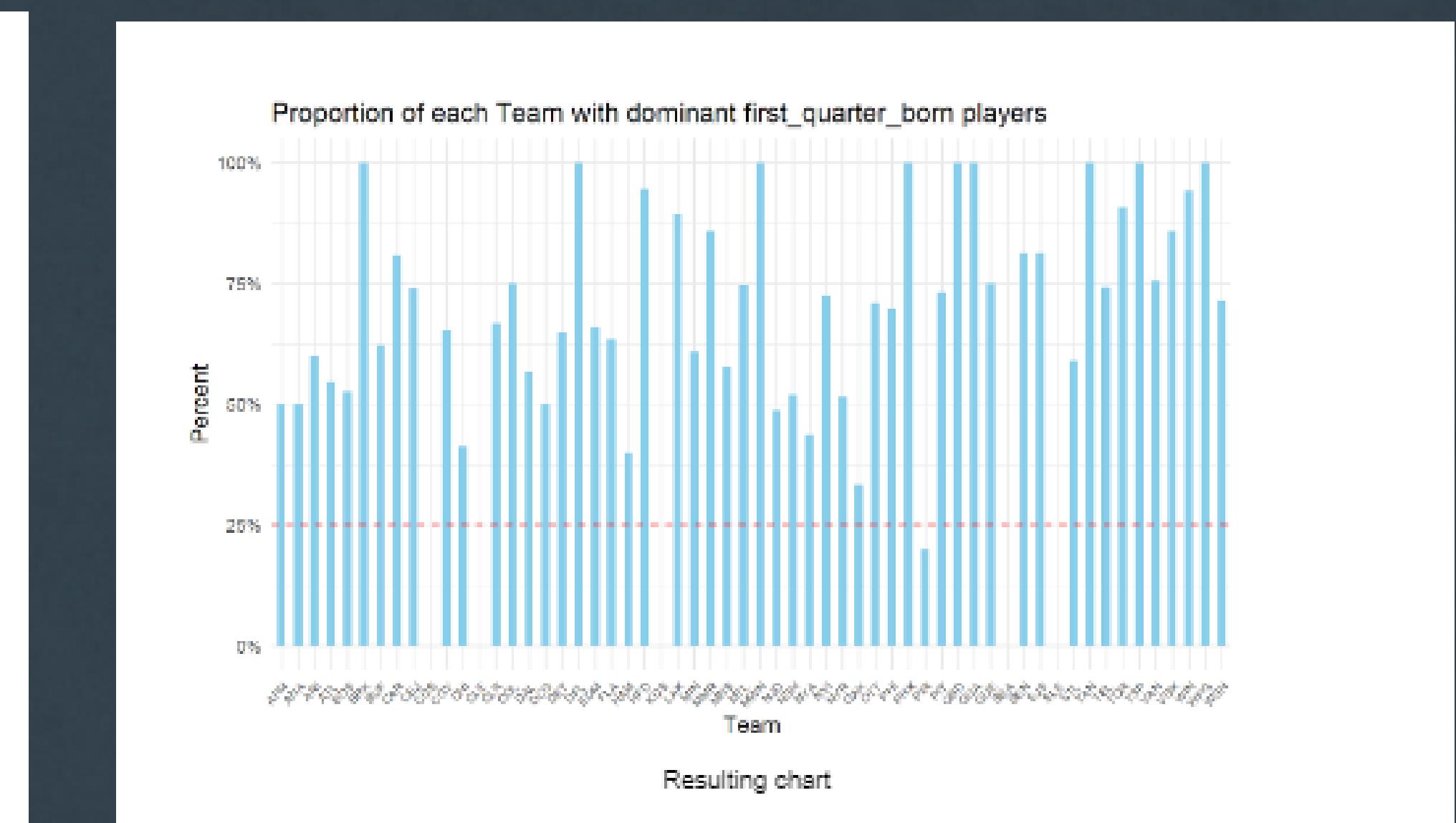
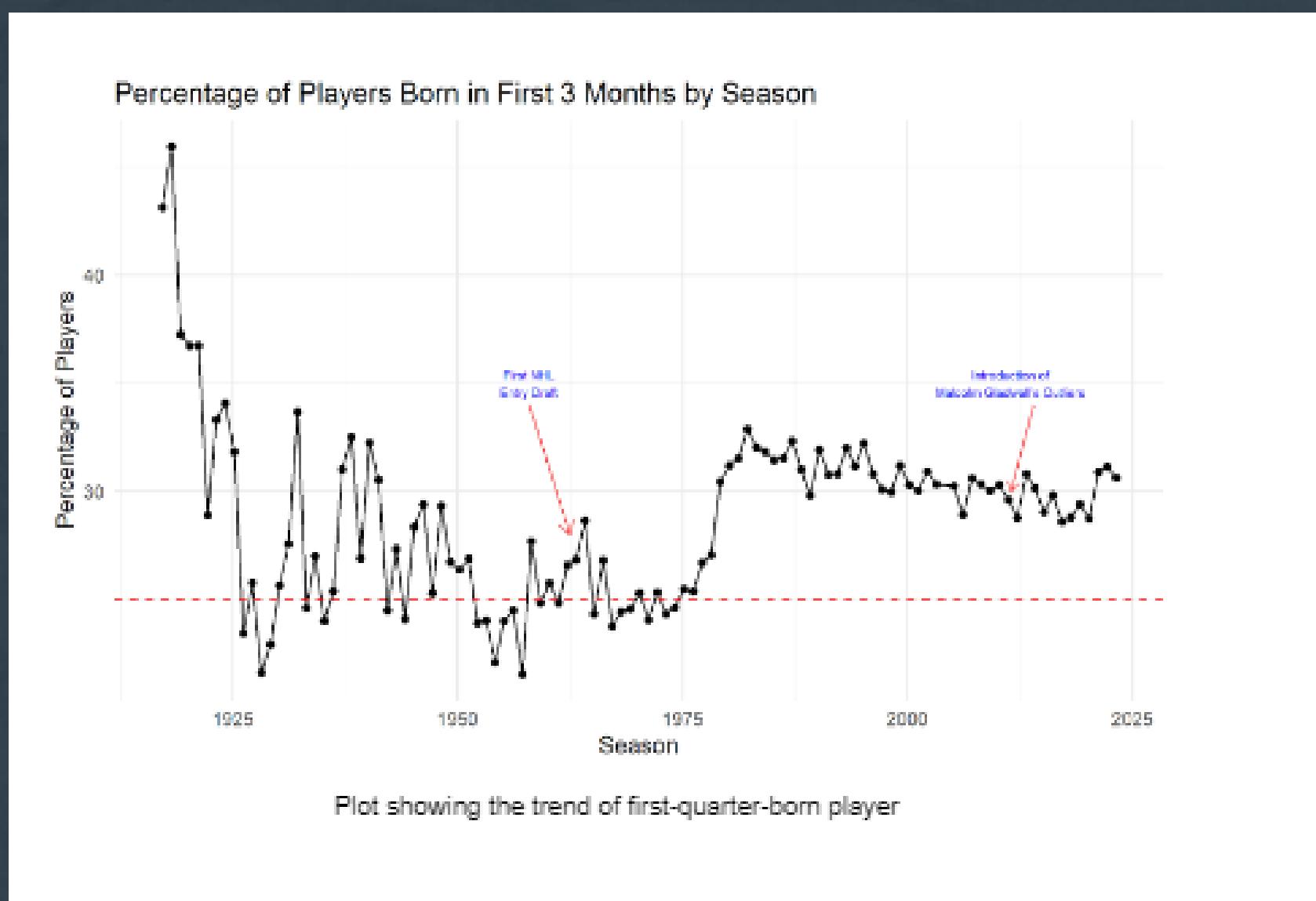


MERGE THE TWO DATA FILES USING A UNIQUE IDENTIFIER LIKE PLAYER\_ID TO CREATE A COMPREHENSIVE DATASET.

REMOVE EXTRANEOUS COLUMNS NOT RELEVANT TO THE ANALYSIS TO STREAMLINE THE DATASET.

CREATE A NEW DATAFRAME CALCULATING THE PERCENTAGE OF PLAYERS BORN IN THE FIRST QUARTER (JANUARY TO MARCH) COMPARED TO TOTAL PLAYERS.

PREPARE THE DATA FOR VISUALIZATION.



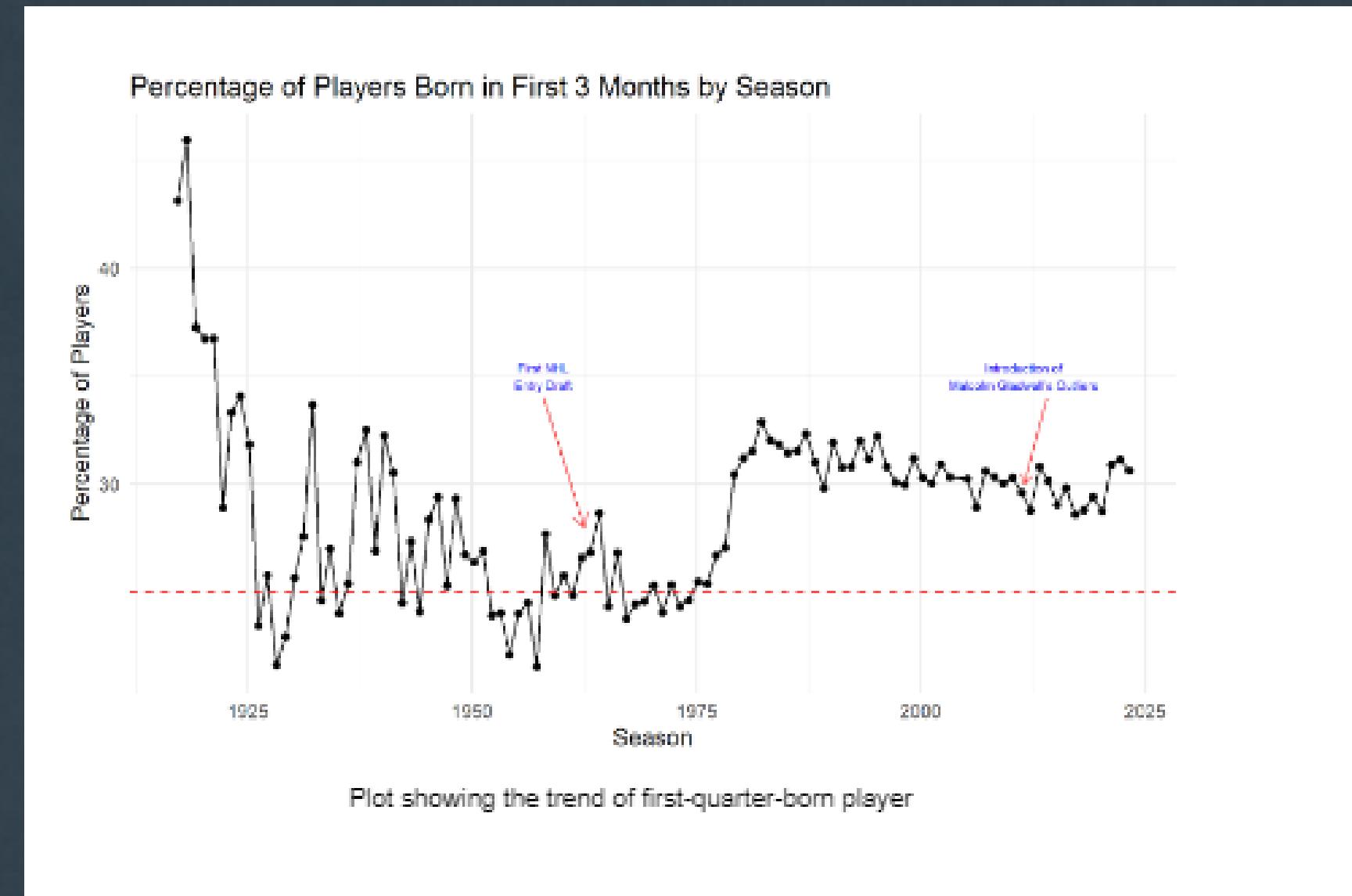
# DISCUSSION



## QUESTION 2: WHETHER MALCOLM GLADWELL'S ASSERTION IS STILL VALID?

FIRST GRAPH: EXAMINING MALCOLM'S ASSERTION OVER TIME.

- CONSISTENT PATTERN: >25% PLAYERS BORN IN FIRST QUARTER.
- STABILIZATION AT ~30% ATTRIBUTED TO NHL DRAFT EVENT IN 1964.
- NHL DRAFT BROUGHT STABILITY TO TEAM COMPOSITIONS.
- BEFORE 1964: ERRATIC PLAYER ACQUISITION, FLUCTUATING PROPORTIONS.
- NHL DRAFT'S IMPACT UNDERSCORES ITS PROFOUND INFLUENCE ON SHAPING NHL TEAM DEMOGRAPHICS.



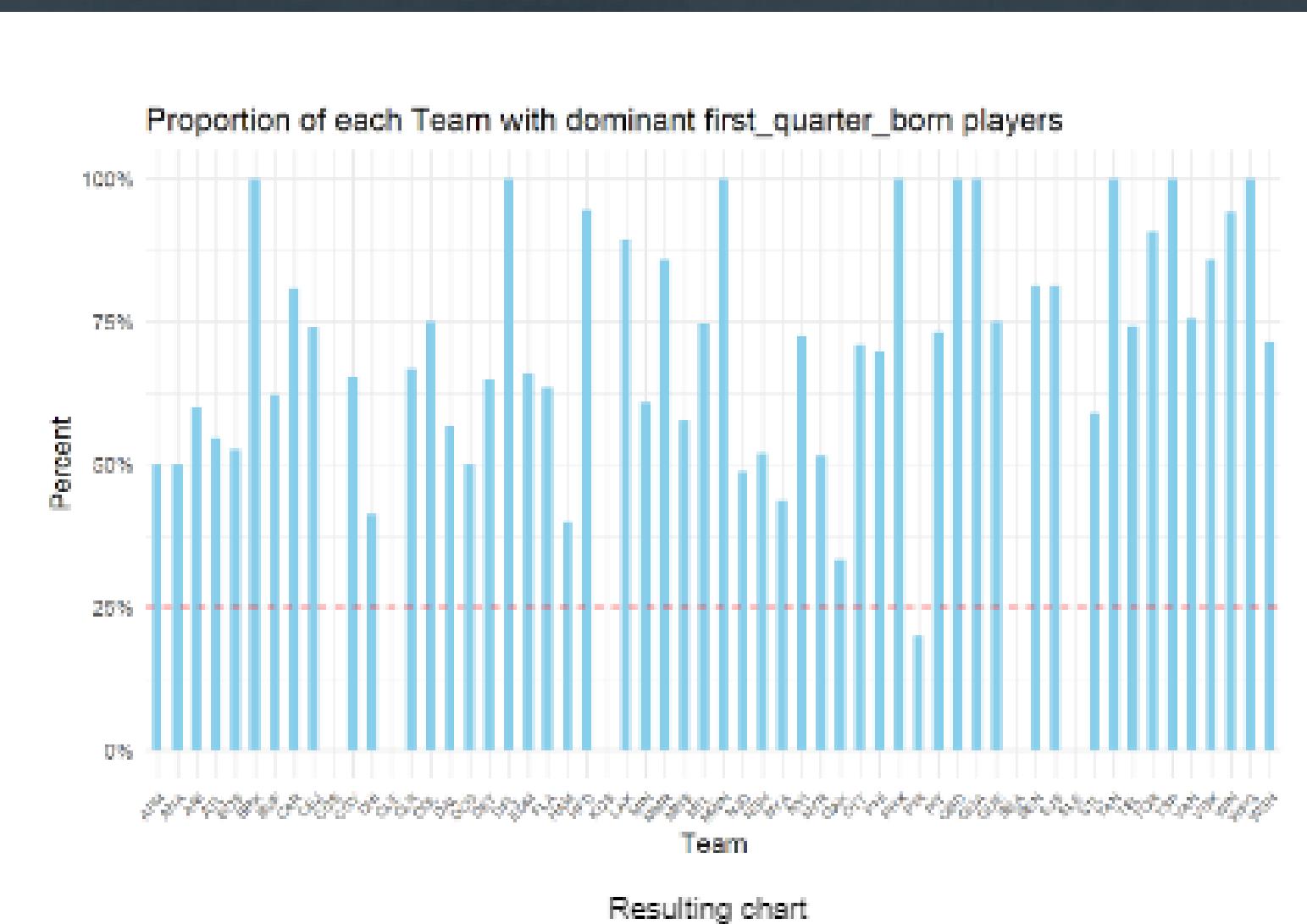
# DISCUSSION



## QUESTION 2: WHETHER MALCOLM GLADWELL'S ASSERTION IS STILL VALID?

SECOND GRAPH: INVESTIGATING MALCOLM'S ASSERTION ACROSS TEAMS.

- MOST TEAMS EXHIBIT PREVALENCE OF FIRST-QUARTER-BORN PLAYERS (>25% OF ROSTER).
- INTRIGUING OUTLIERS: SOME TEAMS WITH 100% REPRESENTATION OF FIRST-QUARTER-BORN PLAYERS.
- DOMINANCE OF INDIVIDUALS BORN WITHIN FIRST THREE MONTHS OF THE YEAR.
- WIDESPREAD TREND ACROSS TEAMS SUPPORTS VALIDITY OF MALCOLM'S ASSERTION ON BIRTH MONTH SIGNIFICANCE IN RECRUITMENT AND TEAM COMPOSITION.



# THANK YOU FOR LISTENING!

