Concussions from 2012-2014 NFL Seasons

Serena Amro Gazze, Arjun Vadlamudi, Anirudhh Venkat

2025-02-17

```r
#Load libraries

suppressPackageStartupMessages({
library(tidyverse)
library(ggplot2)
library(dplyr)
library(forcats)
})
```

Serena Amro Gazze: Analysis of NFL Concussions relative to Positions

```r
#Load NFL Concussions dataset

concussions <- read.csv("C:/Users/samro/Downloads/Concussion Injuries 2012-2014.csv")

concussions <- concussions %>%
  mutate(Position = ifelse(Position == "Comerback", "Cornerback", Position)) %>%
  filter(Position != "Long Snapper")

#Make new column for offensive vs defensive positions
offense_pos <- c("Center", "Offensive Tackle", "Quarterback", "Running Back", "Tight End", "Wide
Receiver", "Guard", "Full Back")
defense_pos <- c("Defensive End", "Defensive Tackle", "Linebacker", "Safety", "Cornerback")

concussions <- concussions %>%
  mutate(Side = case_when(
    Position %in% offense_pos ~ "Offense",
    Position %in% defense_pos ~ "Defense"
  ))

#head(concussions)
#str(concussions)
```

```r
#Filter Concussions dataset into Positions dataset
positions <- concussions %>% select(ID, Position, Side, Reported.Injury.Type, Unknown.Injury., P
re.Season.Injury., Weeks.Injured, Games.Missed, Play.Time.After.Injury)

#head(positions)
```

```
#Write out all the unique positions
position <- positions %>% select(Position)

unique <- unique(position)
#unique

#str(position)
```

```
#Make offense dataset
offense <- positions %>% filter(Side == "Offense")

#head(offense)
#str(offense)
```

```
#Make defense dataset
defense <- positions %>% filter(Side == "Defense")

#head(defense)
#str(defense)
```

Question 1a: Do offensive or defensive football positions result in more head injuries?

```
#Make data frame of numbers of offense vs defensive concussions
position_counts <- data.frame(
  Side = c("Offense", "Defense"),
  Count = c(nrow(offense), nrow(defense))
)

side_counts_plot <- ggplot(position_counts, aes(x = Side, y = Count, fill = Side)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = Count), vjust = -0.5) +
  scale_fill_manual(values = c("Offense" = "blue", "Defense" = "red")) +
  labs(title = "Number of Concussions: Offense vs. Defense",
       x = "Position Group",
       y = "Count of Head Injuries") +
  theme_minimal()

side_counts_plot
```

Number of Concussions: Offense vs. Defense
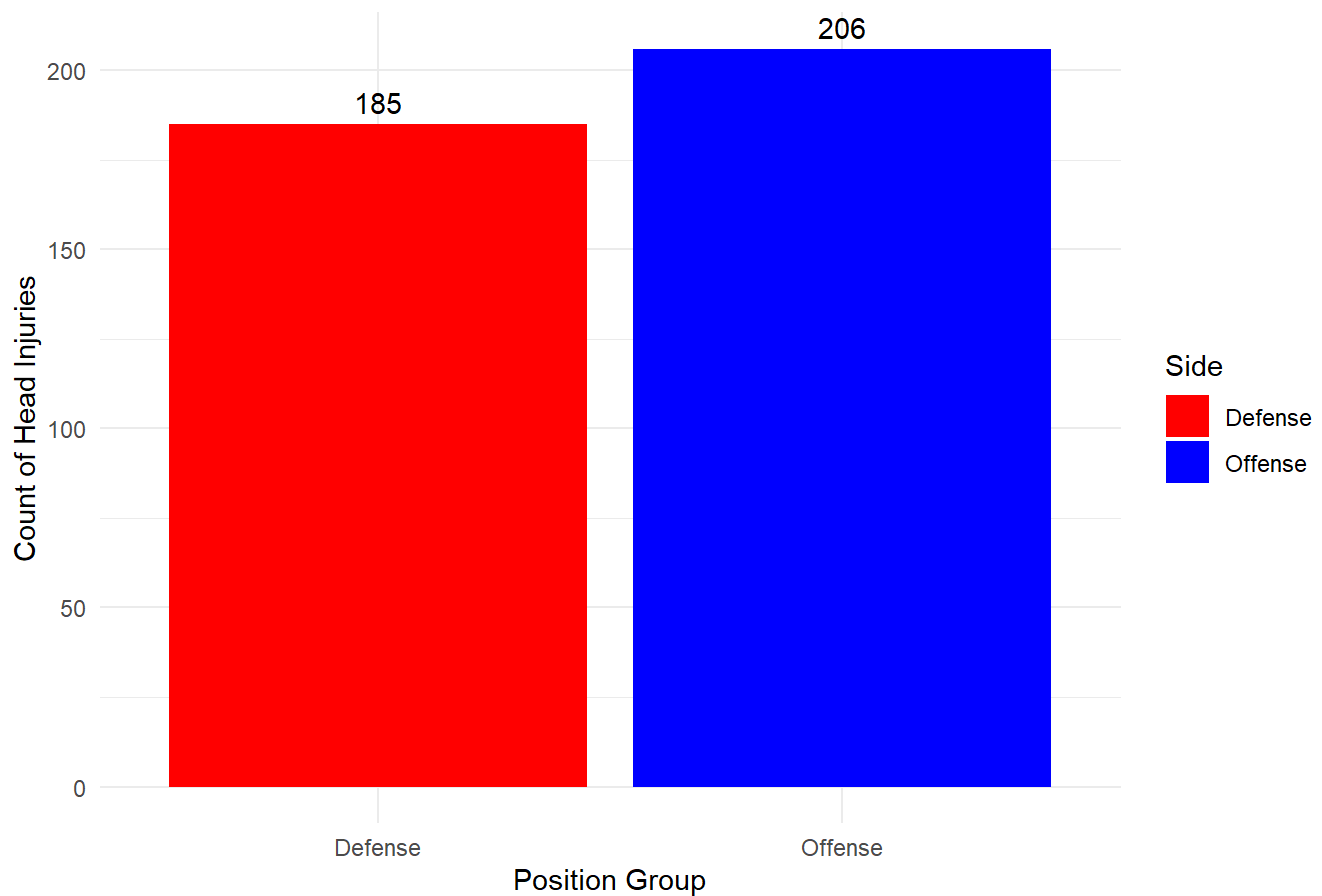
Question 1b: Which positions have the most and least head injuries?

```
#Make full positions plot
#Separate bars by color of offense vs defense

positions_plot <- ggplot(positions, aes(x = fct_infreq(Position), fill = Side)) +
  geom_bar(stat = "count") +
  geom_text(stat = "count", aes(label = after_stat(count)), vjust = -0.5) +
  scale_fill_manual(values = c("Offense" = "blue", "Defense" = "red")) +
  labs(title = "Number of Concussions per Position",
       x = "Position",
       y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

positions_plot
```

Number of Concussions per Position
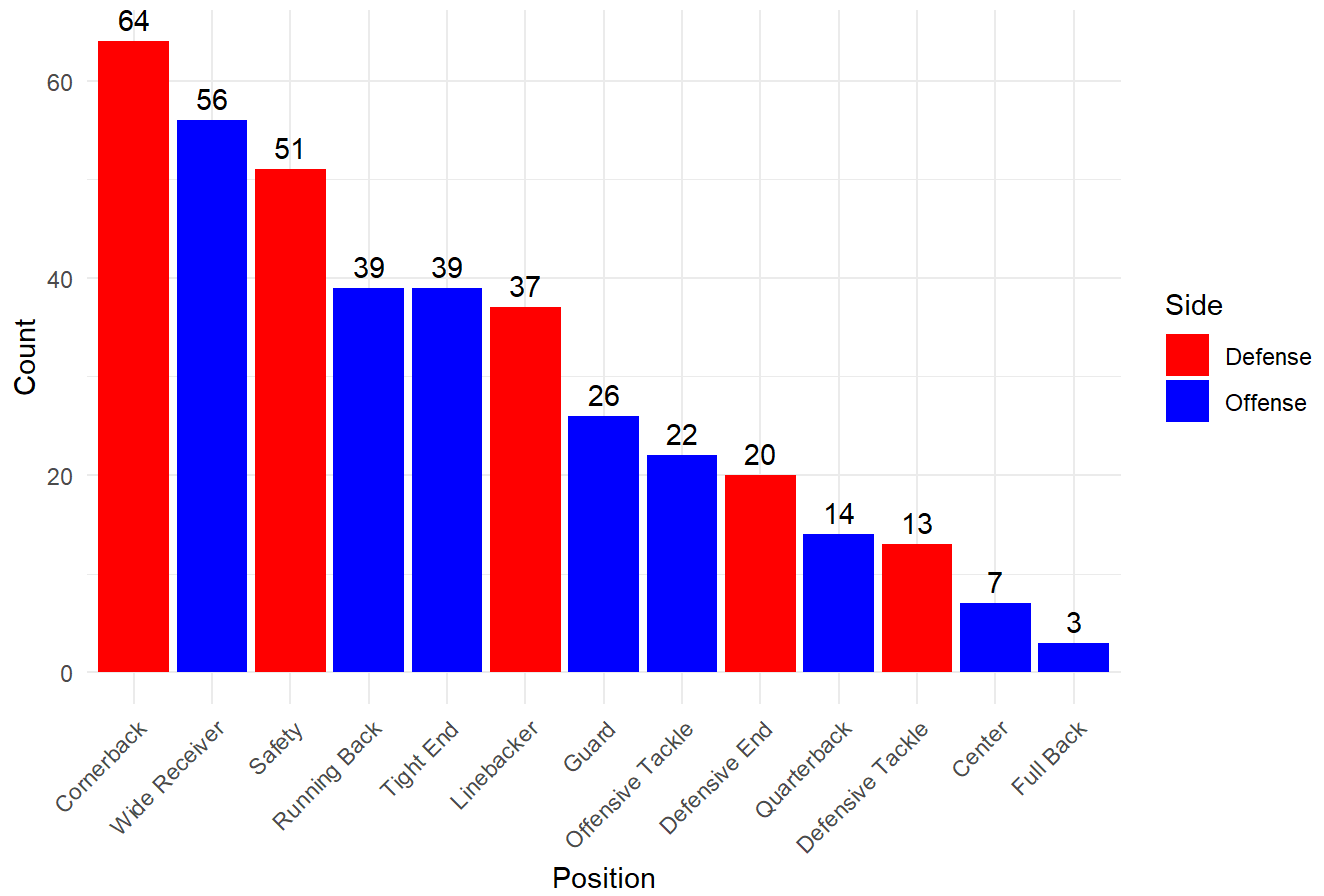
```
#Make data frame of most injured positions
top_injured <- positions %>%
  count(Position, name = "Injury.Count") %>%
  arrange(desc(Injury.Count)) %>%
  slice_head(n = 5)

top_injured <- top_injured %>%
  mutate(Side = case_when(
    Position %in% offense_pos ~ "Offense",
    Position %in% defense_pos ~ "Defense",
  )) %>%
  select(Position, Side, Injury.Count)

#top_injured

#Make data frame of least injured positions
bottom_injured <- positions %>%
  count(Position, name = "Injury.Count") %>%
  arrange(Injury.Count) %>%
  slice_head(n = 5)

bottom_injured <- bottom_injured %>%
  mutate(Side = case_when(
    Position %in% offense_pos ~ "Offense",
    Position %in% defense_pos ~ "Defense",
  )) %>%
  select(Position, Side, Injury.Count)

#bottom_injured
```
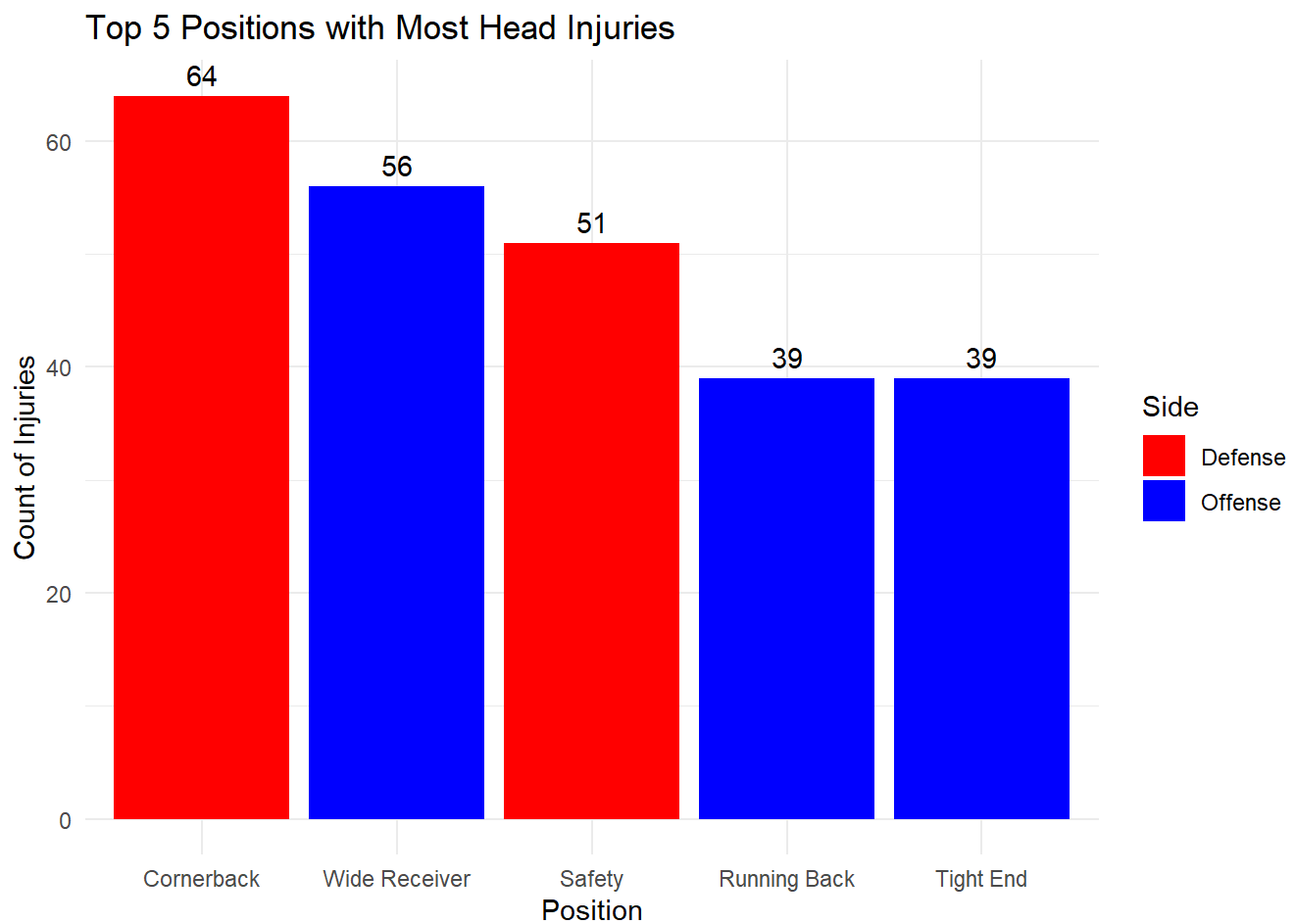
Question 1c: Which positions result in the top 5 numbers of head injuries?

```
#Make top injured plot
top_plot <- ggplot(top_injured, aes(x = reorder(Position, -Injury.Count), y = Injury.Count, fill
= Side)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = Injury.Count), vjust = -0.5) +
  scale_fill_manual(values = c("Offense" = "blue", "Defense" = "red")) +
  labs(title = "Top 5 Positions with Most Head Injuries",
       x = "Position",
       y = "Count of Injuries") +
  theme_minimal()

top_plot
```
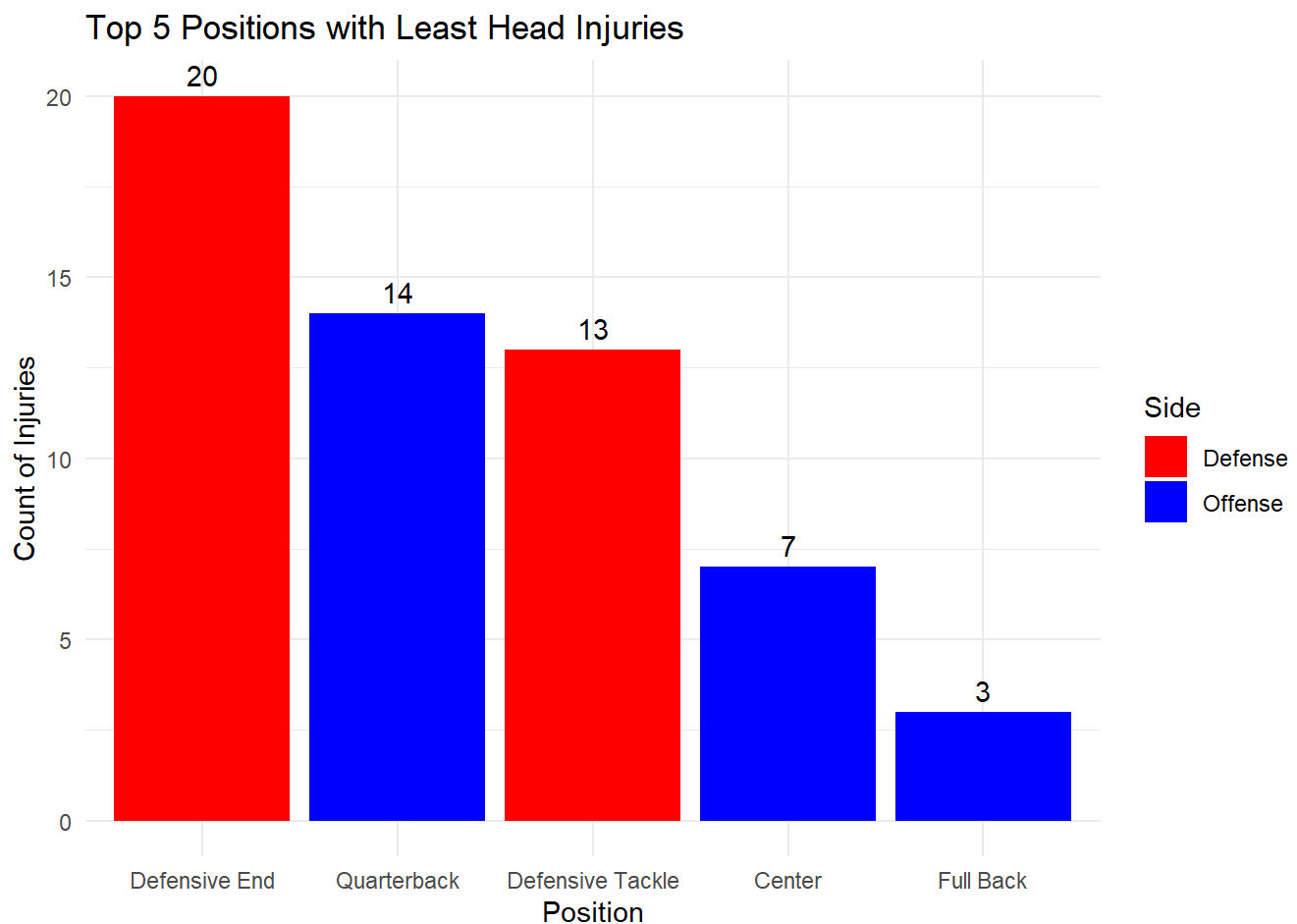
Top 5 Positions with Most Head Injuries

Question 1d: Which positions result in the bottom 5 numbers of head injuries?

```
#Make bottom injured plot
bottom_plot <- ggplot(bottom_injured, aes(x = reorder(Position, -Injury.Count), y = Injury.Count, fill = Side)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = Injury.Count), vjust = -0.5) +
  scale_fill_manual(values = c("Offense" = "blue", "Defense" = "red")) +
  labs(title = "Top 5 Positions with Least Head Injuries",
       x = "Position",
       y = "Count of Injuries") +
  theme_minimal()

bottom_plot
```
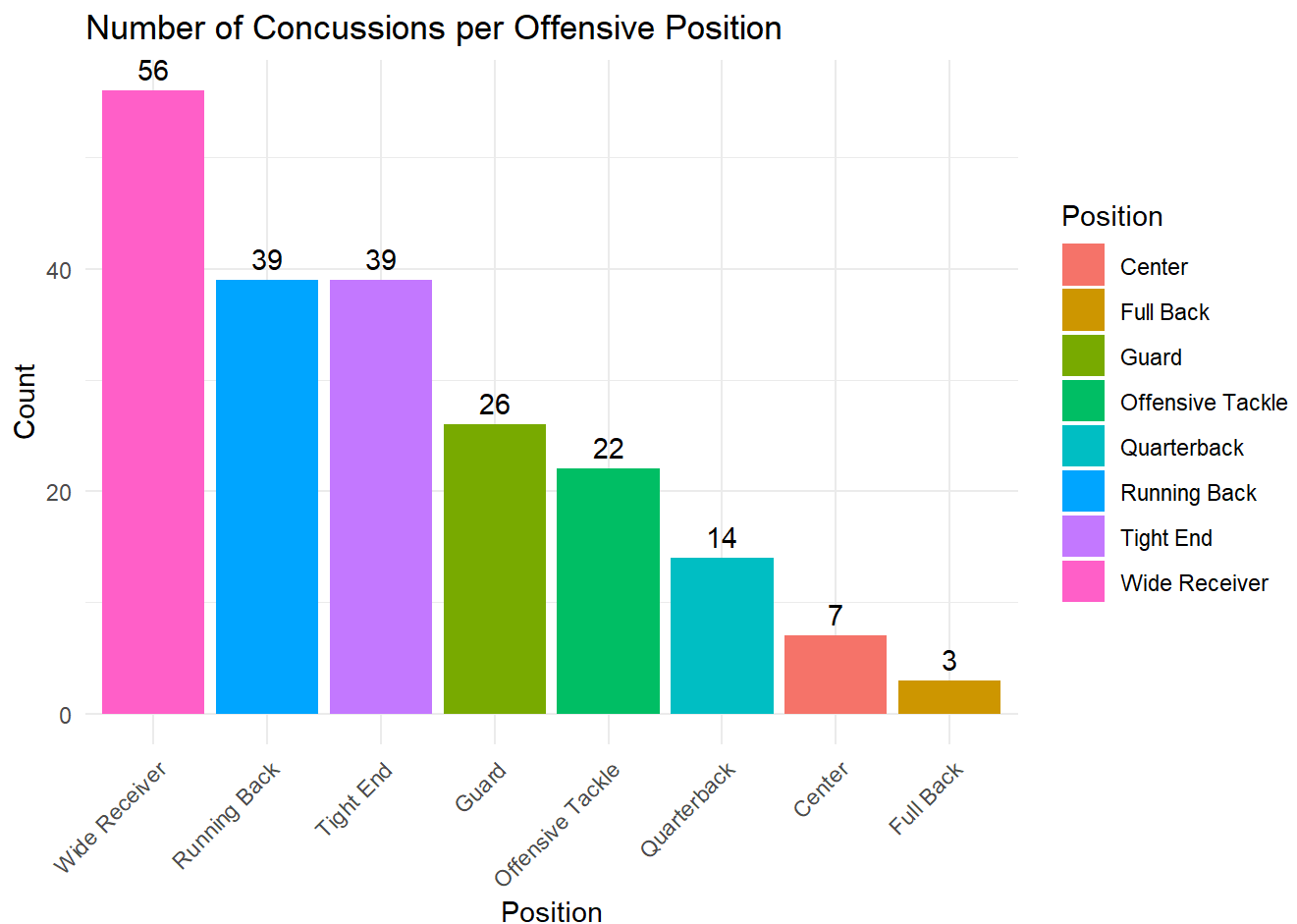
Top 5 Positions with Least Head Injuries

Question 1e: Which offensive positions result in the most and least head injuries?

```
#Make offensive positions plot
library(forcats)

offensive_plot <- ggplot(offense, aes(x = fct_infreq(Position), fill = Position)) +
  geom_bar(stat = "count") +
  geom_text(stat = "count", aes(label = after_stat(count)), vjust = -0.5) +
  labs(title = "Number of Concussions per Offensive Position",
       x = "Position",
       y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

offensive_plot
```

Number of Concussions per Offensive Position

Question 1f: Which defensive positions result in the most and least head injuries?

```
#Make defensive positions plot
defensive_plot <- ggplot(defense, aes(x = fct_infreq(Position), fill = Position)) +
  geom_bar(stat = "count") +
  geom_text(stat = "count", aes(label = after_stat(count)), vjust = -0.5) +
  labs(title = "Number of Concussions per Defensive Position",
       x = "Position",
       y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

defensive_plot
```

Number of Concussions per Defensive Position
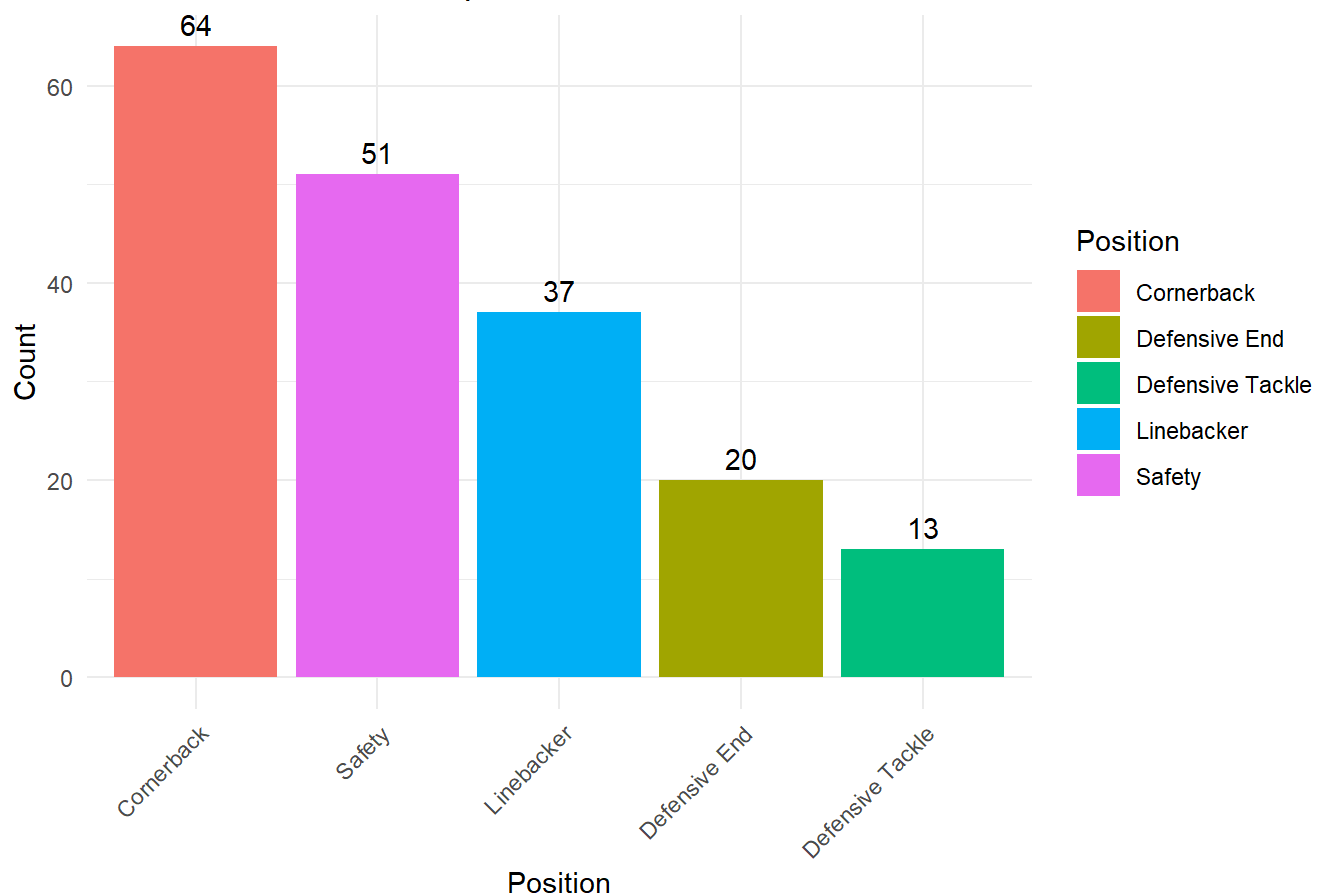

```r
#Get number of injuries per season
twelve <- concussions %>% select(ID, Season, Position, Side, Reported.Injury.Type, Unknown.Injury., Pre.Season.Injury., Weeks.Injured, Games.Missed, Play.Time.After.Injury)
twelve <- twelve %>% filter(Season == "2012/2013")
twelve_count <- nrow(twelve)
#twelve_count

thirteen <- concussions %>% select(ID, Season, Position, Side, Reported.Injury.Type, Unknown.Injury., Pre.Season.Injury., Weeks.Injured, Games.Missed, Play.Time.After.Injury)
thirteen <- thirteen %>% filter(Season == "2013/2014")
thirteen_count <- nrow(thirteen)
#thirteen_count

fourteen <- concussions %>% select(ID, Season, Position, Side, Reported.Injury.Type, Unknown.Injury., Pre.Season.Injury., Weeks.Injured, Games.Missed, Play.Time.After.Injury)
fourteen <- fourteen %>% filter(Season == "2014/2015")
fourteen_count <- nrow(fourteen)
#fourteen_count
```

```
#Make data frame of number of injuries per season
year_counts <- data.frame(
    Season = c("2012/2013", "2013/2014", "2014/2015"),
    Concussion_Count = c(twelve_count, thirteen_count, fourteen_count)
)

#year_counts
```

Question 1g: Is there a trend in concussion frequency over different seasons (2012-2014 seasons)?

```
#Plot concussions per season
year_plot <- ggplot(year_counts, aes(x = reorder(Season, -Concussion_Count), y = Concussion_Coun
t, fill = Season)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = Concussion_Count), vjust = -0.5) +
  labs(title = "Number of Concussions Per Season",
       x = "Season",
       y = "Concussion Count") +
  theme_minimal()

year_plot
```
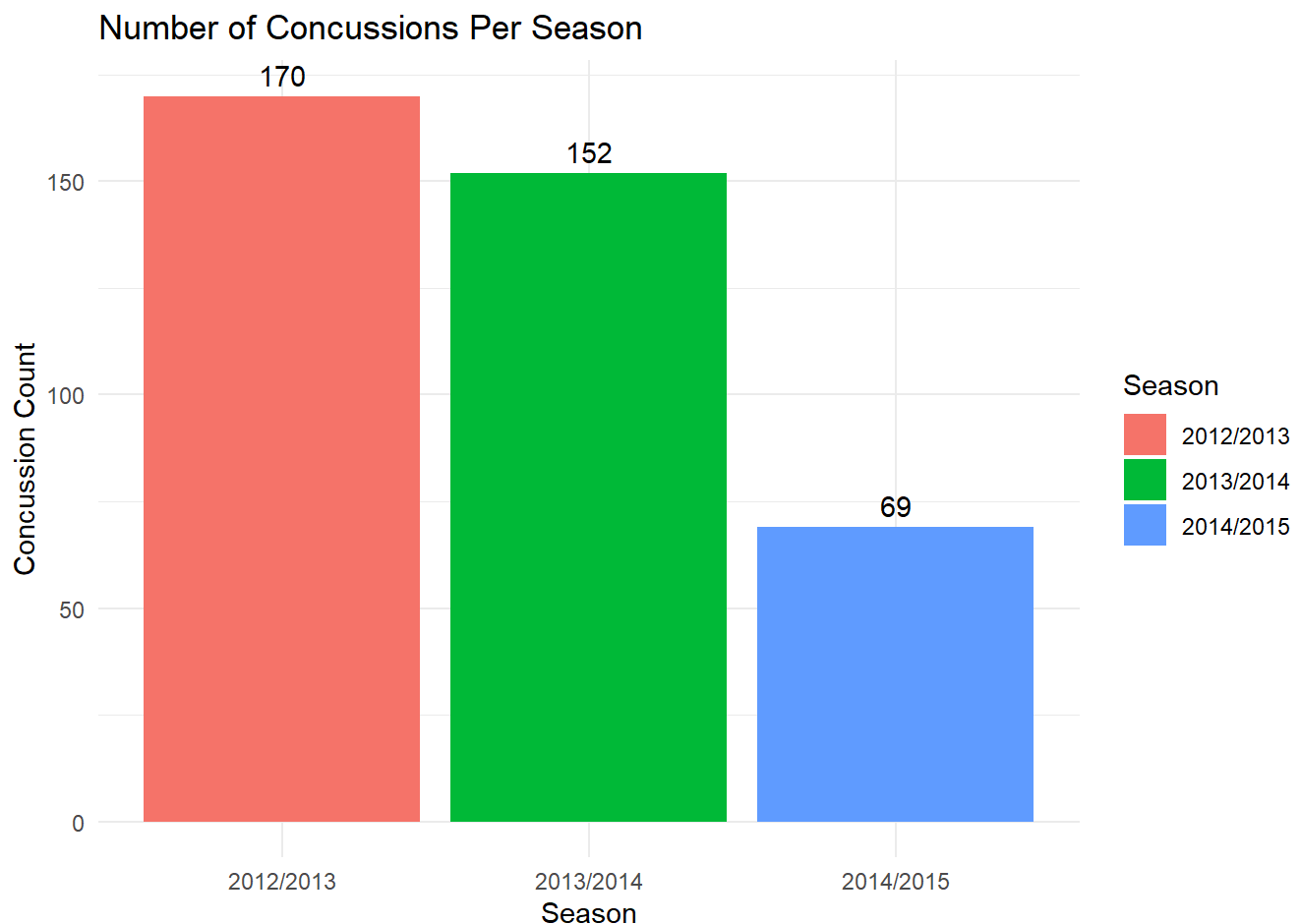
```
#Get number of specifically offense or defense positions per year
twelve_off <- twelve %>% filter(Side == "Offense")
twelve_off_count <- nrow(twelve_off)
#twelve_off_count

twelve_def <- twelve %>% filter(Side == "Defense")
twelve_def_count <- nrow(twelve_def)
#twelve_def_count

thirteen_off <- thirteen %>% filter(Side == "Offense")
thirteen_off_count <- nrow(thirteen_off)
#thirteen_off_count

thirteen_def <- thirteen %>% filter(Side == "Defense")
thirteen_def_count <- nrow(thirteen_def)
#thirteen_def_count

fourteen_off <- fourteen %>% filter(Side == "Offense")
fourteen_off_count <- nrow(fourteen_off)
#fourteen_off_count

fourteen_def <- fourteen %>% filter(Side == "Defense")
fourteen_def_count <- nrow(fourteen_def)
#fourteen_def_count
```

```
#Make data frames of number of specifically offense or defense positions per year
off_year_counts <- data.frame(
  Season = c("2012/2013", "2013/2014", "2014/2015"),
  Concussion_Count = c(twelve_off_count, thirteen_off_count, fourteen_off_count)
)

def_year_counts <- data.frame(
  Season = c("2012/2013", "2013/2014", "2014/2015"),
  Concussion_Count = c(twelve_def_count, thirteen_def_count, fourteen_def_count)
)

#off_year_counts
#def_year_counts
```
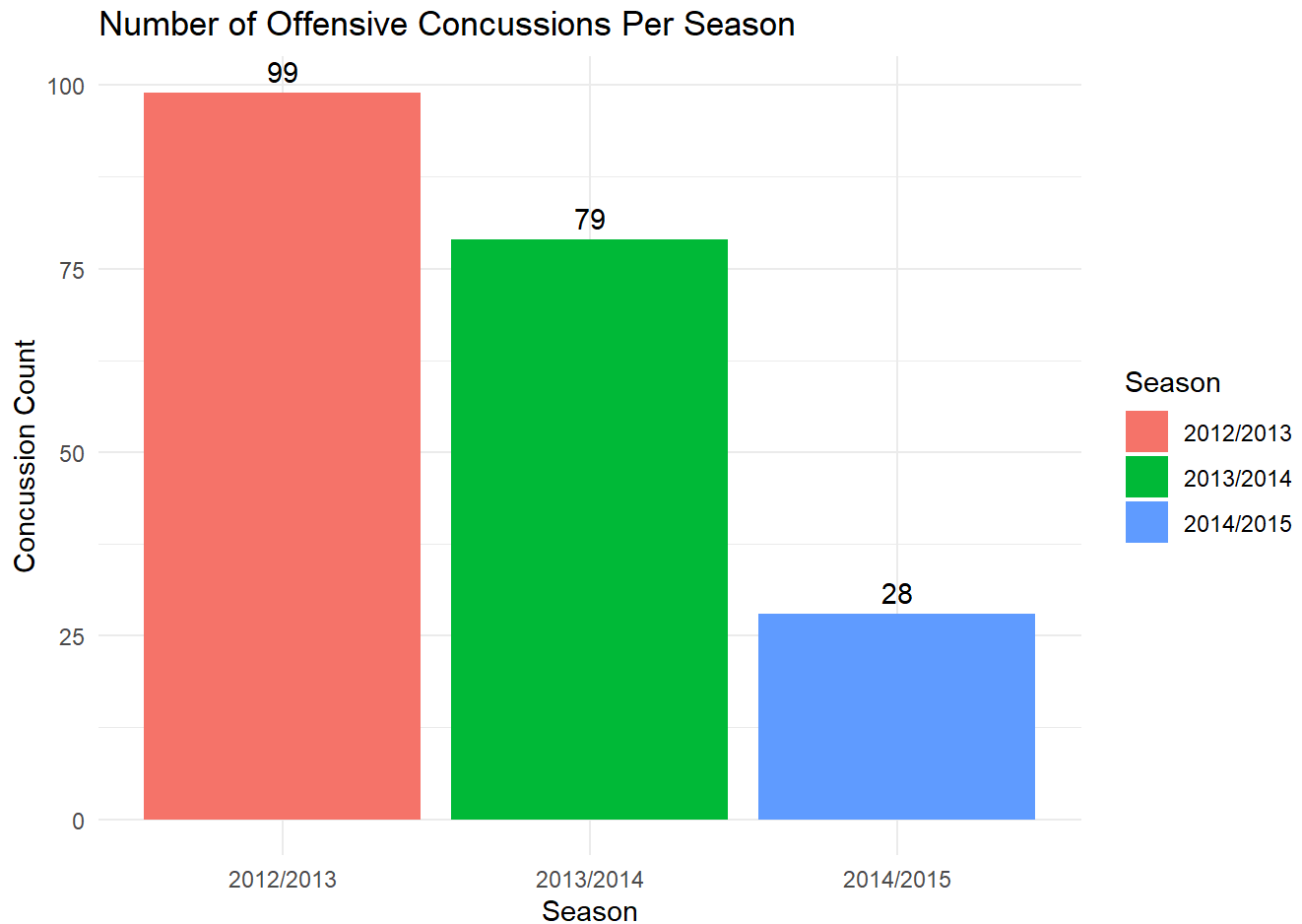
Question 1h: Is there a trend in concussion frequency for offensive positions over different seasons (2012-2014)?

```
#Make plot of numbers of offensive injuries per season
off_years_plot <- ggplot(off_year_counts, aes(x = Season, y = Concussion_Count, fill = Season))
+
  geom_bar(stat = "identity") +
  geom_text(aes(label = Concussion_Count), vjust = -0.5) +
  labs(title = "Number of Offensive Concussions Per Season",
       x = "Season",
       y = "Concussion Count") +
  theme_minimal()

off_years_plot
```


Question 1i: Is there a trend in concussion frequency for defensive positions over different seasons (2012-2014)?

```
#Make plot of numbers of defensive injuries per season
def_years_plot <- ggplot(def_year_counts, aes(x = Season, y = Concussion_Count, fill = Season))
+
  geom_bar(stat = "identity") +
  geom_text(aes(label = Concussion_Count), vjust = -0.5) +
  labs(title = "Number of Defensive Concussions Per Season",
       x = "Season",
       y = "Concussion Count") +
  theme_minimal()

def_years_plot
```
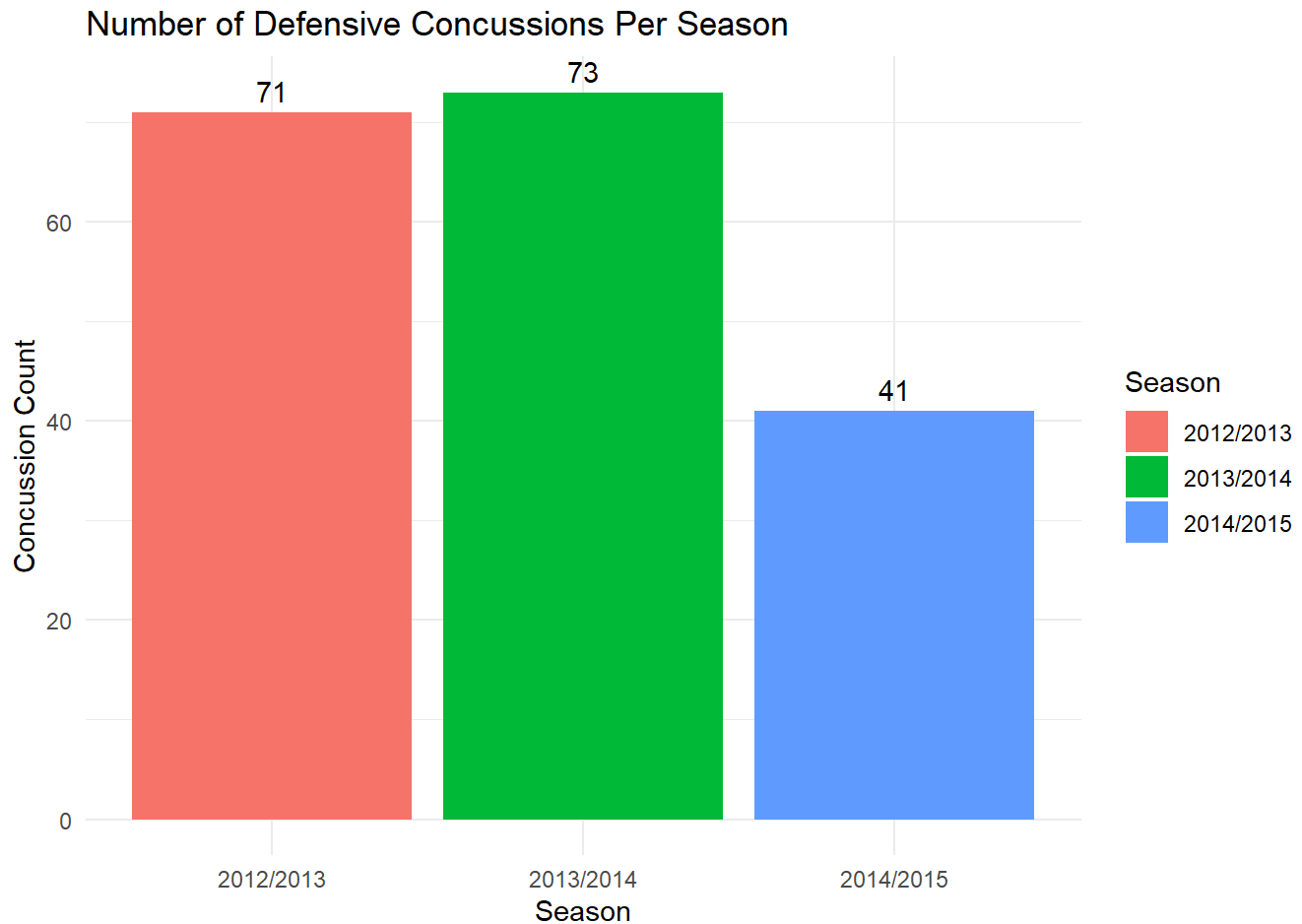

Arjun Vadlamudi: Analysis of Injury Type relative to Seasons

Question 2a: How does the season correlate with injury type?

```
# Using group_by function to group data by season and injury type
injury_counts_by_season <- concussions %>%
  group_by(Season, Reported.Injury.Type) %>%
  count(Season, name = "InjuryCount")

# Create bar chart
ggplot(injury_counts_by_season, aes(x = Season, y = InjuryCount, fill = Reported.Injury.Type)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Number of Injury Types per Season",
       x = "Season",
       y = "Number of Injuries",
       fill = "Injury Type") +
         theme_bw()
```


Number of Injury Types per Season
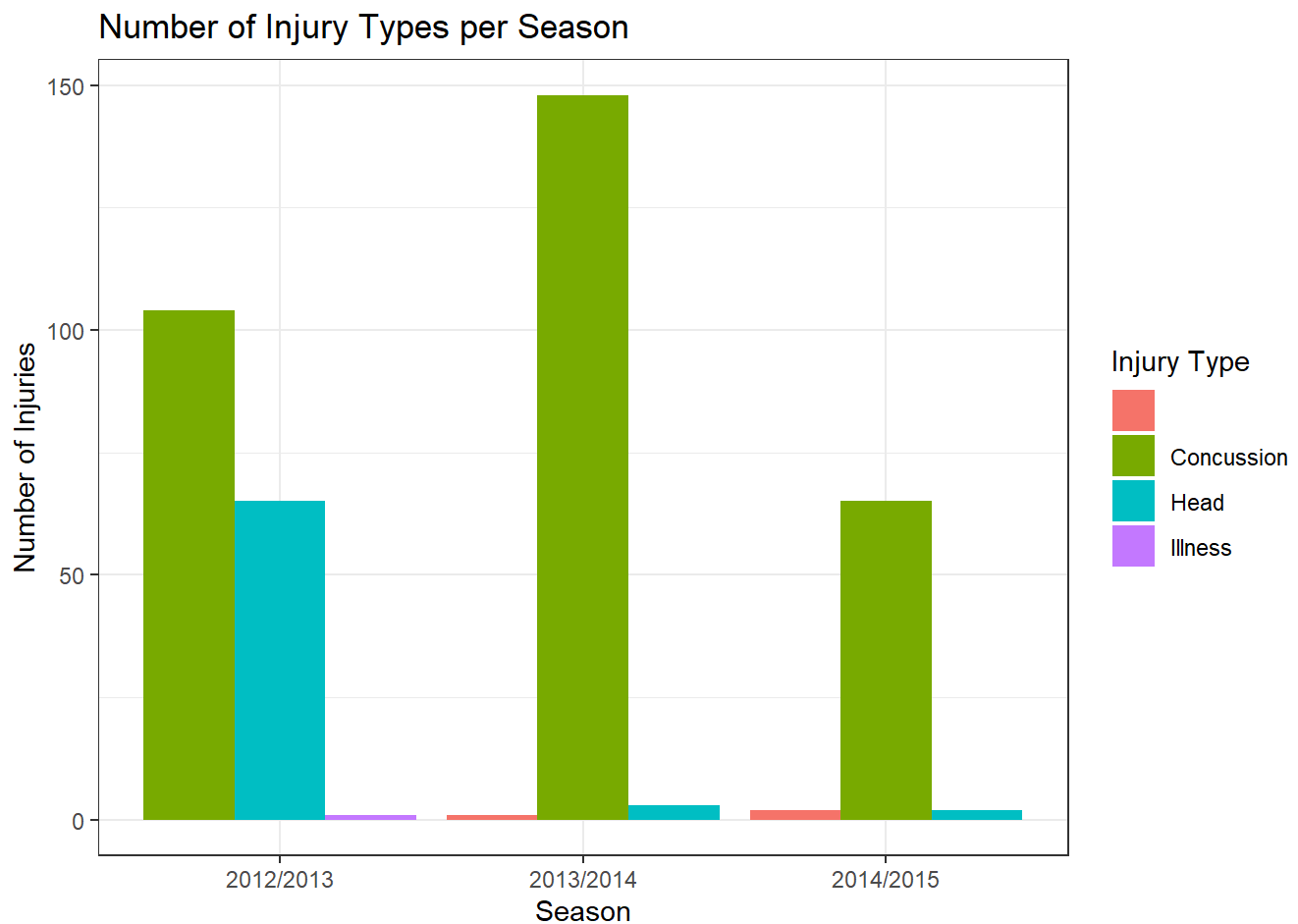
Question 2b: How does injury type affect team's performance?

```
# Using group_by function to group data by Winning Team and injury type
injury_win_counts <- concussions %>%
  group_by(Winning.Team., Reported.Injury.Type) %>%
  count(Winning.Team., name = "InjuryCount")

# Create bar chart
ggplot(injury_win_counts, aes(x = Reported.Injury.Type, y = InjuryCount, fill = Winning.Team.))
+
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Injury Type vs. Winning Team",
       x = "Injury Type",
       y = "Number of Injuries",
       fill = "Winning Team") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```
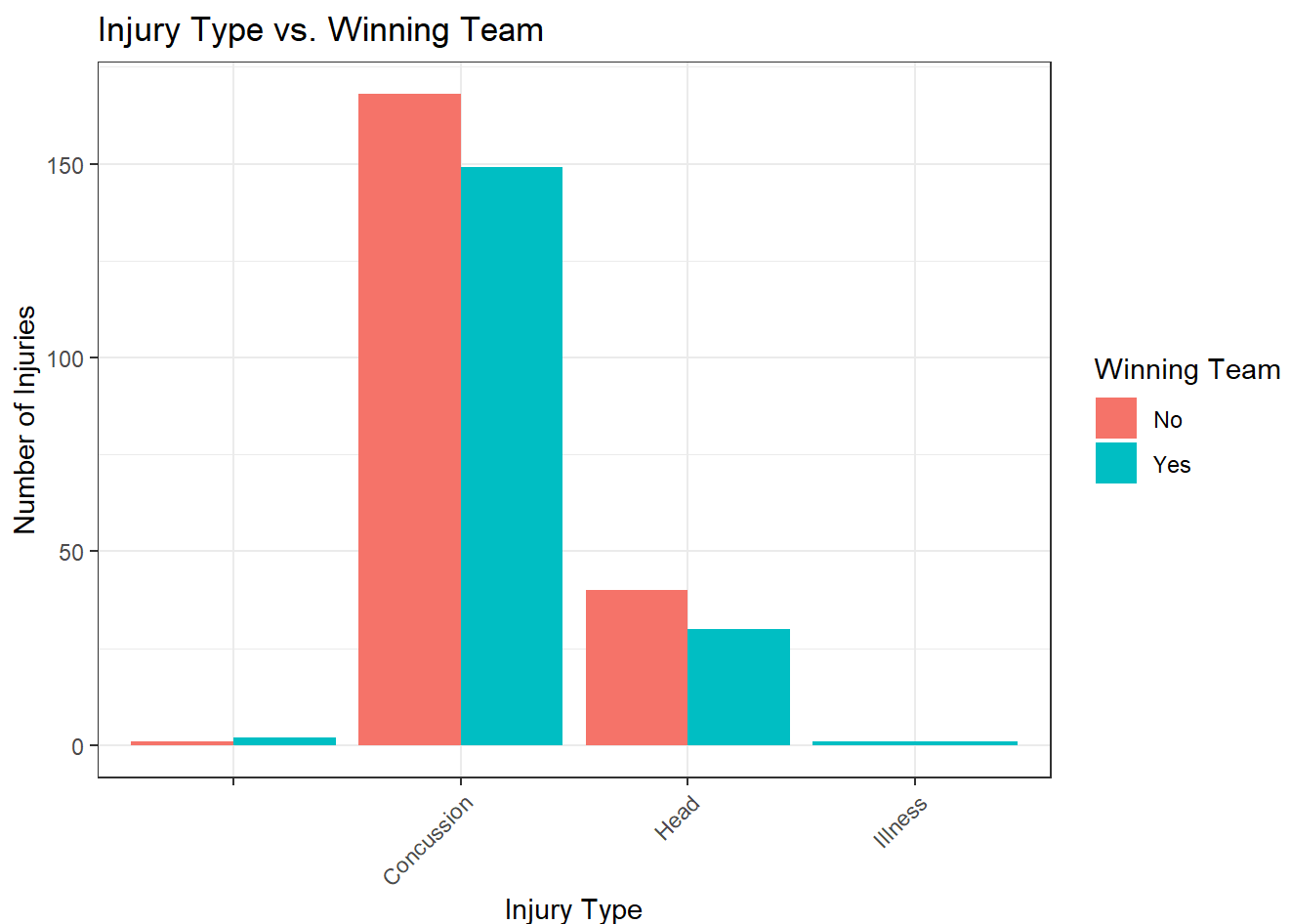

Anirudhh Venkat: Analysis of Teams Causing the Most Head Injuries

Question 3a: Which team has the most injuries?

```r
# create "other" injuries
concussions <- concussions %>%
  mutate(Reported.Injury.Type = ifelse(Reported.Injury.Type == "",
                                        "Other", Reported.Injury.Type))

# get top 10 teams
top_teams <- concussions %>%
  group_by(Team) %>%
  summarize(TotalInjuries = n(), .groups = "drop") %>%
  arrange(desc(TotalInjuries)) %>%
  slice_head(n = 10)

# filter for top 10 teams and by injury
injury_counts_by_team_injury <- concussions %>%
  filter(Team %in% top_teams$Team) %>%
  group_by(Team, Reported.Injury.Type) %>%
  summarize(InjuryCount = n(), .groups = "drop") %>%
  left_join(top_teams, by = "Team")

# plot bar plot
ggplot(injury_counts_by_team_injury, aes(x = reorder(Team, -TotalInjuries), y = InjuryCount, fil
l = Reported.Injury.Type)) +
  geom_bar(stat = "identity") +
  labs(title = "Top 10 Teams with Most Injuries",
       x = "Team",
       y = "Number of Injuries",
       fill = "Injury Type") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```
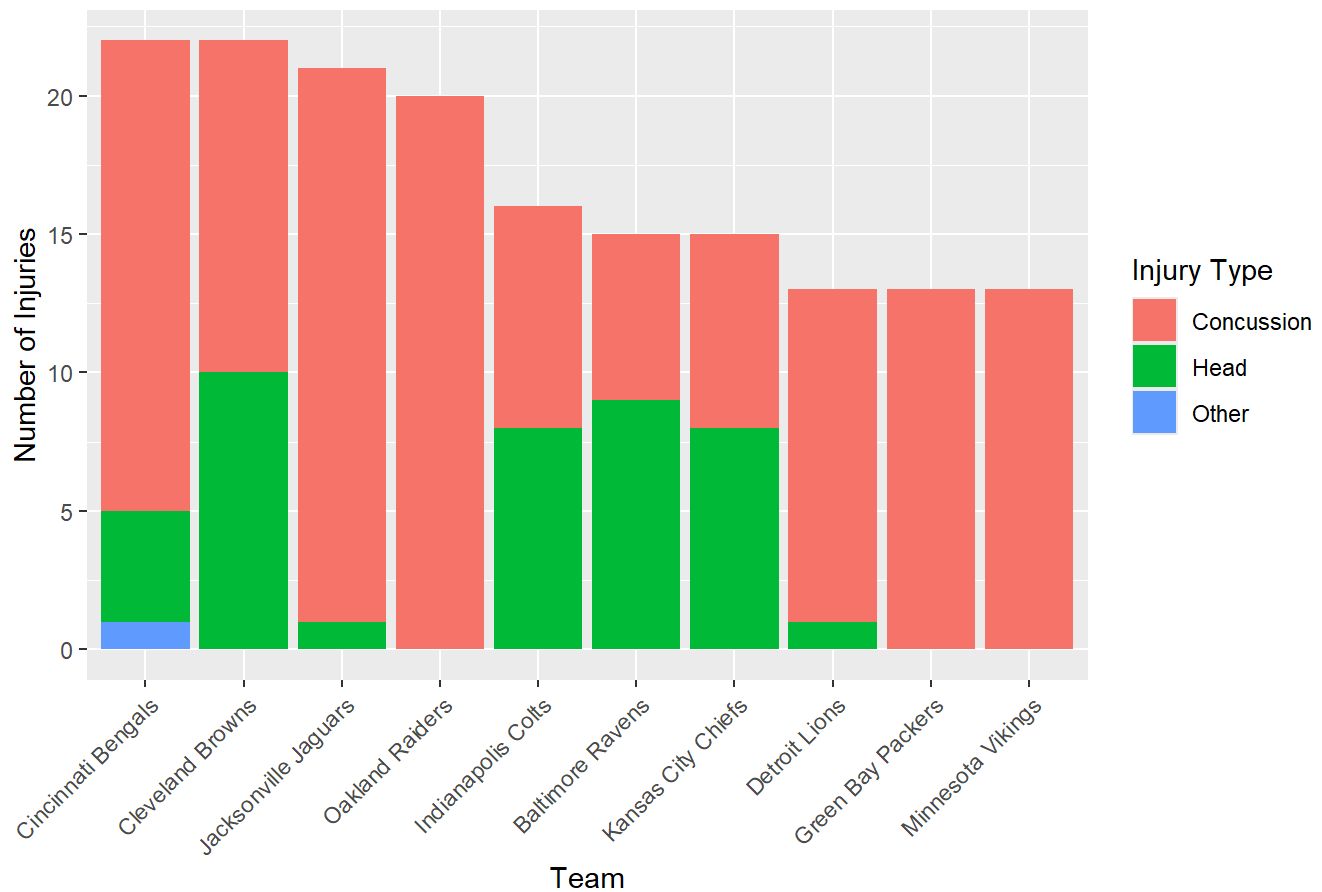
Top 10 Teams with Most Injuries

Question 3b: Which team causes the most injuries?

```r
# get top 10 injury counts by opposing teams
team_injury_count <- concussions %>%
  group_by(Opposing.Team) %>%
  summarise(Injury_Count = n()) %>%
  arrange(desc(Injury_Count)) %>%
  slice_head(n = 10)

# plot bubble chart
ggplot(team_injury_count, aes(x = reorder(Opposing.Team, -Injury_Count), y = Injury_Count, size
= Injury_Count)) +
  geom_point(alpha = 0.7) +
  labs(title = "Injury Counts by Opposing Team",
       x = "Opposing Team",
       y = "Injury Count") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  # control bubble sizes
  scale_size_continuous(range = c(5, 15))
```

Injury Counts by Opposing Team
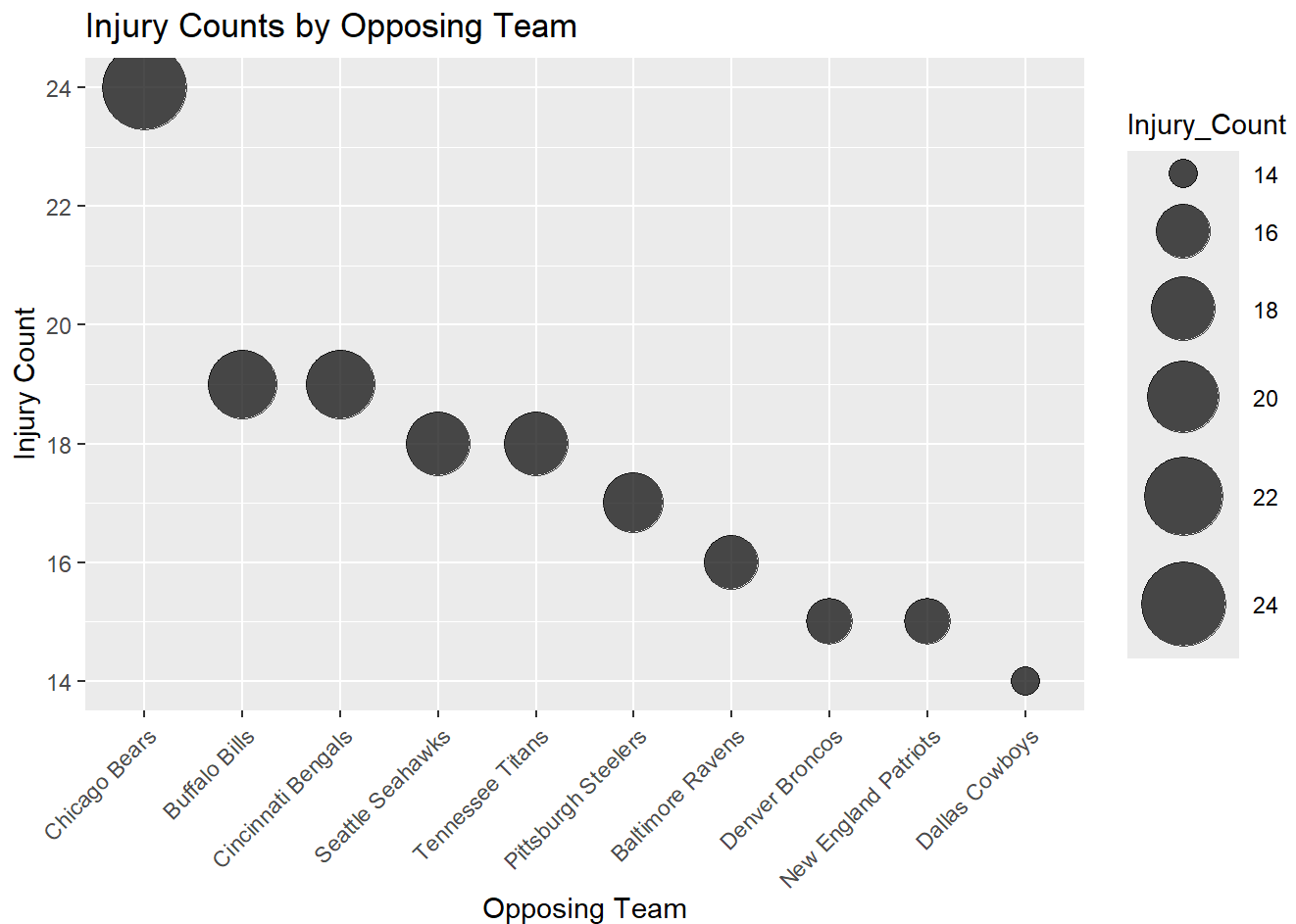
Conclusion

This analysis of NFL concussions from the 2012-2014 seasons provides valuable insights into how head injuries are distributed across positions, injury types, and teams within the NFL. By examining the data from multiple perspectives, it allows us to better understand the patterns and trends associated with concussions in professional football.

Regarding the question of whether offensive or defensive positions result in more head injuries in the NFL, we hypothesized that offensive positions would cause more head injuries as offensive players are frequently tackled aggressively by defensive players. All together, the data supported our hypothesis, as all offensive positions combined had more total head injuries than all defensive positions combined. However, there were more offensive positions inlcuded in the dataset. So, if we look at the average number of head injuries per offensive or defensive position, defensive positions had more head injuries per position on average.

Additionally, our data analysis clearly showed that the number of head injuries decreased per season between the 2012 through 2014 seasons. This trend could be attributed to increased awareness of concussions, improved safety protocols, and rule changes aimed at reducing head injuries in the NFL. The decrease in head injuries over time suggests that the league's efforts to address player safety and concussion prevention have been effective. However, the continued prevalence of concussions in high-contact roles underscores the need for continuous evaluation of protective equipment and concussion protocols.

When analyzing the impact of head injuries on team performance, we hypothesized that players with head injuries would cause their teams to lose more games. The results confirmed this expectation, showing that for both concussions and head injuries, the head injuries led to that team losing more often than winning. This finding highlights the significant impact of head injuries on player availability and team performance, underscoring the importance of player safety and injury prevention in professional football.

We were also very curious to see which teams accumulated the most amount of injuries and which teams caused the most amount of injuries during these NFL seasons. The data showed that the Cincinnati Bengals and Cleveland Browns had the most amount of players with head injuries, while the Chicago Bears and Buffalo Bills caused the most amount of head injuries. This data could be used to help teams understand where they need to improve in terms of player safety and injury prevention.

Regarding problems and solutions in our code, we encountered some issues with the data cleaning process, such as correcting typos in the Position column and dealing with missing and duplicate values within the dataset. We then resolved some of these data cleaning issues using code such as count() and identity(). We also discover how to use some new functions while debugging, such as mutate(), arrange(), and slice(). For some of the plots, we also had to learn how to make/add new columns to the dataset and make whole new data frames based off of the info from the concussions dataset. All of these problems that we had to adapt to and solve helped us to better understand the data and how to manipulate it to get our desired results.

For future data analyses projects, we would like to look at more recent concussion data from the NFL and compare the current numbers to the data from the 2012-2014 seasons. This would allow us to see if the trends we observed in our analysis have continued or changed over time. We would also like to explore the long-term effects of concussions on player health and career longevity (such as developing CTE), as this is a critical issue facing the NFL and other contact sports leagues.

The findings emphasize the importance of ongoing research and policy adjustments in the NFL to mitigate risks associated with head injuries. As concussion awareness continues to grow, leagues, teams, and medical professionals must work collaboratively to enhance player safety while maintaining the integrity of the sport.

Thank you! The end!