

Tendencias de peleas en la NHL a lo largo del tiempo

Tania Gómez

2024-11-08

Análisis de Tendencias de peleas en la NHL a lo largo del tiempo

Contexto y Objetivo

Este análisis tiene como objetivo explorar las tendencias y patrones en las peleas ocurridas en la NHL durante las últimas 25 temporadas. A través del análisis de datos históricos, se analizan varios aspectos clave:

- Peleas por temporada: la cantidad de peleas en las temporadas regulares y de playoffs, observando si ha habido una tendencia creciente o decreciente a lo largo del tiempo.
- Peleas por jugador: se identifican los jugadores más involucrados en peleas, desde los delanteros y “ejecutores” tradicionales hasta las estrellas del hockey con un estilo de juego más técnico.
- Peleas por equipo: se investiga qué equipos han sido los más agresivos, y si esta tendencia ha influido en su rendimiento global a lo largo de las temporadas.
- Evolución de peleas por periodos y tiempo de juego: se examina cómo se distribuyen las peleas a lo largo de los diferentes periodos del partido y cómo varían según el tiempo de juego, destacando así los momentos más propensos a la agresión física.

Preguntas de análisis

- ¿Hay una tendencia creciente o decreciente en las peleas a lo largo de los años?
- ¿Cuáles son los jugadores y equipos más involucrados en peleas?
- ¿Quién gana más peleas, y hay alguna tendencia entre los equipos?
- ¿Cuáles son los equipos más involucrados en peleas?
- ¿Quién gana más peleas, y hay alguna tendencia entre los equipos?
- ¿En qué periodo del partido ocurren más peleas?
- ¿Cuándo se dan más peleas, al inicio, mitad o final del partido?

Datos utilizados

El conjunto de datos proviene de Kaggle y se ha obtenido a través de scrapping en [hockeyfights.com](https://www.kaggle.com/datasets/tgomez123/nhl-fights). “All NHL regular season fights from 99/00 to 23/24” (CCO: Public Domain) contiene todas las peleas de las últimas 25 temporadas con todos los detalles necesarios. Tiene una actualización anual y la última, en cara a este proyecto, fue el 01/11/2024.

Los datos contienen las siguientes columnas:

- season: años de la temporada
- season_type: tipo de temporada
- fight: participantes en la pelea

- player_1_name: nombre del jugador uno
- player_2_name: nombre del jugador dos
- player_1_team: equipo del jugador uno
- player_2_team: equipo del jugador dos
- date: fecha de la pelea
- period: período durante el cual ocurrió la pelea
- gametime: tiempo dentro del período
- winner: resultado de la encuesta de hockeyfights.com
- rating: calificaciones de los espectadores de hockeyfights.com
- vote_count: número de votos de los espectadores de hockeyfights.com
- fight_id: identificador único para el registro

Los datos se encuentran en archivos CSV y han sido descargados desde la plataforma de Kaggle. Los datos se mantienen en una sola hoja de cálculo en formato ancho.

No representan un riesgo de privacidad. La seguridad se mantiene al trabajar en un entorno protegido, sin datos de identificación personal sensible.

Para verificar la integridad de los datos, estos se limpiaron de duplicados, se validaron formatos y se verificaron columnas problemáticas cómo pueden ser caracteres extraños.

Procesamiento de los datos

Pasos de procesamiento

08/11/2024

- Eliminar duplicados y verificar registros nulos.
- Convertir formatos de fecha y hora.
- Estandarizar nombres de columnas.

Análisis de datos

08/11/2024

- Carga de los datos
- Análisis de los datos.
- Creación de visualizaciones.
- Resumen del análisis.
- Hallazgos clave.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr    1.5.1
## v ggplot2    3.5.1      v tibble     3.2.1
## v lubridate  1.9.3      v tidyr      1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(tinytex)
library(ggplot2)
library(readr)
```

```
peleas <- read_csv("~/NHL_regular_fights/NHL_regular_fights.csv")
```

```
## Rows: 11535 Columns: 14
## -- Column specification -----
## Delimiter: ","
## chr  (11): season, season_type, fight, player_1_name, player_2_name, player_...
## dbl  (2): vote_count, fight_id
## time (1): gametime
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
view(peleas)
head(peleas)
```

```
## # A tibble: 6 x 14
##   season    season_type fight      player_1_name player_2_name player_1_team
##   <chr>      <chr>      <chr>      <chr>      <chr>      <chr>
## 1 1999-2000 reg      B. Berard (TO~ B. Berard      P. Poulin      Toronto Mapl~
## 2 1999-2000 reg      R. Salei (ANA~ R. Salei      B. Hull        Anaheim Ducks
## 3 1999-2000 reg      D. Langdon (N~ D. Langdon      G. Laraque      New York Ran~
## 4 1999-2000 reg      S. Passmore (~ S. Passmore      S. Shields      Chicago Blac~
## 5 1999-2000 reg      D. Zmolek (CH~ D. Zmolek      P. Marleau      Chicago Blac~
## 6 1999-2000 reg      R. McKay (NJD~ R. McKay      C. Tamer        New Jersey D~
## # i 8 more variables: player_2_team <chr>, date <chr>, period <chr>,
## #   gametime <time>, winner <chr>, rating <chr>, vote_count <dbl>,
## #   fight_id <dbl>
```

```
str(peleas)
```

```
## spc_tbl_ [11,535 x 14] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ season      : chr [1:11535] "1999-2000" "1999-2000" "1999-2000" "1999-2000" ...
##  $ season_type : chr [1:11535] "reg" "reg" "reg" "reg" ...
##  $ fight       : chr [1:11535] "B. Berard (TOR) vs. P. Poulin (MON)" "R. Salei (ANA) vs. B. Hull (D..."
##  $ player_1_name: chr [1:11535] "B. Berard" "R. Salei" "D. Langdon" "S. Passmore" ...
##  $ player_2_name: chr [1:11535] "P. Poulin" "B. Hull" "G. Laraque" "S. Shields" ...
##  $ player_1_team: chr [1:11535] "Toronto Maple Leafs" "Anaheim Ducks" "New York Rangers" "Chicago Bl..."
##  $ player_2_team: chr [1:11535] "Montreal Canadiens" "Dallas Stars" "Edmonton Oilers" "San Jose Shar..."
##  $ date        : chr [1:11535] "2/10/1999" "2/10/1999" "1/10/1999" "4/10/1999" ...
##  $ period      : chr [1:11535] "2nd Period" "2nd Period" "1st Period" "3rd Period" ...
##  $ gametime    : 'hms' num [1:11535] 16:32:00 04:00:00 13:01:00 18:45:00 ...
##  ..- attr(*, "units")= chr "secs"
##  $ winner      : chr [1:11535] "Bryan Berard" "Brett Hull" "Darren Langdon" "Steve Passmore" ...
##  $ rating      : chr [1:11535] "4,00" "7,00" "5,00" "0,00" ...
##  $ vote_count  : num [1:11535] 0 6 2 0 0 0 0 0 0 ...
##  $ fight_id    : num [1:11535] 1 2 3 4 5 6 7 8 9 10 ...
##  - attr(*, "spec")=
```

```
## .. cols(
## ..   season = col_character(),
## ..   season_type = col_character(),
## ..   fight = col_character(),
## ..   player_1_name = col_character(),
## ..   player_2_name = col_character(),
## ..   player_1_team = col_character(),
## ..   player_2_team = col_character(),
## ..   date = col_character(),
## ..   period = col_character(),
## ..   gametime = col_time(format = ""),
## ..   winner = col_character(),
## ..   rating = col_character(),
## ..   vote_count = col_double(),
## ..   fight_id = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
summary(peleas)
```

```
##      season      season_type      fight      player_1_name
## Length:11535      Length:11535      Length:11535      Length:11535
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##      player_2_name      player_1_team      player_2_team      date
## Length:11535      Length:11535      Length:11535      Length:11535
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##      period      gametime      winner      rating
## Length:11535      Length:11535      Length:11535      Length:11535
## Class :character  Class1:hms      Class :character  Class :character
## Mode  :character  Class2:difftime  Mode  :character  Mode  :character
##                  Mode  :numeric
##
##
##      vote_count      fight_id
## Min.   : 0.0      Min.   : 1
## 1st Qu.: 5.0      1st Qu.: 2884
## Median : 37.0      Median : 5768
## Mean   : 53.5      Mean   : 5768
## 3rd Qu.: 76.0      3rd Qu.: 8652
## Max.   :672.0      Max.   :11535
```

```
sapply(peleas, function(x) length(unique(x)))
```

Cálculos previos

```
##      season  season_type      fight player_1_name player_2_name
##      24         1        10844         1553         1508
## player_1_team player_2_team      date      period      gametime
##      33         33        3383         4         1199
##      winner      rating  vote_count      fight_id
##      1375         750        367        11535
```

```
unique(peleas$season)
```

```
## [1] "1999-2000" "2000-2001" "2001-2002" "2002-2003" "2003-2004" "2005-2006"
## [7] "2006-2007" "2007-2008" "2008-2009" "2009-2010" "2010-2011" "2011-2012"
## [13] "2012-2013" "2013-2014" "2014-2015" "2015-2016" "2016-2017" "2017-2018"
## [19] "2018-2019" "2019-2020" "2020-2021" "2021-2022" "2022-2023" "2023-2024"
```

```
unique(peleas$season_type)
```

```
## [1] "reg"
```

```
colSums(is.na(peleas))
```

```
##      season  season_type      fight player_1_name player_2_name
##      0         0         0         0         0
## player_1_team player_2_team      date      period      gametime
##      0         0         0         0         0
##      winner      rating  vote_count      fight_id
##      0         0         0         0
```

```
table(peleas$season)
```

```
##
## 1999-2000 2000-2001 2001-2002 2002-2003 2003-2004 2005-2006 2006-2007 2007-2008
##      573      684      803      656      789      466      497      664
## 2008-2009 2009-2010 2010-2011 2011-2012 2012-2013 2013-2014 2014-2015 2015-2016
##      719      714      645      546      347      469      391      344
## 2016-2017 2017-2018 2018-2019 2019-2020 2020-2021 2021-2022 2022-2023 2023-2024
##      372      280      226      195      196      314      334      311
```

```
table(peleas$season_type)
```

```
##
## reg
## 11535
```

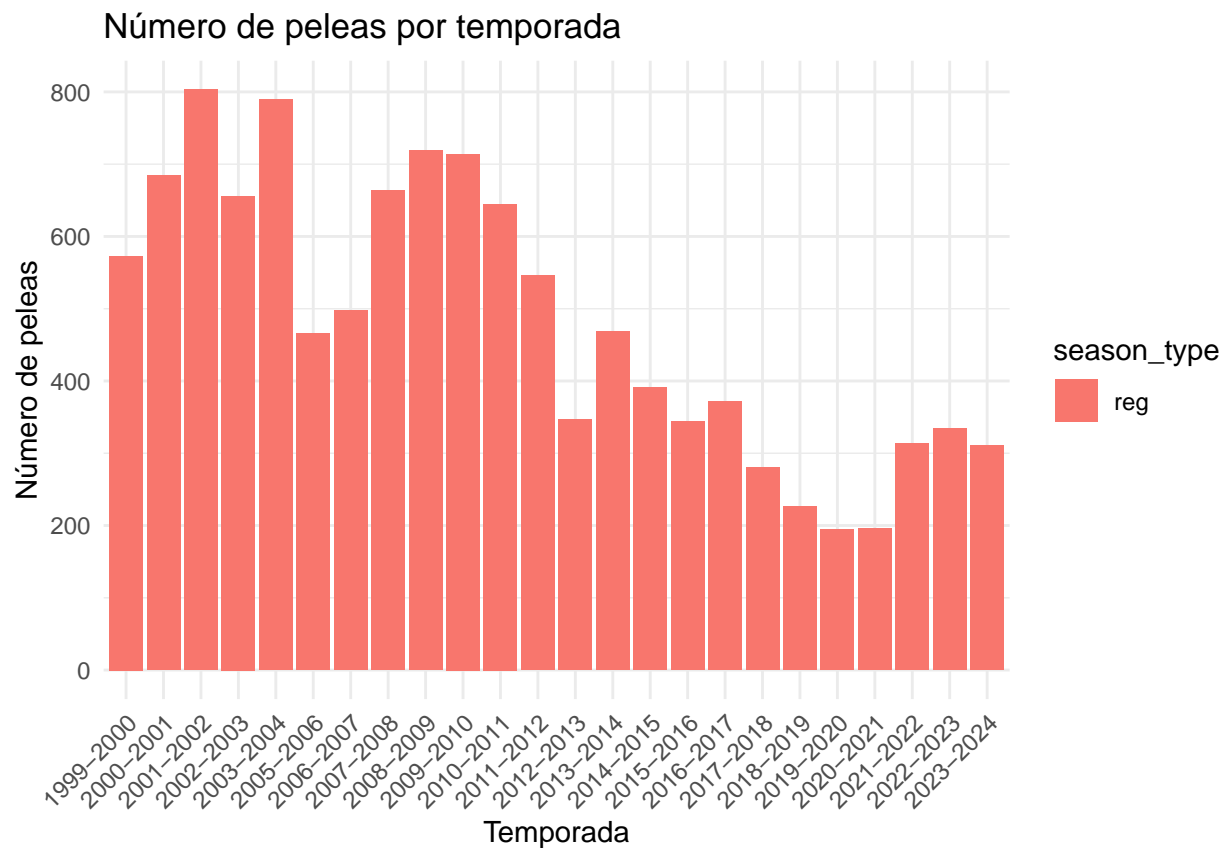
Análisis y visualizaciones

```
peleas_por_temporada <- peleas %>%
  group_by(season, season_type) %>%
  summarise(total_peleas = n())
```

'summarise()' has grouped output by 'season'. You can override using the
'.groups' argument.

```
view(peleas_por_temporada)
```

```
ggplot(peleas_por_temporada, aes(x = season, y = total_peleas, fill = season_type)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Número de peleas por temporada",
       x = "Temporada",
       y = "Número de peleas") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



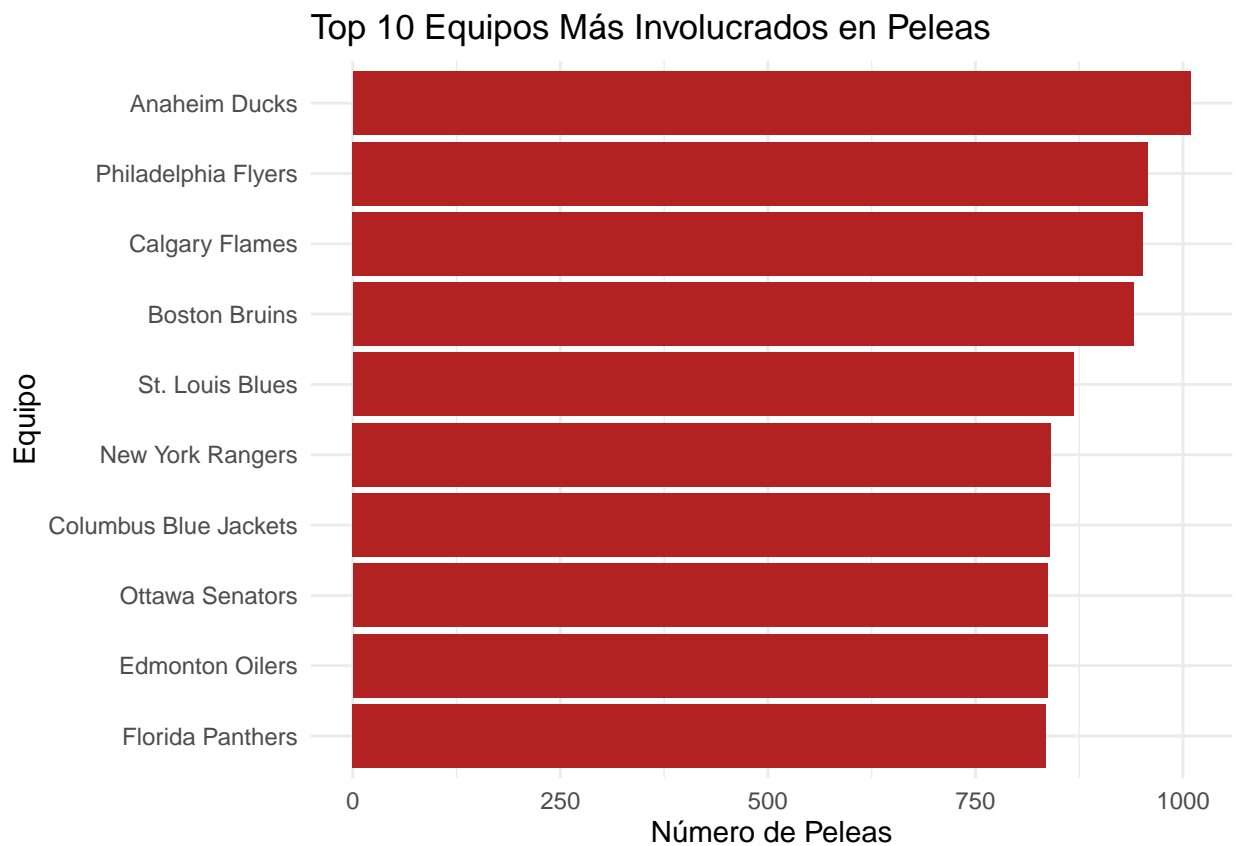
```
peleas_por_jugador <- peleas %>%
  gather(key = "jugador", value = "nombre_jugador", player_1_name, player_2_name) %>%
  group_by(nombre_jugador) %>%
  summarise(total_peleas = n()) %>%
  arrange(desc(total_peleas)) %>%
  head(10)
```

```

peleas_por_equipo <- peleas %>%
  gather(key = "equipo", value = "nombre_equipo", player_1_team, player_2_team) %>%
  group_by(nombre_equipo) %>%
  summarise(total_peleas = n()) %>%
  arrange(desc(total_peleas)) %>%
  head(10)

ggplot(peleas_por_equipo, aes(x = reorder(nombre_equipo, total_peleas), y = total_peleas)) +
  geom_bar(stat = "identity", fill = "firebrick") +
  coord_flip() +
  labs(title = "Top 10 Equipos Más Involucrados en Peleas",
       x = "Equipo",
       y = "Número de Peleas") +
  theme_minimal()

```

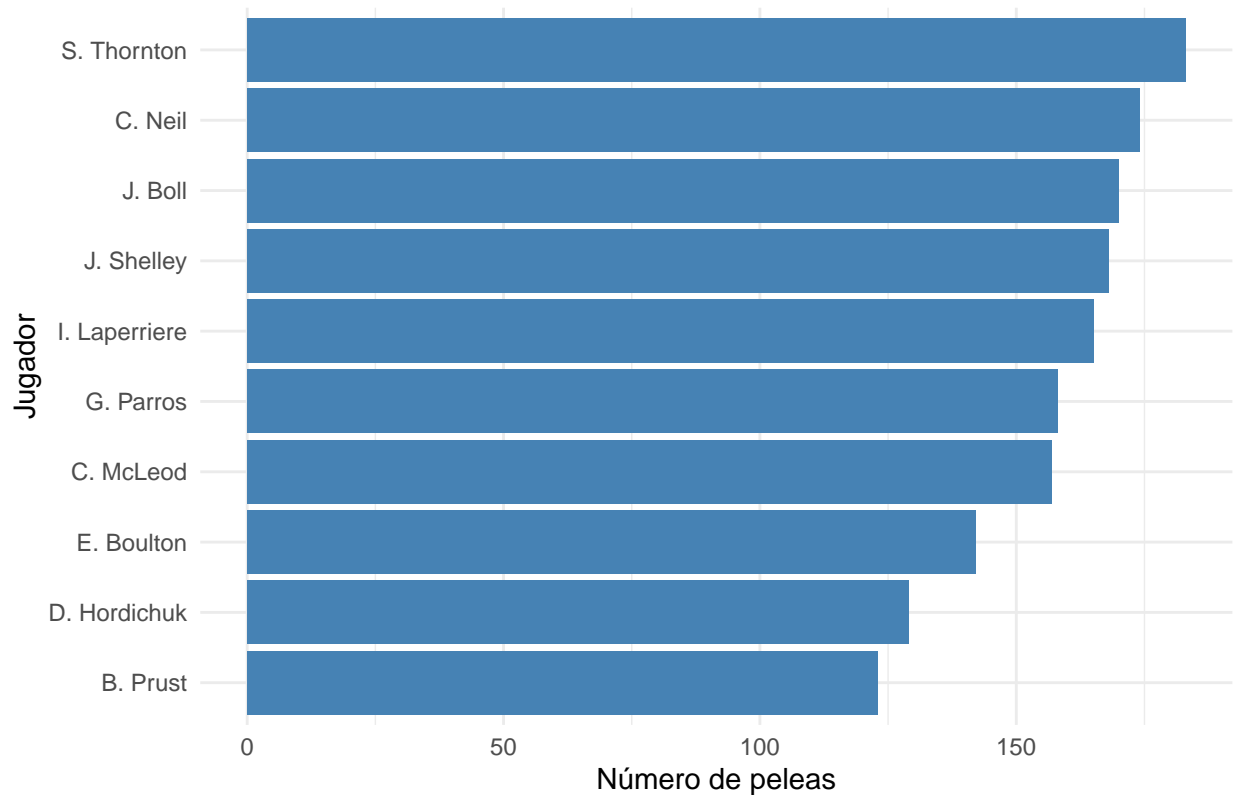


```

ggplot(peleas_por_jugador, aes(x= reorder(nombre_jugador, total_peleas), y = total_peleas)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() + #Invertir gráfico
  labs(title = "Top 10 jugadores más involucrados en peleas",
       x = "Jugador",
       y = "Número de peleas") +
  theme_minimal()

```

Top 10 jugadores más involucrados en peleas

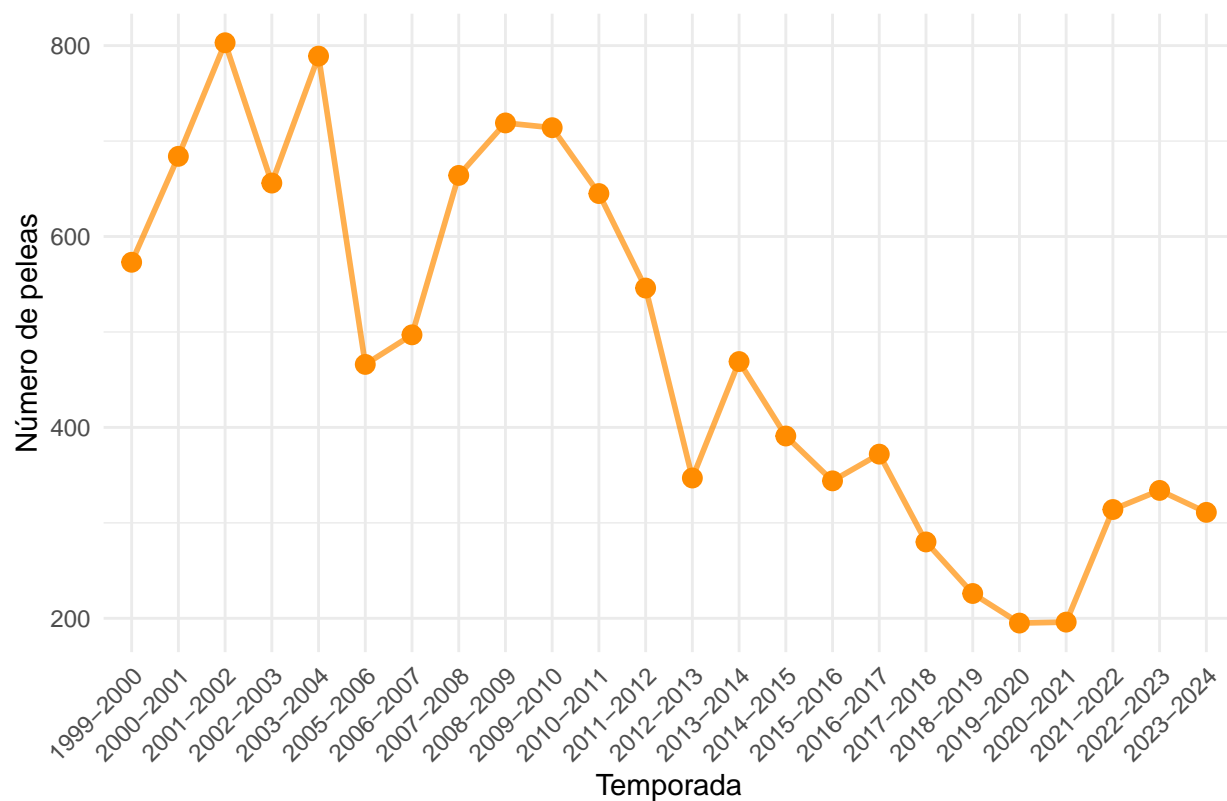


```
peleas_por_año <- peleas %>%
  group_by(season) %>%
  summarise(total_peleas = n())

ggplot(peleas_por_año, aes(x = season, y = total_peleas, group = 1)) +
  geom_point(color = "darkorange", size = 3) +
  geom_line(color = "darkorange", size = 1, alpha = 0.7) +
  labs(title = "Evolución de las peleas por temporada",
       x = "Temporada",
       y = "Número de peleas") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

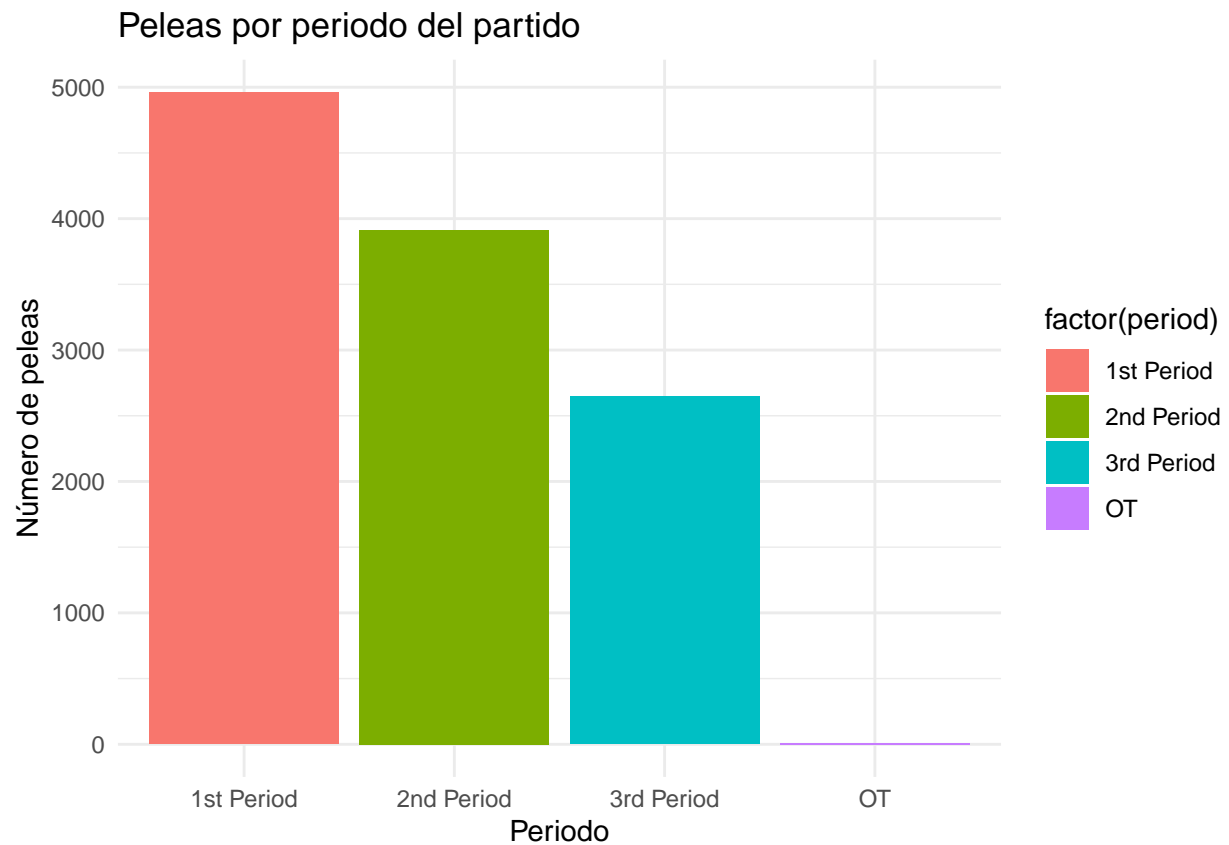
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```


Evolución de las peleas por temporada

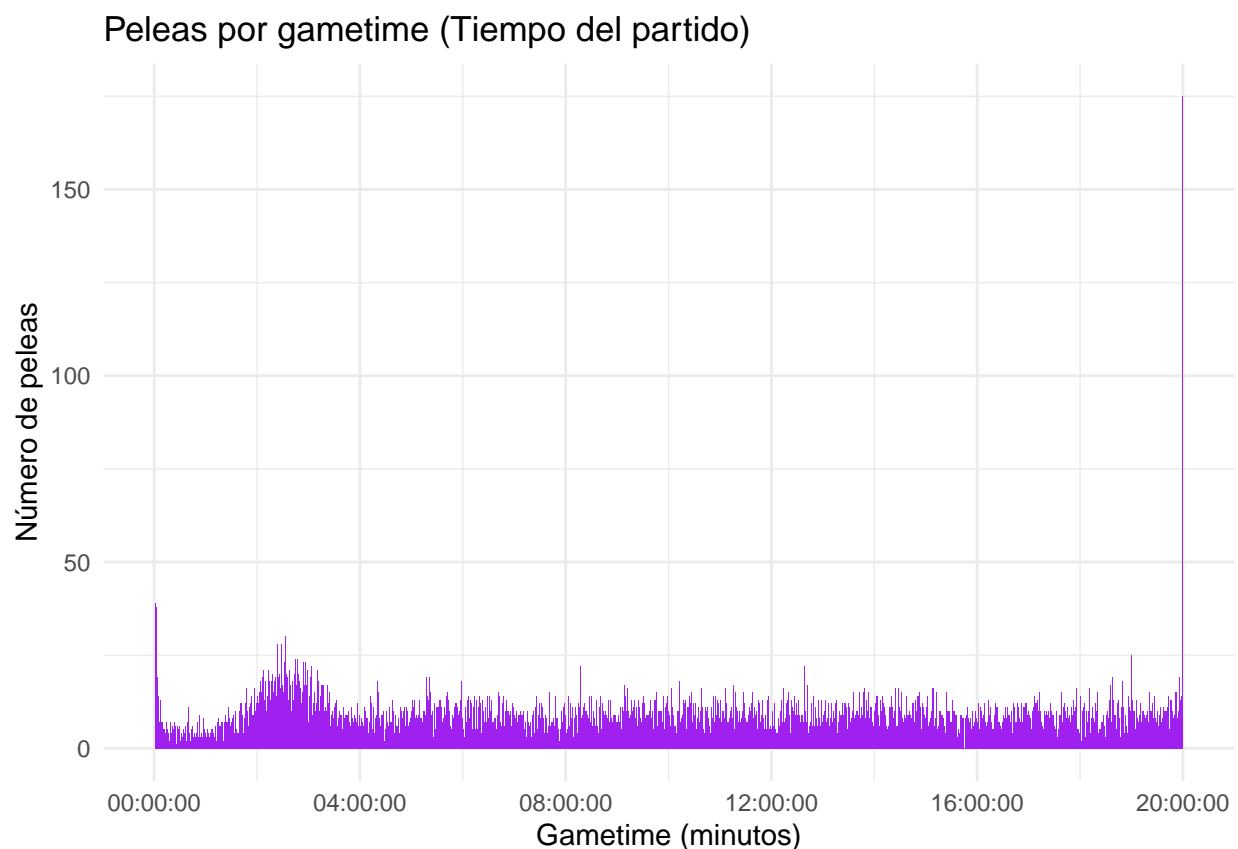


```
peleas$date <- as.Date(peleas$date, format="%d/%m/%Y")
peleas$period <- as.factor(peleas$period)

ggplot(peleas, aes(x = factor(period), fill = factor(period))) +
  geom_bar() +
  labs(title = "Peleas por periodo del partido",
       x = "Periodo",
       y = "Número de peleas") +
  theme_minimal()
```



```
ggplot(peleas, aes(x = gametime)) +  
  geom_histogram(binwidth = 5, fill = "purple") + # Aquí cambiamos el binwidth a 5  
  labs(title = "Peleas por gametime (Tiempo del partido)",  
        x = "Gametime (minutos)",  
        y = "Número de peleas") +  
  theme_minimal()
```



Resumen del análisis

Los hallazgos de este análisis revelan una disminución significativa en la cantidad de peleas en la NHL a lo largo de los años, impulsada por factores como cambios en la cultura de la liga y nuevas regulaciones. Sin embargo, la agresividad sigue siendo una característica importante en ciertos contextos del juego, especialmente al inicio de los partidos y en momentos críticos.

Hallazgos clave

Al analizar la evolución de las peleas a lo largo de las temporadas, se puede observar una clara tendencia a la baja en el número de peleas registradas. En las temporadas regulares, el pico máximo se alcanzó a principios de los años 2000, llegando a casi 800 peleas por temporada. Desde entonces, se ve una disminución constante, llegando a menos de 300 peleas por temporada en los últimos años.

Esta tendencia decreciente puede deberse a varios factores: - Un cambio en la cultura de la liga, con mayor énfasis en el juego técnico y habilidoso y menos tolerancia a los comportamientos agresivos. - La implementación de nuevas reglas y sanciones más estrictas para limitar las peleas y promover un juego más seguro. - Un mayor enfoque en la prevención de lesiones y la protección de los jugadores, lo que lleva a una menos aceptación de las peleas.

Es interesante notar que, en comparación con las temporadas regulares, los playoffs presentan un número de peleas mucho más bajo y estable a lo largo de los años. Esto nos sugiere que en momentos críticos de la temporada, los equipos y jugadores tienen a priorizar estrategias y habilidades por encima de la agresividad física.

Al analizar el top 10 de jugadores más involucrados en peleas, destacan ciertos patrones:

- Varios de los jugadores, como Colton Orr, George Parros y Jody Shelley tenían el papel de delanteros. Su estilo de juego agresivo y dispuesto a pelear les valió un lugar destacado en la lista.
- Por otro lado, la presencia de jugadores como Zdeno Chára y Shea Weber sugiere que la agresividad no se limita a los delanteros. Estos defensores de élite también han demostrado una disposición a defender a sus compañeros de equipo mediante peleas cuando es necesario.

La correlación vista entre el número de peleas y el éxito en el hielo es discutible. Algunos de estos jugadores han tenido carreras destacadas, pero otros han sido más conocido por su papel “luchador” que por su habilidad técnica. Por ejemplo, Colton Orr era más conocido por ser un “ejecutor”, dispuesto a quitarse los guantes, antes que su toque goleador.

En general, este Top 10 refleja la diversidad de perfiles que pueden verse involucrados en las peleas, desde los especialistas en enfrentamientos físicos hasta las estrellas de la NHL que de forma ocasional tienen a defender a sus compañeros.

En cuánto a los equipos, al analizar el Top 10 de los más involucrados en peleas observamos algunos puntos interesantes:

- Los equipos líderes, como los Anaheim Ducks, Philadelphia Flyers y Calgary Flames, han sido tradicionalmente reconocidos por su juego físico y agresivo. Esta agresividad parece una parte integral del estilo de juego e identidad.
- El éxito de estos equipos en el rink de hockey es variado. Algunos han logrado importantes logros, mientras que otros han tenido temporadas más irregulares.

Esto nos sugiere que la agresividad, tal y cómo hemos visto con los jugadores, aunque puede que sea una herramienta útil en ciertos momentos, no es en sí misma una garantía de éxito. Para que un equipo destaque debe de haber un equilibrio entre la intensidad física y la habilidad técnica. Equipos como los Rangers de Nueva York han demostrado que se puede tener un juego agresivo sin necesariamente ser uno de los equipos más peleones de la liga.

Los dos últimos gráficos muestran la distribución de las peleas a lo largo de los diferentes periodos de juego, en los que vemos:

- El primer periodo es el que presenta un mayor número de peleas, con un pico de alrededor 4.500 peleas.
- El segundo periodo, si bien tiene un número elevado, es menor que el primero.
- En el tercer periodo las peleas disminuyen de forma considerable, hasta llegar a los 2.000. Más de la mitad.
- En la categoría de Tiempo Extra (OT) es cuando muchas menos peleas suceden, apenas llegando a los 500.

Estos datos nos sugieren que la mayor parte de la actividad física y agresiva se concentra al inicio del partido, probablemente debido a la alta tensión y a la rivalidad entre los equipos. A medida que el juego avanza parece haber una disminución en la propensión a las peleas, posiblemente porque los equipos se enfocan más en la estrategia y ejecución técnica para lograr la victoria.

El segundo gráfico muestra las peleas por minuto de juego. Aquí observamos un patrón similar. Hay picos de actividad en los primeros minutos del partido, con algunas fluctuaciones a lo largo del mismo, pero en general una tendencia a la baja conforme avanza el tiempo de juego.

Esto nos sugiere que los momentos más álgidos y propensos a las peleas se dan al inicio de los partidos, cuando los equipos tienen más energía y están ávidos por imponer su dominio físico. Conforme el juego progresa, los jugadores se concentran más en la estrategia y la disciplina para evitar sanciones que puedan comprometer las posibilidades de ganar.

Así pues, el análisis de la evolución de las peleas por periodo y tiempo de juego revela patrones interesantes relacionados con la dinámica y la tensión inherente a los partidos de hockey.

Conclusión

En resumen, la agresividad en la NHL ha disminuido con el tiempo, pero sigue siendo un elemento importante en ciertos contextos y momentos del juego. Los datos reflejan una liga en evolución, que se mueve hacia un juego más técnico y táctico, aunque la historia de las peleas sigue siendo una parte esencial del carácter de la NHL.