

# Auckland Rugby Project - Trend Analysis of GPS Data

Ian Chen

The project will be available on GitHub at <https://github.com/ichen98/2021-UoA-DATASCI-792-Project>. If there are any questions about this analysis, please contact me via email at [ian.chen1201@gmail.com](mailto:ian.chen1201@gmail.com).

```
library(tidyverse)
```

First, I load the datasets in. The 2018 data notably has many more variables than the other years' data. These additional variables may be of use for the analysis, but because there is less data on these variables, they may necessitate a separate model that trains on only the 2018 data.

```
# Loading the 2018 .csv files in
master2018 <-
  list.files(path = "./2018_csvs/", pattern = "*.CSV", full.names = T) %>%
  map_df(~read.csv(., skip = 4, header = TRUE))

master2019 <-
  list.files(path = "./2019_csvs/", pattern = "*.CSV", full.names = T) %>%
  map_df(~read.csv(., skip = 4, header = TRUE, colClasses = rep("character", 17)))

master2020 <-
  list.files(path = "./2020_csvs/", pattern = "*.CSV", full.names = T) %>%
  map_df(~read.csv(., skip = 4, header = TRUE))
```

I begin with cleaning the 2018 .csv files.

The column names have varying degrees of spacing before and after words. They are cleaned up to have consistent names in proper English.

```
# Cleaning up column names
correctColumnNames <- c("Athlete",
  "Team",
  "Date",
  "Start Time",
  "Duration Total (s)",
  "Duration Speed Hi-Inten (s)",
  "Duration HR Hi-Inten (s)",
  "Distance Total (m)",
  "Distance Rate (m/min)",
  "Distance Speed Hi-Inten (m)",
  "Distance HR Hi-Inten (m)",
  "Speed Max (km/h)",
  "Sprints Total (num)",
  "Sprints Hi-Inten (num)",
  "Sprints HR Hi-Inten (num)",
  "HR Max Total (bpm)",
  "% Max HR",
  "Work Recovery Ratio",
```

```

"Speed Duration Total (s)",
"HR Duration Total (s)",
"Athlete Load",
"Metabolic PowerPeak",
"Hi Int Acceleration (num)",
"Hi Int Deceleration (num)",
"Impact Rate (imp/min)",
"Body Impacts (num)",
"Hi Intensity Effort (num)",
"HIE Rate",
"Distance Speed Zone 1 (m)",
"Distance Speed Zone 2 (m)",
"Distance Speed Zone 3 (m)",
"Distance Speed Zone 4 (m)",
"Distance Speed Zone 5 (m)",
"Sprints Speed Zone 3 (num)",
"Sprints Speed Zone 4 (num)",
"Sprints Speed Zone 5 (num)",
"Duration HR Zone 4 (s)",
"Duration HR Zone 5 (s)",
"Accelerations Zone 3 (num)",
"Accelerations Zone 4 (num)",
"Accelerations Zone 5 (num)",
"Decelerations Zone 3 (num)",
"Decelerations Zone 4 (num)",
"Decelerations Zone 5 (num)",
"Body Impacts in Body Impacts Zone Total (num)",
"Body Impacts Grade 1 (num)",
"Body Impacts Grade 2 (num)",
"Body Impacts Grade 3 (num)",
"Body Impacts Grade 4 (num)",
"Body Impacts Grade 5 (num)"
colnames(master2018) <- correctColumnNames

```

Each individual .csv includes four opening rows that do not provide any meaningful information (which are skipped when the .csv is read into R), and three rows at the end that provide details about the average (mean), maximum and minimum values for each column. These are not useful for this analysis, so they should be removed here.

Furthermore, there are many data points that have missing data (represented by two asterisks - "\*\*"). These need to be converted into NA values, which are easier to work around than a string of two asterisks forcing numeric columns into character columns.

```

# Removing excess rows
master2018 <- subset(master2018, Athlete != "Avg" & Athlete != "Highest" & Athlete != "Lowest")

# Replacing all missing values with NA
master2018 <- na_if(master2018, "**")

```

The cells that were initially occupied by "\*\*" strings forcibly converted their respective columns into character columns during the dataset import. These columns need to be converted into their proper class such that they can be useful for modeling.

First, the dates are imported into R as characters. For ease of reading, the data frame is sorted by date, from earliest to latest. This involves the conversion of the Date column into Date class objects, which requires all values in the Date column to be of a certain format.

The last column appears to be an error, not existing in the actual .csv files, so it is additionally dropped.

```
# Converting dates into something usable
library(lubridate)

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##      date, intersect, setdiff, union

master2018$Date[21:38] <- "27/10/2018"
master2018$Date <- parse_date_time(master2018$Date, c("%d/%m/%Y"))
# Sorting by date, dropping redundant column `X`
master2018 <- master2018[order(as.Date(master2018$Date)), -51]
```

Columns 11 and 15 (Distance HR Hi-Inten (m) and Sprints HR Hi-Inten (num) respectively) are numeric values that were also imported into R as characters. These are transformed back into numeric variables. This is necessary for the proportional standardisation that is applied later.

```
# Converting columns 11 and 15 back into numeric vectors
for (i in c(11, 15)) {
  master2018[, i] <- as.numeric(master2018[, i])
}
```

All durations are imported into R as character strings, as R can't parse the "MM:SS" format. Some preprocessing will need to be done with the times in the dataset, and by converting them into numeric values, manipulation of them will become a lot simpler. Therefore, all times in the data are converted to numeric values.

In this case, converting them to seconds is an easy way of standardising all of the times, making them integers. Integers make things easy to calculate without having to deal with fractions of a minute (which are in base 60).

```
minsec_to_sec <- function(strvec) {
  # All durations are in "MM:SS" format; durations > 1 hr simply have MM > 59
  prelength <- ifelse(nchar(strvec) == 6, 3, ifelse(nchar(strvec) == 5, 2, 1))
  pre <- as.numeric(substr(strvec, 1, prelength))
  suf <- as.numeric(substr(strvec, nchar(strvec) - 1, nchar(strvec)))
  strvec <- pre * 60 + suf
  return(strvec)
}
master2018[, c(5:7, 19:20, 37:38)] <- lapply(master2018[, c(5:7, 19:20, 37:38)], minsec_to_sec)
```

A rugby union match goes for two 40-minute halves, with a halftime of a maximum length of 15 minutes. This sets a match at roughly a maximum of 95 minutes long. In the Mitre 10 Cup, should a semi-final or final match be tied at the end of regulation time, two 10-minute halves of extra time are played. This is the longest extension a Mitre 10 Cup game can have. Because much of the data's time values are abnormally high, a hard limit is set at 95 minutes (roughly the length of a regular match, including halftime), with the exception of the 2018 final, which went to extra time (resulting in a total of 120 minutes being played, so a hard limit of 120 minutes will be applied exclusively for that match).

95 minutes is equal to  $95 \times 60 = 5700$  seconds, while 120 minutes is equal to  $120 \times 60 = 7200$  seconds, so 5700 and 7200 will be the hard limits imposed on the minutes played.

Other duration variables may also have abnormally high values, so they will need to be adjusted too. These anomalous values are likely due to errors with the time tracking device, as it appears that many of the duration values are problematic.

If a player's total minutes played is cut down to the set ceiling, then the other duration variables are adjusted by calculating a proportion of the original minutes played, and using this proportion as a multiplier for the other duration variables. For instance, if a player has 100 minutes (6000 seconds) played in a non-2018-final match, that player's corresponding proportion is  $5700/6000 = 0.95$ , which then multiplies by the player's other duration values to give their adjusted values.

```
# Calculating proportion by the above method
master2018$Proportion <- ifelse(
  as.character(master2018[, 3]) == "2018-10-27",
  7200 / master2018$`Duration Total (s)`,
  5700 / master2018$`Duration Total (s)`)
# Only interested in adjusting values that have a `Proportion` value < 1
master2018$Proportion[which(master2018$Proportion > 1)] <- 1
for (j in c(5, 7:8, 10:11, 13:15, 19:24, 26:27, 29:50)) {
  master2018[, j] <- master2018[, j] * master2018$Proportion
}
```

Column 17, % Max HR, contains a percentage symbol in each of the values. Because all of these values should be numeric, the percentage symbol is removed and % Max HR is converted to numeric.

```
# Removing percentage symbols
master2018[, 17] <- as.numeric(substr(master2018[, 17], 1, nchar(master2018[, 17]) - 1))
```

Column 18, Work Recovery Ratio, contains a small set of unique values. This can be recoded into a factor.

```
# Recoding Work Recovery Ratio into a factor
master2018[, 18] <- as.factor(master2018[, 18])
```

Player names are misspelled in different ways across each dataset. These must be standardised to allow for simpler merging of additional information.

```
# Every name from every dataset combined
currentNames <- sort(unique(c(master2018$Athlete, master2019$Athlete, master2020$Athlete)))

# The incorrectly-recorded names
problematicNames <- c("Able, Rob",
  "Hallem Ewes, Liam",
  "Hodgmen, Alex",
  "Lemalu, Faatungu",
  "Liaana, Desma",
  "Liana, Desma",
  "Lundenmuth, Ezeikeil",
  "Reidler Kapa, Waimana",
  "Ruru, Jonathon",
  "Schwenke, Lief",
  "Scraffton, Scott",
  "Sosene, Mike",
  "Sotutu, Hoksins")

# The corrections to the above names
correctedNames <- c("Abel, Robbie",
  "Hallam-Eames, Liam",
  "Hodgman, Alex",
  "Lemalu, Fa'atiga",
  "Liaina, Desma",
  "Liaina, Desma",
  "Lindenmuth, Ezi",
  "Riedlinger-Kapa, Waimana",
```

```

        "Ruru, Jonathan",
        "Schwenke, Leif",
        "Scrafton, Scott",
        "Sosene-Feagai, Mike",
        "Sotutu, Hoskins")

# A function for name correction
nameCorrection <- function(data) {
  for (k in 1:length(problematicNames)) {
    data[which(data[, 1] == problematicNames[k]), 1] <- correctedNames[k]
  }
  return(data)
}

# Applying the function
master2018 <- nameCorrection(master2018)

```

Win margins will be used as a one-size-fits-all metric for measuring how good a player's performance in a match is i.e. the response variable for any fitted model. This is added to the main dataset.

```

# Dates of matches
matchDates <- as.Date(c("2018-08-18",
                        "2018-08-26",
                        "2018-08-30",
                        "2018-09-07",
                        "2018-09-16",
                        "2018-09-22",
                        "2018-09-28",
                        "2018-10-04",
                        "2018-10-10",
                        "2018-10-14",
                        "2018-10-20",
                        "2018-10-27",
                        "2019-08-09",
                        "2019-08-15",
                        "2019-08-24",
                        "2019-08-31",
                        "2019-09-08",
                        "2019-09-14",
                        "2019-09-22",
                        "2019-09-27",
                        "2019-10-05",
                        "2019-10-11",
                        "2019-10-19",
                        "2020-09-12",
                        "2020-09-20",
                        "2020-09-27",
                        "2020-10-02",
                        "2020-10-10",
                        "2020-10-17",
                        "2020-10-24",
                        "2020-10-31",
                        "2020-11-07",
                        "2020-11-15",

```

```

        "2020-11-21",
        "2020-11-28"))
# Match win margins by date
margins <- c(4, 16, 18, 26, 5, 1, -5, 5, 48, 16, 21, 7,
            0, 33, 6, 0, -10, 15, -19, -40, 57, 24, -9,
            32, -18, 38, 4, 1, 21, -1, 21, 4, -1, 5, -1)
# Combining date and win margins into one dataframe
winMargins <- data.frame(Date = matchDates, margins)
# Combining win margins into the main dataframe, merging by Date
master2018 <- left_join(master2018, winMargins)

```

```
## Joining, by = "Date"
```

I created two supplementary files to provide additional necessary variables. The first is `positional_data_by_match.csv`, which contains each match's game day squad. This provides the position that each player named in the squad for that matchup played at. The replacements (wearing jerseys 16-23) were labelled as 16, as the replacement jersey number does not provide exact positional information.

The second supplementary file is `positional_data.csv`, which contains the preferred position for each player. This was determined by selecting the position in the starting XV that they appeared in the most over the matches represented in the dataset. For those that did not make any appearances in the starting XV, some Googling and some clarification with Paul Downes, my Auckland Rugby representative filled in their preferred position.

The positional data in `positional_data_by_match.csv` is added to the master dataset for the players that were named in the starting XV for each of the matches played in 2018. The preferred positions in `positional_data.csv` is added to the master dataset for any players that were named as replacements, to show what position they would typically fill in if they had started the match with the starting XV.

```

# Positional data by match
matchPos <- read.csv("positional_data_by_match.csv", skip = 4)
# Rename columns to be consistent with the data
colnames(matchPos) <- c("Athlete", as.character(matchDates))
matchPos <- nameCorrection(matchPos)
# Initialise the position column
master2018$Position <- 0
# Go through each match day
for (l in 2:36) {
  currentDate <- colnames(matchPos)[1]
  # Get the squad that played on/was named for this day
  activeSquad <- matchPos[which(matchPos[, 1] != 0), c(1, 1)]
  playersOnThisDate <- which(master2018$Date == as.Date(currentDate) & master2018$Athlete %in% activeSq
  # Add the position for each player
  for (m in playersOnThisDate) {
    player <- which(activeSquad[, 1] == master2018[m, 1])
    master2018[m, 53] <- activeSquad[player, 2]
  }
}

# Now to deal with the replacements, which are all labelled 16
# The preferred positions are in "positional_data.csv"
preferredPos <- read.csv("positional_data.csv")[1:65, ]
colnames(preferredPos) <- c("Name",
                           "1 - Loosehead prop",
                           "2 - Hooker",
                           "3 - Tighthead prop",

```

```

        "4 - Left lock",
        "5 - Right lock",
        "6 - Blindside flanker",
        "7 - Openside flanker",
        "8 - Number 8",
        "9 - Scrum-half",
        "10 - Fly-half",
        "11 - Left wing",
        "12 - Inside centre",
        "13 - Outside centre",
        "14 - Right wing",
        "15 - Fullback",
        "16-23 - Replacement",
        "Pref. pos. (number)",
        "Pref. pos. (text)",
        "Pref. group")
preferredPos <- nameCorrection(preferredPos)
# Coercing the columns of the preferred positional data into ideal classes
for (n in 2:18) {
  preferredPos[, n] <- as.numeric(preferredPos[, n])
}
# Getting the rows that correspond to replacements
replacements <- which(master2018$Position == 16)
# Giving the replacements their preferred position
for (o in replacements) {
  replacementName <- master2018[o, 1]
  master2018[o, 53] <- as.numeric(preferredPos[which(preferredPos[, 1] == replacementName), 18])
}
# Converting the positional data into a factor
master2018[, 53] <- as.factor(master2018[, 53])

```

Finally, excess rows are removed from the master dataset. These include rows where, for a given match, a player is listed multiple times, as well as players who are in the data but were not named to the 23-man match-day squad for that match. For the latter, they did not contribute to the win margin that corresponds to that match, so their data is not useful for prediction.

```

# Removing duplicate rows
for (p in unique(master2018$Date)) {
  matchPlayers <- master2018[which(master2018$Date == p), 1]
  for (q in matchPlayers[duplicated(matchPlayers)]) {
    dupes <- master2018[which(master2018$Date == p & master2018$Athlete == q),]
    notHighestMinutes <- as.numeric(rownames(dupes[which(dupes$`Duration Total (s)` != max(dupes$`Durat
    master2018 <- master2018[-notHighestMinutes,]
  }
}
# Deleting players that didn't play in the game, since there would be no win margin associated with the
master2018 <- master2018[-which(master2018$Position == 0),]

```

That should be all the preliminary cleaning and preprocessing that needs to be done. The same methods are applied to the 2019 data, although in a modified manner. The 2019 data has a far smaller subset of the variables in the 2018 data, so the column names here are different than the 2018 columns. Otherwise, the same cleaning is applied to the 2019 data, to keep the data consistent across each year.

```

# 2019 and 2020 datasets have fewer variables
colnames2019_20 <- c("Athlete",

```

```

      "Team",
      "Date",
      "Start Time",
      "Duration Total (s)",
      "Distance Total (m)",
      "Speed Max (km/h)",
      "Hi Int Acceleration (num)",
      "Distance Speed Zone 1 (m)",
      "Distance Speed Zone 2 (m)",
      "Distance Speed Zone 3 (m)",
      "Distance Speed Zone 4 (m)",
      "Distance Speed Zone 5 (m)",
      "Body Impacts in Body Impacts Zone Total (num)",
      "Sprints Speed Zone 3 (num)",
      "Sprints Speed Zone 4 (num)",
      "Sprints Speed Zone 5 (num)"
    )
colnames(master2019) <- colnames2019_20

master2019 <- subset(master2019, Athlete != "Avg" & Athlete != "Highest" & Athlete != "Lowest")
master2019 <- na_if(master2019, "**")

master2019$Date <- parse_date_time(master2019$Date, c("%d/%m/%Y"))
master2019 <- master2019[order(as.Date(master2019$Date)), -18]

for (i in 6:17) {
  master2019[, i] <- as.numeric(master2019[, i])
}

master2019[, 5] <- minsec_to_sec(master2019[, 5])

master2019$Proportion <- 5700 / master2019$`Duration Total (s)`
master2019$Proportion[which(master2019$Proportion > 1)] <- 1
for (j in c(5:6, 8:17)) {
  master2019[, j] <- master2019[, j] * master2019$Proportion
}

master2019 <- nameCorrection(master2019) %>%
  left_join(winMargins)

## Joining, by = "Date"
master2019$Position <- 0
for (l in 2:36) {
  currentDate <- colnames(matchPos)[l]
  activeSquad <- matchPos[which(matchPos[, 1] != 0), c(1, l)]
  playersOnThisDate <- which(master2019$Date == as.Date(currentDate) & master2019$Athlete %in% activeSq
  for (m in playersOnThisDate) {
    player <- which(activeSquad[, 1] == master2019[m, 1])
    master2019[m, 20] <- activeSquad[player, 2]
  }
}
replacements <- which(master2019$Position == 16)
for (o in replacements) {
  replacementName <- master2019[o, 1]

```



```

    master2019[o, 20] <- as.numeric(preferredPos[which(preferredPos[, 1] == replacementName), 18])
  }
master2019[, 20] <- as.factor(master2019[, 20])

for (p in unique(master2019$Date)) {
  matchPlayers <- master2019[which(master2019$Date == p), 1]
  for (q in matchPlayers[duplicated(matchPlayers)]) {
    dupes <- master2019[which(master2019$Date == p & master2019$Athlete == q),]
    notHighestMinutes <- as.numeric(rownames(dupes[which(dupes$`Duration Total (s)` != max(dupes$`Durat
    master2019 <- master2019[-notHighestMinutes,]
  }
}
master2019 <- master2019[-which(master2019$Position == 0),]

```

The same is done with the 2020 dataset, which closely resembles the 2019 dataset in structure and format, including the number and name of columns.

```

colnames(master2020) <- colnames2019_20

master2020 <- subset(master2020, Athlete != "Avg" & Athlete != "Highest" & Athlete != "Lowest")
master2020 <- na_if(master2020, "**")

master2020$Date <- parse_date_time(master2020$Date, c("%d/%m/%Y"))
master2020 <- master2020[order(as.Date(master2020$Date)), -18]

for (i in 6:17) {
  master2020[, i] <- as.numeric(master2020[, i])
}

master2020[, 5] <- minsec_to_sec(master2020[, 5])

master2020$Proportion <- 5700 / master2020$`Duration Total (s)`
master2020$Proportion[which(master2020$Proportion > 1)] <- 1
for (j in c(5:6, 8:17)) {
  master2020[, j] <- master2020[, j] * master2020$Proportion
}

master2020 <- nameCorrection(master2020) %>%
  left_join(winMargins)

## Joining, by = "Date"
master2020$Position <- 0
for (l in 2:36) {
  currentDate <- colnames(matchPos)[l]
  activeSquad <- matchPos[which(matchPos[, 1] != 0), c(1, 1)]
  playersOnThisDate <- which(master2020$Date == as.Date(currentDate) & master2020$Athlete %in% activeSq
  for (m in playersOnThisDate) {
    player <- which(activeSquad[, 1] == master2020[m, 1])
    master2020[m, 20] <- activeSquad[player, 2]
  }
}
replacements <- which(master2020$Position == 16)
for (o in replacements) {
  replacementName <- master2020[o, 1]

```

```

  master2020[o, 20] <- as.numeric(preferredPos[which(preferredPos[, 1] == replacementName), 18])
}
master2020[, 20] <- as.factor(master2020[, 20])

for (p in unique(master2020$Date)) {
  matchPlayers <- master2020[which(master2020$Date == p), 1]
  for (q in matchPlayers[duplicated(matchPlayers)]) {
    dupes <- master2020[which(master2020$Date == p & master2020$Athlete == q),]
    notHighestMinutes <- as.numeric(rownames(dupes[which(dupes$`Duration Total (s)` != max(dupes$`Duration Total (s)`), ]))
    master2020 <- master2020[-notHighestMinutes,]
  }
}
master2020 <- master2020[-which(master2020$Position == 0),]

```

The 2019 and 2020 datasets are combined, since they share the same columns. The 2018 dataset is joined with them also, but only at the columns that are shared with the 2019 and 2020 datasets.

```
combinedData <- rbind(master2018[, c(1:5, 8, 12, 23, 29:36, 45, 51:53)], master2019, master2020)
```

Finally, it appears that **Speed Duration Total (s)** and **HR Duration Total (s)** are not needed, since they measure the total duration of data collected beginning when the GPS unit locks (and when heart rate is detected for **HR Duration Total (s)**.) These are extremely correlated with **Duration Total (s)**, so these can be safely removed.

Additionally, **Body Impacts in Body Impact Zones Total (num)** is equal to the **Body Impacts (num)** measure in the 2018 dataset, appearing to capture the same information. Because the former is in all three datasets, the latter is removed, and the former is renamed to the simpler **Body Impacts (num)**.

```

master2018 <- master2018[, -c(19, 20, 26)]
colnames(master2018)[42] <- "Body Impacts (num)"
colnames(combinedData)[17] <- "Body Impacts (num)"

```

## Exploring the data

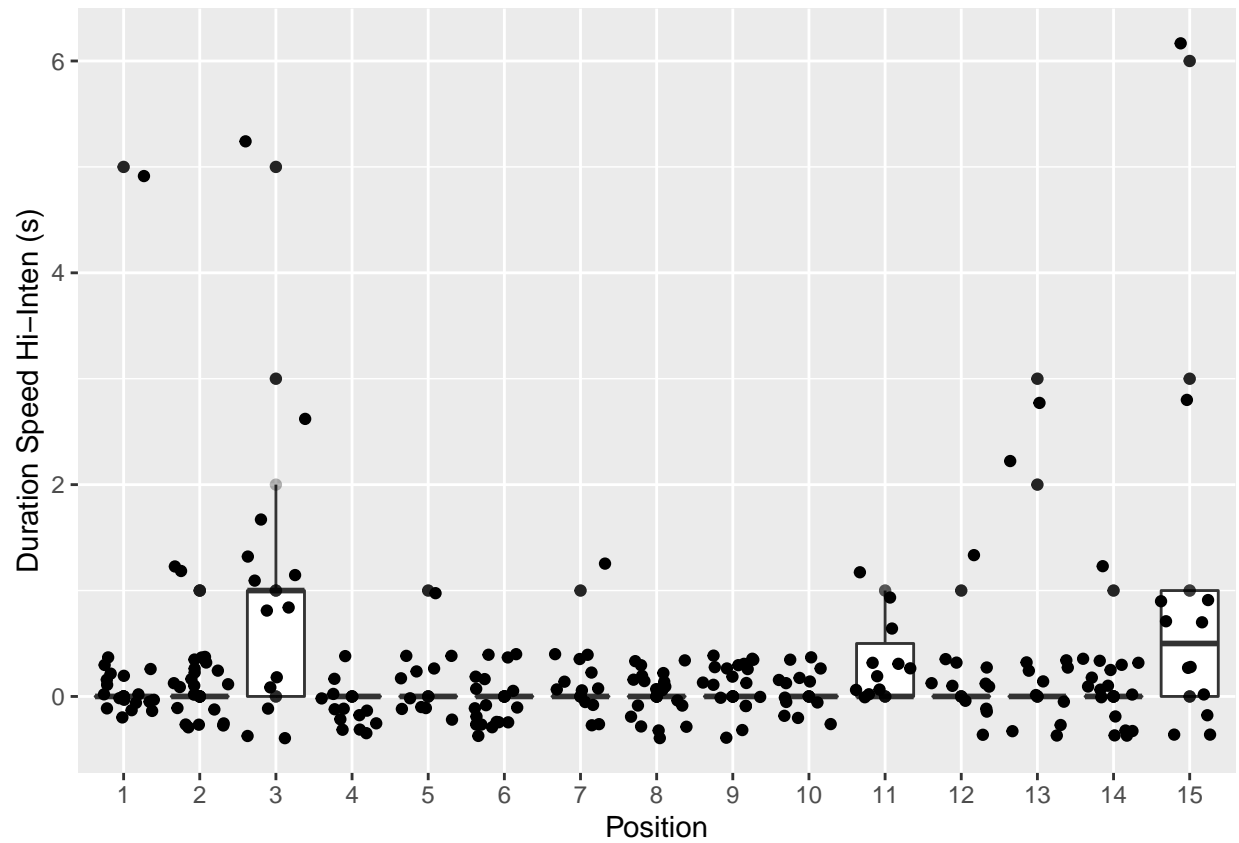
I want to explore the data to see if there are any interesting relationships between positional groups for each of the variables present in the datasets.

First, the variables unique to the 2018 dataset are plotted.

```

# 2018 dataset-unique variable visualisation
for (u in c(6:7, 9:11, 13:17, 19:20, 22:25, 34:41, 43:47)) {
  print(ggplot(master2018, aes(Position, master2018[, u])) +
    geom_boxplot() +
    geom_point(alpha = 0.3) +
    ylab(colnames(master2018)[u]) +
    geom_jitter())
}

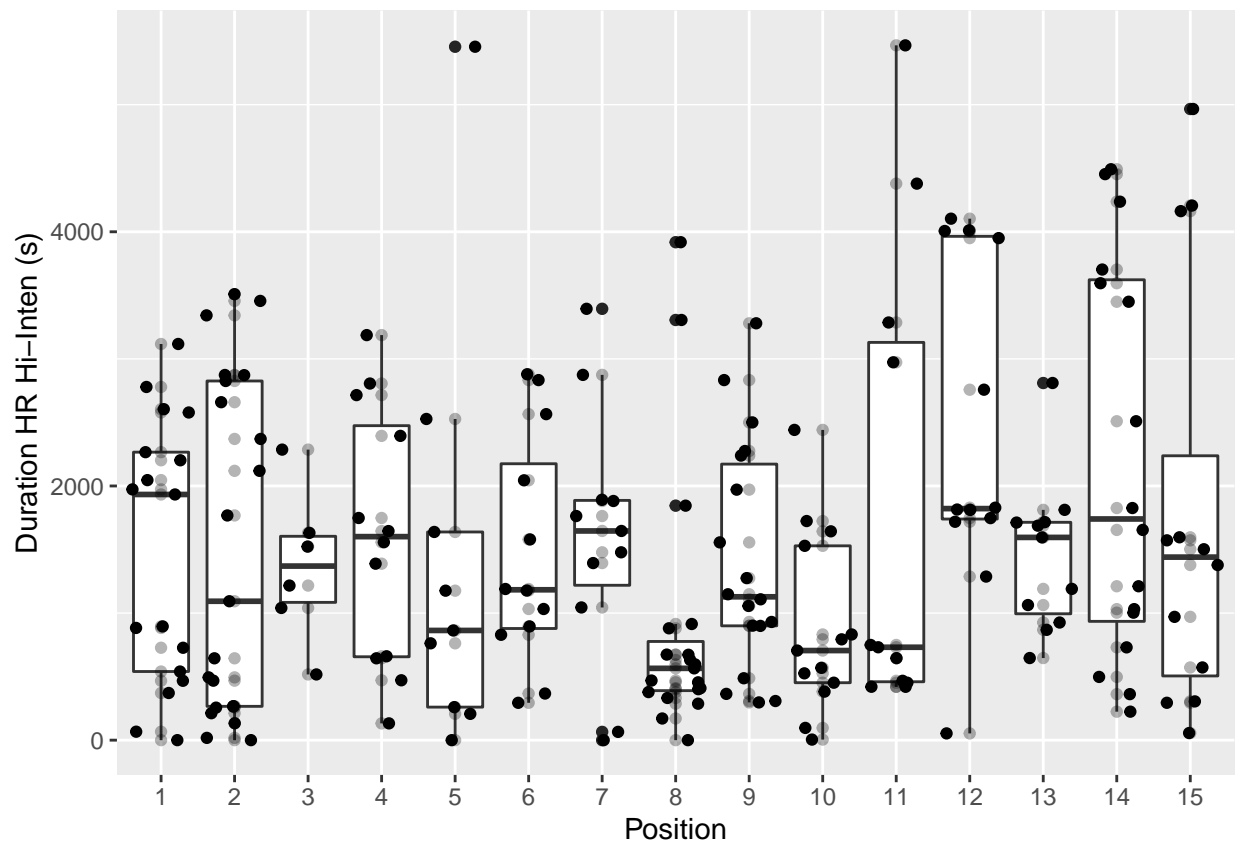
```

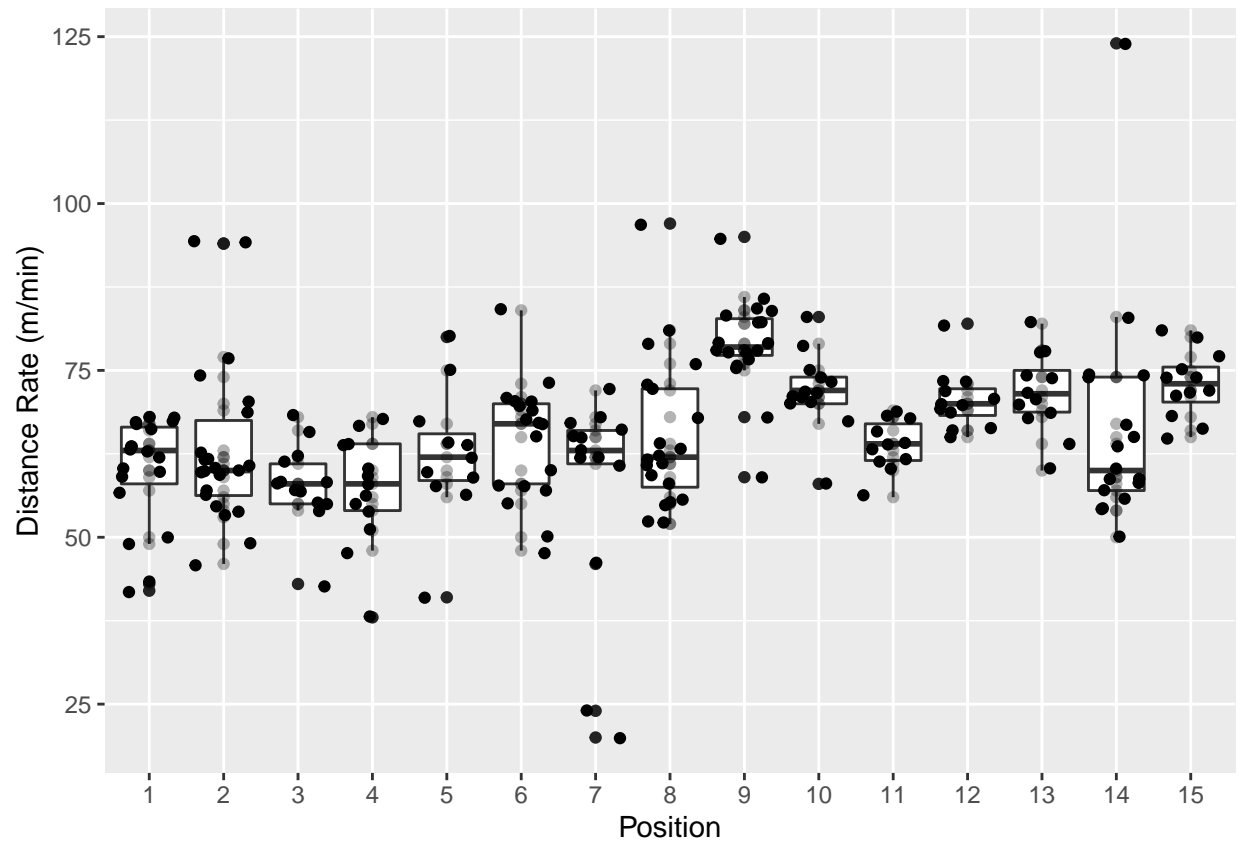


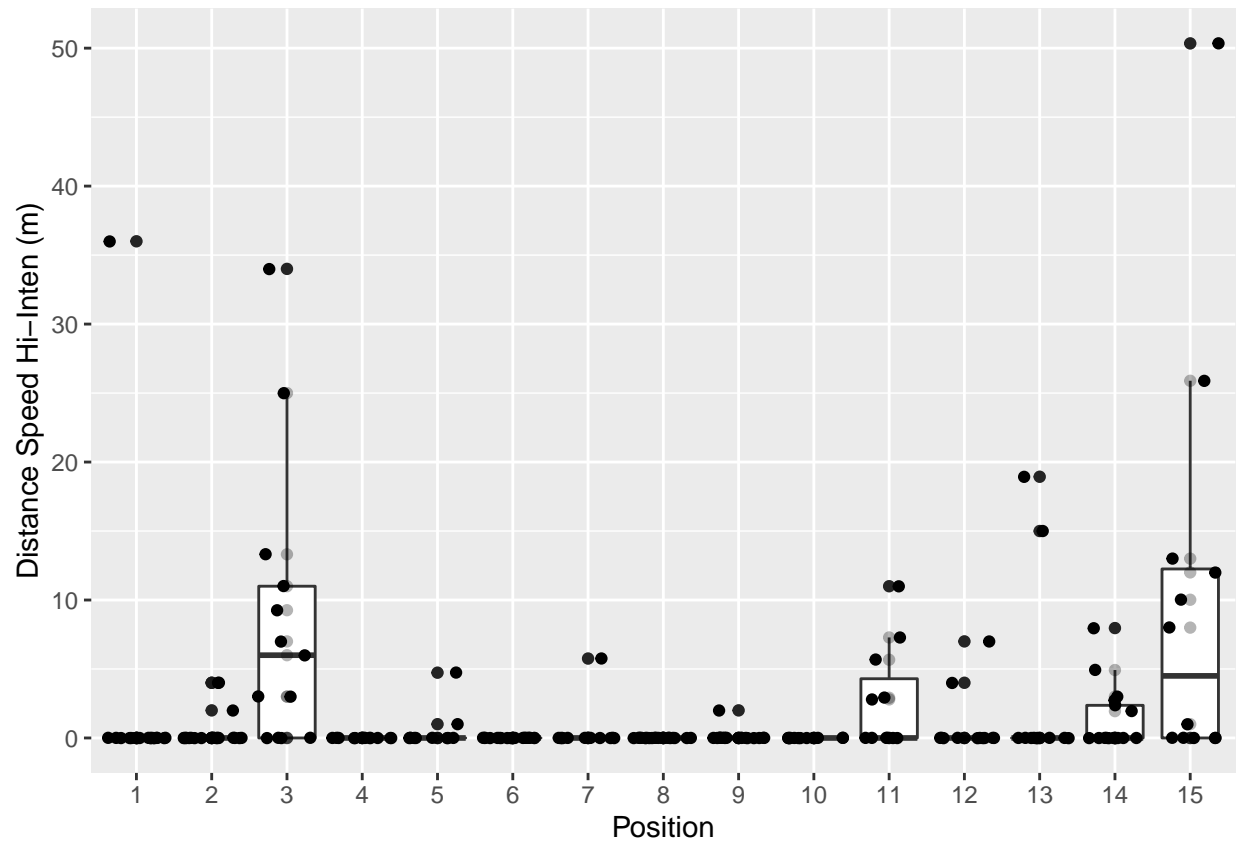
```
## Warning: Removed 24 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 24 rows containing missing values (geom_point).
```

```
## Warning: Removed 24 rows containing missing values (geom_point).
```



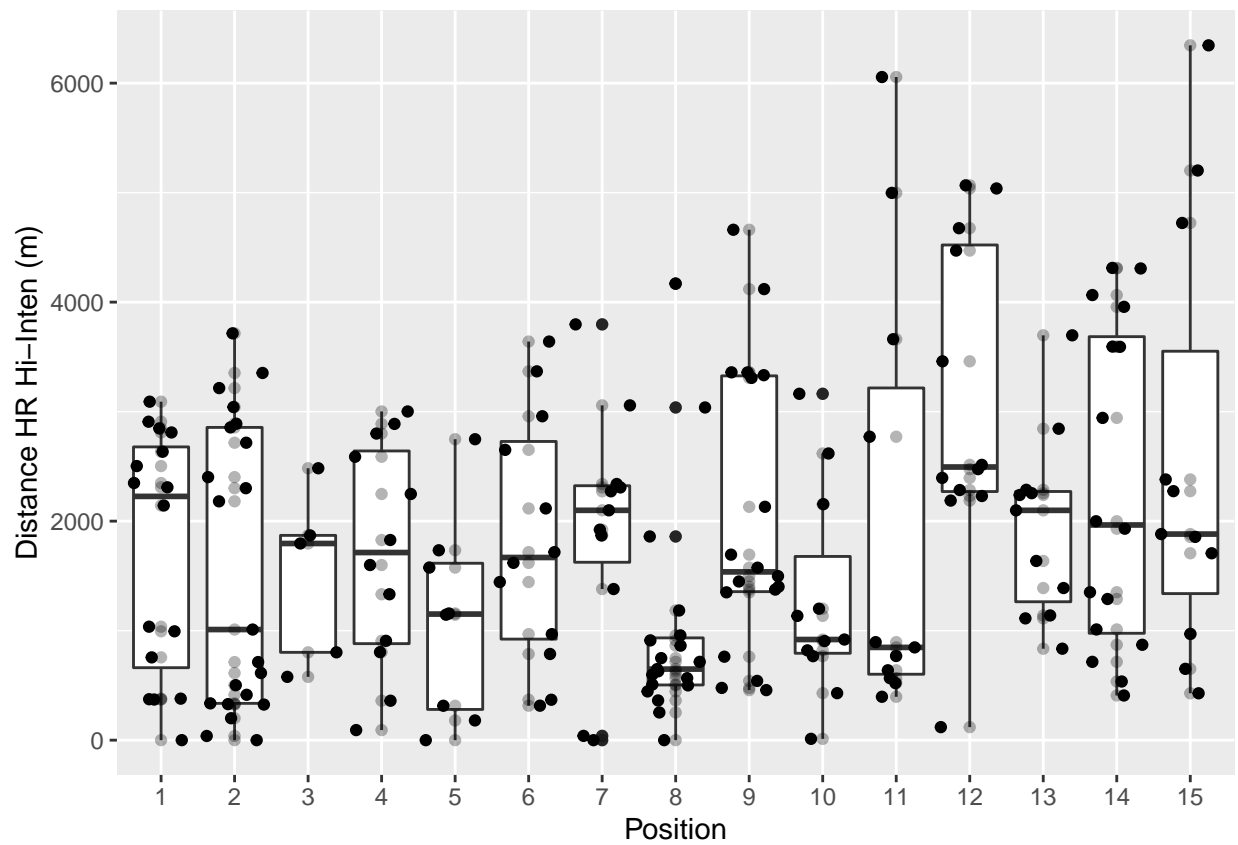


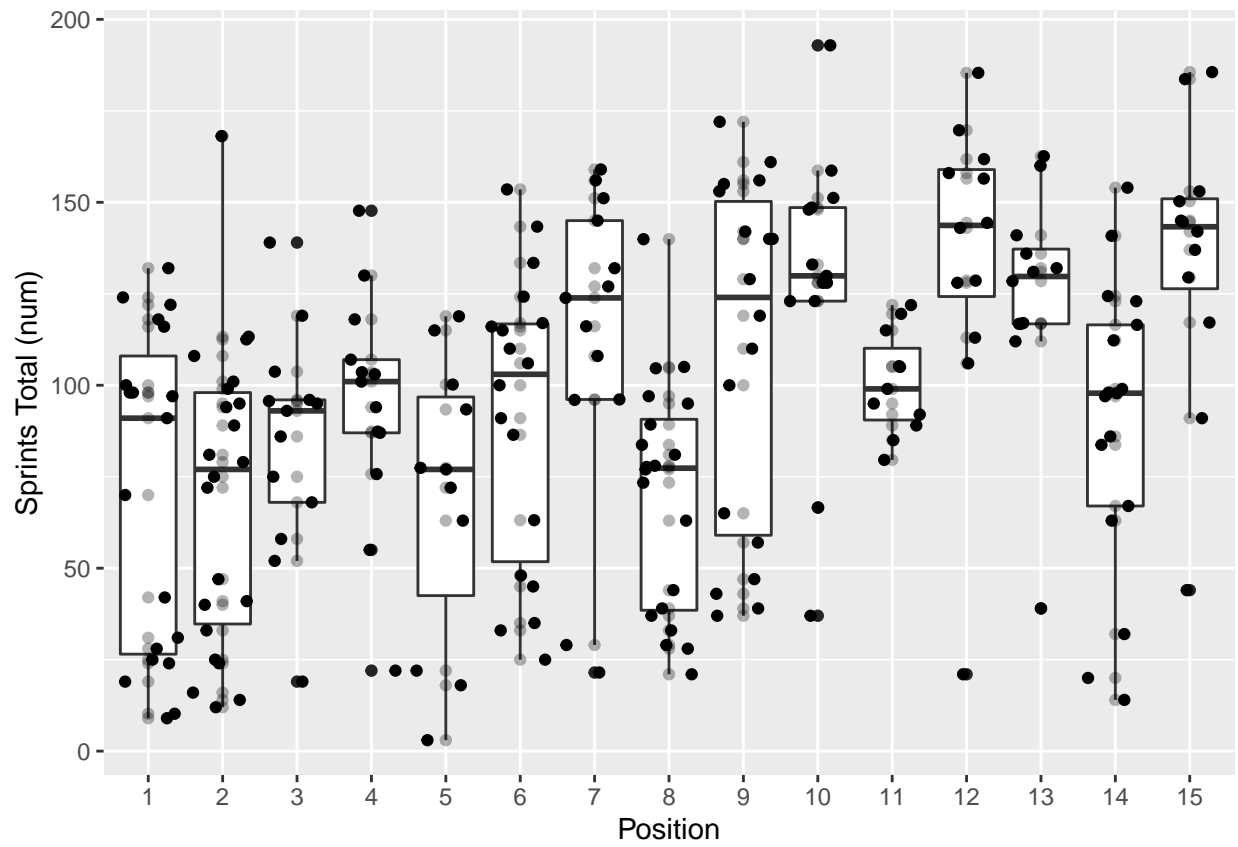


```
## Warning: Removed 30 rows containing non-finite values (stat_boxplot).
```

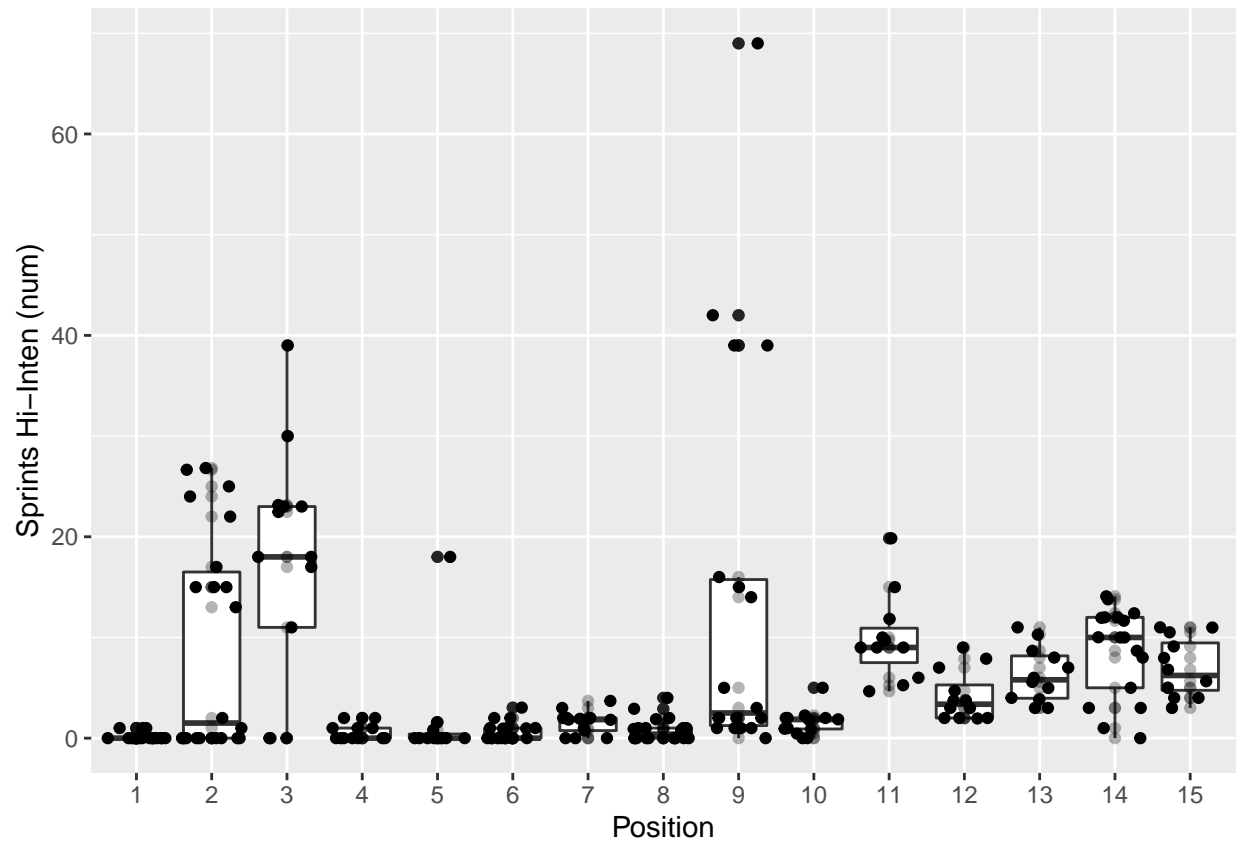
```
## Warning: Removed 30 rows containing missing values (geom_point).
```

```
## Warning: Removed 30 rows containing missing values (geom_point).
```





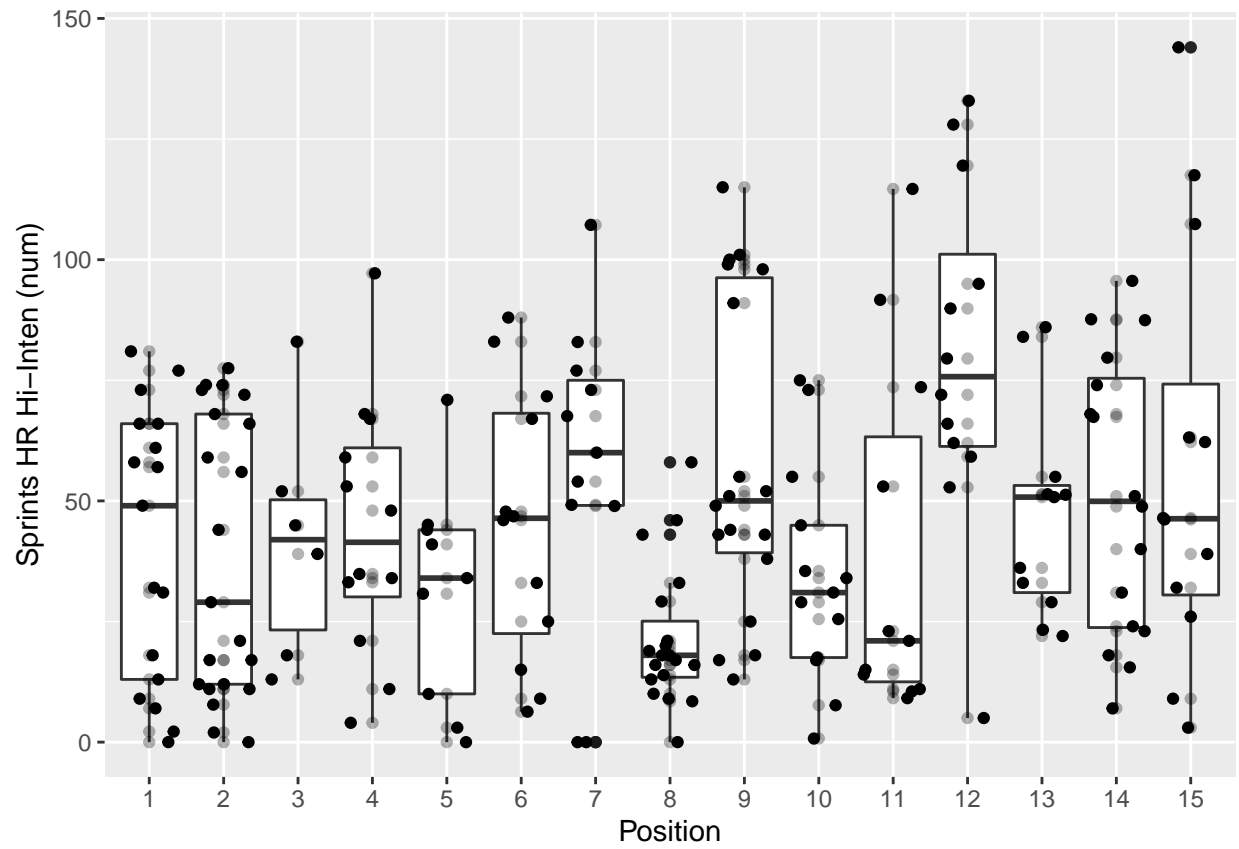




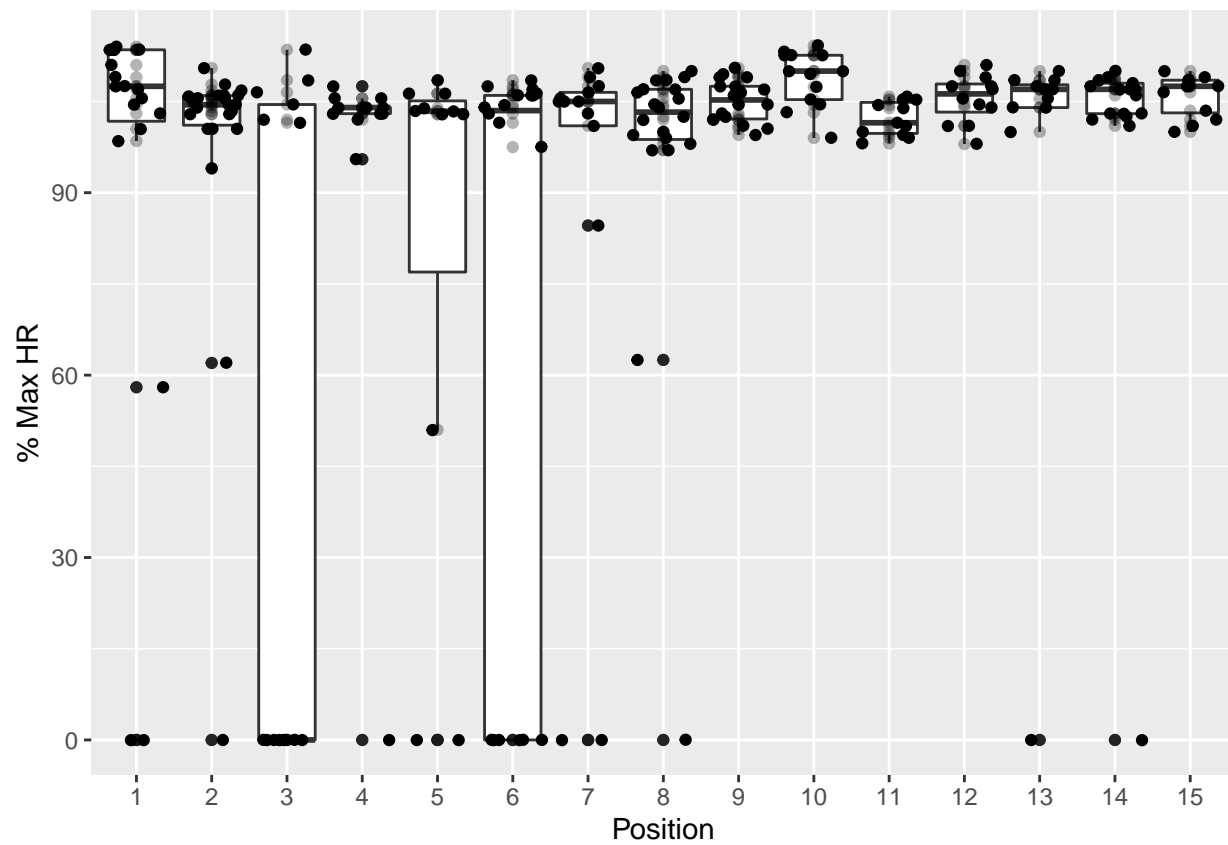
```
## Warning: Removed 24 rows containing non-finite values (stat_boxplot).
```

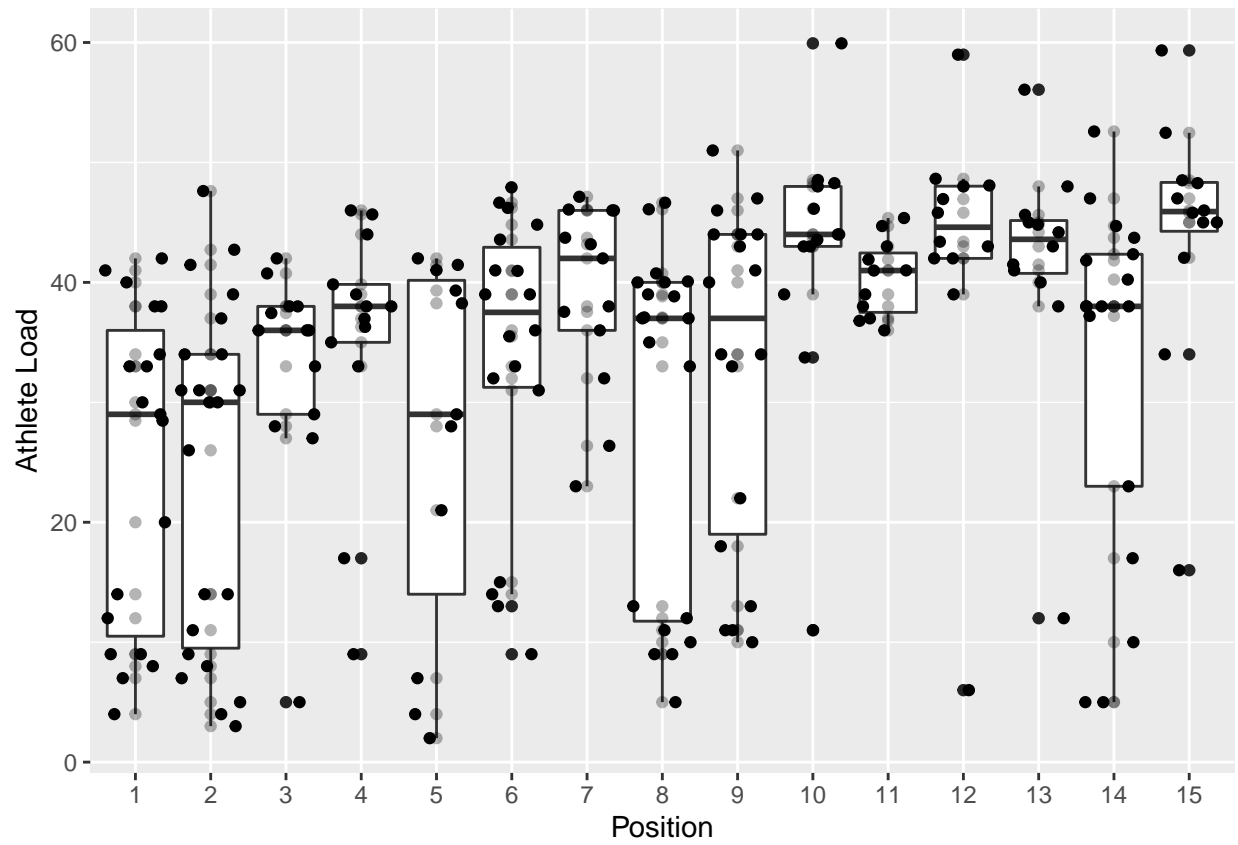
```
## Warning: Removed 24 rows containing missing values (geom_point).
```

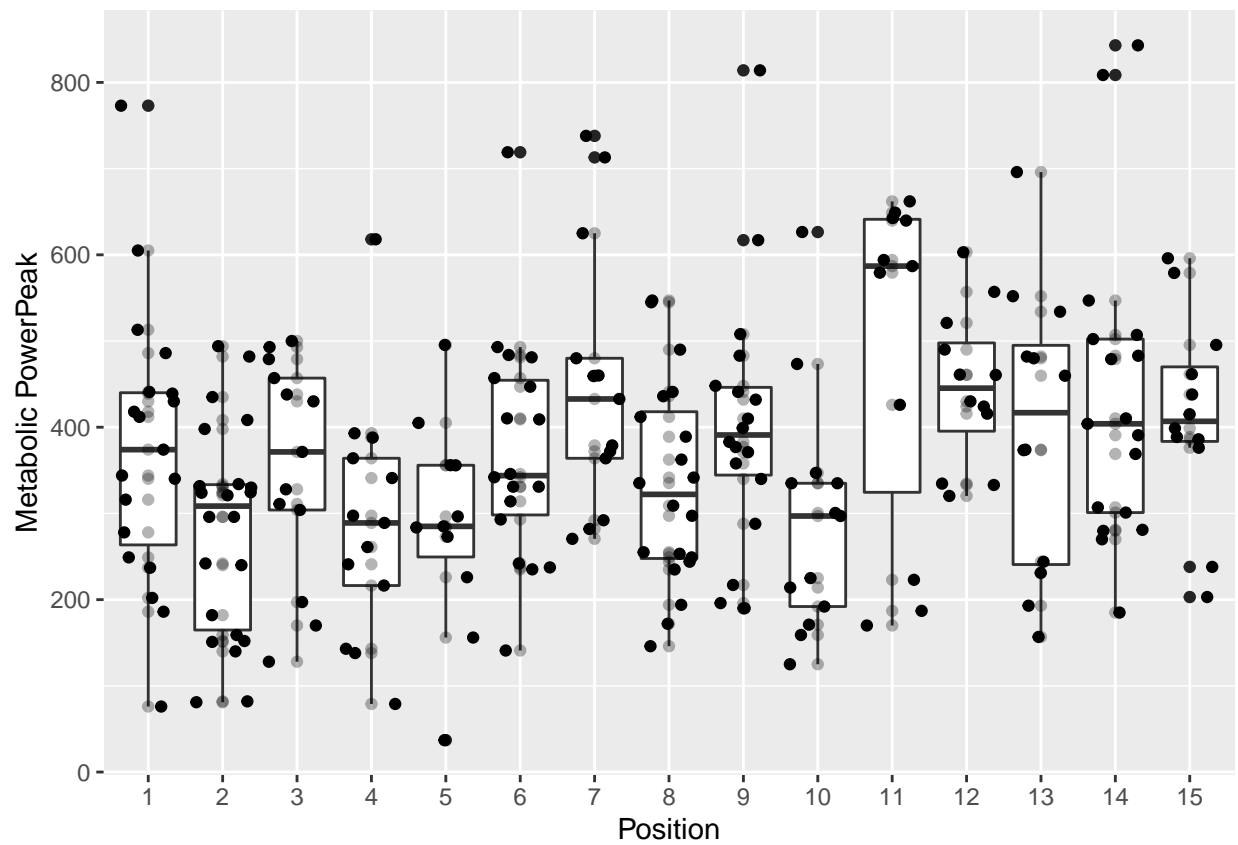
```
## Warning: Removed 24 rows containing missing values (geom_point).
```

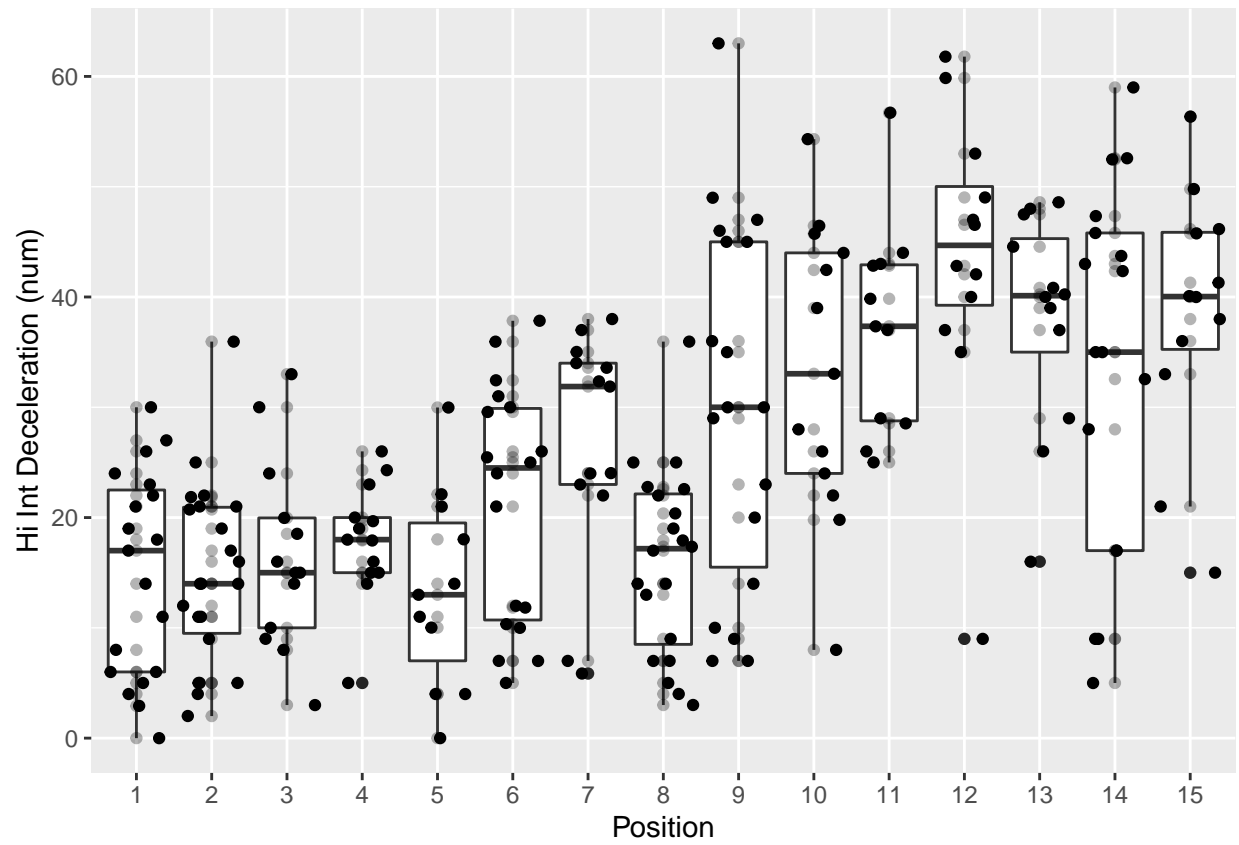


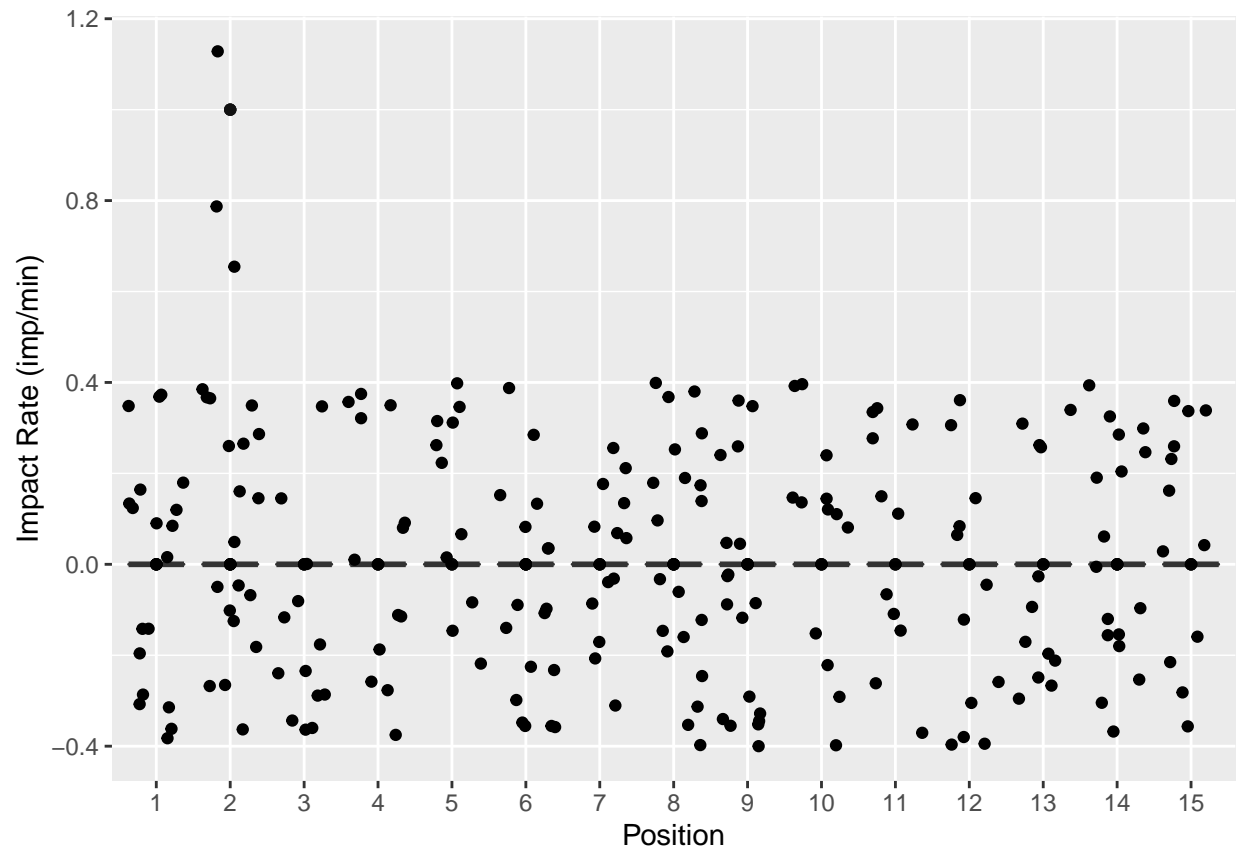




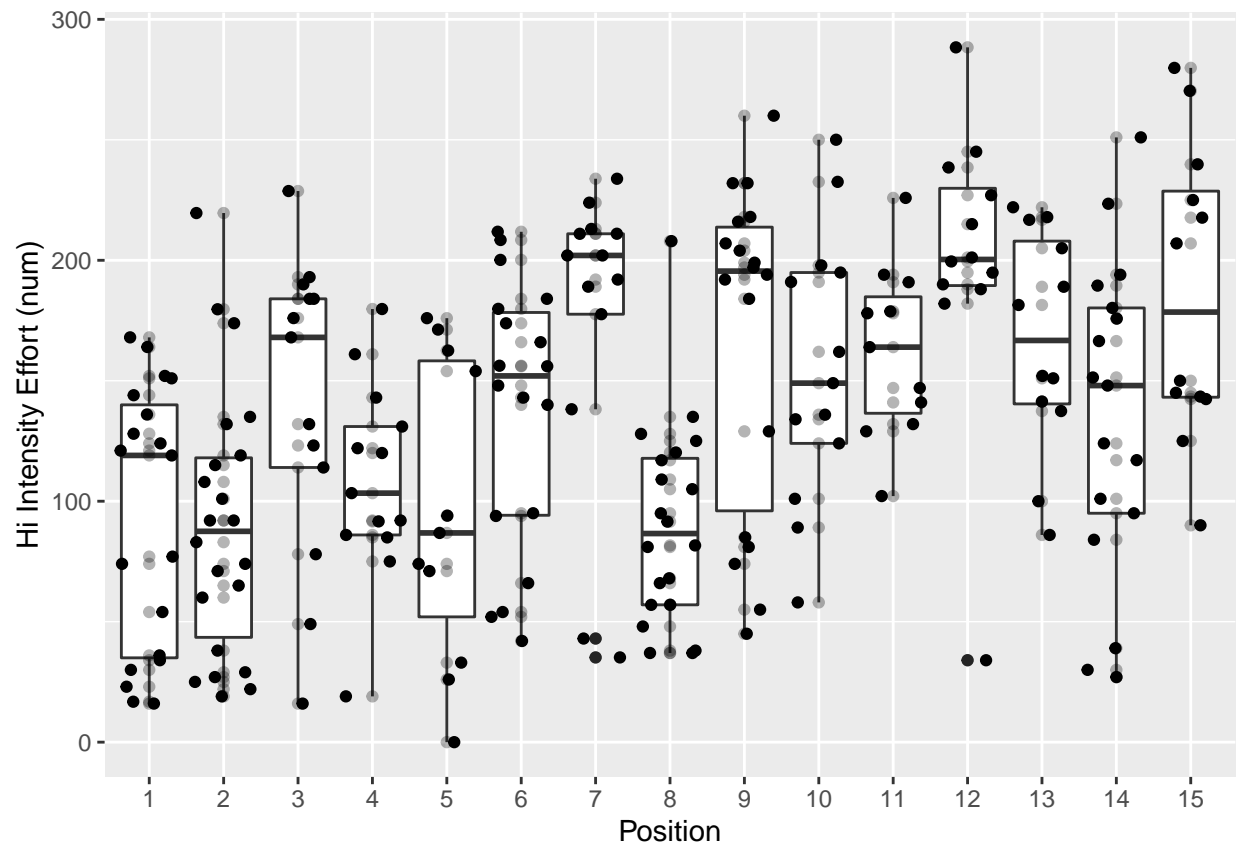


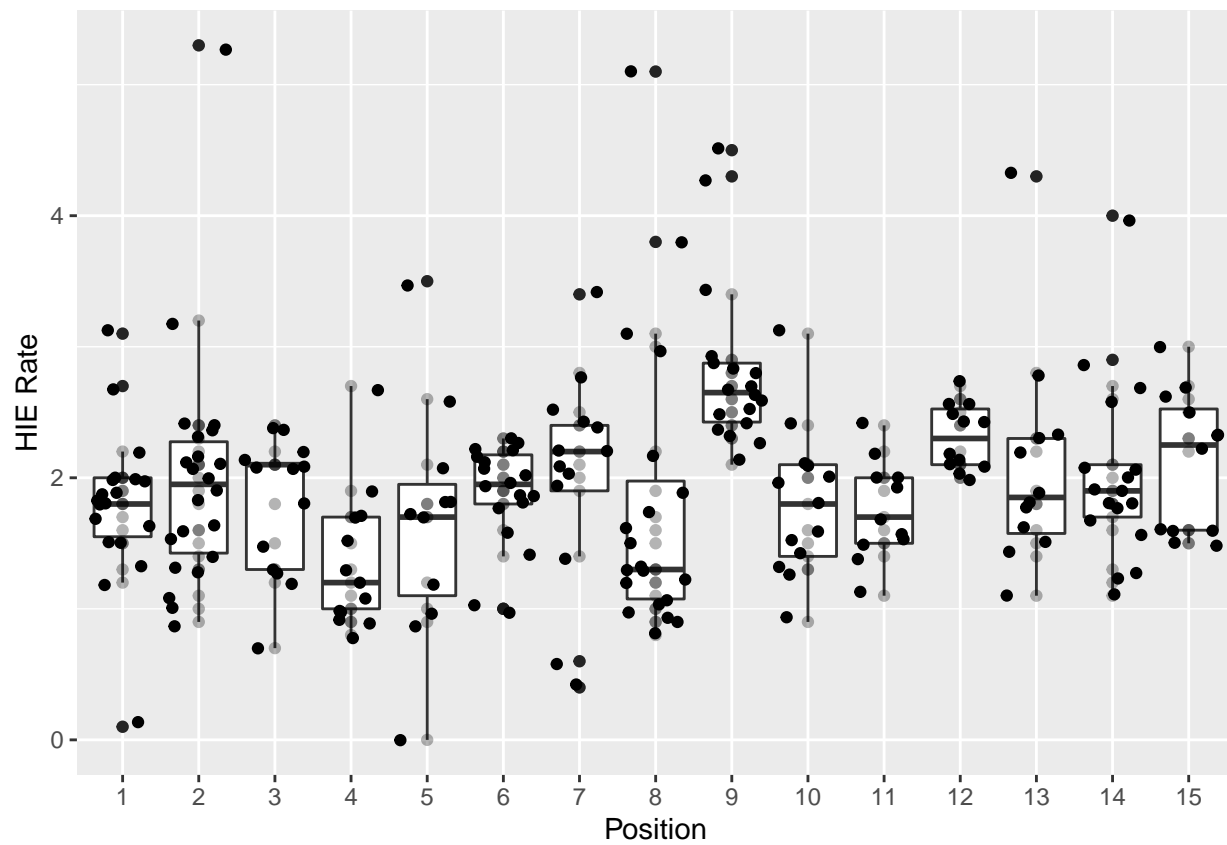


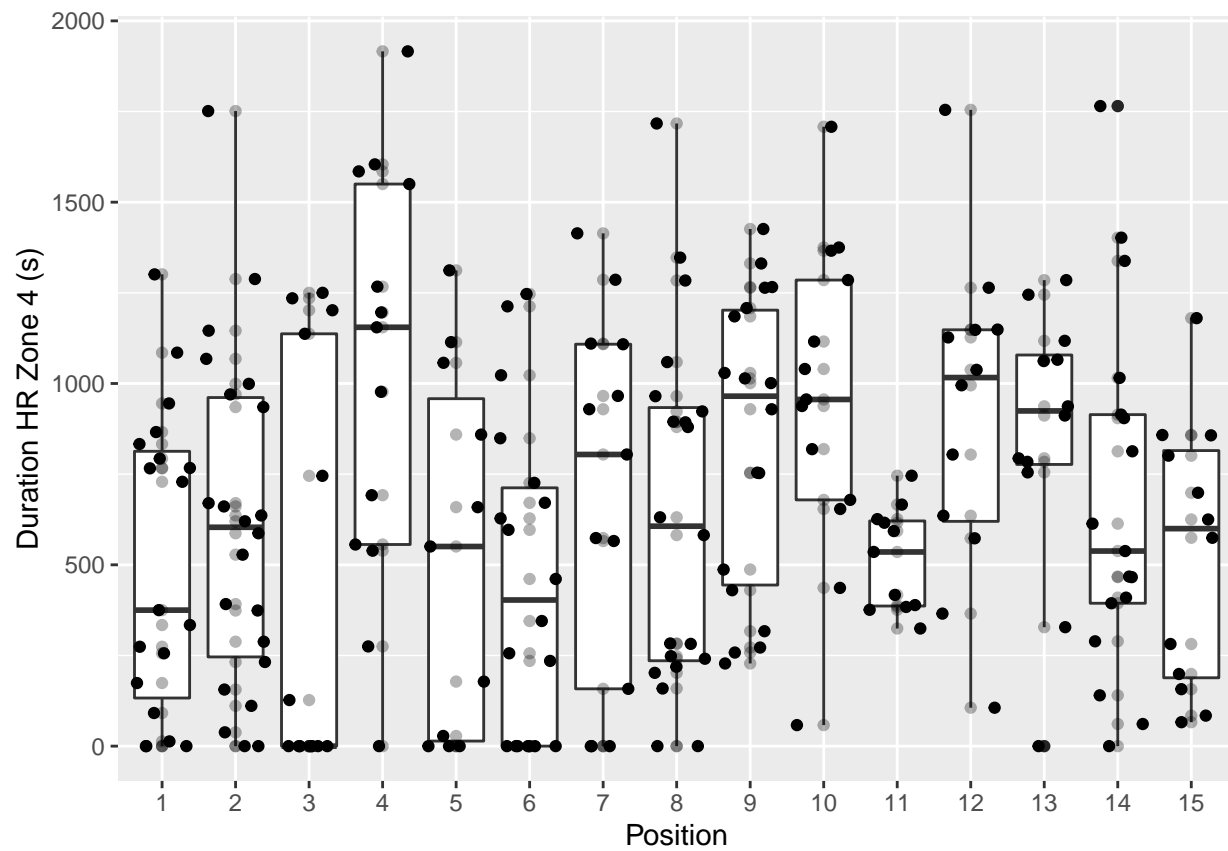


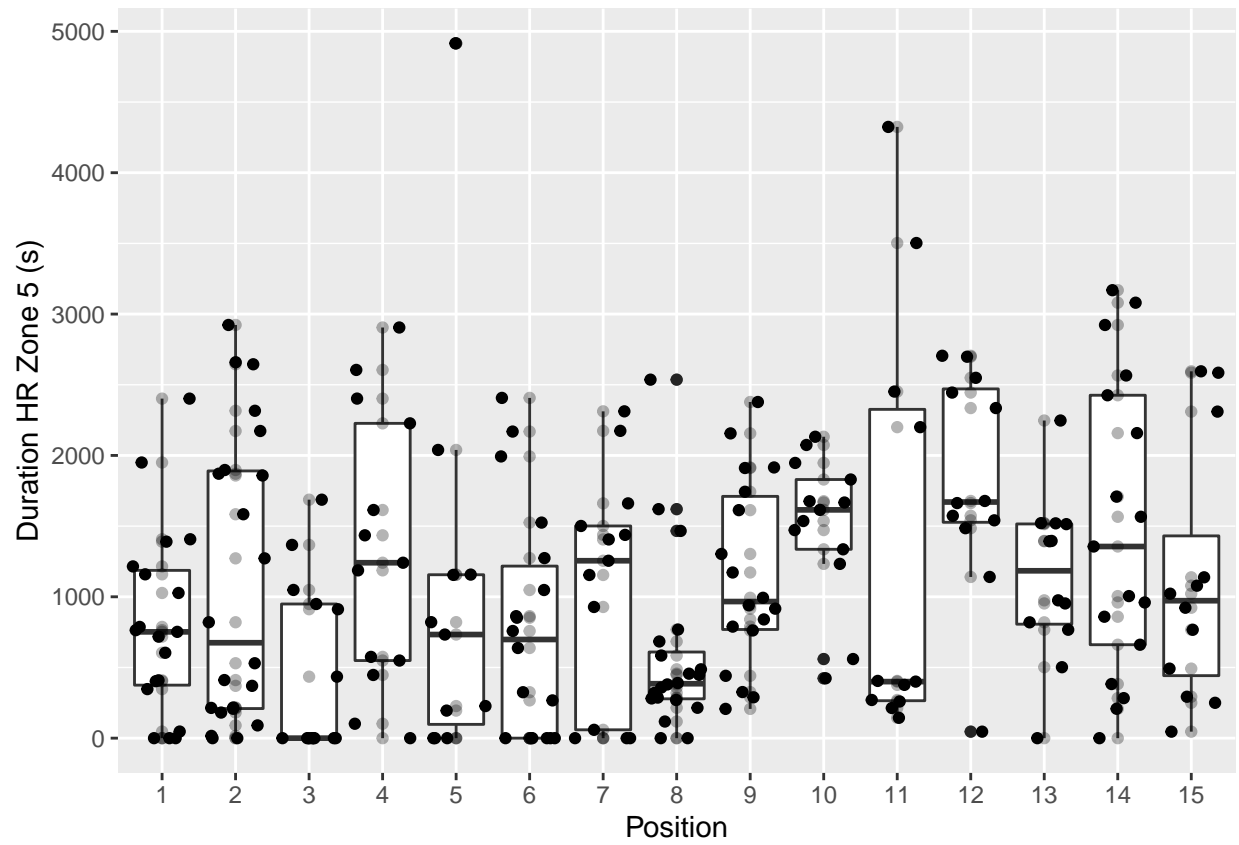


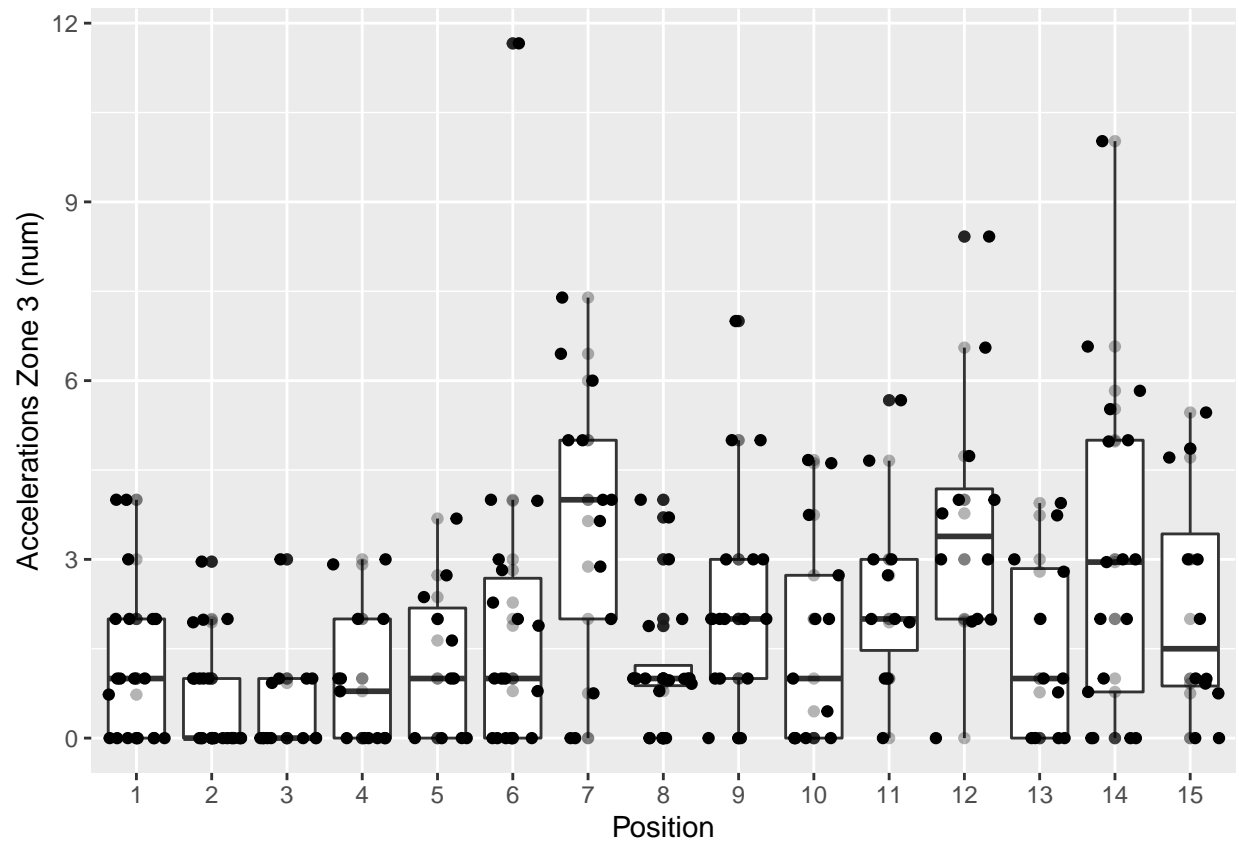


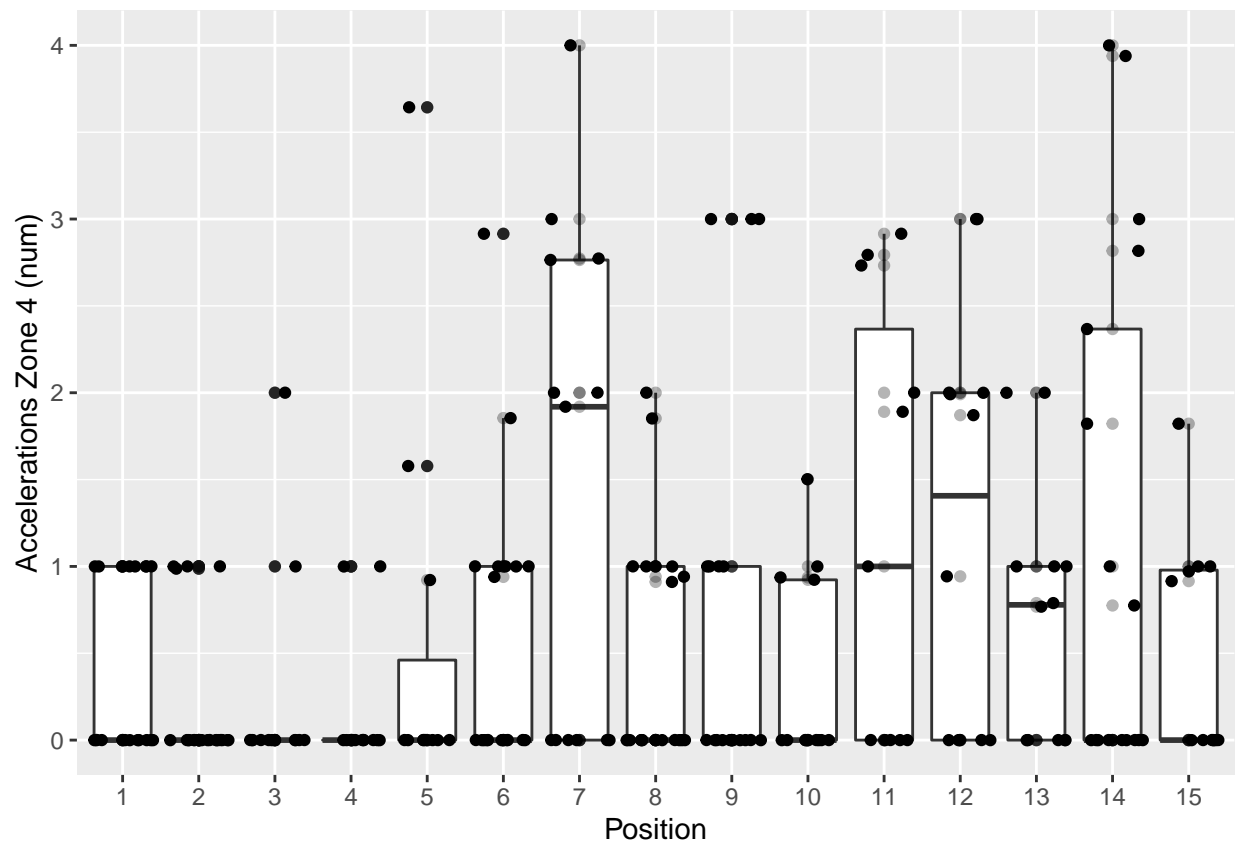


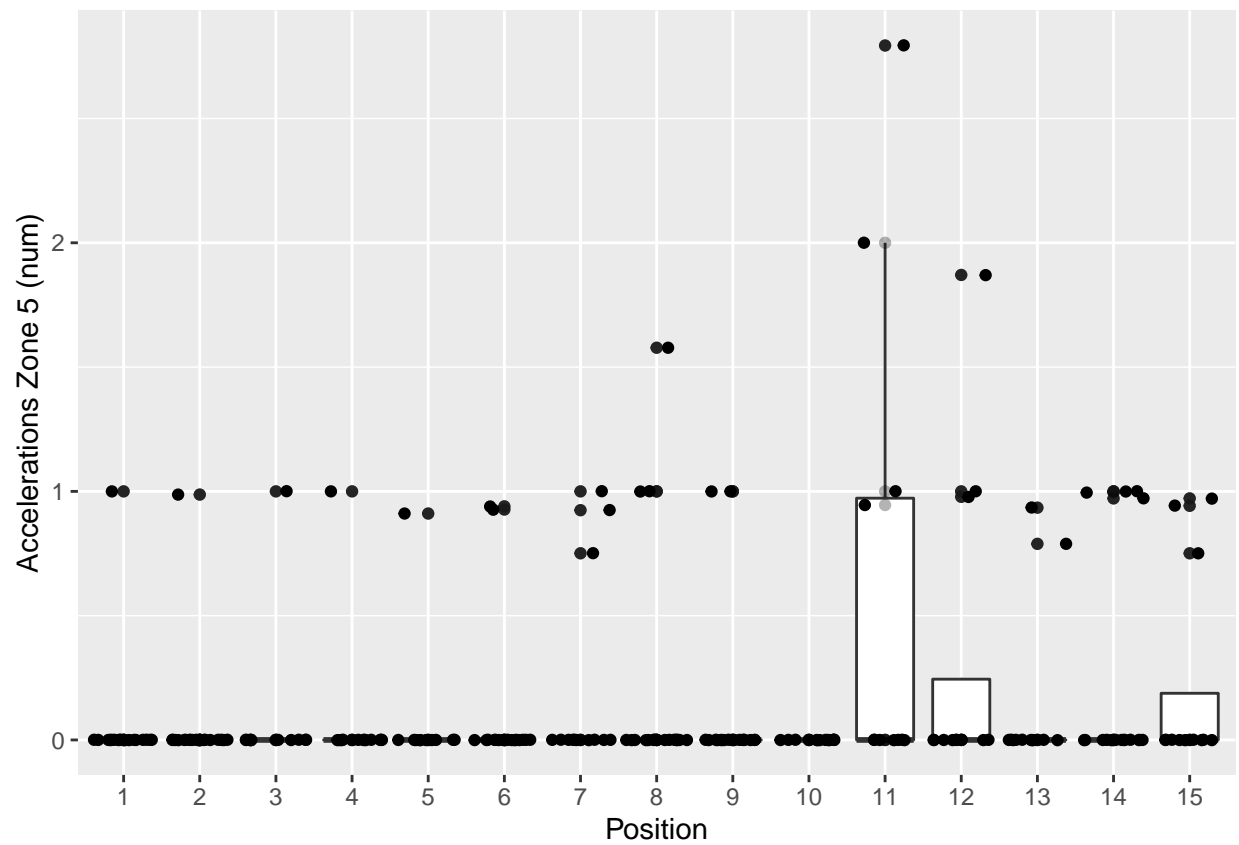


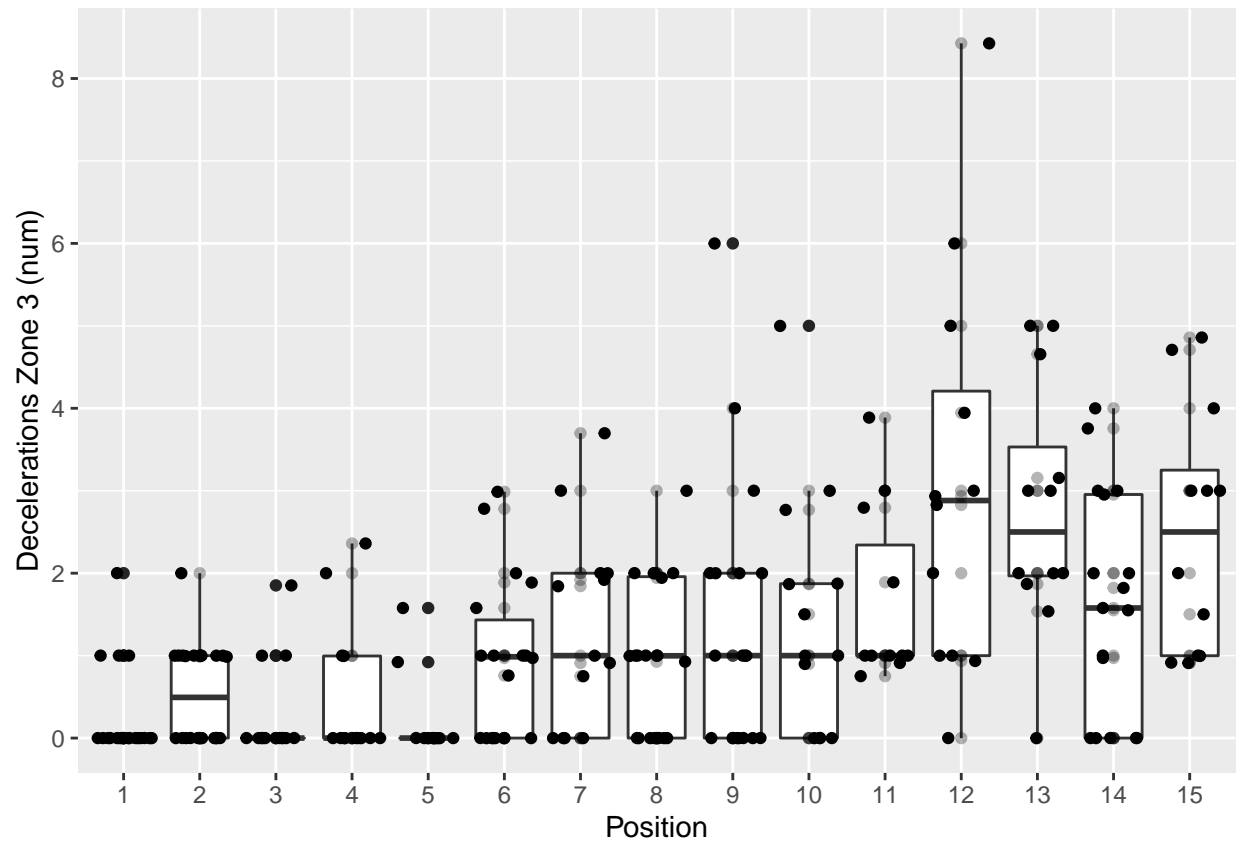




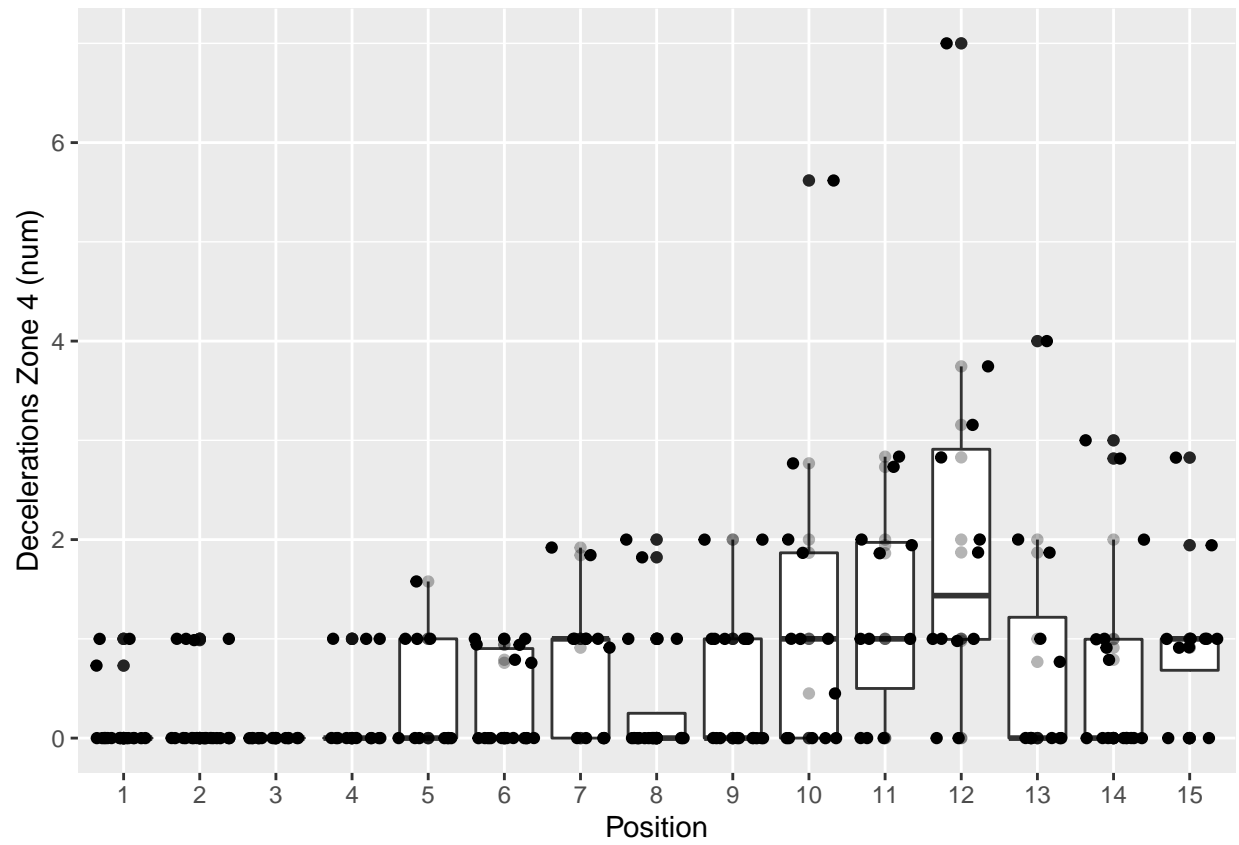




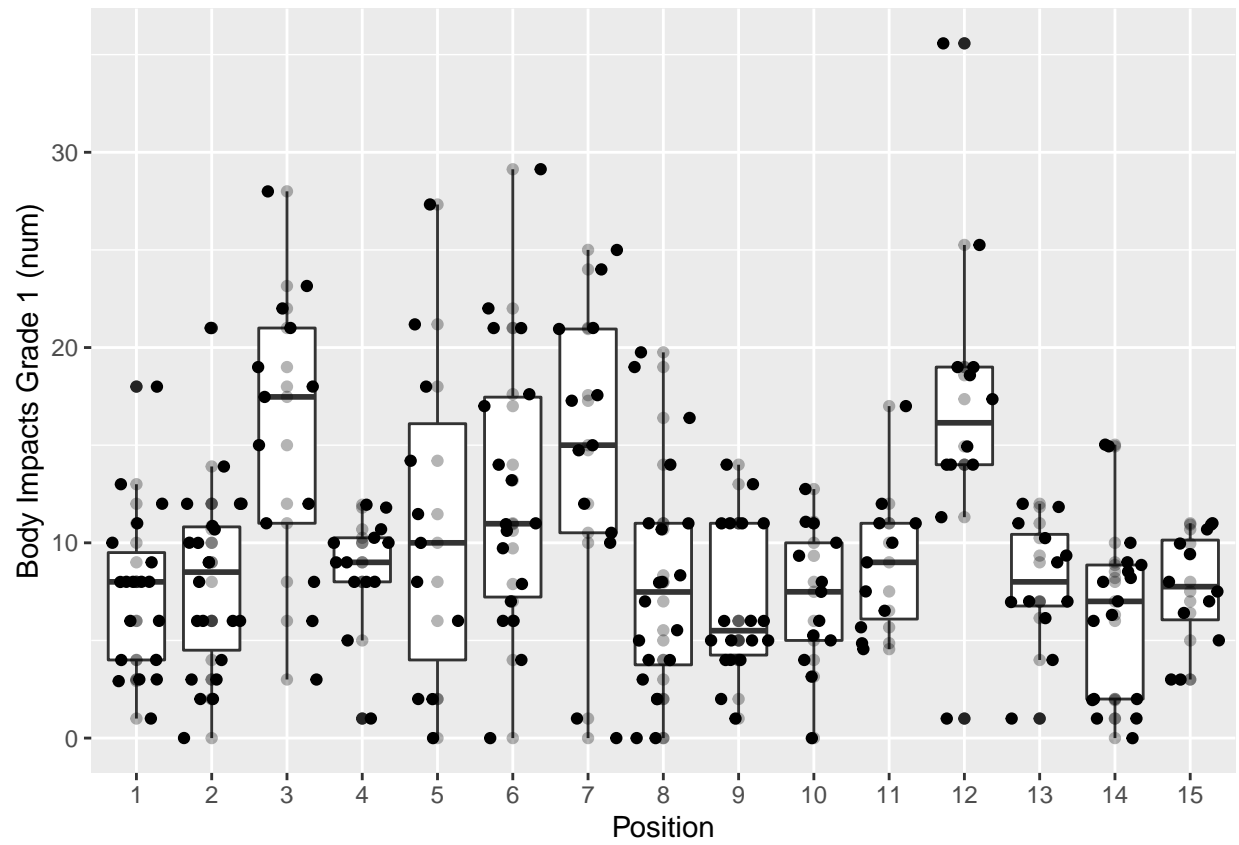


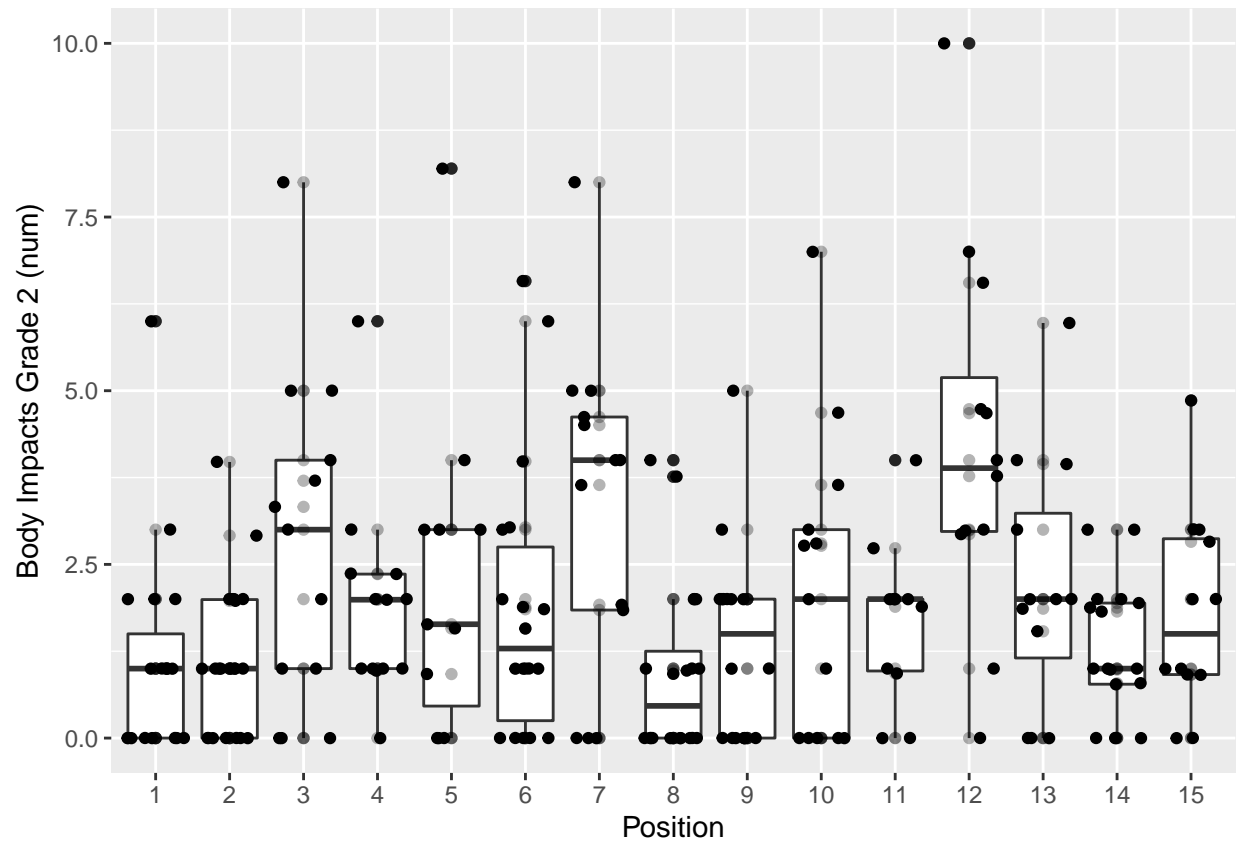


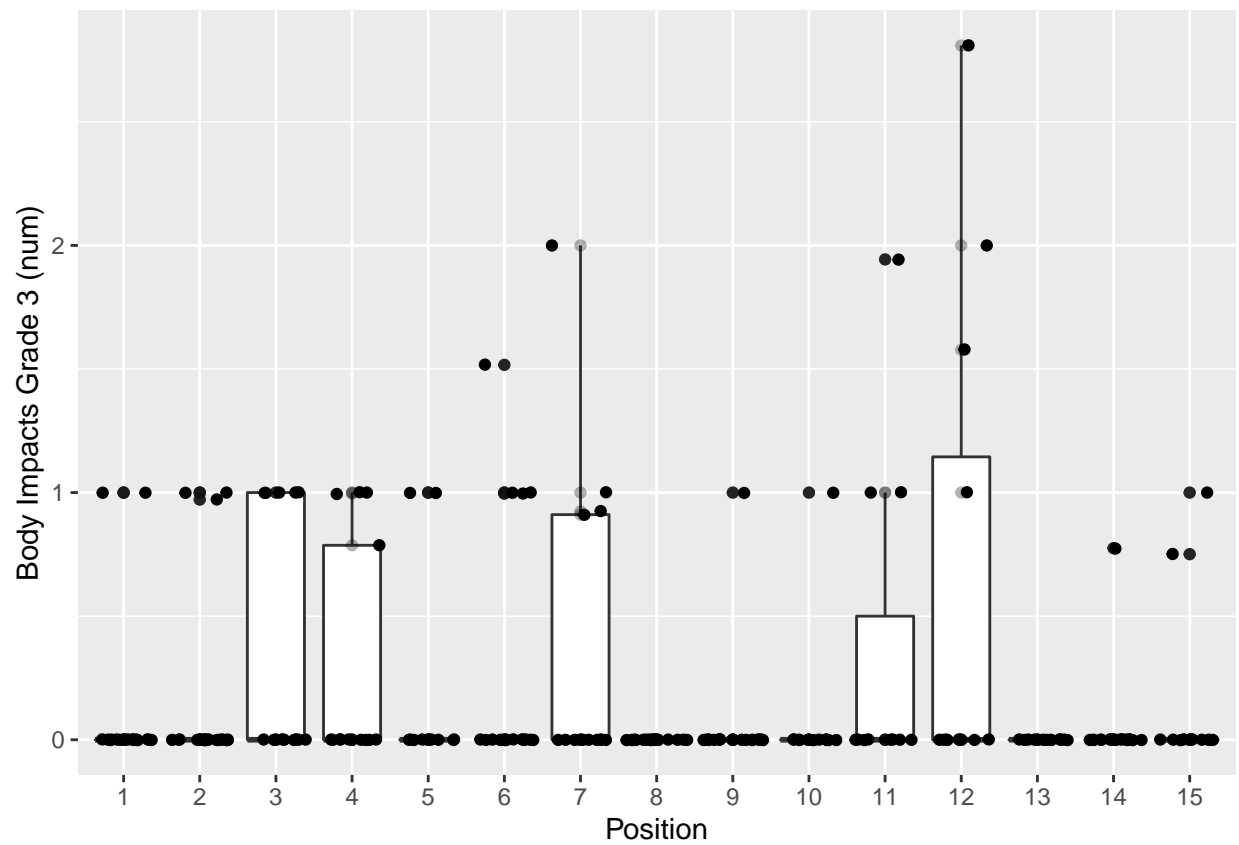


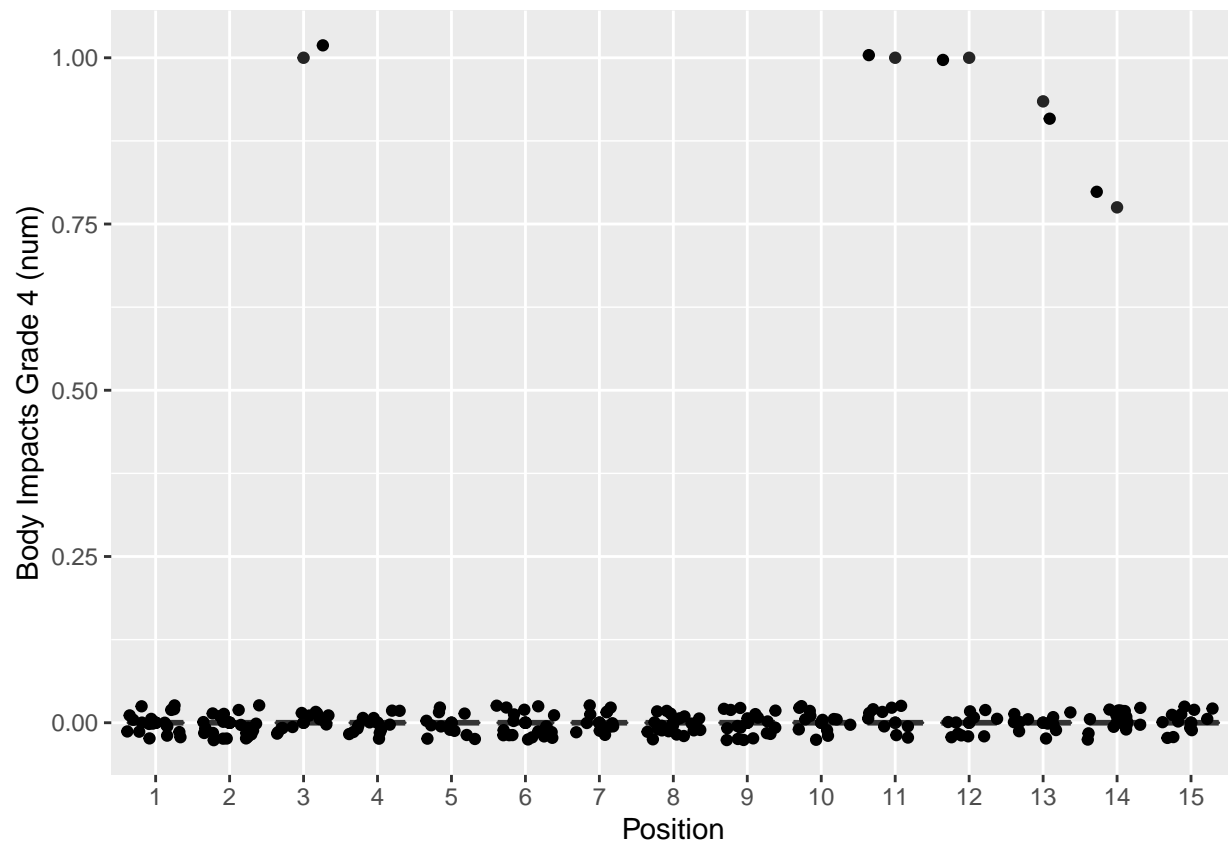


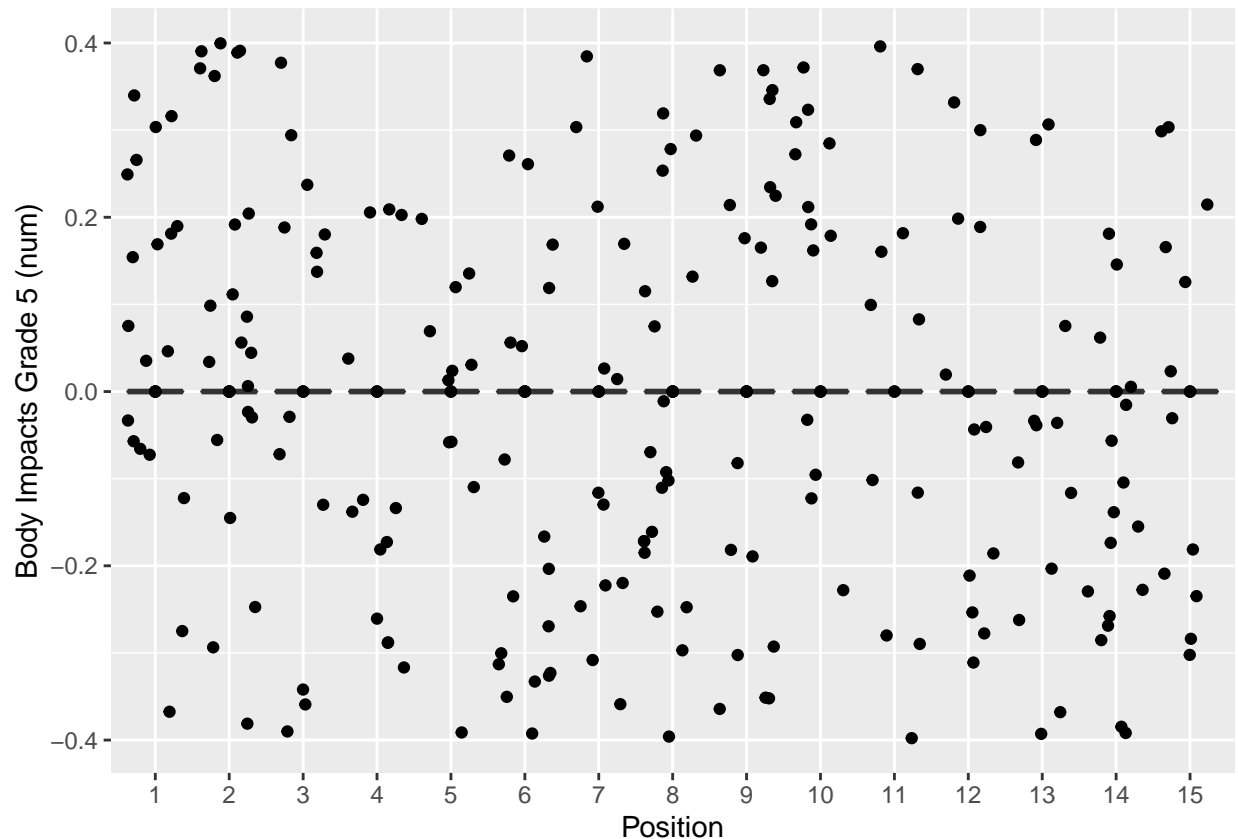












There are a lot of plots here, but the main points I gleaned from these were that:

Backs have high values for most sprint, distance and deceleration stats;

Forwards tend to have high values for low grade body impacts, and low values for energy related measures;

Heart rate threshold-related measures (e.g. `Duration HR Hi-Inten (s)`, `Distance HR Hi-Inten (m)`, `Sprints HR Hi-Inten (num)`) do not have any particular trend outside of inside centre (#12) and fullback (#15) registering high values;

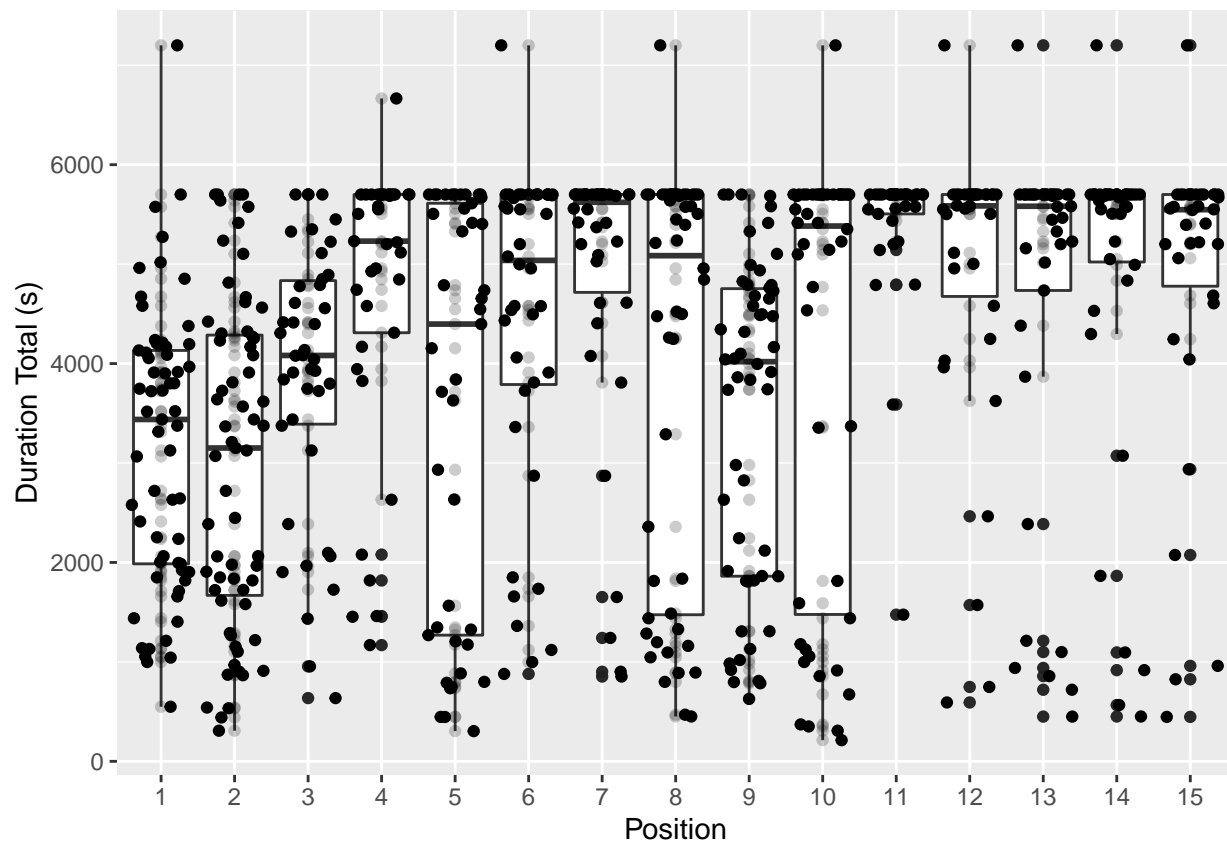
There are some observations for the heart rate measures `HR Max Total (bpm)` and `% Max HR` that are zeroes, which are unrealistic (as this would imply the players died - more likely to be due to a malfunction in the GPS unit);

Scrum-half (#9) has a much higher high-intensity sprint number than every other position;

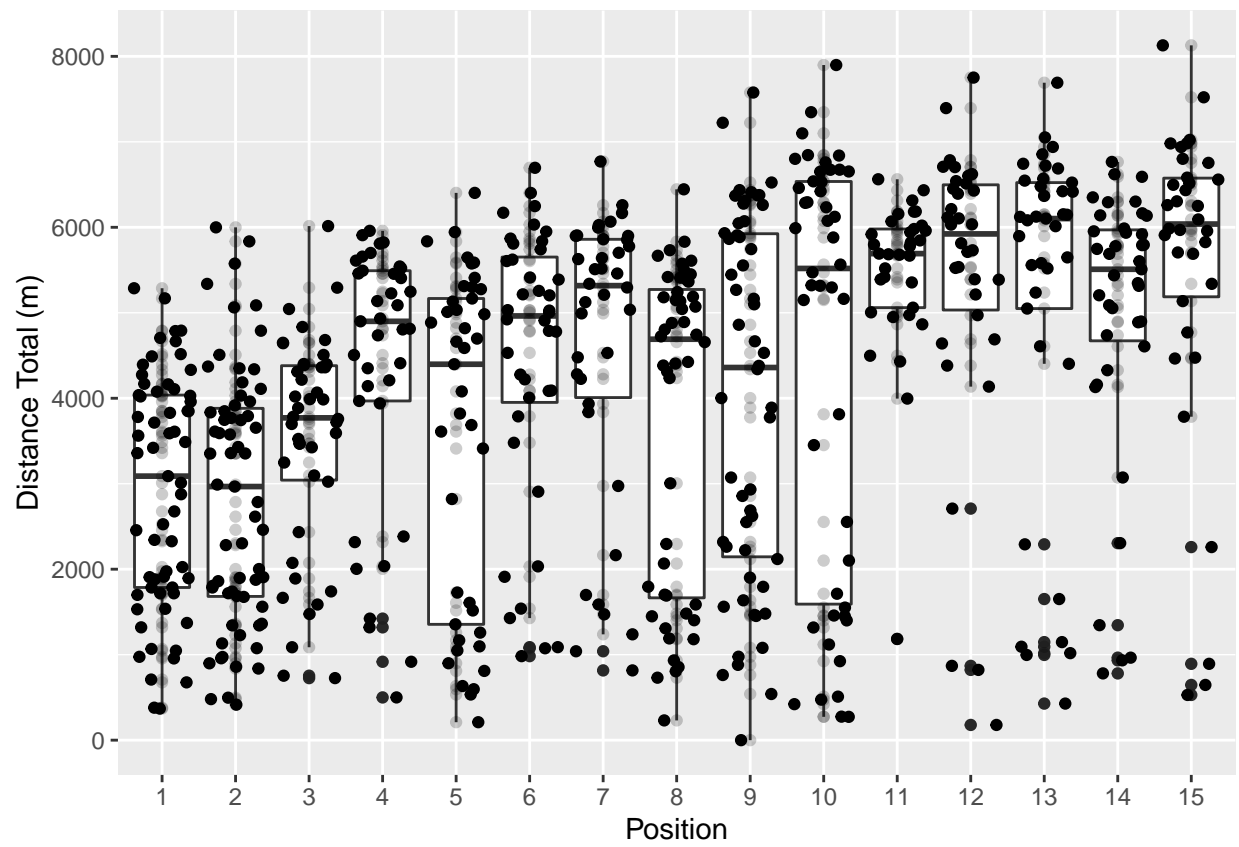
Inside centre (#12) registers particularly high values for accelerations, decelerations and body impacts, but also high values overall for most of the other variables.

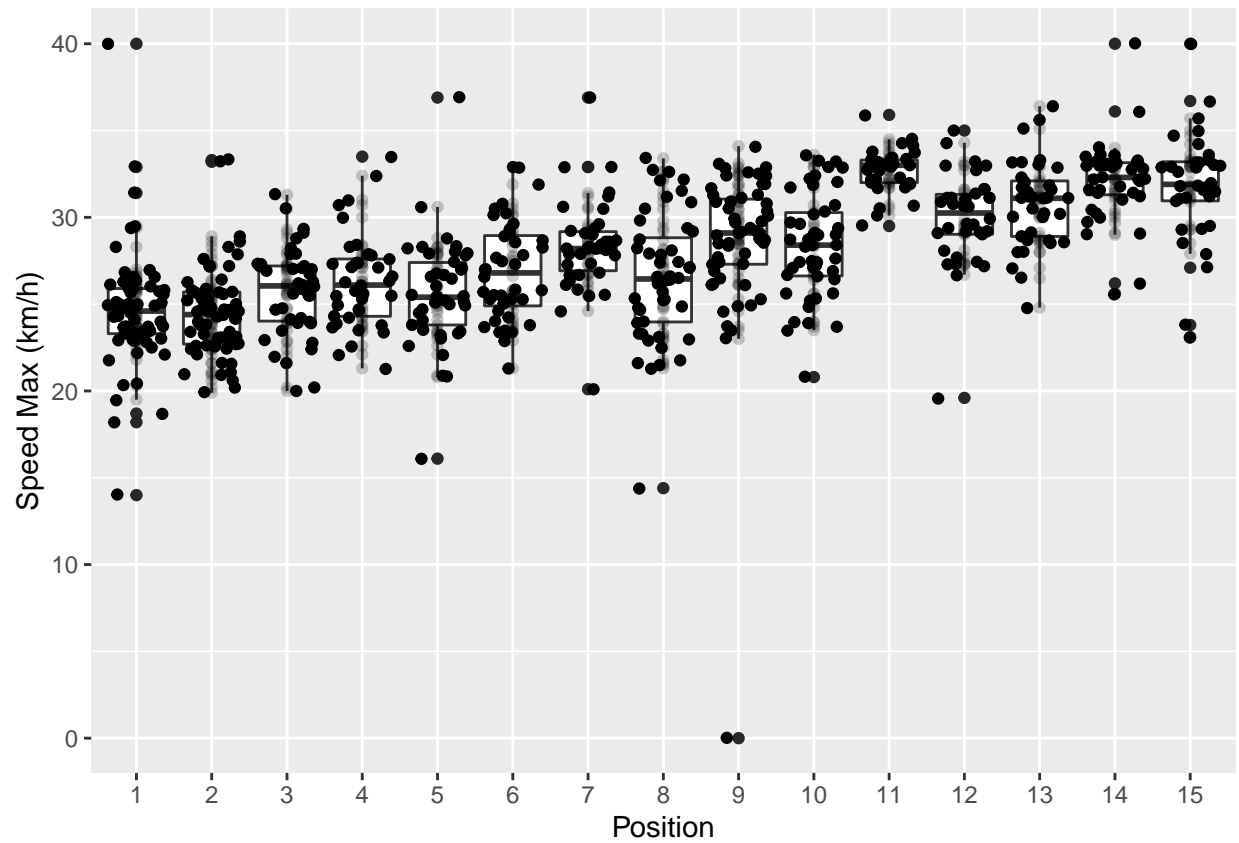
Now, the remaining variables shared between the 2018, 2019 and 2020 datasets are plotted.

```
# Plotting the other variables
for (v in 5:17) {
  print(ggplot(combinedData, aes(Position, combinedData[, v])) +
    geom_boxplot() +
    geom_point(alpha = 0.2) +
    ylab(colnames(combinedData)[v]) +
    geom_jitter())
}
```





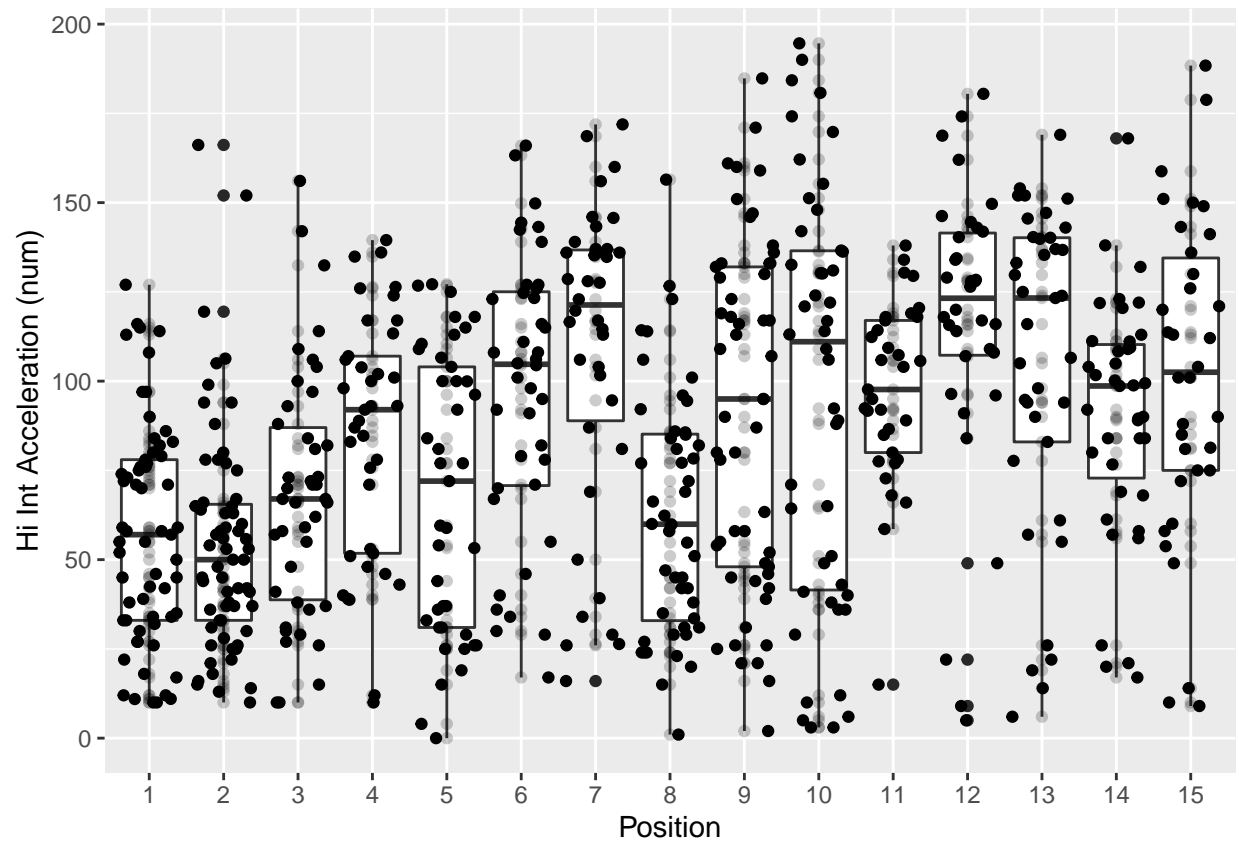




```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

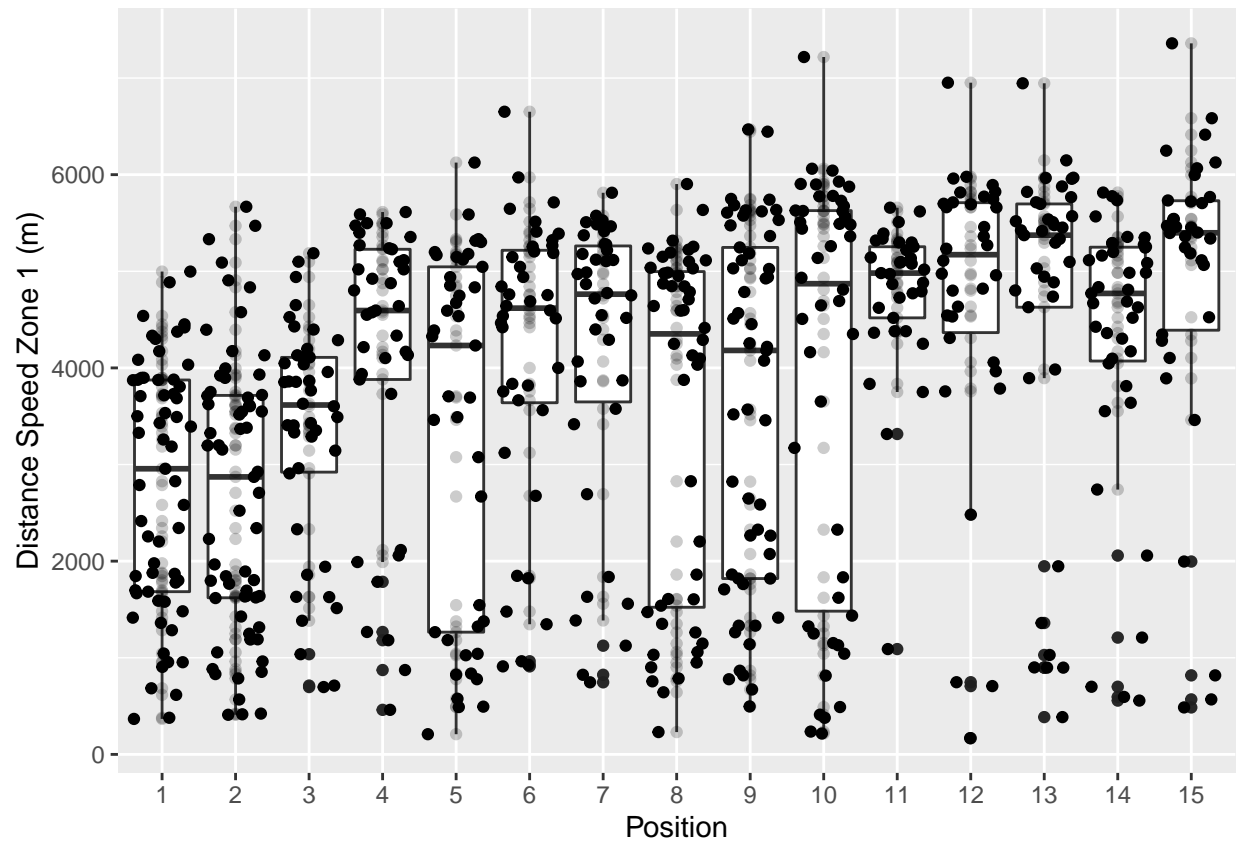
```
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

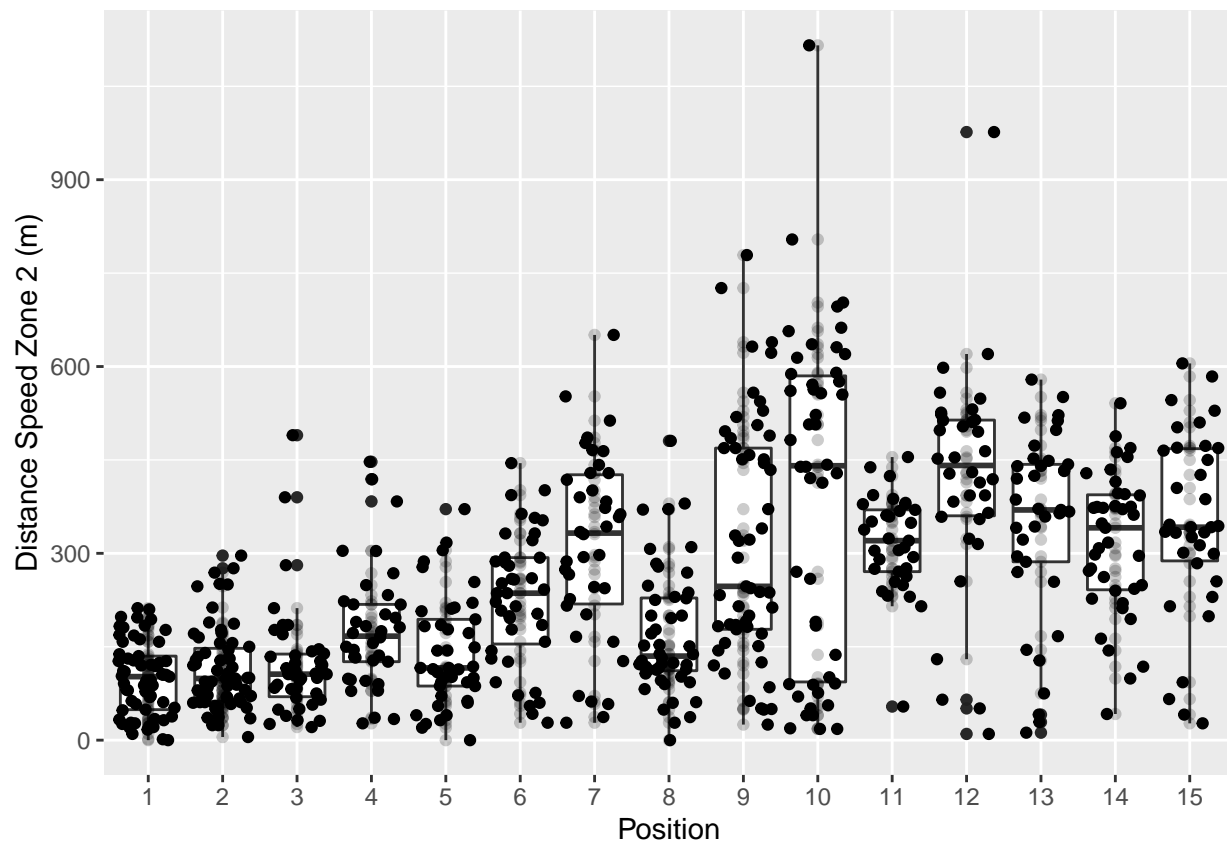
```
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

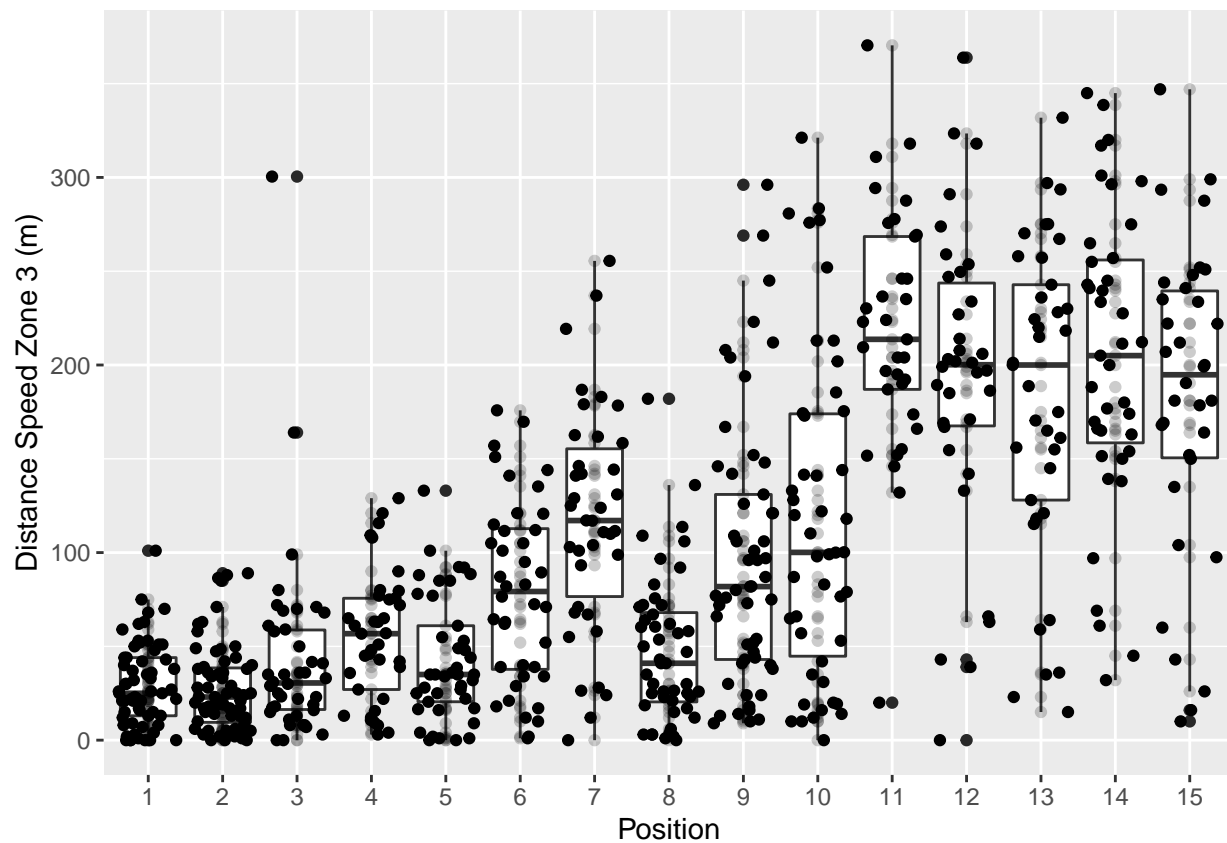
```
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

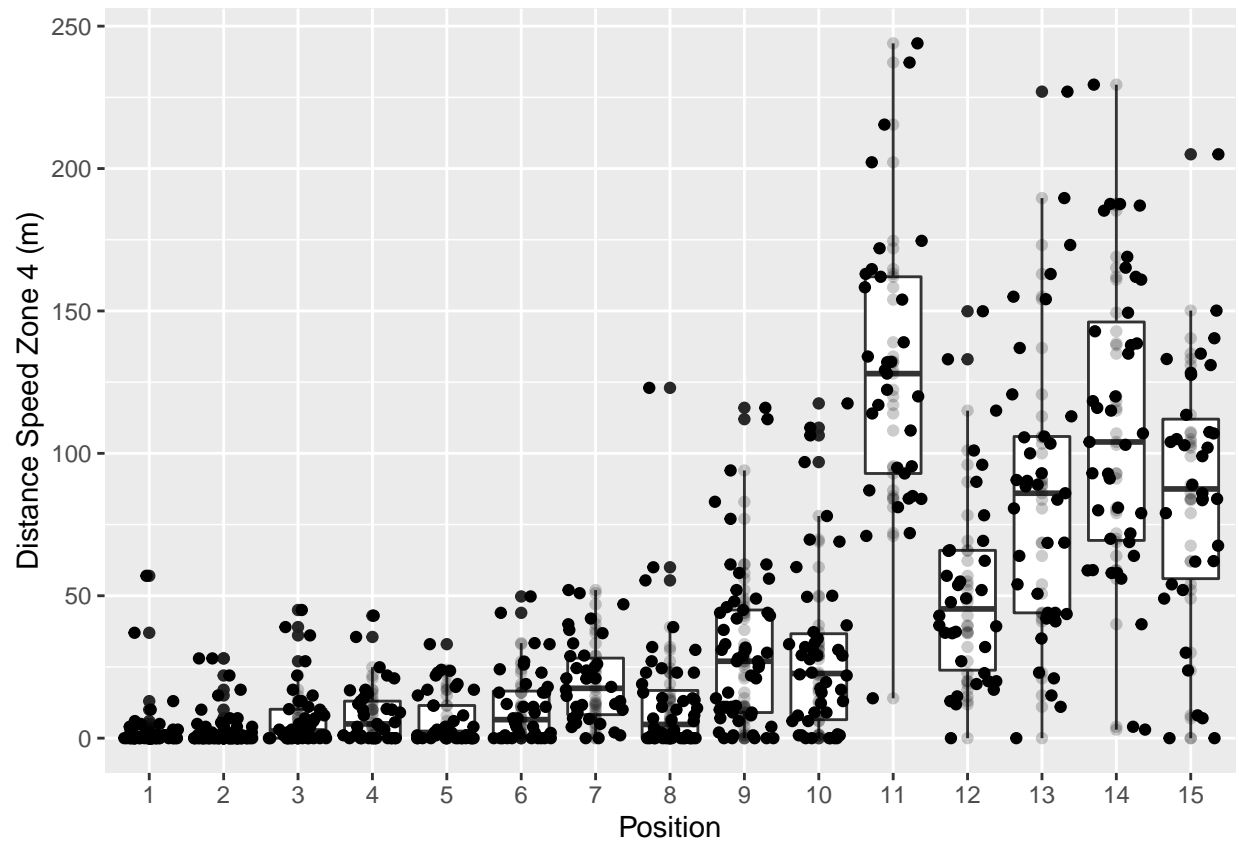
```
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

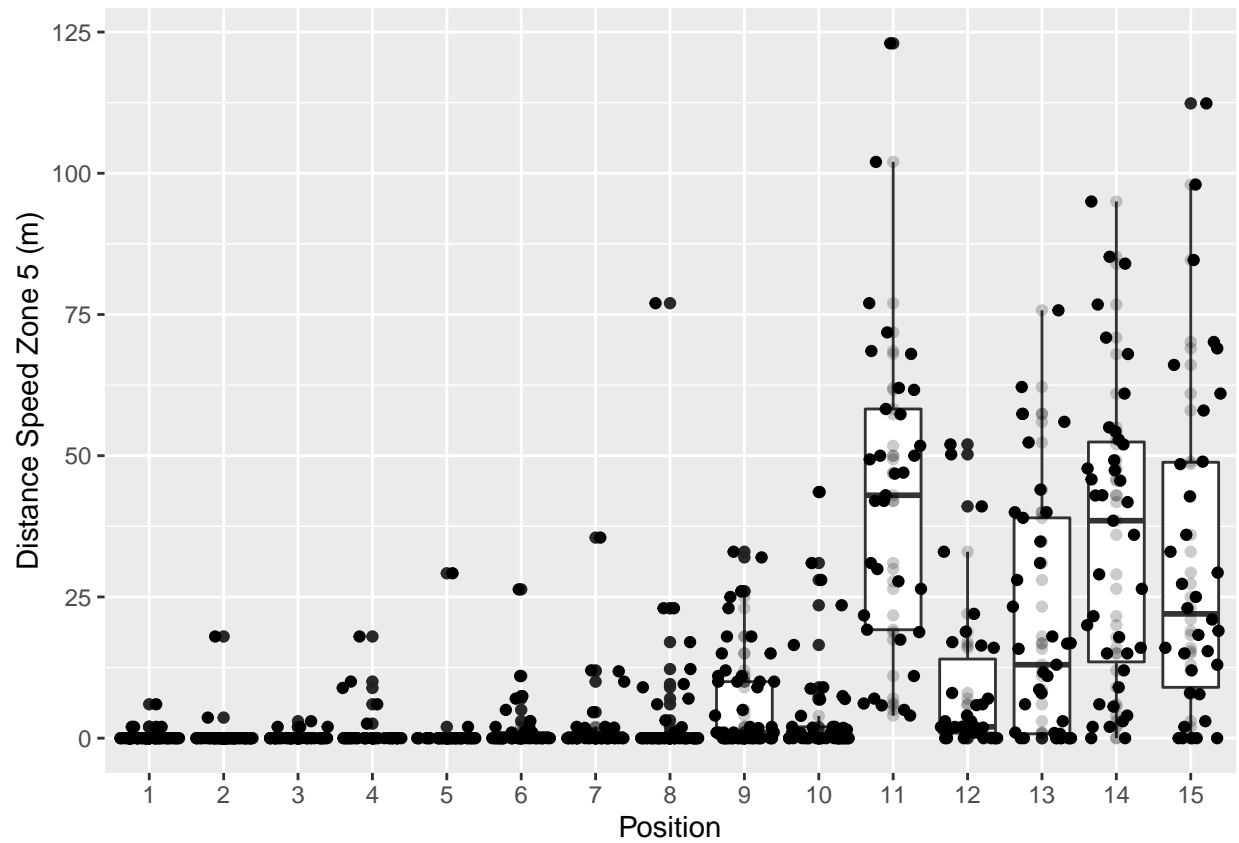
```
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

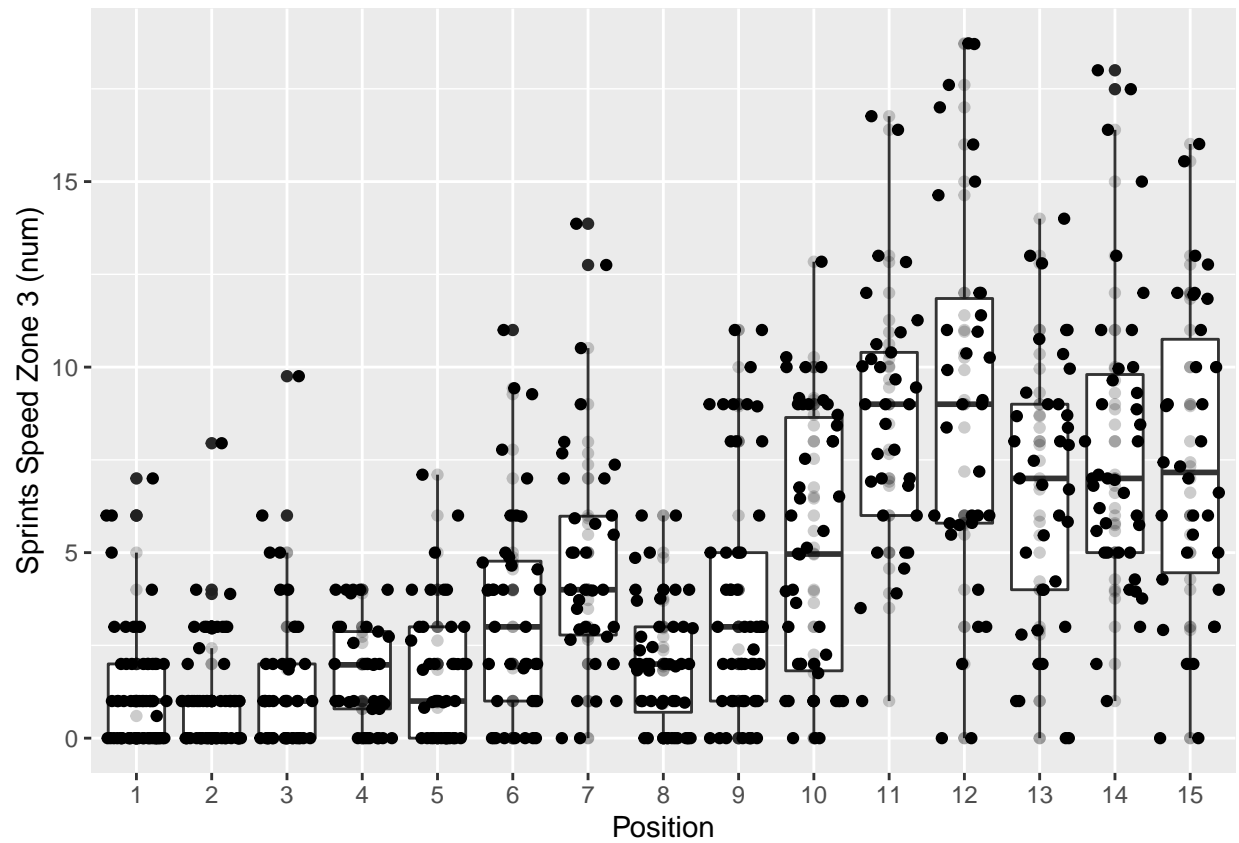


```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

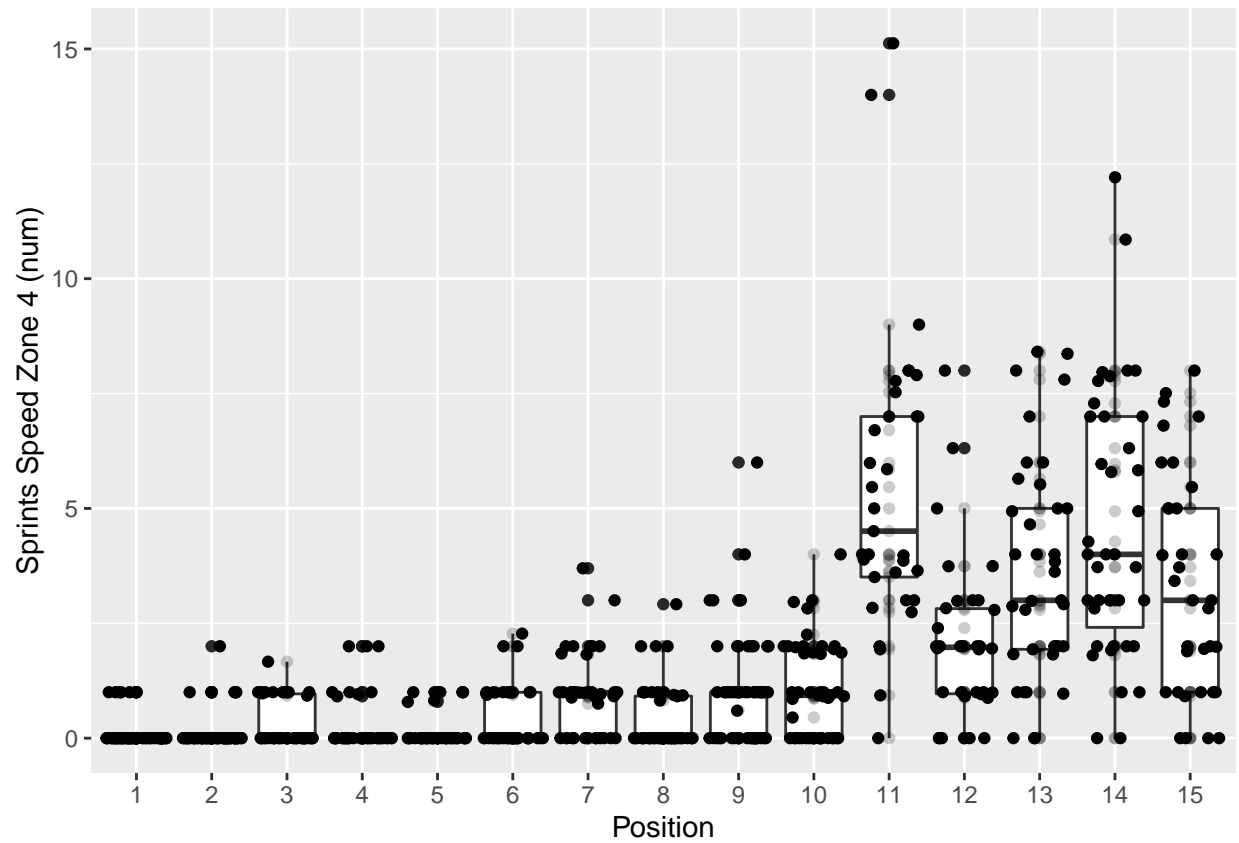




```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

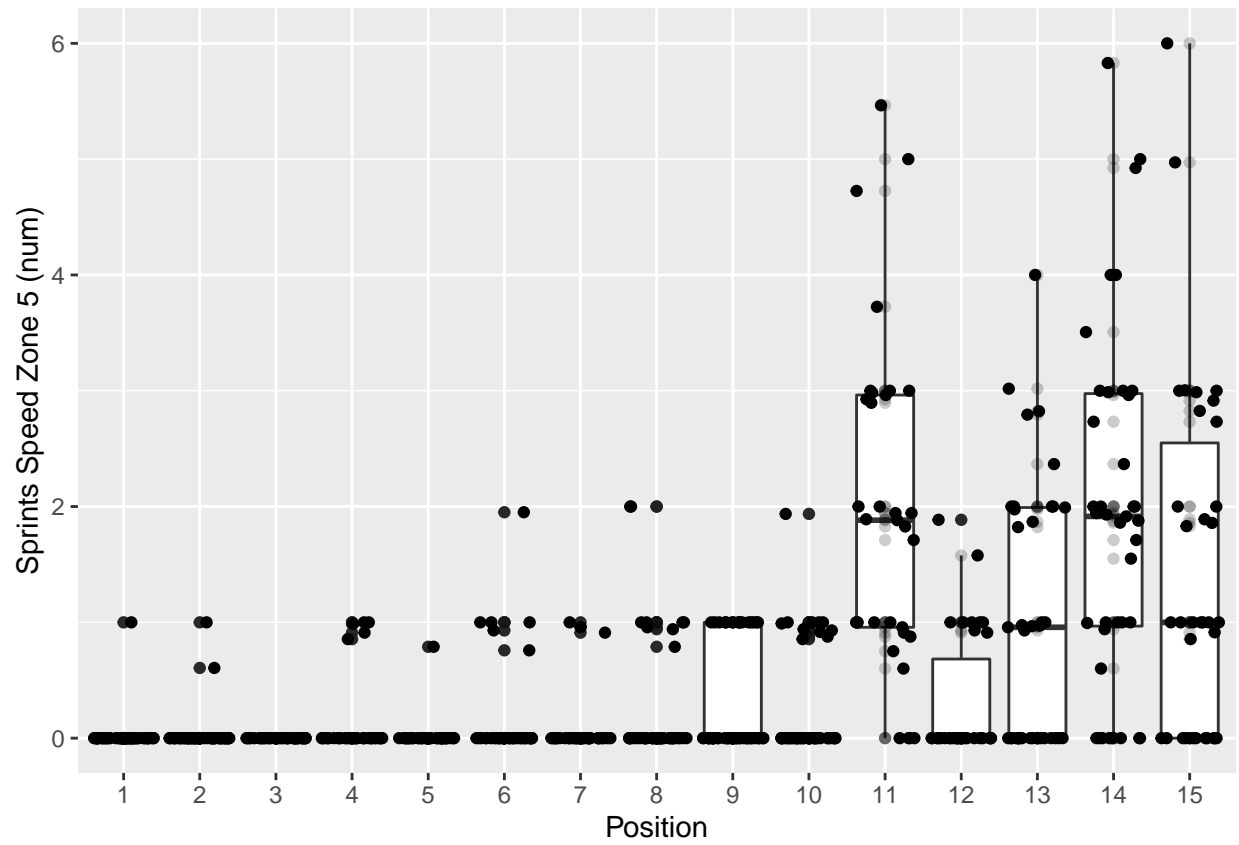
```
## Warning: Removed 1 rows containing missing values (geom_point).
```

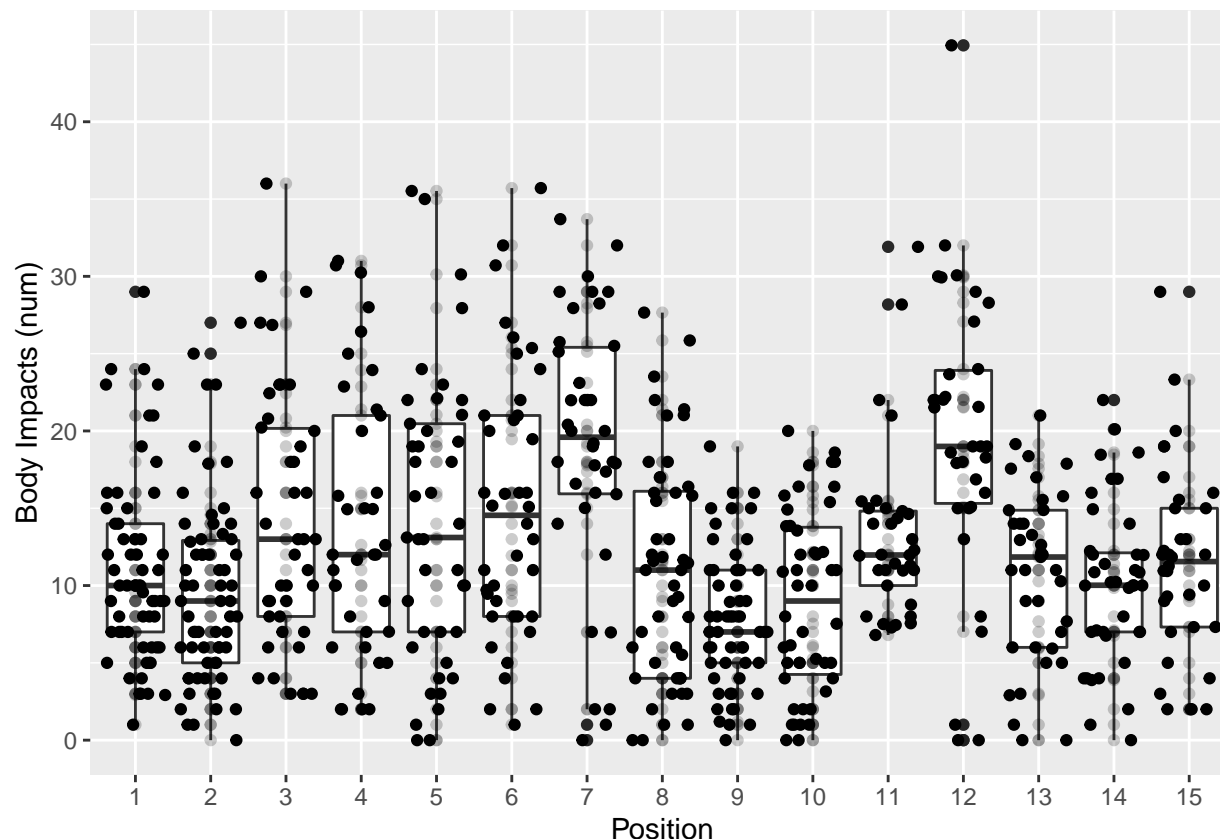


```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```





These plots mostly reinforce what is in the 2018 dataset-unique variable plots, but additionally:

Front rowers (#1, #2, #3) have lower minutes and distance values than the other positions;

For Body Impacts in Body Impacts Zone Total (num), inside centre (#12) has a high distribution centre relative to the other backs. Overall, 12 has the second-highest distribution centre, which also contains the data points with the highest values across all positions.

Regarding the heart rate issues, the distributions are all centred quite similarly across all positions, and for the most part are quite narrow. I am quite comfortable with removing these variables completely, as they do not appear to show any meaningful trend.

```
master2018 <- master2018[, -c(16, 17)]
```

## Finding the positional maximum, minimum and mean data for each marker

The 2018 dataset contains 53 variables, 33 of which are not shared by the 2019 and 2020 datasets. Within the 20 variables that are shared, 7 of them are not performance markers; they are either redundant information, unique identifiers for each individual row, or variables I created during the cleaning process.

Therefore, the 2018 dataset must be separated by position, and then the positional maximum, minimum and mean can be found. These values are printed. Variable 18, HIE Rate is not numeric, so it is left out here. Any performance markers that are shared by the 2018, 2019 and 2020 datasets are then found and printed afterwards.

```
for (r in 1:15) {  
  # Finding the positional minimum, mean and maximum for performance markers exclusive to the 2018 data
```

```

positionalData <- master2018[which(master2018$Position == r),]
cat(paste0("POSITION: ", as.character(r)), "\n")
for (s in c(6:7, 9:11, 13:15, 17:18, 20:23, 32:39, 41:45)) {
  print(paste0("Variable ", as.character(s), " - ", colnames(positionalData)[s],
    " - MIN: ", min(positionalData[, s], na.rm = TRUE),
    " | MEAN: ", mean(positionalData[, s], na.rm = TRUE),
    " | MAX: ", max(positionalData[, s], na.rm = TRUE)))
}
cat("\n")
# Finding the positional minimum, mean and maximum for performance markers shared between the 2018, 2019 and 2020
positionalSmallData <- combinedData[which(combinedData$Position == r),]
for (t in 5:17) {
  print(paste0("Variable ", as.character(t), " - ", colnames(positionalSmallData)[t],
    " - MIN: ", min(positionalSmallData[, t], na.rm = TRUE),
    " | MEAN: ", mean(positionalSmallData[, t], na.rm = TRUE),
    " | MAX: ", max(positionalSmallData[, t], na.rm = TRUE)))
}
cat("\n", "\n", "\n")
}

```

```

## POSITION: 1
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0.263157894736842 | MAX: 5"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 0 | MEAN: 1496.90804871939 | MAX: 3117"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 42 | MEAN: 59.9473684210526 | MAX: 68"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 1.89473684210526 | MAX: 36"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 0 | MEAN: 1718.9375 | MAX: 3091"
## [1] "Variable 13 - Sprints Total (num) - MIN: 9 | MEAN: 71.2747921314135 | MAX: 132"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 0 | MEAN: 0.157894736842105 | MAX: 1"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 0 | MEAN: 41.1876603003805 | MAX: 81"
## [1] "Variable 17 - Athlete Load - MIN: 4 | MEAN: 24.7091013886369 | MAX: 42"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 76 | MEAN: 374.695524554644 | MAX: 773"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 0 | MEAN: 14.9431736916821 | MAX: 30"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 16 | MEAN: 93.0416697798034 | MAX: 168"
## [1] "Variable 23 - HIE Rate - MIN: 0.1 | MEAN: 1.79473684210526 | MAX: 3.1"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 0 | MEAN: 505.382072601906 | MAX: 1301"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 0 | MEAN: 809.511831698492 | MAX: 2402"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 1.24895131765736 | MAX: 4"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 0.421052631578947 | MAX: 1"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.0526315789473684 | MAX: 1"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 0.263157894736842 | MAX: 2"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0.143688159762619 | MAX: 1"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.210526315789474 | MAX: 1"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 1 | MEAN: 7.52212106010311 | MAX: 18"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 1.10526315789474 | MAX: 6"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.105263157894737 | MAX: 1"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 549 | MEAN: 3177.49180327869 | MAX: 7200"
## [1] "Variable 6 - Distance Total (m) - MIN: 369.417967957818 | MEAN: 2911.0083922572 | MAX: 5287"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 14 | MEAN: 24.6590163934426 | MAX: 40"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 10 | MEAN: 57.6785295636835 | MAX: 127"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 367.227742851349 | MEAN: 2767.94466483495 | MAX: 4

```

```

## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 0 | MEAN: 99.0268662515762 | MAX: 212"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 0 | MEAN: 29.2609954340411 | MAX: 101"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 2.81307762130047 | MAX: 57"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 0.229508196721311 | MAX: 6"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 1.23930712949719 | MAX: 7"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 0.114754098360656 | MAX: 1"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 0.0163934426229508 | MAX: 1"
## [1] "Variable 17 - Body Impacts (num) - MIN: 1 | MEAN: 10.9423616152659 | MAX: 29"
##
##
## POSITION: 2
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0.0909090909090909 | MAX: 1"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 0 | MEAN: 1506.91168899195 | MAX: 3509"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 46 | MEAN: 63.4090909090909 | MAX: 94"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 0.454545454545455 | MAX: 4"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 0 | MEAN: 1578.61386801341 | MAX: 3715"
## [1] "Variable 13 - Sprints Total (num) - MIN: 12 | MEAN: 69.9503668342423 | MAX: 168.104330037504"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 0 | MEAN: 9.20362878330435 | MAX: 26.8258671779676"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 0 | MEAN: 38.2528264218704 | MAX: 77.49694962523"
## [1] "Variable 17 - Athlete Load - MIN: 3 | MEAN: 24.0817167039389 | MAX: 47.6133651551313"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 81 | MEAN: 281.981425245348 | MAX: 494"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 2 | MEAN: 14.7973602583021 | MAX: 35.95294919877"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0.136363636363636 | MAX: 1"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 19 | MEAN: 90.0065153062512 | MAX: 219.604500511"
## [1] "Variable 23 - HIE Rate - MIN: 0.9 | MEAN: 1.99545454545455 | MAX: 5.3"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 0 | MEAN: 611.337445800863 | MAX: 1751.26428818843"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 0 | MEAN: 1102.49607094331 | MAX: 2923"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 0.631457077986083 | MAX: 2.9615517838"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 0.181235633088768 | MAX: 1"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.0448719967251315 | MAX: 0.987183927"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 0.544578843826087 | MAX: 2"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0.181235633088768 | MAX: 1"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 0 | MEAN: 8.02079303369265 | MAX: 21"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 1.1301670177218 | MAX: 3.974202544884"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.135077333168025 | MAX: 1"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 308 | MEAN: 3008.87301587302 | MAX: 5700"
## [1] "Variable 6 - Distance Total (m) - MIN: 415 | MEAN: 2848.3633271742 | MAX: 5999.28400954654"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 19.9 | MEAN: 24.5222222222222 | MAX: 33.3"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 10 | MEAN: 54.7252307305186 | MAX: 166.1609273781"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 410 | MEAN: 2704.08308687971 | MAX: 5667.93385612"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 5 | MEAN: 113.216557894861 | MAX: 296.3689055574"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 0 | MEAN: 27.6564561939912 | MAX: 89"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 2.73694296754858 | MAX: 28"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 0.343446620379514 | MAX: 18"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 1.13050032140242 | MAX: 7.94840508976"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 0.126984126984127 | MAX: 2"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 0.0254950716505539 | MAX: 1"
## [1] "Variable 17 - Body Impacts (num) - MIN: 0 | MEAN: 9.51792050426162 | MAX: 27"
##

```

```

##
##
## POSITION: 3
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 1.15384615384615 | MAX: 5"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 517 | MEAN: 1368.89619520265 | MAX: 2286"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 43 | MEAN: 57.8461538461538 | MAX: 68"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 8.58251452870832 | MAX: 34"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 578 | MEAN: 1505.8 | MAX: 2483"
## [1] "Variable 13 - Sprints Total (num) - MIN: 19 | MEAN: 84.5692319421972 | MAX: 139"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 0 | MEAN: 17.2781371464679 | MAX: 39"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 13 | MEAN: 41.6546246290079 | MAX: 83"
## [1] "Variable 17 - Athlete Load - MIN: 5 | MEAN: 32.7836268985947 | MAX: 42"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 128 | MEAN: 354.349193357073 | MAX: 500"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 3 | MEAN: 16.5761088548634 | MAX: 33"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 16 | MEAN: 141.221243830969 | MAX: 228.740861088"
## [1] "Variable 23 - HIE Rate - MIN: 0.7 | MEAN: 1.78461538461538 | MAX: 2.4"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 0 | MEAN: 438.189831916733 | MAX: 1250"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 0 | MEAN: 492.178008825214 | MAX: 1687"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 0.532775104667875 | MAX: 3"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 0.230769230769231 | MAX: 2"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.0769230769230769 | MAX: 1"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 0.29631944010498 | MAX: 1.85215272136"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 3 | MEAN: 15.6633700714761 | MAX: 28"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 2.77171436610197 | MAX: 8"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.384615384615385 | MAX: 1"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0.0769230769230769 | MAX: 1"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 636 | MEAN: 3895.42857142857 | MAX: 5700"
## [1] "Variable 6 - Distance Total (m) - MIN: 726 | MEAN: 3507.06320710674 | MAX: 6014.11945918193"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 20 | MEAN: 25.7285714285714 | MAX: 31.3"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 10 | MEAN: 67.9466777843583 | MAX: 156.0842033202"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 697 | MEAN: 3333.12595895898 | MAX: 5186.87318158"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 21 | MEAN: 122.414436958762 | MAX: 489.714187917"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 0 | MEAN: 43.6848209028421 | MAX: 300.4620913914"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 7.45850542642828 | MAX: 45"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 0.213120298607206 | MAX: 3"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 1.56208131973519 | MAX: 9.75526270751"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 0.299180792612661 | MAX: 1.6639906582"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 17 - Body Impacts (num) - MIN: 3 | MEAN: 14.2693522978664 | MAX: 36"
##
##
## POSITION: 4
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 132 | MEAN: 1612.33176180076 | MAX: 3187"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 38 | MEAN: 57.0769230769231 | MAX: 68"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 92 | MEAN: 1703.99741436061 | MAX: 3001.585407432"
## [1] "Variable 13 - Sprints Total (num) - MIN: 22 | MEAN: 94.7175969370542 | MAX: 147.698602113877"

```

```

## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 0 | MEAN: 0.692307692307692 | MAX: 2"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 4 | MEAN: 44.1791707474111 | MAX: 97.17013296965"
## [1] "Variable 17 - Athlete Load - MIN: 9 | MEAN: 35.2130869787205 | MAX: 46"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 79 | MEAN: 289.910323451374 | MAX: 618"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 5 | MEAN: 17.913493690885 | MAX: 26"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 19 | MEAN: 108.361340525177 | MAX: 179.764745993"
## [1] "Variable 23 - HIE Rate - MIN: 0.8 | MEAN: 1.36153846153846 | MAX: 2.7"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 0 | MEAN: 1023.9982521874 | MAX: 1915.93291404612"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 0 | MEAN: 1329.92781063912 | MAX: 2905"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 0.977065651653179 | MAX: 3"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 0.153846153846154 | MAX: 1"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.0769230769230769 | MAX: 1"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 0.488927351929813 | MAX: 2.3602484472"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0.230769230769231 | MAX: 1"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.153846153846154 | MAX: 1"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 1 | MEAN: 8.66889982773846 | MAX: 11.9496855345"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 1.89925738766055 | MAX: 6"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.290965893098789 | MAX: 1"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 1169 | MEAN: 4723.54054054054 | MAX: 6666"
## [1] "Variable 6 - Distance Total (m) - MIN: 500 | MEAN: 4342.04564313501 | MAX: 5958.77300613497"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 21.3 | MEAN: 26.3216216216216 | MAX: 33.5"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 10 | MEAN: 84.6205743647886 | MAX: 139.5002567174"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 461 | MEAN: 4097.55722739379 | MAX: 5613.07624890"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 27 | MEAN: 179.50165089735 | MAX: 447"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 3 | MEAN: 55.0671750846275 | MAX: 129"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 8.55870017312879 | MAX: 43"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 1.2291368648365 | MAX: 18"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 1.74399401171273 | MAX: 4"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 0.293746641612146 | MAX: 2"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 0.12849440364037 | MAX: 1"
## [1] "Variable 17 - Body Impacts (num) - MIN: 2 | MEAN: 13.8527471541677 | MAX: 31"
##
##
##
## POSITION: 5
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0.0909090909090909 | MAX: 1"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 0 | MEAN: 1432.11707653224 | MAX: 5455.87812589825"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 41 | MEAN: 62.3636363636364 | MAX: 80"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 0.521233092167348 | MAX: 4.733564013"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 0 | MEAN: 1106.84004068974 | MAX: 2748.1978341684"
## [1] "Variable 13 - Sprints Total (num) - MIN: 3 | MEAN: 69.0772313710288 | MAX: 118.845967350897"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 0 | MEAN: 1.85427778100771 | MAX: 18"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 0 | MEAN: 30.9736980513676 | MAX: 70.93906578309"
## [1] "Variable 17 - Athlete Load - MIN: 2 | MEAN: 26.6417086134977 | MAX: 42"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 37 | MEAN: 288.146090222937 | MAX: 495.493767976989"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 0 | MEAN: 13.3756101901928 | MAX: 29.97923875432"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 0 | MEAN: 95.3233527382152 | MAX: 175.9657346048"
## [1] "Variable 23 - HIE Rate - MIN: 0 | MEAN: 1.66363636363636 | MAX: 3.5"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 0 | MEAN: 523.414269120335 | MAX: 1311.91207370293"

```



```

## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 0 | MEAN: 1021.95428402095 | MAX: 4915.20551882725"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 1.31116659253708 | MAX: 3.68514627444"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 0.558407069921314 | MAX: 3.6433365292"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.0828031029373311 | MAX: 0.910834132"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 0.227194658171989 | MAX: 1.5778546712"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0.416168606480025 | MAX: 1.5778546712"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.162629757785467 | MAX: 1"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 0 | MEAN: 10.9258290878338 | MAX: 27.3250239693"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 2.3031863882088 | MAX: 8.197507190795"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.181818181818182 | MAX: 1"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 304 | MEAN: 3581.9756097561 | MAX: 5700"
## [1] "Variable 6 - Distance Total (m) - MIN: 210 | MEAN: 3550.16731952325 | MAX: 6402.64402031168"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 16.1 | MEAN: 25.6073170731707 | MAX: 36.9"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 0 | MEAN: 68.1882284984665 | MAX: 127.13754646840"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 209 | MEAN: 3358.8747919027 | MAX: 6125.179478199"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 0 | MEAN: 140.723754894935 | MAX: 371"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 0 | MEAN: 42.761438695081 | MAX: 133"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 6.71720468069078 | MAX: 33"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 0.760739302894759 | MAX: 29.1903114186"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 1.76402958201665 | MAX: 7.10034602076"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 0.185564103791334 | MAX: 1"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 0.0192421301375644 | MAX: 0.788927335"
## [1] "Variable 17 - Body Impacts (num) - MIN: 0 | MEAN: 14.1804907331458 | MAX: 35.5225311601151"
##
##
##
## POSITION: 6
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 294 | MEAN: 1473.66255818301 | MAX: 2879"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 48 | MEAN: 64.4444444444444 | MAX: 84"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 314.782006920415 | MEAN: 1829.335603442 | MAX: 36"
## [1] "Variable 13 - Sprints Total (num) - MIN: 25 | MEAN: 91.3961297056411 | MAX: 153.528810092056"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 0 | MEAN: 0.66356227669271 | MAX: 3.0335284725918"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 6.31141868512111 | MEAN: 44.8805679134632 | MAX:"
## [1] "Variable 17 - Athlete Load - MIN: 9 | MEAN: 33.756186176748 | MAX: 47.9248238057948"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 141 | MEAN: 372.918113610756 | MAX: 719"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 5 | MEAN: 21.1915960360053 | MAX: 37.84067085953"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 42 | MEAN: 137.233213080121 | MAX: 211.830889873"
## [1] "Variable 23 - HIE Rate - MIN: 1 | MEAN: 1.87222222222222 | MAX: 2.3"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 0 | MEAN: 458.307394860648 | MAX: 1246.75711449371"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 0 | MEAN: 784.373687525423 | MAX: 2407"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 1.96738853007014 | MAX: 11.6604159563"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 0.594931467230508 | MAX: 2.9151039890"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.103704579128447 | MAX: 0.9397024275"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 0.942359310756141 | MAX: 2.9874213836"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0.357227551758362 | MAX: 1"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.0553226182157 | MAX: 0.995807127882"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 0 | MEAN: 12.6735603672789 | MAX: 29.1307752545"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 1.8840364856914 | MAX: 6.577916992952"

```

```

## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.250698409121028 | MAX: 1.5167642362"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 879 | MEAN: 4441.3 | MAX: 7200"
## [1] "Variable 6 - Distance Total (m) - MIN: 983 | MEAN: 4458.99581413471 | MAX: 6696.31949882537"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 21.3 | MEAN: 26.9175 | MAX: 32.9"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 17 | MEAN: 97.092850245423 | MAX: 165.98279285241"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 911 | MEAN: 4140.50214624009 | MAX: 6651.21378230"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 28 | MEAN: 226.743308649874 | MAX: 444.839979462"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 1 | MEAN: 78.9691713807999 | MAX: 175.8283009629"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 11.0123639256839 | MAX: 49.75183980831"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 1.67024057137139 | MAX: 26.33920931028"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 3.35304783480104 | MAX: 11"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 0.429503904216778 | MAX: 2.2751463544"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 0.165993578258851 | MAX: 1.9510525415"
## [1] "Variable 17 - Body Impacts (num) - MIN: 1 | MEAN: 15.0577041292301 | MAX: 35.708692247455"
##
##
##
## POSITION: 7
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0.0769230769230769 | MAX: 1"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 0 | MEAN: 1584.57539654151 | MAX: 3393.93546294795"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 20 | MEAN: 57 | MAX: 72"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 0.442890442890443 | MAX: 5.757575757"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 0 | MEAN: 1916.64418674828 | MAX: 3796.9190854548"
## [1] "Variable 13 - Sprints Total (num) - MIN: 21.4873200822481 | MEAN: 112.359715741115 | MAX: 158.9"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 0 | MEAN: 1.54090137291866 | MAX: 3.69709745419167"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 0 | MEAN: 56.3499533812517 | MAX: 107.2158261715"
## [1] "Variable 17 - Athlete Load - MIN: 23 | MEAN: 39.0020681933792 | MAX: 47.1379925409437"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 270.544893762851 | MEAN: 451.358227278207 | MAX: 738"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 5.86017820424949 | MEAN: 26.7475259904441 | MAX: "
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 35.1610692254969 | MEAN: 174.748306479482 | MAX: "
## [1] "Variable 23 - HIE Rate - MIN: 0.4 | MEAN: 2.02307692307692 | MAX: 3.4"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 0 | MEAN: 685.738426406166 | MAX: 1414"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 0 | MEAN: 1068.29908950838 | MAX: 2311.88166828322"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 3.6244043864559 | MAX: 7.394194908383"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 1.41971703910621 | MAX: 4"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.205781813419659 | MAX: 1"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 1.31699105504861 | MAX: 3.69709745419"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0.744069003887125 | MAX: 1.919191919"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.362779996536119 | MAX: 1"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 0 | MEAN: 14.5414263206331 | MAX: 25"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 3.2717005023577 | MAX: 8"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.371931422758351 | MAX: 2"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 853 | MEAN: 4873.07894736842 | MAX: 5700"
## [1] "Variable 6 - Distance Total (m) - MIN: 816 | MEAN: 4610.9245565816 | MAX: 6769.93521274733"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 20.1 | MEAN: 28.3368421052632 | MAX: 36.9"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 16 | MEAN: 108.800275027653 | MAX: 171.9150316199"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 746 | MEAN: 4154.65556958085 | MAX: 5812.78234985"

```

```

## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 28 | MEAN: 312.182376360463 | MAX: 650.744177902"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 0 | MEAN: 117.896225341909 | MAX: 255.5069164769"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 19.7537256729293 | MAX: 52"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 2.22638388818134 | MAX: 35.5050505050505"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 4.70401858712875 | MAX: 13.8641154532"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 0.965468602683914 | MAX: 3.6970974541"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 0.0755452968233962 | MAX: 1"
## [1] "Variable 17 - Body Impacts (num) - MIN: 0 | MEAN: 19.0128557297344 | MAX: 33.6993243243243"
##
##
##
## POSITION: 8
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 0 | MEAN: 889.540305075034 | MAX: 3918"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 52 | MEAN: 65.3 | MAX: 97"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 0 | MEAN: 997.690320106124 | MAX: 4169"
## [1] "Variable 13 - Sprints Total (num) - MIN: 21 | MEAN: 69.7798746810894 | MAX: 139.924991476304"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 0 | MEAN: 0.824811733787481 | MAX: 4"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 0 | MEAN: 21.4972556668689 | MAX: 58"
## [1] "Variable 17 - Athlete Load - MIN: 5 | MEAN: 28.9737667069005 | MAX: 46.6416638254347"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 146 | MEAN: 332.688424119736 | MAX: 547"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 3 | MEAN: 16.0477421376015 | MAX: 35.95294919877"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 37 | MEAN: 90.2149680502521 | MAX: 207.944084555"
## [1] "Variable 23 - HIE Rate - MIN: 0.8 | MEAN: 1.795 | MAX: 5.1"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 0 | MEAN: 640.49286974983 | MAX: 1716.99624957382"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 0 | MEAN: 583.401906648697 | MAX: 2536"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 1.26264722934843 | MAX: 4"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 0.485183573151499 | MAX: 2"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.178892733564014 | MAX: 1.5778546712"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 0.943264307397904 | MAX: 3"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0.341083413231064 | MAX: 2"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.193338048501848 | MAX: 1"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 0 | MEAN: 7.93308707116915 | MAX: 19.7543767964"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 0.933025806389864 | MAX: 4"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 451 | MEAN: 3977.6875 | MAX: 7200"
## [1] "Variable 6 - Distance Total (m) - MIN: 231 | MEAN: 3741.22888377951 | MAX: 6444.32321854756"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 14.4 | MEAN: 26.5145833333333 | MAX: 33.4"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 1 | MEAN: 62.8898706444858 | MAX: 156.44391408114"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 231 | MEAN: 3506.7368350704 | MAX: 5903.085577906"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 0 | MEAN: 169.720946678763 | MAX: 480.5597964376"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 0 | MEAN: 47.6803982631724 | MAX: 182"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 13.0796696295287 | MAX: 123"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 3.95631283589847 | MAX: 77"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 1.96472229199339 | MAX: 6"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 0.385741928707293 | MAX: 2.9151039890"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 0.222648174269814 | MAX: 2"
## [1] "Variable 17 - Body Impacts (num) - MIN: 0 | MEAN: 10.9464325546025 | MAX: 27.6555189741813"
##

```

```

##
##
## POSITION: 9
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 297 | MEAN: 1412.611111111111 | MAX: 3280"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 59 | MEAN: 78.94444444444444 | MAX: 95"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 0.1111111111111111 | MAX: 2"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 457 | MEAN: 2047.111111111111 | MAX: 4661"
## [1] "Variable 13 - Sprints Total (num) - MIN: 37 | MEAN: 109.1666666666667 | MAX: 172"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 0 | MEAN: 14.11111111111111 | MAX: 69"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 13 | MEAN: 58.44444444444444 | MAX: 115"
## [1] "Variable 17 - Athlete Load - MIN: 10 | MEAN: 32.55555555555556 | MAX: 51"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 190 | MEAN: 404 | MAX: 814"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 7 | MEAN: 30.27777777777778 | MAX: 63"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 45 | MEAN: 166.8888888888889 | MAX: 260"
## [1] "Variable 23 - HIE Rate - MIN: 2.1 | MEAN: 2.816666666666667 | MAX: 4.5"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 228 | MEAN: 841.7777777777778 | MAX: 1426"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 207 | MEAN: 1149.666666666667 | MAX: 2378"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 2.277777777777778 | MAX: 7"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 0.7222222222222222 | MAX: 3"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.1111111111111111 | MAX: 1"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 1.388888888888889 | MAX: 6"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0.5555555555555556 | MAX: 2"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.1666666666666667 | MAX: 1"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 1 | MEAN: 6.888888888888889 | MAX: 14"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 1.333333333333333 | MAX: 5"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.0555555555555556 | MAX: 1"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 628 | MEAN: 3464.666666666667 | MAX: 5700"
## [1] "Variable 6 - Distance Total (m) - MIN: 0 | MEAN: 3995.15648250463 | MAX: 7576.81715182151"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 0 | MEAN: 28.5074074074074 | MAX: 34.1"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 2 | MEAN: 90.7387485675077 | MAX: 184.80041833711"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 496 | MEAN: 3616.1566087964 | MAX: 6468.014641798"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 25 | MEAN: 318.955913622277 | MAX: 778.943698797"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 9 | MEAN: 97.4520600807118 | MAX: 296.0780895938"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 31.3613092946132 | MAX: 116"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 6.31672768505906 | MAX: 33"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 3.64778574758242 | MAX: 11"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 1.03014475894283 | MAX: 6"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 0.264150943396226 | MAX: 1"
## [1] "Variable 17 - Body Impacts (num) - MIN: 0 | MEAN: 7.76180185507276 | MAX: 19"
##
##
##
## POSITION: 10
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 4.50533526544592 | MEAN: 899.70106970556 | MAX: 24"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 58 | MEAN: 71.9230769230769 | MAX: 83"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 12.0142273745225 | MEAN: 1283.95058566739 | MAX: 3"
## [1] "Variable 13 - Sprints Total (num) - MIN: 37 | MEAN: 128.293951400784 | MAX: 192.898946547015"

```

```

## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 0 | MEAN: 1.55861941968264 | MAX: 5"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 0.750889210907654 | MEAN: 34.2961414506259 | MAX: 5"
## [1] "Variable 17 - Athlete Load - MIN: 11 | MEAN: 42.4746188327812 | MAX: 59.9297698010144"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 125.067087608524 | MEAN: 292.344399539444 | MAX: 626.4"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 8 | MEAN: 33.2907695452161 | MAX: 54.31135388216"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 58 | MEAN: 155.333606300526 | MAX: 250.019508388"
## [1] "Variable 23 - HIE Rate - MIN: 0.9 | MEAN: 1.80769230769231 | MAX: 3.1"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 58 | MEAN: 956.191160008163 | MAX: 1707.99843932891"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 423.501514951917 | MEAN: 1500.07369801408 | MAX: 21"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 1.63143543758393 | MAX: 4.66677583101"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 0.335458394042468 | MAX: 1.5017784218"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 1.45456857103811 | MAX: 5"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 1.20794938493392 | MAX: 5.61841591884"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.566545641728341 | MAX: 1.8216682646"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 0 | MEAN: 7.15809100444045 | MAX: 12.7516778523"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 2.06874995664175 | MAX: 7"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.0769230769230769 | MAX: 1"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 213 | MEAN: 4056.73913043478 | MAX: 7200"
## [1] "Variable 6 - Distance Total (m) - MIN: 273 | MEAN: 4503.50740867205 | MAX: 7898.55637924307"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 20.8 | MEAN: 28.3673913043478 | MAX: 33.6"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 3 | MEAN: 96.6657486910692 | MAX: 194.61525394938"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 216 | MEAN: 3926.13425719632 | MAX: 7217.79165040"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 18 | MEAN: 393.484645601956 | MAX: 1115.73621772"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 0 | MEAN: 116.879346673234 | MAX: 321.1967899511"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 30.4027319391951 | MAX: 117.5359556402"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 4.33970533067534 | MAX: 43.57057924239"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 5.04406029879308 | MAX: 12.8400623808"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 1.04852561630707 | MAX: 4"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 0.248843033177613 | MAX: 1.9364701885"
## [1] "Variable 17 - Body Impacts (num) - MIN: 0 | MEAN: 8.9312498478014 | MAX: 20"
##
##
##
## POSITION: 11
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0.272727272727273 | MAX: 1"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 419.747068897378 | MEAN: 1817.11961917282 | MAX: 5"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 56 | MEAN: 63.7272727272727 | MAX: 69"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 2.69692178093397 | MAX: 11"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 396 | MEAN: 2010.74063549825 | MAX: 6055.64268666"
## [1] "Variable 13 - Sprints Total (num) - MIN: 79.5942563562113 | MEAN: 100.568704618029 | MAX: 121.9"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 4.65610194412678 | MEAN: 9.93796711402402 | MAX: 19"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 9.10834132310642 | MEAN: 39.684026982983 | MAX: 1"
## [1] "Variable 17 - Athlete Load - MIN: 36 | MEAN: 40.3408997237654 | MAX: 45.3656110097828"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 170 | MEAN: 487.236129202779 | MAX: 662"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 25 | MEAN: 37.2055359640892 | MAX: 56.7070137622"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 102.120932683441 | MEAN: 162.058839911418 | MAX: 1"
## [1] "Variable 23 - HIE Rate - MIN: 1.1 | MEAN: 1.75454545454545 | MAX: 2.4"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 324.384139112106 | MEAN: 515.661935762434 | MAX: 74"

```

```

## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 144 | MEAN: 1322.49251647698 | MAX: 4324.07091714968"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 2.45479167060679 | MAX: 5.67070137622"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 1.21195466768836 | MAX: 2.91510398908"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.612616187501201 | MAX: 2.7936611664"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0.750889210907654 | MEAN: 1.65749305641408 | MAX: 2.7936611664"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 1.30669968382611 | MAX: 2.83535068811"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.612026210291844 | MAX: 2"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 4.55417066155321 | MEAN: 9.01007395610117 | MAX: 2.7936611664"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 1.68672332525741 | MAX: 4"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.358491150853919 | MAX: 1.9434026593"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0.0909090909090909 | MAX: 1"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 1475 | MEAN: 5376.09090909091 | MAX: 5700"
## [1] "Variable 6 - Distance Total (m) - MIN: 1184 | MEAN: 5464.7692499405 | MAX: 6561.94660918587"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 29.5 | MEAN: 32.7060606060606 | MAX: 35.9"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 15 | MEAN: 97.1295130951609 | MAX: 137.9870668214"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 1090 | MEAN: 4756.42620236099 | MAX: 5658.2168428"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 54 | MEAN: 316.314840938528 | MAX: 454.601226993"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 20 | MEAN: 218.250849173772 | MAX: 370.418730301"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 14 | MEAN: 129.840356963518 | MAX: 243.961185236"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 4 | MEAN: 42.5090043310638 | MAX: 123"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 1 | MEAN: 8.550401164549 | MAX: 16.761966998856"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 5.41712679879978 | MAX: 15.1218703365"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 1.93877634225981 | MAX: 5.46500479386"
## [1] "Variable 17 - Body Impacts (num) - MIN: 6.80190930787589 | MEAN: 13.1843516107464 | MAX: 31.908"
##
##
##
## POSITION: 12
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0.0833333333333333 | MAX: 1"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 53 | MEAN: 2424.02367233332 | MAX: 4102.9440952696"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 65 | MEAN: 70.5 | MAX: 82"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 0.916666666666667 | MAX: 7"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 119 | MEAN: 3076.24369708262 | MAX: 5066.66666666667"
## [1] "Variable 13 - Sprints Total (num) - MIN: 21 | MEAN: 134.614217396678 | MAX: 185.383615084525"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 1.95607412491421 | MEAN: 4.16942814786824 | MAX: 9"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 5 | MEAN: 80.1498888124403 | MAX: 132.9308633807"
## [1] "Variable 17 - Athlete Load - MIN: 6 | MEAN: 42.6540822300487 | MAX: 58.9856957087126"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 320.304498269896 | MEAN: 445.866403387882 | MAX: 603"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 9 | MEAN: 43.591904920599 | MAX: 61.794538361508"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 34 | MEAN: 200.291184725746 | MAX: 288.374512353"
## [1] "Variable 23 - HIE Rate - MIN: 2 | MEAN: 2.30833333333333 | MAX: 2.7"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 106 | MEAN: 913.222646458276 | MAX: 1754.5903771131"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 46 | MEAN: 1821.56538252425 | MAX: 2705"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 3.61870192392369 | MAX: 8.41811617984"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 1.23375656840755 | MAX: 3"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.320727462239 | MAX: 1.8706924844109"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 3.08907819235904 | MAX: 8.42652795838"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 2.04815655571944 | MAX: 7"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.576115275388546 | MAX: 3"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 1 | MEAN: 17.0018749666007 | MAX: 35.5786736020"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 4.22140685905019 | MAX: 10"

```

```

## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.615558110339676 | MAX: 2.80884265279"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0.0833333333333333 | MAX: 1"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 592 | MEAN: 4908.32352941176 | MAX: 7200"
## [1] "Variable 6 - Distance Total (m) - MIN: 178 | MEAN: 5385.2341447149 | MAX: 7751.46944083225"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 19.6 | MEAN: 29.9558823529412 | MAX: 35"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 5 | MEAN: 115.775807283737 | MAX: 180.47236008899"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 168 | MEAN: 4707.2533908691 | MAX: 6951.885565669"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 10 | MEAN: 422.791948280264 | MAX: 976.501797022"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 0 | MEAN: 193.380877874623 | MAX: 363.8712989902"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 52.1026386086186 | MAX: 149.8961937716"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 9.41215679251559 | MAX: 52"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 9.0312020275368 | MAX: 18.72561768530"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 2.04053573218407 | MAX: 8"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 0.30308331244429 | MAX: 1.88554416142"
## [1] "Variable 17 - Body Impacts (num) - MIN: 0 | MEAN: 19.6189815932908 | MAX: 44.9414824447334"
##
##
##
## POSITION: 13
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0.416666666666667 | MAX: 3"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 646 | MEAN: 1456.83067046643 | MAX: 2810"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 60 | MEAN: 71.6666666666667 | MAX: 82"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 2.82785467128028 | MAX: 18.934256055"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 835 | MEAN: 1957.39709313467 | MAX: 3697"
## [1] "Variable 13 - Sprints Total (num) - MIN: 39 | MEAN: 124.379624697539 | MAX: 162.59571706684"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 3 | MEAN: 6.28042386002171 | MAX: 11"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 22 | MEAN: 47.4411032592424 | MAX: 86"
## [1] "Variable 17 - Athlete Load - MIN: 12 | MEAN: 41.5980508679472 | MAX: 56.0674886437378"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 156.73271330368 | MEAN: 397.985816627163 | MAX: 696"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 16 | MEAN: 38.0590592756435 | MAX: 48.5918234912"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 86 | MEAN: 166.661774355969 | MAX: 222"
## [1] "Variable 23 - HIE Rate - MIN: 1.1 | MEAN: 2.08333333333333 | MAX: 4.3"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 0 | MEAN: 857.030758244707 | MAX: 1285"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 0 | MEAN: 1133.84523816302 | MAX: 2247"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 1.52036900292997 | MAX: 3.94463667820"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 0.713102079156155 | MAX: 2"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.14361545664187 | MAX: 0.93445814406"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 2.76811023360661 | MAX: 5"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0.80310115852986 | MAX: 4"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.212752673146372 | MAX: 0.9958071278"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 1 | MEAN: 7.96157888011266 | MAX: 12"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 2.19320962096787 | MAX: 5.97484276729"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0.0778715120051914 | MAX: 0.934458144"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 450 | MEAN: 4668.48648648649 | MAX: 7200"
## [1] "Variable 6 - Distance Total (m) - MIN: 428 | MEAN: 5212.95383632044 | MAX: 7691.52498377677"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 24.8 | MEAN: 30.8189189189189 | MAX: 36.4"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 6 | MEAN: 105.33180767057 | MAX: 169"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 386 | MEAN: 4572.8870982046 | MAX: 6945.827384815"

```

```

## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 12 | MEAN: 348.126243208506 | MAX: 579"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 15 | MEAN: 183.020218063916 | MAX: 331.849315068"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 85.8047695342868 | MAX: 227"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 20.8766847149781 | MAX: 75.73702422145"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 6.67697612717188 | MAX: 14"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 3.61955232741941 | MAX: 8.41012329656"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 1.01247419038426 | MAX: 4"
## [1] "Variable 17 - Body Impacts (num) - MIN: 0 | MEAN: 10.8009858301121 | MAX: 21"
##
##
##
## POSITION: 14
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 0.0588235294117647 | MAX: 1"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 224 | MEAN: 2186.13152690256 | MAX: 4492"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 50 | MEAN: 66.5882352941177 | MAX: 124"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 1.34900733740742 | MAX: 7.9664570230"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 408 | MEAN: 2305.06326763412 | MAX: 4312.42225293"
## [1] "Variable 13 - Sprints Total (num) - MIN: 14 | MEAN: 89.9116574531769 | MAX: 154"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 0 | MEAN: 8.62238498211491 | MAX: 14.0845070422535"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 7 | MEAN: 51.1318065642845 | MAX: 95.59748427672"
## [1] "Variable 17 - Athlete Load - MIN: 5 | MEAN: 33.0358899297577 | MAX: 52.5821596244131"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 185 | MEAN: 433.389187857743 | MAX: 843"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 5 | MEAN: 33.7539284501242 | MAX: 59"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 27 | MEAN: 135.107606749586 | MAX: 251"
## [1] "Variable 23 - HIE Rate - MIN: 1.1 | MEAN: 2.02941176470588 | MAX: 4"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 0 | MEAN: 678.278056047185 | MAX: 1765"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 0 | MEAN: 1488.8932231342 | MAX: 3169"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 3.09724453838742 | MAX: 10.0191754554"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 1.15997791688003 | MAX: 4"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.233382850445833 | MAX: 1"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0 | MEAN: 1.62539164555973 | MAX: 4"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0.677204117899064 | MAX: 3"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.72133579089587 | MAX: 4"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 0 | MEAN: 6.45894394478018 | MAX: 15.0234741784"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 1.24657747347948 | MAX: 3"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.0455934345454255 | MAX: 0.775088387"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0.0455934345454255 | MAX: 0.775088387"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 451 | MEAN: 4907.61538461538 | MAX: 7200"
## [1] "Variable 6 - Distance Total (m) - MIN: 781 | MEAN: 4965.82327944301 | MAX: 6764.98465734743"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 25.6 | MEAN: 32.0948717948718 | MAX: 40"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 17 | MEAN: 90.1958394759506 | MAX: 168"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 557 | MEAN: 4299.39076202336 | MAX: 5815.02347417"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 42 | MEAN: 316.718956032795 | MAX: 541"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 32 | MEAN: 203.733436102847 | MAX: 344.953972042"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 3 | MEAN: 108.967245407316 | MAX: 229.4574982722"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 35.8327151626073 | MAX: 95"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 1 | MEAN: 7.89462136610802 | MAX: 18"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 4.56572078243502 | MAX: 12.2065727699"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 1.91373337582454 | MAX: 5.83020797817"
## [1] "Variable 17 - Body Impacts (num) - MIN: 0 | MEAN: 10.1103959084512 | MAX: 22"
##

```



```

##
##
## POSITION: 15
## [1] "Variable 6 - Duration Speed Hi-Inten (s) - MIN: 0 | MEAN: 1.08333333333333 | MAX: 6"
## [1] "Variable 7 - Duration HR Hi-Inten (s) - MIN: 56 | MEAN: 1798.33879026649 | MAX: 4966"
## [1] "Variable 9 - Distance Rate (m/min) - MIN: 65 | MEAN: 72.9166666666667 | MAX: 81"
## [1] "Variable 10 - Distance Speed Hi-Inten (m) - MIN: 0 | MEAN: 10.0212859480131 | MAX: 50.345270595"
## [1] "Variable 11 - Distance HR Hi-Inten (m) - MIN: 428 | MEAN: 2583.51487985049 | MAX: 6345"
## [1] "Variable 13 - Sprints Total (num) - MIN: 44 | MEAN: 135.231494217563 | MAX: 185.594953972042"
## [1] "Variable 14 - Sprints Hi-Inten (num) - MIN: 3 | MEAN: 6.91875588733108 | MAX: 11"
## [1] "Variable 15 - Sprints HR Hi-Inten (num) - MIN: 3 | MEAN: 57.9865083863352 | MAX: 144"
## [1] "Variable 17 - Athlete Load - MIN: 16 | MEAN: 44.1215164444487 | MAX: 59.3406593406593"
## [1] "Variable 18 - Metabolic PowerPeak - MIN: 203 | MEAN: 414.637442715333 | MAX: 596"
## [1] "Variable 20 - Hi Int Deceleration (num) - MIN: 15 | MEAN: 38.5372429921935 | MAX: 56.3586771224"
## [1] "Variable 21 - Impact Rate (imp/min) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 22 - Hi Intensity Effort (num) - MIN: 90 | MEAN: 186.292985935194 | MAX: 279.849982952"
## [1] "Variable 23 - HIE Rate - MIN: 1.5 | MEAN: 2.11666666666667 | MAX: 3"
## [1] "Variable 32 - Duration HR Zone 4 (s) - MIN: 66 | MEAN: 531.839967367149 | MAX: 1180.21978021978"
## [1] "Variable 33 - Duration HR Zone 5 (s) - MIN: 46 | MEAN: 1124.91449830244 | MAX: 2595.07311289685"
## [1] "Variable 34 - Accelerations Zone 3 (num) - MIN: 0 | MEAN: 2.22459603963097 | MAX: 5.46500479386"
## [1] "Variable 35 - Accelerations Zone 4 (num) - MIN: 0 | MEAN: 0.475728179217081 | MAX: 1.8216682646"
## [1] "Variable 36 - Accelerations Zone 5 (num) - MIN: 0 | MEAN: 0.22204214735281 | MAX: 0.97170132969"
## [1] "Variable 37 - Decelerations Zone 3 (num) - MIN: 0.910834132310642 | MEAN: 2.49098925207717 | MAX: 2.8257456828"
## [1] "Variable 38 - Decelerations Zone 4 (num) - MIN: 0 | MEAN: 0.966279252573286 | MAX: 2.8257456828"
## [1] "Variable 39 - Decelerations Zone 5 (num) - MIN: 0 | MEAN: 0.360287223869716 | MAX: 1.8216682646"
## [1] "Variable 41 - Body Impacts Grade 1 (num) - MIN: 3 | MEAN: 7.65936831438312 | MAX: 11"
## [1] "Variable 42 - Body Impacts Grade 2 (num) - MIN: 0 | MEAN: 1.79248491595686 | MAX: 4.85850664848"
## [1] "Variable 43 - Body Impacts Grade 3 (num) - MIN: 0 | MEAN: 0.145907434242304 | MAX: 1"
## [1] "Variable 44 - Body Impacts Grade 4 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
## [1] "Variable 45 - Body Impacts Grade 5 (num) - MIN: 0 | MEAN: 0 | MAX: 0"
##
## [1] "Variable 5 - Duration Total (s) - MIN: 447 | MEAN: 4850 | MAX: 7200"
## [1] "Variable 6 - Distance Total (m) - MIN: 529 | MEAN: 5528.17954158002 | MAX: 8128.7284144427"
## [1] "Variable 7 - Speed Max (km/h) - MIN: 23.1 | MEAN: 31.7676470588235 | MAX: 40"
## [1] "Variable 8 - Hi Int Acceleration (num) - MIN: 9 | MEAN: 101.156079178673 | MAX: 188.38304552590"
## [1] "Variable 9 - Distance Speed Zone 1 (m) - MIN: 486 | MEAN: 4875.81169803507 | MAX: 7359.18367346"
## [1] "Variable 10 - Distance Speed Zone 2 (m) - MIN: 27 | MEAN: 348.39442953005 | MAX: 605"
## [1] "Variable 11 - Distance Speed Zone 3 (m) - MIN: 10 | MEAN: 182.925720360573 | MAX: 347"
## [1] "Variable 12 - Distance Speed Zone 4 (m) - MIN: 0 | MEAN: 85.6605532734465 | MAX: 205"
## [1] "Variable 13 - Distance Speed Zone 5 (m) - MIN: 0 | MEAN: 31.8691395668902 | MAX: 112.3691556175"
## [1] "Variable 14 - Sprints Speed Zone 3 (num) - MIN: 0 | MEAN: 7.47398448247788 | MAX: 16.0125588697"
## [1] "Variable 15 - Sprints Speed Zone 4 (num) - MIN: 0 | MEAN: 3.31716237427918 | MAX: 8"
## [1] "Variable 16 - Sprints Speed Zone 5 (num) - MIN: 0 | MEAN: 1.46428239293726 | MAX: 6"
## [1] "Variable 17 - Body Impacts (num) - MIN: 2 | MEAN: 11.3959025874786 | MAX: 29"
##
##
##

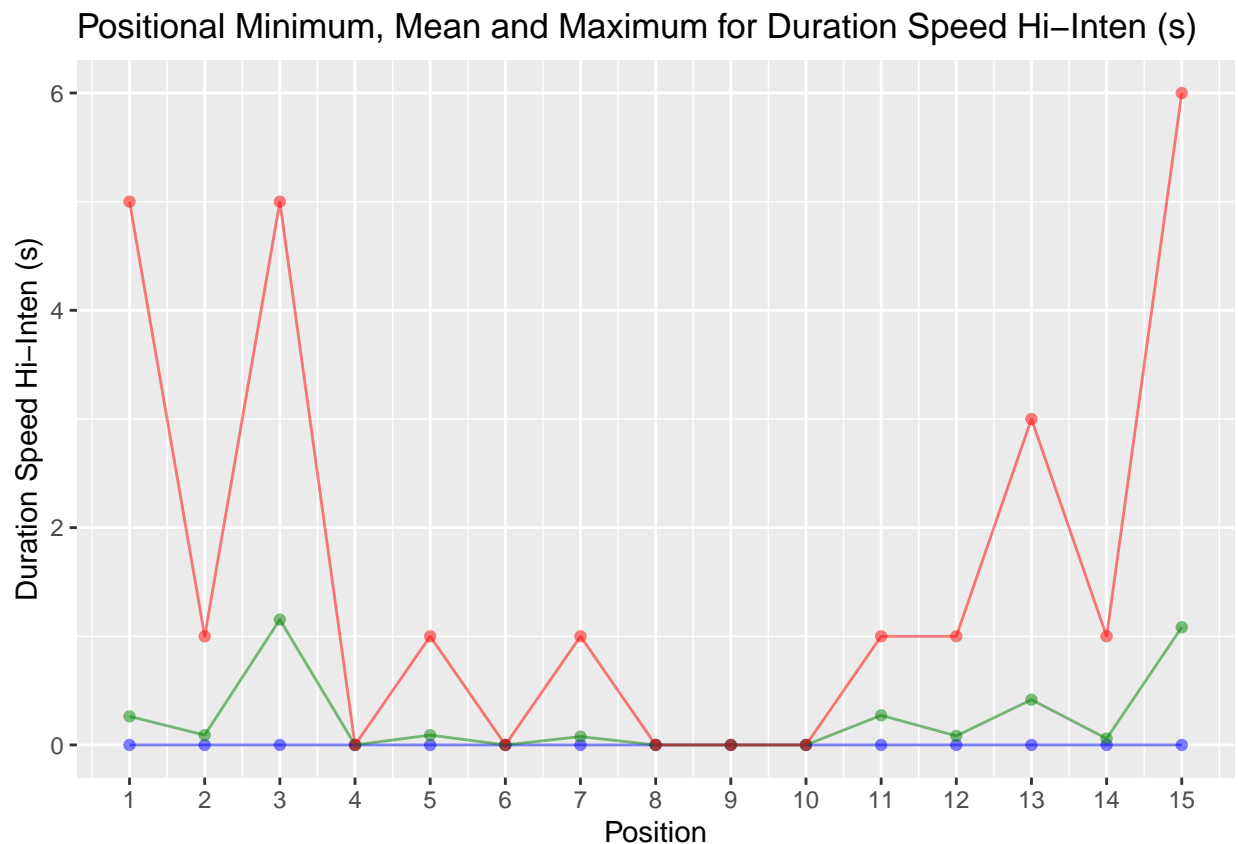
```

## Plotting the positional values of interest

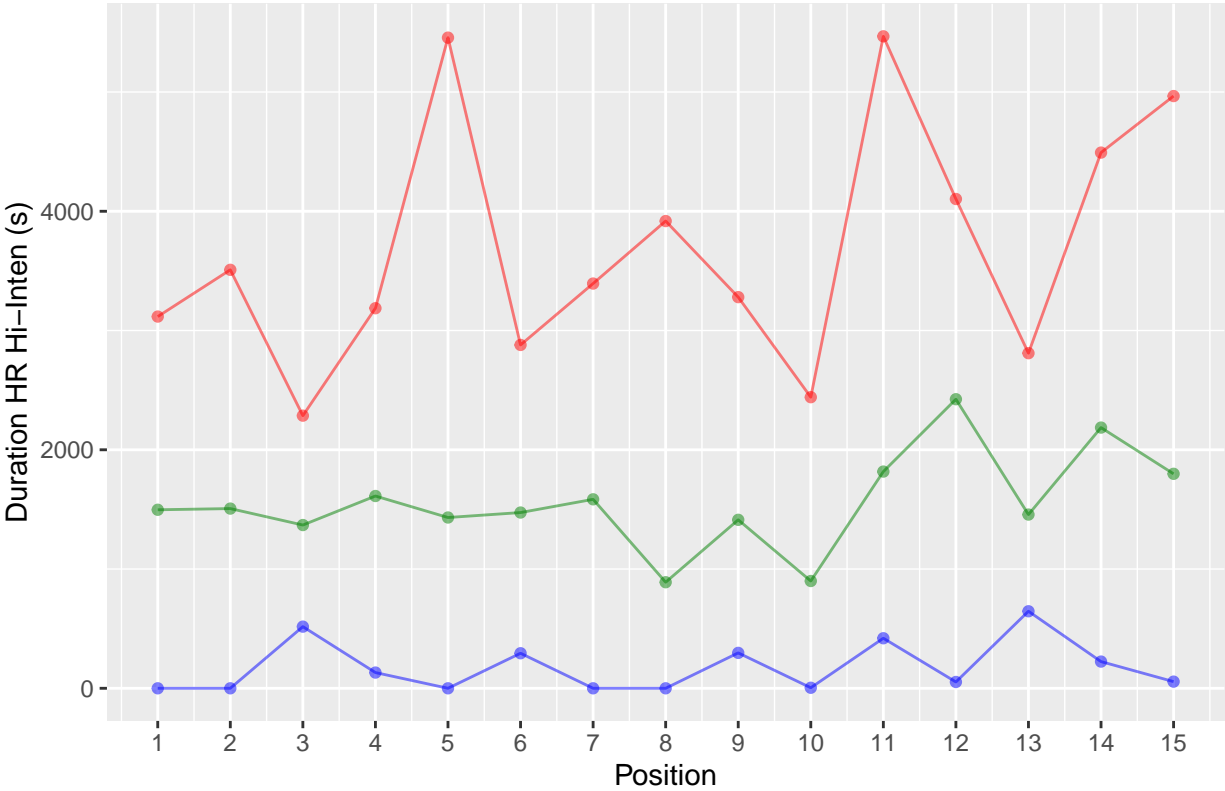
Here, each variable is plotted against position, but only the minimum, mean and maximum values are plotted. Lines are shown to show differences in values visually. A blue line with blue points indicates the minimum value for the variable across each position. A dark green line with dark green points indicates the mean value for the variable across each position. A red line with red points indicates the maximum value for the variable

across each position.

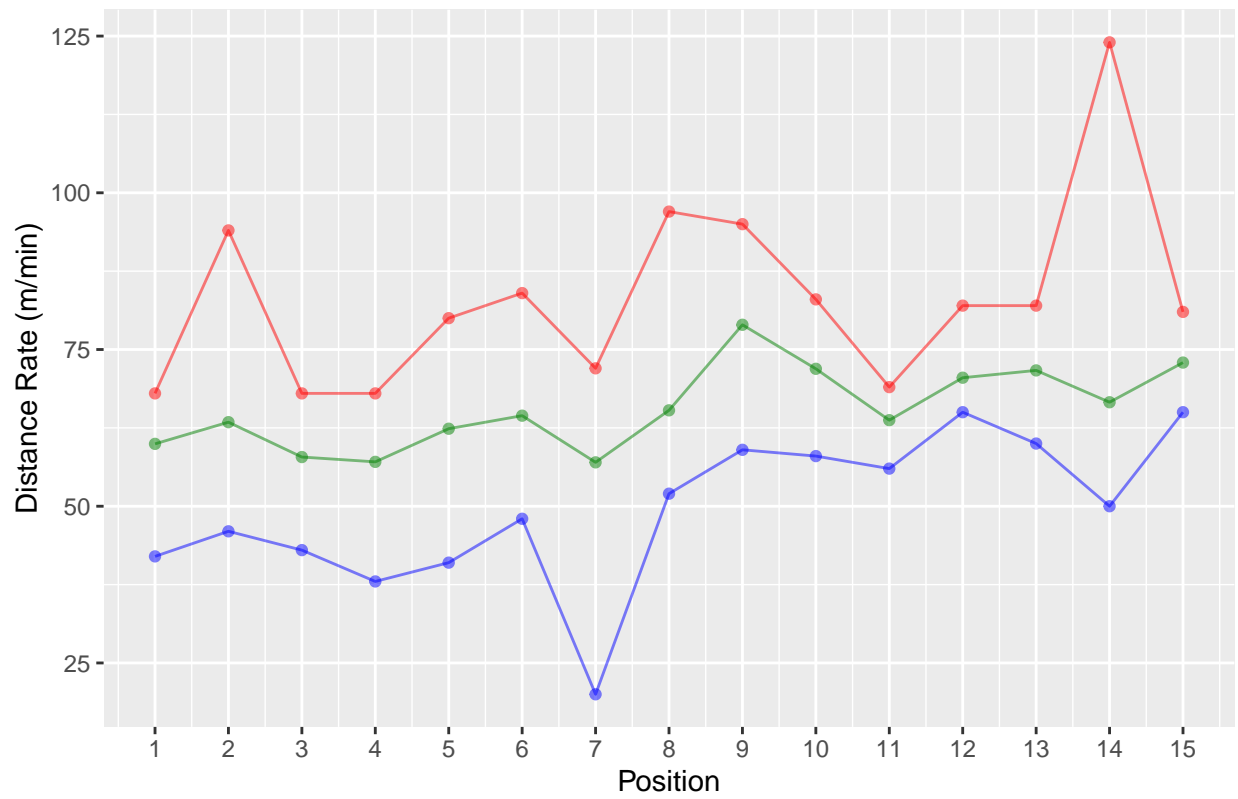
```
# Plotting positional minimum, mean and maximum for the 2018-exclusive variables
for (var in c(6:7, 9:11, 13:15, 17:18, 20:23, 32:39, 41:45)) {
  valuesOfInterest <- data.frame(Position = sort(unique(master2018$Position)), Minimum = 0, Mean = 0, Maximum = 0)
  for (pos in 1:15) {
    positionalVector <- master2018[which(master2018$Position == pos), var]
    valuesOfInterest[pos, 2] <- min(positionalVector, na.rm = TRUE)
    valuesOfInterest[pos, 3] <- mean(positionalVector, na.rm = TRUE)
    valuesOfInterest[pos, 4] <- max(positionalVector, na.rm = TRUE)
  }
  print(ggplot(valuesOfInterest, aes(x = as.numeric(Position))) +
    geom_point(aes(y = Minimum), col = "blue", alpha = 0.5) +
    geom_point(aes(y = Mean), col = "#008000", alpha = 0.5) +
    geom_point(aes(y = Maximum), col = "red", alpha = 0.5) +
    geom_line(aes(y = Minimum), col = "blue", alpha = 0.5) +
    geom_line(aes(y = Mean), col = "#008000", alpha = 0.5) +
    geom_line(aes(y = Maximum), col = "red", alpha = 0.5) +
    scale_x_continuous(breaks = 2:16, labels = as.character(1:15)) +
    xlab("Position") +
    ylab(colnames(master2018)[var]) +
    ggtitle(paste0("Positional Minimum, Mean and Maximum for ", colnames(master2018)[var]))
  )
}
```



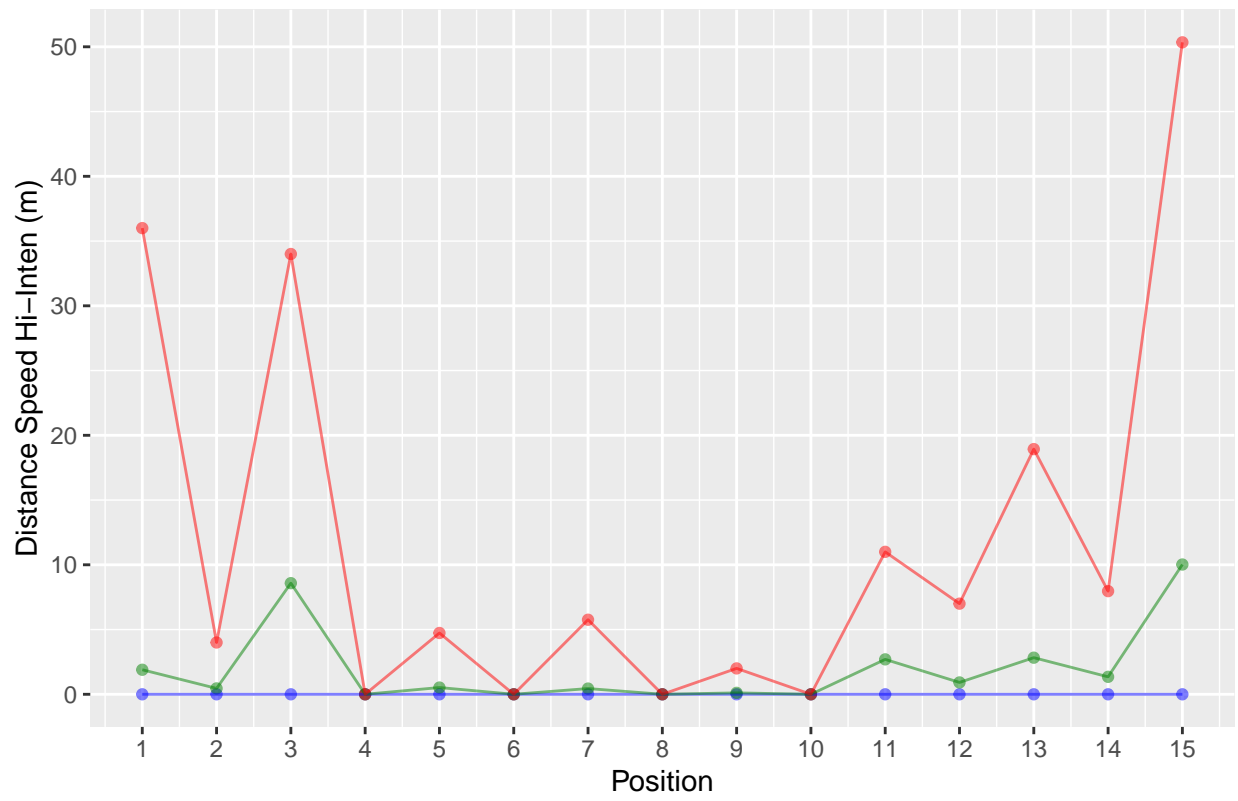
Positional Minimum, Mean and Maximum for Duration HR Hi-Inten (s)



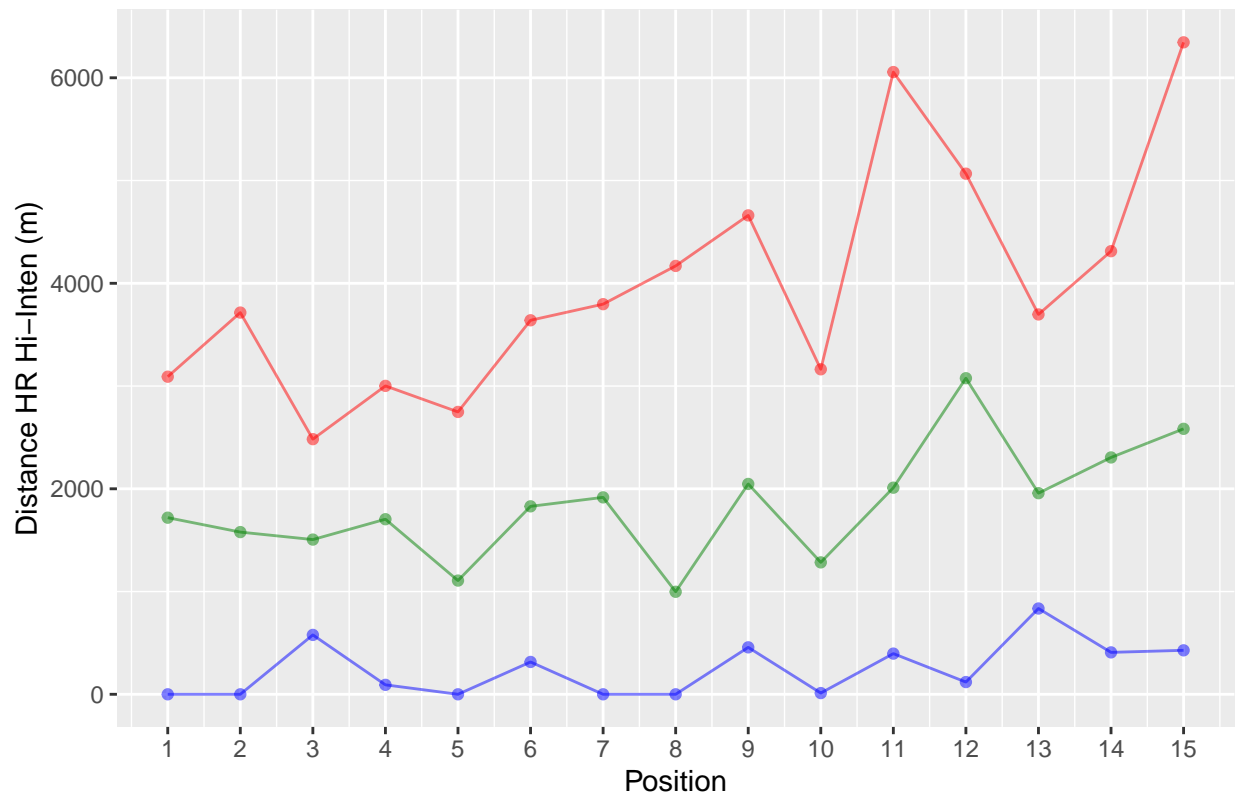
Positional Minimum, Mean and Maximum for Distance Rate (m/min)

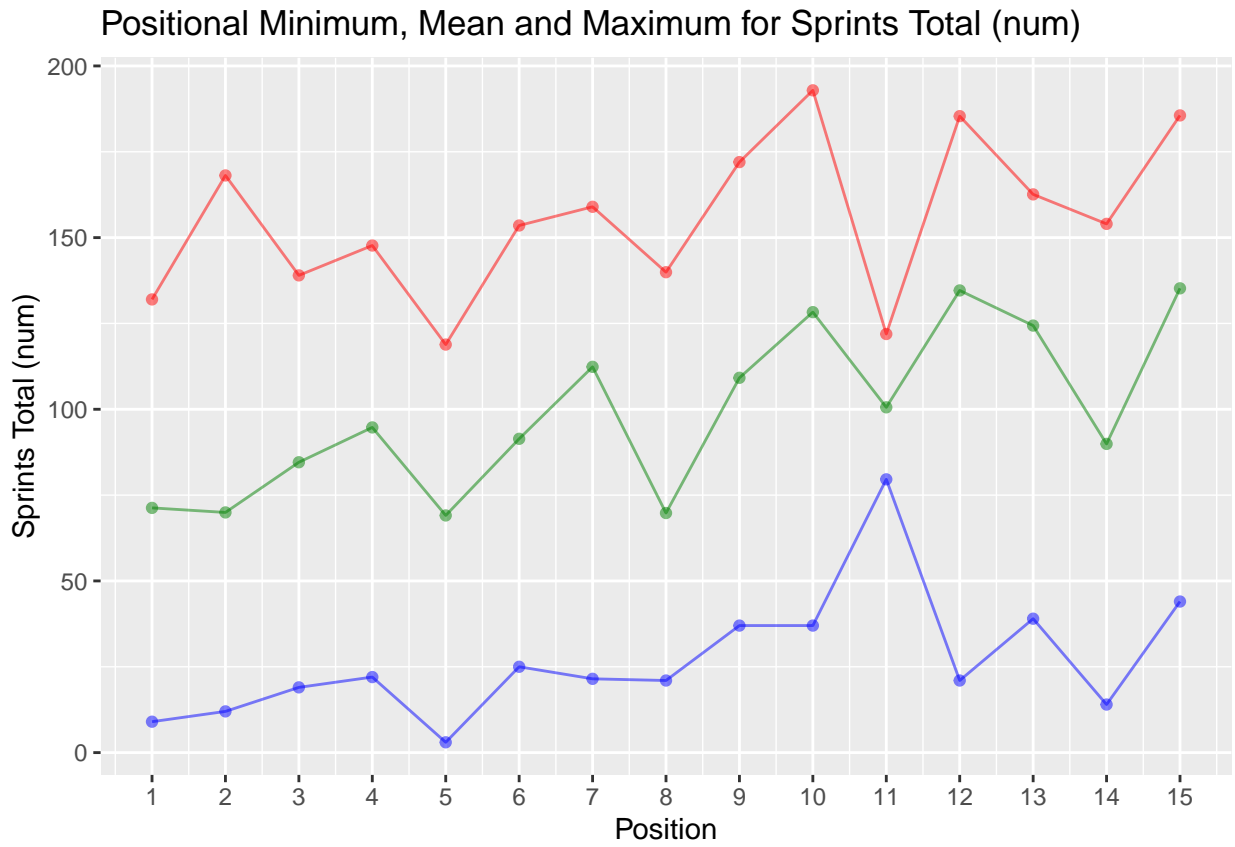


Positional Minimum, Mean and Maximum for Distance Speed Hi-Inten (m)

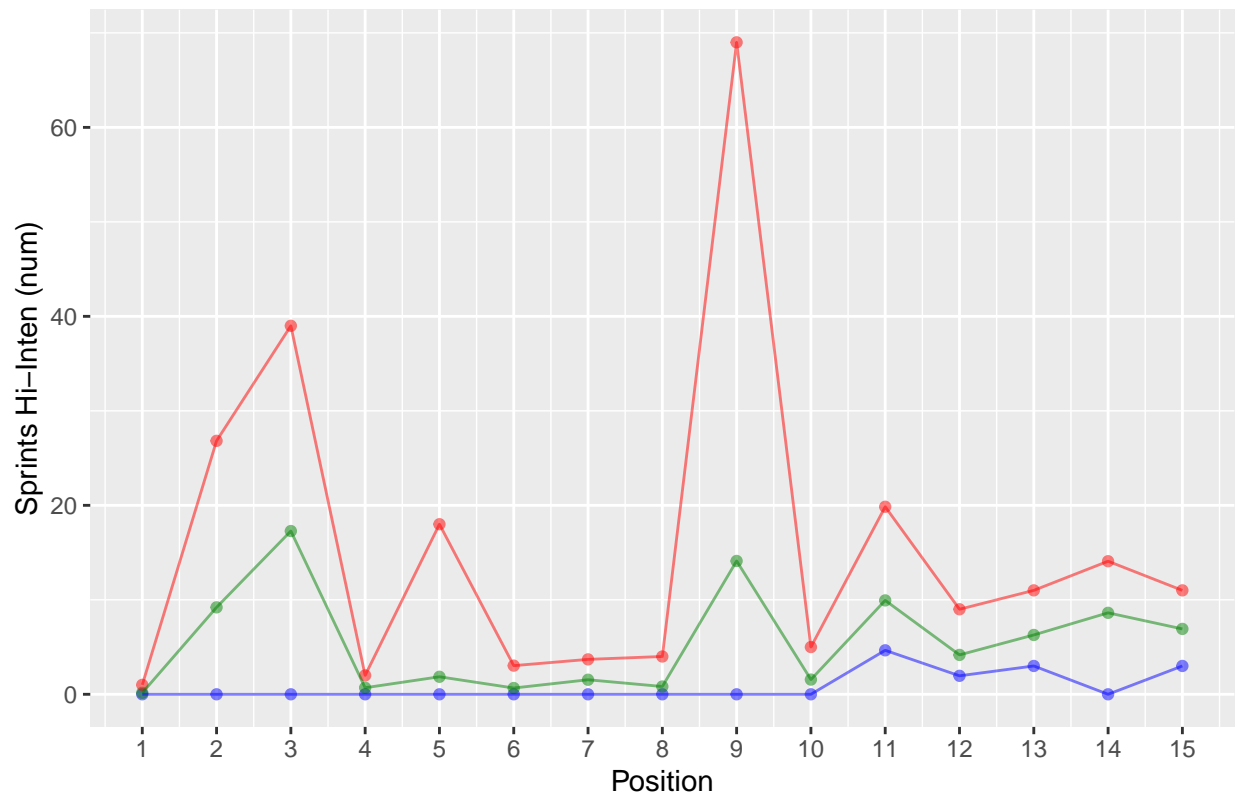


Positional Minimum, Mean and Maximum for Distance HR Hi-Inten (m)

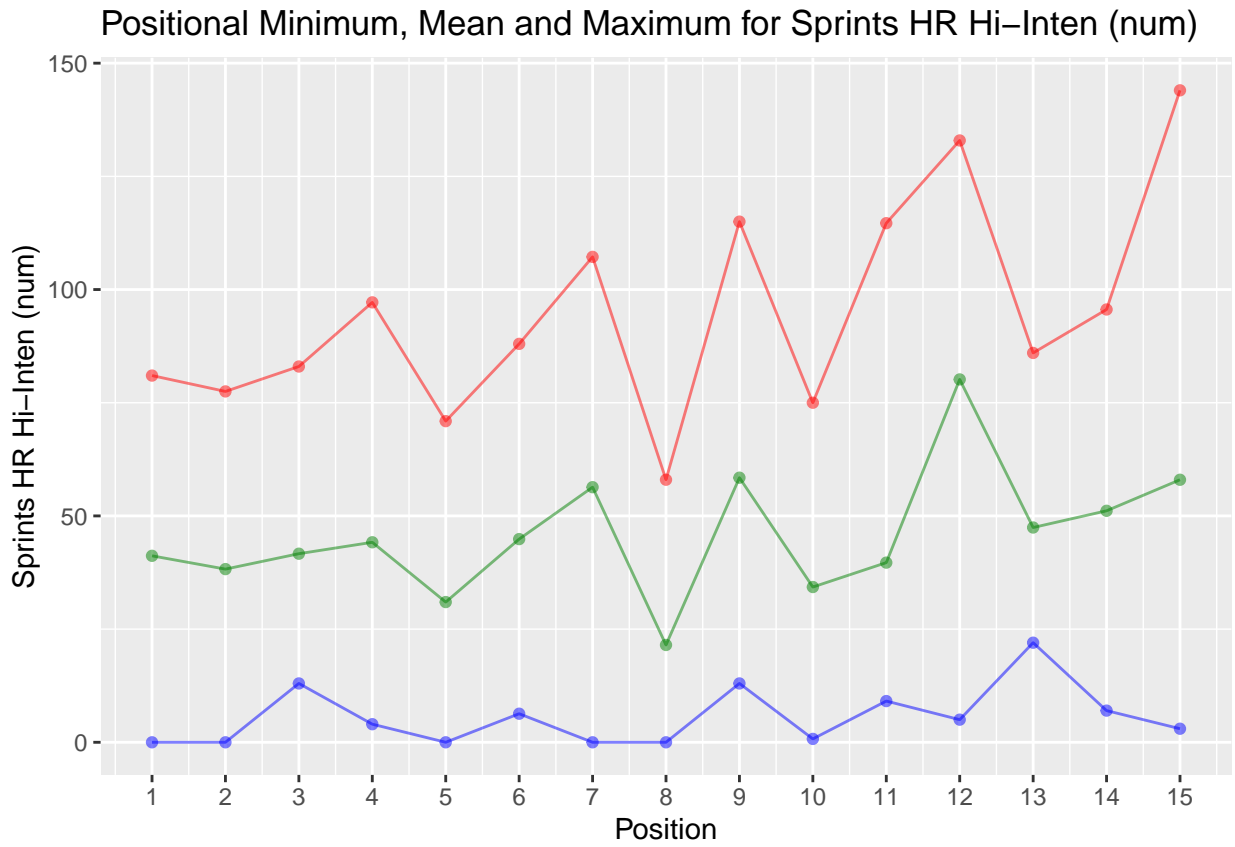




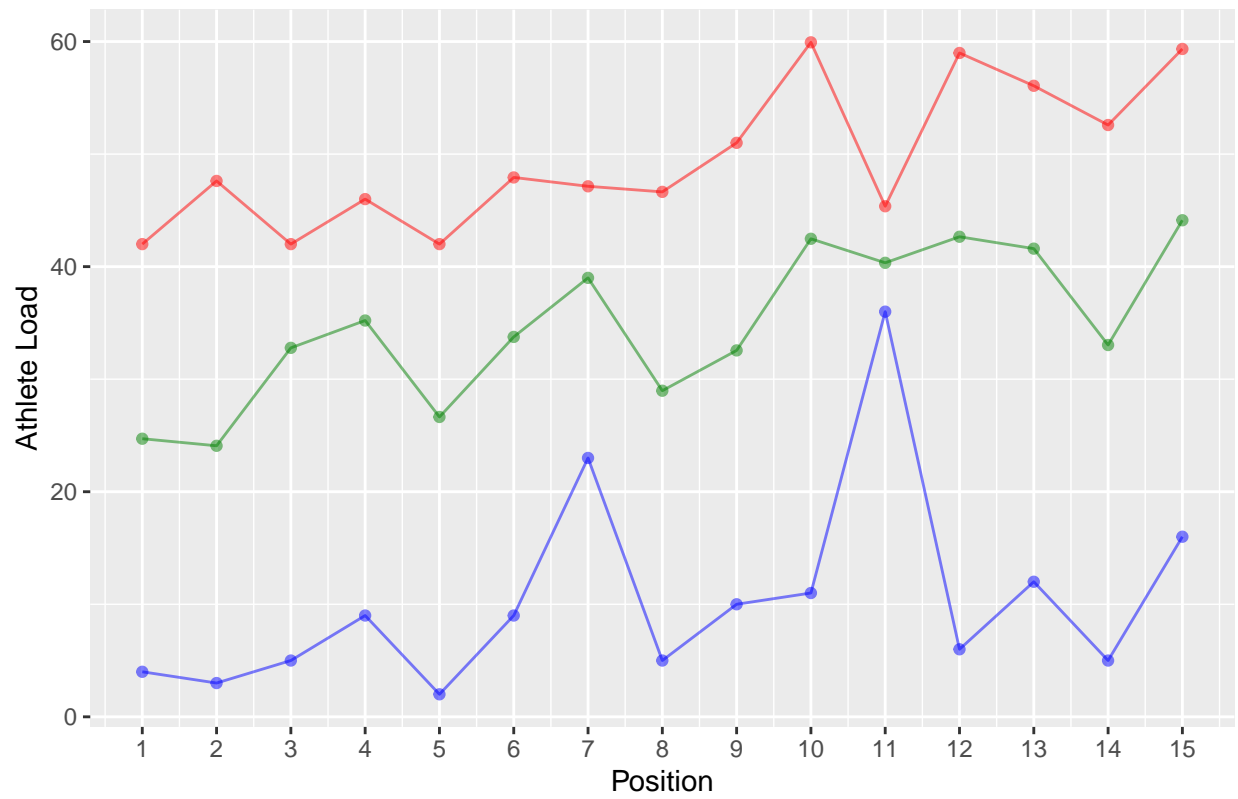
Positional Minimum, Mean and Maximum for Sprints Hi-Inten (num)



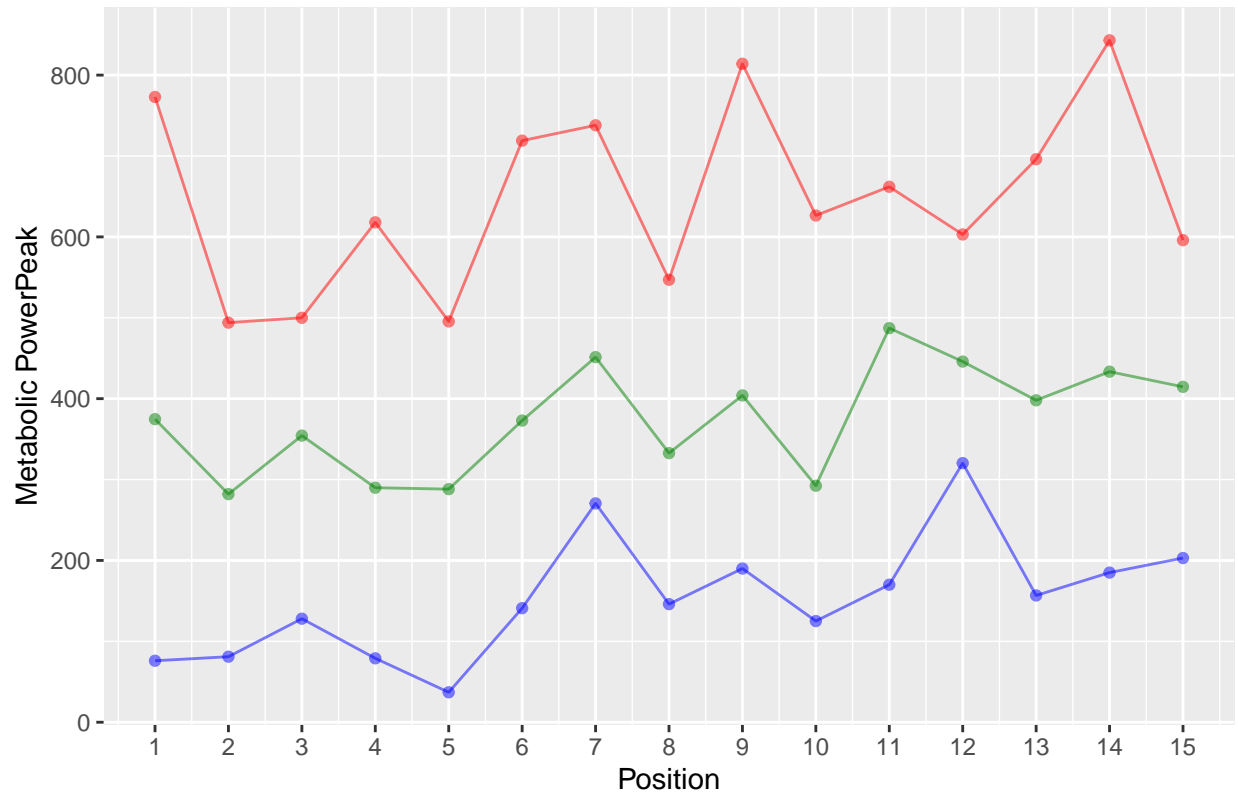




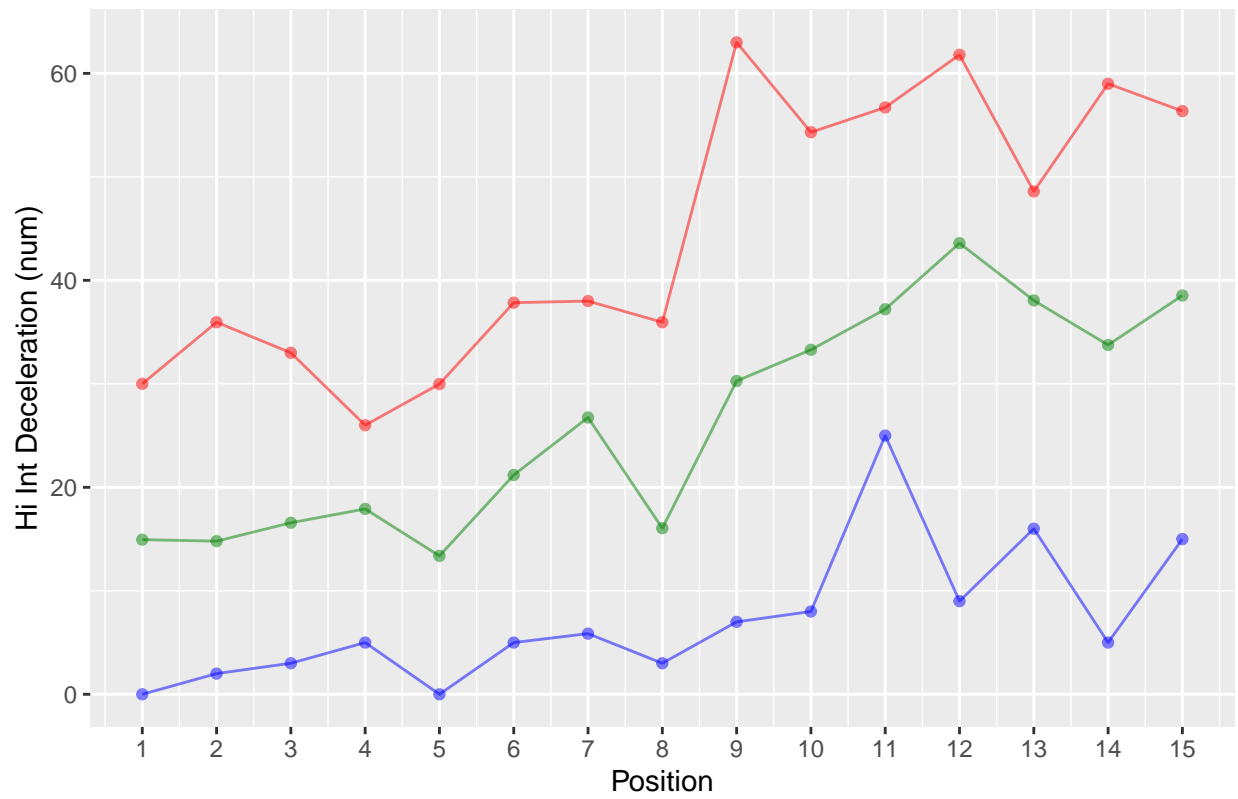
Positional Minimum, Mean and Maximum for Athlete Load

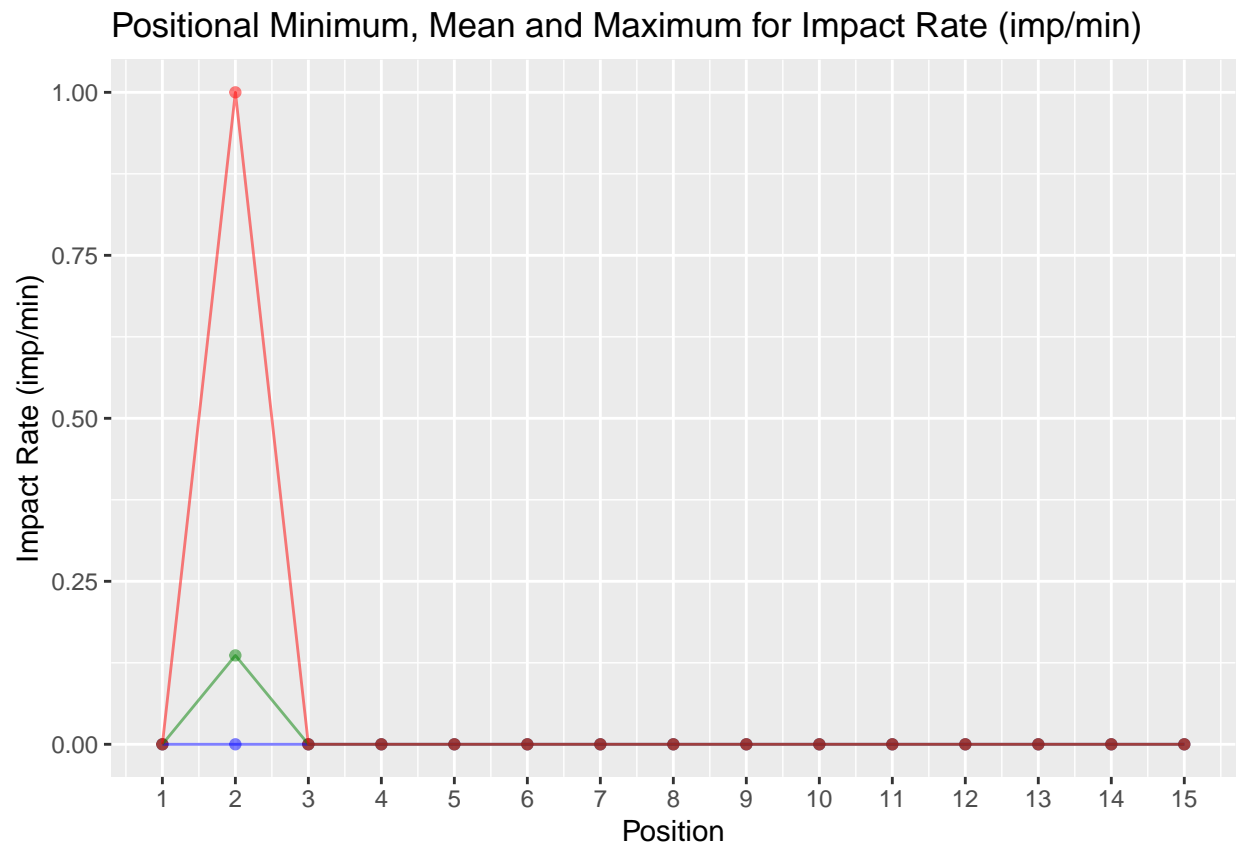


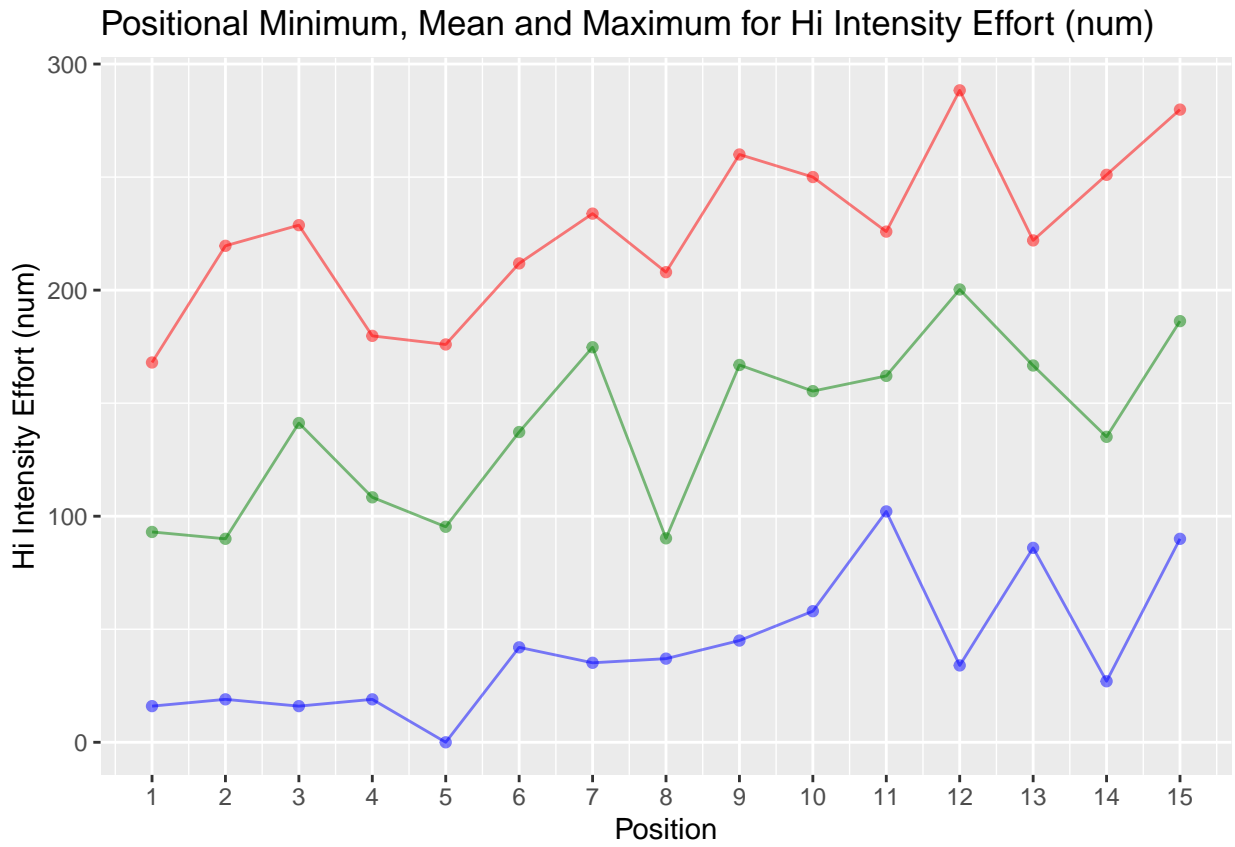
Positional Minimum, Mean and Maximum for Metabolic PowerPeak



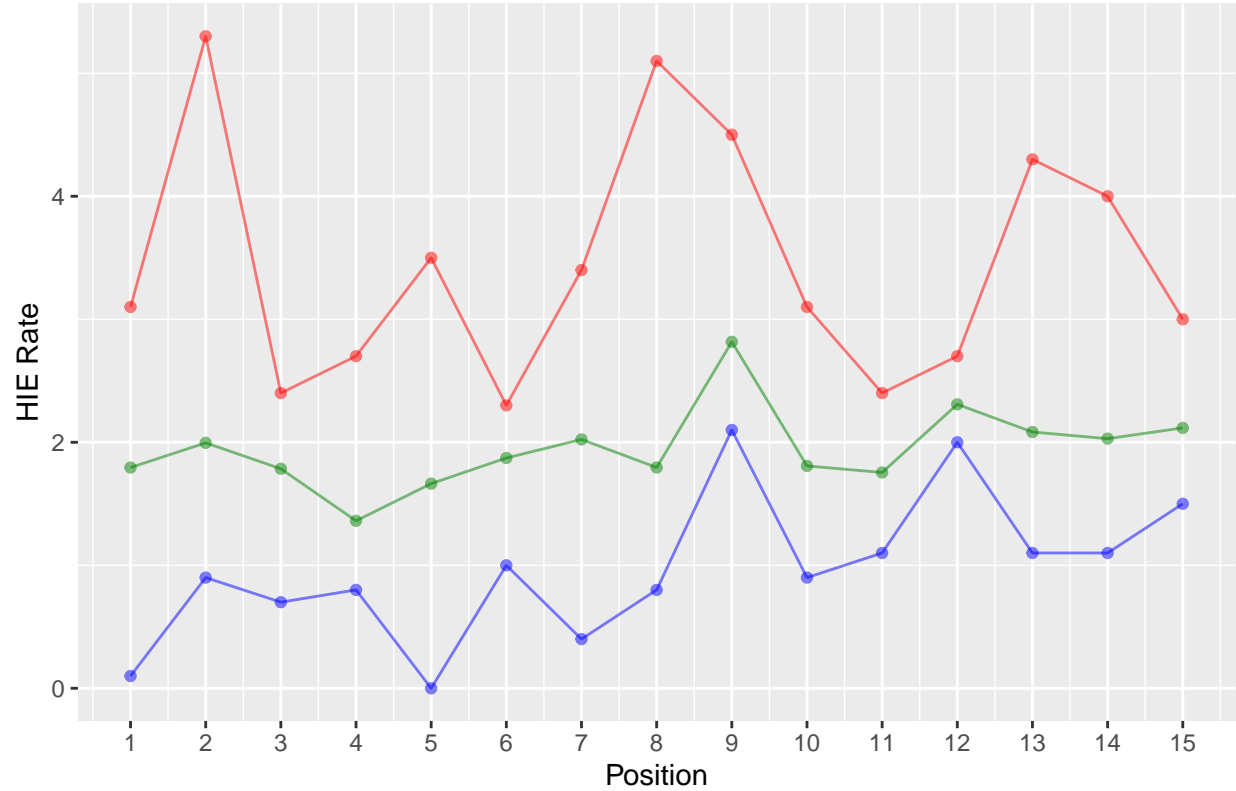
Positional Minimum, Mean and Maximum for Hi Int Deceleration (num)

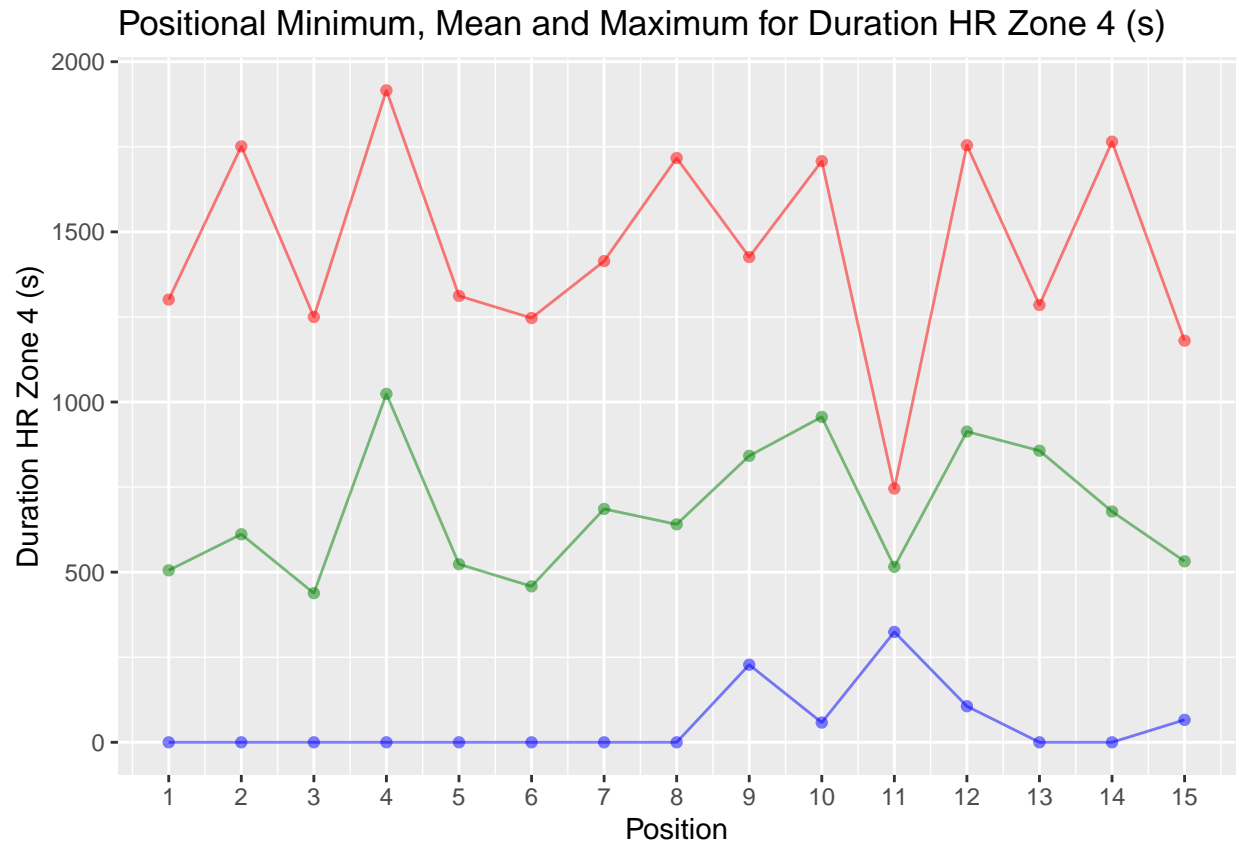






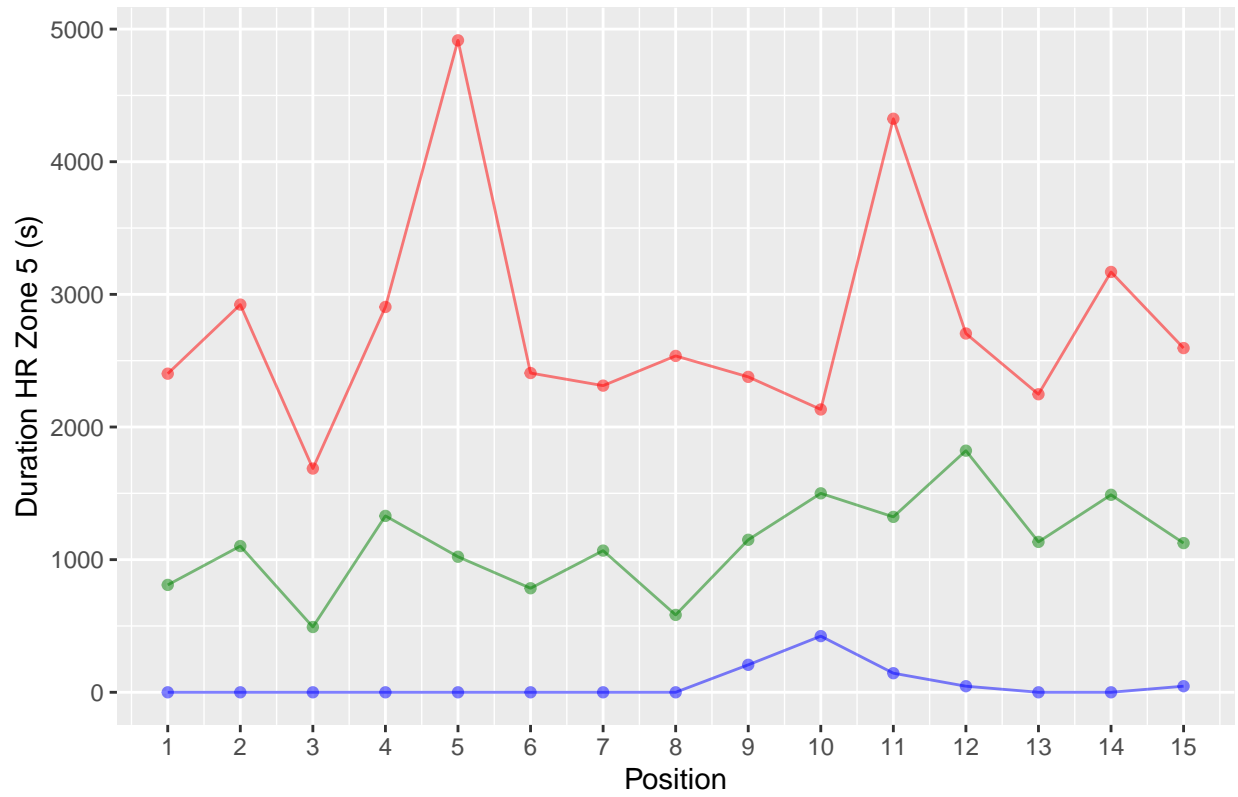
Positional Minimum, Mean and Maximum for HIE Rate



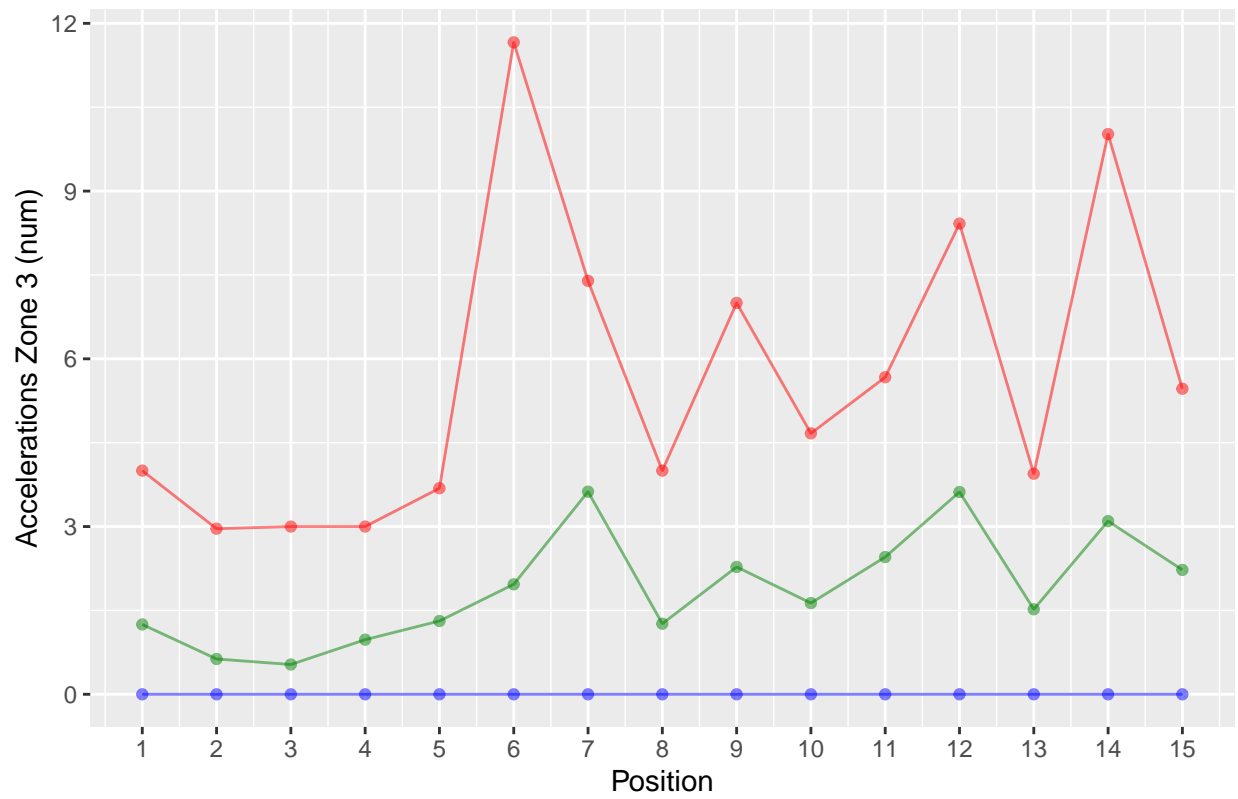


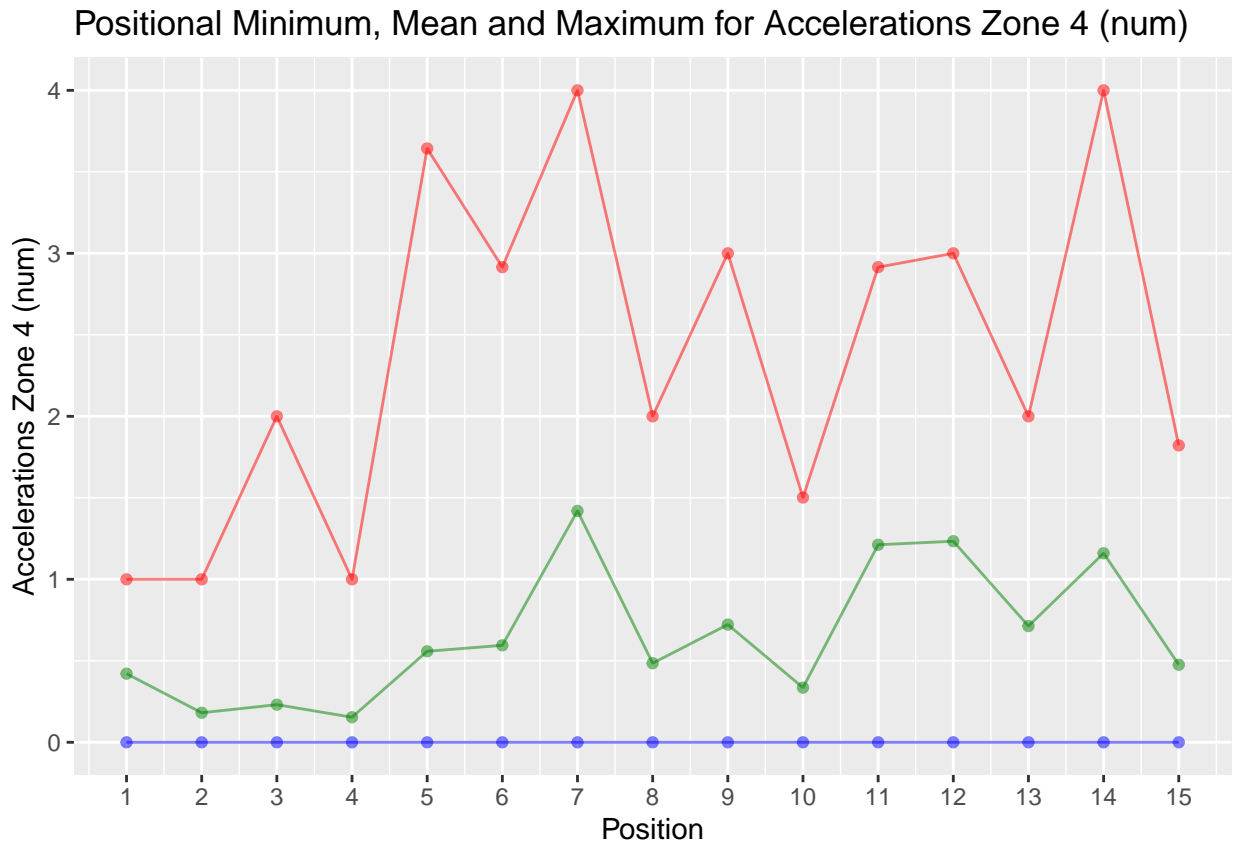


Positional Minimum, Mean and Maximum for Duration HR Zone 5 (s)

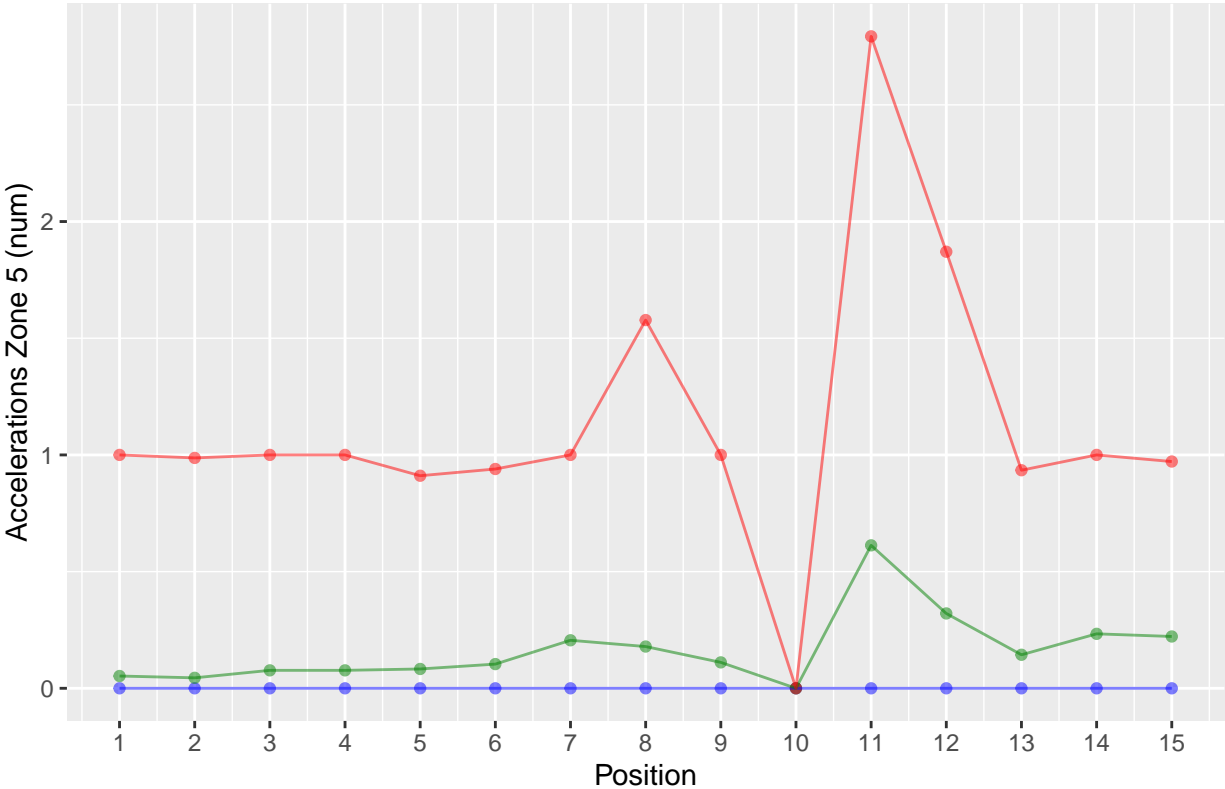


Positional Minimum, Mean and Maximum for Accelerations Zone 3 (num)

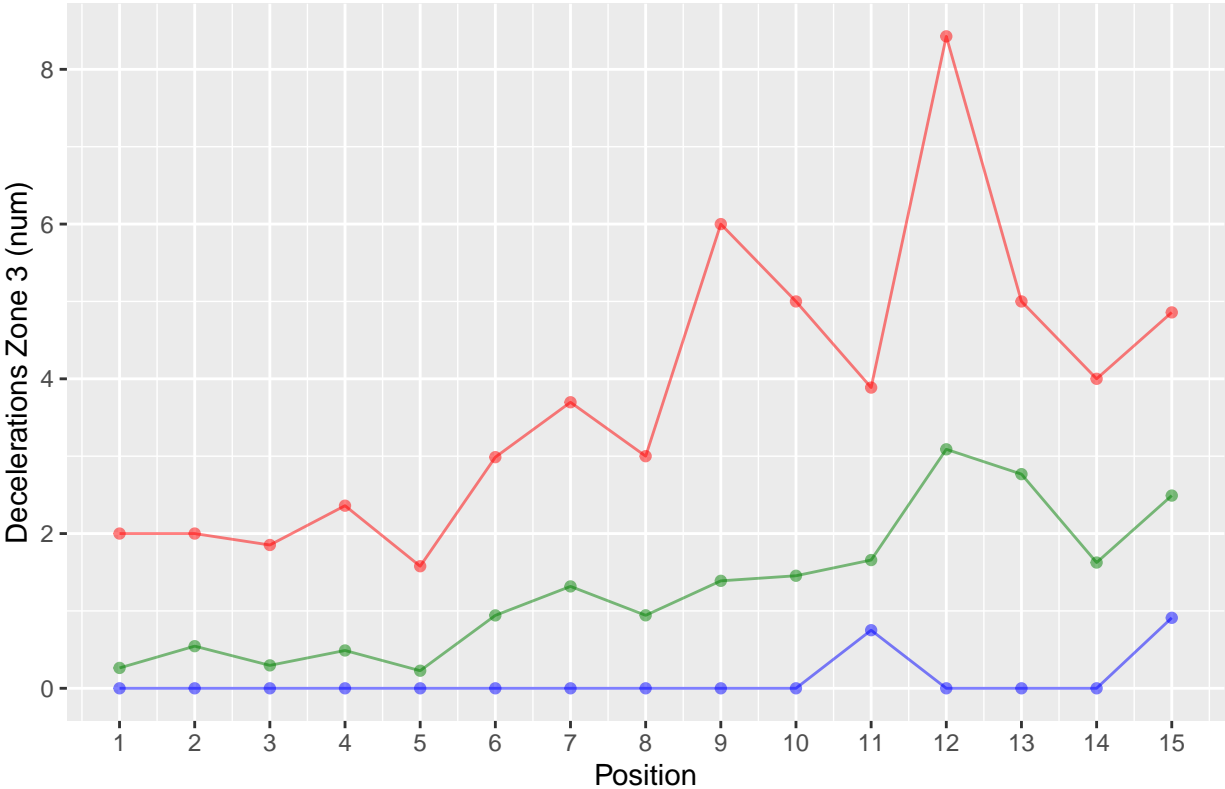




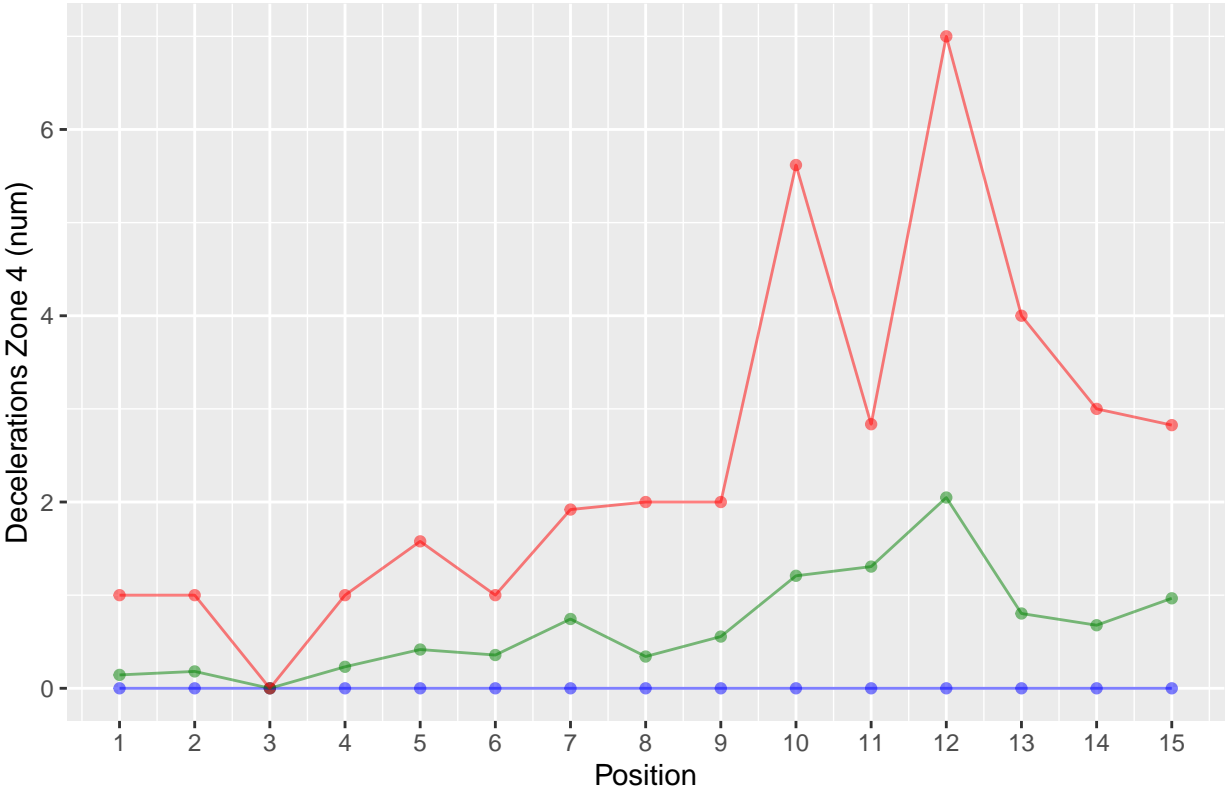
Positional Minimum, Mean and Maximum for Accelerations Zone 5 (num)



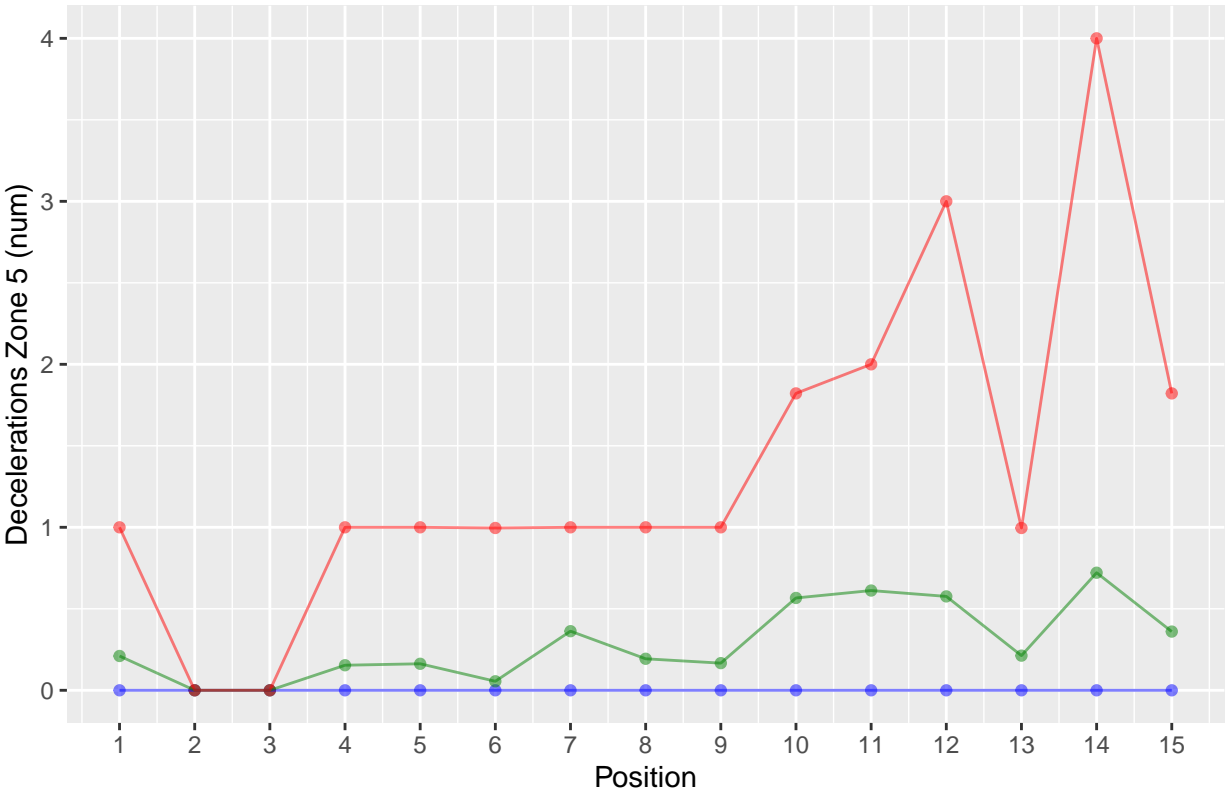
Positional Minimum, Mean and Maximum for Decelerations Zone 3 (num)



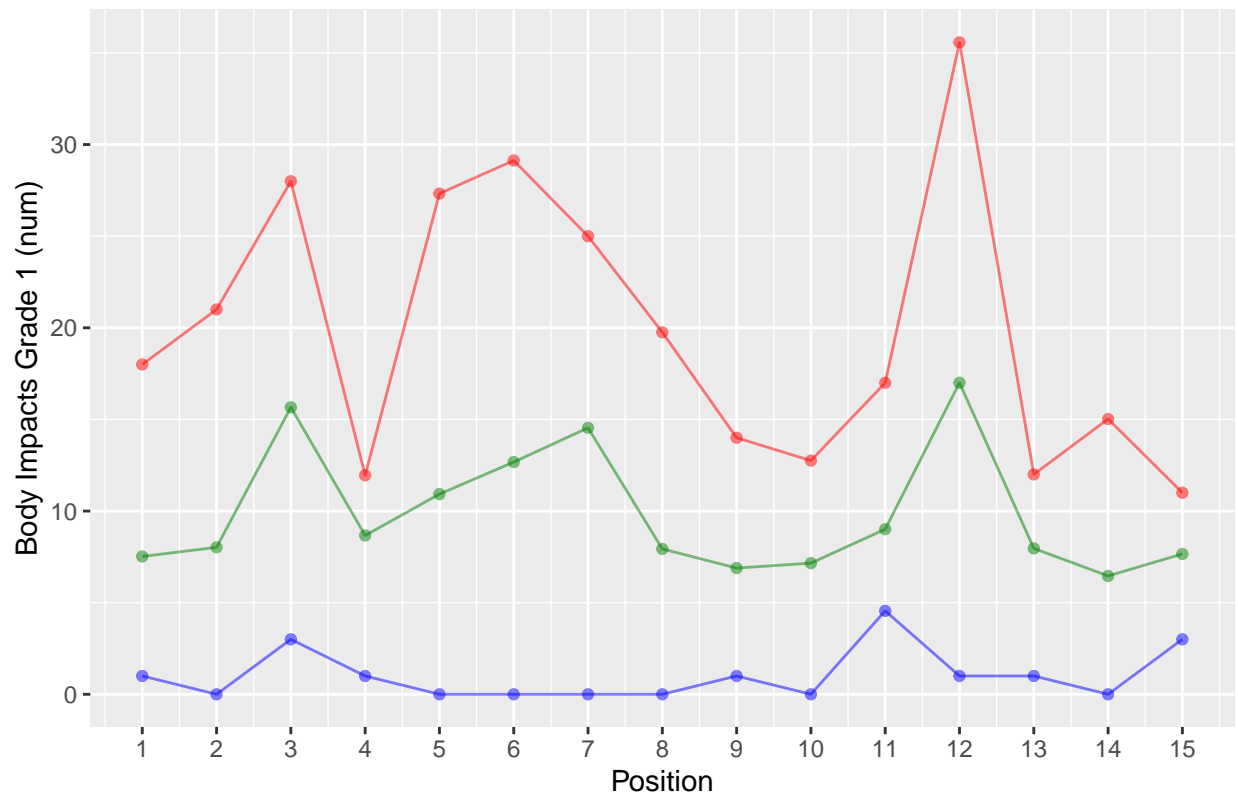
Positional Minimum, Mean and Maximum for Decelerations Zone 4 (num)



Positional Minimum, Mean and Maximum for Decelerations Zone 5 (num)

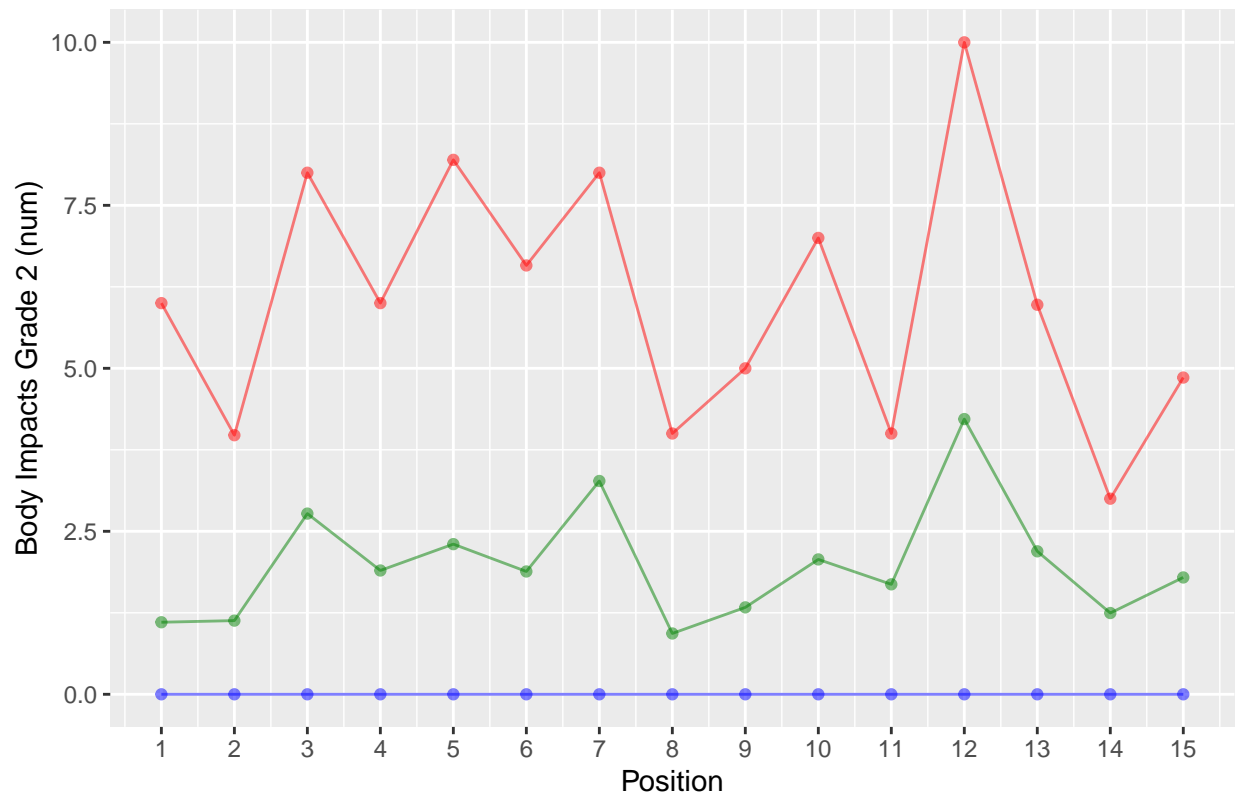


Positional Minimum, Mean and Maximum for Body Impacts Grade 1 (num)

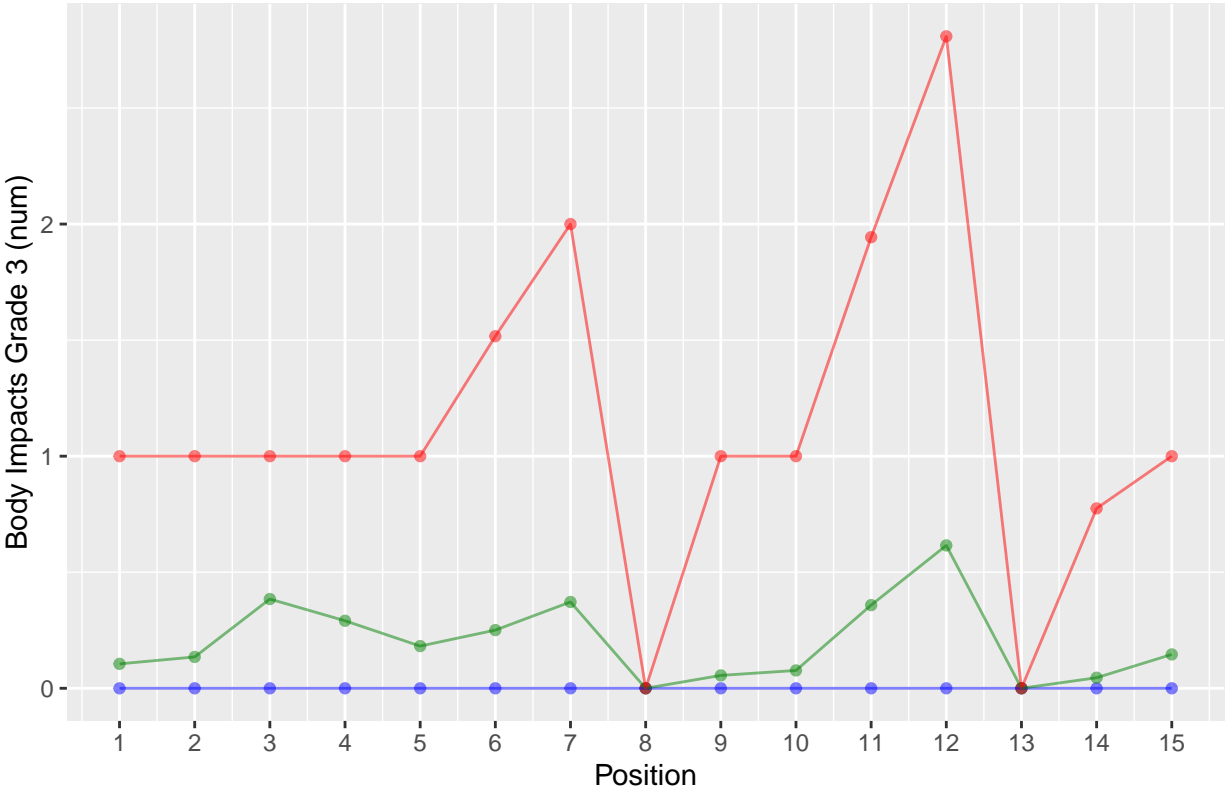




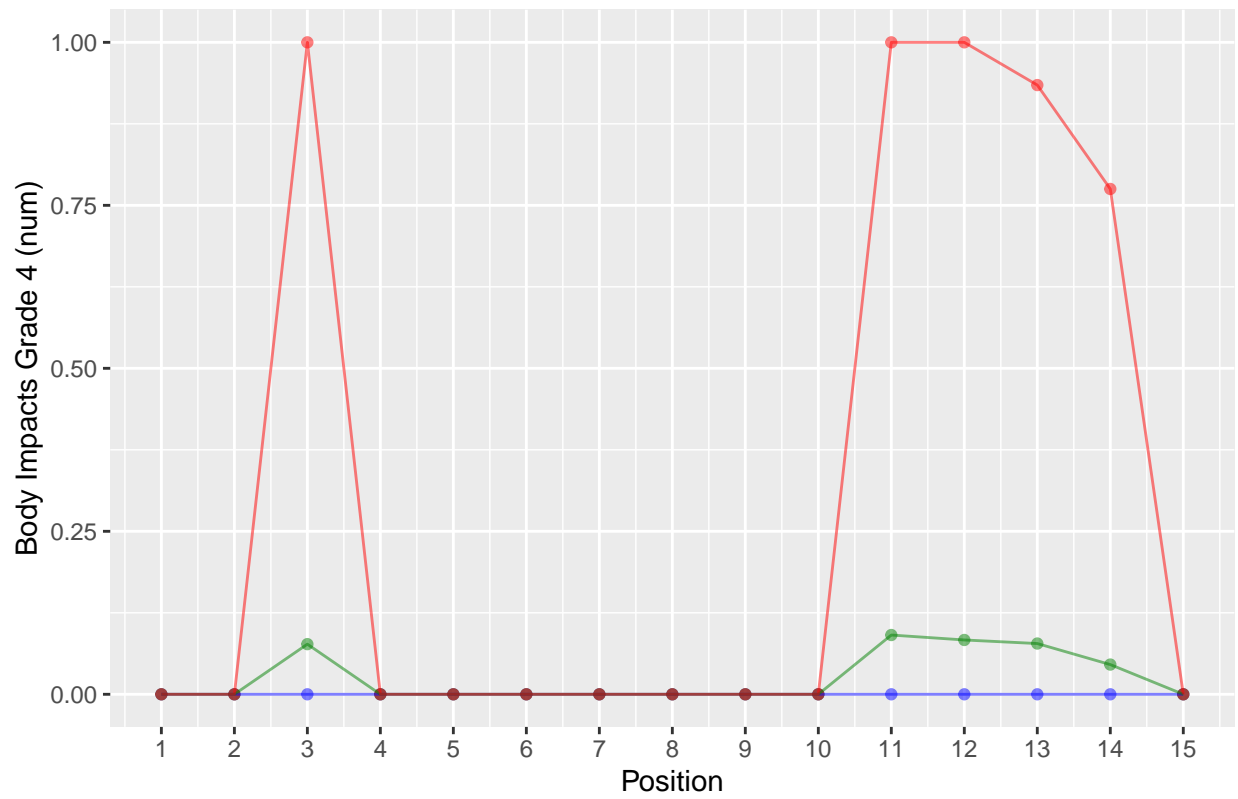
Positional Minimum, Mean and Maximum for Body Impacts Grade 2 (num)

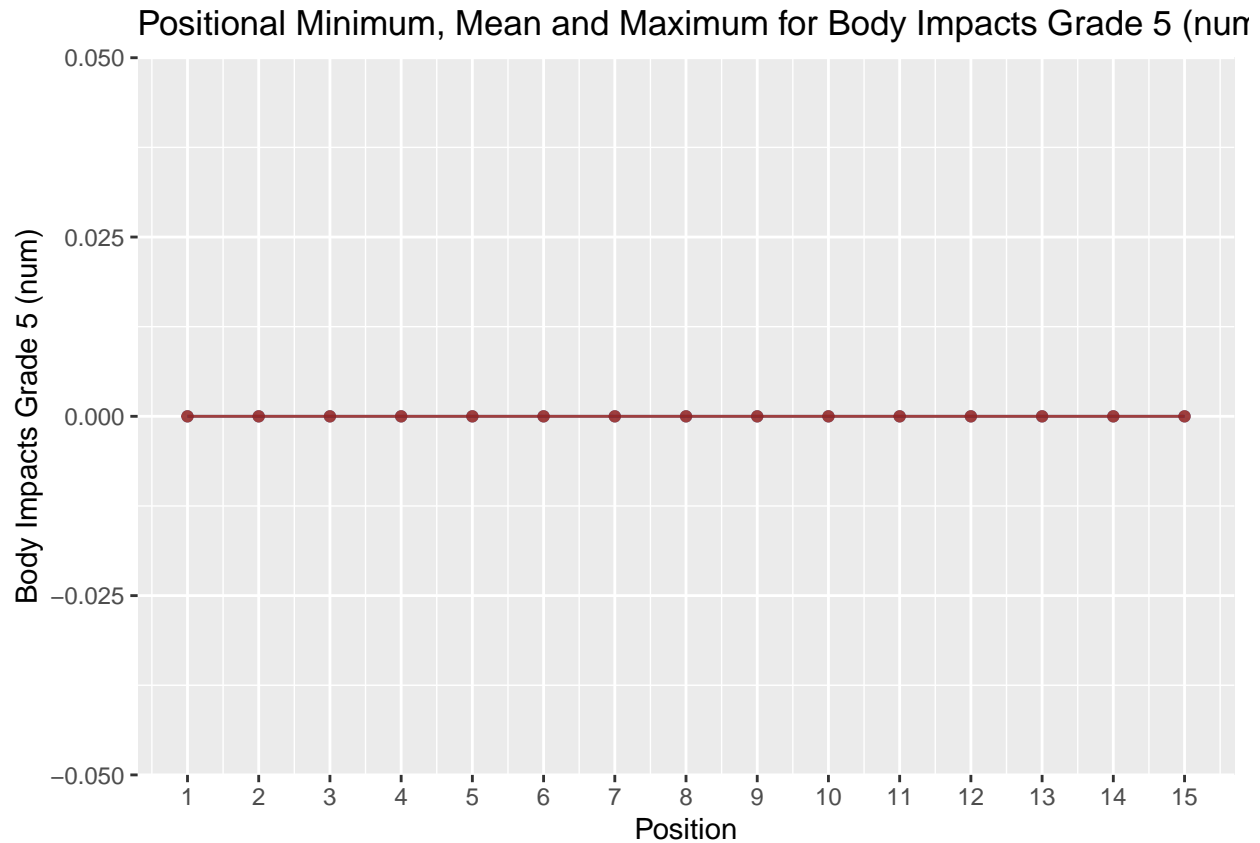


Positional Minimum, Mean and Maximum for Body Impacts Grade 3 (num)



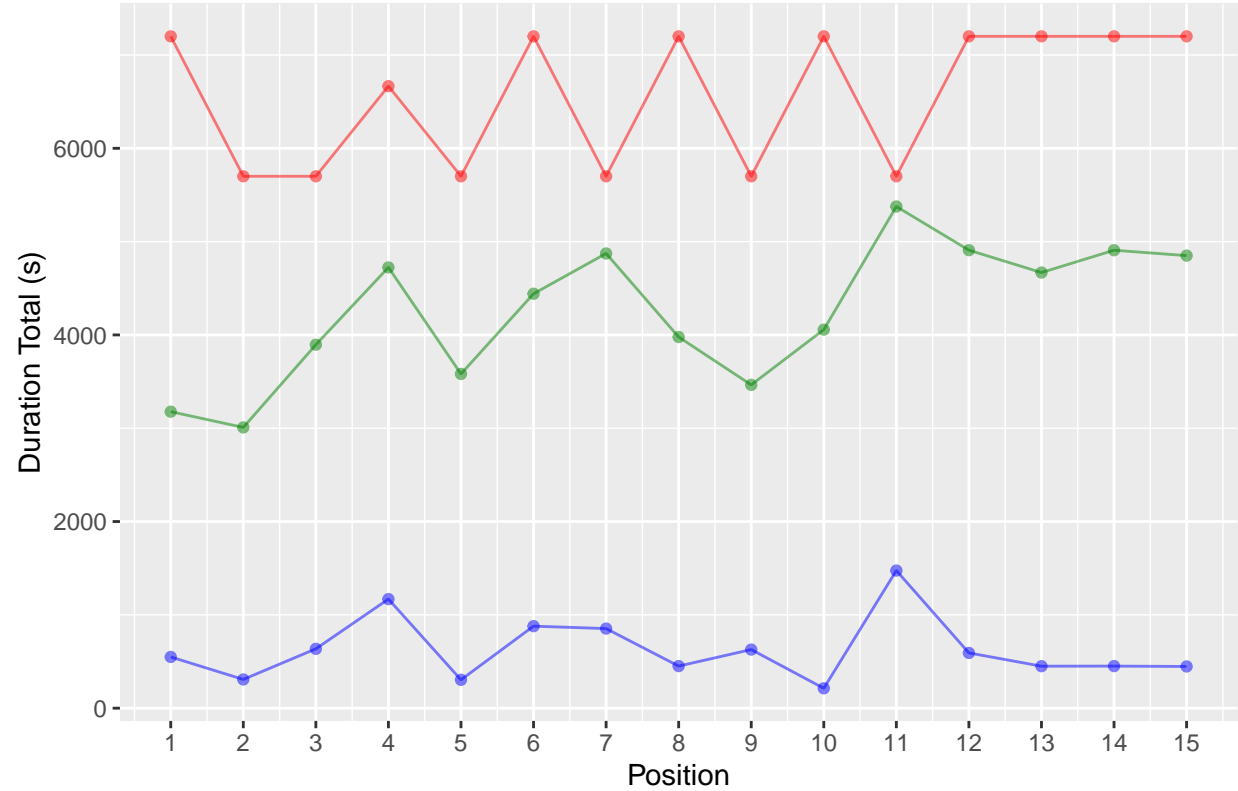
Positional Minimum, Mean and Maximum for Body Impacts Grade 4 (num)



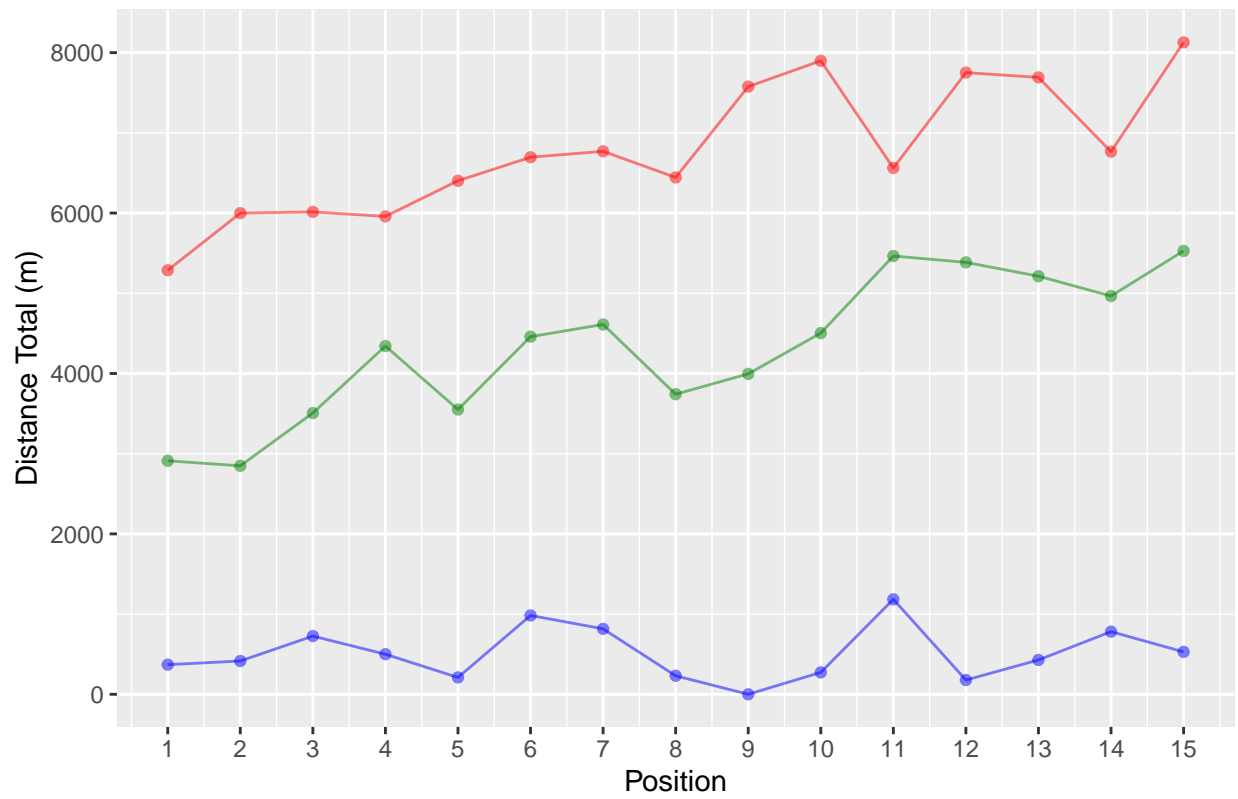


```
# Plotting positional minimum, mean and maximum for the variables shared between the 2018, 2019 and 2020
for (var in 5:17) {
  valuesOfInterest <- data.frame(Position = sort(unique(combinedData$Position)), Minimum = 0, Mean = 0,
  for (pos in 1:15) {
    positionalVector <- combinedData[which(combinedData$Position == pos), var]
    valuesOfInterest[pos, 2] <- min(positionalVector, na.rm = TRUE)
    valuesOfInterest[pos, 3] <- mean(positionalVector, na.rm = TRUE)
    valuesOfInterest[pos, 4] <- max(positionalVector, na.rm = TRUE)
  }
  print(ggplot(valuesOfInterest, aes(x = as.numeric(Position))) +
    geom_point(aes(y = Minimum), col = "blue", alpha = 0.5) +
    geom_point(aes(y = Mean), col = "#008000", alpha = 0.5) +
    geom_point(aes(y = Maximum), col = "red", alpha = 0.5) +
    geom_line(aes(y = Minimum), col = "blue", alpha = 0.5) +
    geom_line(aes(y = Mean), col = "#008000", alpha = 0.5) +
    geom_line(aes(y = Maximum), col = "red", alpha = 0.5) +
    scale_x_continuous(breaks = 2:16, labels = as.character(1:15)) +
    xlab("Position") +
    ylab(colnames(combinedData)[var]) +
    ggtitle(paste0("Positional Minimum, Mean and Maximum for ", colnames(combinedData)[var]))
  )
}
```

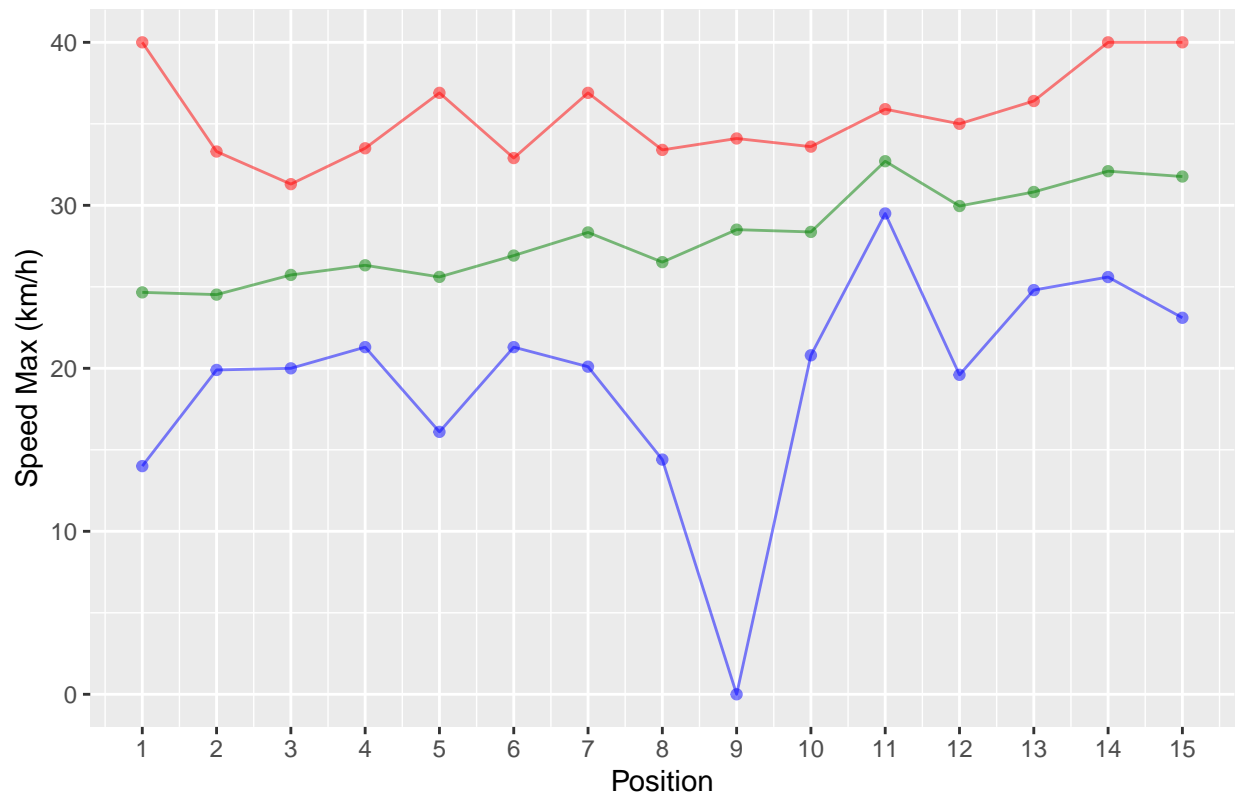
Positional Minimum, Mean and Maximum for Duration Total (s)

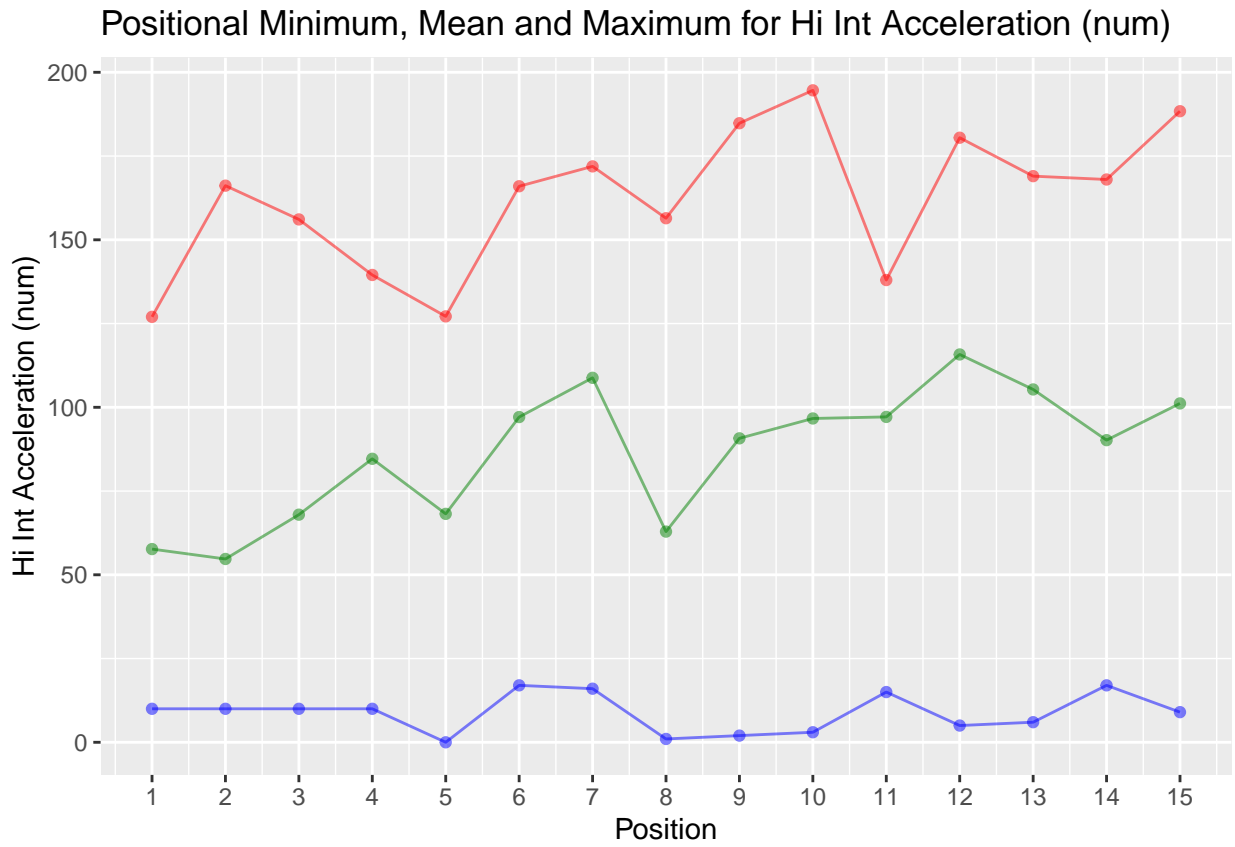


Positional Minimum, Mean and Maximum for Distance Total (m)



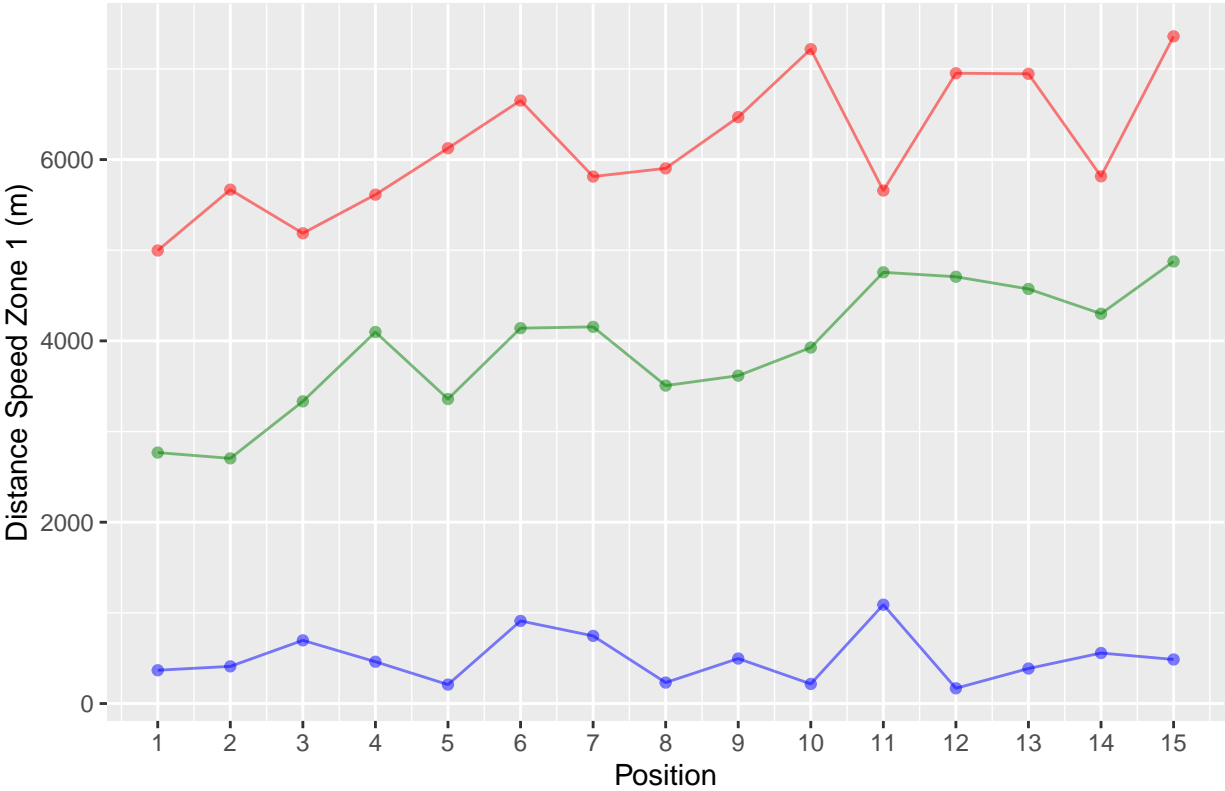
Positional Minimum, Mean and Maximum for Speed Max (km/h)



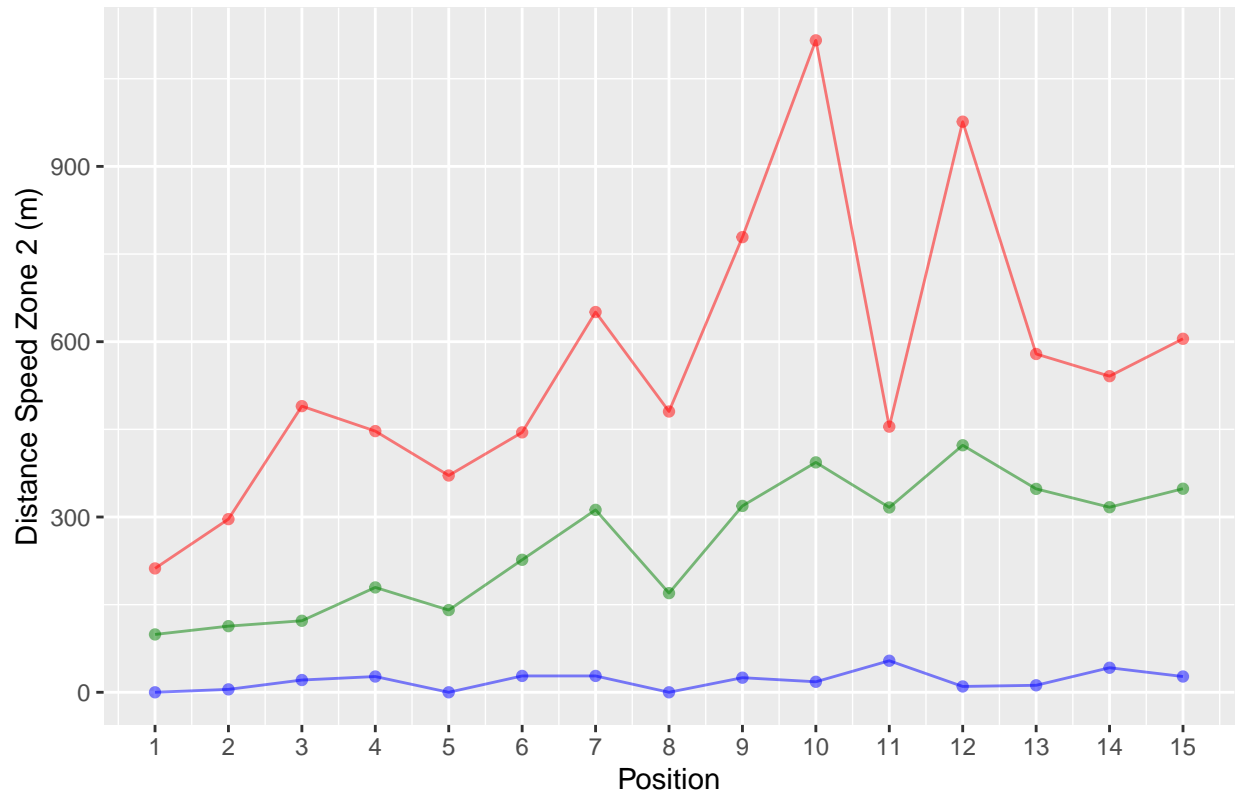




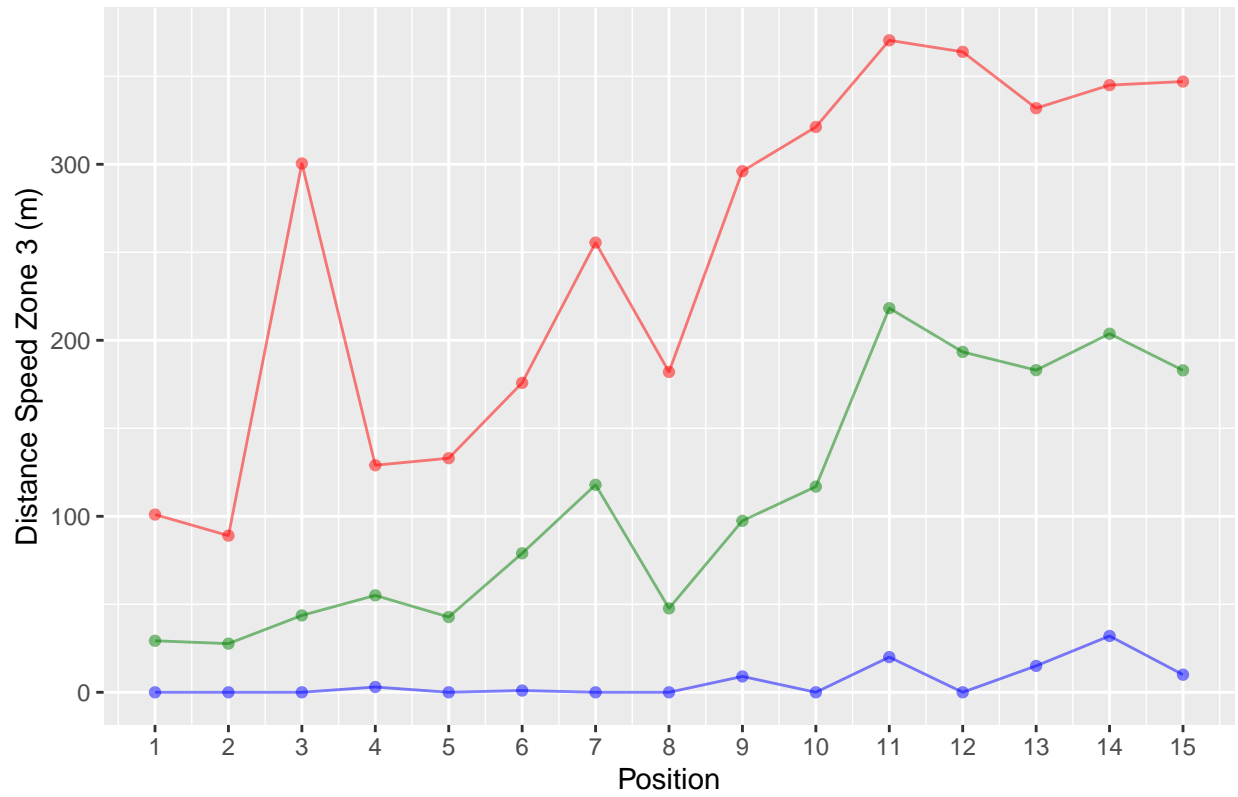
Positional Minimum, Mean and Maximum for Distance Speed Zone 1 (m)

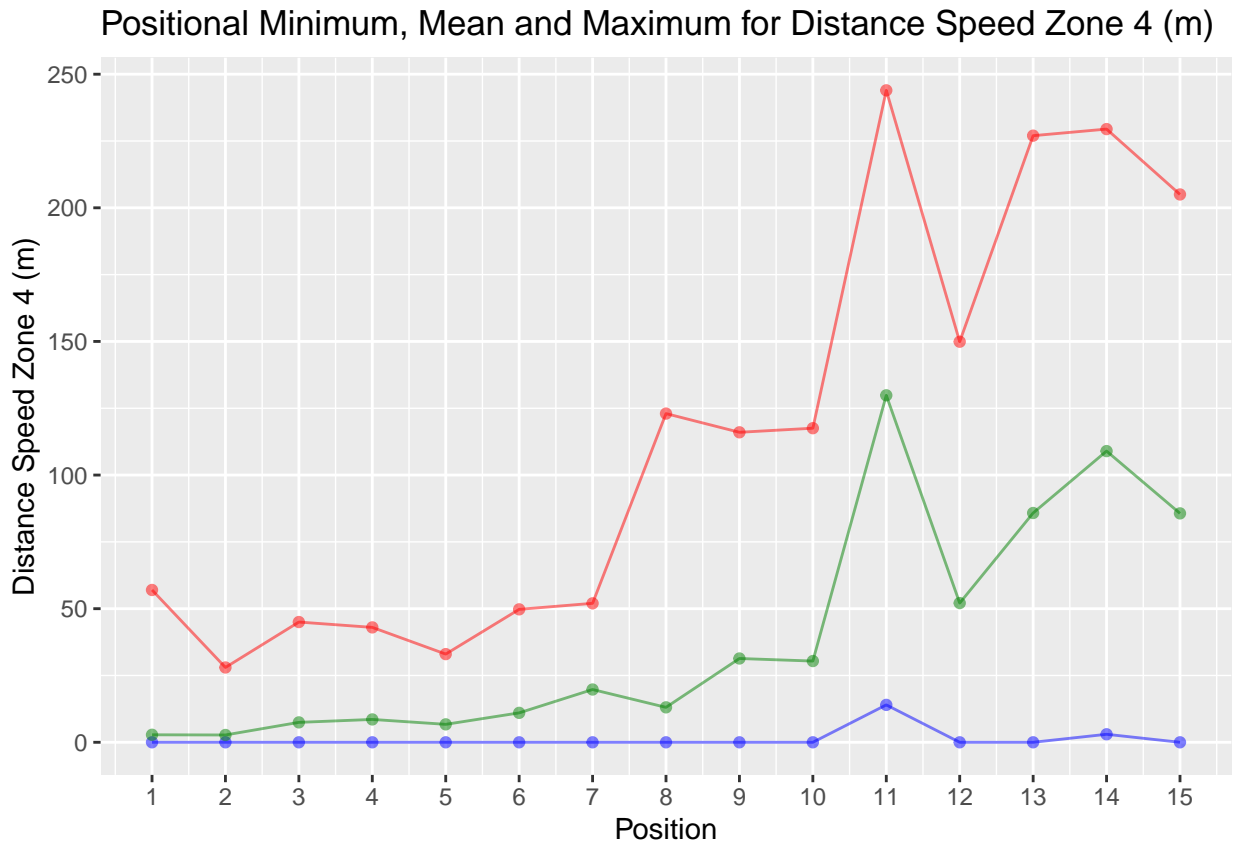


Positional Minimum, Mean and Maximum for Distance Speed Zone 2 (m)

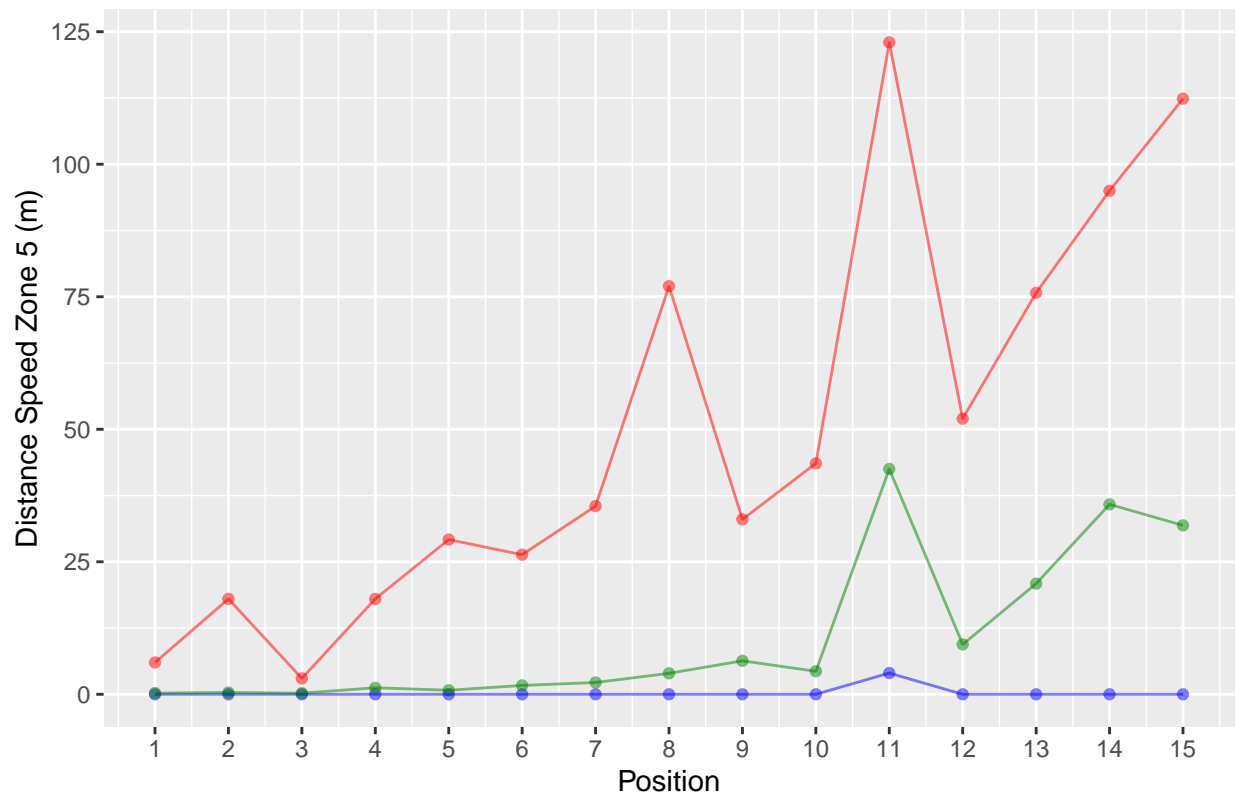


Positional Minimum, Mean and Maximum for Distance Speed Zone 3 (m)

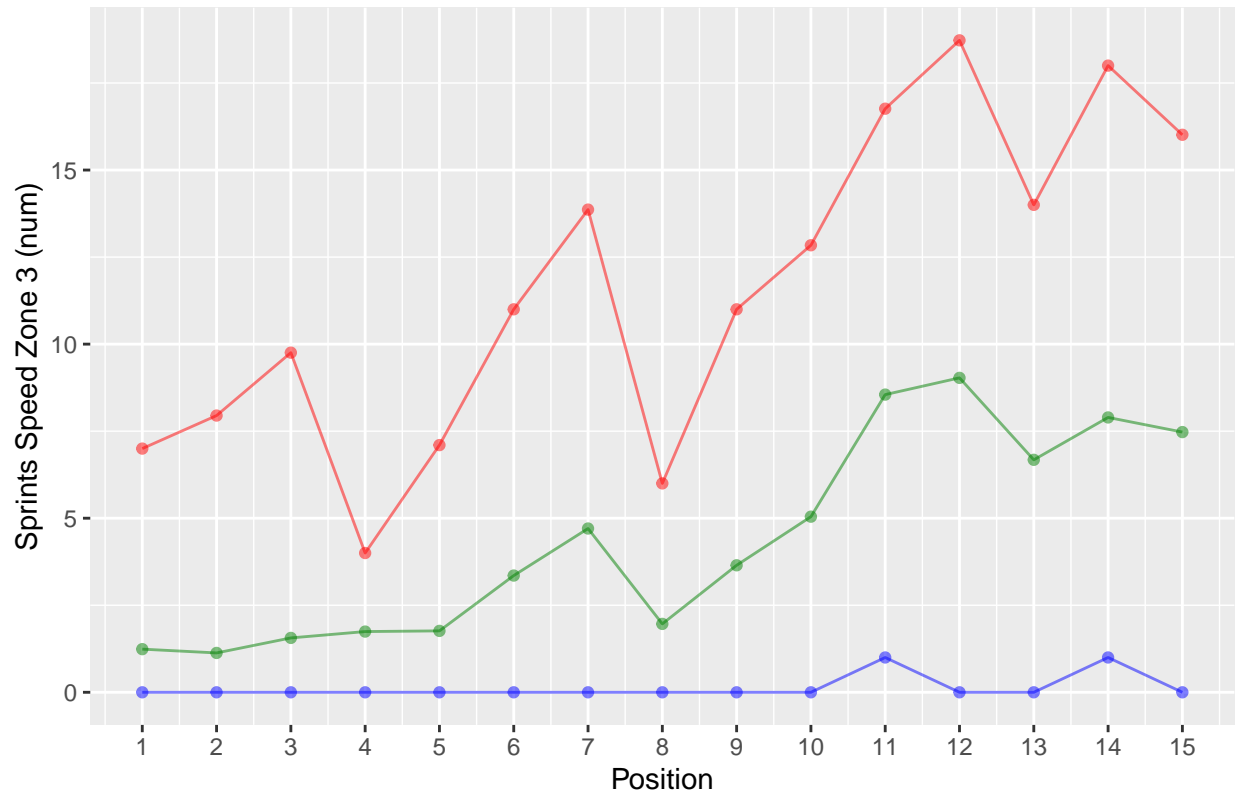




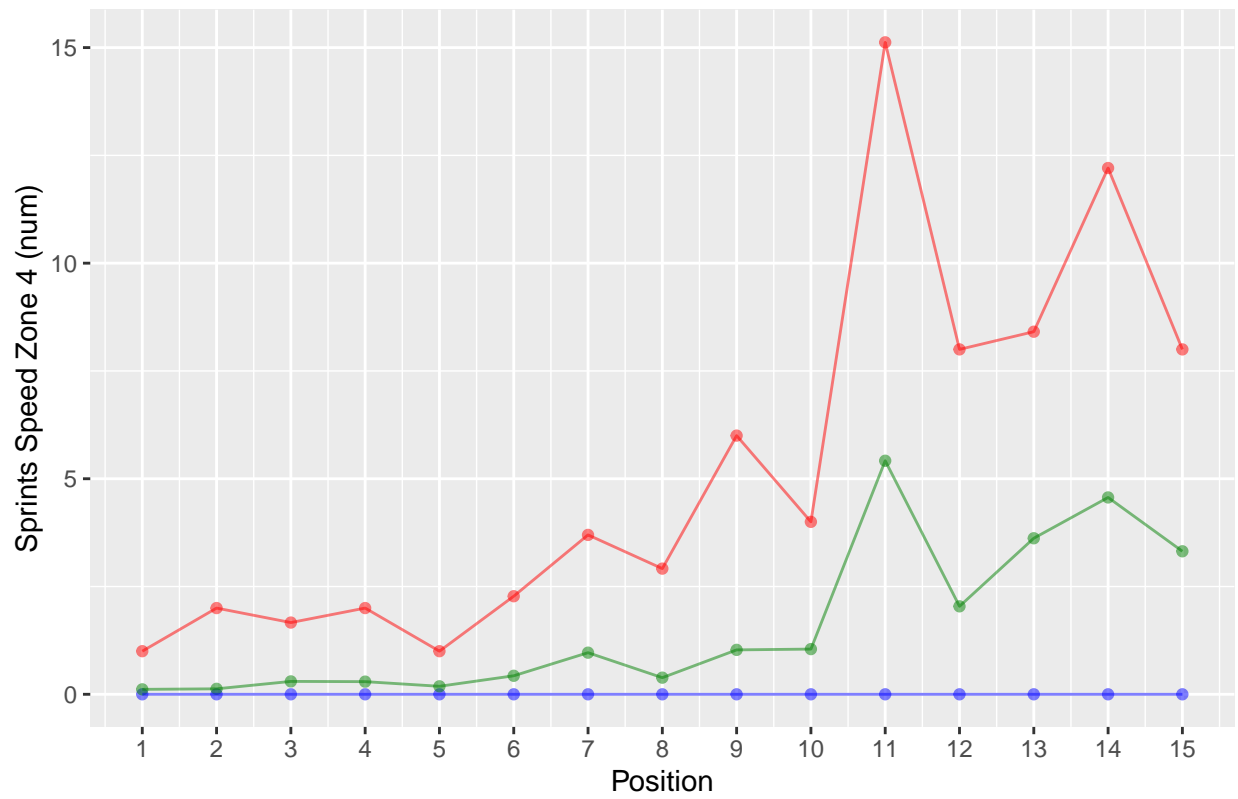
Positional Minimum, Mean and Maximum for Distance Speed Zone 5 (m)

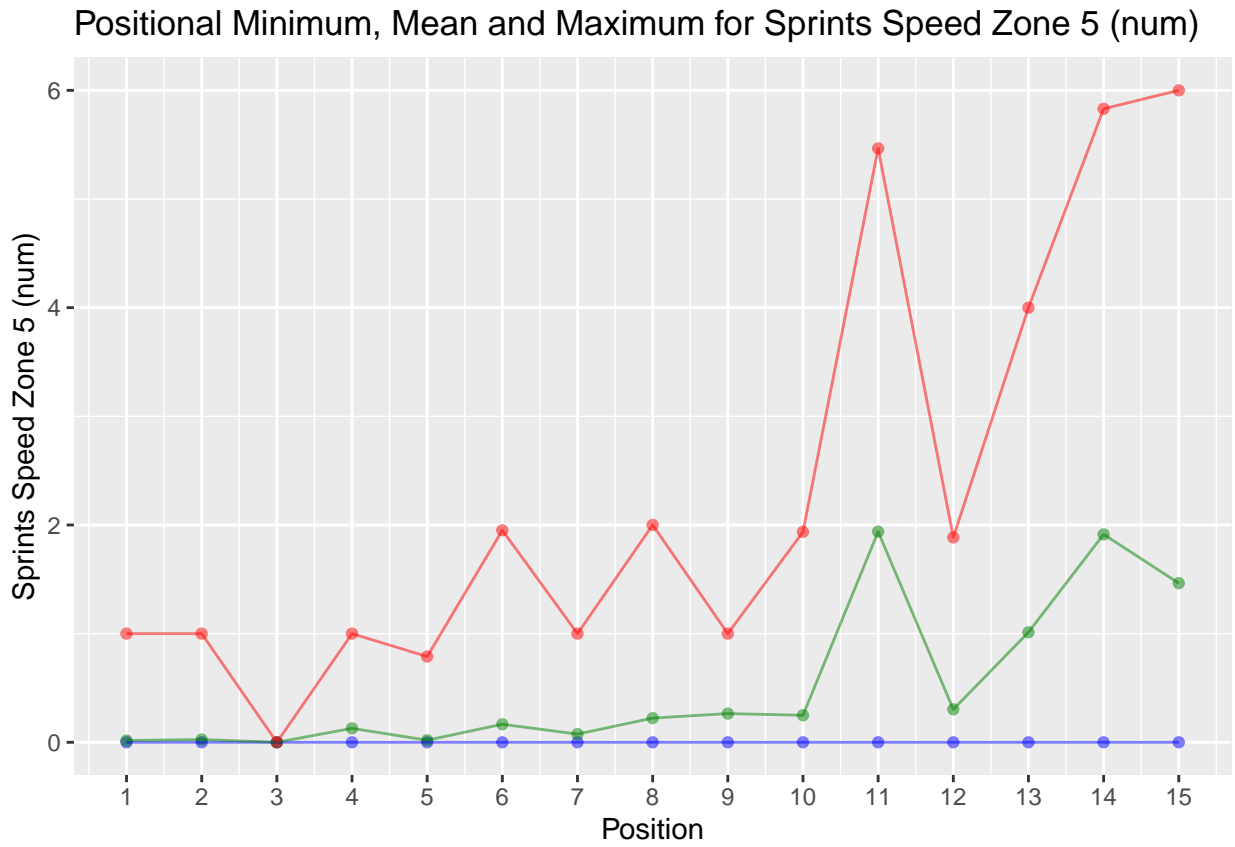


Positional Minimum, Mean and Maximum for Sprints Speed Zone 3 (num)



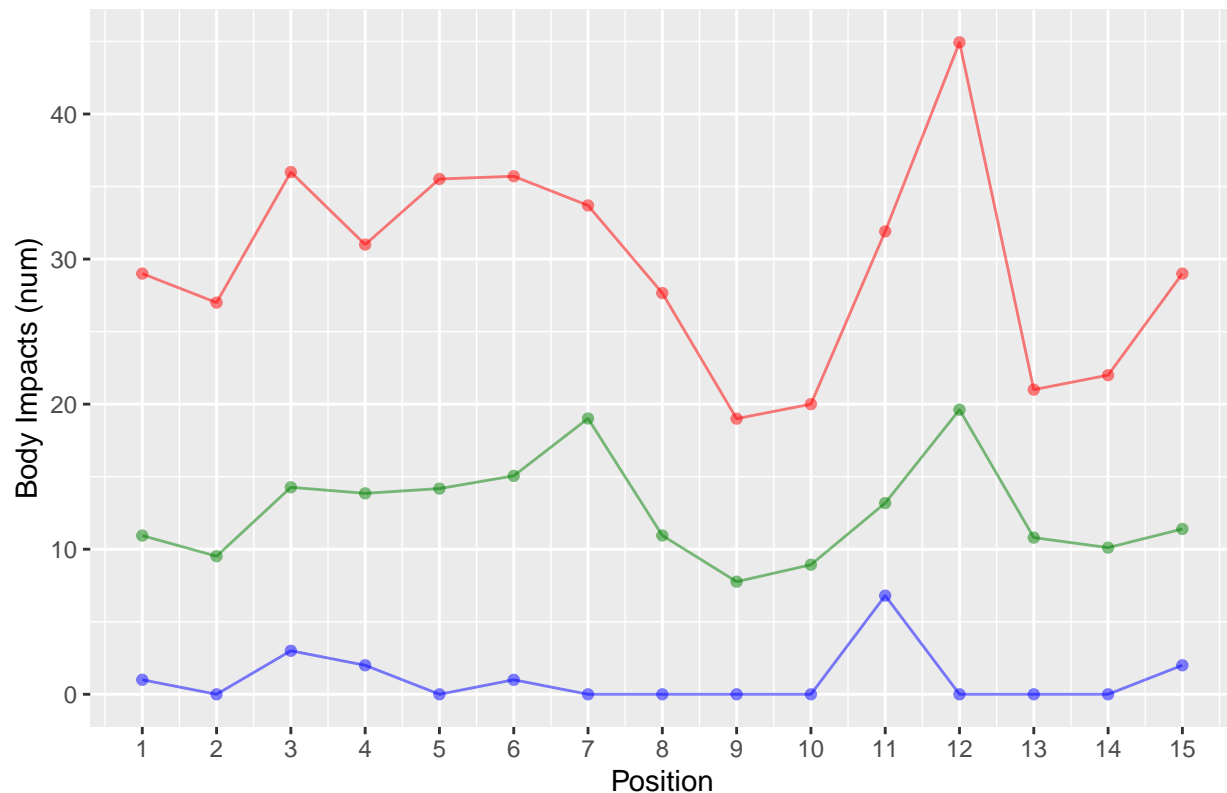
Positional Minimum, Mean and Maximum for Sprints Speed Zone 4 (num)







## Positional Minimum, Mean and Maximum for Body Impacts (num)



## Fitting models to find top variables by position

To find the top variables by position, models need to be fitted. The easiest way to do this is to fit a separate model for the data filtered by each position.

First, the data is preprocessed one more time. **Position** previously had a factor 16, used to represent replacements. This is no longer used, and as such is removed. The factor **Work Recovery Ratio** is problematic, as it has a large number of **NA** values, even among the 2018 data. These values are replaced by “Not Applicable”, and this is set as the reference level.

Some variables are very sparse or simply do not have much variance. These are removed with `nearZeroVar()` from the `caret` package.

Finally, values are imputed for variables initially not present in the 2019 and 2020 data. Median imputation is used to impute median values for each variable in for all the missing values in the 2019 and 2020 data.

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.0.5
```

```
## Loading required package: lattice
```

```
##
```

```
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
```

```
##
```

```
## lift
```

```

# Combining the 2018, 2019 and 2020 datasets
fullyCombined <- full_join(master2018, combinedData)

## Joining, by = c("Athlete", "Team", "Date", "Start Time", "Duration Total (s)", "Distance Total (m)",
dim(fullyCombined)

## [1] 647 48

# Removing the unused levels for position
fullyCombined$Position <- droplevels(fullyCombined$Position)
# Changing NA for Work Recovery Ratio to an actual level
levels(fullyCombined$`Work Recovery Ratio`) <- c(levels(fullyCombined$`Work Recovery Ratio`), "Not Applicable")
fullyCombined[which(is.na(fullyCombined$`Work Recovery Ratio`)), 16] <- "Not Applicable"
# Setting "Not Applicable" as the reference level
fullyCombined$`Work Recovery Ratio` <- relevel(fullyCombined$`Work Recovery Ratio`, "Not Applicable")
fullyCombined$`Work Recovery Ratio` <- droplevels(fullyCombined$`Work Recovery Ratio`)
# Removing variables with almost zero variance
fullyCombined <- fullyCombined[, -c(nearZeroVar(fullyCombined))]
dim(fullyCombined)

## [1] 647 44

# Imputing values for the variables that were not initially present in the 2019 and 2020 data
for (u in 1:15) {
  imputations <- preprocess(fullyCombined[which(fullyCombined$Position == u), ], method = "medianImpute")
  fullyCombined[which(fullyCombined$Position == u), ] <- predict(imputations, fullyCombined[which(fullyCombined$Position == u), ])
}

```

## Backward stepwise selection

For backward stepwise selection, the datasets are split by position. A full model is fitted for each split to obtain the coefficients for the variables when all are taken into account. Some variables may result in singularities, which are most likely due to highly correlated variables coexisting in the dataset. By creating a correlation matrix with `cor()`, and finding variables with correlations beyond a certain cutoff using `findCorrelation()`, these variables can be singled out and removed from the data.

The top five variables by backward stepwise selection are then determined with `regsubsets(..., method = "backward")`. The most important variable is removed last, and as such is the only variable in the one-variable model. The second-most important variable is removed penultimately, and as such is the variable that differs between the one-variable and two-variable models, etc. The full model coefficients for these variables can then be determined and analysed.

The results are presented as an ordered list, from most important variable to fifth-most important variable. Next to each selected variable is its coefficient estimate and its corresponding p-value. Underneath each selected variable is at least one bullet point that provides in plain English an interpretation of the 95% confidence interval for the variable's coefficient estimate.

### Position 1: Loosehead prop

```

library(leaps)

## Warning: package 'leaps' was built under R version 4.0.5

pos1data <- fullyCombined[which(fullyCombined$Position == 1), -c(1:4)]
pos1data$`Work Recovery Ratio` <- droplevels(pos1data$`Work Recovery Ratio`)
pos1data <- pos1data[, -c(38, 40)]

```

```

# Checking for correlated variables, which would cause singularities
corr1 <- cor(pos1data[, -c(11, 38)])
print("Which variables have high correlations with other variables?")

## [1] "Which variables have high correlations with other variables?"
findCorrelation(corr1, cutoff = 0.999)

## integer(0)
findCorrelation(corr1, cutoff = 0.99)

## [1] 4
findCorrelation(corr1, cutoff = 0.95)

## [1] 15 4 10
findCorrelation(corr1, cutoff = 0.9)

## [1] 15 13 8 4 10
findCorrelation(corr1, cutoff = 0.85)

## [1] 15 13 8 14 4 17 10 3 20
findCorrelation(corr1, cutoff = 0.8)

## [1] 15 13 8 14 4 17 10 3 20 24
findCorrelation(corr1, cutoff = 0.75)

## [1] 15 13 8 14 4 17 10 3 35 20 24
sort(findCorrelation(corr1, cutoff = 0.7))

## [1] 3 4 8 10 11 13 14 15 17 20 23 24 35

# Removing variables that are causing singularities
pos1data <- pos1data[, -c(3, 4, 8, 10:11, 13:15, 17, 20, 23:24, 35)]

# Performing backward stepwise selection
model1.1 <- regsubsets(margins ~ ., data = pos1data, method = "backward", nvmax = 100)
coef(model1.1, c(1:5))

## [[1]]
##           (Intercept) `Distance Speed Zone 5 (m)`
##           7.525538           1.781586
##
## [[2]]
##           (Intercept) `Distance Speed Zone 4 (m)`
##           8.2929498           -0.9094958
## `Distance Speed Zone 5 (m)`
##           9.5855343
##
## [[3]]
##           (Intercept) `Duration Speed Hi-Inten (s)`
##           8.999785           -6.832199
## `Distance Speed Zone 4 (m)` `Distance Speed Zone 5 (m)`
##           -1.420701           15.211657
##

```

```
## [[4]]
##               (Intercept) `Duration Speed Hi-Inten (s)`
##               8.329844                -7.235827
## `Distance Speed Zone 4 (m)` `Distance Speed Zone 5 (m)`
##               -1.569319                16.778622
## `Decelerations Zone 3 (num)`
##               9.289885
##
## [[5]]
##               (Intercept) `Duration Speed Hi-Inten (s)`
##               15.343848                -9.036364
## `Distance Speed Zone 4 (m)` `Distance Speed Zone 5 (m)`
##               -1.704121                17.975167
## `Accelerations Zone 3 (num)` `Decelerations Zone 3 (num)`
##               -6.722629                15.171390

# Full model
full1.1 <- lm(margins ~ . , data = pos1data)
summary(full1.1)

##
## Call:
## lm(formula = margins ~ . , data = pos1data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -42.477  -6.877  -0.117   6.844  60.150
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -5.426520   56.180152  -0.097  0.92359
## `Duration Total (s)` -0.002828    0.006207  -0.456  0.65145
## `Duration Speed Hi-Inten (s)` -40.817216  24.948925  -1.636  0.11055
## `Distance Rate (m/min)`    1.724976    1.291942   1.335  0.19020
## `Distance HR Hi-Inten (m)`    0.008725    0.012019   0.726  0.47256
## `Speed Max (km/h)`      -1.976248    2.009690  -0.983  0.33199
## `Sprints Hi-Inten (num)`    29.456847   25.198145   1.169  0.25008
## `Athlete Load`      -1.575403    1.512804  -1.041  0.30464
## `Hi Intensity Effort (num)` -0.119672    0.526072  -0.227  0.82134
## `Distance Speed Zone 1 (m)`    0.007879    0.008243   0.956  0.34550
## `Distance Speed Zone 2 (m)` -0.050515    0.091484  -0.552  0.58424
## `Distance Speed Zone 4 (m)`  -1.589990    0.836829  -1.900  0.06546 .
## `Distance Speed Zone 5 (m)`   31.574431   10.313818   3.061  0.00415 **
## `Sprints Speed Zone 5 (num)` -74.172149   46.694740  -1.588  0.12093
## `Duration HR Zone 4 (s)`    0.043991    0.036502   1.205  0.23601
## `Duration HR Zone 5 (s)`   -0.034270    0.026490  -1.294  0.20401
## `Accelerations Zone 3 (num)` -19.386848    7.580832  -2.557  0.01491 *
## `Accelerations Zone 4 (num)` -18.764728   16.262963  -1.154  0.25617
## `Accelerations Zone 5 (num)`  43.659301   45.840262   0.952  0.34723
## `Decelerations Zone 3 (num)`  26.571951   16.783568   1.583  0.12212
## `Decelerations Zone 4 (num)`  33.135837   45.078415   0.735  0.46706
## `Decelerations Zone 5 (num)`  27.068828   36.619293   0.739  0.46458
## `Body Impacts (num)`      -0.769519    0.778173  -0.989  0.32932
## `Body Impacts Grade 2 (num)`  20.218964   15.259800   1.325  0.19353
## `Body Impacts Grade 3 (num)` -30.231075   26.866546  -1.125  0.26794
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.53 on 36 degrees of freedom
## Multiple R-squared:  0.3738, Adjusted R-squared:  -0.04375
## F-statistic: 0.8952 on 24 and 36 DF,  p-value: 0.6058
# Confidence interval for the top 5 variables
confint(full1.1)[c(13, 12, 3, 20, 17), ]

##                2.5 %    97.5 %
## `Distance Speed Zone 5 (m)`    10.657039 52.4918222
## `Distance Speed Zone 4 (m)`   -3.287159  0.1071784
## `Duration Speed Hi-Inten (s)` -91.415981  9.7815485
## `Decelerations Zone 3 (num)`  -7.466704 60.6106048
## `Accelerations Zone 3 (num)` -34.761488 -4.0122082
```

These five models suggest that the most important GPS variables for a loosehead prop are, beginning from the most important:

Distance Speed Zone 5 (m) | +31.6, p-value = 0.00415

Every additional metre a loosehead prop covers in Speed ZOne 5 contributes between +10.7 and +52.5 points to the win margin

Distance Speed Zone 4 (m) | -1.59, p-value = 0.06546

Every additional metre a loosehead prop covers in Speed Zone 4 contributes between -3.29 and +0.11 points to the win margin

Duration Speed Hi-Inten (s) | -40.8, p-value = 0.11055

Every additional second a loosehead prop manages to maintain a speed beyond the high intensity speed threshold contributes between -91.4 and +9.8 points to the win margin

Decelerations Zone 3 (num) | +26.6, p-value = 0.12212

Every additional deceleration a loosehead prop performs in Deceleration Zone 3 contributes between -7.5 and +60.6 points to the win margin

Accelerations Zone 3 (num) | -19.4, p-value = 0.01491

Every additional acceleration a loosehead prop performs in Acceleration Zone 3 contributes between -34.8 and -4.0 points to the win margin

## Position 2: Hooker

```
pos2data <- fullyCombined[which(fullyCombined$Position == 2), -c(1:4)]
pos2data$`Work Recovery Ratio` <- droplevels(pos2data$`Work Recovery Ratio`)
# All zeroes for Decelerations Zone 5 (num)
pos2data <- pos2data[, -c(33, 38, 40)]

corr2 <- cor(pos2data[, -c(11, 38)])
print("Which variables have high correlations with other variables?")

## [1] "Which variables have high correlations with other variables?"
findCorrelation(corr2, cutoff = 0.999)

## integer(0)
```

```

findCorrelation(corr2, cutoff = 0.99)

## [1] 17 6
findCorrelation(corr2, cutoff = 0.98)

## [1] 17 10 6
# Removing variables that are causing singularities
pos2data <- pos2data[, -c(6, 10, 17)]

model1.2 <- regsubsets(margins ~ ., data = pos2data, method = "backward", nvmax = 100)
coef(model1.2, 1:5)

## [[1]]
##              (Intercept) `Distance Speed Zone 2 (m)`
##              5.68548764              0.02212564
##
## [[2]]
##              (Intercept)      `Distance Total (m)`
##              5.770346e+00      -6.086522e-05
## `Distance Speed Zone 2 (m)`
##              2.290739e-02
##
## [[3]]
##              (Intercept)      `Distance Total (m)`
##              4.6988531      -0.2245959
## `Distance Speed Zone 1 (m)` `Distance Speed Zone 2 (m)`
##              0.2262065              0.2785913
##
## [[4]]
##              (Intercept)      `Distance Total (m)`
##              4.9891651      -0.2824289
## `Distance Speed Zone 1 (m)` `Distance Speed Zone 2 (m)`
##              0.2842678              0.3363632
## `Distance Speed Zone 4 (m)`
##              0.3273400
##
## [[5]]
##              (Intercept)      `Distance Total (m)`
##              5.1042182      -0.6985495
## `Distance Speed Zone 1 (m)` `Distance Speed Zone 2 (m)`
##              0.7001610              0.7570799
## `Distance Speed Zone 3 (m)` `Distance Speed Zone 4 (m)`
##              0.4239357              0.7589175

full1.2 <- lm(margins ~ ., data = pos2data)
summary(full1.2)

##
## Call:
## lm(formula = margins ~ ., data = pos2data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -45.363  -7.173  -0.084   1.514  50.374

```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      15.534836  255.752542   0.061   0.952
## `Duration Total (s)`      -0.003782   0.023731  -0.159   0.875
## `Duration Speed Hi-Inten (s)` -44.921005  172.371822  -0.261   0.796
## `Duration HR Hi-Inten (s)`   -0.006848   0.086788  -0.079   0.938
## `Distance Total (m)`      -7.391956   8.126266  -0.910   0.371
## `Distance Rate (m/min)`     0.608872   3.225604   0.189   0.852
## `Speed Max (km/h)`        -0.862061   3.673213  -0.235   0.816
## `Sprints Total (num)`      -3.511722  13.987190  -0.251   0.804
## `Sprints Hi-Inten (num)`     3.293258   7.794640   0.423   0.676
## `Work Recovery Ratio`1:1    23.267959  116.381217   0.200   0.843
## `Work Recovery Ratio`2:3    29.116538  111.198668   0.262   0.795
## `Athlete Load`           8.452787   35.767273   0.236   0.815
## `Metabolic PowerPeak`      -0.215221   0.436053  -0.494   0.625
## `Hi Int Acceleration (num)` -0.089457   0.382435  -0.234   0.817
## `Hi Int Deceleration (num)`  1.600566  13.280490   0.121   0.905
## `Hi Intensity Effort (num)`  1.108994   1.172989   0.945   0.353
## `Distance Speed Zone 1 (m)`  7.396889   8.127771   0.910   0.371
## `Distance Speed Zone 2 (m)`  7.373193   8.109100   0.909   0.371
## `Distance Speed Zone 3 (m)`  7.474230   8.039358   0.930   0.360
## `Distance Speed Zone 4 (m)`  8.371263   7.928536   1.056   0.300
## `Distance Speed Zone 5 (m)`  4.390022  12.887842   0.341   0.736
## `Sprints Speed Zone 3 (num)` -6.995196   8.708268  -0.803   0.429
## `Sprints Speed Zone 4 (num)` -0.483591  31.406961  -0.015   0.988
## `Sprints Speed Zone 5 (num)` 52.648316  156.745615   0.336   0.739
## `Duration HR Zone 4 (s)`    -0.060436   0.161068  -0.375   0.710
## `Duration HR Zone 5 (s)`    -0.033535   0.129651  -0.259   0.798
## `Accelerations Zone 3 (num)` -28.042104  100.038226  -0.280   0.781
## `Accelerations Zone 4 (num)`  54.895903  195.461043   0.281   0.781
## `Accelerations Zone 5 (num)` -51.372600  138.025747  -0.372   0.713
## `Decelerations Zone 3 (num)` -13.539708  34.915285  -0.388   0.701
## `Decelerations Zone 4 (num)` 12.852524  142.988183   0.090   0.929
## `Body Impacts (num)`       0.555045   1.038093   0.535   0.597
## `Body Impacts Grade 1 (num)` -0.540235   8.141335  -0.066   0.948
## `Body Impacts Grade 2 (num)` 14.657972  64.123737   0.229   0.821
## `Body Impacts Grade 3 (num)` 22.708301  88.270686   0.257   0.799
##
```

```
## Residual standard error: 23.84 on 28 degrees of freedom
## Multiple R-squared:  0.2759, Adjusted R-squared:  -0.6033
## F-statistic: 0.3138 on 34 and 28 DF,  p-value: 0.9992
```

```
confint(full1.2)[c(18, 5, 17, 20, 19), ]
```

```
##              2.5 %    97.5 %
## `Distance Speed Zone 2 (m)` -9.237546  23.983932
## `Distance Total (m)`      -24.037857   9.253946
## `Distance Speed Zone 1 (m)` -9.252094  24.045873
## `Distance Speed Zone 4 (m)` -7.869606  24.612132
## `Distance Speed Zone 3 (m)` -8.993648  23.942108
```

These five models suggest that the most important GPS variables for a hooker are, beginning from the most important:

Distance Speed Zone 2 (m) | +7.37, p-value = 0.371

Every additional metre a hooker covers in Speed Zone 2 contributes between -9.24 and +23.98 points to the win margin

Distance Total (m) | -7.39, p-value = 0.371

Every additional metre a hooker covers in a match contributes between -24.04 and +9.25 points to the win margin

Distance Speed Zone 1 (m) | +7.40, p-value = 0.371

Every additional metre a hooker covers in Speed Zone 1 contributes between -9.25 and +24.05 points to the win margin

Distance Speed Zone 4 (m) | +8.37, p-value = 0.300

Every additional metre a hooker covers in Speed Zone 4 contributes between -7.87 and +24.61 points to the win margin

Distance Speed Zone 3 (m) | +7.47, p-value = 0.360

Every additional metre a hooker covers in Speed Zone 3 contributes between -8.99 and +23.94 points to the win margin

It should be noted that every single variable in this top 5 is a distance measure.

### Position 3: Tighthead prop

```
pos3data <- fullyCombined[which(fullyCombined$Position == 3), -c(1:4)]
pos3data$`Work Recovery Ratio` <- droplevels(pos3data$`Work Recovery Ratio`)
# All zeroes for Sprints Speed Zone 5 (num), Decelerations Zones 4 and 5 (num)
pos3data <- pos3data[, -c(25, 32, 33, 38, 40)]
```

```
corr3 <- cor(pos3data[, -c(11, 38)])
print("Which variables have high correlations with other variables?")
```

```
## [1] "Which variables have high correlations with other variables?"
```

```
findCorrelation(corr3, cutoff = 0.999)
```

```
## integer(0)
```

```
findCorrelation(corr3, cutoff = 0.99)
```

```
## [1] 4
```

```
findCorrelation(corr3, cutoff = 0.9)
```

```
## [1] 4 17 6 3 25
```

```
findCorrelation(corr3, cutoff = 0.85)
```

```
## [1] 13 15 4 17 14 8 6 3 31 18 25
```

```
findCorrelation(corr3, cutoff = 0.8)
```

```
## [1] 13 15 4 17 14 8 6 3 31 18 25 22
```

```
findCorrelation(corr3, cutoff = 0.75)
```

```
## [1] 13 15 4 17 14 8 6 3 31 18 25 22 20
```

```
findCorrelation(corr3, cutoff = 0.7)
```

```
## [1] 13 15 4 17 14 8 6 3 16 31 18 25 22 20
```



```

findCorrelation(corr3, cutoff = 0.67)

## [1] 13 15 4 17 14 8 6 3 16 31 18 25 22 20
# Removing variables that are causing singularities
pos3data <- pos3data[, -c(3, 4, 6, 8, 13:18, 20, 22, 25, 31)]

model1.3 <- regsubsets(margins ~ ., data = pos3data, method = "backward", nvmax = 100)

## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in =
## force.in, : 2 linear dependencies found

## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in =
## force.in, : nvmax reduced to 19

## Warning in rval$lopt[] <- rval$vorder[rval$lopt]: number of items to replace is
## not a multiple of replacement length

coef(model1.3, 1:5)

## [[1]]
##              (Intercept) `Sprints HR Hi-Inten (num)`
##              -13.9376737              0.5244621
##
## [[2]]
##              (Intercept) `Sprints HR Hi-Inten (num)`
##              -26.4988213              0.7900766
##      `Work Recovery Ratio`1:1
##              19.9733682
##
## [[3]]
##              (Intercept) `Sprints HR Hi-Inten (num)`
##              -19.5000646              0.6050193
##      `Work Recovery Ratio`1:1      `Work Recovery Ratio`2:3
##              18.0945307              12.5018951
##
## [[4]]
##              (Intercept)      `Duration Total (s)`
##              -12.717914218      -0.002539032
##      `Sprints HR Hi-Inten (num)`      `Work Recovery Ratio`1:1
##              0.674644569              17.547374918
##      `Work Recovery Ratio`2:3
##              15.706188017
##
## [[5]]
##              (Intercept)      `Duration Total (s)`
##              -12.782083771      -0.002693672
##      `Sprints HR Hi-Inten (num)`      `Work Recovery Ratio`1:1
##              0.678885124              17.991289338
##      `Work Recovery Ratio`2:3 `Accelerations Zone 5 (num)`
##              16.378585940              17.180609903

full1.3 <- lm(margins ~ ., data = pos3data)
summary(full1.3)

##
## Call:

```

```
## lm(formula = margins ~ ., data = pos3data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -35.279  -7.221   0.000   2.231  50.881
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -8.471e+02  1.122e+03  -0.755   0.458
## `Duration Total (s)` -2.275e-03  4.386e-03  -0.519   0.609
## `Duration Speed Hi-Inten (s)` 3.612e+01  4.034e+01   0.895   0.380
## `Distance Rate (m/min)`  8.696e+00  1.122e+01   0.775   0.446
## `Speed Max (km/h)`    -1.333e+00  2.153e+00  -0.619   0.542
## `Sprints Hi-Inten (num)` -6.321e+00  7.782e+00  -0.812   0.425
## `Sprints HR Hi-Inten (num)` 1.048e+00  1.206e+00   0.869   0.394
## `Work Recovery Ratio`1:1 3.688e+02  4.682e+02   0.788   0.439
## `Work Recovery Ratio`2:3 4.692e+02  5.886e+02   0.797   0.434
## `Athlete Load`       9.681e+00  1.266e+01   0.765   0.453
## `Distance Speed Zone 2 (m)` -1.348e-02  8.148e-02  -0.165   0.870
## `Distance Speed Zone 4 (m)` -2.985e-01  9.907e-01  -0.301   0.766
## `Sprints Speed Zone 3 (num)` 3.075e+00  3.675e+00   0.837   0.412
## `Sprints Speed Zone 4 (num)` 2.904e+00  2.127e+01   0.137   0.893
## `Duration HR Zone 5 (s)`    -3.295e-01  4.302e-01  -0.766   0.452
## `Accelerations Zone 3 (num)` -1.973e+01  3.998e+01  -0.494   0.626
## `Accelerations Zone 4 (num)` 5.431e+01  6.870e+01   0.791   0.438
## `Accelerations Zone 5 (num)` 9.696e+01  1.123e+02   0.864   0.397
## `Decelerations Zone 3 (num)` -5.782e+00  2.411e+01  -0.240   0.813
## `Body Impacts Grade 1 (num)` 4.269e+00  5.955e+00   0.717   0.481
## `Body Impacts Grade 2 (num)`      NA         NA      NA      NA
## `Body Impacts Grade 3 (num)`      NA         NA      NA      NA
##
## Residual standard error: 21.25 on 22 degrees of freedom
## Multiple R-squared:  0.2738, Adjusted R-squared:  -0.3534
## F-statistic: 0.4366 on 19 and 22 DF,  p-value: 0.9637
```

Work Recovery Ratio has two dummy variables represented in the five-variable model. We will look into larger models until a fifth non-dummy variable is found.

```
coef(model1.3, 6)
```

```
##              (Intercept)          `Duration Total (s)`
##          6.113504032          -0.001945393
##          `Speed Max (km/h)`    `Sprints HR Hi-Inten (num)`
##        -0.851484034          0.680664324
##          `Work Recovery Ratio`1:1    `Work Recovery Ratio`2:3
##          17.899975027          16.459506300
##          `Accelerations Zone 5 (num)`
##          18.153716278
```

```
confint(full1.3)[c(7:9, 2, 18, 5), ]
```

```
##              2.5 %          97.5 %
## `Sprints HR Hi-Inten (num)` -1.45364011 3.549925e+00
## `Work Recovery Ratio`1:1 -602.16169440 1.339696e+03
## `Work Recovery Ratio`2:3 -751.48413990 1.689807e+03
## `Duration Total (s)` -0.01137205 6.821670e-03
```

```
## `Accelerations Zone 5 (num)` -135.88164475 3.298111e+02
## `Speed Max (km/h)` -5.79874677 3.132361e+00
```

These six models suggest that the most important GPS variables for a tighthead prop are, beginning from the most important:

Sprints HR Hi-Inten (num) | +1.05, p-value = 0.394

Every additional sprint a tighthead prop performs while over the high intensity HR benchmark contributes between -1.45 and +3.55 points to the win margin

Work Recovery Ratio | 1:1 -> +369, p-value = 0.439; 2:3 -> +469, p-value = 0.434

If a tighthead prop has a Work Recovery Ratio of 1:1 in a match, it will contribute between -602 and +1340 points(!) to the win margin

If a tighthead prop has a Work Recovery Ratio of 2:3 in a match, it will contribute between -751 and +1690 points(!) to the win margin

Duration Total (s) | -0.00228, p-value = 0.609

Every additional second a tighthead prop plays in a match contributes between -0.01137 and +0.00682 points to the win margin

Accelerations Zone 5 (num) | +97.0, p-value = 0.397

Every additional acceleration a tighthead prop performs in Acceleration Zone 5 contributes between -135.9 and +329.8 points(!) to the win margin

Speed Max (km/h) | -1.33, p-value = 0.542

For every kilometre per hour in a tighthead prop's maximum speed in a match, between -5.80 and +3.13 points are added to the win margin

The tighthead prop data contains some surprisingly high coefficient magnitudes. This may be due to the number of variables removed from the dataset to deal with singularities in the full model. Even then, some singularities remain, but the correlation cutoff has reached close to 0.5, and the offending variables were not singled out for removal.

#### Position 4: Left lock

```
pos4data <- fullyCombined[which(fullyCombined$Position == 4), -c(1:4)]
pos4data$`Work Recovery Ratio` <- droplevels(pos4data$`Work Recovery Ratio`)
# All zeroes in Duration Speed Hi-Inten (s)
pos4data <- pos4data[, -c(2, 38, 40)]
```

```
corr4 <- cor(pos4data[, -c(10, 37)])
print("Which variables have high correlations with other variables?")
```

```
## [1] "Which variables have high correlations with other variables?"
```

```
findCorrelation(corr4, cutoff = 0.999)
```

```
## integer(0)
```

```
findCorrelation(corr4, cutoff = 0.9)
```

```
## [1] 5 10 2 16
```

```
findCorrelation(corr4, cutoff = 0.8)
```

```
## [1] 14 5 9 10 2 12 16 1 15 20
```

```

findCorrelation(corr4, cutoff = 0.75)

## [1] 14 5 9 13 10 2 12 16 1 15 8 23
sort(findCorrelation(corr4, cutoff = 0.7))

## [1] 1 2 5 6 8 9 10 12 13 14 15 16 19 20 26
# Removing variables that are causing singularities
pos4data <- pos4data[, -c(1:2, 5:6, 8:10, 12:16, 19, 20, 26)]

model1.4 <- regsubsets(margins ~ ., data = pos4data, method = "backward", nvmax = 100)
coef(model1.4, 1:5)

## [[1]]
##              (Intercept) `Distance Speed Zone 1 (m)`
##              2.465755135                0.001739778
##
## [[2]]
##              (Intercept)          `Distance Total (m)`
##              6.22571438                -0.04932264
## `Distance Speed Zone 1 (m)`
##              0.05308774
##
## [[3]]
##              (Intercept)          `Distance Total (m)`
##              11.5024133                -0.2077127
## `Distance Speed Zone 1 (m)` `Distance Speed Zone 2 (m)`
##              0.2098620                0.2232217
##
## [[4]]
##              (Intercept)          `Distance Total (m)`
##              34.3814877                -0.2236625
## `Athlete Load` `Distance Speed Zone 1 (m)`
##              -0.7758081                0.2274910
## `Distance Speed Zone 2 (m)`
##              0.2391613
##
## [[5]]
##              (Intercept)          `Distance Total (m)`
##              38.0670914                -0.2550060
## `Sprints Total (num)`          `Athlete Load`
##              0.7259407                -2.7620887
## `Distance Speed Zone 1 (m)` `Distance Speed Zone 2 (m)`
##              0.2585583                0.2777415

full1.4 <- lm(margins ~ ., data = pos4data)
summary(full1.4)

##
## Call:
## lm(formula = margins ~ ., data = pos4data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.207  -7.024   0.000   4.924  42.553

```

```
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.59856   93.22765  -0.028  0.9781
## `Distance Total (m)`      -0.59058    0.26881  -2.197  0.0441 *
## `Distance Rate (m/min)`    -0.36264    1.85253  -0.196  0.8474
## `Sprints Total (num)`      1.91363    2.67994   0.714  0.4862
## `Athlete Load`      -5.55956    8.85413  -0.628  0.5395
## `Distance Speed Zone 1 (m)`  0.59432    0.26703   2.226  0.0418 *
## `Distance Speed Zone 2 (m)`  0.64347    0.33346   1.930  0.0728 .
## `Distance Speed Zone 5 (m)`  1.86171    4.89226   0.381  0.7089
## `Sprints Speed Zone 3 (num)` 10.57929    6.15163   1.720  0.1060
## `Sprints Speed Zone 4 (num)` 10.72129   18.26691   0.587  0.5660
## `Sprints Speed Zone 5 (num)` -6.22209   44.44627  -0.140  0.8905
## `Duration HR Zone 4 (s)`     0.08332    0.09507   0.876  0.3946
## `Accelerations Zone 3 (num)` -5.04406   53.02607  -0.095  0.9255
## `Accelerations Zone 4 (num)` -154.14943 201.53791  -0.765  0.4562
## `Accelerations Zone 5 (num)`  47.72766 292.87314   0.163  0.8727
## `Decelerations Zone 3 (num)` 13.02749   43.84730   0.297  0.7705
## `Decelerations Zone 4 (num)`  2.35785   67.25718   0.035  0.9725
## `Decelerations Zone 5 (num)` 119.34175 113.48768   1.052  0.3096
## `Body Impacts (num)`       -0.16226    0.88835  -0.183  0.8575
## `Body Impacts Grade 1 (num)` -12.40847   16.38490  -0.757  0.4606
## `Body Impacts Grade 2 (num)`  34.76380   29.72792   1.169  0.2605
## `Body Impacts Grade 3 (num)` -50.82777 129.68372  -0.392  0.7006
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.48 on 15 degrees of freedom
## Multiple R-squared:  0.4611, Adjusted R-squared:  -0.2934
## F-statistic: 0.6112 on 21 and 15 DF,  p-value: 0.8534
confint(full1.4)[c(6, 2, 7, 5, 4), ]

##              2.5 %      97.5 %
## `Distance Speed Zone 1 (m)`  0.02515619  1.16347882
## `Distance Total (m)`      -1.16353511 -0.01762577
## `Distance Speed Zone 2 (m)` -0.06727843  1.35421626
## `Athlete Load`           -24.43168059 13.31256377
## `Sprints Total (num)`      -3.79852686  7.62577844
```

These five models suggest that the most important GPS variables for a left lock are, beginning from the most important:

Distance Speed Zone 1 (m) | +0.594, p-value = 0.0418

Every additional metre a left lock covers in Speed Zone 1 contributes between +0.025 and +1.163 to the win margin

Distance Total (m) | -0.591, p-value = 0.0441

Every additional metre a left lock covers in the match contributes between -1.164 and -0.018 points to the win margin

Distance Speed Zone 2 (m) | +0.643, p-value = 0.0728

Every additional metre a left lock covers in Speed Zone 2 contributes between -0.067 and +1.354 points to the win margin

Athlete Load | -5.56, p-value = 0.5395

Every additional point in Athlete Load a left lock has contributes between -24.43 and +13.31 points to the win margin

Sprints Total (num) | +1.91, p-value = 0.4862

Every additional sprint a left lock performs in the match contributes between -3.80 and +7.63 points to the win margin

### Position 5: Right lock

```
pos5data <- fullyCombined[which(fullyCombined$Position == 5), -c(1:4)]
pos5data$`Work Recovery Ratio` <- droplevels(pos5data$`Work Recovery Ratio`)

pos5data <- pos5data[, -c(38, 40)]

corr5 <- cor(pos5data[, -c(11, 38)])
print("Which variables have high correlations with other variables?")

## [1] "Which variables have high correlations with other variables?"
findCorrelation(corr5, cutoff = 0.999)

## [1] 2
findCorrelation(corr5, cutoff = 0.9)

## [1] 11 10 13 4 1 21 2 26
findCorrelation(corr5, cutoff = 0.8)

## [1] 14 15 11 8 10 13 4 1 17 30 21 2 28 26 29 16
findCorrelation(corr5, cutoff = 0.75)

## [1] 14 15 11 8 10 13 4 1 17 30 25 18 21 2 28 35 26 5
# Removing variables that are causing singularities
pos5data <- pos5data[, -c(1:2, 4:5, 8, 10:11, 13:15, 17:18, 21, 25:26, 28:30, 35)]

model1.5 <- regsubsets(margins ~ ., data = pos5data, method = "backward", nvmax = 100)
coef(model1.5, 1:5)

## [[1]]
##           (Intercept) `Sprints Hi-Inten (num)`
##           6.489683           2.300478
##
## [[2]]
##           (Intercept)           `Sprints Hi-Inten (num)`
##           11.822684           2.593344
## `Sprints Speed Zone 3 (num)`
##           -3.105786
##
## [[3]]
##           (Intercept)           `Sprints Hi-Inten (num)`
##           12.165203           2.607797
## `Sprints Speed Zone 3 (num)` `Decelerations Zone 3 (num)`
##           -3.521746           6.300704
##
```

```
## [[4]]
##               (Intercept)      `Sprints Hi-Inten (num)`
##               10.5233007          2.5465385
## `Sprints Speed Zone 3 (num)` `Decelerations Zone 3 (num)`
##               -3.5100896          6.6323986
## `Body Impacts Grade 2 (num)`
##               0.8980811
##
## [[5]]
##               (Intercept)      `Sprints Hi-Inten (num)`
##               15.11427525        2.49155987
## `Hi Intensity Effort (num)` `Sprints Speed Zone 3 (num)`
##               -0.08109202        -3.46800686
## `Decelerations Zone 3 (num)` `Body Impacts Grade 2 (num)`
##               11.99592290        2.14289631
```

```
full1.5 <- lm(margins ~ ., data = pos5data)
summary(full1.5)
```

```
##
## Call:
## lm(formula = margins ~ ., data = pos5data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -33.335  -6.446   0.000   3.553  50.444
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -25.56093    49.27285  -0.519   0.609
## `Duration HR Hi-Inten (s)`    -0.12715     0.13225  -0.961   0.347
## `Distance HR Hi-Inten (m)`    -0.01911     0.07788  -0.245   0.808
## `Speed Max (km/h)`           3.50994     2.81116   1.249   0.225
## `Sprints Hi-Inten (num)`       5.65231     3.71757   1.520   0.143
## `Athlete Load`              -0.11656    10.06454  -0.012   0.991
## `Hi Intensity Effort (num)`    -0.67773     2.24114  -0.302   0.765
## `Distance Speed Zone 2 (m)`     0.03833     0.08167   0.469   0.643
## `Distance Speed Zone 3 (m)`    -0.06092     0.30940  -0.197   0.846
## `Distance Speed Zone 5 (m)`    -8.32163    12.02872  -0.692   0.496
## `Sprints Speed Zone 3 (num)`   -5.97488     5.58191  -1.070   0.296
## `Sprints Speed Zone 4 (num)`   -3.08603    13.85653  -0.223   0.826
## `Duration HR Zone 5 (s)`       0.14001     0.16442   0.852   0.404
## `Decelerations Zone 3 (num)`  163.20562   226.57514   0.720   0.479
## `Decelerations Zone 4 (num)`   19.25228    44.71754   0.431   0.671
## `Decelerations Zone 5 (num)`  -25.37264    42.58889  -0.596   0.557
## `Body Impacts (num)`          -0.21834     0.47689  -0.458   0.652
## `Body Impacts Grade 2 (num)`   25.73845    15.45601   1.665   0.110
## `Body Impacts Grade 3 (num)` -37.36833    39.21446  -0.953   0.351
##
## Residual standard error: 20.03 on 22 degrees of freedom
## Multiple R-squared:  0.3437, Adjusted R-squared:  -0.1932
## F-statistic: 0.6402 on 18 and 22 DF,  p-value: 0.8299
```

```
confint(full1.5)[c(5, 11, 14, 18, 7), ]
```

```
##                2.5 %    97.5 %
## `Sprints Hi-Inten (num)`    -2.057467  13.362089
## `Sprints Speed Zone 3 (num)` -17.551048   5.601294
## `Decelerations Zone 3 (num)` -306.682470 633.093705
## `Body Impacts Grade 2 (num)`  -6.315342  57.792247
## `Hi Intensity Effort (num)`  -5.325564   3.970097
```

These five models suggest that the most important GPS variables for a right lock are, beginning from the most important:

Sprints Hi-Inten (num) | +5.65, p-value = 0.143

Every additional sprint a right lock performs above the high intensity sprint benchmark contributes between -2.06 and +13.36 points to the win margin

Sprints Speed Zone 3 (num) | -5.97, p-value = 0.296

Every additional sprint a right lock performs in Speed Zone 3 contributes between -17.55 and +5.60 points to the win margin

Decelerations Zone 3 (num) | +163, p-value = 0.479

Every additional deceleration a right lock performs in Deceleration Zone 3 contributes between -306 and +633 points(!) to the win margin

Body Impacts Grade 2 (num) | +25.7, p-value = 0.110

Every additional Grade 2 body impact a right lock performs contributes between -6.3 and +57.8 points to the win margin

Hi Intensity Effort (num) | -0.678, p-value = 0.765

For every additional effort a right lock performs that falls under any of the five high intensity categories (Hi-Int Sprints, Hi-Int Accelerations, Hi-Int Decelerations, Body Impacts and Jumps), between -5.326 and +3.970 points are added to the win margin

## Position 6: Blindside flanker

```
pos6data <- fullyCombined[which(fullyCombined$Position == 6), -c(1:4)]
pos6data$`Work Recovery Ratio` <- droplevels(pos6data$`Work Recovery Ratio`)
# All zeroes in Duration Speed Hi-Inten (s)
pos6data <- pos6data[, -c(2, 38, 40)]
```

```
corr6 <- cor(pos6data[, -c(10, 37)])
print("Which variables have high correlations with other variables?")
```

```
## [1] "Which variables have high correlations with other variables?"
```

```
findCorrelation(corr6, cutoff = 0.999)
```

```
## integer(0)
```

```
findCorrelation(corr6, cutoff = 0.92)
```

```
## [1] 14  3 16  9  5 20
```

```
findCorrelation(corr6, cutoff = 0.9)
```

```
## [1] 12 14  3 16  7  9  5 20
```



```
# Removing variables that are causing singularities
pos6data <- pos6data[, -c(3, 5, 7, 9, 12, 14, 16)]

modell1.6 <- regsubsets(margins ~ ., data = pos6data, method = "backward", nvmax = 100)
coef(modell1.6, 1:5)
```

```
## [[1]]
##              (Intercept) `Distance Speed Zone 2 (m)`
##              0.99671035              0.03551721
##
## [[2]]
##              (Intercept) `Distance Speed Zone 2 (m)`
##              0.8588275              0.0299237
## `Sprints Speed Zone 4 (num)`
##              3.2739498
##
## [[3]]
##              (Intercept)      `Speed Max (km/h)`
##              40.67925201      -1.69055333
## `Distance Speed Zone 2 (m)` `Sprints Speed Zone 4 (num)`
##              0.04800387              6.96538479
##
## [[4]]
##              (Intercept)      `Speed Max (km/h)`
##              43.86317544      -1.98214782
## `Distance Speed Zone 2 (m)` `Sprints Speed Zone 4 (num)`
##              0.06328289              6.61434614
## `Decelerations Zone 4 (num)`
##              8.40690457
##
## [[5]]
##              (Intercept)      `Speed Max (km/h)`
##              55.63305072      -2.60934396
## `Work Recovery Ratio`2:3 `Distance Speed Zone 2 (m)`
##              -18.36474346              0.08234539
## `Sprints Speed Zone 4 (num)` `Decelerations Zone 4 (num)`
##              9.05761652              21.07595360
```

```
full1.6 <- lm(margins ~ ., data = pos6data)
summary(full1.6)
```

```
##
## Call:
## lm(formula = margins ~ ., data = pos6data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.617  -5.922   0.000   0.303  34.368
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      64.08059   389.30573    0.165   0.873
## `Duration Total (s)`    -0.01367    0.03815   -0.358   0.728
## `Duration HR Hi-Inten (s)`  0.01164    0.17322    0.067   0.948
```

```
## `Distance Rate (m/min)`      0.94043    3.52754    0.267    0.796
## `Speed Max (km/h)`          -5.65093    10.64625   -0.531    0.608
## `Sprints Hi-Inten (num)`     -4.14389   117.50606   -0.035    0.973
## `Work Recovery Ratio`1:1     -27.08416   103.84198   -0.261    0.800
## `Work Recovery Ratio`2:3     -62.35147   163.11759   -0.382    0.711
## `Athlete Load`              2.65298    12.31436    0.215    0.834
## `Hi Int Acceleration (num)`  -0.33242    0.73329   -0.453    0.661
## `Hi Intensity Effort (num)`  -1.12333    2.86426   -0.392    0.704
## `Distance Speed Zone 1 (m)`   0.02385    0.04780    0.499    0.630
## `Distance Speed Zone 2 (m)`   0.26312    0.30778    0.855    0.415
## `Distance Speed Zone 3 (m)`  -0.27803    0.71377   -0.390    0.706
## `Distance Speed Zone 4 (m)`  -0.57575    1.26532   -0.455    0.660
## `Distance Speed Zone 5 (m)`   3.92417    7.15445    0.548    0.597
## `Sprints Speed Zone 3 (num)`   4.43557    9.95313    0.446    0.666
## `Sprints Speed Zone 4 (num)`  18.82721   29.31145    0.642    0.537
## `Sprints Speed Zone 5 (num)` -33.31820   94.57492   -0.352    0.733
## `Duration HR Zone 4 (s)`      -0.04629    0.28912   -0.160    0.876
## `Duration HR Zone 5 (s)`      0.05442    0.15877    0.343    0.740
## `Accelerations Zone 3 (num)`   4.30610   30.98646    0.139    0.893
## `Accelerations Zone 4 (num)`   0.39055   44.67241    0.009    0.993
## `Accelerations Zone 5 (num)` -46.88160  143.98839   -0.326    0.752
## `Decelerations Zone 3 (num)`  12.27514   115.01552    0.107    0.917
## `Decelerations Zone 4 (num)`  31.71683   133.55584    0.237    0.818
## `Decelerations Zone 5 (num)` -97.97550   599.70295   -0.163    0.874
## `Body Impacts (num)`          -0.28300    1.53153   -0.185    0.857
## `Body Impacts Grade 1 (num)` -1.96159   11.25798   -0.174    0.866
## `Body Impacts Grade 2 (num)`  15.86879   75.13676    0.211    0.837
## `Body Impacts Grade 3 (num)`  44.07383   214.87403    0.205    0.842
```

```
##
```

```
## Residual standard error: 30.11 on 9 degrees of freedom
```

```
## Multiple R-squared: 0.4689, Adjusted R-squared: -1.301
```

```
## F-statistic: 0.2649 on 30 and 9 DF, p-value: 0.9972
```

```
confint(full1.6)[c(13, 18, 5, 26, 8), ]
```

```
##              2.5 %      97.5 %
## `Distance Speed Zone 2 (m)`  -0.4331233  0.9593664
## `Sprints Speed Zone 4 (num)` -47.4798847  85.1343144
## `Speed Max (km/h)`          -29.7344066  18.4325556
## `Decelerations Zone 4 (num)` -270.4074834  333.8411371
## `Work Recovery Ratio`2:3     -431.3490992  306.6461538
```

These five models suggest that the most important GPS variables for a blindside flanker are, beginning from the most important:

Distance Speed Zone 2 (m) | +0.263, p-value = 0.415

Every additional metre a blindside flanker covers in Speed Zone 2 contributes between -0.433 and +0.959 points to the win margin

Sprints Speed Zone 4 (num) | +18.8, p-value = 0.537

Every additional sprint a blindside flanker performs in Speed Zone 4 contributes between -47.5 and +85.1 points to the win margin

Speed Max (km/h) | -5.65, p-value = 0.608

For every kilometre per hour in a blindside flanker's maximum speed in a match, between -29.73 and +18.43

points are added to the win margin

Decelerations Zone 4 (num) | +31.7, p-value = 0.818

Every additional deceleration a blindside flanker performs in Deceleration Zone 4 contributes between -270.4 and +333.8 points(!) to the win margin

Work Recovery Ratio | 2:3 -> -62.4, p-value = 0.711

If a blindside flanker has a Work Recovery Ratio of 2:3 in a match, it will contribute between -431.3 and +306.6 points(!) to the win margin

### Position 7: Openside flanker

```
pos7data <- fullyCombined[which(fullyCombined$Position == 7), -c(1:4)]
pos7data$`Work Recovery Ratio` <- droplevels(pos7data$`Work Recovery Ratio`)

pos7data <- pos7data[, -c(38, 40)]

corr7 <- cor(pos7data[, -c(11, 38)])
print("Which variables have high correlations with other variables?")

## [1] "Which variables have high correlations with other variables?"
findCorrelation(corr7, cutoff = 0.999)

## integer(0)
findCorrelation(corr7, cutoff = 0.9)

## [1] 8 10 15 6 5 13 4
findCorrelation(corr7, cutoff = 0.85)

## [1] 8 10 15 6 5 11 13 16 4 18 2
findCorrelation(corr7, cutoff = 0.8)

## [1] 8 10 15 6 5 11 13 16 4 17 18 2
findCorrelation(corr7, cutoff = 0.77)

## [1] 8 10 15 6 5 11 3 14 13 16 4 17 18 2 24
# Removing variables that are causing singularities
pos7data <- pos7data[, -c(2:6, 8, 10:11, 13, 15:18)]

model1.7 <- regsubsets(margins ~ ., data = pos7data, method = "backward", nvmax = 100)
coef(model1.7, 1:5)

## [[1]]
##              (Intercept) `Accelerations Zone 5 (num)`
##              5.496038              19.120534
##
## [[2]]
##              (Intercept) `Sprints Speed Zone 4 (num)`
##              10.228878              -5.136992
## `Accelerations Zone 5 (num)`
##              22.341665
##
## [[3]]
```

```
##              (Intercept) `Sprints Speed Zone 4 (num)`
##              10.653936          -6.593079
## `Accelerations Zone 5 (num)` `Body Impacts Grade 3 (num)`
##              21.352090          8.255397
##
## [[4]]
##              (Intercept) `Sprints Speed Zone 4 (num)`
##              7.473199414          -6.654240283
## `Duration HR Zone 4 (s)` `Accelerations Zone 5 (num)`
##              0.004092438          21.496297679
## `Body Impacts Grade 3 (num)`
##              9.074061285
##
## [[5]]
##              (Intercept) `Sprints Speed Zone 4 (num)`
##              10.78048198          -7.53364051
## `Duration HR Zone 4 (s)` `Accelerations Zone 5 (num)`
##              0.01447122          23.53992544
## `Body Impacts Grade 2 (num)` `Body Impacts Grade 3 (num)`
##              -3.02034395          15.36352594
```

```
full1.7 <- lm(margins ~ ., data = pos7data)
summary(full1.7)
```

```
##
## Call:
## lm(formula = margins ~ ., data = pos7data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.273  -2.839   0.000   4.077  25.770
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -5.499e+01  2.000e+02  -0.275  0.7877
## `Duration Total (s)`    6.982e-04  4.703e-03   0.148  0.8843
## `Speed Max (km/h)`    -1.302e-01  5.848e+00  -0.022  0.9826
## `Sprints Hi-Inten (num)`  3.021e+02  2.131e+02   1.418  0.1797
## `Athlete Load`    1.562e+00  3.689e+00   0.423  0.6789
## `Hi Int Acceleration (num)`  2.719e-01  2.965e-01   0.917  0.3758
## `Distance Speed Zone 2 (m)` -7.179e-02  7.242e-02  -0.991  0.3397
## `Distance Speed Zone 3 (m)`  1.520e-01  1.592e-01   0.955  0.3571
## `Distance Speed Zone 4 (m)`  1.156e+00  5.183e-01   2.230  0.0440 *
## `Distance Speed Zone 5 (m)`  1.357e-01  2.311e+00   0.059  0.9541
## `Sprints Speed Zone 3 (num)` -1.612e+00  4.013e+00  -0.402  0.6944
## `Sprints Speed Zone 4 (num)` -3.198e+01  1.204e+01  -2.657  0.0198 *
## `Sprints Speed Zone 5 (num)` -1.955e+01  2.550e+01  -0.767  0.4569
## `Duration HR Zone 4 (s)`    4.671e-01  4.276e-01   1.092  0.2945
## `Duration HR Zone 5 (s)`   -1.647e-01  1.480e-01  -1.113  0.2859
## `Accelerations Zone 3 (num)`  1.143e+02  9.746e+01   1.173  0.2617
## `Accelerations Zone 4 (num)` -2.642e+02  2.042e+02  -1.294  0.2183
## `Accelerations Zone 5 (num)`  1.024e+03  7.600e+02   1.348  0.2008
## `Decelerations Zone 3 (num)` -1.700e+02  1.246e+02  -1.364  0.1956
## `Decelerations Zone 4 (num)`  4.809e+02  3.919e+02   1.227  0.2416
## `Decelerations Zone 5 (num)` -6.021e+02  4.789e+02  -1.257  0.2308
```

```
## `Body Impacts (num)` -1.347e+00 9.164e-01 -1.470 0.1653
## `Body Impacts Grade 1 (num)` -2.552e+01 2.128e+01 -1.199 0.2519
## `Body Impacts Grade 2 (num)` -1.504e+02 1.190e+02 -1.264 0.2286
## `Body Impacts Grade 3 (num)` 3.437e+02 2.811e+02 1.223 0.2431
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.55 on 13 degrees of freedom
## Multiple R-squared: 0.6245, Adjusted R-squared: -0.06875
## F-statistic: 0.9008 on 24 and 13 DF, p-value: 0.603
confint(full1.7)[c(18, 12, 25, 14, 24), ]

## 2.5 % 97.5 %
## `Accelerations Zone 5 (num)` -617.6489128 2666.176686
## `Sprints Speed Zone 4 (num)` -57.9887750 -5.973988
## `Body Impacts Grade 3 (num)` -263.5698306 951.002358
## `Duration HR Zone 4 (s)` -0.4566217 1.390908
## `Body Impacts Grade 2 (num)` -407.6081775 106.759426
```

These five models suggest that the most important GPS variables for an openside flanker are, beginning from the most important:

Accelerations Zone 5 (num) | +1024, p-value = 0.2008

Every additional acceleration an openside flanker performs in Acceleration Zone 5 contributes between -617 and 2666 points(!) to the win margin

Sprints Speed Zone 4 (num) | -32.0, p-value = 0.0198

Every additional sprint an openside flanker performs in Speed Zone 4 contributes between -58.0 and -6.0 points(!) to the win margin

Body Impacts Grade 3 (num) | +344, p-value = 0.2431

Every additional Grade 3 body impact an openside flanker performs contributes between -264 and +951 points(!) to the win margin

Duration HR Zone 4 (s) | +0.467, p-value = 0.2945

Every additional second an openside flanker spends in HR Zone 4 contributes between -0.457 and +1.391 points to the win margin

Body Impacts Grade 2 (num) | -150, p-value = 0.2286

Every additional Grade 2 body impact an openside flanker performs contributes -408 and +107 points(!) to the win margin

Like the tighthead prop model, there are coefficients here that are also very large in magnitude. Singularities have been removed here, so this could be due to some variables being mostly zero-valued, but with a very small portion of non-zero values that are associated with a large-magnitude win margin.

## Position 8: Number 8

```
pos8data <- fullyCombined[which(fullyCombined$Position == 8), -c(1:4)]
pos8data$`Work Recovery Ratio` <- droplevels(pos8data$`Work Recovery Ratio`)
# All zeroes for Duration Speed Hi-Inten (s) and Body Impacts Grade 3 (num)
pos8data <- pos8data[, -c(2, 37, 38, 40)]
```

```

corr8 <- cor(pos8data[, -c(10, 36)])
print("Which variables have high correlations with other variables?")

## [1] "Which variables have high correlations with other variables?"
findCorrelation(corr8, cutoff = 0.999)

## integer(0)
findCorrelation(corr8, cutoff = 0.9)

## [1] 3 16 7 14 15 5 25 9
# Removing variables that are causing singularities
pos8data <- pos8data[, -c(3, 5, 7, 14:16)]

modell1.8 <- regsubsets(margins ~ ., data = pos8data, method = "backward", nvmax = 100)
coef(modell1.8, 1:5)

## [[1]]
##                (Intercept) `Duration HR Zone 4 (s)`
##                1.25938675          0.01260577
##
## [[2]]
##                (Intercept) `Hi Int Acceleration (num)`
##                4.03900132          -0.08307131
## `Duration HR Zone 4 (s)`
##                0.01654467
##
## [[3]]
##                (Intercept) `Hi Int Acceleration (num)`
##                1.03797460          -0.23760140
## `Distance Speed Zone 2 (m)` `Duration HR Zone 4 (s)`
##                0.06262507          0.01991303
##
## [[4]]
##                (Intercept) `Hi Int Acceleration (num)`
##                -1.13461300          -0.27809141
## `Distance Speed Zone 2 (m)` `Sprints Speed Zone 3 (num)`
##                0.06085583          2.11998124
## `Duration HR Zone 4 (s)`
##                0.02128914
##
## [[5]]
##                (Intercept) `Hi Int Acceleration (num)`
##                -2.24094283          -0.32225671
## `Distance Speed Zone 2 (m)` `Distance Speed Zone 3 (m)`
##                0.11888706          -0.23751923
## `Sprints Speed Zone 3 (num)` `Duration HR Zone 4 (s)`
##                5.07451765          0.02057206

full1.8 <- lm(margins ~ ., data = pos8data)
summary(full1.8)

##
## Call:
## lm(formula = margins ~ ., data = pos8data)

```

```

##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -37.269  -4.322   0.000   0.000  42.834
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.387e+01  1.503e+02   0.225   0.8248
## `Duration Total (s)` -9.322e-04  7.143e-03  -0.131   0.8979
## `Duration HR Hi-Inten (s)`  1.178e-02  1.474e-01   0.080   0.9374
## `Distance Rate (m/min)` -8.985e-01  2.241e+00  -0.401   0.6942
## `Speed Max (km/h)`      1.281e+00  2.136e+00   0.600   0.5578
## `Sprints Hi-Inten (num)` -5.043e-01  2.572e+01  -0.020   0.9846
## `Sprints HR Hi-Inten (num)`  1.449e+00  4.396e+00   0.330   0.7462
## `Work Recovery Ratio`1:1 -2.793e+01  4.504e+01  -0.620   0.5446
## `Work Recovery Ratio`1:2  1.045e+01  8.068e+01   0.129   0.8987
## `Work Recovery Ratio`2:3 -5.607e+01  6.083e+01  -0.922   0.3712
## `Work Recovery Ratio`3:1  1.709e+01  7.820e+01   0.219   0.8300
## `Athlete Load`        -1.477e+00  2.869e+00  -0.515   0.6143
## `Metabolic PowerPeak`    -4.990e-02  2.672e-01  -0.187   0.8543
## `Hi Int Acceleration (num)` -4.250e-01  4.480e-01  -0.949   0.3578
## `Distance Speed Zone 1 (m)` -4.103e-03  1.050e-02  -0.391   0.7015
## `Distance Speed Zone 2 (m)`  1.881e-01  1.019e-01   1.846   0.0847 .
## `Distance Speed Zone 3 (m)` -8.599e-01  3.923e-01  -2.192   0.0446 *
## `Distance Speed Zone 4 (m)` -1.281e-01  1.606e+00  -0.080   0.9374
## `Distance Speed Zone 5 (m)`  6.645e-01  2.674e+00   0.249   0.8071
## `Sprints Speed Zone 3 (num)`  1.850e+01  8.551e+00   2.164   0.0470 *
## `Sprints Speed Zone 4 (num)`  6.910e+00  1.920e+01   0.360   0.7240
## `Sprints Speed Zone 5 (num)`  1.649e+00  3.113e+01   0.053   0.9585
## `Duration HR Zone 4 (s)`    4.982e-02  9.430e-02   0.528   0.6050
## `Duration HR Zone 5 (s)`   -3.683e-02  2.371e-01  -0.155   0.8786
## `Accelerations Zone 3 (num)`  1.002e+01  2.575e+01   0.389   0.7028
## `Accelerations Zone 4 (num)` -9.517e+00  7.739e+01  -0.123   0.9038
## `Accelerations Zone 5 (num)` -3.270e+00  6.601e+01  -0.050   0.9611
## `Decelerations Zone 3 (num)`  1.654e+01  3.336e+01   0.496   0.6273
## `Decelerations Zone 4 (num)`  1.678e+01  7.200e+01   0.233   0.8189
## `Decelerations Zone 5 (num)` -3.328e+00  1.691e+02  -0.020   0.9846
## `Body Impacts (num)`       8.569e-01  1.157e+00   0.741   0.4702
## `Body Impacts Grade 1 (num)`  6.407e-01  1.264e+01   0.051   0.9602
## `Body Impacts Grade 2 (num)`  9.196e-01  9.721e+00   0.095   0.9259
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.31 on 15 degrees of freedom
## Multiple R-squared:  0.5471, Adjusted R-squared:  -0.419
## F-statistic: 0.5663 on 32 and 15 DF,  p-value: 0.9132
confint(full1.8)[c(23, 14, 16, 20, 17), ]

##              2.5 %      97.5 %
## `Duration HR Zone 4 (s)` -0.15117629  0.25081777
## `Hi Int Acceleration (num)` -1.37979664  0.52978416
## `Distance Speed Zone 2 (m)` -0.02906099  0.40534053
## `Sprints Speed Zone 3 (num)`  0.27610348 36.72910782
## `Distance Speed Zone 3 (m)` -1.69615840 -0.02367729

```

These five models suggest that the most important GPS variables for a number 8 are, beginning from the most important:

Duration HR Zone 4 (s) | +0.0498, p-value = 0.6050

Every additional second a number 8 spends in HR Zone 4 contributes between -0.1512 and +0.2508 points to the win margin

Hi Int Acceleration (num) | -0.425, p-value = 0.3578

Every additional acceleration a number 8 performs over the high intensity acceleration benchmark contributes between -1.380 and +0.530 points to the win margin

Distance Speed Zone 2 (m) | +0.188, p-value = 0.0847

Every additional metre a number 8 covers in Speed Zone 2 contributes between -0.029 and +0.405 points to the win margin

Sprints Speed Zone 3 (num) | +18.5, p-value = 0.0470

Every additional sprint a number 8 performs in Speed Zone 3 contributes between +0.3 and +36.7 points to the win margin

Distance Speed Zone 3 (m) | -0.860, p-value = 0.0446

Every additional metre a number 8 covers in Speed Zone 3 contributes between -1.696 and -0.024 points to the win margin

## Position 9: Scrum-half

```
pos9data <- fullyCombined[which(fullyCombined$Position == 9), -c(1:4)]
pos9data$`Work Recovery Ratio` <- droplevels(pos9data$`Work Recovery Ratio`)
# All zeroes for Duration Speed Hi-Inten (s)
pos9data <- pos9data[, -c(2, 38, 40)]

corr9 <- cor(pos9data[, -c(10, 37)])
print("Which variables have high correlations with other variables?")

## [1] "Which variables have high correlations with other variables?"
findCorrelation(corr9, cutoff = 0.999)

## integer(0)
findCorrelation(corr9, cutoff = 0.9)

## [1] 10 9 7 25 3 16 5
# Removing variables that are causing singularities
pos9data <- pos9data[, -c(3, 7, 9, 10, 16, 25)]

modell1.9 <- regsubsets(margins ~ ., data = pos9data, method = "backward", nvmax = 100)
coef(modell1.9, 1:5)

## [[1]]
##           (Intercept) `Decelerations Zone 4 (num)`
##           7.018293              8.701220
##
## [[2]]
##           (Intercept) `Distance Speed Zone 3 (m)`
##           3.04577335              0.04033654
```



```
## `Decelerations Zone 4 (num)`
##      8.98888805
##
## [[3]]
##      (Intercept)  `Distance Speed Zone 3 (m)`
##      2.93088296      0.09515159
## `Sprints Speed Zone 3 (num)` `Decelerations Zone 4 (num)`
##      -1.59088861      12.08304716
##
## [[4]]
##      (Intercept)  `Duration HR Hi-Inten (s)`
##      -3.004025551      0.006489414
## `Distance Speed Zone 3 (m)` `Sprints Speed Zone 3 (num)`
##      0.098313573      -2.219030986
## `Decelerations Zone 4 (num)`
##      11.952154030
##
## [[5]]
##      (Intercept)  `Duration HR Hi-Inten (s)`
##      -8.87049151      0.01133749
## `Distance Speed Zone 3 (m)` `Sprints Speed Zone 3 (num)`
##      0.13450994      -3.49573560
## `Decelerations Zone 4 (num)` `Decelerations Zone 5 (num)`
##      10.69474725      23.32085658
```

```
full1.9 <- lm(margins ~ ., data = pos9data)
summary(full1.9)
```

```
##
## Call:
## lm(formula = margins ~ ., data = pos9data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -37.721  -0.667   0.000   5.523  28.165
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -95.679209  233.069084  -0.411  0.6852
## `Duration Total (s)`    -0.006542   0.006135  -1.066  0.2973
## `Duration HR Hi-Inten (s)`  0.075774   0.192013   0.395  0.6968
## `Distance Rate (m/min)`    2.335652   4.236743   0.551  0.5868
## `Distance HR Hi-Inten (m)`  0.013037   0.124840   0.104  0.9177
## `Speed Max (km/h)`    1.159365   0.893947   1.297  0.2075
## `Sprints Hi-Inten (num)`  -0.098614   0.827163  -0.119  0.9061
## `Athlete Load`    -4.273749   9.189136  -0.465  0.6462
## `Metabolic PowerPeak`  -0.584408   0.823750  -0.709  0.4852
## `Hi Int Acceleration (num)` -0.649664   0.491436  -1.322  0.1992
## `Hi Int Deceleration (num)` -4.140207   3.766888  -1.099  0.2831
## `Hi Intensity Effort (num)`  2.065383   3.190856   0.647  0.5239
## `Distance Speed Zone 1 (m)`  0.031844   0.014502   2.196  0.0385 *
## `Distance Speed Zone 2 (m)` -0.127548   0.064773  -1.969  0.0611 .
## `Distance Speed Zone 3 (m)`  0.405767   0.195480   2.076  0.0493 *
## `Distance Speed Zone 4 (m)` -0.040678   0.252778  -0.161  0.8736
## `Distance Speed Zone 5 (m)`  0.664078   0.664876   0.999  0.3283
```

```
## `Sprints Speed Zone 3 (num)` -6.796103 3.715817 -1.829 0.0804 .
## `Sprints Speed Zone 4 (num)` -1.839188 5.000890 -0.368 0.7164
## `Sprints Speed Zone 5 (num)` -29.068835 15.365831 -1.892 0.0712 .
## `Duration HR Zone 5 (s)` -0.105192 0.111813 -0.941 0.3566
## `Accelerations Zone 3 (num)` -6.218425 20.443054 -0.304 0.7637
## `Accelerations Zone 4 (num)` -5.582236 46.780110 -0.119 0.9061
## `Accelerations Zone 5 (num)` 45.391847 160.201820 0.283 0.7794
## `Decelerations Zone 3 (num)` 28.585982 60.816130 0.470 0.6428
## `Decelerations Zone 4 (num)` 50.775187 29.027958 1.749 0.0936 .
## `Decelerations Zone 5 (num)` 91.507425 148.735777 0.615 0.5444
## `Body Impacts (num)` -1.407481 1.329695 -1.058 0.3008
## `Body Impacts Grade 1 (num)` 2.332537 5.490568 0.425 0.6749
## `Body Impacts Grade 2 (num)` -25.408840 46.603982 -0.545 0.5909
## `Body Impacts Grade 3 (num)` 155.248391 293.715706 0.529 0.6022
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.04 on 23 degrees of freedom
## Multiple R-squared: 0.5305, Adjusted R-squared: -0.08194
## F-statistic: 0.8662 on 30 and 23 DF, p-value: 0.6485
confint(full1.9)[c(26, 15, 18, 3, 27), ]

##                2.5 %      97.5 %
## `Decelerations Zone 4 (num)` -9.273719e+00 110.8240940
## `Distance Speed Zone 3 (m)` 1.385104e-03 0.8101480
## `Sprints Speed Zone 3 (num)` -1.448286e+01 0.8906510
## `Duration HR Hi-Inten (s)` -3.214359e-01 0.4729829
## `Decelerations Zone 5 (num)` -2.161760e+02 399.1908233
```

These five models suggest that the most important GPS variables for a scrum-half are, beginning from the most important:

Decelerations Zone 4 (num) | +50.8, p-value = 0.0936

Every additional deceleration a scrum-half performs in Deceleration Zone 4 contributes between -9.3 and +110.8 points(!) to the win margin

Distance Speed Zone 3 (m) | +0.406, p-value = 0.0493

Every additional metre a scrum-half covers in Speed Zone 3 contributes between +0.001 and +0.810 points to the win margin

Sprints Speed Zone 3 (num) | -6.80, p-value = 0.0804

Every additional sprint a scrum-half performs in Speed Zone 3 contributes between -14.48 and +0.89 points to the win margin

Duration HR Hi-Inten (s) | +0.0758, p-value = 0.6968

Every additional second a scrum-half spends over the high intensity heart rate benchmark contributes between -0.3214 and +0.4730 points to the win margin

Decelerations Zone 5 (num) | +91.5, p-value = 0.5444

Every additional deceleration a scrum-half performs in Deceleration Zone 5 contributes between -216.2 and +399.2 points(!) to the win margin

## Position 10: Fly-half

```
pos10data <- fullyCombined[which(fullyCombined$Position == 10), -c(1:4)]
pos10data$`Work Recovery Ratio` <- droplevels(pos10data$`Work Recovery Ratio`)
# All zeroes for Duration Speed Hi-Inten (s) and Accelerations Zone 5 (num)
pos10data <- pos10data[, -c(2, 30, 38, 40)]

corr10 <- cor(pos10data[, -c(10, 36)])
print("Which variables have high correlations with other variables?")

## [1] "Which variables have high correlations with other variables?"
findCorrelation(corr10, cutoff = 0.999)

## integer(0)
findCorrelation(corr10, cutoff = 0.9)

## [1] 14 12 3 7 16 9
findCorrelation(corr10, cutoff = 0.8)

## [1] 14 13 12 3 7 16 10 1 9 2 20
findCorrelation(corr10, cutoff = 0.75)

## [1] 14 13 12 3 7 16 10 1 18 9 28 2 20
findCorrelation(corr10, cutoff = 0.73)

## [1] 14 13 12 3 7 16 10 21 1 22 18 9 28 23 2 4
# Removing variables that are causing singularities
pos10data <- pos10data[, -c(1:4, 7, 9, 10, 12:14, 16, 18, 20:23, 28)]

modell1.10 <- regsubsets(margins ~ ., data = pos10data, method = "backward", nvmax = 100)
coef(modell1.10, 1:5)

## [[1]]
##                (Intercept) `Sprints Hi-Inten (num)`
##                26.06091                -10.18663
##
## [[2]]
##                (Intercept)      `Sprints Hi-Inten (num)`
##                28.605360                -15.531702
## `Decelerations Zone 3 (num)`
##                6.102755
##
## [[3]]
##                (Intercept)      `Sprints Hi-Inten (num)`
##                37.101009                -19.635535
## `Decelerations Zone 3 (num)` `Decelerations Zone 4 (num)`
##                11.153360                -6.568018
##
## [[4]]
##                (Intercept)      `Sprints Hi-Inten (num)`
##                34.31298646                -22.09670121
## `Distance Speed Zone 3 (m)` `Decelerations Zone 3 (num)`
##                0.05037379                12.73109876
```

```
## `Decelerations Zone 4 (num)`
##          -7.07560218
##
## [[5]]
##          (Intercept)      `Sprints Hi-Inten (num)`
##          40.8328496          -24.9805575
## `Distance Speed Zone 3 (m)` `Decelerations Zone 3 (num)`
##          0.1014444          14.7942455
## `Decelerations Zone 4 (num)`      `Body Impacts (num)`
##          -7.0026377          -1.0979361
```

```
full1.10 <- lm(margins ~ ., data = pos10data)
summary(full1.10)
```

```
##
## Call:
## lm(formula = margins ~ ., data = pos10data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -44.263  -8.483   0.000   8.219  41.338
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.752e+02  2.992e+02  -0.586   0.5630
## `Distance HR Hi-Inten (m)` -4.693e-02  7.605e-02  -0.617   0.5424
## `Speed Max (km/h)`        8.608e-01  2.012e+00   0.428   0.6721
## `Sprints Hi-Inten (num)`  -6.715e+01  6.087e+01  -1.103   0.2797
## `Athlete Load`          5.395e+00  8.061e+00   0.669   0.5091
## `Hi Intensity Effort (num)` 1.533e+00  2.168e+00   0.707   0.4857
## `Distance Speed Zone 1 (m)` -3.136e-03  3.937e-03  -0.797   0.4326
## `Distance Speed Zone 3 (m)` 1.551e-01  7.727e-02   2.007   0.0548
## `Sprints Speed Zone 5 (num)` -9.950e+00  1.193e+01  -0.834   0.4117
## `Duration HR Zone 4 (s)`   -8.522e-02  1.416e-01  -0.602   0.5522
## `Duration HR Zone 5 (s)`    6.745e-02  1.093e-01   0.617   0.5422
## `Accelerations Zone 3 (num)` -7.468e+00  1.606e+01  -0.465   0.6455
## `Decelerations Zone 3 (num)` 3.764e+01  3.910e+01   0.963   0.3443
## `Decelerations Zone 4 (num)` -4.843e+01  5.786e+01  -0.837   0.4099
## `Decelerations Zone 5 (num)` -5.815e+01  9.401e+01  -0.619   0.5414
## `Body Impacts (num)`       -6.537e-01  1.143e+00  -0.572   0.5722
## `Body Impacts Grade 1 (num)` -1.442e+01  1.721e+01  -0.838   0.4093
## `Body Impacts Grade 2 (num)` -1.084e-01  6.272e+00  -0.017   0.9863
## `Body Impacts Grade 3 (num)` 3.658e+01  8.748e+01   0.418   0.6791
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.25 on 27 degrees of freedom
## Multiple R-squared:  0.338, Adjusted R-squared:  -0.1034
## F-statistic: 0.7658 on 18 and 27 DF,  p-value: 0.7184
confint(full1.10)[c(4, 13, 14, 8, 16), ]

##              2.5 %      97.5 %
## `Sprints Hi-Inten (num)` -1.920380e+02  57.7476343
## `Decelerations Zone 3 (num)` -4.258785e+01 117.8637242
```

```
## `Decelerations Zone 4 (num)` -1.671593e+02 70.2928785
## `Distance Speed Zone 3 (m)` -3.454265e-03 0.3136396
## `Body Impacts (num)` -2.999142e+00 1.6917644
```

These five models suggest that the most important GPS variables for a fly-half are, beginning from the most important:

Sprints Hi-Inten (num) | -67.2, p-value = 0.2797

Every additional sprint a fly-half performs above the high intensity sprint benchmark contributes between -192.0 and +57.7 points(!) to the win margin

Decelerations Zone 3 (num) | +37.6, p-value = 0.3443

Every additional deceleration a fly-half performs in Deceleration Zone 3 contributes between -42.6 and +117.9 points(!) to the win margin

Decelerations Zone 4 (num) | -48.4, p-value = 0.4099

Every additional deceleration a fly-half performs in Deceleration Zone 4 contributes between -167.2 and +70.3 points(!) to the win margin

Distance Speed Zone 3 (m) | +0.155, p-value = 0.0548

Every additional metre a fly-half covers in Speed Zone 3 contributes between -0.003 and +0.314 points to the win margin

Body Impacts (num) | 0.654, p-value = 0.5722

Every additional body impact a fly-half makes in a match contributes between -2.999 and +1.692 points to the win margin

### Position 11: Left wing

```
pos11data <- fullyCombined[which(fullyCombined$Position == 11), -c(1:4)]
pos11data$`Work Recovery Ratio` <- droplevels(pos11data$`Work Recovery Ratio`)
```

```
pos11data <- pos11data[, -c(38, 40)]
```

```
corr11 <- cor(pos11data[, -c(11, 38)])
print("Which variables have high correlations with other variables?")
```

```
## [1] "Which variables have high correlations with other variables?"
```

```
findCorrelation(corr11, cutoff = 0.999)
```

```
## [1] 26
```

```
findCorrelation(corr11, cutoff = 0.9)
```

```
## [1] 6 10 26 15 4
```

```
findCorrelation(corr11, cutoff = 0.8)
```

```
## [1] 6 10 26 15 8 16 13 4 24 1
```

```
findCorrelation(corr11, cutoff = 0.75)
```

```
## [1] 6 10 26 15 8 16 13 4 24 1
```

```
findCorrelation(corr11, cutoff = 0.7)
```

```
## [1] 6 10 26 3 15 28 27 8 16 13 4 9 17 24
```

```

findCorrelation(corr11, cutoff = 0.66)

## [1] 6 10 26 3 15 28 27 8 16 13 5 14 32 4 9 17 24 18
# Removing variables that are causing singularities
pos11data <- pos11data[, -c(1, 3:6, 8:10, 13:17, 24, 26:28, 32)]

model1.11 <- regsubsets(margins ~ ., data = pos11data, method = "backward", nvmax = 100)
coef(model1.11, 1:5)

## [[1]]
##              (Intercept) `Distance Speed Zone 3 (m)`
##              -16.7044473                0.1109714
##
## [[2]]
##              (Intercept)      `Work Recovery Ratio`2:3
##              -22.4821815                13.5759857
## `Distance Speed Zone 3 (m)`
##              0.1261345
##
## [[3]]
##              (Intercept)      `Work Recovery Ratio`2:3
##              -32.785999                16.307532
## `Distance Speed Zone 3 (m)` `Decelerations Zone 3 (num)`
##              0.125344                8.185697
##
## [[4]]
##              (Intercept)      `Work Recovery Ratio`2:3
##              -40.6424593                23.7480951
## `Distance Speed Zone 3 (m)` `Decelerations Zone 3 (num)`
##              0.1433303                12.8755247
## `Decelerations Zone 5 (num)`
##              -15.3893722
##
## [[5]]
##              (Intercept)      `Work Recovery Ratio`2:3
##              -40.3230538                23.5581557
## `Distance Speed Zone 3 (m)` `Distance Speed Zone 5 (m)`
##              0.1128178                0.1501208
## `Decelerations Zone 3 (num)` `Decelerations Zone 5 (num)`
##              12.8491644                -15.2660507

full1.11 <- lm(margins ~ ., data = pos11data)
summary(full1.11)

##
## Call:
## lm(formula = margins ~ ., data = pos11data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -39.200  -9.009   0.000   3.505  45.159
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)

```

```
## (Intercept) -4.982e+02 9.954e+02 -0.500 0.626
## `Duration Speed Hi-Inten (s)` 6.565e+01 1.821e+02 0.361 0.725
## `Speed Max (km/h)` 3.891e+00 8.391e+00 0.464 0.651
## `Work Recovery Ratio`1:1 -7.093e+01 2.568e+02 -0.276 0.787
## `Work Recovery Ratio`2:3 2.902e+01 1.144e+02 0.254 0.804
## `Athlete Load` 5.475e+00 1.936e+01 0.283 0.782
## `Distance Speed Zone 1 (m)` -6.652e-03 1.456e-02 -0.457 0.656
## `Distance Speed Zone 2 (m)` 1.042e-01 1.879e-01 0.554 0.589
## `Distance Speed Zone 3 (m)` 4.965e-02 2.183e-01 0.227 0.824
## `Distance Speed Zone 4 (m)` 8.309e-02 2.035e-01 0.408 0.690
## `Distance Speed Zone 5 (m)` 1.907e-01 3.829e-01 0.498 0.627
## `Sprints Speed Zone 3 (num)` 2.163e-01 3.815e+00 0.057 0.956
## `Sprints Speed Zone 5 (num)` -5.118e-01 7.442e+00 -0.069 0.946
## `Accelerations Zone 4 (num)` -1.612e+01 6.481e+01 -0.249 0.808
## `Accelerations Zone 5 (num)` 1.313e+00 4.350e+01 0.030 0.976
## `Decelerations Zone 3 (num)` 7.597e+01 1.423e+02 0.534 0.603
## `Decelerations Zone 5 (num)` -3.512e+01 6.012e+01 -0.584 0.570
## `Body Impacts (num)` -4.940e-02 1.089e+00 -0.045 0.965
## `Body Impacts Grade 1 (num)` 8.732e+00 1.835e+01 0.476 0.643
## `Body Impacts Grade 2 (num)` -1.016e+01 2.030e+01 -0.501 0.626
## `Body Impacts Grade 3 (num)` -3.659e+01 9.549e+01 -0.383 0.708
##
```

```
## Residual standard error: 25.4 on 12 degrees of freedom
## Multiple R-squared: 0.4525, Adjusted R-squared: -0.4601
## F-statistic: 0.4958 on 20 and 12 DF, p-value: 0.9202
```

```
confint(full1.11)[c(9, 5, 16:17, 11), ]
```

```
##                2.5 %      97.5 %
## `Distance Speed Zone 3 (m)` -0.4260927 0.5253936
## `Work Recovery Ratio`2:3 -220.1288288 278.1747652
## `Decelerations Zone 3 (num)` -234.1163818 386.0660940
## `Decelerations Zone 5 (num)` -166.1168911 95.8675680
## `Distance Speed Zone 5 (m)` -0.6435470 1.0249782
```

These five models suggest that the most important GPS variables for a left wing are, beginning from the most important:

Distance Speed Zone 3 (m) | +0.0497, p-value = 0.824

Every additional metre a left wing covers in Speed Zone 3 contributes between -0.4261 and +0.5254 points to the win margin

Work Recovery Ratio | 2:3 -> +29.0, p-value = 0.804

If a left wing has a Work Recovery Ratio of 2:3 in a match, it will contribute between -220.1 to +278.2 points(!) to the win margin

Decelerations Zone 3 | +76.0, p-value = 0.603

Every additional deceleration a left wing performs in Deceleration Zone 3 contributes between -234.1 and +386.1 points(!) to the win margin

Decelerations Zone 5 | -35.1, p-value = 0.570

Every additional deceleration a left wing performs in Deceleration Zone 5 contributes between -166.1 and +95.9 points(!) to the win margin

Distance Speed Zone 5 (m) | +0.191, p-value = 0.627

Every additional metre a left wing performs in Speed Zone 5 contributes between -0.644 and 1.024 points to the win margin

## Position 12: Inside centre

```
pos12data <- fullyCombined[which(fullyCombined$Position == 12), -c(1:4)]
pos12data$`Work Recovery Ratio` <- droplevels(pos12data$`Work Recovery Ratio`)

pos12data <- pos12data[, -c(38, 40)]

corr12 <- cor(pos12data[, -c(11, 38)])
print("Which variables have high correlations with other variables?")

## [1] "Which variables have high correlations with other variables?"
findCorrelation(corr12, cutoff = 0.999)

## integer(0)
findCorrelation(corr12, cutoff = 0.9)

## [1] 15 11 1 26 4 17 6 10
findCorrelation(corr12, cutoff = 0.8)

## [1] 15 11 1 14 26 4 17 13 34 35 19 6 3 32
findCorrelation(corr12, cutoff = 0.75)

## [1] 15 11 1 14 26 4 17 13 34 35 19 6 3 20 32
findCorrelation(corr12, cutoff = 0.7)

## [1] 15 11 1 14 26 4 8 17 13 34 22 35 19 6 10 23 20 32
findCorrelation(corr12, cutoff = 0.68)

## [1] 15 11 1 14 26 4 8 17 13 34 22 35 19 6 10 23 9 20 32
# Removing variables that are causing singularities
pos12data <- pos12data[, -c(1, 3:4, 6, 8:11, 13:15, 17, 19:20, 22:23, 26, 32, 34:35)]

model1.12 <- regsubsets(margins ~ ., data = pos12data, method = "backward", nvmax = 100)
coef(model1.12, 1:5)

## [[1]]
##           (Intercept) `Distance Speed Zone 1 (m)`
##           20.737275745                -0.002580919
##
## [[2]]
##           (Intercept)                `Athlete Load`
##           -5.112445661                0.798870443
## `Distance Speed Zone 1 (m)`
##           -0.004541884
##
## [[3]]
##           (Intercept)                `Athlete Load`
##           -10.926570984                0.891613877
## `Distance Speed Zone 1 (m)` `Sprints Speed Zone 4 (num)`
##           -0.005339584                2.693653493
```



```
##
## [[4]]
##           (Intercept)           `Athlete Load`
##           -15.575421974           1.335766856
## `Distance Speed Zone 1 (m)` `Sprints Speed Zone 4 (num)`
##           -0.005947993           4.920707903
## `Body Impacts Grade 2 (num)`
##           -4.129661351
##
## [[5]]
##           (Intercept)           `Athlete Load`
##           -17.118190370           1.496380315
## `Distance Speed Zone 1 (m)` `Sprints Speed Zone 4 (num)`
##           -0.006123201           4.313514443
## `Decelerations Zone 5 (num)` `Body Impacts Grade 2 (num)`
##           8.516673284           -5.422884416
```

```
full1.12 <- lm(margins ~ ., data = pos12data)
summary(full1.12)
```

```
##
## Call:
## lm(formula = margins ~ ., data = pos12data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.575  -5.709   0.000   4.647  36.540
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -35.780624  284.731190  -0.126   0.9016
## `Duration Speed Hi-Inten (s)` -18.729118  301.143116  -0.062   0.9512
## `Distance Rate (m/min)`      -0.399072   3.623277  -0.110   0.9137
## `Speed Max (km/h)`           0.560619   2.326423   0.241   0.8126
## `Athlete Load`           1.369521  10.906271   0.126   0.9016
## `Hi Intensity Effort (num)`    0.443029   3.533500   0.125   0.9018
## `Distance Speed Zone 1 (m)`   -0.006408   0.003177  -2.017   0.0608
## `Distance Speed Zone 4 (m)`   -0.353824   0.312726  -1.131   0.2746
## `Sprints Speed Zone 4 (num)`  16.812864  10.953425   1.535   0.1443
## `Sprints Speed Zone 5 (num)`  -5.176114  14.522068  -0.356   0.7262
## `Duration HR Zone 5 (s)`     -0.026442   0.097368  -0.272   0.7894
## `Accelerations Zone 3 (num)`   3.874734  21.453745   0.181   0.8589
## `Accelerations Zone 4 (num)`   5.339406  25.798751   0.207   0.8386
## `Accelerations Zone 5 (num)` -55.574834  76.371354  -0.728   0.4773
## `Decelerations Zone 3 (num)`   3.368323  15.313844   0.220   0.8287
## `Decelerations Zone 5 (num)`  27.148015  52.814404   0.514   0.6143
## `Body Impacts Grade 2 (num)` -14.178829   8.128753  -1.744   0.1003
## `Body Impacts Grade 3 (num)` -19.978556  31.119354  -0.642   0.5300
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.96 on 16 degrees of freedom
## Multiple R-squared:  0.5173, Adjusted R-squared:  0.004487
## F-statistic: 1.009 on 17 and 16 DF, p-value: 0.4951
```

```
confint(full1.12)[c(7, 5, 9, 17, 16), ]
```

```
##                2.5 %          97.5 %
## `Distance Speed Zone 1 (m)` -0.01314359 3.268043e-04
## `Athlete Load` -21.75074112 2.448978e+01
## `Sprints Speed Zone 4 (num)` -6.40735949 4.003309e+01
## `Body Impacts Grade 2 (num)` -31.41101519 3.053357e+00
## `Decelerations Zone 5 (num)` -84.81351944 1.391095e+02
```

These five models suggest that the most important GPS variables for an inside centre are, beginning from the most important:

Distance Speed Zone 1 (m) | -0.00641, p-value = 0.0608

Every additional metre an inside centre covers in Speed Zone 1 contributes between -0.01314 and +0.00033 points to the win margin

Athlete Load | +1.37, p-value = 0.9016

Every additional point in Athlete Load an inside centre has contributes between -21.75 and +24.49 points to the win margin

Sprints Speed Zone 4 (num) | +16.8, p-value = 0.1443

Every additional sprint an inside centre performs in Speed Zone 4 contributes between -6.4 and +40.0 points(!) to the win margin

Body Impacts Grade 2 (num) | -14.2, p-value = 0.1003

Every additional Grade 2 body impact an inside centre performs contributes between -31.4 and +3.1 points(!) to the win margin

Decelerations Zone 5 (num) | +27.1, p-value = 0.6143

Every additional deceleration an inside centre performs in Deceleration Zone 5 contributes between -84.8 and +139.1 points(!) to the win margin

### Position 13: Outside centre

```
pos13data <- fullyCombined[which(fullyCombined$Position == 13), -c(1:4)]
pos13data$`Work Recovery Ratio` <- droplevels(pos13data$`Work Recovery Ratio`)
# All zeroes in Body Impacts Grade 3 (num)
pos13data <- pos13data[, -c(37:38, 40)]
```

```
corr13 <- cor(pos13data[, -c(11, 37)])
print("Which variables have high correlations with other variables?")
```

```
## [1] "Which variables have high correlations with other variables?"
```

```
findCorrelation(corr13, cutoff = 0.999)
```

```
## integer(0)
```

```
findCorrelation(corr13, cutoff = 0.9)
```

```
## [1] 11 1 4 17 6 3
```

```
findCorrelation(corr13, cutoff = 0.8)
```

```
## [1] 11 8 1 4 17 19 6 13 3 21
```

```

findCorrelation(corr13, cutoff = 0.75)

## [1] 11 8 1 4 17 19 14 20 6 13 3 16 24 21 12
findCorrelation(corr13, cutoff = 0.7)

## [1] 11 8 1 4 17 19 14 20 6 13 3 15 16 24 21 12 27
findCorrelation(corr13, cutoff = 0.67)

## [1] 11 8 1 4 17 19 14 20 6 13 3 15 16 24 26 7 12 27
# Removing variables that are causing singularities
pos13data <- pos13data[, -c(1, 3:4, 6:8, 11:17, 19:21, 24, 26:27)]

modell1.13 <- regsubsets(margins ~ ., data = pos13data, method = "backward", nvmax = 100)
coef(modell1.13, 1:5)

## [[1]]
##              (Intercept) `Sprints Speed Zone 3 (num)`
##              0.9076027              1.0257868
##
## [[2]]
##              (Intercept) `Sprints Speed Zone 3 (num)`
##              6.1968834              1.7244803
##      `Body Impacts (num)`
##      -0.9216233
##
## [[3]]
##              (Intercept) `Sprints Speed Zone 3 (num)`
##              14.0388077              1.9465138
##      `Decelerations Zone 3 (num)`      `Body Impacts (num)`
##      -3.5697384              -0.9299268
##
## [[4]]
##              (Intercept) `Sprints Speed Zone 3 (num)`
##              7.8670202              2.1273760
##      `Accelerations Zone 4 (num)` `Decelerations Zone 3 (num)`
##              7.8079322              -4.0580790
##      `Body Impacts (num)`
##      -0.9008527
##
## [[5]]
##              (Intercept) `Sprints Speed Zone 3 (num)`
##              -5.798098              2.363818
##      `Accelerations Zone 4 (num)` `Decelerations Zone 3 (num)`
##              11.088195              -6.390728
##      `Body Impacts (num)` `Body Impacts Grade 1 (num)`
##              -1.171847              2.324058

full1.13 <- lm(margins ~ ., data = pos13data)
summary(full1.13)

##
## Call:
## lm(formula = margins ~ ., data = pos13data)
##

```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -42.393  -9.484   0.000   2.266  44.448
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.478e+02  9.301e+02   0.912   0.373
## `Duration Speed Hi-Inten (s)`  5.914e+01  7.240e+01   0.817   0.424
## `Distance Rate (m/min)`      -1.248e+01  1.364e+01  -0.915   0.372
## `Sprints Hi-Inten (num)`      3.721e+01  4.339e+01   0.858   0.402
## `Sprints HR Hi-Inten (num)`  -5.252e+00  5.275e+00  -0.996   0.332
## `Distance Speed Zone 1 (m)`  -1.858e-03  3.708e-03  -0.501   0.622
## `Distance Speed Zone 5 (m)`  -3.398e-01  5.476e-01  -0.620   0.542
## `Sprints Speed Zone 3 (num)`  3.302e+00  2.239e+00   1.475   0.157
## `Sprints Speed Zone 5 (num)`  1.023e+01  1.168e+01   0.875   0.392
## `Accelerations Zone 3 (num)` -7.368e+00  2.850e+01  -0.259   0.799
## `Accelerations Zone 4 (num)`  5.346e+01  3.438e+01   1.555   0.136
## `Accelerations Zone 5 (num)` -3.011e+02  4.062e+02  -0.741   0.468
## `Decelerations Zone 3 (num)` -8.072e+01  8.431e+01  -0.957   0.350
## `Decelerations Zone 4 (num)`  7.709e+01  8.628e+01   0.893   0.383
## `Decelerations Zone 5 (num)` -3.021e+02  3.099e+02  -0.975   0.342
## `Body Impacts (num)`        -1.244e+00  1.086e+00  -1.146   0.266
## `Body Impacts Grade 1 (num)`  2.259e+01  2.018e+01   1.119   0.277
## `Body Impacts Grade 2 (num)`  4.505e+01  4.598e+01   0.980   0.339
##
## Residual standard error: 22.24 on 19 degrees of freedom
## Multiple R-squared:  0.3128, Adjusted R-squared:  -0.302
## F-statistic: 0.5088 on 17 and 19 DF,  p-value: 0.9164
confint(full1.13)[c(8, 16, 13, 11, 17), ]

##              2.5 %      97.5 %
## `Sprints Speed Zone 3 (num)`  -1.384461  7.988569
## `Body Impacts (num)`         -3.517997  1.029226
## `Decelerations Zone 3 (num)` -257.180025  95.740118
## `Accelerations Zone 4 (num)` -18.494345 125.410147
## `Body Impacts Grade 1 (num)` -19.655318  64.834048
```

These five models suggest that the most important GPS variables for an outside centre are, beginning from the most important:

Sprints Speed Zone 3 (num) | +3.30, p-value = 0.157

Every additional sprint an outside centre performs in Speed Zone 3 contributes between -1.38 and +7.99 points to the win margin

Body Impacts (num) | -1.24, p-value = 0.266

Every additional body impact an outside centre performs in a match contributes between -3.52 and +1.03 points to the win margin

Decelerations Zone 3 (num) | -80.7, p-value = 0.350

Every additional deceleration an outside centre performs in Deceleration Zone 3 contributes between -257.2 and +95.7 points(!) to the win margin

Accelerations Zone 4 (num) | +53.5, p-value = 0.136

Every additional acceleration an outside centre performs in Acceleration Zone 4 contributes between -18.5 and +125.4 points(!) to the win margin

Body Impacts Grade 1 (num) | +22.6, p-value = 0.277

Every additional Grade 1 body impact an outside centre performs contributes between -19.7 and +64.8 points(!) to the win margin

#### Position 14: Right wing

```
pos14data <- fullyCombined[which(fullyCombined$Position == 14), -c(1:4)]
pos14data$`Work Recovery Ratio` <- droplevels(pos14data$`Work Recovery Ratio`)

pos14data <- pos14data[, -c(38, 40)]

corr14 <- cor(pos14data[, -c(11, 38)])
print("Which variables have high correlations with other variables?")

## [1] "Which variables have high correlations with other variables?"
findCorrelation(corr14, cutoff = 0.999)

## integer(0)
findCorrelation(corr14, cutoff = 0.9)

## [1] 11 14 8 1 4 10 6 5 26
findCorrelation(corr14, cutoff = 0.87)

## [1] 11 14 8 15 1 4 10 6 3 5
# Removing variables that are causing singularities
pos14data <- pos14data[, -c(1, 4:6, 8, 10:11, 14:15, 26)]

modell1.14 <- regsubsets(margins ~ ., data = pos14data, method = "backward", nvmax = 100)
coef(modell1.14, 1:5)

## [[1]]
## (Intercept) `HIE Rate`
##      5.281690      2.293763
##
## [[2]]
##              (Intercept)              `HIE Rate`
##              3.457321              3.110810
## `Body Impacts Grade 3 (num)`
##              11.366068
##
## [[3]]
##              (Intercept)              `HIE Rate`
##              -8.458637550              6.434066595
## `Duration HR Zone 5 (s)` `Body Impacts Grade 3 (num)`
##              0.003711459              19.761183198
##
## [[4]]
##              (Intercept)              `HIE Rate`
##              -22.447300895              10.047572235
## `Duration HR Zone 5 (s)` `Body Impacts Grade 2 (num)`
```

```
##              0.005182309              4.235475160
## `Body Impacts Grade 3 (num)`
##              27.252470425
##
## [[5]]
##              (Intercept)              `HIE Rate`
##              -41.671785579              16.003910145
##      `Duration HR Zone 5 (s)` `Accelerations Zone 4 (num)`
##              0.008303351              -4.123528227
## `Body Impacts Grade 2 (num)` `Body Impacts Grade 3 (num)`
##              8.724921775              40.926050293
```

```
full1.14 <- lm(margins ~ ., data = pos14data)
summary(full1.14)
```

```
##
## Call:
## lm(formula = margins ~ ., data = pos14data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -32.036   0.000   0.000   3.478  48.800
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.648e+02  2.491e+02  -1.063   0.311
## `Duration Speed Hi-Inten (s)`  3.205e+02  2.219e+02   1.444   0.177
## `Duration HR Hi-Inten (s)`    -4.871e-03  5.946e-02  -0.082   0.936
## `Speed Max (km/h)`           1.070e+00  3.224e+00   0.332   0.746
## `Sprints Hi-Inten (num)`     -1.080e+01  1.428e+01  -0.757   0.465
## `Athlete Load`              -5.394e+00  7.254e+00  -0.744   0.473
## `Metabolic PowerPeak`       -2.833e-01  2.469e-01  -1.147   0.276
## `Hi Intensity Effort (num)`   1.666e+00  1.407e+00   1.184   0.262
## `HIE Rate`                 1.022e+02  9.254e+01   1.105   0.293
## `Distance Speed Zone 1 (m)`   7.506e-03  1.007e-02   0.745   0.472
## `Distance Speed Zone 2 (m)`   1.940e-02  1.417e-01   0.137   0.894
## `Distance Speed Zone 3 (m)`   3.129e-02  2.790e-01   0.112   0.913
## `Distance Speed Zone 4 (m)`   8.531e-02  3.884e-01   0.220   0.830
## `Distance Speed Zone 5 (m)`   6.656e-02  4.172e-01   0.160   0.876
## `Sprints Speed Zone 3 (num)`  -2.677e+00  3.864e+00  -0.693   0.503
## `Sprints Speed Zone 4 (num)`  -1.460e+00  7.473e+00  -0.195   0.849
## `Sprints Speed Zone 5 (num)`  -2.827e+00  1.003e+01  -0.282   0.783
## `Duration HR Zone 5 (s)`       7.610e-02  5.197e-02   1.464   0.171
## `Accelerations Zone 3 (num)`  -1.857e+00  7.377e+00  -0.252   0.806
## `Accelerations Zone 4 (num)`  -2.927e+01  3.260e+01  -0.898   0.389
## `Accelerations Zone 5 (num)`  -8.958e+01  6.062e+01  -1.478   0.168
## `Decelerations Zone 3 (num)`   1.944e+01  4.175e+01   0.466   0.651
## `Decelerations Zone 4 (num)`  -4.698e+00  3.131e+01  -0.150   0.883
## `Decelerations Zone 5 (num)`   9.318e+00  2.285e+01   0.408   0.691
## `Body Impacts (num)`         -8.638e-01  1.660e+00  -0.520   0.613
## `Body Impacts Grade 1 (num)`   1.674e+00  1.586e+01   0.106   0.918
## `Body Impacts Grade 2 (num)`   6.769e+01  7.581e+01   0.893   0.391
## `Body Impacts Grade 3 (num)`   2.824e+02  1.907e+02   1.481   0.167
##
## Residual standard error: 25.66 on 11 degrees of freedom
```

```
## Multiple R-squared:  0.4775, Adjusted R-squared:  -0.8049
## F-statistic: 0.3724 on 27 and 11 DF,  p-value: 0.9821
confint(full1.14)[c(9, 28, 18, 27, 20), ]
```

```
##                2.5 %      97.5 %
## `HIE Rate`      -101.46859098 305.8947800
## `Body Impacts Grade 3 (num)` -137.30675378 702.1576077
## `Duration HR Zone 5 (s)`     -0.03829273  0.1904906
## `Body Impacts Grade 2 (num)` -99.16137563 234.5325434
## `Accelerations Zone 4 (num)` -101.01977773 42.4891787
```

These five models suggest that the most important GPS variables for a right wing are, beginning from the most important:

HIE Rate | +102, p-value = 0.293

Every additional high intensity effort a right wing completes per unit of time contributes between -102 and +306 points(!) to the win margin

Body Impacts Grade 3 (num) | +282, p-value = 0.167

Every additional Grade 3 body impact a right wing performs contributes between -137 and +702 points(!) to the win margin

Duration HR Zone 5 (s) | +0.0761, p-value = 0.171

Every additional second a right wing remains in HR Zone 5 contributes between -0.0383 and +0.1905 points to the win margin

Body Impacts Grade 2 (num) | +67.7, p-value = 0.391

Every additional Grade 2 body impact a right wing performs contributes between -99.2 and +234.5 points to the win margin

Accelerations Zone 4 (num) | -29.3, p-value = 0.389

Every additional acceleration a right wing performs in Acceleration Zone 4 contributes between -101.0 and +42.5 points to the win margin

## Position 15: Fullback

```
pos15data <- fullyCombined[which(fullyCombined$Position == 15), -c(1:4)]
pos15data$`Work Recovery Ratio` <- droplevels(pos15data$`Work Recovery Ratio`)
```

```
pos15data <- pos15data[, -c(38, 40)]
```

```
corr15 <- cor(pos15data[, -c(11, 38)])
print("Which variables have high correlations with other variables?")
```

```
## [1] "Which variables have high correlations with other variables?"
```

```
findCorrelation(corr15, cutoff = 0.999)
```

```
## integer(0)
```

```
findCorrelation(corr15, cutoff = 0.9)
```

```
## [1]  4 17  8 10 26  3
```

```
findCorrelation(corr15, cutoff = 0.8)
```

```
## [1] 13 15  4 17 19 14  8 10  6  3 21
```

```

findCorrelation(corr15, cutoff = 0.7)

## [1] 13 15 4 17 19 14 23 8 1 9 35 10 6 33 3 21
findCorrelation(corr15, cutoff = 0.65)

## [1] 13 15 4 17 19 14 23 8 22 5 9 18 35 10 6 25 3 24
findCorrelation(corr15, cutoff = 0.64)

## [1] 13 15 4 17 19 14 23 8 34 22 5 9 18 35 31 10 6 25 3 24
# Removing variables that are causing singularities
pos15data <- pos15data[, -c(1, 3:6, 8:10, 13:15, 17:27, 31, 33:35)]

modell1.15 <- regsubsets(margins ~ ., data = pos15data, method = "backward", nvmax = 100)
coef(modell1.15, 1:5)

## [[1]]
##                (Intercept) `Body Impacts Grade 2 (num)`
##                3.204045                3.908699
##
## [[2]]
##                (Intercept)      `Work Recovery Ratio`1:2
##                -0.4870871                26.4870871
## `Body Impacts Grade 2 (num)`
##                5.7250947
##
## [[3]]
##                (Intercept)      `Work Recovery Ratio`1:2
##                -2.506406                28.506406
## `Accelerations Zone 3 (num)` `Body Impacts Grade 2 (num)`
##                2.141890                4.601940
##
## [[4]]
##                (Intercept)      `Work Recovery Ratio`1:2
##                10.9895057                30.6506150
## `Hi Intensity Effort (num)` `Accelerations Zone 3 (num)`
##                -0.1078629                4.0983218
## `Body Impacts Grade 2 (num)`
##                6.1963714
##
## [[5]]
##                (Intercept) `Duration Speed Hi-Inten (s)`
##                15.6685817                4.0157185
##      `Work Recovery Ratio`1:2      `Hi Intensity Effort (num)`
##                38.9730812                -0.1975287
## `Accelerations Zone 3 (num)` `Body Impacts Grade 2 (num)`
##                6.2537125                9.1336825

full1.15 <- lm(margins ~ ., data = pos15data)
summary(full1.15)

##
## Call:
## lm(formula = margins ~ ., data = pos15data)
##

```



```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -47.739  -6.845   0.000   0.000  48.975
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      9.2151    131.6525   0.070   0.945
## `Duration Speed Hi-Inten (s)` 25.5426    88.8624   0.287   0.777
## `Speed Max (km/h)`      0.4777     1.6737   0.285   0.778
## `Work Recovery Ratio`1:1  -35.7169   117.9214  -0.303   0.765
## `Work Recovery Ratio`1:2   95.7592   204.9469   0.467   0.645
## `Work Recovery Ratio`2:3  -97.7137   435.5723  -0.224   0.825
## `Athlete Load`          -1.3182     6.1969  -0.213   0.834
## `Hi Intensity Effort (num)` -0.2522     0.8832  -0.286   0.778
## `Accelerations Zone 3 (num)` 30.0837    99.5539   0.302   0.766
## `Accelerations Zone 4 (num)` -2.1002    45.8353  -0.046   0.964
## `Accelerations Zone 5 (num)` 25.1804   113.6642   0.222   0.827
## `Decelerations Zone 4 (num)`  4.2319    62.1100   0.068   0.946
## `Body Impacts Grade 2 (num)` 17.5638    17.0303   1.031   0.315
## `Body Impacts Grade 3 (num)` 32.8372   282.8975   0.116   0.909
##
## Residual standard error: 22.82 on 20 degrees of freedom
## Multiple R-squared:  0.1884, Adjusted R-squared:  -0.3391
## F-statistic: 0.3572 on 13 and 20 DF,  p-value: 0.9694
confint(full1.15)[c(13, 5, 9, 8, 2), ]

##              2.5 %      97.5 %
## `Body Impacts Grade 2 (num)` -17.960732  53.088369
## `Work Recovery Ratio`1:2    -331.752487  523.270926
## `Accelerations Zone 3 (num)` -177.582222  237.749554
## `Hi Intensity Effort (num)`  -2.094467   1.590103
## `Duration Speed Hi-Inten (s)` -159.821162  210.906307
```

These five models suggest that the most important GPS variables for a fullback are, beginning from the most important:

Body Impacts Grade 2 (num) | +17.6, p-value = 0.315

Every additional Grade 2 body impact a fullback performs contributes between -18.0 and +53.1 points(!) to the win margin

Work Recovery Ratio | 1:2 -> +95.8, p-value = 0.645

If a fullback has a Work Recovery Ratio of 1:2 in a match, it will contribute between -331.8 and +523.3 points(!) to the win margin

Accelerations Zone 3 (num) | +30.1, p-value = 0.766

Every additional acceleration a fullback performs in Acceleration Zone 3 contributes between -177.5 and +237.7 points(!) to the win margin

Hi Intensity Effort (num) | -0.252, p-value = 0.778

For every additional effort a fullback performs that falls under any of the five high intensity categories (Hi-Int Sprints, Hi-Int Accelerations, Hi-Int Decelerations, Body Impacts and Jumps), between -2.094 and +1.590 points are added to the win margin

Duration Speed Hi-Inten (s) | +25.5, p-value = 0.777

Every additional second a fullback spends above the high intensity speed benchmark contributes between -159.8 and +210.9 points(!) to the win margin

## Random forest model

Using the `caret` package in conjunction with the `ranger` package, random forest models can be fitted on the data that was already cleaned of variables to remove singularities. It also has the capability of performing  $k$ -fold cross-validation; here,  $k = 10$ .

Permutation importance is used as the variable importance measure, as it generally performs better than Gini impurity or Actual Impurity Reduction (AIR) importance. Permutation importance determines a variable's importance by measuring the amount of error that is created when the values of that variable are randomly permuted. A larger error created in this scenario is indicative of greater variable importance.

```
library(ranger)

## Warning: package 'ranger' was built under R version 4.0.5

library(janitor)

## Warning: package 'janitor' was built under R version 4.0.5

##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test

# Setting up the cross-validation conditions
ctrl <- trainControl(method = "cv",
                     number = 10,
                     savePredictions = TRUE)
```

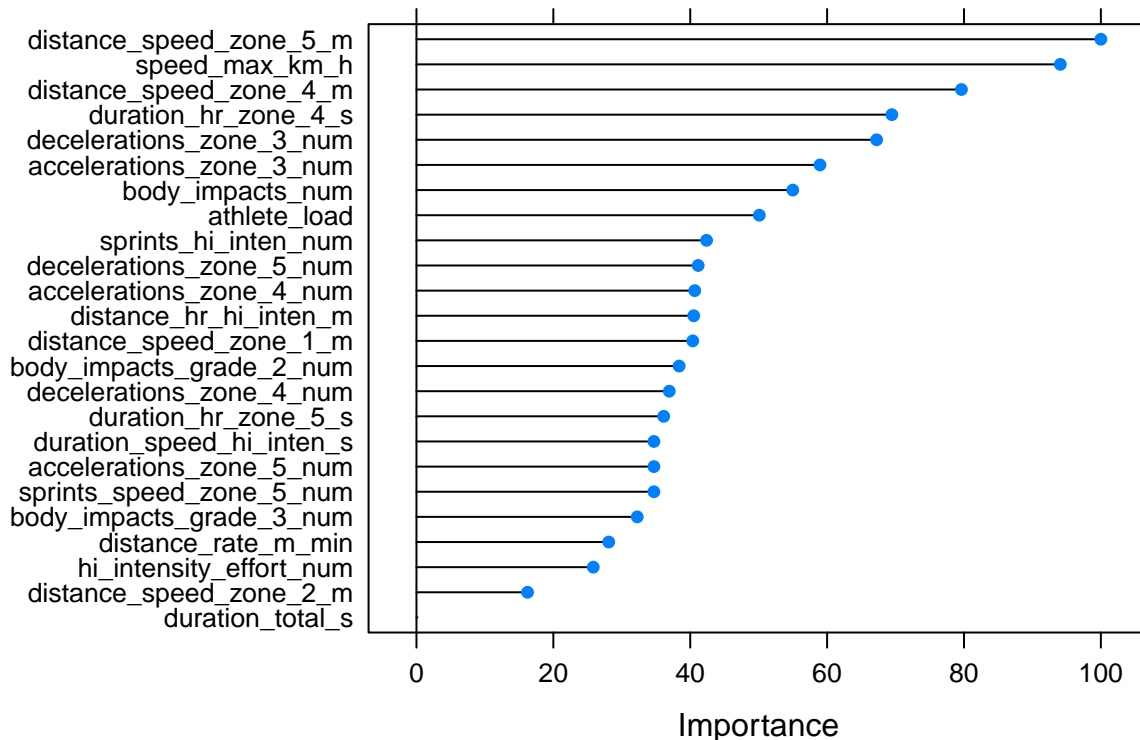
### Position 1: Loosehead prop

```
set.seed(1)

# Fitting the random forest model
ranger2.1 <- train(margins ~ .,
                  data = clean_names(pos1data),
                  method = "ranger",
                  importance = "permutation",
                  trControl = ctrl,
                  verbose = TRUE)

# Plotting variable importance
plot(varImp(ranger2.1), main = "Random Forest Variable Importance for Loosehead Props")
```

## Random Forest Variable Importance for Loosehead Props



According to this variable importance plot, the top 5 variables for a loosehead prop by permutation importance are:

Distance Speed Zone 5 (m)

Speed Max (km/h)

Distance Speed Zone 4 (m)

Duration HR Zone 4 (s)

Decelerations Zone 3 (num)

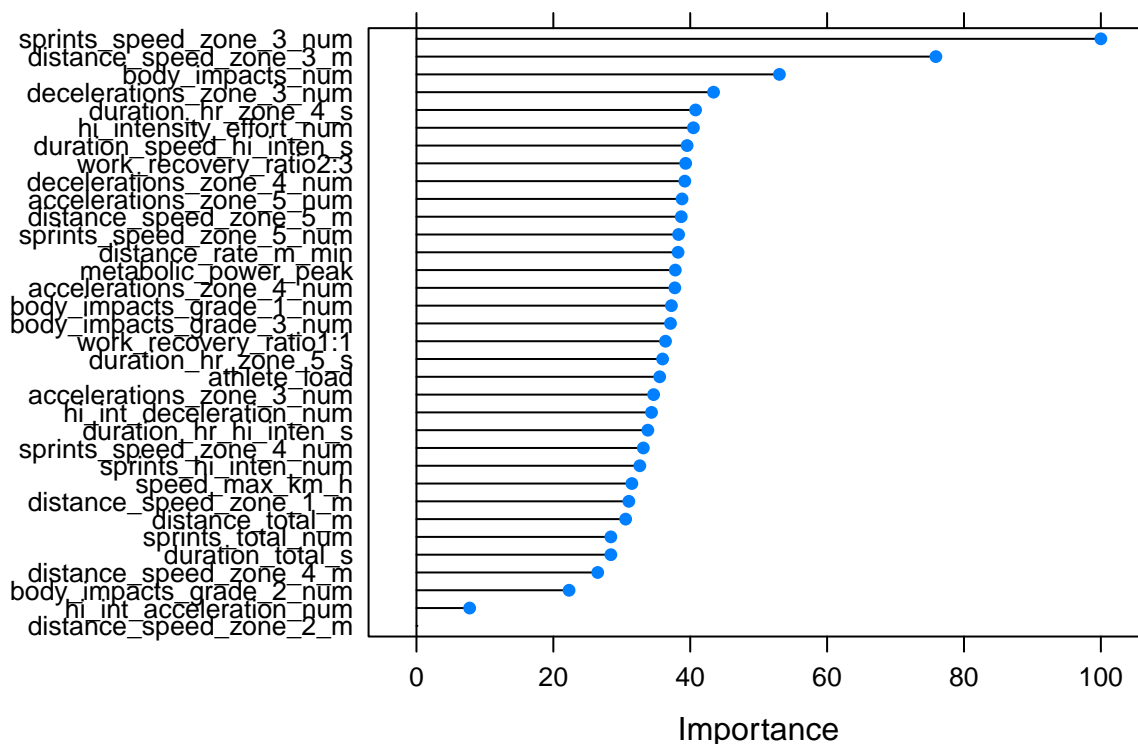
Comparing this with the top 5 variables from backward stepwise selection, #1, #2 and #4 are all represented in the random forest model as the most, 3rd-most and 5th-most important variables respectively. #3 Duration Speed Hi-Inten (s) is considered the 17th-most important variable in this random forest model, while #5 Accelerations Zone 3 (num) is considered the 6th-most important here.

### Position 2: Hooker

```
set.seed(1)

ranger2.2 <- train(margins ~ .,
  data = clean_names(pos2data),
  method = "ranger",
  importance = "permutation",
  trControl = ctrl,
  verbose = TRUE)
plot(varImp(ranger2.2), main = "Random Forest Variable Importance for Hookers")
```

## Random Forest Variable Importance for Hookers



According to this variable importance plot, the top 5 variables for a hooker by permutation importance are:

Sprints Speed Zone 3 (num)

Distance Speed Zone 3 (m)

Body Impacts (num)

Decelerations Zone 3 (num)

Duration HR Zone 4 (s)

Comparing this with the top 5 variables from backward stepwise selection, only one of the top 5 is found in this random forest model (#5 Distance Speed Zone 3 (m) at 2nd-most important). For the hooker, this is probably a better selection of different variables that are important, as opposed to the backward stepwise selection, which determined that distance measures are more important than all others.

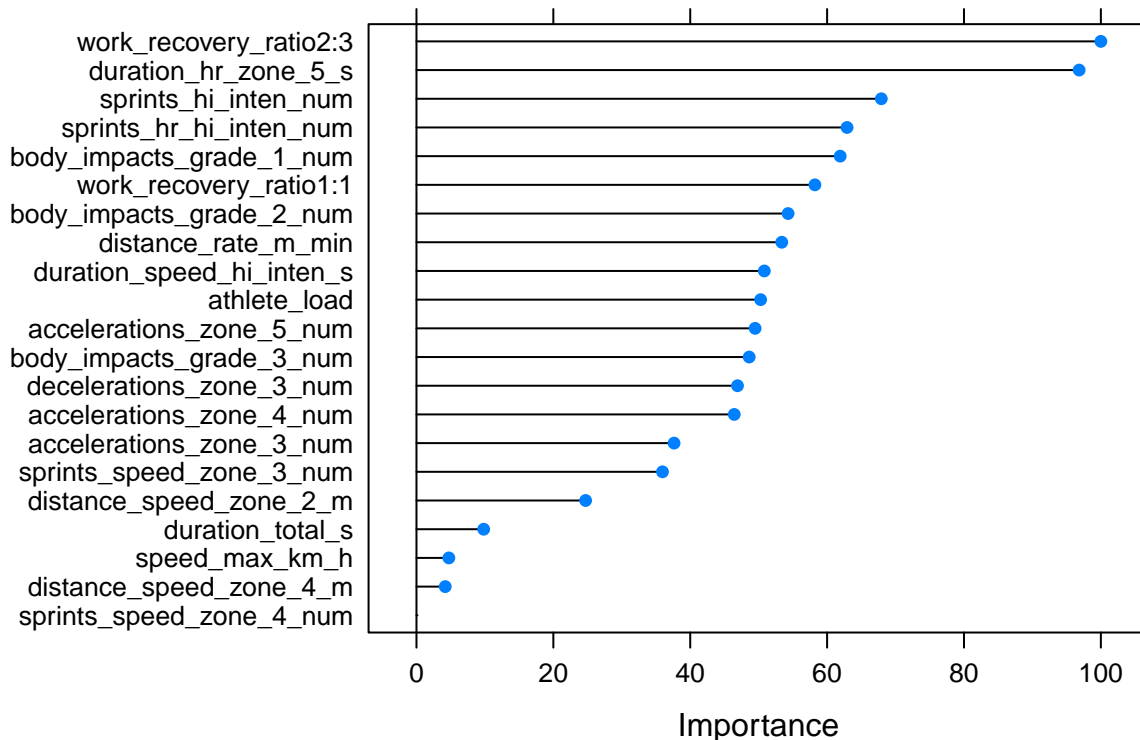
### Position 3: Tighthead prop

```
set.seed(1)

ranger2.3 <- train(margins ~ .,
                   data = clean_names(pos3data),
                   method = "ranger",
                   importance = "permutation",
                   trControl = ctrl,
                   verbose = TRUE)

plot(varImp(ranger2.3), main = "Random Forest Variable Importance for Tighthead Props")
```

## Random Forest Variable Importance for Tighthead Props



According to this variable importance plot, the top 5 variables for a tighthead prop by permutation importance are:

Work Recovery Ratio | 2:3

Duration HR Zone 5 (s)

Sprints Hi-Inten (num)

Sprints HR Hi-Inten (num)

Body Impacts Grade 1 (num)

Comparing this with the top 5 variables from backward stepwise selection, #1 and #2 are represented in this random forest model as the 4th-most and most important variables respectively. #3 **Duration Total** (s) is considered the 18th-most important variable in this random forest model, while #4 **Accelerations Zone 5** (num) is considered the 11th-most important variable and #5 **Speed Max** (km/h) is considered the 19th-most important variable. Of note, the dummy variable **Work Recovery Ratio** | 1:1 was considered among the top 5 variables from backward stepwise selection, and is found to be the 6th-most important variable in this random forest model.

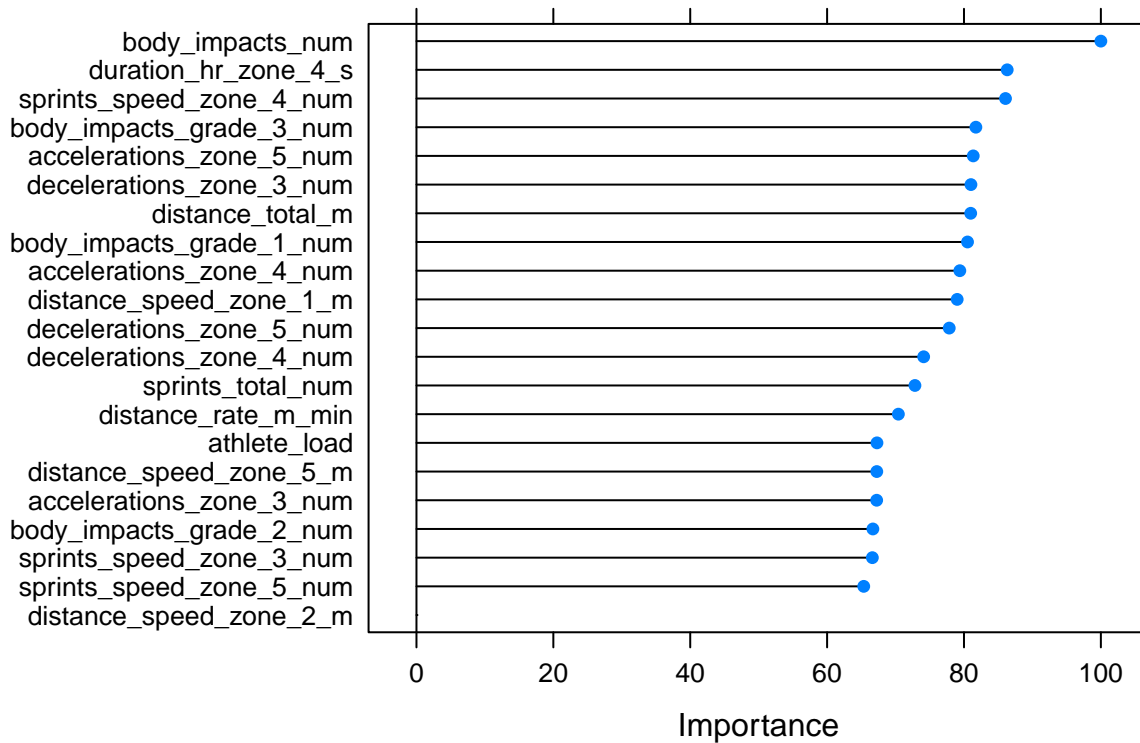
### Position 4: Left lock

```
set.seed(1)

ranger2.4 <- train(margins ~ .,
                   data = clean_names(pos4data),
                   method = "ranger",
                   importance = "permutation",
```

```
trControl = ctrl,
verbose = TRUE)
plot(varImp(ranger2.4), main = "Random Forest Variable Importance for Left Locks")
```

## Random Forest Variable Importance for Left Locks



According to this variable importance plot, the top 5 variables for a left lock by permutation importance are:

Body Impacts (num)

Duration HR Zone 4 (s)

Sprints Speed Zone 4 (num)

Body Impacts Grade 3 (num)

Accelerations Zone 5 (num)

Comparing this with the top 5 variables from backward stepwise selection, no variables are shared between the two methods. #1 Distance Speed Zone 1 (m) is considered the 10th-most important variable in the random forest model. #2 Distance Total (m) is found to be the 7th-most important variable, #3 Distance Speed Zone 2 (m) the 21st-most important variable (or least in this subset of variables), #4 Athlete Load the 15th-most important variable and #5 Sprints Total (num) the 13th-most important variable in this random forest model.

### Position 5: Right lock

```
set.seed(1)

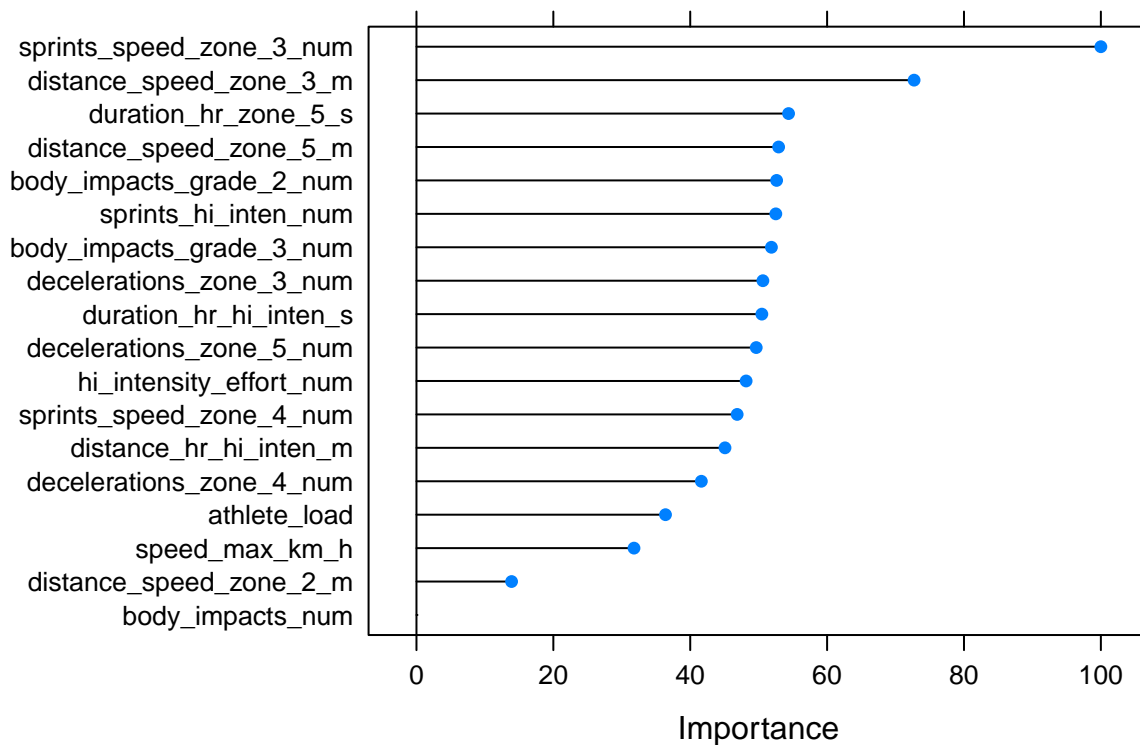
ranger2.5 <- train(margins ~ .,
  data = clean_names(pos5data),
```

```

method = "ranger",
importance = "permutation",
trControl = ctrl,
verbose = TRUE)
plot(varImp(ranger2.5), main = "Random Forest Variable Importance for Right Locks")

```

## Random Forest Variable Importance for Right Locks



According to this variable importance plot, the top 5 variables for a right lock by permutation importance are:

Sprints Speed Zone 3 (num)

Distance Speed Zone 3 (m)

Duration HR Zone 5 (s)

Distance Speed Zone 5 (m)

Body Impacts Grade 2 (num)

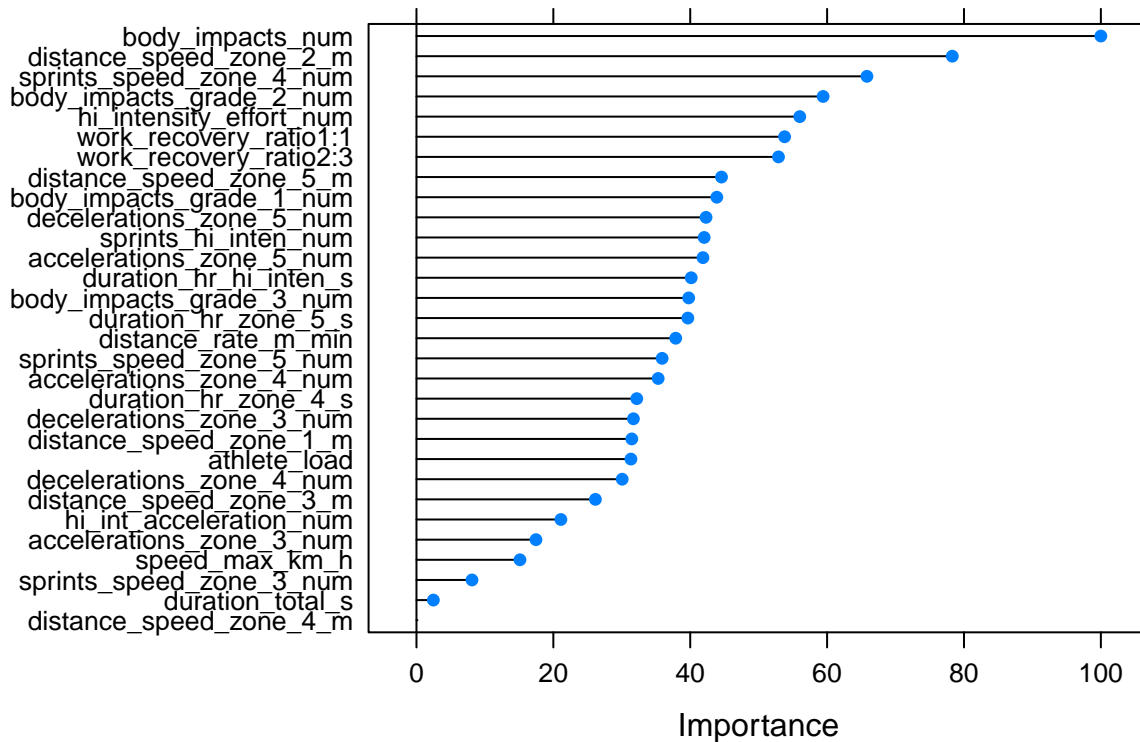
Comparing this with the top 5 variables from backward stepwise selection, #2 Sprints Speed Zone 3 (num) is most important, and #4 Body Impacts Grade 2 (num) is 5th-most important in this random forest model. #1 Sprints Hi-Inten (num) is considered the 6th-most important, #3 Decelerations Zone 3 (num) is considered the 8th-most important, and #5 Hi Intensity Effort (num) is considered the 11th-most important.

### Position 6: Blindside flanker

```
set.seed(1)
```

```
ranger2.6 <- train(margins ~ .,
  data = clean_names(pos6data),
  method = "ranger",
  importance = "permutation",
  trControl = ctrl,
  verbose = TRUE)
plot(varImp(ranger2.6), main = "Random Forest Variable Importance for Blindside Flankers")
```

## Random Forest Variable Importance for Blindside Flankers



According to this variable importance plot, the top 5 variables for a blindside flanker by permutation importance are:

Body Impacts (num)

Distance Speed Zone 2 (m)

Sprints Speed Zone 4 (num)

Body Impacts Grade 2 (num)

Hi Intensity Effort (num)

Comparing this with the top 5 variables from backward stepwise selection, the top 2 variables **Distance Speed Zone 2 (m)** and **Sprints Speed Zone 4 (num)** are present on the top 5 list for the random forest model as the second-most and third-most important variables. #3 **Speed Max (km/h)** is the 27th-most important, #4 **Decelerations Zone 4 (num)** is the 23rd-most important and #5 **Work Recovery Ratio | 2:3** is the 7th-most important variable in this random forest model.



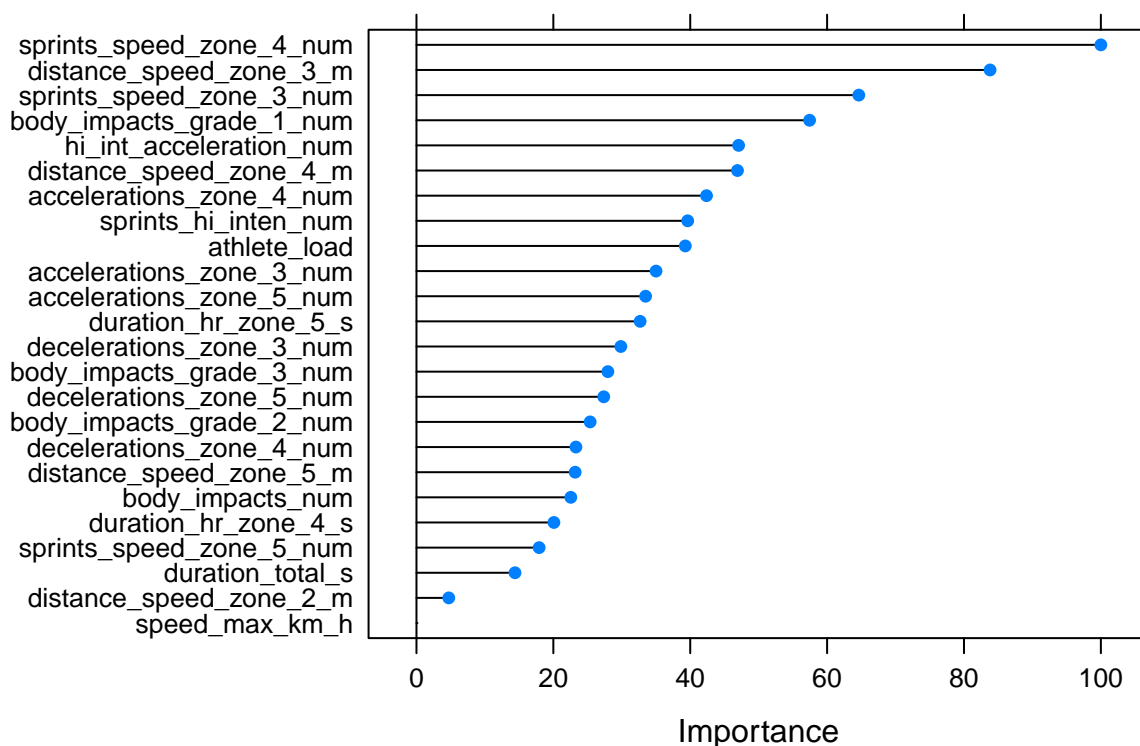
## Position 7: Openside flanker

```
set.seed(1)

ranger2.7 <- train(margins ~ .,
  data = clean_names(pos7data),
  method = "ranger",
  importance = "permutation",
  trControl = ctrl,
  verbose = TRUE)

plot(varImp(ranger2.7), main = "Random Forest Variable Importance for Openside Flankers")
```

### Random Forest Variable Importance for Openside Flankers



According to this variable importance plot, the top 5 variables for an openside flanker by permutation importance are:

Sprints Speed Zone 4 (num)

Distance Speed Zone 3 (m)

Sprints Speed Zone 3 (num)

Body Impacts Grade 1 (num)

Hi Int Acceleration (num)

Comparing this with the top 5 variables from backward stepwise selection, only #2 Sprints Speed Zone 4 (num) is present on the top 5 list for the random forest model, appearing as the most-important variable. #1 Accelerations Zone 5 (num) is the 11th-most important, #3 Body Impacts Grade 3 (num) is the 14th-most important, #4 Duration HR Zone 4 (s) is the 20th-most important, and #5 Body Impacts

Grade 2 (num) is the 16th-most important.

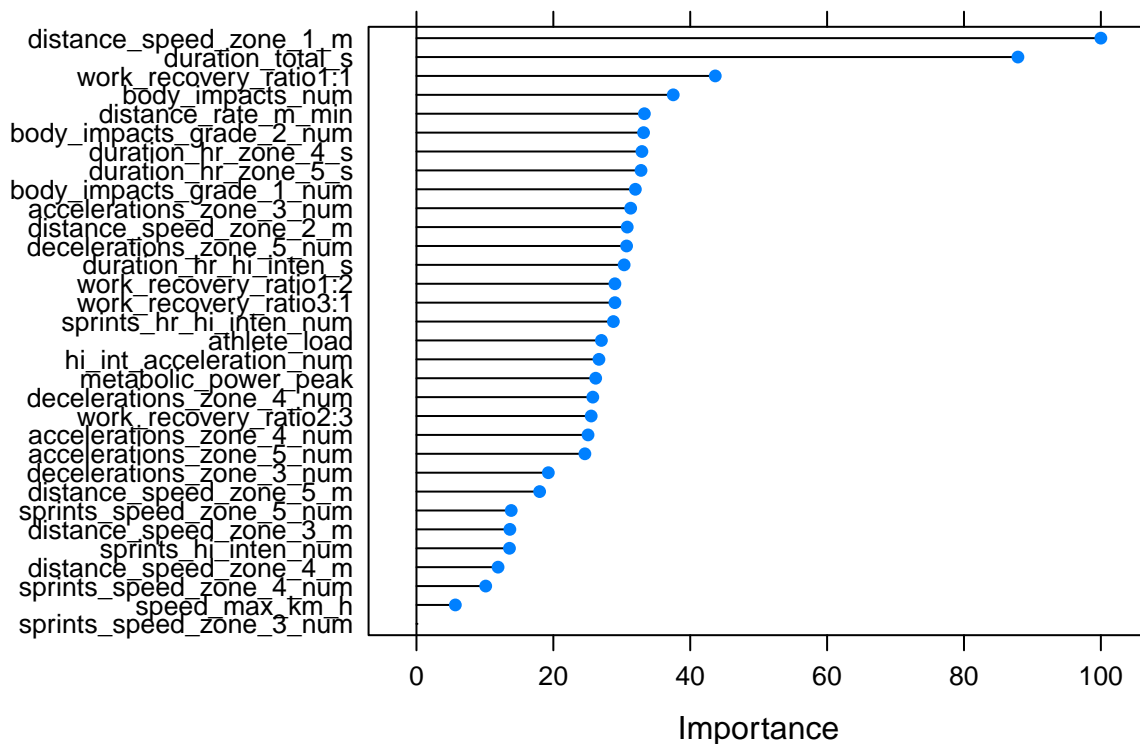
#### Position 8: Number 8

```
set.seed(1)

ranger2.8 <- train(margins ~ .,
                   data = clean_names(pos8data),
                   method = "ranger",
                   importance = "permutation",
                   trControl = ctrl,
                   verbose = TRUE)

plot(varImp(ranger2.8), main = "Random Forest Variable Importance for Number 8s")
```

### Random Forest Variable Importance for Number 8s



According to this variable importance plot, the top 5 variables for a number 8 by permutation importance are:

Distance Speed Zone 1 (m)

Duration Total (s)

Work Recovery Ratio | 1:1

Body Impacts (num)

Distance Rate (m/min)

Comparing this with the top 5 variables from backward stepwise selection, no variables are shared between the two methods. #1 Duration HR Zone 4 (s) is considered the 7th-most important, #2 Hi Int Acceleration (num) the 18th-most important, #3 Distance Speed Zone 2 (m) the 11th-most important, #4 Sprints

Speed Zone 3 (num) the 32nd-most important (or least in this subset of variables), and #5 Distance Speed Zone 3 (m) the 27th-most important.

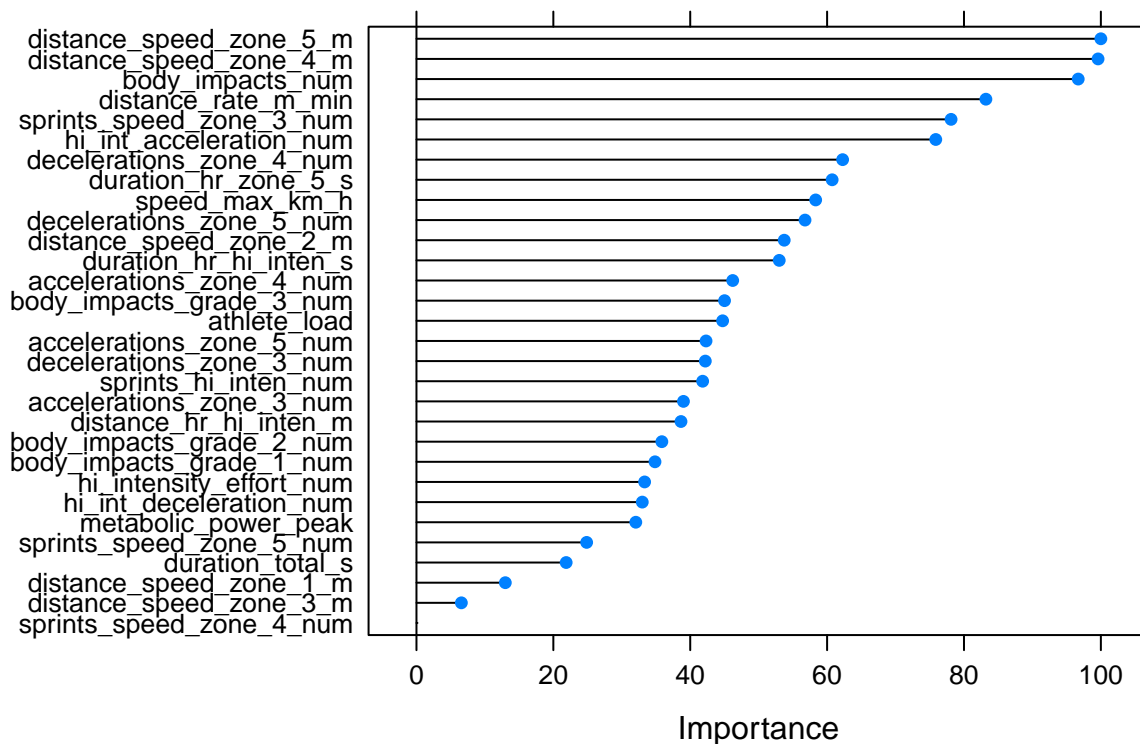
### Position 9: Scrum-half

```
set.seed(1)

ranger2.9 <- train(margins ~ .,
  data = clean_names(pos9data),
  method = "ranger",
  importance = "permutation",
  trControl = ctrl,
  verbose = TRUE)

plot(varImp(ranger2.9), main = "Random Forest Variable Importance for Scrum-halves")
```

### Random Forest Variable Importance for Scrum-halves



According to this variable importance plot, the top 5 variables for a scrum-half by permutation importance are:

Distance Speed Zone 5 (m)

Distance Speed Zone 4 (m)

Body Impacts (num)

Distance Rate (m/min)

Sprints Speed Zone 3 (num)

Comparing this with the top 5 variables from backward stepwise selection, only #3 Sprints Speed Zone 3

(num) is shared, being the 5th-most important variable in this random forest model. #1 Decelerations Zone 4 (num) is considered the 7th-most important variable, #2 Distance Speed Zone 3 (m) the 29th-most important (and second-least in this subset of variables), #4 Duration HR Hi-Inten (s) the 12th-most important, and #5 Decelerations Zone 5 (num) the 10th-most important.

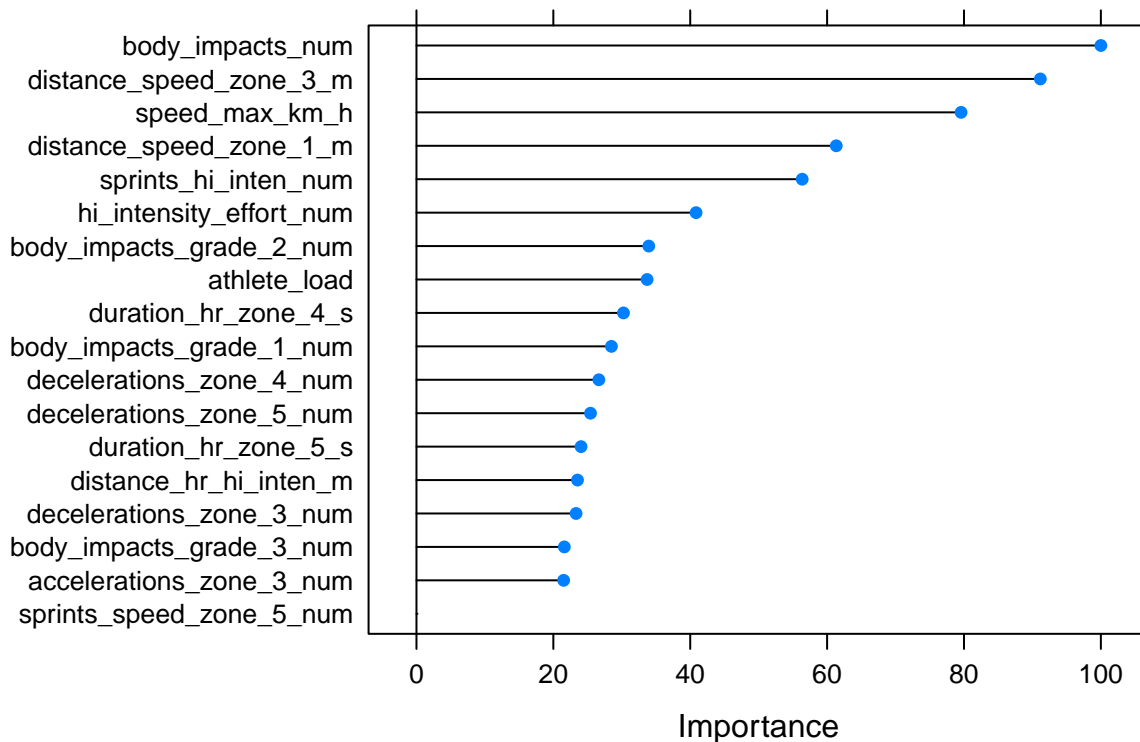
### Position 10: Fly-half

```
set.seed(1)

ranger2.10 <- train(margins ~ .,
                    data = clean_names(pos10data),
                    method = "ranger",
                    importance = "permutation",
                    trControl = ctrl,
                    verbose = TRUE)

plot(varImp(ranger2.10), main = "Random Forest Variable Importance for Fly-halves")
```

### Random Forest Variable Importance for Fly-halves



According to this variable importance plot, the top 5 variables for a fly-half by permutation importance are:

Body Impacts (num)

Distance Speed Zone 3 (m)

Speed Max (km/h)

Distance Speed Zone 1 (m)

Sprints Hi-Inten (num)

Comparing this with the top 5 variables from backward stepwise selection, #1 Sprints Hi-Inten (num), #4 Distance Speed Zone 3 (m) and #5 Body Impacts (num) appear in the top 5 for the random forest model (at 5th-most important, 2nd-most important and most important respectively). #2 Decelerations Zone 3 (num) is considered the 15th-most important variable, and #3 Decelerations Zone 4 (num) is considered the 11th-most important.

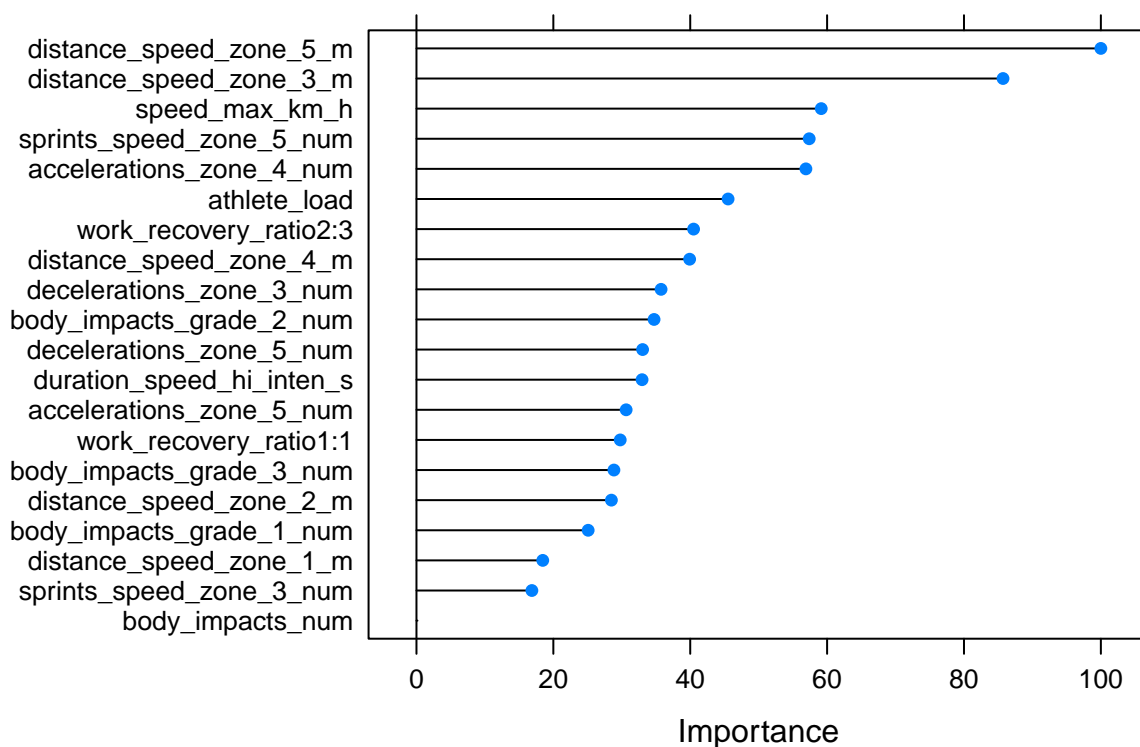
#### Position 11: Left wing

```
set.seed(1)

ranger2.11 <- train(margins ~ .,
                    data = clean_names(pos11data),
                    method = "ranger",
                    importance = "permutation",
                    trControl = ctrl,
                    verbose = TRUE)

plot(varImp(ranger2.11), main = "Random Forest Variable Importance for Left Wings")
```

### Random Forest Variable Importance for Left Wings



According to this variable importance plot, the top 5 variables for a left wing by permutation importance are:

- Distance Speed Zone 5 (m)
- Distance Speed Zone 3 (m)
- Speed Max (km/h)
- Sprints Speed Zone 5 (num)
- Accelerations Zone 4 (num)

Comparing this with the top 5 variables from backward stepwise selection, #1 Distance Speed Zone 3 (m) and #5 Distance Speed Zone 5 (m) are present on the top 5 for the random forest model, at 2nd-most and most important. #2 Work Recovery Ratio | 2:3 is considered the 7th-most important variable, #3 Decelerations Zone 3 the 9th-most important, and #4 Decelerations Zone 5 the 11th-most important.

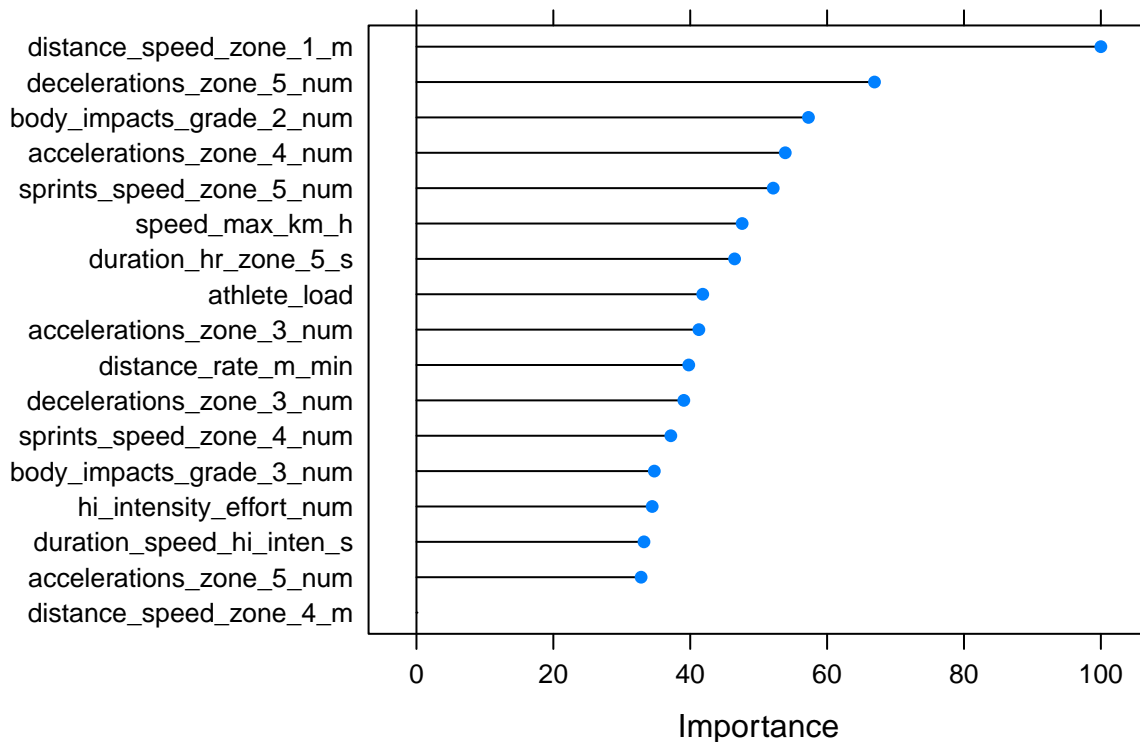
#### Position 12: Inside centre

```
set.seed(1)

ranger2.12 <- train(margins ~ .,
                    data = clean_names(pos12data),
                    method = "ranger",
                    importance = "permutation",
                    trControl = ctrl,
                    verbose = TRUE)

plot(varImp(ranger2.12), main = "Random Forest Variable Importance for Inside Centres")
```

### Random Forest Variable Importance for Inside Centres



According to this variable importance plot, the top 5 variables for an inside centre by permutation importance are:

- Distance Speed Zone 1 (m)
- Decelerations Zone 5 (num)
- Body Impacts Grade 2 (num)
- Accelerations Zone 4 (num)
- Sprints Speed Zone 5 (num)

Comparing this with the top 5 variables from backward stepwise selection, #1 Distance Speed Zone 1 (m), #4 Body Impacts Grade 2 (num) and #5 Decelerations Zone 5 (num) are present on the top 5 for the random forest model, at most, 3rd-most and 2nd-most important. #2 Athlete Load is considered the 8th-most important variable, and #3 Sprints Speed Zone 4 (num) the 12th-most important.

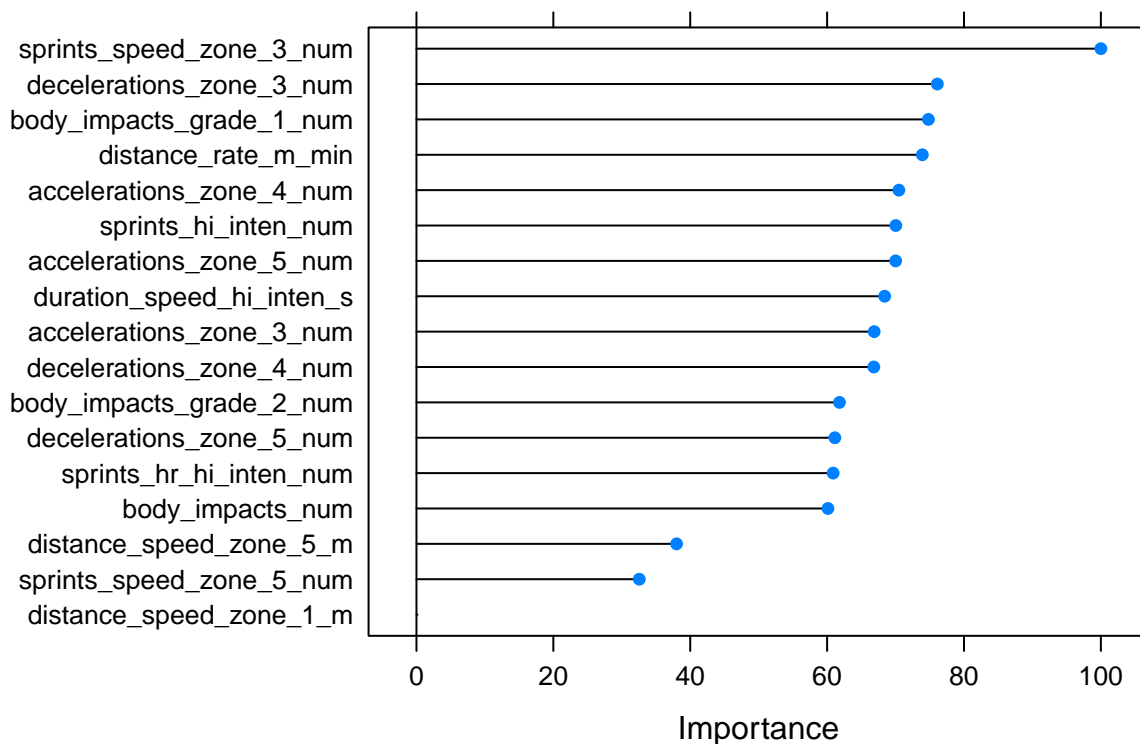
### Position 13: Outside centre

```
set.seed(1)

ranger2.13 <- train(margins ~ .,
                    data = clean_names(pos13data),
                    method = "ranger",
                    importance = "permutation",
                    trControl = ctrl,
                    verbose = TRUE)

plot(varImp(ranger2.13), main = "Random Forest Variable Importance for Outside Centres")
```

### Random Forest Variable Importance for Outside Centres



According to this variable importance plot, the top 5 variables for an outside centre by permutation importance are:

- Sprints Speed Zone 3 (num)
- Decelerations Zone 3 (num)
- Body Impacts Grade 1 (num)
- Distance Rate (m/min)
- Accelerations Zone 4 (num)

Comparing this with the top 5 variables from backward stepwise selection, #1 Sprints Speed Zone 3 (num), #3 Decelerations Zone 3 (num), #4 Accelerations Zone 4 (num) and #5 Body Impacts Grade 1 (num) are present on the top 5 for the random forest model, at most, 2nd-most, 5th-most and 3rd-most important. #2 Body Impacts (num) is considered the 14th-most important variable here.

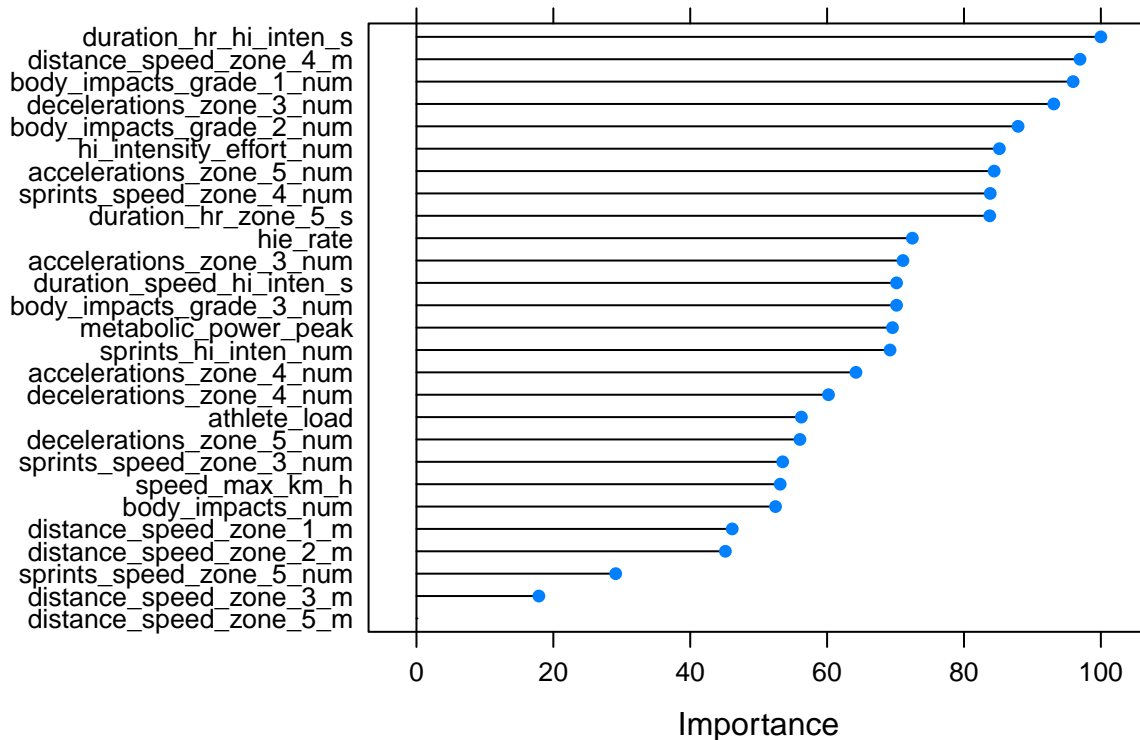
#### Position 14: Right wing

```
set.seed(1)

ranger2.14 <- train(margins ~ .,
                    data = clean_names(pos14data),
                    method = "ranger",
                    importance = "permutation",
                    trControl = ctrl,
                    verbose = TRUE)

plot(varImp(ranger2.14), main = "Random Forest Variable Importance for Right Wings")
```

### Random Forest Variable Importance for Right Wings



According to this variable importance plot, the top 5 variables for an right wing by permutation importance are:

Duration HR Hi-Inten (s)

Distance Speed Zone 4 (m)

Body Impacts Grade 1 (num)

Decelerations Zone 3 (num)

Body Impacts Grade 2 (num)



Comparing this with the top 5 variables from backward stepwise selection, only #4 Body Impacts Grade 2 (num) made the top 5 for the random forest model, at 5th-most important. #1 HIE Rate is considered the 10th-most important variable, #2 Body Impacts Grade 3 (num) the 13th-most important, #3 Duration HR Zone 5 (s) the 9th-most important, and #5 Accelerations Zone 4 (num) the 16th-most important.

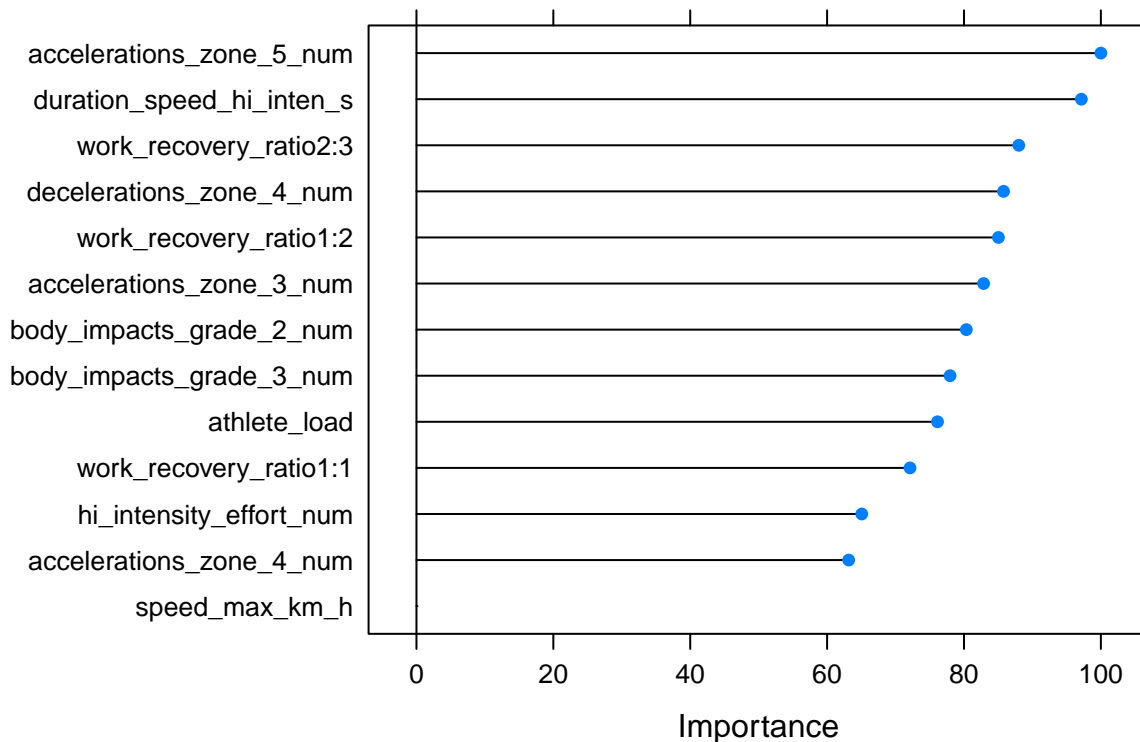
#### Position 15: Fullback

```
set.seed(1)

ranger2.15 <- train(margins ~ .,
                    data = clean_names(pos15data),
                    method = "ranger",
                    importance = "permutation",
                    trControl = ctrl,
                    verbose = TRUE)

plot(varImp(ranger2.15), main = "Random Forest Variable Importance for Fullbacks")
```

### Random Forest Variable Importance for Fullbacks



According to this variable importance plot, two dummy variables for Work Recovery Ratio are in the top 5 variables by permutation importance. So instead, the top 6 variables are taken, to obtain five unique variables. The top 6 variables for a fullback by permutation importance are:

Accelerations Zone 5 (num)  
 Duration Speed Hi-Inten (s)  
 Work Recovery Ratio | 2:3  
 Decelerations Zone 4 (num)

Work Recovery Ratio | 1:2

Accelerations Zone 3 (num)

Comparing this with the top 5 variables from backward stepwise selection, #2 Work Recovery Ratio | 1:2, #3 Accelerations Zone 3 (num) and #5 Duration Speed Hi-Inten (s) are present in the top 6 for the random forest model, at 5th-most, 6th-most and 2nd-most important. #1 Body Impacts Grade 2 (num) is considered the 7th-most important variable, and #4 Hi Intensity Effort (num) the 11th-most important.

## Concluding thoughts

The top five variables for the front row appear to be dominated by acceleration, deceleration and distance measures. Body impact measures are also considered important. Distance measures being important, particularly in Speed Zones, is interesting, considering the front row is not necessarily expected to make quick long runs.

The top five variables for the back row appear to be dominated by body impact measures, sprints, speed, acceleration and deceleration measures. Body impacts being considered important is expected, since the back row are a good combination of size, physicality and speed, and so are able to make more tackles against larger opponents.

The top five variables for the halves appear to be dominated by speed, sprints, distance and body impact measures. Scrum-half surprisingly registered several data points of Sprints Hi-Inten (num) that are all higher than every other value, skewing the distribution significantly.

The top five variables for the centres appear to be dominated by body impact measures, sprints, speed, acceleration and deceleration measures. Inside centre, in particular was found to be particularly strong for body impact, acceleration and deceleration measures - higher distribution than back row players for body impacts, and higher distribution centres than some of the wings and fullback acceleration and deceleration measures.

The top five variables for the wings and fullback appear to be speed, acceleration and deceleration measures. This makes sense, since they are expected to cover large distances very quickly. The acceleration and deceleration measures being considered important may have to do with their ability to sidestep players during their runs.

I wanted to apply XGBoost, lasso regression and possibly elastic net models to this data, but initial testing presented me with errors in running the code. I estimated that the time it would take to troubleshoot these would take much longer than I can afford, unfortunately. I have enjoyed getting to work with this data, however, pushing my personal skills beyond what I've done in my courses.

Thank you, Auckland Rugby, for this opportunity to work with real world data in an industry setting. It has been valuable experience, and I hope you are satisfied with what has been presented here.