NHL Goalie and Skater Heat Maps

Sasank Vishnubhatla January 10, 2019

Last Update: 2019-01-15 14:01:20

GitHub Repository: sv4u/goalie-and-skater-heat-maps

Introduction

In the NHL, players have their *sweet-spots*. For skaters, it can be the top of a certain circle, or right in-front of the net. For goalies, it could be where their vision is best and where they have the best angle to cut down a shot. All players have these sweet-spots, but it is difficult to analytically say where they are. By using shot location data, we can determine these locations and create models show where goalies and skaters need improvement and where they succeed.

Before we jump in, let's clean up our R environment and also load in some libraries we will be using.

```
rm(list = ls())
library(purrr)
library(ggplot2)
```

Data Formatting

To start, we need to read in our data. Our data is formatted nicely in CSV format. We have data from the, 2015-2016 season, 2016-2017 season, 2017-2018 season, and 2018-2019 season (up to 1/7/19). This data was downloaded from MoneyPuck. Let's first start by loading in all three seasons of data:

```
data.2015 = read.csv("data/2015.csv")
data.2016 = read.csv("data/2016.csv")
data.2017 = read.csv("data/2017.csv")
data.2018 = read.csv("data/2018.csv")
```

Note: this will take a *relatively* long time to compute as the datasets are large. Each dataset contains all shot data (**including** playoffs).

We'll only look at regular season data. The playoffs in the NHL are a beast of their own.

```
get.regular.season = function(data) {
    subset(data, isPlayoffGame == 0)
}
season.2015 = get.regular.season(data.2015)
season.2016 = get.regular.season(data.2016)
season.2017 = get.regular.season(data.2017)
season.2018 = get.regular.season(data.2018)
```

Now that we have our data, we can remove extraneous columns. Here is a table of what columns we are keeping, and what we are renaming them to:

Old Column	New Column
xCordAdjusted	x
yCordAdjusted	У
goal	goal
${\it shot}$ Angle Adjusted	angle
goalie Name For Shot	$goalie_name$
shooterName	$skater_name$
game_id	game

Now, here is the R code to do this subsetting of the original dataset.

Now, we have all the data we need.

Function Definitions

Generic

From our data, we can calculate some very important statistics like the following:

- Goal Percent: goals per total shots
- Save Percent: saves (total shots goals) per total shots
- Shots per Goal: total shots per goal

Additionally, we can break up our data by game. There are some generic functions we can write to help for both goalies and skaters. Let's write them now!

```
get.goal.percent = function(data) {
    shots = length(data$goal)
    temp = subset(data, goal == 1)
    goals = length(temp$goal)
    goals / shots
}

get.save.percent = function(data) {
    shots = length(data$goal)
    temp = subset(data, goal == 1)
    goals = length(temp$goal)
    (shots - goals) / shots
```

```
get.shots.per.goal = function(data) {
    shots = length(data$goal)
    temp = subset(data, goal == 1)
    goals = length(temp$goal)
    shots / goals
}
```

Note: when using get.shots.per.goal, if there were no goals scored, R will handle the division by zero by returning infinity. This will be problematic when graphing data. I am still working on a good solution to this problem. Earlier, I used 200 as a substitute value. However, 200 still skews graphs, which is unideal.

```
get.games = function(data) {
    unique(data$game)
}

get.single.game = function(data, game_id) {
    subset(data, game == game_id)
}

get.all.games = function(data) {
    games = get.games(data)
    Map(function(x) get.single.game(data, x), games)
}
```

Now, we can create our game by game statistic functions:

```
get.game.goal.percent = function(data) {
    gameframe = get.all.games(data)
    games.gp = map(gameframe, function(x) get.goal.percent(x))
    unlist(games.gp, use.names = FALSE)
}

get.game.save.percent = function(data) {
    gameframe = get.all.games(data)
    games.sp = map(gameframe, function(x) get.save.percent(x))
    unlist(games.sp, use.names = FALSE)
}

get.game.shots.per.goal = function(data) {
    gameframe = get.all.games(data)
    games.spg = map(gameframe, function(x) get.shots.per.goal(x))
    unlist(games.spg, use.names = FALSE)
}
```

Also, we'll need a function to get match-ups between a specific goalie and skater. Let's write that here, instead of in our goalies and our skaters sections.

```
get.matchup.data = function(data, goalie, skater) {
    subset(data, goalie_name == goalie & skater_name == skater)
}
```

We've now written our generic data handling functions.

Goalies

Let's first start with a function to get data for a specific goalie.

```
get.goalie.data = function(data, name) {
    subset(data, goalie_name == name)
}
```

Skaters

Let's first start with a function to get data for a specific skater.

```
get.skater.data = function(data, name) {
    subset(data, skater_name == name)
}
```

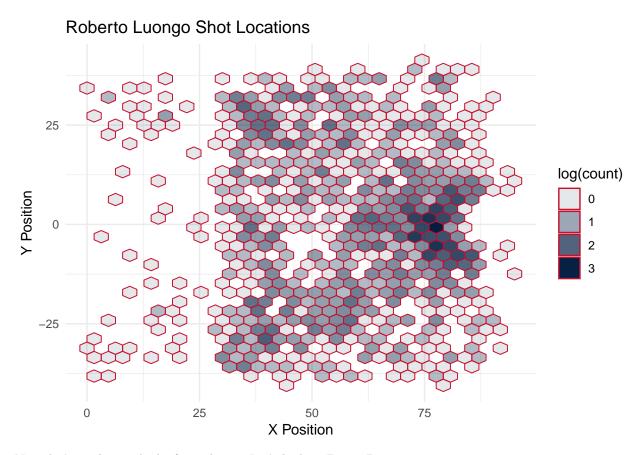
Graphing

Given specific data, we should be able to graph the location of shots. Let's write a function that uses ggplot to do so.

```
graph.shot.locations = function(data, primary, secondary, name) {
   plot = ggplot(data) +
        geom_hex(aes(x = x, y = y, alpha = log(..count..)), fill = primary, color = secondary) +
        labs(title = paste(name, "Shot Locations", sep = " "), x = "X Position", y = "Y Position") +
        theme_minimal()
   plot
}
```

To see a test of what this does, let's quickly make a graph of Roberto Luongo's shots against him.

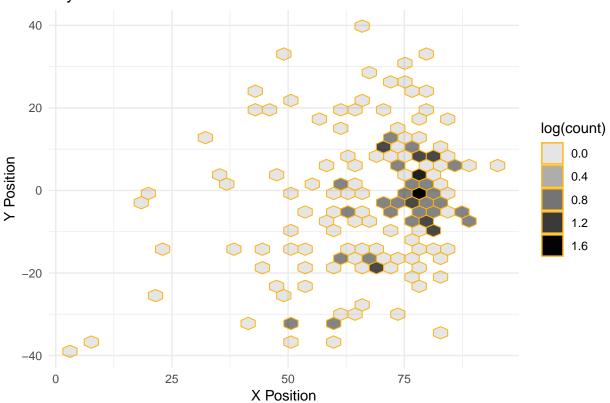
```
luongo = get.goalie.data(analysis.2017, "Roberto Luongo")
plot = graph.shot.locations(luongo, "#041E42", "#C8102E", "Roberto Luongo")
plot
```



Now, let's see how it looks for a skater. Let's look at Bryan Rust.

```
rust = get.skater.data(analysis.2017, "Bryan Rust")
plot = graph.shot.locations(rust, "#000000", "#FCB514", "Bryan Rust")
plot
```





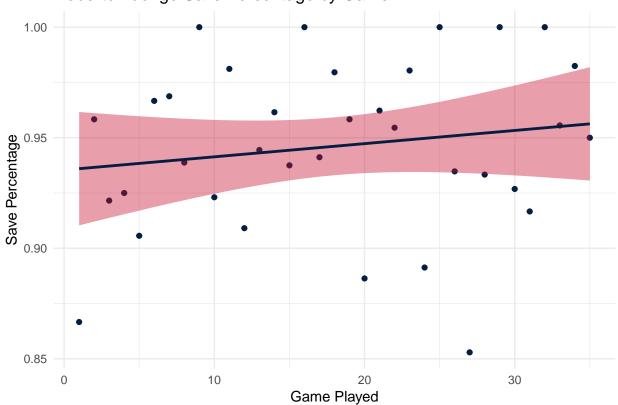
Furthermore, we should be able to graph trends in certain statistics.

```
graph.trend = function(trend, type, primary, secondary, name) {
    frame = data.frame(x = c(1:length(trend)), y = trend)
    disp = paste(name, type, "by Game", sep = " ")
    plot = ggplot(frame) +
        geom_point(aes(x = x, y = y), color = primary) +
        geom_smooth(aes(x = x, y = y), method = "lm", color = primary, fill = secondary) +
        labs(title = disp, x = "Game Played", y = type) +
        theme_minimal()
    plot
}
```

So, let's go back to Luongo and look at his save percentage per game.

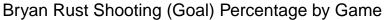
```
luongo.game.sp = get.game.save.percent(luongo)
plot = graph.trend(luongo.game.sp, "Save Percentage", "#041E42", "#C8102E", "Roberto Luongo")
plot
```

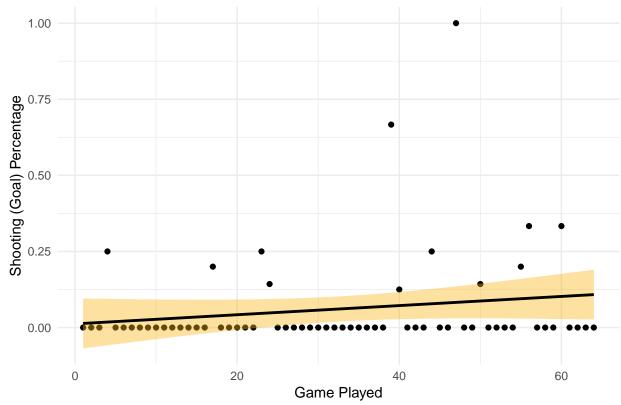




We can also see the shooter's perspective. Let's look at Rust's shooting percentage (goal percent).

```
rust.game.gp = get.game.goal.percent(rust)
plot = graph.trend(rust.game.gp, "Shooting (Goal) Percentage", "#000000", "#FCB514", "Bryan Rust")
plot
```





Analysis

Goalies

Matt Murray

Matt Murray is a 24 year old phenomenon who has already won 2 Stanley Cups. Let's take a look at how he's done it.

```
murray.2015 = get.goalie.data(analysis.2015, "Matt Murray")
murray.2016 = get.goalie.data(analysis.2016, "Matt Murray")
murray.2017 = get.goalie.data(analysis.2017, "Matt Murray")
murray.2018 = get.goalie.data(analysis.2018, "Matt Murray")
```

Let's start by calculating some of his stats for each season and then tabularizing.

```
murray.2015.sp = get.save.percent(murray.2015)
murray.2016.sp = get.save.percent(murray.2016)
murray.2017.sp = get.save.percent(murray.2017)
murray.2018.sp = get.save.percent(murray.2018)

murray.2015.gp = get.goal.percent(murray.2015)
murray.2016.gp = get.goal.percent(murray.2016)
murray.2017.gp = get.goal.percent(murray.2017)
murray.2018.gp = get.goal.percent(murray.2018)
```

```
murray.2015.spg = get.shots.per.goal(murray.2015)
murray.2016.spg = get.shots.per.goal(murray.2016)
murray.2017.spg = get.shots.per.goal(murray.2017)
murray.2018.spg = get.shots.per.goal(murray.2018)
```

Now, let's put it in a table:

Season	Save Percent	Goal Percent	Shots Per Goal
2015-2016	0.9460043	0.0539957	18.52
2016-2017	0.9429599	0.0570401	17.5315315
2017-2018	0.9301837	0.0698163	14.3233083
2018-2019	0.9374185	0.0625815	15.9791667

We can look at his game by game data also.

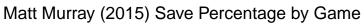
```
murray.2015.games.sp = get.game.save.percent(murray.2015)
murray.2015.games.gp = get.game.goal.percent(murray.2015)
murray.2015.games.spg = get.game.shots.per.goal(murray.2015)

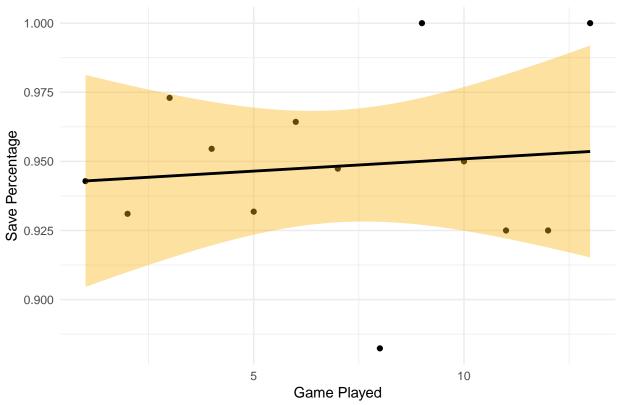
murray.2016.games.sp = get.game.save.percent(murray.2016)
murray.2016.games.gp = get.game.goal.percent(murray.2016)
murray.2016.games.spg = get.game.shots.per.goal(murray.2016)

murray.2017.games.sp = get.game.save.percent(murray.2017)
murray.2017.games.spg = get.game.shots.per.goal(murray.2017)
murray.2017.games.spg = get.game.shots.per.goal(murray.2017)
murray.2018.games.spg = get.game.save.percent(murray.2018)
murray.2018.games.spg = get.game.shots.per.goal(murray.2018)
murray.2018.games.spg = get.game.shots.per.goal(murray.2018)
```

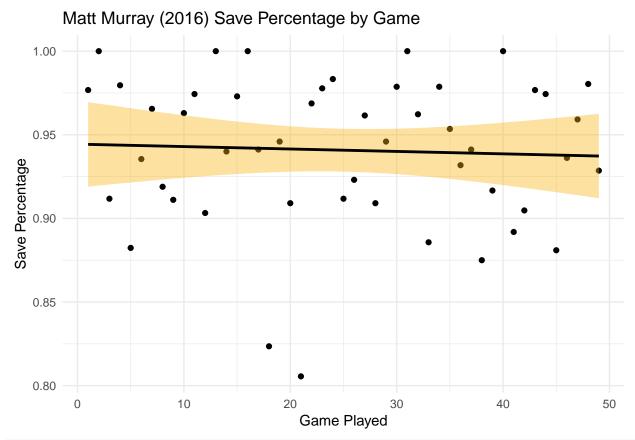
Let's look at his save percentage graphs:

```
murray.2015.sp.plot = graph.trend(murray.2015.games.sp, "Save Percentage", "#000000", "#FCB514", "Matt !
murray.2015.sp.plot
```

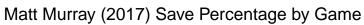


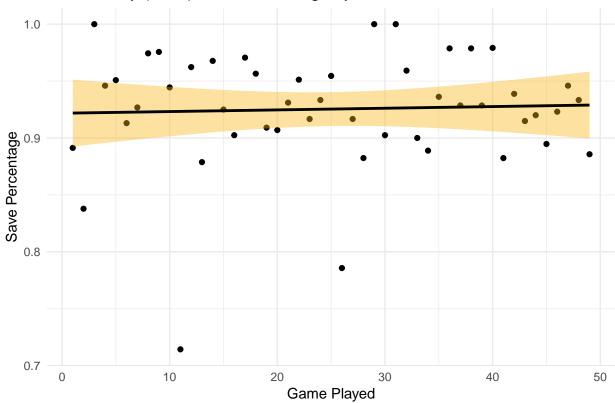


murray.2016.sp.plot = graph.trend(murray.2016.games.sp, "Save Percentage", "#000000", "#FCB514", "Matt I murray.2016.sp.plot

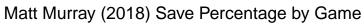


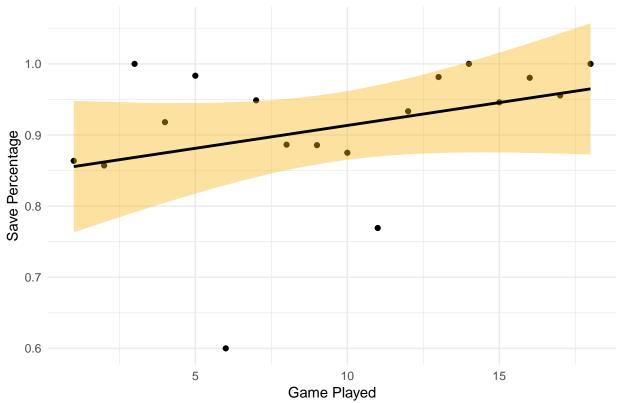
murray.2017.sp.plot = graph.trend(murray.2017.games.sp, "Save Percentage", "#000000", "#FCB514", "Matt I murray.2017.sp.plot





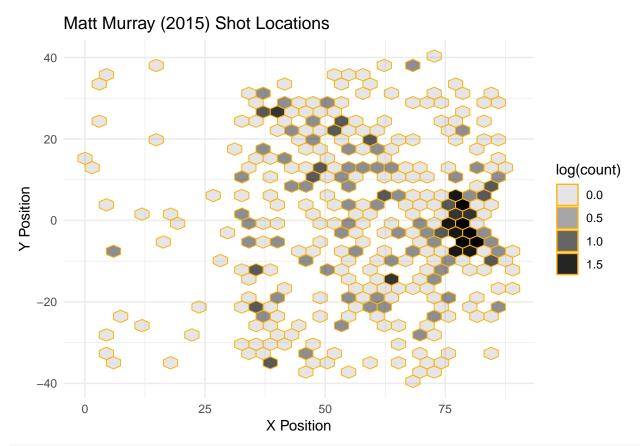
murray.2018.sp.plot = graph.trend(murray.2018.games.sp, "Save Percentage", "#000000", "#FCB514", "Matt I murray.2018.sp.plot



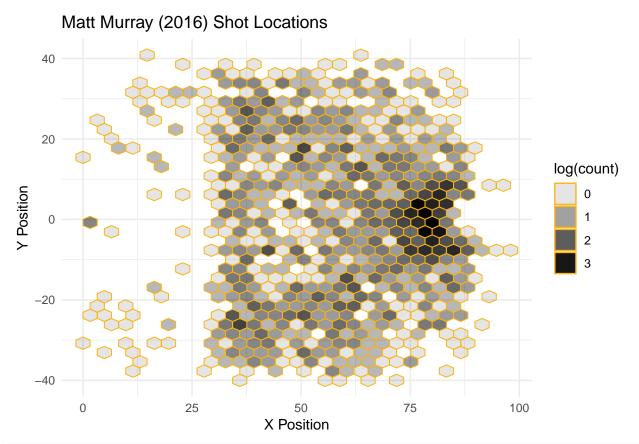


And now, let's take a look at his shot location data.

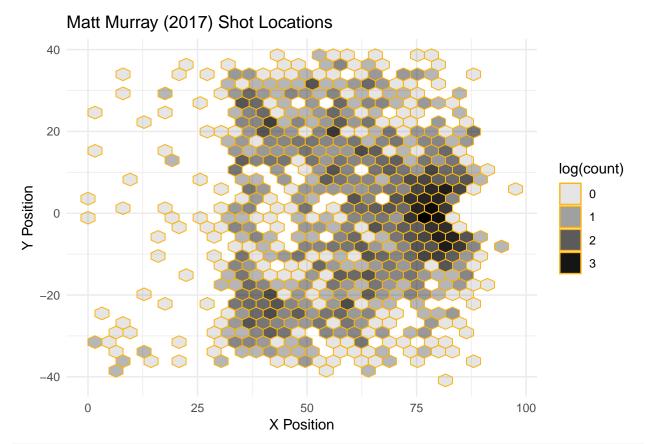
murray.2015.locations.plot = graph.shot.locations(murray.2015, "#000000", "#FCB514", "Matt Murray (2015
murray.2015.locations.plot



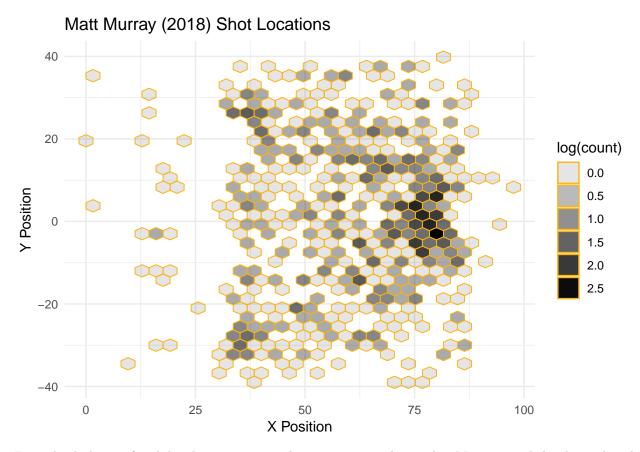
murray.2016.locations.plot = graph.shot.locations(murray.2016, "#000000", "#FCB514", "Matt Murray (2016
murray.2016.locations.plot



murray.2017.locations.plot = graph.shot.locations(murray.2017, "#000000", "#FCB514", "Matt Murray (2017
murray.2017.locations.plot



murray.2018.locations.plot = graph.shot.locations(murray.2018, "#000000", "#FCB514", "Matt Murray (2018
murray.2018.locations.plot



From the darkness of each hex bin, we can use that as a percent chance that Murray saved the shot. Though the darkness is not exactly the save percentage of the shots, it's the log chance of Murray seeing the shot. This log chance is a good likelihood function of Murray saving the shot. We can see that Murray's weakness is very close to him. Those shots are most likely to be tipped or deflected shots, not actual shots. We can see his weakness is the top of the circles, based on the X and Y coordinates.

From the graph, we can see that Murray has struggled with shots near the circles. Disregarding tips, it seems like that Murray is positionally sound with goals against near him. However, compared to the right wing, Murray seems weak against shots coming from the left wing side. If teams were to exploit this, it would be beneficial to shoot towards Murray's glove side from the left wing circle.

Casey DeSmith

On January 11th, Casey DeSmith inked a 3 year extension with the Penguins with an average annual value of \$1.25 million. His rise to earn this paycheck started with exemplary play during the 2017-2018 season and that has carried over into this season. Let's see how he got here!

```
desmith.2017 = get.goalie.data(analysis.2017, "Casey DeSmith")
desmith.2018 = get.goalie.data(analysis.2018, "Casey DeSmith")
```

Let's start by calculating some of his stats for each season and then tabularizing.

```
desmith.2017.sp = get.save.percent(desmith.2017)
desmith.2018.sp = get.save.percent(desmith.2018)

desmith.2017.gp = get.goal.percent(desmith.2017)
desmith.2018.gp = get.goal.percent(desmith.2018)
```

```
desmith.2017.spg = get.shots.per.goal(desmith.2017)
desmith.2018.spg = get.shots.per.goal(desmith.2018)
```

Now, let's put it in a table:

Season	Save Percent	Goal Percent	Shots Per Goal
2017-2018	0.9427403	0.0572597	17.4642857
2018-2019	0.9422886	0.0577114	17.3275862

We can look at his game by game data also.

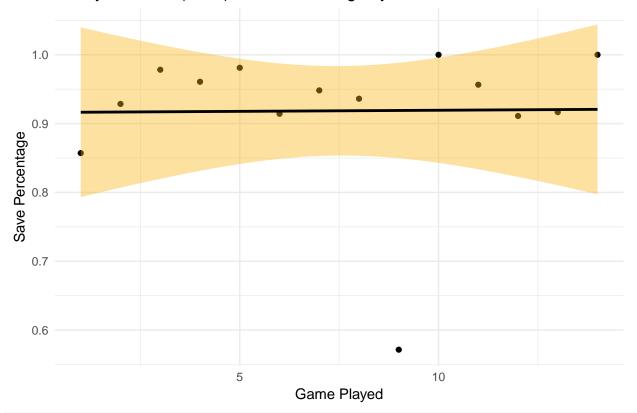
```
desmith.2017.games.sp = get.game.save.percent(desmith.2017)
desmith.2017.games.gp = get.game.goal.percent(desmith.2017)
desmith.2017.games.spg = get.game.shots.per.goal(desmith.2017)

desmith.2018.games.sp = get.game.save.percent(desmith.2018)
desmith.2018.games.gp = get.game.goal.percent(desmith.2018)
desmith.2018.games.spg = get.game.shots.per.goal(desmith.2018)
```

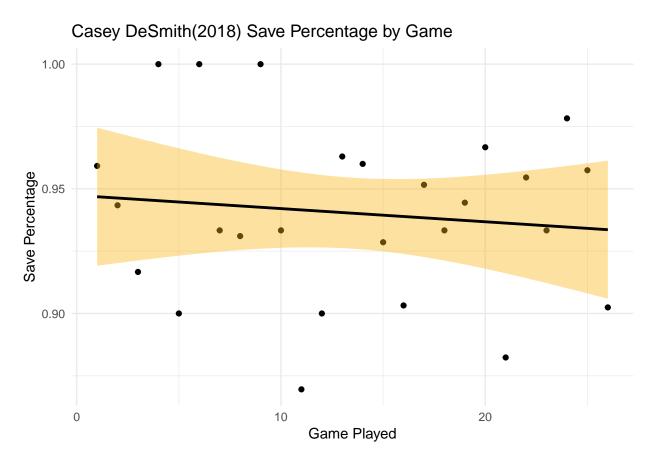
Let's look at his save percentage graphs:

```
desmith.2017.sp.plot = graph.trend(desmith.2017.games.sp, "Save Percentage", "#000000", "#FCB514", "Cas
desmith.2017.sp.plot
```

Casey DeSmith (2017) Save Percentage by Game

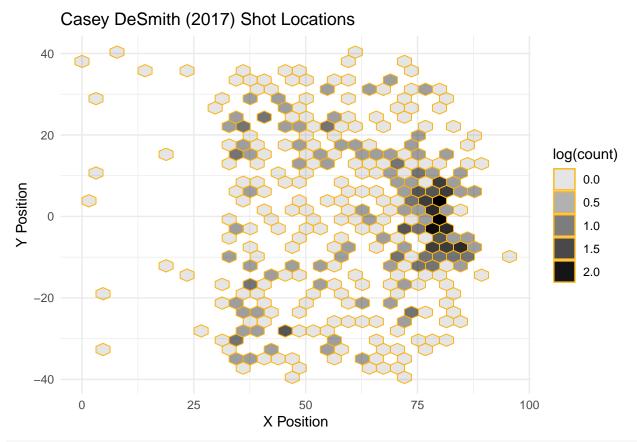


desmith.2018.sp.plot = graph.trend(desmith.2018.games.sp, "Save Percentage", "#000000", "#FCB514", "Cas
desmith.2018.sp.plot

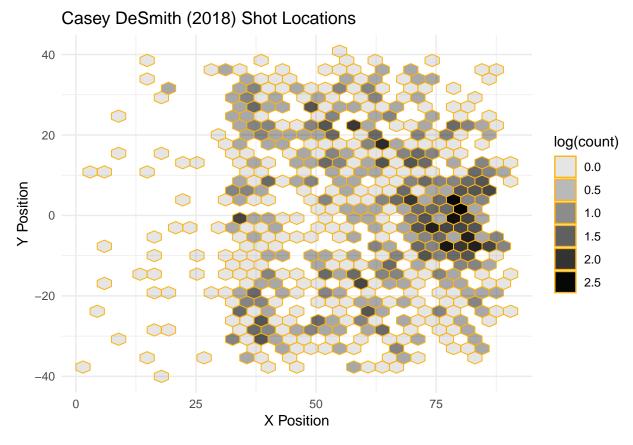


And now, let's take a look at his shot location data.

desmith.2017.locations.plot = graph.shot.locations(desmith.2017, "#000000", "#FCB514", "Casey DeSmith (
desmith.2017.locations.plot



desmith.2018.locations.plot = graph.shot.locations(desmith.2018, "#000000", "#FCB514", "Casey DeSmith (desmith.2018.locations.plot



As we can see, DeSmith has played pretty phenomenally. He's earned his new extension and his play has been pretty consistent. He is the perfect complement for Matt Murray.

Marc-Andre Fleury

Before he was picked by the Vegas Golden Knights in the most recent NHL Expansion Draft, Marc-Andre Fleury was the franchise goaltender for the Pittsburgh Penguins. Drafted first overall in the 2003 NHL Entry Draft, Marc-Andre "Flower" Fleury made his NHL debut in October of 2003. Since thenm he's been the main net-protector for the Penguins (until 2017 that is). His athleticism plus is exceptionally quick reactions makes Flower a premier NHL goaltender. With 3 Stanley Cup victories (2009, 2016, and 2017), Fleury has cemented his place as a top-10 all-time goaltender. With 429 regular season wins, Fleury sits at number 9 for career regular season wins. Only two active goaltenders, Roberto Luongo and Henrik Lundqvist, sit higher above him.

Fleury has been playing phenomenally the past few season. What seems like a second-prime for the Canadian born goaltender, Fleury showed shades of his 2008-2009 season, when he led the Penguins to their third Stanley Cup

```
fleury.2015 = get.goalie.data(analysis.2015, "Marc-Andre Fleury")
fleury.2016 = get.goalie.data(analysis.2016, "Marc-Andre Fleury")
fleury.2017 = get.goalie.data(analysis.2017, "Marc-Andre Fleury")
fleury.2018 = get.goalie.data(analysis.2018, "Marc-Andre Fleury")
```

Let's start by calculating some of his stats for each season and then tabularizing.

```
fleury.2015.sp = get.save.percent(fleury.2015)
fleury.2016.sp = get.save.percent(fleury.2016)
fleury.2017.sp = get.save.percent(fleury.2017)
```

```
fleury.2018.sp = get.save.percent(fleury.2018)

fleury.2015.gp = get.goal.percent(fleury.2015)
fleury.2016.gp = get.goal.percent(fleury.2016)
fleury.2017.gp = get.goal.percent(fleury.2017)
fleury.2018.gp = get.goal.percent(fleury.2018)

fleury.2015.spg = get.shots.per.goal(fleury.2015)
fleury.2016.spg = get.shots.per.goal(fleury.2016)
fleury.2017.spg = get.shots.per.goal(fleury.2017)
fleury.2018.spg = get.shots.per.goal(fleury.2018)
```

Now, let's put it in a table:

Season	Save Percent	Goal Percent	Shots Per Goal
2015-2016	0.9427197	0.0572803	17.4580153
2016-2017	0.9314803	0.0685197	14.5943396
2017-2018	0.9475891	0.0524109	19.08
2018-2019	0.9365621	0.0634379	15.7634409

We can look at his game by game data also.

```
fleury.2015.games.sp = get.game.save.percent(fleury.2015)
fleury.2015.games.gp = get.game.goal.percent(fleury.2015)
fleury.2015.games.spg = get.game.shots.per.goal(fleury.2015)

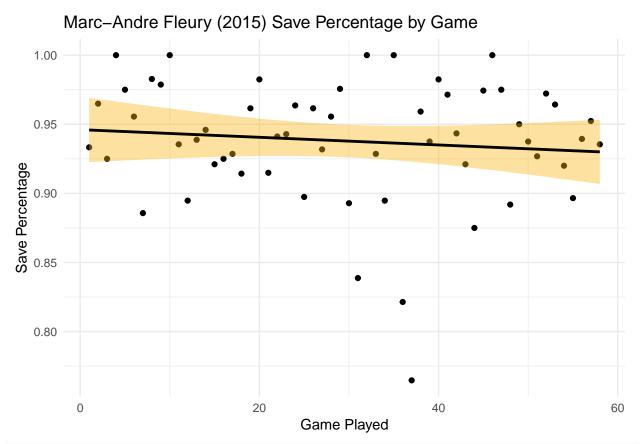
fleury.2016.games.sp = get.game.save.percent(fleury.2016)
fleury.2016.games.gp = get.game.goal.percent(fleury.2016)
fleury.2016.games.spg = get.game.shots.per.goal(fleury.2016)

fleury.2017.games.sp = get.game.save.percent(fleury.2017)
fleury.2017.games.gp = get.game.goal.percent(fleury.2017)
fleury.2017.games.spg = get.game.shots.per.goal(fleury.2017)

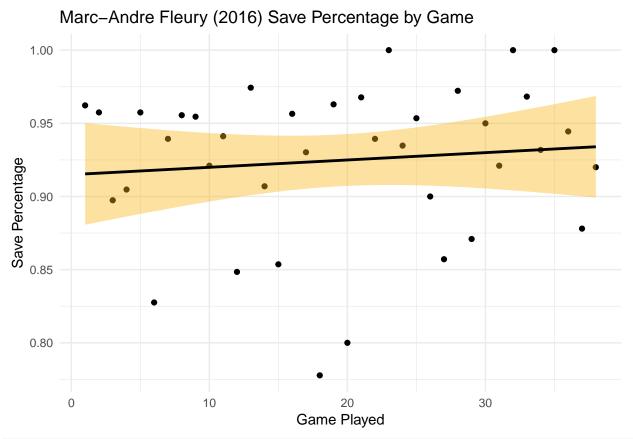
fleury.2018.games.sp = get.game.save.percent(fleury.2018)
fleury.2018.games.spg = get.game.shots.per.goal(fleury.2018)
fleury.2018.games.spg = get.game.shots.per.goal(fleury.2018)
```

Let's look at his save percentage graphs:

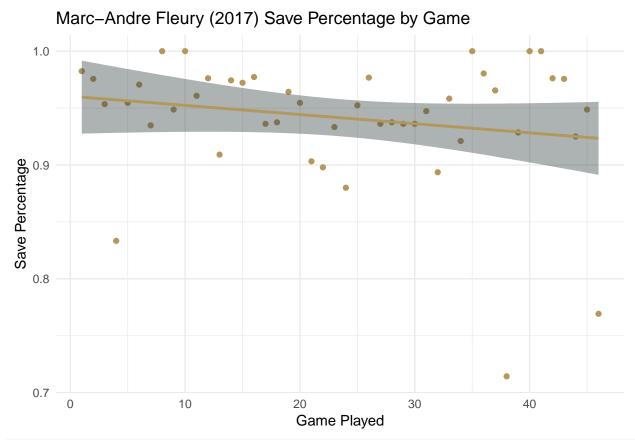
```
fleury.2015.sp.plot = graph.trend(fleury.2015.games.sp, "Save Percentage", "#000000", "#FCB514", "Marc-fleury.2015.sp.plot
```



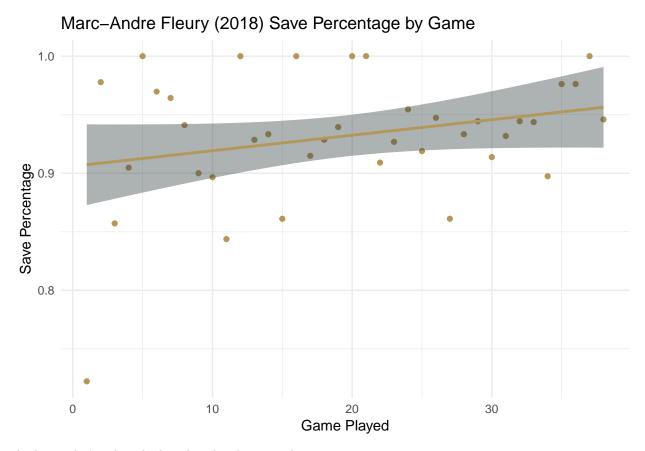
fleury.2016.sp.plot = graph.trend(fleury.2016.games.sp, "Save Percentage", "#000000", "#FCB514", "Marcfleury.2016.sp.plot



fleury.2017.sp.plot = graph.trend(fleury.2017.games.sp, "Save Percentage", "#B4975A", "#333F42", "Marcfleury.2017.sp.plot

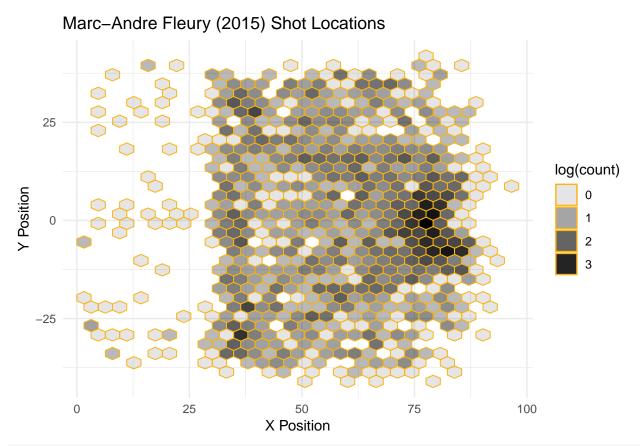


fleury.2018.sp.plot = graph.trend(fleury.2018.games.sp, "Save Percentage", "#B4975A", "#333F42", "Marcfleury.2018.sp.plot

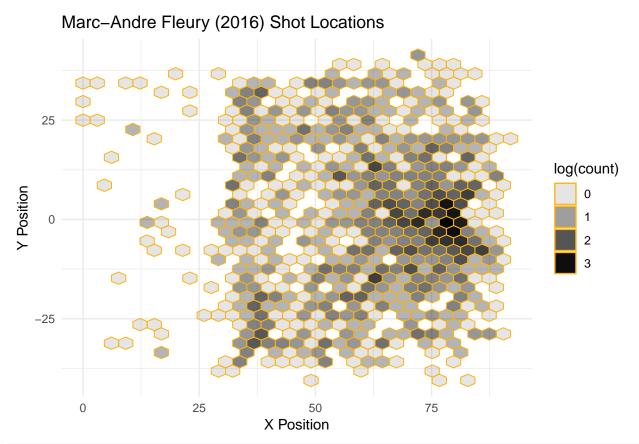


And now, let's take a look at his shot location data.

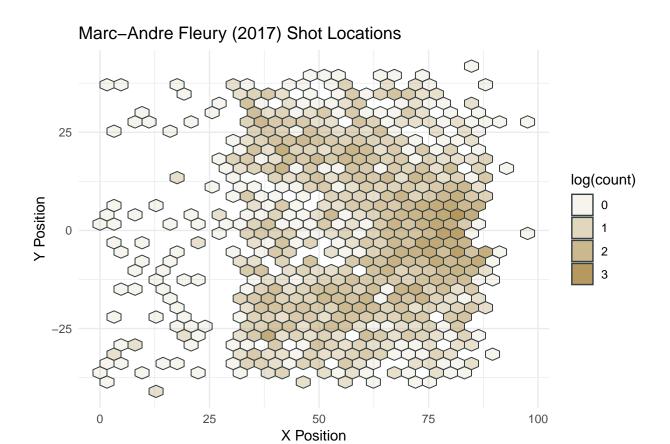
fleury.2015.locations.plot = graph.shot.locations(fleury.2015, "#000000", "#FCB514", "Marc-Andre Fleury
fleury.2015.locations.plot



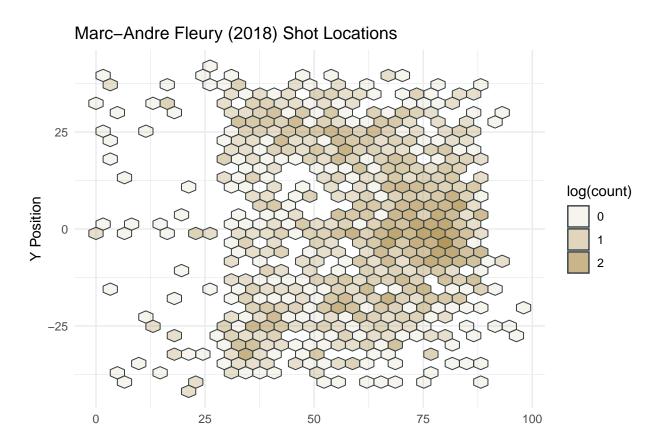
fleury.2016.locations.plot = graph.shot.locations(fleury.2016, "#000000", "#FCB514", "Marc-Andre Fleury fleury.2016.locations.plot



fleury.2017.locations.plot = graph.shot.locations(fleury.2017, "#B4975A", "#333F42", "Marc-Andre Fleury fleury.2017.locations.plot



fleury.2018.locations.plot = graph.shot.locations(fleury.2018, "#B4975A", "#333F42", "Marc-Andre Fleury
fleury.2018.locations.plot



Fleury has shown exceptional consistency. For any young goaltender breaking into the league, he should be a top-role model.

X Position

Braden Holtby

After exercising his playoff demons during the 2018 Stanley Cup Playoffs, Braden Holtby has lived up to his expectations. His play has been phenomenal during both the regular season and playoffs. Let's see how he's been since 2015.

```
holtby.2015 = get.goalie.data(analysis.2015, "Braden Holtby")
holtby.2016 = get.goalie.data(analysis.2016, "Braden Holtby")
holtby.2017 = get.goalie.data(analysis.2017, "Braden Holtby")
holtby.2018 = get.goalie.data(analysis.2018, "Braden Holtby")
```

Let's start by calculating some of his stats for each season and then tabularizing.

```
holtby.2015.sp = get.save.percent(holtby.2015)
holtby.2016.sp = get.save.percent(holtby.2016)
holtby.2017.sp = get.save.percent(holtby.2017)
holtby.2018.sp = get.save.percent(holtby.2018)

holtby.2015.gp = get.goal.percent(holtby.2015)
holtby.2016.gp = get.goal.percent(holtby.2016)
holtby.2017.gp = get.goal.percent(holtby.2017)
holtby.2018.gp = get.goal.percent(holtby.2018)
holtby.2015.spg = get.shots.per.goal(holtby.2015)
```

```
holtby.2016.spg = get.shots.per.goal(holtby.2016)
holtby.2017.spg = get.shots.per.goal(holtby.2017)
holtby.2018.spg = get.shots.per.goal(holtby.2018)
```

Now, let's put it in a table:

Season	Save Percent	Goal Percent	Shots Per Goal
2015-2016	0.9448141	0.0551859	18.1205674
2016-2017	0.9476936	0.0523064	19.1181102
2017-2018	0.9348104	0.0651896	15.3398693
2018-2019	0.9345048	0.0654952	15.2682927

We can look at his game by game data also.

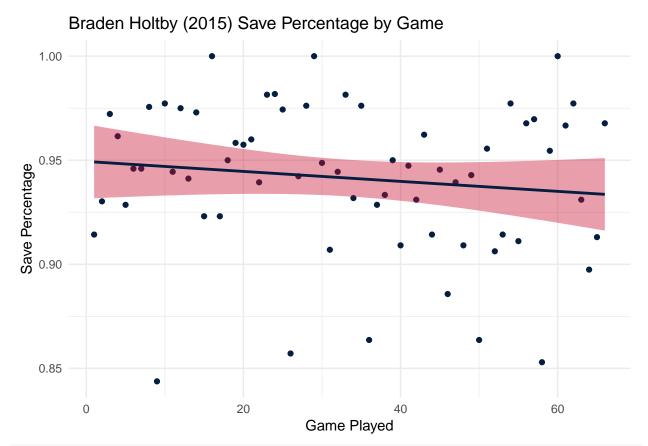
```
holtby.2015.games.sp = get.game.save.percent(holtby.2015)
holtby.2015.games.gp = get.game.goal.percent(holtby.2015)
holtby.2015.games.spg = get.game.shots.per.goal(holtby.2015)

holtby.2016.games.sp = get.game.save.percent(holtby.2016)
holtby.2016.games.gp = get.game.goal.percent(holtby.2016)
holtby.2016.games.spg = get.game.shots.per.goal(holtby.2016)

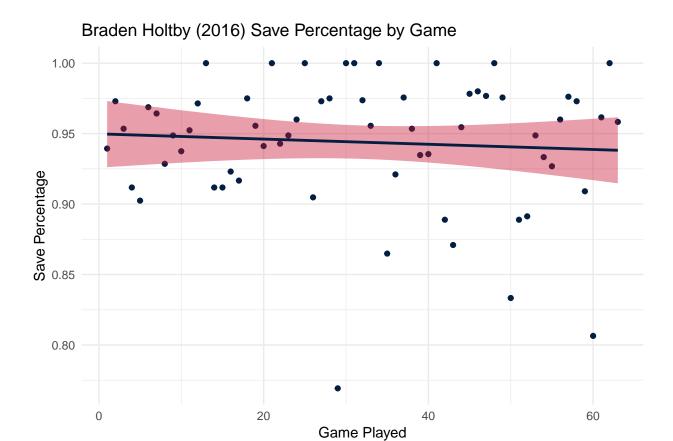
holtby.2017.games.sp = get.game.save.percent(holtby.2017)
holtby.2017.games.gp = get.game.goal.percent(holtby.2017)
holtby.2017.games.spg = get.game.shots.per.goal(holtby.2017)
holtby.2018.games.spg = get.game.save.percent(holtby.2018)
holtby.2018.games.spg = get.game.soal.percent(holtby.2018)
holtby.2018.games.spg = get.game.shots.per.goal(holtby.2018)
```

Let's look at his save percentage graphs:

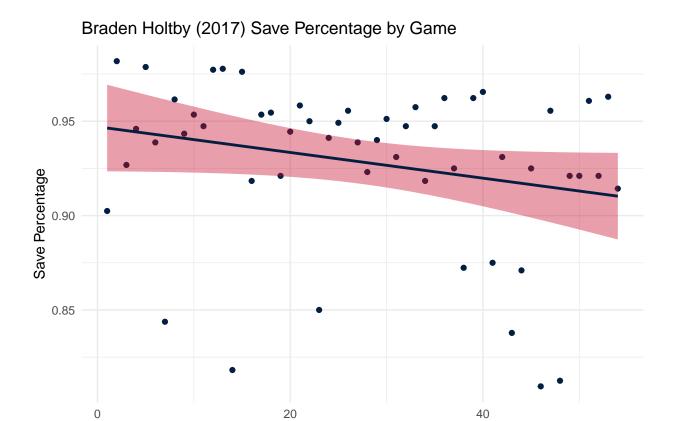
```
holtby.2015.sp.plot = graph.trend(holtby.2015.games.sp, "Save Percentage", "#041E42", "#C8102E", "Brade holtby.2015.sp.plot
```



holtby.2016.sp.plot = graph.trend(holtby.2016.games.sp, "Save Percentage", "#041E42", "#C8102E", "Brade holtby.2016.sp.plot

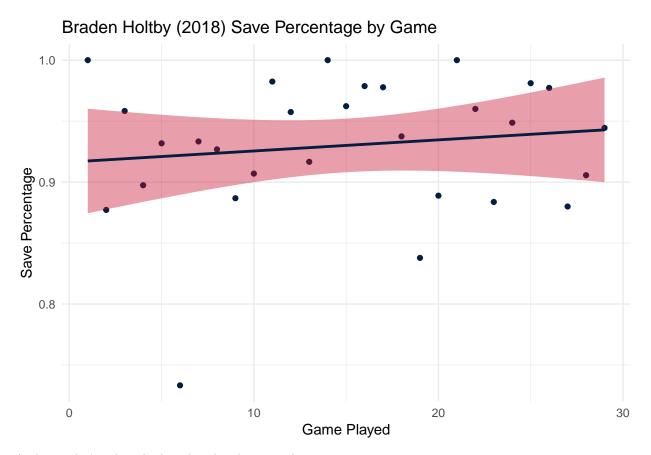


holtby.2017.sp.plot = graph.trend(holtby.2017.games.sp, "Save Percentage", "#041E42", "#C8102E", "Brade holtby.2017.sp.plot



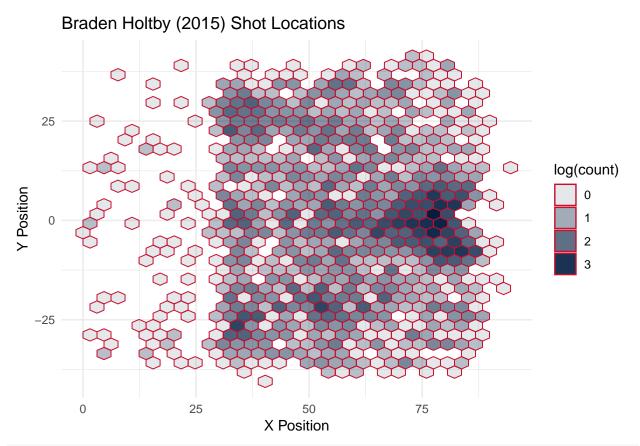
holtby.2018.sp.plot = graph.trend(holtby.2018.games.sp, "Save Percentage", "#041E42", "#C8102E", "Brade holtby.2018.sp.plot

Game Played

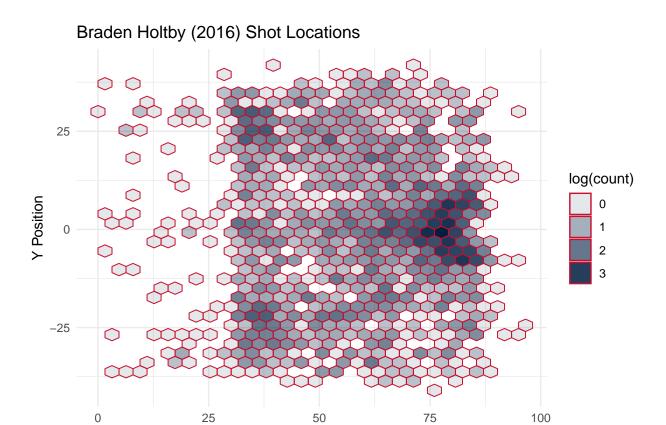


And now, let's take a look at his shot location data.

holtby.2015.locations.plot = graph.shot.locations(holtby.2015, "#041E42", "#C8102E", "Braden Holtby (20 holtby.2015.locations.plot

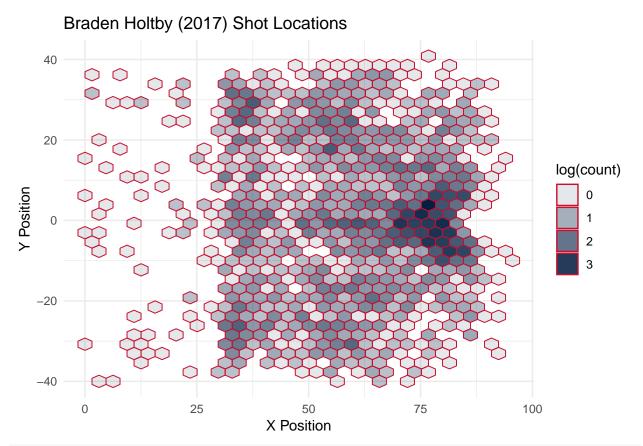


holtby.2016.locations.plot = graph.shot.locations(holtby.2016, "#041E42", "#C8102E", "Braden Holtby (20 holtby.2016.locations.plot

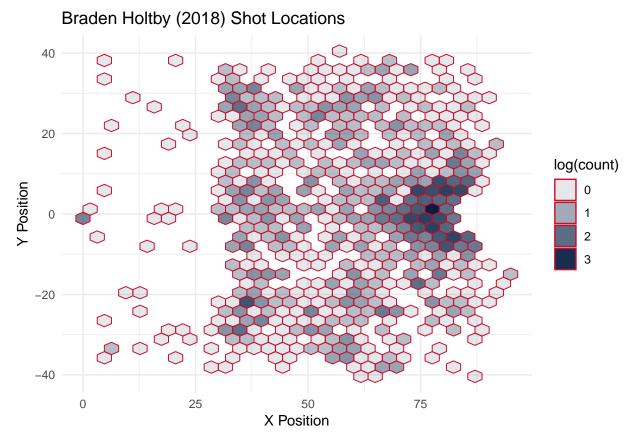


X Position

holtby.2017.locations.plot = graph.shot.locations(holtby.2017, "#041E42", "#C8102E", "Braden Holtby (20 holtby.2017.locations.plot



holtby.2018.locations.plot = graph.shot.locations(holtby.2018, "#041E42", "#C8102E", "Braden Holtby (20 holtby.2018.locations.plot



Holtby has had the trend of playing poorly near the end of the season. However, it seems like Holtby is currently going against his negative trend this season. His main weakness seems to be shots from the left circle. This is probably aimed towards his glove hand, which is significantly weaker than his positional play.

John Gibson

Skaters

Sidney Crosby

Evgeni Malkin

Alex Ovechkin

Jake Guentzel

Erik Karlsson

Kris Letang