Recently you might have seen an interesting [ABC](https://www.abc.net.au/news/2020-06-11/how-richmond-tigers-afl-player-dustin-martin-destroys-compeition/12335750?section=sport) piece on [Dustin Martin](https://en.wikipedia.org/wiki/Dustin_Martin)

Now more of you might be thinking oh but the AFL website itself has some more data available, but its just a pain to use. Its UI is pretty horrid, so its a little off putting having to manually copy and paste data into spreadsheets or its too much of a hurdle to go and learn how to scrape a website using R or python. But thankfully fryzigg has heard your frustrations and come along and help us make it more accessible for all.

So you might have read the Dustin Martin article and thought oh that’s cool but maybe I want to highlight some different players? While you can hover your mouse over the visualisation to label other players, that might not help you if you have no idea where your player of interest sits!

So How would you go about it and what are some things I think might be cool to explore a little differently.

In saying that, I want to make 3 changes to the graphs.

1 – I want the players I am interested in to to stand out more visually

2 – I want to be able to highlight players I am interested in

3 – I want to add a Season element to the plots, that is I want to be able to compare 2017 Dusty to 2016 Dusty and so on.

Lets first download the latest development version from github and create a Season column using lubridate::year

library(fitzRoy)

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.0 ✓ purrr 0.3.4

## ✓ tibble 3.0.1 ✓ dplyr 1.0.0

## ✓ tidyr 1.1.0 ✓ stringr 1.4.0

## ✓ readr 1.3.1 ✓ forcats 0.5.0

## ── Conflicts ────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──

## x dplyr::filter() masks stats::filter()

## x dplyr::lag() masks stats::lag()

df<-fitzRoy::get\_fryzigg\_stats(start=1897, end=2020)

## Returning cached data from 1897 to 2020

## This may take some time.

names(df)

## [1] "venue\_name" "match\_id"

## [3] "match\_home\_team" "match\_away\_team"

## [5] "match\_date" "match\_local\_time"

## [7] "match\_attendance" "match\_round"

## [9] "match\_home\_team\_goals" "match\_home\_team\_behinds"

## [11] "match\_home\_team\_score" "match\_away\_team\_goals"

## [13] "match\_away\_team\_behinds" "match\_away\_team\_score"

## [15] "match\_margin" "match\_winner"

## [17] "match\_weather\_temp\_c" "match\_weather\_type"

## [19] "player\_id" "player\_first\_name"

## [21] "player\_last\_name" "player\_height\_cm"

## [23] "player\_weight\_kg" "player\_is\_retired"

## [25] "player\_team" "guernsey\_number"

## [27] "kicks" "marks"

## [29] "handballs" "disposals"

## [31] "effective\_disposals" "disposal\_efficiency\_percentage"

## [33] "goals" "behinds"

## [35] "hitouts" "tackles"

## [37] "rebounds" "inside\_fifties"

## [39] "clearances" "clangers"

## [41] "free\_kicks\_for" "free\_kicks\_against"

## [43] "brownlow\_votes" "contested\_possessions"

## [45] "uncontested\_possessions" "contested\_marks"

## [47] "marks\_inside\_fifty" "one\_percenters"

## [49] "bounces" "goal\_assists"

## [51] "time\_on\_ground\_percentage" "afl\_fantasy\_score"

## [53] "supercoach\_score" "centre\_clearances"

## [55] "stoppage\_clearances" "score\_involvements"

## [57] "metres\_gained" "turnovers"

## [59] "intercepts" "tackles\_inside\_fifty"

## [61] "contest\_def\_losses" "contest\_def\_one\_on\_ones"

## [63] "contest\_off\_one\_on\_ones" "contest\_off\_wins"

## [65] "def\_half\_pressure\_acts" "effective\_kicks"

## [67] "f50\_ground\_ball\_gets" "ground\_ball\_gets"

## [69] "hitouts\_to\_advantage" "hitout\_win\_percentage"

## [71] "intercept\_marks" "marks\_on\_lead"

## [73] "pressure\_acts" "rating\_points"

## [75] "ruck\_contests" "score\_launches"

## [77] "shots\_at\_goal" "spoils"

## [79] "subbed" "player\_position"

df$Season<-lubridate::year(df$match\_date)

Next part, is we want to create the data to plot, to do this we want reproduce the plot that has the average number of centre clearances won on the x axis and the average number of stoppage clearances won on the y axis. The other things we want to do is filter our data by number of games played in season, I am going to set this number to 15

We will do this and call that dataframe p

p<-df%>%

group\_by(Season,player\_id, player\_first\_name, player\_last\_name)%>%

summarise(mean\_centre\_clearances=mean(centre\_clearances, na.rm=TRUE), mean\_clear=mean(stoppage\_clearances,na.rm=TRUE),no\_games=n())%>%

filter(Season>2011)%>%

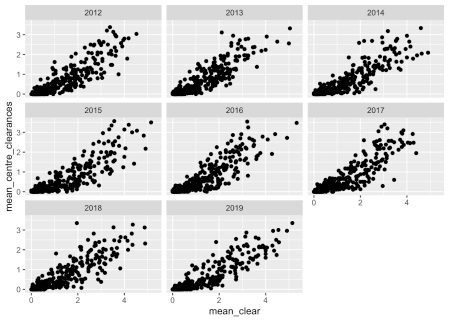
filter(no\_games>14)

## `summarise()` regrouping output by 'Season', 'player\_id', 'player\_first\_name' (override with `.groups` argument)

Now if we were to take p and use facet\_wrap(~Season) we get a heap of black dots geom\_point().

p%>%

ggplot(aes(x=mean\_clear, y=mean\_centre\_clearances,label=paste(player\_first\_name, player\_last\_name)))+geom\_point() +facet\_wrap(~Season)



Now we want to create a dataset with our players of interest one way to do this is using their unique player IDS which has been kindly provided by Fryzigg. These align to the official champion data IDS where possible, which is really cool, if you are so lucky it might mean you can append all the secret sauce AFL statistics that are withheld from fans to your insights.

p\_subset<-p%>%

filter(player\_id %in% c(11706, # Patrick Dangerfield

11801, # Dustin Martin

11844, # Nat Fyfe

12269, # Patrick Cripps

12061, # Lachie Neale

11813, # Luke Shuey

12058, # Adam Treloar

12223, # Brodie Grundie

11506, # Scott Pendlebury

12605, # Tim Kelly

12277 , # Marcus Bontempelli

11170 #Gary Ablett Jr

))

p\_subset\_dusty<-p%>%

filter(player\_id %in% c(11801))

So how do we make our players of interest pop out a bit more on the graph?

Well lets plot all the relevant players in our dataset in a lighter colour say grey and over that, we plot our players of interest in a darker colour say black.

p%>%

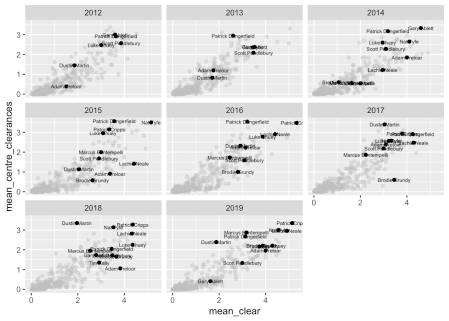
ggplot(aes(x=mean\_clear, y=mean\_centre\_clearances,label=paste(player\_first\_name, player\_last\_name)))+

geom\_point(colour="grey", alpha=0.4)+ # all the data

geom\_point(data=p\_subset, colour="black")+ # subset of players of interest

geom\_text(data=p\_subset, size=2)+

facet\_wrap(~Season)



p%>%

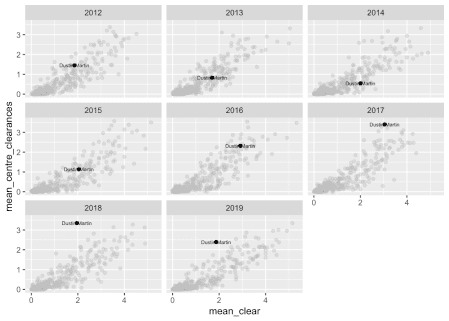
ggplot(aes(x=mean\_clear, y=mean\_centre\_clearances,label=paste(player\_first\_name, player\_last\_name)))+

geom\_point(colour="grey", alpha=0.4)+ # all the data

geom\_point(data=p\_subset\_dusty, colour="black")+ # subset of players of interest

geom\_text(data=p\_subset\_dusty, size=2)+

facet\_wrap(~Season)



So I think that is a pretty handy template if you want to explore so what exactly can you explore?

names(df)

## [1] "venue\_name" "match\_id"

## [3] "match\_home\_team" "match\_away\_team"

## [5] "match\_date" "match\_local\_time"

## [7] "match\_attendance" "match\_round"

## [9] "match\_home\_team\_goals" "match\_home\_team\_behinds"

## [11] "match\_home\_team\_score" "match\_away\_team\_goals"

## [13] "match\_away\_team\_behinds" "match\_away\_team\_score"

## [15] "match\_margin" "match\_winner"

## [17] "match\_weather\_temp\_c" "match\_weather\_type"

## [19] "player\_id" "player\_first\_name"

## [21] "player\_last\_name" "player\_height\_cm"

## [23] "player\_weight\_kg" "player\_is\_retired"

## [25] "player\_team" "guernsey\_number"

## [27] "kicks" "marks"

## [29] "handballs" "disposals"

## [31] "effective\_disposals" "disposal\_efficiency\_percentage"

## [33] "goals" "behinds"

## [35] "hitouts" "tackles"

## [37] "rebounds" "inside\_fifties"

## [39] "clearances" "clangers"

## [41] "free\_kicks\_for" "free\_kicks\_against"

## [43] "brownlow\_votes" "contested\_possessions"

## [45] "uncontested\_possessions" "contested\_marks"

## [47] "marks\_inside\_fifty" "one\_percenters"

## [49] "bounces" "goal\_assists"

## [51] "time\_on\_ground\_percentage" "afl\_fantasy\_score"

## [53] "supercoach\_score" "centre\_clearances"

## [55] "stoppage\_clearances" "score\_involvements"

## [57] "metres\_gained" "turnovers"

## [59] "intercepts" "tackles\_inside\_fifty"

## [61] "contest\_def\_losses" "contest\_def\_one\_on\_ones"

## [63] "contest\_off\_one\_on\_ones" "contest\_off\_wins"

## [65] "def\_half\_pressure\_acts" "effective\_kicks"

## [67] "f50\_ground\_ball\_gets" "ground\_ball\_gets"

## [69] "hitouts\_to\_advantage" "hitout\_win\_percentage"

## [71] "intercept\_marks" "marks\_on\_lead"

## [73] "pressure\_acts" "rating\_points"

## [75] "ruck\_contests" "score\_launches"

## [77] "shots\_at\_goal" "spoils"

## [79] "subbed" "player\_position"

## [81] "Season"

Now that you know and hopefully have some script to one, what would you as a fan like to explore?