DAR F23 Assignment 5

Hockey Analytics'

Caleb Smith

2023-11-02

Contents

Weekly Work Summary

NOTE: Follow an outline format; use bullets to express individual points.

- RCS ID: **Always** include this!

smithc22

- Project Name: **Always** include this!

Hockey Analytics

- Summary of work since last week
  - Describe the important aspects of what you worked on and accomplished

  I worked on plotting plays on the rink and prototyping an app to do this for us

- NEW: Summary of github issues added and worked
  - Issues that you've submitted
  - Issues that you've self-assigned and addressed

  N/A

- Summary of github commits
  - include branch name(s) dar-smithc22
  - include browsable links to all external files on github
  - Include links to shared Shiny apps https://lp01.idea.rpi.edu/shiny/smithc22/Play_Plotter/

- List of presentations, papers, or other outputs
  - Include browsable links

  Basic explanation of the play plotter: https://docs.google.com/document/d/19RMmRwWlqKi4pS0m0VTqR2eSwVaxh_ciWOWHYY0eH_E/edit

- List of references (if necessary)
- Indicate any use of group shared code base

I took the code from Amy's notebook that she was using to create the rink plot and draw arrows

- Indicate which parts of your described work were done by you or as part of joint efforts
- **Required:** Provide illustrating figures and/or tables

Personal Contribution

- Clearly defined, unique contribution(s) done by you: code, ideas, writing...
- Include github issues you've addressed

Main contribution is the app

Analysis: Question 1 Plotting a cluster

Question being asked

Provide in natural language a statement of what question you're trying to answer

I'm trying to visualize the clusters in an intuitive, explainable way since UMAP is neither of thse things.

Data Preparation

Provide in natural language a description of the data you are using for this analysis

Code taken from Amy

Include a step-by-step description of how you prepare your data for analysis

Split the shots into the 5 clusters so I can plot them seperately and load the neccesary data and images to make the half rink plots.

If you're re-using dataframes prepared in another section, simply re-state what data you're using

```
# Include all data processing code (if necessary), clearly commented
# Include all data processing code (if necessary), clearly commented
# Size of rink image and of all plots

# Plotting
library(jpeg)
library(grid)
library(ggnewscale)
library(scales)
# Goal shot stats
library(reshape2)
#library(randomForest)
# library(glmnet)
#library(MASS)
#library(Boruta)
#library(MLmetrics)
#library(kernelshap)
#library(shapviz)
library(knitr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.3     v purrr   1.0.2
## v tibble  3.2.1     v dplyr   1.1.2
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.3     v forcats 0.5.2
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x readr::col_factor() masks scales::col_factor()
## x purrr::discard()    masks scales::discard()
## x dplyr::filter()     masks stats::filter()
## x dplyr::lag()        masks stats::lag()
```

```
library(tidymodels)
```

```
## -- Attaching packages ------------------------------------- tidymodels 1.1.1 --
## v broom        1.0.5     v rsample      1.2.0
## v dials        1.2.0     v tune         1.1.2
## v infer        1.0.4     v workflows    1.1.3
## v modeldata    1.2.0     v workflowsets 1.0.1
## v parsnip      1.1.1     v yardstick    1.2.0
## v recipes      1.0.8
## -- Conflicts ---------------------------------------- tidymodels_conflicts() --
## x purrr::discard()  masks scales::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
## * Use tidymodels_prefer() to resolve common conflicts.
```

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##     set_names
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
source("../../AnalysisCodeFunc.R")

#Splits it up into clusters
shots_stats.df <- readRDS("../../StudentData/shot_stats_goal_clusters.df.Rds")
cluster1 <- shots_stats.df %>% filter(Cluster == 1)
cluster2 <- shots_stats.df %>% filter(Cluster == 2)
cluster3 <- shots_stats.df %>% filter(Cluster == 3)
cluster4 <- shots_stats.df %>% filter(Cluster == 4)
cluster5 <- shots_stats.df %>% filter(Cluster == 5)
#goals <- shots_stats.df %>% filter(outcomes == 2)

xsize <- 2000
ysize <- 850
# Coordinates to the goal pipes
pipes_x <- 1890
lpipe_y <- 395
rpipe_y <- 455
# This file path should contain the hockey rink images and all the sequences
filepath <- '../../FinalGoalShots/'
```

```
# Read the rink images and format them to a raster used for graphing
rink_raster <- makeRaster(filepath, 'Rink_Template.jpeg')
half_rink_raster <- makeRaster(filepath, 'Half_Rink_Template.jpeg')
```

Analysis: Methods and results

Describe in natural language a statement of the analysis you're trying to do

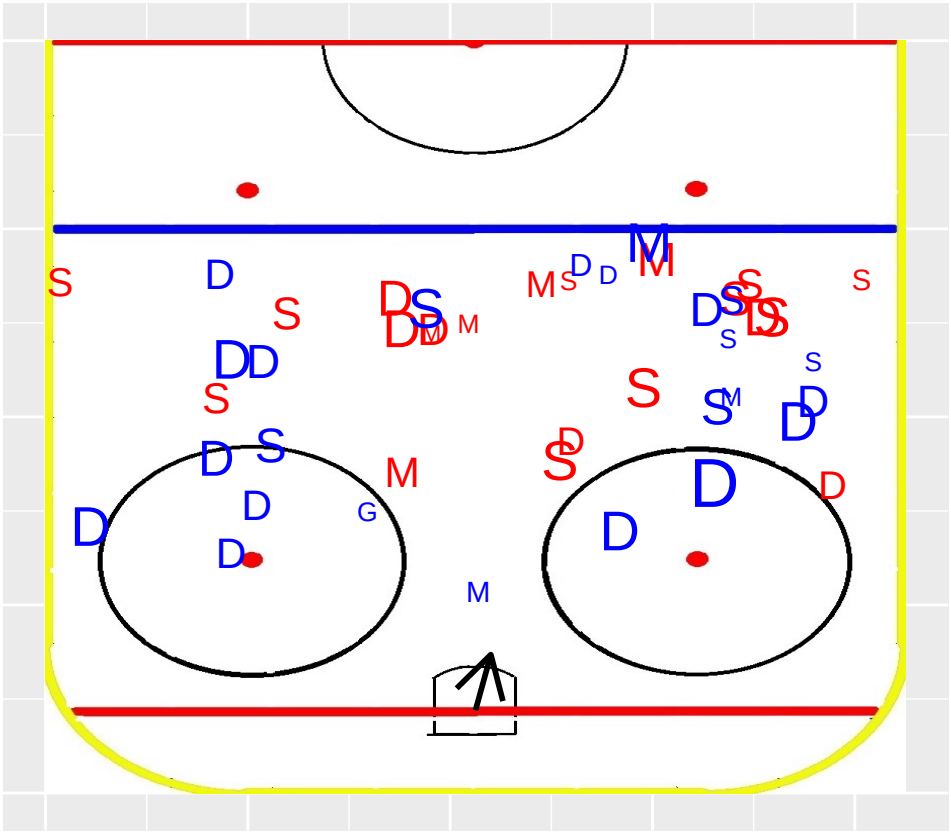
I'm trying to plot each of the clusters with as many features visible as feasible on a half rink graph. The location of the graph allows for the angle and distance to the goal to be plotted, and I intend to utilize color and size as other dimensions. I have (mostly arbitrarily) chosen possesion time for the size dimension and handedness for the color dimension (because it was already in the code)

Provide clearly commented analysis code; include code for tables and figures!

```
# Include all analysis code, clearly commented
# If not possible, screen shots are acceptable.
# If your contributions included things that are not done in an R-notebook,
#   (e.g. researching, writing, and coding in Python), you still need to do
#   this status notebook in R.  Describe what you did here and put any products
#   that you created in github. If you are writing online documents (e.g. overleaf
#   or google docs), you can include links to the documents in this notebook
#   instead of actual text.

plotShots <- function(shots){
ShotType <- shots$shotOutcome
levels(ShotType) <- c('D','G','S','M')

#Going to try and display
angle_means <- shots %>%
  #group_by(puckAngle<=90) %>%
  summarise(mean(goalieAngle), mean(goalieDist)) %>%
  set_names(c('meanAngle', 'meanDist')) %>%
  # Data for graphing
  mutate(xstart = ysize / 2) %>%
  mutate(ystart = pipes_x) %>%
  mutate(radius = meanDist)

#Idea: do the arrows out of the goal for goalie angle and distance.
#For how to divide it: maybe goalie angle average for points on the left side and goalie angle average
#Need to rescale the arrows

halfRinkGraph(shots) +
  # Graph players colored by handedness
  geom_text(aes(label = ShotType,color = as.logical(rightHanded), x = shotStatX(shots),
                y = shotStatY(shots), size = posTime)) +
  scale_color_discrete('Handedness', type = c('red', 'blue'), labels = c('Left', 'Right')) +
  scale_size_continuous(name = 'Posession Time', range = c(3,9))+
  new_scale_color() +
  # Arrow pointing to average goalie angle
  # Length of arrow is average goalie distance
  geom_spoke(data = angle_means, aes(x = xstart, y = ystart, angle =
  torad(meanAngle), radius = radius),
  key_glyph = 'abline', linetype = 'solid', arrow = arrow(), linewidth = 1) +
  #scale_color_discrete('Average angle', type = c('red', 'blue'),
```

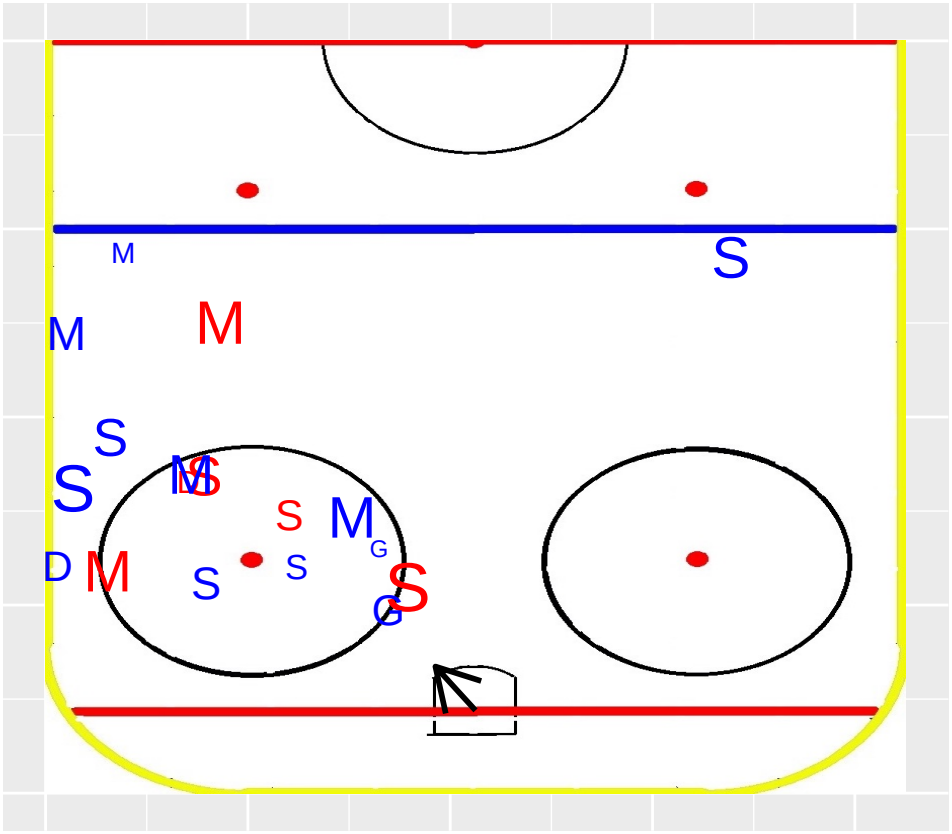
```
  #labels = c('Left', 'Right')) +
  labs(x = NULL, y = NULL)
}
plotShots(cluster1)
```

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




```
plotShots(cluster2)
```

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



Handedness

a Left

a Right

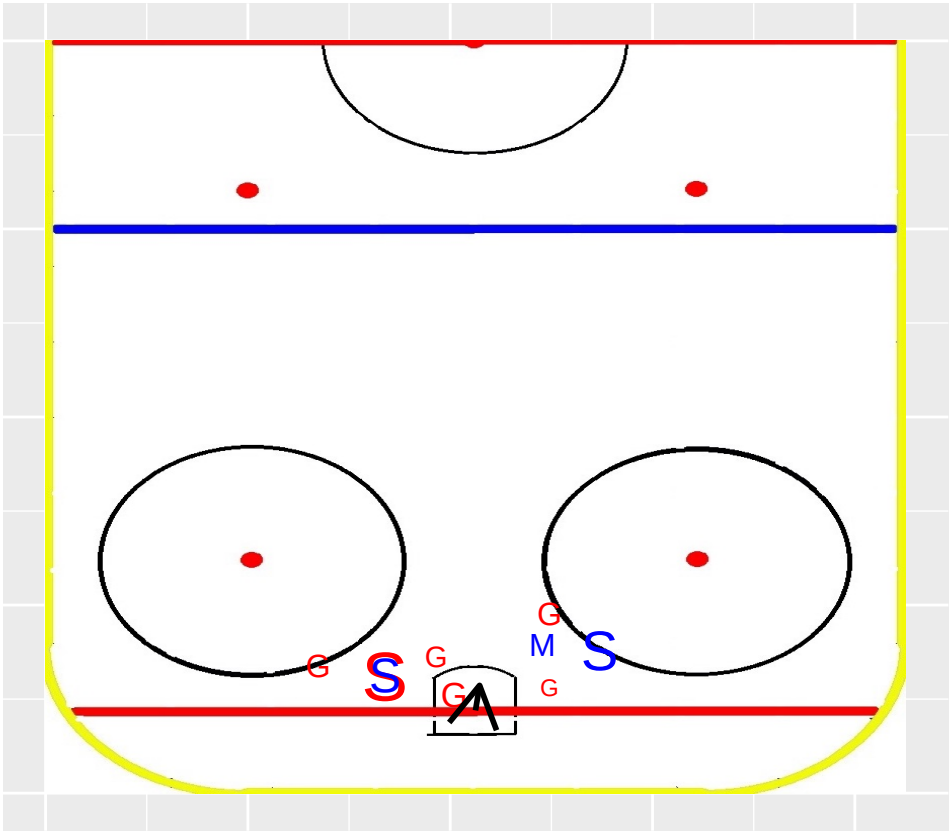
Posession Time

a 20

a 40

a 60

```
plotShots(cluster3)
```



Handedness

a Left

a Right

Posession Time

a 10

a 20

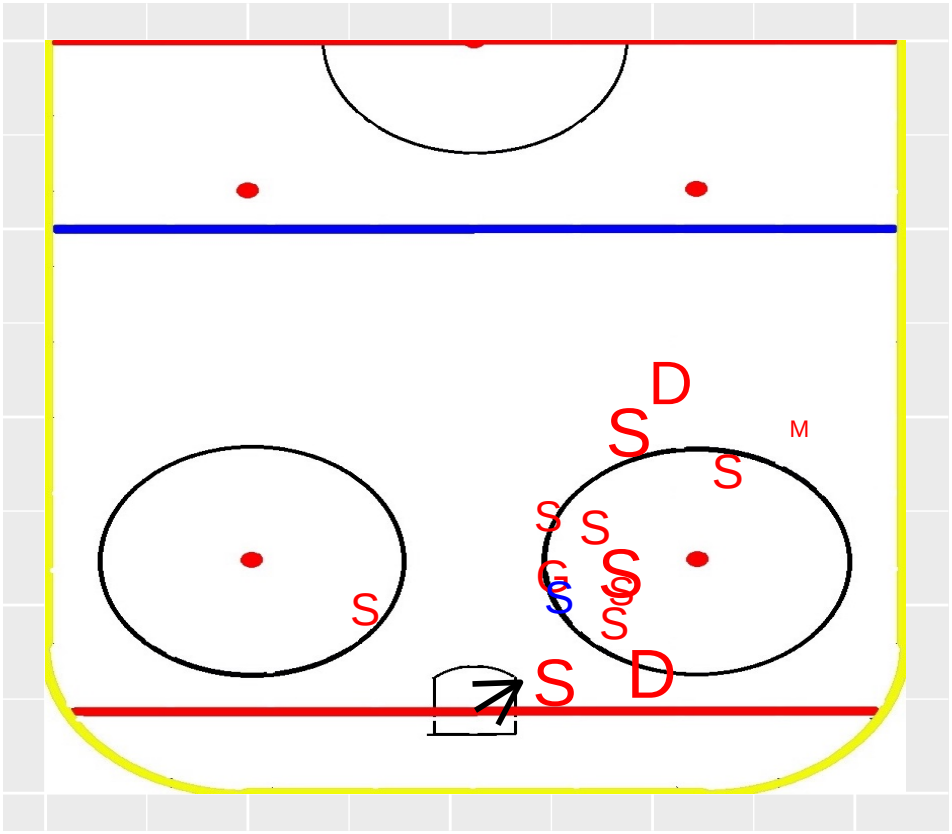
a 30

a 40

a 50

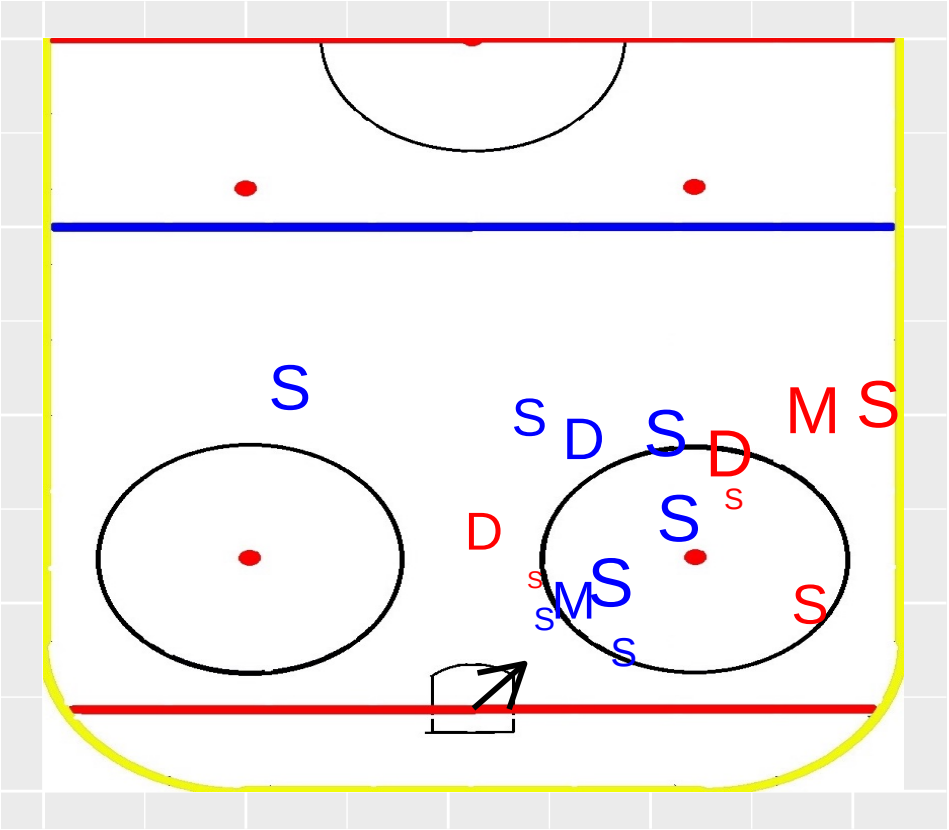
a 60

```
plotShots(cluster4)
```




```
plotShots(cluster5)
```

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




```
# to show the raw numbers used in the creation of the image
#print(angle_means)
#plotShots(goals)
```

Discussion of results

Provide in natural language a clear discussion of your observations.

The rink plots provide a sense of the actual meaning of the data that the numbers lack, especially since the raw data units aren't intuitive. Here, we can see the defender blocks were simply too far away, the panic shots were on the left and had a high possesion time, the goals were snap shots from close in, most of the ones that had a good shot but failed were on the right, and basically all the traffic jam shots took place from within or next to the right circle.

Analysis: Question 2 Graphing a singular point

Question being asked

Provide in natural language a statement of what question you're trying to answer

What is the best way to plot a single play on the rink?

Data Preparation

Provide in natural language a description of the data you are using for this analysis

Include a step-by-step description of how you prepare your data for analysis

All the needed data preperation was done in the first question. I am reusing the half rink picture and the data frames containing the clusters.

If you're re-using dataframes prepared in another section, simply re-state what data you're using

```
# Include all data processing code (if necessary), clearly commented
```

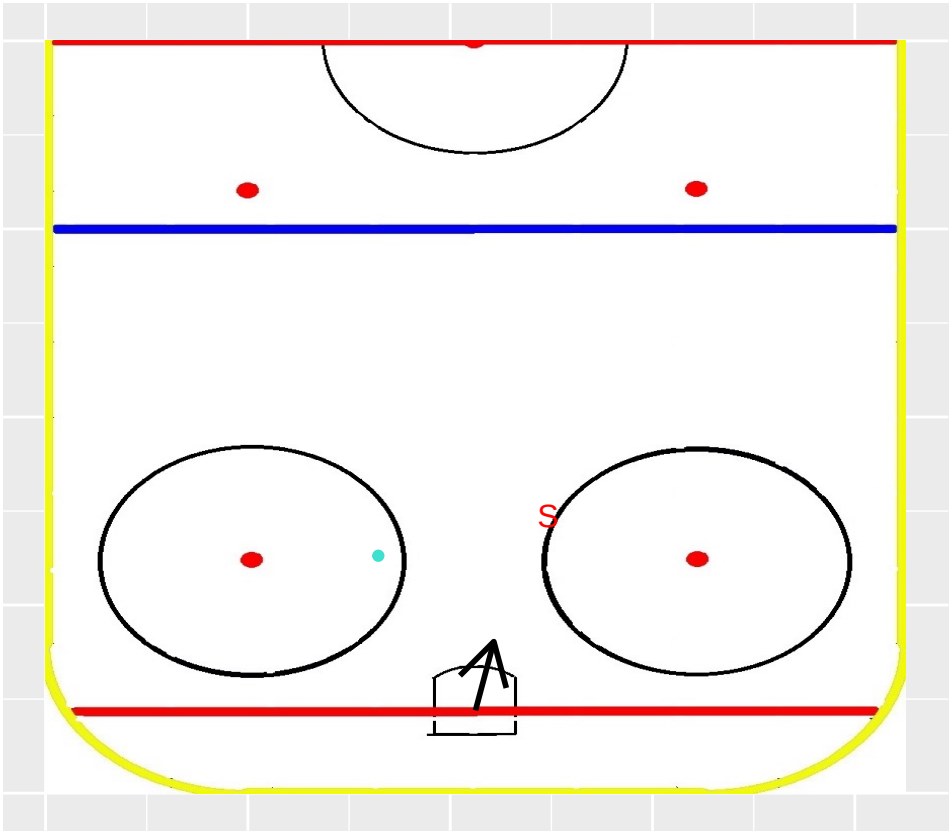
Analysis: Methods and Results

Describe in natural language a statement of the analysis you're trying to do

I am plotting the individual play with the same feature mappings as I did for the clusters, but I am also including a dot for the location of the nearest defender and an arrow pointing to where the goalie is.

Provide clearly commented analysis code; include code for tables and figures!

```
# Include all analysis code, clearly commented
# If not possible, screen shots are acceptable.
# If your contributions included things that are not done in an R-notebook,
#   (e.g. researching, writing, and coding in Python), you still need to do
#   this status notebook in R.  Describe what you did here and put any products
#   that you created in github. If you are writing online documents (e.g. overleaf
#   or google docs), you can include links to the documents in this notebook
#   instead of actual text.

#plotPoint <- function(shot){
shot <- shots_stats.df[101,]
  ShotType <- shot$shotOutcome
levels(ShotType) <- c('D','G','S','M')

#potential code for computing defender location
defPos <- cbind.data.frame(shotStatX(shot) + cos(torad(shot$defAngle))*shot$defDist,shotStatY(shot) + s
colnames(defPos) <- c('x','y')

#Calculates the goalie's position
angle <-  shot %>% mutate(xstart = ysize / 2) %>%
  mutate(ystart = pipes_x) %>%
  mutate(rad = goalieDist) %>%
  mutate(ang = goalieAngle)

  halfRinkGraph(shot) +
  # Graph players colored by handedness
  geom_text(aes(label = ShotType,color = as.logical(rightHanded), x = shotStatX(shot),
                y = shotStatY(shot), size = posTime)) +
  scale_color_discrete('Handedness', type = c('red', 'blue'), labels = c('Left', 'Right')) +
  scale_size_continuous(name = 'Posession Time', limits = c(0,5))+
  new_scale_color() +
  # Arrow pointing to average goalie angle
  # Length of arrow is average goalie distance
  geom_spoke(data = angle, aes(x = xstart, y = ystart, angle =
  torad(ang), radius = rad),
  key_glyph = 'abline', linetype = 'solid', arrow = arrow(), linewidth = 1) +
  geom_point(data = defPos,aes(x = x,y=y,color = "Turquoise"))+
  labs(x = NULL, y = NULL)+
    scale_color_discrete('',type = c('turquoise'))
```




```
#}
#plotPoint(shots_stats.df[29,])
```

Discussion of results

Provide in natural language a clear discussion of your observations.

The main advantage of this is it does a good job of explaining why a shot had a particular outcome in a way that isn't as easy to see in the main graph. For example, the save here is obviously a save because the nearest defender wasn't in a position to block it, allowing it to get to the goal. The goalie was then able to stop it because she was facing the shooter and also had come out of the goalie box towards her.

Analysis: Question 3 Producing an app

Question being asked

Provide in natural language a statement of what question you're trying to answer

I am trying to figure out the best way to display the data in a way that will allow useful insight to be made in the future

Data Preparation

Provide in natural language a description of the data you are using for this analysis

Include a step-by-step description of how you prepare your data for analysis

If you're re-using dataframes prepared in another section, simply re-state what data you're using

```
# Include all data processing code (if necessary), clearly commented
```

Analysis methods used

Describe in natural language a statement of the analysis you're trying to do

Provide clearly commented analysis code; include code for tables and figures!

I am making an app that will allow plays to be plotted. I intend to add more features to allow users to customize data selection and plot features in the future. The link to the current version is at https://lp01.idea.rpi.edu/shiny/smithc22/Play_Plotter/ . I intend to add more features over the next few weeks. Note that the drop down is currently used for cluster selection, the checkboxes don't do anything. You can find the code for the app in the ShinyApps folder under Play_Plotter. Linked is a google doc that explains the app because R doesn't want to let me put screenshots in the notebook for some reason

https://docs.google.com/document/d/19RMmRwWlqKi4pS0m0VTqR2eSwVaxh_ciWOWHYY0eH_E/edit

```
# Include all analysis code, clearly commented
# If not possible, screen shots are acceptable.
# If your contributions included things that are not done in an R-notebook,
#   (e.g. researching, writing, and coding in Python), you still need to do
#   this status notebook in R.  Describe what you did here and put any products
#   that you created in github. If you are writing online documents (e.g. overleaf
#   or google docs), you can include links to the documents in this notebook
#   instead of actual text.

#knitr::include_graphics("~/Hockey_Fall_2023/StudentNotebooks/Assignment05/defaultdisplay.png")
```

Discussion of results

Provide in natural language a clear discussion of your observations.

Summary and next steps

Provide in natural language a clear summary and your proposed next steps.