DAR F23 Project Status Notebook

Hockey Analytics

Amy Enyenihi

2023-10-16

Contents

Weekly Work Summary

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

- RCS ID: enyena

- Project Name: Hockey Analytics

- Summary of work since last week

  - I began to work on visualization of rink and the players.

- NEW: Summary of github issues added and worked

  - N/A

- Summary of github commits

  - branch name: dar-enyena
  - Updated the shots_stats_goal file with the Rds file containing the new features
  - https://github.rpi.edu/DataINCITE/Hockey_Fall_2023/pull/15

- List of presentations, papers, or other outputs

  - N/A

- List of references (if necessary)

- Indicate any use of group shared code base

- Indicate which parts of your described work were done by you or as part of joint efforts

  - I started with code that was originally written by Jeff (density plot) and Mohamed (rink plot)

- **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

Analysis: Question 1 - Density Plot by Outcome

Question being asked

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

How does the existing box plot, made by Jeff, look with the new, categorized outcome feature?

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

I am using the data frame that was created by Mohamed and Dr. Morgan. This will be the same data frame I use throughout the notebook.

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

#read in the data frame
shots_stats.df <- readRDS("../../StudentData/shots_stats_goal.df.Rds")
```

### Analysis: Methods and results

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!

The original density plot was only categorized by goal and no goal. I made this recently after distinguishing all four outcomes and wanted to see the effects on graphs my team members had made.

```
# 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.

# density plot for puckDist
# code was originally written by Jeff
ggplot(shots_stats.df, aes(x = puckDist, fill = shotOutcome)) +
  geom_density(alpha = 0.3) +
  labs(title = "Density Plot of Puck Distance by Outcome",
       x = "Puck Distance", y = "Density")
```
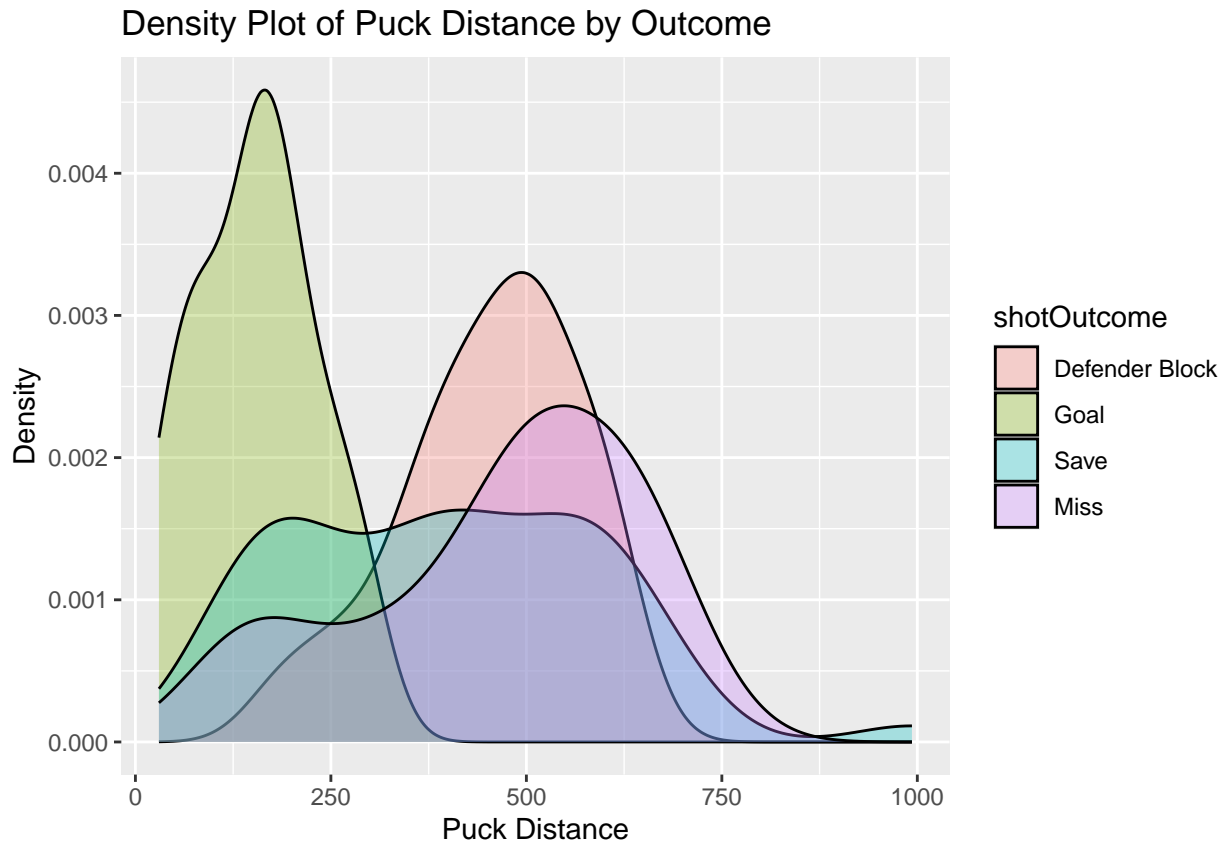
Density Plot of Puck Distance by Outcome

Discussion of results

Provide in natural language a clear discussion of your observations.

It was already made clear that the goals were made when the puck was closer to the goal. Not much new information was provided as all non goals are still found at a further distance, but it was nice to have all four outcomes displayed and see the new feature visualized.

Analysis: Question 2 - Visualizing Players by Handedness

Question being asked

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

How can I edit the rink image showing players handedness in relation to their average shooting angle?

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

I will be using the same data frame from above. It is already loaded in, so there is no processing necessary. Now that I am displaying the rink, I included the appropriate numbers for the rink size and coordinates and loaded in the jpegs.

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

# Size of rink image and of all plots
```

```
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

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

I wanted to take the image that was originally created by Mohamed and try to add more features to it. It also became the basis for my work going forward since I am now working on visualization. To the original image, I added a feature that depicts the average puck distance for each handedness. This is represented in the image as the length of the arrow. I also wanted to make the image more color blind friendly.

```
# 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.

angle_means <- shots_stats.df %>%
  group_by(rightHanded) %>%
  summarise(mean(puckAngle), mean(puckDist)) %>%
  set_names(c('rightHanded', 'meanAngle', 'meanDist')) %>%
  # Data for graphing
  mutate(xstart = ysize / 2) %>%
  mutate(ystart = pipes_x) %>%
  mutate(radius = meanDist)

halfRinkGraph(shots_stats.df) +
  # Graph players colored by handedness
  geom_point(aes(color = as.logical(rightHanded), x = shotStatX(shots_stats.df),
             y = shotStatY(shots_stats.df))) +
  scale_color_discrete('Handedness', type = c('red', 'blue'), labels = c('Left', 'Right')) +
  new_scale_color() +
  # Arrow pointing to average direction
  # Length of arrow is average puck distance
  geom_spoke(data = angle_means, aes(x = xstart, y = ystart, angle =
  torad(meanAngle), radius = radius, color = as.logical(rightHanded)),
  key_glyph = 'abline', linetype = 'solid', arrow = arrow(), linewidth = 1) +
```
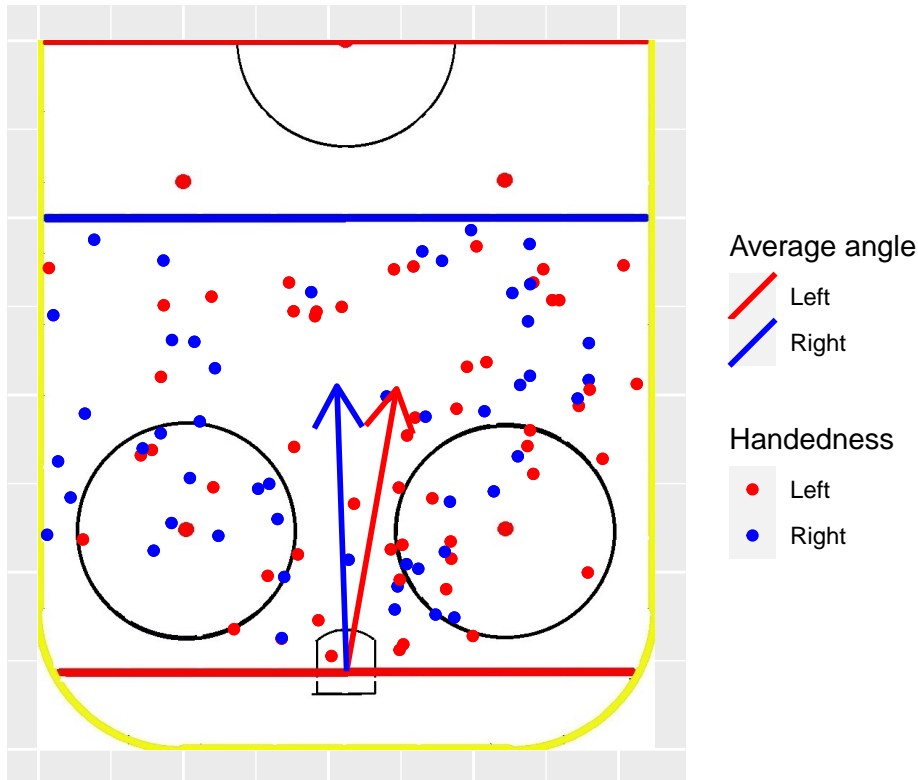
4

```
    scale_color_discrete('Average angle', type = c('red', 'blue'),
    labels = c('Left', 'Right')) +
    labs(x = NULL, y = NULL)
```

Warning: Removed 6 rows containing missing values (`geom_point()`).


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

```
## # A tibble: 2 x 6
##   rightHanded meanAngle meanDist xstart ystart radius
##         <dbl>     <dbl>    <dbl>  <dbl>  <dbl>  <dbl>
## 1           0      80.2     405.    425   1890   405.
## 2           1      91.9     403.    425   1890   403.
```

Discussion of results

Provide in natural language a clear discussion of your observations.

This figure is a graph of the rink with the handedness of players labeled, arrows representing the average angle of the puck for each handedness, and the length of the arrow representing the average puck distance for each handedness. The dots represent the players location on the ice at the time of the shot. Left-handed players average shooting from their left of the goal, but right-handed players average shooting from the center. Oppositely, the average shooting distance was nearly the same between left-handed and right-handed players.

Analysis: Question 3 - Visualizing Shots by Outcome

Question being asked

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

How can I recreate the previous image but categorizing the shooters by their shot outcome rather than handedness?

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

I will be using the same data frame from above as well as the rink info that was already loaded in.

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

#N/A
```

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 want to create a similar image to the handedness diagram but with a different feature. The dots still represent the players location and this time they will be sorted by the four outcomes: goal, save, defender block, and miss. Keeping the same format, the arrows will depict the average angle and its length will model the average puck distance.

```r
# 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.

shot_outcomes <- shots_stats.df %>%
  group_by(shotOutcome) %>%
  summarise(mean(puckAngle), mean(puckDist)) %>%
  set_names(c('outcome', 'meanAngle', 'meanDist')) %>%
  # Data for graphing
  mutate(xstart = ysize / 2) %>%
  mutate(ystart = pipes_x) %>%
  mutate(radius = meanDist)

halfRinkGraph(shots_stats.df) +
  # Graph players colored by shot type
  geom_point(aes(color = shotOutcome, x = shotStatX(shots_stats.df), y =
  shotStatY(shots_stats.df))) + scale_color_discrete('Outcome', type = c('red',
  'blue', 'green', 'black'), labels = c('DB', 'Goal', 'Save','Miss')) +
  new_scale_color() +
  # Arrow pointing to average direction
  # Length of arrow is average puck distance
  geom_spoke(data = shot_outcomes, aes(x = xstart, y = ystart, angle =
  torad(meanAngle), radius = radius, color = outcome), key_glyph = 'abline',
  linetype = 'solid', arrow = arrow(), linewidth = 1) +
  scale_color_discrete('Average Angle', type = c('red', 'blue', 'green', 'black'),
```

6

```
    labels = c('DB', 'Goal', 'Save','Miss')) +
  labs(x = NULL, y = NULL)
```

Warning: Removed 6 rows containing missing values (`geom_point()`).
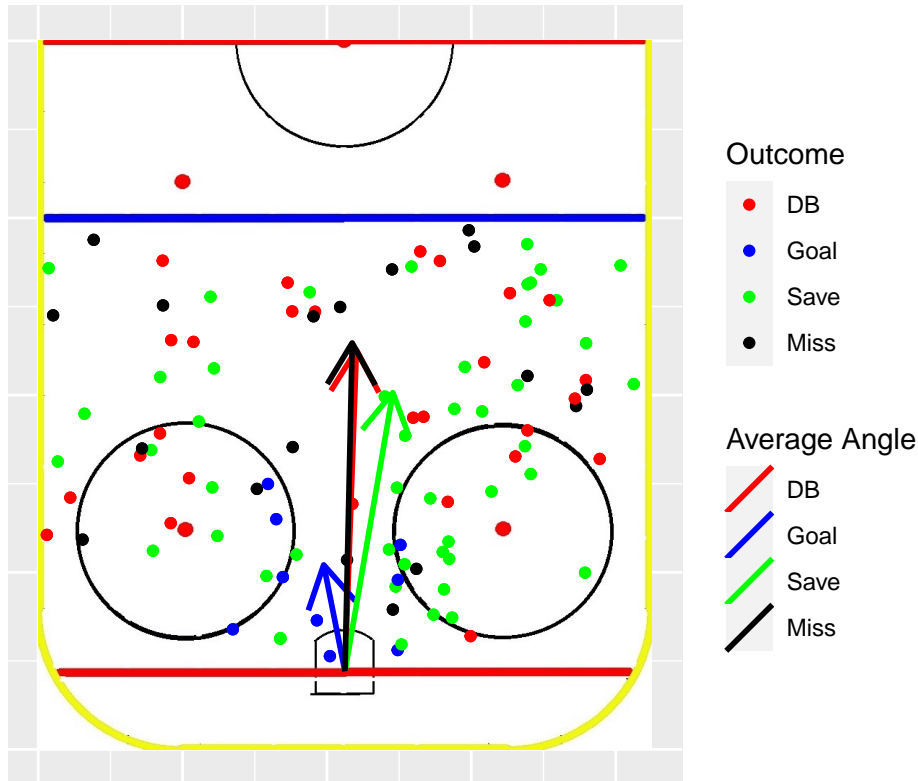

```
# to show the raw numbers used in the creation of the image
print(shot_outcomes)
```

```
## # A tibble: 4 x 6
##   outcome        meanAngle meanDist xstart ystart radius
##   <fct>              <dbl>    <dbl>  <dbl>  <dbl>  <dbl>
## 1 Defender Block      88.0     454.    425   1890    454.
## 2 Goal               101.      153.    425   1890    153.
## 3 Save                80.4     399.    425   1890    399.
## 4 Miss                88.8     464.    425   1890    464.
```

Discussion of results

Provide in natural language a clear discussion of your observations.

This figure is a graph of the rink with the outcome of the shot labeled, arrows representing the average angle of the puck for each outcome, and the length of the arrow representing the average puck distance for each outcome. As expected, the goals were located closer to the goal and all non goals had a larger average puck distance. This aligns with other quick analyses, like the density plot from above, and is expected given general hockey knowledge. Additionally, the average angle for all goals is greater than those of the non goals. The resulting average leaned towards the right side of the ice while all non goals were close to the center or to the left. While this was not common knowledge to me, Dr. Morgan reported that this is expected.

Summary and next steps

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

I will work on making my diagrams more color blind friendly. During the most recent Monday stand up, Dr. Bennett provided me with information about color brewer and I will use it going forward. Dr. Morgan also informed me that the sequences data frame contains more features than the shots_stats data frame. I plan on looking at those features and figuring out what type of visualization is appropriate for them. I also started to work on individual graphs of the players' locations and outcome. This is not entirely necessary but I thought it might help if anyone finds the original image crowded.