

# Age and Decay

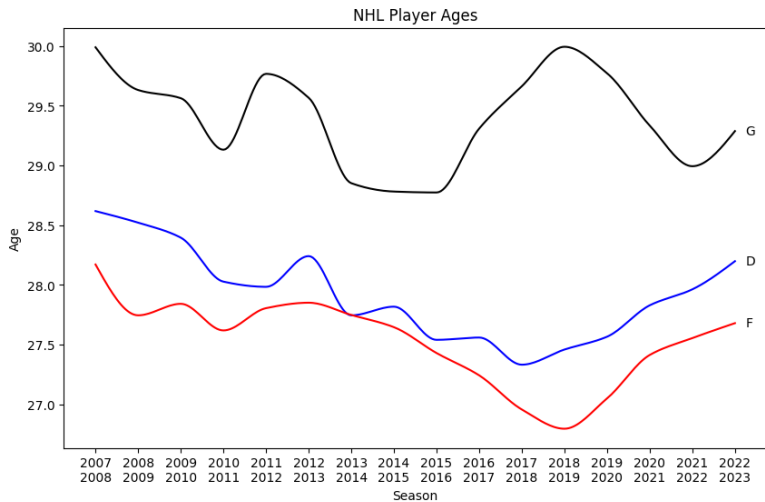
Micah Blake McCurdy  
hockeyviz.com

Seattle, Washington  
Seattle Hockey Analytics Conference  
December 4, 2022

# Introduction

- ▶ How does age affect year-to-year ability among NHL players?

# Age by Position



# Introduction

- ▶ How does age affect year-to-year ability among NHL players?

# Introduction

- ▶ How does age affect year-to-year ability among NHL players?
  - ▶ Key is relationship between the players we get to see and the players we don't.

# Introduction

- ▶ How does age affect year-to-year ability among NHL players?
  - ▶ Key is relationship between the players we get to see and the players we don't.
  - ▶ We don't see them (in the NHL) but they are seen.

# Historical Work

“What does not get observed can be used to make age curves stronger: estimating player age curves using regression and imputation.”

- ▶ Michael Schuckers
- ▶ Michael Lopez
- ▶ Brian Macdonald

# Deceptively simple model

**Observed** One term for each age in years,  $A_{23}^{\bullet}$

**Hidden** One term for each age in years,  $A_{23}^{\circ}$

**Desired** Aging curve:  $wA_{23}^{\bullet} + (1 - w)A_{23}^{\circ}$

Ages run from 18 to 50 (but today you only get to see up to 40)



# Cohort

What cohort of people do we want to discuss the aging of?

# Cohort

What cohort of people do we want to discuss the aging of?

- ▶ All players who have played 1000 minutes of regular season hockey since the start of 2007-2008. (It's 1,589 guys)

# Cohort

What cohort of people do we want to discuss the aging of?

- ▶ All players who have played 1000 minutes of regular season hockey since the start of 2007-2008. (It's 1,589 guys)
- ▶ Ability data from 2007-2008 and 2022-2023 is **excluded**.

## Data: Mid-career years

Player ability result of  $Y_{23}$ , encoded as  $A_{23}^\bullet$

## Data: Mid-career years

Player ability result of  $Y_{23}$ , encoded as  $A_{23}^\bullet$

- ▶ The most elaborate way to compute an average you've ever seen.

## Data: Career ending years

- ▶ Player plays at age 38 and then retires.

## Data: Career ending years

- ▶ Player plays at age 38 and then retires.
- ▶ Could just encode result  $Y_{38}$  as  $A_{38}^{\bullet}$

## Data: Career ending years

- ▶ Player plays at age 38 and then retires.
- ▶ Could just encode result  $Y_{38}$  as  $A_{38}^{\bullet}$
- ▶ CANNOT encode result  $Y_{39}$  as anything, because it doesn't exist.



## Data: Career ending years

- ▶ Player plays at age 38 and then retires.
- ▶ Could just encode result  $Y_{38}$  as  $A_{38}^{\bullet}$
- ▶ CANNOT encode result  $Y_{39}$  as anything, because it doesn't exist.
- ▶ If  $Y_{39}$  did exist, we would expect it to satisfy:

$$Y_{38} - Y_{39} \sim A_{38}^{\bullet} - A_{39}^{\bullet}$$

## Data: Career ending years

- ▶ Player plays at age 38 and then retires.
- ▶ Could just encode result  $Y_{38}$  as  $A_{38}^{\bullet}$
- ▶ CANNOT encode result  $Y_{39}$  as anything, because it doesn't exist.
- ▶ If  $Y_{39}$  did exist, we would expect it to satisfy:

$$Y_{38} - Y_{39} \sim A_{38}^{\bullet} - A_{39}^{\bullet}$$

- ▶ So we cheat:

$$Y_{38} - A_{39}^{\circ} \sim A_{38}^{\bullet} - A_{39}^{\bullet}$$

## Data: Career ending years

- ▶ Player plays at age 38 and then retires.
- ▶ Could just encode result  $Y_{38}$  as  $A_{38}^{\bullet}$
- ▶ CANNOT encode result  $Y_{39}$  as anything, because it doesn't exist.
- ▶ If  $Y_{39}$  did exist, we would expect it to satisfy:

$$Y_{38} - Y_{39} \sim A_{38}^{\bullet} - A_{39}^{\bullet}$$

- ▶ So we cheat:

$$Y_{38} - A_{39}^{\circ} \sim A_{38}^{\bullet} - A_{39}^{\bullet}$$

- ▶ That is:

$$Y_{38} \sim A_{39}^{\circ} + A_{38}^{\bullet} - A_{39}^{\bullet}$$

## Data: Career beginning years

The same thing only the other way around.

## Data: Career beginning years

The same thing only the other way around.

- ▶ When there are more entries/exits near certain ages, the hidden and observed values for those ages are tied together more tightly.

# Missing data

What do we do about the data that is hidden from us?

# Missing data

What do we do about the data that is hidden from us?

- ▶ We guess.

# Missing data

What do we do about the data that is hidden from us?

- ▶ We impute.



# Missing data

What do we do about the data that is hidden from us?

- ▶ We impute.
- ▶ Simulate from normal distribution:
  - ▶ Centred at the **observed** average
  - ▶ With a suitably inflated variance
  - ▶ Truncated at the 75th percentile

# Missing data

What do we do about the data that is hidden from us?

- ▶ We impute.
- ▶ Simulate from normal distribution:
  - ▶ Centred at the **observed** average
  - ▶ With a suitably inflated variance
  - ▶ Truncated at the 75th percentile
    - ▶ This percentile controls how much we trust the gatekeepers to do their jobs properly.

# Fitting

- ▶ Linear regression (on maps, or numbers, as you please) with ridge penalties.

# Fitting

- ▶ Linear regression (on maps, or numbers, as you please) with ridge penalties.
  - ▶ Mild smoothing penalties for observed/hidden.

# Fitting

- ▶ Linear regression (on maps, or numbers, as you please) with ridge penalties.
  - ▶ Mild smoothing penalties for observed/hidden.
  - ▶ Smoothing penalties for the **full** aging curve

# Fitting

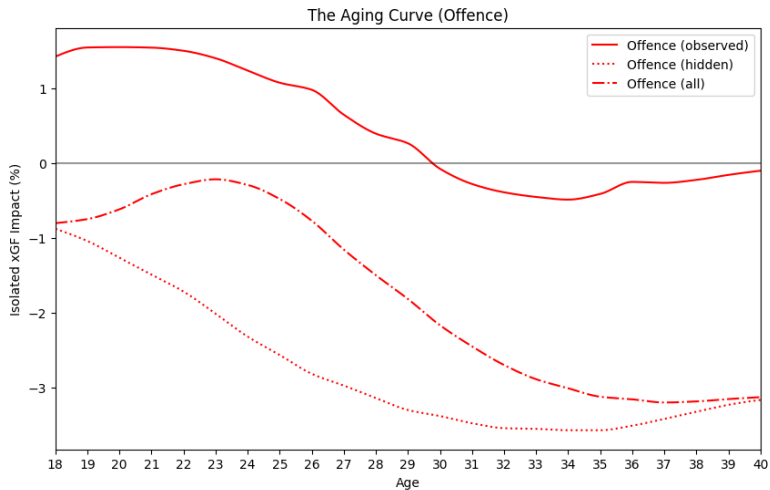
- ▶ Linear regression (on maps, or numbers, as you please) with ridge penalties.
  - ▶ Mild smoothing penalties for observed/hidden.
  - ▶ Smoothing penalties for the **full** aging curve
  - ▶ For instance: full cohort ability at age 23 should be similar to full cohort ability at age 24.

# Fitting

- ▶ Linear regression (on maps, or numbers, as you please) with ridge penalties.
  - ▶ Mild smoothing penalties for observed/hidden.
  - ▶ Smoothing penalties for the **full** aging curve
  - ▶ For instance: full cohort ability at age 23 should be similar to full cohort ability at age 24.
  - ▶ That is:

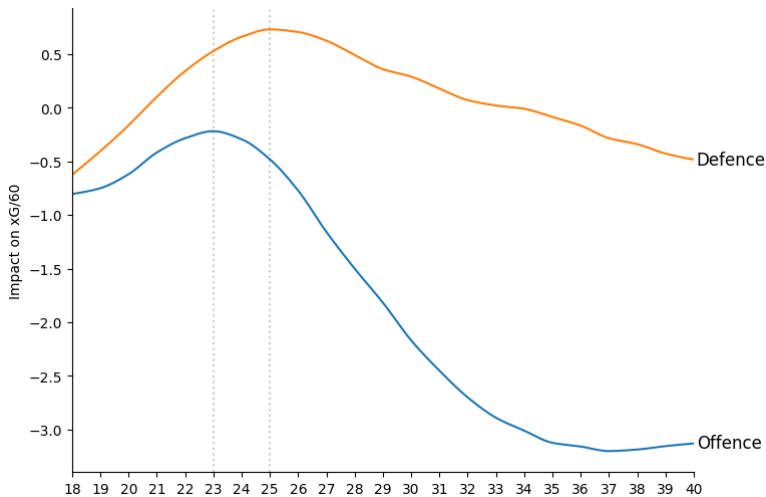
$$w_{23}A_{23}^{\bullet} + (1 - w_{23})A_{23}^{\circ} \sim w_{24}A_{24}^{\bullet} + (1 - w_{24})A_{24}^{\circ}$$

# Offence

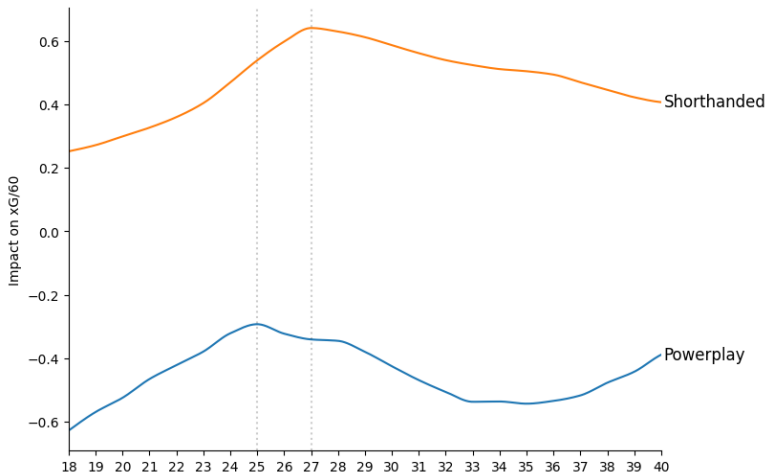




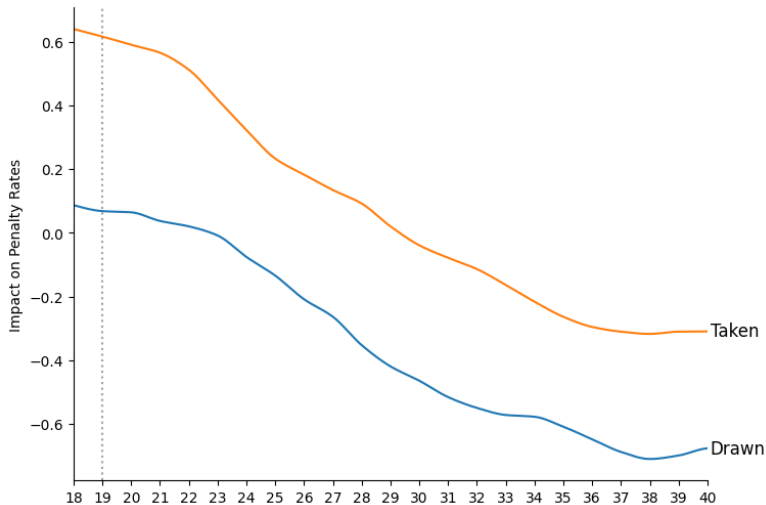
## 5v5 xG Rates



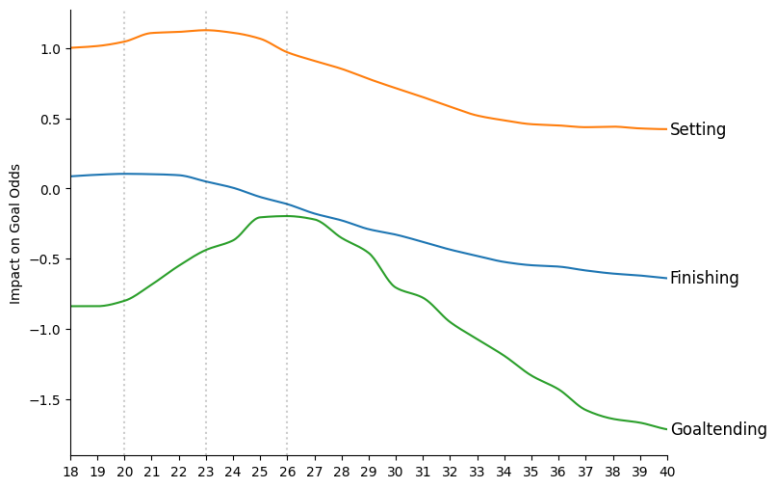
# Special Teams xG Rates



# Penalties



# Goal Threat



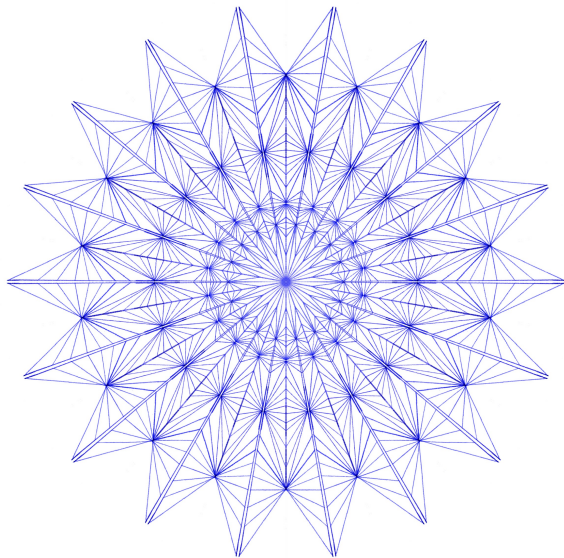
# Conclusions

- ▶ Peaks and slopes are different for different skills.
- ▶ Some players will peak later, but then some will peak earlier.
- ▶ Average declines are slowwww.

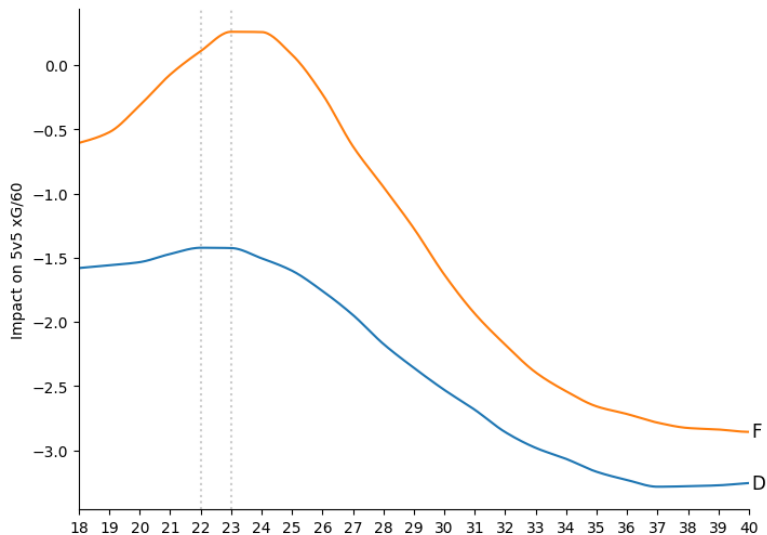
# Future Work

- ▶ Specific percentiles
- ▶ Replace imputation structure with a more sophisticated model.
- ▶ Experience vs wear & tear.
- ▶ Individual / subgroup distinctions.

Thanks!

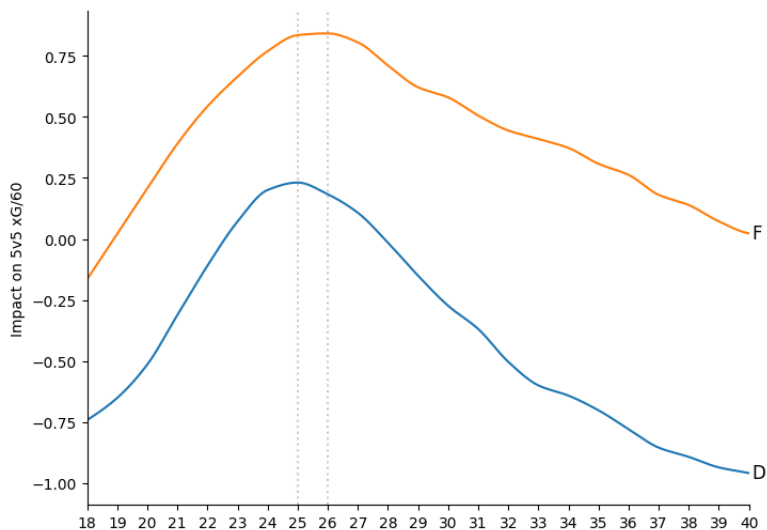


# Offence

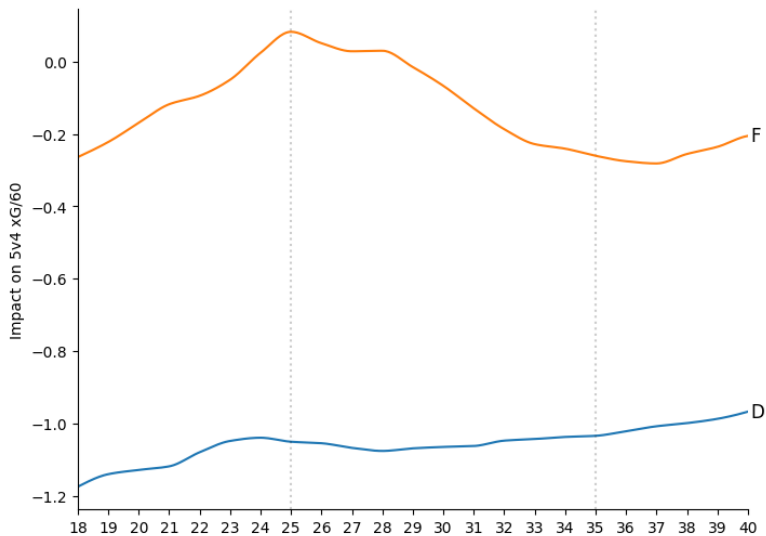




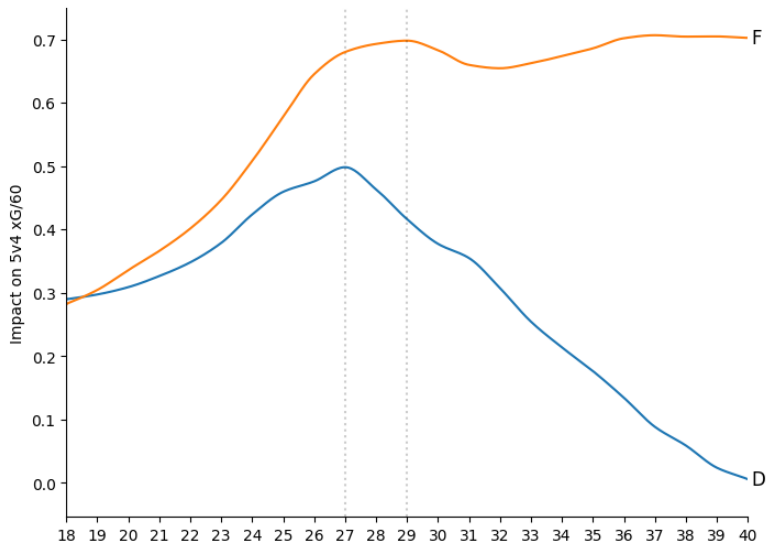
# Defence



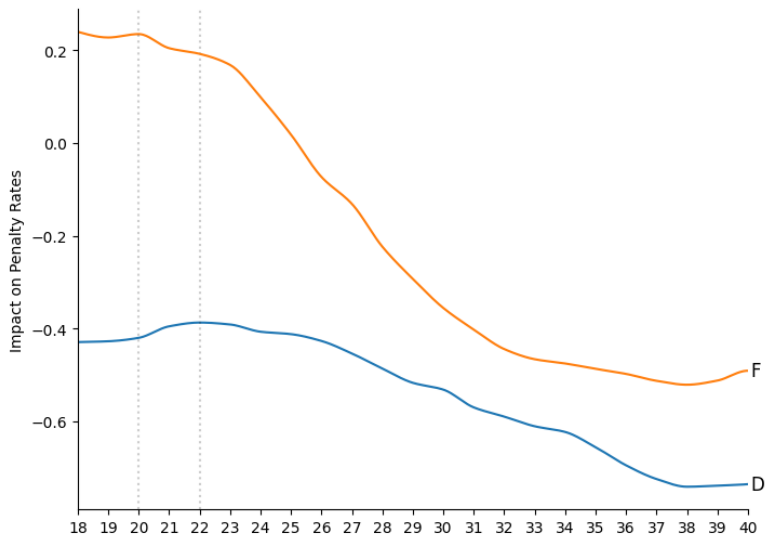
# Powerplay



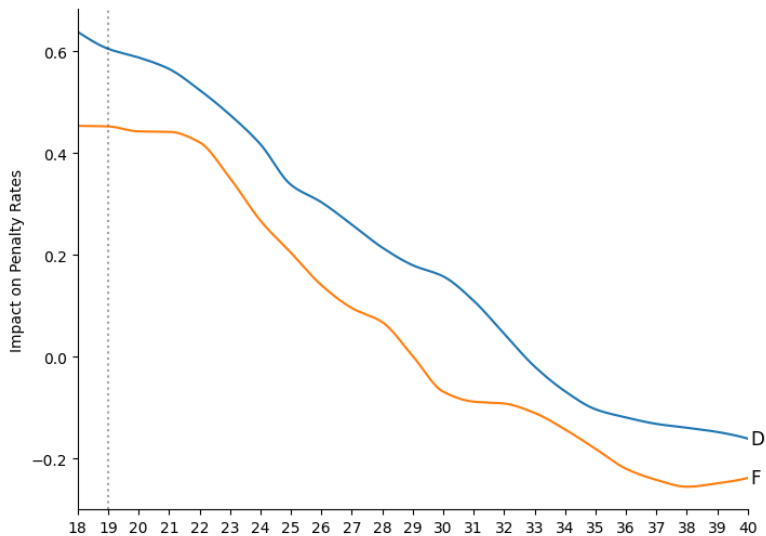
# Shorthanded



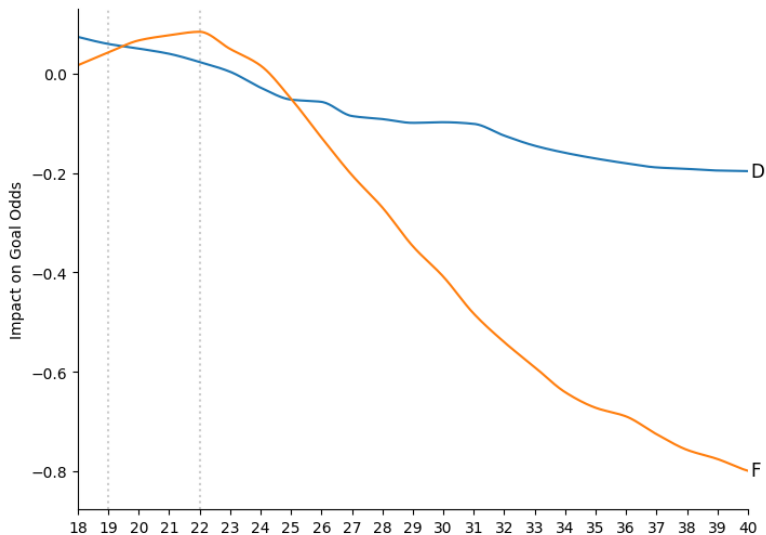
# Drawn



# Taken



# Finishing



# Setting

