DEVELOPMENT POINT SHARES

JOSH POHLKAMP-HARTT, NOV 20 2024

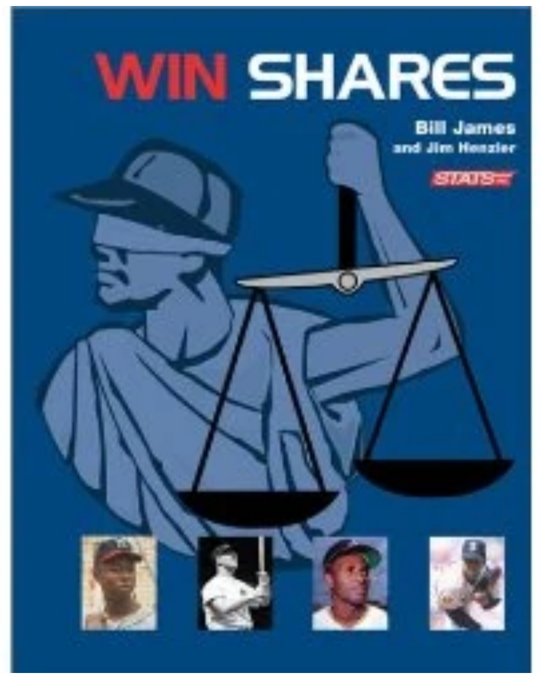
Implementation of Point Share methodology for player development.
Project by Research Fellow Claire Dudley

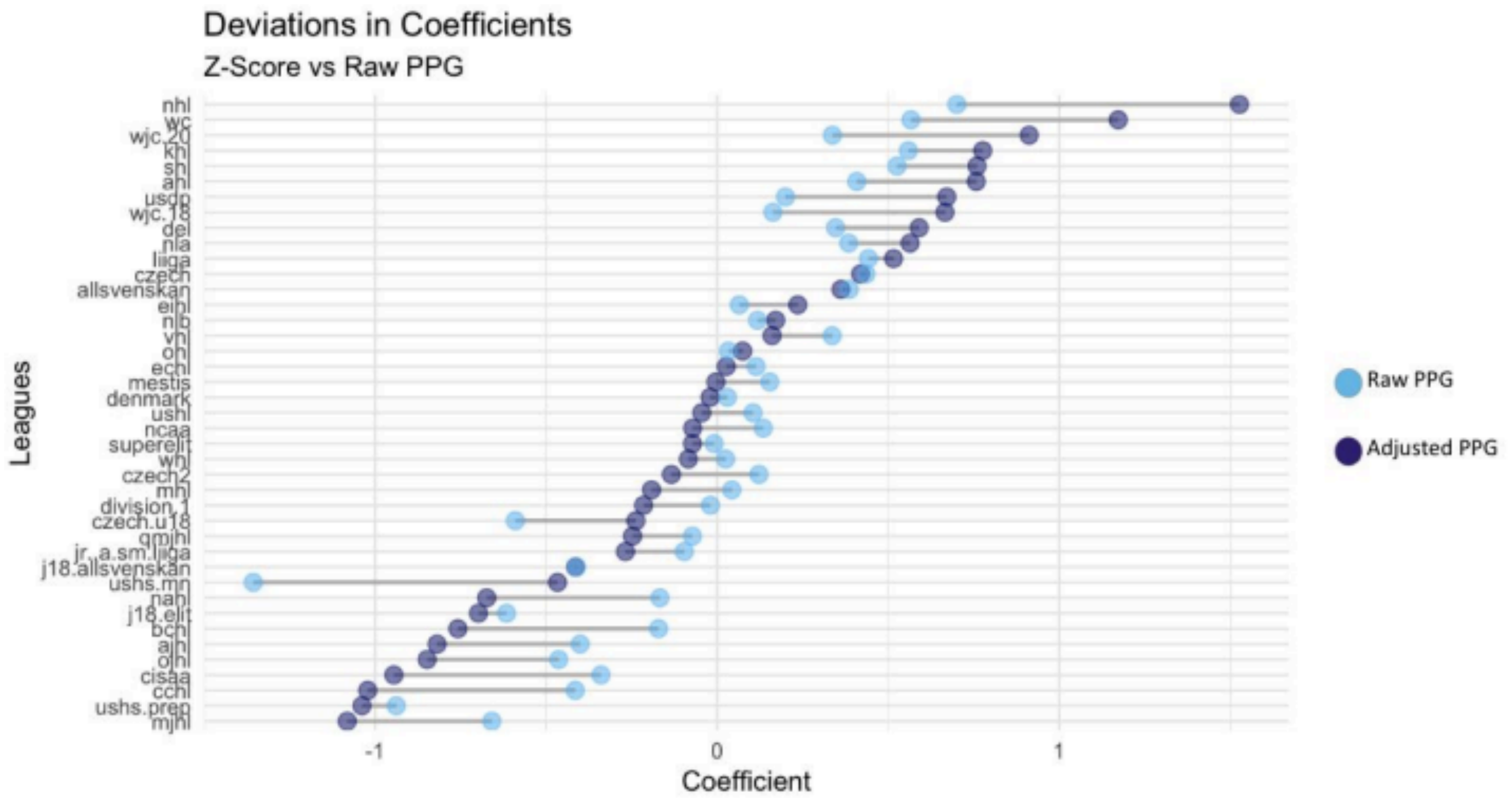
HOW DO WE MEASURE PLAYER DEVELOPMENT

- Unlike a lot of sports, once a player is drafted they do not normally join their pro organization. Instead they play 1-4 years outside a teams control.
- Players are in leagues around the world with varying levels of skill and data availability.
- We have dedicated staff for coaching our developing players. They would like insights both at a micro and macro level on player performance for actionable feedback.
- Request: build league and age agnostic value of players as they develop.

LITERATURE REVIEW

- Win Shares (Baseball) - Bill James (2001, 2002)
- Point Shares (NBA) - Justin Kubatko (2016)
- Point Shares (NHL) - Justin Kubatko (2022)
- League Equivalencies (NHL) - Gabriel Desjardins (2004)
- Which League Is Best (NHL) - Katerina Wu & Madeline Gal (2020)





PROCESS

1. Build Point Shares By League
2. Normalize Point Shares within League
3. Standardize Point Shares Across Leagues
4. Build Development Age Curves
5. Model Probability of Becoming NHL Player

DATA SOURCE

- We use EliteProspects.com as our data source.
- They offer an API for accessing their data.
- They have the widest range of leagues and seasons.
- Their data can be broken down by team or player at a season level.
- This data does not include TOI.
- Ex. David Pastrnak

S ▲	TEAM	LEAGUE	GP	G	A	TP	PIM	+/-	POST	GP	G	A	TP	PIM	+/-
2012-13	Södertälje SK J20	J20 SuperElit	36	12	17	29	67	6	Playoffs	4	2	2	4	10	2
	Södertälje SK	HockeyAllsvenskan	11	2	1	3	0	0	Kvalserien SHL	5	0	0	0	0	-3
2013-14	Södertälje SK J20	J20 SuperElit	1	1	1	2	0	-1	Playoffs	2	0	0	0	0	-5
	Södertälje SK	HockeyAllsvenskan	36	8	16	24	24	7							
2014-15	Boston Bruins	NHL	46	10	17	27	8	12							
	Providence Bruins	AHL	25	11	17	28	12	15	Playoffs	3	0	0	0	0	0
2015-16	Boston Bruins	NHL	51	15	11	26	20	3							
	Providence Bruins	AHL	3	1	3	4	2	1							
2016-17	Boston Bruins	NHL	75	34	36	70	34	11	Playoffs	6	2	2	4	6	1
2017-18	Boston Bruins	NHL	82	35	45	80	37	10	Playoffs	12	6	14	20	8	2
2018-19	Boston Bruins	NHL	66	38	43	81	32	6	Playoffs	24	9	10	19	4	0
2019-20	Boston Bruins	NHL	70	48	47	95	40	21	Playoffs	10	3	7	10	2	-3
2020-21	Boston Bruins	NHL	48	20	28	48	24	10	Playoffs	11	7	8	15	8	3
2021-22	Boston Bruins "A"	NHL	72	40	37	77	20	13	Playoffs	7	3	3	6	2	1
2022-23	Boston Bruins "A"	NHL	82	61	52	113	38	34	Playoffs	7	5	0	5	2	-2
2023-24	Boston Bruins "A"	NHL	82	47	63	110	47	21	Playoffs	13	4	4	8	25	0
2024-25	Boston Bruins "A"	NHL	19	8	9	17	18	-2							

POINT SHARES OVERVIEW

- The Point Shares system is based on a player's proportion of production or production suppression relative to their team's performance.
- Our units are standings points while our calculations are in goals, so we will be normalizing by the rate of goal scoring within a league.
- Players can have negative point shares, justified by the idea that a player can "take away points" his teammates generated (by being bad).
- Offensive Point Shares are calculated using goals created relative to your team.
- Defensive Point Shares are calculated using weighted games, plus/minus adjustment (Δ) to estimate goals against and compared to your team.
- Overall Point Shares are the sum of Offensive and Defensive Point Shares.
  Ie, $Point\ Shares = PS_O + PS_D$

POINT SHARES OFFENSE

$$PS_O = \frac{MGF}{Goals_{league}/points_{league}}$$

$$GC = (goals + 0.5 * assists) * \frac{goals_{team}}{goals_{team} + 0.5 * assists_{team}}$$

$$MGF = GC - \frac{7}{12} * GP * \frac{GC_{team}}{GP_{team}}$$

POINT SHARES DEFENSE

$$PS_D = \frac{MGA}{Goals_{league}/points_{league}}$$

$$MGA_{team} = \left(1 + \frac{7}{12}\right) * GP_{team} * GPG_{league} - GA_{team}$$

$$\Delta = \frac{1}{7} - \partial \; * (+/-) - GP * \frac{+/-_{team,\, position}}{GP_{team,\, position}}$$

$$MGA = weighted\%GP * \frac{5}{7} * \partial * MGA_{team} + \Delta^2$$

- position adjustments (∂) for defensemen are 10/7 and for forwards are 5/7
- the use of 5/7 in MGA represents constant proportion of team MGA assigned to skaters

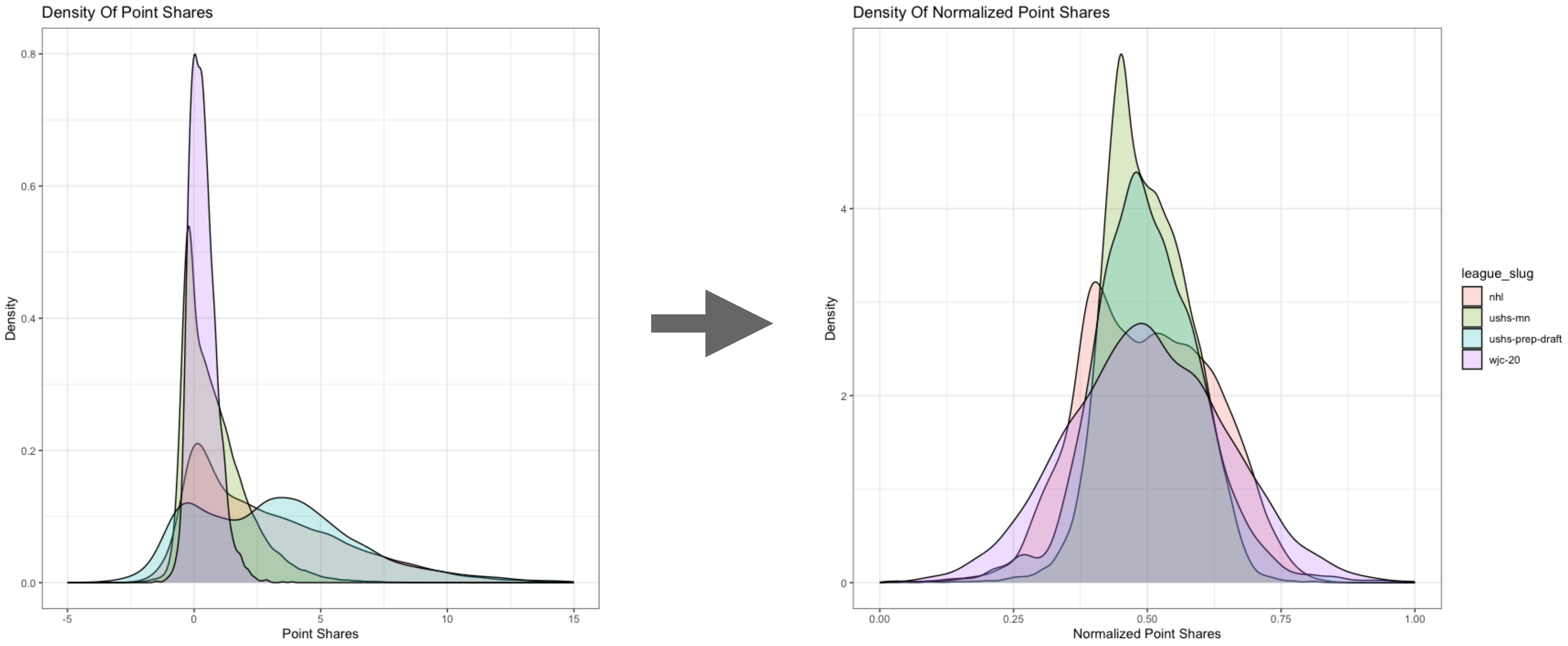
A NOTE ON ASSUMPTIONS AND IMPUTATIONS

- Elite Prospects Data is “slightly” less than complete for some league.
  Ex. Canadian HS does not have +/-, Czech U17 is missing half their players per team.
- We have to make a variety assumptions and imputation choices.
  Here is a rough list:
  - Assume leagues with multiple sub-seasons can be combined and are the same quality.
  - Adjust GP and goals for teams in leagues where there is missing data to be at an expected amount.
  - Some league-wide rates are missing, fill with global averages.
  - Assume teams have enough players to play a game and expand their team totals to match.
  - Make some reasonable guardrails for normalization factors for small sample leagues.

LEAGUE NORMALIZATION

- By league and position we transform the calculated point shares to be approximately a [0,1] bounded Gaussian distribution.
- This is done so we have scores on a scale that our non-analytics staff can understand and that the conversions from league to league can be simple fractions.
- Steps:
  - Yeo-Johnson Transformation to make the data more Normal
  - Left align the data to 0 by removing the min value
  - Divide by the range
- $PS_{Norm} = \dfrac{PS_{Yj} - \min(PS_{Yj})}{\max(PS_{Yj}) - \min(PS_{Yj})}$,
  $PS_{Yj} = \text{Yeo-Johnson}(PS)$

LEAGUE NORMALIZATION



LEAGUE STANDARDIZATION

- Some leagues are stronger than others and we want to be able to compare them.
- To do so, we will make each league's distribution relative to the NHL.
  To be exact, if you were playing in the NHL next year what would your normalized point share be give your current normalized point share?
- We match players' careers into season, league pairs and create zero-intercept linear regression models for each pair of leagues.

	player_id	age	season	old_league	new_league	old_PS	new_PS
	<dbl>	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>
1	1280	23	1999-2000	ncaa	ahl	0.784	0.402
2	1329	23	2000-2001	ncaa	ahl	0.768	0.598
3	2616	22	2002-2003	ncaa	ahl	0.837	0.696
4	3210	20	2002-2003	ncaa	ahl	0.806	0.544
5	3592	23	1998-1999	ncaa	ahl	0.762	0.431
6	3603	22	1998-1999	ncaa	ahl	0.768	0.572
7	3735	21	1999-2000	ncaa	ahl	0.812	0.772
8	4202	22	2008-2009	ncaa	ahl	0.905	0.706
9	4253	21	2000-2001	ncaa	ahl	0.763	0.484
10	4375	21	1999-2000	ncaa	ahl	0.907	0.694

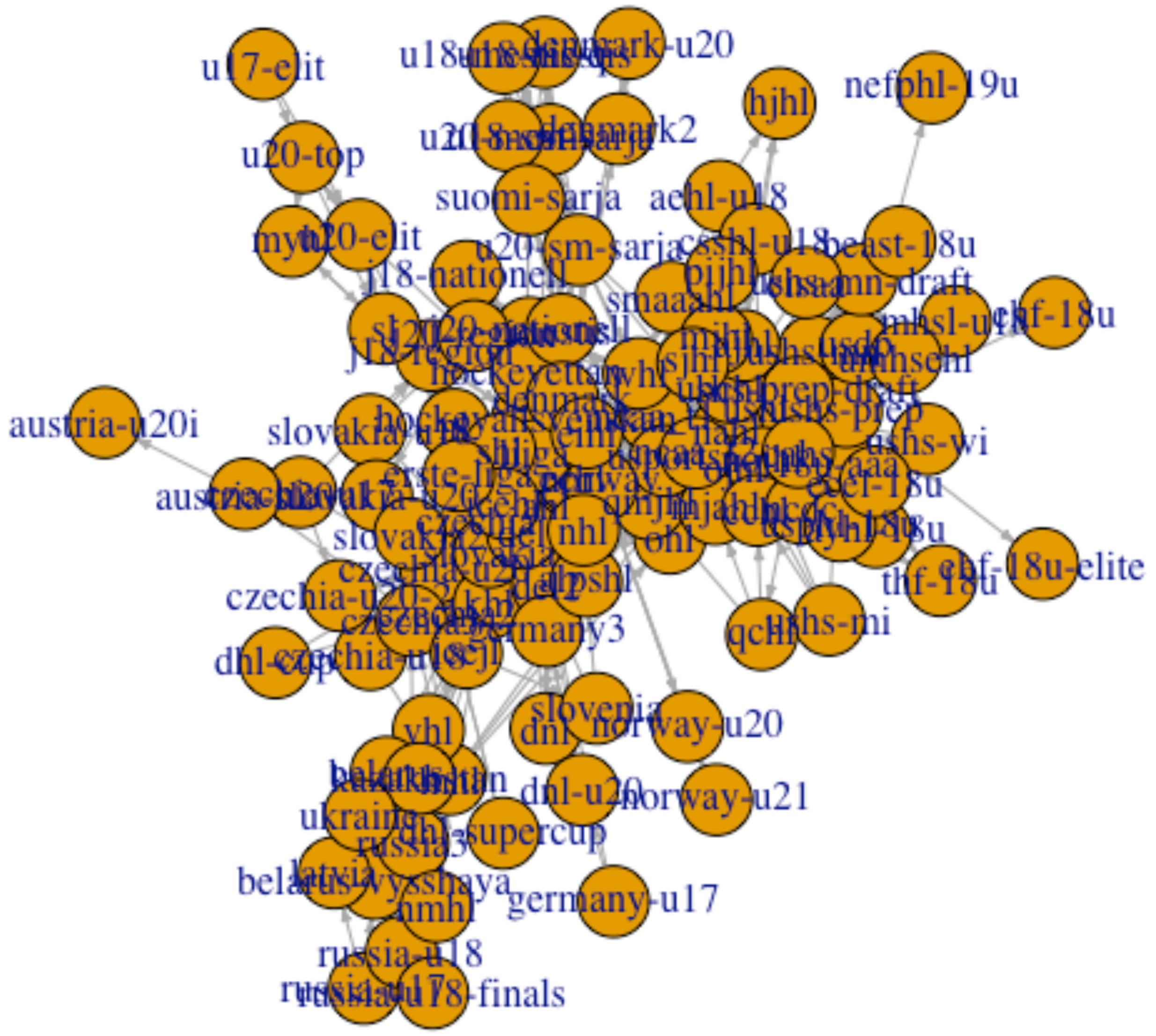
ℹ 2,104 more rows

{NCAA, AHL} = .6311

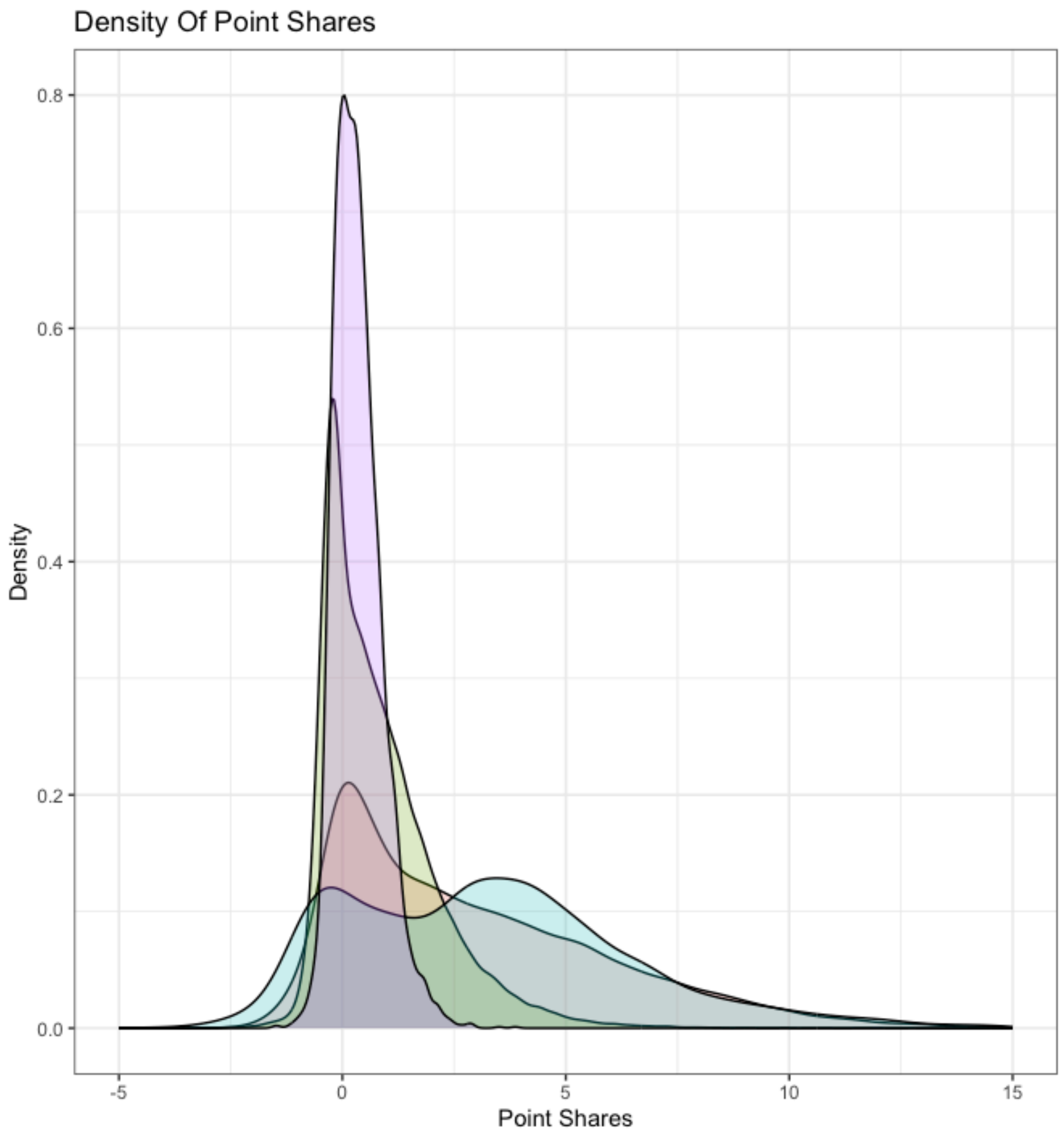
- We weight the samples in the regression such that more recent seasons are more important.
- There must be 10 samples for us to build a coefficient.

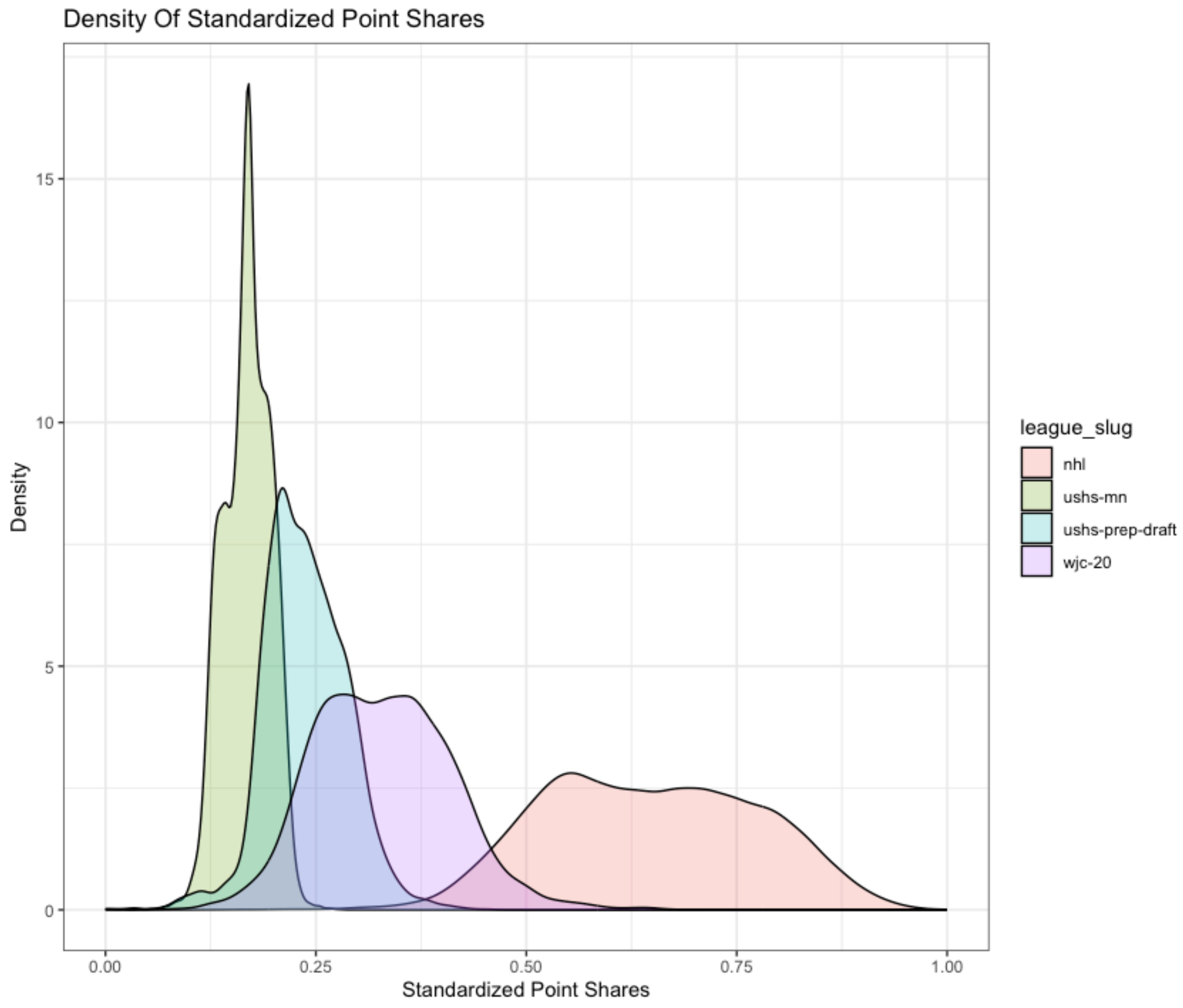
LEAGUE STANDARDIZATION

- Not all leagues feed directly to the NHL so we use the “Wilson Method” which is to assume that we can treat the strengths as transitive and chain them between linked league to get to the NHL.
  Ex: ECHL ➤ AHL ➤ NHL, {ECHL, NHL} = {ECHL, AHL} x {AHL, NHL} = .72 x .84 = .605
- To do this at scale, we need to find the shortest/strongest paths between each league and the NHL and perform the Wilson Method on that path.
- This is done by building a directed graph between leagues with the inverse of the # of players being the edge weights and then finding the lowest weight path.
- To avoid edge cases, we assume all leagues travel through the AHL.
- We also treat tournaments like the Olympics separately.



LEAGUE STANDARDIZATION

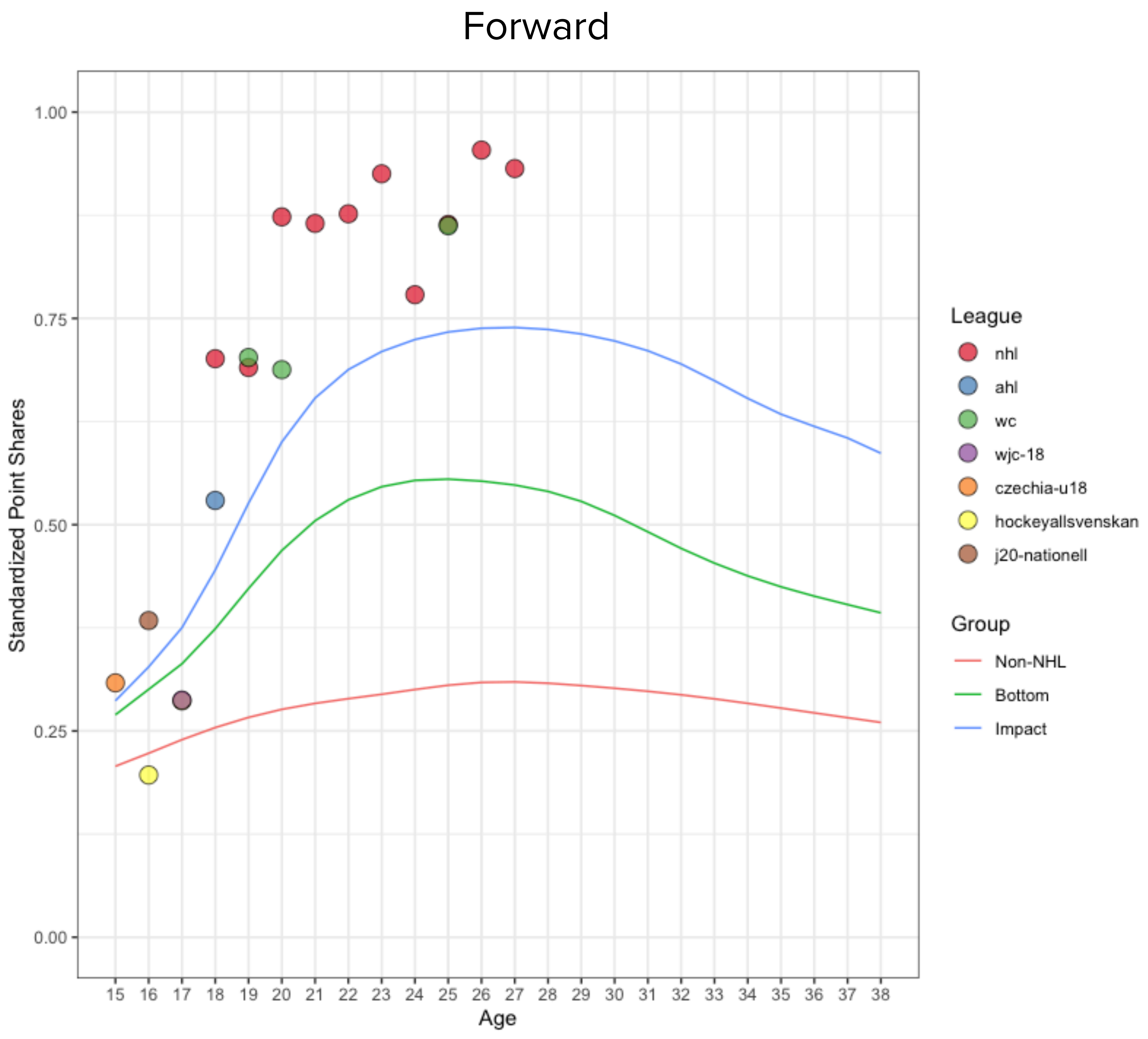




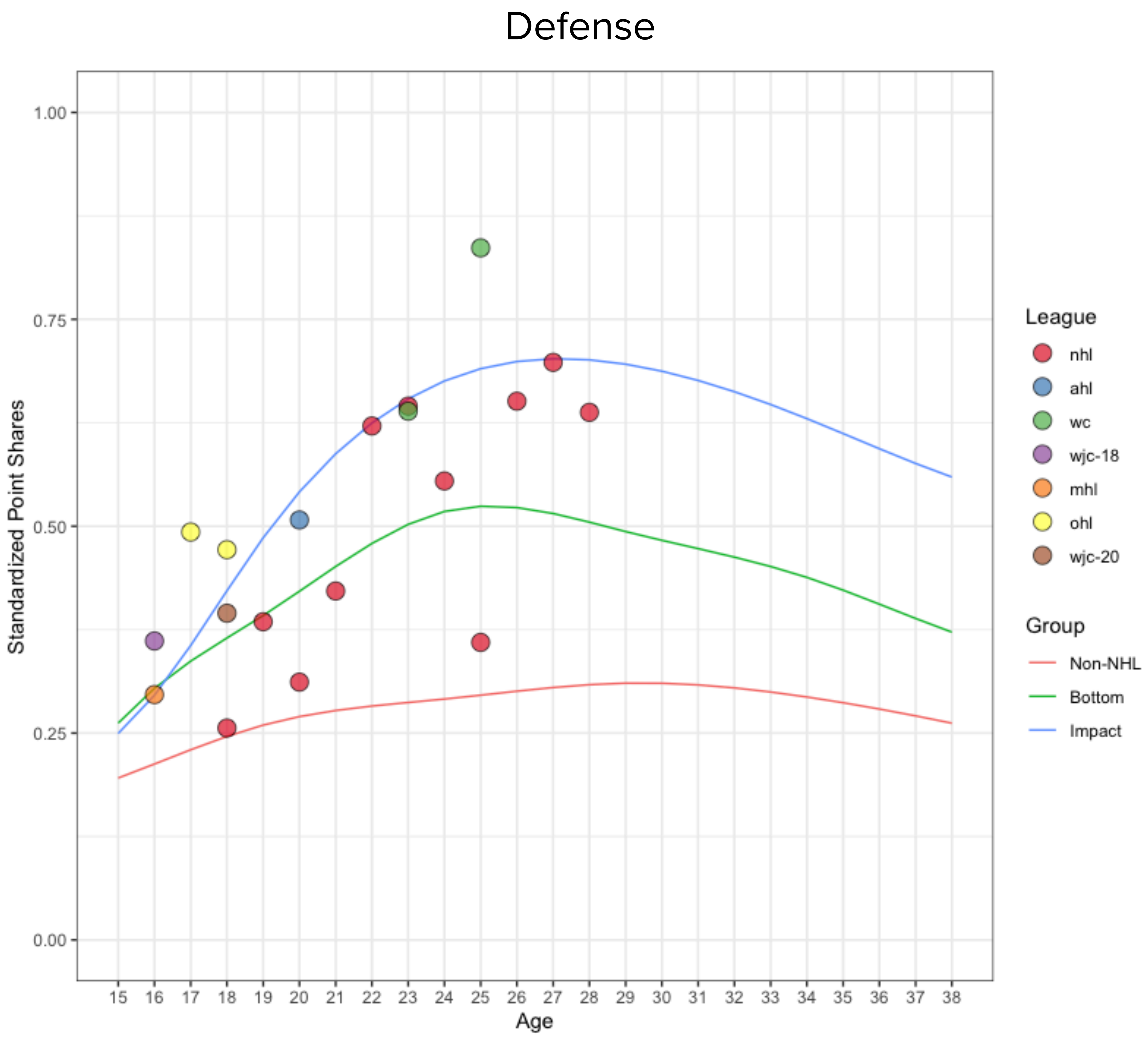
AGE CURVES

- To model development we want to see how players age. Players that are successful in the NHL will age differently than those don't.
- For this we first categorized players into 3 groups:
  NHL impact players: more than a third of their career is as a top 3rd player by PS
  NHL bottom of lineup players: not impact but 5 or more seasons
  Non-NHL players: less than 5 NHL seasons (99% of players)
- We then build mixed effects spline regression models for each group with players as random effects and age modeled by B-Splines.

AGE CURVES



Forward
Standardized Point Shares
1.00
0.75
0.50
0.25
0.00
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
Age
League
nhl
ahl
wc
wjc-18
czechia-u18
hockeyallsvenskan
j20-nationell
Group
Non-NHL
Bottom
Impact

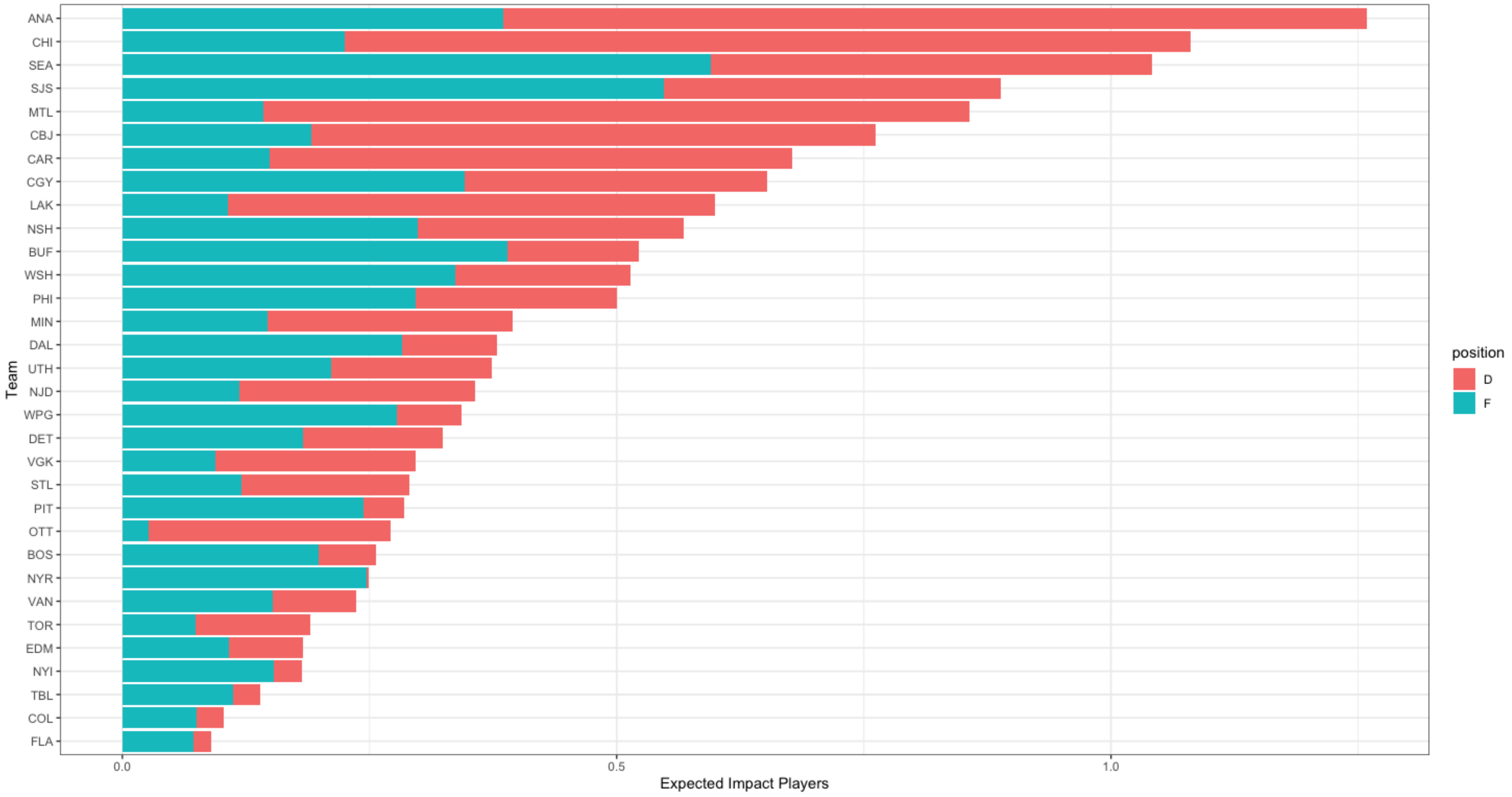


Defense
Standardized Point Shares
1.00
0.75
0.50
0.25
0.00
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
Age
League
nhl
ahl
wc
wjc-18
mhl
ohl
wjc-20
Group
Non-NHL
Bottom
Impact

NHL PROBABILITY MODEL

- For each player that is actively in development, we would like to know what the probability that they will be an impact NHL player.
- To determine this we use build hierarchical probability model where we are first predicting if a player will play 5+ NHL seasons and then if so are they an impact player. Ie. P(Impact) = P(Impact | NHL) x P(NHL).
  This is done instead of having a multinomial regression so that we have larger and more balanced sample sizes.
- We build the models using XGBOOST and treat forwards and defense separately. Our features are the distance from the impact curve of the player's point shares for the last 3 seasons and their age.
- Downsampling is done to balance our classes for NHL vs Non-NHL so we have to calibrate our probabilities, which we do with polynomial regression.
- With these probabilities we can value each player within a team's development system.

NHL PROBABILITY MODEL



IMPLEMENTATION IN PRODUCTION

R script that runs in Docker.

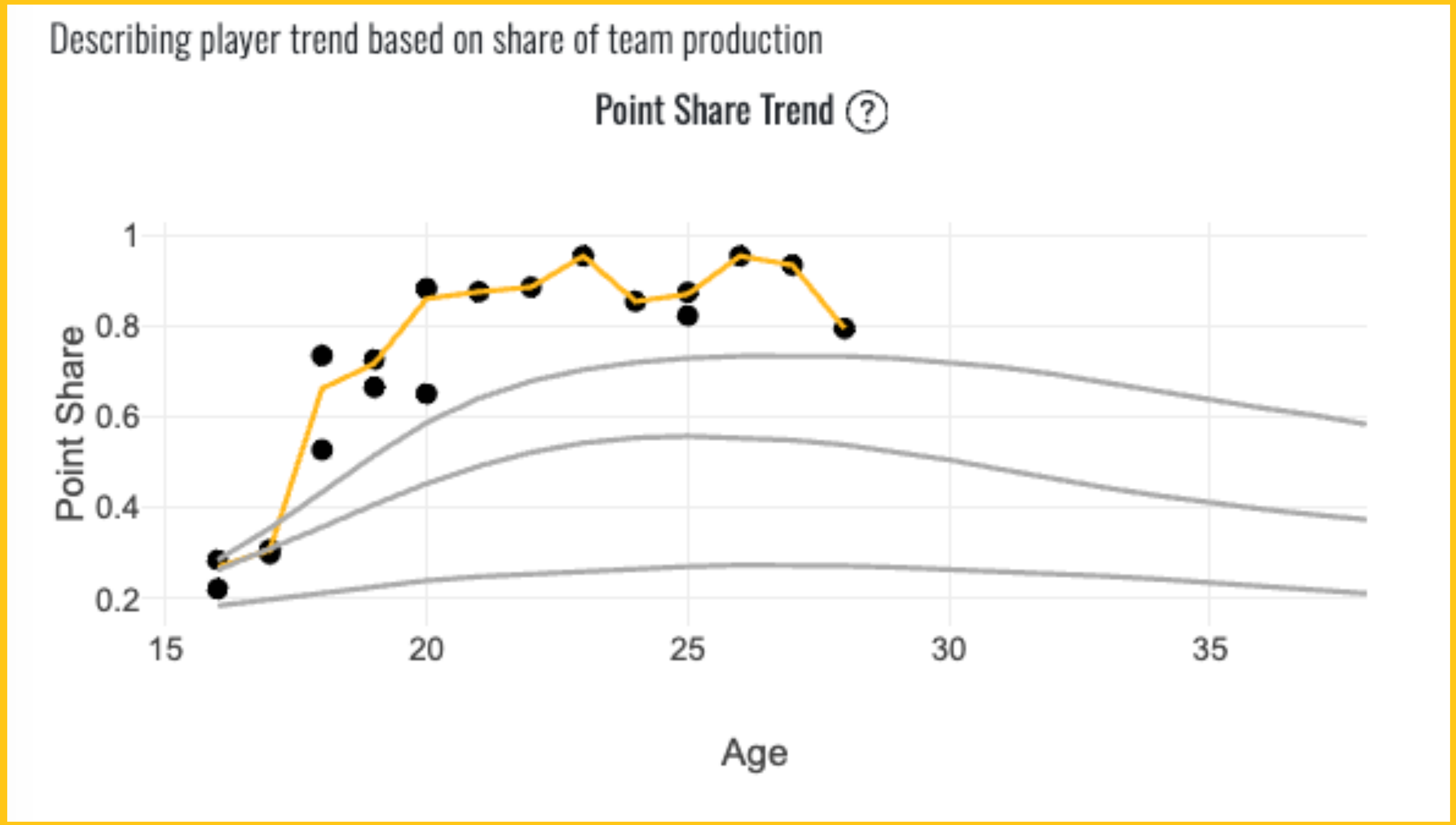
EP Tables pulled from API and stored in S3.

Cron job.

R output is stored back in S3.

Internal Website hosts the data.

Prob model runs in a custom pipeline.



THANK YOU, QUESTIONS?