# Evaluating Overagers + The NHL Teams Who Draft Them

Namita Nandakumar
@nnstats

| Overall | Player | Age |
|---|---|---|
| 62 | Joonas Korpisalo | 18 |
| 63 | Jujhar Khaira | 18 |
| 64 | Tim Bozon | 18 |
| 65 | Adam Pelech | 18 |
| 66 | Jimmy Vesey | 19 |
| 67 | Mackenzie MacEachern | 18 |
| 68 | John Draeger | 18 |
| 69 | Daniel Altshuller | 18 |
| 70 | Scott Kosmachuk | 18 |
| 71 | Tanner Richard | 19 |
| 72 | Troy Bourke | 18 |
| 73 | Justin Kea | 18 |
| 74 | Esa Lindell | 18 |
| 75 | Jon Gillies | 18 |
| 76 | Chris Driedger | 18 |
| 77 | Chandler Stephenson | 18 |
| 78 | Shayne Gostisbehere | 19 |

*2012 draft snapshot*
*via Hockey Reference*

# Stuff to know before we continue.

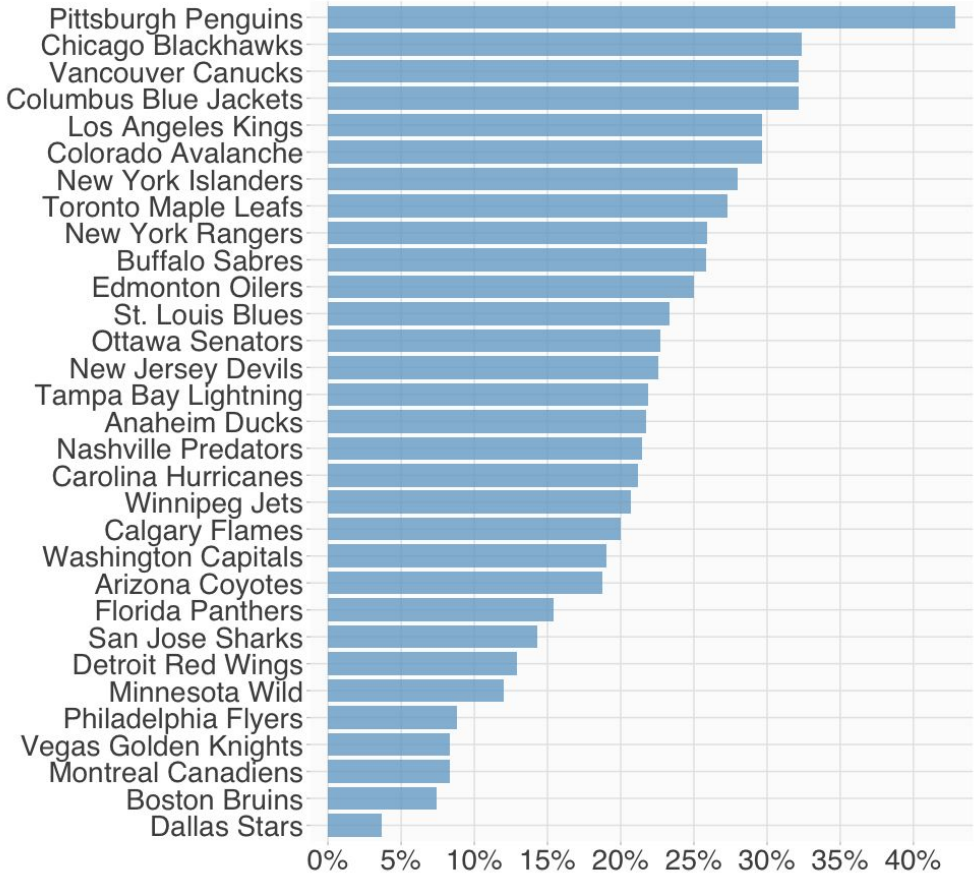
- Hockey is a sport, 31 NHL teams, 7 draft rounds, etc. 🏒
- If a prospect turns 18 by September 15$^{th}$ of the year of the draft, he is eligible to be drafted.
- Most drafted players are in their first year of eligibility. Some (~20%) are in their second (or third, or fourth...).
- Notable recent overager picks include Viktor Arvidsson, Freddy Andersen, Shayne Gostisbehere, and more.

—

# % of Picks Spent on Overagers

***2014-17 NHL Drafts***

*Try to spot a team that's acting a little differently than the rest.*

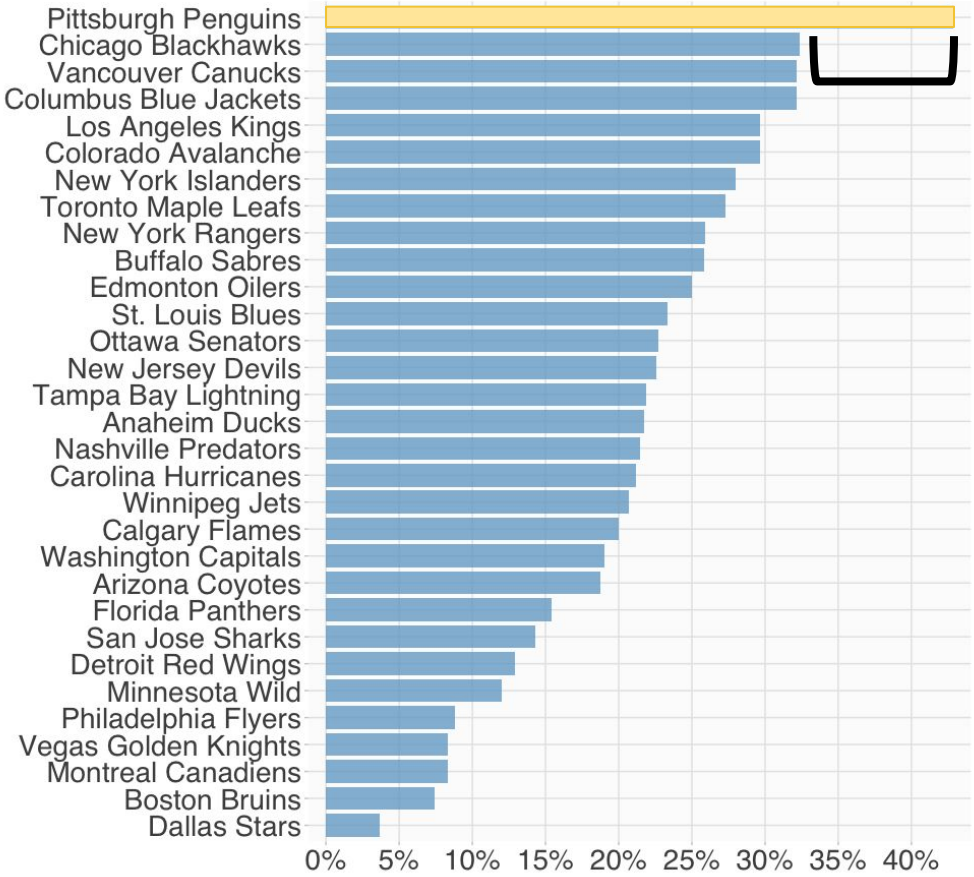


—

# % of Picks Spent on Overagers

***2014-17 NHL Drafts***

*Try to spot a team that's acting a little differently than the rest.*



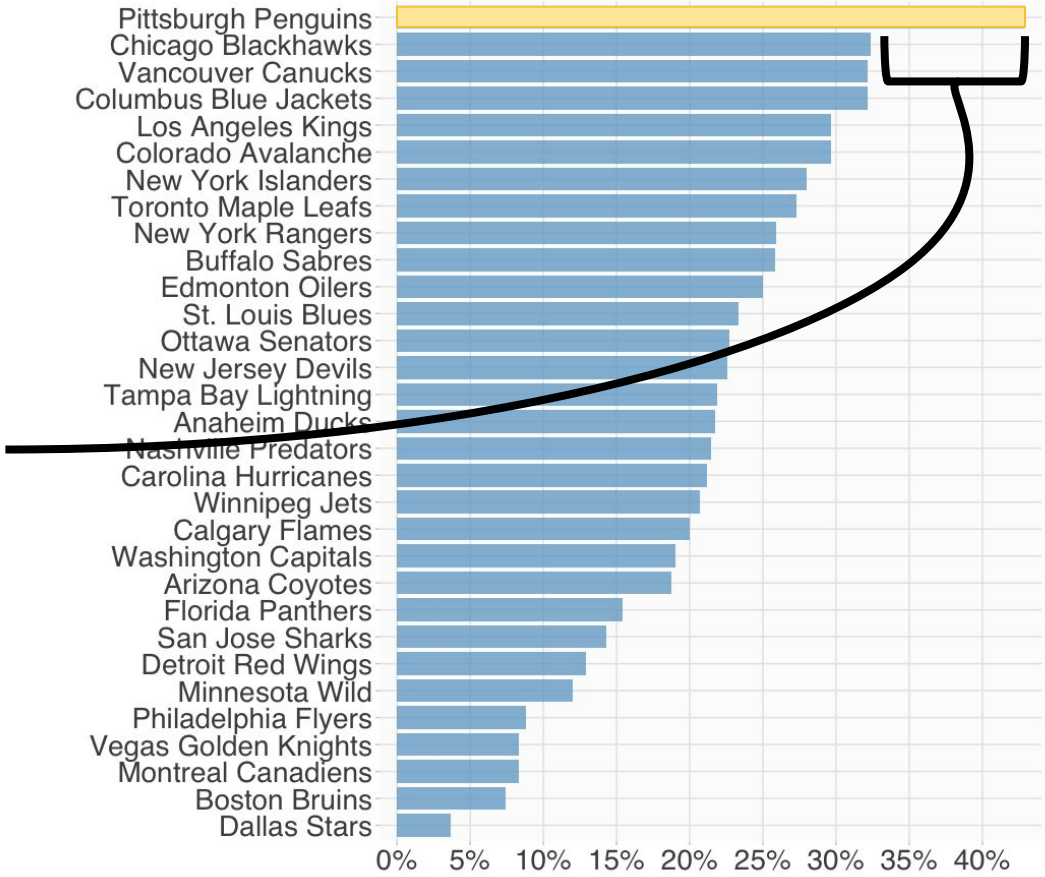
—

# % of Picks Spent on Overagers

***2014-17 NHL Drafts***

*Try to spot a team that's acting a little differently than the rest.*

But wait, can this be explained by:



—

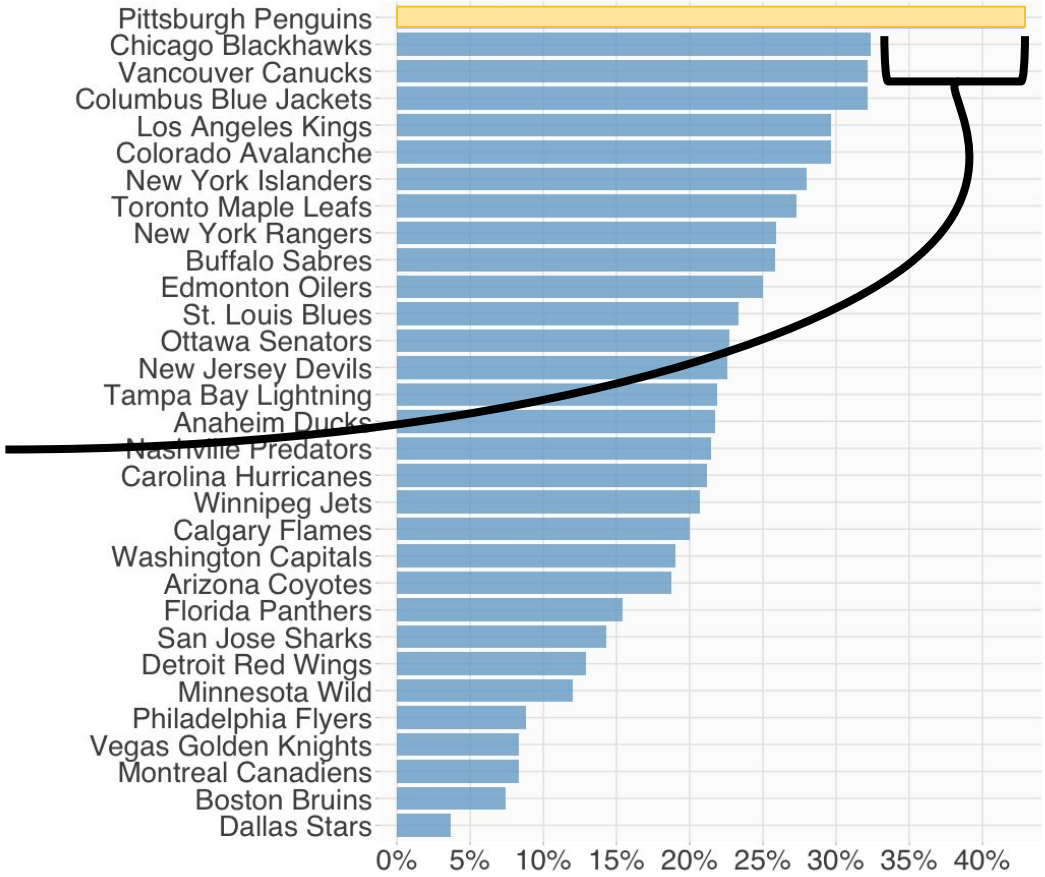
# % of Picks Spent on Overagers

***2014-17 NHL Drafts***

*Try to spot a team that's acting a little differently than the rest.*

But wait, can this be explained by:

- having mostly late round picks for years



—

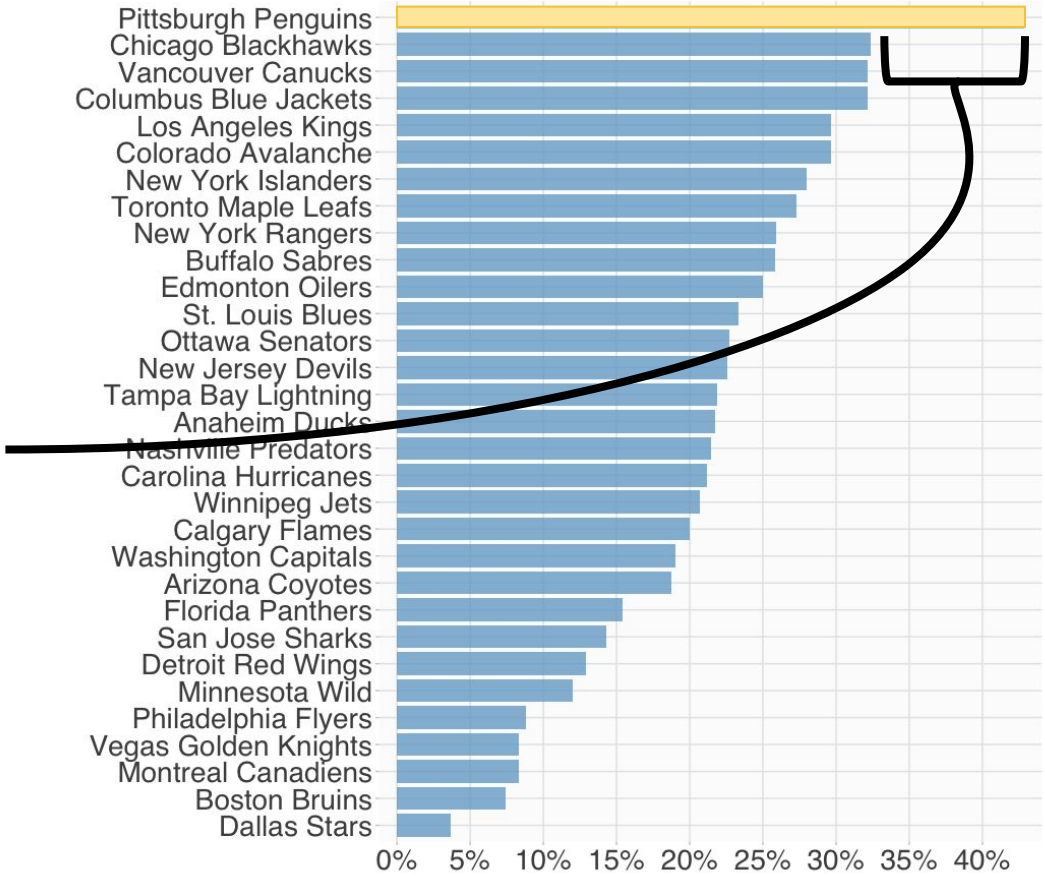
# % of Picks Spent on Overagers

***2014-17 NHL Drafts***

*Try to spot a team that's acting a little differently than the rest.*

But wait, can this be explained by:

- having mostly late round picks for years
- the ever-present chaos and sheer randomness that typifies human existence



—

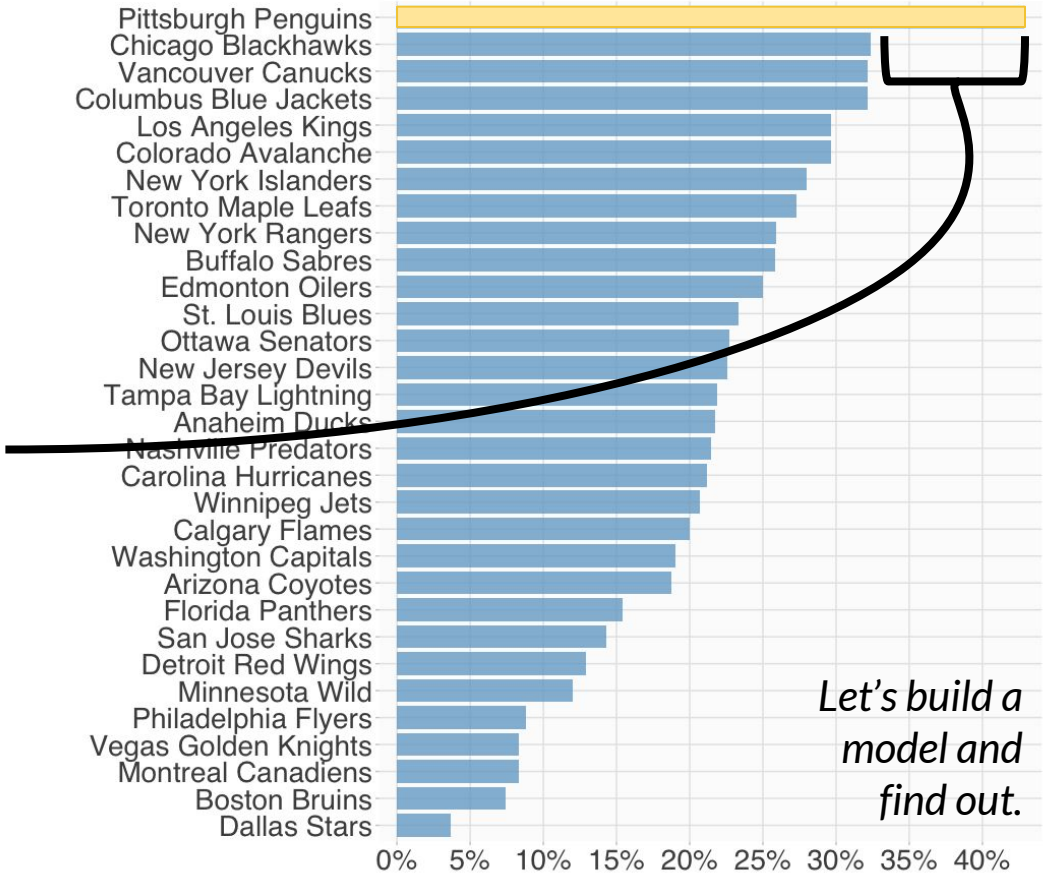
# % of Picks Spent on Overagers

***2014-17 NHL Drafts***

*Try to spot a team that's acting a little differently than the rest.*

But wait, can this be explained by:

- having mostly late round picks for years
- the ever-present chaos and sheer randomness that typifies human existence
- all of the above ???



# *Tired:* logistic regression

*Tired:* logistic regression

*Wired:* logistic regression,
but Bayesian,
but it's still easy
to implement

# Building A Model

- We'd like to predict Pr(selecting an overage prospect) as a function of overall pick #.
- Using the R package rstanarm, we can easily build a logistic model with very weakly informative normal priors centered on 0 for the coefficients.
- Which means the coefficient point estimates will be ~ equivalent to what we see in the frequentist version, with a dash of regularization.

—

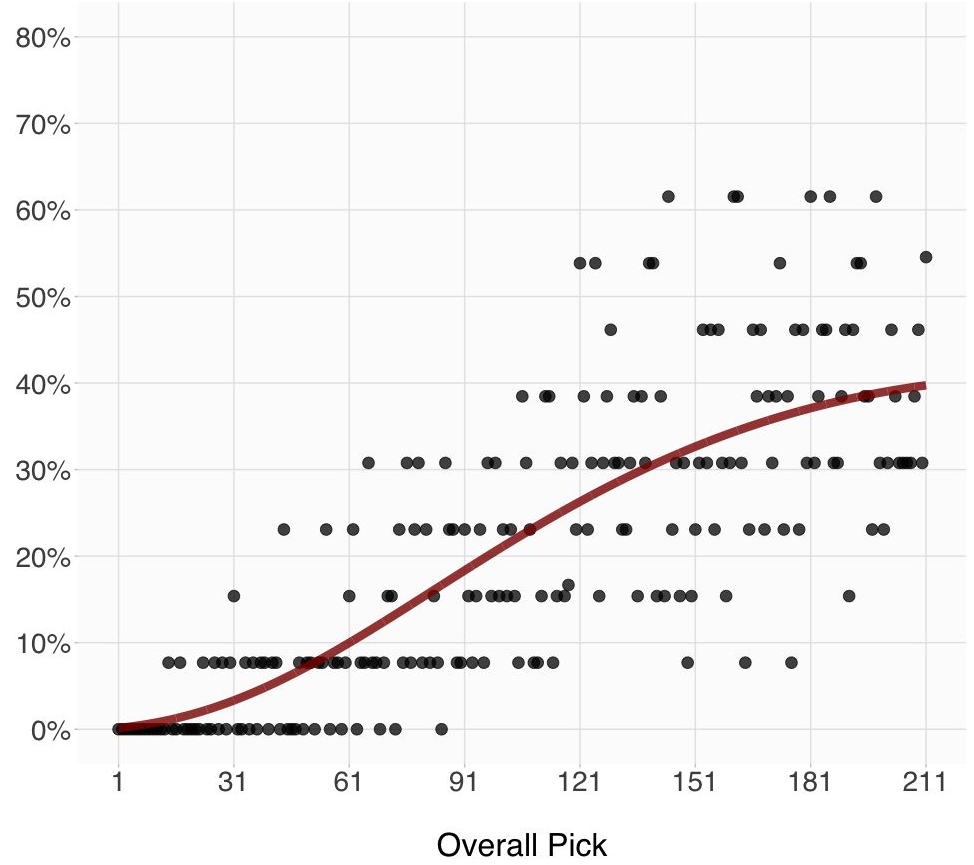
## % of Overage Selections at Every Pick

*2005-17 NHL Drafts*

First things first:

there is a clear relationship between pick # and propensity to select an overager.

Also, our fitted curve looks pretty good.



# Let's compare 2 teams.

- **Penguins** 🐧
- 2014-17 Picks:

| Round # | | | | | | |
|---|---|---|---|---|---|---|
| | 1 \| | 22 | | | | |
| | 2 \| | 46 | 51 | 55 | 61 | |
| | 3 \| | 77 | 93 | | | |
| | 4 \| | 113 | 121 | | | |
| | 5 \| | 137 | 145 | 151 | 152 | 155 |
| | 6 \| | 167 | 173 | 181 | 186 | |
| | 7 \| | 197 | 203 | 217 | | |

- Actual Overager %: *43%*
  Expected Overager %: *26%*
  Difference: **+17%**

- **Golden Knights** 👑
- 20~~14~~-17 Picks:

| Round # | | | | |
|---|---|---|---|---|
| | 1 \| | 6 | 13 | 15 |
| | 2 \| | 34 | 62 | |
| | 3 \| | 65 | | |
| | 4 \| | 96 | | |
| | 5 \| | 127 | 142 | |
| | 6 \| | 158 | 161 | |
| | 7 \| | 189 | | |

- Actual Overager %: *8%*
  Expected Overager %: *18%*
  Difference: **-10%**

—

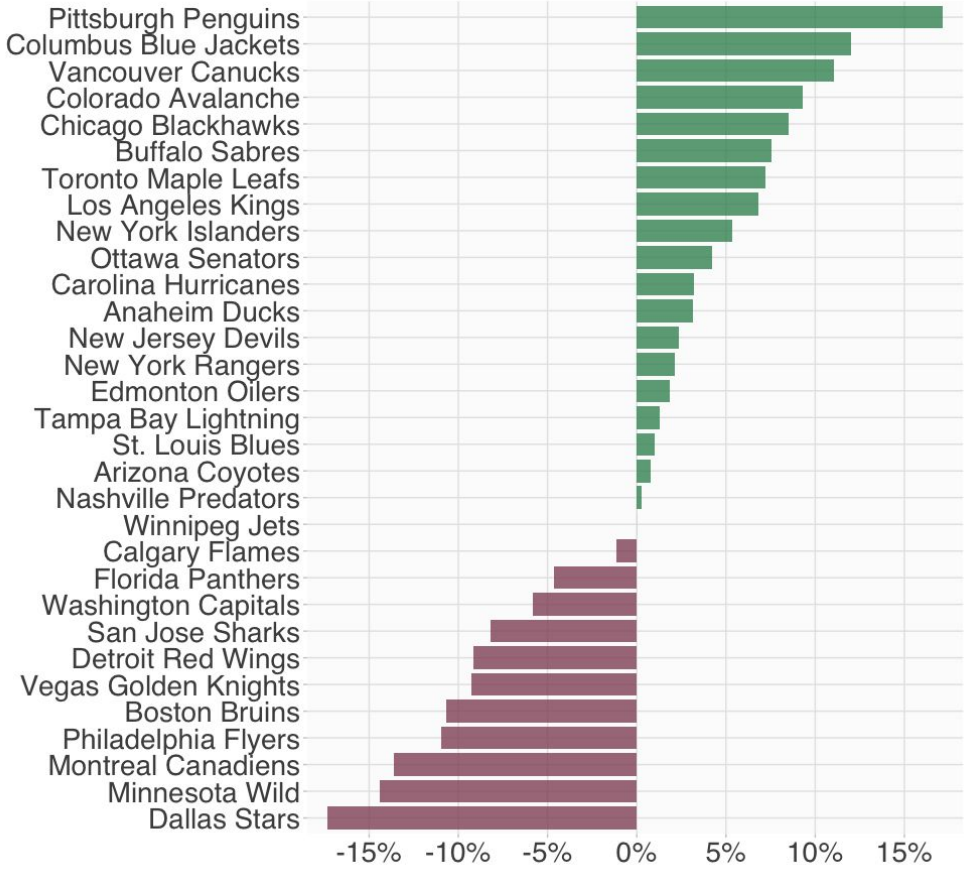
## % of Picks Spent on Overagers > Expected

*2014-17 NHL Drafts*

Second things second:

even after adjusting for the lateness of their picks, the Penguins picked lots of overagers.

But is this meaningful over only 4 draft years and 21 picks?

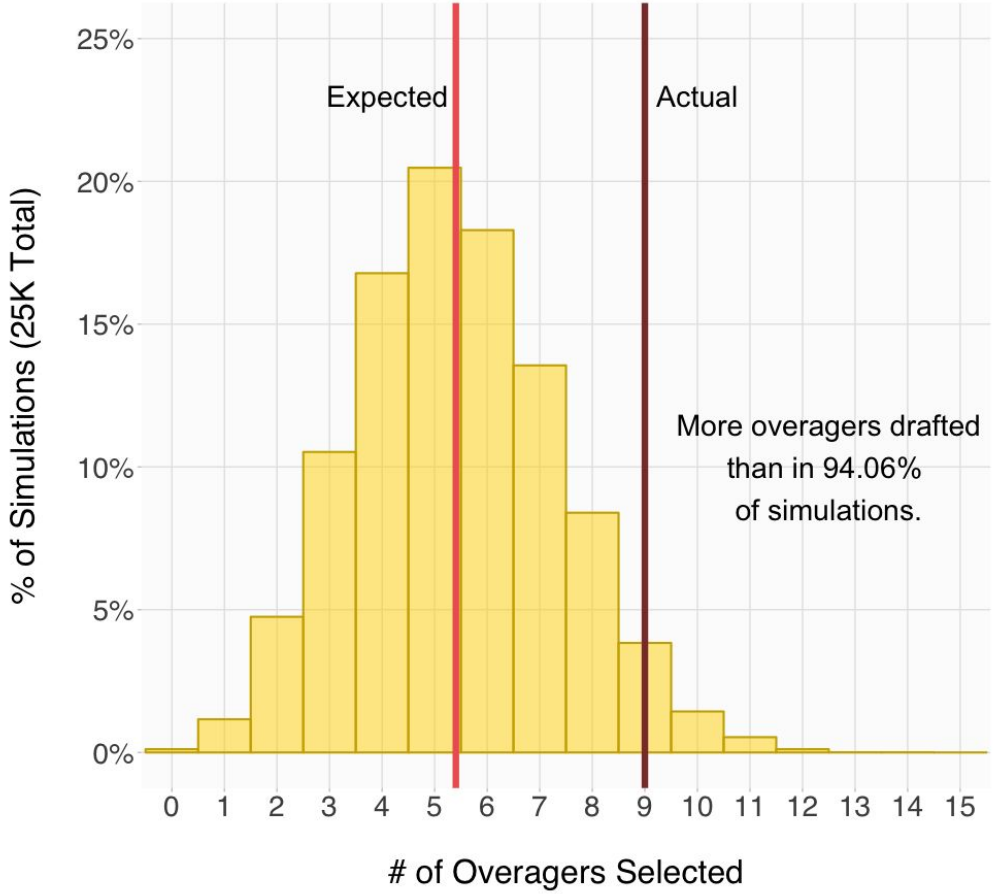


—

# Simulated Distribution of Overager Selections

***2014-17 Pittsburgh Penguins Picks***

Fortunately, a key benefit of this model is being well-suited to simulation.



—

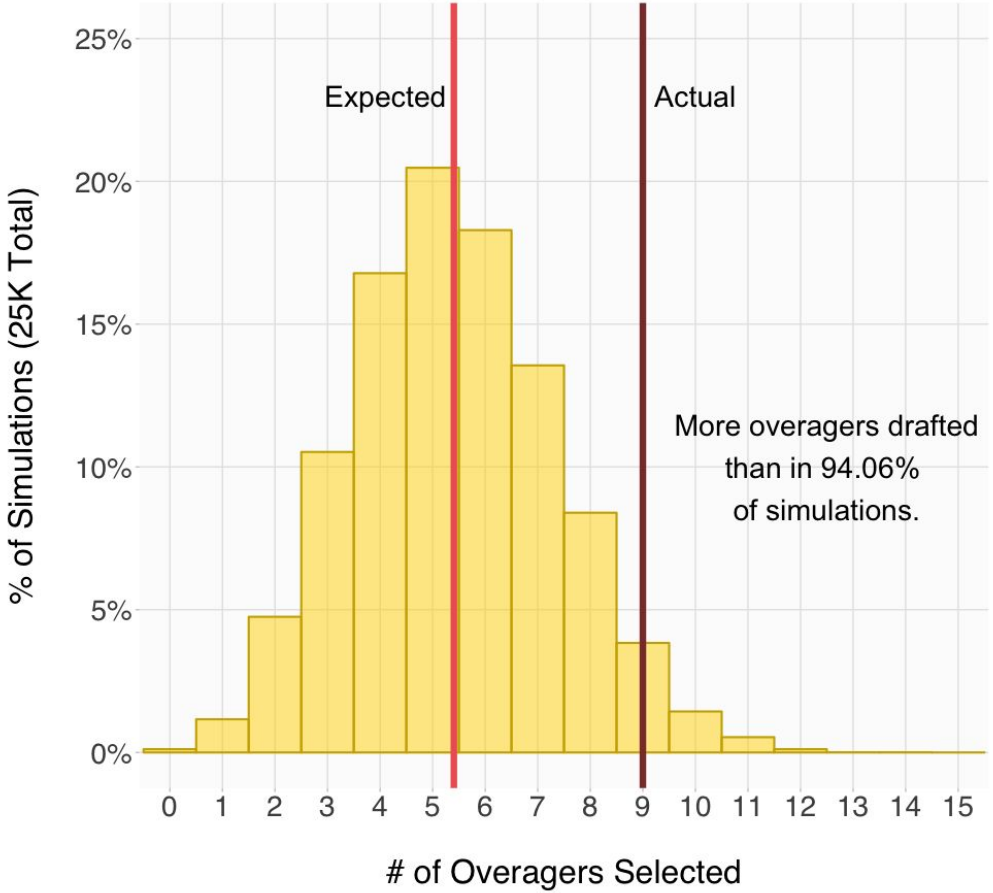
## Simulated Distribution of Overager Selections

***2014-17 Pittsburgh Penguins Picks***

Fortunately, a key benefit of this model is being well-suited to simulation.

Is this a small sample? *Yes.*

Is this still a compelling trend? *Also yes.*



> My message to our staff is, *‘You've got to be open minded.’* I don't think there's a set age with a player. A guy like Dominik Simon, he went through the draft three years.

~ Patrik Allvin,
Penguins Director of Amateur Scouting
(link)

So, we basically know they’re doing this on purpose.
But why?

# Overagers: A Market Inefficiency?

- My hunch was that overagers are *more likely* to play NHL games for your team, but *less likely* to be impactful players.
- To test this, I built gradient-boosted tree models using the xgboost package in R to predict

  1) total NHL games played 7 seasons after the draft
  2) career NHL Point Shares per game for players with 10+ career games played

- *Variables*: age, pick #, draft round, height, weight, nationality.
  *Draft years*: 2005-11.
  *Hyperparameters*: chosen to minimize 5-fold CV testing error.

*Data used is current as of 10/12/2018.*

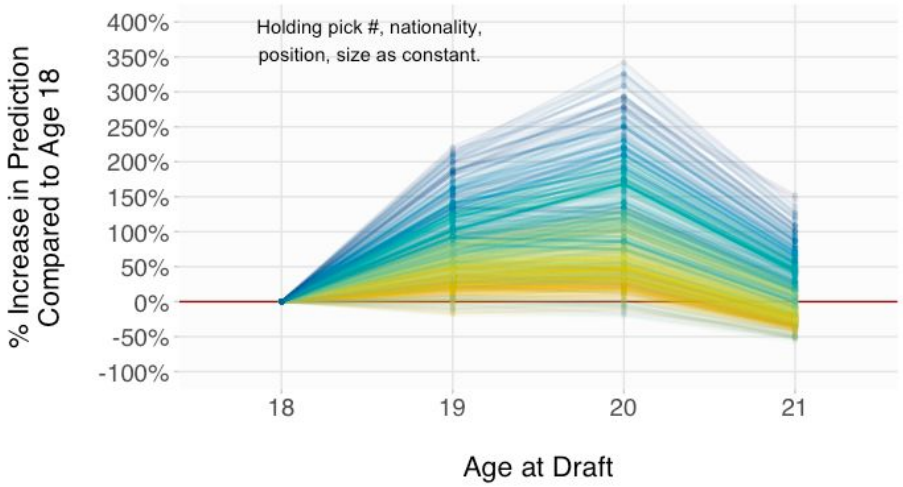
# We can interpret ~ML~ models by comparing predictions across different values of a variable.

*They call these "individual conditional expectation plots."*

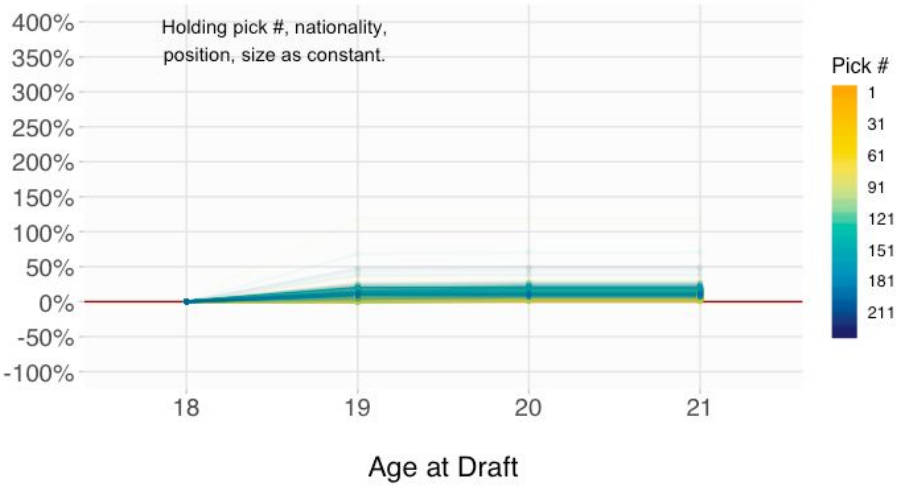
- *Step 1:* Generate 4 predictions for each prospect, assuming they're 18, 19, 20, or 21 years old.
- *Step 2:* Compare the predictions across ages for each prospect.
- *Step 3:* See how the age effects vary between models.
- *Step 4:* Feel overwhelmed by how many follow-up questions you have for yourself and then go to bed.

# Individual Conditional Expectation Plots

## Games Played Model




## Value Per Game Model



# Another look:

| Draft Round | Age 18 → 19 Median Predicted Increase in GP | Age 18 → 19 Median Predicted Increase in Value |
|---|---|---|
| 1 | 23.3% | 4.1% |
| 2 | 39.1% | 7.2% |
| 3 | 59.7% | 9.5% |
| 4 | 69.9% | 11.2% |
| 5 | 91.5% | 10.7% |
| 6 | 104.6% | 10.8% |
| 7 | 139.9% | 11.1% |

# I was right! (kind of).

- Lots of evidence that overage prospects play more games on average than their 18-year-old counterparts (after holding pick #, size, and nationality constant) within 7 years of the draft.
- Some evidence that they may be *slightly* more valuable in terms of impact on a per-game basis, although this is where we miss having coefficients with confidence intervals.
- Lots of evidence that these effects are highly dependent on where you are in the draft. Being 19+ doesn't really move the needle much re: value in rounds 1-3, but it does in rounds 4-7.

# Overagers: A Needed Investment?

*Overage prospects are ~50% more likely to enter the NHL at any time than **comparable** 18-year-old prospects.*

- The Penguins are experiencing what some people like to call a “Cup window.”
- In order to make the best of it, they need very cheap depth to fill out a roster and plug in next to Sidney Crosby and Jack Johnson.
- Overagers should theoretically enter the league sooner than 18-year-olds, as they saw with 21-year-old Dominik Simon.
- I was able to confirm this with my survival analysis of NHL prospect timelines.

# What I Tell An NHL GM

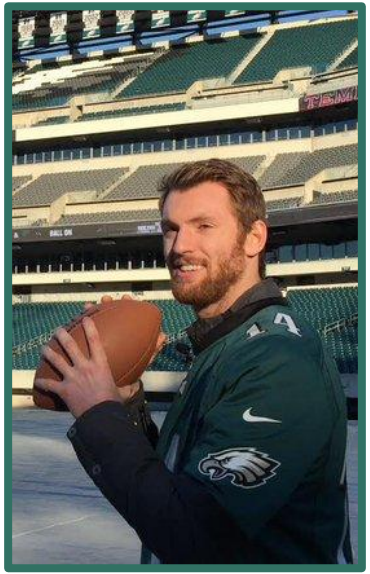
The Penguins have drafted a lot of overagers recently, and I would guess that it's likely they continue to do so.

Drafting overagers is not an inherently amazing or terrible strategy, and it depends on cap situation and "Cup windows."

But it might be worth a look in later draft rounds, when you don't expect to draft a particularly impactful player anyway.

It's also worth doing research to target players with projected development timelines that complement your personnel needs.

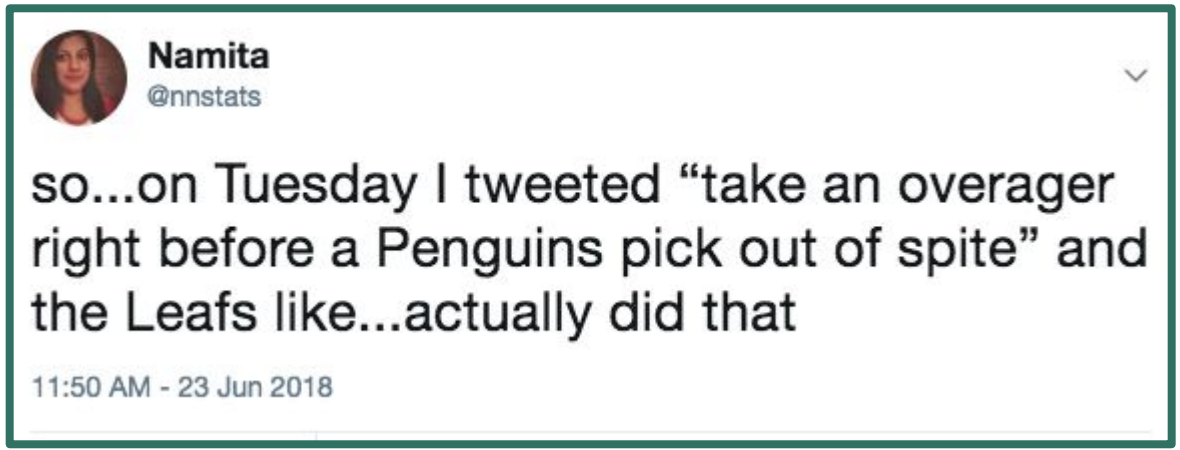
# Thank you for listening!



- Questions/comments/concerns can be directed to me now or emailed to namitanandakumar@gmail.com later.
- In the spirit of reproducible research, I will tweet out all of the R code used to generate these analyses.
- (I'm @nnstats on Twitter.)
- All data courtesy of NHL.com and Hockey Reference.
- #GoBirds 🏈 📊

# Various Appendices

# My Impact™

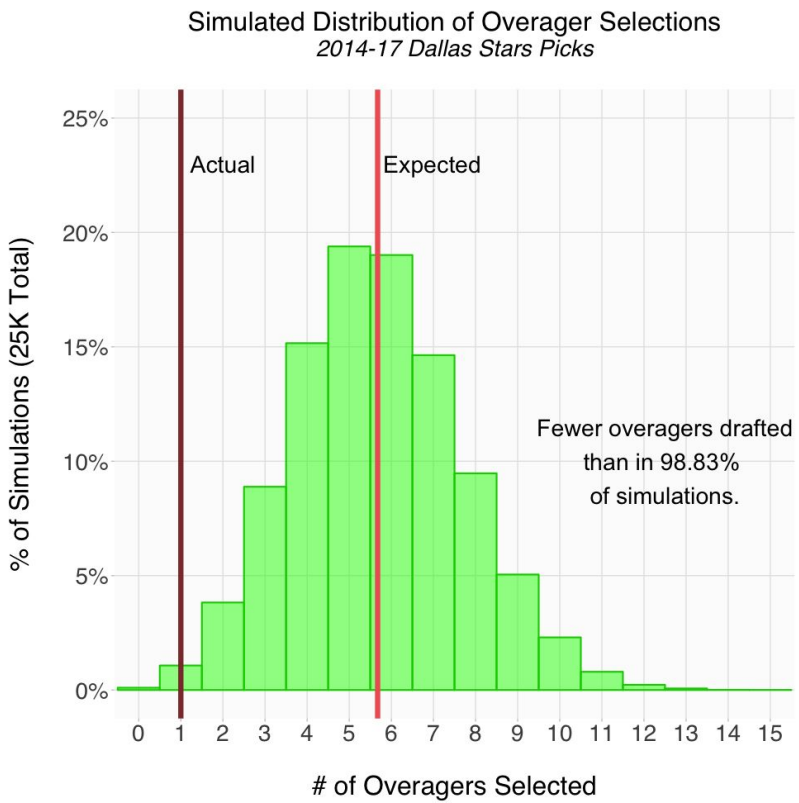

Namita
@nnstats
so...on Tuesday I tweeted “take an overager right before a Penguins pick out of spite” and the Leafs like...actually did that
11:50 AM - 23 Jun 2018

# Team Model Notes

- The rstanarm default weakly informative priors are:

  N(0, 10) for the intercept
  N(0, 0.041) for overall pick #
  N(0, 0.730) for (overall pick #)$^2$

- Simulations assume independence between a team's picks. It's theoretically possible to pick, like, 7 overagers in a row, but teams probably wouldn't go for that, so I think the "real" distributions would be slightly tighter.




# *Why did I call out the Penguins and not the Stars?*

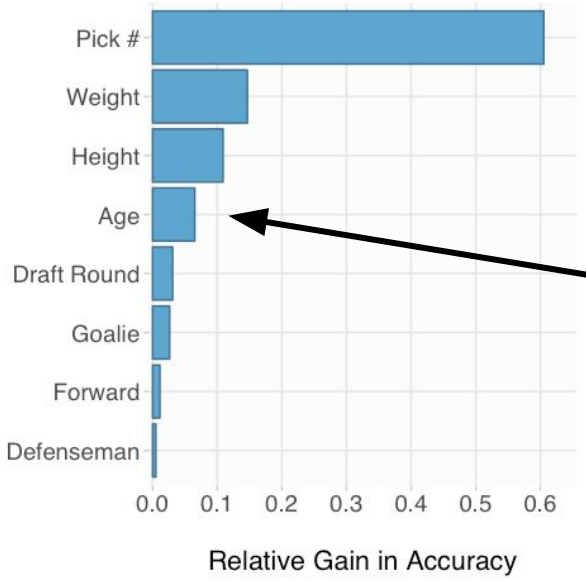
- Prioritizing first year eligible players is basically the default position to take.
- Meanwhile, the Penguins' perceived tendency gives us a better handle on what they might do moving forward.
- And it has pretty compelling potential reasons behind it.
- Also, I don't like the Penguins.

# Player Model Notes

- Overagers playing more NHL games in the first 7 years of their career could potentially be explained away by the fact of them entering the league sooner.
- It'd be interesting to see if overagers end up playing more games on average over the course of their whole careers, or if it eventually ends up being a wash.
- The 10+ games played cutoff for the per-game value model was chosen as the minimum cutoff that seemed to get rid of obvious small sample issues, but this can be fiddled with.

# Feature Importance

## Games Played Model




## Value Per Game Model