Namita Nandakumar
@nnstats

Does (NHL Player) Size Matter?

Some Questions

Does height and/or weight matter for...

- being drafted?
- playing in the NHL?
- playing *well* in the NHL?
- excelling in the playoffs?

Some Reminders

It's possible that size affects :

- player quality

which affects :

- player opportunity
- performance conditional on opportunity

- scouting + team perception of player quality

which affects :

- player opportunity

Some Questions Written in Math

Does height and/or weight matter for...

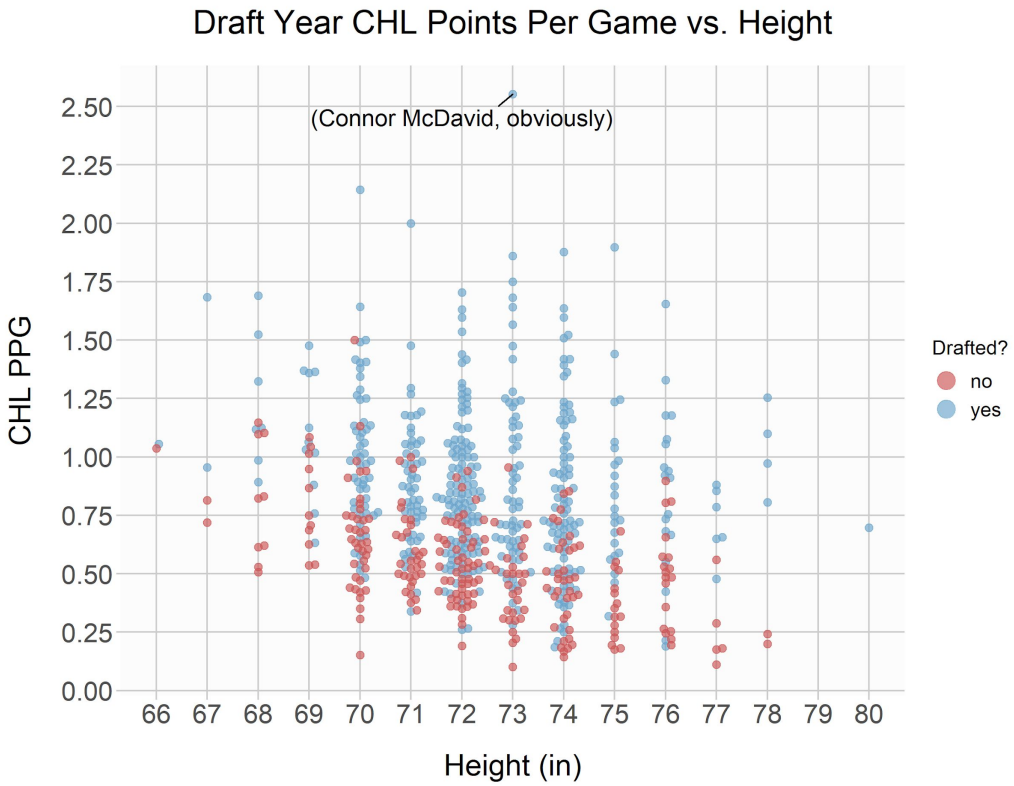
- P(drafted | ranked)
- E(NHL games played | drafted)
- E(NHL value | NHL)
- P(good in playoffs | NHL)

1. P(drafted | ranked)

NORTH AMERICAN SKATERS

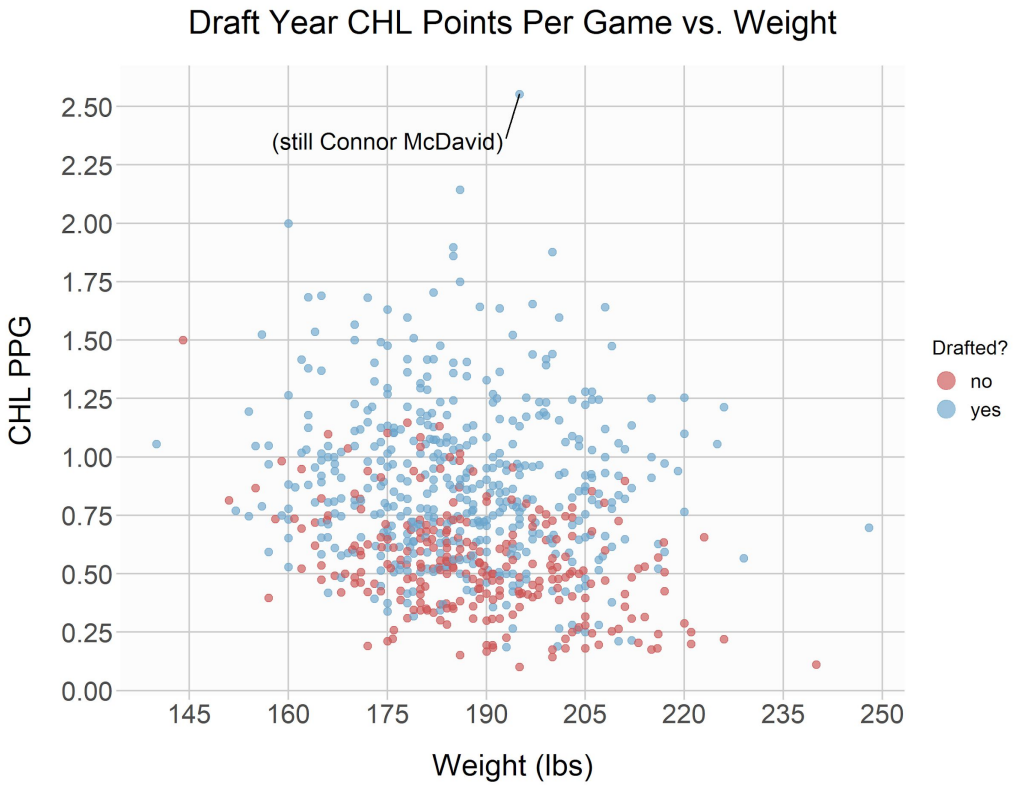
Final Rank↑	Midterm Rank	Player	Height	Weight
1	1	MCDAVID, CONNOR	6' 1"	195
2	2	EICHEL, JACK	6' 2"	196
3	3	HANIFIN, NOAH	6' 3"	203
4	5	STROME, DYLAN	6' 3"	185
5	4	CROUSE, LAWSON	6' 4"	215
6	7	MARNER, MITCHELL	5' 11"	160
7	10	PROVOROV, IVAN	6' 0"	201
8	8	ZACHA, PAVEL	6' 3"	210

- 2010-19 draft classes
- ranked by NHL Central Scouting
- first year eligible Canadian Hockey League forwards
- 10+ games played in draft year



Let's compare apples to apples.

- 2010-19 draft classes
- ranked by NHL Central Scouting
- first year eligible Canadian Hockey League forwards
- 10+ games played in draft year



Let's compare apples to apples.

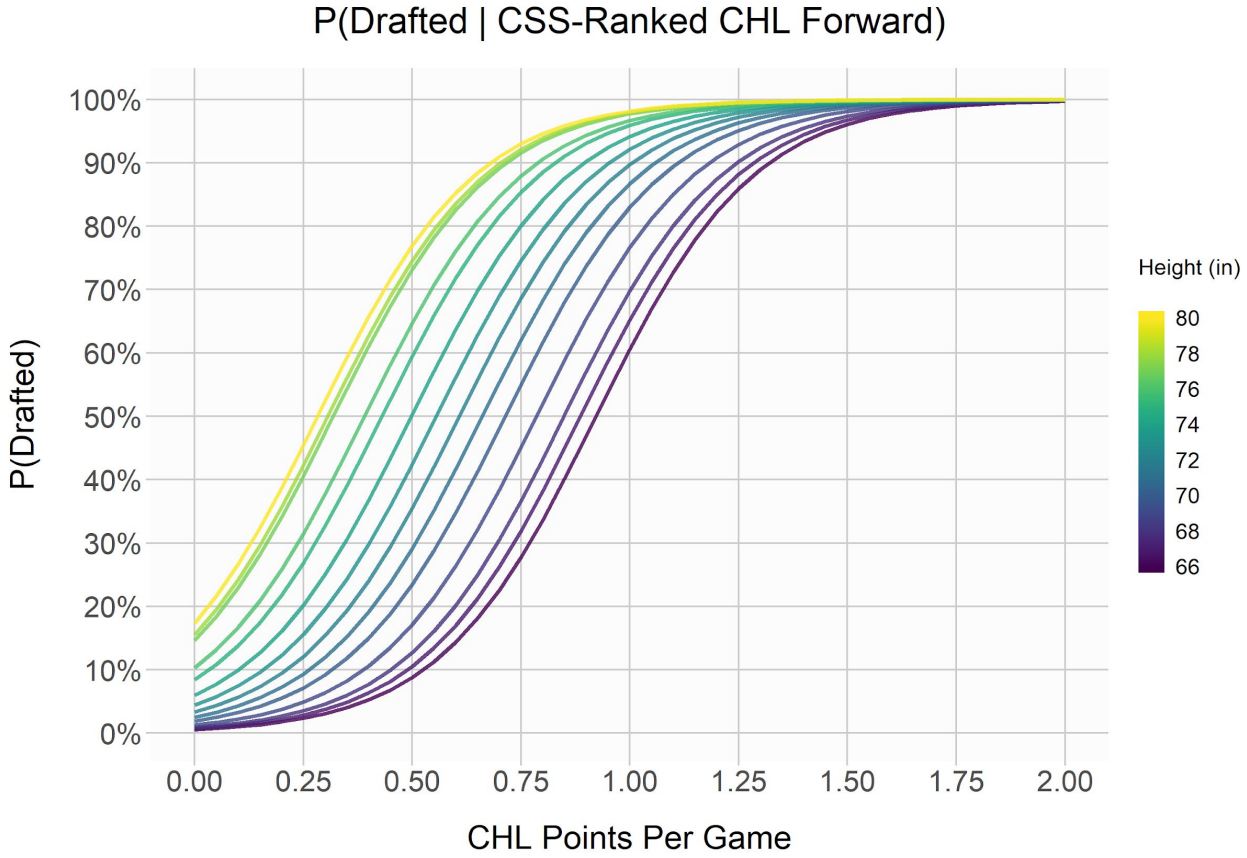
It’s time for logistic regression.

What does this output mean?

- I'm literally asking.
- We can see that height and weight are "*significant*" predictors of being drafted but there's much more to understand here!

```
Call:
glm(formula = drafted ~ height + weight + pts_gp, family = "binomial",
    data = css_chl_info_clean)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1871  -0.8235   0.2803   0.7593   2.2044

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -28.469102   4.180632  -6.810 9.78e-12 ***
height        0.408516   0.065773   6.211 5.26e-10 ***
weight       -0.023294   0.008293  -2.809  0.00497 **
pts_gp        5.533241   0.471307  11.740  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 916.71  on 695  degrees of freedom
Residual deviance: 662.12  on 692  degrees of freedom
AIC: 670.12

Number of Fisher Scoring iterations: 5
```

- Let's create a toy dataset of every height + the median weight for that height + the model prediction and then graph it:



- And then let's find the PPG rate that gives you ~50/50 odds of being drafted based on your height + median weight:

Height	5'6	5'7	5'8	5'9	5'10	5'11	6'0	6'1	6'2	6'3	6'4	6'5	6'6
PPG	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.50	0.45	0.40	0.35	0.35

- And finally let's check out the model calibration:

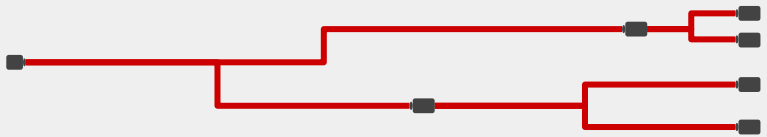
Probability Bin	0-10 %	10-20 %	20-30 %	30-40 %	40-50 %	50-60 %	60-70 %	70-80 %	80-90 %	90-100 %
% Drafted	17%	19%	24%	32%	46%	54%	68%	74%	84%	97%
# of Guys	6	43	68	65	63	56	65	78	96	156

2. **E(NHL GP | drafted)**
3. **E(value | NHL)**

Working with subsets of draft data is interesting but limiting...

How do we build models for 18-year-old Canadian wingers + 19-year-old American goalies + 20-year-old Swedish defensemen?

Gradient-boosted trees can help.



- Tree models allow for all sorts of variable interactions.
- It's a key benefit of *~machine learning~*.

We can use xgboost to predict:

- total NHL games played within 7 seasons of being drafted
- career NHL Point Shares per game for players with 10+ career NHL games played

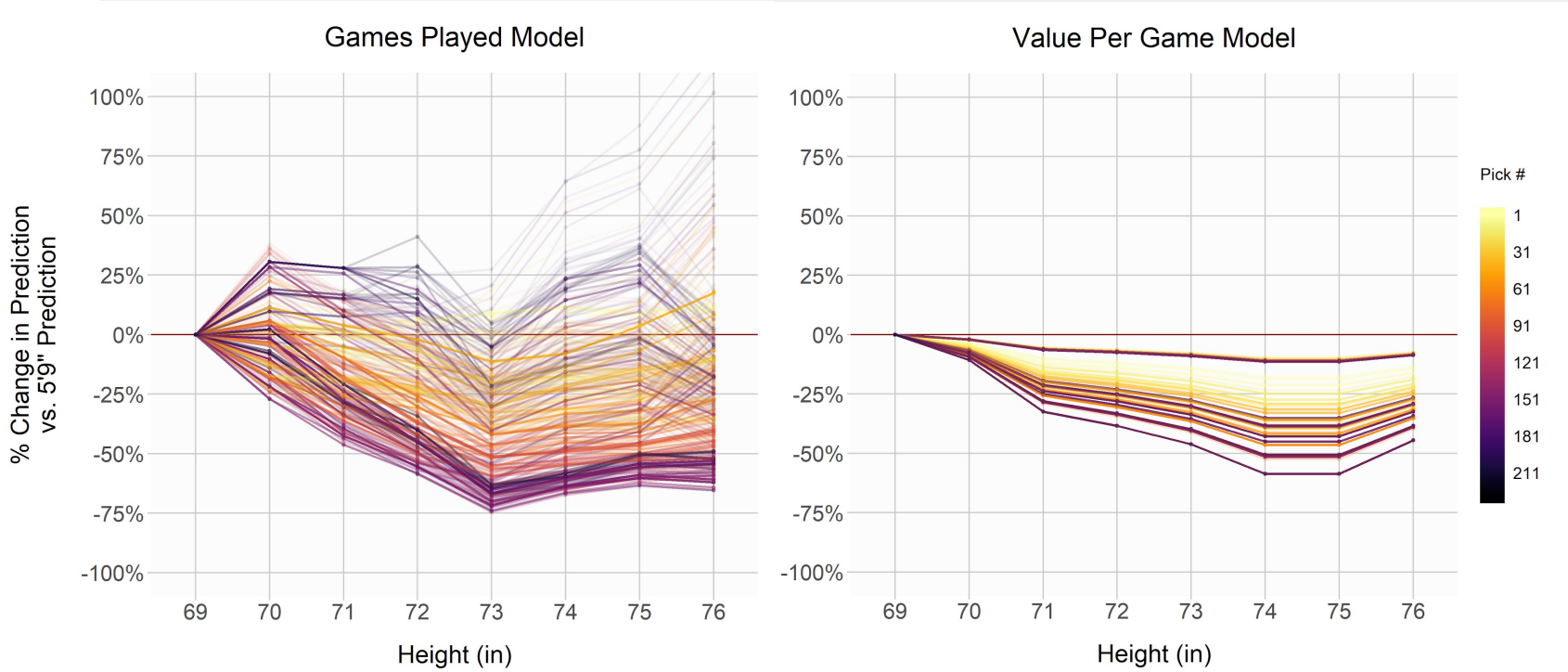
This doesn't have to be an entirely black box!

model details :

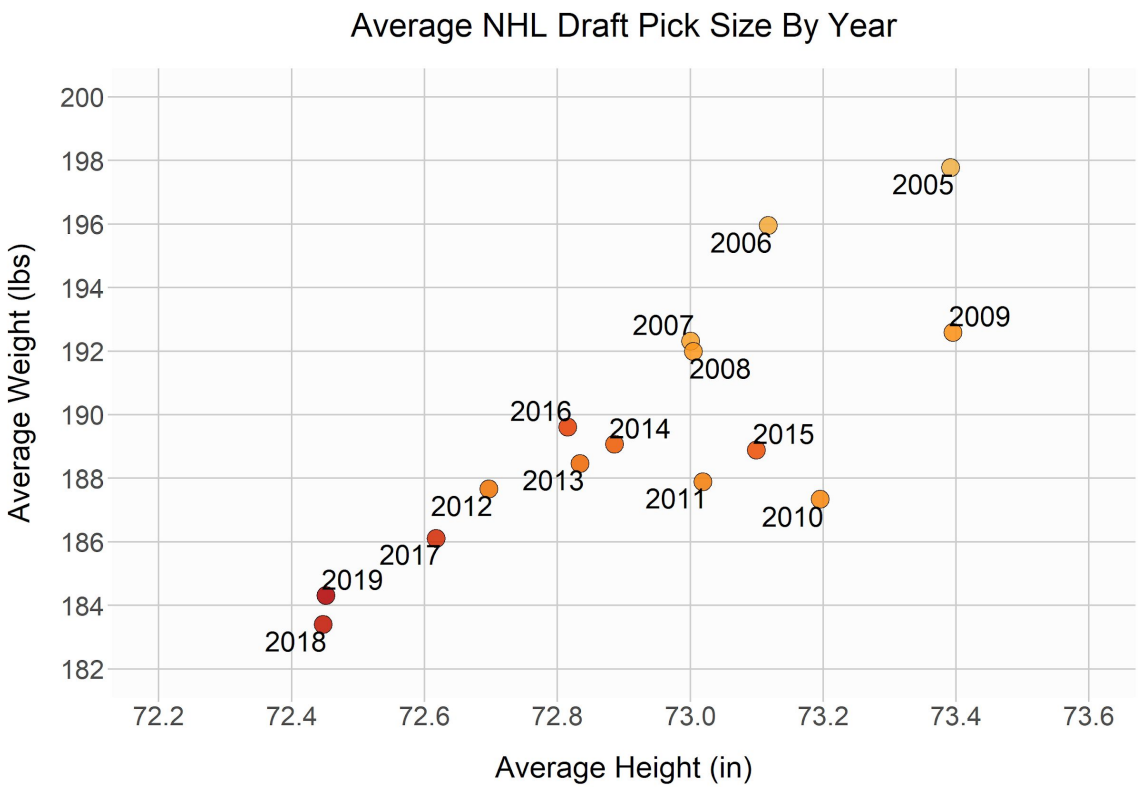
- trained on age, height, weight, position, draft pick #, Euro/NA in 2005-11 drafts
- hyperparameters chosen to minimize 5-fold CV testing error

We can evaluate our predictions by:

- Making 8 copies of each prospect.
- Pretending they have a different height (5'9" - 6'4") and weight (170 - 210 lbs) each time.
- Generating xgboost predictions.
- Comparing each model's predictions across hypothetical prospect sizes.



Individual Conditional Expectation Plots



Teams may already be reacting...

4. P(good in playoffs | NHL)

[REDACTED]

One of the reasons NHL scouts still hesitate to take players under 5-10 high in the draft is the concern that come playoff time it gets harder for them to through the checking. Gaudreau and Arvidsson combined, for example, have 9 goals in 74 playoff games. They scored 70 this yr.

6:54 PM · Apr 20, 2019 · Twitter Web Client

Namita
@nnstats

fact check: there's 8 fwds < 5'10" who've played 50+ min in the playoffs this season.

Gaudreau, Arvidsson, Marchand, T. Johnson have fewer goals/60 in the playoffs.
Atkinson, Gourde, Marchessault, Zuccarello have more goals/60 in the playoffs.

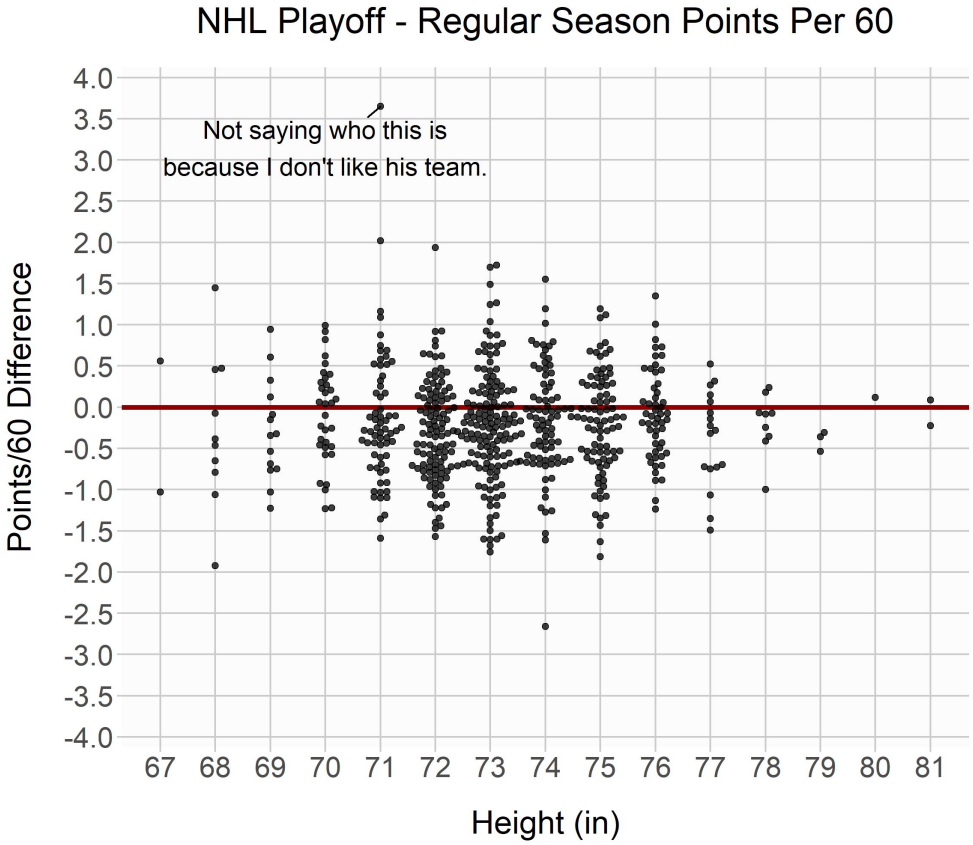
you can cherry pick any narrative.

...I could've done better.

I stand by my rebuttal philosophically, *but* it was far from comprehensive.

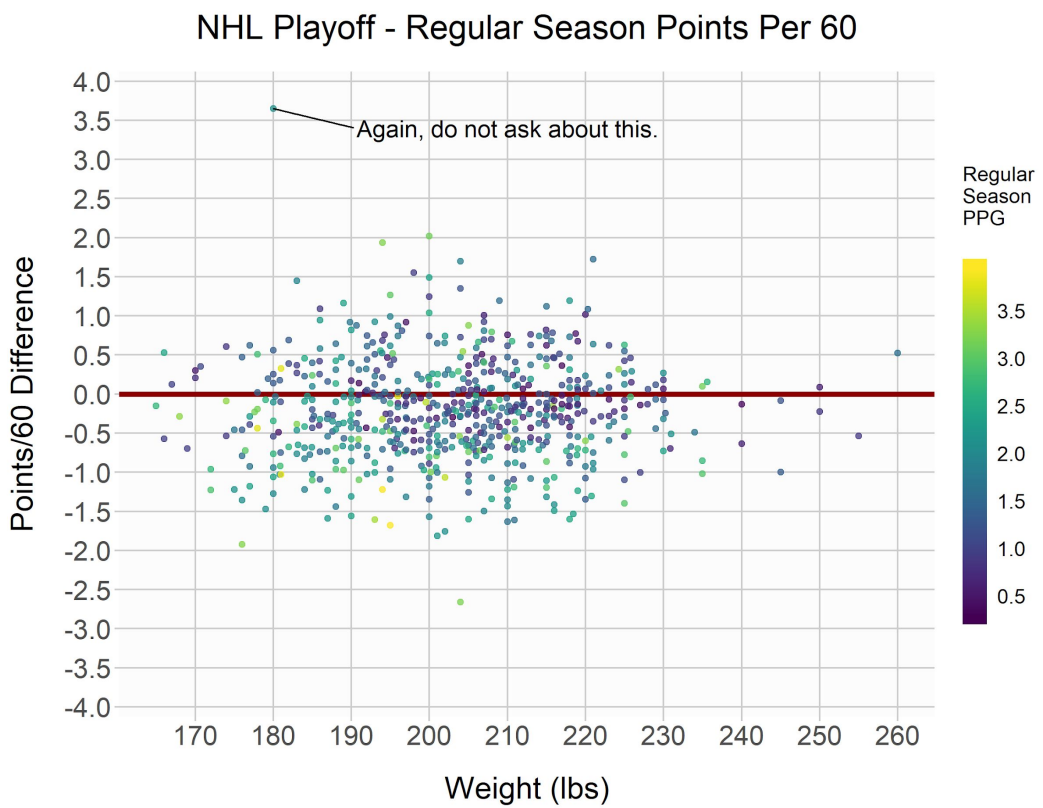
- I didn't check the baseline difference between regular season and playoff performance.
- I used a random height cutoff and only looked at last season's playoffs.
- It was straight up 8 dudes.

- 2015-19 seasons
- players with 41+ regular season games and 10+ playoff games
- n o c o r r e l a t i o n
- try to regress that, I dare you



A closer, more comprehensive look...

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.2577357  1.0259216  -0.251    0.802
height      -0.0019284  0.0187198  -0.103    0.918
weight       0.0009384  0.0026534   0.354    0.724

Residual standard error: 0.6577 on 626 degrees of freedom
Multiple R-squared:  0.0003,    Adjusted R-squared:  -0.002894
F-statistic: 0.09394 on 2 and 626 DF,  p-value: 0.9104
```



Even closer...

So what was the point?

Nothing.

- Size matters for NHLers sometimes, in part because powerful people think so.
- Null results might seem boring but you can still use them to fight people on the internet.
- A data scientist's job is to answer every question with the phrase "It depends."

Thank you!

Data via Hockey Reference + NHL.com (yeah, seriously).

*Tweeting this out **@nnstats**.*

Appendix :

- As one might expect, heights and weights are correlated.