


# Pucks, Penalties, and...Percentages?

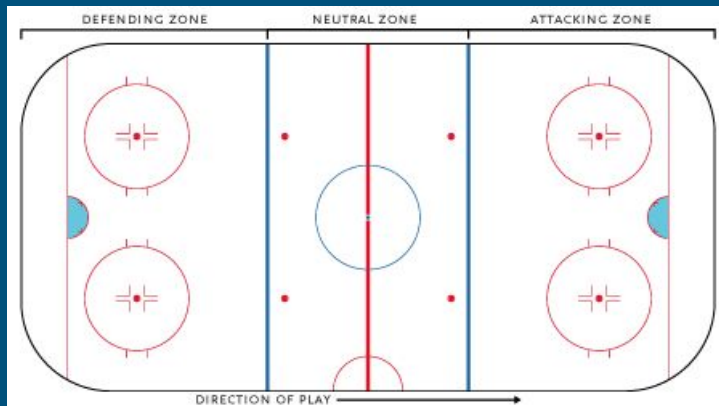


Mark Brennan and Kristina Barounis



# Setting up

- Question: Which factors contribute to an NHL team's success in the regular season?
- Our expectation: Many aspects of the game would need to be included in our final model given the structure and dynamics of the game



The traditional dump-and-chase strategy may not be as effective as it was once thought to be



# Our process



Data gathering



Data cleaning and organization

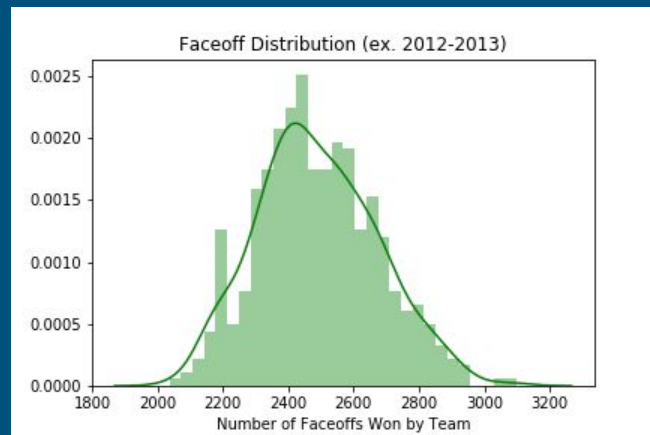
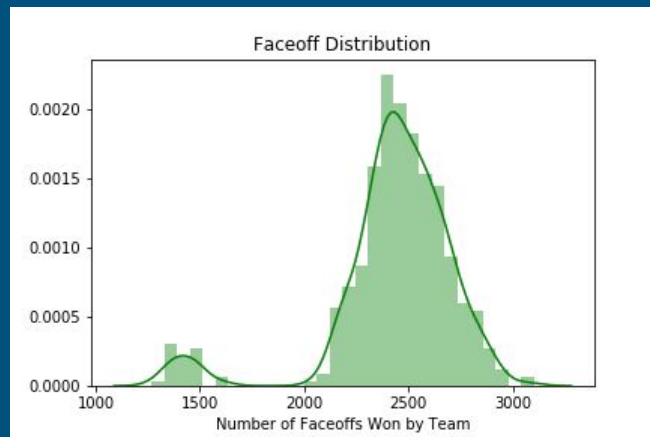


Finalizing an  
OLS model

Data exploration

# The data

- Data sources: NHL API & web scraping NHL website
- 20 seasons and 31 teams
- But we had to make some cuts
  - 2004-2005
  - 2012-2013
- 516 individual observations

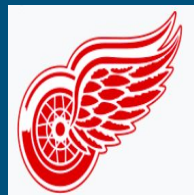
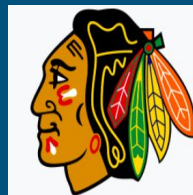


# Independent variables

---

- Categorical

- “Original 6”



- Interaction

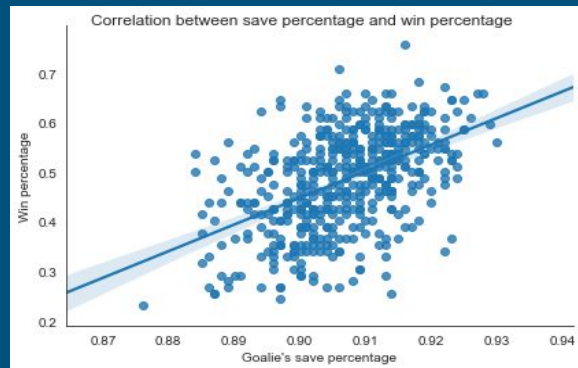
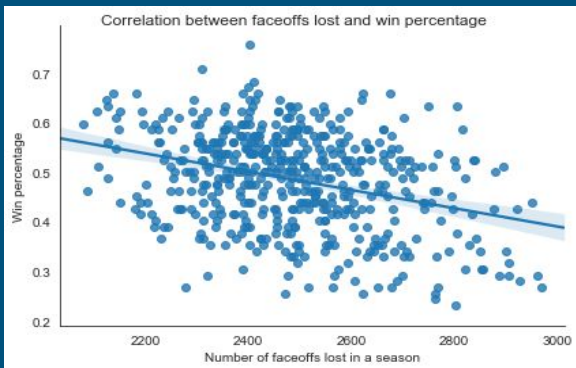
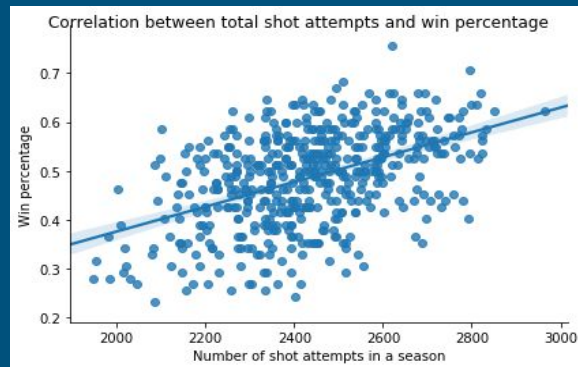
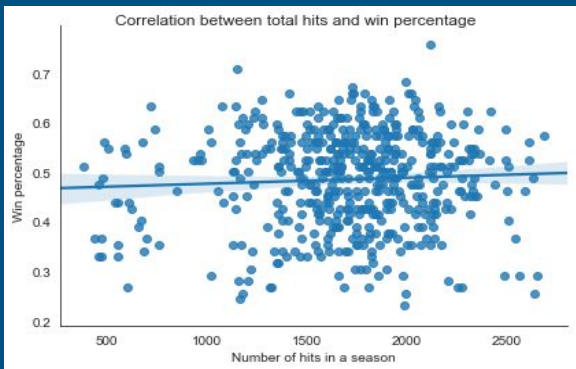
- Puck possession - shots attempted, a.k.a. “Corsi”
  - Defensive strength - hits, takeaways, penalty kill percentage, shots allowed

---

Every fan loves a  
good hockey brawl  
but...



# Taking a look at correlations





# Final models

<b>Dep. Variable:</b>	win_percentage	<b>R-squared:</b>	0.400
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.395
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	85.06
<b>Date:</b>	Fri, 23 Aug 2019	<b>Prob (F-statistic):</b>	2.43e-55
<b>Time:</b>	10:56:47	<b>Log-Likelihood:</b>	606.94
<b>No. Observations:</b>	516	<b>AIC:</b>	-1204.
<b>Df Residuals:</b>	511	<b>BIC:</b>	-1183.
<b>Df Model:</b>	4		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	0.4855	0.003	147.049	0.000	0.479	0.492
faceOffsLost	-0.0202	0.003	-5.786	0.000	-0.027	-0.013
powerPlayOpportunities	0.0224	0.004	6.036	0.000	0.015	0.030
savePctg	0.0518	0.004	14.664	0.000	0.045	0.059
puck_possession	0.0272	0.004	7.521	0.000	0.020	0.034

<b>Dep. Variable:</b>	win_percentage	<b>R-squared:</b>	0.870
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.869
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	568.2
<b>Date:</b>	Fri, 23 Aug 2019	<b>Prob (F-statistic):</b>	6.51e-222
<b>Time:</b>	10:41:03	<b>Log-Likelihood:</b>	1001.8
<b>No. Observations:</b>	516	<b>AIC:</b>	-1990.
<b>Df Residuals:</b>	509	<b>BIC:</b>	-1960.
<b>Df Model:</b>	6		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	0.4855	0.002	315.485	0.000	0.482	0.489
faceOffsLost	-0.0148	0.002	-8.729	0.000	-0.018	-0.011
goalsAgainstPerGame	-0.0470	0.002	-19.500	0.000	-0.052	-0.042
goalsPerGame	0.0601	0.002	36.802	0.000	0.057	0.063
giveaways	-0.0074	0.002	-4.179	0.000	-0.011	-0.004
savePctg	0.0091	0.002	3.805	0.000	0.004	0.014
puck_possession	0.0132	0.002	6.883	0.000	0.009	0.017

# Getting better for next game

---

- NHL's partnership with SAP introduced a new data platform in 2015
- Expand our data selection
  - Shot types
  - Individual player statistics