Comparing NHL Players' Shots and Goals by Algorithmically Decomposing Shot Intensity Surfaces

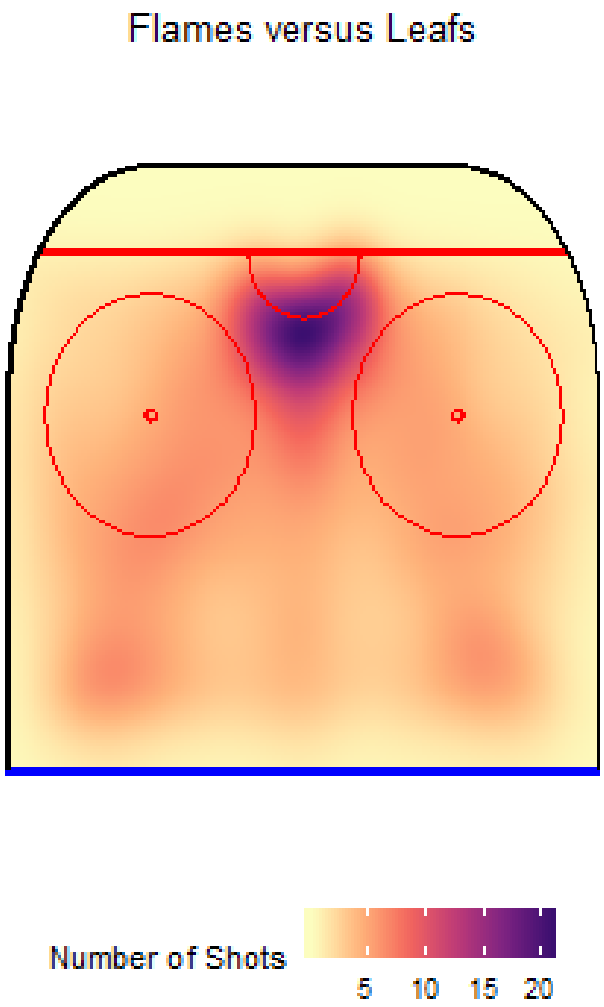
Devan Becker, Ph.D. Candidate
Department of Statistical and Actuarial Sciences
The University of Western Ontario, Canada

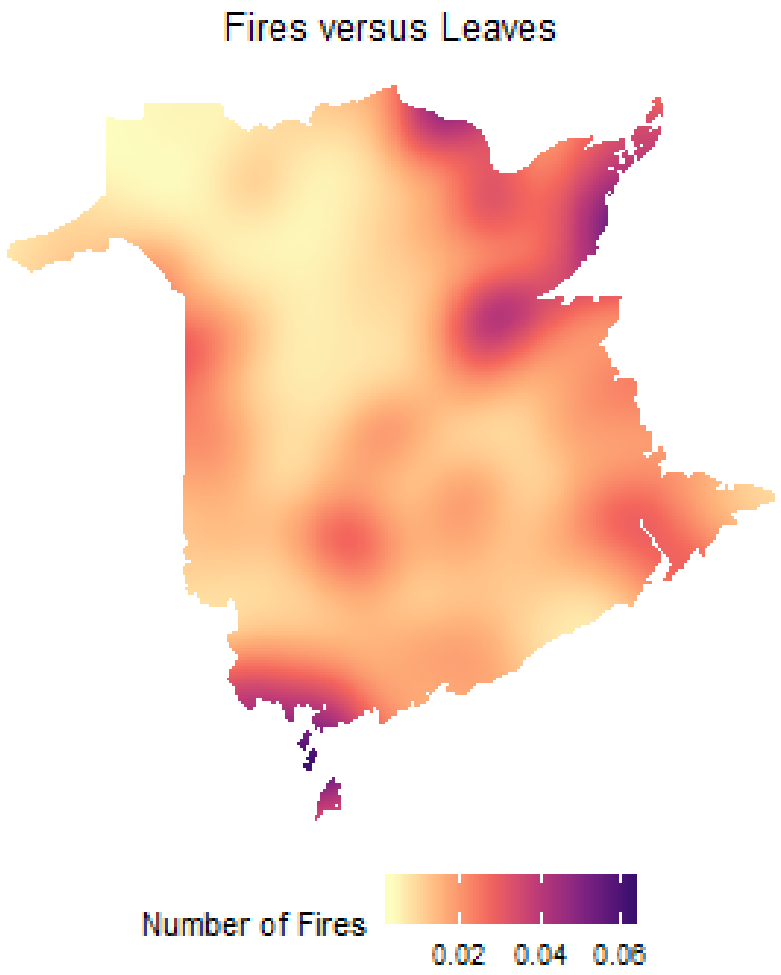
Collaborators: Charmaine Dean and Douglas Woolford

New England Symposium on Statistics in Sports
Harvard University
September 28, 2019

Introduction

A Song of Ice Hockey and Forest Fire





Outline

- Describe the NMF-LGCP algorithm in general.
- Apply it to NHL shot data.
- Use these estimates to describe shot efficiency (with a twist).

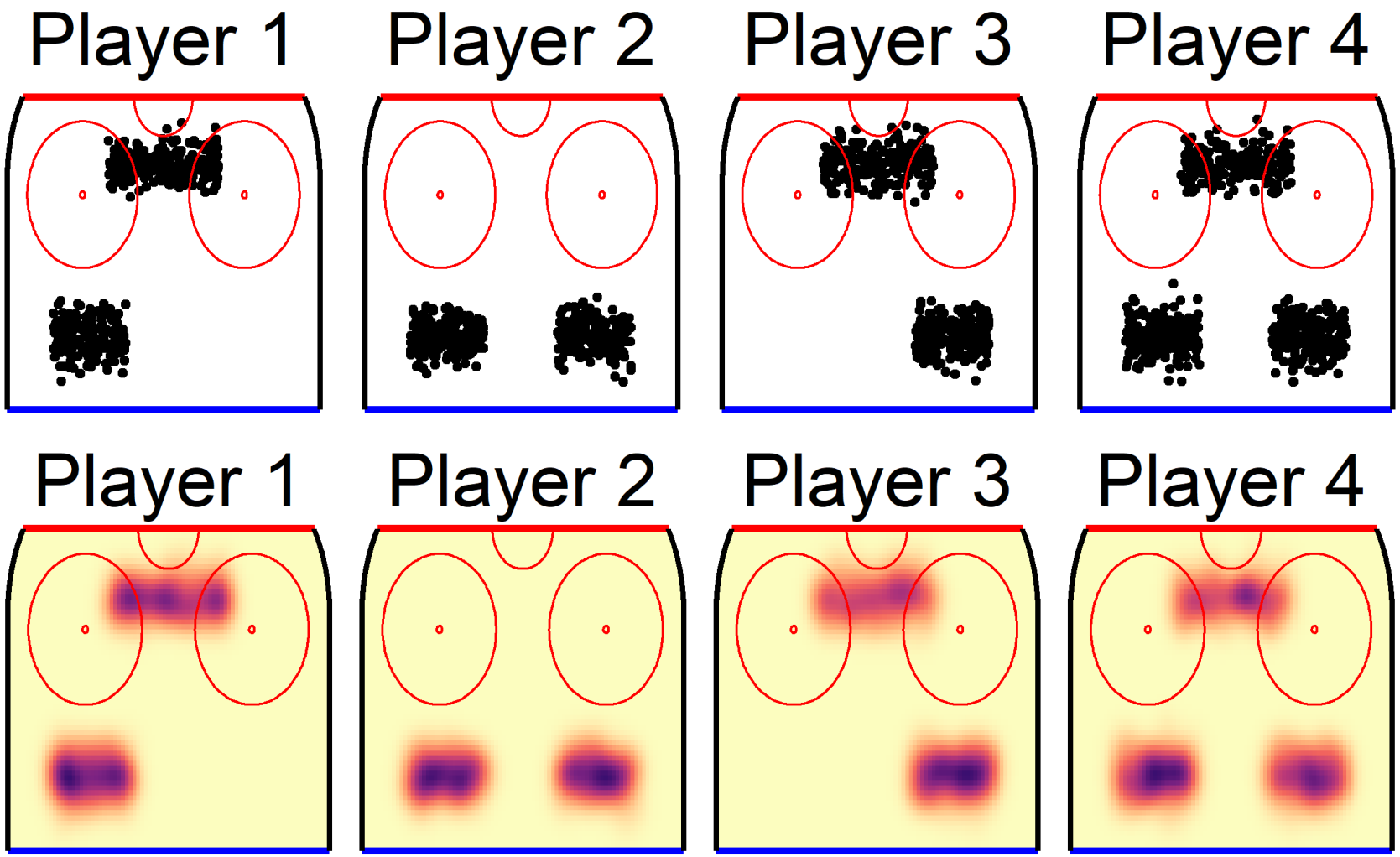
Image Recognition and Spatial Estimates

Idea: Convert one big matrix into two smaller matrices

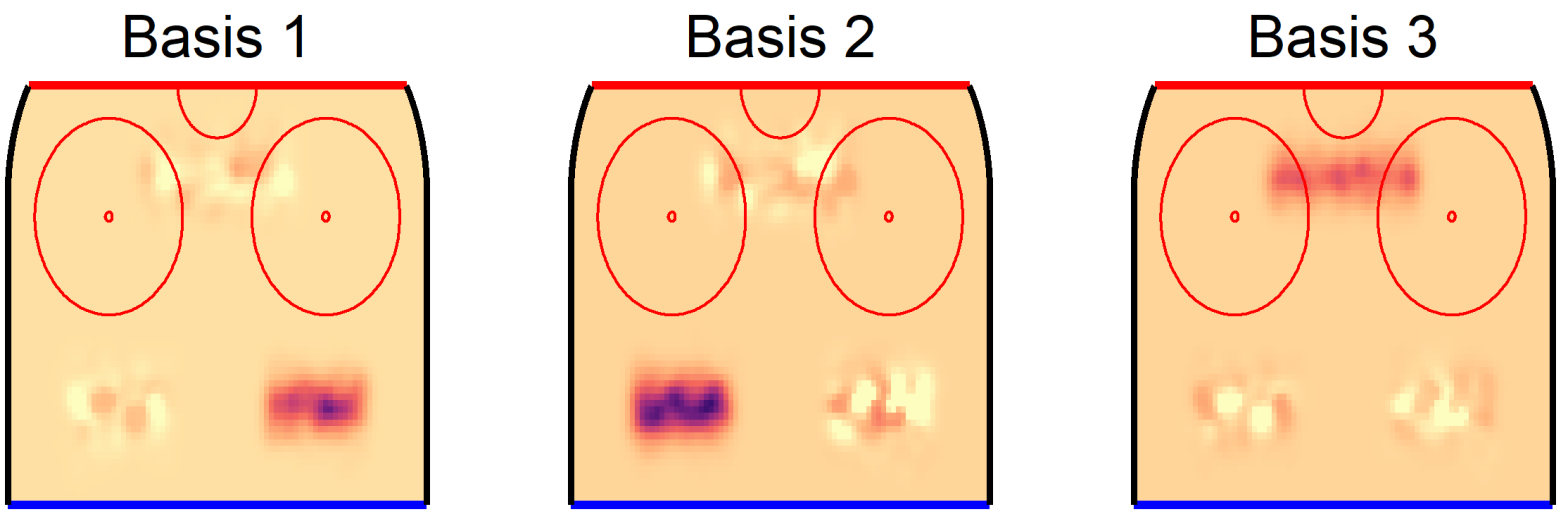
$$X \approx WH$$

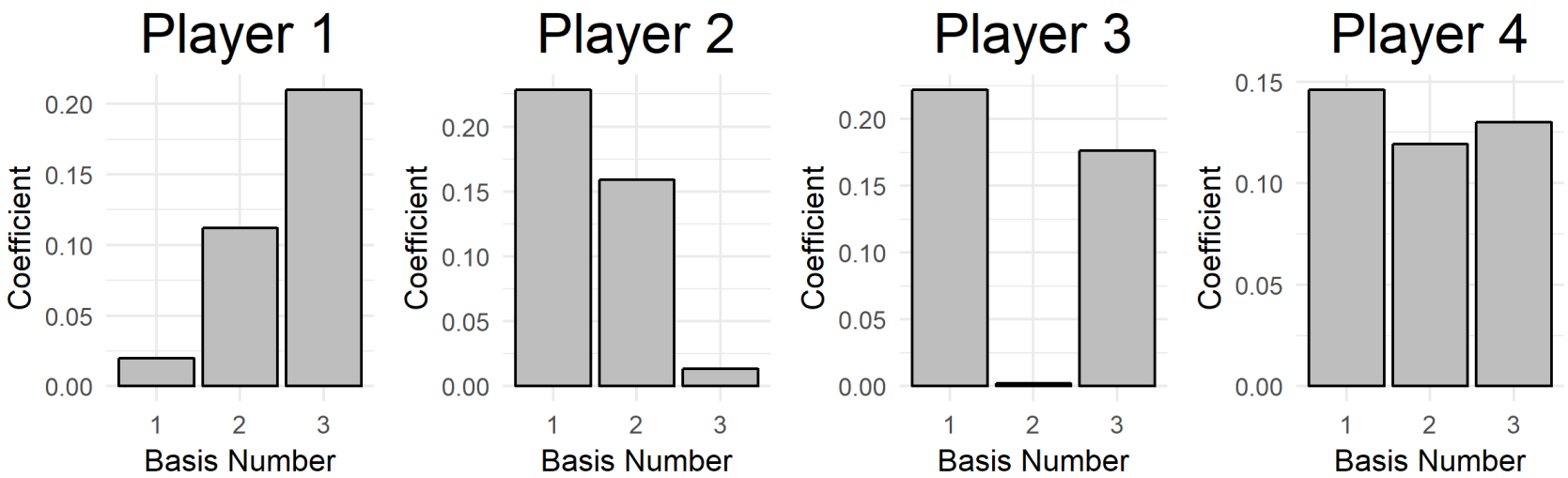
- X is a giant collection of images
  - In this case, LGCP estimates of the shot locations.
- W defines the "bases"
  - Building blocks for all images
- H defines the "coefficients"
  - How much each image uses each building block

Dummy Data and Estimates



Dummy Bases and Coefficients





Choosing the number of bases

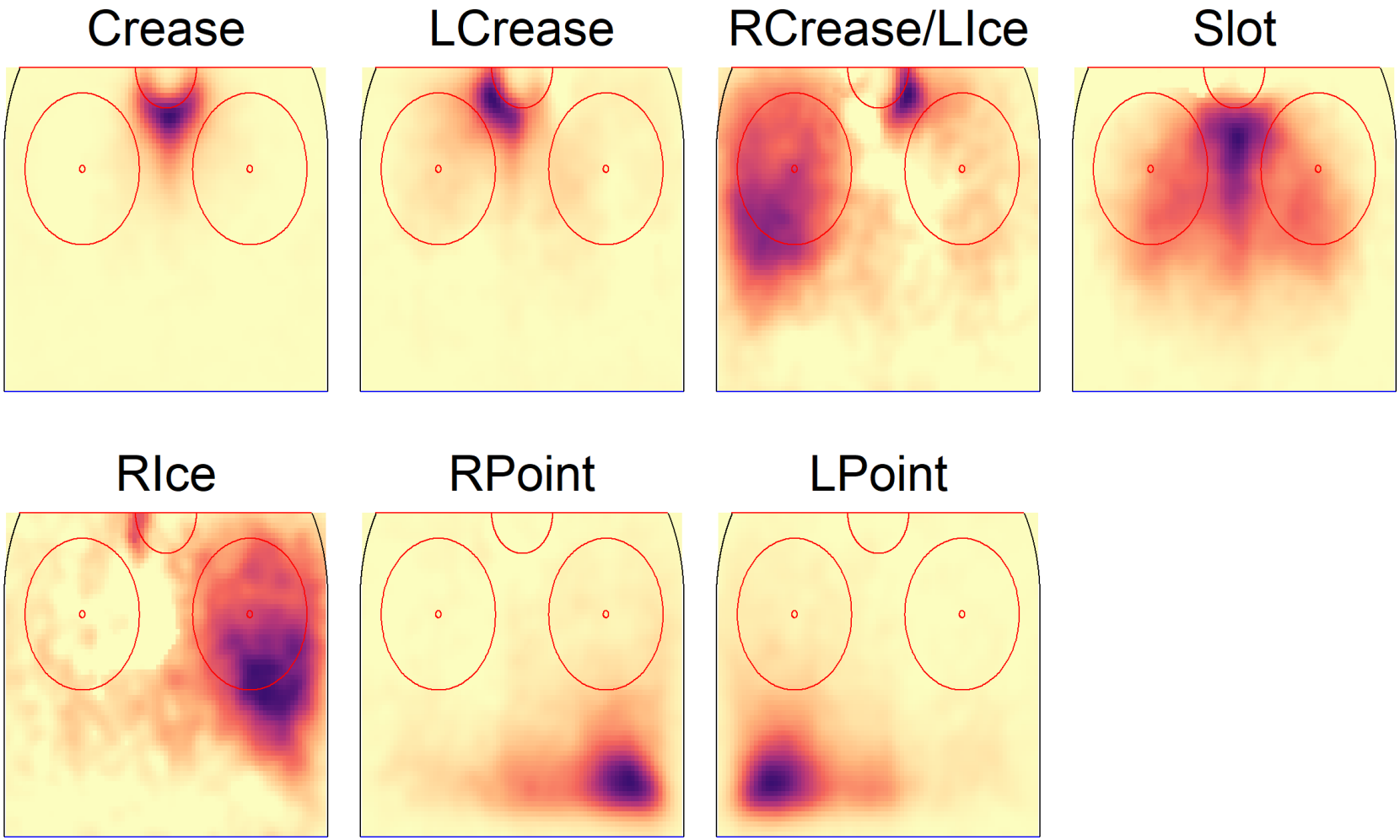
- Too many bases and you're modelling the noise.
- Too few and your residuals are too large.
- Balance: Interpretability and accuracy.

NHL Shot Location Data

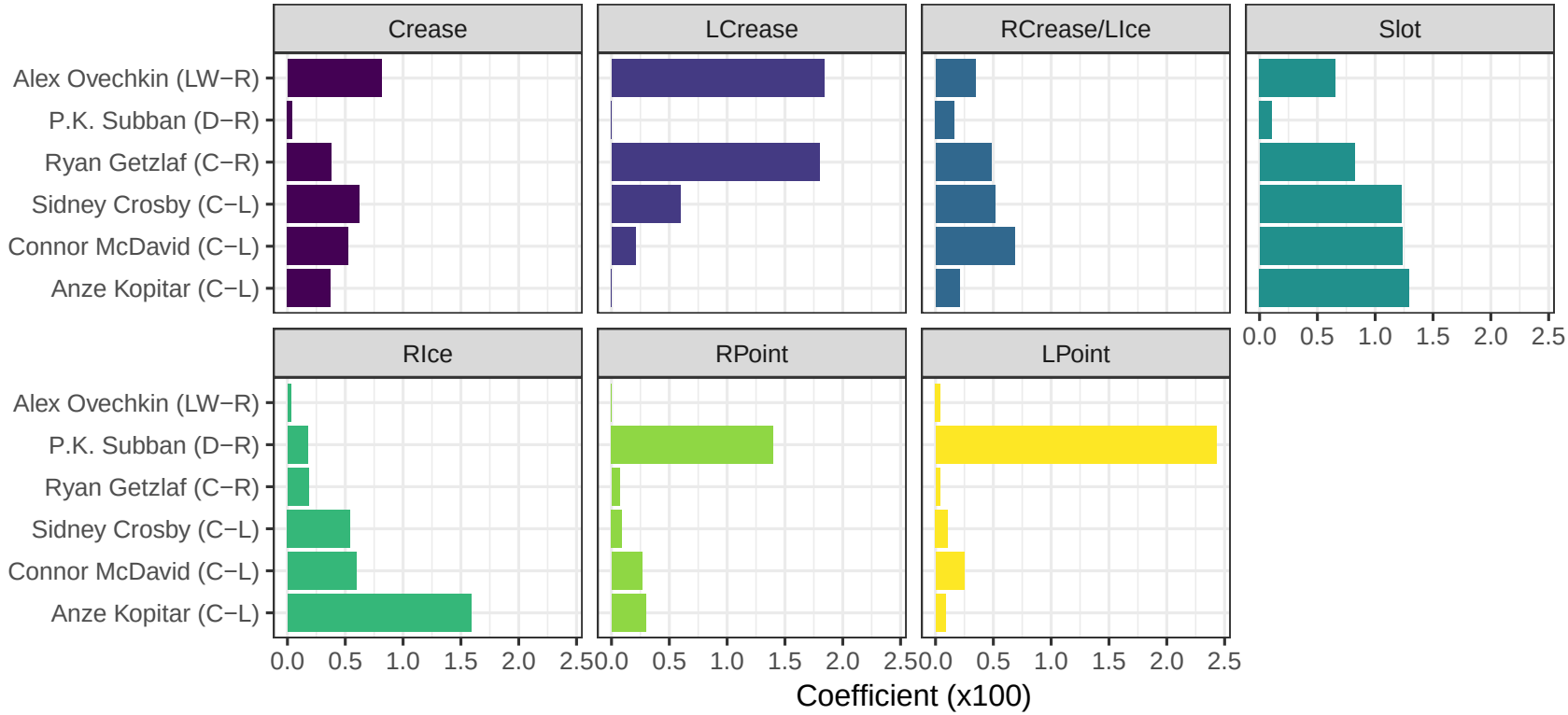
- Scraped from https://statsapi.web.nhl.com/api/v1/...
  - Incomplete version available from a Kaggle data set.
  - Always check robots.txt before scraping.
- Only looked at players with > 500 shots.
- Removed rebounds.
  - Focused on shot choice, not opportunity.

In all, 510,847 shots and goals from 466 players.

Estimated Bases for NHL Data

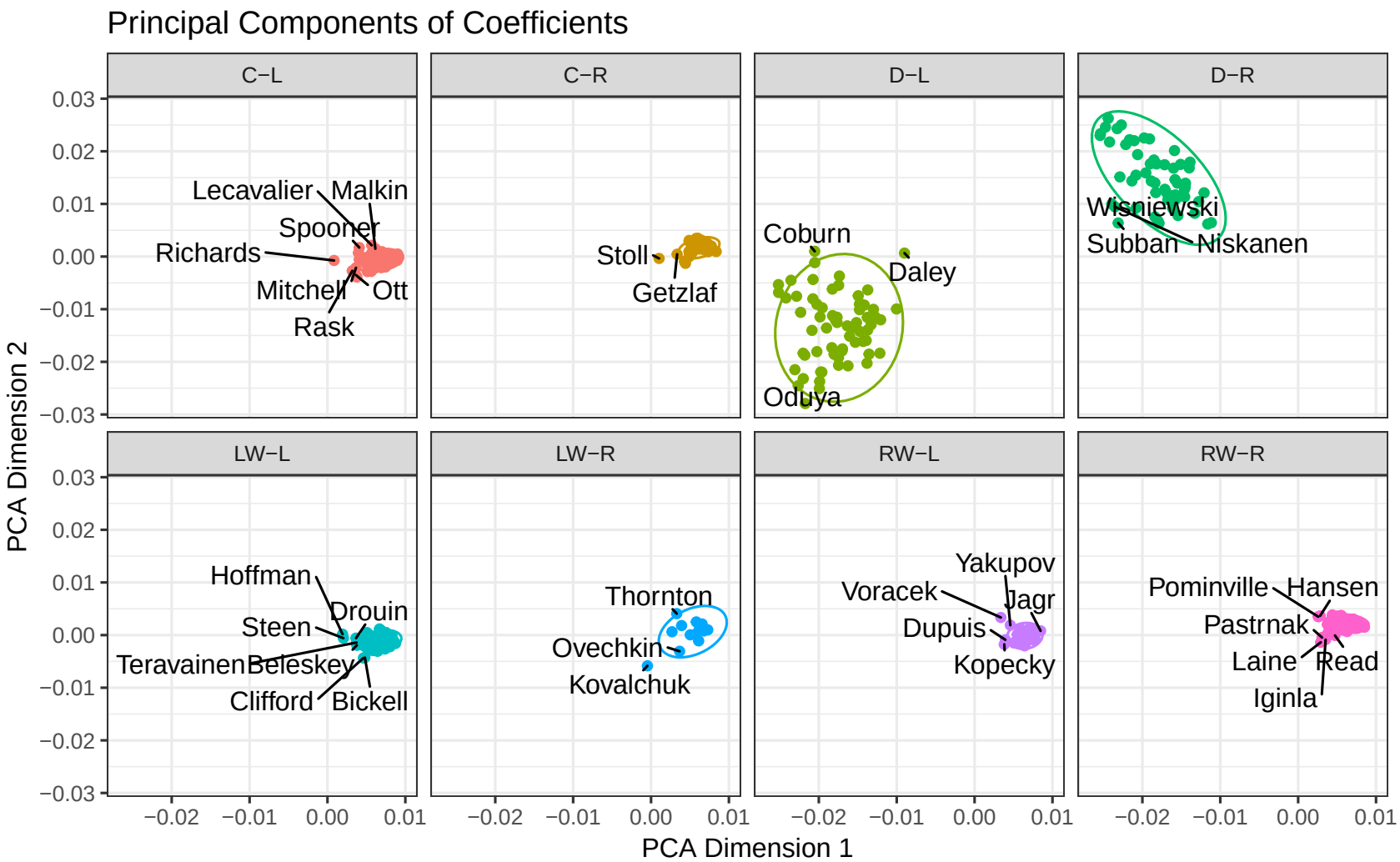


Some Select Players



Different position/handedness shoot from different places, but same position/handedness can be different too.

Basis Results - Outliers



Shot Quality

Basketball versus Hockey

For players who played at least half of the 2017-2018 season:

Basketball

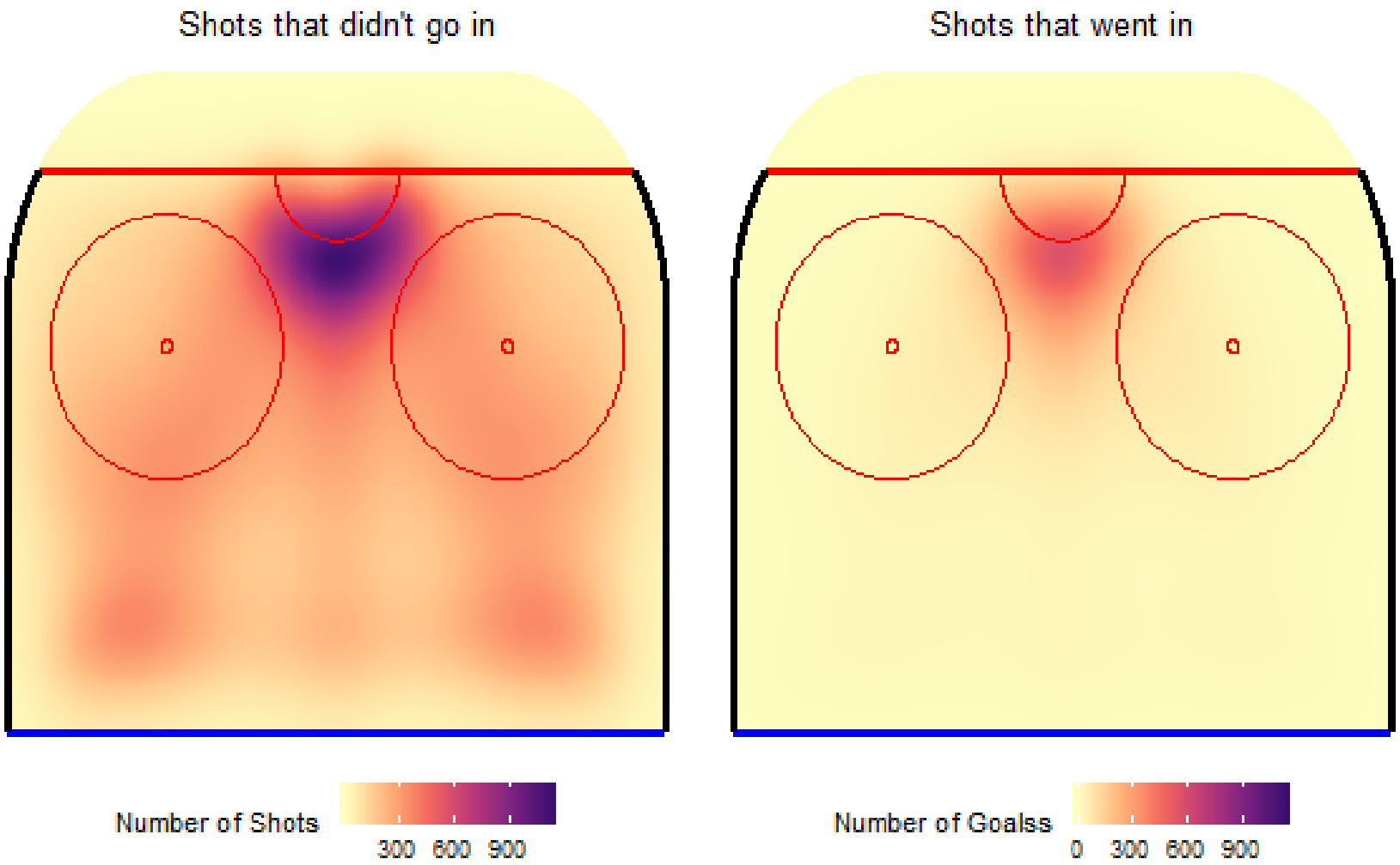
- WORST 3-point percentage: 29.8% (Russell Westbrook)

Hockey

- BEST shooting percentage: 23.5% (Alexander Kerfoot)

Hockey just doesn't have enough goals per player!

Where do Goals Come From?

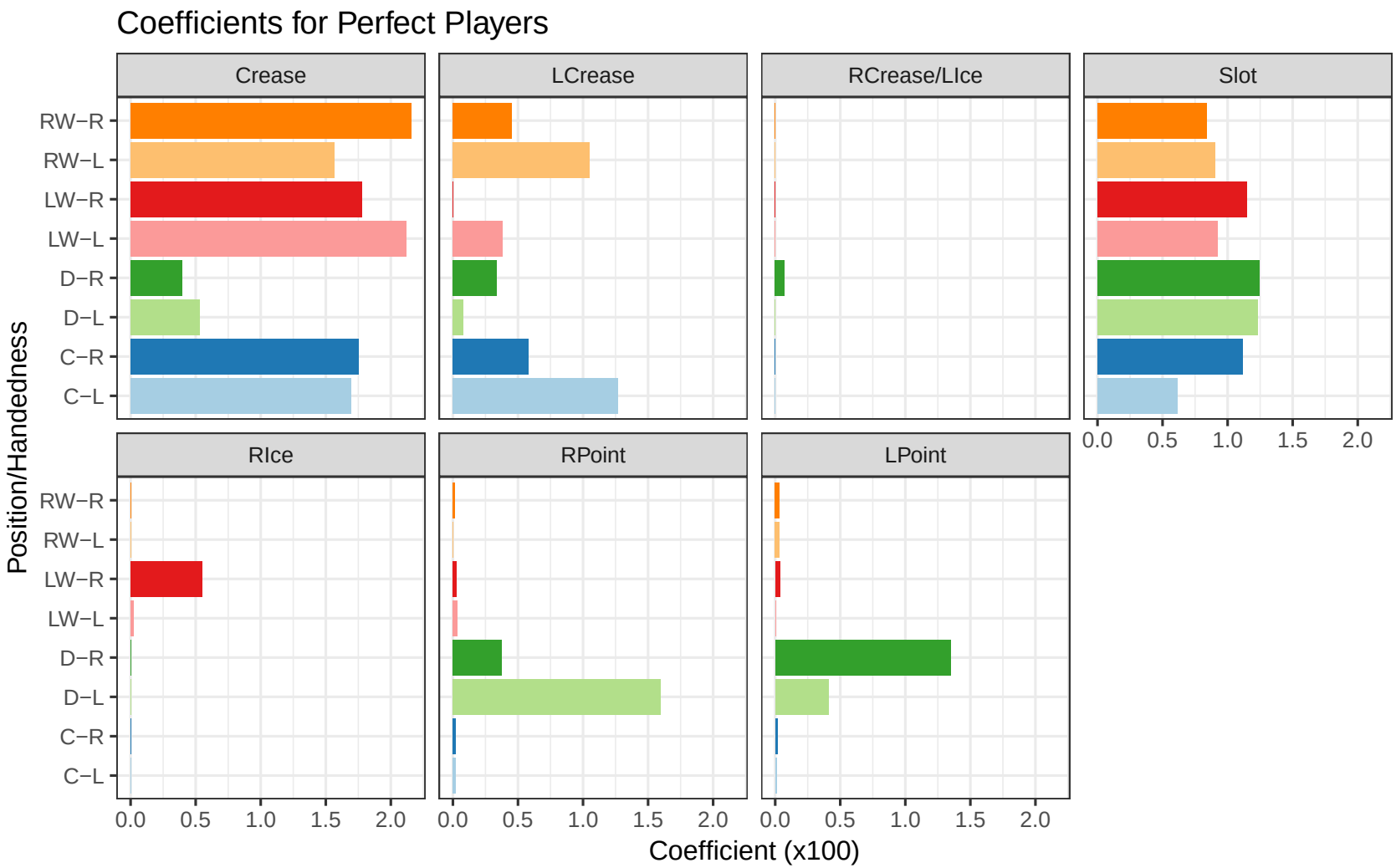


Perfect Players

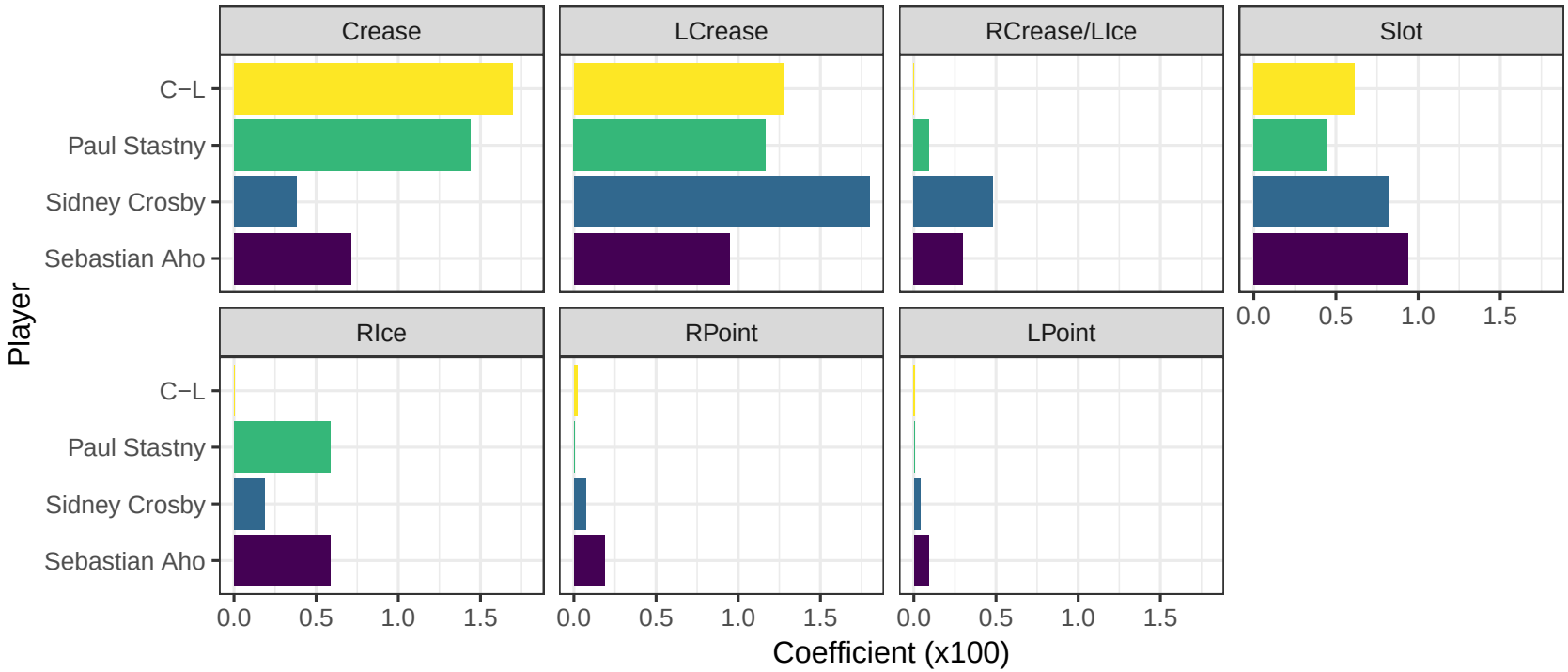
Instead of looking at a player's goals, we use the goals as a player.

- Goals are split up based on position/handedness.
- Name of player is replaced by the position/handedness.
- These "perfect players" are added to the algortihm.

Coefficients for the Perfect Players

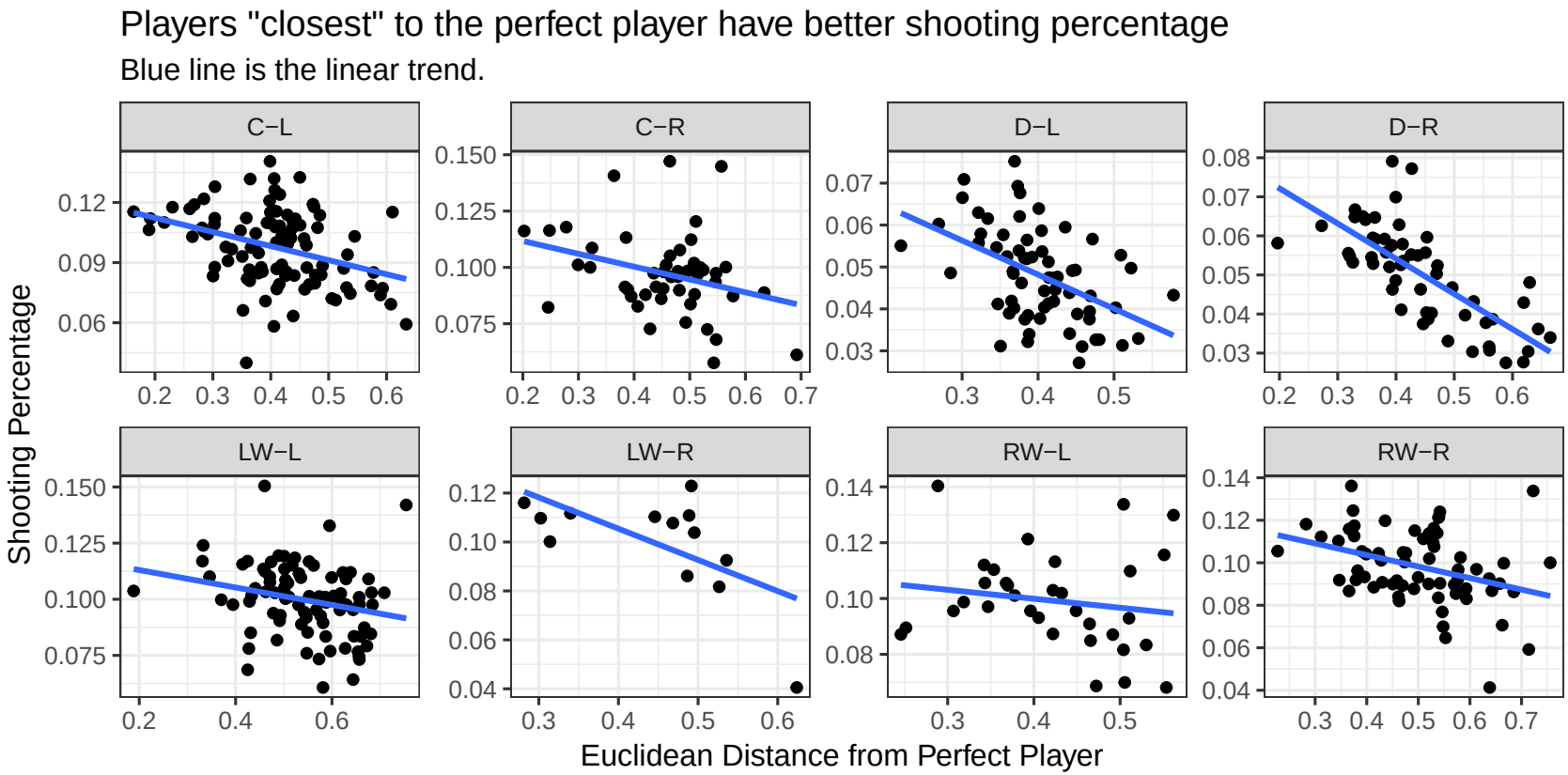


Comparing Stastny, Crosby, and Aho to the Perfect Player



- Stastny: Follows the Perfect Player, "Could stand to shoot the puck more." - EliteProspects.com
- Crosby: Shoots from different place, but scores.
- Aho: "Very good passer." - EliteProspects.com

Euclidean distance as shot quality



Recall: estimates were normalized to **not** include the number of shots or goals.

Conclusions

- Bases provide good foundations for heuristic advice.
  - "Shoot more when you're at the point, less from the faceoff circle."
- Coefficients simplify the comparison of player shooting strategy.
- Perfect players give us a measure of quality when we don't have enough data.
- Playing like a perfect player does not mean you'll always score.

Acknowledgements

NSERC CRSNG

CANSSI INCASS