

# Clustering and Analyzing 5v5 NHL Shot Location Data

Brendan Kumagai, McMaster University / University of Toronto

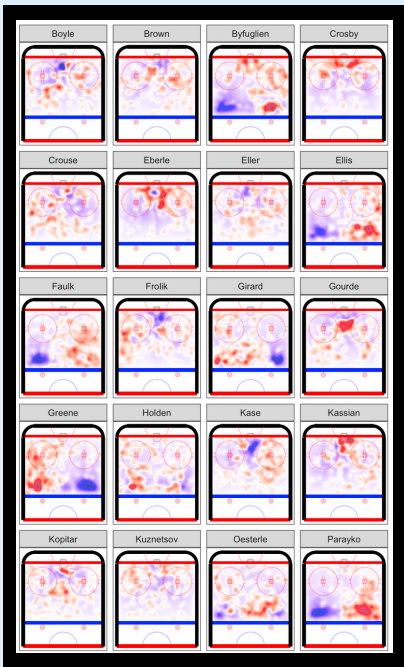
Twitter: @BrendanKumagai  
Website: kumihockey.com  
Email: brendankumagai@gmail.com

## Introduction

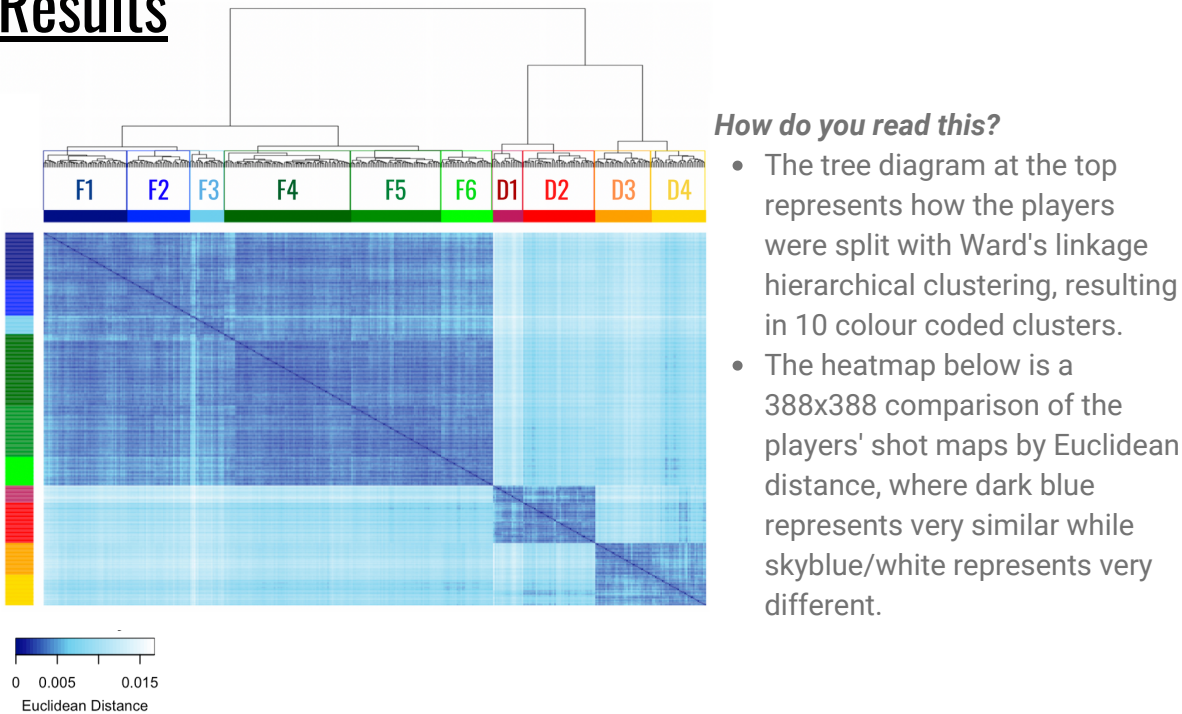
**Problem:** Heatmap-style visualizations are a common and intuitive way to display patterns in shot location data for individual players or teams. However, analyzing these shot maps becomes exponentially more difficult as the number of players/teams increases. For example of this issue, zoom in on the shot heatmaps on the right, now try imagine analyzing these and about 360 more players at once. That will get difficult really quickly.

**Solution:** I propose a framework for grouping NHL players by 5v5 shot location data through fitting a shot density polygrid for each player and performing Ward's linkage hierarchical clustering [1] as outlined in the Methodology section.

Once I have the hierarchical clustering results, I will analyze the composition of each cluster and compare differences between clusters to search for any interesting observations and patterns in NHL 5v5 shot maps.



## Results



How do you read this?

- The tree diagram at the top represents how the players were split with Ward's linkage hierarchical clustering, resulting in 10 colour coded clusters.
- The heatmap below is a 388x388 comparison of the players' shot maps by Euclidean distance, where dark blue represents very similar while skyblue/white represents very different.

## Key Facts

- 10 clusters
- 3 subgroups
  - Home plate forwards
  - Perimeter forwards
  - Defencemen
- Results separate:
  - Perimeter and home plate shooters
  - "Dynamic" D-men who produce more shots deep in the zone from "static" D-men who are very highly concentrated in their corner.
- See "Cluster Composition Observations" for more interesting trends.

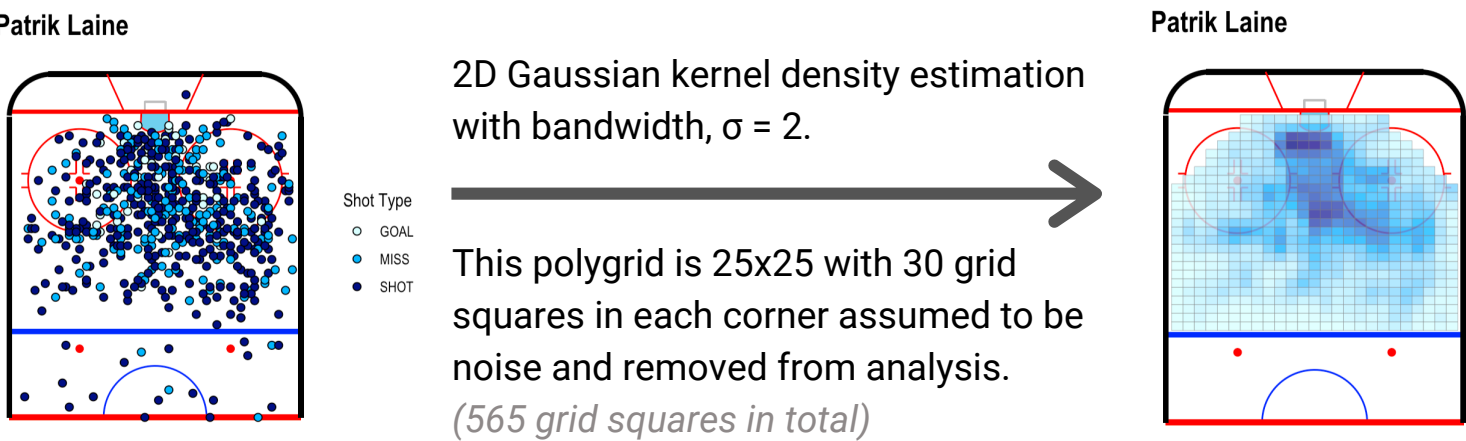
## Methodology

### 1 Obtain and prepare NHL 5v5 shot location data

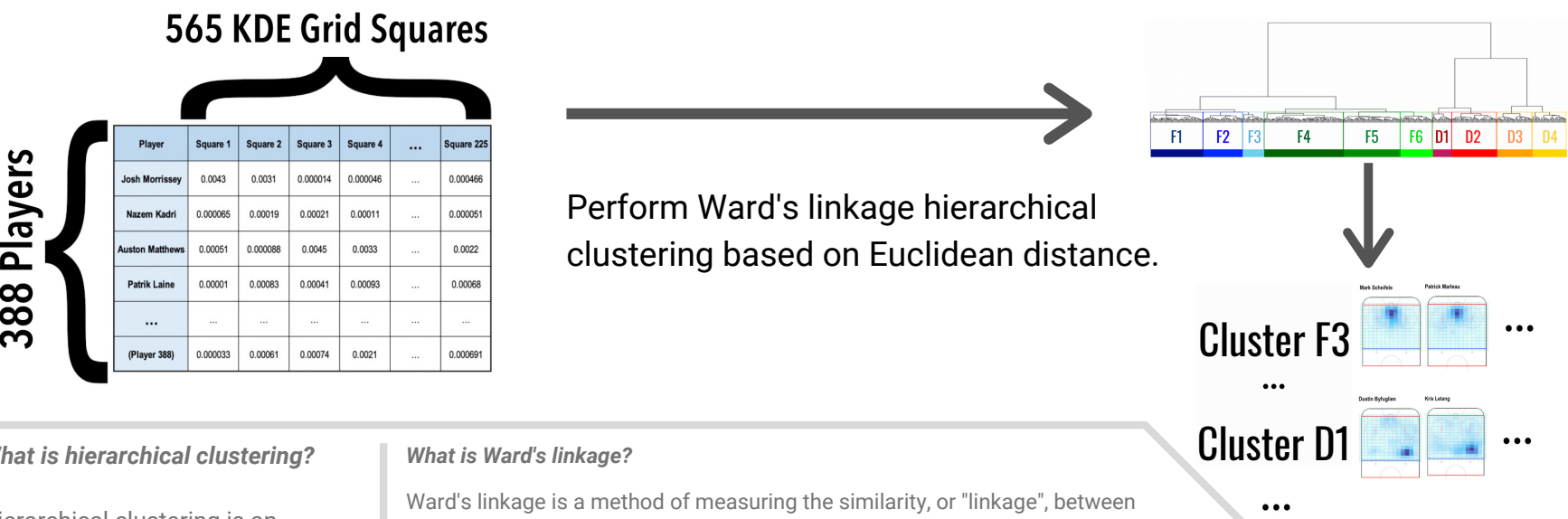
Player	x-Coords	y-Coords
Josh Morrissey	27	35
Nazem Kadri	-10	77
Auston Matthews	2	81
Patrik Laine	25	66
...		

- Shot location data obtained from evolving-hockey.com [2].
  - All unblocked shots
  - At 5v5
  - 3 NHL regular seasons (2017-18 to 2019-20)
- I will be analyzing the 388 players with at least 300 5v5 shots in this dataset.

### 2 Create a shot density polygrid for each player



### 3 Group players based on their shot maps with hierarchical clustering



#### What is hierarchical clustering?

Hierarchical clustering is an algorithm that groups together objects (players, in this case) based on their features (shot density polygrid squares).

#### Algorithm:

- Start with each player in his own cluster (388 clusters).
- Link the two most similar clusters.
- Repeat 1 & 2 until you reach the desired number of clusters.

#### What is Ward's linkage?

Ward's linkage is a method of measuring the similarity, or "linkage", between two clusters in hierarchical clustering.

Before using Ward's linkage, I define the (Euclidean) distance between player x and player y as  $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^{565} (x_j - y_j)^2}$  where  $x_j$  and  $y_j$  represents the jth grid square density for player x and y, respectively.

Ward's linkage determines which two clusters will lead to the **minimal increase in the sum of squared distances** between each player and the mean value of the cluster when you form these two clusters together.

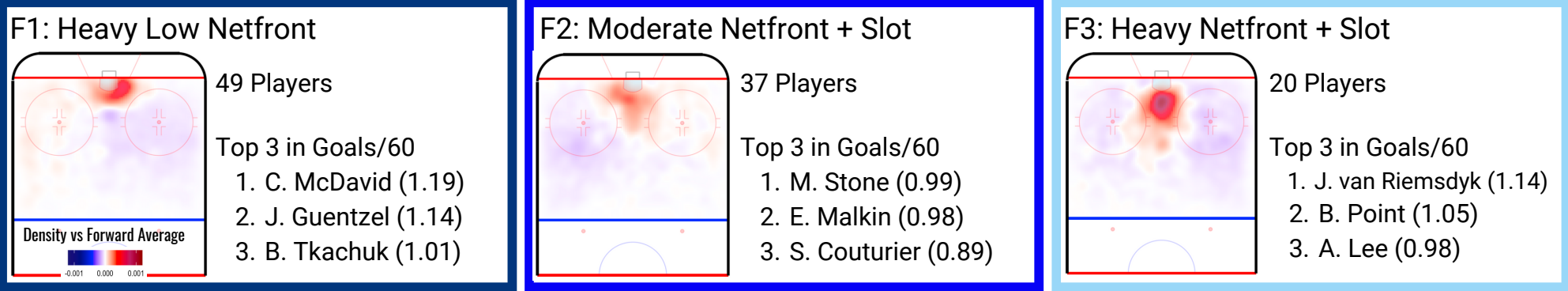
Determining the best two clusters to link via Ward's linkage:

- Find the mean value of each cluster. The mean of cluster k is  $\mu_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_i$  where  $\mathbf{x}_1, \dots, \mathbf{x}_{n_k}$  are the vectors of grid square densities for all  $n_k$  players in the cluster.
- For each cluster, calculate the distance between the cluster mean and each player, then square these values and sum them up. This is called the "sum of squared distances of the cluster and is defined as  $SS_k = \sum_{i=1}^{n_k} d(\mathbf{x}_i, \mu_k)^2$  for cluster k.
- For each pair of clusters, assume they are a single cluster and repeat Steps 1 and 2.
- For each pair of clusters, take the sum of squared distances value for the merged cluster from step 3 and subtract the individual sum of squared distances values found in Step 2. This is written as  $E_{k_1, k_2} = SS_{k_1 \cup k_2} - (SS_{k_1} + SS_{k_2})$
- Select the pair of clusters that has the minimum value from Step 4. Mathematically, this is expressed as  $\arg \min_{k_1, k_2} E_{k_1, k_2}$

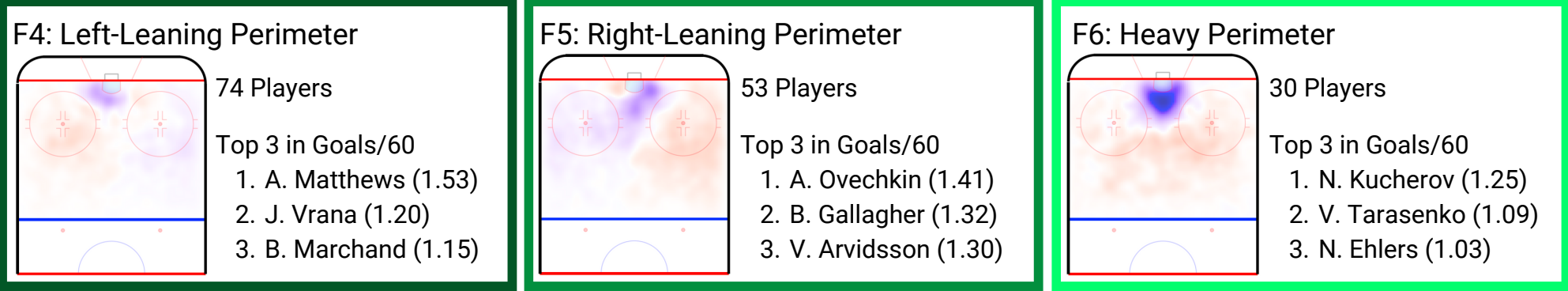
## References

- [1] Ward, J.H., 1963. Hierarchical grouping to optimize an objective function, Journal of the American Statistical Association, 58, 236–244
- [2] Younggren, J. & Younggren, L. NHL BPP Query. Retrieved August 20, 2020. <https://www.evloving-hockey.com>
- [3] Yu, D., Boucher, C., Bornn, L., & Javan, M. Playing Fast Not Loose: Evaluating team-level pace of play in ice hockey using spatio-temporal possession data. arXiv preprint arXiv:1902.02020 (2019).

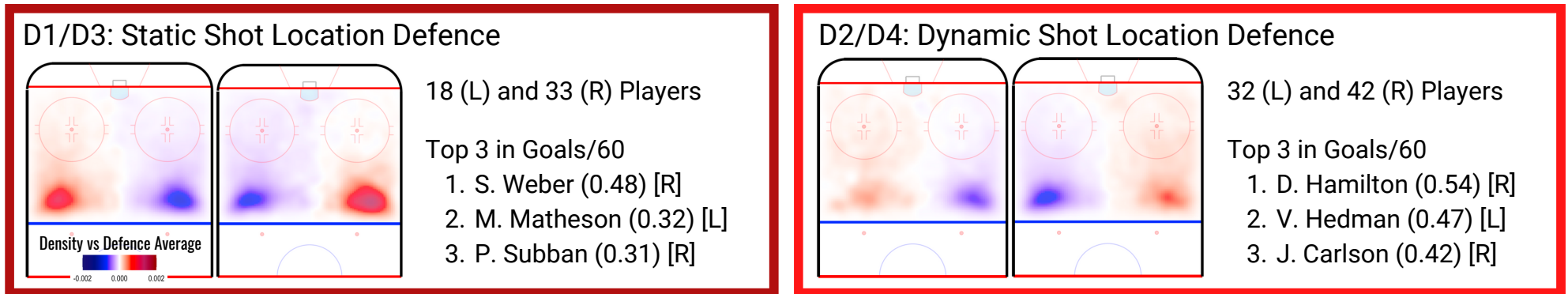
## Home Plate Forward Clusters



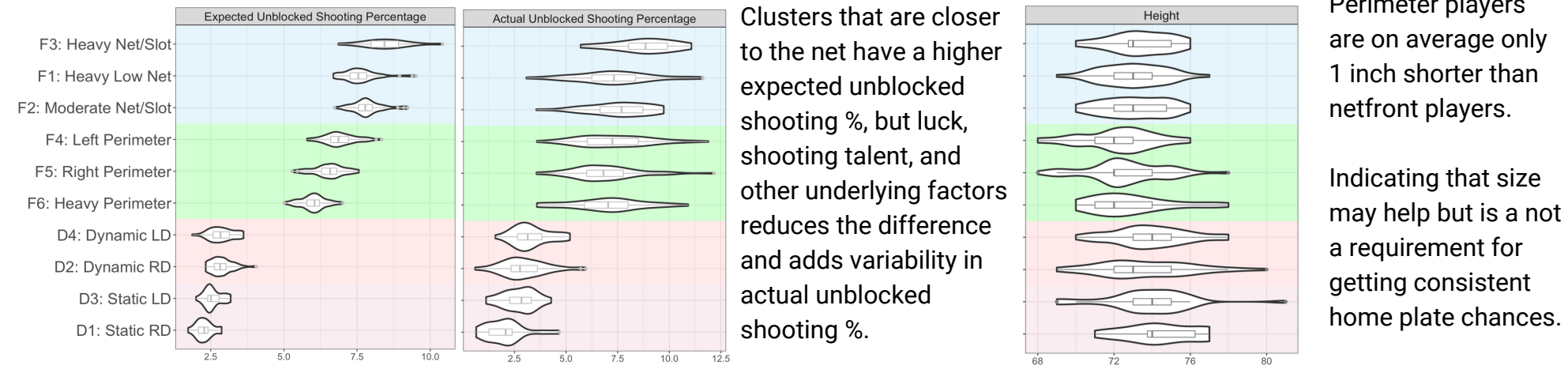
## Perimeter Forward Clusters



## Defence Clusters



## Cluster Composition Observations



## Main Takeaways

- Through fitting a density polygrid to each player's shot location data and using hierarchical clustering, I identify 10 clusters of players based on their shooting patterns.
- These clusters can be separated into subgroups of perimeter forwards, home plate forwards and defencemen. Furthermore, defencemen can be divided into dynamic and static shooting patterns.
- The distribution of any statistic, metric or value among players in a cluster can be determined and compared to draw meaningful insights.

## Future Work

- Apply this framework to any other sports location data.
- Continue to explore other high-dimensional clustering techniques for this problem.
- Develop "shot map similarities" comparing each individual pair of players with some similarity measurement

# Appendix: Cluster Results

<div><div>F1: Heavy Low Net</div><div><div>Andrew Copp</div><div>Marcus Foligno</div><div>Paul Stastny</div><div>Artem Anisimov</div><div>Jake Guentzel</div><div>Brandon Saad</div><div>Connor McDavid</div><div>Anthony Duclair</div><div>Milan Lucic</div><div>Nick Ritchie</div><div>Patrick Maroon</div><div>Jordan Martinook</div><div>Matthew Tkachuk</div><div>Tobias Rieder*</div><div>Zack Kassian</div><div>Anze Kopitar</div><div>Scott Laughton</div><div>Matt Calvert</div><div>Nick Bonino</div><div>Antoine Roussel</div><div>Evander Kane</div><div>Jamie Benn</div><div>Phillip Danault</div><div>William Carrier</div><div>Artturi Lehkonen</div><div>Brock McGinn</div><div>Chris Kreider</div><div>Sebastian Aho</div><div>Gabriel Landeskog</div><div>Nick Cousins</div><div>Matt Duchene</div><div>Tomas Nosek</div><div>Kevin Hayes</div><div>Sam Bennett</div><div>Filip Chytil</div><div>Bo Horvat</div><div>Pavel Buchnevich</div><div>Carl Soderberg</div><div>Jimmy Vesey</div><div>Adam Henrique</div><div>Derick Brassard</div><div>Nick Foligno</div><div>Eric Staal</div><div>Kevin Fiala</div><div>Joel Eriksson Ek</div><div>Zach Parise</div><div>Tyler Ennis</div><div>Andrei Svechnikov</div><div>Brady Tkachuk</div></div></div>	<div><div>F2: Moderate Netfront + Slot</div><div><div>Evgeni Malkin</div><div>Tyler Bozak</div><div>Oskar Sundqvist</div><div>Stefan Noesen</div><div>Sidney Crosby</div><div>Sam Reinhart</div><div>Ryan Strome</div><div>Colton Sissons</div><div>Micheal Ferland</div><div>Derek Ryan</div><div>Chris Tierney</div><div>Jonathan Toews</div><div>Sean Couturier</div><div>Tomas Hertl</div><div>Joonas Donskoi</div><div>Patric Hornqvist</div><div>Joe Pavelski</div><div>Tom Wilson</div><div>Wayne Simmonds</div><div>Ryan Carpenter</div><div>Ryan Johansen</div><div>Noel Acciari</div><div>T.J. Oshie</div><div>Warren Foegele</div><div>Mark Stone</div><div></div><div>Ryan Dzingel</div><div></div><div>Luke Glendening</div><div></div><div>Nino Niederreiter</div><div></div><div>Charlie Coyle</div><div></div><div>Bryan Rust</div><div></div><div>Chris Wagner</div><div></div><div>Corey Perry</div><div></div><div>Pierre-Luc Dubois</div><div></div><div>Mathew Barzal</div><div></div><div>Jordan Eberle</div><div></div><div>Alex Radulov</div><div></div><div>Mattias Janmark</div><div></div></div></div>	<div><div>F3: Heavy Netfront + Slot</div><div><div>Patrick Marleau</div><div></div><div>James Van Riemsdyk</div><div></div><div>Mark Scheifele</div><div></div><div>Sean Monahan</div><div></div><div>Zemgus Girgensons</div><div></div><div>Michael Grabner</div><div></div><div>Anders Lee</div><div></div><div>Brock Nelson</div><div></div><div>Josh Bailey</div><div></div><div>Anthony Beauvillier</div><div></div><div>Brayden Point</div><div></div><div>Radek Faksa</div><div></div><div>Zach Hyman</div><div></div><div>Vladislav Namestnikov</div><div></div><div>Nico Hischier</div><div></div><div>Alex Kerfoot</div><div></div><div>Jordan Staal</div><div></div><div>Christian Dvorak</div><div></div><div>Jesper Fast</div><div></div><div>Tyler Bertuzzi</div><div></div></div></div>	<div><div>F4: Left-Leaning Perimeter</div><div><div>Nazem Kadri</div><div>Mikael Granlund</div><div>J.T. Miller</div><div>Auston Matthews</div><div>Riley Sheahan</div><div>Mats Zuccarello</div><div>Mathieu Perreault</div><div>Anthony Mantha</div><div>Tyler Johnson</div><div>Conor Sheary</div><div>Mikko Koivu</div><div>Taylor Hall</div><div>Brayden Schenn*</div><div>Tomas Tatar</div><div>Jesper Bratt</div><div>Carl Hagelin</div><div>Jason Zucker</div><div>Blake Coleman</div><div>Jaden Schwartz</div><div>Richard Panik</div><div>Mikko Rantanen*</div><div>Johnny Gaudreau</div><div>Andrew Cogliano*</div><div>Jeff Skinner</div><div>Mikael Backlund</div><div>Max Domi</div><div>P.E. Bellemare</div><div>Leon Draisaitl</div><div>Clayton Keller</div><div>Ryan Nugent-Hopkins</div><div>Logan Couture</div><div>Brendan Perlini</div><div>Timo Meier</div><div>Jake Debrusk</div><div>Alex Iafallo</div><div>Brad Marchand</div><div>Filip Forsberg</div><div>Tanner Pearson</div><div>Miles Wood</div><div>Jonathan Drouin</div><div>Kyle Clifford*</div><div>Brandon Tanev</div><div>Max Pacioretty</div><div>John Tavares</div><div>Lawson Crouse</div><div>Alex Galchenyuk</div><div>Aleksander Barkov</div><div>Danton Heinen</div><div>Ryan O'Reilly</div><div>Ondrej Palat**</div><div>Yanni Gourde**</div><div>Tyson Jost*</div><div>Alex Killorn**</div><div>Kyle Connor</div><div>Jakub Vrana</div><div>Evgeny Dadonov</div><div>Alex Steen</div><div>Lars Eller</div><div>Steven Stamkos**</div><div>Boone Jenner</div><div>Evgeny Kuznetsov</div><div>Jonathan Huberdeau</div><div>Andreas Athanasiou</div><div>Zack Smith</div><div>Cody Eakin</div><div>Mark Jankowski*</div><div>Nicklas Backstrom</div><div>Erik Haula</div><div>Dominik Simon</div><div>Dylan Larkin</div><div>William Karlsson</div><div>Ryan Donato</div><div>Gustav Nyquist</div><div>Reilly Smith</div><div></div></div></div>	<div><div>F5: Right-Leaning Perimeter</div><div><div>Connor Brown</div><div>Jakob Silfverberg</div><div>William Nylander*</div><div>Rickard Rakell*</div><div>Travis Konecny</div><div>Ondrej Kase</div><div>Jakub Voracek</div><div>Tyler Toffoli</div><div>Kevin Labanc</div><div>Trevor Lewis</div><div>David Pastrnak</div><div>Dustin Brown</div><div>Craig Smith</div><div>Cam Atkinson*</div><div>Viktor Arvidsson</div><div>Oliver Bjorkstrand</div><div>Calle Jarnkrok</div><div>Nick Bjugstad</div><div>Brendan Gallagher*</div><div>Colton Sceviour</div><div>Kyle Okposo</div><div>Tyler Pitlick</div><div>Jason Pominville</div><div>Jason Spezza</div><div>Jack Eichel</div><div>Jonathan Marchessault</div><div>Nathan MacKinnon</div><div>Tyler Seguin</div><div>Blake Comeau</div><div>Elias Lindholm</div><div>Mika Zibanejad*</div><div>Justin Williams*</div><div>Brett Connolly</div><div>Sam Gagner</div><div>Alex Ovechkin</div><div>Jeff Carter</div><div>Bobby Ryan</div><div>David Krejci</div><div>Jean-Gabriel Pageau</div><div>Ryan Getzlaf</div><div>Kyle Turris</div><div>Josh Anderson</div><div>Ryan Hartman</div><div>Kyle Palmieri</div><div>Brock Richardson*</div><div>Brock Boeser</div><div>Christian Fischer*</div><div>Joel Armia</div><div>Derek Stepan</div><div>Alex Tuch</div><div>Kasperii Kapanen</div><div>David Backes</div><div>Vinnie Hinostroza</div><div></div></div></div>
<div><div>F6: Heavy Perimeter</div><div><div>Nikolaj Ehlers</div><div>Teuvo Teravainen</div><div>Blake Wheeler</div><div>Michael Frolik</div><div>Patrik Laine</div><div>Jake Virtanen</div><div>Vladimir Tarasenko</div><div>Patrice Bergeron</div><div>Phil Kessel</div><div>Brian Boyle</div><div>Claude Giroux</div><div>Evan Rodrigues</div><div>Sean Kuraly</div><div></div><div>Frank Vatrano</div><div></div><div>J.T. Compher</div><div></div><div>Andre Burakovsky</div><div></div><div>Mike Hoffman</div><div></div><div>Frans Nielsen</div><div></div><div>Darren Helm</div><div></div><div>Alex Debrincat</div><div></div><div>Patrick Kane</div><div></div><div>Adrian Kempe</div><div></div><div>Artemi Panarin</div><div></div><div>Vincent Trocheck</div><div></div><div>Nikita Kucherov</div><div></div><div>Jared McCann</div><div></div><div>David Perron</div><div></div><div>James Neal</div><div></div><div>Mitch Marner</div><div></div><div>Joakim Nordstrom</div><div></div></div></div>	<div><div>D1: Static Shooting Right D</div><div><div>Jacob Trouba</div><div></div><div>Adam Larsson</div><div></div><div>Brent Burns</div><div></div><div>Dylan Demelo</div><div></div><div>Justin Braun</div><div></div><div>P.K. Subban</div><div></div><div>Brandon Carlo</div><div></div><div>Radko Gudas</div><div></div><div>Johnny Boychuk</div><div></div><div>Colin Miller</div><div></div><div>Anton Stralman</div><div></div><div>Scott Mayfield</div><div></div><div>Shea Weber</div><div></div><div>Deryk Engelland</div><div></div><div>Travis Hamonic</div><div></div><div>Erik Gudbranson</div><div></div><div>Ryan Pulock</div><div></div><div>Greg Pateryn</div><div></div></div></div>	<div><div>D2: Dynamic Shooting Right D</div><div><div>Dustin Byfuglien</div><div>Brent Seabrook</div><div>Kris Letang</div><div>Josh Manson</div><div>Alex Pietrangelo</div><div>Brandon Montour</div><div>Colton Parayko*</div><div>Drew Doughty</div><div>Dougie Hamilton</div><div>David Savard</div><div>Matt Benning</div><div>Seth Jones</div><div>Charlie McAvoy</div><div>Mark Pysyk</div><div>Rasmus Ristolainen</div><div>Aaron Ekblad</div><div>Tony Deangelo</div><div>John Klingberg</div><div>Kevin Shattenkirk</div><div>Nate Schmidt</div><div>Erik Johnson</div><div>Damon Severson</div><div>Tyson Barrie</div><div>Jeff Petry</div><div>Cody Ceci*</div><div>Justin Faulk</div><div>Matt Niskanen</div><div>Brett Pesce</div><div>John Carlson</div><div>Trevor Van Riemsdyk</div><div>Mike Green</div><div>Troy Stecher</div><div>Nick Jensen</div><div>Tyler Myers</div><div>Jared Spurgeon</div><div>Erik Karlsson</div><div>Matt Dumba</div><div>Alex Biega</div><div>Connor Murphy</div><div>Sami Vatanen</div><div>Neal Pionk</div><div>Ryan Ellis</div></div></div>	<div><div>D3: Static Shooting Left D</div><div><div>Josh Morrissey</div><div>Jacob Slavin*</div><div>Olli Maatta</div><div>Noah Hanifin*</div><div>Joel Edmundson</div><div>Vince Dunn</div><div>Oscar Klefbom</div><div>Torey Krug</div><div>Mark Giordano</div><div>Alex Edler</div><div>Darnell Nurse</div><div>Ben Hutton</div><div>Brenden Dillon</div><div>Will Butcher</div><div>Matt Grzelcyk</div><div>Jonas Brodin</div><div>Zdeno Chara</div><div>Marc-Edouard Vlasic</div><div>Marco Scandella</div><div>Brett Kulak</div><div>Ryan Suter</div><div>Hampus Lindholm</div><div>Duncan Keith</div><div>Ben Chiarot</div><div>Michal Kempny</div><div>Jakob Chychrun</div><div>Oliver Ekman-Larsson</div><div></div><div>Jake Muzzin</div><div></div><div>Jack Johnson</div><div></div><div>Keith Yandle</div><div></div><div>Mike Matheson*</div><div></div><div>Brayden McNabb</div><div></div><div>Andy Greene</div><div></div></div></div>	<div><div>D4: Dynamic Shooting Left D</div><div><div>Morgan Rielly</div><div>Esa Lindell</div><div>Brian Dumoulin</div><div>Jake Gardiner</div><div>Ian Cole</div><div>John Moore</div><div>T.J. Brodie</div><div>Nikita Zadorov</div><div>Ivan Provorov</div><div>Alex Goligoski</div><div>Robert Hagg</div><div>Nick Holden</div><div>Shayne Gostisbehere</div><div>Samuel Girard</div><div>Mattias Ekholm</div><div>Thomas Chabot</div><div>Roman Josi</div><div>Markus Nutivaara</div><div>Jordie Benn*</div><div>Jordan Oesterle</div><div>Brady Skjei</div><div>Shea Theodore</div><div>Ryan McDonagh</div><div>Miro Heiskanen</div><div>Dmitry Orlov</div><div></div><div>Travis Sanheim</div><div></div><div>Adam Pelech</div><div></div><div>Nick Leddy</div><div></div><div>Zach Werenski</div><div></div><div>Victor Hedman</div><div></div><div>Mikhail Sergachev</div><div></div><div>Braydon Coburn*</div><div></div></div></div>

\* = Player is not the best fit in their given cluster by my "eye test" on their individual shot map

\*\* = A group of Tampa Bay forwards who ended up in the "Left-Leaning Perimeter" cluster, but had an abundance of shots *juuust* slightly above the "Heavy Netfront + Slot" cluster, pushing them out into a perimeter cluster.