Towards a spatial analysis of shooting in Philippine basketball

Applications in the University Athletics Association of the Philippines (Season 81)

Ben Hur Pintor, MS Geomatics Engineering, University of the Philippines

FOSS4G BUENOS AIRES 2021

October 1, 13:30 GMT -3, Salta Room

BN HR openness. data. geospatial.

@bnhr.xyz @BNHRdotXYZ BNHRdotXYZ https://bnhr.xyz

About me

Ben Hur Pintor

geospatial generalist. open stuff advocate. maptivist/datactivist. sports nerd.

Proprietor

BN HR

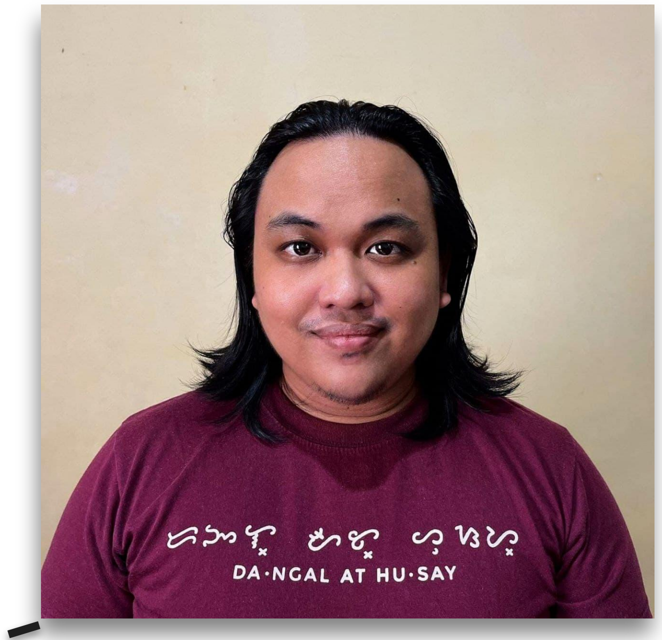
openness.
data.
geospatial.

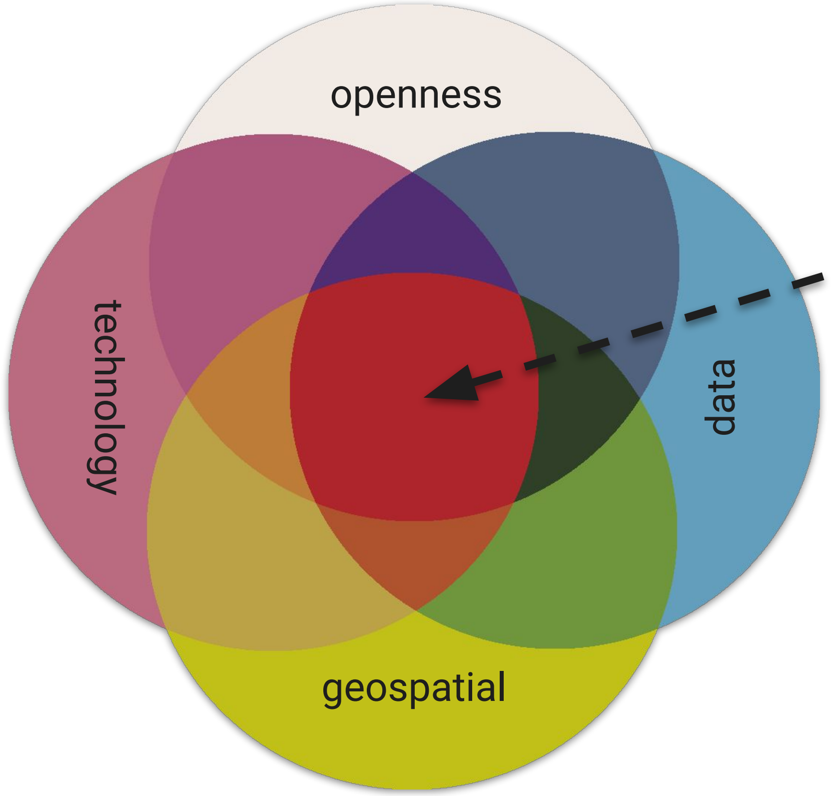
Data Training Lead

Open Knowledge Foundation

Chief Technology Officer

SmartCT
Citizens x Technology





BN HR

openness.
data.
geospatial.

- Established in 2019
- Part enterprise, part advocacy
- Provides training, support, and consulting services on open data, open source, data literacy, and free and open source software for geospatial applications (FOSS4G) -- QGIS, GRASS, GeoNode, etc.

QGIS Certifying Organization

- One of two from the Philippines

QGIS Sustaining Member

- Currently the only (and probably first) sustaining member from the Philippines

Some people I've worked with:

ADB

PISTA NG MAPA


Certificate of Completion

BN HR

Juan de la Cruz

Has attended and completed the course:

Bite-sized QGIS (Beginner) 2nd Run!!!

With a trained competence in:

QGIS Concepts; Layer Creation, Editing, & Styling; Attribute Table & Spatial Queries; Vector & Raster Processing and Analysis

From 21 November 2020 to 6 December 2020

Convened by Mr. Ben Hur Pintor at Quezon City, Philippines

Marco Bernasocchi
Project Representative

Mr. Ben Hur Pintor
Course Convener

ID: QGIS-DEMO

You can verify this certificate by visiting http://changelog.qgis.org/en/qgis/certificate/QGIS-DEMO/.

PISTA NG MAPA

OPEN DATA & SOFTWARE · FREE MAPS · COMMUNITY + TALKS · WORKSHOPS

NOVEMBER 13, 16 & 20 2021 · ONLINE

https://pistangmapa.github.io/2021/

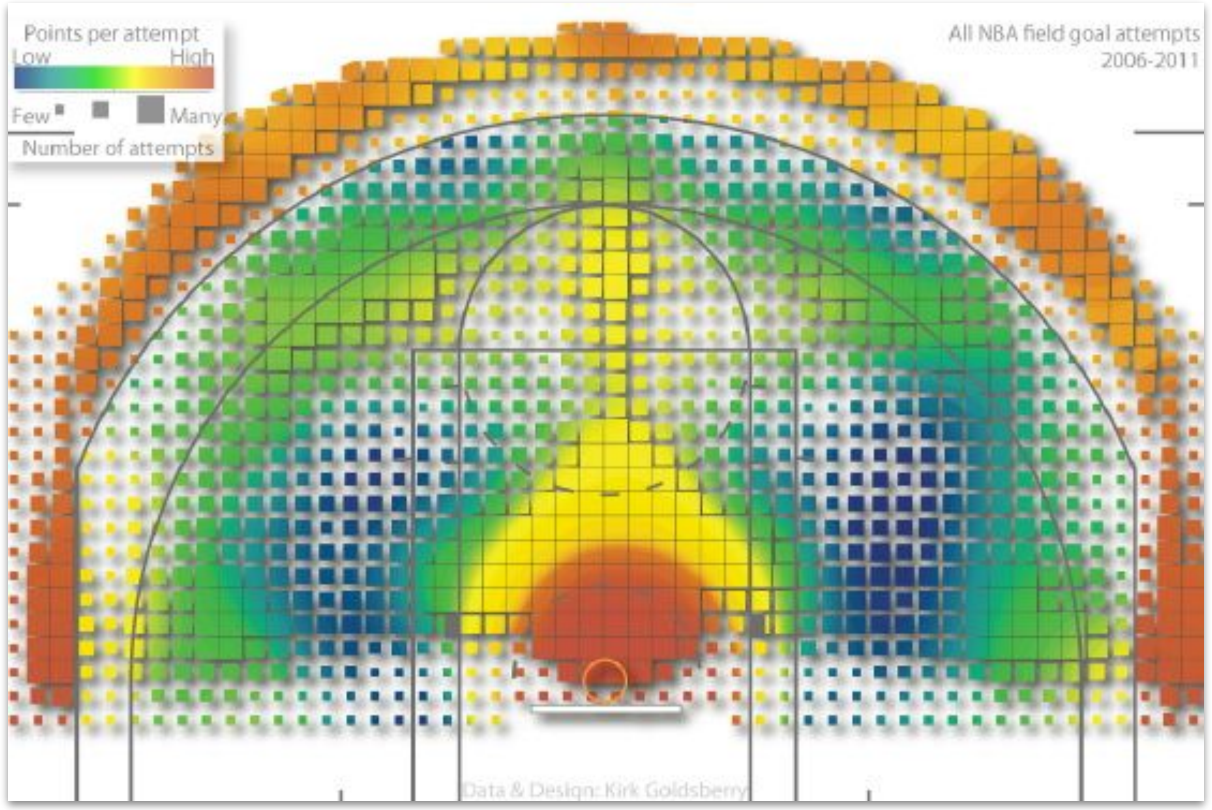
I’m a graduate student taking up MS Geomatics Engineering.

Oh.

And I’m also a big basketball fan.

Why basketball and spatial?

Basketball *is* spatial

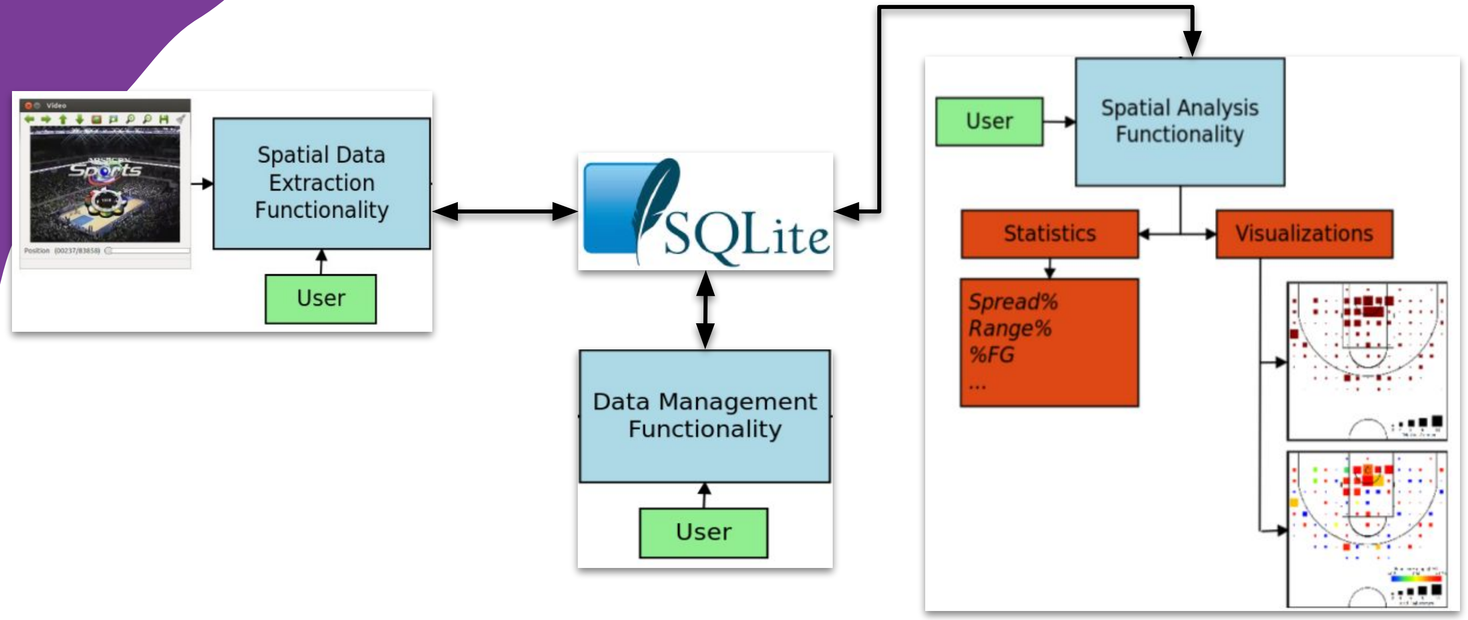


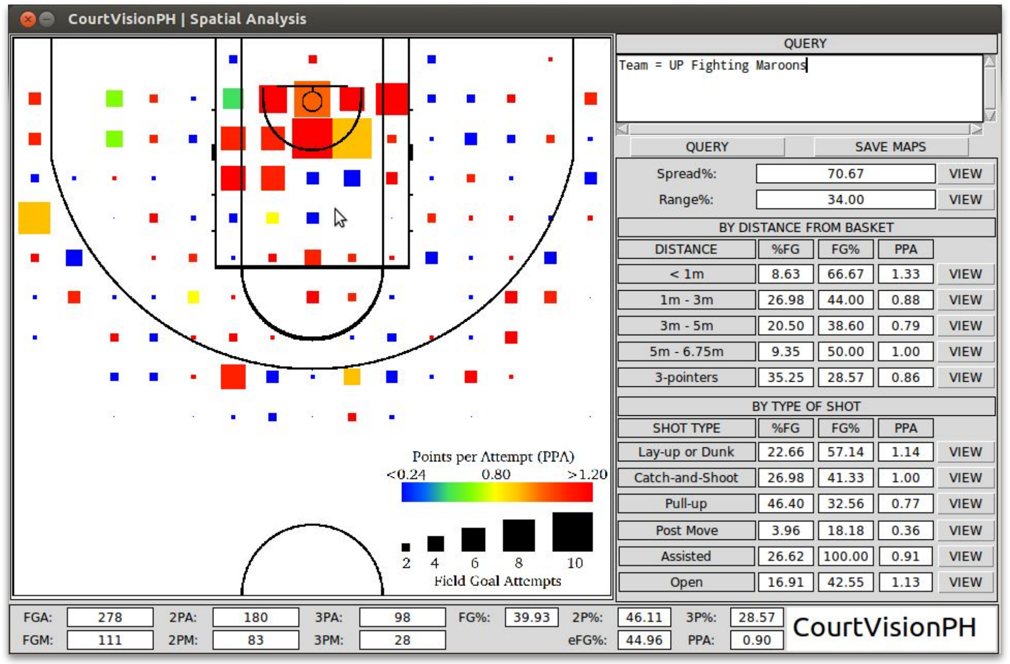
CourtVision: New Visual and Spatial Analytics for the NBA

Dr. Kirk Goldsberry,

2012 MIT-Sloan Sports Analytics Conference

article:
https://web.archive.org/web/20191229154247/http://www.sloansportsconference.com/wp-content/uploads/2012/02/Goldsberry_Sloan_Submission.pdf
video:
https://www.youtube.com/watch?time_continue=3&v=CLhlIVLd0kE&feature=emb_title





CourtVisionPH: A system for the extraction of FGA locations and spatial analysis of shooting using broadcast basketball videos

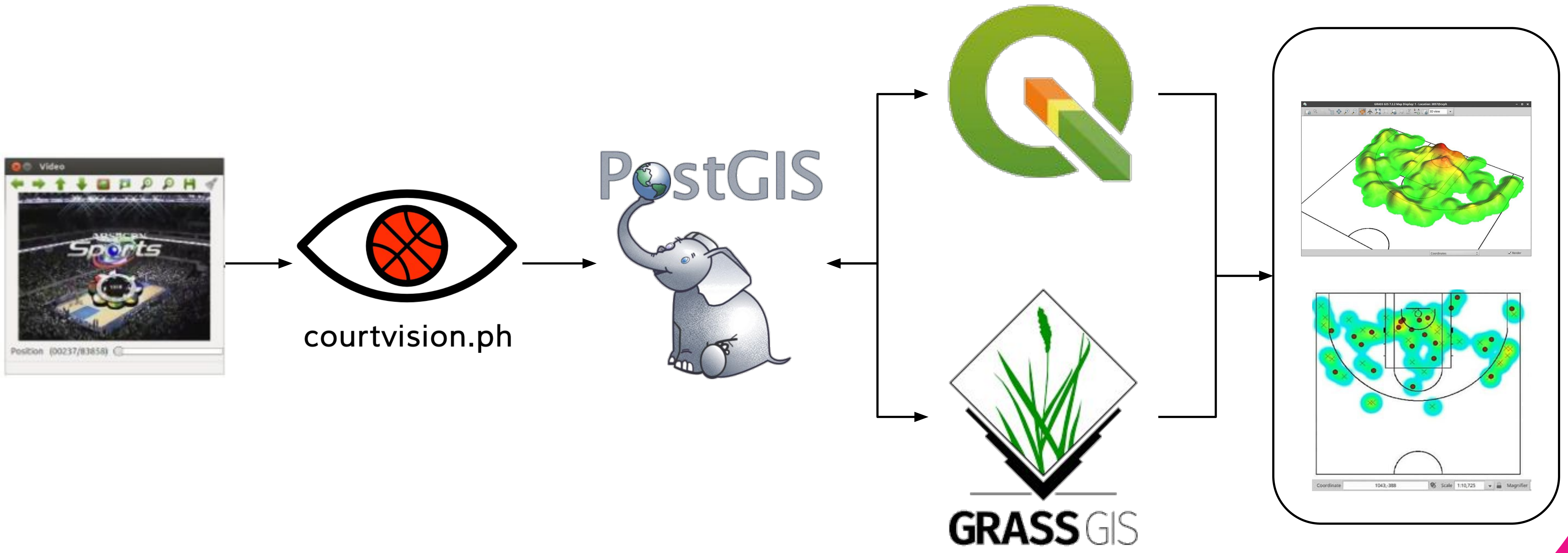
Pintor and Cataniag

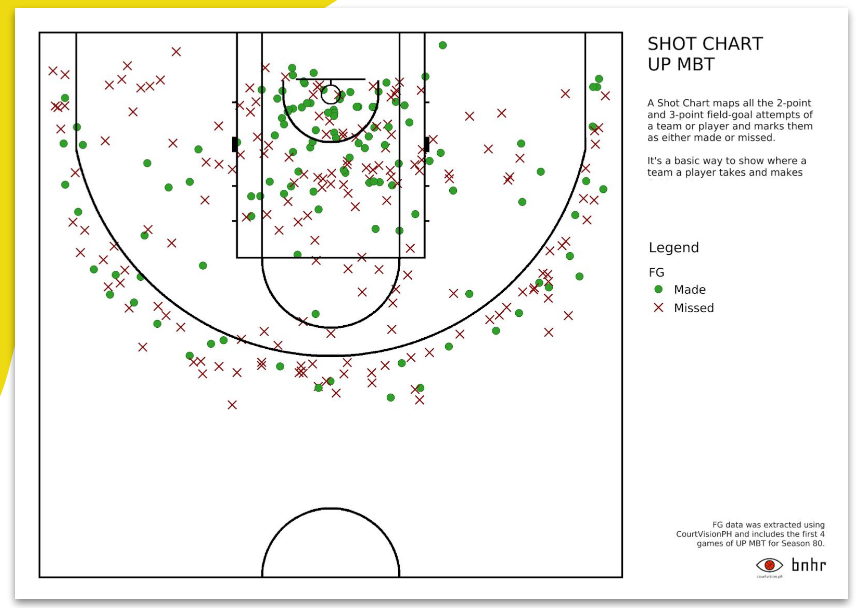
BS GE Undergraduate Thesis (2014)

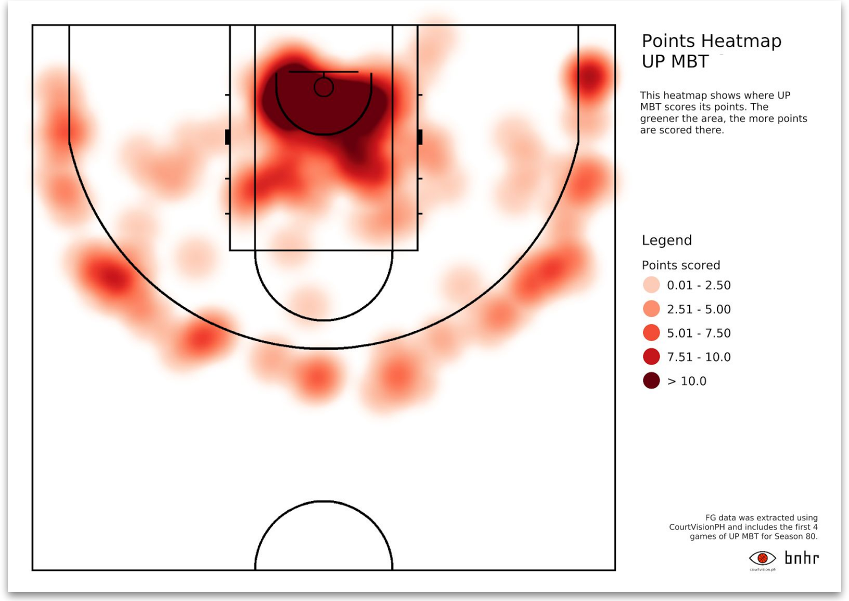
manuscript:
https://github.com/benhur07b/bs-thesis-courtvisionph/blob/main/CourtVisionPH-manuscript-PINTOR-CATANIAG-UPDGE-2014.pdf

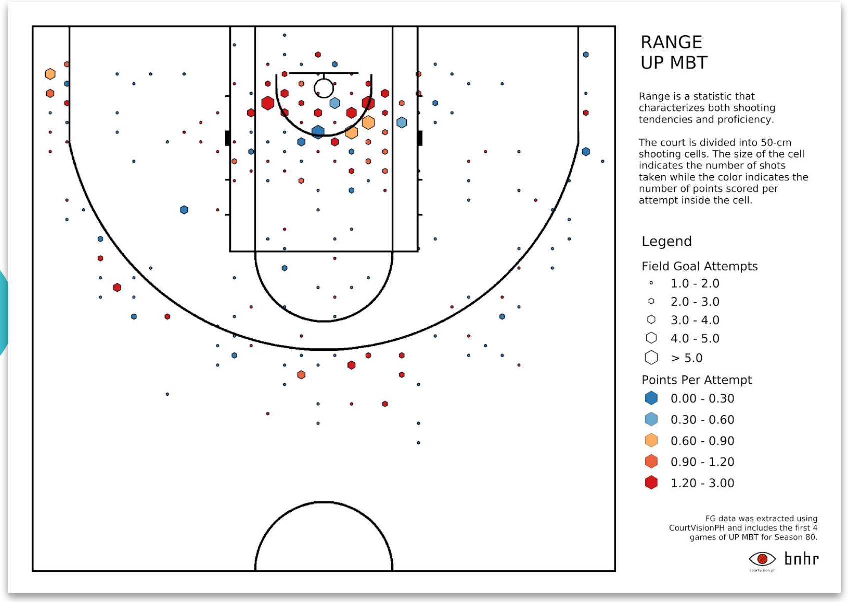
FOSS4G Seoul 2015

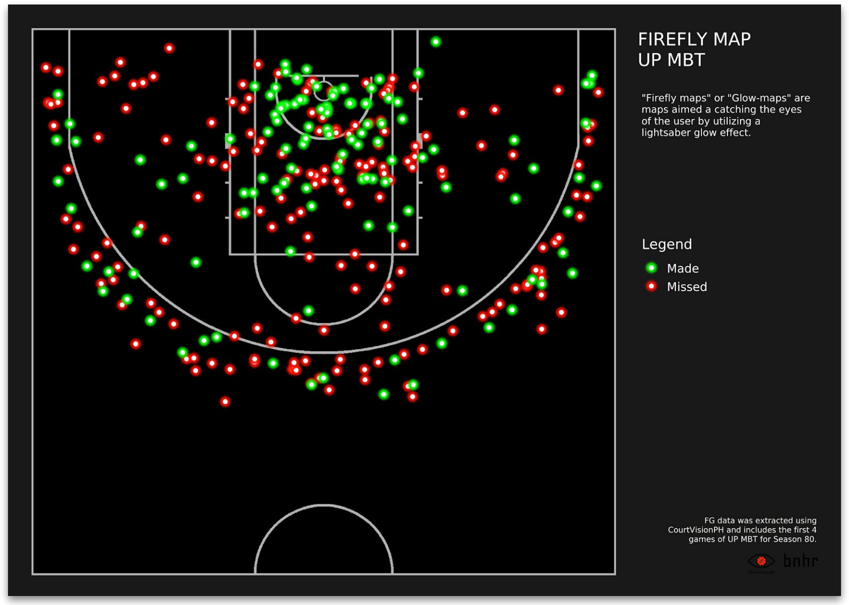
Combining basketball with spatial...











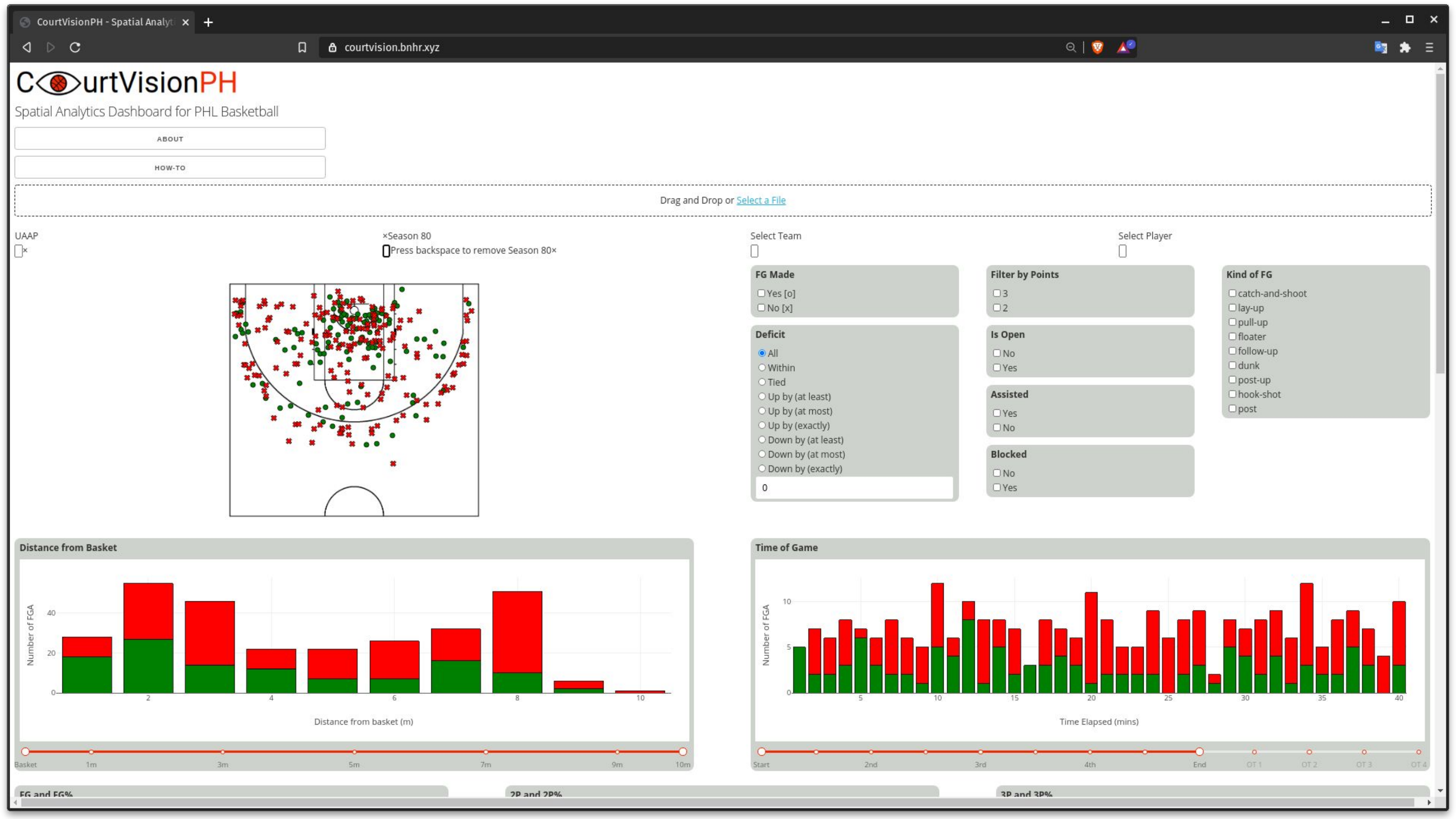
Exploring Spatial Visualizations of Shooting in Basketball using GIS

Final Project, GmE 210 - Spatial Visualizations (2018)

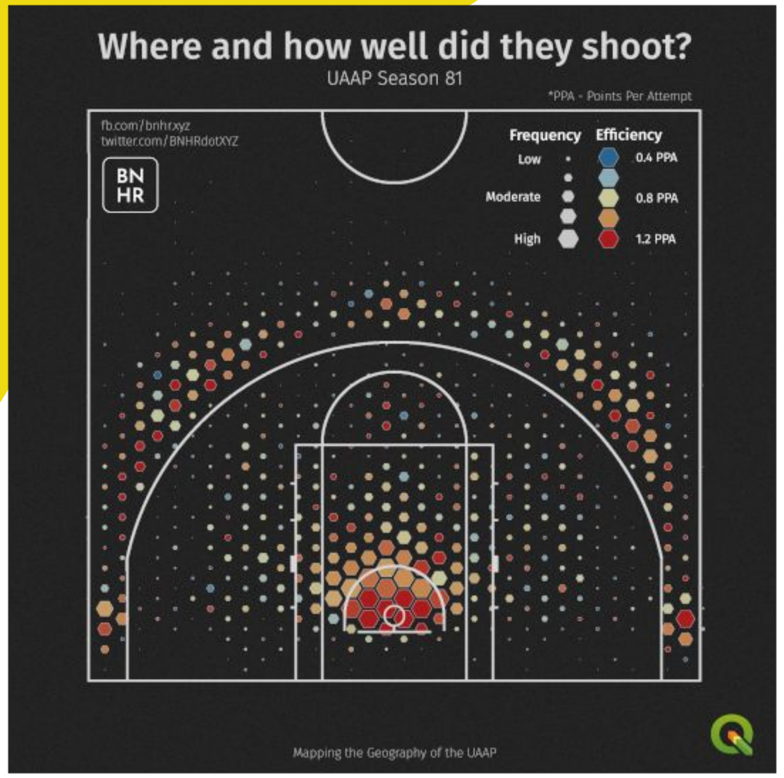
https://bnhr.xyz/2018/05/10/exploring-spatial-visualization-of-shooting-basketball-using-gis.html

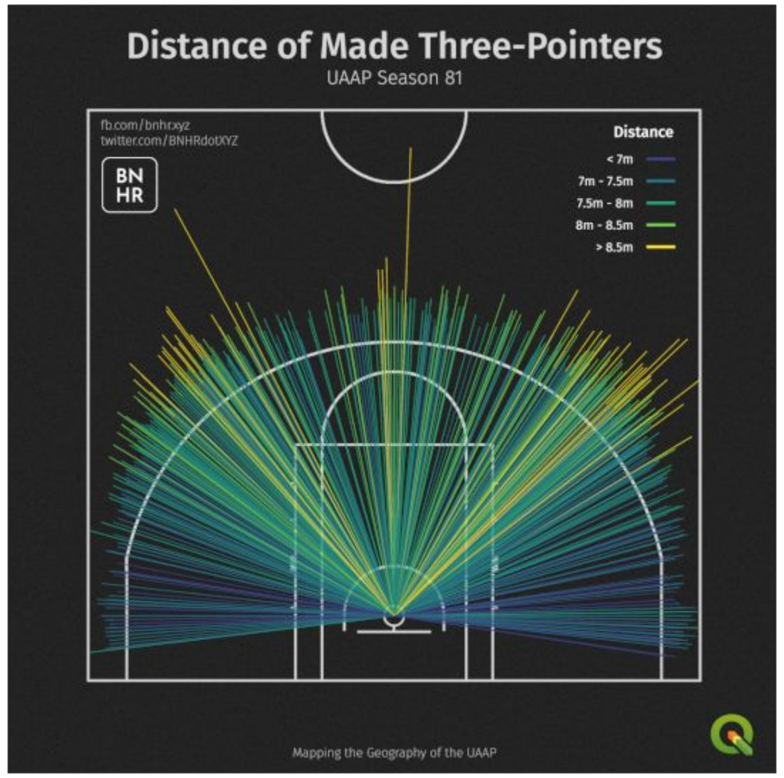
CourtVisionPH Dashboard

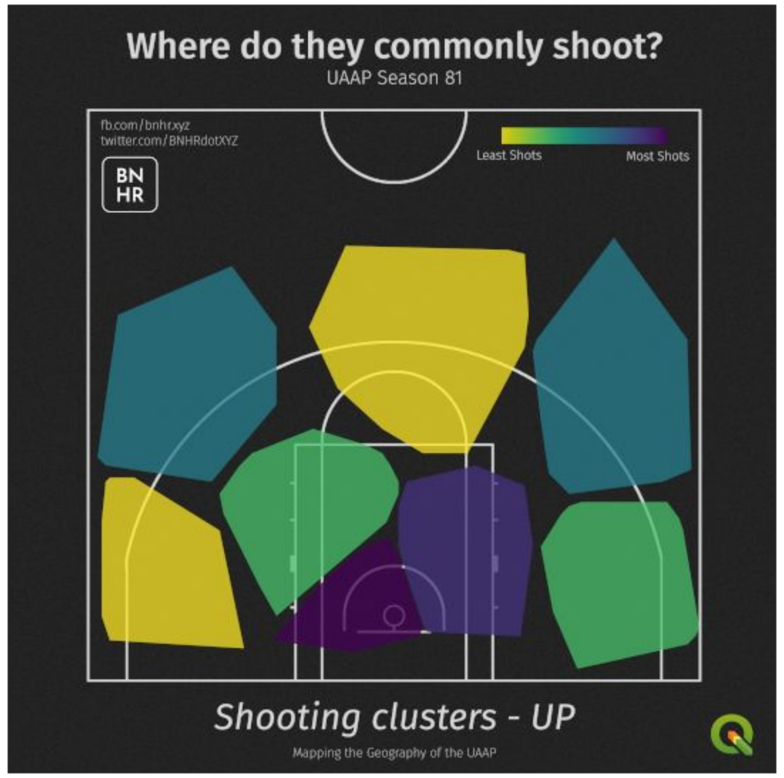
Final Project, GmE 222 - Advanced GIS (2018)

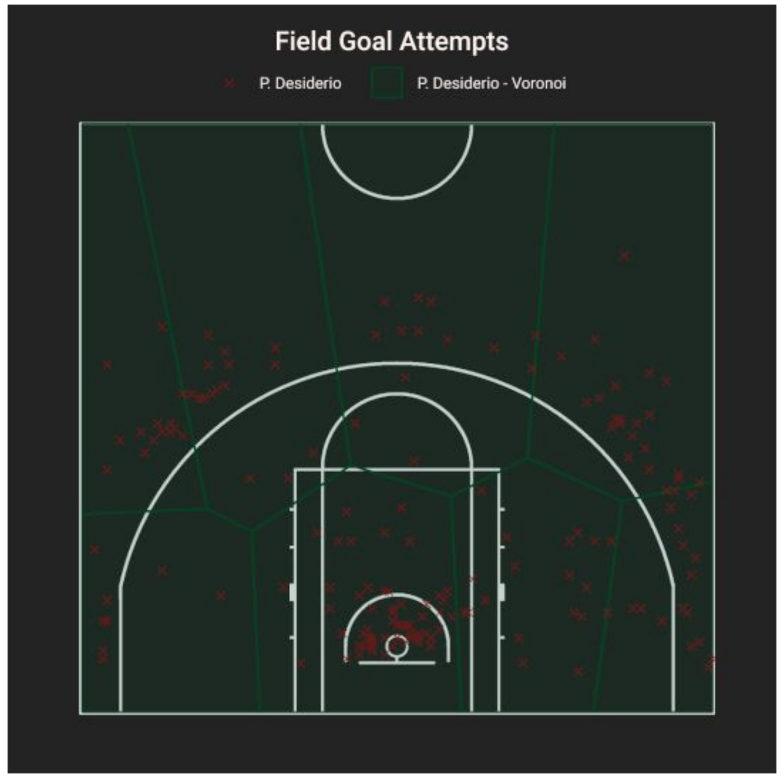


courtvision.bnhr.xyz









Qasketball

Spatial analysis and visualization of basketball with QGIS

video: https://www.youtube.com/watch?v=HILlJz9sghk

Which brings me to...

Towards a spatial analysis of shooting in Philippine basketball

Applications in the University Athletics Association of the Philippines (Season 81)

BN
HR

openness. data. geospatial.

@bnhr.xyz @BNHRdotXYZ BNHRdotXYZ https://bnhr.xyz

FOSS4G BUENOS AIRES 2021

What do I want to accomplish?

- Find court divisions empirically using the field goal dataset.
- Find similar players based on their shooting habits at different areas on the court.
- Compute for spatially-aware shooting metrics for comparing and analysing shooting performance.

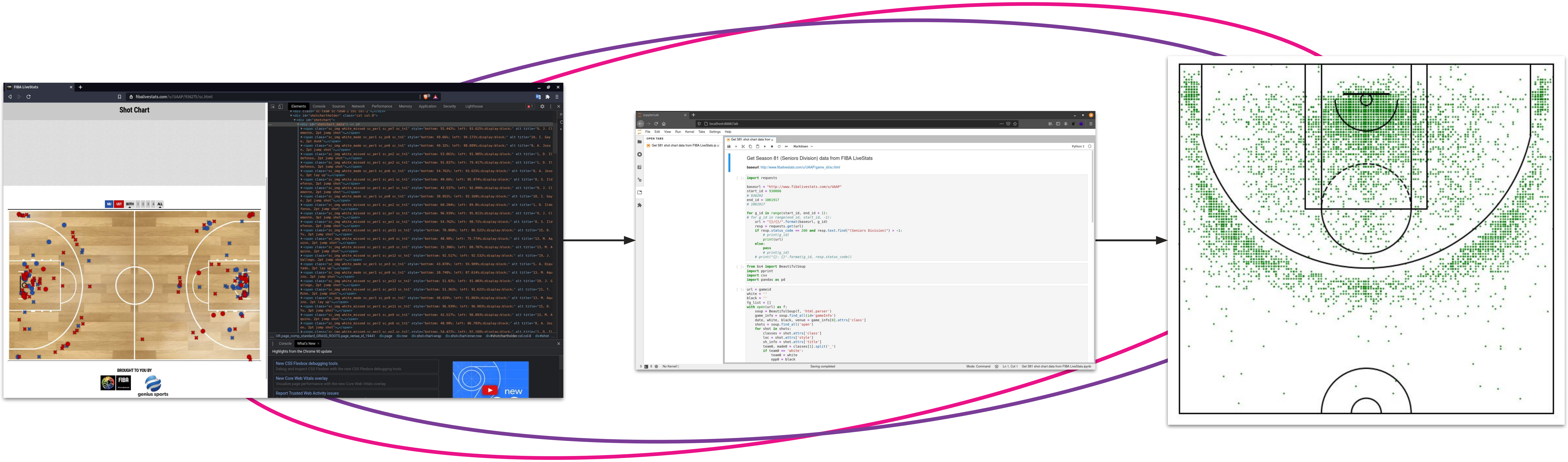
What do I need?

- Field goal dataset with location.

The problem? No such dataset exists for the Philippines (that is easily downloadable).

The solution? Build a scraper of FIBA LiveStats data.

FIBA LiveStats scraper

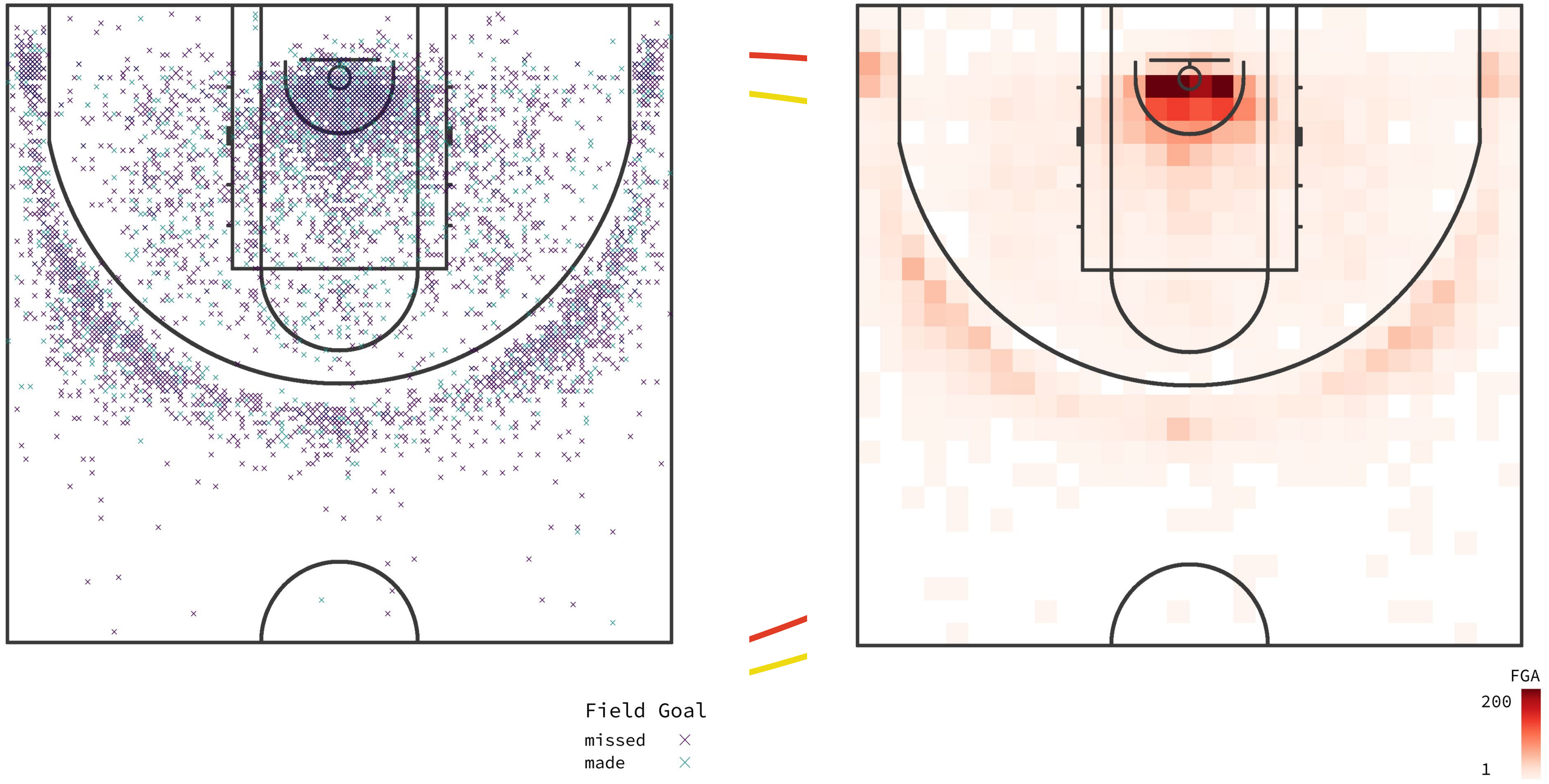


Data

7619 FGA

55 games (1 missing game)

120+ players



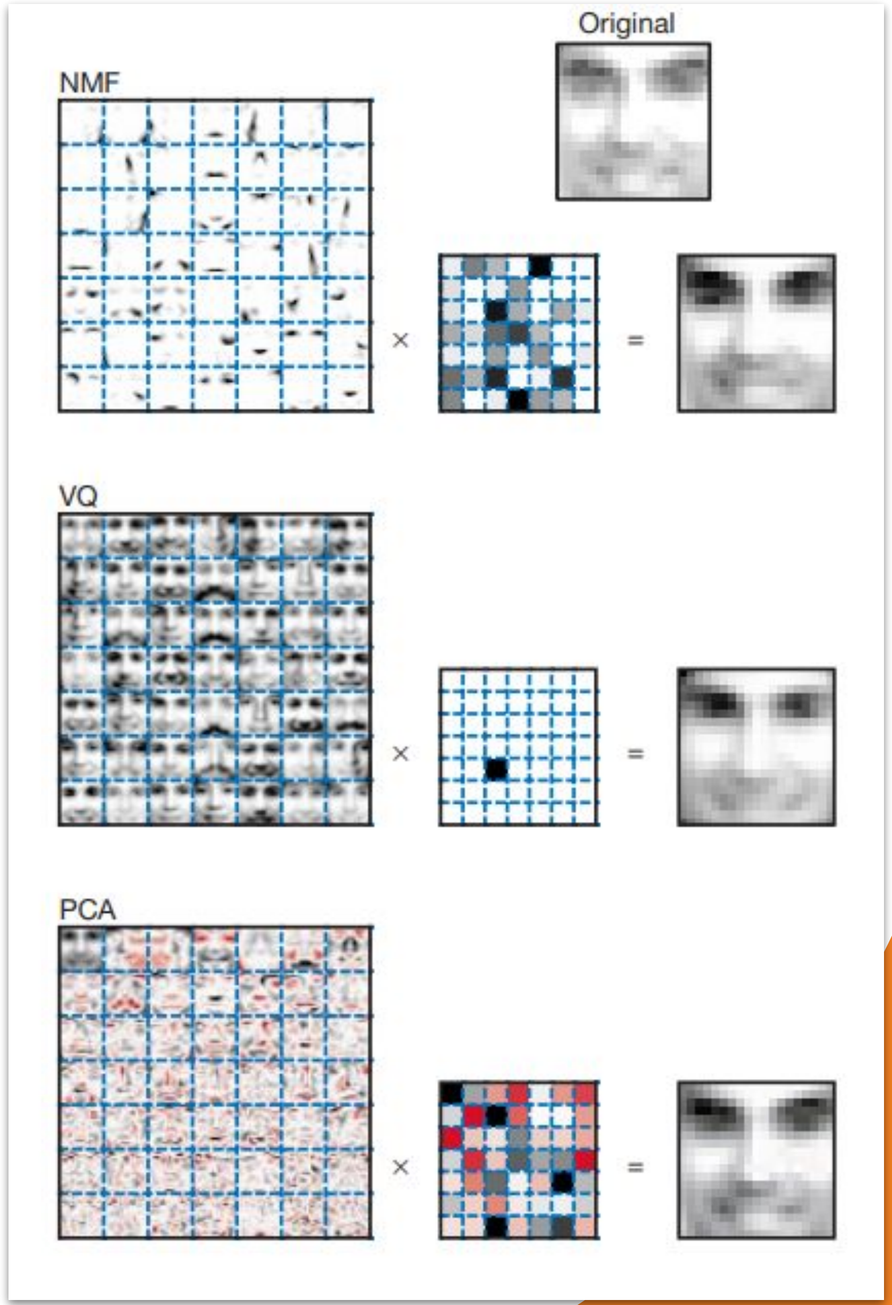
Finding “shot types” or “shooting zones”

What could I do?

- Clustering algorithms
  - K-means
- Matrix decomposition algorithms
  - PCA
  - SVD
  - **NMF**

NMF or Non-negative Matrix Factorization

- **V = WH**
  - V, W, and H are **non-negative** and W, H are lower rank than V
  - Resulting components exhibit a **“parts-based” decomposition** of the original matrix
  - Sparser and more interpretable (i.e. they correspond to **frequently occurring patterns** in the sample)
  - Useful in learning **parts of objects**



Basis vectors for faces using NMF, VQ, and PCA (Lee & Seung, 1999)

NMF makes sense for basketball

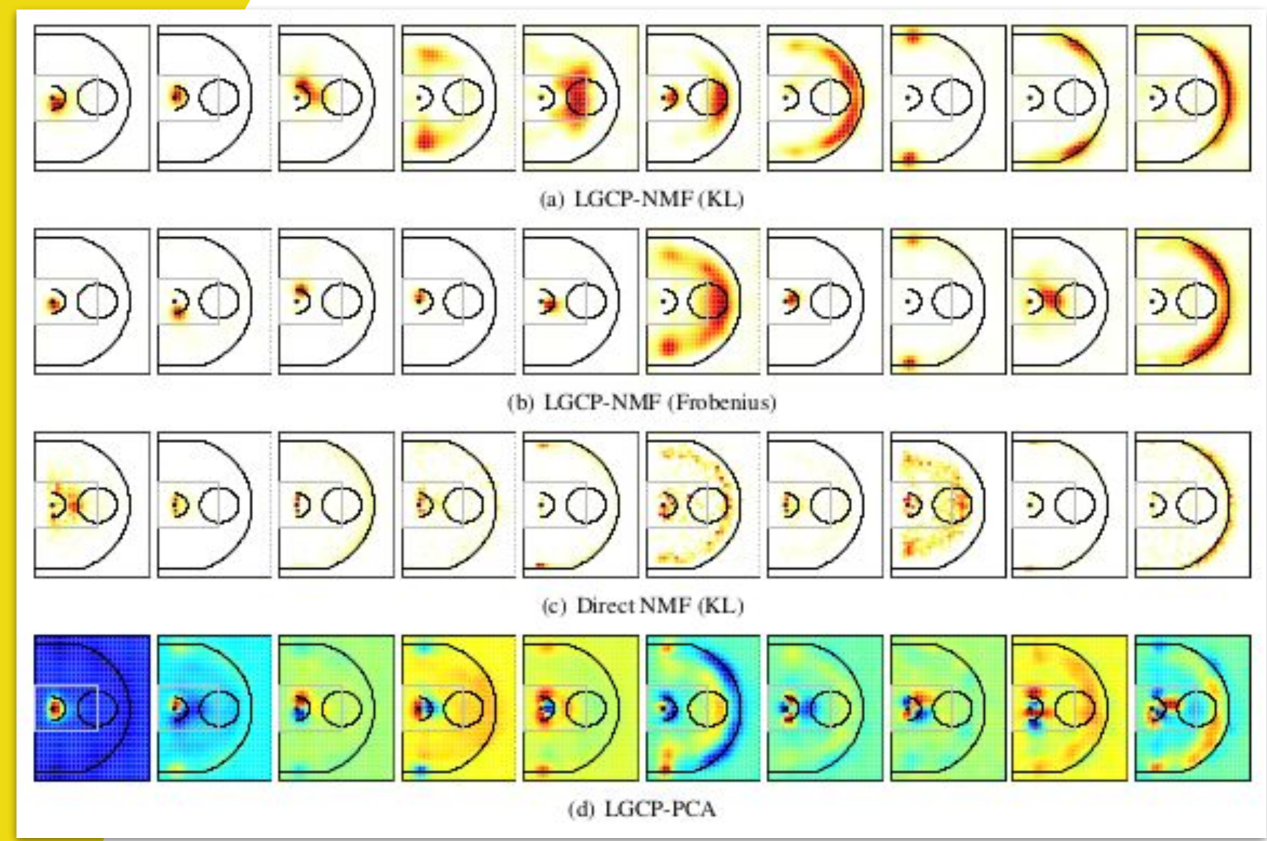
Why it makes sense:

1. the intensity surface of **field goals is always non-negative** because you can't take negative number of field goals in a location
2. the output matrices of the decomposition intuitively correspond to basketball concepts:
   a. **H - represents shooting zones / areas**
   b. **W - represents shooting tendency / frequency**

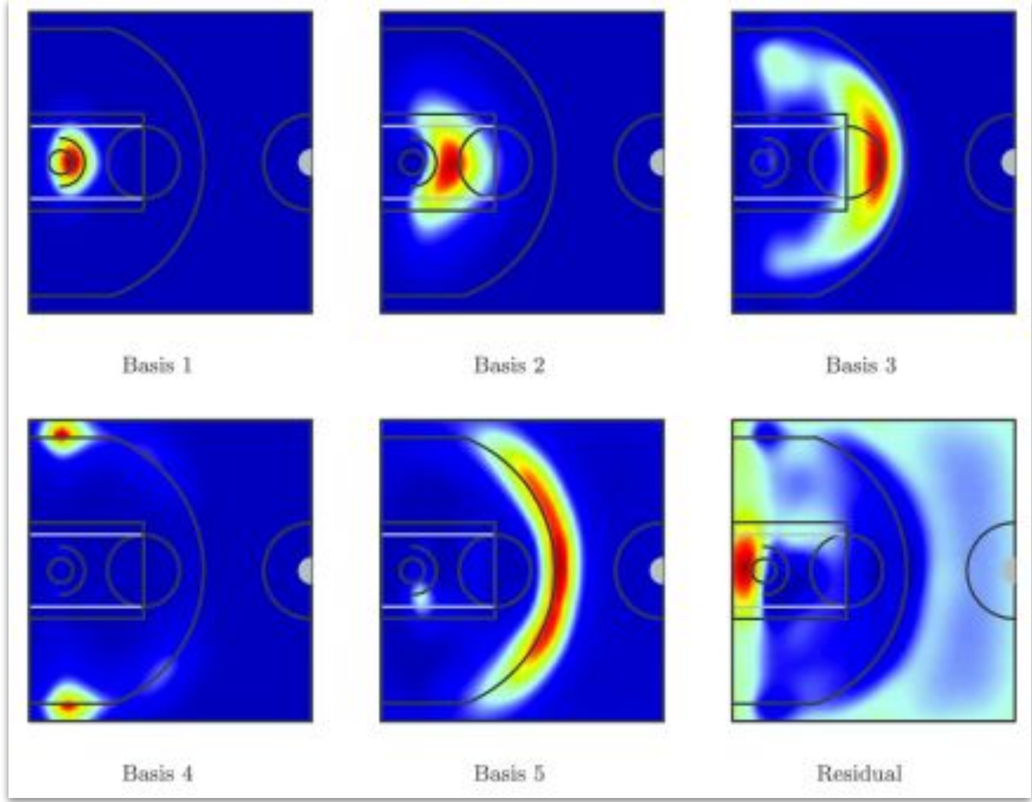
How it's done:

1. **Discretize the basketball court** using a regular tessellation (shooting cells)
2. For each player, **fit an intensity surface** of his field goals on the discretized court
3. **Generate the field goal matrix V** using the intensity surfaces of the players
   a. **Each row** of the matrix is an **array of a player's intensity surface**
   b. **Each column** in the player array corresponds to **one shooting cell**
4. Use NMF to find W, H such that **V ~ WH**.

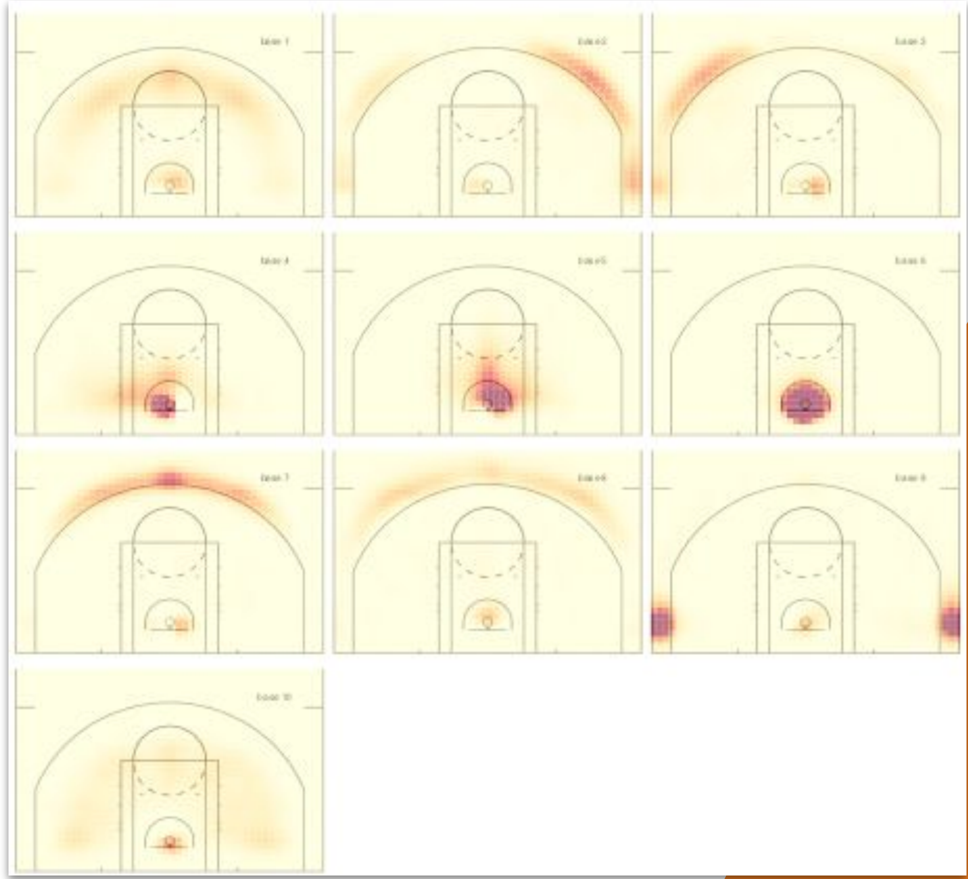
NMF makes sense for basketball



Spatial basis vectors of field goals using LGCP-NMF with B=10 using the Kullback-Leibler (KL), Frobenius loss, direct NMF, and LGCP-PCA. (Miller, 2014)

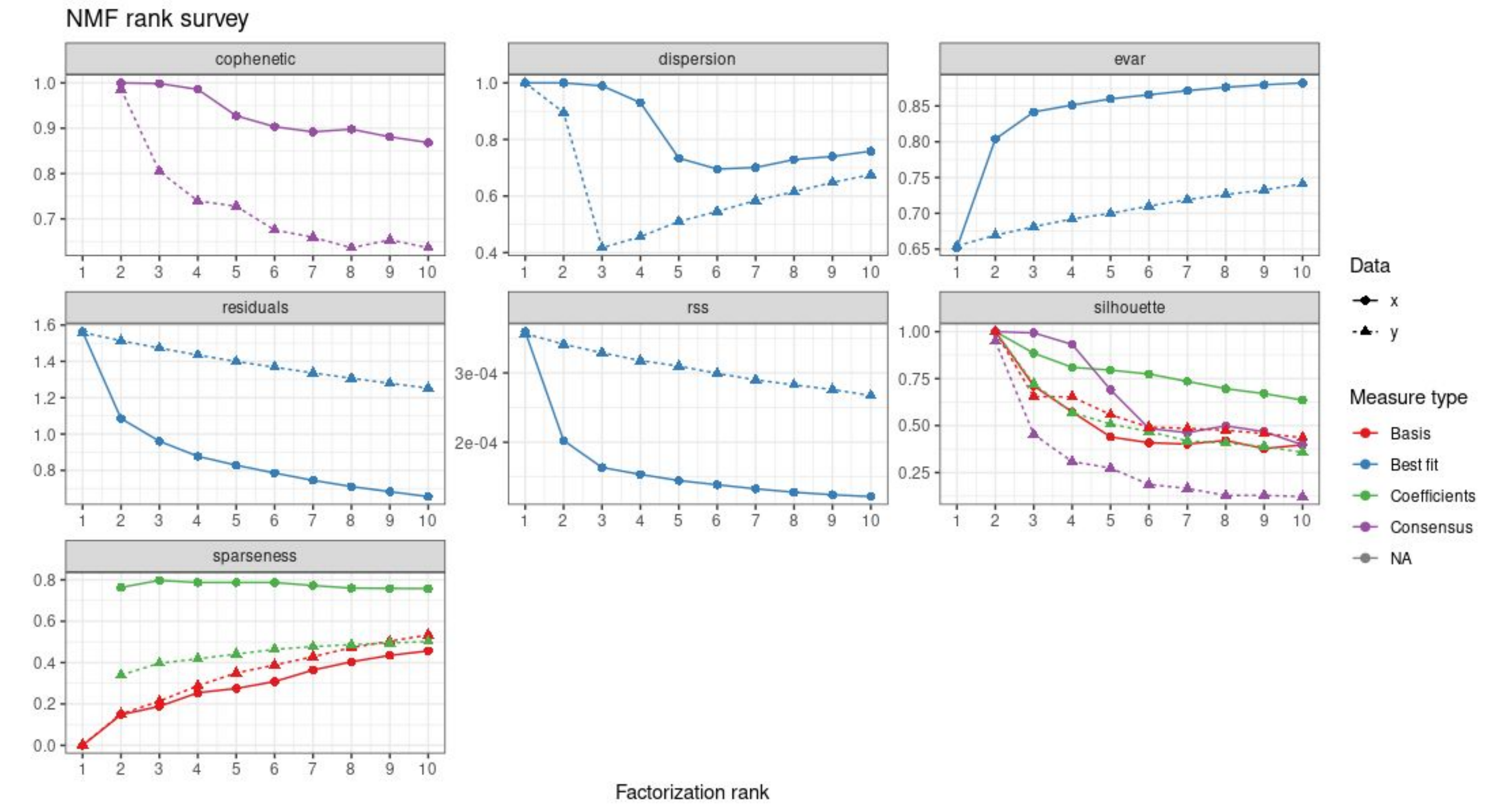


Spatial basis vectors identified by LGCP-NMF for B=6. (Franks, 2015)



Spatial basis vectors of field goals using kernel estimation and NMF (Jiao, 2020)

Finding the optimal rank (number of basis)

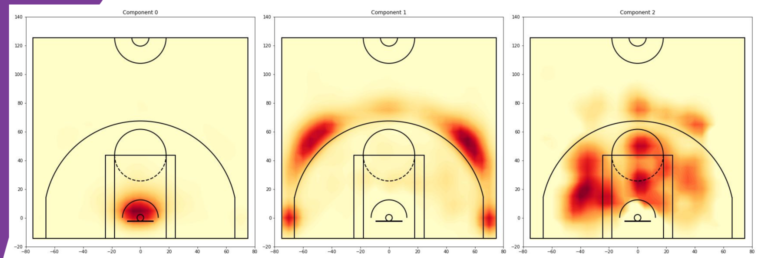


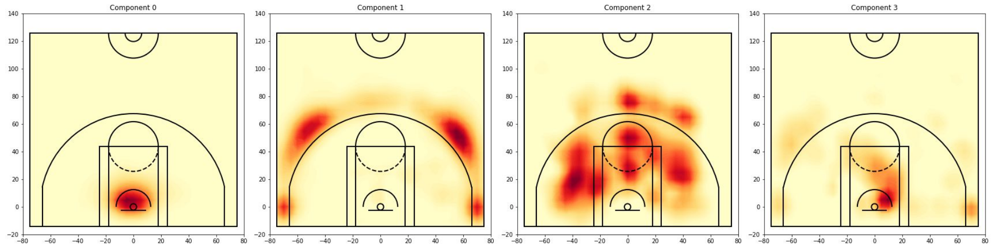
Three methods for estimating the optimal value of K (number of basis) for NMF from the application of NMF on omic (e.g. genomic) data by looking at different NMF rank measures

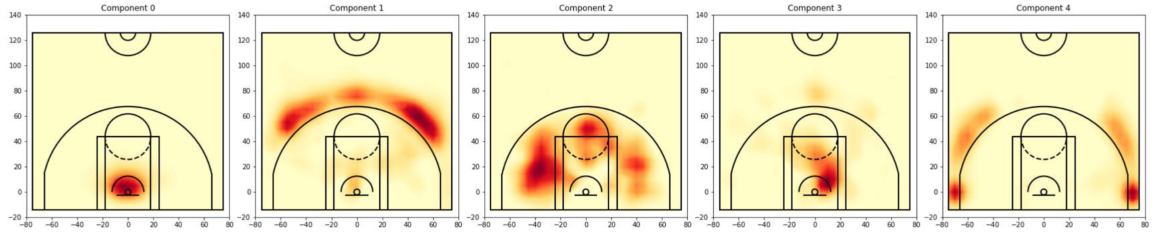
- Brunet (2004) - the first value of K for which the cophenetic coefficient starts decreasing
- Hutchins (2008) - the first inflection point of the residual sum of squares (rss) curve
- Frigyesi (2008) - the smallest value of K where the decrease in RSS computed from the actual data is lower than the decrease in RSS computed from random/permuted data

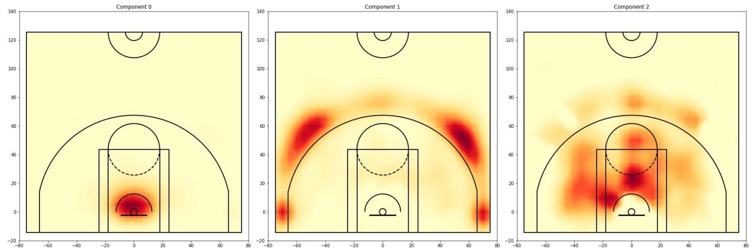
Based on the graphs and rank measures, the optimal number of bases in the field goal dataset is **between 3-5**.

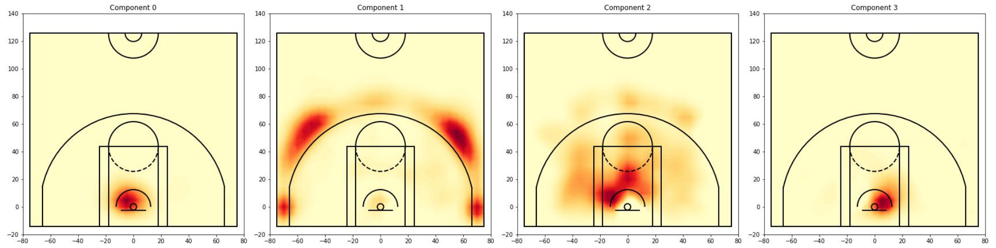
Finding the optimal rank (number of basis)

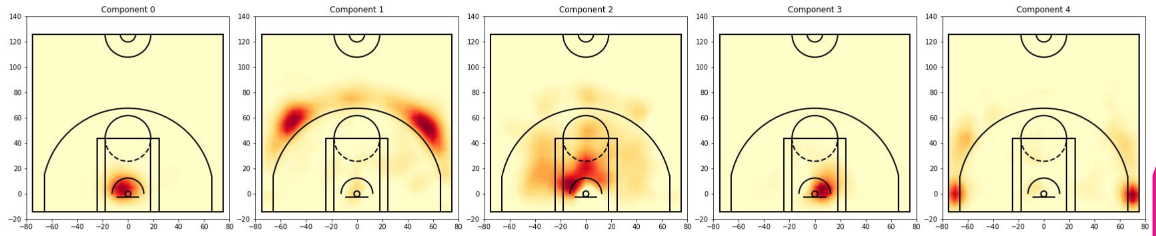


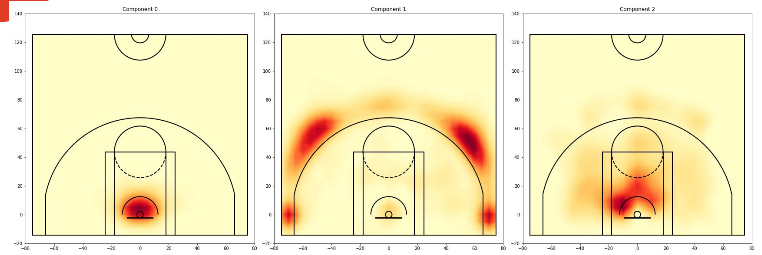


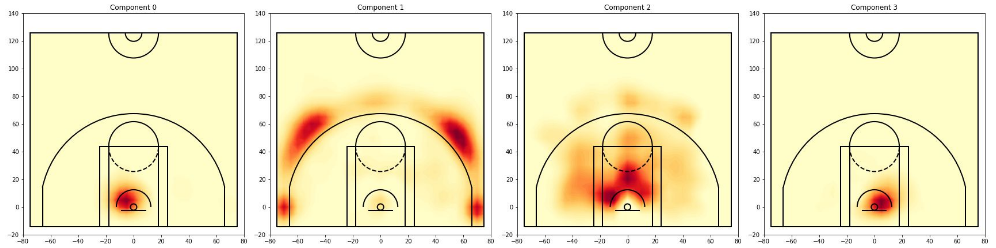


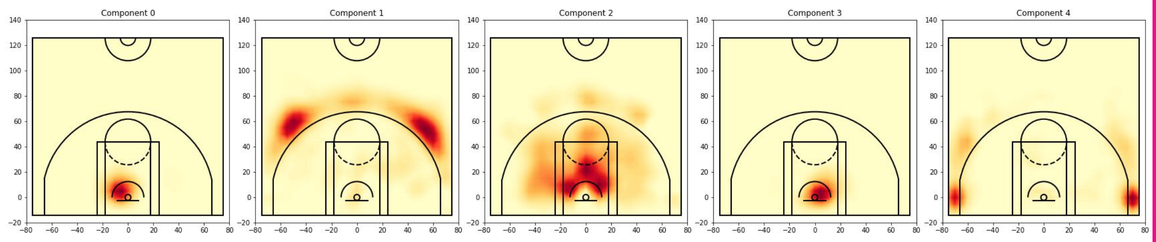




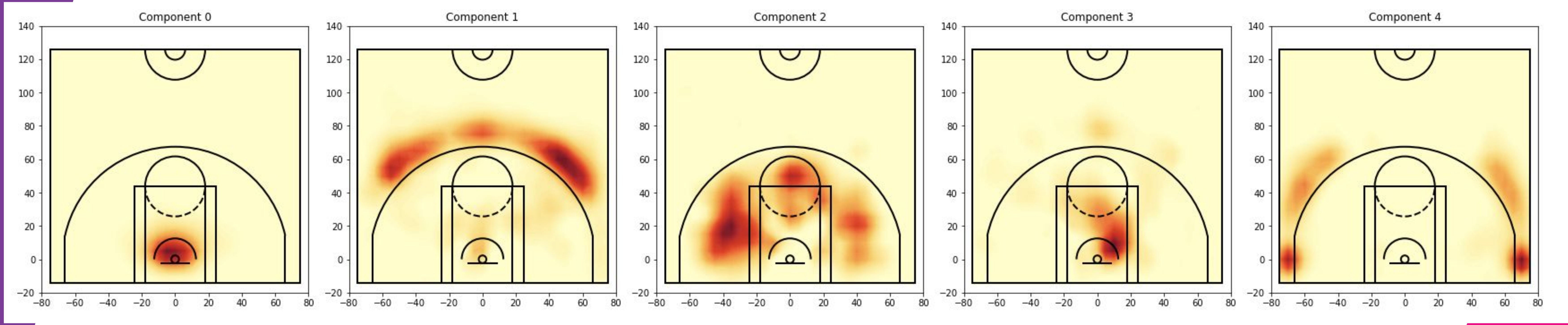






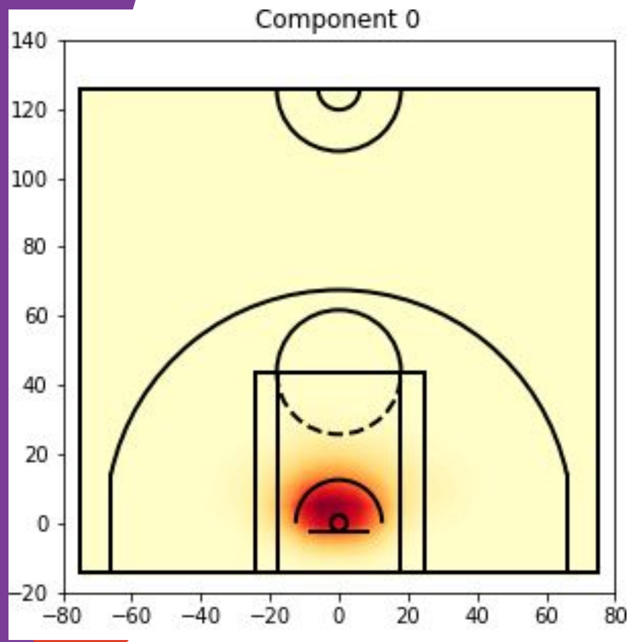


Spatial basis vectors of FG using NMF

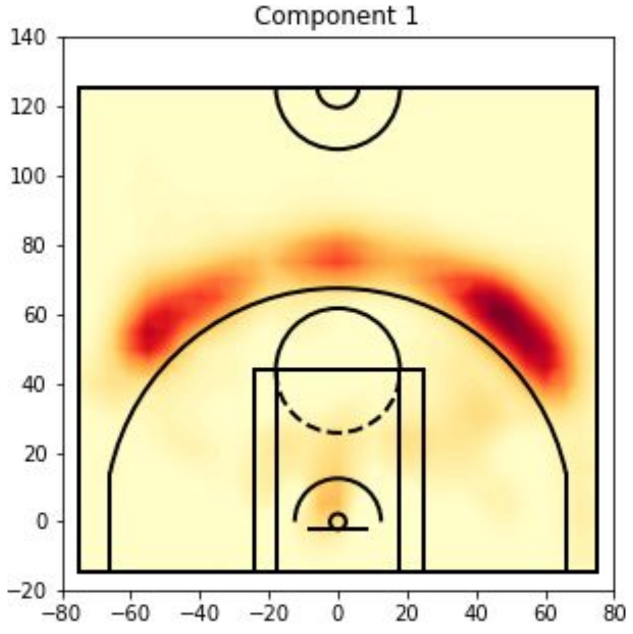


Kullback-Leibler, K=5, solver=multiplicative update, init=nddsvda (Nonnegative Double Singular Value Decomposition (NNDSVD) initialization with the average of X instead of zeros)

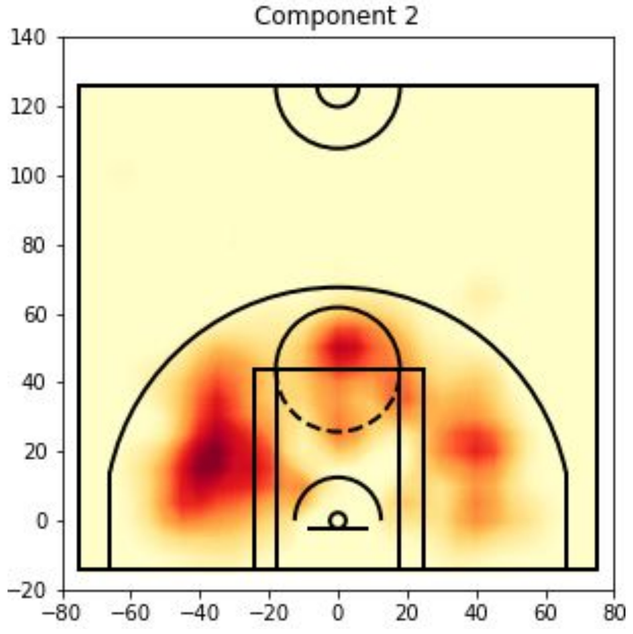
“Shot types” or “Shooting zones” per basis



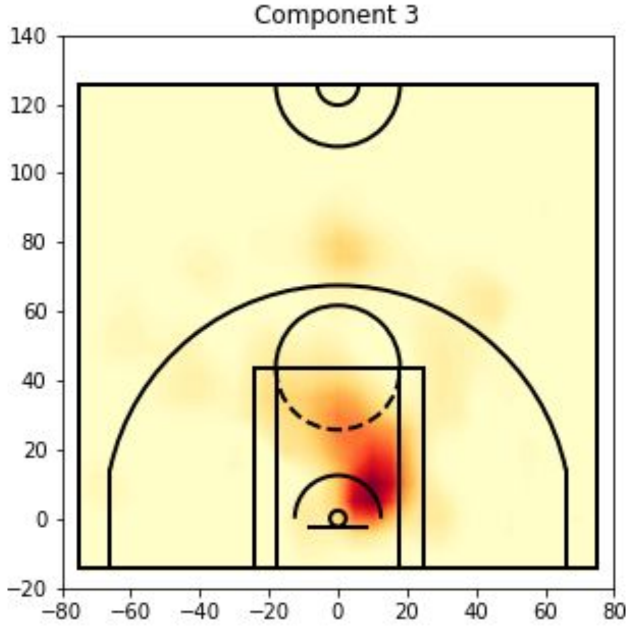
paint



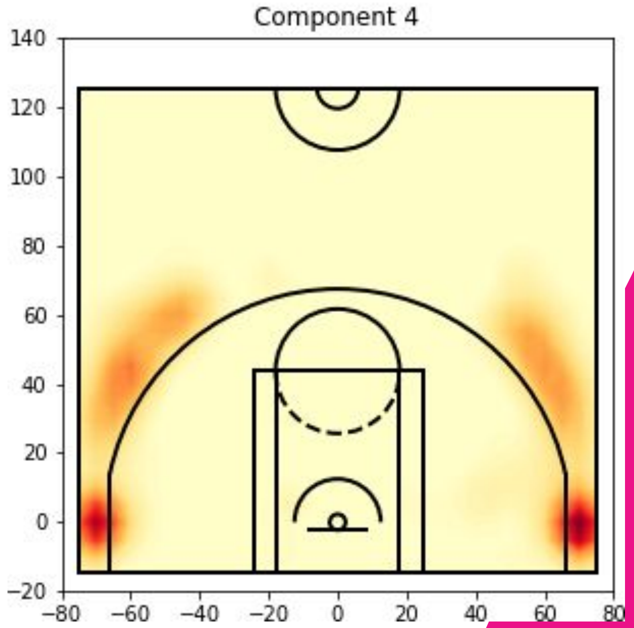
wing + key three-pointers



mid-range + some paint



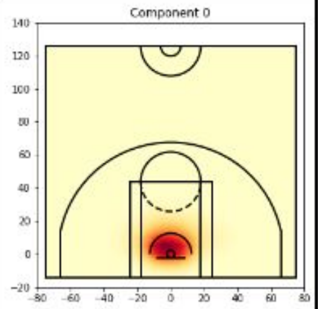
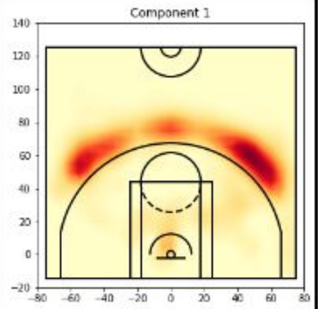
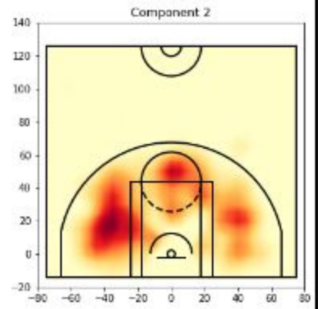
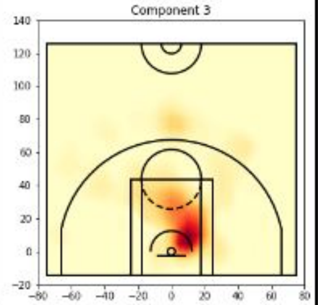
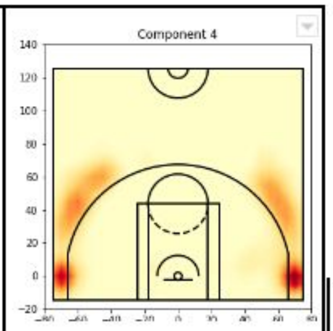
left block + some paint



corner + some wing three-pointers

Player shooting habits

Player shooting habits (basis weights)

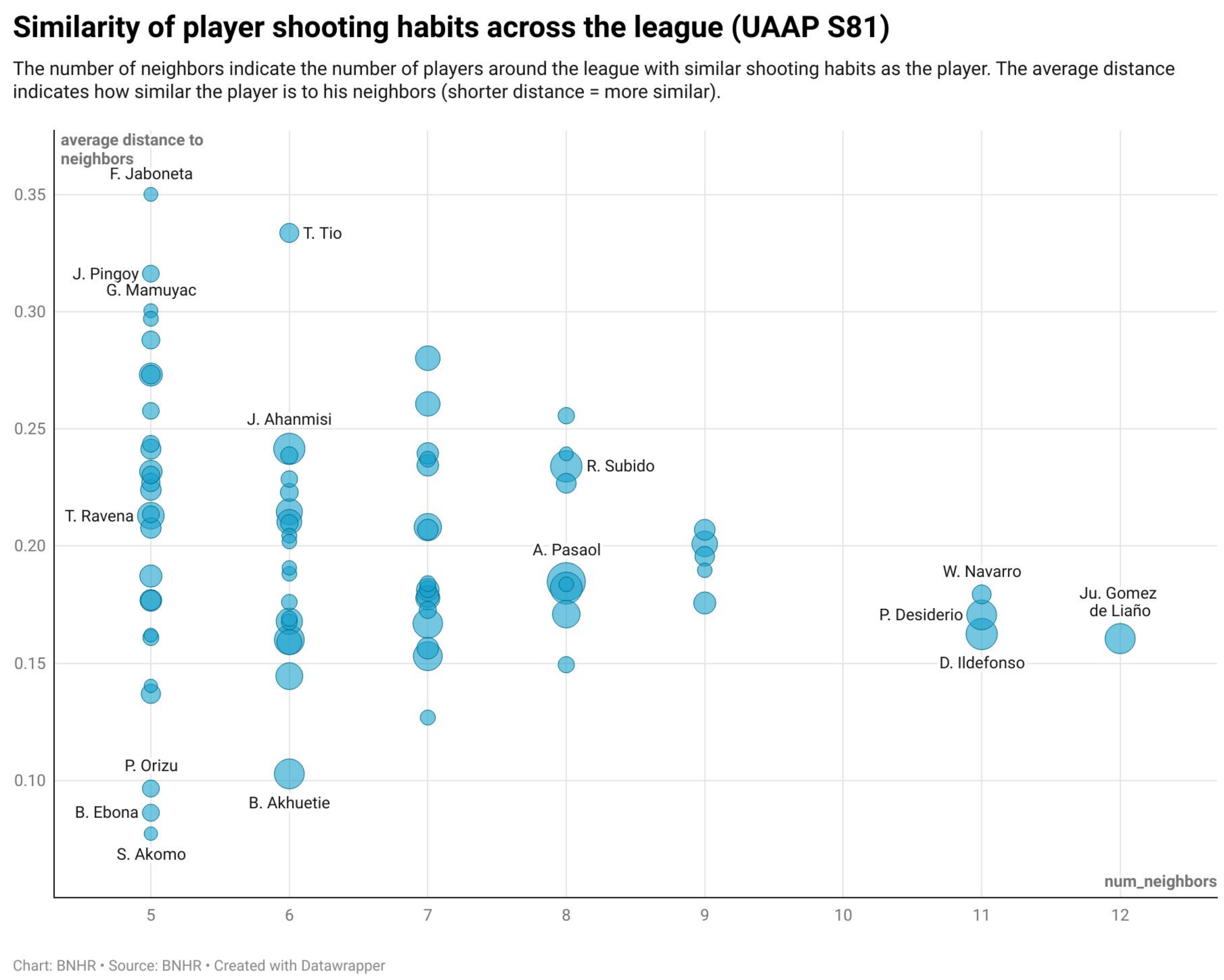
					
A. Pasaol	**0.47**	**0.27**	0.11	0.00	0.14
A. Melecio	0.18	**0.27**	**0.20**	**0.12**	**0.22**
R. Subido	0.08	**0.41**	0.12	0.10	**0.29**
D. Ildefonso	0.31	**0.26**	0.11	0.09	**0.22**
J. Ahanmisi	0.04	**0.32**	**0.16**	**0.36**	0.12
Ju. Gomez de Liaño	0.26	**0.21**	**0.21**	0.06	**0.27**
B. Akhuetie	**0.94**	0.00	0.03	0.00	0.03
P. Desiderio	0.29	**0.26**	0.07	0.11	**0.28**
M. Lee	0.05	**0.25**	0.07	0.06	**0.57**
S. Manganti	0.27	0.15	0.14	**0.19**	**0.25**

Finding similar players

What I did...

- Players with similar shooting habits were determined by computing the Euclidean distance of the player basis weights.
- Using this distance, a player's five nearest neighbors were identified.
- If player k is the neighbor of player l, it is assumed that player l is also a neighbor of player k to enforce symmetry.

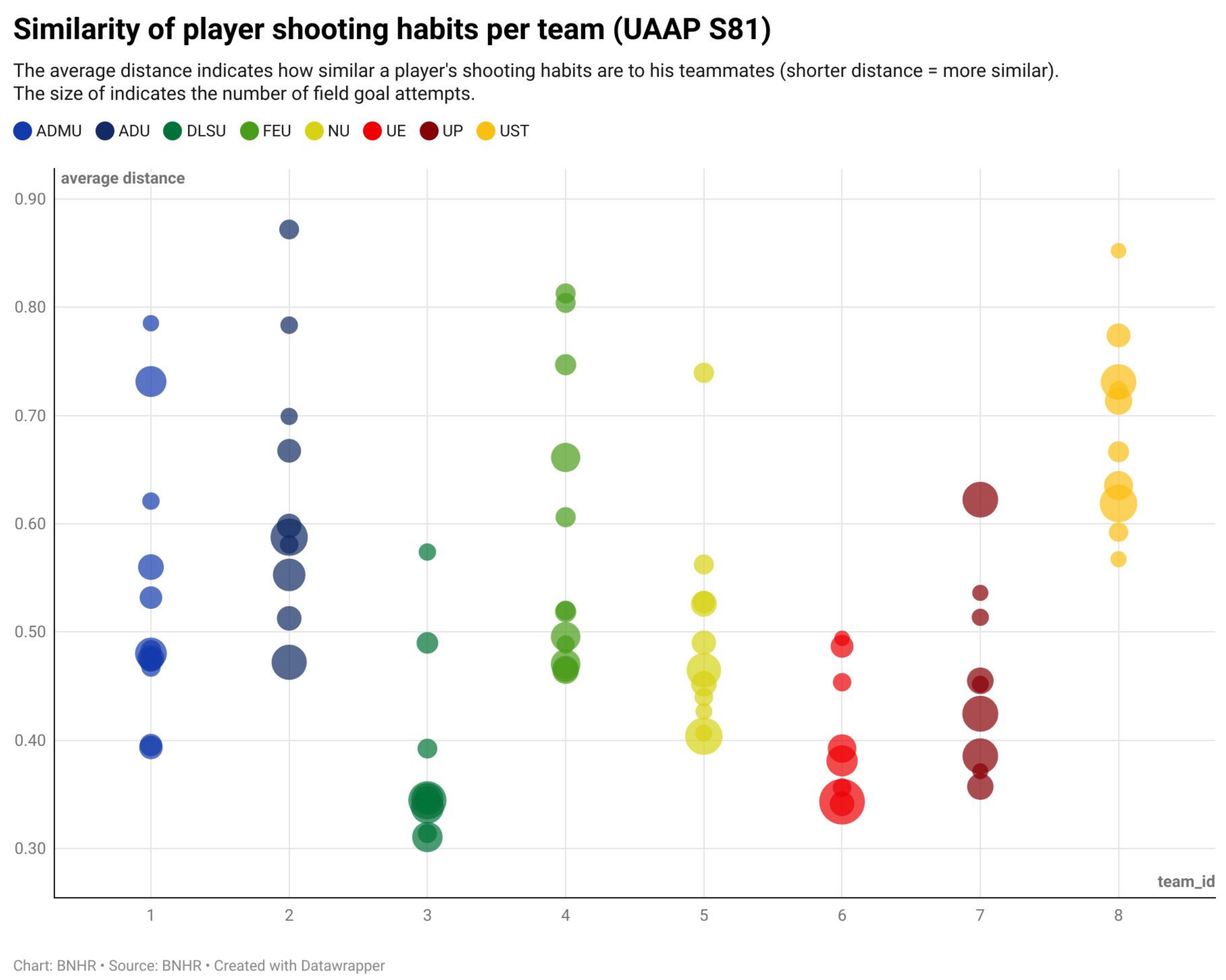
What I found...



Two groups of players that are interesting.

- Group of players with **average distance near or less than 0.1**. This indicates that they have **very similar shooting habits to their neighbors**. What's characteristic of these players is that t**hey almost exclusively take shots from near the basket**. Players from this group include P. Orizu, B. Ebona. S. Akomo, and B. Akhuetie.
- Second group are those whose **average distance to their neighbors is greater than 0.3.** All the players that belong to this group only have five to six neighbors each. This combination of low number of neighbors and long distance between them and their neighbors could indicate that the players belonging to this group have **uncommon shooting habits relative to the rest of the league**. The players in this group are G. Mamuyac, J. Pingoy, F. Jaboneta, and T. Tio.

What I found...



UE Red Warriors and DLSU Green Archers

- In the case of both teams, the distances between the shooting habits of their players with the rest of their teammates are small which means that their players have similar shooting habits and can indicate that both teams run a predictable offense regardless of the combination of players on the court.
- For both teams, the players with the most field goal attempts are also those with the most similar shooting habits to the rest of their team indicated by the large circles at the lower end of their graphs.

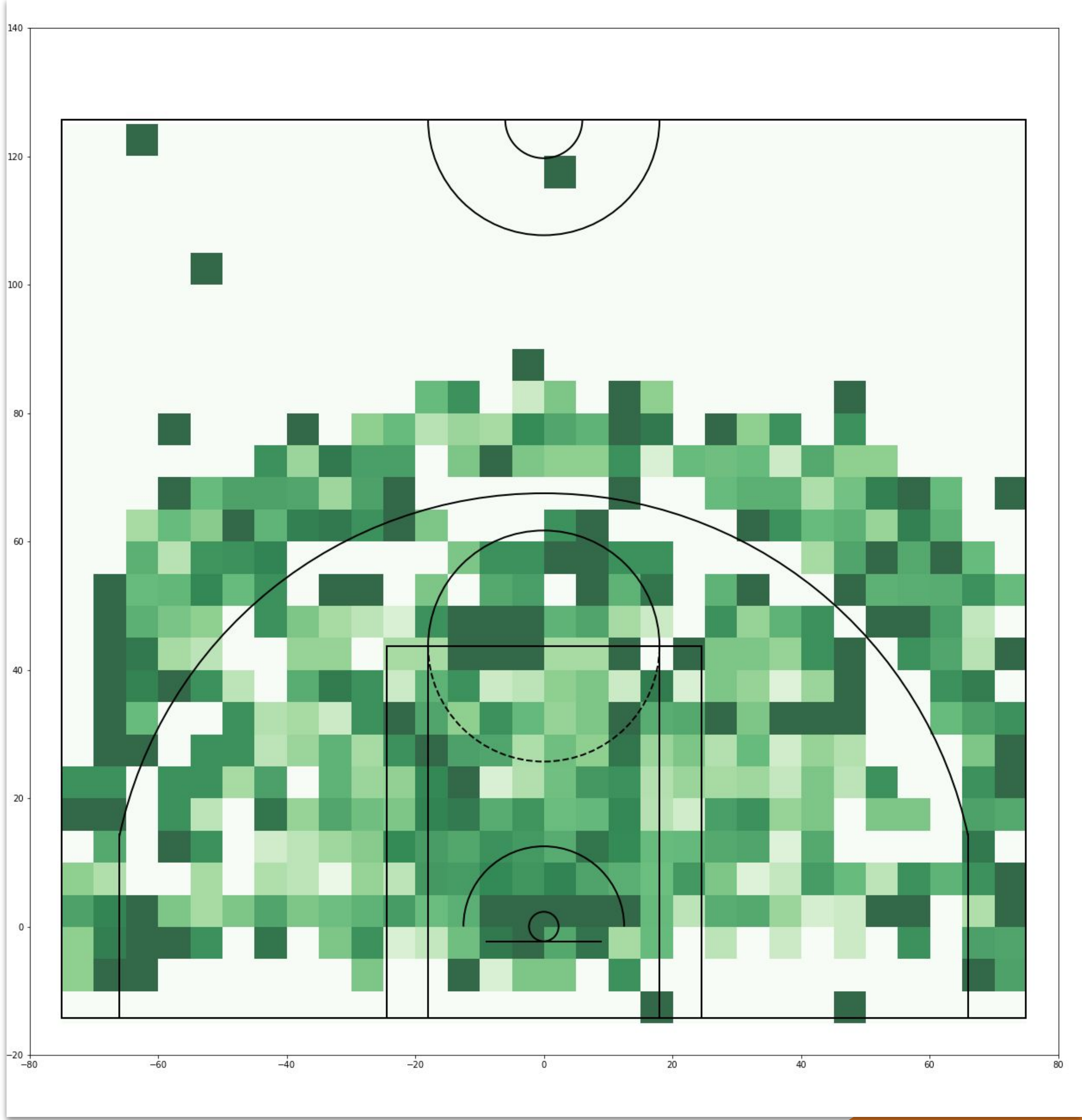
In contrast, the other teams have a significantly larger range of values in terms of the similarities in the shooting habits of their players.

Spatially-aware shooting metrics

Better model of scoring ability

What's the problem?

1. Divide the court into 50cmx50cm cells.
2. Not all cells on the court have field goal attempts.
3. Some cells have low field goal attempts that may not properly indicate the actual scoring rate at that location.
4. This results in a peppered and noisy Points Per Attempt map.



What's the solution?

1. Compute for the Empirical Bayes estimate of PPA at each cell.
2. Prior distribution at cell i includes:
   a. nearby cells within distance j from cell n
   b. cells that are equidistant (within k from distance d) to the basket
   c. increasing values of j and k for cells beyond 7 meters from the basket

$$PPA_i = \frac{PTS_i}{FGA_i} \qquad (1)$$

Points Per Attempt
PTS = points scored at cell **i**
FGA = number of FG at cell **i**

$$\hat{\gamma}_i = \frac{\sum PTS_j}{\sum FGA_j} \qquad (2)$$

prior mean for cell i
j are the cells in the local neighborhood

$$\hat{\varphi}_i = \frac{\sum n_j (p_j - \hat{\gamma}_j)^2}{2} - \frac{\hat{\gamma}_i}{\bar{n}_j} \qquad (3)$$

prior variance
$\mathbf{n_j}$ is the number of shots taken in cell j, $\mathbf{p_j}$ is the raw PPA observed in cell j, and $\bar{\mathbf{n}}_j$ is the sample mean of the number of shots taken within all neighborhood cells j

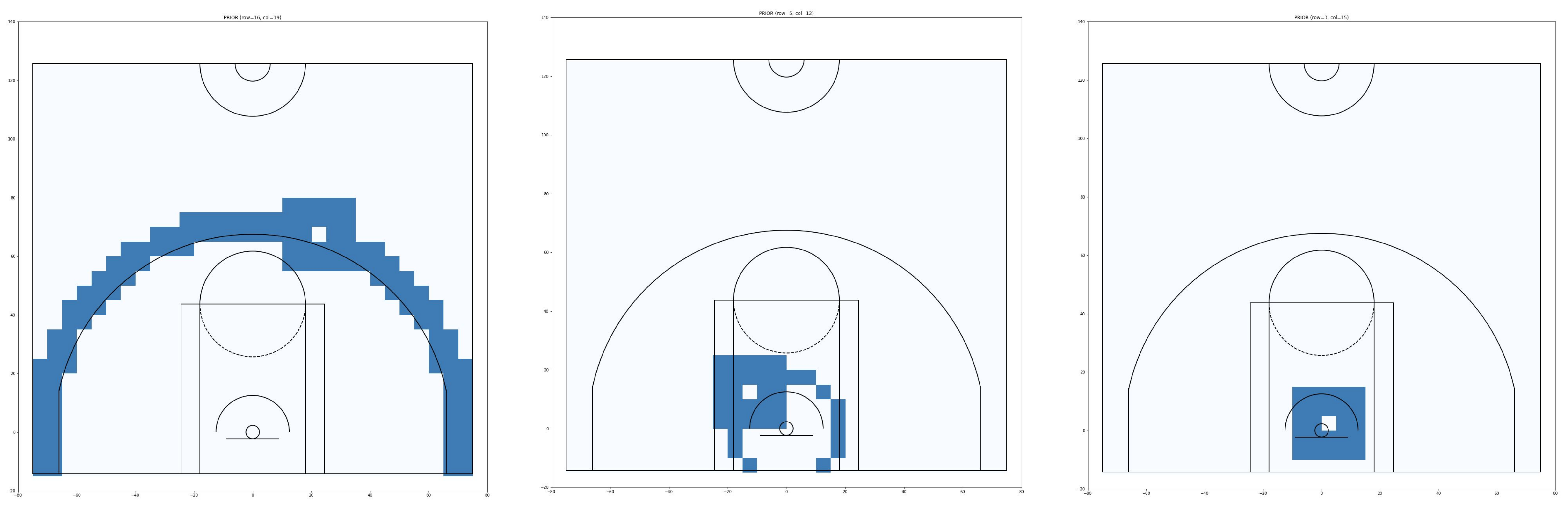
$$\hat{W}_i = \frac{\hat{\varphi}_i}{\hat{\varphi}_i + \hat{\gamma}_i / n_i} \qquad (4)$$

weighting factor
$\mathbf{n_i}$ is the number of shot attempts from i

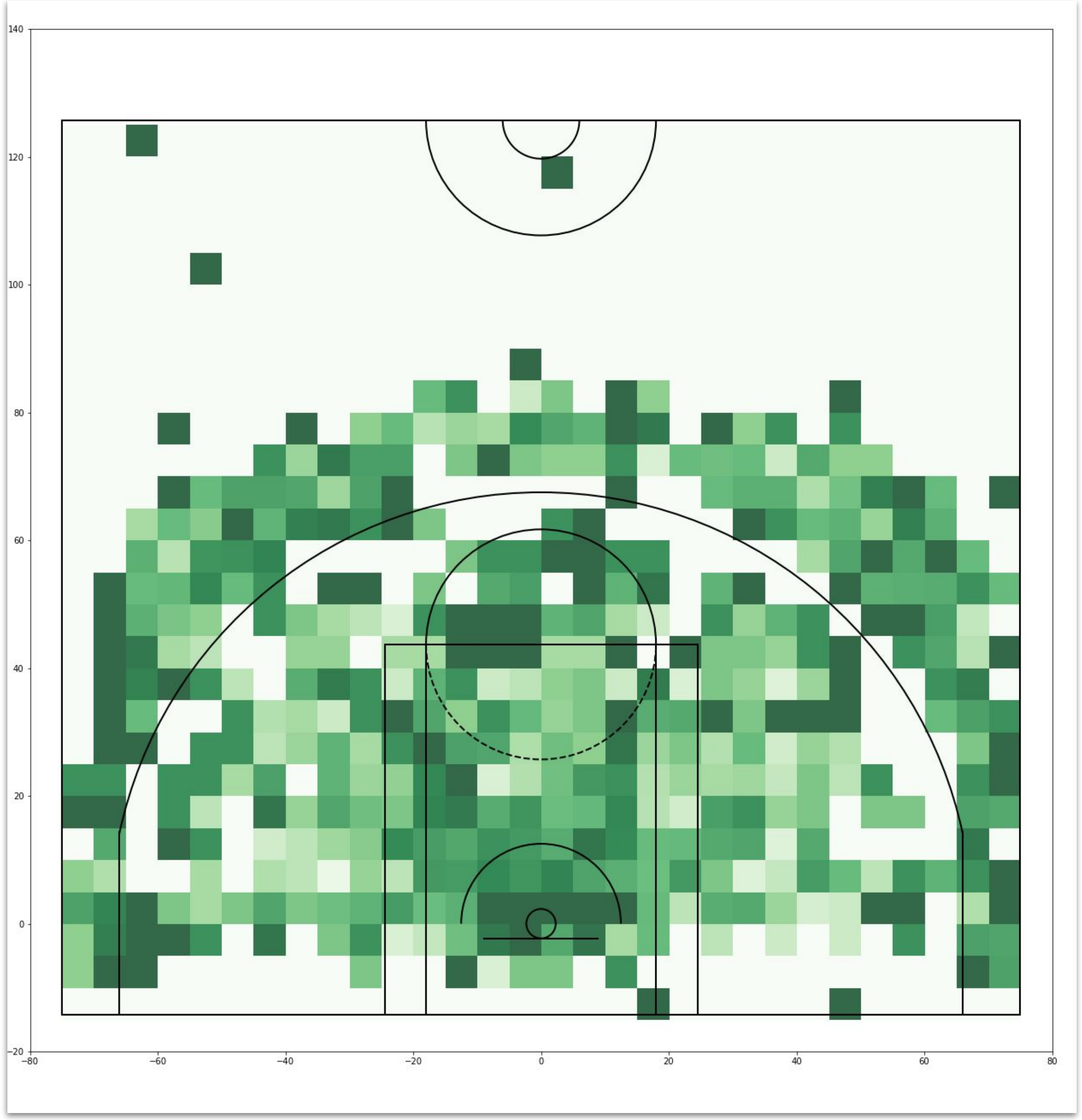
$$\hat{\theta}_i = \hat{W}_i p_i + (1 - \hat{W}_i)\hat{\gamma}_i \qquad (5)$$

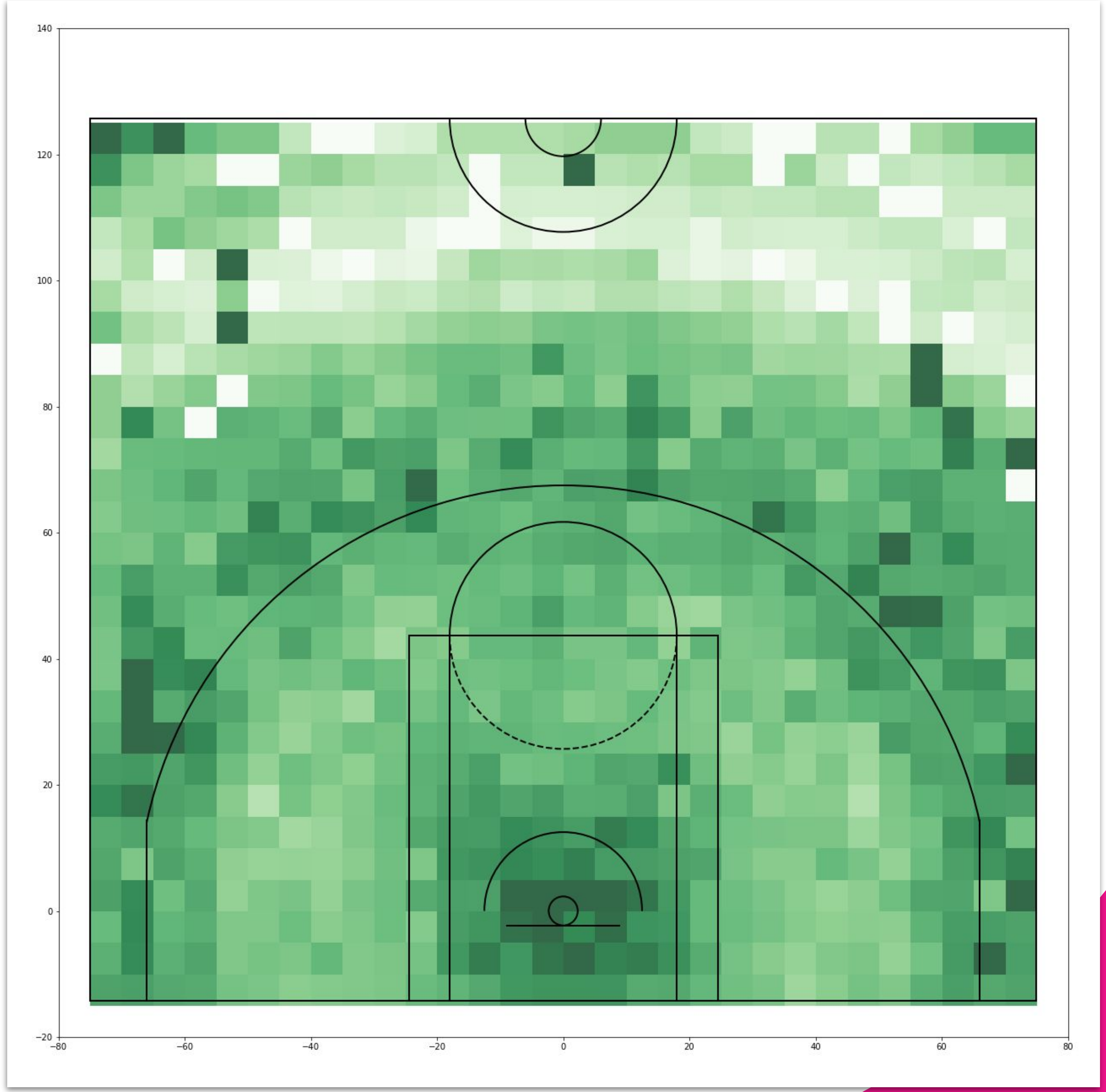
Empirical Bayes estimate of the PPA
weighted combination of the local raw PPA $\mathbf{p_i}$ and the Bayesian local average $\hat{\boldsymbol{\gamma}}_i$

Neighborhood/prior cells of sample cells



Empirical Bayes estimate of PPA





What’s next?

Spatial Scoring Effectiveness (SScE)

Spatial Scoring Effectiveness (SScE)

1. Similar to the Spatial Shooting Effectiveness metric introduced by Shortridge (2014).
2. A measure of scoring effectiveness based on the spatial distribution of field goals defined as the difference between the player's points per attempt (PPA) and his expected points per attempt (EPPA) based on the EB-estimated PPA.

Spatial Scoring Effectiveness (SScE)

1. Get a player's field goal locations.
2. Compute for the player's Expected Points Per Attempt (EPPA) based on his field goal locations and the EB-estimate PPA at each location.
   a. *EPPA can be used as a summary metric indicating if a player takes shots from areas that are easy or difficult to score (low EPPA = difficult shots).*
3. Compute for the player's actual PPA.
   a. for Global SScE: total PPA
   b. for local SScE: PPA per cell/location
4. **SScE = PPA - EPPA; local SScE = PPA at cell - EPPA at cell**

Spatial Scoring Effectiveness (SScE)

Global SScE

- Positive values indicate players are scoring more than expected while negative values indicate the opposite.

Local SScE

- SScE computed per cell/location
- Can be mapped to show the spatial distribution of a player's SScE—i.e. **at what areas on the court is a player scoring more or less than expected**.

Points Relative to League Average (PRLA)

Points Relative to League Average (PRLA)

1. Similar to the Points Over League Average metric introduced by Shortridge (2014).
2. A measure of scoring effectiveness based on the spatial distribution of field goals defined as the difference between the player's total points scored (PTS) and expected points (EPTS) based on the EB-estimated PPA.

Points Relative to League Average (PRLA)

1. Get a player's field goal locations.
2. Compute for the player's Expected Points (EPTS) which is the sum of his Expected Local Points (ELPTS) based on his field goal locations and the EB-estimate PPA at each location.
3. Compute for the player's actual PTS scored.
   a. for Global SScE: total PTS scored
   b. for local SScE: PTS per cell/location
4. **PRLA = PTS - EPTS; local PRLA = PTS at cell - ELPTS at cell**

Points Relative to League Average (PRLA)

Global PRLA

- Positive values indicate players are scoring more than expected while negative values indicate the opposite.

Local PRLA

- PRLA computed per cell/location
- Can be mapped to show the spatial distribution of a player's PRLA—i.e. **at what areas on the court is a player scoring more or less than expected**.

What else?

- Comparison of teams and players using SScE and PRLA.
- Comparison of SScE and PRLA with conventional shooting metrics.
- Sniff test with fans, players, coaches, teams, etc.

FOSS4G applications and other open source tools used


QGIS
pandas
python™
PostGIS
ggplot2
R
scikit learn

If you want to learn more or collaborate...

https://github.com/benhur07b/ms-thesis-spatial-analysis-shooting-philippine-basketball


TORONTO

Thank you!

References

Bailey, N., Bhuwalka, K., Lee, H., & Zhong, T. (2018), Understanding Features of Successful 3 Point Shots in the NBA, Massachusetts Institute of Technology, MIT. Available at: http://www.mit.edu/~nbailey/files/3ptreport.pdf

Bornn L, Daly-Grafstein D. (2019). Using in-game shot trajectories to better understand defensive impact in the NBA. arXiv preprint arXiv:1905.00822

Brunet J-P., et al. (2004) Metagenes and molecular pattern discovery using matrix factorization. Proc. Natl Acad. Sci. USA, 2004, vol. 101 (pg. 4164-4169). DOI:10.1073/pnas.0308531101.

Cervone D., D'Amour A., Bornn L., & Goldsberry K. (2016). A multiresolution stochastic process model for predicting basketball possession outcomes. Journal of the American Statistical Association 111:585–599.

Cervone, D., Bornn, L., & Goldsberry, K. (2016a). NBA Court Realty. 10th MIT Sloan Sports Analytics Conference, 2016.

Daly-Grafstein D, & Bornn L. (2019). Rao-Blackwellizing field goal percentage, Journal of Quantitative Analysis in Sports 15:85–95.

D'Amour, A., Cervone, D., Bornn, L., & Goldsberry, K. (2015). Move or Die: How Ball Movement Creates Open Shots in the NBA. 9th MIT Sloan Sports Analytics Conference, 2015.

Franks, A., Miller, A., Bornn, L., & Goldsberry, K. (2015). Characterizing the spatial structure of defensive skill in professional basketball. Annals of Applied Statistics 9 (2015), no. 1, 94-121. DOI:10.1214/14-AOAS799.

Frigyesi, A., & Höglund, M. (2008). Non-negative matrix factorization for the analysis of complex gene expression data: identification of clinically relevant tumor subtypes. Cancer informatics, 6, 275–292. DOI:10.4137/cin.s606.

Gaujoux, R. & Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 2010, 11:367. http://www.biomedcentral.com/1471-2105/11/367.

Goldsberry, K. (2012) CourtVision: New Visual and Spatial Analytics for the NBA. MIT Sloan Sports Analytics Conference, 2012.

Goldsberry, K., & Weiss, E. (2013), The Dwight Effect: A New Ensemble of Interior Defense Analytics for the NBA, MIT Sloan Sports Analytics Conference, 2013.

Goldsberry, K. (2019). Sprawlball: a visual tour of the new era of the NBA. Boston: Houghton Mifflin Harcourt.

Harris, C. R. et al. (2020). Array programming with NumPy, Nature, 585, 357–362, DOI:10.1038/s41586-020-2649-2.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering, 9, 90-95, DOI:10.1109/MCSE.2007.55

Hutchins, L., Murphy, S., Singh, P., & Graber, J. (2008). Position-Dependent Motif Characterization Using Nonnegative Matrix Factorization. Bioinformatics (Oxford, England). 24. 2684-90. DOI:10.1093/bioinformatics/btn526.

Jiao, J., Hu, G., & Jan, Y. (2019). A Bayesian Marked Spatial Point Processes Model for Basketball Shot Chart. https://arxiv.org/abs/1908.05745v2.

Lee, D. & Seung, S. (1999), Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755):788–791, 1999.

Maheswaran, R., Chang, Y.H., Henehan, A., & Danesis, S. (2012). Deconstructing the Rebound with Optical Tracking Data. MIT Sloan Sports Analytics Conference, 2012.

Maheswaran, R., et al. (2014). The Three Dimensions of Rebounding. MIT Sloan Sports Analytics Conference, 2014.

Marty R. (2018). High-resolution shot capture reveals systematic biases and an improved method for shooter evaluation, In Proceedings of the 2018 MIT Sloan Sports Analytics Conference.

McKinney, W. (2010). Data Structures for Statistical Computing in Python, Proceedings of the 9th Python in Science Conference, 51-56. DOI:10.25080/Majora-92bf1922-00a

Miller, A., Bomn, L., Adams, R., & Goldsberry, K. (2014). Factorized point process intensities: A spatial analysis of professional basketball. In 31st International Conference on Machine Learning, ICML 2014 (pp. 398-414). (31st International Conference on Machine Learning, ICML 2014; Vol. 1). International Machine Learning Society (IMLS).

Murillo M.A.S. (2019, July 22). Sports analytics in the Moneyball era. BusinessWorld. https://www.bworldonline.com/sports-analytics-in-the-moneyball-era/

Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825-2830. URL http://jmlr.org/papers/v12/pedregosa11a.html

Pintor, B., & Cataniag, N. (2014). CourtVisionPH: A System for the Extraction of Field Goal Attempt Locations from Broadcast Basketball Videos and Spatial Analysis of Shooting. Undergraduate Research. University of the Philippines College of Engineering.

Python Software Foundation (2021). Python. URL https://www.python.org/.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robinson, David (2017). ebbr: Empirical Bayes on binomial data. R package version 0.1.

RStudio Team (2021). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Rolland, G., Vuillemot, R., Bos, W., & Rivière, N. (2020). Characterization of Space and Time-Dependence of 3-Point Shots in Basketball. MIT Sloan Sports Analytics Conference, 2020.

Sandholtz, N., Mortensen, J., & Bornn, L. (2019). Measuring Spatial Allocative Efficiency in Basketball. https://arxiv.org/abs/1912.05129v1.

Virtanen, P. et al. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17(3), 261-272. DOI:doi.org/10.1038/s41592-019-0686-2.

Shortridge, A., Goldsberry, K., & Adams, M. (2014). Creating space to shoot: quantifying spatial relative field goal efficiency in basketball. Journal of Quantitative Analysis in Sports, 10(3), 303-313. DOI:doi.org/10.1515/jqas-2013-0094.

Socamos, K. (2018, January 16). PBA News: Alaska analytics coach Layug on…. spin.ph. https://www.spin.ph/basketball/pba/more-than-just-numbers-as-analytics-coach-paolo-layug-tries-to-help-alaska-relive-lost-glory

Tenner, Z., & Franks A. (2020). Modeling Player and Team Performance in Basketball.