

STATS 601

# Caught Looking: Analyzing Variations in Umpire Strike Zones

Simon Fontaine, Moritz Korte-Stapff and Brian Manzo

{simfont, kortest, bmanzo}@umich.edu



April 20, 2020



**MOTIVATION**

# Motivation

## Problem Setting

In baseball, the home plate umpire is tasked with calling every pitch a ball or strike, unless the batter swings

MLB provides PITCHf/x data which includes information about every pitch such as location, umpire's call, speed, game situation, etc.[3]

We want to use this data to see how close umpires come to calling the true strike zone

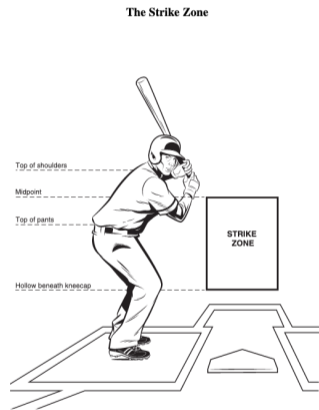


Figure: Official MLB strike zone[2]

# Motivation

## A brief detour into the rules of baseball

- If a batter gets 3 strikes before 4 balls, he is out. Otherwise he gets to walk to first base
- Therefore, the ball/strike call is more consequential in certain situations, leading umpires to possibly expand or shrink their strike zones

|           |    |      |   |        |   |     |   |   |   |    |      |      |     |
|-----------|----|------|---|--------|---|-----|---|---|---|----|------|------|-----|
| AT<br>BAT | 37 | BALL | 1 | STRIKE | 2 | OUT | 1 | H | E | 3  |      |      |     |
|           |    |      |   |        |   |     |   |   |   |    |      |      |     |
|           | 1  | 2    | 3 | 4      | 5 | 6   | 7 | 8 | 9 | 10 | RUNS | HITS | ERR |
| GUEST     | 1  | 0    | 3 | 1      | 0 | 1   | 2 | 0 | 0 | 1  | 8    | 9    | 1   |
| HOME      | 0  | 2    | 2 | 1      | 1 | 1   | 2 | 0 | 1 | 1  | 9    | 12   | 1   |

**Our primary question is whether different umpires have different strike zones in counts with [0, 1] strikes, 2 strikes, [0, 2] balls, 3 balls. We consider umpires with at least 30 games behind the plate in 2018.**

# Motivation

## Actionable items

- Evaluating umpire ability to determine who gets promotions, playoff assignments, etc.
- Assessing the need for automatic strike zone calls/robot umpires
- Delivering insights to pitchers and batters



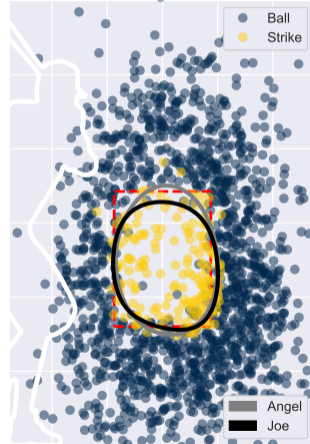
**This is a problem of interest to league officials, teams, and fans alike**

# Motivation

## Quantifying Differences in Strike Zones

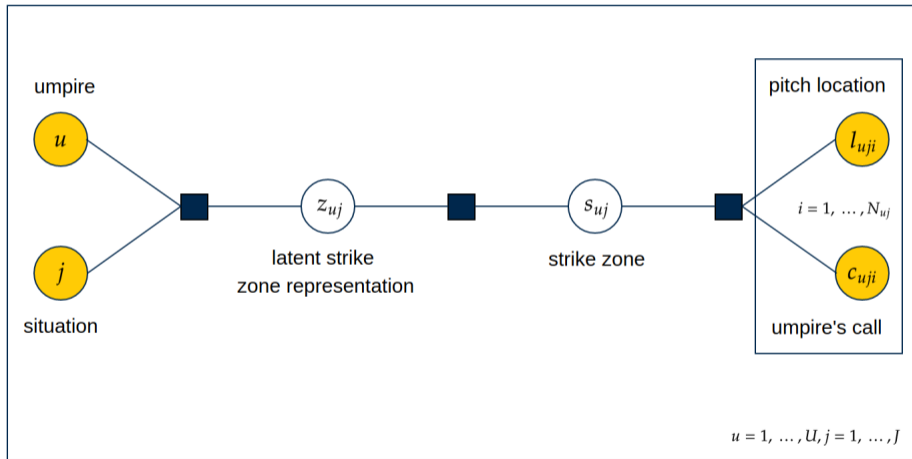
The plot on the right shows that Angel Hernandez and Joe West have different strike zones in counts with 2 strikes and <3 balls, but we need to develop a method to quantify the difference

Angel Hernandez vs. Joe West Strike Zone



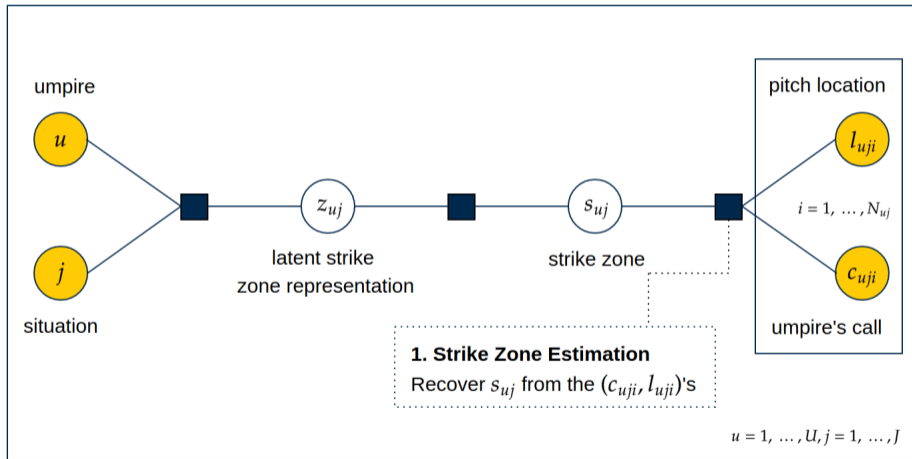
# Motivation

A graphical depiction of our problem



# Motivation

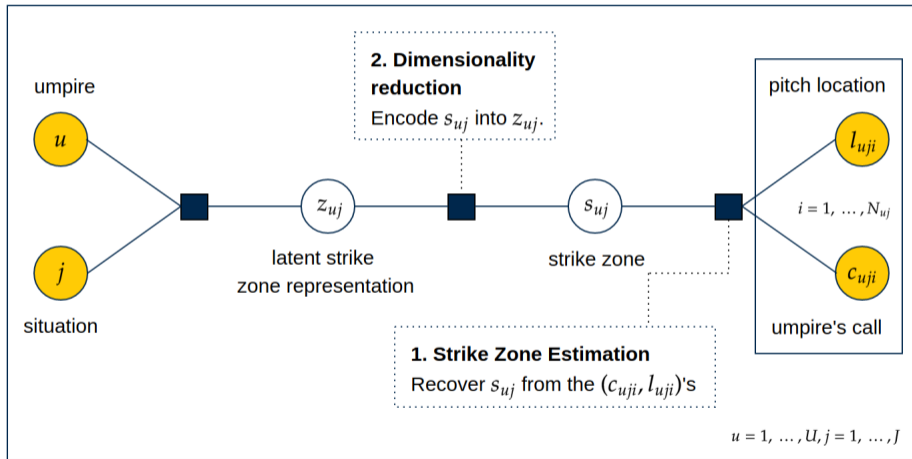
A graphical depiction of our problem





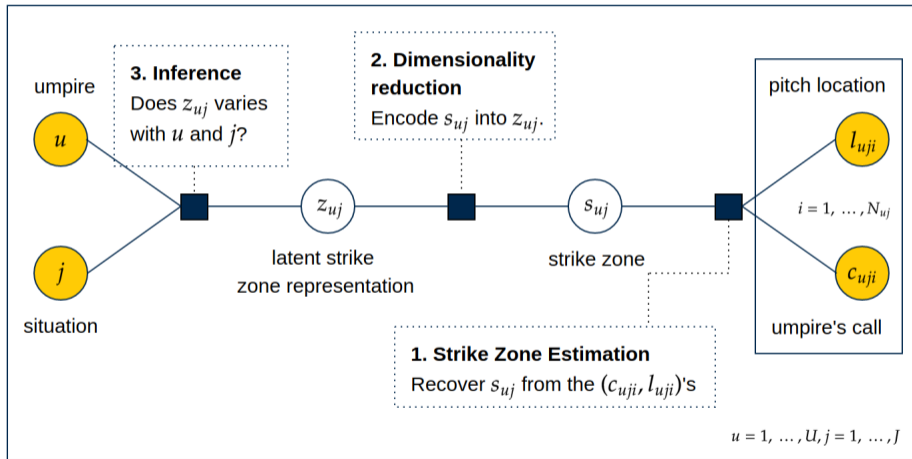
# Motivation

A graphical depiction of our problem



# Motivation

A graphical depiction of our problem



A photograph of a baseball player in profile, facing right. He has long, wavy blonde hair and is wearing a blue New York Mets uniform with orange piping and the 'Mets' logo on the front. He is also wearing a blue cap with a red brim and the Mets logo. He is looking down at a baseball he is holding in both hands. The background is a blurred stadium setting.

**CLASSIFICATION**

# Classification

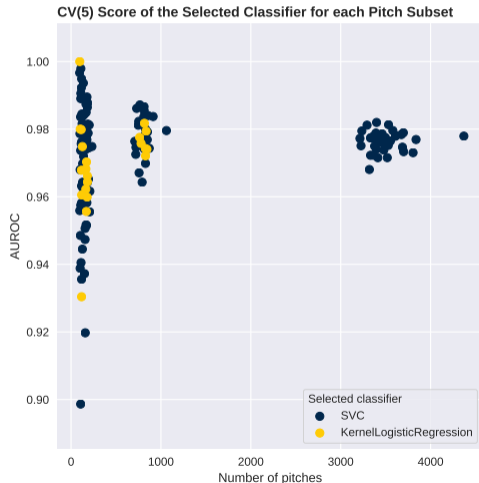
## Choosing a model to learn each umpire's strike zone

- We want to learn the strike zone boundary for each umpire in a variety of game situations
- Challenge is finding good boundary for specific game situations, where sample size may be small, without overfitting
- Some methods we tried included
  - **Kernel Logistic regression**
  - Logistic GAM
  - Neural network
  - **Kernel SVM**
  - Tree-based methods (AdaBoost, CART, Random Forest, Gradient Boosting)

# Classification

## Cross validation results

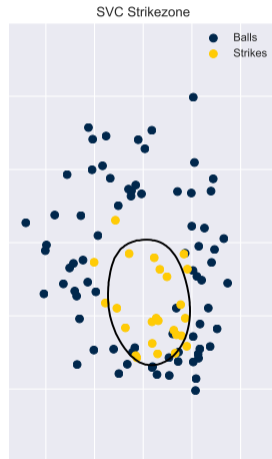
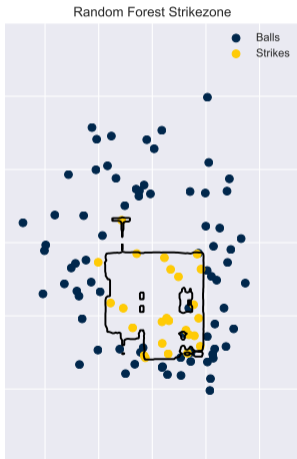
As some umpire/situation combinations have small sample size, we use cross validation to determine the best classifier (and tune them) for each subsample of the data. We use the AUROC score since it is less sensitive to unbalanced classes.



# Classification

## Comparison of Kernel and Ensemble methods

Although these methods have similar error rates, SVC produces a more realistic boundary

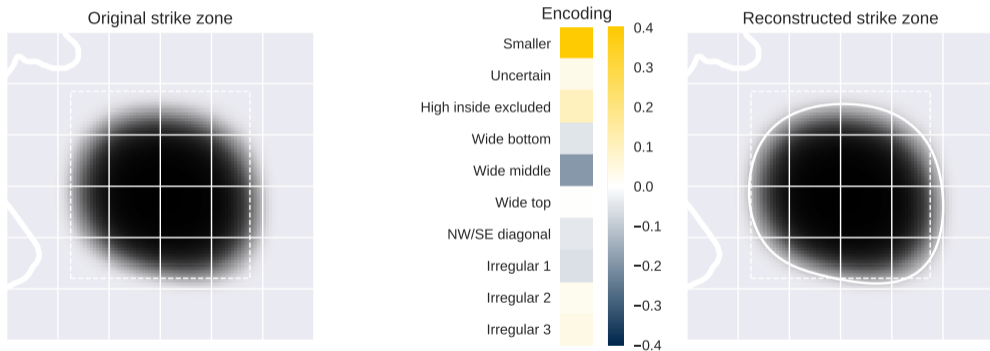


An aerial view of a baseball stadium at dusk, filled with spectators. The field is illuminated, and the city skyline is visible in the background. The text "DIMENSIONALITY REDUCTION" is overlaid in large white letters across the center of the image. The stadium is packed with fans, and the field is clearly visible with players on the field. The sky is a mix of orange and grey, indicating sunset. In the background, the Gateway Arch is prominent, along with various city buildings and stadium lights. The Cardinals logo is visible on the field.

# DIMENSIONALITY REDUCTION

# Dimensionality Reduction

## Encoding





# Dimensionality Reduction

## Model and number of components selection

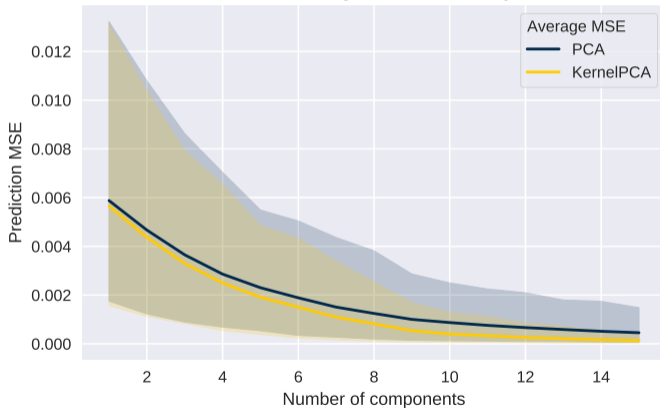
### PCA and Kernel PCA

- **orthogonal** embeddings are desirable for inference

### CNN Autoencoder

- Natural choice for image encoding
- Similar prediction error
- **non-orthogonal** embeddings

Encoders' Prediction Error by Number of Components



# Dimensionality Reduction

## Component interpretation

| <b>Component</b> | <b>Interpretation</b> |
|------------------|-----------------------|
| 1                | Overall size          |
| 2                | Overall uncertainty   |
| 3                | High inside           |
| 4                | Lower width           |
| 5                | Middle width          |
| 6                | Upper width           |
| 7                | Diagonal direction    |
| 8                | Irregular shape 1     |
| 9                | Irregular shape 2     |
| 10               | Irregular shape 3     |

Visualization app



**INFERENCE**

# Inference

## Multivariate Analysis of Variance

Model : components ~ umpire + ball count \* strike count

---

### MANOVA results

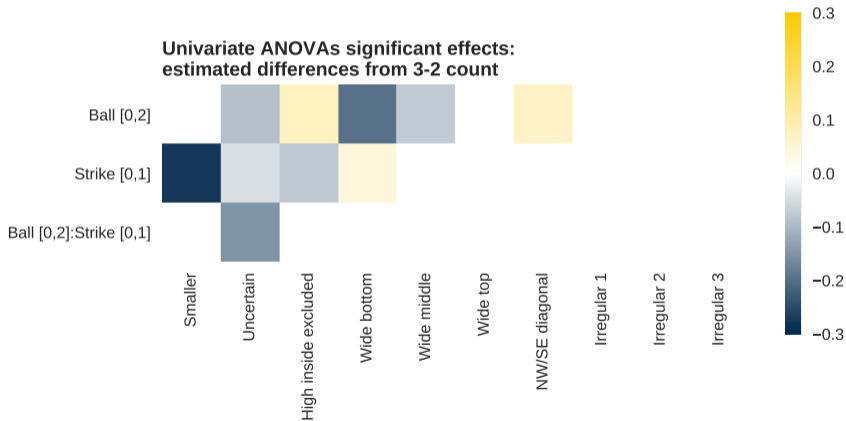
| Term                    | Wilks' Lambda | Num. df | Den. df | F value | p-value  |
|-------------------------|---------------|---------|---------|---------|----------|
| Umpire                  | 0.0142        | 380     | 1046    | 1.51813 | 1.80e-07 |
| Ball count              | 0.4112        | 10      | 105     | 15.0334 | 2.65e-16 |
| Strike count            | 0.3534        | 10      | 105     | 19.0555 | 1.74e-19 |
| Ball count:Strike count | 0.7675        | 10      | 105     | 3.1805  | 0.0013   |

---

# Inference

## Univariate Analyses of Variance

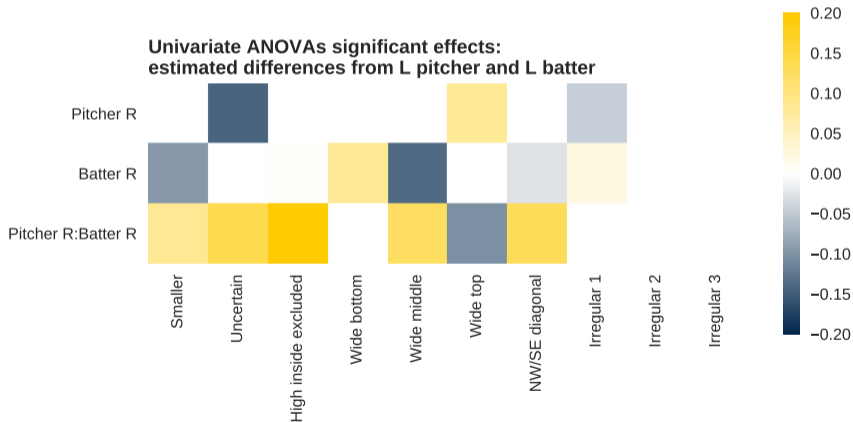
Model : component ~ umpire + ball count \* strike count



# Inference

## Univariate Analyses of Variance

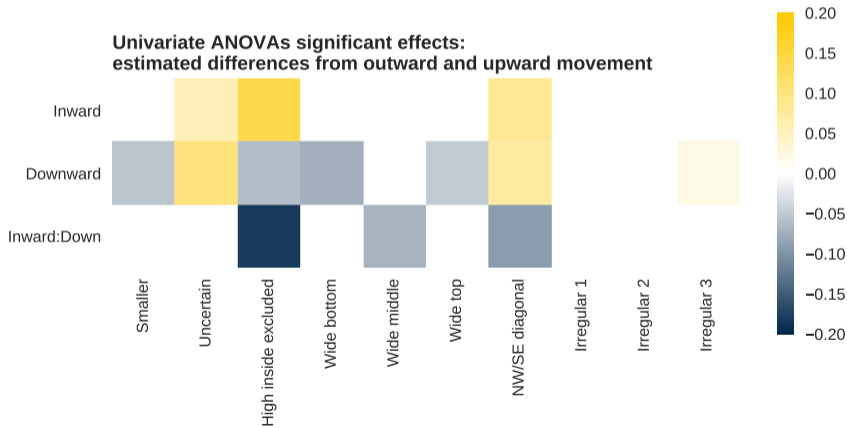
Model : component ~ umpire + batter \* pitcher



# Inference

## Univariate Analyses of Variance

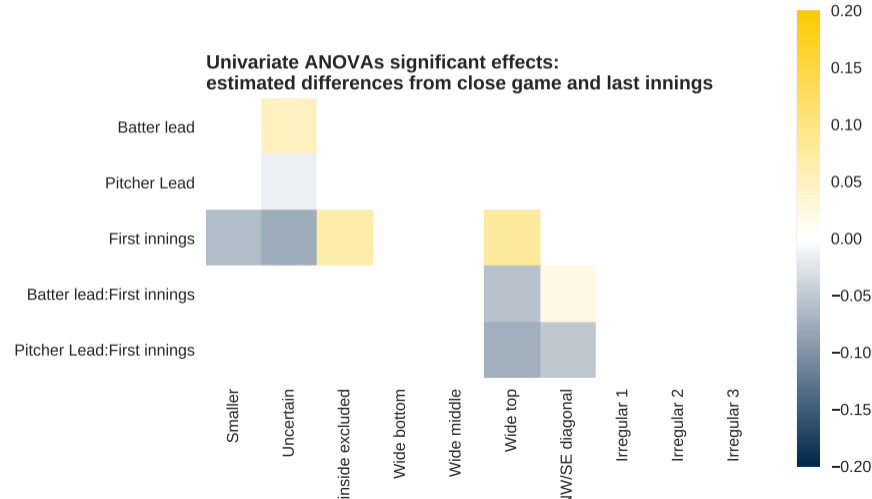
Model : component ~ umpire + horiz. move \* vert. move



# Inference

## Univariate Analyses of Variance

Model : component ~ umpire + score \* inning





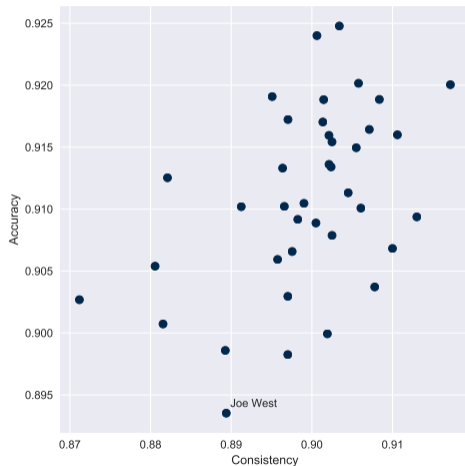


# RANKING UMPIRES

# Ranking Umpires

A metric that accounts for consistency and accuracy

- Umpires are frequently ranked by overall accuracy, but our classification procedure allows for umpires to be ranked on consistency, i.e., how similar is an umpire's strike zone across different game situations?
- We construct umpire ratings that weight consistency and accuracy 25%/75%, respectively.



# Ranking Umpires

How do we compare to Bloomberg's "Umpire Auditor"?

## Our top 5 umpires from 2018

- 1 Mark Wegner
- 2 Vic Carapazza
- 3 Pat Hoberg
- 4 John Tumpane
- 5 Alfonso Marquez

## Bloomberg's top 5[1]

- 1 Mark Wegner
- 2 Pat Hoberg
- 3 Alfonso Marquez
- 4 Nic Lentz
- 5 Sam Holbrook

And the worst, according to our scoring, is **"Country Joe" West.**



**CONCLUSION**

# Conclusion

## Findings

- Low-dimensional encoding of strike zones
- Balls and strikes count has a measurable effect
- Variability between umpires

## Remarks

- Sequential analysis
- Principal component regression
- Multiple testing

## Further analyses

We also considered the following situations:

- Pitcher arm (L/R) and batter stand (L/R)
- Score and inning
- Pitch movement (up/down and inward/outward)

Most yield positive results (omitted for brevity)

Analysis of the variability between umpires



**THANK YOU!**

# References

- [1] *2018 Umpire Ranking*. URL: <https://www.bloomberg.com/businessweek/graphics/baseballs-worst-call-of-the-day/#/umpires/ranking/2018> (visited on 04/18/2020).
- [2] *2019 MLB Official Rule Book*. URL: [https://content.mlb.com/documents/2/2/4/305750224/2019\\_Official\\_Baseball\\_Rules\\_FINAL\\_.pdf](https://content.mlb.com/documents/2/2/4/305750224/2019_Official_Baseball_Rules_FINAL_.pdf) (visited on 04/17/2020).
- [3] Paul Schale. *MLB Pitch Data 2015-2018*. 2019. URL: <https://www.kaggle.com/pschale/mlb-pitch-data-20152018>.

A baseball player in a red jersey and white pants is crouching on a field, looking down at a yellow ball on the ground. The word "QUESTIONS?" is overlaid in white text across the center of the image. The player is wearing white gloves and has a focused expression. The background is a blurred crowd of spectators in a stadium.

**QUESTIONS?**