# TOUCHDOWNS AND THROWDOWNS

## Analyzing Post-Game Crime

By David Boudia and RJ Edgerly

## Our motivation:

Rivalries are one of the most thrilling parts of college football. Fanbases' collective mood is made or broken by their team's performance, from the elation of a big win to the crushing defeat of a loss.

No matter the sport, the Michigan-Michigan State game can make even the most docile Michigander's blood run hot. While many fans keep their celebrating civil, sometimes, post-game revelry can get out-of-hand.

Michigan fans often accuse Spartan fans of acts such as flipping cars or burning couches, but, we wonder how accurate our perception is that Michigan is the more law-abiding program. We want to investigate if correlation exists between Michigan and Michigan State football and basketball game outcomes and campus crimes, or other unlawful behavior, and if one university shows a stronger relationship to game outcomes than the other.

## Our questions:

Is Michigan State (MSU) more prone to incidents, correlated to the timing of sporting events, compared to Michigan (UM)?

1. What offenses correlate with sport event outcomes?
2. Do games at certain times of the season produce more incidents?
3. Do games at certain times of day have a stronger correlation to incidents than other game times?

We will explore ten seasons (2009-2019) of college football and basketball game data for both universities, examining the number of incidents involving law enforcement and associated offenses for both college campuses.

## Outcome:

This analysis can provide nuance in understanding each school's fanbases' likely reactions to football and basketball game outcomes. Insights gleaned could be shared with local law enforcement for proactive mitigation of crimes based on findings.

# METHODOLOGY

## Ethics

The results within are reflective of *reported* incidents - not necessarily an indication that a crime was committed, but only that it was reported.

Our list of NIBRS incidents is not the full set. Some crimes are not related to the outcome of a sporting event. They just happened to occur within the time window after a game.

We ask our readers to keep these points in mind as they interpret our results, so they do not derive conclusions that are beyond the scope of our analysis.

## Data

We approached comparisons between Michigan and Michigan State with the intent to keep each as equally representative as possible.

We represent data by percentages and correlations because the number of incidents between Ann Arbor and East Lansing is unequal, along with their populations.

For comparison between Michigan and MSU, we tend to represent findings as ratios rather than absolutes. This allows for accurate comparison between the two, given their difference in population sizes.

## Analysis

Code used to produce analysis may be found in our GitHub repository.

Our analysis was built with reproducibility in mind. NIBRS data used for this report exists in its current state within the repository.

Outputs were produced in order of notebooks in the 03_notebooks/ directory.

### By The Numbers

- 10 Seasons: 2009 - 2019, exclusive
- 993 Football and Basketball Games
- 6,716 reported incidents

# DATA SOURCES

## NIBRS (National Incident-Based Reporting System - FBI)

Our analysis used 11 years of NIBRS Michigan crime data. Each year contained the following .csv tables:

**Incidents** – Each row contains a record for when a law enforcement agency was called to investigate potential unlawful activity, similar to an incident report an agency would file. Each incident has an associated Agency.

**Agencies** – The enforcement agency responsible for investigation of an incident. We focused on the campus and city police agencies for Ann Arbor and East Lansing.

**Offenses** – Each incident cites an offense, or multiple offenses, and shares ID values with Incidents .csv.

**Offense Types** – A list of offense names for each offense ID.

## CFBD API

An API service for NCAA college football game data. JSON data from 2009 to 2019 was downloaded - 8,966 rows, 30 Columns.

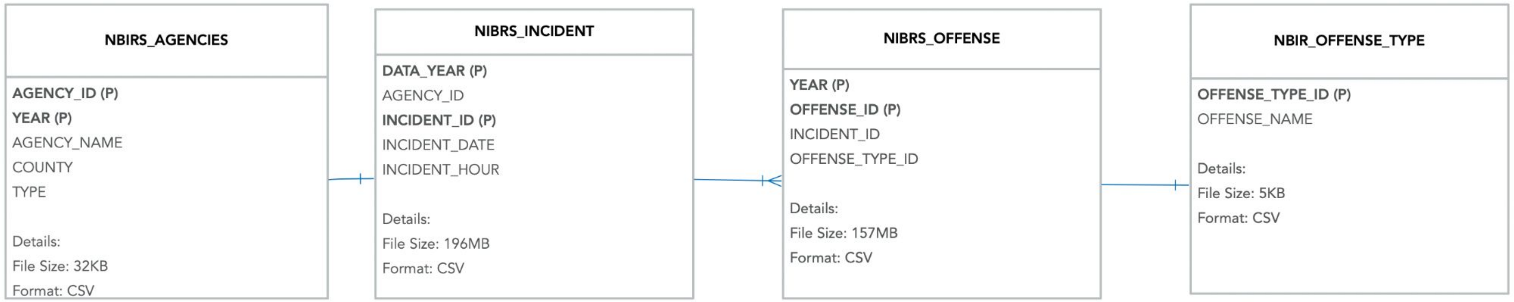
## sports-reference.com

We scraped each program's page using Python Requests library. We collected AP poll rankings from 2009 to 2019 for University of Michigan and Michigan State University. Data was placed into a .csv file.

## mgoblue.com & msuspartans.com

These two websites provided additional basketball details, like game start times. We used Selenium to scrape each season's data to a .csv file. 726 rows - each row represents one game.

NIBRS Entity-Relationship Diagram



# DATA MANIPULATION

## NIBRS Data

For each type of data topic (Incidents, Offenses, Agencies, and Offense Types) we evaluated the respective file for each year to compare column (attribute differences).

We pulled columns that were consistent across the 11 years or, in the case of some files missing the year as a column, we created it.

Each topic was concatenated year-by-year. In order attribute data to a sporting events, we needed to combine NIBRS incident date and hour into a full timestamp.

## Sports Data

The format of football data differed from basketball, so we created a school column containing either UM or MSU, and a column indicating the opponent. Games where neither Michigan nor Michigan State played were excluded.

We attributed rankings (AP or Playoff/BCS) to both teams for football based off which ranking was available and appropriate to use depending on the time of the season. We created columns for result (W=1, L=0) and overtime (Yes=1, No=0).

One challenge was half of basketball games had missing start times. We obtained start times from each University's basketball site using Selenium, and converted start times for Michigan basketball back to Eastern Standard Time, if the game was played in a different timezone.

Reconciling start times for basketball games was necessary because we had to create time windows for incident attribution.

We used an "end incident window" time of 10 hours (2 hour game + 8 hours after) for basketball and 11 hours (3 hour game + 8 hours after) for football.

## Joining NIBRS and Sports Data

We used **Apache Spark SQL** to attribute incidents/offenses to each game.

SQL allowed us to perform complex joins and evaluate whether incidents occurred between game start and end incident window.

This created a dataset of games that had at least one associated incident/offense. We took this dataset and left joined it with the preceding games' incidents based on game_id to return a complete dataset of games with, and without, incidents/offenses.

# DATA MANIPULATION

**Program rank**
Early in our analysis, we identified poll rank as a proxy for the importance of a game. Given the difference in source data for football and basketball, we had to manipulate data for each to create a consistent poll ranking for all programs.

**Basketball:** Each program had its poll rank included in their game string. We extracted rankings with a regular expression.

**Football:** Poll rank was nested in JSON format. To extract poll rank, we looped through each team's ranking and added conditions based on week and year. The polls people follow change as the season passes. An example would be the switch from the AP Poll to the College Football Playoff poll around week 9 of each season after 2013.

**Rivalry**
Our second proxy for game importance was whether or not a game was a rivalry.

We represented rivalries as a binary encoding column: 0 or 1, for each game played. We assigned Michigan and Michigan State as each other's rival for fair comparison:
**Michigan**: Michigan State
**Michigan State**: Michigan

**Offense Grouping**
Of the 71 Group A NIBRS offenses, we wanted to know how offenses of similar type contributed toward incidents in context of game outcomes.

We mapped 7 classes of offenses to a new column: Physical, Property, Scam, Sexual, Substance Violation, Theft, Weapon Violation

**Venue**
We wanted to know how incident occurrences were affected by home or away games. We used LabelEncoder() to create a column with numeric values for each venue.

We assigned Michigan and Michigan State the highest values in the list: If there was a positive correlation between venue and offense group, it was tied to home games.

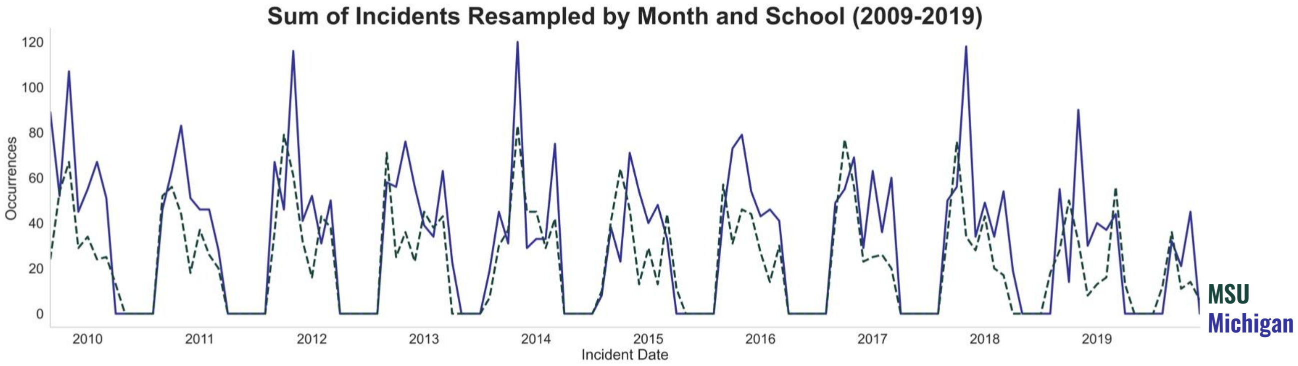
**Score Difference**
We wanted to know how the severity of a win or loss affected number of incidents after a game, and their linked offenses.

We took the point difference for each game and used qcut() to bin them into quartiles.

# OFFENSES BY SCHOOL

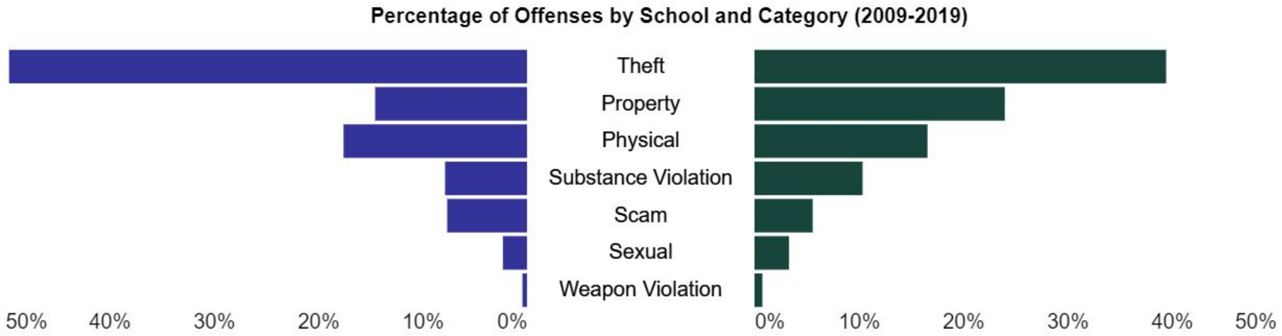
Our analysis examined incidents involving local law enforcement of both schools from 2009 to 2019 football and basketball seasons. The line plot below shows incidents across both sports' seasons. Both Michigan  and MSU show large peaks near October and November, falling near the end of the fall  semester. Michigan appears to have more pronounced peaks near the end of basketball season**.**  Michigan also has a higher incidence rate, most likely due to the larger population in Ann Arbor compared to East Lansing.

Looking at the barchart, **Michigan demonstrated the single-highest number of offenses from theft** occurring almost 2,000 times (~50% of their total offenses)  compared to MSU's 1,000 times (39%) though this represented the largest category for each school. Theft includes many offenses including larceny, shoplifting, theft from building and motor vehicle theft.




**MSU has a higher percentage and number of property-related offenses** (23%, 667) to Michigan (14%, 567). This category includes destruction of property and vandalism.

Michigan and Michigan State were **comparable in percentage of physical offenses** (16.5% and 17.5%), respectively. This category includes both simple and aggravated assault and intimidation.



# GAME IMPORTANCE

We built two features from our data, **poll ranking** and **rivalry**, as a proxy to measure a game's importance. We wanted to know if bigger games lead to more incidents. The table below illustrates how many incidents come from top games by poll rank, and how many come from rivalry games, indicating out of all wins or losses for that category, what the percentage was out of only wins or losses.

MSU football wins add incidents, so do Michigan football rivalry losses:

| Percentages | Top 25 | Top 20 | Top 15 | Top 10 | Top 5 | Rivalry |
|---|---|---|---|---|---|---|
| Michigan Football | W: 12.4<br>L: 38.5 | W: 11.9<br>L: 31.0 | W: 3.3<br>L: 20.4 | W: 1.6<br>L: 12.6 | W: 0.0<br>L: 2.6 | W: 5.6<br>**L: 25.1** |
| MSU Football | W: 24.2<br>L: 32.8 | W: 14.1<br>L: 28.3 | W: 9.5<br>L: 14.5 | W: 9.0<br>L: 4.1 | W: 2.3<br>L: 0.0 | **W: 15.5**<br>L: 13.6 |
| Michigan Basketball | W: 8.5<br>L: 17.2 | W: 5.4<br>L: 13.0 | W: 2.4<br>L: 9.1 | W: 0.8<br>L: 6.7 | W: 0.0<br>L: 0.9 | W: 0.5<br>L: 5.6 |
| MSU Basketball | W: 16.7<br>L: 27.6 | W: 13.9<br>L: 25.6 | W: 8.1<br>L: 12.9 | W: 3.8<br>L: 7.2 | W: 2.1<br>L: 1.1 | W: 2.6<br>L: 1.7 |

Many incidents for Michigan State come from games played against Michigan at Spartan Stadium in which Michigan State won.

Among the five games with the most incidents, Michigan State has two; both were played at Spartan Stadium where they beat Michigan.

Among the ten games with the most incidents, nine of those ten were Michigan vs. Michigan State - at Spartan Stadium or Michigan Stadium.

Incidents from football rivalry games in which Michigan lost to Michigan State make up a large percentage (25.1%) of incidents for football games in which Michigan lost. Many incidents occur for Michigan after a loss at the hand of a rival in football.

It is uncertain how represented each university is, given not all incidents belong to students. Some students from both might also visit each other (e.g. MSU students could initiate crime on Michigan's campus).

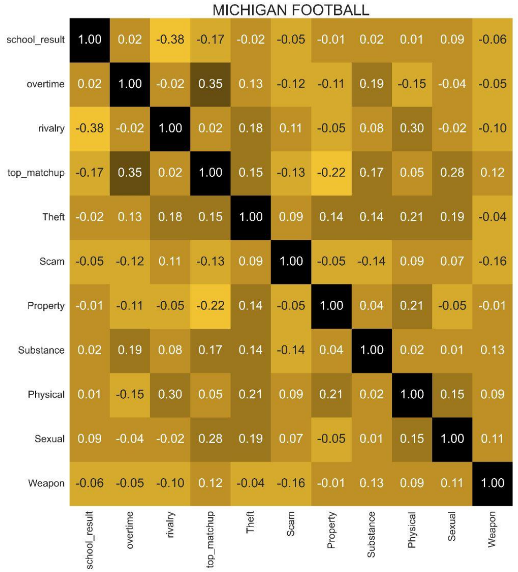
# CORRELATION

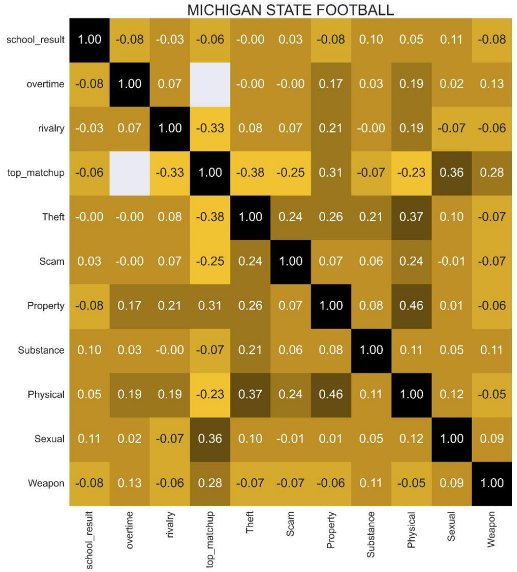
We ran Spearman rank-order correlations for both schools and their sports teams, evaluating relationships between crimes for overtime games, rivalry games, and top matchups.

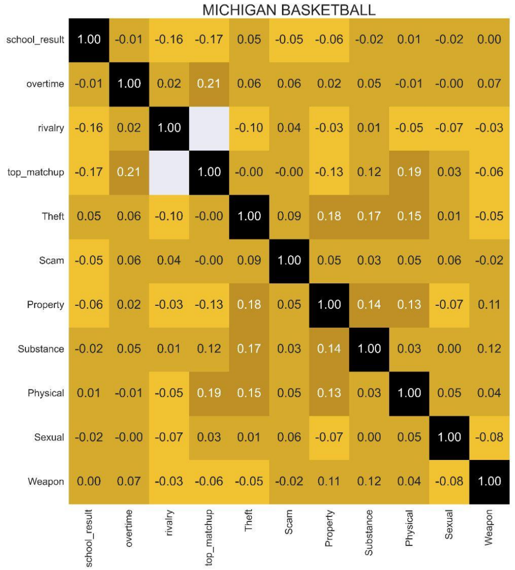
For **overtime**, we saw a slight positive correlation between Michigan Football and substance offenses (.19). MSU has a slight relationship between overtime and physical offenses (.19) as well as overtime and property crimes (.17), implying more instances of these offenses when games go into overtime.

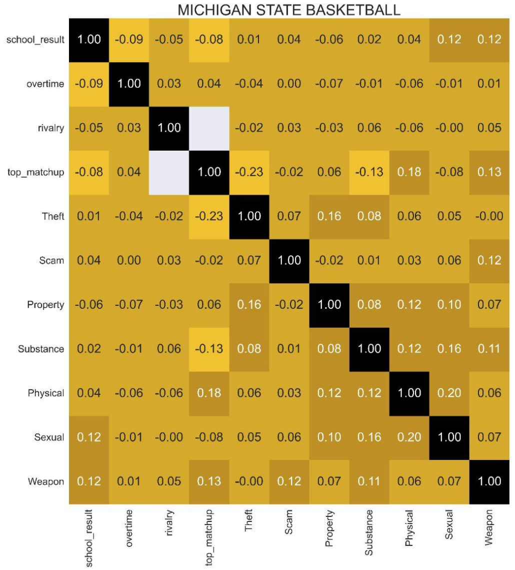
For **rivalry games** between Michigan and MSU, Michigan football had a moderately strong positive correlation with physical offenses (.30) and a less positive correlation with theft (.18). MSU Football showed a weaker correlation than Michigan to physical offenses (.19) and a positive relationship to property offenses (.21). Michigan is more likely to commit physical incidents when it comes to games against Michigan State, and Michigan State tends to have more property incidents when it faces Michigan.

**Top matchups** for both football programs had moderately strong positive correlations for sexual offenses (MSU: .36 and Michigan: .28). In addition, MSU Football showed moderately positive coefficients for both weapons (.28) and property offenses (.31). This suggests that as both teams and their opponents have better rankings, more instances of these offenses occur.



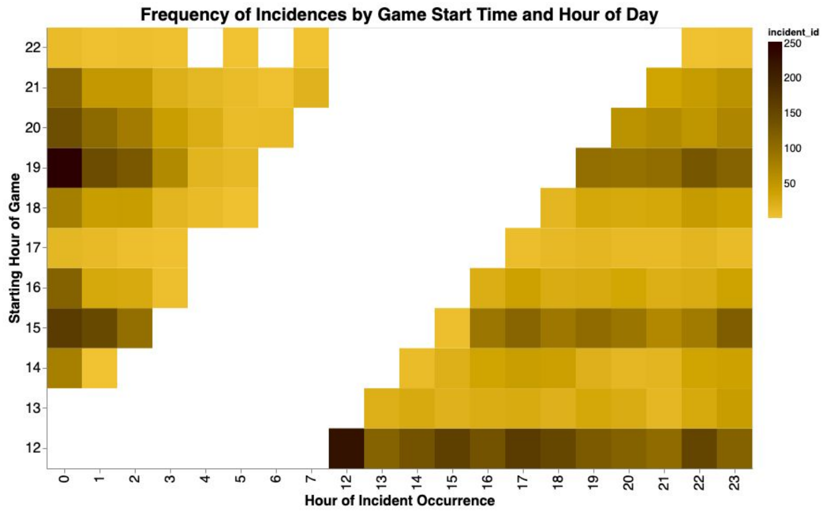



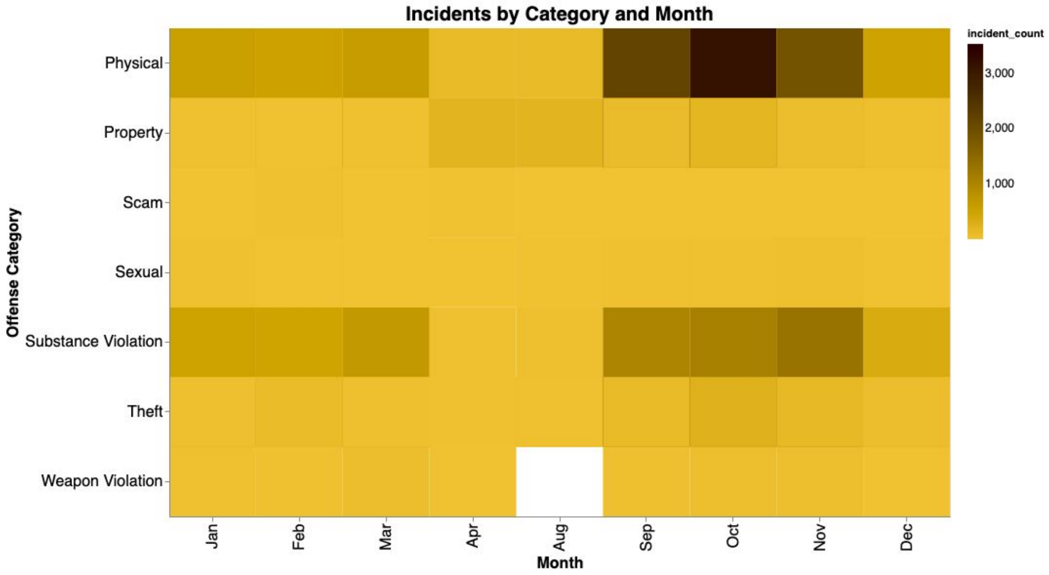



# TIME

The graph at the top shows number of incidents by the hour of day they occur. The associated game start times, by hour, are along the y-axis. Whitespace indicates times that occurred before games started and are outside of the incident window.

Taking the first y-value, 12 (noon start), we see that the most incidents occur within same hour and a steady number occur throughout the remainder of the day. For games starting in the 7 o'clock hour (y=19), we see a high number of offenses around midnight.




We also looked at offenses by month for all ten seasons for football and basketball. Physical offenses are the most frequently occurring category and are most prevalent between September and November, dropping off in December.

Substance offenses occur frequently throughout the football and basketball seasons as well.

# CONCLUSION

## The Takeaway

Most general sporting events do not correlate to incidents in the data we examined. But when we look further into the rivalry between Michigan and Michigan State, we can see some relationships form.

Football showed a higher incident rate for rivalry games and top matchups.

Michigan and Michigan State also showed trends among the types of offenses when they met each other on the football field: Michigan trends more toward physical offenses while Michigan State is trends more toward property offenses.

Physical and substance incidents are frequent, likely from football, in the fall. In winter during basketball season, there is a decrease in incidents. In general, basketball shows a weaker correlation to crime than football.

## Next Steps

- Expand the scope of our data to look at the Big Ten and the Midwest, if not Division I.
- Compare differences among strata of football and basketball: NFL/NBA, NCAA Div. I P5, NCAA Div. I G5.
- Examine the proximity of opponent to venue to try to determine how many attendees are home vs. away.
- Examine the age values of offenders/arrestees to make stronger associations with students.
- Pull in arrestee data which includes incidents where an arrest was made, e.g. drunk driving.
- Investigate possibilities with using the data to make prediction models for incident frequency and type.

## Final Notes

We hope our analysis added insight into how fanbases react to team matchups, specifically how Michigan's and Michigan State's fanbases typically react.

We would be interested in replicating this analysis for all 131 programs in NCAA Division I, and use this data to raise awareness for specific types of crimes likely to occur following rivalry games and top matchups.

Our analysis here was our first attempt at an ambitious data science problem, and we hope you enjoyed exploring the results as much as we did discovering them.

Thank you!

# STATEMENT OF WORK

## Contributions:

**David Boudia**
NIBRS data pull, college basketball scraping from sports-reference.com, merging of sport + NIBRS, baseline analysis, visualizations, report writing, data validation.

**RJ Edgerly**
GitHub repo setup + structure, Selenium basketball scraping, College Football API data pull, exploratory data analysis, report writing, report design, editing/proofing, notebook formatting.

## Sources:


https://api.collegefootballdata.com/api/docs/?url=/api-docs.json#/

https://ucr.fbi.gov/nibrs/2012/resources/nibrs-offense-definitions

https://www.icpsr.umich.edu/web/pages/NACJD/NIBRS/concepts.html

https://www.sports-reference.com

GitHub repository