# Sprawlball NFL – Utilizing GIS for the NFL's Benefit

**Inspired by the works of Kirk Goldsberry, Sarah Mallepalle, Ron Yurko,**

**Konstantinos Pelechrinis, and Samuel L. Ventura**

## Abstract:

Baseball has long used geographical data on a small scale to track information like pitching zones. Recently, the sport's use of geography has grown both in prevalence and scale with the growing popularity of spray charts to plot against the fear-mongering defensive shift. Following baseball was basketball, which now uses geographical data to track shooting zones. The NFL has recently hopped on the geography train, however it gatekeeps its source data behind the veil of Next Gen Statistics. Fortunately, this data can be gathered through the practice of image analysis, allowing it to be analyzed in new ways by independent sources. Which teams use the same play calls on repeat? Whose patterns are redundant and predictable? Did Matt Canada actually keep Kenny Pickett from throwing the ball over the middle of the field? Geographical data will answer all of these questions and more.

## Introduction:

I first discovered geographical information systems (GIS) in college. Combining two of my greatest passions, data and geography, I quickly fell in love with the subject and decided to study it as a minor. My journey with GIS did not end there, however. Shortly after graduating I ran into the work of Kirk Goldsberry during a routine scroll of Instagram and fell in love with the subject even more. Goldsberry, a doctor of data visualization and cartography (GIS before they had to spice it up with a modern name) and former vice president for strategic research of the San Antonio Spurs, is famous for his work combining GIS and sports. His work helped me to realize the true potential of GIS; not only is it the combination of data and geography, but could also be used in conjunction with sports, a third passion of mine.

While Goldsberry focuses his skills on the hardwood, I wish to utilize mine on the gridiron. After looking through Goldsberry's social media and book, *Sprawlball*, I realized there was a lack of similar work in football. Thus, the idea was born to track NFL passing locations GIS-style. I scoured the internet looking for readily available data to no avail. The data science world has not caught on in football yet like it has in basketball or baseball. Despite this, I was able to find a study from 2019 by four Ph.D students, Sarah Mallepalle, Ron Yurko, Konstantinos Pelechrinis, and Samuel L. Ventura that had aimed at producing this exact data, however the project's code had been wiped from the web. All that remained was the research paper and a batch of sample data from the 2017-2020 seasons. Fortunately, this information was enough to help me build a project of my own in order to get modern data. For more information on my coding process, view the Behind the Code essay for this write up on my website: ShermanData.com.

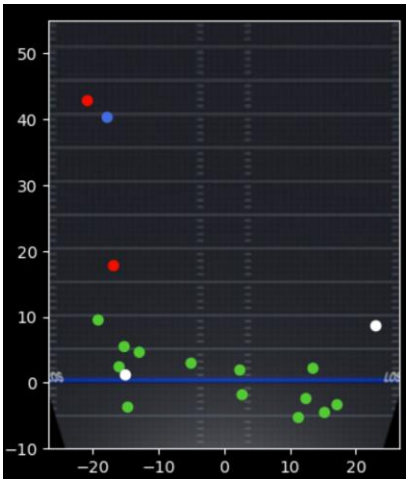
## Goal and Results:

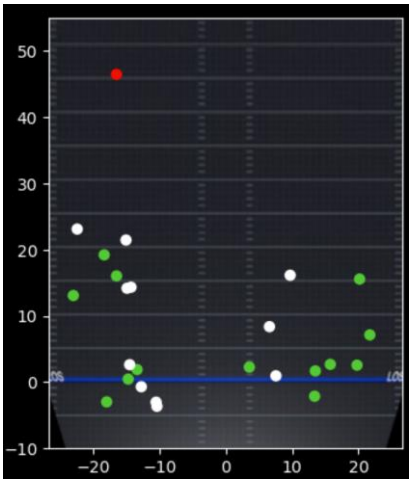
The locations of quarterback's passes can be tracked via the practice of image analysis on the NFL's own Next Gen Statistics passing charts. The process, which includes keystone correction, color thresholding, and multiple clustering algorithms, is reliable, though not to a tee. Below are two examples – one totally correct, and one with inaccuracies:

Original image:



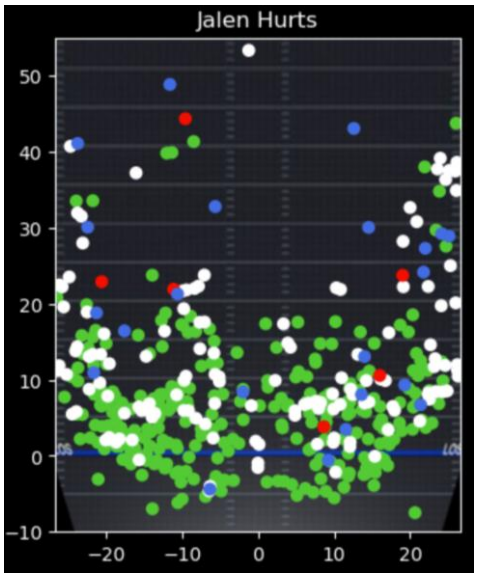
Data Gathered:



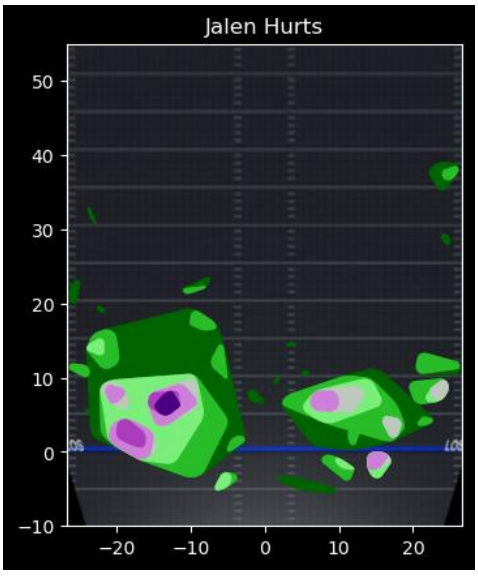

Notice the double white dots at roughly (-10, -4) and (-15, 14). These occur as a result of passes thrown out of bounds, which are still counted as incompletions, but do not appear on the graphics. Most of these incompletions are weeded out through a custom pixel variation algorithm, however a few still sneak through. Other small imperfections occur when two dots overlap one another, or when the blue touchdown streak (as shown in the first original image) paints over other dots, altering their color. That being said, these mishaps are minimal and are not prevalent enough to taint the data when used in full. Using these images on a full season's scale, graphics like the following are created:

## Jalen Hurts:




These are messy – but incredibly useful. The above picture is every pass thrown inbounds by Jalen Hurts in the 2022-2023 season. Small trends can be seen; deep shots are nearly-strictly along the sidelines, there is a strange hole around the breaking point of 10-yard out routes to the left sideline, and the middle of the field is surprisingly empty at all depths.

With a flare of GIS, however, these can be cleaned up to find even more trends. In geography, topographic maps showcase linear growths elevation by layering polygons. Using this same style on the chart above, polygons can be layered based on linear growth in pass density. Below is the end result:

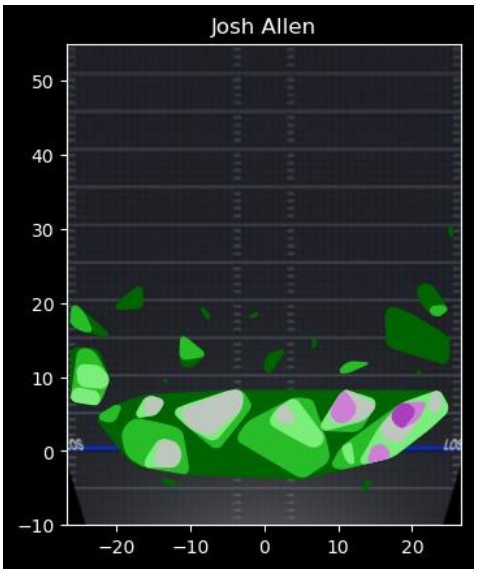


The diverging color scale in use ranges from dark green to light green in lower density areas and to light purple to dark purple in higher density areas. 'No density' areas have no associated color. Clusters are created if three points sit within 1.75 yards of one another radially. Further clusters are created at intervals of two points. That is, the deepest green requires 3 points within 1.75 yards, the next layer requires 5 points within 1.75 yards, the next layer requires 7 points, etc. until a maximum of 19+ points within this same distance.

First and foremost, all of the aforementioned trends are not only still visible, but more apparent than ever. Deep clusters are ONLY along the sideline. Dense clusters exist above and below the 10-yard line, but the 10-yard line itself has only low to no density clusters. Finally, the middle of the field is even more empty than thought priorly.

More importantly, the new and improved topographic charts also expose new trends. First of all, Hurts' hot spots tend to be 6-8 yards up on out routes. The right-handed quarterback also heavily utilizes right sided dump-offs. Opposing teams should be on the lookout for halfback swings, screens, or even pick plays to this section of the field. Despite all of this, the majority of his passes still go to the left side of the field. Hurts' deep balls are strictly sideline-bound, however the most important piece of intel picked up by this graph is the area of medium density about 37 yards up on the right sideline. It is extremely rare to get any density up this high, let alone two layers of it. Many of the league's most prominent deep ball throwers don't showcase such consistency in their passes. There was clearly a shot play drawn up in the Eagles' playbook designed to get the ball to this area of the field. Because of charts like this, defenses would know to be weary of deep balls to this location.

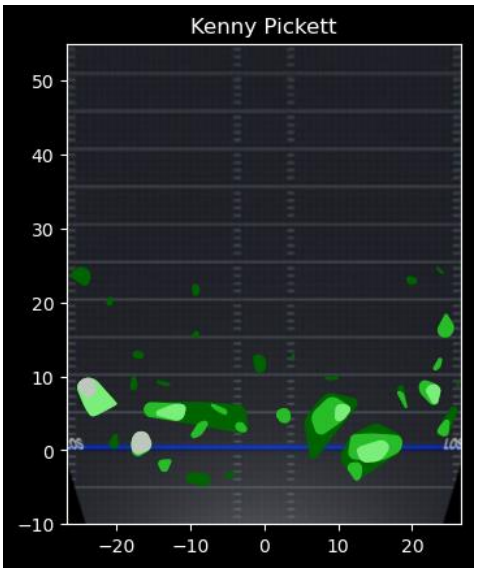
## Josh Allen:




Everyone knows that Josh Allen uses his big arm to push the ball downfield for big gains. What is less known is how he makes it so much more difficult for defenses by spreading his attempts across the field. In fact, his passing attempts show no discernable pattern beyond ~20 yards with the exception of a tiny cluster ~30 yards up on the right sideline. As stated, this isn't to say that Allen doesn't throw the ball deep – he certainly does – but rather that his ability to do so is so unique that there is seemingly no pattern to it at all.

Furthermore, unlike Hurts, Allen's height and strength allow him to utilize the middle of the field. Some of his highest density clusters are located partially within the hashes. Also in opposition to Hurts, Allen tends to favor the right side of the field in his short passing game. He also heavily utilizes the left sideline ~10 yards up. With Allen at the helm, the Bills double-down with dump-offs to both the left and right sides of the field.

## Kenny Pickett:




Pickett only played in 13 games (12 starts) commanding a run focused offense, so his data is limited. Limited data can still be analyzed. To begin, the rumors of the Steelers not throwing the ball over the middle of the field are true. Much like with Hurts' chart, the field in-between the hashes is nearly barren. With the Steelers being run heavy, better utilizing the middle of the field in the passing game could create personnel nightmares for opposing defenses Despite Pickett's apparent shortcomings, his arm strength doesn't appear to be a major hindrance. Pickett's limited chart lacks any purple clusters, but quite a few green ones can be seen forming around the 25-yard line. Much like most right-handed quarterbacks, Pickett's dump offs and screens are primarily to the right side of the field. Similar to Hurts, Picket consistently throws mid-range passes to the left side of the field.

## Conclusion:

The NFL regularly puts out passing charts mapping quarterback's passes. Despite this, the raw data is kept secret and made unavailable to the public. With the use of image analysis,

data science, and GIS, teams can scrape that data and utilize it in a number of ways. While topographic maps are typically used to show elevation data, they can be used in football to visualize passing trends. These visualizations can not only help defenses prepare for upcoming offenses, but also help offenses break trends and catch defenses by surprise.