Dribble-metrics

Shooting Off the Dribble: Understanding the Stigma

Alan Lo

Problem Statement

With this project, we decided to target the topic of off the dribble shooting. Although we understand it is commonly understood that there is a stigma surrounding pull up shots as opposed to catching and shooting, we wanted to not only statistically verify this stigma, but also look into **why** shooting off the dribble proves to be less effective. This project could offer insight into potentially further disincentivizing off the dribble shots amongst players.

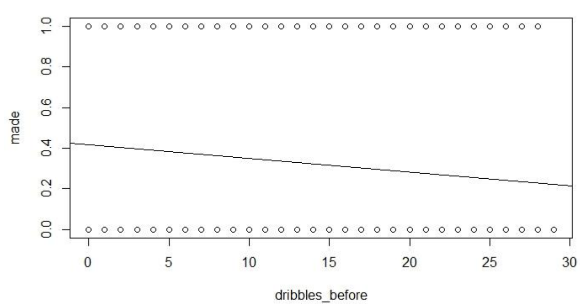
Data Source

The data we look at in this project is 10,000 attempted field goals by the Oklahoma City Thunder during the 2015-2016 season. The data provides advanced metrics such as the defender's angle relative to the shooter, the distance between the shooter and the defender, and other similar metrics that offer more insight into the specific details of each shot measured.

Methodology

We started this project by defining a set of expected relations predicated on metrics' relations on field goal percentage. These metrics are the following: (1) defender distance, (2) shooter velocity, (3) defender velocity, and (4) dribbles before. Trying to find the regressions of these metrics against field goal percentage proved to be difficult with the given metrics, because the statistic given to us was given in a binary manner, with 1 being a made shot and 0 being a missed shot; because of this aspect of the given metrics, the plots ended up looking like this using the linear modeling command in R:

To remedy this, we wrote a set of functions in Python to take in the data and output a set of metrics that would



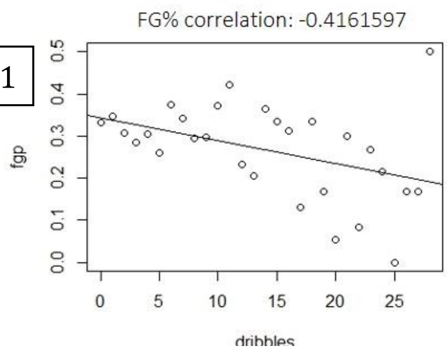
be more easily worked with and interpreted. What made these new metrics more easy to work with was that they established increments of the aforementioned metrics (i.e. defender distance and defender angle), and allowed for a field goal percentage to be measured within the given increments. As a result, we could run clean regressions between these metrics and field goal percentage, essentially allowing for the project to have a two-fold impact; first: we could verify whether or not expected trends between the previously established metrics and field goal percentage actually do exist, and second: we could determine the correlation between pre-shot dribbles and the metrics that are expected to factor into field goal percentage. Basically, the project boiled down to verifying that shooting off the dribble is ineffective, and then delving deeper into the reasoning behind this phenomenon.

Final Results

What we hoped to see were correlations that align with common basketball sense: (1) The faster a defender moves towards the shooter implies worse defensive positioning as the defender scrambles to contest the shot, and this worse defensive positioning entails a higher field goal percentage by the shooter. (2) The greater distance between the defender and the shooter translates to a higher field goal percentage, as the defender is not able to contest the shot as well as if he were guarding the shot closely. (3) A shooter moving at a higher velocity is less balanced, and has a less fundamentally sound shot, thereby decreasing field goal percentage.

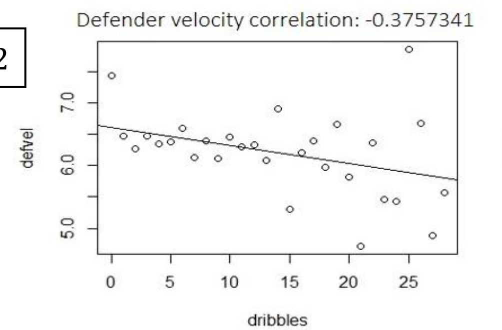
The figures below are the initial regressions that I ran, along with their corresponding correlation coefficients and model summaries.

1



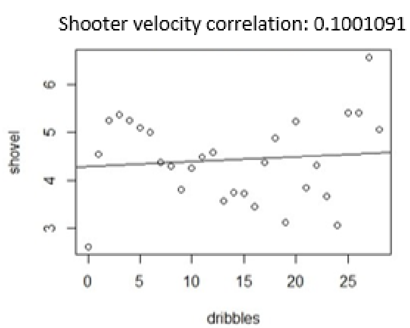

```
Residual standard error: 0.1036 on 27 degrees of freedom
Multiple R-squared:  0.1732,	Adjusted R-squared:  0.1426
F-statistic: 5.656 on 1 and 27 DF,  p-value: 0.02474
```

2



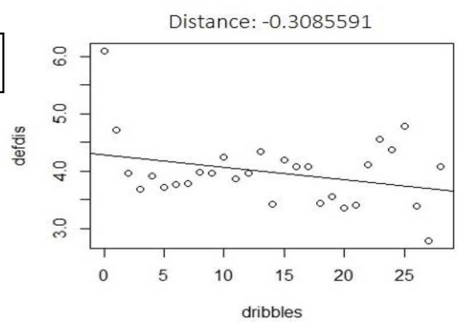

```
Residual standard error: 0.6219 on 27 degrees of freedom
Multiple R-squared:  0.1412,	Adjusted R-squared:  0.1094
F-statistic: 4.438 on 1 and 27 DF,  p-value: 0.04457
```

3



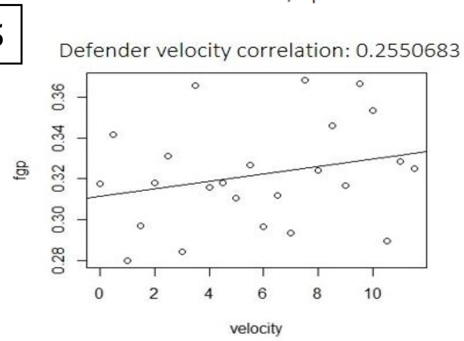

```
Residual standard error: 0.8923 on 27 degrees of freedom
Multiple R-squared:  0.01002,	Adjusted R-squared:  -0.02664
F-statistic: 0.2733 on 1 and 27 DF,  p-value: 0.6054
```

4

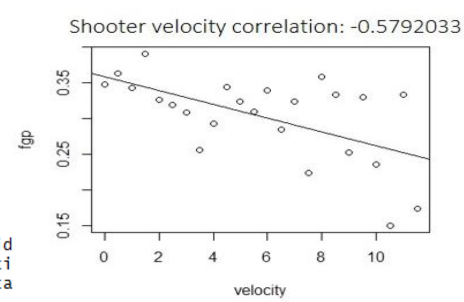



```
Residual standard error: 0.5774 on 27 degrees of freedom
Multiple R-squared:  0.09521,	Adjusted R-squared:  0.0617
F-statistic: 2.841 on 1 and 27 DF,  p-value: 0.1034
```

5



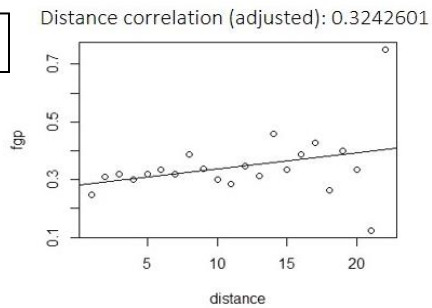
6




```
Resid                                                om
Multi                                                3053
F-sta
```

```
Residual standard error: 0.025 on 22 degrees of freedom
Multiple R-squared:  0.06506,   Adjusted R-squared:  0.02256
F-statistic: 1.531 on 1 and 22 DF,  p-value: 0.229
```

7

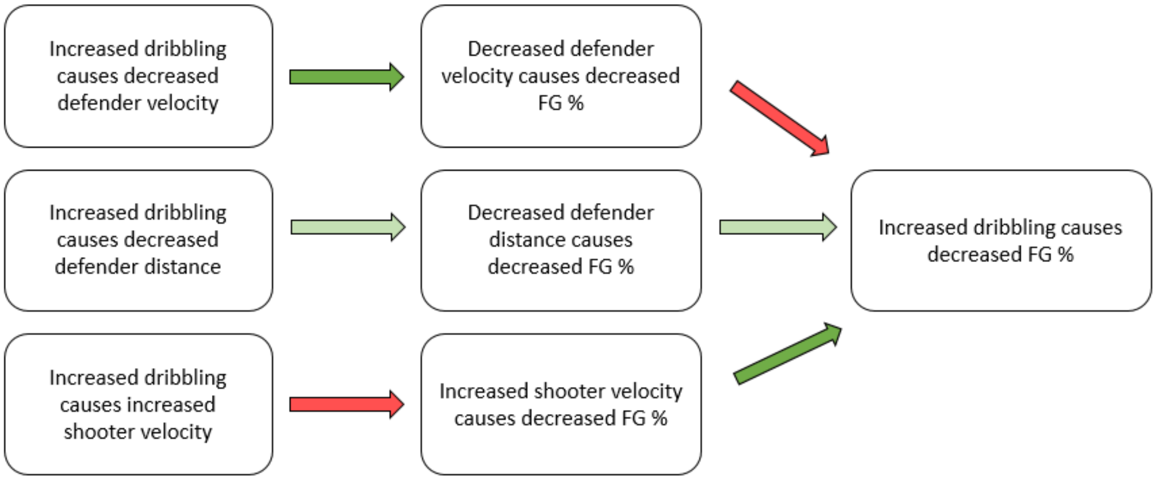



```
Residual standard error: 0.1094 on 20 degrees of freedom
Multiple R-squared:  0.1051,    Adjusted R-squared:  0.0604
F-statistic:  2.35 on 1 and 20 DF,  p-value: 0.141
```

Just as a quick run-through of these linear regressions, graph 1 is telling us that there is indeed a strong correlation between an increase and dribbling and a decrease in field goal percentage; this is merely verification that the phenomenon that we're trying to investigate is indeed a phenomenon that proves true with Oklahoma City's statistics. Graphs 2-4 are the relations between dribbling and the metrics of defender velocity (defvel), shooter velocity (shovel), and defender distance (defdis). Graphs 5-7 are the relations between the aforementioned metrics and field goal percentage. The goal here was to draw a syllogistic link chain of logic: If we can prove that increased dribbling decreases defender distance, and then prove that decreased defender distance decreases field goal percentage, then that effectively proves that increased dribbling decreases field goal percentage. This was the type of logic we looked to prove with all 3 metrics. However, in order to prove these correlations, we needed to first study how strong these correlations actually were.

	FGP V. DRIBBLES	DEFVEL V. DRIBBLES	SHOVEL V. DRIBBLES	DEFDIS V. DRIBBLES	FGP V. DEFVEL	FGP V. SHOVEL	FGP V. DEFDIS
RESIDUAL STANDARD ERROR	Close to 0, suggesting low variability of our data.	Relatively low	Relatively low	Low	Very close to 0! Good.	Very close to 0! Good	Close to 0
R SQUARED	Low. Nevertheless, because human behaviors are generally hard to predict and has R^2 of below 0.5, our low R^2 value does not necessarily suggest that our regression is statistically insignificant.	Low. Nevertheless, because human behaviors are generally hard to predict and has R^2 of below 0.5, our low R^2 value does not necessarily suggest that our regression is statistically insignificant.	Very low to the extent that it sufficiently suggests that the statistic is insignificant	Low. Nevertheless, because human behaviors are generally hard to predict and has R^2 of below 0.5, our low R^2 value does not necessarily suggest that our regression is statistically insignificant.	Low. Nevertheless, because human behaviors are generally hard to predict and has R^2 of below 0.5, our low R^2 value does not necessarily suggest that our regression is statistically insignificant.	Same, but this one is actually high considering the fact that this is the statistics on human behavior, which generally have R^2 lower than 0.5	Low. Nevertheless, because human behaviors are generally hard to predict and has R^2 of below 0.5, our low R^2 value does not necessarily suggest that our regression is statistically insignificant.
F-STATISTIC	F-statistic > F-table value (approximately 1.90 when alpha = 0.05. Thus, reject H0, which claims that there is no relationship between fgp and dribble.	F-statistic > F-table value (approximately 1.90 when alpha = 0.05. Thus, reject H0, which claims that there is no relationship between fgp and dribble.	F-statistic < F-table value (approximately 1.9 with alpha = 0.05) Thus, fail to reject H0	F-statistic > F-table value (approximately 1.90 when alpha = 0.05. Thus, reject H0, which claims that there is no relationship between fgp and dribble.	F-statistic < F-table value (approximately 2.05 when alpha = 0.05. Thus, fail to reject H0	F-statistic > F-table value (approximately 2.05 when alpha = 0.05. Thus, reject H0	F-statistic > F-table value (approximately 2.12 when alpha = 0.05. Thus, reject H0
P-VALUE	Low	Low	high	Fairly low	Fairly high	Very low	Fairly low
CONCLUSION	Strongly supports our H1, which is our expectation.	Strongly supports our H1, which is our expectation.	Insufficient to support H1	Sufficiently supports H1	Insufficient to support H1	Strongly supports our H1, which is our expectation.	Sufficiently supports H1

From the table above, you can see we take the statistical metrics from the summary of the regression models into consideration in deciding how strong a correlation truly is. What we end up finding is that there is a strong correlation between increased dribbling and a decreased defender velocity, as well as between shooter velocity and a decreased FG%. We find a sufficient correlation between increased dribbling and decreased defender distance, as well as decreased defender distance and decreased FG%, and we find that the correlations are weak for the relation between increased dribbling and increased shooter velocity and between decreased defender velocity and decreased FG%. To completely capture all of these relationships, the diagram below shows the link chain in logic that this project looked to prove true, and depicts how well the data actually validated the logic. A strong green represents a very strong correlation, a lighter green represents a weaker but sufficient correlation, and then the red represents a weak correlation that was insufficient in saying that a correlation exists within the relationship. The only relationship that actually ended up cleanly extending through the link chain of logic was the syllogistic logic that increased dribbling decreases defender distance, and decreased defender distance causes decreased FG%.



The findings of this project are far-reaching; what all of this tells you is that we not only understand that shooting off the dribble is typically a poor decision, but also that certain aspects inherent in taking a pull up shot as opposed to catching and shooting can be *causally* linked as to why shooting off the dribble decreases field goal percentage.