# Los Angeles Lakers Shot Chart

Andrew Sang
Stats 416
2019.11.28

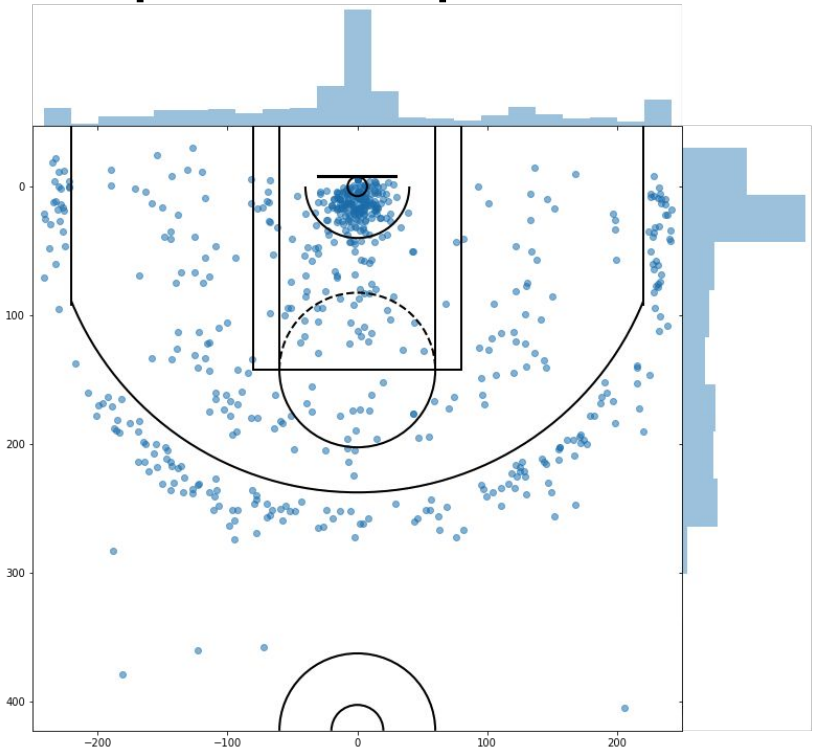
# Outline

1. Description of data
2. Plot data
3. Kernel Smoothing
4. Candidate Model #1
5. Candidate Model #2

# Data

- The Los Angeles Lakers are a professional basketball team in the NBA.
- The data used in this project is from the first 7 games of the 2019-2020 NBA Season that the Los Angeles Lakers played.
- Each data point represents an attempted shot and contains the location on the basketball court where the shot was attempted.
- There are 625 data points (attempted shots) taken by a total of 13 players.

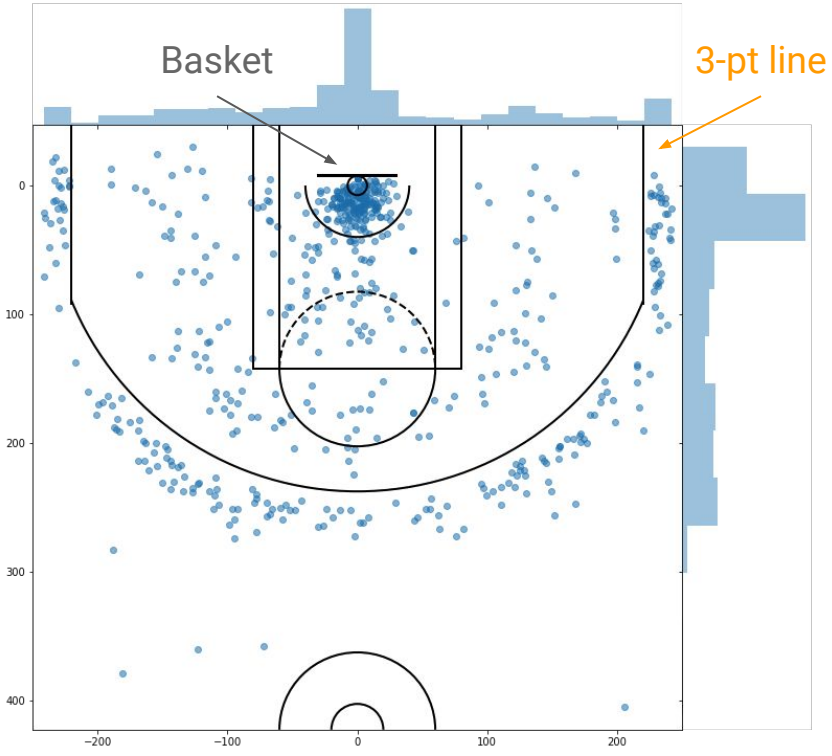
# Graphical Representation of Data





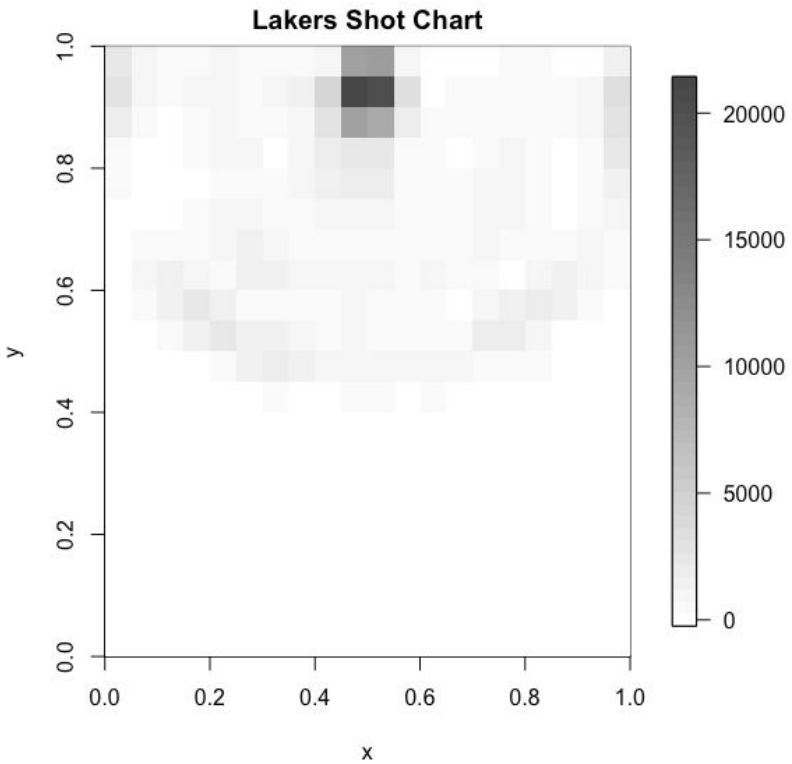
from VirtualStadiumTours

# Plot



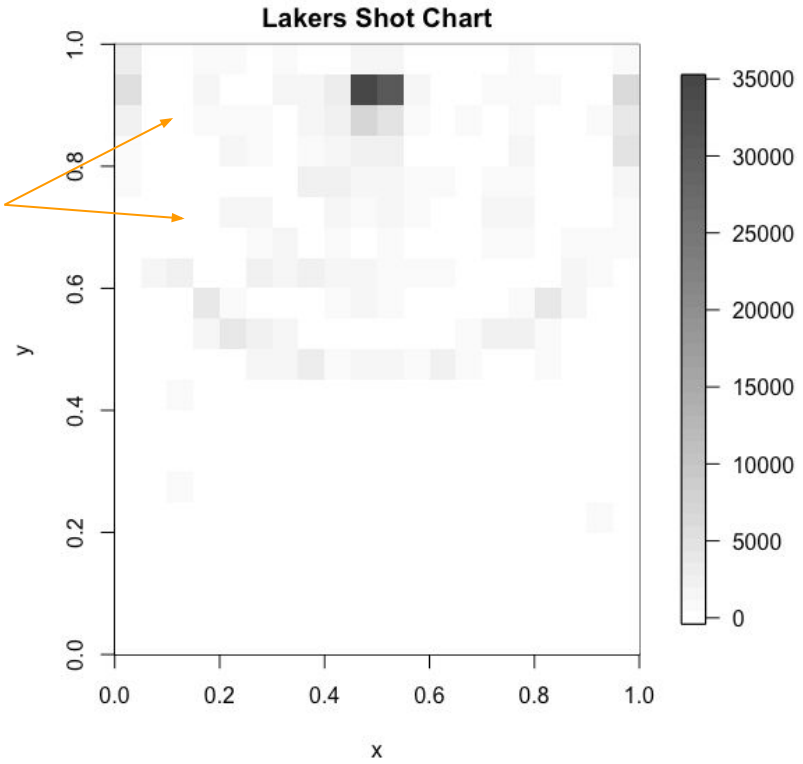

- This is a plot of the data where each point represents an attempted shot. We only look at ½ the court.
- A large percentage of shots are taken very close to the basket.
- A smaller number of shots are taken outside of the three point line.
- Not surprisingly, there is a lack of shots taken right inside of the three point line.

# Kernel Smoothing



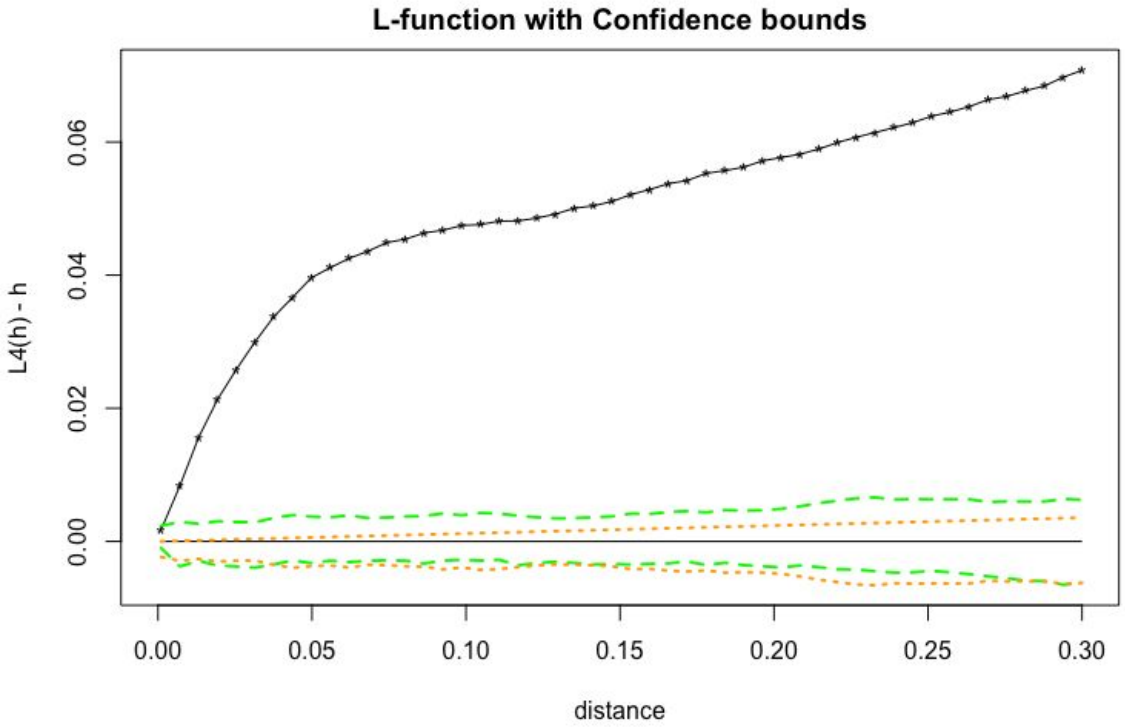

Bandwidth of 0.05

Gaps




Bandwidth of 0.025

# L-function

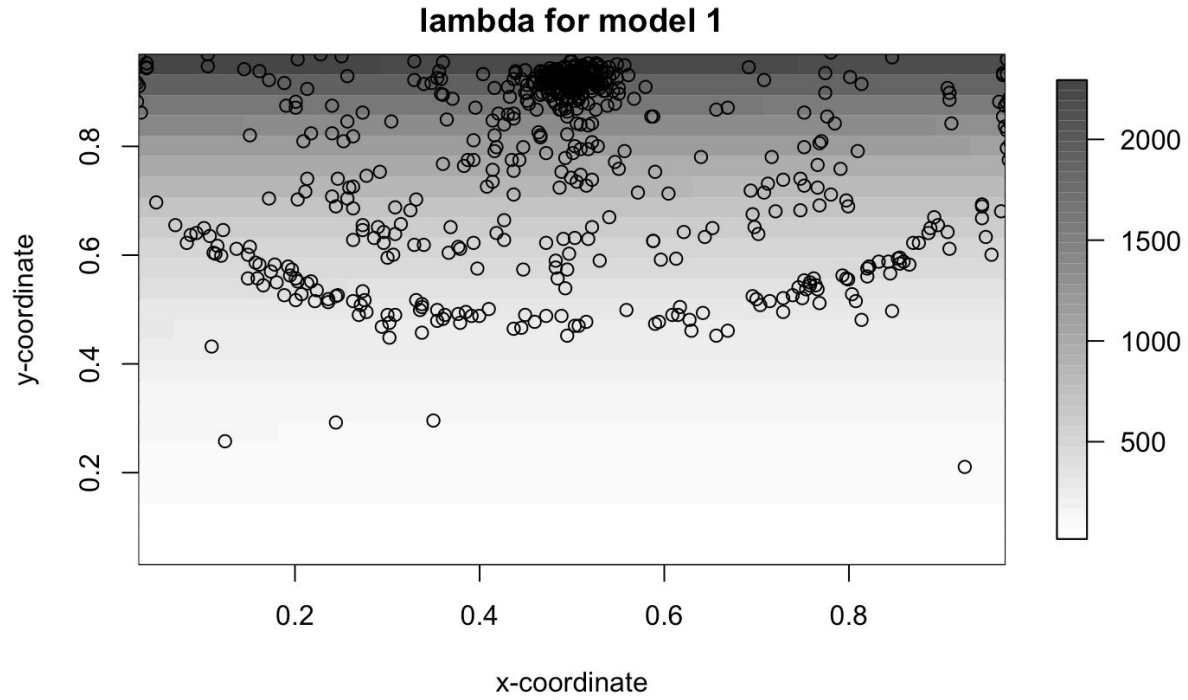



- There seems to be a great deal of clustering in the data, as the number of points for a given distance is much higher than both the theoretical and simulated points for that radius.
- Here, 0 is no evidence of clustering or inhibition.

```
fit1 = ppm(pp1, ~x + y)
```

# Candidate Model #1

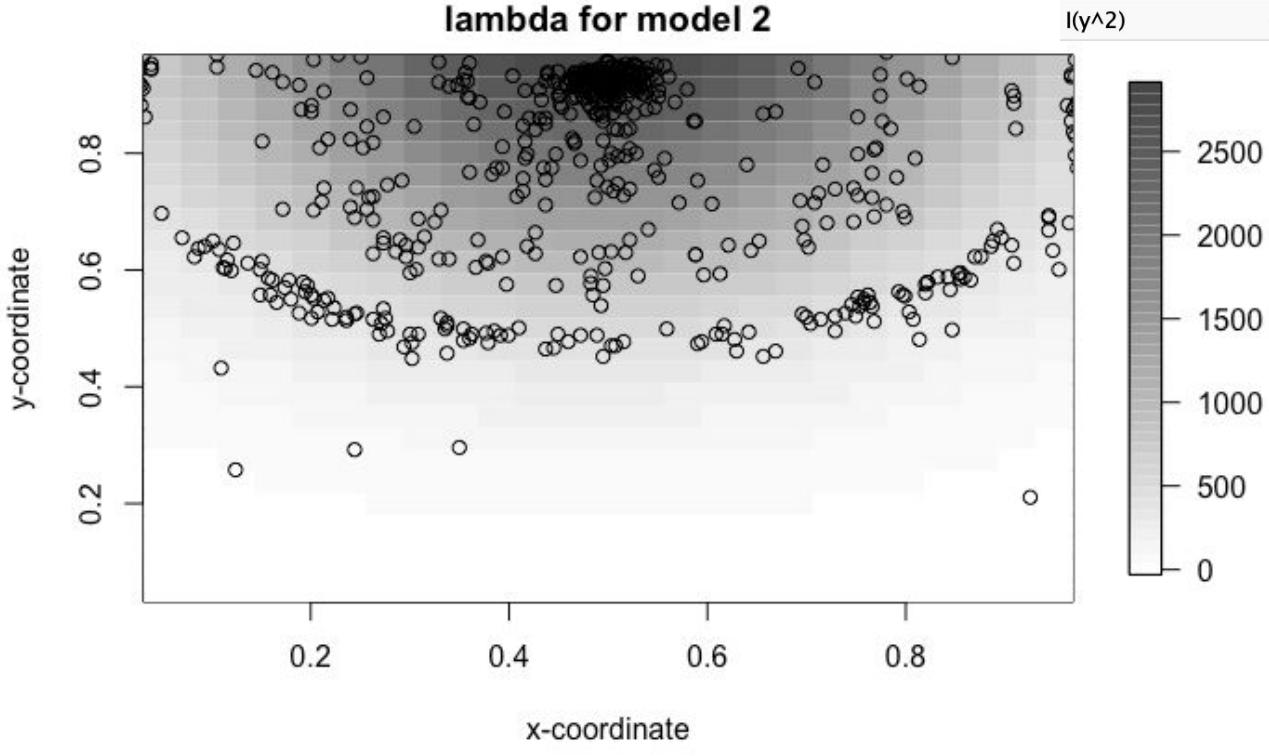
| | Estimate <dbl> | S.E. <dbl> | CI95.lo <dbl> | CI95.hi <dbl> |
|---|---|---|---|---|
| (Intercept) | 3.58893265 | 0.1766170 | 3.2427697 | 3.9350956 |
| x | -0.04845452 | 0.1385896 | -0.3200852 | 0.2231761 |
| y | 4.35852696 | 0.2013356 | 3.9639165 | 4.7531374 |



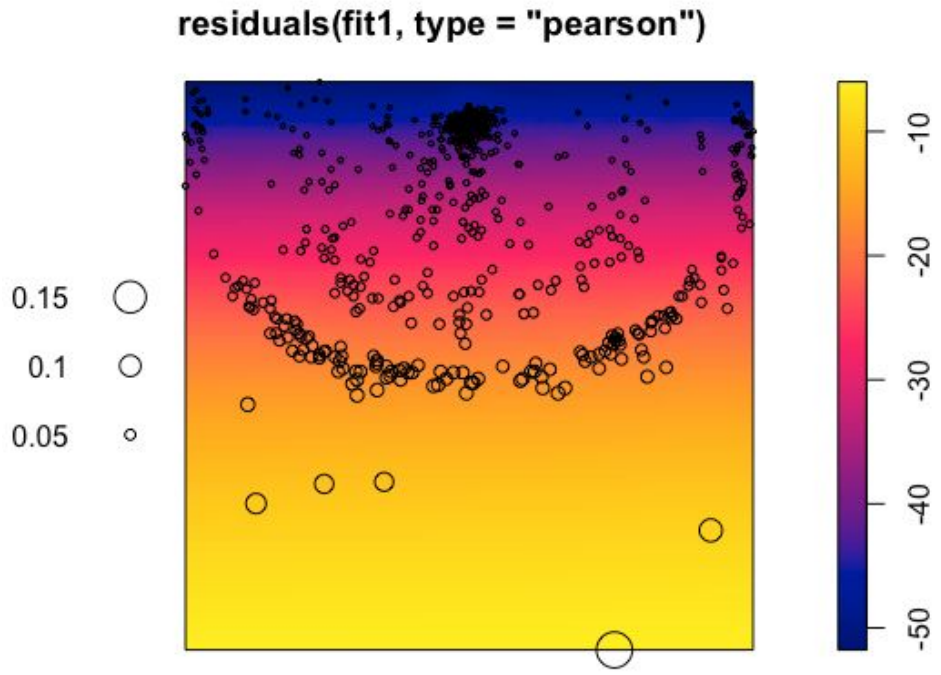
# Candidate Model #2

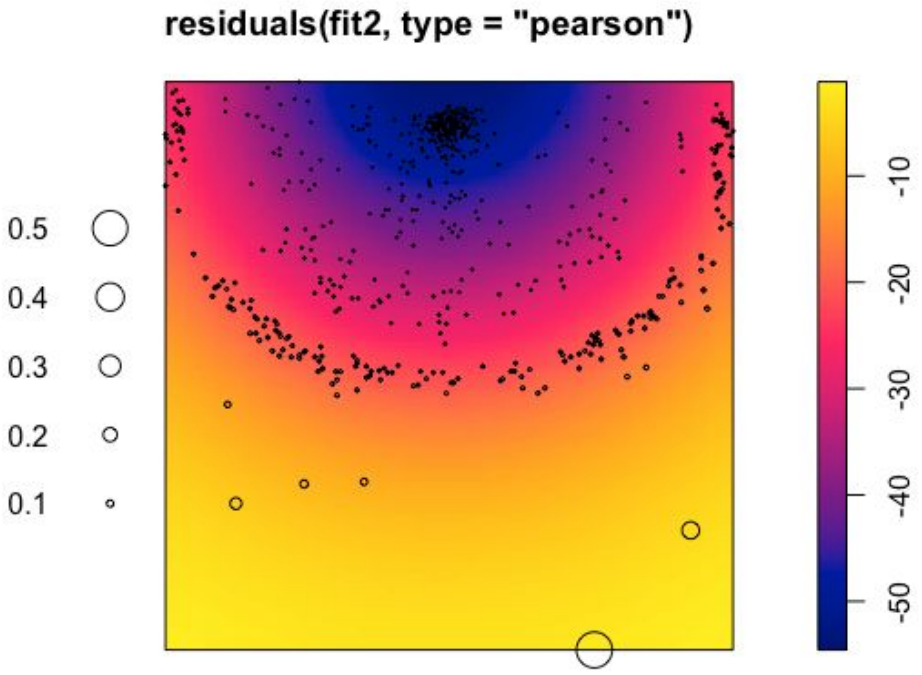
```
fit2 = ppm(pp1, ~polynom(x,y,2))
```

| | Estimate <dbl> | S.E. <dbl> | CI95.lo <dbl> | CI95.hi <dbl> |
|---|---|---|---|---|
| (Intercept) | 0.2886069 | 0.6549560 | -0.9950833 | 1.572297 |
| x | 6.1818164 | 1.0465110 | 4.1306926 | 8.232940 |
| y | 11.5786992 | 1.5479304 | 8.5448115 | 14.612587 |
| I(x^2) | -6.3530594 | 0.6615121 | -7.6495993 | -5.056520 |
| I(x * y) | 0.1224843 | 1.0197973 | -1.8762816 | 2.121250 |
| I(y^2) | -5.4103078 | 1.0414486 | -7.4515095 | -3.369106 |



# Comparison of Models




| | pearson <dbl> | logLik <dbl> | deviance <dbl> | AIC <dbl> |
|---|---|---|---|---|
| model1 | -0.82823696 | 3736.828 | 2967.880 | -7467.656 |
| model2 | 0.06178036 | 3807.684 | 2826.168 | -7603.368 |

# Next Steps

1. **Try Covariates:** Would have been interesting to see if there was a way to set up a model which would have taken into account the expected value of a shot attempt.

   FG Success Rate * Value of Shot

2. **Try different models:** Would have liked to see if there was a way to take into account the three point line to represent the lack of shots directly inside the line. Tried polynomial terms but think that something closer to “piecewise” regression might have worked better.