Statistical Analysis of Kobe Bryant's Shots

Team 5
潘萱尹、陳鼎文、鍾冠毅、劉國勳、陳章豪

Outlines

- Data introduction
- Comparisons and hypothesis tests
- Shots’ locations comparison
- KNN classifier
- Annual shots’ locations
- Future works

Data Introduction

- From the Kaggle
- Containing all the Kobe’s shots during his career
- 30697 shots and 25 variables
- 25697 training data and 5000 test data

Data Introduction (cont.)

- Response: Whether a shot is made or not
- 1 and 0 are denoted for a shot is made or not
- NA is denoted for the test data

Data Introduction (cont.)

- Predictors: 24 variables
- Factors: shot types, shot zone, match up, playoffs…
- Continuous: location(x, y), time remaining…

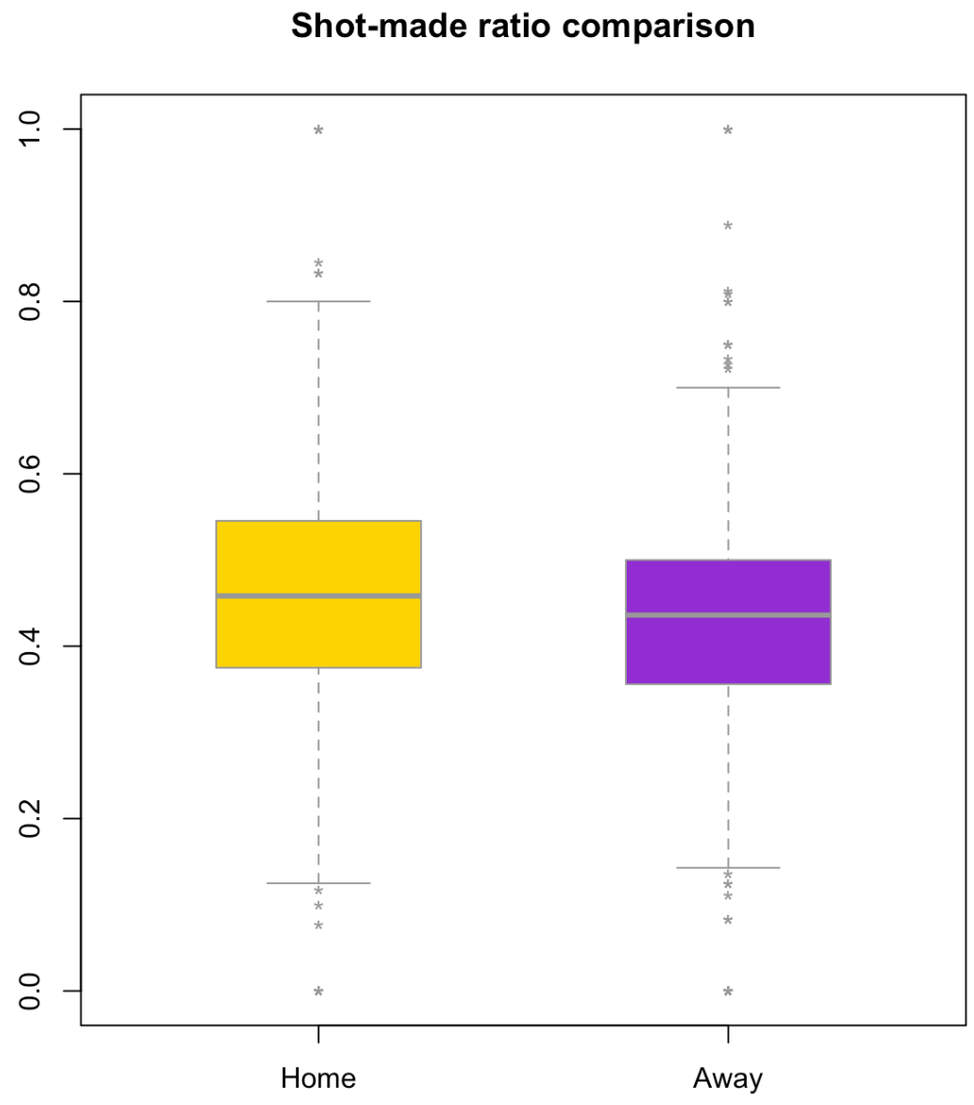
Data Introduction (cont.)

```
> summary(df)
     season           game_date        game_id                     action_type
 2005   : 1924   2016-04-13:   43   21501228:   43   Jump Shot           :15836
 2002   : 1852   2007-03-30:   41   20601081:   41   Layup Shot          : 2154
 2008   : 1851   2002-11-07:   39   20200069:   39   Driving Layup Shot  : 1628
 2007   : 1819   2006-01-22:   39   20500591:   39   Turnaround Jump Shot:  891
 2009   : 1772   2008-01-14:   37   20700553:   37   Fadeaway Jump Shot  :  872
 2001   : 1708   2010-01-08:   36   20900527:   36   Running Jump Shot   :  779
 (Other):14771   (Other)   :25462   (Other) :25462   (Other)            : 3537
combined_shot_type     loc_x              loc_y        shot_distance   three_pt
Bank Shot:  120    Min.   :-250.000   Min.   :-44.00   Min.   : 0.00   0:20285
Dunk     : 1056    1st Qu.: -67.000   1st Qu.:  4.00   1st Qu.: 5.00   1: 5412
Hook Shot:  127    Median :   0.000   Median : 74.00   Median :15.00
Jump Shot:19710    Mean   :   7.148   Mean   : 91.26   Mean   :13.46
Layup    : 4532    3rd Qu.:  94.000   3rd Qu.:160.00   3rd Qu.:21.00
Tip Shot :  152    Max.   : 248.000   Max.   :791.00   Max.   :79.00

time_remaining ot        home       playoffs     opponent     shot_made_flag
Min.   :   0   0:25380   0:13212   0:21939   SAS    : 1638   0:14232
1st Qu.: 720   1:  317   1:12485   1: 3758   PHX    : 1535   1:11465
Median :1378                                 HOU    : 1399
Mean   :1397                                 SAC    : 1397
3rd Qu.:2171                                 DEN    : 1352
Max.   :2874                                 POR    : 1292
                                             (Other):17084
```

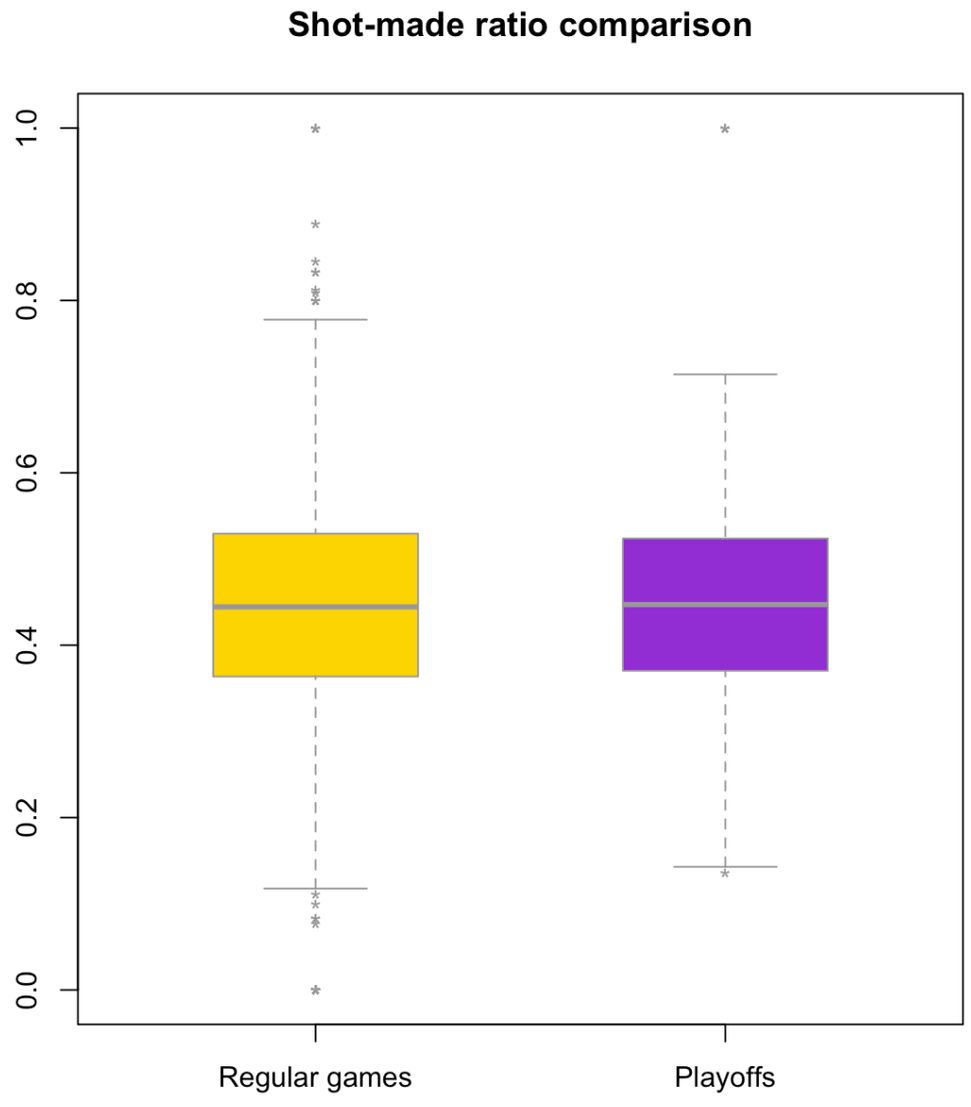
Comparisons and Hypothesis Tests

- home v.s. away
- paired: FALSE
- p-value of variance test: 0.7318
- var.equal: TRUE
- p-value of two sample t-test: 0.0007103
- result: different means



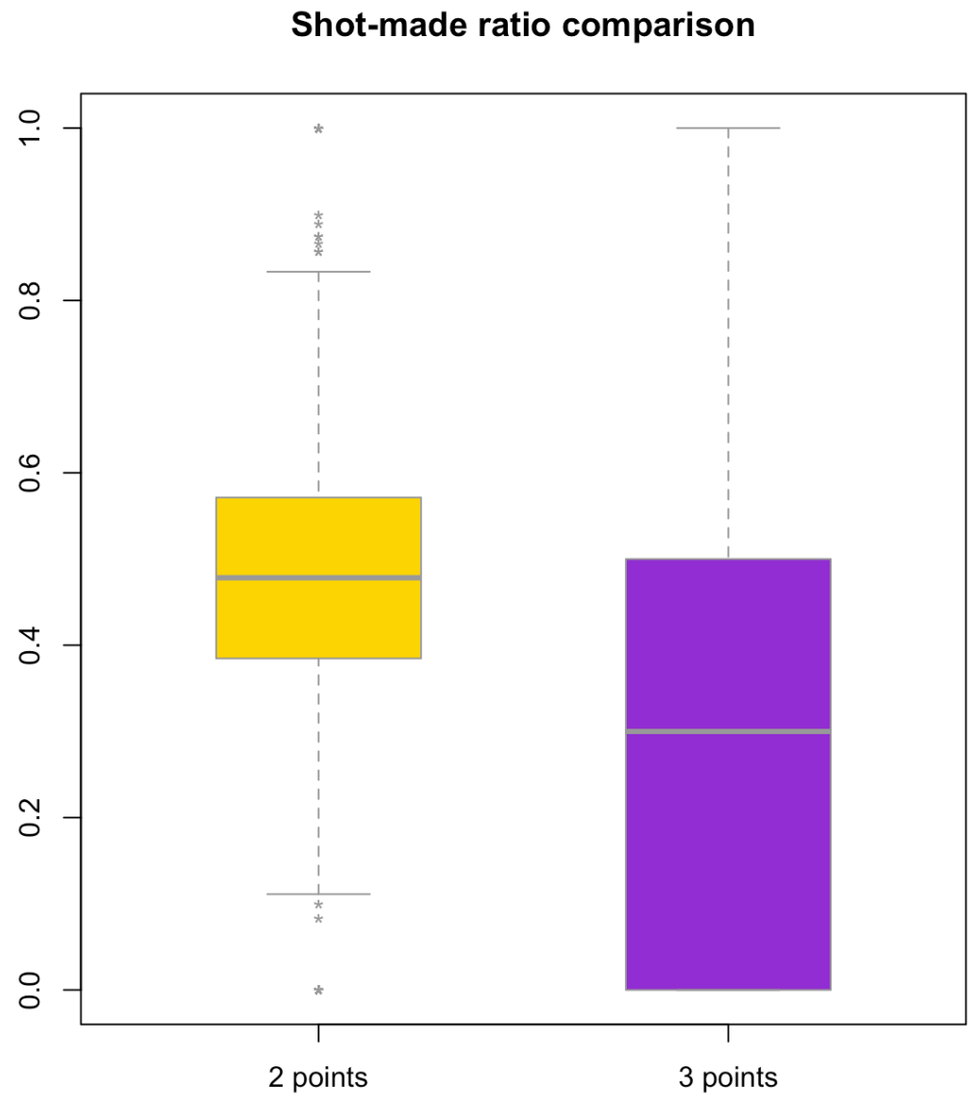
Comparisons and Hypothesis Tests (cont.)

- regular games v.s. playoffs
- paired: FALSE
- p-value of variance test: 0.2206
- var.equal: TRUE
- p-value of two sample t-test: 0.5658
- result: same means



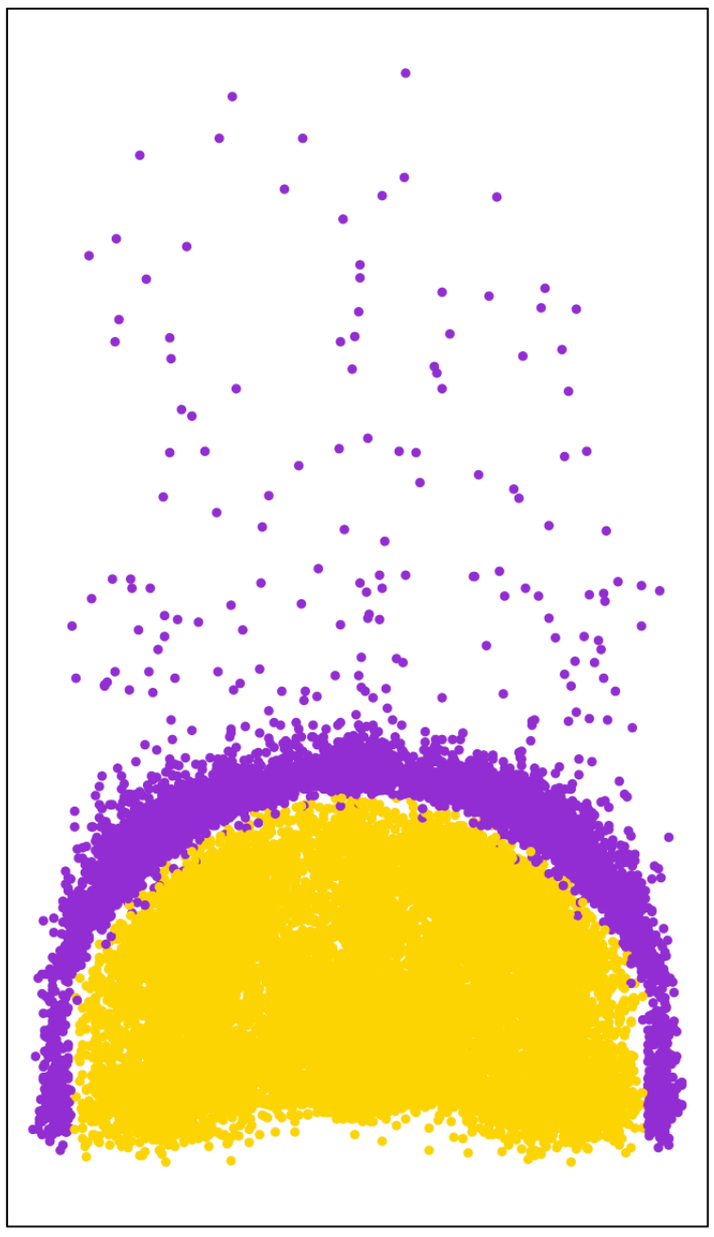
Comparisons and Hypothesis Tests (cont.)

- 2 pt. v.s. 3 pt.
- paired: TRUE
- p-value of variance test:
  < 2.2e-16
- var.equal: FALSE
- p-value of two sample t-test:
  < 2.2e-16
- result: different means

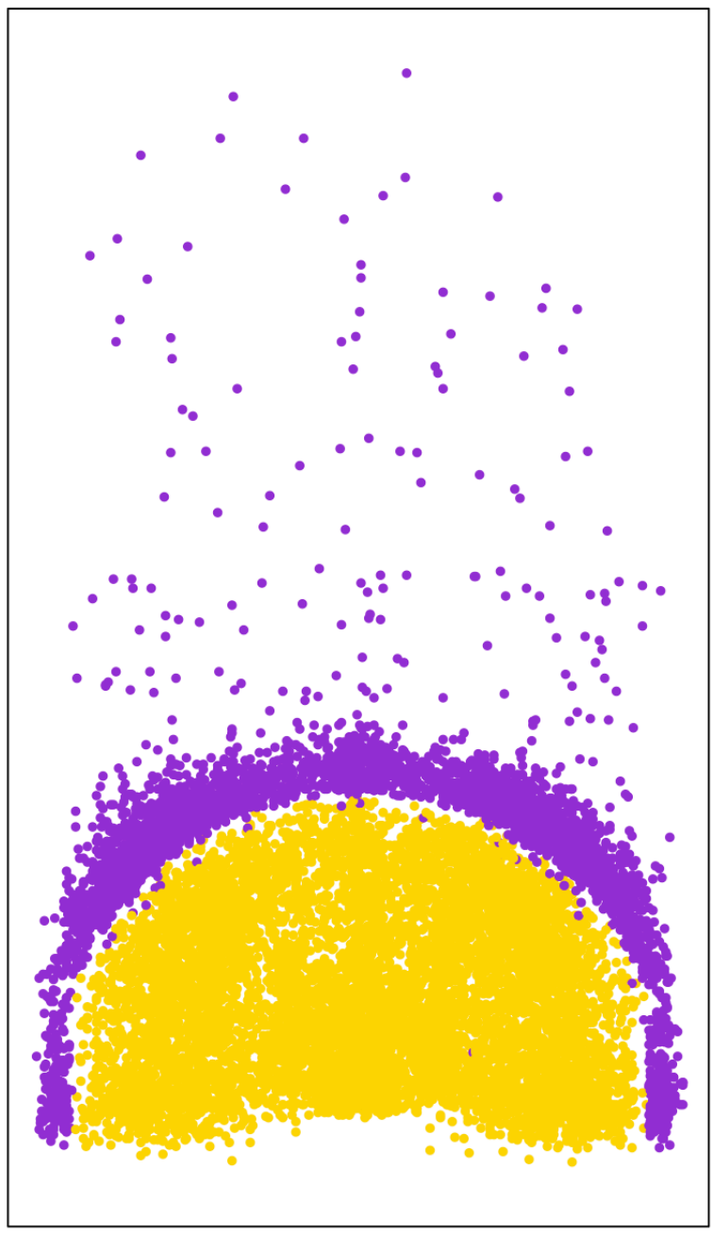


Shots’ Locations Comparison

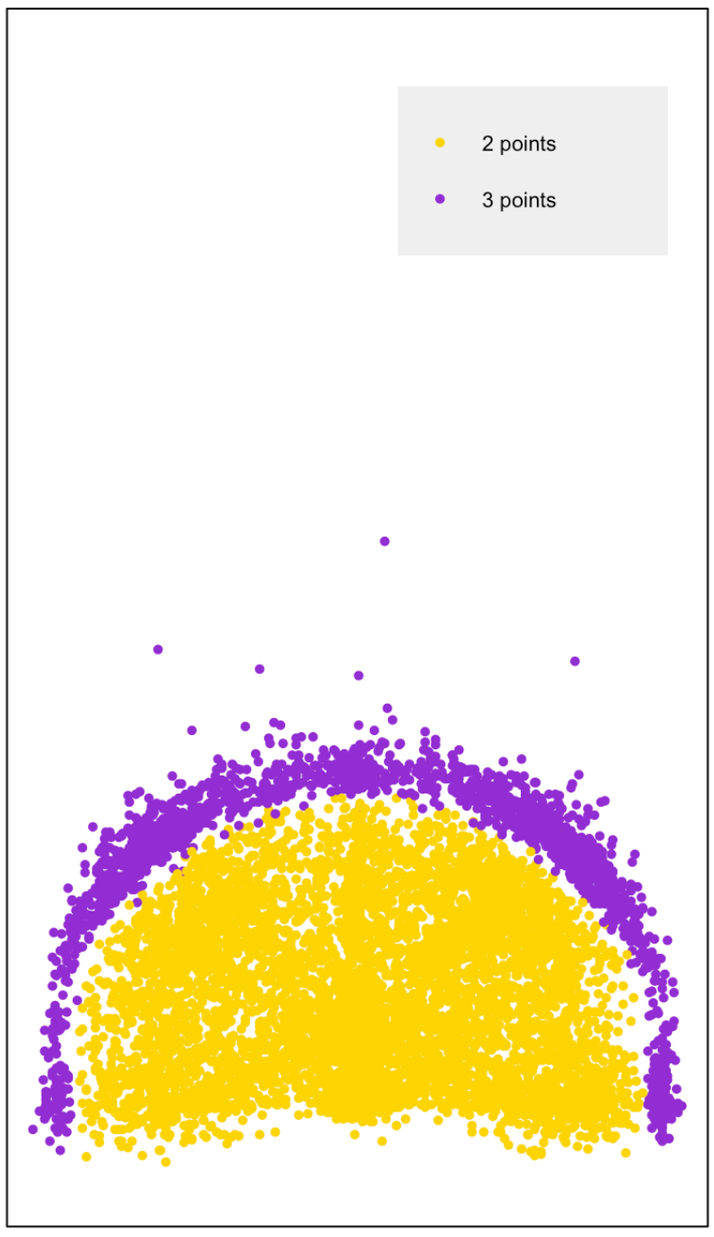
All the shots



All the failure shots



All the made shots



KNN Classifier

$$Pr(Y = j|X = x_0) = \frac{1}{K} \sum_{i \in N_0} 1_{\{y_i = j\}}$$

- The K nearest neighbors classifier

- N_0 is a set collecting K points in the training data closet to the test observation x_0 .

KNN Classifier (cont.)

KNN Classifier (cont.)

1-nearest neighbour

7-nearest neighbour

21-nearest neighbour

49-nearest neighbour

KNN Classifier (cont.)

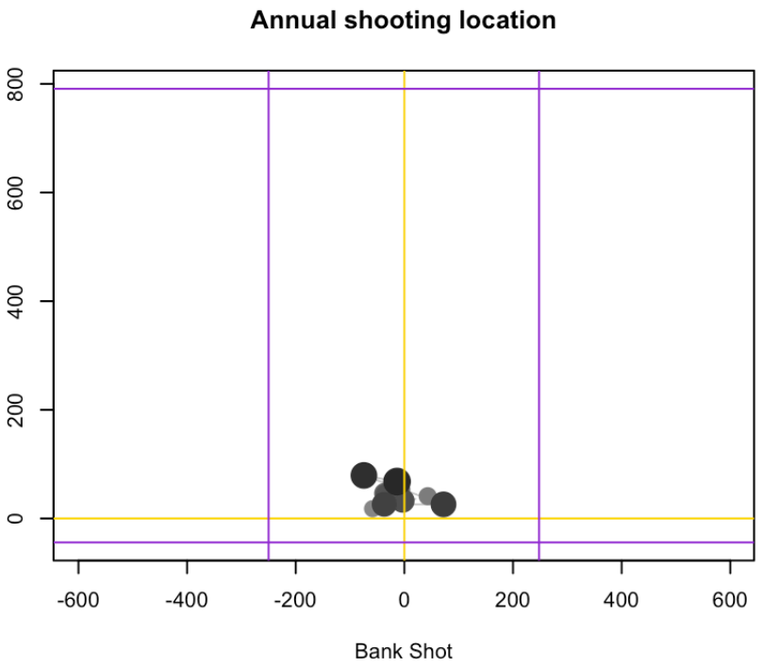
1-nearest neighbour

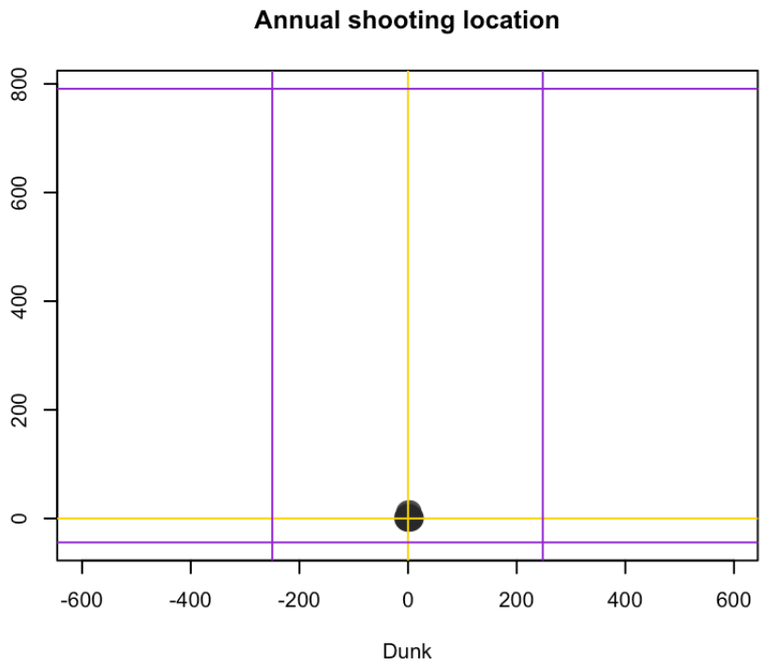
7-nearest neighbour

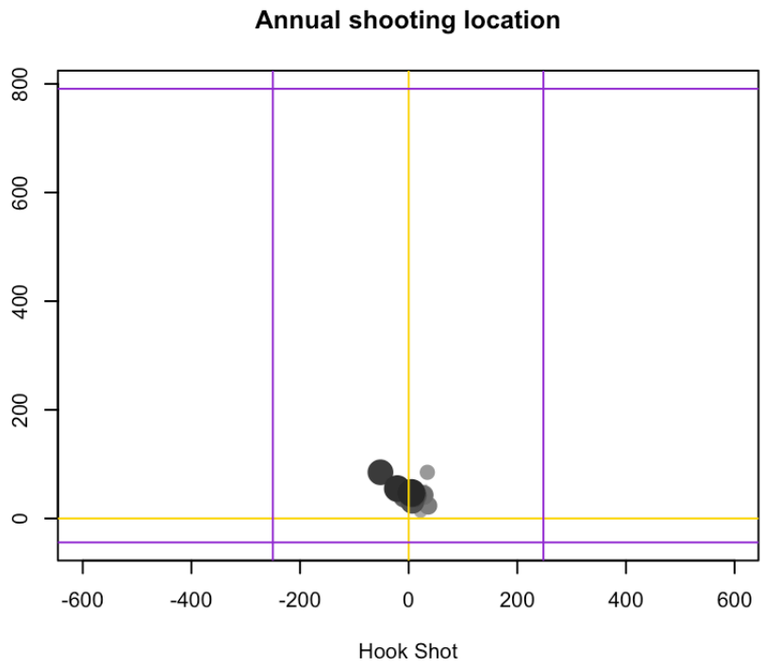
21-nearest neighbour

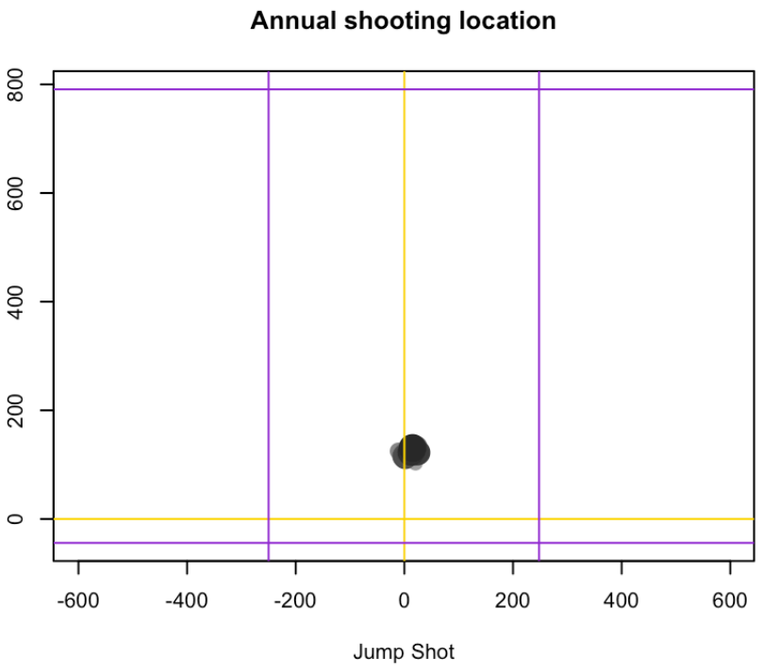
49-nearest neighbour

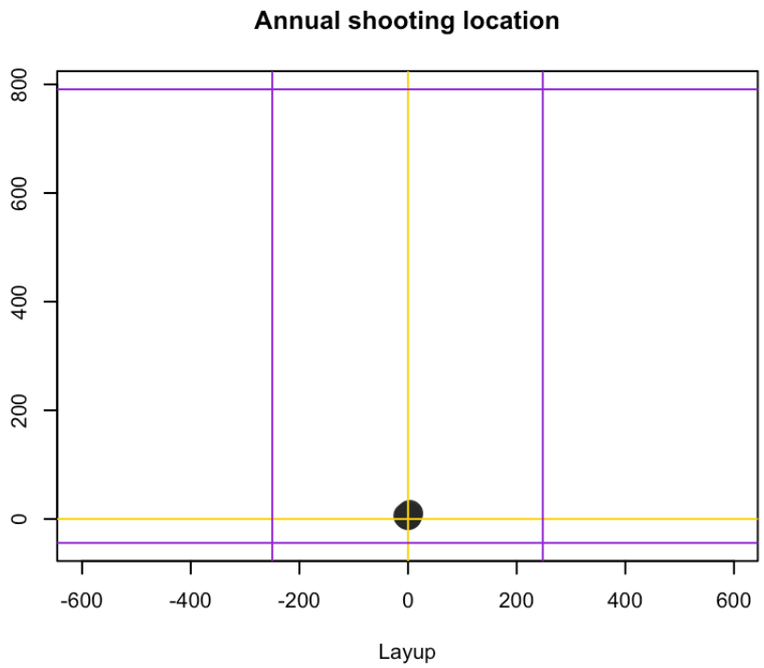
Annual Shots' Locations

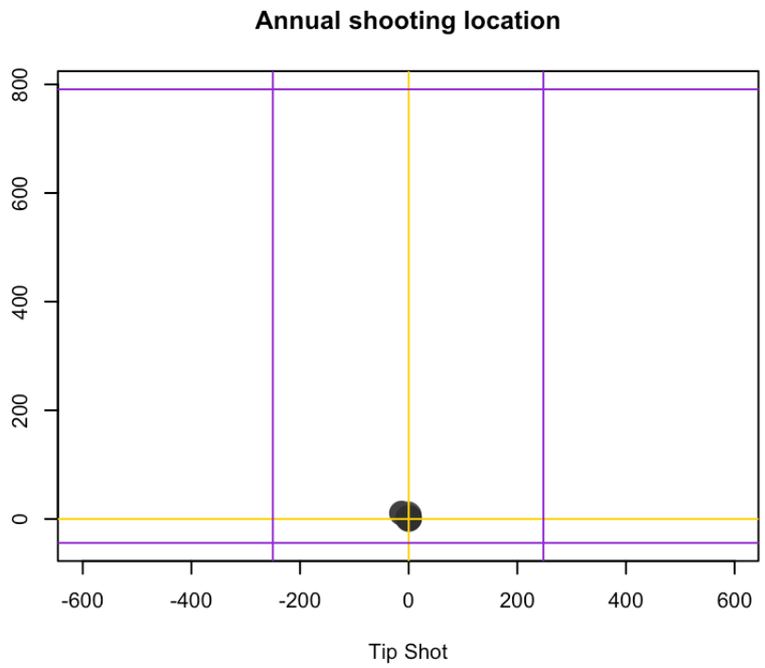




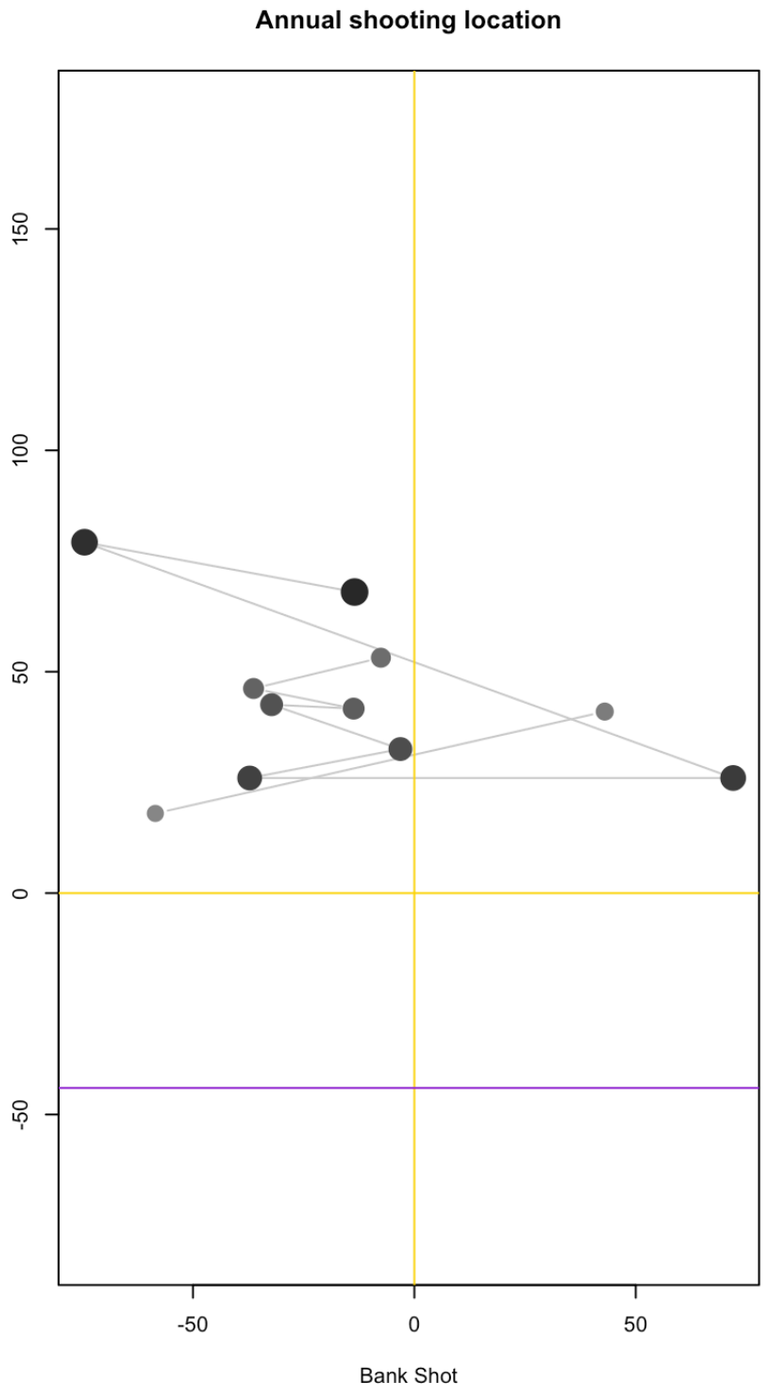


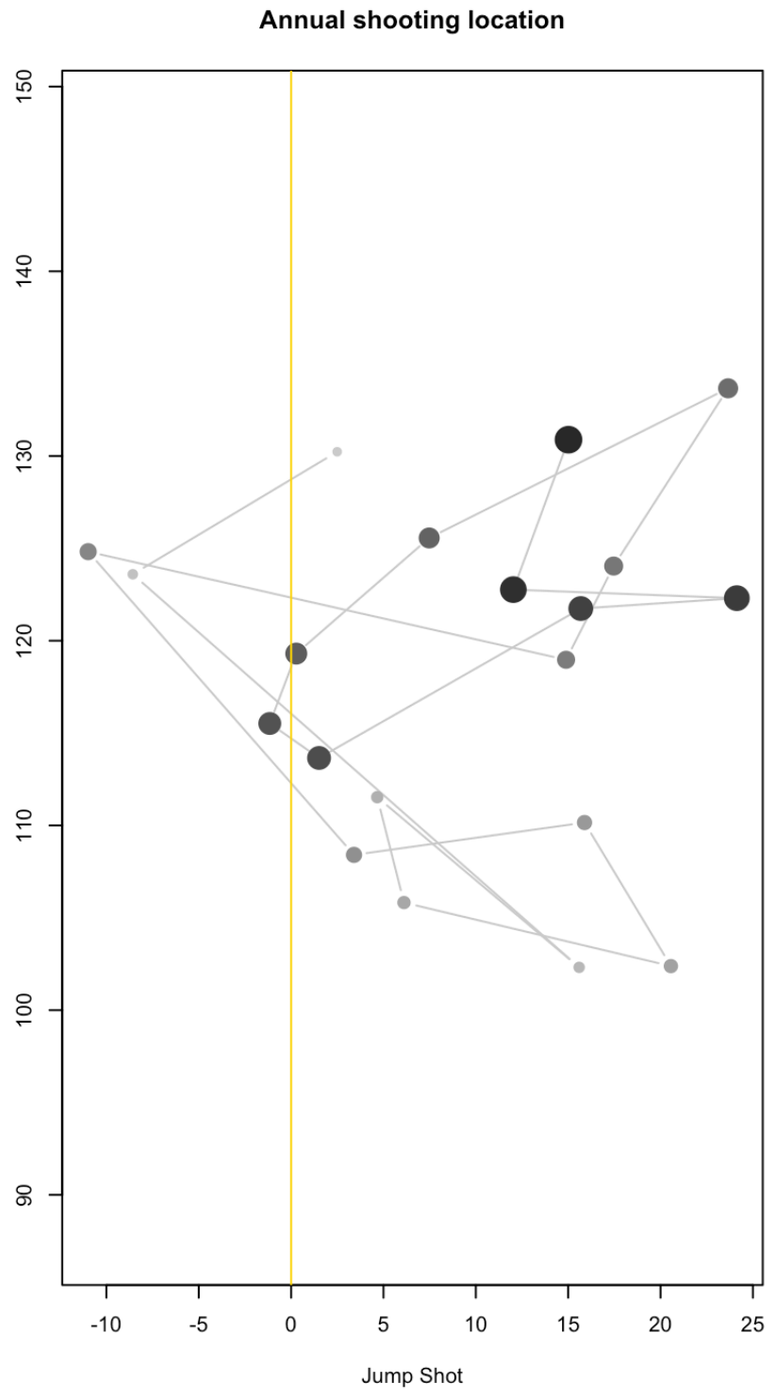


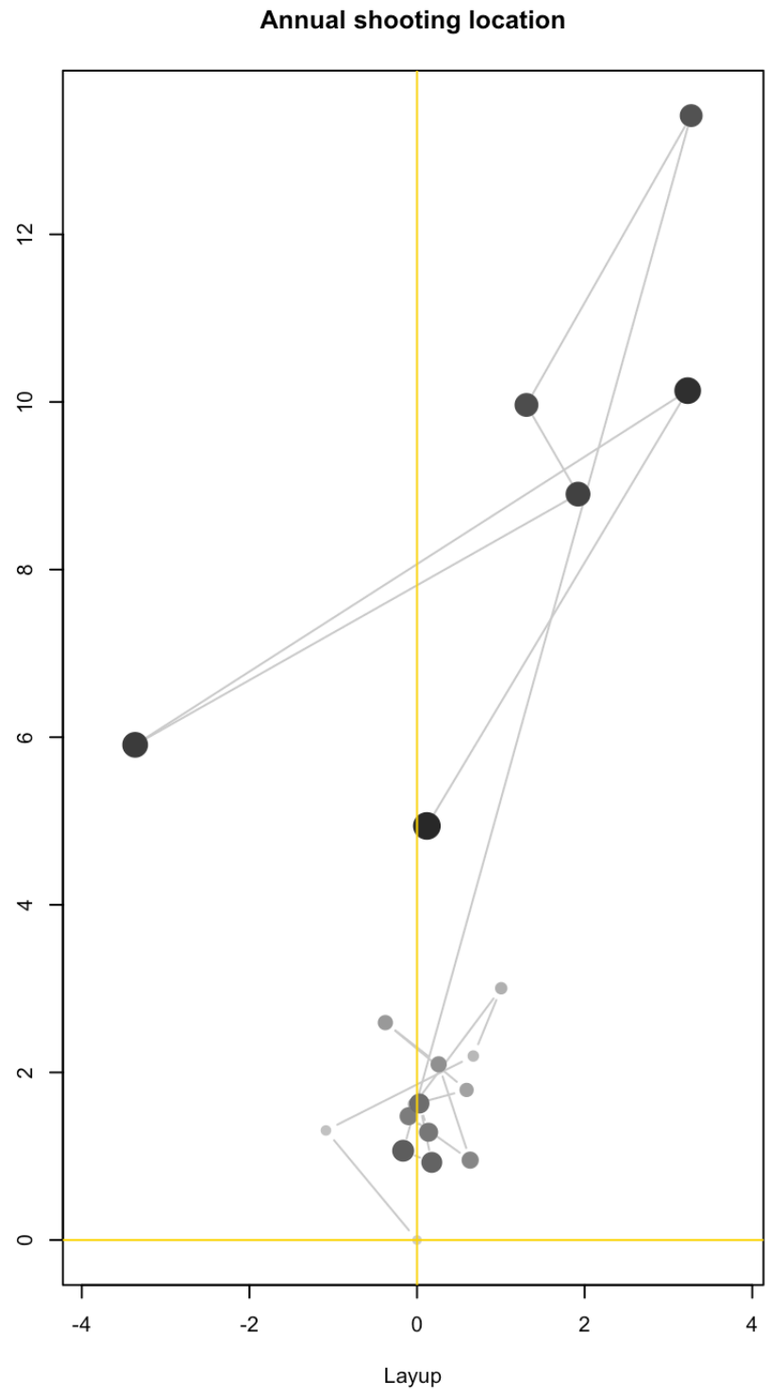




Annual Shots' Locations (cont.)







Future Works

- Logistic regression: model selection, PCA, collinearity analysis
- LDA/ QDA(linear/ quadratic discriminant analysis)
- Decision trees
- Time series analysis

Reference

- G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer, New York, 2013

- Kaggle, https://www.kaggle.com/c/kobe-bryant-shot-selection