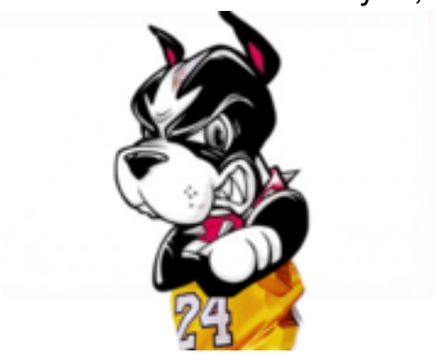
How Did Kobe Score?

BA810 - Fall 2019 - Team Project

Shangkun Zuo

10/17/2019

Team 6 Members: Alvaro Chinchayan, Leighton Li, Andrey Lifar, Yoki Liu, Yue Ping, Sherry Zuo

Our Logo: 

Overview

Our team is interested to analyze Kobe Bryant's performance as he is one of the most valuable players in NBA. Specifically, we take a look at all his shots from 1996 to 2016. Our original dataset has 25 columns including information about shot types, ranges, game time and so on. After data cleaning, we run several models on 94 columns and caculate MSEs to determine the best fit.
We use Dataset Kobe (https://www.kaggle.com/c/kobe-bryant-shot-selection/data)

Table of Content

    - Full model
- Stepwise regression
    - Forward selection
    - Backward selection
- Penalized regression
    - Ridge
    - Lasso
- Trees
- Random Forest
- Boosting

3. Summary

Dataset Understanding

Set Up

```
#loading libraries
library(tidyverse)
library(dplyr)
library(tidyr)
library(magrittr)
library(ggplot2)
library(ggthemes)
library(glmnet)
library(lubridate)
library(fastDummies)
library(MASS)
library(rpart)
library(rpart.plot)
library(randomForest)
library(gbm)
theme_set(theme_bw())
library(caret)
library(leaps)
library(ggvis)
library(dvmisc)
```

Load the Kobe Dataset

```
kobe<-read_csv("data.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   action_type = col_character(),
##   combined_shot_type = col_character(),
##   season = col_character(),
##   shot_type = col_character(),
##   shot_zone_area = col_character(),
##   shot_zone_basic = col_character(),
##   shot_zone_range = col_character(),
##   team_name = col_character(),
##   game_date = col_date(format = ""),
##   matchup = col_character(),
##   opponent = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
dim(kobe)
```

```
## [1] 30697    25
```

Data Cleaning

```
any(is.na(kobe))
complete.cases(kobe)
Kobe<-na.omit(kobe)
```

```
## add new column time_remaining as seconds
Kobe<-Kobe%>%mutate(time_remaining=minutes_remaining*60+seconds_remaining)
## add new columns year, month, and day
Kobe<-Kobe %>% mutate(year=year(game_date), month=month(game_date),
                 day=day(game_date))
## add new column home to see does Lakers is home or not
Kobe<-Kobe %>% mutate(home = case_when( grepl("@", matchup) ~ 1, grepl("vs.", mat
chup) ~ 0))
## add angle for the shot
Kobe<-Kobe%>%mutate(angle=(atan(abs(loc_y/loc_x))*180/pi))
## add angle range for the shot
Kobe<-Kobe %>%
  mutate(angle_range = case_when(angle <= 30 ~ "0-30 degrees",
                                 angle <= 60 ~ "30-60 degrees",
                                 angle <= 90 ~ "60-90 degrees",
                                 is.na(angle)==TRUE ~ "basket"))
Kobe$angle_range<-factor(Kobe$angle_range, levels=c("0-30 degrees", "30-60 degree
s", "60-90 degrees", "basket"))
## change the order of levels
Kobe$shot_zone_range<-factor(Kobe$shot_zone_range, levels=c("Less Than 8 ft.", "8
-16 ft.", "16-24 ft.", "24+ ft.", "Back Court Shot"))
```

Preview of Data

Summary Stats

```
dim(Kobe)
```

```
## [1] 25697    32
```

Our dataset has 25697 rows, 32 columns.

```
colnames(Kobe)
```

```
##  [1] "action_type"        "combined_shot_type" "game_event_id"
##  [4] "game_id"            "lat"                "loc_x"
##  [7] "loc_y"              "lon"                "minutes_remaining"
## [10] "period"             "playoffs"           "season"
## [13] "seconds_remaining"  "shot_distance"      "shot_made_flag"
## [16] "shot_type"          "shot_zone_area"     "shot_zone_basic"
## [19] "shot_zone_range"    "team_id"            "team_name"
## [22] "game_date"          "matchup"            "opponent"
## [25] "shot_id"            "time_remaining"     "year"
## [28] "month"              "day"                "home"
## [31] "angle"              "angle_range"
```

```
head(Kobe)
```

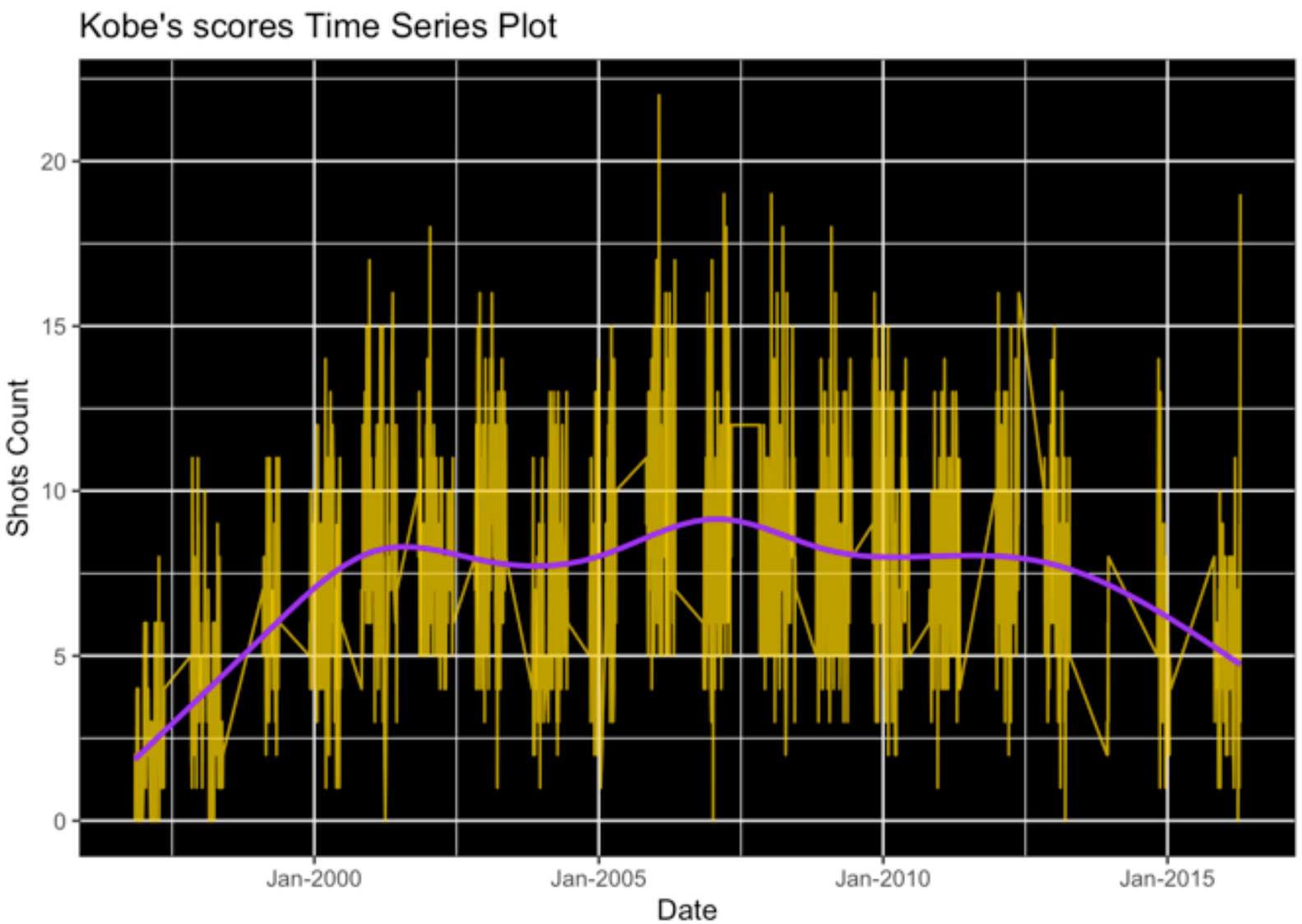
```
## # A tibble: 6 x 32
##   action_type combined_shot_t… game_event_id game_id   lat loc_x loc_y
##   <chr>       <chr>                   <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1 Jump Shot   Jump Shot                  12  2.00e7  34.0  -157     0
## 2 Jump Shot   Jump Shot                  35  2.00e7  33.9  -101   135
## 3 Jump Shot   Jump Shot                  43  2.00e7  33.9   138   175
## 4 Driving Du… Dunk                      155  2.00e7  34.0     0     0
## 5 Jump Shot   Jump Shot                 244  2.00e7  34.1  -145   -11
## 6 Layup Shot  Layup                    251  2.00e7  34.0     0     0
## # … with 25 more variables: lon <dbl>, minutes_remaining <dbl>,
## #   period <dbl>, playoffs <dbl>, season <chr>, seconds_remaining <dbl>,
## #   shot_distance <dbl>, shot_made_flag <dbl>, shot_type <chr>,
## #   shot_zone_area <chr>, shot_zone_basic <chr>, shot_zone_range <fct>,
## #   team_id <dbl>, team_name <chr>, game_date <date>, matchup <chr>,
## #   opponent <chr>, shot_id <dbl>, time_remaining <dbl>, year <dbl>,
## #   month <dbl>, day <int>, home <dbl>, angle <dbl>, angle_range <fct>
```

```
str(Kobe)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    25697 obs. of  32 variables:
##  $ action_type       : chr  "Jump Shot" "Jump Shot" "Jump Shot" "Driving Dunk
Shot" ...
##  $ combined_shot_type: chr  "Jump Shot" "Jump Shot" "Jump Shot" "Dunk" ...
##  $ game_event_id     : num  12 35 43 155 244 251 265 294 309 4 ...
##  $ game_id           : num  2e+07 2e+07 2e+07 2e+07 2e+07 ...
##  $ lat               : num  34 33.9 33.9 34 34.1 ...
##  $ loc_x             : num  -157 -101 138 0 -145 0 -65 -33 -94 121 ...
##  $ loc_y             : num  0 135 175 0 -11 0 108 125 238 127 ...
##  $ lon               : num  -118 -118 -118 -118 -118 ...
##  $ minutes_remaining : num  10 7 6 6 9 8 6 3 1 11 ...
##  $ period            : num  1 1 1 2 3 3 3 3 3 1 ...
##  $ playoffs          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ season            : chr  "2000-01" "2000-01" "2000-01" "2000-01" ...
##  $ seconds_remaining : num  22 45 52 19 32 52 12 36 56 0 ...
##  $ shot_distance     : num  15 16 22 0 14 0 12 12 25 17 ...
##  $ shot_made_flag    : num  0 1 0 1 0 1 1 0 0 1 ...
##  $ shot_type         : chr  "2PT Field Goal" "2PT Field Goal" "2PT Field Goal"
"2PT Field Goal" ...
##  $ shot_zone_area    : chr  "Left Side(L)" "Left Side Center(LC)" "Right Side
Center(RC)" "Center(C)" ...
##  $ shot_zone_basic   : chr  "Mid-Range" "Mid-Range" "Mid-Range" "Restricted Ar
ea" ...
##  $ shot_zone_range   : Factor w/ 5 levels "Less Than 8 ft.",..: 2 3 3 1 2 1 2
2 4 3 ...
##  $ team_id           : num  1.61e+09 1.61e+09 1.61e+09 1.61e+09 1.61e+09 ...
##  $ team_name         : chr  "Los Angeles Lakers" "Los Angeles Lakers" "Los Ang
eles Lakers" "Los Angeles Lakers" ...
##  $ game_date         : Date, format: "2000-10-31" "2000-10-31" ...
##  $ matchup           : chr  "LAL @ POR" "LAL @ POR" "LAL @ POR" "LAL @ POR" ..
.
##  $ opponent          : chr  "POR" "POR" "POR" "POR" ...
##  $ shot_id           : num  2 3 4 5 6 7 9 10 11 12 ...
##  $ time_remaining    : num  622 465 412 379 572 532 372 216 116 660 ...
##  $ year              : num  2000 2000 2000 2000 2000 2000 2000 2000 2000 2000
...
##  $ month             : num  10 10 10 10 10 10 10 10 10 11 ...
##  $ day               : int  31 31 31 31 31 31 31 31 31 1 ...
##  $ home              : num  1 1 1 1 1 1 1 1 1 0 ...
##  $ angle             : num  0 53.2 51.74 NaN 4.34 ...
##  $ angle_range       : Factor w/ 4 levels "0-30 degrees",..: 1 2 2 4 1 4 2 3 3
2 ...
```

Describing the Data

```
Kobe %>%
  group_by(game_date)%>%
  summarize(shots = sum(shot_made_flag))%>%
  arrange(game_date)%>%
  ggplot(aes(x=game_date, y=shots)) +
    geom_line(color="gold", alpha=0.7) +
    geom_smooth(se=FALSE, col="purple")+
    labs(title="Kobe's scores Time Series Plot", x="Date",y="Shots Count")+
    scale_x_date(date_labels = "%b-%Y")+
    theme(panel.background = element_rect(fill="black"))
```



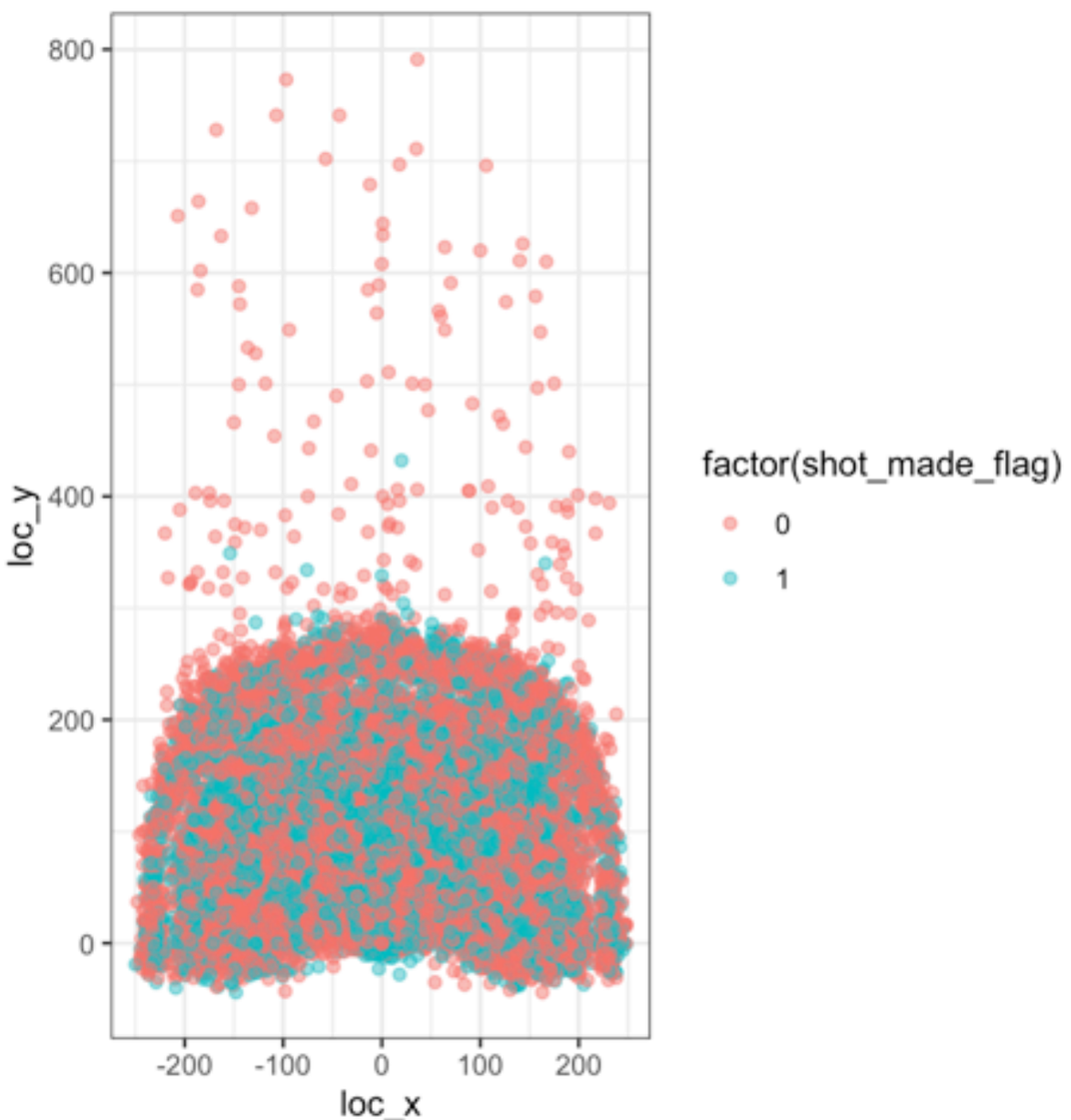

```
Kobe %>%
  group_by(year)%>%
  summarize(shots_percentage = sum(shot_made_flag)/n())%>%
  arrange(year)%>%
  ggplot(aes(x=year, y=shots_percentage)) +
    geom_line(color="purple", alpha=0.7) +
    geom_smooth(se=FALSE, col="gold")+
    labs(title="Kobe's scores percentage Time Series Plot", x="Year",y="Shots Per
centage")+
    theme(panel.background = element_rect(fill="black"))
```

Look through those categorical variables:
What's action_type?

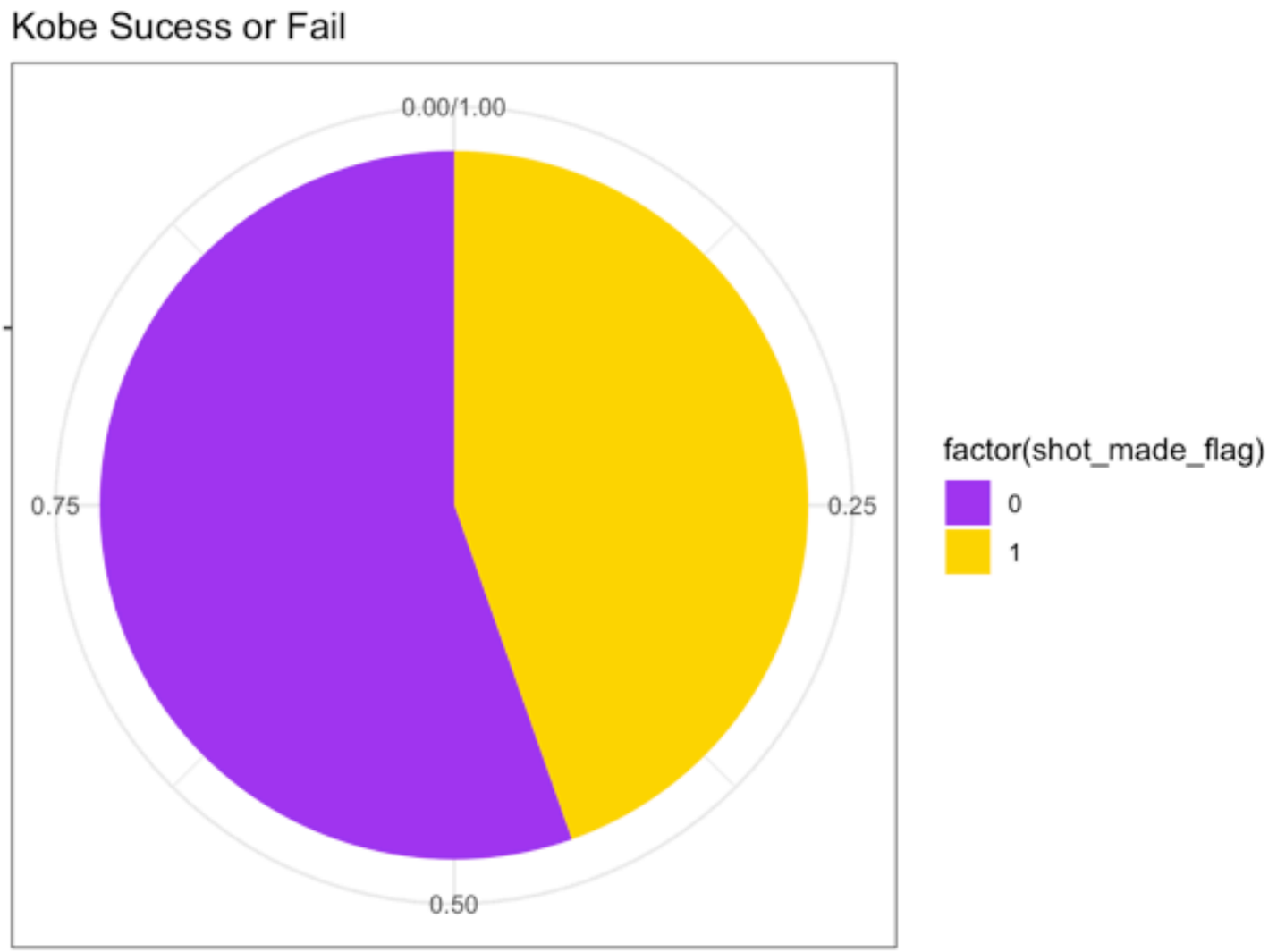
```
Kobe%>%
  group_by(action_type)%>%
  summarize(shots=sum(shot_made_flag))%>%
  arrange(desc(shots))
```

```
## # A tibble: 55 x 2
##    action_type          shots
##    <chr>                <dbl>
##  1 Jump Shot             5177
##  2 Driving Layup Shot    1207
##  3 Layup Shot             830
##  4 Running Jump Shot      582
##  5 Turnaround Jump Shot   533
##  6 Fadeaway Jump Shot     501
##  7 Slam Dunk Shot         328
##  8 Pullup Jump shot       291
##  9 Driving Dunk Shot      251
## 10 Jump Bank Shot         224
## # … with 45 more rows
```

```
ggplot(Kobe, aes(loc_x, loc_y, color=factor(shot_made_flag)))+
  geom_point(alpha=0.5)+
  theme(aspect.ratio = 1.9)
```



```
Kobe %>%
  group_by(shot_made_flag)%>%
  summarise(success=n())%>%
  ggplot(aes(x="", y=success, fill=factor(shot_made_flag)))+
  geom_bar(width = 1, stat="identity", position="fill")+
  coord_polar("y")+
  scale_fill_manual(values = c("purple", "gold"))+
  theme(panel.background = element_rect(fill="white"))+
  labs(title = "Kobe Sucess or Fail",x='',y='')
```

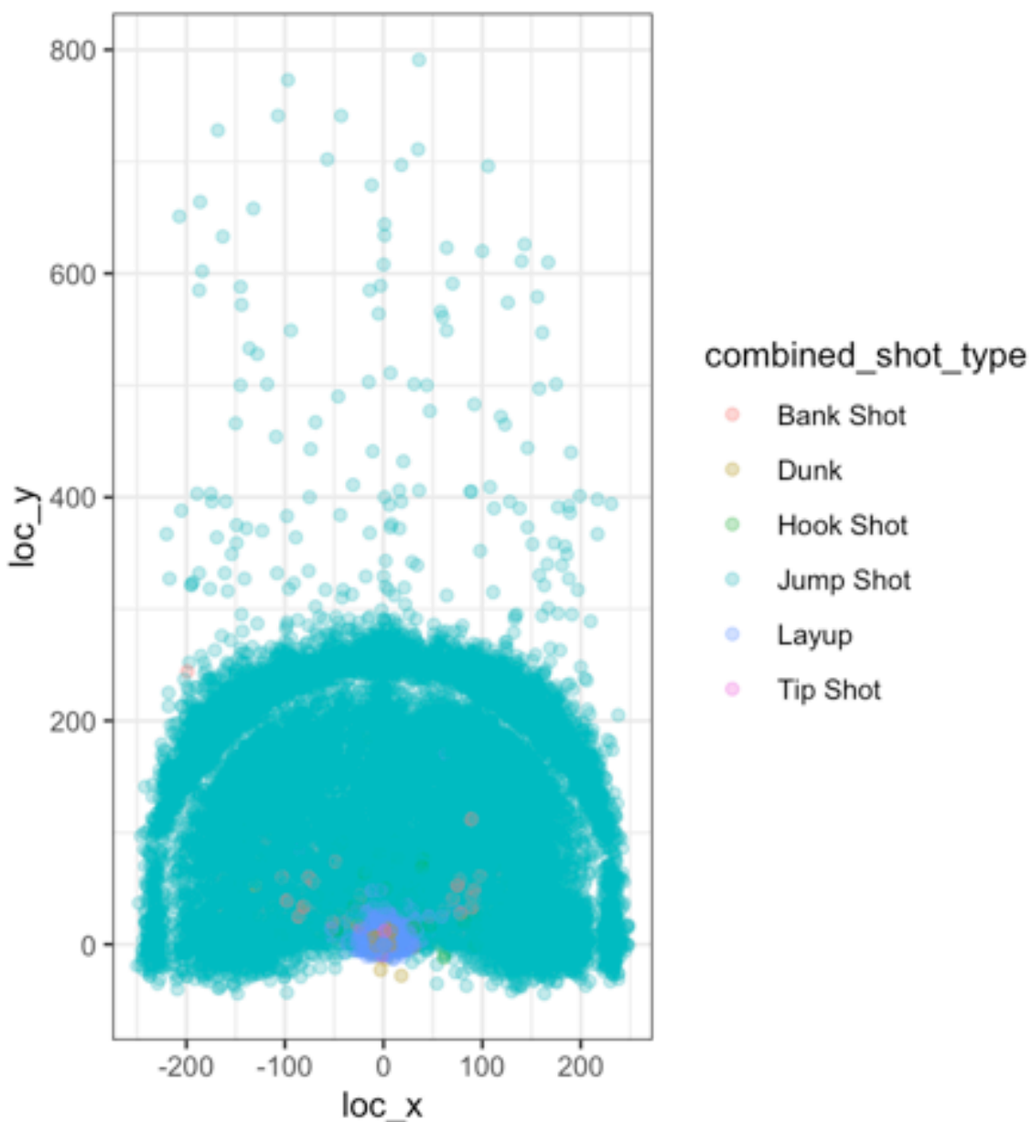


What's combined_shot_type

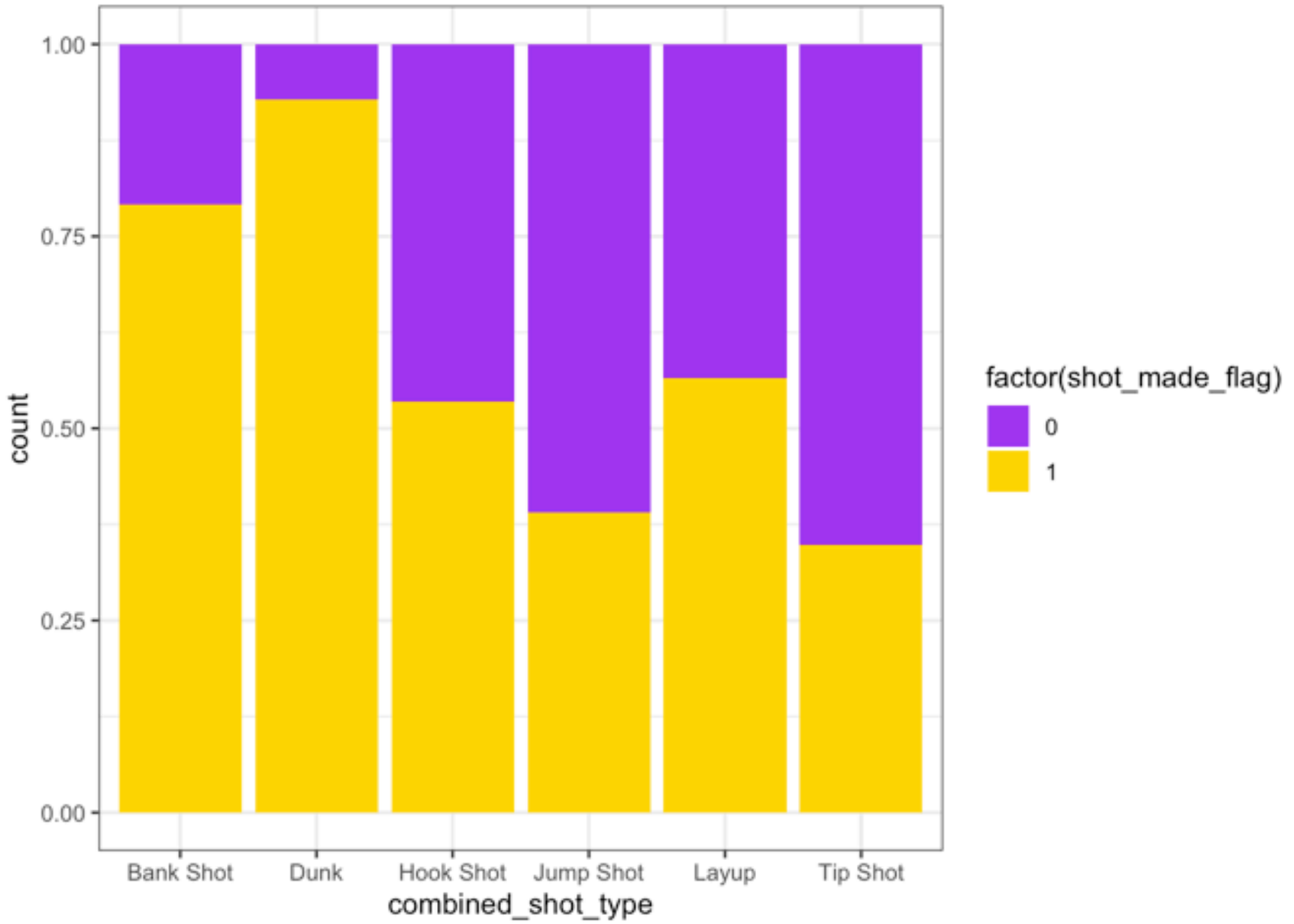
```
Kobe%>%
  group_by(combined_shot_type)%>%
  summarize(shots=sum(shot_made_flag))%>%
  arrange(desc(shots))
```

```
## # A tibble: 6 x 2
##   combined_shot_type shots
##   <chr>              <dbl>
## 1 Jump Shot           7708
## 2 Layup               2561
## 3 Dunk                 980
## 4 Bank Shot             95
## 5 Hook Shot             68
## 6 Tip Shot              53
```

```
ggplot(Kobe, aes(loc_x, loc_y, color=combined_shot_type))+
  geom_point(alpha=0.3)+
  theme(aspect.ratio = 1.9)
```




```
ggplot(Kobe, aes(x=combined_shot_type, fill=factor(shot_made_flag)))+
  geom_bar(position="fill")+
  scale_fill_manual(values = c("purple", "gold"))
```

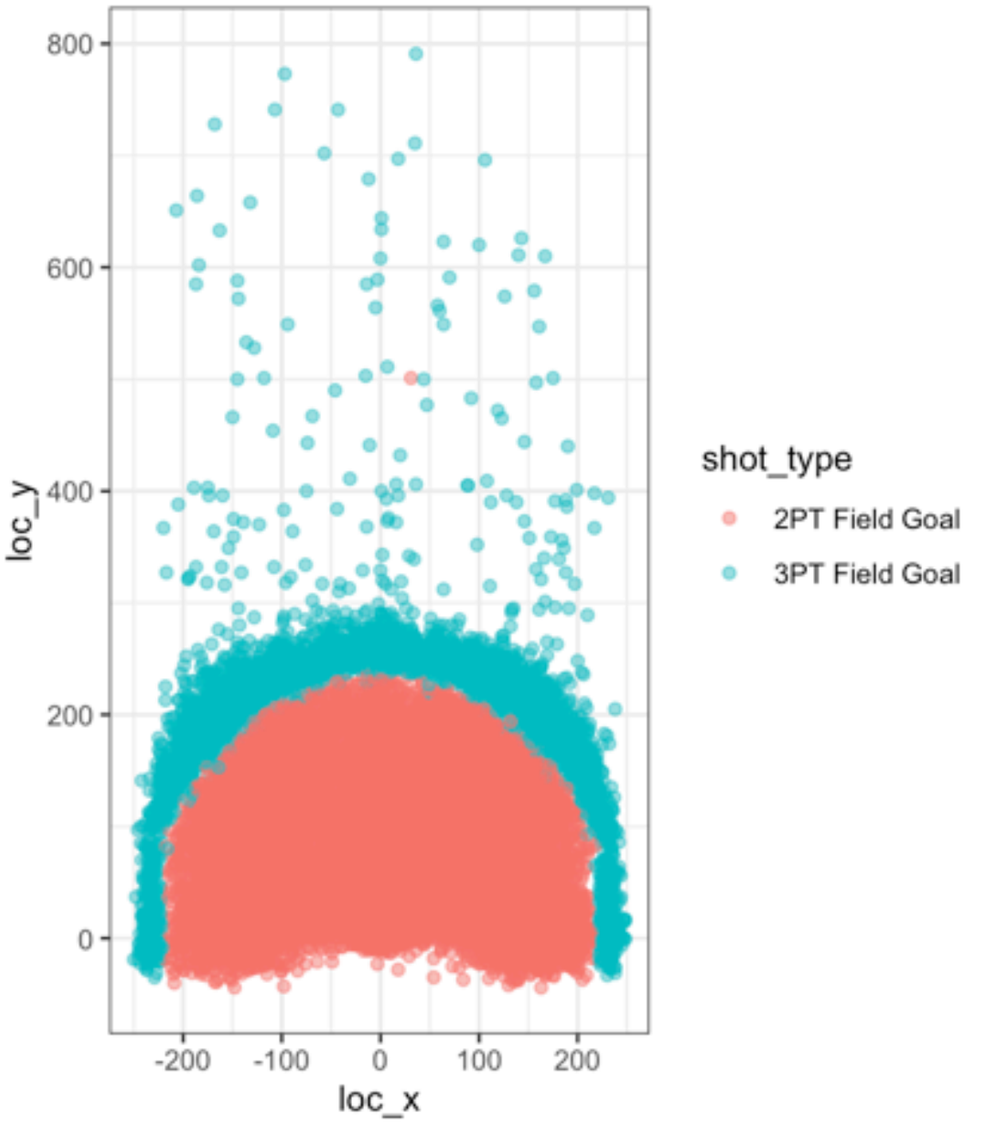


What's shot_type?

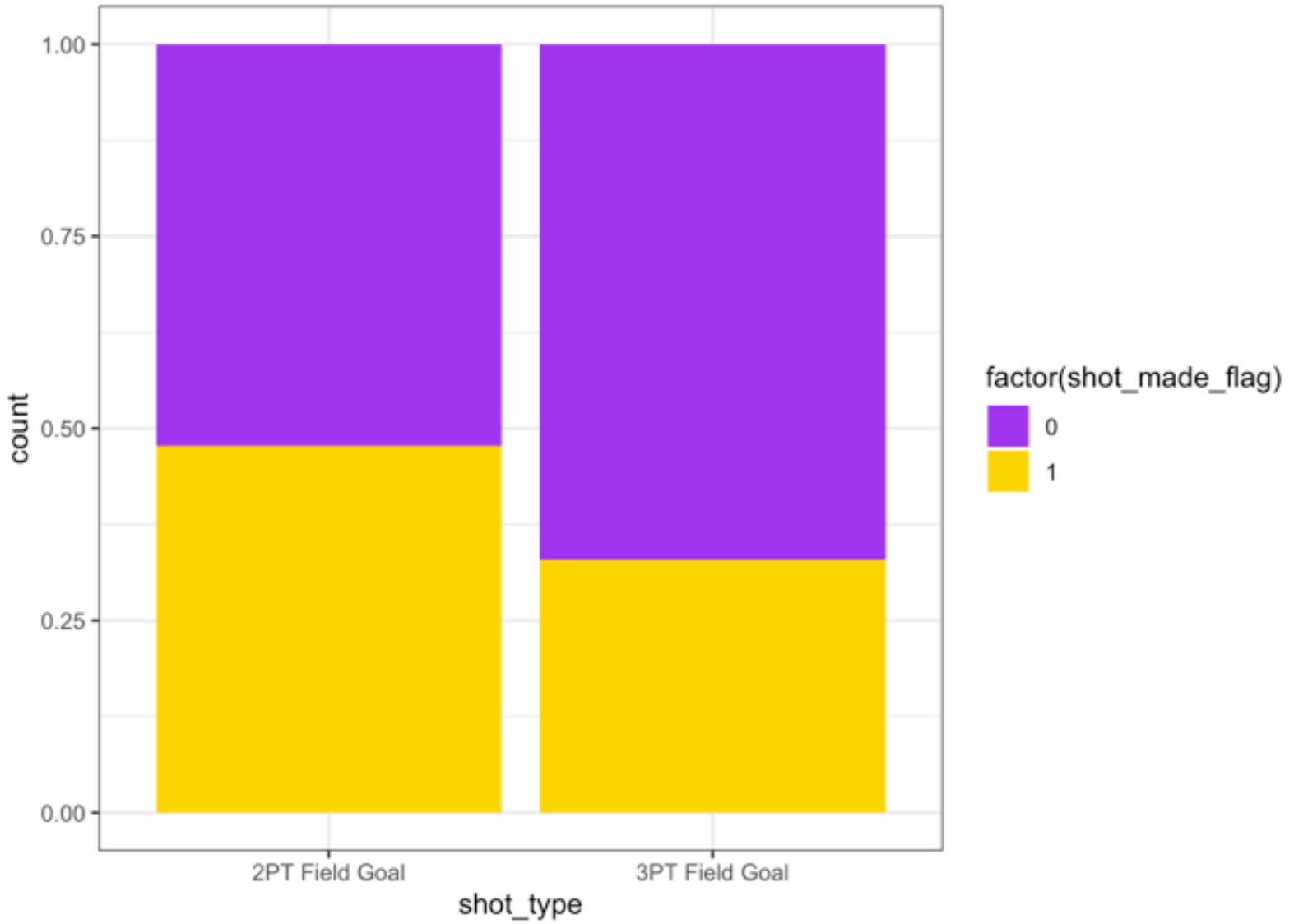
```
Kobe%>%
  group_by(shot_type)%>%
  summarize(shots=sum(shot_made_flag))%>%
  arrange(desc(shots))
```

```
## # A tibble: 2 x 2
##   shot_type      shots
##   <chr>          <dbl>
## 1 2PT Field Goal  9683
## 2 3PT Field Goal  1782
```

```
ggplot(Kobe, aes(loc_x, loc_y, color=shot_type))+
  geom_point(alpha=0.5)+
  theme(aspect.ratio = 1.9)
```




```
ggplot(Kobe, aes(x=shot_type, fill=factor(shot_made_flag)))+
  geom_bar(position="fill")+
  scale_fill_manual(values = c("purple", "gold"))
```

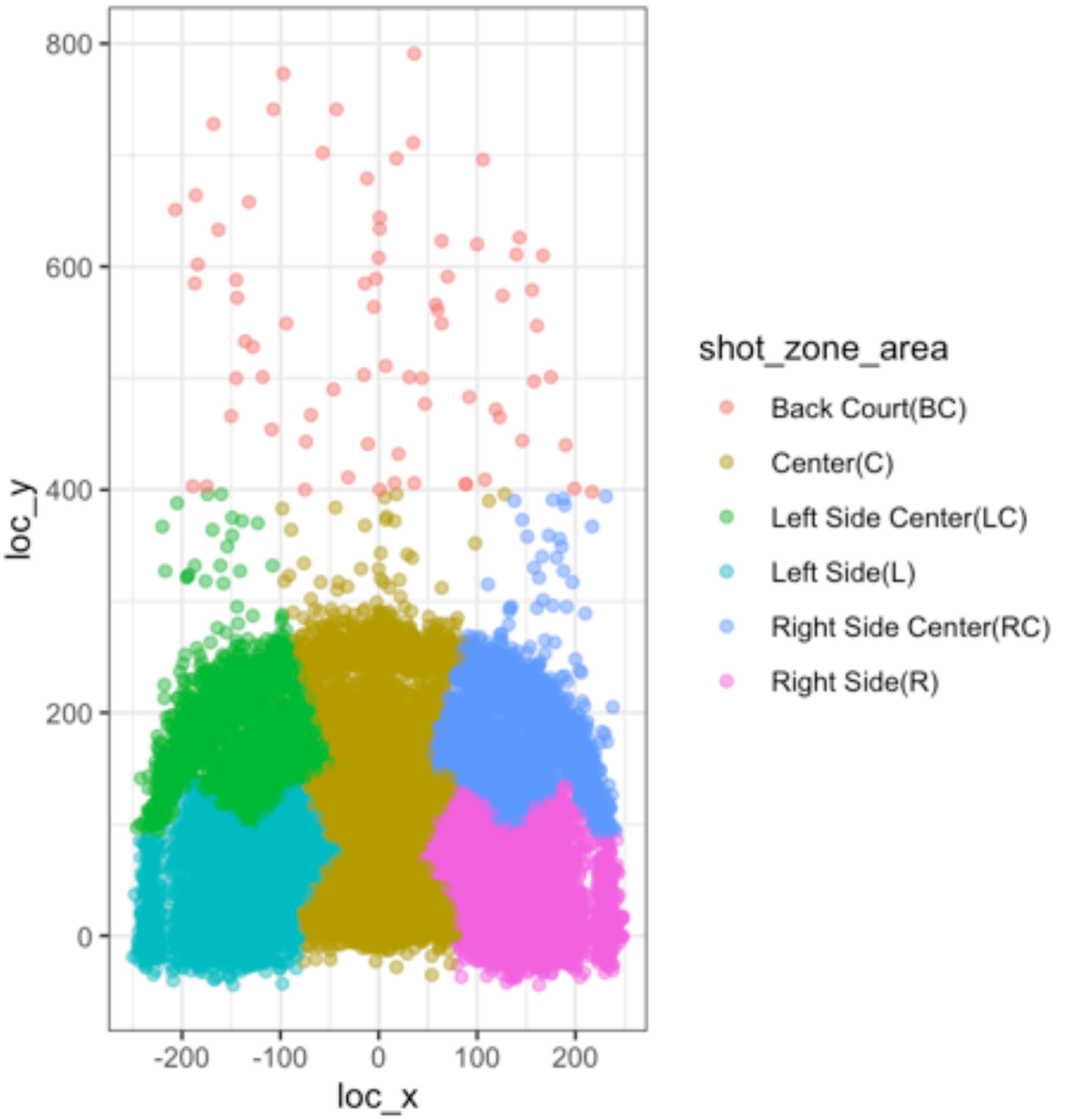


What's shot_zone_area?

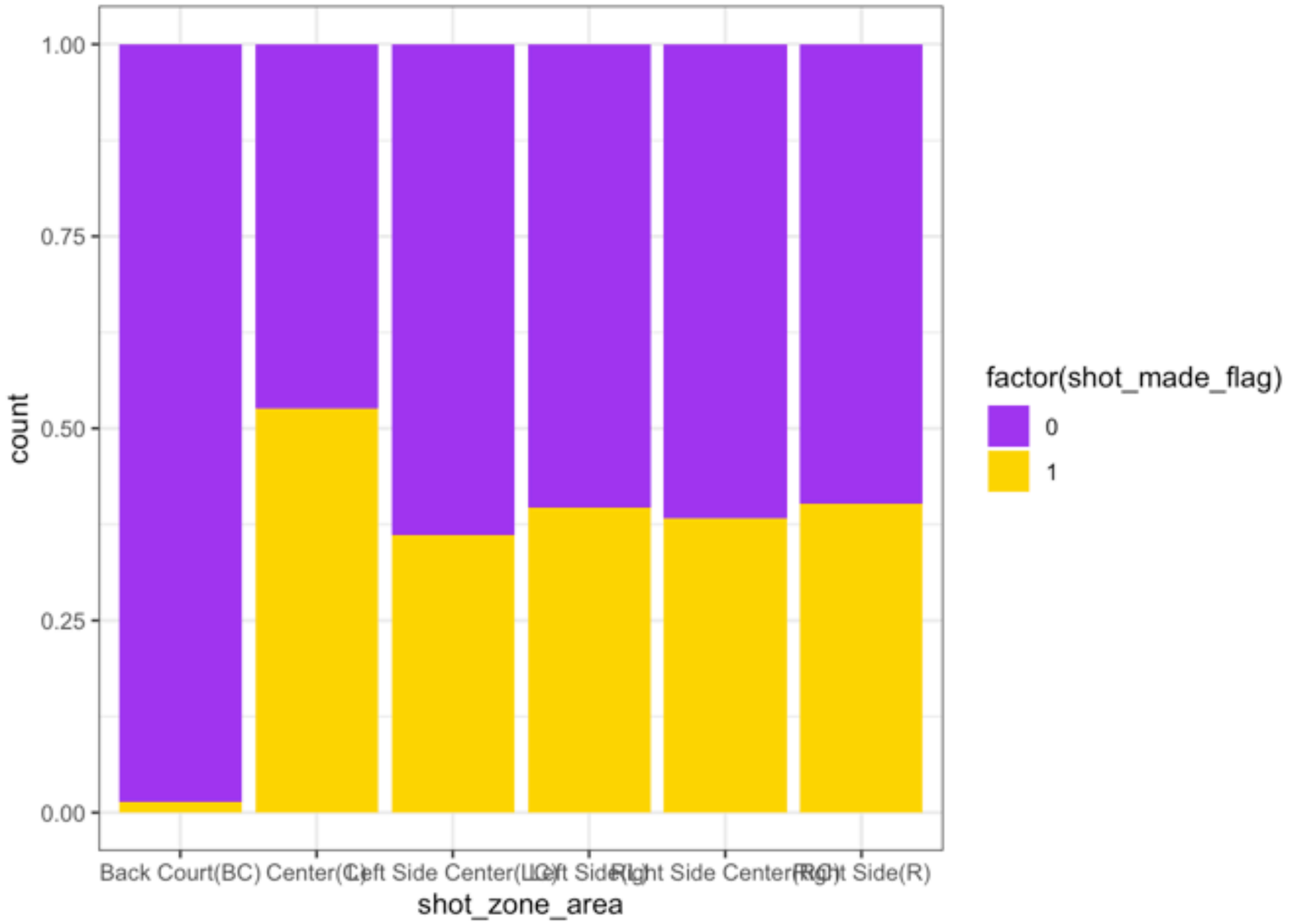
```
Kobe%>%
  group_by(shot_zone_area)%>%
  summarize(shots=sum(shot_made_flag))%>%
  arrange(desc(shots))
```

```
## # A tibble: 6 x 2
##   shot_zone_area        shots
##   <chr>                 <dbl>
## 1 Center(C)              5933
## 2 Right Side(R)          1550
## 3 Right Side Center(RC)  1523
## 4 Left Side(L)           1243
## 5 Left Side Center(LC)   1215
## 6 Back Court(BC)            1
```

```
ggplot(Kobe, aes(loc_x, loc_y, color=shot_zone_area))+
  geom_point(alpha=0.5)+
  theme(aspect.ratio = 1.9)
```




```
ggplot(Kobe, aes(x=shot_zone_area, fill=factor(shot_made_flag)))+
  geom_bar(position="fill")+
  scale_fill_manual(values = c("purple", "gold"))
```

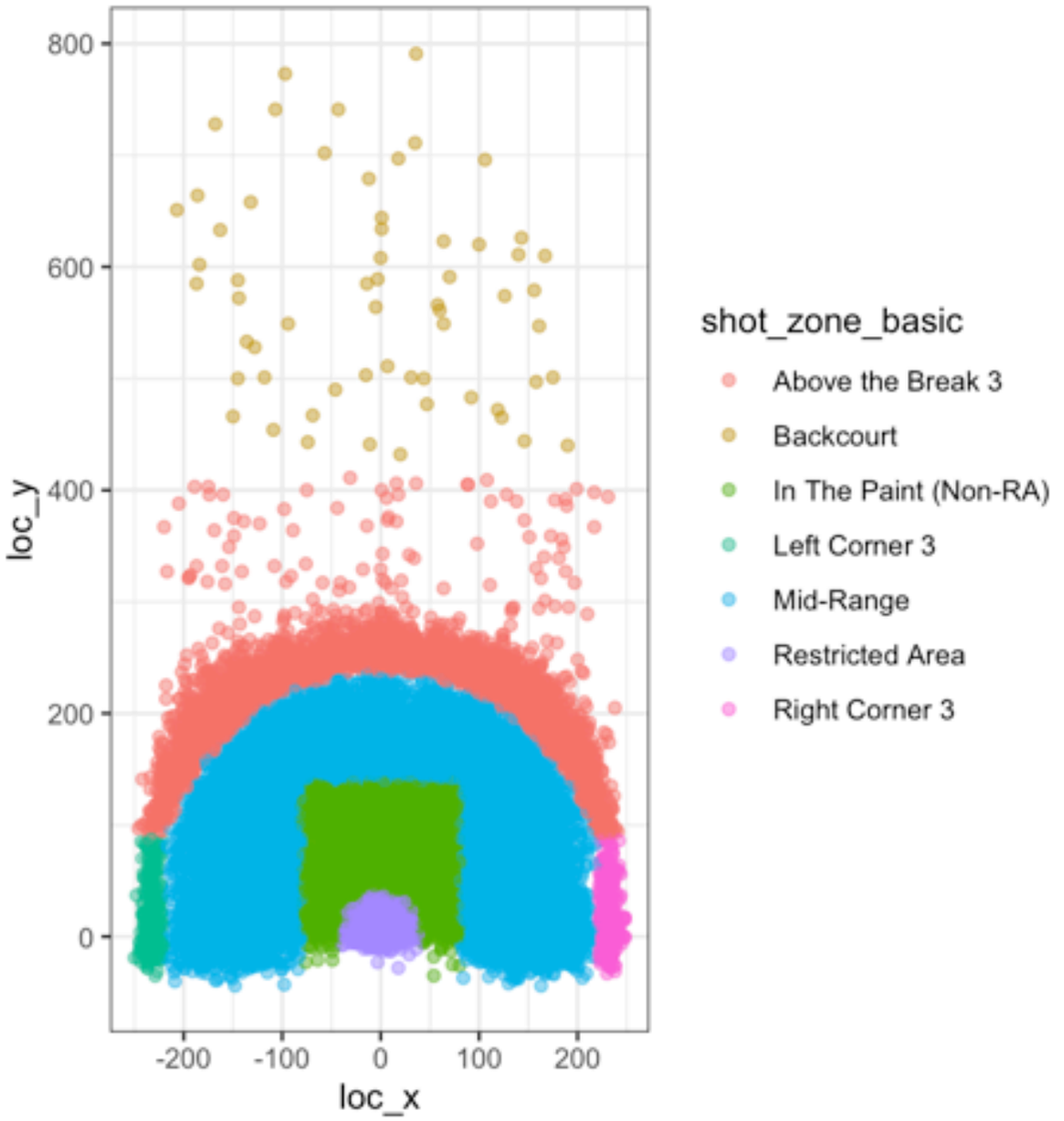


What's shot_zone_basic?

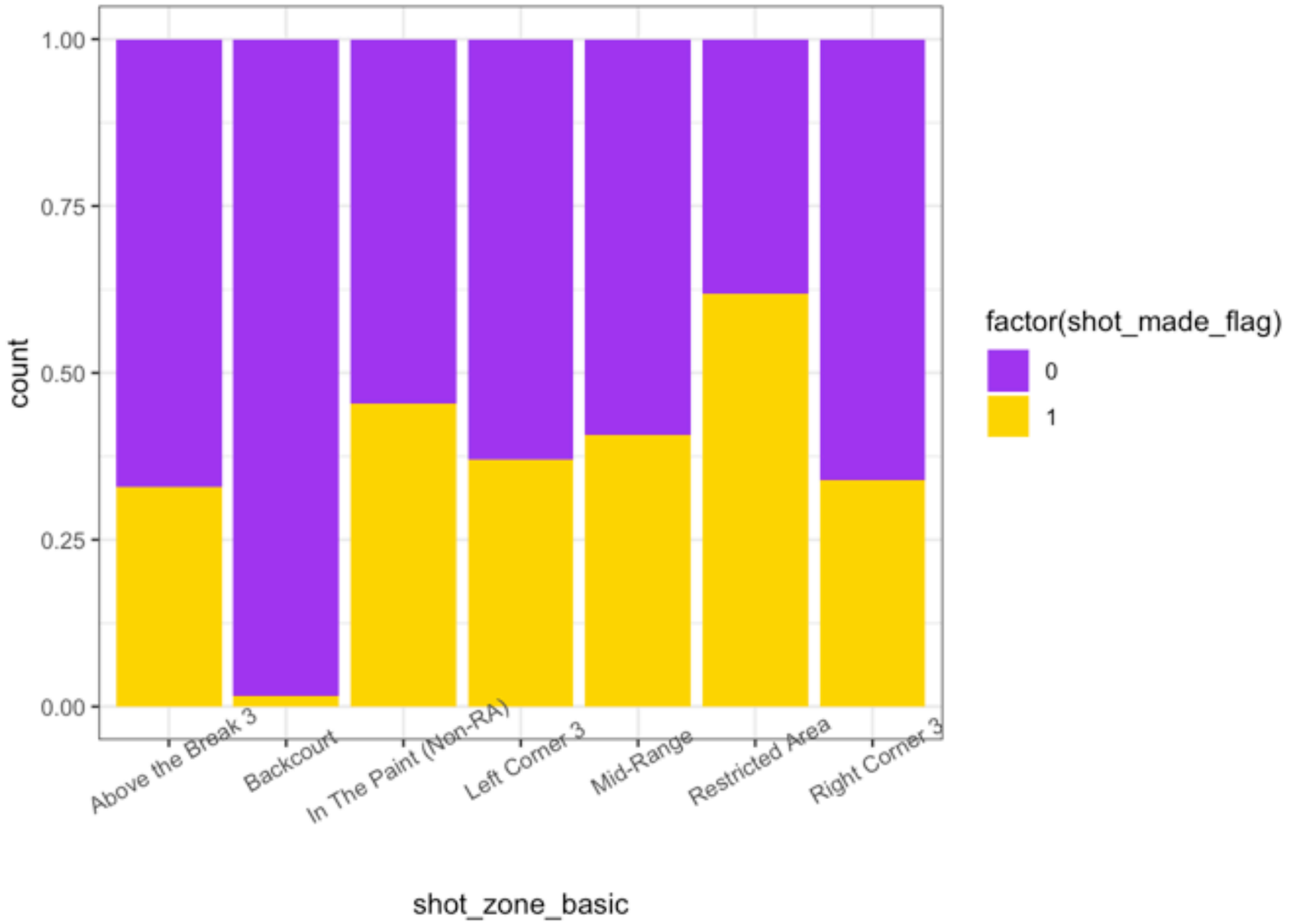
```
Kobe%>%
  group_by(shot_zone_basic)%>%
  summarize(shots=sum(shot_made_flag))%>%
  arrange(desc(shots))
```

```
## # A tibble: 7 x 2
##   shot_zone_basic       shots
##   <chr>                 <dbl>
## 1 Mid-Range              4279
## 2 Restricted Area        3666
## 3 In The Paint (Non-RA)  1763
## 4 Above the Break 3      1554
## 5 Right Corner 3          113
## 6 Left Corner 3            89
## 7 Backcourt                 1
```

```
ggplot(Kobe, aes(loc_x, loc_y, color=shot_zone_basic))+
  geom_point(alpha=0.5)+
  theme(aspect.ratio = 1.9)
```




```
ggplot(Kobe, aes(x=shot_zone_basic, fill=factor(shot_made_flag)))+
  geom_bar(position="fill")+
  scale_fill_manual(values = c("purple", "gold"))+
  theme(axis.text.x=element_text(angle=30))
```

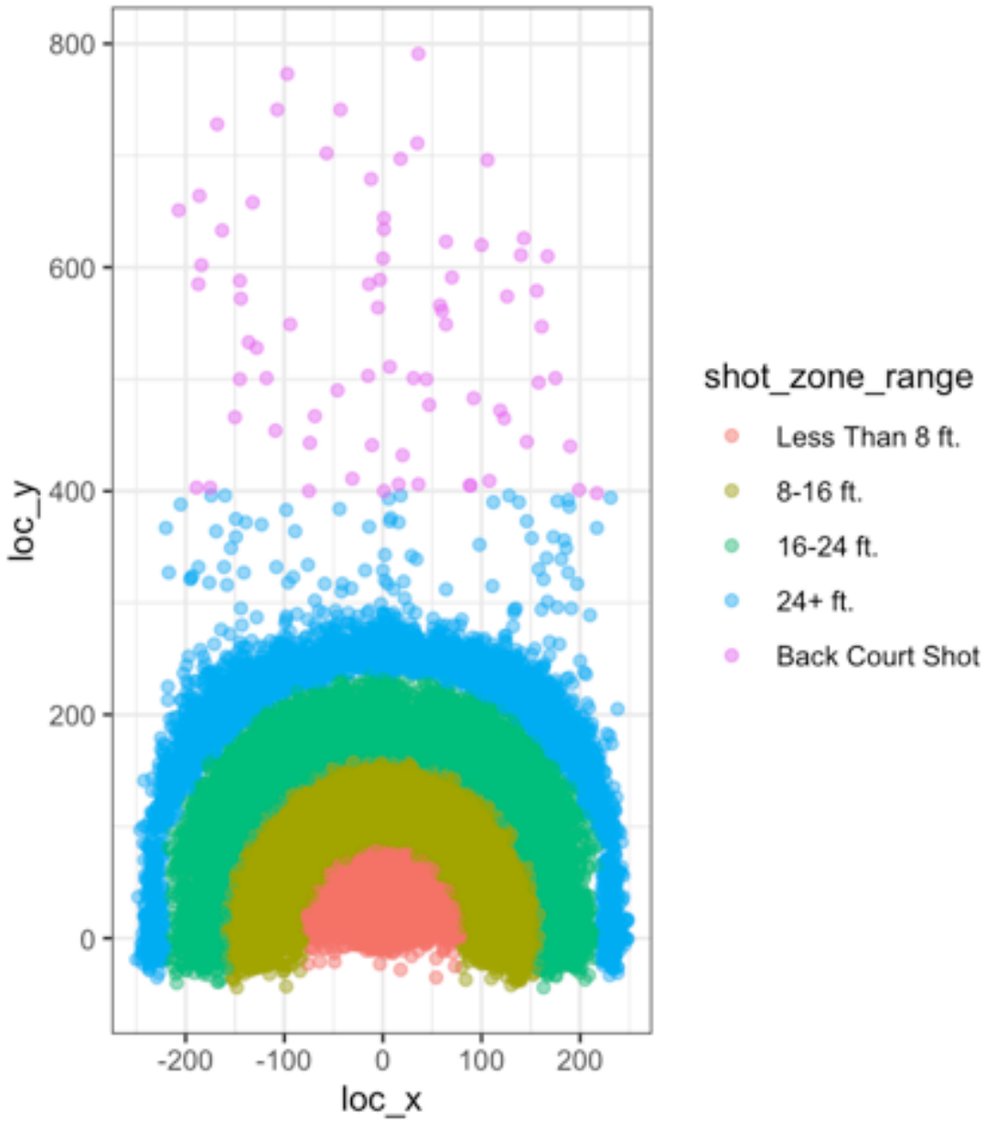


What's shot_zong_range?

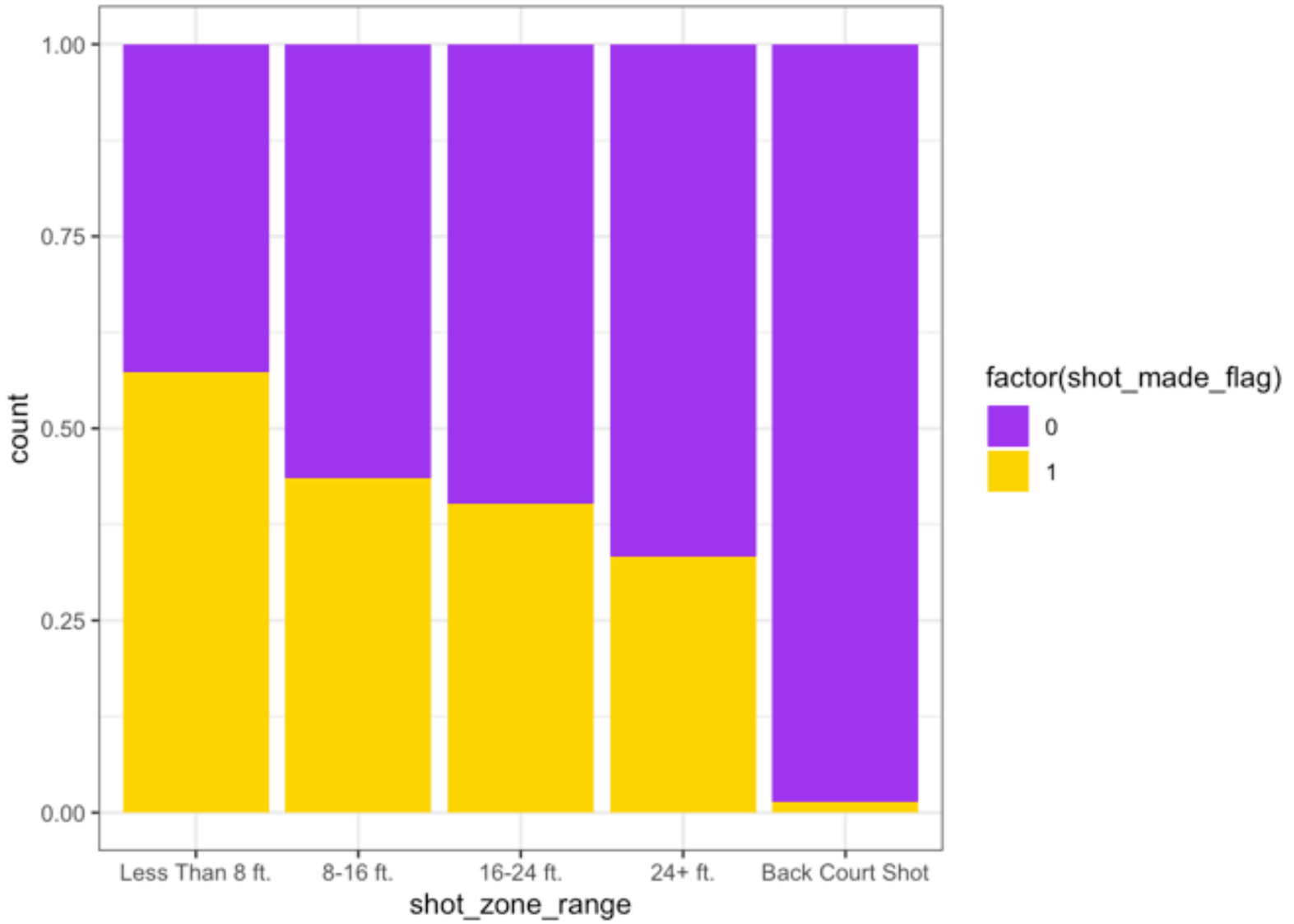
```
Kobe%>%
  group_by(shot_zone_range)%>%
  summarize(shots=sum(shot_made_flag))%>%
  arrange(desc(shots))
```

```
## # A tibble: 5 x 2
##   shot_zone_range shots
##   <fct>           <dbl>
## 1 Less Than 8 ft.  4503
## 2 16-24 ft.        2775
## 3 8-16 ft.         2430
## 4 24+ ft.          1756
## 5 Back Court Shot     1
```

```
ggplot(Kobe, aes(loc_x, loc_y, color=shot_zone_range))+
  geom_point(alpha=0.5)+
  theme(aspect.ratio = 1.9)
```




```
ggplot(Kobe, aes(x=shot_zone_range, fill=factor(shot_made_flag)))+
  geom_bar(position="fill")+
  scale_fill_manual(values = c("purple", "gold"))
```

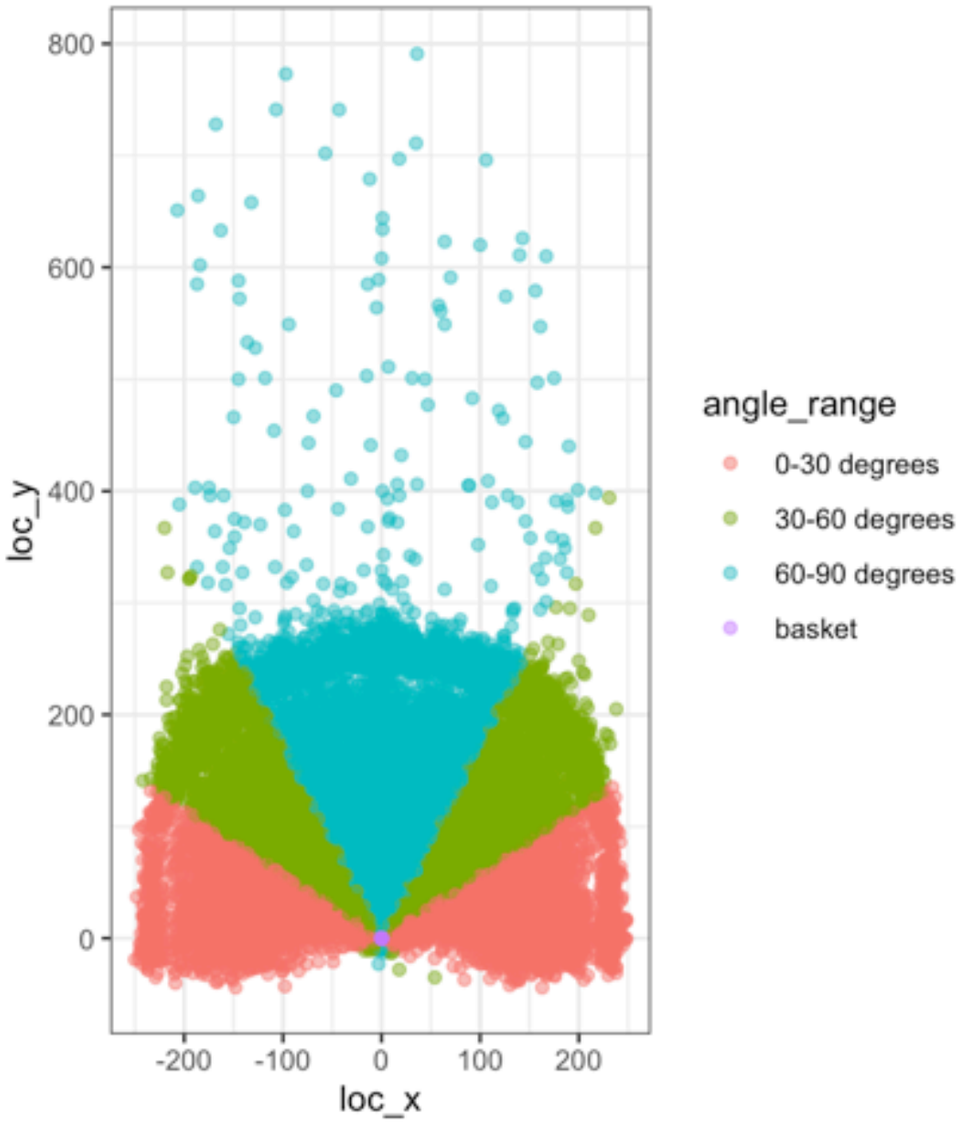


What's angle_range?

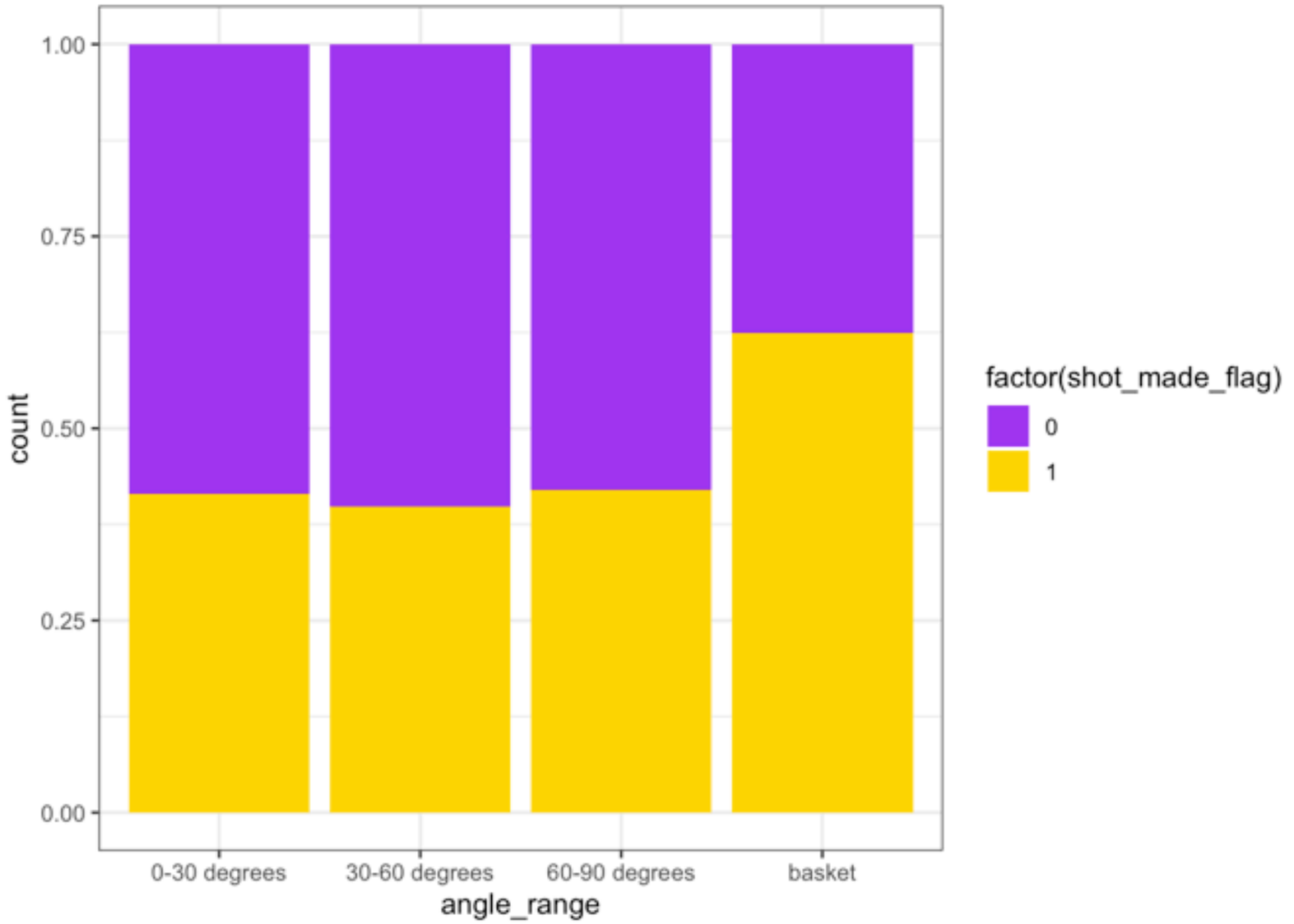
```
Kobe%>%
  group_by(angle_range)%>%
  summarize(shots=sum(shot_made_flag))%>%
  arrange(desc(shots))
```

```
## # A tibble: 4 x 2
##   angle_range   shots
##   <fct>         <dbl>
## 1 30-60 degrees  3283
## 2 60-90 degrees  2870
## 3 basket         2696
## 4 0-30 degrees   2616
```

```
ggplot(Kobe, aes(loc_x, loc_y, color=angle_range))+
  geom_point(alpha=0.5)+
  theme(aspect.ratio = 1.9)
```




```
ggplot(Kobe, aes(x=angle_range, fill=factor(shot_made_flag)))+
  geom_bar(position="fill")+
  scale_fill_manual(values = c("purple", "gold"))
```



Get working dataset **dd**

```
## add dummies by column combined_shot_type
dd<-dummy_cols(Kobe, select_columns = c("action_type", "combined_shot_type","shot
_type","shot_zone_area", "shot_zone_basic", "shot_zone_range", "angle_range"))
dd <- dd[-c(1:5,8,9,12,13,16:25,27:29,31,32)]
```

```
head(dd)
```

```
## # A tibble: 6 x 93
##   loc_x loc_y period playoffs shot_distance shot_made_flag time_remaining
##   <dbl> <dbl>  <dbl>    <dbl>         <dbl>           <dbl>          <dbl>
## 1  -157     0      1        0            15               0            622
## 2  -101   135      1        0            16               1            465
## 3   138   175      1        0            22               0            412
## 4     0     0      2        0             0               1            379
## 5  -145   -11      3        0            14               0            572
## 6     0     0      3        0             0               1            532
## # … with 86 more variables: home <dbl>, `action_type_Jump Shot` <int>,
## #   `action_type_Driving Dunk Shot` <int>, `action_type_Layup Shot` <int>,
## #   `action_type_Running Jump Shot` <int>, `action_type_Reverse Dunk
## #   Shot` <int>, `action_type_Slam Dunk Shot` <int>, `action_type_Driving
```

```
## #   Layup Shot` <int>, `action_type_Turnaround Jump Shot` <int>,
## #   `action_type_Reverse Layup Shot` <int>, `action_type_Tip Shot` <int>,
## #   `action_type_Running Hook Shot` <int>, `action_type_Alley Oop Dunk
## #   Shot` <int>, `action_type_Dunk Shot` <int>, `action_type_Alley Oop
## #   Layup shot` <int>, `action_type_Running Dunk Shot` <int>,
## #   `action_type_Driving Finger Roll Shot` <int>, `action_type_Running
## #   Layup Shot` <int>, `action_type_Finger Roll Shot` <int>,
## #   `action_type_Fadeaway Jump Shot` <int>, `action_type_Follow Up Dunk
## #   Shot` <int>, `action_type_Hook Shot` <int>, `action_type_Turnaround
## #   Hook Shot` <int>, `action_type_Jump Hook Shot` <int>,
## #   `action_type_Running Finger Roll Shot` <int>, `action_type_Jump Bank
## #   Shot` <int>, `action_type_Turnaround Finger Roll Shot` <int>,
## #   `action_type_Hook Bank Shot` <int>, `action_type_Driving Hook
## #   Shot` <int>, `action_type_Running Tip Shot` <int>,
## #   `action_type_Running Reverse Layup Shot` <int>, `action_type_Driving
## #   Finger Roll Layup Shot` <int>, `action_type_Fadeaway Bank shot` <int>,
## #   `action_type_Pullup Jump shot` <int>, `action_type_Finger Roll Layup
## #   Shot` <int>, `action_type_Turnaround Fadeaway shot` <int>,
## #   `action_type_Driving Reverse Layup Shot` <int>, `action_type_Driving
## #   Slam Dunk Shot` <int>, `action_type_Step Back Jump shot` <int>,
## #   `action_type_Turnaround Bank shot` <int>, `action_type_Reverse Slam
## #   Dunk Shot` <int>, `action_type_Floating Jump shot` <int>,
## #   `action_type_Putback Slam Dunk Shot` <int>, `action_type_Running Bank
## #   shot` <int>, `action_type_Driving Bank shot` <int>,
## #   `action_type_Driving Jump shot` <int>, `action_type_Putback Layup
## #   Shot` <int>, `action_type_Putback Dunk Shot` <int>,
## #   `action_type_Running Finger Roll Layup Shot` <int>,
## #   `action_type_Pullup Bank shot` <int>, `action_type_Running Slam Dunk
## #   Shot` <int>, `action_type_Cutting Layup Shot` <int>,
## #   `action_type_Driving Floating Jump Shot` <int>, `action_type_Running
## #   Pull-Up Jump Shot` <int>, `action_type_Tip Layup Shot` <int>,
## #   `action_type_Driving Floating Bank Jump Shot` <int>,
## #   `combined_shot_type_Jump Shot` <int>, combined_shot_type_Dunk <int>,
## #   combined_shot_type_Layup <int>, `combined_shot_type_Tip Shot` <int>,
## #   `combined_shot_type_Hook Shot` <int>, `combined_shot_type_Bank
## #   Shot` <int>, `shot_type_2PT Field Goal` <int>, `shot_type_3PT Field
## #   Goal` <int>, `shot_zone_area_Left Side(L)` <int>, `shot_zone_area_Left
## #   Side Center(LC)` <int>, `shot_zone_area_Right Side Center(RC)` <int>,
## #   `shot_zone_area_Center(C)` <int>, `shot_zone_area_Right
## #   Side(R)` <int>, `shot_zone_area_Back Court(BC)` <int>,
## #   `shot_zone_basic_Mid-Range` <int>, `shot_zone_basic_Restricted
## #   Area` <int>, `shot_zone_basic_In The Paint (Non-RA)` <int>,
## #   `shot_zone_basic_Above the Break 3` <int>, `shot_zone_basic_Right
## #   Corner 3` <int>, shot_zone_basic_Backcourt <int>,
## #   `shot_zone_basic_Left Corner 3` <int>, `shot_zone_range_Less Than 8
## #   ft.` <int>, `shot_zone_range_8-16 ft.` <int>, `shot_zone_range_16-24
## #   ft.` <int>, `shot_zone_range_24+ ft.` <int>, `shot_zone_range_Back
## #   Court Shot` <int>, `angle_range_0-30 degrees` <int>,
## #   `angle_range_30-60 degrees` <int>, `angle_range_60-90 degrees` <int>,
```

```
## #   angle_range_basket <int>
```

```
dim(dd)
```

```
## [1] 25697    93
```

```
colnames(dd)
```

```
##  [1] "loc_x"
##  [2] "loc_y"
##  [3] "period"
##  [4] "playoffs"
##  [5] "shot_distance"
##  [6] "shot_made_flag"
##  [7] "time_remaining"
##  [8] "home"
##  [9] "action_type_Jump Shot"
## [10] "action_type_Driving Dunk Shot"
## [11] "action_type_Layup Shot"
## [12] "action_type_Running Jump Shot"
## [13] "action_type_Reverse Dunk Shot"
## [14] "action_type_Slam Dunk Shot"
## [15] "action_type_Driving Layup Shot"
## [16] "action_type_Turnaround Jump Shot"
## [17] "action_type_Reverse Layup Shot"
## [18] "action_type_Tip Shot"
## [19] "action_type_Running Hook Shot"
## [20] "action_type_Alley Oop Dunk Shot"
## [21] "action_type_Dunk Shot"
## [22] "action_type_Alley Oop Layup shot"
## [23] "action_type_Running Dunk Shot"
## [24] "action_type_Driving Finger Roll Shot"
## [25] "action_type_Running Layup Shot"
## [26] "action_type_Finger Roll Shot"
## [27] "action_type_Fadeaway Jump Shot"
## [28] "action_type_Follow Up Dunk Shot"
## [29] "action_type_Hook Shot"
## [30] "action_type_Turnaround Hook Shot"
## [31] "action_type_Jump Hook Shot"
## [32] "action_type_Running Finger Roll Shot"
## [33] "action_type_Jump Bank Shot"
## [34] "action_type_Turnaround Finger Roll Shot"
## [35] "action_type_Hook Bank Shot"
## [36] "action_type_Driving Hook Shot"
## [37] "action_type_Running Tip Shot"
## [38] "action_type_Running Reverse Layup Shot"
## [39] "action_type_Driving Finger Roll Layup Shot"
```

```
## [40] "action_type_Fadeaway Bank shot"
## [41] "action_type_Pullup Jump shot"
## [42] "action_type_Finger Roll Layup Shot"
## [43] "action_type_Turnaround Fadeaway shot"
## [44] "action_type_Driving Reverse Layup Shot"
## [45] "action_type_Driving Slam Dunk Shot"
## [46] "action_type_Step Back Jump shot"
## [47] "action_type_Turnaround Bank shot"
## [48] "action_type_Reverse Slam Dunk Shot"
## [49] "action_type_Floating Jump shot"
## [50] "action_type_Putback Slam Dunk Shot"
## [51] "action_type_Running Bank shot"
## [52] "action_type_Driving Bank shot"
## [53] "action_type_Driving Jump shot"
## [54] "action_type_Putback Layup Shot"
## [55] "action_type_Putback Dunk Shot"
## [56] "action_type_Running Finger Roll Layup Shot"
## [57] "action_type_Pullup Bank shot"
## [58] "action_type_Running Slam Dunk Shot"
## [59] "action_type_Cutting Layup Shot"
## [60] "action_type_Driving Floating Jump Shot"
## [61] "action_type_Running Pull-Up Jump Shot"
## [62] "action_type_Tip Layup Shot"
## [63] "action_type_Driving Floating Bank Jump Shot"
## [64] "combined_shot_type_Jump Shot"
## [65] "combined_shot_type_Dunk"
## [66] "combined_shot_type_Layup"
## [67] "combined_shot_type_Tip Shot"
## [68] "combined_shot_type_Hook Shot"
## [69] "combined_shot_type_Bank Shot"
## [70] "shot_type_2PT Field Goal"
## [71] "shot_type_3PT Field Goal"
## [72] "shot_zone_area_Left Side(L)"
## [73] "shot_zone_area_Left Side Center(LC)"
## [74] "shot_zone_area_Right Side Center(RC)"
## [75] "shot_zone_area_Center(C)"
## [76] "shot_zone_area_Right Side(R)"
## [77] "shot_zone_area_Back Court(BC)"
## [78] "shot_zone_basic_Mid-Range"
## [79] "shot_zone_basic_Restricted Area"
## [80] "shot_zone_basic_In The Paint (Non-RA)"
## [81] "shot_zone_basic_Above the Break 3"
## [82] "shot_zone_basic_Right Corner 3"
## [83] "shot_zone_basic_Backcourt"
## [84] "shot_zone_basic_Left Corner 3"
## [85] "shot_zone_range_Less Than 8 ft."
## [86] "shot_zone_range_8-16 ft."
## [87] "shot_zone_range_16-24 ft."
## [88] "shot_zone_range_24+ ft."
```

```
## [89] "shot_zone_range_Back Court Shot"
## [90] "angle_range_0-30 degrees"
## [91] "angle_range_30-60 degrees"
## [92] "angle_range_60-90 degrees"
## [93] "angle_range_basket"
```

```
colnames(dd) <- gsub(' ', '_', colnames(dd))
colnames(dd)
```

```
##  [1] "loc_x"
##  [2] "loc_y"
##  [3] "period"
##  [4] "playoffs"
##  [5] "shot_distance"
##  [6] "shot_made_flag"
##  [7] "time_remaining"
##  [8] "home"
##  [9] "action_type_Jump_Shot"
## [10] "action_type_Driving_Dunk_Shot"
## [11] "action_type_Layup_Shot"
## [12] "action_type_Running_Jump_Shot"
## [13] "action_type_Reverse_Dunk_Shot"
## [14] "action_type_Slam_Dunk_Shot"
## [15] "action_type_Driving_Layup_Shot"
## [16] "action_type_Turnaround_Jump_Shot"
## [17] "action_type_Reverse_Layup_Shot"
## [18] "action_type_Tip_Shot"
## [19] "action_type_Running_Hook_Shot"
## [20] "action_type_Alley_Oop_Dunk_Shot"
## [21] "action_type_Dunk_Shot"
## [22] "action_type_Alley_Oop_Layup_shot"
## [23] "action_type_Running_Dunk_Shot"
## [24] "action_type_Driving_Finger_Roll_Shot"
## [25] "action_type_Running_Layup_Shot"
## [26] "action_type_Finger_Roll_Shot"
## [27] "action_type_Fadeaway_Jump_Shot"
## [28] "action_type_Follow_Up_Dunk_Shot"
## [29] "action_type_Hook_Shot"
## [30] "action_type_Turnaround_Hook_Shot"
## [31] "action_type_Jump_Hook_Shot"
## [32] "action_type_Running_Finger_Roll_Shot"
## [33] "action_type_Jump_Bank_Shot"
## [34] "action_type_Turnaround_Finger_Roll_Shot"
## [35] "action_type_Hook_Bank_Shot"
## [36] "action_type_Driving_Hook_Shot"
## [37] "action_type_Running_Tip_Shot"
## [38] "action_type_Running_Reverse_Layup_Shot"
## [39] "action_type_Driving_Finger_Roll_Layup_Shot"
```

```
## [40] "action_type_Fadeaway_Bank_shot"
## [41] "action_type_Pullup_Jump_shot"
## [42] "action_type_Finger_Roll_Layup_Shot"
## [43] "action_type_Turnaround_Fadeaway_shot"
## [44] "action_type_Driving_Reverse_Layup_Shot"
## [45] "action_type_Driving_Slam_Dunk_Shot"
## [46] "action_type_Step_Back_Jump_shot"
## [47] "action_type_Turnaround_Bank_shot"
## [48] "action_type_Reverse_Slam_Dunk_Shot"
## [49] "action_type_Floating_Jump_shot"
## [50] "action_type_Putback_Slam_Dunk_Shot"
## [51] "action_type_Running_Bank_shot"
## [52] "action_type_Driving_Bank_shot"
## [53] "action_type_Driving_Jump_shot"
## [54] "action_type_Putback_Layup_Shot"
## [55] "action_type_Putback_Dunk_Shot"
## [56] "action_type_Running_Finger_Roll_Layup_Shot"
## [57] "action_type_Pullup_Bank_shot"
## [58] "action_type_Running_Slam_Dunk_Shot"
## [59] "action_type_Cutting_Layup_Shot"
## [60] "action_type_Driving_Floating_Jump_Shot"
## [61] "action_type_Running_Pull-Up_Jump_Shot"
## [62] "action_type_Tip_Layup_Shot"
## [63] "action_type_Driving_Floating_Bank_Jump_Shot"
## [64] "combined_shot_type_Jump_Shot"
## [65] "combined_shot_type_Dunk"
## [66] "combined_shot_type_Layup"
## [67] "combined_shot_type_Tip_Shot"
## [68] "combined_shot_type_Hook_Shot"
## [69] "combined_shot_type_Bank_Shot"
## [70] "shot_type_2PT_Field_Goal"
## [71] "shot_type_3PT_Field_Goal"
## [72] "shot_zone_area_Left_Side(L)"
## [73] "shot_zone_area_Left_Side_Center(LC)"
## [74] "shot_zone_area_Right_Side_Center(RC)"
## [75] "shot_zone_area_Center(C)"
## [76] "shot_zone_area_Right_Side(R)"
## [77] "shot_zone_area_Back_Court(BC)"
## [78] "shot_zone_basic_Mid-Range"
## [79] "shot_zone_basic_Restricted_Area"
## [80] "shot_zone_basic_In_The_Paint_(Non-RA)"
## [81] "shot_zone_basic_Above_the_Break_3"
## [82] "shot_zone_basic_Right_Corner_3"
## [83] "shot_zone_basic_Backcourt"
## [84] "shot_zone_basic_Left_Corner_3"
## [85] "shot_zone_range_Less_Than_8_ft."
## [86] "shot_zone_range_8-16_ft."
## [87] "shot_zone_range_16-24_ft."
## [88] "shot_zone_range_24+_ft."
```

```
## [89] "shot_zone_range_Back_Court_Shot"
## [90] "angle_range_0-30_degrees"
## [91] "angle_range_30-60_degrees"
## [92] "angle_range_60-90_degrees"
## [93] "angle_range_basket"
```

```
paste(colnames(dd), collapse = " + ")

colnames(dd)[colnames(dd) == 'shot_zone_basic_Mid-Range'] <- 'shot_zone_basic_Mid
Range'

colnames(dd)[colnames(dd) == 'shot_zone_basic_In_The_Paint_(Non-RA)'] <- 'shot_zo
ne_basic_InThePaint(NonRA)'

colnames(dd)[colnames(dd) == 'shot_zone_range_8-16_ft.'] <- 'shot_zone_range_8to1
6ft'

colnames(dd)[colnames(dd) == 'shot_zone_range_16-24_ft.'] <-
'shot_zone_range_16to24ft'

colnames(dd)[colnames(dd) == 'shot_zone_range_24+_ft.'] <- 'shot_zone_range_24plu
sft'

colnames(dd)[colnames(dd) == 'angle_range_0-30_degrees'] <-
'angle_range_0to30_degrees'

colnames(dd)[colnames(dd) == 'angle_range_30-60_degrees'] <-
'angle_range30to60_degrees'

colnames(dd)[colnames(dd) == 'angle_range_60-90_degrees'] <-
'angle_range_60to90_degrees'

colnames(dd)[colnames(dd) == 'action_type_Running_Pull-Up_Jump_Shot'] <-
'action_type_Running_PullUp_Jump_Shot'

colnames(dd)[colnames(dd) == 'shot_zone_area_Left_Side(L)'] <-
'shot_zone_area_Left_Side'

colnames(dd)[colnames(dd) == 'shot_zone_area_Left_Side_Center(LC)'] <-
'shot_zone_area_Left_Side_Center'

colnames(dd)[colnames(dd) == 'shot_zone_area_Right_Side_Center(RC)'] <-
'shot_zone_area_Right_Side_Center'

colnames(dd)[colnames(dd) == 'shot_zone_area_Center(C)'] <-
'shot_zone_area_Center'

colnames(dd)[colnames(dd) == 'shot_zone_area_Right_Side(R)'] <-
'shot_zone_area_Right_Side'
```

```
colnames(dd)[colnames(dd) == 'shot_zone_area_Back_Court(BC)'] <-
'shot_zone_area_Back_Court'

colnames(dd)[colnames(dd) == 'shot_zone_basic_InThePaint(NonRA)'] <-
'shot_zone_basic_InThePaint'
```

```
paste(colnames(dd), collapse = " + ")
```

```
colnames(dd)
```

```
##  [1] "loc_x"
##  [2] "loc_y"
##  [3] "period"
##  [4] "playoffs"
##  [5] "shot_distance"
##  [6] "shot_made_flag"
##  [7] "time_remaining"
##  [8] "home"
##  [9] "action_type_Jump_Shot"
## [10] "action_type_Driving_Dunk_Shot"
## [11] "action_type_Layup_Shot"
## [12] "action_type_Running_Jump_Shot"
## [13] "action_type_Reverse_Dunk_Shot"
## [14] "action_type_Slam_Dunk_Shot"
## [15] "action_type_Driving_Layup_Shot"
## [16] "action_type_Turnaround_Jump_Shot"
## [17] "action_type_Reverse_Layup_Shot"
## [18] "action_type_Tip_Shot"
## [19] "action_type_Running_Hook_Shot"
## [20] "action_type_Alley_Oop_Dunk_Shot"
## [21] "action_type_Dunk_Shot"
## [22] "action_type_Alley_Oop_Layup_shot"
## [23] "action_type_Running_Dunk_Shot"
## [24] "action_type_Driving_Finger_Roll_Shot"
## [25] "action_type_Running_Layup_Shot"
## [26] "action_type_Finger_Roll_Shot"
## [27] "action_type_Fadeaway_Jump_Shot"
## [28] "action_type_Follow_Up_Dunk_Shot"
## [29] "action_type_Hook_Shot"
## [30] "action_type_Turnaround_Hook_Shot"
## [31] "action_type_Jump_Hook_Shot"
## [32] "action_type_Running_Finger_Roll_Shot"
## [33] "action_type_Jump_Bank_Shot"
## [34] "action_type_Turnaround_Finger_Roll_Shot"
## [35] "action_type_Hook_Bank_Shot"
## [36] "action_type_Driving_Hook_Shot"
```

```
## [37] "action_type_Running_Tip_Shot"
## [38] "action_type_Running_Reverse_Layup_Shot"
## [39] "action_type_Driving_Finger_Roll_Layup_Shot"
## [40] "action_type_Fadeaway_Bank_shot"
## [41] "action_type_Pullup_Jump_shot"
## [42] "action_type_Finger_Roll_Layup_Shot"
## [43] "action_type_Turnaround_Fadeaway_shot"
## [44] "action_type_Driving_Reverse_Layup_Shot"
## [45] "action_type_Driving_Slam_Dunk_Shot"
## [46] "action_type_Step_Back_Jump_shot"
## [47] "action_type_Turnaround_Bank_shot"
## [48] "action_type_Reverse_Slam_Dunk_Shot"
## [49] "action_type_Floating_Jump_shot"
## [50] "action_type_Putback_Slam_Dunk_Shot"
## [51] "action_type_Running_Bank_shot"
## [52] "action_type_Driving_Bank_shot"
## [53] "action_type_Driving_Jump_shot"
## [54] "action_type_Putback_Layup_Shot"
## [55] "action_type_Putback_Dunk_Shot"
## [56] "action_type_Running_Finger_Roll_Layup_Shot"
## [57] "action_type_Pullup_Bank_shot"
## [58] "action_type_Running_Slam_Dunk_Shot"
## [59] "action_type_Cutting_Layup_Shot"
## [60] "action_type_Driving_Floating_Jump_Shot"
## [61] "action_type_Running_PullUp_Jump_Shot"
## [62] "action_type_Tip_Layup_Shot"
## [63] "action_type_Driving_Floating_Bank_Jump_Shot"
## [64] "combined_shot_type_Jump_Shot"
## [65] "combined_shot_type_Dunk"
## [66] "combined_shot_type_Layup"
## [67] "combined_shot_type_Tip_Shot"
## [68] "combined_shot_type_Hook_Shot"
## [69] "combined_shot_type_Bank_Shot"
## [70] "shot_type_2PT_Field_Goal"
## [71] "shot_type_3PT_Field_Goal"
## [72] "shot_zone_area_Left_Side"
## [73] "shot_zone_area_Left_Side_Center"
## [74] "shot_zone_area_Right_Side_Center"
## [75] "shot_zone_area_Center"
## [76] "shot_zone_area_Right_Side"
## [77] "shot_zone_area_Back_Court"
## [78] "shot_zone_basic_MidRange"
## [79] "shot_zone_basic_Restricted_Area"
## [80] "shot_zone_basic_InThePaint"
## [81] "shot_zone_basic_Above_the_Break_3"
## [82] "shot_zone_basic_Right_Corner_3"
## [83] "shot_zone_basic_Backcourt"
## [84] "shot_zone_basic_Left_Corner_3"
## [85] "shot_zone_range_Less_Than_8_ft."
```

```
## [86] "shot_zone_range_8to16ft"
## [87] "shot_zone_range_16to24ft"
## [88] "shot_zone_range_24plusft"
## [89] "shot_zone_range_Back_Court_Shot"
## [90] "angle_range_0to30_degrees"
## [91] "angle_range30to60_degrees"
## [92] "angle_range_60to90_degrees"
## [93] "angle_range_basket"
```

dd2

```
dd2<-Kobe[-c(1,3:5,8,9,12,13,20:25,27:29,31)]
colnames(dd2)
```

```
##  [1] "combined_shot_type" "loc_x"              "loc_y"
##  [4] "period"             "playoffs"           "shot_distance"
##  [7] "shot_made_flag"     "shot_type"          "shot_zone_area"
## [10] "shot_zone_basic"    "shot_zone_range"    "time_remaining"
## [13] "home"               "angle_range"
```

```
dd2$combined_shot_type<-as.factor(dd2$combined_shot_type)
dd2$shot_type<-as.factor(dd2$shot_type)
dd2$shot_zone_area<-as.factor(dd2$shot_zone_area)
dd2$shot_zone_basic<-as.factor(dd2$shot_zone_basic)
```

```
str(dd2)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    25697 obs. of  14 variables:
##  $ combined_shot_type: Factor w/ 6 levels "Bank Shot","Dunk",..: 4 4 4 2 4 5 4
4 4 4 ...
##  $ loc_x             : num  -157 -101 138 0 -145 0 -65 -33 -94 121 ...
##  $ loc_y             : num  0 135 175 0 -11 0 108 125 238 127 ...
##  $ period            : num  1 1 1 2 3 3 3 3 3 1 ...
##  $ playoffs          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ shot_distance     : num  15 16 22 0 14 0 12 12 25 17 ...
##  $ shot_made_flag    : num  0 1 0 1 0 1 1 0 0 1 ...
##  $ shot_type         : Factor w/ 2 levels "2PT Field Goal",..: 1 1 1 1 1 1 1 1
2 1 ...
##  $ shot_zone_area    : Factor w/ 6 levels "Back Court(BC)",..: 4 3 5 2 4 2 4 2
3 5 ...
##  $ shot_zone_basic   : Factor w/ 7 levels "Above the Break 3",..: 5 5 5 6 5 6
3 3 1 5 ...
##  $ shot_zone_range   : Factor w/ 5 levels "Less Than 8 ft.",..: 2 3 3 1 2 1 2
2 4 3 ...
##  $ time_remaining    : num  622 465 412 379 572 532 372 216 116 660 ...
##  $ home              : num  1 1 1 1 1 1 1 1 1 0 ...
##  $ angle_range       : Factor w/ 4 levels "0-30 degrees",..: 1 2 2 4 1 4 2 3 3
2 ...
```

```
dim(dd2)
```

```
## [1] 25697    14
```

Model Choosing

Split Dataset

```
set.seed(666)
test_index <- sample(nrow(dd), 5140) # assign 5140 random rows to the test set(ar
ound 20% of our dataset)
# now split
dd.test <- dd[test_index,]
dd.train <- dd[-test_index,]
```

Linear Regression

Preparation

```
# Intercept
intercept <- lm(shot_made_flag ~ 1, data=dd.train)
summary(intercept)
```

```
##
## Call:
## lm(formula = shot_made_flag ~ 1, data = dd.train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4451 -0.4451 -0.4451  0.5548  0.5548
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.445153   0.003466   128.4   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.497 on 20556 degrees of freedom
```

```
get_mse(intercept, var.estimate = FALSE)
```

```
## [1] 0.2470038
```

```
intercept.test <- lm(shot_made_flag ~1, data = dd.test)
```

```
get_mse(intercept.test, var.estimate = FALSE)
```

```
## [1] 0.2475676
```

```
# Fit the full model
full.model <- lm(shot_made_flag ~., data = dd.train)
summary(full.model)
```

```
get_mse(full.model, var.estimate = FALSE)
```

```
## [1] 0.2102874
```

```
full.model.test <- lm(shot_made_flag ~., data = dd.test)
```

```
get_mse(full.model.test, var.estimate = FALSE)
```

```
## [1] 0.21321
```

Stepwise Regression

Forward Selection

```
set.seed(666)
forward.model <- stepAIC(intercept, direction = "forward", scope=list(upper=full.
model, lower=intercept))
```

```
forward.model$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## shot_made_flag ~ 1
##
## Final Model:
## shot_made_flag ~ action_type_Jump_Shot + action_type_Layup_Shot +
##     combined_shot_type_Dunk + combined_shot_type_Tip_Shot + action_type_Fadeaw
ay_Jump_Shot +
##     action_type_Turnaround_Jump_Shot + action_type_Hook_Shot +
##     shot_zone_range_16to24ft + shot_zone_area_Back_Court + action_type_Turnaro
und_Fadeaway_shot +
##     action_type_Dunk_Shot + period + shot_zone_area_Right_Side_Center +
##     angle_range_60to90_degrees + shot_zone_range_Less_Than_8_ft. +
##     time_remaining + shot_zone_area_Center + action_type_Step_Back_Jump_shot +
##     action_type_Pullup_Bank_shot + action_type_Finger_Roll_Shot +
##     action_type_Reverse_Layup_Shot + action_type_Pullup_Jump_shot +
##     action_type_Driving_Finger_Roll_Shot + action_type_Driving_Finger_Roll_Lay
up_Shot +
##     shot_zone_area_Left_Side + action_type_Running_Finger_Roll_Shot +
##     action_type_Driving_Jump_shot + angle_range_basket + shot_zone_basic_Restr
icted_Area +
##     shot_zone_basic_InThePaint + action_type_Turnaround_Hook_Shot +
##     shot_zone_basic_Right_Corner_3 + shot_zone_basic_Left_Corner_3 +
##     action_type_Tip_Layup_Shot + action_type_Driving_Floating_Jump_Shot +
##     action_type_Fadeaway_Bank_shot + shot_zone_area_Left_Side_Center +
##     loc_y + combined_shot_type_Bank_Shot + action_type_Putback_Slam_Dunk_Shot
##
##
##                                             Step Df    Deviance Resid. Df
## 1                                                                   20556
## 2                        + action_type_Jump_Shot  1 481.7726062     20555
## 3                       + action_type_Layup_Shot  1 138.8754358     20554
## 4                     + combined_shot_type_Dunk  1  43.3015512     20553
## 5                  + combined_shot_type_Tip_Shot  1  13.4299900     20552
## 6               + action_type_Fadeaway_Jump_Shot  1   9.3895729     20551
## 7            + action_type_Turnaround_Jump_Shot  1   9.2043915     20550
```

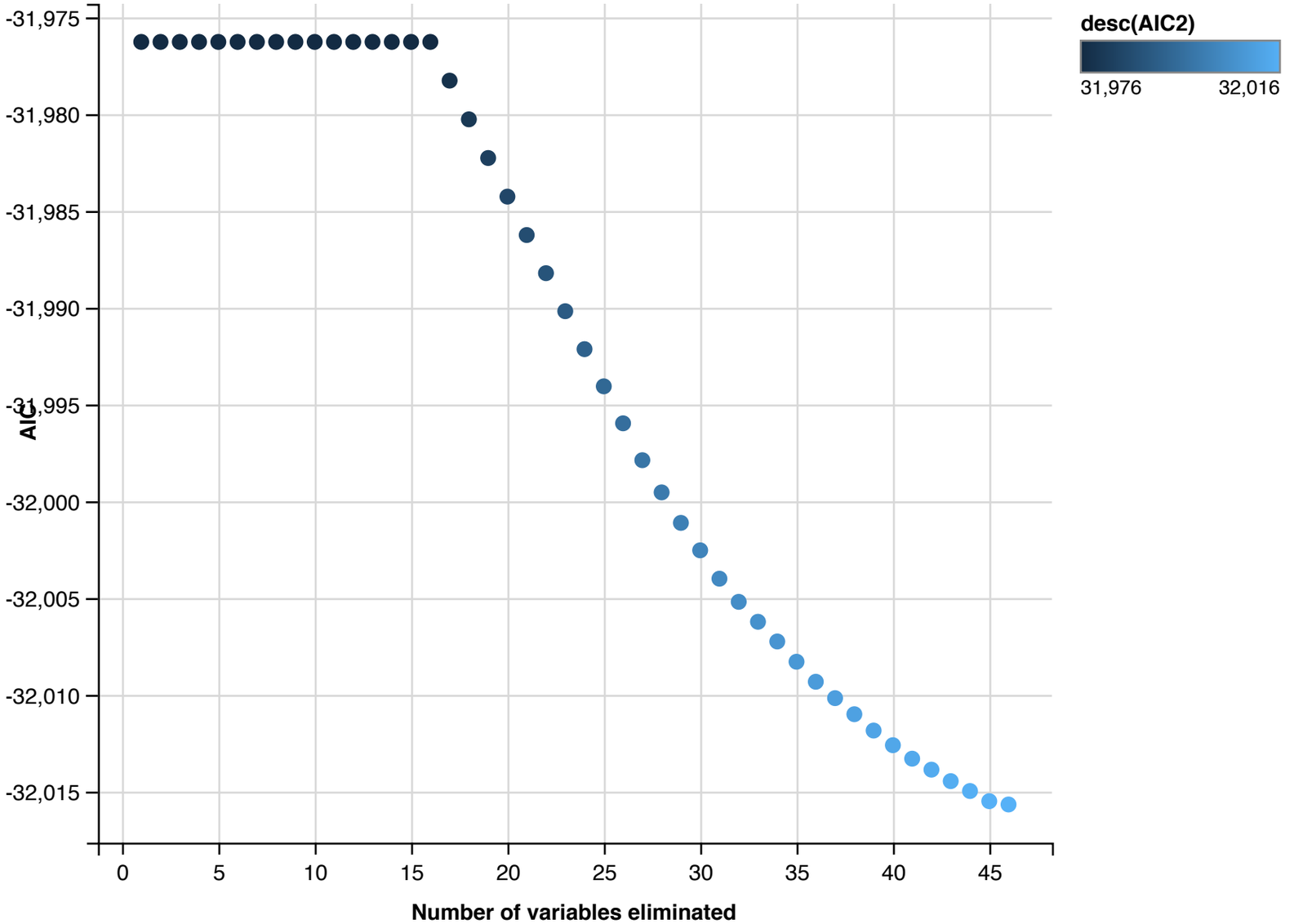
```
## 8                       + action_type_Hook_Shot  1   8.3904168     20549
## 9                     + shot_zone_range_16to24ft  1   6.8884237     20548
## 10                   + shot_zone_area_Back_Court  1   6.1691849     20547
## 11       + action_type_Turnaround_Fadeaway_shot  1   6.1670185     20546
## 12                       + action_type_Dunk_Shot  1   5.3552946     20545
## 13                                     + period  1   3.9050764     20544
## 14            + shot_zone_area_Right_Side_Center  1   3.6341492     20543
## 15                  + angle_range_60to90_degrees  1   2.6518899     20542
## 16             + shot_zone_range_Less_Than_8_ft.  1   2.0171166     20541
## 17                               + time_remaining  1   1.8589092     20540
## 18                       + shot_zone_area_Center  1   1.5405493     20539
## 19             + action_type_Step_Back_Jump_shot  1   1.4803916     20538
## 20                + action_type_Pullup_Bank_shot  1   1.5087321     20537
## 21                + action_type_Finger_Roll_Shot  1   1.4273192     20536
## 22              + action_type_Reverse_Layup_Shot  1   1.4837336     20535
## 23                + action_type_Pullup_Jump_shot  1   1.4204116     20534
## 24       + action_type_Driving_Finger_Roll_Shot  1   1.1151279     20533
## 25 + action_type_Driving_Finger_Roll_Layup_Shot  1   1.1227299     20532
## 26                     + shot_zone_area_Left_Side  1   1.0718028     20531
## 27       + action_type_Running_Finger_Roll_Shot  1   1.0071250     20530
## 28                + action_type_Driving_Jump_shot  1   0.9821474     20529
## 29                             + angle_range_basket  1   1.0323227     20528
## 30             + shot_zone_basic_Restricted_Area  1   1.4851845     20527
## 31                  + shot_zone_basic_InThePaint  1   0.8729998     20526
## 32            + action_type_Turnaround_Hook_Shot  1   0.8320646     20525
## 33              + shot_zone_basic_Right_Corner_3  1   0.7708261     20524
## 34               + shot_zone_basic_Left_Corner_3  1   0.6347565     20523
## 35                  + action_type_Tip_Layup_Shot  1   0.6174544     20522
## 36     + action_type_Driving_Floating_Jump_Shot  1   0.5292472     20521
## 37               + action_type_Fadeaway_Bank_shot  1   0.5022056     20520
## 38             + shot_zone_area_Left_Side_Center  1   0.4464685     20519
## 39                                      + loc_y  1   0.7549780     20518
## 40                + combined_shot_type_Bank_Shot  1   0.4433227     20517
## 41          + action_type_Putback_Slam_Dunk_Shot  1   0.4208339     20516
##    Resid. Dev       AIC
## 1    5077.409 -28744.92
## 2    4595.637 -30792.32
## 3    4456.761 -31421.11
## 4    4413.460 -31619.82
## 5    4400.030 -31680.47
## 6    4390.640 -31722.38
## 7    4381.436 -31763.52
## 8    4373.045 -31800.93
## 9    4366.157 -31831.34
## 10   4359.988 -31858.40
## 11   4353.821 -31885.50
## 12   4348.466 -31908.80
## 13   4344.560 -31925.27
## 14   4340.926 -31940.47
```

```
## 15   4338.274 -31951.04
## 16   4336.257 -31958.60
## 17   4334.398 -31965.41
## 18   4332.858 -31970.72
## 19   4331.377 -31975.74
## 20   4329.869 -31980.90
## 21   4328.441 -31985.68
## 22   4326.958 -31990.73
## 23   4325.537 -31995.48
## 24   4324.422 -31998.78
## 25   4323.299 -32002.12
## 26   4322.228 -32005.21
## 27   4321.220 -32008.01
## 28   4320.238 -32010.68
## 29   4319.206 -32013.59
## 30   4317.721 -32018.66
## 31   4316.848 -32020.82
## 32   4316.016 -32022.78
## 33   4315.245 -32024.45
## 34   4314.610 -32025.48
## 35   4313.993 -32026.42
## 36   4313.463 -32026.94
## 37   4312.961 -32027.33
## 38   4312.515 -32027.46
## 39   4311.760 -32029.06
## 40   4311.317 -32029.17
## 41   4310.896 -32029.18
```

```
get_mse(forward.model)
```

```
## [1] 0.2101236
```

```
AIC <- as.data.frame(forward.model$anova$AIC)
names(AIC) <- "AIC"
AIC %>%
    ggvis(x=~ c(1:41), y=~AIC) %>%
    layer_points(fill = ~ AIC) %>%
    add_axis("y", title = "AIC") %>%
    add_axis("x", title = "Number of variables")
```




```
yhat_test_forward <- predict(forward.model,dd.test)
```

```
mse_test_forward <- mean((dd.test$shot_made_flag-yhat_test_forward)^2)
mse_test_forward
```

```
## [1] 0.2148059
```

Backward Selection

```
set.seed(666)
backward.model <- stepAIC(full.model, direction = "backward")
```

```
backward.model$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## shot_made_flag ~ loc_x + loc_y + period + playoffs + shot_distance +
```

```
##     time_remaining + home + action_type_Jump_Shot + action_type_Driving_Dunk_S
hot +
##     action_type_Layup_Shot + action_type_Running_Jump_Shot +
##     action_type_Reverse_Dunk_Shot + action_type_Slam_Dunk_Shot +
##     action_type_Driving_Layup_Shot + action_type_Turnaround_Jump_Shot +
##     action_type_Reverse_Layup_Shot + action_type_Tip_Shot + action_type_Runnin
g_Hook_Shot +
##     action_type_Alley_Oop_Dunk_Shot + action_type_Dunk_Shot +
##     action_type_Alley_Oop_Layup_shot + action_type_Running_Dunk_Shot +
##     action_type_Driving_Finger_Roll_Shot + action_type_Running_Layup_Shot +
##     action_type_Finger_Roll_Shot + action_type_Fadeaway_Jump_Shot +
##     action_type_Follow_Up_Dunk_Shot + action_type_Hook_Shot +
##     action_type_Turnaround_Hook_Shot + action_type_Jump_Hook_Shot +
##     action_type_Running_Finger_Roll_Shot + action_type_Jump_Bank_Shot +
##     action_type_Turnaround_Finger_Roll_Shot + action_type_Hook_Bank_Shot +
##     action_type_Driving_Hook_Shot + action_type_Running_Tip_Shot +
##     action_type_Running_Reverse_Layup_Shot + action_type_Driving_Finger_Roll_L
ayup_Shot +
##     action_type_Fadeaway_Bank_shot + action_type_Pullup_Jump_shot +
##     action_type_Finger_Roll_Layup_Shot + action_type_Turnaround_Fadeaway_shot
+
##     action_type_Driving_Reverse_Layup_Shot + action_type_Driving_Slam_Dunk_Sho
t +
##     action_type_Step_Back_Jump_shot + action_type_Turnaround_Bank_shot +
##     action_type_Reverse_Slam_Dunk_Shot + action_type_Floating_Jump_shot +
##     action_type_Putback_Slam_Dunk_Shot + action_type_Running_Bank_shot +
##     action_type_Driving_Bank_shot + action_type_Driving_Jump_shot +
##     action_type_Putback_Layup_Shot + action_type_Putback_Dunk_Shot +
##     action_type_Running_Finger_Roll_Layup_Shot + action_type_Pullup_Bank_shot
+
##     action_type_Running_Slam_Dunk_Shot + action_type_Cutting_Layup_Shot +
##     action_type_Driving_Floating_Jump_Shot + action_type_Running_PullUp_Jump_S
hot +
##     action_type_Tip_Layup_Shot + action_type_Driving_Floating_Bank_Jump_Shot +
##     combined_shot_type_Jump_Shot + combined_shot_type_Dunk +
##     combined_shot_type_Layup + combined_shot_type_Tip_Shot +
##     combined_shot_type_Hook_Shot + combined_shot_type_Bank_Shot +
##     shot_type_2PT_Field_Goal + shot_type_3PT_Field_Goal + shot_zone_area_Left_
Side +
##     shot_zone_area_Left_Side_Center + shot_zone_area_Right_Side_Center +
##     shot_zone_area_Center + shot_zone_area_Right_Side + shot_zone_area_Back_Co
urt +
##     shot_zone_basic_MidRange + shot_zone_basic_Restricted_Area +
##     shot_zone_basic_InThePaint + shot_zone_basic_Above_the_Break_3 +
##     shot_zone_basic_Right_Corner_3 + shot_zone_basic_Backcourt +
##     shot_zone_basic_Left_Corner_3 + shot_zone_range_Less_Than_8_ft. +
##     shot_zone_range_8to16ft + shot_zone_range_16to24ft + shot_zone_range_24plu
sft +
##     shot_zone_range_Back_Court_Shot + angle_range_0to30_degrees +
```

```
##     angle_range30to60_degrees + angle_range_60to90_degrees +
##     angle_range_basket
##
## Final Model:
## shot_made_flag ~ loc_y + period + time_remaining + action_type_Jump_Shot +
##     action_type_Layup_Shot + action_type_Running_Jump_Shot +
##     action_type_Driving_Layup_Shot + action_type_Turnaround_Jump_Shot +
##     action_type_Reverse_Layup_Shot + action_type_Tip_Shot + action_type_Dunk_S
hot +
##     action_type_Alley_Oop_Layup_shot + action_type_Running_Layup_Shot +
##     action_type_Finger_Roll_Shot + action_type_Fadeaway_Jump_Shot +
##     action_type_Hook_Shot + action_type_Turnaround_Hook_Shot +
##     action_type_Running_Finger_Roll_Shot + action_type_Jump_Bank_Shot +
##     action_type_Driving_Hook_Shot + action_type_Running_Tip_Shot +
##     action_type_Running_Reverse_Layup_Shot + action_type_Pullup_Jump_shot +
##     action_type_Turnaround_Fadeaway_shot + action_type_Driving_Reverse_Layup_S
hot +
##     action_type_Step_Back_Jump_shot + action_type_Turnaround_Bank_shot +
##     action_type_Floating_Jump_shot + action_type_Running_Bank_shot +
##     action_type_Driving_Jump_shot + action_type_Putback_Layup_Shot +
##     action_type_Pullup_Bank_shot + action_type_Driving_Floating_Jump_Shot +
##     action_type_Running_PullUp_Jump_Shot + action_type_Tip_Layup_Shot +
##     shot_zone_area_Left_Side + shot_zone_area_Left_Side_Center +
##     shot_zone_area_Right_Side_Center + shot_zone_area_Center +
##     shot_zone_area_Right_Side + shot_zone_basic_Restricted_Area +
##     shot_zone_basic_Above_the_Break_3 + shot_zone_range_Less_Than_8_ft. +
##     shot_zone_range_8to16ft + angle_range_0to30_degrees + angle_range30to60_de
grees +
##     angle_range_60to90_degrees
##
##
##                                             Step Df     Deviance Resid. Df
## 1                                                                    20479
## 2                               - angle_range_basket  0 0.000000e+00     20479
## 3                  - shot_zone_range_Back_Court_Shot  0 0.000000e+00     20479
## 4                        - shot_zone_range_24plusft  0 0.000000e+00     20479
## 5                        - shot_zone_range_16to24ft  0 0.000000e+00     20479
## 6                  - shot_zone_basic_Left_Corner_3  0 0.000000e+00     20479
## 7                      - shot_zone_area_Back_Court  0 0.000000e+00     20479
## 8                      - shot_type_3PT_Field_Goal  0 0.000000e+00     20479
## 9                   - combined_shot_type_Bank_Shot  0 0.000000e+00     20479
## 10                  - combined_shot_type_Hook_Shot  0 0.000000e+00     20479
## 11                   - combined_shot_type_Tip_Shot  0 0.000000e+00     20479
## 12                      - combined_shot_type_Layup  0 0.000000e+00     20479
## 13                       - combined_shot_type_Dunk  0 0.000000e+00     20479
## 14                  - combined_shot_type_Jump_Shot  0 0.000000e+00     20479
## 15 - action_type_Driving_Floating_Bank_Jump_Shot  0 0.000000e+00     20479
## 16     - action_type_Turnaround_Finger_Roll_Shot  0 0.000000e+00     20479
## 17          - action_type_Running_Slam_Dunk_Shot  1 6.309959e-05     20480
```

```
## 18                  - action_type_Hook_Bank_Shot  1 5.544149e-05     20481
## 19                  - action_type_Slam_Dunk_Shot  1 3.433617e-04     20482
## 20                   - shot_zone_basic_Backcourt  1 1.053745e-03     20483
## 21              - shot_zone_basic_Right_Corner_3  1 3.241881e-03     20484
## 22                   - shot_zone_basic_MidRange  1 6.286446e-03     20485
## 23                                       - loc_x  1 7.172346e-03     20486
## 24          - action_type_Reverse_Slam_Dunk_Shot  1 8.424946e-03     20487
## 25               - action_type_Driving_Dunk_Shot  1 1.657519e-02     20488
## 26                   - shot_type_2PT_Field_Goal  1 1.877350e-02     20489
## 27          - action_type_Driving_Slam_Dunk_Shot  1 1.971689e-02     20490
## 28             - action_type_Follow_Up_Dunk_Shot  1 7.061869e-02     20491
## 29                                   - playoffs  1 8.887196e-02     20492
## 30              - action_type_Fadeaway_Bank_shot  1 1.219754e-01     20493
## 31             - action_type_Alley_Oop_Dunk_Shot  1 1.120163e-01     20494
## 32                                       - home  1 1.682945e-01     20495
## 33               - action_type_Reverse_Dunk_Shot  1 2.033728e-01     20496
## 34               - action_type_Running_Dunk_Shot  1 2.058090e-01     20497
## 35        - action_type_Driving_Finger_Roll_Shot  1 1.997193e-01     20498
## 36              - action_type_Cutting_Layup_Shot  1 2.022488e-01     20499
## 37               - action_type_Putback_Dunk_Shot  1 2.418387e-01     20500
## 38  - action_type_Driving_Finger_Roll_Layup_Shot  1 2.452080e-01     20501
## 39               - action_type_Driving_Bank_shot  1 2.422158e-01     20502
## 40  - action_type_Running_Finger_Roll_Layup_Shot  1 2.606784e-01     20503
## 41               - action_type_Running_Hook_Shot  1 2.731281e-01     20504
## 42                  - action_type_Jump_Hook_Shot  1 3.004177e-01     20505
## 43          - action_type_Finger_Roll_Layup_Shot  1 2.948547e-01     20506
## 44                               - shot_distance  1 3.123309e-01     20507
## 45                  - shot_zone_basic_InThePaint  1 3.093964e-01     20508
## 46          - action_type_Putback_Slam_Dunk_Shot  1 3.834492e-01     20509
##     Resid. Dev       AIC
## 1     4306.476 -31976.27
## 2     4306.476 -31976.27
## 3     4306.476 -31976.27
## 4     4306.476 -31976.27
## 5     4306.476 -31976.27
## 6     4306.476 -31976.27
## 7     4306.476 -31976.27
## 8     4306.476 -31976.27
## 9     4306.476 -31976.27
## 10    4306.476 -31976.27
## 11    4306.476 -31976.27
## 12    4306.476 -31976.27
## 13    4306.476 -31976.27
## 14    4306.476 -31976.27
## 15    4306.476 -31976.27
## 16    4306.476 -31976.27
## 17    4306.476 -31978.27
## 18    4306.476 -31980.27
## 19    4306.477 -31982.26
```

```
## 20   4306.478 -31984.26
## 21   4306.481 -31986.24
## 22   4306.487 -31988.21
## 23   4306.495 -31990.18
## 24   4306.503 -31992.14
## 25   4306.520 -31994.06
## 26   4306.538 -31995.97
## 27   4306.558 -31997.88
## 28   4306.629 -31999.54
## 29   4306.718 -32001.12
## 30   4306.840 -32002.53
## 31   4306.952 -32004.00
## 32   4307.120 -32005.19
## 33   4307.323 -32006.22
## 34   4307.529 -32007.24
## 35   4307.729 -32008.29
## 36   4307.931 -32009.32
## 37   4308.173 -32010.17
## 38   4308.418 -32011.00
## 39   4308.660 -32011.84
## 40   4308.921 -32012.60
## 41   4309.194 -32013.30
## 42   4309.494 -32013.86
## 43   4309.789 -32014.46
## 44   4310.102 -32014.97
## 45   4310.411 -32015.49
## 46   4310.794 -32015.66
```

```
get_mse(backward.model)
```

```
## [1] 0.2101904
```

```
AIC2 <- as.data.frame(backward.model$anova$AIC)
names(AIC2) <- "AIC2"
AIC2 %>%
    ggvis(x=~ c(1:46), y=~AIC2 ) %>%
    layer_points(fill = ~ desc(AIC2)) %>%
    add_axis("y", title = "AIC") %>%
    add_axis("x", title = "Number of variables eliminated")
```

```
yhat_test_backward <- predict(backward.model,dd.test)
```

```
mse_test_backward <- mean((dd.test$shot_made_flag-yhat_test_backward)^2)
mse_test_backward
```

```
## [1] 0.2155354
```

Penalized regression

```
x_data <- model.matrix( ~ -1 + . -shot_made_flag, dd)
x_train <- x_data[-test_index, ]
y_train <- dd$shot_made_flag[-test_index]
x_test <- x_data[test_index, ]
y_test <- dd$shot_made_flag[test_index]
```

Ridge regression

```
fit_ridge <- cv.glmnet(x_train, y_train, alpha = 0, nfolds = 10)
```

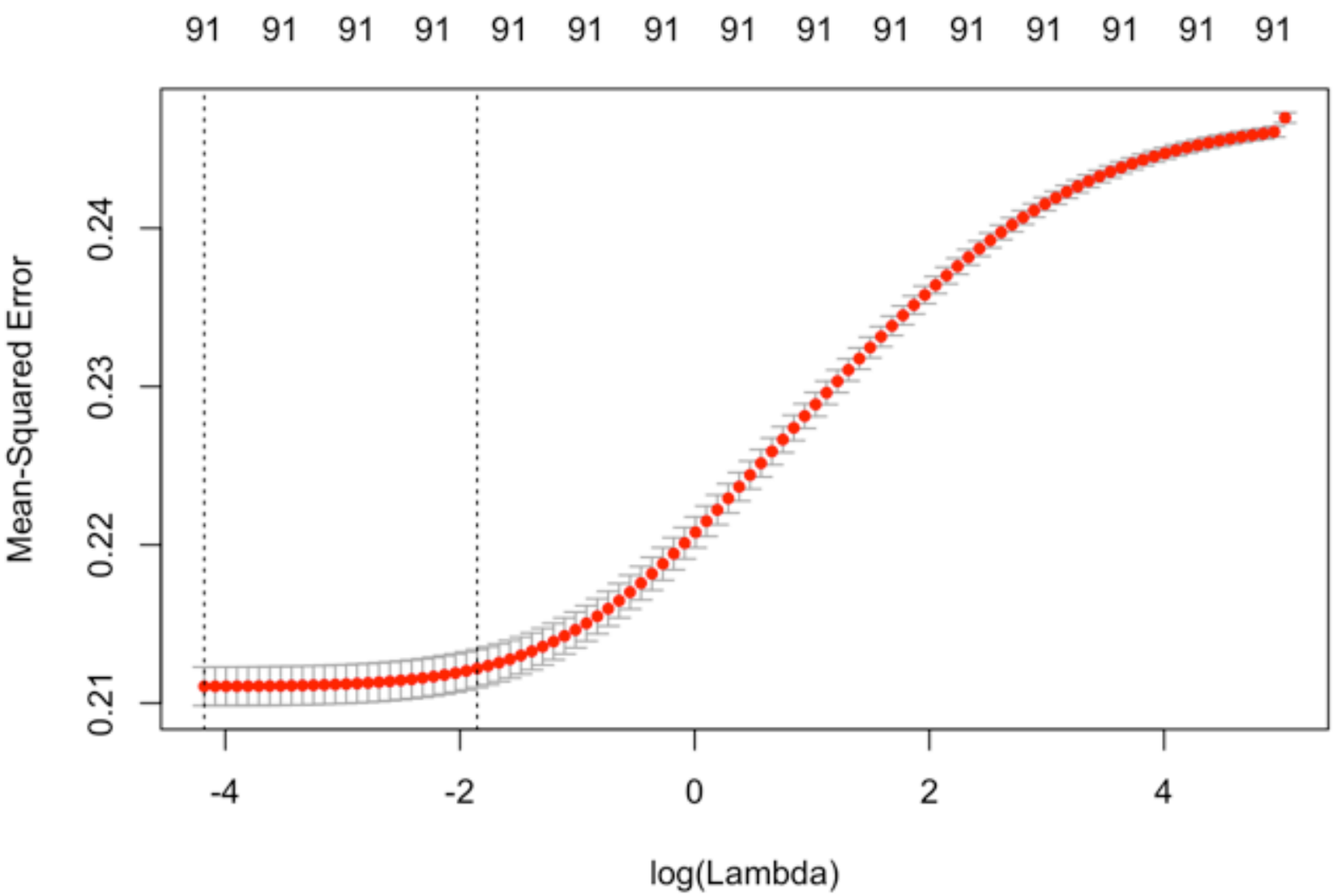
```
yhat_train_ridge <- predict(fit_ridge, x_train, s = fit_ridge$lambda.min)
mse_train_ridge <- mean((y_train - yhat_train_ridge)^2)
mse_train_ridge
```

```
## [1] 0.2095502
```

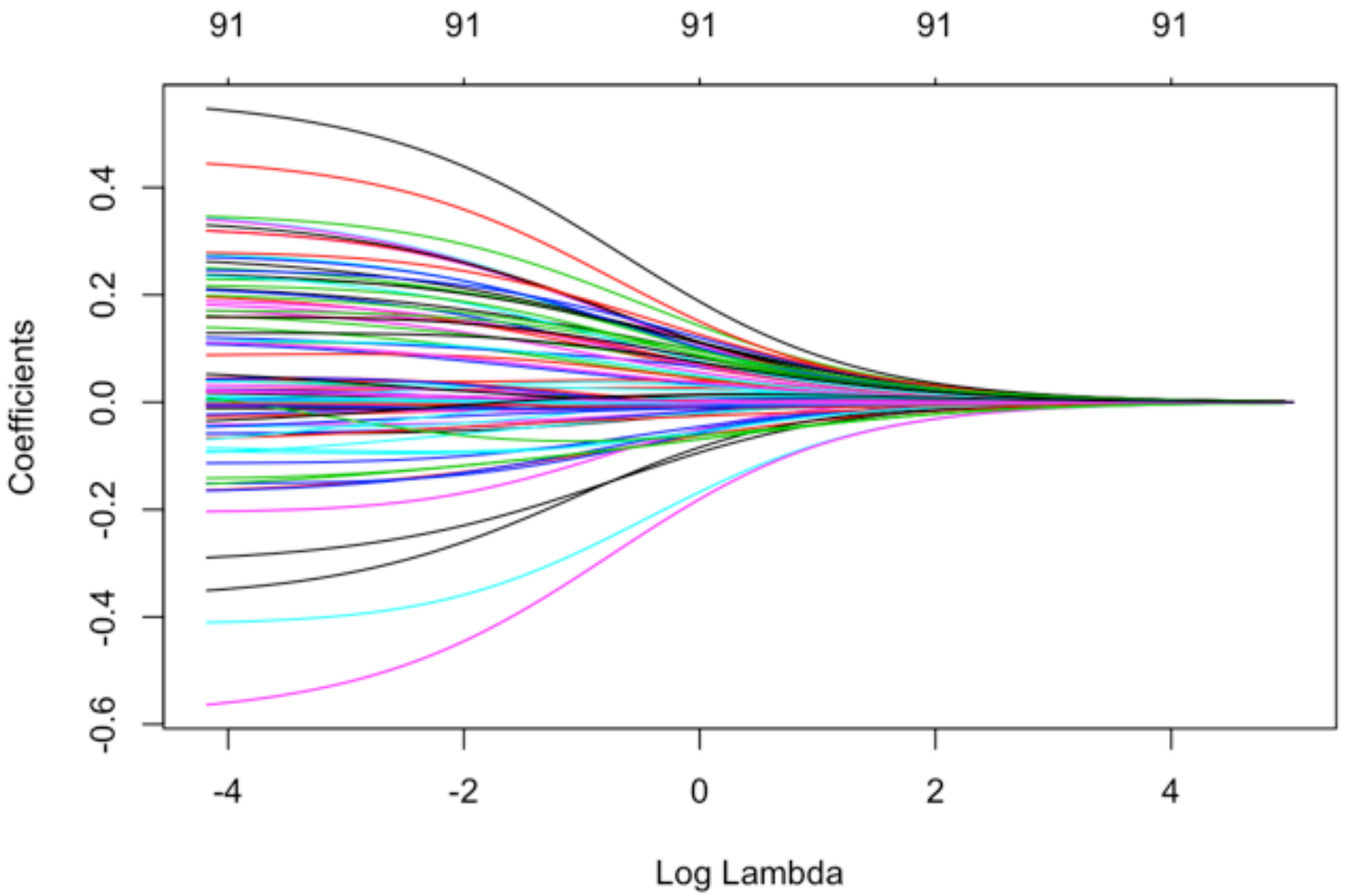
```
yhat_test_ridge <- predict(fit_ridge, x_test, s = fit_ridge$lambda.min)
mse_test_ridge <- mean((y_test - yhat_test_ridge)^2)
mse_test_ridge
```

```
## [1] 0.2148982
```

```
# Plot cross-validation results
plot(fit_ridge)
```




```
fit_ridge2 <- glmnet(x_train, y_train, alpha = 0)
plot(fit_ridge2, xvar = "lambda")
```



Lasso regression

```
fit_lasso <- cv.glmnet(x_train, y_train, alpha = 1, nfolds = 10)
```

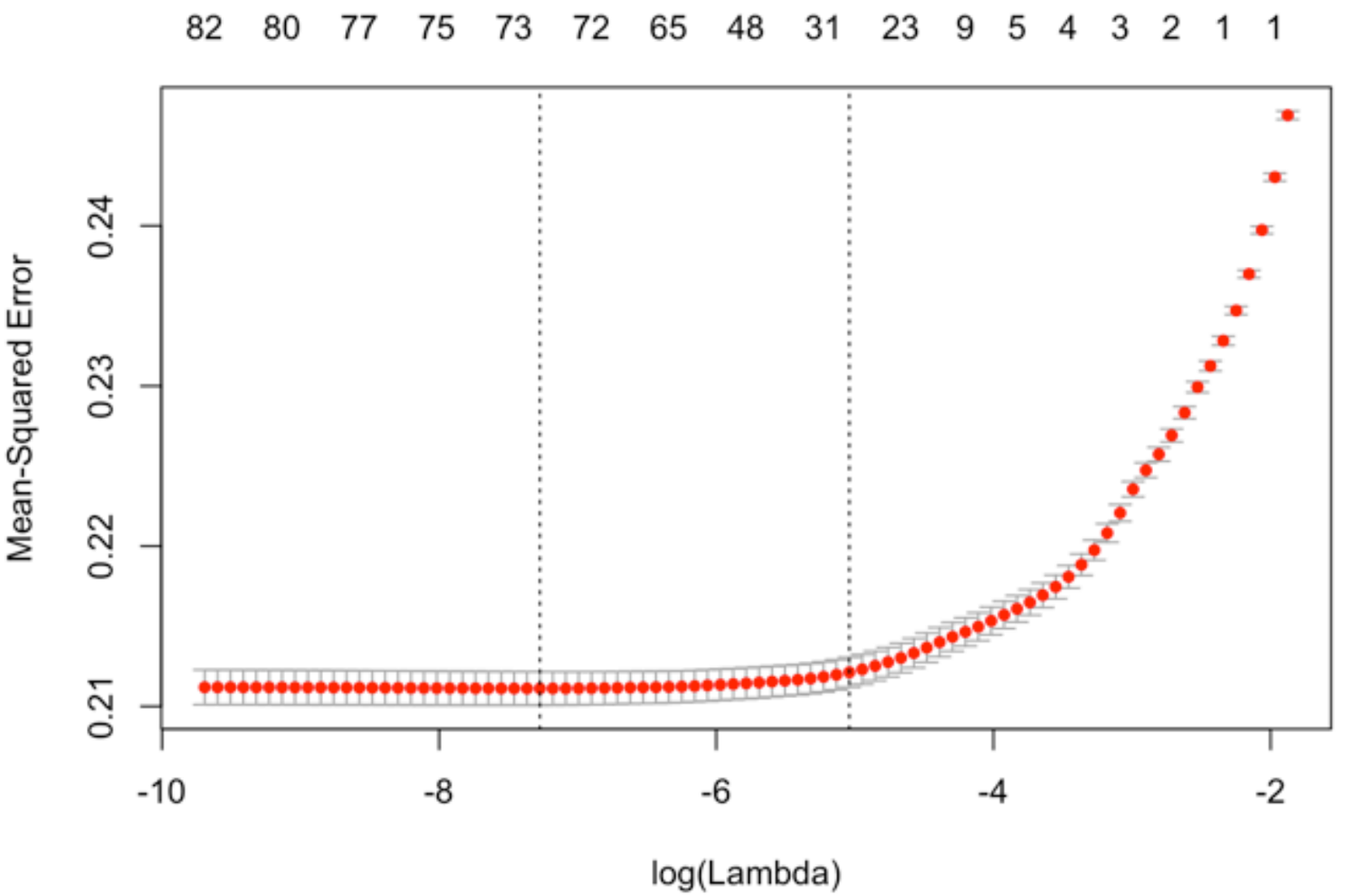
```
yhat_train_lasso <- predict(fit_lasso, x_train, s = fit_lasso$lambda.min)
mse_train_lasso <- mean((y_train - yhat_train_lasso)^2)
mse_train_lasso
```

```
## [1] 0.2095722
```

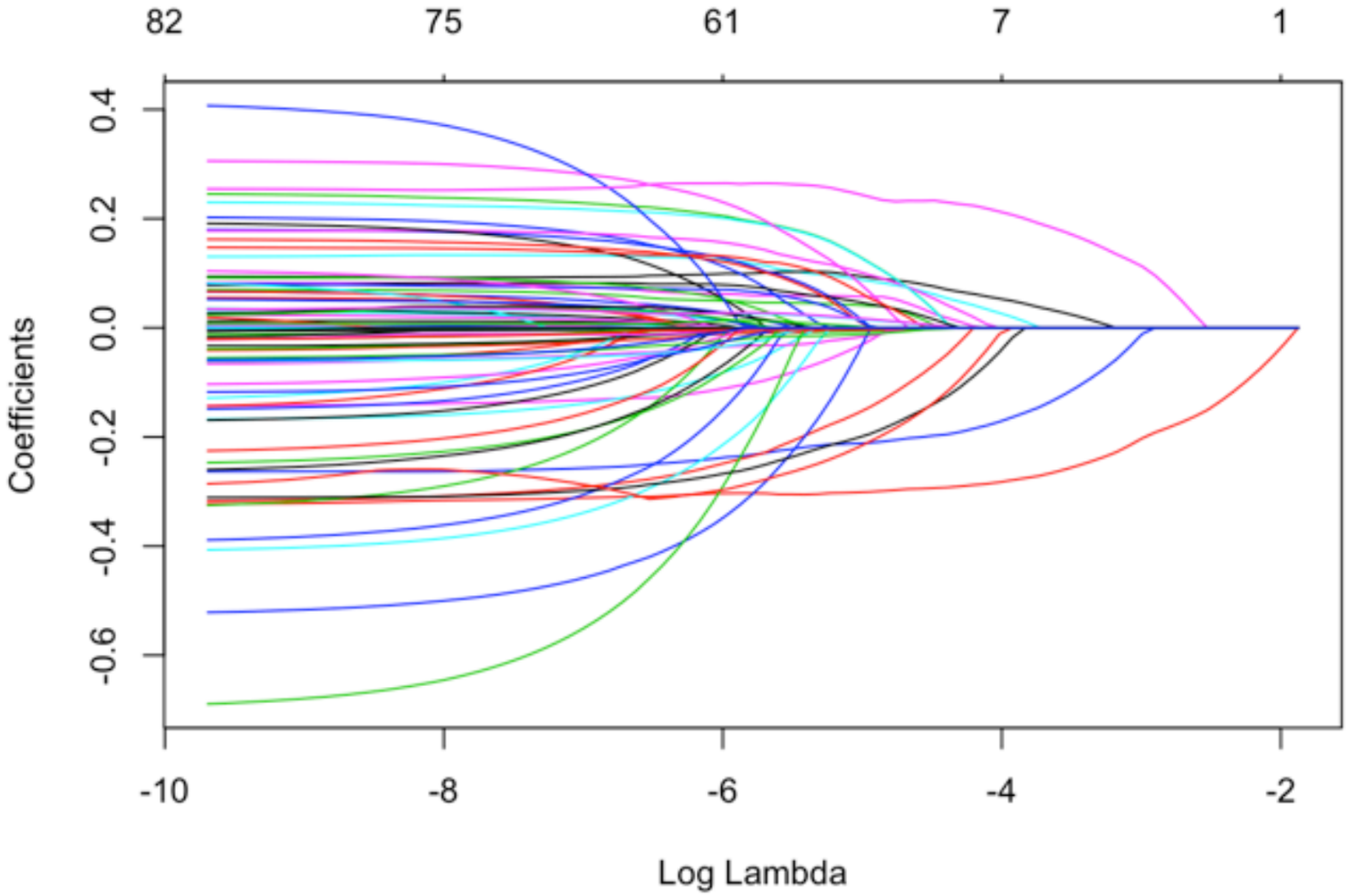
```
yhat_test_lasso <- predict(fit_lasso, x_test, s = fit_lasso$lambda.min)
mse_test_lasso <- mean((y_test - yhat_test_lasso)^2)
mse_test_lasso
```

```
## [1] 0.2147144
```

```
# Plot cross-validation results
plot(fit_lasso)
```




```
fit_lasso2 <- glmnet(x_train, y_train, alpha = 1)
plot(fit_lasso2, xvar = "lambda")
```



Notice: Lasso has a better mse than Ridge in our dataset.

```
##
x_data_1 <- model.matrix( ~ -1 + loc_x + loc_y + playoffs + period + shot_distanc
e + time_remaining + home +  action_type_Jump_Shot + action_type_Layup_Shot + com
bined_shot_type_Dunk, dd)
x_train_1 <- x_data_1[-test_index, ]
y_train_1 <- dd$shot_made_flag[-test_index]
x_test_1 <- x_data_1[test_index, ]
y_test_1 <- dd$shot_made_flag[test_index]
```

```
fit_ridge_1 <- cv.glmnet(x_train_1, y_train_1, alpha = 0, nfolds = 10)
```

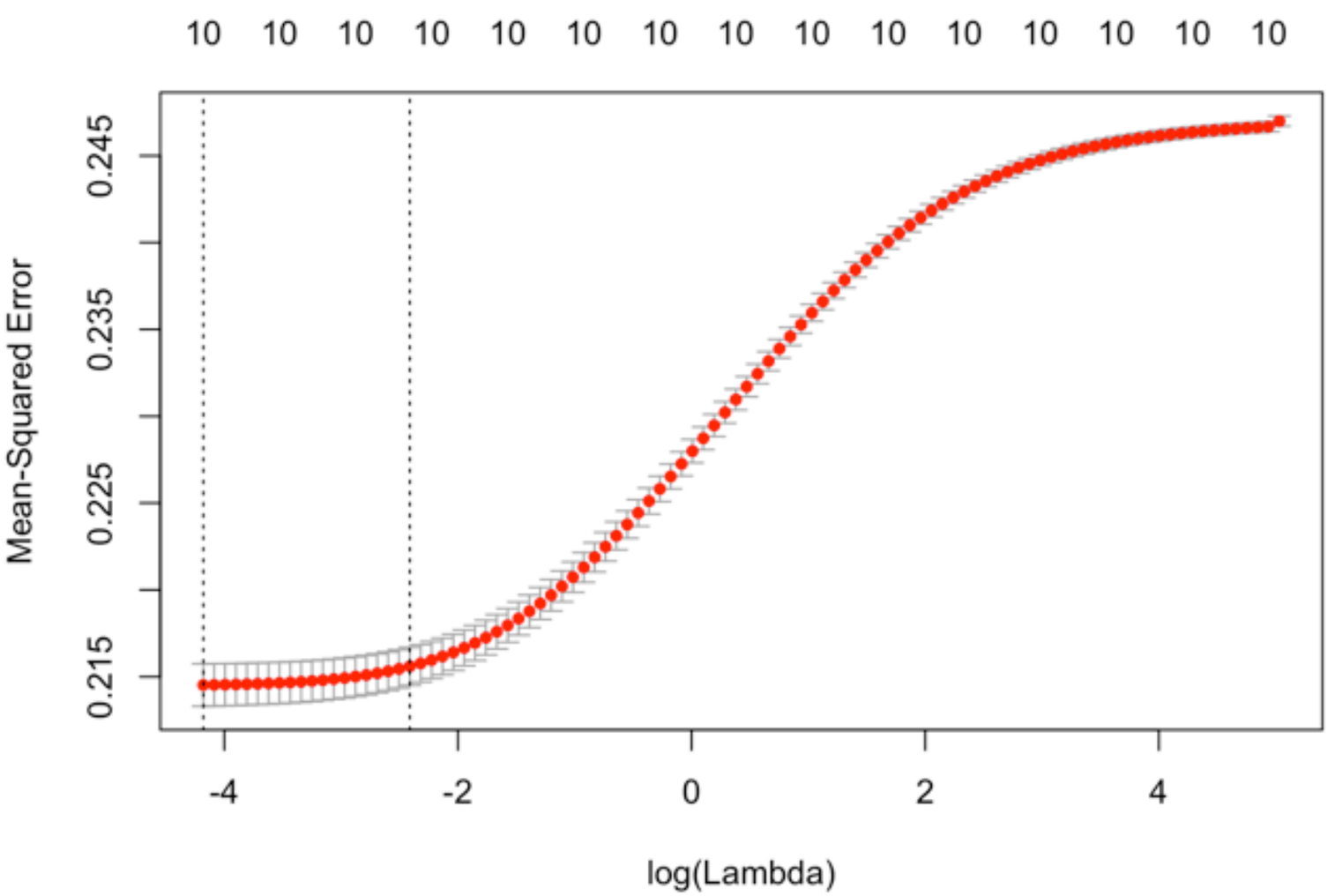
```
yhat_train_ridge_1 <- predict(fit_ridge_1, x_train_1, s = fit_ridge_1$lambda.min)
mse_train_ridge_1 <- mean((y_train_1 - yhat_train_ridge_1)^2)
mse_train_ridge_1
```

```
## [1] 0.214269
```

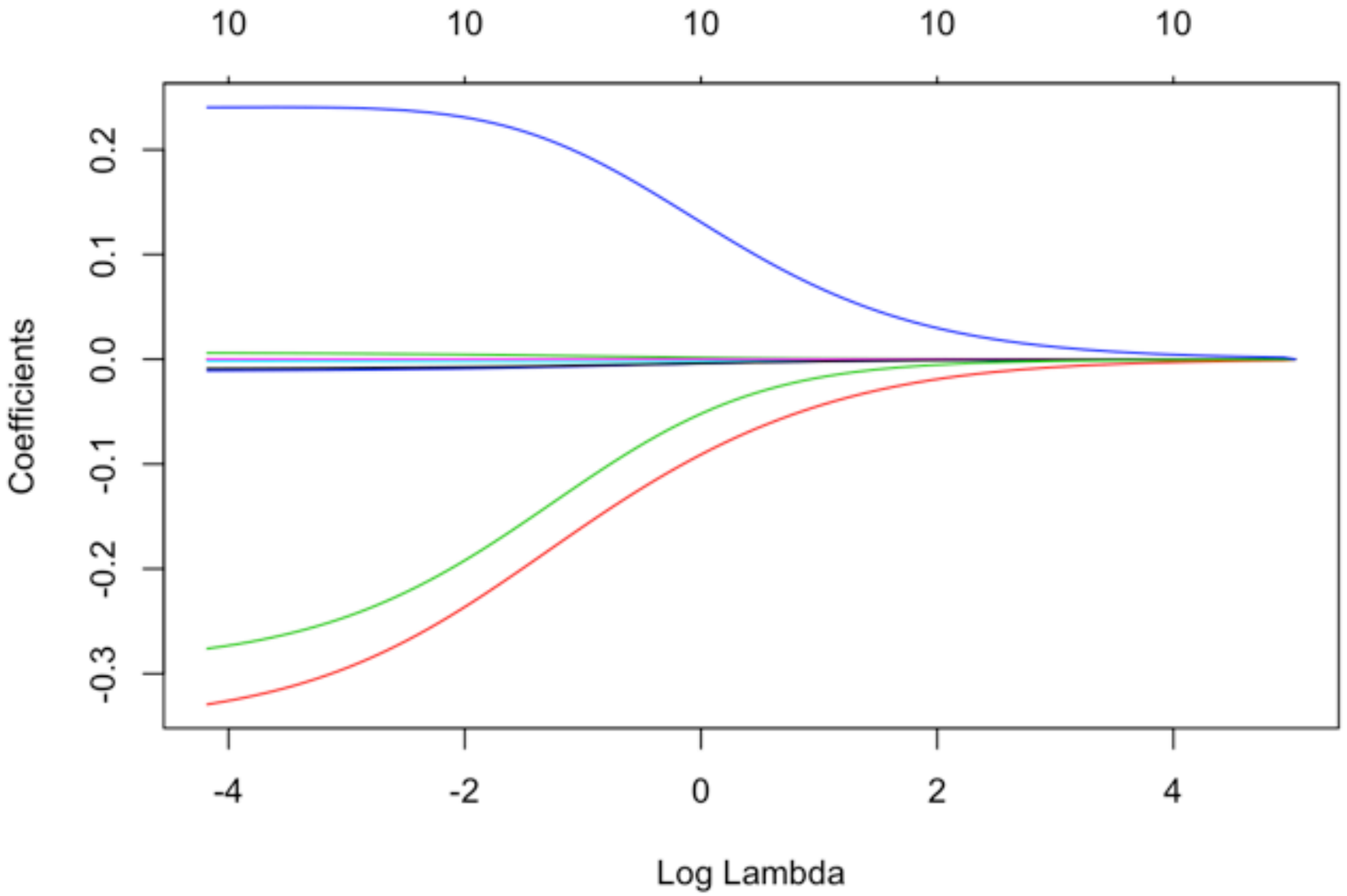
```
yhat_test_ridge_1 <- predict(fit_ridge_1, x_test_1, s = fit_ridge_1$lambda.min)
mse_test_ridge_1 <- mean((y_test_1 - yhat_test_ridge_1)^2)
mse_test_ridge_1
```

```
## [1] 0.2164744
```

```
plot(fit_ridge_1)
```




```
fit_ridge3 <- glmnet(x_train_1, y_train_1, alpha = 0)
plot(fit_ridge3, xvar = "lambda")
```




```
fit_lasso_1 <- cv.glmnet(x_train_1, y_train_1, alpha = 1, nfolds = 10)
```

```
yhat_train_lasso_1 <- predict(fit_lasso_1, x_train_1, s = fit_lasso_1$lambda.min)
mse_train_lasso_1 <- mean((y_train_1 - yhat_train_lasso_1)^2)
mse_train_lasso_1
```

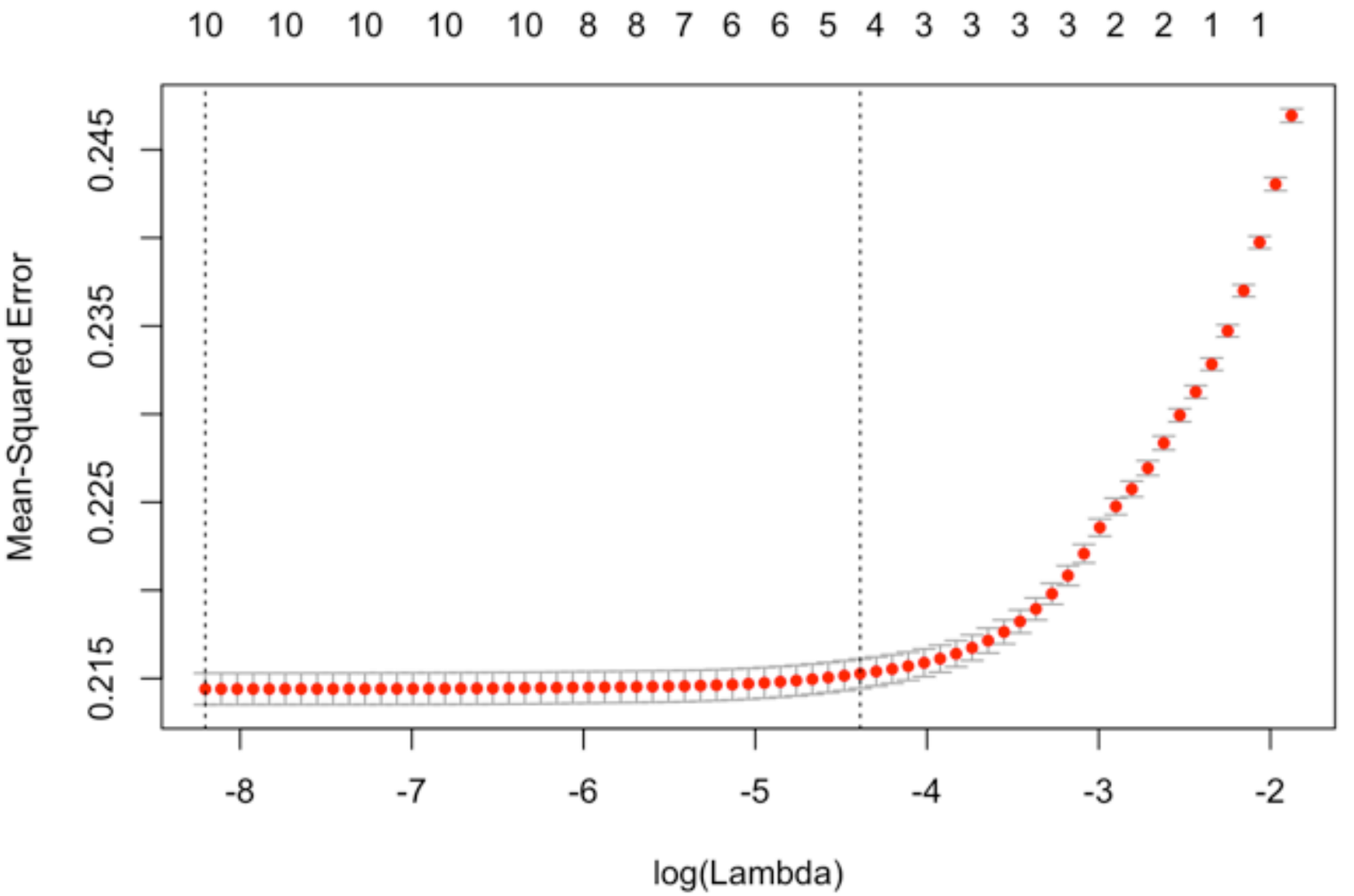
```
## [1] 0.2142151
```

Notice: Lasso has a better mse than Ridge in our dataset.

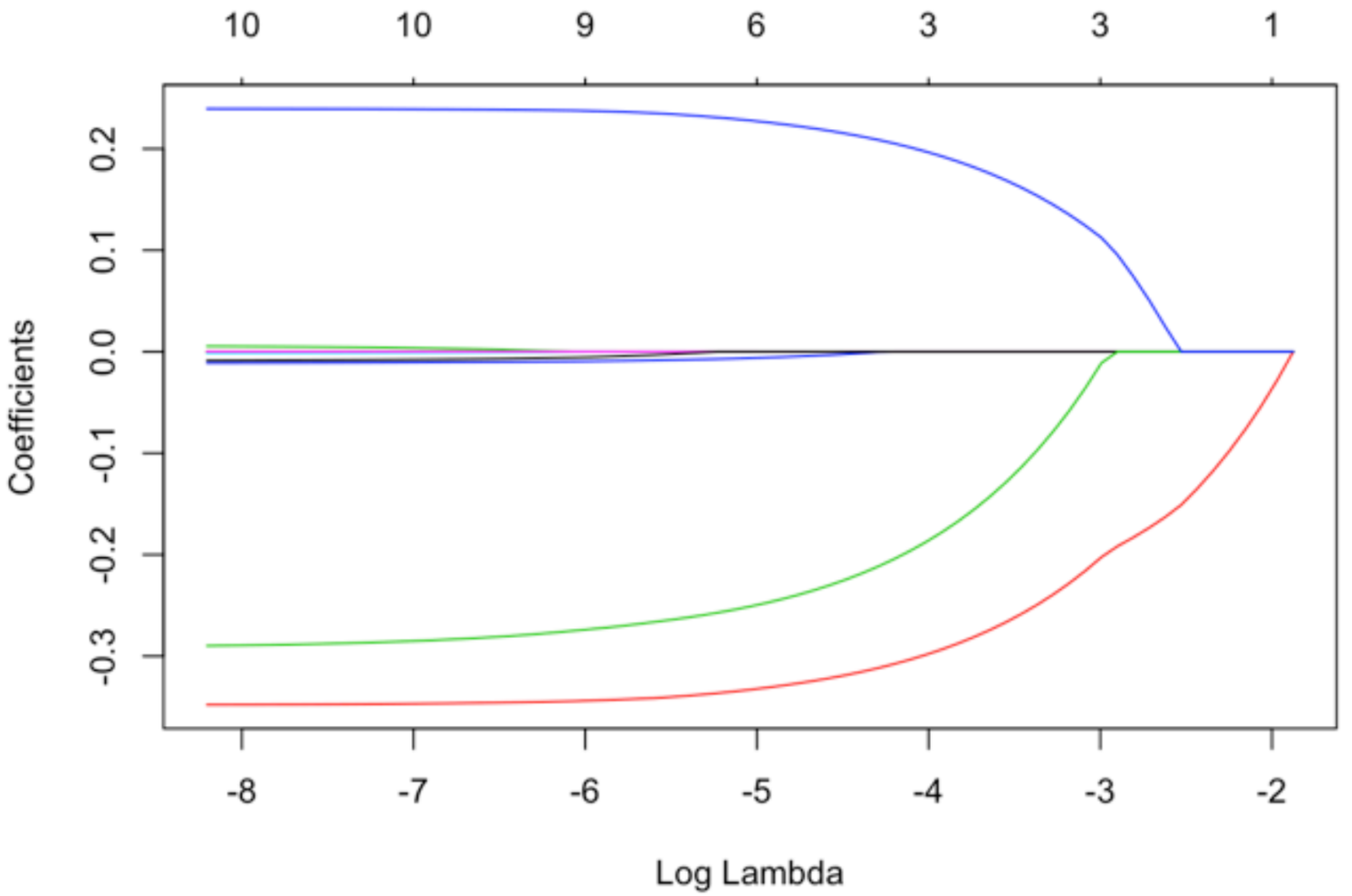
```
yhat_test_lasso_1 <- predict(fit_lasso_1, x_test_1, s = fit_lasso_1$lambda.min)
mse_test_lasso_1 <- mean((y_test_1 - yhat_test_lasso_1)^2)
mse_test_lasso_1
```

```
## [1] 0.2165744
```

```
plot(fit_lasso_1)
```




```
fit_lasso3 <- glmnet(x_train_1, y_train_1, alpha = 1)
plot(fit_lasso3, xvar = "lambda")
```



Trees

```
set.seed(666)
test_index2 <- sample(nrow(dd2), 5140) # assign 5140 random rows to the test set( around 20% of our dataset)
# now split
dd.test2 <- dd2[test_index2,]
dd.train2 <- dd2[-test_index2,]
```

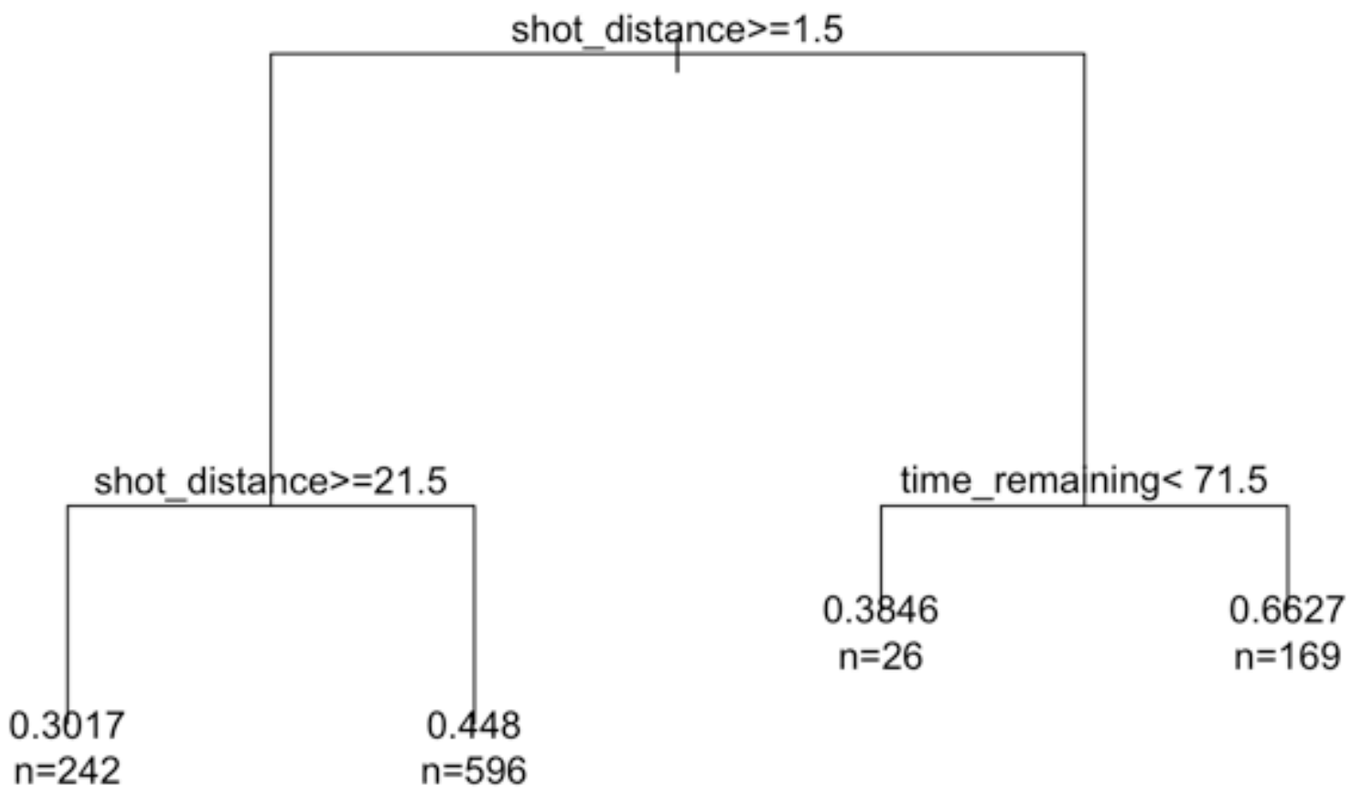
```
dd_test <- dd.test2 %>% filter(sample(c(0,1),nrow(dd.test2),replace=TRUE,prob=c(0.95,0.05))==1)
dd_train <- dd.train2 %>% filter(sample(c(0,1),nrow(dd.train2),replace=TRUE,prob=c(0.95,0.05))==1)
```

```
y_train2 <- dd2$shot_made_flag[-test_index2]
y_test2 <- dd2$shot_made_flag[test_index2]
```

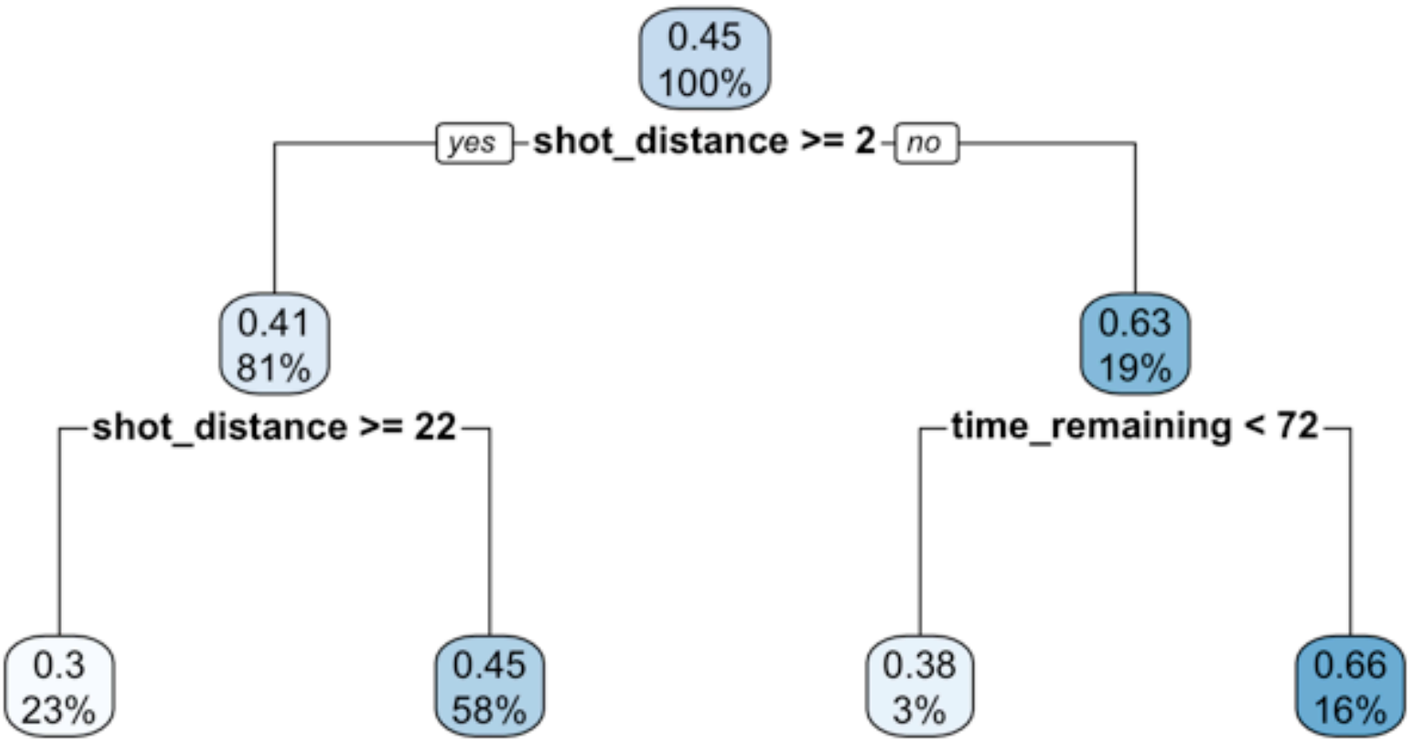
```
f1 <- as.formula(shot_made_flag ~ shot_distance+time_remaining+period+shot_type)
```

```
fit.tree <- rpart(f1, dd_train,
control = rpart.control(cp = 0.006))
```

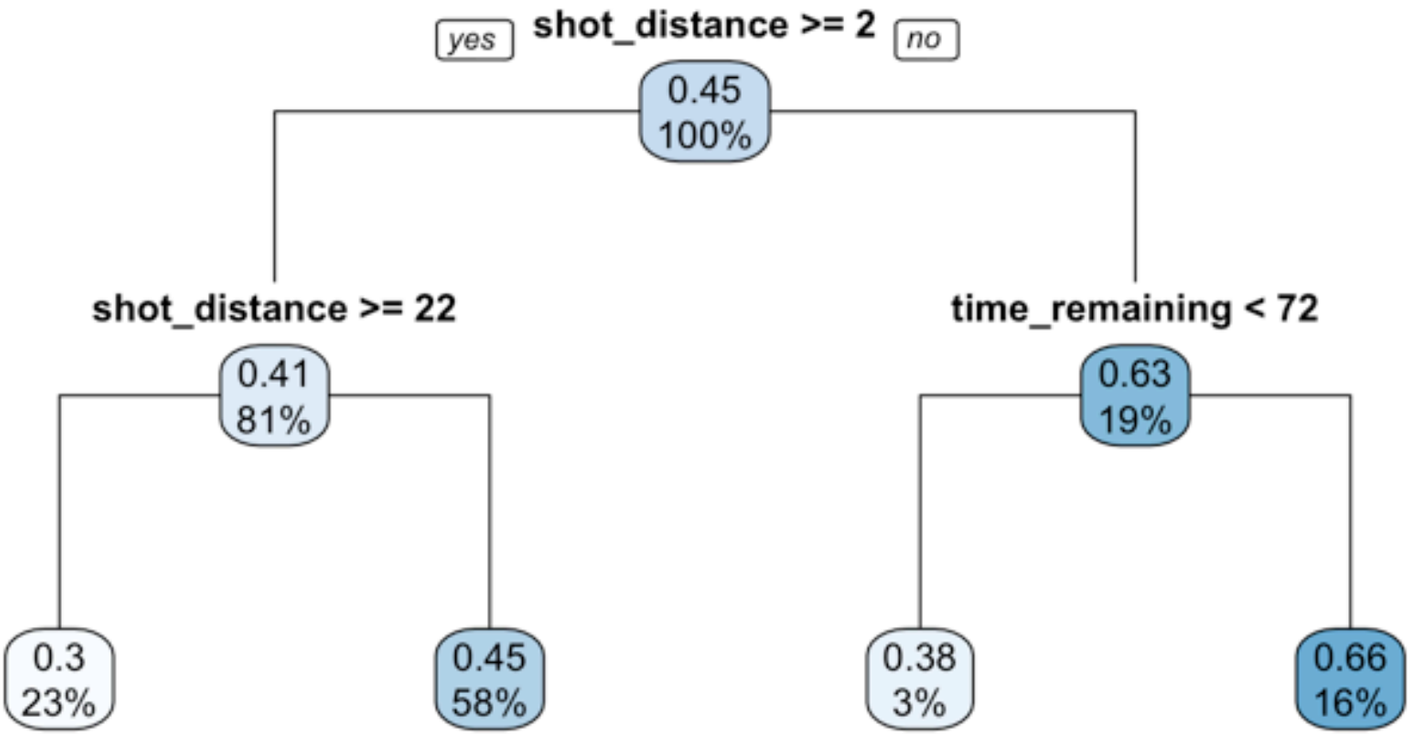
```
par(xpd = TRUE)
plot(fit.tree, compress=TRUE)
text(fit.tree, use.n=TRUE)
```




```
rpart.plot(fit.tree)
```




```
rpart.plot(fit.tree, type = 1)
```




```
yhat.train.tree <- predict(fit.tree, dd_train)
mse.train.tree <- mean((dd_train$shot_made_flag - yhat.train.tree)^2)
mse.train.tree
```

```
## [1] 0.2345556
```

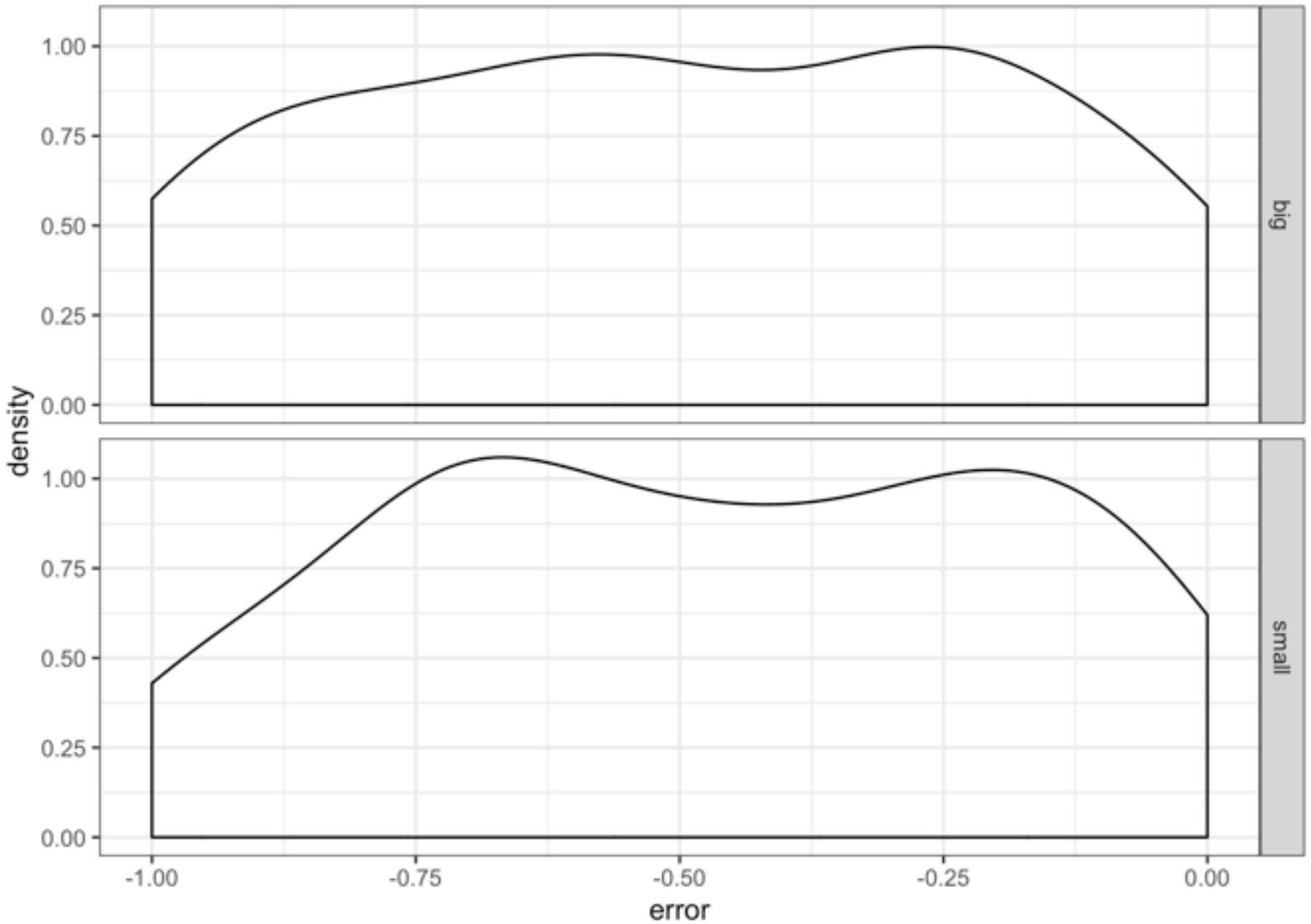
```
yhat.test.tree <- predict(fit.tree, dd_test)
mse.test.tree <- mean((dd_test$shot_made_flag - yhat.test.tree)^2)
mse.test.tree
```

```
## [1] 0.2372982
```

```
##bias-variance tradeoff using trees
f2 <- as.formula(shot_made_flag ~ .)
```

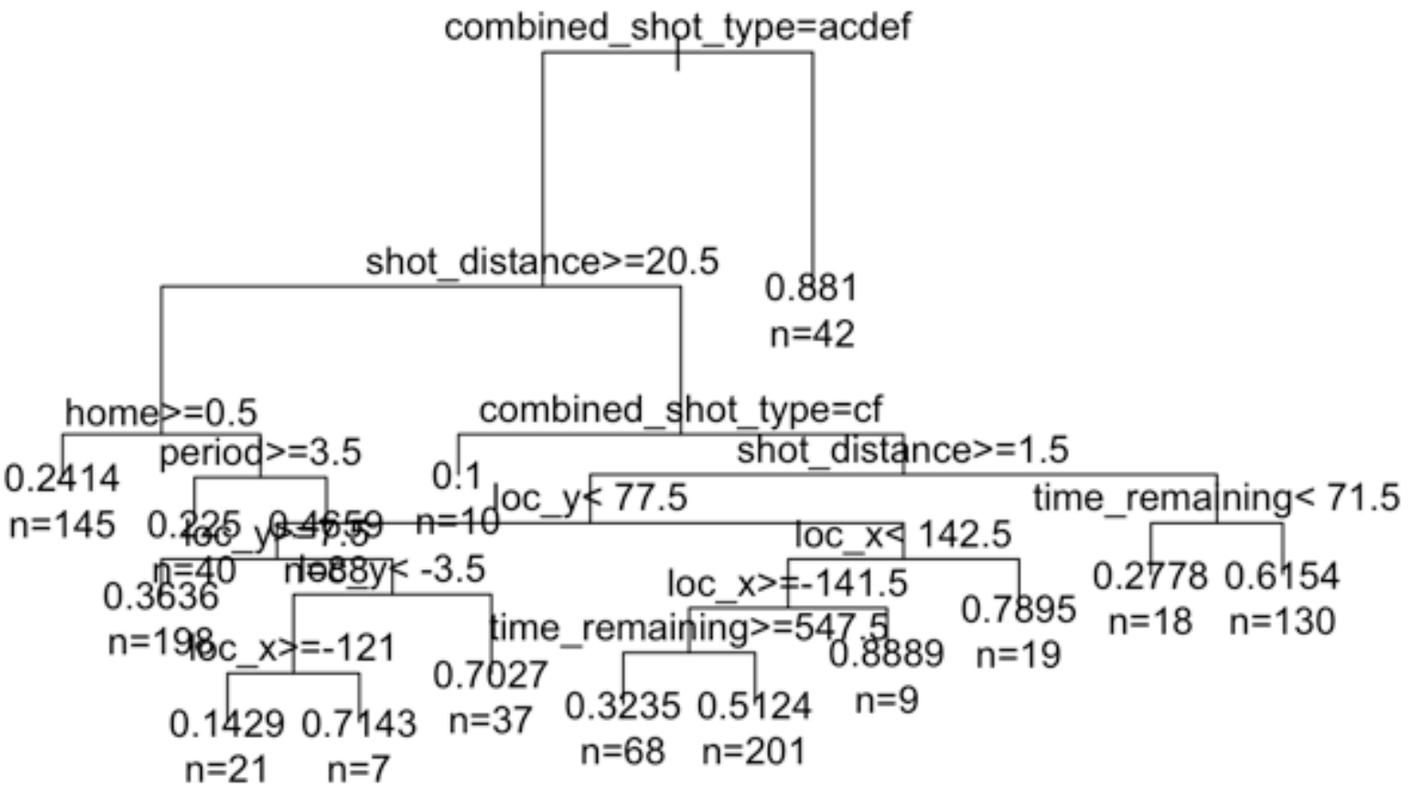
```
set.seed(666)
x0 <- dd_train[1,]
dd_train_1 <- dd_train[-1,]
yhat_small_tree <- c()
for (i in seq(100)) {
  fit_tree <- rpart(f2,
    dd_train_1 %>% sample_frac(size = .1),
    control = rpart.control(cp = 0.001))
  yhat <- predict(fit_tree, x0)
  yhat_small_tree <- c(yhat_small_tree, yhat)
}
# this is vector of predictions for the big trees
yhat_big_tree <- c()
#fit small trees with cp = 0.0001
for (i in seq(100)) {
  fit_tree <- rpart(f2,
    dd_train_1 %>% sample_frac(size = .1),
    control = rpart.control(cp = 0.0001))
  yhat <- predict(fit_tree, x0)
  yhat_big_tree <- c(yhat_big_tree, yhat)
}
# make a data frame with the errors from our two trials
errors <- data.frame(
  "error"= (x0$shot_made_flag - c(yhat_small_tree, yhat_big_tree)),
  "flexibility"= c(rep("small", length(yhat_small_tree)), rep("big",length(yhat_b
ig_tree))))

# in each plot notice the bias and variance
ggplot(errors,aes(error)) +geom_density()+facet_grid(flexibility~.)
```

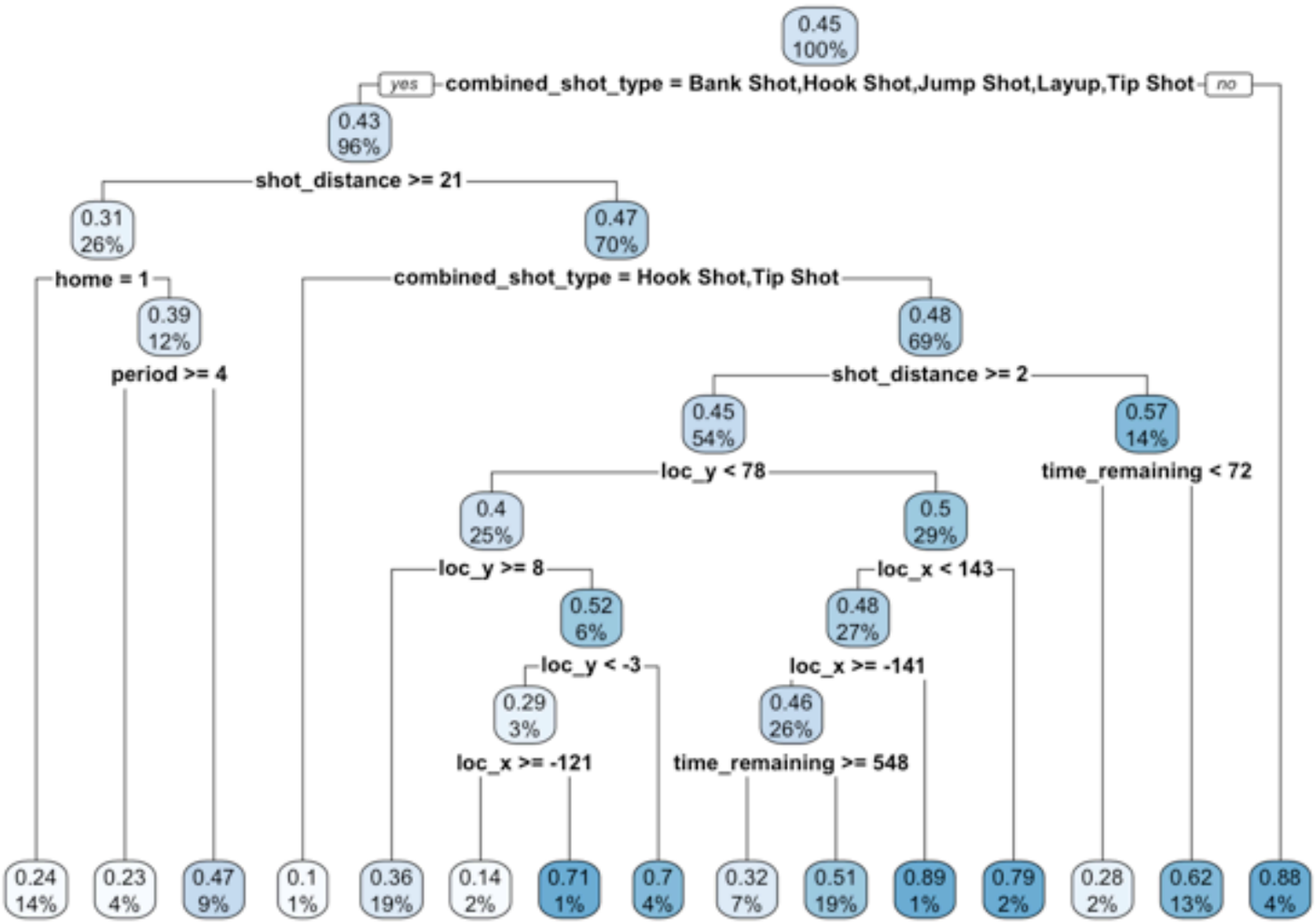



```
fit.tree_2 <- rpart(f2, dd_train,
control = rpart.control(cp = 0.006))
```

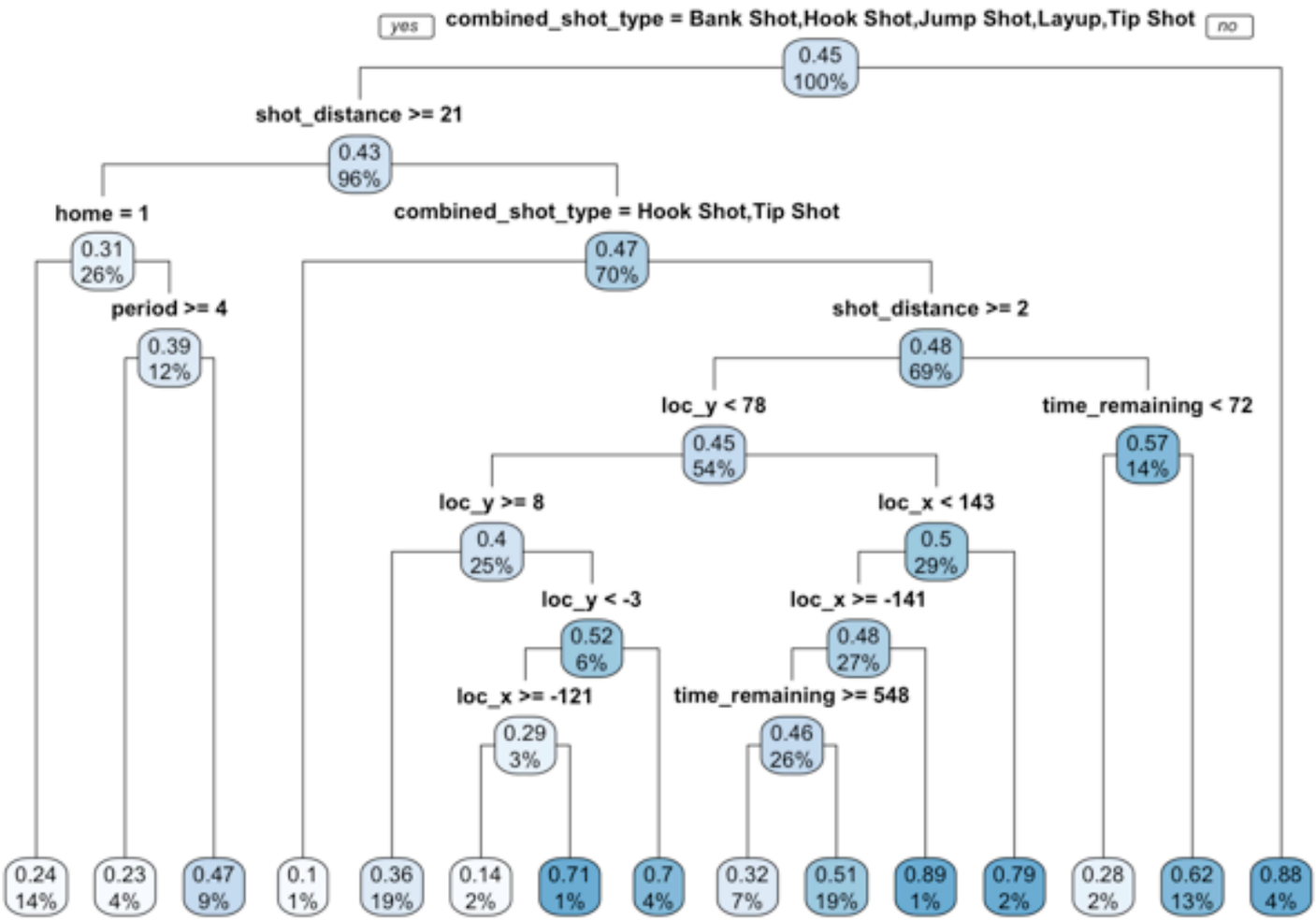
```
par(xpd = TRUE)
plot(fit.tree_2, compress=TRUE)
text(fit.tree_2, use.n=TRUE)
```




```
rpart.plot(fit.tree_2)
```




```
rpart.plot(fit.tree_2, type = 1)
```




```
yhat.train.tree_2 <- predict(fit.tree_2, dd_train)
mse.train.tree_2 <- mean((dd_train$shot_made_flag - yhat.train.tree_2)^2)
mse.train.tree_2
```

```
## [1] 0.2147198
```

```
yhat.test.tree_2 <- predict(fit.tree_2, dd_test)
mse.test.tree_2 <- mean((dd_test$shot_made_flag - yhat.test.tree_2)^2)
mse.test.tree_2
```

```
## [1] 0.2694032
```

Random Forests

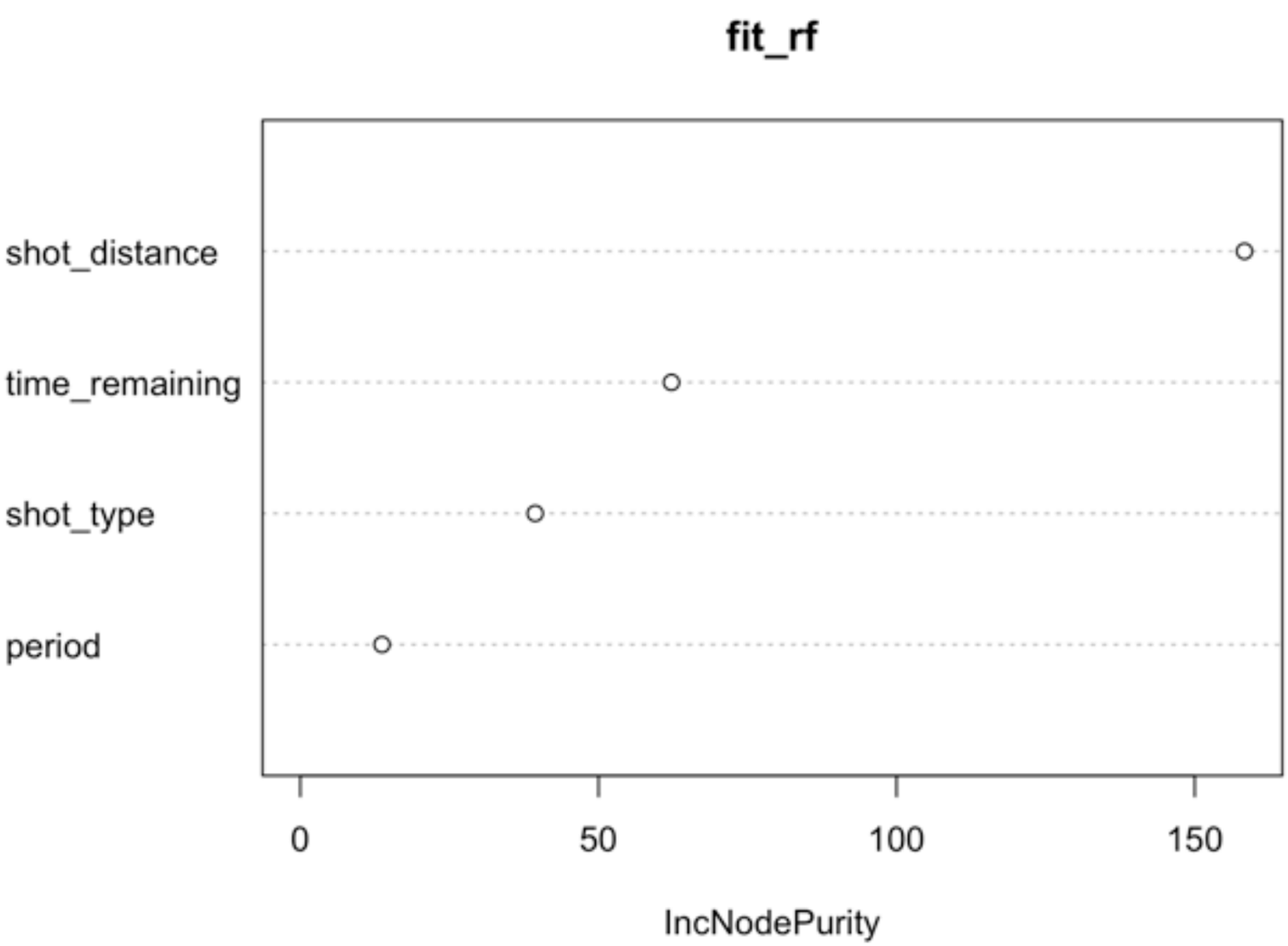
```
y_train2 <- dd2$shot_made_flag[-test_index2]
y_test2 <- dd2$shot_made_flag[test_index2]
```

```
# the [, -1] means take all columns of the matrix except the first column, # whic
h is an intercept added by default
x1_train <- model.matrix(f1, dd.train2)[, -1]
x1_test <- model.matrix(f1, dd.test2)[, -1]
```

```
fit_rf <- randomForest(f1, dd.train2,
                   ntree=100,
                   do.trace=F)
```

```
## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?
```

```
## Variable Importance Plot
varImpPlot(fit_rf)
```




```
yhat_rf <- predict(fit_rf, dd.train2)
mse_rf_small <- mean((yhat_rf - y_train2) ^ 2)
print(mse_rf_small)
```

```
## [1] 0.2317838
```

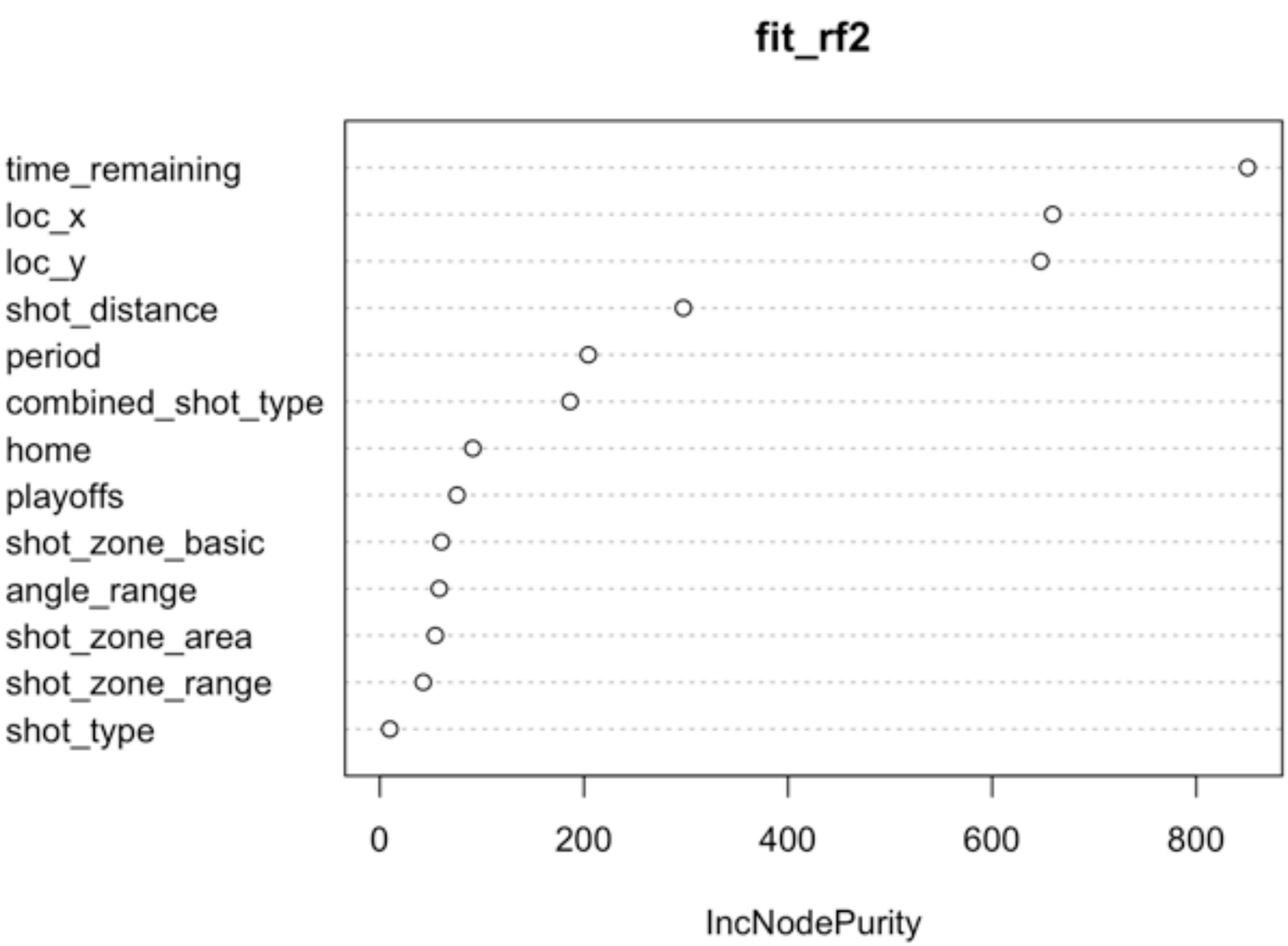
```
yhat_rf_test <- predict(fit_rf, dd.test2)
mse_rf_small_test <- mean((yhat_rf_test - y_test2) ^ 2)
print(mse_rf_small_test)
```

```
## [1] 0.236729
```

```
fit_rf2 <- randomForest(f2, dd.train2,
                  ntree=100,
                  do.trace=F)
```

```
## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?
```

```
varImpPlot(fit_rf2)
```



```
yhat_rf2 <- predict(fit_rf2, dd.train2)
mse_rf_big <- mean((yhat_rf2 - y_train2) ^ 2)
print(mse_rf_big)
```

```
## [1] 0.1030894
```

```
yhat_rf2_test <- predict(fit_rf2, dd.test2)
mse_rf_big_test <- mean((yhat_rf2_test - y_test2) ^ 2)
print(mse_rf_big_test)
```

```
## [1] 0.2395238
```

Notice: We want to get a large tree size in Random Forest

Boosted trees

```
fit_btree_small <- gbm(f1, data = dd.train2,
      distribution = "gaussian",
      n.trees = 100,
      interaction.depth = 2,
      shrinkage = 0.001)
```

```
relative.influence(fit_btree_small)
```

```
## n.trees not given. Using 100 trees.
```

```
##  shot_distance time_remaining         period      shot_type
##      9279.2153       801.3290         0.0000       282.6724
```

```
yhat_btree_small <- predict(fit_btree_small, dd.train2, n.trees = 100)
mse_btree_small <- mean((yhat_btree_small - y_train2) ^ 2)
print(mse_btree_small)
```

```
## [1] 0.2450115
```

```
yhat_btree_small_test <- predict(fit_btree_small, dd.test2, n.trees = 100)
mse_btree_small_test <- mean((yhat_btree_small_test - y_test2) ^ 2)
print(mse_btree_small_test)
```

```
## [1] 0.2456894
```

```
fit_btree_big <- gbm(f2, data = dd.train2,
      distribution = "gaussian",
      n.trees = 100,
      interaction.depth = 2,
      shrinkage = 0.001)
```

```
relative.influence(fit_btree_big)
```

```
## n.trees not given. Using 100 trees.
```

```
## combined_shot_type              loc_x              loc_y
##           13764.31               0.00               0.00
##             period           playoffs      shot_distance
##               0.00               0.00               0.00
##          shot_type     shot_zone_area    shot_zone_basic
##               0.00               0.00               0.00
##    shot_zone_range     time_remaining               home
##               0.00               0.00               0.00
##        angle_range
##               0.00
```

```
yhat_btree_big <- predict(fit_btree_big, dd.train2, n.trees = 100)
mse_btree_big <- mean((yhat_btree_big - y_train2) ^ 2)
print(mse_btree_big)
```

```
## [1] 0.2443108
```

```
yhat_btree_big_test <- predict(fit_btree_big, dd.test2, n.trees = 100)
mse_btree_big_test <- mean((yhat_btree_big_test - y_test2) ^ 2)
print(mse_btree_big_test)
```

```
## [1] 0.2448016
```

Notice: We want to get a small tree size in Boosting

Summary

After comparing all models of our dataset, we find out the best model is Lasso since it has the lowest MSE and did not overfitting.