Modeling Shot Efficiency in the NBA

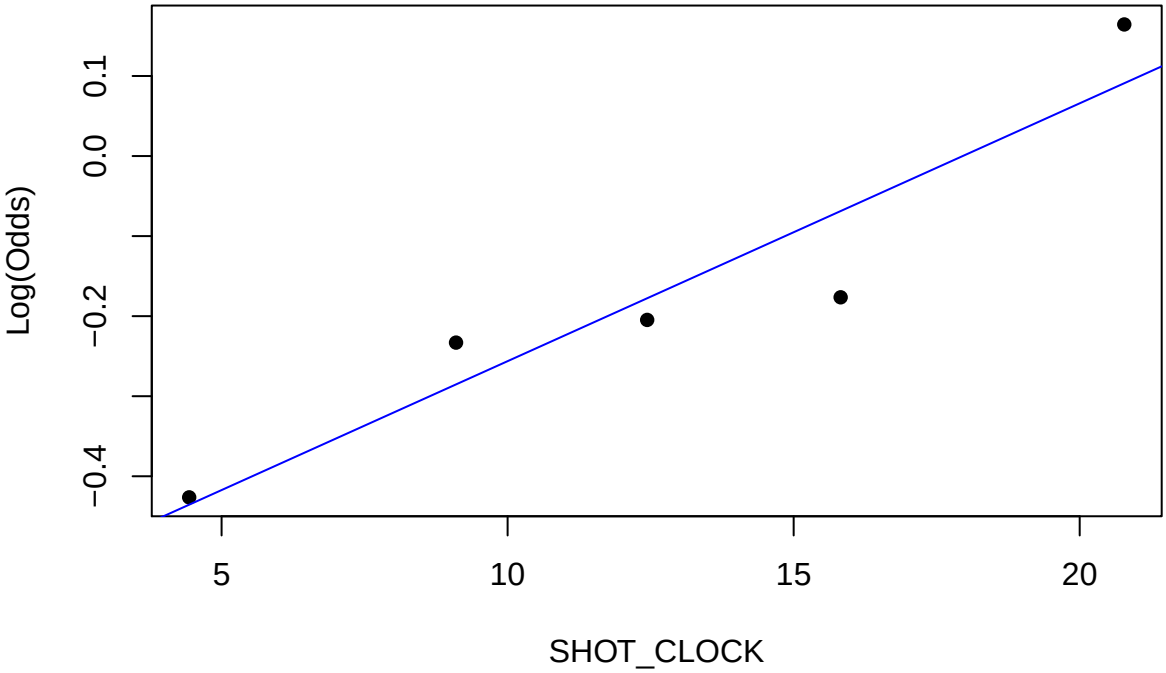
Stat guys: Lewis Eatherton, Team member 2, Team member 3, Team member 4

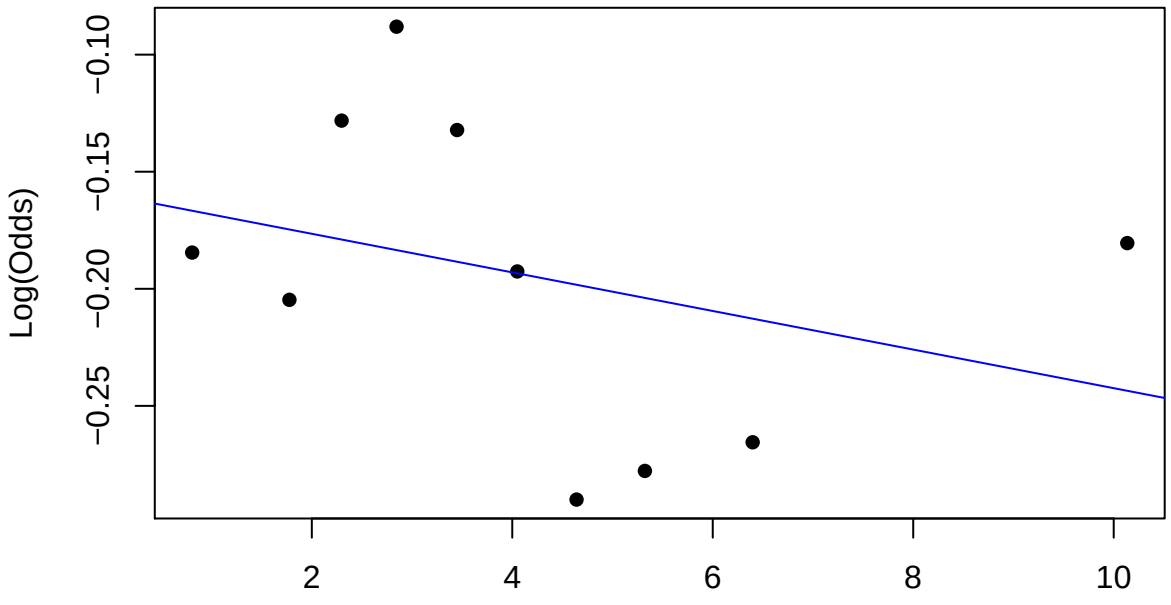
10/28/20

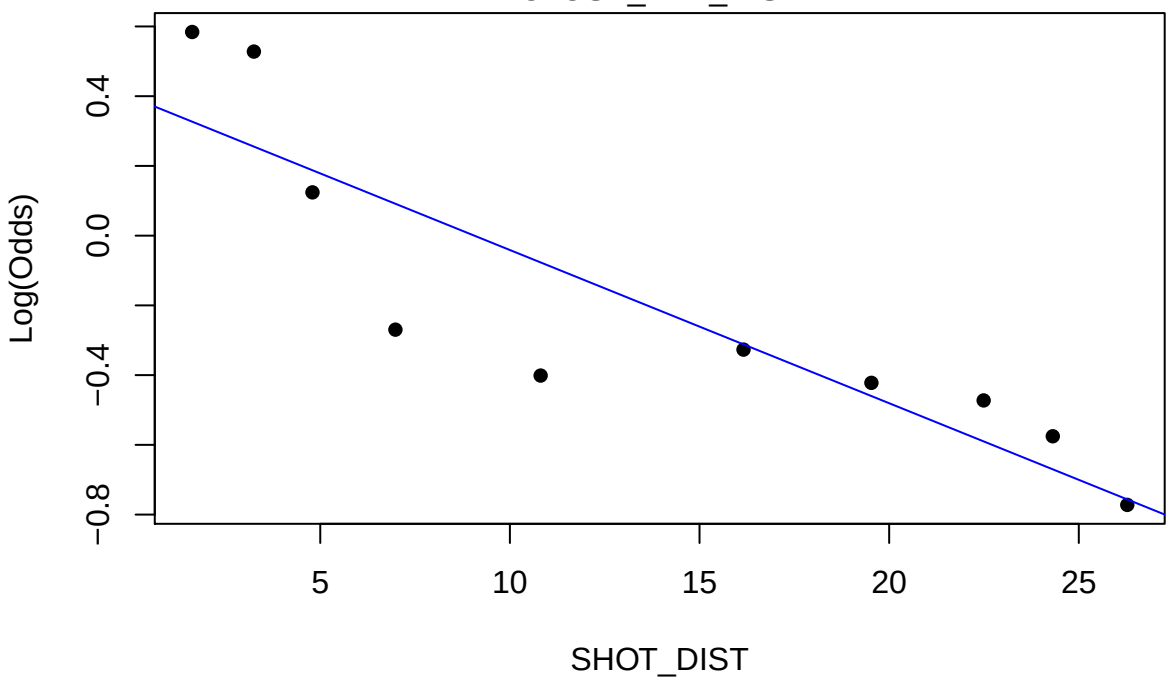
Your written report goes here! Before you submit, make sure your code chunks are turned off with `echo = FALSE` and there are no warnings or messages with `warning = FALSE` and `message = FALSE`

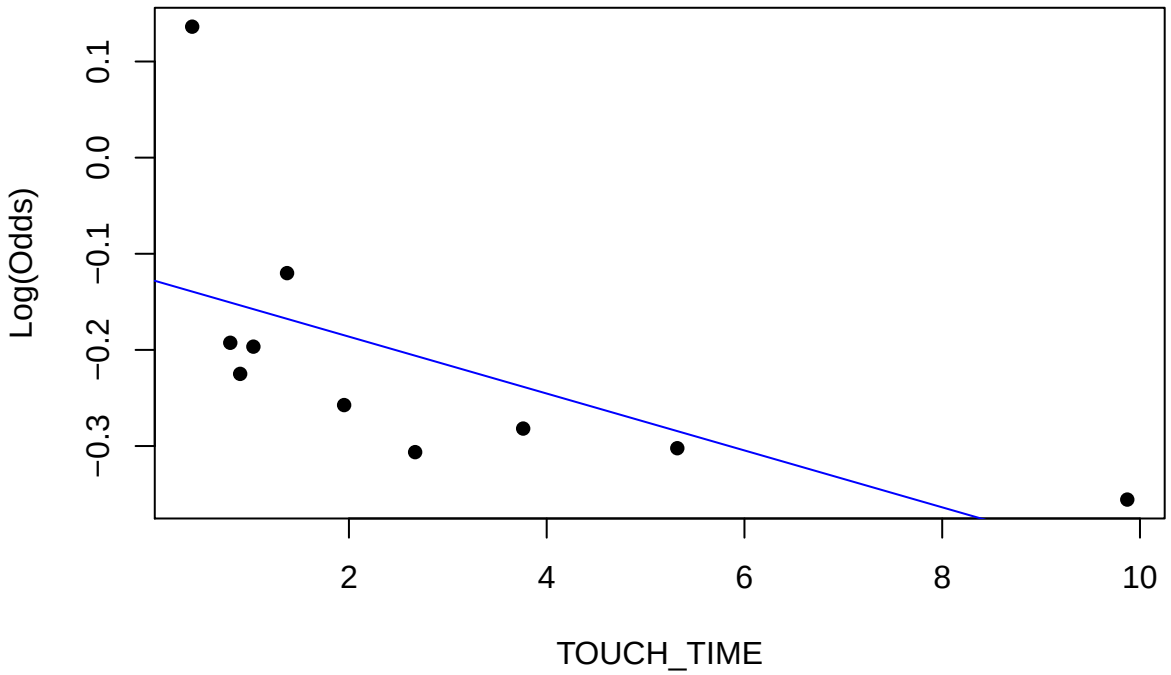
Introduction and EDA

talk about why we chose this topic and what our expectations our... similar to proposal

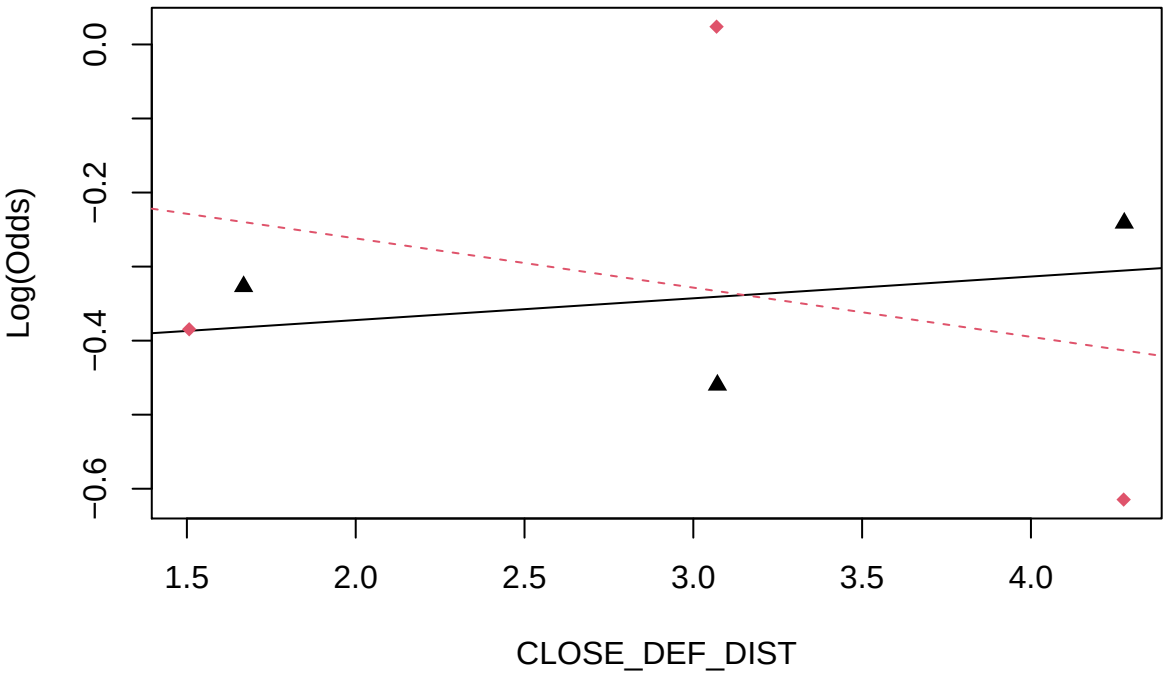



Log(Odds)
−0.10
−0.15
−0.20
−0.25
2
4
6
8
10
CLOSE_DEF_DIST


Log(Odds)
0.4
0.0
−0.4
−0.8
5
10
15
20
25
SHOT_DIST




```
## # A tibble: 470 x 5
## # Groups:   CLOSEST_DEFENDER [470]
##    CLOSEST_DEFENDER   Make      n  prop emp_logit
##    <chr>              <fct> <int> <dbl>     <dbl>
##  1 Acy, Quincy        1       118 0.428    -0.292
##  2 Adams, Jordan      1        16 0.533     0.134
##  3 Adams, Steven      1       215 0.444    -0.224
##  4 Adrien, Jeff       1        40 0.548     0.192
##  5 Afflalo, Arron     1       191 0.417    -0.335
##  6 Ajinca, Alexis     1       114 0.465    -0.139
##  7 Aldemir, Furkan    1        33 0.465    -0.141
##  8 Aldrich, Cole      1       142 0.532     0.128
##  9 Aldridge, LaMarcus 1       302 0.461    -0.156
## 10 Allen, Lavoy       1       141 0.449    -0.205
## # ... with 460 more rows
```



Linearity is not totally satisfied... graphs aren't linear

independence and randomness seems fine accorind to how data was collected

we should also look into collinearity

Creating Model

```
## Single term deletions
##
## Model:
## Make ~ SHOT_CLOCK + DRIBBLES + TOUCH_TIME + SHOT_DIST + CLOSE_DEF_DIST
##                Df Deviance    AIC
## <none>              162325 162337
## SHOT_CLOCK      1   162532 162542
## DRIBBLES        1   162354 162364
## TOUCH_TIME      1   162434 162444
## SHOT_DIST       1   167561 167571
## CLOSE_DEF_DIST  1   163764 163774
```

```
## Single term deletions
##
## Model:
## Make ~ SHOT_CLOCK + DRIBBLES + TOUCH_TIME + SHOT_DIST + CLOSE_DEF_DIST +
##     CLOSEST_DEFENDER
##                   Df Deviance    AIC
## <none>                 855.93 869.93
## SHOT_CLOCK         1   858.72 870.72
## DRIBBLES           1   855.93 867.93
## TOUCH_TIME         1   855.94 867.94
## SHOT_DIST          1   872.34 884.34
## CLOSE_DEF_DIST     1   865.25 877.25
## CLOSEST_DEFENDER   1   855.98 867.98
```

```
## Single term deletions
```

```
##
## Model:
## Make ~ SHOT_CLOCK + TOUCH_TIME + SHOT_DIST + CLOSE_DEF_DIST +
##     CLOSEST_DEFENDER
##                   Df Deviance    AIC
## <none>                 855.93 867.93
## SHOT_CLOCK         1   858.81 868.81
## TOUCH_TIME         1   855.99 865.99
## SHOT_DIST          1   872.34 882.34
## CLOSE_DEF_DIST     1   865.25 875.25
## CLOSEST_DEFENDER   1   855.98 865.98
```

```
## Single term deletions
##
## Model:
## Make ~ SHOT_CLOCK + TOUCH_TIME + SHOT_DIST + CLOSE_DEF_DIST
##                 Df Deviance    AIC
## <none>               855.98 865.98
## SHOT_CLOCK       1   858.90 866.90
## TOUCH_TIME       1   856.02 864.02
## SHOT_DIST        1   872.45 880.45
## CLOSE_DEF_DIST   1   865.36 873.36
```

```
## Single term deletions
##
## Model:
## Make ~ SHOT_CLOCK + SHOT_DIST + CLOSE_DEF_DIST
##                 Df Deviance    AIC
## <none>               856.02 864.02
## SHOT_CLOCK       1   859.04 865.04
## SHOT_DIST        1   872.60 878.60
## CLOSE_DEF_DIST   1   865.68 871.68
```

```
## # A tibble: 2 x 5
##   Resid..Df Resid..Dev    df Deviance p.value
##       <dbl>      <dbl> <dbl>    <dbl>   <dbl>
## 1       645       856.    NA   NA      NA
## 2       643       855.     2    0.813   0.666
```

term	estimate	std.error	statistic	p.value
(Intercept)	-0.737	0.294	-2.508	0.012
SHOT_CLOCK	0.026	0.015	1.733	0.083
SHOT_DIST	-0.048	0.012	-4.013	0.000
CLOSE_DEF_DIST	0.243	0.079	3.067	0.002

talk about final model outcome and how we came to it

###Discussion

Talk about our results, the limitations of these results, and what'd we do differently...