THE UNIVERSITY OF TEXAS AT AUSTIN
McCOMBS SCHOOL OF BUSINESS

Interactions 2

Lecture 16

STA 371G

NBA data

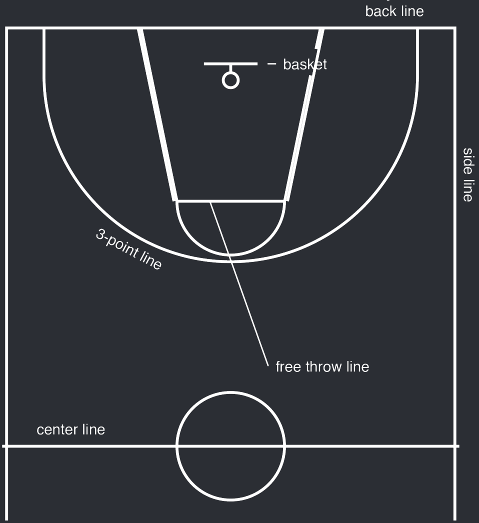
Basketball-Reference.com provides detailed data on NBA teams and players. We'll look at team data for 4 seasons ending in 2016; each of these metrics is the average across the season:

- **PTS**: Total points
- **PCT3P**: Percentage of 3-point shots made
- **N3PA**: Number of 3-point shots attempted

There are 30 NBA teams $\times$ 4 seasons = 120 cases in this file.

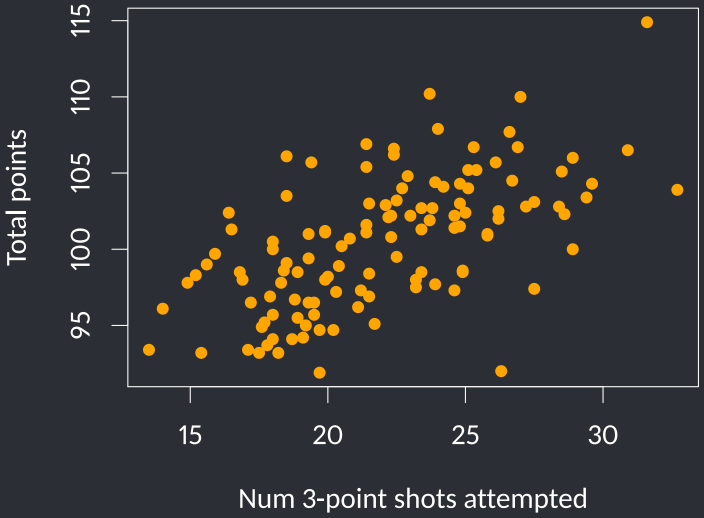
NBA data

In basketball, there are three ways to score:



- **1 point** for free throws made after a foul by the other team
- **2 points** for shots made inside the 3-point line
- **3 points** for shots made outside the 3-point line

```r
plot(nba$N3PA, nba$PTS, pch=16, col='orange',
  xlab='Num 3-point shots attempted', ylab='Total points')
```



```r
model1 <- lm(PTS ~ N3PA, data=nba)
summary(model1)
```

```
Call:
lm(formula = PTS ~ N3PA, data = nba)

Residuals:
    Min      1Q  Median      3Q     Max 
-11.245  -2.511   0.055   2.225   8.640 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  86.1920     1.7746   48.57  < 2e-16 ***
N3PA          0.6484     0.0794    8.17  3.9e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.5 on 118 degrees of freedom
Multiple R-squared:  0.361,Adjusted R-squared:  0.356 
F-statistic: 66.8 on 1 and 118 DF,  p-value: 3.89e-13
```

Can we do better?

$R^2 = 36\%$, so we can explain 36% of the variance in total points based only on knowing the number of 3-point attempts.

Can we do better?

$R^2 = 36\%$, so we can explain 36% of the variance in total points based only on knowing the number of 3-point attempts.

This means that **most** of the variance (64%) in total points is **not** explained by the number of 3-point attempts.

Can we do better?

$R^2 = 36\%$, so we can explain 36% of the variance in total points based only on knowing the number of 3-point attempts.

This means that **most** of the variance (64%) in total points is **not** explained by the number of 3-point attempts.

Let's add another variable to our model — why might 3-point percentage be useful as another predictor?

Can we do better?

```
model2 <- lm(PTS ~ N3PA + PCT3P, data=nba)
summary(model2)



Call:
lm(formula = PTS ~ N3PA + PCT3P, data = nba)

Residuals:
   Min     1Q Median     3Q    Max 
-8.349 -2.139 -0.079  1.869  9.190 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  62.0049     5.6140   11.04  < 2e-16 ***
N3PA          0.5647     0.0759    7.44  1.8e-11 ***
PCT3P         0.7342     0.1629    4.51  1.6e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.2 on 117 degrees of freedom
Multiple R-squared:  0.456,Adjusted R-squared:  0.447 
F-statistic:   49 on 2 and 117 DF,  p-value: 3.48e-16
```

Can we do even better?

It would make sense that the **impact** of the number of 3-pointers taken on total points would **depend on** how well the team shoots the 3!

Can we do even better?

It would make sense that the **impact** of the number of 3-pointers taken on total points would **depend on** how well the team shoots the 3!

This sounds like an interaction — let's make a model with an interaction between the two predictors!

```r
model3 <- lm(PTS ~ N3PA * PCT3P, data=nba)
summary(model3)
```

```


Call:
lm(formula = PTS ~ N3PA * PCT3P, data = nba)

Residuals:
   Min     1Q Median     3Q    Max
-7.263 -2.276  0.115  1.970  9.376

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 122.8490    30.5894    4.02  0.00011 ***
N3PA         -2.1190     1.3290   -1.59  0.11356
PCT3P        -0.9841     0.8646   -1.14  0.25740
N3PA:PCT3P    0.0756     0.0374    2.02  0.04542 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.2 on 116 degrees of freedom
Multiple R-squared:  0.474,Adjusted R-squared:  0.461
F-statistic: 34.9 on 3 and 116 DF,  p-value: 3.8e-16
```

Model 3 corresponds to the regression equation

$$\widehat{\text{PTS}} = 122.85 - 2.12 \cdot \text{N3PA} - 0.98 \cdot \text{PCT3P} + 0.08 \cdot \text{N3PA} \cdot \text{PCT3P}.$$

Model 3 corresponds to the regression equation

$$\widehat{\text{PTS}} = 122.85 - 2.12 \cdot \text{N3PA} - 0.98 \cdot \text{PCT3P} + 0.08 \cdot \text{N3PA} \cdot \text{PCT3P}.$$

We interpret the coefficients as follows:

- **Intercept** (122.85) is our prediction of total points when $\text{N3PA} = \text{PCT3P} = 0$. (Meaningless in this context!)

Model 3 corresponds to the regression equation

$$\widehat{\text{PTS}} = 122.85 - 2.12 \cdot \text{N3PA} - 0.98 \cdot \text{PCT3P} + 0.08 \cdot \text{N3PA} \cdot \text{PCT3P}.$$

We interpret the coefficients as follows:

- **Intercept** (122.85) is our prediction of total points when $\text{N3PA} = \text{PCT3P} = 0$. (Meaningless in this context!)
- **N3PA** (-2.12) is the predicted increase in total points for each additional 3-pointer taken, when $\text{PCT3P} = 0$.

Model 3 corresponds to the regression equation

$$\widehat{\text{PTS}} = 122.85 - 2.12 \cdot \text{N3PA} - 0.98 \cdot \text{PCT3P} + 0.08 \cdot \text{N3PA} \cdot \text{PCT3P}.$$

We interpret the coefficients as follows:

- **Intercept** (122.85) is our prediction of total points when $\text{N3PA} = \text{PCT3P} = 0$. (Meaningless in this context!)
- **N3PA** (-2.12) is the predicted increase in total points for each additional 3-pointer taken, when $\text{PCT3P} = 0$.
- **PCT3P** (-0.98) is the predicted increase in total points for each additional percentage point of 3-point shooting accuracy, when $\text{N3PA} = 0$.

Model 3 corresponds to the regression equation

$$\widehat{\text{PTS}} = 122.85 - 2.12 \cdot \text{N3PA} - 0.98 \cdot \text{PCT3P} + 0.08 \cdot \text{N3PA} \cdot \text{PCT3P}.$$

We interpret the coefficients as follows:

- **Intercept** (122.85) is our prediction of total points when $\text{N3PA} = \text{PCT3P} = 0$. (Meaningless in this context!)
- **N3PA** (-2.12) is the predicted increase in total points for each additional 3-pointer taken, when $\text{PCT3P} = 0$.
- **PCT3P** (-0.98) is the predicted increase in total points for each additional percentage point of 3-point shooting accuracy, when $\text{N3PA} = 0$.
- **N3PA · PCT3P** (0.08) can be interpreted in two ways:

Model 3 corresponds to the regression equation

$$\widehat{\text{PTS}} = 122.85 - 2.12 \cdot \text{N3PA} - 0.98 \cdot \text{PCT3P} + 0.08 \cdot \text{N3PA} \cdot \text{PCT3P}.$$

We interpret the coefficients as follows:

- **Intercept** (122.85) is our prediction of total points when $\text{N3PA} = \text{PCT3P} = 0$. (Meaningless in this context!)
- **N3PA** (-2.12) is the predicted increase in total points for each additional 3-pointer taken, when $\text{PCT3P} = 0$.
- **PCT3P** (-0.98) is the predicted increase in total points for each additional percentage point of 3-point shooting accuracy, when $\text{N3PA} = 0$.
- **N3PA · PCT3P** (0.08) can be interpreted in two ways:

Model 3 corresponds to the regression equation

$$\widehat{\mathrm{PTS}} = 122.85 - 2.12 \cdot \mathrm{N3PA} - 0.98 \cdot \mathrm{PCT3P} + 0.08 \cdot \mathrm{N3PA} \cdot \mathrm{PCT3P}.$$

We interpret the coefficients as follows:

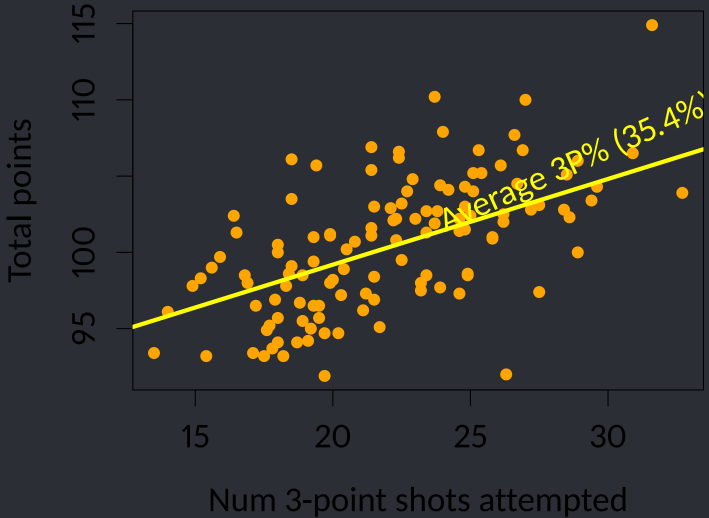
- **Intercept** (122.85) is our prediction of total points when $\mathrm{N3PA} = \mathrm{PCT3P} = 0$. (Meaningless in this context!)
- **N3PA** (-2.12) is the predicted increase in total points for each additional 3-pointer taken, when $\mathrm{PCT3P} = 0$.
- **PCT3P** (-0.98) is the predicted increase in total points for each additional percentage point of 3-point shooting accuracy, when $\mathrm{N3PA} = 0$.
- **N3PA · PCT3P** (0.08) can be interpreted in two ways:
  - the increase in the *slope coefficient* for N3PA for each 1-unit increase of PCT3P.

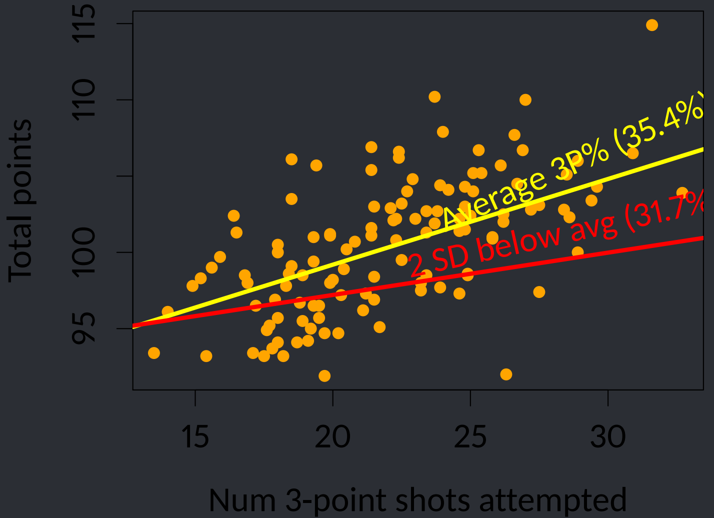
Model 3 corresponds to the regression equation

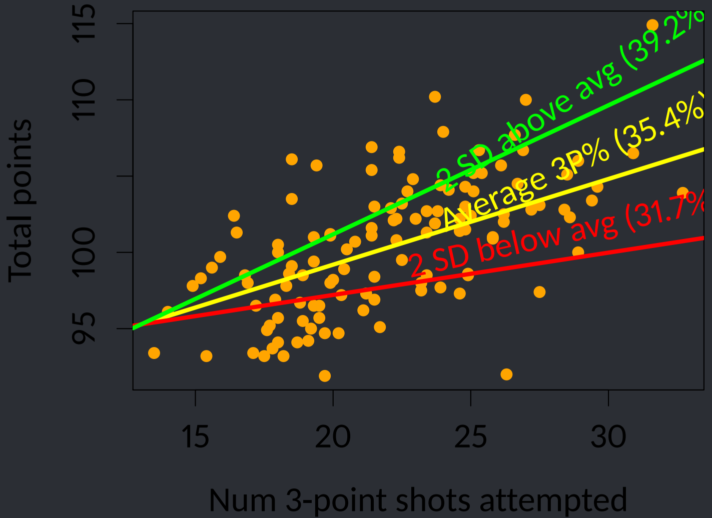
$$\widehat{\text{PTS}} = 122.85 - 2.12 \cdot \text{N3PA} - 0.98 \cdot \text{PCT3P} + 0.08 \cdot \text{N3PA} \cdot \text{PCT3P}.$$

We interpret the coefficients as follows:

- **Intercept** (122.85) is our prediction of total points when N3PA $=$ PCT3P $= 0$. (Meaningless in this context!)
- **N3PA** (-2.12) is the predicted increase in total points for each additional 3-pointer taken, when PCT3P $= 0$.
- **PCT3P** (-0.98) is the predicted increase in total points for each additional percentage point of 3-point shooting accuracy, when N3PA $= 0$.
- **N3PA · PCT3P** (0.08) can be interpreted in two ways:
  - the increase in the *slope coefficient* for N3PA for each 1-unit increase of PCT3P.
  - the increase in the *slope coefficient* for PCT3P for each 1-unit increase of N3PA.


Total points
115
110
100
95
Average 3P% (35.4%)
15
20
25
30
Num 3-point shots attempted


Average 3P% (35.4%)
2 SD below avg (31.7%)
Total points
95
100
110
115
15
20
25
30
Num 3-point shots attempted



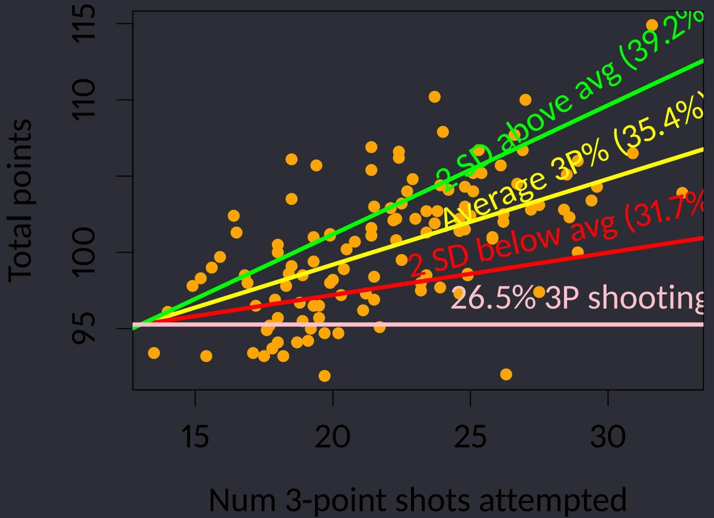
Total points
115
110
100
95
15
20
25
30
Num 3-point shots attempted
2 SD above avg (39.2%
Average 3P% (35.4%)
2 SD below avg (31.7%

$$\widehat{\text{PTS}} = 122.85 - 2.12 \cdot \text{N3PA} - 0.98 \cdot \text{PCT3P} + 0.08 \cdot \text{N3PA} \cdot \text{PCT3P}.$$

- How many points per game do you predict for a team that shoots 3-pointers at the NBA average rate (35.4) and that takes 30 3-pointers per game?

$$\widehat{\text{PTS}} = 122.85 - 2.12 \cdot \text{N3PA} - 0.98 \cdot \text{PCT3P} + 0.08 \cdot \text{N3PA} \cdot \text{PCT3P}.$$

- How many points per game do you predict for a team that shoots 3-pointers at the NBA average rate (35.4) and that takes 30 3-pointers per game?
- How bad would a team have to shoot the 3 before taking 3-point shots start to have a negative impact on total points?


2 SD above avg (39.2%
Average 3P% (35.4%)
2 SD below avg (31.7%
26.5% 3P shooting
Total points
95
100
110
115
15
20
25
30
Num 3-point shots attempted