

Lebron James Turnovers

@jonmichaelstroh

12/17/2021

On the NBA podcast The Lowe Post, an ESPN writer, I believe Dave McMenamin, said that LeBron's turnovers decrease as the season progresses: prior to LBJ's 37th b-day.

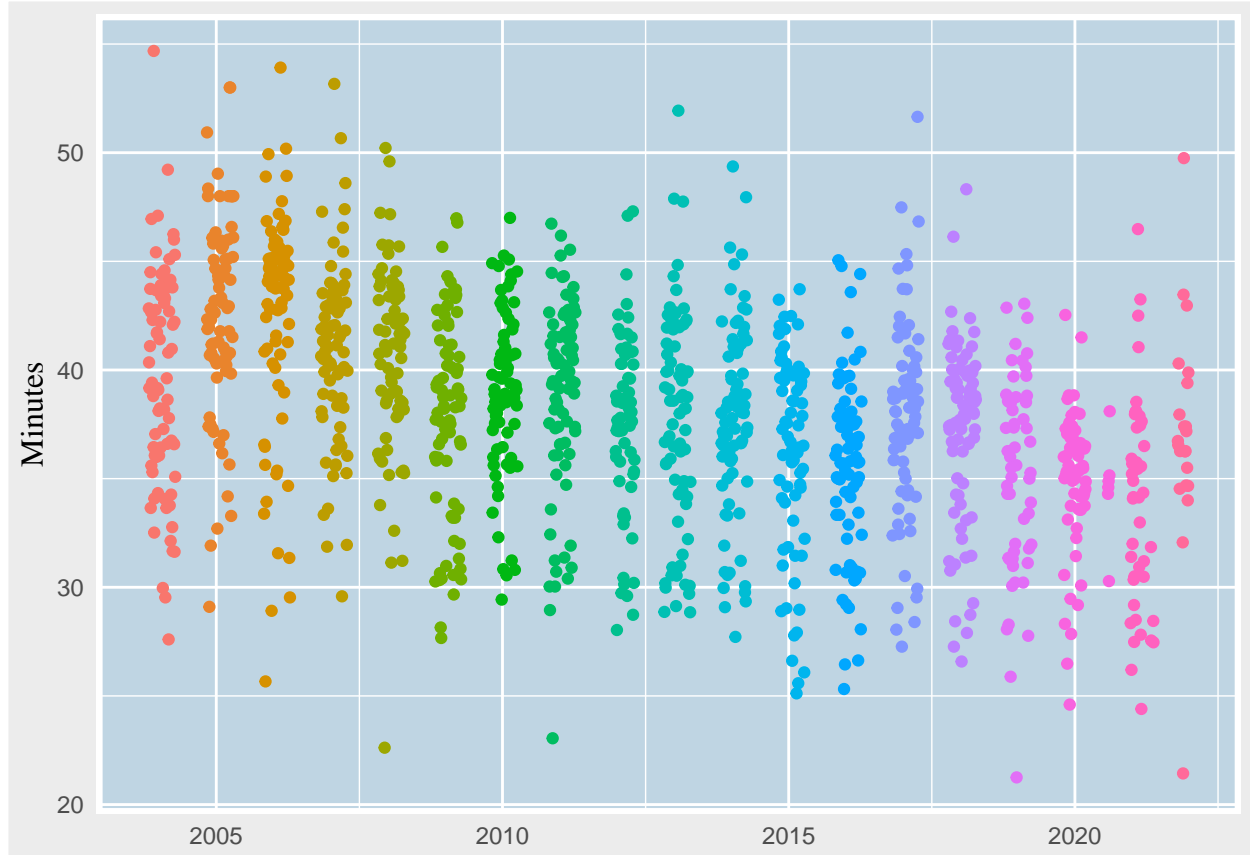
Our data is a complete game log for LeBron to 12/25/21: right before his 37th birthday.

```
james_log <- james_log %>% drop_na(G)
james_log <- james_log %>% mutate(Season = as.factor(Season)) %>% mutate(G = as.numeric(G)) %>% mutate(G = as.numeric(G))
```

```
james_log <- james_log %>% filter(MP >= 20)
```

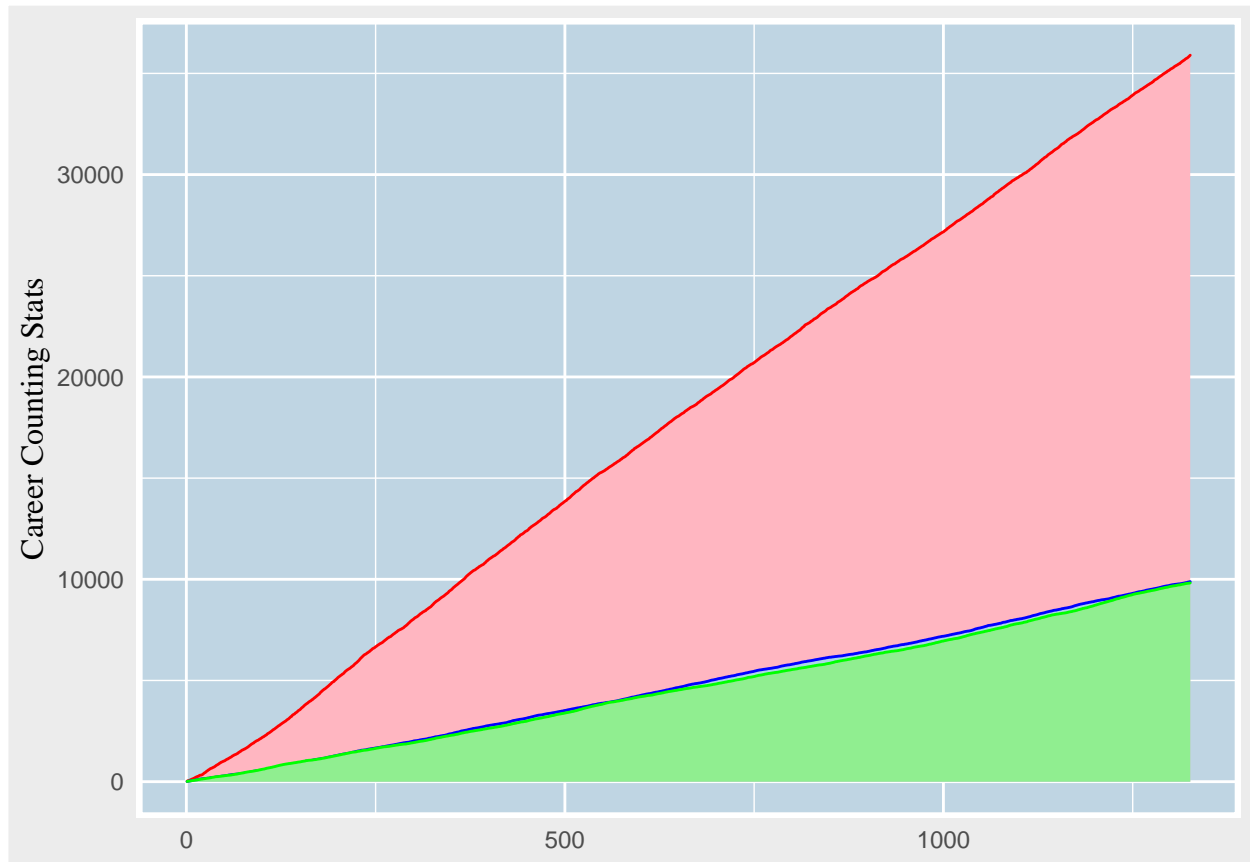
Distribution of regular season minutes per game over time. Kinda neat. Clearly, James was worked quite hard early in his career and has seen his minute load decrease with age:

```
ggplot(data = james_log, aes(x = Date, y = MP)) + geom_point(aes(color = Season)) + theme(legend.position = "right")
```



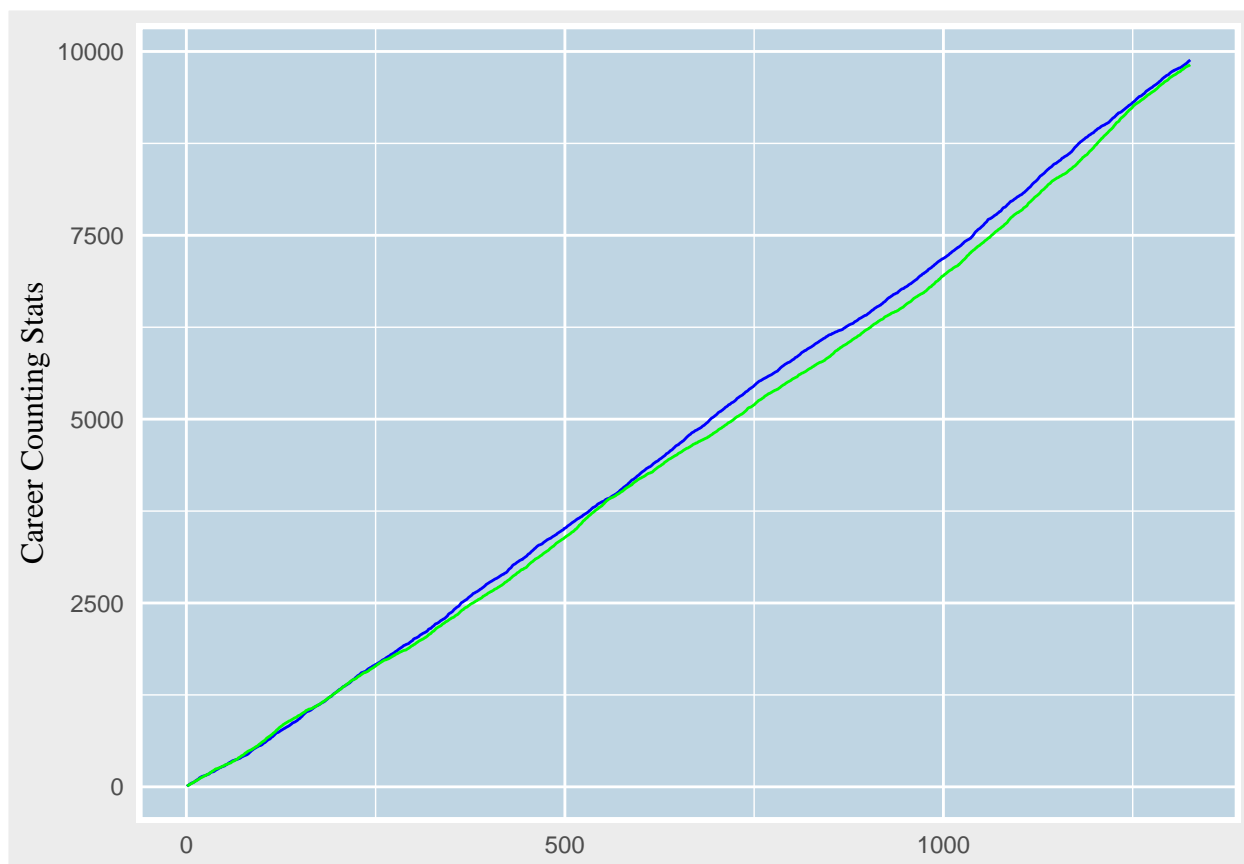
```
james_log <- james_log %>% mutate(careerpoints = cumsum(PTS)) %>% mutate(careergames = cumsum(G)) %>% m

ggplot(data = james_log, aes(x= careergames)) + geom_area(aes(y = careerpoints), color = 'red', fill =
```



Red is clearly points, blue is rebounds, and green is assists. Absolutely wild how close LBJ's assist and rebound career totals have stayed throughout his career.

```
ggplot(data = james_log, aes(x= careergames)) + geom_line(aes(y = careerrebounds), color = 'blue') + ge
```



```
james_log_same <- james_log %>% filter(careerassists == careerrebounds)
```

The last time LeBron had the same career assists and rebounds was the 2006 season when both were 1464. Rebounds have held the edge throughout most of his career. However, after LBJ's league leading 10.2 assist 2019-20 season, things are getting a bit tighter. We need the lines to intersect!!

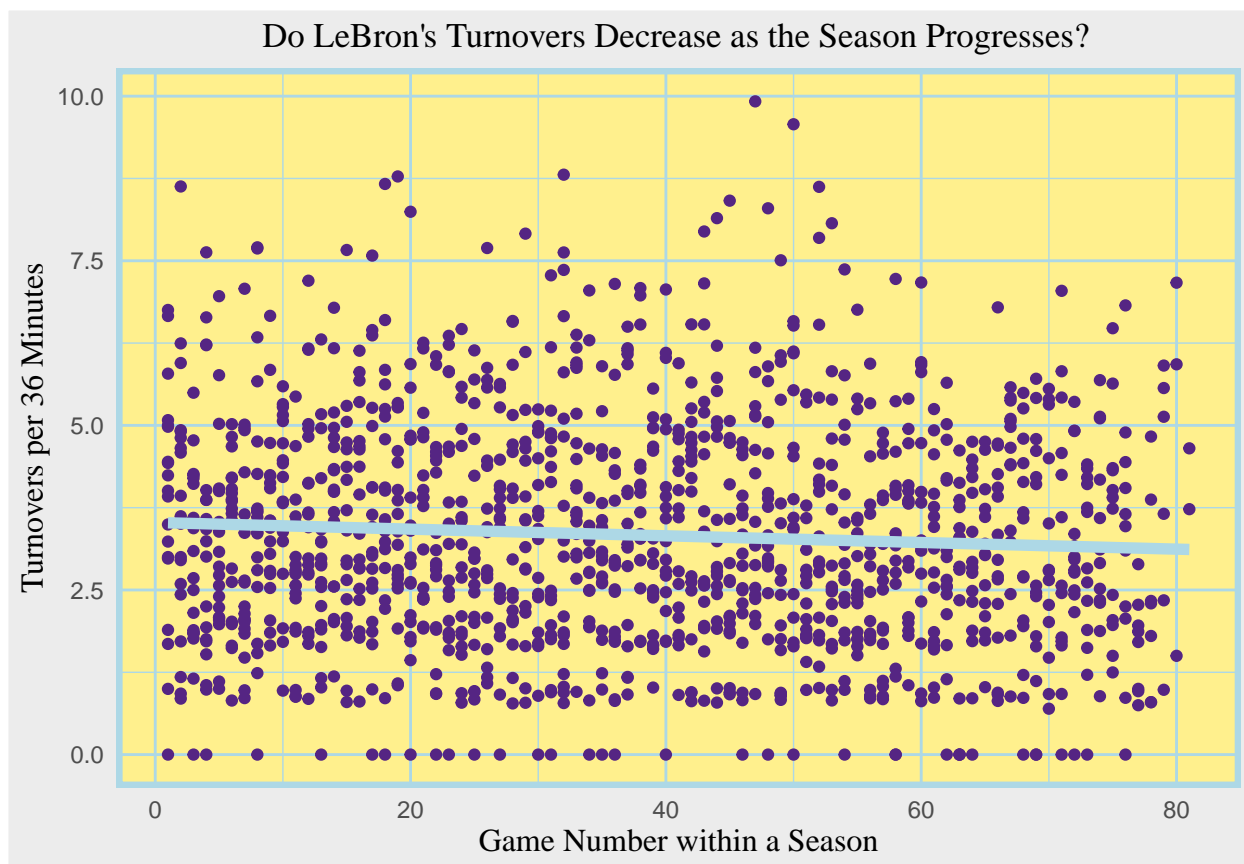
Now the main course – do LBJ's turnovers decrease as the season progresses:

```
james_log_over30 <- james_log %>% filter(MP >= 30)
```

Hmmm maybe they do slightly...

```
ggplot(data = james_log, aes(x= G, y = TOV/MP * 36)) + geom_point(color = '#552583') + geom_smooth(meth
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



```
james_log %>% summarise(mean(TOV/MP * 36))
```

```
## # A tibble: 1 x 1
##   'mean(TOV/MP * 36)'
##   <dbl>
## 1         3.34
```

```
james_log_firstx <- james_log %>% filter(G >= 63)
james_log_firstx %>% summarise(mean(TOV/MP * 36))
```

```
## # A tibble: 1 x 1
##   'mean(TOV/MP * 36)'
##   <dbl>
## 1         3.09
```

LBJ's career turnovers per 36 min is 3.34, after game 63 it is 3.09

```
tovsdecrease <- lm(data= james_log, TOV/MP*36 ~ G)
summary(tovsdecrease)
```

```
##
## Call:
## lm(formula = TOV/MP * 36 ~ G, data = james_log)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.520 -1.325 -0.157  1.211  6.633
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.525305   0.092657  38.047  <2e-16 ***
## G            -0.005062   0.002150  -2.355   0.0187 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.721 on 1324 degrees of freedom
## Multiple R-squared:  0.00417,    Adjusted R-squared:  0.003418
## F-statistic: 5.545 on 1 and 1324 DF,  p-value: 0.01868
```

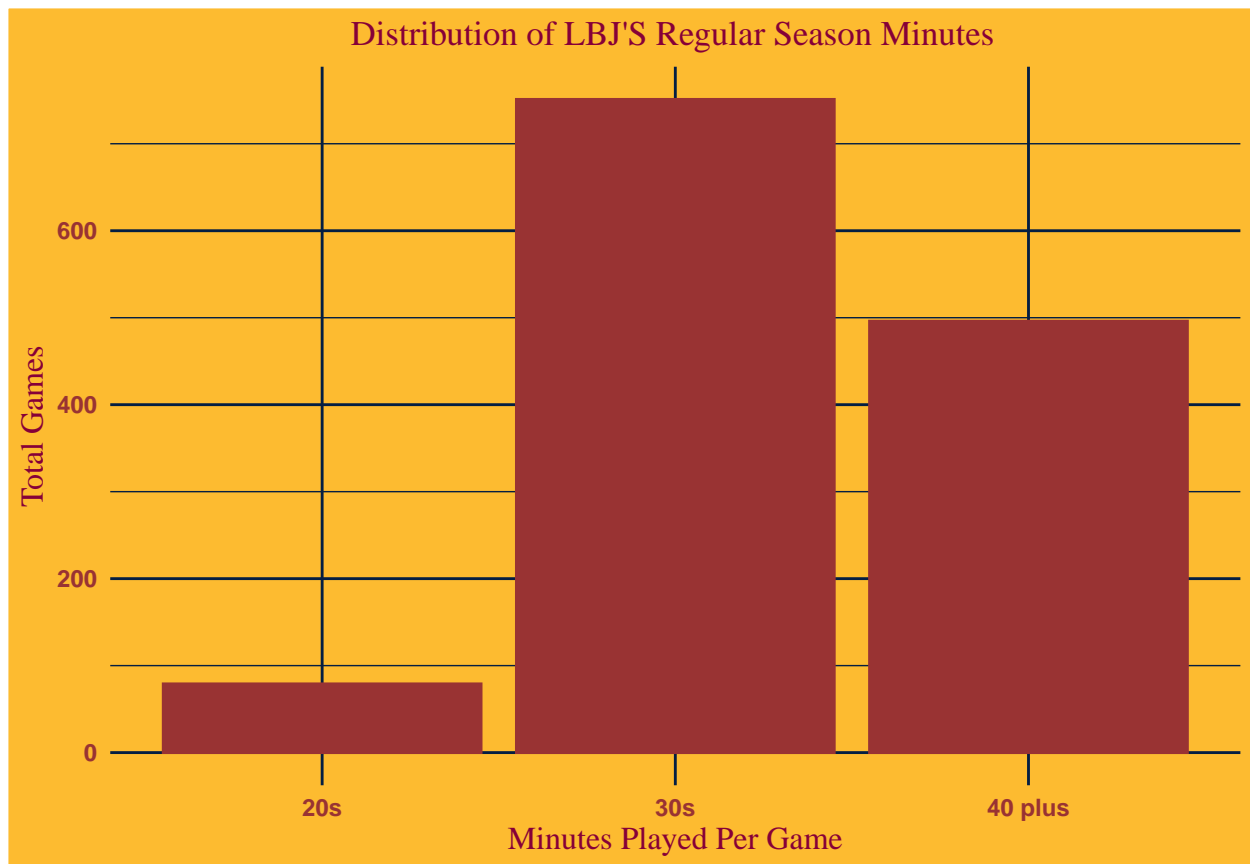
P-value is .059 which is so barely above the traditional statistical significance threshold.

Proportion of different games by minutes

```
james_log <- james_log %>% mutate(overx = if_else(MP >= 40, "40 plus", if_else(MP >= 30, "30s", if_else(MP >= 20, "20s", "Less than 20"))))
james_log$overx <- factor(james_log$overx, ordered = TRUE, levels = c('Less than 20', '20s', '30s', '40 plus'))
```

Yeah, just some fun graphs -> this dude has played a ton of high minute games:

```
ggplot(data = james_log, aes(x= overx)) + geom_bar(color = "#993333", fill = "#993333")+ theme(legend.position = "none")
```



```
my_summary_data <- james_log %>%  
  group_by(overx) %>%  
  summarise(Count = n())
```

```
ggplot(data = james_log, aes(x= MP)) + geom_density(fill = "#98002E", color = "#98002E") + theme(legend
```

