its_unfair

Joe Parker

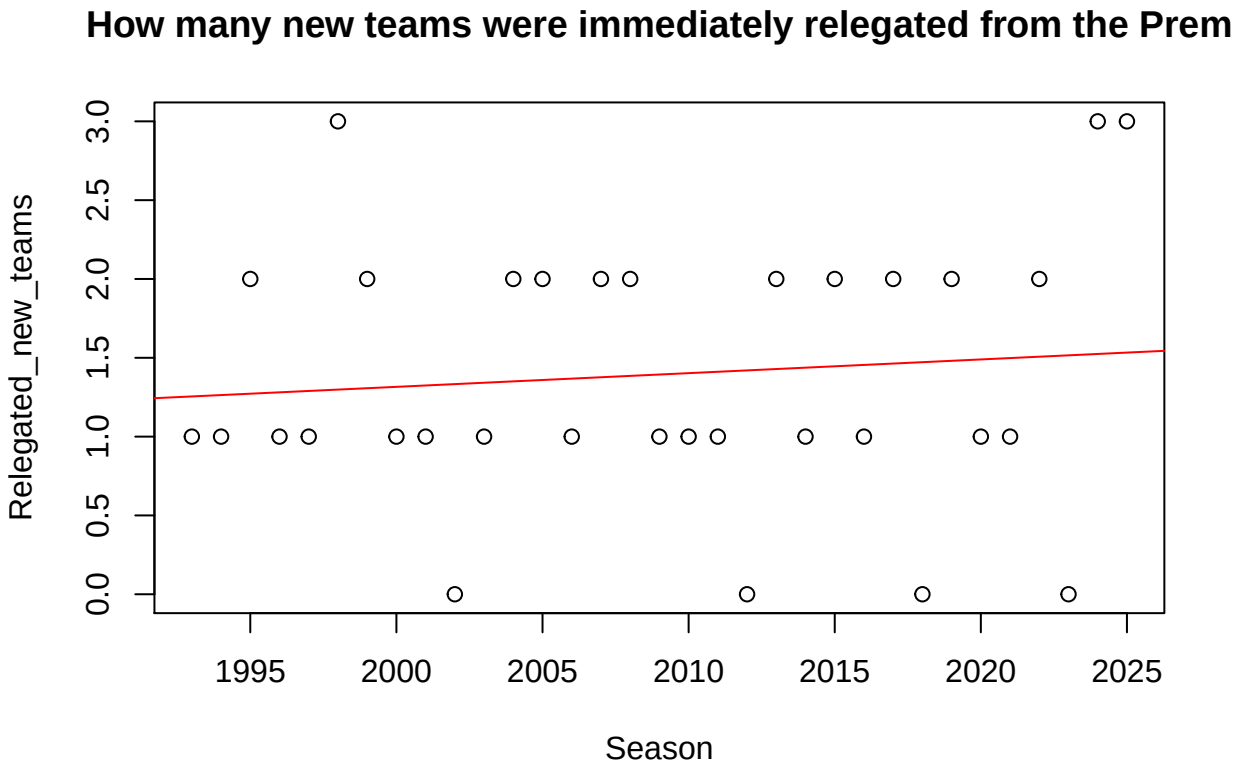
2025-05-16

is the prem a stitchup these days?

I went to https://www.11v11.com/league-tables/premier-league/2025/ to look up all the historical prem data for 1992-93 up to now (2024-25, Fri 16th May with 36 games played.) These I saved in a .csv at `relegation_stats.csv`. So now I can import data from prem historical. this includes:

1. season
2. number of new teams who were immediately relegated
3. moving average of three years of (2)
4. moving average of five years of (2)
5. Pts for the champions
6. Pts for the wooden-spoon
7. Gap between relegated teams (18th vs 17th place)
8. Points spread (best-worst)
9. Notes

```
prem_title='How many new teams were immediately relegated from the Prem'
relegation=read.csv('relegation_stats.csv',header=T)
my_lm=lm(Relegated_new_teams ~ Season, data=relegation)
plot(Relegated_new_teams ~ Season, data=relegation, main=prem_title)
abline(my_lm,col='Red')
```




```
summary(my_lm)
```

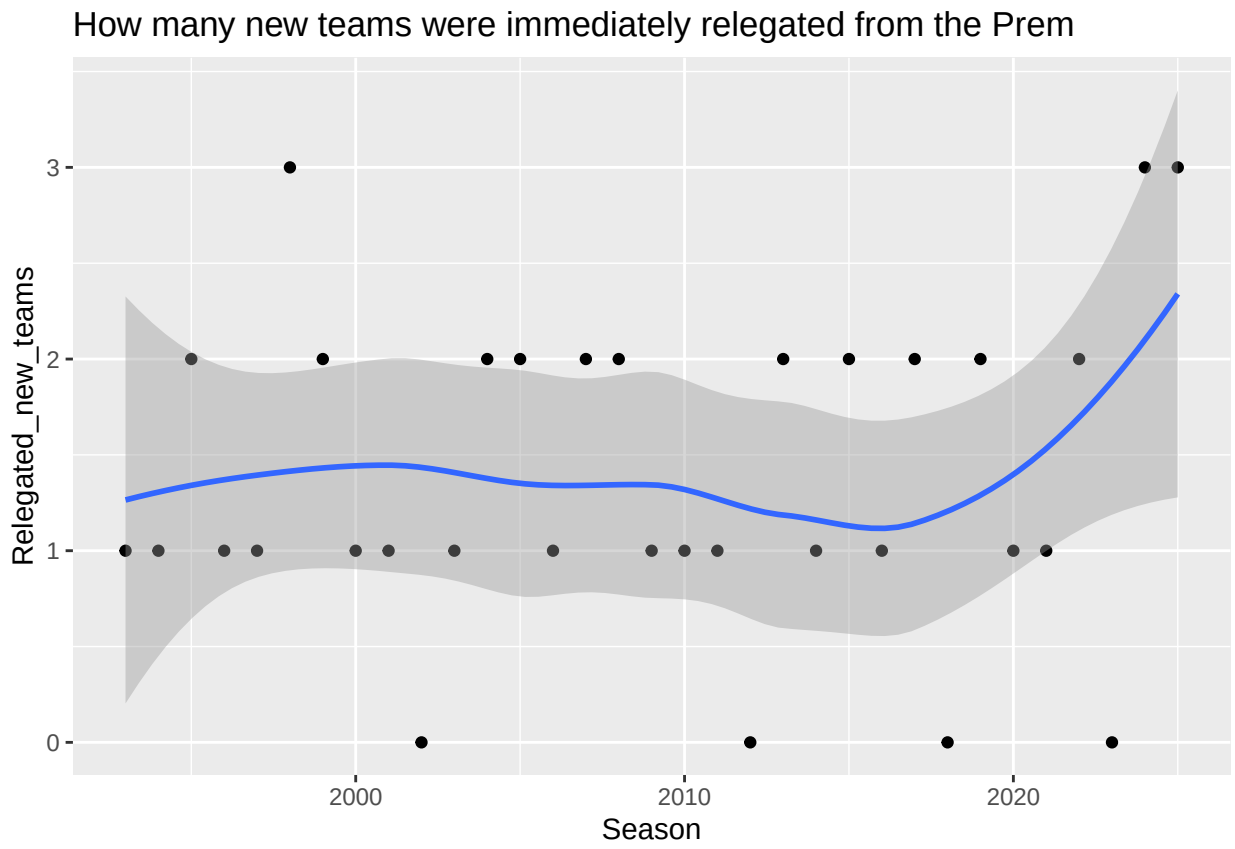
```
##
## Call:
## lm(formula = Relegated_new_teams ~ Season, data = relegation)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5156 -0.4113 -0.2810  0.6148  1.7017
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -16.06395   30.69592  -0.523    0.604
## Season        0.00869    0.01528   0.569    0.574
##
## Residual standard error: 0.8358 on 31 degrees of freedom
## Multiple R-squared:  0.01033,    Adjusted R-squared:  -0.0216
## F-statistic: 0.3235 on 1 and 31 DF,  p-value: 0.5736
```

The regression isn't significant. But it does go up so with the jaundiced eye of the TOTALLY IMPARTIAL saints fan we can say for definite that the prem's got harder over time.

But - there isn't much signal in that. So let's look at some averages, and also the points spread and relegation gap. We'll use ggplot2:

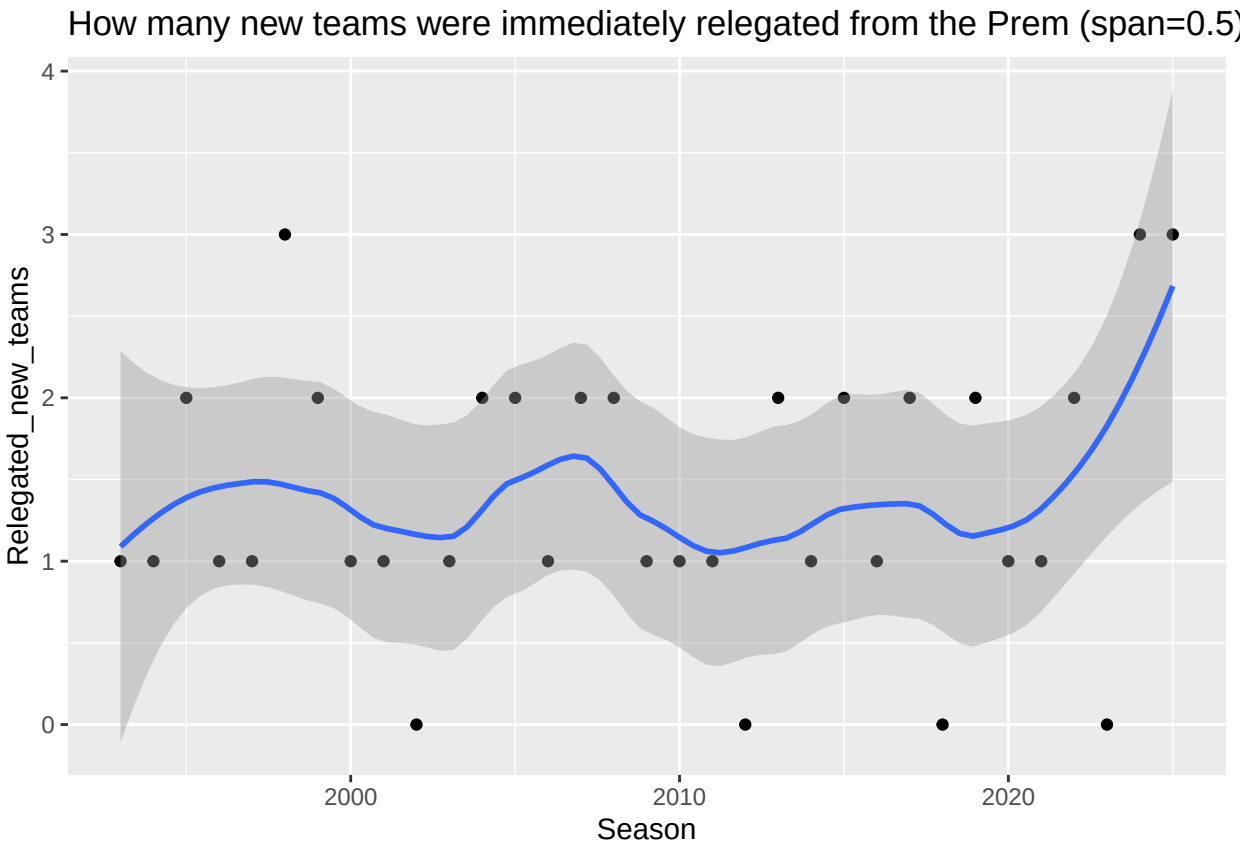
```
library(ggplot2)
ggplot(relegation, aes(y=Relegated_new_teams, x=Season)) + geom_point() + geom_smooth() + ggtitle(prem_
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



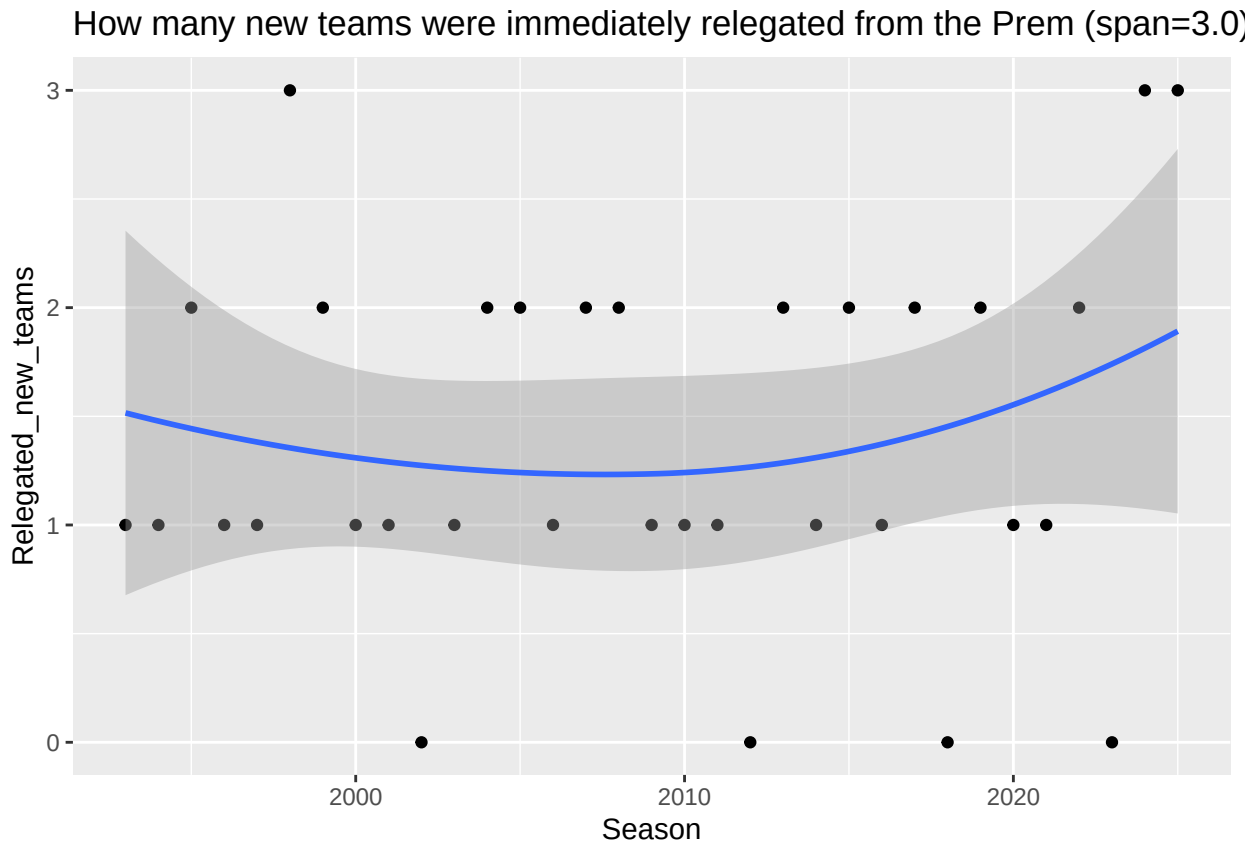

```
# play with the span width a bit to check param sensitivity
ggplot(relegation, aes(y=Relegated_new_teams, x=Season)) + geom_point() + geom_smooth(span=0.5) + ggtitl
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



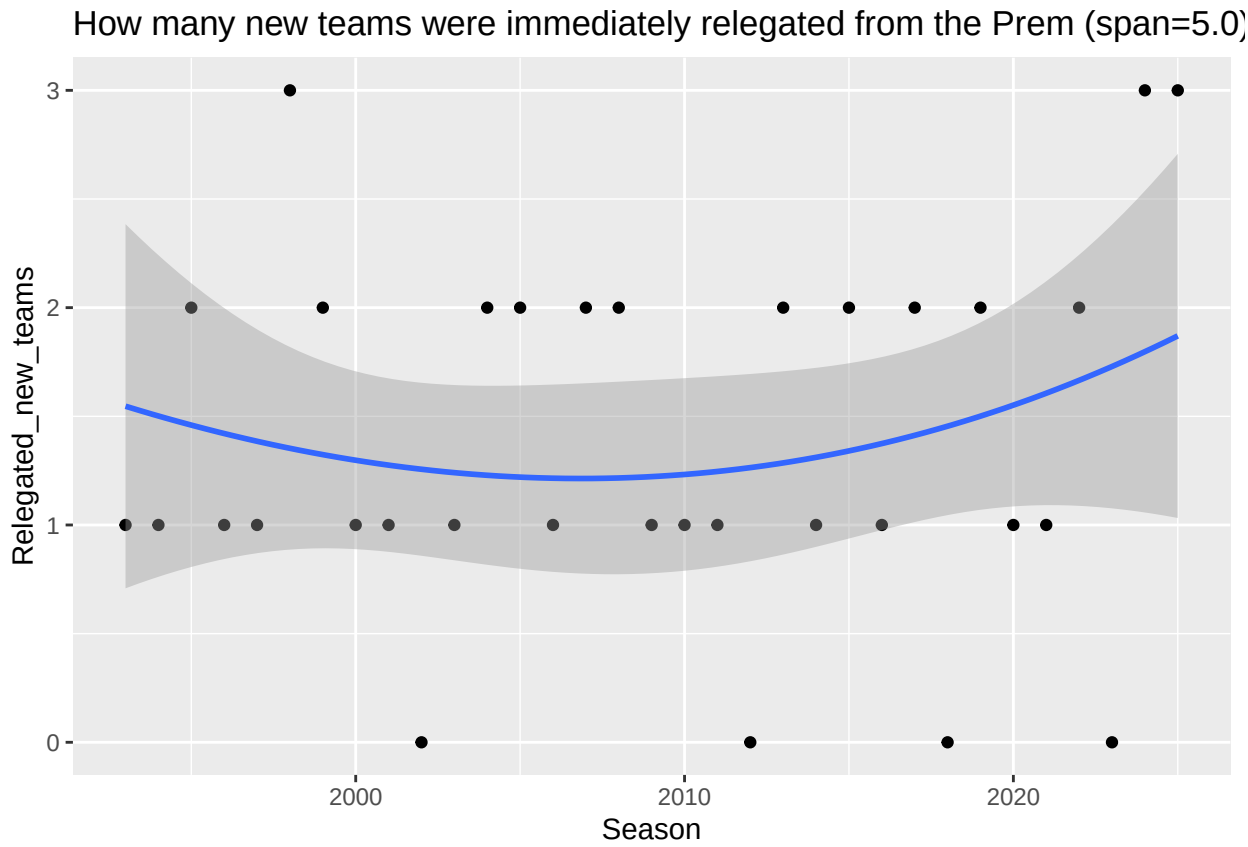

```
ggplot(relegation, aes(y=Relegated_new_teams, x=Season)) + geom_point() + geom_smooth(span=3) + ggtitle
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
ggplot(relegation, aes(y=Relegated_new_teams, x=Season)) + geom_point() + geom_smooth(span=5) + ggtitle
```

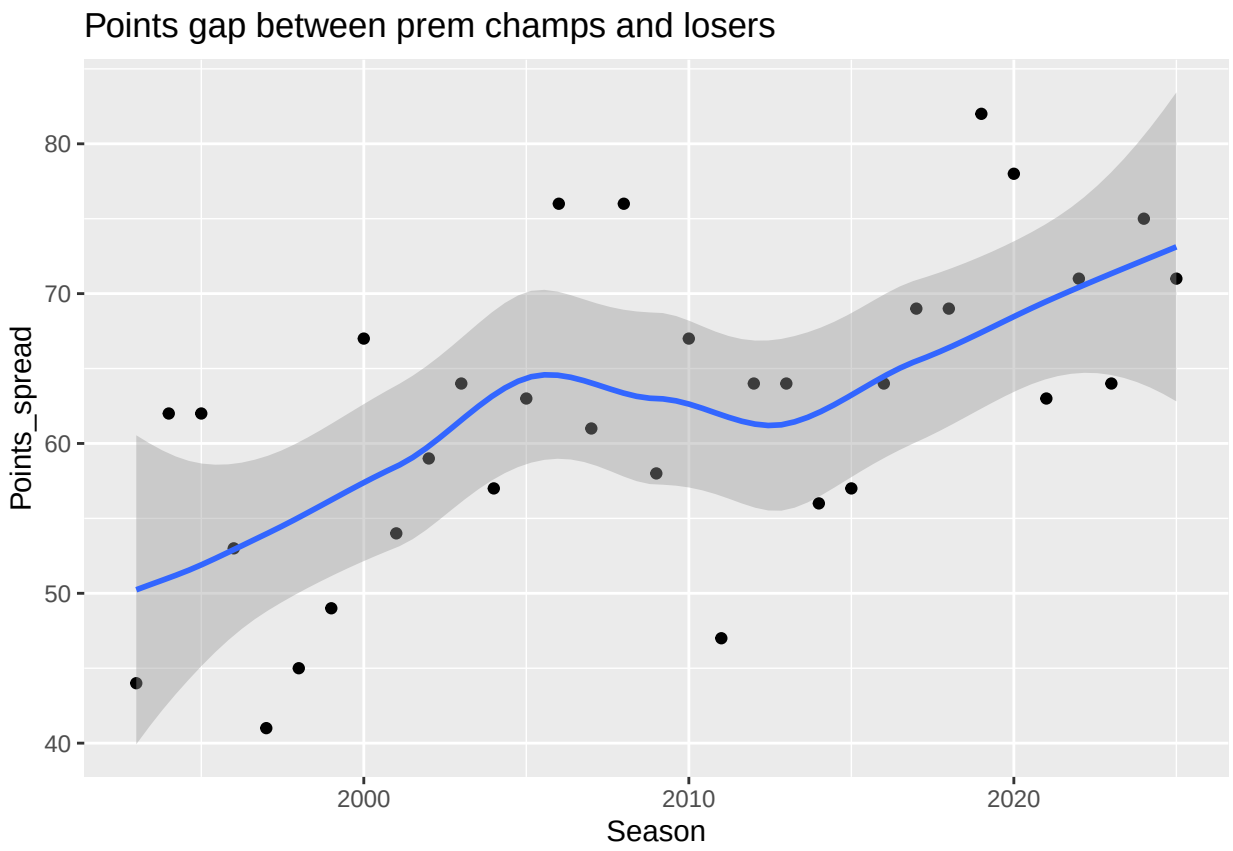
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Well, that's clearly INCREDIBLY convincing. Let's try the same with points spread and relegation gap:

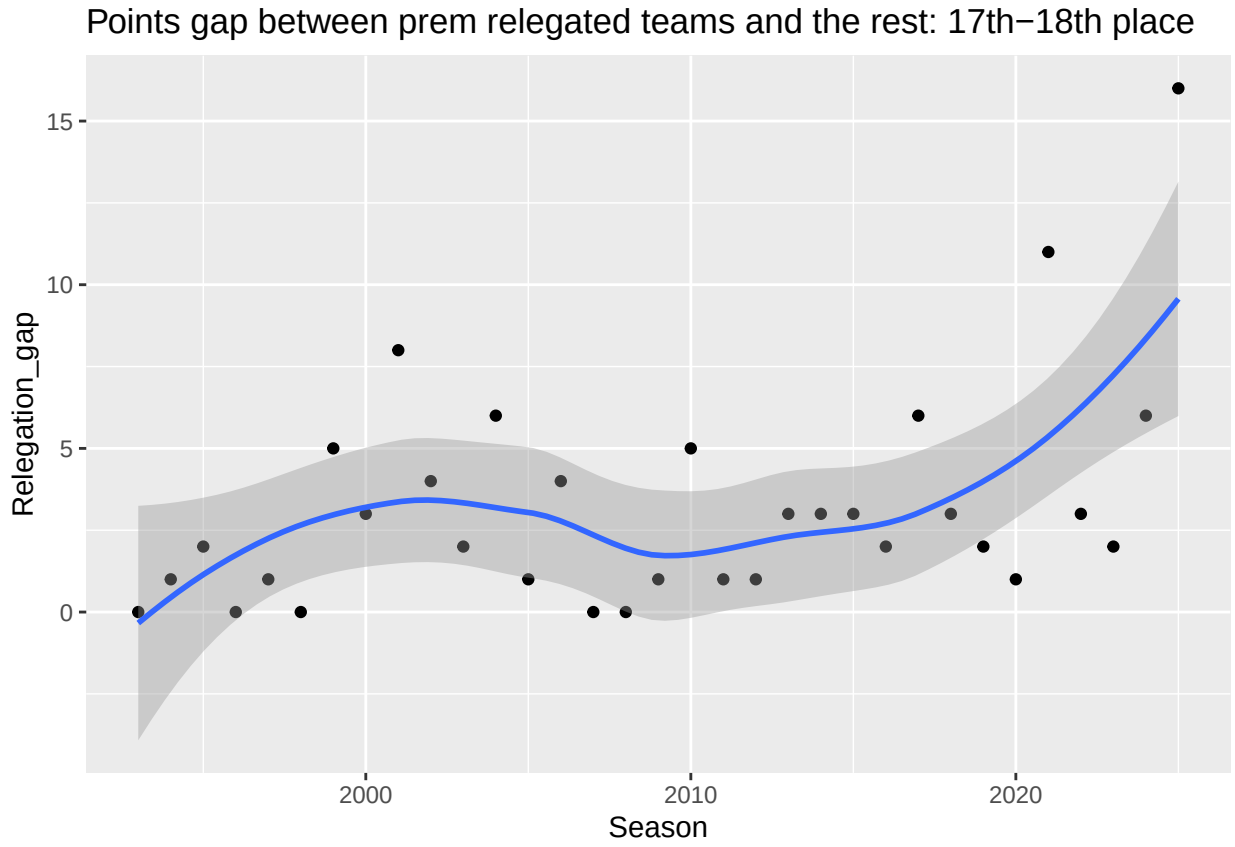
```
# points spread
ggplot(relegation, aes(y=Points_spread, x=Season)) + geom_point() + geom_smooth() + ggtitle("Points gap
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
# relegation gap
ggplot(relegation, aes(y=Relegation_gap, x=Season)) + geom_point() + geom_smooth() + ggtitle("Points ga
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



GLM modelling

OK, we're now in a position to do some proper modelling on this. Fit maximal model then drop terms:

```
max_lm = glm(Relegated_new_teams ~ Season + Points_spread + Relegation_gap, data=relegation)
summary(aov(max_lm))
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## Season          1  0.226  0.2259   0.323  0.574
## Points_spread   1  0.072  0.0717   0.102  0.751
## Relegation_gap  1  1.268  1.2678   1.810  0.189
## Residuals      29 20.313  0.7005
```

```
# drop Points spread
summary(aov( glm(Relegated_new_teams ~ Season  + Relegation_gap, data=relegation)))
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## Season          1  0.226  0.2259   0.332  0.569
## Relegation_gap  1  1.229  1.2289   1.805  0.189
## Residuals      30 20.424  0.6808
```

```
# drop season
summary(aov(glm(Relegated_new_teams ~  Relegation_gap, data=relegation)))
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## Relegation_gap  1  1.454  1.4541   2.207  0.147
## Residuals      31 20.425  0.6589
```

```
# Nah, still not significant. But - I suppose we're more interested in
# relegation gap (how close/exciting/fucking terrifying the run-in was really):
summary(aov( glm(Relegation_gap ~ Season  + Relegated_new_teams + Points_spread, data=relegation)))
```

```
##                     Df Sum Sq Mean Sq F value Pr(>F)
## Season               1  64.12   64.12   6.448 0.0167 *
## Relegated_new_teams  1  17.45   17.45   1.755 0.1957
## Points_spread        1   1.58    1.58   0.159 0.6931
## Residuals           29 288.37    9.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Significant! Drop points_spread
summary(aov( glm(Relegation_gap ~ Season  + Relegated_new_teams , data=relegation)))
```

```
##                     Df Sum Sq Mean Sq F value Pr(>F)
## Season               1  64.12   64.12   6.634 0.0152 *
## Relegated_new_teams  1  17.45   17.45   1.805 0.1892
## Residuals           30 289.95    9.66
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Drop relegated new teams
summary(aov( glm(Relegation_gap ~ Season  , data=relegation)))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Season       1  64.12   64.12   6.466 0.0162 *
## Residuals   31 307.40    9.92
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

So, most significant model has `relegation gap` between 17th and 18th placed-teams (league stich-up-ness) explained by `Season` with $P =< 0.0162$.

Conclusion

It's ALL ABOUT THE PETROSTATES and the dollar jockeys. Money is increasingly wrecking football. Saints are the best and the last defenders of decent proper football against billionaire tourists. QED.

Ahem.