R Notebook

```
source("scriptDavid.R")
```

```
## -- Attaching packages ---------------------------------------------------------------------------------------

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

## Warning: package 'stringr' was built under R version 4.0.3

## Warning: package 'forcats' was built under R version 4.0.3

## -- Conflicts ------------------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

## Warning: package 'readxl' was built under R version 4.0.3

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
gbsoccer <- read_excel("data/partcipation_statistics_12_01_202009_22.xlsx")
```

```
gbsoccer1 <- gbsoccer %>%
  group_by(Year) %>%
  summarise(Boys_School = sum(`Boys School`),
            Girls_School = sum(`Girls School`)) %>%
  mutate(Boys_change = (Boys_School/lag(Boys_School) - 1) * 100) %>%
  mutate(Girls_change = (Girls_School/lag(Girls_School) - 1) * 100) %>%
  na.omit()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
head(gbsoccer1, 5)
```

Table 1: US High Schools with Soccer Programs

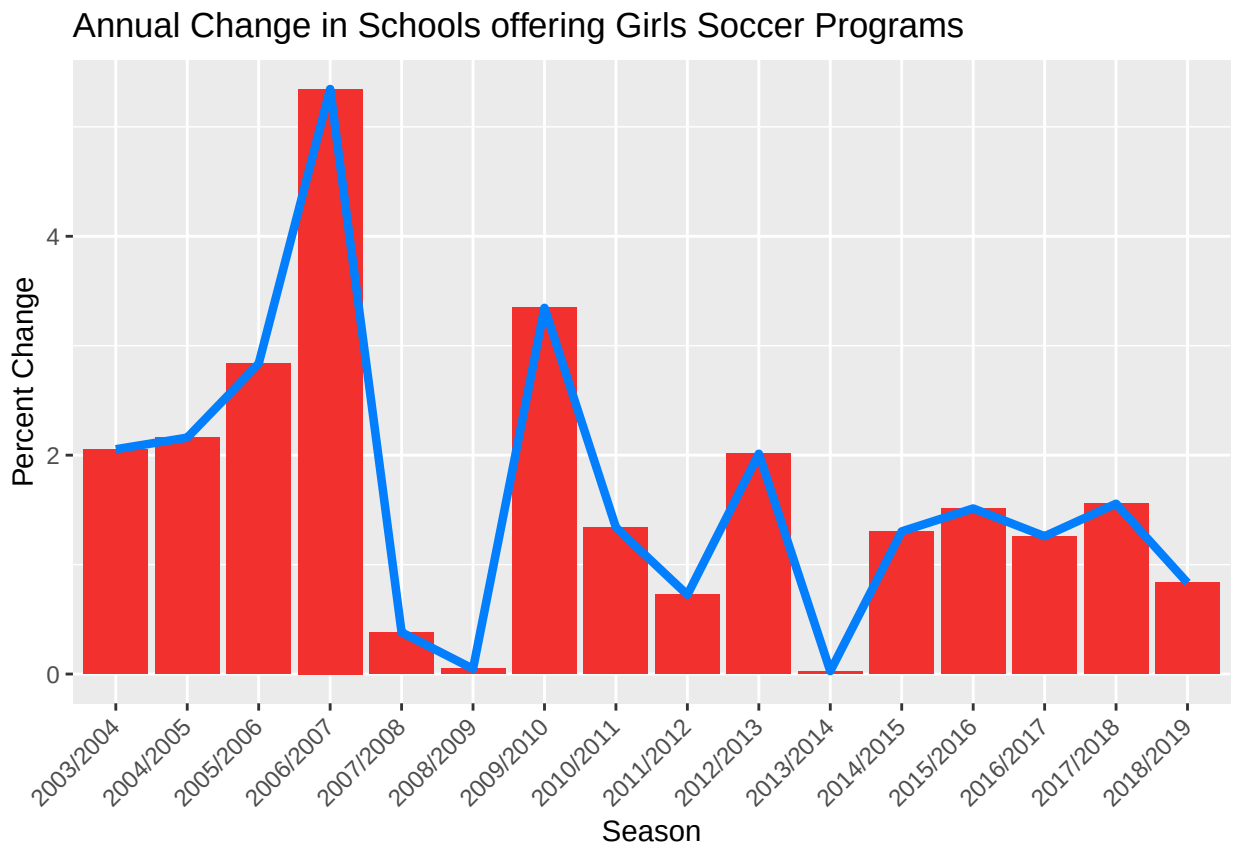
Year	Boys_School	Girls_School	Boys_change	Girls_change
2003/2004	10219	9490	1.1481738	2.0539843
2004/2005	10392	9695	1.6929249	2.1601686
2005/2006	10580	9970	1.8090839	2.8365137
2006/2007	11066	10503	4.5935728	5.3460381
2007/2008	11122	10543	0.5060546	0.3808436
2008/2009	11139	10548	0.1528502	0.0474248
2009/2010	11375	10901	2.1186821	3.3466060
2010/2011	11503	11047	1.1252747	1.3393267
2011/2012	11600	11127	0.8432583	0.7241785
2012/2013	11626	11351	0.2241379	2.0131212
2013/2014	11718	11354	0.7913298	0.0264294
2014/2015	11838	11502	1.0240655	1.3035054
2015/2016	12054	11676	1.8246325	1.5127804
2016/2017	12188	11823	1.1116642	1.2589928
2017/2018	12393	12007	1.6819823	1.5562886
2018/2019	22818	12107	84.1200678	0.8328475

```
## # A tibble: 5 x 5
##   Year      Boys_School Girls_School Boys_change Girls_change
##   <chr>           <dbl>        <dbl>       <dbl>        <dbl>
## 1 2003/2004       10219         9490       1.15         2.05
## 2 2004/2005       10392         9695       1.69         2.16
## 3 2005/2006       10580         9970       1.81         2.84
## 4 2006/2007       11066        10503       4.59         5.35
## 5 2007/2008       11122        10543       0.506        0.381
```

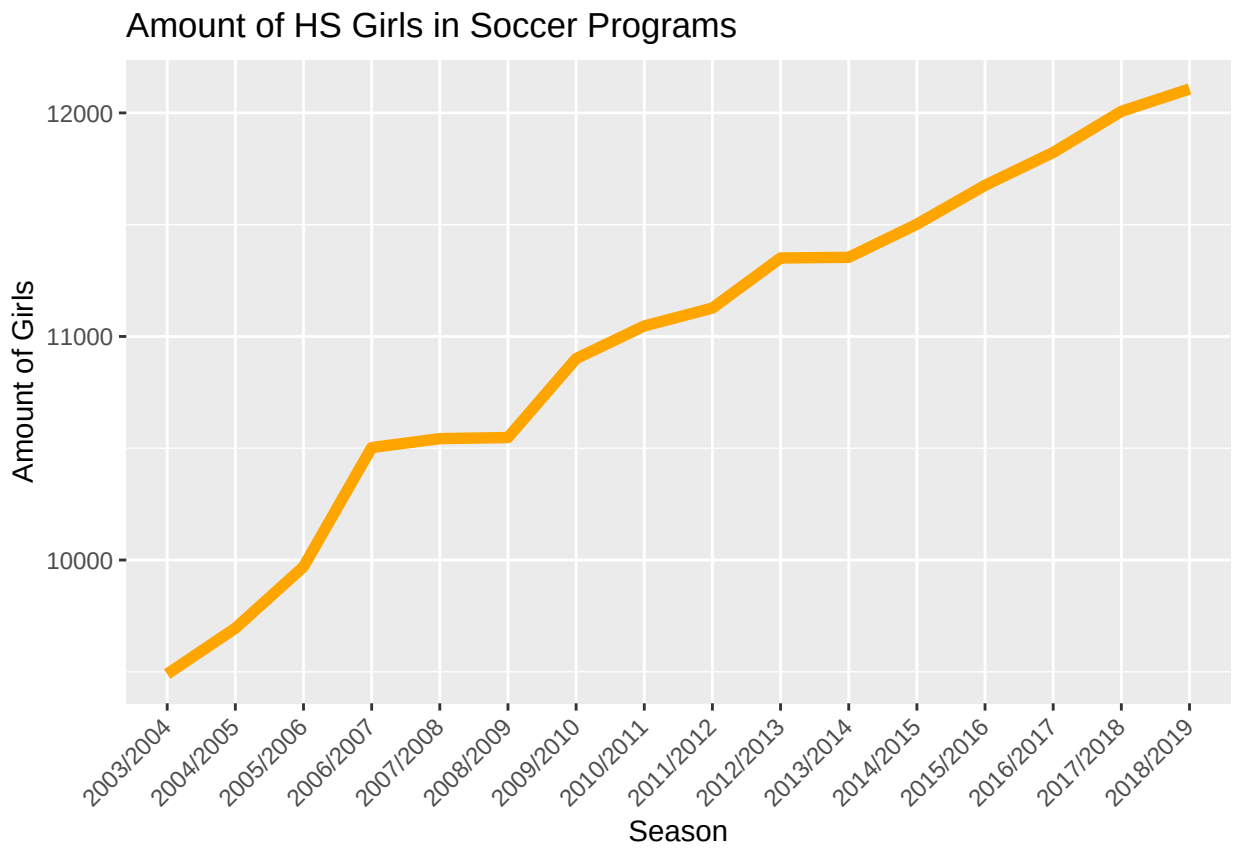
```
gbsoccer1a <- gbsoccer1 %>%
  mutate(Girls_School = (Girls_School/1000))
```

```
gbsoccer1 %>%
  kbl(caption = "US High Schools with Soccer Programs") %>%
  kable_classic(position = "center")
```

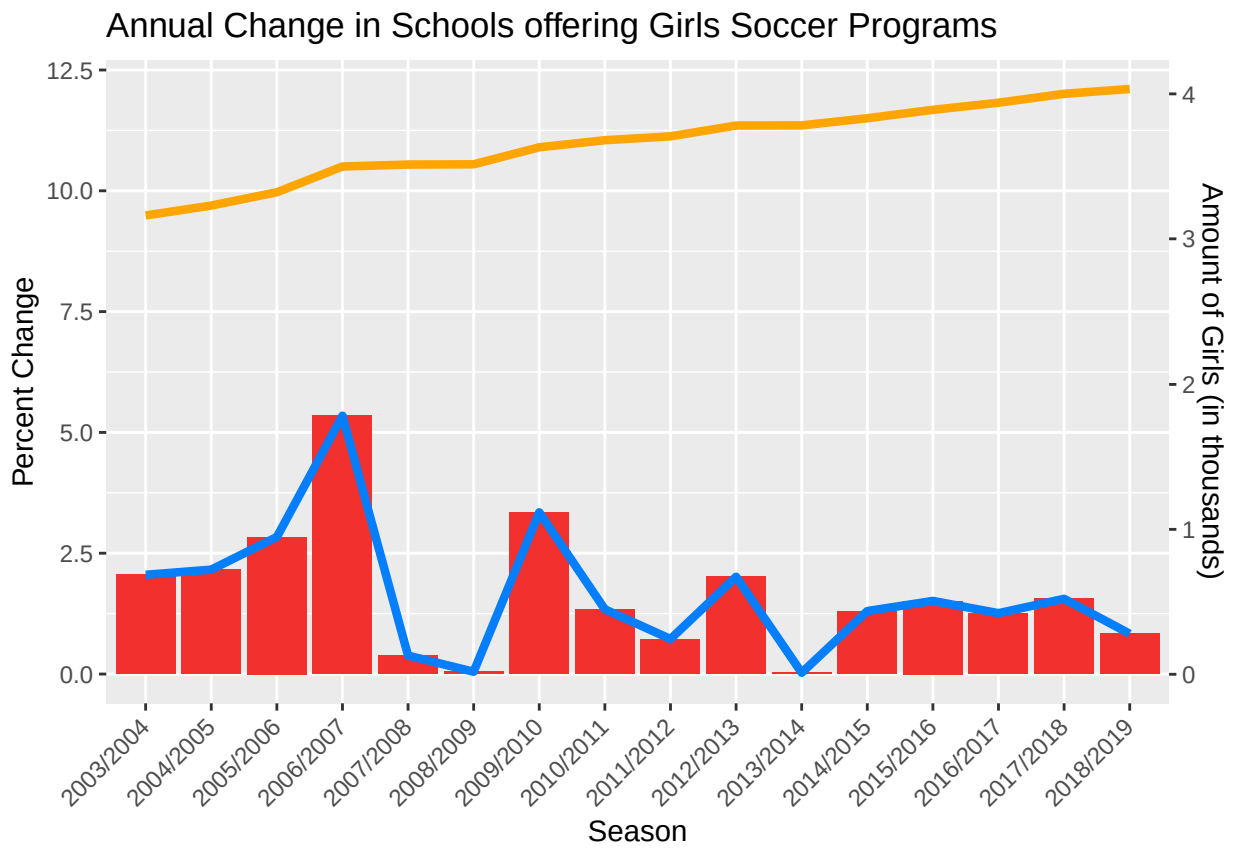
```
ggplot(gbsoccer1) +
  geom_col(aes(x = Year, y = Girls_change), fill = "#F2312E") +
  guides(fill = FALSE) +
  geom_line(aes(x = Year, y = Girls_change, group = 1), color= "#037EFC", size = 1.5) +
    theme(axis.text.x = element_text(angle = 45, vjust = (1.1), hjust = (1.1))) +
  labs(title = "Annual Change in Schools offering Girls Soccer Programs", x = "Season", y = "Percent Cha
```



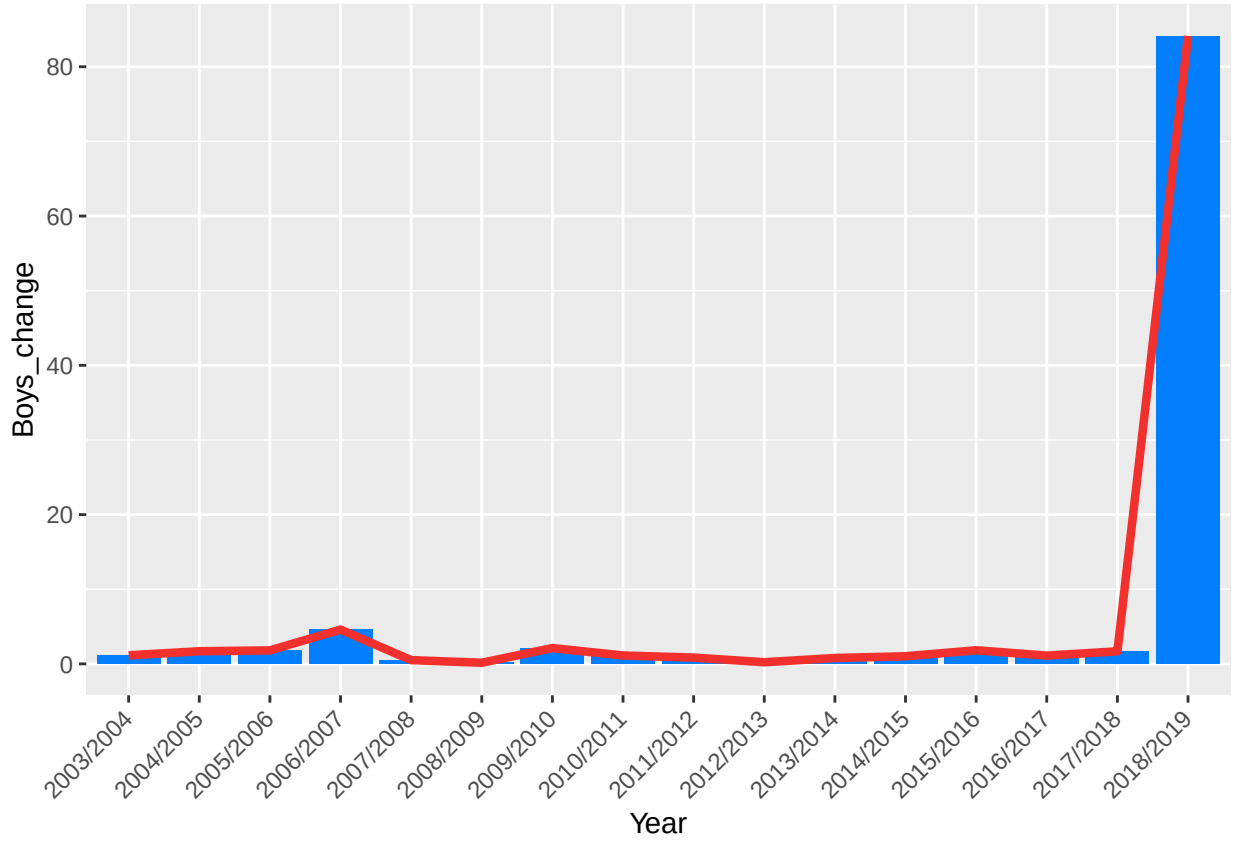

```
ggplot(gbsoccer1) +
  geom_line(aes(x = Year, y = Girls_School, group = 1), color = "orange", size = 2) +
  theme(axis.text.x = element_text(angle = 45, vjust = (1.1), hjust = (1.1)))+
  labs(title = "Amount of HS Girls in Soccer Programs", x = "Season", y = "Amount of Girls")
```




```
ggplot(gbsoccer1a) +
  geom_col(aes(x = Year, y = Girls_change), fill = "#F2312E") +
  guides(fill=FALSE) +
  geom_line(aes(x = Year, y = Girls_change, group = 1), color= "#037EFC", size = 1.5) +
    theme(axis.text.x = element_text(angle = 45, vjust = (1.1), hjust = (1.1))) +
  labs(title = "Annual Change in Schools offering Girls Soccer Programs", x = "Season", y = "Percent Cha
  geom_line(aes(x = Year, y = Girls_School, group = 1), color = "orange", size = 1.5) +
    scale_y_continuous(sec.axis = sec_axis(~./3, name = "Amount of Girls (in thousands)"))
```



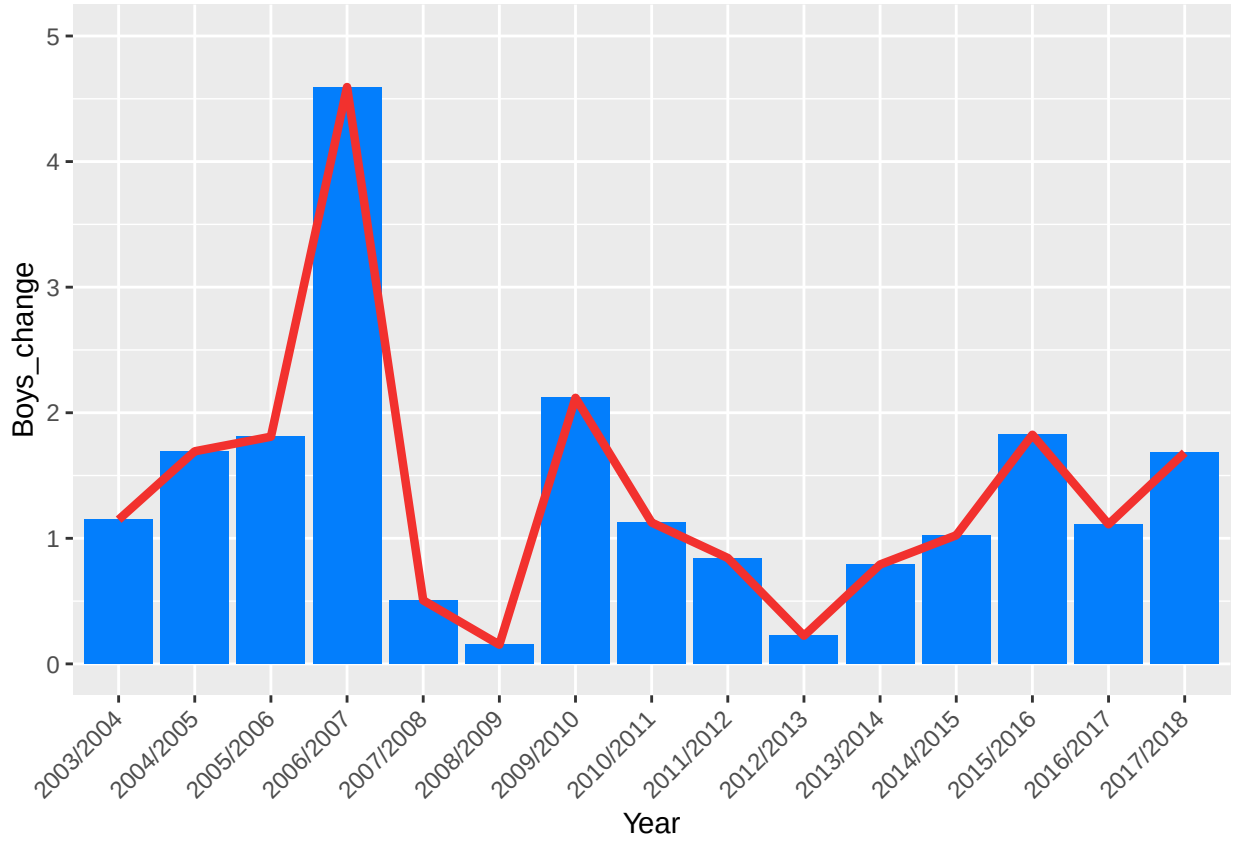

```
ggplot(gbsoccer1) +
  geom_col(aes(x = Year, y = Boys_change), fill = "#037EFC") +
  guides(fill = FALSE) +
  geom_line(aes(x = Year, y = Boys_change, group = 1), color= "#F2312E", size= 1.5) +
    theme(axis.text.x = element_text(angle = 45, vjust = (1.1), hjust = (1.1)))
```



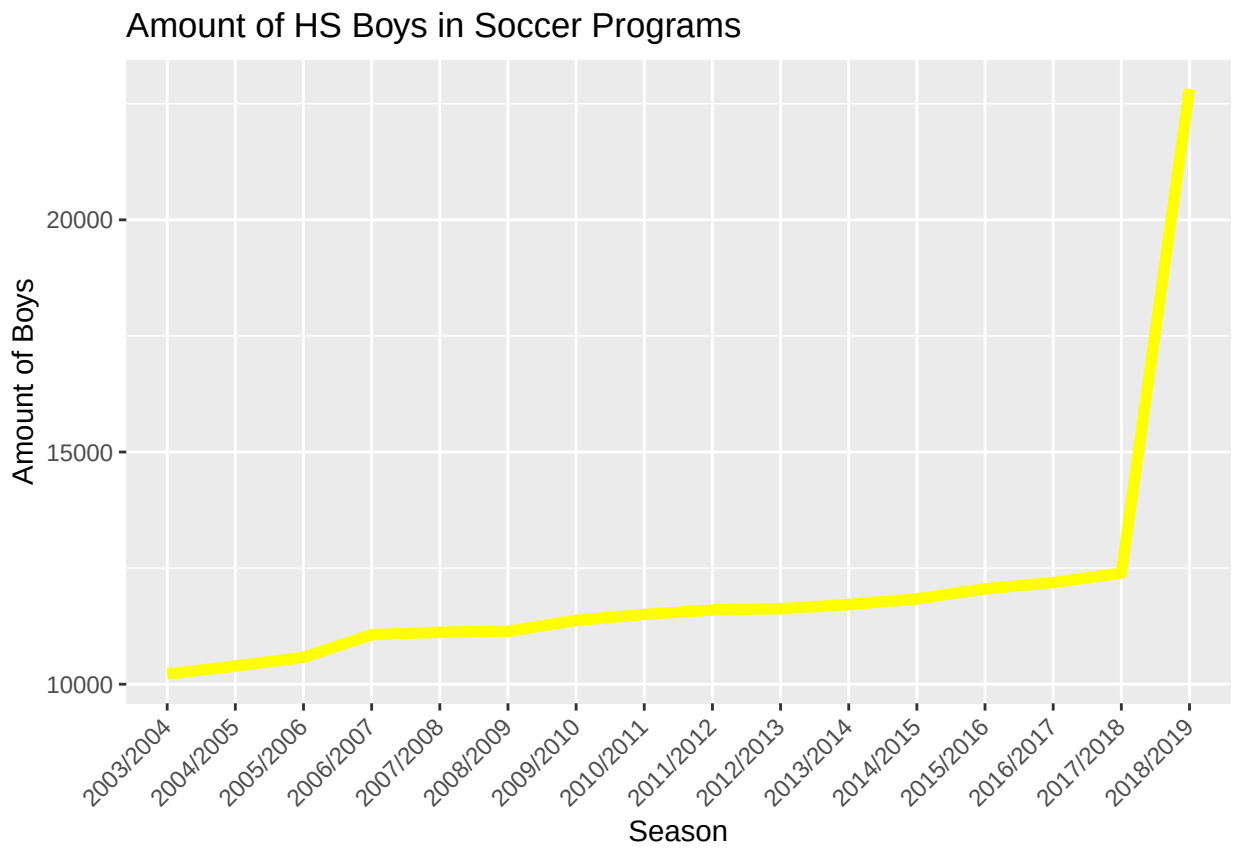

```
ggplot(gbsoccer1) +
  geom_col(aes(x = Year, y = Boys_change), fill = "#037EFC") +
  guides(fill = FALSE) +
  ylim(0, 5) +
  xlim("2003/2004", "2004/2005", "2005/2006",
       "2006/2007","2007/2008", "2008/2009", "2009/2010", "2010/2011", "2011/2012", "2012/2013", "2013/
      "2014/2015","2015/2016", "2016/2017", "2017/2018") +
  geom_line(aes(x = Year, y = Boys_change, group = 1), color= "#F2312E", size= 1.5) +
    theme(axis.text.x = element_text(angle = 45, vjust = (1.1), hjust = (1.1)))
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```




```
ggplot(gbsoccer1) +
  geom_line(aes(x = Year, y = Boys_School, group = 1), color = "yellow", size = 2) +
  theme(axis.text.x = element_text(angle = 45, vjust = (1.1), hjust = (1.1)))+
  labs(title = "Amount of HS Boys in Soccer Programs", x = "Season", y = "Amount of Boys")
```


Amount of HS Boys in Soccer Programs
Amount of Boys
20000
15000
10000
2003/2004
2004/2005
2005/2006
2006/2007
2007/2008
2008/2009
2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
2016/2017
2017/2018
2018/2019
Season