NFL career length

Ryan Menezes

July 31, 2018

```
setwd('/home/ryan/sports/nfl-life-death/')
```

```
library('tidyverse')
library('zoo')
```

Data import and cleaning

Bring in all the player data from pro-football-reference.com

```
careers <- read_csv('data/nfl-players.csv')
```

```
## Parsed with column specification:
## cols(
##   name = col_character(),
##   srid = col_character(),
##   active = col_character(),
##   hof = col_character(),
##   positions = col_character(),
##   start = col_integer(),
##   end = col_integer(),
##   letter = col_character()
## )
```

```
head(careers)
```

```
## # A tibble: 6 x 8
##   name            srid      active hof   positions      start   end letter
##   <chr>           <chr>     <chr>  <chr> <chr>          <int> <int> <chr>
## 1 Isaako Aaitui   AaitIs00  False  False NT              2013  2014 A
## 2 Joe Abbey       AbbeJo20  False  False E               1948  1949 A
## 3 Fay Abbott      AbboFa20  False  False BB-FB-TB-QB-W~  1921  1929 A
## 4 Vince Abbott    abbotvin~ False  False K               1987  1988 A
## 5 Jared Abbreder~ AbbrJa00  True   False WR              2014  2017 A
## 6 Duke Abbruzzi   AbbrDu20  False  False HB-DB           1946  1946 A
```

Unfurl the data into separate rows for each position listed

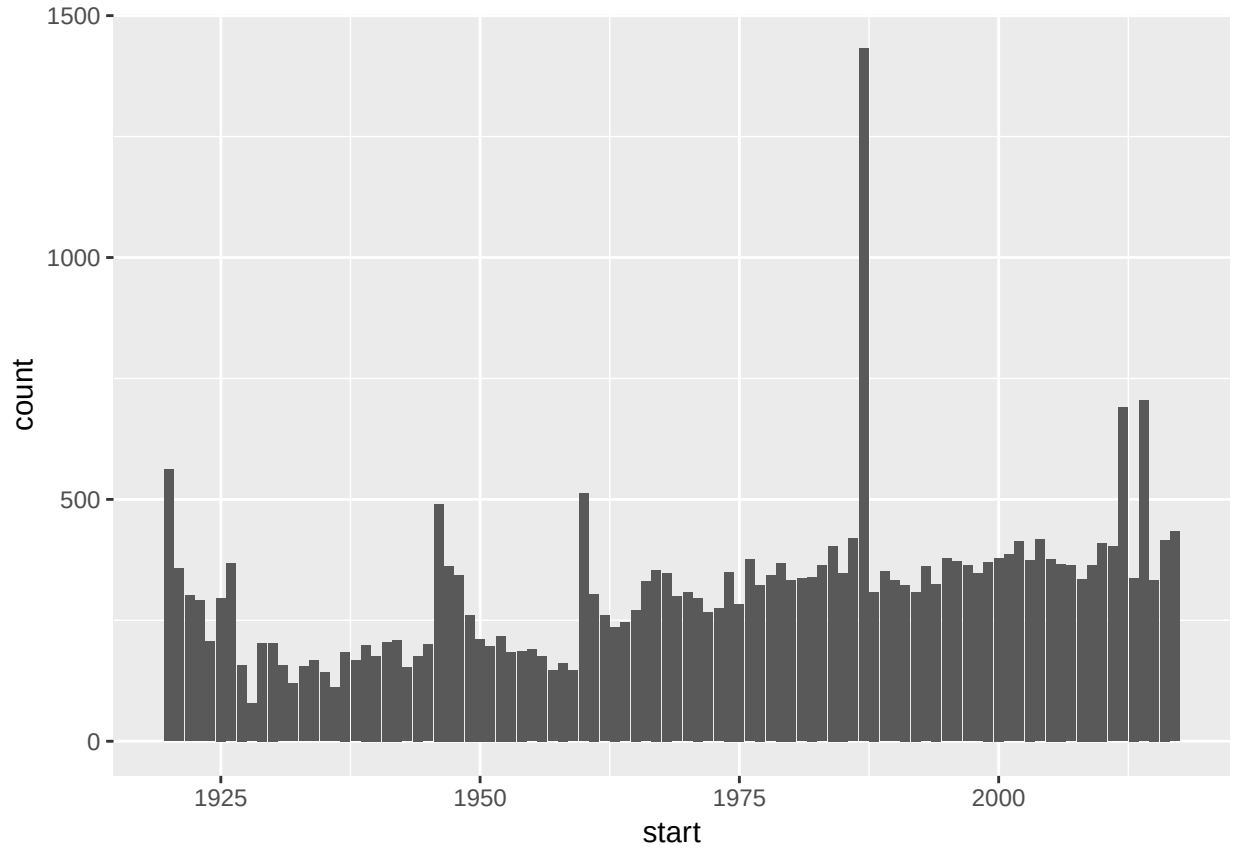
```
allpositions <- careers %>%
  select(name, srid, active, positions, start, end) %>%
  replace_na(list(positions = '')) %>%
  mutate(positions = str_replace(positions, ',|/', '-')) %>%
  separate_rows(positions, sep = '-')
head(allpositions)
```

```
## # A tibble: 6 x 6
##   name          srid     active positions start   end
##   <chr>         <chr>    <chr>  <chr>     <int> <int>
## 1 Isaako Aaitui AaitIs00 False  NT         2013  2014
```

```
## 2 Joe Abbey     AbbeJo20 False  E          1948  1949
## 3 Fay Abbott    AbboFa20 False  BB         1921  1929
## 4 Fay Abbott    AbboFa20 False  FB         1921  1929
## 5 Fay Abbott    AbboFa20 False  TB         1921  1929
## 6 Fay Abbott    AbboFa20 False  QB         1921  1929
```

What positions exist over time?

```
ggplot(data = allpositions) +
  geom_bar(mapping = aes(x = start))
```



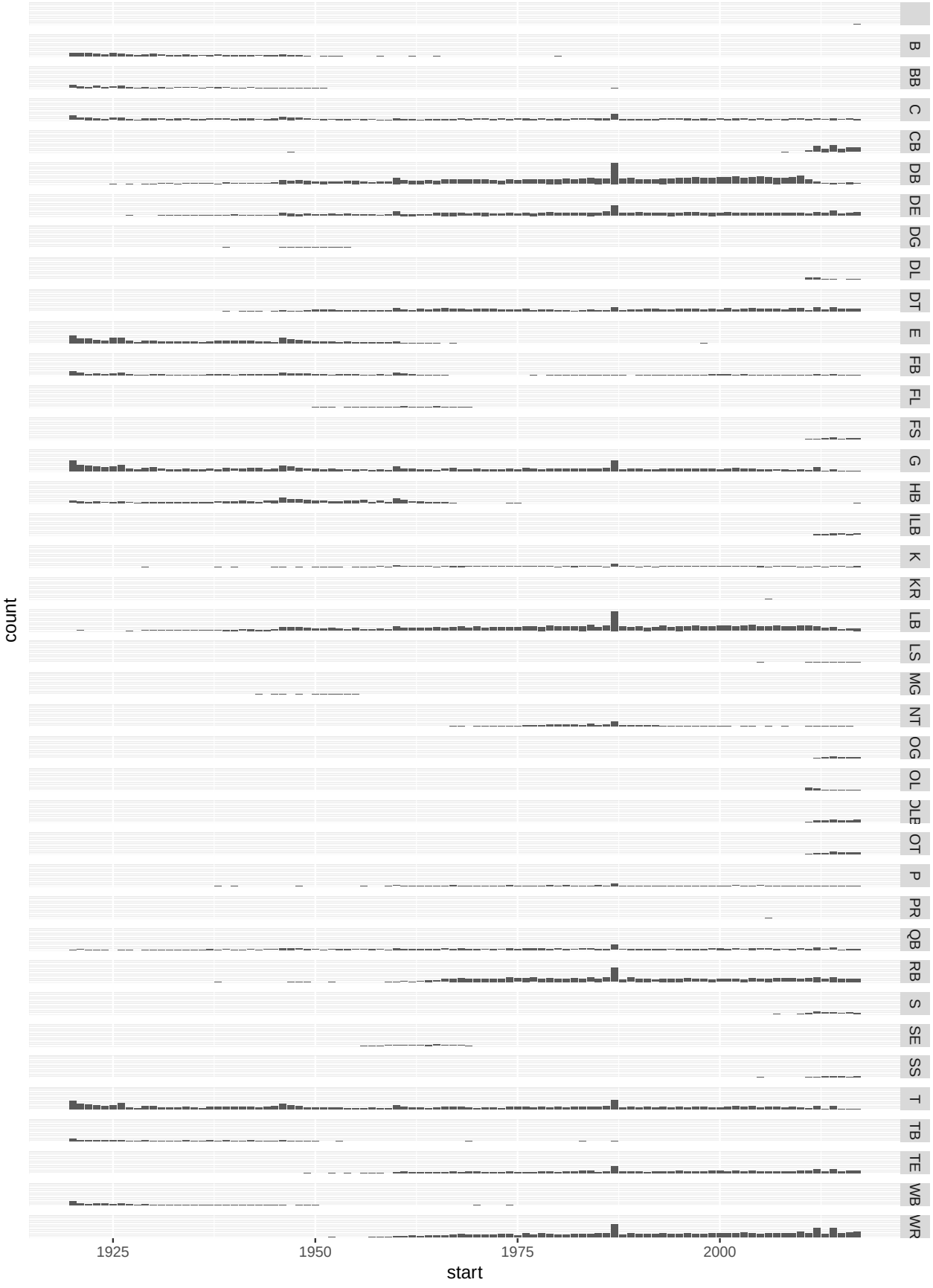

```
allpositions %>%
  group_by(start) %>%
  count() %>%
  arrange(desc(n)) %>%
  head()
```

```
## # A tibble: 6 x 2
## # Groups:   start [6]
##   start     n
##   <int> <int>
## 1  1987  1432
## 2  2014   704
## 3  2012   690
## 4  1920   562
## 5  1960   512
## 6  1946   490
```

Why the bump in 1987? Strike season!

What positions exist at what times?

```
ggplot(data = allpositions) +
  geom_bar(mapping = aes(x = start)) +
  facet_grid(positions ~ .) +
  theme(axis.text.y = element_blank(),
        axis.ticks.y = element_blank())
```


count
start
1925
1950
1975
2000
B
BB
C
CB
DB
DE
DG
DL
DT
E
FB
FL
FS
G
HB
ILB
K
KR
LB
LS
MG
NT
OG
OL
OLB
OT
P
PR
QB
RB
S
SE
SS
T
TB
TE
WB
WR

There are a lot of "backs" (wing backs, blocking backs, etc.) in earlier years that are probably not worth including now.

PFR links: positions glossary

Spit out the positions so I can classify them

```
posfreq <- allpositions %>%
  count(positions)
write_csv(posfreq, 'data/positions.csv')
```

Bring in the "simple" positions

```
simplepos <- read_csv('data/positions-categorized.csv')
```

```
## Parsed with column specification:
## cols(
##   positions = col_character(),
##   n = col_integer(),
##   category = col_character()
## )
```

```
knitr::kable(simplepos %>% arrange(category, desc(n)))
```

positions	n	category
DB	3558	defback
CB	314	defback
S	85	defback
FS	69	defback
SS	62	defback
DE	2098	defline
DT	1668	defline
DL	50	defline
DG	21	defline
LB	3096	linebacker
OLB	146	linebacker
ILB	75	linebacker
MG	25	linebacker
G	2845	offline
T	2712	offline
C	1392	offline
NT	359	offline
OT	116	offline
OG	71	offline
OL	61	offline
QB	1019	quarterback
RB	2077	runback
HB	939	runback
FB	816	runback
TB	278	runback
TE	1258	tightend
WR	2399	widerec
E	1175	NA
B	551	NA
K	380	NA
P	331	NA

positions	n	category
WB	281	NA
BB	267	NA
FL	91	NA
SE	68	NA
LS	26	NA
KR	1	NA
PR	1	NA
NA	1	NA

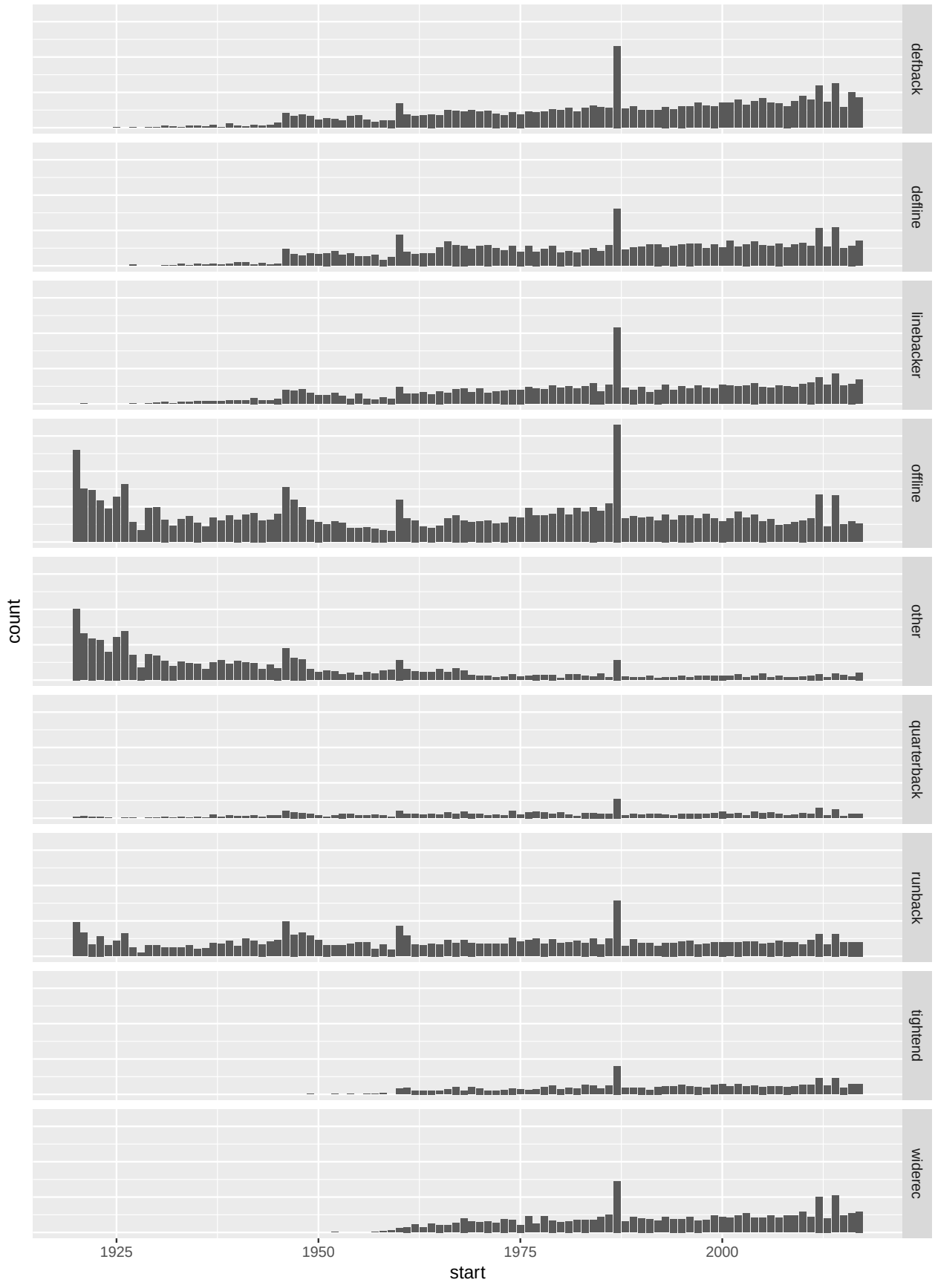
Join the simple positions back to the data

```
allpositions <- allpositions %>%
  left_join(simplepos %>% filter(!is.na(category)) %>% select(positions, category)) %>%
  replace_na(list(category = 'other'))
```

```
## Joining, by = "positions"
```

Now chart positions over time

```
ggplot(data = allpositions) +
  geom_bar(mapping = aes(x = start)) +
  facet_grid(category ~ .) +
  theme(axis.text.y = element_blank(),
        axis.ticks.y = element_blank())
```


defback
defline
linebacker
offline
other
quarterback
runback
tightend
widerec
count
1925
1950
1975
2000
start

Very little of "other" in recent years

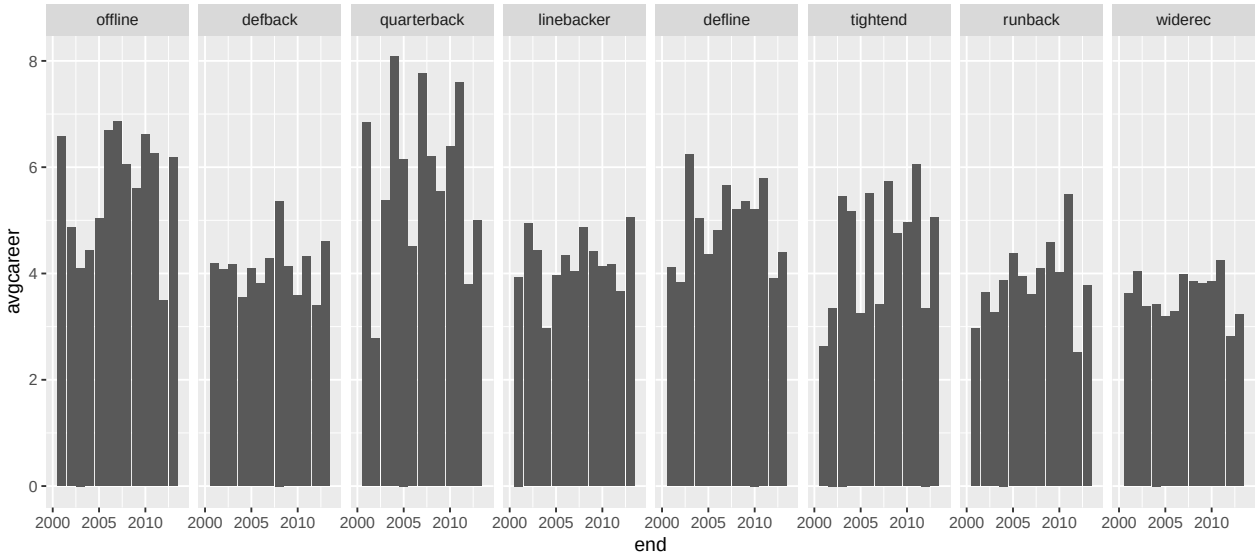
Let's filter down to "retired" players (where active is False, according to PFR)

```
retired <- allpositions %>%
  filter(active == "False" & category != 'other') %>%
  mutate(career = (end - start))
```

Now calculate raw average for each position in each year and plot it

```
# lining it up to match WSJ plot
retired$category <- factor(
  x = retired$category,
  levels = c('offline','defback','quarterback','linebacker','defline','tightend','runback','widerec')
)

careers.by.pos <- retired %>%
  group_by(category, end) %>%
  summarise(avgcareer = mean(career))

ggplot(data = careers.by.pos) +
  geom_bar(mapping = aes(x = end, y = avgcareer), stat = 'identity') +
  facet_grid(. ~ category) +
  scale_x_continuous(limits = c(2000, 2014))
```

```
## Warning: Removed 541 rows containing missing values (position_stack).
```

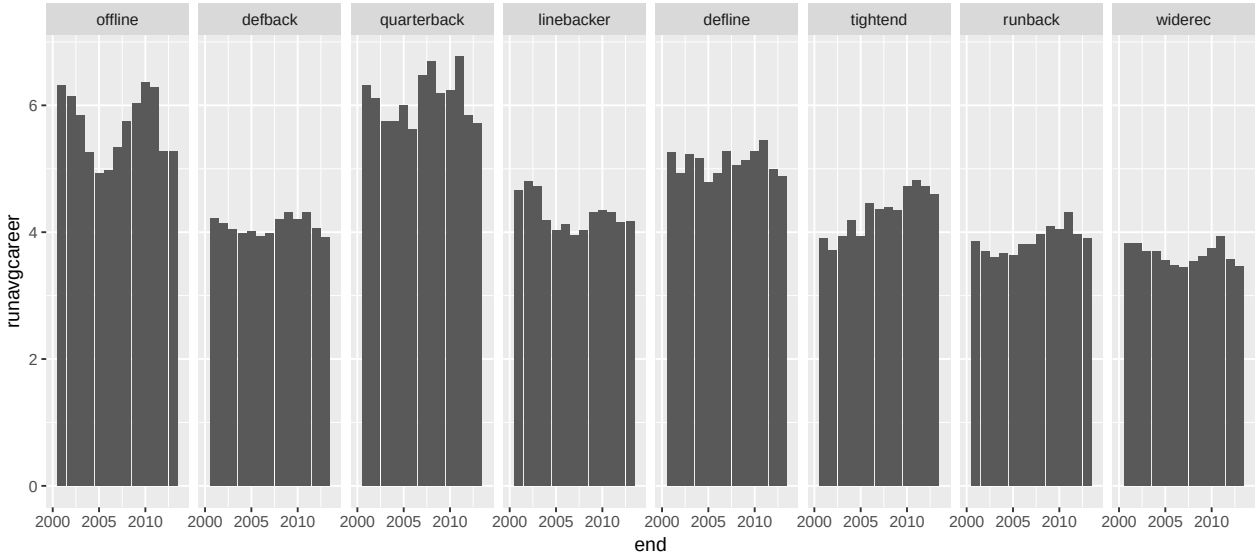


There is some noise. Will a moving average smooth it out?

```
window <- 5 # years

run.careers.by.pos <- retired %>%
  group_by(category, end) %>%
  summarise(careers = sum(career),
            players = n()) %>%
  arrange(category, end) %>%
  group_by(category) %>%
  mutate(sumcareers = rollsumr(careers, k = window, na.pad = TRUE),
```

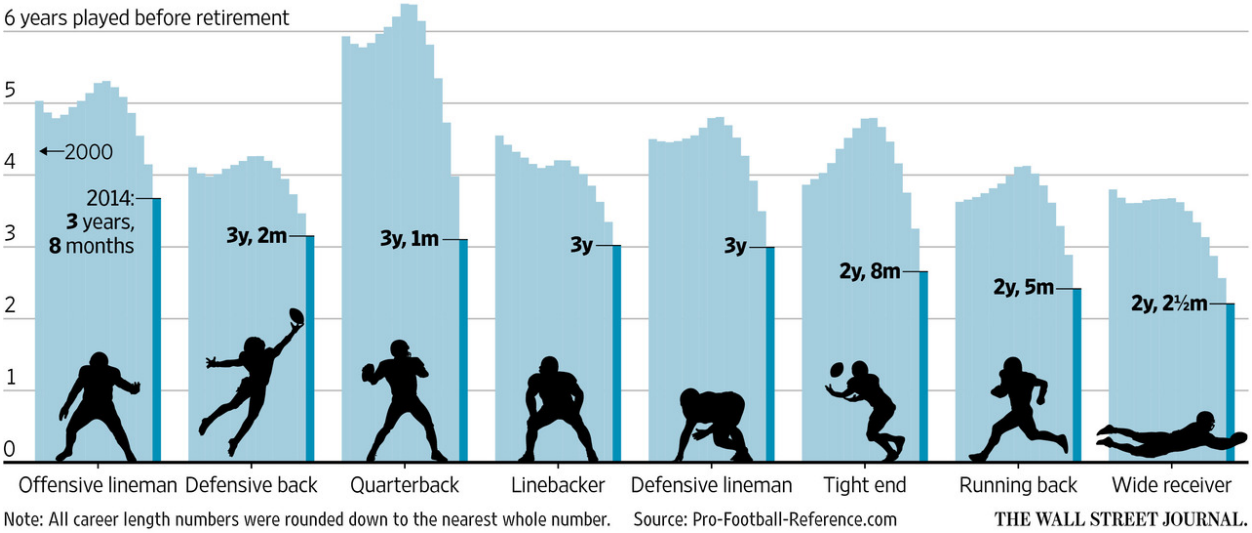
```
        sumplayers = rollsumr(players, k = window, na.pad = TRUE),
        runavgcareer = sumcareers / sumplayers)

ggplot(data = run.careers.by.pos) +
  geom_bar(mapping = aes(x = end, y = runavgcareer), stat = 'identity') +
  facet_grid(. ~ category) +
  scale_x_continuous(limits = c(2000, 2014))
```

```
## Warning: Removed 541 rows containing missing values (position_stack).
```



Time in the NFL

The average length of an NFL career has seen a sharp drop since 2000. A look at the average number of years played before retirement, sorted by position.



Note: All career length numbers were rounded down to the nearest whole number. Source: Pro-Football-Reference.com THE WALL STREET JOURNAL.