

Compile NFL data

Ryan Menezes

July 30, 2018

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

library('tidyverse')

players <- 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(players)

## # 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

dates <- read_csv('data/nfl-dob-dod.csv', col_types = 'ccc')
dates$dod <- str_replace(dates$dod, '-00$', '-01')
dates <- type_convert(dates)

## Parsed with column specification:
## cols(
##   srid = col_character(),
##   dob = col_date(format = ""),
##   dod = col_date(format = "")
## )

head(dates)

## # A tibble: 6 x 3
##   srid    dob      dod
##   <chr>  <date>   <date>
## 1 JaneCl20 1911-04-01 1990-01-16
## 2 JohnHe21 1928-07-10 NA
## 3 Henda.00 1993-04-08 NA
## 4 FrizWi20 1962-09-08 NA
```

```
## 5 HarrNi00 1994-12-07 NA
## 6 BollDo20 1927-07-16 2001-12-29
```

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

```
## Parsed with column specification:
## cols(
##   srid = col_character(),
##   team = col_character(),
##   seasons = col_integer(),
##   games = col_integer()
## )
```

```
head(careers)
```

```
## # A tibble: 6 x 4
##   srid      team      seasons games
##   <chr>    <chr>      <int> <int>
## 1 AaitIs00 Washington Redskins      1      0
## 2 AaitIs00 Miami Dolphins      1      4
## 3 AbbeJo20 Chicago Bears      1     12
## 4 AbbeJo20 New York Bulldogs      1      6
## 5 AbboFa20 Dayton Triangles      9     57
## 6 abbotvin01 San Diego Chargers      2     23
```

```
gmsns <- careers %>%
  group_by(srid) %>%
  summarise(
    seasons = sum(seasons),
    games = sum(games)
  )
head(gmsns)
```

```
## # A tibble: 6 x 3
##   srid      seasons games
##   <chr>      <int> <int>
## 1 AaitIs00      2      4
## 2 AbbeJo20      2     18
## 3 AbboFa20      9     57
## 4 abbotvin01      2     23
## 5 AbbrDu20      1      3
## 6 AbbrJa00      4     21
```

```
life_careers <- players %>%
  inner_join(gmsns) %>%
  inner_join(dates)
```

```
## Joining, by = "srid"
## Joining, by = "srid"
```

```
head(life_careers)
```

```
## # A tibble: 6 x 12
##   name srid active hof positions start end letter seasons games
##   <chr> <chr> <chr> <chr> <chr>    <int> <int> <chr>    <int> <int>
## 1 Isaa~ Aait~ False False NT      2013 2014 A        2      4
## 2 Joe ~ Abbe~ False False E       1948 1949 A        2     18
## 3 Fay ~ Abbo~ False False BB-FB-TB~ 1921 1929 A        9     57
```

```
## 4 Vinc~ abbo~ False  False K          1987  1988 A          2    23
## 5 Jare~ Abbr~ True   False WR          2014  2017 A          4    21
## 6 Duke~ Abbr~ False  False HB-DB       1946  1946 A          1     3
## # ... with 2 more variables: dob <date>, dod <date>

write_csv(life_careers, path='data/life-careers.csv')
```