

# Create fiscal years

Anisha Dubhashi

11/13/2017

## load data

```
df <- read_csv("catalog sales data.csv")

## Parsed with column specification:
## cols(
##   targdol = col_double(),
##   datead6 = col_character(),
##   datelp6 = col_character(),
##   lpuryear = col_integer(),
##   slstyr = col_integer(),
##   slslyr = col_integer(),
##   sls2ago = col_integer(),
##   sls3ago = col_integer(),
##   slshist = col_integer(),
##   ordtyr = col_integer(),
##   ordlyr = col_integer(),
##   ord2ago = col_integer(),
##   ord3ago = col_integer(),
##   ordhist = col_integer(),
##   falord = col_integer(),
##   sprord = col_integer(),
##   train = col_integer()
## )

df$datead6 <- as.Date(df$datead6,format = "%m/%d/%Y")
df$datelp6 <- as.Date(df$datelp6,format = "%m/%d/%Y")
```

## Create year\_ordyr

```
df$year_ordyr <- 1980
#needs to be in order of least to most recent for overwriting
df$year_ordyr[df$ord3ago > 0] <- 2009
df$year_ordyr[df$ord2ago > 0] <- 2010
df$year_ordyr[df$ordlyr > 0] <- 2011
df$year_ordyr[df$ordtyr > 0] <- 2012
```

## Create recentseason function

```
findSeason <- function(date, cutoff) {
  if (month(date) < cutoff) {
    season <- "Spring"
  }
  else {
```

```

    season <- "Fall"
  }
  return(season)
}

findSeasons <- function(dataframe, cutoff) {
  seasons <- sapply(dataframe$datelp6, findSeason, cutoff = cutoff)
  return(seasons)
}

```

Create year\_lp6yr

```

df$recentseason <- findSeasons(df, cutoff = 7)

df$year_lp6yr <- year(df$datelp6)

# Jan-June of Year x -> fiscalyear.a = x
df$year_lp6yr[df$recentseason == "Spring"] <- year(df$datelp6[df$recentseason == "Spring"])

# July+ of Year x -> fiscalyear.b = x+1
df$year_lp6yr[df$recentseason == "Fall"] <- year(df$datelp6[df$recentseason == "Fall"]) + 1

#clean the 3 2013s
df$year_lp6yr[df$year_lp6yr == 2013] <- df$year_lp6yr[df$year_lp6yr == 2013] - 1

```

```
table(df$year_lp6yr, df$year_ordyr, dnn = c("lp6 after", "ordyr"))
```

```
##           ordyr
## lp6 after 1980  2009  2010  2011  2012
##      1980    17     0     0     1     0
##      2002     2     0     0     0     0
##      2003 1268     1     0     3     5
##      2004 2181     0     0     3     0
##      2005 3755     0     0     0     3
##      2006 5948     0     0     4     6
##      2007 7639     8     3     3     1
##      2008 9816    607     9     4    11
##      2009     3 12146    653    17    11
##      2010     2     1 15081    367    30
##      2011     1     2     0 17890    462
##      2012   337    220    320    549 22142
```

```
#table(year(df$datelp6), df$year_ordyr, dnn = c("lp6 before", "ordyr"))
```

```
head(df[(df$year_ordyr != df$year_lp6yr) & df$year_ordyr > 1980 & df$targdol, ], 20)
```

```
## # A tibble: 20 x 20
##   targdol   datead6   datelp6 lpuryear slstyr slslyr sls2ago sls3ago
##   <dbl>   <date>   <date>   <int>  <int>  <int>  <int>  <int>
## 1 14.949997 2010-05-03 2012-06-29     1     0    13     0     0
## 2 19.000000 2006-02-05 2012-03-01     3     0     0     0    44
## 3 67.849976 2007-10-21 2012-03-01     3     0     0    70     0
## 4 57.699982 2008-07-06 2012-03-01     3     0    41     0    10
## 5 213.000000 2007-03-04 2012-03-01     3     0     0     0    62
```

```
## 6 51.799988 2007-10-27 2012-03-01      3      0      82      0     107
## 7 48.899994 2009-10-25 2012-03-01      3      0      50     112      0
## 8 23.899994 2011-03-29 2012-03-01      3      0      13      0      0
## 9 56.849976 2005-12-02 2012-03-01      3      0      0      20      0
## 10 44.949982 2007-01-21 2012-03-01      3      0      0      30     19
## 11 18.949997 2009-08-16 2012-03-01      3      0      0      7      0
## 12 59.449982 2010-01-05 2012-03-01      3      0      0      34      0
## 13 12.949997 2008-10-27 2012-03-01      3      0      0      17     23
## 14 19.949997 2009-10-11 2009-11-15      9     23      0      34      0
## 15 9.949997 2008-12-02 2009-03-01      9      0     15      0     63
## 16 13.949997 2009-11-29 2012-03-01      3      0      0      20      0
## 17 79.000000 2010-11-06 2012-03-01      3      0     40      0      0
## 18 17.849991 2011-02-11 2012-03-01      3      0     23      0      0
## 19 85.000000 2008-10-11 2012-03-01      3      0      0     72     10
## 20 38.849976 2009-10-24 2012-03-01      3      0      0     55      0
## # ... with 12 more variables: slshist <int>, ordtyr <int>, ordlyr <int>,
## #   ord2ago <int>, ord3ago <int>, ordhist <int>, falord <int>,
## #   sprord <int>, train <int>, year_ordyr <dbl>, recentseason <chr>,
## #   year_lp6yr <dbl>
```

```
df$max_year <- pmax(df$year_lp6yr, df$year_ordyr)
table(df$max_year, df$year_ordyr, dnn = c("max", "ordyr"))
```

```
##      ordyr
## max      1980  2009  2010  2011  2012
## 1980      17    0    0    0    0
## 2002       2    0    0    0    0
## 2003    1268    0    0    0    0
## 2004    2181    0    0    0    0
## 2005    3755    0    0    0    0
## 2006    5948    0    0    0    0
## 2007    7639    0    0    0    0
## 2008    9816    0    0    0    0
## 2009       3 12762    0    0    0
## 2010       2    1 15746    0    0
## 2011       1    2    0 18292    0
## 2012     337   220   320   549 22671
```

```
table(df$max_year, df$year_lp6yr, dnn = c("max", "lp6"))
```

```
##      lp6
## max      1980  2002  2003  2004  2005  2006  2007  2008  2009  2010  2011
## 1980      17    0    0    0    0    0    0    0    0    0    0
## 2002       0    2    0    0    0    0    0    0    0    0    0
## 2003       0    0 1268    0    0    0    0    0    0    0    0
## 2004       0    0    0 2181    0    0    0    0    0    0    0
## 2005       0    0    0    0 3755    0    0    0    0    0    0
## 2006       0    0    0    0    0 5948    0    0    0    0    0
## 2007       0    0    0    0    0    0 7639    0    0    0    0
## 2008       0    0    0    0    0    0    0 9816    0    0    0
## 2009       0    0    1    0    0    0    8  607 12149    0    0
## 2010       0    0    0    0    0    0    3    9   653 15084    0
## 2011       1    0    3    3    0    4    3    4    17   367 17893
## 2012       0    0    5    0    3    6    1   11   11    30   462
##      lp6
```

##	max	2012
##	1980	0
##	2002	0
##	2003	0
##	2004	0
##	2005	0
##	2006	0
##	2007	0
##	2008	0
##	2009	0
##	2010	0
##	2011	0
##	2012	23568