Homework 6

Aidan Pizzo

10/6/21

#data_munged.R

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)

get_pitching <- function()
{
  p <- read.csv("../data/Pitching.csv", header = T, stringsAsFactors = F)
  return(p)
}

get_salaries <- function()
{
  s <- read.csv("../data/Salaries.csv", header = T, stringsAsFactors = F)
  return(s)
}

get_inflation <- function()
{
  i <- read.csv("../data/inflation.csv", header = T, stringsAsFactors = F)
  return(i)
}

p <- get_pitching()
s <- get_salaries()
inf <- get_inflation()
inf <- as.list(inf)
#i <- enframe(i)

make_df <- function(p, s, inf)
{
  join_df <- dplyr::inner_join(p, s, by=c("playerID", "yearID"))
```

```
  join_df <- dplyr::filter(join_df, yearID < 2015 && yearID > 1984)
  join_df <- dplyr::select(join_df, yearID, playerID, salary, ERA)
  mapYear <- unique(join_df$yearID)
  map_salary <- Map(function(year){
    year_df <- dplyr::filter(join_df, yearID == year)
    multiplier <- inf$inflation2015[inf$year == year]
    era <- sort(year_df$ERA)
    if (year == 2015)
    {
      multiplier = 1
    }
    for (i in 1:length(year_df$ERA))
    {
      year_df$salary[i] <- year_df$salary[i]*multiplier
      ndx <- grep(year_df$ERA[i], era)
      ndx <- ndx[1]
      era_perc <- ndx/length(era)
      year_df$ERA[i] <- era_perc
    }
    return(year_df)
  }, mapYear)
  return(map_salary)
}

df_list <- make_df(p, s, inf)
df <- do.call(rbind, df_list)
names(df) <-c("year", "pitcher", "salary.adjusted", "ERA.adjusted")

plot_df <- Map(function(thisYear){
  year_df <- dplyr::filter(df, year == thisYear)
  year <- thisYear
  avg_sal <- mean(year_df$salary.adjusted)
  avg_era <- mean(year_df$ERA.adjusted, na.rm = TRUE)
  out_df <- data.frame(year= thisYear,
                       salary.average= avg_sal,
                       ERA.average =avg_era,
                       stringsAsFactors = F)
  return(out_df)
}, unique(df$year))

plot_df <- do.call(rbind, plot_df)
```

#analysis.R

```
plot_df <- Map(function(thisYear){
  year_df <- dplyr::filter(df, year == thisYear)
  year <- thisYear
  avg_sal <- mean(year_df$salary.adjusted)
  avg_era <- mean(year_df$ERA.adjusted, na.rm = TRUE)
  out_df <- data.frame(year= thisYear,
                       salary.average= avg_sal,
                       ERA.average =avg_era,
                       stringsAsFactors = F)
  return(out_df)
```

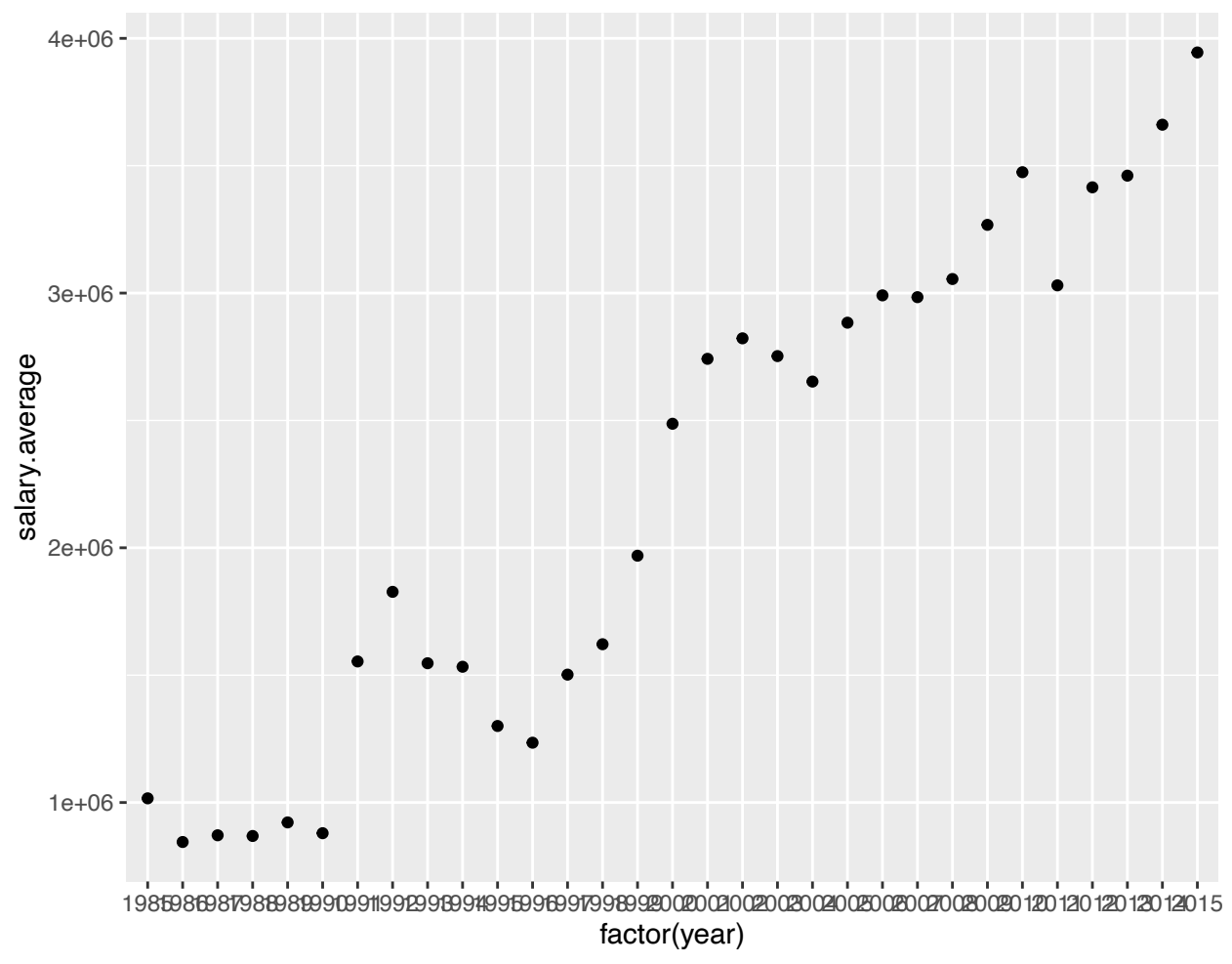
```
}, unique(df$year))

plot_df <- do.call(rbind, plot_df)
```

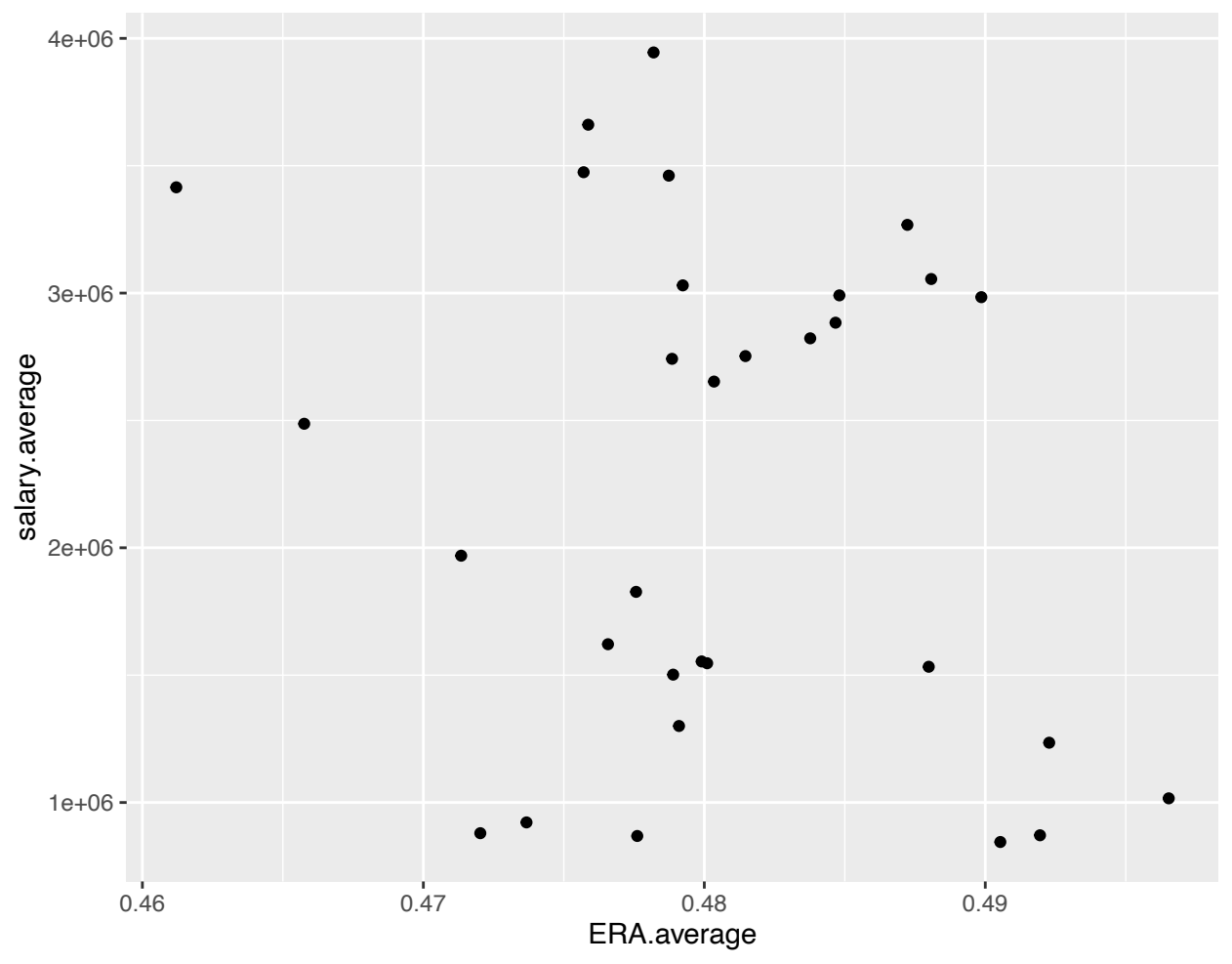
#presentation.R

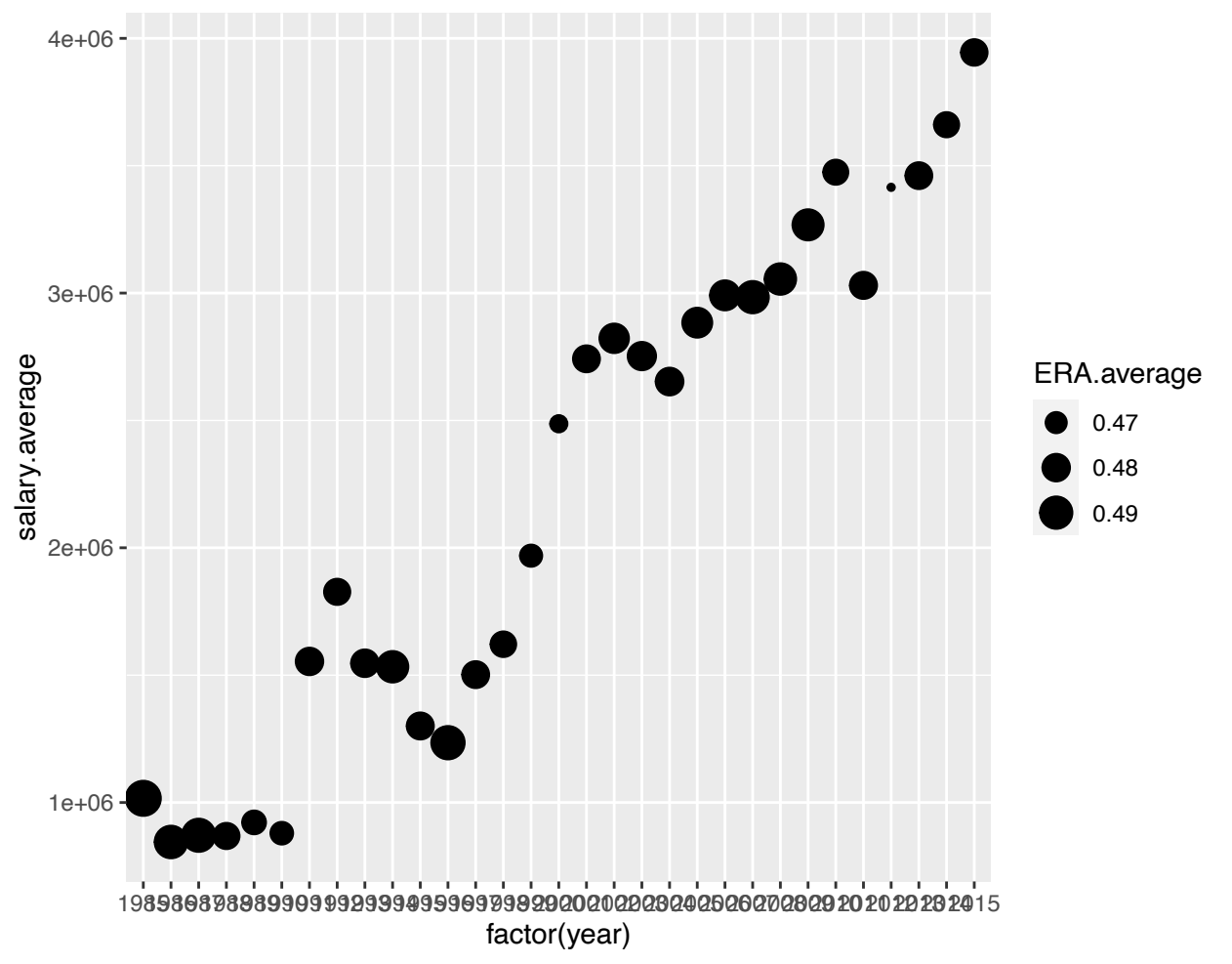
```
source("configuration.R")
g <- ggplot(data=plot_df) +
  geom_point(mapping=aes(x=factor(year),
                         y=salary.average))

g2 <- ggplot(data=plot_df) +
  geom_point(mapping=aes(x=ERA.average,
                         y=salary.average))

g3 <- ggplot(data=plot_df) +
  geom_point(mapping=aes(x=factor(year),
                         y=salary.average,
                         size = ERA.average))
```

#configuration.R

```
source("data_munged.R")
source("analysis.R")
```


4e+06
3e+06
2e+06
1e+06
salary.average
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
factor(year)


4e+06
3e+06
2e+06
1e+06
salary.average
0.46
0.47
0.48
0.49
ERA.average


4e+06
3e+06
2e+06
1e+06
salary.average
factor(year)
ERA.average
0.47
0.48
0.49