Project 1

Lu Beyer

Overview:

I'm reading data from a text file and organizing into a usable csv with excluding irrelevant information. This data is based on chess rankings with several calculated fields.

```
#load tidyverse
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
#read txt as a string
test <- read_file("tournamentinfo.txt")
```

```
#split lines delim by "---"
x <- str_split(test, "---------------------------------------------------------------------------------
```

```
y <- tibble(text = x) %>%
  #establish column names separated by |
  separate(col = text, into = c("pair", "player_name", "total", "Round_1", "Round_2", "Round_3", "Round_
  #remove extra lines from beginning and end
  slice(4:n()-1) %>%
  #remove non-integers from each round 1-7
  mutate(Round_1 = gsub("[^0-9.-]", "", Round_1),
         Round_2 = gsub("[^0-9.-]", "", Round_2),
         Round_3 = gsub("[^0-9.-]", "", Round_3),
         Round_4 = gsub("[^0-9.-]", "", Round_4),
         Round_5 = gsub("[^0-9.-]", "", Round_5),
         Round_6 = gsub("[^0-9.-]", "", Round_6),
         Round_7 = gsub("[^0-9.-]", "", Round_7)) %>%
  #rename column, extract the five characters after R:
  mutate(score = str_extract(USCF_ID_Rtg_pre, "(?<=R:\\s).{5}")) %>%
  #only keep numbers for score and remove blank spaces and stray characters
  mutate(score = gsub("[^0-9.-]", "", score)) %>%
  #column values as numbers instead of characters
  mutate(Round_1 = as.numeric(Round_1),
         Round_2 = as.numeric(Round_2),
         Round_3 = as.numeric(Round_3),
         Round_4 = as.numeric(Round_4),
         Round_5 = as.numeric(Round_5),
         Round_6 = as.numeric(Round_6),
         Round_7 = as.numeric(Round_7),
         pair = as.numeric(pair)) %>%
  #selecting relevant columns
   select(pair, player_name, player_state, score, total, Round_1, Round_2, Round_3, Round_4, Round_5, R
  #matching the pair ID from the Round columns to the opponents scores
  mutate(Round_1 = score[match(Round_1, pair)],
         Round_2 = score[match(Round_2, pair)],
         Round_3 = score[match(Round_3, pair)],
         Round_4 = score[match(Round_4, pair)],
         Round_5 = score[match(Round_5, pair)],
         Round_6 = score[match(Round_6, pair)],
         Round_7 = score[match(Round_7, pair)]) %>%
  #for some reason had to make the rounds numbers again
  mutate(Round_1 = as.numeric(Round_1),
         Round_2 = as.numeric(Round_2),
         Round_3 = as.numeric(Round_3),
         Round_4 = as.numeric(Round_4),
         Round_5 = as.numeric(Round_5),
         Round_6 = as.numeric(Round_6),
         Round_7 = as.numeric(Round_7),
         pair = as.numeric(pair)) %>%
  #finding the mean for opponents scores of all rounds and outputting as a new column
 mutate(avg_opp = rowMeans(select(., Round_1, Round_2, Round_3, Round_4, Round_5, Round_6, Round_7), na
  #selecting relevant columns
  select(player_name, player_state, total, score, avg_opp)
```

```
## Warning: Expected 20 pieces. Additional pieces discarded in 65 rows [2, 3, 4, 5, 6, 7,
## 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

## Warning: Expected 20 pieces. Missing pieces filled with ‘NA‘ in 2 rows [1, 67].
```

```
write.csv(y, "607_Project1.csv")
```

Conclusion:

Using tidyverse, I was able to extract and organize the data into columns that allows the players names, state, total number of points, pre-rating, and average pre chess rating of opponents easily read. In the future, we can perform the same operations on this Tournament data over multiple years, and use that to see how players perform over time.