Tidyverse

Randall Thompson

12/1/2019

Importing CSV files with readr

Using the nba dataset from fivethirtyeight we're going to take a look at how importing csv files in the readr package differs from the base r importing fuctions.

First we will import using the base r function read.csv followed by the readr function read_csv

```
library(readr)
basernba<-read.csv('https://raw.githubusercontent.com/fivethirtyeight/data/master/nba-draymond/draymond
```

```
readrnba<-read_csv('https://raw.githubusercontent.com/fivethirtyeight/data/master/nba-draymond/draymond
```

```
## Parsed with column specification:
## cols(
##   season = col_double(),
##   player = col_character(),
##   possessions = col_double(),
##   DRAYMOND = col_double()
## )
```

The first difference we notice is that read_csv states how each column was parsed.

Let's see what other differences we can find using a function from the dplyr package called glimpse. glimpse() shows you each column, the data type, and shows you as much data as it can within your console.

```
library(dplyr)
glimpse(basernba)
```

```
## Observations: 3,009
## Variables: 4
## $ season      <int> 2017, 2014, 2015, 2014, 2015, 2016, 2017, 2018, 20...
## $ player      <fct> AJ Hammons, AJ Price, AJ Price, Aaron Brooks, Aaro...
## $ possessions <dbl> 331.02580, 211.71560, 633.51860, 3257.93400, 3984....
## $ DRAYMOND    <dbl> -0.1766801, 5.9121720, -1.7909210, -0.9529003, -0....
```

```
glimpse(readrnba)
```

```
## Observations: 3,009
## Variables: 4
## $ season      <dbl> 2017, 2014, 2015, 2014, 2015, 2016, 2017, 2018, 20...
## $ player      <chr> "AJ Hammons", "AJ Price", "AJ Price", "Aaron Brook...
## $ possessions <dbl> 331.02580, 211.71560, 633.51860, 3257.93400, 3984....
## $ DRAYMOND    <dbl> -0.1766801, 5.9121720, -1.7909210, -0.9529003, -0....
```

The default assignment for strings in read.csv are factors while the default assignment for strings in read_csv are characters. This is an important difference to know for text analysis.

Another difference is that read.csv parses numbers with no decimals as integers while read_csv parses all numbers as doubles for added precision.

There are many other notable differences between these two functions so hopefully I made this super easy for someone's part 2.