Classwork

Immanuel Williams Ph.D.

8/19/2020

Every Time You Create A Dataframe, Explore It

Wednesday: String Manipulations & R Shiny Basics

Homework Questions

HW1. Within the mtcars dataframe, make the rownames a column. Name the column an appropriate name. Use the following pattern '[0-9-]' (more regex code) to detect if a number is present in the car name variable. Provide a dataframe where all the car name variable does not have numbers. Provide the dataframe with an appropriate name.

HW2. Within the mtcars dataframe, make the rownames a column. Name the column an appropriate name. Replace the word 'Merc' with 'Mercedes Benz'. Provide the dataframe with an appropriate name.

HW3 Within the mtcars dataframe, make the rownames a column. Name the column an appropriate name. Using the str_to_lower function, make all the words in the car names variable lowercase. Keep the variable name the same. Save the dataframe with an appropriate name.

Every Time You Create A New Dataframe, Explore It

Run Previous Code, we will begin with **mod1_great_players_df**.

Q1. Install the following package. It is used to manipulate date data.

```
# install.packages('lubridate')
library(lubridate)

##
## Attaching package: 'lubridate'

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

Q2. What type of data is R recognizing the variable Date as? 'Date' is another data type that exists in R. Convert the *Date* variable into a 'Date' data type using the *as_data()* function (Keep the name the same). Verify that you changed the data type for the *Date* variable. Save the dataframe with **mod2_** in front of dataframe name.

Q3. Create a new variable named *month_game* by using the *Date* variable as an input into the *months()* function. Create another new variable named *year_game* by using the *Date* variable as an input into the *year()* function. Make the *year_game* variable a character. Why would I ask you to do that? Verify that you changed the data type for the *year_game* variable. Save the dataframe with **mod3_** in front of dataframe name.

Using the Separate Function

Below there is a variable in a dataframe where there is a symbol separating two pieces of information.

```
self_df = data.frame(day_feelings = c('Mon_Happy for math', 'Tues_Chill to re
ad','Wed_Excited for Data Science'))
```

We can separate the pieces of information into two variables using the *separate function*. When remove is **FALSE** the original variable is kept. It is traditionally ran as **TRUE**.

```
self_df %>%
  separate(col = day_feelings, into = c('day', 'feelings'), sep = '_' ,remove
= FALSE)

##                   day_feelings  day                 feelings
## 1           Mon_Happy for math  Mon           Happy for math
## 2           Tues_Chill to read Tues            Chill to read
## 3 Wed_Excited for Data Science  Wed Excited for Data Science
```

Q4. Part a Turn the *Age* variable two variables, *age_years* and *age_days*. Convert both new variables into numeric variables. Verify that you changed the data type for the *age_years* and *age_days* variables. Save the dataframe with **mod4_** in front of dataframe name.

Q4. Part b Create another new variable named *Age_Numeric* by adding *age_days* divided by 365 plus *age_years*. Save the dataframe with **mod5_** in front of dataframe name.

Q5. Separate Game outcome and point difference into two separate variables. Write out plan first then execute. Save the dataframe with **mod6_** in front of dataframe name.

Q6. Turn Minutes Played (*MP*) into a numeric variable. Write out plan first then execute. (*Hint turn seconds into a decimal*.) Save the dataframe with **mod7_** in front of dataframe name.

Q7. Turn the remaining numeric variables to be numeric. Save the dataframe with **mod8_** in front of dataframe name.

Congratulations, YOUR DATA is clean

Q8. How many games in MJ win and lose by division? How many games in MJ win and lose by conference?

Q9. How many games in LJ win and lose by division? How many games in LJ win and lose by conference?

Q10. How many games did each player's team win by more than 10 points?

Q11. Which division was the toughest to play for each player's team? (*Define tough as the average point difference.*) Create a new variable based on this rank and plot how did this player perform in terms *GmSc*.

Q12. - Q14. Create three plots based variables that you believe will have a relationship. Use variables that make sense in the context in basketball. Make sure to label the x and y axes as well as the title. Make sure the title is centered. What pattern do you notice in this plot?