Data Science 101 - Class Notes

Dr. Immanuel Williams

Concept 6

Description of a Game dataframe

Description of the a Game Dataframes

There are dataframes regarding two versions of a Game (Game 1 and Game 2). Various information is gathered regarding the players that either bought the game via cd or downloading it that are presented in these dataframes. Other variables such as age, system, and days as well as many others to track how players play these games.

Install Game Dataframes

```
install.packages("devtools")
devtools::install_git("https://github.com/jamesijw23/stat_calpoly_package.git")
```

Bind Columns

Combining Dataframes

Method 1: bind_cols()

Binding columns is when two dataframes of the same row lengths are combine. This is very important when new variables are collected on existing observations and the order of the observations are the same in both dataframes.

bind_cols()

There is a dataframe named **basic_info_teen_df** for teenagers that have the cd and downloaded versions of Game 1.

```
basic_info_teen_df
## # A tibble: 3 x 4
##      id       group_age      age      gender
##   <chr>       <chr>        <dbl>     <chr>
## 1 id_333      teenage        17      Female
## 2 id_122      teenage        13      Female
## 3 id_530      teenage        14      Male
```

There is another dataframe named **game_info_teen_df** for the same individuals above.

```
game_info_teen_df
## # A tibble: 3 x 3
##   game          cd       download
##   <chr>       <dbl>      <dbl>
## 1 game_1        1            1
## 2 game_1        1            1
## 3 game_1        1            1
```

We can bind these columns from the **basic_info_teen_df** and **game_info_teen_df** by using the following function, where order just places one dataframe in front of another:

```
bind_cols(basic_info_teen_df,game_info_teen_df)
## # A tibble: 3 x 7
##      id      group_age    age    gender   game      cd      download
##   <chr>      <chr>      <dbl>   <chr>   <chr>    <dbl>     <dbl>
## 1 id_333   teenage        17    Female   game_1     1         1
## 2 id_122   teenage        13    Female   game_1     1         1
## 3 id_530   teenage        14    Male     game_1     1         1
```

This also could be done using the pipe function:

```
basic_info_teen_df %>%
  bind_cols(game_info_teen_df)
## # A tibble: 3 x 7
##      id      group_age    age    gender   game      cd      download
##   <chr>      <chr>      <dbl>   <chr>   <chr>    <dbl>     <dbl>
## 1 id_333   teenage        17    Female   game_1     1         1
## 2 id_122   teenage        13    Female   game_1     1         1
## 3 id_530   teenage        14    Male     game_1     1         1
```

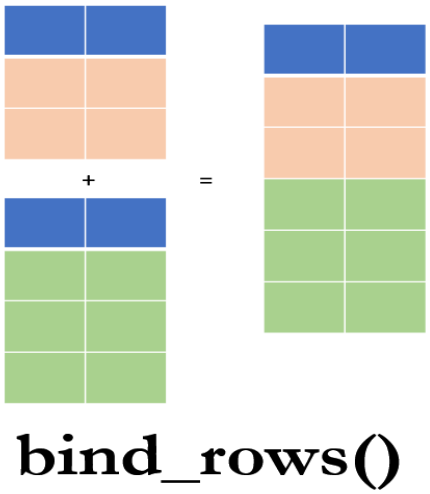
Question about Binds Columns

There are two dataframes named, **basic_info_adult_df** and **game_info_adult_df**. Combine the columns of these dataframes, provide an appropriate name for the dataframe, and specifiy the variable types in the combined dataframe.

Bind Rows

Method 2: bind_rows()

Binding rows is when two dataframes of the same column length are combine. This is very important when new observations are collected with respect existing variables that already have observations.



There is a dataframe named **basic_info_teen_df** for teenagers that have the cd and downloaded versions of Game 1.

```
basic_info_teen_df
## # A tibble: 3 x 4
##      id      group_age      age      gender
##   <chr>       <chr>      <dbl>     <chr>
## 1 id_333      teenage      17       Female
## 2 id_122      teenage      13       Female
## 3 id_530      teenage      14       Male
```

There is another dataframe named **basic_info_adult_df** for adults that have the cd and downloaded versions of Game 1.

```
basic_info_adult_df
## # A tibble: 3 x 4
##      id      group_age      age      gender
##   <chr>       <chr>      <dbl>     <chr>
## 1 id_237      adult        23       Male
## 2 id_225      adult        27       Female
## 3 id_309      adult        19       Male
```

We can bind these rows from the **basic_info_teen_df** and **basic_info_adult_df** by using the following function, where order just places one dataframe on top of the nother:

```
bind_rows(basic_info_teen_df,basic_info_adult_df)
## # A tibble: 6 x 4
##       id      group_age    age       gender
##    <chr>      <chr>      <dbl>     <chr>
## 1 id_333     teenage       17      Female
## 2 id_122     teenage       13      Female
## 3 id_530     teenage       14      Male
## 4 id_237     adult         23      Male
## 5 id_225     adult         27      Female
## 6 id_309     adult         19      Male
```

This also could be done using the pipe function:

```
basic_info_teen_df %>%
  bind_rows(basic_info_adult_df)
## # A tibble: 6 x 4
##       id      group_age    age       gender
##    <chr>      <chr>      <dbl>     <chr>
## 1 id_333     teenage       17      Female
## 2 id_122     teenage       13      Female
## 3 id_530     teenage       14      Male
## 4 id_237     adult         23      Male
## 5 id_225     adult         27      Female
## 6 id_309     adult         19      Male
```

Question about Bind Rows

There are two dataframes named, **game_info_teen_df** and **game_info_adult_df**. Combine the rows of these dataframes, provide an appropriate name for the dataframe, and specifiy the variable types in the combined dataframe.