Inclass Assignment (6)

AksharPatel

3/16/2022

# Run this line first

if (!require('tidyverse')) install.packages('tidyverse'); library('tidyverse')

## Loading required package: tidyverse

## Warning: package 'tidyverse' was built under R version 4.1.3

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.6 v dplyr 1.0.8  
## v tidyr 1.2.0 v stringr 1.4.0  
## v readr 2.1.2 v forcats 0.5.1

## Warning: package 'tidyr' was built under R version 4.1.3

## Warning: package 'readr' was built under R version 4.1.3

## Warning: package 'forcats' was built under R version 4.1.3

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

Q1. Update the author name and date in YAML part above.

Q2. A line of code to import the student.csv file from your Desktop folder and create department data for this assignment

department <- read\_csv("C:/Users/Akshar/Desktop/Inclass\_Assigments/student.csv")

## Rows: 1192 Columns: 6  
## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): sex, race  
## dbl (4): earn, height, ed, age  
##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

Q3. A line of code to return column names of department data

colnames(department)

## [1] "earn" "height" "sex" "ed" "age" "race"

Q4. Two lines of code to divide height column of department data into height\_a and height\_b

department %>%  
 separate(height, into = c("height\_a", "height\_b"))

## # A tibble: 1,192 x 7  
## earn height\_a height\_b sex ed age race   
## <dbl> <chr> <chr> <chr> <dbl> <dbl> <chr>   
## 1 50000 74 4244387818035 male 16 45 white   
## 2 60000 65 5375428255647 female 16 58 white   
## 3 30000 63 6291977374349 female 16 29 white   
## 4 50000 63 1085616752971 female 16 91 other   
## 5 51000 63 4024835710879 female 17 39 white   
## 6 9000 64 3995075440034 female 15 26 white   
## 7 29000 61 6563258264214 female 12 49 white   
## 8 32000 72 6985437364783 male 17 46 white   
## 9 2000 72 0394668497611 male 15 21 hispanic  
## 10 27000 72 2349325611722 male 12 26 white   
## # ... with 1,182 more rows

Q5. Two lines of code to combine sex and race column into demo column (with an underscore)

department %>%  
 unite(demo, sex, race)

## # A tibble: 1,192 x 5  
## earn height demo ed age  
## <dbl> <dbl> <chr> <dbl> <dbl>  
## 1 50000 74.4 male\_white 16 45  
## 2 60000 65.5 female\_white 16 58  
## 3 30000 63.6 female\_white 16 29  
## 4 50000 63.1 female\_other 16 91  
## 5 51000 63.4 female\_white 17 39  
## 6 9000 64.4 female\_white 15 26  
## 7 29000 61.7 female\_white 12 49  
## 8 32000 72.7 male\_white 17 46  
## 9 2000 72.0 male\_hispanic 15 21  
## 10 27000 72.2 male\_white 12 26  
## # ... with 1,182 more rows

Q6. Two lines of code to combine sex and race column into demo column (WITHOUT an underscore)

department %>%  
 unite(demo, sex, race, sep = "")

## # A tibble: 1,192 x 5  
## earn height demo ed age  
## <dbl> <dbl> <chr> <dbl> <dbl>  
## 1 50000 74.4 malewhite 16 45  
## 2 60000 65.5 femalewhite 16 58  
## 3 30000 63.6 femalewhite 16 29  
## 4 50000 63.1 femaleother 16 91  
## 5 51000 63.4 femalewhite 17 39  
## 6 9000 64.4 femalewhite 15 26  
## 7 29000 61.7 femalewhite 12 49  
## 8 32000 72.7 malewhite 17 46  
## 9 2000 72.0 malehispanic 15 21  
## 10 27000 72.2 malewhite 12 26  
## # ... with 1,182 more rows

#run this code to make a stock data

stock <- tibble(  
 time = c(2015, 2015, 2015, 2015, 2016, 2016, 2016),  
 division = c( 1, 2, 3, 4, 2, 3, 4),  
 profit = c(23, 452, 888, NA, 3234, 279, NA)  
)

Q7. How many implicit and explicit NA values does stock data have?

# implicit NA = 1  
# explicit NA = 2

Q8. Two lines of code to change implicit NA values into explicit ones using pivot\_wider() (\*tip: time and profit column)

stock %>%  
 pivot\_wider(names\_from = time , values\_from = profit)

## # A tibble: 4 x 3  
## division `2015` `2016`  
## <dbl> <dbl> <dbl>  
## 1 1 23 NA  
## 2 2 452 3234  
## 3 3 888 279  
## 4 4 NA NA

Q9. Two lines of code to change implicit NA values to explicit ones using complete() (\*tip: time and division column)

stock %>%  
 complete(time, division)

## # A tibble: 8 x 3  
## time division profit  
## <dbl> <dbl> <dbl>  
## 1 2015 1 23  
## 2 2015 2 452  
## 3 2015 3 888  
## 4 2015 4 NA  
## 5 2016 1 NA  
## 6 2016 2 3234  
## 7 2016 3 279  
## 8 2016 4 NA

Q10. Three lines of code to fill in the NA value with the most recent non-missing value using fill()

stock %>%  
 complete(time, division)%>%  
 fill(profit)

## # A tibble: 8 x 3  
## time division profit  
## <dbl> <dbl> <dbl>  
## 1 2015 1 23  
## 2 2015 2 452  
## 3 2015 3 888  
## 4 2015 4 888  
## 5 2016 1 888  
## 6 2016 2 3234  
## 7 2016 3 279  
## 8 2016 4 279