R Notebook

GUOTAI SUN

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
v <- c(1,4,4,3,2,2,3)
print(v)
```

```
## [1] 1 4 4 3 2 2 3
```

```
print(v[3])
```

```
## [1] 4
```

```
print(v[c(2,3,4)])
```

```
## [1] 4 4 3
```

```
print(v[2:4])
```

```
## [1] 4 4 3
```

```
print(v[c(2,4,3)])
```

```
## [1] 4 3 4
```

```
df <- read.table(header=T, text='
 subject sex size
       1   M    7
       2   F    6
       3   F    9
       4   M   11
 ')

print(df)
```

```
##   subject sex size
## 1       1   M    7
## 2       2   F    6
## 3       3   F    9
## 4       4   M   11
```

```
print(df[1,3])
```

```
## [1] 7
```

```
print(df[1:2, ])
```

```
##   subject sex size
## 1       1   M    7
## 2       2   F    6
```

```
print(df[1:2, 2])
```

```
## [1] "M" "F"
```

```
df[1:2, c("sex","size")]
```

```
##   sex size
## 1   M    7
## 2   F    6
```

```
df$sex
```

```
## [1] "M" "F" "F" "M"
```

```
f <- df$size >= 9
f
```

```
## [1] FALSE FALSE  TRUE  TRUE
```

```
df[f, ]
```

```
##   subject sex size
## 3       3   F    9
## 4       4   M   11
```

```
df[df$size >= 9, ]
```

```
##   subject sex size
## 3       3   F    9
## 4       4   M   11
```

```
print(v)
```

```
## [1] 1 4 4 3 2 2 3
```

```
print(v[-1])
```

```
## [1] 4 4 3 2 2 3
```

```
print(v[-1:-3])
```

```
## [1] 3 2 2 3
```

```
print(v)
```

```
## [1] 1 4 4 3 2 2 3
```

```
print(length(v))
```

```
## [1] 7
```

```
print(v[length(v)])
```

```
## [1] 3
```

```
print(tail(v, 1))
```

```
## [1] 3
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages --------------------------------------- tidyve
rse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'tibble' was built under R version 4.0.5
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## -- Conflicts ------------------------------------------ tidyverse_co
nflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
ICEdata <- read_csv("ICE1_Data.csv")
```

```
##
## -- Column specification --------------------------------------------
------------
## cols(
##   DBN = col_character(),
##   Quality_Review_Score = col_character(),
##   `Progress_Rpt_10-11` = col_character(),
##   `Student_Progress_10-11` = col_character(),
##   `graduation 2010-11` = col_double(),
##   `college enroll 2010-11` = col_double()
## )
```

```
ICEdata
```

```
## # A tibble: 422 x 6
##    DBN    Quality_Review_S~ `Progress_Rpt_10~ `Student_Progres~ `gra
duation 201~
##    <chr>  <chr>             <chr>             <chr>
```

```
       <dbl>
##  1 01M292 Developing        C                 C
       0.563
##  2 01M448 Developing        C                 B
       0.707
##  3 01M450 Well Developed    A                 B
       0.716
##  4 01M509 Proficient        C                 C
       0.564
##  5 01M539 Proficient        A                 A
       0.953
##  6 01M696 Well Developed    B                 C
       0.976
##  7 02M047 Proficient        C                 D
       0.696
##  8 02M288 Proficient        A                 B
       0.82
##  9 02M294 Well Developed    B                 B
       0.675
## 10 02M296 Proficient        A                 A
       0.793
## # ... with 412 more rows, and 1 more variable: college enroll 2010-1
1 <dbl>
```

```
new_ICE <- select(ICEdata, `DBN`, `Quality_Review_Score`, `Progress_Rpt
_10-11`)
new_ICE
```

```
## # A tibble: 422 x 3
##    DBN    Quality_Review_Score `Progress_Rpt_10-11`
##    <chr>  <chr>                <chr>
##  1 01M292 Developing           C
##  2 01M448 Developing           C
##  3 01M450 Well Developed       A
##  4 01M509 Proficient           C
##  5 01M539 Proficient           A
##  6 01M696 Well Developed       B
##  7 02M047 Proficient           C
##  8 02M288 Proficient           A
##  9 02M294 Well Developed       B
## 10 02M296 Proficient           A
## # ... with 412 more rows
```

```
collegeGraduation <- mutate(ICEdata, colllegeGraduationRate = `college
enroll 2010-11` / `graduation 2010-11`)
collegeGraduation
```

```
## # A tibble: 422 x 7
##    DBN    Quality_Review_S~ `Progress_Rpt_10~ `Student_Progres~ `gra
duation 201~
##    <chr>  <chr>             <chr>             <chr>
```

```
       <dbl>
##  1 01M292 Developing        C                 C
       0.563
##  2 01M448 Developing        C                 B
       0.707
##  3 01M450 Well Developed    A                 B
       0.716
##  4 01M509 Proficient        C                 C
       0.564
##  5 01M539 Proficient        A                 A
       0.953
##  6 01M696 Well Developed    B                 C
       0.976
##  7 02M047 Proficient        C                 D
       0.696
##  8 02M288 Proficient        A                 B
       0.82
##  9 02M294 Well Developed    B                 B
       0.675
## 10 02M296 Proficient        A                 A
       0.793
## # ... with 412 more rows, and 2 more variables: college enroll 2010-
11 <dbl>,
## #   colllegeGraduationRate <dbl>
```

```
filter(ICEdata, `graduation 2010-11` > 0.8)
```

```
## # A tibble: 103 x 6
##    DBN    Quality_Review_S~ `Progress_Rpt_10~ `Student_Progres~ `gra
duation 201~
##    <chr>  <chr>             <chr>             <chr>
       <dbl>
##  1 01M539 Proficient        A                 A
       0.953
##  2 01M696 Well Developed    B                 C
       0.976
##  3 02M288 Proficient        A                 B
       0.82
##  4 02M298 Well Developed    A                 A
       0.915
##  5 02M316 Proficient        B                 C
       0.803
##  6 02M400 Well Developed    B                 C
       0.832
##  7 02M407 Well Developed    A                 A
       0.833
##  8 02M408 Proficient        B                 B
       0.931
##  9 02M411 Well Developed    A                 A
       1
```

```
## 10 02M412 Proficient        A                 B
       0.964
## # ... with 93 more rows, and 1 more variable: college enroll 2010-11
 <dbl>
```

```
filter(ICEdata, `graduation 2010-11` > 0.8 & `Quality_Review_Score` ==
"Proficient")
```

```
## # A tibble: 45 x 6
##    DBN    Quality_Review_S~ `Progress_Rpt_10~ `Student_Progres~ `gra
duation 201~
##    <chr>  <chr>             <chr>             <chr>
       <dbl>
##  1 01M539 Proficient        A                 A
       0.953
##  2 02M288 Proficient        A                 B
       0.82
##  3 02M316 Proficient        B                 C
       0.803
##  4 02M408 Proficient        B                 B
       0.931
##  5 02M412 Proficient        A                 B
       0.964
##  6 02M414 Proficient        A                 A
       0.98
##  7 02M420 Proficient        B                 C
       0.824
##  8 03M541 Proficient        B                 B
       0.875
##  9 04M555 Proficient        A                 A
       0.806
## 10 04M610 Proficient        A                 B
       0.984
## # ... with 35 more rows, and 1 more variable: college enroll 2010-11
 <dbl>
```

```
new_ICE <- select(ICEdata, `Quality_Review_Score`, `Student_Progress_10
-11`, `graduation 2010-11`)
filter(new_ICE, `graduation 2010-11` > 0.8)
```

```
## # A tibble: 103 x 3
##    Quality_Review_Score `Student_Progress_10-11` `graduation 2010-11
`
##    <chr>                <chr>                                    <dbl
>
##  1 Proficient           A                                        0.95
3
##  2 Well Developed       C                                        0.97
6
##  3 Proficient           B                                        0.82
```

```
##  4 Well Developed       A                                       0.91
5
##  5 Proficient           C                                       0.80
3
##  6 Well Developed       C                                       0.83
2
##  7 Well Developed       A                                       0.83
3
##  8 Proficient           B                                       0.93
1
##  9 Well Developed       A                                       1

## 10 Proficient           B                                       0.96
4
## # ... with 93 more rows
```

```
ICEdata %>%
  select(`Quality_Review_Score`, `Student_Progress_10-11`, `graduation
2010-11`) %>%
  filter(`graduation 2010-11` > 0.8)
```

```
## # A tibble: 103 x 3
##    Quality_Review_Score `Student_Progress_10-11` `graduation 2010-11
`
##    <chr>                <chr>                                   <dbl
>
##  1 Proficient           A                                       0.95
3
##  2 Well Developed       C                                       0.97
6
##  3 Proficient           B                                       0.82

##  4 Well Developed       A                                       0.91
5
##  5 Proficient           C                                       0.80
3
##  6 Well Developed       C                                       0.83
2
##  7 Well Developed       A                                       0.83
3
##  8 Proficient           B                                       0.93
1
##  9 Well Developed       A                                       1

## 10 Proficient           B                                       0.96
4
## # ... with 93 more rows
```

```
16 %>% sqrt()
```

```
## [1] 4
```

```
16 %>% sqrt() %>% log2()
```

```
## [1] 2
```

```
16 %>% sqrt() %>% log(base = 2)
```

```
## [1] 2
```

```
ICEdata %>% select(`Quality_Review_Score`, `Student_Progress_10-11`, `g
raduation 2010-11`) %>% filter(`graduation 2010-11` > 0.8)
```

```
## # A tibble: 103 x 3
##    Quality_Review_Score `Student_Progress_10-11` `graduation 2010-11
`
##    <chr>                <chr>                                   <dbl
>
##  1 Proficient           A                                       0.95
3
##  2 Well Developed       C                                       0.97
6
##  3 Proficient           B                                       0.82

##  4 Well Developed       A                                       0.91
5
##  5 Proficient           C                                       0.80
3
##  6 Well Developed       C                                       0.83
2
##  7 Well Developed       A                                       0.83
3
##  8 Proficient           B                                       0.93
1
##  9 Well Developed       A                                       1

## 10 Proficient           B                                       0.96
4
## # ... with 93 more rows
```

```
v <- c(1,4,4,3,2,2,3)
v
```

```
## [1] 1 4 4 3 2 2 3
```

```
c(v, 9)
```

```
## [1] 1 4 4 3 2 2 3 9
```

```
append(v, 9)
```

```
## [1] 1 4 4 3 2 2 3 9
```

```r
#install.packages("nycflights13")
library(nycflights13)
```

```
## Warning: package 'nycflights13' was built under R version 4.0.5
```

```r
data(flights)
data(weather)

flights <- flights %>% select(carrier, flight,
                              origin, dest, time_hour)
weather <- weather %>% select(temp, wind_speed, precip, origin, time_ho
ur)
mergeCols <- c("origin", "time_hour")

left_dplyr  <- left_join(flights,  weather, by = mergeCols)
right_dplyr <- right_join(flights, weather, by = mergeCols)
inner_dplyr <- inner_join(flights, weather, by = mergeCols)
full_dplyr  <- full_join(flights,  weather, by = mergeCols)

g <- ICEdata %>%
  filter(`Quality_Review_Score` == "Proficient") %>%
  summarise(average = mean(`graduation 2010-11`, na.rm = TRUE),
            standardDeviation = sd(`graduation 2010-11`, na.rm = TRUE))
g
```

```
## # A tibble: 1 x 2
##   average standardDeviation
##     <dbl>             <dbl>
## 1   0.730             0.136
```

```r
g$average
```

```
## [1] 0.7295541
```

```r
ICEdata %>% group_by(Quality_Review_Score)
```

```
## # A tibble: 422 x 6
## # Groups:   Quality_Review_Score [7]
##    DBN    Quality_Review_S~ `Progress_Rpt_10~ `Student_Progres~ `gra
duation 201~
##    <chr>  <chr>             <chr>             <chr>
      <dbl>
##  1 01M292 Developing        C                 C
      0.563
##  2 01M448 Developing        C                 B
      0.707
##  3 01M450 Well Developed    A                 B
      0.716
##  4 01M509 Proficient        C                 C
      0.564
##  5 01M539 Proficient        A                 A
```

```
       0.953
##  6 01M696 Well Developed    B                 C
       0.976
##  7 02M047 Proficient        C                 D
       0.696
##  8 02M288 Proficient        A                 B
       0.82
##  9 02M294 Well Developed    B                 B
       0.675
## 10 02M296 Proficient        A                 A
       0.793
## # ... with 412 more rows, and 1 more variable: college enroll 2010-1
1 <dbl>
```

```
ICEdata %>%
  group_by(Quality_Review_Score) %>%
  summarize(GraduationAverage = mean(`graduation 2010-11`, na.rm = TRUE
),
            CollegeEnrollAverage = mean(`college enroll 2010-11`, na.rm
 = TRUE))
```

```
## # A tibble: 7 x 3
##   Quality_Review_Score                        GraduationAverage Colle
geEnrollAve~
##   <chr>                                                  <dbl>
       <dbl>
## 1 Developing                                             0.633
       0.403
## 2 Outstanding (only an option in 2007-8)                 0.864
       0.748
## 3 Proficient                                             0.730
       0.521
## 4 Underdeveloped                                         0.550
       0.350
## 5 Underdeveloped with Proficient Features (~           NaN
     NaN
## 6 Well Developed                                         0.823
       0.626
## 7 <NA>                                                   0.689
       0.442
```

```
ICEdata %>%
  group_by(`Quality_Review_Score`, `Progress_Rpt_10-11`) %>%
  summarize(count = n())
```

```
## `summarise()` has grouped output by 'Quality_Review_Score'. You can
override using the `.groups` argument.
```

```
## # A tibble: 22 x 3
## # Groups:   Quality_Review_Score [7]
##    Quality_Review_Score                    `Progress_Rpt_10-11` count
```

```
##    <chr>                                  <chr>                <int>
##  1 Developing                             A                        4
##  2 Developing                             B                       11
##  3 Developing                             C                       19
##  4 Developing                             D                       16
##  5 Developing                             F                        3
##  6 Developing                             <NA>                    11
##  7 Outstanding (only an option in 2007-8) A                        3
##  8 Proficient                             A                       48
##  9 Proficient                             B                       65
## 10 Proficient                             C                       39
## # ... with 12 more rows
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.