Midterm II Winter 2021

Data Science for Managers

Concepts and Code Interpretation

# Exam parameters:

* Closed book
* No time limit
* Complete in one sitting
* Submit before starting the second part of the exam
* Due: 23 March before class
* Submit a word document

# Concepts

1. Explain why pivot\_longer and pivot\_wider are opposite operations.
2. Compare and contrast mutating and filtering joins.
3. First define factors and explain what is meant by a factor’s level and label.
4. Explain when and why you should write an R function.
5. Explain the relationship between the cut function and if else conditional execution.

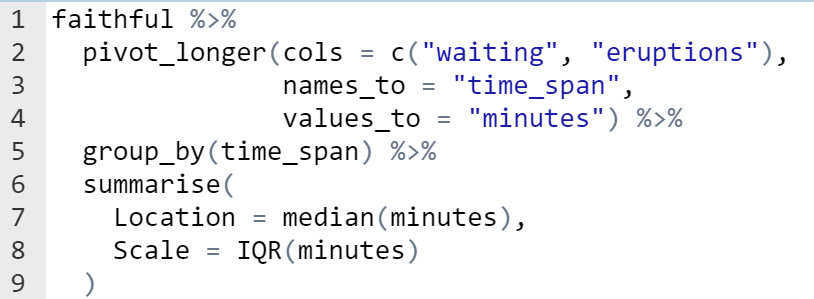
# Code Interpretation

As you interpret the following code, feel free to copy and paste it into RStudio so you can experiment with the outcomes for each line of code. Make sure that you explain your code by identifying each line of code you are interpreting. Make sure that your explanation focuses on what the code is accomplishing and doesn’t just repeat the code.

## Code Chunk #1

### Interpret lines 2 thru 8

Please recall that the faithful tibble has two columns or variables. They are the waiting time between eruptions (waiting) and the duration of the eruption of Old Faithful (eruptions).



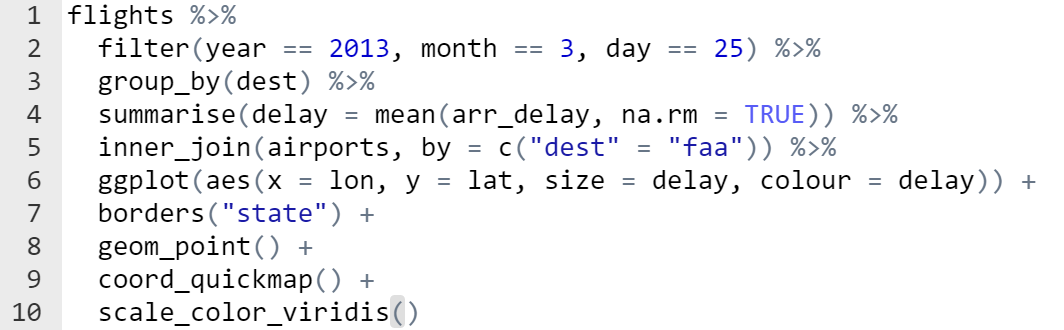
### Output

## # A tibble: 2 x 3  
## time\_span Location Scale  
## <chr> <dbl> <dbl>  
## 1 eruptions 4 2.29  
## 2 waiting 76 24

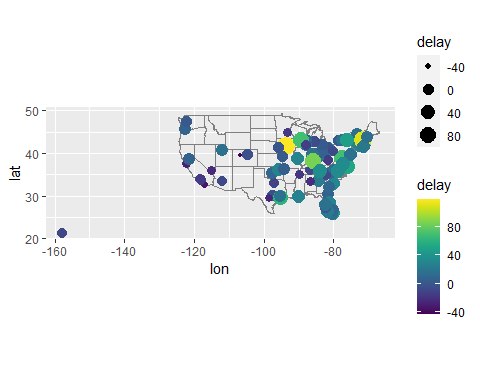
## Code Chunk #2

### Interpret lines 2 thru 6

Please remember that the flights tibble reports all flights leaving from one of the three New York City airports during the year 2013. The airports tibble include information about each of the airports in the FAA database.

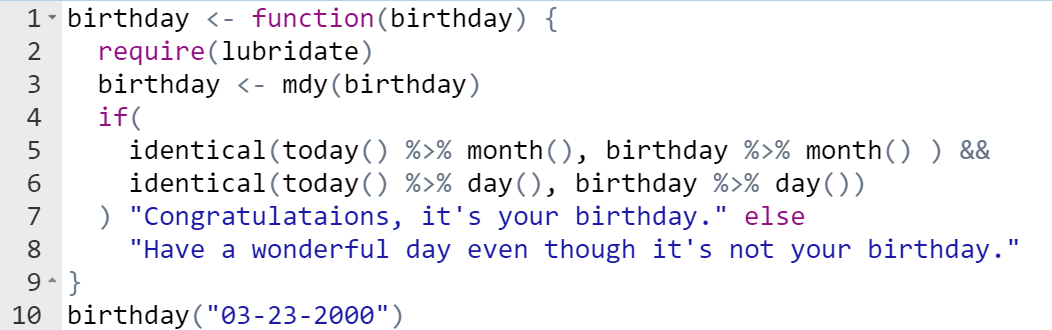


### Output



## Code Chunk #3

### Interpret lines 1 thru 10

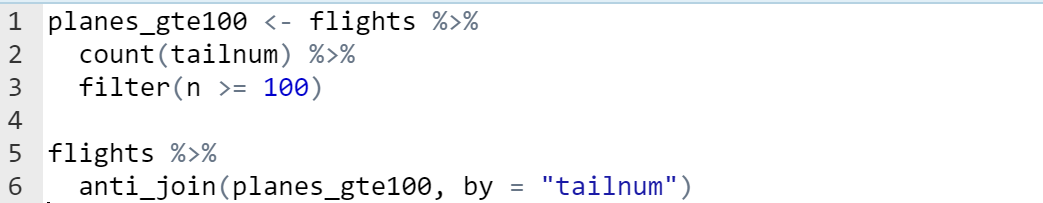


### Output

## [1] "Have a wonderful day even though it's not your birthday."

## Code Chunk #4

### Interpret lines 2 thru 6

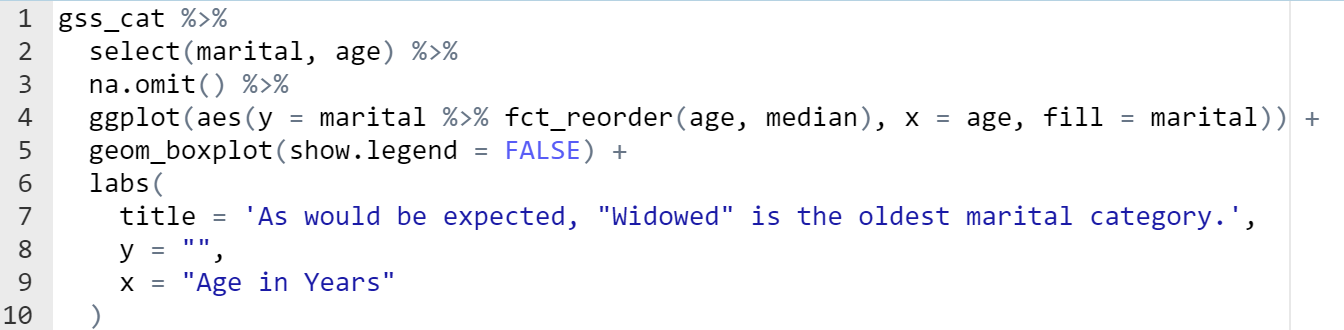


### Output

## # A tibble: 105,874 x 19  
## year month day dep\_time sched\_dep\_time dep\_delay arr\_time sched\_arr\_time  
## <int> <int> <int> <int> <int> <dbl> <int> <int>  
## 1 2013 1 1 542 540 2 923 850  
## 2 2013 1 1 554 600 -6 812 837  
## 3 2013 1 1 558 600 -2 753 745  
## 4 2013 1 1 558 600 -2 924 917  
## 5 2013 1 1 559 600 -1 941 910  
## 6 2013 1 1 602 610 -8 812 820  
## 7 2013 1 1 606 610 -4 858 910  
## 8 2013 1 1 611 600 11 945 931  
## 9 2013 1 1 622 630 -8 1017 1014  
## 10 2013 1 1 623 610 13 920 915  
## # ... with 105,864 more rows, and 11 more variables: arr\_delay <dbl>,  
## # carrier <chr>, flight <int>, tailnum <chr>, origin <chr>, dest <chr>,  
## # air\_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>, time\_hour <dttm>

## Code Chunk #5

### Interpret lines 2 thru 4



### Output

