**nycflights13**

For this assignment, you must name your R file nycflights13.R

* For all questions you should load tidyverse and nycflights13. You should not need to use any other libraries.
  + If the tidyverse package is not installed, you’ll need to do a *one-time installation* from the Console Window in RStudio like this:  
    install.packages("tidyverse")
  + If the nycflights13 package is not installed, you’ll need to do a *one-time installation* from the Console Window in RStudio like this:  
    install.packages("nycflights13")
  + Load tidyverse with: suppressPackageStartupMessages(library(tidyverse))
  + Load nycflights13 with:  
    suppressPackageStartupMessages(library(nycflights13))
  + The actual data set is called flights.  
    See the [nycflights13 package page](https://nycflights13.tidyverse.org/) and [chapter 5 from the textbook](https://r4ds.had.co.nz/transform.html) for more info.
  + **You can not attempt to *install* packages in CodeGrade.**
* Round all float/dbl values to two decimal places.
  + If your rounding does not work the way you expect, convert the tibble to a dataframe by using as.data.frame()
* All statistics should be run with variables in the order I state
  + E.g., “Run a regression predicting mileage from mpg, make, and type” would be:

lm(mileage ~ mpg + make + type...)

**Before attempting to answer these, or if you lose points upon an attempt, please review all CodeGrade information provided in the CodeGrade overview submodule - if you do not you are likely to lose points.**

1. What is the mean distance of flights for each of the carriers AA, EV, and FL?

* This answer should be assigned to Q1.

1. For the month with the highest number of flights, what is that value? Hint: use head(1).

* This answer should be assigned to Q2.

1. Find the five shortest minimum distances, called min\_dist, by origin/destination combination.

* This answer should be assigned to Q3 and appear like this:

|  |
| --- |
| origin dest min\_dist  <chr> <chr> <dbl> 1 [value] [value] [value] 2 [value] [value] [value] 3 [value] [value] [value] 4 [value] [value] [value] 5 [value] [value] [value] |

1. What five days of the year had the highest mean distance when leaving from JFK? Sort in descending order.

* The answer should be assigned to Q4 and appear like this:

|  |
| --- |
| month day mean\_distance 1 [value] [value] [value] 2 [value] [value] [value] 3 [value] [value] [value] 4 [value] [value] [value] 5 [value] [value] [value] |

1. Calculate the maximum arrival delay for flights to Boston and Atlanta, separately.

* The answer should be assigned to Q5 and appear like this:

|  |
| --- |
| dest max\_arr\_delay  <chr> <dbl> 1 [value] [value] 2 [value] [value] |
|  |

**fastfood**

For this assignment, you must name your R file fastfood.R

* For all questions you should load tidyverse and openintro.  You should not need to use any other libraries.
  + If the openintro package is not installed, you’ll need to do a *one-time installation* from the Console Window in RStudio like this:  
    install.packages("openintro")
* Load libraries with

suppressPackageStartupMessages(library(tidyverse))

suppressPackageStartupMessages(library(openintro))

* Round all float/dbl values to two decimal places.
* All statistics should be run with variables in the order I state
  + E.g., “Run a regression predicting mileage from mpg, make, and type” would be:

lm(mileage ~ mpg + make + type...)

* Once submitted to CodeGrade you will see some are hidden tests.  You will not see the results until the end of the term.  This is done to make sure students do not attempt to come up with workarounds to solutions.

**Before attempting to answer these, or if you lose points upon an attempt, please review all CodeGrade information provided in the CodeGrade overview submodule - if you do not you are likely to lose points.**

To access the fastfood data, run the following:

|  |
| --- |
| fastfood <- openintro::fastfood |

1. Looking only at Burger King and Chick-Fil-A, which item has the highest calories?  The answer should be saved as Q1.

1. What is the mean sugar amount for all items from Subway?  Save the value as Q2.

1. What is the mean value of calories for all items from Taco Bell?  Save the value as Q3.

1. Create a variable equal to total\_fat x sugar called fatXsugar.  Produce a tibble that has the restaurant, item, and fatXsugar for the top 3 items, from highest to lowest.

* Your answer should be in a 3 X 3 tibble called Q4 with formatting like this:

|  |
| --- |
| # A tibble: 1 x 1   restaurant item   fatXsugar   <chr>     <chr>      <dbl> 1 [name]    [name]   [value]  2 [name]    [name]   [value]  3 [name]    [name]   [value] |

1. How many restaurants have an average saturated fat over 10?

* Your answer should be one integer stored in a variable called Q5.

**pizza**

For this assignment, name your R file pizza.R

* For all questions you should load tidyverse.  You should not need to use any other libraries.
* Load tidyverse with suppressPackageStartupMessages(library(tidyverse))
* Download the pizza.csv file from Brightspace and place it in the same folder/directory as your script file. Then in RStudio, set your Working Directory to your Source File location:  
  Graphical user interface, application

  Description automatically generated
* Load the pizza.csv file like this:  
  pizza <- read\_csv('pizza.csv')
* Round all float/dbl values to two decimal places.
  + If your rounding does not work the way you expect, convert the tibble to a dataframe by using as.data.frame()
* All statistics should be run with variables in the order I state
  + E.g., “Run a regression predicting mileage from mpg, make, and type” would be:

lm(mileage ~ mpg + make + type...)

* In each of these you must use at least two dplyr functions.  You may use Google to look up how to do certain aspects.

**Before attempting to answer these, or if you lose points upon an attempt, please review all CodeGrade information provided in the CodeGrade overview submodule - if you do not you are likely to lose points.**

1. Create a dataframe containing driver names of instances where free\_wine = 1, discount\_customer = 1, and the order contained more than 4 pizzas.  (There will be repeated names).

* The answers should look like the following:

|  |
| --- |
| 1 [value] 2 [value] 3 [value] 4 [value] 5 [value]     6 [value]  7 [value]  8 [value]  9 [value] |

* If your CodeGrade output is <fct> instead of <chr>, you can use as.character(driver) to convert it
* Assign that to Q1

1. Create a variable that is the ratio of bill to pizza, called ratio.  What is the mean of that value (call the value mean\_ratio)?

* Assign this to Q2

1. For each day of the week, what is the variance in pizzas?

* The created values should be called var\_pizzas.
* The answer should be assigned to Q3 and show look like the following:

|  |
| --- |
| 1 Friday          [value] 2 Monday          [value] 3 Saturday        [value] 4 Sunday          [value] 5 Thursday        [value] 6 Tuesday         [value] 7 Wednesday       [value] |

1. Which operator had the higher average bill?

* The answer should be assigned to Q4.

1. What was the highest amount of free wine given by day/driver combination?  (For instance, Friday Bruno was 13, while Wednesday Salvator was 12)

* The answer should be assigned to Q5 and look like the following:

|  |
| --- |
| # A tibble: 1 x 3 # Groups:   day, driver [1]   day     driver     n   <chr>   <chr>  <int> 1 [day]   [name] [value] |

Depending on how you do this, you might need to convert a <dbl> to <int>.   You can convert a variable using as.integer().

**msleep**

For this assignment, you must name your R file msleep.R

* For all questions you should load tidyverse.  You should not need to use any other libraries.
* Load tidyverse with suppressPackageStartupMessages(library(tidyverse))
* The actual data set is called msleep
* Round all float/dbl values to two decimal places.
  + If your rounding does not work the way you expect, convert the tibble to a dataframe by using as.data.frame()
* All statistics should be run with variables in the order I state
  + E.g., “Run a regression predicting mileage from mpg, make, and type” would be:

lm(mileage ~ mpg + make + type...)

**Before attempting to answer these, or if you lose points upon an attempt, please review all CodeGrade information provided in the CodeGrade overview submodule - if you do not you are likely to lose points.**

1. What is the variance in total sleep for carni vores and those whose conservation status is lc?

* Your answer should be in a 1 X 1 data frame with a value called ‘var’ with formatting like this:

|  |
| --- |
| var       1 [value] |

1. What rodent (order Rodentia) has the highest total sleep/rem sleep ratio?

* Your answer should be in a 1 X 1 tibble assigned to Q2 with formatting like this:

|  |
| --- |
| # A tibble: 1 x 1   name    <chr>       1 [name] |

1. How many primates have a bodyweight/brainwt ratio greater than 100?

1. Create two new variables, the mean of sleep total and variance of sleep total, grouped by conservation, and removing missing values.  The names of the variables should correspond to those in the example below.

* The answer should be assigned to Q4:

|  |
| --- |
| conservation mean\_sleep var\_sleep  1 cd                [value]   [value]  2 domesticated      [value]   [value]  3 en                [value]   [value]  4 lc                [value]   [value]  5 nt                [value]   [value]  6 vu                [value]   [value] |

1. Provide the name(s) of all the domesticated herbivores that sleep more than 12 hours.

* The answer should be assigned to Q5