````
---
title: "R code for Data Science for Beginners"
subtitle: " Day 4: Individual Exercise"
author: "Benjamin Cook"  #If multiple, 'c("A", "B")'
date: "2024-9-12"  #r Sys.Date()
output:
  pdf_document: default
  html_document: default
---

# Clean up your workspace
```{r}
rm(list=ls(all=TRUE)) # remove all the named objects visible in the environment
cat("\014") # clean your console
```

## 1. Let's do more exercises with dplyr (with a different dataset)

Please download the nycflights13 data by installing this package called `nycflights13`

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

### 1-1: Please find all March flights in the data (the dataset is named "flights")
flights

WRITE YOUR ANSWER (code) HERE  > library(nycflights13)
>
>
> march_flights <- flights %>%
+     filter(month == 3)
>
>
> print(head(march_flights))
# A tibble: 6 × 19
   year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time arr_delay
carrier
  <int> <int> <int>    <int>          <int>     <dbl>    <int>          <int>     <dbl>
<chr>
1  2013     3     1        4           2159       125      318             56       142 B6
2  2013     3     1       50           2358        52      526            438        48 B6
3  2013     3     1      117           2245       152      223           2354       149 B6
4  2013     3     1      454            500        -6      633            648       -15 US
5  2013     3     1      505            515       -10      746            810       -24 UA
6  2013     3     1      521            530        -9      813            827       -14 UA
# ℹ 9 more variables: flight <int>, tailnum <chr>, origin <chr>, dest <chr>, air_time
<dbl>,
#   distance <dbl>, hour <dbl>, minute <dbl>, time_hour <dttm>

### 1-2 :Create a new variable as date with a format like this 1/1/2013, using the
`mutate()` function

WRITE YOUR ANSWER (code) HERE  > flights_with_date <- flights %>%
+     mutate(date = paste(month, day, year, sep = "/"))
>
> print(head(flights_with_date))

### 1-3: Change column name tailnum to tail_number

WRITE YOUR ANSWER (code) HERE  > flights_renamed <- flights %>%
+     rename(tail_number = tailnum)
>
````

```
> print(head(flights_renamed))

### 1-4: Group flights by their origins

WRITE YOUR ANSWER (code) HERE  > flights_count_by_origin <- flights %>%
+     group_by(origin) %>%
+     summarize(count = n())
>
> print(flights_count_by_origin)

### 1-5: Count how many flights departing from JFK on 2013-12-31?

WRITE YOUR ANSWER (code) HERE  > jfk_flights_dec_31 <- flights %>%
+     filter(origin == "JFK", year == 2013, month == 12, day == 31) %>%
+     summarize(count = n())
>
> print(jfk_flights_dec_31)

### 1-6: Calculate the average hours of delay in departure for all flights from JFK

WRITE YOUR ANSWER (code) HERE  > avg_dep_delay_jfk <- flights %>%
+     filter(origin == "JFK") %>%
+     summarize(avg_delay_hours = mean(dep_delay, na.rm = TRUE) / 60)
>
> print(avg_dep_delay_jfk)

Finally, execute the entire contents of this file. Make sure that you don't get any error
message. If you get an error message, it's probably because you forgot to comment out
something.
```