

Homework 3

Prepare your answers as a **single PDF file**.

Group work: You may work in groups of 1-3. Include all group member names in the PDF file. You may work with students in both sections (375-01, -02). Only one person in the group should submit to Canvas.

Worked on it by myself

Name: Serop Kelkelian

Due: check on Canvas.

Load the `nycflights13` library (will have to install the `nycflights13` package first) which contains flight arrival and departure data in a table called `flights`. Apply the tidyverse's data wrangling verbs to answer these questions. For each question, **give only the code (as one data pipeline with multiple "verbs" one after the other) beginning with `flights %>%` ...**

1. List data only for flights that departed on March 12, 2013.
2. List data only for flights that were delayed (both arrival and departure) by more than 2 hours.
3. List data only for flights that were delayed (either arrival or departure) by more than 2 hours.
4. List data only for flights that were operated by United, American, or Delta.
5. Sort data in order of fastest flights (`air_time`).
6. Sort data in order of longest duration flights (`air_time`).
7. Show only the origin and destination of flights sorted by longest flights.
8. Add a new variable that indicates the total delay (both departure and arrival delay).
9. Show only the origin and destination of flights sorted by descending order of total delay.
10. Show only the origin and destination of 10 most delayed flights [Hint: there are multiple ways of solving this. Some additional functions that you will find useful are `head()`, `slice()`, `min_rank()`.]

Question 1

List data only for flights that departed on March 12, 2013.

```
flights %>% filter(month==3, day==12)
```

Question 2

List data only for flights that were delayed (both arrival and departure) by more than 2 hours.

```
flights %>% filter(dep_delay>120, arr_delay>120)
```

or

```
flights %>% filter(dep_delay>120 & arr_delay>120)
```

Question 3

```
# List data only for flights that were delayed (either arrival or departure) by more than 2 hours.
flights %>% filter(dep_delay>120 | arr_delay>120)
```

```
# Question 4
```

```
# List data only for flights that were operated by United, American, or Delta.
```

```
flights %>% filter(carrier=="AA" | carrier=="DL" | carrier=="UA")
```

```
# Question 5
```

```
# Sort data in order of fastest flights (air_time).
```

```
flights %>% arrange(air_time)
```

```
# Question 6
```

```
# Sort data in order of longest duration flights (air_time).
```

```
flights %>% arrange(desc(air_time))
```

```
# Question 7
```

```
# Show only the origin and destination of flights sorted by longest flights.
```

```
flights %>% arrange(desc(air_time)) %>% select(origin, dest)
```

```
# Question 8
```

```
# Add a new variable that indicates the total delay (both departure and arrival delay).
```

```
flights %>% mutate(total_delay = dep_delay+arr_delay) %>% View()
```

```
# Question 9
```

```
# Show only the origin and destination of flights sorted by descending order of total delay.
```

```
flights %>% mutate(total_delay = dep_delay+arr_delay) %>% arrange(desc(total_delay)) %>%
select(origin, dest) %>% View()
```

```
# Question 10
```

```
# Show only the origin and destination of 10 most delayed flights [Hint: there are multiple
# ways of solving this. Some additional functions that you will find useful are head(), slice(),
min_rank().]
```

```
flights %>% mutate(total_delay = dep_delay+arr_delay) %>% arrange(desc(total_delay)) %>%
select(origin, dest) %>% head(10)
```