Jake Gadaleta | Homework 2

5.2.4

Find all flights that

All answers also available in Script.R

1. Had an arrival delay of two or more hours

```
flights %>%
filter(arr_delay > (60 * 2))
```

2. Flew to Houston (IAH or HOU)

```
flights %>%
filter(dest == 'IAH' | dest == 'HOU')
```

3. Were operated by United, American, or Delta

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

4. Departed in summer (July, August, and September)

```
flights %>%
filter(month >= 7 & month <= 9)
```

5. Arrived more than two hours late, but didn't leave late

```
flights %>%
filter( arr_delay > (60 * 2) & dep_delay <= 0)
```

6. Were delayed by at least an hour, but made up over 30 minutes in flight

```
flights %>%
filter(dep_delay >= 60 & dep_delay - arr_delay > 30)
```

1. Departed between midnight and 6am (inclusive)

```
flights %>% filter(dep_time < (6 * 60))
```

5.4.1

2. What happens if you include the name of a variable multiple times in a select() call?

```
A: The second call is automatically dropped
```

4. Does the result of running the following code surprise you? How do the select helpers deal with cases by default? How can you change that default?

```
select(flights, contains("TIME"))
```

```
It returned back all variables (headers, points whatever data scientist
need to decided on one name), that contained the time in any form, I was a
little shocked at the lack of case sensitivity.


Essentially the functions like contains (or helpers) let you do more while
writing less code which everyone and their mothers is a fan of.

You could also write the question as an excluder cause. Thus this now
returns everything except for what is in time
```

```
    select(flights, -contains("TIME"))
```

5.5.2

2. Compare air_time with arr_time - dep_time. What do you expect to see? What do you see? What do you need to do to fix it?

```
flights %>%
mutate(true_air = arr_time - dep_time) %>%
select(air_time, true_air)
```

```
You would expect the 2 numbers to the same but they are not, I will blame
time zones for us
```

3. Compare dep_time, sched_dep_time, and dep_delay. How would you expect those three numbers to be related?

```
flights %>%
select(sched_dep_time, dep_delay, dep_time,)
```

```
the scheduled time + delay = dep_time, as one would expect
```

5.6.7

5. Which carrier has the worst delays?

```
flights %>%
group_by(carrier) %>%
mutate(total_delay = dep_delay - arr_delay) %>%
summarise(mean_delay=mean(total_delay,na.rm=TRUE)) %>%
arrange(desc(mean_delay))
```

```
AS, 15.8 mean_delay
```

5.7.1

3. What time of day should you fly if you want to avoid delays as much as possible?

'''r flights %>% mutate(total_delay = dep_delay - arr_delay) %>% mutate(approx_hour = floor(sched_dep_time / 100)) %>% group_by(approx_hour) %>% summarise(mean_delay=mean(total_delay,na.rm=TRUE)) %>% arrange(mean_delay) %>% select(approx_hour, mean_delay) '''

```
After making my own of grouping by hours in the second mutate function, I
decided to sort by then check vrs those hours and I found that the
aproxmate best time to leave is in the 23 hour (11pm) with average total
delays being approx 2.27 minutes
```