

62070501064 อรวิภา คุณเจริญไพศาล CPE

## LAB 1 CPE231 Data Models

### R- Script

```
flights <- read.csv("flights.csv")
airlines <- read.csv("airlines.csv")

library(tidyr)
library(dplyr)
flights %>%
  select(carrier, dep_delay) %>%
  filter(!is.na(dep_delay)) %>%
  group_by(carrier) %>%
  summarise(mean_delay = mean(dep_delay)) %>%
  arrange(desc(mean_delay)) %>%
  left_join(select(airlines, -X), flights, by = "carrier")
```

### Result

1.

Console	Terminal x	Jobs x
D:/OUM/112-2/CPE213 Data Models/wk 2/lab/ ↗		
<pre>&gt; flights &lt;- read.csv("flights.csv") &gt; airlines &lt;- read.csv("airlines.csv") &gt;  </pre>		

2.

```
> flights %>%
+   select(carrier, dep_delay)
  carrier dep_delay
1      UA         2
2      UA         4
3      AA         2
4      B6        -1
5      DL        -6
6      UA        -4
7      B6        -5
8      EV        -3
9      B6        -3
10     AA        -2
```

3.

```
+   filter(!is.na(dep_delay))
  carrier dep_delay
1      UA         2
2      UA         4
3      AA         2
4      B6        -1
5      DL        -6
6      UA        -4
7      B6        -5
8      EV        -3
9      B6        -3
10     AA        -2
```

4.

```
+ group_by(carrier) %>%
+ summarise(mean_delay = mean(dep_delay))
# A tibble: 16 x 2
  carrier mean_delay
*   <chr>         <dbl>
1  9E             16.7
2  AA              8.59
3  AS              5.80
4  B6             13.0
5  DL              9.26
6  EV             20.0
7  F9             20.2
8  FL             18.7
9  HA              4.90
10 MQ             10.6
```

5.

```
+ arrange(desc(mean_delay))
# A tibble: 16 x 2
  carrier mean_delay
  <chr>         <dbl>
1  F9             20.2
2  EV             20.0
3  YV             19.0
4  FL             18.7
5  WN             17.7
6  9E             16.7
7  B6             13.0
8  VX             12.9
9  OO             12.6
10 UA             12.1
11 MQ             10.6
12 DL              9.26
13 AA              8.59
14 AS              5.80
15 HA              4.90
16 US              3.78
```

6.

```
+ left_join(select(airlines, -X), flights, by = "carrier")
# A tibble: 16 x 3
  carrier mean_delay name
  <chr>         <dbl> <chr>
1  F9             20.2 Frontier Airlines Inc.
2  EV             20.0 ExpressJet Airlines Inc.
3  YV             19.0 Mesa Airlines Inc.
4  FL             18.7 AirTran Airways Corporation
5  WN             17.7 Southwest Airlines Co.
6  9E             16.7 Endeavor Air Inc.
7  B6             13.0 JetBlue Airways
8  VX             12.9 Virgin America
9  OO             12.6 SkyWest Airlines Inc.
10 UA             12.1 United Air Lines Inc.
11 MQ             10.6 Envoy Air
12 DL              9.26 Delta Air Lines Inc.
13 AA              8.59 American Airlines Inc.
14 AS              5.80 Alaska Airlines Inc.
15 HA              4.90 Hawaiian Airlines Inc.
16 US              3.78 US Airways Inc.
```