Natthanan Bhukan 61070507206

King Mongkut's University of Technology Thonburi

Faculty of Engineering, Department of Computer Engineering

CPE 213 Data Models, 2/2020

LAB Lecture 2: R Tutorial

Assign Date: 29 Jan 2021, Due Date: 4 Feb 2021

---

1) Load two data frames "flights.csv" and "airlines.csv"
   - Hint: read.csv

Load flights.csv and airline.csv to dataframe

```
[75] df_flights <- read.csv("flights.csv")
     df_airlines <- read.csv("airlines.csv")
```

```
[76] df_flights
```

A data.frame: 336776 × 20

X	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight	tailnum	origin	dest	air_time	distance
<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<int>	<int>
1	2013	1	1	517	515	2	830	819	11	UA	1545	N14228	EWR	IAH	227	1400
2	2013	1	1	533	529	4	850	830	20	UA	1714	N24211	LGA	IAH	227	1416
3	2013	1	1	542	540	2	923	850	33	AA	1141	N619AA	JFK	MIA	160	1089
4	2013	1	1	544	545	-1	1004	1022	-18	B6	725	N804JB	JFK	BQN	183	1576
5	2013	1	1	554	600	-6	812	837	-25	DL	461	N668DN	LGA	ATL	116	762
6	2013	1	1	554	558	-4	740	728	12	UA	1696	N39463	EWR	ORD	150	719
7	2013	1	1	555	600	-5	913	854	19	B6	507	N516JB	EWR	FLL	158	1065
8	2013	1	1	557	600	-3	709	723	-14	EV	5708	N829AS	LGA	IAD	53	229

```
[77] df_airlines
```

A data.frame: 16 × 3

X	carrier	name
<int>	<chr>	<chr>
1	9E	Endeavor Air Inc.
2	AA	American Airlines Inc.
3	AS	Alaska Airlines Inc.
4	B6	JetBlue Airways
5	DL	Delta Air Lines Inc.
6	EV	ExpressJet Airlines Inc.
7	F9	Frontier Airlines Inc.
8	FL	AirTran Airways Corporation
9	HA	Hawaiian Airlines Inc.
10	MQ	Envoy Air
11	OO	SkyWest Airlines Inc.
12	UA	United Air Lines Inc.
13	US	US Airways Inc.
14	VX	Virgin America
15	WN	Southwest Airlines Co.
16	YV	Mesa Airlines Inc.

King Mongkut's University of Technology Thonburi

Faculty of Engineering, Department of Computer Engineering

CPE 213 Data Models, 2/2020

LAB Lecture 2: R Tutorial

Assign Date: 29 Jan 2021, Due Date: 4 Feb 2021

---

2) Select two columns (carrier, dep_delay)

▾ **Select two columns (carrier, dep_delay)**

```
[78] str(df_flights)
```

```
'data.frame':   336776 obs. of  20 variables:
 $ X             : int  1 2 3 4 5 6 7 8 9 10 ...
 $ year          : int  2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 ...
 $ month         : int  1 1 1 1 1 1 1 1 1 1 ...
 $ day           : int  1 1 1 1 1 1 1 1 1 1 ...
 $ dep_time      : int  517 533 542 544 554 554 555 557 557 558 ...
 $ sched_dep_time: int  515 529 540 545 600 558 600 600 600 600 ...
 $ dep_delay     : int  2 4 2 -1 -6 -4 -5 -3 -3 -2 ...
 $ arr_time      : int  830 850 923 1004 812 740 913 709 838 753 ...
 $ sched_arr_time: int  819 830 850 1022 837 728 854 723 846 745 ...
 $ arr_delay     : int  11 20 33 -18 -25 12 19 -14 -8 8 ...
 $ carrier       : chr  "UA" "UA" "AA" "B6" ...
 $ flight        : int  1545 1714 1141 725 461 1696 507 5708 79 301 ...
 $ tailnum       : chr  "N14228" "N24211" "N619AA" "N804JB" ...
 $ origin        : chr  "EWR" "LGA" "JFK" "JFK" ...
 $ dest          : chr  "IAH" "IAH" "MIA" "BQN" ...
 $ air_time      : int  227 227 160 183 116 150 158 53 140 138 ...
 $ distance      : int  1400 1416 1089 1576 762 719 1065 229 944 733 ...
 $ hour          : int  5 5 5 5 6 5 6 6 6 6 ...
 $ minute        : int  15 29 40 45 0 58 0 0 0 0 ...
 $ time_hour     : chr  "2013-01-01 05:00:00" "2013-01-01 05:00:00" "2013-01-01 05:00:00" "2013-01-01 05:00:00" ...
```

```
[79] df_flights_select <- df_flights %>%
                          select(carrier, dep_delay)
```

```
[80] df_flights_select
```

A data.frame:
336776 × 2

carrier <chr>	dep_delay <int>
UA	2
UA	4
AA	2
B6	-1
DL	-6
UA	-4
B6	-5
EV	-3
B6	-3
AA	-2
B6	-2
B6	-2
UA	-2
UA	-2
AA	-1
B6	0
UA	-1
B6	0
MQ	0
B6	1
DL	-8

King Mongkut's University of Technology Thonburi

Faculty of Engineering, Department of Computer Engineering

CPE 213 Data Models, 2/2020

LAB Lecture 2: R Tutorial

Assign Date: 29 Jan 2021, Due Date: 4 Feb 2021

---

3) Filter NA out of dep_delay

▾ **Filter NA out of dep_delay**

```
[81] df_flights_select_out_NA <- df_flights_select %>%
                                   drop_na(dep_delay)
```

```
[82] df_flights_select_out_NA
```

A data.frame: 328521 × 2

carrier	dep_delay
<chr>	<int>
UA	2
UA	4
AA	2
B6	-1
DL	-6
UA	-4
B6	-5
EV	-3
B6	-3
AA	-2
B6	-2
B6	-2
UA	-2
UA	-2
AA	-1
B6	0
UA	-1
B6	0
MQ	0
B6	1

4) Use carrier as a group and calculate mean of dep_delay

▾ **Use carrier as a group and calculate mean of dep_dealy**

```
[83] df_mean_delay_of_each_carrier <- df_flights_select_out_NA %>%
                                        group_by(carrier) %>%
                                        summarise(mean_delay = mean(dep_delay))
```

```
[84] df_mean_delay_of_each_carrier
```

A tibble: 16 × 2

	carrier	mean_delay
	<chr>	<dbl>
1	9E	16.725769
2	AA	8.586016
3	AS	5.804775
4	B6	13.022522
5	DL	9.264505
6	EV	19.955390
7	F9	20.215543
8	FL	18.726075
9	HA	4.900585
10	MQ	10.552041
11	OO	12.586207
12	UA	12.106073
13	US	3.782418
14	VX	12.869421
15	WN	17.711744
16	YV	18.996330

King Mongkut's University of Technology Thonburi

Faculty of Engineering, Department of Computer Engineering

CPE 213 Data Models, 2/2020

LAB Lecture 2: R Tutorial

Assign Date: 29 Jan 2021, Due Date: 4 Feb 2021

---

5) Sort the worst carrier by departure delay (dep_delay)

▾ Sort the wort carrier by departure delay (dep_deplay)

```
[85] df_mean_delay_of_each_carrier_sort <- df_mean_delay_of_each_carrier %>%
                                           arrange(desc(mean_delay))
```

```
[86] df_mean_delay_of_each_carrier_sort
```

A tibble: 16 × 2

carrier	mean_delay
<chr>	<dbl>
F9	20.215543
EV	19.955390
YV	18.996330
FL	18.726075
WN	17.711744
9E	16.725769
B6	13.022522
VX	12.869421
OO	12.586207
UA	12.106073
MQ	10.552041
DL	9.264505
AA	8.586016
AS	5.804775
HA	4.900585
US	3.782418

6) Join airlines data to flight data using 'carrier' as a key

▾ Join airlines data to flight data using "carrier" as a key

```
[87] df_mean_delay_of_each_carrier_sort %>%
       left_join(df_airlines, by="carrier") %>%
       select(carrier, mean_delay, name)
```

A tibble: 16 × 3

carrier	mean_delay	name
<chr>	<dbl>	<chr>
F9	20.215543	Frontier Airlines Inc.
EV	19.955390	ExpressJet Airlines Inc.
YV	18.996330	Mesa Airlines Inc.
FL	18.726075	AirTran Airways Corporation
WN	17.711744	Southwest Airlines Co.
9E	16.725769	Endeavor Air Inc.
B6	13.022522	JetBlue Airways
VX	12.869421	Virgin America
OO	12.586207	SkyWest Airlines Inc.
UA	12.106073	United Air Lines Inc.
MQ	10.552041	Envoy Air
DL	9.264505	Delta Air Lines Inc.
AA	8.586016	American Airlines Inc.
AS	5.804775	Alaska Airlines Inc.
HA	4.900585	Hawaiian Airlines Inc.
US	3.782418	US Airways Inc.

Natthanan Bhukan 61070507206

King Mongkut's University of Technology Thonburi

Faculty of Engineering, Department of Computer Engineering

CPE 213 Data Models, 2/2020

LAB Lecture 2: R Tutorial

Assign Date: 29 Jan 2021, Due Date: 4 Feb 2021

Compare result and expected result

Result

A tibble: 16 × 3

carrier	mean_delay	name
<chr>	<dbl>	<chr>
F9	20.215543	Frontier Airlines Inc.
EV	19.955390	ExpressJet Airlines Inc.
YV	18.996330	Mesa Airlines Inc.
FL	18.726075	AirTran Airways Corporation
WN	17.711744	Southwest Airlines Co.
9E	16.725769	Endeavor Air Inc.
B6	13.022522	JetBlue Airways
VX	12.869421	Virgin America
OO	12.586207	SkyWest Airlines Inc.
UA	12.106073	United Air Lines Inc.
MQ	10.552041	Envoy Air
DL	9.264505	Delta Air Lines Inc.
AA	8.586016	American Airlines Inc.
AS	5.804775	Alaska Airlines Inc.
HA	4.900585	Hawaiian Airlines Inc.
US	3.782418	US Airways Inc.

Expected result

```
# A tibble: 16 x 3
   carrier mean_delay                      name
     <chr>      <dbl>                     <chr>
 1      F9  20.215543      Frontier Airlines Inc.
 2      EV  19.955390    ExpressJet Airlines Inc.
 3      YV  18.996330          Mesa Airlines Inc.
 4      FL  18.726075 AirTran Airways Corporation
 5      WN  17.711744      Southwest Airlines Co.
 6      9E  16.725769           Endeavor Air Inc.
 7      B6  13.022522             JetBlue Airways
 8      VX  12.869421              Virgin America
 9      OO  12.586207       SkyWest Airlines Inc.
10      UA  12.106073       United Air Lines Inc.
11      MQ  10.552041                   Envoy Air
12      DL   9.264505       Delta Air Lines Inc.
13      AA   8.586016      American Airlines Inc.
14      AS   5.804775       Alaska Airlines Inc.
15      HA   4.900585      Hawaiian Airlines Inc.
16      US   3.782418            US Airways Inc.
```