

STATS 506 Problem Set #4

Haiming Li

Tidyverse

```
library(tidyverse)
library(nycflights13)
```

a. First table

```
flights %>%
  group_by(origin) %>%
  summarise(mean_delay=mean(arr_delay, na.rm=TRUE),
            median_delay=median(arr_delay, na.rm=TRUE),
            n_flights=n()) %>%
  ungroup() %>%
  filter(n_flights >= 10) %>%
  left_join(airports, by=join_by(origin == faa)) %>%
  select(name, mean_delay, median_delay) %>%
  arrange(desc(mean_delay))
```

```
# A tibble: 3 x 3
  name                mean_delay median_delay
  <chr>                <dbl>         <dbl>
1 Newark Liberty Intl    9.11           -4
2 La Guardia            5.78           -5
3 John F Kennedy Intl    5.55           -6
```

Second table

```

flights %>%
  group_by(dest) %>%
  summarise(mean_delay=mean(arr_delay, na.rm=TRUE),
            median_delay=median(arr_delay, na.rm=TRUE),
            n_flights=n()) %>%
  ungroup() %>%
  filter(n_flights >= 10) %>%
  left_join(airports, by=join_by(dest == faa)) %>%
  mutate(name = coalesce(name, dest)) %>%
  select(name, mean_delay, median_delay) %>%
  arrange(desc(mean_delay)) %>%
  print(n=102)

```

A tibble: 102 x 3

	name <chr>	mean_delay <dbl>	median_delay <dbl>
1	"Columbia Metropolitan"	41.8	28
2	"Tulsa Intl"	33.7	14
3	"Will Rogers World"	30.6	16
4	"Jackson Hole Airport"	28.1	15
5	"Mc Ghee Tyson"	24.1	2
6	"Dane Co Rgnl Truax Fld"	20.2	1
7	"Richmond Intl"	20.1	1
8	"Akron Canton Regional Airport"	19.7	3
9	"Des Moines Intl"	19.0	0
10	"Gerald R Ford Intl"	18.2	1
11	"Birmingham Intl"	16.9	-2
12	"Theodore Francis Green State"	16.2	1
13	"Greenville-Spartanburg International"	15.9	-0.5
14	"Cincinnati Northern Kentucky Intl"	15.4	-3
15	"Savannah Hilton Head Intl"	15.1	-1
16	"Manchester Regional Airport"	14.8	-3
17	"Eppley Afld"	14.7	-2
18	"Yeager"	14.7	-1.5
19	"Kansas City Intl"	14.5	0
20	"Albany Intl"	14.4	-4
21	"General Mitchell Intl"	14.2	0
22	"Piedmont Triad"	14.1	-2
23	"Washington Dulles Intl"	13.9	-3
24	"Cherry Capital Airport"	13.0	-10
25	"James M Cox Dayton Intl"	12.7	-3
26	"Louisville International Airport"	12.7	-2

27	"Chicago Midway Intl"	12.4	-1
28	"Sacramento Intl"	12.1	4
29	"Jacksonville Intl"	11.8	-2
30	"Nashville Intl"	11.8	-2
31	"Portland Intl Jetport"	11.7	-4
32	"Greater Rochester Intl"	11.6	-5
33	"Hartsfield Jackson Atlanta Intl"	11.3	-1
34	"Lambert St Louis Intl"	11.1	-3
35	"Norfolk Intl"	10.9	-4
36	"Baltimore Washington Intl"	10.7	-5
37	"Memphis Intl"	10.6	-2.5
38	"Port Columbus Intl"	10.6	-3
39	"Charleston Afb Intl"	10.6	-4
40	"Philadelphia Intl"	10.1	-3
41	"Raleigh Durham Intl"	10.1	-3
42	"Indianapolis Intl"	9.94	-3
43	"Charlottesville-Albemarle"	9.5	-5
44	"Cleveland Hopkins Intl"	9.18	-5
45	"Ronald Reagan Washington Natl"	9.07	-2
46	"Burlington Intl"	8.95	-4
47	"Buffalo Niagara Intl"	8.95	-5
48	"Syracuse Hancock Intl"	8.90	-5
49	"Denver Intl"	8.61	-2
50	"Palm Beach Intl"	8.56	-3
51	"BQN"	8.25	-1
52	"Bob Hope"	8.18	-3
53	"Fort Lauderdale Hollywood Intl"	8.08	-3
54	"Bangor Intl"	8.03	-9
55	"Asheville Regional Airport"	8.00	-1
56	"PSE"	7.87	0
57	"Pittsburgh Intl"	7.68	-5
58	"Gallatin Field"	7.6	-2
59	"NW Arkansas Regional"	7.47	-2
60	"Tampa Intl"	7.41	-4
61	"Charlotte Douglas Intl"	7.36	-3
62	"Minneapolis St Paul Intl"	7.27	-5
63	"William P Hobby"	7.18	-4
64	"Bradley Intl"	7.05	-10
65	"San Antonio Intl"	6.95	-9
66	"South Bend Rgnl"	6.5	-3.5
67	"Louis Armstrong New Orleans Intl"	6.49	-6
68	"Key West Intl"	6.35	7
69	"Eagle Co Rgnl"	6.30	-4

70	"Austin Bergstrom Intl"	6.02	-5
71	"Chicago Ohare Intl"	5.88	-8
72	"Orlando Intl"	5.45	-5
73	"Detroit Metro Wayne Co"	5.43	-7
74	"Portland Intl"	5.14	-5
75	"Nantucket Mem"	4.85	-3
76	"Wilmington Intl"	4.64	-7
77	"Myrtle Beach Intl"	4.60	-13
78	"Albuquerque International Sunport"	4.38	-5.5
79	"George Bush Intercontinental"	4.24	-5
80	"Norman Y Mineta San Jose Intl"	3.45	-7
81	"Southwest Florida Intl"	3.24	-5
82	"San Diego Intl"	3.14	-5
83	"Sarasota Bradenton Intl"	3.08	-5
84	"Metropolitan Oakland Intl"	3.08	-9
85	"General Edward Lawrence Logan Intl"	2.91	-9
86	"San Francisco Intl"	2.67	-8
87	"SJU"	2.52	-6
88	"Yampa Valley"	2.14	2
89	"Phoenix Sky Harbor Intl"	2.10	-6
90	"Montrose Regional Airport"	1.79	-10.5
91	"Los Angeles Intl"	0.547	-7
92	"Dallas Fort Worth Intl"	0.322	-9
93	"Miami Intl"	0.299	-9
94	"Mc Carran Intl"	0.258	-8
95	"Salt Lake City Intl"	0.176	-8
96	"Long Beach"	-0.0620	-10
97	"Martha\\'s Vineyard"	-0.286	-11
98	"Seattle Tacoma Intl"	-1.10	-11
99	"Honolulu Intl"	-1.37	-7
100	"STT"	-3.84	-9
101	"John Wayne Arpt Orange Co"	-7.87	-11
102	"Palm Springs Intl"	-12.7	-13.5

b. Here's the table

```
flights %>%
  left_join(planes, by = "tailnum") %>%
  mutate(mph=60*distance/air_time) %>%
  group_by(model) %>%
  summarize(avg_mph = mean(mph, na.rm = TRUE),
            n_flights = n()) %>%
```

```
arrange(desc(avg_mph)) %>%  
slice_head(n=1)
```

```
# A tibble: 1 x 3  
  model    avg_mph n_flights  
  <chr>    <dbl>    <int>  
1 777-222    483.         4
```