R Practice - ggplot2 and NYC Flights

See if you can make some plots from the NYC Flights data set
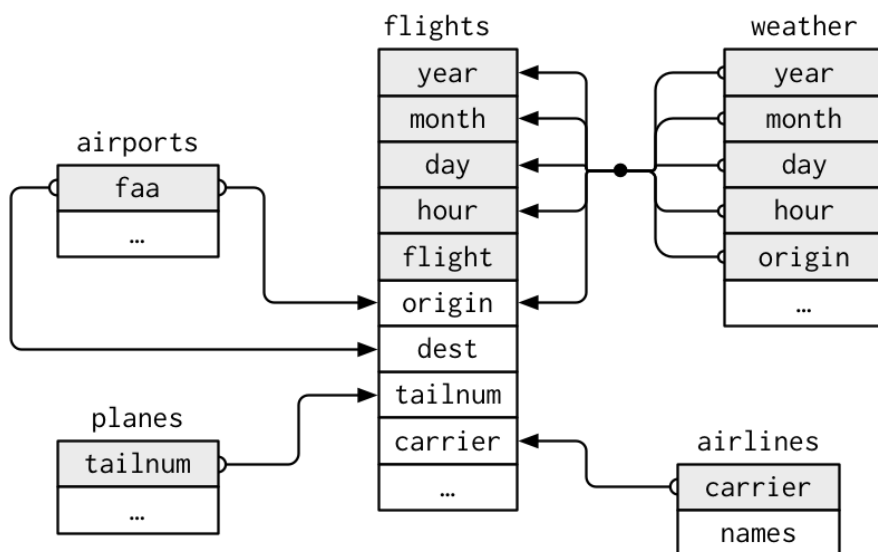
Accessing the data

Make sure you load the data by loading the nycflights13 library

```
library(nycflights13)
```

And also be sure to load ggplot2 and dplyr

```
library(ggplot2)
library(dplyr)
```

Recall that the library includes the following tables

Since this data set is quite large, let's works with just a subset of the data. Create a table for just the data from January

```
jan <- flights %>% filter(month==1)
```

Use ggplot2 to make some plots that address these questions

Remember to only use the data from January for these questions

1) Show how many flights took place each hour

2) Is there a relationship between the scheduled hour of the flight and the distance flown?

3) Plot the arrival and departure delays for each flight colored by carrier for the top three carriers B6 (JetBlue), UA (United Airlines), EV (ExpressJet)

4) Compare the distributions of departure delays from the three origins (only look at delays less than 120 minutes). Try overlaying the estimated densities in a single plot. Now try using facets.

5) First let's make a map of the USA. You may need to install the following packages

```
#install.packages(c("maps","mapproj"))
ggplot(map_data("state"), aes(long, lat)) +
    geom_polygon(aes(group=group), colour="white") +
    coord_map("albers",  at0 = 45.5, lat1 = 29.5)
```

This should make a nice map. Note how it uses long and latitude for the coordinates. The "airports" table contains the longitude and latitude for each of the airports in the dataset. Add a layer to the above plot that places a circle on each airport with a size relative to the number of flights to that location

6) Create a plot of your own design!

Possible Answers

There are no "right" answers to these questions. You may interpret the question differently or have different approaches and that's OK.

1)

```
ggplot(jan, aes(hour)) + geom_bar()
```

2)

```
ggplot(jan, aes(factor(hour), distance)) + geom_boxplot()
ggplot(jan, aes(distance, hour)) + geom_hex()
```

3)

```
ggplot(jan %>% filter(carrier %in% c("B6","UA","EV")), aes(dep_delay, arr_delay, color=carrier)) + geom_point()
```

4)

```
ggplot(jan %>% filter(dep_delay<120), aes(dep_delay)) + geom_density(aes(color=origin), alpha=.2
```

```
gplot(jan %>% filter(dep_delay<120), aes(dep_delay)) + geom_density() + facet_wrap(~origin, ncol=1)
```

5)

```
dests <- jan %>% count(dest) %>%
        inner_join(airports, c("dest"="faa")) %>% rename(long=lon)
ggplot(map_data("state"), aes(long, lat)) +
        geom_polygon(aes(group=group), color="white") +
        geom_point(data=dests, aes(color=n, size=n)) +
        coord_map("albers",  at0 = 45.5, lat1 = 29.5)
```