Assignment 2
Due **February 5^{th}, 2018**

Total 44 points.

Please submit a well-commented R script, documenting all code used in this problem set, along with a write up answering all questions below. Use figures as appropriate to support your answers, and when required by the problem.

Data: In this problem set we will use the `nycflights13` R package that contains data for all flights that departed NYC (i.e. JFK, LGA or EWR) in 2013. You can load this data as follows:

```
install.packages("nycflights13") # install package if you have not
already
library(nycflights13)            # load library
data(flights)                    # load data on flights
```

1. Let's explore flights from NYC to Seattle. Use the flights dataset to answer the following questions.

a) (2 pt) How many flights were there to and from NYC in 2013?

b) (2 pt) How many flights were there from NYC airports to Seattle (SEA) in 2013?

c) (2 pt) How many airlines fly from NYC to Seattle? Hint: look at the function `unique()`

d) (2 pt) What is the average arrival delay for flights from NYC to Seattle? Hint: if there are missing values in the vector you want the mean of, you have to remove them. Try the following:

```
mean(c(1,2,3,NA))
```

and

```
mean(c(1,2,3,NA), na.rm=T)
```

2. Flights are often delayed. Consider the following questions exploring delay patterns.

a) (4 pt) What is the mean arrival delay time? What is the median arrival delay time?

b) (2 pt) What does a negative arrival delay mean?

c) (4 pt) Plot a histogram of arrival delay times. Does the answers you obtained in (a) consistent with the shape of the delay time distribution?

d) (4 pt) Is there seasonality in departure delays? Try

```
by(flights$dep_delay, flights$month, function(x) mean(x, na.rm=T))
```

and describe what patterns you see. Is there a best month to leave New York? A worst? Why might this be?

3. EDA

a) (4 pt) Plot a histogram of the total air flight time with 100 breaks. (look at the help for `hist()`). How many peaks do you see in this distribution? What is an explanation for this?

b) (4 pt) What time of day do flights most commonly depart? Why might there be two most popular times of day to depart?

c) (4 pt) Plot a box plot of departure delays and hour of departure. What pattern do you see? What is an explanation for this?

4. (10 pt) Develop one research question you can address using the nycflights2013 dataset. Provide two visualizations to support your exploration of this question. Discuss what you find.