MIS2502: Data Analytics
Assignment: Getting Familiar with R/RStudio

For this assignment you'll be doing a simple analysis on a data set by modifying the R script you used in the last In-Class Exercise (Descriptives.r).

The data set – OnTimeAirport-Jan14.csv –contains actual data regarding on-time flight statistics for 84,656 flights, by airline and airport, for January 2014.

The metadata for the spreadsheet is below:

Variable Name	Variable Description
FlightDate	The date of the flight (mm/dd/yyyy)
UniqueCarrier	The unique carrier code
AirlineFullName	The full name of the airline
AirlineID	The numeric ID of the airline
Origin	The origin airport of the flight
OriginCityName	The origin city of the flight
DestCityName	The destination city of the flight
DepDelayMinutes	The delay in departing from the origin gate
TaxiOut	The minutes spent taxiing out to the runway at origin
TaxiIn	The minutes spent taxiing in from the runway at destination
ArrDelayMinutes	The delay in arrive to the destination gate
Cancelled	Whether the flight was cancelled (0 = no, 1 = yes)
Distance	The total distance of the flight

To complete the assignment, modify the Descriptives.r script to perform an analysis of departure delays by origin airport:

1) Use OnTimeAirport-Jan14.csv as the input file.
2) Present the number of flights, grouped by origin airport.
3) Present summary statistics for departure delay, grouped by origin airport.
4) Compare, using a t-test, the departure delays from Philadelphia (PHL) versus Los Angeles (LAX).
5) Create a histogram, properly labeled, of the overall distribution of departure delays for all flights.

These first five things can be done by simply modifying the existing script. Of course, you'll need to understand what the script does in order for you to successfully modify it.

Once you've completed this part, add several new lines to the script that does the following: (NOTE: Make sure you add these lines right before the sink() function so that the results are included in your text file output.)

6) Use describeBy() to compare the cancellation rates across origin airports.
7) Use describeBy() to compare the cancellation rates across airlines.
8) Answer this question using a t-test: Do planes spend more time taxiing to the runway in Phoenix (PHX) or Chicago (ORD)?

What to submit:

Send a single email to your instructor with the following attachments:

- The completed, working R script that produced the analysis in items 1 through 8.
- The output files – descriptivesOutput.txt and histogram.pdf.
- The completed worksheet provided on the next page.

Answer Sheet for Assignment: Getting Familiar with R/RStudio

Name _____

Answer the questions below based on your script output

Question		Answer
1	How many total flights (including cancelled flights) came out of the Philadelphia airport (PHL) during January 2014?	
2	What was the average departure delay for the Pittsburgh airport (PIT) during January 2014?	
3	What was the longest departure delay for La Guardia airport (LGA) during January 2014?	
4	On average, which airport experienced greater departure delays: Philadelphia's airport (PHL) or Los Angeles' airport (LAX)?	
5	For question #4, was this difference statistically significant? What is the p-value? (answer both questions in the blank to the right)	
6	Which airport(s) had the highest cancelled flight percentage? (you can list more than one if it's a tie)	
7	Which airport(s) had the lowest cancelled fight percentage? (you can list more than one if it's a tie)	
8	Which airline(s) had the highest cancelled flight percentage? (you can list more than one if it's a tie)	
9	Which airline(s) had the lowest cancelled fight percentage? (you can list more than one if it's a tie)	
10	On average, which airport experienced greater taxi out times: Chicago's airport (ORD) or Phoenix' airport (PHX)?	
11	For question #10, was this difference statistically significant? What is the p-value? (answer both questions in the blank to the right)	
12	Looking at the histogram, are most flights delayed less than 30 minutes or more than 30 minutes?	