Assignment 3-1

File Name: Assign31_LastName.doc

Using R, please conduct the following tasks. Create your report in a Word document, and upload it to Blackboard. In your report, provide the source code of R, the screen capture of the output, and some descriptions of the outputs.

Instructions:

Install "dplyr" and "hflights" packages. These packages enable use of the functions and data for this assignment. The "dplyr" package provides some useful functions, including filter(), select(), mutate(), arrage(), and summaries().

Using the database that contains information about flights at "Geroge Bush Intercontinental Airport." This dataset is included in the package named "hflights" and contains 227,496 flights that departed from the airport in Houston in 2011.

1) Explore the highlights of the data that display information about the data. The output may look like

```
'data.frame':   227496 obs. of  21 variables:
 $ Year              : int  2011 2011 2011 2011 2011 2011 2011 2011 2011 2011 ...
 $ Month             : int  1 1 1 1 1 1 1 1 1 1 ...
 $ DayofMonth        : int  1 2 3 4 5 6 7 8 9 10 ...
 $ DayOfWeek         : int  6 7 1 2 3 4 5 6 7 1 ...
 $ DepTime           : int  1400 1401 1352 1403 1405 1359 1359 1355 1443 1443 ...
 $ ArrTime           : int  1500 1501 1502 1513 1507 1503 1509 1454 1554 1553 ...
 $ UniqueCarrier     : chr  "AA" "AA" "AA" "AA" ...
```

Also, plyr can work with data frames as is, but if you're dealing with large data, it's worthwhile to convert them to a tbl_df: this is a wrapper around a data frame that won't accidentally print a lot of data to the screen (for example, t_hflights <- tbl_df (hflights). The output looks like

```
# A tibble: 227,496 x 21
    Year Month DayofMonth DayOfWeek DepTime ArrTime UniqueCarrier FlightNum TailNum ActualElapsedTime
*  <int> <int>      <int>     <int>   <int>   <int>         <chr>     <int>   <chr>             <int>
1   2011     1          1         6    1400    1500            AA       428  N576AA                60
2   2011     1          2         7    1401    1501            AA       428  N557AA                60
3   2011     1          3         1    1352    1502            AA       428  N541AA                70
4   2011     1          4         2    1403    1513            AA       428  N403AA                70
5   2011     1          5         3    1405    1507            AA       428  N492AA                62
6   2011     1          6         4    1359    1503            AA       428  N262AA                64
7   2011     1          7         5    1359    1509            AA       428  N493AA                70
8   2011     1          8         6    1355    1454            AA       428  N477AA                59
9   2011     1          9         7    1443    1554            AA       428  N476AA                71
10  2011     1         10         1    1443    1553            AA       428  N504AA                70
# ... with 227,486 more rows, and 11 more variables: AirTime <int>, ArrDelay <int>, DepDelay <int>,
#   Origin <chr>, Dest <chr>, Distance <int>, TaxiIn <int>, TaxiOut <int>, Cancelled <int>,
#   CancellationCode <chr>, Diverted <int>
```

2) Display the data on the first day of January for all years in the dataset. The output may look like the following:

	Year	Month	DayofMonth	DayOfWeek	DepTime	ArrTime	UniqueCarrier	FlightNum	TailNum	ActualElapsedTime
1	2011	1	1	6	1400	1500	AA	428	N576AA	60
2	2011	1	1	6	728	840	AA	460	N520AA	72
3	2011	1	1	6	1631	1736	AA	1121	N4WVAA	65
4	2011	1	1	6	1756	2112	AA	1294	N3DGAA	136
5	2011	1	1	6	1012	1347	AA	1700	N3DAAA	155
6	2011	1	1	6	1211	1325	AA	1820	N593AA	74
7	2011	1	1	6	557	906	AA	1994	N3BBAA	129
8	2011	1	1	6	1824	2106	AS	731	N614AS	282
9	2011	1	1	6	654	1124	B6	620	N324JB	210
10	2011	1	1	6	1639	2110	B6	622	N324JB	211
11	2011	1	1	6	942	1356	CO	1	N69063	494
12	2011	1	1	6	1845	1947	CO	5	N29717	62
13	2011	1	1	6	1533	1634	CO	6	N47414	61
14	2011	1	1	6	1459	1602	CO	33	N62631	63

TIP: The function, filter(), is useful for processing the data.

3) Arrange your data in order of ArrDealy, Month, and Year using the arrange( ) function. For example,

	Year	Month	DayofMonth	DayOfWeek	DepTime	ArrTime	UniqueCarrier	FlightNum	TailNum	ActualElapsedTime
1	2011	7	3	7	1914	2039	XE	2804	N12157	85
2	2011	12	25	7	741	926	OO	4591	N814SK	165
3	2011	8	21	7	935	1039	OO	2001	N767SK	184
4	2011	8	31	3	934	1039	OO	2040	N783SK	185
5	2011	8	26	5	2107	2205	OO	2003	N713SK	178
6	2011	12	24	6	2129	2337	CO	1552	N37437	248

4) Using the select() function, display the observations with only “Year,” “Month,” and “DayofWeek.”

5) Split the complete dataset into individual planes and then summarize each plane by counting the number of flights (count = n()) and computing the average distance (dist =

mean(Distance, na.rm = TRUE)) and arrival delay (delay = mean(ArrDelay, na.rm = TRUE)). The output looks like

```
##       Year          Month          DayofMonth      DayOfWeek
##  Min.   :2011   Min.   : 1.000   Min.   : 1.00   Min.   :1.000
##  1st Qu.:2011   1st Qu.: 4.000   1st Qu.: 8.00   1st Qu.:2.000
##  Median :2011   Median : 7.000   Median :16.00   Median :4.000
##  Mean   :2011   Mean   : 6.514   Mean   :15.74   Mean   :3.948
##  3rd Qu.:2011   3rd Qu.: 9.000   3rd Qu.:23.00   3rd Qu.:6.000
##  Max.   :2011   Max.   :12.000   Max.   :31.00   Max.   :7.000
##
##     DepTime        ArrTime     UniqueCarrier        FlightNum
##  Min.   :   1   Min.   :   1   Length:227496      Min.   :   1
##  1st Qu.:1021   1st Qu.:1215   Class :character   1st Qu.: 855
##  Median :1416   Median :1617   Mode  :character   Median :1696
##  Mean   :1396   Mean   :1578                      Mean   :1962
##  3rd Qu.:1801   3rd Qu.:1953                      3rd Qu.:2755
##  Max.   :2400   Max.   :2400                      Max.   :7290
##  NA's   :2905   NA's   :3066
##    TailNum          ActualElapsedTime    AirTime         ArrDelay
##  Length:227496      Min.   : 34.0     Min.   : 11.0   Min.   :-70.000
##  Class :character   1st Qu.: 77.0     1st Qu.: 58.0   1st Qu.: -8.000
##  Mode  :character   Median :128.0     Median :107.0   Median :  0.000
##                     Mean   :129.3     Mean   :108.1   Mean   :  7.094
##                     3rd Qu.:165.0     3rd Qu.:141.0   3rd Qu.: 11.000
##                     Max.   :575.0     Max.   :549.0   Max.   :978.000
##                     NA's   :3622      NA's   :3622    NA's   :3622
##     DepDelay          Origin             Dest              Distance
##  Min.   :-33.000   Length:227496      Length:227496      Min.   :  79.0
##  1st Qu.: -3.000   Class :character   Class :character   1st Qu.: 376.0
##  Median :  0.000   Mode  :character   Mode  :character   Median : 809.0
##  Mean   :  9.445                                         Mean   : 787.8
##  3rd Qu.:  9.000                                         3rd Qu.:1042.0
##  Max.   :981.000                                         Max.   :3904.0
##  NA's   :2905
```