Lab Session 10 STATS 220 Data Technologies Due 12:00pm Monday 30 May

The point of this lab is to get more practice writing R code and specifically to practice subsetting and writing for loops and working with data structures in R.

We will work with twelve CSV files that contain data on public transport use in Auckland. The file for 2015 is shown in Figure 1.

The data are in files called:
`patronage-2005.csv`,
`patronage-2006.csv`,
`patronage-2007.csv`,
`patronage-2008.csv`,
`patronage-2009.csv`,
`patronage-2010.csv`,
`patronage-2011.csv`,
`patronage-2012.csv`,
`patronage-2013.csv`,
`patronage-2014.csv`,
`patronage-2015.csv`, and
`patronage-2016.csv`.

There are links to these files on the STATS 220 web site and a zip file containing all files at once is also available.

```
"Month","Total","BusTotal","BusRapid","BusOther","Train","Ferry"
"Jan",5328.8,3872.1,184.1,3688,863.6,593.1
"Feb",6683,4917.3,226.7,4690.6,1209.9,555.8
"Mar",8394.8,6282.6,300.3,5982.3,1564.8,547.4
"Apr",6286.2,4674.8,232.6,4442.2,1134.5,476.9
"May",7311.8,5535.1,269,5266.1,1344.3,432.4
"Jun",6743.1,5100,245.7,4854.3,1265.5,377.6
"Jul",6748.6,5019.8,313.9,4705.9,1328.6,400.2
"Aug",7276.5,5453.6,326.7,5126.9,1419.4,403.5
"Sep",6985.2,5218.3,314,4904.3,1362.3,404.5
"Oct",7082,5254.8,338.7,4916.1,1359.6,467.6
"Nov",6900.6,5037.1,347,4690.1,1377.4,486.1
"Dec",5836.6,4111.6,272,3839.6,1149.8,575.2
```

Figure 1: Patronage of public transport in Auckland (2015).

1. Write R code to read in the first CSV file, `"patronage-2015.csv"`, and assign the resulting data frame to the symbol `pat2015`.

   The symbol `pat2015` should print like this:

```
   Month  Total BusTotal BusRapid BusOther  Train Ferry
1    Jan 5328.8   3872.1    184.1   3688.0  863.6 593.1
2    Feb 6683.0   4917.3    226.7   4690.6 1209.9 555.8
3    Mar 8394.8   6282.6    300.3   5982.3 1564.8 547.4
4    Apr 6286.2   4674.8    232.6   4442.2 1134.5 476.9
5    May 7311.8   5535.1    269.0   5266.1 1344.3 432.4
6    Jun 6743.1   5100.0    245.7   4854.3 1265.5 377.6
7    Jul 6748.6   5019.8    313.9   4705.9 1328.6 400.2
8    Aug 7276.5   5453.6    326.7   5126.9 1419.4 403.5
9    Sep 6985.2   5218.3    314.0   4904.3 1362.3 404.5
10   Oct 7082.0   5254.8    338.7   4916.1 1359.6 467.6
11   Nov 6900.6   5037.1    347.0   4690.1 1377.4 486.1
12   Dec 5836.6   4111.6    272.0   3839.6 1149.8 575.2
```

2. Write R code to determine in which months the `BusTotal` patronage was more than 10 times the `Ferry` patronage and assign the result to the symbol `bigBus`.

   The symbol `bigBus` should print like this:

```
 [1] FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
```

3. Write R code to subset the `BusTotal` and `Ferry` columns from the `pat2015` data frame for the months when the `BusTotal` patronage was more than 10 times the `Ferry` patronage and assign the result to the symbol `bigMonths`.

   The symbol `bigMonths` should print like this:

```
   BusTotal Ferry
3    6282.6 547.4
5    5535.1 432.4
6    5100.0 377.6
7    5019.8 400.2
8    5453.6 403.5
9    5218.3 404.5
10   5254.8 467.6
11   5037.1 486.1
```

4. Write a for loop that writes out a message for each file reporting how many months `BusTotal` patronage was more than 10 times the `Ferry` patronage.

   The output of your code should look like this:

```
patronage-2005.csv contains 5 months in which bus patronage was more than 10 times Ferry patronage
patronage-2006.csv contains 8 months in which bus patronage was more than 10 times Ferry patronage
patronage-2007.csv contains 8 months in which bus patronage was more than 10 times Ferry patronage
patronage-2008.csv contains 8 months in which bus patronage was more than 10 times Ferry patronage
patronage-2009.csv contains 8 months in which bus patronage was more than 10 times Ferry patronage
patronage-2010.csv contains 8 months in which bus patronage was more than 10 times Ferry patronage
patronage-2011.csv contains 8 months in which bus patronage was more than 10 times Ferry patronage
patronage-2012.csv contains 9 months in which bus patronage was more than 10 times Ferry patronage
patronage-2013.csv contains 9 months in which bus patronage was more than 10 times Ferry patronage
patronage-2014.csv contains 9 months in which bus patronage was more than 10 times Ferry patronage
patronage-2015.csv contains 8 months in which bus patronage was more than 10 times Ferry patronage
patronage-2016.csv contains 0 months in which bus patronage was more than 10 times Ferry patronage
```

   Remember that it is a good idea to start simple and build up complexity *and* it is a good idea to print out values within the loop so that you can see what is happening.

[EXTRA for EXPERTS - NO MARKS]

Create a matrix that records `TRUE` or `FALSE` for each month and for each year, based on whether `BusTotal` patronage was more than 10 times the `Ferry` patronage.

HINT: this is pretty nasty and the course content from the final week might help. Doing STATS 380 next semester would help even more :)

```
       Jan   Feb   Mar   Apr  May  Jun  Jul  Aug  Sep   Oct  Nov   Dec
2005    NA    NA    NA    NA   NA   NA TRUE TRUE TRUE  TRUE TRUE FALSE
2006 FALSE FALSE  TRUE FALSE TRUE TRUE TRUE TRUE TRUE  TRUE TRUE FALSE
2007 FALSE FALSE  TRUE FALSE TRUE TRUE TRUE TRUE TRUE  TRUE TRUE FALSE
2008 FALSE FALSE FALSE  TRUE TRUE TRUE TRUE TRUE TRUE  TRUE TRUE FALSE
2009 FALSE FALSE  TRUE  TRUE TRUE TRUE TRUE TRUE TRUE FALSE TRUE FALSE
2010 FALSE FALSE  TRUE FALSE TRUE TRUE TRUE TRUE TRUE  TRUE TRUE FALSE
2011 FALSE FALSE  TRUE  TRUE TRUE TRUE TRUE TRUE TRUE FALSE TRUE FALSE
2012 FALSE  TRUE  TRUE FALSE TRUE TRUE TRUE TRUE TRUE  TRUE TRUE FALSE
2013 FALSE FALSE  TRUE  TRUE TRUE TRUE TRUE TRUE TRUE  TRUE TRUE FALSE
2014 FALSE FALSE  TRUE  TRUE TRUE TRUE TRUE TRUE TRUE  TRUE TRUE FALSE
2015 FALSE FALSE  TRUE FALSE TRUE TRUE TRUE TRUE TRUE  TRUE TRUE FALSE
2016 FALSE    NA    NA    NA   NA   NA   NA   NA   NA    NA   NA    NA
```

NOTE: You should submit a file containing R code that creates the appropriate objects. I will run the code in your file and then check the value of the objects.

NOTE: Your file should ONLY contain valid R code, properly **indented**, and with **comments**. You should be able to copy-and-paste your entire file of R code into R and get no errors.

NOTE: You should submit your answers via the submission form in the "Submissions" section of the STATS 220 web site.