Lab Session 9 STATS 220 Data Technologies Due 4:00pm Mon 26 May

The point of this lab is to get more practice writing R code and specifically to practice subsetting and writing for loops 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-2005.csv"`, and assign the resulting data frame to the symbol `pat2005`.

   The symbol `pat2005` should print like this:

```
  Month  Total BusTotal BusRapid BusOther Train Ferry
1   Jul 4083.4   3433.7       NA       NA 365.8 284.0
2   Aug 4912.3   4226.7       NA       NA 427.0 258.7
3   Sep 4380.3   3752.7       NA       NA 397.4 230.1
4   Oct 4301.2   3617.7       NA       NA 363.1 320.3
5   Nov 4395.4   3633.2     14.2   3619.0 418.0 344.2
6   Dec 3295.0   2641.4     18.7   2622.7 315.8 337.9
```

2. Write R code to subset the `Total` column from the `pat2005` data frame and assign the result to the symbol `pat2005total`.

   The symbol `pat2005total` should print like this:

```
[1] 4083.4 4912.3 4380.3 4301.2 4395.4 3295.0
```

3. Write a conditional expression that compares the `Total` patronage in July 2005 with the `Total` patronage in August 2005 and assigns a value to the symbol `monthChange`: `"increased"` if August is bigger and `"decreased"` if July is bigger.

   The symbol `monthChange` should print like this:

```
[1] "increased"
```

4. Write R code to loop over the values in the `Total` column of the `pat2005` data frame and print a message to the screen saying whether the patronage has increased or decreased.

   Your code should produce output like this:

```
The patronage in Aug increased to 4912.3
The patronage in Sep decreased to 4380.3
The patronage in Oct decreased to 4301.2
The patronage in Nov increased to 4395.4
The patronage in Dec decreased to 3295
```

[EXTRA for EXPERTS - NO MARKS]

Write R code to determine the largest percentage increase in bus patronage between 2005 and 2015.

```
The largest percentage increase, from 3246.7 to 4624.1 (42%), occurred in Feb, 2012
```

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.