

## CT5102: Programing for Data Analytics 2018/19

### Assignment 7: Exploring energy data using dplyr

The aim of this assignment is to investigate the relationship between weather data and energy production. The datasets are from 2017, and record energy generation (Ireland), and wind speed data (Mace Head, Galway).

(1) Include the following libraries

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(lubridate)
library(readxl)
```

(2) Load in the energy dataset, contained in the CT5102 github account<sup>1</sup>.

```
ener <- read_excel("datasets/Energy/IrelandData January 2017.xlsx")
```

```
> ener
# A tibble: 2,784 x 8
  DateTime           Demand Generation Wind    CO2 NetImports EWIC Moyle
  <dtm>           <dbl>     <dbl> <dbl> <dbl>    <dbl> <dbl> <dbl>
1 2017-01-29 00:00:00   3834     4041   449   552    -145    -33   -112
2 2017-01-29 00:15:00   3785     4041   505   548    -200   -108    -92
3 2017-01-29 00:30:00   3708     4130   521   544    -294   -183   -111
4 2017-01-29 00:45:00   3634     4181   492   543    -419   -258   -161
5 2017-01-29 01:00:00   3581     4211   538   555    -503   -333   -170
6 2017-01-29 01:15:00   3552     4278   561   531    -598   -379   -219
7 2017-01-29 01:30:00   3491     4133   484   545    -516   -374   -142
8 2017-01-29 01:45:00   3435     4143   474   551    -581   -365   -216
9 2017-01-29 02:00:00   3374     4158   442   550    -653   -373   -280
10 2017-01-29 02:15:00  3329     4135   421   550    -676   -377   -299
# ... with 2,774 more rows
```

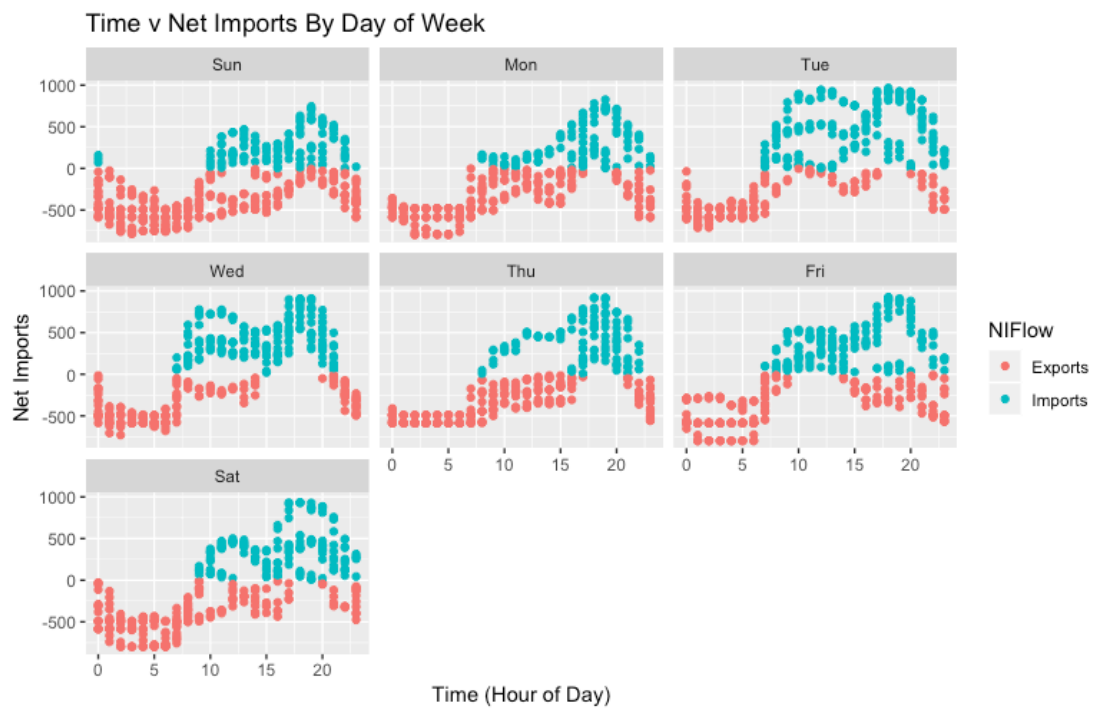
(3) Add new features to the energy data set, based on the original values, and make use of **dplyr** and **tidyr** functions. These include: Date (String), Time (String), HourOfDay, Minute of Day, DayOfWeek and NIFlow. NIFlow should be "Exporting" if the NetImports is negative, otherwise it should be "Importing."

```
# A tibble: 2,784 x 14
  DateTime           Date Time HourOfDay MinuteOfDay DayOfWeek NIFlow Demand
  <dtm>           <chr> <chr>    <int>      <int>    <ord>    <chr>    <dbl>
1 2017-01-29 00:00:00 2017... 00:0...      0          0 Sun      Expor...  3834
2 2017-01-29 00:15:00 2017... 00:1...      0         15 Sun      Expor...  3785
3 2017-01-29 00:30:00 2017... 00:3...      0         30 Sun      Expor...  3708
4 2017-01-29 00:45:00 2017... 00:4...      0         45 Sun      Expor...  3634
5 2017-01-29 01:00:00 2017... 01:0...      1          0 Sun      Expor...  3581
6 2017-01-29 01:15:00 2017... 01:1...      1         15 Sun      Expor...  3552
7 2017-01-29 01:30:00 2017... 01:3...      1         30 Sun      Expor...  3491
8 2017-01-29 01:45:00 2017... 01:4...      1         45 Sun      Expor...  3435
9 2017-01-29 02:00:00 2017... 02:0...      2          0 Sun      Expor...  3374
10 2017-01-29 02:15:00 2017... 02:1...      2         15 Sun      Expor...  3329
# ... with 2,774 more rows, and 6 more variables: Generation <dbl>, Wind <dbl>,
# CO2 <dbl>, NetImports <dbl>, EWIC <dbl>, Moyle <dbl>
```

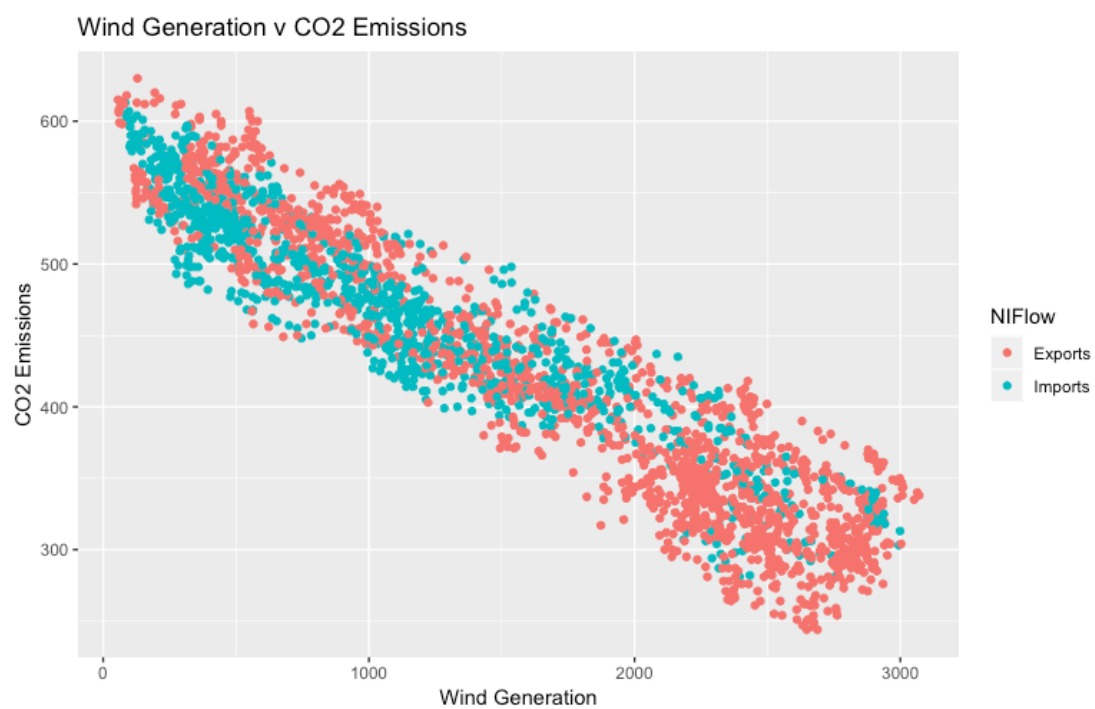
---

<sup>1</sup> <https://github.com/JimDuggan/CT5102>

(4) Plot the net import data over time, colour by NIFlow, and facet by day of the week.



(5) Plot the wind generation vs CO<sub>2</sub> Emissions



(6) Load in the weather data set.

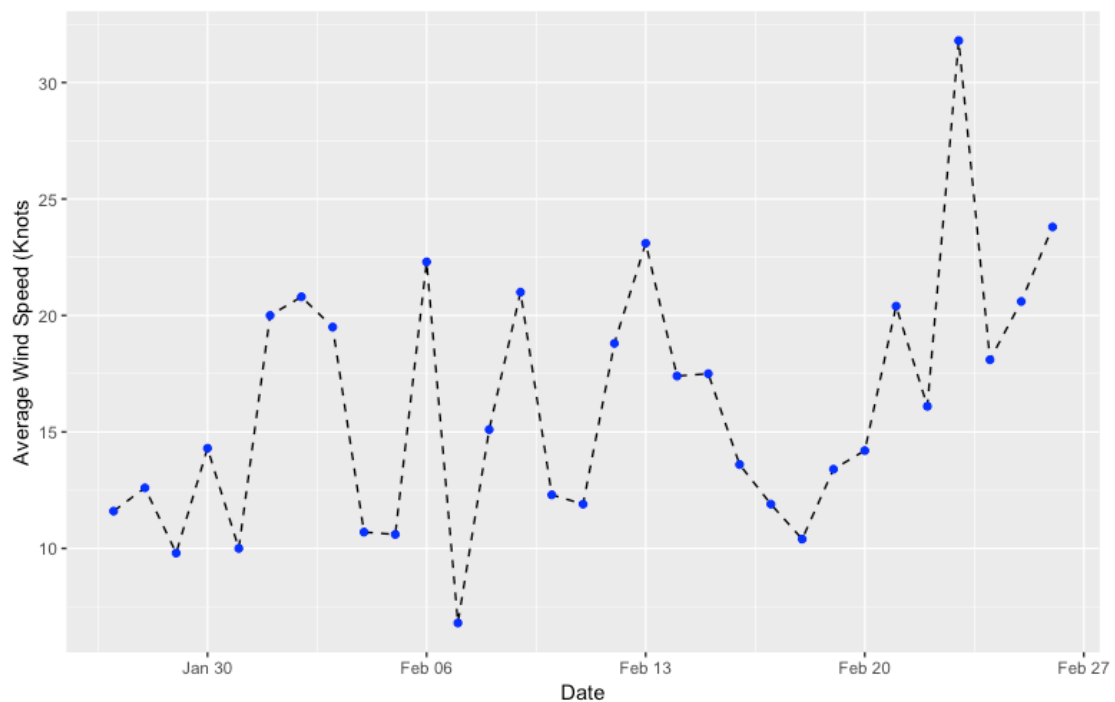
```
weather <- read_excel("datasets/Energy/Mac Head Wind Data.xlsx")
```

```
> weather
# A tibble: 31 x 7
  Date           Rainfall MaxTemp MinTemp GrassMinTemp AVRWind MaxWindGust
  <dtm>          <dbl>   <dbl>   <dbl>         <dbl>   <dbl>         <dbl>
1 2017-01-27 00:00:00    7.9     8.7     4.3          -0.7    11.6           NA
2 2017-01-28 00:00:00    3.5     8       4.5           2.9    12.6           NA
3 2017-01-29 00:00:00    4.7     9       4.9           3.7     9.8           NA
4 2017-01-30 00:00:00    7.8    11.2     7.1           5.8    14.3           NA
5 2017-01-31 00:00:00     0     10.3     7.3           5.8    10            NA
6 2017-02-01 00:00:00    0.6    10.2     6.1           5.2    20            38
7 2017-02-02 00:00:00    4.9    11.2     7.4           6.4    20.8          45
8 2017-02-03 00:00:00    2.2     8.5     3.6           2.1    19.5          46
9 2017-02-04 00:00:00    5.3     6.5     1.8          -1.3    10.7           NA
10 2017-02-05 00:00:00    6.9     7.5     2.2          -1.4    10.6           NA
# ... with 21 more rows
```

(7) Convert the Date (dtm) to (date) format.

```
# A tibble: 31 x 7
  Date           Rainfall MaxTemp MinTemp GrassMinTemp AVRWind MaxWindGust
  <date>          <dbl>   <dbl>   <dbl>         <dbl>   <dbl>         <dbl>
1 2017-01-27      7.9     8.7     4.3          -0.7    11.6           NA
2 2017-01-28      3.5     8       4.5           2.9    12.6           NA
3 2017-01-29      4.7     9       4.9           3.7     9.8           NA
4 2017-01-30      7.8    11.2     7.1           5.8    14.3           NA
5 2017-01-31      0     10.3     7.3           5.8    10            NA
6 2017-02-01      0.6    10.2     6.1           5.2    20            38
7 2017-02-02      4.9    11.2     7.4           6.4    20.8          45
8 2017-02-03      2.2     8.5     3.6           2.1    19.5          46
9 2017-02-04      5.3     6.5     1.8          -1.3    10.7           NA
10 2017-02-05      6.9     7.5     2.2          -1.4    10.6           NA
# ... with 21 more rows
```

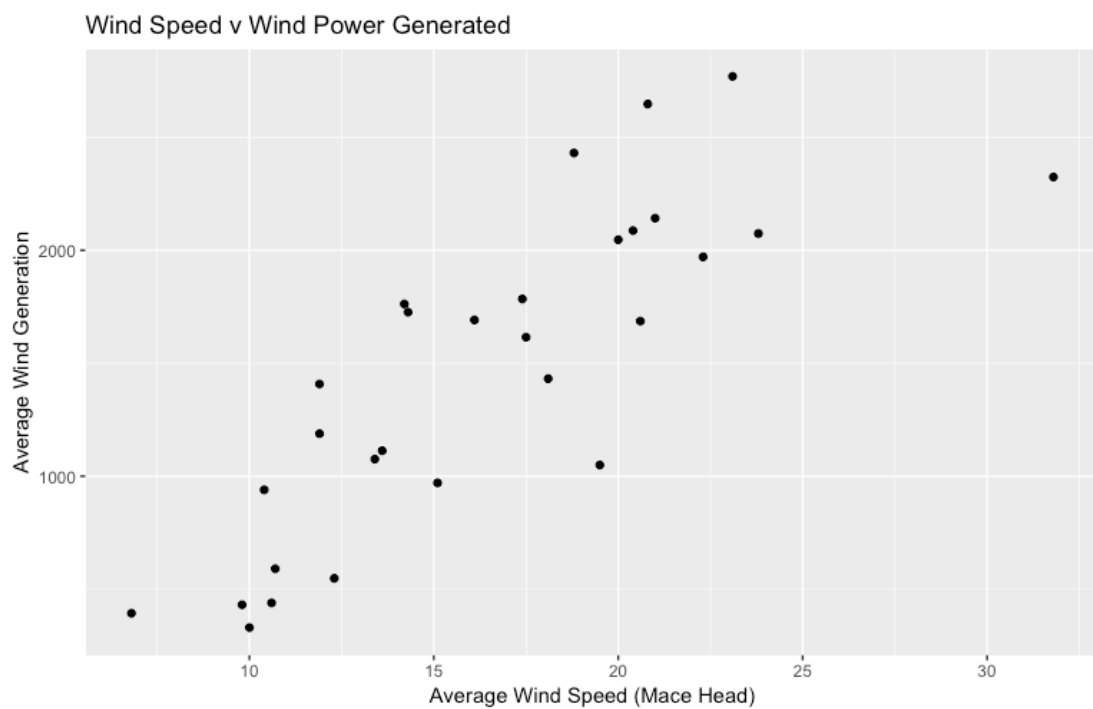
(8) Plot the average wind speed



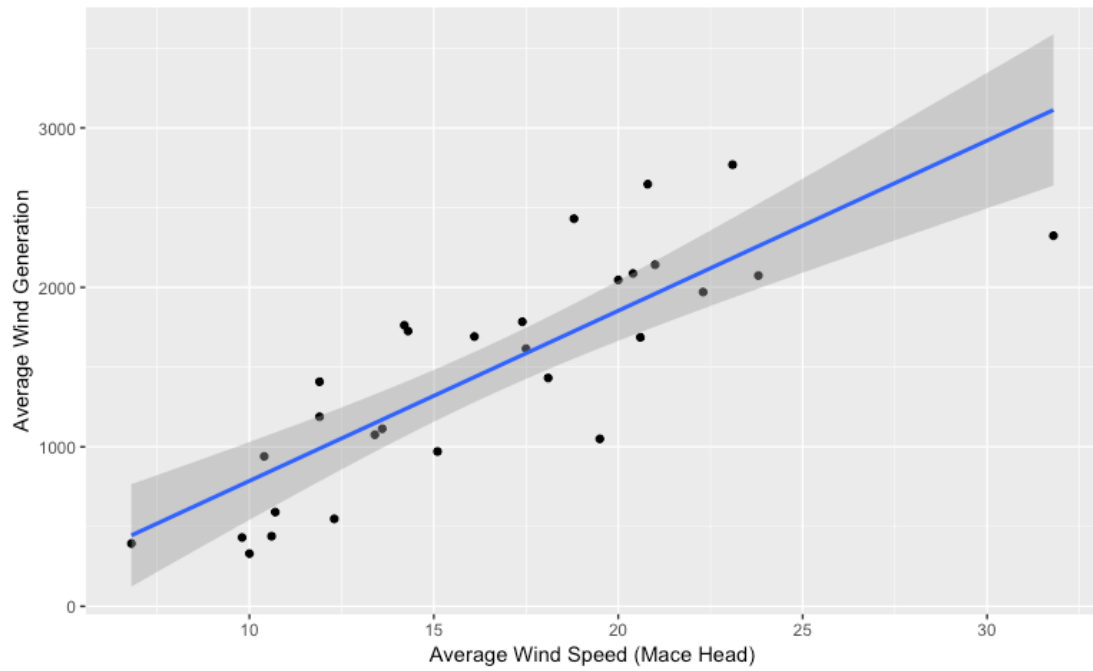
- (9) Generate the average daily wind generation from the energy data, and ensure that the Date column is of type <date>.

```
> avr_daily_wind
# A tibble: 29 x 2
  Date           AverageWindGeneration
  <date>              <dbl>
1 2017-01-29         431.
2 2017-01-30        1726.
3 2017-01-31         330.
4 2017-02-01        2047.
5 2017-02-02        2647.
6 2017-02-03        1050.
7 2017-02-04         591.
8 2017-02-05         439.
9 2017-02-06        1971.
10 2017-02-07         394.
```

- (10) Join the new datasets and produce the following plots:



Wind Speed v Wind Power Generated, with linear model



Wind Speed v Wind Power Generated, with loess model

