9. Linking Data and Linear Modelling

An Energy Case Study

Transmission System (2017)
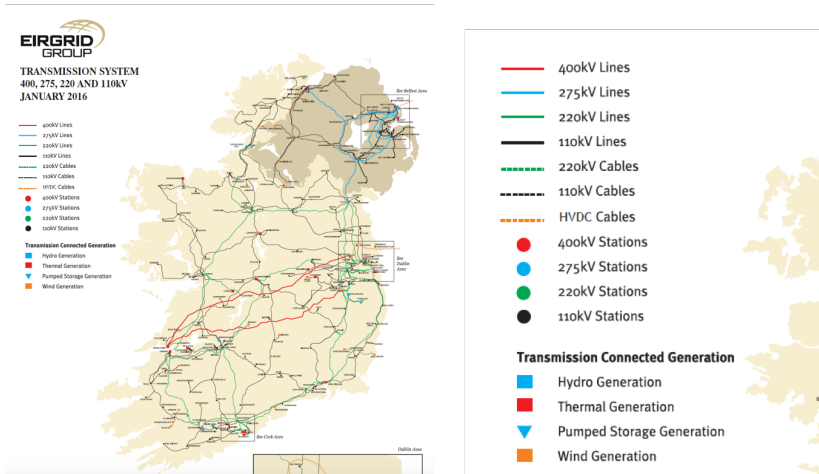

Figure 1: Electricity Transmission System

Generation Information
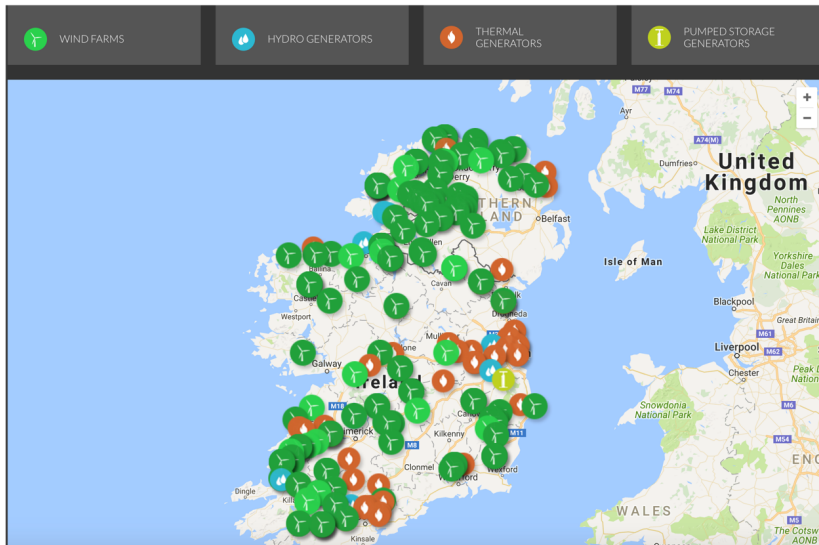

Figure 2: Generation Information

Sample Data

Actual System Generation

System Generation represents the total electricity production on the system, including system losses, but net of generators' requirements. System Generation is shown in 15 minute intervals.
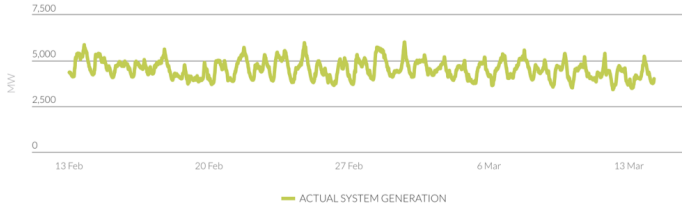
| DAY | WEEK | MONTH |

COMPARE WITH OTHER DATA ∨

<   Last 30 Days (13/02/2017 - 14/03/2017)   >

7,500

5,000

MW

2,500

0

13 Feb   20 Feb   27 Feb   6 Mar   13 Mar

— ACTUAL SYSTEM GENERATION

Figure 3: Generation Information

Accessing Data

```
ener <- read_excel("../../datasets/energy/IrelandData Janua
glimpse(ener)

## Observations: 2,784
## Variables: 8
## $ DateTime   <dttm> 2017-01-29 00:00:00, 2017-01-29 00:1
## $ Demand     <dbl> 3834, 3785, 3708, 3634, 3581, 3552, 3
## $ Generation <dbl> 4041, 4041, 4130, 4181, 4211, 4278, 4
## $ Wind       <dbl> 449, 505, 521, 492, 538, 561, 484, 47
## $ CO2        <dbl> 552, 548, 544, 543, 555, 531, 545, 55
## $ NetImports <dbl> -145, -200, -294, -419, -503, -598, -
## $ EWIC       <dbl> -33, -108, -183, -258, -333, -379, -3
## $ Moyle      <dbl> -112, -92, -111, -161, -170, -219, -1
```

Processing Dates - lubridate

Date component	Accessor
Year	`year()`
Month	`month()`
Week	`week()`
Day of year	`yday()`
Day of month	`mday()`
Day of week	`wday()`
Hour	`hour()`
Minute	`minute()`
Second	`second()`
Time zone	`tz()`

Extracting information

```
ener$DateTime[1]

## [1] "2017-01-29 UTC"

year(ener$DateTime[1])

## [1] 2017

wday(ener$DateTime[1])

## [1] 1
```

Adding New Columns

```r
ener <- ener %>% mutate(Date=ymd(DateTime),
                        HourOfDay=hour(DateTime),
                        MinuteOfDay=minute(DateTime),
                        DayOfWeek=wday(DateTime,label=T))
```

```
## Warning: All formats failed to parse. No formats found.
```

```r
glimpse(ener)
```

```
## Observations: 2,784
## Variables: 12
## $ DateTime   <dttm> 2017-01-29 00:00:00, 2017-01-29 00:
## $ Demand     <dbl> 3834, 3785, 3708, 3634, 3581, 3552,
## $ Generation <dbl> 4041, 4041, 4130, 4181, 4211, 4278,
## $ Wind       <dbl> 449, 505, 521, 492, 538, 561, 484, 4
## $ CO2        <dbl> 552, 548, 544, 543, 555, 531, 545, 5
## $ NetImports <dbl> -145, -200, -294, -419, -503, -598,
## $ EWIC       <dbl> -33, -108, -183, -258, -333, -379, -
## $ Moyle      <dbl> -112, -92, -111, -161, -170, -219, -
```

Split out date and time (need for join later)

```r
ener <- ener %>% separate(DateTime,c("Date","Time"),
                          sep=" ", remove=F) %>%
        mutate(Date=ymd(Date))
glimpse(ener)
```

```
## Observations: 2,784
## Variables: 13
## $ DateTime    <dttm> 2017-01-29 00:00:00, 2017-01-29 00:
## $ Date        <date> 2017-01-29, 2017-01-29, 2017-01-29,
## $ Time        <chr> "00:00:00", "00:15:00", "00:30:00",
## $ Demand      <dbl> 3834, 3785, 3708, 3634, 3581, 3552,
## $ Generation  <dbl> 4041, 4041, 4130, 4181, 4211, 4278,
## $ Wind        <dbl> 449, 505, 521, 492, 538, 561, 484, 4
## $ CO2         <dbl> 552, 548, 544, 543, 555, 531, 545, 5
## $ NetImports  <dbl> -145, -200, -294, -419, -503, -598,
## $ EWIC        <dbl> -33, -108, -183, -258, -333, -379, -
## $ Moyle       <dbl> -112, -92, -111, -161, -170, -219, -
## $ HourOfDay   <int> 0, 0, 0, 0, 1, 1, 1, 1, 2, 2, 2, 2,
## $ MinuteOfDay <int> 0, 15, 30, 45, 0, 15, 30, 45, 0, 15
```
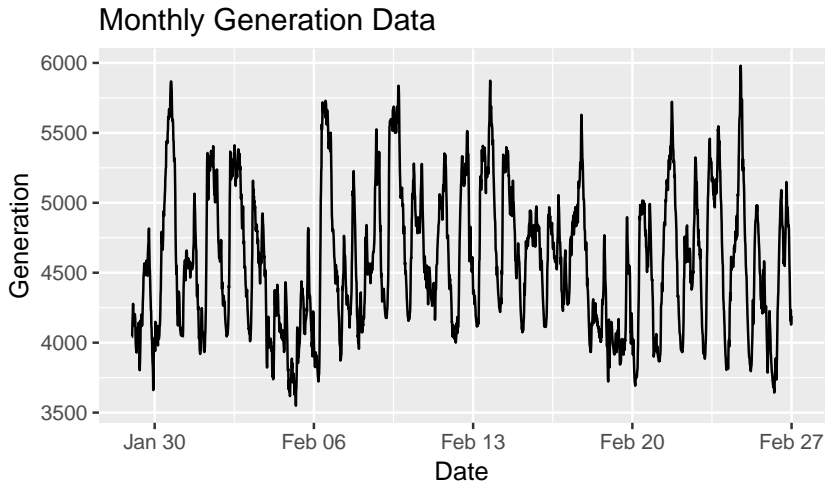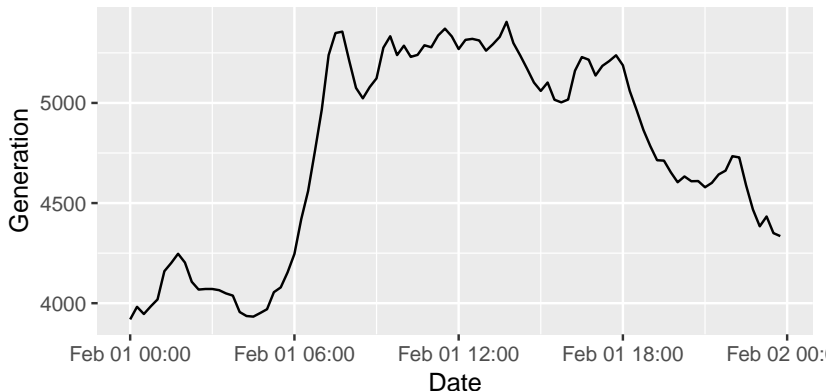
Plot time series

```
ggplot(data = ener,aes(x=DateTime, y=Generation)) +
  geom_line() + xlab("Date") + ylab("Generation") +
  ggtitle("Monthly Generation Data")
```

Value for 1/2/2017
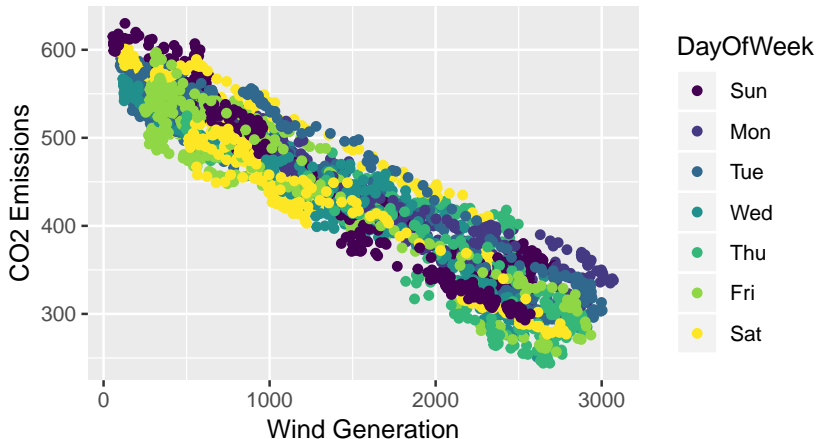
```
ggplot(data = filter(ener,Date=="2017-02-01"),
       aes(x=DateTime, y=Generation)) +
  geom_line() + xlab("Date") + ylab("Generation") +
  ggtitle("Generation Data for Feb 1st 2017")
```


Generation Data for Feb 1st 2017

Wind Generation v CO2 Emissions

```
ggplot(data = ener) +
  geom_point(aes(x=Wind,y=CO2,colour=DayOfWeek))+
  xlab("Wind Generation") + ylab("CO2 Emissions")
```

Linking weather to wind generation

Monthly Data

Please choose a monthly data report from any station by clicking one of the links below:

|			
1	Athenry	14	Knock Airport
2	Ballyhaise	15	Mace Head
3	Belmullet	16	Malin Head
4	Carlow Oakpark	17	Markree
5	Baldonnel - Casement Aerodrome	18	Mount Dillon
6	Claremorris	19	Mullingar
7	Cork Airport	20	Newport
8	Dublin Airport	21	Phoenix Park
9	Dunsany	22	Roches Point
10	Fermoy Moorepark	23	Shannon Airport
11	Finner	24	Sherkin Island
12	Gurteen Agri College	25	Valentia Observatory
13	Johnstown		
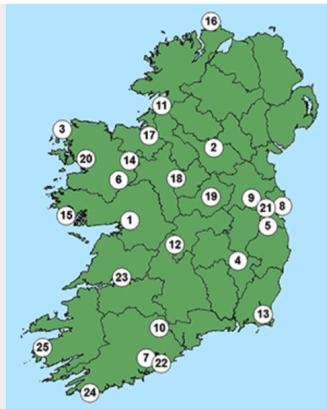
Sample Data

Daily Data

Weather Station Data - From 13/03/2015 to 12/03/2017

Please Select a Station and Date from the menu on the right.

REPORTS FROM MACE HEAD (A)

Date	Rainfall (mm)	Max Temp (°C)	Min Temp (°C)	Grass Min Temp (°C)	Mean Wind Speed (knots)	Maximum Gust (if >= 34 knots)	Sunshine (hours)
31/1/2017	0	10.3	7.3	5.8	10		

HOURLY VALUES (UTC) 31 Jan 2017 Mace Head
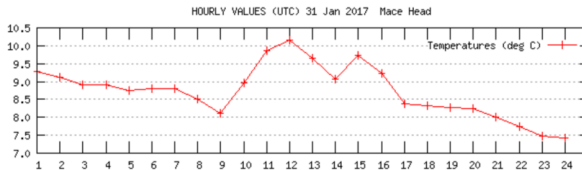
Select Station & Date

Station

Mace Head (A)

Date

31/01/2017

Go ->

Synoptic Stations 2011

http://www.met.ie

Mace Head Daily Data
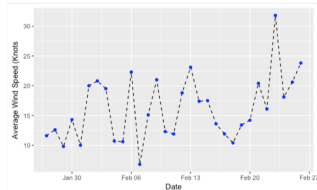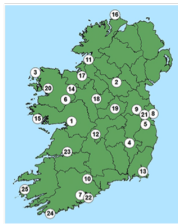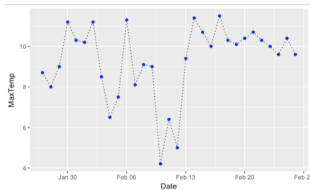
Date	Rainfall	MaxTemp	MinTemp	GrassMinTemp	AVRWind	MaxWindGust
27/01/17	7.9	8.7	4.3	-0.7	11.6	
28/01/17	3.5	8	4.5	2.9	12.6	
29/01/17	4.7	9	4.9	3.7	9.8	
30/01/17	7.8	11.2	7.1	5.8	14.3	
31/01/17	0	10.3	7.3	5.8	10	
01/02/17	0.6	10.2	6.1	5.2	20	38
02/02/17	4.9	11.2	7.4	6.4	20.8	45
03/02/17	2.2	8.5	3.6	2.1	19.5	46
04/02/17	5.3	6.5	1.8	-1.3	10.7	

Weather Data

```r
wd <- read_excel("../../datasets/energy/Mac Head Wind Data.
wd <- mutate(wd,Date=ymd(Date))
slice(wd,1:7)
```

```
## # A tibble: 7 x 7
##   Date       Rainfall MaxTemp MinTemp GrassMinTemp AVRW
##   <date>        <dbl>   <dbl>   <dbl>        <dbl>  <dl
## 1 2017-01-27      7.9     8.7     4.3         -0.7    11
## 2 2017-01-28      3.5     8       4.5          2.9    12
## 3 2017-01-29      4.7     9       4.9          3.7     9
## 4 2017-01-30      7.8    11.2     7.1          5.8    14
## 5 2017-01-31      0      10.3     7.3          5.8    10
## 6 2017-02-01      0.6    10.2     6.1          5.2    20
## 7 2017-02-02      4.9    11.2     7.4          6.4    20
```

Select Required Columns - Generation

```
gd <- select(ener,DateTime,Date,Wind) %>%
      arrange(DateTime)
slice(gd,1:7)

## # A tibble: 7 x 3
##   DateTime            Date         Wind
##   <dttm>              <date>       <dbl>
## 1 2017-01-29 00:00:00 2017-01-29    449
## 2 2017-01-29 00:15:00 2017-01-29    505
## 3 2017-01-29 00:30:00 2017-01-29    521
## 4 2017-01-29 00:45:00 2017-01-29    492
## 5 2017-01-29 01:00:00 2017-01-29    538
## 6 2017-01-29 01:15:00 2017-01-29    561
## 7 2017-01-29 01:30:00 2017-01-29    484
```

Select Required Columns - Weather

```
wd1 <- select(wd,Date,AVRWind) %>%
       arrange(Date)
slice(wd1,1:7)

## # A tibble: 7 x 2
##   Date        AVRWind
##   <date>        <dbl>
## 1 2017-01-27     11.6
## 2 2017-01-28     12.6
## 3 2017-01-29      9.8
## 4 2017-01-30     14.3
## 5 2017-01-31     10
## 6 2017-02-01     20
## 7 2017-02-02     20.8
```

Approach

Need to find the average generation by wind from grid data

```
avr_wd1 <- gd %>% group_by(Date) %>%
  summarise(AvrWindGeneration=mean(Wind))
slice(avr_wd1,1:7)
```

```
## # A tibble: 7 x 2
##   Date        AvrWindGeneration
##   <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.
```

Join the tables

```
join_t <- left_join(avr_wd1,wd1)

## Joining, by = "Date"
slice(join_t,1:7)

## # A tibble: 7 x 3
##   Date       AvrWindGeneration AVRWind
##   <date>                 <dbl>   <dbl>
## 1 2017-01-29              431.     9.8
## 2 2017-01-30             1726.    14.3
## 3 2017-01-31              330.    10
## 4 2017-02-01             2047.    20
## 5 2017-02-02             2647     20.8
## 6 2017-02-03             1050.    19.5
## 7 2017-02-04              591.    10.7
```
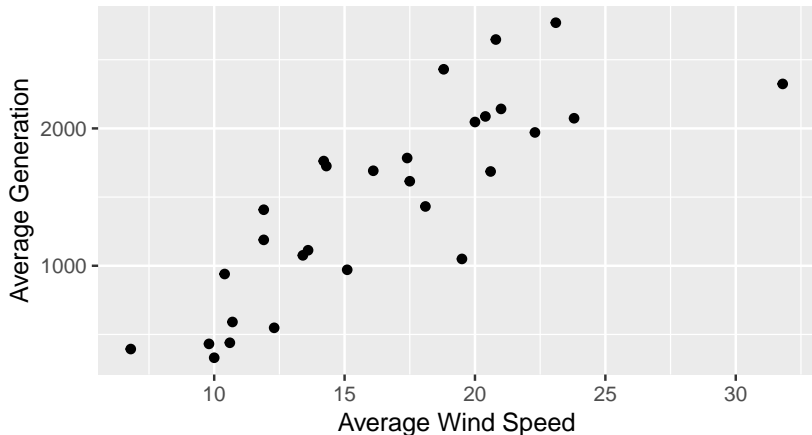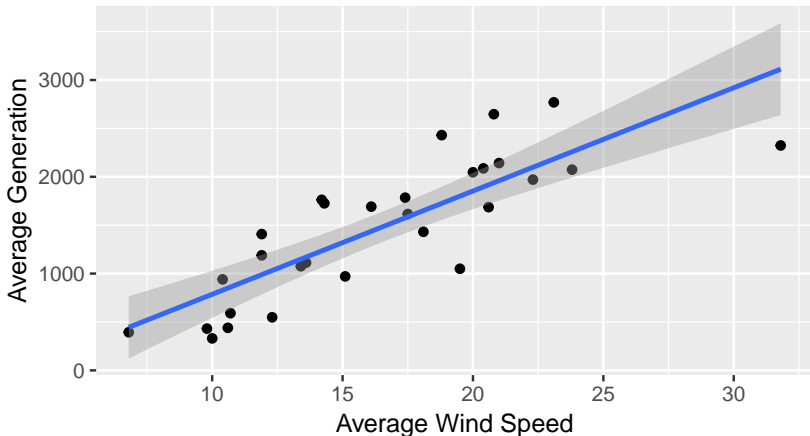
Plot Avr Wind Speed v Avr Wind Generation

```
ggplot(data = join_t) +
  geom_point(aes(x=AVRWind,y=AvrWindGeneration))+
  xlab("Average Wind Speed") + ylab("Average Generation")
```

Visualise Linear Model

```
ggplot(data = join_t,
       aes(x=AVRWind,y=AvrWindGeneration)) +
  geom_point()+xlab("Average Wind Speed") + ylab("Average (
  geom_smooth(method="lm")
```

Generate Model

```
mod <- lm(data=join_t,AvrWindGeneration~AVRWind)
mod
```

```
##
## Call:
## lm(formula = AvrWindGeneration ~ AVRWind, data = join_t)
##
## Coefficients:
## (Intercept)        AVRWind
##      -280.8          106.7
```
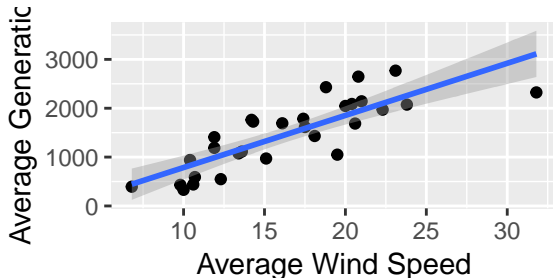
Predicting Values

```r
ggplot(data = join_t,
       aes(x=AVRWind,y=AvrWindGeneration)) +
  geom_point()+xlab("Average Wind Speed") + ylab("Average (
  geom_smooth(method="lm")
```



```r
predict(mod, newdata = data.frame(AVRWind=25))
```
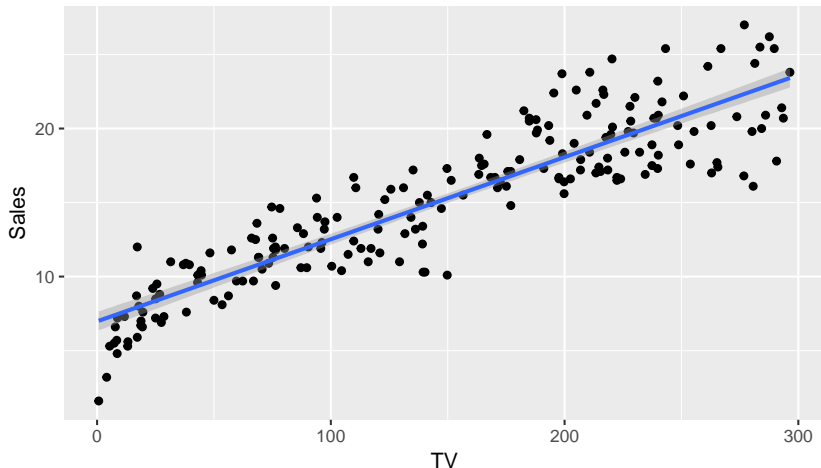
```
##        1
## 2386.727
```

Challenge

Generate linear models with the advertising data

```
adv <- read_excel("../../datasets/Advertising/AdvertisingDa
slice(adv,1:8)
```

```
## # A tibble: 8 x 4
##       TV Radio Newspaper Sales
##    <dbl> <dbl>     <dbl> <dbl>
## 1 230.   37.8      69.2  22.1
## 2  44.5  39.3      45.1  10.4
## 3  17.2  45.9      69.3  12
## 4 152.   41.3      58.5  16.5
## 5 181.   10.8      58.4  17.9
## 6   8.7  48.9      75     7.2
## 7  57.5  32.8      23.5  11.8
## 8 120.   19.6      11.6  13.2
```

Visualise Relationship

```
ggplot(data = adv,
       aes(x=TV,y=Sales)) +
  geom_point()+geom_smooth(method="lm")
```

Test Slide with Plot