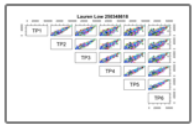
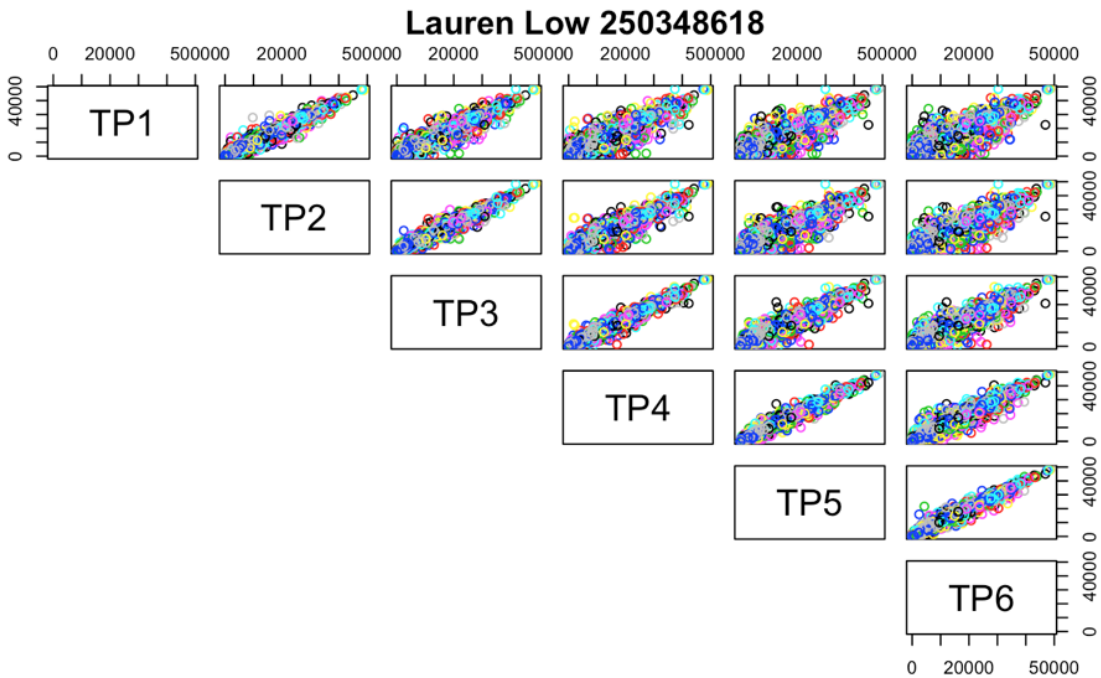
```
---
title: "Assignment 3"
author: "Lauren Low"
date: "9/19/2019"
output:
  html_document: default
  pdf_document: default
---

```{r setup, include=FALSE}

knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(readr)
library(lubridate)
library(leaflet)
generation = read.csv("~/Downloads/generation.csv", header = TRUE)
regions = read.csv("~/Downloads/regions.csv")
plants = read.csv("~/Downloads/plants.csv")
```

#Question 2

##part a
```{r}
generation$Fuel <- as.factor(plants$Fuel[match(generation$Node, plants$Node)])
generation$Region <- as.factor(plants$Region[match(generation$Node, plants$Node)])
generation$Renewable <- as.factor(plants$Type[match(generation$Node, plants$Node)])
generation
```
```

Node <fctr>	Date <fctr>	TP1 <dbl>	TP2 <dbl>	TP3 <dbl>	TP4 <dbl>	TP5 <dbl>	TP6 <dbl>	TP7 <dbl>
ARA2201	1/01/2018	6110.0000	6140.0000	5900.0000	5950.0000	6030.0000	6110.0000	5920.0000
ARA2201	2/01/2018	5670.0000	5430.0000	5400.0000	5410.0000	5290.0000	5180.0000	5470.0000
ARA2201	3/01/2018	5870.0000	6090.0000	5930.0000	6000.0000	6090.0000	5960.0000	6060.0000
ARA2201	4/01/2018	5810.0000	5960.0000	5980.0000	5860.0000	5780.0000	5810.0000	5750.0000
ARA2201	5/01/2018	5520.0000	5670.0000	5500.0000	5670.0000	5730.0000	5760.0000	5770.0000
ARA2201	6/01/2018	6270.0000	6670.0000	6370.0000	6390.0000	6920.0000	7560.0000	7590.0000
ARA2201	7/01/2018	5380.0000	5440.0000	5530.0000	5460.0000	5550.0000	5620.0000	5720.0000
ARA2201	8/01/2018	5750.0000	5870.0000	5930.0000	5860.0000	5970.0000	6000.0000	6010.0000
ARA2201	9/01/2018	11120.0000	11300.0000	11260.0000	9810.0000	10350.0000	8980.0000	9390.0000
ARA2201	10/01/2018	13110.0000	13030.0000	12480.0000	12440.0000	12470.0000	12420.0000	12450.0000

1-10 of 25,852 rows | 1-9 of 55 columns

````
##part b
```{r}
windgen <- generation %>%
  filter(Fuel == "Wind")
windgen
```
````

Node <fctr>	Date <fctr>	TP1 <dbl>	TP2 <dbl>	TP3 <dbl>	TP4 <dbl>	TP5 <dbl>	TP6 <dbl>	TP7 <dbl>
BPE0331	1/01/2018	4263.1475	3367.8923	3220.8250	3851.0001	2891.2927	2048.0407	1869.2268
BPE0331	2/01/2018	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BPE0331	3/01/2018	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BPE0331	4/01/2018	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BPE0331	5/01/2018	13867.5649	13314.4973	14079.8786	13841.9346	13027.9022	13464.0297	14371.6307
BPE0331	6/01/2018	5635.7140	10094.8378	10762.3702	10165.4268	7857.5765	4810.8622	2510.3863
BPE0331	7/01/2018	3665.1108	4247.6352	3545.2622	3388.8710	3028.7164	2995.7219	4017.5093
BPE0331	8/01/2018	8707.2954	6946.2108	6479.9768	6638.3379	8010.1103	9491.2728	8334.4752
BPE0331	9/01/2018	8.5709	590.1052	2062.4596	4061.8595	5151.4923	4299.1640	2145.6214
BPE0331	10/01/2018	1865.7407	1381.6016	1820.8748	1875.7866	2454.8351	3908.2428	4186.4526

1-10 of 3,285 rows | 1-9 of 55 columns

Previous 1 2 3 4 5 6 ... 100 Next

````
##part c
```{r}
windgen_graph <- pairs(windgen[c(3:8)], col = windgen$Date, lower.panel = NULL, cex.labels = 2) +
  title("Lauren Low 250348618")
windgen_graph
#In general, closer trading periods had stronger positive correnlations of energy (measured in MWh) collegected by wind turbines (i.e. the positive correlation between TP1 and TP2 was stronger than between TP1 and TP6)
```
````






```
46
47 ##part d
48 ```{r}
49 generation_gather <- gather(generation, key = TP, value = MWh, starts_with("TP"))
50 generation_gather
51 ```
```

Node <fctr>	Date <fctr>	Fuel <fctr>	Region <fctr>	Renewable <fctr>	TP <chr>	MWh <dbl>
ARA2201	1/01/2018	Hydro	USI	Renewable	TP1	6110.0000
ARA2201	2/01/2018	Hydro	USI	Renewable	TP1	5670.0000
ARA2201	3/01/2018	Hydro	USI	Renewable	TP1	5870.0000
ARA2201	4/01/2018	Hydro	USI	Renewable	TP1	5810.0000
ARA2201	5/01/2018	Hydro	USI	Renewable	TP1	5520.0000
ARA2201	6/01/2018	Hydro	USI	Renewable	TP1	6270.0000
ARA2201	7/01/2018	Hydro	USI	Renewable	TP1	5380.0000
ARA2201	8/01/2018	Hydro	USI	Renewable	TP1	5750.0000
ARA2201	9/01/2018	Hydro	USI	Renewable	TP1	11120.0000

```
##part e
```{r}
generation_fuel <- generation_gather %>%
  group_by(Fuel) %>%
  summarize(MWh = sum(MWh, na.rm = TRUE))
generation_fuel
```
```

Fuel <fctr>	MWh <dbl>
Coal	1172104497
Diesel	6442272
Gas	5152530887
Geo	7186851782
Hydro	25303567080
Wind	1658591069
Wood	220054591

7 rows

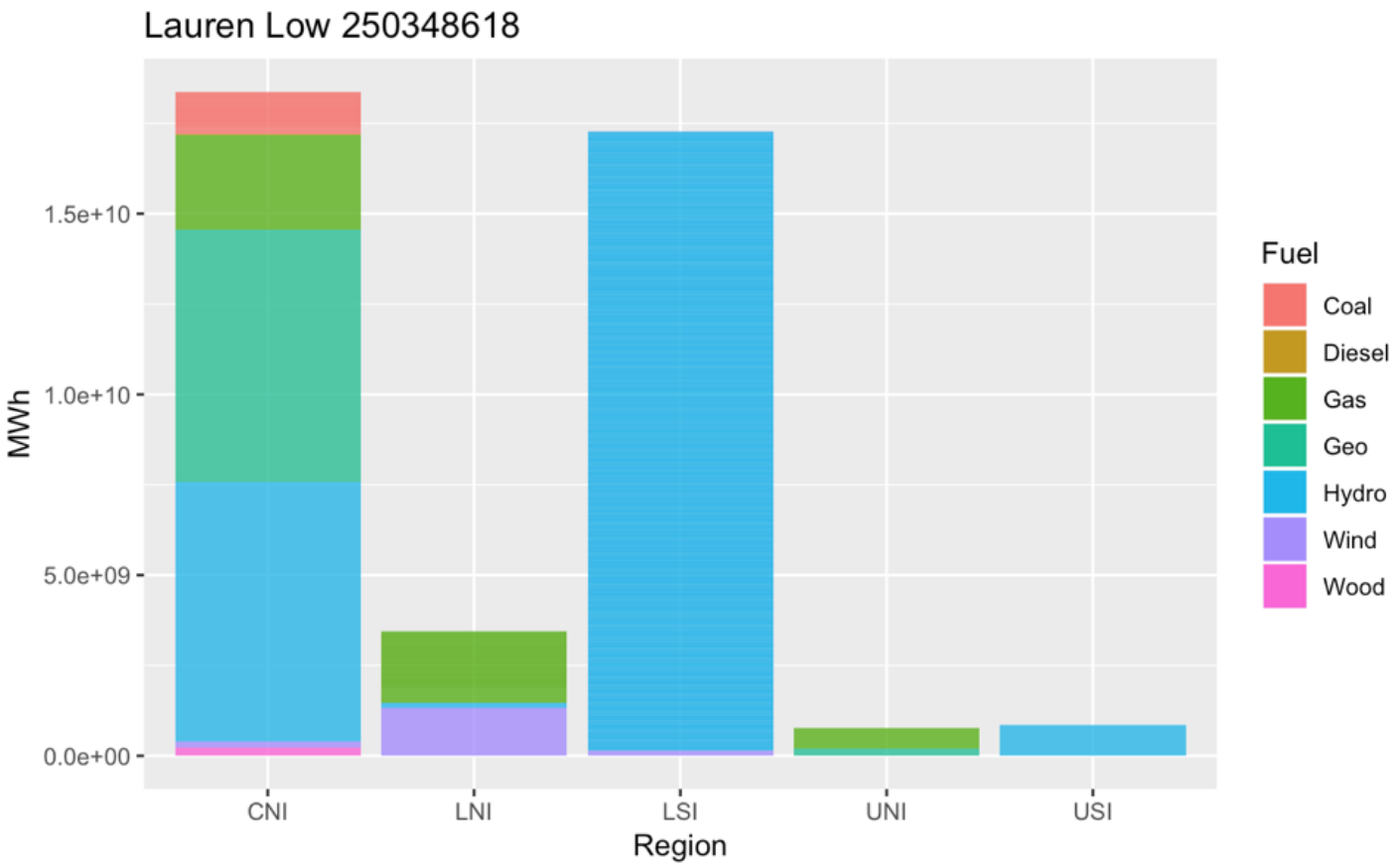
```
##part f
```{r}
generation_regions <- generation_gather %>%
  group_by(Region) %>%
  summarize(MWh = sum(MWh, na.rm = TRUE))
generation_regions
```
```

Region <fctr>	MWh <dbl>
CNI	18370360249
LNI	3439389683
LSI	17275698106
UNI	771300086
USI	843394053

5 rows

```{r}
stacked_plot <- ggplot(data = generation_gather) +
  geom_bar(mapping = aes(x = Region, y = MWh, fill = Fuel), stat = "identity") +
  ggtitle("Lauren Low 250348618")
stacked_plot
#In region CNI, the highest source of energy generation was pretty evenly split between geo and hydro, in region LNI the highest source of energy generation was gas, in region LSI the highest source of energy generation was hydro, in region UNI the highest source of energy generation was gas, and in region USI the highest source of energy generation was hydro.  Regions CNI and LNI generated the most energy while regions UNI and USI generatd the least energy.
```

Removed 51704 rows containing missing values (position_stack).



Converting TP values to integers:

```{r}
generation_gather$TP = as.integer(gsub("TP", "", generation_gather$TP))
generation_gather
```

Node <fctr>	Date <fctr>	Fuel <fctr>	Region <fctr>	Renewable <fctr>	TP <int>	MWh <dbl>
ARA2201	1/01/2018	Hydro	USI	Renewable	1	6110.0000
ARA2201	2/01/2018	Hydro	USI	Renewable	1	5670.0000
ARA2201	3/01/2018	Hydro	USI	Renewable	1	5870.0000
ARA2201	4/01/2018	Hydro	USI	Renewable	1	5810.0000
ARA2201	5/01/2018	Hydro	USI	Renewable	1	5520.0000
ARA2201	6/01/2018	Hydro	USI	Renewable	1	6270.0000
ARA2201	7/01/2018	Hydro	USI	Renewable	1	5380.0000
ARA2201	8/01/2018	Hydro	USI	Renewable	1	5750.0000
ARA2201	9/01/2018	Hydro	USI	Renewable	1	11120.0000
ARA2201	10/01/2018	Hydro	USI	Renewable	1	13110.0000

1-10 of 1,292,600 rows

Previous 1 2 3 4 5 6 ... 100 Next

Parsing date:

```{r}
generation_gather$Date = parse_date_time(generation_gather$Date,"d/m/Y")
generation_gather
```

Node <fctr>	Date <S3: POSIXct>	Fuel <fctr>	Region <fctr>	Renewable <fctr>	TP <int>	MWh <dbl>
ARA2201	2018-01-01	Hydro	USI	Renewable	1	6110.0000
ARA2201	2018-01-02	Hydro	USI	Renewable	1	5670.0000
ARA2201	2018-01-03	Hydro	USI	Renewable	1	5870.0000
ARA2201	2018-01-04	Hydro	USI	Renewable	1	5810.0000
ARA2201	2018-01-05	Hydro	USI	Renewable	1	5520.0000
ARA2201	2018-01-06	Hydro	USI	Renewable	1	6270.0000
ARA2201	2018-01-07	Hydro	USI	Renewable	1	5380.0000
ARA2201	2018-01-08	Hydro	USI	Renewable	1	5750.0000
ARA2201	2018-01-09	Hydro	USI	Renewable	1	11120.0000
ARA2201	2018-01-10	Hydro	USI	Renewable	1	13110.0000

1-10 of 1,292,600 rows

Previous 1 2 3 4 5 6 ... 100 Next

```r
##part g
```{r}
generation_gather$month <- format(as.Date(generation$Date), "%m")
generation_gather
```
```

Node <fctr>	Date <S3: POSIXct>	Fuel <fctr>	Region <fctr>	Renewable <fctr>	TP <int>	MWh <dbl>	month <chr>
ARA2201	2018-01-01	Hydro	USI	Renewable	1	6110.0000	01
ARA2201	2018-01-02	Hydro	USI	Renewable	1	5670.0000	01
ARA2201	2018-01-03	Hydro	USI	Renewable	1	5870.0000	01
ARA2201	2018-01-04	Hydro	USI	Renewable	1	5810.0000	01
ARA2201	2018-01-05	Hydro	USI	Renewable	1	5520.0000	01
ARA2201	2018-01-06	Hydro	USI	Renewable	1	6270.0000	01
ARA2201	2018-01-07	Hydro	USI	Renewable	1	5380.0000	01
ARA2201	2018-01-08	Hydro	USI	Renewable	1	5750.0000	01
ARA2201	2018-01-09	Hydro	USI	Renewable	1	11120.0000	01
ARA2201	2018-01-10	Hydro	USI	Renewable	1	13110.0000	01

1-10 of 1,292,600 rows

Previous 1 2 3 4 5 6 ... 100 Next

```r
```{r}
avgs_table <- generation_gather %>%
  filter(TP <= 48) %>%
  group_by(month, TP) %>%
  summarize(avg_MWh = mean(MWh))
avgs_table
```
```

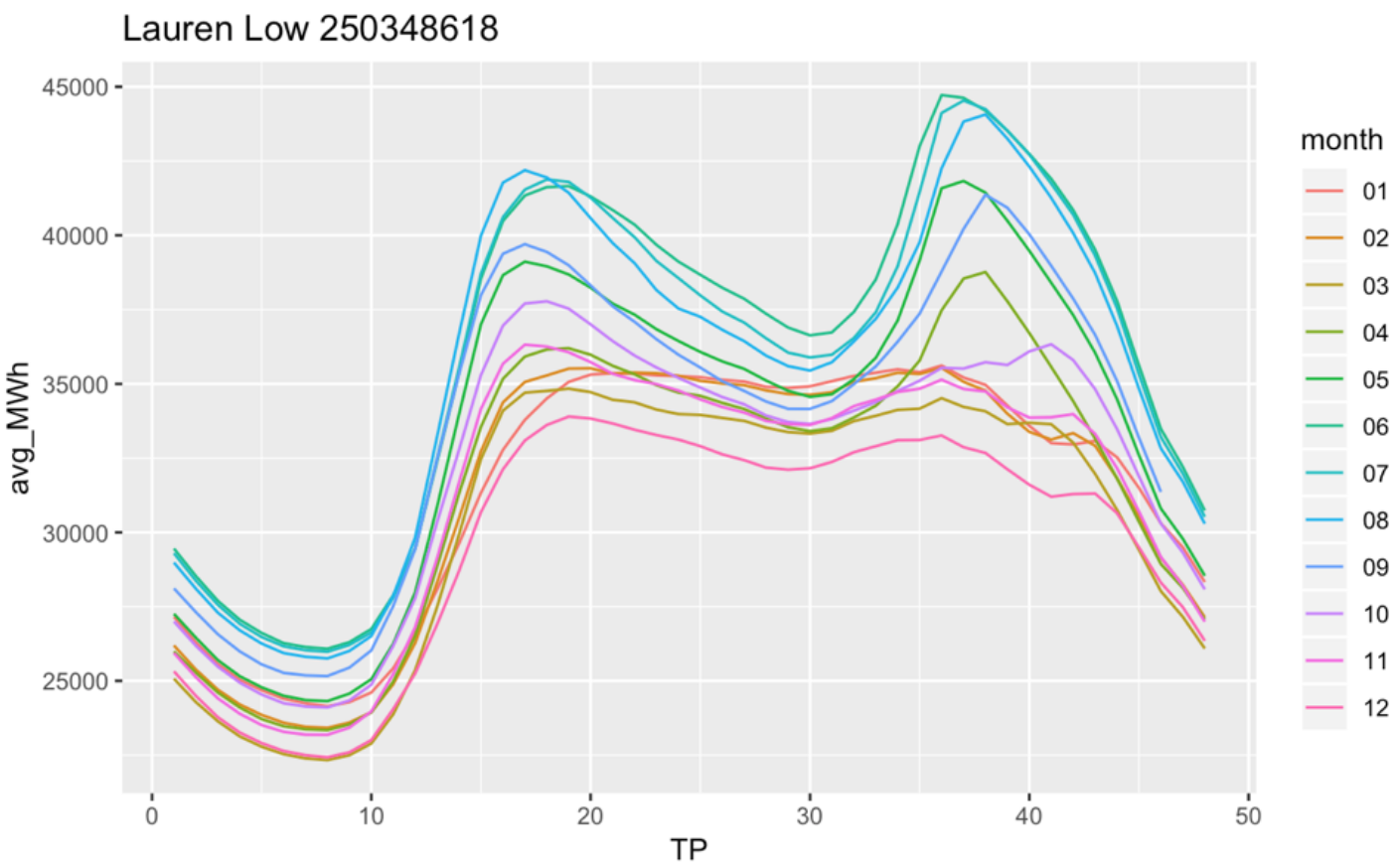
month <chr>	TP <int>	avg_MWh <dbl>
01	1	27160.36
01	2	26296.83
01	3	25569.48
01	4	25042.27
01	5	24687.00
01	6	24403.90
01	7	24250.12
01	8	24146.94
01	9	24285.83
01	10	24610.84

1-10 of 576 rows

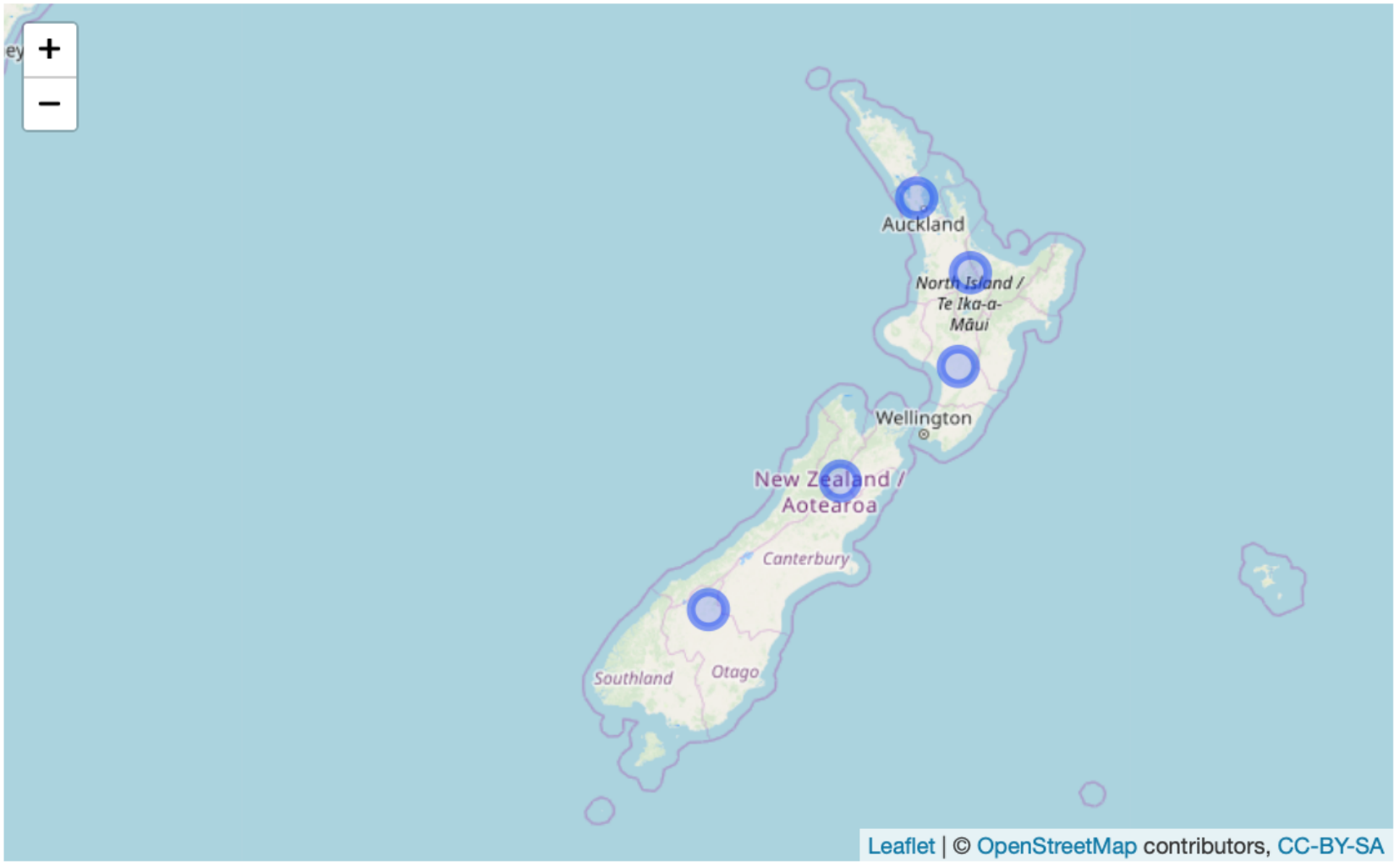
Previous 1 2 3 4 5 6 ... 58 Next

```{r}
line_graph <- ggplot(data = avgs_table) +
  geom_line(mapping = aes(x = TP, y = avg_MWh, color = month)) +
  ggtitle("Lauren Low 250348618")
line_graph
#In general, during winter months in New Zealand, the energy generation was higher than the summer months.  Further, on average there were energy generation spikes around trading periods 17 and 37 of each month and low amounts of energy generation around trading period 8
```

Removed 2 rows containing missing values (geom_path).



```
##part h
```{r}
map <- leaflet(regions) %>%
  addTiles() %>%
  addCircleMarkers(lng = regions$Lng, lat = regions$Lat)
map
```
```



```
##part i
```{r}
generation_gather2 <- gather(generation, key = TP, value = MWh, starts_with("TP"))
```

```{r}
totalgen <- generation_gather2 %>%
  group_by(Region, Renewable) %>%
  summarize(Total = sum(MWh, na.rm = TRUE))
totalgen
```
```

Region <fctr>	Renewable <fctr>	Total <dbl>
CNI	FossilFuel	3798556266
CNI	Renewable	14571803983
LNI	FossilFuel	1964406004
LNI	Renewable	1474983679
LSI	Renewable	17275698106
UNI	FossilFuel	568115385
UNI	Renewable	203184701
USI	Renewable	843394053

8 rows

```
##part j
```{r}
totalgen.spread <- totalgen %>%
  spread(key = Renewable, value = Total)
totalgen.spread
```
```

Region <fctr>	FossilFuel <dbl>	Renewable <dbl>
CNI	3798556266	14571803983
LNI	1964406004	1474983679
LSI	*NA*	17275698106
UNI	568115385	203184701
USI	*NA*	843394053

5 rows

replacing NA values with 0

```{r}
totalgen.spread[is.na(totalgen.spread)]=0
totalgen.spread
```

Region <fctr>	FossilFuel <dbl>	Renewable <dbl>
CNI	3798556266	14571803983
LNI	1964406004	1474983679
LSI	0	17275698106
UNI	568115385	203184701
USI	0	843394053

5 rows

##part k

```{r}
totalgen.spread$Lat <- as.numeric(regions$Lat[match(totalgen.spread$Region, regions$Region)])
totalgen.spread$Lng <- as.numeric(regions$Lng[match(totalgen.spread$Region, regions$Region)])
totalgen.spread$TotalGenerarion <- totalgen.spread$FossilFuel + totalgen.spread$Renewable
totalgen.spread$PTotalGenerarion <- 40*((totalgen.spread$FossilFuel + totalgen.spread$Renewable) / 18370360249)
totalgen.spread$PRenewable <- totalgen.spread$Renewable / totalgen.spread$TotalGenerarion
totalgen.spread
```

Region <fctr>	FossilFuel <dbl>	Renewable <dbl>	Lat <dbl>	Lng <dbl>	TotalGenerarion <dbl>	PTotalGenerarion <dbl>	PRenewable <dbl>
CNI	3798556266	14571803983	-38.13938	175.9367	18370360249	40.000000	0.7932236
LNI	1964406004	1474983679	-39.96193	175.6443	3439389683	7.488998	0.4288504
LSI	0	17275698106	-44.51435	169.3139	17275698106	37.616460	1.0000000
UNI	568115385	203184701	-36.64952	174.5874	771300086	1.679445	0.2634315
USI	0	843394053	-42.15198	172.6827	843394053	1.836424	1.0000000

5 rows

```{r}
colorN = colorNumeric(palette = "viridis", domain = c(0,1))
proportional_map <- leaflet(totalgen.spread) %>%
  addTiles() %>%
  addCircleMarkers(lng = totalgen.spread$Lng, lat = totalgen.spread$Lat, color = ~colorN(totalgen.spread$PRenewable), radius = totalgen.spread$PTotalGenerarion) %>%
  addLegend("bottomright", pal = colorN, values = ~PRenewable, title = "Proportion of Renewable")
 proportional_map
```