Assignment 1

Setup:
To collect the data we used the command read.csv() this gave us a neat table of 9 variables. From there we changed the object type of Date to a POSIXlt, which gave us the ability to easily distinguish month, day and year. Using Subset() we obtained a table of May 7, 2007, to May 13, 2007, which is the 19th week of 2007.

Part 1:

- **Arithmetic mean:**

  The arithmetic mean was obtained using mean().

	Global active power	Global reactive power	Voltage
Arithmetic mean	1.10992977678571	0.169166468253968	240.484470238095

- **Geometric mean:**

  The geometric mean computed using exp(mean(log()))

	Global active power	Global reactive power	Voltage
Geometric mean	0.753	0	240.476

- **Median:**

  The median was obtained using median().

	Global active power	Global reactive power	Voltage
Median	0.75925	0.158	240.55

- **Standard deviation:**

  standard deviation was obtained using sd().

	Global active power	Global reactive power	Voltage
Stand deviation	0.9896	0.1191	1.9744

- **Mode:**

  To get the mode we used a function that first eliminated all duplicates giving us a list of unique numbers. We then compared against the original list to find which data point occurred the most.

	Global active power	Global reactive power	Voltage
Mode	0.15	0	240.65

- **Min and Max:**

  To start we subsetted the weekdays from the weekends, then decided that day would be from 6am to 9pm, and night would be 9pm to 12am. We then used the subset() function to create another subset for days and nights of weekdays and weekends. This left us with 4 different subsets. From there we applied min() and max function() to each set.

	Global active power	Global reactive power
Max weekday day	7.572	0.44
Max weekday night	7.572	0.786
Max weekend day	4.422	0.462
Max weekend night	6.024	0.724
Min weekday day	0.11	0
Min weekday night	0.11	0
Min weekend day	0.144	0
Min weekend night	0.142	0

Part 2:

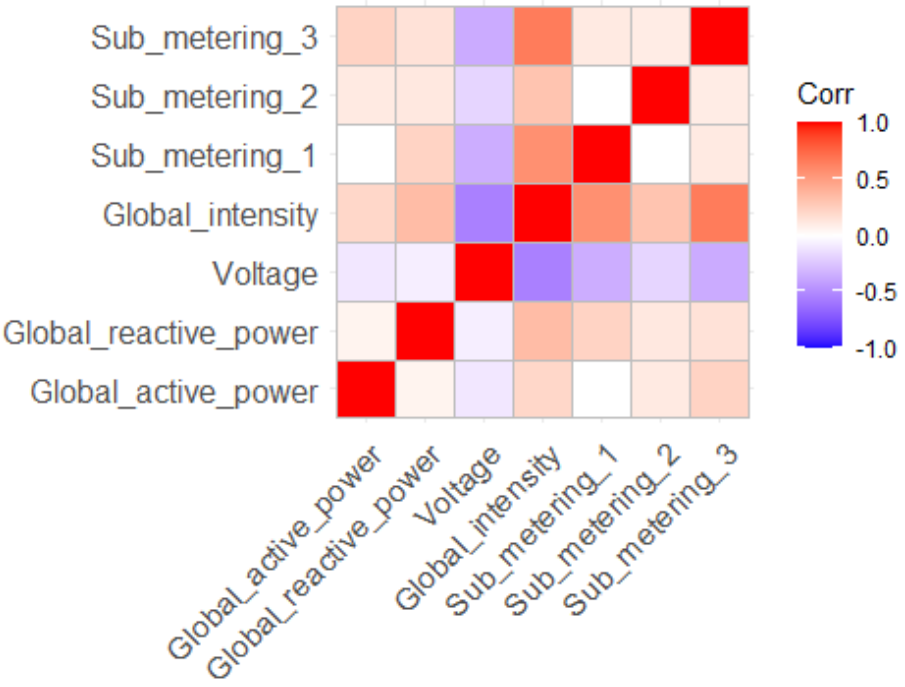
- **Correlation:**

  To calculate the correlation of each disjoint pair we used cor() with the Pearson method.

	A	B	C	D	E	F	G
A	1						
B	0.058	1					
C	-0.1	-0.075	1				
D	0.208	0.351	-0.548	1			
E	-0.004	0.232	-0.353	0.572	1		
F	0.105	0.118	-0.182	0.312	0.002	1	
G	0.233	0.155	-0.356	0.662	0.111	0.097	1

Correlation matrix:

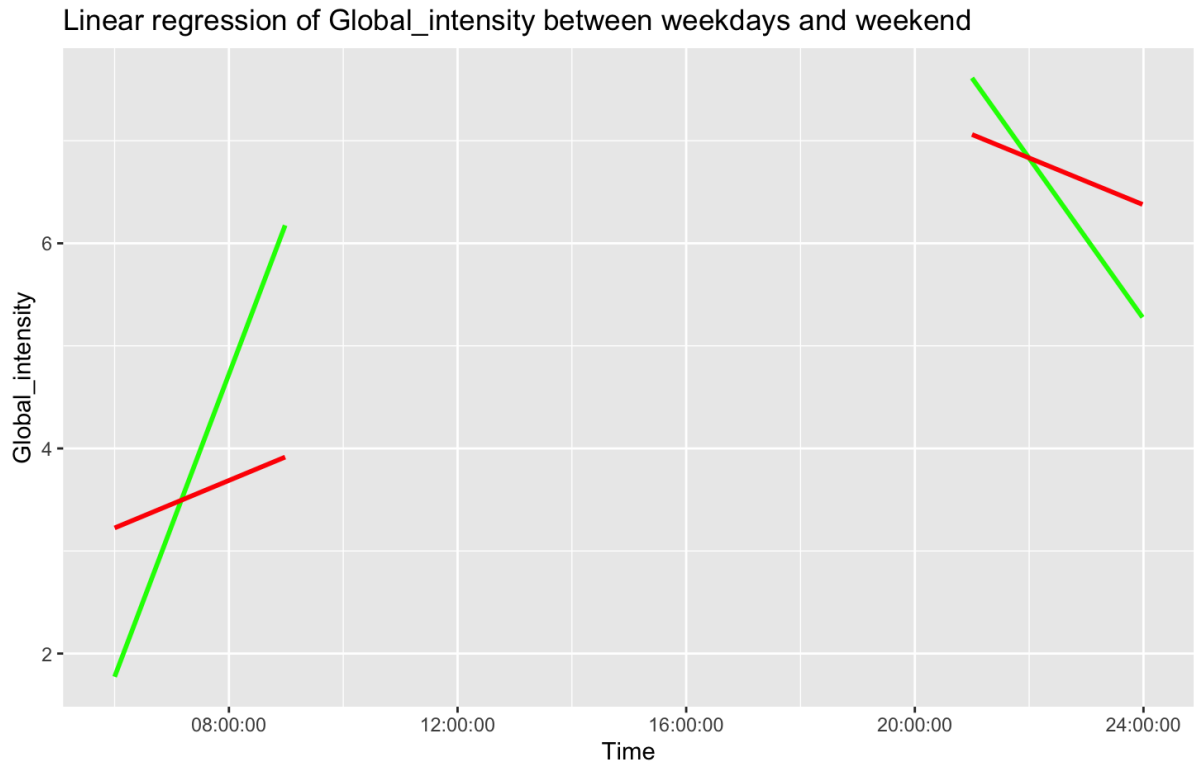
To we plotted the correlation matrix using ggcorrplot()



Part 3:

For the time we chose the day to be 6am to 9am and night to be 9pm to 12pm. We then created two functions blackboxDay and blackboxNight to get the mean of global intensity for each weekdays and weekends minutes and put it in its own list. From there we added it to the subset and created a graph using ggplot() with a ggplot_smooth() to add linear and polynomial regression.

Linear Regression:
Green line stands for weekends, red line stands for weekdays.



Polynomial Regression:
Green line stands for weekends, red line stands for weekdays.