Final

Data Description

This dataset is a collection of buildings in Chicago and how much energy they consume. Other characteristics of the buildings, like their type, are provided.

Energy usage is given separately by month (columns).

Part 1

I am looking to see how energy usage varies by month. Interestingly, winter months like January and February show low levels of energy used (despite probably using heat), yet December's usage is high. The summer months all show high energy usage probably because of air conditioning.

The Therm chart shows almost the opposite of the KWH one. The winter months have the highest usage and the summer months show some of the lowest.

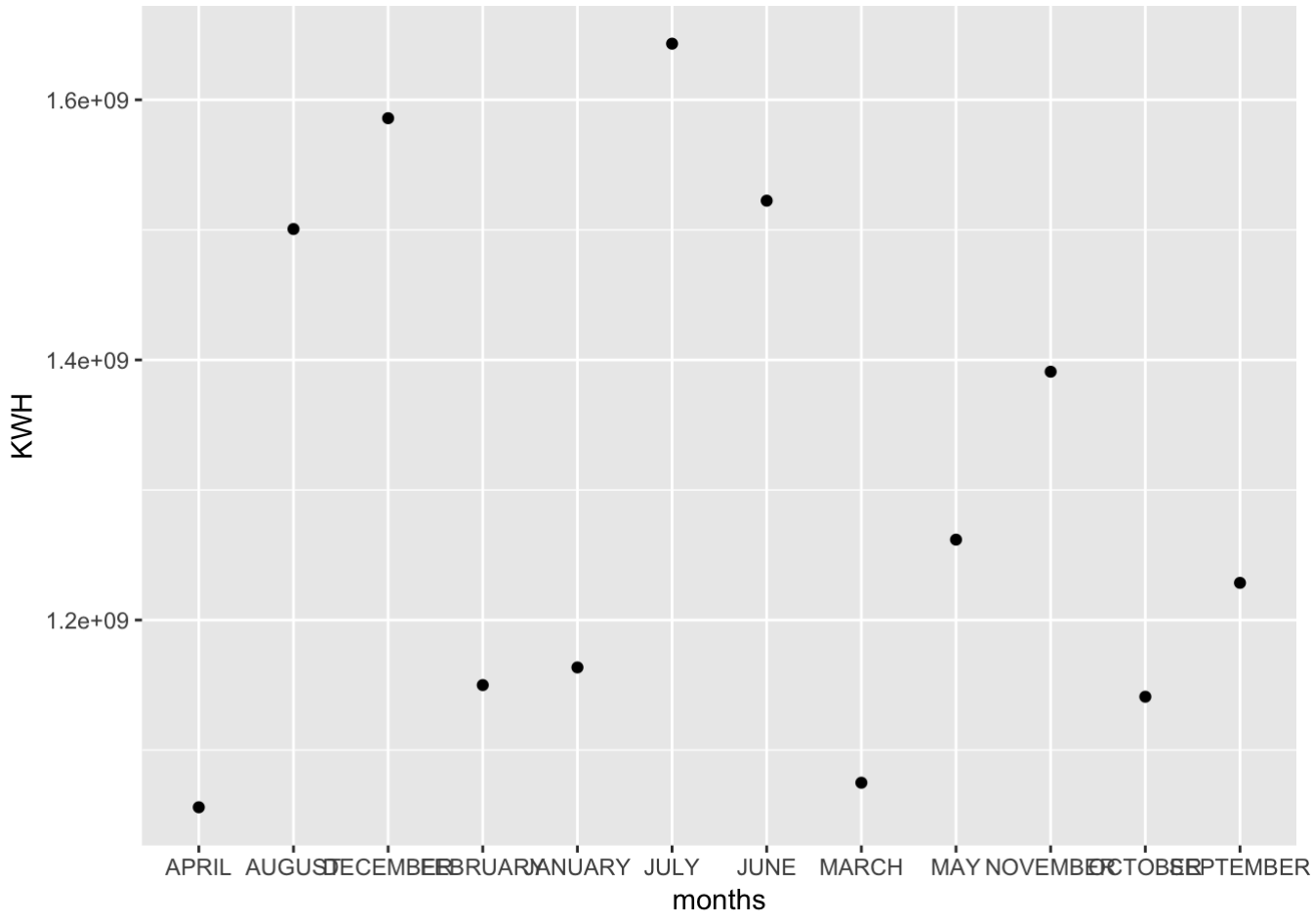
```
source("./Final/Final/Tracy/Final.R")
```

```
##
## Attaching package: 'dplyr'
```

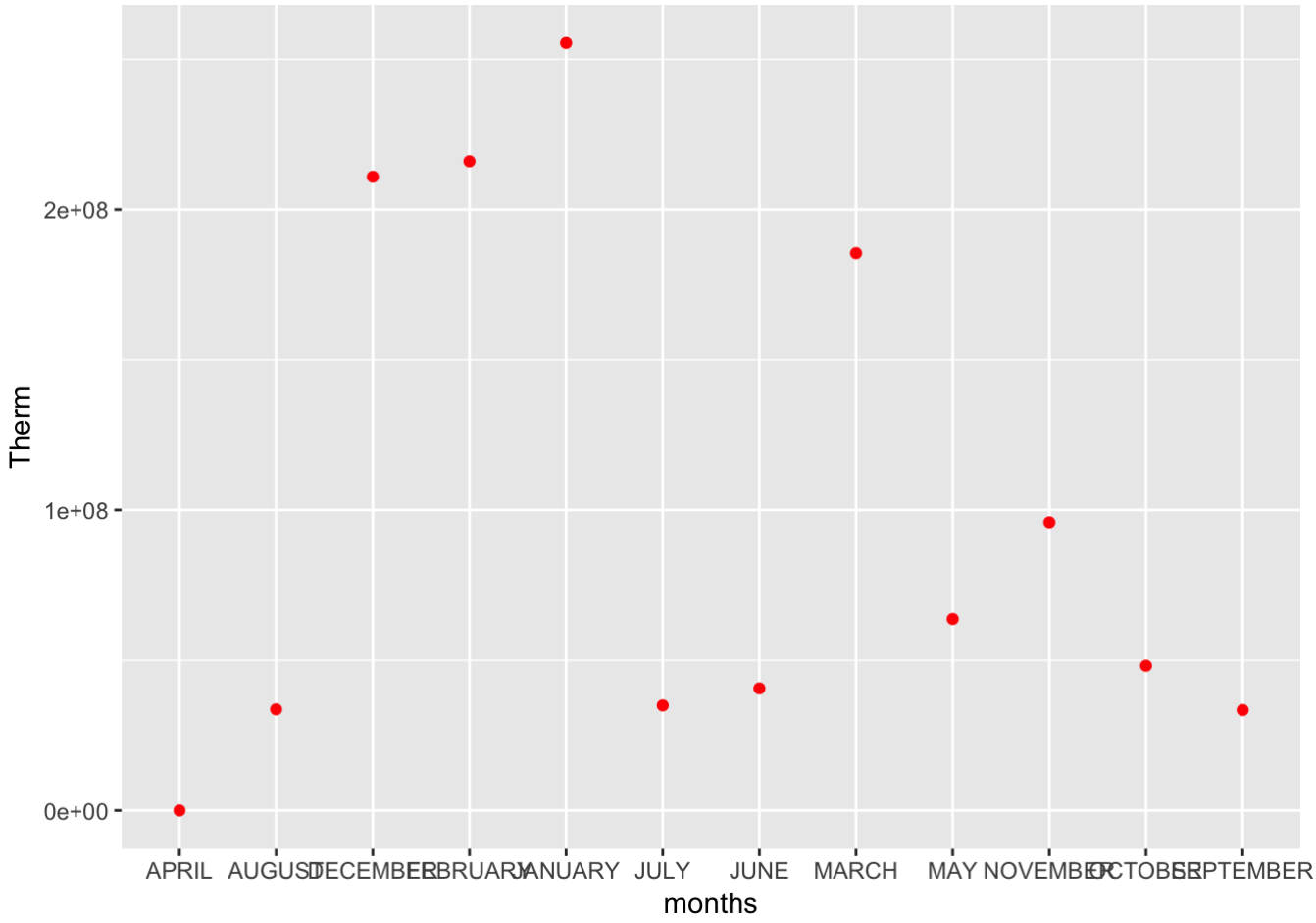
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(dplyr)
library(ggplot2)

months <- get_months()

totals_KWH <- rep(NA, length(months))
for (i in 1:length(months)) {
  cm <- months[i]
  total <- get_total_KWH_month(cm, e)
  totals_KWH[i] <- total
}

totals_therm <- rep(NA, length(months))
for (i in 1:length(months)) {
  cm <- months[i]
  total <- get_total_therm_month(cm)
  totals_therm[i] <- total
}

monthly_usage <- data.frame(months=months, KWH=totals_KWH, Therm=totals_therm,
                            stringsAsFactors=F)

p <- ggplot()
p <- p + geom_point(mapping=aes(x=months, y=KWH), data=monthly_usage)
print(p)
```



```
p <- ggplot()
p <- p + geom_point(mapping=aes(x=months, y=Therm), color="red", data=monthly_usage)
print(p)
```

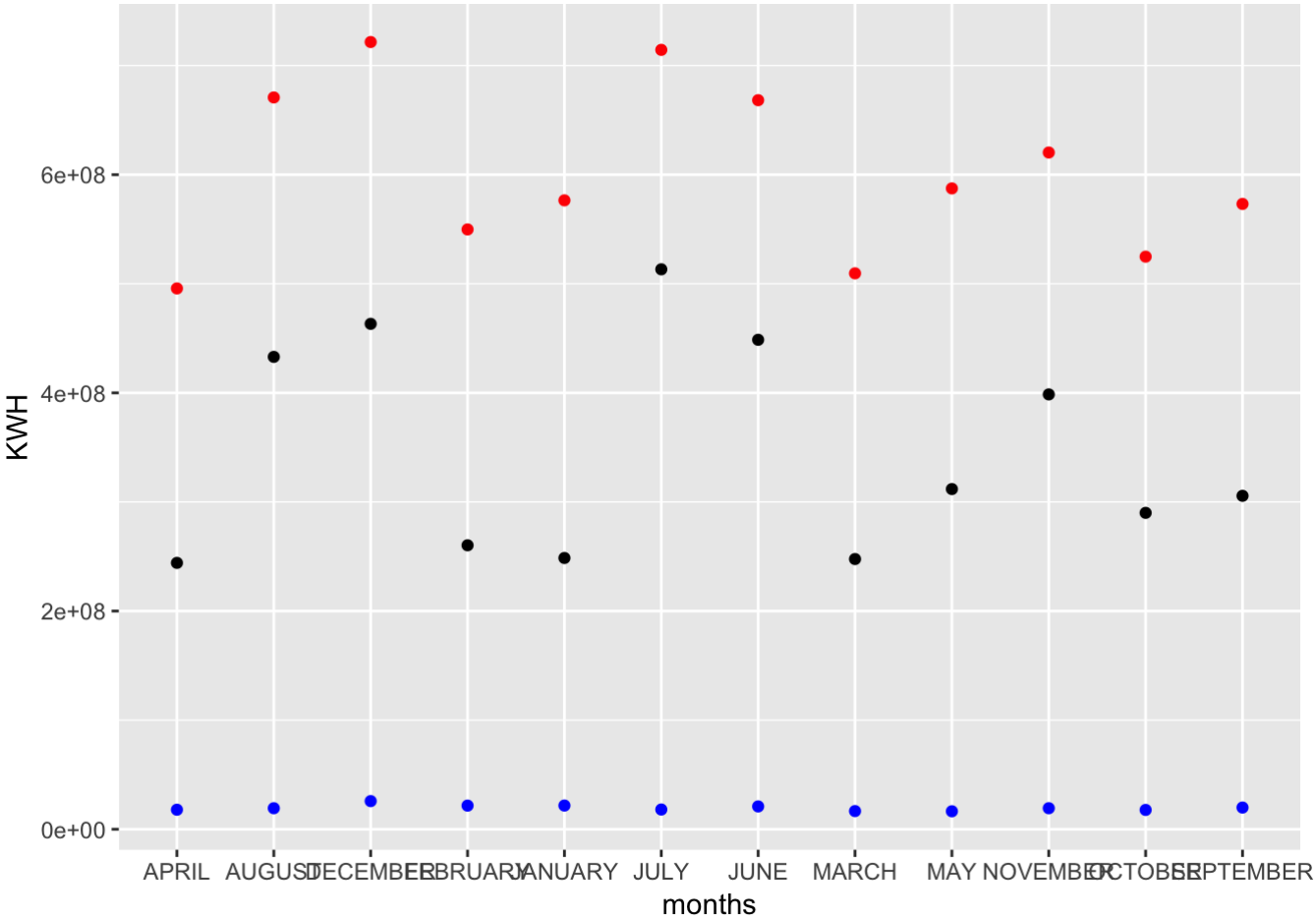


Part 2

I am exploring the differences in energy usage based off building type. Commercial buildings use the most, which is expected since business, shopping, and other public buildings tend to use HVAC and lighting much more liberally.

Interestingly enough, industrial buildings used significantly less energy than I would have thought while residential buildings fell in the middle though still much higher than industrial buildings. This was not controlled for building size or occupation, however, so those could omitted variables.

```
res_totals_KWH <- rep(NA, length(months))
for (i in 1:length(months)) {
  cm <- months[i]
  res_filt_e <- get_building_type("Residential")
  total <- get_total_KWH_month(cm, res_filt_e)
  res_totals_KWH[i] <- total
}

com_totals_KWH <- rep(NA, length(months))
for (i in 1:length(months)) {
  cm <- months[i]
  com_filt_e <- get_building_type("Commercial")
  total <- get_total_KWH_month(cm, com_filt_e)
  com_totals_KWH[i] <- total
}

ind_totals_KWH <- rep(NA, length(months))
for (i in 1:length(months)) {
  cm <- months[i]
  ind_filt_e <- get_building_type("Industrial")
  total <- get_total_KWH_month(cm, ind_filt_e)
  ind_totals_KWH[i] <- total
}

building_type_KWH_df <- data.frame(months=months, residential=res_totals_KWH,
                                   commercial=com_totals_KWH, industrial=ind_totals_KWH,
                                   stringsAsFactors = F)

p <- ggplot()
p <- p + geom_point(mapping=aes(x=months, y=residential),
                    data=building_type_KWH_df)
p <- p + geom_point(mapping=aes(x=months, y=commercial),
                    data=building_type_KWH_df, color="red")
p <- p + geom_point(mapping=aes(x=months, y=industrial),
                    data=building_type_KWH_df, color="blue")
p <- p + labs(y="KWH")
print(p)
```


6e+08
4e+08
2e+08
0e+00
KWH
APRIL
AUGUST
DECEMBER
FEBRUARY
JANUARY
JULY
JUNE
MARCH
MAY
NOVEMBER
OCTOBER
SEPTEMBER
months