R1 report

Author: Richard T. Watson

Date: 2020-06-14

```
library(dplyr)
library(readr)
library(lubridate)
```

a.

Download data Atlanta weather (a timestamp, air temperature, humidity, and precipitation), and electricity prices (a timestamp and cost in cents per kWh). Merge the files for electricity, price, and weather.

```
url <-  "http://www.richardtwatson.com/data/ATLweather.csv"
w <- read_delim(url,delim=',')
```

```
## Parsed with column specification:
## cols(
##   Timestamp = col_datetime(format = ""),
##   Temperature = col_double(),
##   Humidity = col_double(),
##   Precipitation = col_double()
## )
```

```
url <-  "http://www.richardtwatson.com/data/electricityprices.csv"
e <- read_delim(url,delim=',')
```

```
## Parsed with column specification:
## cols(
##   timestamp = col_datetime(format = ""),
##   cost = col_double()
## )
```

```
m <- inner_join(w,e,by=c("Timestamp" = 'timestamp'))
```

b.

Compute the correlation between temperature and electricity price. What do you conclude?

```r
cor.test(m$Temperature,m$cost)
```

```
##
##  Pearson's product-moment correlation
##
## data:  m$Temperature and m$cost
## t = 42.434, df = 52506, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1738064 0.1903459
## sample estimates:
##      cor
## 0.182089
```

Conclusion

- As the p-value is less than .05, conclude that there is a relationship between temperature and electricity cost. As the temperature increases, electricity prices increase.
- The correlation is *small* (0.18) as it is between .1 and .3, as shown in the following table:

Correlation coefficient	Effect size
.10 - .30	Small
.30 - .50	Moderate
$>$ .50	Large

- Given the small correlation, there are likely other factors that influence electricity cost, such as generation capacity

Note: The p-value tells you whether there is a relationship and the correlation coefficient indicates the size of that relationship.

c.

Extract the data for July through September (Summer) and redo the correlation What do you conclude?

```r
Summer <- m %>% filter((month(Timestamp) >= 7 & month(Timestamp) <= 9))
cor.test(Summer$Temperature,Summer$cost)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Summer$Temperature and Summer$cost
## t = 66.584, df = 12493, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4987230 0.5246084
## sample estimates:
##       cor
## 0.5117819
```

Conclusion

- As the p-value is less than .05, conclude that there is a relationship between temperature and electricity cost during summer.
- The correlation is *large* (0.51). There appears to be generation capacity problem in summer, usually the hottest months of the year, so electricity prices are raised to dampened demand on hot days.

d.

Extract the data for January through March (Winter) and redo the correlation What do you conclude?

```
Winter <- m %>% filter((month(Timestamp) >= 1 & month(Timestamp) <= 3))
cor.test(Winter$Temperature,Winter$cost)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Winter$Temperature and Winter$cost
## t = -39.705, df = 13710, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3360679 -0.3060433
## sample estimates:
##        cor
## -0.3211363
```

Conclusion

- As the p-value is less than .05, conclude that there is a relationship between temperature and electricity cost during winter
- The correlation is negative and *moderate* (-.32). There appears to be generation capacity problem in winter, usually the coldest months of the year, so electricity prices are raised to dampened demand on cold days.

f.

Download the Athens data for solar (a timestamp and solar radiation in watts/m2), Using the Athens solar radiation data, compute the average (one value), min and max for solar radiation.

Average, min, and max

```
url <- "http://www.richardtwatson.com/data/SolarRadiationAthens.csv"
s <- read_delim(url,delim=',')
```

```
## Parsed with column specification:
## cols(
##   TimeStamp = col_datetime(format = ""),
##   SolarWatt = col_double()
## )
```

```r
mean(s$SolarWatt)
```

```
## [1] 193.2395
```

```r
min(s$SolarWatt)
```

```
## [1] 0
```

```r
max(s$SolarWatt)
```

```
## [1] 1457
```

Monthly average

```r
s %>%
  group_by(month(TimeStamp)) %>%
  summarize(meanMonth = mean(SolarWatt)*24/1000)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 12 x 2
##    `month(TimeStamp)` meanMonth
##                 <dbl>     <dbl>
##  1                  1      2.89
##  2                  2      3.61
##  3                  3      4.63
##  4                  4      6.42
##  5                  5      6.26
##  6                  6      6.51
##  7                  7      6.16
##  8                  8      4.91
##  9                  9      5.05
## 10                 10      4.27
## 11                 11      2.60
## 12                 12      2.29
```

e.

Assuming the total area for capturing solar energy by PV cells is 25m2 and panels are 20% efficient. How much electricity will be generated in a day?

```r
round(mean(s$SolarWatt)*24/1000*25*.2,2)
```

```
## [1] 23.19
```

Findings

About 25kW will be generated each day. More in summer than winter.