Energy Informatics R3 - Time series

Author: Richard T. Watson

Date: 2020-03-07

```r
library(readr)
library(dplyr)
library(ggplot2)
library(eia)
library(EIAdata)
```

```r
url <- "~/Dropbox/R/EIAAPIkey.txt"
key <- read_delim(url, delim = ',')
eia_set_key(key$key, store = c("env", "options", "sysenv"))
```

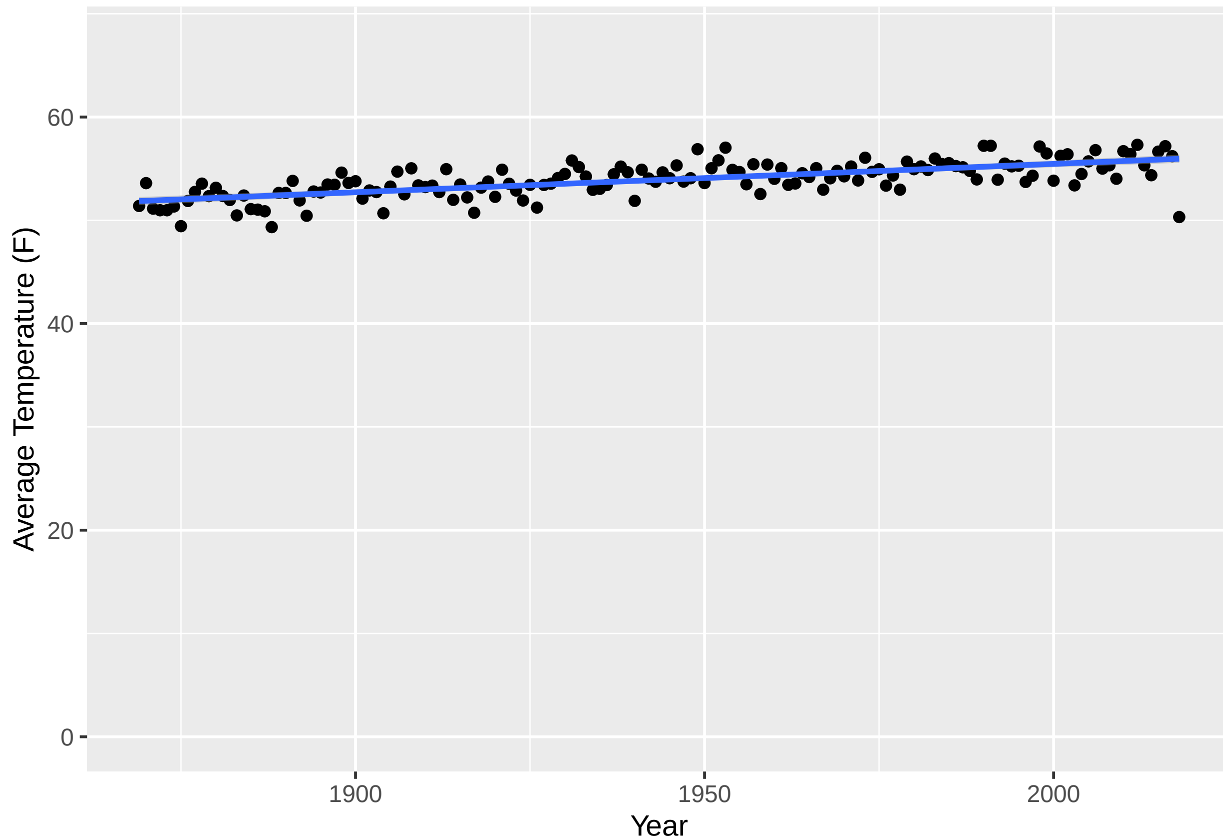
a.

Read the temperature data for Central Park. Compute the average temperature for each year, and create a scatter graph with a linear regression line.

```r
url <- "https://www.richardtwatson.com/data/centralparktemps.txt"
t <- read_delim(url, delim = ',')
```

```
## Parsed with column specification:
## cols(
##   year = col_double(),
##   month = col_double(),
##   temperature = col_double()
## )
```

```r
a <- t %>%
  group_by(year) %>%
  summarize(mean = mean(temperature))
ggplot(a,aes(year,mean)) +
  geom_point() +
  geom_smooth(method = lm) +
  ylim(0,max(a$mean + 10)) +
  xlab('Year') +
  ylab('Average Temperature (F)')
```



b.

Is the linear relationship between mean temperature and year significant? What do you conclude?

```r
mod <- lm(a$mean ~ a$year)
summary(mod)
```

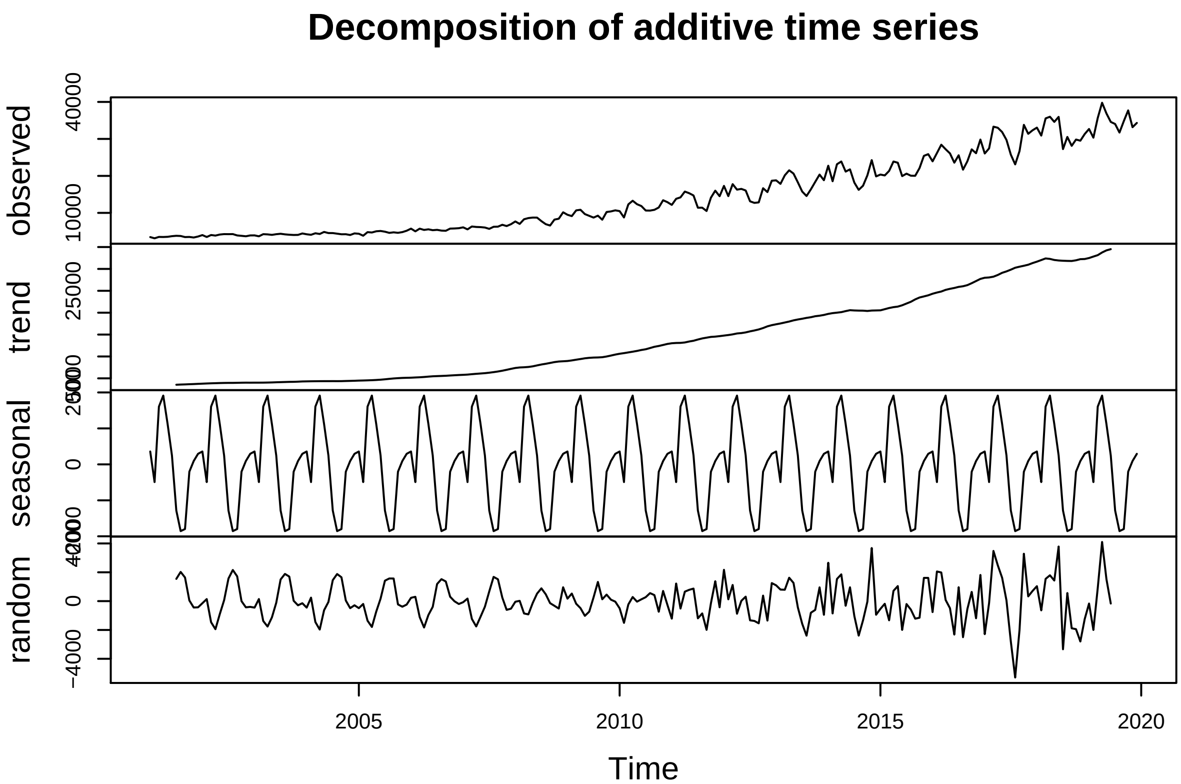
```
##
## Call:
## lm(formula = a$mean ~ a$year)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6393 -0.6568  0.1168  0.7927  2.8621
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.548316   4.496423   0.122    0.903
## a$year      0.027457   0.002313  11.871   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.227 on 148 degrees of freedom
## Multiple R-squared:  0.4877, Adjusted R-squared:  0.4843
## F-statistic: 140.9 on 1 and 148 DF,  p-value: < 2.2e-16
```

- As the p-value is less than .05, conclude that there is a significan linear relationship between temperature and year.
- The linear relationship explains 48% of the variation in temperature.

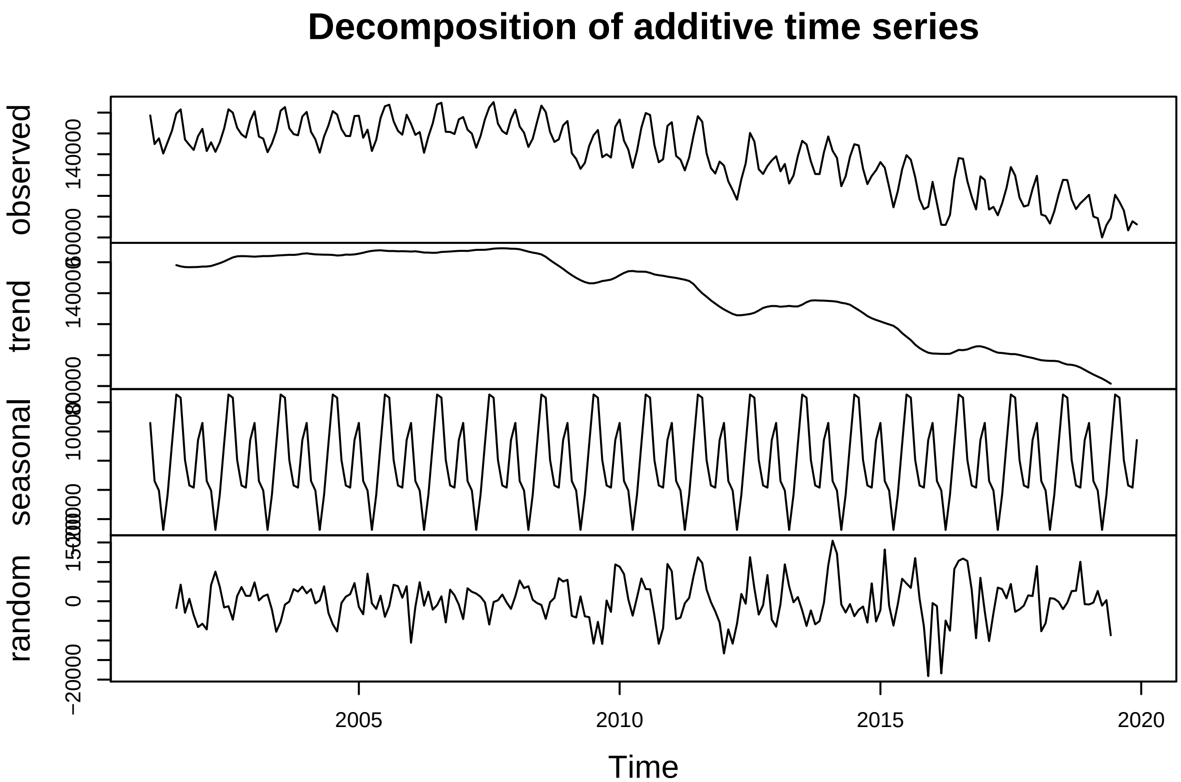
c.

Read the EIA data on electricity generated by renewables (ELEC.GEN.AOR-US-98.M). Create a time series for monthly MWh. Decompose the time series. What are your conclusions?

```
id <- "ELEC.GEN.AOR-US-98.M"
d <- eia_series(id)
d1 <-  d$data[[1]] %>% arrange(date)
ts <-  ts(d1$value,start=c(2001),frequency=12)
d2 <-  decompose(ts)
plot(d2)
```




```
id <- "ELEC.GEN.COW-US-99.M"
d <- eia_series(id)
d1 <-  d$data[[1]] %>% arrange(date)
ts <-  ts(d1$value,start=c(2001),frequency=12)
d2 <-  decompose(ts)
plot(d2)
```



Conclusion: Decomposition shows a long-term upward trend for renewables and a downward trend for coal starting around 2008. Both have a seasonal effect, but the causes are different. You can control the seasonality of burning coal for electricity, but the seasonality of renewables is determined by a locality's climate.