R3 - Time series analysis

Richard T. Watson

2020-12-27 10:23:33

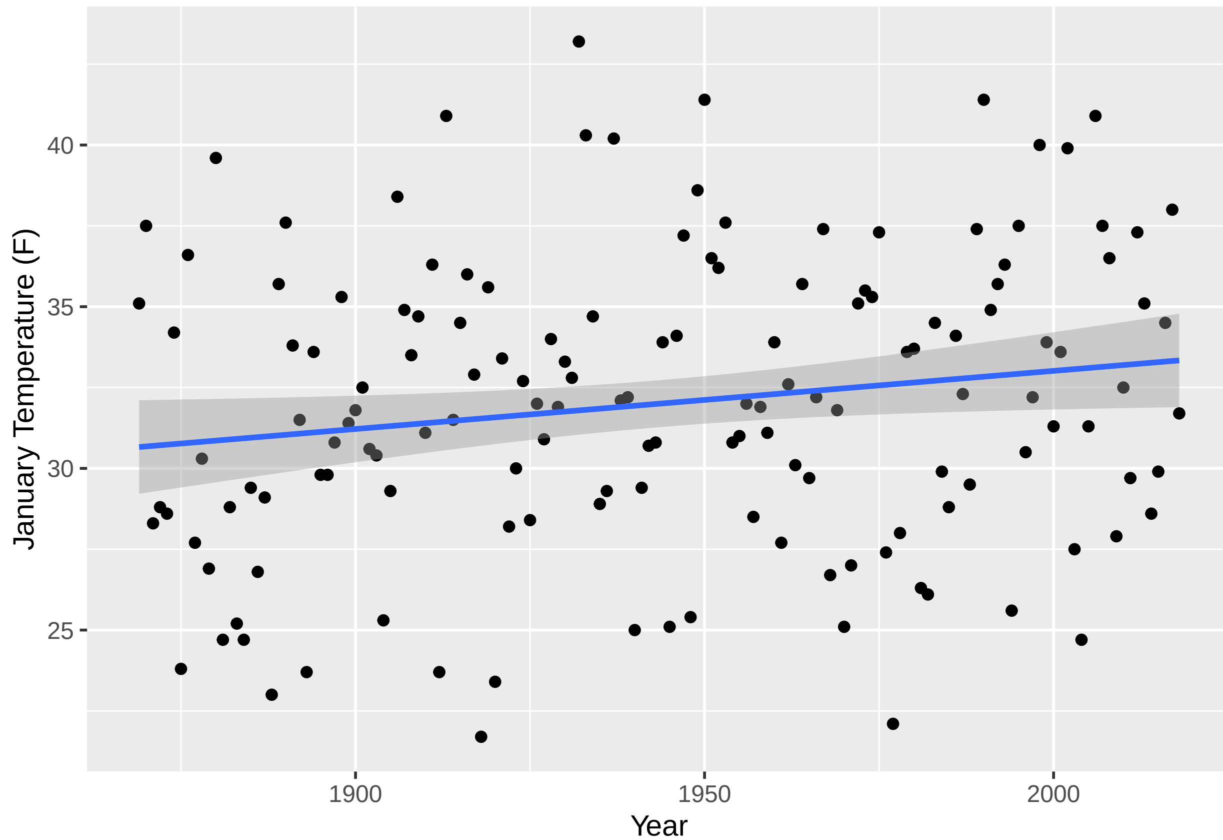
```
library(tidyverse)
library(eia)
library(EIAdata)
library(car)
```

```
url <- "~/Dropbox/R/API keys/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. Extract the mean temperature for January, and create a scatter graph with a linear regression line for temperature and year.

```
url <- "https://www.richardtwatson.com/data/centralparktemps.txt"
t <- read_delim(url, delim = ',')
jan <- t %>%
  select(month, year, temperature) %>%
  filter(month == 1)
ggplot(jan,aes(year,temperature)) +
  geom_point() +
  geom_smooth(method = lm) +
  xlab('Year') +
  ylab('January Temperature (F)')
```



b.

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

```
mod <- lm(jan$temperature ~ jan$year)
summary(mod)
```

```
##
## Call:
## lm(formula = jan$temperature ~ jan$year)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5021  -2.9174  -0.0871   3.2431  11.4067
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.932726  16.510240  -0.178    0.859
## jan$year     0.017974   0.008493   2.116    0.036 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.504 on 148 degrees of freedom
## Multiple R-squared:  0.02937,    Adjusted R-squared:  0.02282
## F-statistic: 4.479 on 1 and 148 DF,  p-value: 0.03599
```

```r
durbinWatsonTest(mod)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1      0.03097258      1.930597   0.582
##  Alternative hypothesis: rho != 0
```

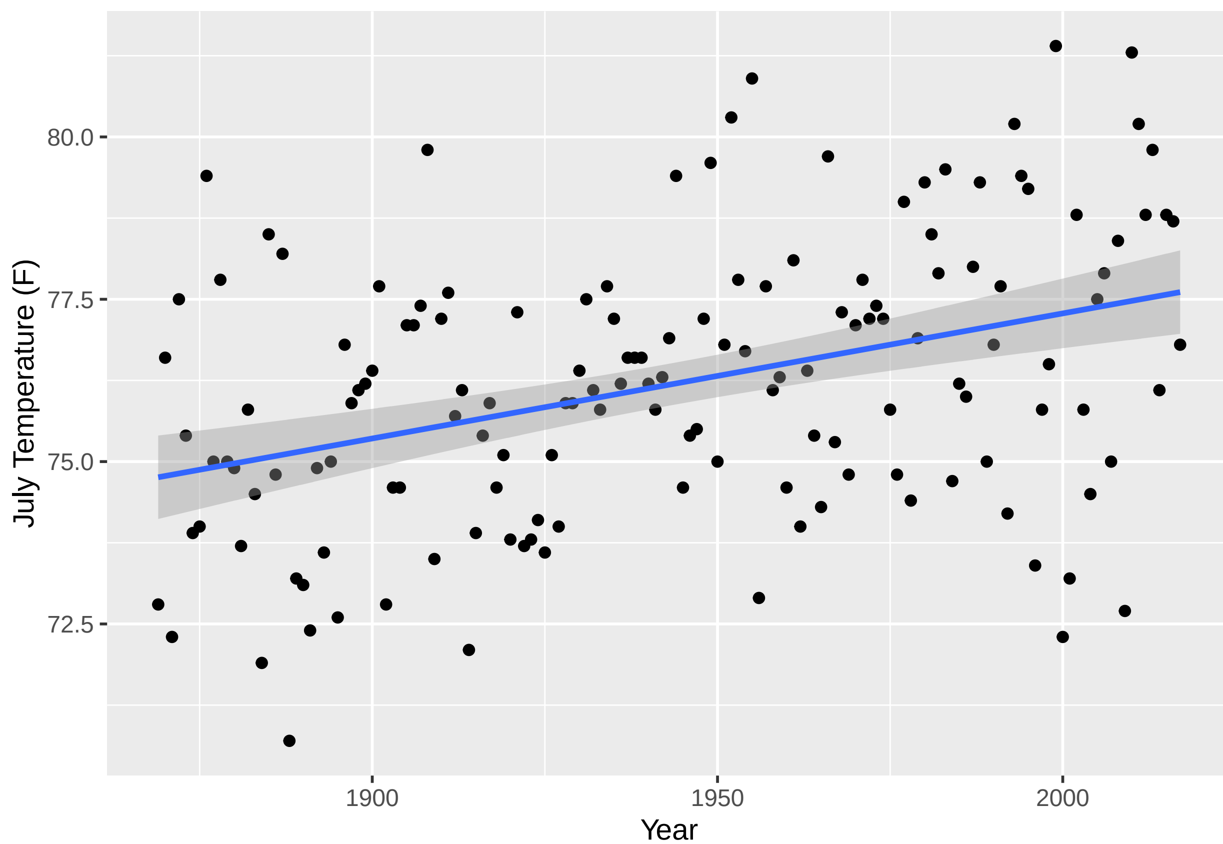
Conclusion

- As the p-value is less than .05, conclude that there is a significant linear relationship between January's mean temperature and year.
- The linear relationship explains 2% of the variation in temperature.
- January's mean temperature in Central Park is increasing by 0.018º F per year
- There is no autocorrelation of the residuals (p = 0.602), so model is OK.

c.

Read the temperature data for Central Park. Extract the mean temperature for July, and create a scatter graph with a linear regression line for temperature and year.

```r
url <- "https://www.richardtwatson.com/data/centralparktemps.txt"
t <- read_delim(url, delim = ',')
jul <- t %>%
  select(month, year, temperature) %>%
  filter(month == 7)
ggplot(jul,aes(year,temperature)) +
  geom_point() +
  geom_smooth(method = lm) +
  xlab('Year') +
  ylab('July Temperature (F)')
```



d.

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

```
mod <- lm(jul$temperature ~ jul$year)
summary(mod)
```

```
##
## Call:
## lm(formula = jul$temperature ~ jul$year)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.9825 -1.4522  0.0867  1.2864  4.5060
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.758288   7.381571   5.251 5.21e-07 ***
## jul$year     0.019262   0.003798   5.071 1.17e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.994 on 147 degrees of freedom
## Multiple R-squared:  0.1489, Adjusted R-squared:  0.1431
## F-statistic: 25.72 on 1 and 147 DF,  p-value: 1.172e-06
```

```
durbinWatsonTest(mod)
```

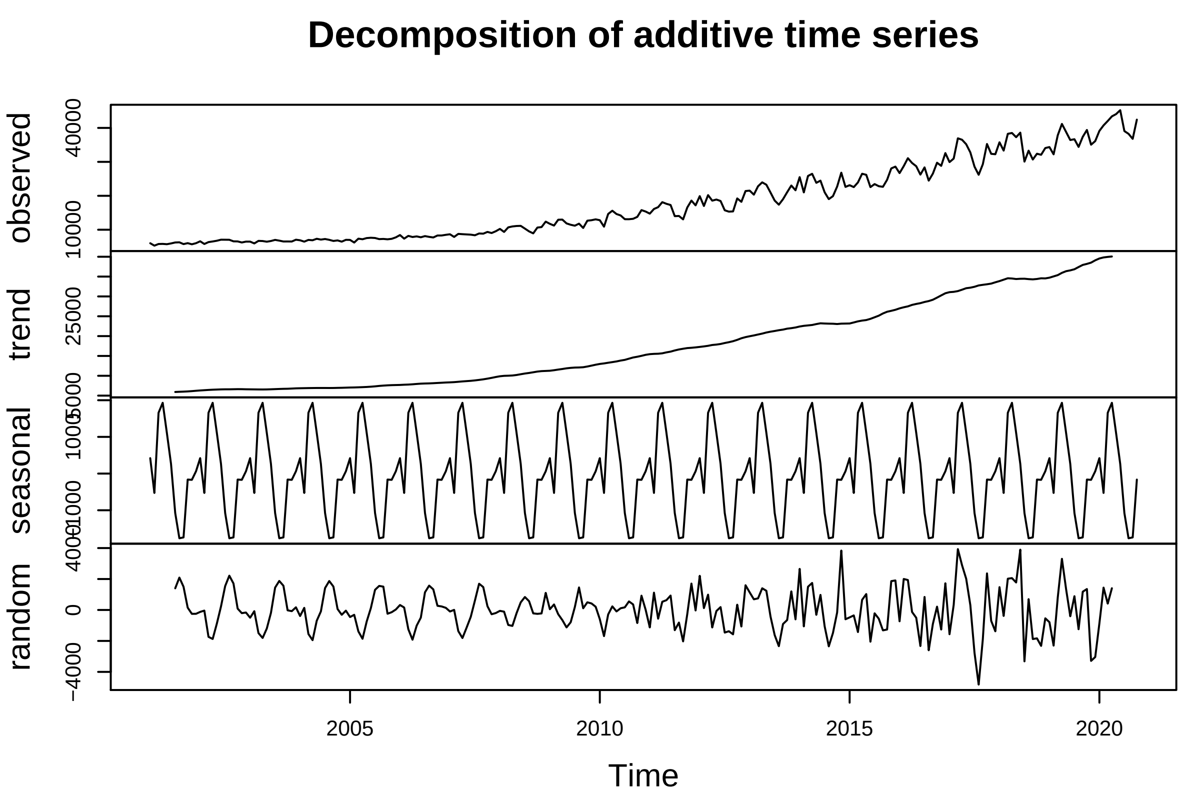
```
##  lag Autocorrelation D-W Statistic p-value
##    1      -0.1034269      2.199165   0.268
##  Alternative hypothesis: rho != 0
```

- As the p-value is less than .05, conclude that there is a significant linear relationship between July's mean temperature and year.
- The linear relationship explains 14% of the variation in temperature.
- July's mean temperature in Central Park is increasing by 0.019º F per year
- There is no autocorrelation of the residuals (p = 0.264), so model looks OK.

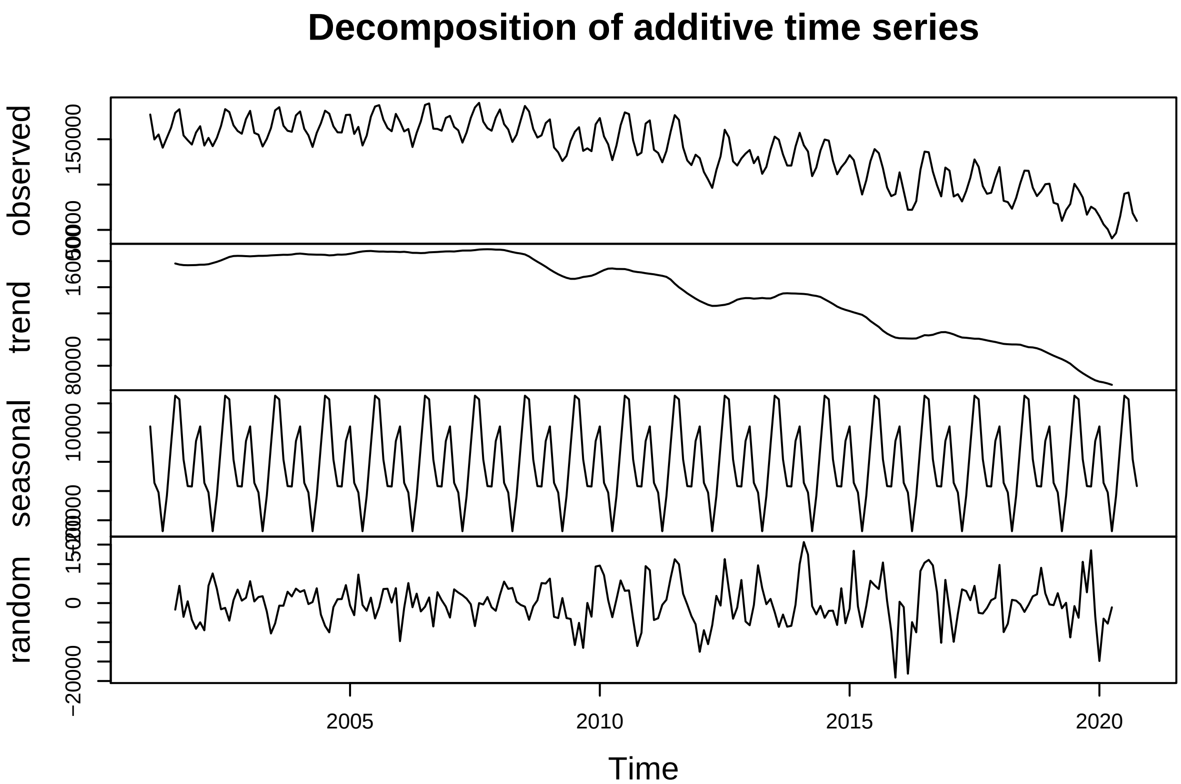
e.

Read the US EIA data on electricity generated by renewables (ELEC.GEN.AOR-GA-99.M). Create a time series for monthly MWh. Decompose the time series. What are your conclusions? Do the same for electricity generated by coal (ELEC.GEN.COW-GA-99.M). What are your conclusions?

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



```r
id <- "ELEC.GEN.COW-US-99.M"
d <- eia_series(id)
d1 <-  d$data[[1]] %>% arrange(date)
ts <-  ts(d1$value,start=c(2001,1),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. Renewables can't adjust to demand, but coal can, which is why we need a solution to the intermittency problem.