Are carbon dioxide concentrations and global surface temperatures correlated?

Nomdeploom

2018-11-01

One of the claims often made by those who would deny climate change is that changes in global temperature are not related to changes in levels of carbon dioxide. While we often see graphs showing changes in temperature over time, or changes in carbon dioxide levels over time, we seldom if ever see analyses of the relationship between carbon dioxide levels and temperature, taking time out of the mix.

In this post, I'm going to try to show how temperature and carbon dioxide are correlated, ignoring time.

What is correlation and how do we measure it?

In its broadest sense, correlation is a relationship or a connection between two or more things. In a statistical sense, we try to find out to what extent two or more quantities change together.

A measure of correlation is a numerical measure of the relationship between 2 variables (X and Y), where the measurements are taken in pairs - one measurement for each variable. By convention, such measures are crafted so that they meet the following requirements:

- They lie between -1 and 1;
- If larger values of X tend to group with larger values of Y and vice versa, the measure will be positive. If larger values of X tend to group with smaller values of Y and vice versa, the measure will be negative;
- The stronger the tendency to group large with large and small with small, the closer the measure will be to 1; and
- The stronger the tendency to group large with small, the closer the measure will be to -1.

Our process will consist of two phases: First, we will perform an exploratory review of the data to see if possible correlation surfaces. Second, we will calculate various correlation measures and see if they indicate correlation.

```
## [1] 140  19
```

```
## [1] 60 19
```

```
##    Year   Jan     Feb     Mar     Apr     May    Jun    Jul    Aug    Sep    Oct    Nov    Dec
## 80 "1959" " 0.08" " 0.09" " 0.18" " 0.13" ".04"  ".02"  ".06"  "-.02" "-.06" "-.09" "-.09" "-.01"
## 81 "1960" " 0.00" " 0.16" "-0.34" "-0.14" "-.07" "-.04" "-.03" ".01"  ".07"  ".07"  "-.12" ".19"
## 82 "1961" " 0.07" " 0.19" " 0.09" " 0.12" ".11"  ".11"  ".00"  ".03"  ".07"  ".00"  ".03"  "-.15"
## 83 "1962" " 0.07" " 0.14" " 0.12" " 0.05" "-.04" ".05"  ".02"  "-.01" ".01"  ".00"  ".07"  ".00"
## 84 "1963" "-0.02" " 0.20" "-0.14" "-0.06" "-.04" ".05"  ".08"  ".27"  ".20"  ".15"  ".15"  "-.01"
## 85 "1964" "-0.08" "-0.13" "-0.23" "-0.31" "-.24" "-.02" "-.03" "-.21" "-.29" "-.31" "-.21" "-.30"
##    J.D    D.N    DJF    MAM    JJA    SON
## 80 ".03"  ".03"  ".06"  ".12"  ".02"  "-.08"
## 81 "-.02" "-.04" ".05"  "-.18" "-.02" ".01"
## 82 ".06"  ".08"  ".15"  ".11"  ".05"  ".03"
## 83 ".04"  ".03"  ".02"  ".04"  ".02"  ".03"
## 84 ".07"  ".07"  ".06"  "-.08" ".14"  ".17"
## 85 "-.20" "-.17" "-.07" "-.26" "-.09" "-.27"
```

```
##     Year   Jan     Feb     Mar     Apr     May    Jun    Jul    Aug     Sep    Oct     Nov     Dec
## 134 "2013" " 0.67" " 0.56" " 0.66" " 0.53" ".58" ".66" ".58" ".67"  ".77" ".67"  ".78"  ".65"
## 135 "2014" " 0.73" " 0.52" " 0.76" " 0.77" ".85" ".66" ".56" ".81"  ".88" ".81"  ".66"  ".77"
```

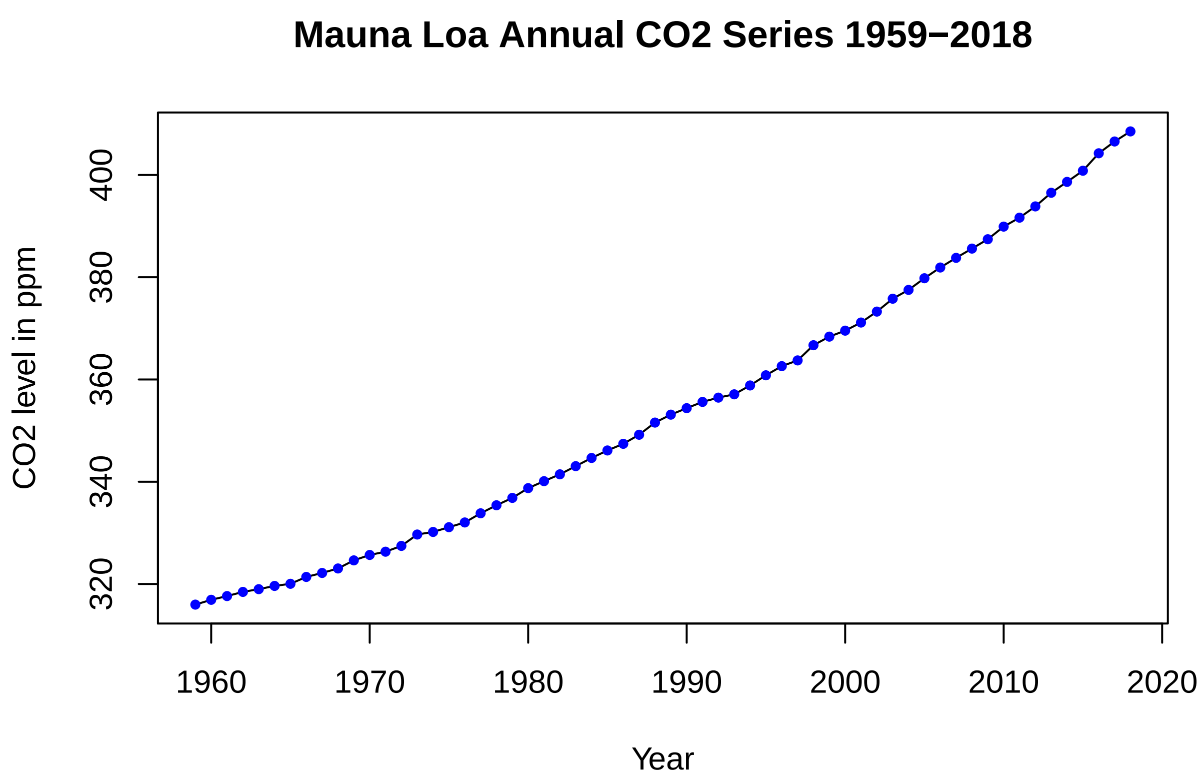
```
## 136 "2015" " 0.81" " 0.87" " 0.90" " 0.75" ".75" ".79" ".71" ".78"  ".81" "1.07" "1.03" "1.10"
## 137 "2016" " 1.15" " 1.35" " 1.30" " 1.07" ".91" ".77" ".82" "1.01" ".88" ".90"  ".91"  ".83"
## 138 "2017" " 0.97" " 1.13" " 1.12" " 0.92" ".89" ".69" ".82" ".87"  ".76" ".88"  ".86"  ".88"
## 139 "2018" " 0.77" " 0.84" " 0.91" " 0.87" ".81" ".75" ".78" ".73"  ".76" ".98"  ".78"  ".89"
##     J.D   D.N    DJF    MAM    JJA   SON
## 134 ".65" ".64"  ".58"  ".59"  ".63" ".74"
## 135 ".73" ".72"  ".63"  ".79"  ".68" ".79"
## 136 ".87" ".84"  ".82"  ".80"  ".76" ".97"
## 137 ".99" "1.01" "1.20" "1.09" ".87" ".90"
## 138 ".90" ".89"  ".98"  ".98"  ".79" ".83"
## 139 ".82" ".82"  ".83"  ".86"  ".75" ".84"

## [1] 60
```

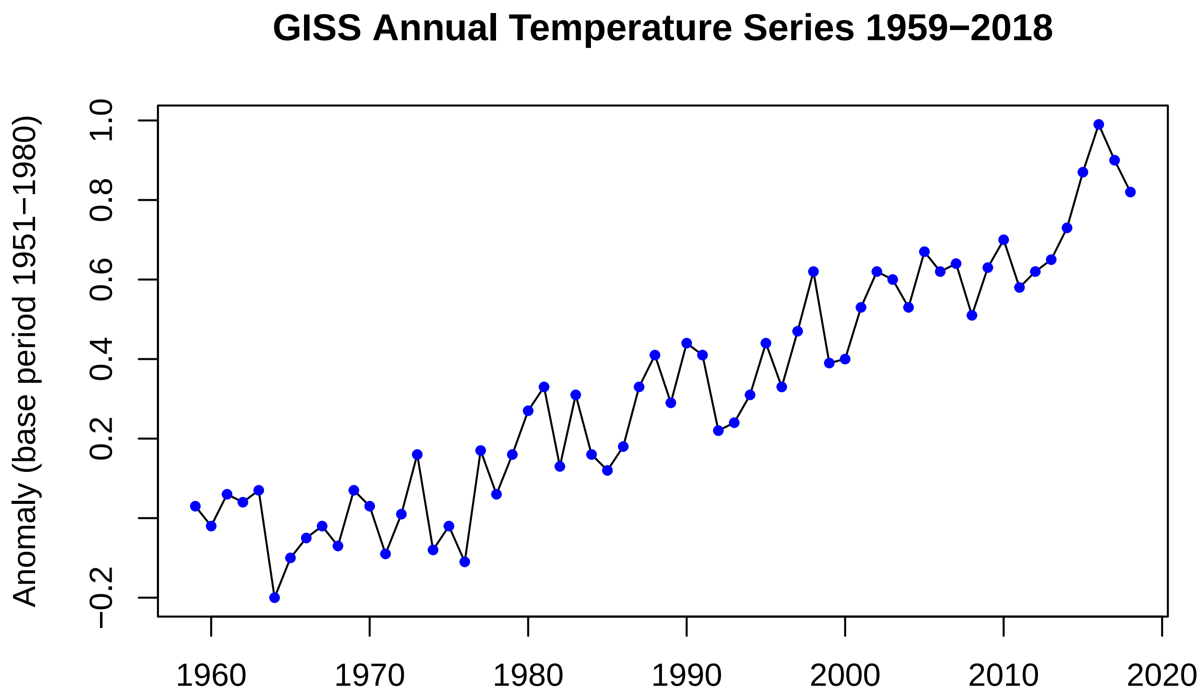
Exploratory review of the time series and their relationship

Generally, we will first look to graphical representations of the data pairs to see if the two variables pair more or less in one of the ways described above (large with large or large with small). Side-by-side time series can reveal similar patterns over time. In addition, scatterplots of the data pairs can reveal the correlation relationship.

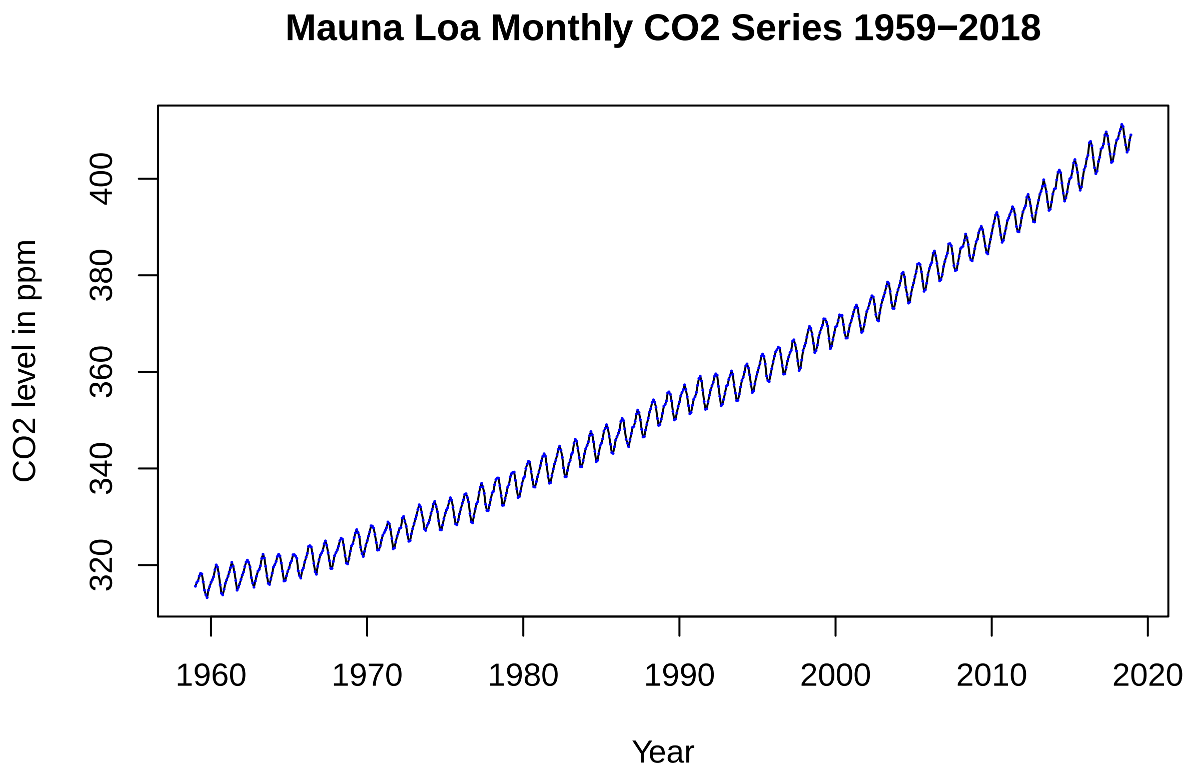
Our first step is to explore the data series. The first thing we can do is to look at a chart of the annual CO_2 data next to the annual temperature data and the monthly CO_2 data next to the monthly temperature data.

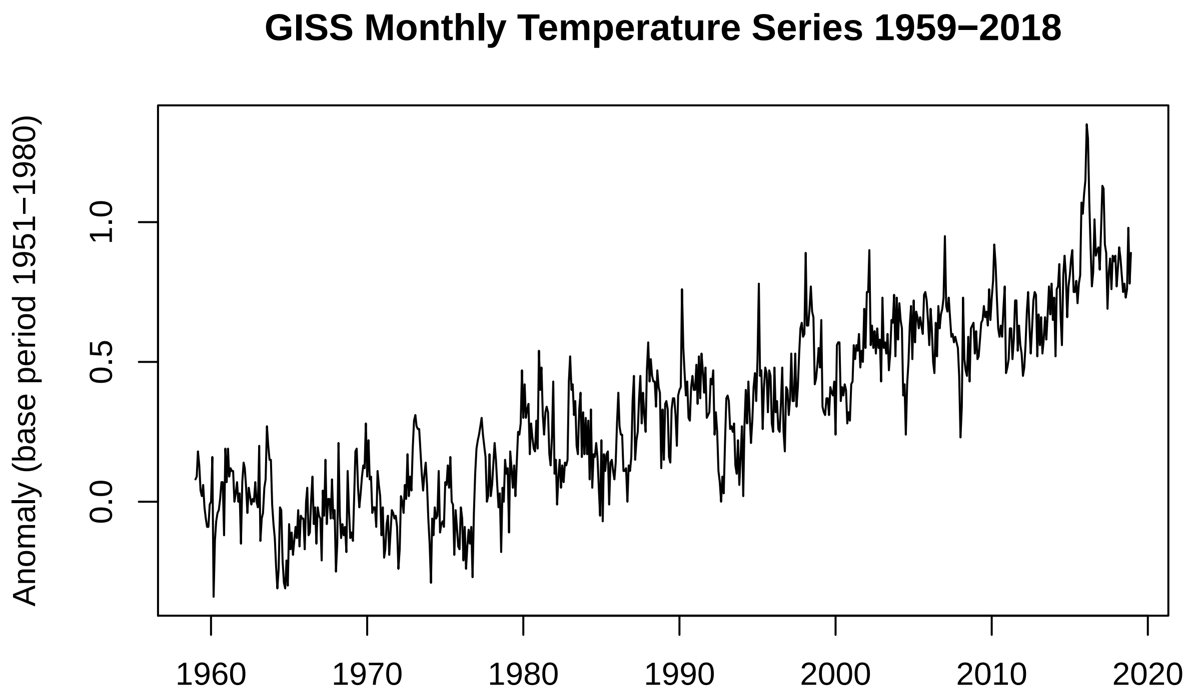
First, we look at the annual data. We can see that both of the series show a relatively steady increase over time. The carbon dioxide series is much smoother, though both are increasing in roughly similar fashion on average over time.





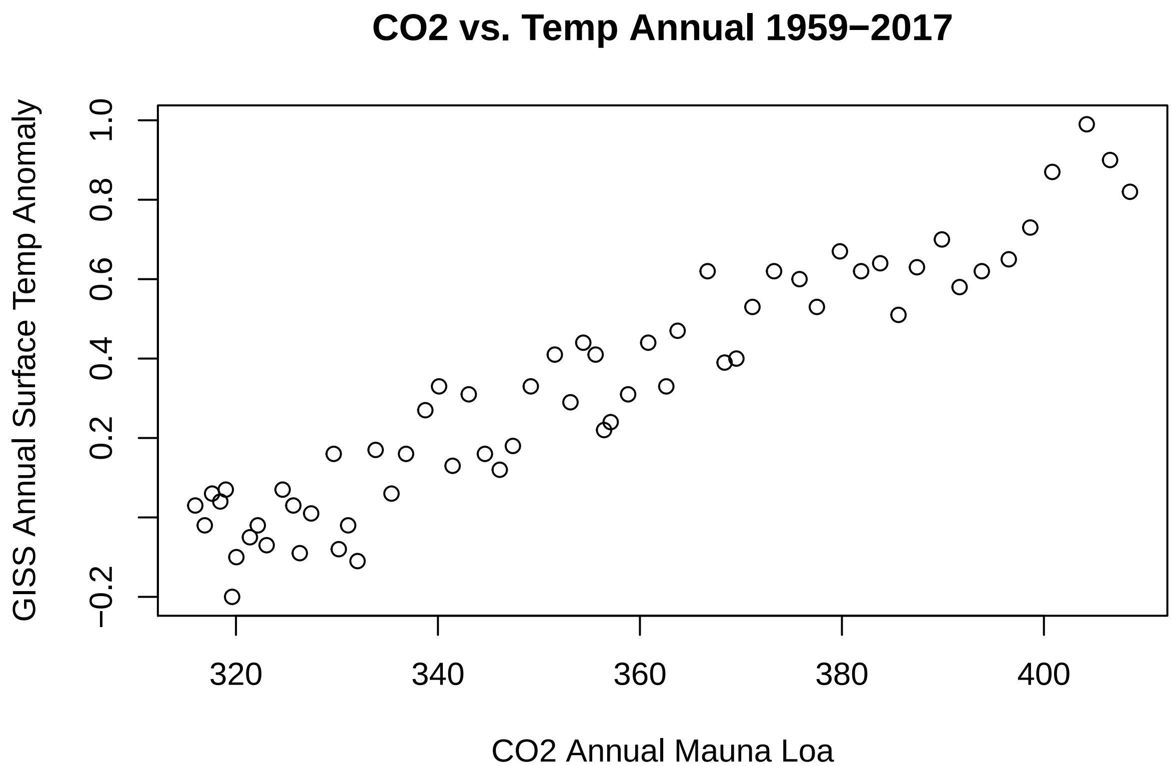
Then we can look at the two monthly series. Obviously, monthly series are more jagged than annual. We can see that the monthly carbon dioxide series reveals a seasonal fluctuation that reflects higher concentrations during northern hemisphere winters. Nevertheless, both series still reflect the same general ascending pattern.

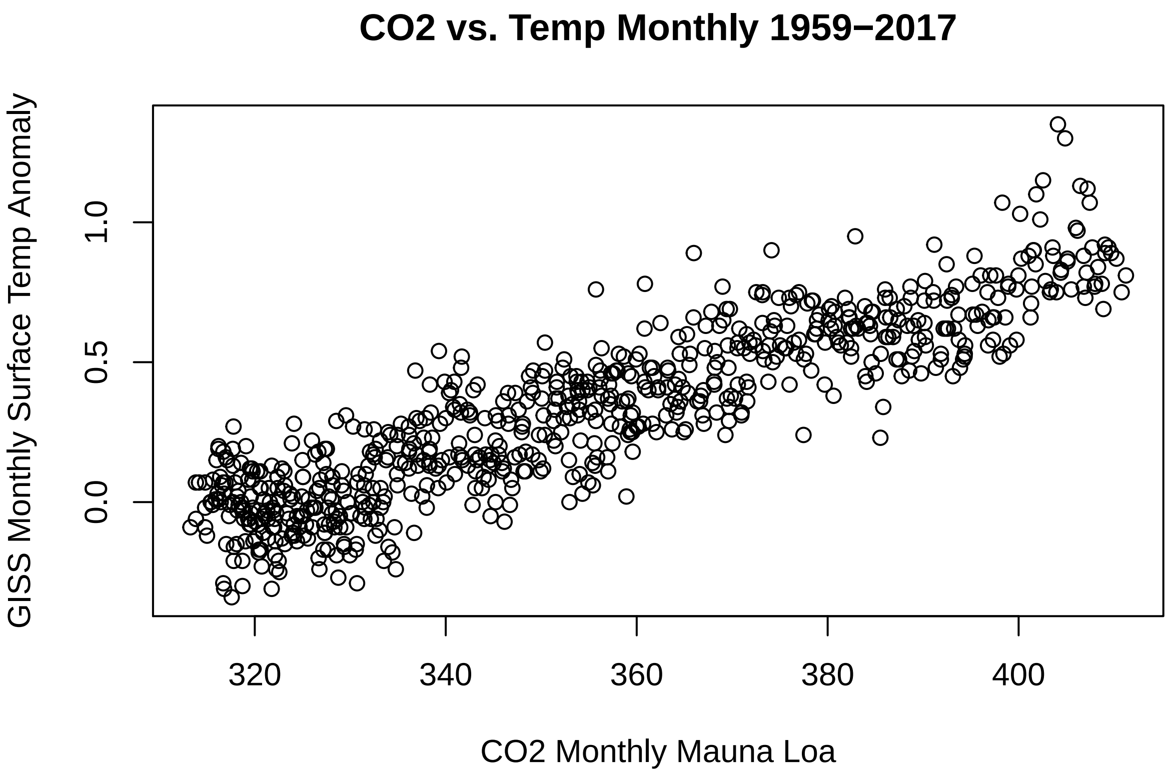

Mauna Loa Monthly CO2 Series 1959−2018
CO2 level in ppm
320
340
360
380
400
1960
1970
1980
1990
2000
2010
2020
Year



Scatterplots of carbon dioxide concentration vs. surface temperature

Scatterplots are different from the time series above. Here we will take each pair of measurements taken at the same time and will plot carbon dioxide concentration against surface temperature. We are looking to see if there is a linear relationship between the two numbers (in other words, we are looking to see if surface temperature increases when carbon dioxide concentration increases.) In both cases, we can see that there is a very strong linear relationship between the two variables.


CO2 vs. Temp Annual 1959−2017
GISS Annual Surface Temp Anomaly
−0.2
0.2
0.4
0.6
0.8
1.0
320
340
360
380
400
CO2 Annual Mauna Loa



Various correlation coefficients

Pearson's correlation coefficient:

$$\rho_{x,y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x \sigma_y}$$

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

Some say that Pearson's correlation coefficient requires the data to be normal. Others say it is not required but if the data is bivariate normal, "Pearson's correlation provides a complete description of the association."\

Spearman's Rho is Pearson's r computed on ranks and average ranks:

$$\rho = \frac{\sum_{i=1}^{n} R(X_i)R(Y_i) - n(\frac{n+1}{2})^2}{(\sum_{i=1}^{n} R(X_i)^2 - n(\frac{n+1}{2})^2)^{\frac{1}{2}}(\sum_{i=1}^{n} R(Y_i)^2 - n(\frac{n+1}{2})^2)^{\frac{1}{2}}}$$

Kendall's Tau is another measure of rank correlation. We are given a bivariate sample of size n in the form of (x_i, y_i) for $i = 1, 2, \ldots, n$. We compare each pair of observations, i.e., $\binom{n}{2}$ pairs of observations. We determine whether each pair is discordant or concordant, where concordant means that the two numbers in one member of the pair differ in the same direction from the two numbers in the other member of the pair. Discordant is the reverse. If there are no ties, Kendall's Tau is then:

$$\tau = \frac{N_c - N_d}{n(n-1)/2}$$

where N_c is the number of concordant pairs and N_d is the number of discordant pairs. If there are ties, we use:

$$\tau = \frac{N_c - N_d}{N_c + N_d}$$

where we consider only pairs where $x_i \neq x_j$.

Correlation coefficients CO_2 to Surface Temperature 1959-2017

We can calculate the Pearson, Spearman and Kendall correlation coefficients comparing the Mauna Loa carbon dioxide data to the GISS surface temperature data. All of the coefficients are in excess of 0.7. This is generally considered to be strong correlation.

	Pearson	Spearman	Kendall
Annual Data	0.950	0.935	0.780
Monthly Data	0.894	0.893	0.701

```
cor(mlann,gissann,method="pearson");cor(mlann,gissann,method="spearman");cor(mlann,gissann,method="kenda
```

```
## [1] 0.9502095

## [1] 0.9345806

## [1] 0.7800039
```

```
cor(mlmon,gissmo,method="pearson");cor(mlmon,gissmo,method="spearman");cor(mlmon,gissmo,method="kendall
```

```
## [1] 0.8940144

## [1] 0.8926873

## [1] 0.7011426
```

```
shapiro.test(mlann)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mlann
## W = 0.94085, p-value = 0.005913
```

```
shapiro.test(gissann)
```

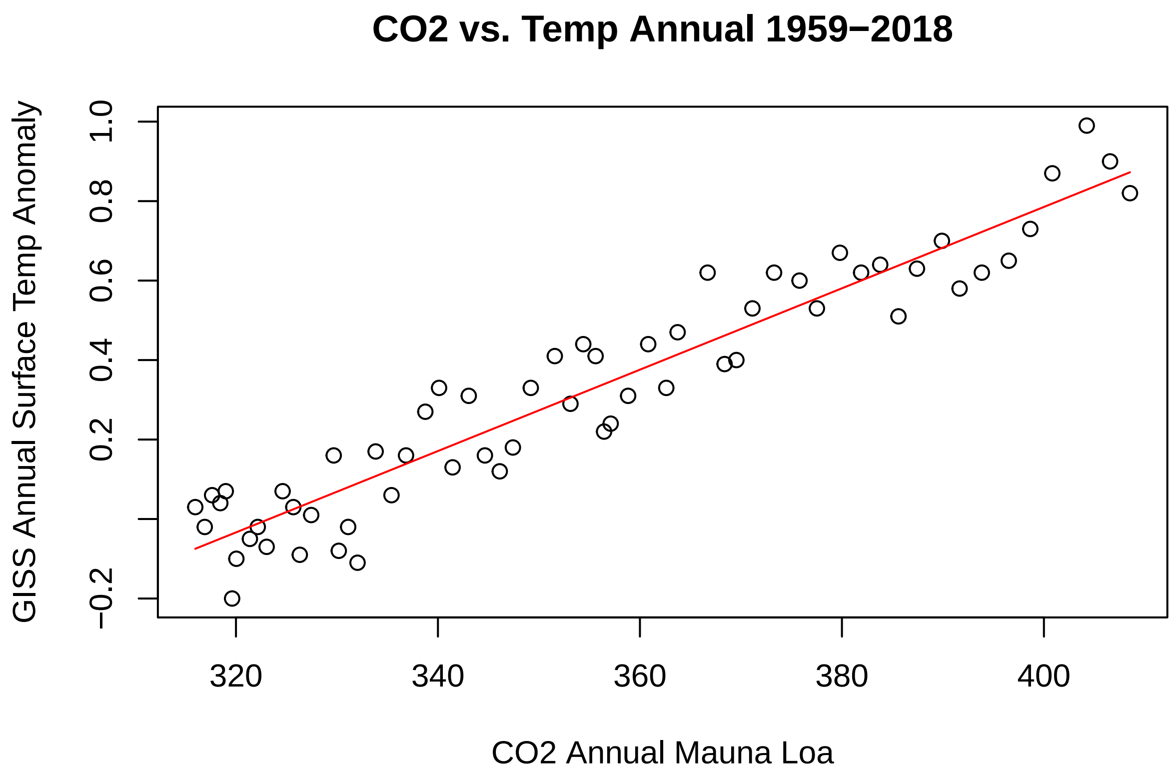
```
##
##  Shapiro-Wilk normality test
##
## data:  gissann
## W = 0.96443, p-value = 0.07768
```

```
shapiro.test(mlmon)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mlmon
## W = 0.94367, p-value = 6.686e-16
```

```r
shapiro.test(gissmo)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  gissmo
## W = 0.98126, p-value = 5.683e-08
```




```r
#Properties of the regression line (the red trend line above)
summary(lm1)
```

```
##
## Call:
## lm(formula = gissann ~ mlann)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.199608 -0.068512 -0.001482  0.076902  0.175568
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.3095742  0.1566912  -21.12   <2e-16 ***
## mlann        0.0102373  0.0004408   23.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

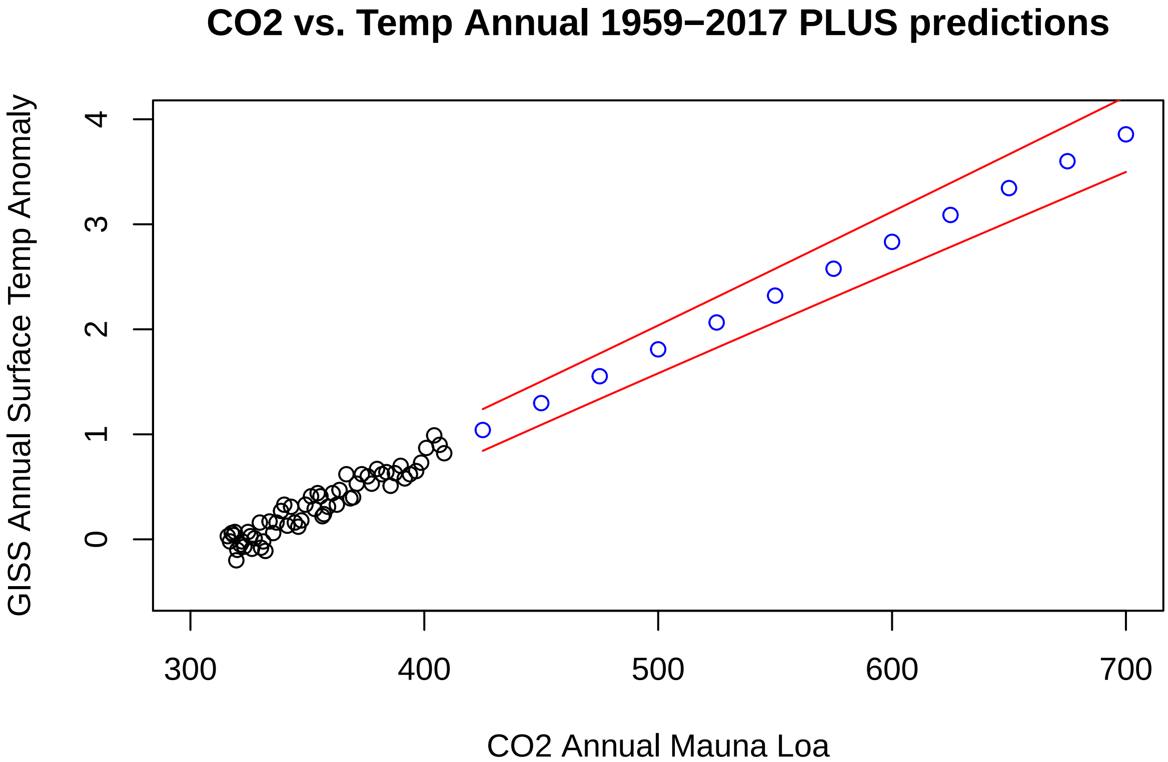
```
## Residual standard error: 0.09333 on 58 degrees of freedom
## Multiple R-squared:  0.9029, Adjusted R-squared:  0.9012
## F-statistic: 539.3 on 1 and 58 DF,  p-value: < 2.2e-16
```

Can we predict surface temperature increases from anticipated carbon dioxide increases?

```
#Predictions for CO2 concentrations of 425 to 700, increments of 25
new <- data.frame(mlann = seq(425,700,by=25))
predictions <- predict(lm1,newdata=new,interval="prediction")
round(predictions,2)
```

```
##     fit  lwr  upr
## 1  1.04 0.84 1.24
## 2  1.30 1.09 1.50
## 3  1.55 1.34 1.77
## 4  1.81 1.58 2.04
## 5  2.06 1.82 2.31
## 6  2.32 2.07 2.58
## 7  2.58 2.31 2.85
## 8  2.83 2.55 3.12
## 9  3.09 2.78 3.39
## 10 3.34 3.02 3.67
## 11 3.60 3.26 3.94
## 12 3.86 3.50 4.21
```

```
#plot(c(mlann,seq(425,700,25)),c(gissann,predictions[,1]),
plot(mlann,gissann,
     main="CO2 vs. Temp Annual 1959-2017 PLUS predictions",
     xlab="CO2 Annual Mauna Loa",ylab="GISS Annual Surface Temp Anomaly",
     xlim=c(300,700),ylim=c(-.5,4))
points(seq(425,700,25),predictions[,1],col="blue")
lines(seq(425,700,25),predictions[,2],type="l",col="red")
lines(seq(425,700,25),predictions[,3],type="l",col="red")
```


CO2 vs. Temp Annual 1959−2017 PLUS predictions
GISS Annual Surface Temp Anomaly
0
1
2
3
4
300
400
500
600
700
CO2 Annual Mauna Loa