Lab 2: Managing uncertainty

Theresa Ong
12/27/2019

Coupled Human-Natural Systems

ENVS 80.10/ EEES 181 Winter 2020

Dartmouth College

Due: Wednesday 1/23/2020 at 4:30pm to Blackboard

Your submission should be a single PDF or Word file (no need to submit your code). A printed copy of the answer key will be made available before the deadline. You will self-assess your assignment and post your edited version of the lab exercises to Blackboard by the due date.

Please consult the syllabus for any accessibility concerns.

This week we discussed the difficulties of managing uncertainty in socio-ecological systems. In this lab we will take a closer look into some of these uncertainties using data on Climate Change as an example.

Setup

First, we need some data.

We will install the R package dslabs, which has compiled temperature and greehouse gas emission data from various sources including the The National Oceanic and Atmospheric Administration: https://www.ncdc.noaa.gov/cdo-web/

```
install.packages("dslabs")

update.packages("dslabs")

library(dslabs)
```

We can also setup a colorblind safe color palette:

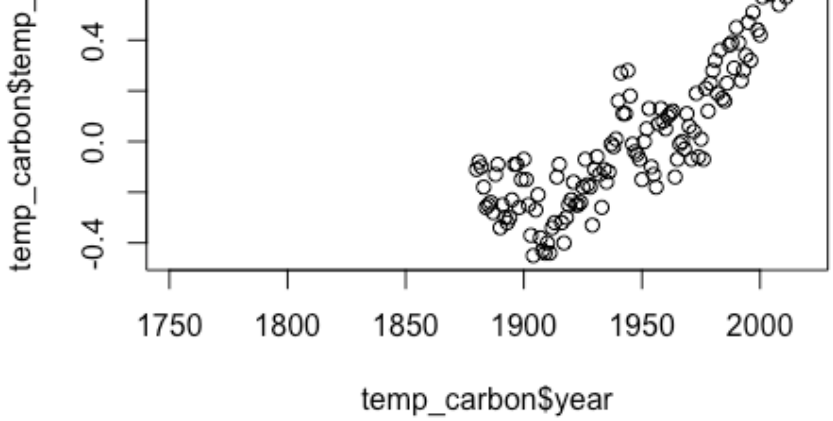
```
# set colorblind-friendly color palette
cb_palette <- c("black", "#E69F00", "#56B4E9", "#009E73",
                "#CC79A7", "#F0E442", "#0072B2", "#D55E00")
```

We will be using three datasets available in the dslabs package:

```
data(temp_carbon)
data(historic_co2)
data(greenhouse_gases)
head(temp_carbon)
##   year temp_anomaly land_anomaly ocean_anomaly carbon_emissions
## 1 1880       -0.11       -0.48        -0.01              236
## 2 1881       -0.08       -0.40         0.01              243
## 3 1882       -0.10       -0.48         0.00              256
## 4 1883       -0.18       -0.66        -0.04              272
## 5 1884       -0.26       -0.69        -0.14              275
## 6 1885       -0.25       -0.56        -0.17              277
```

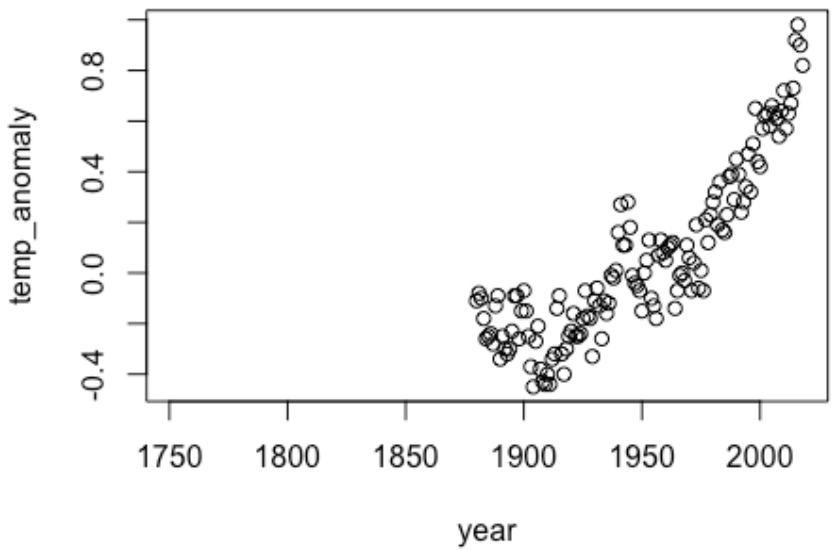
The head function produces a chunk of the dataset so we can easily see the format and headers of each column. Let's first plot temperature anomalies by year:

```
plot(temp_carbon$year,temp_carbon$temp_anomaly)
```



The $ tells R to look into temp_carbon for a column called year. Some but not all functions in R will let you avoid this by specifiying a dataset in the function, but plot does not. We can use the attach and detach functions to avoid using $ call here:

```
attach(temp_carbon)
plot(year,temp_anomaly)
```




```
detach(temp_carbon)
```

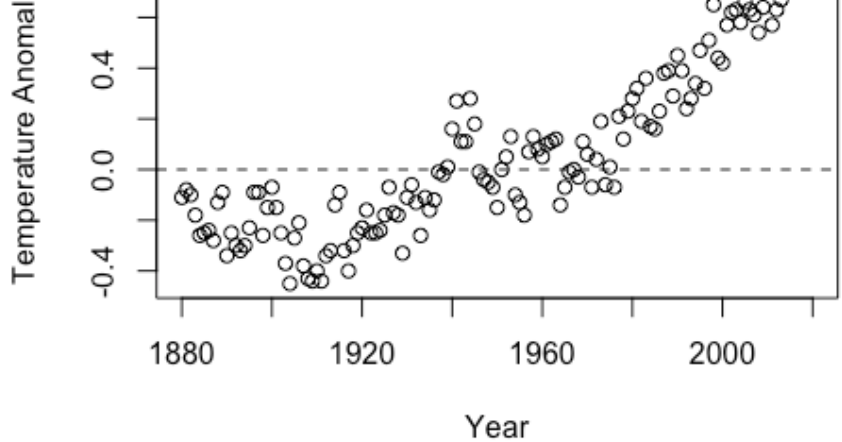
Care must be taken when using the attach/detach function because if you forget to detach the data you can end up mixing up or rewritting datasets that you don't intend to. How you choose to run your code will come down to personal preference.

Let's clean up this graph a bit by adding axes labels and including a 0 line. First, we need to know what the units are. Find it by looking up the help file for the dataset:

```
?temp_carbon
```

To get the ° symbol, we use the expression function and a paste function to combine text and symbols:

```
plot(temp_carbon$year,temp_carbon$temp_anomaly, xlab="Year",
   ylab=expression(paste("Temperature Anomaly (",degree,"C)")),xlim=c(1880,2020))
abline(h=0,lty=2)
```

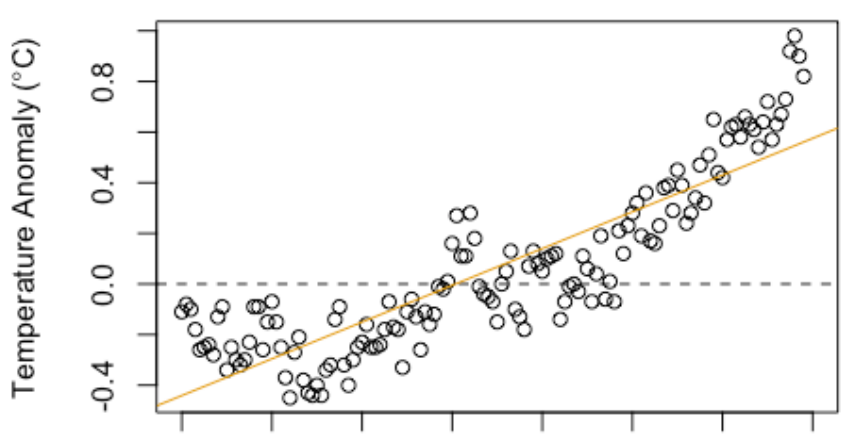


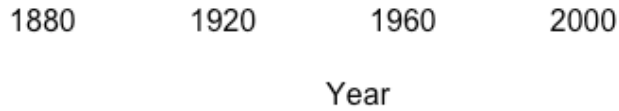
How do we know if temperature anomalies (deviations from mean global temperatures in the 20th century) have significantly changed over time? Let's start with a simple linear regression using the lm function and summarize the results with the summary function to see if there is any relationship between these two variables:

```
model1<-lm(temp_anomaly~year,data=temp_carbon)
summary(model1)
##
## Call:
## lm(formula = temp_anomaly ~ year, data = temp_carbon)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32598 -0.11846 -0.01062  0.11185  0.43367
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.409e+01  6.750e-01  -20.87   <2e-16 ***
## year         7.259e-03  3.463e-04   20.96   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1638 on 137 degrees of freedom
##   (129 observations deleted due to missingness)
## Multiple R-squared:  0.7623, Adjusted R-squared:  0.7606
## F-statistic: 439.4 on 1 and 137 DF,  p-value: < 2.2e-16
```

This output gives us a summary of the linear regression, showing a signiciant relationship with a Pr(>|t|) value close to 0. The positive and large Estimate indicates a strong and positive relationship between year and temperature anomalies. The Adjusted R-squared value tells us that about 76% of the variation in temperature anomalies can be explained by year. Let's add the regression line to our plot:

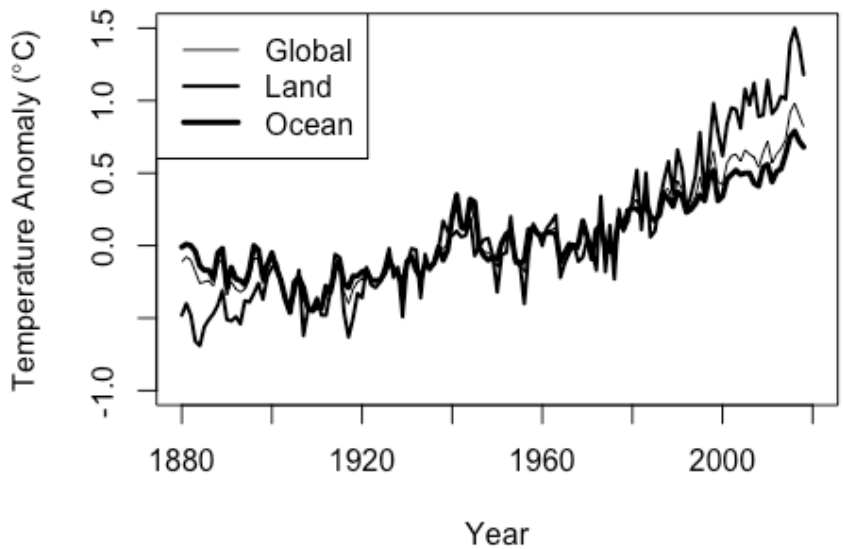
```
plot(temp_carbon$year,temp_carbon$temp_anomaly, xlab="Year",
   ylab=expression(paste("Temperature Anomaly (",degree,"C)")),xlim=c(1880,2020))
abline(h=0,lty=2)
abline(model1,col=cb_palette[2])
```





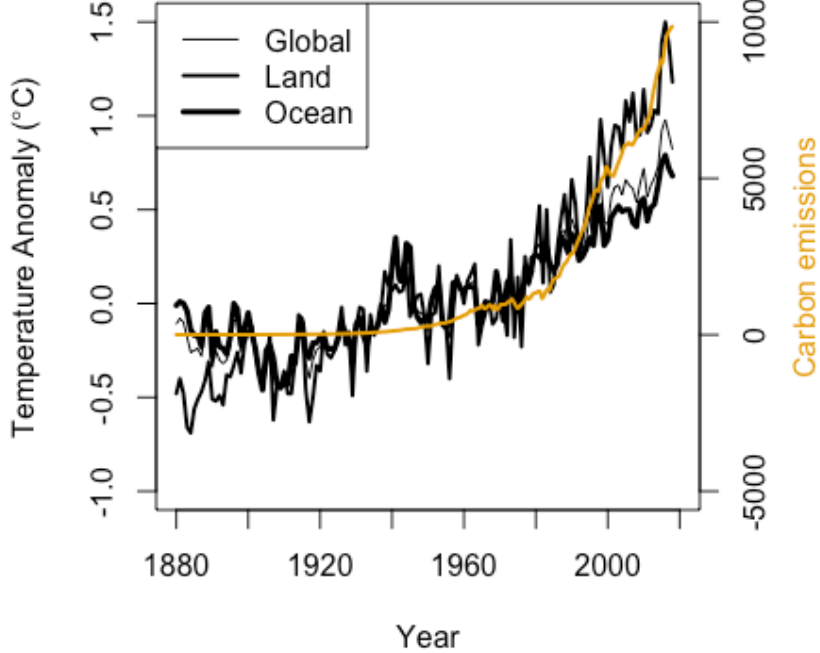
Let's see how this relationship changes when we look at temperature anomalies from the land or sea:

```
head(temp_carbon)
##   year temp_anomaly land_anomaly ocean_anomaly carbon_emissions
## 1 1880       -0.11       -0.48        -0.01            236
## 2 1881       -0.08       -0.40         0.01            243
## 3 1882       -0.10       -0.48         0.00            256
## 4 1883       -0.18       -0.66        -0.04            272
## 5 1884       -0.26       -0.69        -0.14            275
## 6 1885       -0.25       -0.56        -0.17            277
plot(temp_carbon$year,temp_carbon$temp_anomaly, xlab="Year",
   ylab=expression(paste("Temperature Anomaly (",degree,"C)")),
   type="l",xlim=c(1880,2020), ylim=c(-1,1.5))
points(temp_carbon$year,temp_carbon$land_anomaly,type="l", lwd=2, xlim=c(1880,2020))
points(temp_carbon$year,temp_carbon$ocean_anomaly,type="l", lwd=3, xlim=c(1880,2020))


legend("topleft",lwd=c(1,2,3),c("Global","Land","Ocean"))
```



Let's add carbon emissions and see if there is an obvious correlation:

```
#this sets the margins so that we can include a second axis later
par(mai=c(1,1,0.2,1))
#plot the first graph
plot(temp_carbon$year,temp_carbon$temp_anomaly, xlab="Year",
   ylab=expression(paste("Temperature Anomaly (",degree,"C)")),type="l",
   xlim=c(1880,2020),ylim=c(-1,1.5))
points(temp_carbon$year,temp_carbon$land_anomaly,type="l",
    lwd=2, xlim=c(1880,2020))
points(temp_carbon$year,temp_carbon$ocean_anomaly,type="l",
    lwd=3, xlim=c(1880,2020))
#plot the second graph
par(new=TRUE)
plot(temp_carbon$year,temp_carbon$carbon_emissions,type="l", lwd=2,
   col=cb_palette[2],axes=F,xlab="",ylab="",ylim=c(-5000,10000))
#add the axis for Carbon emissions
axis(4)
mtext("Carbon emissions",4,padj=4,col=cb_palette[2])
#add a legend
legend("topleft",lwd=c(1,2,3),c("Global","Land","Ocean"))
```

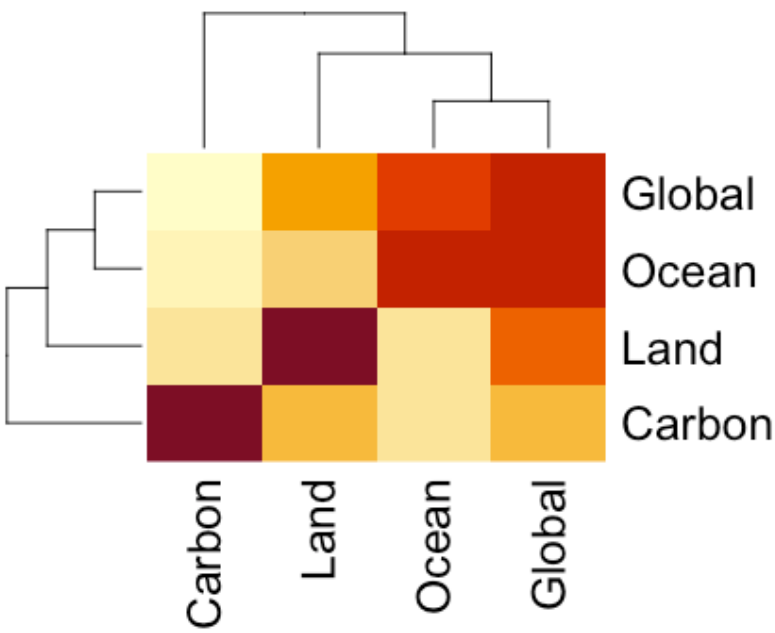


At first glance, that looks like a pretty tight correlation. Let's analyze this with statistics. First we will create a matrix of the relevant variables to correlate where each row is a year. We will stop at the 135th row so that all columns have the same number of data and there are no missing data:

```
tcMat<-temp_carbon[1:135,c(2,3,4,5)]
```

Then, we can create a cross-correlation matrix and plot the results in a heat map:

```
cor(tcMat)
##                temp_anomaly land_anomaly ocean_anomaly carbon_emissions
## temp_anomaly       1.0000000    0.9611008     0.9853432       0.8997974
## land_anomaly       0.9611008    1.0000000     0.9010434       0.9016064
## ocean_anomaly      0.9853432    0.9010434     1.0000000       0.8648896
## carbon_emissions   0.8997974    0.9016064     0.8648896       1.0000000
heatmap(cor(tcMat),labCol=c("Global","Land","Ocean","Carbon"),
      labRow=c("Global","Land","Ocean","Carbon"),margins=c(9,4))
```



We will then run a correlation test between each of the temperature anomalies and carbon emissions. We could run correlation tests separately:

```
cor.test(tcMat[,4],tcMat[,1])
##
##  Pearson's product-moment correlation
##
```

```
##
## data:  tcMat[, 4] and tcMat[, 1]
## t = 23.784, df = 133, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8618675 0.9277164
## sample estimates:
##       cor
## 0.8997974
```

But to save time, we can construct a for loop:

```
for(i in 1:4){
print(cor.test(tcMat[,4],tcMat[,i]))
 }
##
##  Pearson's product-moment correlation
##
## data:  tcMat[, 4] and tcMat[, i]
## t = 23.784, df = 133, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8618675 0.9277164
## sample estimates:
##       cor
## 0.8997974
##
##
##  Pearson's product-moment correlation
##
## data:  tcMat[, 4] and tcMat[, i]
## t = 24.038, df = 133, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8643124 0.9290402
## sample estimates:
##       cor
## 0.9016064
##
##
##  Pearson's product-moment correlation
##
## data:  tcMat[, 4] and tcMat[, i]
## t = 19.871, df = 133, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8150325 0.9020332
## sample estimates:
##       cor
## 0.8648896
##
##
##  Pearson's product-moment correlation
##
## data:  tcMat[, 4] and tcMat[, i]
## t = Inf, df = 133, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  1 1
## sample estimates:
## cor
##   1
```

Exercise 1: Evidence for the effects of carbon emissions on global temperatures

a.) Which of the temperature anomaly metrics is most tightly correlated with carbon emissions? What is your evidence and can you declare the relationship statistically significant?

b.) What else was signficantly correlated in this datset? Why is it important to specify where temperature anomalies were calculated?

c.) Do you take these results as signficant evidence for anthropogenic-caused climate change? Why or why not? (Note: There is no right answer, just answer honestly)

Exercise 2: Correlation or causation?

Let's repeat the above exercise on a dataset called divorce_margarine that includes divorces per 1000 people in Maine and the US per capita consumption of margarine in pounds.

```
?divorce_margarine
```

a.) Plot Maine divorce rates by margarine consumption and run a correlation test. Report your r (cor) and p values. **Make sure to label your axes**.

b.) Add a linear regression to your plot and report the slope, and P value.

c.) Can you provide a reasonable mechanism to explain these results? Do the results influence your answer to 1c? Why or why not?

As scientists we often rely on statistics to determine signals in the noise but those statistics are not infallible. There is always the chance that we may misinterpret results in any individual study, which is why science relies on a broad community to confirm or refute results and provide greater evidence through consensus. Let's take a look at some other sources of data on climate change to see if we confirm our previous observations.

Rather than look at temperature changes injust the last couple centuries, let's expand our analysis to much longer timescale using the historic_co2 dataset:

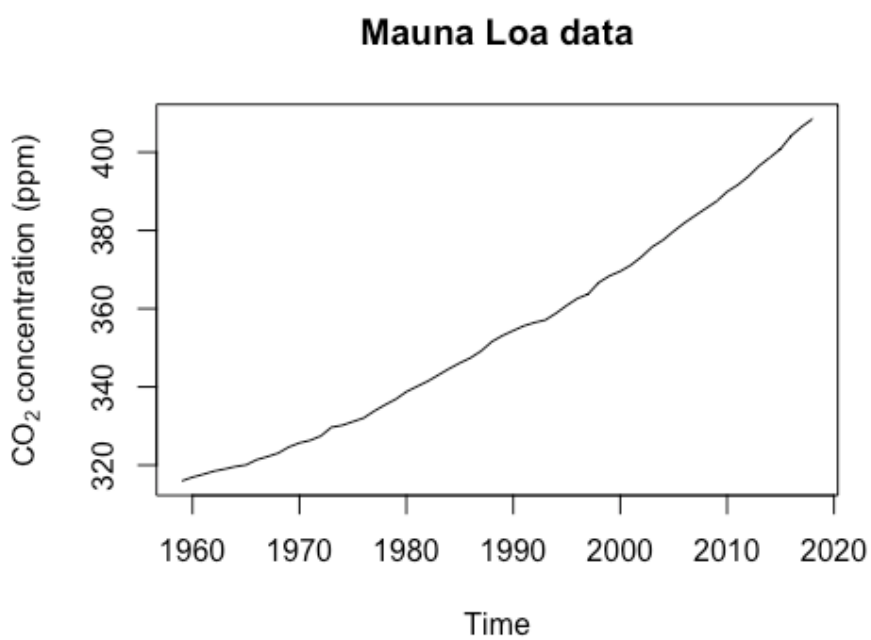
```
?historic_co2
```

This dataset includes timeseries data of from two sources: direct measurements done at Mauna Loa and indirect measurements taken fro Ice Cores. Let's first subset these two sources of data into new vectors we will call:

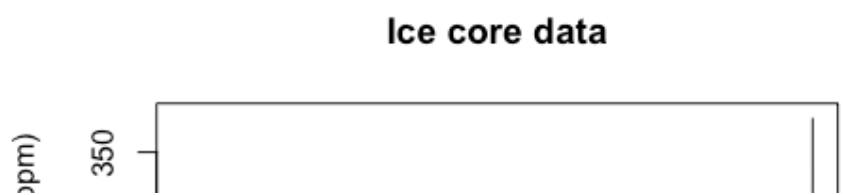
```
MaunaLoa<-subset(historic_co2,source=="Mauna Loa")
IceCores<-subset(historic_co2,source=="Ice Cores")
```

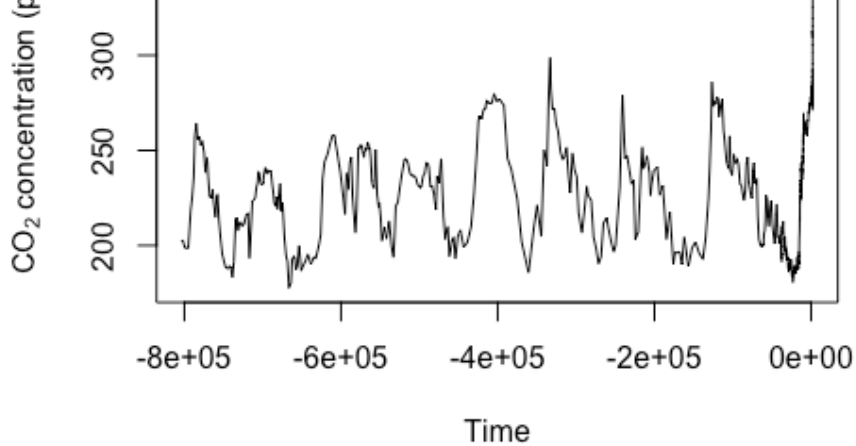
Then let's plot the two time series on separate graphs:

```
plot(MaunaLoa$year,MaunaLoa$co2,type ="l", xlab="Time",
  ylab=expression(paste('CO'[2], ' concentration (ppm)')),main="Mauna Loa data")
```




```
plot(IceCores$year,IceCores$co2,type ="l",  xlab="Time",
  ylab=expression(paste('CO'[2], ' concentration (ppm)')), main="Ice core data")
```



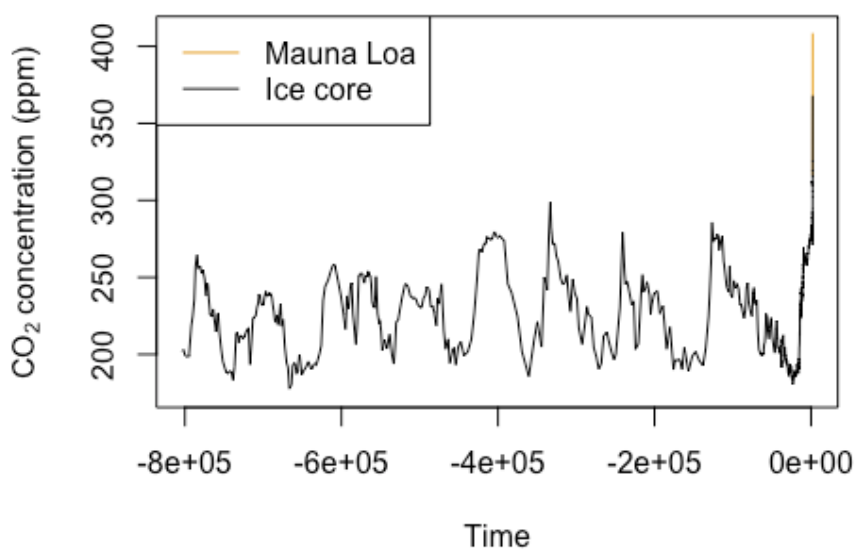


We can see that the data range for the two sets of data are quite different

```
summary(historic_co2$year)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -803182 -470498  -43278 -219753   -8924    2018
summary(historic_co2$co2)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   177.7   206.7   236.9   245.9   271.8   408.5
```

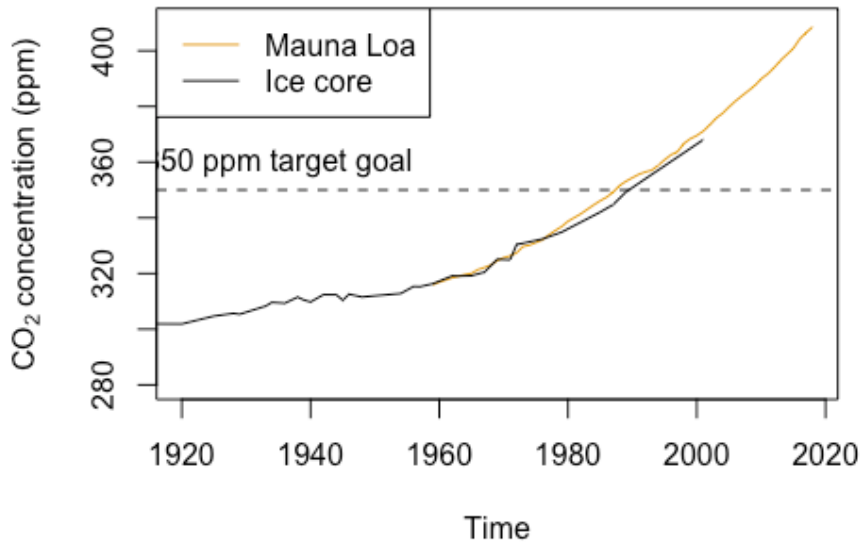
We will use the min and max as the xlim and ylim in order to plot the two datasets on the same graph:

```
plot(MaunaLoa$year,MaunaLoa$co2,type ="l", xlab="Time",
   ylab=expression(paste('CO'[2], ' concentration (ppm)')), xlim=c(-804000,2018),
   ylim=c(175,410),col=cb_palette[2])
points(IceCores$year,IceCores$co2,type ="l",col="black")
legend("topleft",lty=1,col=c(cb_palette[2],"black"),c("Mauna Loa","Ice core"))
```



That looks like a pretty sharp deviation from historical fluctuations in ! We can take closer look at the timeframe where the data overlaps:

```
plot(MaunaLoa$year,MaunaLoa$co2,type ="l", xlab="Time",
   ylab=expression(paste('CO'[2], ' concentration (ppm)')), xlim=c(1920,2018),
   ylim=c(280,410),col=cb_palette[2])
points(IceCores$year,IceCores$co2,type ="l",col="black")
legend("topleft",lty=1,col=c(cb_palette[2],"black"),c("Mauna Loa","Ice core"))
#Bill McKibben's organization, 350.org gets its name from the target concentration
#of CO2 thought to preserve most biodiversity and ecosystem function.
#We will mark that line here:
abline(h=350,lty=2)
text(1935,360,"350 ppm target goal")
```



Notice that the 350 mark has not been crossed previously anytime in the last ~800,000 years!

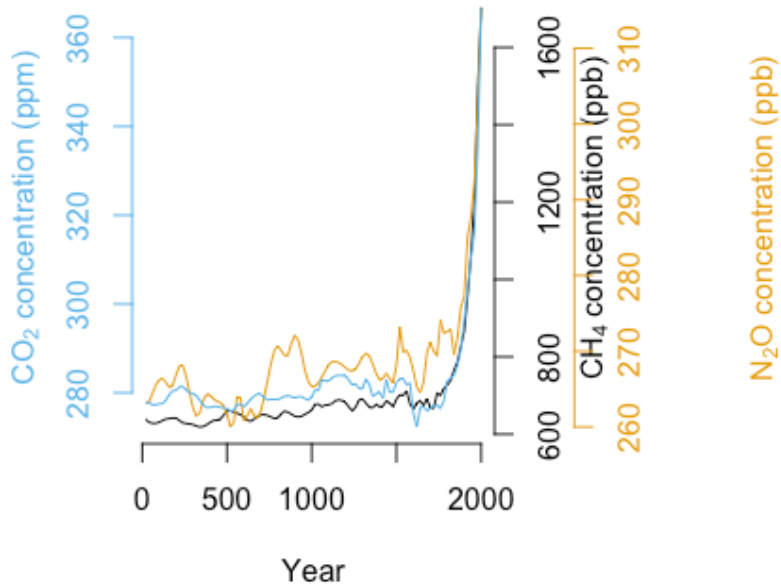
Remember that our linear regression suggested that 76% of the variance in temperature could be explained by C02:

```
summary(model1)
##
## Call:
## lm(formula = temp_anomaly ~ year, data = temp_carbon)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32598 -0.11846 -0.01062  0.11185  0.43367
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.409e+01  6.750e-01  -20.87   <2e-16 ***
## year         7.259e-03  3.463e-04   20.96   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1638 on 137 degrees of freedom
##   (129 observations deleted due to missingness)
## Multiple R-squared:  0.7623, Adjusted R-squared:  0.7606
## F-statistic: 439.4 on 1 and 137 DF,  p-value: < 2.2e-16
```

Other greenhouse gases are thought to contribute to the remaining variance. Let's see how methane () and nitrous oxide () have increased alongside in the greenhouse_gases database:

```
#First subset by gas

CH4gg<-subset(greenhouse_gases,gas=="CH4")
N2Ogg<-subset(greenhouse_gases,gas=="N2O")
CO2gg<-subset(greenhouse_gases,gas=="CO2")

#Then plot
par(mai=c(1,1,.5,2))
plot(CH4gg$year,CH4gg$concentration,type="l",col=cb_palette[1],
   xlab="Year", ylab="",axes=F)
axis(1)
axis(4,col=cb_palette[1])
mtext(expression(paste('CH'[4], ' concentration (ppb)')),side=4,padj=3)
par(new=TRUE)
plot(N2Ogg$year,N2Ogg$concentration,type="l",col=cb_palette[2],axes=FALSE,
   xlab="",ylab="")
axis(4,col=cb_palette[2],col.axis=cb_palette[2],pos=2550)
mtext(expression(paste('N'[2], 'O concentration (ppb)')),side=4,padj=9,col=cb_palette[2])
par(new=TRUE)
plot(CO2gg$year,CO2gg$concentration,type="l",col=cb_palette[3],axes=F,
```

```
plot(CO2gg$year,CO2gg$concentration,type="l",col=cb_palette[3],axes=F,
  xlab="",ylab="")
axis(2,col=cb_palette[3],col.axis=cb_palette[3])
mtext(expression(paste('CO'[2], ' concentration (ppm)')),side=2,col=cb_palette[3],padj=-3)
```



Exercise 3: Detecting signals in the noise.

a.) Run a cross-correlation matrix between CO2, NO2, and CH4. Which, if any are significantly correlated?

b.) How does the rise in greenhouse gases relate to the carbon emissions data in the temp_carbon dataset? Based on your results, is there consistent evidence to suggest that the rise in global temperatures is driven by human-caused green house gas emissions? What other analyses could you run to confirm or refute your hypothesis? What data would you need and how would it need to be collected and analyzed?

All models are wrong, some models are useful.

One way to confirm our suspicions that temperature changes are induced by carbon emissions is to construct a simple model of the mechanism behind this. Since the mechanism in this case is quite complicated and beyond the scope of this course, we will use a model called prophet, which helps predict future time series based on historical trends only. First we will install and load the package:

```
install.packages("prophet")
```

Say no when prompted for the binary version that requires compilation

```
library(prophet, warn.conflicts=F, quietly=T)
```

The prophet function requires data in the form of a dataframe (df) with two columns: date formatted-time and the y variable of interest.

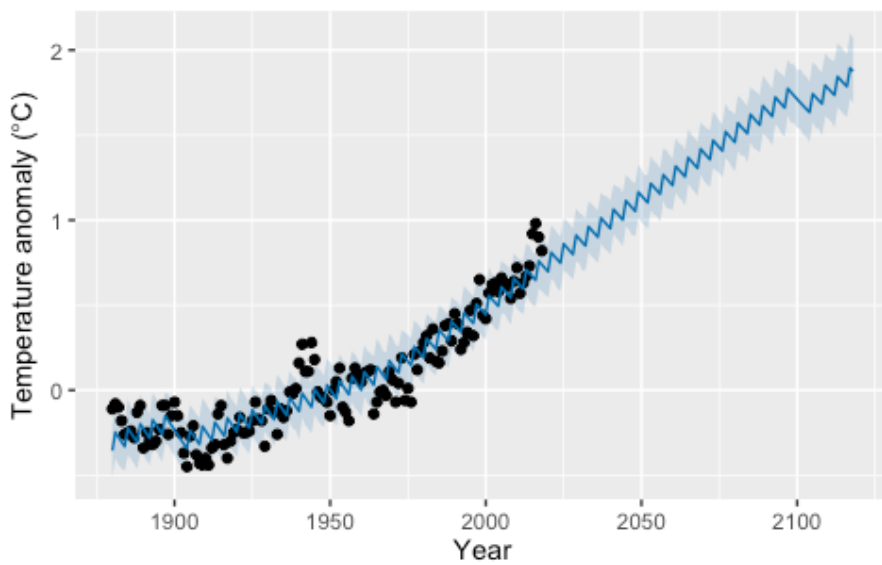
```
?prophet

#We first need the time in the format YYYY-MM-DD. Because we only have the year,
#we will just add 01-01 to each year:

dates<-rep(0,length(temp_carbon$year))
for(i in 1:length(temp_carbon$year)){
dates[i]<-paste(temp_carbon$year[i],"0101",sep="")
}

dates2<-as.Date(dates,"%Y%m%d")
CCdf<-data.frame(ds=dates2,y=temp_carbon$temp_anomaly)
```

We ended up with some NAs in this dataframe so let's get rid of them:

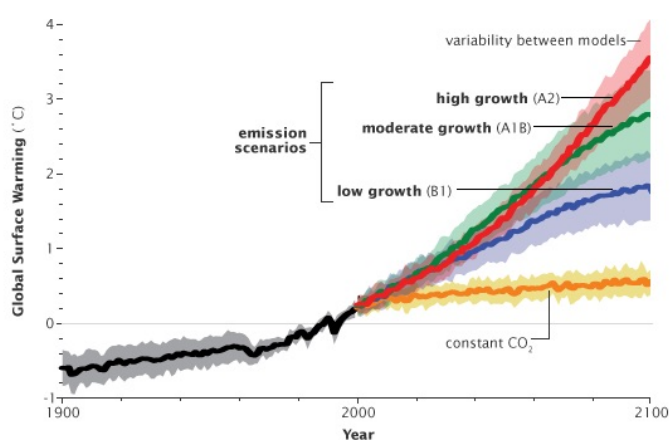
```
CCdf<-na.omit(CCdf)
```

Then we can run prophet

```
prophet_CCdf<-prophet(CCdf)
future<-make_future_dataframe(prophet_CCdf,freq="year",periods=100)
tail(future)
##           ds
## 234 2113-01-01
## 235 2114-01-01
## 236 2115-01-01
## 237 2116-01-01
## 238 2117-01-01
## 239 2118-01-01
forecast<-predict(prophet_CCdf,future)
tail(forecast[c('ds', 'yhat', 'yhat_lower', 'yhat_upper')])
##           ds     yhat yhat_lower yhat_upper
## 234 2113-01-01 1.843897   1.665554   2.040727
## 235 2114-01-01 1.824194   1.645097   2.020629
## 236 2115-01-01 1.804887   1.619649   1.999649
## 237 2116-01-01 1.786017   1.609007   1.974409
## 238 2117-01-01 1.894626   1.718804   2.099004
## 239 2118-01-01 1.874922   1.688451   2.067217
plot(prophet_CCdf, forecast,xlab="Year",
  ylab=expression(paste("Temperature anomaly (",degree,"C)")))
```



Let's compare our model results to NASA's current predictions:



NASA temperature anomaly predictions

Exercise 4: Comparing models

NASA's models are created from a combination of similar techniques used here by you today and more thorough mechanistic understandings/assumptions about climate systems and future human behavior in regards to GHG emissions.

a.) How does our model compare to NASA's predictions? Are you surprised?

a.) How does our model compare to NASA's predictions? Are you surprised?

b.) Predict future CO2 concentrations for the next 500 years based on the data from the `greenhouse_gases' dataset starting at the year 1000 and add your plot, with labels, here.

c.) What do you predict carbon dioxide concentrations will be for the year 2400? What happens to the confidence intervals at this time scale? Do you think this model is accurate? What is the model missing?