Stat 240 - Climate Data Extraction

Dr. Dave Campbell

March 25, 2019

Contents

1 CO_2 data **1**
1.1 CO_2 at surface level averaged over the Northern Hemisphere . . 1
1.2 CO_2 at surface level averaged over the world 3

2 Canadian Weather Data **4**
2.1 Monthly Snow 4
2.2 Total Precipitation 5
2.3 Max Temp 6
2.4 Mean Temp 7
2.5 Min Temp 8
2.6 Wind Speed 9

1 CO_2 data

1.1 CO_2 at surface level averaged over the Northern Hemisphere

- Date Range: 1979-01-01, 2018-01-01
- Degree Latitude Range: 30.0000, 90.0000

If this data is used in your assignment, you are required to reference the source. Please reference these data as: Dlugokencky, E.J., K.W. Thoning, P.M. Lang, and P.P. Tans (2017), NOAA Greenhouse Gas Reference from Atmospheric Carbon Dioxide Dry Air Mole Fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network. Data Path: `ftp://aftp.cmdl.noaa.gov/data/trace_gases/co2/flask/surface/`

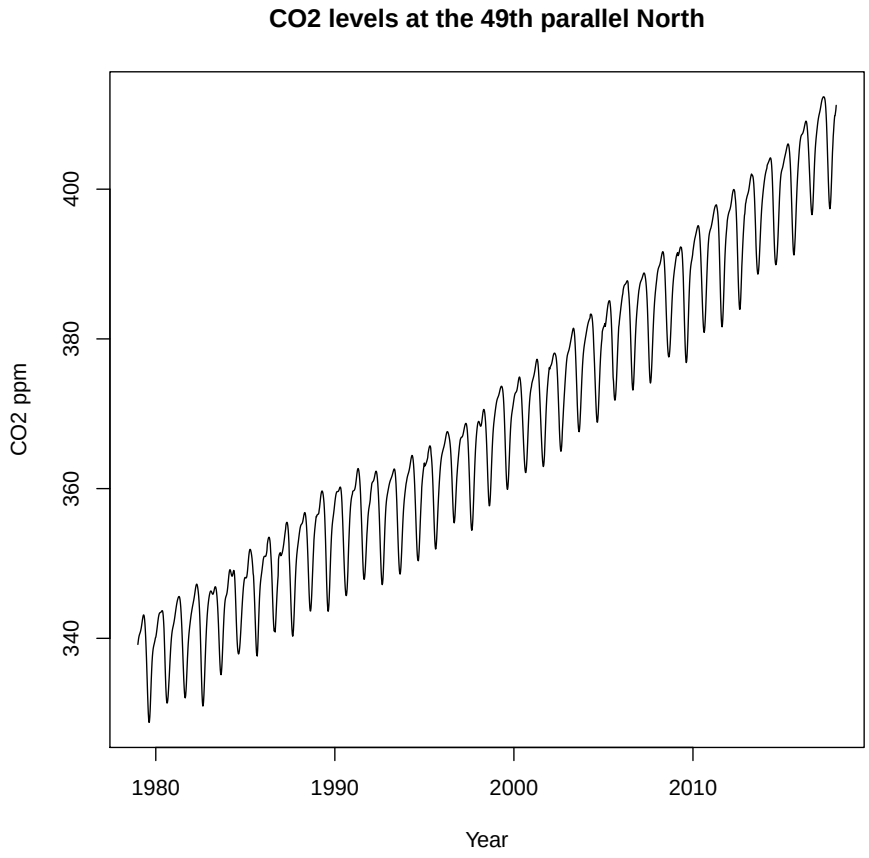
Data columns are the year in decimal form, the CO_2 value and its uncertainty at the sine of the latitude. Here I convert back to Latitude in building the column names.

```
Co2North = read.table("co2_GHGreference.1408404262_surface_Northern.txt",skip = 78,header=FA

Names2Use = c("YearDecimal",
              paste("Latitude",rep(round(asin(c(seq(.5,1,by=.05)))/(pi/2)*90),each=2),
              c("value","uncertainty"),sep=""))

colnames(Co2North) = Names2Use

plot(Co2North$YearDecimal,Co2North$Latitude49value,type="l",
     main = "CO2 levels at the 49th parallel North",
     xlab = "Year",ylab="CO2 ppm")
```




```
save(Co2North,file="CO2NorthernHemisphere.Rdata")
```

1.2 CO_2 at surface level averaged over the world

- Date Range: 1979-01-01, 2017-01-01
- Worldwide surface average

If this data is used in your assignment, you are required to reference the source. Please reference these data as (same as above): Dlugokencky, E.J., K.W. Thoning, P.M. Lang, and P.P. Tans (2017), NOAA Greenhouse Gas Reference from Atmospheric Carbon Dioxide Dry Air Mole Fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network. Data Path: `ftp://aftp.cmdl.noaa.gov/data/trace_gases/co2/flask/surface/`

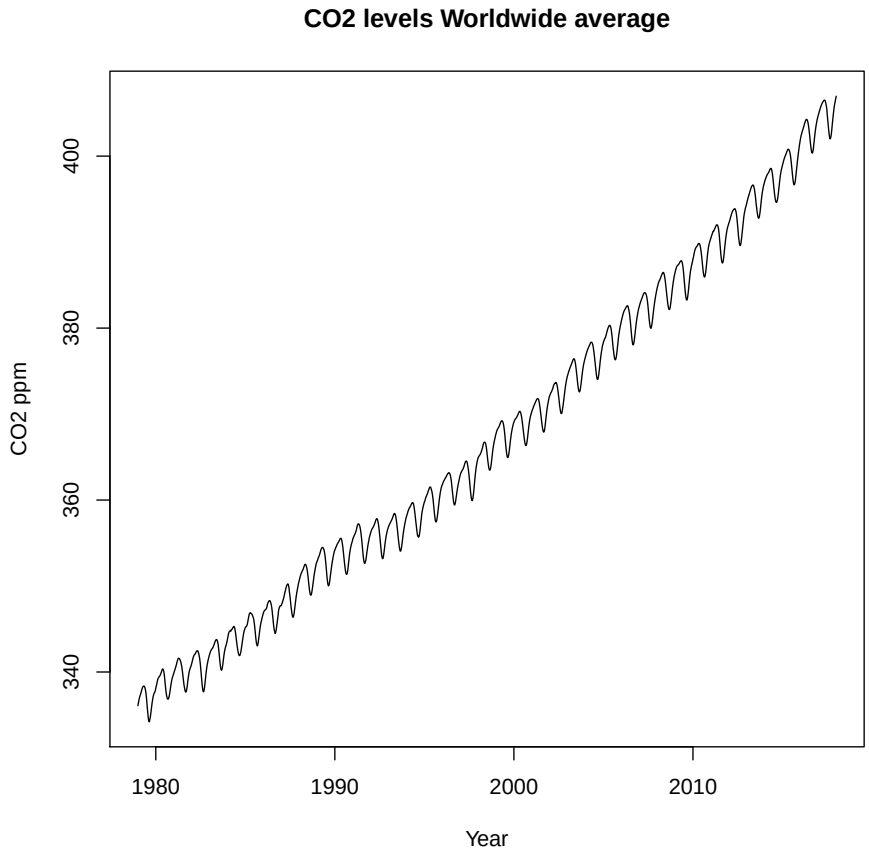
Data columns are the year, month, day, year as a decimal, value, uncertainty.

```
Co2World = read.table("co2_GHGreference.1692038112_zonal-90+90.txt",skip = 72,header=TRUE)

Names2Use = c("Year","Month","Day","YearDecimal","Value","Uncertainty")

colnames(Co2World) = Names2Use

plot(Co2World$YearDecimal,Co2World$Value,type="l",
     main = "CO2 levels Worldwide average",
     xlab = "Year",ylab="CO2 ppm")
```




```
save(Co2World,file="CO2Worldwide.Rdata")
```

2 Canadian Weather Data

Data can be obtained from the data source: `http://data.ec.gc.ca/data/climate/scientificknowledge/adjusted-and-homogenized-canadian-climate-data-ahccd/` You must reference the appropriate paper listed on that site when you use its dataset.

2.1 Monthly Snow

```
setwd("Canadian Climate Data from canada.ca/Adj_monthly_snow")
#getwd()
```

```r
#Find all files in the folder:
files2use = list.files(pattern = ".txt")

#Prep the data storage:
lp = 1
ColNames     = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
ColNamesMain = gsub(ColNames[which(ColNames!=" ")],pattern="\\s",replacement="")
#######This could be made much more efficient by pre-allocating the memory for the full mat

AllSnow = data.frame(matrix(NA,nrow=0,ncol=length(ColNamesMain)+3,dimnames=list(NULL,c(ColNa
fails = rep(NA,length(files2use))
for(lp in 1:length(files2use)){
    InfoTemp = gsub(scan(file=files2use[lp],what="character",nlines=1,sep=","),pattern="(\\s

    ColNames = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
    ColNames = gsub(ColNames,pattern="\\s",replacement="")
    DataTable = read.table(files2use[lp],skip=4,sep = ",")

    colnames(DataTable) = ColNames
    ColNames2Use = ColNames[ColNames!=""]
    if(all(ColNamesMain%in%ColNames2Use)){
      AllSnow =  rbind(AllSnow,cbind(DataTable[,ColNamesMain],InfoTemp[1],InfoTemp[2],InfoTe
      fails[lp]=FALSE
    }else{
      fails[lp]=TRUE
      #just find out when there are extra column names
    }
}


if(all(fails==FALSE)){
  print("Snow is complete")
  save(AllSnow,file="CanadianAvgSnow.Rdata")
}else{
  print("Snow Failed")
}

## [1] "Snow is complete"

setwd("../..")
```

2.2 Total Precipitation

```r
setwd("Canadian Climate Data from canada.ca/Adj_monthly_total_prec")
#getwd()
#Find all files in the folder:
files2use = list.files(pattern = ".txt")

#Prep the data storage:
lp = 1
ColNames     = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
ColNamesMain = gsub(ColNames[which(ColNames!=" ")],pattern="\\s",replacement="")
#######This could be made much more efficient by pre-allocating the memory for the full mat

AllPrecip = data.frame(matrix(NA,nrow=0,ncol=length(ColNamesMain)+3,dimnames=list(NULL,c(Col
fails = rep(NA,length(files2use))
for(lp in 1:length(files2use)){
    InfoTemp = gsub(scan(file=files2use[lp],what="character",nlines=1,sep=","),pattern="(\\s

    ColNames = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
    ColNames = gsub(ColNames,pattern="\\s",replacement="")
    DataTable = read.table(files2use[lp],skip=4,sep = ",")

    colnames(DataTable) = ColNames
    ColNames2Use = ColNames[ColNames!=""]
    if(all(ColNamesMain%in%ColNames2Use)){
      AllPrecip =  rbind(AllPrecip,cbind(DataTable[,ColNamesMain],InfoTemp[1],InfoTemp[2],In
      fails[lp]=FALSE
    }else{
      fails[lp]=TRUE
      #just find out when there are extra column names
    }
}


if(all(fails==FALSE)){
  print("Precip is complete")
  save(AllPrecip,file="CanadianPrecip.Rdata")
}else{
  print("Precip Failed")
}

## [1] "Precip is complete"

setwd("../..")
```

2.3 Max Temp

```r
setwd("Canadian Climate Data from canada.ca/Homog_monthly_max_temp")
#getwd()
#Find all files in the folder:
files2use = list.files(pattern = ".txt")

#Prep the data storage:
lp = 1
ColNames     = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
ColNamesMain = gsub(ColNames[which(ColNames!=" ")],pattern="\\s",replacement="")
#######This could be made much more efficient by pre-allocating the memory for the full mat

MaxTemp = data.frame(matrix(NA,nrow=0,ncol=length(ColNamesMain)+3,dimnames=list(NULL,c(ColNa
fails = rep(NA,length(files2use))
for(lp in 1:length(files2use)){
    InfoTemp = gsub(scan(file=files2use[lp],what="character",nlines=1,sep=","),pattern="(\\s

    ColNames = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
    ColNames = gsub(ColNames,pattern="\\s",replacement="")
    DataTable = read.table(files2use[lp],skip=4,sep = ",")

    colnames(DataTable) = ColNames
    ColNames2Use = ColNames[ColNames!=""]
    if(all(ColNamesMain%in%ColNames2Use)){
      MaxTemp =  rbind(MaxTemp,cbind(DataTable[,ColNamesMain],InfoTemp[1],InfoTemp[2],InfoTe
      fails[lp]=FALSE
    }else{
      fails[lp]=TRUE
      #just find out when there are extra column names
    }
}


if(all(fails==FALSE)){
  print("Max Temp is complete")
  save(MaxTemp,file="CanadianMaxTemp.Rdata")
}else{
  print("Max Temp Failed")
}

## [1] "Max Temp is complete"

setwd("../..")
```

2.4 Mean Temp

```r
setwd("Canadian Climate Data from canada.ca/Homog_monthly_mean_temp")
#getwd()
#Find all files in the folder:
files2use = list.files(pattern = ".txt")

#Prep the data storage:
lp = 1
ColNames     = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
ColNamesMain = gsub(ColNames[which(ColNames!=" ")],pattern="\\s",replacement="")
#######This could be made much more efficient by pre-allocating the memory for the full mat

MeanTemp = data.frame(matrix(NA,nrow=0,ncol=length(ColNamesMain)+3,dimnames=list(NULL,c(ColN
fails = rep(NA,length(files2use))
for(lp in 1:length(files2use)){
    InfoTemp = gsub(scan(file=files2use[lp],what="character",nlines=1,sep=","),pattern="(\\s

    ColNames = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
    ColNames = gsub(ColNames,pattern="\\s",replacement="")
    DataTable = read.table(files2use[lp],skip=4,sep = ",")

    colnames(DataTable) = ColNames
    ColNames2Use = ColNames[ColNames!=""]
    if(all(ColNamesMain%in%ColNames2Use)){
      MeanTemp =  rbind(MeanTemp,cbind(DataTable[,ColNamesMain],InfoTemp[1],InfoTemp[2],Info
      fails[lp]=FALSE
    }else{
      fails[lp]=TRUE
      #just find out when there are extra column names
    }
}


if(all(fails==FALSE)){
  print("MeanTemp is complete")
  save(MeanTemp,file="CanadianMeanTemp.Rdata")
}else{
  print("MeanTemp Failed")
}

## [1] "MeanTemp is complete"

setwd("../..")
```

2.5 Min Temp

```r
setwd("Canadian Climate Data from canada.ca/Homog_monthly_min_temp")
#getwd()
#Find all files in the folder:
files2use = list.files(pattern = ".txt")

#Prep the data storage:
lp = 1
ColNames     = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
ColNamesMain = gsub(ColNames[which(ColNames!=" ")],pattern="\\s",replacement="")
#######This could be made much more efficient by pre-allocating the memory for the full mat

MinTemp = data.frame(matrix(NA,nrow=0,ncol=length(ColNamesMain)+3,dimnames=list(NULL,c(ColNa
fails = rep(NA,length(files2use))
for(lp in 1:length(files2use)){
    InfoTemp = gsub(scan(file=files2use[lp],what="character",nlines=1,sep=","),pattern="(\\s

    ColNames = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
    ColNames = gsub(ColNames,pattern="\\s",replacement="")
    DataTable = read.table(files2use[lp],skip=4,sep = ",")

    colnames(DataTable) = ColNames
    ColNames2Use = ColNames[ColNames!=""]
    if(all(ColNamesMain%in%ColNames2Use)){
      MinTemp =  rbind(MinTemp,cbind(DataTable[,ColNamesMain],InfoTemp[1],InfoTemp[2],InfoTe
      fails[lp]=FALSE
    }else{
      fails[lp]=TRUE
      #just find out when there are extra column names
    }
}


if(all(fails==FALSE)){
  print("MinTemp is complete")
  save(MinTemp,file="CanadianMinTemp.Rdata")
}else{
  print("MinTemp Failed")
}

## [1] "MinTemp is complete"

setwd("../..")
```

2.6 Wind Speed

```
#setwd("Canadian Climate Data from canada.ca/Homog_monthly_wind_speed")
#getwd()
#Find all files in the folder:
files2use = list.files(pattern = ".txt")

#Prep the data storage:
lp = 1
ColNames     = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
ColNamesMain = gsub(ColNames[which(ColNames!=" ")],pattern="\\s",replacement="")
#######This could be made much more efficient by pre-allocating the memory for the full mat

WindSpeed = data.frame(matrix(NA,nrow=0,ncol=length(ColNamesMain)+3,dimnames=list(NULL,c(Col
fails = rep(NA,length(files2use))
for(lp in 1:length(files2use)){
    InfoTemp = gsub(scan(file=files2use[lp],what="character",nlines=1,sep=","),pattern="(\\s

    ColNames = scan(file=files2use[lp],what="character",nlines=1,skip=2,sep=",")
    ColNames = gsub(ColNames,pattern="\\s",replacement="")
    DataTable = read.table(files2use[lp],skip=4,sep = ",")

    colnames(DataTable) = ColNames
    ColNames2Use = ColNames[ColNames!=""]
    if(all(ColNamesMain%in%ColNames2Use)){
      WindSpeed =  rbind(WindSpeed,cbind(DataTable[,ColNamesMain],InfoTemp[1],InfoTemp[2],In
      fails[lp]=FALSE
    }else{
      fails[lp]=TRUE
      #just find out when there are extra column names
    }
}


if(all(fails==FALSE)){
  print("WindSpeed is complete")
  save(WindSpeed,file="CanadianWindSpeed.Rdata")
}else{
  print("WindSpeed Failed")
}

## [1] "WindSpeed is complete"

setwd("../..")
```