# Climate Exercise

*Chandler Sutherland & Vanessa Garcia*

## Unit I: Climate Change Module

### Examining CO2 trends in R

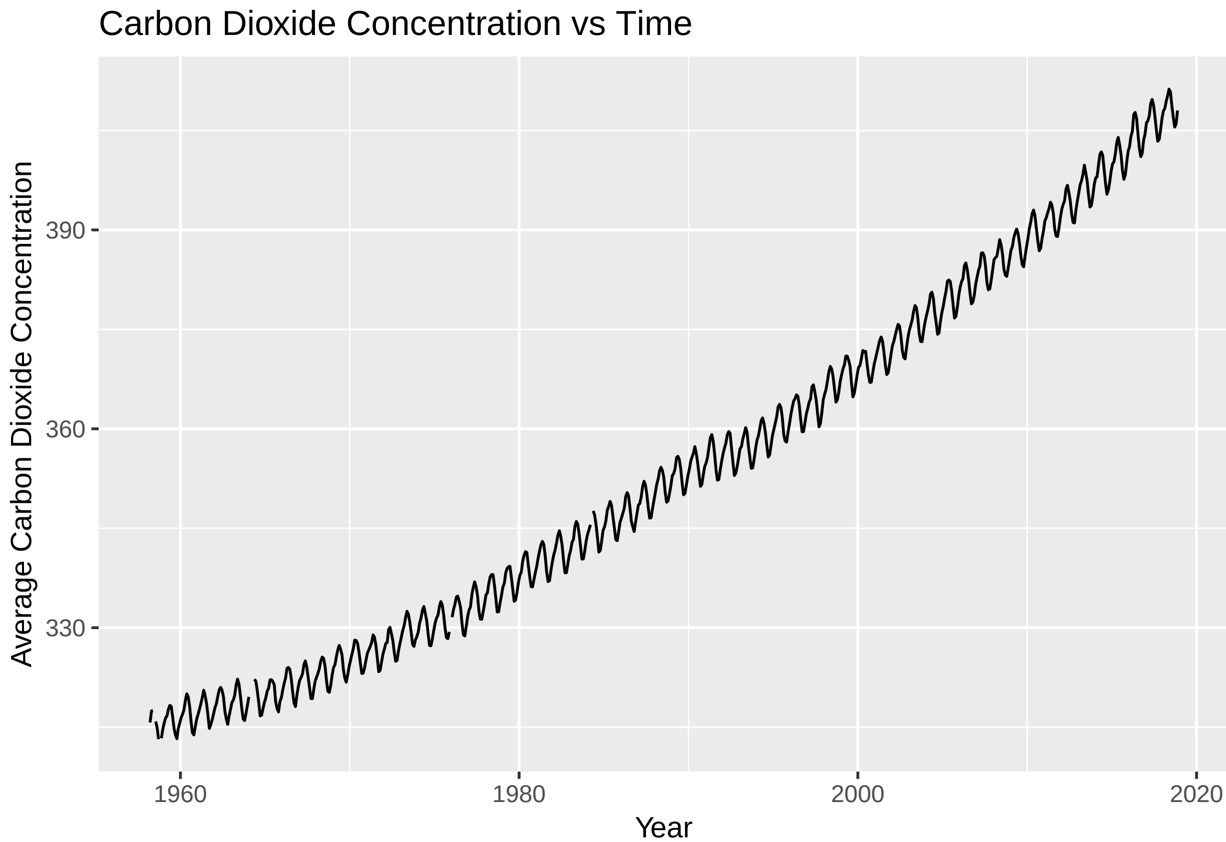
- Example from http://climate.nasa.gov/vital-signs/carbon-dioxide/
- Raw data from ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_mm_mlo.txt

```
library(tidyverse)
library(RcppRoll)
library(ggplot2)
library(lubridate)
```

```
co2 <-
readr::read_table("ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_mm_mlo.txt",
                  comment="#",
                  col_names = c("year", "month", "decimal_date",
                                "average", "interpolated", "trend",
                                "days"),
                  na = c("-1", "-99.99"))
co2
```

```
## # A tibble: 729 x 7
##     year month decimal_date average interpolated trend  days
##    <int> <int>        <dbl>   <dbl>        <dbl> <dbl> <int>
##  1  1958     3        1958.    316.         316.  315.    NA
##  2  1958     4        1958.    317.         317.  315.    NA
##  3  1958     5        1958.    318.         318.  315.    NA
##  4  1958     6        1958.     NA          317.  315.    NA
##  5  1958     7        1959.    316.         316.  315.    NA
##  6  1958     8        1959.    315.         315.  316.    NA
##  7  1958     9        1959.    313.         313.  316.    NA
##  8  1958    10        1959.     NA          313.  316.    NA
##  9  1958    11        1959.    313.         313.  315.    NA
## 10  1958    12        1959.    315.         315.  316.    NA
## # ... with 719 more rows
```

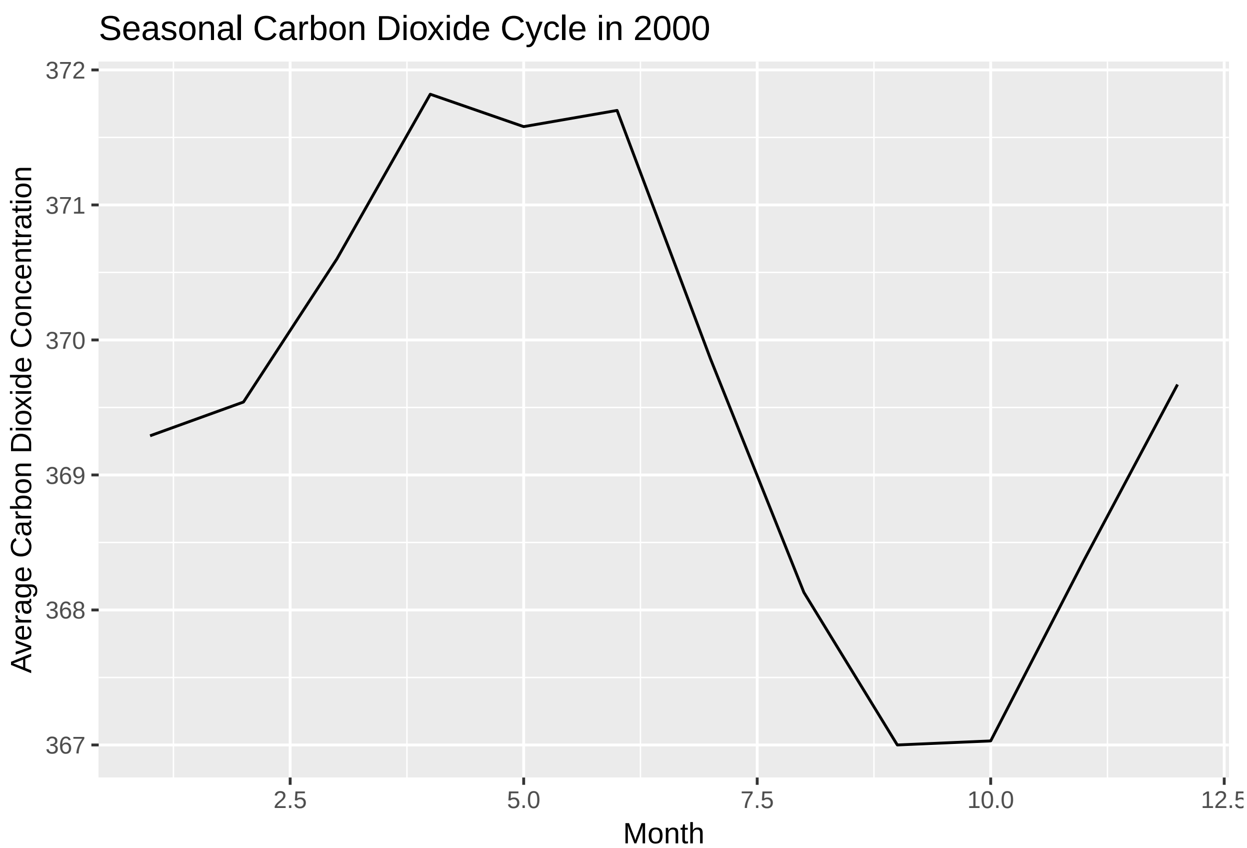
```
ggplot(co2, aes(x = decimal_date, y = average)) + geom_line() +
  labs(title = 'Carbon Dioxide Concentration vs Time', x = 'Year', y = 'Average Carbon Dioxide Concentra
```




CO2 values are at their minimum in September and October and their maximum in April and May. In April and May decomposition of plants releases a lot of CO2 into the atmosphere and the springtime burst of photosyntehsis is just beginning. In fall, decomposition hasn't started and photosynthesis is at its peak.

This can be observed simply by looking through the data from 2018 and 2017. Below is the year 2000 as an example, with the average CO2 emmisions by month.

```
co2_2002 <- co2 %>%
  filter(year == 2000)
ggplot(co2_2002, aes(x = month, y = average)) + geom_line()+
  labs(title = 'Seasonal Carbon Dioxide Cycle in 2000', x = 'Month', y = 'Average Carbon Dioxide Concen
```




The rolling average used to compute the trend line is the average seasonal cycle in the seven years surrounding each monthly value, allowing for change in seasonal cycles. The trend is then computed by removing the seasonal cycle, allowing the trend to show changes in co2 indepdent of the natural changes with each season. Missing trend values are linearly interpolated.

---

# Global Temperature Data

Each of the last years has consecutively set new records on global climate. In this section we will analyze global mean temperature data.

Data from: http://climate.nasa.gov/vital-signs/global-temperature

## Description of Global Temperature Data Set

The global temperature data set from the NASA global climate change archive includes 3 columns: the year of the measurement, the annual mean global surface temperature anomaly, and the lowess smoothing value, with units of years, degrees celsius, and degrees celsius, respectively. They are all doubles.

The measurements are the change in average of the global surface temperatures for a single year, or the mean global anomaly. Therefore we lose information about temperature distribution across the surface and the variations within a year due to seasonal change. Because our dataset is a change in averages, there is some associated uncertainty but I don't know enough about stats to articulate it.

The resolution of the data is to the .01 in degree celsius for Lowess Smoothing and global temperature anomally. The years are integers. From scrolling through the data in table form, I don't think there are missing points but if there were (and from the framing of this question I think there are) we should use the trend line prediction.

### Importing the Global Temperature Data Set

Construct the necessary R code to import and prepare for manipulation the following data set: http://climate.nasa.gov/system/internal_resources/details/original/647_Global_Temperature_Data_File.txt

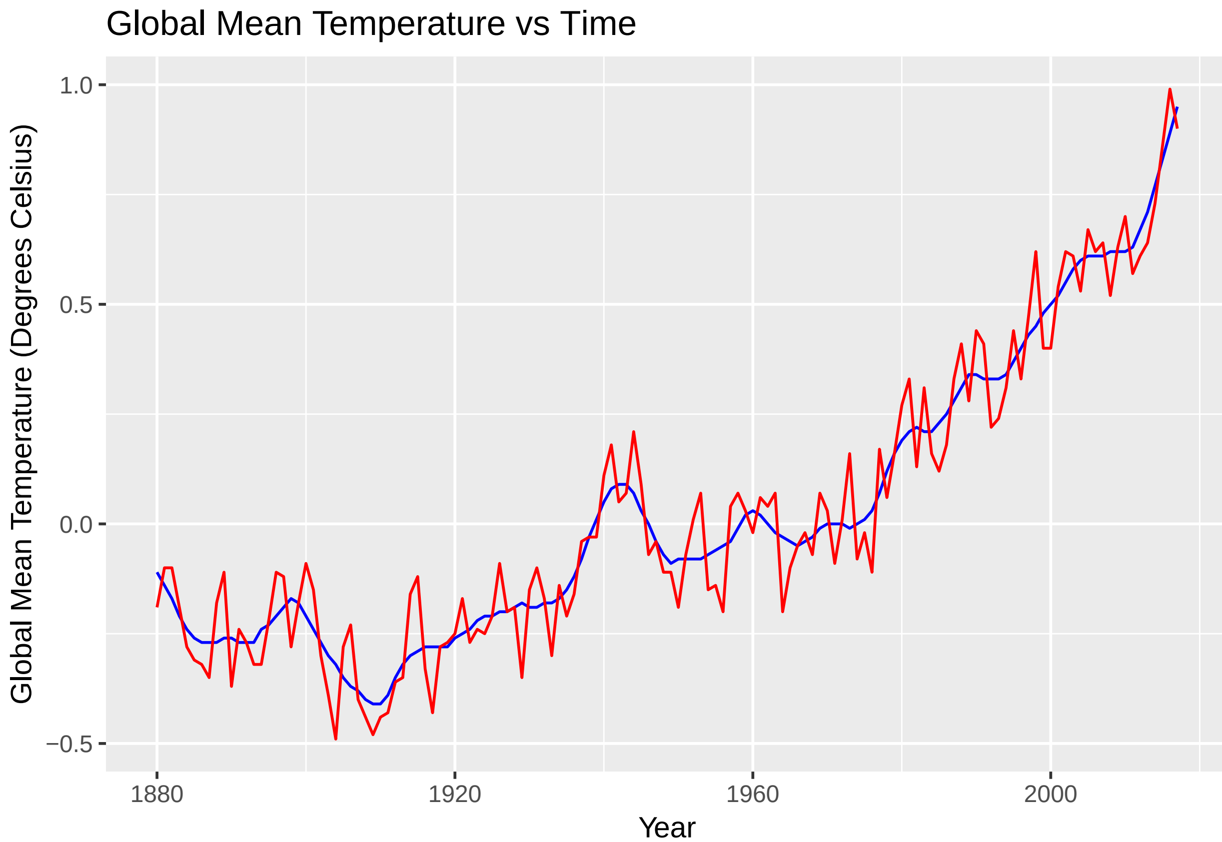
```
temperature <- 'http://climate.nasa.gov/system/internal_resources/details/original/647_Global_Temperatu

temp_table <- read_table2(temperature,
                          col_names = c("Year", "Annual_Mean","Lowess_Smoothing" ),
                          col_types = 'idd')
temp_table
```

```
## # A tibble: 138 x 3
##     Year Annual_Mean Lowess_Smoothing
##    <int>       <dbl>            <dbl>
##  1  1880       -0.19            -0.11
##  2  1881       -0.1             -0.14
##  3  1882       -0.1             -0.17
##  4  1883       -0.19            -0.21
##  5  1884       -0.28            -0.24
##  6  1885       -0.31            -0.26
##  7  1886       -0.32            -0.27
##  8  1887       -0.35            -0.27
##  9  1888       -0.18            -0.27
## 10  1889       -0.11            -0.26
## # ... with 128 more rows
```

### Visualizing Global Mean Temperature

```
temp_table %>%
  ggplot(aes(x = Year)) +
  geom_line(aes(y= Lowess_Smoothing), color = 'blue') +
  geom_line(aes(y = Annual_Mean), color = 'red') +
  labs(title = 'Global Mean Temperature vs Time',
       y = 'Global Mean Temperature (Degrees Celsius)')
```




In the plot we see an upward trend of the change in average global surface temperature starting in approximately 1970, the great acceleration. This pattern shows increasing surface temperatures relative to the previous years.

## Evaluating the evidence for a "Pause" in warming

The 2013 IPCC Report included a tentative observation of a "much smaller increasing trend" in global mean temperatures since 1998 than was observed previously. This led to much discussion in the media about the existence of a "Pause" or "Hiatus" in global warming rates, as well as much research looking into where the extra heat could have gone. (Examples discussing this question include articles in The Guardian, BBC News, and Wikipedia).

By examining the data here, what evidence do you find or not find for such a pause? Present an analysis of this data (using the tools & methods we have covered in Foundation course so far) to argue your case.

The articles concur that global surface temperature as a metric for measuring global warming is not comprehensive. Some theories for this so-called hiatus include cooler Atlantic water currents surfacing and absorbing heat, which would not show up on global surface temperature metrics. We find that the evidence of a pause is just incomplete analysis of temperature.

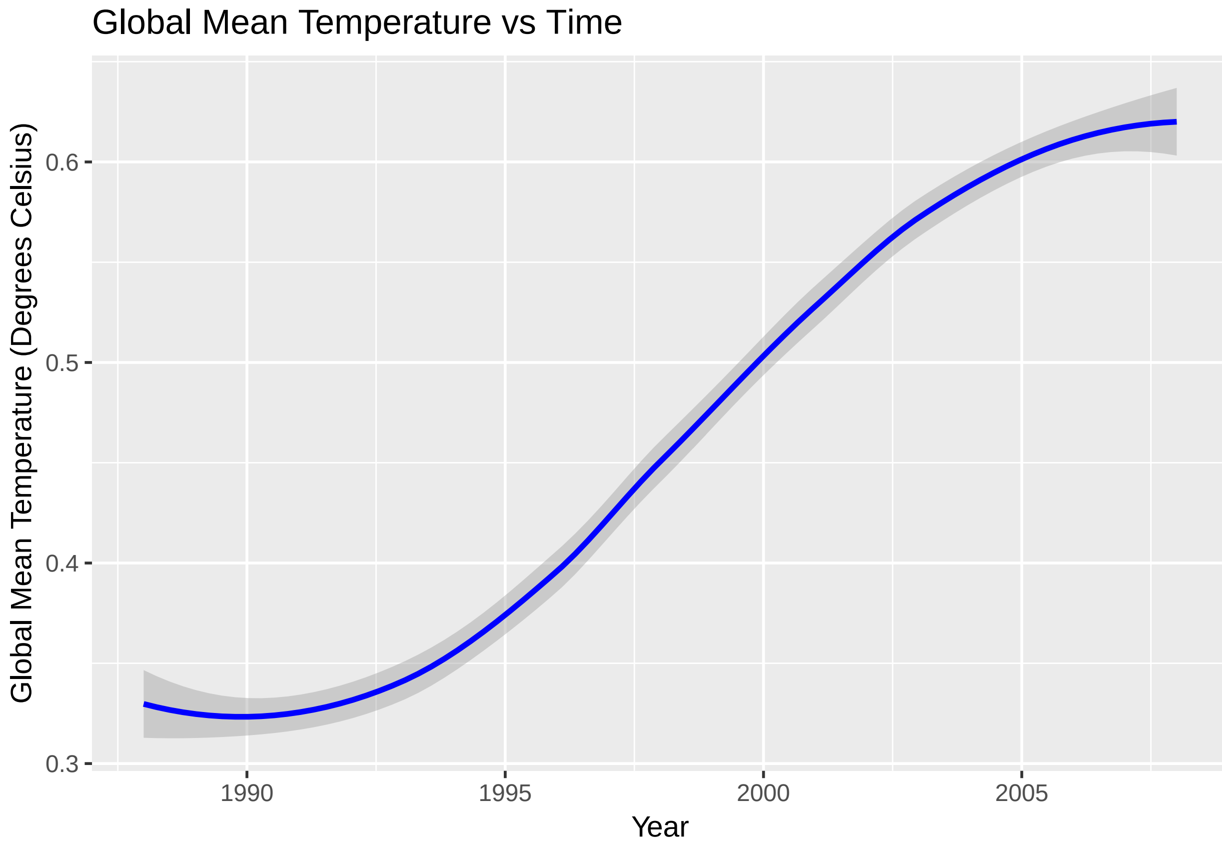
The plot below examines the twenty year interval surrounding 1998, and a decrease in rate is clear.

```
hiatus <- temp_table %>%
  filter(Year >= 1988) %>%
  filter(Year <= 2008)

ggplot(hiatus, aes(x = Year)) +
  geom_smooth(aes(y= Lowess_Smoothing), color = 'blue') +
  labs(title = 'Global Mean Temperature vs Time',
```

```
       y = 'Global Mean Temperature (Degrees Celsius)')
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




A data source documenting average ocean temperature would prove that the hiatus was more of a redistribution of heat than an actual slow down of global warming. Incorporating this and saline levels would be a more comprehensive view of change in temperature.

### Rolling averages as a Measurement Metric

A 5 year average is the average of the data of the surrounding five years, whereas an annual average is the average of data over one year.

Constructing 5 year, 10 year, and 20 year averages from the annual data:

```
temp_table
```

```
## # A tibble: 138 x 3
##     Year Annual_Mean Lowess_Smoothing
##    <int>       <dbl>            <dbl>
##  1  1880       -0.19            -0.11
##  2  1881       -0.1             -0.14
##  3  1882       -0.1             -0.17
##  4  1883       -0.19            -0.21
##  5  1884       -0.28            -0.24
##  6  1885       -0.31            -0.26
##  7  1886       -0.32            -0.27
##  8  1887       -0.35            -0.27
```

```
## 9   1888       -0.18            -0.27
## 10  1889       -0.11            -0.26
## # ... with 128 more rows
```

```
temp_table_average <- temp_table %>%
  mutate(five_yr_avg = roll_mean(Annual_Mean, n = 5, align = "right", fill = NA)) %>%
  mutate(ten_yr_avg = roll_mean(Annual_Mean, n = 10, align = "right", fill = NA)) %>%
  mutate(twenty_yr_avg = roll_mean(Annual_Mean, n = 20, align = "right", fill = NA))


temp_table_average
```

```
## # A tibble: 138 x 6
##     Year Annual_Mean Lowess_Smoothing five_yr_avg ten_yr_avg twenty_yr_avg
##    <int>       <dbl>            <dbl>       <dbl>      <dbl>         <dbl>
##  1  1880       -0.19            -0.11      NA         NA                NA
##  2  1881       -0.1             -0.14      NA         NA                NA
##  3  1882       -0.1             -0.17      NA         NA                NA
##  4  1883       -0.19            -0.21      NA         NA                NA
##  5  1884       -0.28            -0.24      -0.172     NA                NA
##  6  1885       -0.31            -0.26      -0.196     NA                NA
##  7  1886       -0.32            -0.27      -0.24      NA                NA
##  8  1887       -0.35            -0.27      -0.29      NA                NA
##  9  1888       -0.18            -0.27      -0.288     NA                NA
## 10  1889       -0.11            -0.26      -0.254     -0.213            NA
## # ... with 128 more rows
```

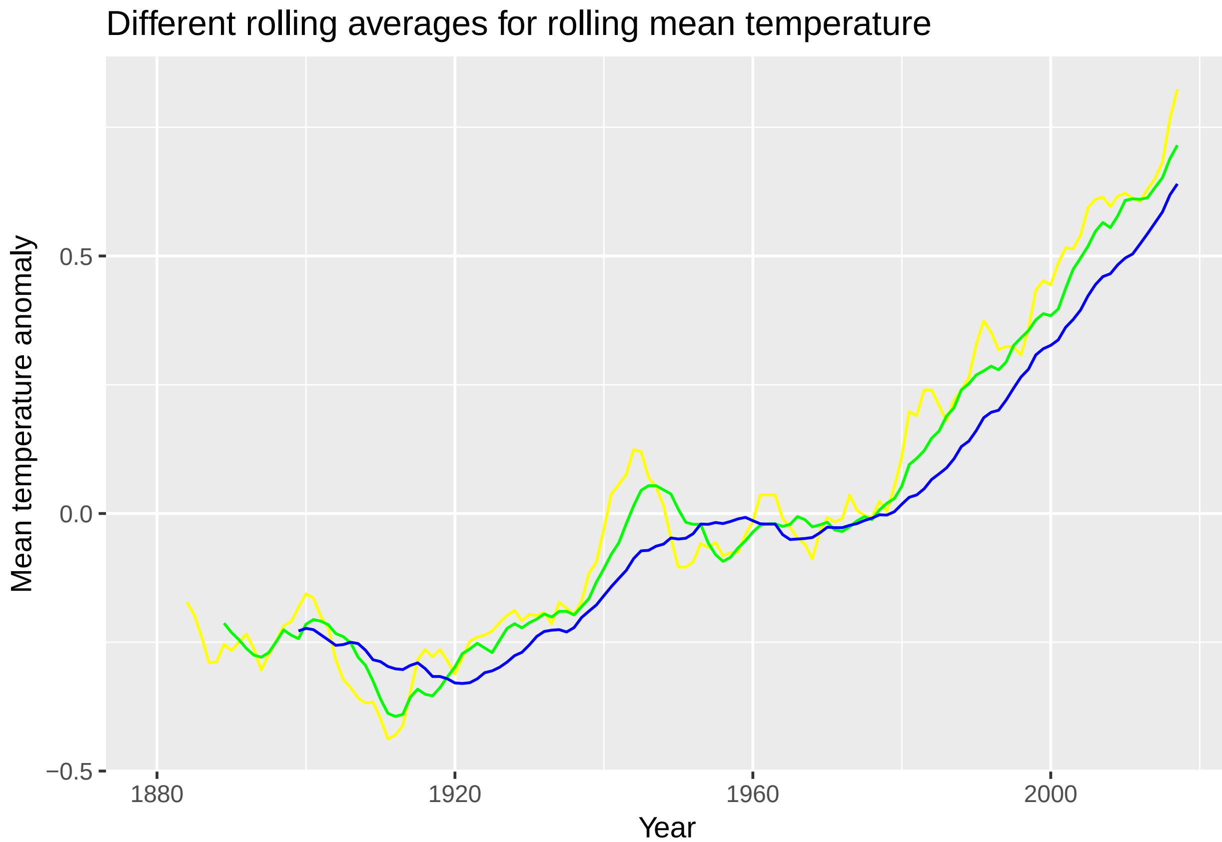
- Plot the different averages and describe what differences you see and why.

```
ggplot(temp_table_average, aes(x = Year)) +
  geom_line(aes(y = five_yr_avg), color = "yellow") +
  geom_line(aes(y = ten_yr_avg), color = "green") +
  geom_line(aes(y = twenty_yr_avg), color = "blue") +
  labs(title = "Different rolling averages for rolling mean temperature",
       x = "Year",
       y = "Mean temperature anomaly")
```

```
## Warning: Removed 4 rows containing missing values (geom_path).
```

```
## Warning: Removed 9 rows containing missing values (geom_path).
```

```
## Warning: Removed 19 rows containing missing values (geom_path).
```




As you increase the number of years to take an average from, the line gets smoother. There is less variability in points. The trend of values is clearer as you expand the range of averages. However, we do loose information about the initial values because you cannot calculate a rolling average for the earlier years. The ten year average is closest to the Lowess Smoothing line.

## Melting Ice Sheets

- Data description: http://climate.nasa.gov/vital-signs/land-ice/
- Raw data file: http://climate.nasa.gov/system/internal_resources/details/original/499_GRN_ANT_mass_changes.csv

### Description of Melting Ice Sheet Data Set

There are three columns in the NASA climate data set: time, measured in decimal years, greenland's mass of ice sheets in gigatonnes, and Antartica's mass of ice sheets in gigatonnes. The measurements are gathered from NASA's GRACE satellites, with GIA correction.

The uncertainty in the measurement has to do with the accuracy of the satelitte in determing the mass of the ice sheets, as that is a prediction and not a direct measurement. The resolution of the year is to .01 decimal years, and the resolution of the mass is .01 gigatonnes for both Antartica and Greenland. There were some data points that had a lower resolution than others, and I would guess that they were calculated from a trend line to fill a missing value.

### Importing Melting Ice Sheet Data Set

Construct the necessary R code to import this data set as a tidy `Table` object.

```r
melting_ice <- readr::read_csv("http://climate.nasa.gov/system/internal_resources/details/original/499_(

                  col_names = c("time_decimal", "greenland_mass",
                                "antartica_mass") ,
                  skip = 10)
```

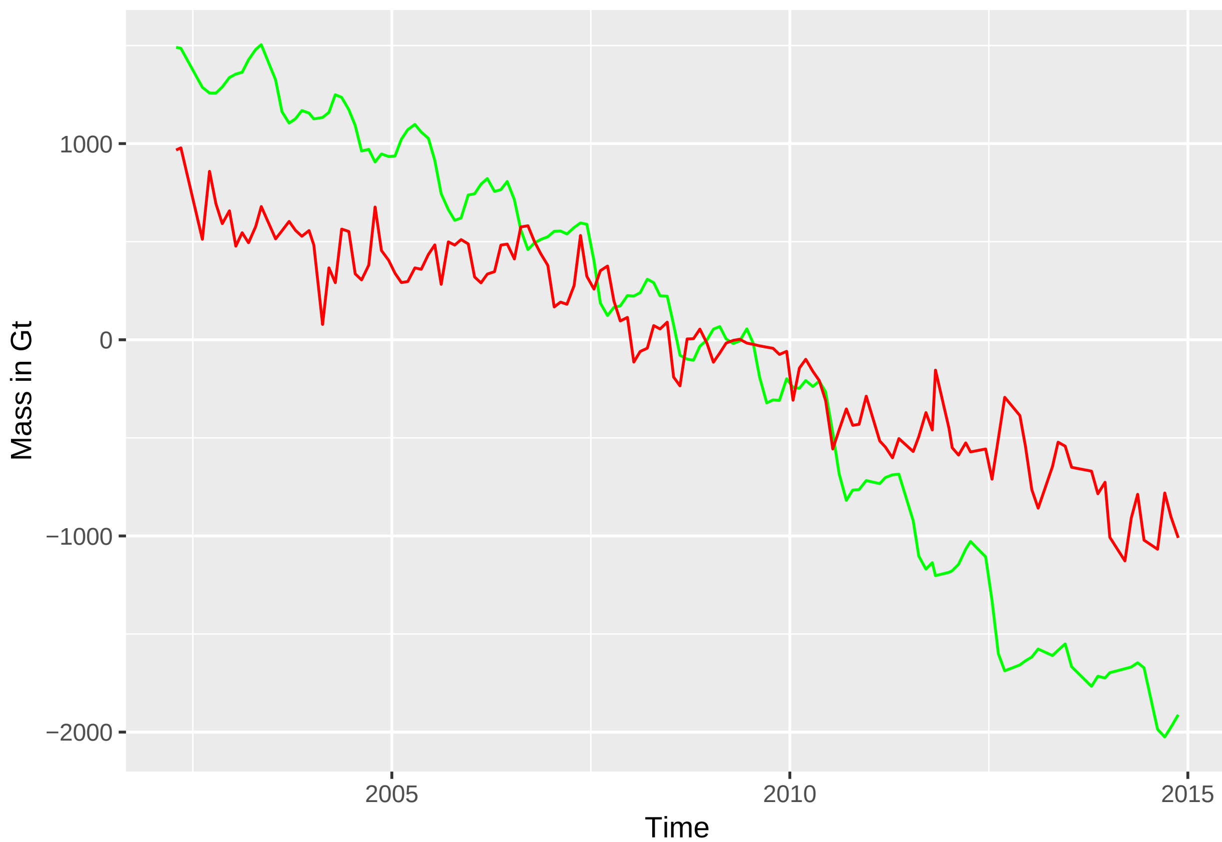
```
## Parsed with column specification:
## cols(
##   time_decimal = col_double(),
##   greenland_mass = col_double(),
##   antartica_mass = col_double()
## )
```

```r
melting_ice
```

```
## # A tibble: 140 x 3
##    time_decimal greenland_mass antartica_mass
##           <dbl>          <dbl>          <dbl>
##  1        2002.          1491.           967.
##  2        2002.          1486.           979.
##  3        2003.          1287.           512.
##  4        2003.          1258.           859.
##  5        2003.          1257.           694.
##  6        2003.          1288.           592.
##  7        2003.          1337.           658.
##  8        2003.          1354.           477.
##  9        2003.          1363.           546.
## 10        2003.          1427.           494.
## # ... with 130 more rows
```

### Melting Ice Sheet Visualization

```r
ggplot(melting_ice, aes(x=time_decimal)) +
  geom_line(aes(y = greenland_mass), color = "green") +
  geom_line(aes(y = antartica_mass), color = "red") +
  labs(x = "Time", y = "Mass in Gt")
```




## Rising Sea Levels

- Data description: http://climate.nasa.gov/vital-signs/sea-level/
- Raw data file: http://climate.nasa.gov/system/internal_resources/details/original/121_Global_Sea_Level_Data_File.txt

### Rising Sea Level Data Description

There are 12 columns in this NASA data set: describing altimeter type(0=dual-frequency 999=single frequency), year and fraction of year, the number of obseravtions, the number of weighted observations, GMSL variation(GLobal mean sea level) with the Global Isostatic Adjustment (GIA) not applied) with respect to TOPEX collinear mean reference in (mm), the SD of GMSL variation (with GIA not applied) in mm, the smoothed (60-day Gaussian type filter) GMSL variation (with GIA not applied) in mm, the GMSL(with GIA applied) with respect to TOPEX in mm, theSD of GMSL (with GIA applied) in mm , the smoothed (60-day Gaussian type filter) GMSL variation (GIA applied) in mm, and the smoothed (60-day Gaussian type filter) GMSL variation (with GIA applied) in mm with annual and semi-annual signal removed. Al th ecolumn have numerical values.

These GMSL variations calcualtions come from NASA Goddard Space Flight Center under the auspices of the NASA MEASUREs program and they used the GMSL that was generated using the Integrated Multi-Mission Ocean Altimeter Data for Climate Research.

The uncertainty is ±4 mm, the resolution is that sea levels have been rising and they have attributed that primarily to two factors related to global warming: the added water from melting ice sheets and glaciers and the expansion of seawater as it warms. There wasn't any missing data, the flag for missing data was "99900.000" and it never showed up.

### Importing Sea Level Data

Constructing the necessary R code to import this data set as a tidy `Table` object.

```r
sea_levels <- readr::read_table('http://climate.nasa.gov/system/internal_resources/details/original/121_
                                col_names = c('altimeter_type',
                                'merged_file_cycle_number',
                                'year_and_fraction_of_year',
                                'number_of_observations',
                                'number_of_weighted_observations',
                                'GMSL', 'sd_of_GMSL','smoothed_GMSL',
                                'GMSL_GIA_applied',
                                'sd_of_GMSL_GIA_applied',
                                'smoothed_GMSL_GIA_applied',
                                'smoothed_GMSL_GIA_applied_removed') ,
                                skip = 47)
```

```
## Parsed with column specification:
## cols(
##   altimeter_type = col_integer(),
##   merged_file_cycle_number = col_integer(),
##   year_and_fraction_of_year = col_double(),
##   number_of_observations = col_integer(),
##   number_of_weighted_observations = col_double(),
##   GMSL = col_double(),
##   sd_of_GMSL = col_double(),
##   smoothed_GMSL = col_double(),
##   GMSL_GIA_applied = col_double(),
##   sd_of_GMSL_GIA_applied = col_double(),
##   smoothed_GMSL_GIA_applied = col_double(),
##   smoothed_GMSL_GIA_applied_removed = col_double()
## )
```

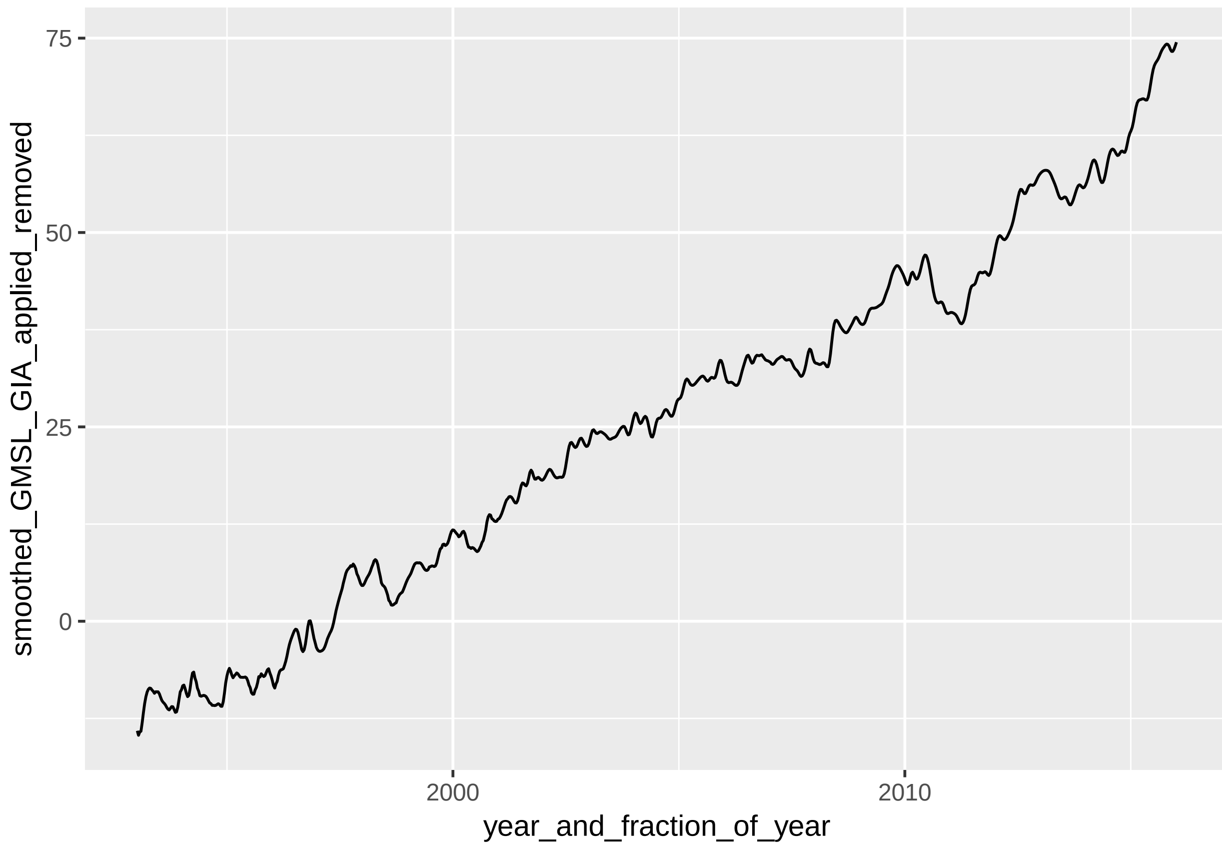
```r
sea_levels
```

```
## # A tibble: 847 x 12
##    altimeter_type merged_file_cyc~ year_and_fracti~ number_of_obser~
##             <int>            <int>            <dbl>            <int>
##  1              0               11            1993.           463892
##  2              0               12            1993.           458154
##  3              0               13            1993.           469524
##  4              0               14            1993.           419112
##  5              0               15            1993.           456793
##  6              0               16            1993.           414055
##  7              0               17            1993.           465235
##  8              0               18            1993.           463257
##  9              0               19            1993.           458542
## 10            999               20            1993.           464921
## # ... with 837 more rows, and 8 more variables:
## #   number_of_weighted_observations <dbl>, GMSL <dbl>, sd_of_GMSL <dbl>,
## #   smoothed_GMSL <dbl>, GMSL_GIA_applied <dbl>,
## #   sd_of_GMSL_GIA_applied <dbl>, smoothed_GMSL_GIA_applied <dbl>,
## #   smoothed_GMSL_GIA_applied_removed <dbl>
```

### Visualize Sea Level Data

Plot the data and describe the trends you observe.

```
ggplot(sea_levels, aes(x=year_and_fraction_of_year, y =smoothed_GMSL_GIA_applied_removed)) +geom_line()
```




## Arctic Sea Ice

- http://nsidc.org/data/G02135
- ftp://sidads.colorado.edu/DATASETS/NOAA/G02135/north/daily/data/N__seaice__extent__daily__v3.0.csv

### Question 1:

- Describe the data set: what are the columns and units?
- Where do these data come from?
- What is the uncertainty in measurement? Resolution of the data? Interpretation of missing values?

### Question 2:

Construct the necessary R code to import this data set as a tidy `Table` object.

```
library(tidyverse)

arctic_sea <- readr::read_csv("ftp://sidads.colorado.edu/DATASETS/NOAA/G02135/north/daily/data/N_seaice_

                   col_names = c("year", "month",
```

```
                              "day", "extent", "missing", "data_source") ,
                 skip = 2)
```

```
## Parsed with column specification:
## cols(
##   year = col_integer(),
##   month = col_character(),
##   day = col_character(),
##   extent = col_double(),
##   missing = col_double(),
##   data_source = col_character()
## )
```

```
arctic_sea
```

```
## # A tibble: 13,001 x 6
##     year month day   extent missing data_source
##    <int> <chr> <chr>  <dbl>   <dbl> <chr>
##  1  1978 10    26      10.2       0 ['ftp://sidads.colorado.edu/pub/DATAS~
##  2  1978 10    28      10.4       0 ['ftp://sidads.colorado.edu/pub/DATAS~
##  3  1978 10    30      10.6       0 ['ftp://sidads.colorado.edu/pub/DATAS~
##  4  1978 11    01      10.7       0 ['ftp://sidads.colorado.edu/pub/DATAS~
##  5  1978 11    03      10.8       0 ['ftp://sidads.colorado.edu/pub/DATAS~
##  6  1978 11    05      11.0       0 ['ftp://sidads.colorado.edu/pub/DATAS~
##  7  1978 11    07      11.1       0 ['ftp://sidads.colorado.edu/pub/DATAS~
##  8  1978 11    09      11.2       0 ['ftp://sidads.colorado.edu/pub/DATAS~
##  9  1978 11    11      11.3       0 ['ftp://sidads.colorado.edu/pub/DATAS~
## 10  1978 11    13      11.5       0 ['ftp://sidads.colorado.edu/pub/DATAS~
## # ... with 12,991 more rows
```

‘

### Question 3:

Plot the data and describe the trends you observe.

## Longer term trends in CO2 Records

The data we analyzed in the unit introduction included CO2 records dating back only as far as the measurements at the Manua Loa observatory. To put these values into geological perspective requires looking back much farther than humans have been monitoring atmospheric CO2 levels. To do this, we need another approach.

Ice core data:

Vostok Core, back to 400,000 yrs before present day

- Description of data set: http://cdiac.esd.ornl.gov/trends/co2/vostok.html
- Data source: http://cdiac.ornl.gov/ftp/trends/co2/vostok.icecore.co2

### Questions / Tasks:

- Describe the data set: what are the columns and units? Where do the numbers come from? There are four columns in the dataset: Depth (m), Age of the ice, (yr BP), Mean age of the air (yr BP), and CO2 concentration(ppmv). These numbers come from a 1998 ice drilling project that recovered the

deepest ice core ever. The air in the ice is about 6000 years younger than the ice itself, and the CO2 concentrations are gathered by gas chromatography.
- What is the uncertainty in measurment? Resolution of the data? Interpretation of missing values? The exact depth at which air bubbles close during the firn-ice transition and the exact age of the ice and the air is somewhat uncertain because the scientists are working backwards to determine timing. These numbers are calculated uring semiempirical models, and may not be the precise year. They are a best guess.

The resolution of the depth is to the .1 meter, the age of the ice and mean age of the air are to the 1 yr BP, and CO2 concentration is to the .1 ppmv.

Because the age data is calculated using a model, there are not missing values to interpret. The depths were also all recorded, and so were the CO2 concentrations. Therefore, this dataset does not have missing values to interpret.

- Read in and prepare data for analysis.

```
ice_core <- readr::read_delim('http://cdiac.ornl.gov/ftp/trends/co2/vostok.icecore.co2',
                              delim ='\t',
                              col_names = c('depth', 'age_ice', 'mean_age_air',
                              'co2_concentration'),
                              comment = '*',
                              skip = 4)
```

```
## Parsed with column specification:
## cols(
##   depth = col_double(),
##   age_ice = col_integer(),
##   mean_age_air = col_integer(),
##   co2_concentration = col_double()
## )
```

```
ice_core
```

```
## # A tibble: 363 x 4
##     depth age_ice mean_age_air co2_concentration
##     <dbl>   <int>        <int>             <dbl>
##   1  149.    5679         2342              285.
##   2  173.    6828         3634              273.
##   3  177.    7043         3833              268.
##   4  229.    9523         6220              262.
##   5  250.   10579         7327              255.
##   6  266    11334         8113              260.
##   7  303.   13449        10123              262.
##   8  321.   14538        11013              264.
##   9  332.   15208        11326              245.
##  10  342.   15922        11719              238.
## # ... with 353 more rows
```

- Reverse the ordering to create a chronological record.

```
arrange(ice_core, desc(age_ice))
```

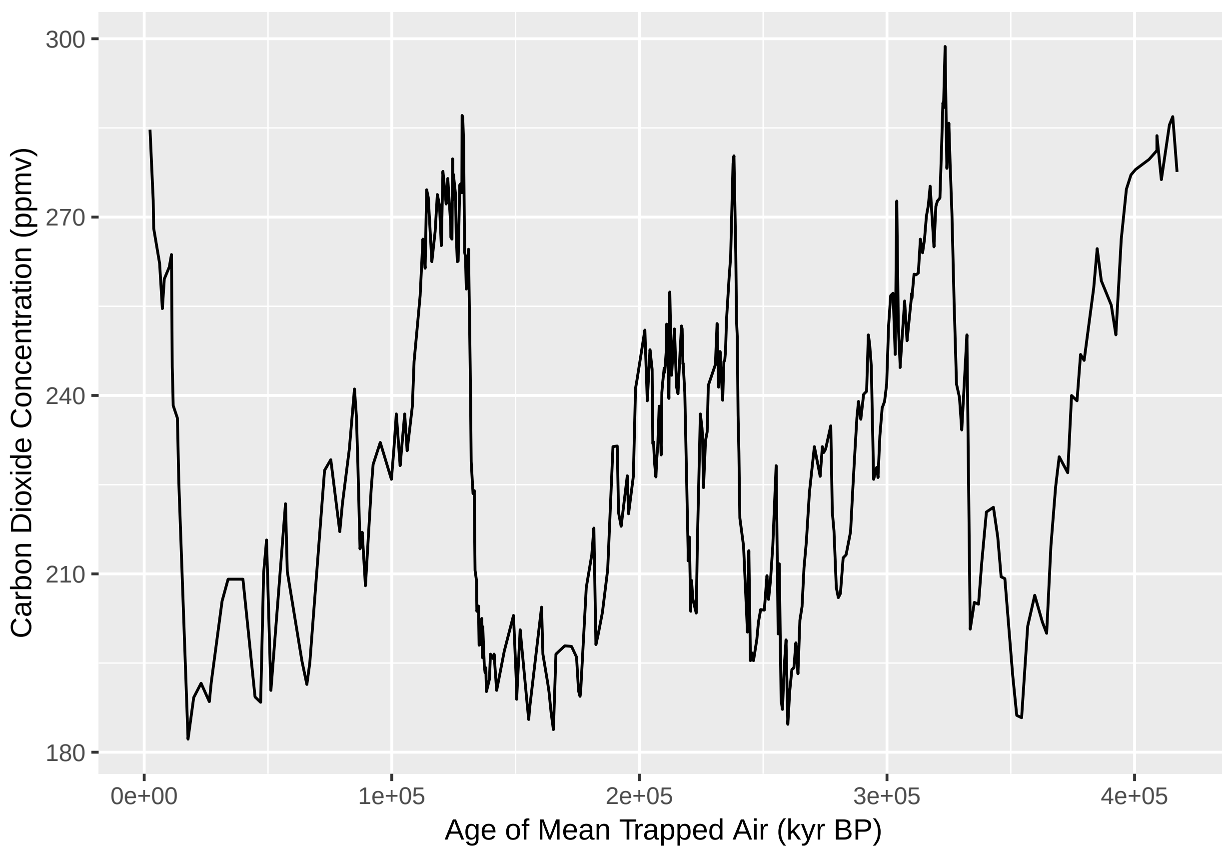
```
## # A tibble: 363 x 4
##     depth age_ice mean_age_air co2_concentration
##     <dbl>   <int>        <int>             <dbl>
##   1 3304.  419328       417160              278.
##   2 3301.  417638       415434              287.
```

```
##   3 3299.   416332       414085              286.
##   4 3293.   413010       410831              276.
##   5 3289.   411202       409022              281.
##   6 3289.   411202       409022              284.
##   7 3284.   408236       405844              280.
##   8 3274.   403173       400390              278
##   9 3271.   401423       398554              277.
## 10 3268.   399733       396713              275.
## # ... with 353 more rows
```

I assumed that the plot I created intially was in chronological order, going from earliest to latest year, but here is the reverse of that ordering.

- Plot data

```
ggplot(data = ice_core, aes(x=mean_age_air, y = co2_concentration))+
  geom_line()+
  labs(x= 'Age of Mean Trapped Air (kyr BP)', y = 'Carbon Dioxide Concentration (ppmv)')
```



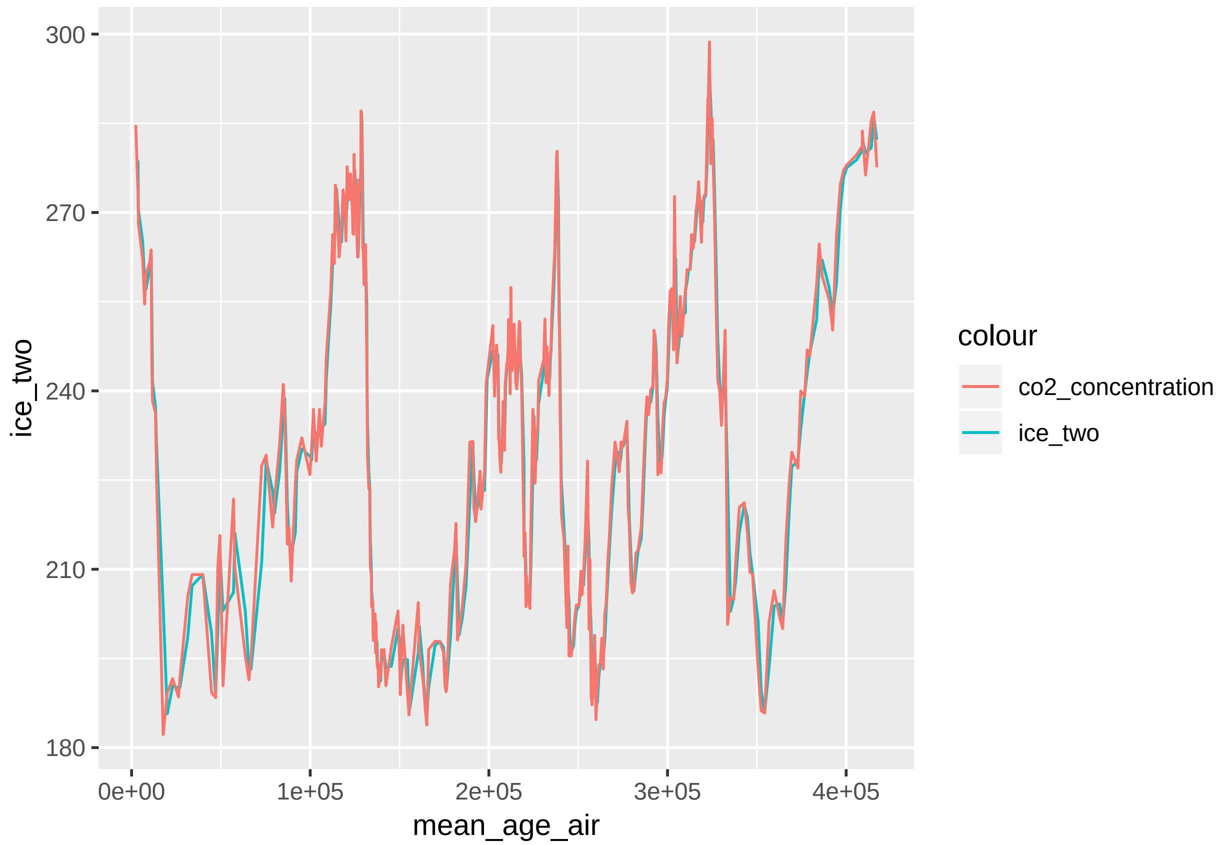

- Consider various smoothing windowed averages of the data.

```
ice_core_co2_averages <- ice_core %>%
  mutate(ice_two = roll_mean(co2_concentration, n = 2, align = "right", fill = NA)) %>%
  mutate(ice_five = roll_mean(co2_concentration, n = 5, align = "right", fill = NA)) %>%
  mutate(ice_ten = roll_mean(co2_concentration, n = 10, align = "right", fill = NA)) %>%
  mutate(ice_twenty = roll_mean(co2_concentration, n =  20, align = 'right', fill = NA))

ggplot(ice_core_co2_averages, aes(x=mean_age_air))+
```

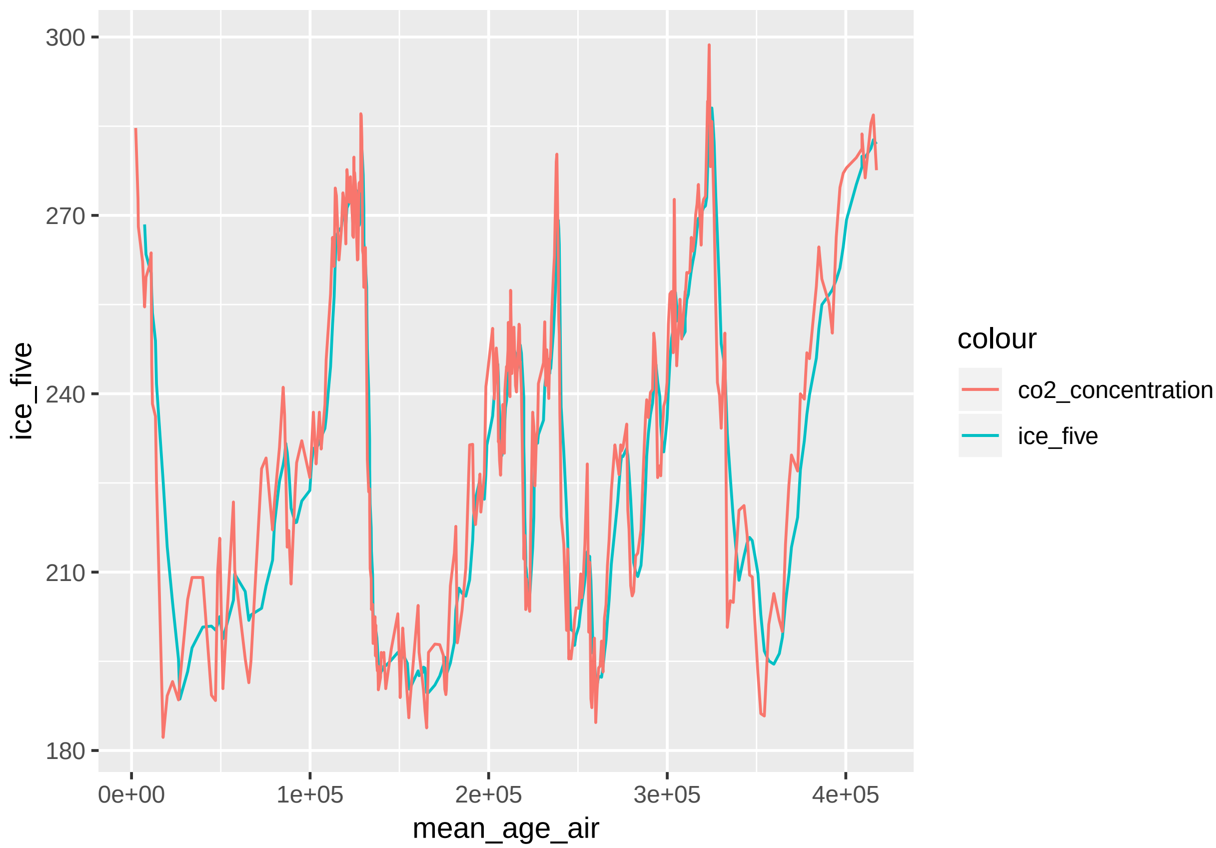
```r
  geom_line(aes(y=ice_two, color = 'ice_two'))+
  geom_line(aes(y = co2_concentration, color = 'co2_concentration'))
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



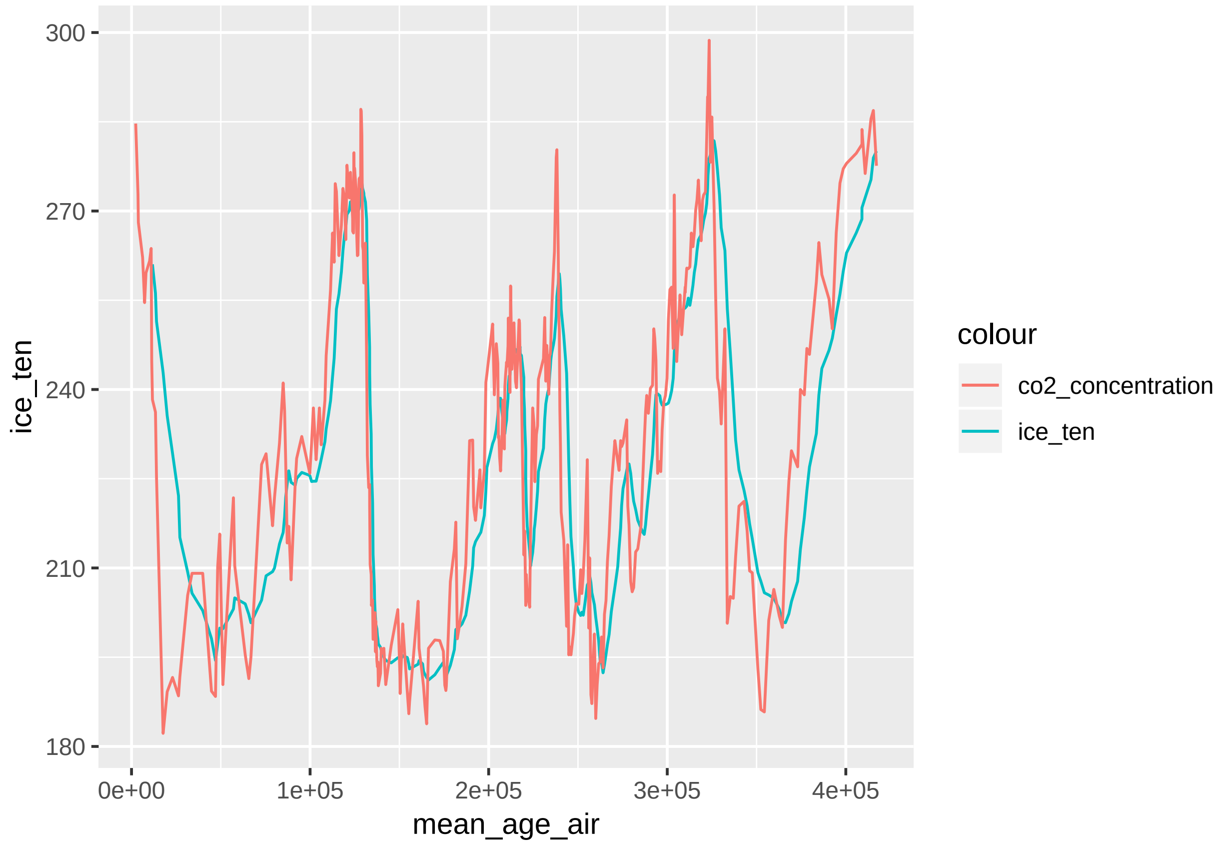

```r
ggplot(ice_core_co2_averages, aes(x=mean_age_air))+
  geom_line(aes(y=ice_five, color = 'ice_five'))+
  geom_line(aes(y = co2_concentration, color = 'co2_concentration'))
```

```
## Warning: Removed 4 rows containing missing values (geom_path).
```



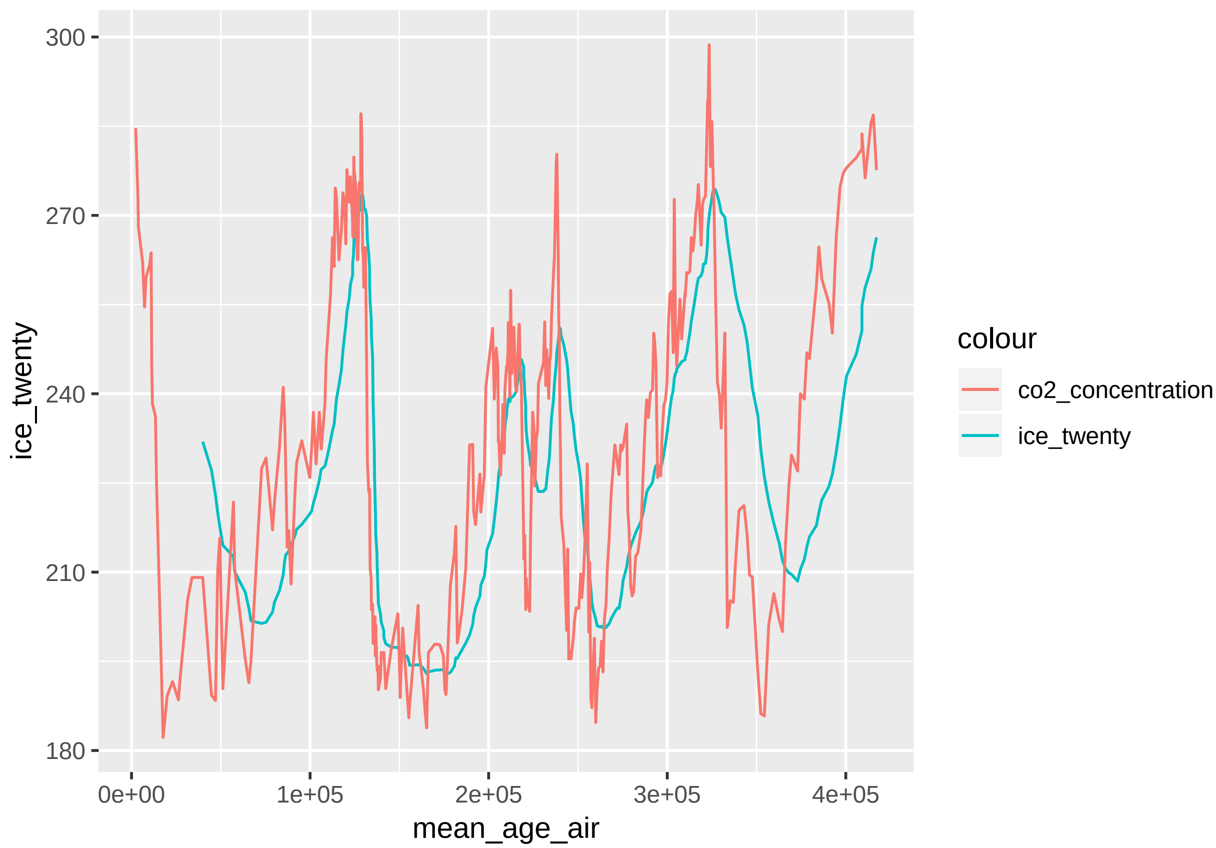

```
ggplot(ice_core_co2_averages, aes(x=mean_age_air))+
  geom_line(aes(y=ice_ten, color = 'ice_ten'))+
  geom_line(aes(y = co2_concentration, color = 'co2_concentration'))
```

```
## Warning: Removed 9 rows containing missing values (geom_path).
```




```
ggplot(ice_core_co2_averages, aes(x=mean_age_air))+
  geom_line(aes(y=ice_twenty, color = 'ice_twenty'))+
  geom_line(aes(y = co2_concentration, color = 'co2_concentration'))
```

```
## Warning: Removed 19 rows containing missing values (geom_path).
```


300
270
240
210
180
ice_twenty
0e+00
1e+05
2e+05
3e+05
4e+05
mean_age_air
colour
co2_concentration
ice_twenty