

Module 1: Climate change

```
library(tidyverse)
```

Team members, etc

List your team member(s): Nikhil Shah

Include a link to your forked GH repository: <https://github.com/nikhilshah7/CLES131-module1-climate#>

Include a link to your .qmd file: <https://github.com/nikhilshah7/CLES131-module1-climate/blob/main/module1.qmd>

CO2 trends from Mauna Loa

We will begin at the global scale by plotting the famed Mauna Loa CO₂ curve at monthly intervals, with the twin objectives of (1) making layered plots with ‘ggplot2’, a package within the ‘tidyverse’ collection widely adopted by the R data science community, and (2) refreshing understanding of the seasonal cycle in atmospheric [CO₂].

With data you’ve collected yourself, it is more common to store locally as a csv. However, we can use tidyverse tools to read directly from files stored online and will do so here to get the latest data.

```
co2 <- read_table("https://gml.noaa.gov/webdata/ccgg/trends/co2/co2_mm_mlo.txt",
  comment="#",
  col_names = c("year", "month", "decimal_date", "average",
    "interpolated", "trend", "days"),
  na = c("-1", "-99.99"))
```

You can examine these data in several ways. One option is to use your mouse in the Environment tab and click on the dataset name. Alternatively, you can type `View(co2)` in your

console to achieve the same outcome. Below are some other common options to understand the data structure at a glance.

```
str(co2)
```

```
spc_tbl_ [810 x 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ year      : num [1:810] 1958 1958 1958 1958 1958 ...
 $ month     : num [1:810] 3 4 5 6 7 8 9 10 11 12 ...
 $ decimal_date: num [1:810] 1958 1958 1958 1958 1959 ...
 $ average   : num [1:810] 316 317 318 317 316 ...
 $ interpolated: num [1:810] 314 315 315 315 315 ...
 $ trend     : num [1:810] NA NA NA NA NA NA NA NA NA ...
 $ days      : num [1:810] -9.99 -9.99 -9.99 -9.99 -9.99 -9.99 -9.99 -9.99 -9.99 ...
- attr(*, "problems")= tibble [810 x 5] (S3: tbl_df/tbl/data.frame)
 ..$ row      : int [1:810] 1 2 3 4 5 6 7 8 9 10 ...
 ..$ col      : chr [1:810] NA NA NA NA ...
 ..$ expected: chr [1:810] "7 columns" "7 columns" "7 columns" "7 columns" ...
 ..$ actual   : chr [1:810] "8 columns" "8 columns" "8 columns" "8 columns" ...
 ..$ file     : chr [1:810] "'https://gml.noaa.gov/webdata/ccgg/trends/co2/co2_mm_mlo.txt'"
- attr(*, "spec")=
 .. cols(
 ..   year = col_double(),
 ..   month = col_double(),
 ..   decimal_date = col_double(),
 ..   average = col_double(),
 ..   interpolated = col_double(),
 ..   trend = col_double(),
 ..   days = col_double()
 .. )
```

```
head(co2)
```

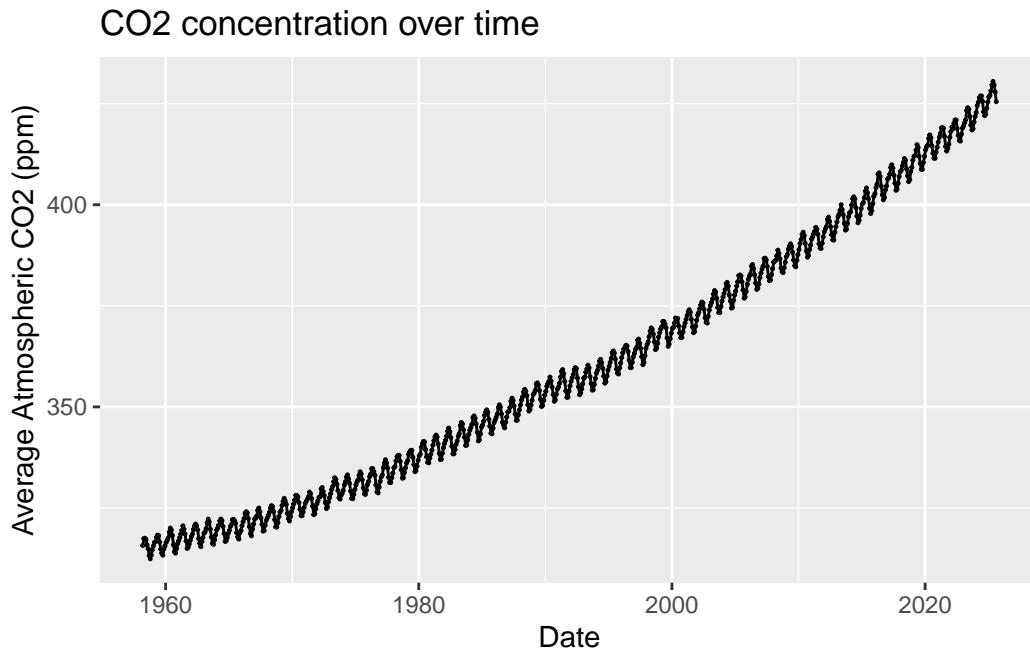
```
# A tibble: 6 x 7
  year month decimal_date average interpolated trend  days
  <dbl> <dbl>      <dbl>    <dbl>      <dbl> <dbl> <dbl>
1  1958     3      1958.    316.        314.    NA  -9.99
2  1958     4      1958.    317.        315.    NA  -9.99
3  1958     5      1958.    318.        315.    NA  -9.99
4  1958     6      1958.    317.        315.    NA  -9.99
5  1958     7      1959.    316.        315.    NA  -9.99
6  1958     8      1959.    315.        316.    NA  -9.99
```

Q1 (1 point)

Create a timeseries plot of average $[\text{CO}_2]$. Use `as.Date()` to create a formatted date for the x-axis. Label both the y-axis and add a line to connect the points.

```
co2 <- mutate(co2, date = as.Date(str_c(year, month, '15', sep = '-'),
                                       format = "%Y-%m-%d"))

ggplot(co2, aes(x = date,
                y = average)) +
  geom_point(size = 0.2) +
  geom_line() +
  labs(x = 'Date',
       y = 'Average Atmospheric CO2 (ppm)',
       title = 'CO2 concentration over time')
```

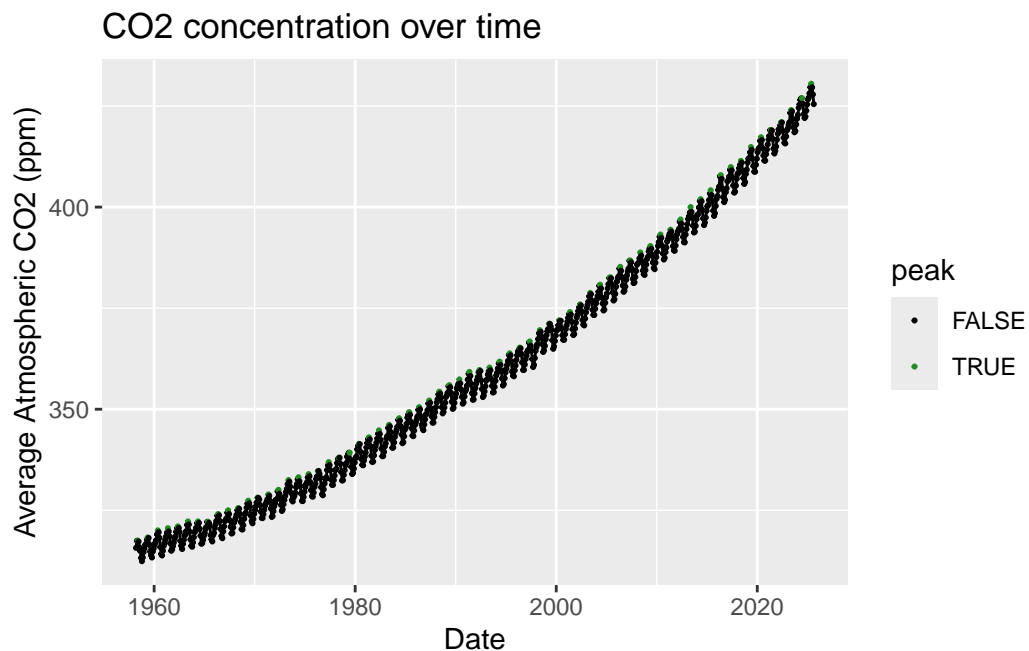


Q2 (1 point)

What months are the $[\text{CO}_2]$ values at the maximum? Create another dataframe with these months selected and create a plot like above, but with the maximum months colored in “forest-green”.

```
co2 <- co2 |>
  group_by(year) |>
  mutate(peak = average == max(average)) |>
  ungroup()

ggplot(co2, aes(x = date,
                 y = average)) +
  geom_line() +
  geom_point(aes(color = peak),
             size = 0.4) +
  scale_color_manual(values = c('black', 'forestgreen')) +
  labs(x = 'Date',
       y = 'Average Atmospheric CO2 (ppm)',
       title = 'CO2 concentration over time')
```



Q3 (1 point)

What explains the seasonal variation in $[\text{CO}_2]$? After reading Lewandowsky et al. 2016, how might you go about separating the long term trend and the seasonal variation in $[\text{CO}_2]$?

- $[\text{CO}_2]$ is affected by photosynthesis, which does not occur at a constant rate year-round because of plant growth cycles. Fluctuations in photosynthetic activity would cause

correlated fluctuations in the amount of CO₂ in the atmosphere.

- Lewandowsky et al differentiate annual 15-year trends against the 44-year average global temperature increase. You could do something similar here, use the average [CO₂] increase over a long period of time and take the measurement at each month and see how it deviates from that mean, and then plot that difference or a z-score, isolating the seasonal fluctuation.

Global temperature

Next, we turn to a global temperature anomaly dataset from NASA GISS, similar to the GMST used by Lewandowsky et al. 2016.

Q4 (1 point)

Read some [background](#) as well as the [simplified documentation](#) for this dataset.

How are the measurements made? Describe each column in the dataset and its units. What are the resolution of the data?

- Meteorological data is gathered from stations on both the surface of the land and sea every year. Gaps are filled in using patterns from stations up to 1200km away.
- Column 1 is the year, which is just a year. Column 2 is the difference between the actual temperature value for that year, and the 1951-1980 average temperature, in degrees C. Column 3 is used to smooth the line chart, and is the 5 year rolling average, also in degrees C.
- Annual, to the nearest 0.01 C

Q5 (1 point)

Construct code to import dataset in a format that can be used for transforming and visualization.

```
giss <- read_table(file = 'data/giss.txt',  
                  skip = 5,  
                  col_names = c("Year", "No_Smoothing", "Lowess5"))
```

```
-- Column specification -----  
cols(  
  Year = col_double(),  
  No_Smoothing = col_double(),
```

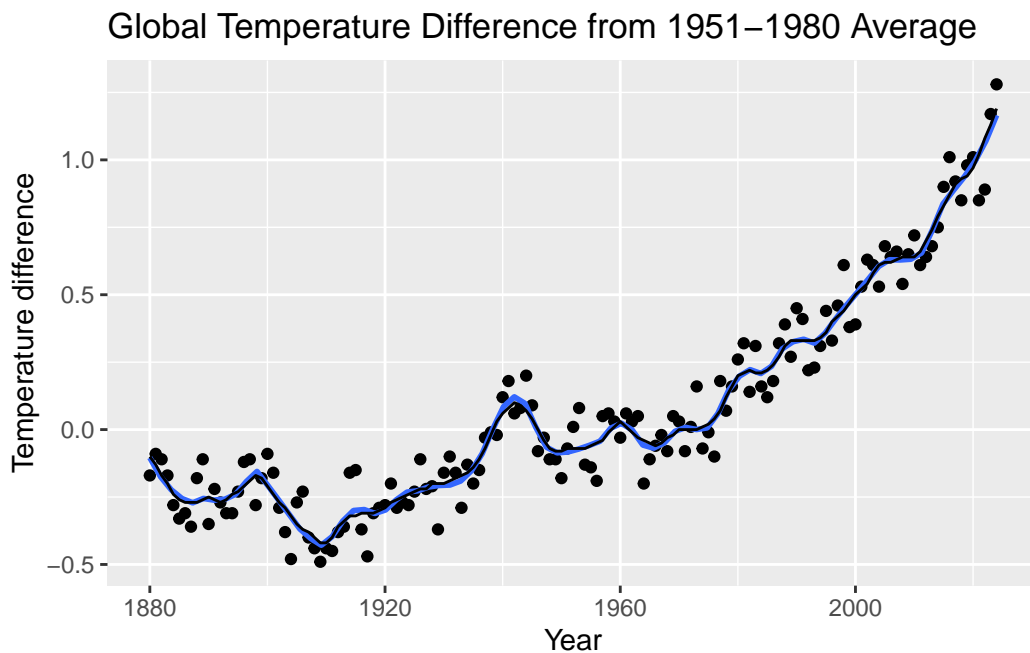
```
Lowess5 = col_double()
)
```

Q6 (1 point)

Plot the trend in global mean temperature over time. Describe what you see and how you interpret the patterns you observe.

```
ggplot(giss) +
  geom_point(aes(x = Year,
                 y = No_Smoothing)) +
  geom_smooth(aes(x = Year,
                 y = No_Smoothing),
             se = F,
             span = 0.12) +
  geom_line(aes(x = Year,
               y = Lowess5)) +
  labs(x = 'Year',
       y = 'Temperature difference',
       title = 'Global Temperature Difference from 1951–1980 Average')
```

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



Q7 (1 point)

Consider the “pause” analyzed in Lewandowsky et al. 2016. Do you see evidence for such a pause?

- The smoothing used by NASA shows the decadal fluctuation in global temperatures, where independent of human-caused global warming, it increases and decreases marginally every 12-15 years. Both effects combined results 6-8 year periods of rapid temperature increase, followed by 6-8 years of slow or zero temperature increase. However, long term patterns show that while in 2005 the temperature might not have been increasing year-to-year, the long-term pattern shows that global warming would not be stopping, and the temperature experienced its highest ever slope immediately after, disproving the idea of global warming stopping.

Bonus 1 (1 point)

Recreate the analysis from Lewandowsky’s Fig. 1.

```
giss_sample <- giss |>
  mutate('fifteen' = (lead(No_Smoothing, n = 7) - lag(No_Smoothing, n = 7))) |>
  mutate('fifteen_decadal' = round(((fifteen / 1.5)), 2)) |>
  filter(Year >= 1970 & Year <= 2014) |>
  mutate('z' = (fifteen_decadal - 0.1601) / sd(fifteen_decadal, na.rm = T)) |>
  mutate(direction = ifelse(z >= 0, 'above 0', 'below 0'))

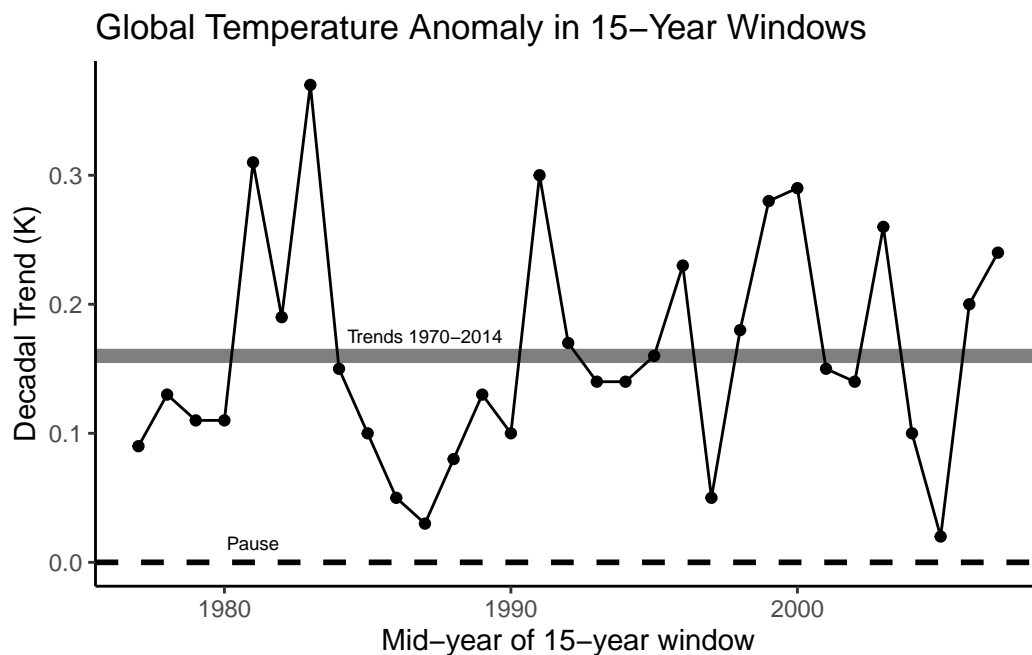
giss_sample |>
  filter(Year >= 1977 & Year <= 2007) |>
  ggplot(aes(x = Year,
             y = fifteen_decadal)) +
  geom_point() +
  geom_line() +
  geom_hline(yintercept = 0.16,
             size = 2.5,
             alpha = 0.5) +
  geom_hline(yintercept = 0,
             linetype = 'dashed',
             size = 1) +
  annotate(geom = "text",
          x = 1981,
          y = 0.015,
          label = 'Pause',
          size = 2.5) +
```

```

annotate(geom = "text",
  x = 1987,
  y = 0.175,
  label = 'Trends 1970-2014',
  size = 2.5) +
labs(x = 'Mid-year of 15-year window',
  y = 'Decadal Trend (K)',
  title = 'Global Temperature Anomaly in 15-Year Windows') +
theme_classic()

```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
 i Please use `linewidth` instead.



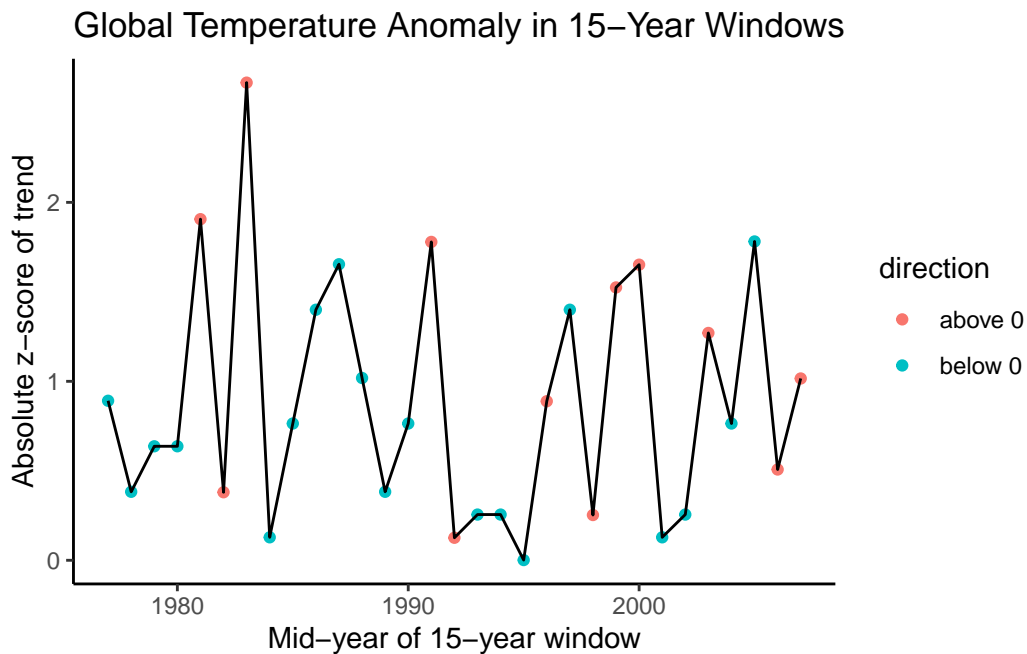
```

giss_sample |>
  filter(Year >= 1977 & Year <= 2007) |>
  ggplot(aes(x = Year,
    y = abs(z))) +
  geom_point(aes(color = direction)) +
  geom_line() +
  labs(x = 'Mid-year of 15-year window',
    y = 'Absolute z-score of trend',

```



```
title = 'Global Temperature Anomaly in 15-Year Windows' +  
theme_classic()
```



Long term records of CO₂

The data from Mauna Loa go back as far as modern instrumentation. To put these values into geological perspective, we must go back much further than humans have been monitoring [CO₂]. Ice sheets contain layers of snow with trapped air bubbles and can be cored and dated like tree rings.

Previously hosted by Oak Ridge National Lab, the [data](#) associated with the famed Vostok Ice Core is now archived on Zenodo. Skim [Barnola et al. 1987](#) as needed.

After downloading this dataset, create a folder entitled **data** and save the file within the folder with a .txt extension. Read this file into R, making sure you understand the column names and units.

Q8 (1 point)

After downloading this dataset, create a folder entitled **data** and save the file within the folder with a .txt extension. Read this file into R. What are the columns and units? How were these data obtained?

```
vostok <- read_table(file = 'data/vostok.icecore.co2.txt',
                     skip = 20,
                     col_names = c("Depth",
                                   "IceAge",
                                   "AirTime",
                                   "average"))
```

```
-- Column specification -----
cols(
  Depth = col_double(),
  IceAge = col_double(),
  AirTime = col_double(),
  average = col_double()
)
```

- Depth in the ice in meters
- Age of the ice and of the air in years before present
- Concentration of carbon dioxide in parts per million by volume
- Data was obtained by taking ice samples from the Vostok Ice Core and looking at pores in the ice to sample air. Dating is done by looking at yearly layers of the ice.

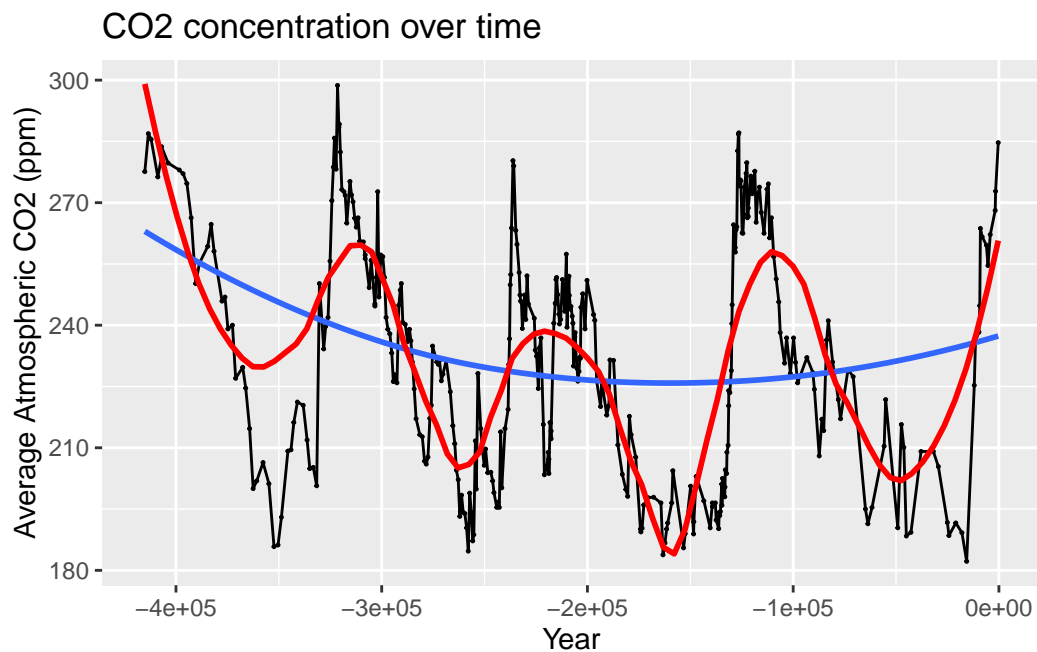
Q9 (1 point)

Transform the data as needed to create a plot in chronological order. Add at least two different smoothing window averages to your plot.

```
vostok <- vostok |>
  mutate(year = 2025 - AirTime)

ggplot(data = vostok,
       mapping = aes(x = year,
                     y = average)) +
  geom_point(size = 0.2) +
  geom_line() +
  geom_smooth(se = F, span = 2) +
  geom_smooth(se = F, span = 0.3, color = 'red') +
  labs(x = 'Year',
       y = 'Average Atmospheric CO2 (ppm)',
       title = 'CO2 concentration over time')
```

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



Q10 (1 point)

Join your data with the Mauna Loa data and plot together, considering different colors for the different datasets.

```
co2_yearly <- co2 |>
  group_by(year) |>
  summarize(yearly_average = mean(average)) |>
  mutate(source = 'Mauna Loa')

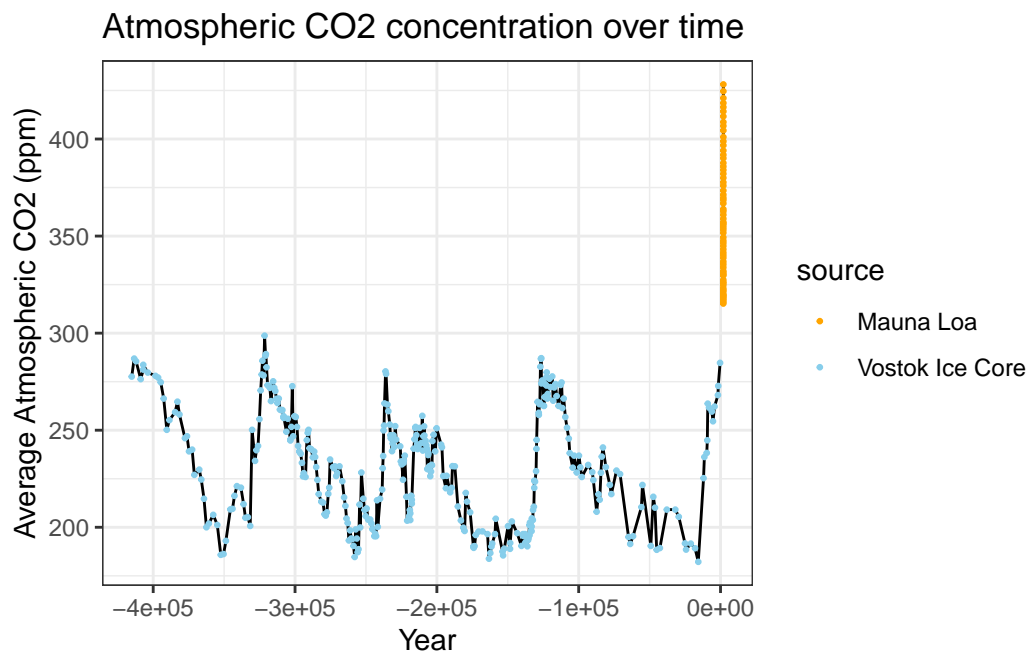
vostok <- vostok |>
  mutate(source = 'Vostok Ice Core')

ggplot() +
  geom_line(data = vostok,
            mapping = aes(x = year,
                          y = average)) +
  geom_point(data = co2_yearly,
             mapping = aes(x = year,
```

```

        y = average,
        color = source),
    size = 0.5) +
geom_line(data = co2_yearly,
    mapping = aes(x = year,
        y = yearly_average)) +
geom_point(data = co2_yearly,
    mapping = aes(x = year,
        y = yearly_average,
        color = source),
    size = 0.5) +
labs(x = 'Year',
    y = 'Average Atmospheric CO2 (ppm)',
    title = 'Atmospheric CO2 concentration over time',
    legend = 'Data Source') +
scale_color_manual(values = c("Vostok Ice Core" = "skyblue",
    "Mauna Loa" = "orange")) +
theme_bw()

```



Bonus 2 (1 point)

Write a catchy title and short description in the style of a NYtimes infographic, contrasting the rate of anthropogenic vs. background climate change for a general public audience.

Atmospheric CO₂ changes naturally. Humans are increasing it 500 times faster than usual.

The atmosphere naturally fluctuates between about 200 and 300 parts per million (ppm) of CO₂ over long periods of time. Each of these cycles takes about 100,000 years. According to historical samples of the atmosphere extracted from ice cores, in the most recent cycle, the atmosphere gained 100 ppm of CO₂ in a span of about 25,000 years. Humans have managed to inject another 100 ppm into the atmosphere in just the last 53 years, and the level of carbon in the atmosphere is now 44% higher than its peak in the last 400,000 years.