

What is time series data?

MANIPULATING TIME SERIES DATA IN R

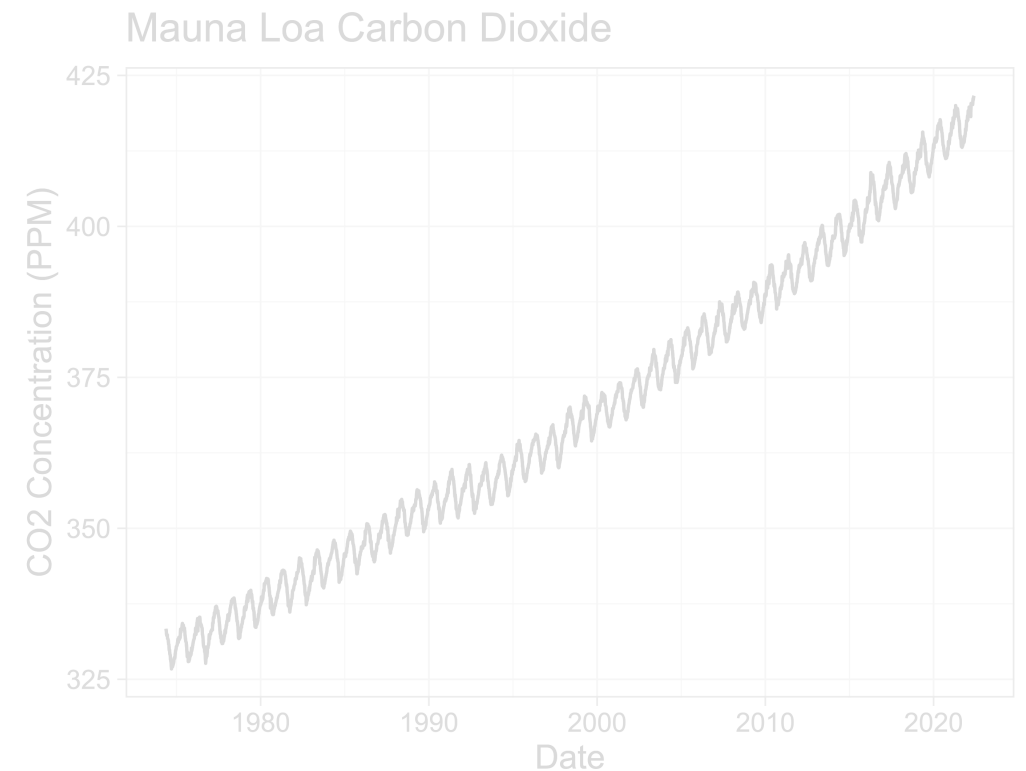
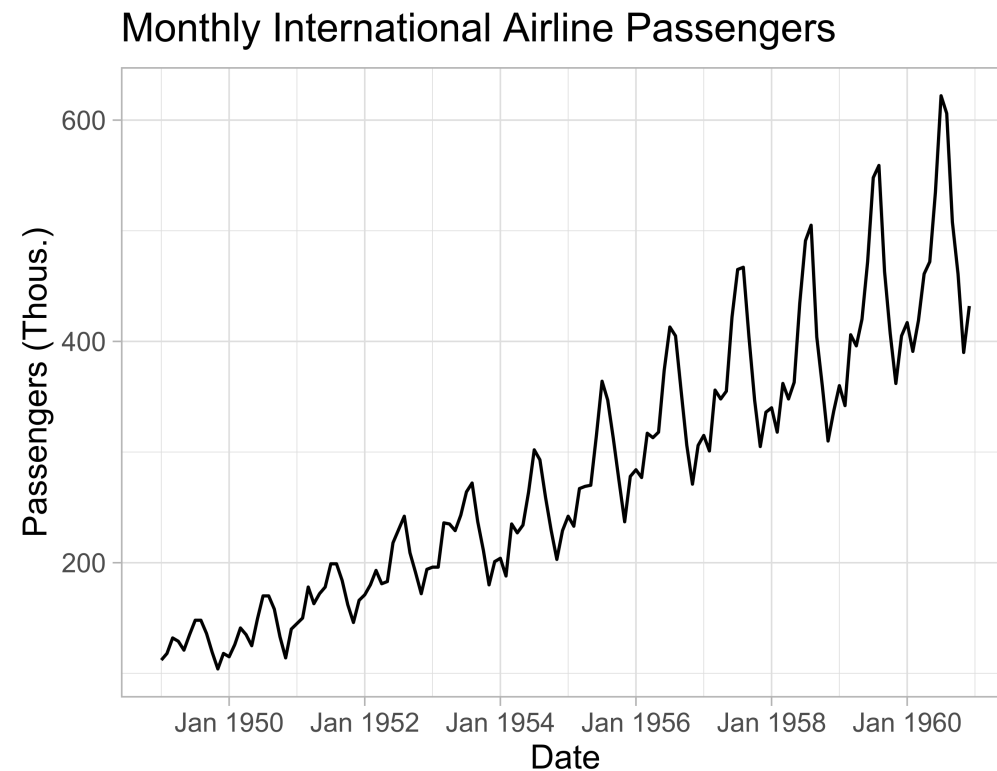


Harrison Brown

Graduate Researcher in Geography

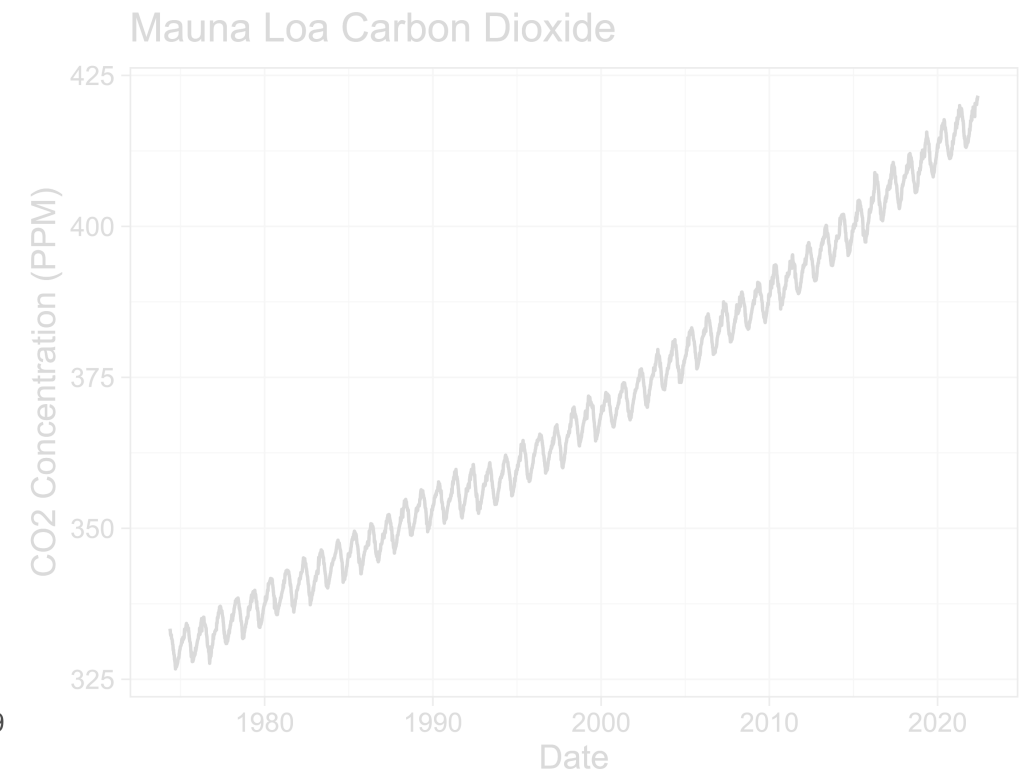
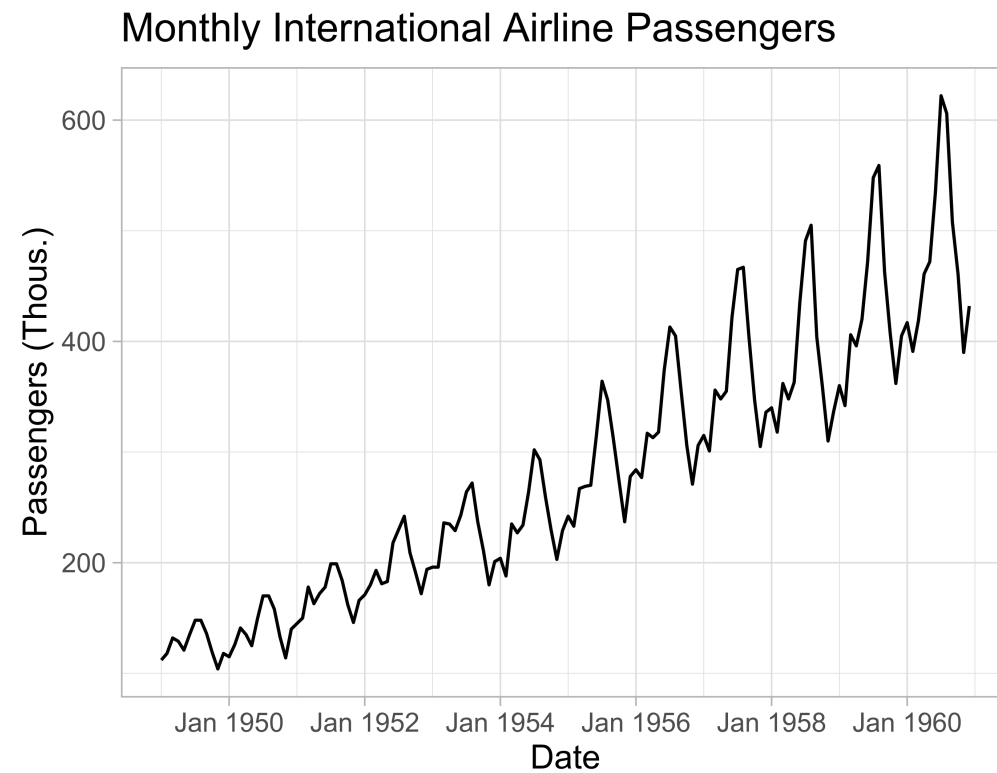
What is a time series?

- A time series is a collection of data points ordered sequentially over time.



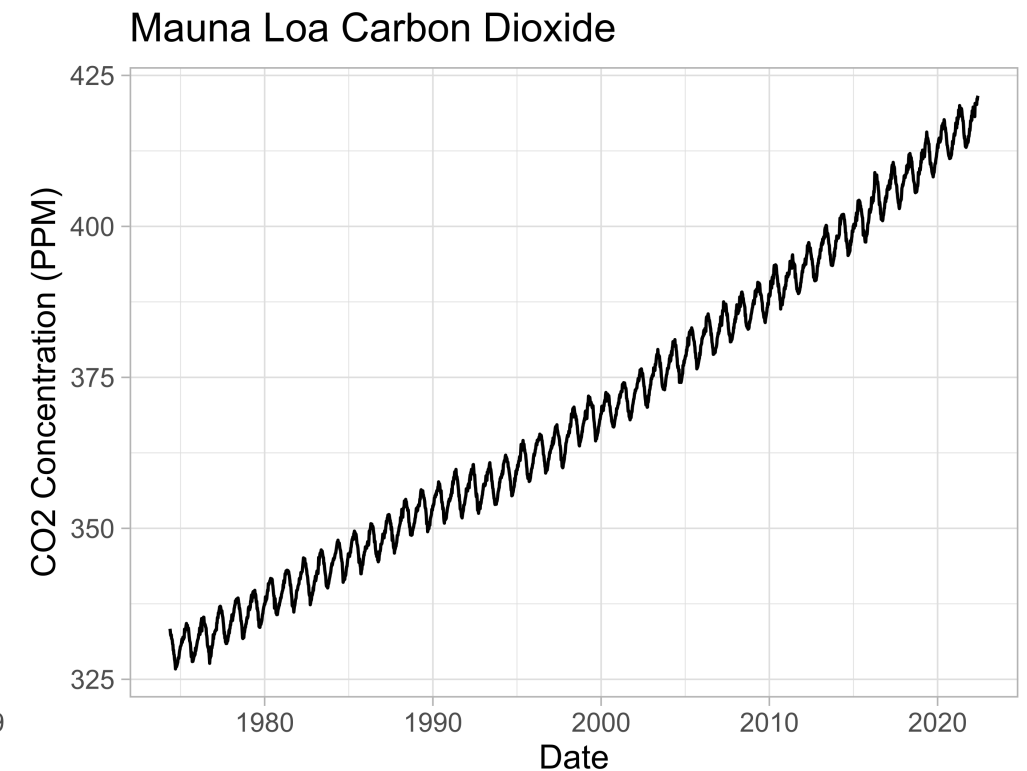
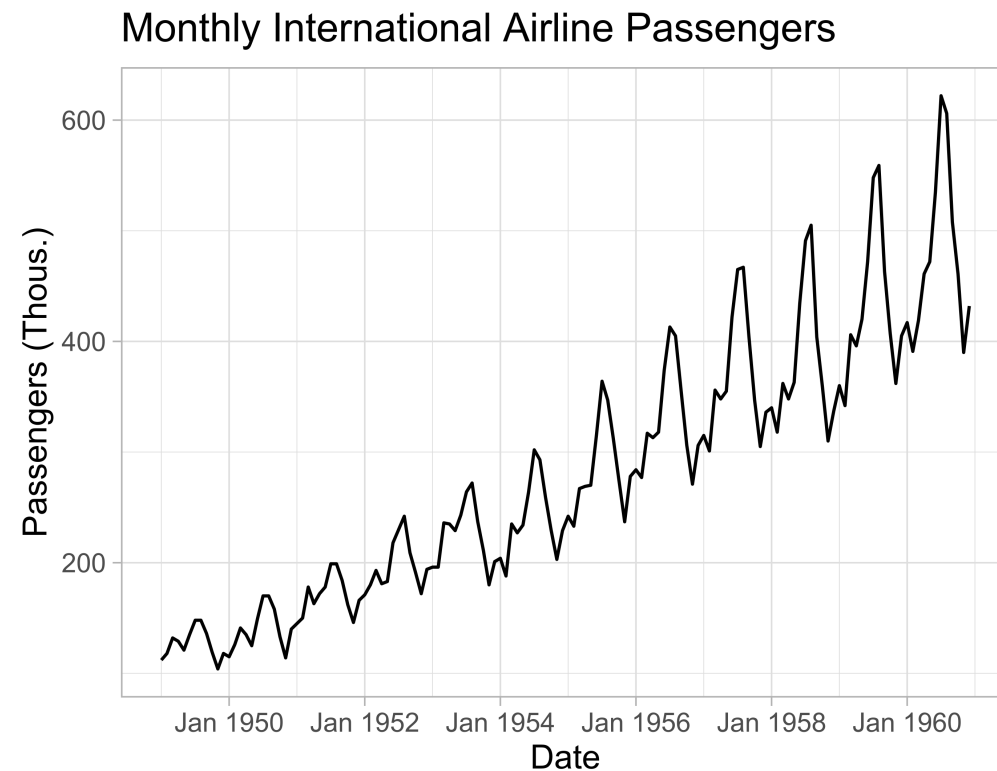
What is a time series?

- A time series is a collection of data points ordered sequentially over time.



What is a time series?

- A time series is a collection of data points ordered sequentially over time.



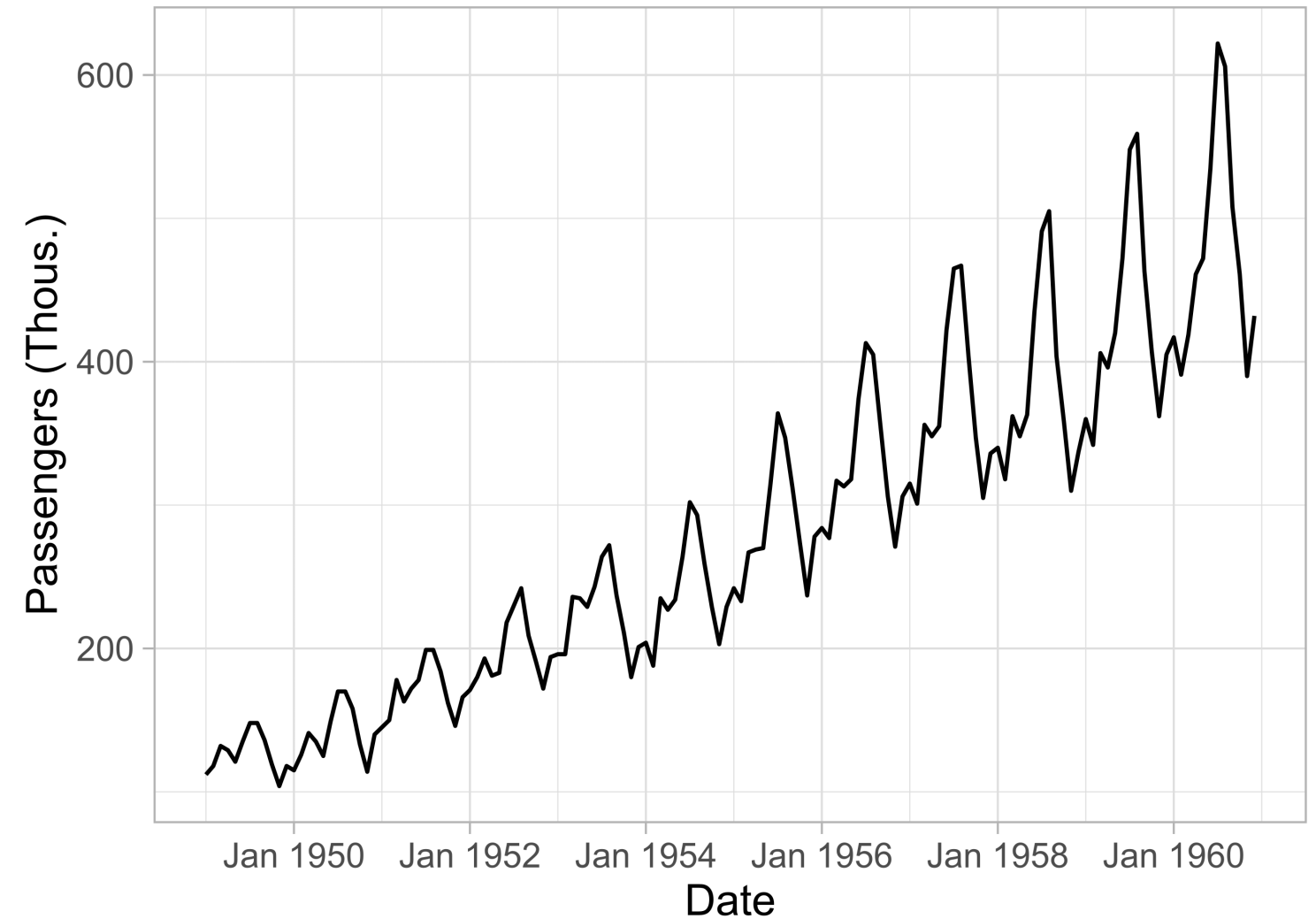
Time series in R

- In R, time series datasets are usually `ts` or `zoo` objects.

AirPassengers

```
      Jan Feb Mar Apr May Jun Jul Aug ...  
1949 112 118 132 129 121 135 148 148 ...  
1950 115 126 141 135 125 149 170 170 ...  
1951 145 150 178 163 172 178 199 199 ...  
1952 171 180 193 181 183 218 230 242 ...  
1953 196 196 236 235 229 243 264 272 ...  
...
```

Monthly International Airline Passengers



Summary statistics

```
AirPassengers
```

```
      Jan Feb Mar Apr May Jun Jul Aug ...  
1949 112 118 132 129 121 135 148 148 ...  
1950 115 126 141 135 125 149 170 170 ...  
1951 145 150 178 163 172 178 199 199 ...  
1952 171 180 193 181 183 218 230 242 ...  
1953 196 196 236 235 229 243 264 272 ...  
...
```

```
base::summary(AirPassengers)
```

```
Min.      104  
1st Qu.   180  
Median   265.5  
Mean     280.2986  
3rd Qu.  360.5  
Max.      622
```

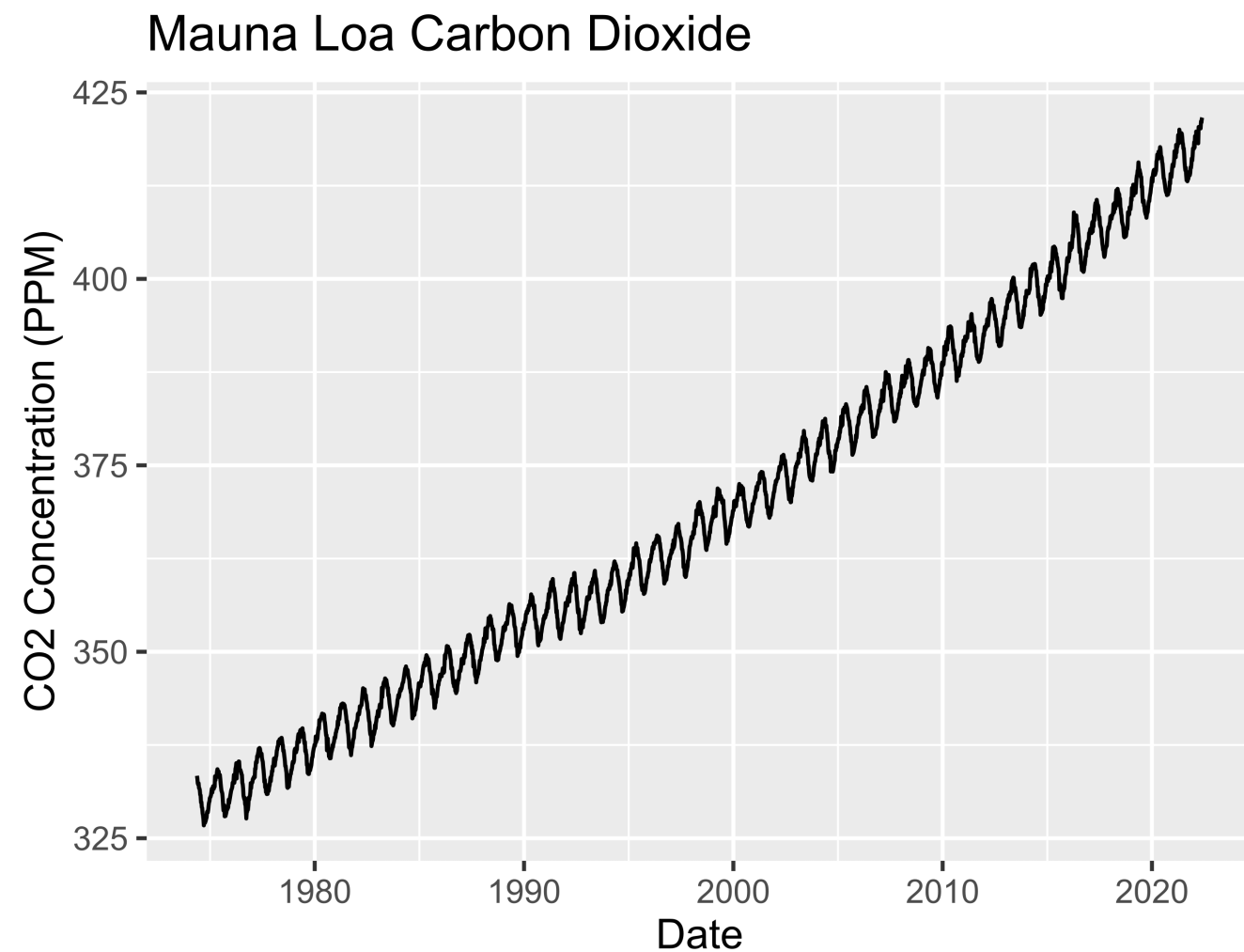
Why use time series objects in R?

Time series objects:

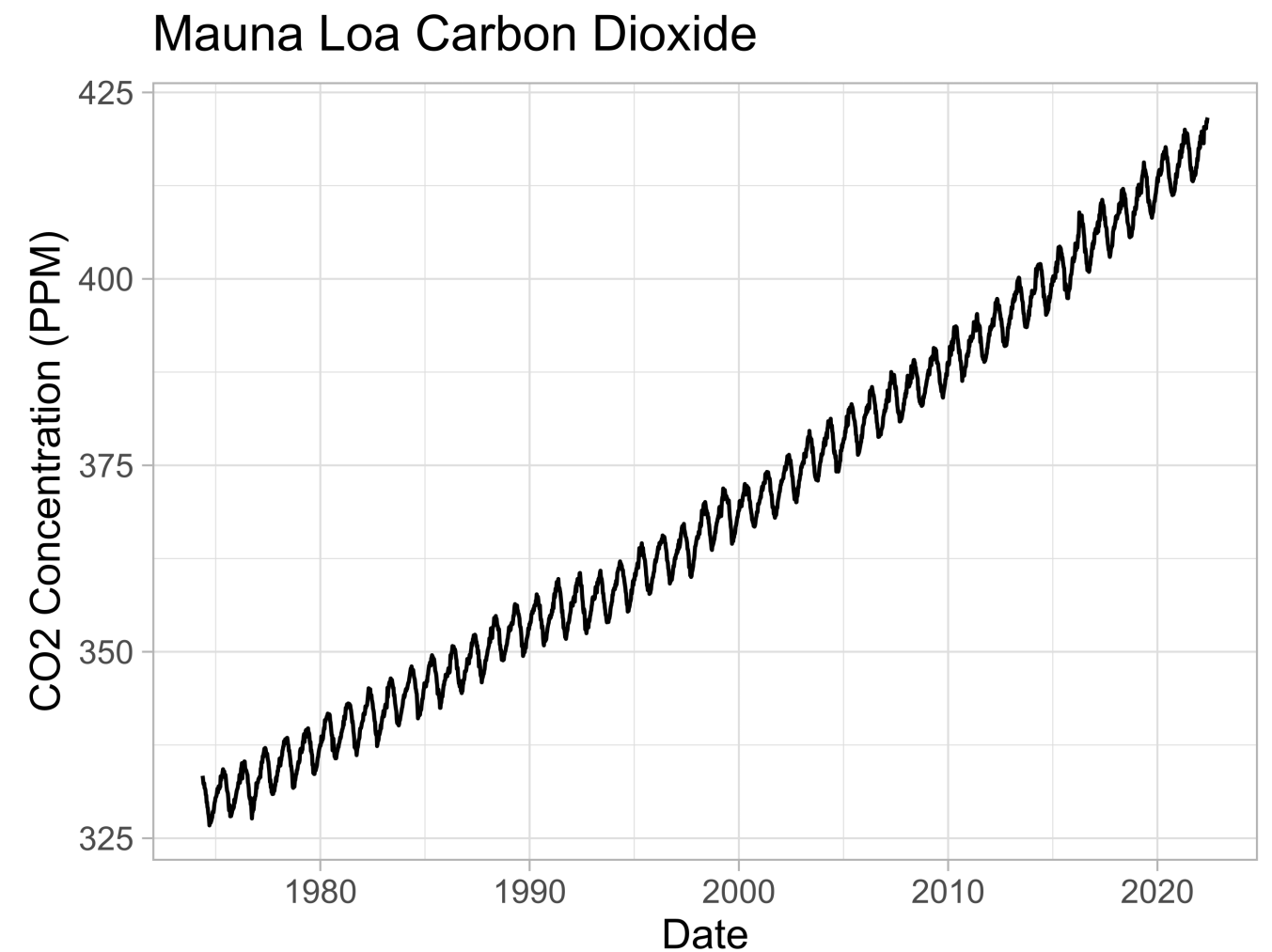
- work better with specialized tools,
- keep track of date and time,
- aid in smoother workflows!

Plotting with autoplot

```
autoplot(mauna_loa)
```



```
autoplot(mauna_loa) +  
  theme_light()
```



Let's practice!

MANIPULATING TIME SERIES DATA IN R

Temporal data classes in R

MANIPULATING TIME SERIES DATA IN R



Harrison Brown

Graduate Researcher in Geography

Date-time classes

- `numeric`
 - Integers, floating point, etc.
 - August 9, 2022 = `19223`
 - Number of days since Jan. 1, 1970
- `character`
 - Strings of text, names, etc.
 - August 9, 2022 = `"2022-08-09"`
 - August 9, 2022 = `"August 9, 2022"`

Date-time classes

- `Date`

- Dates, days of the year
- August 9, 2022 = `"2022-08-09"`

```
start_date <- as_date("2022-08-01")
end_date <- as_date("2022-08-09")

end_date - start_date
```

Time difference of 8 days

- `POSIXct`

- Dates **and** times, time zones
- August 9, 2022, 4:17 p.m. =
 - `"2022-08-09 20:17:00 UTC"`

```
start_time <-
  as.POSIXct("2022-08-01 08:57:00 EDT")
end_time <-
  as.POSIXct("2022-08-09 15:30:00 EDT")
end_time - start_time
```

Time difference of 8.272917 days

Lubridate

```
lubridate::as_date()
```

```
my_date <- as_date("2022-01-20")
```

```
my_date
```

```
[1] "2022-01-20"
```

```
class(my_date)
```

```
[1] "Date"
```

`as_date()` versus `base::as.Date()`

- Better support for time zones
- Warnings for invalid date formats
- Easier conversions from `numeric`
- ...

Testing data classes

```
class("2022-08-01")
```

```
[1] "character"
```

```
is.numeric(10494)
```

```
[1] TRUE
```

```
is.character("April 01, 2003")
```

```
[1] TRUE
```

```
is.Date(lubridate::today())
```

```
[1] TRUE
```

```
is.POSIXct(lubridate::now())
```

```
[1] TRUE
```

Let's practice!

MANIPULATING TIME SERIES DATA IN R

Formatting dates in R

MANIPULATING TIME SERIES DATA IN R



Harrison Brown

Graduate Researcher in Geography

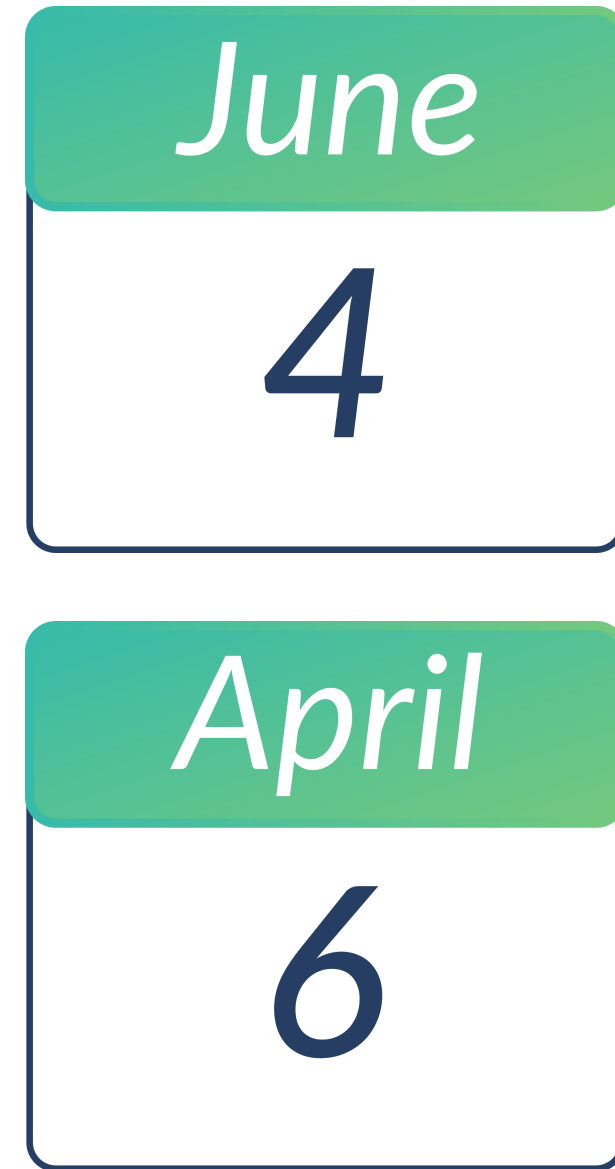
Order of time elements

Order of day, month, and year:

- U.S.: 12/20/2022
- U.K.: 20/12/2022
- Ambiguous most of the year:
 - 6/4/2010 : June 4th or April 6th?

"Time elements":

- Day, Month, Year
- Hour, Minute, Second



ISO 8601

- Time elements arranged largest-to-smallest
 - Year -> month -> day -> ...
 - 2022-06-04
 - 2022-06-04 = June 4th, 2022
- Time elements separated by specific characters
 - Hyphens (-) between date elements
 - 2022-06-04
- Ensures legibility and clarity
 - 2022-06-04 vs. 20220604

Formatting dates and times

```
lubridate::parse_date_time()
```

```
earthday <- "April 22, 2022"

parse_date_time(earthday,
               orders = "%B %d, %Y")
```

```
[1] "2022-08-22 UTC"
```

`?strptime` to view conversion specifications

Common conversion specifications:

Time Element	Conversion Spec.
Year (<code>YYYY</code>)	<code>%Y</code>
Year (<code>yy</code>)	<code>%y</code>
Day (<code>dd</code>)	<code>%d</code>
Month (<code>mm</code>)	<code>%m</code>
Month (August)	<code>%B</code>
Month (Aug)	<code>%b</code>

Parsing multiple date formats

```
dates_vector <- c("12/20/2022",  
                  "2022-12-21",  
                  "December 22, 2022")  
  
dates_vector
```

```
[1] "12/20/2022"  
[2] "2022-12-21"  
[3] "December 22, 2022"
```

```
library(lubridate)  
parse_date_time(  
  dates_vector,  
  orders = c("%m/%d/%Y",  
             "%Y-%m-%d",  
             "%B %d, %Y"))
```

```
[1] "2022-12-20 UTC"  
[2] "2022-12-21 UTC"  
[3] "2022-12-22 UTC"
```

Let's practice

MANIPULATING TIME SERIES DATA IN R