Capstone Exercise

# Capstone Exercise

The final exercise for this course involves performing a time series analysis on real-world data: Carbon Dioxide concentration at the Mauna Loa Observatory, from early 1959 to Present. You’ll go through the process of importing the data, converting to a time series object (with zoo), imputing missing values, and plotting the resulting data. Additionally, you will create an aggregate of the data, as well as a rolling window average of the data.

## Importing the Data

# The following libraries are included for you  
  
library(zoo)  
# Sample data from the Mauna Loa Observatory  
# https://gml.noaa.gov/webdata/ccgg/trends/co2/co2\_mm\_mlo.csv  
  
# Data is already pre-processed as a `zoo` object. It contains missing values,  
# so we'll need to impute those!  
  
# This will be hidden from the users, of course.  
missing\_co2 <- readRDS("data/missing.Rds")

## Visualizing the Data

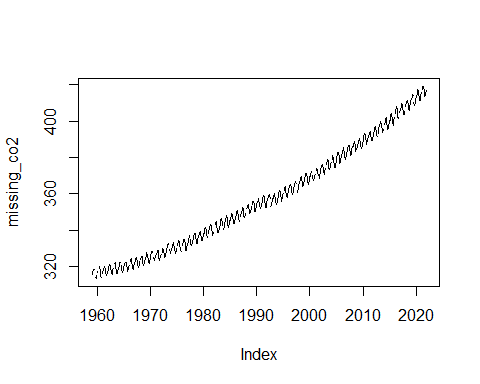
Perform basic data exploration by:

1. Printing the first 20 observations with head(), and,
2. Creating a plot of the data with plot()

head(missing\_co2, n = 20)

## Jan 1959 Feb 1959 Mar 1959 Apr 1959 May 1959 Jun 1959 Jul 1959 Aug 1959   
## 315.58 316.48 316.65 317.72 318.29 318.15 NA NA   
## Sep 1959 Oct 1959 Nov 1959 Dec 1959 Jan 1960 Feb 1960 Mar 1960 Apr 1960   
## 313.84 313.33 314.81 NA 316.43 316.98 NA NA   
## May 1960 Jun 1960 Jul 1960 Aug 1960   
## 320.04 319.59 318.18 315.90

plot(missing\_co2)



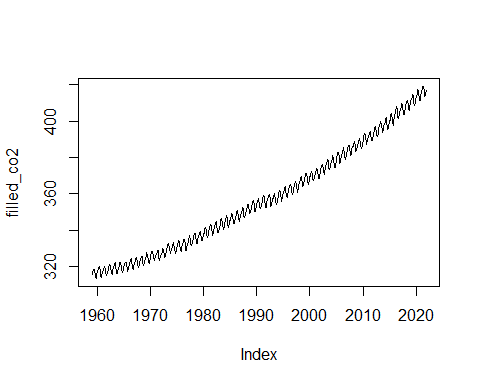
Notice that there are “holes” in the data? this suggests that we’ve got NA values, which is apparent when we view the first few observations with head().

## Imputing the Missing Values

Impute the missing values with a *cubic spline* interpolation, then plot the results

filled\_co2 <- na.spline(missing\_co2)

plot(filled\_co2)

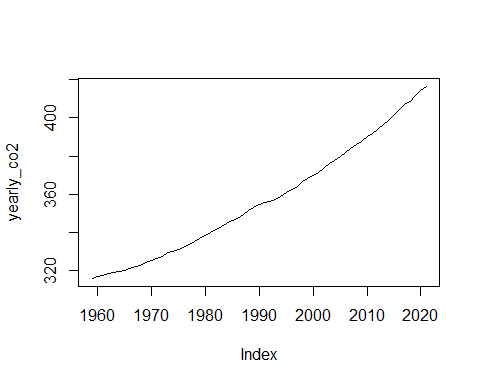


## Yearly Aggregate

Using aggregate(), create a yearly mean of the data, then plot the data

\*Hint: use lubridate::year()

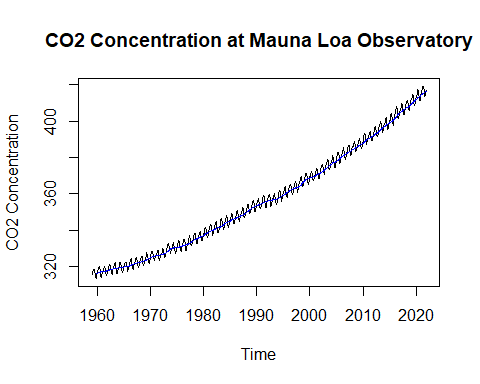
yearly\_co2 <- aggregate(filled\_co2,  
 by = lubridate::year,  
 FUN = mean  
)  
  
  
plot(yearly\_co2)



## Rolling Window

Calculate a 12-month rolling window average of the data, then overlay the results (in blue) on top of the original data. Label the x-axis as “Time” and the y-axis as “CO2 Concentration”, and give your graph the title “Monthly CO2 Concentration at Mauna Loa Observatory”

roll <- rollapplyr(filled\_co2,  
 FUN = mean,  
 width = 12  
)  
  
plot(filled\_co2,  
 xlab = "Time",  
 ylab = "CO2 Concentration",  
 main = "CO2 Concentration at Mauna Loa Observatory")  
  
lines(roll, col = "blue")



## Given Code

The following code is given to the learners at the beginning of the exercise:

# Question 1: Explore the Data  
\_\_\_(missing\_co2, \_\_\_ = \_\_\_)  
  
\_\_\_(missing\_co2)

# Question 2: Impute Missing Values  
  
filled\_co2 <- \_\_\_(\_\_\_)  
  
\_\_\_(filled\_co2)

# Question 3: Find Yearly Mean Aggregate  
yearly\_co2 = aggregate(\_\_\_,  
 by = \_\_\_,  
 FUN = \_\_\_)  
  
\_\_\_(\_\_\_)

# Question 4: Calculate a Rolling Window  
roll <- rollapplyr(filled\_co2,  
 FUN = \_\_\_,  
 width = 12)  
  
plot(\_\_\_,  
 xlab = \_\_\_,  
 ylab = \_\_\_,  
 main = "CO2 Concentration at Mauna Loa Observatory")  
  
lines(\_\_\_, col = \_\_\_)