

429Project

Vladislav Fedorov, Sasha Matveev

2025-10-23

```
library(dplyr)
library(tidyverse)
library(astsa)
library(TSA)
library(fpp2)
```

All group member's names and sections (UG or GR)

Vladislav Fedorov (GR Section), Sasha Matveev (GR Section)

Data Sources:

Original Emissions Data source: <https://ourworldindata.org/co2-emissions> Compiled Emissions Data: <https://www.kaggle.com/datasets/shreyanshdangi/co-emissions-across-countries-regions-and-sectors> (primary data) Compiled Temperature Data: <https://www.kaggle.com/datasets/giabchnguyn/average-temperature-from-1900-to-2023> (supplementary temperature data)

#A paragraph that describes the nature of the data

The original emissions data comes from Our World in Data and was compiled by Shreyansh Dangi for use on kaggle. The data includes global CO2 and other greenhouse gas emissions across countries, and sectors while also combining variables describing the countries in the dataset such as population, GDP, and energy consumption. Each row is a country and a given year so the time unit is yearly. For our analysis we will be focusing on the United States from (1900-2023). The variables of interest will include time (year), the population, GDP (inflation adjusted and cost-of-living adjusted in dollars), CO2 emissions (million tons), methane emissions (tons), nitrous oxide emissions (tons), as well as the average annual temperature in Fahrenheit from the supplementary dataset. The supplementary dataset containing the average annual temperature in the US comes from the National Centers for Environmental Information (NCEI) and was made available on kaggle by Gia Bách Nguyễn. The data is based on observations from a network of thousands of weather stations across the United States in the same exact time period (1900-2023) as the primary dataset containing emissions data.

What questions we want to address

Through our analysis we would like to address the following questions:

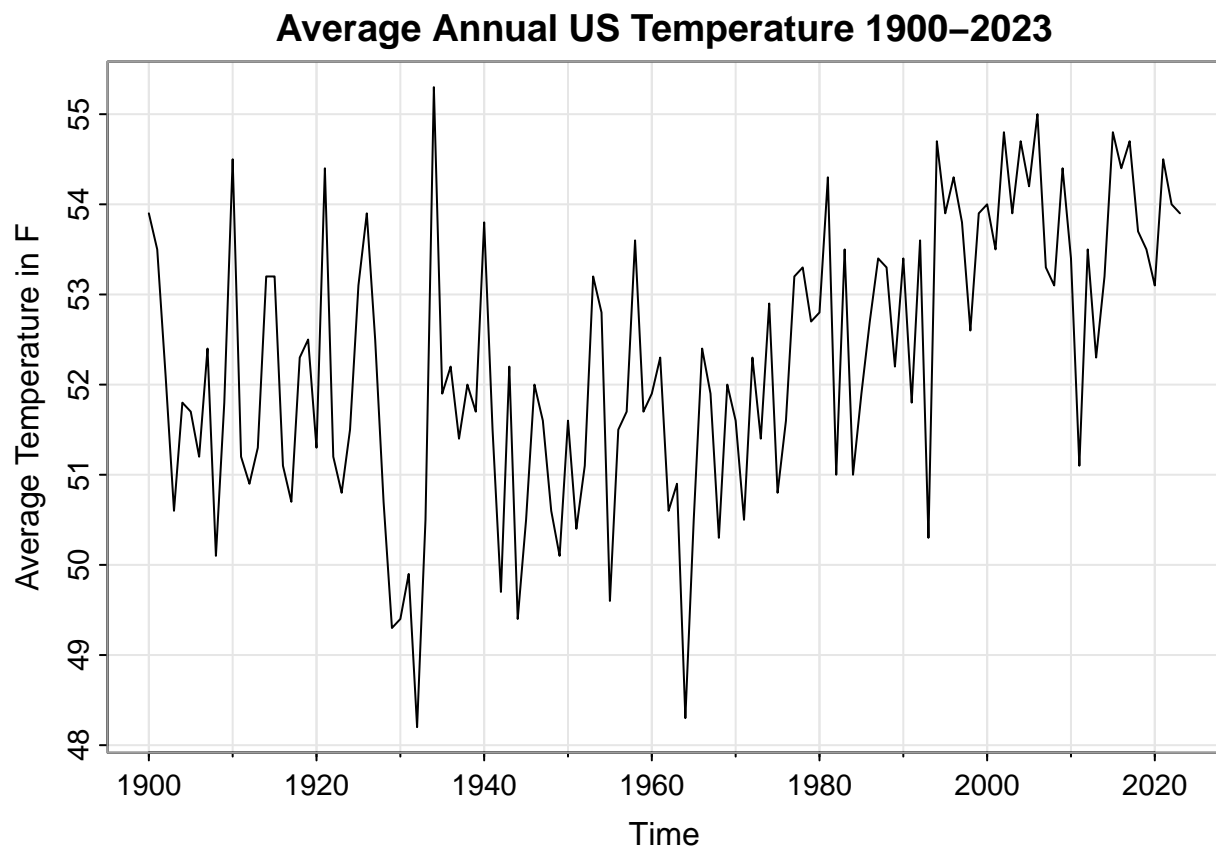
- What factors influence the average temperature in the United States the most?

- How do these factors affect the change in average temperature in the United States year to year (over time)?
- What variables are the strongest predictors of the CO2 emissions for the United States?

Time series plot and any other visual presentations

```
co2 <- read_csv("data/Data.csv", show_col_types = FALSE)
temp <- read_csv("data/Temperature.csv", show_col_types = FALSE)
co2 <- co2|>
  filter(Name == "United States" & year >= 1900)
```

```
temp_ts <- ts(temp[2], start = 1900, frequency = 1)
tsplot(temp_ts, ylab = "Average Temperature in F", main = "Average Annual US Temperature 1900-2023")
```



Plans for analysis (choose analysis A or B)

Analysis A:

We will first fit a model to predict the US average temperature using time, population, GDP, CO2, nitro

GR section students should present basic ideas about analysis C

We will first try to work with our current data and use either Long Memory ARMA and Fractional Differencing (chapter 5.1) or the Lagged Regression methods (mentioned in chapter 5.5), since these two methods seem to be the most fitting considering our data's nature. However, if there aren't any new significant findings or the methods do not work, we will find a dataset which involves clearly periodic observations and work on studying its periodicities using spectral analysis.