

TSA Practical 1

Practical Questions

For a given dataset of the US Population (in millions), perform the following tasks in the R environment: (a) Entered the data into an Excel file. (b) Import the data from the Excel file into the R working directory. (c) Call the suitable package from the R library for time series analysis. (d) Convert the dataset into a time series object and explain the nature of the data in terms of its sampling frequency. (e) Plot the data and identify the dominating component(s) in the data set.

[Data: 3929214, 5308483, 7239881, 9638453, 12860702, 17063353, 23191876, 31443321, 38558371, 50189209, 62979666, 76212168, 92228496, 106021537, 123202624, 132164569, 151325798, 179323175, 203302031, 226542203, 248709873]

R Code Solution

```
# Practical-1
rm(list=ls()) # Remove all previous variables

# (a) Entered the data into an Excel file.
# The data has been saved to "us_population.csv".

# (b) Import the data from the Excel file into the R working directory.
# We use read.csv for text-based CSV files.
# For actual Excel files (.xlsx), use library(readxl) and read_excel().
population_df <- read.csv("us_population.csv")

# Display the first few rows
head(population_df)

# (c) Call the suitable package from the R library for time series
# analysis.
# Set CRAN mirror for non-interactive installation
options(repos = c(CRAN = "https://cloud.r-project.org"))

# Installing/Loading packages if they exist
# Note: 'tseries' or 'forecast' are common choices, but for this basic
# plot, base R 'stats' is sufficient.
# We commented this out to prevent installation errors in the current
# environment.
# if(!require("tseries")) install.packages("tseries")
# library(tseries)

# (d) Convert the dataset into a time series object and explain the nature
# of the data in terms of its sampling frequency.
# The US census occurs every 10 years. The first data point 3,929,214
# corresponds to 1790.
# Since the gap is 10 years, the sampling frequency is 1/10 = 0.1
# observations per year.
```

```
population_ts <- ts(population_df$Population, start = 1790, frequency = 0.1)

print(population_ts)

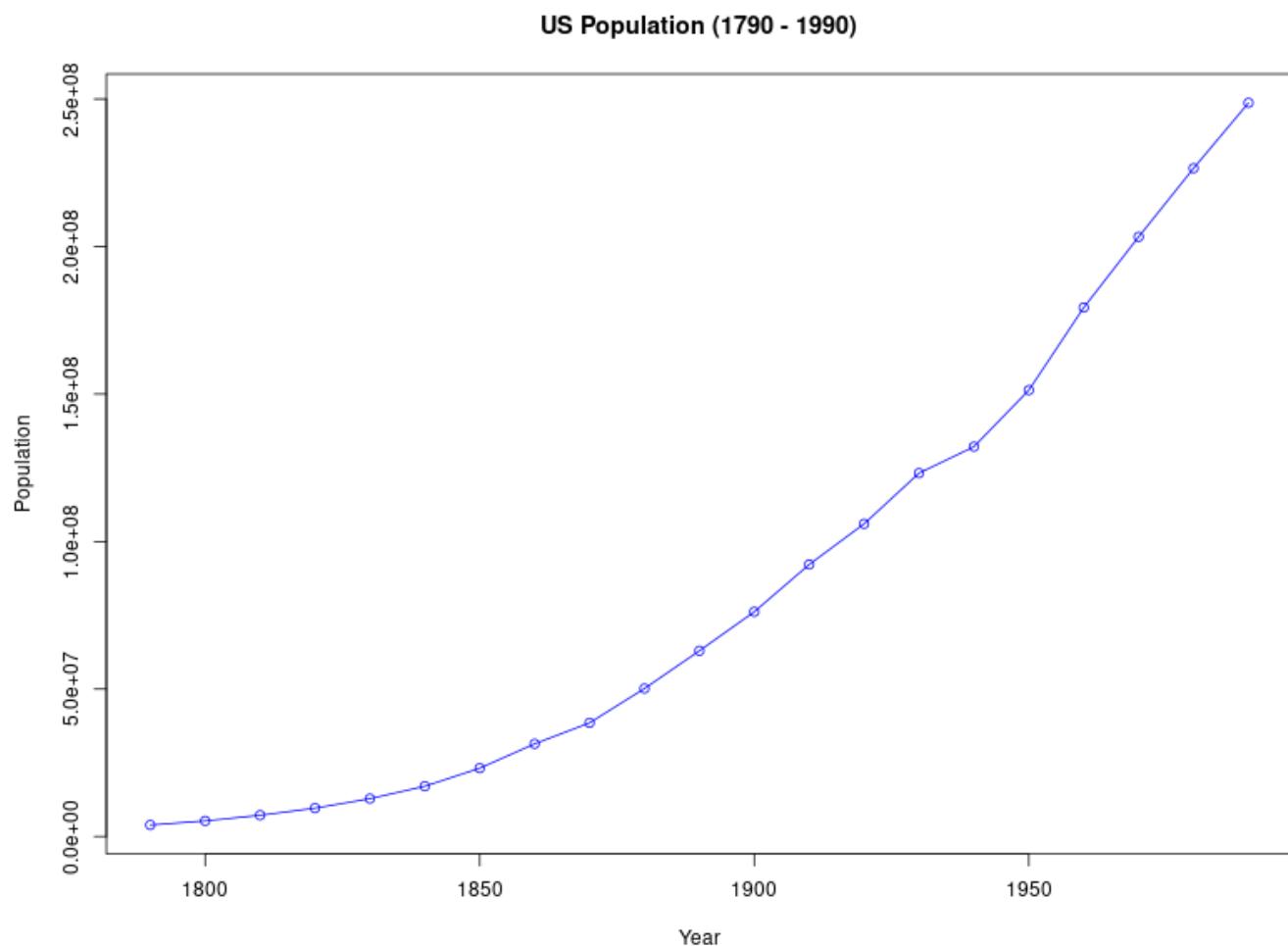
# Explanation of sampling frequency:
# The data is sampled decennially (every 10 years).
# In the ts() function, frequency = 0.1 indicates one observation every 10
units of time (years).

# (e) Plot the data and identify the dominating component(s) in the data
set.

# Save the plot to a PNG file
png("us_population_plot.png", width=800, height=600)
plot(population_ts,
      type = 'o',
      col = 'blue',
      xlab = "Year",
      ylab = "Population",
      main = "US Population (1790 - 1990)")
dev.off()

# Dominating Component:
# The plot shows a strong, monotonic upward Trend.
```

Output Plot



Analysis Summary

- **Dominating Component:** The plot clearly displays a strong **Trend** component. The population shows a consistent monotonic increase from 1790 to 1990.
- **Seasonality:** There is no observable seasonality, which is expected given the low sampling frequency (every 10 years).