ST566D Lab Activity 1

Objective

- Explore time series data in R.
- Create time plot for time series data.

Data of interest

We will use the GNP data set as an example. This data contains the annual gross national product (GNP) deflator of United States from Year 1889 to 1970. The original data can be found here.

Dowload "GNP.csv" file from Canvas. Make sure the file is saved under your R working directory. Read the data into R:

```
GNP <- read.csv("GNP.csv", head = T)
```

```
head(GNP,n=10)
```

```
##    Year  GNP
## 1  1889 25.9
## 2  1890 25.4
## 3  1891 24.9
## 4  1892 24.0
## 5  1893 24.5
## 6  1894 23.0
## 7  1895 22.7
## 8  1896 22.1
## 9  1897 22.2
## 10 1898 22.9
```

Explore GNP data

In time series analysis, it is more convenient to work with a time series data object. In R, ts() function can be used to create time-series objects. We first transfrom the GNP data into a time series object:

```
GNP_ts <- ts(GNP[,2], start = 1889, end = 1970, frequency = 1)
```

Here **start** specifies the time of the first observation, **end** specifies the time of the last observation, and **frequency** can be used to specify the number of observations per unit of time. For example, specify **frequency=12** when the data are sampled monthly and the natural time period is a year; and **frequency=4** for quarterly data and the time period is a year. The command ts(data, frequency = 4, start = c(1959, 2)) specfies a quarterly data with time starting at 2nd Quarter of 1959.

We can also check whether an object is a time series object via function *is.ts()*:

```
is.ts(GNP_ts)
```

```
## [1] TRUE
```

We can display the data using print(). It gives the Start, End, Frequency of the data along with the observations.

```
print(GNP_ts)
```

```
## Time Series:
## Start = 1889
## End = 1970
## Frequency = 1
##  [1]  25.9  25.4  24.9  24.0  24.5  23.0  22.7  22.1  22.2  22.9  23.6
## [12]  24.7  24.5  25.4  25.7  26.0  26.5  27.2  28.3  28.1  29.1  29.9
## [23]  29.7  30.9  31.1  31.4  32.5  36.5  45.0  52.6  53.8  61.3  52.2
## [34]  49.5  50.7  50.1  51.0  51.2  50.0  50.4  50.6  49.3  44.8  40.2
## [45]  39.3  42.2  42.6  42.7  44.5  43.9  43.2  43.9  47.2  53.0  56.8
## [56]  58.2  59.7  66.7  74.6  79.6  79.1  80.2  85.6  87.5  88.3  89.6
## [67]  90.9  94.0  97.5 100.0 101.6 103.3 104.6 105.8 107.2 108.8 110.9
## [78] 113.9 117.6 122.3 128.2 135.3
```

We can also check the number of observations in the data set:

```
length(GNP_ts)
```

```
## [1] 82
```

Sometimes, the time series is too long. The functions *head(), tail()* will only print the first or last few items of the series. Here n is the number of items to display.

```
head(GNP_ts, n = 10)
```

```
##  [1] 25.9 25.4 24.9 24.0 24.5 23.0 22.7 22.1 22.2 22.9
```

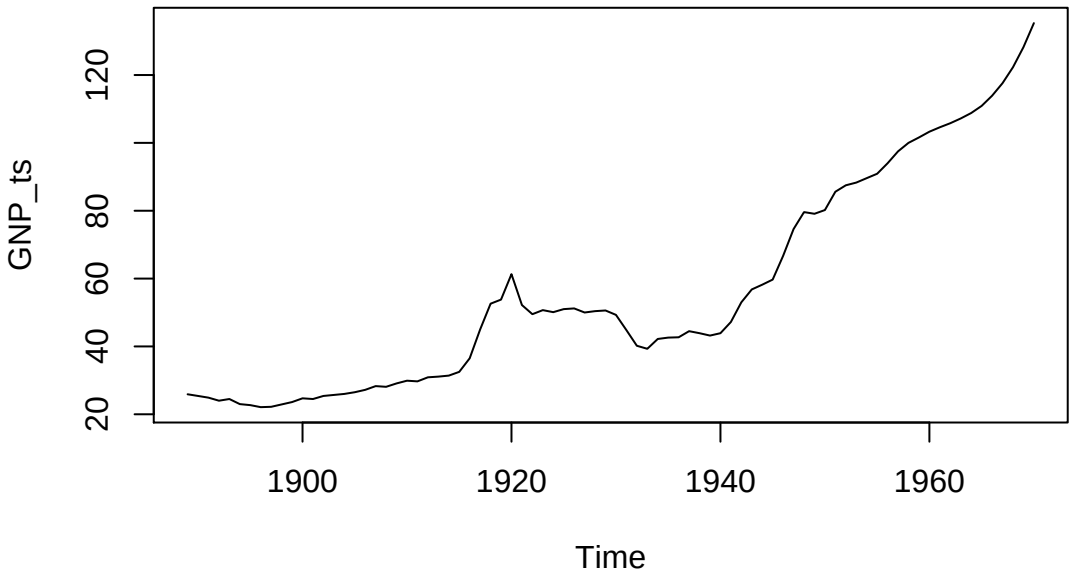
```
tail(GNP_ts, n = 5)
```

```
## [1] 113.9 117.6 122.3 128.2 135.3
```

Time plot

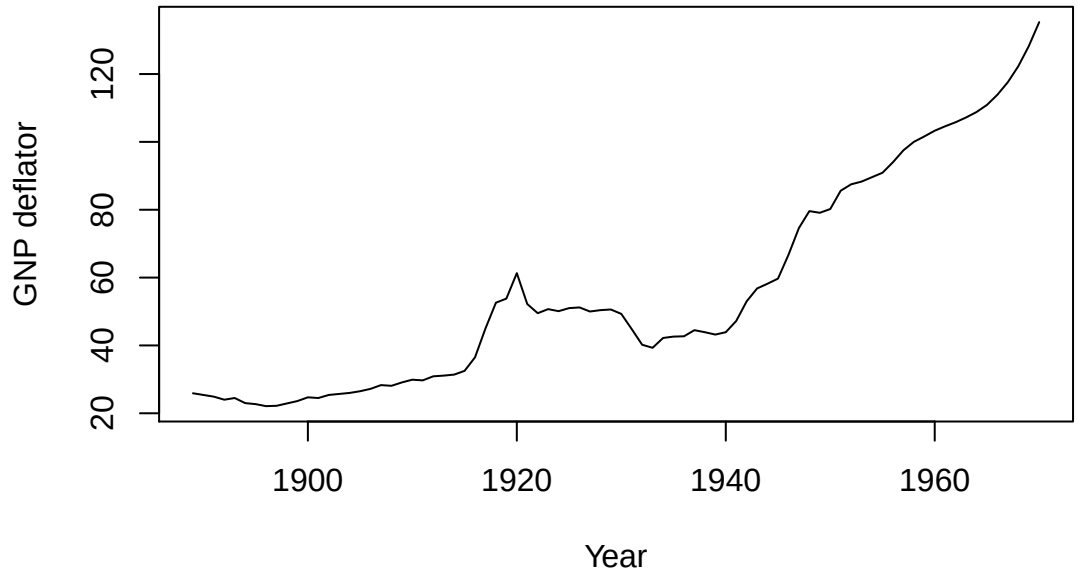
Time plot is a very useful data visualizaton tool in time series analysis. For a time-series object, like "GNP_ts", we can directly use the plot() function in R:

```
plot(GNP_ts)
```



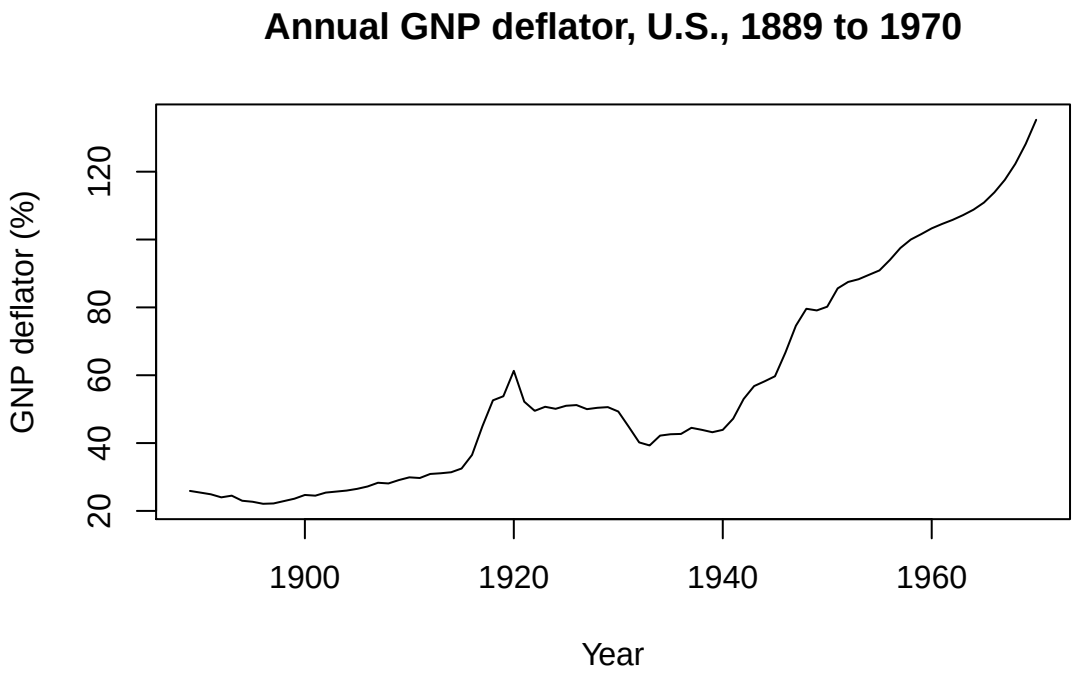
We can also change the labels for this figure:

```
plot(GNP_ts, xlab = "Year", ylab = "GNP deflator")
```



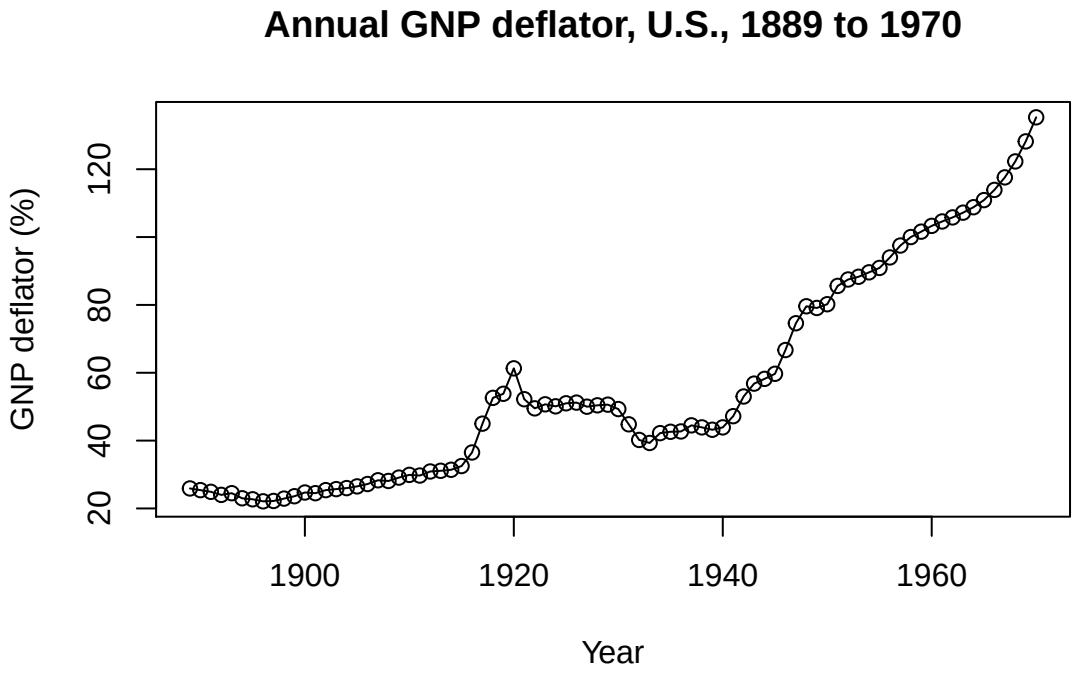
Add a title:

```
plot(GNP_ts, xlab = "Year", ylab = "GNP deflator (%)",
     main = "Annual GNP deflator, U.S., 1889 to 1970")
```



Add circles to hightlight data points:

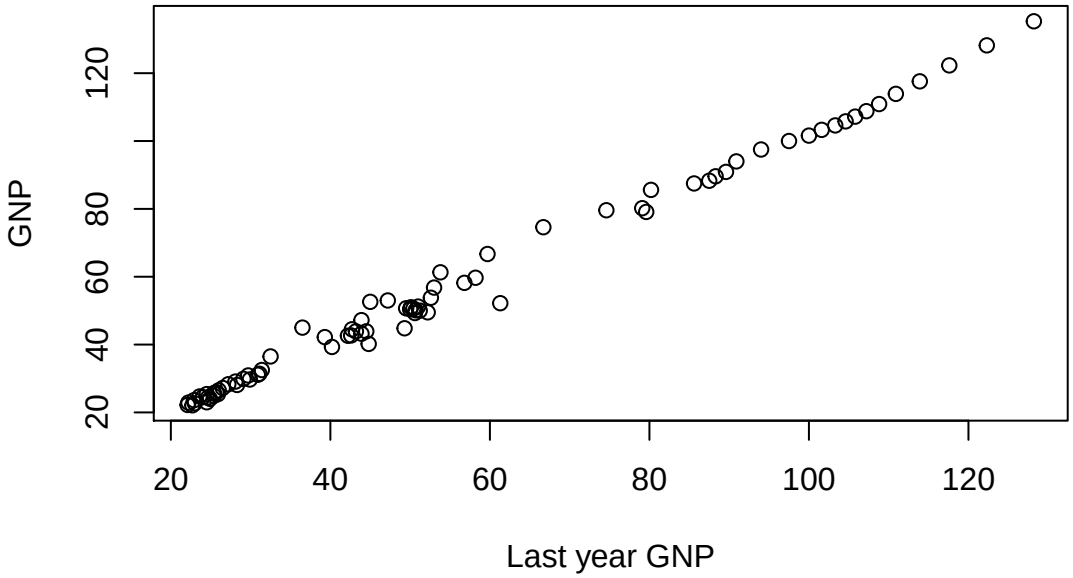
```
plot(GNP_ts, xlab = "Year", ylab = "GNP deflator (%)",
     main = "Annual GNP deflator, U.S., 1889 to 1970",type='o')
```



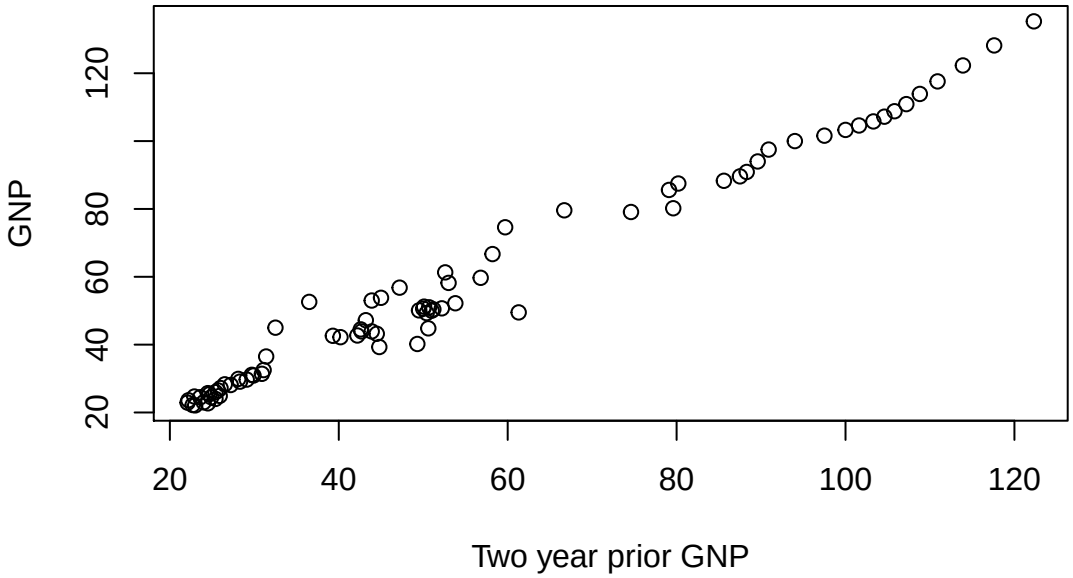
Understand correlation in time series

A main feature of time series data is that the present value of a series is often correlated with its past (or lagged) values. To explore the correlation in time series, we plot the original series against its lagged values. In R, zlag() in package("TSA") can compute the lagged values of a time series.

```r
# install.packages("TSA")
library("TSA")
lag1<-zlag(GNP_ts,d=1)
plot(lag1,GNP_ts, xlab="Last year GNP", ylab="GNP")
```




```r
lag2<-zlag(GNP_ts,d=2)
plot(lag2,GNP_ts, xlab="Two year prior GNP", ylab="GNP")
```



Straight lines in both plots suggest that there exist strong correlations betwen current year GNP and GNP from previous years.

Plot time series using ggplot (Optional)

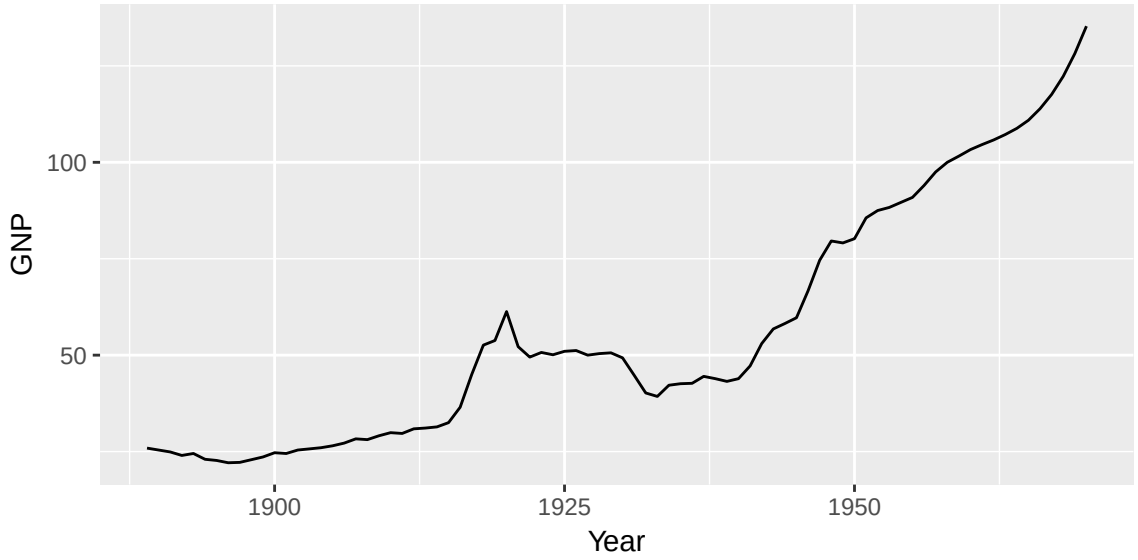
Note that ggplot can also be used to create time series graphs. First you need to load a package "ggplot2". (You need to install the packages. If you haven't done so, run the command *install.packages("ggplot2")* to

install it.)

```
# install.packages("ggplot2")
library(ggplot2)
```

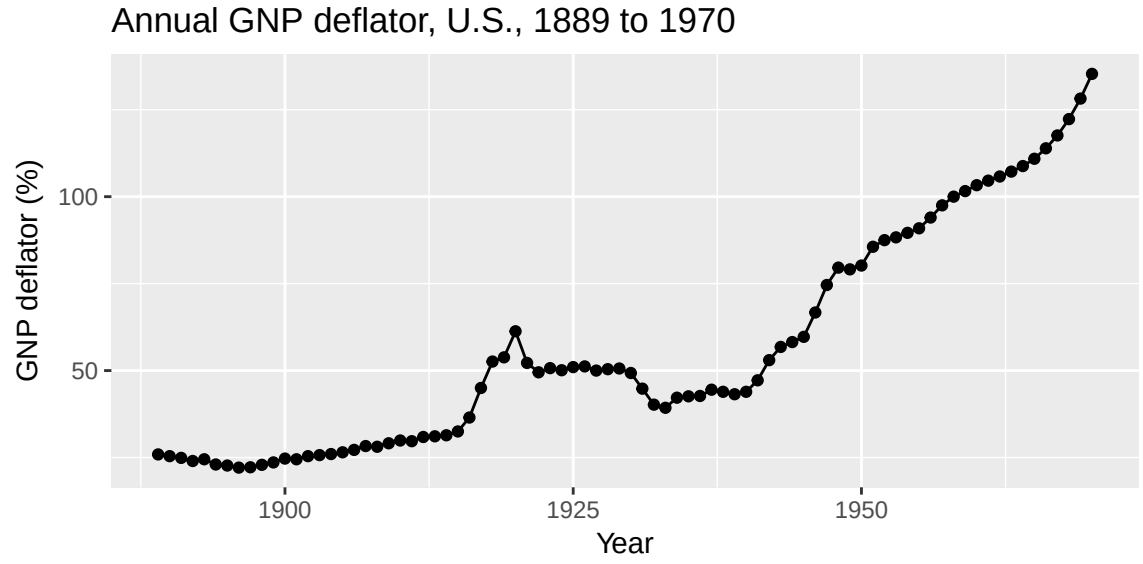
Note that ggplot() is best at handling dataframes. We don't need to transform the original data into a time series onject for ggplot. So in the following, we will use "GNP" that we have read from the csv file.

```
ggplot(GNP, aes(Year, GNP)) + geom_line()
```



Also we can add labels and title to the above figure:

```
ggplot(GNP, aes(Year, GNP)) + geom_line() + geom_point() + xlab("Year") +
  ylab("GNP deflator (%)") + ggtitle("Annual GNP deflator, U.S., 1889 to 1970")
```



Summary

In this lab, we have introduced how to

1. transform a raw data into a time series object
2. obtain the length and a subset of the time series
3. plot the time series with customized labels and title
4. create lagged values of a time series
5. use ggplot for time plot (Optional)

The fucntions we have coverd include:

- read.csv(file, head =)
- ts(data = , start = , end = , frequency =)
- length()
- head(, n =)
- tail(, n =)
- zlag()
- plot(, xlab = "", ylab = "", title = "")

```
knitr::opts_chunk$set(echo = T, warning = F, message = F,
                      fig.align = "center", fig.width = 6, fig.height = 4)
GNP <- read.csv("GNP.csv", head = T)
head(GNP,n=10)

GNP_ts <- ts(GNP[,2], start = 1889, end = 1970, frequency = 1)
is.ts(GNP_ts)
print(GNP_ts)
length(GNP_ts)
head(GNP_ts, n = 10)
tail(GNP_ts, n = 5)
plot(GNP_ts)
plot(GNP_ts, xlab = "Year", ylab = "GNP deflator")
plot(GNP_ts, xlab = "Year", ylab = "GNP deflator (%)",
     main = "Annual GNP deflator, U.S., 1889 to 1970")
plot(GNP_ts, xlab = "Year", ylab = "GNP deflator (%)",
     main = "Annual GNP deflator, U.S., 1889 to 1970",type='o')
# install.packages("TSA")
library("TSA")
lag1<-zlag(GNP_ts,d=1)
plot(lag1,GNP_ts, xlab="Last year GNP", ylab="GNP")
lag2<-zlag(GNP_ts,d=2)
plot(lag2,GNP_ts, xlab="Two year prior GNP", ylab="GNP")
# install.packages("ggplot2")
library(ggplot2)
ggplot(GNP, aes(Year, GNP)) + geom_line()
ggplot(GNP, aes(Year, GNP)) + geom_line() + geom_point() + xlab("Year") +
  ylab("GNP deflator (%)") + ggtitle("Annual GNP deflator, U.S., 1889 to 1970")
```