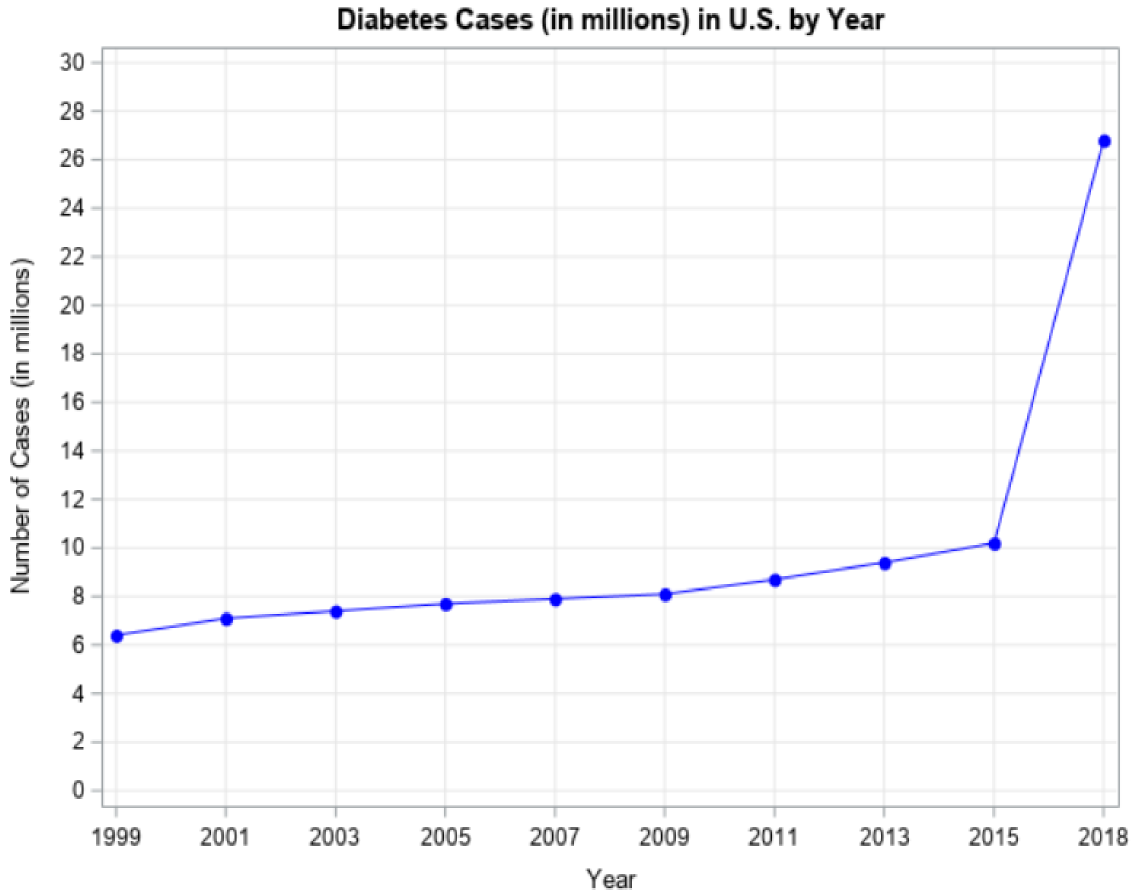
3.3 Time Series Plot

Continuous observations collected longitudinally may be plotted against time producing a **time series plot**. Such plots are described in terms of overall trends: increases, decreases, periodicity, and seasonality.

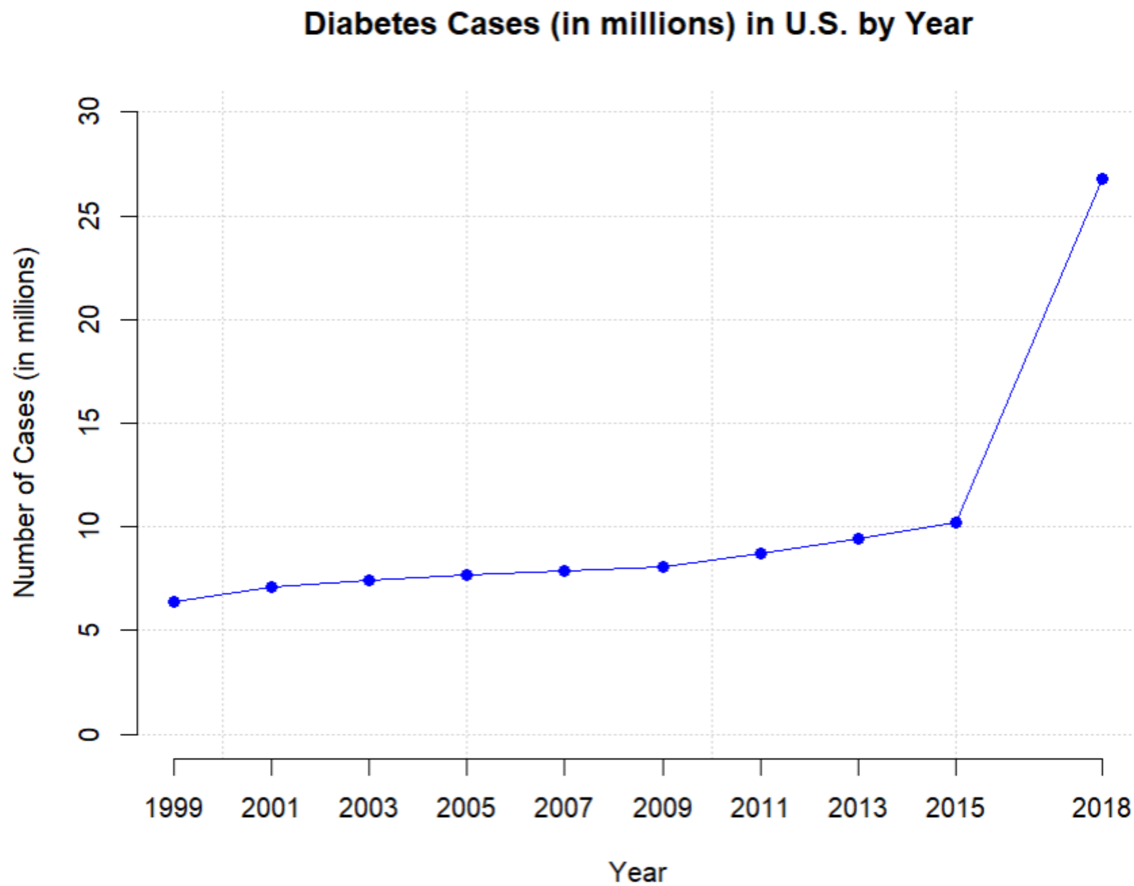
Example. The data set for this example contains the number of diagnosed cases of diabetes (in millions) in U.S. between 1999 and 2018. The codes below produce the time-series plots in SAS and R. In SAS:

```
proc import out=diabetes3 datafile="./diabetes_data3.csv" dbms=csv
replace;
run;

title 'Diabetes Cases (in millions) in U.S. by Year';
proc sgplot data=diabetes3;
series x=year y=diabetes/ legendlabel='diagnosed' markers
markerattrs=(color=blue symbol=circlefilled) lineattrs=(color=blue
pattern=solid);
xaxis label='Year' type=discrete grid;
yaxis label='Number of Cases (in millions)' grid values=(0 to 30 by 1);
run;
```



From the graph, the number of diabetes cases steadily increased between 1999 and 2015, and exhibited a drastic increase (about 2.7 times) between 2015 and 2018. In R:

```
diabetes3.data<- read.csv(file="./diabetes_data3.csv", header=TRUE, sep=",")

plot(diabetes3.data$year, diabetes3.data$diabetes, type="l", col="blue",
main="Diabetes Cases (in millions) in U.S. by Year", xlim=c(1999,2018),
ylim=c(0,30), xlab="Year", ylab="Number of Cases (in millions)", axes=FALSE,
panel.first=grid())
axis(side=1, at=c(seq(1999,2015,by=2), 2018))
axis(side=2)
points(diabetes3.data$year, diabetes3.data$diabetes, pch=16, col="blue")
#pch=plot character 16=dot
```



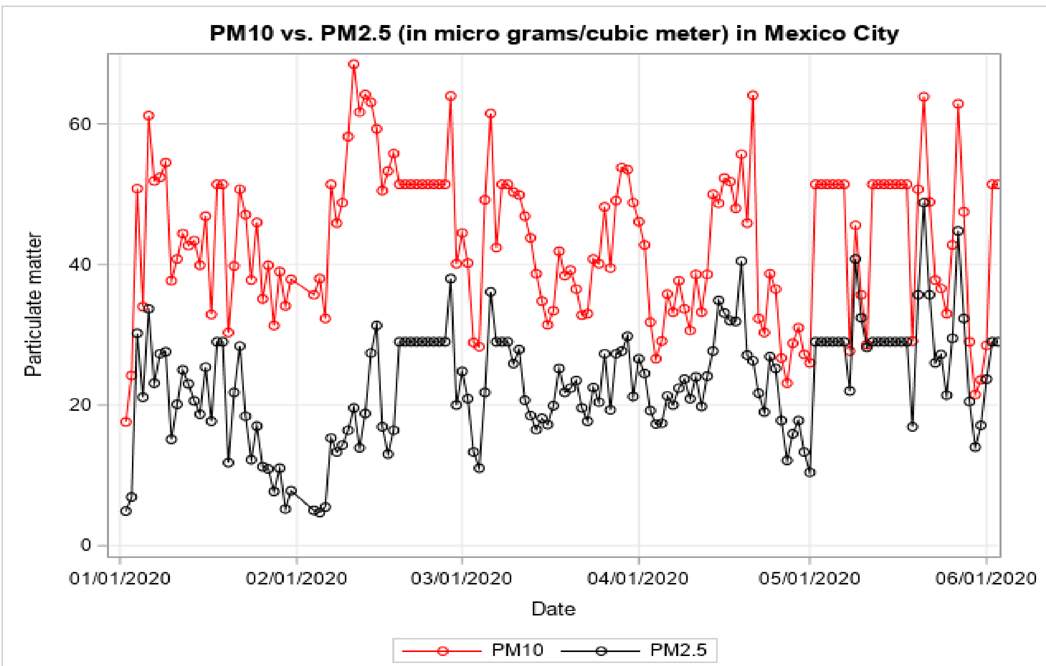
□

Example. In this example we plot particulate matter PM10 vs. PM2.5 (in micro grams per cubic meter) that were recorded in Mexico City between January 2, 2020, and June 3, 2020.

A note: **PM2.5** refers to atmospheric particulate matter (PM) that has a diameter of fewer than 2.5 micrometers. **PM10** refers to atmospheric particulate matter that has a diameter of fewer than 10 micrometers. The difference between the two particulate matters lies in where they reside in the human body. PM2.5 is likelier to travel into and deposit on the surface of the deeper parts of the lung, while PM10 is likelier to deposit on the surfaces of the larger airways of the upper region of the lung. In SAS:

```
proc import out=pollution datafile="./pollution_data.csv" dbms=csv
```

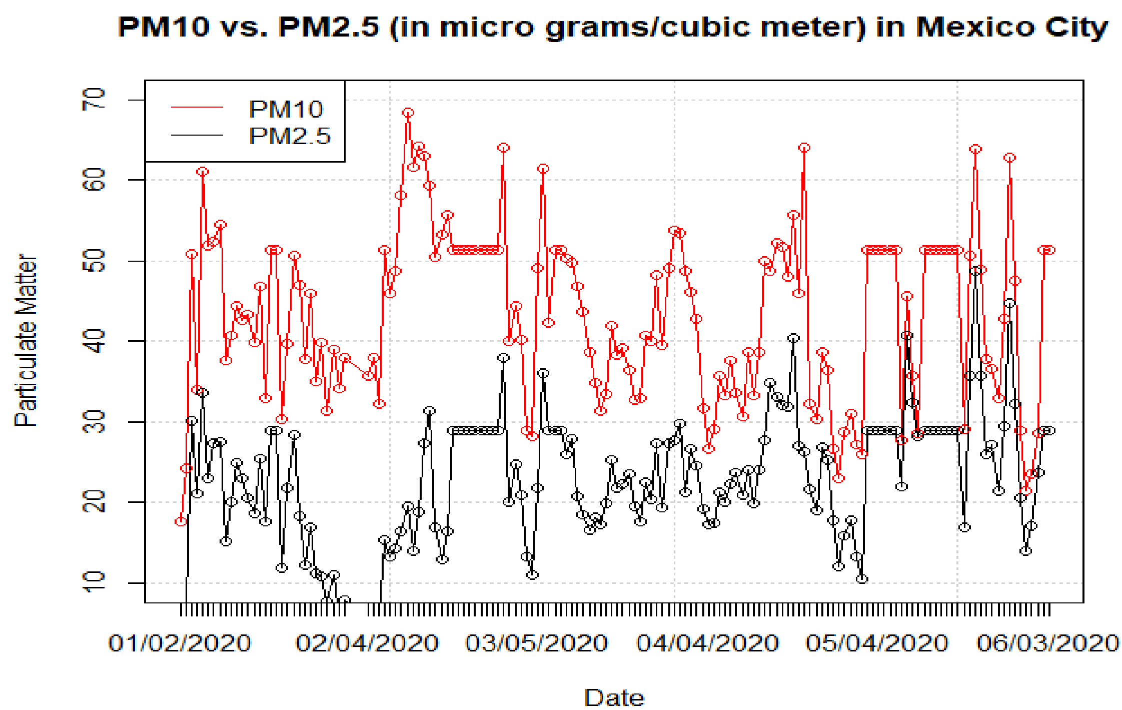
```
replace;
run;

title 'PM10 vs. PM2.5 (in micro grams/cubic meter) in Mexico City ';
proc sgplot data=pollution;
series x=date y=PM10/ legendlabel='PM10' markers
 markerattrs=(color=red symbol=circle) lineattrs=(color=red pattern=solid);
  series x=date y=PM2_5/ legendlabel='PM2.5' markers
    markerattrs=(color=black symbol=circle) lineattrs=(color=black
    pattern=solid);
 xaxis label='Date' values=('1jan20'd to '1jun20'd by month) grid;
   yaxis label='Particulate matter' grid;
run;
```



We can see on the graph that PM10 and PM2.5 display similar fluctuations and that PM2.5 is roughly half the magnitude of PM10. In R:

```
pollution.data<- read.csv(file="./pollution_data.csv", header=TRUE,
sep=",")

plot(as.Date(pollution.data$Date, "%m/%d/%Y"), pollution.data$PM10,
type="l", col="red", main="PM10 vs. PM2.5 (in micro grams/cubic meter)
```

```
in Mexico City", ylim=c(10,70), xlab="Date", ylab="Particulate Matter",
xaxt="n", panel.first=grid())
lines(as.Date(pollution.data$Date, "%m/%d/%Y"), pollution.data$PM2.5,
col="black")
points(as.Date(pollution.data$Date, "%m/%d/%Y"), pollution.data$PM10,
pch=1, col="red")
points(as.Date(pollution.data$Date, "%m/%d/%Y"), pollution.data$PM2.5,
pch=1, col="black")
legend("topleft", c("PM10", "PM2.5"), lty=1, col=c("red", "black"))
axis(1, at=as.Date(pollution.data$Date, "%m/%d/%Y"),
labels=format(as.Date(pollution.data$Date, "%m/%d/%Y"),"%m/%d/%Y"))
```



□