3.1 Graphical Presentation

Frequency, proportion, or percent of diseased individuals are typically presented graphically on a bar graph, histogram, or time-series plot. Below we consider three examples with SAS and R codes and graphically outputs.

Example. The data given in the table below contain percent of diagnosed diabetes cases by race/ethnicity and gender among U.S. adults aged 18 years or older, in 2017–2018.

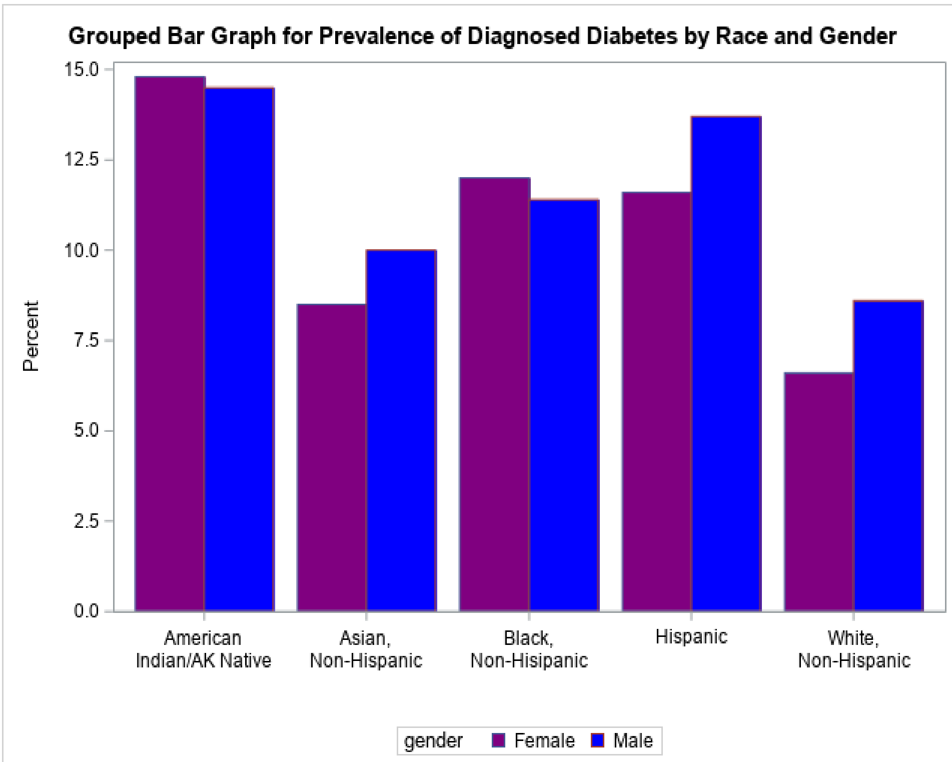
Race/ethnicity	Gender	
	Male	Female
American Indian/Alaska Native	14.5	14.8
Asian, Non-Hispanic	10.0	8.5
Black, Non-Hispanic	11.4	12.0
Hispanic	13.7	11.6
White, Non-Hispanic	8.6	6.6

The following SAS and R codes produce bar graphs for these data.

In SAS:

```
data diabetes1;
input race gender percent @@;
cards;
1 1 14.5 1 0 14.8 2 1 10.0 2 0 8.5  3 1 11.4 3 0 12.0
4 1 13.7 4 0 11.6  5 1 8.6  5 0 6.6
;

proc format;
value racefmt 1='American Indian/AK Native'
              2='Asian, Non-Hispanic'
              3='Black, Non-Hispanic'
              4='Hispanic'
              5='White, Non-Hispanic'
   ;
value genderfmt 1='Male' 0='Female';
run;

title 'Grouped Bar Graph for Prevalence of Diagnosed Diabetes by
Race and Gender';
proc sgplot data=diabetes1; *SG=Statistical Graphics;
```

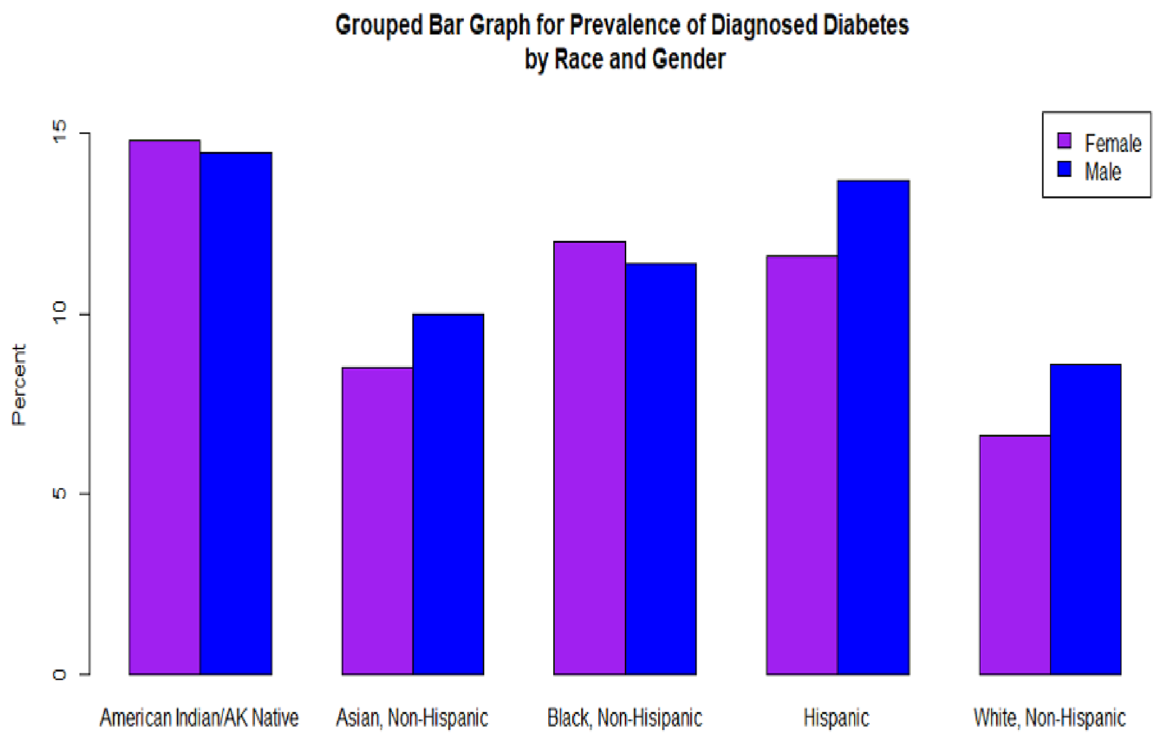
```
  styleattrs datacolors=(purple blue);
   vbar race /response=percent group=gender groupdisplay=cluster;
    format race racefmt. gender genderfmt.;
      xaxis label=' ';
       yaxis label='Percent';
run;
```



From the side-by-side bar graphs, American Indians/Alaskan Natives have the highest percentages diagnosed with diabetes, for both men and women. Whites have the smallest such percentages. Fewer percentage of women than men are diagnosed for Asians, Hispanics, and Whites. For American Indians/Alaskan Natives and Blacks, higher percentage of women have diabetes than men.

In R:
diabetes1<- read.csv(file="./diabetes1.csv", header=TRUE, sep=",")
table <- xtabs(percent ~ gender+race, data=diabetes1)
barplot(table, main="Grouped Bar Graph for Prevalence of Diagnosed Diabetes by Race and Gender", ylim=c(0,16), xlab="", ylab="Percent",

col=c("purple", "blue"), legend.text=rownames(table), beside = TRUE)



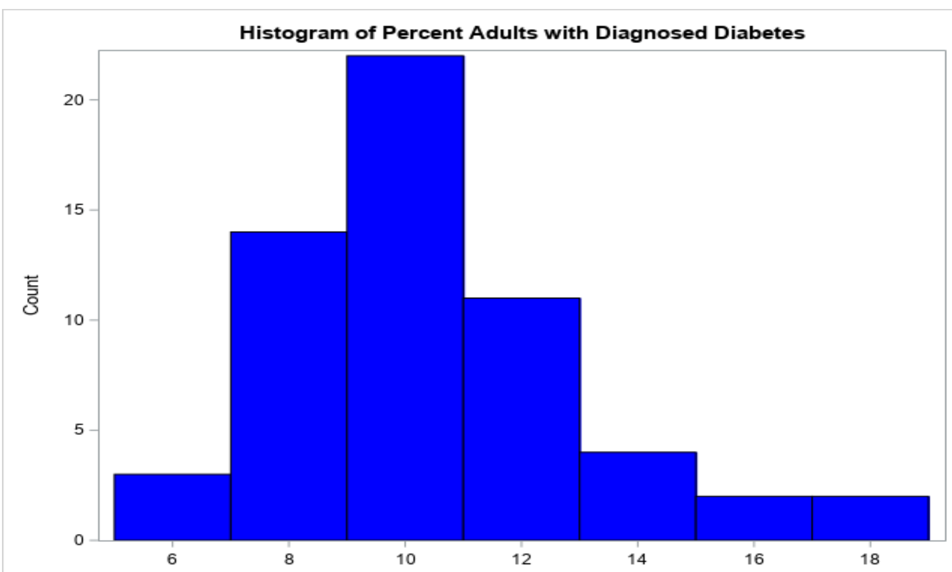
□

Example. In this example, the data represent percentage of adults aged 20 and above with diagnosed diabetes in 2017 by county in California. The codes below plot histograms for the data.

In SAS:

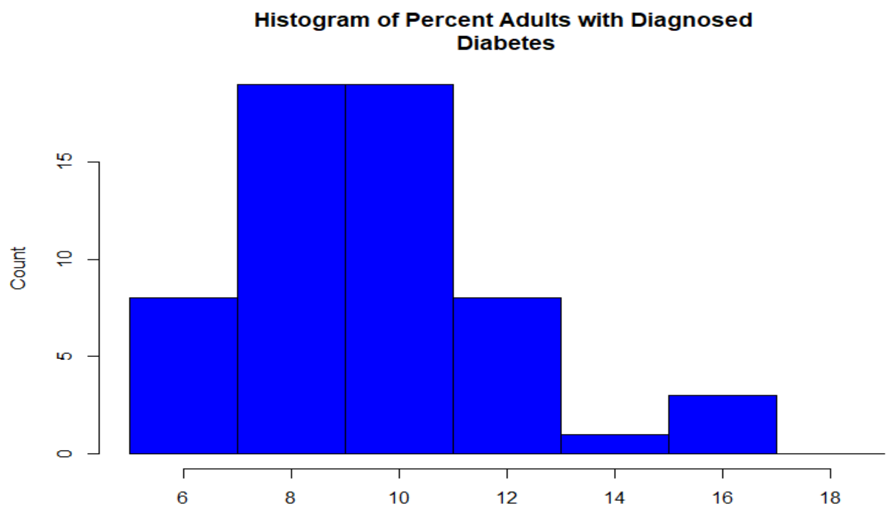
```
data diabetes2;
length county $15;
input county $ percent @@;
cards;
Alameda         7   Alpine          7   Amador          17
Butte           10  Calaveras       11  Colusa          16
Contra_Costa    9   Del_Norte       17  El_Dorado       8
Fresno          10  Glenn           11  Humboldt        10
Imperial        9   Inyo            12  Kern            10
```

```
Kings            13  Lake           11  Lassen          15
Los_Angeles      9   Madera         12  Marin           8
Mariposa         8   Mendocino      11  Merced          13
Modoc            12  Mono           11  Monterey        9
Napa             10  Nevada         6   Orange          9
Placer           7   Plumas         7   Riverside       10
Sacramento       9   San_Benito     10  San_Bernardino  10
San_Diego        8   San_Francisco  6   San_Joaquin     10
San_Luis_Obispo  9   San_Mateo      8   Santa_Barbara   8
Santa_Clara      9   Santa_Cruz     6   Shasta          12
Sierra           7   Siskiyou       10  Solano          9
Sonoma           8   Stanislaus     10  Sutter          13
Tehama           11  Trinity        10  Tulare          11
Tuolumne         8   Ventura        9   Yolo            8
Yuba             13
;

title 'Histogram of Percent Adults with Diagnosed Diabetes';
proc sgplot data=diabetes2;
 histogram percent/ scale=count binstart=6 binwidth=2
  showbins fillattrs=(color=blue);
   xaxis display=(nolabel);
run;
```



The histogram is unimodal and right-skewed.

In R:
diabetes2<- read.csv(file="./diabetes2.csv", header=TRUE, sep=",")
hist(diabetes2$percent, main="Histogram of Percent Adults with Diagnosed Diabetes", breaks=seq(6,19,by=2), col="blue", xlab="", ylab="Count")



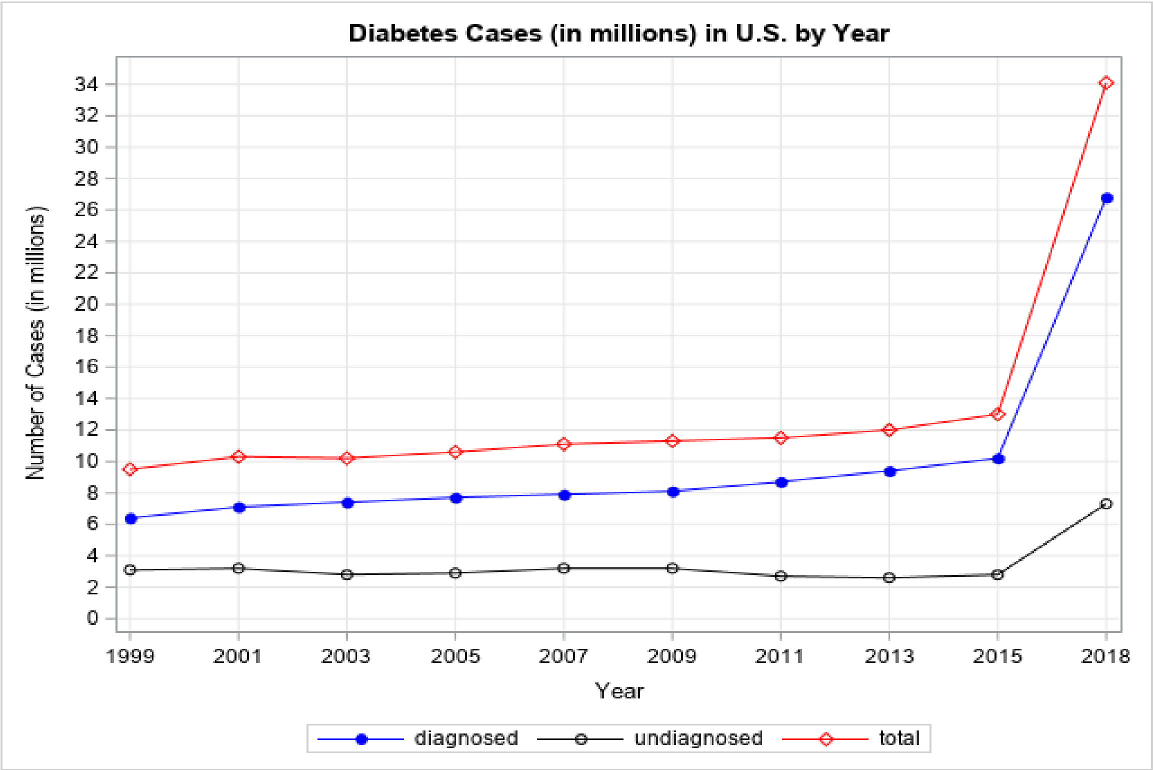
□

Example. The data set for this example contains the number of diagnosed, undiagnosed, and total 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:

```
data diabetes3;
input year diagnosed undiagnosed total @@;
cards;
1999 6.4  3.1 9.5   2001 7.1 3.2 10.3  2003 7.4  2.8 10.2
2005 7.7  2.9 10.6  2007 7.9 3.2 11.1  2009 8.1  3.2 11.3
2011 8.7  2.7 11.5  2013 9.4 2.6 12.0  2015 10.2 2.8 13.0
2018 26.8 7.3 34.1
;

title 'Diabetes Cases (in millions) in U.S. by Year';
```

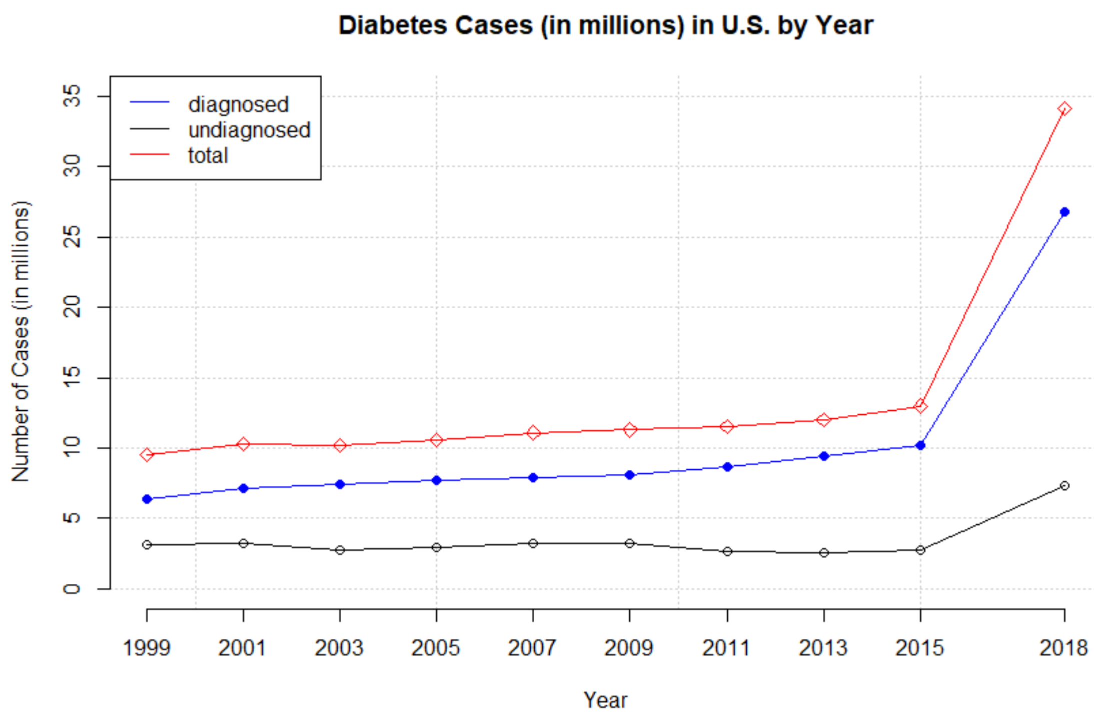
```
proc sgplot data=diabetes3;
series x=year y=diagnosed/ legendlabel='diagnosed' markers
markerattrs=(color=blue symbol=circlefilled) lineattrs=(color=blue pattern=solid);
 series x=year y=undiagnosed/ legendlabel='undiagnosed' markers
  markerattrs=(color=black symbol=circle) lineattrs=(color=black pattern=solid);
   series x=year y=total/ legendlabel='total' markers
    markerattrs=(color=red symbol=diamond) lineattrs=(color=red pattern=solid);
 xaxis label='Year' type=discrete grid;
   yaxis label='Number of Cases (in millions)' grid values=(0 to 35 by 1);
run;
```



In R:
diabetes3.data<- read.csv(file="./diabetes3.csv", header=TRUE, sep=",")

lot(diabetes3.data$year, diabetes3.data$diagnosed, type="l", col="blue", main="Diabetes Cases (in millions) in U.S. by Year", xlim=c(1999,2018), ylim=c(0,35), 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)

lines(diabetes3.data$year, diabetes3.data$undiagnosed, col="black")
lines(diabetes3.data$year, diabetes3.data$total, col="red")

points(diabetes3.data$year, diabetes3.data$diagnosed, pch=16, col="blue")
#pch=plot character 16=dot
points(diabetes3.data$year, diabetes3.data$undiagnosed, pch=1, col="black")
#1=circle
points(diabetes3.data$year, diabetes3.data$total, pch=5, col="red")
#5=diamond
legend("topleft", c("diagnosed", "undiagnosed", "total"), lty=1,
col=c("blue", "black", "red"))
```

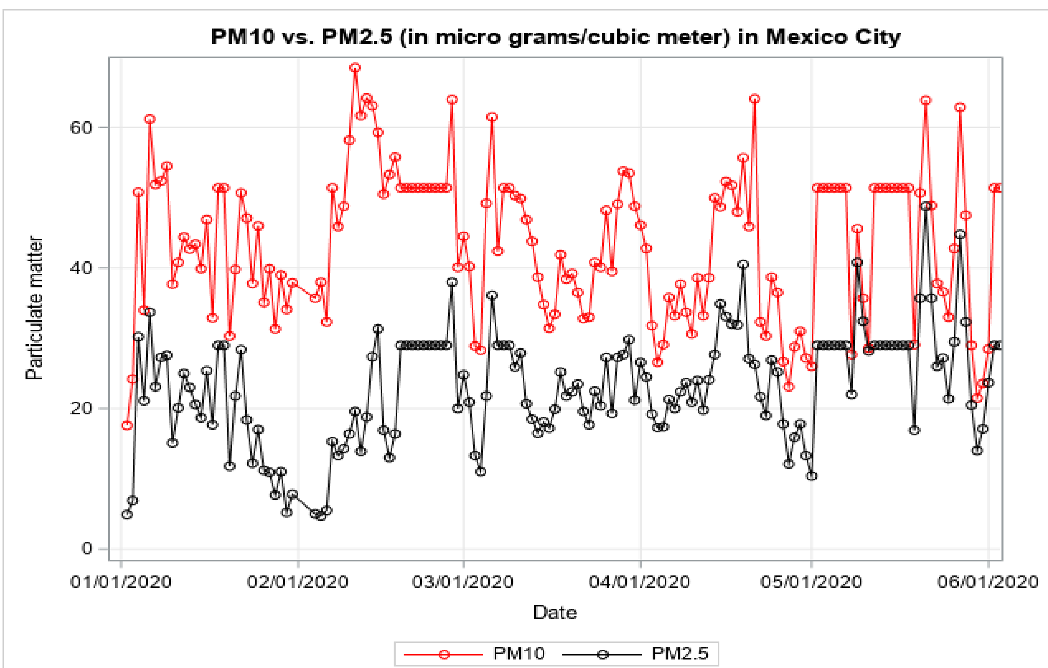


Example. In this example we plot particulate matter PM10 vs. PM2.5 (in micro grams per cubic meter) that were recorded in Mexico City between January 2,2020, and June 3, 2020.

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.

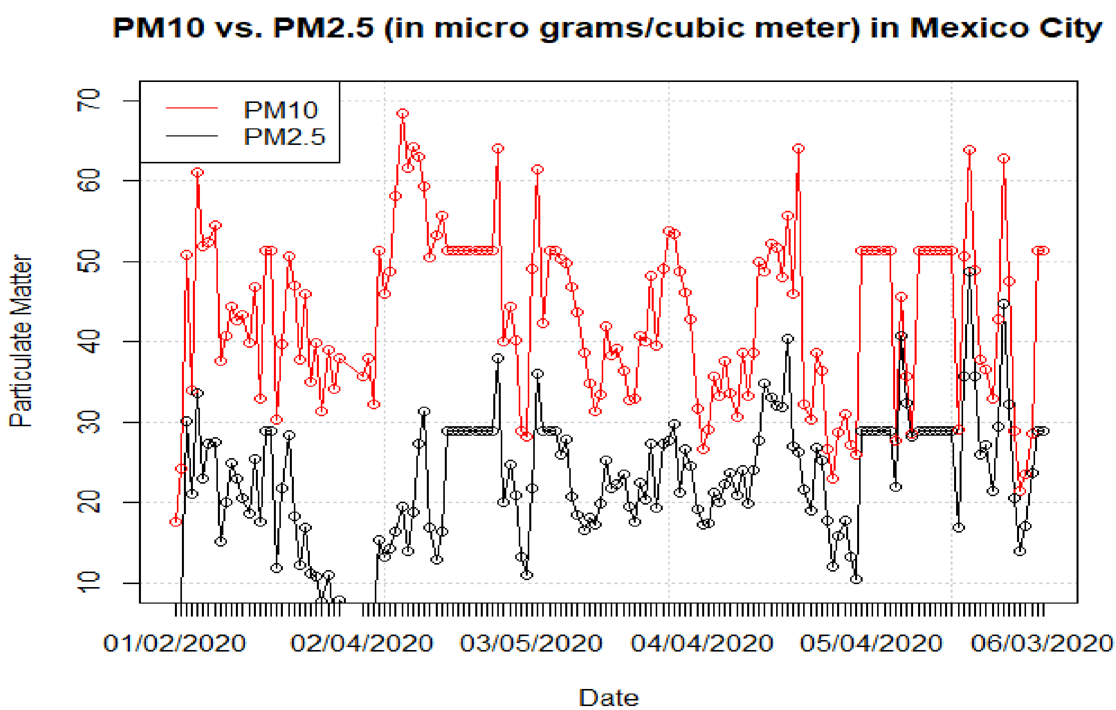
In SAS:

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

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;
```



In R:

```
pollution.data<- read.csv(file="/pollution.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"))
```



□