Plotting Data

- plots in sciences important for interpreting and communicating information in an effective manner
- Specialized plots used in all sciences.
- In the earth sciences, triangular and polar-style plots used to interpret multivariate and spatial data.
- plots express relationships between data, the distribution of a single parameter, and organizational structure.

Univariate Plots

- histograms, stem and leaf diagrams, dot plots, and box and whisker plots
- In some cases the information one wants to communicate are average values and the spread of the data.
- could just provide a mean and a standard deviation, there are several potential problems with this.
  - assumed data distribution
  - no information about extreme values
  - lack of visual feedback
- To overcome these limitations, extra effort of plotting the data is needed.

Stem and leaf plots, dot plots

- simple graphing methods to evaluate the distribution of data.
- Starting with some basic data: chloride concentrations measured in ground-water samples collected at local landfill:

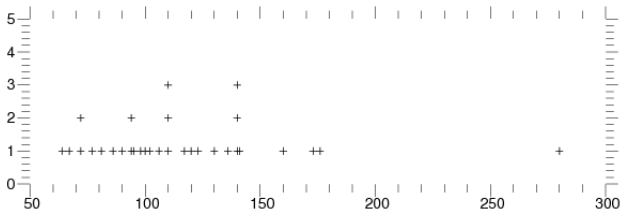
  64, 67, 72, 72, 77, 81, 86, 90, 94, 94, 95, 98, 100, 102, 106, 110, 110, 110, 117, 120, 123, 130, 136, 140, 140, 140, 141, 160, 173, 176, 280
- stem and leaf diagram uses the tens column of the number as bins, and plots the number from the ones column adjacent to it, as shown on the next page.

Stem and leaf plots

6 47 7 227 8 16 9 04458 10 026 11 0007 12 03 13 06 14 0001 15 16 0 17 36 18 . . . 28 0

Dot plots

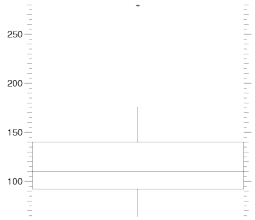
- plotted along a number line, instead of numbers (similar to stem/impulse plot).
- if identical numbers are present in a data set the dots are stacked.



- similar to the familiar histogram, a column or bar graph that displays the frequency of different value ranges.

Box and Whisker Plot

- compact way of displaying data distribution is the box and whisker plot.
- box outlines 50% of the data, from the first to third quartile (interquartile range)
- Whiskers extend to furthest data point within $1.5 \cdot IQR$
- Data outside this range are plotted as individual points



Box and Whisker Plot

- do not contain information about the number of measurements used
- notch is sometimes added to a box plot that conveys information about the sample number
- distance the notch extends above and below the median is:

$$\pm 1.57 \frac{IQR}{\sqrt{n}}$$

- notch provides an initial indication that distributions are similar or different.
- if notches overlap, the two box and whisker plots have similar data distributions

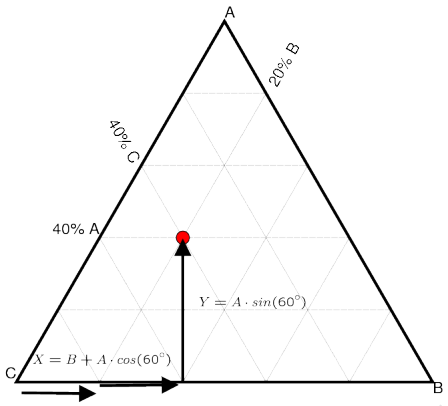
Scatter plots

- difference between scatter or xy plots and line plots
- used throughout geology to visualize relationships between parameters
- convention, the scatter plot is constructed with the independent variable plotted in the x direction and the dependent variable plotted in the y direction.
- In many cases, it is not clear which variable is dependent and independent.
- to simplify relationships, graphs or plotted using logarithmic scales.
- Regardless of the data one has, logarithmic transformations will tend to linearize data and will distort the data, obscuring the true trends in the data.

Multivariate plots

- to plot more than two variables in a two dimensional graphs:
  - use a glyph/glyph size/color to represent third data dimension.
  - triangular plots allow one to plot data with three variables

Trilinear Plots

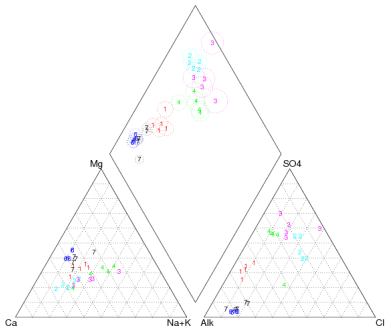


Trilinear Plots

How can we cast this problem into vector addition?

- origin:where two variables are zero
  - lower left corner for A and B
  - lower right corner for A and C
- use this corner as starting point
- vectors extend from this vertex toward two other vertexes
- vector 'A' [$A \cdot cos(60°)$,$A \cdot sin(60°)$]
- vector 'B' [B,0]
- add vectors gives final position in cartesian coordinates

Data from multiple triangular and scatter plots can be projected into a seperate field, creating unique plots.



Directional Data

- vector/quiver plots
- polar graph simply plots the ends of vectors all starting at the center of a plot.
- stereonets project data from the skin of a sphere onto a 2-D plot (circle).
  - Great circle: circle on sphere befined by the intersection of a plane that passes through the center of the sphere and the sphere.
  - Small circle: any other circle on the sphere

- an equal interval stereonet 'unwraps and flattens' the surface of the sphere, maintaining the spacing between lines of equal dip. The distance from the center to a point on the plot is proportional to a dip angle.

$$r = \frac{R(90^\circ - \phi)}{90}$$

- an equal angle (Wulff) stereonet preserves angular relationships, and preserves circles on surface of sphere.

$$r = \frac{Rtan(90^\circ - \phi)}{2}$$

- an equal area (Schmidt) stereonet preserves areas on surface of sphere.

$$r = \frac{\sqrt{2}R\sin(90^\circ - \phi)}{2}$$

Example Script:Read datafile,make plots

```
import pylab as P
from dateutil import parser#handles dates
dataobj=open('./WXDailyHistory.txt','r')#access file (pointer to location)

date=[]
aveTF=[]
PressureMaxIn=[]
PressureMinIn=[]
Precip=[]
for line in dataobj:
    Words=line.split(',')
    if Words[0]=='Date':
        header=Words
    else:
        date.append(parser.parse(Words[0]))
        aveTF.append(float(Words[2]))
        PressureMaxIn.append(float(Words[10]))
        PressureMinIn.append(float(Words[11]))
        Precip.append(float(Words[-1]))

fig=P.figure()
sp1=P.subplot(211)
P.plot(date,aveTF,color='black',linestyle=':',marker='o', linewidth=.2)
sp1.set_ylabel('temp F')
labels = sp1.get_xticklabels()#hide labels (repeated on other subplot)
for label in labels:
    label.set_visible(False)
sp1.set_position([0.1,0.48,.8,.5])#change position and size of plot
sp1_2=P.twinx()#two y axes
sp1_2.plot(date,PressureMaxIn,linewidth=.1)
sp1_2.plot(date,PressureMinIn,linewidth=.1)
sp1_2.set_ylabel('pressure')#label axis

sp2 = P.subplot(212)
sp2.set_position([0.1,0.2,.8,.2])
P.vlines(date,[0],Precip)
labels = sp2.get_xticklabels()#rotate labels 30 degrees
for label in labels:
    label.set_horizontalalignment('right')
    label.set_rotation(30)
P.ylim([3,0])#set range of y axis from 3 to 0 (reversed)
P.ylabel('precip')

P.show()
```

Example Script:Polar Plots

```
from dateutil.parser import parse
import numpy as N
import pylab as P

datafile=open('./WgScreenscrape.csv','r')
date=[]
direction=[]
windspeed=[]
for line in datafile:
      try:
            words=line.split(',')
            date.append(parse(words[0]))
            direction.append(float(words[5]))
            windspeed.append(float(words[6]))
      except ValueError:
            print line
#remove 'bad' data
direction=N.array(direction)
index=N.where(direction>0)##Note numpy command 'where', two forms of command
d2=direction.take(index)[0]
d2=d2*N.pi/180. ##need to convert to radians

### lots of gotcha's with this....when plotting on polar plot
### in radians
### bar plots must be centered
sp=P.subplot(111,polar=True)
bins=[i*2*N.pi/8. for i in range(9)]
a,b=N.histogram(d2,bins=bins) #a is freq, b is list of **edge** values for each bin
width=b[1]-b[0]
b2=[(b[i]+b[i+1])/2 for i in range(len(b)-1)]
#sp.stem(b2,a)
sp.hist(d2,bins=[i*2.*N.pi/8. for i in range(9)])
P.show()
```