## Multivariate visualization

C. Andrews

## Univariate



## Univariate: Histogram



## Univariate: Stem and Leaf

Stem

## Leaf

$$
\begin{array}{r|l}
4 & 12 \\
5 & 7 \\
6 & 34458 \\
7 & 1222566666778999 \\
8 & 0111122339 \\
9 & 1235555678899 \\
10 & 00
\end{array}
$$

## Univariate: Box and Whiskers



## Adding a variable...



## Univariate questions

Which is the biggest?

Which is the smallest?

What is the "center"?

What is the "shape" of the data?
What are the outliers?

## Multivariate questions

Which items are most alike?

Which items are most exceptional?
How can these items be combined into logical groups based on similarity?

## Bivariate



## Trivariate



## Hypervariate



## MultiD Scatterplots


or just add another axis

## Scatterplot matrix

| $\begin{array}{ll}  & 15 \\ \frac{8}{2} & 10 \\ \frac{8}{2} & \\ \text { 卷 } & 5 \end{array}$ |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| \％ 10 <br> \％  <br> 骨 5 <br>   <br>  0 |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | $\int_{+*}^{+*}+{ }^{+}+$ |
| 年 |  |  |  |  |  |  |  |  |  |
|  |  |  | $\stackrel{+t_{i}}{+\underbrace{+}_{i}}$ ${ }^{\text {Chy }}$ edthet |  |  |  |  |  |  |
|  0.2 <br> \＆  <br> en 0.1 <br> 0  <br>  0.0 |  |  |  |  | ${ }_{+}^{+\infty+t^{+\infty}}$ |  |  |  |  |
|  |  |  |  |  |  | $\therefore+{ }_{+}^{2+5}$ |  |  | $\underbrace{t}_{*}$ |
|  |  |  |  |  |  |  |  |  |  |
|  | $\begin{array}{lccc} 0 & 5 & 10 \\ \\ \text { Alpha Acid } \end{array}$ | $\begin{array}{lcc} 0 & 5 \\ \text { Beta Acids } \end{array}{ }^{10}$ | $\begin{array}{llll} 0 & 20 & 40 & 60 \\ \text { Cohumulone Acids } \end{array}$ | $0 \begin{array}{cccc} 0 & 1 & 2 & 3 \\ 0 i l \\ \hline \end{array}$ | $\begin{array}{rcr} 0.0 & 0.20 \\ \text { Oil Humulene } \end{array}$ | $\begin{array}{lc} 0.0 & 0.1 \\ & \text { Oill Myrcene } \end{array}$ |  | Oill Caryophyllene | $0.000 \quad 0.005 \quad 0$. Oill Farnesene |

## Scatterplot matrix

Matrix Plot
Sources: CDC (2009), Census Bureau (2000)

| OBESE (bmi 30.0 - 99.8 ) |  |  |  |
| :---: | :---: | :---: | :---: |
|  | \% in Poverty |  |  |
|  |  | \% of all $\mathrm{HH}=$ Single Parent |  |
|  | ${ }^{\circ} 0^{\circ}{ }^{\circ}{ }^{\circ}$ क $80 \%{ }^{\circ}$ <br> 8 |  |  |

http://www.statsoft.com/support/blog/entryid/212/finding-the-right-pieces-to-the-puzzle

http://junkcharts.typepad.com/junk_charts/2010/06/the-scatterplot-matrix-a-great-tool.html

Trellis plot


## Trellis plot



## Trellis graph - not just scatterplots



## Small multiples



## Small multiples



Tufte．The Visual Display of Quantitative Information

Chaiyya Chaiyya Dance Moves

1 か 8 \＆$\%$＊
2 介 кк人入入

3

4 in if o \＆

5
－会务和

## Small multiples



## Embedded visualization



## Multidimensional scaling

Calculate the similarity of all pairs of records using some distance function

Create a map that maps each record into our 2 (or 3) dimensional space

Calculate the similarity of all pairs of points
Compute the stress on the system as function of the difference between the similarity of the points and the similarity of the original records

If the stress is above some threshold, move points to reduce stress and repeat

## Multidimensional scaling

## MDS of Census Data



