Multivariate visualization

C. Andrews

2016-03-25
Univariate
Univariate: Histogram

CS101 Midterm Grades Histogram
Univariate: Stem and Leaf

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1 2</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>3 4 4 5 8</td>
</tr>
<tr>
<td>7</td>
<td>1 2 2 2 5 6 6 6 6 6 6 7 7 8 9 9 9 9</td>
</tr>
<tr>
<td>8</td>
<td>0 1 1 1 1 2 2 3 3 9</td>
</tr>
<tr>
<td>9</td>
<td>1 2 3 5 5 5 5 6 7 8 8 9 9</td>
</tr>
<tr>
<td>10</td>
<td>0 0</td>
</tr>
</tbody>
</table>
Univariate: Box and Whiskers

25th percentile  median  75th percentile

low value  high value
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

http://www.doka.ch/Excel3Dscatterplot.htm
Scatterplot matrix
Scatterplot matrix

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

- OBESE (BMI 30.0 - 99.8)
- % in Poverty
- % of all HH = Single Parent
- % Driving Alone

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

Trellis plot
Trellis plot
Trellis graph - not just scatterplots
Small multiples

Heer et al., “A Tour Through the Visualization Zoo”
Small multiples

Tufte. The Visual Display of Quantitative Information

http://jeffreymilanetti.wordpress.com/2012/11/
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