Visualization of Data and Statistical Models Using R

Yihui XIE
School of Statistics, RUC
Outline

1. Get started by a map
2. Introduction to R
   a) What's R
   b) R graphics
3. Two examples
   a) Regression tree
   b) Geographical maps
1. Get started

☐ Why visualization?
  ■ Intuitional – data are complicated

☐ How?
  ■ Compute statistics – information
  ■ Technique: combine as many variables as possible in a single plot
1.1 A Famous Map

C. J. Minard's Map: Napoleon's March of 1812

Source: http://www.ddg.com/LIS/InfoDesignF96/Kelvin/Napoleon/map.html
1.1 A Famous Map (cont'd)

Crossing the Bérézina, 26th November. Of 40,000 men, 25,000 were lost.
1.1 A Famous Map (cont'd)

Map of Napoleon's March of 1812 (animated)

Source: http://www.ddg.com/LIS/InfoDesignF96/Kelvin/Napoleon/map.html
1.2 Summary – six variables

- Size of the French army
  - Band width
- Location on a 2D surface
  - From Poland to Moscow
- Direction of invasion and retreat
  - Upper and lower band
- Temperature during the retreat
  - Cold
2. Introduction to R

- Free, open source
- For statistical computation & graphics
  - Flexible & powerful
  - Very quick implementations of advanced theories in statistics
- Many contributors (strong teamwork)
- Homepage: [www.R-project.org](http://www.R-project.org)
2. Introduction to R (cont'd)

- Demonstrations of R graphics
  - Contour plot, level plot
  - Conditioning plot
  - Scatterplot matrix
  - 3D surfaces, lines, dots, etc
  - Symbols plot
  - Maps
  - ...

3. Two examples

- Regression tree
  - Export competitiveness: how to select explanatory variables? And what information can we learn from the tree?

- Geographical maps
  - Combining agricultural export competitiveness with geographical information
3.1 Examples: regression tree

- Classification And Regression Tree
  - CART, known as decision tree earlier
  - "Classify" the response variable by splitting explanatory variables
  - R package: `rpart`
### 3.1 Examples: regression tree

<table>
<thead>
<tr>
<th>Country / Region</th>
<th>export</th>
<th>protectionism</th>
<th>culture</th>
<th>law</th>
<th>politics</th>
<th>credit</th>
<th>transport</th>
<th>rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARGENTINA</td>
<td>12.83</td>
<td>4.13</td>
<td>6.34</td>
<td>2.40</td>
<td>4.47</td>
<td>26.40</td>
<td>0.21</td>
<td>0.44</td>
</tr>
<tr>
<td>AUSTRALIA</td>
<td>9.445</td>
<td>7.56</td>
<td>8.22</td>
<td>6.49</td>
<td>9.71</td>
<td>87.20</td>
<td>0.22</td>
<td>1.69</td>
</tr>
<tr>
<td>AUSTRIA</td>
<td>28.33</td>
<td>7.92</td>
<td>7.28</td>
<td>6.83</td>
<td>9.42</td>
<td>91.40</td>
<td>0.53</td>
<td>2.26</td>
</tr>
<tr>
<td>BAVARIA</td>
<td>18.285</td>
<td>6.84</td>
<td>6.58</td>
<td>4.26</td>
<td>9.23</td>
<td>91.40</td>
<td>0.54</td>
<td>2.90</td>
</tr>
<tr>
<td>BELGIUM</td>
<td>51.515</td>
<td>7.08</td>
<td>7.30</td>
<td>3.40</td>
<td>7.87</td>
<td>89.40</td>
<td>0.78</td>
<td>1.85</td>
</tr>
<tr>
<td>BRAZIL</td>
<td>8.55</td>
<td>4.95</td>
<td>7.59</td>
<td>2.90</td>
<td>6.30</td>
<td>48.20</td>
<td>0.06</td>
<td>0.93</td>
</tr>
<tr>
<td>BULGARIA</td>
<td>30.965</td>
<td>4.92</td>
<td>7.50</td>
<td>4.00</td>
<td>5.50</td>
<td>55.00</td>
<td>0.21</td>
<td>0.50</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
3.1 Examples: regression tree

transport $< 0.525$

law $< 4.335$

credit $< 62.4$

rd $\geq 1.205$

16.82, n=25

12.05, n=5

19.48, n=15

37.22, n=8

54.76, n=8

node 3 node 5 node 7 node 8 node 9

node 3

node 5

node 7

node 8

node 9
3.2 Examples: geographical maps

Agricultural products competitiveness

-0.997

0.878
3.2 Examples: geographical maps
Thanks!

☐ A "THANKS" drawn by hand 😊