The 7th KNIME User Group Meeting
- Welcome & Introduction -

Michael R. Berthold
KNIME.com AG, Zurich, Switzerland
Our Stars

7* club:
Thorsten Naumann
(Taka Ohshima)

6*:
Volker Eyrich
Greg (Gregor, Gregory) Landrum

And our 5*ers:
Fabian Bös
Jean-Christophe Mozziconacci
Jürgen Reymann
2004-2014

10 Years of KNIME-ing
After 10 Years...

- 1.7 Mio Lines of Code
- 22,000+ Users
- 400+ Customers
- 50+ community and partner extensions
Once upon a time...

January 2004:

interface DataTable {
    String getID();
    RowIterator getRowIterator();
    BlockIteratorgetBlockIterator(...);
    int getNumColumns();
    DataCell getColumnName(int column);
    DataCell getColumnType(int column);
    int getNumRows(); [-1,0,...,n]
}

interface RowIterator {
    void reset();
    boolean end();
    DataRow next();
}

interface DataRow {
    int getSize();
    DataCell getRowType();
    DataCell getRowkey();
    DataCell getCellAt(int index);
    int contains(DataCell cell); [-1,0,...,n]
    DataCell clone();
}
Our Goals in 2004

A Data Analysis Environment that is...

• Intuitive
• Stable
• Scalable
• Extensible
• Open Source
• Application Agnostic
Our Users like in 2014

An Open Platform that is

• Integrative
• Transparent
• Flexible (Agile)
• Collaborative
• Powerful
Open = Integrative

- IT Integration
- Data Integration
- Tool Integration
- Scalability Integration
Open = Transparent

- Intuitive GUI
- Open Source
- And: a transparent business model.
Open = Agile

- Visual documentation...
- ...Reproducibility
- ...Reusability
- Easy to extend
- Easy to evaluate
- Easy to adopt.
Open = Collaborative

- Combine best of breed tools (community, partners, inhouse, legacy...)
- Learn from the Community
- Share your tools
- Share your results
Open = Powerful

- The platform is the sum of all of its parts!
- Powerful ETL & Analytics: choice of data, tools, and science.
“...You can no longer win with a closed-source platform, and you can’t build a successful stand-alone company purely on open source. [...] We believe we need to build value — real, differentiated intellectual property that makes customers successful — as part of a closed-source, proprietary IP portfolio. We can’t do it at the platform level, because platform software has to be open source. We need to look elsewhere, at complementary tools and layered offerings that let customers run that pure open source platform better, more affordably, getting more value out of it.”

Mike Olson, Cloudera Chief Strategy Officer
KNIME Today
KNIME is growing

• User base more than doubled in 2013
• LinkedIn Profiles listing KNIME as a skill: >1000 (from ~0 in 2012)
Top in User Satisfaction: 2012 and also 2013!

<table>
<thead>
<tr>
<th>Software or platform Type:</th>
<th>Overall</th>
<th>IBM SPSS Statistics</th>
<th>IBM SPSS Modeler</th>
<th>KNIME</th>
<th>R</th>
<th>Rapid Miner</th>
<th>SAS</th>
<th>SAS Enterprise Miner</th>
<th>STATISTICA (StatSoft)</th>
<th>Weka</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Proprietary</td>
<td>Proprietary</td>
<td>Open Source</td>
<td>Open Source</td>
<td>Open Source</td>
<td>Proprietary</td>
<td>Proprietary</td>
<td>Proprietary</td>
<td>Open Source</td>
</tr>
<tr>
<td>Quality and accuracy of model performance</td>
<td>4.34</td>
<td>3.93</td>
<td>4.32</td>
<td>4.32</td>
<td>4.43</td>
<td>4.37</td>
<td>4.34</td>
<td>4.26</td>
<td>4.69</td>
<td>4.21</td>
</tr>
<tr>
<td>Dependability/Stability of software</td>
<td>4.25</td>
<td>4.19</td>
<td>4.05</td>
<td>4.43</td>
<td>4.34</td>
<td>4.22</td>
<td>4.32</td>
<td>4.44</td>
<td>4.56</td>
<td>3.73</td>
</tr>
<tr>
<td>Variety of available algorithms</td>
<td>4.20</td>
<td>3.69</td>
<td>4.30</td>
<td>4.48</td>
<td>4.72</td>
<td>4.54</td>
<td>4.01</td>
<td>4.00</td>
<td>4.63</td>
<td>4.33</td>
</tr>
<tr>
<td>Ease of use</td>
<td>4.19</td>
<td>4.23</td>
<td>4.60</td>
<td>4.55</td>
<td>3.56</td>
<td>4.47</td>
<td>3.69</td>
<td>4.00</td>
<td>4.49</td>
<td>4.05</td>
</tr>
<tr>
<td>Ability to automate repetitive tasks</td>
<td>4.17</td>
<td>3.75</td>
<td>3.96</td>
<td>4.39</td>
<td>4.39</td>
<td>4.40</td>
<td>4.31</td>
<td>4.00</td>
<td>4.45</td>
<td>3.71</td>
</tr>
<tr>
<td>Data manipulation capabilities</td>
<td>4.15</td>
<td>4.00</td>
<td>4.32</td>
<td>4.35</td>
<td>4.10</td>
<td>4.27</td>
<td>4.45</td>
<td>3.82</td>
<td>4.41</td>
<td>3.52</td>
</tr>
<tr>
<td>Quality of output / Ease of interpretation</td>
<td>4.10</td>
<td>3.91</td>
<td>4.04</td>
<td>4.39</td>
<td>4.04</td>
<td>4.36</td>
<td>3.69</td>
<td>4.00</td>
<td>4.53</td>
<td>3.66</td>
</tr>
<tr>
<td>Good metrics of model quality</td>
<td>4.10</td>
<td>3.85</td>
<td>3.96</td>
<td>4.05</td>
<td>4.13</td>
<td>4.28</td>
<td>4.08</td>
<td>4.18</td>
<td>4.50</td>
<td>3.85</td>
</tr>
<tr>
<td>Good variable discovery, profiling and selection</td>
<td>4.03</td>
<td>3.70</td>
<td>4.06</td>
<td>4.17</td>
<td>3.98</td>
<td>4.33</td>
<td>3.81</td>
<td>4.35</td>
<td>4.44</td>
<td>3.69</td>
</tr>
<tr>
<td>Quality of user interface</td>
<td>4.03</td>
<td>4.11</td>
<td>4.53</td>
<td>4.62</td>
<td>3.36</td>
<td>4.45</td>
<td>3.58</td>
<td>3.91</td>
<td>4.49</td>
<td>3.59</td>
</tr>
<tr>
<td>Ease of model deployment (scoring models to other data sets)</td>
<td>4.03</td>
<td>3.61</td>
<td>4.13</td>
<td>4.43</td>
<td>3.82</td>
<td>4.20</td>
<td>3.90</td>
<td>4.21</td>
<td>4.46</td>
<td>3.77</td>
</tr>
<tr>
<td>Speed</td>
<td>4.02</td>
<td>3.84</td>
<td>4.13</td>
<td>4.12</td>
<td>3.58</td>
<td>3.90</td>
<td>4.03</td>
<td>3.57</td>
<td>4.48</td>
<td>3.53</td>
</tr>
<tr>
<td>Data quality assessment and data preparation capabilities</td>
<td>4.00</td>
<td>3.94</td>
<td>4.22</td>
<td>4.33</td>
<td>3.76</td>
<td>4.20</td>
<td>4.05</td>
<td>3.68</td>
<td>4.38</td>
<td>3.53</td>
</tr>
<tr>
<td>Ability to handle very large data sets</td>
<td>3.99</td>
<td>3.82</td>
<td>4.21</td>
<td>4.35</td>
<td>2.95</td>
<td>3.74</td>
<td>4.41</td>
<td>4.44</td>
<td>4.58</td>
<td>3.03</td>
</tr>
<tr>
<td>Ability to modify algorithm options to fine-tune analyses</td>
<td>3.95</td>
<td>3.17</td>
<td>3.59</td>
<td>3.98</td>
<td>4.33</td>
<td>4.23</td>
<td>3.97</td>
<td>3.97</td>
<td>4.33</td>
<td>3.88</td>
</tr>
<tr>
<td>Enables mining within one’s database</td>
<td>3.94</td>
<td>3.54</td>
<td>4.26</td>
<td>4.12</td>
<td>3.75</td>
<td>4.10</td>
<td>3.92</td>
<td>4.00</td>
<td>4.19</td>
<td>3.61</td>
</tr>
<tr>
<td>Ability to easily incorporate data at different levels of granularity (e.g. transaction data and customer data)</td>
<td>3.90</td>
<td>3.56</td>
<td>4.05</td>
<td>4.24*</td>
<td>3.77</td>
<td>3.99</td>
<td>4.14</td>
<td>3.94</td>
<td>4.24*</td>
<td>5.29</td>
</tr>
<tr>
<td>Useful help menu, demos and tutorials</td>
<td>3.87</td>
<td>3.83</td>
<td>3.99</td>
<td>3.93</td>
<td>3.88</td>
<td>3.90</td>
<td>3.76</td>
<td>3.79</td>
<td>4.35</td>
<td>3.61</td>
</tr>
<tr>
<td>Cost of software</td>
<td>3.79</td>
<td>3.16</td>
<td>3.00</td>
<td>4.35</td>
<td>4.90</td>
<td>4.82</td>
<td>2.33</td>
<td>2.74</td>
<td>3.90</td>
<td>4.88</td>
</tr>
<tr>
<td>Overall</td>
<td>4.04</td>
<td>3.76</td>
<td>4.07</td>
<td>4.52</td>
<td>3.99</td>
<td>4.25</td>
<td>3.89</td>
<td>3.98</td>
<td>4.44</td>
<td>3.74</td>
</tr>
</tbody>
</table>
Recognition

In 2010 KNIME was named Gartner Cool Vendor for Business Intelligence and Analytics.
Recognition

...this year, don’t miss Gartner’s Upcoming Report on “Advanced Analytic Platforms”...
The first 2014 Highlight: The User Group Meeting!