From Visual Data Exploration and Analysis to Scientific Conclusions

Alexandra Vamvakidou, PhD

September 15th, 2016
The Power of a Visual Data
We Collect Data to Compare and Add Knowledge
How Companies Learn Your Secrets

By CHARLES DUHIGG
Published: February 16, 2012 | 570 Comments
WE'VE DECIDED TO TAKE BIG DATA TO THE NEXT LEVEL...

HUMONGOUS DATA

Data is Everywhere
<table>
<thead>
<tr>
<th>Price</th>
<th>Change</th>
<th>Percentage</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.200</td>
<td>0.73%</td>
<td>27.430</td>
<td>7200</td>
</tr>
<tr>
<td>0.460</td>
<td>1.07%</td>
<td>43.630</td>
<td>1900</td>
</tr>
<tr>
<td>0.026</td>
<td>0.25%</td>
<td>1.052</td>
<td>0</td>
</tr>
<tr>
<td>0.200</td>
<td>0.73%</td>
<td>27.430</td>
<td>7200</td>
</tr>
<tr>
<td>0.040</td>
<td>0.52%</td>
<td>7.700</td>
<td>39300</td>
</tr>
<tr>
<td>460</td>
<td>1.07%</td>
<td>43.630</td>
<td>1900</td>
</tr>
</tbody>
</table>
Not looking at this kind of Data
Visuals Help us with Data Exploration
As Humans we Can…

- Recognize people and objects
- Navigate through obstacles
- Understand mood in the scene
- Imagine stories

Scientists:
- Identify outliers
- See patterns and trends
- Create new hypotheses

“One picture is worth a thousand words”
Data Exploration

- Traditional data analysis is very static
  - Takes time and requires highly focused questions
  - Obvious disadvantage
    - We miss out on important trends that we weren’t looking for

Unable to interpret the data to it’s full extent
What if…

- data exploration was dynamic?
- getting answers to new questions was instantaneous?
Things you DON'T know

Questions you are asking

BUSINESS INTELLIGENCE

Questions you are NOT asking

DATA DISCOVERY
Science Data is Deep Data
Scientific data involves **Data Mashup**, combining data from various types and sources. You can work with HCS data, chemistry with structures, plate based assay or genomics data.
**SciSTREAM** is a generic data loader to batch import files, aggregate and arrange raw instrument data and metadata in **Spotfire**.
<table>
<thead>
<tr>
<th>Plate information</th>
<th>Plate</th>
<th>Repeat</th>
<th>Barcode</th>
<th>Measured height</th>
<th>Chamber temp</th>
<th>Chamber temp</th>
<th>Humidity at start</th>
<th>Humidity at end</th>
<th>Ambient temp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>ESP0005125</td>
<td>13.7</td>
<td>29.6</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>29.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Results for AlphaScreen xxxx</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>49220</td>
<td>45004</td>
<td>41750</td>
<td>45565</td>
<td>47559</td>
<td>49656</td>
<td>48635</td>
<td>47270</td>
</tr>
<tr>
<td>B</td>
<td>44771</td>
<td>47869</td>
<td>47512</td>
<td>48104</td>
<td>48930</td>
<td>49335</td>
<td>45819</td>
<td>53325</td>
</tr>
<tr>
<td>C</td>
<td>51239</td>
<td>51144</td>
<td>43925</td>
<td>48242</td>
<td>51913</td>
<td>49247</td>
<td>52570</td>
<td>47566</td>
</tr>
<tr>
<td>D</td>
<td>45776</td>
<td>45191</td>
<td>47543</td>
<td>51047</td>
<td>48380</td>
<td>52571</td>
<td>46570</td>
<td>46188</td>
</tr>
<tr>
<td>E</td>
<td>61124</td>
<td>50706</td>
<td>52661</td>
<td>49984</td>
<td>51238</td>
<td>50510</td>
<td>55322</td>
<td>49154</td>
</tr>
<tr>
<td>F</td>
<td>49261</td>
<td>45250</td>
<td>47842</td>
<td>48027</td>
<td>49152</td>
<td>48125</td>
<td>47996</td>
<td>48507</td>
</tr>
<tr>
<td>G</td>
<td>49002</td>
<td>49944</td>
<td>50631</td>
<td>49300</td>
<td>54376</td>
<td>49531</td>
<td>53048</td>
<td>50303</td>
</tr>
<tr>
<td>H</td>
<td>49567</td>
<td>50376</td>
<td>44074</td>
<td>50208</td>
<td>49343</td>
<td>54021</td>
<td>47040</td>
<td>47630</td>
</tr>
<tr>
<td>I</td>
<td>73781</td>
<td>50098</td>
<td>54653</td>
<td>44689</td>
<td>52447</td>
<td>49603</td>
<td>44678</td>
<td>49290</td>
</tr>
<tr>
<td>J</td>
<td>50475</td>
<td>50997</td>
<td>49826</td>
<td>53526</td>
<td>49918</td>
<td>52821</td>
<td>50584</td>
<td>48215</td>
</tr>
<tr>
<td>K</td>
<td>55558</td>
<td>49475</td>
<td>57282</td>
<td>48802</td>
<td>49335</td>
<td>51191</td>
<td>51119</td>
<td>49143</td>
</tr>
<tr>
<td>L</td>
<td>47731</td>
<td>49434</td>
<td>50372</td>
<td>50372</td>
<td>37770</td>
<td>45336</td>
<td>48972</td>
<td>46215</td>
</tr>
<tr>
<td>M</td>
<td>46589</td>
<td>50012</td>
<td>49671</td>
<td>49359</td>
<td>54017</td>
<td>49362</td>
<td>47011</td>
<td>40805</td>
</tr>
<tr>
<td>N</td>
<td>47727</td>
<td>49851</td>
<td>47223</td>
<td>52020</td>
<td>47288</td>
<td>56122</td>
<td>45389</td>
<td>48452</td>
</tr>
<tr>
<td>O</td>
<td>61320</td>
<td>51148</td>
<td>54660</td>
<td>50923</td>
<td>55480</td>
<td>50338</td>
<td>51114</td>
<td>49655</td>
</tr>
<tr>
<td>P</td>
<td>47755</td>
<td>51641</td>
<td>51162</td>
<td>51549</td>
<td>51534</td>
<td>52708</td>
<td>52241</td>
<td>51830</td>
</tr>
</tbody>
</table>
LEAD DISCOVERY is an add-on for analysis of chemistry data. It enables the import, rendering, filtering and analysis of data with chemical structures.

Visualizations + Chemical Intelligence
The dataset to the right contains over 1 M structures:

- Structure Filter Performance is optimized for datasets of up to 25k compounds.
- Datasets up to 1M compounds have been used in testing.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Registration Number</th>
<th>IUPAC CAS Name</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AR-2547657-1</td>
<td>4-(2,4-dimethoxyphenyl)piperazin-1-yl-1-phenyl-1H-pyrazolo[3,4-b]pyridin-2-one</td>
</tr>
<tr>
<td></td>
<td>AR-2540553-1</td>
<td>1H-[1,2,3]-triazolo[4,5-b][1,4]benzodiazepine</td>
</tr>
<tr>
<td></td>
<td>AR-2529876-1</td>
<td>4-[(4-ethyl-4H-pyran-2-yl)methyl]-3-(3-fluorophenyl)-1H-indazole</td>
</tr>
<tr>
<td></td>
<td>AR-2525006-1</td>
<td>N-(1,3-benzodioxol-5-yl)-2-[4-(4-quinoxalinyl)ethyl]-4-piperidinecarboxamide</td>
</tr>
</tbody>
</table>
**OmicsOffice** is an add-on for the analysis of genomics data using guided workflows as well as a rich toolkit that includes advanced statistics, data integration tools and functional analysis methods.
**Extensibility to statistical engines.** This allows statisticians and programmers to build analytic workflows more quickly and enable others to use them.
Selecting the Right Chart Type for Your Data

Comparison Data

Transition Data

Composition Data
Labels and Tool Tips

- Mark and display additional information
- Hover over item to show different values
- Display images, structures, links etc.
- Trellis to display data with more granularity
Filtering Schemes

Comments
First mark compounds below or on another page. Each panel shows dose-response curves for a single compound. Clicking on a compound will select it for analysis.

Dose Response Curves

- **Corr_ID**: SF0001, SF0002, SF0003, SF0004, SF0005, SF0006, SF0007

Data limited by:
- **Marking**
- **Logistic Regression Curve Fit**
- Assay1: Reference, Assay2: Inflection Point, Assay3: Inflection Point

Logistic Regression Curve Fit:
- Assay1: Per compound
- Assay2: Per compound
- Assay3: Per compound

Dose Range:
- 0.01 to 1
- Data points at 0.01, 0.1, 1

Percent Activity:
- Data points at 10, 50, 90
- Data points at 10, 50, 90
- Data points at 10, 50, 90

Dose Response
- Assay1
- Assay2
- Assay3

RLU (cps)
- 155
- 39570

Criteria
- Percent Activity
- Corr_ID
Creating a Self-Explanatory Drill-Down Interface

- Details on Demand, Details Visualizations
  - Data selection depends on marked data in other visualization
    - Use trellis to provide title
    - Color settings are not reset anymore
Summarizing and Annotating Charts: Text Areas

- Extra information on the analysis
- Shortcuts to specific selections and markings
- To improve the overall layout

Correlation analysis:
- Low correlation between Drug 1 and Drug 2
- High correlation between Drug 3 and Drug 5
- Outlier or interesting virus?
Collaborations

Brand A sales vs Class sales

![Graph showing sales comparison](image)

**Browse conversations**

<table>
<thead>
<tr>
<th>Time</th>
<th>User</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:39 AM</td>
<td>FTAC</td>
<td>This curve says a lot about what kind of stores that is important to us.</td>
</tr>
<tr>
<td>11:39 AM</td>
<td>FTAC</td>
<td>Yes, but how about profits? Is it distributed in the same way?</td>
</tr>
<tr>
<td>11:27 AM</td>
<td>FTAC</td>
<td>Can we discuss sales in these states in our meeting this afternoon?</td>
</tr>
<tr>
<td>11:32 AM</td>
<td>FTAC</td>
<td>We dont make to much from these so called high value stores.</td>
</tr>
</tbody>
</table>
Annotations

- Available in:
  - Spotfire Business Author and Spotfire Analyst,

- You can:
  - Add textual descriptions to visualizations in the form of a text box appearing on top of a visualization.
  - Draw attention to the key points in your visualization.
  - Specify the color and width of the text box,
Using Multiple Data Tables

• Different data sources
  ◦ One-to-many relations between different data tables
  ◦ Visualizations generally use only 1 data table
• Same data in different pivot state

<table>
<thead>
<tr>
<th>Drug</th>
<th>Measurement 1</th>
<th>Measurement 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPV</td>
<td>50.40</td>
<td>45.80</td>
</tr>
<tr>
<td>IDV</td>
<td>23.20</td>
<td>28.30</td>
</tr>
<tr>
<td>Reference</td>
<td>45.50</td>
<td>23.20</td>
</tr>
</tbody>
</table>

• Unpivot
• New column: 
  first(if([Drug]="Reference", [Value], null)) OVER [Category]

Curve from Data Table
In Summary

- Interactive Visualizations
  - Access corporate and local data sources and spreadsheets
  - Analyze and explore data with intuitive, interactive visualizations
  - Capture and collaborate around analysis workflows
  - Distribute analyses to colleagues

- R&D Applications
  - Easy functional & biological analyses: Allowing researchers to interpret biological relevance of their data
  - Availability of specific guided workflows designed for different genomic technologies with Click and Go® technology
  - Powerful visualizations: Powered by TIBCO® Spotfire® platform, OmicsOffice offers the best visualization and data mining options in the market
Thank you!

alexandra.vamvakidou@perkinelmer.com

Join Us at the Interactive Stations to Learn More