How To Become a Data Scientist With Spotfire 5

Today is a Webcast about What's New with Spotfire Version 4.5 for Big Data / Extreme Information, Pervasive Prediction, and Enterprise-Class Mobile Apps. While I am really looking forward to that, I have already started to think about a version 5 of Spotfire that will enable one to become a Data Scientist.
A **data scientist** is a job title for an employee or business intelligence (BI) consultant who excels at analyzing data, particularly large amounts of data, to help a business gain a competitive edge. Perhaps the most important skill a data scientist possesses, however, is the ability to explain the significance of data in a way that can be easily understood by others. Data scientists are in great demand today.

While TIBCO **Spotfire** is a comprehensive software platform that allows customers, like Fortune 500 companies, to analyze data, using predictive and complex statistics, there is also need for a version for individuals that want to analyze data and become data scientists, thus enter **Silver Spotfire** several years ago.

In fact Silver Spotfire helped me to became a Data Scientist several years ago when I was still working for the Federal Government. Some highlights of my career path were:

- Trained in data science, statistics, and visualization
- Worked with Ben Shneiderman and Christopher Ahlberg on very wide tables from EPA with the forerunner of Spotfire
- Used the original S-Plus and Spotfire before they were acquired by TIBCO
- Used Spotfire Professional 3.1 on US EPA and Federal Data Science Products
- Encouraged development of Silver Spotfire 1 and 2
- Built MindTouch Knowledgebase and Spotfire Library of Data Science Blogging and Journalism for AOL Government and Semantic Community

Now my expert advice from all my experience with both Spotfire and Silver Spotfire, briefly distilled into bullets, for getting to a Spotfire 5 for Data Scientists is as follows:

- Use Spotfire on Spotfire Content for Business Intelligence and Business Analytics:
  - Trends and Outliers: TIBCO Spotfire’s Business Intelligence Blog
  - Demo and Template Gallery
  - Reuse Data from Individual Spotfire Files
- Show Spotfire Blogging Team How to Build Data Science Products:
  - Pre-condition data sets
  - Value added data sets that use Spotfire Analytics
  - New data types: Text, APIs, RDF, etc.
- A New Way to Market Spotfire:
  - Big data in memory in the cloud front end
  - Open government and social media data analysis tool
  - Faceted search web browser like embedded widgets

Now Spotfire Team has certainly been moving in this direction early on with the Spotfire Business Intelligence Blog for June 8, 2010: [The 2010 FIFA World Cup – Can You Predict the Winner?](http://semanticommunity.info/A_Spotfire_Gallery) where the blog says:

TIBCO has put together World Cup data trends dating back to 1930 and developed a neat analytic solution. It allows you to examine country statistics, including World Cup appearances, wins, games played, goals scored, penalties, attendance rates, referee nationality, and more. The application, which uses the TIBCO Spotfire WebPlayer, combined with its Spotfire Server and Statistic Services, also predicts the bracket finalists. You can take Tibco’s word for it, or you can go one step further and upgrade or downgrade particular teams to impact the results based on your own set of
opinions or hunches. Have some fun checking out all the data and visual displays. Who know – you might be able to predict the winner for yourself, and win a million Euro.

and contains the link to an actual Spotfire Web Player application. We just need more of these. So I did one quickly to demonstrate what I mean following the bullets above:

I did the first two major bullets above by building a new value-added spreadsheet that shows Spotfire Analytics on Spotfire’s own Social Business Intelligence data in the Web Player. Interestingly Business Intelligence (234), Business Analytics (152), and Data Analytics (125) accounted for for 511 of the 1429 total topics (about a third), and with Big Data (54) and Data Scientist (21) showing small numbers because they are hot topics now.

I have also tried to build a data set that uses all 15 Spotfire visualizations. See my Heritage Provider Network Health Prize Application for the upcoming Health Datapalooza, June 5-6, 2012.

In the broader perspective of Business Intelligence Software there are ten leaders according to Datamation which says: "The top five vendors below are ranked in terms of market share (about 75%). The remaining five are ranked by the extremely unscientific measure of how likely I (Jeff Vance of Datamation) think they are to challenge the leaders."

• 1. SAP AG
• 2. SAS
• 3. Oracle
• 4. IBM
• 5. Microsoft
• 6. MicroStrategy
• 7. TIBCO Spotfire
• 8. Information Builders
• 9. QlikTech
• 10. Tableau Software

In my view, Spotfire is the "Swiss Army Knife of BI" and Spotfire Version 5.0 for the Data Scientist: would do the following:

• Implement My Expert Advice (recall above bullets)
• Work with the Nine Other BI Software Formats
• Work with Advanced Processing Frameworks (e.g. Cray Graph Computer, Hadoop, Watson, etc.)
• Work with Be Informed for Dynamic Case Management
• Work with Social BPM for Semantic Analytics for Big Social Data

A simple example of last two bullets that follows a Business Process (really more Dynamic Case Management):

• The Problem
• See The Data
• Map The Data
• Exploratory Data Analysis: State Data
• Exploratory Data Analysis: School Data
• Facet Search and Curve Fitting
to transform an existing Spotfire Demo Gallery example ([Student Debt Crisis 2012](#)) into a Data Science Product in the new [Gallery](#) and [Web Player](#) is shown. I also provides [slides](#) to show this data story of How To Become a Data Scientist With Spotfire 5. More examples will follow.

---

**Spotfire Inventory and Visualization of Assets**

Also see [Inventory and Visualization of Spotfire Assets](#)

For Internet Explorer Users and Those Wanting Full Screen Display Use: [Web Player](#) Get [Spotfire](#) for [iPad App](#)

---

**Announcements**

**November 17, 2011**: TIBCO Spotfire 4.0 Webcast to Showcase the Power of 'Free Dimensional' Analytics and Collective Intelligence

**November 17, 2011**, PushToTest Heralds TIBCO as a Leader in Composition Approach to Enterprise-Scale SOA

---

**Concepts**

TIBCO Federal Needs Data Science and Data Scientists to Market Spotfire and SOA. See [DoD DCMO RFI](#) and [DoD OIG RFI](#) for example.

Show Me the Data (and the Metadata - Data Dictionary) and Show Me the Results in Side-by-Side Comparisons (Edward Tufte) and with Interactivity (Spotfire)

Data Preconditioning for TIBCO Spotfire Using Data Science by a Data Scientist

Does the Data have Georeferencing or Not? If so, select appropriate Spotfire Boundary Files to start. If not, just drag and drop the Spreadsheet on top of the open Spotfire interface.

Data Preconditioning for TIBCO SOA Using Data Science by a Data Scientist

Excel to Spotfire [Information Designer](#) and TIBCO BPM and TIBCO SOA

Excel to [Be Informed Be Structured](#) and Be Informed DMAD and DTAP and Be Informed Dynamic Case Management

DMAD: Design = Model = Application = Documentation

DTAP: Deployment, Testing (Functionality and Usability, Acceptation, and Production (Upload to the Cloud)
Spotfire and TIBCO SOA Used for Dynamic Case Management (Brand Niemann on Be Informed)

MindTouch WOA Used for Social Knowledgebase

---

**Events**

December 15, 2011: Pilot for NCOIC - NGA Kickoff Meeting

April 3, 2012: Pilot for 13th Federal SOA for e-Gov Conference at MITRE

---

**Trends and Outliers**

Source: [http://spotfireblog.tibco.com/?page_id=2](http://spotfireblog.tibco.com/?page_id=2)

TIBCO Spotfire's Business Intelligence Blog

How To Become a Data Scientist With Spotfire 5: [Slides](#)

For Internet Explorer Users and Those Wanting Full Screen Display Use: [Web Player](#) Get [Spotfire](#) for [iPad App](#)

---

**About Blog**

08/25/2009

retweet

Welcome! Before we describe what you can get from reading our new blog on business intelligence software, we want to introduce ourselves to anyone new to this space. Spotfire is a leading provider of enterprise analytics software. Our mission is to help companies improve decision-making across the enterprise, through the use of user-driven business intelligence solutions, analytics and more.

Between mergers and acquisitions, innovation, economic pressures and more, the business intelligence market is fast-moving and (it feels like) constantly changing. With this blog, we aspire to inform you, maybe provoke you, and possibly even entertain you.

First, let’s cover a few relevant definitions:

- **Business intelligence** is critical to successful competition. It enables end users to quickly and easily see patterns, trends, outliers and unexpected relationships that basic reporting misses.
• **Data visualization** helps users interpret critical relationships in multidimensional data. **Information visualization** allows users to easily query and comprehend complex data. Ideally, these technologies should feature open architectures that integrate with other technologies, including **data mining**, warehousing and CRM systems.

• **Enterprise analytics**, sometimes called **business analytics** or **business intelligence** solutions, supply valuable insights to professionals in a variety of business processes, including **financial analytics**, manufacturing analytics and, sales and marketing analytics.

• Decision analysis software speeds confident business decisions. Decision analysis applications organize and deliver critical information for rapid insight.

What can you find at the Spotfire blog? Our blogging team will cover a wide-range of business intelligence topics, including news, interviews with business intelligence luminaries and influencers, information on trends, and more.

Check back on a regular basis to read about:

• Innovations in business intelligence software
• What business intelligence software is doing for other companies
• Technology and business news related to the business intelligence market and players
• Fun uses of business intelligence software (yes, they exist!)

Lastly, we’ll touch upon how business intelligence and analytics are used in specific verticals, including:

• Business intelligence in life sciences and pharmaceuticals
• Business intelligence in energy
• Business intelligence in financial services
• Business intelligence in manufacturing
• Business intelligence in government and government intelligence
• Business intelligence in academic institutions, semiconductor producers, and consumer goods

So, enjoy reading, and don’t forget to comment. We want to hear from you. Also, subscribe to our blog to stay informed on data analytics topics.

---

**Spotfire Demo and Template Gallery**

**Student Debt Analysis**

For Internet Explorer Users and Those Wanting Full Screen Display Use: [Web Player](#) Get [Spotfire for iPad App](#)

---

**Model Economic Capital Based on Two Risk Factors: Market Risk and Mortality Risk**

This analysis demonstrates a scenario-based view of Economic Capital that is driven by market risk and mortality risk. The user specifies parameters that determine the joint distribution of these two risk factors, and their contribution to overall Economic Capital. TIBCO Spotfire Pre-sales, July 2009, Revised March 2010.
S+ Integration Example: Loess

A simple example incorporating an S+ script that smooths a trend line.

The Data: Measurements (blue) fluctuate daily and need to be smoothed to see the trend.

There are two smoothing techniques:

A rolling average (brown). This technique is useful, but for a low average window, 10 days for example, large fluctuations remain. However for larger windows, say 30 days, the trend develops a lag. Use the slider below to adjust the value.

A trend line (black dashes - - - ) using a technique called Loess Smoothing. Here the general trend is Modeled (calculated) using all the available data, rather than just the set available in a moving window. This produces more accurate identification of general trends.

Airline Incident Analysis Revisited

I have always liked this Spotfire Analysis, but found the recent update using the New Graphical Table (see below) was in error. The number of fatalities is wrong, there are only a total of 77 for the entire United States. Also note that the period is 2004-2007, not 2004-2008. I decided to follow a business process and show the correct number of fatalities, and do Exploratory Data Analysis on relationships between Incidents and Pilot Certifications, State Populations, and Pilot Experience.

Arlington Virginia County Crime Analysis Revisited

This application investigates crime in Arlington County, Virginia. The data, coming from the Arlington County Website, focuses on the Type of Crime, Date/Time, and "notes" from the responding officers.
I found this statement to be incorrect: For example, Monday is the day of week that has the highest count of crime. Perhaps, more interesting, is that on Mondays at 5pm the count of crime events is significantly higher than the count for any other hour during the week. For correct answer see: Cross and Graphical Tables.

Note: This is just a simulation of what could be done if the crime data set was setup for Network Graphics. I do not fully understand what I have done here with the Network Graph and what it means, but it something that should be pursued with crime other types of data. See my previous work.

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

Media, iframe, embed and object tags are not supported inside of a PDF.

US Government Program Effectiveness Analysis Revisited

This is an example of an analysis of both earmarks data and data from the Program Assessment Rating Tool (PART).

The original analysis by the Spotfire Team seems to be overly complicated to me when the basis results can be gotten by just summarizing the data in Summary and Cross Tables. Interestingly, the Earmarks 2008 data set has really nothing to do with the PART Analysis 2008 data set, but provides interesting results, namely the Department of Defense received the largest earmarks (almost $8 Billion), but not have the highest favorable Rating (Effective, Moderately Effective, and Adequate).

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

Media, iframe, embed and object tags are not supported inside of a PDF.

Global Distribution of Natural Disasters (Earthquakes, Tsunamis, and Volcanic Eruptions) Revisited
I find this statement by the creator of this Spotfire file: "Despite an average falling magnitude of Earthquakes (My observation: only since about 1900), the financial impact is ever larger--both in the developed and developing world (My observation: only have damage costs for recent years and due to population growth effect-see Notes below). However, while the average damage has been increasing (My observation: see previous), injury rates have stayed roughly stable (My observation: much more in developing counties since 1900). Fatalities, however show a different story. Here we see a divergence between the developing and developed worlds (My observation: agree, but more in rest of the world for obvious reasons - more people living in poorer construction in more earthquake areas)." to be questionable.

Notes: These data covers Year 10 to Year 2010! The Summary Table shows that Fatalities (1172) and Injuries (789) are reported in less than 50% of the events (5466) so one has to be cautious in interpreting those statistics. For example: Haiti Earthquake 2010 has over 220,000 Fatalities, but No Damage Category given!

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

H1N1 Spread

An overview of the spread of the Summer '09 outbreak of the H1N1 virus in the continental United States.

This was done in an earlier version of Spotfire (3.2) and needs to be upgraded and have the Business Process added as follows:

See The Data
Map The Data By State
Map The Data By County
Map The Data With Background Image
Summary, Cross, & Graphical Tables, and Exploratory Data Analysis: Faceted Search & Plots

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

World Energy Survey Analysis Revisited

The Problem: I thought this Spotfire analysis should be done with a business process in mind to answer the questions below as follows:

–See The Data
–Map The Data
–Summary, Cross, and Graphical Tables
–Exploratory Data Analysis: Faceted Search and Plots
<table>
<thead>
<tr>
<th>Description</th>
<th>Metadata</th>
<th>Data</th>
<th>Analytics</th>
<th>Screen Capture</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>TUCON 2011 Attendee Analysis</td>
<td>MindTouch</td>
<td>Excel</td>
<td>Spotfire</td>
<td>Image</td>
<td></td>
</tr>
<tr>
<td>Hurricane Facts and Analysis</td>
<td>None</td>
<td>Excel</td>
<td>Spotfire</td>
<td>Image</td>
<td></td>
</tr>
<tr>
<td>World Facts Analysis</td>
<td>MindTouch</td>
<td>Excel</td>
<td>Spotfire</td>
<td>Image</td>
<td>AOL Government Story</td>
</tr>
<tr>
<td>DoD Presidential Budget Request</td>
<td>MindTouch</td>
<td>Excel</td>
<td>Spotfire</td>
<td>Image</td>
<td>AOL Government Story</td>
</tr>
</tbody>
</table>

For Internet Explorer Users and Those Wanting Full Screen Display Use: [Web Player](#) Get [Spotfire](#) for [iPad App](#)
<p>| Person, Programs, Operations and Maintenance Programs, Procurement Programs, and Research, Development, Test &amp; Evaluation Programs. The primary goal of the workflow is to provide a view of the budget that can be easily manipulated and drilled into by any user. |
|---|---|
| Trends and statistics are the key points for analysis in this look at unemployment in Spotfire. The Spotfire Analysis file includes data from 1999 through 2009, available at the Bureau of Labor Statistics. The analysis addresses unemployment over that time period across the states and counties in the United States. |
| Unemployment Rates for the US | World Energy Survey Analysis | MindTouch |
| Excel | Spotfire | Image |
| <a href="http://semanticommunity.info/A_Spotfire_Gallery">http://semanticommunity.info/A_Spotfire_Gallery</a> | | |
| Updated: Sat, 19 Sep 2015 02:32:10 GMT | | |
| Powered by MindTouch | | |</p>
<table>
<thead>
<tr>
<th>Topic</th>
<th>Description</th>
<th>Tools</th>
<th>Platforms</th>
<th>Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseball Stats</td>
<td>Who are the best players in baseball for 2005? Where does it pay to play? Stats of 278 great players from 30 teams are represented here. Check out the competitive advantage interactive visual analytics brings!</td>
<td>MindTouch, Excel</td>
<td>Spotfire, Image</td>
<td><a href="http://semanticommunity.info/A_Spotfire_Gallery">AOL Government Story</a></td>
</tr>
<tr>
<td>Mashing up structured data and unstructured content</td>
<td>Use free text search to explore and analyze the unstructured content in Enron emails</td>
<td>MindTouch, Excel</td>
<td>Spotfire, Image</td>
<td><a href="http://semanticommunity.info/A_Spotfire_Gallery">AOL Government Story</a></td>
</tr>
<tr>
<td>Student Debt Analysis</td>
<td>Student debt is at an all-time high, here is an analysis of student debt in USA for various regions &amp; colleges.</td>
<td>MindTouch, Excel, Spotfire (1, 2, 3, 4)</td>
<td>Image</td>
<td><a href="http://semanticommunity.info/A_Spotfire_Gallery">AOL Government Story</a></td>
</tr>
<tr>
<td>Model Economic Capital Based on Two Risk Factors; Market Risk and Mortality Risk</td>
<td>This analysis demonstrates a scenario-based view of Economic Capital that is driven by market risk and mortality risk. The user specifies parameters that determine the joint distribution of these two risk factors, and</td>
<td>MindTouch, Excel (NA)</td>
<td>Spotfire, Image</td>
<td><a href="http://semanticommunity.info/A_Spotfire_Gallery">AOL Government Story</a></td>
</tr>
<tr>
<td>S+ Integration Example: Loess</td>
<td>A simple example incorporating an S+ script that smooths a trend line.</td>
<td>MindTouch</td>
<td>Excel (NA)</td>
<td>Spotfire</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------</td>
<td>-----------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>Airline Incident Analysis Revisited</td>
<td>I have always liked this Spotfire Analysis, but found the recent update using the New Graphical Table (see below) was in error. The number of fatalities is wrong, there are only a total of 77 for the entire United States. Also note that the period is 2004-2007, not 2004-2008. I decided to follow a business process and show the correct number of fatalities, and do Exploratory Data Analysis on relationships between Incidents and Pilot Certifications, State</td>
<td>MindTouch</td>
<td>Excel (NA)</td>
<td>Spotfire</td>
</tr>
<tr>
<td>Title</td>
<td>Description</td>
<td>Tools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Arlington Virginia County Crime Analysis Revisited</td>
<td>This application investigates crime in Arlington County, Virginia. The data, coming from the Arlington County Website, focuses on the Type of Crime, Date/Time, and &quot;notes&quot; from the responding officers.</td>
<td>MindTouch</td>
<td>Excel (NA)</td>
<td>Spotfire</td>
</tr>
<tr>
<td>US Government Program Effectiveness Analysis Revisited</td>
<td>This is an example of an analysis of both earmarks data and data from the Program Assessment Rating Tool (PART).</td>
<td>MindTouch</td>
<td>Excel (NA)</td>
<td>Spotfire</td>
</tr>
<tr>
<td>Global Distribution of Natural Disasters (Earthquakes, Tsunamis, and Volcanic Eruptions) Revisited</td>
<td>I find this statement by the creator of this Spotfire file: &quot;Despite an average falling magnitude of Earthquakes (My observation: only since about 1900),</td>
<td>MindTouch</td>
<td>Excel (NA)</td>
<td>Spotfire</td>
</tr>
</tbody>
</table>

http://semanticommunity.info/A_Spotfire_Gallery
Updated: Sat, 19 Sep 2015 02:32:10 GMT
Powered by MindTouch™
the financial impact is ever larger—both in the developed and developing world (My observation: only have damage costs for recent years and due to population growth effect—see Notes below). However, while the average damage has been increasing (My observation: see previous), injury rates have stayed roughly stable (My observation: much more in developing counties since 1900). Fatalities, however show a different story. Here we see a divergence between the developing and developed worlds (My
<table>
<thead>
<tr>
<th>Observation: agree, but more in rest of the world for obvious reasons - more people living in poorer construction in more earthquake areas)&quot; to be questionable.</th>
</tr>
</thead>
<tbody>
<tr>
<td>An overview of the spread of the Summer '09 outbreak of the H1N1 virus in the continental United States. This was done in an earlier version of Spotfire (3.2) and needs to be upgraded and have the Business Process added as follows: See The Data Map The Data By State Map The Data By County Map The Data With Background Image Summary, Cross, &amp; Graphical Tables, and Exploratory Data Analysis:</td>
</tr>
</tbody>
</table>
The terms described by the DIA, Collection, Analysis, and Collaboration, were activities during the process of a tasker with a deadline having to spend too much time getting the data and building a beautiful Powerpoint. The problem is that a high amount of the tasker’s time is spent in collection, followed by a low amount of time in analysis (where the actual work is), and then a high amount again in collaboration. This creates a “bathtub” like curve. Hanging on the wall of the CIO’s office at the DIA are posters that address the bathtub problem and the goal of Inverting the Bathtub. This is the problem that Spotfire addresses and, as a result, gives more time to think.
Spotfire inverts your Bathtub and improves communication flow

Activity levels without Spotfire vs. Activity levels with Spotfire

FLOW

Evaluation

Collection

Analysis

Communications

Time to complete task

Data in motion

PowerPoint
Excel
Word
PDF
METAL
Images (e.g., gif, jpg, png, bmp)