Splunk>: Atlanta Meetup
Advanced Visualizations

Kris Hutchinson
July 2014 Meeting
Purpose of this document

• Understand how visualizations affect the way that you interpret data.
• Highlight examples of what is currently possible in Splunk 6.x
• Highlight links and resources available to someone who would like to work with more than “what is out of the box”
• Present summary
Why Visualization?

### Visualization Libraries and Examples
- d3.org
- d3 - GitHub MboStock
- +datavisualization.ch
- Leaflet.js
- The 5 Best Libraries for Building Data Visualizations

### D3
- D3 Gallery
- Plugins
- API Reference
- mbostock's Blocks

### Apps Used
- Splunk 6.x Dashboard Examples
- Splunk Web Framework Toolkit
- Simple XML Examples (Toggle)
- Gantt Chart Visualization
- R Project

### Tutorials
- Splunk Web Framework Overview
- Add a D3 Visualization using Splunk JS Stack
- Dashing D3 Tutorial
- mbostock D3 Tutorial
- Databases of statistical information
- D3 Cookbook

### Old School Splunk
- Afterglow
- Gantt Chart Visualization
- Particle

### Miscellaneous
- Add a D3 Visualization using Splunk JS Stack
creativebloq - design tools
- Splunk Telco Usecase, Leaflet.js and scrubber.
- Big Data Visualization Images on Google
- 3 Errors to Avoid in Big Data Visualizations
- The Urban Observatory
- The Need for Smart Data Visualization
- D3 Google Group
- JS Bin (online testing)
"User Interface Design is important as a discipline of user experience engineering that focuses on interaction between the user and the parts of an application that facilitate the user's desires. To put it simply, UI design allows the user to achieve goals quickly and easily, turning any app — no matter how sophisticated — into a seamless environment that responds intuitively to the user. Truly effective User Interface Design means the user will only rarely need to look at help files, either online or written, because the software is structured to respond to a natural human thought process."

"The greatest value of a picture is when it forces us to notice what we never expected to see."
— John W. Tukey, Exploratory Data Analysis. 1977

"And that is precisely what we’re supposed to be getting from Big Data and Big Content or whatever you choose to call it. It should ideally show us connections and ideas we might not have considered or that would have taken us ages to get to manually with our meager human brains processing it."

"Every picture tells a story, don't it," you begin to see the beauty of the notion--that when you combine a big content repository, data analysis and data visualization, and some really smart people, something truly beautiful happens and you can begin to find greater value in this content then you probably ever dreamed possible.

Thoughtful user experience design is essential to a successful product...

User interface design is important for several reasons. First of all the more intuitive the user interface the easier it is to use, and the easier it is to use and the less expensive to use it. The better the user interface the easier it is to train people to use it, reducing your training costs. The better your user interface the less help people will need to use it, reducing your support costs. The better your user interface the more your users will like to use it, increasing their satisfaction with the work that you have done.
### Why Visualization?

#### Tables

<table>
<thead>
<tr>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
<th>Set 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
<th>Set 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>8.0</td>
<td>5.76</td>
<td>8.0</td>
<td>5.76</td>
</tr>
<tr>
<td>8.0</td>
<td>7.71</td>
<td>8.0</td>
<td>7.71</td>
</tr>
<tr>
<td>8.0</td>
<td>8.84</td>
<td>8.0</td>
<td>8.84</td>
</tr>
<tr>
<td>8.0</td>
<td>8.47</td>
<td>8.0</td>
<td>8.47</td>
</tr>
<tr>
<td>8.0</td>
<td>7.04</td>
<td>8.0</td>
<td>7.04</td>
</tr>
<tr>
<td>8.0</td>
<td>5.25</td>
<td>8.0</td>
<td>5.25</td>
</tr>
<tr>
<td>19.0</td>
<td>12.50</td>
<td>19.0</td>
<td>12.50</td>
</tr>
<tr>
<td>8.0</td>
<td>5.96</td>
<td>8.0</td>
<td>5.96</td>
</tr>
<tr>
<td>8.0</td>
<td>7.91</td>
<td>8.0</td>
<td>7.91</td>
</tr>
<tr>
<td>8.0</td>
<td>6.89</td>
<td>8.0</td>
<td>6.89</td>
</tr>
</tbody>
</table>

#### Descriptive Statistics

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of $x$ in each case</td>
<td>9 (exact)</td>
</tr>
<tr>
<td>Variance of $x$ in each case</td>
<td>11 (exact)</td>
</tr>
<tr>
<td>Mean of $y$ in each case</td>
<td>7.50 (to 2 decimal places)</td>
</tr>
<tr>
<td>Variance of $y$ in each case</td>
<td>4.122 or 4.127 (to 3 decimal places)</td>
</tr>
<tr>
<td>Correlation between $x$ and $y$</td>
<td>0.816 (to 3 decimal places)</td>
</tr>
<tr>
<td>Linear regression line in each case</td>
<td>$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)</td>
</tr>
</tbody>
</table>
Why Visualization?
Why Visualization?

- Highest bandwidth to interact with humans
- Puts the human in the loop:
  - Pattern detection
  - Remembers context
  - Fantastic intuition
  - Can predict

Visualization enables humans to solve problems
Good Visualization is a Process

- Focus on the data – intuitive tools to enable the analyst
- No single visualization exists to handle all data sets
- Workflow to filter and zoom lets you access detail as you need it
- Never lose sight of the raw data
6.x: Simple, Interactive, and Extensible

**POWERFUL ANALYTICS**
- Pivot
- Data Models

**VISUALIZATION EXPLORATION**
- Interactive Forms
- Contextual Drilldown

**CUSTOMIZABLE FRAMEWORK**
- Dashboard Editor
- Web Framework
Explore the Associations in your Data.

- User driven analytics with contextual drilldown
- Build interactive dashboards and user workflows without coding
- Pan and zoom enables more focused analytics with ability to select range of interest on a chart and zoom in for deeper analysis
Extend Splunk native visualizations and search commands with open source libraries and custom code.
Begin creating your very own customized visualizations... giving careful thought to your use case!

The majority of the following slides show visualizations, demos, and tutorials from our Splunk 6.x Dashboard examples app. Others are more advanced but fun to integrate.

Download now!
Sankey
Sankey diagrams are a specific type of flow diagram, in which the width of the arrows is shown proportionally to the flow quantity. They are typically used to visualize energy or material or cost transfers between processes.

Parallel Coordinates
Parallel coordinates is a common way of visualizing high-dimensional geometry and analyzing multivariate data. To show a set of points in an n-dimensional space, a backdrop is drawn consisting of n parallel lines, typically vertical and equally spaced. A point in n-dimensional space is represented as a polyline with vertices on the parallel axes; the position of the vertex on the ith axis corresponds to the ith coordinate of the point. This visualization is closely related to time series visualization, except that it is applied to data where the axes do not correspond to points in time, and therefore do not have a natural order. Therefore, different axis arrangements may be of interest.
A Gantt chart is a type of bar chart, developed by Henry Gantt in the 1910s, that illustrates a project schedule. Gantt charts illustrate the start and finish dates of the terminal elements and summary elements of a project. Terminal elements and summary elements comprise the work breakdown structure of the project. Modern Gantt charts also show the dependency (i.e., precedence network) relationships between activities. Gantt charts can be used to show current schedule status using percent-complete shadings and a vertical "TODAY" line as shown here.

The Bubble Chart visualizes the magnitude of data values as circular areas. Comparing one value to another might be harder at first glance in a bubble chart than in a bar chart, but bubble charts are good for giving an overall impression of a large number of inputs in a small space. Data can optionally be grouped and colored by secondary fields to convey more information in the same area.

Require path: splunk_wftoolkit/components/bubblechart/bubblechart

The Bubble Chart works best with data that has magnitude and is grouped by category. For example, Splunk searches that include a stats count by ... tend to work well.

Example Searches

```
index=_internal | head 10000 | stats count by sourcetype clientip | inputlookup faa.demo.csv | stats count by OriginCityName OriginState cancellation
```
The Force-Directed Graph displays information about connectivity and clustering between nodes in a network. This view displays nodes and their connections to other nodes and lays them out based on a physical simulation. Note that because the layout is based on a simulation, the final result is not guaranteed to be the same every time.

The Force-Directed Graph is best for visualizing interconnected data where clustering is interesting.

Example searches

| inputlookup faa.demo.csv | head 200 | stats count by OriginCityName, DestCityName, DestStateName | sort count

A tag cloud (word cloud, or weighted list in visual design) is a visual representation for text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text. Tags are usually single words, and the importance of each tag is shown with font size or color.[2] This format is useful for quickly perceiving the most prominent terms and for locating a term alphabetically to determine its relative prominence. When used as website navigation aids, the terms are hyperlinked to items associated with the tag.
Sunburst | This is a d3.js sunburst implementation. See http://bl.ocks.org/kerryrodden/7090426 for more information about the d3.js implementation.

A sequence sunburst may be useful to show the clickstream path for web statistics.
required inputs for the diagram are following two Splunk fields:

- path (in form of a string-sequence like "node1-node2-node3-node4")
- count (the sequence count)

Chart Overlay / Dual Axis
Specify fields to be displayed as a line chart on your charts. Perfect to show bounds for your data.

Chart Overlay Options:

```xml
<option name="charting.chart.overlayFields">average</option>
```
The Calendar Heatmap displays time-series data on a timeline as a heatmap. This view is good for visualizing time-based events and activity and can be configured extensively.

Require path: splunk_wftoolkit/components/calendarheatmap/calendarheatmap

The Calendar Heatmap works best with Splunk searches that use the timechart command.

The Calendar Heatmap displays a different timeline for each field in the results that are returned from a timechart search.

Example searches

```
index=_internal sourcetype=splunkd_access | head 2000 | timechart span=1m count by status
```

Punchcard shows a count relationship between two data dimensions.

Require path: splunk_wftoolkit/components/punchcard/punchcard

The Punchcard displays a data dimension counted by another data dimension, such as some event type by minute or hour, or events by host. The Punchcard works especially well with things counted by time.

Example searches

```
index=_internal | head 100000 | stats count by date_hour sourcetype
```
This example shows how to display data from the sample FAA dataset in a couple of NVD3 charts.
This example shows you how to plot geographical data on integrated maps.

Related examples:

Drilldown Maps
My Example: Image Overlay

This example shows you how to display workflow status visually using an image, and groups single value elements. I will use this example from the dashboard example app to create my own “image overlay” Executive Home page.

Related examples:

Single Value Element
Make Machine Data Accessible, Usable and Valuable to IT and Business Users
Insights Across Machine Data

Energy IT Data Sources
- WAM Systems
- SCADA / Sensor Data
- Smart Meters
- ETRM Systems
- EHS Systems

Core IT
- Web Services
- Devices
- Security
- Developers
- App Support
- Desktops
- Telecommunications
- Networking
- Messaging
- Storage

Customer-Facing Data
- Mobile Apps
- Customer Portals
- Clickstream
- Billing
- Social Media
- CDR

Data Sources
- ETRM Systems
- Social Media
- Intelligent Metering
- Smart Meters
- Cloud
- Virtual
Easy to Get Started

Download and install in minutes

1. Download
2. Eat your Machine Data
3. Start Splunking
Splunk is a 74.1 MB download. 1 product. 2 main components. Simple works. Simple handles 100TB+ a day real-time collection, analytics, reporting and automation of your critical business processes.
Interested in Getting Started with Splunk but don't want us meddling? We know that there a lot of you out there that like to take things into your own hands. If you have any issues at all please don't hesitate to contact me athutch@splunk.com 404-664-7009. Happy Splunking!

LMGTFY - Let me google that for you. This is great internet slang that is especially true with Splunk. If you have a question about something... don’t spend a lot of time figuring it out for yourself... don’t immediately call support... don’t stress... just GOOGLE it. Splunk has amazing online resources and an amazing community. There is a 90% chance you will find something to get you over the hump. If you can’t get there then reach out to me hutch@splunk.com.

Download Splunk:  
http://www.splunk.com/download

Getting Started Splunk Tutorial:  

Installation Guide:  

Getting Data into Splunk:  
http://docs.splunk.com/Documentation/Splunk/latest/Data/What-Splunk-can-monitor

Common Search Queries/Functions:  

Architecture Reference Hardware:  
http://docs.splunk.com/Documentation/Splunk/6.0.2/Deploy/Referencehardware

Splunk Architecture Images  
Images on Google

Pollution is Bad.  
http://blogs.splunk.com/2010/03/08/pollution-is-bad/

Basic Searching  
- Creating alerts  
- Creating and using event types  
- Saving and sharing searches  
- Creating Dashboards  
- and many more  

Splunk Search Cheat Sheet:  

Splunk Online Blogs:  
http://blogs.splunk.com/
Splunk Useful Links

Great “Go To” Links

The Splunk Book: Exploring Splunk - Search Processing Language (SPL) Primer and Cookbook
- [http://splunkbook.com](http://splunkbook.com) (Bought this myself on amazon and use it all the time!)

Dashboards and Visualizations

Forwarder Management (New)

Introduction to Pivot

Splunk Online Blogs:

Search Command Reference

All Search Commands

Security
[http://www.sans.edu/research/security-laboratory/article/sixtoplogcategories](http://www.sans.edu/research/security-laboratory/article/sixtoplogcategories)

Enterprise Security App

Advanced Persistent Threats

Hadoop (Hunk, Splunk for Hadoop, Splunk Enterprise integration with Hadoop)


Splunkbase!

David Carasso’s Personal Blog
[innovato.com](http://www.innovato.com)

Splunk Customers talking about how great we are!

Splunk Videos
[http://www.splunk.com/videos](http://www.splunk.com/videos)

Weekly Web Demo w/ Todd Gow
I also wanted to mention a few best practices when getting started.

1. **When first testing a new data source**, it is often easier to just have the file monitored on the local Splunk instance. However, eventually be sure to use the Universal Forwarder to collect the data.
   - [http://www.splunk.com/download/universalforwarder](http://www.splunk.com/download/universalforwarder)
   - [http://docs.splunk.com/Documentation/Splunk/5.0.1/Data/Usingforwardingagents](http://docs.splunk.com/Documentation/Splunk/5.0.1/Data/Usingforwardingagents)

2. **When testing inputs**, be sure to use a test index so you can easily clean the data if there is a problem. This link describes the process:

3. **The main things you want to make sure Splunk handles correctly at index time are:**

   **Event breaking** – Splunk often does this automatically, but it is best correctly detected at the beginning and end of an event. You can search the index and if you see that the count is different than what you expect. If it goes wrong see this:

   **Time stamp** – Splunk also generally detects the time stamp but if your log format has a strange time stamp you may need to manually configure this. When you load data, look at the events closely and make sure the event time is correct. If not, look at this:

   **Host** - Generally dedicated from the universal forwarder so really only a concern if you are using a syslog server with data from multiple hosts

   **Source type** - Splunk knows about many source types (i.e. access_combined, IIS, syslog, etc) but if you have a custom data source you should set this so that you can easily find it later on by doing queries like sourcetype=my_mobile_api etc...
   - [http://docs.splunk.com/Special:SplunkSearch/docs?q=sourcetype](http://docs.splunk.com/Special:SplunkSearch/docs?q=sourcetype)

   **Indexes Separation/Customizing** - Use a test index when starting out. Eventually, you will want separate indexes based on whether the data has specific retention requirements or if you want to prevent certain users from being able to query the data. Almost everything else can be dealt with after the data has been indexed during search time. This includes field extractions, tagging, and reporting.

   **Information** -
   - [http://docs.splunk.com/Documentation/Splunk/latest/Admin/Aboutmanagingindexes](http://docs.splunk.com/Documentation/Splunk/latest/Admin/Aboutmanagingindexes)
   - [http://docs.splunk.com/Documentation/Splunk/latest/admin/Setupmultipleindexes](http://docs.splunk.com/Documentation/Splunk/latest/admin/Setupmultipleindexes)