**Paper: Data Visualization: Past, Present & Future**

1. Giới thiệu:

Trực quan dữ liệu, cách dùng hình ảnh để biểu diễn thông tin, ngày nay đã trở nên phổ dụng bởi những lợi ích nó mang lại cho doanh nghiệp. Nó cung cấp một cách mạnh mẽ những khả năng về hiển thị cũng như cách truyền đạt những thông điệp. Mặc cho tiềm năng của nó, Trực quan dữ liệu vẫn chưa được đánh giá cao bởi sự thiếu hiểu biết về nó. Nhiều xu hướng ngày nay về Trực quan dữ liệu thực sự đã gây ra những tác dụng phụ mang tính tiêu cực, đó là sự nhầm lẫn thay vì sự thấu hiểu về Trực quan dữ liệu. Ngày nay trong lĩnh vực kinh doanh thông minh (Business intelligence) không có gì có thể mang chúng ta lại gần hơn với những hứa hẹn về sự tương tác thông minh hơn là Trực quan dữ liệu. Nhưng điều này chỉ xảy ra khi chúng ta thực sự hiểu và dùng nó một cách đúng đắn. Và chúng ta phải thực sự hành động và vứt bỏ những quan niệm chưa đúng về Trực quan dữ liệu.

1. Lịch sử của Trực quan dữ liệu:

Để hiểu về xu hướng hiện tại cũng như tương lai về Trực quan dữ liệu, sẽ tốt hơn nếu chúng ta biết về lịch sử của nó. Mặc dù những người tiền nhiệm đã biết về Trực quan dữ liệu từ những năm thế kỷ thứ 2 Sau Công Nguyên, nhưng phần lớn sự phát triển tập trung ở 2.5 thế kỷ cuối, chủ yếu trong 30 năm qua.

Bảng dữ liệu cổ nhất được bảo tồn cho tới ngày nay được tạo ra từ thế kỷ thứ 2 ở Ai Cập, nhằm sắp xếp các thông tin về thiên văn để chỉ đường. Bảng này chủ yếu biểu diễn dữ liệu dưới dạng văn bản, nhưng nó sử dụng các thuộc tính trực quan của các liên kết, khoảng trắng, và các luật về thời gian (các đường dọc và ngang) để sắp xếp dữ liệu thành các hàng và cột. Bảng (table), cùng với đồ thị (graph) và lược đồ (diagram) đều được xếp vào lớp biểu diễn dữ liệu gọi là biểu đồ (chart). Mặc dù các bảng chủ yếu là văn bản, nhưng sự sắp xếp dữ liệu thành các hàng, cột là một bước đầu mạnh mẽ đối với sự phát triển về sau, chuyển dần từ văn bản sang dạng trực quan trong biểu diễn dữ liệu.

Biểu diễn trực quan của các dữ liệu mang tính định lượng liên quan tới hệ toạ độ 2-chiều, là hình thức phổ biến nhất mà chúng ta thường gọi là đồ thị (graph), được tạo ra từ thế kỷ 17. Rene Descartes, nhà triết gia người Pháp hay được biết tới là một nhà toán học vĩ đại với câu nói “Tôi suy nghĩ, vì thế tôi tồn tại”, đã phát minh ra phương pháp này, ban đầu không phải dùng để trình bày dữ liệu mà dùng để biểu diễn các công thức toán học dựa trên hệ trục toạ độ. Tuy nhiên về sau, phương pháp này được công nhận như một cách hiệu quả để biểu diễn thông tin.

Sau phát kiến của Descartes, mãi cho đến cuối thế kỷ 18 và đầu thế kỷ 19, nhiều loại đồ thị mà chúng ta vẫn dùng ngày nay như biểu đồ cột (bar chart) và biểu đồ mảng (pie chart), mới được phát mình hay cải tiến đáng kể bởi một nhà khoa học xã hội người Scotland tên là William Playfair.

Hơn một thế kỷ trôi qua, tuy nhiên, trước khi giá trị của những kỹ thuật này được công nhận vào thời điểm đó, các khoá học về dữ liệu đồ hoạ (graphing data) đã được giới thiệu, đầu tiên tại trường Đại Học Bang Iowa vào năm 1913.

Người đã giới thiệu cho chúng ta về sức mạnh của Trực quan dữ liệu như một phương tiện để khám phá và tạo nên ý nghĩa của dữ liệu là giáo sư về thống kê John Tukey của Đại Học Princeton, người đã phát triển phương pháp hướng trực quan để khám phá và phân tích dữ liệu, được gọi là phân tích dữ liệu thăm dò (*exploratory data analysis*).

Năm 1983, một người đam mê Trực quan dữ liệu tên là Edward Tufte đã xuất bản cuốn sách mang tính đột phá “The Visual Display of Quantitative Information, đã cho chúng ta thấy nhiều cách hiệu quả để biểu diễn thông tin một cách trực quan và sau đó có những cách mà hầu hết chúng ta đều sử dụng, đáng buồn thay là các cách đó thiếu hiệu quả. Một năm sau, vào năm 1984, trong khi chúng ta đang xem Super Bowl, Apple Computer đã giới thệu chiếc máy tính đầu tiên phổ biến với giá phải chăng tập trung vào đồ hoạ như một phương thức tương tác và hiển thị. Điều này đã mở ra một cách sử dụng Trực quan dữ liệu mà chúng ta có thể xem và tương tác với chúng khi sử dụng máy tính.

Với sự sẵn có của những chiếc máy tính giá cả phải chăng cùng đồ hoạ mạnh mẽ, một nghiên cứu mới nổi lên như một đặc sản trong giới hàn lâm, được đặt tên là “Trực quan thông tin” (Information visualization). Năm 1999, cuốn sách “Readings in Information Visualization: Using Vision to Think” đã tập hợp tác phẩm này thành một khối và làm cho nó dễ dàng tiếp cận hơn, vượt ra ngoài bức tường học thuật.

Ngoài những cột mốc trong sự phát triển của Trực quan dữ liệu, một sự kiện khác trong nửa sau thế kỷ 20 đã ảnh hưởng lớn đến chất lượng của Trực quan dữ liệu, nhưng lại sai hướng đi: Sự tăng trưởng của PC hãng IBM. Trước khi máy tính cá nhân trở nên phổ biến ở nơi công sở, nếu bạn cần biểu diễn dữ liệu dưới dạng đồ thị, bạn phải đối mặt với một khối lượng lớn công việc liên quan đến hình vuông chữ T, các bản thảo, và một bộ các bút chì và bút bi. Thường thì việc này tốn hàng giờ để tạo ra một đồ thị để trình bày trong một cuộc họp hay để đính kèm vào một bản báo cáo. Khi quá trình này tốn nhiều thời gian và nỗ lực, để đảm bảo trách nhiệm trong công việc, thường một cá nhân phải tốn nhiều thời gian hơn để phát triển kĩ năng giao tiếp thông qua các đồ thị. Nhưng với sự ra đời của PC và bùng nổ của các phần mềm doanh nghiệp như bảng tính điện tử, điều này đã thay đổi. Với PC, các cú nhấp chuột có thể chuyển đổi một loạt các số liệu thành đồ thị, và những người không biết gì về thiết kế đồ thị cũng có thể trở thành các Rembrandt của giao tiếp đồ thị. Mặc cho những nỗ lực của Edward Tufte từ những năm 1980, chất lượng của Trực quan dữ liệu đã bị bỏ ngỏ nhiều, đặc biệt trong hình thái đồ thị dùng cho doanh nghiệp, mặc cho tiềm năng phát triển của nó.

Dựa trên bối cảnh lịch sử như trên, chúng ta hãy nhìn vào những gì đang tiếp diễn ngày nay.

1. Những xu hướng hiện tại trong Trực quan dữ liệu:

Ngày nay, Trực quan dữ liệu ngày càng đóng vai trò quan trọng trong kinh doanh thông minh. Bao gồm, nghiên cứu, yêu cầu bởi những người làm việc với dữ liệu, mua bởi những người nắm giữ hầu bao, và dùng bởi một tỷ lệ ngày càng tăng trong lực lượng lao động trí óc, đặc biệt là các nhà phân tích. Đó là những tin mang tính khởi sắc. Nhưng tin xấu là, trong giới doanh nghiệp, Trực quan dữ liệu vẫn còn bị bỏ ngỏ, hiểu nhầm, dùng chưa hiệu quả, và thường bị làm lệch đi bởi các nhà cung cấp sản xuất và bán phần mềm trực quan.

* 1. Những hướng đi tích cực:

Trực quan dữ liệu đã trở thành một lĩnh vực nghiên cứu trong những năm gần đây. Nhiều trường đại học đã mở các Khoa chyên nghiên cứu về Trực quan và có một số chương trình ưu việt phục vụ nhu cầu nghiên cứu và chế tạo các nguyên mẫu của sinh viên. Cộng đồng nghiên cứu này bao gồm các cá nhân đến từ nhiều lĩnh vực không chỉ riêng Khoa học Máy tính, như tâm thần học và thậm chí các doanh nghiệp, nơi tạo môi trường lý tưởng để phát triển lý thuyết Trực quan một cách thực tế và mang tính cách mạng nhất.

Chúng ta đang bắt đầu nhận thấy các sản phẩm Trực quan dữ liệu đang thực sự tiện ích như thế nào. Tuy vẫn đang là thiểu số nhưng là một thiểu số đang lớn mạnh. Hầu hết các phần mềm Trực quan dữ liệu thương mại đều xuất phát từ các nghiên cứu. Các nỗ lực nhằm dần thu hẹp khoảng cách giữa các nhà nghiên cứu với các ý tưởng tuyệt vời và các nhà sản xuất kinh doanh với khả năng xây dựng và phát triển các ý tưởng này thành các sản phẩm thương mại.

Một trong những xu hướng mới đáng khích lệ trong kinh doanh thông minh là sự gia tăng trong việc công nhận những lợi ích đáng kể do Trực quan dữ liệu trong hình thức phân tích (form of analytics) mang lại. Phần mềm phân tích trực quan cho phép chúng ta không chỉ biểu diễn dữ liệu dưới dạng đồ thì, mà còn có thể tương tác với chúng để thay đổi tính tự nhiên vốn có, lọc ra những gì không liên quan, đi sâu vào chi tiết hơn, làm nổi bật các tập con dữ liệu trên nhiều đồ thị cùng một lúc. Điều này hiệu quả hơn đối với não bộ và mắt của chúng ta, hơn là cách tiếp cận truyền thống. Những đồ thị tĩnh (static graphic) trên giấy, hay hiển thị trên màn hình máy tính giúp chúng ta truyền đạt thông tin một cách rõ ràng và sáng tỏ, đó là một lợi ích mà ta không nên đánh giá thấp, đặc biệt là các nhà doanh nghiệp.

Một trong những kỹ thuật mạnh mẽ nhất của phân tích trực quan liên quan đến biểu diễn nhiều đồ thị đồng thời, có chức năng biểu diễn các tập dữ liệu con khác nhau trích xuất từ tập dữ liệu lớn hơn, hoặc những góc nhìn khác nhau về một tập dữ liệu chia sẻ. Edward Tufte đã phổ biến một hình thức biểu diễn mà ông gọi là bội số nhỏ (small multiples), dùng một loạt các đồ thị nhỏ được sắp xếp cùng nhau trong một nháy mặt để có thể so sánh. Mỗi đồ thị đại diện cho một tập dữ liệu con khác nhau thuộc về một tập dữ liệu đầy đủ, chẳng hạn một loạt các đồ thị dòng (line graphs) biểu diễn chi phí của một công ty qua thời gian, với một đồ thị riêng biệt cho mỗi bộ phận. Bội số nhỏ mở rộng số lượng các biến (số chiều) để xem xét và so sánh. Một cách tiếp cận khác trong biểu diễn đồng bộ đa đồ thị là kiểm tra từng đồ thi để phân tích các khía cạnh của một tập dữ liệu phổ biến. Ví dụ, một số đồ thị, với nhiều loại khác nhau (đồ thị thanh, đồ thị dòng, đồ thị phân tán,..) có thể được biểu diễn cùng nhau để xem xét một số khía cạnh của tập dữ liệu, cho phép chúng ta khám phá các kết nối trong dữ liệu mà có thể không bao giờ xuất hiện nếu biểu diễn riêng lẻ. Các sản phẩm phân tích trực quan hỗ trợ hiển thị như vậy đang được công nhận nhanh chóng bởi các nhà phân tích.

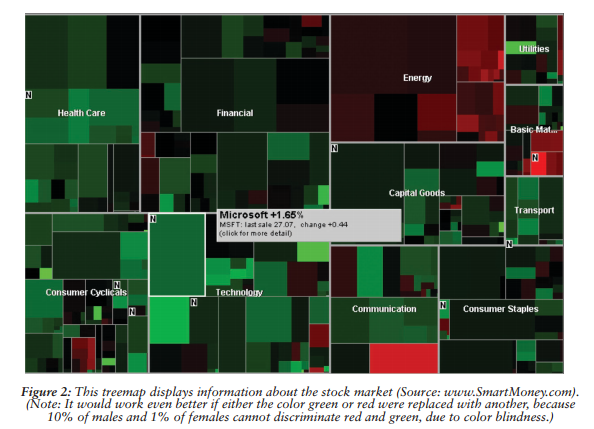
Bên cạnh sự phổ biến của phân tích trực quan, thì không thể bỏ qua ý nghĩa của việc dùng các đồ thị truyền thống để trao đổi thông tin. Khi có một vấn đề liên quan tới dữ liệu cần trình bày cho đối tác, việc biểu diễn trực quan như đồ thị hay lược đồ thường là phương tiện tối ưu nhất, nhưng chỉ khi bạn hiểu ngôn ngữ đó. Giao tiếp trực quan liên quan đến cả ngữ nghĩa lẫn cú pháp, giống như ngôn ngữ nói vậy. Bạn phải nắm các quy tắc đồ thị để trình bày hiệu quả. Ngày nay, qua những nghiên cứu tiên phong của Edward Tufte và William Cleveland từ những năm 1980, và gần đây là những nỗ lực của Gene Zelazny, Naomi Robbins, và bản thân tôi (Stephen Few), việc nhận thức phải nắm bắt những kỹ năng trong giao tiếp thông qua đồ thị phải được giáo dục. Tôi tin rằng những kỹ năng này khá dễ dàng để tìm hiểu, nhưng việc này không đòi hỏi phải có góc nhìn trực quan, mà yêu cầu những nỗ lực và nguồn lực đúng đắn.

Không có ví dụ về Trực quan dữ liệu chiếm vị trí nổi bật hơn trong suy nghĩ của các nhà kinh doanh ngày nay bằng trang tổng quan (dashboard). Những màn hình kết hợp các thông tin cần thiết một cách nhanh chóng để theo dõi một khía cạnh của doanh nghiệp trên một màn hình duy nhất, là những bổ sung mạnh mẽ cho các doanh nghiệp thông minh. Khi thiết kế phù hợp cho việc giao tiếp trực quan hiệu quả, bàng điều khiển hỗ trợ một mức độ trong nhận thức – bức tranh biểu diễn những vấn đề đang xảy ra – mà không bao giờ có thể thể hiện thông qua kiểu báo cáo truyền thống. Không may các sản phẩm cũng như các nhà sản xuất và phân phối các thiết bị trang tổng quan hầu hết đều không tận dụng được tối đa thế mạnh của Trực quan dữ liệu. Thay vào đó, các trang tổng quan có xu hướng trông giống về bề ngoài lẫn chức năng như những trò chơi điện tử hơn là những màn hình biểu diễn thông tin. Thực tế, nhiều sản phẩm trang tổng quan, trong khi phải nâng cao tầm nhìn của Trực quan dữ liệu thì nó lại đưa ra những mặt xấu bởi các thiết kế kém.

Một biểu hiện của Trực quan dữ liệu đã nắm lấy trí tưởng tượng của nhiều người trong giới kinh doanh nhiều năm gần đây, đó là trực quan không gian địa lý (geo-sptial visualization). Sự phổ biến của Google Earth và các dịch vụ Web tương tự đã đóng góp rất nhiều cho lĩnh vực này. Phần lớn các thông tin mà doanh nghiệp phải theo dõi và nắm bắt được gắn với vị trí địa lý. Ví dụ, đôi khi thông tin giảm giá chỉ có thể được biết tới nếu bạn biết nơi đang bán các mặt hàng đó. Trong những trường hợp này, khả năng ước lượng doanh thu bán hàng trên bản đồ cho ta thêm một góc nhìn để hiểu rõ hơn tầm quan trọng của nó. Khả năng tận dụng lợi thế về thông tin địa lý thực ra đã có sẵn trong hệ thống của bạn, chẳng hạn mã Zip của khách hàng, dùng để hiện thị các thông tin liên quan như các đợt giảm giá trên bản đồ ngày càng trở nên cần thiết hơn trong các phần mềm kinh doanh thông minh, và tốt hơn là được tích hợp các báo cáo và phân tích tổng quan mỗi ngày.

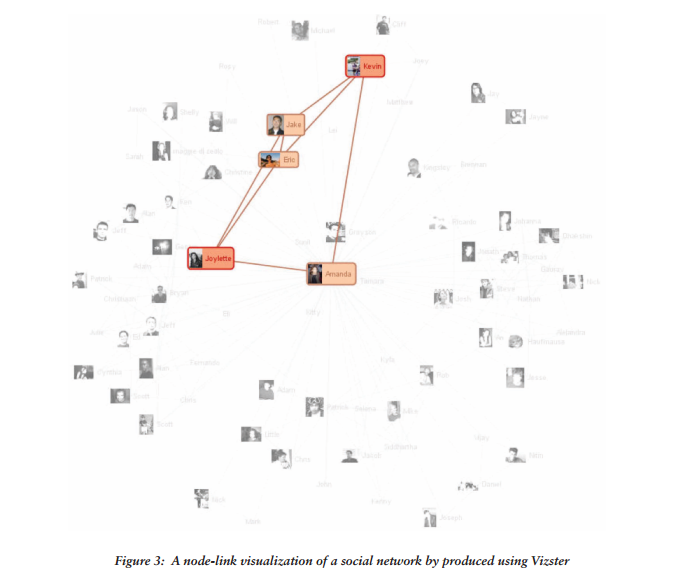
Một xu hướng khác đang bắt đầu tìm đường tiếp cận lĩnh vực ứng dụng kinh doanh thông minh liên quan đến việc sử dụng các hình ảnh động trực quan (visual animation) – sự chuyển động của các đối tượng trong đồ thị - để thể hiện sự thay đổi theo thời gian. Trước đây chúng ta thường thị đồ thị dòng để thể hiện sự thay đổi theo thời gian. Việc này mang tính ưu việt đối với các thông tin phụ thuộc về thời gian, như ước lượng lưu lượng truy câp Web ở các khoảng thời gian như nhau, chẳng hạn như hàng ngày. Nhưng nếu bạn muốn khảo sát một mối quan hệ khác nhau giữa các giá trị, và đồng thời kiểm chứng sự biến thiên của chúng theo thời gian thì làm thế nào ?

Hãy xem xét mối tương quan giữa chi phí tiếp thị và kết quả thu được từ việc bán hàng. Cách tốt nhất để khảo sát sự tương quan này nằm ở một điểm cụ thể theo thời gian thể hiện bởi biểu đồ phân tán (scatter plot), với chi phí tiếp thị thể hiện bằng trục X, lợi nhuận bán hàng thể hiện bằng trục Y và các điểm riêng biệt thể hiện cho từng hàng hoá khác nhau với tổng tất cả có 50 điểm. Nếu chúng ta muốn xem bản chất của mối tương quan này thay đổi theo thời gian, thì việc dùng đồ thị dòng sẽ không hiệu quả, vậy chúng ta phải biểu diễn như thế nào ? Câu trả lời là chuyển động đồ thị phân tán, cho phép các điểm dữ liệu biểu diễn 2 thông tin trên ứng với các trạng thái khác nhau dịch chuyển bên trong đồ thị để chỉ ra sự thay đổi theo thời gian của chúng như thế nào. Một vài thí dụ điển hình được phát triển bởi [www.GapMinder.org](http://www.GapMinder.org), website này đã sử dụng công nghệ trên để biểu diễn dữ liệu quan trọng tầm quốc tế, như mối quan hệ giữa thu nhập của một quốc gia và tỷ lệ tử vong trẻ sơ sinh, và cách thế giới thay đổi ở lĩnh vực này trong 30 năm qua.

Một xu hướng khác đã được tạo ra một cuộc hành trình trong những năm gần đây, chuyển từ các công trình nghiên cứu thành các sản phẩm thương mại, giải quyết vấn đề biểu diễn dữ liệu định lượng của các tập hợp lớn trong một màn hình giới hạn về không gian. Ví dụ phổ biến nhất là bản đồ cây (treemap), lần đầu tạo ra bởi Ben Shneiderman của Đại Học Maryland. Treemap được thiết kế để hiển thị lên tới 2 biến định lượng khác nhau ở những mức độ khác nhau của cây phân cấp. Ví dụ, có thể bạn quan tâm đến lượng cổ phiếu đang được giao dịch trên sàn chứng khoán New York, treemap cho phép bạn so sánh giá cả cũng như biến thiên trong giá với ngày hôm qua của các cổ phiếu. Đó thực sự là một lượng dữ liệu khổng lồ. Bạn có thể thử sử dụng 2 biểu đồ thanh ngang (horizontal bar graph) được sắp xếp cạnh nhau, một cho giá cổ phiếu và một cho sự biến thiên về giá, nhưng bạn sẽ nhanh chóng chạm tới giới hạn của số thanh được biểu diễn trên biểu đồ. Treemap cung cấp một phương tiện tối đa hoá số lượng thông tin được hiển thị trên màn hình, lấp đầy khoảng không với các thông tin.

Trong hình 2, có thể thấy một hình chữ nhật cho mỗi cổ phiếu, được xếp trong hình chữ nhật lớn hơn, gom thành các ngành công nghiệp. Kích thước (vùng 2 chiều) của mỗi hình chữ nhật đại diện cho giá của ngành đó và sự thay đổi giá được đại diện bởi các màu, với các giá trị âm khoảng màu từ sáng tới đỏ đậm, giá trị dương khoảng màu từ áng đến xanh lá đậm, và màu đen cho giá trị không đổi. Treemap không có nghĩa là so sánh chính xác về giá trị, không thể dùng màu sắc cũng như hình 2 chiều để mã hoá giá cả, nhưng có thể nhanh chóng tìm ra lượng thông tin lớn hay các xu hướng đang chiếm ưu thế (ví dụ trong hình thì sự thay đổi tích cực nhất là ở mảng công nghệ). Khi sử dụng cho mục đích này, treemap vẫn là thành viên mới trong lĩnh vực phần mềm Trực quan dữ liệu nhưng đã tạo ra một bước đi đáng kể.

Xu hướng cuối cùng được tạo ra từ các phòng nghiên cứu khoá học đã chuyển thành các phần mềm thương mại có hơi khác so với sự Trực quan dữ liệu mà tôi đề cập đến bởi vì nó tập trung nhiều hơn vào biểu diễn mối quan hệ giữa các thực thể (ví dụ như các công ty hay Websites) hơn là các giá trị định lượng. Các mối quan hệ giữa những điều làm chúng ta quan tâm kinh doanh có thể được miêu tả như một mạng lưới, với các liên kết kết nối các đối tượng trong một sự sắp xếp phức tạp. Ví dụ, Internet bao gồm các Websites hình thành nên một mạng lưới kết nối đa dạng và phức tạp.

Hình 3 thể hiện Internet dưới dạng nút và các liên kết trực quan. Mỗi nút là một Website và các đường kết nối nằm giữa chúng. Đường càng nhỏ, số lượng liên kết càng lớn. Bởi sự trực quan này được nhìn từ màn hình máy tính và được thiết kế để có thể tương tác, bạn có thể dễ dàng tập trung vào các từng phần trong cái toàn phần để xem xét một cách chi tiết hơn. Ngoài ra có thể lọc ra các Website không mong muốn dựa trên một số thuộc tính thiết lập sẵn như quốc gia,..

Mọi người trong công ty cũng có thể liên kết với nhau trong một mạng lưới kết nối phức tạp, có thể được khảo sát dựa trên email. Bên cạnh trực quan mạng lưới Internet, trực quan mạng lưới đã tạo ra một bước tiến trong biểu diễn mối quan hệ giữa người và người, được sử dụng để nghiên cứu các mạng xã hội. Điều này đặc biệt được quan tâm bởi các nhà xã hội học, nhưng các ứng dụng liên quan đến doanh nghiệp cũng bắt đầu xuất hiện.Hình 3 cho thấy một mạng lưới bạn bè.Mỗi người được đánh dấu riêng biệt, vì mỗi 2 người được chọn sẽ nổi bật ra những bạn bè có điểm chung giữa họ thông qua một tính năng của ứng dụng.

Chúng ta đã xem qua một số hướng đi tích cực trong Trực quan dữ liệu, bên cạnh đó vẫn còn một số mặt không được tốt đáng xem xét.

* 1. Các hướng đi tiêu cực:

Bất cứ khi nào một xu hướng trong công nghệ thông tin nắm bắt được sự quan tâm của mọi người và trở nên phổ biến, cùng theo đó cũng sẽ có một lượng lớn những người vội chạy theo xu hướng ấy nhưng với vốn hiểu biết rất ít về nó là gì hay nó hoạt động như thế nào. Nhiểu nhà sản xuất đã chạy theo xu hướng Trực quan dữ liệu nhằm thương mại hoá nó nhưng lại không bỏ thời gian để nghiên cứu hay phát triển nó một cách đúng đắn. Các chiến lược kinh doanh thỉnh thoảng mang đến những điều tồi tệ cho Trực quan dữ liệu, thay thế bằng đèn chiếu hay mê hoặc người dùng bằng những chức năng màu mè kém hiệu quả, và cuối cùng đem đến cho Trực quan dữ liệu một cái nhìn tệ hại khi người dùng cố gắng sử dụng và thất bại. Trang tổng quan là một ví dụ điển hình. Hầu hết các sản phẩm đều trông rất thu hút nhưng lại làm việc rất kém. Và bởi sự phổ biến của nó, mỗi nhà cung cấp gọi bất cứ thứ gì mà họ bán có kết hợp nhiều đồ thị trên một màn hình đều là Trang tổng quan. Sự lạm dụng thuật ngữ này đã cướp đi của Trang tông quan một định nghĩa rõ ràng. Một thiết kế chính xác cho Trang tông quan nhằm miêu tả những gì đang xảy ra, hơi khác so với thiết kế một bộ sưu tập đặc biệt các đồ thị có thể xem hay tương tác đồng thời trên màn hình máy tính để phân tích dữ liêu.

Trực quan dữ liệu không phải là làm cho mọi thứ trông dễ thương hay đẹp đẽ hơn. Nó cũng không trang hoàng cho bài thuyết trình của bạn nhằm mê hoặc người xem. Kiểu trực quan này được đặt tên là “thị trường” (marketing). Trực quan dữ liệu chỉ bao hàm một ít định nghĩa về nghệ thuật, mặc dù tính thẩm mỹ có liên quan tới việc trình bày dữ liệu một cách hiệu quả. Thay vào đó, nó liên quan nhiều hơn đến khoa học, tập hợp các quy tắc dựa trên những gì chúng ta biết về nhận thức và sự liên hệ một cách trực quan, có thể làm theo để hiển thị thông tin hiệu quả. Khi những luật này được hiểu rõ và các phần mềm được thiết kế hỗ trợ chúng, và chỉ như thế, những điều tuyệt vời với Trực quan dữ liệu mới có thể xảy ra.

1. Hướng đi trong tương lai của Trực quan dữ liệu:

Tôi sẽ bắt đầu phần này với những bộc lộ về hy vọng ở tương lai, hơn là dự đoán điều gì sẽ xảy ra. Những tiến bộ hiện đại của Trực quan dữ liệu bắt nguồn từ các nghiên cứu khoa học, mà nguồn gốc sơ khai từ những nghiên cứu về nhận thức tri giá và con người (visual perception and human cognition). Những nghiên cứu này đã khai phá khả năng và giới hạn của cả việc tạo ra các phương thức Trực quan dữ liệu và các ứng dụng nhằm đạt được tối ưu tiềm năng của chúng nhằm khắc phục các hạn chế mà chúng ta mắc phải. Như vậy, Trực quan dữ liệu có thể được xem đảm nhận vai trò trung tâm trong kinh doanh thông minh, được tạo ra để hỗ trợ.

Thực tế này là hiển nhiên đối với bất kỳ ai đã dành thời gian để tìm hiểu nhu cầu thực sự cũng như khả năng của các doanh nghiệp những người thưởng xuyên làm việc với dữ liệu và xem nó như mục tiêu của kinh doanh thông minh. Để đạt được những nguyện vọng trên, chúng ta phải thay đổi sự tập trung từ công nghệ sang con người, từ máy móc hỗ trợ sang những người nắm được cốt lõi để dựa vào đó đạt được thành công. Với sự thay đổi này, chúng ta sẽ thấy các phần mềm kinh doanh thông minh dần được lược bỏ các chức năng dư thừa và được thiết kế để dần ẩn đi, che mờ đi trong ý thức và khi đó tâm trí ta có thể tập trung hoàn toàn vào việc khám phá dữ liệu. Đó chính là tương lai mà chúng ta kỳ vọng và phải đấu tranh để đạt được.

Cộng đồng nghiên cứu Trực quan dữ liệu và các nhà cung cấp phần mềm thương mại đã tham gia vào cuộc đua đầy năng động và sáng tạo. Phần lớn đều bị phớt lờ bởi các nhà cung cấp, những nhà nghiên cứu thường dành nhiều thời gian theo đuổi các dự án có rất ít hay thậm chí không có tính ứng dụng, hay thoả mãn nhu cầu thực sự của các nhà doanh nghiệp. Nếu không đối diện với nhu cầu của những người thực sự cần, vậy tại sao các nhà nghiên lại theo đuổi những gì thu hút họ mà không có bất kỳ mục tiêu nào ngoài khám phá? Điều này phụ thuộc vào chúng ta, những người thực sự làm việc với các thông tin để thấu hiểu nó và trao đổi những phát hiện của mình đối với người khác để vượt qua trở ngại này bằng cách lắng nghe các nhu cầu thự sự. Đối với các nhà cung cấp chúng ta phải rõ ràng một điều rằng chúng ta không tạo ra các phần mềm kém hiệu quả, ít tập trung, và khó sử dụng. Chúng ta phải ngăn chặn việc ảnh hưởng của các tính năng bề nổi, thiếu sâu sắc mà không mang lại nhiều giá trị thực sự. Khi làm điều này, các nhà cung cấp sẽ chuyển sự chú ý sang các cộng đồng nghiên cứu để tìm nguồn cảm hứng và sự tập trung của họ sẽ chuyển thành các hành động đáp ứng được nhu cầu thực sự.

Trong nhiều vấn đề liên quan đến công nghệ, chính phủ có vẻ tụt hậu dần so với công nghiệp, ngoại trừ việc lãnh đạo. Từ năm 2005, một chương trình từ chính phủ Mỹ liên quan đến Trung tâm Phân tích Trực quan Quốc gia (National Visual Analytics Centers NVACs) đã thành lập các trung tâm tại nhiều trường đại học để tìm hiểu và phát triển các cách để áp dụng Trực quan dữ liệu nhằm phân tích các nhu cầu từ các cơ quan chinh phủ, đặc biệt là các cơ quan tình báo. Những trung tâm này hợp tác với các nhà cung cấp phần mềm thương mại nhằm mang các ý tưởng trở thành hiện thực. Đó là thời điểm giới kinh doanh nhận ra sự chậm trễ và phải nỗ lực để tạo tiếng nói và tầm vóc cho nó.

Tôi hy vọng Trực quan dữ liệu sẽ tiếp tục thu hút và dần trưởng thành hơn so với các xu hướng trước đây. Trang tổng quan, phân tích trực quan, và thậm chí cả những đồ thị đơn giản sẽ tiếp tục phát triển và phù hợp so với thực tế. Tôi cũng được chứng kiến những bằng chứng chỉ ra những nỗ lực mới nổi sẽ sớm phát triển theo hướng toàn diện.

Một trong những nỗ lực liên quan đến Trang tổng quan trên các màn hình nhỏ hơn. Như một công cụ để kiểm soát các vấn đề đang xảy ra, Trang tổng quan thường luôn sẵn sàng dù bạn không ngồi trước màn hình máy tính, điều này có nghĩa là bạn luôn có thể truy cập từ các thiết bị di động như điện thoại hay PDA. Giới hạn về diện tích màn hình càng lớn, các thiết kế dành cho Trang tổng quan sẽ càng tốt hơn, thậm chị sẽ giúp các nghiên cứu theo chiều hướng hiệu quả hơn để hiển thị tối đa nội dung trong một không gian nhỏ. Các nhóm nghiên cứu đang thực sự hiện thực điều này.

Đối lập với các yếu tố thu hẹp là các yếu tố mở rộng: không gian xung quanh, chẳng hạn như các phòng trang bị video nhằm mở rộng số lượng thông tin có thể truy cập để kiểm soát và phân tích. Nếu bạn từng xem bộ phim khoa học viễn tưởng Minority Report, có lẽ bạn khá ấn tượng khi Tom Cruise đeo một chiếc tai nghe thực tế ảo (virtual reality) và sau đó thấy một màn hình lớn, giúp anh ta có thể thao tác với các thông tin thông qua suy nghĩ. Việc tạo ra các môi trường đa dạng đang trở thành hiện thực, một số cảnh trong bộ phim này có thể không lâu nữa sẽ xuất hiện ngoài đời thực. Nhằm mục đích tiếp cận thông tin dễ dàng hơn, so sánh và tương tác nhanh hơn để hiểu được và sau đó đưa ra quyết định tốt hơn.

Nhiều tháng trước, tôi được xem một bộ phim, trong đó thể hiện cách con người hợp tác để phân tích thông tin trong tương lai. Bộ phim dựa trên sự đóng góp của nhiều chuyên gia về thông tin, bao gồm các chuyên gia trong lĩnh vực Trực quan dữ liệu. Nó gợi tôi nhớ về Minority Report. Mọi người mặc những bộ cảm biến theo dõi các khía cạnh sinh lý của họ để phát hiện stress, hệ thống cho phép tự động điều chỉnh để hiển thị dữ liệu trên màn hình sao cho việc nắm bắt và xử lý của họ là đơn giản nhất. Một nhóm các nhà phân tích ngồi quanh bàn tròn đối điện với các màn hình cá nhân trong suốt. Bên cạnh dự án này, một nghiên cứu đang nổi lên nhằm tìm ra cách giúp con người hợp tác trong việc thăm dò và tạo ra ý nghĩa cho các thông tin. Hiện tại một nhóm các chuyên gia về Trực quan dữ liệu tại IBM đang phát triển nghiên cứu này. Sự hợp tác, hiển nhiên không yêu cầu tất cả mọi người phải ngồi chung trong một căn phòng. Công nghệ cho phép chúng ta có thể chia sẻ thông tin, thảo luận và đưa ra quyết định kể cả khi chúng ta ở khắp mọi nơi trên thế giới.

Một trong những vấn đề phiền nhiễu việc khám phá và phân tích dữ liệu liên quan đến lượng lớn dữ liệu khó có thể trực quan theo nhiều cách nhằm che đi những gì lộn xộn đằng sau nó. Đây là vấn đề nhận được nhiều sự chú ý từ giới nghiên cứu. Các phương pháp và giải thuật giả lập được đưa ra để hạn chế số lượng dữ liệu bằng cách giảm đi số lượng nhưng không ảnh hưởng đến ý nghĩa, hay giảm đi sự rối rắm khi thể hiện trực quan thông qua phương pháp sắp đặt các đối tượng dữ liệu, sử dụng màu sắc, hay các thuộc tính trực quan khác như độ trong suốt. Công trình này thực sự lý tưởng và có thể thương mại hoá.