



Visualizações De Dados Disponíveis Na Internet e Novas Possibilidades Na Comunicação¹

Matias PERUYERA²
Adriana AMARAL³

Faculdade Internacional de Curitiba – Facinter - PR

Resumo

A internet, especialmente desde o advento da chamada Web 2.0, tornou-se uma fonte gigante de dados criados por usuários, que podem ser analisados e cruzados de diversas maneiras. O grande volume de dados faz com que a visualização através de recursos gráficos seja o melhor, senão o único meio para sua análise. Aqui, é analisado o trabalho do programador Jim Bumgardner, que apesar de seu viés artístico, reúne várias características fundamentais para trabalhos científicos ou jornalísticos.

Palavras-chave: visualização de dados, infografia, estatística

Introdução

A chamada Web 2.0 se caracteriza por ser um espaço onde é o usuário quem cria e alimenta os sites com seu conteúdo. Essa interação gera bancos de dados gigantescos com as informações vindas do conteúdo do usuário, seja no conteúdo em si ou em metadados⁴ embutidos nos arquivos enviados.

Muitos desses sites abrem seu conteúdo para que programadores possam desenvolver aplicativos que usem os bancos de dados. Geralmente são aplicativos que acrescentam novas funcionalidades aos sites, mas alguns programadores têm aproveitado a quantidade de dados como ponto de partida para transformá-los em visualizações⁵ através das quais é possível ver novas informações.

Este trabalho trata brevemente da representação gráfica, ou visualização, de dados disponíveis através dos sites, e sua influência sobre a infografia jornalística e a arte.

¹ Trabalho submetido ao XVII Intercom Júnior 2010 na categoria IJ 8 – Estudos Interdisciplinares da Comunicação.

² Estudante de graduação do 3º. semestre do Curso de Comunicação Social - Jornalismo da Facinter, Curitiba, email: matiasperuyera@gmail.com

³ Orientadora do trabalho e professora do curso de Comunicação Social da Facinter – email: adriamaral@yahoo.com

⁴ Dados embutidos em arquivos digitais, referentes ao próprio arquivo, como informações sobre a exposição de uma fotografia ou o artista de um arquivo de música.

⁵ O termo visualização é usado neste artigo como sinônimo de gráfico, já que é o termo geralmente usado para se referir a gráficos de alta densidade de dados.



Para isto, será analisado as visualizações produzidas pelo programador e artista Jim Bumgardner.

Aprendendo a linguagem

Como afirmam Unwin, Theus e Hofmann, “Visualização de dados é boa para limpeza dos dados, para explorá-los, para identificar tendências e agrupamentos, para encontrar padrões locais, para avaliar modelos de representação de gráficos e para apresentar resultados.”⁶



Fig. 1: Gráficos do site GraphJam.com, que permite aos usuários criar gráficos, geralmente com sentido humorístico.

Essas qualidades estão presentes até em gráficos mais simples, como os usados em infografia jornalística, que seguem códigos já conhecidos. Por exemplo, a densidade demográfica de um país costuma ser representada com mapas coloridos que seguem uma escala de cores associada a valores; alguns dados econômicos, como pontuação de bolsas, são representados por gráficos de linhas que mostram sua evolução ao longo do tempo; pesquisas de opinião usam gráficos de pizza. A linguagem dos gráficos se tornou comum a ponto de poder ser parodiada (Fig. 1).

Dados mais incomuns exigem novas formas de visualização, possivelmente combinando duas ou mais formas já conhecidas. Por exemplo, grandes volumes de informação, como as informações vindas de institutos de estatística de governos. Gráficos onde cada elemento tem mais de um atributo que representa uma informação:

6 UNWIN, THEUS E HOFMANN, Graphics of Large Datasets, pág. 2.

por exemplo, seu tamanho representa uma variável, sua posição outra variável e sua cor uma terceira.

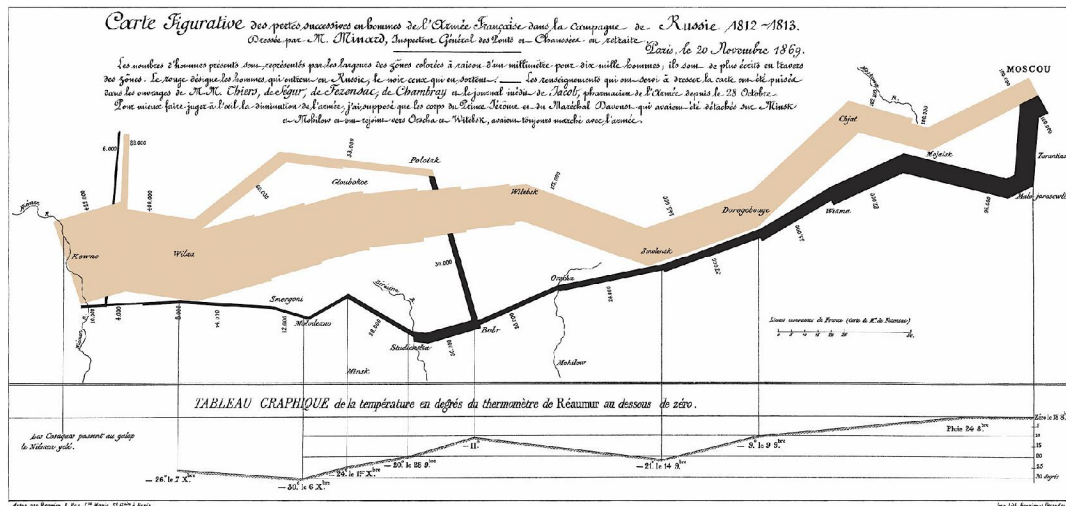


Fig. 2: Gráfico sobre as tropas de Napoleão durante a invasão da Rússia, feito por Minard em 1869.

Um exemplo antigo de gráfico complexo é o que narra o caminho e as baixas sofridas pelas tropas de Napoleão durante a tentativa de invadir a Rússia, feito por Minard em 1869 (Fig. 2). Ele mostra várias informações ao mesmo tempo. A ida é representada em bege, a volta em preto. A largura da linha representa o número de homens do contingente. O caminho deles é representado de maneira esquemática, assim como as bifurcações no caminho. A temperatura enfrentada pelas tropas no caminho de volta é mostrada em um gráfico separado, mas relacionado com o caminho das tropas. Este gráfico é apontado por Edward Tufte como o melhor gráfico estatístico da história: “Minard’s graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time.”⁷

Enquanto o mapa de Minard guarda certa relação com a realidade, outro tipo de dados exige uma abstração maior. Exemplo disso são os “Butterfly Graphs” (Fig. 3), feitos pela primeira vez pelo astrônomo inglês Edward W. Maudner em 1904, nos quais ele organizou informação sobre ciclos de manchas solares observadas em diferentes latitudes. Esse tipo de gráfico foi aperfeiçoada com o uso de cores e associando o gráfico a outro gráfico que mostra a porcentagem do sol coberta por manchas solares, resultando em um gráfico de densidade alta que, apesar de precisar de um conhecimento

⁷ TUFTE, The Visual Display of Quantitative Information, pág. 40



prévio para ser lido, permite várias leituras que não seriam possíveis se a informação não estivesse registrada graficamente.

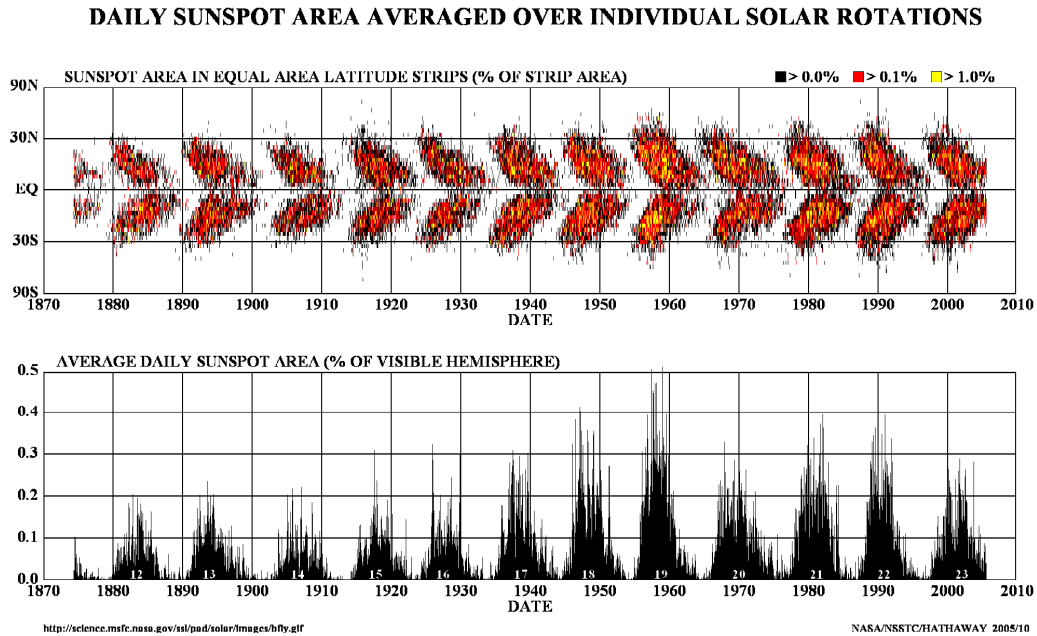


Fig. 3: “Butterfly Graphs” modernos, em uma releitura dos criados por Maudner.

Abundância de dados na web 2.0

A abertura para a participação do público é a principal inovação da chamada Web 2.0, como afirma Henrique Antoun (2008). Alguns sites, além de permitir upload de arquivos, permitem que programadores tenham acesso a suas bases de dados através da abertura das APIs,⁸ o que permite que programadores criem suas próprias formas de visualizar o conteúdo do site. Exemplos disso são o BigHugeLabs⁹, que é uma interface alternativa que também oferece algumas ferramentas para o conteúdo do Flickr, usando os metadados das fotografias, que tem informações de abertura e exposição da foto, geolocalização, nome do dono da câmera; e informações acrescentadas posteriormente, como tags e comentários sobre as fotos; além das informações da própria fotografia, como cores e sua distribuição. Ou os vários sites que analisam perfis de usuários do Twitter acessando diretamente a base de dados dos *tweets*¹⁰ dos usuários.

Considerando o número gigantesco de usuários que alimentam as bases de dados de

⁸ Interface de Programação de Aplicativos, na sigla em inglês.

⁹ www.bighugelabs.com

¹⁰ Termo que se refere a cada um dos textos postados no Twitter.



sites como Twitter¹¹ e Flickr¹², temos um universo de informações que ainda não foram devidamente exploradas. Por outra parte, o volume de informação é gigantesco, e a forma mais adequada de gerar novas informações a partir desses dados é a visualização através de gráficos.



Fig. 4: visualização realizada por Jer Thorp, mapeando os “tweets” dizendo “Good Morning”.

Talvez por não encontrar utilidade nesses dados, a maioria dos projetos de visualização caminha entre a paródia e a análise do site, como o trabalho de Jer Thorp, que mapeou os twits dizendo “bom dia” (Fig. 4)¹³. Esse trabalho parodia o uso dado ao Twitter, e também questiona a qualidade do conteúdo gerado e classificado pelos usuários.

Por exemplo, as limitações no uso de tags ou etiquetas¹⁴ apontadas por Begelman, Keller e Smadja (2008): “As pessoas pensam e etiquetam de maneiras diferentes. Isto cria um tag-space com ruídos, fazendo mais difícil encontrar material etiquetado por outras pessoas.” Apesar dessa limitação, a informação presente nas tags pode gerar informação interessante quando visualizada de maneira adequada, como no trabalho de Jim Bumgardner.

¹¹ www.twitter.com

¹² www.flickr.com

¹³ Disponível em <<http://blog.blprnt.com/blog/blprnt/goodmorning>>. Acesso:10/12/2009.

¹⁴ Palavras-chave associadas a conteúdos digitais.



Jim Bumgardner e sua série “Time Graphs”

Jim Bumgardner teve contato com programação enquanto estudava vídeo e música no California Institute of the Arts, entre 1980 e 1984. Atualmente, ele continua desenvolvendo aplicativos para a web, trabalhando em sites (atualmente trabalha no site de música *Topspin*¹⁵ e escrevendo, ocasionalmente, música.

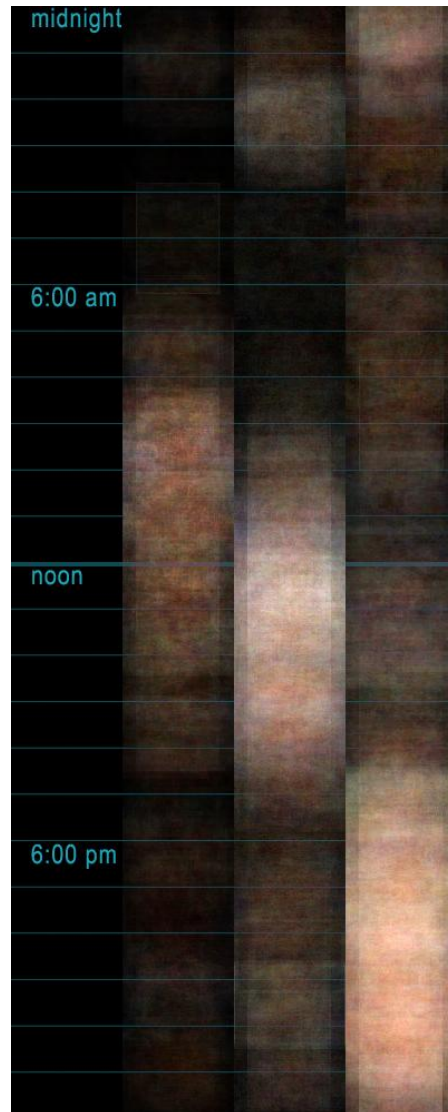


Fig. 5: Visualização das fotografias com os tags “breakfast”, “lunch” e “dinner” do Flickr (café da manhã, almoço e janta, respectivamente), organizadas pelo horário do dia em que foram tiradas (6 da manhã, meio-dia e 6 da tarde, de cima para baixo).

Sua relação com arte se manifesta mesmo em seus projetos mais ligados à programação. “Tenho tendência a associar arte com falta de utilidade, e a maioria dos meus projetos pessoais são gloriosamente sem utilidade”, diz, aproximando seu

¹⁵ www.topspin.com



trabalho do campo artístico. Entre seu trabalho, a série “Time Graphs”¹⁶ é a que mais transmite informação. Eles es aproximam da estrutura de um gráfico estatístico convencional, com um grid¹⁷ onde o eixo horizontal representa os meses do ano, e o vertical representa as horas do dia. Nesse grid são organizadas miniaturas de fotos disponíveis no Flickr, segundo as informações presentes nos metadados do arquivo.

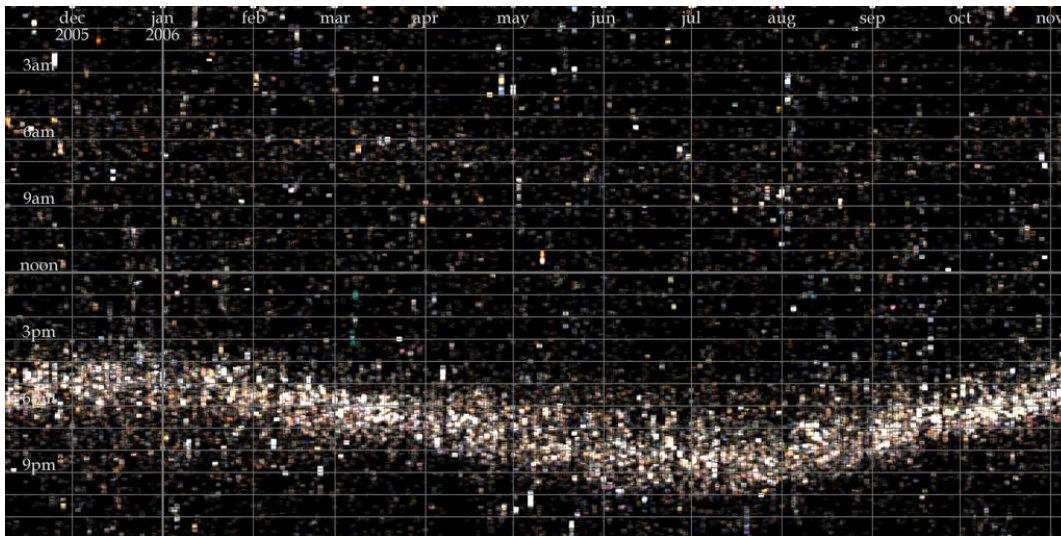


Fig.6: Visualização construída com o tag “sunset” (pôr do sol).

As fotografias são filtradas segundo suas tags. Em alguns casos, ele filtra também dados de diferentes regiões do mundo. Cada gráfico reúne aproximadamente 30 mil fotografias. Cada uma delas é transparente, o que permite que pontos de alta concentração de imagens fiquem mais brilhantes.

Para realizar estes gráficos, Bumgardner escreveu um script em Perl e Imagemagick¹⁸, que se conecta à base de dados do Flickr através de sua API para baixar as fotos e gerar os gráficos. O design dos gráficos também é impecável: as linhas que separam os meses e as horas são da cor e espessura certa para não atrapalhar a visualização das fotos, assim como seus rótulos¹⁹ e o preto é o fundo ideal para contrastar com a cor clara das imagens. “Há uma técnica para fazer gráficos que também é uma forma de arte. Mas neste caso, eu também os fiz de uma maneira brincalhona, que eu identificaria claramente como um processo artístico, similar à maneira que eu faço algumas músicas” diz. Podemos afirmar que o conhecimento em programação de

¹⁶ Disponível em <<http://www.flickr.com/photos/krazydad/sets/140323/>>. Acesso em 10/12/2009.

¹⁷ Série de eixos horizontais e verticais para organizar informação.

¹⁸ Perl é uma linguagem de programação. ImageMagick é um software de manipulação de imagens.

¹⁹ TUFTE, *Envisioning Information*, pág. 59



Bumgardner é o meio através do qual ele faz arte; isso é mais evidente em outros projetos, como em seus caleidoscópios²⁰.

O resultado é um gráfico intrigante, de alta densidade de dados²¹, que permite leituras micro e macro²². Uma leitura macro é a leitura do gráfico como um todo. No caso do gráfico feito a partir da tag “sunset” (Fig. 6), uma das conclusões que podemos tirar é que no verão o sol se põe mais tarde, como indica a curva que desce na parte do gráfico referente aos meses de junho e julho, verão no hemisfério norte. Como essa curva reflete as estações do hemisfério norte, o gráfico também pode indicar que no Flickr há uma predominância de fotos dessa região. Em uma leitura micro, o que se vê são os detalhes; em agosto a curva fica mais densa, coincidindo com o período que o Flickr deu maior suporte a geolocalização; o usuário do site GustavoG observou um ritmo mensal, que poderia coincidir com o momento que o Flickr “zera” seu limite de upload para usuários gratuitos²³, que seria quando os usuários podem postar novamente.

Essa leitura micro evidencia algumas inconsistências. Por exemplo, a presença de fotos com a tag “pôr-do-sol” feitas às nove da manhã pode ser explicada pelo horário incorreto da câmera ou pelo fotógrafo ter etiquetado a foto com essa palavra por outros motivos (problemas típicos da folksonomia, apontados por Begelman, Keller e Smadja).

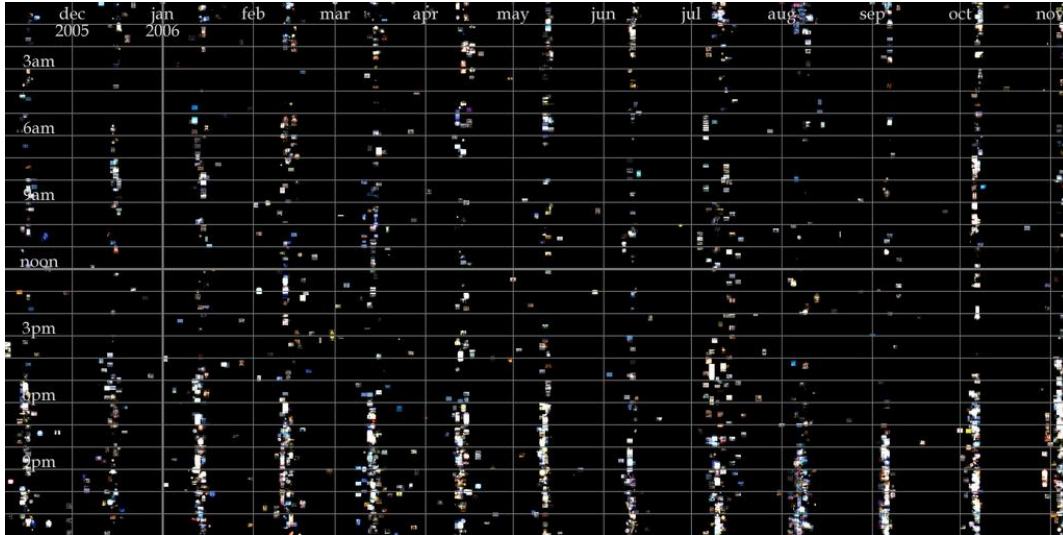


Fig. 7: Visualização construída com o tag “full moon” (lua cheia), onde é visível o ciclo da lua ao longo dos meses.

²⁰ Disponíveis em <http://www.krazydad.com/bestiary/bestiary_yinyang_scope.html>. Acesso: 10/12/2009.

²¹ TUFTE, *Envisioning Information*, pág 49–51.

²² *Idem*, capítulo 3.

²³ Comentário visível em <http://www.flickr.com/photos/krazydad/292424841/>



Sobre a informação revelada nos gráficos, Bumgardner diz: “Os fenômenos que os gráficos mostram não são necessariamente de interesse científico (por exemplo, o fato que fotos de café da manhã serem tiradas mais cedo que as de almoço (Fig. 5) não é grande novidade), então sua falta de utilidade é também evidência de suas propriedades artísticas.” Apesar da precisão científica do método, ele não vê os gráficos como algo útil, e sim como algo brincalhão. “Eu fiz estes gráficos para minha diversão – não necessariamente para sua elucidação” disse Jim, no fórum do site de Edward Tufte, onde estava sendo discutido seu trabalho.²⁴

Essa aparente contradição entre a concepção divertida dos projetos e a exatidão científica do processo e da técnica artística é o que faz o trabalho de Bumgardner ser difícil de encaixar nos modelos de gráficos mais conhecidos. Estamos acostumados a gráficos estatísticos de assuntos mais “relevantes” e de fontes mais confiáveis. A série “Time Graphs” rompe com vários parâmetros atuais e abre novos caminhos para o design de informação, a arte, e visualização de dados científicos.

Conclusão

A disponibilidade de informação estatística revoluciona a maneira de procurar informação. Com toda essa informação disponível, é necessário mostrá-la de maneira adequada para serem compreendidos pelas pessoas a quem a informação é destinada. E essa linguagem, que apesar de não ser tão nova, está cada vez mais popular no repertório de artistas; e não só a linguagem visual, mas os recursos, que podem ser usados para criar arte.

A série “Time Graphs”, de Jim Bumgardner reúne todas essas características. Usa informação de uma maneira nova, com um método exato. O resultado é rico em informação, e permite, a quem estiver disposto a fazer novas leituras, criar novas informações. Mostra as possibilidades artísticas da programação e uso de computadores, que ainda são vistos pelos artistas como uma técnica nova, que não é ensinada no contexto da arte. E, como bônus, é feita com humor.

Trabalhos como este influenciam várias áreas, e aos poucos abrem caminho para novas experiências na arte, no design, no jornalismo e em outras áreas da comunicação, fazendo com que o público se habitue a novas linguagens e maneiras de ler e entender informação.

²⁴ Disponível em <http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001iT>. Acesso em 10/12/2009.



Referências bibliográficas

ANTOUN, Henrique (org.). **Web 2.0: participação e vigilância na era da comunicação distribuída**. Rio de Janeiro: Mauad X, 2008.

BEGELMAN, Grigory; KELLER, Philipp; SMADJA, Frank. **Automated Tag Clustering: Improving search and exploration in the tag space**. Escócia, 2006. Disponível em: <http://www.pui.ch/phred/automated_tag_clustering/>. Acesso: 10/12/2009.

TUFTE, Edward. **The Visual Display of Quantitative Information**. Connecticut: Graphic Press, 2001.

_____, _____. **Envisioning Information**. Connecticut: Graphic Press, 1990.

UNWIN, Antony THEUS, Martin, HOFMANN, Heike. **Graphics of Large Datasets**. Nova Iorque: Springer, 2006.



Anexo: entrevista com Jim Bumgardner, via email.

I'd like to know about your relationship with art. You have worked with music and poetry. Do you consider that your "Time graphs" series are art too?

There isn't a clear boundary between "art" and "not art". There is a craft to making graphs which is a form of art. But in this case, I also made them in a playful way that I would clearly identify as an artistic process, similar to the way I make some music. Also, the phenomena that the graphs illustrate are not necessarily of scientific interest (e.g. the fact that photos taken during breakfast occur earlier than lunch isn't groundbreaking news), so their uselessness is also evidence of their artistic properties. I tend to associate art with lack of utility, and most of my personal projects are gloriously useless.

Did you intend making an exhibition with them, other than showing on Internet? (I personally would like to see them printed big)

I have indeed displayed some of them (at the Data+Art show at the Pasadena Museum of California Art, last year). See this link. Incidentally, the pieces in that art show explored the boundary between data and art, in a manner which you would find interesting. Having said that, I didn't originally create these with the intention of showing them in an art show. They were just personal experiments that I shared on Flickr. I was invited to the show.

The "Time graphs" have more design elements than other of your works. For example, you chose a logical way to arrange the photos, and didn't change them after it was done; I see it as a design decision, rather than an artistically one. What can you say about that?

I don't think "design decisions" are different than "artistic decisions". Design is a subset of art, and all design decisions are artistic decisions. Some artistic decisions involve form and layout, and some involve semantics, messaging and other things.

How did you choose the tags to get the photos? It was like "I want to see what happens if I choose this tag" or more like "I bet this tag would make an interesting graph"?

It's usually a bit of both. In some cases, I specifically chose tags that I knew were time correlated and might be interesting (such as "sunset", "breakfast lunch dinner", "full moon"), in other cases, (such as "accidents") I was intentionally looking for things that weren't highly time correlated, because I was interested to find out what I would get. "What happens if I choose this" was the question I asked when I built my "color of" graph.



(<<http://www.flickr.com/photos/krazydad/5035125>>)

What kind of tags worked better?

Well, “better” is subjective. The time-correlated tags tend to be more interesting to me, but they can also be kind of boring, because of the predictability of them. I think “sunsets” and “breakfast lunch dinner” are the strongest pieces, personally, because the subject matter pertains to matters which are core to human experience.

Do somebody discovered something that you hadn't realized, other than GustavoG in the “Sunsets in different parts of the world” graph?

I think there was another photographer who recognized a strip of his photos in one graph. When Edward Tufte first saw them, he liked their use of “images as data points”, and commented that they were a counter example to Priit P’s assertion that “montage probably cannot be used as evidence in the strict sense”

Can you name your influences in this style of work?

Edward Tufte, Eadward Muybridge, Harold Edgerton, John Whitney Sr., John Cage.