

Tabela de Conteúdos

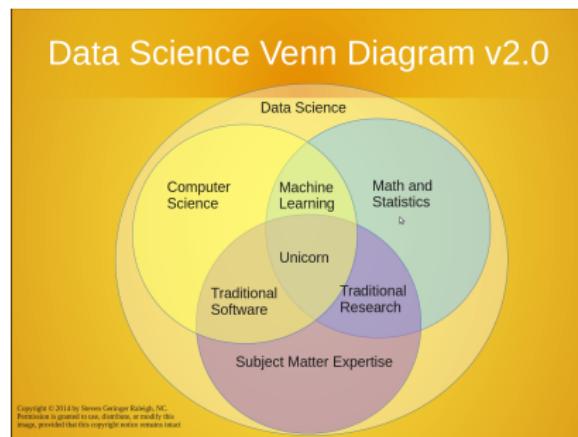
1 Introdução

2 Tendências

3 Considerações Finais

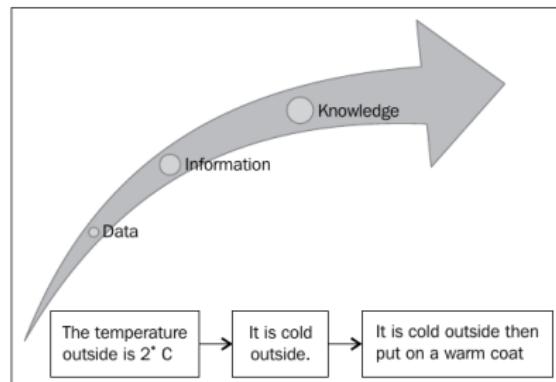
Cientistas & Unicórnios

- “Data Scientist” is a Data Analyst who lives in California.
- A data scientist is someone *who is better at statistics than any software engineer and better at software engineering than any statistician.* (Wills, Cloudera)



Dados, Informação e Conhecimento

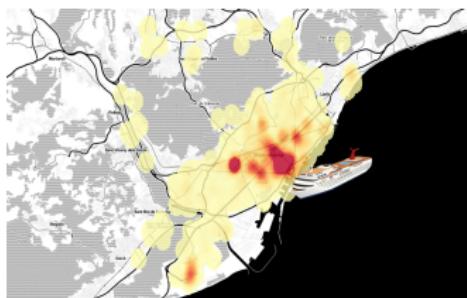
- Dados são os factos que descrevem o mundo.
- A informação surge quando transformamos esses valores em algo relevante.
- Conhecimento implica uma generalização dos dados e da informação, de forma a criar regras.



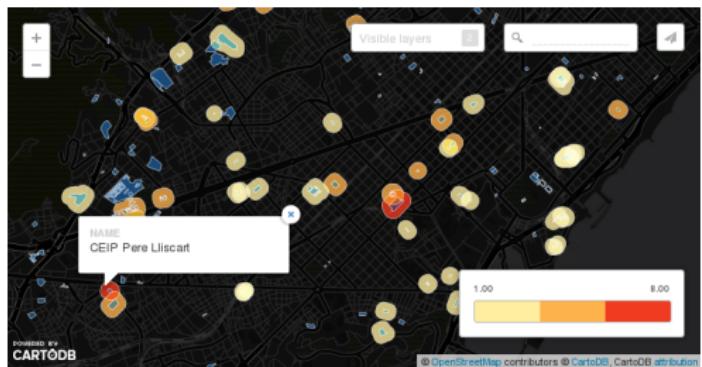
Algo espacial...

Lei de Tobler:

Everything is related to everything else, but near things are more related to each other.



Heatmap de Tweets perto de um Cruzeiro.



Buffers de acidentes ao redor de escolas.

Problemas Fundamentais

Longley *ET AL* (2005):

Problemas Fundamentais

Longley *ET AL* (2005):

- O comportamento espacial actual, muitas vezes reflecte padrões passados.

Problemas Fundamentais

Longley *ET AL* (2005):

- O comportamento espacial actual, muitas vezes reflecte padrões passados.
- A explicação no tempo apenas necessita de olhar para o passado, mas a explicação no espaço necessita de olhar em todas as direções simultaneamente.

Problemas Fundamentais

Longley *ET AL* (2005):

- O comportamento espacial actual, muitas vezes reflecte padrões passados.
- A explicação no tempo apenas necessita de olhar para o passado, mas a explicação no espaço necessita de olhar em todas as direções simultaneamente.
- Embora alguns fenômenos espaciais variem de forma gradual através do espaço, outros podem exibir uma extrema irregularidade.

Problemas Fundamentais

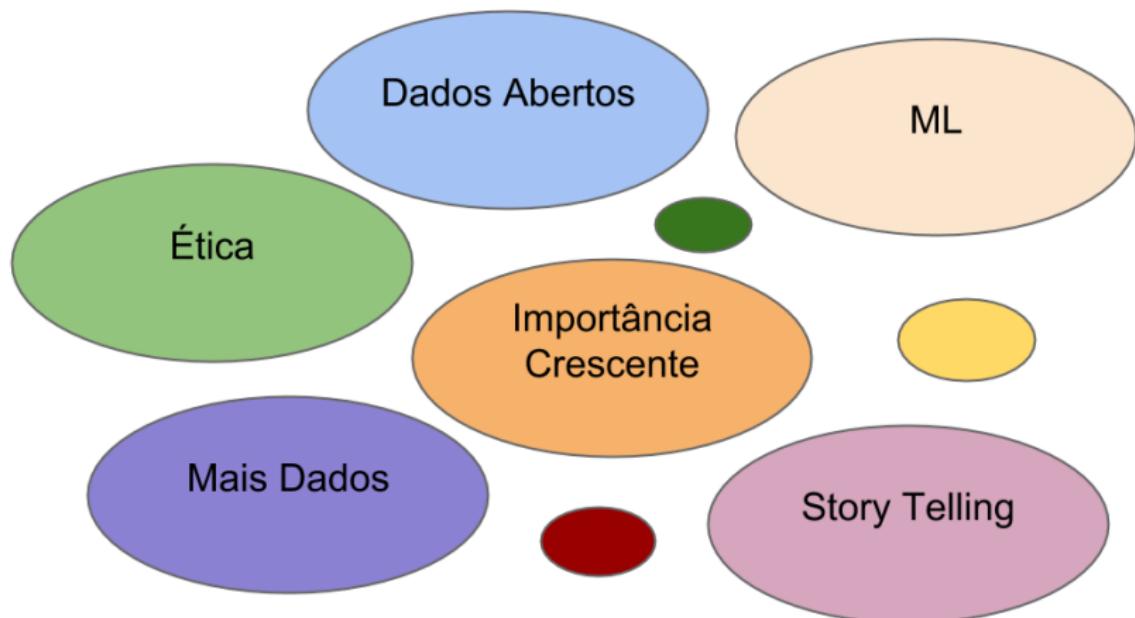
Longley *ET AL* (2005):

- O comportamento espacial actual, muitas vezes reflecte padrões passados.
- A explicação no tempo apenas necessita de olhar para o passado, mas a explicação no espaço necessita de olhar em todas as direções simultaneamente.
- Embora alguns fenômenos espaciais variem de forma gradual através do espaço, outros podem exibir uma extrema irregularidade.
- Embora a autocorrelação espacial nos ajude a construir representações, ela pode frustrar os nossos esforços de predição.

Onde Vamos?

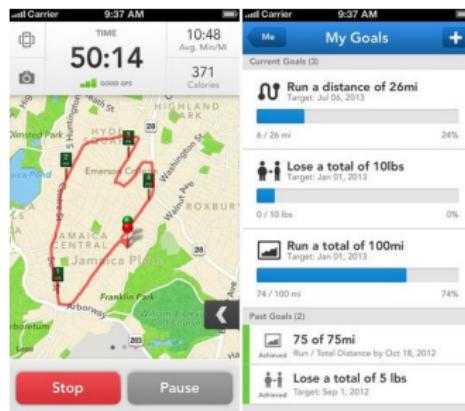


(Algumas) Tendências



Importância Crescente

- A ciencia de dados ganha torna-se cada vez mais importante, a medida que entra em novos campos.



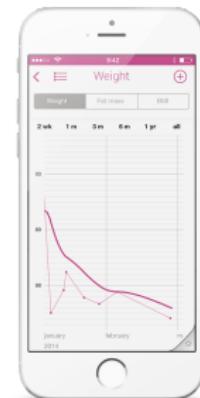
Importância Crescente

- A ciencia de dados ganha torna-se cada vez mais importante, a medida que entra em novos campos.
- A Geografia tambem beneficia desse impulso, a medida que o publico em geral toma consciencia de que a maior parte das coisas, acontecem num lugar.



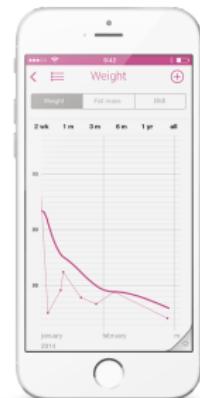
Importância Crescente (cont.)

- *Whitings* é uma balança “inteligente”.



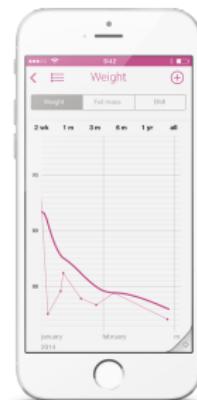
Importância Crescente (cont.)

- *Whitings* e uma balança “inteligente”.
- Recolhe medições corporais precisas: peso, massa gorda e batimentos cardíacos.



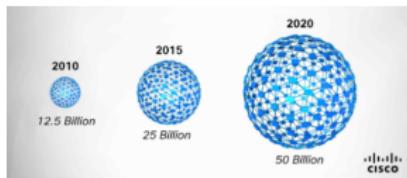
Importância Crescente (cont.)

- *Whitings* é uma balança “inteligente”.
- Recolhe medições corporais precisas: peso, massa gorda e batimentos cardíacos.
- Uma app analisa estes dados, mostra tendências e permite gerar planos e monitorizar metas.



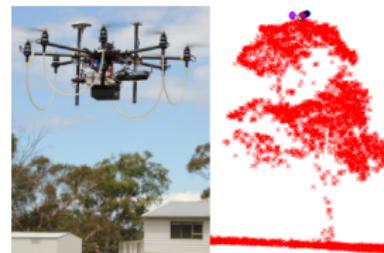
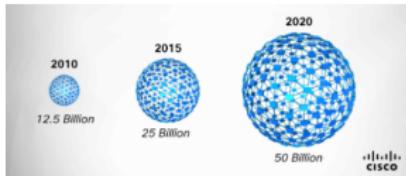
Mais Dados

- Existe uma explosão na quantidade de dados gerados por sensores (IoT, IoE).



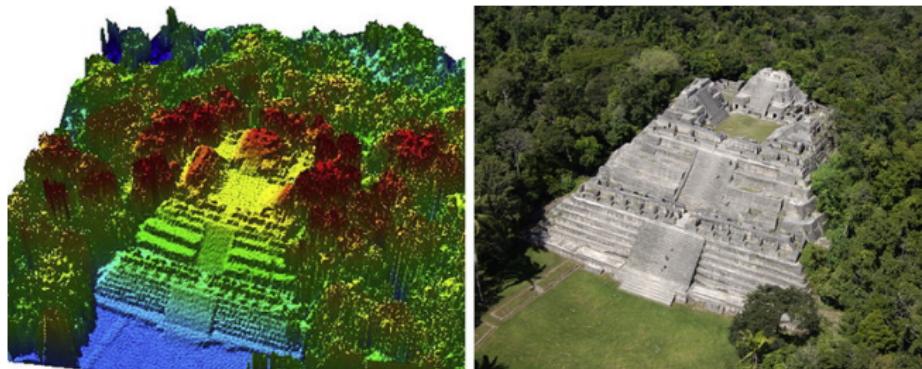
Mais Dados

- Existe uma explosão na quantidade de dados gerados por sensores (IoT, IoE).
- Graças a tecnologias de posicionamento mais baratas e generalizadas (e.g.: receptores de GPS), uma proporção grande destes dados está georeferenciada.



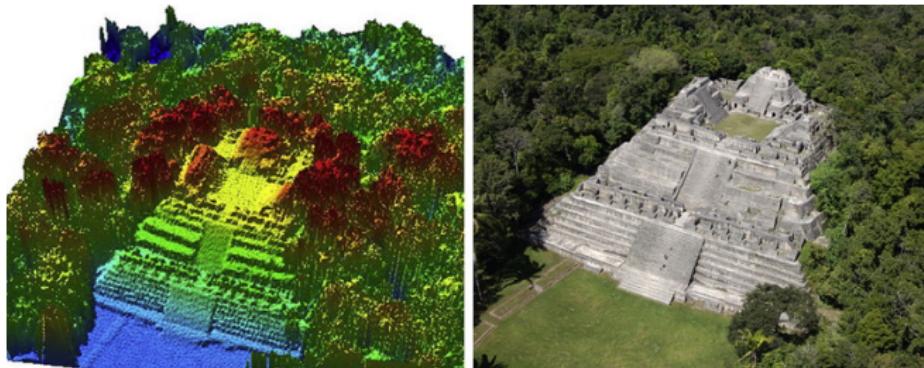
Mais Dados (cont.)

- LIDAR: mede propriedades da luz reflectida de modo a obter a distância ou outra informação a respeito de um determinado objecto distante.



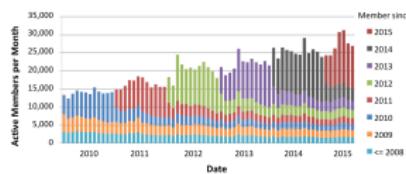
Mais Dados (cont.)

- LIDAR: mede propriedades da luz reflectida de modo a obter a distância ou outra informação a respeito de um determinado objecto distante.
- O UAV LIDAR aplica esta técnica a partir de um drone.



E Ainda Mais Dados

- Existe uma explosão na quantidade de UGC.
- voluntário: citizens as sensors e VGI (p.e: OSM).
- Não-voluntário (?): gerados nas redes sociais (p.e: Twitter).

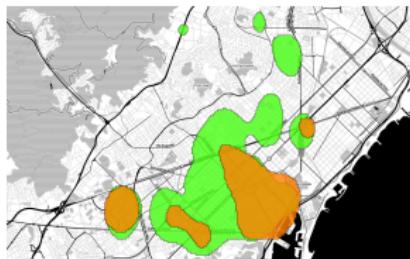


Evolução do número de membros do OSM.

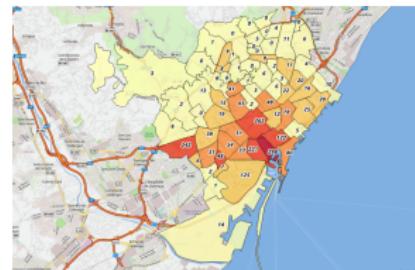


Mapa com a localização do último milhão de Tweets.

E Ainda Mais Dados (cont.)



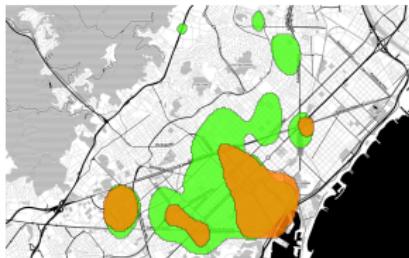
Clusters de Tweets de locais e estrangeiros.



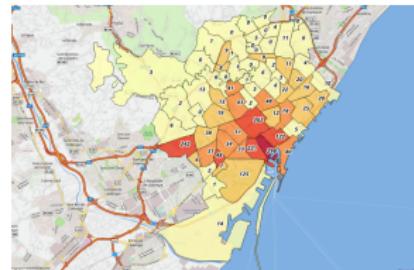
Distribuição de densidades de Tweets de estrangeiros.

E Ainda Mais Dados (cont.)

- Um método de detecção de origem, permite diferenciar os Tweets gerados por locais dos gerados por estrangeiros.



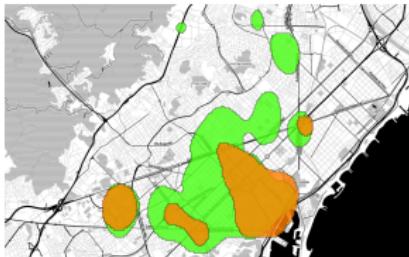
Clusters de Tweets de locais e estrangeiros.



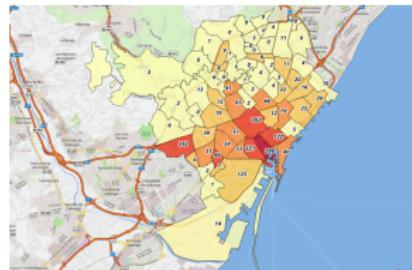
Distribuição de densidades de Tweets de estrangeiros.

E Ainda Mais Dados (cont.)

- Um metodo de detecção de origem, permite diferenciar os Tweets gerados por locais dos gerados por estrangeiros.
- Podemos observar que ha um acrescimo no numero de Tweets gerados por estrangeiros, apos a chegada dos cruzeiros ao Porto de Barcelona.



Clusters de Tweets de locais e estrangeiros.



Distribuição de densidades de Tweets de estrangeiros.

“Arqueologia” de Dados

- Existe uma necessidade crescente de “recuperar” fontes de dados de legacy (por exemplo em formato analógico)



Mapa de casos de colera em Londres, produzido por John Snow (1864).

“Arqueología” de Dados

- Existe uma necessidade crescente de “recuperar” fontes de dados de legacy (por exemplo em formato analogico)
 - Embora isto implique um desafio tecnologico, muitas vezes estes conjuntos de dados sao extremamente valiosos.



Mapa de casos de colera em Londres, produzido por John Snow (1864).

“Arqueologia” de Dados (cont.)

- Dados de informacao de pescas (biomassas), foram guardados sem metadados.

“Arqueología” de Dados (cont.)

- Dados de informacao de pescas (biomassas), foram guardados sem metadados.
 - Para capturar estes dados, foi necessario efectuar cruzeiros com um custo elevadissimo.

```
Path list file is "[ /home/joaaa/projects/no/test2/eng1 ]", "/home/joaaa/projects/no/test2/eng1" [  
  Reader list file is "[ \"eng1\", \"arg1\" ]"  
  Reader duplicates:  
    00773-26771196-474a-b2069224ad [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp]  
    2015e-7c95eab2-02bc-2d3288d8bf [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng2.asp]  
    2015e-7c95eab2-02bc-2d3288d8bf [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng3.asp]  
    2015e-7c95eab2-02bc-2d3288d8bf [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng4.asp]  
    66dd0-2a62cd2a-2d2a-1bd1c12c1a [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp]  
    66dd0-2a62cd2a-2d2a-1bd1c12c1a [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng2.asp]  
    66dd0-2a62cd2a-2d2a-1bd1c12c1a [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng3.asp]  
    66dd0-2a62cd2a-2d2a-1bd1c12c1a [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng4.asp]  
    False duplicate: [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_Pelagic_Gelon_Mantisia/Copy of awww888.asp]  
    False duplicate: [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_Pelagic_Gelon_Mantisia/Copy of awww888.asp]  
    False duplicate: [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_Pelagic_Gelon_Mantisia/Copy of awww888.asp]  
    False duplicate: [/home/joaaa/projects/no/test2/eng1/2000048_eng1.asp?xscript=eng1.asp] -> [/home/joaaa/projects/no/test2/eng1/2000048_Pelagic_Gelon_Mantisia/Copy of awww888.asp]
```

“Arqueologia” de Dados (cont.)

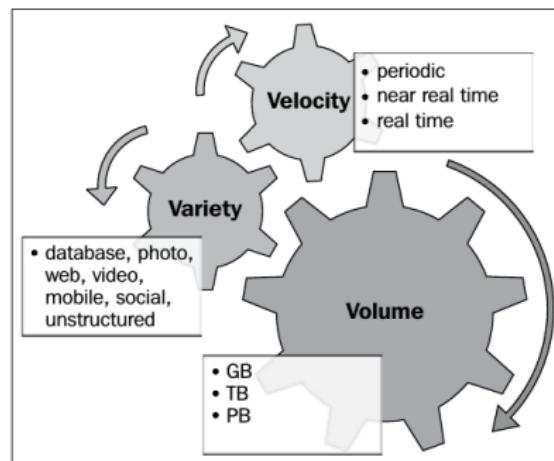
- Dados de informacao de pescas (biomassas), foram guardados sem metadados.
 - Para capturar estes dados, foi necessario efectuar cruzeiros com um custo elevadissimo.
 - A informacao capturada e irrepetivel.

“Arqueología” de Dados (cont.)

- Dados de informacao de pescas (biomassas), foram guardados sem metadados.
 - Para capturar estes dados, foi necessario efectuar cruzeiros com um custo elevadissimo.
 - A informacao capturada e irrepetivel.
 - Os metadados estao encodificados na path do ficheiro (i.e.: timestamp, nome do cruzeiro).

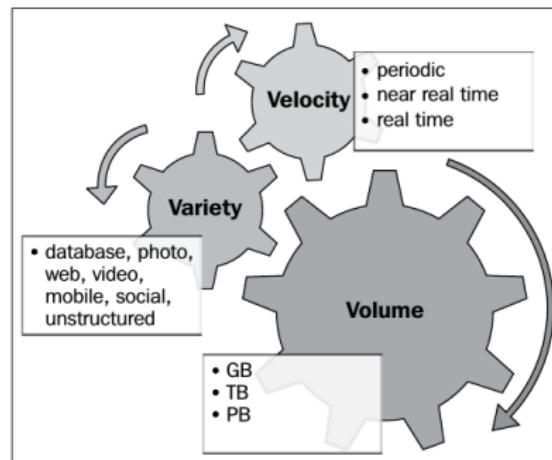
Tecnologias de *Big Data*

- 3 Vs.



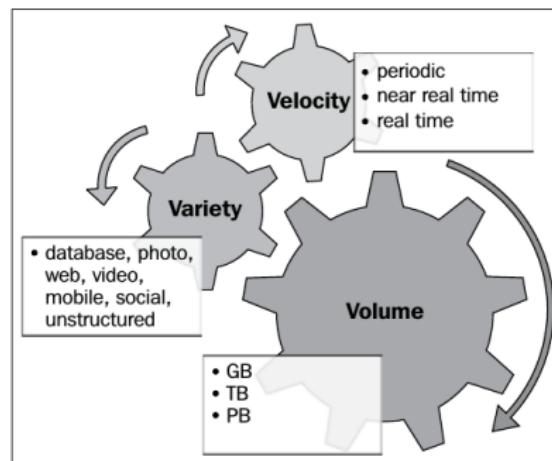
Tecnologias de *Big Data*

- 3 Vs.
- Computação em ambiente Cloud, NoSQL, Processamento em tempo-real ,Linked Data.



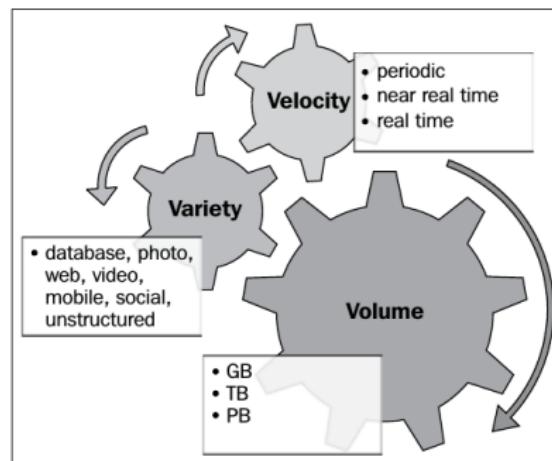
Tecnologias de *Big Data*

- 3 Vs.
- Computação em ambiente Cloud, NoSQL, Processamento em tempo-real ,Linked Data.

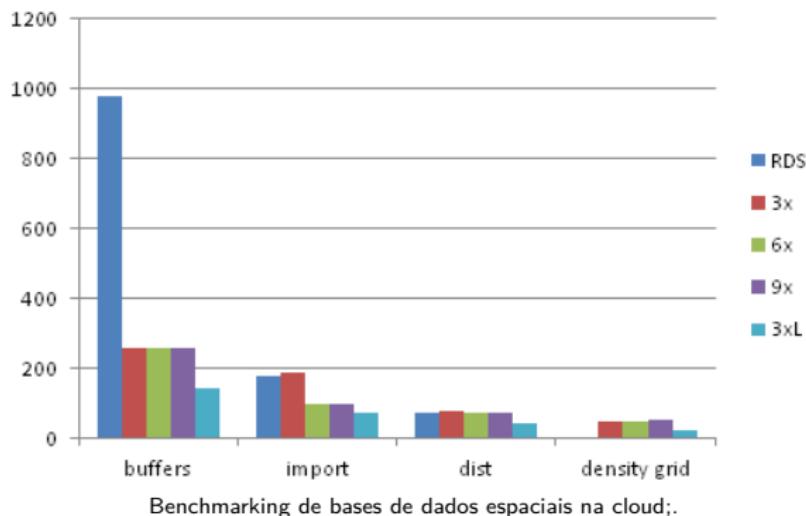


Tecnologias de *Big Data*

- 3 Vs.
- Computação em ambiente Cloud, NoSQL, Processamento em tempo-real ,Linked Data.
- Big Spatial Data.

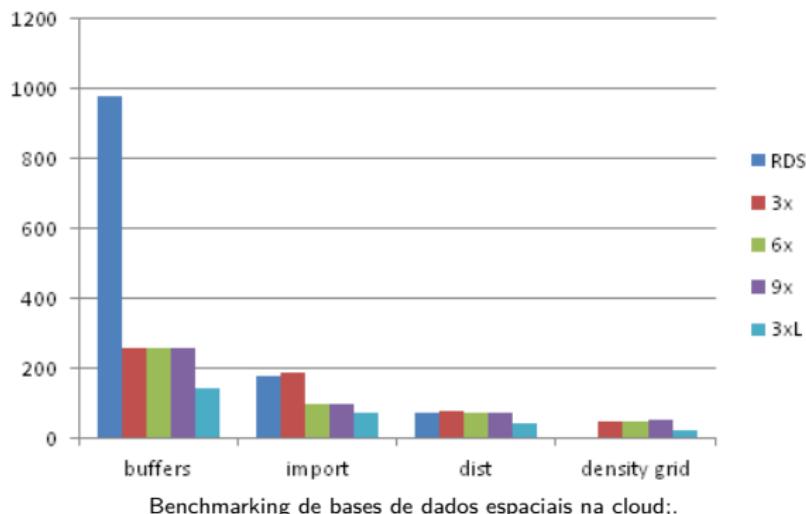


Desenvolvimento de tecnologias espaciais escaláveis e distribuídas.



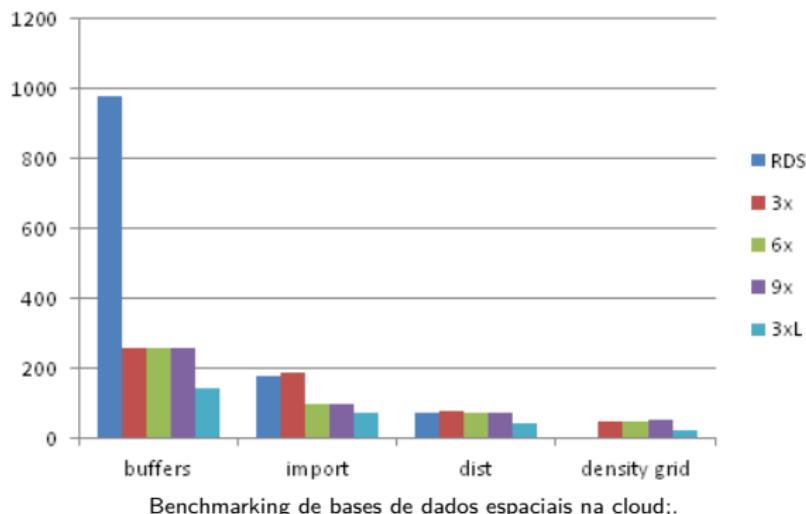
Desenvolvimento de tecnologias espaciais escaláveis e distribuídas.

- Suporte espacial ainda limitado.



Desenvolvimento de tecnologias espaciais escaláveis e distribuídas.

- Suporte espacial ainda limitado.
- Necessidade crescente de saber em que situações utilizar cada tecnologia.



Uso Cada vez Mais Generalizado de ML

A disponibilidade crescente de dados e os progressos computacionais vão impulsionar o uso de Machine Learning.

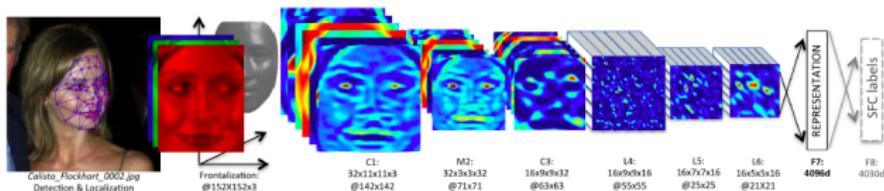


Figure 2. Outline of the DeepFace architecture. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate outputs for each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

Arquitectura de uma rede de Deep Learning para reconhecimento facial.

Uso Cada vez Mais Generalizado de ML

A disponibilidade crescente de dados e os progressos computacionais vão impulsionar o uso de Machine Learning.

- Ascendência de algoritmos computacionalmente intensivos (p.e.: Deep Learning)

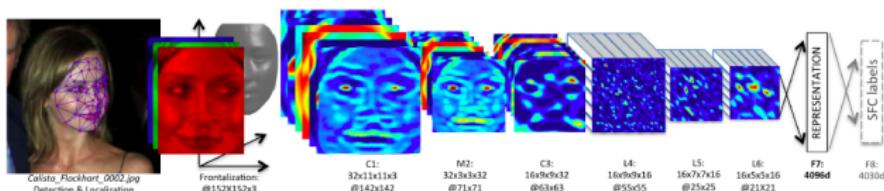


Figure 2. Outline of the *DeepFace* architecture. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate outputs for each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

Arquitectura de uma rede de Deep Learning para reconhecimento facial.

Uso Cada vez Mais Generalizado de ML

A disponibilidade crescente de dados e os progressos computacionais vão impulsionar o uso de Machine Learning.

- Ascendência de algoritmos computacionalmente intensivos (p.e.: Deep Learning)
- Modelos Ensemble.

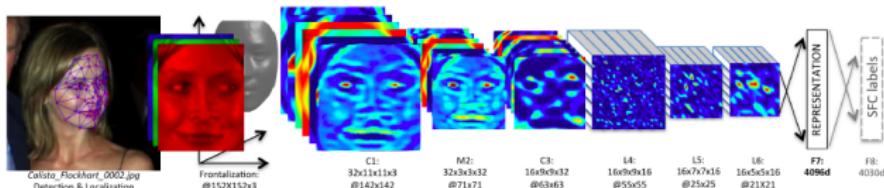


Figure 2. Outline of the *DeepFace* architecture. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate outputs for each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

Arquitectura de uma rede de Deep Learning para reconhecimento facial.

Uso Cada vez Mais Generalizado de ML

A disponibilidade crescente de dados e os progressos computacionais vão impulsionar o uso de Machine Learning.

- Ascendência de algoritmos computacionalmente intensivos (p.e.: Deep Learning)
- Modelos Ensemble.
- Modelos espaciais.

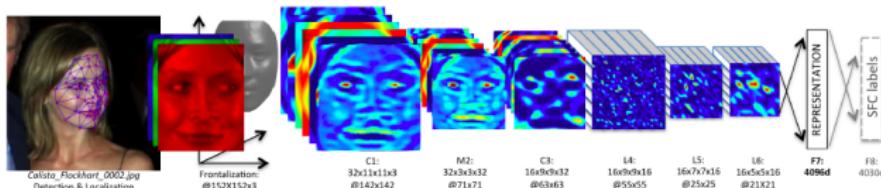


Figure 2. Outline of the DeepFace architecture. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate outputs for each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

Arquitectura de uma rede de Deep Learning para reconhecimento facial.

Uso Cada vez Mais Generalizado de ML

A disponibilidade crescente de dados e os progressos computacionais vão impulsionar o uso de Machine Learning.

- Ascendência de algoritmos computacionalmente intensivos (p.e.: Deep Learning)
- Modelos Ensemble.
- Modelos espaciais.
- Descriptivo → Predictivo

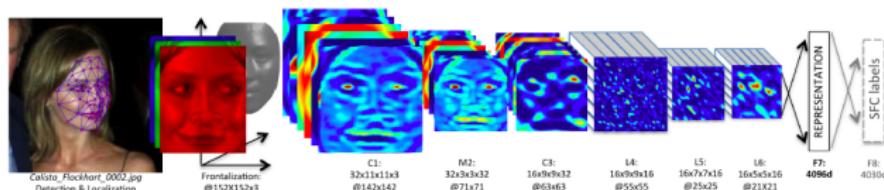
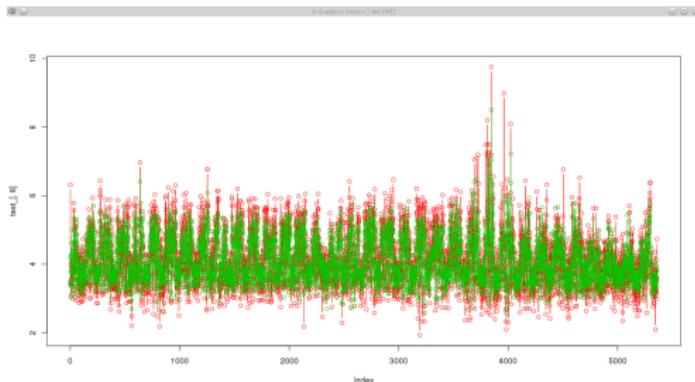


Figure 2. Outline of the *DeepFace* architecture. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate outputs for each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

Arquitectura de uma rede de Deep Learning para reconhecimento facial.

Uso Cada vez Mais Generalizado de *ML* (cont.)

Trafico é um domínio onde tradicionalmente se utilizam modelos de micro-simulação.

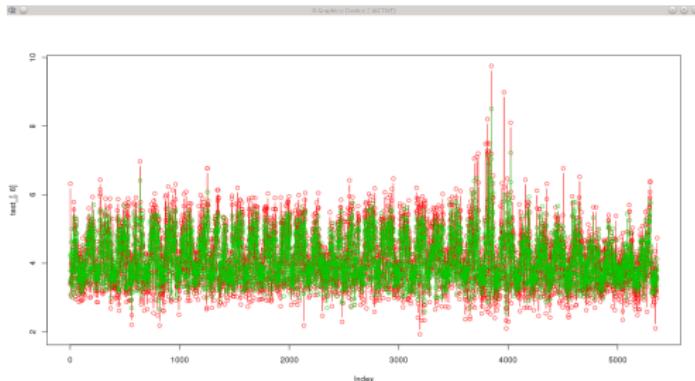


Ajuste entre as previsões do SVM (verde) e os valores observados (vermelho).

Uso Cada vez Mais Generalizado de *ML* (cont.)

Trafico é um domínio onde tradicionalmente se utilizam modelos de micro-simulação.

- Foram aplicados modelos de ML para prever os tempos de viagem na cidade de Barcelona.

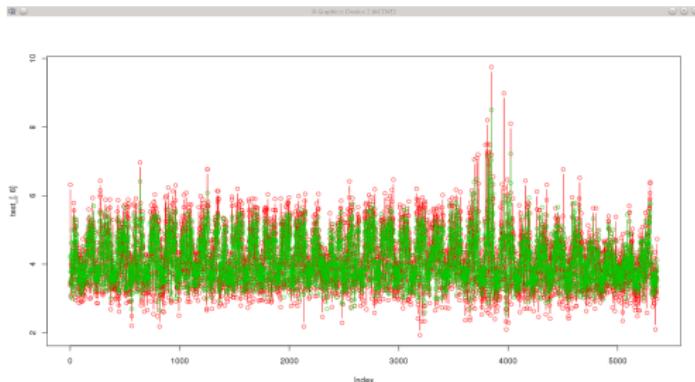


Ajuste entre as previsões do SVM (verde) e os valores observados (vermelho).

Uso Cada vez Mais Generalizado de *ML* (cont.)

Traffic é um domínio onde tradicionalmente se utilizam modelos de micro-simulação.

- Foram aplicados modelos de ML para prever os tempos de viagem na cidade de Barcelona.
- Os modelos foram treinados e validados com um ano de dados.

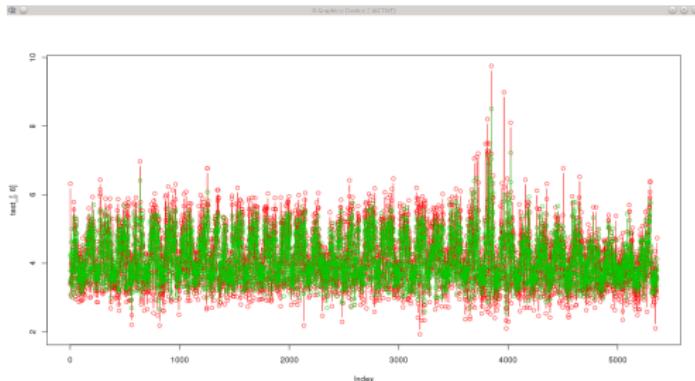


Ajuste entre as previsões do SVM (verde) e os valores observados (vermelho).

Uso Cada vez Mais Generalizado de *ML* (cont.)

Trafico é um domínio onde tradicionalmente se utilizam modelos de micro-simulação.

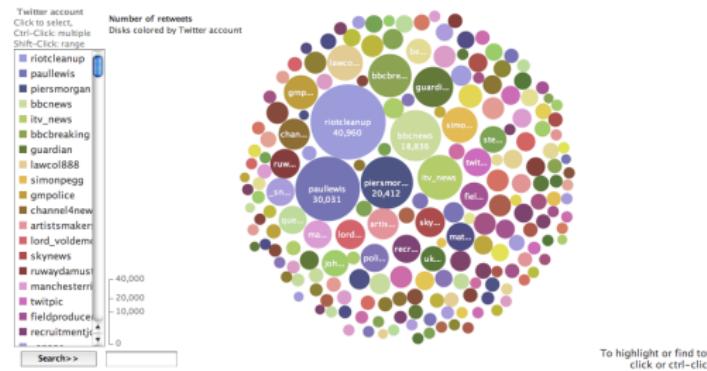
- Foram aplicados modelos de ML para prever os tempos de viagem na cidade de Barcelona.
- Os modelos foram treinados e validados com um ano de dados.
- A ANN foi descartada a favor de uma SVM.



Ajuste entre as previsões do SVM (verde) e os valores observados (vermelho).

Jornalismo de Dados & Story Telling

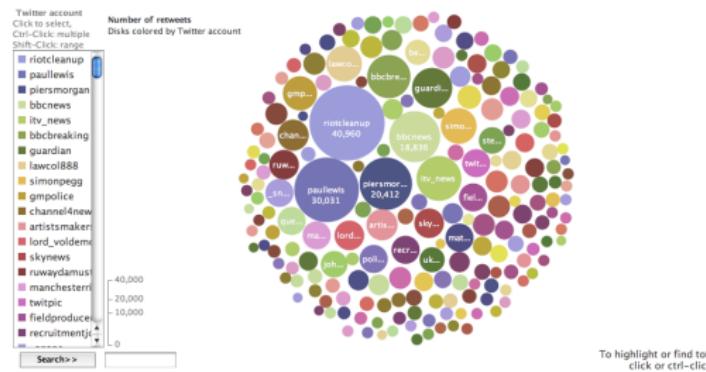
- Os dados quantitativos tem um papel cada vez mais valorizado na produção e distribuição de informação.



Visualização de reweets de um rumor, por Twitter account

Jornalismo de Dados & Story Telling

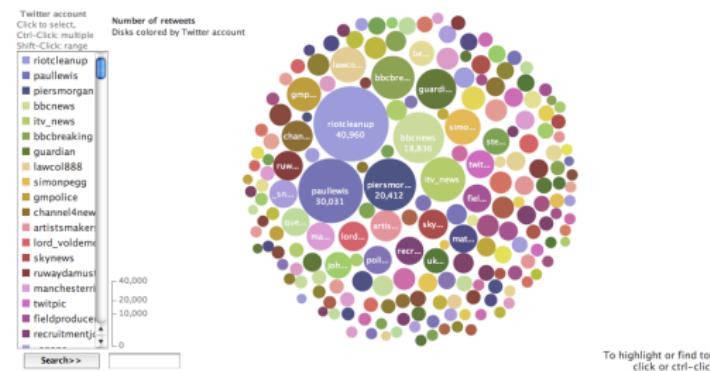
- Os dados quantitativos tem um papel cada vez mais valorizado na produção e distribuição de informação.
- Existe uma necessidade crescente de apresentar “histórias” de dados a uma audiência mais vasta (infografia).



Visualização de reweets de um rumor, por Twitter account

Jornalismo de Dados & Story Telling

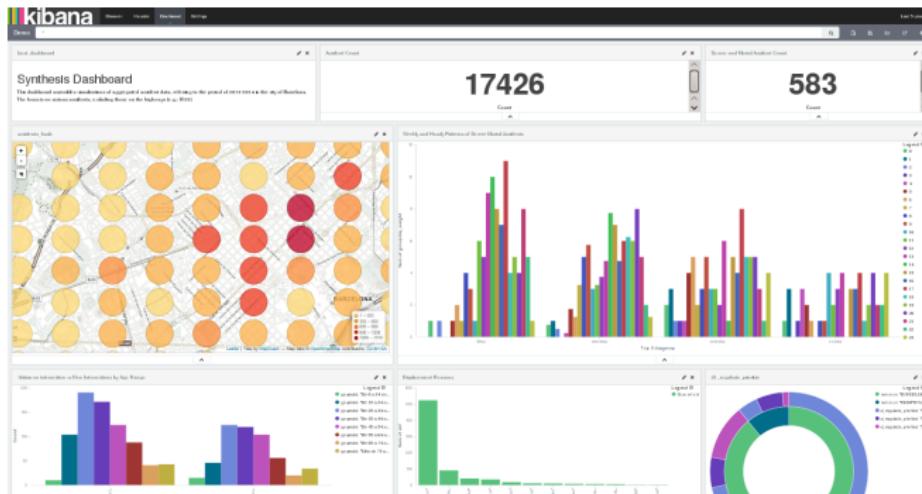
- Os dados quantitativos tem um papel cada vez mais valorizado na produção e distribuição de informação.
- Existe uma necessidade crescente de apresentar “histórias” de dados a uma audiência mais vasta (infografia).
- As visualizações (incluindo mapas) podem ser uma ferramenta valiosa.



Visualização de reweets de um rumor, por Twitter account

Jornalismo de Dados & Story Telling (cont.)

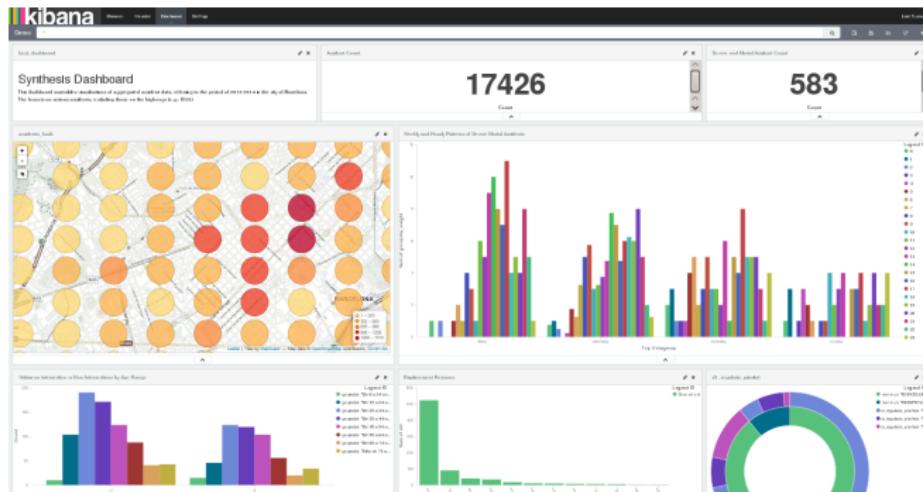
As dashboard dão a possibilidade de visualizar os dados de forma dinâmica, e em tempo real.



Jornalismo de Dados & Story Telling (cont.)

As dashboard dao a possibilidade de visualizar os dados de forma dinamica, e em tempo real.

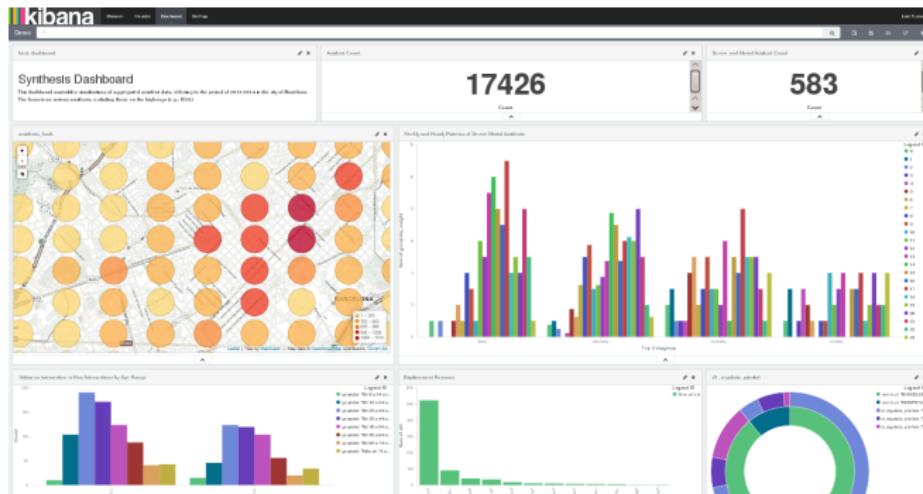
- Uma dashboard foi desenvolvida para transmitir os insights sobre acidentes graves e mortais na cidade de Barcelona.



Jornalismo de Dados & Story Telling (cont.)

As dashboard dao a possibilidade de visualizar os dados de forma dinamica, e em tempo real.

- Uma dashboard foi desenvolvida para transmitir os insights sobre acidentes graves e mortais na cidade de Barcelona.
- A stack utilizada (ELK) permite uma actualizacao em tempo real.



Mais e Melhores Dados Abertos

- “Revolução” dos dados abertos.



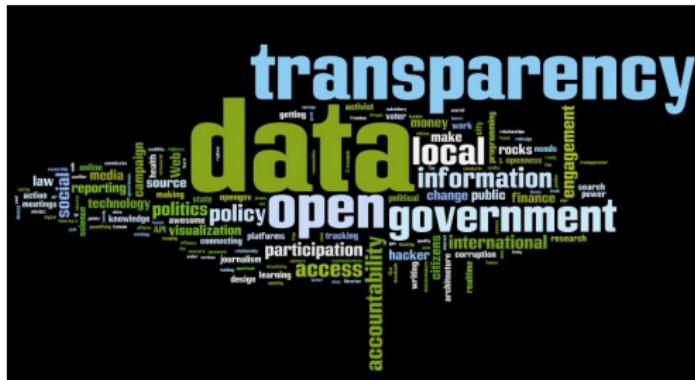
Mais e Melhores Dados Abertos

- “Revolução” dos dados abertos.
- Cada vez mais instituições são “pressionadas” para abrir os seus dados.

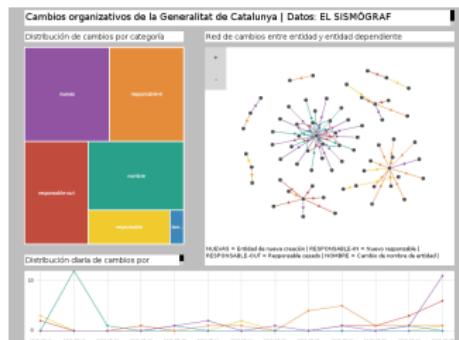


Mais e Melhores Dados Abertos

- “Revolução” dos dados abertos.
- Cada vez mais instituições são “pressionadas” para abrir os seus dados.
- Qualidade dos dados.

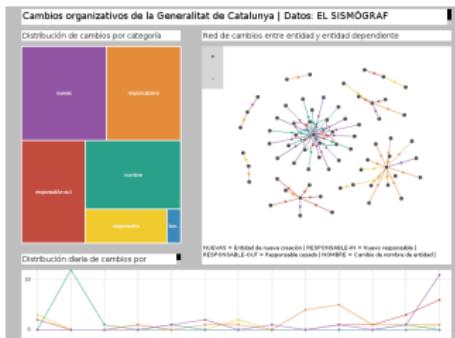


Mais e Melhores Dados Abertos (cont.)



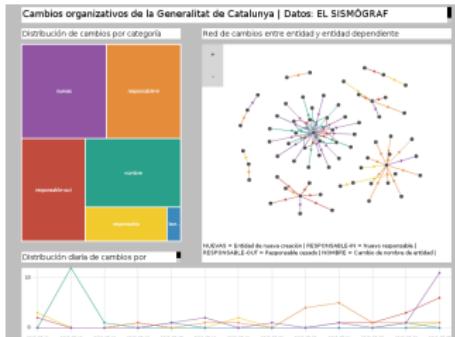
Mais e Melhores Dados Abertos (cont.)

- 2013: ficheiro actualizado varias vezes por semana com as 8191 entidades que compoem a estrutura do governo e Admininstracao Publica Catala.



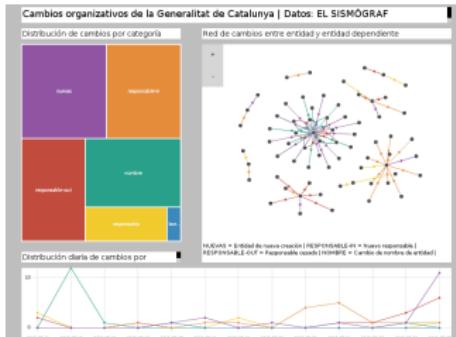
Mais e Melhores Dados Abertos (cont.)

- 2013: ficheiro actualizado varias vezes por semana com as 8191 entidades que compoem a estrutura do governo e Admininstracao Publica Catala.
- Sismografo da Catalunya @OpenGovCat: permite visualizar estes dados.

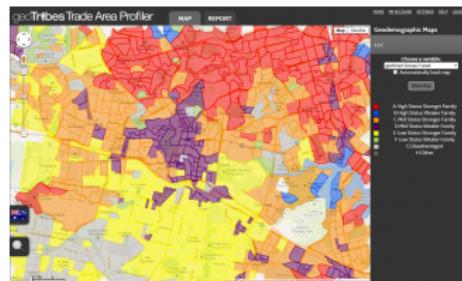


Mais e Melhores Dados Abertos (cont.)

- 2013: ficheiro actualizado varias vezes por semana com as 8191 entidades que compoem a estrutura do governo e Admininstracao Publica Catala.
- Sismografo da Catalunya @OpenGovCat: permite visualizar estes dados.
 - *Porque foram criadas 32 entidades novas e 57 entidades ficaram subitamente sem responsavel?*



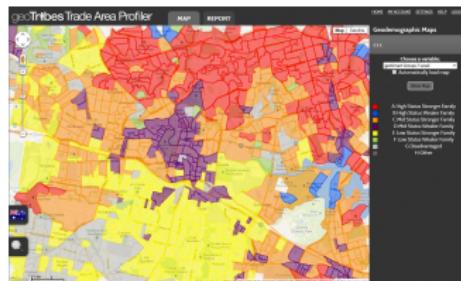
Ética



Os mapas de geodemografia suscitam algumas questões de privacidade.

Ética

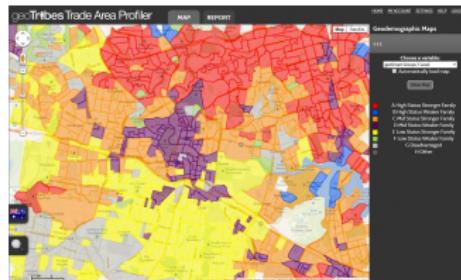
- As pessoas estão cada vez mais preocupadas com o uso dos seus dados pessoais, o que está reflectido em leis de privacidade.



Os mapas de geodemografia suscitam algumas questões de privacidade.

Ética

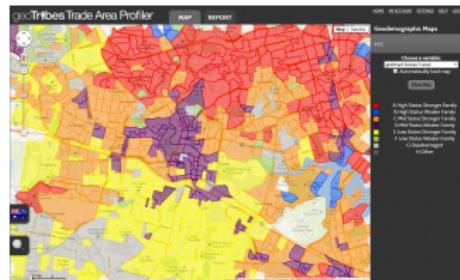
- As pessoas estão cada vez mais preocupadas com o uso dos seus dados pessoais, o que está reflectido em leis de privacidade.
- Devido à sua natureza, os dados de localização são particularmente sensíveis.



Os mapas de geodemografia suscitam algumas questões de privacidade.

Ética

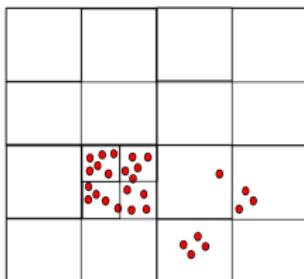
- As pessoas estão cada vez mais preocupadas com o uso dos seus dados pessoais, o que está reflectido em leis de privacidade.
- Devido à sua natureza, os dados de localização são particularmente sensíveis.
- Existe uma necessidade de desenvolver métodos que abordem estas preocupações.



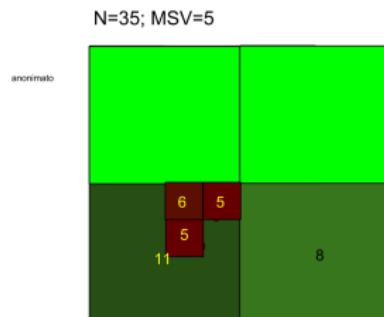
Os mapas de geodemografia suscitam algumas questões de privacidade.

Ética (cont.)

- Obfuscation spatial: altera, substitui ou generaliza a localização dos utilizadores, por forma a esconder a sua localização real.

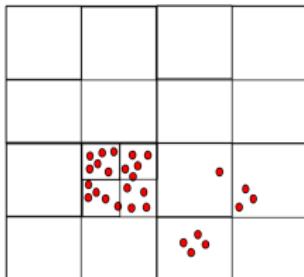


Raw points vs
Ofuscated points.

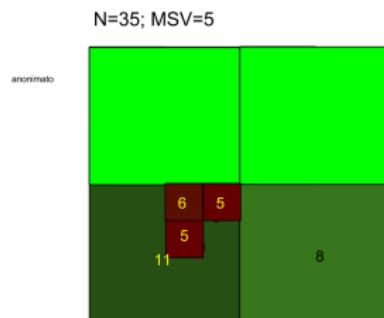


Ética (cont.)

- Obfuscation spatial: altera, substitui ou generaliza a localização dos utilizadores, por forma a esconder a sua localização real.

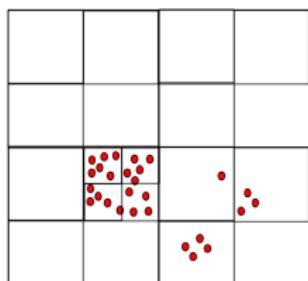


Raw points vs
Ofuscated points.

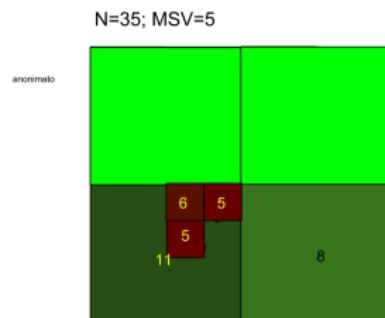


Ética (cont.)

- Obfuscacao espacial: altera, substitui ou generaliza a localizacao dos utilizadores, por forma a esconder a sua localizacao real.
- Maximizar a qualidade da informacao, para um grau minimo de anonimidade

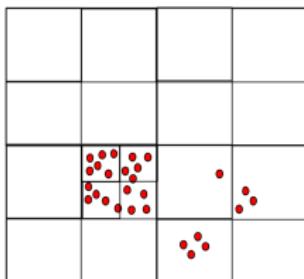


Raw points vs
Ofuscated points.

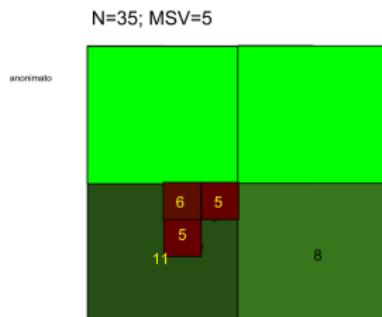


Ética (cont.)

- Obfuscacão espacial: altera, substitui ou generaliza a localização dos utilizadores, por forma a esconder a sua localização real.
- Maximizar a qualidade da informação, para um grau mínimo de anonimidade

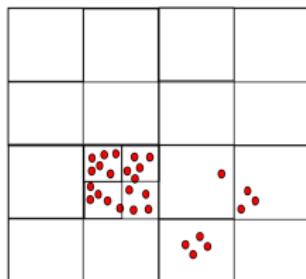


Raw points vs
Ofuscated points.

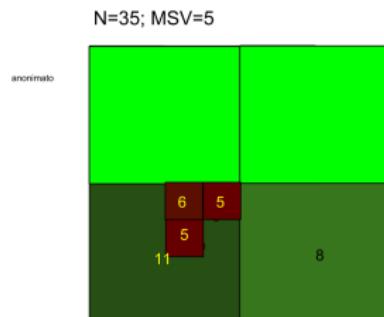


Ética (cont.)

- Obfuscacão espacial: altera, substitui ou generaliza a localização dos utilizadores, por forma a esconder a sua localização real.
- Maximizar a qualidade da informação, para um grau mínimo de anonimidade

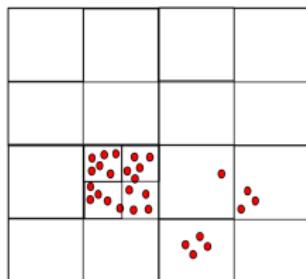


Raw points vs
Ofuscated points.

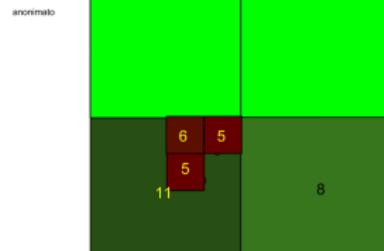


Ética (cont.)

- Obfuscacao espacial: altera, substitui ou generaliza a localizacao dos utilizadores, por forma a esconder a sua localizacao real.
- Maximizar a qualidade da informacao, para um grau minimo de anonimidade
- Transformar os pontos originais (raw points) numa quadricula de densidades, respeitando o MSV.

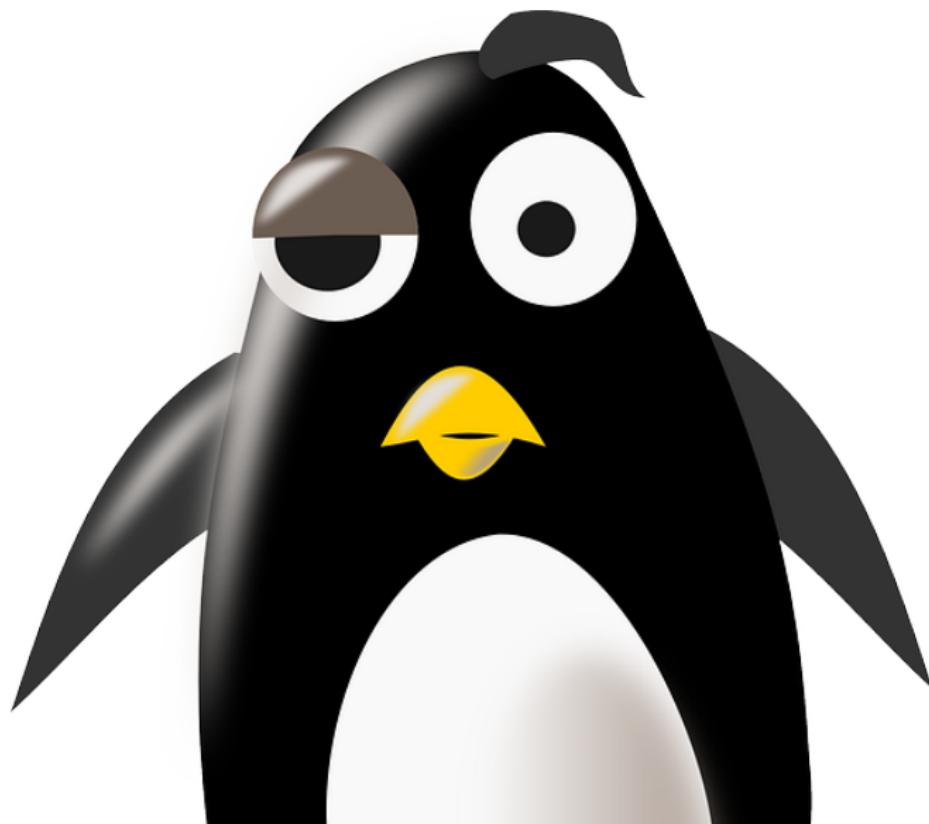


N=35; MSV=5



Raw points vs
Ofuscated points.

SIG Open-Source?



Importância destas Tendências para o FOSS4G

Mais aplicações e mais quantidades de dados: maior comunidade

- Infraestruturas colaborativas de dados.

Importância destas Tendências para o FOSS4G

Mais aplicações e mais quantidades de dados: maior comunidade

- Infraestruturas colaborativas de dados.
- Bibliotecas de ML com capacidades espaciais.
- Visualização (interativa) de dados espaciais (3D).
- Infraestrutura e processamento de Big Spatial Data.
- Ética e privacidade.

Obrigada pela vossa Atenção



Esta apresentação encontra-se disponível em:
<http://tinyurl.com/nfbrhv1>