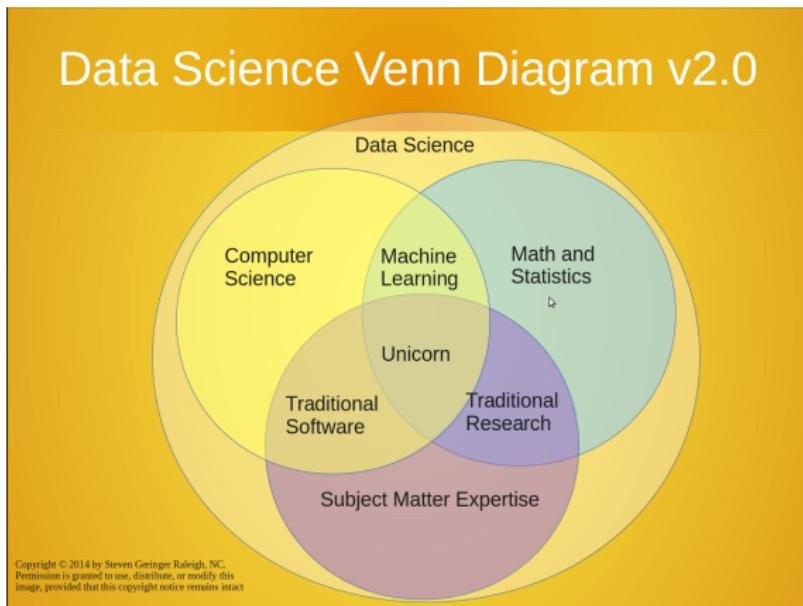
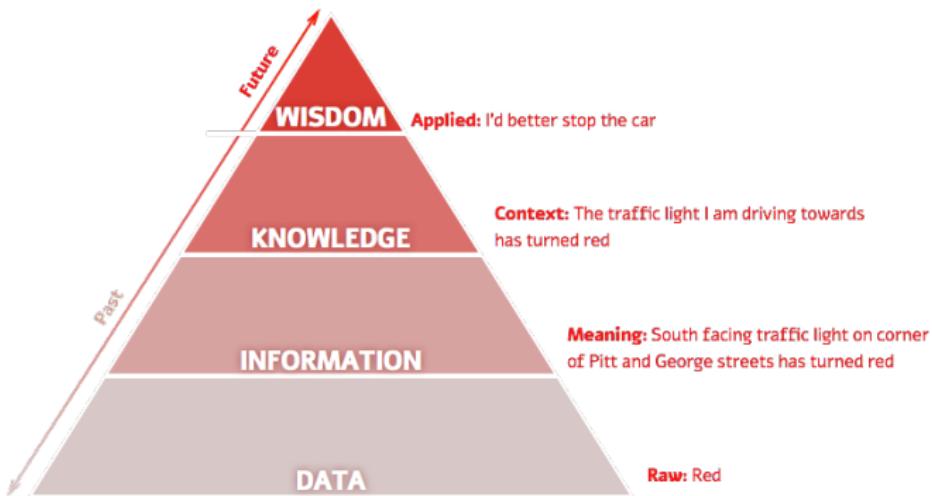


# Cientistas & Unicórnios

- “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)



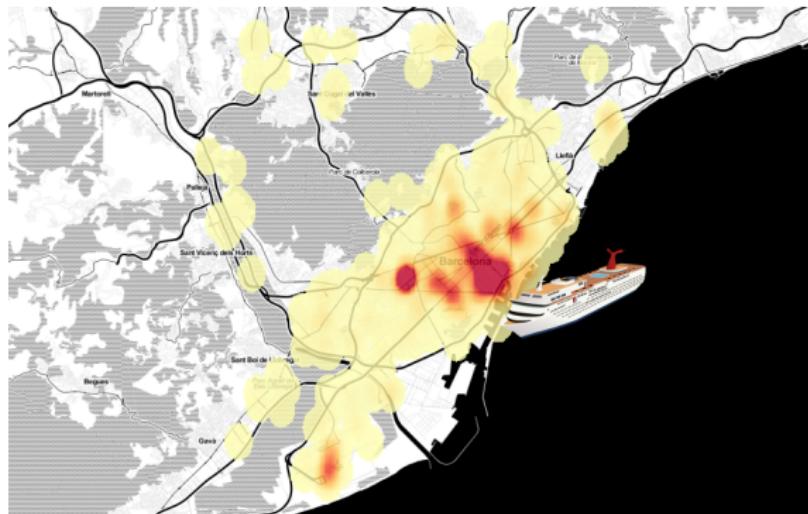
# Dados, Informação e Conhecimento



# Algo espacial...

Lei de Tobler:

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

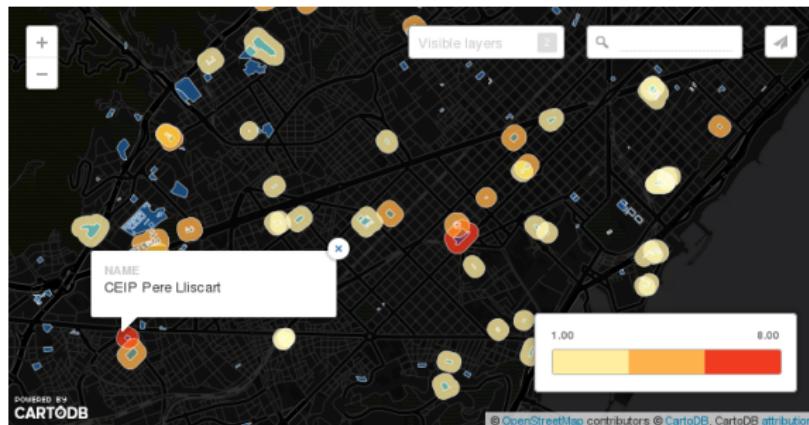


Heatmap de Tweets perto de um Cruzeiro (Eurecat, unpublished).

# Algo espacial...

Lei de Tobler:

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

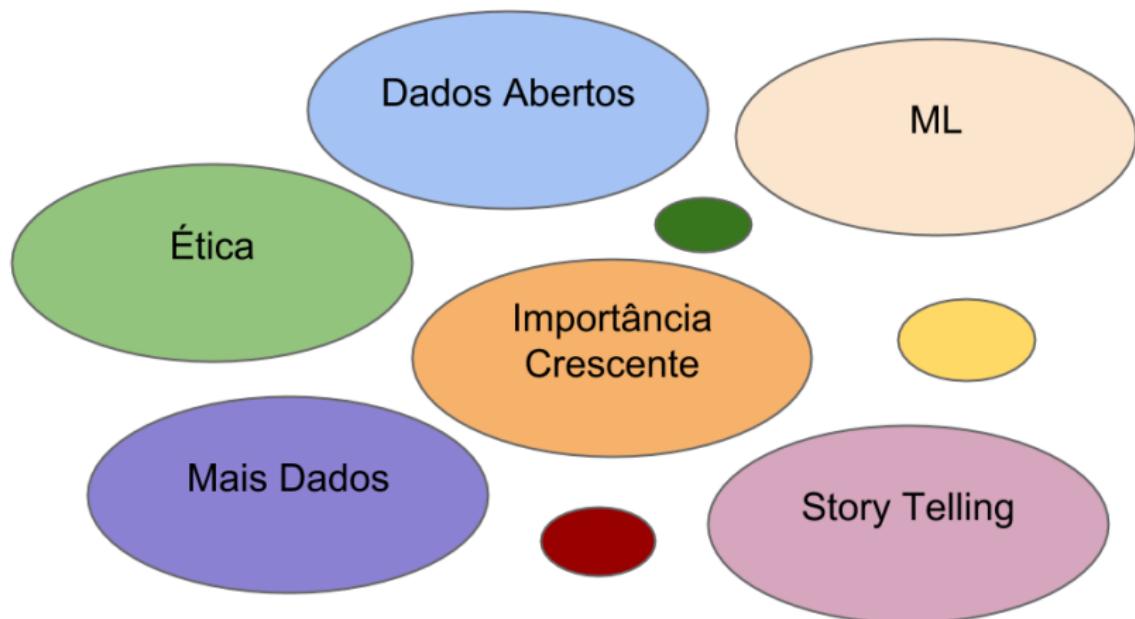


Buffers de acidentes à volta das escolas (Eurecat, unpublished).

# Onde Vamos?



# (Algumas) Tendências



# Importância Crescente

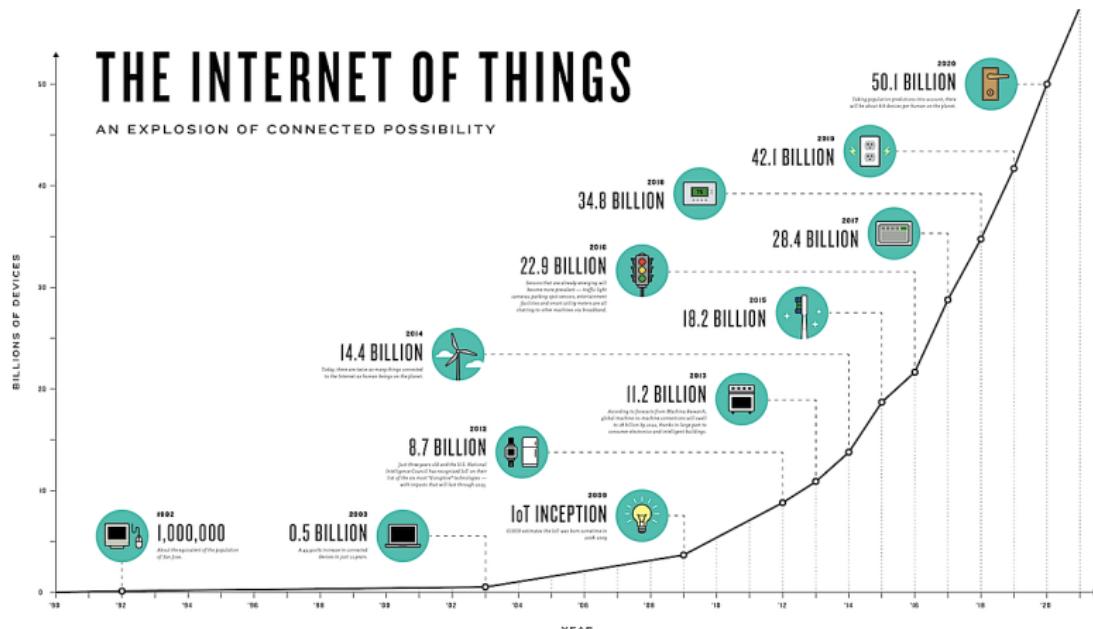


Runkeeper: fitness track app <https://runkeeper.com/>



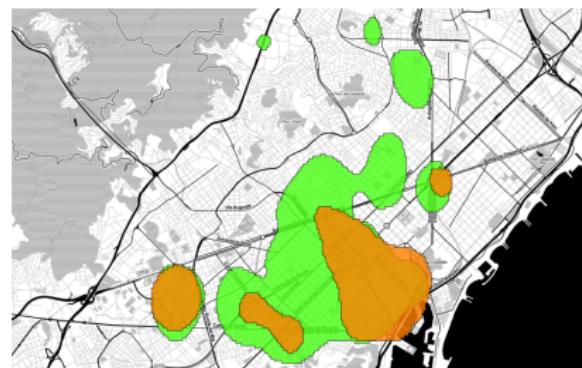
Withings: balança "inteligente" <http://www2.withings.com/eu/en>

# Mais Dados



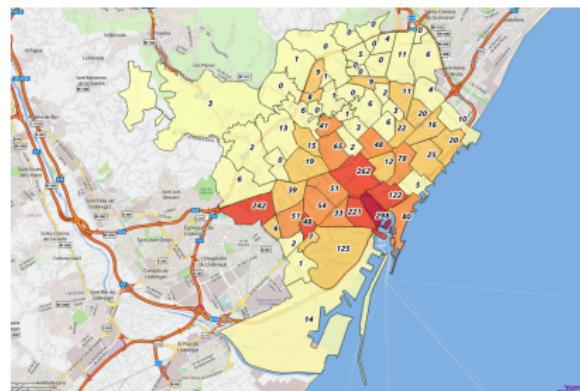
Infografia baseada en dados da Cisco <http://www.i-scoop.eu/internet-of-things/>

# E Ainda Mais Dados



Clusters de Tweets enviados por locais e estrangeiros (Eurecat, unpublished).

## E Ainda Mais Dados



Distribuição de densidades de Tweets enviados por estrangeiros (Eurecat, unpublished).

## “Arqueologia” de Dados



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

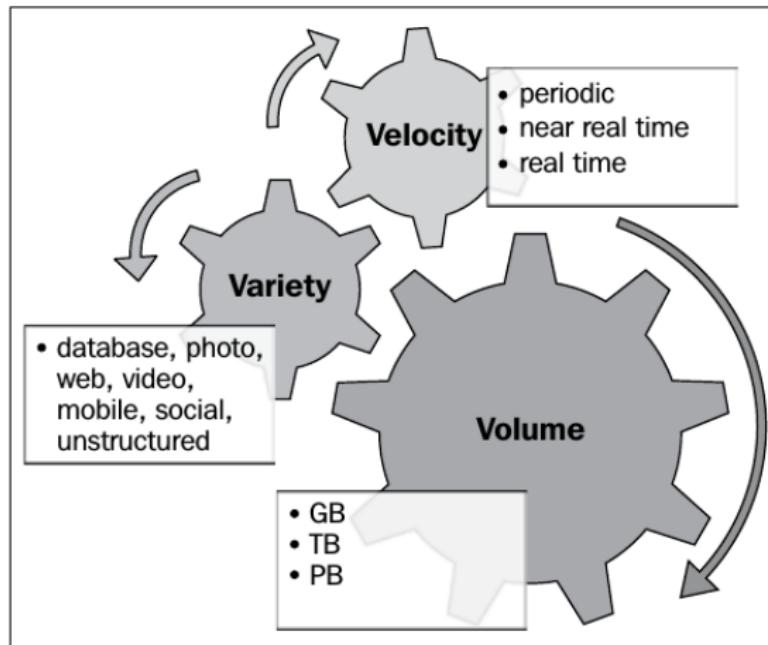
## “Arqueologia” de Dados



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

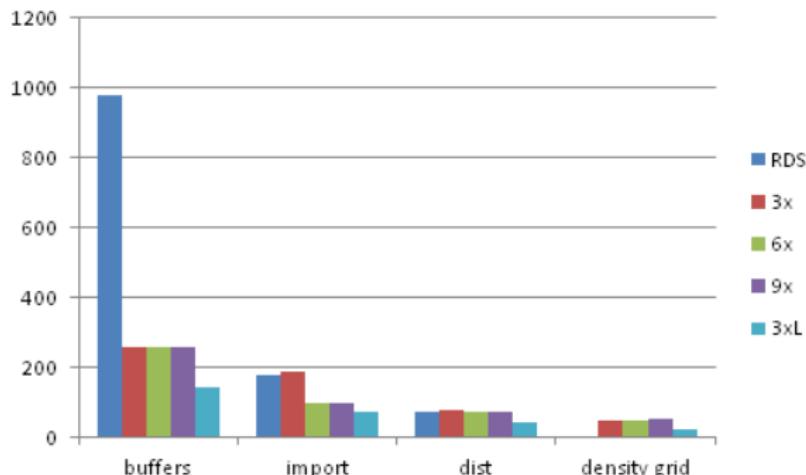
Descodificação de metadados através da estrutura de directórios (Institute of Marine Research, unpublished).

# Tecnologias de *Big Data*



Os 3 Vs de Big Data (Cuesta, H., 2013).

# Tecnologias de *Big Data*



Benchmarking de bases de dados espaciais na cloud (Simoes, 2015).

# Uso Cada vez Mais Generalizado de *ML*

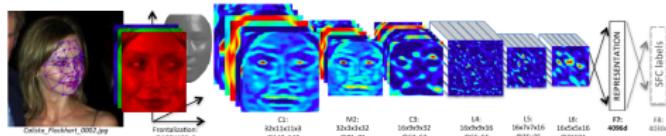
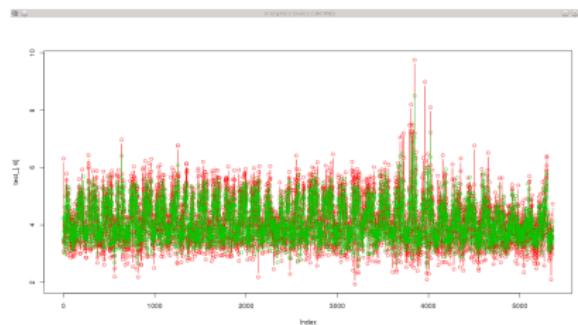


Figure 2. Outline of the DeepFace architecture. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate outputs for each layer. The net includes more than 120 million parameters,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

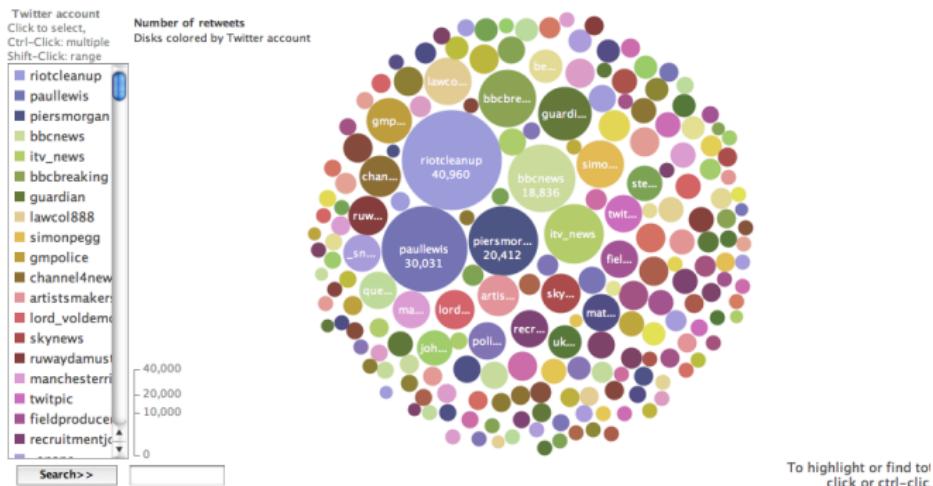
Arquitectura de uma red de Deep Learning para reconhecimento facial

<https://gigaom.com/2015/03/06/how-paypal-uses-deep-learning-and-detective-work-to-fight-fraud/>.



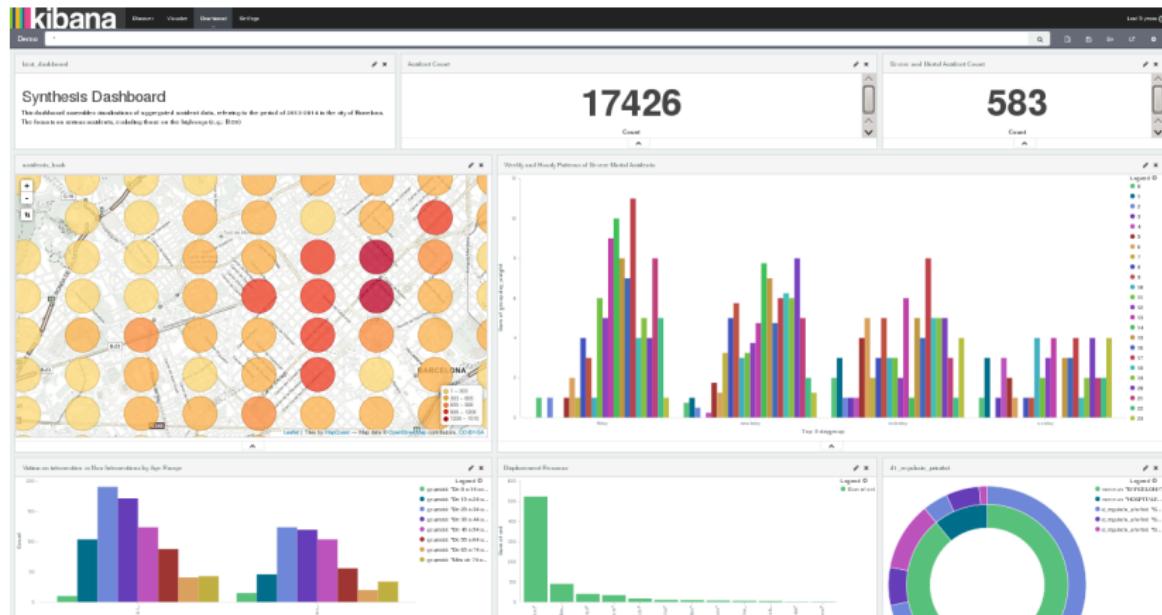
Tempos de viagem na cidade de Barcelona: ajuste entre as previsões SVM (verde) e os valores observados (vermelho) (Eurecat, unpublished).

# Jornalismo de Dados & Story Telling



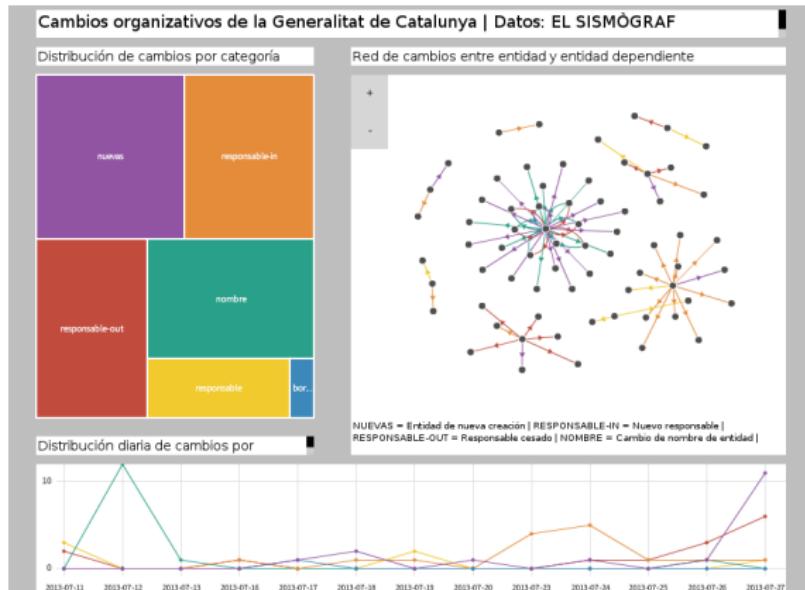
How riot rumours spread on Twitter (Guardian Interactive team, 2011).

# Jornalismo de Dados & Story Telling



Dashboard sobre acidentes graves e mortais na cidade de Barcelona (Eurecat, unpublished).

# Mais (e Melhores) Dados Abertos



Sismógrafo da Catalunya, @OpenGovCat (OpenGov, 2015)

# Ética

Ofuscação espacial (Genovese, A., 2012):

# Ética

Ofuscação espacial (Genovese, A., 2012):

- Anonymity set

# Ética

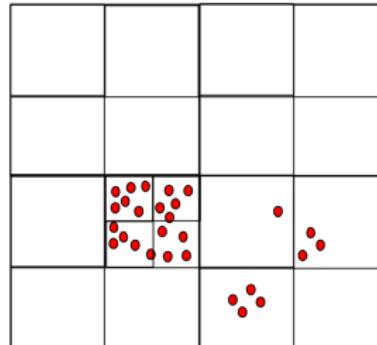
Ofuscação espacial (Genovese, A., 2012):

- Anonymity set
- MSV

# Ética

Ofuscação espacial (Genovese, A., 2012):

- Annonimity set
- MSV



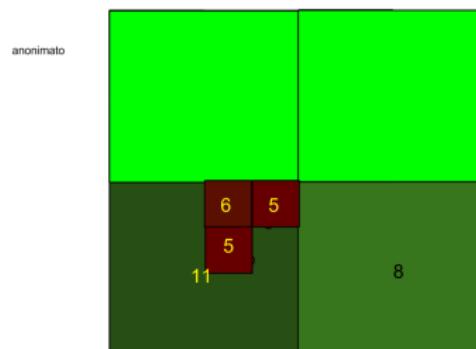
Pontos originais (Eurecat, unpublished).

# Ética

Ofuscação espacial (Genovese, A., 2012):

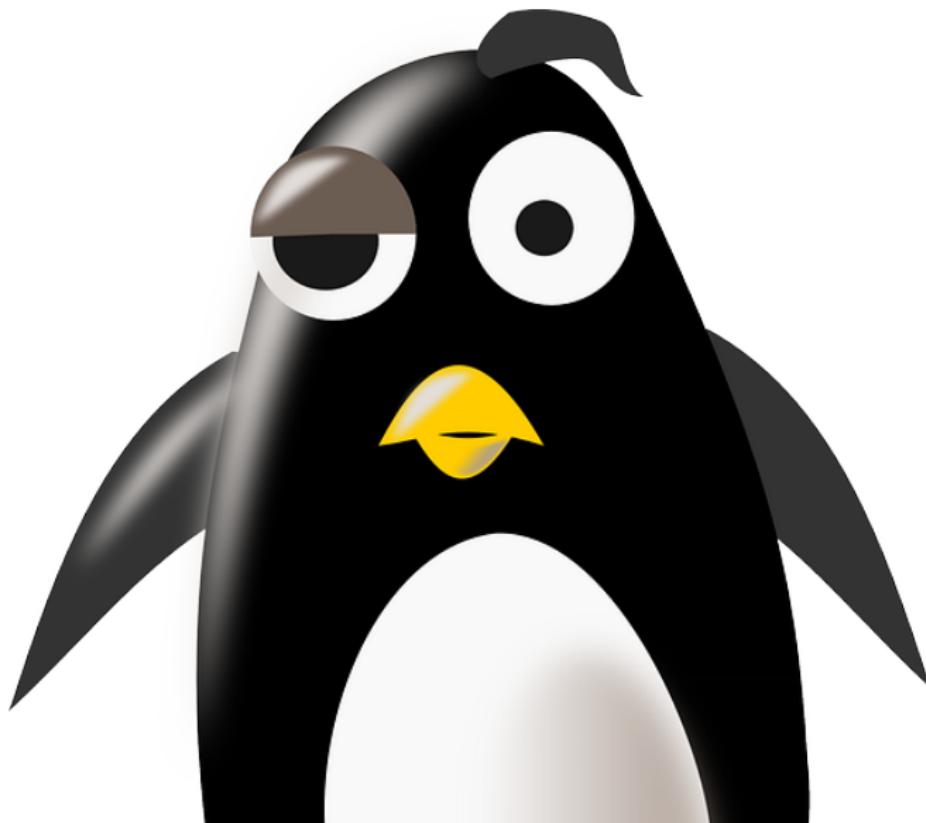
- Annonimity set
- MSV

N=35; MSV=5



Pontos ofuscados (Eurecat, unpublished).

# SIG Open-Source?



# Importância destas Tendências para o FOSS4G

Mais aplicações + maior quantidades de dados = maior comunidade



# Importância destas Tendências para o FOSS4G

Mais aplicações + maior quantidades de dados = maior comunidade

- Infra estruturas colaborativas de dados.
- Bibliotecas de ML com capacidades espaciais.
- Visualização (interactiva) de dados espaciais (3D).
- Infra estrutura e processamento de Big Spatial Data.
- Ética e privacidade.



# Obrigada pela vossa Atenção

Esta apresentação encontra-se disponível em:

<http://tinyurl.com/nfbrhvl>



# Referências

- Simoes, J., 2015. *Visualizing Geolocated Tweets: A Spatial Data Mining Approach*. Presentation at the 4th Data beers BCN, in Barcelona. Available at:  
[https://github.com/doublebyte1/data\\_beers/blob/master/data\\_beers.pdf](https://github.com/doublebyte1/data_beers/blob/master/data_beers.pdf)
- Johnson, S., 2006. The Ghost Map: The Story of London's Most Terrifying Epidemic —and How it Changed Science, Cities and the Modern World. Riverhead Books. ISBN 1-59448-925-4
- Cuesta, H., 2013. Practical Data Analysis. PACKT Publishing.
- Simoes, J., Giménez, R., Planagumà, M., 2015. *Big Data y Bases de Datos Espaciales: una análisis comparativo*. Presentación en las 9as Jornadas SIG Libre, Girona.
- Guardian Interactive team, 2011. *Behind the rumours: how we built our Twitter riots interactive*. Available at: <http://www.theguardian.com/news/datablog/2011/dec/08/twitter-riots-interactive>
- OpenGov, 2015. *The seismograph's odyssey and transparency for show*. Available at:  
<http://opengov.cat/en/2015/08/the-seismographs-odyssey-and-transparency-for-show/>
- Genovese, A. 2012. Obfuscation-based techniques for privacy protection in location-based systems: a comparison of recent methods. Dissertation presented for the degree of Doctor in Computer Science, in the Universita Degli Studio di Milano.