

Big Data

Stan Matwin

Canada Research Chair

Director

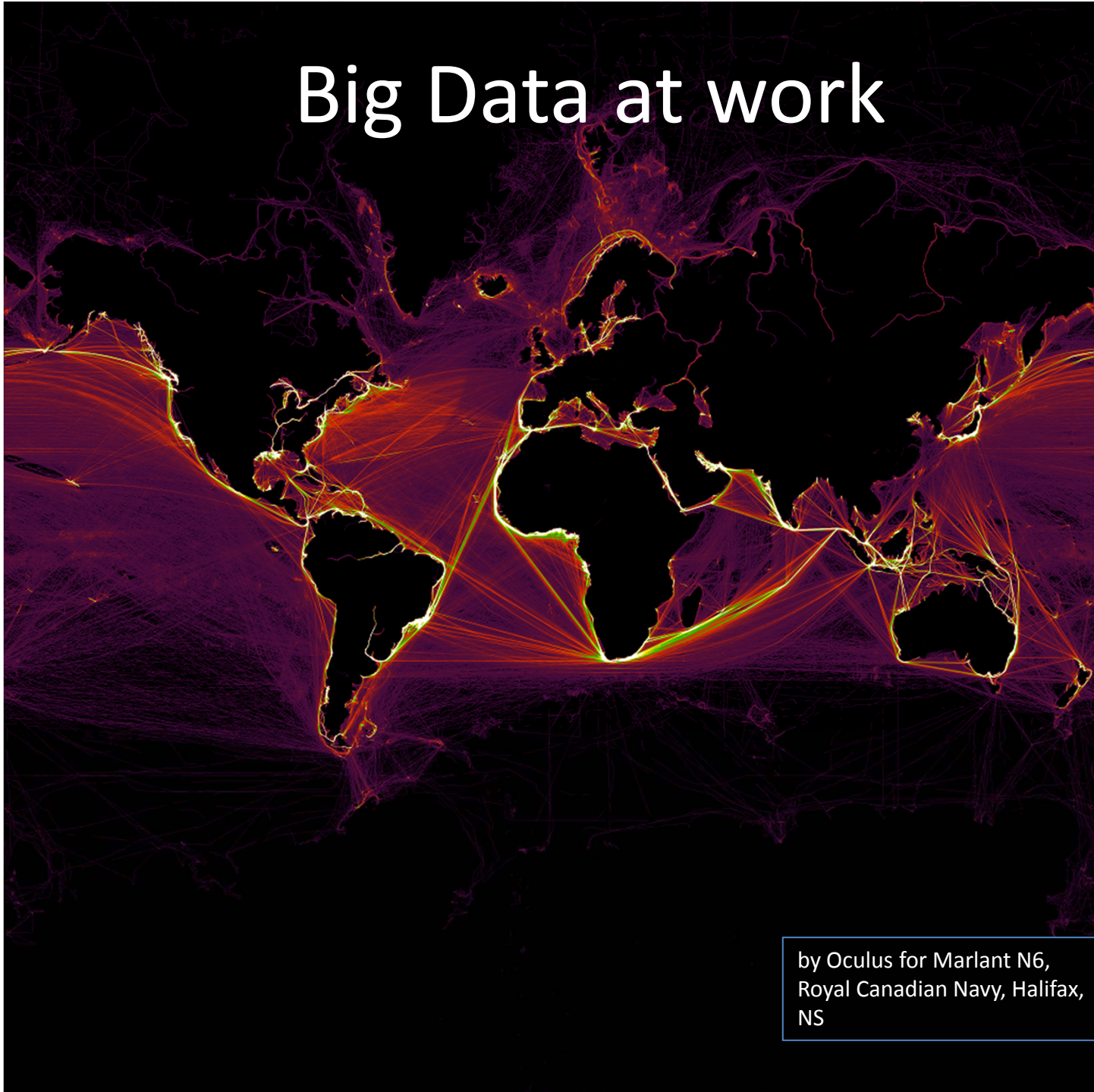
Big Data Analytics Institute

Productivity Design, May 23, 2014

stan@cs.dal.ca



Big Data at work



by Oculus for Marlant N6,
Royal Canadian Navy, Halifax,
NS



1 NEW DEFINITION IS ADDED ON Urban Dictionary

1,600+ READS ON Scribd

13,000+ HOURS MUSIC STREAMING ON PANDORA

12,000+ NEW ADS POSTED ON craigslist

370,000+ MINUTES VOICE CALLS ON skype

98,000+ TWEETS

20,000+ NEW POSTS ON tumblr.

13,000+ iPhone APPLICATIONS DOWNLOADED

320+ NEW twitter ACCOUNTS

100+ NEW LinkedIn ACCOUNTS

1 associatedcontent NEW ARTICLE IS PUBLISHED

THE WORLD'S LARGEST COMMUNITY CREATED CONTENT!!

QUESTIONS ASKED ON THE INTERNET...

100+ Answers.com 40+ YAHOO! ANSWERS

25+ HOURS TOTAL DURATION

600+ NEW VIDEOS

70+ DOMAINS REGISTERED

60+ NEW BLOGS

1,500+ BLOG POSTS

168 MILLION EMAILS ARE SENT

694,445 SEARCH QUERIES

1,700+ Firefox DOWNLOADS

695,000+ facebook STATUS UPDATES

50+ WORDPRESS DOWNLOADS

79,364 WALL POSTS

125+ PLUGIN DOWNLOADS

510,040 COMMENTS

- McKinsey anticipates shortage of 140,000-190,000 “deep analytical positions” in the US by 2018
- Davos World Economic Forum – “big data” creates unprecedented opportunities for development for international development

What?

- Maturing of several technologies:
 - Data capture
 - Data bases
 - Data mining
 - High-Performance Computing, particularly **the cloud**
 - Visualization
- Coming together to “assetize” data

What...

Machine learning....

1980s

Data mining....

2000...

Big data...

2013...

5 V's of Big Data

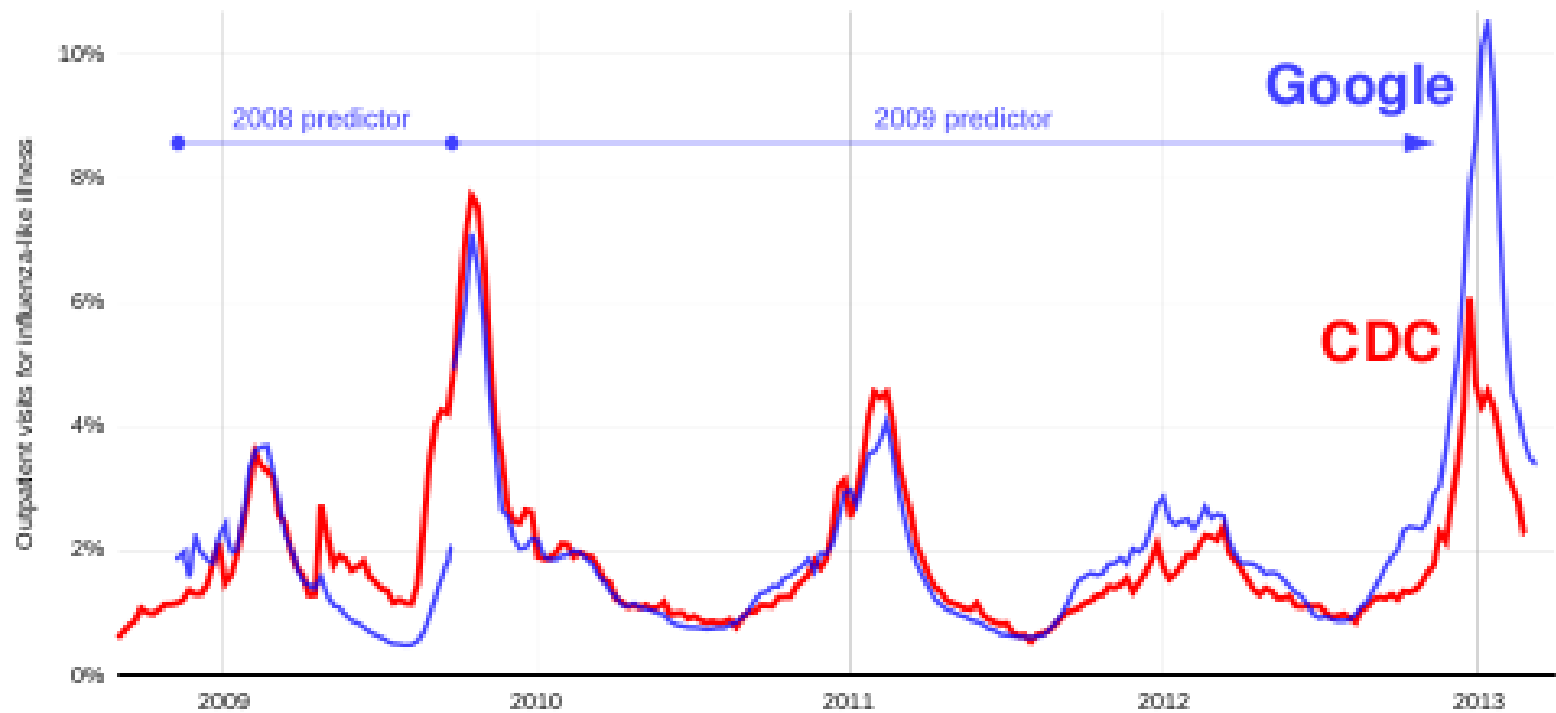
- Volume
- Velocity
- Variety
- Veracity
- ... and Value

Some successful Big Data examples...

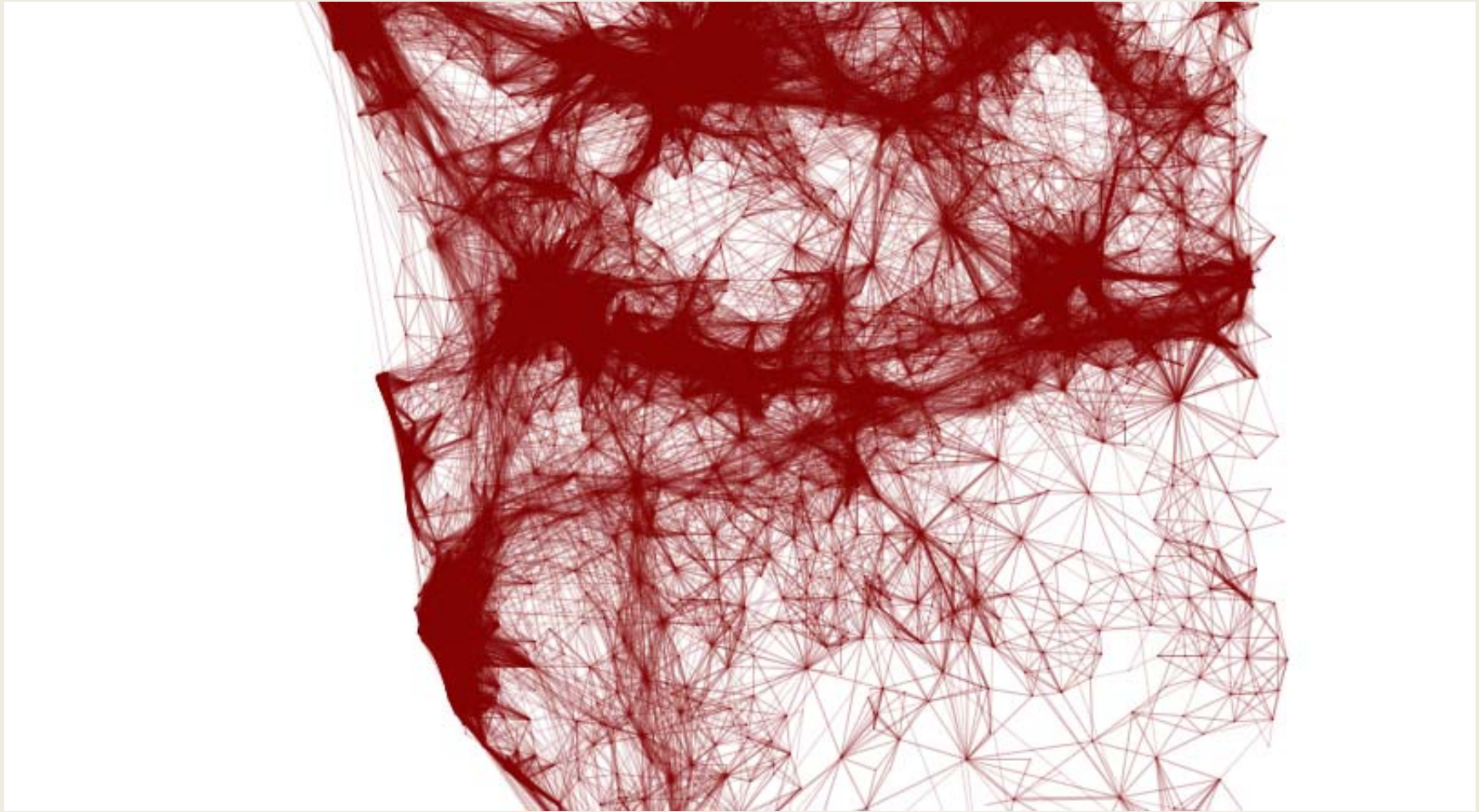
- Google flu [search engine]
- Tuscan regions [GPS/mobile phones]
- Happiest city [twitter]

Google flu

Second divergence in 2012–2013 for U.S.



People mobility in Tuscany



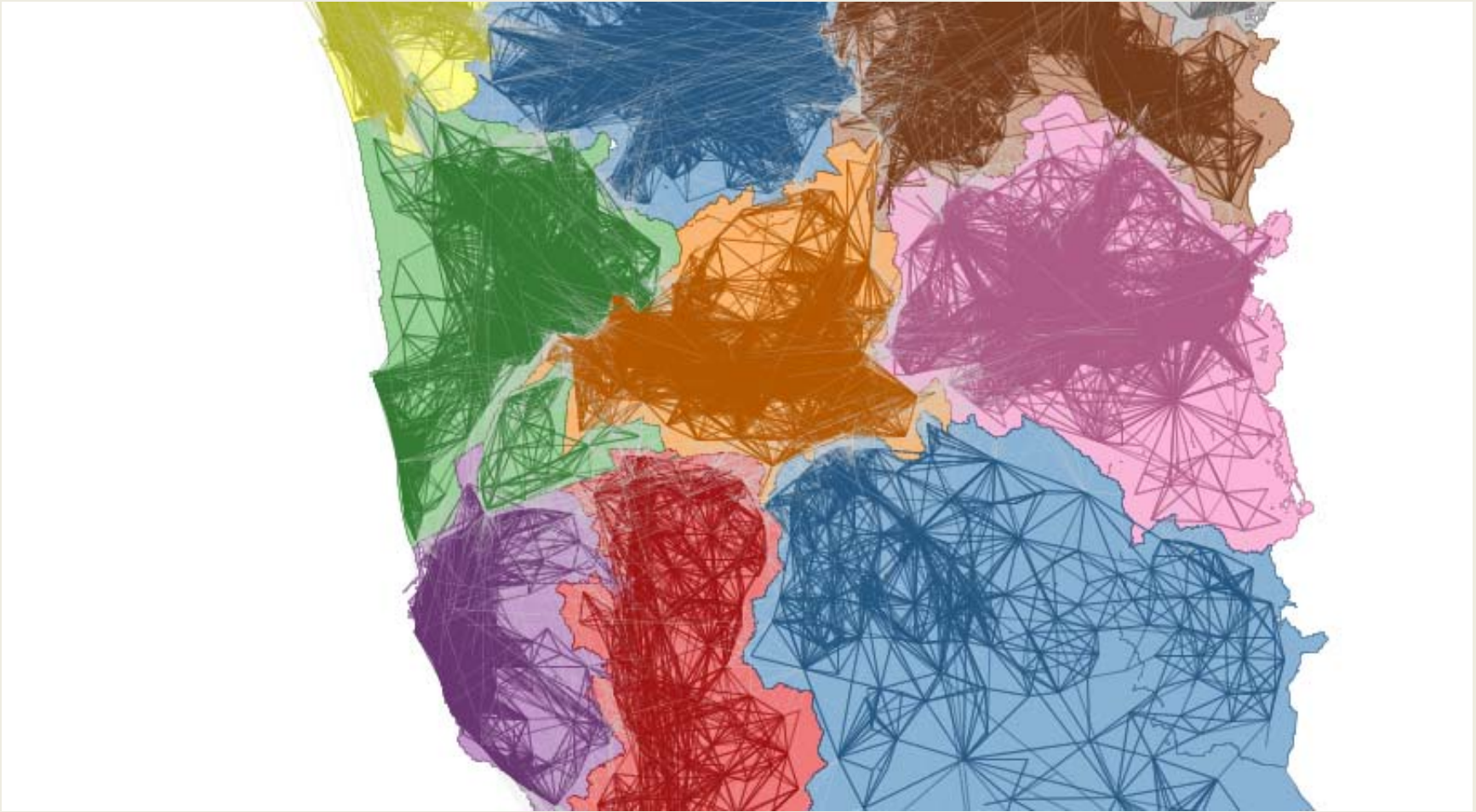
Community detection



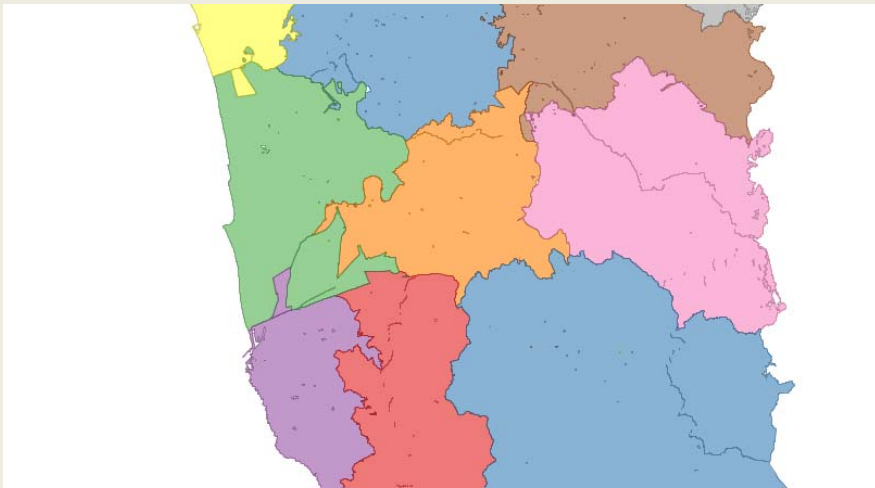
back to geography



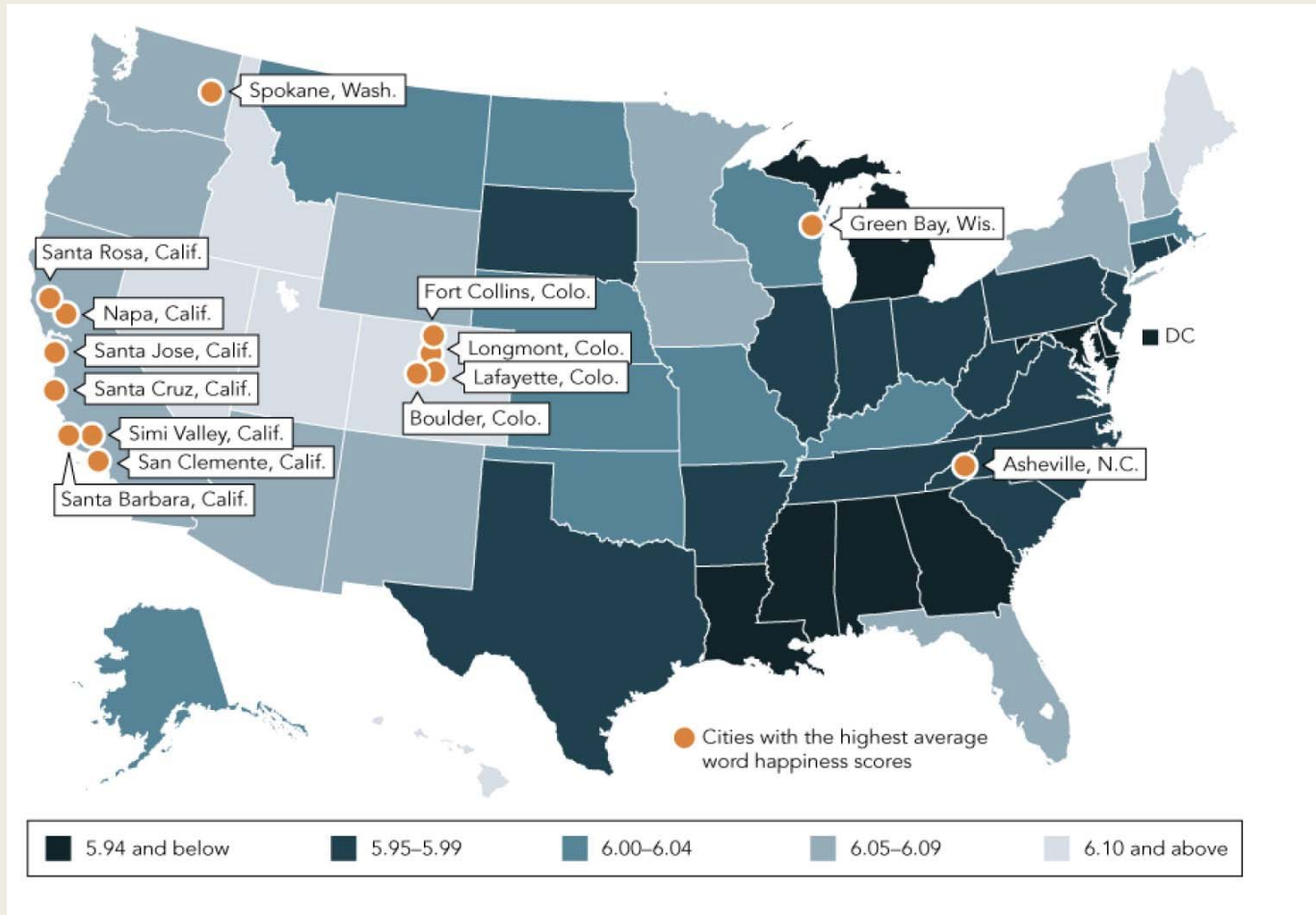
draw borders



Final result



Happiest city - sentiment analysis on twitter

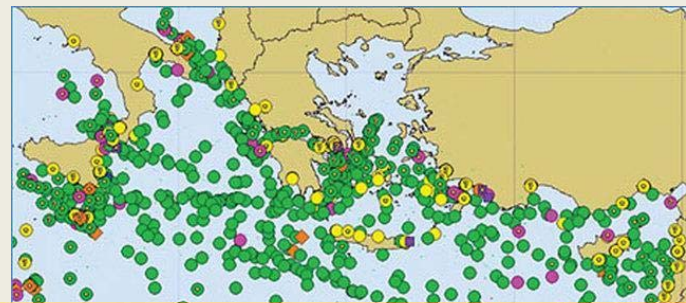


How - Our model

- You have the data and a question/problem
- We help you answering it using state-of-the-art tools we have
- Do all this in a privacy-respectful way
- We are interested in the research aspect of such projects
- We train students through thesis topics inspired by your R&D needs

Intelligent Maritime Solutions from AIS Satellite Data

- AIS data is produced 24/7
- How can it be used to answer questions such:
 - What is a typical pattern of movement (trajectory) of ship fishing for tuna (as opposed to halibut)?
 - what is the projected traffic of ships at the Dubai port entrance in 48 hours?
 - Has this ship been to area X in the last 90 days?
- Challenges of this project
 - Parsing the data
 - Volume of the data...
- Brazilian collaborators



Tissue sales...

- POS data
- How to increase tissue sales?
- Some of the questions
 - Demographics
 - Flu/allergies
- Big challenge – the Variety V



TRIBE: Training in Big Text Data: five pillars

- New grad program in CS at Dal
- A Graduate Certificate
 - Structured curriculum
 - Project
 - Industrial internship
 - Student mobility
 - Respect for data privacy

Q&A, discussion...

