NAMSOR DEEP LEARNING FOR MIGRATION ANALYTICS : DECRYPTING IDENTITY IN SPACE AND TIME THROUGH PERSONAL NAMES, GEOGRAPHIC, SEMANTIC, SOCIAL GRAPH

2018-01

Elian CARSENAT, NamSor

Founder Bio

Elian CARSENAT, a computer scientist trained at ENSIIE/INRIA, started his career at JP Morgan in Paris in 1997. He later worked as consultant and managed business & IT projects in London, Paris, Moscow and Shanghai.

In 2012, Elian created **NamSor**, a piece of sociolinguistics software to mine the 'Big Data' and better understand international flows of money, ideas and people.

http://fr.linkedin.com/in/eliancarsenat/en

NamSor sorts Names

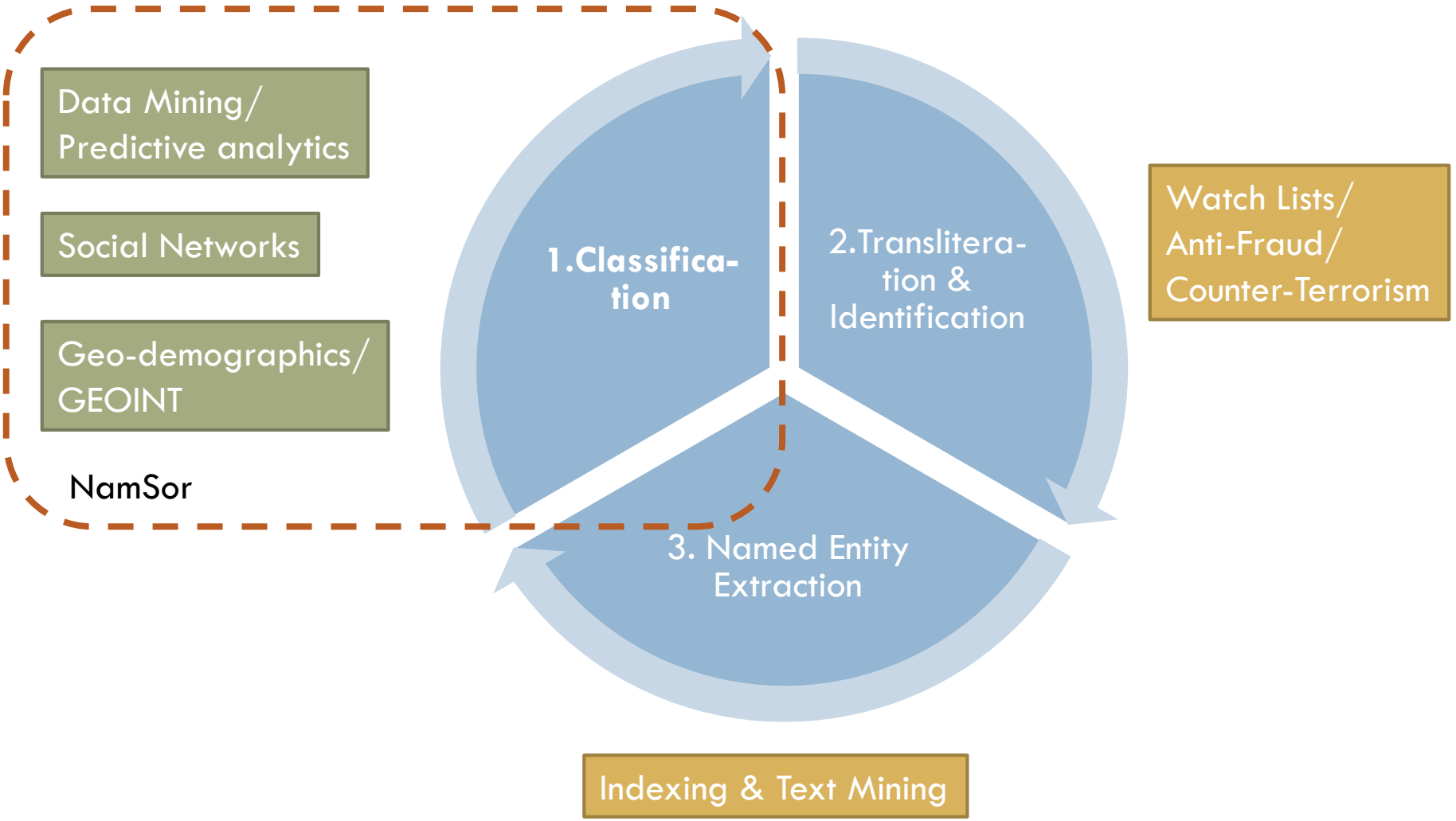
NamSor Origin & Diasporas

- Classification with various taxonomies
  - Gender (female/ male / unknown)
  - Script (LATIN, ARABIC, GUJARATI,...)
  - Origine (Country ex. France vs. Inde)
  - Region (ex. Gujarat vs. Andhra Pradesh)
  - Diaspora (ex. Indian Diaspora in US vs Indian Diaspora in Mauricius)
- Sorting according to a numerical score, allowing combining NamSor with other algorithm (graph, semantics, predictive …)
- Flexibility to learn new taxonomies (machine learn.)
- Ease of integration (NamSor API, Java/Python SDK, ESRI, RapidMiner, NationBuilder …)

A global coverage -142+ countries

DIMENSION	CURRENT COVERAGE
SCRIPT (22)	LATIN, ARABIC, CYRILLIC, ARMENIAN, BENGALI, DEVANAGARI, GEORGIAN, GREEK, GUJARATI, GURMUKHI, HAN, HANGUL, HEBREW, HIRAGANA, KANNADA, KATAKANA, MALAYALAM, MYANMAR, ORIYA, TAMIL, TELUGU, THAI
COUNTRY (142+)	AE, AF, AL, AM, AO, AR, AT, AZ, BA, BD, BE, BF, BG, BH, BI, BJ, BN, BR, BT, BW, BY, CA, CD, CF, CG, CH, CI, CL, CM, CN, CO, CR, CV, CY, CZ, DE, DK, DZ, EE, EG, ER, ES, ET, FI, FJ, FR, GA, GB, GE, GH, GM, GN, GR, HK, HR, HT, HU, ID, IE, IL, IN, IQ, IR, IS, IT, JO, JP, KE, KG, KH, KM, KP, KR, KW, KZ, LA, LB, LK, LR, LS, LT, LU, LV, LY, MA, MD, ME, MG, MK, ML, MM, MN, MR, MU, MV, MW, MX, MY, MZ, NA, NE, NG, NL, NO, NP, OM, PE, PH, PK, PL, PS, PT, QA, RO, RS, RU, RW, SA, SD, SE, SI, SK, SN, SO, SR, SY, TD, TG, TH, TJ, TM, TN, TO, TR, TT, TW, TZ, UA, UG, US, UZ, VE, VN, YE, ZA, ZM, ZW
COUNTRY/ REGION (15)	RU (80), IN (~30), FR (22), IT (17), LB (14), BF (13), CD (8), TR (7), ID (7), GB (4), ES (17), ML (50), GN (8), CI (34), AF(16)
COUNTRY/ DIASPORA	US, CA, SG, GB, (EU)

NamSor can enrich any nominative data



Two complementary approaches

NamSor CORE (Origin, Diaspora)

- Optimized for global coverage : coding names to a large multi-class taxonomy (all countries / regions /ethnicities)
- The **only input is NAMES** : not other information is required

NamSor ML

- Deep-learning capability to re-train models towards a focused research or a customized taxonomy (binary classifier, or just a few classes)
- **Name information is combined with other data** (geographic, behavioural, semantic …)

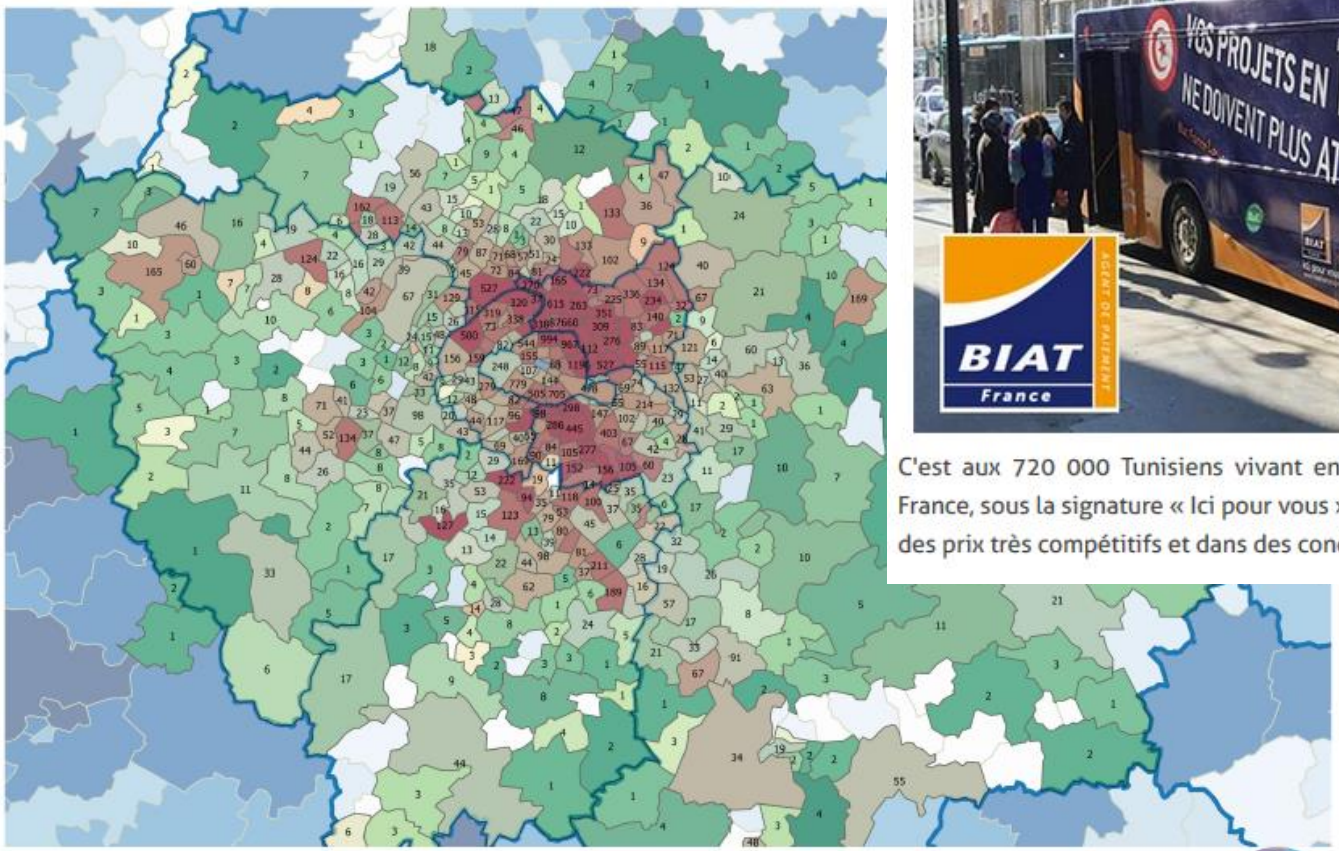
NamSor Core : Mapping Tunisian Diaspora

La BIAT lance un road show pour les tunisiens en île de France

Le 17 mars 2016, BIAT France lance le « BIAT France Tour » et part à la rencontre des Tunisiens résidents à Paris et en région parisienne afin de leur présenter les produits et services qui leur sont destinés.

181 1

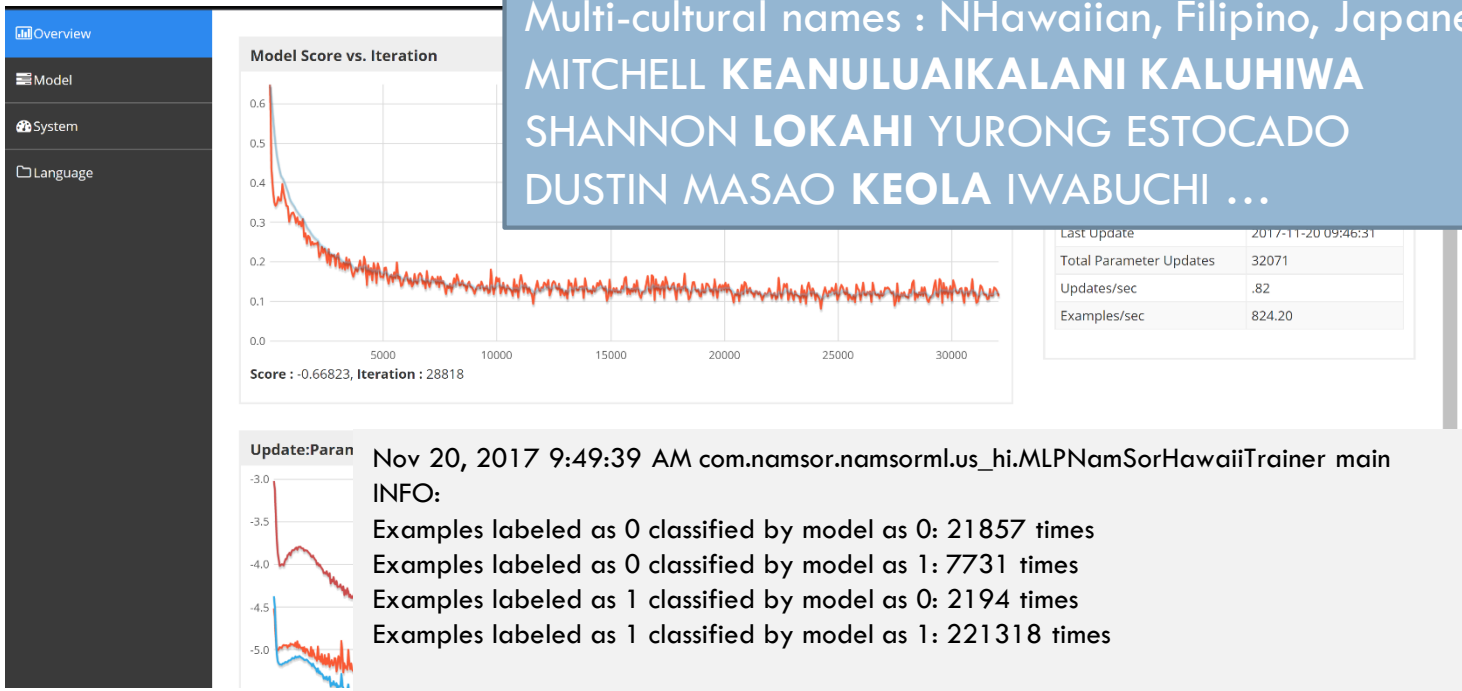
C'est aux 720 000 Tunisiens vivant en France que la BIAT s'adresse avec sa filiale BIAT France, sous la signature « Ici pour vous », un service de transfert d'argent leur est proposé à des prix très compétitifs et dans des conditions de rapidité et de sécurité exemplaires.



NamSor

NamSor ML :
Native Hawaiian Names - A Binary Classifier Example

Multi-cultural names : NHawaiian, Filipino, Japanese, Anglo-Saxon …
MITCHELL **KEANULUAIKALANI KALUHIWA**
SHANNON **LOKAHI** YURONG ESTOCADO
DUSTIN MASAO **KEOLA** IWABUCHI …



Nov 20, 2017 9:49:39 AM com.namsor.namsorml.us_hi.MLPNamSorHawaiiTrainer main
INFO:
Examples labeled as 0 classified by model as 0: 21857 times
Examples labeled as 0 classified by model as 1: 7731 times
Examples labeled as 1 classified by model as 0: 2194 times
Examples labeled as 1 classified by model as 1: 221318 times

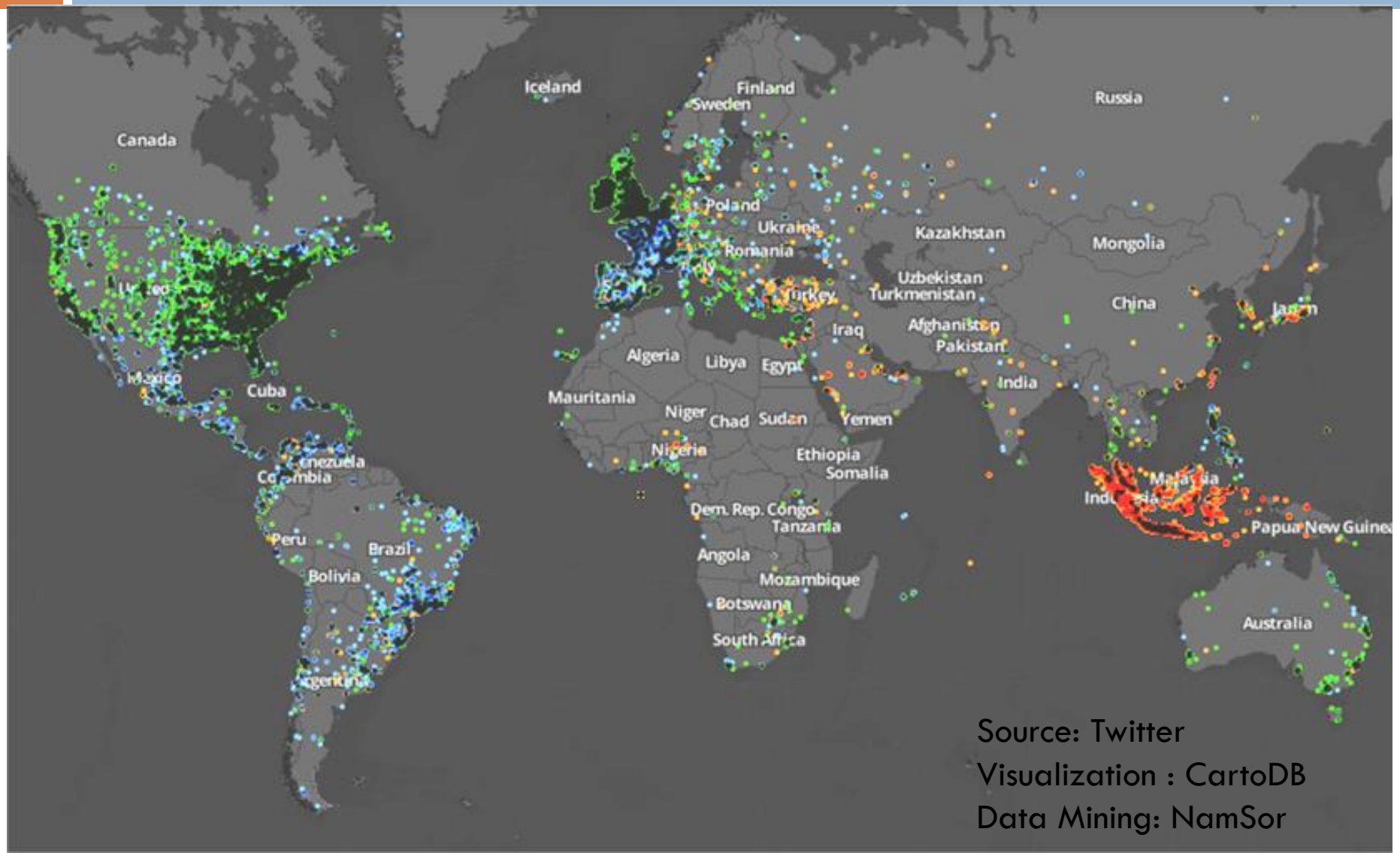
==========Scores==========
# of classes: 2
Accuracy: 0.9608
Precision: 0.9375
Recall: 0.8644
F1 Score: 0.9781
==========

Nov 20, 2017 9:49:39 AM com.namsor.namsorml.us_hi.MLPNamSorHawaiiTrainer main
INFO: **********Example finished**********

- NamSor use cases

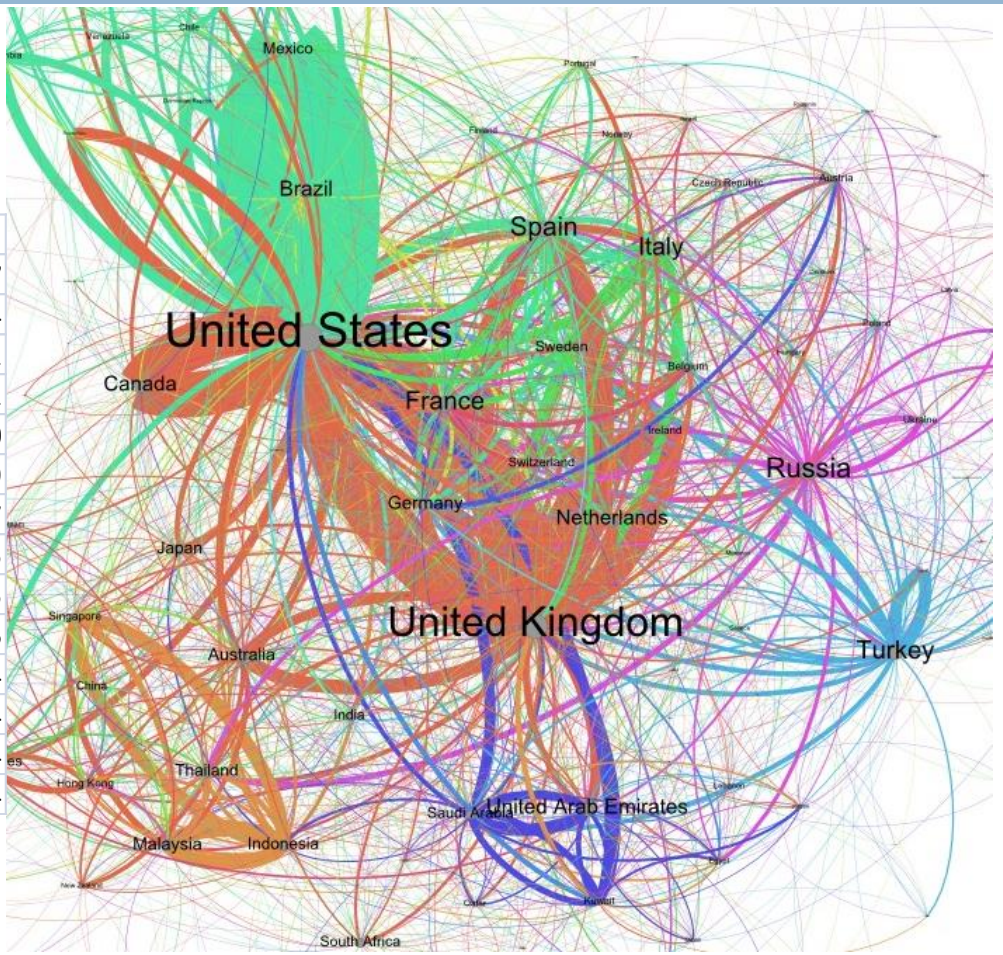
Mining 3M twitter names to map ***Diasporas***

Who *are they,* ***where*** *are they and* ***what*** *are they doing?*



Flow view – who travels where?

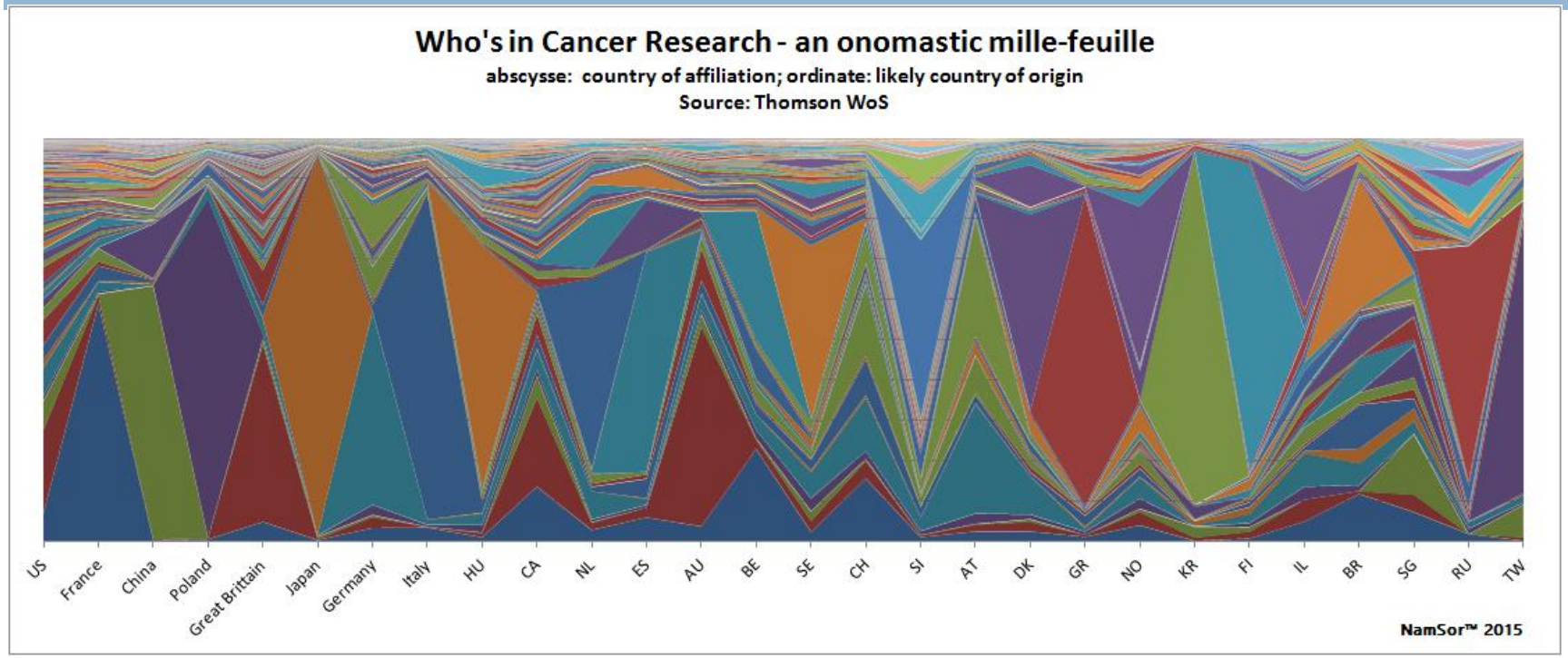
Source	Target	Type	Id	Onoma	Weight
United Kingdom	**France**	**Directed**	**16**	**Great Britain**	**37**
Spain	France	Directed	55	Spain	14
United States	France	Directed	75	Great Britain	12
Turkey	France	Directed	79	Turkey	11
Brazil	France	Directed	87	Portugal	10
United Kingdom	France	Directed	112	Ireland	9
Italy	France	Directed	152	Italy	7
Switzerland	France	Directed	226	France	5
Belgium	France	Directed	247	France	5
United Kingdom	**France**	**Directed**	**258**	**France**	**5**
Mexico	France	Directed	287	Spain	4
Ireland	France	Directed	317	Great Britain	4
United Kingdom	France	Directed	333	Italy	4
United States	France	Directed	375	France	4



Source: Twitter
Visualization : Gephi
Data Mining: NamSor

Mapping Talents in Cancer Research

(in collaboration with French INSERM)



Thomson Reuters WebOfScience (6 countries, 250k scientists, 50k papers)

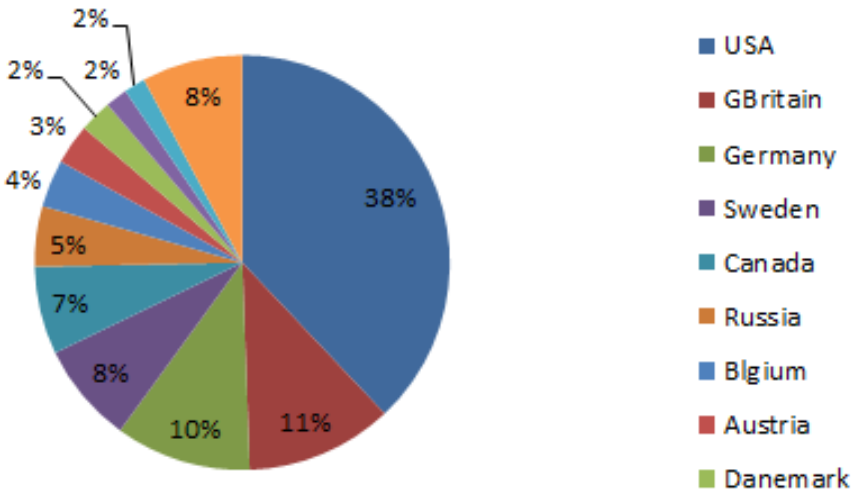
"Analysts uncovered amazing patterns in the way scientists' names correlate with whom they publish, and who they cite in their papers - not just in case of a particular country, but globally. Tania Vichnevskaia of the French National Institute for Health (INSERM) presented the paper 'Applying onomastics to scientometrics' at IREG International symposium 2015 organised by University of Maribor and Shanghai Jiao Tong University. The paper was prepared jointly with NamSor, a private start-up company specialized in mapping international Diasporas."

Source: WoS; Data Mining: INSERM with NamSor

Cancer Research in Poland and Slovenia

Examining the 'brain drain'

The Polish "brain drain"

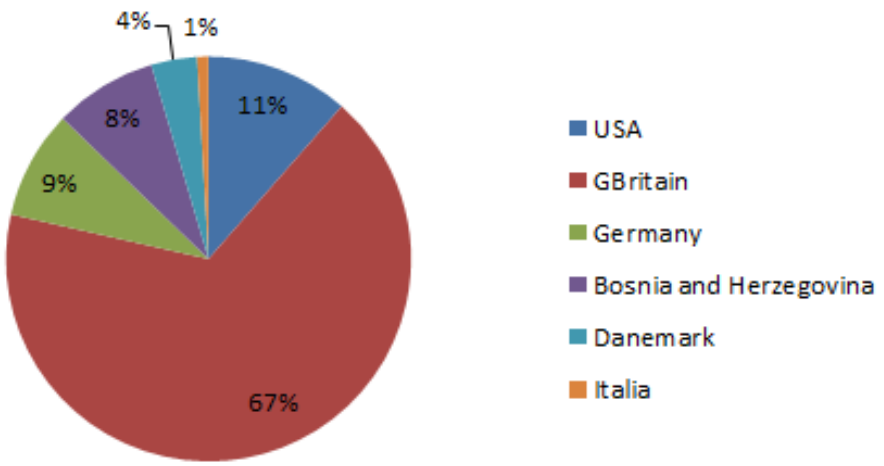


In the Polish Corpus, we look at co-authors with Polish names, affiliated abroad.

Top countries:

1. **US,**
2. **Great-Britain,**
3. **Germany.**

The Slovenien "brain drain"



In the Slovenian Corpus, we look at co-authors with Slovenian names, affiliated abroad.

Top countries:

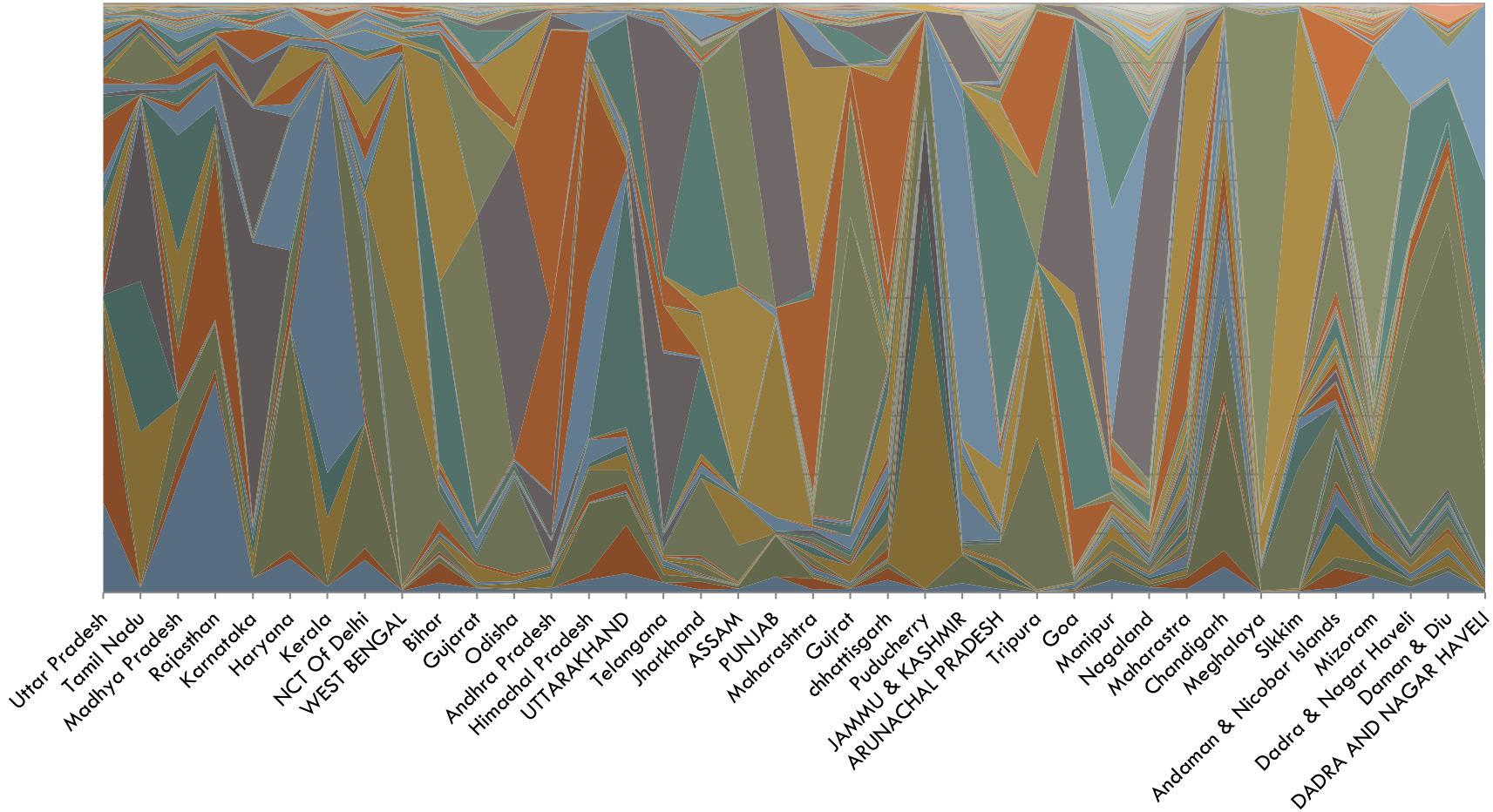
1. **Great-Britain,**
2. **US,**
3. **Germany.**

Source: WoS; Data Mining: INSERM with NamSor

“Incredible India” – 1.2 BN People

Indian onomastics by State/Union Territory

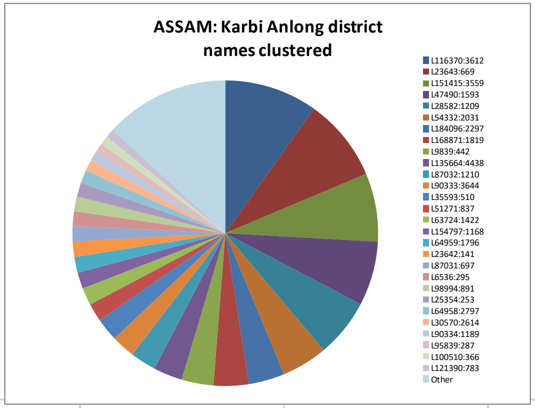
Names in LATIN, BENGALI, DEVANAGARI, GUJARATI, GURMUKHI, KANNADA, MALAYALAM, ORIYA, TAMIL, TELUGU, ARABIC



ASSAM: Karbi Anglong, within district

Inter-caste marriages ?

output		Input			Input	
clusterId	clusterParentId	Firstname	LastName	parent is	FirstParen	LastParent
L25354:25:	L64958:2797	Aà[\ ¹	¹}[ššã	husband	¤à¡[Wª	[W}>à
L47490:15!	L64958:2797	¤àK[¹	[W}>à	father	¤à¡[Wª	[W}>à
L28582:12(	L47490:1593	[³>à	Ò}[t ¡šã	husband	¤àK[¹	[W}>à
L23643:66!	L35593:510	™à∫}à	[W}>à šã	father	¤à¡[Wª	[W}>à
L23643:66!	L35593:510	³ à>àÒü	[W}>à šã	father	¤à¡[Wª	[W}>à
L47490:15!	L35593:510	Vàl=¢	[W}>à	father	Vå¤	[W}>à
L23643:66!	L35593:510	Aà¹	t ¡à¹ïšã	husband	Vå¤	[W}>à
L35593:51(	L47490:1593	[ƒ[º š	[W}>à	father	Vàl¤	[W}>à
L23643:66!	L47490:1593	[¹>à	[W}>à šã	father	Vàl¤	[W}>à



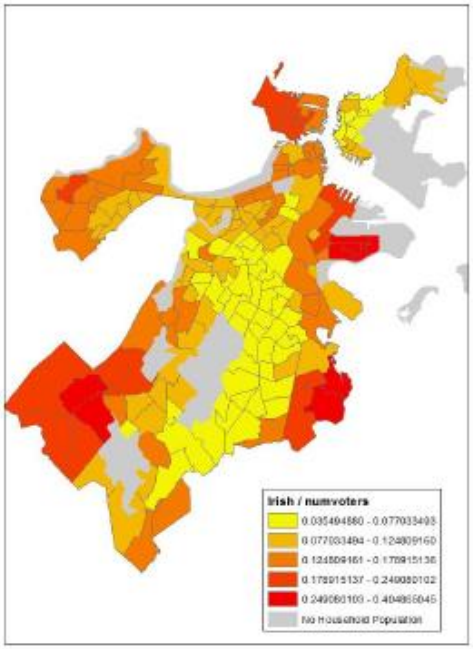
parent is	husband						
Count of serial	**Column Labels**						
Row Labels	**L47490:1593**	**L116370:3612**	**L54332:2031**	**L184096:2297**	**L35593:510**	**L168871:1819**	**L135664:4438**	**L51271:837**
L23643:669	6931	84	5099	15	2069	28	791	1924
L151415:3559	18	212	11	6446	19	1217	55	6
L28582:1209	5132	68	3565	10	1494	17	592	1323
L116370:3612	66	10283	38	72	40	321	137	29
L9839:442	2491	60	1851	9	774	11	321	660
L168871:1819	7	263	6	361	8	2730	24	4
L23642:141	1198	8	822	2	375	4	156	332
L25354:253	1181	12	932		375	7	100	323
L135664:4438	20	154	5	22	19	44	2212	3
L87032:1210	11	315	13	51	14	141	37	9
L90333:3644	3	204	2	31		190	5	
L184096:2297		13		1735	3	84	11	1
L87031:697	4	136	4	12	3	137	4	5
L14495:131	614	10	432		167	4	68	163
L63724:1422	17	83	10	34	34	28	96	6
L98994:891	31	161	46	21	19	59	21	5

Source: Voters List; Data Mining: NamSor

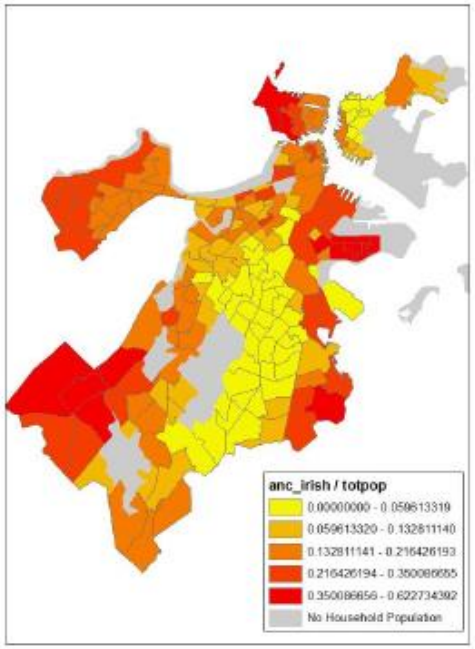
Boston geo-demographics 1/2

BOSTON REDEVELOPMENT AUTHORITY

Irish Share, namsor



Irish Share, 2010-2014 ACS

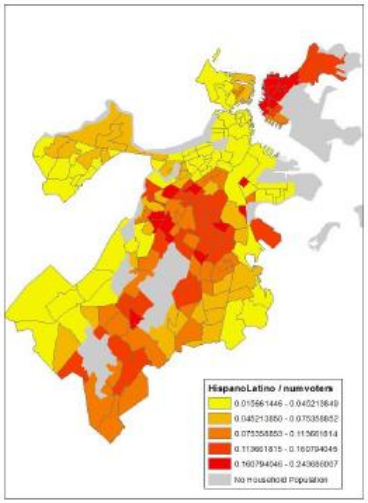


Source: Boston Voters List ; Visualization : ESRI ; Data Mining: NamSor

Boston geo-demographics 2/2

BOSTON REDEVELOPMENT AUTHORITY

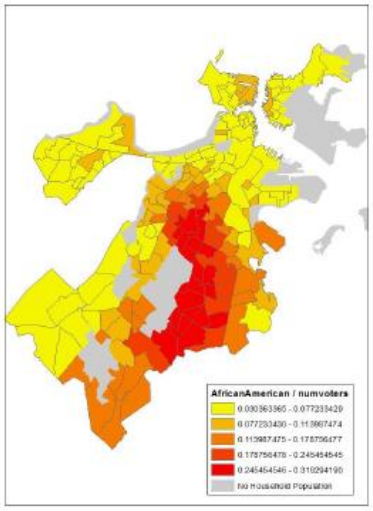
Hispanic/Latino Share, namsor



March 7, 2016
Presentation Title

BOSTON REDEVELOPMENT AUTHORITY

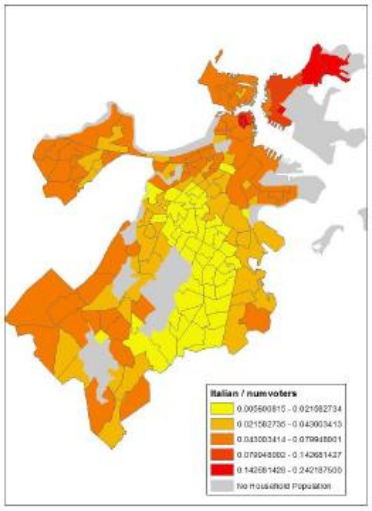
Black/African-American Share, nams



March 7, 2016
Presentation Title

BOSTON REDEVELOPMENT AUTHORITY

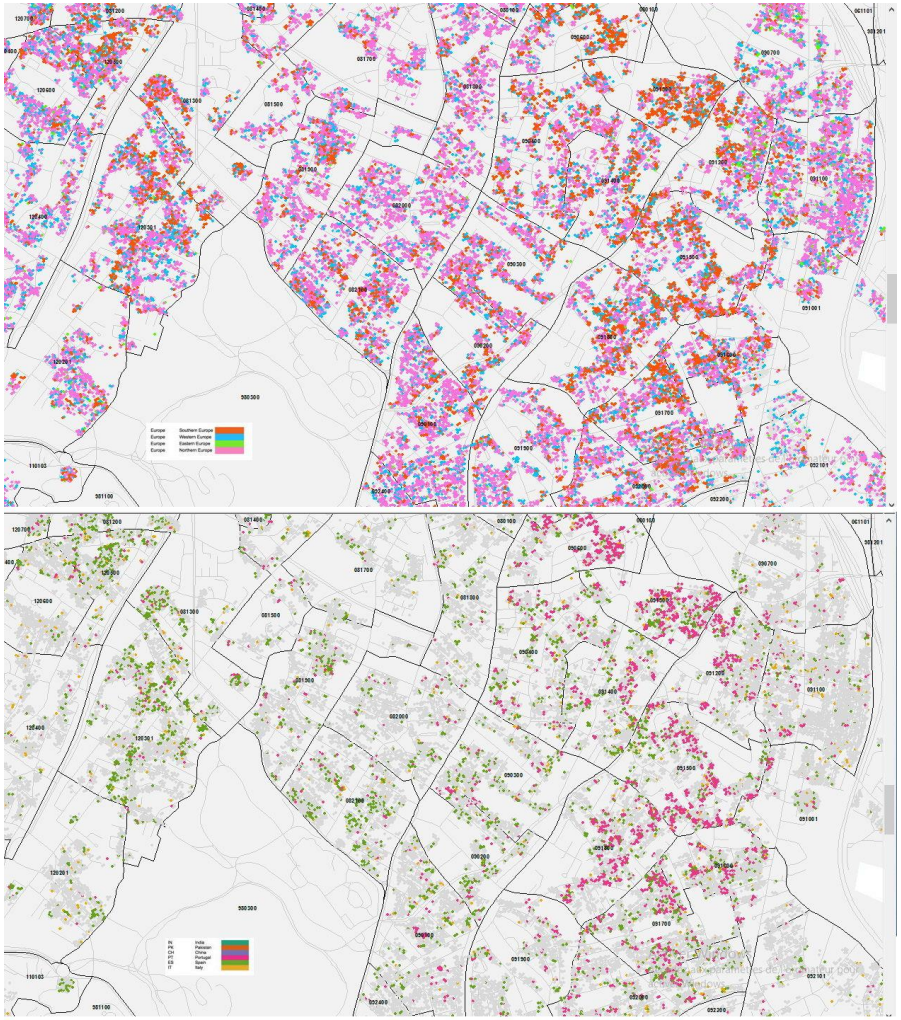
Italian Share, namsor

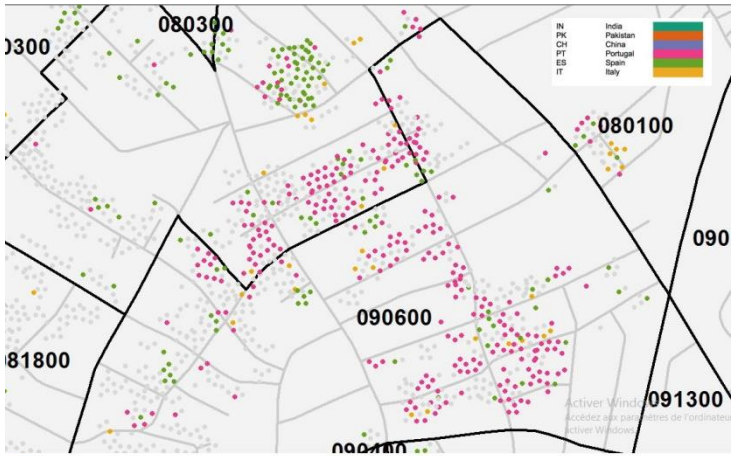


March 7, 2016
Presentation Title

Source: Boston Voters List
Visualization : ESRI
Data Mining: NamSor

Breaking down ‘White’ and ‘Asian’ into Portuguese, Spanish, Italian, India, Pakistan, China, …





Source: Boston Voters List
Visualization : ESRI
Data Mining: NamSor

PATENT DATABASES

US AID PROJECT

dateextract	linkedinurl	parentse	fullname	titlename	orgname	locality	languages	connections	descriptor	profilesummary	experience	organizations	volunteering	projects	education	awards	skills	recommandations	groups	publications	score	UA link	titlena	orgnan	localit	langua	connec	descrip	profile	experi	organi	volunti	project	educat	award	skills3	resum	groups	publica	isUkrCa	TOT
06/02/2016	https://www.linkedin.com/in/vyacheslav-dubchenko-	https://ww	Vyacheslav Dubchenko	Unemployed		Winnipeg, Manitoba, Canada	Russian,Ukrainian,Engli	68 connections	Winnipeg, Ma	Summary Summary	Experience	Manager of engineer	Volunteer Experience & Cau	Education Donskoy Stati	Skills Agriculture Agribusiness L	Groups Agricultural Job opportur	4.200801	9	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16.2008
06/02/2016	https://www.linkedin.com/in/viktoriia-osmak-419081	https://ww	Viktoriia Osmak	Certified Salesforce Developer	Freelancer	Bakersfield, California	English,Russian,Ukrainia	323 connection	Bakersfield, C	Summary Summary	Experience	Certified Salesforce Developer Freelancer July 2	Education Kyiv National Skills MVC	Recommendations A	Groups Ukrainian Physics and Ma	0.603375	10	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	1	0	x						15.60338
07/02/2016	https://ca.linkedin.com/in/yuriy-diakunchak-6a3bb61		Yuriy Diakunchak	Director of Marketing at Ukrainian Credit	Ukrainian Credit Unio	Toronto, Canada Area	English,Ukrainian,	500+ connectio	Toronto, Can	Summary Experien	Experience	Dir Organizations KU	Volunteer Experience & Cau	Education M.M. Robinso	Skills Proc	Recommendations A	Groups Markian Silecky Real Este	2.368793	7	1	1	0	1	0	0	0	1	1	0	0	0	0	0	1	0	0	x						15.36879
05/02/2016	https://www.linkedin.com/in/svitlanakozhushko	https://ww	Svitlana (Bilyk) Kozhushko	Director - Transfer Pricing and Internatio	PricewaterhouseCoo	New York, New York	Russian,Ukrainian,Engli	500+ connectio	New York, Ne	Summary I have 19	Experience	Director -Transfer P	Volunteer Experience & Cau	Education Kyiv State Eco	Skills Inte	Recommendations A	Groups ECJ Leading Cases Busine	3.272521	7	0	0	0	1	0	0	1	1	0	0	0	0	0	1	1	0	x						15.27252
05/02/2016	https://www.linkedin.com/in/sergiilytvynenko	https://ww	Sergii Lytvynenko	Business Development Manager at MGID Inc.		Santa Monica, California	English,Ukrainian,Russi	279 connection	Santa Monica	Summary - Providir	Experience	Re Organizations Tc	Volunteer Experience & Cau	Education Kyi	Honors &	Skills Creative Marketer Market	View Sergii's full profile to…	See	4.174871	7	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	x						15.17487
08/02/2016	https://be.linkedin.com/in/vbondarchuk	https://be.li	Vitalii Bondarchuk	Business Development Director at Rader	Rademade	Greater New York City Area	English,Russian,Ukraini	500+ connectio	Greater New	Summary 4 years of	Experience	Business Development Director Rademade Octo	Education Kyiv National	Skills Teac	Recommendations A	Groups KNEU alumni MJM and Ic	4.142169	9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	x						15.14217
05/02/2016	https://www.linkedin.com/in/yevhen-nesin-6141a13	https://ww	Yevhen Nesin	Research Lead		Other	English,Ukrainian,Russi	500+ connectio	Other		Experience	Mz Organizations Sc	Volunteer Experience & Cau	Education Kyiv National	Skills Financial Modeling Emerg	Groups Fryday Business Events a	1.098148	9	0	0	0	1	0	0	0	1	1	0	0	1	0	0	0	1	0	x						15.09815
11/02/2016	https://www.linkedin.com/in/yuriygnatyuk		Yuriy Gnatyuk	CBO at KindGeek	KindGeek	San Francisco Bay Area	English,Ukrainian,Russi	500+ connectio	San Francisco	Summary What abo	Experience	CBO KindGeek May	Volunteer Expe	Projects ST	Education National Univ	Skills Start-ups Product Manage	Groups Outsourc	Published by Yu	3.963097	8	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	x						14.9634
06/02/2016	https://www.linkedin.com/in/oleksandra-mykolaichu	https://ww	Oleksandra Mykolaichuk	Researcher	Forte Group Inc.	Chicago, Illinois	Ukrainian,Russian,Engli	99 connections	Chicago, Illin	Summary I'm active	Experience	Researcher Forte Group Inc. April 2015 – Present	Education Ternopil Volo	Skills Recruiting Candidates Sea	Groups JavaScript Discussion / Jo	2.876263	8	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	x						14.87626
06/02/2016	https://www.linkedin.com/in/sskakun	https://ww	Sergii Skakun	Assistant Research Professor at Universit	University of Maryla	College Park, Maryland	Ukrainian,Russian,Engli	250 connection	College Park,	Summary	Experience	Assistant Research Professor Univer	Projects St	Education Sp	Honors &	Skills Computer Science Resear	Groups Horizon2020 - H2020 Wor	1.869683	8	0	0	0	1	0	0	0	1	1	1	0	0	0	0	0	1	0	x						14.86868
08/02/2016	https://be.linkedin.com/in/ksoroka	https://be.li	Kseniia Soroka	Business Development Manager at Rade	Rademade	San Francisco Bay Area	English,Ukrainian,Russi	495 connection	San Francisco	Summary An ambiti	Experience	Bu Organizations BE	Volunteer Experience & Cau	Education National Tech	Skills Teamwork English Microso	Groups Luxury Travel & Tourism	1.612316	9	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	x						14.61232
06/02/2016	https://www.linkedin.com/in/vfedoroff	https://ww	Vadym Fedorov	Solutions Architect at SoftServe	SoftServe	Austin, Texas Area	English,	500+ connectio	Austin, Texas	Summary A multi-i	Experience	Solutions Architect	Volunteer Expe	Projects Pe	Education Sc	Honors &	Skills OOP	Recommendations A	Groups H1B Jobs	Published by Vi	3.466105	8	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	x						14.4661
11/02/2016	https://www.linkedin.com/in/saikevych		Ievgenii Saikevych	Senior Software Engineer at EPAM Syster	EPAM Systems	San Francisco Bay Area	English,Russian,Ukraini	500+ connectio	San Francisco	Summary Specialis	Experience	Senior Software Engineer EPAM Systems April 2	Education LGU , Econom	Skills Wel	Recommendations A	Groups Semantic Web Open Sou	3.386101	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	x						14.3861
06/02/2016	https://www.linkedin.com/in/artem-zinchenko-14966	https://ww	Artem Zinchenko	Student at Taras Shevchenko National Ur	Google	San Francisco Bay Area	Russian,Ukrainian,Engli	118 connection	San Francisco Bay Area		Experience	Software Engineer Google October 2015 – Prese	Education Kyiv National	Skills Algorithms Data Structure	Groups Deep Learning Kyiv Natio	4.246047	9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	x						14.24605

seedname	https://ca.linkedin.com/in/yuriy-diakunchak-6a3bb6[...]_-1035308383.linkedin	**Explanations**
datetime	2016-02-07T12:49+0000	
dateextract	07/02/2016	
linkedinurl	https://ca.linkedin.com/in/yuriy-diakunchak-6a3bb61	**LinkedIn URL and Profile data**
parentseed		-
fullname	Yuriy Diakunchak	-
titlename	Director of Marketing at Ukrainian Credit Union Limited	-
orgname	Ukrainian Credit Union Limited	-
locality	Toronto, Canada Area	-
languages	English,Ukrainian,	-
connections	500+ connections	-
descriptor	Toronto, Canada Area	-
profilesummary	Summary Experienced marketing manager. Focused on branding and brand management, executing programs and campaigns targeted at building br	-
experience	Experience Director of Marketing Ukrainian Credit Union Limited August 2010 – Present (5 years 7 months) Heading up the marketing department. R	-
organizations	Organizations KUMF Art Gallery Director Starting March 2013 Member of Board of Directors of KUMF Art Gallery (Ukrainian Canadian Art Foundation)	-
volunteering	Volunteer Experience & Causes Causes Yuriy cares about: Arts and Culture Civil Rights and Social Action Politics	-
projects		-
education	Education M.M. Robinson McMaster University BComm, Marketing, Finance Pragmatic Marketing , Product Marketing Course Ryerson University Bacl	-
awards		-
skills	Skills Product Management Marketing Automation Brand Management Product Marketing Demand Generation Multi-channel Marketing Marketing S	-
recommandations	Recommendations A preview of what LinkedIn members have to say about Yuriy: I had the pleasure of working with Yuriy since 2006 on various vide	-
groups	Groups Markian Silecky Real Estate - Babiak Team Royal LePage Real Estate Services Ltd, Brokerage Cloud Computing Taxonomy & Information Mana	-
publications		-
score	2.368793324	**< how much Ukrainian the nams sounds**
count	7	**< how many Ukrainian connections (max :10)**
titlename2	0	
orgname3	0	
locality2	0	
languages3	1	**< Ukrainian appears in Languages**
connections4	0	
descriptor5	0	
profilesummary6	0	
experience7	1	
organizations8	0	
volunteering9	0	
projects10	1	
education11	0	
awards12	0	
skills13	0	
recommandations14	0	
groups15	1	**< Ukraine appears in Goups**
publications16	0	
TOT	13.98341639	**< Overall score taking into account factors from L2**

Indian diaspora names - a global airline use case

'For 93% of our customers, when NamSor recognizes an Indian name, the client has travelled to India in the past.'

At state level : ~50%

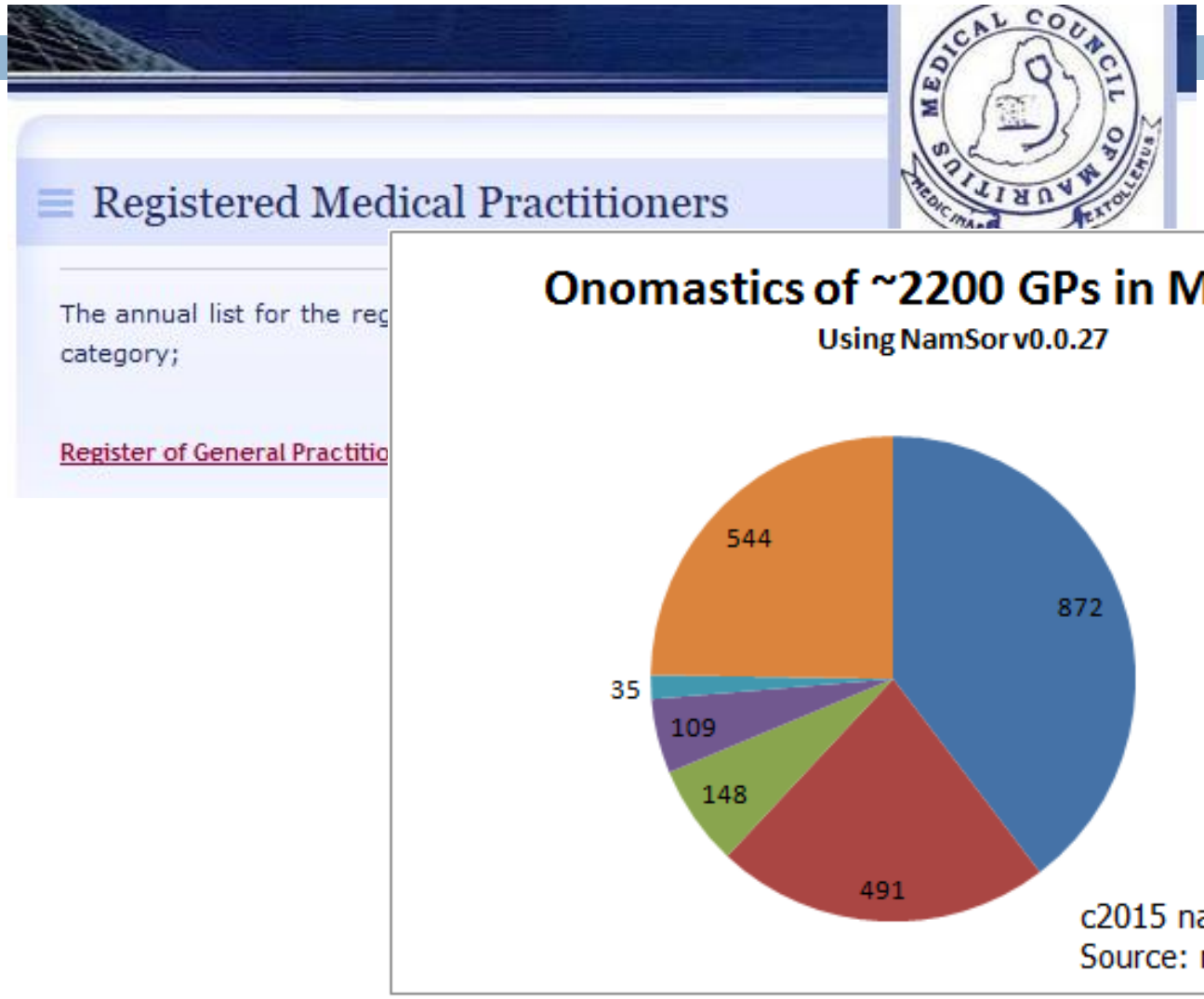
Analysis of NamSor's **First** Choice Country Compared to Historic Travel on [redacted]

NamSor's **First** Choice Country	Count of Unique Individuals who Have Travelled to NamSor's **First** Choice Country			% of Unique Individuals who Have Travelled to NamSor's **First** Choice Country		
	No	Yes	Grand Total	No	Yes	Grand Total
India	1,633	20,315	21,948	7%	93%	100%
Italy	281	869	1,150	24%	76%	100%
Bangladesh	524	1,456	1,980	26%	74%	100%
Ethiopia	3	8	11	27%	73%	100%
Iran	701	1,657	2,358	30%	70%	100%
Saudi Arabia	679	771	1,450	47%	53%	100%
Afghanistan	21	23	44	48%	52%	100%
Pakistan	2,171	2,309	4,480	48%	52%	100%
Jordan	148	124	272	54%	46%	100%
Kuwait	51	37	88	58%	42%	100%
Qatar	3	2	5	60%	40%	100%

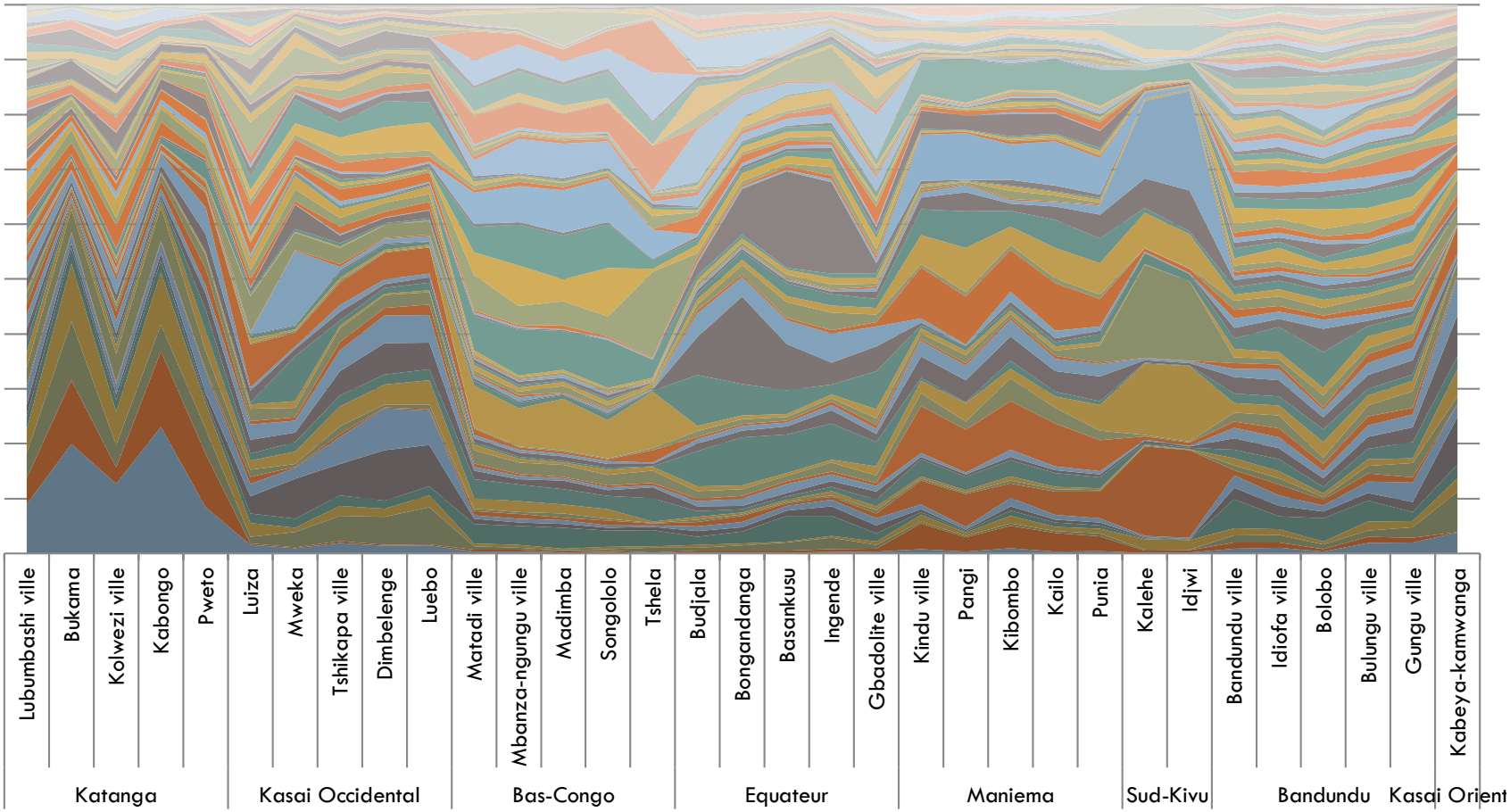
Analysis of NamSor's **Region Rounded** Score Compared to Historic Travel on [redacted] for India

Customer has Flown to NamSor's Region	Count of Unique Individuals who Have Travelled to NamSor's **Region**			% of Unique Individuals who Have Travelled to NamSor's **Region**		
	No	Yes	Grand Total	No	Yes	Grand Total
5		3	3	0%	100%	100%
4	569	696	1,265	45%	55%	100%
3	2,202	2,774	4,976	44%	56%	100%
2	2,861	3,226	6,087	47%	53%	100%
1	2,442	2,523	4,965	49%	51%	100%
0	1,686	1,423	3,109	54%	46%	100%
-1	702	519	1,221	57%	43%	100%
-2	180	109	289	62%	38%	100%
-3	23	9	32	72%	28%	100%
-4	1		1	100%	0%	100%
Grand Total	10,666	11,282	21,948	49%	51%	100%

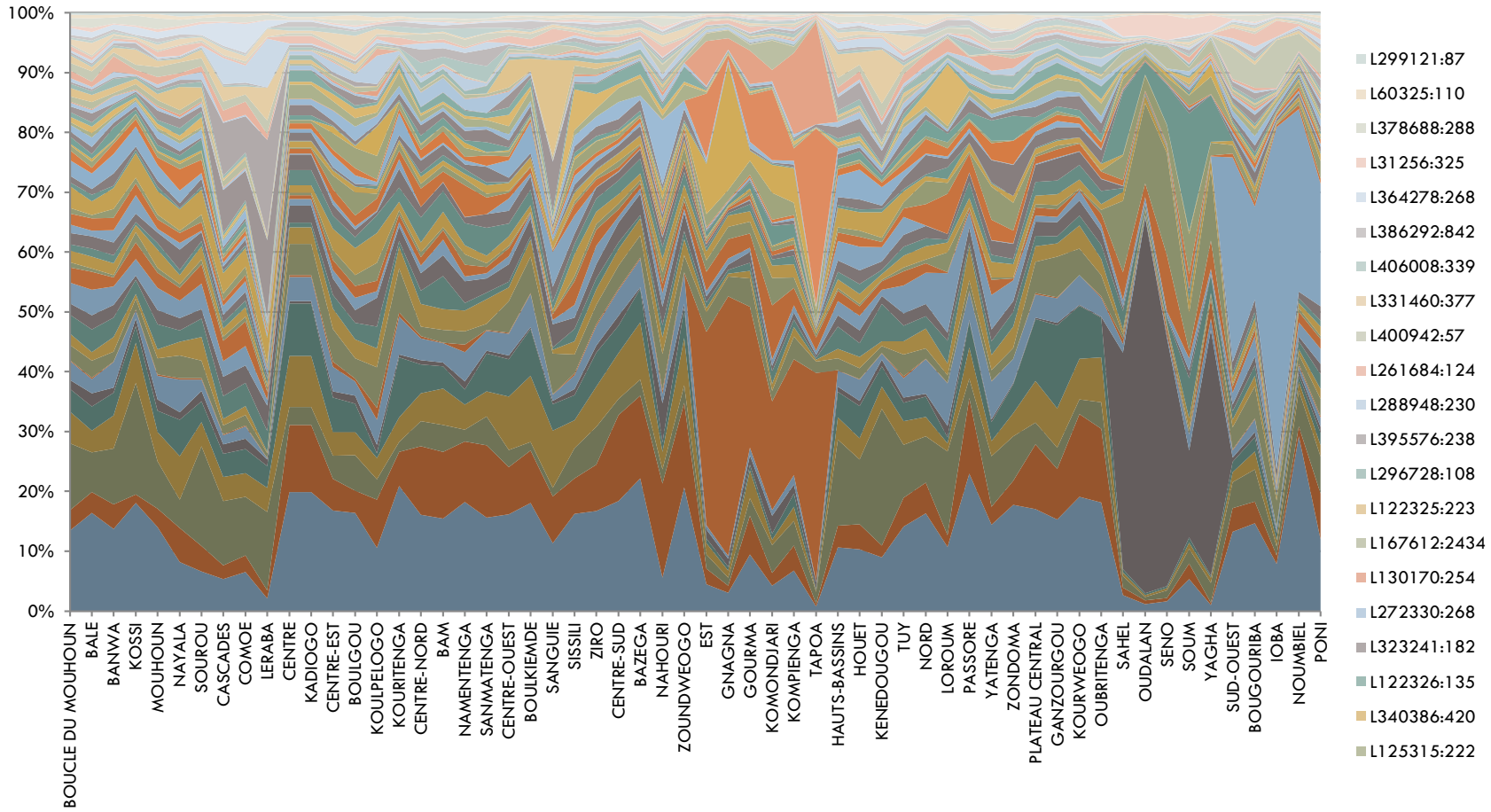
Indian diaspora in Mauritius



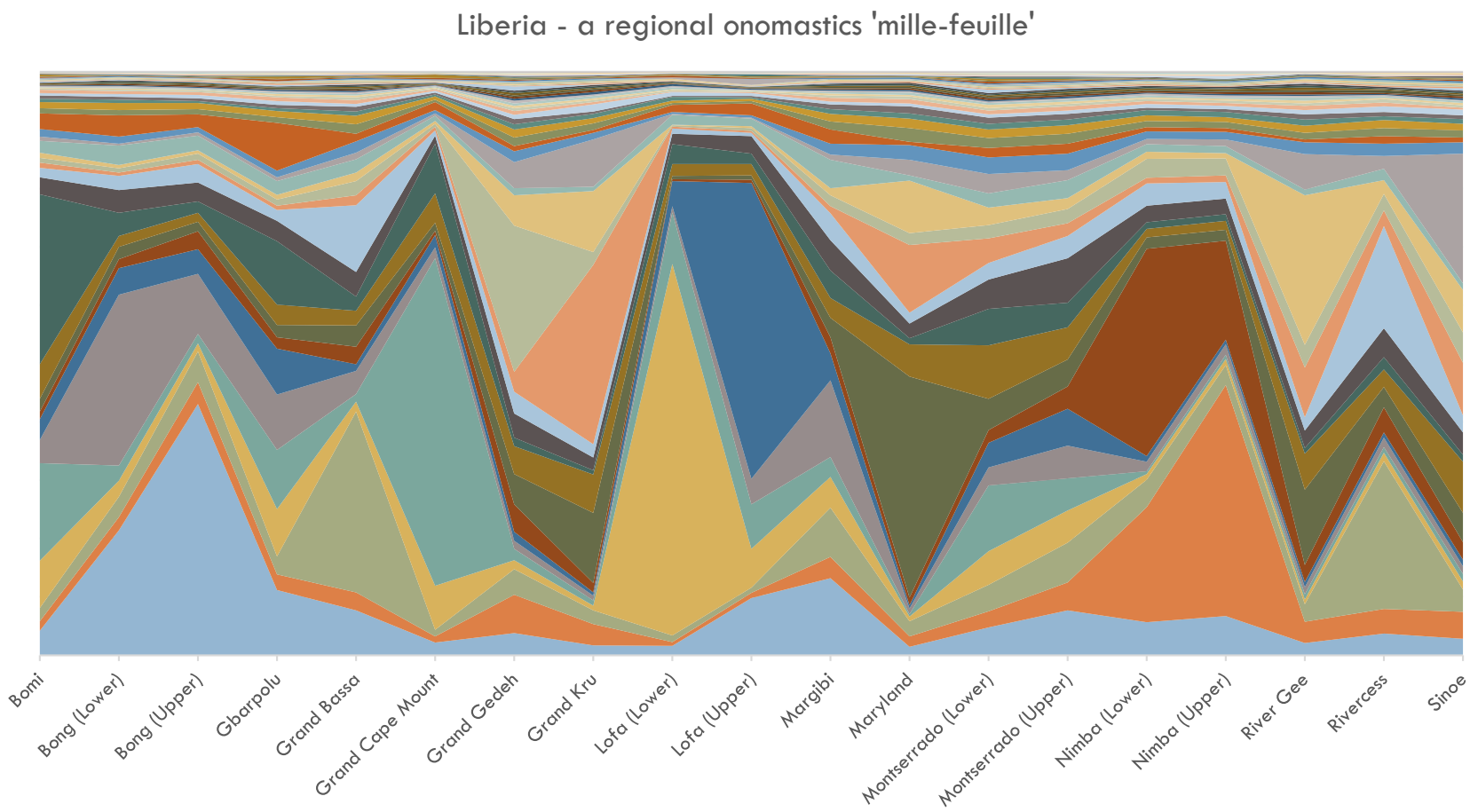
Africa: complex identities (Congo RDC)



Africa: complex identities (Burkina Faso)



Africa: complex identities (Liberia)



Thank you !

Elian CARSENAT,

elian.carsenat@namsor.com

Phone : +33 6 52 77 99 07