BILL & MELINDA GATES foundation

HEALTHY BIRTH, GROWTH & DEVELOPMENT

knowledge integration

December 5, 2017

Country Segmentation Tool

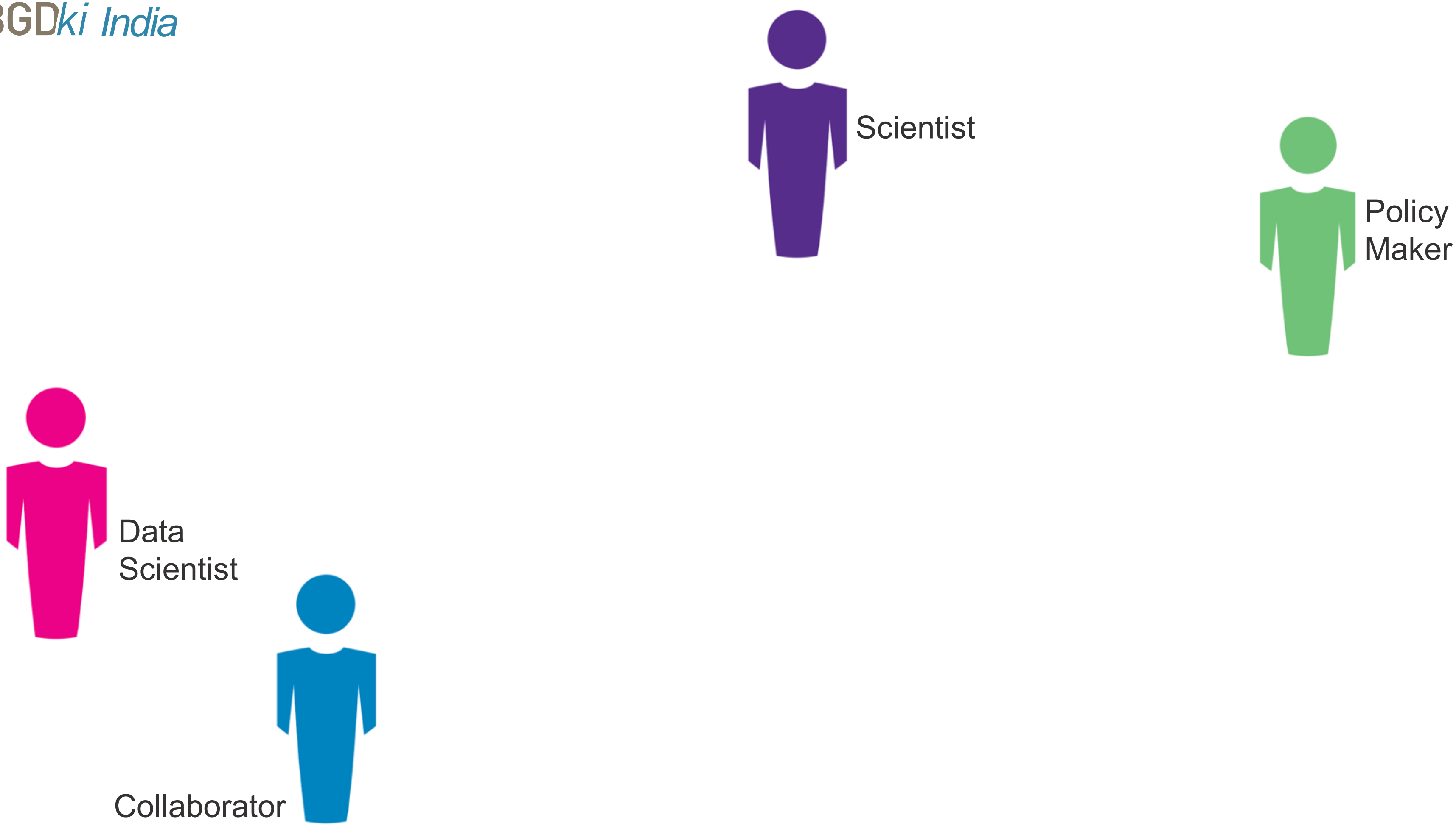
Dave King

सत्यमेव जयते
Department of Biotechnology Govt. of India

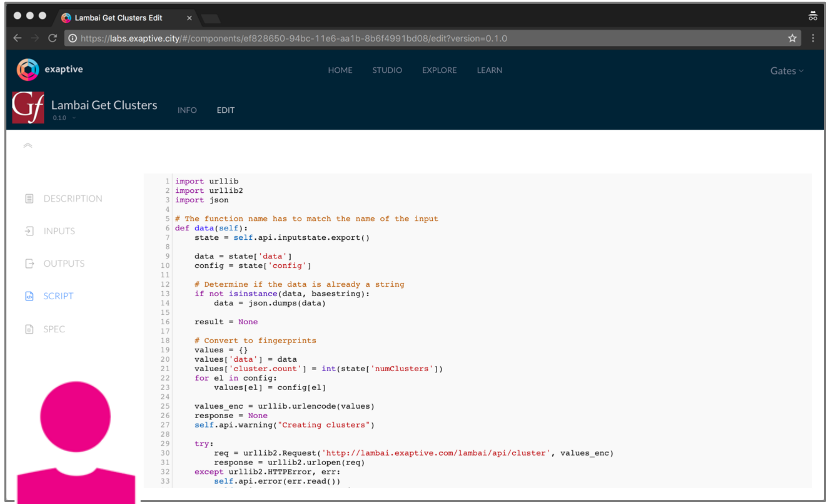
birac
Ignite Innovate Incubate

Grand Challenges India

exaptive

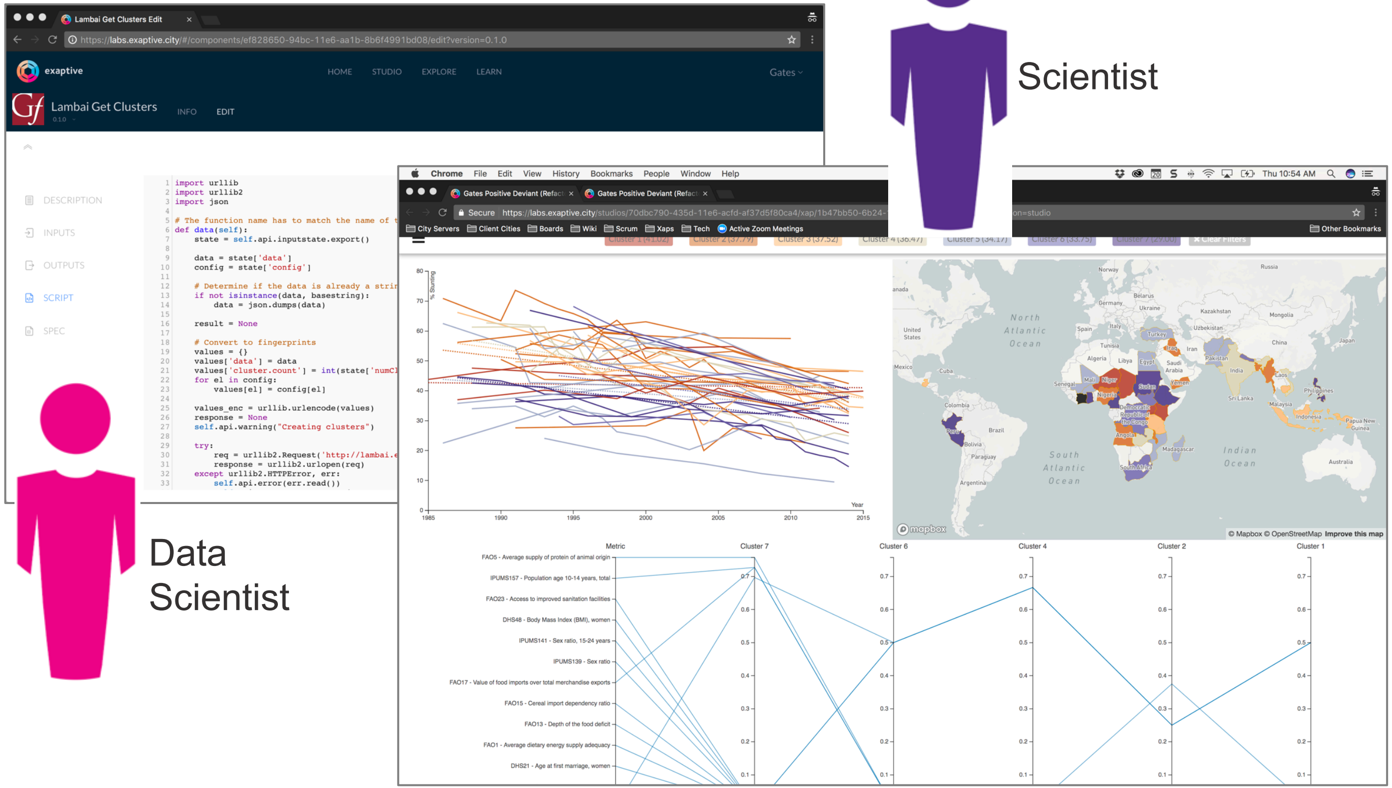

Scientist
Policy Maker
Data Scientist
Collaborator

HBGDki India



Data
Scientist

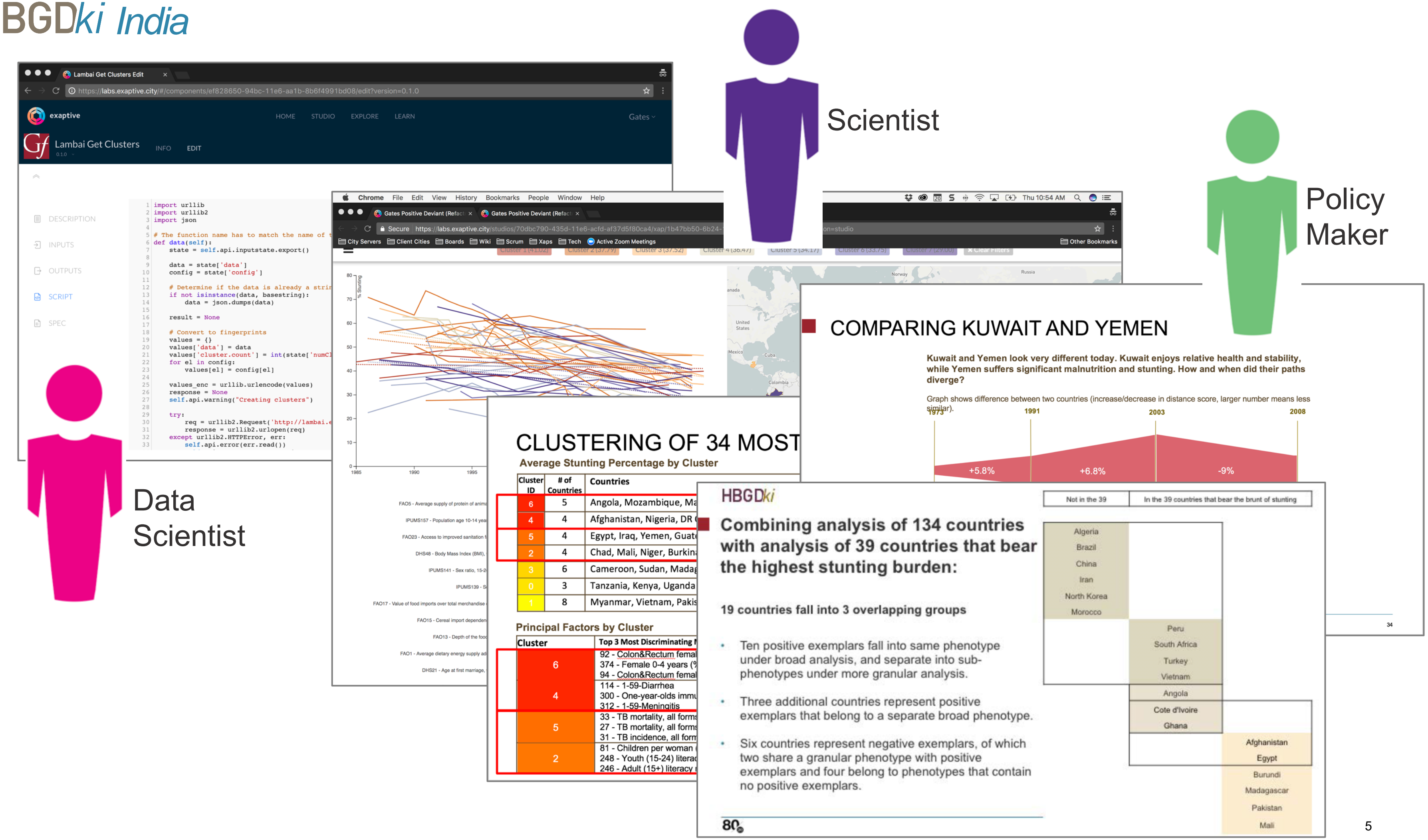
HBGDki India



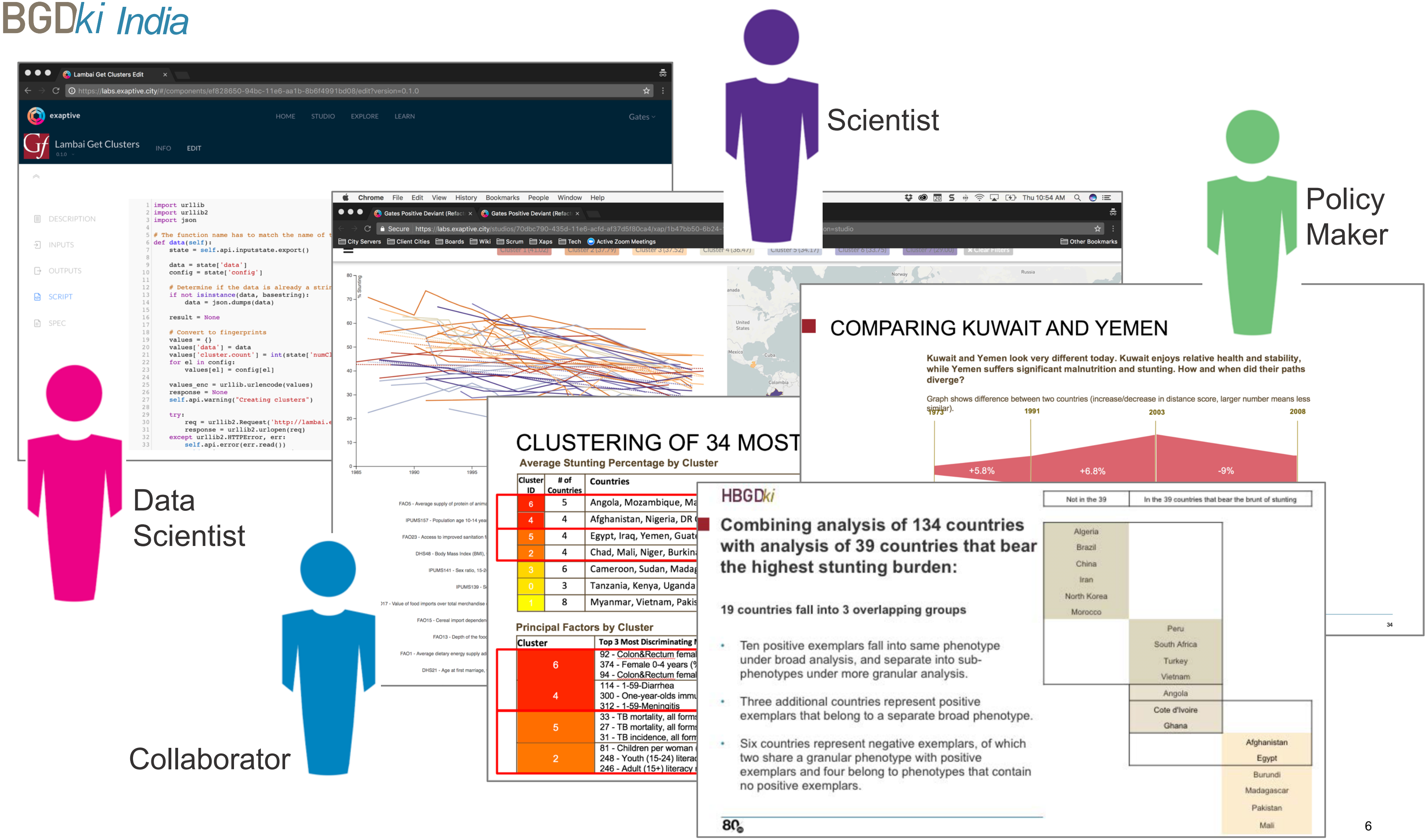
Scientist

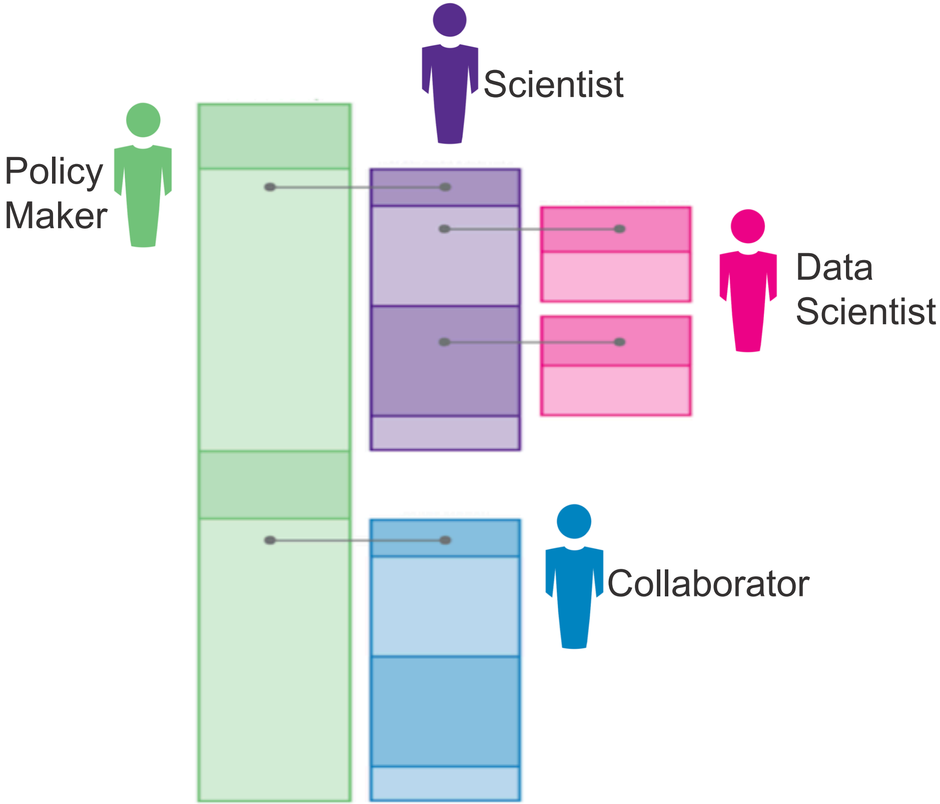
Data Scientist

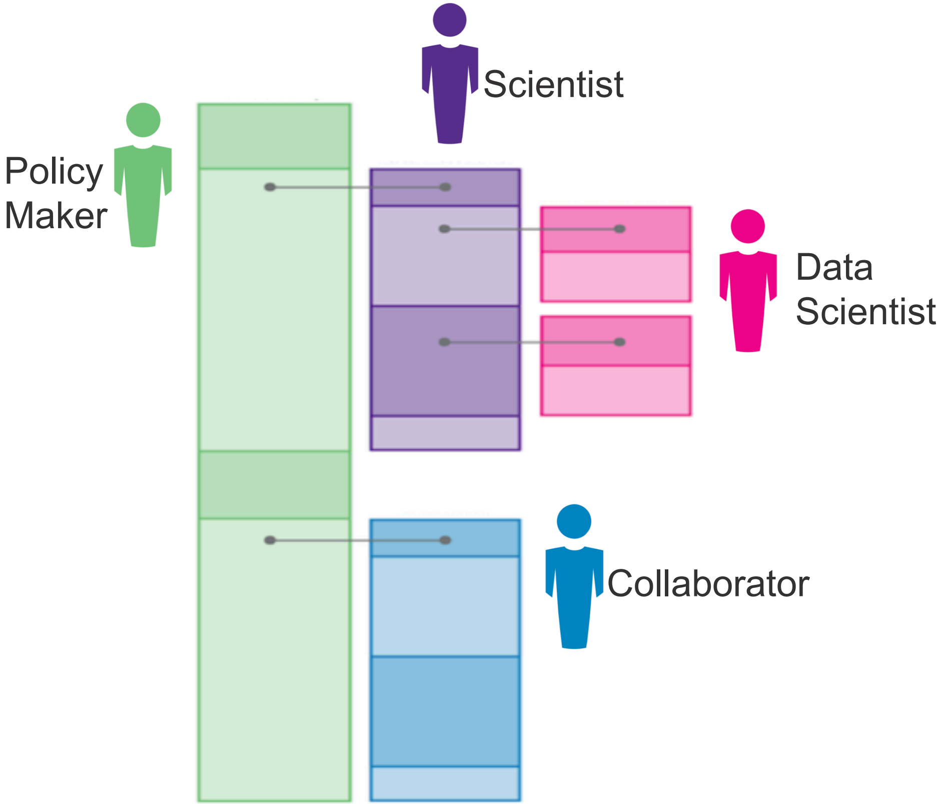
HBGDki India



HBGD*ki* India




Scientist
Policy
Maker
Data
Scientist
Collaborator



https://labs.exaptive.city/x/gates/gcgdatastory

HBGDki Country Genotyping

> "Happy families are all alike; every unhappy family is unhappy in its own way" -- Leo Tolstoy, from Anna Karenina

This famous quote has come to be used for more than just describing families, it has come to be referred to as "The Anna Karenina Principle" - the recognition that just because certain communities exhibit similar outward signs of distress doesn't mean that the inner root causes of that dysfunction are the same. In the case of global stunted growth there are many unhappy countries. 90% of the world's stunting burden is shouldered by 38 countries. But just because these countries are similar in one problem area, doesn't mean that they should all be grouped together. We know that stunting interventions have a wide range of effectiveness and that interventions that may have been effective in one pilot study may not be as effective when attempted again in a different place. What we don't know, however, is in what to attribute this variance in efficacy. Is the intervention just not consistently effective, or is that the environments in which it has been introduced are not consistent? The questions that the Country Genotyping project seeks to answer are:

By better understanding the similarities and differences between countries, can we make better decisions about how to spend our limited research resources? Can we better target intervention recommendations to the communities where they will have the maximum impact?

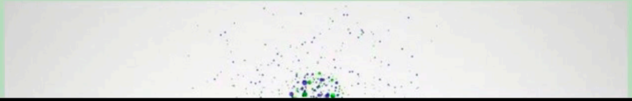
While the Country Genotyping methodology can be applied in a wide variety of situations, we've focused on using it to help researchers decide the best set of "positive exemplar" countries to perform in-depth studies on.

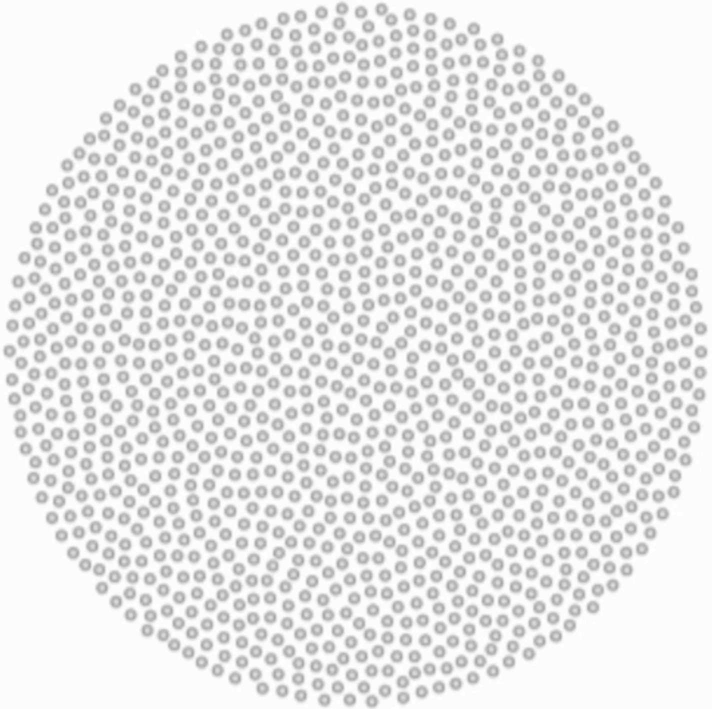
The SickKids Global Center for Child Health, in partnership with the Gates Foundation and bgC3, is in the process of planning a series of fieldwork studies in countries that have improved stunting more than their peers. Unfortunately, there are more countries worthy of study than resources available, so a subset of countries must be selected. How to make this selection? There are logistical criteria that certainly factor in - it is hard to study a country which does not allow researchers access, and it is easier to study a country where researchers may already have connections to local organizations and government. However, we believe that the best criteria for selecting a country for study is to what degree it is believed that anything learned from that study will have applicability for other countries working to improve healthy birth, growth, and development.

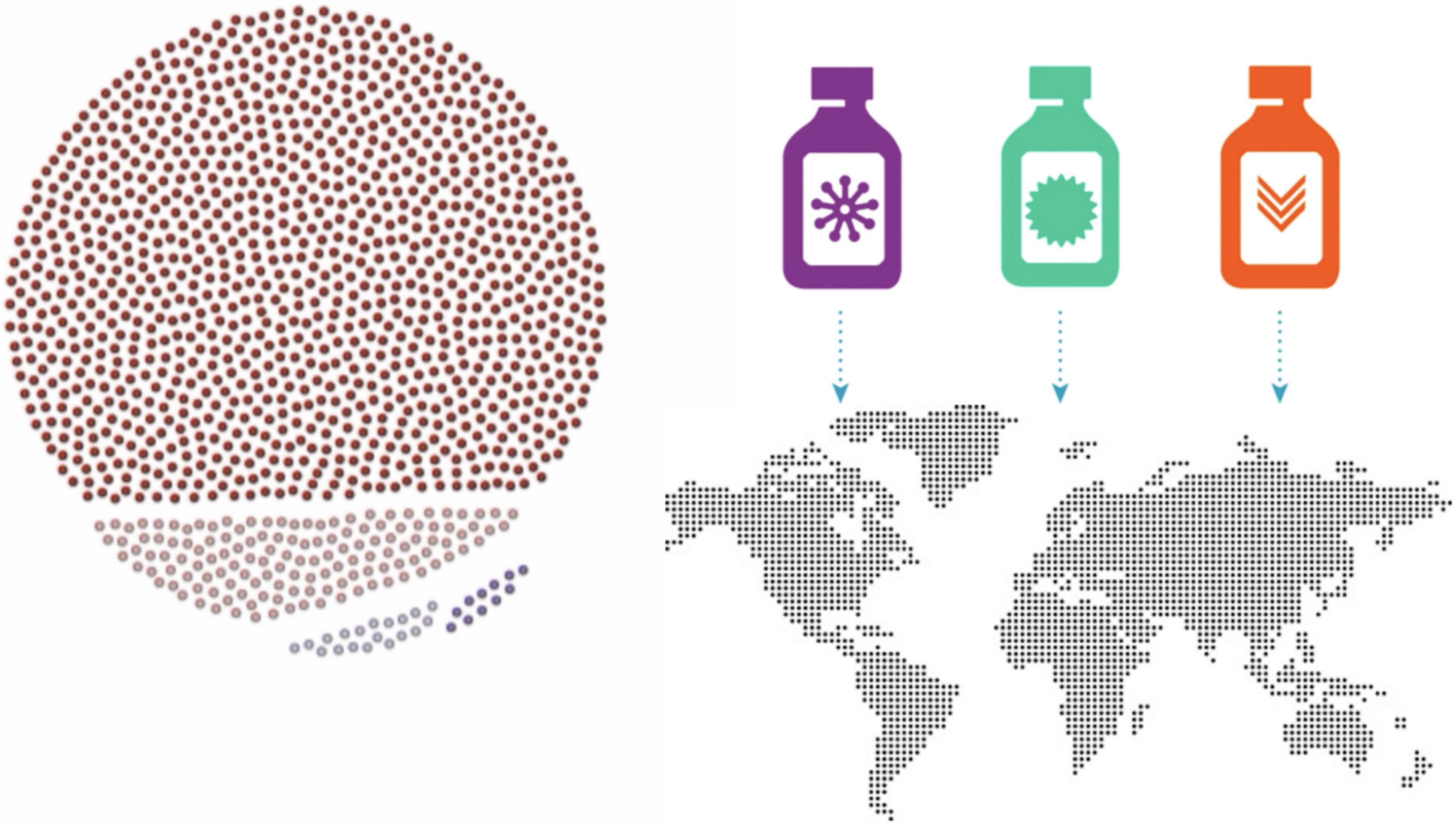
The Country Genotyping Project, applied to the question of which "positive exemplar" countries to study further, aims to identify, through an impartial data-driven method, the best subset of positive exemplars suited to applying their lessons learned to other countries.

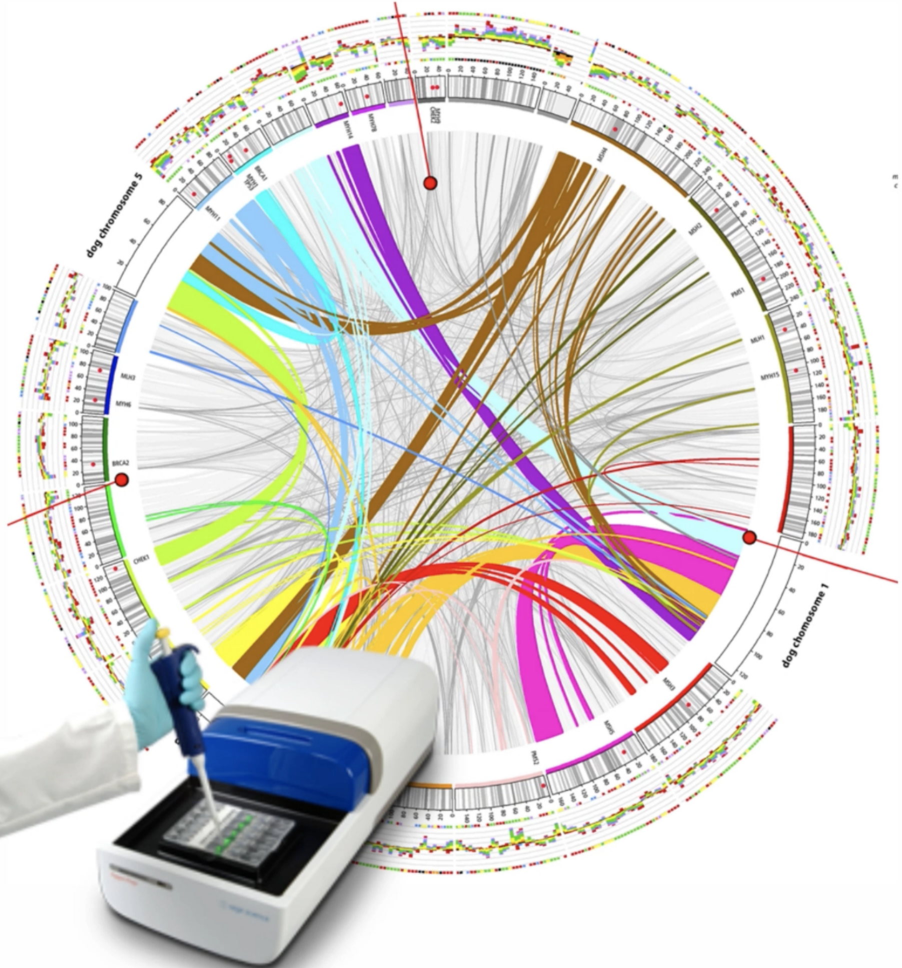
General Approach

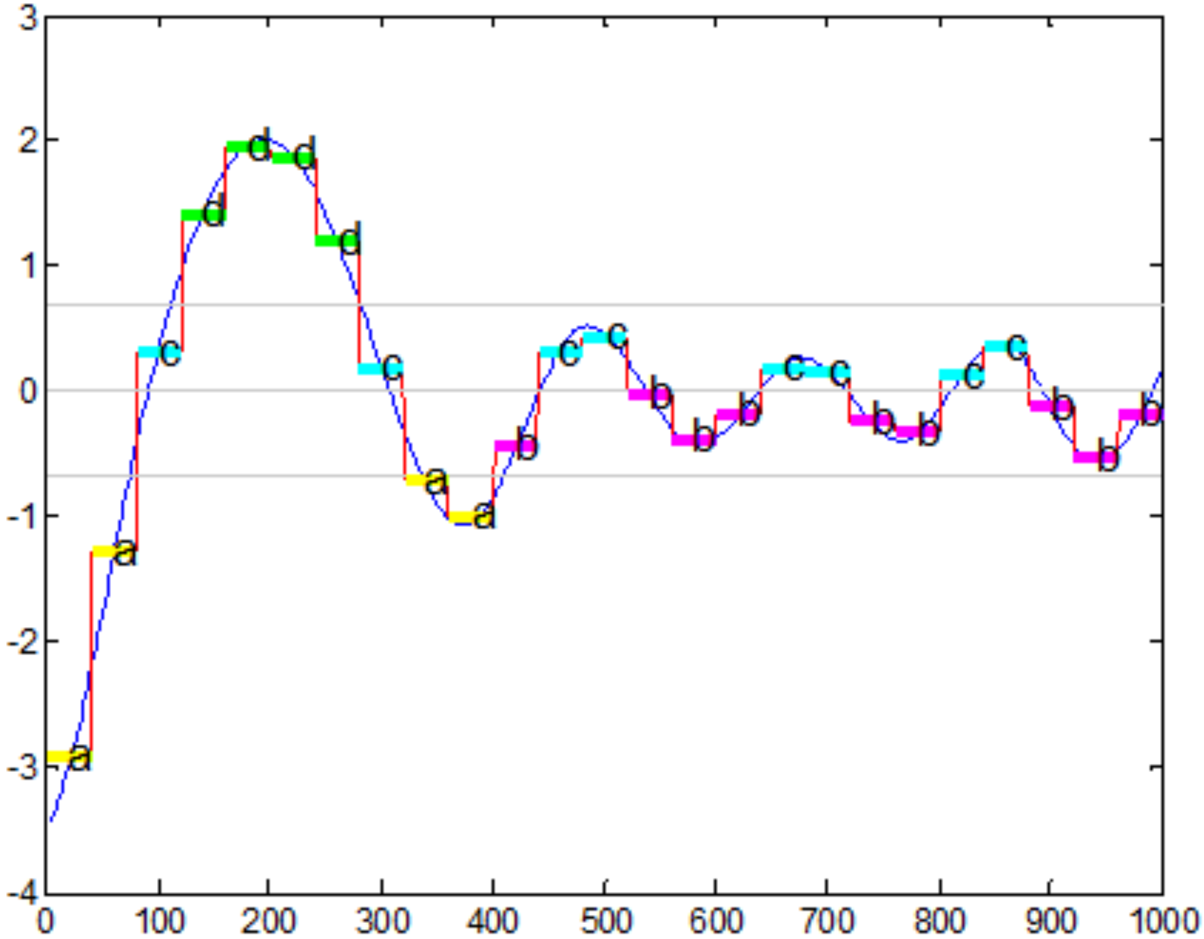
To tackle this problem, we've borrowed an approach from precision medicine that we believe can yield dividends when applied to countries instead of just people. The whole field of precision medicine is based on the Anna Karenina principle - the idea that people exhibit similar symptoms for different reasons, and that the key to improving patient outcomes is to match the right treatment



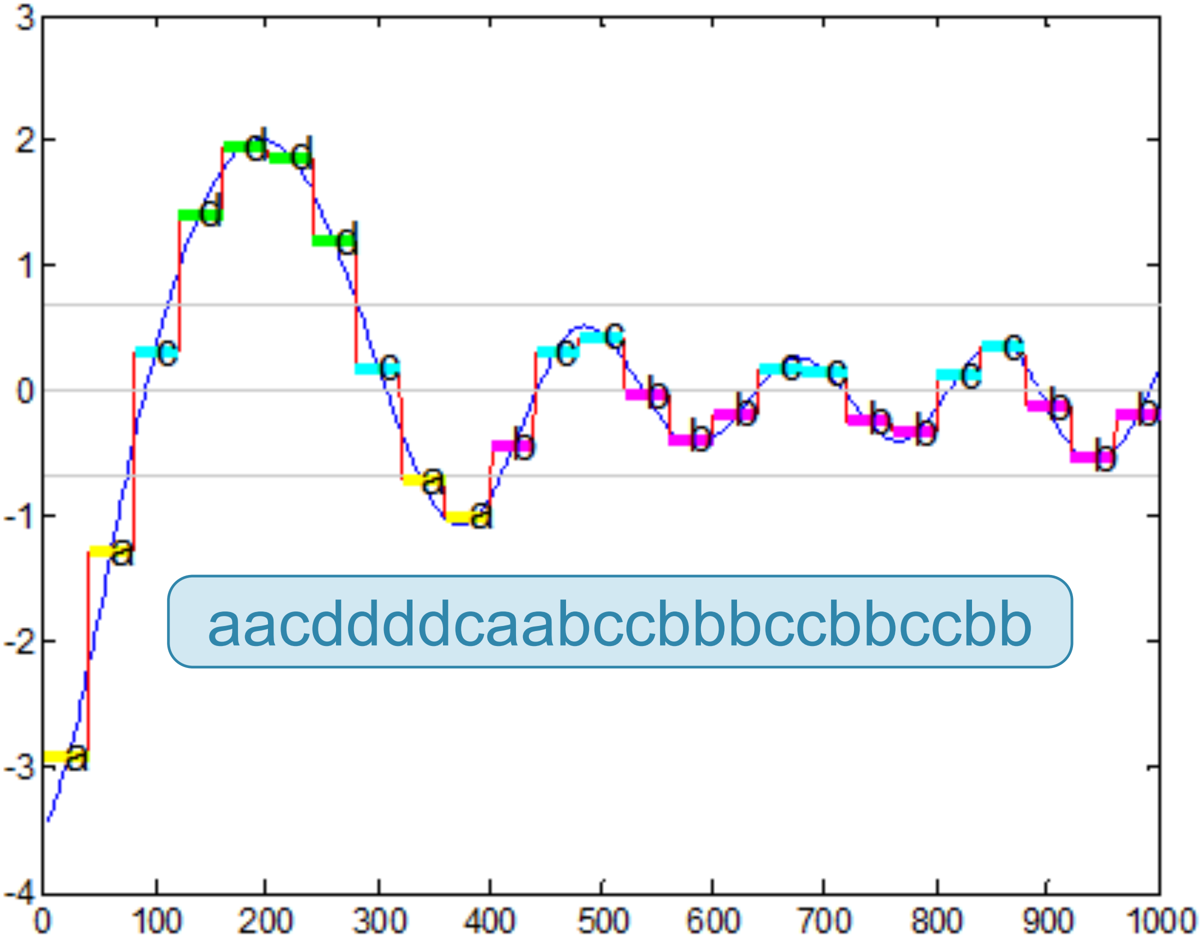


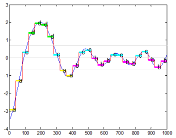



dog chromosome 5
dog chromosome 1
BRCA2

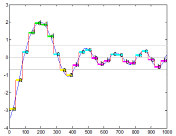


3
2
1
0
-1
-2
-3
-4
0
100
200
300
400
500
600
700
800
900
1000
a
a
c
d
d
d
d
c
a
a
b
c
c
b
b
b
c
c
b
b
c
c
b
b
b

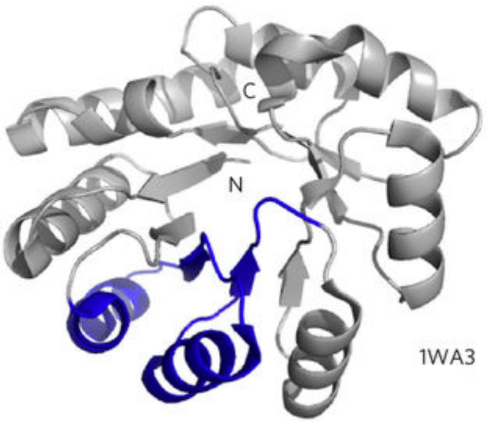

3
2
1
0
-1
-2
-3
-4
0
100
200
300
400
500
600
700
800
900
1000
aacddddcaabccbbbccbbccbb

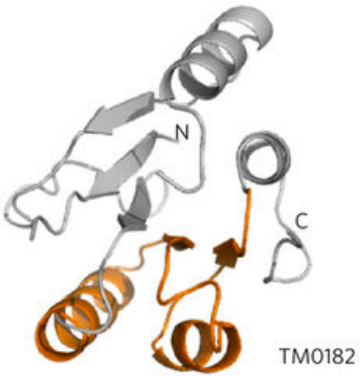


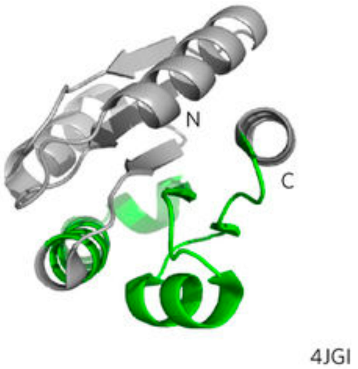
aacddddcaabccbbbccbbccbb

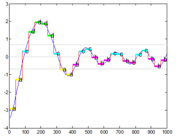


aacddddcaabccbbbccbbccbb





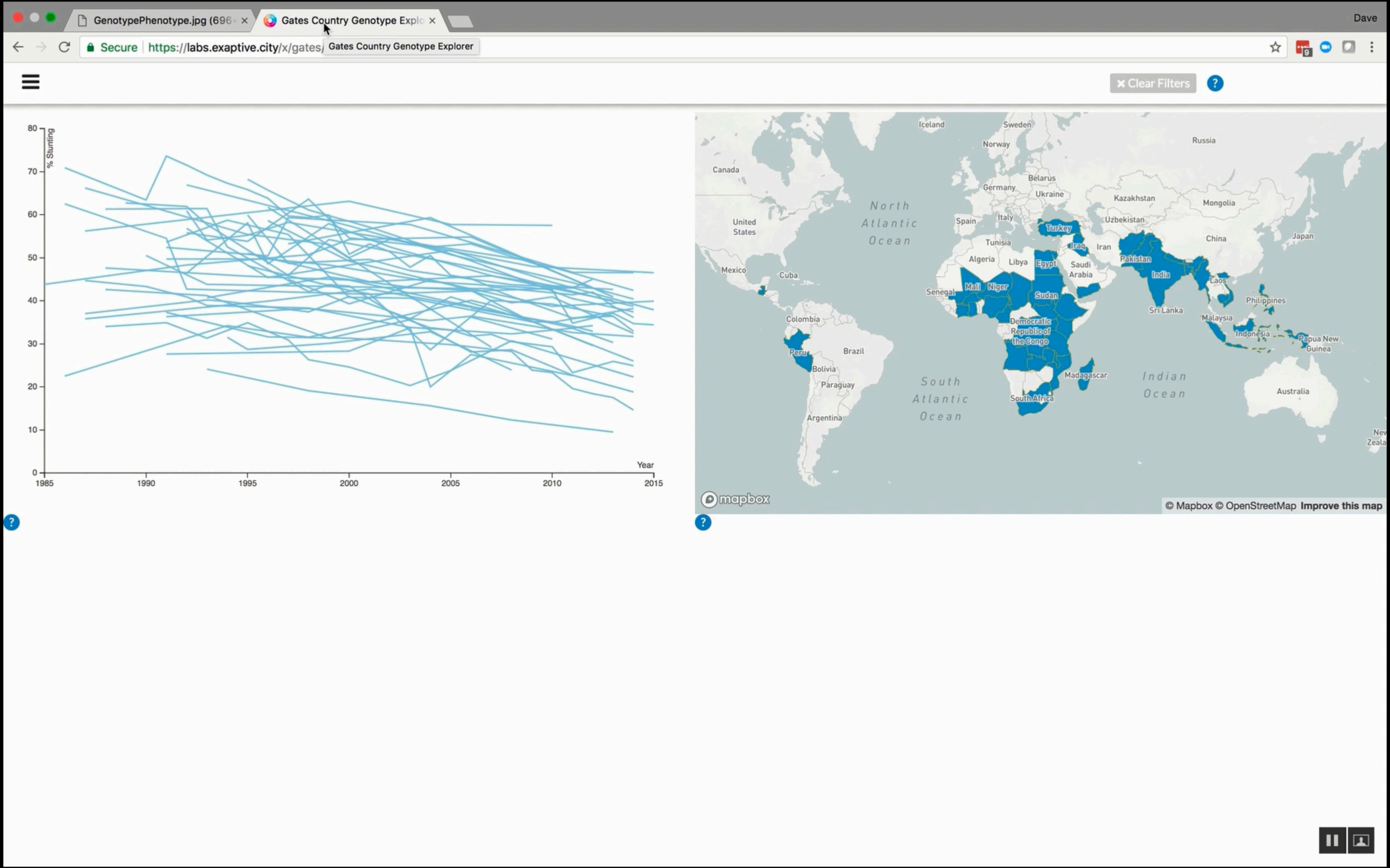


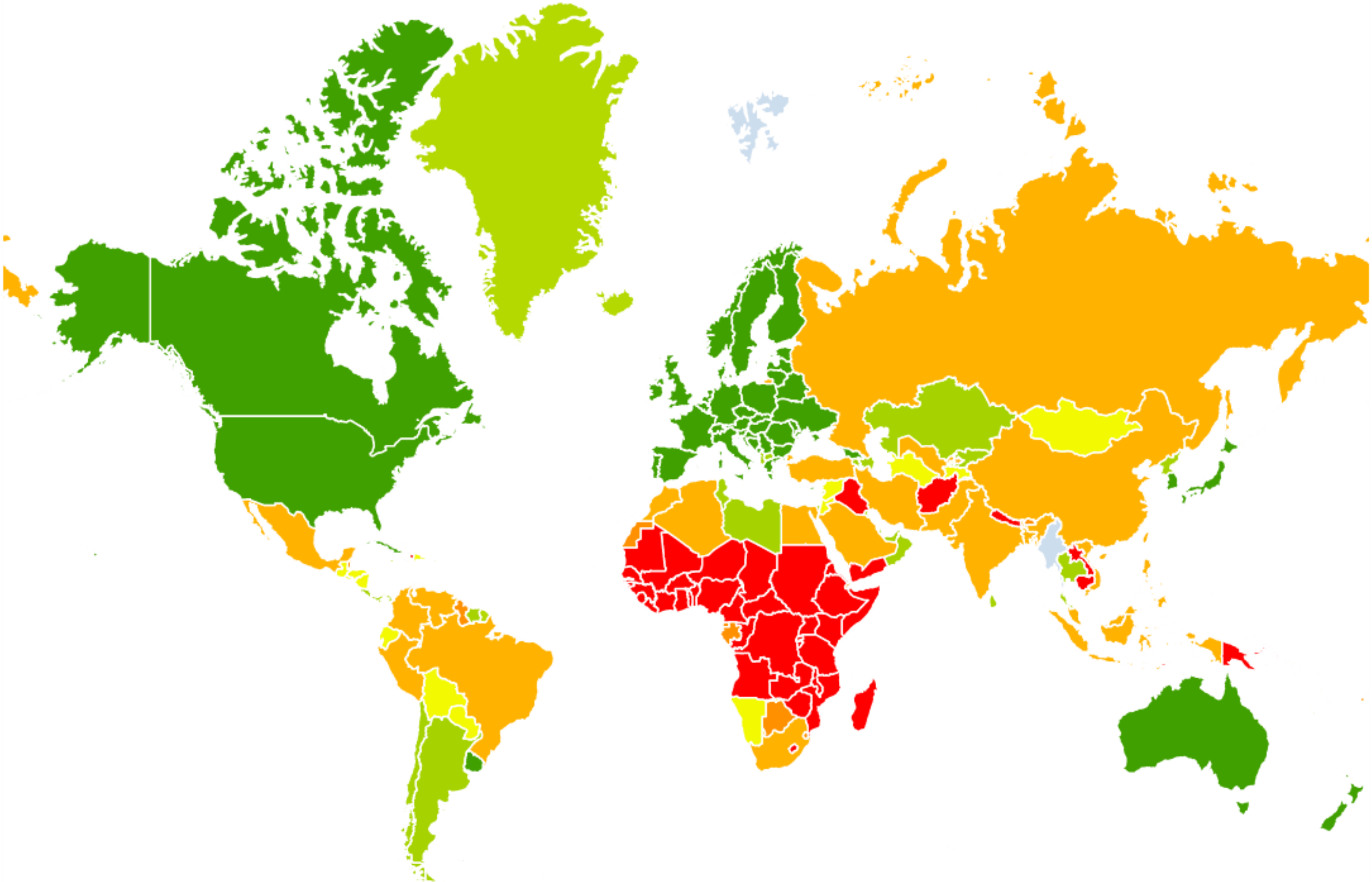


aacddddcaabccbbbccbbccbb

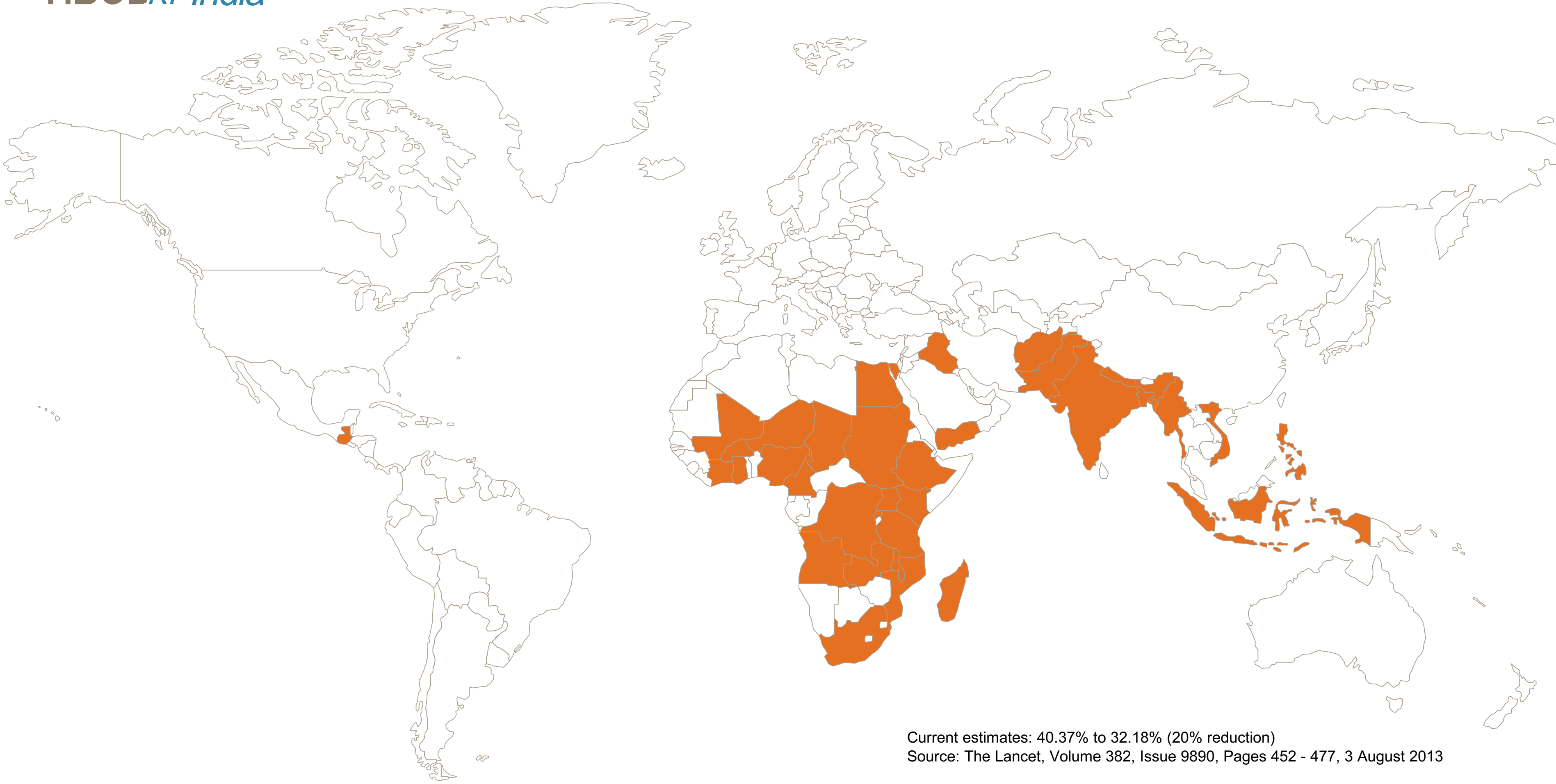

```
1WA3      11 KIVAVLR---ANsveeak------------------------ekalaVF----EGGv---HLIEITFTvP- 46
              ++                                                         ++  +|
TM0182     1 MYILFRE---MK-NNWY--SLAALLSTiysrhldVEARPV----KFEEI----KKFPpeKTIVAYSFMSF- 56
               |+                   +  +        |+  +        |    |       ||  | +
4JGI      90 AKIVLATvegDLhDIGKniFRTMAEASg------FEVFDLgidvPVKIIvdkvKEVN--PEIVGLSGVLTl 152


1WA3      47 DADTVIKELSFLKEK-GA---IIGAGT--VTsveqCr-kAVE-SGAEFIVSpHldeeisqfckekgVFYMPG 110
             |+|||  |+  |||+ |     + ||   ||         +   | + + +   +| |  |          |
TM0182    57 DLDTVREEVKTLKER-GY---TLIAGGphVTa---DpegCLR-MGFDHVFTgDGeENILKFLMGErKKIFDG 120
              ||++|| |  ||         +|+||  |             +| |   |       +|
4JGI     153 ALDSMRETVDALKAEgLRndlKVIIGGvpVNe---N---VCQrVGADDFST-NA-ADGVKICQRW-vg---- 211
```


GenotypePhenotype.jpg (696×
Gates Country Genotype Explo
Dave
Secure
https://labs.exaptive.city/x/gates/
Gates Country Genotype Explorer
Clear Filters
% Stunting
80
70
60
50
40
30
20
10
0
Year
1985
1990
1995
2000
2005
2010
2015
Iceland
Sweden
Norway
Russia
Canada
Belarus
Germany
Ukraine
Kazakhstan
Mongolia
North Atlantic Ocean
United States
Spain
Italy
Uzbekistan
Turkey
China
Japan
Tunisia
Iraq
Iran
Algeria
Libya
Egypt
Saudi Arabia
Pakistan
Mexico
Cuba
India
Laos
Senegal
Mali
Niger
Sudan
Philippines
Colombia
Sri Lanka
Democratic Republic of the Congo
Malaysia
Indonesia
Papua New Guinea
Peru
Brazil
Bolivia
Paraguay
South Atlantic Ocean
Madagascar
Indian Ocean
Australia
South Africa
Argentina
mapbox
© Mapbox © OpenStreetMap Improve this map

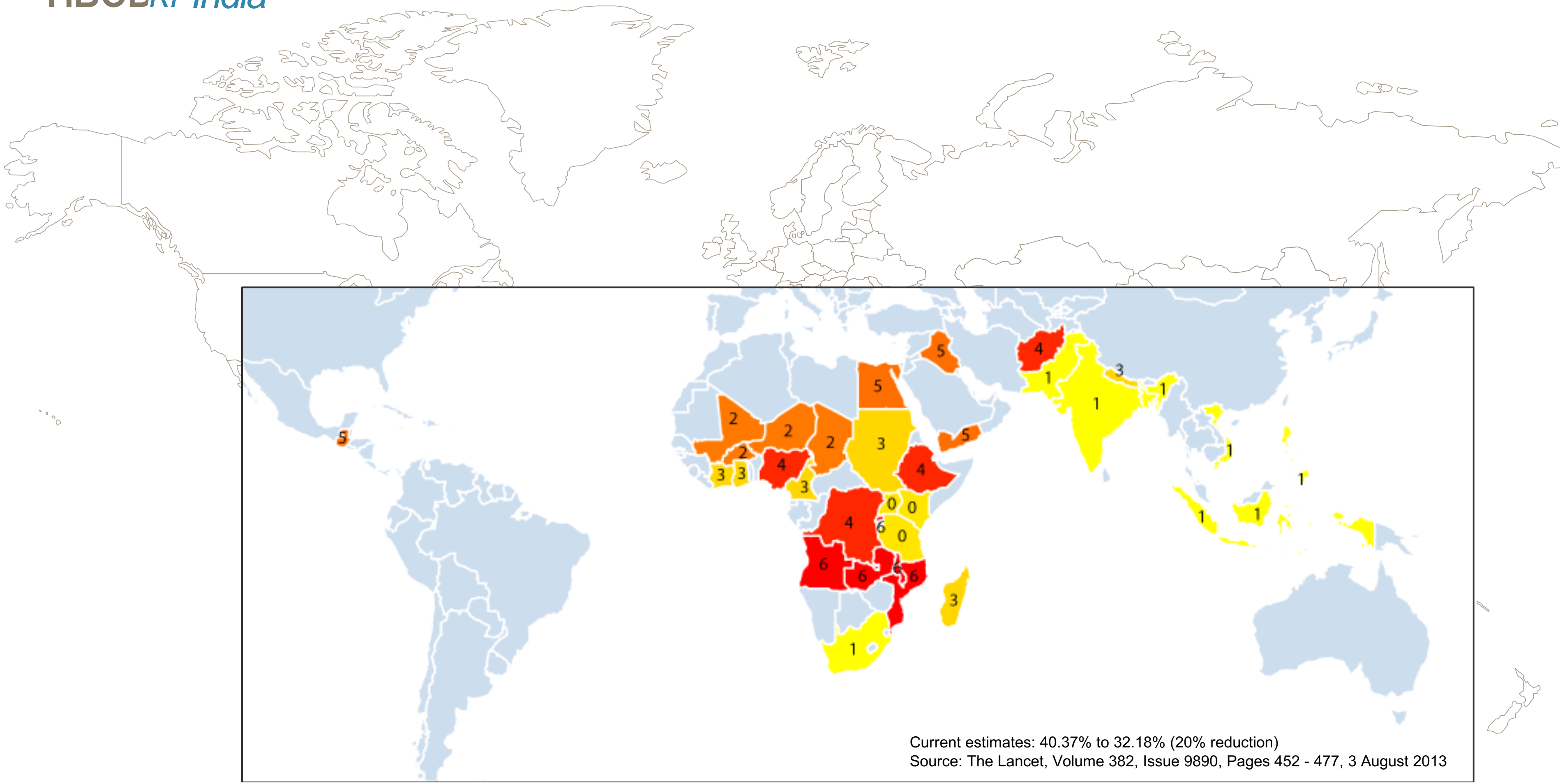


HBGDki India



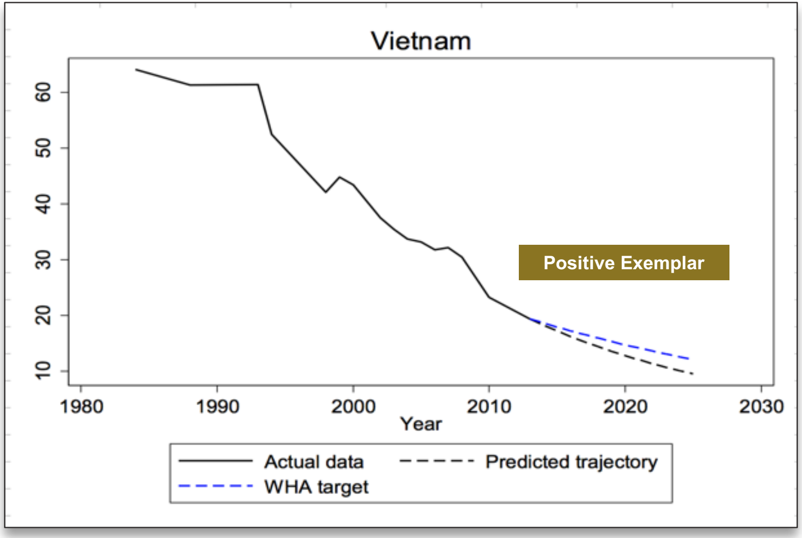
Current estimates: 40.37% to 32.18% (20% reduction)
Source: The Lancet, Volume 382, Issue 9890, Pages 452 - 477, 3 August 2013

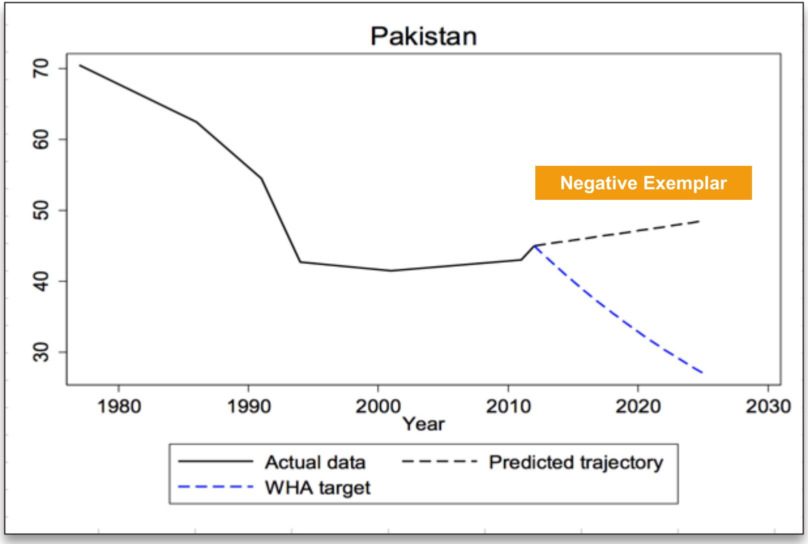
HBGDki India



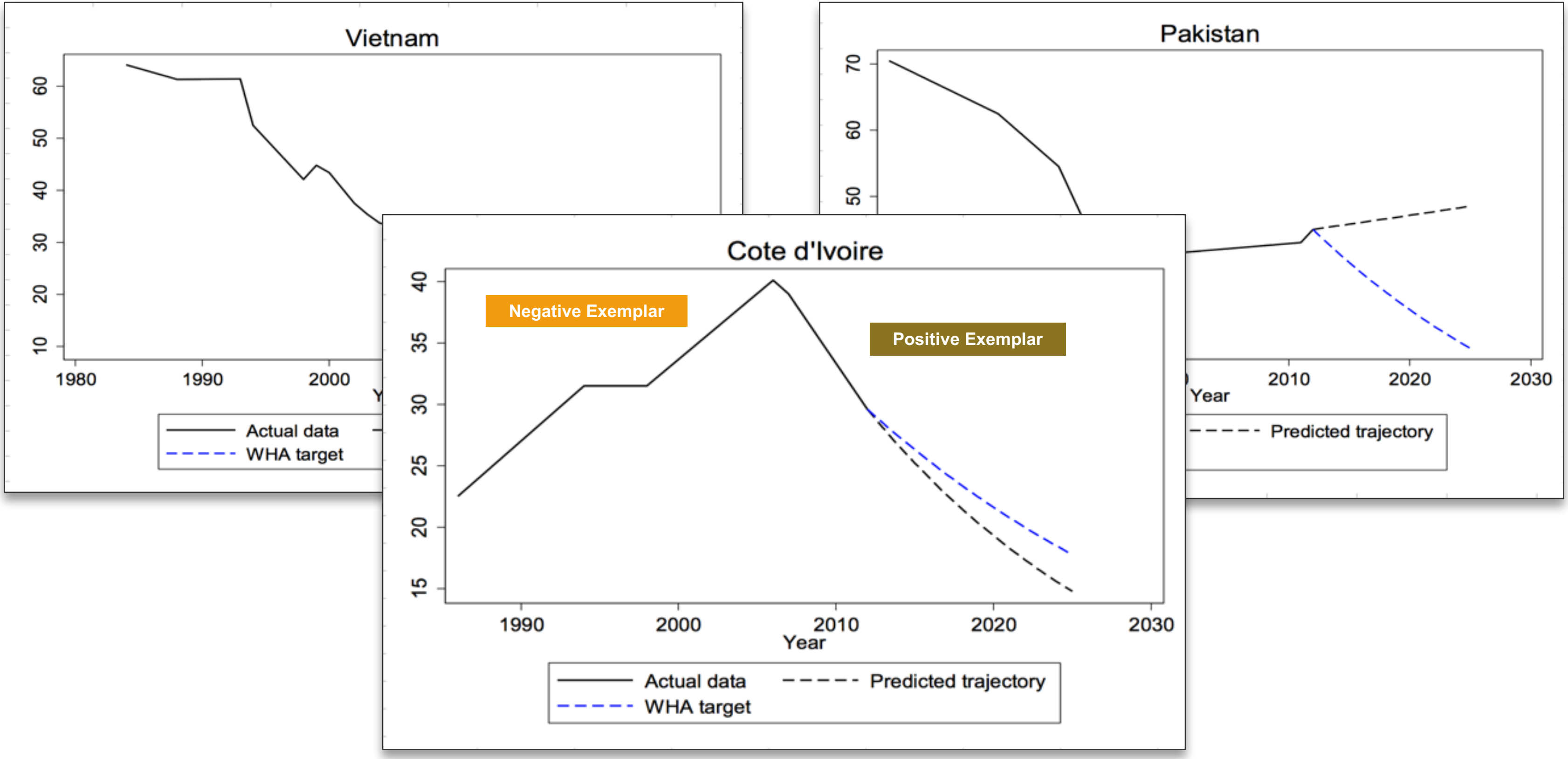
Current estimates: 40.37% to 32.18% (20% reduction)
Source: The Lancet, Volume 382, Issue 9890, Pages 452 - 477, 3 August 2013

Positive/Negative exemplar analysis:

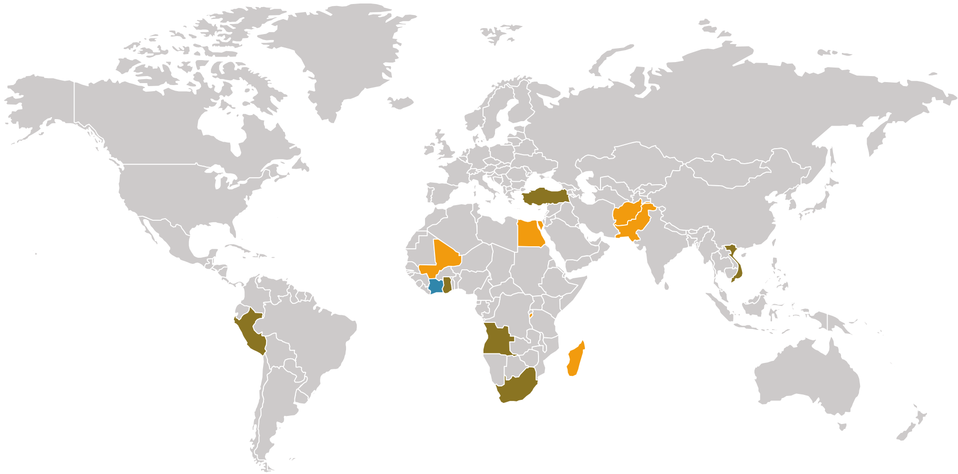




Positive/Negative exemplar analysis:



Exemplars among the 39 countries that bear the greatest burden of stunting:



Positive Exemplars

Peru | Ghana | Angola | South Africa | Turkey | Vietnam

Negative Exemplars

Mali | Egypt, Arab Rep. | Burundi | Madagascar | Afghanistan | Pakistan

Mixed

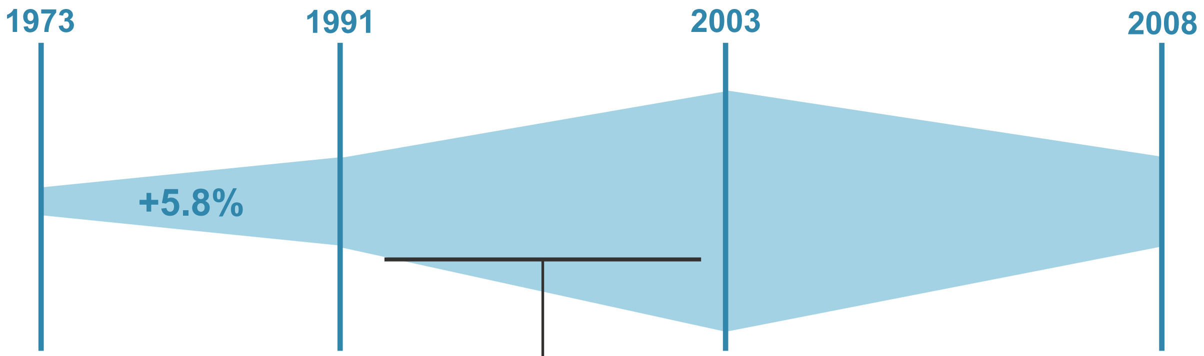
Côte d'Ivoire

		(2) SAX Fingerprint Segmentation on Time-Series Trajectories of Non-Stunting Metrics						
		Segmentation cluster 1 FAO1 - Average dietary energy supply adequacy FAO12 - Share of food expenditure of the poor FAO16 - Percent of arable land equipped for irrigation FAO10 - Domestic food price index	**Segmentation cluster 2** FAO12 - Share of food expenditure of the poor FAO16 - Percent of arable land equipped for irrigation FAO19 - Domestic food price volatility	**Segmentation cluster 3** FAO3 - Share of dietary energy supply derived from cereals, roots and tubers FAO12 - Share of food expenditure of the poor FAO10 - Domestic food price index	**Segmentation cluster 4** IYCF2 - Ever Breastfed %: Female IYCF52 - Continued Breastfeeding 1 Year: Female IYCF53 - Continued Breastfeeding 1 Year: Male	**Segmentation cluster 5** FAO5 - Average supply of protein of animal origin FAO12 - Share of food expenditure of the poor FAO19 - Domestic food price volatility	**Segmentation cluster 6** FAO1 - Average dietary energy supply adequacy FAO4 - Average protein supply FAO3 - Share of dietary energy supply derived from cereals, roots and tubers	**Segmentation cluster 7** FAO18 - Political stability and absence of violence/terrorism FAO16 - Percent of arable land equipped for irrigation FAO10 - Domestic food price index
(1) Linear Regression on Log(Stunting)	+							
	-							
	+/-							

		(2) SAX Fingerprint Segmentation on Time-Series Trajectories of Non-Stunting Metrics						
		Segmentation clusters with Positive Exemplars				Segmentation clusters with Negative Exemplars		Mixed
		Segmentation cluster 7 FAO18 - Political stability and absence of violence/terrorism FAO16 - Percent of arable land equipped for irrigation FAO10 - Domestic food price index	**Segmentation cluster 6** FAO1 - Average dietary energy supply adequacy FAO4 - Average protein supply FAO3 - Share of dietary energy supply derived from cereals, roots and tubers	**Segmentation cluster 4** IYCF2 - Ever Breastfed %: Female IYCF52 - Continued Breastfeeding 1 Year: Female IYCF53 - Continued Breastfeeding 1 Year: Male	**Segmentation cluster 2** FAO12 - Share of food expenditure of the poor FAO16 - Percent of arable land equipped for irrigation FAO19 - Domestic food price volatility	**Segmentation cluster 1** FAO1 - Average dietary energy supply adequacy FAO12 - Share of food expenditure of the poor FAO16 - Percent of arable land equipped for irrigation FAO10- Domestic food price index	**Segmentation cluster 3** FAO3 - Share of dietary energy supply derived from cereals, roots and tubers FAO12 - Share of food expenditure of the poor FAO10 - Domestic food price index	**Segmentation cluster 5** FAO5 - Average supply of protein of animal origin FAO12 - Share of food expenditure of the poor FAO19 - Domestic food price volatility
(1) Linear Regression on Log(Stunting)	+	Peru South Africa	Turkey	Vietnam	Angola			Ghana
	-					Burundi Madagascar Pakistan	Mali	Afghanistan Egypt
	+/-							Ivory Coast

HBGDki India

Looking at shifts in segmentation distance over time: Kuwait vs. Yemen



US liberates Kuwait, increases aid; US ceases aid to Yemen

US begins military intervention in Yemen

Yemen's malnutrition more than doubles during this time, while Kuwait's decreases considerably

Kuwait in Iraq cluster
Yemen in Afghanistan cluster

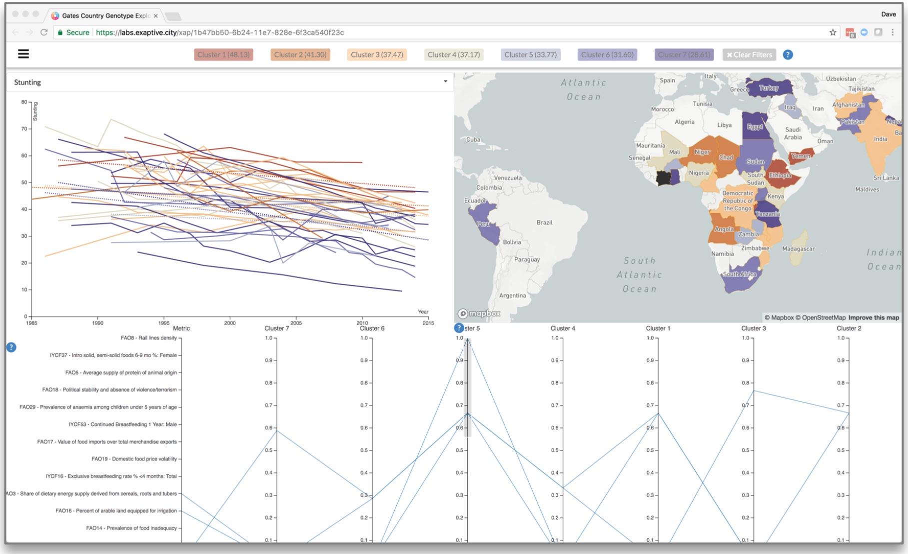
Kuwait joins cluster with India and Egypt

Yemen joins cluster with Ghana and Cote d'Ivoire

Use it Yourself:

Hands-on training to perform India-specific analysis

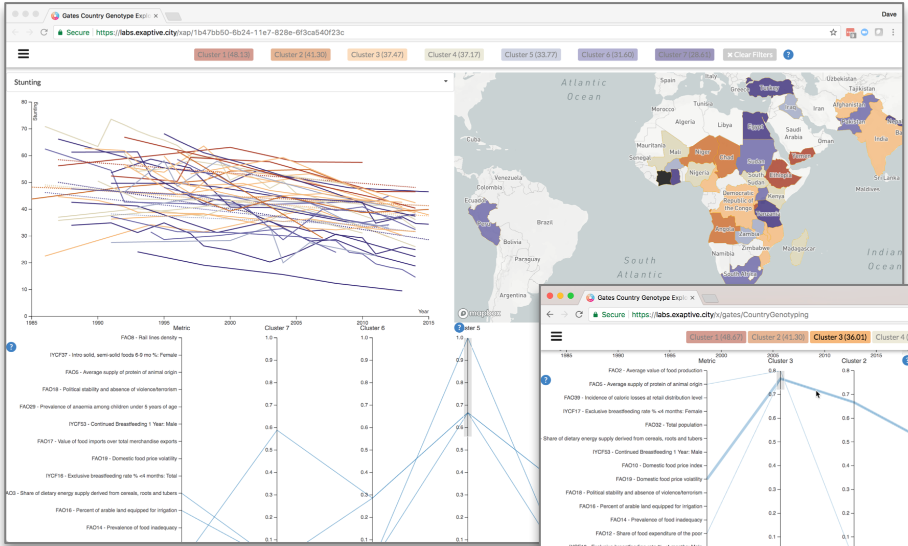
Country Segmentation Tool



Explore clusters

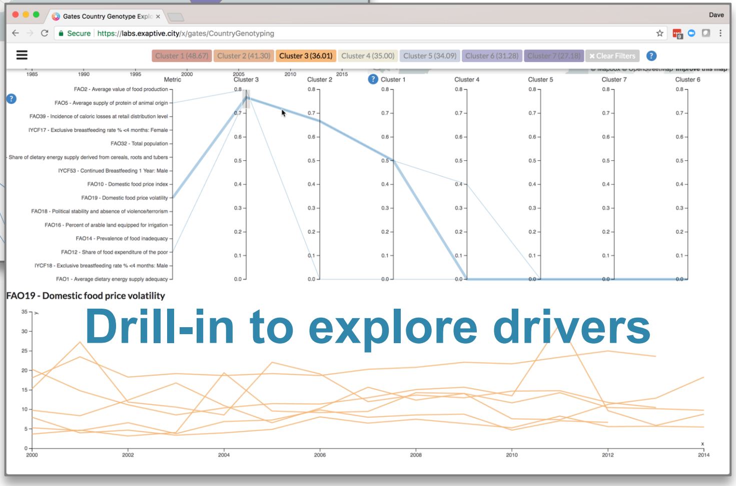
“Cluster Mode”

Country Segmentation Tool



Explore clusters

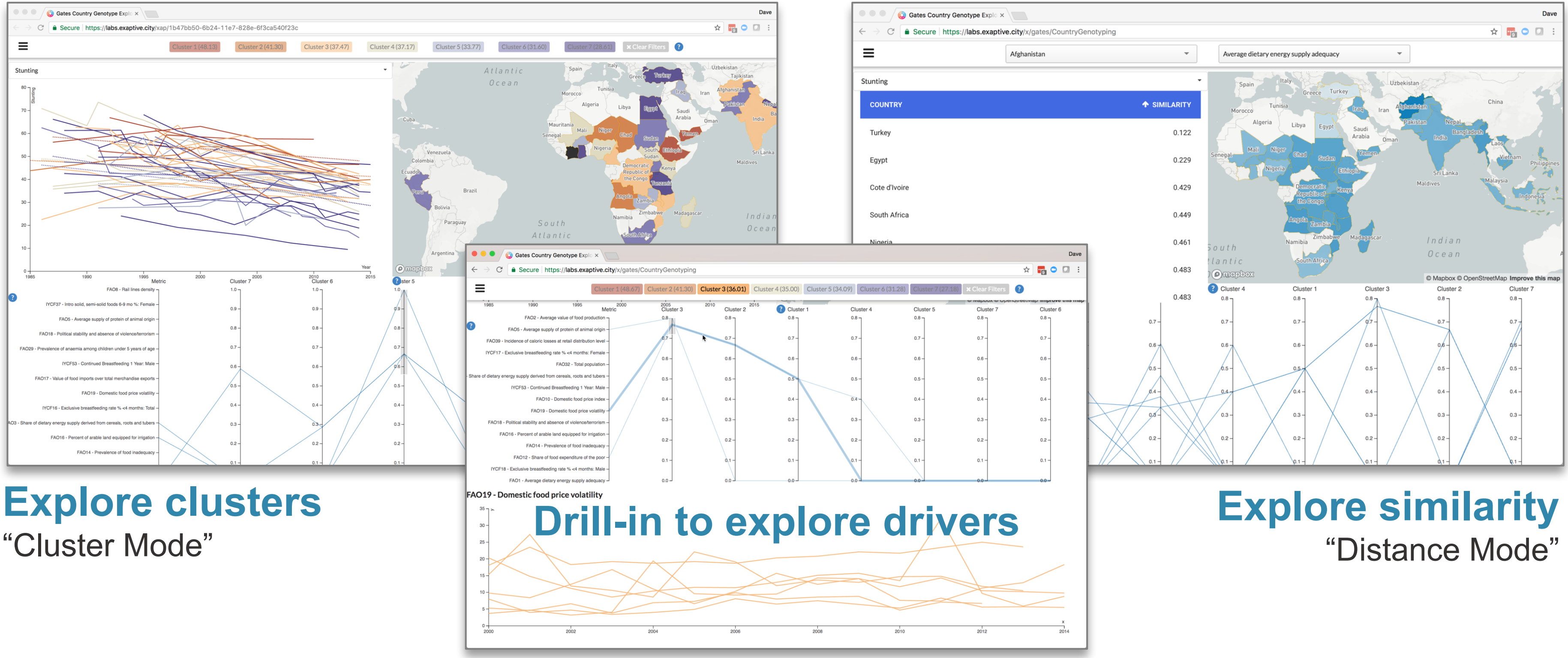
“Cluster Mode”



Drill-in to explore drivers

HBGDki India

Country Segmentation Tool

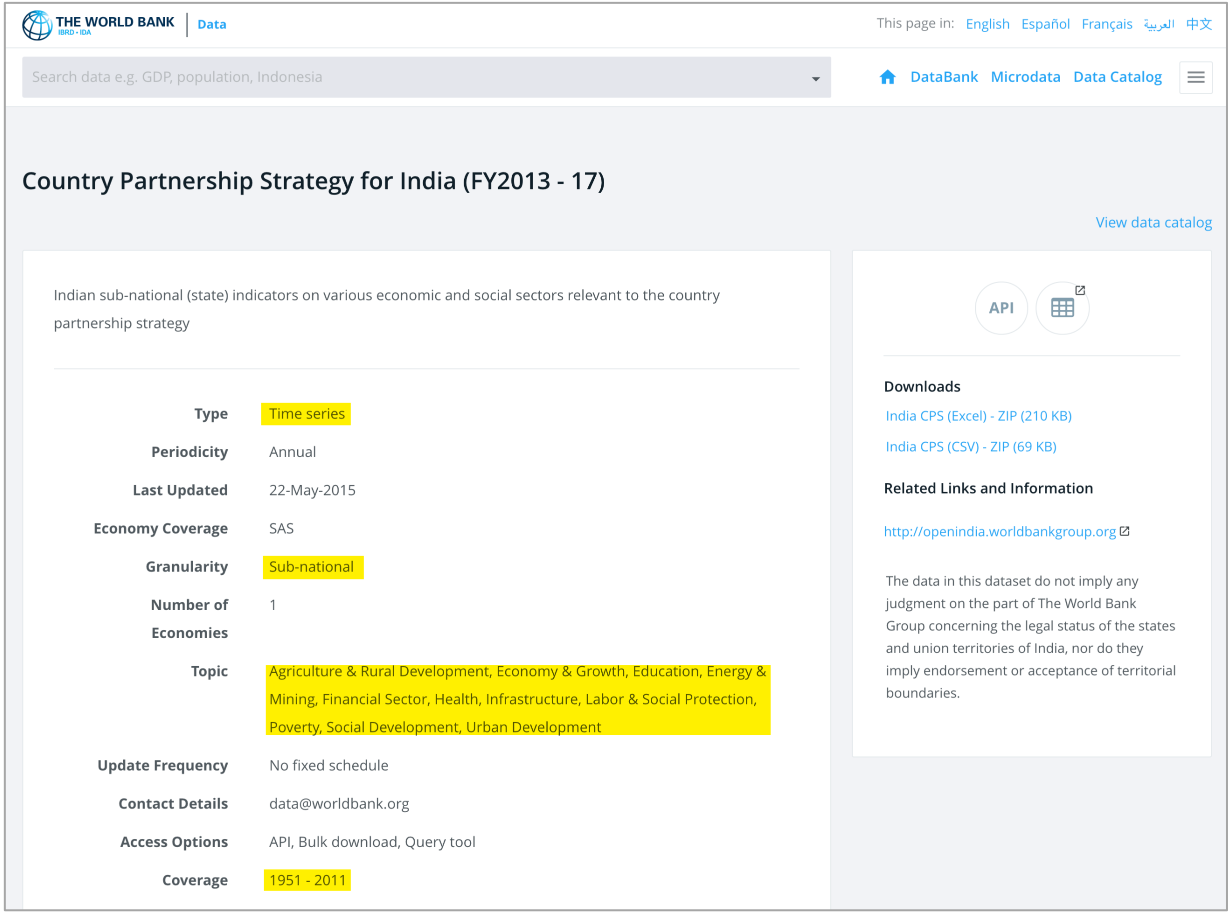


Explore clusters
"Cluster Mode"

Drill-in to explore drivers

Explore similarity
"Distance Mode"

Can "country segmentation" algorithm can also be used for regional segmentation?



THE WORLD BANK IBRD • IDA | Data

This page in: English Español Français العربية 中文

Search data e.g. GDP, population, Indonesia

DataBank Microdata Data Catalog

Country Partnership Strategy for India (FY2013 - 17)

View data catalog

Indian sub-national (state) indicators on various economic and social sectors relevant to the country partnership strategy

Type	Time series
Periodicity	Annual
Last Updated	22-May-2015
Economy Coverage	SAS
Granularity	Sub-national
Number of Economies	1
Topic	Agriculture & Rural Development, Economy & Growth, Education, Energy & Mining, Financial Sector, Health, Infrastructure, Labor & Social Protection, Poverty, Social Development, Urban Development
Update Frequency	No fixed schedule
Contact Details	data@worldbank.org
Access Options	API, Bulk download, Query tool
Coverage	1951 - 2011

API

Downloads

India CPS (Excel) - ZIP (210 KB)

India CPS (CSV) - ZIP (69 KB)

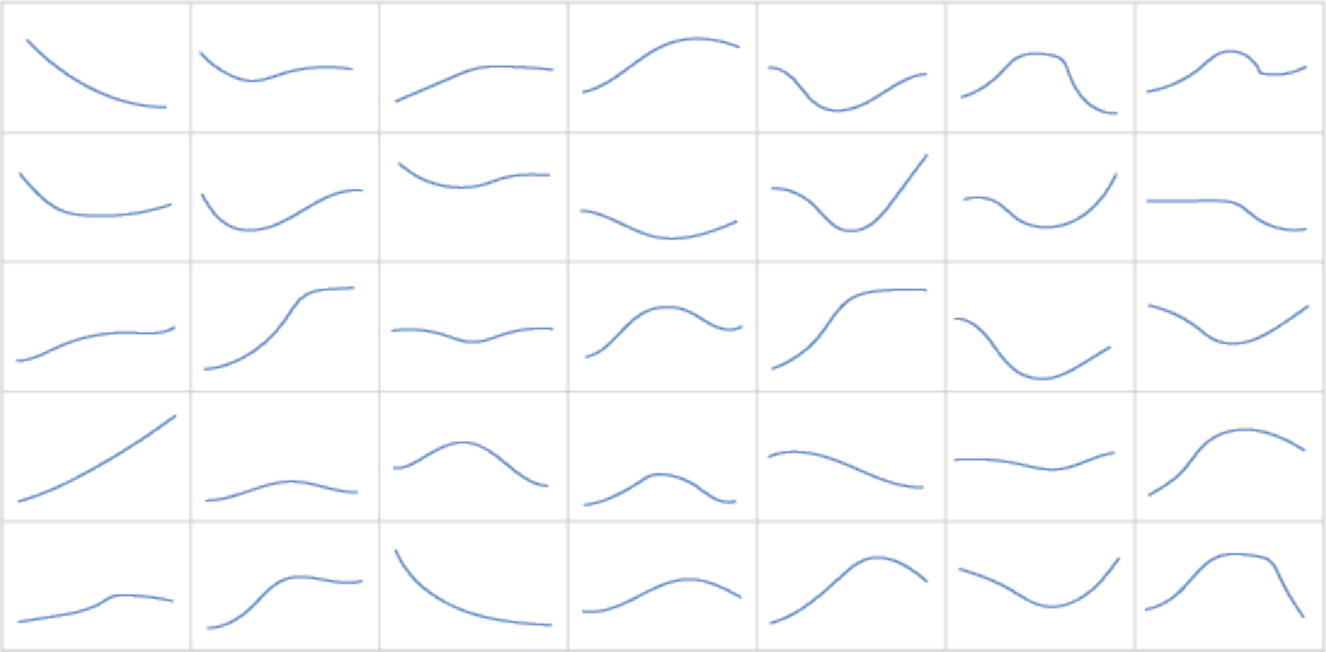
Related Links and Information

http://openindia.worldbankgroup.org

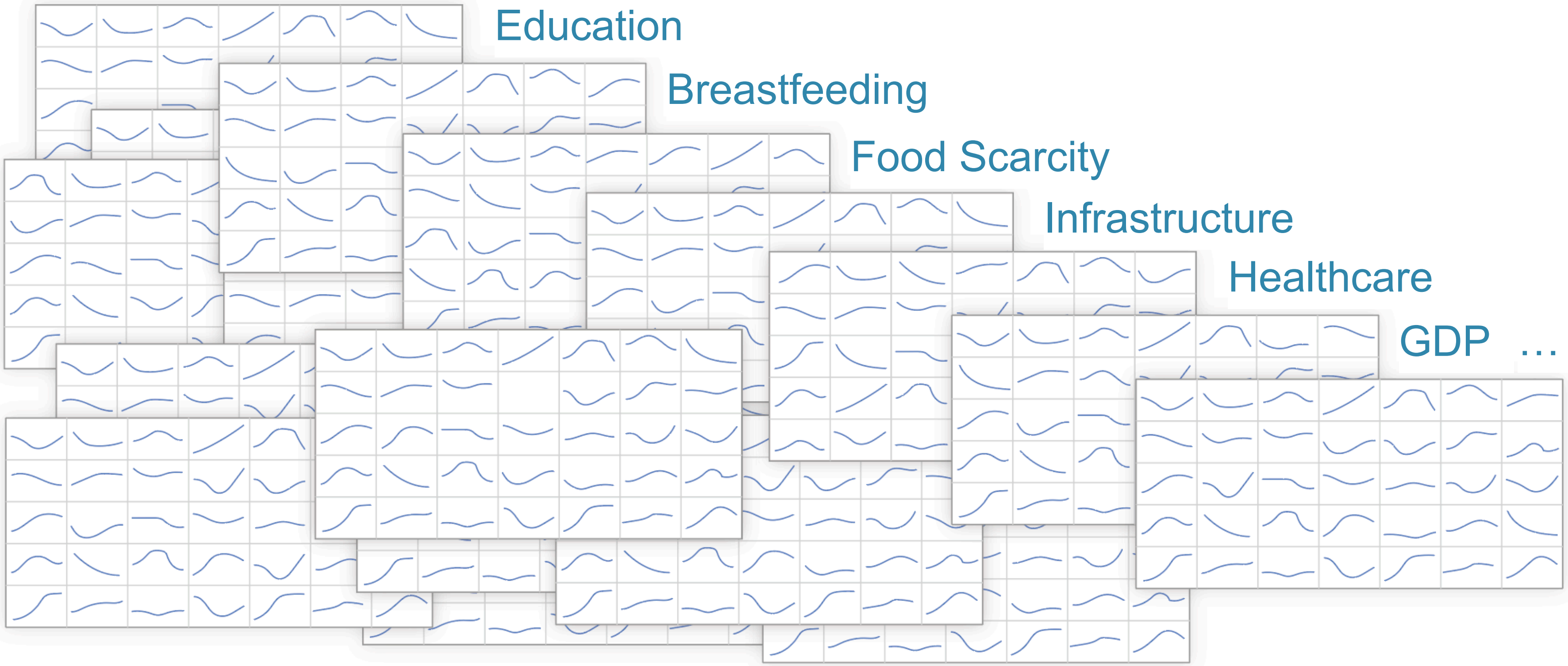
The data in this dataset do not imply any judgment on the part of The World Bank Group concerning the legal status of the states and union territories of India, nor do they imply endorsement or acceptance of territorial boundaries.

Appendix:

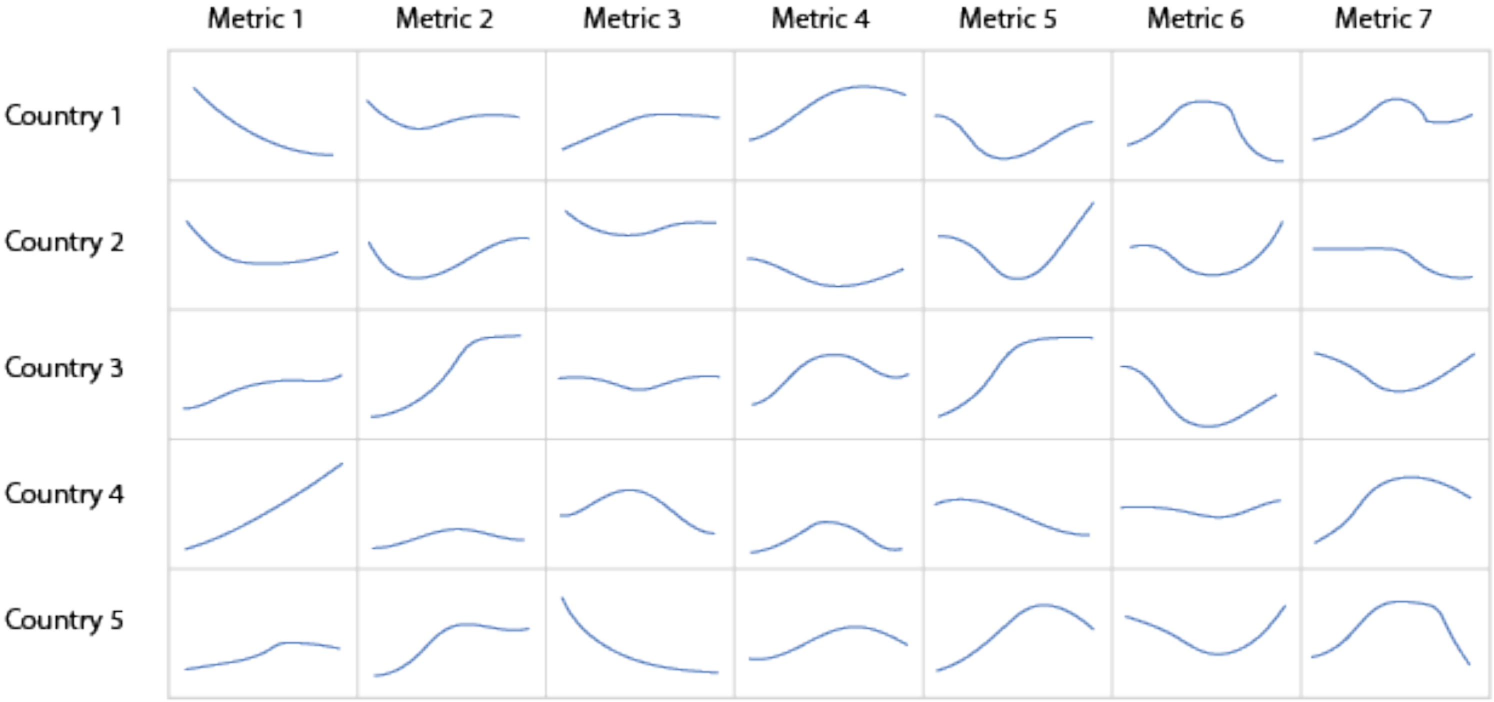
Algorithmic Details



HBGDki India

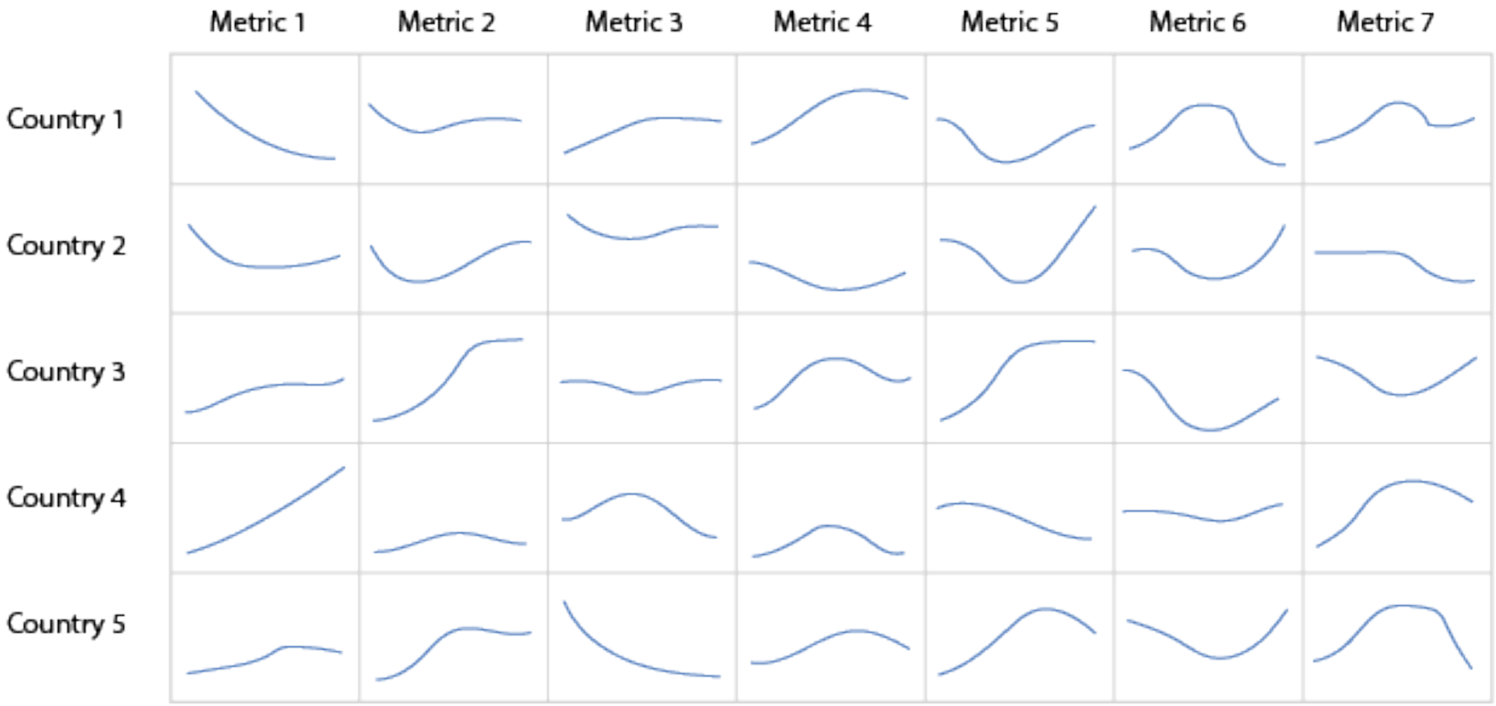


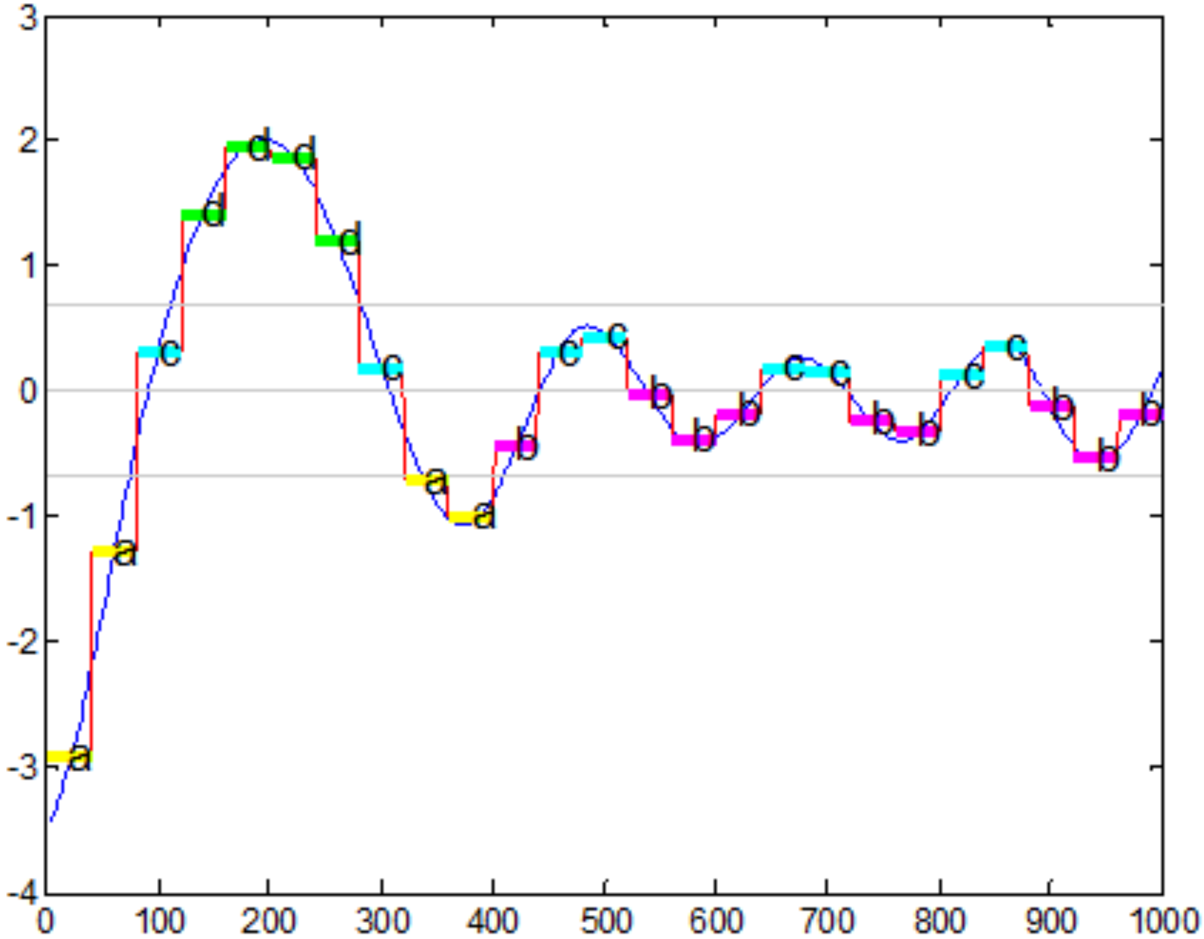
HBGDki India



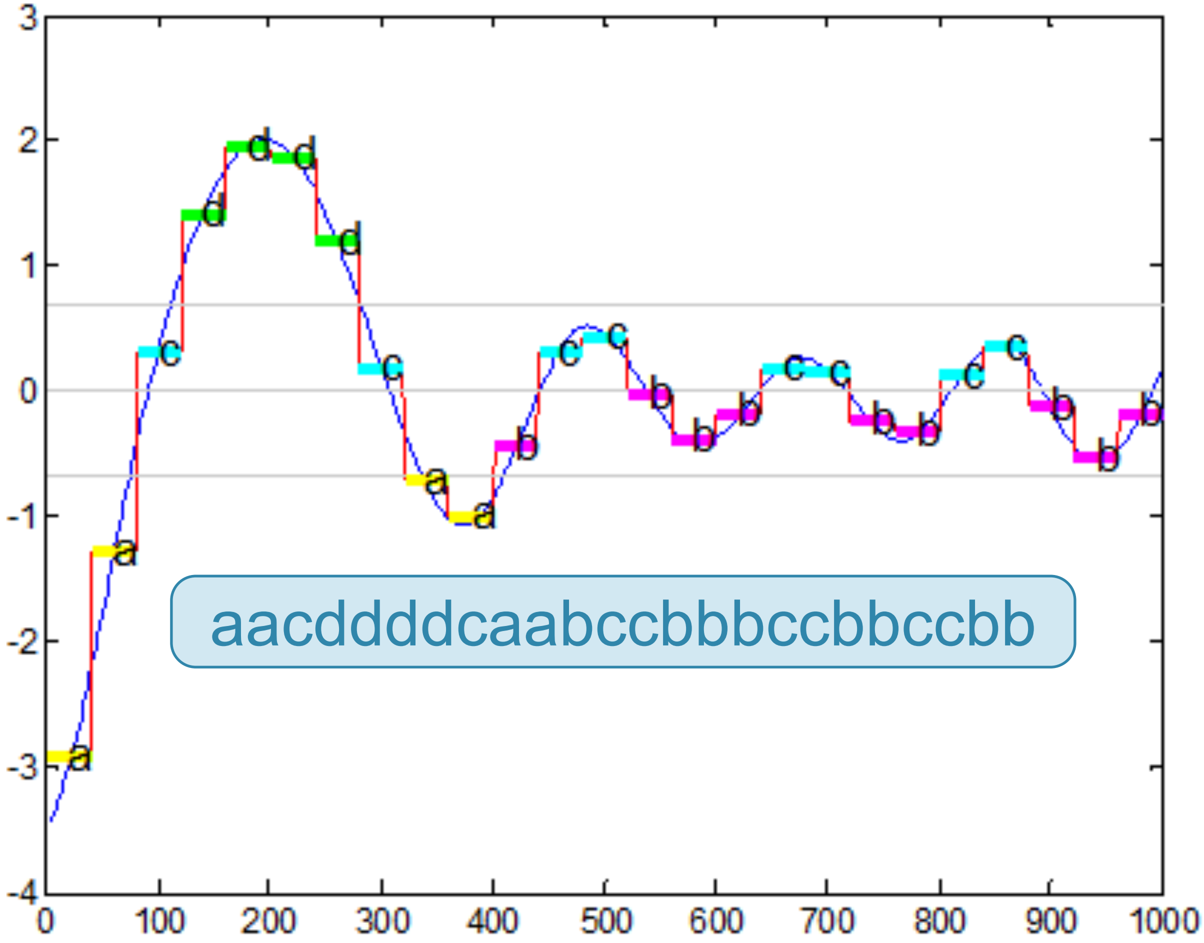
HBGDki India

	Metric 1	Metric 2	Metric 3	Metric 4	Metric 5	Metric 6	Metric 7
Country 1	1	3	5	3	4	1	5
Country 2	2	2	3	1	5	1	2
Country 3	2	3	4	2	4	1	2
Country 4	1	2	1	3	3	2	3
Country 5	3	1	4	5	5	4	4


Metric 1
Metric 2
Metric 3
Metric 4
Metric 5
Metric 6
Metric 7
Country 1
Country 2
Country 3
Country 4
Country 5



3
2
1
0
-1
-2
-3
-4
0
100
200
300
400
500
600
700
800
900
1000
a a c d d d d d c a a b c c b b b c c b b c c b b b



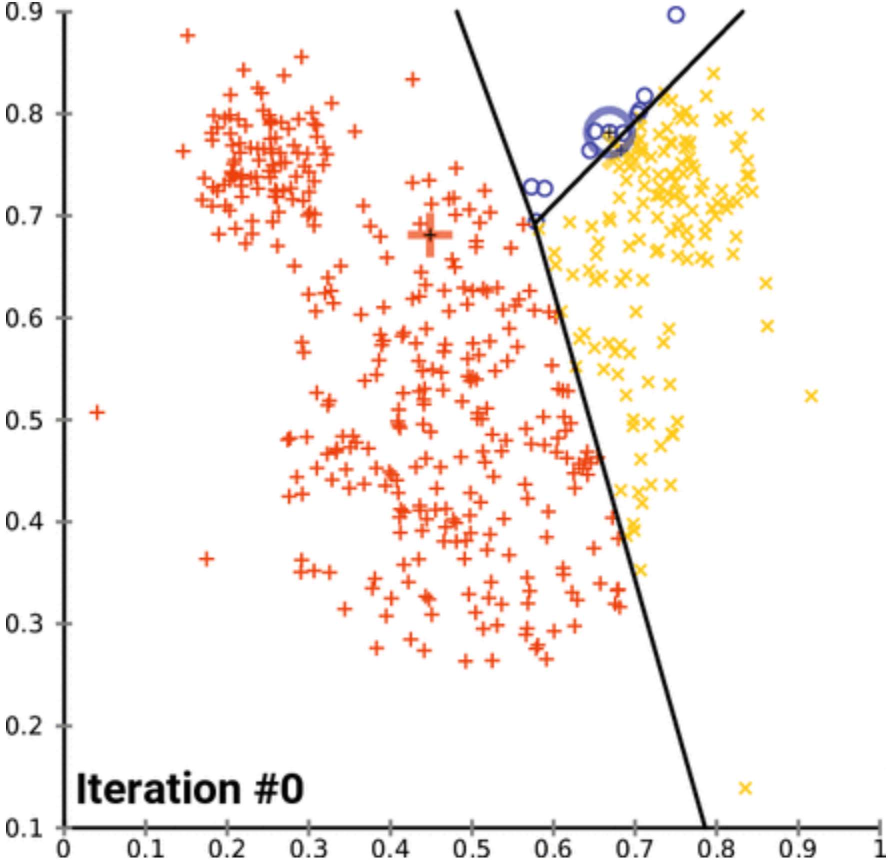
3
2
1
0
-1
-2
-3
-4
0
100
200
300
400
500
600
700
800
900
1000
aacddddcaabccbbbccbbccbb

	Metric 1	Metric 2	Metric 3	Metric 4	Metric 5	Metric 6	Metric 7
Country 1	aabcdaffg	ddedghh	jjhhhssd	aabbccdd	aajjkllsss	ccddeefd	dkifdded
Country 2	daffeghg	deghhslif	aaddssd	aaassdg	eeddffss	ccddeeffd	edfkggh
Country 3	aabcdaffg	ddsseefh	eedfjkk	aabcdaffg	aabcdaffg	aabcdaffg	ddeeffgh
Country 4	aabcdaffg	aabbees	bbcccdde	aabcdaffg	aabcdaffg	ggffdkkk	cccddeeh
Country 5	aabcdaffg	jjdffhgl	llefdggs	aabcdaffg	aabcdaffg	aabcdaffg	dddeffg

Empty String

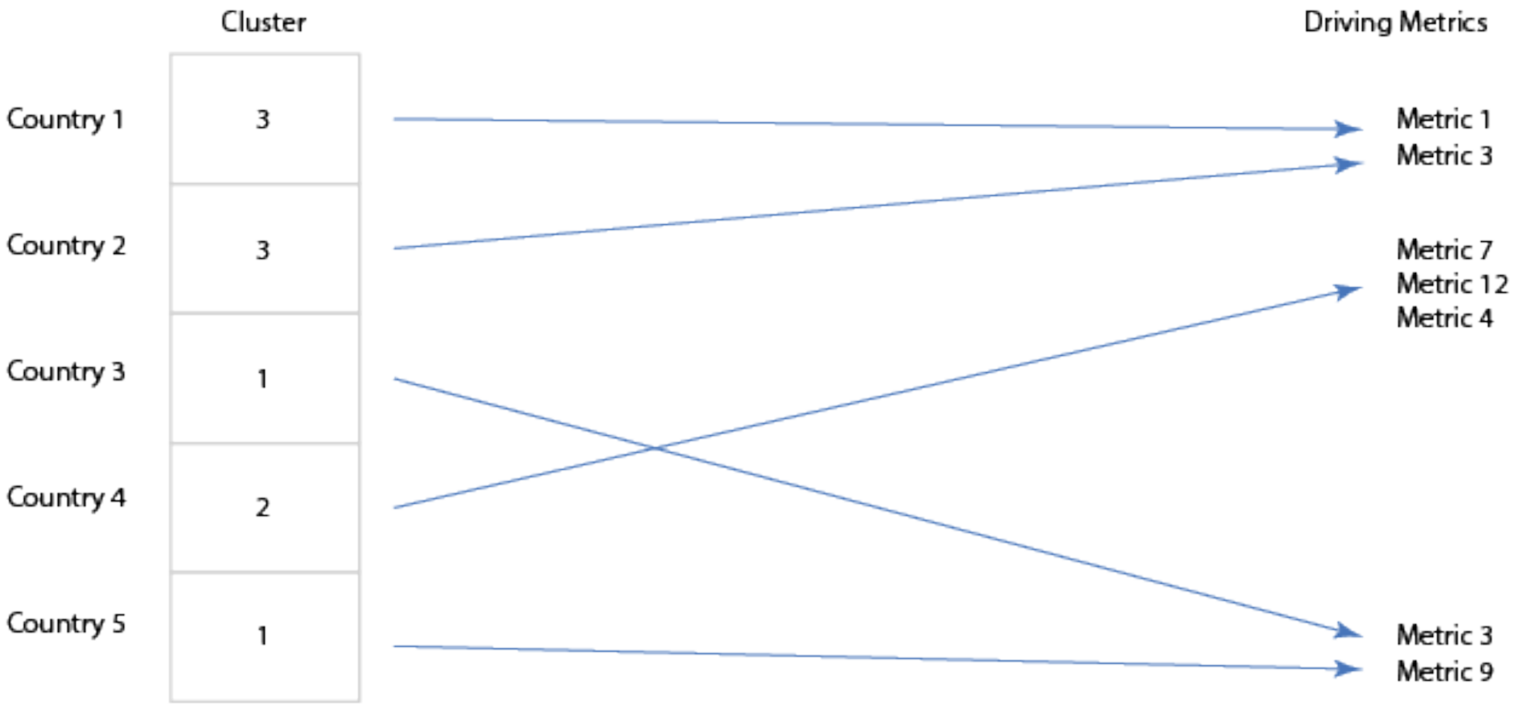
String "B"

		""	A	L	T	R	U	I	S	M
Empty String	""	0	1	2	3	4	5	6	7	8
String "A"	P	1	1	2	3	4	5	6	7	8
	L	2	2	1	2	3	4	5	6	7
	A	3	2	2	2	3	4	5	6	7
	S	4	3	3	3	3	4	5	5	6
	M	5	4	4	4	4	4	5	6	5
	A	6	5	5	5	5	5	5	6	6


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Iteration #0

	Metric 1	Metric 2	Metric 3	Metric 4	Metric 5	Metric 6	Metric 7
Country 1	1	ddedghh	jjhhhssd	aabbccdd	aajjkllsss	ccddeefd	dkifdded
Country 2	2	deghhslif	aaddssd	aaassdg	eeddffss	ccddeeffd	edfkggh
Country 3	2	ddsseefh	eedfjkk	aabcdaffg	aabcdaffg	aabcdaffg	ddeeffgh
Country 4	1	aabbees	bbcccdde	aabcdaffg	aabcdaffg	ggffdkkk	cccddeeh
Country 5	3	jjdffhgl	llefdggs	aabcdaffg	aabcdaffg	aabcdaffg	dddeffg

	Metric 1	Metric 2	Metric 3	Metric 4	Metric 5	Metric 6	Metric 7
Country 1	1	3	5	3	4	1	5
Country 2	2	2	3	1	5	1	2
Country 3	2	3	4	2	4	1	2
Country 4	1	2	1	3	3	2	3
Country 5	3	1	4	5	5	4	4


Cluster
Driving Metrics
Country 1
3
Country 2
3
Country 3
1
Country 4
2
Country 5
1
Metric 1
Metric 3
Metric 7
Metric 12
Metric 4
Metric 3
Metric 9