The Development of Historical Data Visualization – *a spatiotemporal web application supporting teaching and learning at the Harvard Business School*

Giovanni Zambotti
Benjamin Lewis
Weihe Wendy Guan*

Center for Geographic Analysis
Harvard University

Harvard Business School Collaborators

- **Geoffrey Jones**, Isidor Straus Professor of Business History
- **Walter Friedman**, Director, Harvard Business History Initiative, Lecturer of Business Administration
- **Patrick Clapp** – Harvard Business School Library, Information Research Specialist
- **Jeffrey Cronin** - Harvard Business School Library, Information Research Specialist
- HBS Educational Technology Services Team

Project Objectives

- Support a course in "Entrepreneurship and Global Capitalism"
- Support research in "Business History Initiative"

- Animate trends in the global economy
- Illustrate the history of capitalism

Themes on the Time Maps

- **International commodities**
  - bananas, coffee, petroleum
- **Global corporations**
  - BCG, Carrefour, Coca-Cola, Ford, Huawei, IBM, McDonald's, McKinsey, Pampers, Royal Dutch Shell, Singer, Tesco
- **International transportation and communication**
  - broadband subscriptions, mail handled, merchant shipping, mobile phones, personal computers, railroad lines, sea cables, and telephone rates
- **Political economy**
  - GATT, GDP, interstate conflict, League of Nations, life expectancy, literary rates, political institutions, population density, population growth rate, spread of communism, spread of imperialism, trade union membership, unemployment rates, urbanization, world's ten largest cities, women's suffrage, WTO
- **Business history of the United States**
  - foreign born, home ownership, medium home prices, plumbing

User Requirements

- Web-based, public facing
- Users select a theme among many layers and categories
- Interactive map for zoom and pan
- Report variable and value at user's click
- Interactive time bar
- Login protection when needed

Other Functions of the Website

- Branding of the website
- Introduction of the project
- Tabular data download
- Metadata viewing and download
- Scanned historical maps viewing
  - from Baker Library's collection, showing trade routes, oil exports, and other information

Platform Selection

- There was not much choice in existing temporal mapping platforms
  - MapStory is not organized for creating branded single page maps.
  - Google Earth can't show the whole world in one view;
  - Google Map API requires all geometry to be client-side and we have too much data for that even in the simple layers.
- ArcGIS Server is the only available choice

System Design

- GIS map publishing services
  - ArcGIS Server
- Temporal map rendering
  - ArcGIS Server time map client
- Platform hosting
  - Amazon EC2
- Data preparation and organization
  - ArcGIS desktop
- Historic nation state boundaries since 1875
  - Yuri Zhukov dataset
    - http://hdl.handle.net/1902.1/21595
    - http://scholar.harvard.edu/zhukov/files/neighbors_isq_preprint.pdf

Zhukov Dataset (1875-1998) - 1 global shape file per year, 124 shape files totaling 2GB

d1925.dbf

	A	B	C	D	E	F	G	H	I	J	K	L
1	SP_ID	CNTRY_CODE	COUNTRY	OID_	cyear	MAP	MAP_CCOI	year	numid	MAP_CNTRY	bbb	scode
2	2	206923.0000000000	Afghanistan	0	7001925	d1921-1939	206923	1925	700	Afghanistan	AFG	AFG
3	3	206924.0000000000	Albania	1	3391925	d1921-1939	206924	1925	339	Albania	ALB	ALB
4	10	206932.0000000000	Argentina	2	1601925	d1921-1939	206932	1925	160	Argentina	ARG	ARG
5	12	206937.0000000000	Australia	3	9001925	d1921-1939	206937	1925	900	Australia	AUL	AUL
6	13	206938.0000000000	Austria	4	3051925	d1921-1939	206938	1925	305	Austria	AUS	AUS
7	18	206945.0000000000	Belgium	5	2111925	d1921-1939	206945	1925	211	Belgium	BEL	BEL
8	23	206951.0000000000	Bolivia	6	1451925	d1921-1939	206951	1925	145	Bolivia	BOL	BOL
9	25	206954.0000000000	Brazil	7	1401925	d1921-1939	206954	1925	140	Brazil	BRA	BRA
10	28	206957.0000000000	Bulgaria	8	3551925	d1921-1939	206957	1925	355	Bulgaria	BUL	BUL
11	34	206963.0000000000	Canada	9	201925	d1921-1939	206963	1925	20	Canada	CAN	CAN
12	39	206969.0000000000	Chile	10	1551925	d1921-1939	206969	1925	155	Chile	CHL	CHL
13	245	206970.0000000000	China	11	7101925	d1921-1939	206970	1925	710	China	CHN	CHN
14	43	206974.0000000000	Colombia	12	1001925	d1921-1939	206974	1925	100	Colombia	COL	COL
15	47	206979.0000000000	Costa Rica	13	941925	d1921-1939	206979	1925	94	Costa Rica	COS	COS
16	48	206981.0000000000	Cuba	14	401925	d1921-1939	206981	1925	40	Cuba	CUB	CUB
17	50	206984.0000000000	Denmark	16	3901925	d1921-1939	206984	1925	390	Denmark	DEN	DEN
18	53	206987.0000000000	Dominican Republic	17	421925	d1921-1939	206987	1925	42	Dominican Republic	DOM	DOM
19	54	206988.0000000000	Ecuador	18	1301925	d1921-1939	206988	1925	130	Ecuador	ECU	ECU
20	55	206989.0000000000	Egypt	19	6511925	d1921-1939	206989	1925	651	Egypt	EGY	EGY
21	56	206990.0000000000	El Salvador	20	921925	d1921-1939	206990	1925	92	El Salvador	SAL	SAL
22	58	206992.0000000000	Estonia	21	3661925	d1921-1939	206992	1925	366	Estonia	EST	EST
23	59	206993.0000000000	Ethiopia	22	5301925	d1921-1939	206993	1925	530	Ethiopia	ETH	ETH
24	63	206997.0000000000	Finland	23	3751925	d1921-1939	206997	1925	375	Finland	FIN	FIN
25	64	206998.0000000000	France	24	2201925	d1921-1939	206998	1925	220	France	FRN	FRN
26	243	207005.0000000000	Germany	25	2551925	d1921-1939	207005	1925	255	Germany	GMY	GMY
27	71	207008.0000000000	Greece	26	3501925	d1921-1939	207008	1925	350	Greece	GRC	GRC
28	76	207013.0000000000	Guatemala	27	901925	d1921-1939	207013	1925	90	Guatemala	GUA	GUA
29	81	207018.0000000000	Honduras	28	911925	d1921-1939	207018	1925	91	Honduras	HON	HON
30	82	207020.0000000000	Hungary	29	3101925	d1921-1939	207020	1925	310	Hungary	HUN	HUN
31	86	207025.0000000000	Iran	30	6301925	d1921-1939	207025	1925	630	Iran	IRN	IRN

System Development

- Develop data creation workflow and guide the HBS staff to provide tabular data in required format
- Install and configure server software
- Develop web client application
- Load data to system, check for errors, make corrections
- Test system for performance and stability, make adjustments as needed

Data Creation Work Flow

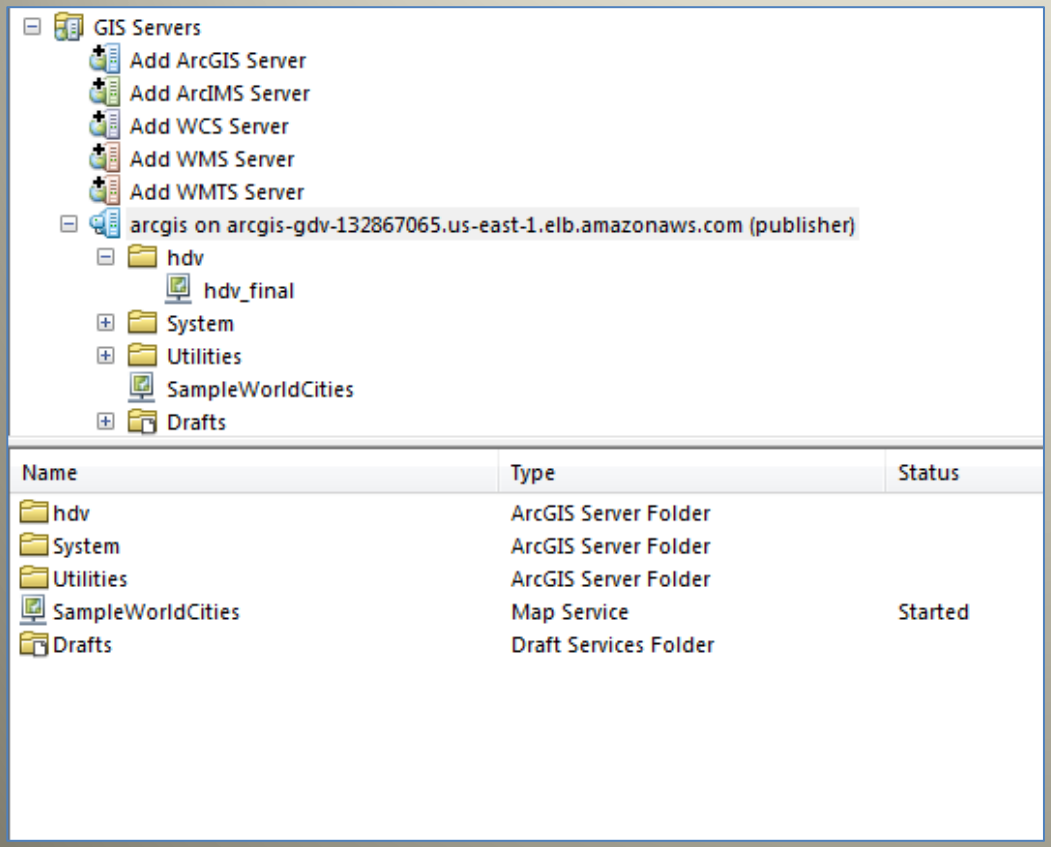
- Append all Zhukov shapefiles into a single Feature Class.
- Export the merged Feature Class attribute table in .csv format, containing unique IDs for country and year.
- Provide the table to historians to add their variables.
- Join tables with historical variables back to the Feature Class by the unique ID.
- Extract just the features required to render each variable (of different time range) into a separate Feature Class.
- Symbolize in ArcGIS Desktop, serve out from ArcGIS Server.
- Review, correct data, adjust symbology.

Setting up ArcGIS Server on Amazon EC2

- Create an AWS account
- Request ESRI to connect the AWS account with their AWS AMI (Amazon Machine Images) stack
- Create an Access Key ID and Secret access key on the AWS account (needed for connection between ArcGIS Server and Cloud Builder)
- Install Cloud Builder on a local desktop (needed for managing the ArcGIS Server on AWS AMI)
- Set up auto-scaling on the AWS (optional)

Publish to AWS ArcGIS Server

Connect to AWS Linux Server from Desktop ArcCatalog



Publish the map server on AWS ArcGIS Server (services directory tree)

Spatiotemporal Web Client Development

Built from scratch in JavaScript --

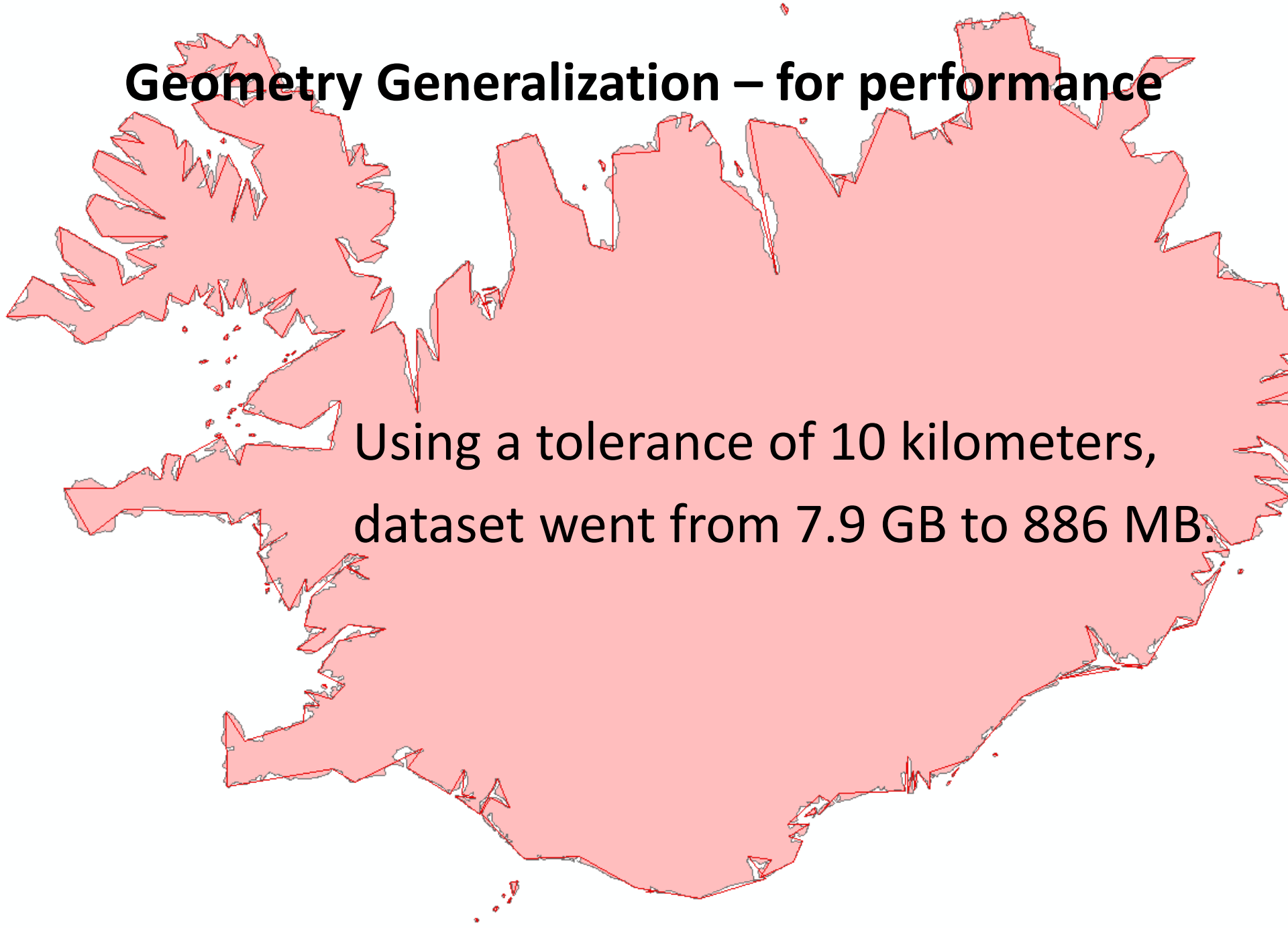
- **jQuery UI** https://jqueryui.com/ to control the UI behavior
- **ArcGIS JS API** https://developers.arcgis.com/javascript/ to connect to ArcGIS Server and control the Map behavior

Data Challenges

- Varying time increments in the data
- Differences in ways of referring to country boundaries
- Units of measurements from different sources

- Communication between the historians and the GIS programmers
- Physical distance between the two teams

Geometry Generalization – for performance



Using a tolerance of 10 kilometers, dataset went from 7.9 GB to 886 MB.

AWS Auto Scaling

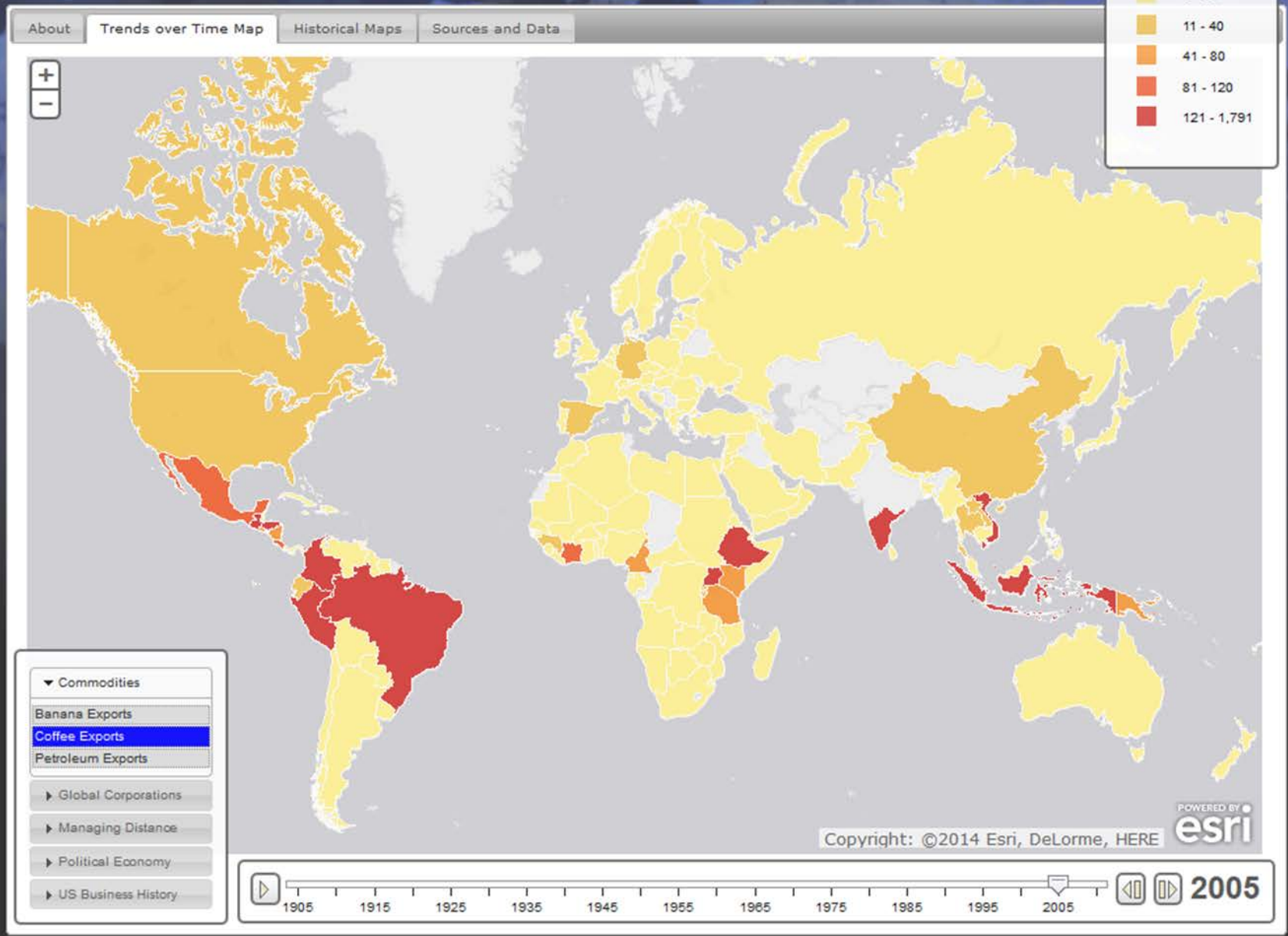
- ArcGIS Server does not cache temporal data(unlike regular mapping layers) so time change in every layer is a new call to the server to render a globe of country polygons.
- This is not a scalable web solution, but we did not have the budget to build a scalable system.
- We installed ArcGIS server on Amazon EC2 to allow the system to automatically spawn additional servers when load reaches a given threshold.
- This worked in testing mode, but has not been needed in production so far.
- There is a minute or so delay in firing up an additional server, therefore would still have performance issues in case a huge number of users hit the site simultaneously.

Lessons Learned and Thoughts for the Future

- Instead of providing an ideal solution for spatiotemporal data animation, this project attempted to address this common need with limited available tools; and by doing so, exposed an unmet need for such a robust system.
- For ArcGIS Server, caching temporal data would be a nice enhancement.
- Given adequate resources, developing an open source solution for generic temporal maps rendering could be worthwhile.

Historical Data Visualization

http://hdv.cga.harvard.edu HDV_user



Copyright © 2014 President and Fellows of Harvard College

HARVARD | BUSINESS | SCHOOL

Thank You

- We welcome your feedback, commentary, and corrections.
- For data content related issues, please contact
  - **wfriedman@hbs.edu**, or
  - **etg@hbs.edu**
- For GIS and web related issues, please contact
  - **gzambotti@cga.harvard.edu**

HARVARD BUSINESS SCHOOL

VE RI TAS

Center for Geographic Analysis
Harvard University