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://hdl.handle.net/1902.1/21595)
Zhukov Dataset (1875-1998) - 1 global shape file per year, 124 shape files totaling 2GB

<table>
<thead>
<tr>
<th>SP ID</th>
<th>COUNTRY_CODE</th>
<th>COUNTRY</th>
<th>OID_year</th>
<th>MAP</th>
<th>MAP_CCOI_year</th>
<th>numid</th>
<th>MAP_CTRY</th>
<th>bbb</th>
<th>scode</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>206923.000000000000000</td>
<td>Afghanistan</td>
<td>0</td>
<td>7001925</td>
<td>d1921-1939</td>
<td>206923</td>
<td>1925</td>
<td>700</td>
<td>AFG</td>
</tr>
<tr>
<td>3</td>
<td>206924.000000000000000</td>
<td>Albania</td>
<td>1</td>
<td>3391925</td>
<td>d1921-1939</td>
<td>206924</td>
<td>1925</td>
<td>339</td>
<td>ALB</td>
</tr>
<tr>
<td>4</td>
<td>206932.000000000000000</td>
<td>Argentina</td>
<td>2</td>
<td>1601925</td>
<td>d1921-1939</td>
<td>206932</td>
<td>1925</td>
<td>160</td>
<td>ARG</td>
</tr>
<tr>
<td>5</td>
<td>206937.000000000000000</td>
<td>Australia</td>
<td>3</td>
<td>9001925</td>
<td>d1921-1939</td>
<td>206937</td>
<td>1925</td>
<td>900</td>
<td>AUL</td>
</tr>
<tr>
<td>6</td>
<td>206938.000000000000000</td>
<td>Austria</td>
<td>4</td>
<td>3051925</td>
<td>d1921-1939</td>
<td>206938</td>
<td>1925</td>
<td>305</td>
<td>AUS</td>
</tr>
<tr>
<td>7</td>
<td>206945.000000000000000</td>
<td>Belgium</td>
<td>5</td>
<td>2111925</td>
<td>d1921-1939</td>
<td>206945</td>
<td>1925</td>
<td>211</td>
<td>BEL</td>
</tr>
<tr>
<td>8</td>
<td>206951.000000000000000</td>
<td>Bolivia</td>
<td>6</td>
<td>1451925</td>
<td>d1921-1939</td>
<td>206951</td>
<td>1925</td>
<td>145</td>
<td>BOL</td>
</tr>
<tr>
<td>9</td>
<td>206954.000000000000000</td>
<td>Brazil</td>
<td>7</td>
<td>1401925</td>
<td>d1921-1939</td>
<td>206954</td>
<td>1925</td>
<td>140</td>
<td>BRA</td>
</tr>
<tr>
<td>10</td>
<td>206957.000000000000000</td>
<td>Bulgaria</td>
<td>8</td>
<td>3551925</td>
<td>d1921-1939</td>
<td>206957</td>
<td>1925</td>
<td>355</td>
<td>BUL</td>
</tr>
<tr>
<td>11</td>
<td>206963.000000000000000</td>
<td>Canada</td>
<td>9</td>
<td>201925</td>
<td>d1921-1939</td>
<td>206963</td>
<td>1925</td>
<td>20</td>
<td>CAN</td>
</tr>
<tr>
<td>12</td>
<td>206969.000000000000000</td>
<td>Chile</td>
<td>10</td>
<td>1551925</td>
<td>d1921-1939</td>
<td>206969</td>
<td>1925</td>
<td>155</td>
<td>CHL</td>
</tr>
<tr>
<td>13</td>
<td>206970.000000000000000</td>
<td>China</td>
<td>11</td>
<td>7101925</td>
<td>d1921-1939</td>
<td>206970</td>
<td>1925</td>
<td>710</td>
<td>CHN</td>
</tr>
<tr>
<td>14</td>
<td>206974.000000000000000</td>
<td>Colombia</td>
<td>12</td>
<td>1001925</td>
<td>d1921-1939</td>
<td>206974</td>
<td>1925</td>
<td>100</td>
<td>COL</td>
</tr>
<tr>
<td>15</td>
<td>206979.000000000000000</td>
<td>Costa Rica</td>
<td>13</td>
<td>941925</td>
<td>d1921-1939</td>
<td>206979</td>
<td>1925</td>
<td>94</td>
<td>COS</td>
</tr>
<tr>
<td>16</td>
<td>206981.000000000000000</td>
<td>Cuba</td>
<td>14</td>
<td>401925</td>
<td>d1921-1939</td>
<td>206981</td>
<td>1925</td>
<td>40</td>
<td>CUB</td>
</tr>
<tr>
<td>17</td>
<td>206984.000000000000000</td>
<td>Denmark</td>
<td>16</td>
<td>3901925</td>
<td>d1921-1939</td>
<td>206984</td>
<td>1925</td>
<td>390</td>
<td>DEN</td>
</tr>
<tr>
<td>18</td>
<td>206987.000000000000000</td>
<td>Dominican Republic</td>
<td>17</td>
<td>421925</td>
<td>d1921-1939</td>
<td>206987</td>
<td>1925</td>
<td>42 Dominican Republic</td>
<td>DOM</td>
</tr>
<tr>
<td>19</td>
<td>206988.000000000000000</td>
<td>Ecuador</td>
<td>18</td>
<td>1301925</td>
<td>d1921-1939</td>
<td>206988</td>
<td>1925</td>
<td>130</td>
<td>ECU</td>
</tr>
<tr>
<td>20</td>
<td>206989.000000000000000</td>
<td>Egypt</td>
<td>19</td>
<td>6511925</td>
<td>d1921-1939</td>
<td>206989</td>
<td>1925</td>
<td>651</td>
<td>EGY</td>
</tr>
<tr>
<td>21</td>
<td>206990.000000000000000</td>
<td>El Salvador</td>
<td>20</td>
<td>921925</td>
<td>d1921-1939</td>
<td>206990</td>
<td>1925</td>
<td>92 El Salvador</td>
<td>SAL</td>
</tr>
<tr>
<td>22</td>
<td>206992.000000000000000</td>
<td>Estonia</td>
<td>21</td>
<td>3661925</td>
<td>d1921-1939</td>
<td>206992</td>
<td>1925</td>
<td>366</td>
<td>EST</td>
</tr>
<tr>
<td>23</td>
<td>206993.000000000000000</td>
<td>Ethiopia</td>
<td>22</td>
<td>5301925</td>
<td>d1921-1939</td>
<td>206993</td>
<td>1925</td>
<td>530</td>
<td>ETH</td>
</tr>
<tr>
<td>24</td>
<td>206997.000000000000000</td>
<td>Finland</td>
<td>23</td>
<td>3751925</td>
<td>d1921-1939</td>
<td>206997</td>
<td>1925</td>
<td>375</td>
<td>FIN</td>
</tr>
<tr>
<td>25</td>
<td>206998.000000000000000</td>
<td>France</td>
<td>24</td>
<td>2201925</td>
<td>d1921-1939</td>
<td>206998</td>
<td>1925</td>
<td>220</td>
<td>FRN</td>
</tr>
<tr>
<td>26</td>
<td>207005.000000000000000</td>
<td>Germany</td>
<td>25</td>
<td>2551925</td>
<td>d1921-1939</td>
<td>207005</td>
<td>1925</td>
<td>255</td>
<td>GMY</td>
</tr>
<tr>
<td>27</td>
<td>207008.000000000000000</td>
<td>Greece</td>
<td>26</td>
<td>3501925</td>
<td>d1921-1939</td>
<td>207008</td>
<td>1925</td>
<td>350</td>
<td>GRE</td>
</tr>
<tr>
<td>28</td>
<td>207013.000000000000000</td>
<td>Guatemala</td>
<td>27</td>
<td>901925</td>
<td>d1921-1939</td>
<td>207013</td>
<td>1925</td>
<td>90</td>
<td>GUA</td>
</tr>
<tr>
<td>29</td>
<td>207018.000000000000000</td>
<td>Honduras</td>
<td>29</td>
<td>911925</td>
<td>d1921-1939</td>
<td>207018</td>
<td>1925</td>
<td>91</td>
<td>HON</td>
</tr>
<tr>
<td>30</td>
<td>207020.000000000000000</td>
<td>Hungary</td>
<td>29</td>
<td>3101925</td>
<td>d1921-1939</td>
<td>207020</td>
<td>1925</td>
<td>310</td>
<td>HUN</td>
</tr>
<tr>
<td>31</td>
<td>207025.000000000000000</td>
<td>Iran</td>
<td>30</td>
<td>6301925</td>
<td>d1921-1939</td>
<td>207025</td>
<td>1925</td>
<td>630</td>
<td>IRN</td>
</tr>
</tbody>
</table>
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)
Built from scratch in JavaScript --

- **jQuery UI** [https://jqueryui.com/](https://jqueryui.com/) to control the UI behavior

- **ArcGIS JS API** [https://developers.arcgis.com/javascript/](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.
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