DiviVis: Exploration into Socio-Economic Factors that can Potentially Affect Individual Internet Usage with Visualization

Shirlett Hall, Peyvand Forouzandeh
CPSC 547
December 12, 2017
Abstract and Goals

By providing illustrations and graphs which enable us to compare socio-economic factors with Internet use, we aim to find answers to the following questions/issues:

1. Internet use within geographic boundaries
2. Trends in Internet usage and potential factors
3. Testing and comparing social and economic factors in relation to the Internet usage

Goals:

- Problem-driven programming project
- Hypothesis generating process
- User guided interactivity
# Current State of Dataset

**Database:** United Nations Data Retrieval System

**Data Sources:**
1. International Labor Organization (ILO) - Unemployment Rate
2. World Health Organization (WHO) - ex. Median Life Expectancy
3. International Telecommunications Union (ITU) - Internet Use
4. United Nation Statistics Division (UNSD) - Gross National Income

## Snapshot of the Factors Dataset:

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Internet_Users_per_100</th>
<th>Mobile_Subs_per_100</th>
<th>Toi_pop</th>
<th>Percent_rural</th>
<th>Percent_urban</th>
<th>GNI_per_cap</th>
<th>Primary_Comp_Rate</th>
<th>Per_Access_Electricity</th>
<th>Per_Adult_Unemployment</th>
<th>Dom_Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>2008</td>
<td>1.840000</td>
<td>29.22037</td>
<td>23511400</td>
<td>77.32360</td>
<td>22.67070</td>
<td>370</td>
<td>58.22502</td>
<td>NA</td>
<td>42.40000</td>
<td>NA</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>2009</td>
<td>3.650000</td>
<td>37.09494</td>
<td>23090800</td>
<td>77.04870</td>
<td>22.96330</td>
<td>470</td>
<td>66.60989</td>
<td>NA</td>
<td>42.70000</td>
<td>NA</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>2010</td>
<td>4.000000</td>
<td>45.77817</td>
<td>24985500</td>
<td>75.70955</td>
<td>23.23937</td>
<td>520</td>
<td>56.57083</td>
<td>NA</td>
<td>42.70000</td>
<td>NA</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>2011</td>
<td>5.000000</td>
<td>50.32932</td>
<td>25887700</td>
<td>76.41852</td>
<td>23.58388</td>
<td>570</td>
<td>59.32795</td>
<td>NA</td>
<td>51.51442</td>
<td>NA</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>2012</td>
<td>6.454846</td>
<td>66.45219</td>
<td>28900100</td>
<td>78.11797</td>
<td>23.82010</td>
<td>720</td>
<td>69.69781</td>
<td>NA</td>
<td>66.10000</td>
<td>NA</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>2013</td>
<td>6.900000</td>
<td>70.00100</td>
<td>20023100</td>
<td>75.84900</td>
<td>24.11500</td>
<td>730</td>
<td>00.02827</td>
<td>NA</td>
<td>75.15437</td>
<td>NA</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>2014</td>
<td>7.000000</td>
<td>74.88284</td>
<td>28556754</td>
<td>75.58650</td>
<td>24.41350</td>
<td>670</td>
<td>60.37448</td>
<td>NA</td>
<td>50.50000</td>
<td>NA</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>2015</td>
<td>8.280000</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>630</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
What: Domain and Abstract Data

Selected Data:
17,875 items
2 Categorical attributes
7 Quantitative and ordinal attributes

Key attribute:
- Internet Usage per 100 inhabitants

Other attributes:
- Median Life Expectancy,
- Gross National Income per capita,
- Percentage of Completed Primary Education,
- Percentage of Adult Unemployment,
- Percentage of Population in Urban Areas,
- Percentage with Access to Electricity
Why: Introduction And Context

Digital Divide and Technology Diffusion: Is an economic and social inequality with regard to access to, use of, or impact of information and communication Technologies (ICT).

Two major divides exist:
- an access divide and
- a skills divide [1]

Target Audience:
Students, Researchers, Data Analysts in Government or Telecom, Policy Makers

➔ Understanding trends across collection of time-varying tabular data
➔ Understanding relationships between variables
➔ Looking at the distribution of the variable across geographic regions
➔ Comparisons between countries and attributes
➔ Measuring the degree of the correlation between the main attribute and specific attributes
Related Work

- Global Web Index: Global social network penetration [4]

- Oxford Internet Institute: Internet penetration and population [2]

- Robertson et al. animation tool showing life expectancy and infant mortality [3]

- International Telecommunication Union: Individual Internet user growth rate [5]
How: Initial Steps We Took

- Choropleth Map
- Bar charts
- Polar charts or radar graphs
Solutions - Tasks

Level 1:
- Dot Plot: to show a general overview of the relative positions of all the countries
Solutions - Tasks

Level 2:

- Geo-Spatial Map visualization (bubble map): distribution, correlation; locate clusters, outliers
- Multiple line charts: to show and compare trends
Solutions - Tasks

Level 3:
- Parallel Coordinates Line Charts: to lookup and compare values, static layers, distinguished with color

Level 4:
- Linear Regression Line and Scatterplot: to find trends, outliers, distribution, correlation;
- Tables: to find detailed information
# Summary of Idioms

<table>
<thead>
<tr>
<th>Idiom</th>
<th>What</th>
<th>How (encode)</th>
<th>Why</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dot chart</td>
<td>One quantitative value attribute, one ordered key attribute</td>
<td>Express value attribute with aligned vertical position and point marks.</td>
<td>Find accumulation</td>
</tr>
<tr>
<td>Parallel Coordinates</td>
<td>Multidimensional table: categorical attribute (Country name), six quantitative value attribute</td>
<td>Line charts, colored by name of the country categorical attribute, hover</td>
<td>Lookup and compare values, static layers, distinguished with color</td>
</tr>
<tr>
<td>Line Charts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geo-Spatial Bubble Map</td>
<td>Two quantitative value attributes</td>
<td>Express values with horizontal and vertical spatial position</td>
<td>Find outliers, distribution, correlation; locate clusters</td>
</tr>
<tr>
<td>Scatterplot</td>
<td>Two quantitative value attributes</td>
<td>Express values with horizontal and vertical spatial position and point marks</td>
<td></td>
</tr>
<tr>
<td>Multiple Line chart</td>
<td>One quantitative value attribute, one ordered key attribute</td>
<td>Dot chart with connection marks between dots</td>
<td>Show and compare trends</td>
</tr>
<tr>
<td>Table</td>
<td>List of quantitative and categorical attributes</td>
<td>Express values in rows and columns</td>
<td>Find detailed information</td>
</tr>
</tbody>
</table>
Implementation

- User Interface - Shiny Web
  Application built on top of R

- Code Languages - R, HTML, CSS, Javascript

<table>
<thead>
<tr>
<th>Other Major Supporting R Libraries</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plotly</td>
<td>Build Interactivity in parallel coordinates plot and dot plot to reduce the effects of occlusion and scrolling</td>
</tr>
<tr>
<td>ColorBrewer</td>
<td>Get consistent color scheme on all graphs for sequential ordering and categories</td>
</tr>
<tr>
<td>Leaflet</td>
<td>Create mobile-friendly interactive map</td>
</tr>
<tr>
<td>Markdown</td>
<td>Create instructions and short descriptions of variables</td>
</tr>
<tr>
<td>DT</td>
<td>Create sortable data frame table</td>
</tr>
</tbody>
</table>
Results - Demo

A live demonstration of the application can be viewed online: https://shirlett.shinyapps.io/worldinternetusage/
Results and Future Works

Strengths

● Simple and clear to use for a large dataset and many countries
● Visualization in multiple scale and details: World, group of countries and single country
● Analysis in multiple scales of details: Internet usage trends, multiple social and economic factors, and linear regression analysis
● Flexibility to see data in tables, on the map or in the charts
● Interactive map and plots

Limitations and future work

● Missing data narrowed the options
● Break-out menu beneath map to choose Country/groups of countries for comparison
● Seventh selection of countries clears the list - clear one by one
● Highlight links between charts and maps when clicking on one country
●Sortable table with bars to show similarity in pattern and trend
Bibliography

Thank You!