

***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:

Country	Year	Internet_Users_per_100	Mobile_Subs_per_100	Tot_pop	Percent_rural	Percent_urban	GNI_per_cap	Median_Life_Exp	Primary_Compl_Rate	Per_Access_Electricity	Per_Adult_Unemployment	Dom_Language
Afghanistan	2008	1.840000	29.22037	23511400	77.32930	22.67070	370	58.22502	NA	42.40000	NA	Persian
Afghanistan	2009	3.550000	37.89494	23993500	77.04670	22.95330	470	58.60368	NA	47.88847	NA	Persian
Afghanistan	2010	4.000000	45.77817	24485600	76.76083	23.23937	520	58.97083	NA	42.70000	NA	Persian
Afghanistan	2011	5.000000	60.32632	24987700	76.47162	23.52838	570	59.32795	NA	61.51442	NA	Persian
Afghanistan	2012	5.454546	65.45219	25500100	76.17970	23.82030	720	59.67961	NA	69.10000	NA	Persian
Afghanistan	2013	5.900000	70.66136	26023100	75.88450	24.11550	730	60.02827	NA	75.15437	NA	Persian
Afghanistan	2014	7.000000	74.88284	26556754	75.58850	24.41350	670	60.37446	NA	89.50000	NA	Persian
Afghanistan	2015	8.280000	NA	NA	NA	NA	630	NA	NA	NA	NA	Persian

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]

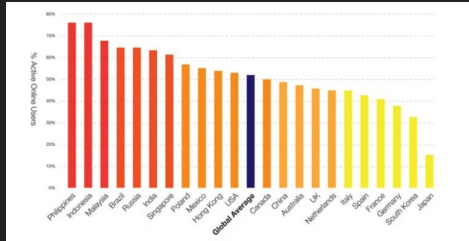
Target Audience:

Students, Researchers, Data Analysts in Government or Telecom, Policy Makers

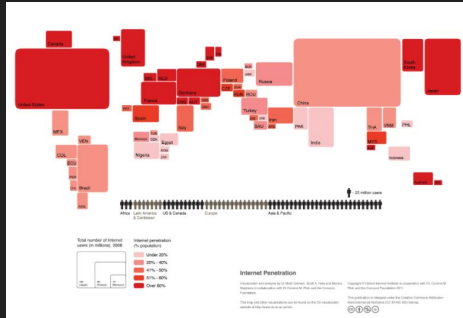
- Understanding trends across collection of time-varying tabular data
- Understanding relationships between variables
- Looking at the distribution of the variable across geographic regions
- Comparisons between countries and attributes
- Measuring the degree of the correlation between the main attribute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

Related Work

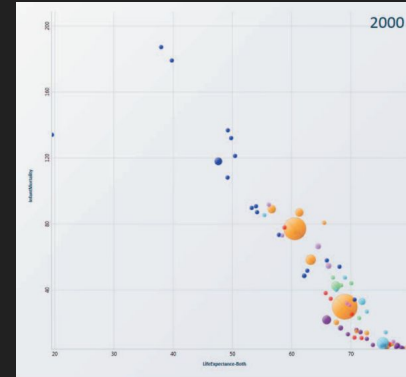
- Global Web Index: Global social network penetration [4]



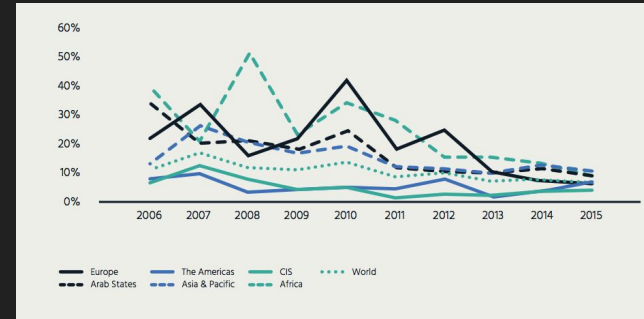
- Oxford Internet Institute: Internet penetration and population [2]



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

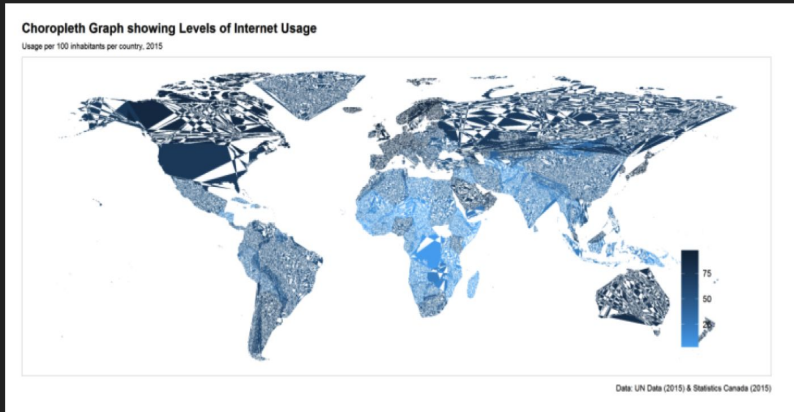


- International Telecommunication Union: Individual internet user growth rate [5]

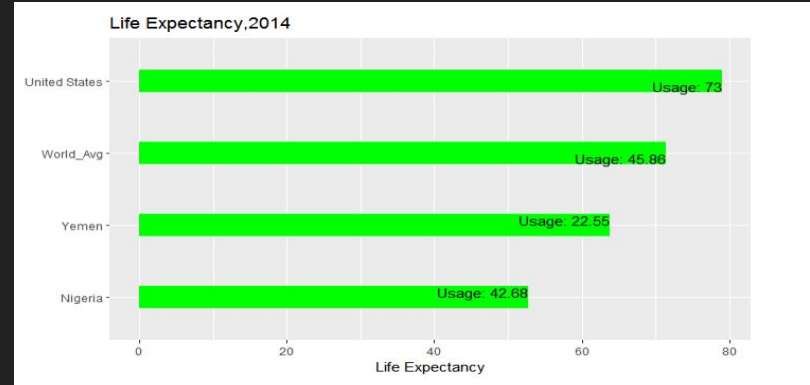


How: Initial Steps We Took

- Choropleth Map



- Bar charts

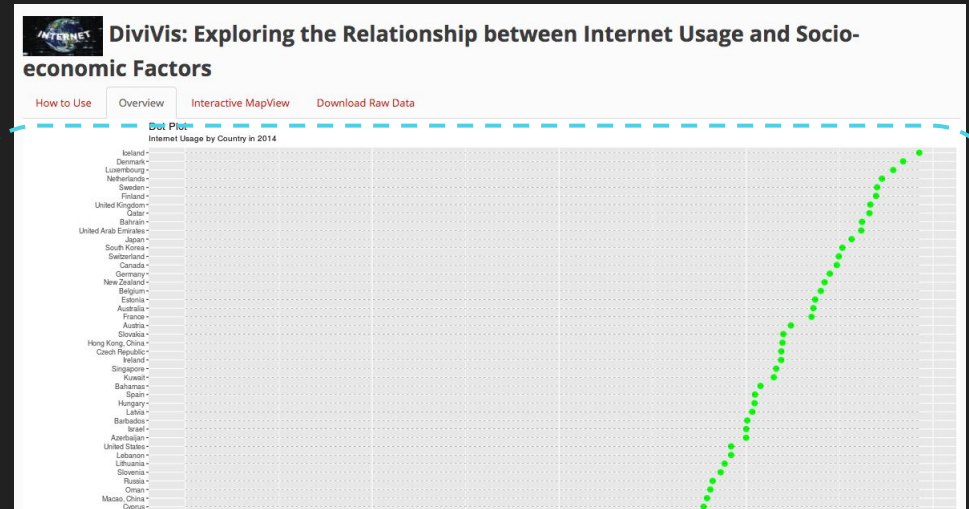


- Polar charts or radar graphs

Solutions - Tasks

Level 1:

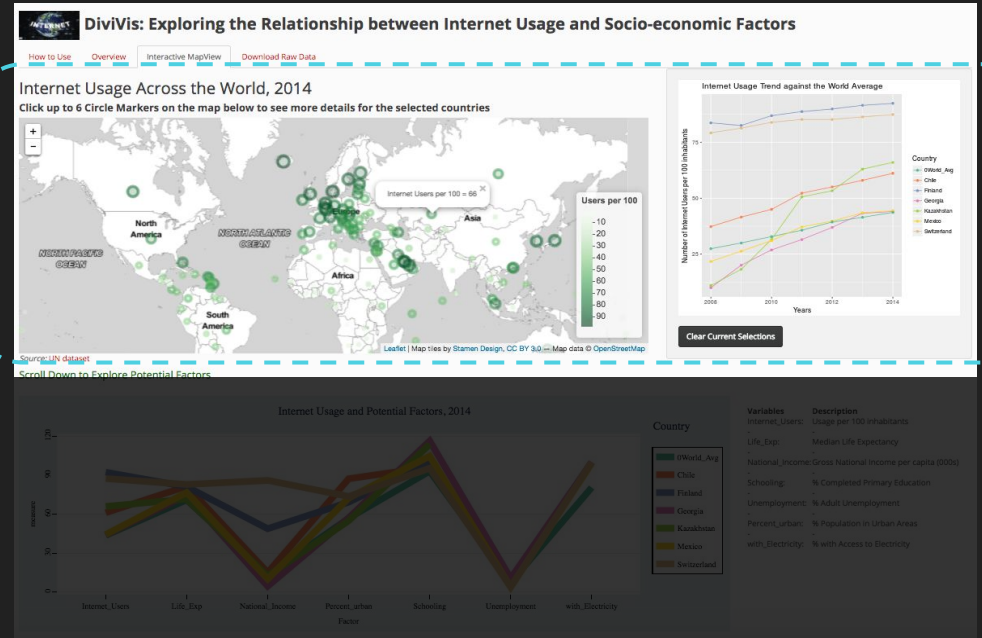
- Dot Plot: to show a general overview of the relative positions of all the countries



Solutions - Tasks

Level 2:

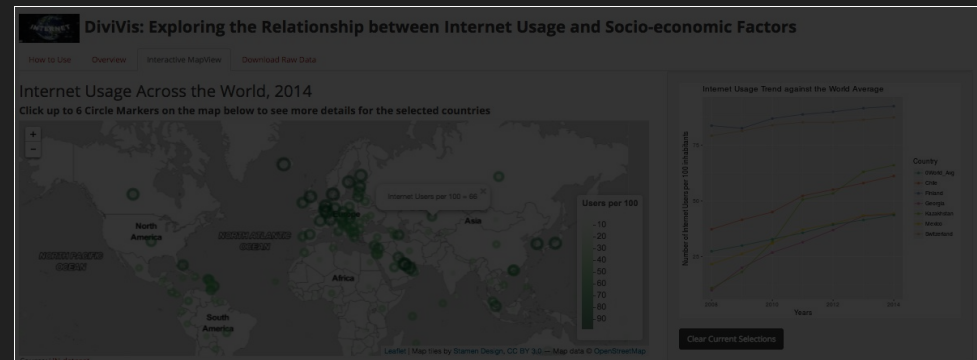
- Geo-Spatial Map visualization (bubble map): distribution, correlation; locate clusters, outliers
- Multiple line charts: to show and compare trends



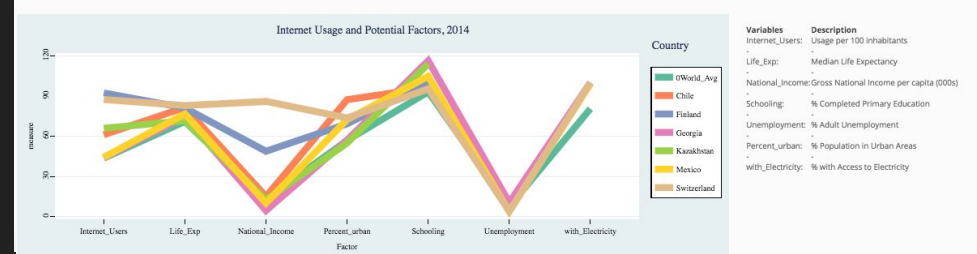
Solutions - Tasks

Level 3:

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



Scroll Down to Explore Potential Factors



Level 4:

- Linear Regression Line and Scatterplot: to find trends, outliers, distribution, correlation;
- Tables: to find detailed information

Linear Regression Analysis

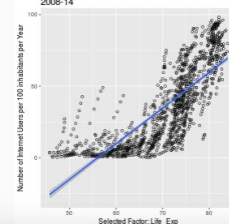
Coefficients of Internet Usage with all Other Factors:

term	estimate	std.error	statistic	p.value
(Intercept)	-108.15	19.17	-5.64	4.090e-08
Life_Exp	1.48	0.31	4.77	2.934e-06
National_Income	0.72	0.07	9.60	4.377e-19
Percent_urban	0.14	0.06	2.38	1.810e-02
Schooling	-0.03	0.12	-0.23	8.212e-01
Unemployment	0.67	0.19	3.47	6.070e-04
with_Electricity	0.24	0.09	2.56	1.105e-02

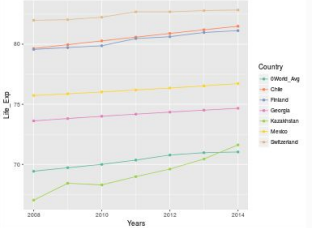
Potential Factors:

Median Life Expectancy

World Internet Usage vs Selected Factor, 2008-14



By Country Trends in Selected Factor: Life_Exp



Summary of Idioms

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

Implementation

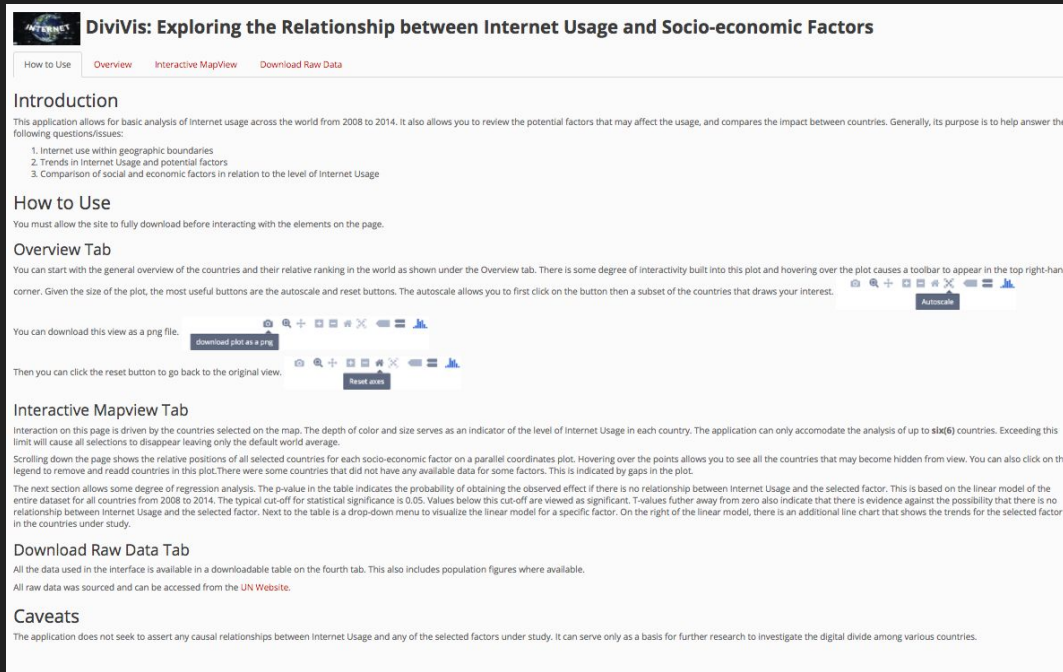
- ❖ User Interface - Shiny Web Application built on top of R
- ❖ Code Languages - R, HTML, CSS,, Javascript

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

Results - Demo

A live demonstration of the application can be viewed online:

<https://shirlett.shinyapps.io/worldinternetusage/>



DiviVis: Exploring the Relationship between Internet Usage and Socio-economic Factors

How to Use Overview Interactive MapView Download Raw Data

Introduction

This application allows for basic analysis of Internet usage across the world from 2008 to 2014. It also allows you to review the potential factors that may affect the usage, and compares the impact between countries. Generally, its purpose is to help answer the following questions/issues:

1. Internet use within geographic boundaries
2. Trends in Internet Usage and potential factors
3. Comparison of social and economic factors in relation to the level of Internet Usage

How to Use

You must allow the site to fully download before interacting with the elements on the page.

Overview Tab

You can start with the general overview of the countries and their relative ranking in the world as shown under the Overview tab. There is some degree of interactivity built into this plot and hovering over the plot causes a toolbar to appear in the top right-hand corner. Given the size of the plot, the most useful buttons are the autoscale and reset buttons. The autoscale allows you to first click on the button then a subset of the countries that draws your interest.

You can download this view as a png file.

Then you can click the reset button to go back to the original view.

Interactive Mapview Tab

Interaction on this page is driven by the countries selected on the map. The depth of color and size serves as an indicator of the level of Internet Usage in each country. The application can only accommodate the analysis of up to **six(6)** countries. Exceeding this limit will cause all selections to disappear leaving only the default world average.

Scrolling down the page shows the relative positions of all selected countries for each socio-economic factor on a parallel coordinates plot. Hovering over the points allows you to see all the countries that may become hidden from view. You can also click on the legend to remove and read countries in this plot. There were some countries that did not have any available data for some factors. This is indicated by gaps in the plot.

The next section allows some degree of regression analysis. The p-value in the table indicates the probability of obtaining the observed effect if there is no relationship between internet Usage and the selected factor. This is based on the linear model of the entire dataset for all countries from 2008 to 2014. The typical cut-off for statistical significance is 0.05. Values below this cut-off are viewed as significant. T-values further away from zero also indicate that there is evidence against the possibility that there is no relationship between Internet Usage and the selected factor. Next to the table is a drop-down menu to visualize the linear model for a specific factor. On the right of the linear model, there is an additional line chart that shows the trends for the selected factor in the countries under study.

Download Raw Data Tab

All the data used in the interface is available in a downloadable table on the fourth tab. This also includes population figures where available.

All raw data was sourced and can be accessed from the [UN Website](#).

Caveats

The application does not seek to assert any causal relationships between Internet Usage and any of the selected factors under study. It can serve only as a basis for further research to investigate the digital divide among various countries.

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

1. France, Belanger and Lemuria, Carter. The Effects of the Digital Divide on E-.government: An Empirical Evaluation; Virginia Polytechnic Institute and State University (2005)
2. Graham, Mark and Sabbato, Stephano. Information Geographies. Oxford Internet Institute. (2013)<http://geography.oii.ox.ac.uk/?page=home>
3. George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, and JohnStasko. “Effectiveness of Animation in Trend Visualization.” IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 08) 14:6 (2008), 1325–1332. (pages 142, 147)
4. How we use the Internet
<https://wearesocial.com/uk/blog/2011/06/world-map-global-social-media-usage>
5. Global Internet Report 2017. ICT Facts and Figures. 2017.
https://www.internetsociety.org/globalinternetreport/2016/wp-content/uploads/2016/11/SOC_GIR_2016-v1.pdf



Thank You!