

Rob Acheson

racheson@fas.harvard.edu

CS-171 Project 1 Proposal

2/14/2013

Title

Effects of Internet Connections on Education and Employment

Questions and Hypotheses

Referencing data from the World Bank API, this visualization will explore Internet usage statistics, educational benchmarks and the unemployment rates, globally and on an individual county basis. Through this exploration we hope to answer the following questions.

- What percentage of the population has access to the Internet?

- How many students complete primary education?

- What is the literacy rate?

- How many are employed?

- Are there correlations between Internet penetration and education, employment?

- Are there trends based on geographic regions?

Without having conducted any prior research, one would assume that an increase in Internet penetration would be directly proportional to the overall education level of the populous. Additionally, one could hypothesize that this increase in would cause a decrease in unemployment.

Motivation

First and foremost, the goal is to learn more about users of the Internet. As someone who's career is focused on development of mobile and web applications I am interested to learn more about the effect of these tools on the general population. After discovering the World Bank data API and realizing the breadth of information available, I see its great potential for use as a live data source in dynamic web applications.

Data Source and Process

The World Bank API will provide all of the data sets required for this exploratory visualization. Sourced from 214 countries and dating back to 1960, this data should reveal long-term trends in the areas of exploration. The API is a REST style service and can be queried with additional arguments to filter the data. Responses will be returned as JSON data. Using a python script to query the database, data will be parsed, and saved in CSV format. Data sets will be merged using Google Fusion Tables when needed, and sanitized with Google Refine.

Potential Visualizations

Visualizations will require much exploration before decisions can be made. For the majority, however, simple line graphs of percentage data on the y-axis vs. time on the x-axis. Some may be small-multiples showing individual county data in a grid. There is also potential for a multiple-country chart with one specific data charted to allow us to see larger trends and potential to color code groups by region. Below are some preliminary sketches of the charts.

