Ryan Enderby (rae83), Radhika Bhanu (rk539), Yufeng Wang (yw756)

Story

We take Internet access for granted. At worst, we find connection outages to be an inconvenience. However, Internet access is not so ubiquitous around the world. There exist stark differences between levels of access in different countries. Through this project, we looked to explore levels of Internet access around the world, as well as possible reasons why nations have experienced different rates of Internet adoption.

Our choropleth shows that across the world, Internet usage per 100 people is highly variable, even as recently as 2014. For the most part, areas of high adoption and areas of low adoption are clustered together. Western Europe has high adoption, the US and Canada has high adoption, all standing in contrast to Africa's almost entirely low adoption. Given the international differences illuminated by this map, we wanted to visualize the paths that different countries took to get to this point. In terms of related and potentially contributing factors, mobile phone usage and GDP per capita seemed like promising areas to explore.

GDP per capita is related in the sense that low GDP countries generally have low Internet penetration, while high GDP have high penetration, respectively. However, increases in Internet penetration in low GDP per capita nations outpaces the growth in GDP per capita, meaning that although related to existing levels of Internet usage, GDP per capita is not a factor limiting future growth in these nations.

Similarly, mobile phone usage has tracked Internet usage in low-penetration countries. Surprisingly enough, mobile phone usage is actually converging to a level comparable to that of the US, even as Internet penetration and GPD per capita do not converge. Libya is a case of mobile phone growth so explosive that we had to check several different sources just to make sure that our results were not some kind of error! Through our visualizations we can see that although commonplace in the US, Internet usage is not so across the world. Low GDP countries have been slower to adopt the Internet, although this factor is not preventing rapid growth, as both Internet usage and mobile phone adoption have been and continue to accelerate around the world.

Ryan Enderby (rae83), Radhika Bhanu (rk539), Yufeng Wang (yw756)

Data

Source

Much of our data comes from Knoema, an online public data platform.[1] The site provided a JSON dataset containing information on Internet usage per 100 people for each country, for each year since 1960. For GDP per capita and mobile phone subscriptions per 100 people, the World Bank allowed us to pull information dating back to 1981 for each country in terms of USD.[2]

Variables

We used the following variables in our dataset:

Country – variable storing the name of each country

Year – we used data corresponding to each year from 1990-2014. Each of the below variables was reported as a single figure for each year

Internet usage per 100 people – for each country, the number of people who used the Internet in any capacity via any type of device in the last 12 months

Mobile phone subscriptions per 100 people – for each country, the number of mobile subscriptions in any given year per 100 people

Annual GDP per capita in USD – the per capita GDP of each country reported in USD

Reformatting

The JSON dataset did not require reformatting, although we did have to manually update country names so that names in the map JSON matched all names in the Internet data JSON; for example, one listed "Hong Kong," while the other, "Hong Kong, SAR". However, the World Bank data (GDP, mobile phone subscriptions) could only be downloaded in CSV format, so we had to include a function for converting CSV to JSON in our code. We kept each sub-dataset in a unique file and then integrated in code by matching country names.

---

[1] http://knoema.com/WBINU2015/internet-users-1960-2014?regionId=World

[2] http://data.worldbank.org/indicator/NY.GDP.PCAP.CD

Additional data

Instead of the world map shape file presented in class, we used a map JSON found on Github.[3] In this file, countries are identified by name as opposed to numerical code, so we were easily able to map data to country just by using name.

Subset

While we have data for each country going back several decades, we decided to only use and present data beginning in 1990. While initially plotting and exploring our data, we noticed that across the world, there was virtually no penetration or growth prior to 1990. Internationally, 1990 to 2014 was the time of greatest change, and as such, we wanted our visualizations to communicate information about this era, and not get bogged down by long stretches of little change.

For the line plots, we selected a few showcase countries, as plotting the GDP per capita, mobile phone subscriptions and Internet access for every country would be highly impractical. As our countries of study, we chose the United States, India, China, Mexico and Libya. These countries fall at different locations along the spectrum of Internet adoption, and are always present in the news. Similarly, we wanted to have a representative country from different geographic regions. The USA and Mexico were interesting representative nations because they are in such close proximity, but have very different rates of Internet usage. India and China are in a similar situation. Africa as a whole has relatively low Internet usage, and Libya is an interesting country to study as it has tremendous adoption of mobile phone subscriptions.

Mapping to visual elements

Because our numbers were mostly presented in terms of per 100 people and because GDP tends to grow gradually, we did not need to apply any transformations to our data.

---

[3] https://raw.githubusercontent.com/johan/world.geo.json/master/countries.geo.json

For the choropleth, color and location communicate the usage statistics. Location naturally identifies the country onto which we are mapping Internet usage data, and the color gradient identifies the number of Internet users per 100. Lighter shades have relatively less usage than darker countries. For example, a country would be perfectly white if it had 0 out of 100 people using the Internet. We opted for a single color gradient, as the information being presented is a kind of density figure, making it intuitive that darker countries have more users out of 100 than lighter ones.

Originally, we experimented with a scale that put high usage countries in the green spectrum and low usage in red, but we found that this carried unintended connotations. For example, a country with 50 out of 100 people using the Internet would appear brown, meaning that we had red, brown and green nations on the map. This made it seem as though there were three different groups or types of countries, while we were really just trying to communicate where each country fell along a spectrum in terms of a single variable. As such, we opted to display our info as a single color gradient.

For the line plots, each country was identified by a unique color. In each of the three graphs, a country was always kept the same color to facilitate easy comparison between countries. Each line graph featured lines of different style (e.g. solid, dashed, dotted) to communicate the fact that different information was being conveyed in each graph. The coupling of same country color but different line style helps the user understand that they are looking at different information being presented about the same countries.