# Table of Contents

Introduction / Goals ..................................................................................................... 3

Wealth & Water .............................................................................................................. 4
  Discussion .................................................................................................................. 5
  Comparisons .............................................................................................................. 6

Dirty Water means Dead Babies .............................................................................. 8
  Discussion .................................................................................................................. 5

The Flip Side ................................................................................................................. 11

Application Critique .................................................................................................. 13

Suggestions for Improvements .............................................................................. 15

Eight Golden Rules ..................................................................................................... 16

Bugs Found ................................................................................................................... 17

Authors .................................................................................................................................................................... 18
Introduction

*The gauntlet was thrown down. And we accepted.*

The [World Water Day Visualization Challenge](#) is an Information Visualization contest sponsored by GE and Seed Media Group set to coincide with [World Water Day](#) on March 22nd, 2011. It's a rapid-fire competition to create visualizations to explore data sets that highlight global issues and effects related to water. Data for visualization is provided through Google Fusion Tables and Google Public Data Explorer. See the Application Discussion below.

Goals

We wanted to take a world-wide look at the relationship between clean water and the health and wealth of nations, **looking specifically at regional differences**. We believe that looking at data along, among others, a regional axis promotes a more pragmatic analysis geared to help high-level public policy. Focusing on a specific country may promote a narrower vision and ignore broader problems. Of course, it's important to keep outliers in mind.
Wealth & Water: The World's Water is Getting Better

A History of Improved Water Access by GDP and Region in the Last Two Decades

A Five Axis Presentation:
1. X: Improved Water Source (% of Population With Access)
2. Y: GDP, Per Capita, US $ (logarithmic scale)
3. Bubble Size: Population, Total
4. Color: Region (discreet)
5. Time: Year (1990-2008)

View Live Chart
Discussion

Five axes is a lot! As a 2-D scatter-plot, we see the positive correlation between GDP per capita and Improved Water Sources. This probably doesn't surprise anyone, but upon further study it's certainly interesting that there aren't even many outliers. Looking at 2008 (the latest year), we see that Angola is the one example of a country with high GDP but bad water supply. Burundi is on the other side with good water but a relatively low GDP. Keep in mind that these are per-capita values: it's interesting that countries with great total wealth don't break the curve! The GDP is of course on a log scale, which presents a much more important view of that data.

The size of the bubbles is of course the total population. You can see China, India, and Indonesia somewhere in the middle of the pack, with Brazil faring slightly better, and the United States close to the top. Note that these countries alone comprise nearly half of the world's population.

The color signifies the region of the world, as shown by the legend. Note that Sub-Saharan Africa is segregated from North Africa. You can mouse over the legend to see the clustering of the countries: North America is of course at the top, and Europe fares about as well, except for a trail of former Soviet Bloc countries. Tajikistan, Azerbaijan, Uzbekistan, the Kyrgyz Republic, and others are woefully behind. South Asia is in the middle except for Afghanistan, which is one of the worst water countries in the world, as you'll soon see. North Africa is clustered somewhere in the top half except for Yemen, and some countries are better than Europe: Kuwait, Qatar, and UAE are stellar. Egypt is an interesting outlier with low GDP but great water, almost certainly owing to the Nile Valley. Latin America is also in the middle, but with one tragic outlier: Haiti. Now we get to the bottom of the chart: Sub-Saharan Africa. Astonishingly low access to fresh water coupled with pervasively low GDP has wreaked havoc on the subcontinent. Some of the most populous countries are the worst: Congo, Ethiopia, and Nigeria. We'll see even more views into the area soon.

Now let's go to the video tape! Playing the animation, we see an uneven but steady progress along the curve for almost every country! This is incredibly promising; hopefully in a few decades nearly every country will have adequate access to fresh water.
Data Notes
As you can see, we have data going back to 1990, which is limited by the Water information available on the countries. Information like GDP and Population go back much further.

More Information / Data Sources:

Access to an improved water source refers to the percentage of the population with reasonable access to an adequate amount of water from an improved source, such as a household connection, public standpipe, borehole, protected well or spring, and rainwater collection. Unimproved sources include vendors, tanker trucks, and unprotected wells and springs. Reasonable access is defined as the availability of at least 20 liters a person a day from a source within one kilometer of the dwelling. Data courtesy The World Bank, http://data.worldbank.org/indicator/SH.H2O.SAFE.ZS?cid=GPD_24


GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars. Data courtesy The World Bank, http://data.worldbank.org/indicator/NY.GDP.PCAP.CD?cid=GPD_57

Source World Bank national accounts data, and OECD National Accounts data files.

Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship—except for refugees not permanently settled in the country of asylum, who are generally considered part of the population of their country of origin. The values shown are midyear estimates. Data courtesy The World Bank, http://data.worldbank.org/indicator/SP.POP.TOTL?cid=GPD_1

Dirty Water means Dead Babies

Regional Infant Mortality dictated by Water and Wealth

A Four Axis Presentation

1. X: Improved Water Source (% of Population With Access)
2. Y: Infant Mortality Rate (per 1,000 live births) (logarithmic scale)
3. Bubble Size: GNI Per Capita, PPP (Current International $)
4. Color: Region (discreet)
5. Time: Year (1990-2008), ignored. (See Discussion)

View Live Chart
Discussion

Now we turn our attention from Wealth to Death. Infant Mortality rates are well-documented throughout the world, and are one of the primary indicators of public health. Looking at the 2-D scatter-plot, we see a very strong correlation between the percentage of people with access to improved water and the Infant Mortality Rate. For the purposes of our analysis here, we will ignore the time dimension, though you may take a look if interested.

Unlike the last chart, here we've made the size of the circle correspond with the per-capita GDP. This results in a beautiful effect where the larger, "heavier" bubbles appear to "sink" towards lower mortality!

You can see a third-party visualization of Infant Mortality Rates here:

More Information

Infant mortality rate is the number of infants dying before reaching one year of age, per 1,000 live births in a given year. Data courtesy The World Bank, http://data.worldbank.org/indicator/SP.DYN.IMRT.IN?cid=GPD_55


GNI per capita based on purchasing power parity (PPP). PPP GNI is gross national income (GNI) converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GNI as a U.S. dollar has in the United States. GNI is the sum of value added by all resident producers plus any product taxes (less subsidies) not included in the valuation of output plus net receipts of primary income (compensation of employees and property income) from abroad. Data are in current international dollars. Data courtesy The World Bank, http://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD?cid=GPD_8

Source World Bank, International Comparison Program database.
The Flip Side: Internal Renewable Water Wanes

Slow decline of internal freshwater may have future consequences.

A Simple Two Axis Presentation
1. X: Year
2. Y: Cubic meters of renewable freshwater, per capita

Discussion
This chart was chosen to simply highlight that not everything is slowly improving: Internal renewable freshwater is a measure of the available rivers and rainfall in a particular country. In this case we’re looking at the entire world, so the “internal” qualifier doesn’t apply.

Although this data only shows from 2002 on, the trend is unmistakable. It’s possible that Global Warming is having an alarming effect on total water. There’s a fantastic paper about “Peak Water” by Palaniappan and Gleick at [http://www.worldwater.org/data20082009/ch01.pdf](http://www.worldwater.org/data20082009/ch01.pdf)
More Information
Renewable internal freshwater resources flows refer to internal renewable resources (internal river flows and groundwater from rainfall) in the country. Renewable internal freshwater resources per capita are calculated using the World Bank's population estimates.

Source Food and Agriculture Organization, AQUASTAT data.
Application Critique:  
Google Public Data Explorer

"Data Visualizations for a Changing World" -- Google

Google Public Data Explorer (PDE) is designed as a visualization tool to explore large data sets, focusing on public data in categories like health, economy, environment, global issues, people, and geography. The visualizations (specifically the Bubble Charts) are based on the same technology behind GapMinder World called "Trendalyzer" and share many of its features and stylings. It will be interesting to see this tool evolve separately from GapMinder as Google puts its own work into it.

Google Fusion Tables is a data-warehousing and sharing site with several very simple visualization tools. It is intended more for dissemination of raw data, which can be exported in various formats.

Google Public Data Explorer:  
High-Level Critique

Benefits

Powerful Multi-Dimensional Visualizations
Like GapMinder, the Bubble Charts provide five dimensions of visualization, leaps and bounds more useful than simple bar or line graphs. It takes advantage of 2-D position, color, size, and animation. All this without using confusing 3-D trickery.

Time-dimensional Animations
An incredibly powerful tool to view large-scale changes, PDE provides smooth animations of data points changing over time. It also provides "trails" for selected data points which provide even more insight into specific comparisons.
Easy To Use
PDE is incredibly easy to use compared to tools with a "kitchen-sink" approach such as Tableau and Spotfire. Tableau and Spotfire certainly provide more power and flexibility, but at the cost of a steep learning curve.

Free
Tableau and Spotfire are oriented to business users and cost an incredible sum of money ($2000 for Tableau Desktop). PDE is instead created to help make the world a better place; something anyone can use to create visualizations for any purpose.

Data Liberation
Unlike with Gapminder, anyone can create and upload datasets to PDE in a specific format. These datasets can then be shared with the world for anyone to use.

Powerful Data Format
Dataset Publishing Language (DSPL), which is the only format accepted for upload, has powerful features like language internationalization, geocoding, linkable datasets, and concept relationships.

Accessible
Unlike with Tableau or Spotfire, this tool works from ANY computer without installing anything! (Except Flash)

Drawbacks

Limited charting options
Most of the functionality and "wow" possibilities are concentrated in the Bubble Charting tool. While PDE also provides standard charts like Line and Bar graphs, they provide fewer axes and insight.

Can't use Raw Comma-Separated Values
PDE can only accept data in a specific (but open) format called Dataset Publishing Language (DSPL), which makes use of XML. In its defense, you can also take an existing .csv file and add XML annotations to it.
Suggestions for Improvements

"Computers are good at following instructions, but not at reading your mind."

-- Donald Knuth

High-Level
Choice of Services, Intended Users, etc.

- It may be prudent to expand the intended audience to include business interests, akin to Tableau or Spotfire. Powerful free visualization tools from Google would probably gain rapid acceptance for various business use cases.
- The map chart interface (perhaps the system as a whole) should allow a spatial/time dynamics, such as the spatial spread of a disease or migration phenomena.
- Although PDE generalizes well to sub-country areas, it doesn’t have any specific features like Gapminder lab, who already has a US state version of the bubble/map charts: http://www.gapminder.org/labs/gapminder-usa/

Mid-Level
Use of Color, Interaction Techniques, etc.

- An Export/Save Chart feature that would create an image or PDF.
- It would be interesting if the user could zoom in/out (that is, allow changes in axis resolution) to better visualize cluttered parts of the bubble chart interface. Also desirable in bar and line mode. Note that this is allowed in map chart mode.
- An interesting feature should be allow adding text (or even other media) that is connected to a country and year, to point out outstanding behavior. When run, this information would pop up in front of the user.
- The line chart interface should allow more than one curve to permit inspecting different data sources or trends simultaneously.

Low-Level
Terminology, Label Placement, Color, Fonts, etc.

- The user should have the ability to relabel axes.
- It would be great to have the option to specify colors or color schemes.
- A "Select All" button as counterpoint to "Clear selections" on the left is also desirable.
- The user should have more options to scale axes instead of just "log" or "linear".
• The system should allow you a custom year range or granularity. For example, the user might want to see biennial trends in between 2000 and 2008.
• The system should allow you to control the speed of motion in bubble chart mode. This is a feature of GapMinder that, for some reason, was not kept in Google PDE.
• Years that have interpolated data should be signalized somehow by the system, since sometimes it is difficult to see the little stars in the chart.

Eight Golden Rules of Interface Design

From "Designing the User Interface" by Ben Shneiderman

Consistency
Color schemes are loaded at random every time the charts are loaded, which is bad. Apart from that, the usage is pretty much the same, and involves interaction with the axis and bubbles.

Shortcuts
The operations are not very complex, so there is small need for shortcuts. However, it would be interesting for those who are often concerned with the same group of countries to be able to have a shortcut for selecting those (something like storing it as a preference setting).

Feedback
Overall, the system presents prompt responses to user interactions (e.g. changes in bubble colors, sizes, and position according to new choices of data and smooth axis scale changes).

Closure
Groups of actions are relatively simple and the sequences are intuitive. Perhaps the most complex group of interactions is axis/color/size changes, in which the user has to choose the corresponding data in a hierarchical list on the left and then press OK/Cancel on the right.

Error-Handling
It is very difficult to make a mistake, since most of the options come from drop boxes.

Reversion
The system completely lacks UNDO or REDO options.

Control
Users are active most of the time, even when the plot is evolving in chart mode, since they can select/deselect groups, fetch information by mouse-over, and so on.
Memory Load
The user can choose among the three different visualizations by opening tabs that have self-explained icons and points of interactions are so intuitive that I doubt someone would ever need to memorize how to interact with the displays and set options.

Bugs Found

"A documented bug is not a bug; it is a feature." --James P. MacLennan

- The colors used to fill the bubbles will most likely change with every reload of the same chart. In fact, it looks like it loads a color scheme randomly every time the chat is accessed.
- Interpolated values appear to move inconsistently.
- Labeled pointers to data points (country names) are not re-positioned to avoid overlap whenever possible.
- In bubble chart mode, if one selects or deselects a country while the bubbles are in motion, the system moves back one year, moves further to the present year, and then stops the action.
Authors

Leonardo Claudino <claudino@cs.umd.edu>

Nishant Patel <Nishant.s.patel@gmail.com>

Graduate Students, University of Maryland - College Park

Department of Computer Science