Dominic Delmolino Project 2 – Final Presentation

Sarah Kelley INFO W18 – Python Bridge 2

Andrea Pope

Introduction:

This paper summarizes our analysis of data provided by the World Health Organization (WHO). Our analysis explored life expectancy for various countries over different time periods and cross referenced with various factors to determine any possible correlations with life expectancy. Those factors include environmental factors, HIV prevalence, how health services are provided (government vs private) and aid provided to countries.

Methodology:

Data sets were downloaded from the World Health Organization (“WHO”) website. The core set of data for our analysis was the Life Expectancy data set. We concentrated on three (3) key statistics, specifically:

* Life Expectancy at Birth
* Life Expectancy at Age 60
* Healthy Life Expectancy (HALE) at birth

The data was provided for the years 1990, 2000, 2012, and 2013 for 194 different countries. The HALE data set takes into account degraded health due to disease or injury.

Against the Life Expectancy data, we compared 4 different factors that could impact Life Expectancy, specifically environmental factors (access to clean drinking water and sanitation capabilities), HIV prevalence, how health services are provided (government vs private) and aid provided to each country. In each case, we have conducted investigative analysis of the individual data set, and then compared it to the Life Expectancy data set.

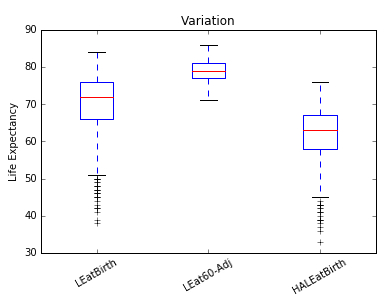
For all of the data sets, initial data cleaning was completed to structure the data such that is allowed easier analysis, and to rid the data of any superfluous characters. Examples of this include separating a “Country;Date” field into 2 separate fields, simplifying column names, managing null values, and dropping years of data without any corresponding datasets.

For a complete list of all columns included, please see the Appendix included at the end of this document.

Result:

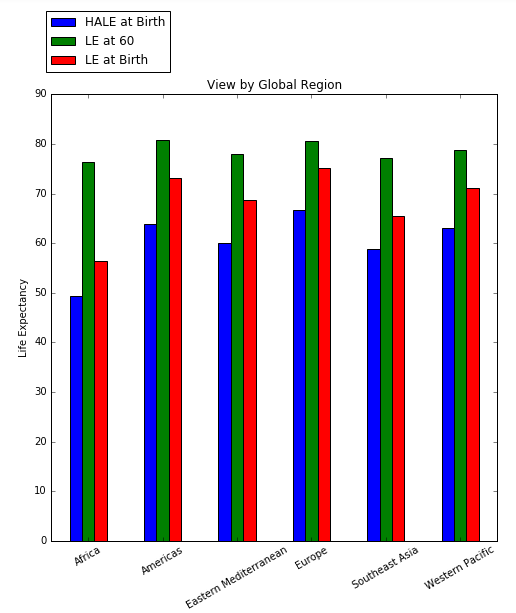
##### Life Expectancy:

The boxplot below shows the variation both between the various Life Expectancy measures, but also the variation of the measures across the dataset of 194 countries. As anticipated, life expectancy at age 60 has a higher mean, as the starting point takes into account they have reached age 60. Similarly, when you take into account a Healthy Lifestyle (via the HALE measure), the mean and upper quartile drop.

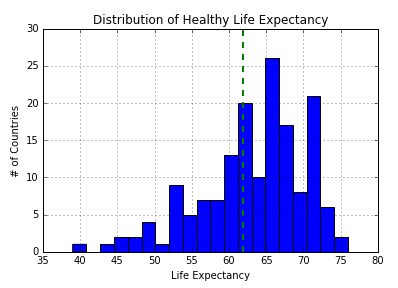
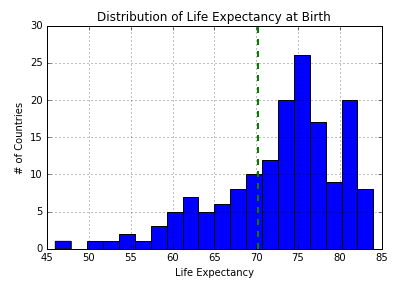


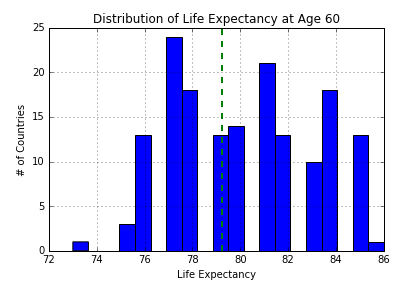
##### Further delving into these statistics, the chart below maps each of these statistics for each region of the world. Africa has the lowest Life Expectancy at Birth and Healthy Life Expectancy, but if once at the age of 60 in Africa, your life expectancy is not significantly different than other regions of the world.

As anticipated, Europe and the Americas share honors for the longest life expectancy.

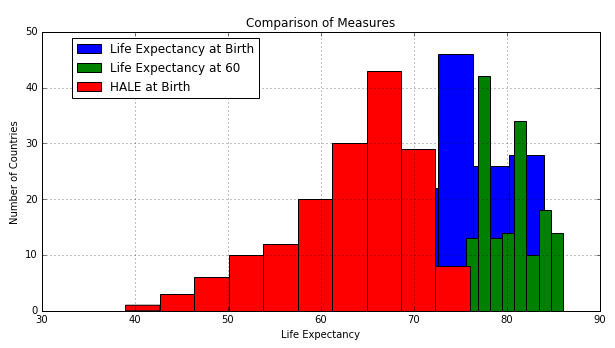


##### The next 3 charts show the distribution of these 3 measures, including an indicator of the mean (green line) across all of the countries.

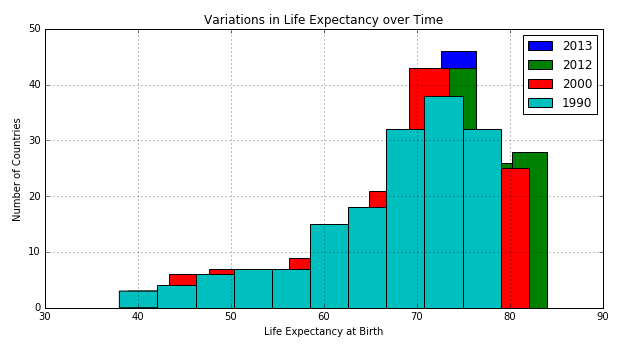
 



Combined together, the charts follow our original assumptions. Healthy life expectancy trends lower than the life expectancy measures, and once the age of 60 is reached, the number that reach over 80 is significantly high.

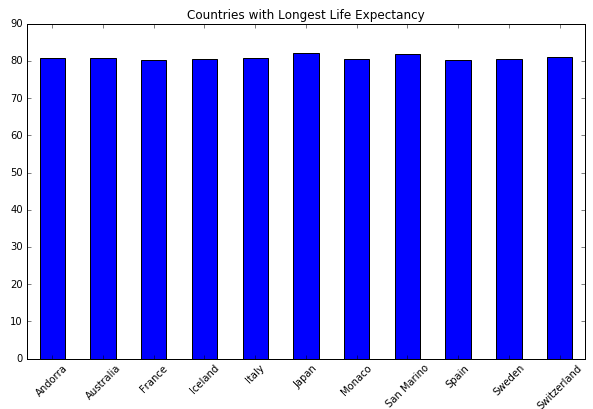


For reference purposes, the chart below shows Life Expectancy at Birth for each year of data provided.



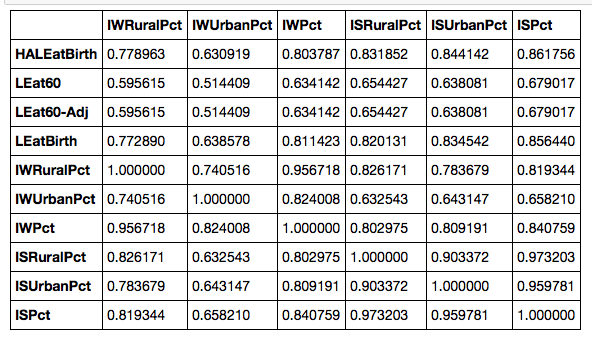
Diving into country specific details for Life Expectancy at Birth (for 2013), there are only 2 countries with life expectancies under 50: Zambia and Sierra Leone.

However, there are eleven countries with a life expectancy (at Birth) over 80.



##### Environmental Factors (Hygiene):

Correlations of Life Expectancy against Environmental Hygiene factors and cross correlation between Hygiene factors:



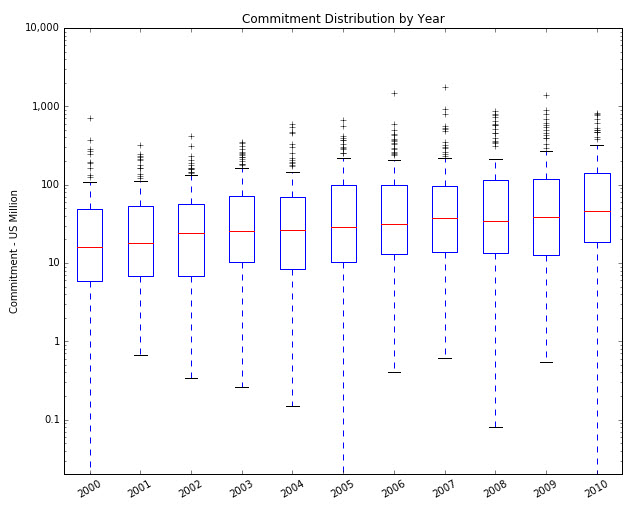
While access to sanitation facilities had the highest correlation to Life Expectancies (and among Life Expectancies, most highly correlated with “At Birth” expectancies), both rural ***and***urban sanitation access is most highly correlated with rural access to clean drinking water. In fact, even urban sanitation access showed higher correlation with rural access to clean drinking water than to urban access to clean drinking water.

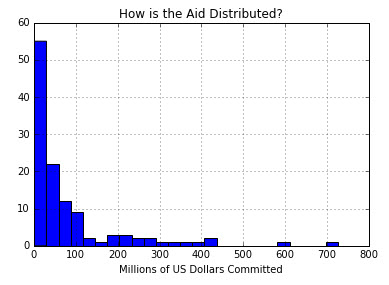
Comparing rural access to clean drinking water with urban access to clean drinking water, across all life expectancies the rural access was more highly correlated than the urban access.

Given rural access to clean drinking water having such high correlations to sanitation and life expectancy, it appears that efforts to improve rural access to clean drinking water have the largest environmental / hygiene beneficial impact on life expectancy.

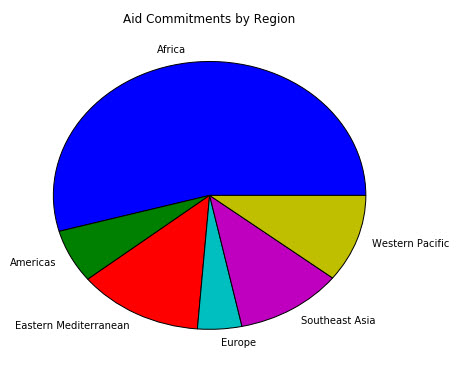
##### AID:

Analysis of aid committed and disbursed to 119 different countries between the years 2000 and 2010 shows a fairly consistent distribution, with the total increase over time. Please note the y-axis below is logarithmic.

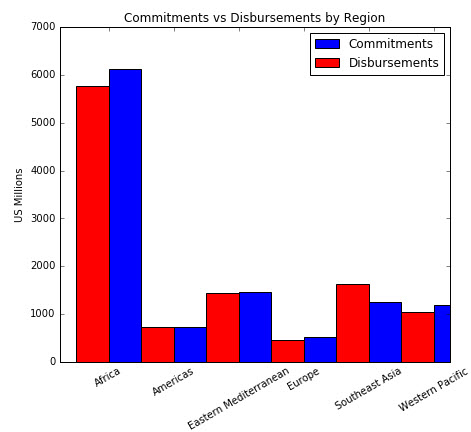


Most countries that receive aid, receive under $25 Million, as shown in this histogram. 

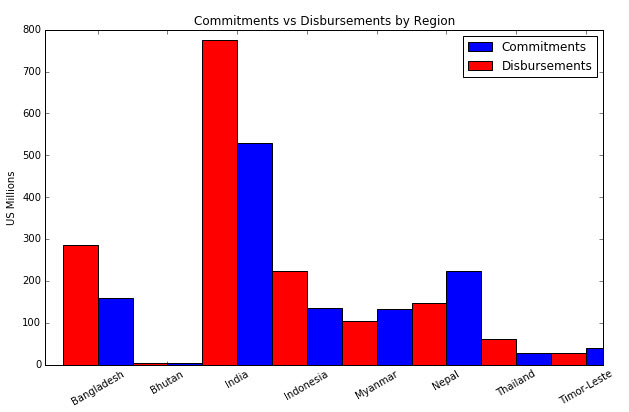
As anticipated, countries in Africa represent the largest recipients, as illustrated below:



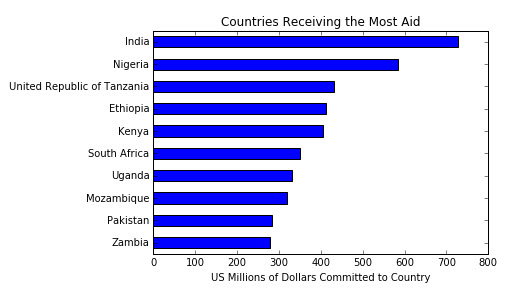
A quick review of commitments vs disbursements, for 2010 (latest year with available data) for each region, shows they are fairly consistent.



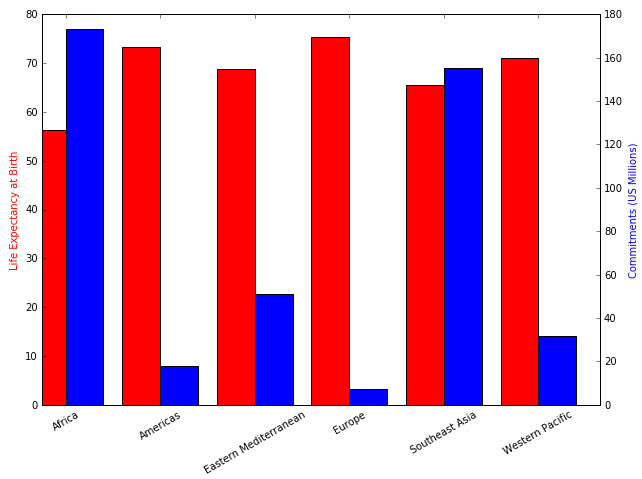
The biggest variation, by percentage, is in Southeast Asia. Below shows that broken by country. India has the greatest disparity between committed and disbursed.



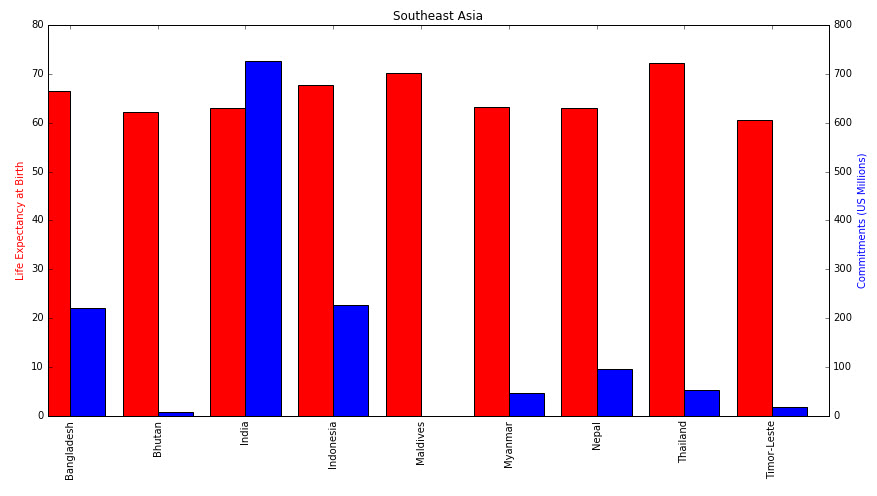
India, in fact, has the most amount of aid committed, as illustrated below in the chart showing the top aid recipients; those receiving on average over $250 million per year.

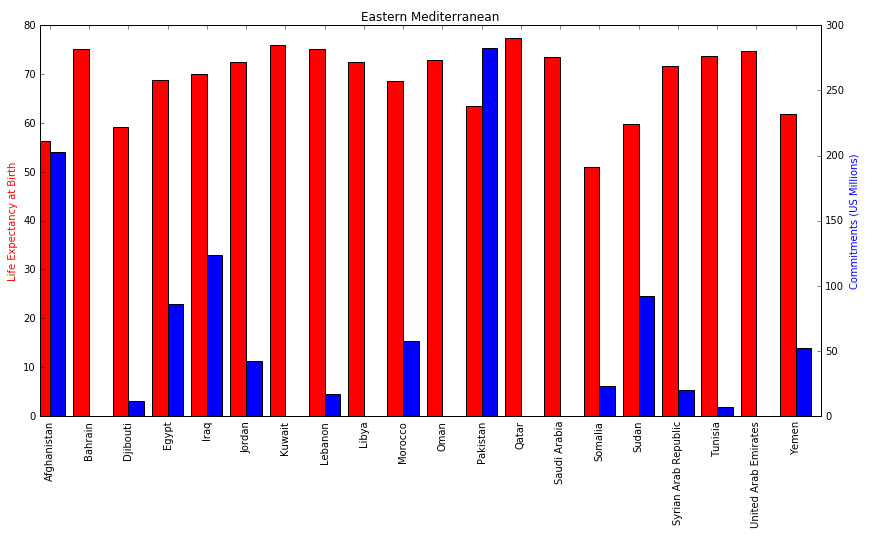


There is not a strong positive or negative correlation between life expectancy within a country, and the amount of aid received, with the exception of each end of the life expectancy range. The region with the highest life expectancy receives the least aid, and the region with the lowest life expectancy receives the most aid. This negative correlation was anticipated. Please note the dual y-axis on the chart below.



One surprising result was Southeast Asia, given their fairly high life expectancy but also very large aid commitments, especially in comparison to the Eastern Mediterranean region, that has a fairly similar life expectancy, but receives much less aid. Below is a breakout by country for each of these regions.





Again we see India with a very high aid commitment, and only one Southeast Asia country that receives no aid. Separately, in the Eastern Mediterranean region, a smaller percentage of countries within the region are receiving aid, however, for those that are, the life expectancies are lower, and more in line with Southeast Asia. This again supports the negative correlation between the amount of aid committed and Life Expectancy.

Conclusion:

Sources:

Aid: <http://apps.who.int/gho/data/node.main.A1626?lang=en>

Environmental Factors: <http://apps.who.int/gho/data/node.main.167?lang=en>

Healthcare Services: <http://apps.who.int/gho/data/node.main.75?lang=en>

HIV: <http://apps.who.int/gho/data/node.main.622?lang=en>

Life Expectancy: <http://apps.who.int/gho/data/node.main.3?lang=en>

*Source files for Region/Country Mapping:*

Africa:<http://www.afro.who.int/en/countries.html>

Americas: <http://www.paho.org/hq/index.php?option=com_wrapper&view=wrapper&Itemid=2005&lang=en>

Eastern Mediterranean: <http://www.emro.who.int/countries.html>

Europe: <http://www.euro.who.int/en/countries>

Southeast Asia: <http://www.searo.who.int/countries/en/>

Western Pacific: <http://www.wpro.who.int/countries/en/>

Appendix:

List of tables here…

##### 

##### 

##### 

##### FROM ORIGINAL WRITE-UP

##### Environmental Factors (Hygiene):

***Overview:***  Access to improved drinking water and improved sanitation facilities has data for 192 countries and 3 sample years (1990,2000,2015) for Rural, Urban and Total populations as a percentage.

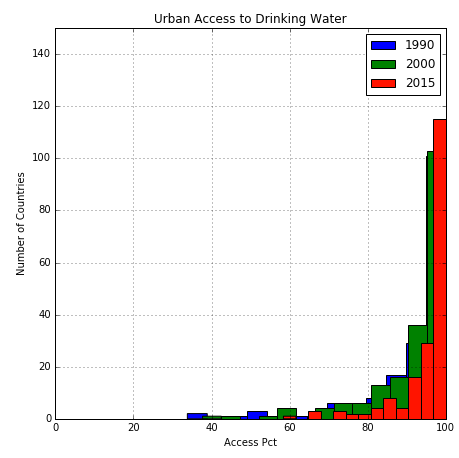
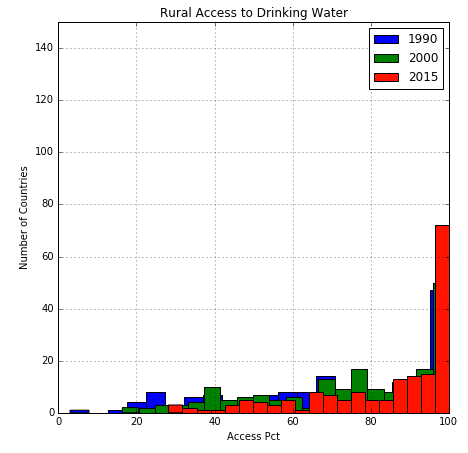
***Source:*** http://apps.who.int/gho/data/node.main.167?lang=en

***Total records:*** 558 ***Total Columns:*** 7

***Planned Analysis:*** For analysis in hygiene and sanitary conditions, we will check if life expectancy of the total population has any significant correlation with sanitary access for the different populations. To do this, we will look at Country, Population (Rural, Urban, Total) and data from 1990, 2000 and 2015 (as it matches up to the Life Expectancy in the years 1990, 2000 and 2013). Our initial hypothesis is that countries with higher percentage of access will have higher life expectancy. Within the file itself, we will compare rural vs urban improvement over time.

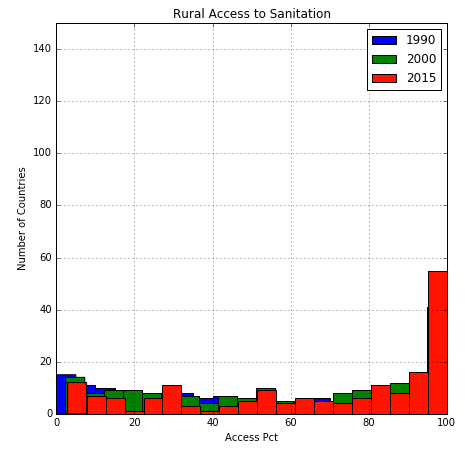
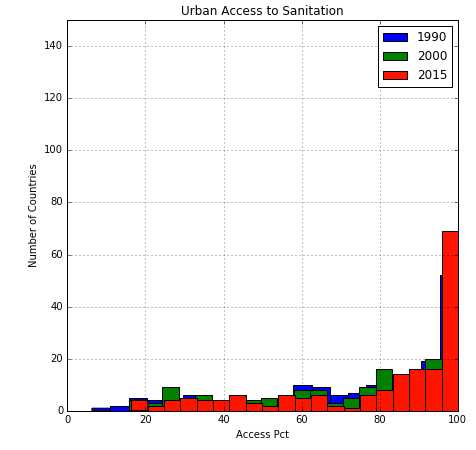
***Initial Analysis***

The initial data set was fairly clean, and only required splitting up the Country;Year column into separate columns. We began our analysis looking at how rapid the changes were for increases in access to drinking water for both the rural and urban populations across all countries. Increases in rural access to drinking water showed significant improvement from 1990 to 2015, moving from having some countries to less than 30% access in 1990 to all countries being better than that in 2015.

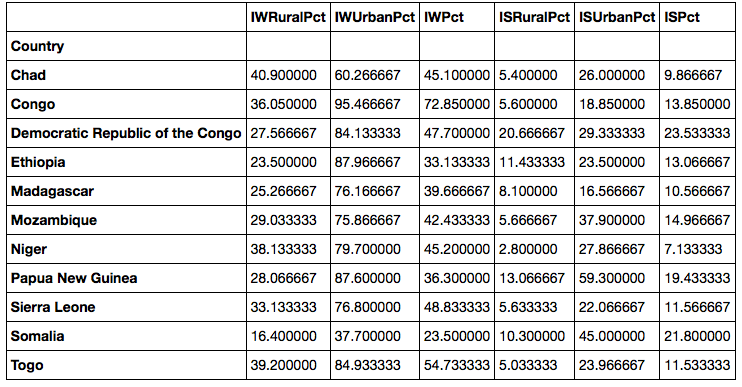


Urban access also improved, but mostly at the high end, moving from many countries with between 80 and 90% access in 1990 to most countries going to greater than 95% access in 2015.

Sanitation access did not show as much movement as drinking water access.

All of the countries with the lowest access to clean drinking water are in Africa:



#### Columns for each data source:

Life Expectancy Columns:

Country; Year

Life expectancy at birth (years); Both sexes

Life expectancy at birth (years); Female

Life expectancy at birth (years); Male

Life expectancy at age 60 (years); Both sexes

Life expectancy at age 60 (years); Female

Life expectancy at age 60 (years); Male

Healthy life expectancy (HALE) at birth (years); Both sexes

Healthy life expectancy (HALE) at birth (years); Female

Healthy life expectancy (HALE) at birth (years); Male

*Initial manipulation: The first column naturally splits into 2 separate columns, one for Country and Year.*

Environmental Factors (Hygiene) Columns:

Country; Year

Access to Clean Drinking Water (rural)

Access to Clean Drinking Water (urban)

Access to Clean Drinking Water (total)

Access to Improved Sanitation Facilities (rural)

Access to Improved Sanitation Facilities (urban)

Access to Improved Sanitation Facilities (total)

*Initial manipulation: The first column naturally splits into 2 separate columns, one for Country and Year.*

HIV Columns:

Country

HIV prevalence 2001

HIV prevalence 2005

HIV prevalence 2009

HIV prevalence 2013

Healthcare Services (applicable) Columns:

Country

Total expenditure on health as a percentage of gross domestic product 2013

General government expenditure on health as a percentage of total expenditure on health 2013

Private expenditure on health as a percentage of total expenditure on health 2013

Out-of-pocket expenditure as a percentage of total expenditure on health 2013

Total expenditure on health as a percentage of gross domestic product 2012

General government expenditure on health as a percentage of total expenditure on health 2012

Private expenditure on health as a percentage of total expenditure on health 2012

Out-of-pocket expenditure as a percentage of total expenditure on health 2012

Total expenditure on health as a percentage of gross domestic product 2000

General government expenditure on health as a percentage of total expenditure on health 2000

Private expenditure on health as a percentage of total expenditure on health 2000

Out-of-pocket expenditure as a percentage of total expenditure on health 2000

Aid – Commitments and Disbursement Columns:

Country; Year

Commitments to recipient countries (Million, constant 2009 US$

Disbursements to recipient countries (Million, constant 2009 US$

*Initial manipulation: The first column naturally splits into 2 separate columns, one for Country and Year.*

Data for Region/Country Maps were pulled from specific web pages. Please see lab notebooks for details.