Dominic Delmolino Project 2 – Final Presentation

Sarah Kelley INFO W18 – Python Bridge 2

Andrea Pope

Abstract:

<<summary of paper>>

Editing online test....

Cool – I can see your edits!

Introduction:

<<rationale behind project – introduce what we are trying to do

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.

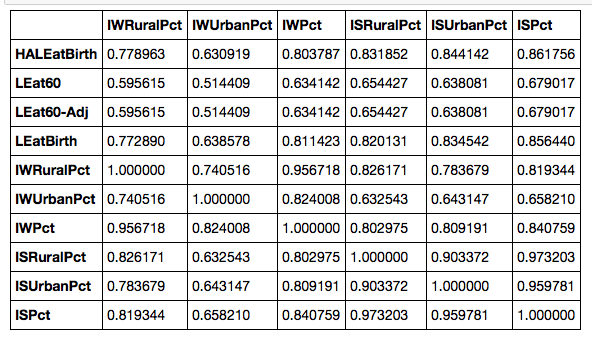
For all of the data sets, initial data cleaning was completed to structure the data such that is allowed easier analysis, and to rid the data of any superfluous characters. Examples of this include separating a “Country;Date” field into 2 separate fields, simplifying column names, managing null values, 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:

##### 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.

Discussion:

Our project

Conclusion:

Sources:

Appendix:

List of tables here…

Data:

Each section below describes the data sets included in our project. The appendix includes specific column names from each data set.

##### Life Expectancy:

***Overview:*** The life expectancy data set looks at life expectancy at birth and at age 60, for both males and females in 194 countries, for the years 1999, 2000, 2012, and 2013. The data set also includes Healthy Life Expectancy (HALE) at birth. This measure takes into account degraded health due to disease or injury. HALE numbers are only available for 2000, 2012, and 2013.

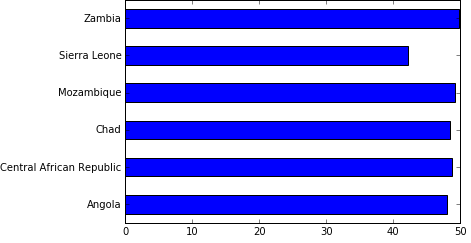
***Data source:***<http://apps.who.int/gho/data/node.main.3?lang=en>

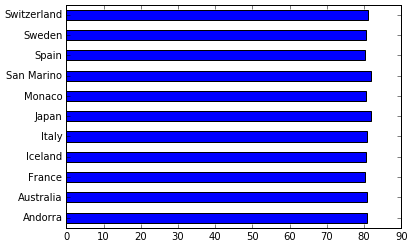
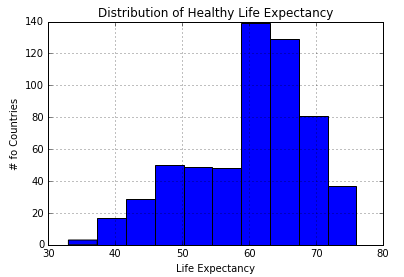
***Total rows:*** 776 ***Total Columns:*** 10

***Planned Analysis:*** This data set will be used as the baseline on which we compare the other factors.

Within this file, introductory analysis will review male vs. female life expectancy, and a detailed comparison by country, noting outliers. A final analysis on this file will look at life expectancy at birth, vs at age 60, and highlight those countries with the biggest discrepancies.

***Initial Analysis***

As our primary data set, some initial analysis has been completed. 

1. Life expectancy for both sexes ranges from a low of 38 years to a high of 87 years.
2. Countries where the expected life expectancy at birth (both sexes) is less than 50:
3. Countries where the expected life expectancy at birth (both sexes) is greater than 80:
4. Normal distribution of HALE for both sexes.

##### 

##### 

##### 

##### 

##### 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.

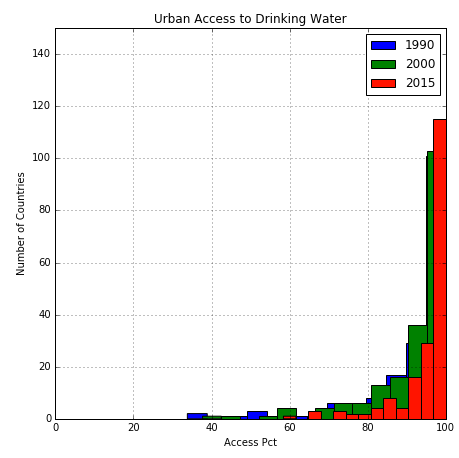
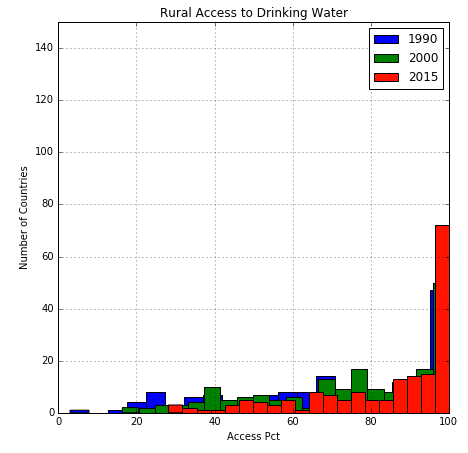
***Source:*** http://apps.who.int/gho/data/node.main.167?lang=en

***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.

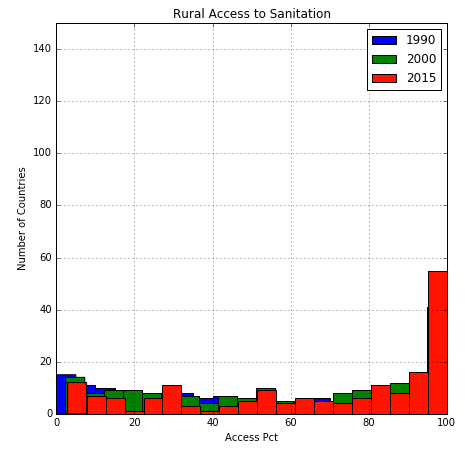
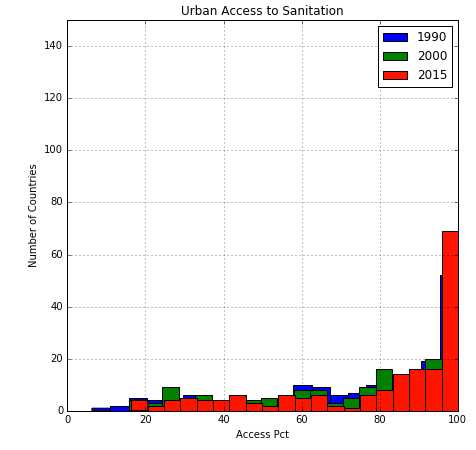
***Initial Analysis***

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.

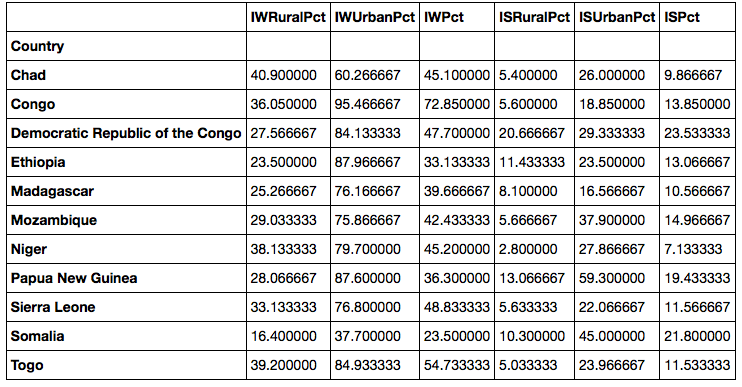


Urban access also improved, but mostly at the high end, moving from many countries with between 80 and 90% access in 1990 to most countries going to greater than 95% access in 2015.

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:



##### HIV:

***Overview:***  This data provides HIV prevalence statistics for 158 countries for 2001, 2005, 2009, and 2013. We will match up data as closely to Life Expectancy dates as possible (2001 with 2000 and 2013 with 2013/2015). In practice, the HIV prevalence does not shift extraordinarily rapidly.

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

***Total rows:*** 158 ***Total Columns:*** 5

***Planned Analysis:*** Overall, we will analyze the prevalence of HIV and life expectancy.

Within the file itself, we will review changes over time between 2000 and 2015 and at the distribution of HIV across countries.

##### Healthcare Services:

***Overview:***  This data provides indicators of the % of GDP spent on healthcare for 191 countries for the years 2013, 2012, and 2000. Additional indicators include the type of healthcare spending (government, private, out of pocket).

***Source:*** http://apps.who.int/gho/data/node.main.75?lang=en.

***Total Rows:*** 191 ***Total Columns:*** 13

**Planned Analysis:** Overall, we will analyze how health care spending relates to life expectancy. We will explore the distribution of healthcare spending as a % of GDP and the cross national differences between government and private spending. Finally, we will compare health care spending patterns with countries with high HIV prevalence.

##### Aid – Commitments and Disbursement:

***Overview:***  This data set documents aid committed and disbursed to 119 different countries between the years 2000 and 2010.

***Source:*** http://apps.who.int/gho/data/node.main.A1626?lang=en

***Total rows:*** 1295 ***Total columns:*** 3

***Planned Analysis:*** Initial analysis will look at the difference between commitments and disbursements by country. Additionally, we will review the distribution of aid, and highlight those receiving the most, the least, and how those numbers have changed over time. Finally, we will analyze aid with life expectancy, and determine if there is a noticeable effect over time, after receiving aid.

### 

### Appendix:

#### 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.*