Maddy Butts
ECON 428
Data Sources
Spring 2014

I. Malaria Data from WHO:

World Health Organization. World malaria report 2013 . World Health Organization, 2013. Retrieved from http://www.who.int/malaria/publications/world_malaria_report_2013/en/

Download excel spreadsheet for Annex 6D in the 2013 World Malaria Report from the WHO and name *original_malaria*

Copy data in the WHO region labeled "Africa" into *importable_alldata* and delete rows for each country labeled "No Pf," "No Pv," and "No other," so only "Suspected" rows for each country should be left.

Delete the following countries due to lack of data (for most) and deleted South Africa because it is not considered to be a developing country:

Cabo Verde, Cameroon, Comoros, Equatorial Guinea, Liberia, Mayotte, France, Sao Tome and Principe, South Africa, Swaziland, Delete mainland and Zanzibar under Tanzania, and Zimbabwe.

This leads to the final list of 33 African countries:

Algeria	Ethiopia	Namibia
Angola	Gabon	Niger
Benin	Gambia	Nigeria
Botswana	Ghana	Rwanda
Burkina Faso	Guinea	Senegal
Burundi	Guinea-Bissau	Sierra Leone
Central African Republic	Kenya	Swaziland
Chad	Madagascar	Tanzania
Cote D'Ivoire	Malawi	Togo

Democratic Republic of the Congo	Mali	Uganda
Eritrea	Mauritania	Zambia

Finally, in *importable_alldata* divide all data points by 1,000 so that the data becomes total cases per 1,000 people. Arrange data points by country and year.

II. FAO Data:

1) Land data:
FAOSTAT, Food and Agriculture Organization of the United Nations (2014). Retrieved from http://faostat3.fao.org/faostat-gateway/go/to/download/R/RL/E

Under Resources → Land I downloaded Forest Area, Agricultural Area, and total area equipped for irrigation for the above 33 countries from 2001-2009.

There are 4 boxes to select information on the web page: **Countries**, select the 33 countries in the above table. For **Items**, select

- Agricultural area
- Forest area
- Total area equipped for irrigation

Under **elements**, select "area" and under **years** select 2001-2009.

Under **Output Options,** deselect "show flags" then download as excel workbook named *original_land*

Copy all data to *importable_alldata* arrange data points by country and year.

2) Cereals data:
FAOSTAT, Food and Agriculture Organization of the United Nations (2014). Retrieved from http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E

Then download cereal data under Production →Crops. For **countries,** select the 33 countries in the table above. Then, choose **items aggregated** and select "Cereals, Total + (Total)." Select "yield" under **elements** and select 2001-2009 under **years.**

Deselect "show flags" under Output options and download as excel workbook named *original_cereal*

Copy all data to *importable_alldata*. Divide all data points by 10,000 to convert from hectograms/hectare to tons/hectare. Arrange data points by country and year.

III. World Bank:

The world bank, World Development Indicators (2014). Africa Development Indicators. Retrieved from http://databank.worldbank.org/data/views/variableselection/selectvariables.aspx?source=africa-development-indicators#

In this link on the World Bank website, first select the 33 countries in the above table. Under series, select the following 10 variables:

1) GDP per capita, PPP (Constant 2005 international $)
2) CO2 emissions from cement production (thousand metric tons)
3) Rural population (%) of total population
4) Population, total
5) Population ages 0-14 (% of total)
6) Terrestrial protected areas (sq km)
7) Life expectancy at birth, total (years)
8) Health expenditure, public (% of GDP)
9) Gross public investment (% of GDP)
10) Labor force participation rate, total (% of total population ages 15-64)

Under **years**, select 2001-2009. Download as excel workbook named *original_worldbank*

Copy all data to *importable_alldata* and make the following changes:

- Divide all *Terrestrial protected areas* variables by 10 to convert from square kilometers to area per 1000 hectare for better comparison with other land variables

Then arrange data points by country and year.

IV. World Governance Indicators

Daniel Kaufmann, Aart Kraay and Massimo Mastruzzi (2010). Worldwide Governance Indicators. The World Bank. Retrieved from http://info.worldbank.org/governance/wgi/index.aspx#home

Select "download full dataset (excel)" and name as *original_wgi*

Only "Regulatory quality" and "Government effectiveness" data sheets need to be kept. Delete years before 2001 and after 2009 for both variables. Use filter feature in excel to select the 33 countries in the above table and delete all other countries.

Then, for each year in both variable sheets, delete the columns: *StdErr, NumSrc, P-Rank, Lower,* and *Upper*. Only the column *estimate* should remain for each year.

Then arrange Regulatory quality and Government Effectiveness data into *importable_alldata by country and year.*

V. Population living within 100 km of coast

UNEP (2014): The UNEP Environmental Data Explorer, as compiled from UNEP/DEWA/GRID-Geneva . United Nations Environment Programme. Retrieved from http://geodata.grid.unep.ch/mod_download/download.php

From link above, select 2001-2007 under **define years** then select "Population Within 100 Kilometers of Coast" under **define dataset** and select "national" under **define options.** Then select download as Microsoft excel file and save as *original_coast*

Copy data in tab labeled *national* into another sheet. Use filter feature in excel to select the 33 countries in the above table and delete all other countries. Delete columns: *Sovereign, GEO Region, GEO Subregion, GEO ID, ISO 2 Code, ISO 3 Code, UN Code, Developed, Least Developed, OECD, Sub-saharan, Small Island Dev. States,* and *Arab World.*

Arrange all data points in *importable_alldata* by country and year. Create new variable in *importable_alldata* named *poppercoast* by dividing "Population Within 100 Kilometers of Coast" by "Population" to make the units on population data consistent.

(n.b. data points for this variable only from 2001-2007 so add … to all empty cells)