Econ 494 - Homework #1

Due

September, 3 2019

Background

Read the following article, "Making data dance", The Economist. 9 December 2010, link, and consider whether you agree or disagree with the following claims.

1. Today it no longer makes sense to consider the world as divided between developing and industrialized countries.
2. People everywhere respond similarly to increasing levels of wealth and health, with higher material aspirations and smaller families.

Assignment

Import and summarize the Gapminder data provided in class to investigate claims 1 & 2, above.

Data Source

A nonstandard source of economic data, Gapminder produces free teaching resources with the goal of making the world understandable through reliable statistics, learn more here.

Methodology

1. Download the **gapdata.zip** files used in class from Canvas or the course website.
   (a) Zip-file is approx. 0.3 MB in size.
   (b) Unzipped file contains four files; gpd.txt, lifespan.csv, pop.csv, and region.raw.
2. Write a do-file to import, clean, and save each file into a seperate Stata .dta file.
   (a) **Insheet** should be used to import all .csv and .txt files.
   (b) **Infile** should be used to import all .raw files.
   (c) **Reshape** long the variable *tp* by *cc* and *year*, and **rename** *tp pop*.
   (d) **Reshape** the variable *gdp*, and **replace** its value with $ln(gdp)$.
   (e) **Sort** all variables first by *cc* and then by *year*.
   (f) **Save** a copy file as a .dta file of the same name.

3. **Merge** the four saved .dta files created above into a single .dta file.
   (a) Use a **merge 1:1** on *cc* and *year* to combine lifespan, pop, and gdp data.
   (b) Decide to use either a **m:1** or **1:m** merge to combine the regions data.
   (c) Check the results of each merge in the results window.
   (d) Save the combined files as **gapdata.dta**.
4. Begin exploring the data by producing some basic summary statistics and tables.
   (a) **Describe** displays variable names, storage types, and display formats.
   (b) **Summarize** displays basic or detailed summary statistics; mean, min, max, sd.
   (c) **Tabulate** allows you to create one or two-way frequencies tables.
   (d) **Tabulate, summarize(·)** adds summary statistics to one- and two-way tables.
   (e) **Table** is another option to create tables of summary statistics.
5. The following commands demonstrate different ways to summarize the data. Try to find one or two tables <u>not</u> listed here to support or refute the above claims.

```
describe *
tabulate group6
summarize life pop gdp
sort group6
by group6: su life pop gdp
tabulate group6, summarize(life)
tabulate group6 if year == 2016, summarize(life)
tabulate group6 if year == 1948, summarize(life)
table group6, c(m gdp)
table group6, c(m life m gdp m pop)
table group6 if year>1899 & year<1948, c(sd life sd gdp)
```

6. Graphs are another convenient way to summarize and display data. Try to find one or two graphs to support or refute the above claims.

```
graph twoway line life year if cc=="United States"
twoway line life year if cc=="United Kingdom"
line life year if cc=="United States" || line life year if cc=="United Kingdom"

twoway scatter life gdp if year==1940
twoway scatter life gdp if year==1960, mlabel(cc)
twoway scatter life gdp if year==1980, mlabel(id) mlabpos(0) ms(i)
twoway scatter life gdp if year==2000 [w=pop]
```

7. Submit a .zip folder containing the following to the online Homework folder.
   (a) A do-file containing all relevant code and comments.
   (b) A summary of your findings, including any relevant tables or charts.