Data Management and Visualization

**Assignment – Week 1**

**Selecting a research question**

**by mapolarbear@gmail.com**

**Step 1. Choose a data set that you would like to work with.**

* I’m very interested in demographics and I think **GAPMINDER** dataset fits my research question. I’ve also checked [www.gapminder.org](http://www.gapminder.org) containing tons of data, and I’m excited to dig and use it.

**STEP 2. Identify a specific topic of interest**

* My **main specific topic** of interest is suicide rate, and if variables **suicider100TH** (suicide rate) correlates with various **unemployrates** in datasets provided by [www.gapminder.org](http://www.gapminder.org) among different countries.

**STEP 3. Prepare a codebook of your own**

The GAPMINDER dataset has limited data, that’s why I think it would be beneficial to add detailed statistical data to complete my research. The additional data will contain 2002 – 2004 time series aggregated data from various code book. Data will be also joined by reported country and retried indicators/variables will be renamed according to data meaning.

**GAPMINDER** - Gapeminder base dataset

**Country** - Name of reported country variable

**Suicider100TH** - 2002-2004 combined (2005) suicide age adjusted, per 100 000 mortalities due to self-Inflicted injury, per 100 000 standard population, age adjusted

In addition, following copybooks and corresponding variables/indicators chosen for the purpose of my research:

**Population** - Total population per country per year dataset

**Country -** Reported countryvariable

**2002** - Year 2002 total population per reported country variable

**2004** - Year 2004 total population per reported country variable

**WB Data GDP** - Real GDP per capita PPP 2005 dataset

**Country -** Reported country variable

**2002** - Year 2002 GDP per Capita per reported country variable

**2004** - Year 2004 GDP per Capita per reported country variable

**Total suicides** - Total suicides per reported country dataset

**Country -** Reported country variable

**2002** - Year 2002 total suicides per reported country variable

**2004** - Year 2002 total suicides per reported country variable

**Suicide female indicator file UL 20100916** - Suicides female indicator dataset

**Country -** Reported country variable

**2002** - Year 2002 female suicides for reported country (%) variable

**2004** - Year 2004 female suicides for reported country (%) variable

**Suicide male indicator file UL 20100916** - Suicides male indicator dataset

**Country -** Reported country variable

**2002** - Year 2002 male suicides for reported country (%) variable

**2004** - Year 2004 male suicides for reported country (%) variable

**indicator\_t long term unemploy** - Total long term unemployment rate per reported country dataset

**Country -** Reported country variable

**2002** - Year 2002 total reported unemployed (%) variable

**2004** - Year 2004 total reported unemployed (%) variable

**indicator\_f long term unemploy** - Total long term unemployed female rate per reported country dataset

**Country -** Reported country variable

**2002** - Year 2002 total reported female unemployed (%) variable

**2004** - Year 2004 total reported female unemployed (%) variable

**indicator\_m long term unemploy** - Total long term unemployed male rate per reported country dataset

**Country -** Reported country variable

**2002** - Year 2002 total reported male unemployed (%) variable

**2004** - Year 2004 total reported male unemployed (%) variable

**STEP 4. Identify a second topic that you would like to explore in terms of its association with your original topic.**

* My **second specific** topic of interest is if long **term unemployment** among **female** and **male** correlates with suicide rate. If indicators are correlating, is there **difference between female** **and male suicide ratings**

**STEP 5. Add questions/items/variables documenting this second topic to your personal codebook.**

* Variables in question are listed in STEP 4.

**Questions:**

1. Does correlation between suicides and long term unemployment exists?
2. What is the difference (if any) in suicidal deaths between unemployed females and males?
3. Is it true that males are more suicidal than females?
4. It there any difference in suicidal ratings between 5 reported countries with highest GDP per capita and 5 lowest?

**STEP 6. Perform a literature review to see what research has been previously done on this topic. Use sites such as Google Scholar (http://scholar.google.com) to search for published academic work in the area(s) of interest. Try to find multiple sources, and take note of basic bibliographic information.**

* **Related literature and research studies (http://scholar.google.com):**

[**http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0051333**](http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0051333)

[**http://www.aerzteblatt.de/pdf/DI/110/23/m413.pdf**](http://www.aerzteblatt.de/pdf/DI/110/23/m413.pdf)

[**http://www.bmj.com/content/345/bmj.e5142.full.pdf+html**](http://www.bmj.com/content/345/bmj.e5142.full.pdf+html)

[**http://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.94.12.2155**](http://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.94.12.2155)

**STEP 7. Based on your literature review, develop a hypothesis about what you believe the association might be between these topics. Be sure to integrate the specific variables you selected into the hypothesis.**

* My hypothesis is that there is a correlation between suicide deaths and long term unemployment. I think when people don’t have job for a long time they feel disconnected from society and don’t see purpose of existence.