

# BIOS 600 · Global Topics Group Project I

Due 16 September 2011

**Submission Instructions:** This is a group project. Each group will submit 1 electronic copy and 1 paper copy of its final product by Friday, 16 September 2011. The electronic copy should be in pdf format. Please submit the electronic copy by email to [tg5@email.unc.edu](mailto:tg5@email.unc.edu).

**Honor Code:** Your final product should reflect the work of group members.

**Project Goal:** A good table or plot can add context and clarity to concepts difficult to communicate in any other way. The goal of this project is for you to thoughtfully consider

- (a) When is a plot/table helpful?
- (b) What makes a good plot/table good?
- (c) What makes a poor plot/table poor?

A secondary goal of this project is to expose you to the work of serious researchers dedicated to creating innovative and effective data displays. Hopefully you will see how these plots can aid you as you develop models or communicate results.

This project is also part of the School of Public Health's initiative to expose you—the student—to global health topics.

**Background:** Many studies have investigated the association between male circumcision and HIV prevalence. One of the first studies (Bongaarts et al., 1989) examined the relationship between the percentage of males who were circumcised in 37 African countries and the HIV seroprevalence in those countries based on estimates from the capital city. The article by Drain et al. (2006) explores the same relationship but with slightly different data.

**Instructions:** Obtain a copy of both articles, and work together to answer the following questions. Some the questions will ask you to create tables and plots. You may use Excel to create the tables and plots, but I encourage those students seeking an H to work in a scripted language. (This project is an opportunity for you to learn many of the skills required by the Computer Skills project.)

**Question 1:** Create a concise outline of each article. Each outline should include a clear statement of the study's (a) primary research question, (b) data sources, and (c) conclusions. Be careful to not overstate the strength of the results. Both of these studies are *Ecological studies* which means the data are collected at a population-wide level. Such studies are helpful when looking at population level relationships, but they do not warrant individual specific conclusions.

**Question 2:** Recreate Figure 2 of Bongaarts et al. (1989); the data for the plot are reported in Table 1 of the article. Add a text label to 3 extreme points.

**Question 3:** Edward Tufte is a statistician known for his work in data visualization. In Tufte (2007), he wrote:

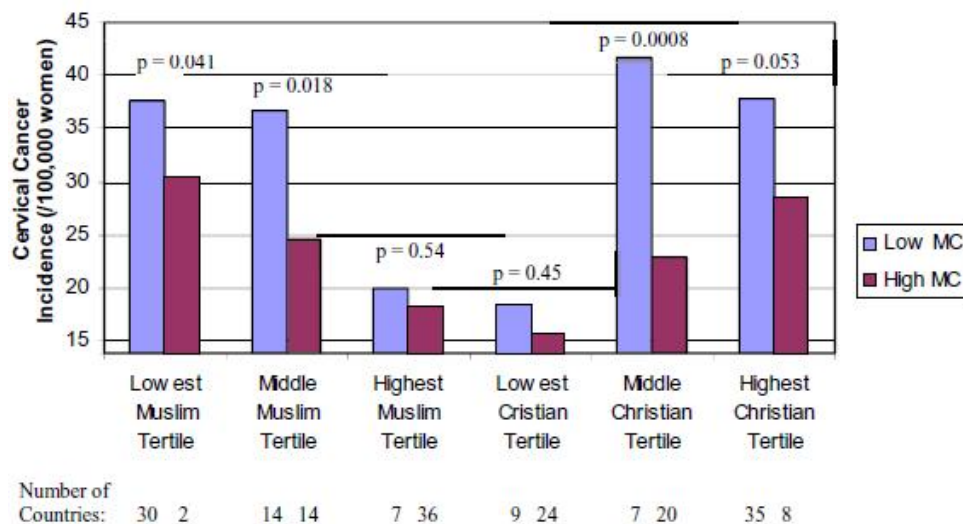
Excellence in statistical graphics consists of complex ideas communicated with clarity, precision, and efficiency. Graphical displays should

- show the data
- induce the viewer to think about the substance rather than about the methodology, graphic design, the technology of graphic production, or something else
- avoid distorting what the data have to say
- make large data sets coherent
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the statistical and verbal descriptions of a data set.

Graphics reveal data. Indeed graphics can be more precise and revealing than conventional statistical computations....

Evaluate Figure 1 of Bongaarts et al. (1989) in light of Tufte's statement. If you find a shortcoming in the figure, be sure to suggest how to fix it. Is there a better way to display the same data? Does this data display add value to the article? What exactly does the author want the reader to get from the figure? Does it offer the reader context? (You do not need to answer these questions directly; I offer them as way to generate discussion.)

**Question 4:** Consider Figure 1 of Drain et al. (2006). Suppose the authors chose to report the following bar chart in order to save space.



Would this more compact figure communicate the same information as the original? If you answer yes, explain how. If you answer no, explain why not.

**Question 5:** The HIV prevalence data used in Drain et al. (2006) is publicly available. You can download the data from the UNAIDS website ([www.unaids.org/en/dataanalysis/epidemiology/](http://www.unaids.org/en/dataanalysis/epidemiology/)). Use the HIV prevalence data, the data found in the articles, or other publicly available and reputable datasets to create a data display that communicates a complex idea. The plot should be no larger than a single page. In a few sentences, explain the purpose of your display and why your display achieves that purpose.

There are many data plots not discussed in chapter 3 of the textbook, and you are free to consider those as well. For example, one of plot is called a *parallel coordinate plot* or a *slopegraph*. For an example published in Tufte (2007), see [charliepark.org/slopegraphs/](http://charliepark.org/slopegraphs/).

**WARNING:** Admittedly, this is a very open ended question, and it has the potential to be very time consuming if you allow it. Do not allow it to consume your time. Be decisive about the data and the plot. Do not spend time worrying about the WOW factor. I simply want to see that you are thinking about the principles of good graphical displays.

## References

- Bongaarts, J., Reining, P., Way, P., and Conant, F. (1989), "The relationship between male circumcision and HIV infection in African populations," *AIDS*, 3, 373– 378.
- Drain, P., Halperin, D., Hughes, J., Klausner, J., and Bailey, R. (2006), "Male circumcision, religion, and infectious diseases: an ecologic analysis of 118 developing countries," *BMC Infectious Diseases*, 6, 172.
- Tufte, E. R. (2007), *The Visual Display of Quantitative Information*, Graphics Press LLC, 2nd ed.