Lab 02: Measurement Scales,  Statements and Big Data

**Handout date:** Monday, September 09, 2019

**Due date:** Wednesday, September 18, 2019 at the beginning of the lecture as hardcopy

*This lab counts 4 % toward your total grade*

**Objectives:** In this lab   
[a] you will identify and justify the measurement scale of several variables.

[b] you will practice some data manipulation techniques as well as  statements, and  
[c] read article on big data and answer some questions.

**Format of answer:** Your answers (graphs and verbal description) should be handed in as ***hard-copy*** in ***one*** document. Add a running title into the header of the document with the following information: ***your name***, ***GISC6301-Lab02*** and ***page numbers***. Label each answer properly starting with its task number. Maintain the sequence of questions. Format any code and computer output properly before inserting it into the document with your answer. -code and text output need to be in a ***monospaced*** font (i.e., fixed-pitch font) such as Courier New so proper spacing and alignments are preserved. Excessive, but irrelevant, output will lead to a deduction of your accumulated points.

## Task 1: Identify and justify the measurement levels of several statistical variables (1 point)

Justify your selection of the measurement scale. You may want to look up Wikipedia for some of the variables

1. The longitude and latitude in degrees on the earth’s spherical surface. Be cautious in your arguments with regards to the longitude. That is, are distances between two longitudes constant all over the globe? (0.2 points)
2. The temperature on the Celisus scale. (0.1 points)
3. The wind direction in degrees. (0.1 points)
4. Number of break-ins in a neighborhood. (0.1 points)
5. The hierarchical classification of U.S. census enumeration units. (0.1 points)
6. The decibel in acoustics (0.1 points)
7. Income (0.1 points)
8. Income brackets for taxation purposes (0.1 points)
9. Elevation above sea-level at a fixed point in time. (0.1 points)

## Task 2: Working with Data (2 points)

For all tasks below show your properly formatted code. You find the necessary code for the examples in Lander. Only if asked show also the output.

1. Import the SPSS data-file **Concord1.sav** in the Lab02 folder as ***data-frame*** into the R environment by using a function from the library **foreign**. Make sure to name your data-frame properly. (0.2 points)
2. Show the ***tail*** of the last 6 records of the imported data-frame. (0.1 points)
3. Show the ***summary*** information for the imported data-frame. (0.1 points)
4. Discuss the summary: How did the ***average*** water consumption change from 1979 to 1981? (0.1 points)
5. Discuss the summary: Which variable has ***missing*** observations? (0.1 points)
6. Discuss the summary: Which variable is a factor? (0.1 points)
7. List all ***case numbers***, which have at least for one variable missing data. Show also the code. (0.2 points)
8. Which ***class*** are the following data selections: [a] **Concord$retire**, [b] **Concord["retire"]**, [c] **Concord[ , "retire"]**, and [d] **Concord[["retire"]]**? Show code and the output. (0.2 points)
9. Calculate the ***average*** of the household’s water consumption for the 3 years and save it the new variable **meanWater** into the data-frame. (0.2 points)
10. Remove the variable **case** from the data-frame by assigning it **<- NULL**. Show the code (0.1 point)
11. Add a new **caseID** variable by labeling each record by its record number ranging from 1 to the number of observations. Show the code. (0.1 point
12. ***Bind*** the two variables **peop80** and **peop81** together into a ***matrix***. Show the code. (0.1 points)
13. Give a code example of the use of the **ifelse** statement. (0.1 points)
14. Give an example of the **while** statement (0.1 points)
15. What are [a] positional, [b] named and [c] default arguments of a function? (0.3 points)

## Task 3: Critical discussion of big data analyses (1 point)

Read the document **BigDataAndStatitics.pdf** in the **Lecture01** folder.

[a] Summarize in your own words in a few sentences an example where and why big data analysis failed. (0.5 points)

[b] In a few sentences provide some arguments why theory and statistics are still important in the era of big data analysis. (0.5 points)