**Name:Zhan Zhao**

**Lab #1: Introduction to R and the TANGO dataset**

People who use drug (PWUD) are at increased risk of malnutrition through changes in eating habits due to several factors, including living a drug-dependent lifestyle, using drugs that induce anorexia, having limited financial resources, and/or having chronic diseases such as hepatitis B or C, tuberculosis and/or HIV-1 infection. In Argentina, it is estimated that approximately 133,000 individuals are currently living with HIV/AIDS. The epidemic is largely concentrated in urban areas, primarily (~80%) in and around Buenos Aires. It is reported that Argentina has one of the highest rates of HIV incidence among PWUDs in Latin America. In 2006, we enrolled former and current drug users living in Buenos Aires into a cross-sectional study (the Tango study). In this exercise, you will begin to examine this dataset more closely.

For this exercise, you will need to answer the questions below. In addition to the answers, you must also create an **R** **file** which shows all the R commands you used to obtain the answers.

Open the **tango.rds** data, you may assign it any name, such as “**tg**”, keep this data name short, as you will be referring to it frequently.

### How many variables (columns) are there?

39 variables

### How many observations (rows) are there?

205 observations

### What are the labeled responses for the **drinker\_cat** variable?

"Heavy Drinker" "Light Drinker" "Moderate Drinker" "Non-Drinker"

### What are the labeled responses for the **sex** variable?

"Female" "Male"

### What percent of study participants ever injected drugs (idu\_ever)?

42.43902

### How many men ever injected drugs?

78

### What percentage of men injected drugs?

43.33333%

### What percentage of injection drug users were women?

36%

### What are the mean, minimum and maximum values for the variable, **weight\_kg**?

Mean = 70.2961 , minimum = 46.5 , maximum = 146.

### Describe the shape of the variable, **weight\_kg**?

Right skew.

### Make an appropriate plot that can answer these questions: What is the median for weight? Are there any outliers?

Median = 69.0 , There are 1 outlier, number”146”.

### Compare the distribution of weight between males and females. Provide a narrative description, and include graphical evidence.

The male distribution has a higher median and mean than the female distribution, and males also show a wider spread.For males, the median is 69.50, which is lower than the mean (71.29). This suggests that the distribution is slightly right-skewed. The typical range (from the first to the third quartile) is 47.7 to 77.1, but the graph also shows outlier values above this range.For females, the mean (63.17) is also higher than the median (60.50), so the distribution is also right-skewed. The typical range is 46.5 to 69.0, and there is an outlier at 91.5.