EDUC 641 Lab: Applied Statistics in Education and Human Services I

Lab 3: 10/18 and 10/19

## Focus on describing and summarizing one continuous variable and recall using filter() to look at individual observations

**Goals:**

1. **Describe summary statistics of one continuous variable using the functions:** mean, median, range, IQR, variance, standard deviation, summary
2. **Visualize the distribution of one continuous variable**: using **geom\_histogram()** and **geom\_density()**
3. Recall using **filter()** function to look at an **individual observation** (e.g., one teacher)

***What’s NEW?!***

**Summary functions:**

**mean(), median(), range(), IQR(), sd(), var(), summary()**

**Distribution functions using ggplot():**

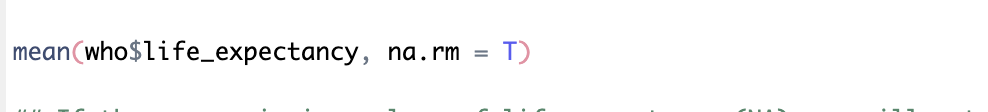
**geom\_histogram()** and **geom\_density()** to create visualizations of the distribution of our data.

**Missing variables:**

You can set the **na.rm** parameter to **TRUE (to remove any missing values if they exist)**

**Ex.**

**‘ na.rm** **=** T **‘**



**Worksheet**

**Don’t forget to load packages: library(**tidyverse**)**

**Measures of Central Tendency**

1. **Suppose you want to know the average or measures of central tendency of your variable.** 
   1. **Use the mean() function to find the average**

**mean(**your\_data$variable\_name, na.rm = T**)**

* 1. **Use the median() function as another measure for central tendency**

**median(**your\_data$variable\_name, na.rm = T**)**

**Measures of Variation**

1. **If you want to know how much your variable ranges, try:**
   1. **range(**your\_data$variable\_name, na.rm = T**)**
   2. **IQR(**your\_data$variable\_name, na.rm = T**)**

**Measures of Spread**

* 1. **var(**your\_data$variable\_name, na.rm = T**)**
  2. **sd(**your\_data$variable\_name, na.rm = T**)**
  3. **summary(**your\_data$variable\_name**)**

\*\*The **summary** function returns a bunch of summary statistics of one continuous variable including, minimum value, maximum value, mean, median, first quartile, third quartile, and the number of NAs.

**Try yourself:** *Summarize the descriptive statistics of the* ***life\_expectancy*** *variable from the* ***who*** *dataset.*

**Distribution**

1. **To visualize the distribution of your data:**
   1. **‘ geom\_histogram() ‘**

**ggplot(**your\_data**, aes(x =** your\_variable**)) +**

**geom\_histogram() +**

**labs(**title = “Your title distribution of variable\_name”,

x = “Title of x-axis”,

y = “Frequency *or* Count”**)**

* 1. **‘ geom\_density() ‘**

**ggplot(**your\_data**, aes(x =** your\_variable**)) +**

**geom\_density() +**

**labs(**title = “Your title distribution of variable\_name”,

x = “Title of x-axis”,

y = “Density”**)**

**Try yourself:***Represent the distribution of the* ***life\_expectancy*** *variable using* ***geom\_histogram()*** *and* ***geom\_density()*** *from the* ***who*** *dataset.*

1. **Suppose you want to look at the data relevant to only one individual teacher** 
   1. **subset\_teacher1 <- filter(**your\_data**,** teacher\_variable =="individual\_teacher")

**Try yourself:** *Filter the* ***who*** *dataset to look at only Canada. What is the* ***life\_expectancy*** *in Canada?*