**Lab 1.1**

**Introduction to Plots in Rstudio**

**Background of the Study**

We will focus on a random sample of 100 teens between the age 14-18 from the CDC Data. By the end of this lab, we hope to gain some insights about child obesity in the USA.

**Types of Variables**

To load the data into R's working memory; follow the below instructions.

load('~/Dropbox/IDS Curriculum Design/Folder-James/Unit 1/Data/cdc.rda')

View(cdc)

The data set cdc that shows up in your workspace is a **data matrix**, with each row representing a case (a teen) and each column representing a variable. R calls this data format a **data frame**.

What story can you tell about the first case (teen) in our **data set**?

To view the names of the variables, type the command:

names(cdc)

This returns the names of the variables. Each one of these variables corresponds to a question that was asked in the survey. For example, for general\_health, respondents were asked to evaluate their general health, responding either excellent, very good, good, fair or poor.

1- How many cases are there? How many variables? For each variable, identify its data type (e.g. categorical, numerical).

There is an easy way to find the number of rows and the number of columns in our data. Try:

dim(cdc)

**Graphical Summaries**

To create a plot we can use the following command:

plot(~general\_health, data = cdc)

plot(~height, data = cdc)

Comment on your plots! Is this what you expected to see?

**Two Variables**

**Scatter Plot**

When we're interested in learning about the relationship between two numerical variables, we use scatter plot to describe this relationship. For example:

plot(weight ~ height, data = cdc)

What story can you tell about the relationship between height and weight?

**Side by Side Plots**

When we're interested in learning about the relationship between one numerical and one categorical variables:

plot(weight ~ depressed, data = cdc)

How many plots there are in your plot? Can you guess why? Also, can you guess the type of the plots?

**Segmented Bar Chart**

When we're interested in learning about the relationship between two categorical variables:

plot(general\_health ~ depressed, data = cdc)

Describe what you see. Why do you think there are 5 segments in each bar?