

AR HW#3

Calvin Makelky, A01872013

November 11, 2016

```
options(repos="https://cran.rstudio.com" )
#install.packages("RCurl")
library(RCurl)

## Loading required package: bitops

#install.packages("XML")
library(XML)

cig_wiki = "https://en.wikipedia.org/wiki/List_of_countries_by_cigarette_consumption_per_capita"
cig_url= getURL(cig_wiki)
cig = readHTMLTable(cig_url, which=1, stringsAsFactors=FALSE)
colnames(cig) = c("Ranking", "Country", "Cigarette Use")

life_wiki = "https://en.wikipedia.org/wiki/List_of_countries_by_life_expectancy"
life_url= getURL(life_wiki)
life= readHTMLTable(life_url, which=1, stringsAsFactors=FALSE)
colnames(life)= c("Country", "Both Sexes Rank", "Both Sexes Life Expectancy",
"Female Rank", "Female Life Expectancy", "Male Rank", "Male Life Expectancy",
"Both Sexes Rank HALE", "Both Sexes Life Expectancy HALE")

cig = cig[order(cig$Country),]
life = life[order(life$Country),]

cig_life <- merge(cig,life,by="Country")

install.packages("RColorBrewer")

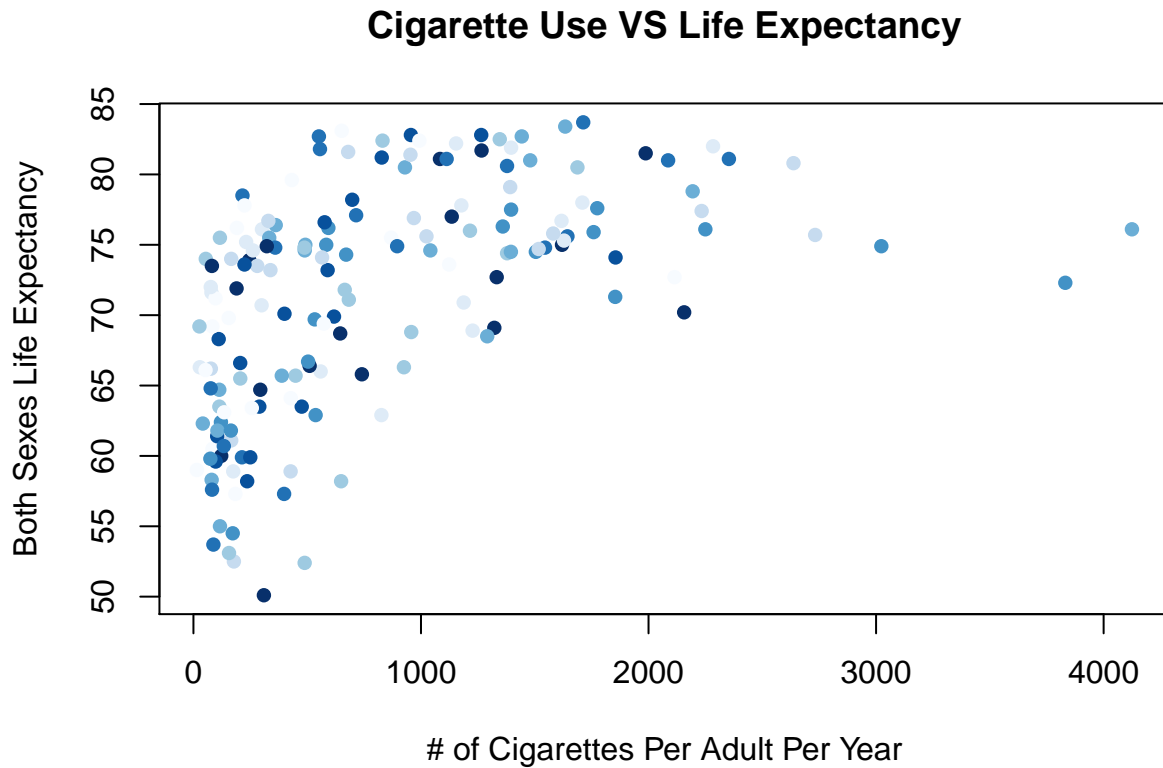
## Installing package into 'C:/Users/CJ/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'RColorBrewer' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\CJ\AppData\Local\Temp\RtmpmITwEb\downloaded_packages

library(RColorBrewer)
mycolor = brewer.pal(10, "Blues")

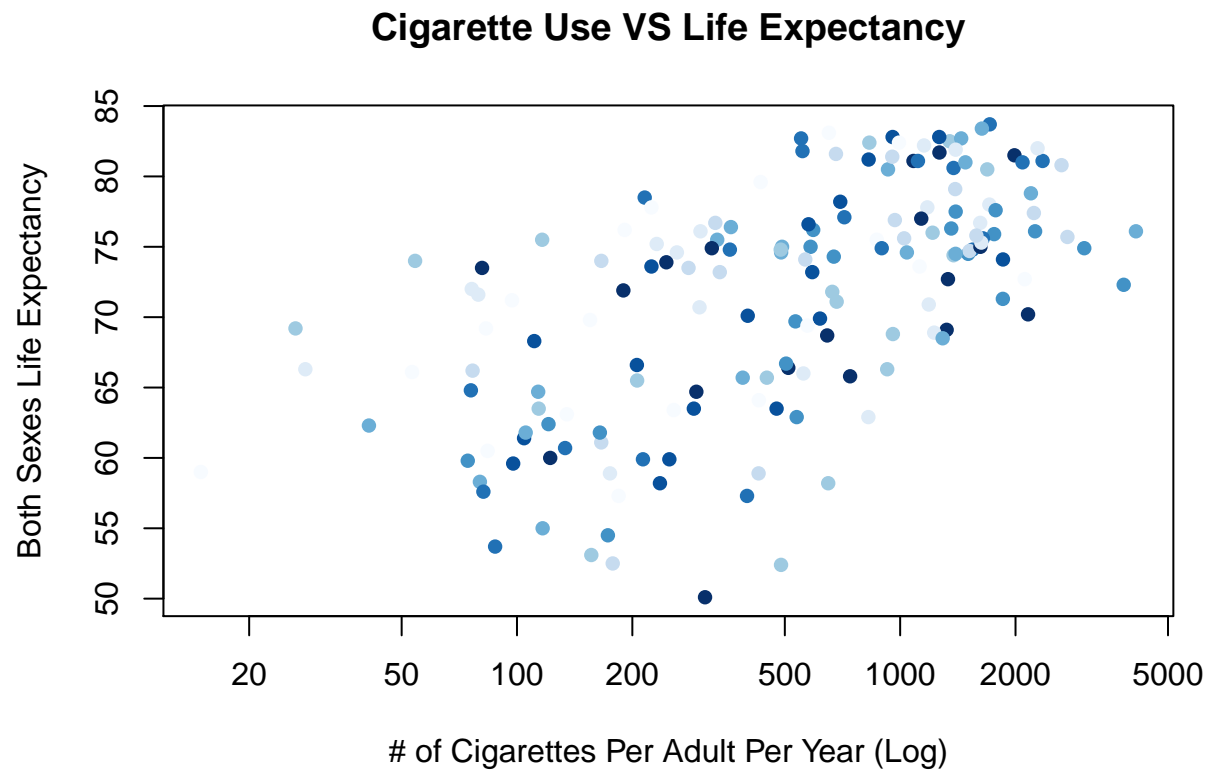
## Warning in brewer.pal(10, "Blues"): n too large, allowed maximum for palette Blues is 9
## Returning the palette you asked for with that many colors

plot(cig_life$`Cigarette Use`, cig_life$`Both Sexes Life Expectancy`,
     main="Cigarette Use VS Life Expectancy", ylab = "Both Sexes Life Expectancy",
     xlab = "# of Cigarettes Per Adult Per Year", pch=16, col=mycolor)
```



The first plot shows that there does seem to be an positive association between cigaretatte use and life expectancy in countries, although it is not linear in nature. Typically the higher cigarette use, the higher life expectancy for a country. There are many countries with low life expectancies and low cigarette usage, but there is no countries with low life expectancies and high cigarette usage. It's surprising that there is a positive association instead of a negative association between these two variables. Maybe there is a confounding variable like health care or violence. For example, countries that smoke more tend to have better health care so this leads to better life expectancy. Or maybe countries with higher cigarette use tend to have less violence, hence, long life expectancy.

```
plot(cig_life$`Cigarette Use`, cig_life$`Both Sexes Life Expectancy`,
     main="Cigarette Use VS Life Expectancy", ylab = "Both Sexes Life Expectancy",
     xlab = "# of Cigarettes Per Adult Per Year (Log)", log="x", pch=16, col=mycolor)
```



Due to the first plot having a distribution similar to a log distribution rather than a linear distribution, I plotted the log of cigarette use on the x-axis in the second plot. This worked pretty well, as the plot banks at 45 degree angle that we want.