STA 404/504 Homework 5 - Categorical Displays

Submission through Canvas - Due Monday 3/11/19 by 5:00 PM

Learning Objectives:

- Cleaning data for categorical displays
- Various types of multivariate categorical displays
- Comparison of plot types
- Effectively supporting data narratives

Assignment:

For this assignment you will be continuing to use the BRFSS data to build a few different categorical variable displays. This time you will be using the data directly from the source (CDC website) to download the data from the full 2017 survey. Please download from the link labeled "2017 BRFSS Data (SAS Transport Format)" [ZIP size 101 MB], extract the files then move the file named "LLCP2017.XPT" to your homework 5 folder. Note this file is BIG (1.4 GB) so it will take time to load into R. Use the attached code titled “Homework5_Startup.R” to get started.

Note for MAC users: Several mac using students had issues with reading in the data. After some digging we figured out the the mac software tool that extracts from the zip format added a “sneaky space” to the end of the file extension type. After you set the working directory you can run the *list.files()* function to look up all the files in your wd. You will likely find that instead of "LLCP2017.XPT", it is listed as "LLCP2017.XPT ". Thus to read it in you would need:

```
full_data <- foreign::read.xport("LLCP2017.XPT ")
```

Once you have the data loaded you need to make the following plots:

1. Heatmap: general health (GENHLTH) vs sex (SEX)
2. Proportionally stacked barplot: pregnancy (PREGNANT) rates across months (IMONTH) for female responses
3. Mosaic plot: joint pain (JOINPAI1) within BMI categories (X_BMI5CAT)
4. Your Choice of Plot: association between number of children (X_CHLDCNT) and education (X_EDUCAG)

For each of the plots above, your write-up should include the plot image and one paragraph summarizing what the plot tells us in the context of the data.

Submission Format: **R code file** showing your work **+ a pdf or word document** contain your written solutions. Plot images need to be included in your write-up! (R markdown is an acceptable alternative for those who know how to use it; if used please submit the .Rmd and pdf document.)

What will we look for in grading?

[2 pts] Code - well commented/structured/organized/documented

[4 pts] Structure of plot - Correct plot type and correct use of aesthetics

[4 pts] Plot design and descriptions - axis labels and titles, labeled units, appropriate use of color, proper grammar in write-up, contextual insight into data.