HW5

INSTRUCTIONS

Complete the below data visualization tasks in R Markdown and submit the Markdown file or the Notebook HTML file on Blackboard. Don't forget to use promising practices in your coding. This homework continues to use the BRFSS.48K.csv that we used in class.

0. Read in data and load libraries

1. Bar plot with calculated order

Create a bar plot of counts of `income` column. Order the columns by descending order, so that the most counts (tallest bar) is on the left.

2. Bar plots of given height

Suppose we want to calculate the total weight of all individuals per state. We can add up all weights for each state with the following,

```
library(plyr)
state.weight = ddply(BRFSS, "state", summarise, tot.weight=sum(weight))
head(state.weight)
```

Create a bar plot with the total weight for each state indicated by bar height. Label your Y variable with units.

Hint Note `geom_bar` has the default argument, `geom_bar(stat="identity")`.

3. Scatter plot with age

Create a scatter plot of height and weight, with age indicated for each point by some other aesthetic. Try using aesthetics such as color, fill, size, shape, and alpha (opacity) to emphasize features which may be interesting. Label your X and Y variables with units.

4. Compare sexes

Create a scatter plot of height and weight that illustrates the distibution of the two sexes. Label your X and Y variables with units.

- Use aesthetics of your choice
- Draw two separate plots, one for each sex
- **Bonus (0.1 extra points)**: Use *facets* to draw separate plots for each sex

5. Bubble plots

Create a bubble plot, which displays counts of observations two categorical variables as the size of a geom_point. We can obtain counts of income and employed status, for instance, with the below code.

```r
library(plyr)
income.employed.count = count(BRFSS,  c("employed", "income"))
head(income.employed.count)
```

6. Superpose boxplots over violin plots

a. Draw box plots (geom_boxplot) instead of violins for weight and diabetes
b. Superimpose the boxplots over the top of the violin plots. Label your X and Y variables with units.

7. Indicate significant differences (Bonus problem 0.1 extra points)

Evaluate whether the weight distribution between any two pairs of diabetes status categories on the violin plots is significantly different. Devise a way to indicate which pairs of distributions are significantly different.