Stat 145, Mon 13-Sep-2021 -- Mon 13-Sep-2021
Biostatistics
Spring 2021

Monday, September 13th 2021

Wk 3, Mo
Topic:: R Markdown source file
Due:: Quiz Ch. 1 ends at 10 pm

population
 - like taking a census
 - for many quantitative vars
    does not have the choppiness imposed by histograms
    has a symmetric bell-shaped (normal, Gaussian) distribution
      gf_dhistogram(~ TotChol, data=NHANES, color="black")   with
      gf_density(~ TotChol, data=NHANES)

 - std deviation
    same units as that of quantitative variable
    notation: sigma vs. s
    visualizing
    sigma as a "unit" of measure
      visualizing
      standardizing a score
        Z = ((unstandardized score) - mean) / (standard deviation)

Q4: Who performed better?
     Millie with score of 1410 on the SAT (mean = 1026, sd = 209), or
     Michal with score of 27 on the ACT (mean = 20.8, sd = 4.8), or

R Markdown
 - getting started with a template
 - initialization cell
 - compiling
    document options
      .pdf preferred for hw? must download file to local computer
    commands and their results are displayed
 - formatting

```
    blank line starts a new paragraph
    *word* italicizes
    **word** emboldens
    ### Heading    after blank line gives section heading
    ## Heading     after blank line gives subsection heading

Task for remainder of class:
 1. In your .Rmd file, load some non-Lock 5 datasets
     <nameYouChoose> <- read.csv( <full url, in quotation marks> )
    EXAMPLE:
      ssurv <- read.csv("http://scofield.site/teaching/data/csv/ssurv.csv")
    Other data files available at
      http://scofield.site/teaching/data/csv/index.html

 2. Choose from commands below, and make for yourself an example of its use.
    Whenever it makes sense, try these in univariate and bivariate settings.
     read.csv()
     head()
     help()
     dim()
     names()
     tally()
     addmargins()
     mean()
     median()
     sd()
     favstats()
     filter()
     gf_box()
       try replacing gf_box() with gf_percents() and note the difference
     gf_point()
     gf_histogram()
       try replacing with gf_dhistogram() and note the difference
     gf_density()
```