Complete the following tasks:

* Find a dataset that you find interesting. You can look at a variety of sources for data including:
  + Religion, health, etc.: [www.Thearda.com](http://www.Thearda.com)
  + Census, Education, Wealth, Time-Use, etc: [www.ipums.org](http://www.ipums.org)
  + Education: [www.nces.ed.gov](http://www.nces.ed.gov)
  + Mental Health: [www.cdc.gov/mentalhealth/data\_publications/index.htm](http://www.cdc.gov/mentalhealth/data_publications/index.htm)
  + Many Other Sources
* Once you have identified a dataset, explain how you accessed it. What links did you click? Was there any registration required? Did you download directly or was there an online system you navigated:
  + [https://www.census.gov](https://www.census.gov/)
  + Topics>Education> Education Attainment>Education Attainment in the United States:2017>Table 1. Education Attainment of the Population 18 years and Over, by Age, Sex, Race, and Hispanic Origin: 2017> All Races.
  + There was no registration needed and document was opened in excel. I deleted both genders and male to focus on female data. I reformatted table to take unnecessary spacing and imported to R from my computer using an excel document.
* In R:
  + Import the data
  + Provide a list of variables in the dataset
    - Copy and paste this list in your word document:
    - [1] "Age" "Total"
    - [3] "None" "1st - 4th grade"
    - [5] "5th - 6th grade" "7th - 8th grade"
    - [7] "9th grade" "10th grade"
    - [9] "11th grade2" "High school graduate"
    - [11] "Some college, no degree" "Associate's degree, occupational"
    - [13] "Associate's degree, academic" "Bachelor's degree"
    - [15] "Master's degree" "Professional degree"
    - [17] "Doctoral degree"
  + What is the structure of the data? Which variables are character, and which are numeric:
    - Copy and paste this list in your word document:
    - $ Age : chr ".18 years and over" "..18 to 24 years" ".25 years and over" "..25 to 29 years" ...
    - $ Total : num 127155 14559 112597 11335 10795 ...
    - $ None : num 410 29 380 5 22 33 28 25 36 41 ...
    - $ 1st - 4th grade : num 761 24 737 29 28 35 43 56 57 75 ...
    - $ 5th - 6th grade : num 1575 27 1548 37 94 ...
    - $ 7th - 8th grade : num 1800 59 1740 82 129 118 145 135 117 126 ...
    - $ 9th grade : num 1746 120 1626 110 157 ...
    - $ 10th grade : num 2056 271 1785 161 110 ...
    - $ 11th grade2 : num 4923 1536 3388 323 261 ...
    - $ High school graduate : num 35010 3794 31216 2678 2359 ...
    - $ Some college, no degree : num 24247 5721 18525 2126 1860 ...
    - $ Associate's degree, occupational: num 5399 364 5035 492 481 ...
    - $ Associate's degree, academic : num 8278 654 7624 832 734 ...
    - $ Bachelor's degree : num 26145 1788 24357 3264 2864 ...
    - $ Master's degree : num 11680 137 11543 969 1294 ...
    - $ Professional degree : num 1426 19 1407 110 167 ...
    - $ Doctoral degree : num 1700 16 1685 116 235 ...
  + Describe the data through Central Tendencies:
    - Give me the mean, median, and mode of **five variables**:
    - > mean(Education\_table$`High school graduate`)
    - [1] 7231
    - > mean(Education\_table$`Some college, no degree`)
    - [1] 4787
    - > mean(Education\_table$`Bachelor's degree`)
    - [1] 5474.714
    - > mean(Education\_table$`Master's degree`)
    - [1] 2493
    - > mean(Education\_table$`Doctoral degree`)
    - [1] 363.1429
    - > median(Education\_table$`High school graduate`)
    - [1] 2897
    - > median(Education\_table$`Some college, no degree`)
    - [1] 1792
    - > median(Education\_table$`Bachelor's degree`)
    - [1] 2353.5
    - > median(Education\_table$`Master's degree`)
    - [1] 1030.5
    - > median(Education\_table$`Doctoral degree`)
    - [1] 155
    - \*couldn’t get mode
  + Describe the data through variation:
    - Choose one variable and give me:
    - The variance, range, and standard deviation:
    - #variance:
    - var(Degree)
    - > #range
    - > max(Degree)-min(Degree)
    - [1] 1684
    - > #standard deviation:
    - > a <- c(1700, 16, 1685, 116, 235, 192, 200, 189, 170, 137, 140, 108, 79, 117)
    - > sd(a)
    - [1] 565.916
    - A histogram and either box-plot or stem-leaf plot
    - \*couldn’t make histogram/box-plot unsure?

**What to turn in**

* Publish your homework to GitHub under your user account, the appropriate format (check syllabus for instructions). This means turn in a separate document from your code that answers the questions above.
* Upload the code you used to GitHub.
* Provide me the links of these two files via e-mail no later than Thursday, September 20th at 6:00pm.