# Jodi Pafford – MSDS 6306 – Unit 5 – Codebook

This project took two files with popular children's names from 2015 and 2016 (source unknown). After the files were cleaned, they were merged together. The final step was to remove all Male Names and report only the top 10 girl names with both years of data totaled.

**Initial download of files needed to complete this analysis** (found in github site: https://github.com/jpafford/SMU_MSDS_Homework/tree/master/Week_5)

- yob2015.txt: Popular Children's names from children born in the USA in 2015
  - Name, Gender, Number of Children born with that name in 2015
  - File is comma (,) separated
- yob2016.txt: Popular Children's names from children born in the USA in 2015
  - Name, Gender, Number of Children born with that name in 2016
  - File is semi-colon (;) separated

**Description of data from the above txt files:**

- Name: Called-"First_Name"
  - Alphabetic characters only. Each line is unique
- Gender: Called-"Gender"
  - F = Female
  - M = Male
- Number of Children born with that name in201_: Called-"Number_of_Children"
  - Numeric characters only.

**List of data sets created in this project:**

- "df" – import of file, yob2016.txt.
- "names(df)" – named the columns, as listed above
- "y2015" – import of file yob2015.txt.
- "final" – merge of yob2015.txt and yob2016.txt. Merging occurred with the "First_Name" column. If there were any lines that did not match, they were removed from the file.
  - New columns were named:
    - "Gender.x" – yob2015.txt
    - "Number_of_Chilrden.x" – yob2015.txt
    - "Gender.y" – yob2016.txt
    - "Number_of_Children.y" – yob2016.txt
- "Total" column was added in order to sum the two "Number_of_Children" columns and then file was sorted by the highest "Total".
- "final_Girl" – file with all the 'M' (males) removed from the "Gender.x" column
- "final_10_Girl" – file with just the top 10 girl names and two columns:
  - "First_Name"
  - "Total": sum of both 2015 and 2016 files.
- Final CSV file created to display only the top 10 names with no row names is "Top 10 Girl Names 15-16".