Baby Names Part II - Unix Utilities and RegEx (56pts)
due: Nov 10, 2014 11:59pm

Important Notes

- Submission command: submit -c 246 -p **7** -d YourDirectory

- **This assignment is to be done on your own.** If you need help, see the instructor or TA.

- Please start the assignment as soon as possible and get your questions answered early.

- Read through this specification completely before you start.

- Some aspects of this specification are subject to change, in response to issues detected by students or the course staff.

Based on the baby names data files given in Part I, find the appropriate combinations of Unix utilty programs to answer the following questions. Only unix commands, pipes and redirections are allowed. No shell scripts or C programs. For each question, record what commands and the output you used to get the answer.

1. Find top 25 female names of 2012

2. Find bottom 25 male names of 2012

3. Find all female names used from 1900-2012, sorted alphabetically, with no repeats

4. How many distinct female names used from 1900-2012?

5. Find all unique male variations on the name "John"

6. Find all unique female variations on the name "John"

7. Find all unique variations on the name "John", male and female. Can you do this with one command line (i.e. just one return)? Can you do this without storing intermediate results?

8. Find all female/male names that start with the first letter of your name and end with the last letter of your name (1900-2012).

9. How many babies have the same name as you (pick a name to match if yours isn't in the files) in the last decade (2000-2009)?

10. Find the total number of female babies contributing to these statistics files (1900-2012) (you might want to check the Unix utility gawk)

11. Find the total number of unique male names (1900-2012).

12. Find top 25 males names used from 1900-2012. Can this be done easily with Unix utility programs alone? Why and why not?

13. Find the longest name (1900-2012).

14. Find the longest male name (1900-2012).

What to hand in: Electronic text file consisting of Unix commands (the ones you used to obtain the results) and the corresponding output for each problem.