Lab 09 Option 1:
Baby Names Part II - Unix Utilities and RegEx

Important Notes

- Submission location (Mercurial repository): /rd/cs246s2016/$USER/lab09

- **This assignment can be done in pairs.** If you need help, see the instructor or TA.

- 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 Assignment 6, Baby Names 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. If you do this option, then you must also do option 1 for lab10.

1. Find top 25 female names of 2014

2. Find bottom 25 male names of 2014

3. Find all female names used from 1900-2014, sorted alphabetically, with no repeats

4. How many distinct female names used from 1900-2014?

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 (1900-2014).

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 (1900-2014) (you might want to check the Unix utility gawk)

11. Find the total number of unique male names (1900-2014).

12. Find top 25 males names used from 1900-2014. Can this be done easily with Unix utility programs alone? Why and why not?

13. Find the longest name (1900-2014).

14. Find the longest male name (1900-2014).

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