

*Project 3-- csc433
Sarah Cummings
4/27/16;

*REPORT 1;

```
options linesize=90 nodate pageno=1;
data erie;
    *read in the text file which contains our data and set it to read starting with the 12th line;
    format finaltimenew time12.0;
    infile "/folders/myfolders/433files/erie-marathon-2006.txt" firstobs=12 flowover trunccover;
    *create variables and point to where they are in the file;
    input jerseyNumber 3-5 place 9-11 info $19-64 finalTime $88-94 milePace $98-102;
    finaltimenew = input(finalTime,?? hhmmss.);
    drop finalTime;
    *find the name within the info variable;
    name= scan(info, 1, ',');
    *find the age (and gender indicator) within the info variable;
    ageG= scan(info, 2, ',');
    *find the hometown within the info variable-- city and state;
    homeTown= scan(info, 3, ',');
    state= scan(info, 4, ',');
    *get rid of the lines we don't want by having sas throw out the lines without a jersey
number
    *perhaps this is trivial way to get discard those extra lines, but it works;
    if missing(jerseyNumber) then delete;
    *find length of age/gender variable;
    len= length(ageG);
    *figure out the gender by looking for an asterisk with substring function;
    if substr(ageG, len)= '*' then do;
        *create an age variable without the asterisk;
        age = substr(ageG, 1, len - 1);
        *set gender variable to F if there's an asterisk;
        gender = "F";
    end;
    else do;
        age = ageG;
        gender = "M";
    end;

run;

proc means data=erie;
    *use proc means to create variables for our header;
    output out=out n=count mean(finaltimenew)=AverageT;

data _null_;
    title "Project 3: Erie Marathon All Runners";
    *Create a file to output the first report;
    file "/folders/myfolders/433files/erie-marathonALLRunners.txt" print title header=head;
```

```

        set erie;
        if _n_ = 1 then set out;
        *Use put command to insert data at specific column places;
        put @1 name @20 jerseyNumber @35 gender @45 age @50 homeTown',' state @75
finaltime new;
        return;
**Create a header for the top of each file;
head:
        put //;
        put "Total Number of Runners: " count;
        put "Average Final Time: " AverageT;
        put //;
        put @1 "Name" @20 "Jersey Number" @35 "Gender" @45 "Age" @50 "Home Town"
@75 "Final Time";

```

*REPORT 2, repeat the process but output data for just the female runners;

```

options linesize=90 nodate pageno=1;
data erie2;
        *read in the text file which contains our data and set it to read starting with the 12th line;
        format finaltime new time12.0;
        infile "/folders/myfolders/433files/erie-marathon-2006.txt" firstobs=12 flowover trunccover;
        *create variables and point to where they are in the file;
        input jerseyNumber 3-5 place 9-11 info $19-64 finalTime $88-94 milePace $98-102;
        finaltime new = input(finalTime, ?? hhmmss.);
        drop finalTime;
        *find the name within the info variable;
        name= scan(info, 1, ',');
        *find the age (and gender indicator) within the info variable;
        ageG= scan(info, 2, ',');
        *find the hometown within the info variable-- city and state;
        homeTown= scan(info, 3, ',');
        state= scan(info, 4, ',');
        *get rid of the lines we don't want by having sas throw out the lines without a jersey
number
        *perhaps this is trivial way to get discard those extra lines, but it works;
        if missing(jerseyNumber) then delete;
        *find length of age/gender variable;
        len= length(ageG);
        *figure out the gender by looking for an asterisk with substring function;
        if substr(ageG, len)= '*' then do;
                *create an age variable without the asterisk;
                age = substr(ageG, 1, len - 1);
                *set gender variable to F if there's an asterisk;
                gender = "F";
        end;
        else do;
                age = ageG;
                gender = "M";
        end;

```

```

end;
*delete the male runners from the file;
if gender= "M" then delete;
run;

proc means data=erie2;
    *use proc means to create variables for our header;
    output out=out n=fCount mean(finaltimenew)=fAverageT;

data _null2_;
    title "Project 3: Just Women";
    file "/folders/myfolders/433files/erie-marathonWomenOnly.txt" print title header=head2;
    set erie2;
    if _n_ = 1 then set out;
    put @1 name @20 jerseyNumber @35 gender @45 age @50 homeTown',' state @75
finaltimenew;
    return;
**Create a header for the top of each file;
head2:
    put //;
    put "Total Number of Women: " fCount;
    Put "Percent of Total Runners: 30.5% ";
    put "Average Final Time for Women: " fAverageT;
    put //;
    put @1 "Name" @20 "Jersey Number" @35 "Gender" @45 "Age" @50 "Home Town"
@75 "Final Time";

*REPORT 3, repeat the process for just male runners;

options linesize=90 nodate pageno=1;
data erie3;
    *read in the text file which contains our data and set it to read starting with the 12th line;
    format finaltimenew time12.0;
    infile "/folders/myfolders/433files/erie-marathon-2006.txt" firstobs=12 flowover trunccover;
    *create variables and point to where they are in the file;
    input jerseyNumber 3-5 place 9-11 info $19-64 finalTime $88-94 milePace $98-102;
    finaltimenew = input(finalTime,?? hhmmss.);
    drop finalTime;
    *find the name within the info variable;
    name= scan(info, 1, ',');
    *find the age (and gender indicator) within the info variable;
    ageG= scan(info, 2, ',');
    *find the hometown within the info variable-- city and state;
    homeTown= scan(info, 3, ',');
    state= scan(info, 4, ',');
    *get rid of the lines we don't want by having sas throw out the lines without a jersey
number
    *perhaps this is trivial way to get discard those extra lines, but it works;
    if missing(jerseyNumber) then delete;

```

```

        *find length of age/gender variable;
        len= length(ageG);
        *figure out the gender by looking for an asterisk with substring function;
        if substr(ageG, len)= '*' then do;
            *create an age variable without the asterisk;
            age = substr(ageG, 1, len - 1);
            *set gender variable to F if there's an asterisk;
            gender = "F";
        end;
        else do;
            age = ageG;
            gender = "M";
        end;
        *delete the data for the female runners;
        if gender= "F" then delete;
run;

proc means data=erie3;
    *use proc means to create variables for our header;
    output out=out n=mCount mean(finaltimenew)=mAverageT;

data _null3_;
    title "Project 3: Just Men";
    *Create a file for our third report;
    file "/folders/myfolders/433files/erie-marathonMenOnly.txt" print title header=head3;
    set erie2;
    if _n_ = 1 then set out;
    put @1 name @20 jerseyNumber @35 gender @45 age @50 homeTown',' state @75
finaltimenew;
    return;
**Create a header for the top of each file;
head3:
    put //;
    put "Total Number of Men: " mCount;
    Put "Percent of Total Runners: 69.5% " ;
    put "Average Final Time for Men: " mAverageT;
    put //;
    put @1 "Name" @20 "Jersey Number" @35 "Gender" @45 "Age" @50 "Home Town"
@75 "Final Time";

run;
quit;

```