**Class 1 Homework Answers**

Question 1: How many variables are included in this data set?

Three variables are included in this dataset namely “year”, “boys” and “girls”.

Question 2: What years are included in this dataset?

The dataset includes information from the year 1940 to 2002.

Question 3: What command would you use to view just the counts of girls born each year?

‘present$girls’ command can be used to view the counts of girls for each year.

Question 4: Is there an apparent trend in the number of girls born over the years? How would you describe it?

Approximately, from the line graph we can see that after every 4-5 years there is a slight dip in the number of girls born and then again it increases. It looks a like a sinusoidal pattern.

Question 5: Check out the help file of which.max() function, and find out in what year did we see the most total number of births in the U.S.?

In the year 1961, we saw highest number of total births i.e 4268326 (index 22).

Question 6a: Now, make a plot of the proportion of boys over time, and based on the plot determine if the following statement is true or false: The proportion of boys born in the US has decreased over time.

Yes, from the plot we can see that proportion of boys has decreased over the years.

Question 6b: Do you think the trend you saw in question 6a is a meaningful one or just represent random variation (or noise in the data)?

Although, from the plot we can see the decrease in proportion of boys over the years but there is not much significance difference between the values in the years of 1940 and 2000 respectively. So, it’s not of much importance from the analysis point of view.

Question 7: How many TRUEs did you see after typing the previous expression?

63 true values are displayed which for the all the years(1940-2002) boys outnumbered girls.

Question 8: How does the output for this code compare to your answer for Question 7? What does this code do?

In the previous expression it just prints TRUE for all 63 rows, by using sum() function it actually treats TRUE as 1 and sums it up to 63.

Question 9 [MULTIPLE CHOICE]: Make a plot that displays the boy-to-girl ratio for every year in the data set. What do you see?

* There appears to be no trend in the boy-to-girl ratio from 1940 to 2002. - False
* There is initially an increase in boy-to-girl ratio, which peaks around 1960. After 1960 there is a decrease in the boy-to-girl ratio, but the number begins to increase in the mid 1970s. – False
* There is initially a decrease in the boy-to-girl ratio, and then an increase between 1960 and 1970, followed by a decrease.- True
* The boy-to-girl ratio has increased over time. - False
* There is an initial decrease in the boy-to-girl ratio born but this number appears to level around 1960 and remain constant since then – False

Question 10: Calculate the absolute differences between number of boys and girls born in each year, and determine which year out of the present data had the biggest absolute difference in the number of girls and number of boys born?

The year of 1963 has the largest absolute difference in the number of girls and number of boys born.