SEIS631

Dan Ward

Assignment 1

**Question 1: How many variables are included in this data set? What type of variables are each of these (i.e. categorical ordinal, numeric discrete, numeric continuous, etc.)**

There are 3 variables included in this data set.

Variable 1: Year – Categorical Ordinal

Variable 2: Boys Born – Numerical Continuous

Variable 3: Girls Born – Numerical Continuous

**Question 2: What years are included in this dataset? (Note: can list as a range (ex. 1990 – 2402))**

Years 1940 – 2002

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

Present$girls

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

There are some apparent trends. Between 1940 and 1960 there was a rapid increase in the number of girls born, between 1960 and 1975 there was a significant decrease in girls born, followed by a gradual increase in girls born between 1975 and 2002.

**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.?**

Which.max(present$total) yields 22. 22 is the id for 1961 when 4268326 births took place.

**Question 6: 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, it is true that the proportion of boys born in the US has decreased over time. While the proportion of boys over time has increased and decreased over time between 1940 and 2002, the general overall trend is decreasing.

**Question 7: How many TRUE did you see after typing the previous expression? Hint: use sum() function.**

There are 63 TRUE values

**Question 8 [MULTIPLE CHOICE]: Make a plot that displays the boy-to-girl ratio for every year in the data set. What do you see?**

(c) There is initially a decrease in the boy-to-girl ratio, and then an increase between 1960 and 1970, followed by a decrease.

**Question 9: Calculate 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?**

Which.max(abs(present$boys-present$girls)) yields 24 which is the id for 1963 which had a difference of 105,244.

#returns the year of 1963 which is the year that the max difference occurs

present$year[which.max(abs(present$boys-present$girls))]