Lab 0

Mu Lu

2015-09-21

The Dataframe

Read in your data set and assign to the object:

```
present <- read.table("http://assets.datacamp.com/course/dasi/present.txt")
```

`present` will print your data set in the console:

```
##    year    boys   girls
## 1  1940 1211684 1148715
## 2  1941 1289734 1223693
## 3  1942 1444365 1364631
## 4  1943 1508959 1427901
## 5  1944 1435301 1359499
## 6  1945 1404587 1330869
## 7  1946 1691220 1597452
## 8  1947 1899876 1800064
## 9  1948 1813852 1721216
## 10 1949 1826352 1733177
## 11 1950 1823555 1730594
## 12 1951 1923020 1827830
## 13 1952 1971262 1875724
## 14 1953 2001798 1900322
## 15 1954 2059068 1958294
## 16 1955 2073719 1973576
## 17 1956 2133588 2029502
## 18 1957 2179960 2074824
## 19 1958 2152546 2051266
## 20 1959 2173638 2071158
## 21 1960 2179708 2078142
## 22 1961 2186274 2082052
## 23 1962 2132466 2034896
## 24 1963 2101632 1996388
## 25 1964 2060162 1967328
## 26 1965 1927054 1833304
## 27 1966 1845862 1760412
## 28 1967 1803388 1717571
## 29 1968 1796326 1705238
## 30 1969 1846572 1753634
## 31 1970 1915378 1816008
## 32 1971 1822910 1733060
## 33 1972 1669927 1588484
## 34 1973 1608326 1528639
## 35 1974 1622114 1537844
## 36 1975 1613135 1531063
## 37 1976 1624436 1543352
```

```
## 38 1977 1705916 1620716
## 39 1978 1709394 1623885
## 40 1979 1791267 1703131
## 41 1980 1852616 1759642
## 42 1981 1860272 1768966
## 43 1982 1885676 1794861
## 44 1983 1865553 1773380
## 45 1984 1879490 1789651
## 46 1985 1927983 1832578
## 47 1986 1924868 1831679
## 48 1987 1951153 1858241
## 49 1988 2002424 1907086
## 50 1989 2069490 1971468
## 51 1990 2129495 2028717
## 52 1991 2101518 2009389
## 53 1992 2082097 1982917
## 54 1993 2048861 1951379
## 55 1994 2022589 1930178
## 56 1995 1996355 1903234
## 57 1996 1990480 1901014
## 58 1997 1985596 1895298
## 59 1998 2016205 1925348
## 60 1999 2026854 1932563
## 61 2000 2076969 1981845
## 62 2001 2057922 1968011
## 63 2002 2057979 1963747
```

Print the number of rows and variables:

```
dim(present)
```

```
## [1] 63  3
```

Print the names of the variables of the data frame:

```
names(present)
```

```
## [1] "year"  "boys"  "girls"
```

We can tell there are 3 vaibales (year,boys,girls) in the present data set.

Find the number of boys and girls born each year, and assign answers to variables:

```
num_boys <-present$boys
num_girls <- present$girls
num_boys
```

```
##  [1] 1211684 1289734 1444365 1508959 1435301 1404587 1691220 1899876
##  [9] 1813852 1826352 1823555 1923020 1971262 2001798 2059068 2073719
## [17] 2133588 2179960 2152546 2173638 2179708 2186274 2132466 2101632
## [25] 2060162 1927054 1845862 1803388 1796326 1846572 1915378 1822910
## [33] 1669927 1608326 1622114 1613135 1624436 1705916 1709394 1791267
```

```
## [41] 1852616 1860272 1885676 1865553 1879490 1927983 1924868 1951153
## [49] 2002424 2069490 2129495 2101518 2082097 2048861 2022589 1996355
## [57] 1990480 1985596 2016205 2026854 2076969 2057922 2057979
```

```
num_girls
```

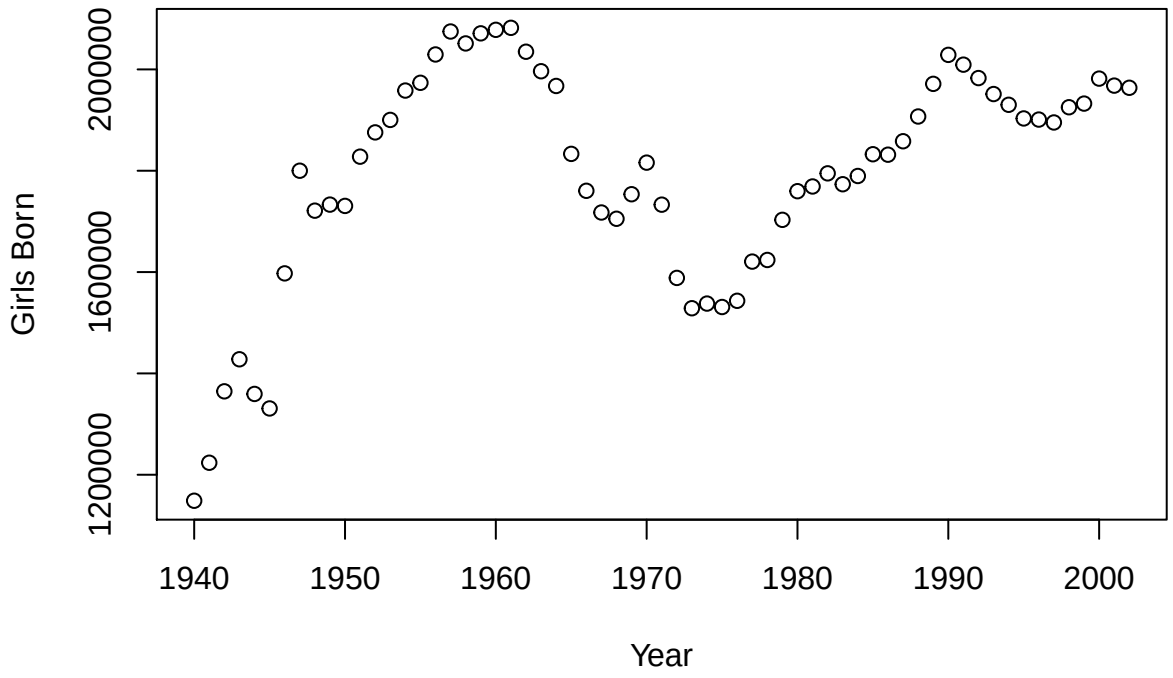
```
##  [1] 1148715 1223693 1364631 1427901 1359499 1330869 1597452 1800064
##  [9] 1721216 1733177 1730594 1827830 1875724 1900322 1958294 1973576
## [17] 2029502 2074824 2051266 2071158 2078142 2082052 2034896 1996388
## [25] 1967328 1833304 1760412 1717571 1705238 1753634 1816008 1733060
## [33] 1588484 1528639 1537844 1531063 1543352 1620716 1623885 1703131
## [41] 1759642 1768966 1794861 1773380 1789651 1832578 1831679 1858241
## [49] 1907086 1971468 2028717 2009389 1982917 1951379 1930178 1903234
## [57] 1901014 1895298 1925348 1932563 1981845 1968011 1963747
```

The result is in data type **vector**.

Arithmetic calculation on dataframe and plottting

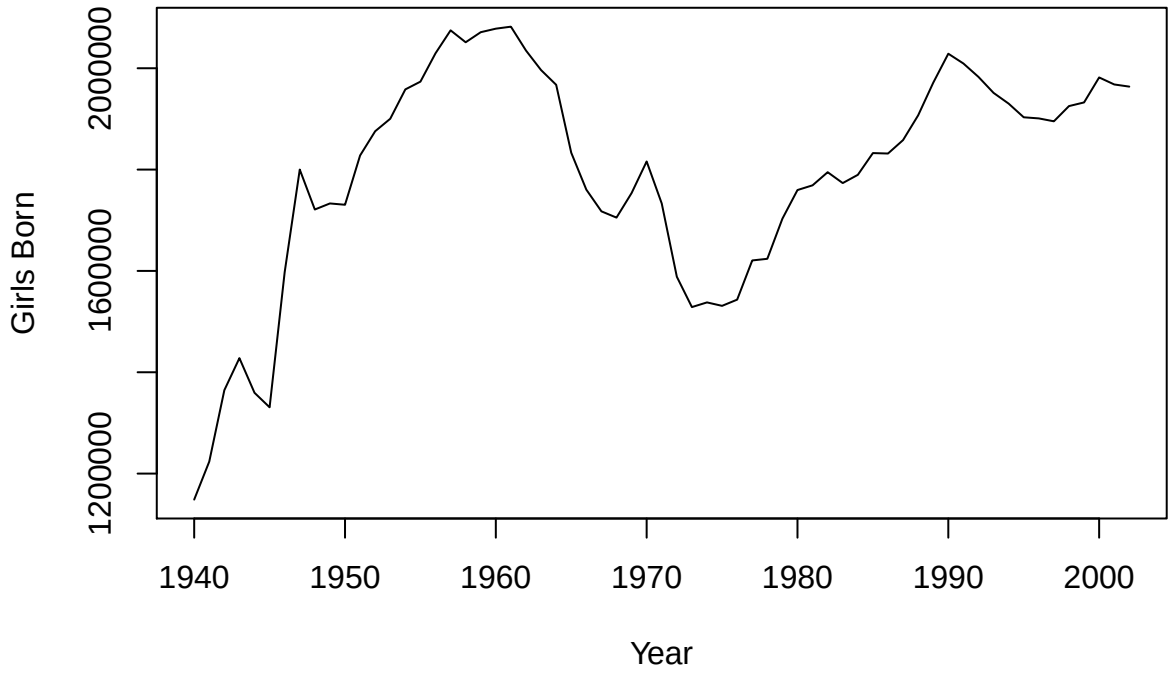
Plot the data:

```
plot(x=present$year,y=present$girls,xlab = "Year" ,ylab = "Girls Born")
```



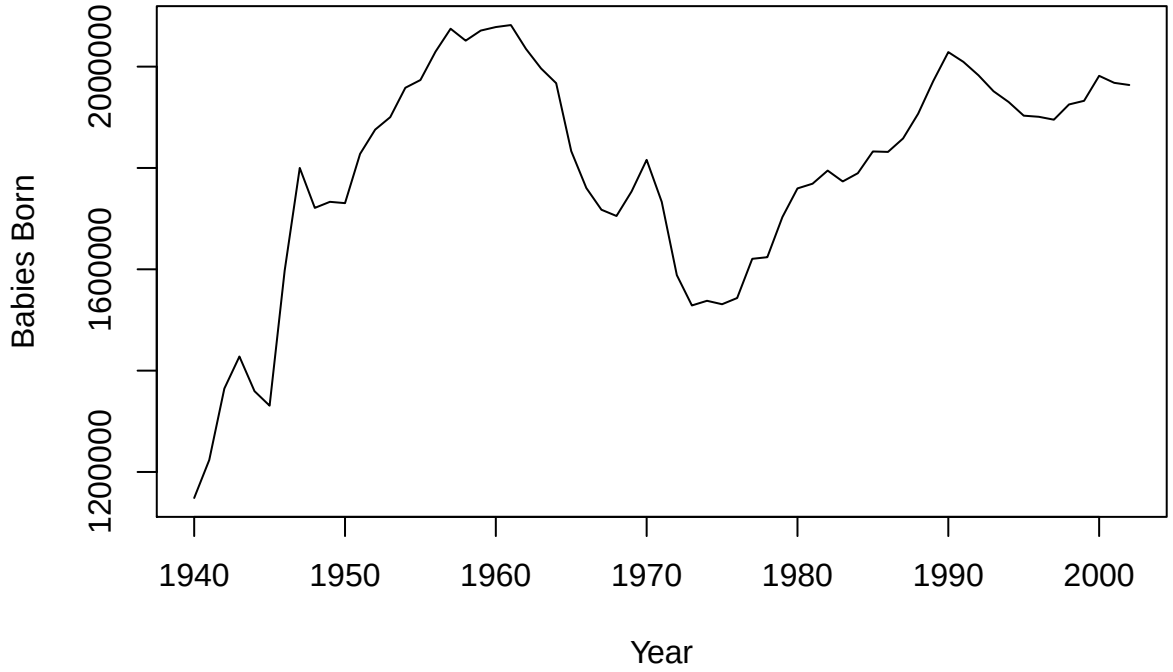
Line plot:

```
plot(x=present$year,y=present$girls,xlab = "Year" ,ylab = "Girls Born", type = "l")
```



Plotting the total number of new born babies by year:

```
babies <- present$boys + present$girls
plot(x=present$year,y=present$girls,xlab = "Year" ,ylab = "Babies Born", type = "l")
```



Compute the ratio of the number of boys divided by the number of girls born:

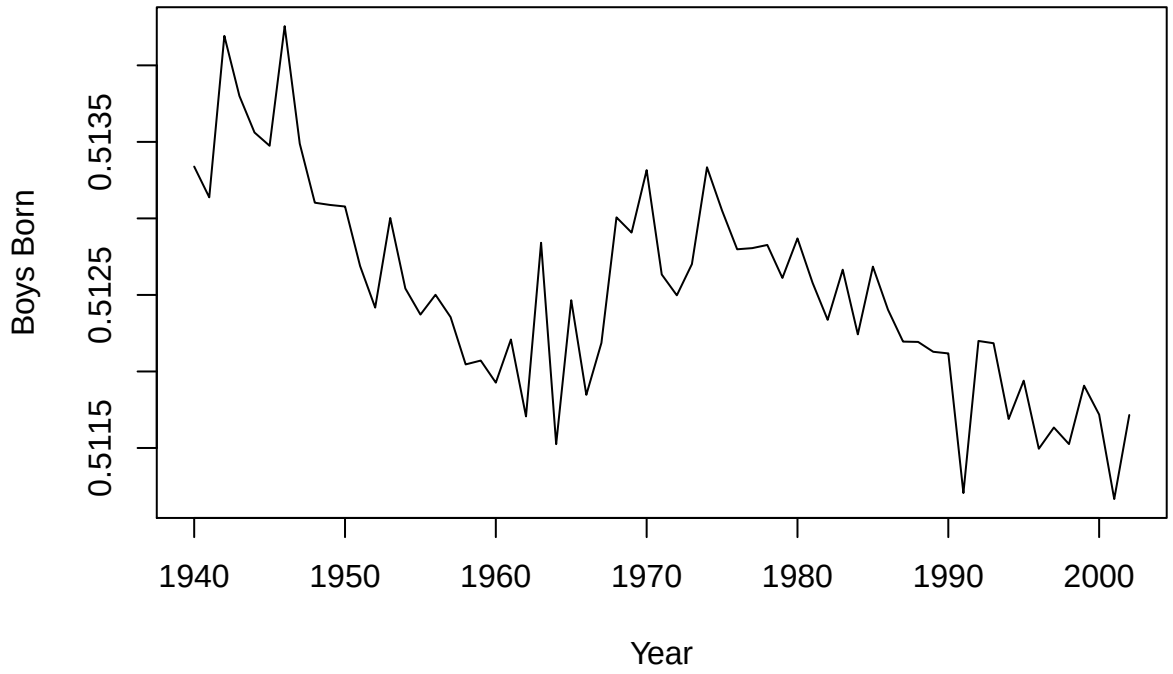
```
present$boys/present$girls
```

```
##  [1] 1.054817 1.053969 1.058429 1.056767 1.055757 1.055391 1.058698
```

```
##  [8] 1.055449 1.053820 1.053760 1.053716 1.052078 1.050934 1.053399
## [15] 1.051460 1.050742 1.051286 1.050672 1.049374 1.049480 1.048873
## [22] 1.050057 1.047948 1.052717 1.047188 1.051137 1.048540 1.049964
## [29] 1.053417 1.052997 1.054719 1.051845 1.051271 1.052129 1.054797
## [36] 1.053605 1.052538 1.052569 1.052657 1.051749 1.052837 1.051615
## [43] 1.050597 1.051976 1.050199 1.052061 1.050876 1.050000 1.049991
## [50] 1.049720 1.049676 1.045849 1.050017 1.049955 1.047877 1.048928
## [57] 1.047062 1.047643 1.047190 1.048791 1.047998 1.045686 1.047986
```

Calculate the proportion of male newborns for all years and plot it:

```
plot(x=present$year,y=present$boys/(present$boys + present$girls),xlab = "Year" ,ylab = "Boys Born", ty
```



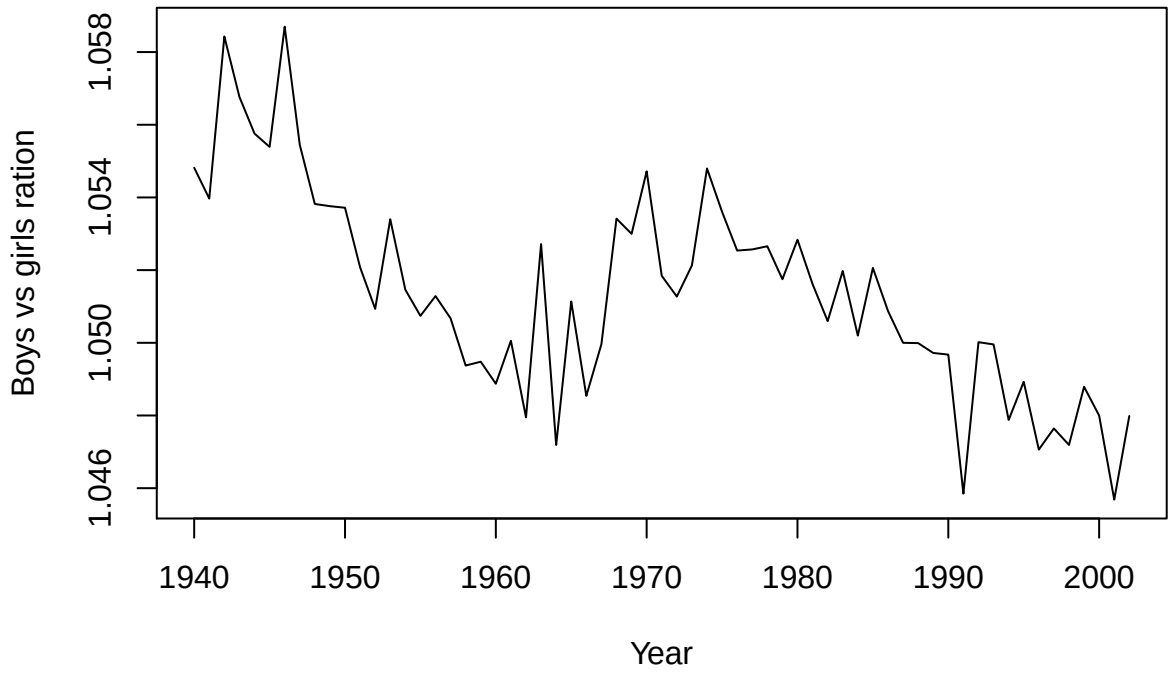
In addition to simple mathematical operators like subtraction and division, you can ask R to make comparisons: greater than with `>`, less than with `<`, and equality with `==`. We can ask if boys outnumber girls in each year:

```
present$boys > present$girls
```

```
##  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [15] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [29] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [43] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [57] TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

Plotting boys-to-girls ration for every year:

```
plot(x=present$year,y=present$boys/present$girls,xlab = "Year" ,ylab = "Boys vs girls ration", type = "
```



Plotting absolute differences between number of boys and girls born in each year:

```
plot(x=present$year,y=abs(present$boys-present$girls),xlab = "Year" ,ylab = "Boys/girls diff", type = "
```