STAT40620 Assignment 1

Brian Buckley 14203480

Brian Buckley 14203480

STAT40730 Assignment 1

Task 1: manipulation

1 Set working directory to the folder that contains the names data and read in a file

2 Answer some specific questions about this data set

(a) How many children were born in 1950?

- According to these data there were **10301** children born in 1950

(b) Which were the 10 most popular names for each sex?

- The 10 most popular female names in 1950 were **Linda, Mary, Patricia, Barbara, Susan, Nancy, Deborah, Sandra, Carol, Kathleen**
- The 10 most popular male names in 1950 were **James, Robert, John, Michael, David, William, Richard, Thomas, Charles, Gary**

(c) Are there any names in the data set that were only given once?

```
## [1] "There are no names given less than 5 times"
```

3 Which female and male name had the biggest rise/fall compared with 1950

- I interpret this question as the name that had the most/least difference from their position in 1950 to their position in 2010

Which female name had the biggest rise/fall

- The biggest female rise was **Kathleen**
- The biggest female fall was **Linda**

Which male name had the biggest rise/fall

- The biggest male rise was **Charles**
- The biggest male fall was **Robert**

Task 2: analysis

1 What year was the maximum for your name?

- **1972** was the year my name, **Brian** was most popular

2 Do males or females tend to have higher birth rates?

- Female Summary Statistics

```
##       year          number
##  Min.   :1880   Min.   :  90994
##  1st Qu.:1913   1st Qu.: 607795
##  Median :1946   Median :1426197
##  Mean   :1946   Mean   :1225439
##  3rd Qu.:1978   3rd Qu.:1769403
##  Max.   :2011   Max.   :2044048
```

- Male Summary Statistics

```
##       year          number
##  Min.   :1880   Min.   : 100747
##  1st Qu.:1913   1st Qu.: 491870
##  Median :1946   Median :1492673
##  Mean   :1946   Mean   :1244912
##  3rd Qu.:1978   3rd Qu.:1904012
##  Max.   :2011   Max.   :2155583
```

- Male birth rates have both a higher mean and median than female birth rates so overall males tend to have higher birth rates
- Specifically females have an approximately 98% mean and 96% median birth rate compared with males over the whole data set

3 Which last letter(s) stand out as having the biggest increase/decrease?

- The letter **a** has the biggest increase for **females**.
- The letter **q** has the biggest decrease for **females**.
- The letter **n** has the biggest increase for **males**.
- The letter **q** has the biggest decrease for **males**.

4 Which are the most popular palindromic names?

	name	number
26	Anna	859618
70	Hannah	388654
38	Ava	155700
49	Bob	93135
23	Ana	92054
3	Ada	89557

4.1 Are such names on the increase?

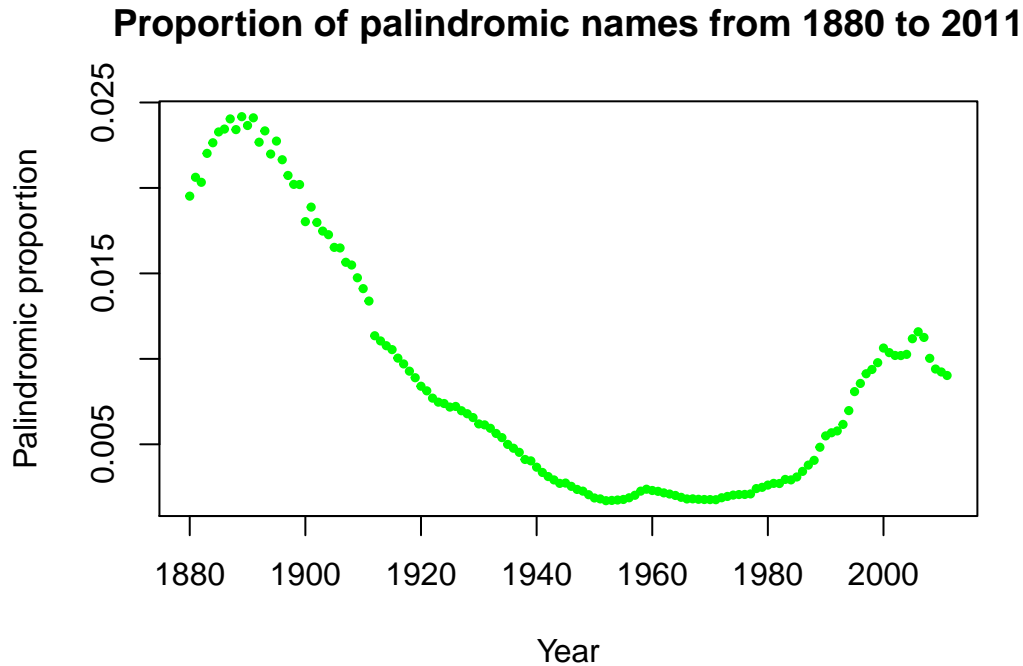- The plot of the proportion of palindromic names over time (Figure 1) shows that such names are **not** on the increase

Proportion of palindromic names from 1880 to 2011

Figure 1:

Task 3: creativity

1 Plot female and male birth rates to get a finer level of granularity than summary data

- From the plot in Figure 2 we see that prior to 1940 females tended to have higher birth rates but after 1940 males tended to have higher birth rates.
- The dramatic rise in both birth rates from ~1940 to ~1960 corresponds to the 'baby bommer years' following World War II

2 Plot my name's popularity

- The plot in Figure 3 shows that my name was very unpopular until ~1940 when it grew significantly until the early 1970s when it declined markedly in popularity. However, at the end of the data set (2011) it was still as popular as in around 1950. These data is therefore skewed to the left.
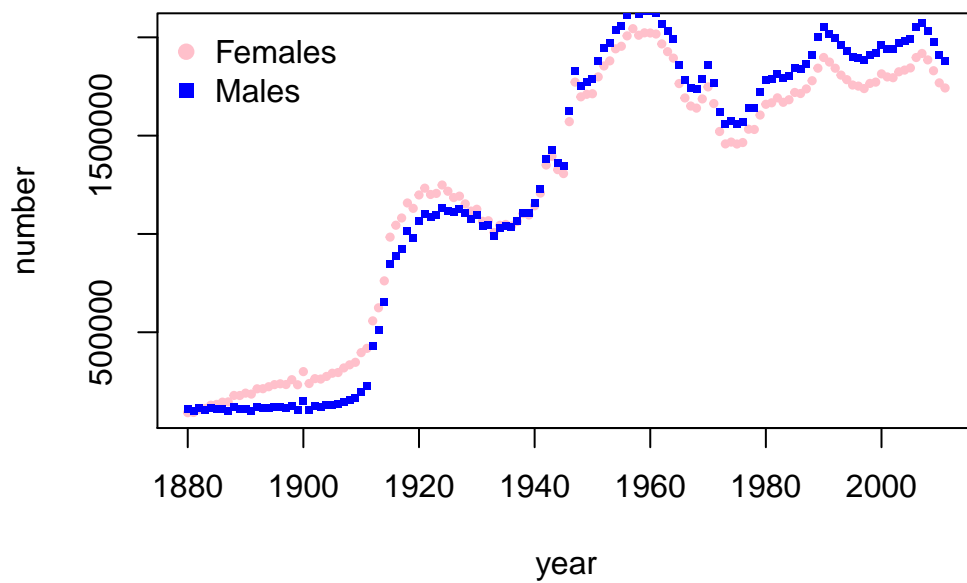
Birth rates of US females and males names from 1880 to 2(

Figure 2:

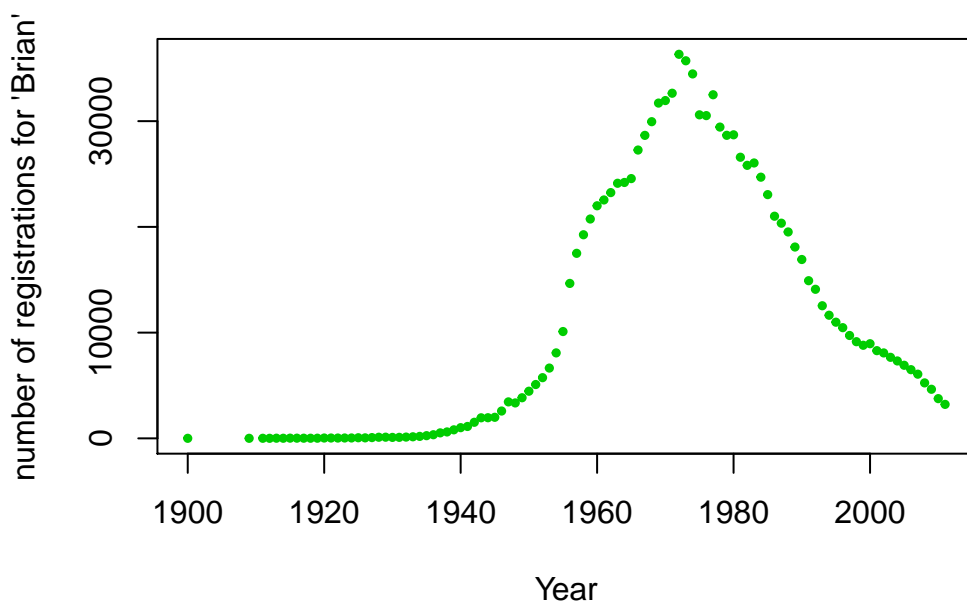**Popularity of the name 'Brian' from 1900 to 2011**

Figure 3:

4