Assignment 4

Chi-Yu, Chen

2018/7/28

Load worldrecord.csv and answer the following?

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
setwd("/Users/ChiYuChen/Data Analysis, Intro/Assignment 4")
worldrecords <- read.csv("WorldRecords.csv")
```

1. How many different types of events (e.g. "Mens 100m", "Womens Shotput" etc) are represented in the dataset

```
q1.1 <- group_by(worldrecords, Event)
nrow(summarise(q1.1))
```

```
## [1] 10
```

2. In what year did Usain Bolt first break the world record for the Men's 100m?

```
q2.1 <- filter(worldrecords, Event=="Mens 100m", Athlete=="Usain Bolt")
q2.1[1,]$Year
```

```
## [1] 2008
```

3. Which variable tells us the record setting time or distance? The variable name in the data set is? What type of the variable is this?

```
typeof(worldrecords$Type)
```

```
## [1] "integer"
```

```
class(worldrecords$Type)
```

```
## [1] "factor"
```

4. Create a subset of the dataset that contains only the world record cases for men's shotput and women's shotput

```
q4 <- filter(worldrecords, Event=="Mens Shotput" | Event=="Womens Shotput")
q4
```

```
##             Event     Type Record                       Athlete
## 1    Mens Shotput distance  17.68      \xcaCharlie Fonville\xca
## 2    Mens Shotput distance  17.79             \xcaJim Fuchs\xca
## 3    Mens Shotput distance  17.82             \xcaJim Fuchs\xca
## 4    Mens Shotput distance  17.90             \xcaJim Fuchs\xca
## 5    Mens Shotput distance  17.95             \xcaJim Fuchs\xca
## 6    Mens Shotput distance  18.00         \xcaParry O'Brien\xca
## 7    Mens Shotput distance  18.04         \xcaParry O'Brien\xca
## 8    Mens Shotput distance  18.42         \xcaParry O'Brien\xca
## 9    Mens Shotput distance  18.43         \xcaParry O'Brien\xca
## 10   Mens Shotput distance  18.54         \xcaParry O'Brien\xca
## 11   Mens Shotput distance  18.62         \xcaParry O'Brien\xca
## 12   Mens Shotput distance  18.69         \xcaParry O'Brien\xca
## 13   Mens Shotput distance  19.06         \xcaParry O'Brien\xca
## 14   Mens Shotput distance  19.25         \xcaParry O'Brien\xca
## 15   Mens Shotput distance  19.25           \xcaDallas Long\xca
## 16   Mens Shotput distance  19.30         \xcaParry O'Brien\xca
## 17   Mens Shotput distance  19.38           \xcaDallas Long\xca
## 18   Mens Shotput distance  19.45           \xcaBill Nieder\xca
## 19   Mens Shotput distance  19.67           \xcaDallas Long\xca
## 20   Mens Shotput distance  19.99           \xcaBill Nieder\xca
## 21   Mens Shotput distance  20.06           \xcaBill Nieder\xca
## 22   Mens Shotput distance  20.08           \xcaDallas Long\xca
## 23   Mens Shotput distance  20.10           \xcaDallas Long\xca
## 24   Mens Shotput distance  20.20           \xcaDallas Long\xca
## 25   Mens Shotput distance  20.68           \xcaDallas Long\xca
## 26   Mens Shotput distance  21.52          \xcaRandy Matson\xca
## 27   Mens Shotput distance  21.78          \xcaRandy Matson\xca
## 28   Mens Shotput distance  21.82           \xcaAl Feuerbach\xca
## 29   Mens Shotput distance  21.85        \xcaTerry Albritton\xca
## 30   Mens Shotput distance  22.00 \xcaAleksandr Baryshnikov\xca
## 31   Mens Shotput distance  22.15             \xcaUdo Beyer\xca
## 32   Mens Shotput distance  22.22             \xcaUdo Beyer\xca
## 33   Mens Shotput distance  22.62        \xcaUlf Timmermann\xca
## 34   Mens Shotput distance  22.64             \xcaUdo Beyer\xca
## 35   Mens Shotput distance  22.72     \xcaAlessandro Andrei\xca
## 36   Mens Shotput distance  22.84     \xcaAlessandro Andrei\xca
## 37   Mens Shotput distance  22.91     \xcaAlessandro Andrei\xca
## 38   Mens Shotput distance  23.06        \xcaUlf Timmermann\xca
## 39   Mens Shotput distance  23.12          \xcaRandy Barnes\xca
## 40 Womens Shotput distance  14.59    \xcaTatyana Sevryukova\xca
## 41 Womens Shotput distance  14.86     \xcaKlavdia Tochonova\xca
## 42 Womens Shotput distance  15.02         \xcaAnna Andreeva\xca
## 43 Womens Shotput distance  15.28         \xcaGalina Zybina\xca
## 44 Womens Shotput distance  15.37         \xcaGalina Zybina\xca
## 45 Womens Shotput distance  15.42         \xcaGalina Zybina\xca
## 46 Womens Shotput distance  16.20         \xcaGalina Zybina\xca
```

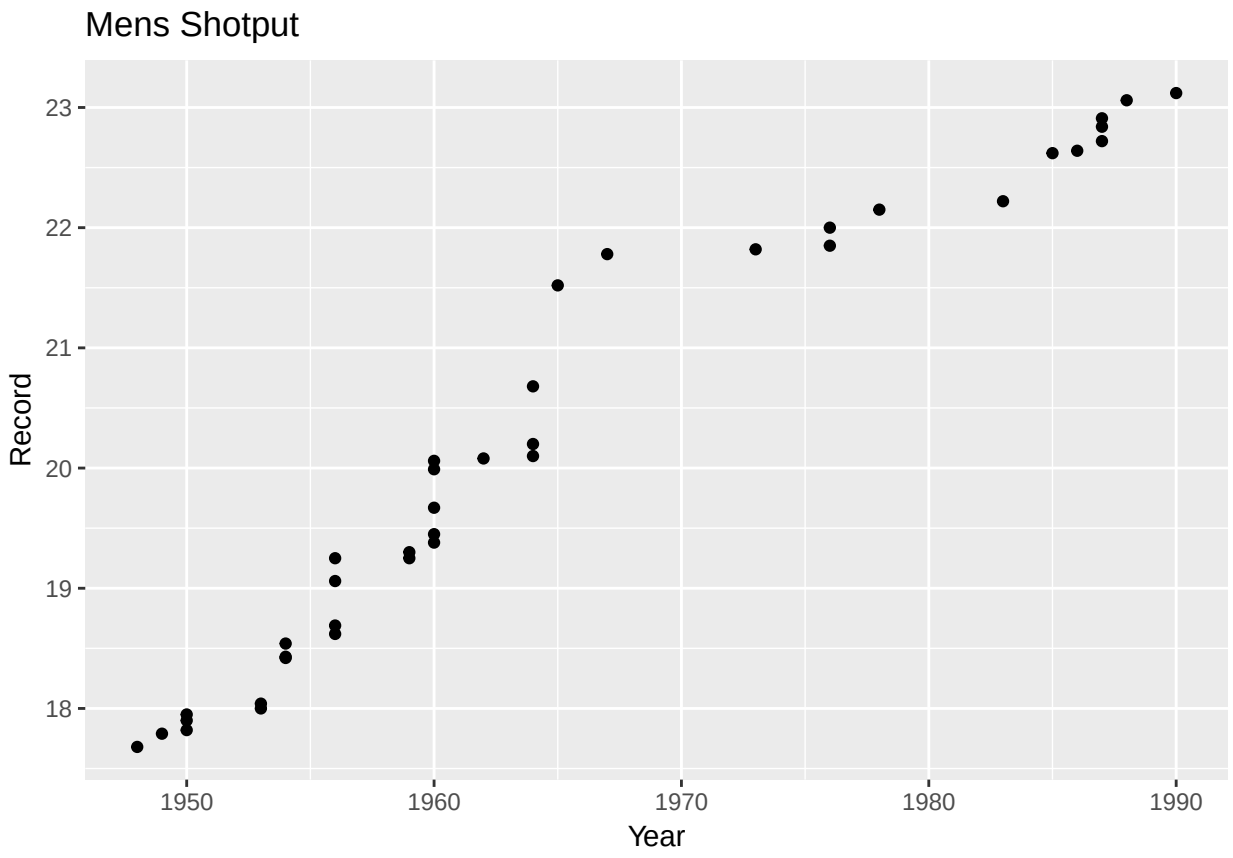
```
## 47 Womens Shotput distance  16.28         \xcaGalina Zybina\xca
## 48 Womens Shotput distance  16.28         \xcaGalina Zybina\xca
## 49 Womens Shotput distance  16.67         \xcaGalina Zybina\xca
## 50 Womens Shotput distance  16.76         \xcaGalina Zybina\xca
## 51 Womens Shotput distance  17.25          \xcaTamara Press\xca
## 52 Womens Shotput distance  17.42          \xcaTamara Press\xca
## 53 Womens Shotput distance  17.78          \xcaTamara Press\xca
## 54 Womens Shotput distance  18.55          \xcaTamara Press\xca
## 55 Womens Shotput distance  18.55          \xcaTamara Press\xca
## 56 Womens Shotput distance  18.59          \xcaTamara Press\xca
## 57 Womens Shotput distance  18.67     \xcaNadezhda Chizhova\xca
## 58 Womens Shotput distance  18.87       \xcaMargitta Gummel\xca
## 59 Womens Shotput distance  19.07       \xcaMargitta Gummel\xca
## 60 Womens Shotput distance  19.61       \xcaMargitta Gummel\xca
## 61 Womens Shotput distance  19.72     \xcaNadezhda Chizhova\xca
## 62 Womens Shotput distance  20.09     \xcaNadezhda Chizhova\xca
## 63 Womens Shotput distance  20.10       \xcaMargitta Gummel\xca
## 64 Womens Shotput distance  20.10     \xcaNadezhda Chizhova\xca
## 65 Womens Shotput distance  20.43     \xcaNadezhda Chizhova\xca
## 66 Womens Shotput distance  20.43     \xcaNadezhda Chizhova\xca
## 67 Womens Shotput distance  20.63     \xcaNadezhda Chizhova\xca
## 68 Womens Shotput distance  21.03     \xcaNadezhda Chizhova\xca
## 69 Womens Shotput distance  21.20     \xcaNadezhda Chizhova\xca
## 70 Womens Shotput distance  21.60         \xcaMarianne Adam\xca
## 71 Womens Shotput distance  21.67         \xcaMarianne Adam\xca
## 72 Womens Shotput distance  21.87      \xcaIvanka Khristova\xca
## 73 Womens Shotput distance  21.89      \xcaIvanka Khristova\xca
## 74 Womens Shotput distance  21.99 \xcaHelena Fibingerov\x87\xca
## 75 Womens Shotput distance  22.32 \xcaHelena Fibingerov\x87\xca
## 76 Womens Shotput distance  22.36       \xcaIlona Slupianek\xca
## 77 Womens Shotput distance  22.45       \xcaIlona Slupianek\xca
## 78 Womens Shotput distance  22.53    \xcaNatalya Lisovskaya\xca
## 79 Womens Shotput distance  22.60    \xcaNatalya Lisovskaya\xca
## 80 Womens Shotput distance  22.63    \xcaNatalya Lisovskaya\xca
##    Nationality                      Location Year
## 1          USA                 Lawrence, U.S. 1948
## 2          USA                   Oslo, Norway 1949
## 3          USA              Los Angeles, U.S. 1950
## 4          USA                  Visby, Sweden 1950
## 5          USA             Eskilstuna, Sweden 1950
## 6          USA                   Fresno, U.S. 1953
## 7          USA                  Compton, U.S. 1953
## 8          USA              Los Angeles, U.S. 1954
## 9          USA              Los Angeles, U.S. 1954
## 10         USA              Los Angeles, U.S. 1954
## 11         USA           Salt Lake City, U.S. 1956
## 12         USA              Los Angeles, U.S. 1956
## 13         USA                   Eugene, U.S. 1956
## 14         USA              Los Angeles, U.S. 1956
## 15         USA            Santa Barbara, U.S. 1959
## 16         USA              Albuquerque, U.S. 1959
## 17         USA              Los Angeles, U.S. 1960
## 18         USA                Palo Alto, U.S. 1960
## 19         USA              Los Angeles, U.S. 1960
```

```
## 20         USA                  Austin, U.S. 1960
## 21         USA                  Walnut, U.S. 1960
## 22         USA             Los Angeles, U.S. 1962
## 23         USA             Los Angeles, U.S. 1964
## 24         USA             Los Angeles, U.S. 1964
## 25         USA             Los Angeles, U.S. 1964
## 26         USA          College Station, U.S. 1965
## 27         USA          College Station, U.S. 1967
## 28         USA                 San Jose, U.S. 1973
## 29         USA         Honolulu, Hawaii, U.S. 1976
## 30         URS                  Paris, France 1976
## 31         GDR             Gothenburg, Sweden 1978
## 32         GDR             Los Angeles, U.S. 1983
## 33         GDR               Berlin, Germany 1985
## 34         GDR               Berlin, Germany 1986
## 35         ITA              Viareggio, Italy 1987
## 36         ITA              Viareggio, Italy 1987
## 37         ITA              Viareggio, Italy 1987
## 38         GDR                Chania, Greece 1988
## 39         USA             Los Angeles, U.S. 1990
## 40         URS           Moscow, Soviet Union 1948
## 41         URS          Tbilisi, Soviet Union 1949
## 42         URS              Ploie?ti, Romania 1950
## 43         URS              Helsinki, Finland 1952
## 44         URS           Frunze, Soviet Union 1952
## 45         URS           Frunze, Soviet Union 1952
## 46         URS               Malm\x9a, Sweden 1953
## 47         URS             Kiev, Soviet Union 1954
## 48         URS        Leningrad, Soviet Union 1955
## 49         URS          Tbilisi, Soviet Union 1955
## 50         URS         Tashkent, Soviet Union 1956
## 51         URS          Nalchik, Soviet Union 1959
## 52         URS           Moscow, Soviet Union 1960
## 53         URS           Moscow, Soviet Union 1960
## 54         URS          Leipzig, East Germany 1962
## 55         URS            Beograd, Yugoslavia 1962
## 56         URS           Kassel, West Germany 1965
## 57         URS            Sochi, Soviet Union 1968
## 58         GDR Frankfurt (Oder), East Germany 1968
## 59         GDR            Mexico City, Mexico 1968
## 60         GDR            Mexico City, Mexico 1968
## 61         URS           Moscow, Soviet Union 1969
## 62         URS             Chorz\x97w, Poland 1969
## 63         GDR      East Berlin, East Germany 1969
## 64         URS                 Athens, Greece 1969
## 65         URS                 Athens, Greece 1969
## 66         URS           Moscow, Soviet Union 1971
## 67         URS            Sochi, Soviet Union 1972
## 68         URS           Munich, West Germany 1972
## 69         URS             Lvov, Soviet Union 1973
## 70         GDR      East Berlin, East Germany 1975
## 71         GDR  Karl-Marx-Stadt, East Germany 1976
## 72         BUL             Belmeken, Bulgaria 1976
## 73         BUL             Belmeken, Bulgaria 1976
```

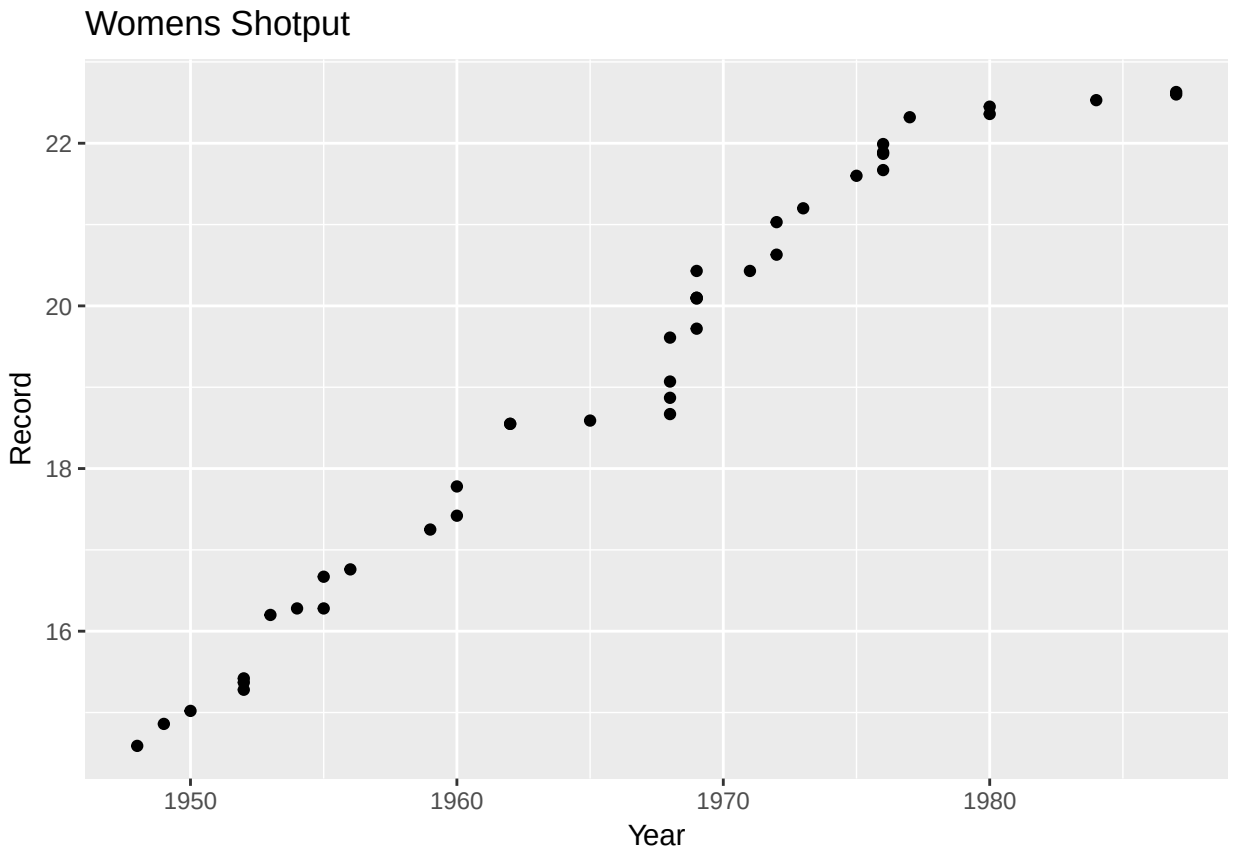
```
## 74         TCH          Opava, Czechoslovakia 1976
## 75         TCH          Nitra, Czechoslovakia 1977
## 76         GDR             Celje, Yugoslavia 1980
## 77         GDR          Potsdam, East Germany 1980
## 78         URS            Sochi, Soviet Union 1984
## 79         URS           Moscow, Soviet Union 1987
## 80         URS           Moscow, Soviet Union 1987
```

5. Create a scatter plot of the year and record shotput distance one for men and one for women.

```
library(ggplot2)
q5.m <- filter(q4, Event=="Mens Shotput")
ggplot(q5.m, aes(x = Year, y = Record))+ geom_point() + labs(title="Mens Shotput")
```




```
q5.w <- filter(q4, Event=="Womens Shotput")
ggplot(q5.w, aes(x = Year, y = Record))+ geom_point() + labs(title="Womens Shotput")
```



6. Find the average/mean time for each event. How many athletes have time more than average in each event.

```
q6.1 <- filter(worldrecords, Type=="time")
q6.2 <- group_by(q6.1, Event)
summarise(q6.2, Avg=mean(Record))
```

```
## # A tibble: 6 x 2
##   Event          Avg
##   <fct>        <dbl>
## 1 Mens 100m     9.85
## 2 Mens 800m   105.
## 3 Mens Mile   237.
## 4 Womens 100m  10.9
## 5 Womens 800m 128.
## 6 Womens Mile 264.
```

```
q6.4 <- mutate(q6.2, Avg=mean(Record))
q6.5 <- filter(q6.4, Record > Avg)
summarise(q6.5, Athletes=n())
```

```
## # A tibble: 6 x 2
##   Event       Athletes
##   <fct>          <int>
## 1 Mens 100m          9
## 2 Mens 800m         10
```

```
## 3 Mens Mile         15
## 4 Womens 100m        4
## 5 Womens 800m       13
## 6 Womens Mile        5
```

7. Select the athlete who took most time in men's 100m and women's event.

```
q7.m.1 <- filter(worldrecords, Event=="Mens 100m")
q7.m.2 <- filter(q7.m.1, Record==max(Record))
q7.m.2
```

```
##       Event Type Record   Athlete   Nationality     Location Year
## 1 Mens 100m time  10.06 Bob Hayes United States Tokyo, Japan 1964
```

```
q7.w.1 <- filter(worldrecords, Event=="Womens 100m")
q7.w.2 <- filter(q7.w.1, Record==max(Record))
q7.w.2
```

```
##         Event Type Record        Athlete       Nationality
## 1 Womens 100m time  11.07    Wyomia Tyus \xcaUnited States
## 2 Womens 100m time  11.07 Renate Stecher  \xcaEast Germany
##               Location Year
## 1  Mexico City, Mexico 1968
## 2 Munich, West Germany 1972
```

8. Which country won maximum times of men's 100m event?

```
q8.1 <- filter(worldrecords, Event=="Mens 100m")
q8.2 <- group_by(q8.1, Nationality)
q8.3 <- summarize(q8.2, NumberOfRecords=n())
filter(q8.3, NumberOfRecords == max(NumberOfRecords))
```

```
## # A tibble: 1 x 2
##   Nationality   NumberOfRecords
##   <fct>                  <int>
## 1 United States             11
```

9. How many athletes are there in each event?

```
q9.1 <- group_by(worldrecords, Event)
summarise(q9.1, Athletes=n_distinct(Athlete))
```

```
## # A tibble: 10 x 2
##    Event           Athletes
##    <fct>              <int>
##  1 Mens 100m             11
##  2 Mens 800m             18
##  3 Mens Mile             23
##  4 Mens Polevault        22
##  5 Mens Shotput          13
##  6 Mens TripleJump       18
```

```
##  7 Womens 100m            8
##  8 Womens 800m           21
##  9 Womens Mile           10
## 10 Womens Shotput        12
```

10. Which country has maximum wins?

```
q10.1 <- group_by(worldrecords, Nationality)
q10.2 <- summarize(q10.1, NumberOfRecords=n())
filter(q10.2, NumberOfRecords == max(NumberOfRecords))
```

```
## # A tibble: 1 x 2
##   Nationality NumberOfRecords
##   <fct>                 <int>
## 1 URS                      49
```