HW#2

Milad Tatari

9/18/2019

```
surveys <- read.csv("surveys.csv", header = T, sep = ",")
library('rmarkdown')
```

1. Write R code to extract the survey observations for the first three months of 1990 using the filter() function. (5 points)

The command filter enables us to extract what we need out of the data frame.

```
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
```

```
surveys <- read.csv("surveys.csv", header = T, sep = ",")
head(surveys)
```

```
##   record_id month day year plot_id species_id sex hindfoot_length weight
## 1         1     7  16 1977       2         NL   M              32     NA
## 2         2     7  16 1977       3         NL   M              33     NA
## 3         3     7  16 1977       2         DM   F              37     NA
## 4         4     7  16 1977       7         DM   M              36     NA
## 5         5     7  16 1977       3         DM   M              35     NA
## 6         6     7  16 1977       1         PF   M              14     NA
```

```
filtered1 <-filter(surveys, year == 1990, month %in% c(1,2,3))
```

2. Sort the 1990 winter surveys data by descending order of record ID, then by ascending order of weight.

```
sorted2 <- arrange(filtered1, desc(record_id), weight)
head(sorted2)
```

```
##   record_id month day year plot_id species_id sex hindfoot_length weight
## 1     17369     3  30 1990       8         DM   F              36     39
## 2     17368     3  30 1990      11         DM   F              35     41
## 3     17367     3  30 1990       4         DM   M              37     44
## 4     17366     3  30 1990      11         DM   M              37     46
## 5     17365     3  30 1990       4         DM   F              38     48
## 6     17364     3  30 1990       8         DM   M              36     51
```

```
sorted3 <- sorted2[with(sorted2, order(weight)), ]
head(sorted3)
```

```
##     record_id month day year plot_id species_id sex hindfoot_length weight
## 154     17216     2  25 1990      10         RM   M              17      6
## 455     16915     1   6 1990      19         PF   F              16      6
## 24      17346     3  30 1990       3         PF   F              16      7
## 52      17318     3  30 1990      15         PF   M              15      7
## 59      17311     3  30 1990       3         PF   F              16      7
## 84      17286     3  29 1990       6         PF   F              15      7
```

3. Write code that returns the record_id, sex and weight of all surveyed individuals of Reithrodontomys montanus (RO)

```
surveysRO <- filter(surveys, species_id == "RO")
select(surveysRO, record_id, sex, weight)
```

```
##   record_id sex weight
## 1     18871   F     11
## 2     33397   M      8
## 3     33556   M      9
## 4     33565   F      8
## 5     34517   M     11
## 6     35402   F     12
## 7     35420   M     10
## 8     35487   F     13
```

4. Write code that returns the average weight and hindfoot length of Dipodomys merriami (DM) individuals observed in each month (irrespective of the year). Make sure to exclude NA values.

```
surveysDM <- filter(surveys, species_id == "DM")
surveysDM1 <- group_by(surveysDM, month)
summarize(surveysDM1, avg_wgt = mean(weight, na.rm = TRUE),
          avg_hfl = mean(hindfoot_length, na.rm = TRUE))
```

```
## # A tibble: 12 x 3
##    month avg_wgt avg_hfl
##    <int>   <dbl>   <dbl>
##  1     1    42.9    36.1
##  2     2    44.0    36.2
##  3     3    45.2    36.1
##  4     4    44.8    36.2
##  5     5    43.2    35.8
##  6     6    41.5    36.0
##  7     7    41.9    35.7
##  8     8    41.8    35.8
##  9     9    43.3    35.8
## 10    10    42.5    36.0
## 11    11    42.4    35.9
## 12    12    43.0    36.0
```

5. Write code that determines the number of individuals by species observed in the winter of 1990.

```
filtered1 <-filter(surveys, year == 1990, month %in% c(1,2,3))
table(filtered1$species_id)
```

```
##
##      AB  AH  AS  BA  CB  CM  CQ  CS  CT  CU  CV  DM  DO  DS  DX  NL  OL
##   1  25   4   0   3   0   0   0   0   0   0   0 132  65   6   0  10   7
##  OT  OX  PB  PC  PE  PF  PG  PH  PI  PL  PM  PP  PU  PX  RF  RM  RO  RX
##  22   0   0   7  37  19   4   3   0   0   0   1   0   0  10 115   0   0
##  SA  SC  SF  SH  SO  SS  ST  SU  UL  UP  UR  US  ZL
##   0   0  13   7   0   0   0   0   0   0   0   0   0
```

Questions 6-10

```
library("dplyr")
#install.packages("gapminder")
library("gapminder")
data1 <- as.data.frame(gapminder)
```

6. Create a dataframe named gapminder_df and mutate it to contain a column that contains the gross domestic product for each row in the data frame.

```
gapminder_df <- mutate(data1, GDP = gdpPercap*pop)
head(gapminder_df)
```

```
##       country continent year lifeExp      pop gdpPercap         GDP
## 1 Afghanistan      Asia 1952  28.801  8425333  779.4453  6567086330
## 2 Afghanistan      Asia 1957  30.332  9240934  820.8530  7585448670
## 3 Afghanistan      Asia 1962  31.997 10267083  853.1007  8758855797
## 4 Afghanistan      Asia 1967  34.020 11537966  836.1971  9648014150
## 5 Afghanistan      Asia 1972  36.088 13079460  739.9811  9678553274
## 6 Afghanistan      Asia 1977  38.438 14880372  786.1134 11697659231
```

7. Calculate the Mean GDP for Cambodia for the years within the dataset. (15 points)

```
gapminder_df %>%
  filter(country == "Cambodia")%>%
  summarize(mean_gdp = mean(GDP))
```

```
##     mean_gdp
## 1 6596612377
```

8. Find the year with the maximum life expectancy for countries in Asia and arrange them in descending order by year, The result should contain the country's name, the year and the life expectancy.

```
x<-gapminder_df %>%
  filter(continent == "Asia")%>%
  group_by(country) %>%
  filter(lifeExp == max(lifeExp))%>%
  select(country,continent,year,lifeExp)%>%
  arrange(desc(year))
head(x)
```

```
## # A tibble: 6 x 4
## # Groups:   country [6]
##   country          continent  year lifeExp
##   <fct>            <fct>     <int>   <dbl>
## 1 Afghanistan      Asia       2007    43.8
## 2 Bahrain          Asia       2007    75.6
## 3 Bangladesh       Asia       2007    64.1
## 4 Cambodia         Asia       2007    59.7
## 5 China            Asia       2007    73.0
## 6 Hong Kong, China Asia       2007    82.2
```

9. Count the number of observations per continent.

```
y<- gapminder_df %>%
  group_by(continent) %>%
  summarize(n = n())
head(y)
```

```
## # A tibble: 5 x 2
##   continent     n
##   <fct>     <int>
## 1 Africa      624
## 2 Americas    300
## 3 Asia        396
## 4 Europe      360
## 5 Oceania      24
```

10. Compute the average and median life expectancy and GDP per capita by continent for the years 1952 and 2007. Should we be optimistic given the results?

```
# Summarize the median GDP and median life expectancy per continent in 2007
by_continent_2007 <- gapminder_df %>%
  filter(year == 2007) %>%
  group_by(continent) %>%
  summarize(medianGdpPercap = median(gdpPercap),
            medianLifeExp = median(lifeExp), meanGdpPercap= mean(gdpPercap), meanLifeExp=mean(li
feExp))
head(by_continent_2007)
```

```
## # A tibble: 5 x 5
##   continent medianGdpPercap medianLifeExp meanGdpPercap meanLifeExp
##   <fct>               <dbl>         <dbl>         <dbl>       <dbl>
## 1 Africa              1452.          52.9         3089.        54.8
## 2 Americas            8948.          72.9        11003.        73.6
## 3 Asia                4471.          72.4        12473.        70.7
## 4 Europe             28054.          78.6        25054.        77.6
## 5 Oceania            29810.          80.7        29810.        80.7
```

```
by_continent_1952 <- gapminder_df %>%
  filter(year == 1952) %>%
  group_by(continent) %>%
  summarize(medianGdpPercap = median(gdpPercap),
            medianLifeExp = median(lifeExp), meanGdpPercap= mean(gdpPercap), meanLifeExp=mean(li
feExp))
head(by_continent_1952)
```

```
## # A tibble: 5 x 5
##   continent medianGdpPercap medianLifeExp meanGdpPercap meanLifeExp
##   <fct>               <dbl>         <dbl>         <dbl>       <dbl>
## 1 Africa               987.          38.8         1253.        39.1
## 2 Americas            3048.          54.7         4079.        53.3
## 3 Asia                1207.          44.9         5195.        46.3
## 4 Europe              5142.          65.9         5661.        64.4
## 5 Oceania            10298.          69.3        10298.        69.3
```

Yes, we have to be happy that average expected life has been increased from 1952 to 2007 in all continents.