Dsouza_Clinton_Assignment #2

Clinton Dsouza

9/14/2019

Lets load the libraries required for this assignment

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------- tidyvers

## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------------------- tidyverse_con
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(gapminder)
```

Lets convert gapminder data to a data frame as below:

```
gapminder <- as.data.frame(gapminder)
```

Lets load the surveys.csv data as shown below:

```
setwd("C:\\Users\\dsouz\\Desktop")
surveys <- read.csv("surveys.csv", header = T, sep = ",")
head(surveys,40)
```

```
##    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
## 7          7     7  16 1977       2         PE   F              NA     NA
## 8          8     7  16 1977       1         DM   M              37     NA
## 9          9     7  16 1977       1         DM   F              34     NA
## 10        10     7  16 1977       6         PF   F              20     NA
## 11        11     7  16 1977       5         DS   F              53     NA
## 12        12     7  16 1977       7         DM   M              38     NA
## 13        13     7  16 1977       3         DM   M              35     NA
## 14        14     7  16 1977       8         DM                  NA     NA
## 15        15     7  16 1977       6         DM   F              36     NA
## 16        16     7  16 1977       4         DM   F              36     NA
```

```
## 17        17     7  16 1977       3         DS   F              48     NA
## 18        18     7  16 1977       2         PP   M              22     NA
## 19        19     7  16 1977       4         PF                  NA     NA
## 20        20     7  17 1977      11         DS   F              48     NA
## 21        21     7  17 1977      14         DM   F              34     NA
## 22        22     7  17 1977      15         NL   F              31     NA
## 23        23     7  17 1977      13         DM   M              36     NA
## 24        24     7  17 1977      13         SH   M              21     NA
## 25        25     7  17 1977       9         DM   M              35     NA
## 26        26     7  17 1977      15         DM   M              31     NA
## 27        27     7  17 1977      15         DM   M              36     NA
## 28        28     7  17 1977      11         DM   M              38     NA
## 29        29     7  17 1977      11         PP   M              NA     NA
## 30        30     7  17 1977      10         DS   F              52     NA
## 31        31     7  17 1977      15         DM   F              37     NA
## 32        32     7  17 1977      10         DM   F              35     NA
## 33        33     7  17 1977      11         DM   F              36     NA
## 34        34     7  17 1977      17         DM                  NA     NA
## 35        35     7  17 1977      13         DM   F              38     NA
## 36        36     7  17 1977      16         OT   F              22     NA
## 37        37     7  17 1977      11         DM   F              35     NA
## 38        38     7  17 1977      17         NL   M              33     NA
## 39        39     7  17 1977      11         DM   F              36     NA
## 40        40     7  18 1977      20         DM   M              36     NA
```

QUESTION 1: Extract surveys observation for the first 3 months of 1990

```
first_3_1990 <- surveys %>%
              filter(year == 1990 & (month ==1 | month == 2 | month ==3))
head(first_3_1990, 40)
```

```
##    record_id month day year plot_id species_id sex hindfoot_length weight
## 1      16879     1   6 1990       1         DM   F              37     35
## 2      16880     1   6 1990       1         OL   M              21     28
## 3      16881     1   6 1990       6         PF   M              16      7
## 4      16882     1   6 1990      23         RM   F              17      9
## 5      16883     1   6 1990      12         RM   M              17     10
## 6      16884     1   6 1990      24         RM   M              17      9
## 7      16885     1   6 1990      12         SF   M              25     35
## 8      16886     1   6 1990      24         SH   F              30     73
## 9      16887     1   6 1990      12         SF   M              28     44
## 10     16888     1   6 1990      17         DO   M              36     55
## 11     16889     1   6 1990      21         SF   M              29     55
## 12     16890     1   6 1990      12         OT   M              22     23
## 13     16891     1   6 1990      12         DO   F              36     53
## 14     16892     1   6 1990      21         AB                  NA     NA
## 15     16893     1   6 1990      12         OT   F              21     24
## 16     16894     1   6 1990       1         OT   F              21     20
## 17     16895     1   6 1990      12         SF   F              27     75
## 18     16896     1   6 1990      12         RM   M              19     11
## 19     16897     1   6 1990      21         SF   F              29     46
```

```
## 20     16898     1   6 1990      23         RM   M              18     11
## 21     16899     1   6 1990      17         DO   M              36     47
## 22     16900     1   6 1990      19         RM   M              17     10
## 23     16901     1   6 1990      12         AH                  NA     NA
## 24     16902     1   6 1990       7         RM   F              17      9
## 25     16903     1   6 1990       1         OL   M              22     34
## 26     16904     1   6 1990      18         RM   F              17     13
## 27     16905     1   6 1990      21         RM   M              18     11
## 28     16906     1   6 1990      12         DO   M              36     57
## 29     16907     1   6 1990      20         RM   M              17     10
## 30     16908     1   6 1990      19         RM   F              17     12
## 31     16909     1   6 1990      24         RM   M              17     11
## 32     16910     1   6 1990      20         RM   M              16      7
## 33     16911     1   6 1990      18         PE   M              22     20
## 34     16912     1   6 1990      22         DM   M              37     52
## 35     16913     1   6 1990       6         RM   M              17      8
## 36     16914     1   6 1990      17         DM   M              NA     NA
## 37     16915     1   6 1990      19         PF   F              16      6
## 38     16916     1   6 1990      17         NL   F              32    165
## 39     16917     1   6 1990       6         RM   M              16      9
## 40     16918     1   6 1990      18         RM   M              17     10
```

QUESTION 2: Sort 1990 winter surveys data by descending order of record_id and ascending of weight

Here I have assumed that winter lasts for the month of November, December, January and February

```
winter_data <- surveys %>%
  filter(year == 1990 & month %in% c(11,12,1,2)) %>%
  arrange(desc(record_id), weight)
head(winter_data, 40)
```

```
##    record_id month day year plot_id species_id sex hindfoot_length weight
## 1      18189    12  16 1990       5         RM   M              17      8
## 2      18188    12  16 1990       3         DM   F              37     38
## 3      18187    12  16 1990      14         RM   M              16      8
## 4      18186    12  16 1990      11         DM   F              36     43
## 5      18185    12  16 1990       3         DM   M              37     45
## 6      18184    12  16 1990      11         DM   M              36     45
## 7      18183    12  16 1990       9         DM   F              37     40
## 8      18182    12  16 1990      11         DM   F              37     37
## 9      18181    12  16 1990       9         DM   F              37     42
## 10     18180    12  16 1990       8         DM   F              37     43
## 11     18179    12  16 1990      11         OL   F              20     32
## 12     18178    12  16 1990       4         OT   M              21     24
## 13     18177    12  16 1990      14         DM   M              36     43
## 14     18176    12  16 1990       5         OT   F              20     22
## 15     18175    12  16 1990      13         PE   M              19     16
## 16     18174    12  16 1990       8         DM   M              37     39
## 17     18173    12  16 1990      10         RM   F              16     14
## 18     18172    12  16 1990      11         OT   M              20     20
## 19     18171    12  16 1990       4         DM   F              36     39
## 20     18170    12  16 1990      15         RM   F              17     12
```

```
## 21     18169    12  16 1990       3         RM   M              16     10
## 22     18168    12  16 1990       8         DM   F              35     38
## 23     18167    12  16 1990       4         DM   F              37     45
## 24     18166    12  16 1990      13         PE   F              22     24
## 25     18165    12  16 1990      14         DM   F              37     42
## 26     18164    12  16 1990       9         DM   M              37     45
## 27     18163    12  16 1990      11         DM   F              36     43
## 28     18162    12  16 1990       4         DM   F              37     40
## 29     18161    12  16 1990      15         PF   M              15      7
## 30     18160    12  16 1990      14         DM   M              36     44
## 31     18159    12  16 1990       8         DM   M              36     52
## 32     18158    12  16 1990       4         DM   M              37     39
## 33     18157    12  16 1990      16         RM   F              17     12
## 34     18156    12  16 1990      11         DM   F              34     37
## 35     18155    12  16 1990       6         RM   F              16      9
## 36     18154    12  16 1990      13         RM   F              17     13
## 37     18153    12  16 1990       8         DM   M              37     36
## 38     18152    12  16 1990       3         RM   F              15     11
## 39     18151    12  16 1990      14         DM   F              36     37
## 40     18150    12  16 1990      11         DM   M              36     50
```

QUESTION 3: Return record_id, sex, weight of all individuals of RO montanus

```
surveys %>%
  filter(species_id == "RO")%>%
  select(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
```

QUESTION 4: Return the avg weight and hindfoot length of DM individual for each month

```
surveys %>%
  filter(species_id == "DM") %>%
  group_by(month) %>%
  summarise(avg_weight = mean(weight, na.rm = TRUE), avg_hflength = mean(hindfoot_length, na.rm = TRUE))
```

```
## # A tibble: 12 x 3
##    month avg_weight avg_hflength
##    <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
```

QUESTION 5: Determine number of species observed in winter of 1990

Here I have assumed that winter is from November to February

```
surveys %>%
  filter(year == 1990 & month %in% c(1,2,11,12)) %>%
  group_by(species_id) %>%
  summarise(count = n())
```

```
## # A tibble: 20 x 2
##    species_id count
##    <fct>      <int>
##  1 ""             1
##  2 AB            25
##  3 AH             6
##  4 BA             5
##  5 DM           184
##  6 DO            70
##  7 DS             6
##  8 NL            11
##  9 OL            14
## 10 OT            33
## 11 PC             7
## 12 PE            35
## 13 PF            21
## 14 PG             4
## 15 PH             2
## 16 PP             1
## 17 RF             7
## 18 RM           137
## 19 SF            11
## 20 SH             4
```

QUESTION 6: Mutate to contain a column for gross domesic product for each row

```r
gapminder_df <- gapminder %>%
                  mutate(GDP = gdpPercap*pop)
head(gapminder_df,40)
```

```
##        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  Afghanistan      Asia 1982  39.854 12881816  978.0114  12598563401
## 8  Afghanistan      Asia 1987  40.822 13867957  852.3959  11820990309
## 9  Afghanistan      Asia 1992  41.674 16317921  649.3414  10595901589
## 10 Afghanistan      Asia 1997  41.763 22227415  635.3414  14121995875
## 11 Afghanistan      Asia 2002  42.129 25268405  726.7341  18363410424
## 12 Afghanistan      Asia 2007  43.828 31889923  974.5803  31079291949
## 13     Albania    Europe 1952  55.230  1282697 1601.0561   2053669902
## 14     Albania    Europe 1957  59.280  1476505 1942.2842   2867792398
## 15     Albania    Europe 1962  64.820  1728137 2312.8890   3996988985
## 16     Albania    Europe 1967  66.220  1984060 2760.1969   5476396323
## 17     Albania    Europe 1972  67.690  2263554 3313.4222   7500110047
## 18     Albania    Europe 1977  68.930  2509048 3533.0039   8864476394
## 19     Albania    Europe 1982  70.420  2780097 3630.8807  10094200603
## 20     Albania    Europe 1987  72.000  3075321 3738.9327  11498418358
## 21     Albania    Europe 1992  71.581  3326498 2497.4379   8307722183
## 22     Albania    Europe 1997  72.950  3428038 3193.0546  10945912519
## 23     Albania    Europe 2002  75.651  3508512 4604.2117  16153932130
## 24     Albania    Europe 2007  76.423  3600523 5937.0295  21376411360
## 25     Algeria    Africa 1952  43.077  9279525 2449.0082  22725632678
## 26     Algeria    Africa 1957  45.685 10270856 3013.9760  30956113720
## 27     Algeria    Africa 1962  48.303 11000948 2550.8169  28061403854
## 28     Algeria    Africa 1967  51.407 12760499 3246.9918  41433235247
## 29     Algeria    Africa 1972  54.518 14760787 4182.6638  61739408943
## 30     Algeria    Africa 1977  58.014 17152804 4910.4168  84227416174
## 31     Algeria    Africa 1982  61.368 20033753 5745.1602 115097120653
## 32     Algeria    Africa 1987  65.799 23254956 5681.3585 132119742845
## 33     Algeria    Africa 1992  67.744 26298373 5023.2166 132102425043
## 34     Algeria    Africa 1997  69.152 29072015 4797.2951 139467033682
## 35     Algeria    Africa 2002  70.994 31287142 5288.0404 165447670333
## 36     Algeria    Africa 2007  72.301 33333216 6223.3675 207444851958
## 37      Angola    Africa 1952  30.015  4232095 3520.6103  14899557133
## 38      Angola    Africa 1957  31.999  4561361 3827.9405  17460618347
## 39      Angola    Africa 1962  34.000  4826015 4269.2767  20603593596
## 40      Angola    Africa 1967  35.985  5247469 5522.7764  28980597822
```

QUESTION 7: Calculate mean gdp for cambodia for the years within the dataset

In this example I am assuming it is asking to find the mean GDP for each year over the years in the dataset

```
gapminder_df %>%
  filter(country == "Cambodia") %>%
  group_by(year) %>%
  summarise(mean(GDP))
```

```
## # A tibble: 12 x 2
##     year  `mean(GDP)`
##    <int>        <dbl>
##  1  1952  1729534398.
##  2  1957  2310184671.
##  3  1962  3023033308.
##  4  1967  3643123977.
##  5  1972  3141354496.
##  6  1977  3663574552.
##  7  1982  4541488550.
##  8  1987  5725430805.
##  9  1992  6925441368.
## 10  1997  8652054255.
## 11  2002 11585251106.
## 12  2007 24218877034.
```

If we want to find the mean gdp over all the years in the dataset

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

```
##    mean(GDP)
## 1 6596612377
```

QUESTION 8: Find the year with the max life expectancy for countries in Asia and arrange in descending of year

```
gapminder %>%
  filter(continent == "Asia") %>%
  group_by(country) %>%
  select(country, year, lifeExp) %>%
  filter(lifeExp == max(lifeExp)) %>%
  arrange(desc(year))
```

```
## # A tibble: 33 x 3
## # Groups:   country [33]
##    country           year lifeExp
##    <fct>            <int>   <dbl>
##  1 Afghanistan       2007    43.8
##  2 Bahrain           2007    75.6
##  3 Bangladesh        2007    64.1
##  4 Cambodia          2007    59.7
##  5 China             2007    73.0
```

```
##  6 Hong Kong, China  2007    82.2
##  7 India            2007    64.7
##  8 Indonesia        2007    70.6
##  9 Iran             2007    71.0
## 10 Israel           2007    80.7
## # ... with 23 more rows
```

QUESTION 9: Count number of observations per continent

```
gapminder %>%
  group_by(continent) %>%
  summarise(count = n())
```

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

QUESTION 10: Compute avg and median life expectancy and GDP per capita by continent for year 1952 and 2007

```
gapminder %>%
  filter(year == 1952 | year == 2007) %>%
  group_by(continent, year) %>%
  summarise(avglfexpt = mean(lifeExp), medlyfexpt = median(lifeExp), mean(gdpPercap))
```

```
## # A tibble: 10 x 5
## # Groups:   continent [5]
##    continent  year avglfexpt medlyfexpt `mean(gdpPercap)`
##    <fct>     <int>     <dbl>      <dbl>             <dbl>
##  1 Africa     1952      39.1       38.8             1253.
##  2 Africa     2007      54.8       52.9             3089.
##  3 Americas   1952      53.3       54.7             4079.
##  4 Americas   2007      73.6       72.9            11003.
##  5 Asia       1952      46.3       44.9             5195.
##  6 Asia       2007      70.7       72.4            12473.
##  7 Europe     1952      64.4       65.9             5661.
##  8 Europe     2007      77.6       78.6            25054.
##  9 Oceania    1952      69.3       69.3            10298.
## 10 Oceania    2007      80.7       80.7            29810.
```

The above results show that the averagelife expectancy, median of the life expectancy and the average of the GDP/capita have all increased for all continents in the recent year (2007) as compared to past data (1952). We should be optimistic about the result because there is progress in every domain in the latest data.