Assignment-2

Ashmi

January 18, 2018

# Loading required packages

Used supressMessages with library as without it, I was getting this error while knitting: “! Package inputenc Error: Unicode char √ (U+221A) (inputenc) not set up for use with LaTeX.”

suppressMessages(library(tidyverse))  
suppressMessages(library(ggplot2))  
surveys <- read.csv("surveys.csv", header = T, sep = ",")   
View(surveys)  
suppressMessages(library(gapminder))

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

winter\_surveys\_90 <- surveys %>% filter((year == "1990") , (month %in% c("1","2","3")))   
View(winter\_surveys\_90)  
head(winter\_surveys\_90)

## 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

# 2. Sort the 1990 winter surveys data by descending order of record ID, then by ascending order of weight. (10 points)

Taking the winter surveys data as the data of the first three months

arrange\_record <- arrange(winter\_surveys\_90, desc(record\_id), weight)  
View(arrange\_record)  
head(arrange\_record)

## 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

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

# 3. Write code that returns the record\_id, sex and weight of all surveyed individuals of Reithrodontomys montanus (RO), (10 points)

RO\_data <- surveys %>% filter(species\_id == "RO") %>% select(species\_id, record\_id, sex, weight)  
head(RO\_data)

## species\_id record\_id sex weight  
## 1 RO 18871 F 11  
## 2 RO 33397 M 8  
## 3 RO 33556 M 9  
## 4 RO 33565 F 8  
## 5 RO 34517 M 11  
## 6 RO 35402 F 12

# 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. (10 points)

DM <- filter(surveys, species\_id =="DM" )  
View(DM)  
months\_DM <- group\_by(DM, month)  
View(months\_DM)  
month\_sum <- summarise(months\_DM, weight\_mean = mean(weight, na.rm = T), hindfoot\_length\_mean = mean(hindfoot\_length, na.rm = T))  
month\_sum

## # A tibble: 12 x 3  
## month weight\_mean hindfoot\_length\_mean  
## <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. (15 points)

Taking the winter surveys data as the data of the first three months

Groups <- summarize(group\_by(winter\_surveys\_90, species\_id), num\_individuals =n())  
head(Groups)

## # A tibble: 6 x 2  
## species\_id num\_individuals  
## <fctr> <int>  
## 1 "" 1  
## 2 AB 25  
## 3 AH 4  
## 4 BA 3  
## 5 DM 132  
## 6 DO 65

# 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. (5 points)

gapminder\_df <- as.data.frame(gapminder)  
View(gapminder\_df)  
gapminder\_df <- mutate(gapminder\_df, GDP = pop\*gdpPercap)  
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)

summarize(gapminder\_df, country = "Cambodia", mean\_GDP = mean(GDP))

## country mean\_GDP  
## 1 Cambodia 186809560507

# 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. (15 points)

gapminder\_df %>% filter(continent == "Asia") %>% group\_by(country) %>% filter(lifeExp == max(lifeExp)) %>% select(country,year,lifeExp) %>% arrange(desc(year))

## # A tibble: 33 x 3  
## # Groups: country [33]  
## country year lifeExp  
## <fctr> <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

The year with the maximum life expectancy for countries in Asia is 2007 #9. Count the number of observations per continent. (5 points)

gapminder\_df %>% group\_by(continent) %>% summarise(num\_obs=n())

## # A tibble: 5 x 2  
## continent num\_obs  
## <fctr> <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? (10 points)

gapminder\_df %>% filter(year %in% c(1952,2007)) %>% group\_by(continent, year) %>% summarize(mean\_lifeExp = mean(lifeExp), med\_lifeExp = median(lifeExp), avg\_GDP = mean(gdpPercap), med\_GDP = median(gdpPercap))

## # A tibble: 10 x 6  
## # Groups: continent [?]  
## continent year mean\_lifeExp med\_lifeExp avg\_GDP med\_GDP  
## <fctr> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 Africa 1952 39.1 38.8 1253 987  
## 2 Africa 2007 54.8 52.9 3089 1452  
## 3 Americas 1952 53.3 54.7 4079 3048  
## 4 Americas 2007 73.6 72.9 11003 8948  
## 5 Asia 1952 46.3 44.9 5195 1207  
## 6 Asia 2007 70.7 72.4 12473 4471  
## 7 Europe 1952 64.4 65.9 5661 5142  
## 8 Europe 2007 77.6 78.6 25054 28054  
## 9 Oceania 1952 69.3 69.3 10298 10298  
## 10 Oceania 2007 80.7 80.7 29810 29810

We should be optimistic as the average and median life expectancy and GDP per capita by continent are increasing.