Homework #1

Le Wang

Instruction: Do all the following empirical exercises using R and make your answer using R Markdown. For your convinence, I also posted my R markdown file `ps1.Rmd` on Github so you just need to download it and fill in the code chunks. Refer to the output whenever appropriate when discussing your results.

Question 1.[Basic Matrix Algebra]

For this question, you have to search for some new commands to perform matrix algebra.

1. Create the following matrix, A, and vector, x, and show it by using `print(A)`

$$A = \begin{pmatrix} 3 & 2 \\ 5 & 3 \end{pmatrix} \quad x = \begin{pmatrix} 1 & 2 \end{pmatrix}$$

```
A <-
print(A)
x <-
print(x)
```

2. Calculate xA, the matrix muplication of the matrice A and the vector x. **matrix muplication** cannot simply be calculated using `*`. You need to
3. Find out the dimension of xA.
4. Find out A^{-1}, the inverse of the matrix A.
5. Calculate the muplication of the matrice A and its inverse: $A \cdot A^{-1}$. Note that due to rounding errors, it may not necessarily be the exact **Identity matrix** that you are expecting.
6. In class, we talked about how to evaluate whether a not a number is divisble by 4. Now generate a sequence of integers from 1 to 100 (inclusive) by typing the following command

```
x <- 1:100
```

Then using the logical expressions to select and report those numbers in x that are divisible by 5.

Question 2. [Learn How to Manually Input Data and Use Functions to Analyze Data]

Suppose that we have data on the outcomes of OU games from last season. Source. For example, the structure of your data should look like the following (Note this table is just illustrative and incomplete):

Opponent.Name	OU.Score	Opponenet.Score
FAU	63	14
UCLA	49	21
Iowa State	37	27

1. We can enter each column of the data into R as a vector object and name it accordingly. In other words, you should have three variables called **Opponent.Name**, **OU.Score**, and **Opponent.Score**.

A *vector* or a one-dimensional array simply represents a collection of information stored in a specific order. Note that you have two different types of vectors, one character/string variable, and the other numerical variables.

2. Show the contents of **Opponent.Name** to make sure that this variable indeed contains the observations that we want.

3. Using the function **data.frame()**, combine these three variables (or vectors) together into a dataset called `OUdata`.

4. Suppose that I type in the following command: **OU.Score[c(2, 4)]**. Can you tell me what this command does? We did not discuss this command in class, but I would like you to find out how to perform some basic operations of the objects (or data) on your own.

5. Let's generate a new variable called **OU.Score2** whose element is equal to the corresponding element in **OU.Score** divided by 10. Such operation is called *element-wise operation*.

6. Let's generate a new variable called **OU.Score3** whose element is equal to the corresponding element in **OU.Score** divided by the first element in **OU.Score**. You must use your answer to (3) to refer to the first element, and cannot simply use 41 in the denominator.

7. Generate a variable called spread equal to the difference between **OU.Score** and **Opponent.Score**.

8. What is the number of elements in the vector **OU.Score**. Use the function **length()** to answer this question.

9. Calculate the sum of all the elements in the vector **OU.Score**.

10. Using your answers above, Calculate the average OU scores from last season. What do you think about OU offense last year?

11. Calculate the average Opponent Scores from last season. What do you think about OU defense last year?

Question 3. [Learn How to Read and Export Data and Use Functions to Analyze Data]

Now, let's use `R` to understand a real-life issue – population dynamics. Understanding population dynamics is important for many areas of social science. We will calculate some basic demographic quantities of births and deaths for the world's population from two time periods: 1950 to 1955 and 2005 to 2010. We will analyze the CSV data file called `Sweden.csv` here.

The names and descriptions of the variables in each data set are

Name	Description
`country`	Abbreviated country name
`period`	Period during which data are collected
`age`	Age group

Name	Description
`births`	Number of births in thousands (i.e., number of children born to women of the age group)
`deaths`	Number of deaths in thousands
`py.men`	Person-years for men in thousands
`py.women`	Person-years for women in thousands

Source: United Nations, Department of Economic and Social Affairs, Population Division (2013). *World Population Prospects: The 2012 Revision, DVD Edition.*

The data are collected for a period of 5 years where *person-year* is a measure of the time contribution of each person during the period. For example, a person that lives through the entire 5 year period contributes 5 person-years whereas someone who only lives through the first half of the period contributes 2.5 person-years. Before you begin this exercise, it would be a good idea to directly inspect each data set. In R, this can be done with the `View` function, which takes as its argument the name of a `data.frame` to be examined. In RStudio, double-clicking a `data.frame` in the `Environment` tab will enable you to view the data in a spreadsheet-like view.

1. Load the data into R first. Call it `Sweden`. Also, show me the first six observations of the data by typing

```
Sweden <-
head(Sweden)
```

2. We begin by computing **crude birth rate** (CBR) for a given period. The CBR is defined as:

$$\text{CBR} = \frac{\text{Total number of births during the period}}{\text{Total number of person-years lived during the period}}$$

Compute the CBR for each period. Let me guide you through this process

2.a. calculate total person-years for each period and age group by combining person-years for men and women, and call it `py.total` and save it in the Data `Sweden`.

```
Sweden$py.total <-
```

2.b. We need to sum up the births born to different age groups for each period (1950-1955 and 2005-2010). That is, you need to sum up `births` for each period.

2.c. Calculate the CBR for each period using the formula above, and combine these two values into a vector called `CBR`. This vector would have two values.

3. Export the vector `CBR` to a dataset called `CBR.csv`.