# Homework 01

*DSO 545: Statistical Computing and Data Visualization*

*Fall 2018*

## Instructions

- Use R to answer all question. Write R code for each question.
- Submit your R file to blackboard with comments when needed
- USC won't tolerate any kind of cheating
- Good luck!

## FIFA 2018 Players

The dataset `fifa_players.csv` provides information on the players featured in FIFA 2018. It includes players details on nationality, age, club, their value and wage.

## Metadata

| Variables | Description |
| --- | --- |
| Name | Name of the player |
| Age | Age of the player |
| Nationality | Nationality of teh player |
| Club | The international club for which the player plays |
| Value | Player value in millions (Euro) |
| Wage | Player wage in thousands (Euro) |

1. *(0.5 points)* What is the mean, min, and max age of the players?

2. *(0.5 points)* List the names of all 16-year old players.

3. *(1 points)* How many French are there in the Arsenal club?

4. *(1 points)* How many Manchester United players are in the age group 16-19 (inclusive)?

5. *(1 point)* Extract Manchester United club and Manchester City club into two dataframes called (`manchester_united`, and `manchester_city` respectively). Each dataframe should only have the variables: Name, Age, Nationality

6. *(1 point)* Use the two dataframes created in the previous question to find the (1) percentage of English players in Manchester United, and (2) the percentage of the English players in Manchester City? Which Club has higher percentage of players from England?

7. *(1 point)* Calculate the average age of the players whose value is in the range of 100-150 Millions inclusive.

8. *(1 point)* Create a new variable "Value.Range" as follows:

| Value | Value.Range |
|---|---|
| Value<102 | Below Average |
| Value>= 102 | Above Average |

9. *(0.5 points)* Use the `table()` function to report the number of players in each category for the variable `Value.Range`.

10. *(1.5 point)* Create a new column "Age.Range"

| Age | Age.Range |
|---|---|
| Age<20 | Below 20 years |
| Age>=20 and Age<25 | [20-25) years |
| Age>=25 and Age<30 | [25-30) years |
| Age>=30 | Above 30 years |

11. *(1 point)* Calculate the percentage of players in each age group (`Age.Range`). Which age group has the highest percentage?