Department of Statistics
2020/21 – Semester II

STA272 – STATISTICAL COMPUTING

Assignment 3

Due: 06 - APR - 21 **Time: 12h00**

Instructions:

- All of your work must be typeset using Rmarkdown and submitted online through the course's blackboard shell.
- Any work submitted late would be penalized as follows:
  - any work submitted before midnight of the due date would attract a penalty of up to 10%
  - any work submitted a day late would attract a penalty of up to 25%
  - any work submitted two days late would attract a penalty of up to 50%
  - otherwise you'll be awarded a zero mark.
- You are encouraged to discuss the assignment with others but at the end you must submit your individual work.
- Any form cheating is not allowed and plagiarized work will be awarded a zero mark.

Q1. The SPSS data `credit.sav` (available on Moodle) contains some information about customer behaviour and buying patterns. The data contains among others, the following variables:

- `gender`: Gender
- `card`: Primary credit card
- `type`: Type of transaction
- `items`: Number of items bought
- `spent`: Total amount spent

(a) Import this dataset to `R`, name it `credit` and keep the above mentioned variables only. Then find the summary statistics of each variable in the data.

(b) Create a new dataset, `agg.credit` that contains the averages of number of items bought and amount spent by `type` and `gender`.

Q2. Consider the `weather` data set from the `nycflights13` package.

(a) Use the `dplyr` package to create a data set which consists of the `origin` and the monthly minimum and maximum temperatures.

(b) The temperatures were measured in degrees Fahrenheit, °F. Convert these measurements to degrees celsius, °C given that

$$°\text{C} = \frac{5(°\text{F} - 32)}{9}.$$

Q3. Lets define a canceled flight with the following conditional statement:

```
is.na(dep_delay) | is.na(arr_delay)
```

(a) Obtain the total number of scheduled and canceled flights by `carrier` for each month.

(b) Name three airline carriers which had the highest proportions of canceled flights during 2013.

Q4. The `csv` file named `BWA_QGDP` in STA272 Moodle course shell consist of Botswana's **quarterly** gross domestic product (GDP) in millions of Pula from 2005 – 2018 at current prices.

(a) Use the `dplyr` package in `R` to compute the Botswana's **annual** GDP growth rate given by

$$\text{RATE}_t = \frac{\text{GDP}_t - \text{GDP}_{t-1}}{\text{GDP}_{t-1}} \times 100$$

where $t = 1, 2, \ldots, n$ is the number of years.
`HINT: convert the series to an annual series first. The function diff() is used to find the first difference of a series`

(b) Present the annual growth rates obtained above as a line graph and interpret your graph.

Age	Male	Female	Total
15 – 17	12.3	14.0	13.1
18 – 19	41.0	38.3	39.6
20 – 24	43.5	45.6	44.6
25 – 29	38.3	47.3	43.0
30 – 34	33.6	39.3	36.8
35	29.3	40.6	34.3

Q5. The table above presents NEET (not in employment, education nor training) rate by gender and youth age groups from the Botswana Quarterly Multi-Topic Survey Report: Labour Force Module Quarter 4 Report.

Use `R` to represent this table as a bar graph with proper labels and a legend. Interpret your graph.

Each Question is worth 4 marks!