

ACT3035 Examen Final Hiver 2018

Numéro 1) [/10]

Créez une fonction appelée `nbWeeks` qui prend deux arguments:

1. une date de début
2. et une date de fin

Lorsque cette fonction sera exécutée, elle nous retourne le nombre de semaine ainsi que le nombre de jours entre ces deux dates-là dans une phrase comme suit:

```
nbWeeks("2018-04-30", "2018-07-27")
```

```
## [1] "il y'a 13 semaines, et 4 jours entre le 2018-04-30 et 2018-07-27"
```

Numéro 2)

a) [/10]

Fixez le `seed` de votre générateur de variables aléatoire à la valeur 123. Ensuite, générez 1000 variables pseudo-aléatoires tirées d'une distribution Unifome[-2,2] que vous allez sauvegarder dans un vecteur `x`.

b) [/10]

Utilisez x_i que vous avez généré précédemment afin d'obtenir une variable réponse Y_i définie comme suit;

$$Y_i = x_i^3 + \epsilon_i$$

Où ϵ_i sont des variables i.i.d d'une distribution Normale $N(0.1, 0.1)$. **Sauvegardez** Y_i dans un vecteur `y`.

c) [/10]

Dans un graphique, tracez Y_i .

Numéro 3)

a) [/10]

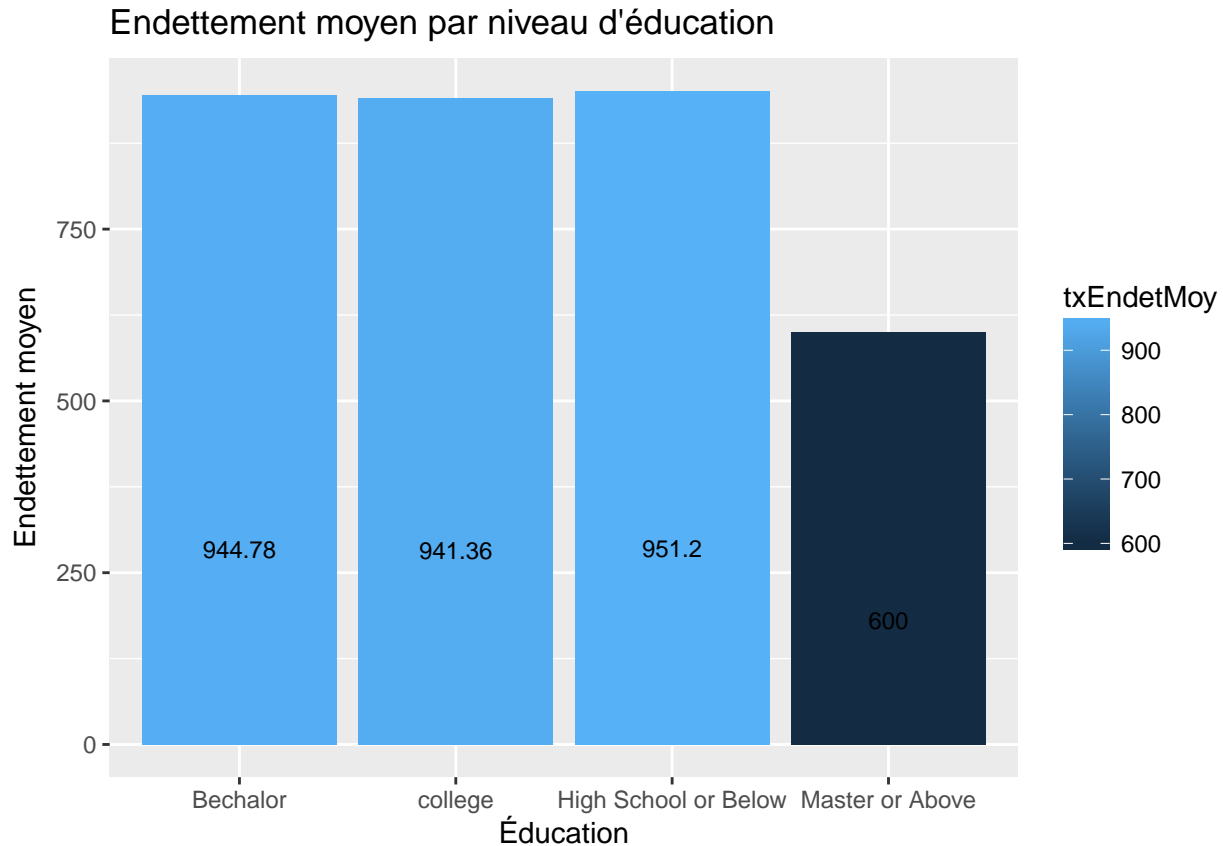
Dans la base de données `Loan payments data.csv`, faite un sommaire qui résume le taux d'endettement moyen (Principal) par niveau d'éducation.

```
## # A tibble: 4 x 2
##   education          txEndetMoy
##   <fctr>             <dbl>
## 1 Bachelor           945
## 2 college            941
## 3 High School or Below 951
```

```
## 4 Master or Above          600
```

b) [/10]

Faites un graphique sur les données du résultat précédent tel qu'illustré ci-dessous;

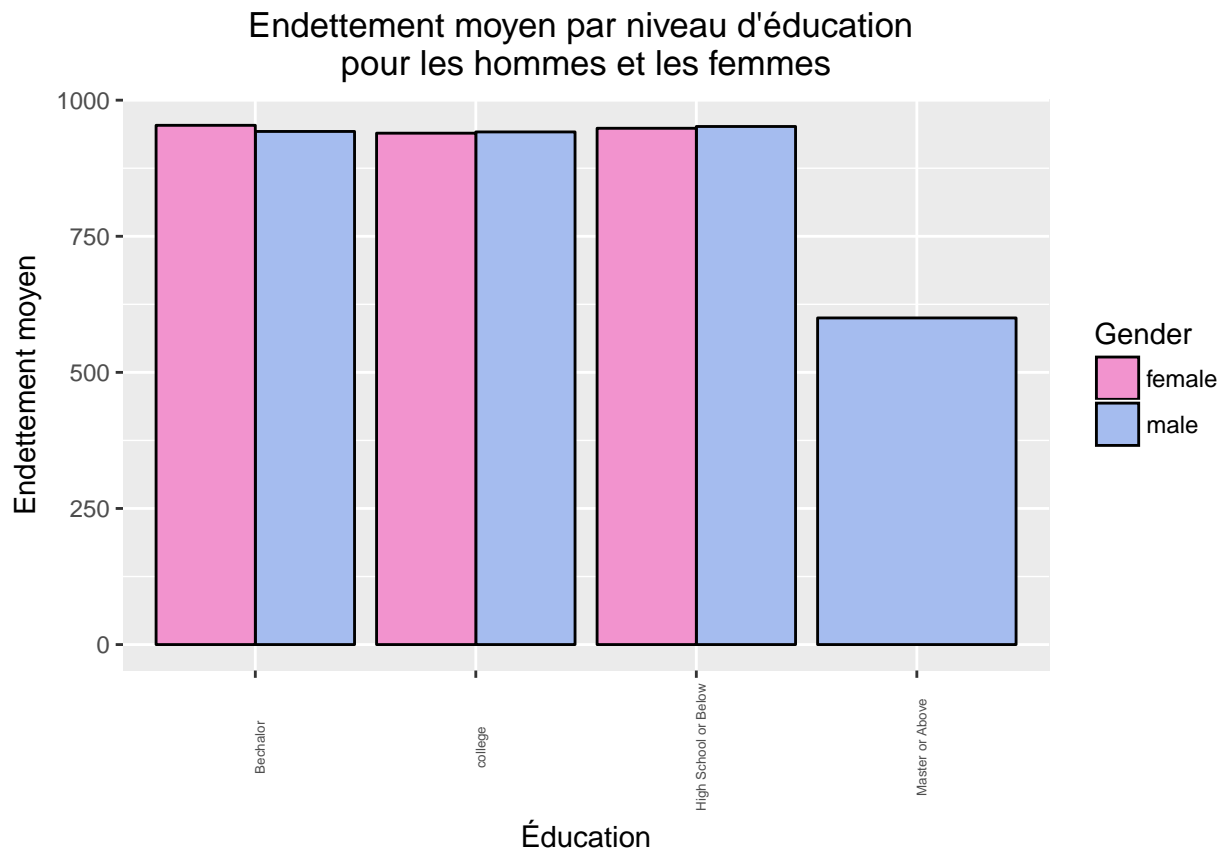


c) [/10]

Faites le même graphique que le numéro précédent, mais cette fois séparez les hommes des femmes. Vous devriez obtenir le graphique ci-dessous où les couleurs utilisées sont "#f293ce" et "#a5bcef". Afin d'obtenir un texte vertical de l'étiquette de l'axe de x , ajoutez la ligne suivante:

```
theme(plot.title = element_text(hjust = 0.5), axis.text.x = element_text(angle = 90, size=5))
```

```
## # A tibble: 7 x 3
## # Groups:   education [?]
##   education      Gender txEndetMoy
##   <fctr>         <fctr>     <dbl>
## 1 Bechalar      female      954
## 2 Bechalar      male       943
## 3 college       female      939
## 4 college       male       942
## 5 High School or Below female      948
## 6 High School or Below male       952
## 7 Master or Above male        600
```



d)[/10]

Supposons que le taux d'intérêt simple que la banque charge à ses clients est de 19.90% annuellement.

Supposons que dans la colonne `past_due_days` on trouve le nombre de jours en retard de paiement en date de l'extraction des données.

Quel serait alors le montant total des intérêts que la banque pourrait encaisser si les clients paient leur dû en cette date? Pour simplifier les calculs, supposer que le taux d'intérêt quotidien est simplement le taux d'intérêt annuel divisé par 365 jours.

Écrivez votre code R qui permet d'obtenir la réponse.

Numéro 5)[/10]

Dans le package `nycflights13`, vous trouvez cinq tables de données sur le trafic aérien de l'aéroport de New York. Écrivez un code R qui fait une jointure entre deux tables (trouvez lesquels) afin d'obtenir une seule table contenant les données suivantes;

```
## # A tibble: 336,776 x 7
##   year month   day hour tailnum carrier name
##   <int> <int> <int> <dbl> <chr>   <chr>   <chr>
## 1  2013     1     1  5.00 N14228   UA      United Air Lines Inc.
## 2  2013     1     1  5.00 N24211   UA      United Air Lines Inc.
## 3  2013     1     1  5.00 N619AA   AA      American Airlines Inc.
## 4  2013     1     1  5.00 N804JB   B6      JetBlue Airways
## 5  2013     1     1  6.00 N668DN   DL      Delta Air Lines Inc.
## 6  2013     1     1  5.00 N39463   UA      United Air Lines Inc.
## 7  2013     1     1  6.00 N516JB   B6      JetBlue Airways
```

