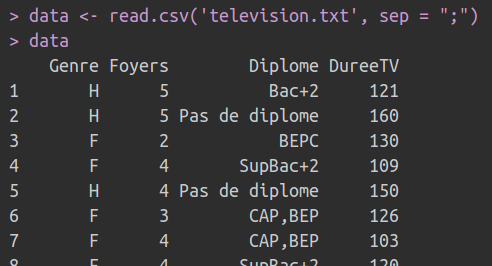
Exercice 1:

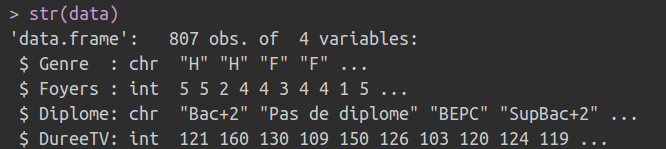
1. Importer les données dans un vecteur intitulé data.



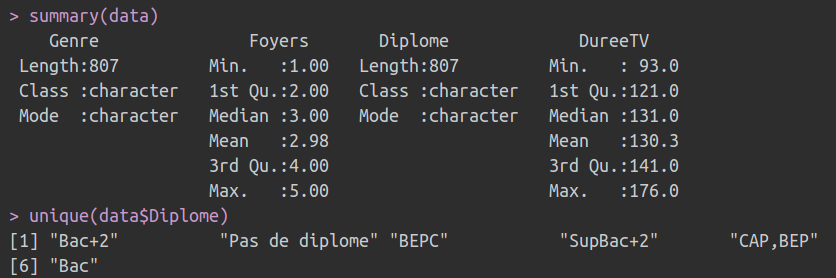
2. Déterminer la taille de l’échantillon.



3. Déterminer la structure de data et préciser les types des variables.

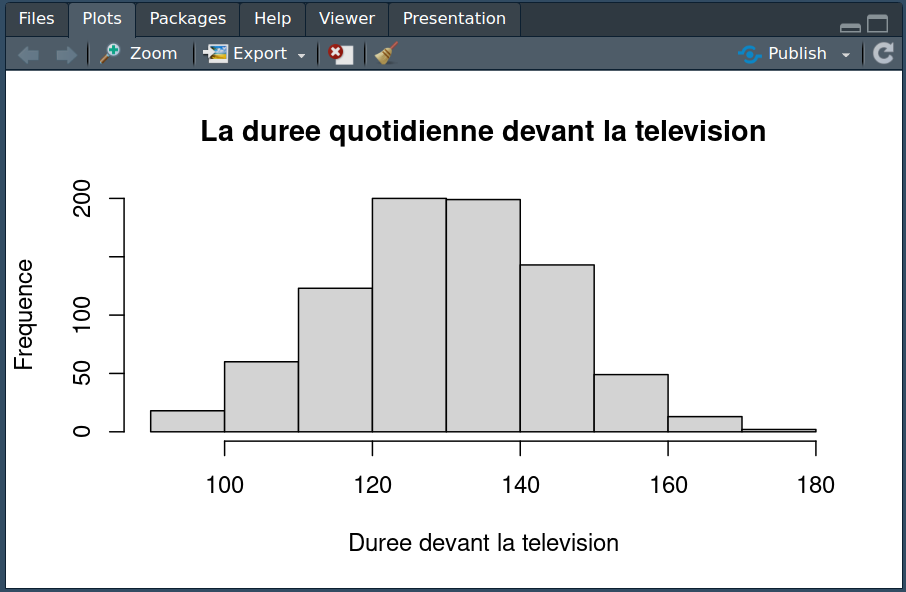


4. Donner un résumé des variables. Donner la liste des diplômes.



5. Effectuer l’histogramme représentant la durée quotidienne devant la télévision.

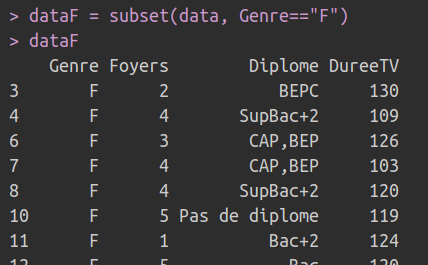




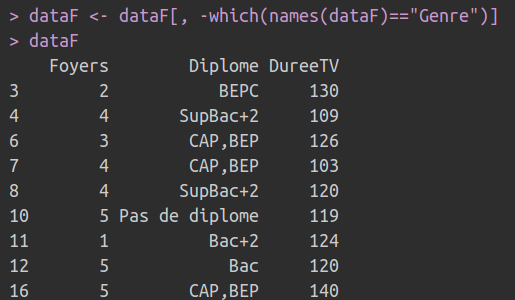
6. Calculer le nombre de personnes qui regardent 2h ou plus par jour en moyenne la télévision.



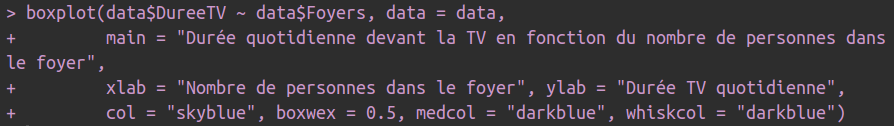
7. Extraire dans un vecteur intitulé dataF les données concernant les femmes.

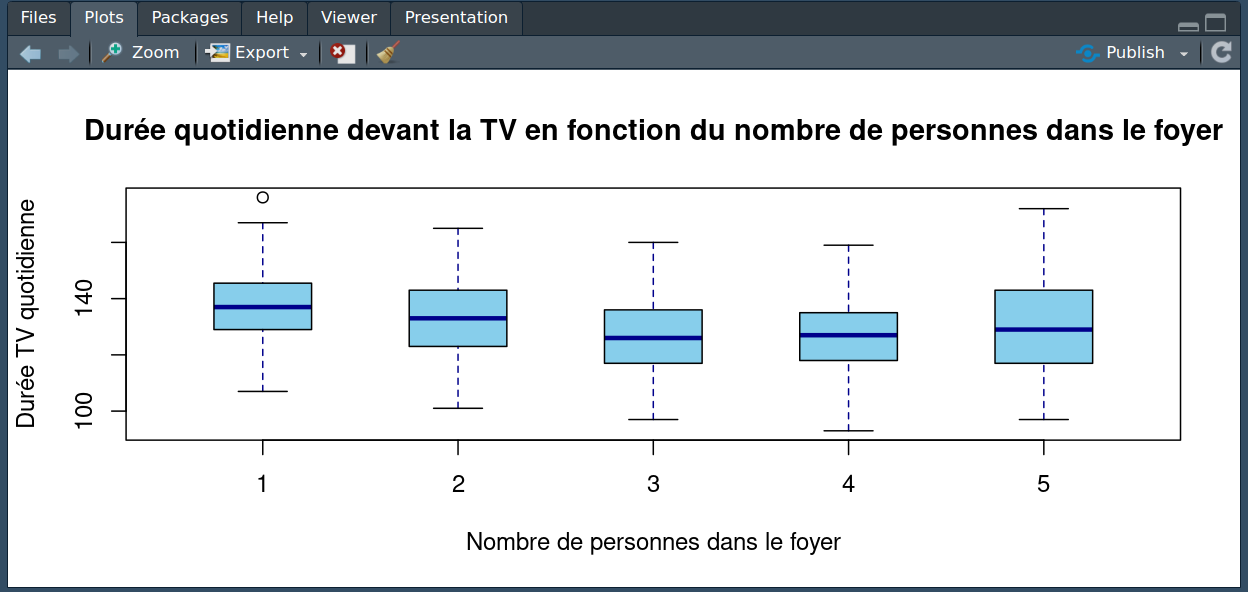


8. Supprimer la colonne Genre de dataF.

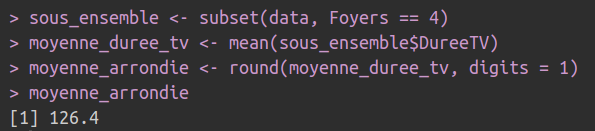


9. Effectuer les diagrammes en boîte de la durée quotidienne devant la TV en fonction du nombre de personnes dans le foyer. Soigner le graphique.

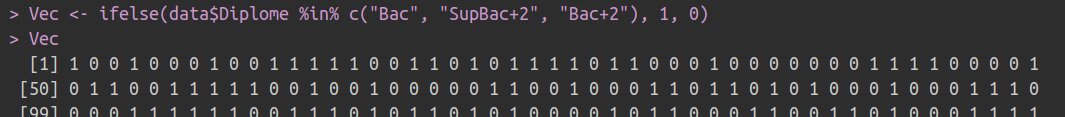




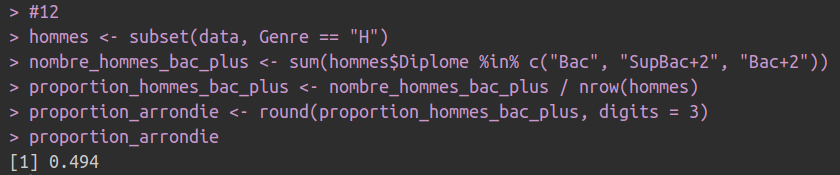
10. Déterminer la moyenne arrondi au dixième de la durée quotidienne devant la TV pour les individus appartenant à un foyer comportant 4 personnes.



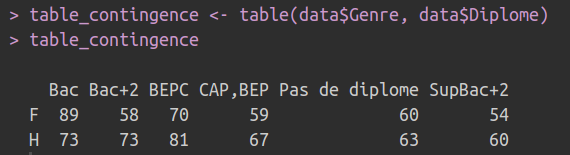
11. Créer un vecteur intitulé Vec qui prend la valeur 1 si l’individu possède le bac ou un diplôme supérieur, et qui prend 0 sinon.



12. Calculer la proportion, arrondie au millième, d’hommes qui ont le bac ou plus dans l’échantillon .

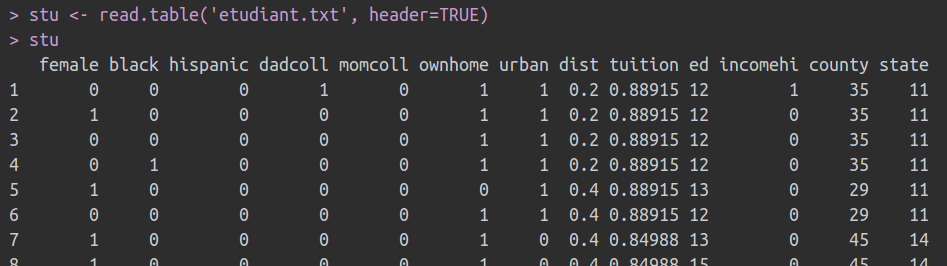


13. Déterminer le tableau de contingence du couple de variables Genre et Diplome.

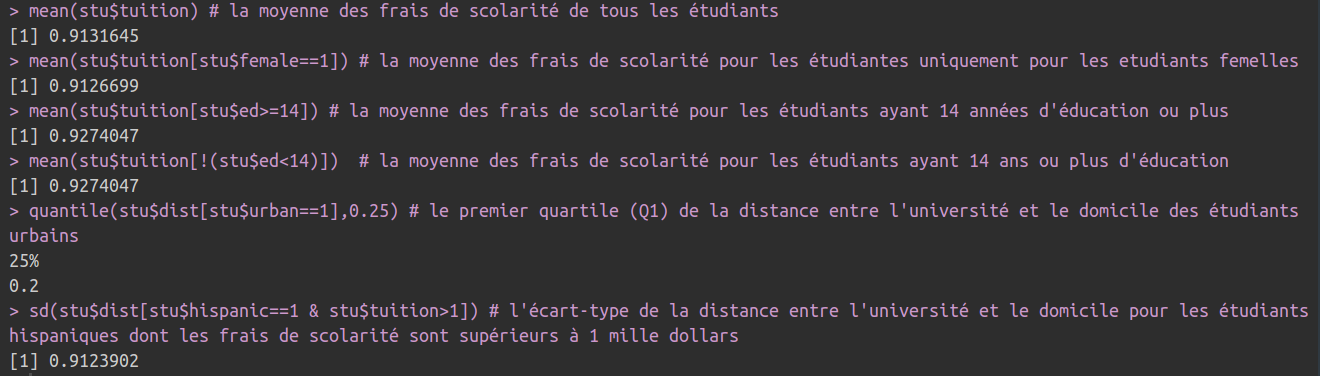


Exercice 2

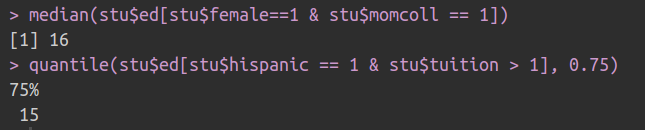
1. Importer les données du fichier etudiant.txt dans un vecteur intitulé stu.



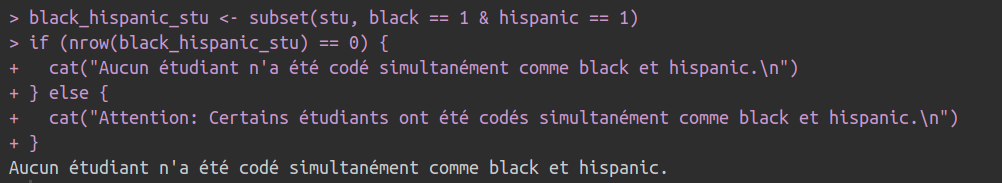
2. Exécuter les commandes ci-dessous et les interpréter.



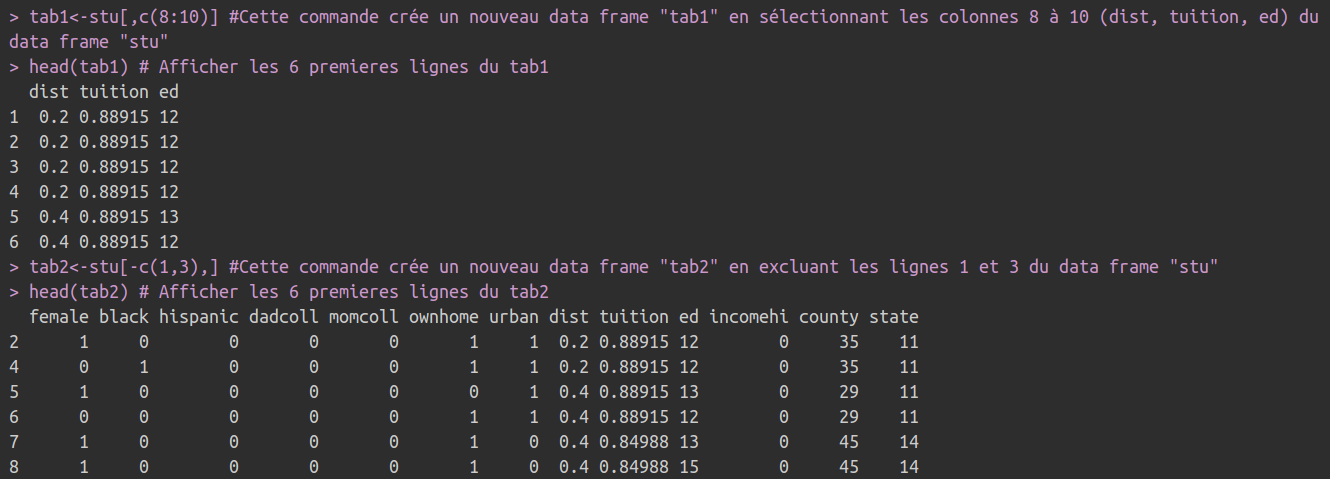
3. Déterminer la médiane du nombre d’années d’étude des femmes dont la mère est diplômée du supérieur ainsi que le troisième quartile du nombre d’années d’étude des hispaniques dont les frais de scolarité sont supérieurs à 1000 $.

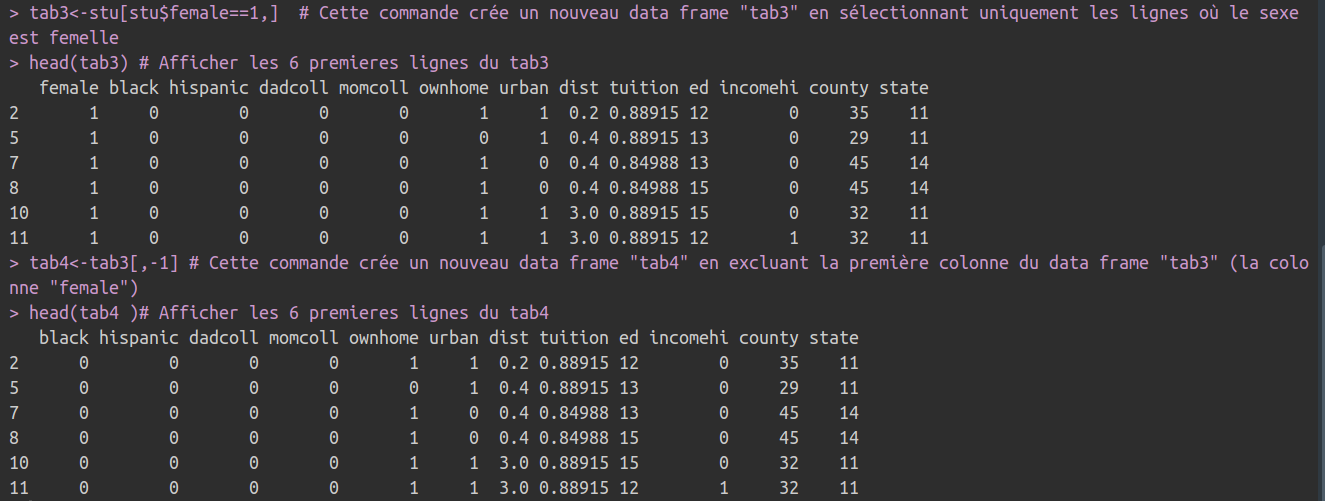


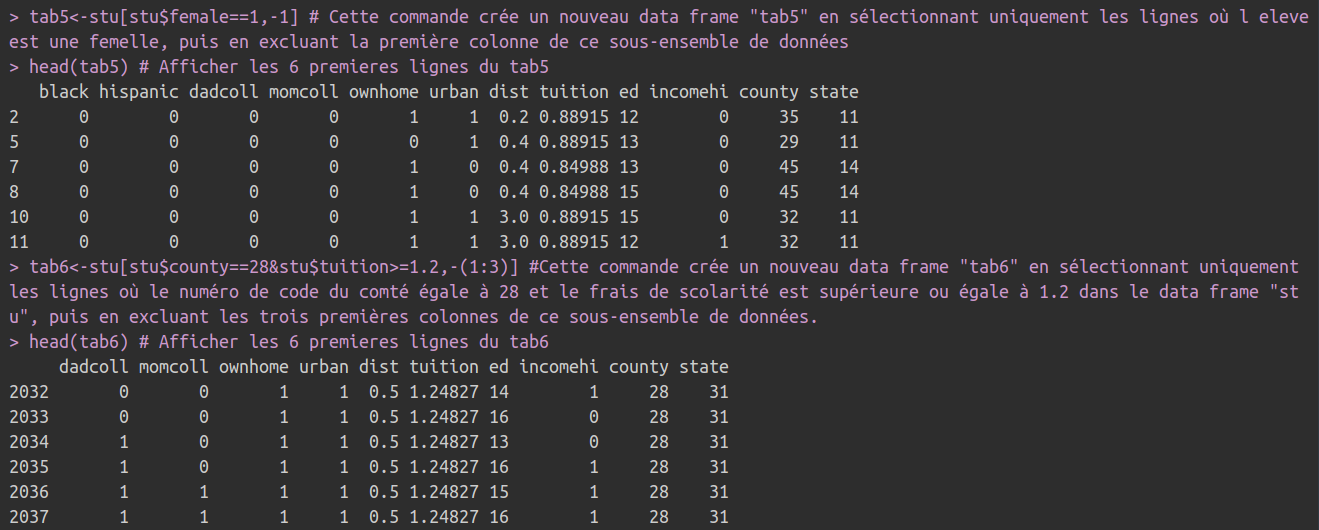
4. Créer un code qui permet de vérifier qu’aucun étudiant n’ait été codé simultanément black et hispanic.



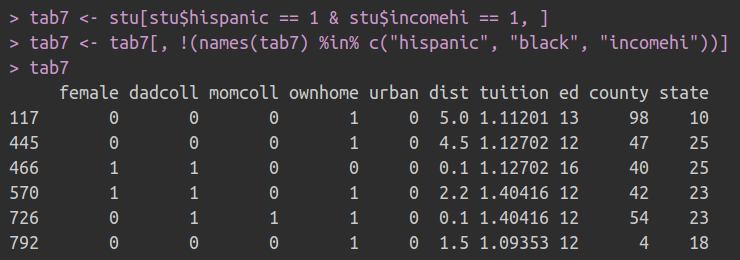
5. Exécuter les commandes suivantes et expliquer ce qu’elles produisent.

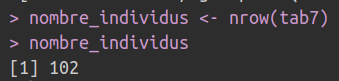




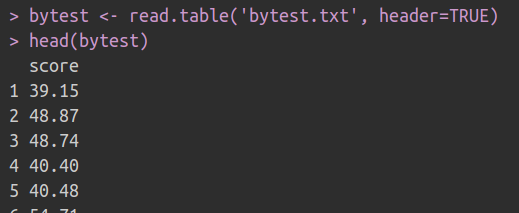


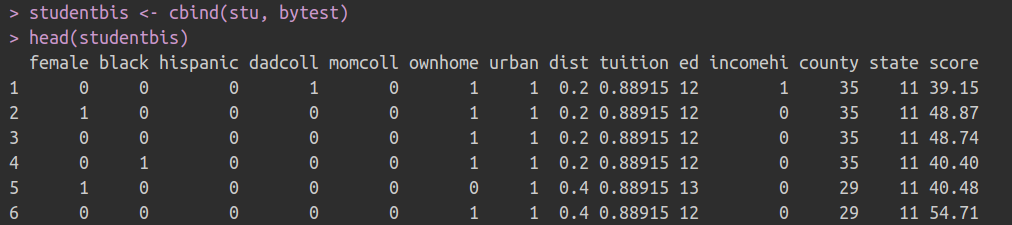
6. Extraire dans un vecteur intitulé tab7 les étudiants hispaniques dont le revenu familial est supérieur à 25.000 $ et supprimer les variables hispanic, black et incomehi. Combien y a-t-il d’individus ?



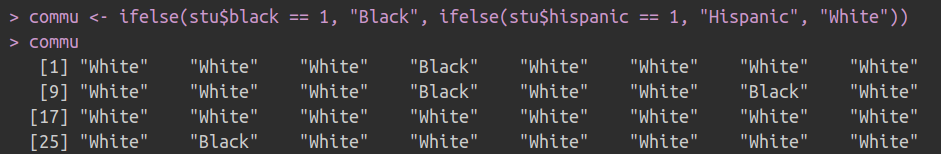


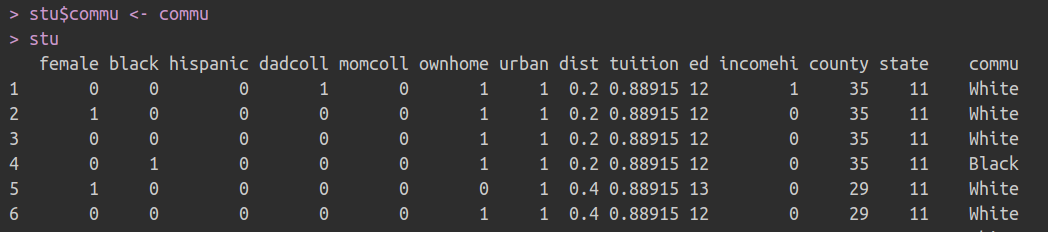
7. On souhaite ajouter au tableau les résultats de ces étudiants au Base Year Composite Test Score, un test évaluant le niveau de l’étudiant. Pour cela, importer les données du fichier bytest.txt et créer un nouveau tableau intitulé studentbis contenant les données de etudiant et celle de bytest.



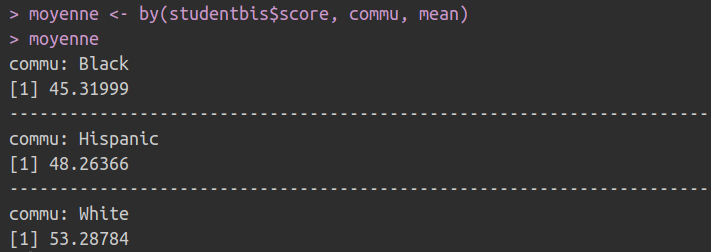


8. Construire un vecteur intitulé commu permettant d’identifier la communauté à laquelle appartient l’étudiant, en codant Hispanic pour hispanic, Black pour black et White pour ni black, ni hispanic. L’ajouter au tableau.

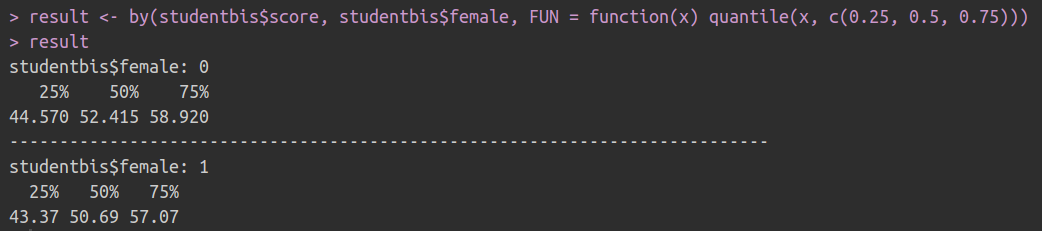




9. Pour chacune des communautés, déterminer la moyenne du bytest avec la commande suivante.

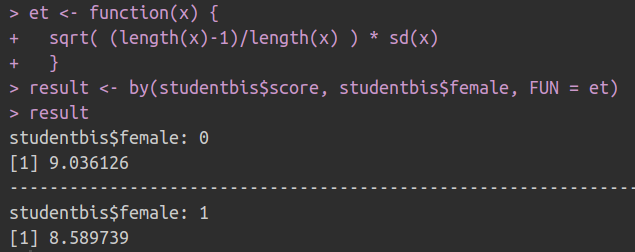


10. En utilisant la fonction by() et son aide, calculer les quartiles des bytest par sexe.



11. On souhaite obtenir les écarts type des bytest par sexe. Cependant la commande sd() calcule l’écart type corrigé. On peut créer la fonction et() calculant l’écart type avec le commande suivante.

L’utiliser pour calculer les écarts type des bytest par sexe.



12. Construire une nouvelle variable intitulée par.sup qui vaut 0 si aucun parent n’est diplômé du supérieur et 1 sinon. Calculer les déciles du bytest pour ces deux groupes.

