Statistics

AyaBH

28/11/2021

#Getting errors at 528

```
# Import required packages
library(readr)
library(plyr)
library(dplyr)
library(plotly)
library(xtable)
library(tufte)
library(summarytools)
library(dplyr)
library(vcd)
#install.packages("multcomp")
library(multcomp)
library(finalfit)
library(ggplot2)
#install.packages("pscl")
library(pscl) #McFadden , pseudo-R2 library
library(survival)
library(survminer)
```

3 Data wrangling, feature engineering

```
# Import effec data files
effec1_df <- read_csv("effec1.quest.compil.csv",
                      locale = locale(encoding = "ISO-8859-1"))
effec2_df <- read_csv("effec2.quest.compil.csv",
                      locale = locale(encoding = "ISO-8859-1"))
effec3_df <- read_csv("effec3.quest.compil.csv",
                      locale = locale(encoding = "ISO-8859-1"))

# rbind(append rows) effec data files
effec_df <- rbind.fill(effec1_df, effec2_df, effec3_df)
head(effec_df)
```

```
##   Student_ID    Gender birth.year    Country                        Diploma
## 1        221      <NA>         NA       <NA>                           <NA>
## 2      19178 une femme       1986     France  Bac+5 (Master ou équivalent)
## 3       1086 une femme       1967     France  Bac+5 (Master ou équivalent)
## 4       1948 une femme       1983  Allemagne             Bac ou équivalent
## 5      16209 une femme         NA Madagascar Bac+3 (Licence ou équivalent)
## 6       6685  un homme       1951       <NA>  Bac+5 (Master ou équivalent)
```

```
##                                                                    Formation
## 1                                                                       <NA>
## 2                                                                      Droit
## 3 Sciences sociales (économie\\, sciences politiques\\, sociologie\\, etc)
## 4                                                                      Droit
## 5 Sciences naturelles (Agronomie\\, biologie\\, physique\\, chimie\\, etc)
## 6                                                  Ingénierie et technologies
##                                       CSP
## 1                                    <NA>
## 2     Cadres et professions intellectuelles
## 3 Artisans, commerçants, chefs d'entreprise
## 4                                 Employés
## 5                Professions intermédiaires
## 6                                 Retraités
##                                How.heard
## 1                                   <NA>
## 2 par un article ou un blog sur Internet
## 3      par une communication de l'EMLYON
## 4          par une communication de Unow
## 5         par un ami ou une connaissance
## 6          par une communication de Unow
##
Exp.crea
## 1
<NA>
## 2                          Je n'ai aucune expérience en création
d'entreprise
## 3              Je suis en train de créer mon entreprise (phase de
démarrage)
## 4                          Je n'ai aucune expérience en création
d'entreprise
## 5 J<U+0092>ai un projet de création d<U+0092>entreprise (phase de
réflexion)
## 6                          Je n'ai aucune expérience en création
d'entreprise
##   Curiosity.MOOC Certif.self.sat Rencontres Certif.work Incitation
## 1           <NA>              NA       <NA>          NA         NA
## 2              4               4          4           1          4
## 3              2               1          1           1          3
## 4              1               3          2           1          1
## 5              1               4          4           1          5
## 6              1               2          1           1          1
##                Temps.Dispo                                        Exp.MOOC
## 1                     <NA>                                            <NA>
## 2 Entre une et deux heures Non, c'est ma première participation à un MOOC
## 3 Entre une et deux heures Non, c'est ma première participation à un MOOC
## 4 Entre une et deux heures            Oui, mais tous suivis partiellement
## 5 Entre une et deux heures Non, c'est ma première participation à un MOOC
## 6       Plus de six heures               Oui, dont certains intégralement
##   Completion.proba                                    Instit.brand
```

```
## 1                NA                                                    <NA>
## 2                 5                                                    <NA>
## 3                 4                                                    <NA>
## 4                 4                                                    <NA>
## 5                 5                                                    <NA>
## 6                 5 2. Oui, c<U+0092>est un paramètre très important
##
motiv.princ
## 1
<NA>
## 2
<NA>
## 3
<NA>
## 4
<NA>
## 5
<NA>
## 6 La satisfaction personnelle d<U+0092>être allé jusqu<U+0092>au bout de
la formation
##                                          diffic  encad.disp
## 1                                          <NA>        <NA>
## 2                                          <NA>        <NA>
## 3                                          <NA>        <NA>
## 4                                          <NA>        <NA>
## 5                                          <NA>        <NA>
## 6 Lenteur ou ruptures de la connexion Internet Disponibles
##                                                    How.contact
## 1                                                         <NA>
## 2                                                         <NA>
## 3                                                         <NA>
## 4                                                         <NA>
## 5                                                         <NA>
## 6 je n<U+0092>ai pas échangé avec les autres participants
##                             entour
## 1                             <NA>
## 2                             <NA>
## 3                             <NA>
## 4                             <NA>
## 5                             <NA>
## 6 Oui, des membres de ma famille
##
entour.inter
## 1
<NA>
## 2
<NA>
## 3
<NA>
## 4
```

```
<NA>
## 5
<NA>
## 6 Regardé des vidéos ensemble,S<U+0092>encourager mutuellement à
poursuivre le MOOC
##   Satisf Eval.diffic Estimated.hours
## 1     NA        <NA>            <NA>
## 2     NA        <NA>            <NA>
## 3     NA        <NA>            <NA>
## 4     NA        <NA>            <NA>
## 5     NA        <NA>            <NA>
## 6      5   Difficile De 4 à 8 heures
##
Part.labo
## 1
<NA>
## 2
<NA>
## 3
<NA>
## 4
<NA>
## 5
<NA>
## 6 Non\\, j<U+0092>ai compris ce qu<U+0092>était le Laboratoire mais je
n<U+0092>y ai pas participé
##          Plat.satisf Peer.eval.relev encad.diffic Country_HDI
Country_HDI.fin
## 1               <NA>            <NA>           NA        <NA>
<NA>
## 2               <NA>            <NA>           NA          TH
TH
## 3               <NA>            <NA>           NA          TH
TH
## 4               <NA>            <NA>           NA          TH
TH
## 5               <NA>            <NA>           NA           B
B
## 6 Très satisfaisante               3           NA        <NA>
<NA>
##                                     CSP.fin      Temps.dispo.fin
Current.Score
## 1                                      <NA>                <NA>
NA
## 2     Cadres et professions intellectuelles Moins de deux heures
NA
## 3 Artisans, commerçants, chefs d'entreprise Moins de deux heures
NA
## 4                                  Employés Moins de deux heures
NA
```

```
## 5                                    Autre Moins de deux heures
NA
## 6                                    Autre   Plus de six heures
NA
##   Section  Mot EMLyon Proba.reco EMLyon.et Certif.bin EMLYON.et age
## 1    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA
## 2    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA
## 3    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA
## 4    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA
## 5    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA
## 6    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA

# Import usages_effec data files
usages_effec1_df <- read_csv("usages.effec1.csv")
usages_effec2_df <- read_csv("usages.effec2.csv")
usages_effec3_df <- read_csv("usages.effec3.csv")

# rbind usages_effec data files
usages_effec_df <- rbind.fill(usages_effec1_df, usages_effec2_df,
                              usages_effec3_df)
head(usages_effec_df)

##   Student_ID Exam.score Exam.bin Assignment.score Assignment.bin
Quizz.1.score
## 1         28         NA        0               NA              0
NA
## 2         36         NA        0               NA              0
NA
## 3         45         NA        0               NA              0
16
## 4         83         NA        0               60              1
13
## 5         84         NA        0               NA              0
18
## 6         87         NA        0               NA              0
NA
##   Quizz.1.bin Quizz.2.score Quizz.2.bin Quizz.3.score Quizz.3.bin
Quizz.4.bin
## 1           0            NA           0            NA           0
0
## 2           0            NA           0            NA           0
0
## 3           1            20           1            18           1
1
## 4           1            20           1            18           1
1
## 5           1            20           1            NA           0
0
## 6           0            NA           0            NA           0
0
```

```
##   Quizz.4.score Quizz.5.bin Quizz.5.score Intro.MOOC Prez.sem.1 S1.L1
S1.L2
## 1            NA           0            NA         NA          1     0
0
## 2            NA           0            NA         NA          0     0
0
## 3            20           1            19         NA          1     1
1
## 4            20           1            13         NA          1     1
1
## 5            NA           0            NA         NA          1     1
1
## 6            NA           0            NA         NA          1     1
0
##   S1.L3 S1.L4 S1.L5 S1.L6 Prez.sem.2 S2.L1 S2.L2 S2.L3 S2.L4 S2.L5 S2.L6
## 1     0     0     0     0          0     0     0     0     0     0     0
## 2     0     0     0     0          0     0     0     0     0     0     0
## 3     1     1     1     1          1     1     1     1     1     1     1
## 4     0     1     1     1          1     1     1     1     1     1     0
## 5     1     1     1     1          1     0     0     0     0     0     0
## 6     0     0     0     0          0     0     0     0     0     0     0
##   Prez.sem.3 S3.L1.1 S3.L1.2 S3.L2 S3.L3 S3.L4 S3.L5 Prez.sem.4 S4.L1.1
S4.L1.2
## 1          0       0       0     0     0     0     0          0       0
0
## 2          0       0       0     0     0     0     0          0       0
0
## 3          1       1       1     1     1     1     0          1       1
1
## 4          1       1       1     1     1     1     1          1       1
1
## 5          0       1       0     0     0     0     0          0       1
0
## 6          0       0       0     0     0     0     0          0       0
0
##   S4.L2 S4.L3 S4.L4 S4.L5 Prez.sem.5 S5.L1.1 S5.L1.2 S5.L2 S5.L3 S5.L4
S5.L5
## 1     0     0     0     0          0       0       0     0     0     0
0
## 2     0     0     0     0          0       0       0     0     0     0
0
## 3     1     0     0     0          1       1       1     1     1     1
0
## 4     1     1     1     1          1       0       0     0     0     0
0
## 5     0     0     0     0          0       0       0     0     0     0
0
## 6     0     0     0     0          0       0       0     0     0     0
0
##   Post.forum.0 view.forum.0 Post.forum.1 Post.forum.1.2 view.forum.1
```

```
## 1            0            0            0              0            0
## 2            0            0            0              0            0
## 3            1            1            0              0            1
## 4            0            1            0              0            1
## 5            0            0            0              0            1
## 6            0            0            0              0            1
##   view.forum.1.2 Post.forum.2 Post.forum.2.2 view.forum.2 view.forum.2.2
## 1              0            0              0            0              0
## 2              0            0              0            0              0
## 3              1            0              0            1              1
## 4              1            0              0            0              1
## 5              0            0              0            0              0
## 6              1            0              0            0              0
##   Post.forum.3 view.forum.3 Post.forum.4 Post.forum.4.2 view.forum.4
## 1            0            0            0              0            0
## 2            0            0            0              0            0
## 3            0            1            0              0            1
## 4            0            0            0              0            1
## 5            0            0            0              0            0
## 6            0            0            0              0            0
##   view.forum.4.2 Post.forum.5 Post.forum.5.2 view.forum.5 view.forum.5.2
## 1              0            0              0            0              0
## 2              0            0              0            0              0
## 3              1            0              0            1              1
## 4              0            0              0            1              0
## 5              0            0              0            0              0
## 6              0            0              0            0              0
##   last.video last.quizz Assignment.choice Post.forum.fonc.cours
## 1          1          0                NA                    NA
## 2          0          0                NA                    NA
## 3         34          5                NA                    NA
## 4         29          5                NA                    NA
## 5         23          2                NA                    NA
## 6          2          0                NA                    NA
##   view.forum.fonc.cours
## 1                    NA
## 2                    NA
## 3                    NA
## 4                    NA
## 5                    NA
## 6                    NA
```

```
# Merge effec_df and usages_effec_df with Student_ID as key
df_no_HDI <- full_join(effec_df, usages_effec_df, by="Student_ID")
head(df_no_HDI)
```

```
##   Student_ID    Gender birth.year    Country                        Diploma
## 1        221      <NA>         NA       <NA>                           <NA>
## 2      19178 une femme       1986     France  Bac+5 (Master ou équivalent)
## 3       1086 une femme       1967     France  Bac+5 (Master ou équivalent)
```

```
## 4       1948 une femme       1983  Allemagne             Bac ou équivalent
## 5      16209 une femme         NA Madagascar Bac+3 (Licence ou équivalent)
## 6       6685  un homme       1951       <NA>  Bac+5 (Master ou équivalent)
##                                                                  Formation
## 1                                                                     <NA>
## 2                                                                    Droit
## 3 Sciences sociales (économie\\, sciences politiques\\, sociologie\\, etc)
## 4                                                                    Droit
## 5 Sciences naturelles (Agronomie\\, biologie\\, physique\\, chimie\\, etc)
## 6                                                Ingénierie et technologies
##                                       CSP
## 1                                    <NA>
## 2     Cadres et professions intellectuelles
## 3 Artisans, commerçants, chefs d'entreprise
## 4                                 Employés
## 5                Professions intermédiaires
## 6                                 Retraités
##                                How.heard
## 1                                   <NA>
## 2 par un article ou un blog sur Internet
## 3      par une communication de l'EMLYON
## 4          par une communication de Unow
## 5         par un ami ou une connaissance
## 6          par une communication de Unow
##
Exp.crea
## 1
<NA>
## 2                          Je n'ai aucune expérience en création
d'entreprise
## 3              Je suis en train de créer mon entreprise (phase de
démarrage)
## 4                          Je n'ai aucune expérience en création
d'entreprise
## 5 J<U+0092>ai un projet de création d<U+0092>entreprise (phase de
réflexion)
## 6                          Je n'ai aucune expérience en création
d'entreprise
##   Curiosity.MOOC Certif.self.sat Rencontres Certif.work Incitation
## 1           <NA>              NA       <NA>          NA         NA
## 2              4               4          4           1          4
## 3              2               1          1           1          3
## 4              1               3          2           1          1
## 5              1               4          4           1          5
## 6              1               2          1           1          1
##                Temps.Dispo                                      Exp.MOOC
## 1                    <NA>                                          <NA>
## 2 Entre une et deux heures Non, c'est ma première participation à un MOOC
## 3 Entre une et deux heures Non, c'est ma première participation à un MOOC
## 4 Entre une et deux heures           Oui, mais tous suivis partiellement
```

```
## 5 Entre une et deux heures Non, c'est ma première participation à un MOOC
## 6       Plus de six heures               Oui, dont certains intégralement
##   Completion.proba                                      Instit.brand
## 1               NA                                              <NA>
## 2                5                                              <NA>
## 3                4                                              <NA>
## 4                4                                              <NA>
## 5                5                                              <NA>
## 6                5 2. Oui, c<U+0092>est un paramètre très important
##
motiv.princ
## 1
<NA>
## 2
<NA>
## 3
<NA>
## 4
<NA>
## 5
<NA>
## 6 La satisfaction personnelle d<U+0092>être allé jusqu<U+0092>au bout de
la formation
##                                          diffic  encad.disp
## 1                                          <NA>        <NA>
## 2                                          <NA>        <NA>
## 3                                          <NA>        <NA>
## 4                                          <NA>        <NA>
## 5                                          <NA>        <NA>
## 6 Lenteur ou ruptures de la connexion Internet Disponibles
##                                                  How.contact
## 1                                                       <NA>
## 2                                                       <NA>
## 3                                                       <NA>
## 4                                                       <NA>
## 5                                                       <NA>
## 6 je n<U+0092>ai pas échangé avec les autres participants
##                            entour
## 1                            <NA>
## 2                            <NA>
## 3                            <NA>
## 4                            <NA>
## 5                            <NA>
## 6 Oui, des membres de ma famille
##
entour.inter
## 1
<NA>
## 2
<NA>
```

```
## 3
<NA>
## 4
<NA>
## 5
<NA>
## 6 Regardé des vidéos ensemble,S<U+0092>encourager mutuellement à
poursuivre le MOOC
##   Satisf Eval.diffic Estimated.hours
## 1     NA        <NA>            <NA>
## 2     NA        <NA>            <NA>
## 3     NA        <NA>            <NA>
## 4     NA        <NA>            <NA>
## 5     NA        <NA>            <NA>
## 6      5   Difficile De 4 à 8 heures
##
Part.labo
## 1
<NA>
## 2
<NA>
## 3
<NA>
## 4
<NA>
## 5
<NA>
## 6 Non\\, j<U+0092>ai compris ce qu<U+0092>était le Laboratoire mais je
n<U+0092>y ai pas participé
##          Plat.satisf Peer.eval.relev encad.diffic Country_HDI
Country_HDI.fin
## 1               <NA>            <NA>           NA        <NA>
<NA>
## 2               <NA>            <NA>           NA          TH
TH
## 3               <NA>            <NA>           NA          TH
TH
## 4               <NA>            <NA>           NA          TH
TH
## 5               <NA>            <NA>           NA           B
B
## 6 Très satisfaisante               3           NA        <NA>
<NA>
##                                     CSP.fin      Temps.dispo.fin
Current.Score
## 1                                      <NA>                <NA>
NA
## 2     Cadres et professions intellectuelles Moins de deux heures
NA
## 3 Artisans, commerçants, chefs d'entreprise Moins de deux heures
```

```
NA
## 4                                  Employés Moins de deux heures
NA
## 5                                     Autre Moins de deux heures
NA
## 6                                     Autre   Plus de six heures
NA
##   Section  Mot EMLyon Proba.reco EMLyon.et Certif.bin EMLYON.et age
Exam.score
## 1    <NA> <NA>   <NA>         NA         NA         NA      <NA>  NA
NA
## 2    <NA> <NA>   <NA>         NA         NA         NA      <NA>  NA
NA
## 3    <NA> <NA>   <NA>         NA         NA         NA      <NA>  NA
NA
## 4    <NA> <NA>   <NA>         NA         NA         NA      <NA>  NA
NA
## 5    <NA> <NA>   <NA>         NA         NA         NA      <NA>  NA
NA
## 6    <NA> <NA>   <NA>         NA         NA         NA      <NA>  NA
NA
##   Exam.bin Assignment.score Assignment.bin Quizz.1.score Quizz.1.bin
## 1        0               NA              0            NA           0
## 2        0               NA              0            NA           0
## 3        0               NA              0            11           1
## 4        0               NA              0            NA           0
## 5        0               NA              0            20           1
## 6        0               70              1            20           1
##   Quizz.2.score Quizz.2.bin Quizz.3.score Quizz.3.bin Quizz.4.bin
Quizz.4.score
## 1            NA           0            NA           0           0
NA
## 2            NA           0            NA           0           0
NA
## 3            20           1         17.33           1           1
20.00
## 4            NA           0            NA           0           0
NA
## 5            20           1         20.00           1           1
20.00
## 6            20           1         18.00           1           1
17.33
##   Quizz.5.bin Quizz.5.score Intro.MOOC Prez.sem.1 S1.L1 S1.L2 S1.L3 S1.L4
S1.L5
## 1           0            NA         NA          1     0     0     0     0
0
## 2           0            NA         NA          1     1     0     0     0
0
## 3           0            NA         NA          1     1     1     1     1
1
```

```
## 4           0            NA          NA          1     1     0     0     0
0
## 5           1            20          NA          0     0     0     0     0
0
## 6           1            19          NA          0     1     0     1     1
0
##   S1.L6 Prez.sem.2 S2.L1 S2.L2 S2.L3 S2.L4 S2.L5 S2.L6 Prez.sem.3 S3.L1.1
## 1     0          0     0     0     0     0     0     0          0       0
## 2     0          0     0     0     0     0     0     0          0       0
## 3     1          1     1     1     1     1     1     1          1       1
## 4     0          0     0     0     0     0     0     0          0       0
## 5     0          0     0     0     0     0     0     0          0       0
## 6     0          1     1     0     0     0     0     1          1       1
##   S3.L1.2 S3.L2 S3.L3 S3.L4 S3.L5 Prez.sem.4 S4.L1.1 S4.L1.2 S4.L2 S4.L3
S4.L4
## 1       0     0     0     0     0          0       0       0     0     0
0
## 2       0     0     0     0     0          0       0       0     0     0
0
## 3       1     1     1     1     1          1       1       1     1     1
1
## 4       0     0     0     0     0          0       0       0     0     0
0
## 5       0     0     0     0     0          0       0       0     0     0
0
## 6       0     0     0     0     0          0       0       0     0     0
0
##   S4.L5 Prez.sem.5 S5.L1.1 S5.L1.2 S5.L2 S5.L3 S5.L4 S5.L5 Post.forum.0
## 1     0          0       0       0     0     0     0     0            0
## 2     0          0       0       0     0     0     0     0            0
## 3     1          1       1       1     1     1     1     1            0
## 4     0          0       0       0     0     0     0     0            0
## 5     0          0       0       0     0     0     0     0            0
## 6     0          0       0       0     0     0     0     0            0
##   view.forum.0 Post.forum.1 Post.forum.1.2 view.forum.1 view.forum.1.2
## 1            0            0              0            0              0
## 2            0            0              0            0              0
## 3            0            0              0            1              1
## 4            0            0              0            0              0
## 5            0            0              0            0              0
## 6            1            0              0            1              1
##   Post.forum.2 Post.forum.2.2 view.forum.2 view.forum.2.2 Post.forum.3
## 1            0              0            0              0            0
## 2            0              0            0              0            0
## 3            0              0            0              1            0
## 4            0              0            0              0            0
## 5            0              0            0              0            0
## 6            0              0            1              1            0
##   view.forum.3 Post.forum.4 Post.forum.4.2 view.forum.4 view.forum.4.2
## 1            0            0              0            0              0
```

```
## 2            0            0             0            0             0
## 3            1            1             0            1             1
## 4            0            0             0            0             0
## 5            0            0             0            0             0
## 6            0            0             0            1             0
##   Post.forum.5 Post.forum.5.2 view.forum.5 view.forum.5.2 last.video
last.quizz
## 1            0              0            0              0          1
0
## 2            0              0            0              0          2
0
## 3            1              0            1              1         35
4
## 4            0              0            0              0          2
0
## 5            0              0            0              0          0
5
## 6            0              0            1              0         16
5
##   Assignment.choice Post.forum.fonc.cours view.forum.fonc.cours
## 1                NA                    NA                    NA
## 2                NA                    NA                    NA
## 3                NA                    NA                    NA
## 4                NA                    NA                    NA
## 5                NA                    NA                    NA
## 6                NA                    NA                    NA
```

```
# Import countries_hdi data file
#Assign headers to each column i.e Country, HDI, and index
countries_HDI_df <- read_csv("countries.HDI.csv",
                             locale = locale(encoding = "ISO-8859-1"),
                             col_names = c("Country","HDI", "Index"))
head(countries_HDI_df)
```

```
## # A tibble: 6 x 3
##   Country               HDI   Index
##   <chr>                 <chr> <dbl>
## 1 Norvège               TH        1
## 2 Australie             TH        2
## 3 Etats-Unis d'Amérique TH        3
## 4 Pays-Bas              TH        4
## 5 Allemagne             TH        5
## 6 Nouvelle-Zélande      TH        6
```

```
# Change H and M HDI to I
##Group together, for the HDI variable, the High and Medium level to create a
#new intermediate level.
levels(countries_HDI_df$HDI) <- c(levels(countries_HDI_df$HDI), "I")
countries_HDI_df$HDI[countries_HDI_df$HDI == "M"]  <- "I"
```

```r
countries_HDI_df$HDI[countries_HDI_df$HDI == "H"]  <- "I"
head(countries_HDI_df)
```

```
## # A tibble: 6 x 3
##   Country               HDI   Index
##   <chr>                 <chr> <dbl>
## 1 Norvège               TH        1
## 2 Australie             TH        2
## 3 Etats-Unis d'Amérique TH        3
## 4 Pays-Bas              TH        4
## 5 Allemagne             TH        5
## 6 Nouvelle-Zélande      TH        6
```

```r
# Merge df_no_HDI and countries_HDI_df
full_df <- full_join(df_no_HDI, countries_HDI_df[c("Country","HDI")], by =
"Country")
head(full_df)
```

```
##   Student_ID    Gender birth.year    Country                        Diploma
## 1        221      <NA>         NA       <NA>                           <NA>
## 2        221      <NA>         NA       <NA>                           <NA>
## 3      19178 une femme       1986     France  Bac+5 (Master ou équivalent)
## 4       1086 une femme       1967     France  Bac+5 (Master ou équivalent)
## 5       1948 une femme       1983  Allemagne             Bac ou équivalent
## 6      16209 une femme         NA Madagascar Bac+3 (Licence ou équivalent)
##                                                                  Formation
## 1                                                                     <NA>
## 2                                                                     <NA>
## 3                                                                    Droit
## 4 Sciences sociales (économie\\, sciences politiques\\, sociologie\\, etc)
## 5                                                                    Droit
## 6 Sciences naturelles (Agronomie\\, biologie\\, physique\\, chimie\\, etc)
##                                        CSP
## 1                                     <NA>
## 2                                     <NA>
## 3     Cadres et professions intellectuelles
## 4 Artisans, commerçants, chefs d'entreprise
## 5                                 Employés
## 6                Professions intermédiaires
##                                  How.heard
## 1                                     <NA>
## 2                                     <NA>
## 3 par un article ou un blog sur Internet
## 4      par une communication de l'EMLYON
## 5          par une communication de Unow
## 6         par un ami ou une connaissance
##
Exp.crea
## 1
<NA>
```

```
## 2
<NA>
## 3                         Je n'ai aucune expérience en création
d'entreprise
## 4              Je suis en train de créer mon entreprise (phase de
démarrage)
## 5                         Je n'ai aucune expérience en création
d'entreprise
## 6 J<U+0092>ai un projet de création d<U+0092>entreprise (phase de
réflexion)
##   Curiosity.MOOC Certif.self.sat Rencontres Certif.work Incitation
## 1           <NA>              NA       <NA>          NA         NA
## 2           <NA>              NA       <NA>          NA         NA
## 3              4               4          4           1          4
## 4              2               1          1           1          3
## 5              1               3          2           1          1
## 6              1               4          4           1          5
##                Temps.Dispo                                        Exp.MOOC
## 1                    <NA>                                            <NA>
## 2                    <NA>                                            <NA>
## 3 Entre une et deux heures Non, c'est ma première participation à un MOOC
## 4 Entre une et deux heures Non, c'est ma première participation à un MOOC
## 5 Entre une et deux heures           Oui, mais tous suivis partiellement
## 6 Entre une et deux heures Non, c'est ma première participation à un MOOC
##   Completion.proba Instit.brand motiv.princ diffic encad.disp How.contact
## 1               NA         <NA>        <NA>   <NA>       <NA>        <NA>
## 2               NA         <NA>        <NA>   <NA>       <NA>        <NA>
## 3                5         <NA>        <NA>   <NA>       <NA>        <NA>
## 4                4         <NA>        <NA>   <NA>       <NA>        <NA>
## 5                4         <NA>        <NA>   <NA>       <NA>        <NA>
## 6                5         <NA>        <NA>   <NA>       <NA>        <NA>
##   entour entour.inter Satisf Eval.diffic Estimated.hours Part.labo
Plat.satisf
## 1   <NA>         <NA>     NA        <NA>            <NA>      <NA>
<NA>
## 2   <NA>         <NA>     NA        <NA>            <NA>      <NA>
<NA>
## 3   <NA>         <NA>     NA        <NA>            <NA>      <NA>
<NA>
## 4   <NA>         <NA>     NA        <NA>            <NA>      <NA>
<NA>
## 5   <NA>         <NA>     NA        <NA>            <NA>      <NA>
<NA>
## 6   <NA>         <NA>     NA        <NA>            <NA>      <NA>
<NA>
##   Peer.eval.relev encad.diffic Country_HDI Country_HDI.fin
## 1            <NA>           NA        <NA>            <NA>
## 2            <NA>           NA        <NA>            <NA>
## 3            <NA>           NA          TH              TH
## 4            <NA>           NA          TH              TH
```

```
## 5            <NA>           NA          TH              TH
## 6            <NA>           NA           B               B
##                                    CSP.fin      Temps.dispo.fin
Current.Score
## 1                                     <NA>                <NA>
NA
## 2                                     <NA>                <NA>
NA
## 3     Cadres et professions intellectuelles Moins de deux heures
NA
## 4 Artisans, commerçants, chefs d'entreprise Moins de deux heures
NA
## 5                                 Employés Moins de deux heures
NA
## 6                                    Autre Moins de deux heures
NA
##   Section  Mot EMLyon Proba.reco EMLyon.et Certif.bin EMLYON.et age
Exam.score
## 1    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA
NA
## 2    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA
NA
## 3    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA
NA
## 4    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA
NA
## 5    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA
NA
## 6    <NA> <NA>   <NA>         NA        NA         NA      <NA>  NA
NA
##   Exam.bin Assignment.score Assignment.bin Quizz.1.score Quizz.1.bin
## 1        0               NA              0            NA           0
## 2        0               NA              0            NA           0
## 3        0               NA              0            NA           0
## 4        0               NA              0            11           1
## 5        0               NA              0            NA           0
## 6        0               NA              0            20           1
##   Quizz.2.score Quizz.2.bin Quizz.3.score Quizz.3.bin Quizz.4.bin
Quizz.4.score
## 1            NA           0            NA           0           0
NA
## 2            NA           0            NA           0           0
NA
## 3            NA           0            NA           0           0
NA
## 4            20           1         17.33           1           1
20
## 5            NA           0            NA           0           0
NA
## 6            20           1         20.00           1           1
```

```
20
##   Quizz.5.bin Quizz.5.score Intro.MOOC Prez.sem.1 S1.L1 S1.L2 S1.L3 S1.L4
S1.L5
## 1           0            NA         NA          1     0     0     0     0
0
## 2           0            NA         NA          1     0     0     0     0
0
## 3           0            NA         NA          1     1     0     0     0
0
## 4           0            NA         NA          1     1     1     1     1
1
## 5           0            NA         NA          1     1     0     0     0
0
## 6           1            20         NA          0     0     0     0     0
0
##   S1.L6 Prez.sem.2 S2.L1 S2.L2 S2.L3 S2.L4 S2.L5 S2.L6 Prez.sem.3 S3.L1.1
## 1     0          0     0     0     0     0     0     0          0       0
## 2     0          0     0     0     0     0     0     0          0       0
## 3     0          0     0     0     0     0     0     0          0       0
## 4     1          1     1     1     1     1     1     1          1       1
## 5     0          0     0     0     0     0     0     0          0       0
## 6     0          0     0     0     0     0     0     0          0       0
##   S3.L1.2 S3.L2 S3.L3 S3.L4 S3.L5 Prez.sem.4 S4.L1.1 S4.L1.2 S4.L2 S4.L3
S4.L4
## 1       0     0     0     0     0          0       0       0     0     0
0
## 2       0     0     0     0     0          0       0       0     0     0
0
## 3       0     0     0     0     0          0       0       0     0     0
0
## 4       1     1     1     1     1          1       1       1     1     1
1
## 5       0     0     0     0     0          0       0       0     0     0
0
## 6       0     0     0     0     0          0       0       0     0     0
0
##   S4.L5 Prez.sem.5 S5.L1.1 S5.L1.2 S5.L2 S5.L3 S5.L4 S5.L5 Post.forum.0
## 1     0          0       0       0     0     0     0     0            0
## 2     0          0       0       0     0     0     0     0            0
## 3     0          0       0       0     0     0     0     0            0
## 4     1          1       1       1     1     1     1     1            0
## 5     0          0       0       0     0     0     0     0            0
## 6     0          0       0       0     0     0     0     0            0
##   view.forum.0 Post.forum.1 Post.forum.1.2 view.forum.1 view.forum.1.2
## 1            0            0              0           0              0
## 2            0            0              0           0              0
## 3            0            0              0           0              0
## 4            0            0              0           1              1
## 5            0            0              0           0              0
## 6            0            0              0           0              0
```

```
##   Post.forum.2 Post.forum.2.2 view.forum.2 view.forum.2.2 Post.forum.3
## 1            0              0            0              0            0
## 2            0              0            0              0            0
## 3            0              0            0              0            0
## 4            0              0            0              1            0
## 5            0              0            0              0            0
## 6            0              0            0              0            0
##   view.forum.3 Post.forum.4 Post.forum.4.2 view.forum.4 view.forum.4.2
## 1            0            0              0            0              0
## 2            0            0              0            0              0
## 3            0            0              0            0              0
## 4            1            1              0            1              1
## 5            0            0              0            0              0
## 6            0            0              0            0              0
##   Post.forum.5 Post.forum.5.2 view.forum.5 view.forum.5.2 last.video
last.quizz
## 1            0              0            0              0          1
0
## 2            0              0            0              0          1
0
## 3            0              0            0              0          2
0
## 4            1              0            1              1         35
4
## 5            0              0            0              0          2
0
## 6            0              0            0              0          0
5
##   Assignment.choice Post.forum.fonc.cours view.forum.fonc.cours HDI
## 1                NA                    NA                    NA   B
## 2                NA                    NA                    NA   B
## 3                NA                    NA                    NA  TH
## 4                NA                    NA                    NA  TH
## 5                NA                    NA                    NA  TH
## 6                NA                    NA                    NA   B
```

```
#export full df as csv
#write.csv(full_df,"H:/Downloads/Datatsets/full_df.csv", row.names = FALSE)

full_df <- read_csv("full_df.csv", locale = locale(encoding = "ISO-8859-1"))
```

4 Describing behaviour of the courses

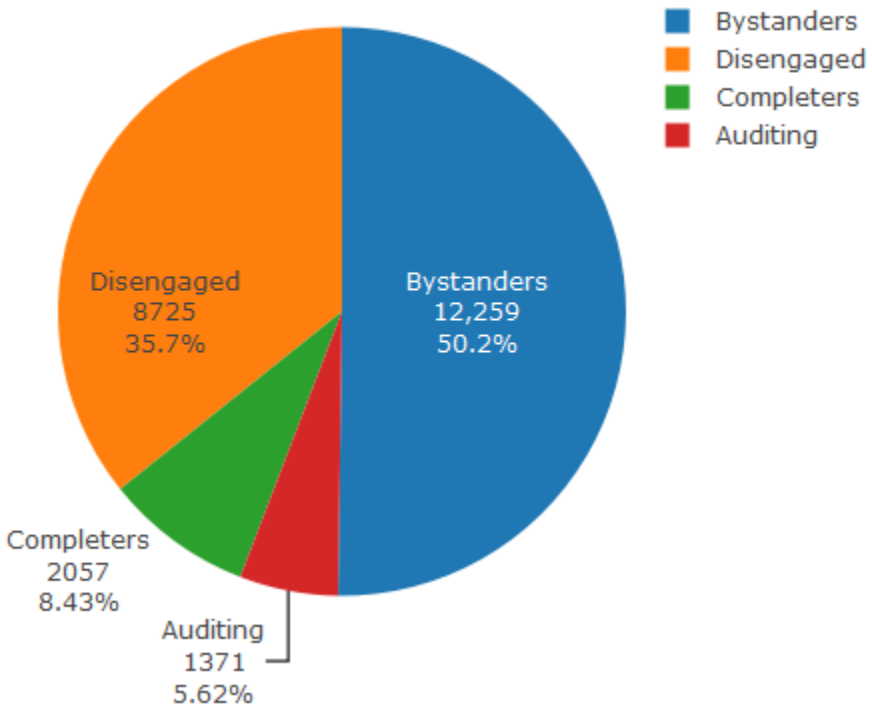
```
#completers , exam bin is used as proxy for completion
completers = nrow(full_df[which(full_df$Exam.bin == 1),])
#get number of videos for each student
full_df$n.videos <- rowSums(full_df[,60:94],na.rm=T)
#auditors
auditing = nrow(full_df %>% filter(Exam.bin == 0  & last.quizz ==0 &
Assignment.bin==0&n.videos/35 >0.1))
```

```r
#bystanders
bystanders = nrow(full_df %>% filter(Exam.bin == 0  & last.quizz ==0 &
Assignment.bin==0&n.videos/35 <=0.1) )
#disengaged learners
disengaged = nrow(full_df %>% filter(Exam.bin == 0 & (Quizz.1.bin == 1 |
Quizz.2.bin == 1 | Quizz.3.bin == 1 | Quizz.4.bin == 1 | Quizz.5.bin == 1 |
Assignment.bin == 1)))

#adding type of learners to our dataframe to use them later in survival
analysis
full_df <- full_df %>%
       mutate(learner = case_when(Exam.bin == 1 ~ "completers",
                                  Exam.bin == 0  & last.quizz ==0 &
Assignment.bin==0&n.videos/35 >0.1 ~ "auditing",
                                  Exam.bin == 0  & last.quizz ==0 &
Assignment.bin==0&n.videos/35 <=0.1 ~ "bystanders",
                                  Exam.bin == 0 & (Quizz.1.bin == 1 |
Quizz.2.bin == 1 | Quizz.3.bin == 1 | Quizz.4.bin == 1 |   Quizz.5.bin == 1 |
Assignment.bin == 1) ~ "disengaged"
       ))
head(full_df)
```

```
## # A tibble: 6 x 126
##   Student_ID Gender    birth.year Country    Diploma  Formation  CSP
How.heard
##        <dbl> <chr>          <dbl> <chr>      <chr>    <chr>      <chr>
<chr>
## 1        221 <NA>              NA <NA>       <NA>      <NA>      <NA>
<NA>
## 2        221 <NA>              NA <NA>       <NA>      <NA>      <NA>
<NA>
## 3      19178 une femme       1986 France     Bac+5 (~ "Droit"    Cadr~ par
un a~
## 4       1086 une femme       1967 France     Bac+5 (~ "Sciences~ Arti~ par
une ~
## 5       1948 une femme       1983 Allemagne  Bac ou ~ "Droit"    Empl~ par
une ~
## 6      16209 une femme         NA Madagascar Bac+3 (~ "Sciences~ Prof~ par
un a~
## # ... with 118 more variables: Exp.crea <chr>, Curiosity.MOOC <dbl>,
## #   Certif.self.sat <dbl>, Rencontres <dbl>, Certif.work <dbl>,
## #   Incitation <dbl>, Temps.Dispo <chr>, Exp.MOOC <chr>,
## #   Completion.proba <dbl>, Instit.brand <chr>, motiv.princ <chr>,
## #   diffic <chr>, encad.disp <chr>, How.contact <chr>, entour <chr>,
## #   entour.inter <chr>, Satisf <dbl>, Eval.diffic <chr>, Estimated.hours
<chr>,
## #   Part.labo <chr>, Plat.satisf <chr>, Peer.eval.relev <chr>, ...
```

```r
#create dataframe of type of learners and their values
df_prop <-
```

```
data.frame(first_column=c('Completers','Auditing','Bystanders','Disengaged'),

second_column=c(completers,auditing,bystanders,disengaged))

#rename columns
colnames(df_prop) <- c("Types","Values")

#plot pie chart in plotly
fig <- plot_ly()


fig <- df_prop %>% plot_ly(type='pie', labels=~Types,
values=~Values,textinfo="label+percent+value",

              insidetextorientation='radial')

fig
```



5.1 From Student's t-test to two-ways ANOVAs

Compare the number of views of videos between genders.

- Assuming equal variance, var = T

```
t.test(n.videos ~ Gender,data=full_df,var.equal=T)

##
##  Two Sample t-test
```

```
##
## data:  n.videos by Gender
## t = -3.544, df = 9929, p-value = 0.000396
## alternative hypothesis: true difference in means between group un homme
and group une femme is not equal to 0
## 95 percent confidence interval:
##  -1.5730798 -0.4526372
## sample estimates:
##  mean in group un homme mean in group une femme
##                15.62396                16.63681
```

- Assuming unequal variance , var = F

```
t.test(n.videos ~ Gender,data=full_df,var.equal=F)

##
##  Welch Two Sample t-test
##
## data:  n.videos by Gender
## t = -3.5174, df = 6247.4, p-value = 0.000439
## alternative hypothesis: true difference in means between group un homme
and group une femme is not equal to 0
## 95 percent confidence interval:
##  -1.5773589 -0.4483581
## sample estimates:
##  mean in group un homme mean in group une femme
##                15.62396                16.63681
```

- Which test should you use to assess whether the difference is statistically significant ?
  - comparing two independent groups

Compare the number of views of videos depending on the HDI of the country of origin. Same questions. Which test should you use to assess whether the difference is statistically significant ?

```
#HDI has more than 2 groups, so we use one-way anova
model1 <- aov(n.videos ~ HDI, data = full_df)
anova(model1)

## Analysis of Variance Table
##
## Response: n.videos
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## HDI           2 1197321  598660  6836.3 < 2.2e-16 ***
## Residuals 28373 2484641      88
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#get latex table
#print(xtable(model1))
```

- What is the difference between the two tests you just used ?
  - difference between independent t-tests and one way ANOVA

Use Gender, HDI and socioeconomic status as explaining variables (lm command in R, lm(y x1+x2)). Introduce an ANOVA table (anova(model) in R) in your report. (socioeconomic status ==> CSP)

```
model2 <- anova(lm(n.videos~HDI,full_df))
model2
```

```
## Analysis of Variance Table
##
## Response: n.videos
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## HDI           2 1197321  598660  6836.3 < 2.2e-16 ***
## Residuals 28373 2484641      88
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#get latex table of model 2
#print(xtable(model2))
```

```
#Gender and HDI- ind.variables
model3 <- anova(lm(n.videos~Gender+HDI,full_df))
model3
```

```
## Analysis of Variance Table
##
## Response: n.videos
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## Gender       1    2252    2252  13.437  0.000248 ***
## HDI          2  102869   51435 306.961 < 2.2e-16 ***
## Residuals 9833 1647626     168
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(xtable(model3))
```

```
## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Mon Dec 13 09:24:56 2021
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## Gender & 1 & 2251.52 & 2251.52 & 13.44 & 0.0002 \\
##   HDI & 2 & 102869.43 & 51434.71 & 306.96 & 0.0000 \\
##   Residuals & 9833 & 1647626.08 & 167.56 &  &  \\
##    \hline
```

```
## \end{tabular}
## \end{table}

#ind var : gender, hdi, csp
model4 <- anova(lm(n.videos~Gender+HDI+CSP,full_df))
model4

## Analysis of Variance Table
##
## Response: n.videos
##             Df  Sum Sq Mean Sq  F value    Pr(>F)
## Gender       1    2104    2104  12.6321  0.000381 ***
## HDI          2  103062   51531 309.3229 < 2.2e-16 ***
## CSP         10    8265     826   4.9609 3.293e-07 ***
## Residuals 9748 1623955     167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

print(xtable(model4))

## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Mon Dec 13 09:24:57 2021
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## Gender & 1 & 2104.43 & 2104.43 & 12.63 & 0.0004 \\
##   HDI & 2 & 103062.48 & 51531.24 & 309.32 & 0.0000 \\
##   CSP & 10 & 8264.55 & 826.46 & 4.96 & 0.0000 \\
##   Residuals & 9748 & 1623955.28 & 166.59 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

5.2 Model refinement, pairwise comparisons

Update the model, and add an interaction parameter in the it (For instance Gender*HDI in R). Use the summary of the model to see the interaction parameter.

```
model5 <- lm(n.videos~Gender+HDI+Gender*HDI,full_df)
model5

##
## Call:
## lm(formula = n.videos ~ Gender + HDI + Gender * HDI, data = full_df)
##
## Coefficients:
##           (Intercept)        Genderune femme                    HDII
##                 8.179                  1.608                 5.165
```

```
##                 HDITH   Genderune femme:HDII  Genderune femme:HDITH
##                 9.355                -3.757                 -1.458

print(summary(model5))

##
## Call:
## lm(formula = n.videos ~ Gender + HDI + Gender * HDI, data = full_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.684 -11.345  -3.535  14.465  26.821
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             8.1794     0.3838  21.313  < 2e-16 ***
## Genderune femme         1.6077     0.9881   1.627  0.10375
## HDII                    5.1653     0.6964   7.418 1.29e-13 ***
## HDITH                   9.3552     0.4250  22.014  < 2e-16 ***
## Genderune femme:HDII   -3.7571     1.3984  -2.687  0.00723 **
## Genderune femme:HDITH  -1.4578     1.0351  -1.408  0.15903
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.94 on 9831 degrees of freedom
##   (18633 observations effacées parce que manquantes)
## Multiple R-squared:  0.06069,    Adjusted R-squared:  0.06022
## F-statistic:   127 on 5 and 9831 DF,  p-value: < 2.2e-16

print(xtable(summary(model5)))

## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Mon Dec 13 09:24:57 2021
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrr}
##   \hline
##  & Estimate & Std. Error & t value & Pr($>$$|$t$|$) \\
##   \hline
## (Intercept) & 8.1794 & 0.3838 & 21.31 & 0.0000 \\
##   Genderune femme & 1.6077 & 0.9881 & 1.63 & 0.1038 \\
##   HDII & 5.1653 & 0.6964 & 7.42 & 0.0000 \\
##   HDITH & 9.3552 & 0.4250 & 22.01 & 0.0000 \\
##   Genderune femme:HDII & -3.7571 & 1.3984 & -2.69 & 0.0072 \\
##   Genderune femme:HDITH & -1.4578 & 1.0351 & -1.41 & 0.1590 \\
##    \hline
## \end{tabular}
## \end{table}
```

Use a stepwise algorithm (step command in R) to assess the performance of various versions of the model (use both forward and backward options).

```r
#convert birth year to integer
full_df$birth.year <- as.integer(full_df$birth.year)

#create age groups
full_df$birth.year[full_df$birth.year<1940] <- NA
full_df$birth.year[full_df$birth.year>2020]<- NA
#calculate age
full_df$age <- 2020-full_df$birth.year
#create seq
seq_1 = seq(0,90,by=3)
#break age into seq1
full_df$age.group <- cut(full_df$age,seq_1)

head(full_df$age.group)
```

```
## [1] <NA>    <NA>    (33,36] (51,54] (36,39] <NA>
## 30 Levels: (0,3] (3,6] (6,9] (9,12] (12,15] (15,18] (18,21] (21,24] ...
(87,90]
```

```r
#remove all Nas in the following variables
full_df_subset =
na.omit(full_df[c('Gender','HDI','n.videos','CSP','age.group')])

model6 <- lm(n.videos~Gender+HDI+CSP+age.group,full_df_subset)

step(model6,direction="both")
```

```
## Start:  AIC=48098.51
## n.videos ~ Gender + HDI + CSP + age.group
##
##             Df Sum of Sq     RSS   AIC
## - Gender     1        25 1563576 48097
## <none>                    1563552 48099
## - CSP       10      7289 1570841 48122
## - age.group 20     11226 1574778 48126
## - HDI        2     77848 1641400 48551
##
## Step:  AIC=48096.65
## n.videos ~ HDI + CSP + age.group
##
##             Df Sum of Sq     RSS   AIC
## <none>                    1563576 48097
## + Gender     1        25 1563552 48099
## - CSP       10      7266 1570842 48120
## - age.group 20     11205 1574781 48124
## - HDI        2     78993 1642569 48555

##
## Call:
## lm(formula = n.videos ~ HDI + CSP + age.group, data = full_df_subset)
```

```
##
## Coefficients:
##                                                                (Intercept)
##                                                                     3.5979
##                                                                       HDII
##                                                                     4.5352
##                                                                      HDITH
##                                                                     8.9506
##                                 CSPArtisans, commerÃ§ants, chefs d'entreprise
##                                                                     3.3687
##                                  CSPArtisans, commerçants, chefs d'entreprise
##                                                                     1.4515
##                                      CSPCadres et professions intellectuelles
##                                                                     2.5882
##                                                                  CSPEmployés
##                                                                     2.9086
##                                                  CSPEn recherche d'emploi
##                                                                     4.6907
##                                                                CSPEtudiants
##                                                                     2.7876
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
##                                                                     5.1031
##                                                                 CSPOuvriers
##                                                                     5.2653
##                                               CSPProfessions intermédiaires
##                                                                     1.1503
##                                                                CSPRetraités
##                                                                     5.2261
##                                                           age.group(21,24]
##                                                                    -0.3869
##                                                           age.group(24,27]
##                                                                     1.7080
##                                                           age.group(27,30]
##                                                                     1.1518
##                                                           age.group(30,33]
##                                                                     0.6700
##                                                           age.group(33,36]
##                                                                     0.5962
##                                                           age.group(36,39]
##                                                                     2.4483
##                                                           age.group(39,42]
##                                                                     2.6719
##                                                           age.group(42,45]
##                                                                     2.1757
##                                                           age.group(45,48]
##                                                                     2.5871
##                                                           age.group(48,51]
##                                                                     2.9964
##                                                           age.group(51,54]
##                                                                     4.3159
```

```
##                                                   age.group(54,57]
##                                                             3.0454
##                                                   age.group(57,60]
##                                                             4.3389
##                                                   age.group(60,63]
##                                                             3.0874
##                                                   age.group(63,66]
##                                                             5.0376
##                                                   age.group(66,69]
##                                                             2.1073
##                                                   age.group(69,72]
##                                                             3.4817
##                                                   age.group(72,75]
##                                                             5.4535
##                                                   age.group(75,78]
##                                                             2.3168
##                                                   age.group(78,81]
##                                                           -15.2746
```

- Age group is divided into too many parts, so we create a smaller group

```
#create second age group
full_df$age.group2 <- cut(full_df$age,c(0,30,50,80,100))

head(full_df$age.group2)

## [1] <NA>    <NA>    (30,50] (50,80] (30,50] <NA>
## Levels: (0,30] (30,50] (50,80] (80,100]

full_df_subset =
na.omit(full_df[c('Gender','HDI','n.videos','CSP','age.group','age.group2')])



model7 <- lm(n.videos~Gender+HDI+CSP+age.group2,full_df_subset)

(summary(step(model7,direction="both")))

## Start:  AIC=48104.1
## n.videos ~ Gender + HDI + CSP + age.group2
##
##              Df Sum of Sq     RSS   AIC
## - Gender      1         8 1570500 48102
## <none>                    1570492 48104
## - CSP        10      6607 1577099 48124
## - age.group2  2      4285 1574778 48126
## - HDI         2     80440 1650933 48569
##
## Step:  AIC=48102.15
## n.videos ~ HDI + CSP + age.group2
##
```

```
##               Df Sum of Sq     RSS   AIC
## <none>                     1570500 48102
## + Gender       1         8 1570492 48104
## - CSP         10      6599 1577099 48122
## - age.group2   2      4281 1574781 48124
## - HDI          2     81794 1652294 48575

##
## Call:
## lm(formula = n.videos ~ HDI + CSP + age.group2, data = full_df_subset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.318 -11.579  -3.096  14.190  27.982
##
## Coefficients:
##
Estimate
## (Intercept)
5.58380
## HDII
4.43266
## HDITH
9.00014
## CSPArtisans, commerÃ§ants, chefs d'entreprise
3.52006
## CSPArtisans, commerçants, chefs d'entreprise
1.49754
## CSPCadres et professions intellectuelles
2.62597
## CSPEmployés
2.57520
## CSPEn recherche d'emploi
4.35589
## CSPEtudiants
1.93036
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
5.00562
## CSPOuvriers
5.23556
## CSPProfessions intermédiaires
1.12703
## CSPRetraités
3.94805
## age.group2(30,50]                                                    -
0.06296
## age.group2(50,80]
1.72834
##                                                                     Std.
Error
```

```
## (Intercept)
4.12482
## HDII
0.62642
## HDITH
0.42513
## CSPArtisans, commerÃ§ants, chefs d'entreprise
4.22115
## CSPArtisans, commerçants, chefs d'entreprise
4.13288
## CSPCadres et professions intellectuelles
4.10892
## CSPEmployés
4.12068
## CSPEn recherche d'emploi
4.11958
## CSPEtudiants
4.12634
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
4.26907
## CSPOuvriers
5.06697
## CSPProfessions intermédiaires
4.18112
## CSPRetraités
4.36406
## age.group2(30,50]
0.49936
## age.group2(50,80]
0.59254
##                                                                      t
value
## (Intercept)
1.354
## HDII
7.076
## HDITH
21.170
## CSPArtisans, commerÃ§ants, chefs d'entreprise
0.834
## CSPArtisans, commerçants, chefs d'entreprise
0.362
## CSPCadres et professions intellectuelles
0.639
## CSPEmployés
0.625
## CSPEn recherche d'emploi
1.057
## CSPEtudiants
0.468
```

```
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
1.173
## CSPOuvriers
1.033
## CSPProfessions intermédiaires
0.270
## CSPRetraités
0.905
## age.group2(30,50]                                                    -
0.126
## age.group2(50,80]
2.917
##
Pr(>|t|)
## (Intercept)
0.17586
## HDII                                                                1.59e-
12
## HDITH                                                                < 2e-
16
## CSPArtisans, commerÃ§ants, chefs d'entreprise
0.40435
## CSPArtisans, commerçants, chefs d'entreprise
0.71710
## CSPCadres et professions intellectuelles
0.52278
## CSPEmployés
0.53202
## CSPEn recherche d'emploi
0.29037
## CSPEtudiants
0.63993
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
0.24101
## CSPOuvriers
0.30150
## CSPProfessions intermédiaires
0.78751
## CSPRetraités
0.36566
## age.group2(30,50]
0.89968
## age.group2(50,80]
0.00354
##
## (Intercept)
## HDII                                                                ***
## HDITH                                                               ***
## CSPArtisans, commerÃ§ants, chefs d'entreprise
## CSPArtisans, commerçants, chefs d'entreprise
```

```
## CSPCadres et professions intellectuelles
## CSPEmployés
## CSPEn recherche d'emploi
## CSPEtudiants
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
## CSPOuvriers
## CSPProfessions intermédiaires
## CSPRetraités
## age.group2(30,50]
## age.group2(50,80]                                                    **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.94 on 9375 degrees of freedom
## Multiple R-squared:  0.06277,    Adjusted R-squared:  0.06137
## F-statistic: 44.85 on 14 and 9375 DF,  p-value: < 2.2e-16

#create second age group
full_df$age.group2 <- cut(full_df$age,c(0,30,50,80,100))

head(full_df$age.group2)

## [1] <NA>    <NA>    (30,50] (50,80] (30,50] <NA>
## Levels: (0,30] (30,50] (50,80] (80,100]

#create subset for linear model

full_df_subset =
na.omit(full_df[c('Gender','HDI','n.videos','CSP','age.group','age.group2')])
#create linear model for HDI,CSP,
model7 <- lm(n.videos~Gender+HDI+CSP+age.group2,full_df_subset)

(summary(step(model7,direction="both")))

## Start:  AIC=48104.1
## n.videos ~ Gender + HDI + CSP + age.group2
##
##              Df Sum of Sq     RSS   AIC
## - Gender      1         8 1570500 48102
## <none>                    1570492 48104
## - CSP        10      6607 1577099 48124
## - age.group2  2      4285 1574778 48126
## - HDI         2     80440 1650933 48569
##
## Step:  AIC=48102.15
## n.videos ~ HDI + CSP + age.group2
##
##              Df Sum of Sq     RSS   AIC
## <none>                    1570500 48102
## + Gender      1         8 1570492 48104
```

```
## - CSP        10      6599 1577099 48122
## - age.group2  2      4281 1574781 48124
## - HDI         2     81794 1652294 48575

##
## Call:
## lm(formula = n.videos ~ HDI + CSP + age.group2, data = full_df_subset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.318 -11.579  -3.096  14.190  27.982
##
## Coefficients:
##
Estimate
## (Intercept)
5.58380
## HDII
4.43266
## HDITH
9.00014
## CSPArtisans, commerÃ§ants, chefs d'entreprise
3.52006
## CSPArtisans, commerçants, chefs d'entreprise
1.49754
## CSPCadres et professions intellectuelles
2.62597
## CSPEmployés
2.57520
## CSPEn recherche d'emploi
4.35589
## CSPEtudiants
1.93036
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
5.00562
## CSPOuvriers
5.23556
## CSPProfessions intermédiaires
1.12703
## CSPRetraités
3.94805
## age.group2(30,50]                                                    -
0.06296
## age.group2(50,80]
1.72834
##                                                                      Std.
Error
## (Intercept)
4.12482
## HDII
```

```
0.62642
## HDITH
0.42513
## CSPArtisans, commerÃ§ants, chefs d'entreprise
4.22115
## CSPArtisans, commerçants, chefs d'entreprise
4.13288
## CSPCadres et professions intellectuelles
4.10892
## CSPEmployés
4.12068
## CSPEn recherche d'emploi
4.11958
## CSPEtudiants
4.12634
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
4.26907
## CSPOuvriers
5.06697
## CSPProfessions intermédiaires
4.18112
## CSPRetraités
4.36406
## age.group2(30,50]
0.49936
## age.group2(50,80]
0.59254
##                                                                    t
value
## (Intercept)
1.354
## HDII
7.076
## HDITH
21.170
## CSPArtisans, commerÃ§ants, chefs d'entreprise
0.834
## CSPArtisans, commerçants, chefs d'entreprise
0.362
## CSPCadres et professions intellectuelles
0.639
## CSPEmployés
0.625
## CSPEn recherche d'emploi
1.057
## CSPEtudiants
0.468
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
1.173
## CSPOuvriers
```

```
1.033
## CSPProfessions intermédiaires
0.270
## CSPRetraités
0.905
## age.group2(30,50]                                                    -
0.126
## age.group2(50,80]
2.917
##
Pr(>|t|)
## (Intercept)
0.17586
## HDII                                                                1.59e-
12
## HDITH                                                                < 2e-
16
## CSPArtisans, commerÃ§ants, chefs d'entreprise
0.40435
## CSPArtisans, commerçants, chefs d'entreprise
0.71710
## CSPCadres et professions intellectuelles
0.52278
## CSPEmployés
0.53202
## CSPEn recherche d'emploi
0.29037
## CSPEtudiants
0.63993
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
0.24101
## CSPOuvriers
0.30150
## CSPProfessions intermédiaires
0.78751
## CSPRetraités
0.36566
## age.group2(30,50]
0.89968
## age.group2(50,80]
0.00354
##
## (Intercept)
## HDII                                                                ***
## HDITH                                                               ***
## CSPArtisans, commerÃ§ants, chefs d'entreprise
## CSPArtisans, commerçants, chefs d'entreprise
## CSPCadres et professions intellectuelles
## CSPEmployés
## CSPEn recherche d'emploi
```

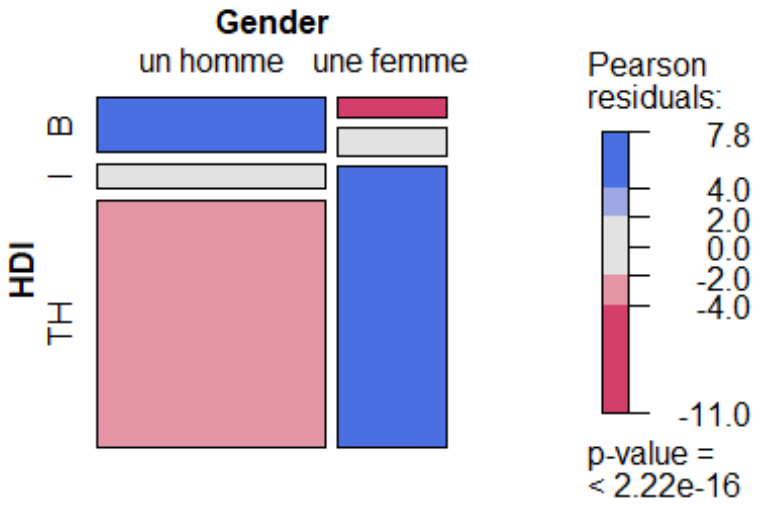
```
## CSPEtudiants
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
## CSPOuvriers
## CSPProfessions intermédiaires
## CSPRetraités
## age.group2(30,50]
## age.group2(50,80]                                                    **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.94 on 9375 degrees of freedom
## Multiple R-squared:  0.06277,    Adjusted R-squared:  0.06137
## F-statistic: 44.85 on 14 and 9375 DF,  p-value: < 2.2e-16

#latex table for figure
print(xtable((summary(step(model7,direction="both")))))

## Start:  AIC=48104.1
## n.videos ~ Gender + HDI + CSP + age.group2
##
##              Df Sum of Sq     RSS   AIC
## - Gender      1         8 1570500 48102
## <none>                    1570492 48104
## - CSP        10      6607 1577099 48124
## - age.group2  2      4285 1574778 48126
## - HDI         2     80440 1650933 48569
##
## Step:  AIC=48102.15
## n.videos ~ HDI + CSP + age.group2
##
##              Df Sum of Sq     RSS   AIC
## <none>                    1570500 48102
## + Gender      1         8 1570492 48104
## - CSP        10      6599 1577099 48122
## - age.group2  2      4281 1574781 48124
## - HDI         2     81794 1652294 48575
## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Mon Dec 13 09:24:59 2021
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrr}
##   \hline
##  & Estimate & Std. Error & t value & Pr($>$$|$t$|$) \\
##   \hline
## (Intercept) & 5.5838 & 4.1248 & 1.35 & 0.1759 \\
##   HDII & 4.4327 & 0.6264 & 7.08 & 0.0000 \\
##   HDITH & 9.0001 & 0.4251 & 21.17 & 0.0000 \\
##   CSPArtisans, commerÃ§ants, chefs d'entreprise & 3.5201 & 4.2212 & 0.83 &
0.4044 \\
##   CSPArtisans, commerçants, chefs d'entreprise & 1.4975 & 4.1329 & 0.36 &
```

```
0.7171 \\
##   CSPCadres et professions intellectuelles & 2.6260 & 4.1089 & 0.64 &
0.5228 \\
##   CSPEmployés & 2.5752 & 4.1207 & 0.62 & 0.5320 \\
##   CSPEn recherche d'emploi & 4.3559 & 4.1196 & 1.06 & 0.2904 \\
##   CSPEtudiants & 1.9304 & 4.1263 & 0.47 & 0.6399 \\
##   CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) &
5.0056 & 4.2691 & 1.17 & 0.2410 \\
##   CSPOuvriers & 5.2356 & 5.0670 & 1.03 & 0.3015 \\
##   CSPProfessions intermédiaires & 1.1270 & 4.1811 & 0.27 & 0.7875 \\
##   CSPRetraités & 3.9481 & 4.3641 & 0.90 & 0.3657 \\
##   age.group2(30,50] & -0.0630 & 0.4994 & -0.13 & 0.8997 \\
##   age.group2(50,80] & 1.7283 & 0.5925 & 2.92 & 0.0035 \\
##    \hline
## \end{tabular}
## \end{table}
```

- Assess the colinearity of all three independant variables of the last model (excluding interaction parameters). To do that, use a chi-test between HDI and Gender, produce a mosaic plot and propose its interpretation (look for residuals below -2 or above 2).
  - referring to the linear model of n.videos ~ Gender + HDI + CSP

```
#references
#https://statsandr.com/blog/chi-square-test-of-independence-in-r/
#http://www.sthda.com/english/wiki/chi-square-test-of-independence-in-r
#For interpretation purposes

full_df_subset2 = na.omit(full_df[c('Gender','HDI','n.videos','CSP')])


chisq <- chisq.test(table(full_df_subset2$Gender,full_df_subset2$HDI))
chisq

##
##  Pearson's Chi-squared test
##
## data:  table(full_df_subset2$Gender, full_df_subset2$HDI)
## X-squared = 215.1, df = 2, p-value < 2.2e-16

#install.packages('summarytools')

# fourth method:
full_df_subset2 %$%
  ctable(Gender, HDI,
    prop = "r", chisq = TRUE, headings = FALSE
  ) %>%
  print(
    method = "render",
    style = "rmarkdown",
```

```
    footnote = NA
  )

mosaic(~ Gender + HDI,
  direction = c("v", "h"),
  data = full_df_subset2,
  shade = TRUE
)
```



Use Tukey HSD, and propose a table, to see the pairwise differences between learners of different socioeconomic status.

```
model8 <- aov(n.videos~age.group2, data=full_df_subset)



TukeyHSD(model8, conf.level=.95)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = n.videos ~ age.group2, data = full_df_subset)
##
## $age.group2
##                       diff       lwr       upr     p adj
## (30,50]-(0,30]  -0.4011337 -1.265986 0.4637187 0.5220194
## (50,80]-(0,30]   2.6541764  1.620870 3.6874825 0.0000000
## (50,80]-(30,50]  3.0553101  2.243034 3.8675866 0.0000000
```

```r
#need to resize plot
plot(TukeyHSD(model8, conf.level=.95),las=3)
```


```r
#new model with gender, hdi, csp and age group 2
model9 <- aov(n.videos~Gender+HDI+CSP+age.group2, data=full_df_subset)


TukeyHSD(model9, conf.level=.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = n.videos ~ Gender + HDI + CSP + age.group2, data =
full_df_subset)
##
## $Gender
##                         diff       lwr      upr     p adj
## une femme-un homme 0.901854 0.3430752 1.460633 0.0015625
##
## $HDI
##          diff      lwr      upr p adj
## I-B  4.217115 2.762443 5.671787     0
## TH-B 8.997273 8.037478 9.957068     0
## TH-I 4.780158 3.580542 5.979773     0
##
## $CSP
##
```

```
diff
## Artisans, commerÃ§ants, chefs d'entreprise-Agriculteurs-exploitants
3.913083483
## Artisans, commerçants, chefs d'entreprise-Agriculteurs-exploitants
1.840001788
## Cadres et professions intellectuelles-Agriculteurs-exploitants
2.871135591
## Employés-Agriculteurs-exploitants
2.498773346
## En recherche d'emploi-Agriculteurs-exploitants
4.404178374
## Etudiants-Agriculteurs-exploitants
1.682999824
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-
Agriculteurs-exploitants                    5.202666120
## Ouvriers-Agriculteurs-exploitants
5.212612562
## Professions intermédiaires-Agriculteurs-exploitants
1.363836560
## Retraités-Agriculteurs-exploitants
5.414356605
## Artisans, commerçants, chefs d'entreprise-Artisans, commerÃ§ants, chefs
d'entreprise                         -2.073081694
## Cadres et professions intellectuelles-Artisans, commerÃ§ants, chefs
d'entreprise                              -1.041947892
## Employés-Artisans, commerÃ§ants, chefs d'entreprise
-1.414310136
## En recherche d'emploi-Artisans, commerÃ§ants, chefs d'entreprise
0.491094892
## Etudiants-Artisans, commerÃ§ants, chefs d'entreprise
-2.230083658
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Artisans,
commerÃ§ants, chefs d'entreprise  1.289582637
## Ouvriers-Artisans, commerÃ§ants, chefs d'entreprise
1.299529080
## Professions intermédiaires-Artisans, commerÃ§ants, chefs d'entreprise
-2.549246923
## Retraités-Artisans, commerÃ§ants, chefs d'entreprise
1.501273122
## Cadres et professions intellectuelles-Artisans, commerçants, chefs
d'entreprise                                1.031133802
## Employés-Artisans, commerçants, chefs d'entreprise
0.658771558
## En recherche d'emploi-Artisans, commerçants, chefs d'entreprise
2.564176586
## Etudiants-Artisans, commerçants, chefs d'entreprise
-0.157001964
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Artisans,
commerçants, chefs d'entreprise   3.362664332
## Ouvriers-Artisans, commerçants, chefs d'entreprise
```

```
3.372610774
## Professions intermédiaires-Artisans, commerçants, chefs d'entreprise
-0.476165228
## Retraités-Artisans, commerçants, chefs d'entreprise
3.574354816
## Employés-Cadres et professions intellectuelles
-0.372362244
## En recherche d'emploi-Cadres et professions intellectuelles
1.533042784
## Etudiants-Cadres et professions intellectuelles
-1.188135766
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Cadres et
professions intellectuelles       2.331530529
## Ouvriers-Cadres et professions intellectuelles
2.341476972
## Professions intermédiaires-Cadres et professions intellectuelles
-1.507299031
## Retraités-Cadres et professions intellectuelles
2.543221014
## En recherche d'emploi-Employés
1.905405028
## Etudiants-Employés
-0.815773522
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Employés
2.703892773
## Ouvriers-Employés
2.713839216
## Professions intermédiaires-Employés
-1.134936787
## Retraités-Employés
2.915583258
## Etudiants-En recherche d'emploi
-2.721178550
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-En
recherche d'emploi                        0.798487745
## Ouvriers-En recherche d'emploi
0.808434188
## Professions intermédiaires-En recherche d'emploi
-3.040341815
## Retraités-En recherche d'emploi
1.010178230
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Etudiants
3.519666296
## Ouvriers-Etudiants
3.529612738
## Professions intermédiaires-Etudiants
-0.319163264
## Retraités-Etudiants
3.731356780
## Ouvriers-Inactif (autre que étudiant, retraité, ou en recherche d'emploi)
```

```
0.009946443
## Professions intermédiaires-Inactif (autre que étudiant, retraité, ou en
recherche d'emploi)                 -3.838829560
## Retraités-Inactif (autre que étudiant, retraité, ou en recherche d'emploi)
0.211690485
## Professions intermédiaires-Ouvriers
-3.848776003
## Retraités-Ouvriers
0.201744042
## Retraités-Professions intermédiaires
4.050520045
##
lwr
## Artisans, commerÃ§ants, chefs d'entreprise-Agriculteurs-exploitants
-9.64006919
## Artisans, commerçants, chefs d'entreprise-Agriculteurs-exploitants
-11.43117022
## Cadres et professions intellectuelles-Agriculteurs-exploitants
-10.32325246
## Employés-Agriculteurs-exploitants
-10.75179878
## En recherche d'emploi-Agriculteurs-exploitants
-8.83154482
## Etudiants-Agriculteurs-exploitants
-11.52662306
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-
Agriculteurs-exploitants                    -8.50870229
## Ouvriers-Agriculteurs-exploitants
-11.06759869
## Professions intermédiaires-Agriculteurs-exploitants
-12.06293635
## Retraités-Agriculteurs-exploitants
-8.57247335
## Artisans, commerçants, chefs d'entreprise-Artisans, commerÃ§ants, chefs
d'entreprise                          -5.61020546
## Cadres et professions intellectuelles-Artisans, commerÃ§ants, chefs
d'entreprise                               -4.27909388
## Employés-Artisans, commerÃ§ants, chefs d'entreprise
-4.87334175
## En recherche d'emploi-Artisans, commerÃ§ants, chefs d'entreprise
-2.91061152
## Etudiants-Artisans, commerÃ§ants, chefs d'entreprise
-5.52877656
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Artisans,
commerÃ§ants, chefs d'entreprise  -3.64892541
## Ouvriers-Artisans, commerÃ§ants, chefs d'entreprise
-8.77184204
## Professions intermédiaires-Artisans, commerÃ§ants, chefs d'entreprise
-6.63158313
## Retraités-Artisans, commerÃ§ants, chefs d'entreprise
```

```
-4.15722936
## Cadres et professions intellectuelles-Artisans, commerçants, chefs
d'entreprise                                -0.67625401
## Employés-Artisans, commerçants, chefs d'entreprise
-1.43907465
## En recherche d'emploi-Artisans, commerçants, chefs d'entreprise
0.56226113
## Etudiants-Artisans, commerçants, chefs d'entreprise
-1.97838437
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Artisans,
commerçants, chefs d'entreprise   -0.73915487
## Ouvriers-Artisans, commerçants, chefs d'entreprise
-6.31596886
## Professions intermédiaires-Artisans, commerçants, chefs d'entreprise
-3.49303909
## Retraités-Artisans, commerçants, chefs d'entreprise
-1.37081549
## Employés-Cadres et professions intellectuelles
-1.91146208
## En recherche d'emploi-Cadres et professions intellectuelles
0.12750548
## Etudiants-Cadres et professions intellectuelles
-2.32184802
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Cadres et
professions intellectuelles       -1.51460914
## Ouvriers-Cadres et professions intellectuelles
-7.24165638
## Professions intermédiaires-Cadres et professions intellectuelles
-4.16613289
## Retraités-Cadres et professions intellectuelles
-2.19202755
## En recherche d'emploi-Employés
0.04494254
## Etudiants-Employés
-2.48043235
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Employés
-1.33077895
## Ouvriers-Employés
-6.94650397
## Professions intermédiaires-Employés
-4.05986124
## Retraités-Employés
-1.97403477
## Etudiants-En recherche d'emploi
-4.26318887
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-En
recherche d'emploi                        -3.18714680
## Ouvriers-En recherche d'emploi
-8.83153144
## Professions intermédiaires-En recherche d'emploi
```

```
-5.89724396
## Retraités-En recherche d'emploi
-3.83905586
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Etudiants
-0.37841670
## Ouvriers-Etudiants
-6.07448564
## Professions intermédiaires-Etudiants
-3.05259668
## Retraités-Etudiants
-1.04617811
## Ouvriers-Inactif (autre que étudiant, retraité, ou en recherche d'emploi)
-10.27335033
## Professions intermédiaires-Inactif (autre que étudiant, retraité, ou en
recherche d'emploi)                  -8.41912924
## Retraités-Inactif (autre que étudiant, retraité, ou en recherche d'emploi)
-5.81593977
## Professions intermédiaires-Ouvriers
-13.74942243
## Retraités-Ouvriers
-10.44607175
## Retraités-Professions intermédiaires
-1.29821193
##
upr
## Artisans, commerÃ§ants, chefs d'entreprise-Agriculteurs-exploitants
17.46623616
## Artisans, commerçants, chefs d'entreprise-Agriculteurs-exploitants
15.11117380
## Cadres et professions intellectuelles-Agriculteurs-exploitants
16.06552364
## Employés-Agriculteurs-exploitants
15.74934547
## En recherche d'emploi-Agriculteurs-exploitants
17.63990157
## Etudiants-Agriculteurs-exploitants
14.89262271
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-
Agriculteurs-exploitants                   18.91403453
## Ouvriers-Agriculteurs-exploitants
21.49282381
## Professions intermédiaires-Agriculteurs-exploitants
14.79060947
## Retraités-Agriculteurs-exploitants
19.40118656
## Artisans, commerçants, chefs d'entreprise-Artisans, commerÃ§ants, chefs
d'entreprise                          1.46404207
## Cadres et professions intellectuelles-Artisans, commerÃ§ants, chefs
d'entreprise                             2.19519809
## Employés-Artisans, commerÃ§ants, chefs d'entreprise
```

```
2.04472148
## En recherche d'emploi-Artisans, commerÃ§ants, chefs d'entreprise
3.89280130
## Etudiants-Artisans, commerÃ§ants, chefs d'entreprise
1.06860924
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Artisans,
commerÃ§ants, chefs d'entreprise  6.22809069
## Ouvriers-Artisans, commerÃ§ants, chefs d'entreprise
11.37090020
## Professions intermédiaires-Artisans, commerÃ§ants, chefs d'entreprise
1.53308928
## Retraités-Artisans, commerÃ§ants, chefs d'entreprise
7.15977560
## Cadres et professions intellectuelles-Artisans, commerçants, chefs
d'entreprise                                2.73852162
## Employés-Artisans, commerçants, chefs d'entreprise
2.75661777
## En recherche d'emploi-Artisans, commerçants, chefs d'entreprise
4.56609205
## Etudiants-Artisans, commerçants, chefs d'entreprise
1.66438044
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Artisans,
commerçants, chefs d'entreprise   7.46448353
## Ouvriers-Artisans, commerçants, chefs d'entreprise
13.06119041
## Professions intermédiaires-Artisans, commerçants, chefs d'entreprise
2.54070864
## Retraités-Artisans, commerçants, chefs d'entreprise
8.51952512
## Employés-Cadres et professions intellectuelles
1.16673759
## En recherche d'emploi-Cadres et professions intellectuelles
2.93858008
## Etudiants-Cadres et professions intellectuelles
-0.05442352
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Cadres et
professions intellectuelles       6.17767020
## Ouvriers-Cadres et professions intellectuelles
11.92461033
## Professions intermédiaires-Cadres et professions intellectuelles
1.15153483
## Retraités-Cadres et professions intellectuelles
7.27846957
## En recherche d'emploi-Employés
3.76586752
## Etudiants-Employés
0.84888531
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Employés
6.73856450
## Ouvriers-Employés
```

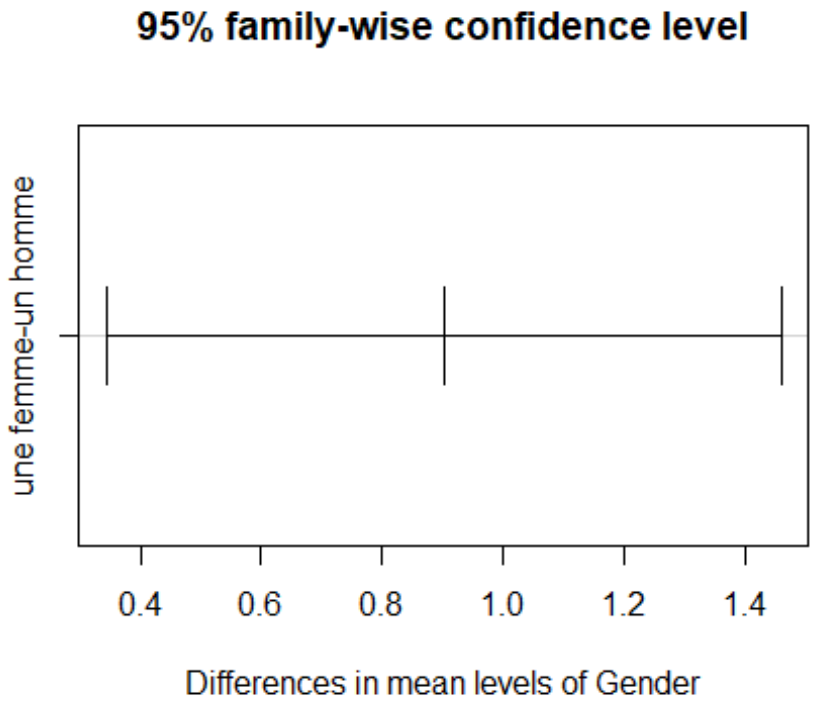
```
12.37418240
## Professions intermédiaires-Employés
1.78998767
## Retraités-Employés
7.80520129
## Etudiants-En recherche d'emploi
-1.17916823
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-En
recherche d'emploi                        4.78412229
## Ouvriers-En recherche d'emploi
10.44839981
## Professions intermédiaires-En recherche d'emploi
-0.18343967
## Retraités-En recherche d'emploi
5.85941232
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Etudiants
7.41774930
## Ouvriers-Etudiants
13.13371112
## Professions intermédiaires-Etudiants
2.41427015
## Retraités-Etudiants
8.50889167
## Ouvriers-Inactif (autre que étudiant, retraité, ou en recherche d'emploi)
10.29324322
## Professions intermédiaires-Inactif (autre que étudiant, retraité, ou en
recherche d'emploi)                   0.74147012
## Retraités-Inactif (autre que étudiant, retraité, ou en recherche d'emploi)
6.23932074
## Professions intermédiaires-Ouvriers
6.05187042
## Retraités-Ouvriers
10.84955983
## Retraités-Professions intermédiaires
9.39925202
##
p adj
## Artisans, commerÃ§ants, chefs d'entreprise-Agriculteurs-exploitants
0.9977157
## Artisans, commerçants, chefs d'entreprise-Agriculteurs-exploitants
0.9999972
## Cadres et professions intellectuelles-Agriculteurs-exploitants
0.9998079
## Employés-Agriculteurs-exploitants
0.9999484
## En recherche d'emploi-Agriculteurs-exploitants
0.9927314
## Etudiants-Agriculteurs-exploitants
0.9999988
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-
```

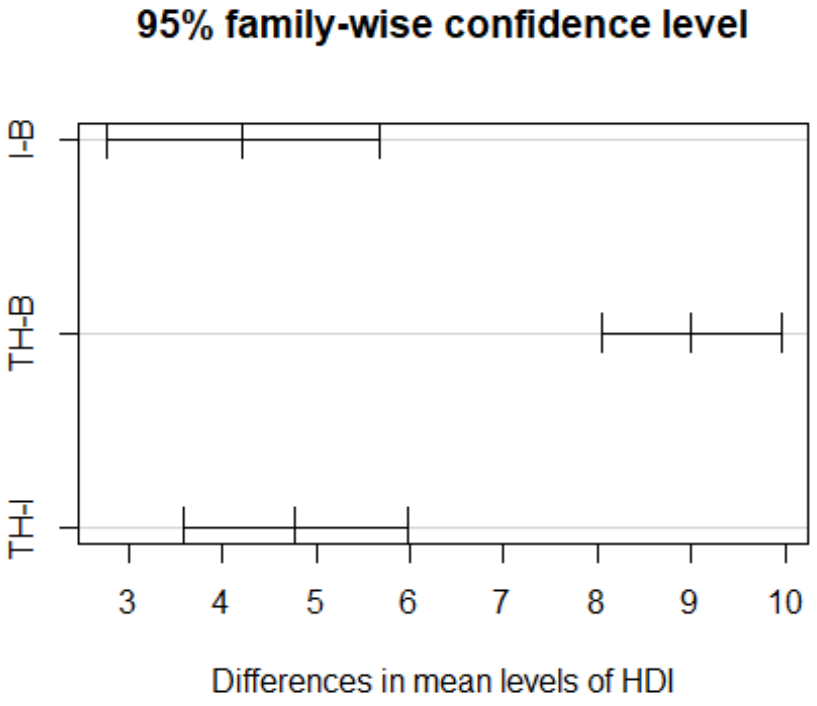
```
Agriculteurs-exploitants                    0.9802095
## Ouvriers-Agriculteurs-exploitants
0.9946531
## Professions intermédiaires-Agriculteurs-exploitants
0.9999999
## Retraités-Agriculteurs-exploitants
0.9771064
## Artisans, commerçants, chefs d'entreprise-Artisans, commerÃ§ants, chefs
d'entreprise                          0.7261598
## Cadres et professions intellectuelles-Artisans, commerÃ§ants, chefs
d'entreprise                               0.9944218
## Employés-Artisans, commerÃ§ants, chefs d'entreprise
0.9662420
## En recherche d'emploi-Artisans, commerÃ§ants, chefs d'entreprise
0.9999959
## Etudiants-Artisans, commerÃ§ants, chefs d'entreprise
0.5217580
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Artisans,
commerÃ§ants, chefs d'entreprise 0.9990307
## Ouvriers-Artisans, commerÃ§ants, chefs d'entreprise
0.9999986
## Professions intermédiaires-Artisans, commerÃ§ants, chefs d'entreprise
0.6416869
## Retraités-Artisans, commerÃ§ants, chefs d'entreprise
0.9988878
## Cadres et professions intellectuelles-Artisans, commerçants, chefs
d'entreprise                                0.6877522
## Employés-Artisans, commerçants, chefs d'entreprise
0.9954297
## En recherche d'emploi-Artisans, commerçants, chefs d'entreprise
0.0018689
## Etudiants-Artisans, commerçants, chefs d'entreprise
1.0000000
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Artisans,
commerçants, chefs d'entreprise  0.2285993
## Ouvriers-Artisans, commerçants, chefs d'entreprise
0.9896610
## Professions intermédiaires-Artisans, commerçants, chefs d'entreprise
0.9999903
## Retraités-Artisans, commerçants, chefs d'entreprise
0.4152139
## Employés-Cadres et professions intellectuelles
0.9995034
## En recherche d'emploi-Cadres et professions intellectuelles
0.0193747
## Etudiants-Cadres et professions intellectuelles
0.0307114
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Cadres et
professions intellectuelles      0.6827337
## Ouvriers-Cadres et professions intellectuelles
```

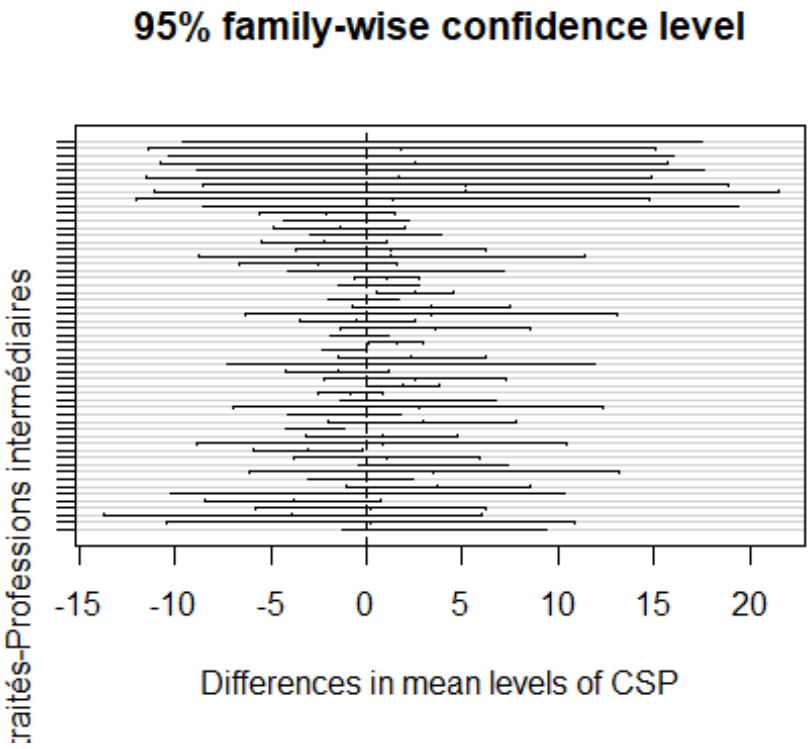
```
0.9994582
## Professions intermédiaires-Cadres et professions intellectuelles
0.7653268
## Retraités-Cadres et professions intellectuelles
0.8207023
## En recherche d'emploi-Employés
0.0392853
## Etudiants-Employés
0.8919294
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Employés
0.5354995
## Ouvriers-Employés
0.9981865
## Professions intermédiaires-Employés
0.9767106
## Retraités-Employés
0.7044271
## Etudiants-En recherche d'emploi
0.0000008
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-En
recherche d'emploi                      0.9999097
## Ouvriers-En recherche d'emploi
1.0000000
## Professions intermédiaires-En recherche d'emploi
0.0258627
## Retraités-En recherche d'emploi
0.9998708
## Inactif (autre que étudiant, retraité, ou en recherche d'emploi)-Etudiants
0.1208022
## Ouvriers-Etudiants
0.9843810
## Professions intermédiaires-Etudiants
0.9999995
## Retraités-Etudiants
0.2959356
## Ouvriers-Inactif (autre que étudiant, retraité, ou en recherche d'emploi)
1.0000000
## Professions intermédiaires-Inactif (autre que étudiant, retraité, ou en
recherche d'emploi)                 0.2005671
## Retraités-Inactif (autre que étudiant, retraité, ou en recherche d'emploi)
1.0000000
## Professions intermédiaires-Ouvriers
0.9763980
## Retraités-Ouvriers
1.0000000
## Retraités-Professions intermédiaires
0.3421942
##
## $age.group2
##                       diff        lwr       upr     p adj
```

```
## (30,50]-(0,30]  -0.3617665 -1.2031265 0.4795936 0.5718639
## (50,80]-(0,30]   1.2487501  0.2435121 2.2539880 0.0100717
## (50,80]-(30,50]  1.6105165  0.8203042 2.4007288 0.0000054
```

```
plot(TukeyHSD(model9, conf.level=.95))
```


95% family-wise confidence level
une femme-un homme
0.4
0.6
0.8
1.0
1.2
1.4
Differences in mean levels of Gender


95% family-wise confidence level
I-B
TH-B
TH-I
3
4
5
6
7
8
9
10
Differences in mean levels of HDI



- In order to get a better understanding of the issue of pairwise comparisons,we designed a dataset with many continuous variables. Use parwise comparisons with the lm model to detect statistically significant relationships between variables. What variables appear to be correlated ? Include a graph in your report and comment it.

- First step/method : use Tukey HSD for pairwise comparisons , we can also use glht method with tukey to produce pairwise comparisons
- Apply bonferroni

```
##### Using glht method

#create model 10

model10 <- lm(n.videos~Gender+HDI+CSP+age.group2,data=full_df_subset)

# running glht()
post.hoc <- glht(model10)

# displaying the result table with summary()
summary(post.hoc)

##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = n.videos ~ Gender + HDI + CSP + age.group2, data =
full_df_subset)
##
## Linear Hypotheses:
##
Estimate
## (Intercept) == 0
5.59271
## Genderune femme == 0                                                    -
0.06330
## HDII == 0
4.44640
## HDITH == 0
9.01420
## CSPArtisans, commerÃ§ants, chefs d'entreprise == 0
3.51477
## CSPArtisans, commerçants, chefs d'entreprise == 0
1.49348
## CSPCadres et professions intellectuelles == 0
2.62227
## CSPEmployés == 0
2.57608
## CSPEn recherche d'emploi == 0
4.35896
## CSPEtudiants == 0
1.92959
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) == 0
5.01234
## CSPOuvriers == 0
5.22251
## CSPProfessions intermédiaires == 0
```

```
1.13363
## CSPRetraités == 0
3.93610
## age.group2(30,50] == 0                                                -
0.06213
## age.group2(50,80] == 0
1.73028
##
Std. Error
## (Intercept) == 0
4.12523
## Genderune femme == 0
0.28973
## HDII == 0
0.62959
## HDITH == 0
0.42999
## CSPArtisans, commerÃ§ants, chefs d'entreprise == 0
4.22144
## CSPArtisans, commerçants, chefs d'entreprise == 0
4.13313
## CSPCadres et professions intellectuelles == 0
4.10916
## CSPEmployés == 0
4.12089
## CSPEn recherche d'emploi == 0
4.11982
## CSPEtudiants == 0
4.12655
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) == 0
4.26940
## CSPOuvriers == 0
5.06758
## CSPProfessions intermédiaires == 0
4.18145
## CSPRetraités == 0
4.36463
## age.group2(30,50] == 0
0.49940
## age.group2(50,80] == 0
0.59264
##                                                                       t
value
## (Intercept) == 0
1.356
## Genderune femme == 0
-0.218
## HDII == 0
7.062
## HDITH == 0
```

```
20.964
## CSPArtisans, commerÃ§ants, chefs d'entreprise == 0
0.833
## CSPArtisans, commerçants, chefs d'entreprise == 0
0.361
## CSPCadres et professions intellectuelles == 0
0.638
## CSPEmployés == 0
0.625
## CSPEn recherche d'emploi == 0
1.058
## CSPEtudiants == 0
0.468
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) == 0
1.174
## CSPOuvriers == 0
1.031
## CSPProfessions intermédiaires == 0
0.271
## CSPRetraités == 0
0.902
## age.group2(30,50] == 0
-0.124
## age.group2(50,80] == 0
2.920
##
Pr(>|t|)
## (Intercept) == 0
0.7054
## Genderune femme == 0
1.0000
## HDII == 0
<0.001
## HDITH == 0
<0.001
## CSPArtisans, commerÃ§ants, chefs d'entreprise == 0
0.9638
## CSPArtisans, commerçants, chefs d'entreprise == 0
0.9999
## CSPCadres et professions intellectuelles == 0
0.9923
## CSPEmployés == 0
0.9932
## CSPEn recherche d'emploi == 0
0.8850
## CSPEtudiants == 0
0.9991
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) == 0
0.8237
## CSPOuvriers == 0
```

```
0.8975
## CSPProfessions intermédiaires == 0
1.0000
## CSPRetraités == 0
0.9454
## age.group2(30,50] == 0
1.0000
## age.group2(50,80] == 0
0.0256
##
## (Intercept) == 0
## Genderune femme == 0
## HDII == 0
***
## HDITH == 0
***
## CSPArtisans, commerÃ§ants, chefs d'entreprise == 0
## CSPArtisans, commerçants, chefs d'entreprise == 0
## CSPCadres et professions intellectuelles == 0
## CSPEmployés == 0
## CSPEn recherche d'emploi == 0
## CSPEtudiants == 0
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) == 0
## CSPOuvriers == 0
## CSPProfessions intermédiaires == 0
## CSPRetraités == 0
## age.group2(30,50] == 0
## age.group2(50,80] == 0                                                    *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)

#apply bonferroni
summary(post.hoc, test = adjusted("bonferroni"))

##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = n.videos ~ Gender + HDI + CSP + age.group2, data =
full_df_subset)
##
## Linear Hypotheses:
##
Estimate
## (Intercept) == 0
5.59271
## Genderune femme == 0                                                      -
0.06330
## HDII == 0
4.44640
```

```
## HDITH == 0
9.01420
## CSPArtisans, commerÃ§ants, chefs d'entreprise == 0
3.51477
## CSPArtisans, commerçants, chefs d'entreprise == 0
1.49348
## CSPCadres et professions intellectuelles == 0
2.62227
## CSPEmployés == 0
2.57608
## CSPEn recherche d'emploi == 0
4.35896
## CSPEtudiants == 0
1.92959
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) == 0
5.01234
## CSPOuvriers == 0
5.22251
## CSPProfessions intermédiaires == 0
1.13363
## CSPRetraités == 0
3.93610
## age.group2(30,50] == 0                                                  -
0.06213
## age.group2(50,80] == 0
1.73028
##
Std. Error
## (Intercept) == 0
4.12523
## Genderune femme == 0
0.28973
## HDII == 0
0.62959
## HDITH == 0
0.42999
## CSPArtisans, commerÃ§ants, chefs d'entreprise == 0
4.22144
## CSPArtisans, commerçants, chefs d'entreprise == 0
4.13313
## CSPCadres et professions intellectuelles == 0
4.10916
## CSPEmployés == 0
4.12089
## CSPEn recherche d'emploi == 0
4.11982
## CSPEtudiants == 0
4.12655
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) == 0
4.26940
```

```
## CSPOuvriers == 0
5.06758
## CSPProfessions intermédiaires == 0
4.18145
## CSPRetraités == 0
4.36463
## age.group2(30,50] == 0
0.49940
## age.group2(50,80] == 0
0.59264
##                                                                    t
value
## (Intercept) == 0
1.356
## Genderune femme == 0
-0.218
## HDII == 0
7.062
## HDITH == 0
20.964
## CSPArtisans, commerÃ§ants, chefs d'entreprise == 0
0.833
## CSPArtisans, commerçants, chefs d'entreprise == 0
0.361
## CSPCadres et professions intellectuelles == 0
0.638
## CSPEmployés == 0
0.625
## CSPEn recherche d'emploi == 0
1.058
## CSPEtudiants == 0
0.468
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) == 0
1.174
## CSPOuvriers == 0
1.031
## CSPProfessions intermédiaires == 0
0.271
## CSPRetraités == 0
0.902
## age.group2(30,50] == 0
-0.124
## age.group2(50,80] == 0
2.920
##
Pr(>|t|)
## (Intercept) == 0
1.0000
## Genderune femme == 0
1.0000
```

```
## HDII == 0
2.81e-11
## HDITH == 0
< 2e-16
## CSPArtisans, commerÃ§ants, chefs d'entreprise == 0
1.0000
## CSPArtisans, commerçants, chefs d'entreprise == 0
1.0000
## CSPCadres et professions intellectuelles == 0
1.0000
## CSPEmployés == 0
1.0000
## CSPEn recherche d'emploi == 0
1.0000
## CSPEtudiants == 0
1.0000
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) == 0
1.0000
## CSPOuvriers == 0
1.0000
## CSPProfessions intermédiaires == 0
1.0000
## CSPRetraités == 0
1.0000
## age.group2(30,50] == 0
1.0000
## age.group2(50,80] == 0
0.0562
##
## (Intercept) == 0
## Genderune femme == 0
## HDII == 0
***
## HDITH == 0
***
## CSPArtisans, commerÃ§ants, chefs d'entreprise == 0
## CSPArtisans, commerçants, chefs d'entreprise == 0
## CSPCadres et professions intellectuelles == 0
## CSPEmployés == 0
## CSPEn recherche d'emploi == 0
## CSPEtudiants == 0
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) == 0
## CSPOuvriers == 0
## CSPProfessions intermédiaires == 0
## CSPRetraités == 0
## age.group2(30,50] == 0
## age.group2(50,80] == 0                                                    .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- bonferroni method)
```

6.1 Producing an Odd-Ratios table (Logistic Regression)

Use a logistic regression model (glm in R, binary family) to test whether completion, in the course, is linked to the user characteristics that you studied earlier. Make an odd-ratio table. Signal the odd-ratios that are significant in terms of p-value (with stars). Interpret the results by providing at least two alternative explanations for how socioeconomic status, or human development index, is linked to completion.

```
# if event is rare, odds ratio and relative risk are almost the same
mod_reg1 = glm(Exam.bin ~ Gender + HDI,data=full_df,family='binomial')
aov(mod_reg1)

## Call:
##    aov(formula = mod_reg1)
##
## Terms:
##                     Gender       HDI Residuals
## Sum of Squares      0.9824    3.9338 1425.2427
## Deg. of Freedom          1         2      9829
##
## Residual standard error: 0.3807937
## Estimated effects may be unbalanced
## 18637 observations effacées parce que manquantes

A=exp(coef(mod_reg1))    # Odd ratios
exp(confint(mod_reg1))  # calculate confidence intervals

##                       2.5 %    97.5 %
## (Intercept)     0.1230143 0.1698055
## Genderune femme 0.9947384 1.2402111
## HDII            0.9299391 1.5506161
## HDITH           1.2927521 1.8219648

summary(mod_reg1)

##
## Call:
## glm(formula = Exam.bin ~ Gender + HDI, family = "binomial", data =
full_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6639  -0.6331  -0.6331  -0.5204   2.0330
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -1.93113    0.08218 -23.498  < 2e-16 ***
## Genderune femme  0.10537    0.05626   1.873   0.0611 .
## HDII             0.18449    0.13032   1.416   0.1569
## HDITH            0.42562    0.08749   4.865 1.15e-06 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 9169.8  on 9832  degrees of freedom
## Residual deviance: 9134.2  on 9829  degrees of freedom
##   (18637 observations effacées parce que manquantes)
## AIC: 9142.2
##
## Number of Fisher Scoring iterations: 4
```

```
anova(mod_reg1)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Exam.bin
##
## Terms added sequentially (first to last)
##
##
##        Df Deviance Resid. Df Resid. Dev
## NULL                    9832     9169.8
## Gender  1    6.683      9831     9163.1
## HDI     2   28.860      9829     9134.2
```

```
#OR table with confidenc intervals
exp(cbind(OR = coef(mod_reg1), confint.default(mod_reg1)))
```

```
##                          OR     2.5 %    97.5 %
## (Intercept)     0.1449847 0.1234150 0.1703242
## Genderune femme 1.1111225 0.9951205 1.2406469
## HDII            1.2025995 0.9315220 1.5525618
## HDITH           1.5305356 1.2893605 1.8168225
```

```
#pseudo-R2 , McFadden
pR2(mod_reg1)
```

```
## fitting null model for pseudo-r2
```

```
##           llh       llhNull            G2      McFadden          r2ML
## -4.567117e+03 -4.584888e+03  3.554273e+01  3.876074e-03  3.608113e-03
##          r2CU
##  5.949548e-03
```

```
#optional
# if we want to change the reference
mod_reg2 = glm(Exam.bin ~ HDI +relevel(as.factor(Gender), ref = "une
femme"),data=full_df,family='binomial')
summary(mod_reg2)
```

```
##
## Call:
## glm(formula = Exam.bin ~ HDI + relevel(as.factor(Gender), ref = "une
femme"),
##     family = "binomial", data = full_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6639  -0.6331  -0.6331  -0.5204   2.0330
##
## Coefficients:
##                                                          Estimate Std. Error
## (Intercept)                                              -1.82576    0.09434
## HDII                                                      0.18449    0.13032
## HDITH                                                     0.42562    0.08749
## relevel(as.factor(Gender), ref = "une femme")un homme -0.10537    0.05626
##                                                          z value Pr(>|z|)
## (Intercept)                                              -19.354  < 2e-16 ***
## HDII                                                       1.416   0.1569
## HDITH                                                      4.865 1.15e-06 ***
## relevel(as.factor(Gender), ref = "une femme")un homme  -1.873   0.0611 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 9169.8  on 9832  degrees of freedom
## Residual deviance: 9134.2  on 9829  degrees of freedom
##   (18637 observations effacées parce que manquantes)
## AIC: 9142.2
##
## Number of Fisher Scoring iterations: 4
```

```
#Model 3 , completion ~ Gender + CSP + HDI
mod_reg3 = glm(Exam.bin ~ Gender + HDI + CSP,data=full_df,family='binomial')

# ORS + confidence intervals
C = exp(cbind(OR = coef(mod_reg3), confint.default(mod_reg3)))


C

##
OR
## (Intercept)
0.2005120
## Genderune femme
1.1321716
## HDII
1.1503313
```

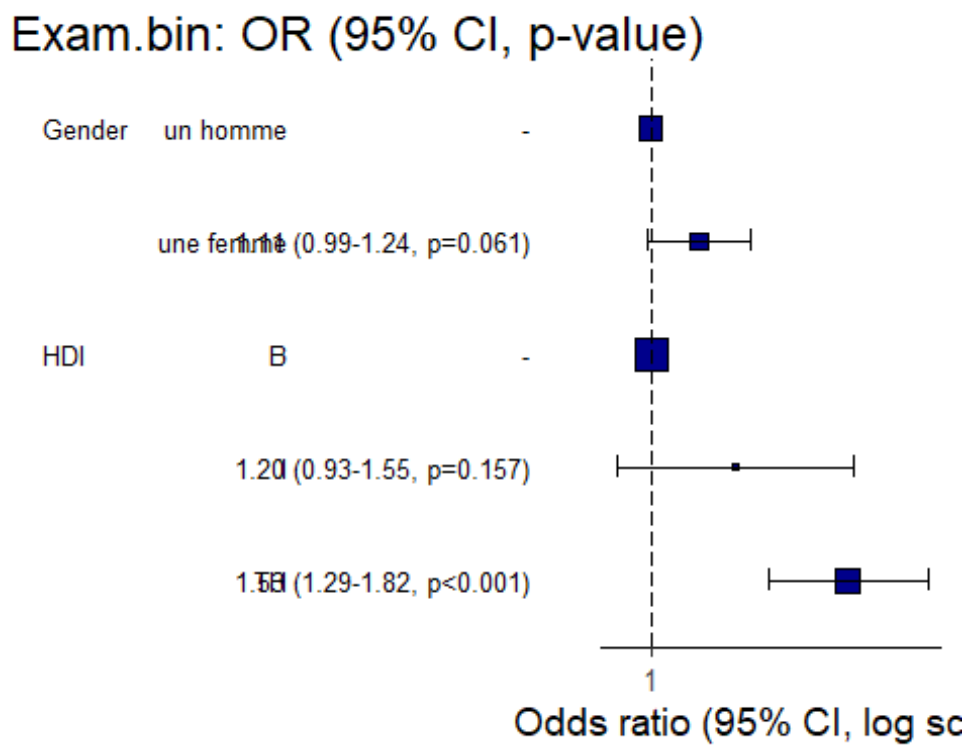
```
## HDITH
1.4558391
## CSPArtisans, commerÃ§ants, chefs d'entreprise
2.9048223
## CSPArtisans, commerçants, chefs d'entreprise
0.4161340
## CSPCadres et professions intellectuelles
0.7963505
## CSPEmployés
0.6656800
## CSPEn recherche d'emploi
0.8487977
## CSPEtudiants
0.6680759
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
0.9474246
## CSPOuvriers
0.9110835
## CSPProfessions intermédiaires
0.5717609
## CSPRetraités
0.9587844
##
2.5 %
## (Intercept)
0.04307495
## Genderune femme
1.01225235
## HDII
0.88704393
## HDITH
1.21845961
## CSPArtisans, commerÃ§ants, chefs d'entreprise
0.60388752
## CSPArtisans, commerçants, chefs d'entreprise
0.08747508
## CSPCadres et professions intellectuelles
0.17002553
## CSPEmployés
0.14129091
## CSPEn recherche d'emploi
0.18051429
## CSPEtudiants
0.14231789
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
0.19122805
## CSPOuvriers
0.13668236
## CSPProfessions intermédiaires
0.11801162
```

```
## CSPRetraités
0.18809322
##
97.5 %
## (Intercept)
0.9333749
## Genderune femme
1.2662974
## HDII
1.4917661
## HDITH
1.7394646
## CSPArtisans, commerÃ§ants, chefs d'entreprise
13.9727884
## CSPArtisans, commerçants, chefs d'entreprise
1.9796211
## CSPCadres et professions intellectuelles
3.7298763
## CSPEmployés
3.1362939
## CSPEn recherche d'emploi
3.9911381
## CSPEtudiants
3.1361163
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
4.6939420
## CSPOuvriers
6.0730085
## CSPProfessions intermédiaires
2.7701554
## CSPRetraités
4.8872977

summary(mod_reg3)

##
## Call:
## glm(formula = Exam.bin ~ Gender + HDI + CSP, family = "binomial",
##     data = full_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1601  -0.6465  -0.5969  -0.5006   2.2644
##
## Coefficients:
##
Estimate
## (Intercept)                                                    -
1.60688
## Genderune femme
```

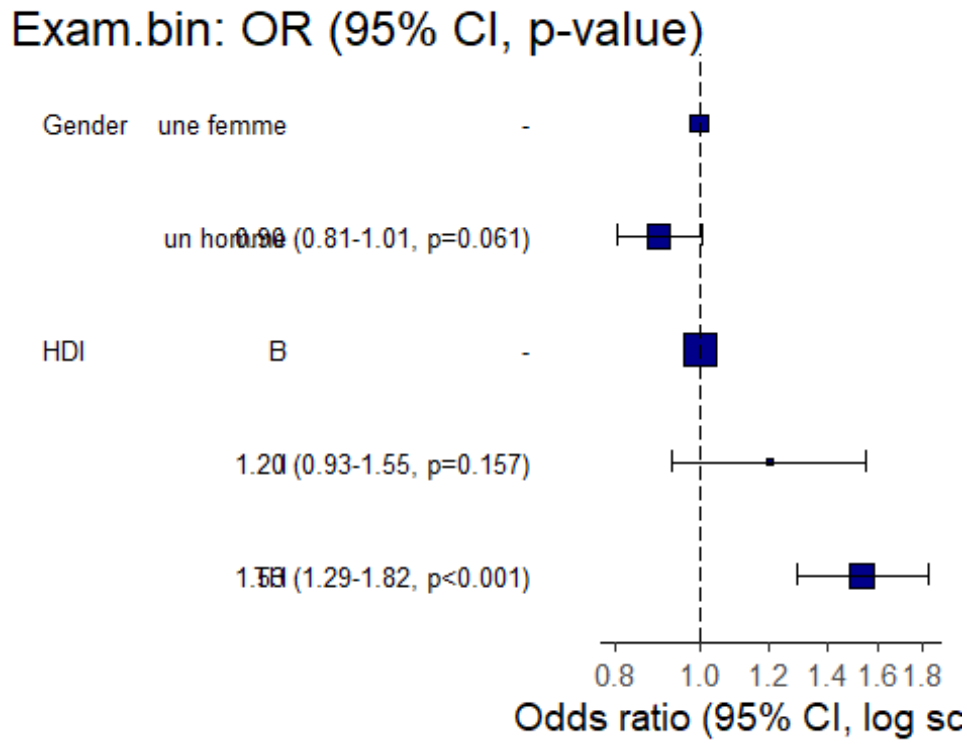
```
0.12414
## HDII
0.14005
## HDITH
0.37558
## CSPArtisans, commerÃ§ants, chefs d'entreprise
1.06637
## CSPArtisans, commerçants, chefs d'entreprise                        -
0.87675
## CSPCadres et professions intellectuelles                            -
0.22772
## CSPEmployés                                                         -
0.40695
## CSPEn recherche d'emploi                                            -
0.16393
## CSPEtudiants                                                        -
0.40335
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi) -
0.05401
## CSPOuvriers                                                         -
0.09312
## CSPProfessions intermédiaires                                       -
0.55903
## CSPRetraités                                                        -
0.04209
##                                                                    Std.
Error
## (Intercept)
0.78467
## Genderune femme
0.05712
## HDII
0.13261
## HDITH
0.09082
## CSPArtisans, commerÃ§ants, chefs d'entreprise
0.80141
## CSPArtisans, commerçants, chefs d'entreprise
0.79576
## CSPCadres et professions intellectuelles
0.78782
## CSPEmployés
0.79082
## CSPEn recherche d'emploi
0.78982
## CSPEtudiants
0.78896
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
0.81648
## CSPOuvriers
```

```
0.96786
## CSPProfessions intermédiaires
0.80509
## CSPRetraités
0.83100
##                                                                   z
value
## (Intercept)                                                      -
2.048
## Genderune femme
2.173
## HDII
1.056
## HDITH
4.136
## CSPArtisans, commerÃ§ants, chefs d'entreprise
1.331
## CSPArtisans, commerçants, chefs d'entreprise                     -
1.102
## CSPCadres et professions intellectuelles                         -
0.289
## CSPEmployés                                                      -
0.515
## CSPEn recherche d'emploi                                         -
0.208
## CSPEtudiants                                                     -
0.511
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)  -
0.066
## CSPOuvriers                                                      -
0.096
## CSPProfessions intermédiaires                                    -
0.694
## CSPRetraités                                                     -
0.051
##
Pr(>|z|)
## (Intercept)
0.0406
## Genderune femme
0.0298
## HDII
0.2909
## HDITH                                                            3.54e-
05
## CSPArtisans, commerÃ§ants, chefs d'entreprise
0.1833
## CSPArtisans, commerçants, chefs d'entreprise
0.2706
## CSPCadres et professions intellectuelles
```

```
0.7725
## CSPEmployés
0.6068
## CSPEn recherche d'emploi
0.8356
## CSPEtudiants
0.6092
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
0.9473
## CSPOuvriers
0.9234
## CSPProfessions intermédiaires
0.4874
## CSPRetraités
0.9596
##
## (Intercept)                                                          *
## Genderune femme                                                      *
## HDII
## HDITH                                                                ***
## CSPArtisans, commerÃ§ants, chefs d'entreprise
## CSPArtisans, commerçants, chefs d'entreprise
## CSPCadres et professions intellectuelles
## CSPEmployés
## CSPEn recherche d'emploi
## CSPEtudiants
## CSPInactif (autre que étudiant, retraité, ou en recherche d'emploi)
## CSPOuvriers
## CSPProfessions intermédiaires
## CSPRetraités
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 9109.8  on 9757  degrees of freedom
## Residual deviance: 8961.7  on 9744  degrees of freedom
##   (18712 observations effacées parce que manquantes)
## AIC: 8989.7
##
## Number of Fisher Scoring iterations: 4

#Odds-ratio plot also known as forest plot
full_df %>%   or_plot('Exam.bin', c('Gender','HDI'),
          breaks = c(0.5, 1, 5, 10, 20, 30),
          table_text_size = 3.5)
```

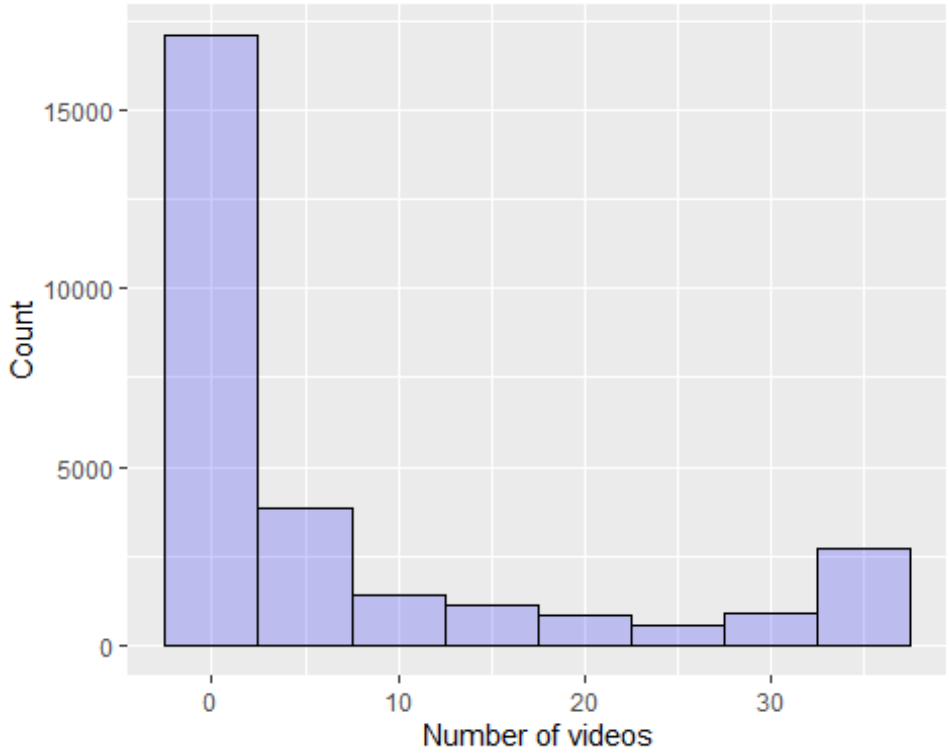



```r
#Forest OR plot with female as reference instead of male

full_df %>% mutate(Gender=factor(Gender,levels=c('une femme','un homme')))
%>%
   or_plot('Exam.bin', c('Gender','HDI'), table_text_size = 3.5)
```



6.2 Poisson model for count data

```
qplot(full_df$n.videos,
      geom="histogram",
      binwidth = 5,
      xlab = "Number of videos",
      ylab="Count",
      fill=I("blue"),
      col=I("black"),
      alpha=I(.2),
    ) +   geom_density()
```

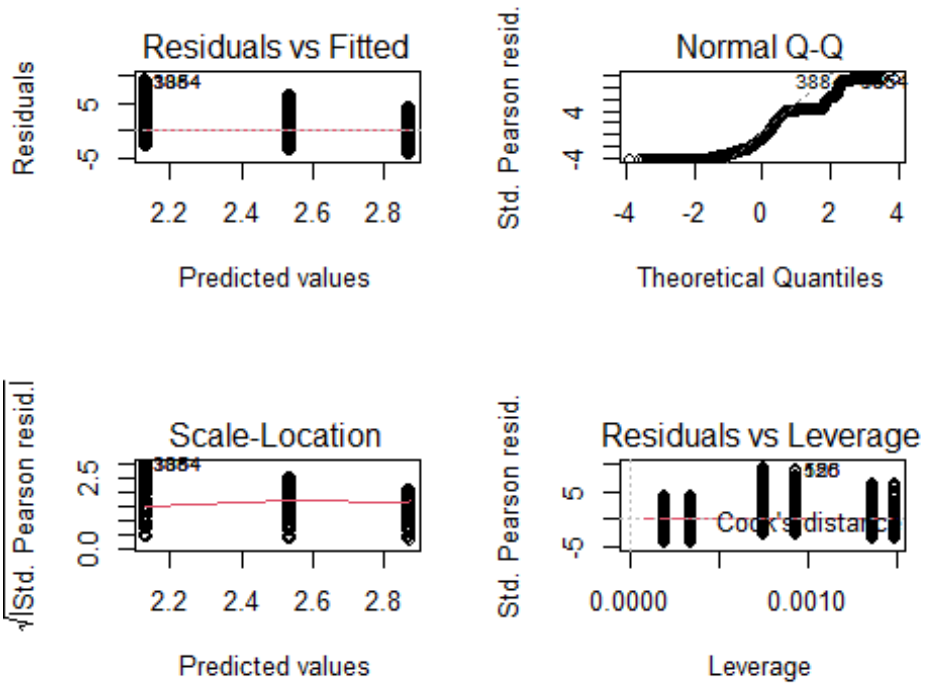



```
#poisson model <=> family="poisson"
mod_reg4 = glm(n.videos ~ Gender+HDI,data=full_df,family='poisson')

summary(mod_reg4)

##
## Call:
## glm(formula = n.videos ~ Gender + HDI, family = "poisson", data = full_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.9404  -3.5607  -0.8802   3.2575   6.8264
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     2.130090   0.009452 225.368   <2e-16 ***
## Genderune femme 0.004977   0.005372   0.926    0.354
## HDII            0.402182   0.013949  28.833   <2e-16 ***
## HDITH           0.735331   0.009858  74.596   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 127057  on 9836  degrees of freedom
## Residual deviance: 119468  on 9833  degrees of freedom
##   (18633 observations effacées parce que manquantes)
```

```
## AIC: 157580
##
## Number of Fisher Scoring iterations: 5

#latex table
#print(xtable(summary(mod4)))
```

Residual analysis of poisson model * Check homoscedasticity of the residuals i.e residual analysis ==> homoscedasticity assumes the residuals are approximately equal for all predicted dependent variable scores , assumes equal variance

```
par(mfrow=c(2,2)) # init 4 charts in 1 panel
plot(mod_reg4)
```




```
#ORs for poisson model
exp(cbind(OR = coef(mod_reg4), confint.default(mod_reg4)))

##                         OR     2.5 %   97.5 %
## (Intercept)     8.415622 8.2611592 8.572973
## Genderune femme 1.004990 0.9944628 1.015628
## HDII            1.495083 1.4547623 1.536521
## HDITH           2.086172 2.0462527 2.126869
```

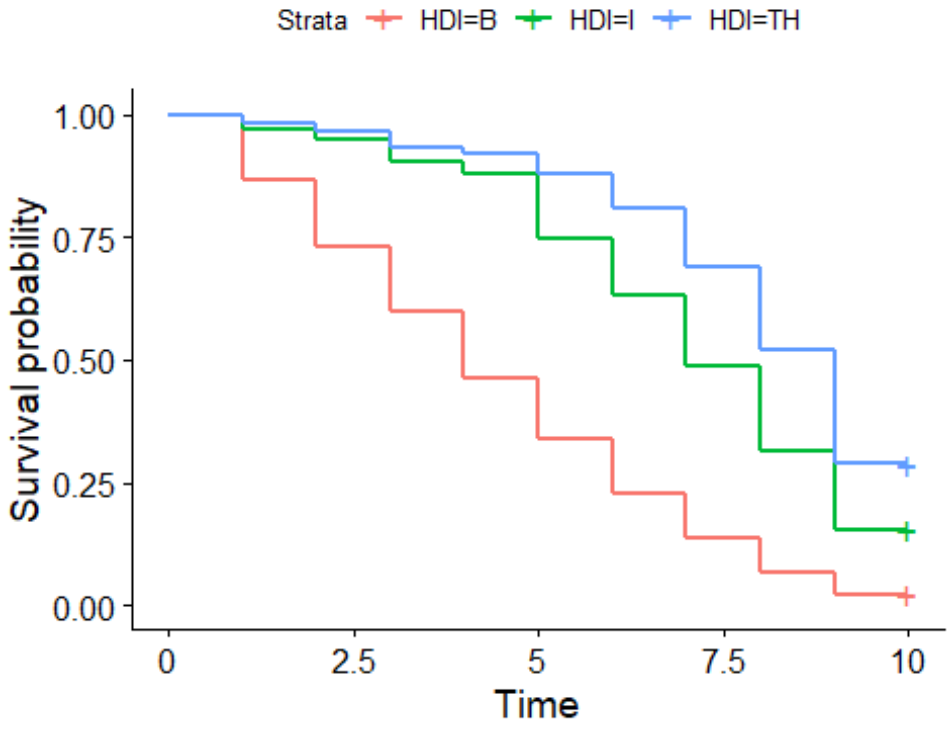
7 Survival Analysis

- You must reason in terms of proportion of the available videos that the learner viewed. Prepare the data so that they are fit for a survival analysis.

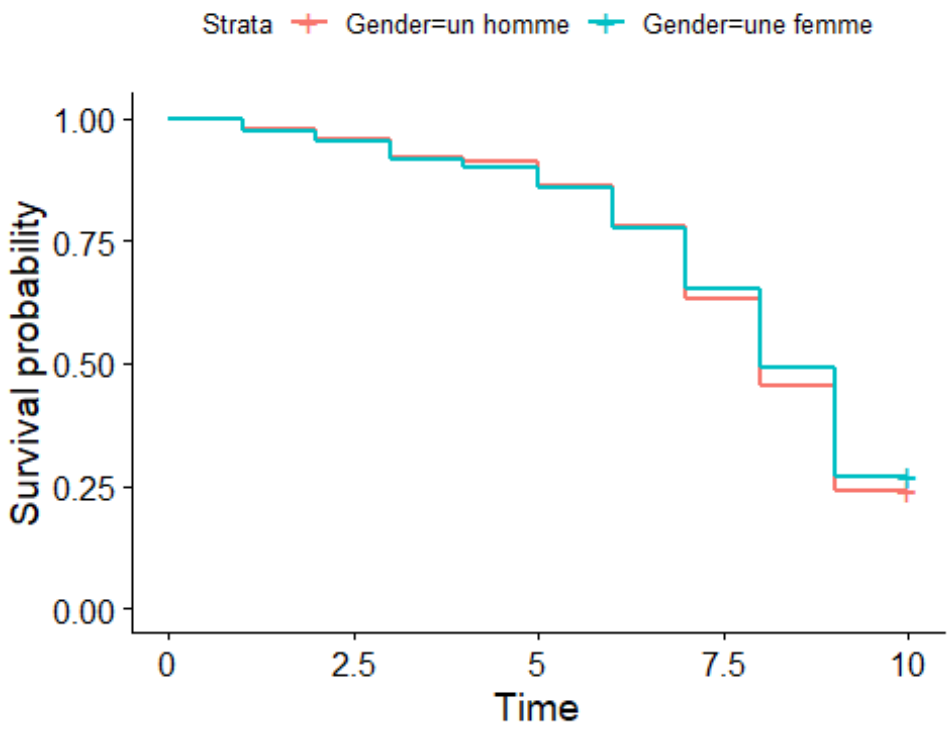
```r
#check deciles for number of videos
n.videos_dec = quantile(full_df$n.videos, probs = seq(.1, .9, by = .1))
#add deciles (new column ) for the number of videos
#using mutate method
full_df<-full_df %>%
  mutate(n.videos.decile = ntile(n.videos, 10))

# add status based on deciles
full_df$status.vid=rep(NA, nrow(full_df))
for (i in 1:nrow(full_df)) {
  if (full_df$n.videos.decile[i]<10) {full_df$status.vid[i]=1}
  if (full_df$n.videos.decile[i]==10) {full_df$status.vid[i]=0}
}
```

- Compare video consumption behavior between auditing and disengaging learners, but this time with a survival analysis (and not the linear model like you did earlier).
- plot the survival curve. Where do you see the most significative drop in terms of video consumption ?

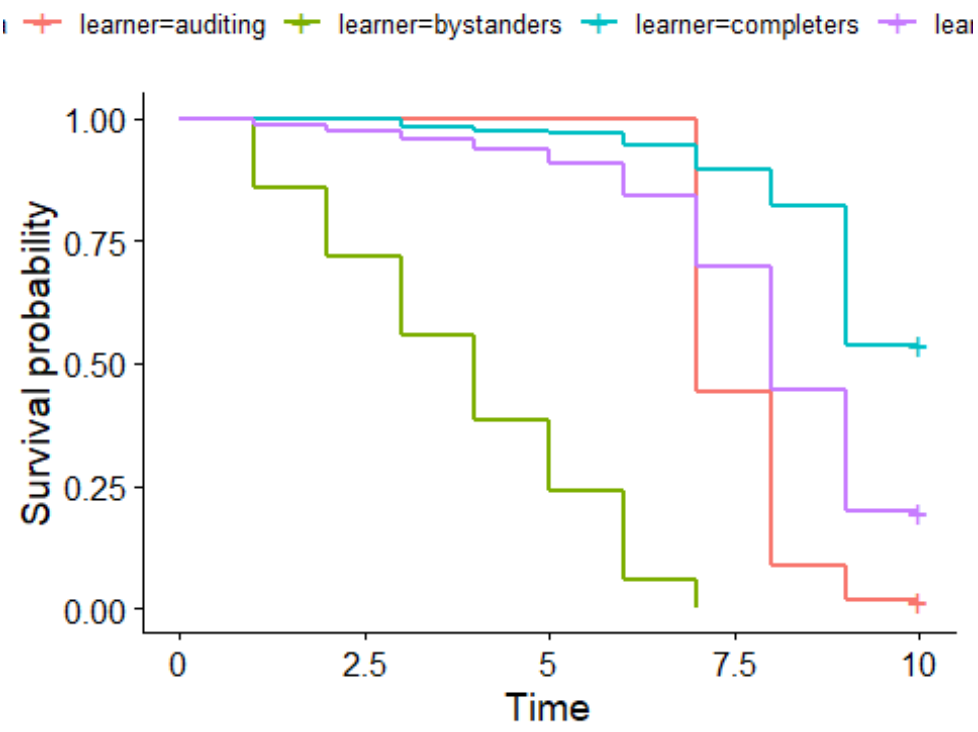
```r
#number of videos survival analysis based on HDI
surv_mod1 <- survfit(Surv(n.videos.decile, status.vid) ~ HDI , data=full_df)
ggsurvplot(surv_mod1, data = full_df)
```




```r
#number of videos survival analysis based on Gender
surv_mod2 <- survfit(Surv(n.videos.decile, status.vid) ~ Gender ,
data=full_df)
ggsurvplot(surv_mod2, data = full_df)
```




```
#number of videos survival analysis based on type of learners(completers,
disengaging etc)
surv_mod3 <- survfit(Surv(n.videos.decile, status.vid) ~ learner ,
data=full_df)
ggsurvplot(surv_mod3, data = full_df)
```

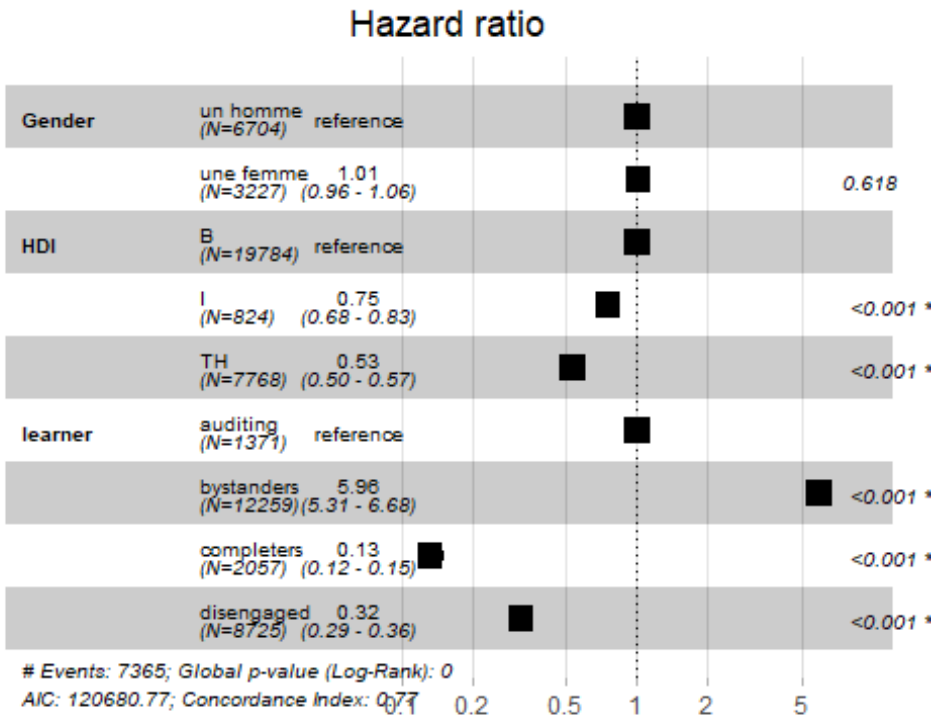


Compute the hazard ratios

```
#Calculate hazard ratios using coxph
mod_cox <- coxph(formula = Surv(n.videos.decile, status.vid) ~
Gender+HDI+learner, data = full_df)

mod_cox

## Call:
## coxph(formula = Surv(n.videos.decile, status.vid) ~ Gender +
##     HDI + learner, data = full_df)
##
##                         coef exp(coef) se(coef)       z        p
## Genderune femme    0.01266   1.01274  0.02538   0.499    0.618
## HDII              -0.28829   0.74954  0.04934  -5.843 5.13e-09
## HDITH             -0.63348   0.53074  0.03247 -19.509  < 2e-16
## learnerbystanders  1.78456   5.95693  0.05845  30.533  < 2e-16
## learnercompleters -2.02005   0.13265  0.06284 -32.147  < 2e-16
## learnerdisengaged -1.13751   0.32062  0.05288 -21.511  < 2e-16
##
## Likelihood ratio test=6846  on 6 df, p=< 2.2e-16
## n= 9833, number of events= 7365
##    (18637 observations effacées parce que manquantes)
```

References are : Male(for gender), Low(For HDI), auditing (for types of learners)

```
#hazard ratios in forest plot
ggforest(mod_cox,data=full_df)
```



Brief interpretation : people from rich countries tend to disengage much slower from the course than people from poor country(H=0.45,ref=poor,p-value<0.001) - do the same for gender and type of learners