Project 1 - Intermediate Statistics

Anh Thu

October 4, 2021

```
# Import library
library(readr)
library(dplyr)
```

```
##
## Attaching package: ’dplyr’

## The following objects are masked from ’package:stats’:
##
##     filter, lag

## The following objects are masked from ’package:base’:
##
##     intersect, setdiff, setequal, union
```

```
library(plyr)
```

```
## ------------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##
## Attaching package: ’plyr’

## The following objects are masked from ’package:dplyr’:
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
library(readxl)
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
## v tibble  3.1.5     v stringr 1.4.0
## v tidyr   1.1.4     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x plyr::arrange()   masks dplyr::arrange()
## x purrr::compact()  masks plyr::compact()
## x plyr::count()     masks dplyr::count()
## x plyr::failwith()  masks dplyr::failwith()
## x dplyr::filter()   masks stats::filter()
## x plyr::id()        masks dplyr::id()
## x dplyr::lag()      masks stats::lag()
## x plyr::mutate()    masks dplyr::mutate()
## x plyr::rename()    masks dplyr::rename()
## x plyr::summarise() masks dplyr::summarise()
## x plyr::summarize() masks dplyr::summarize()
```

```
library(finalfit)
library(survival)
library(survminer)
```

```
## Loading required package: ggpubr
```

```
##
## Attaching package: 'ggpubr'
```

```
## The following object is masked from 'package:plyr':
##
##     mutate
```

```
##
## Attaching package: 'survminer'
```

```
## The following object is masked from 'package:survival':
##
##     myeloma
```

```
library(ggfortify)
```

```
# Import dataset
effec1_quest_compil <- read_csv("Datasets/effec1.quest.compil.csv", locale = locale("fr"), show_col_type
effec2_quest_compil <- read_csv("Datasets/effec2.quest.compil.csv", locale = locale("fr"), show_col_type
effec3_quest_compil <- read_csv("Datasets/effec3.quest.compil.csv", locale = locale("fr"), show_col_type
usages_effec1 <- read_csv("Datasets/usages.effec1.csv", locale = locale("fr"), show_col_types = FALSE)
usages_effec2 <- read_csv("Datasets/usages.effec2.csv", locale = locale("fr"), show_col_types = FALSE)
usages_effec3 <- read_csv("Datasets/usages.effec3.csv", locale = locale("fr"), show_col_types = FALSE)
```

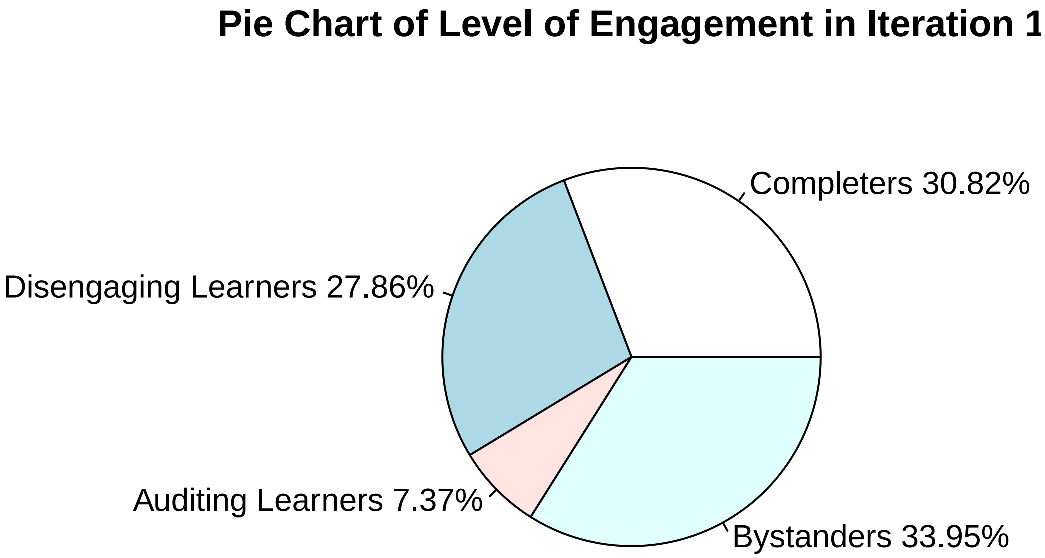
```
# Combine
UE1 <- join_all(list(effec1_quest_compil,usages_effec1),type = 'full', by = 'Student_ID')
UE2 <- join_all(list(effec2_quest_compil,usages_effec2), type = "full", by ="Student_ID")
UE3 <- join_all(list(effec3_quest_compil,usages_effec3), type = "full", by = "Student_ID")
```

```
# Create engagement colum for usages_effect1

usages_effec1 <- distinct(usages_effec1)

usages_effec1 <- usages_effec1 %>% mutate(Engagement_Level = case_when (Assignment.bin == 1 ~ "complete
                                                                           last.quizz > 0 |!is.na(Assignm
                                                                           last.video / 35 > 0.1 ~ "audi
                                                                           TRUE ~ "bystander"))
```

```
Engagement_Level_table1 <- count(usages_effec1, "Engagement_Level")
Engagement_Level_table1
```

```
##   Engagement_Level freq
## 1            audit  587
## 2        bystander 2704
## 3         complete 2455
## 4        disengage 2219
```

```
iteration1<- c(2455,2219,587,2704)
lbls1 <- c("Completers","Disengaging Learners", "Auditing Learners","Bystanders")
pct1 <- round(iteration1/sum(iteration1)*100, digits=2)
lbls1 <- paste(lbls1, pct1)
lbls1 <- paste(lbls1,"%",sep="")
pie(iteration1, labels = lbls1, main="Pie Chart of Level of Engagement in Iteration 1")
```



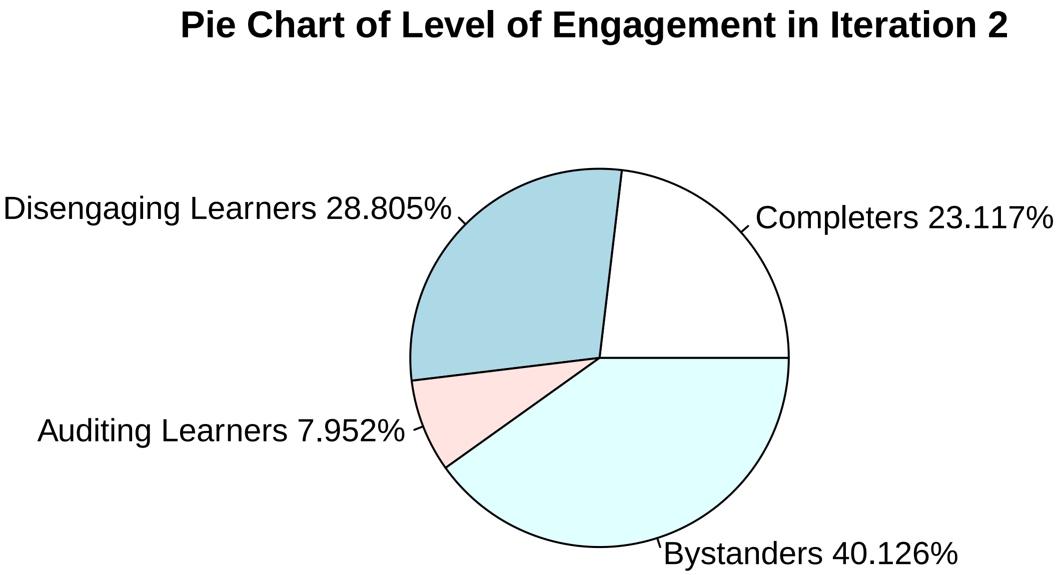
```
UE1 <- UE1 %>% mutate(Engagement_Level = case_when (Assignment.bin == 1 ~ "complete",
                                                                           last.quizz > 0 |!is.na(Assignm
                                                                           last.video / 35 > 0.1 ~ "audi
                                                                           TRUE ~ "bystander"))
```

```
# Create engagement colum for usages_effect2
usages_effec2 <- distinct(usages_effec2)
usages_effec2 <-usages_effec2 %>% mutate(Engagement_Level =case_when (Exam.bin == 1 ~ "complete",
                                                                          last.quizz > 0 |!is.na(Assign
                                                                          last.video / 35 > 0.1 ~ "audi
                                                                          TRUE ~ "bystander"))
```

```
c2 <- count(usages_effec2, "Engagement_Level")
c2
```

```
##   Engagement_Level freq
## 1            audit  302
## 2        bystander 1524
## 3         complete  878
## 4        disengage 1094
```

```
iteration2<- c(878,1094,302,1524)
lbls2 <- c("Completers","Disengaging Learners", "Auditing Learners","Bystanders")
pct2 <- round(iteration2/sum(iteration2)*100, digits=3)
lbls2 <- paste(lbls2, pct2)
lbls2 <- paste(lbls2,"%",sep="")
pie(iteration2, labels = lbls2, main="Pie Chart of Level of Engagement in Iteration 2 ")
```



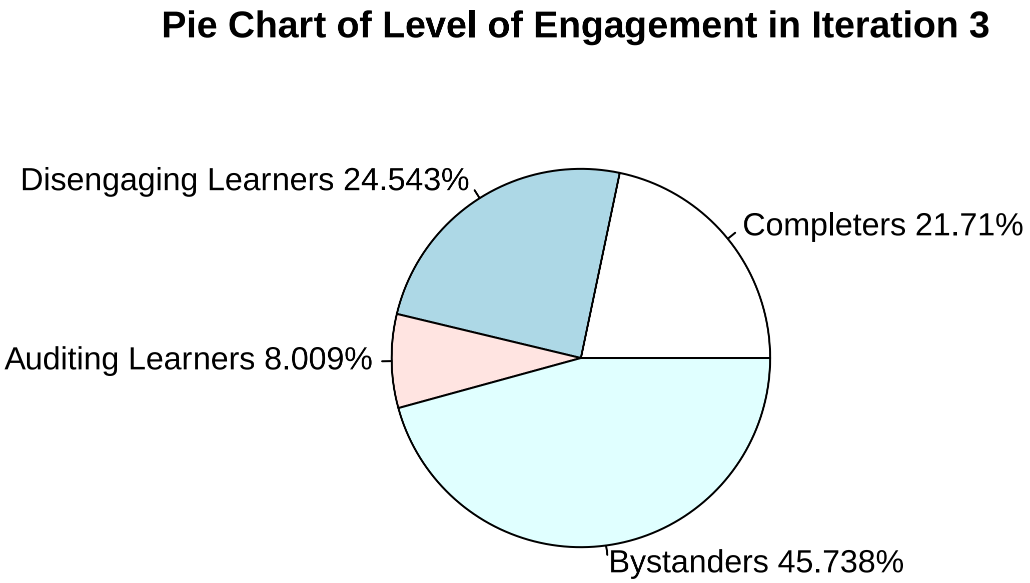
```
UE2 <- UE2 %>% mutate(Engagement_Level =case_when (Exam.bin == 1 ~ "complete",
                                                                          last.quizz > 0 |!is.na(Assign
                                                                          last.video / 35 > 0.1 ~ "audi
                                                                          TRUE ~ "bystander"))
```

```
# Create engagement colum for usages_effect3
usages_effec3 <-usages_effec3 %>% mutate(Engagement_Level =case_when (Exam.bin == 1 ~ "complete",
                                                                          last.quizz > 0 |!is.na(Assign
                                                                          last.video / 35 > 0.1 ~ "audi
                                                                          TRUE ~ "bystander"))
```

```
count(usages_effec3, "Engagement_Level")
```

```
##   Engagement_Level freq
## 1            audit  311
## 2        bystander 1776
## 3         complete  843
## 4        disengage  953
```

```
iteration3<- c(843,953,311,1776)
lbls3 <- c("Completers","Disengaging Learners", "Auditing Learners","Bystanders")
pct3 <- round(iteration3/sum(iteration3)*100, digits=3)
lbls3 <- paste(lbls3, pct3)
lbls3 <- paste(lbls3,"%",sep="")
pie(iteration3, labels = lbls3, main="Pie Chart of Level of Engagement in Iteration 3 ")
```



```
UE3 <- UE3 %>% mutate(Engagement_Level =case_when (Exam.bin == 1 ~ "complete",
                                                   last.quizz > 0 |!is.na(Assignment.score) ~ "disengag
                                                   last.video / 35 > 0.1 ~ "audit",
                                                   TRUE ~ "bystander"))
```

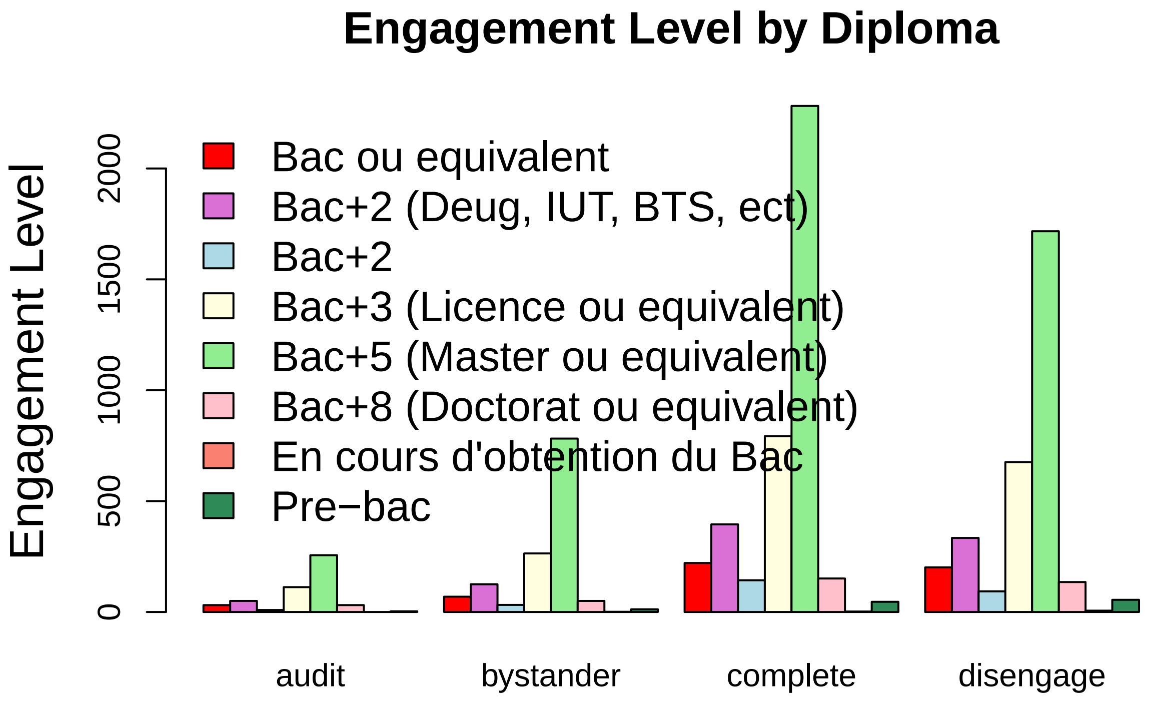
```
library(plyr)
UE_all <- rbind.fill(UE1,UE2, UE3)
Table1 <- table(UE_all$Diploma,UE_all$Engagement_Level)
Table1
```

```
##                                          
##                                           audit bystander complete disengage
##   Bac ou \xe9quivalent                       31        69      221       201
##   Bac+2 (Deug, IUT, BTS ou \xe9quivalent)    50       125      395       334
##   Bac+2 (pr\xe9pas)                           9        32      143        93
##   Bac+3 (Licence ou \xe9quivalent)          112       264      793       676
##   Bac+5 (Master ou \xe9quivalent)           256       782     2282      1717
##   Bac+8 (Doctorat ou \xe9quivalent)          31        50      151       135
##   En cours d'obtention du Bac                       0         1        2         6
##   Pr\xe9-bac                                        3        12       46        55
```

```
Table2 <- table(UE_all$Exp.MOOC, UE_all$Engagement_Level)
print(Table2)
```

```
##
##                                                        audit bystander complete
##   Non, c'est ma premi\xe8re participation à un MOOC      101       378     1028
##   Non, c'est ma premi\xe8re participation \xe0 un MOOC   193       559     1788
##   Oui, dont certains int\xe9gralement                    123       252      948
##   Oui, mais tous suivis partiellement                     72       140      248
##   Oui, que j'ai suivi partiellement                        0         1        0
##
##                                                        disengage
##   Non, c'est ma premi\xe8re participation à un MOOC          782
##   Non, c'est ma premi\xe8re participation \xe0 un MOOC      1353
##   Oui, dont certains int\xe9gralement                        772
##   Oui, mais tous suivis partiellement                        288
##   Oui, que j'ai suivi partiellement                            0
```

```
colours <- c("red", "orchid", "light blue", "light yellow", "light green","pink","salmon","seagreen")
counts <- table(UE_all$Diploma, UE_all$Engagement_Level)
barplot(as.matrix(counts), main="Engagement Level by Diploma", ylab = "Engagement Level", cex.lab = 1.5
legend("topleft", c("Bac ou equivalent", "Bac+2 (Deug, IUT, BTS, ect)","Bac+2","Bac+3 (Licence ou equiva
```



```
head(UE_all)
```

```
##   Student_ID    Gender birth.year    Country                              Diploma
## 1        221      <NA>         NA       <NA>                                 <NA>
## 2      19178 une femme       1986     France  Bac+5 (Master ou \xe9quivalent)
## 3       1086 une femme       1967     France  Bac+5 (Master ou \xe9quivalent)
## 4       1948 une femme       1983  Allemagne             Bac ou \xe9quivalent
## 5      16209 une femme         NA Madagascar Bac+3 (Licence ou \xe9quivalent)
## 6       6685  un homme       1951       <NA>  Bac+5 (Master ou \xe9quivalent)
##                                                                      Formation
## 1                                                                         <NA>
## 2                                                                        Droit
## 3 Sciences sociales (\xe9conomie\\, sciences politiques\\, sociologie\\, etc)
## 4                                                                        Droit
## 5    Sciences naturelles (Agronomie\\, biologie\\, physique\\, chimie\\, etc)
## 6                                               Ing\xe9nierie et technologies
##                                            CSP
## 1                                         <NA>
## 2        Cadres et professions intellectuelles
## 3 Artisans, commer\xe7ants, chefs d’entreprise
## 4                                  Employ\xe9s
## 5                Professions interm\xe9diaires
## 6                                 Retrait\xe9s
##                                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 exp\xe9rience en cr\xe9ation d’entreprise
## 3      Je suis en train de cr\xe9er mon entreprise (phase de d\xe9marrage)
## 4                 Je n’ai aucune exp\xe9rience en cr\xe9ation d’entreprise
## 5 J\x92ai un projet de cr\xe9ation d\x92entreprise (phase de r\xe9flexion)
## 6                 Je n’ai aucune exp\xe9rience en cr\xe9ation 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 premi\xe8re participation \xe0 un MOOC
## 3 Entre une et deux heures Non, c’est ma premi\xe8re participation \xe0 un MOOC
## 4 Entre une et deux heures                  Oui, mais tous suivis partiellement
## 5 Entre une et deux heures Non, c’est ma premi\xe8re participation \xe0 un MOOC
## 6       Plus de six heures                  Oui, dont certains int\xe9gralement
##   Completion.proba                                       Instit.brand
## 1               NA                                               <NA>
## 2                5                                               <NA>
## 3                4                                               <NA>
## 4                4                                               <NA>
## 5                5                                               <NA>
## 6                5 2. Oui, c\x92est un param\xe8tre tr\xe8s important
##                                                                         motiv.princ
## 1                                                                              <NA>
## 2                                                                              <NA>
## 3                                                                              <NA>
## 4                                                                              <NA>
## 5                                                                              <NA>
## 6 La satisfaction personnelle d\x92\xeatre all\xe9 jusqu\x92au 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\x92ai pas \xe9chang\xe9 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 Regard\xe9 des vid\xe9os ensemble,S\x92encourager mutuellement \xe0 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 \xe0 8 heures
##                                                                                   Part.labo
## 1                                                                                      <NA>
## 2                                                                                      <NA>
## 3                                                                                      <NA>
## 4                                                                                      <NA>
## 5                                                                                      <NA>
## 6 Non\\, j\x92ai compris ce qu\x92\xe9tait le Laboratoire mais je n\x92y ai pas particip\xe9
##             Plat.satisf Peer.eval.relev encad.diffic Country_HDI
## 1                  <NA>            <NA>           NA        <NA>
## 2                  <NA>            <NA>           NA          TH
## 3                  <NA>            <NA>           NA          TH
## 4                  <NA>            <NA>           NA          TH
## 5                 <NA>             <NA>           NA           B
## 6 Tr\xe8s satisfaisante                3           NA        <NA>
##   Country_HDI.fin                                      CSP.fin
## 1            <NA>                                         <NA>
## 2              TH        Cadres et professions intellectuelles
## 3              TH Artisans, commer\xe7ants, chefs d’entreprise
## 4              TH                                   Employ\xe9s
## 5               B                                        Autre
## 6            <NA>                                        Autre
##        Temps.dispo.fin Exam.score Exam.bin Assignment.score Assignment.bin
## 1                 <NA>         NA        0               NA              0
## 2 Moins de deux heures         NA        0               NA              0
## 3 Moins de deux heures         NA        0               NA              0
## 4 Moins de deux heures         NA        0               NA              0
## 5 Moins de deux heures         NA        0               NA              0
## 6   Plus de six heures         NA        0               70              1
##   Quizz.1.score Quizz.1.bin Quizz.2.score Quizz.2.bin Quizz.3.score Quizz.3.bin
## 1            NA           0            NA           0            NA           0
## 2            NA           0            NA           0            NA           0
## 3            11           1            20           1         17.33           1
## 4            NA           0            NA           0            NA           0
## 5            20           1            20           1         20.00           1
## 6            20           1            20           1         18.00           1
##   Quizz.4.bin Quizz.4.score Quizz.5.bin Quizz.5.score Intro.MOOC Prez.sem.1
## 1           0            NA           0            NA         NA          1
## 2           0            NA           0            NA         NA          1
## 3           1         20.00           0            NA         NA          1
## 4           0            NA           0            NA         NA          1
## 5           1         20.00           1            20         NA          0
## 6           1         17.33           1            19         NA          0
##   S1.L1 S1.L2 S1.L3 S1.L4 S1.L5 S1.L6 Prez.sem.2 S2.L1 S2.L2 S2.L3 S2.L4 S2.L5
## 1     0     0     0     0     0     0          0     0     0     0     0     0
## 2     1     0     0     0     0     0          0     0     0     0     0     0
## 3     1     1     1     1     1     1          1     1     1     1     1     1
## 4     1     0     0     0     0     0          0     0     0     0     0     0
## 5     0     0     0     0     0     0          0     0     0     0     0     0
## 6     1     0     1     1     0     0          1     1     0     0     0     0
##   S2.L6 Prez.sem.3 S3.L1.1 S3.L1.2 S3.L2 S3.L3 S3.L4 S3.L5 Prez.sem.4 S4.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     1          1       1       0     0     0     0     0          0       0
##   S4.L1.2 S4.L2 S4.L3 S4.L4 S4.L5 Prez.sem.5 S5.L1.1 S5.L1.2 S5.L2 S5.L3 S5.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
##   S5.L5 Post.forum.0 view.forum.0 Post.forum.1 Post.forum.1.2 view.forum.1
## 1     0            0            0            0              0            0
## 2     0            0            0            0              0            0
## 3     1            0            0            0              0            1
## 4     0            0            0            0              0            0
## 5     0            0            0            0              0            0
## 6     0            0            1            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            0              1
## 4              0            0              0            0              0
## 5              0            0              0            0              0
## 6              1            0              0            1              1
##   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            1              0            1
## 4            0            0            0              0            0
## 5            0            0            0              0            0
## 6            0            0            0              0            1
##   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            1              0            1              1
## 4              0            0              0            0              0
## 5              0            0              0            0              0
## 6              0            0              0            1              0
##   last.video last.quizz Engagement_Level Current.Score Section  Mot EMLyon
## 1          1          0        bystander            NA    <NA> <NA>   <NA>
## 2          2          0        bystander            NA    <NA> <NA>   <NA>
## 3         35          4        disengage            NA    <NA> <NA>   <NA>
## 4          2          0        bystander            NA    <NA> <NA>   <NA>
## 5          0          5        disengage            NA    <NA> <NA>   <NA>
## 6         16          5         complete            NA    <NA> <NA>   <NA>
##   Proba.reco EMLyon.et Assignment.choice Certif.bin EMLYON.et age
## 1         NA        NA                NA         NA      <NA>  NA
## 2         NA        NA                NA         NA      <NA>  NA
## 3         NA        NA                NA         NA      <NA>  NA
## 4         NA        NA                NA         NA      <NA>  NA
## 5         NA        NA                NA         NA      <NA>  NA
## 6         NA        NA                NA         NA      <NA>  NA
##   Post.forum.fonc.cours view.forum.fonc.cours
## 1                    NA                    NA
## 2                    NA                    NA
## 3                    NA                    NA
## 4                    NA                    NA
## 5                    NA                    NA
## 6                    NA                    NA
```

Linear Model

```
countries_HDI <- read_csv("Datasets/countries.HDI.csv", locale = locale(encoding = "ISO-8859-1"),
                          col_names = c("Country", "HDI", "Index"), show_col_types = FALSE)
```

##Recode Countries HDI

```
countries_HDI$HDI[countries_HDI$HDI == "M"] <- "I"
countries_HDI$HDI[countries_HDI$HDI == "H"] <- "I"
unique(countries_HDI$HDI)
```

```
## [1] "TH" "I"  "B"  NA
```

Compare the number of views of videos by gender

###Join all

```
df <- join_all(list(UE_all, countries_HDI),type = 'full', by = 'Country')
write.table(df, file = "data.csv",sep = "\t", row.names = T)
```

```
colnames(df)
```

```
##   [1] "Student_ID"            "Gender"                "birth.year"
##   [4] "Country"               "Diploma"               "Formation"
##   [7] "CSP"                   "How.heard"             "Exp.crea"
##  [10] "Curiosity.MOOC"        "Certif.self.sat"       "Rencontres"
##  [13] "Certif.work"           "Incitation"            "Temps.Dispo"
##  [16] "Exp.MOOC"              "Completion.proba"      "Instit.brand"
##  [19] "motiv.princ"           "diffic"                "encad.disp"
##  [22] "How.contact"           "entour"                "entour.inter"
##  [25] "Satisf"                "Eval.diffic"           "Estimated.hours"
##  [28] "Part.labo"             "Plat.satisf"           "Peer.eval.relev"
##  [31] "encad.diffic"          "Country_HDI"           "Country_HDI.fin"
##  [34] "CSP.fin"               "Temps.dispo.fin"       "Exam.score"
##  [37] "Exam.bin"              "Assignment.score"      "Assignment.bin"
##  [40] "Quizz.1.score"         "Quizz.1.bin"           "Quizz.2.score"
##  [43] "Quizz.2.bin"           "Quizz.3.score"         "Quizz.3.bin"
##  [46] "Quizz.4.bin"           "Quizz.4.score"         "Quizz.5.bin"
##  [49] "Quizz.5.score"         "Intro.MOOC"            "Prez.sem.1"
##  [52] "S1.L1"                 "S1.L2"                 "S1.L3"
##  [55] "S1.L4"                 "S1.L5"                 "S1.L6"
##  [58] "Prez.sem.2"            "S2.L1"                 "S2.L2"
##  [61] "S2.L3"                 "S2.L4"                 "S2.L5"
##  [64] "S2.L6"                 "Prez.sem.3"            "S3.L1.1"
##  [67] "S3.L1.2"               "S3.L2"                 "S3.L3"
##  [70] "S3.L4"                 "S3.L5"                 "Prez.sem.4"
##  [73] "S4.L1.1"               "S4.L1.2"               "S4.L2"
##  [76] "S4.L3"                 "S4.L4"                 "S4.L5"
##  [79] "Prez.sem.5"            "S5.L1.1"               "S5.L1.2"
##  [82] "S5.L2"                 "S5.L3"                 "S5.L4"
##  [85] "S5.L5"                 "Post.forum.0"          "view.forum.0"
##  [88] "Post.forum.1"          "Post.forum.1.2"        "view.forum.1"
##  [91] "view.forum.1.2"        "Post.forum.2"          "Post.forum.2.2"
##  [94] "view.forum.2"          "view.forum.2.2"        "Post.forum.3"
##  [97] "view.forum.3"          "Post.forum.4"          "Post.forum.4.2"
## [100] "view.forum.4"          "view.forum.4.2"        "Post.forum.5"
## [103] "Post.forum.5.2"        "view.forum.5"          "view.forum.5.2"
## [106] "last.video"            "last.quizz"            "Engagement_Level"
## [109] "Current.Score"         "Section"               "Mot"
## [112] "EMLyon"                "Proba.reco"            "EMLyon.et"
## [115] "Assignment.choice"     "Certif.bin"            "EMLYON.et"
## [118] "age"                   "Post.forum.fonc.cours" "view.forum.fonc.cours"
## [121] "HDI"                   "Index"
```

```
df$total_views <- rowSums(df[, c(
        'Prez.sem.2', 'S2.L1', 'S2.L2', 'S2.L3', 'S2.L4', 'S2.L5', 'S2.L6', 'Prez.sem.3', 'S3.L1.1', 'S
    )])
df$percent_video <- (df$total_views / 35) *100
```

Compare total views by gender

```
lm_1 <- lm(total_views~Gender, df)
lm_1
```

```
##
## Call:
## lm(formula = total_views ~ Gender, data = df)
##
## Coefficients:
##     (Intercept)  Genderune femme
##         11.8244           0.8331
```

```
ttest <- t.test(total_views~Gender,df)
ttest
```

```
##
##  Welch Two Sample t-test
##
## data:  total_views by Gender
## t = -3.2265, df = 5819.7, p-value = 0.00126
## alternative hypothesis: true difference in means between group un homme and group une femme is not e
## 95 percent confidence interval:
##  -1.3392215 -0.3269076
## sample estimates:
##  mean in group un homme mean in group une femme
##                11.82440                12.65747
```

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

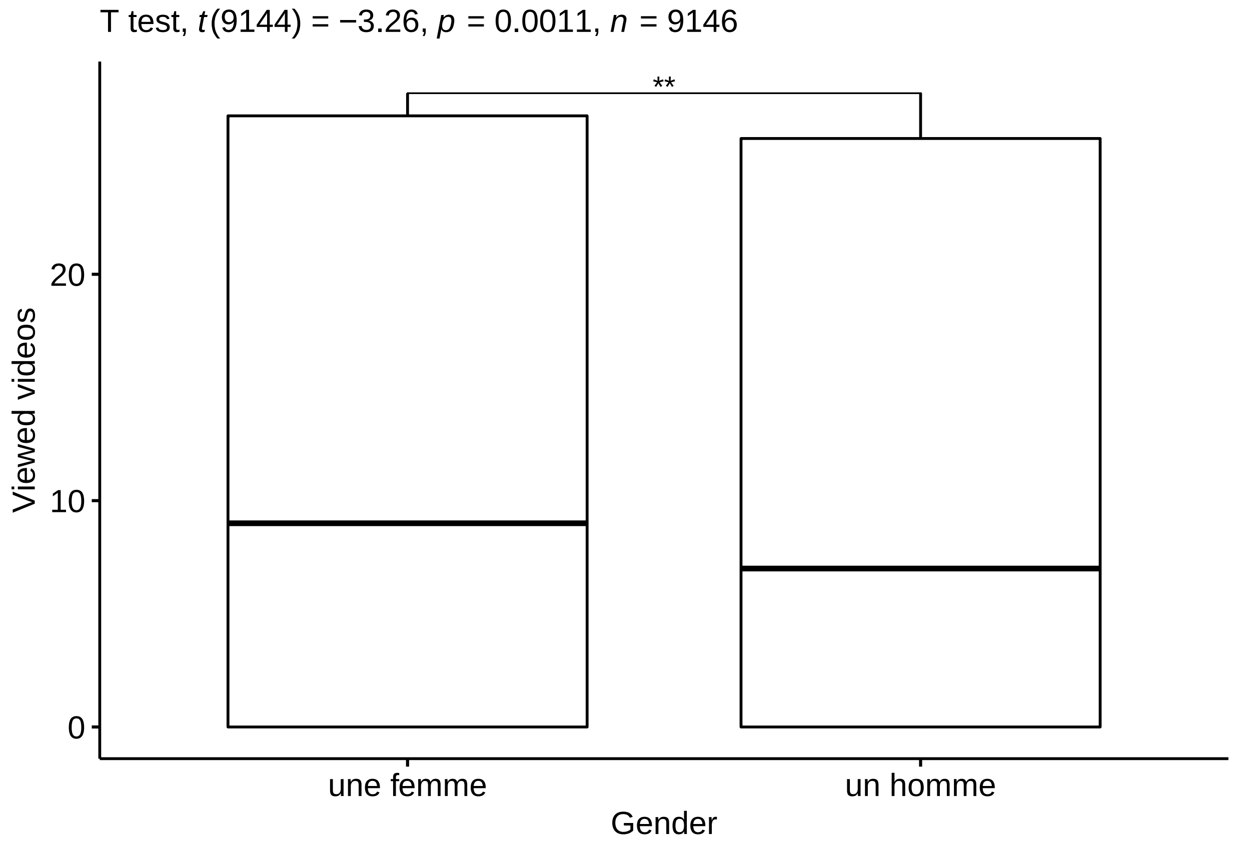
```
library(broom)
library(purrr)
library(xtable)
tab <- map_df(list(ttest), tidy)
xtable(tab)
```

```
## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Wed Jan  5 23:17:00 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrrrrrrll}
##   \hline
##  & estimate & estimate1 & estimate2 & statistic & p.value & parameter & conf.low & conf.high & metho
##   \hline
## 1 & -0.83 & 11.82 & 12.66 & -3.23 & 0.00 & 5819.73 & -1.34 & -0.33 & Welch Two Sample t-test & two.s
##    \hline
## \end{tabular}
## \end{table}
```

```
# Create a box-plot
library(rstatix)
```

```
##
## Attaching package: 'rstatix'

## The following objects are masked from 'package:plyr':
##
##     desc, mutate

## The following object is masked from 'package:stats':
##
##     filter
```

```
df.test <- df %>% drop_na(c(Gender, total_views))
stat <- df.test %>% t_test(total_views ~ Gender,var.equal=TRUE) %>%add_significance()
bxp <- ggboxplot(df.test, x = "Gender", y = "total_views", ylab = "Viewed videos", xlab = "Gender")
stat <- stat %>% add_xy_position(x = "Gender")
bxp + stat_pvalue_manual(stat, tip.length = 0) + labs(subtitle = get_test_label(stat, detailed = TRUE))
```



```
library(xtable)
anova_1 <- anova(lm_1)
print(xtable(anova_1))
```

```
## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Wed Jan  5 23:17:01 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## Gender & 1 & 1400.74 & 1400.74 & 10.62 & 0.0011 \\
##   Residuals & 9144 & 1206387.89 & 131.93 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

###Compare total views by HDI

```
lm_2 <- lm(total_views~HDI, df)
lm_2
```

```
##
## Call:
## lm(formula = total_views ~ HDI, data = df)
```

```
##
## Coefficients:
## (Intercept)         HDII        HDITH
##       1.509        8.236       11.758
```

Chi square test Gender & HDI

```
chi_1 <- chisq.test(df$Gender, df$HDI)
chi_1
```

```
##
##  Pearson's Chi-squared test
##
## data:  df$Gender and df$HDI
## X-squared = 74.738, df = 2, p-value < 2.2e-16
```

##One-Way ANOVA

```
a1 <- aov(total_views~HDI, df)
a1
```

```
## Call:
##    aov(formula = total_views ~ HDI, data = df)
##
## Terms:
##                       HDI Residuals
## Sum of Squares   655808.9 1367370.4
## Deg. of Freedom         2     21274
##
## Residual standard error: 8.017122
## Estimated effects may be unbalanced
## 5314 observations deleted due to missingness
```

```
print(xtable(a1))
```

```
## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Wed Jan  5 23:17:01 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## HDI & 2 & 655808.90 & 327904.45 & 5101.65 & 0.0000 \\
##   Residuals & 21274 & 1367370.37 & 64.27 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

Two-way ANOVA

```
a3 <-anova(lm(total_views ~ Gender + HDI, data = df))
print(xtable(a3))
```

```
## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Wed Jan  5 23:17:01 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## Gender & 1 & 840.12 & 840.12 & 6.47 & 0.0110 \\
##   HDI & 2 & 31896.86 & 15948.43 & 122.80 & 0.0000 \\
##   Residuals & 8181 & 1062452.68 & 129.87 &   &   \\
##    \hline
## \end{tabular}
## \end{table}
```

Two-way ANOVA

```
a4 <- anova(lm(total_views~Gender+HDI+HDI*Gender,df))
a4
```

```
## Analysis of Variance Table
##
## Response: total_views
##              Df  Sum Sq Mean Sq  F value  Pr(>F)
## Gender        1     840   840.1   6.4697 0.01099 *
## HDI           2   31897 15948.4 122.8177 < 2e-16 ***
## Gender:HDI    2     373   186.5   1.4360 0.23794
## Residuals  8179 1062080   129.9
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

```
print(xtable(a4))
```

```
## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Wed Jan  5 23:17:01 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## Gender & 1 & 840.12 & 840.12 & 6.47 & 0.0110 \\
##   HDI & 2 & 31896.86 & 15948.43 & 122.82 & 0.0000 \\
##   Gender:HDI & 2 & 372.94 & 186.47 & 1.44 & 0.2379 \\
##   Residuals & 8179 & 1062079.74 & 129.85 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

```
summary(a4)
```

```
##        Df             Sum Sq           Mean Sq           F value       
##  Min.   :   1.00   Min.   :    372.9   Min.   :  129.9   Min.   :  1.436  
##  1st Qu.:   1.75   1st Qu.:    723.3   1st Qu.:  172.3   1st Qu.:  3.953  
##  Median :   2.00   Median :  16368.5   Median :  513.3   Median :  6.470  
##  Mean   :2046.00   Mean   : 273797.4   Mean   : 4276.2   Mean   : 43.574  
##  3rd Qu.:2046.25   3rd Qu.: 289442.6   3rd Qu.: 4617.2   3rd Qu.: 64.644  
##  Max.   :8179.00   Max.   :1062079.7   Max.   :15948.4   Max.   :122.818  
##                                                          NA's   :1        
##      Pr(>F)        
##  Min.   :0.000000  
##  1st Qu.:0.005496  
##  Median :0.010991  
##  Mean   :0.082976  
##  3rd Qu.:0.124463  
##  Max.   :0.237936  
##  NA's   :1
```

##Model refinement, pairwise comparison

```
model_lm <- lm(total_views~Gender*HDI, df)
print(xtable(summary(model_lm)))
```

```
## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Wed Jan  5 23:17:01 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrr}
##   \hline
##  & Estimate & Std. Error & t value & Pr($>$$|$t$|$) \\
##   \hline
## (Intercept) & 5.4418 & 0.5342 & 10.19 & 0.0000 \\
##   Genderune femme & 2.2974 & 1.3026 & 1.76 & 0.0778 \\
##   HDII & 4.4681 & 0.8142 & 5.49 & 0.0000 \\
##   HDITH & 7.7436 & 0.5599 & 13.83 & 0.0000 \\
##   Genderune femme:HDII & -2.7496 & 1.6757 & -1.64 & 0.1009 \\
##   Genderune femme:HDITH & -2.0289 & 1.3331 & -1.52 & 0.1280 \\
##    \hline
## \end{tabular}
## \end{table}
```

```
step(model_lm,direction = 'forward')
```

```
## Start:  AIC=39837.6
## total_views ~ Gender * HDI
```

```
##
## Call:
## lm(formula = total_views ~ Gender * HDI, data = df)
##
## Coefficients:
##           (Intercept)        Genderune femme                    HDII
##                 5.442                  2.297                   4.468
##                 HDITH   Genderune femme:HDII  Genderune femme:HDITH
##                 7.744                 -2.750                 -2.029
```

```
#step(model_lm,direction = 'backward')
```

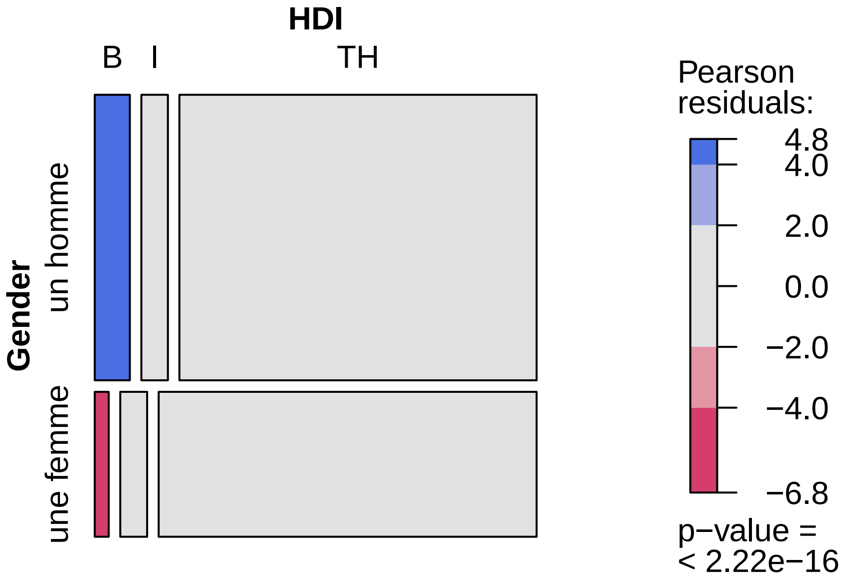
```
chisq.test(df$Gender,df$HDI,correct=FALSE)
```

```
##
##  Pearson’s Chi-squared test
##
## data:  df$Gender and df$HDI
## X-squared = 74.738, df = 2, p-value < 2.2e-16
```

```
ct <- chisq.test(df$Gender,df$HDI,correct=FALSE)
library(xtable)
tab2 <- map_df(list(ct), tidy)
xtable(tab2)
```

```
## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Wed Jan  5 23:17:01 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrl}
##   \hline
##  & statistic & p.value & parameter & method \\
##   \hline
## 1 & 74.74 & 0.00 &   2 & Pearson’s Chi-squared test \\
##    \hline
## \end{tabular}
## \end{table}
```

```
library("grid"); library("vcd")
mosaic(~Gender+HDI, df, shade = TRUE, legend = TRUE)
```



```
tk <- TukeyHSD(a1)
tk
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = total_views ~ HDI, data = df)
##
## $HDI
##           diff       lwr       upr p adj
## I-B   8.236187  7.398123  9.074250     0
## TH-B 11.757502 11.482794 12.032211     0
## TH-I  3.521316  2.669258  4.373373     0
```

```
tab3 <- map_df(list(tk), tidy)
xtable(tab3)
```

```
## % latex table generated in R 4.1.1 by xtable 1.8-4 package
## % Wed Jan  5 23:17:01 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{rllrrrrr}
##   \hline
##  & term & contrast & null.value & estimate & conf.low & conf.high & adj.p.value \\
##   \hline
## 1 & HDI & I-B & 0.00 & 8.24 & 7.40 & 9.07 & 0.00 \\
##   2 & HDI & TH-B & 0.00 & 11.76 & 11.48 & 12.03 & 0.00 \\
##   3 & HDI & TH-I & 0.00 & 3.52 & 2.67 & 4.37 & 0.00 \\
##    \hline
## \end{tabular}
## \end{table}
```

```r
TukeyHSD(aov(total_views ~ Gender + HDI, data = df), "HDI")
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = total_views ~ Gender + HDI, data = df)
##
## $HDI
##          diff      lwr      upr p adj
## I-B  3.811803 2.176471 5.447136 1e-07
## TH-B 7.328076 6.142803 8.513349 0e+00
## TH-I 3.516273 2.303846 4.728699 0e+00
```

```r
pairwise.t.test(df$total_views, df$Gender, p.adj = "none")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  df$total_views and df$Gender
##
##           un homme
## une femme 0.0011
##
## P value adjustment method: none
```

```r
pairwise.t.test(df$total_views, df$Gender, p.adj = "bonf")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  df$total_views and df$Gender
##
##           un homme
## une femme 0.0011
##
## P value adjustment method: bonferroni
```

```r
pairwise.t.test(df$total_views, df$Gender, p.adj = "holm")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  df$total_views and df$Gender
##
##           un homme
## une femme 0.0011
##
## P value adjustment method: holm
```

```
TukeyHSD(a1)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = total_views ~ HDI, data = df)
##
## $HDI
##            diff       lwr       upr p adj
## I-B    8.236187  7.398123  9.074250     0
## TH-B  11.757502 11.482794 12.032211     0
## TH-I   3.521316  2.669258  4.373373     0
```

```
df$age <- 2021 - df$birth.year
df$age_group <- cut(df$age, breaks = seq(0, 100, by = 30))
table(df$age_group)
```

```
##
##  (0,30] (30,60] (60,90]
##    1179    7066     559
```

```
a3 <- aov(total_views~age_group, data = df)
a3
```

```
## Call:
##    aov(formula = total_views ~ age_group, data = df)
##
## Terms:
##                 age_group Residuals
## Sum of Squares     3657.6 1163306.1
## Deg. of Freedom         2      8795
##
## Residual standard error: 11.50083
## Estimated effects may be unbalanced
## 17793 observations deleted due to missingness
```

```
TukeyHSD(a3)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = total_views ~ age_group, data = df)
##
## $age_group
##                      diff        lwr      upr     p adj
## (30,60]-(0,30]  0.3465751 -0.5015949 1.194745 0.6036166
## (60,90]-(0,30]  2.8960536  1.5116365 4.280471 0.0000029
## (60,90]-(30,60] 2.5494785  1.3649497 3.734007 0.0000014
```

#Logistic Regression

```r
colnames(df)
```

```
##   [1] "Student_ID"            "Gender"                "birth.year"
##   [4] "Country"               "Diploma"               "Formation"
##   [7] "CSP"                   "How.heard"             "Exp.crea"
##  [10] "Curiosity.MOOC"        "Certif.self.sat"       "Rencontres"
##  [13] "Certif.work"           "Incitation"            "Temps.Dispo"
##  [16] "Exp.MOOC"              "Completion.proba"      "Instit.brand"
##  [19] "motiv.princ"           "diffic"                "encad.disp"
##  [22] "How.contact"           "entour"                "entour.inter"
##  [25] "Satisf"                "Eval.diffic"           "Estimated.hours"
##  [28] "Part.labo"             "Plat.satisf"           "Peer.eval.relev"
##  [31] "encad.diffic"          "Country_HDI"           "Country_HDI.fin"
##  [34] "CSP.fin"               "Temps.dispo.fin"       "Exam.score"
##  [37] "Exam.bin"              "Assignment.score"      "Assignment.bin"
##  [40] "Quizz.1.score"         "Quizz.1.bin"           "Quizz.2.score"
##  [43] "Quizz.2.bin"           "Quizz.3.score"         "Quizz.3.bin"
##  [46] "Quizz.4.bin"           "Quizz.4.score"         "Quizz.5.bin"
##  [49] "Quizz.5.score"         "Intro.MOOC"            "Prez.sem.1"
##  [52] "S1.L1"                 "S1.L2"                 "S1.L3"
##  [55] "S1.L4"                 "S1.L5"                 "S1.L6"
##  [58] "Prez.sem.2"            "S2.L1"                 "S2.L2"
##  [61] "S2.L3"                 "S2.L4"                 "S2.L5"
##  [64] "S2.L6"                 "Prez.sem.3"            "S3.L1.1"
##  [67] "S3.L1.2"               "S3.L2"                 "S3.L3"
##  [70] "S3.L4"                 "S3.L5"                 "Prez.sem.4"
##  [73] "S4.L1.1"               "S4.L1.2"               "S4.L2"
##  [76] "S4.L3"                 "S4.L4"                 "S4.L5"
##  [79] "Prez.sem.5"            "S5.L1.1"               "S5.L1.2"
##  [82] "S5.L2"                 "S5.L3"                 "S5.L4"
##  [85] "S5.L5"                 "Post.forum.0"          "view.forum.0"
##  [88] "Post.forum.1"          "Post.forum.1.2"        "view.forum.1"
##  [91] "view.forum.1.2"        "Post.forum.2"          "Post.forum.2.2"
##  [94] "view.forum.2"          "view.forum.2.2"        "Post.forum.3"
##  [97] "view.forum.3"          "Post.forum.4"          "Post.forum.4.2"
## [100] "view.forum.4"          "view.forum.4.2"        "Post.forum.5"
## [103] "Post.forum.5.2"        "view.forum.5"          "view.forum.5.2"
## [106] "last.video"            "last.quizz"            "Engagement_Level"
## [109] "Current.Score"         "Section"               "Mot"
## [112] "EMLyon"                "Proba.reco"            "EMLyon.et"
## [115] "Assignment.choice"     "Certif.bin"            "EMLYON.et"
## [118] "age"                   "Post.forum.fonc.cours" "view.forum.fonc.cours"
## [121] "HDI"                   "Index"                 "total_views"
## [124] "percent_video"         "age_group"
```

```r
df$Gender <- relevel(factor(df$Gender), ref ="une femme")
```

```r
lr <- glm(Exam.bin~HDI+Gender, data=df)
lr
```

```
##
## Call:  glm(formula = Exam.bin ~ HDI + Gender, data = df)
```

```
##
## Coefficients:
##    (Intercept)            HDII           HDITH  Genderun homme
##        0.12105         0.07082         0.08350        -0.02025
##
## Degrees of Freedom: 8184 Total (i.e. Null);  8181 Residual
##   (18406 observations deleted due to missingness)
## Null Deviance:       1233
## Residual Deviance: 1228  AIC: 7712
```

```
lra <-aov(lr)
lra
```

```
## Call:
##    aov(formula = lr)
##
## Terms:
##                          HDI    Gender Residuals
## Sum of Squares     3.8812    0.7428 1228.0671
## Deg. of Freedom         2         1      8181
##
## Residual standard error: 0.387443
## Estimated effects may be unbalanced
## 18406 observations deleted due to missingness
```

```
lras <- summary(lra)
lras
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## HDI             2    3.9  1.9406  12.928 2.48e-06 ***
## Gender          1    0.7  0.7428   4.948   0.0261 *
## Residuals   8181 1228.1  0.1501
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 18406 observations deleted due to missingness
```

```
exp(coef(lr))
```

```
##    (Intercept)           HDII          HDITH Genderun homme
##      1.1286787      1.0733930      1.0870798      0.9799551
```

```
data(df)
```

```
## Warning in data(df): data set 'df' not found
```

```
df %>%
  summary_factorlist("Exam.bin", c("Gender","HDI"),
  p=TRUE, add_dependent_label=TRUE) -> t1
```

```
## Note: dependent includes missing data. These are dropped.
```

```r
knitr::kable(t1, row.names=FALSE, align=c("l", "l", "r", "r", "r"))
```

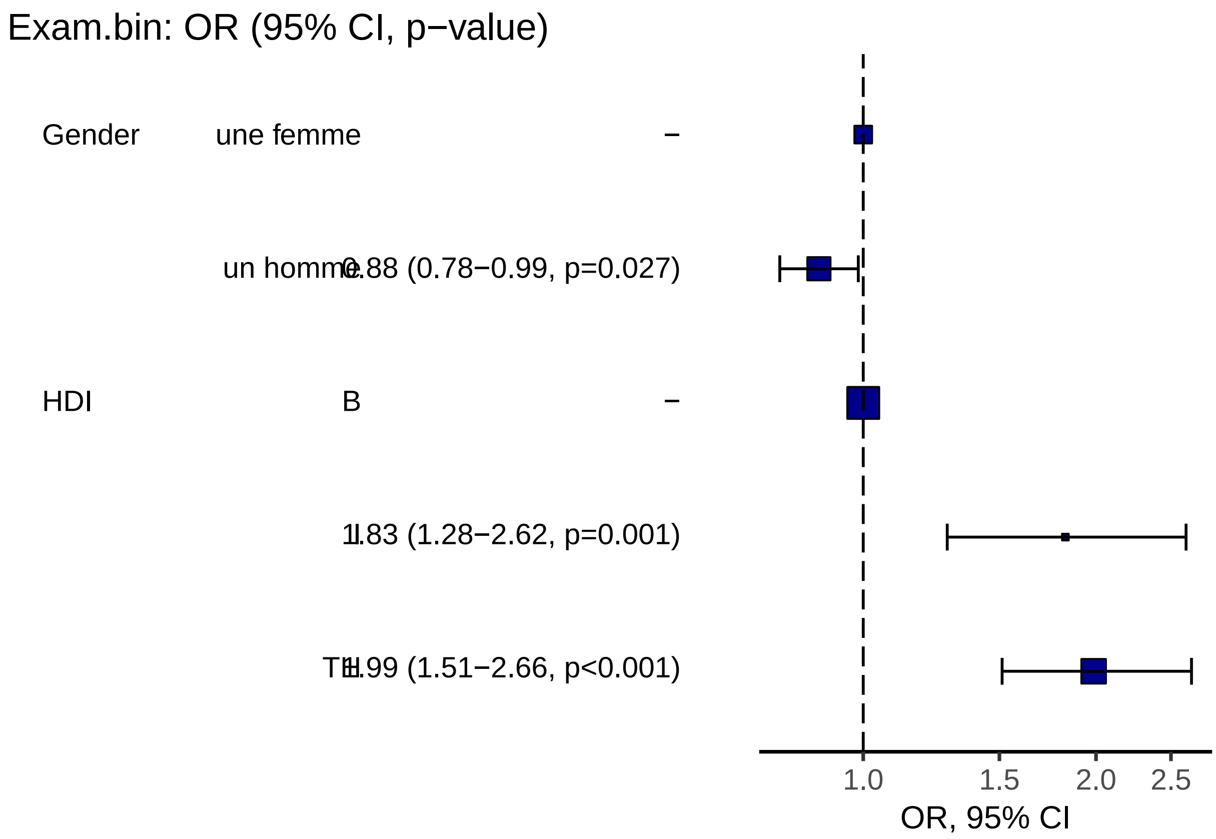
Dependent: Exam.bin		unit	value	p
Gender	une femme	Mean (sd)	0.2 (0.4)	0.015
	un homme	Mean (sd)	0.2 (0.4)	
HDI	B	Mean (sd)	0.0 (0.1)	<0.001
	I	Mean (sd)	0.2 (0.4)	
	TH	Mean (sd)	0.2 (0.4)	

```r
df %>%
  or_plot("Exam.bin", c("Gender","HDI"), table_text_size=4, title_text_size=14,
    plot_opts=list(xlab("OR, 95% CI"), theme(axis.title = element_text(size=12))), ref =)
```

```
## Note: dependent includes missing data. These are dropped.
```

```
## Waiting for profiling to be done...
## Waiting for profiling to be done...
## Waiting for profiling to be done...
```

```
## Warning: Removed 2 rows containing missing values (geom_errorbarh).
```



```r
po <- glm(Exam.bin~CSP.fin+Estimated.hours,df, family="binomial")
po
```

```
##
## Call:  glm(formula = Exam.bin ~ CSP.fin + Estimated.hours, family = "binomial",
##     data = df)
##
## Coefficients:
##                                    (Intercept)
##                                        -4.5789
##                                   CSP.finAutre
##                                        -1.4961
##  CSP.finCadres et professions intellectuelles
##                                        -0.6161
##                             CSP.finEmploy\xe9s
##                                        -1.3317
##                   CSP.finEn recherche d’emploi
##                                         0.5122
##                               CSP.finEtudiants
##                                         0.0835
##         Estimated.hoursDe 1 \xe0\xa0 2 heures
##                                         8.5306
##             Estimated.hoursDe 2 \xe0 4 heures
##                                        -0.4502
##         Estimated.hoursDe 2 \xe0\xa0 4 heures
##                                         9.1855
##     Estimated.hoursDe 30 minutes \xe0 1 heure
##                                        -0.6395
## Estimated.hoursDe 30 minutes \xe0\xa0 1 heure
##                                         7.3130
##             Estimated.hoursDe 4 \xe0 8 heures
##                                         0.6215
##         Estimated.hoursDe 4 \xe0\xa0 8 heures
##                                         8.0970
##            Estimated.hoursMoins de 30 minutes
##                                         2.8912
##               Estimated.hoursPlus de 8 heures
##                                         3.0494
##
## Degrees of Freedom: 3126 Total (i.e. Null);  3112 Residual
##   (23464 observations deleted due to missingness)
## Null Deviance:       2837
## Residual Deviance: 431.9     AIC: 461.9
```

```r
library(gtsummary)
```

```
##
## Attaching package: ’gtsummary’

## The following object is masked from ’package:plyr’:
##
##     mutate
```

```
summary(po)$coefficients
```

```
##                                                       Estimate Std. Error    z value
## (Intercept)                                        -4.57889274  0.5943110 -7.7045395
## CSP.finAutre                                       -1.49610602  0.8148707 -1.8360042
## CSP.finCadres et professions intellectuelles       -0.61605849  0.5922618 -1.0401793
## CSP.finEmploy\xe9s                                 -1.33174543  0.7461197 -1.7848952
## CSP.finEn recherche d'emploi                        0.51215304  0.6255080  0.8187794
## CSP.finEtudiants                                    0.08349917  0.5984863  0.1395173
## Estimated.hoursDe 1 \xe0\xa0 2 heures               8.53060473  0.5133436 16.6177277
## Estimated.hoursDe 2 \xe0 4 heures                  -0.45018653  0.5958735 -0.7555069
## Estimated.hoursDe 2 \xe0\xa0 4 heures               9.18548165  0.6970719 13.1772373
## Estimated.hoursDe 30 minutes \xe0 1 heure          -0.63953710  0.7773207 -0.8227454
## Estimated.hoursDe 30 minutes \xe0\xa0 1 heure       7.31295744  0.5262878 13.8953590
## Estimated.hoursDe 4 \xe0 8 heures                   0.62153604  0.7923318  0.7844391
## Estimated.hoursDe 4 \xe0\xa0 8 heures               8.09700506  1.0943736  7.3987574
## Estimated.hoursMoins de 30 minutes                  2.89115427  0.4842116  5.9708490
## Estimated.hoursPlus de 8 heures                     3.04940535  0.6534582  4.6665651
##                                                       Pr(>|z|)
## (Intercept)                                        1.313159e-14
## CSP.finAutre                                       6.635703e-02
## CSP.finCadres et professions intellectuelles       2.982566e-01
## CSP.finEmploy\xe9s                                 7.427831e-02
## CSP.finEn recherche d'emploi                       4.129123e-01
## CSP.finEtudiants                                   8.890414e-01
## Estimated.hoursDe 1 \xe0\xa0 2 heures              5.186135e-62
## Estimated.hoursDe 2 \xe0 4 heures                  4.499449e-01
## Estimated.hoursDe 2 \xe0\xa0 4 heures              1.186569e-39
## Estimated.hoursDe 30 minutes \xe0 1 heure          4.106528e-01
## Estimated.hoursDe 30 minutes \xe0\xa0 1 heure      6.758378e-44
## Estimated.hoursDe 4 \xe0 8 heures                  4.327825e-01
## Estimated.hoursDe 4 \xe0\xa0 8 heures              1.374647e-13
## Estimated.hoursMoins de 30 minutes                 2.360221e-09
## Estimated.hoursPlus de 8 heures                    3.062766e-06
```

```
#tbl_regression(po, exponentiate = TRUE)
```

```
ll <- glm(total_views~Engagement_Level,df, family="poisson")
ll
```

```
##
## Call:  glm(formula = total_views ~ Engagement_Level, family = "poisson",
##     data = df)
##
## Coefficients:
##               (Intercept)  Engagement_Levelbystander
##                    0.7167                   -19.0193
##  Engagement_Levelcomplete  Engagement_Leveldisengage
##                    2.2879                     1.2175
##
## Degrees of Freedom: 22239 Total (i.e. Null);  22236 Residual
##   (4351 observations deleted due to missingness)
## Null Deviance:       317600 
## Residual Deviance: 93160     AIC: 128800
```

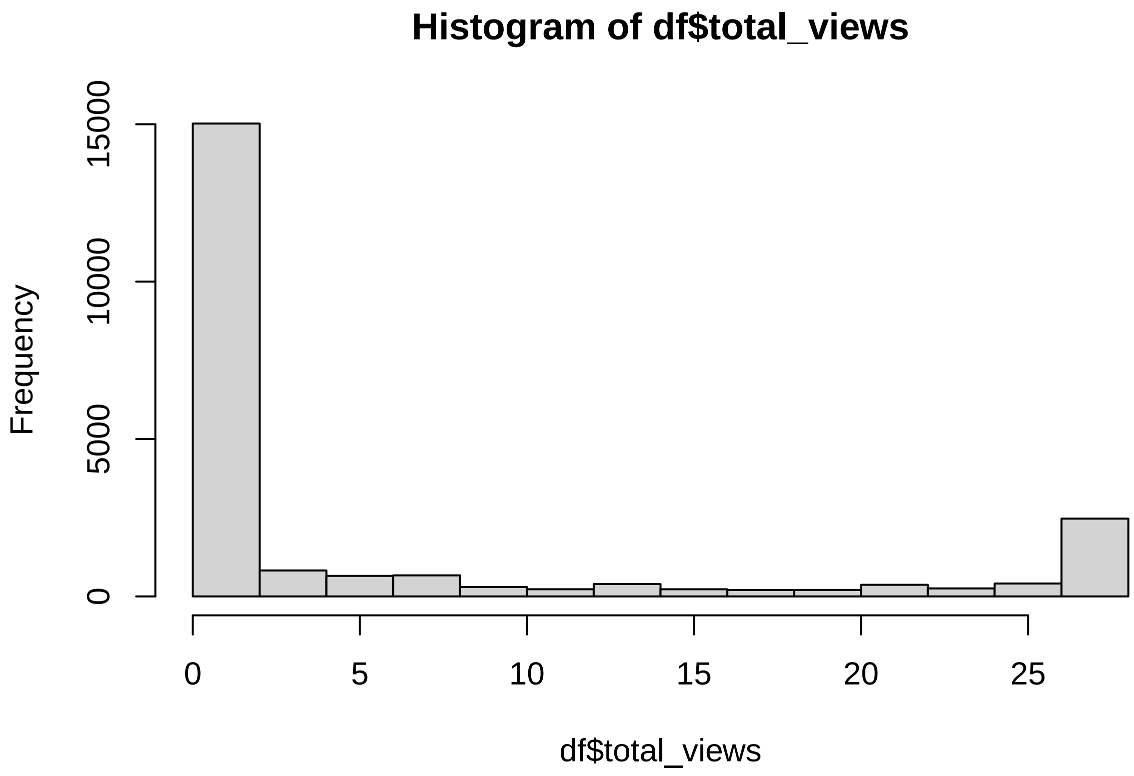
```
exp(coef(po))
```

```
##                                  (Intercept) 
##                                 1.026626e-02 
##                                 CSP.finAutre 
##                                 2.240007e-01 
##  CSP.finCadres et professions intellectuelles 
##                                 5.400689e-01 
##                             CSP.finEmploy\xe9s 
##                                 2.640160e-01 
##                 CSP.finEn recherche d’emploi 
##                                 1.668880e+00 
##                             CSP.finEtudiants 
##                                 1.087084e+00 
##         Estimated.hoursDe 1 \xe0\xa0 2 heures 
##                                 5.067509e+03 
##             Estimated.hoursDe 2 \xe0 4 heures 
##                                 6.375092e-01 
##         Estimated.hoursDe 2 \xe0\xa0 4 heures 
##                                 9.754477e+03 
##     Estimated.hoursDe 30 minutes \xe0 1 heure 
##                                 5.275366e-01 
## Estimated.hoursDe 30 minutes \xe0\xa0 1 heure 
##                                 1.499606e+03 
##             Estimated.hoursDe 4 \xe0 8 heures 
##                                 1.861786e+00 
##         Estimated.hoursDe 4 \xe0\xa0 8 heures 
##                                 3.284616e+03 
##            Estimated.hoursMoins de 30 minutes 
##                                 1.801409e+01 
##               Estimated.hoursPlus de 8 heures 
##                                 2.110279e+01
```

#Survival Analysis

```
# Load required packages
library(survival)
library(survminer)
library(dplyr)
```

```
hist(df$total_views)
```



```
total_views.dec = quantile(df$total_views, probs = seq(.1, .9, by = .1), na.rm = TRUE)

df<-df %>%mutate(total_views.decile = ntile(total_views, 10))
```

```
df$status.vid=rep(NA, nrow(df))
    for (i in 1:nrow(df)) {
    if (is.na(df$total_views.decile[i]<10)) {df$status.vid[i]=1}
    if (is.na(df$total_views.decile[i]==10)) {df$status.vid[i]=0}
    }
```

```
df_s <- read.table('data.csv')
```

```
df$status.vid=rep(NA, nrow(df))
    for (i in 1:nrow(df)) {
    if (is.na(df$total_views.decile[i]<10)) {df$status.vid[i]=1}
    if (is.na(df$total_views.decile[i]==10)) {df$status.vid[i]=0}
    }
```

```
df_s$n_videos <- rowSums(df_s[, c(
        'Prez.sem.2', 'S2.L1', 'S2.L2', 'S2.L3', 'S2.L4', 'S2.L5', 'S2.L6', 'Prez.sem.3', 'S3.L1.1', 'S3
    )])
```

```
df_s$percent_video <- (df_s$n_videos / 35) *100
n_videos_dec = quantile(df_s$n_videos, probs = seq(.1, .9, by = .1), na.rm = TRUE)
df_s <-df_s %>%mutate(n.videos.decile = ntile(n_videos, 10))
df_s$status.vid=rep(NA, nrow(df_s))
df_s <- df_s %>% mutate(status.vid = ifelse(n.videos.decile < 10, 1, 0))
```

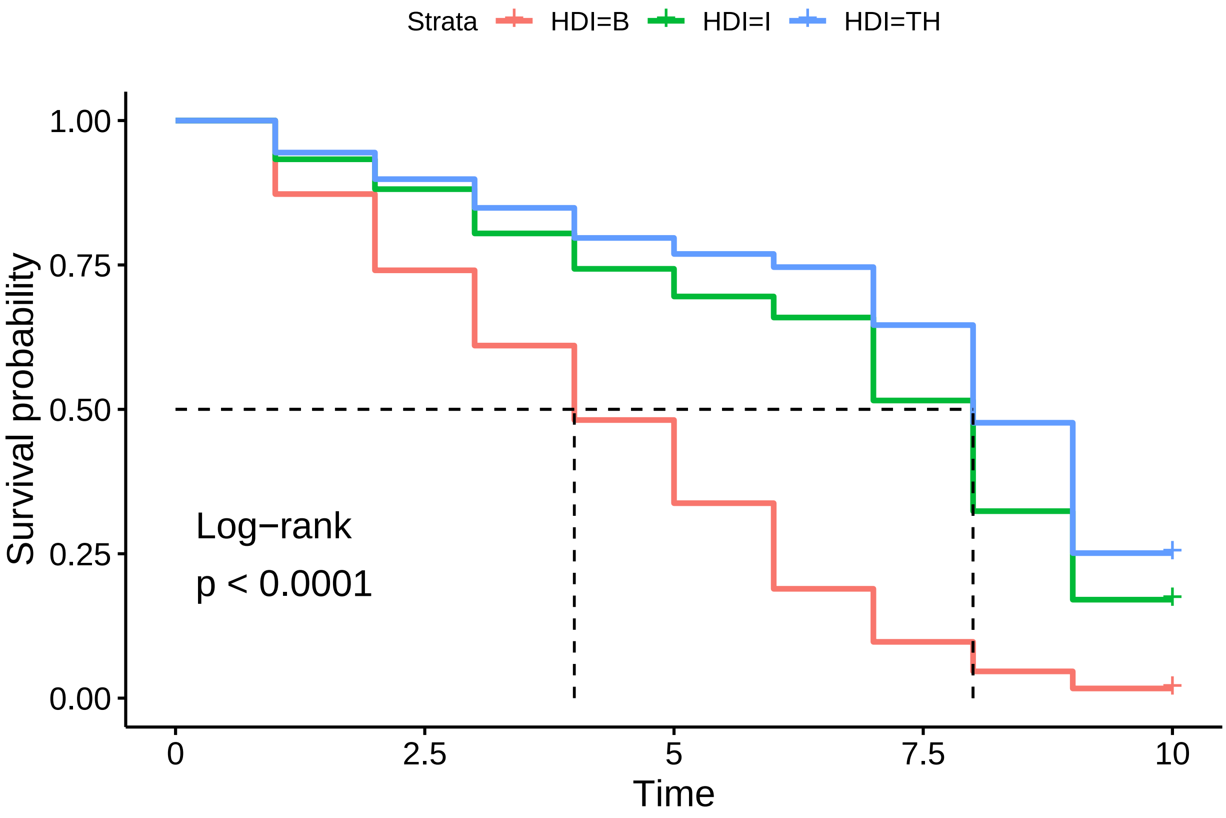
```r
df_s$Group <- factor(df_s$Engagement_Level, levels = c("disengage","audit"))
head(df_s)
```

```
##   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 \xe9quivalent)
## 4       1086 une femme       1967     France  Bac+5 (Master ou \xe9quivalent)
## 5       1948 une femme       1983  Allemagne             Bac ou \xe9quivalent
## 6      16209 une femme         NA Madagascar Bac+3 (Licence ou \xe9quivalent)
##                                                                      Formation
## 1                                                                         <NA>
## 2                                                                         <NA>
## 3                                                                        Droit
## 4 Sciences sociales (\xe9conomie\\, 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, commer\xe7ants, chefs d’entreprise
## 5                                  Employ\xe9s
## 6                Professions interm\xe9diaires
##                                 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 exp\xe9rience en cr\xe9ation d’entreprise
## 4      Je suis en train de cr\xe9er mon entreprise (phase de d\xe9marrage)
## 5                 Je n’ai aucune exp\xe9rience en cr\xe9ation d’entreprise
## 6 J\x92ai un projet de cr\xe9ation d\x92entreprise (phase de r\xe9flexion)
##   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 premi\xe8re participation \xe0 un MOOC
## 4 Entre une et deux heures Non, c’est ma premi\xe8re participation \xe0 un MOOC
## 5 Entre une et deux heures                  Oui, mais tous suivis partiellement
## 6 Entre une et deux heures Non, c’est ma premi\xe8re participation \xe0 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 Exam.score
## 1                                          <NA>                 <NA>         NA
## 2                                          <NA>                 <NA>         NA
## 3        Cadres et professions intellectuelles Moins de deux heures         NA
## 4 Artisans, commer\xe7ants, chefs d’entreprise Moins de deux heures         NA
## 5                                  Employ\xe9s Moins de deux heures         NA
## 6                                        Autre Moins de deux heures         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
##   Engagement_Level Current.Score Section  Mot EMLyon Proba.reco EMLyon.et
## 1        bystander            NA    <NA> <NA>   <NA>         NA        NA
## 2        bystander            NA    <NA> <NA>   <NA>         NA        NA
## 3        bystander            NA    <NA> <NA>   <NA>         NA        NA
## 4        disengage            NA    <NA> <NA>   <NA>         NA        NA
## 5        bystander            NA    <NA> <NA>   <NA>         NA        NA
## 6        disengage            NA    <NA> <NA>   <NA>         NA        NA
##   Assignment.choice Certif.bin EMLYON.et age Post.forum.fonc.cours
## 1                NA         NA      <NA>  NA                    NA
## 2                NA         NA      <NA>  NA                    NA
## 3                NA         NA      <NA>  NA                    NA
## 4                NA         NA      <NA>  NA                    NA
## 5                NA         NA      <NA>  NA                    NA
## 6                NA         NA      <NA>  NA                    NA
##   view.forum.fonc.cours HDI Index n_videos percent_video n.videos.decile
## 1                    NA   B   158        0             0               1
## 2                    NA   B   164        0             0               1
## 3                    NA  TH    20        0             0               1
## 4                    NA  TH    20       28            80              10
## 5                    NA  TH     5        0             0               1
## 6                    NA   B   150        0             0               1
##   status.vid     Group
## 1          1      <NA>
## 2          1      <NA>
## 3          1      <NA>
## 4          0 disengage
## 5          1      <NA>
## 6          1 disengage
```

```
coxph(formula = Surv(n.videos.decile, status.vid) ~ HDI, data = df_s)
```

```
## Call:
## coxph(formula = Surv(n.videos.decile, status.vid) ~ HDI, data = df_s)
##
##           coef exp(coef) se(coef)      z      p
## HDII  -1.05024   0.34986  0.04913 -21.38 <2e-16
## HDITH -1.34263   0.26116  0.01745 -76.95 <2e-16
##
## Likelihood ratio test=6708  on 2 df, p=< 2.2e-16
## n= 21277, number of events= 19170
##    (5314 observations deleted due to missingness)
```

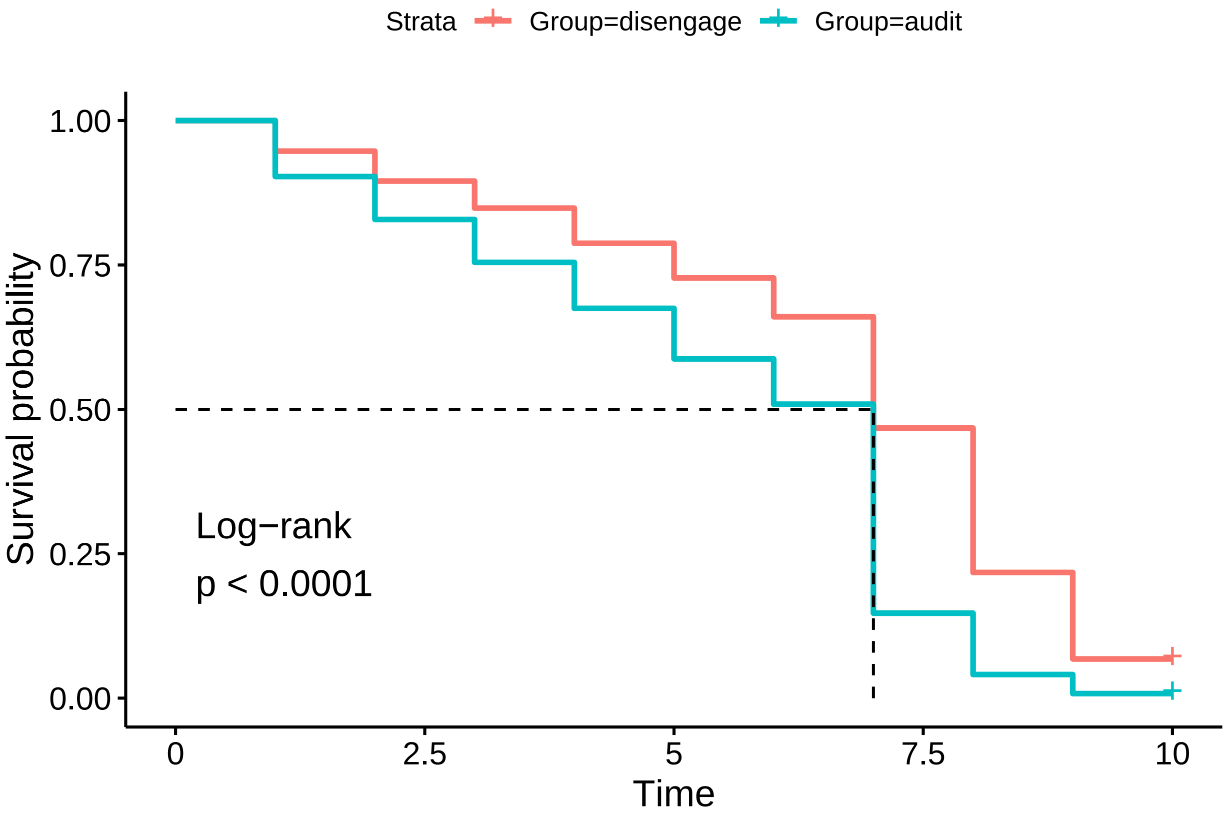
```
survival_HDI <- survfit(Surv(n.videos.decile, status.vid) ~ HDI, data = df_s)

ggsurvplot(
  survival_HDI,
  conf.int = FALSE,
  surv.median.line = c('hv'),
  data = df_s,
  pval = TRUE,
  pval.method = TRUE,
  risk.table = FALSE)
```



```
coxph(formula = Surv(n.videos.decile, status.vid) ~ Group, data = df_s)
```

```
## Call:
## coxph(formula = Surv(n.videos.decile, status.vid) ~ Group, data = df_s)
##
##                coef exp(coef) se(coef)    z      p
## Groupaudit 0.69427   2.00225  0.02789 24.9 <2e-16
##
## Likelihood ratio test=560.9  on 1 df, p=< 2.2e-16
## n= 7233, number of events= 6857
##    (19358 observations deleted due to missingness)
```

```
survival_Group <- survfit(Surv(n.videos.decile, status.vid) ~ Group, data = df_s)
ggsurvplot(
  survival_Group,
  conf.int = FALSE,
  surv.median.line = c('hv'),
  data = df_s,
  pval = TRUE,
  pval.method = TRUE,
  risk.table = FALSE)
```


Strata
Group=disengage
Group=audit
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p < 0.0001
0
2.5
5
7.5
10
Time