Udemy

In this file, we have information about the different courses offered in Udemy.

This is the url where I obtained this data from: https://www.kaggle.com/andrewmvd/udemy-courses

We will start the analysis importing the data.

```
data=read.csv("udemy_courses.csv")
head(data)
```

```
##   course_id                                                course_title
## 1   1070968                          Ultimate Investment Banking Course
## 2   1113822 Complete GST Course & Certification - Grow Your CA Practice
## 3   1006314    Financial Modeling for Business Analysts and Consultants
## 4   1210588          Beginner to Pro - Financial Analysis in Excel 2017
## 5   1011058               How To Maximize Your Profits Trading Options
## 6    192870        Trading Penny Stocks: A Guide for All Levels In 2017
##                                                                                 url
## 1                      https://www.udemy.com/ultimate-investment-banking-course/
## 2                                   https://www.udemy.com/goods-and-services-tax/
## 3 https://www.udemy.com/financial-modeling-for-business-analysts-and-consultants/
## 4       https://www.udemy.com/complete-excel-finance-course-from-beginner-to-pro/
## 5             https://www.udemy.com/how-to-maximize-your-profits-trading-options/
## 6              https://www.udemy.com/trading-penny-stocks-a-guide-for-all-levels/
##   is_paid price num_subscribers num_reviews num_lectures             level
## 1    True   200            2147          23           51         All Levels
## 2    True    75            2792         923          274         All Levels
## 3    True    45            2174          74           51 Intermediate Level
## 4    True    95            2451          11           36         All Levels
## 5    True   200            1276          45           26 Intermediate Level
## 6    True   150            9221         138           25         All Levels
##   content_duration  published_timestamp          subject
## 1              1.5 2017-01-18T20:58:58Z Business Finance
## 2             39.0 2017-03-09T16:34:20Z Business Finance
## 3              2.5 2016-12-19T19:26:30Z Business Finance
## 4              3.0 2017-05-30T20:07:24Z Business Finance
## 5              2.0 2016-12-13T14:57:18Z Business Finance
## 6              3.0 2014-05-02T15:13:30Z Business Finance
```

```
dim(data)
```

```
## [1] 3678   12
```

In the data frame we have the titles of the different courses, the first thing we will do, is to analyze them.

```
head(data$course_title)
```

```
## [1] Ultimate Investment Banking Course
## [2] Complete GST Course & Certification - Grow Your CA Practice
## [3] Financial Modeling for Business Analysts and Consultants
## [4] Beginner to Pro - Financial Analysis in Excel 2017
```

```
## [5] How To Maximize Your Profits Trading Options
## [6] Trading Penny Stocks: A Guide for All Levels In 2017
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

To analyze the titles, we will use the library "stringr". With this library we will be able to see if a certain word is contained in a string. Furthermore, we will be able to search information like the most used words. Since I am sure that we will be using plots and other functionalities from the "tidyverse" library, I will also charge it.

```
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------------------ tidyverse 1.3.0 --
## v ggplot2 3.3.0     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   0.8.5
## v tidyr   1.0.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## -- Conflicts --------------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(stringr)
```

I would like to start by selecting just the most popular courses (above the 80th percentile) and learning about the frequency of use of different words.

In this way, we can search further details about the most successful courses.

```
selection=data[data$num_subscribers>quantile(data$num_subscribers,0.80),]
```

To obtain this information, we will follow a process.

1- The "course title" feature is a "factor", we will transform it to a column of characters.

2- The column has different references; we will collapse all the information in a single string. Furthermore, with the function "tolower" we will force all the capital letters to become lowercase letters.

3- From the big string that we created in the last step, we will now divide it, creating a list of words.

4- We will create a table that shows the frequencies of the words. We will then save the information in a data frame.

5-If we transform the table into a data frame, the resulting words will become its index. Since we want to analyze the words, it would be better if they were part of the data frame and not its index. Because of that, we will now join the index with the information of the data frame. In this way we will obtain a data frame with the words and their frequencies.

6-We will now substitute the actual index of the data frame (the words) with a numeric index.

7-We will name the two existing columns as "Word" and "Freq".

8-Finally, since we want to find the most used words, we sort the data frame by the frequency in descending order.

```
strings=as.character(selection$course_title) #1
strings=tolower(paste(strings, collapse=' ' ))#2
strings=str_split(strings,pattern = " ") #3
strings=data.frame(as.matrix(table(strings)))#4
strings= cbind(name = rownames(strings),strings)#5
rownames(strings) = 1:nrow(strings) #6
colnames(strings)= c("Word","Freq") #7
strings=arrange(strings,desc(Freq)) #8
```

Now that we have the data as we wanted, we can print the top 10 most used words.

As we can see, it includes a lot of prepositions and signs like "-". We will now delete these words and signs because we are searching for a more specific kind of terms.

I prefer not to delete the word "for", because I assume that in many titles, the word "for" will be used to point for whom the course is made for ("for artists", "for beginners", "for managers"...). I think that to keep this word can provide interesting results for this analysis.

```
head(strings,10)
```

```
##        Word Freq
## 1        to  196
## 2         a  133
## 3       for  128
## 4     learn  118
## 5         -  116
## 6       and  108
## 7       the  103
## 8       web  103
## 9      with   85
## 10 complete   75
```

```
strings=strings[-c(which(strings$Word=="to"| strings$Word=="a"| strings$Word=="-"|strings$Word=="the"|s
```

Now that we have just the interesting words in this top 10 list, we can proceed to analyze them.

I think that the words that we can see in this list, have different purposes, I will now explain what, I assume, will be its purpose. We will later examine if my suppositions are correct or not:

-"Generic Keywords": words like "learn", "how" or "course" are words that people tend to search when it comes to find courses or anything related with knowledge.

-"Specific Keywords": words like "web", "website" or "javascript" show the kind of skills that people is demanding.

-"Direction Keywords": words like "for", "complete", "from" or "beginners" show who is the target of the course. Since "complete" and "beginners" are part of this list, I think that many courses will claim to cover all the skills necessary to start the course as a beginner and to become an expert.

```
head(strings,10)
```

```
##          Word Freq
## 3         for  128
## 4       learn  118
## 8         web  103
## 10   complete   75
## 11        how   75
## 13       from   72
## 15     course   69
## 16    website   68
## 17 javascript   67
## 18  beginners   64
```

We will now examine the titles to see if the assumptions I made are correct or not.

To do that, we will use the function "grepl" searching for every word in the top 10 list.

— "For" —

As we said, the word "for" is normally used to explain who the target of the course is. We must highlight that in many cases, as we can see in the titles, the word "for" is followed by the word "Beginners" or similar. This fact happens in 63 out of 153 references.

```
head(selection[grepl("for",tolower(selection$course_title)),"course_title"])
```

```
## [1] Trading Penny Stocks: A Guide for All Levels In 2017
## [2] Forex Trading Secrets of the Pros With Amazon's AWS
## [3] Financial Management Risk and Return For Securities
## [4] Forex Trading Course: Work Smarter Not Harder Proven Results
## [5] Options Trading Stocks: Proven Toolbox For Financial Success
## [6] Forex Trading A-Zâ„¢ - With LIVE Examples of Forex Trading
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

```
length(selection[grepl("for",tolower(selection$course_title)),"course_title"])#Total number of sentence
```

```
## [1] 156
```

```
length(selection[grepl("for beginners",tolower(selection$course_title)),"course_title"])+
length(selection[grepl("for absolute beginners",tolower(selection$course_title)),"course_title"])#Total
```

```
## [1] 63
```

— "Beginners" —

As expected, most of the sentences that include the word "Beginners" also use the word "for". 76 sentences contain the word "Beginners" and, as we have already seen, 63 of them use both words, "for" and "Beginners".

```
head(selection[grepl("beginners",tolower(selection$course_title)),"course_title"])
```

```
## [1] Trading for Beginners - Intermediate Level
## [2] Stock Market Investing for Beginners
## [3] Accounting Basics in 66 Minutes (absolutely for beginners)
## [4] Accounting Is Easy (for Beginners)
## [5] Beginners Binary Options Course
## [6] Trading for Beginners - Entry Level
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

```
length(selection[grepl("beginners",tolower(selection$course_title)),"course_title"])#Total number of se
```

```
## [1] 76
```

— "From" —

Regarding the word "from" we can see from the different titles, that it works similarly to "for". There are many titles that use the words "From Scratch" instead of "for beginners".

```
head(selection[grepl("from",tolower(selection$course_title)),"course_title"])
```

```
## [1] Create A Business From Home Trading Stocks Today In 2017
## [2] Financial Model Basics: Build a model from start to finish
## [3] Build a DCF Model from Scratch
## [4] Ultimate Photoshop Training: From Beginner to Pro
## [5] Sketch 3 from A to Z: Become an App Designer
## [6] Figure Drawing From Life Using The Reilly Technique.
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

— "Complete" —

According to the titles that we can see in the following table, the word "complete" is used to communicate that the course starts with basic knowledge and escalates on difficulty, ending with content for experts.

We can assume that the courses that include the word "complete" tend to be longer than the other courses. To find if this is true, we will calculate the mean duration of the courses that include this word and then the mean of the courses that do not include the word.

As we can see, the duration of those courses who claim to be "complete" is twice as long as the courses that don't include this word.

```
head(selection[grepl("complete",tolower(selection$course_title)),"course_title"])
```

```
## [1] Beginner to Pro in PowerPoint: Complete PowerPoint Training
## [2] Investing 101: The Complete Online Investing Course
## [3] The Complete Bitcoin Course: Get .001 Bitcoin In Your Wallet
## [4] The Complete Investment Banking Course 2017
## [5] The Complete Financial Analyst Course 2017
## [6] Accounting & Financial Statement Analysis: Complete Training
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

```
mean(selection[grepl("complete",tolower(selection$course_title)),"content_duration"]) #Mean of courses
```

```
## [1] 11.67333
```

```
mean(selection[-grepl("complete",tolower(selection$course_title)),"content_duration"])#Mean of courses
```

```
## [1] 5.754807
```

— "Javascript" and "Web" or "Website" —

As we expected, javascript and web or website act as specific keywords for the skills that the students want to learn.

```
#Web
head(selection[grepl("web",tolower(selection$course_title)),"course_title"])
```

```
## [1] Website Investing 101 - Buying & Selling Online Businesses
## [2] DIY Design Professional Web Banners in Photoshop 4 Beginners
## [3] Web Elements Design With Photoshop
## [4] How To Make Graphics For A Website
## [5] How To Build Your Own Web Banner Design Business
## [6] Learn Web Designing & HTML5/CSS3 Essentials in 4-Hours
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

```
#Website
head(selection[grepl("website",tolower(selection$course_title)),"course_title"])
```

```
## [1] Website Investing 101 - Buying & Selling Online Businesses
## [2] How To Make Graphics For A Website
## [3] Learning Dynamic Website Design - PHP MySQL and JavaScript
## [4] How To Make A Wordpress Website 2017 | Divi Theme Tutorial
## [5] Ultimate Web Developer Course Build 10 Websites from Scratch
## [6] Rapid Website Design with Bootstrap
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

```
#Javascript
head(selection[grepl("javascript",tolower(selection$course_title)),"course_title"])
```

```
## [1] Learning Dynamic Website Design - PHP MySQL and JavaScript
## [2] Learn JavaScript for beginners
## [3] Accelerated JavaScript Training
## [4] Advanced Javascript
## [5] Start Writing JavaScript Today - Beginner JavaScript Course
```

```
## [6] Javascript ES6! A Complete Reference Guide to Javascript ES6
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

— "How" —

The word "how" is almost always followed by the word "to", expressing a question that the course promises to solve. For example, "How to create a routine Trading" or "How To Invest With Tiny Capital In Stocks?".

```
head(selection[grepl("how",tolower(selection$course_title)),"course_title"])
```

```
## [1] Options Trading - How to Win with Weekly Options
## [2] How to Win 97% of Your Options Trades
## [3] How to Create Your Personal Budget
## [4] How to Pick The Right Penny Stocks To Invest In 2017
## [5] CPA 101: How To Master Affiliate Marketing In No Time
## [6] Bitcoin or How I Learned to Stop Worrying and Love Crypto
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

```
head(selection[grepl("course",tolower(selection$course_title)),"course_title"])
```

```
## [1] Forex Trading Course: Work Smarter Not Harder Proven Results
## [2] Investing 101: The Complete Online Investing Course
## [3] The Complete Bitcoin Course: Get .001 Bitcoin In Your Wallet
## [4] The Complete Investment Banking Course 2017
## [5] The Complete Financial Analyst Course 2017
## [6] Excel Crash Course: Master Excel for Financial Analysis
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

```
head(selection[grepl("learn",tolower(selection$course_title)),"course_title"])
```

```
## [1] Basic Technical Analysis: Learn the structure of the market
## [2] Learn to Trade for Profit: Find and Trade Winning Stocks
## [3] Learn Accounting. Understand Business.
## [4] Elite Trend Trader: Learn To Trade Stocks, Options & Forex
## [5] Learn to Trade for Profit:Trading with Japanese Candlesticks
## [6] Learn to Trade the Stock Market without Blowing Your Profits
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

— "Course" and "Learn" —

Regarding the words "course" and "learn" both are keywords are, in every one of the tittles analyzed, related with the knowledge/learning field.

```
head(selection[grepl("course",tolower(selection$course_title)),"course_title"])
```

```
## [1] Forex Trading Course: Work Smarter Not Harder Proven Results
## [2] Investing 101: The Complete Online Investing Course
## [3] The Complete Bitcoin Course: Get .001 Bitcoin In Your Wallet
## [4] The Complete Investment Banking Course 2017
## [5] The Complete Financial Analyst Course 2017
## [6] Excel Crash Course: Master Excel for Financial Analysis
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

```
head(selection[grepl("learn",tolower(selection$course_title)),"course_title"])
```

```
## [1] Basic Technical Analysis: Learn the structure of the market
## [2] Learn to Trade for Profit: Find and Trade Winning Stocks
## [3] Learn Accounting. Understand Business.
## [4] Elite Trend Trader: Learn To Trade Stocks, Options & Forex
## [5] Learn to Trade for Profit:Trading with Japanese Candlesticks
```

```
## [6] Learn to Trade the Stock Market without Blowing Your Profits
## 3663 Levels: 'Geometry Of Chance strategy of defeating the roulette." ...
```

Now that we have analyzed the titles, we will find how powerful these words are to attract subscribers.

To do that, we will examine how the data is distributed across different quantiles. If the courses that include one of the top 10 keywords show a notorious difference between their number of subscribers and the rest of the courses, then we will determine that the keywords have a good power of attraction.

The data that we have, show a peculiar distribution. A huge majority of courses have a small number of subscribers and a small quantity of courses accumulate a lot of subscribers.

In other words, the 10% most popular courses have a great number of subscribers but those courses that are not among this percentage have a relatively small number of students enrolled.

To be able to obtain a good understanding of the data, we will examine its distribution with the median and the 10% most popular courses.

To do that, we will start creating a "for loop" that will have 2 objectives: 1- To include in a new data frame the rows where the title has any of the top 10 keywords that we have already seen. 2- For every row in the data frame, the loop will also include the detected keyword in a list that we will join to the same data frame at the end.

```
df_title=data.frame()
detected_word=NULL
for (i in 1:nrow(data)) {
  selec=tolower(as.character(data$course_title[i]))

  for (word in 1:10){
    strings[word,1]
    if (length(grep(as.character(strings[word,1]),selec))>0){
      df_title=rbind(df_title,data[i,])
      detected_word=c(detected_word,as.character(strings[word,1]))
    }
  }
}

df_title=cbind(df_title,data.frame(detected_word))

head(df_title)
```

```
##    course_id                                                course_title
## 1    1070968                            Ultimate Investment Banking Course
## 2    1113822 Complete GST Course & Certification - Grow Your CA Practice
## 21   1113822 Complete GST Course & Certification - Grow Your CA Practice
## 3    1006314    Financial Modeling for Business Analysts and Consultants
## 5    1011058                  How To Maximize Your Profits Trading Options
## 6     192870        Trading Penny Stocks: A Guide for All Levels In 2017
##                                                                        url
## 1                       https://www.udemy.com/ultimate-investment-banking-course/
## 2                                    https://www.udemy.com/goods-and-services-tax/
## 21                                   https://www.udemy.com/goods-and-services-tax/
## 3  https://www.udemy.com/financial-modeling-for-business-analysts-and-consultants/
## 5               https://www.udemy.com/how-to-maximize-your-profits-trading-options/
## 6                https://www.udemy.com/trading-penny-stocks-a-guide-for-all-levels/
##    is_paid price num_subscribers num_reviews num_lectures             level
## 1     True   200            2147          23           51         All Levels
## 2     True    75            2792         923          274         All Levels
```

```
## 21    True    75            2792         923          274         All Levels
## 3     True    45            2174          74           51 Intermediate Level
## 5     True   200            1276          45           26 Intermediate Level
## 6     True   150            9221         138           25         All Levels
##    content_duration  published_timestamp          subject detected_word
## 1               1.5 2017-01-18T20:58:58Z Business Finance        course
## 2              39.0 2017-03-09T16:34:20Z Business Finance      complete
## 21             39.0 2017-03-09T16:34:20Z Business Finance        course
## 3               2.5 2016-12-19T19:26:30Z Business Finance           for
## 5               2.0 2016-12-13T14:57:18Z Business Finance           how
## 6               3.0 2014-05-02T15:13:30Z Business Finance           for
```

Now that we have all the information that we want in a data frame, we will start the analysis.

One thing that we need to take into consideration, is that with the "for loop", we have searched for any string that contains a certain word. Since the word "web" is contained in the word "website" the loop will detect those two words, even if it was not supposed to detect the first one.

In order to solve this problem and considering that "web" and "website" are synonyms and are used to communicate the same idea, we will delete the rows that have detected the word "website".

```
df_title=df_title[-which(df_title$detected_word=="website"),] #We proceed to delete the rows containg t

df_title%>%
  select(num_subscribers,detected_word)%>%
  group_by(detected_word)%>%
  summarise(min=min(num_subscribers),
            median=quantile(num_subscribers,0.50),
            quantile90=quantile(num_subscribers,0.90),
            max=max(num_subscribers),
                    number_of_registers=n())
```

```
## # A tibble: 9 x 6
##   detected_word   min median quantile90    max number_of_registers
##   <fct>         <int>  <dbl>      <dbl>  <int>               <int>
## 1 beginners         0  1364.      9442.  70773                 288
## 2 complete          0  2204      16461. 114512                 189
## 3 course            0  1701      11253. 114512                 252
## 4 for               0  1028       7284. 161029                 768
## 5 from              0  1723      12804. 268923                 212
## 6 how               0  1402       8487.  65576                 270
## 7 javascript      244  2771      15191   84897                 136
## 8 learn             0  1132      10406. 268923                 555
## 9 web               0  2865      17071  121584                 401
```

```
data%>%
  select(num_subscribers)%>%
  summarise(min=min(num_subscribers),
            median=quantile(num_subscribers,0.50),
            quantile90=quantile(num_subscribers,0.90),
            max=max(num_subscribers),
                    number_of_registers=n(),)
```

```
##   min median quantile90    max number_of_registers
## 1   0  911.5     7211.6 268923                3678
```

As we can notice in the first data frame above, the data is grouped by the different "keywords". Each one of

these words have its own minimum, median, 90th quantile, maximum and number of registers.

As we said before, to determine whether this "keywords" are useful to attract subscribers, we have created another dataset with the same information applying it to the whole data frame of titles. We can see this information in the second data frame above. From now on, I will refer to this last data frame as "base information".

From the "base information" we can extract that, the minimum of subscribers in a certain course is 0. The median is 911.5 subscribers and the 90th quantile is 7211.6 subscribers. We Will now proceed to compare this information with the information obtained by the top 10 keywords.

The first information that we can notice is that none of the courses that teach "javascript" has 0 subscribers, the minimum registered is 244. This is an interesting fact since it shows the huge demand of courses about this subject that exist.

Another aspect that we can see is that the word "for" doesn't get better results compared with the "base information". This is a clear indicator than this word is not a relevant keyword. It is just a preposition widely used for the titles but without power to attract new subscribers.

The word "for", is not the only one that obtains this result, "learn", "how" or "beginners" show the same situation. The words "from" and "course" obtain higher median and quantile if we compare them with the "base information". However, the results are not high enough to determine that these words have a significant effect.

Finally, the words that can be considered "keywords" due to the difference between their results and the ones in the base information are "complete", "javascript" and "web".

The analysis that we have just made, show an interesting result. Even if there are words that are frequently used to communicate an idea ("for", "beginners", "learn"...), what people is really interested in, are the specific skills of the courses ("web", "javascript"...). Furthermore, it is interesting to use the word "complete" as a promise that the course will be a good investment for the students.

Now that we have found that the skills offered by the course is what people is more interested about, we will search for the first 5 most used skills according to our data. To do that, we will call the 20 words with the highest frequencies and I will just pick the words that refer to a certain skill.

According to the information we have in the following table, the most offered skills among the most popular courses in the platform are "web", "javascript", "html", "wordpress" and "css".

```
head(strings,20)
```

```
##           Word Freq
## 3          for  128
## 4        learn  118
## 8          web  103
## 10    complete   75
## 11         how   75
## 13        from   72
## 15      course   69
## 16     website   68
## 17  javascript   67
## 18   beginners   64
## 19        html   64
## 20       build   63
## 21      create   56
## 22      design   52
## 23   wordpress   49
## 24     scratch   47
## 25 development   46
```

```
## 26        your   46
## 27         css   45
## 28    beginner   39
```

Now we will continue or analysis focusing on the numeric data.

We will use the library GGally to create a plot that will show the correlation that exist between numeric data and the distribution of every numeric feature.

From the next plot, we can obtain the following conclusions:

1- There is an important correlation between the features "num_lectures" and the "content_duration".

2- There is a weak correlation between the number of reviews and the number of subscribers.

3-The distribution of the prices show that most of the courses are in the low-price tier and the rest are irregularly spread forming smaller groups.

We will start this part of the analysis focusing on the 2nd and 3rd information points. We won't make emphasis on the 1st information point, because it is obvious that the longer a course is, the more lectures it will have.

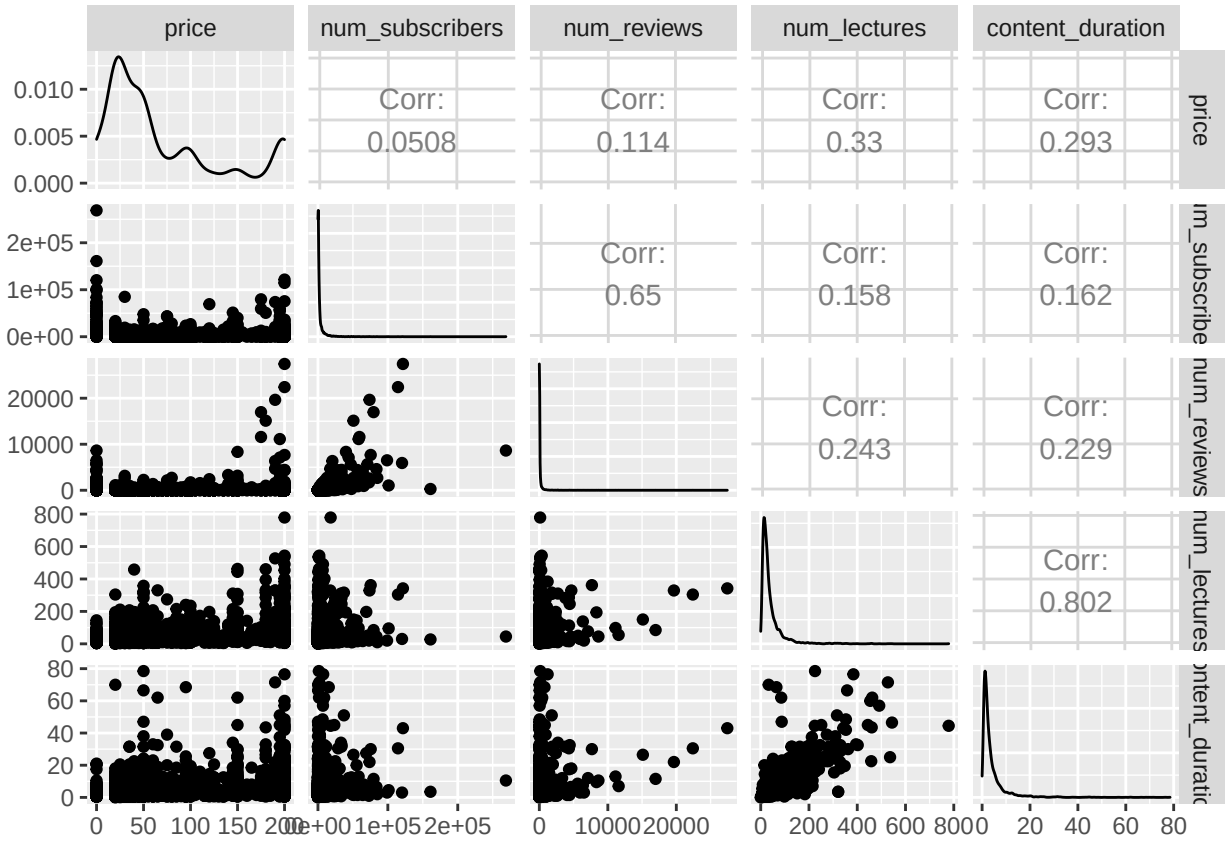
```
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
##
## Attaching package: 'GGally'
```

```
## The following object is masked from 'package:dplyr':
##
##     nasa
```

```
ggpairs(data[,c(5:8,10)])
```



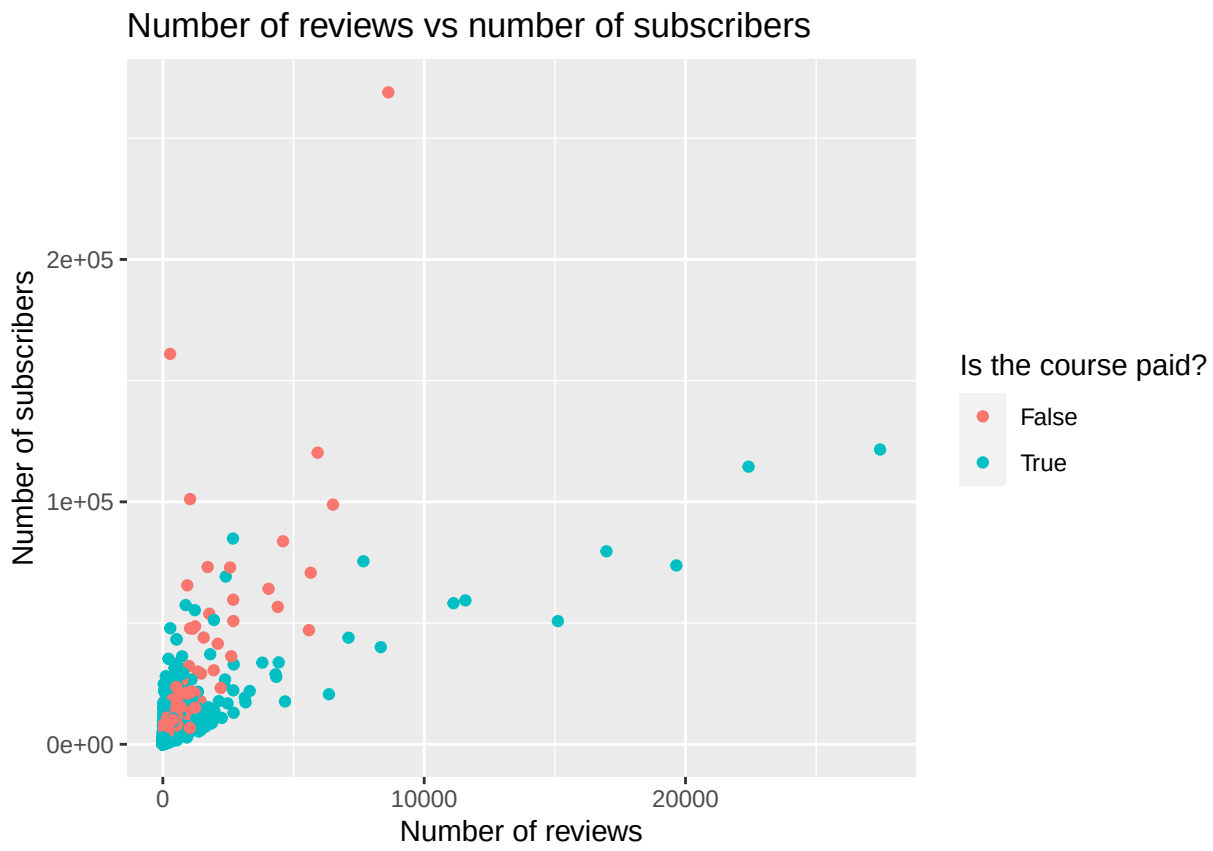
We will now learn more about the relation that exist between the number of reviews and the number of subscribers.

According to psychological principles, it makes sense that some degree of correlation exists between these two features. Many people prefer to take a course that has been already rated by other individuals. In that way, it is easier to determine whether the course will be a good investment or not.

One thing that we need to keep in mind is that this data frame only indicates the number of reviews, not if they are good or bad. If a user sees many reviews but most of them are bad, then he/she will probably not enroll in that course. If the reviews are good, then he/she may take the risk. This explains why the correlation is not higher than 0.65.

Another factor to consider that may be important, is the price. If a course if free, then the user does not take any monetary risk and thus, I assume that it will be easier to see free courses with a low number of reviews and a high level of users. I also assume that the cheaper a course is, the easier will be for a user to pay the price and thus the number of reviews will be less relevant.

```
ggplot(data)+
  geom_point(aes(num_reviews,num_subscribers,color=is_paid))+
  labs(x = "Number of reviews",y="Number of subscribers", colour = "Is the course paid? ",title="Number
```



In the last plot, we can see the points that correspond to the free courses painted in red and the ones that are paid in blue/green.

As we said, in one hand, the courses that are free, obtain many subscribers with just a few reviews.

On the other hand, the courses that need to be payed, show an unclear distribution. Some courses show a strong correlation between the number of reviews and the subscribers, like the ones in the right part of the plot, and other have many subscribers having just a few reviews.

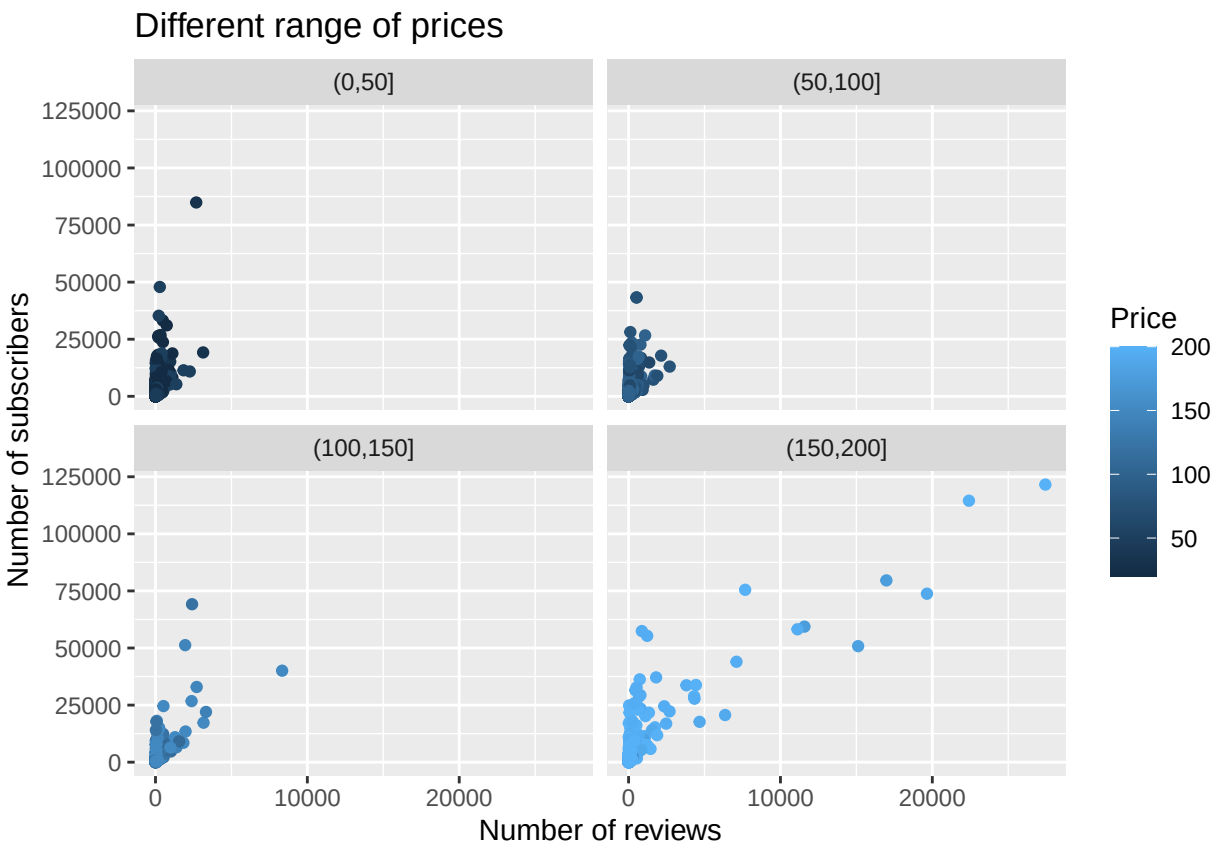
This different behavior may be explained by the price of the courses.

To analyze the price of the courses, we will start by creating a new data frame with just the courses that are not free. Then, we will add a feature with 4 different price ranges (From 0 to 50, from 50 to 100, from 100 to 150 and from 150 to 200). We do this because we will create a scatterplot dividing the graph by the range of price.

As we can see in the following plot, the courses that have a price lower than 100, $reachagoodnumberofsubscribers, evenifthenur$ to 150, $starttoshowaweakcorrelationbetweenthevariables. Finally, thosecoursesthatcostmorethan150$ show a defined relation between the number of comments and the number of subscribers.

We can conclude that the more expensive a course is, the more important are the reviews.

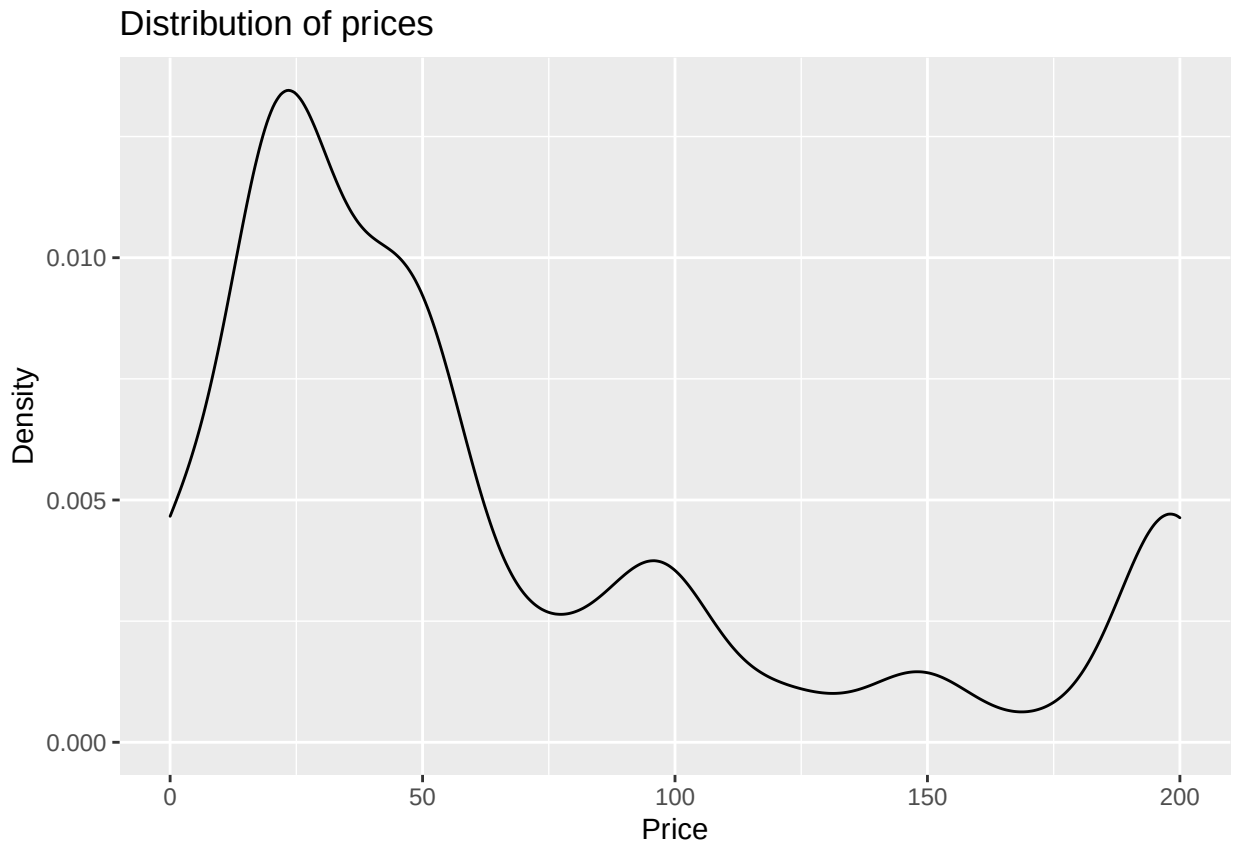
```
data_paid=data[data$is_paid=="True",]
data_paid$prices_cut=cut(data_paid$price,breaks = c(0,50,100,150,200))

ggplot(data_paid)+
  geom_point(aes(num_reviews,num_subscribers,color
                 =price))+
  facet_wrap(~prices_cut)+
  labs(x = "Number of reviews",y="Number of subscribers", colour = "Price",title="Different range of pri
```



We will continue examining the distribution of the prices.

As we can see in the following density plot, the distribution is not uniform. There are certain prices that agglomerate a great part of the courses.

```
ggplot()+
  geom_density(aes(data$price))+
  labs(x = "Price",y="Density",title="Distribution of prices")
```



According to the following proportion table, most of the prices do not group more than 0.6 % of the offered courses. However, there are prices that are so popular that are used by 25.66% of the courses.

The distribution of the prices show a very interesting pattern c that we can divide in behavior of low prices, and behavior of high prices.

```
prop.table(table(data$price))
```

```
##
##           0          20          25          30          35          40
## 0.0842849375 0.2256661229 0.0418705818 0.0451332246 0.0315388798 0.0589994562
##          45          50          55          60          65          70
## 0.0225666123 0.1272430669 0.0095160413 0.0203915171 0.0081566069 0.0070690593
##          75          80          85          90          95         100
## 0.0220228385 0.0078847200 0.0084284937 0.0051658510 0.0413268080 0.0361609570
##         105         110         115         120         125         130
## 0.0029907558 0.0021750952 0.0043501903 0.0100598151 0.0081566069 0.0008156607
##         135         140         145         150         155         160
## 0.0021750952 0.0027188690 0.0073409462 0.0239260468 0.0005437738 0.0010875476
##         165         170         175         180         185         190
## 0.0016313214 0.0013594345 0.0035345296 0.0046220772 0.0016313214 0.0027188690
##         195         200
## 0.0345296357 0.0802066340
```

We will now explore both behaviors. These conducts are commonly applied by business when they are operating in a competitive market.

Udemy is offering many courses from different content creators. Since all the producers want to sell their

courses, they are forced to compete by adopting one of the two strategies.

#BEHAVIOUR OF LOW PRICES.

Most of the courses seem to have a reduced price, as we can see in the following table, 63.73% (0.63730288) of them costs 50$ or less.

According to the business theory, we could say that they are applying a "cost leadership strategy", trying to attract a lot of consumers by lowering the prices.

```
cumsum(prop.table(table(data$price)))
```

```
##          0         20         25         30         35         40         45
## 0.08428494 0.30995106 0.35182164 0.39695487 0.42849375 0.48749320 0.51005982
##         50         55         60         65         70         75         80
## 0.63730288 0.64681892 0.66721044 0.67536705 0.68243611 0.70445895 0.71234367
##         85         90         95        100        105        110        115
## 0.72077216 0.72593801 0.76726482 0.80342577 0.80641653 0.80859163 0.81294182
##        120        125        130        135        140        145        150
## 0.82300163 0.83115824 0.83197390 0.83414899 0.83686786 0.84420881 0.86813486
##        155        160        165        170        175        180        185
## 0.86867863 0.86976618 0.87139750 0.87275693 0.87629146 0.88091354 0.88254486
##        190        195        200
## 0.88526373 0.91979337 1.00000000
```

In the next table, we can see that the courses that have a price of $25, 30$, $35, 40$ and 45$ represent between 3% and 5% of the total number of courses.

The interesting thing to notice here, is that the courses with a cost of 20$ represent a 22.56% and that this is the lowest price that exists (Without including the free courses).

It would be interesting to see how the prices have evolved with time. Nonetheless, considering the existing difference between the percentages, I assume that many courses started costing between 25$ and 50$ and with time, they lowered the cost to be more competitive, reaching the minimum price of 20$.

We can also appreciate, that the courses with a price of 50$ represent a 12.72%. This percentage shows a great difference compared with the courses that cost 45$ and 55$, each one of them representing less than 0.25%.

I presume that the curses that cost 50$ are charging this quantity for 2 reasons: 1- They are courses that were able to obtain a certain degree of differentiation from the cheaper courses and thus do not need to compete that much reducing its price. 2- The producers of the course are applying a technique that in marketing is called "psychological prices".

The "psychological prices" technique has two variants.

-The first one is to reduce the price in order to avoid reaching a certain number. For example, selling a book for 9.99$ instead of 10 dollars.

-The second one, is to charge a quantity that is easy to process for our brain or that looks simpler. For example, the value of 50$ is more appealing and familiar to us than 51.76$.

We can also see the second variant of the "psychological prices" applied in the price of $75, 100$, 150$ and $200. Furthermore, as we will see soon, in the price of 95$ and 195$, it seems that the content creators are applying the first variant of the technique.

#BEHAVIOUR OF HIGH PRICES.

Just 36.26% (1-0.63730288) of the courses are above 50. However, the range of prices of the courses with a higher cost, goes from 50$ until 200$. That wide range show that most of the courses in this section of prices, are not trying to lower the price. Their strategy is the differentiation.

I assume that in this portion of prices we will find the courses that are more specialized, complete, or different from the rest. Their producers know that these courses are offering some added value that the low-price courses are not offering and they charge a higher price because that.

Even if in this range of prices, the cost of the course is not that important, we can also notice how they are applying the "psychological pricing" techniques.

7.74% (0.0413268080 + 0.0361609570) of the courses have a price of 95$ or 100, and11.47 or 200$.

As we have seen before, the prices of 100$ and 200$ may be searching for a round and simple-to-process number. At the same time the 95$ and 195$ and even the 190$ prices, seem to be applying the first variant of the "psychological prices" that we have explained.

```
prop.table(table(data$price))
```

```
##
##           0          20          25          30          35          40
## 0.0842849375 0.2256661229 0.0418705818 0.0451332246 0.0315388798 0.0589994562
##          45          50          55          60          65          70
## 0.0225666123 0.1272430669 0.0095160413 0.0203915171 0.0081566069 0.0070690593
##          75          80          85          90          95         100
## 0.0220228385 0.0078847200 0.0084284937 0.0051658510 0.0413268080 0.0361609570
##         105         110         115         120         125         130
## 0.0029907558 0.0021750952 0.0043501903 0.0100598151 0.0081566069 0.0008156607
##         135         140         145         150         155         160
## 0.0021750952 0.0027188690 0.0073409462 0.0239260468 0.0005437738 0.0010875476
##         165         170         175         180         185         190
## 0.0016313214 0.0013594345 0.0035345296 0.0046220772 0.0016313214 0.0027188690
##         195         200
## 0.0345296357 0.0802066340
```

Now we can proceed to examine the prices dividing its distribution by a column of factors. In this way we will be able to confirm if the assumptions that we made are correct.

We will start by understanding the relation between the subject of the courses and its price.

As we can see, the courses that cost more than 50$ are more focused on technical skills and less in artistic ones.

```
prop.table(table(data[data$price<=50,"subject"]))
```

```
##
##    Business Finance      Graphic Design Musical Instruments     Web Development
##           0.3255119           0.1787543           0.2231229           0.2726109
```

```
prop.table(table(data[data$price>50,"subject"]))
```

```
##
##    Business Finance      Graphic Design Musical Instruments     Web Development
##           0.3238381           0.1379310           0.1176912           0.4205397
```

In fact, there are many cases where, an IT course and an Arts one, even if they have similar duration, they charge different prices.

In the following example we can see the two longest courses on the platform, both with a duration of more than 76 hours. According to its titles, the two courses claim to be the "complete" course in their respective fields.

The IT course costs 200$ and the Graphic Design costs 50$. Even with this price difference, the number of subscribers is higher in the technological one.

```
data[data$content_duration>76,c("course_title","content_duration","price","subject","num_subscribers" )]
```

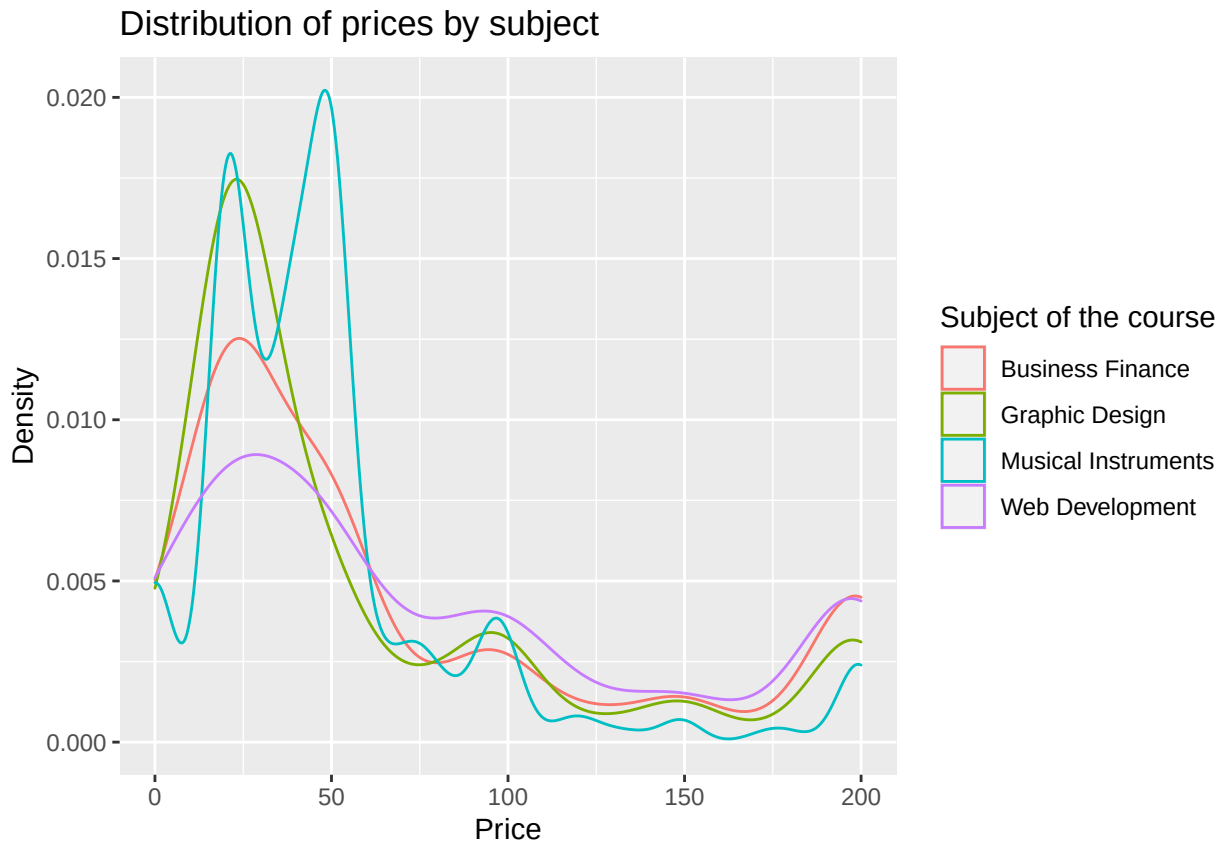
```
##                                                course_title content_duration
## 1659                     The Complete Figure Drawing Course HD             78.5
## 3142 The Complete Web Development Course -  Build 15 Projects             76.5
##      price         subject num_subscribers
## 1659    50  Graphic Design            1323
## 3142   200 Web Development            7501
```

We can find this same information dividing the distribution of the prices by the subject of the courses.

As we can see in the following plot, the artistic courses show a big agglomeration in the lower section of prices. In the rest of the range of prices, the density of artistic courses is, for the most part, lower than the IT and Business courses.

I would like to point out, that the "web Development" courses, show the lowest density among the cheaper rage of prices and the higher density when it comes to the more expensive ones.

```
ggplot()+
  geom_density(aes(data$price,color=data$subject))+
  labs(x = "Price",y="Density", colour = "Subject of the course",title="Distribution of prices by subje
```



Now we will focus on the level of the course to explain the evolution of the prices.

According to the following table, the courses that compete with price, show a higher percentage of "Beginner Level" content.

The courses with a higher price than 50$, have less percentage of Beginners courses and show more percentage of content focused on "Expert Level" and "All levels".

It is important to highlight that according to Udemy's Webpage , in the classification "All levels" we can find the courses that are not for a specific degree of proficiency, courses that mix basic concepts with more advanced ones.

```
prop.table(table(data[data$price<=50,"level"]))
```

```
##
##         All Levels     Beginner Level       Expert Level Intermediate Level
##        0.495307167        0.378839590        0.008105802        0.117747440
```

```
prop.table(table(data[data$price>50,"level"]))
```

```
##
##         All Levels     Beginner Level       Expert Level Intermediate Level
##         0.57571214         0.28635682         0.02923538         0.10869565
```

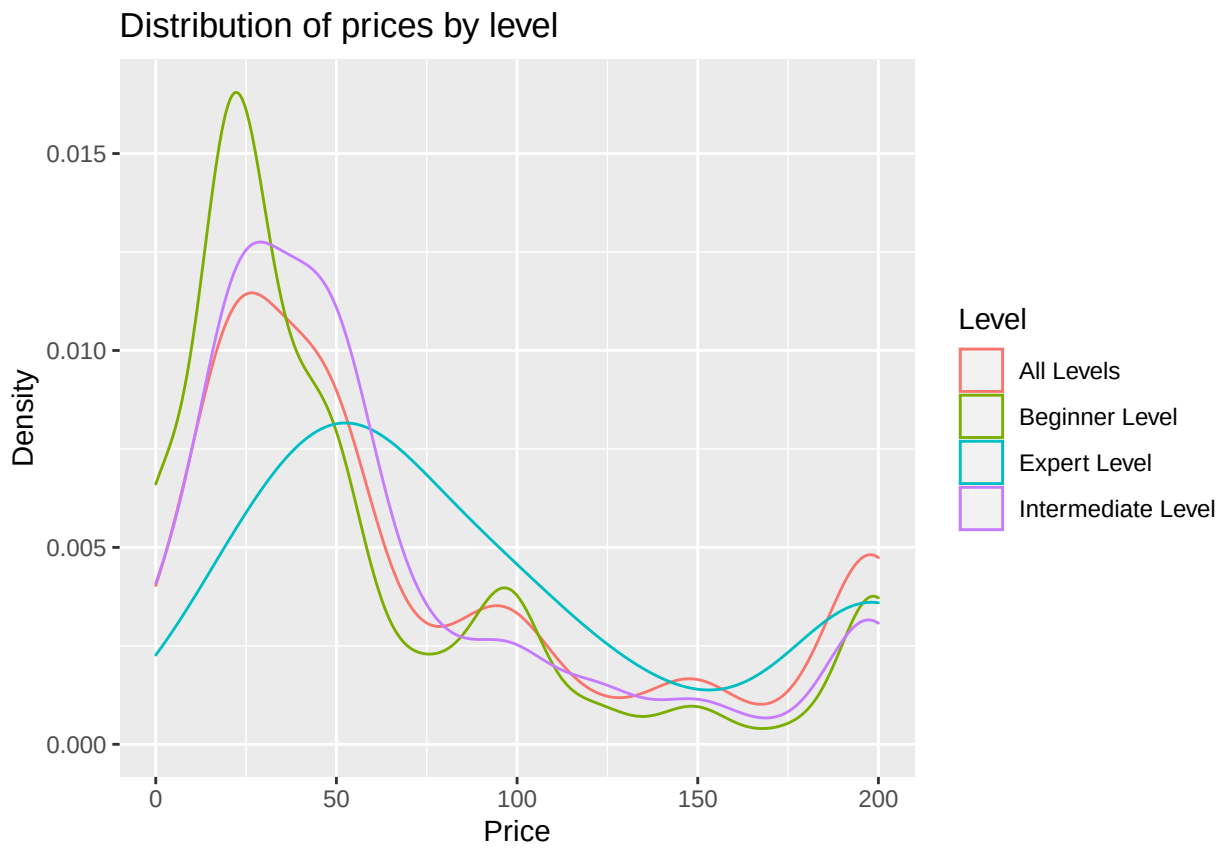
If we plot the prices dividing them by the level, we can see that "All Levels", "Beginner Level" and "Intermediate Level", show a similar behavior. They reach their maximum density between the $20-30$ price, and then, the density of courses drops, experiencing fluctuations around $100, 150$ and 200\$.

Regarding the "Expert Level" courses, they reach their higher density near to 50\$ and then the density decreases until reaching its lower point at $150. From that point, it increases until a local maximum at around 200$.

As we said before, the courses that had a price of 50\$ charged this quantity because, even though they were competing with price, they were offering something that the cheaper courses were not able to.

With this graph we can see that, offering an "Expert level" course, may be the added value that we were talking about. I am sure there are other reasons, but the specialization of the course seems to be an important one.

```
ggplot()+
  geom_density(aes(data$price,color=data$level))+
  labs(x = "Price",y="Density", colour = "Level",title="Distribution of prices by level")
```



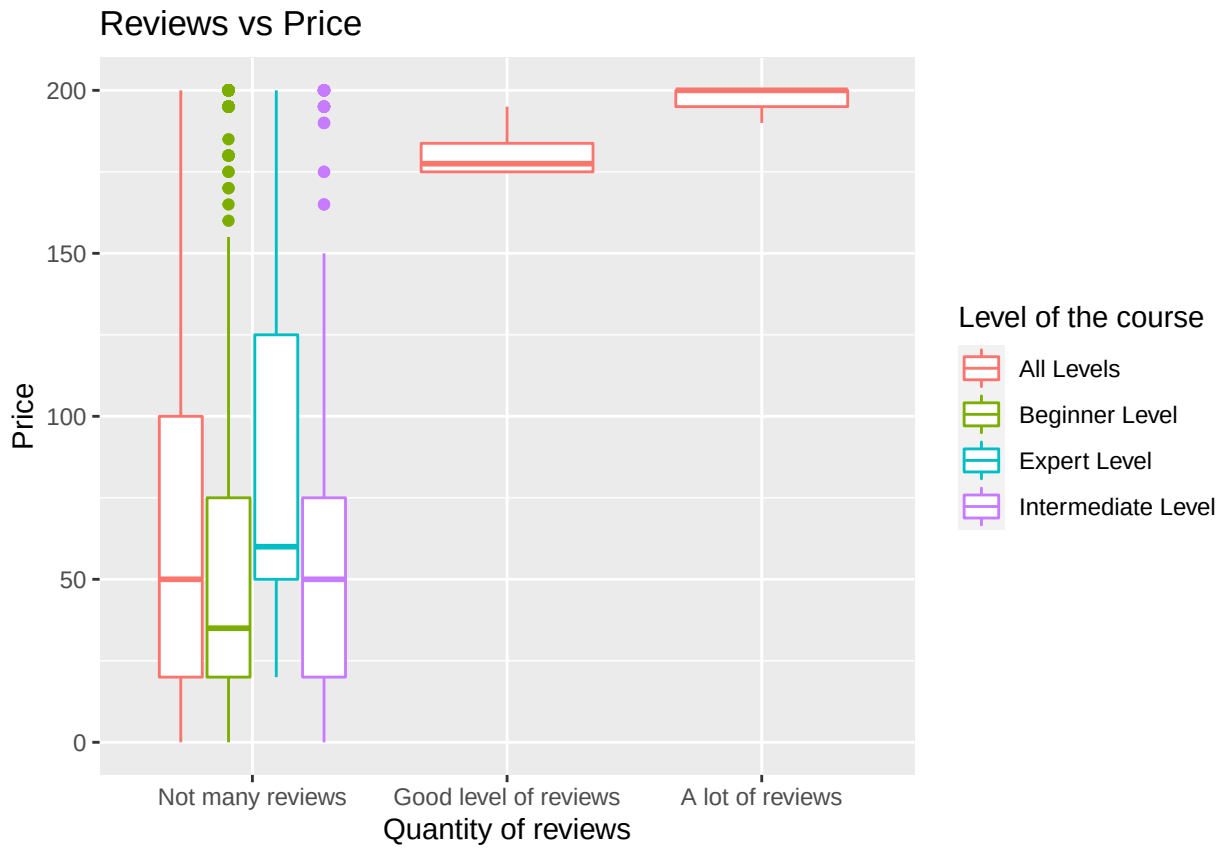
As we have already seen, an "Expert Level" course that is from the IT field, most of the times will be able to differentiate itself and to charge a high price. However, the level and the field of knowledge are not the only variables to take into consideration, there are other ways to differentiate the courses.

In the following plot we have divided the number of reviews that a course has in three categories, "Not many reviews", "Good level of reviews", "A lot of reviews".

As we can see, the courses that obtained a higher number of reviews are not classified as "Expert Level" but "All Levels". Even if they are not "Expert Level" courses, the high number of reviews helped them to differentiate themselves from the rest, allowing them to raise the price.

There are many other ways in which the courses may obtain a good level of differentiation, some examples are: -With a well-known expert teaching the course. -With a clear and simple explanation. -With a teacher that knows how to connect with the students...

```
ggplot()+
  geom_boxplot(aes(cut(data$num_reviews,breaks = 3,labels = c("Not many reviews","Good level of reviews
  labs(x = "Quantity of reviews",y="Price", colour = "Level of the course",title="Reviews vs Price")
```



One last thing that I would like to understand about the level of the courses is their different durations.

In the following table, we have divided the total range or courses length in 13 parts. In this way, we will be able to see which levels tend to have longer courses.

According to the data of this table, the length of the "Beginners" and "Expert" courses are shorter than the "All Levels" and "Intermediate Level" ones.

In fact, 90.23% (0.9023622047) of the Beginners courses and 93.10% (0.9310344828) of the Expert courses last less than 6 hours. I assume that this is the case for two different reasons:

1-The courses that are classified for "Beginner" just show a certain number of basic concepts that does not require much time.

2-I assume that the courses that are made for experts, just focus on a specific knowledge. Since they do not need to teach the basics, the curses tend to be short.

```
prop.table(table(cut(data$content_duration,breaks = 13),data$level),2)
```

```
##
##                   All Levels Beginner Level Expert Level Intermediate Level
##   (-0.0785,6.04] 0.7941938828   0.9023622047 0.9310344828       0.8456057007
##   (6.04,12.1]    0.1207879730   0.0740157480 0.0344827586       0.1021377672
##   (12.1,18.1]    0.0425090721   0.0125984252 0.0344827586       0.0285035629
##   (18.1,24.2]    0.0165889062   0.0070866142 0.0000000000       0.0142517815
##   (24.2,30.2]    0.0088128564   0.0007874016 0.0000000000       0.0047505938
##   (30.2,36.2]    0.0057024365   0.0015748031 0.0000000000       0.0047505938
##   (36.2,42.3]    0.0020736133   0.0000000000 0.0000000000       0.0000000000
##   (42.3,48.3]    0.0036288232   0.0007874016 0.0000000000       0.0000000000
```

```
##   (48.3,54.3]    0.0010368066   0.0000000000 0.0000000000       0.0000000000
##   (54.3,60.4]    0.0010368066   0.0000000000 0.0000000000       0.0000000000
##   (60.4,66.4]    0.0010368066   0.0000000000 0.0000000000       0.0000000000
##   (66.4,72.5]    0.0020736133   0.0000000000 0.0000000000       0.0000000000
##   (72.5,78.6]    0.0005184033   0.0007874016 0.0000000000       0.0000000000
```

The "All Level" category is the one that has the longest courses.

As we said before, according to Udemy's Webpage, the content creators, classify the courses as "All levels" when they are not targeting any particular degree of proficiency, these courses mix basic concepts with more advanced ones.

We have also discovered than the courses that have the word "complete" in their title have a mean duration that is twice as long as the courses that don't include this word.

With these two facts in mind, I assume that most of the courses that include the word "complete" in their title are classified as "All levels". I presume that because the "complete" courses normally touch concepts of different difficulties and because according to the data, they tend to be longer.

Now what we will do is to create two tables. The first one will allow us to see how many courses that include the word "complete" are in every level. The second one will show the percentage that represent the "complete" courses in every level.

In the first table we can notice that, as we said, most of the courses that contain the word "complete" are part of the "All Levels" classification. However, the other levels also have some a considerable number of these courses.

With the second table, we can see that the "complete" courses represent a 7.56% of the totality of "All Levels" courses. Regarding the other levels, the percentage of these kind of courses represent between 2% and 3%.

```
table(data$level,grepl("complete",tolower(data$course_title)))
```

```
##
##                      FALSE TRUE
##   All Levels          1783  146
##   Beginner Level      1238   32
##   Expert Level          56    2
##   Intermediate Level   412    9
```

```
prop.table(table(data$level,grepl("complete",tolower(data$course_title))),1)
```

```
##
##                          FALSE       TRUE
##   All Levels         0.92431312 0.07568688
##   Beginner Level     0.97480315 0.02519685
##   Expert Level       0.96551724 0.03448276
##   Intermediate Level 0.97862233 0.02137767
```

Now we will focus on searching patterns on the dates in which the different courses were published. We will start, by selecting just the date and creating the features "year_published" and "month_published".

These two new variables will respectively contain, the year and the month in which the courses were uploaded.

```
data$published_timestamp=substr(data$published_timestamp, start = 1, stop = 10)
data$year_published=substr(data$published_timestamp, start = 1, stop = 4)
data$month_published=substr(data$published_timestamp, start = 6, stop = 7)
```

Now we will create a table with the number of courses published every year. In this way, we will be able to know more about the evolution of the platform.

As we can see in the following table, the number of courses published increases every year. However, the number of publications that occur during the 2017 is lower than the ones registered during 2016. Considering the evolution that the table shows, I assume that the year 2017 is not complete.

```
table(data$year_published)
```

```
##
## 2011 2012 2013 2014 2015 2016 2017
##    5   45  202  491 1014 1206  715
```

Using the month_published feature and a bar plot, we will now check if, as we said, the year 2017 is uncomplete.

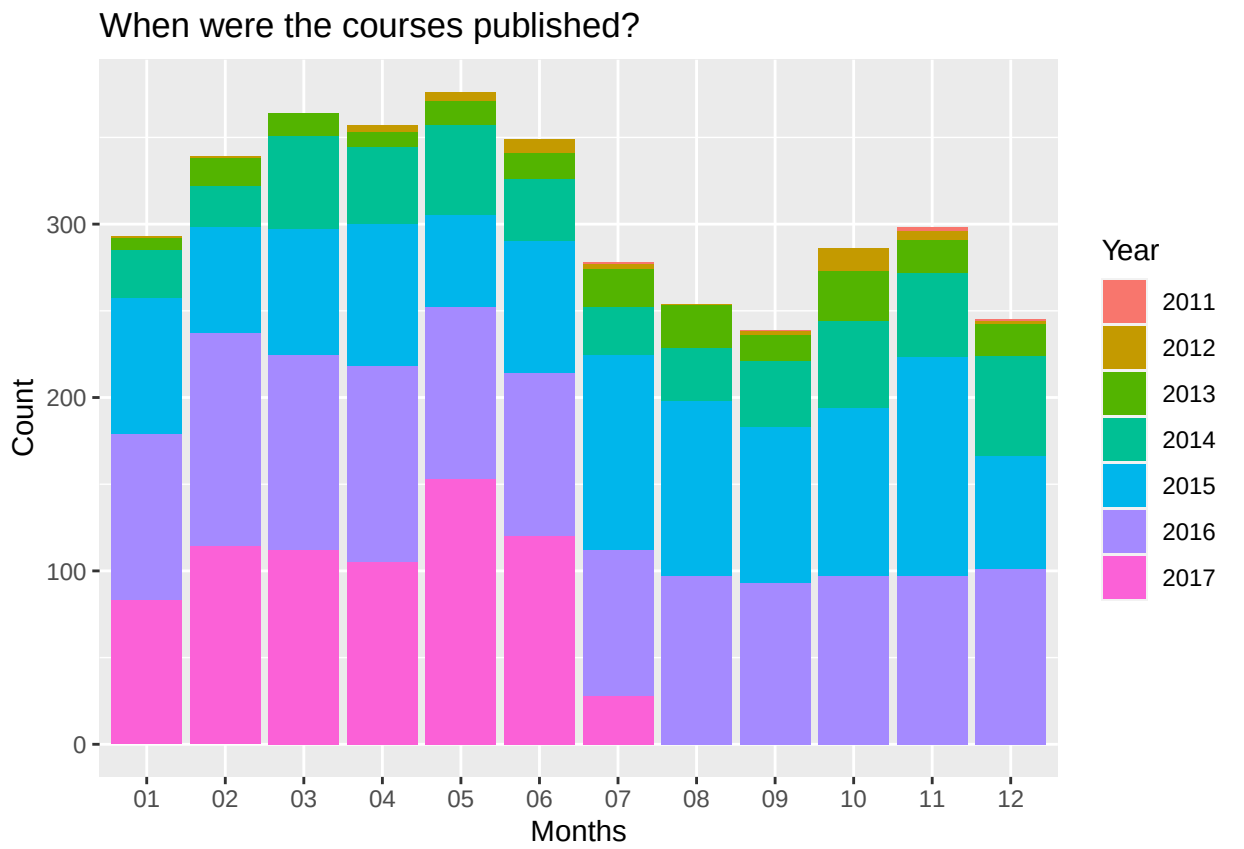
As we can see, the year 2017 has registers just until July and the year 2011 has its first register also during July.

Even if the plot has an interesting shape, we cannot take any conclusion out of it, mainly because of two reasons.

The first one, is that the data that we have, includes information about uncomplete years. That means that we cannot determine that more courses were published during March than during November, because March has courses from the year 2017 and November does not.

The second reason is that during the years that the data includes, the Udemy's platform was still developing. This is an important fact; the data shows that the number of courses published tend to grow as the year goes by. However, the most logic conclusion, is that this phenomenon happens just because the business was growing.

```
ggplot()+
  geom_bar(aes(data$month_published,fill=data$year_published))+
  labs(x = "Months",y="Count", fill = "Year",title="When were the courses published?")
```



Now we will search more information about the day of the week in which the different courses were published.

Before applying the next function, I should clarify that, it transforms a date into the day of the week that the date refers to.

The function has been programed in Spanish and thus, the days of the weeks will be in Spanish. Nonetheless I will offer the following translation to identify which are the days.

#lu is lunes means Monday. #ma is martes means Tuesday. #mi is miercoles means Wednesday. #ju is jueves means Thursday. #vi is viernes means Friday. #sá is sabado means Saturday. #do is domingo means Sunday.

According to the data, on weekends, the number of courses published decreases greatly. This may happen just because the content creators prefer not to work on weekends.

I would need more data to explain with more details why this pattern exists. With the data that we have, I have not found any other useful information.

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:dplyr':
##
##     intersect, setdiff, union
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
Day_con=data%>%
  mutate(wday=wday(published_timestamp,label=TRUE))
```

We will finish this analysis by learning how the most used words have changed during the years.

To do that, we will use the same steps that we used at the beginning of this analysis, but now, using a "for loop".

The resulting data frame is interesting, but it is important to extract the information carefully.

As we can see, the year 2011 and 2012 show very reduced word frequencies. That happens because the number of courses published in those years is extremely limited. It means that if we take conclusions out of these two years, the results will be probably biased.

The observations that we can obtain from this data frame are the following ones:

-Udemy has been offering courses related with arts and IT since its beginning.

-The most famous courses about arts that are regularly appearing in the top 10 most used words are "guitar" and "piano".

-The technological fields have been changing over time. During the first years, there were more courses related with codding like "javascript", "html5", "php"... During the last years, the IT courses are more related with web design.

-During the year 2014, the Business-related words grew in number. That year, the top 10 list, included words like "trading, "financial" and "forex". From that year on, the word "trading" has appeared in every top 10 most used words.

-We can see how the word "complete" has obtained more presence during the last 2 years, showing how the clients start to demand more from the site. The students do not want a course, they want the course from where they will be able to extract all the necessary knowledge.

-As we have seen during our analysis, every year the number of published courses keeps growing. We can also see this trend in the frequencies of the words in this data frame. However, the frequency increase from the year 2015 to the year 2016 is very limited. This happens because the content creators, due to the high level of competition that exists in the platform, are starting to offer courses that cover other fields of knowledge.

```
evolution_words=data.frame(index=c(1,2,3,4,5,6,7,8,9,10))
for (year in 1:length(unique(data$year_published))){


  selection=Day_con[Day_con$year_published==sprintf("201%s", year),]
  strings=as.character(selection$course_title)
  strings=tolower(paste(strings, collapse=' ' ))
  strings=str_split(strings,pattern = " ")
  strings=data.frame(as.matrix(table(strings)))
  strings= cbind(name = rownames(strings),strings)
  rownames(strings) = 1:nrow(strings)
  colnames(strings)= c(sprintf("Word 201%s", year), "Freq")
  strings=arrange(strings,desc(Freq))
  strings=strings[-c(which(strings$Word=="to"| strings[sprintf("Word 201%s", year)]=="a"| strings[sprin
  strings=data.frame(head(strings,10))

  evolution_words=cbind(evolution_words,strings)
}
evolution_words=evolution_words[,-1]
evolution_words
```

```
##    Word.2011 Freq  Word.2012 Freq.1 Word.2013 Freq.2 Word.2014 Freq.3 Word.2015
## 1  beginners    3     guitar      7    guitar     26     learn     73     learn
## 4     become    2      learn      6     learn     21       how     39   trading
## 5  developer    2        web      6      from     16     piano     32       how
## 6       html    2 javascript      5       web     16   trading     32       web
## 7        web    2   beginner      4   lessons     15   aprende     31      your
## 9         an    1  beginners      4     piano     15 financial     30 beginners
## 10 certified    1     design      4       how     14      your     28     build
## 11    course    1      html5      4     level     14     forex     27      from
## 12       css    1      mysql      4   options     14     tocar     26    course
## 13      from    1        php      4     blues     13       web     26    create
##    Freq.4 Word.2016 Freq.5 Word.2017 Freq.6
## 1     159     learn    167     learn     80
## 4      95 beginners     90    course     65
## 5      89   trading     86   trading     62
## 6      65      from     75       how     61
## 7      63    guitar     72 beginners     58
## 9      62     piano     69  complete     46
## 10     59       web     69 photoshop     42
## 11     58  complete     68      from     39
## 12     53    design     66    design     38
## 13     53       how     64       web     38
```

I will now write a conclusion with the main ideas that we have found during the analysis. I will try to summarize all the information in just a few lines. Please keep in mind that in the analysis you can find the explanation of everything I mention in the following text. I would also like to highlight that a great part of the analysis was destined to understand how the data needed to be interpreted.

We have found remarkably interesting information that helped us to reach the conclusions that we will present bellow. However, since this information is related with how the platform works or how the data is structured but not with the courses, I will not include it in the conclusion. If you are interested to know more about this "secondary" information, I would recommend you look at the full analysis.

#Conclusions of the analysis:

We have analyzed Udemy's website during the years 2011 and 2017, a young company that has a platform focused on the online learning. The business started its activity during the year 2010 and according to the data, the first lectures were offered during the year 2011. The organization has experienced an exponential growth, going from a small start up to a highly competitive market of courses focused on the Arts, IT and Business fields.

Nowadays, Udemy has a well-known brand, many content creators are interested in posting their courses in the platform. This fact has increased the level of competition that the publishers must face when they produce a course.

Many content creators are starting to apply different actions to obtain a strategic advantage to sell more than their competitors. As we have seen during the analysis, the main used strategies are the "cost leadership" and the "differentiation", furthermore in many cases they are also applying the "psychological prices" technique. The "differentiation" is a technique that allows the content creator to charge a higher price for the course because its lectures are different from the rest. However, those courses that were not able to differentiate themselves enough, are forced to compete with most of the existing courses by reducing its price.

In one hand, the courses that are focused on the "Expert Level", that obtained many reviews, that are offering IT-related lectures or a demanded skill, are likely to obtain a high level of differentiation. These are the 4 practices that we have been able to measure with the data, however there are other non-measurable ways to differentiate the courses. On the other hand, the courses related with arts, especially those aiming to the beginner or intermediate users, are likely to be forced to apply the "cost leadership" strategy with the price.

Thanks to the analysis that we have done of the popular courses on the platform, we have found that the most effective way to write a good tittle for a course is to include the specific skills that the lectures teach ("web", "javascript", "css"...). Furthermore, we have seen that many of the most used words like "learn", "how to", "for" or "beginners" do not have a clear effect on increasing the sells of the courses.

By examining the evolution of the most used words in the tittle of the courses every year, we have seen that Udemy is experiencing 2 big changes.

The fist one is that the high level of competition is requiring the content creators to start making courses from subjects that were not covered before in order to stay relevant.

The second one is that the number of titles that use the word "complete" has increased during the last years of the analysis. The mean duration of the courses that claim to be "complete" doubles the mean duration of the courses that do not include this word. Knowing that the longer courses are increasing in popularity, seems that the content creators have realized that the students are now searching for courses able to offer all the necessary knowledge that they will need.