Relation between Video Views and Device Usage with Question Correctness

# Data Understanding and Preparation

First of all I had to understand the data that I was given and decide how to go with it. I decided to use the csv files for Video Statistics, Enrollments and Question Responses. I decided to prepare my data as to have a dataframe with the people that got correct answer for the first quiz and how many people watched the the first video before doing the quiz. I prepared it and modeled it using a qplot using ggplot2. But then realized that I did no data wasn’t correct as the amount of people each year for each enrollment was different therefore I had to do calculations to find out not just the value for the amount of total views per video but the amount of total views per video over the amount of people for each enrollment separately. Then I did the model/graph again for (Number of Video Watched over Number of Questions correct relatively to the Enrollments) and my data was looking more understandable and correct. But then as I was going through the data in my csv that some part of the Questions Responses csv were empty. Therefore I had to go back to my munge file again and clear these fields ( do some more data wrangling)

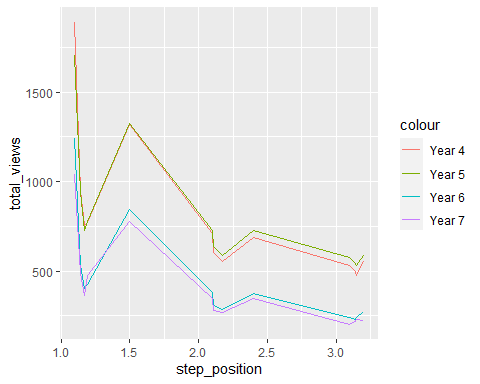
# Modelling

At first my modeling wasn’t done by plotgg or qplot but instead by just plot because i was with the idea the I should process the dataframe from the markdown cause i still wasn’t accustomed with Project Template and the way it worked. I then changed to plotgg and qplot and to use these two you had to use a particular dataframe, therefore I had to go back to Data Preparation and then to munge and make a dataframe that had the appropriate cells that I wanted and use them in my modeling. Later I found out that I could do Geometric Lines and I changed my first graph to that cause i thought it was more visually helpful.

# Evaluation

There were many times that I figured out that my understanding was not correct or up to par. The first thing that I decided to do with my data was to show the relation between what type of usage you had and from what country you where and I only used 1 csv file the Video Statistics. At first I made some data wrangle and came out with some models for my document but I came with the understanding that I didn’t have enought data and didn’t use my data correctly to be able to plot such a model and do assumptions and predictions on it. Therefore i decided to go with another one. Thus choosing the analysis im doing now.

### Total views by each year for each video



Here you can see how total views per video are watched by how many people in each year. As we can see there is a very uniformly way that students decide to watch a video or not, but lets see if what will it be when we do some data wrangling.

I decided to see if there was a relation between people watching the videos and being able to answer the questions correctly. Therefore i took the data from the enrollemnts.csv to check for each year the amount of people that were enrolled. I decided to check for the last 4 years.

## How many people are there in each enrollment by year

### Year 7

## [1] 2342

### Year 6

## [1] 3175

### Year 5

## [1] 3544

### Year 4

## [1] 3992

## How many of those got the quiz correct

I found the amount of people that got the first question correct by each year. I eliminated any any data that wasn’t correctly inputted (NA). ### Year 7

## [1] 570

### Year 6

## [1] 619

### Year 5

## [1] 1040

### Year 4

## [1] 1034

## The average amount of people that watched this videos:

* Welcome to the course
* Why would anyone want your data?
* Preserving privacy in cloud storage: privacy by design
* Staying safe online: personal perspectives
* Privacy online and offline

I decided to measure the views of total views for the 5 videos that were uploaded and you were supposed to see before answering the quiz for each year. I then found the average total views for all the 5 videos combined.

### Year 7

## [1] 629

### Year 6

## [1] 689.4

### Year 5

## [1] 1095.8

### Year 4

## [1] 1137.4

I then decided to divide the mean videos over the enrollment number so all the numbers will be proportionally equal for each year

### Year 7

## [1] 0.2685739

### Year 6

## [1] 0.2171339

### Year 5

## [1] 0.3091986

### Year 4

## [1] 0.2849198

Now the same for the correctness of the Quiz

### Year 7

## [1] 0.2433817

### Year 6

## [1] 0.1949606

### Year 5

## [1] 0.2934537

### Year 4

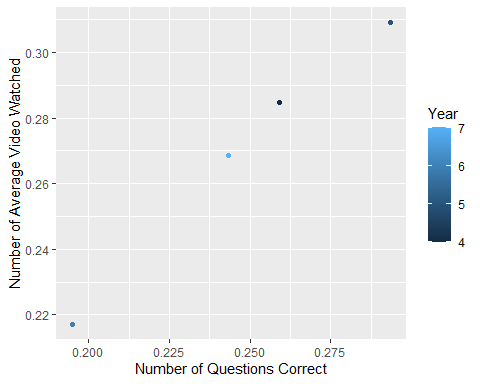
## [1] 0.259018

# Summary

From this data we can observe that most people that watched the people proportionally wise was at the 5th year with the 6th year coming last by a lot. As we can see from the data that i have gathered we can see a very similar result to the previous exercise that I did. This time again i find the Year 5 to be the highest year with the most student that got the result correct and the Year 6 to have the lowest correct results. Furthermore the Year 4 and 7 were again very close in both aspects ( question/size and meanvideo/size) but also both of them are in the same order as in the previous having year 4 coming second and year 7 coming last.

# Plot

I will now plot a graph for Number of Average Video Watched over Number of Questions Asked relative to the number of enrollments to show you the data more clearly.



## Plot Summary

From this my summary is that there is indeed a relation between the time to give to watch a video about the subject and answering the quiz question correctly even if the videos are not all relevant to the question being asked. It seems that people that will do good at the quiz are people that watch most if not all of the videos.

## What type of device did they use and did it have any relation with the outcome?

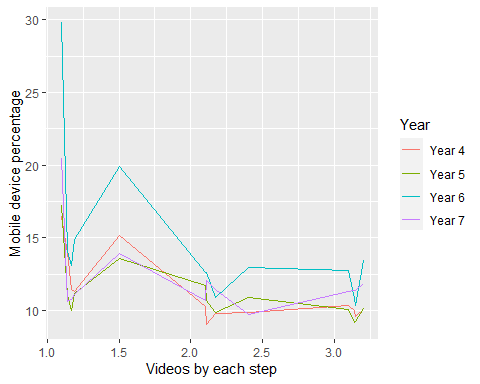
# Summary

I was given data for the video stats from the module cyber-security. I extracted the statistics for device usage.

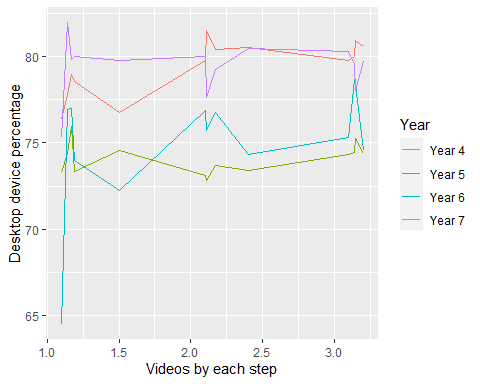
# Assumptions

My Assumption is if there is a relation to what device you are using with the likelihood that you are going to watch the videos.

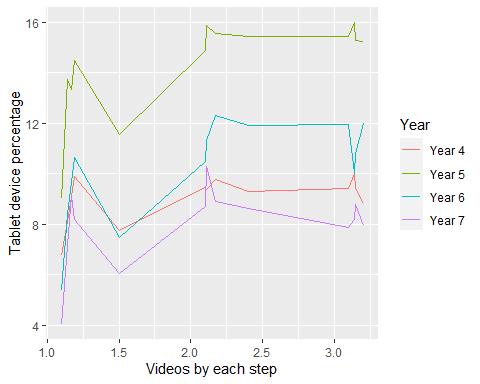
### Mobile device percentage for all videos by each year



### Desktop device percentage for all videos by each year



### Tablet device percentage for all videos by each year



We can observe that the majority of the people watch the videos from their Desktop except for the 1st step in the video which is the introduction of the course therefroe it doesn’t have any important parts they need to take into account. Year 6 was the highest for mobile and the Year 5 was the highest for the tablet. Therefore lets compare how they students did between 2.2 to 3 for tablet and mobile for all years. cause the data is more constant around that region. Therefore the 3.11 quiz i believe is the most optimal one to choose with 3 multiple choices cause we will see the where each student watch each video up to the point of the quiz.

We find out through data wrangling that the number for correct answers based on each year proportionally are the exact.

### Year 7

## [1] 0.3142613

### Year 6

## [1] 0.2513386

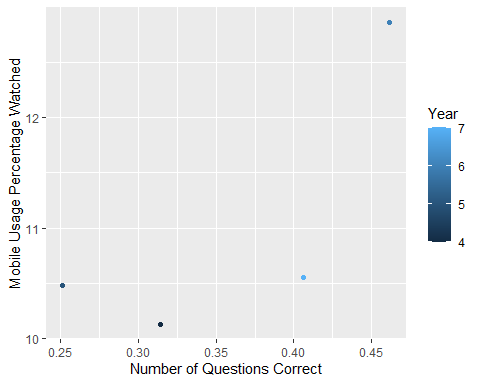
### Year 5

## [1] 0.4616253

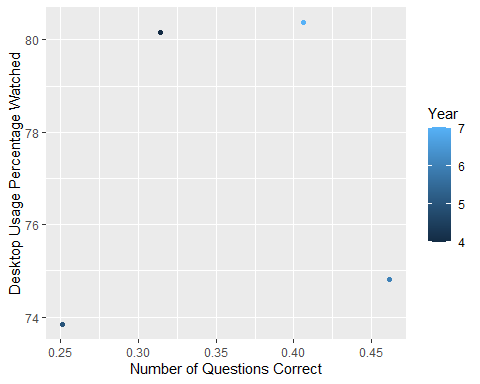
### Year 4

## [1] 0.4063126

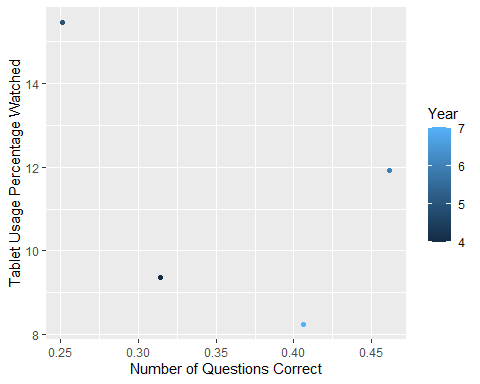
### Plot for if there is a relation between mobile usage and number of correct questions



### Plot for if there is a relation between desktop usage and number of correct questions



### Plot for if there is a relation between tablet usage and number of correct questions



## Summarise

There doesn’t seem to be any relation between what device do you use and how well your results are in the quiz. As it seems more clearly from the graph is that for none of the above devices that i mention have clear indication that as a percentage increase or decrease will have subsequent change in the results of the questions that are being asked in the quiz.