Starters and Finishers

Rachael Sanderson

6 November 2019

# **Who is completing the course?**

## *The Intended Analysis*

This analysis will reflect on the decline of participants within the course, reflecting when individuals are removing themselves from the online course. With this analysis, a further understanding will be gained about which elements of teaching attract the most participents, and so which elements need to be improved to maintain student participation and so overall cohort attainment.

## *The Applied Datasets*

The dataset used was the recorded “Step Activity” for each run of the course. To allow consistany of the temporally based analysis, this was undertaken for runs 3-7 of the course, as these have identical step progressions through the course and so allow a more direct comparison.

## *Data Preparation*

The following function was applied to produce a dataset for analysis;

QuantitiesStarting.function = function(x){ #x = file selected to analyse  
 StepActivity = x  
 Steps=c((StepActivity$week\_number)+((StepActivity$step\_number)/100)) #acknowledging data complication (3.1 matches 3.10)  
 StepActivity = cbind(StepActivity, Steps) #editing the data with the new column  
 S= unique(StepActivity$Steps, incomparables = FALSE) #list the unique steps  
 L =length(S) #How many steps are there?  
   
#making the vector of quantities for each step  
   
 total\_values = vector() #making the vector  
 for(i in 1:L){ #for loop: L = how many steps will be assessed/length of vector  
 Step= StepActivity %>% filter(Steps==S[i]) #filter: group the rows for a single step  
 Quant = nrow(Step) #how many rows were collected for that step?  
 total\_values[i]= Quant} #compile a vector of how many people started each step  
  
#making the dataframe to be able to chart: Step against how many participants.  
DFActivity = data.frame(Step = S, Total = total\_values)  
   
 return(DFActivity)} #return: the final dataframe for further analysis

It was necessary to adapt the dataset to accomodate the steps 1.1 compared to 1.10 (line 3). The function then counts how many people started each step, to produce a data frame that can then be evaluated.

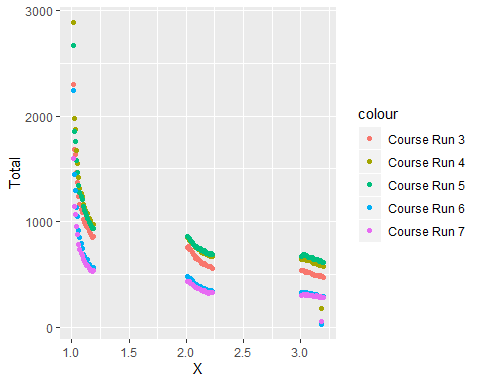
## *Analysis*

### *Part One: How many are starting at each step?*

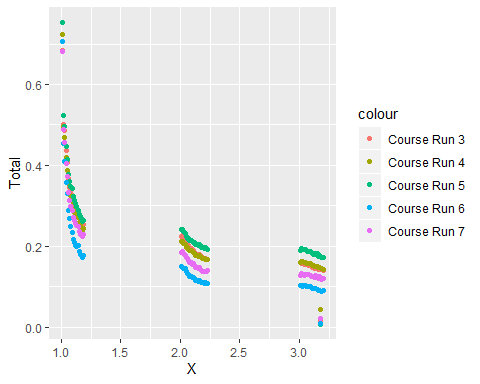
The first analysis was to record how many people are recorded as starting at each step, to reflect on how this declines. This was considered for each course, to consider if the trends were comparable for each run.

This can be graphed as below:

Plot = ggplot (data =CourseS3,aes (x = X, y = Total))  
CourseS3Graph =Plot +geom\_point (aes (x= Step, y = CourseS3$Total, colour = "Course Run 3"))  
CourseS4Graph = CourseS3Graph + geom\_point (aes(x = Step, y = CourseS4$Total, colour = "Course Run 4"))  
CourseS5Graph = CourseS4Graph +geom\_point (aes (x= Step, y = CourseS5$Total, colour = "Course Run 5"))  
CourseS6Graph = CourseS5Graph + geom\_point (aes(x = Step, y = CourseS6$Total, colour = "Course Run 6"))  
CourseS7Graph = CourseS6Graph + geom\_point (aes(x = Step, y = CourseS7$Total, colour = "Course Run 7"))  
  
CourseS7Graph

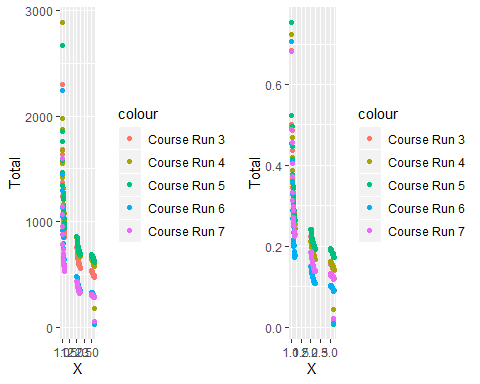


With an awareness that the number of enrolments varies each year, there was a realisation that this data needs to be standardised to acknowledge this. Therefore, the data was represented as a proportion. Within this process, the code as above was repeated, with the y axis instead being the Y value / number of course starters.



When comparing these graphs, the trends show limited differences. The trends show a gradual decline of starters at each step. DISTRIBUTION OF DECLINE? HYPOTHESIS TEST?

grid.arrange(CourseS7Graph, CourseSP7Graph, nrow = 1)



### *Part Two: How many are finishing at each step?*

When reflecting on analysis that focused on the starters, it was realised that there is an additional column of ‘finishers’. This may be a more worthwhile factor to evaluate, as the goal of the educational programme is to have low dropout rates throughout the course. Consequentially, a data frame of finishers was produced in the same way, with the addition of a single line of code at the start:

StepActivity = cyber.security.7.step.activity  
 Filter = StepActivity %>% filter(last\_completed\_at !="")  
 StepActivity=data.frame(Filter)

### *Part Three: Contrasting Starters and Finishers*

### *Discrepencies: What does it mean to “Finish”?*