EDA\_Project report 4

Stanly Wilson Palathingal

20 November 2019

Having seen the week 3 step 18 effect, it is good to compare them with some other files. So this report is making an attempt to see some features of those attended that particular ‘test’. They are compared with the enrollment data set.

#install.packages("ProjectTemplate")  
#install.packages("rmarkdown")  
#install.packages("dplyr")

library(ProjectTemplate)  
library(rmarkdown)  
library(dplyr)  
library(ggplot2)  
# create.project("EDA\_Project")

The required files are loaded.

step\_act\_7= read.csv("data/cyber-security-7\_step-activity.csv")  
step\_act\_6= read.csv("data/cyber-security-6\_step-activity.csv")  
step\_act\_5= read.csv("data/cyber-security-5\_step-activity.csv")  
step\_act\_4= read.csv("data/cyber-security-4\_step-activity.csv")  
step\_act\_3= read.csv("data/cyber-security-3\_step-activity.csv")  
step\_act\_2= read.csv("data/cyber-security-2\_step-activity.csv")  
step\_act\_1= read.csv("data/cyber-security-1\_step-activity.csv")

enrollment\_7= read.csv("data/cyber-security-7\_enrolments.csv")  
enrollment\_6= read.csv("data/cyber-security-6\_enrolments.csv")  
enrollment\_5= read.csv("data/cyber-security-5\_enrolments.csv")  
enrollment\_4= read.csv("data/cyber-security-4\_enrolments.csv")  
enrollment\_3= read.csv("data/cyber-security-3\_enrolments.csv")  
enrollment\_2= read.csv("data/cyber-security-2\_enrolments.csv")  
enrollment\_1= read.csv("data/cyber-security-1\_enrolments.csv")

Here the week 3 step 18 learners are separated with their id and they are inner joined with that f the enrollment file.

step\_7\_1=step\_act\_7 %>% select(learner\_id) %>% filter(step\_act\_7$week\_number==3 & step\_act\_7$step\_number==18)  
new\_7 = inner\_join(enrollment\_7,step\_7\_1,by = "learner\_id")

## Warning: Column `learner\_id` joining factors with different levels, coercing to  
## character vector

step\_6\_1=step\_act\_6 %>% select(learner\_id) %>% filter(step\_act\_6$week\_number==3 & step\_act\_6$step\_number==18)  
new\_6 = inner\_join(enrollment\_6,step\_6\_1,by = "learner\_id")

## Warning: Column `learner\_id` joining factors with different levels, coercing to  
## character vector

step\_5\_1=step\_act\_5 %>% select(learner\_id) %>% filter(step\_act\_5$week\_number==3 & step\_act\_5$step\_number==18)  
new\_5 = inner\_join(enrollment\_5,step\_5\_1,by = "learner\_id")

## Warning: Column `learner\_id` joining factors with different levels, coercing to  
## character vector

step\_4\_1=step\_act\_4 %>% select(learner\_id) %>% filter(step\_act\_4$week\_number==3 & step\_act\_4$step\_number==18)  
new\_4 = inner\_join(enrollment\_4,step\_4\_1,by = "learner\_id")

## Warning: Column `learner\_id` joining factors with different levels, coercing to  
## character vector

step\_3\_1=step\_act\_3 %>% select(learner\_id) %>% filter(step\_act\_3$week\_number==3 & step\_act\_3$step\_number==18)  
new\_3 = inner\_join(enrollment\_3,step\_3\_1,by = "learner\_id")

## Warning: Column `learner\_id` joining factors with different levels, coercing to  
## character vector

step\_2\_1=step\_act\_2 %>% select(learner\_id) %>% filter(step\_act\_2$week\_number==3 & step\_act\_2$step\_number==18)  
new\_2 = inner\_join(enrollment\_2,step\_2\_1,by = "learner\_id")

## Warning: Column `learner\_id` joining factors with different levels, coercing to  
## character vector

step\_1\_1=step\_act\_1 %>% select(learner\_id) %>% filter(step\_act\_1$week\_number==3 & step\_act\_1$step\_number==18)  
new\_1 = inner\_join(enrollment\_1,step\_1\_1,by = "learner\_id")

## Warning: Column `learner\_id` joining factors with different levels, coercing to  
## character vector

a= bind\_rows(new\_1,new\_2,new\_3,new\_4,new\_5,new\_6,new\_7)

Though there are many fields in enrollment file, most of them do not have valid values. So the graph is plotted with the role of the candidate. It is seen that majority are learners and only very few are organizational admins.

ggplot(new\_7, aes(x="", y=role, fill=role)) + geom\_bar(stat="identity", width=1) + coord\_polar("y", start=0)+xlab("")+ylab("")

