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_numbe
new_7 = inner_join(enrollment_7,step_7_1,by = "learner_id")
```

```
step_6_1=step_act_6 %>% select(learner_id) %>% filter(step_act_6$week_number==3 & step_act_6$step_numbe
new_6 = inner_join(enrollment_6,step_6_1,by = "learner_id")
```

```
step_5_1=step_act_5 %>% select(learner_id) %>% filter(step_act_5$week_number==3 & step_act_5$step_numbe
new_5 = inner_join(enrollment_5,step_5_1,by = "learner_id")
```

```
step_4_1=step_act_4 %>% select(learner_id) %>% filter(step_act_4$week_number==3 & step_act_4$step_numbe
new_4 = inner_join(enrollment_4,step_4_1,by = "learner_id")
```

```
step_3_1=step_act_3 %>% select(learner_id) %>% filter(step_act_3$week_number==3 & step_act_3$step_numbe
new_3 = inner_join(enrollment_3,step_3_1,by = "learner_id")
```

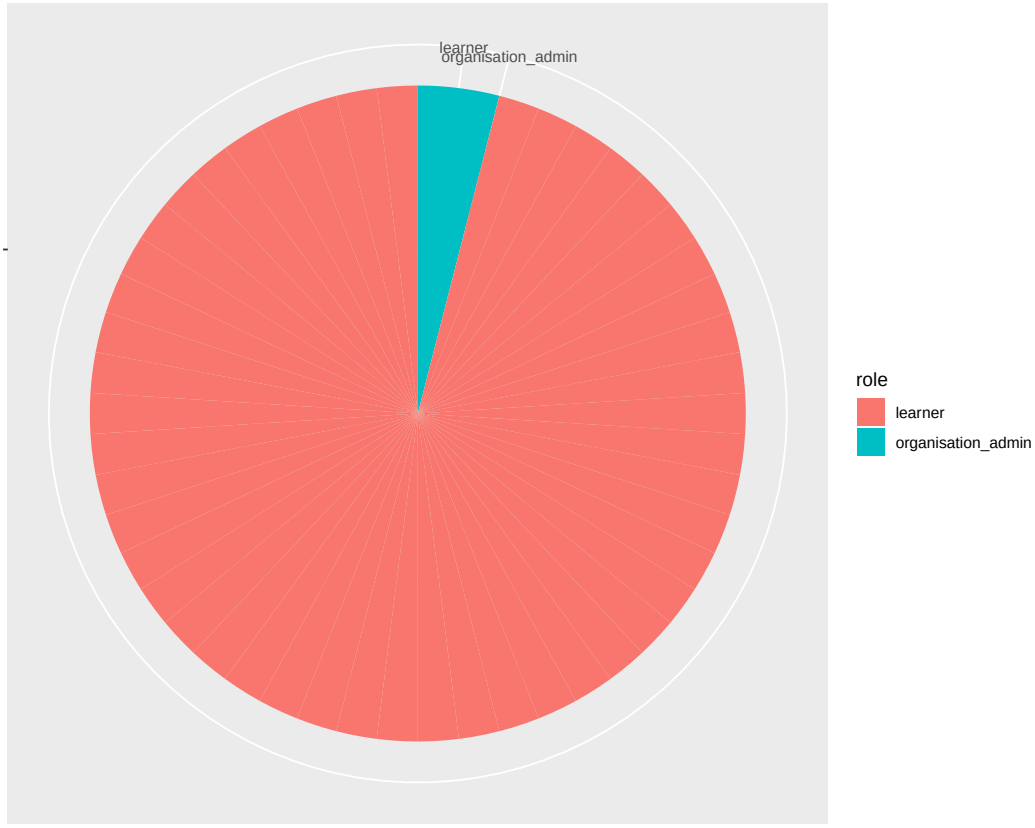
```
step_2_1=step_act_2 %>% select(learner_id) %>% filter(step_act_2$week_number==3 & step_act_2$step_numbe
new_2 = inner_join(enrollment_2,step_2_1,by = "learner_id")
```

```
step_1_1=step_act_1 %>% select(learner_id) %>% filter(step_act_1$week_number==3 & step_act_1$step_numbe
new_1 = inner_join(enrollment_1,step_1_1,by = "learner_id")
```

```
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 va lues. 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("")
```


learner
organisation_admin
role
learner
organisation_admin