

# EDA\_Project Report 3

*Stanly Wilson Palathingal*

*17 November 2019*

Having seen the weekly response, it is good to see in which part of the week the candidates leave. Every week there are various sections or step\_number indicating some modules. So the attempt is made to see how many candidates go through the steps or modules.

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

```
library(ProjectTemplate)
library(rmarkdown)
library(dplyr)
library(ggplot2)
# create.project("EDA_Project")
```

```
load.project()
```

To analyze, step activity files of all the seven years are loaded.

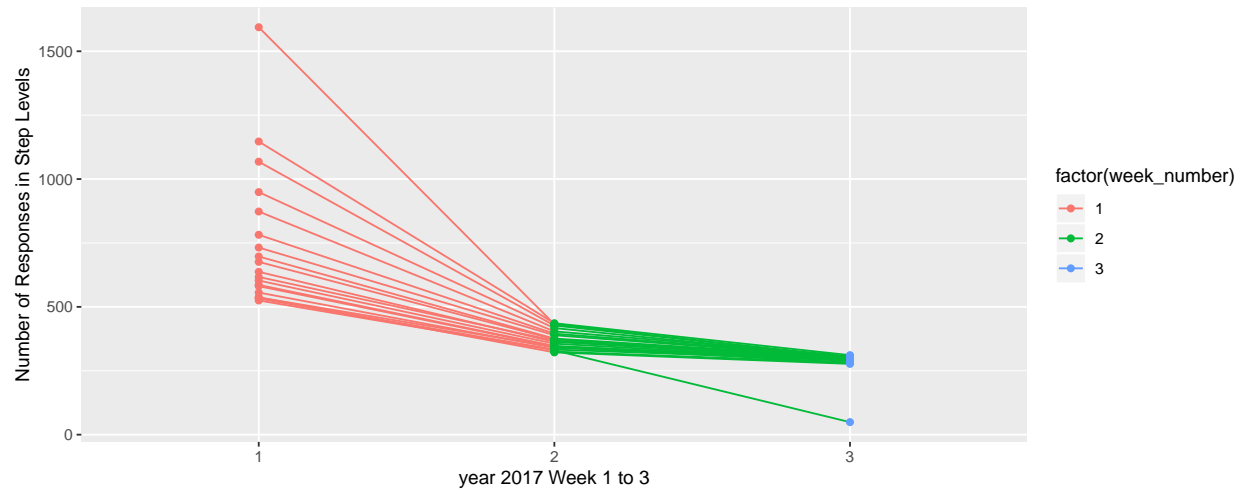
```
step_act_7= read.csv("data/cyber-security-7_step-activity.csv")
step_act_6= read.csv("data/cyber-security-6_step-activity.csv")
step_act_5= read.csv("data/cyber-security-5_step-activity.csv")
step_act_4= read.csv("data/cyber-security-4_step-activity.csv")
step_act_3= read.csv("data/cyber-security-3_step-activity.csv")
step_act_2= read.csv("data/cyber-security-2_step-activity.csv")
step_act_1= read.csv("data/cyber-security-1_step-activity.csv")
```

Group by command from the dplyr package is used to extract the data and they are plotted using a line graph. It can be seen that the one value falls to the lowest with a steep decline in week 3

```
step_7= step_act_7 %>% group_by(week_number, step_number) %>% count()
head(step_7)
```

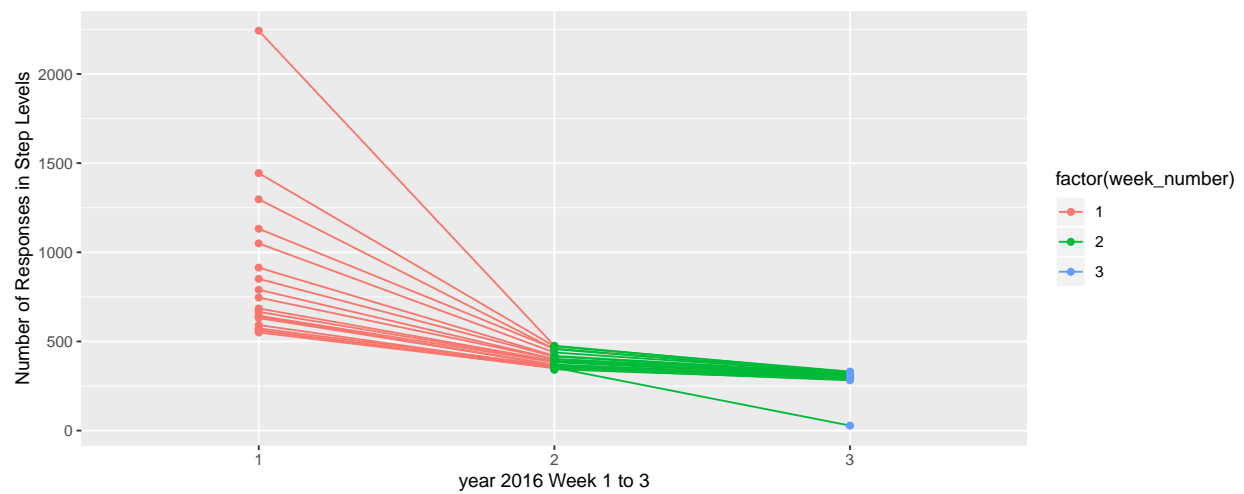
```
## # A tibble: 6 x 3
## # Groups:   week_number, step_number [6]
##   week_number step_number     n
##         <int>         <int> <int>
## 1           1             1 1594
## 2           1             2 1147
## 3           1             3 1068
## 4           1             4  949
## 5           1             5  873
## 6           1             6  782
```

```
ggplot(data=step_7,aes(x=factor(week_number),y=n,group =step_number)) +
  geom_line(aes(color =factor(week_number))) +
  geom_point(aes(color = factor(week_number))) +
  xlab("year 2017 Week 1 to 3") + ylab("Number of Responses in Step Levels")
```



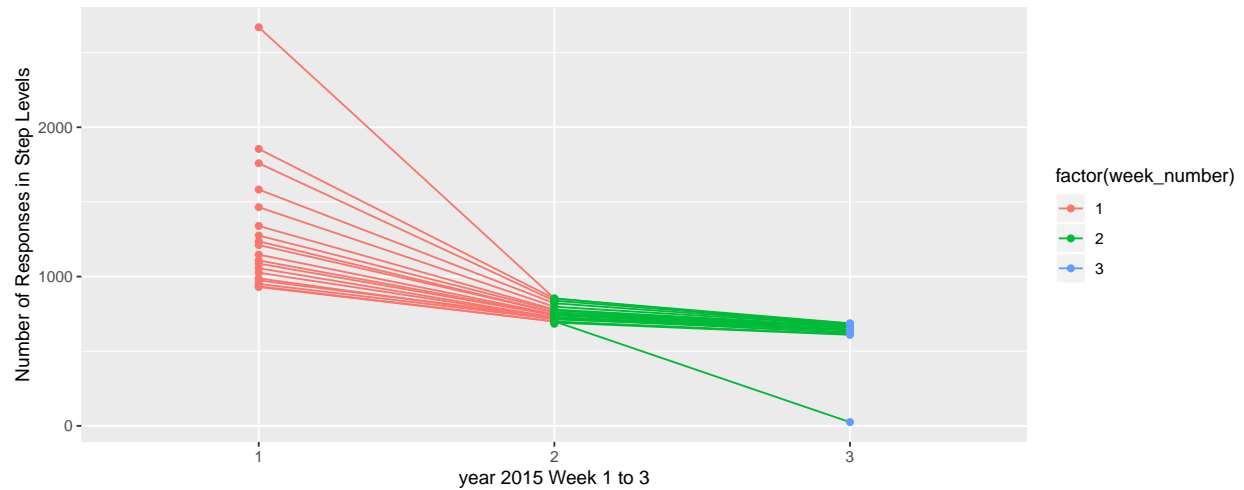
```
step_6= step_act_6 %>% group_by(week_number, step_number) %>% count()

ggplot(data=step_6,aes(x=factor(week_number),y=n,group =step_number)) +
  geom_line(aes(color =factor(week_number))) +
  geom_point(aes(color = factor(week_number))) +
  xlab("year 2016 Week 1 to 3") + ylab("Number of Responses in Step Levels")
```



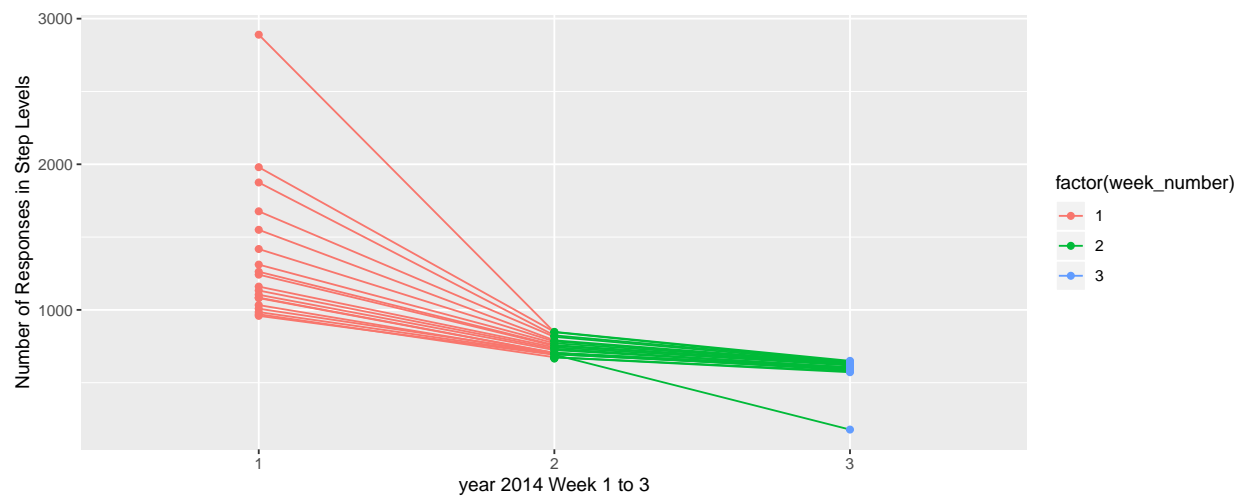
```
step_5= step_act_5 %>% group_by(week_number, step_number) %>% count()

ggplot(data=step_5,aes(x=factor(week_number),y=n,group =step_number)) +
  geom_line(aes(color =factor(week_number))) +
  geom_point(aes(color = factor(week_number))) +
  xlab("year 2015 Week 1 to 3") + ylab("Number of Responses in Step Levels")
```



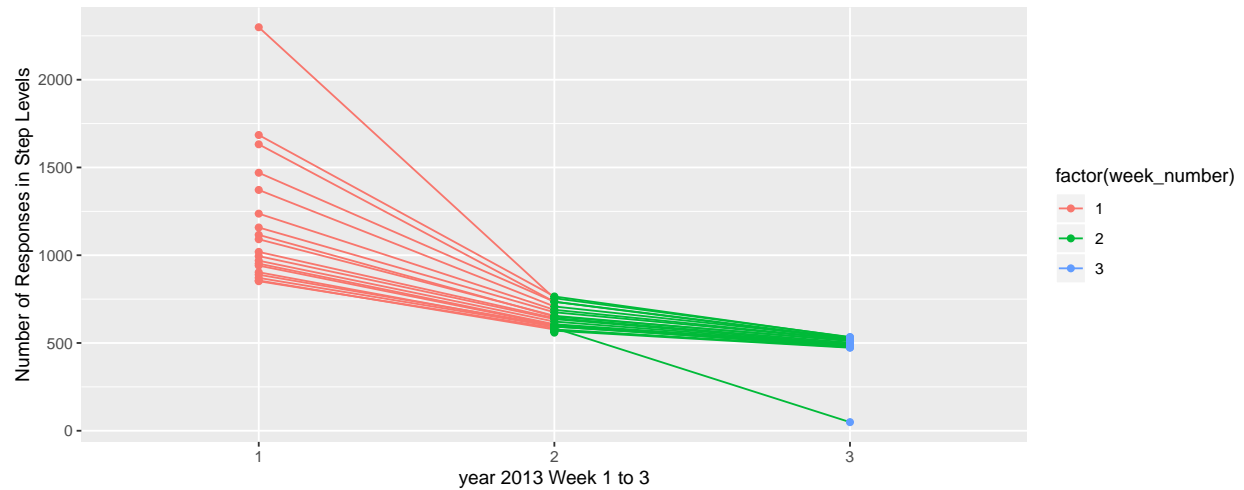
```
step_4= step_act_4 %>% group_by(week_number, step_number) %>% count()

ggplot(data=step_4,aes(x=factor(week_number),y=n,group =step_number)) +
  geom_line(aes(color =factor(week_number))) +
  geom_point(aes(color = factor(week_number))) +
  xlab("year 2014 Week 1 to 3") + ylab("Number of Responses in Step Levels")
```



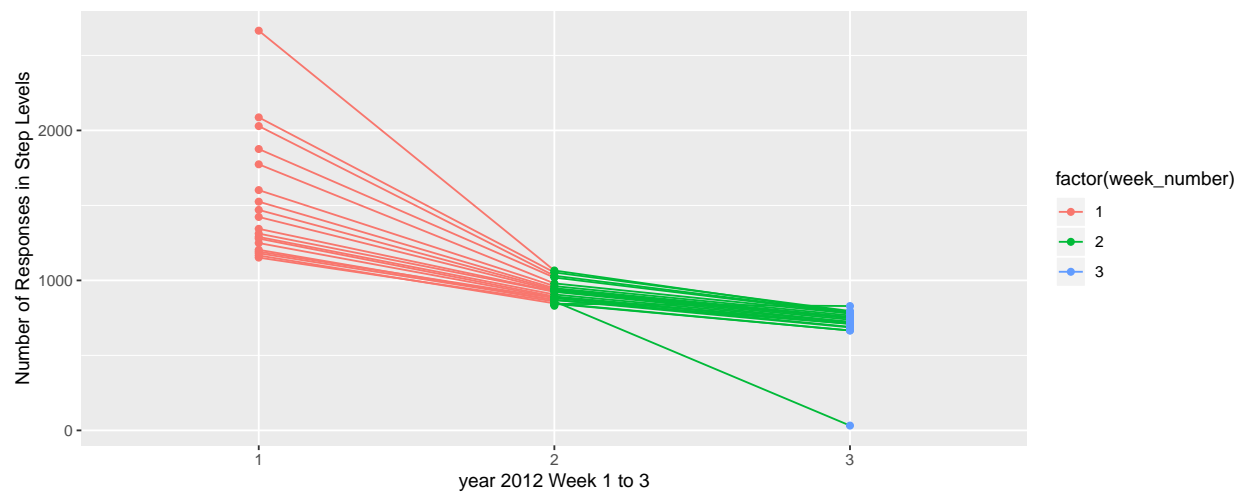
```
step_3= step_act_3 %>% group_by(week_number, step_number) %>% count()

ggplot(data=step_3,aes(x=factor(week_number),y=n,group =step_number)) +
  geom_line(aes(color =factor(week_number))) +
  geom_point(aes(color = factor(week_number))) +
  xlab("year 2013 Week 1 to 3") + ylab("Number of Responses in Step Levels")
```



```
step_2= step_act_2 %>% group_by(week_number, step_number) %>% count()

ggplot(data=step_2,aes(x=factor(week_number),y=n,group =step_number)) +
  geom_line(aes(color =factor(week_number))) +
  geom_point(aes(color = factor(week_number))) +
  xlab("year 2012 Week 1 to 3") + ylab("Number of Responses in Step Levels")
```



```
step_1= step_act_1 %>% group_by(week_number, step_number) %>% count()

ggplot(data=step_1,aes(x=factor(week_number),y=n,group =step_number)) +
  geom_line(aes(color =factor(week_number))) +
  geom_point(aes(color = factor(week_number))) +
  xlab("year 2011 Week 1 to 3") + ylab("Number of Responses in Step Levels")
```



From the line plot, except in year 2011 all the other years there is a sudden fall of value in week 3. The analysis made us understand that it is in week 3 step 18. Those values are taken separately and then plotted as below.

```
step_7_1=step_act_7 %>% select(learner_id) %>% filter(step_act_7$week_number==3 & step_act_7$step_number==18)
st7=data.frame("year"= 2017,"n"= length(step_7_1$learner_id))
```

```
step_6_1=step_act_6 %>% select(learner_id) %>% filter(step_act_6$week_number==3 & step_act_6$step_number==18)
st6=data.frame("year"= 2016,"n"= length(step_6_1$learner_id))
```

```
step_5_1=step_act_5 %>% select(learner_id) %>% filter(step_act_5$week_number==3 & step_act_5$step_number==18)
st5=data.frame("year"= 2015,"n"= length(step_5_1$learner_id))
```

```
step_4_1=step_act_4 %>% select(learner_id) %>% filter(step_act_4$week_number==3 & step_act_4$step_number==18)
st4=data.frame("year"= 2014,"n"= length(step_4_1$learner_id))
```

```
step_3_1=step_act_3 %>% select(learner_id) %>% filter(step_act_3$week_number==3 & step_act_3$step_number==18)
st3=data.frame("year"= 2013,"n"= length(step_3_1$learner_id))
```

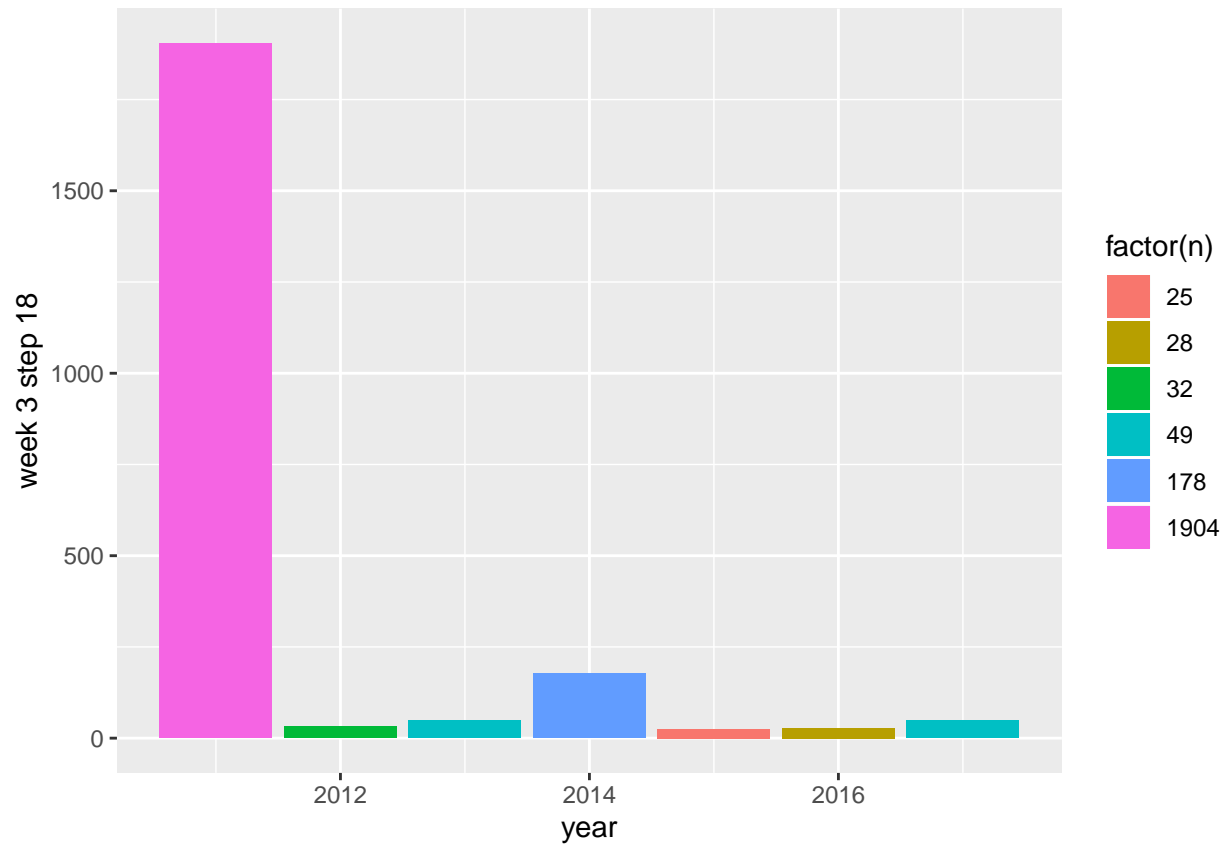
```
step_2_1=step_act_2 %>% select(learner_id) %>% filter(step_act_2$week_number==3 & step_act_2$step_number==18)
st2=data.frame("year"= 2012,"n"= length(step_2_1$learner_id))
```

```
step_1_1=step_act_1 %>% select(learner_id) %>% filter(step_act_1$week_number==3 & step_act_1$step_number==18)
st1=data.frame("year"= 2011,"n"= length(step_1_1$learner_id))
```

```
(step_w3_st18 = rbind(st1,st2,st3,st4,st5,st6,st7))
```

```
##   year    n
## 1 2011 1904
## 2 2012   32
## 3 2013   49
## 4 2014  178
## 5 2015   25
## 6 2016   28
## 7 2017   49
```

```
ggplot(data=step_w3_st18) +
  geom_bar(aes(x=year,y= n,fill=factor(n)), stat = "identity") +
  xlab("year")+ ylab("week 3 step 18")
```



Comparing these with the graph of candidates attending week 3, there is very huge difference. In week 3 year 2017 had 260 candidates and only 49 of them did step 18. By looking into the course content it will be evident that week 3 step 18 is the 'test' module and many did not go through it.