EDA\_Project2

Stanly Wilson Palathingal

12/11/2019

Install the required packages only for the first time.

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

library(ProjectTemplate)

## Warning: package 'ProjectTemplate' was built under R version 3.4.4

library(rmarkdown)  
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.4.4

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

load.project()

## Project name: EDA\_Project

## Loading project configuration

## Autoloading helper functions

## Running helper script: globals.R

## Running helper script: helpers.R

## Autoloading data

## Loading cached data set: cyber.security.1.enrolments

## Loading cached data set: cyber.security.1.leaving.survey.responses

## Loading cached data set: cyber.security.1.question.response

## Loading cached data set: cyber.security.1.step.activity

## Loading cached data set: cyber.security.1.weekly.sentiment.survey.responses

## Loading cached data set: cyber.security.2.archetype.survey.responses

## Loading cached data set: cyber.security.2.enrolments

## Loading cached data set: cyber.security.2.leaving.survey.responses

## Loading cached data set: cyber.security.2.question.response

## Loading cached data set: cyber.security.2.step.activity

## Loading cached data set: cyber.security.2.team.members

## Loading cached data set: cyber.security.2.weekly.sentiment.survey.responses

## Loading cached data set: cyber.security.3.archetype.survey.responses

## Loading cached data set: cyber.security.3.enrolments

## Loading cached data set: cyber.security.3.leaving.survey.responses

## Loading cached data set: cyber.security.3.question.response

## Loading cached data set: cyber.security.3.step.activity

## Loading cached data set: cyber.security.3.team.members

## Loading cached data set: cyber.security.3.video.stats

## Loading cached data set: cyber.security.3.weekly.sentiment.survey.responses

## Loading cached data set: cyber.security.4.archetype.survey.responses

## Loading cached data set: cyber.security.4.enrolments

## Loading cached data set: cyber.security.4.leaving.survey.responses

## Loading cached data set: cyber.security.4.question.response

## Loading cached data set: cyber.security.4.step.activity

## Loading cached data set: cyber.security.4.team.members

## Loading cached data set: cyber.security.4.video.stats

## Loading cached data set: cyber.security.4.weekly.sentiment.survey.responses

## Loading cached data set: cyber.security.5.archetype.survey.responses

## Loading cached data set: cyber.security.5.enrolments

## Loading cached data set: cyber.security.5.leaving.survey.responses

## Loading cached data set: cyber.security.5.question.response

## Loading cached data set: cyber.security.5.step.activity

## Loading cached data set: cyber.security.5.team.members

## Loading cached data set: cyber.security.5.video.stats

## Loading cached data set: cyber.security.5.weekly.sentiment.survey.responses

## Loading cached data set: cyber.security.6.archetype.survey.responses

## Loading cached data set: cyber.security.6.enrolments

## Loading cached data set: cyber.security.6.leaving.survey.responses

## Loading cached data set: cyber.security.6.question.response

## Loading cached data set: cyber.security.6.step.activity

## Loading cached data set: cyber.security.6.team.members

## Loading cached data set: cyber.security.6.video.stats

## Loading cached data set: cyber.security.6.weekly.sentiment.survey.responses

## Loading cached data set: cyber.security.7.archetype.survey.responses

## Loading cached data set: cyber.security.7.enrolments

## Loading cached data set: cyber.security.7.leaving.survey.responses

## Loading cached data set: cyber.security.7.question.response

## Loading cached data set: cyber.security.7.step.activity

## Loading cached data set: cyber.security.7.team.members

## Loading cached data set: cyber.security.7.video.stats

## Loading cached data set: cyber.security.7.weekly.sentiment.survey.responses

## Munging data

## Running preprocessing script: 01-A.R

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

e\_7\_id=data.frame("year"= 2017,"enrolled"= length(unique(enrollment\_7$learner\_id)))  
e\_6\_id=data.frame("year"= 2016,"enrolled"= length(unique(enrollment\_6$learner\_id)))  
e\_5\_id=data.frame("year"= 2015,"enrolled"= length(unique(enrollment\_5$learner\_id)))  
e\_4\_id=data.frame("year"= 2014,"enrolled"= length(unique(enrollment\_4$learner\_id)))  
e\_3\_id=data.frame("year"= 2013,"enrolled"= length(unique(enrollment\_3$learner\_id)))  
e\_2\_id=data.frame("year"= 2012,"enrolled"= length(unique(enrollment\_2$learner\_id)))  
e\_1\_id=data.frame("year"= 2011,"enrolled"= length(unique(enrollment\_1$learner\_id)))  
  
enrolled=rbind(e\_1\_id,e\_2\_id,e\_3\_id,e\_4\_id,e\_5\_id,e\_6\_id,e\_7\_id)

q7\_week1 = filter(q\_response\_7,q\_response\_7$week\_number==1)  
q7\_week2 = filter(q\_response\_7,q\_response\_7$week\_number==2)  
q7\_week3 = filter(q\_response\_7,q\_response\_7$week\_number==3)  
  
(q7=data.frame("year"= 2017,"week1"= length(unique(q7\_week1$learner\_id)), "week2"= length(unique(q7\_week2$learner\_id)), "week3"= length(unique(q7\_week3$learner\_id))))

## year week1 week2 week3  
## 1 2017 603 347 260

#q7=q\_response\_7 %>% select(learner\_id) %>% filter(q\_response\_7$week\_number==1)  
#length(unique(q7$learner\_id))

q6\_week1 = filter(q\_response\_6,q\_response\_6$week\_number==1)  
q6\_week2 = filter(q\_response\_6,q\_response\_6$week\_number==2)  
q6\_week3 = filter(q\_response\_6,q\_response\_6$week\_number==3)  
  
(q6=data.frame("year"= 2016,"week1"= length(unique(q6\_week1$learner\_id)), "week2"= length(unique(q6\_week2$learner\_id)), "week3"= length(unique(q6\_week3$learner\_id))))

## year week1 week2 week3  
## 1 2016 667 368 271

q5\_week1 = filter(q\_response\_5,q\_response\_5$week\_number==1)  
q5\_week2 = filter(q\_response\_5,q\_response\_5$week\_number==2)  
q5\_week3 = filter(q\_response\_5,q\_response\_5$week\_number==3)  
  
(q5=data.frame("year"= 2015,"week1"= length(unique(q5\_week1$learner\_id)), "week2"= length(unique(q5\_week2$learner\_id)), "week3"= length(unique(q5\_week3$learner\_id))))

## year week1 week2 week3  
## 1 2015 1088 726 571

q4\_week1 = filter(q\_response\_4,q\_response\_4$week\_number==1)  
q4\_week2 = filter(q\_response\_4,q\_response\_4$week\_number==2)  
q4\_week3 = filter(q\_response\_4,q\_response\_4$week\_number==3)  
  
(q4=data.frame("year"= 2014,"week1"= length(unique(q4\_week1$learner\_id)), "week2"= length(unique(q4\_week2$learner\_id)), "week3"= length(unique(q4\_week3$learner\_id))))

## year week1 week2 week3  
## 1 2014 1105 706 571

q3\_week1 = filter(q\_response\_3,q\_response\_3$week\_number==1)  
q3\_week2 = filter(q\_response\_3,q\_response\_3$week\_number==2)  
q3\_week3 = filter(q\_response\_3,q\_response\_3$week\_number==3)  
  
(q3=data.frame("year"= 2013,"week1"= length(unique(q3\_week1$learner\_id)), "week2"= length(unique(q3\_week2$learner\_id)), "week3"= length(unique(q3\_week3$learner\_id))))

## year week1 week2 week3  
## 1 2013 999 625 448

q2\_week1 = filter(q\_response\_2,q\_response\_2$week\_number==1)  
q2\_week2 = filter(q\_response\_2,q\_response\_2$week\_number==2)  
q2\_week3 = filter(q\_response\_2,q\_response\_2$week\_number==3)  
  
  
(q2=data.frame("year"= 2012,"week1"= length(unique(q2\_week1$learner\_id)), "week2"= length(unique(q2\_week2$learner\_id)), "week3"= length(unique(q2\_week3$learner\_id))))

## year week1 week2 week3  
## 1 2012 1342 889 668

q1\_week1 = filter(q\_response\_1,q\_response\_1$week\_number==1)  
q1\_week2 = filter(q\_response\_1,q\_response\_1$week\_number==2)  
q1\_week3 = filter(q\_response\_1,q\_response\_1$week\_number==3)  
  
(q1=data.frame("year"= 2011,"week1"= length(unique(q1\_week1$learner\_id)), "week2"= length(unique(q1\_week2$learner\_id)), "week3"= length(unique(q1\_week3$learner\_id))))

## year week1 week2 week3  
## 1 2011 3269 2022 1711

(q=rbind(q1,q2,q3,q4,q5,q6,q7))

## year week1 week2 week3  
## 1 2011 3269 2022 1711  
## 2 2012 1342 889 668  
## 3 2013 999 625 448  
## 4 2014 1105 706 571  
## 5 2015 1088 726 571  
## 6 2016 667 368 271  
## 7 2017 603 347 260

ggplot(data=enrolled) + geom\_bar(aes(x=year,y= enrolled,fill=factor(enrolled)), stat = "identity")

