EDA_Project Report 1

Stanly Wilson Palathingal

The purpose of this report is to analyze the archetypes. In order to do that the archetype files of all the years are loaded. Initial years, 2011 and 2012 do not have archetype files. The intention is to see the level of the candidates.

The project template is used and so it needs to be loaded. The set directory is set in the munge folder

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

The archetype files are loaded to the project

```r
archetype7= read.csv("data/cyber-security-7_archetype-survey-responses.csv")
archetype6= read.csv("data/cyber-security-6_archetype-survey-responses.csv")
archetype5= read.csv("data/cyber-security-5_archetype-survey-responses.csv")
archetype4= read.csv("data/cyber-security-4_archetype-survey-responses.csv")
archetype3= read.csv("data/cyber-security-3_archetype-survey-responses.csv")
```

There are many fields in the field and only some are needed for the analysis. So only what is needed are taken and created as a new data frame.

```r
a7=data.frame("year"= 2017,"learner_id"= archetype7$learner_id,"archetype"= archetype7$archetype)
a6=data.frame("year"= 2016,"learner_id"= archetype6$learner_id,"archetype"= archetype6$archetype)
a5=data.frame("year"= 2015,"learner_id"= archetype5$learner_id,"archetype"= archetype5$archetype)
a4=data.frame("year"= 2014,"learner_id"= archetype4$learner_id,"archetype"= archetype4$archetype)
a3=data.frame("year"= 2013,"learner_id"= archetype3$learner_id,"archetype"= archetype3$archetype)

t=rbind(a3,a4,a5,a6,a7)
head(t)
```

```
##   year                            learner_id  archetype
## 1 2013 0cc306ec-c90b-41be-85c1-0c3351d86241     Fixers
## 2 2013 9924483b-7932-499d-b7c5-aad759a2ede7     Fixers
## 3 2013 32e82b57-ccf7-468c-a86f-a4eb5a9bc62c Vitalisers
## 4 2013 c474fdb2-4b4d-4a68-9343-92851793197b  Explorers
## 5 2013 ffbaa395-4649-49fc-a95d-d8752d4e98aa  Advancers
## 6 2013 4ae58e18-fbda-4024-8c50-ca9632674592 Vitalisers
```

The bar chart shows the level of archetypes for five years. There is no archetype available for 2011 and 2012.

```r
ggplot(data=t) + geom_bar(aes(x=year,fill=archetype), position = position_dodge(preserve = 'single'))
```