Figures for the MOOC Pivot paper

José A. Ruipérez Valiente

12/19/2018

This code uses the CSV files tabular data from the *MOOC_Pivot.ipynb* and creates the **raw** figures that were used for the paper. Note out that figures were edited for final publication.

Imports

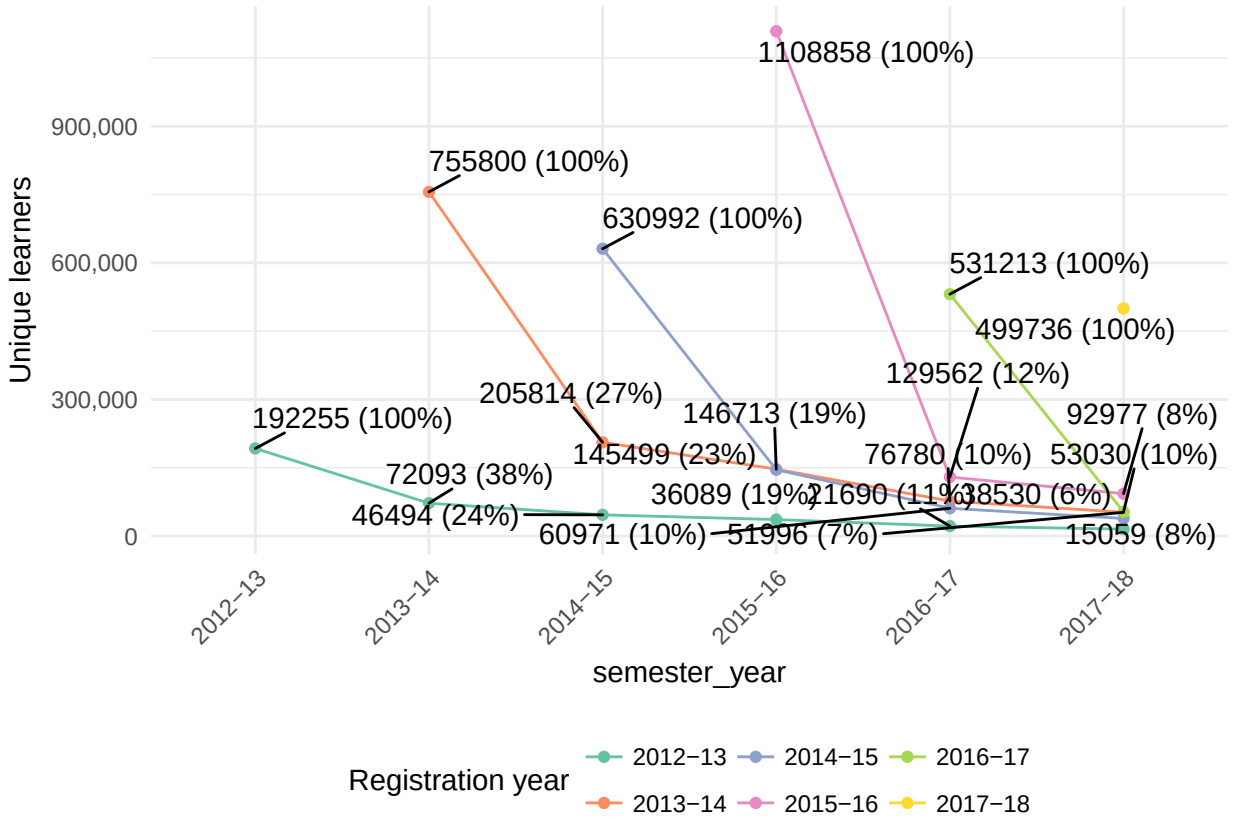
```
library(pander)
library(ggplot2)
library(tidyverse)
library(reshape2)
library(ggrepel)
library(gridExtra)
library(grid)
require(scales)
```

Churn rate

```
churn_rate_data <- read.csv2("Table_S1.csv", sep = ",", dec = ".")

churn_rate_data$registration_year = as.factor(churn_rate_data$registration_year)
churn_rate_data$year = factor(churn_rate_data$year,
                                        labels=c("2012-13", "2013-14", "2014-15", "2015-16", "2016-17",

ggplot(data = churn_rate_data) +
  geom_line(aes(group = registration_year, x = year, y=unique_learners, color=registration_year)) + geo
  geom_text_repel(aes(label=paste(unique_learners, ' (', round(percentage_retention),'%)', sep = ''), x=
  scale_color_brewer(palette="Set2", labels=c("2012-13", "2013-14", "2014-15", "2015-16", "2016-17", "2
  theme_minimal() + theme(legend.position="bottom", plot.title = element_text(hjust = 0.5), axis.text.x
  scale_y_continuous(labels = comma) +
  labs(x = 'semester_year', y = 'Unique learners', color = 'Registration year')
```



Enrolments and Certifications by HDI Category and Year

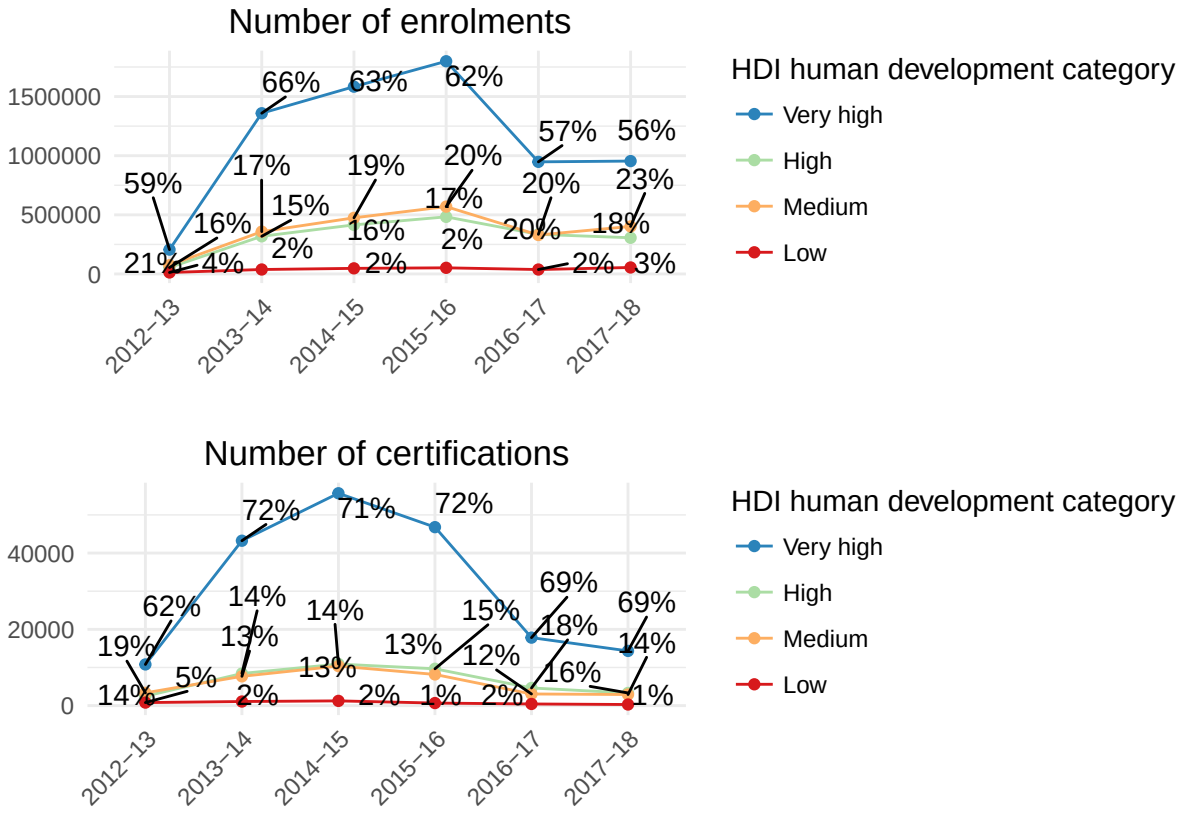
```
yearly_hdi_category_data <- read.csv2("Table_S2.csv", sep = ",", dec = ".")

yearly_hdi_category_data$human_development_category = factor(yearly_hdi_category_data$human_development_
                                                             levels = c("very_high_human_development", "

p_enrolments_hdi <- ggplot(data = yearly_hdi_category_data) +
  geom_line(aes(y=n_enrolments, x=year, color=human_development_category, group = human_development_cate
  geom_text_repel(aes(label=paste(round(p_enrolments_within_year),'%', sep = ''), x=year, y=n_enrolments
  theme(legend.position="right", plot.title = element_text(hjust = 0.5), axis.text.x = element_text(angl
  scale_color_manual(breaks = c("very_high_human_development", "high_human_development", "medium_human_d
                     values = c("#2b83ba", "#abdda4", "#fdae61", "#d7191c"),
                     labels = c("Very high", "High", "Medium", "Low")) +
  labs(x = '', y = '', fill='region', color = 'HDI human development category', title='Number of enrolme

p_certifications_hdi <- ggplot(data = yearly_hdi_category_data) +
  geom_line(aes(y=n_certified, x=year, color=human_development_category, group = human_development_categ
  geom_text_repel(aes(label=paste(round(p_certified_within_year),'%', sep = ''), x=year, y=n_certified)
  theme(legend.position="right", plot.title = element_text(hjust = 0.5), axis.text.x = element_text(angl
  scale_color_manual(breaks = c("very_high_human_development", "high_human_development", "medium_human_d
                     values = c("#2b83ba", "#abdda4", "#fdae61", "#d7191c"),
                     labels = c("Very high", "High", "Medium", "Low")) +
  labs(x = '', y = '', fill='region', color = 'HDI human development category', title='Number of certifi
```

```
grid.arrange(p_enrolments_hdi, p_certifications_hdi, ncol = 1)
```



Completion by Cohort

```
completion_by_cohort <- read.csv2(file = "Table_S3.csv", sep = ",", dec = ".")

completion_by_cohort$cohort <- factor(completion_by_cohort$cohort, levels = c("participants", "intend_c

ggplot(data = completion_by_cohort) +
  geom_bar(aes(y=p_completed, x=year, fill = cohort), position = "dodge", stat = "identity") + scale_fi
  theme(legend.position="bottom", plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(breaks = round(seq(0, 75, by = 5), 1)) +
  labs(x = '', y = 'Percentage completed', fill='cohort', color = 'cohort')
```


Percentage completed
55
50
45
40
35
30
25
20
15
10
5
0
2013−14
2014−15
2015−16
2016−17
2017−18
cohort
participants
intend_complete
verified