Open Univeristy Analysis

Shiwen Xu

October 23, 2017

Pre-analysis

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(scales)
library(tidyr)
library(rpart)
studentdrop=read.csv('student_reg_all.csv')
studenttestdrop=read.csv('student_as_drop.csv')
color=c('#a6cee3','#1f78b4','#b2df8a','#33a02c','#fb9a99','#e31a1c','#fdbf6f')
```

```
studentdrop$sr_date_unregistration=as.numeric(levels(studentdrop$sr_date_unregistration))[studentdrop$s:
```

```
## Warning: NAs introduced by coercion
```

```
levels(studentdrop$si_highest_education)
```

```
## [1] "A Level or Equivalent"      "HE Qualification"
## [3] "Lower Than A Level"         "No Formal quals"
## [5] "Post Graduate Qualification"
```

```
lev = levels(studentdrop$si_highest_education)[c(3,1,2,5,4)]
studentdrop$si_highest_education = factor(studentdrop$si_highest_education,
                                          levels = lev)
```

```
levels(studentdrop$si_highest_education)
```

```
## [1] "Lower Than A Level"         "A Level or Equivalent"
## [3] "HE Qualification"           "Post Graduate Qualification"
## [5] "No Formal quals"
```

Distribution
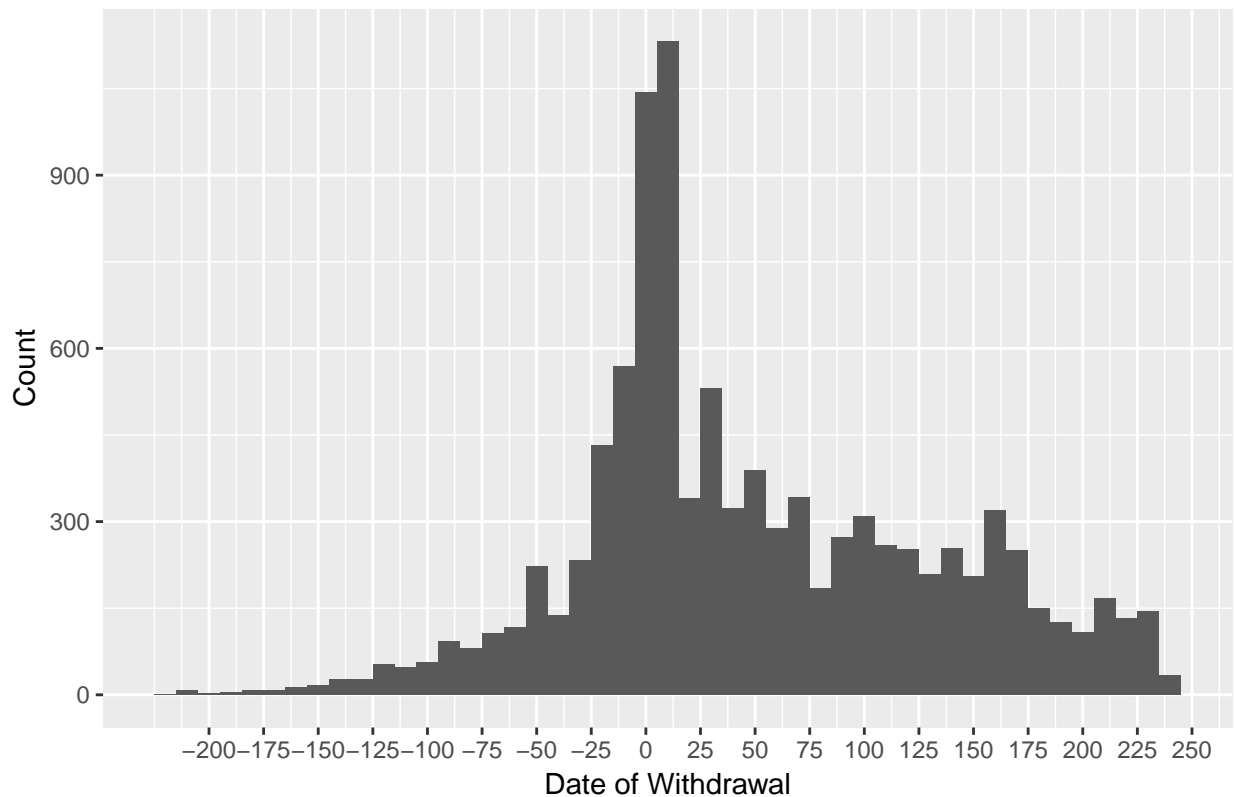
```
## Distribution of withdrawal numbers on days
studentdrop %>%
```

```
filter(si_final_result=='Withdrawn',sr_date_unregistration<250,sr_date_unregistration>-225)%>%
ggplot(aes(x=sr_date_unregistration))+
geom_histogram(binwidth = 10)+
scale_x_continuous(breaks=seq(-200,500,25))+
labs(x='Date of Withdrawal',y='Count',
    title='The Distribution of Withdrawal Days')+
scale_y_continuous(labels = comma)
```


The Distribution of Withdrawal Days

```
## Percentage of withdrawals two week before and after the semester start
studentdrop %>%
  filter(si_final_result=='Withdrawn',sr_date_unregistration<250,sr_date_unregistration>-225)%>%
  summarise(n())
```

```
##     n()
## 1 10025
```

```
studentdrop %>%
  filter(si_final_result=='Withdrawn',sr_date_unregistration<13,sr_date_unregistration>-13)%>%
  summarise(n())
```

```
##    n()
## 1 2694
```

```
2694/10025
```

```
## [1] 0.2687282
```

```
## Student registration date
studentdrop %>%
  filter(si_final_result!='Withdrawn',sr_date_registration<30)%>%
  ggplot(aes(x=sr_date_registration))+
  geom_histogram(binwidth = 10)+
  scale_x_continuous(breaks=seq(-400,30,25))+
  labs(x='Date of Registration',y='Count',
       title='The Distribution of Registration Dates')+
  scale_y_continuous(labels = comma)
```
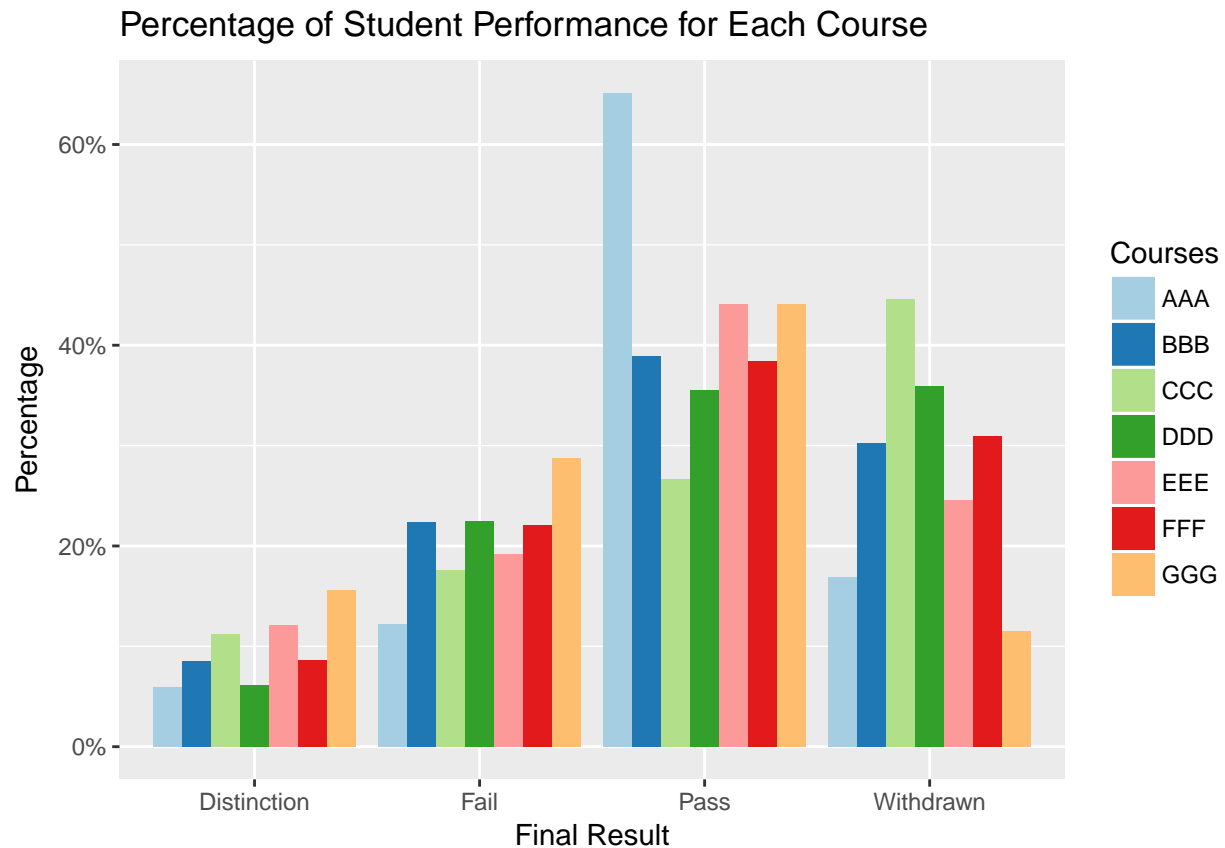
The Distribution of Registration Dates
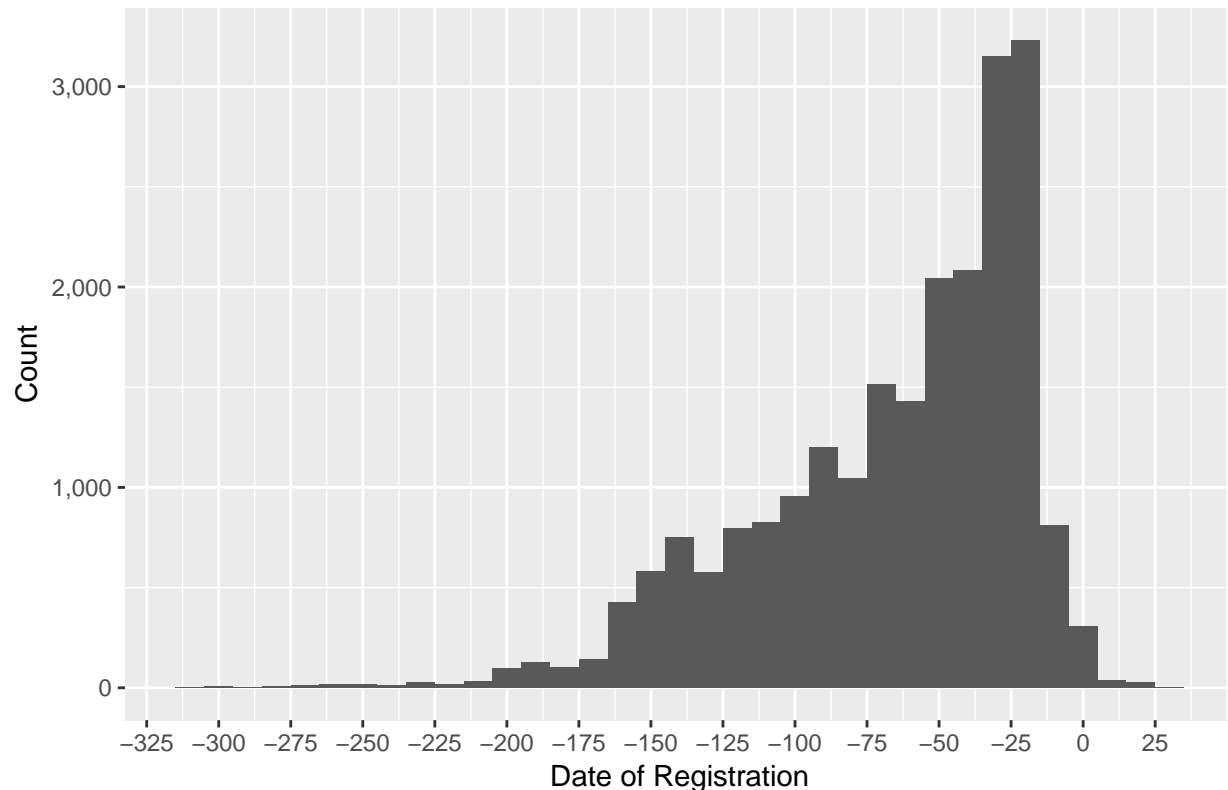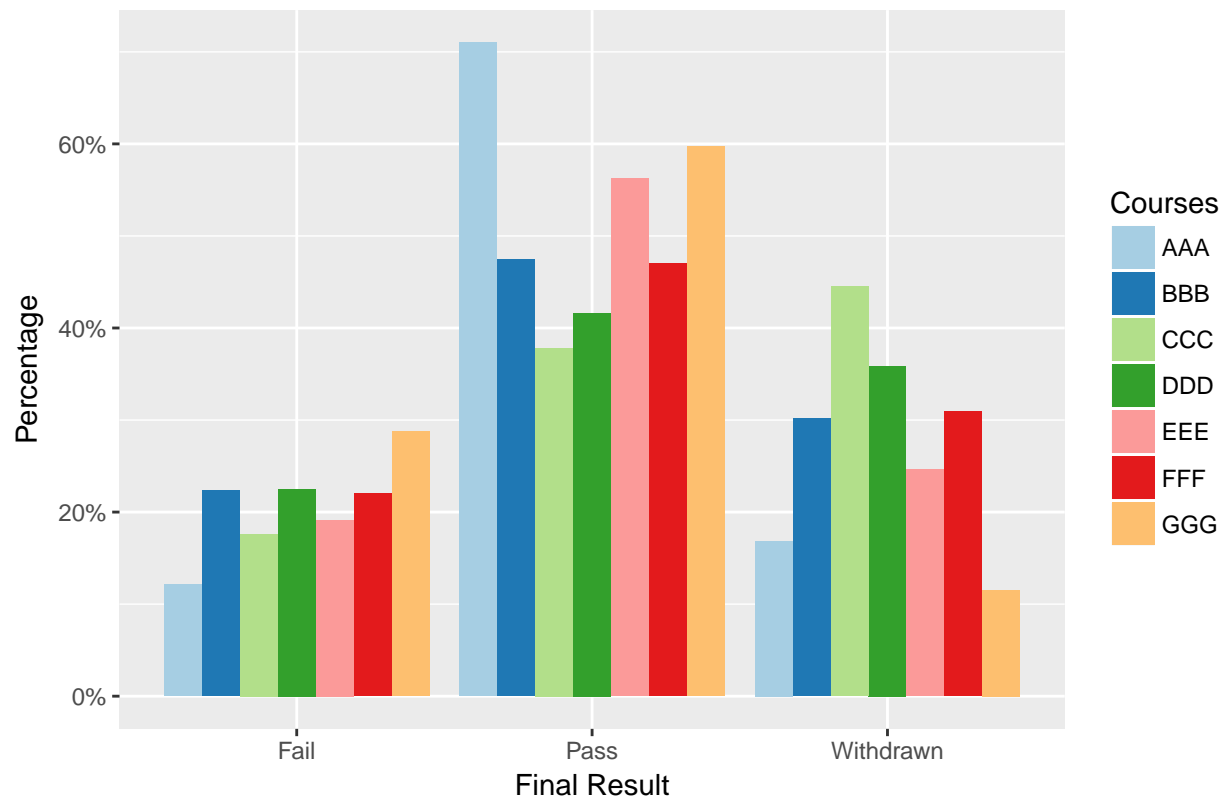

Student Performance

```
## For all Four results
studentdrop %>%
  group_by(si_code_module,si_final_result)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=si_final_result,y=percent,fill=si_code_module))+
  geom_bar(stat='identity',position='dodge')+
  scale_fill_manual(values=color)+
  labs(x='Final Result',y='Percentage',
       title='Percentage of Student Performance for Each Course')+
  guides(fill=guide_legend(title='Courses'))+
```

```
scale_y_continuous(labels = percent)
```

## Percentage of Student Performance for Each Course


```
## For three results, group Distinction and Pass
studentdrop %>%
  mutate(final_result=ifelse(si_final_result %in% c("Pass", "Distinction"), "Pass", as.character(si_fina
  group_by(si_code_module,final_result)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=final_result,y=percent,fill=si_code_module))+
  geom_bar(stat='identity',position='dodge')+
  scale_fill_manual(values=color)+
  labs(x='Final Result',y='Percentage',
       title='Percentage of Student Performance for Each Course')+
  guides(fill=guide_legend(title='Courses'))+
  scale_y_continuous(labels = percent)
```

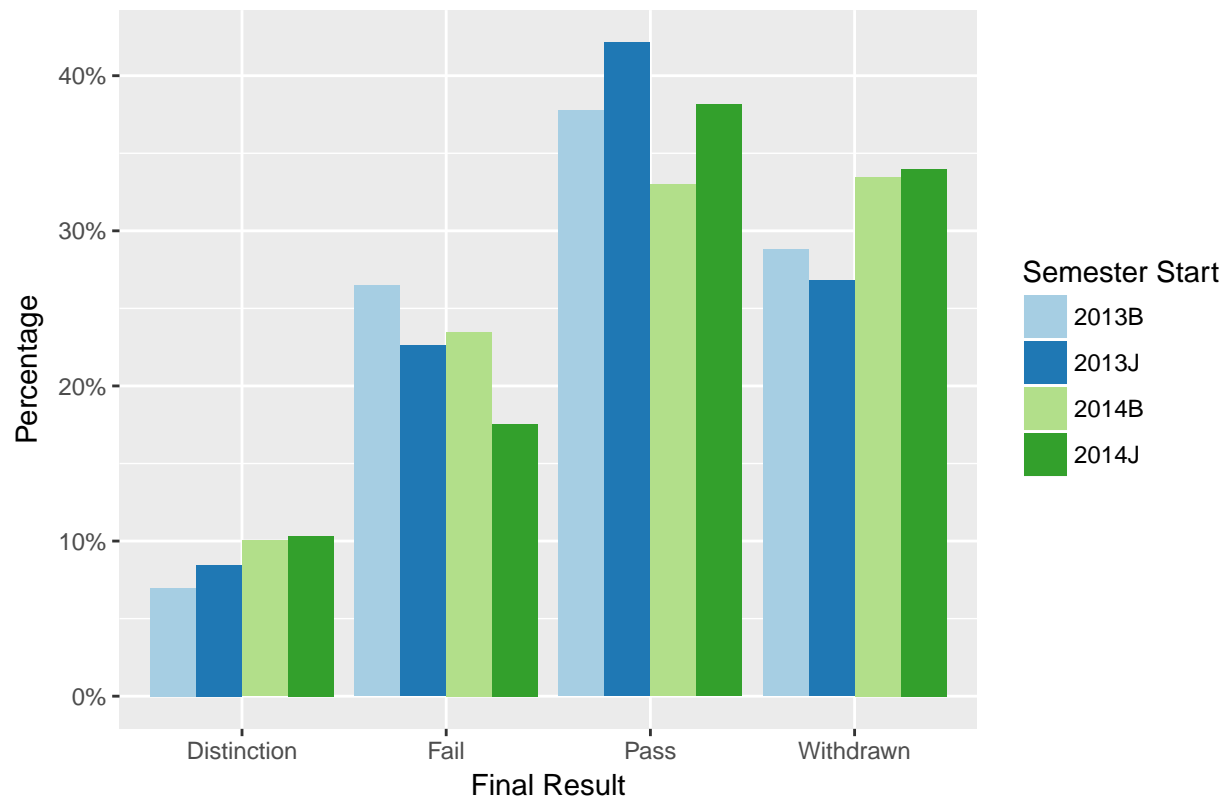## Percentage of Student Performance for Each Course


```
sd=studentdrop%>%
  mutate(Drop=ifelse(!is.na(sr_date_unregistration),1,0))

sd %>%
  group_by(si_code_module,si_num_of_prev_attempts)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=si_num_of_prev_attempts,y=percent,fill=si_code_module))+
  geom_bar(stat='identity',position='dodge')+
  scale_fill_manual(values=color)+
  labs(x='Previous Attempts',y='Percentage',
       title='Percentage of Student Performance for Each Course')+
  guides(fill=guide_legend(title='Courses'))+
  scale_x_continuous(breaks=seq(0,5,1))+
  scale_y_continuous(labels = percent)
```
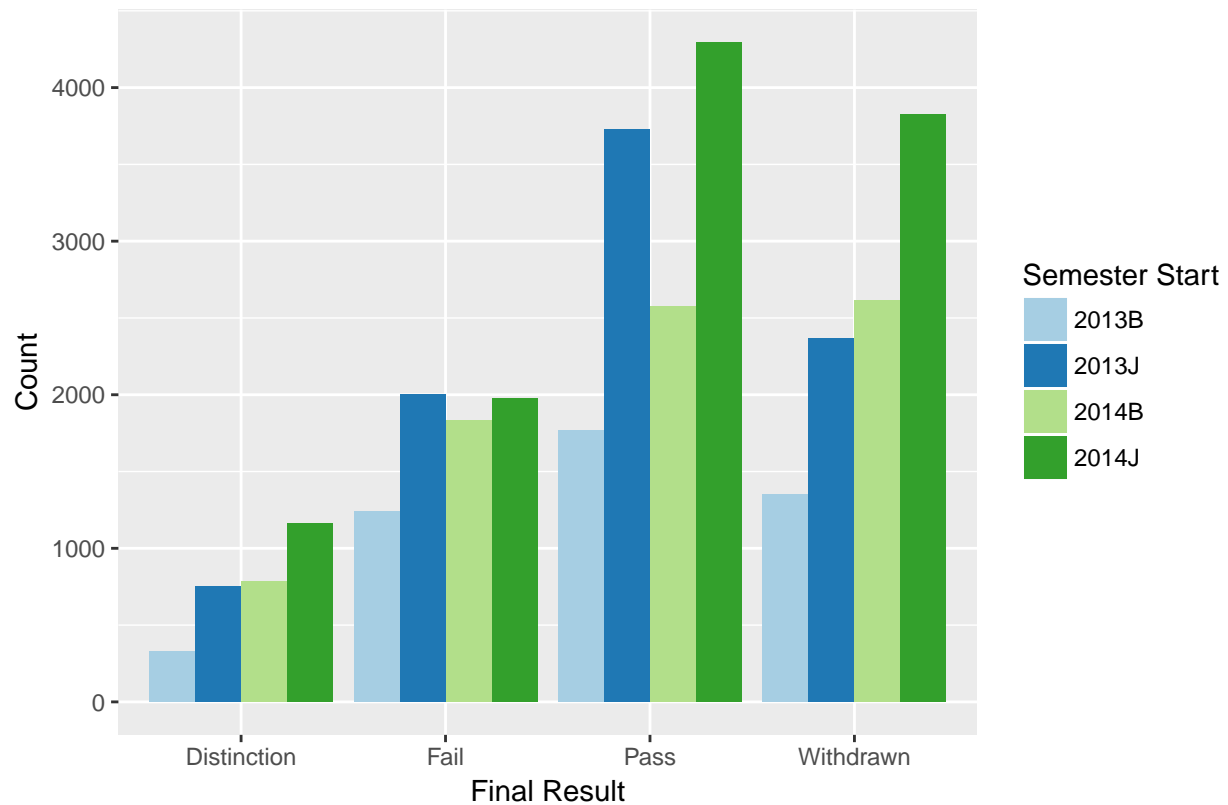
Percentage of Student Performance for Each Course
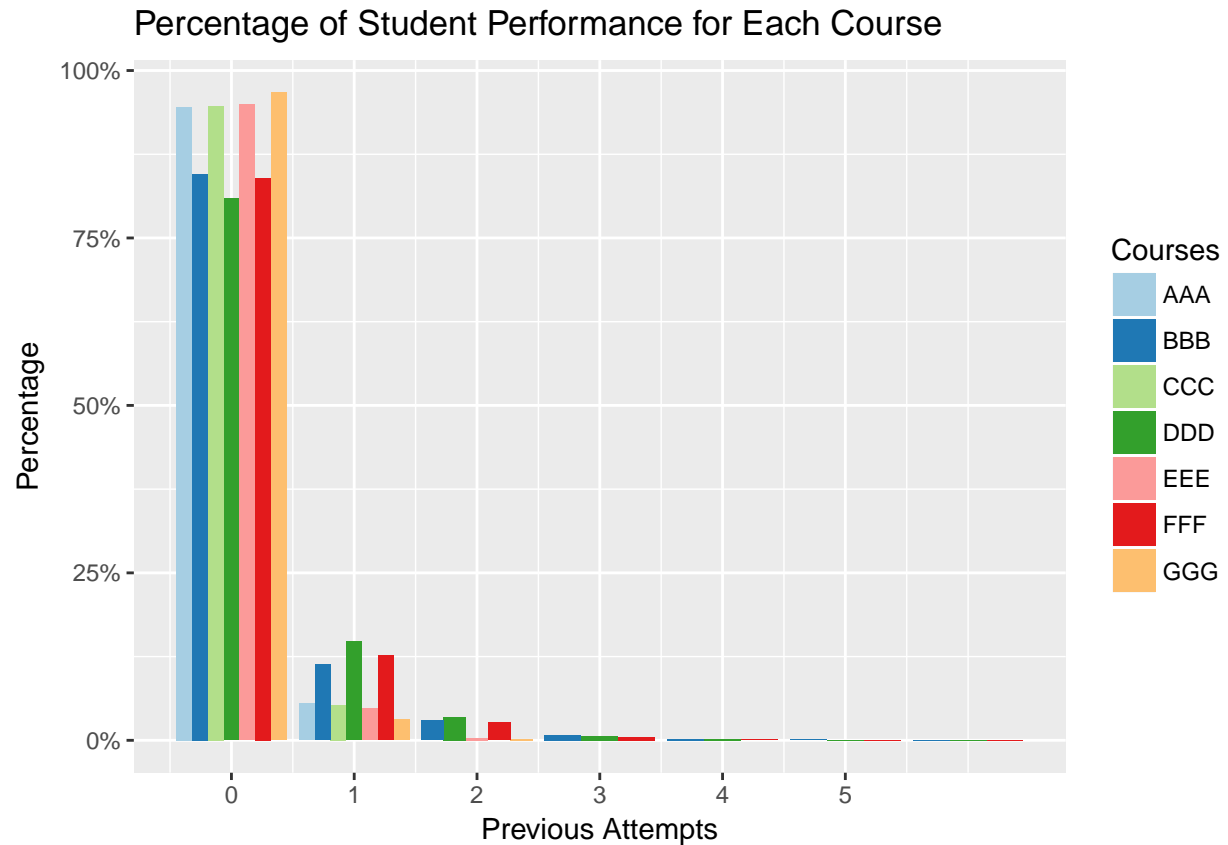

```
studentdrop %>%
  group_by(si_code_presentation,si_final_result)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=si_final_result,y=percent,fill=si_code_presentation))+
  geom_bar(stat='identity',position='dodge')+
  scale_fill_manual(values=color)+
  labs(x='Final Result',y='Percentage',
       title='Percentage of Student Performance for Each Semester')+
  guides(fill=guide_legend(title='Semester Start'))+
  scale_y_continuous(labels = percent)
```

Percentage of Student Performance for Each Semester


```
studentdrop %>%
  ggplot(aes(x=si_final_result,fill=si_code_presentation))+
  geom_bar(position='dodge')+
  scale_fill_manual(values=color)+
  labs(x='Final Result',y='Count',
       title='The Number of Student Performance for Each Semester')+
  guides(fill=guide_legend(title='Semester Start'))
```
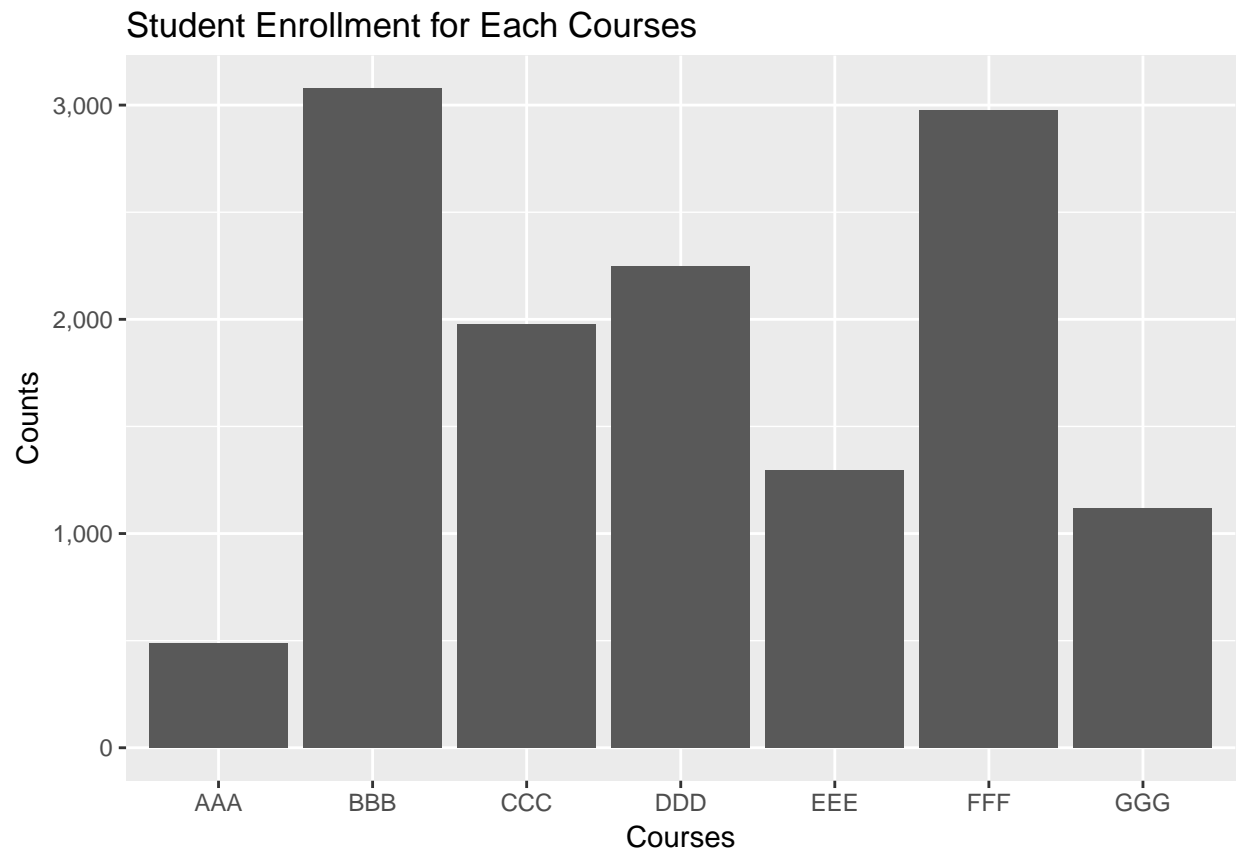
## The Number of Student Performance for Each Semester

Enrollment

```
studentdrop %>%
  group_by(si_code_module,si_final_result)%>%
  summarize(counts=n())%>%
  ggplot(aes(x=si_code_module,y=counts))+
  geom_bar(stat='identity',position='dodge')+
  scale_fill_manual(values=color)+
  labs(x='Courses',y='Counts',
       title='Student Enrollment for Each Courses')+
  scale_y_continuous(labels = comma)
```

Student Enrollment for Each Courses
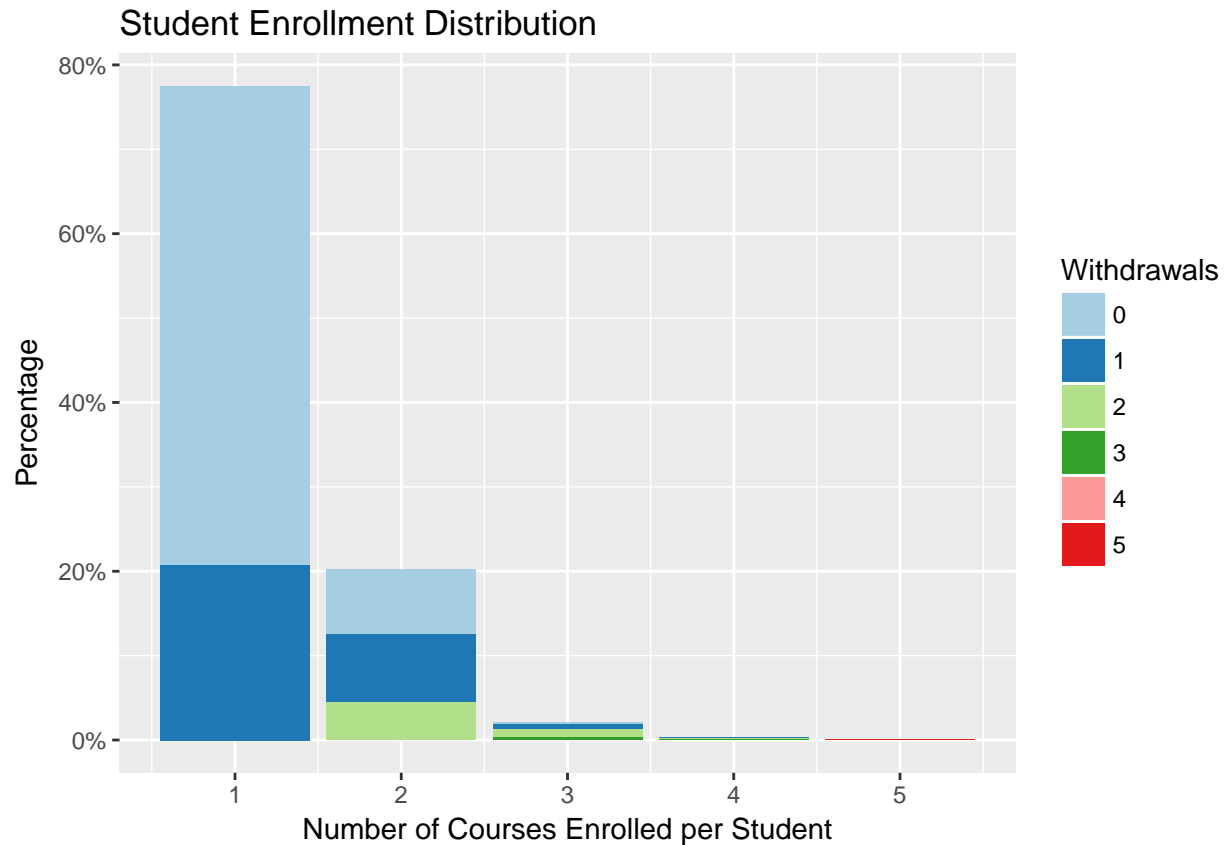


```
sd %>%
  group_by(si_id_student)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=counts,y=percent))+
  geom_bar(stat='identity')+
  labs(x='Number of Courses Enrolled per Student',y='Percentage',
       title='Student Enrollment Distribution')+
  scale_y_continuous(labels = percent)
```

Student Enrollment Distribution
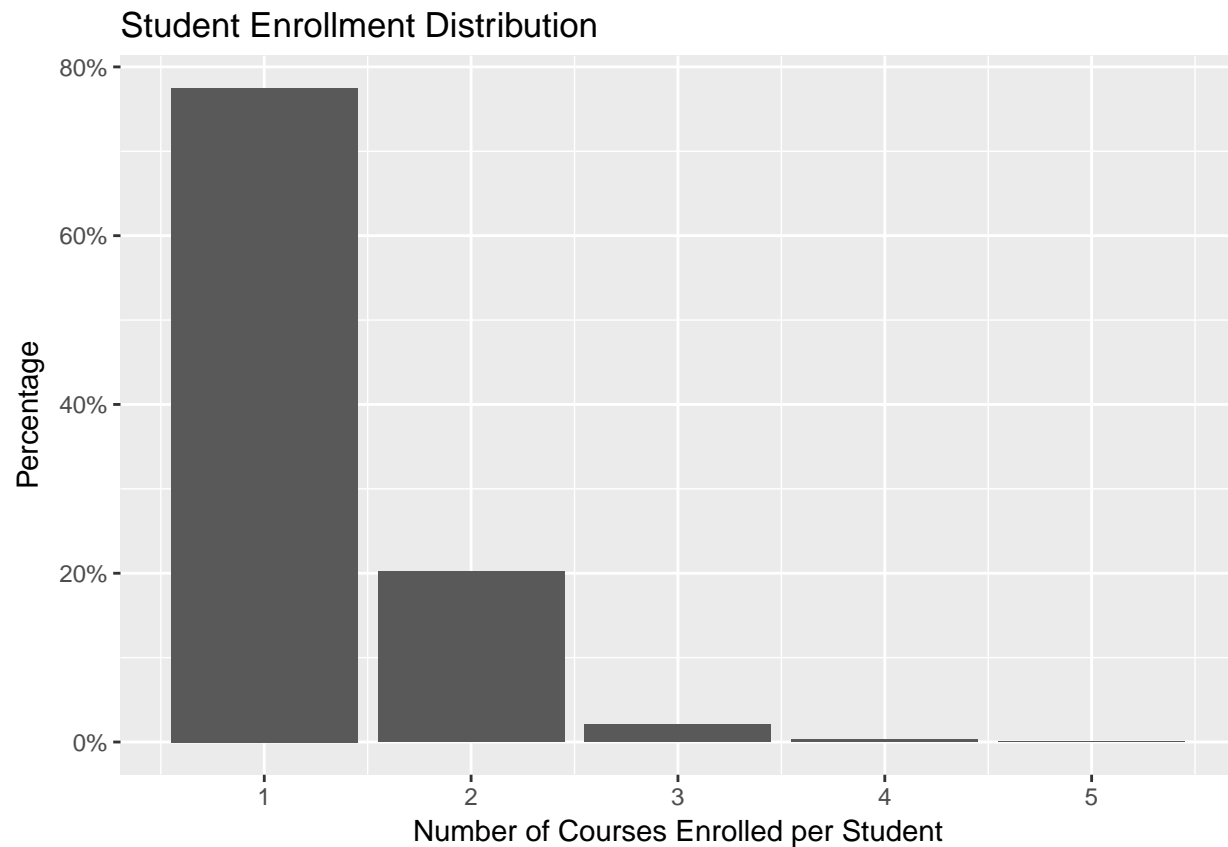

```
sd %>%
  group_by(si_id_student)%>%
  summarize(counts=n(),totaldrop=sum(Drop))%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=counts,y=percent,fill=as.factor(totaldrop)))+
  geom_bar(stat='identity')+
  scale_fill_manual(values=color)+
  labs(x='Number of Courses Enrolled per Student',y='Percentage',
       title='Student Enrollment Distribution')+
  guides(fill=guide_legend(title='Withdrawals'))+
  scale_y_continuous(labels = percent)
```

Student Enrollment Distribution

Compare Course CCC and Course AAA

```
studentdrop %>%
  mutate(final_result=ifelse(si_final_result %in% c("Pass", "Distinction"), "Pass", as.character(si_fina
  filter(si_code_module=='CCC',si_highest_education!='No Formal quals')%>%
  group_by(si_highest_education,final_result)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=final_result,y=percent,fill=si_highest_education))+
  geom_bar(stat='identity',position='dodge')+
  scale_fill_manual(values=color)+
  labs(x='Final Result',y='Percentage',
       title='Percentage of Student Performance at Each Education Level for CCC')+
  guides(fill=guide_legend(title='Highest Education'))+
  scale_y_continuous(labels = percent)
```

Percentage of Student Performance at Each Education Level for CCC
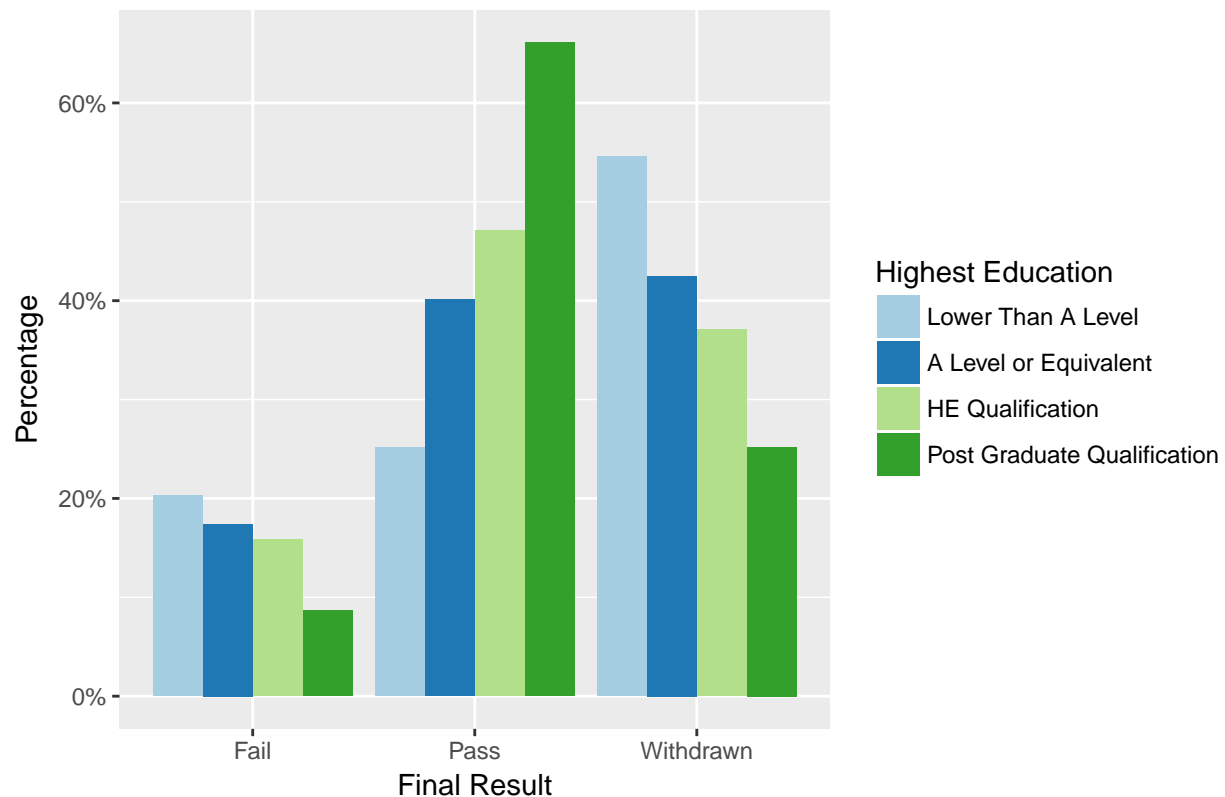


```
studentdrop %>%
  mutate(final_result=ifelse(si_final_result %in% c("Pass", "Distinction"), "Pass", as.character(si_fina
  filter(si_code_module=='AAA',si_highest_education!='No Formal quals')%>%
  group_by(si_highest_education,final_result)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=final_result,y=percent,fill=si_highest_education))+
  geom_bar(stat='identity',position='dodge')+
  scale_fill_manual(values=color)+
  labs(x='Final Result',y='Percentage',
       title='Percentage of Student Performance for Each Education Level at AAA')+
  guides(fill=guide_legend(title='Highest Education'))+
  scale_y_continuous(labels = percent)
```

Percentage of Student Performance for Each Education Level at AAA
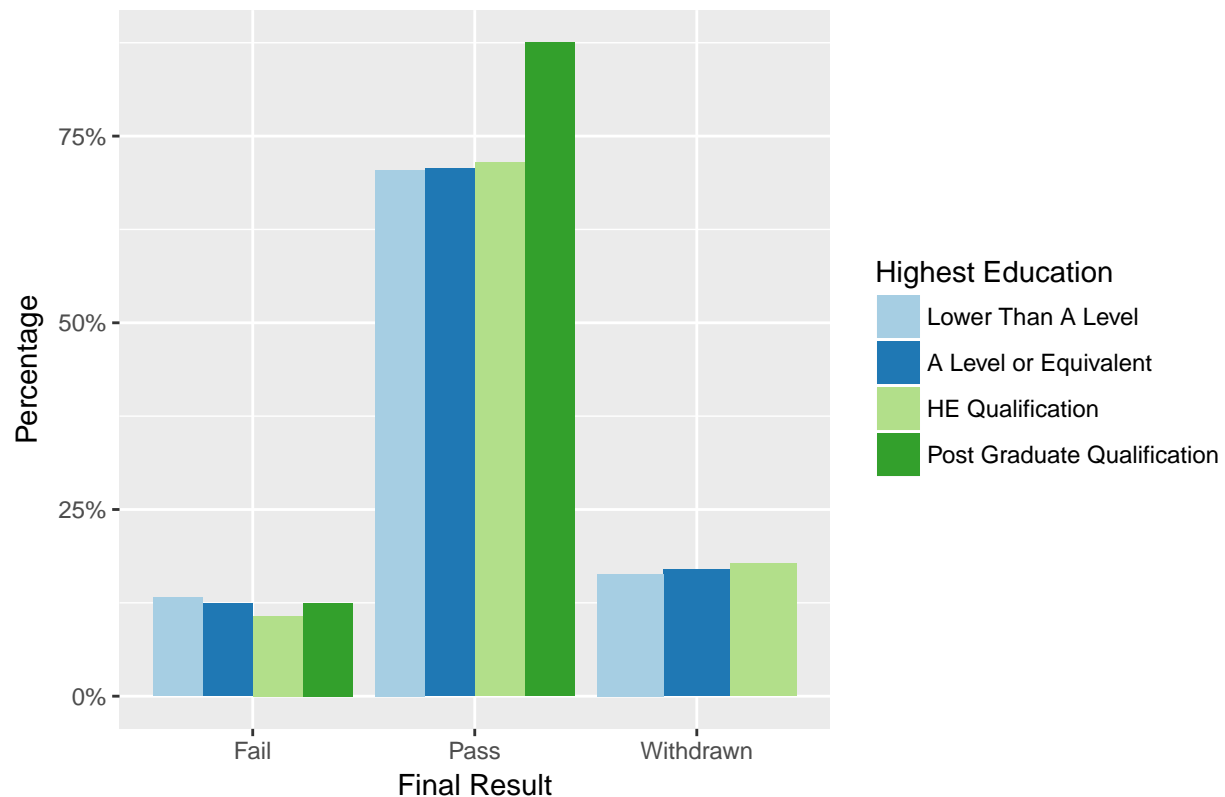


```
studentdrop %>%
  mutate(drop_test=ifelse(sr_date_unregistration<=0,'Drop Before Start',
          ifelse(sr_date_unregistration<50,'Drop After Two Test (10%)',
          ifelse(sr_date_unregistration<150,'Drop After Four Test (25%)',
          ifelse(sr_date_unregistration<175,'Drop After First Exam (22%)',
          ifelse(sr_date_unregistration<205,'Drop After Second Exam (22%)',
          ifelse(sr_date_unregistration<225,'Drop After Final Exam (30%)','')))))))%>%
  filter(si_code_module=='CCC',si_final_result=='Withdrawn',drop_test!='')%>%
  group_by(drop_test,si_code_module,si_final_result)%>%
  summarize(count=n())%>%
  arrange(count)%>%
  ggplot(aes(x=si_final_result,y=count,fill=reorder(drop_test,-count)))+
  geom_bar(stat='identity',position='dodge',width = 0.3)+
  scale_fill_manual(values=color)+
  labs(x='Final Result',y='Count',
       title='The Number of Student Withdrawal After Each Test for Course CCC')+
  guides(fill=guide_legend(title=''))+
  scale_y_continuous(labels = comma)
```

The Number of Student Withdrawal After Each Test for Course CCC
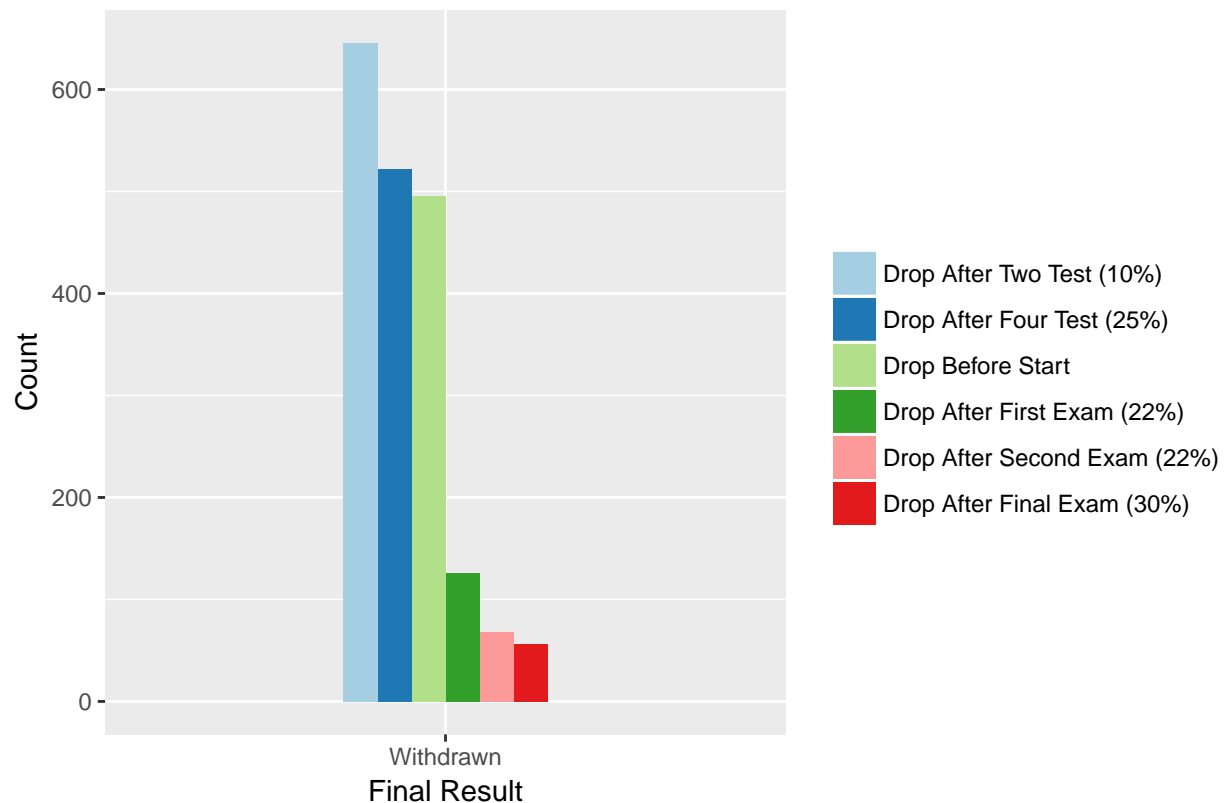

```
studentdrop %>%
  mutate(drop_test=ifelse(sr_date_unregistration<=0,'Drop Before Start',
              ifelse(sr_date_unregistration<25,'Drop After First Test (10%)',
              ifelse(sr_date_unregistration<75,'Drop After Second Test (20%)',
              ifelse(sr_date_unregistration<150,'Drop After Third Test (20%)',
              ifelse(sr_date_unregistration<175,'Drop After Fourth Test (20%)',
   ifelse(sr_date_unregistration<225,'Drop After Final Exam (30%)','')))))))%>%
  filter(si_code_module=='AAA',si_final_result=='Withdrawn',drop_test!='')%>%
  group_by(drop_test,si_code_module,si_final_result)%>%
  summarize(count=n())%>%
  arrange(count)%>%
  ggplot(aes(x=si_final_result,y=count,fill=reorder(drop_test,-count)))+
  geom_bar(stat='identity',position='dodge',width = 0.3)+
  scale_fill_manual(values=color)+
  labs(x='Final Result',y='Count',
       title='The Number of Student Withdrawal After Each Test for Course AAA')+
  guides(fill=guide_legend(title=''))+
  scale_y_continuous(labels = comma)
```

## The Number of Student Withdrawal After Each Test for Course AAA
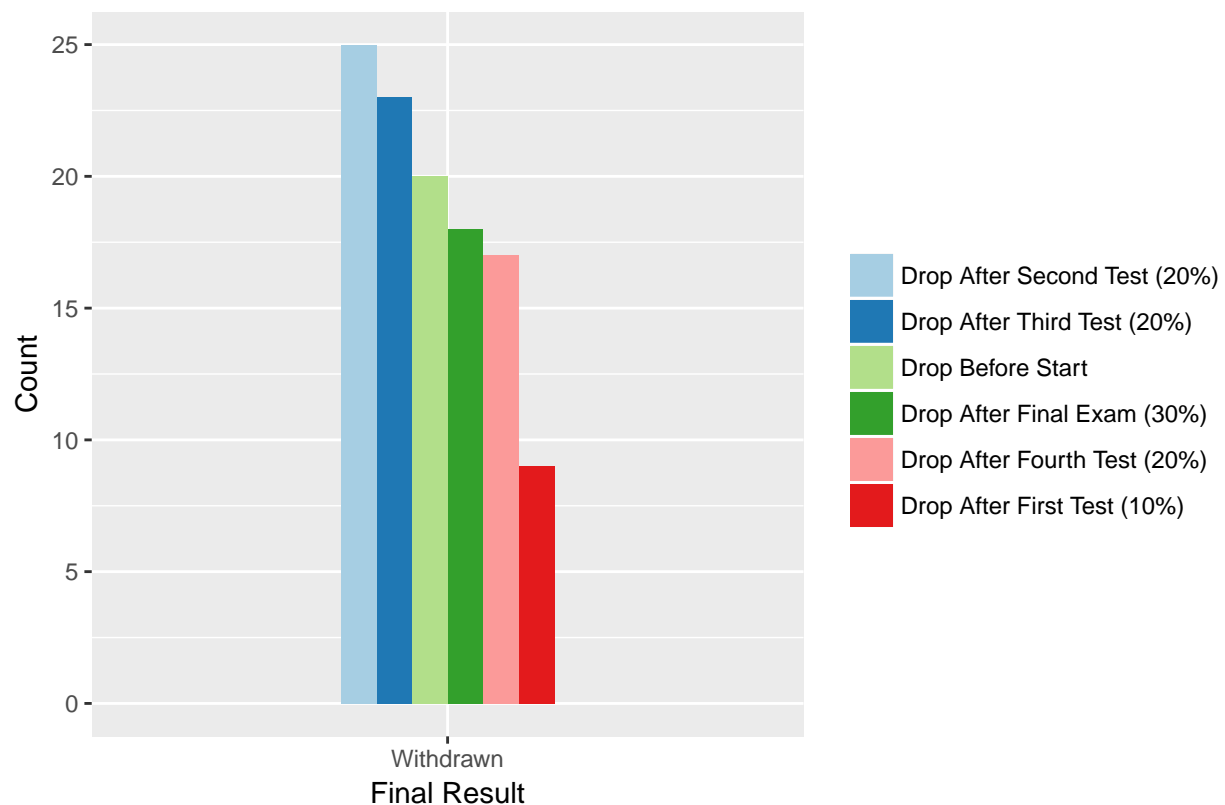

```
## Check the Type of Test & Weight of Test (Choose different date range and repeat)
studenttestdrop %>%
  filter(ou_student_assessment_a_date>200,ou_student_assessment_a_date<225,
         ou_student_assessment_a_code_module=='AAA')%>%
  group_by(ou_student_assessment_a_assessment_type)%>%
  select(ou_student_assessment_a_assessment_type)

studenttestdrop %>%
  filter(ou_student_assessment_a_date>200,ou_student_assessment_a_date<225,
         ou_student_assessment_a_code_module=='AAA')%>%
  select(ou_student_assessment_a_weight)
```

Compare Course CCC and AAA Cont.

```
## Course CCC by Education Count
studentdrop %>%
  group_by(si_code_module,si_final_result,si_highest_education)%>%
  summarize(counts=n())%>%
  filter(si_code_module=='CCC')%>%
  ggplot(aes(x=reorder(si_highest_education,-counts),y=counts))+
  geom_bar(stat='identity')+
  labs(x='Final Result',y='Count',
       title='Student Education Background for Course CCC')+
  theme(axis.text.x=element_text(angle=15,hjust=1))+
```
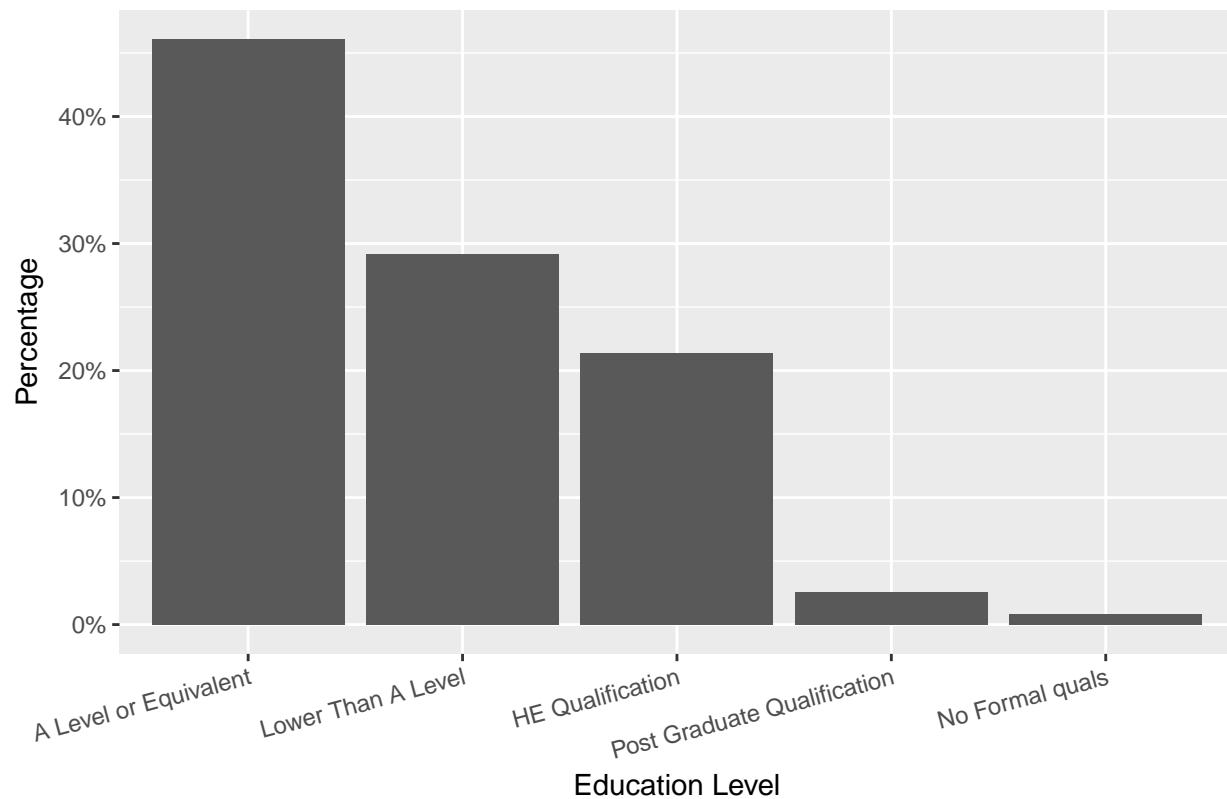
```
guides(fill=guide_legend(title='Highest Education'))
```

Student Education Background for Course CCC
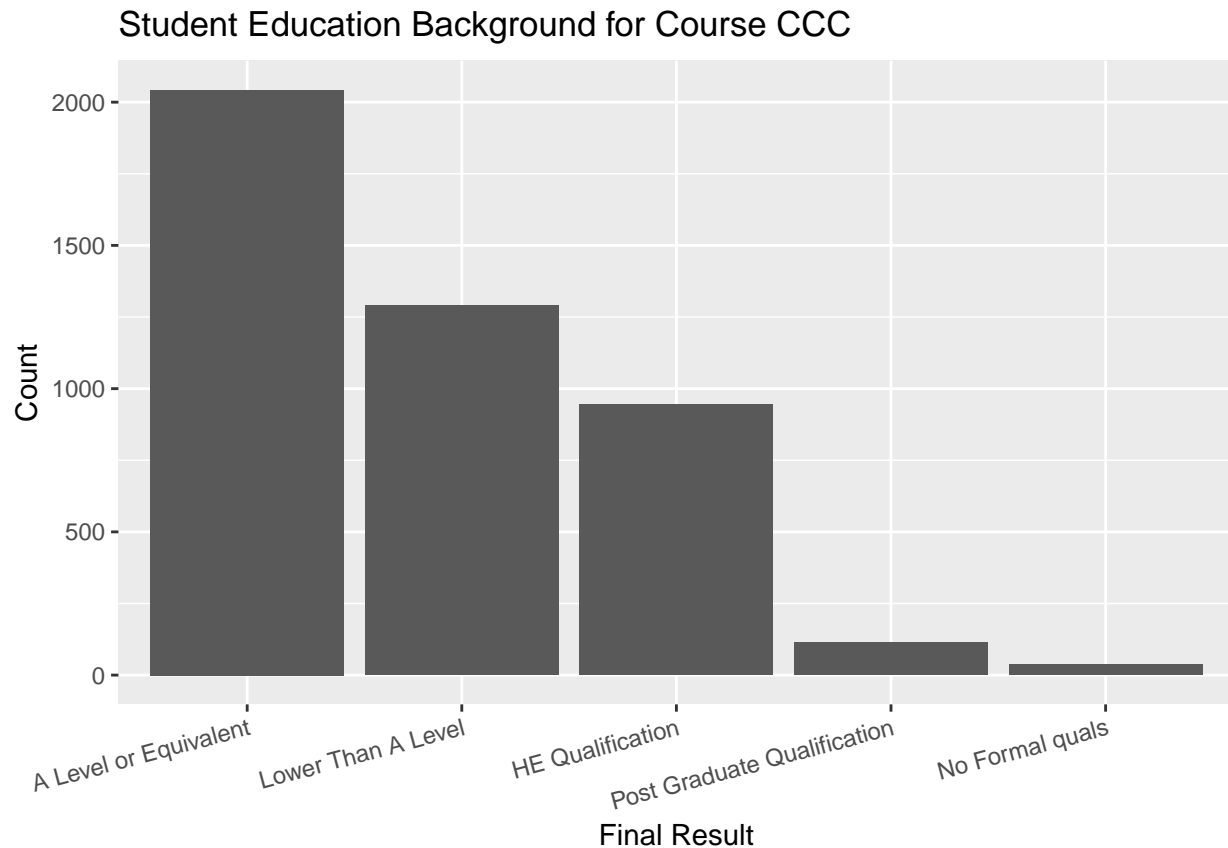

```
## Student Education background in % for Course CCC
studentdrop %>%
  filter(si_code_module=='CCC')%>%
  group_by(si_code_module,si_highest_education)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=reorder(si_highest_education,-percent),y=percent))+
  geom_bar(stat='identity')+
  labs(x='Education Level',y='Percentage',
      title='Student Education Background for Course CCC')+
  theme(axis.text.x=element_text(angle=15,hjust=1))+
  scale_y_continuous(labels = percent)
```

## Student Education Background for Course CCC


```
## Student Education background in % for Course AAA
studentdrop %>%
  filter(si_code_module=='AAA')%>%
  group_by(si_code_module,si_highest_education)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=reorder(si_highest_education,-percent),y=percent))+
  geom_bar(stat='identity')+
  labs(x='Education Level',y='Percentage',
       title='Student Education Background for Course AAA')+
  scale_y_continuous(labels = percent)
```

## Student Education Background for Course AAA
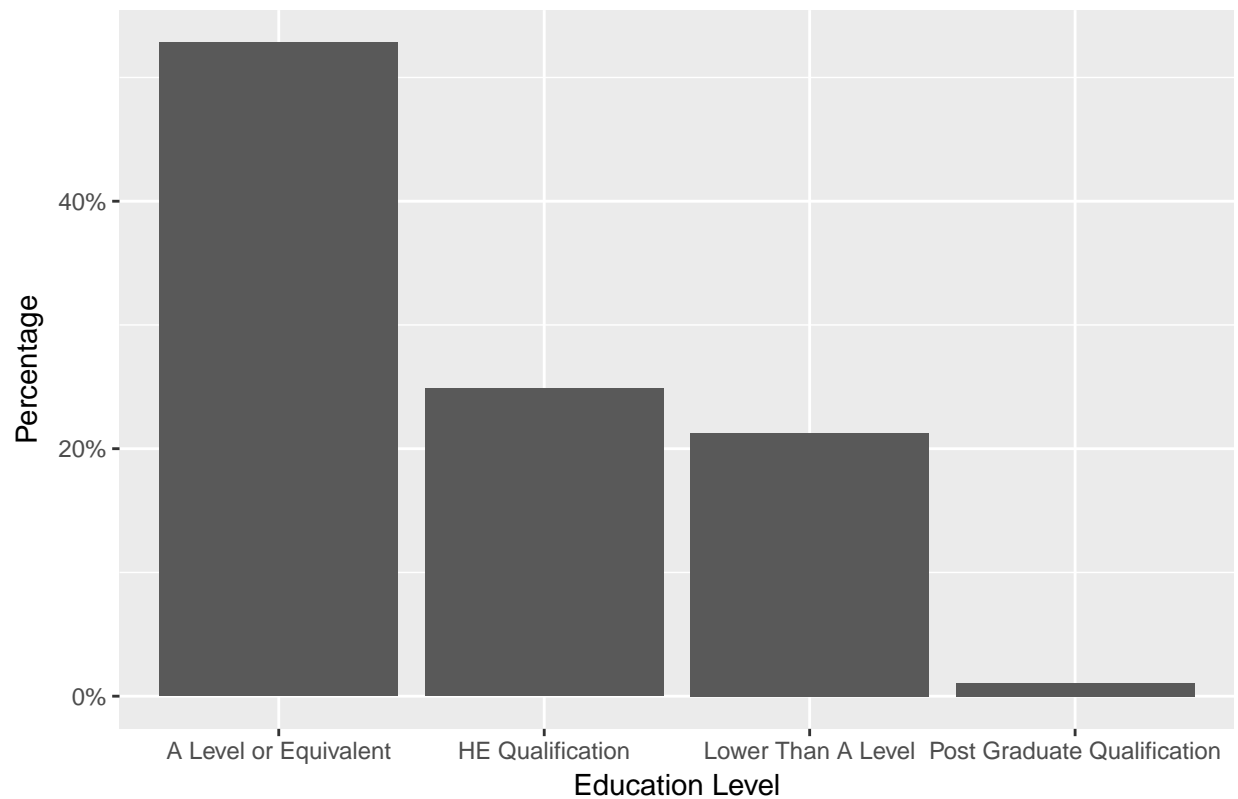

```
## Course CCC withdrawal for A level, lower than A Level Students
studentdrop %>%
  filter(si_final_result=='Withdrawn',
         si_code_module=='CCC',
         sr_date_unregistration<250,sr_date_unregistration>-225,
         si_highest_education %in% c('A Level or Equivalent','Lower Than A Level'))%>%
  ggplot(aes(x=sr_date_unregistration,fill=si_highest_education))+
  geom_histogram(binwidth = 10)+
  scale_x_continuous(breaks=seq(-200,500,25))+
  scale_fill_manual(values=color)+
  labs(x='Withdrawal Date',y='Count',
       title='The Distribution of Withdrawal Days for Course CCC')+
  guides(fill=guide_legend(title='Highest Education'))
```

The Distribution of Withdrawal Days for Course CCC
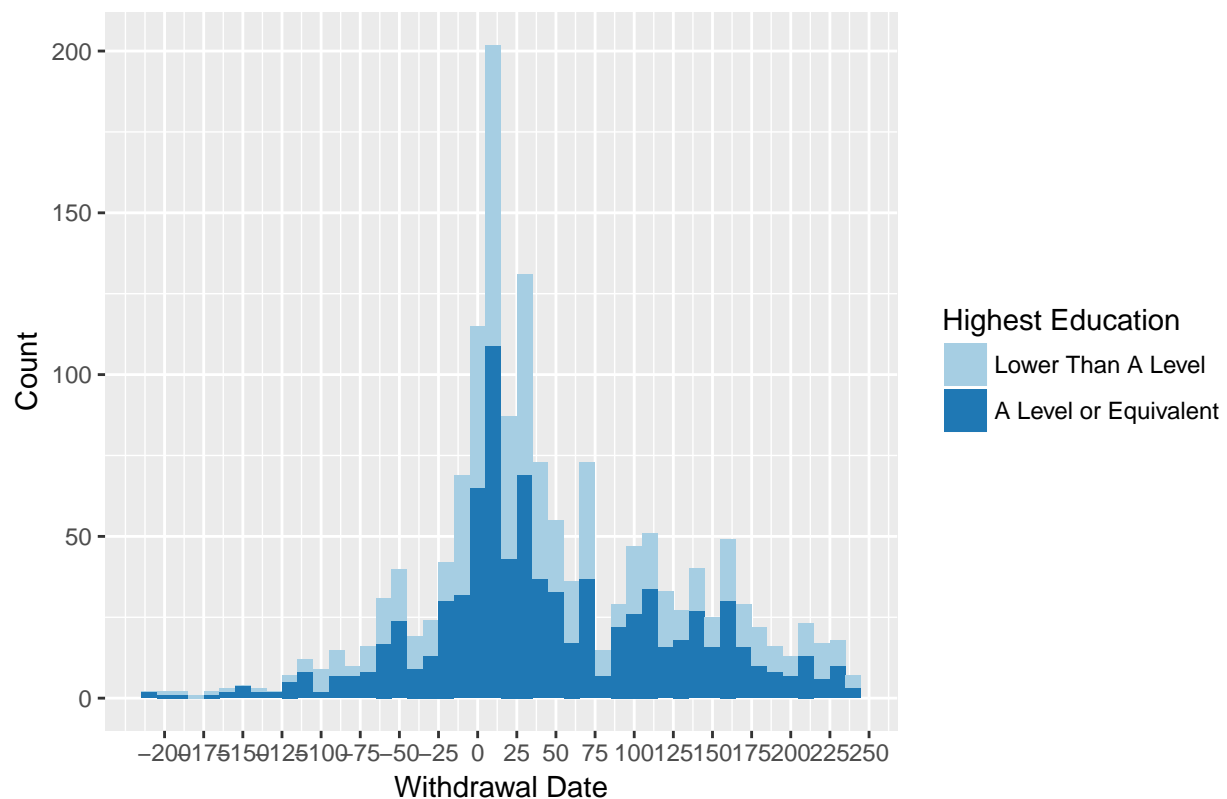

```
## Course AAA by Higest Education Percentage
studentdrop %>%
  mutate(final_result=ifelse(si_final_result %in% c("Pass", "Distinction"), "Pass", as.character(si_fina
  filter(si_code_module=='AAA')%>%
  group_by(si_highest_education,final_result)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=final_result,y=percent,fill=si_highest_education))+
  geom_bar(stat='identity',position='dodge')+
  scale_fill_manual(values=color)+
  labs(x='Final Result',y='Percentage',
       title='Percentage of Student Performance at Each Education Level for AAA')+
  guides(fill=guide_legend(title='Highest Education'))
```

Percentage of Student Performance at Each Education Level for AAA
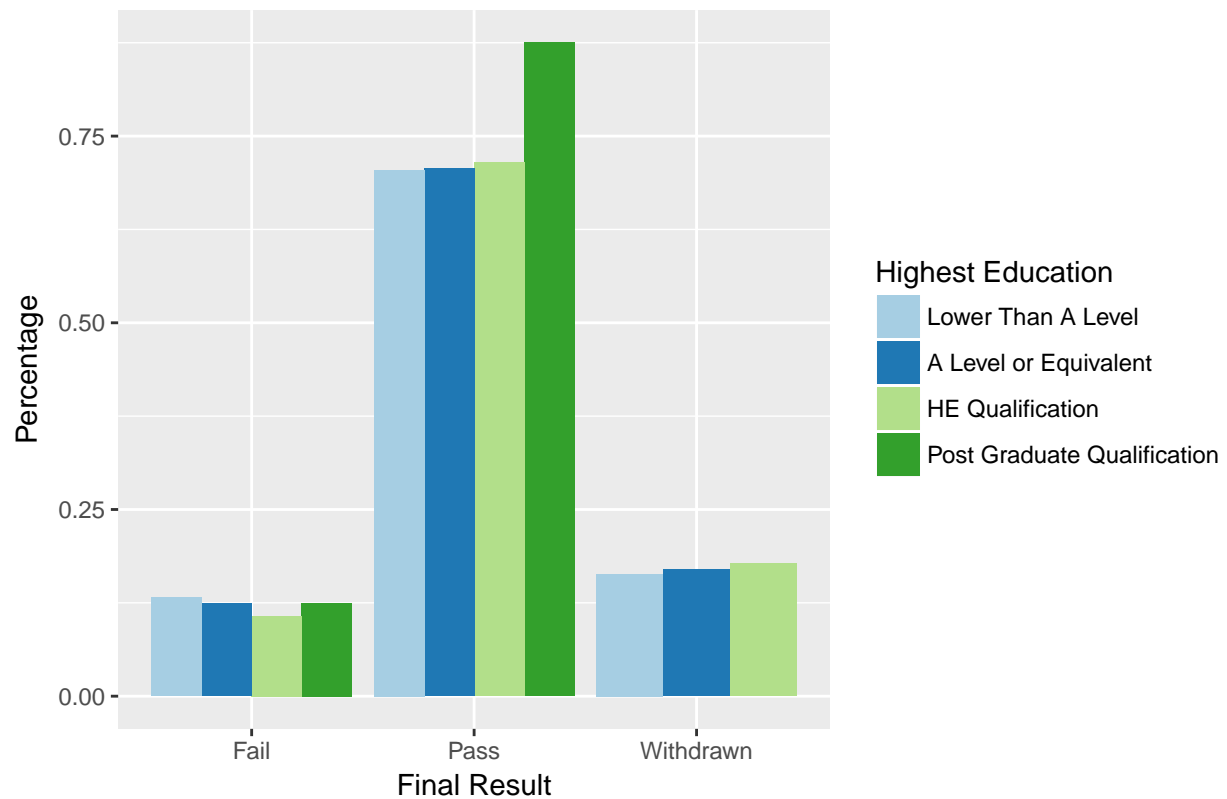

```
## Course AAA by Education Count
studentdrop %>%
  group_by(si_code_module,si_final_result,si_highest_education)%>%
  summarize(counts=n())%>%
  filter(si_code_module=='AAA')%>%
  ggplot(aes(x=reorder(si_highest_education,-counts),y=counts))+
  geom_bar(stat='identity')+
  labs(x='Final Result',y='Count',
       title='Student Education Background for Course AAA')+
  guides(fill=guide_legend(title='Highest Education'))
```

Student Education Background for Course AAA
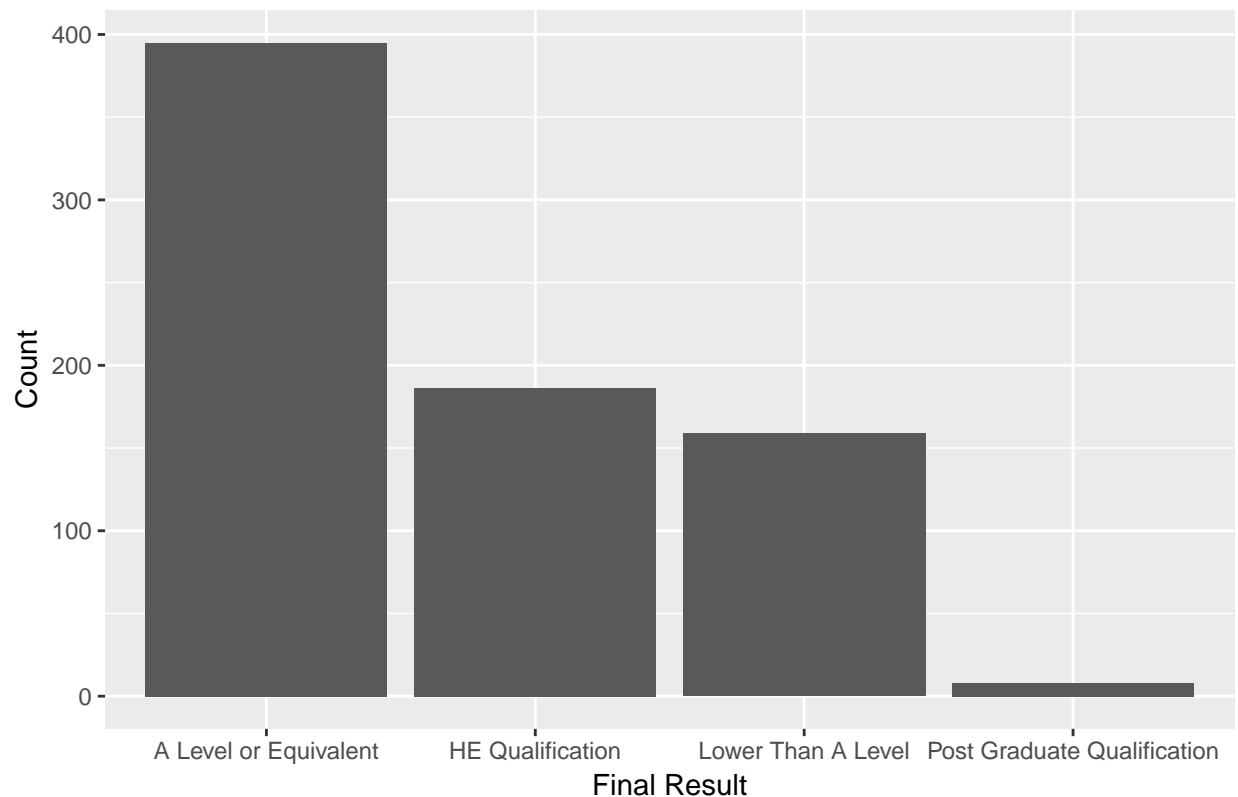
Student Assessment for Course CCC and AAA

```
studenttestdrop=read.csv('student_as_drop.csv')

studenttestdrop$ou_student_drop_date_unregistration=
  as.numeric(levels(studenttestdrop$ou_student_drop_date_unregistration))[studenttestdrop$ou_student_dr
```

```
## Warning: NAs introduced by coercion
```

```
studenttestdrop$ou_student_assessment_a_date=
  as.numeric(levels(studenttestdrop$ou_student_assessment_a_date))[studenttestdrop$ou_student_assessment
```

```
## Warning: NAs introduced by coercion
```

```
studenttestdrop$ou_student_assessment_sa_score=
  as.numeric(levels(studenttestdrop$ou_student_assessment_sa_score))[studenttestdrop$ou_student_assessme
```
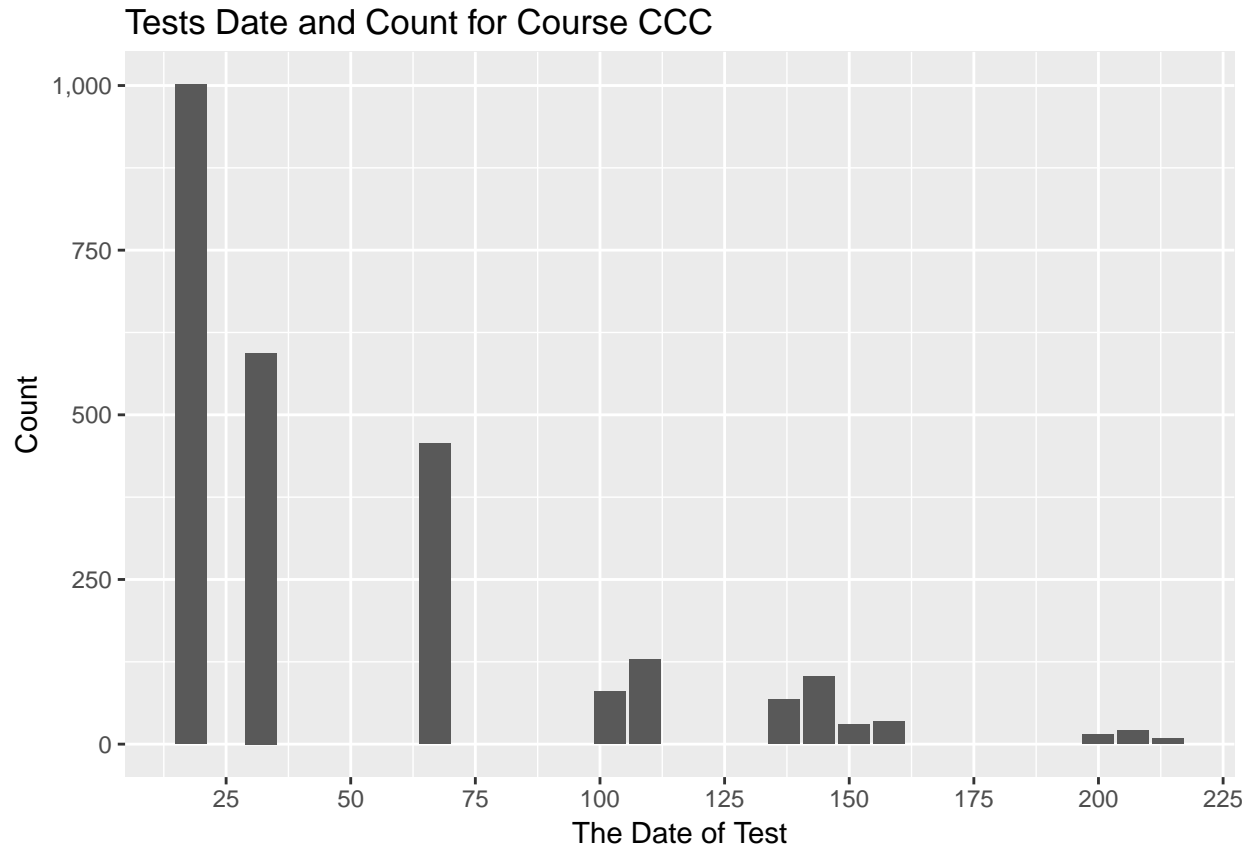
```
## Warning: NAs introduced by coercion
### test date and count for course CCC
studenttestdrop%>%
  filter(ou_student_drop_final_result=='Withdrawn',ou_student_drop_si_code_module=='CCC')%>%
  ggplot(aes(x=ou_student_assessment_a_date))+
  geom_bar()+
  scale_x_continuous(breaks = seq(0,250,25))+
  labs(x='The Date of Test',y='Count',
```

```
            title='Tests Date and Count for Course CCC')+
  scale_y_continuous(labels = comma)
```

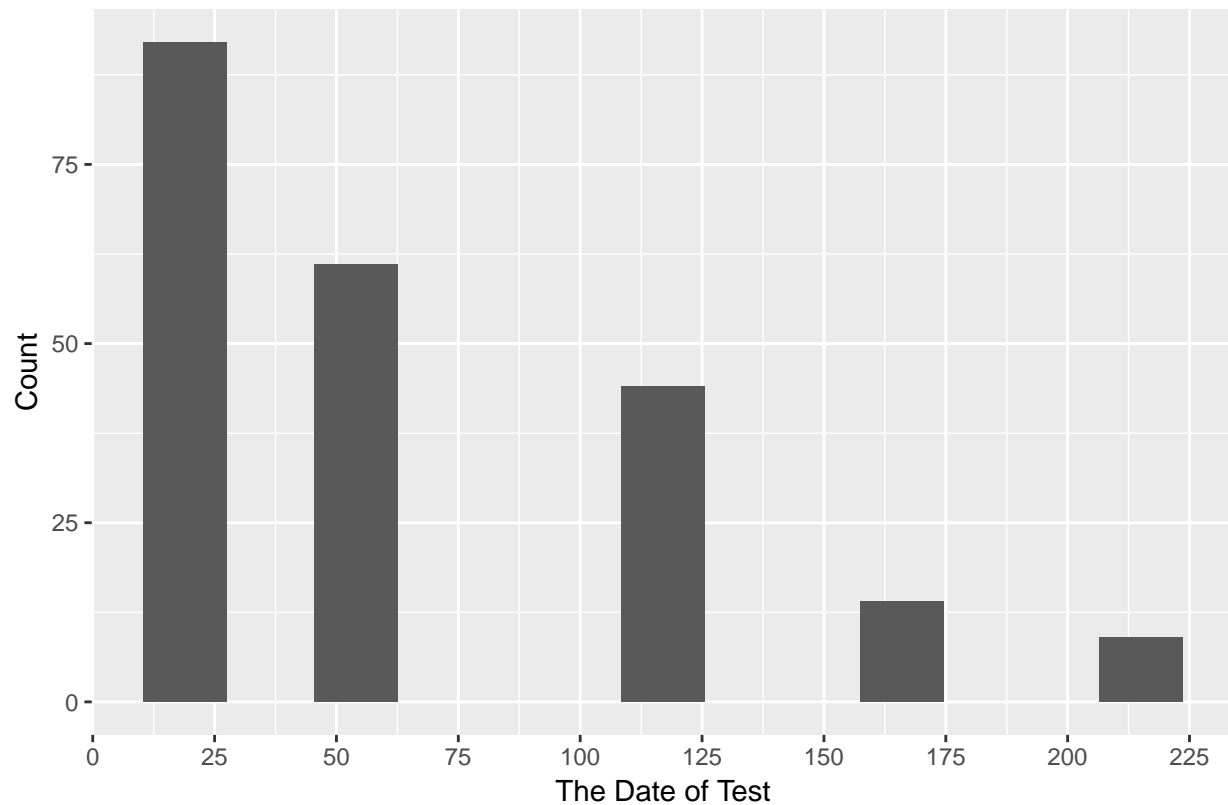## Tests Date and Count for Course CCC


```
### test date and count for course AAA
studenttestdrop%>%
  filter(ou_student_drop_final_result=='Withdrawn',ou_student_drop_si_code_module=='AAA')%>%
  ggplot(aes(x=ou_student_assessment_a_date))+
  geom_bar()+
  scale_x_continuous(breaks = seq(0,250,25))+
  labs(x='The Date of Test',y='Count',
       title='Tests Date and Count for Course AAA')
```

Tests Date and Count for Course AAA
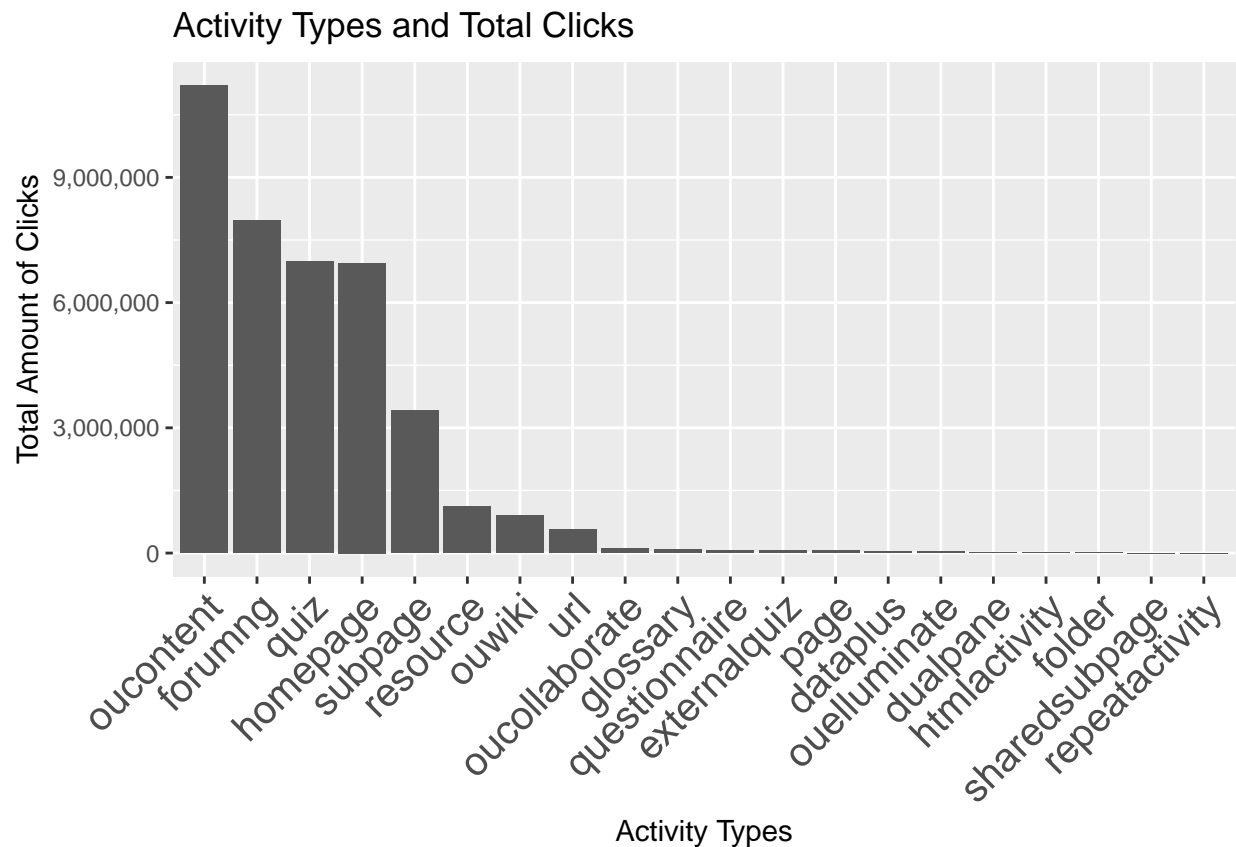
Student Activities and Withdrawal

```
## Student Credit distribution
studentdrop %>%
  group_by(studentid=si_id_student,semester_start=si_code_presentation)%>%
  summarize(credit=max(si_studied_credits))%>%
  filter(credit<360,semester_start=='2014J')%>%
  select(studentid,credit)%>%
  ggplot(aes(credit))+
  geom_histogram(binwidth = 30)+
  scale_x_continuous(breaks = seq(0,360,30))+
  labs(x='Current Credits',y='Number of Students',
       title='The Distribution of Current Credits ')+
  scale_y_continuous(labels = comma)
```

The Distribution of Current Credits
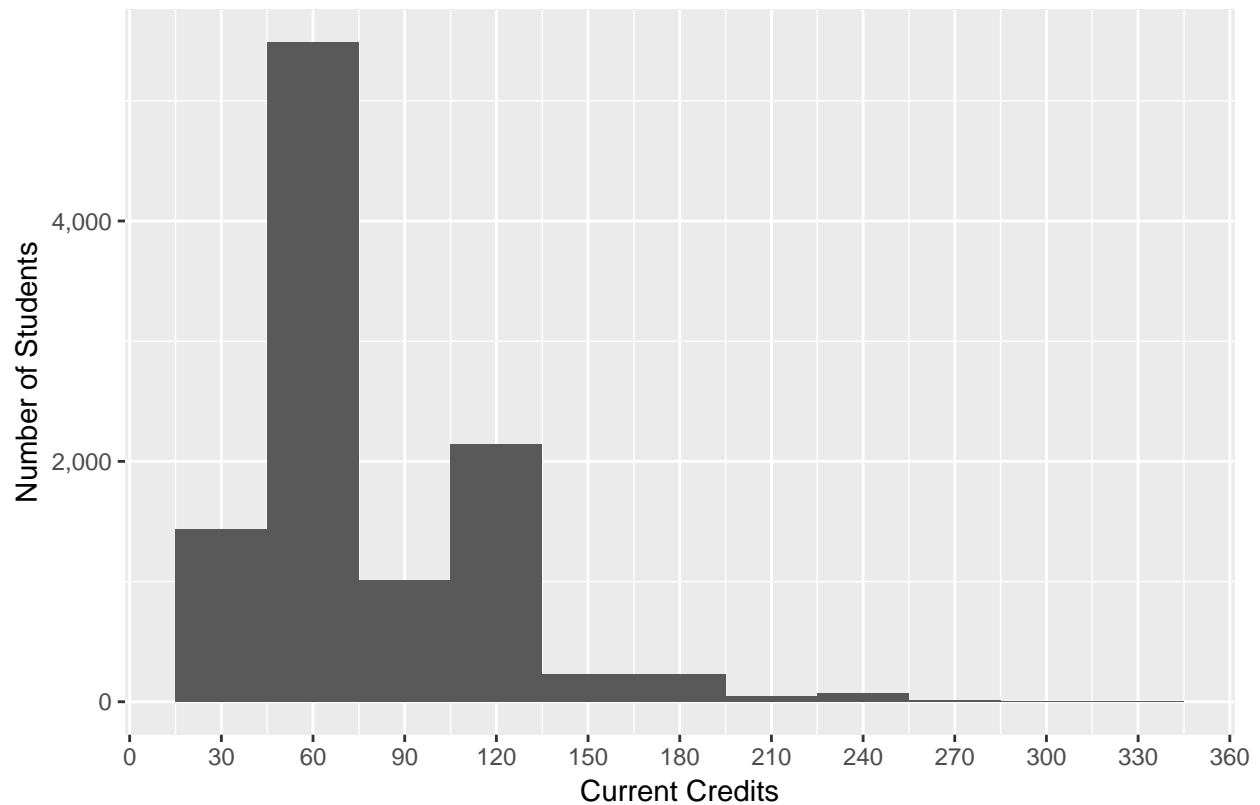

```
## Activity types and Total Amount of Clicks
studentvle=read.csv('student_vle_acti.csv')
studentvle%>%
  group_by(v_activity_type)%>%
  summarise(amount=sum(sv_sum_click))%>%
  ggplot(aes(x=reorder(v_activity_type,-amount),y=amount))+
  geom_bar(stat='identity')+
  theme(axis.text.x = element_text(size=15,angle=45,hjust=1))+
  scale_y_continuous(labels = comma)+
  labs(x='Activity Types',y='Total Amount of Clicks',
       title='Activity Types and Total Clicks')
```

Activity Types and Total Clicks


```r
## Activity types and Total Amount of Clicks for Withdrawal
student_acti_drop=read.csv('student_acti_drop.csv')
student_acti_drop%>%
  filter(final_result=="Withdrawn")%>%
  group_by(activity_type)%>%
  summarise(amount=sum(sum_click)/26022)%>% ## Used SQL to get the total distinct number of students
  ggplot(aes(x=reorder(activity_type,-amount),y=amount))+
  geom_bar(stat='identity')+
  theme(axis.text.x = element_text(size=15,angle=45,hjust=1))+
  scale_y_continuous(labels = comma)+
  labs(x='Activity Types',y='Total Amount of Clicks',
       title='Activity Types and Total Clicks for Withdrawal')
```

Activity Types and Total Clicks for Withdrawal
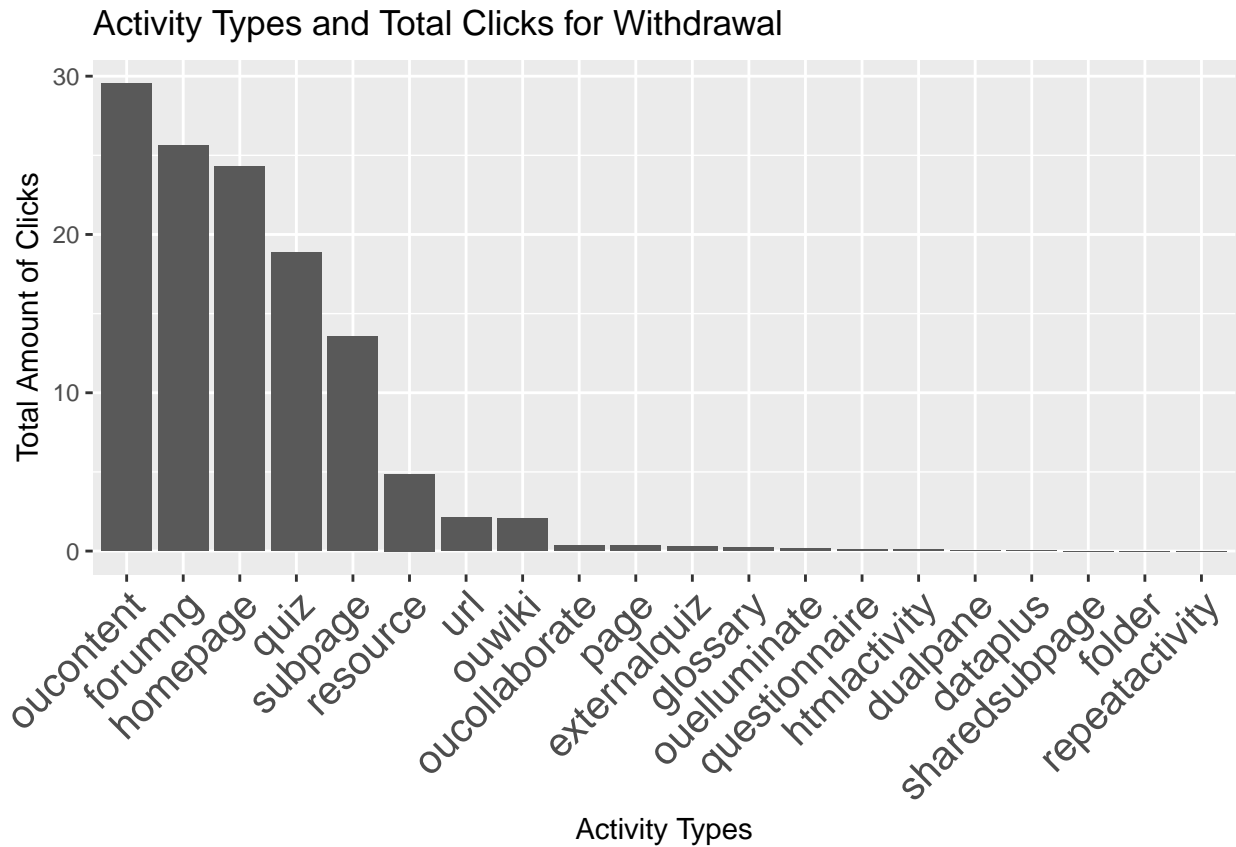

```
## Activity types and Total Amount of Clicks for Withdrawal
student_acti_drop%>%
  filter(final_result!="Withdrawn")%>%
  group_by(activity_type)%>%
  summarise(amount=sum(sum_click)/26022)%>%
  ggplot(aes(x=reorder(activity_type,-amount),y=amount))+
  geom_bar(stat='identity')+
  theme(axis.text.x = element_text(size=15,angle=45,hjust=1))+
  scale_y_continuous(labels = comma)+
  labs(x='Activity Types',y='Total Amount of Clicks',
       title='Activity Types and Total Clicks for Complete')
```

Activity Types and Total Clicks for Complete
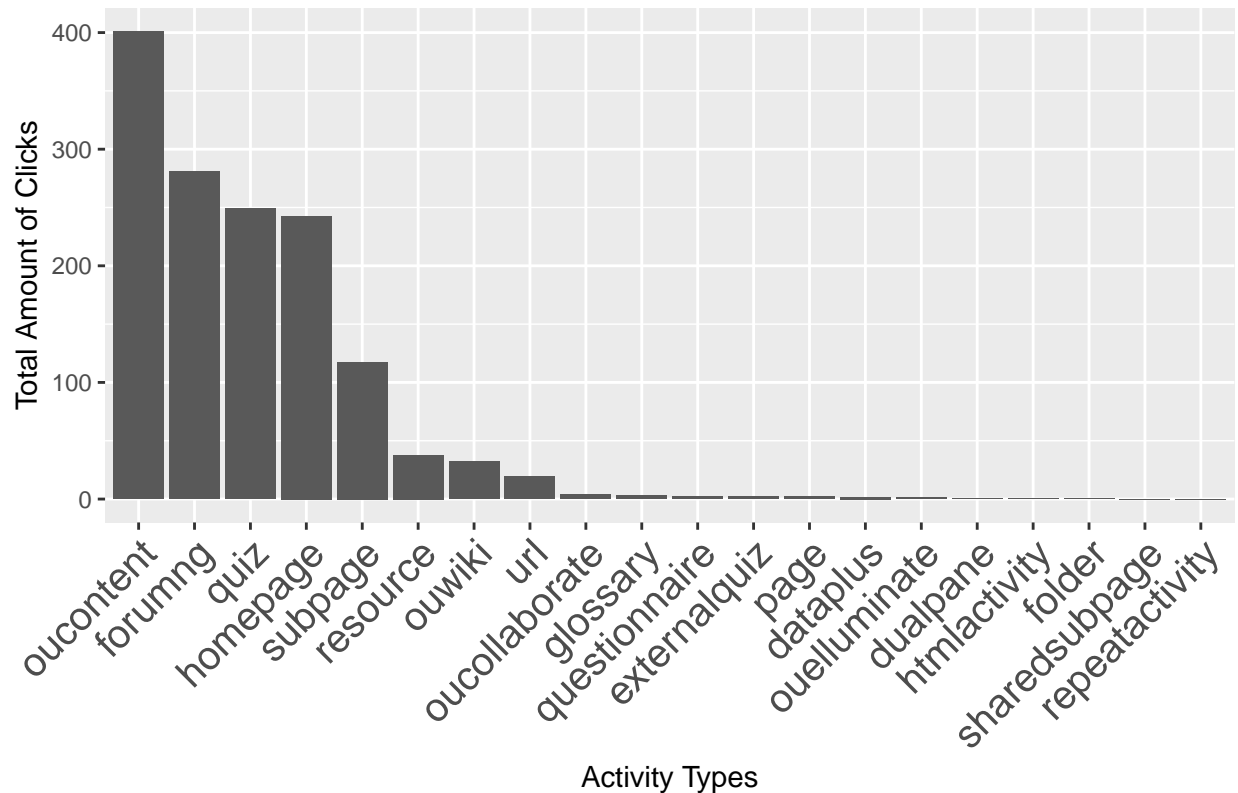

```
student_acti_drop$date_unregistration=
  as.numeric(levels(student_acti_drop$date_unregistration))[student_acti_drop$date_unregistration]
```

```
## Warning: NAs introduced by coercion
```

```
minisad=student_acti_drop%>%
  mutate(drop=ifelse(!is.na(date_unregistration),1,0))%>%
  group_by(id_student,code_module,code_presentation,activity_type)%>%
  summarise(total_click=sum(sum_click),drop=max(drop))
```

```
### Average clicks for Students who drop is lower than who does not
minisad%>%
  group_by(drop)%>%
  summarize(avgclick=mean(total_click))%>%
  ggplot(aes(x=drop,y=avgclick))+
  geom_bar(stat = 'identity')+
  scale_x_continuous(breaks = seq(0,1,1))+
  labs(x='Withdrawal=1',y='Average Clicks',
       title='Average Clicks Between Withdrawal and Complete')+
  scale_y_continuous(labels = comma)
```

Average Clicks Between Withdrawal and Complete
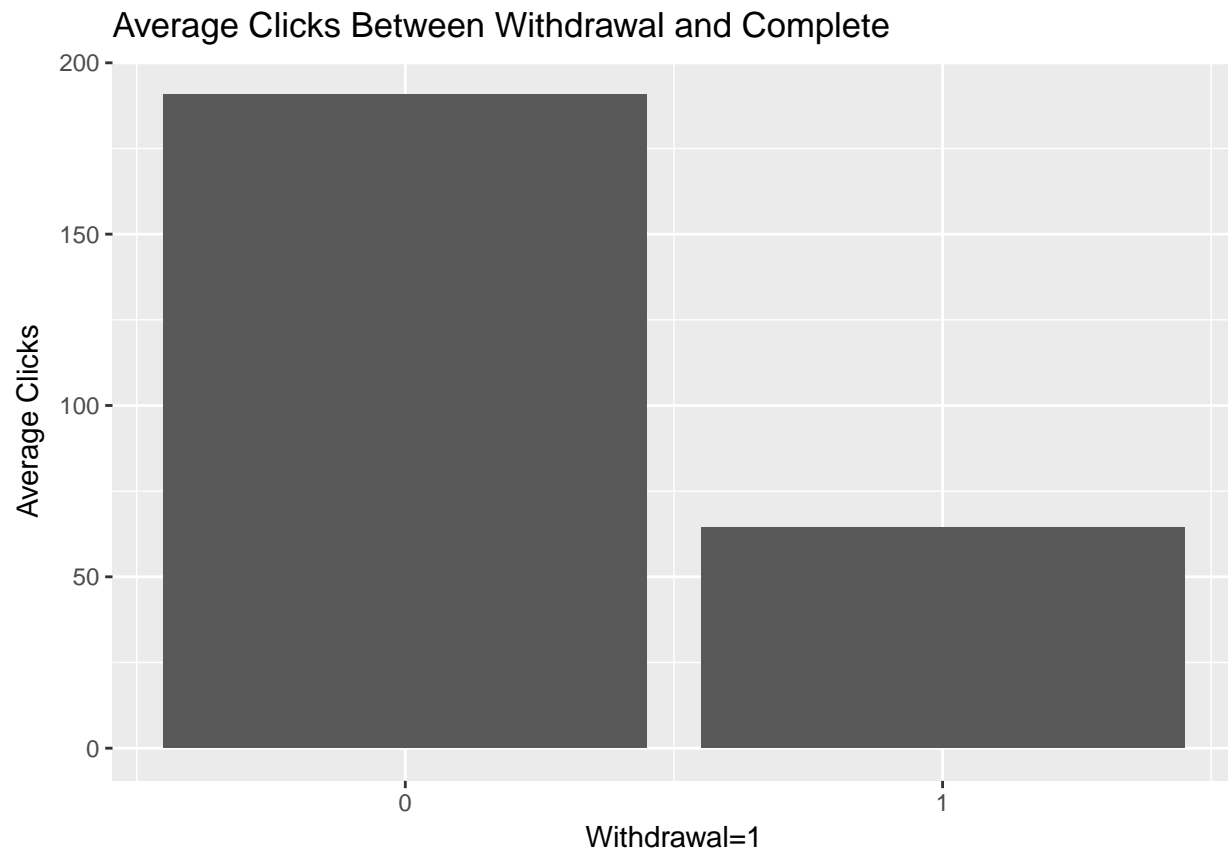

```
### The Average Clicks per Course Between Withdrawal and Complete
minisad%>%
  group_by(code_module,drop)%>%
  summarise(avgclick=mean(total_click))%>%
  ggplot(aes(x=drop,y=avgclick,fill=code_module))+
  geom_bar(stat='identity',position='dodge')+
  scale_x_continuous(breaks = seq(0,1,1))+
  scale_fill_manual(values=color)+
  guides(fill=guide_legend(title='Courses'))+
  labs(x='Withdrawal=1',y='Average Click per Course',
      title='The Average Clicks per Course Between Withdrawal and Complete')+
  guides(fill=guide_legend(title='Course'))+
  scale_y_continuous(labels = comma)
```

# The Average Clicks per Course Between Withdrawal and Complete
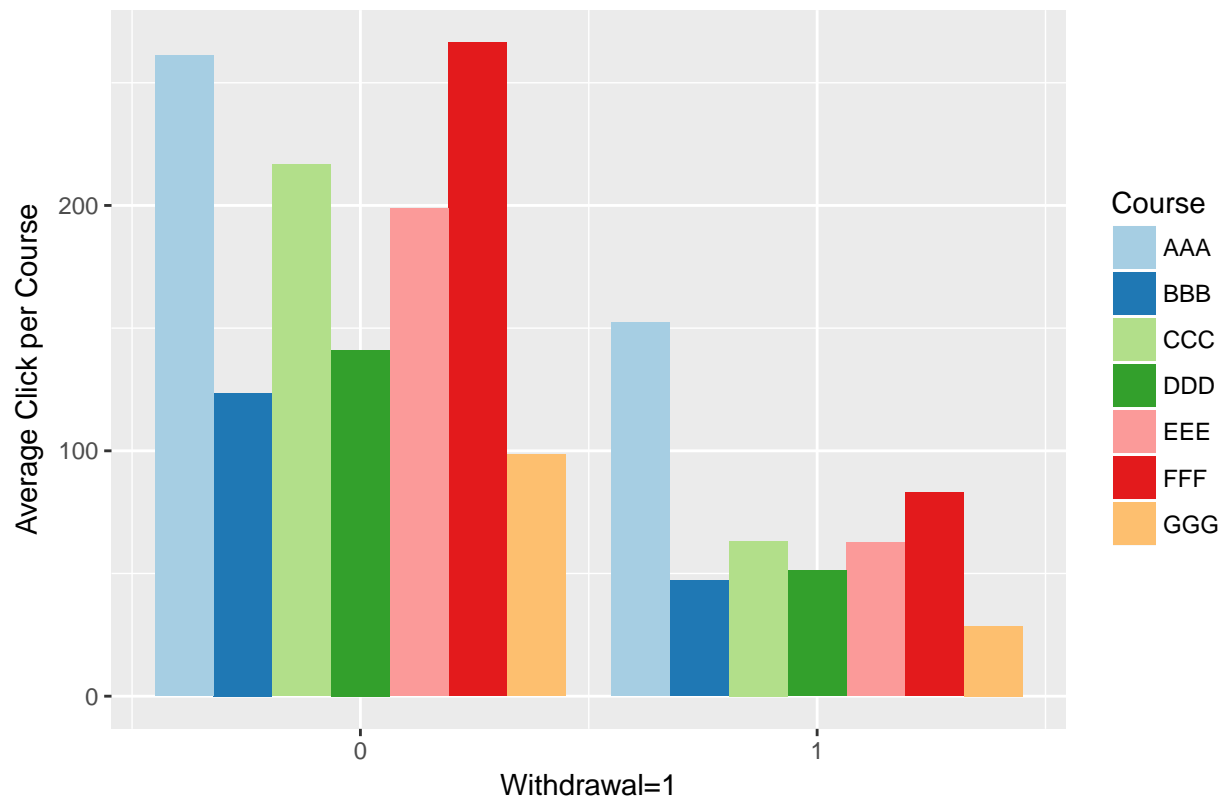
Updated 11/03

```
## Student Performance by Student Education Background
studentdrop %>%
  group_by(si_final_result,si_highest_education)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=reorder(si_final_result,-percent),y=percent,fill=si_highest_education))+
  geom_bar(stat='identity',position='dodge')+
  labs(x='Final Result',y='Percentage',
       title='Student Performance by Student Education Background')+
  guides(fill=guide_legend(title='Highest Education'))+
    scale_fill_manual(values=color)+
  theme(axis.text.x = element_text(size=12),
        axis.text.y=element_text(size=15),
        axis.text = element_text(size = 15),
        axis.title = element_text(size=20),
        legend.text = element_text(size=12))+
  scale_y_continuous(labels = percent)
```

Student Performance by Student Education Background
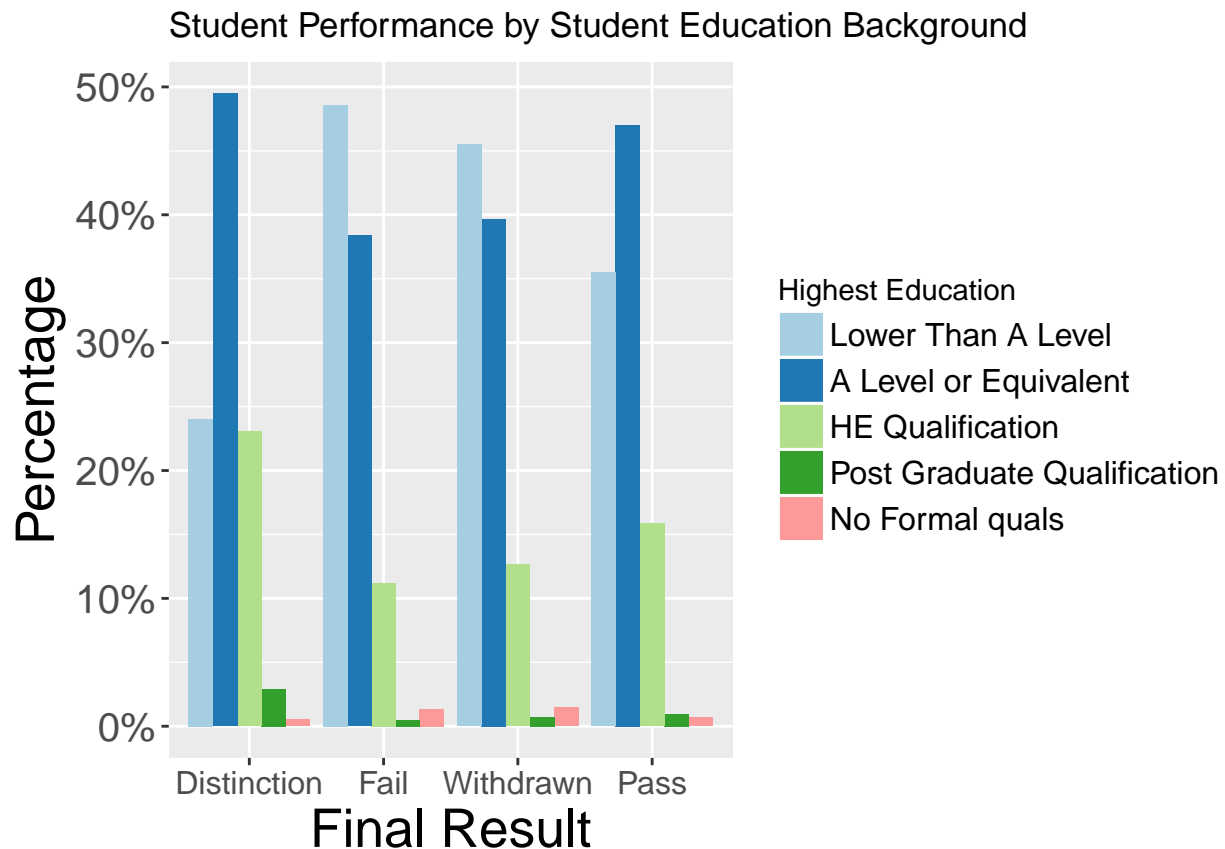

```r
## Student Education Background
studentdrop %>%
  group_by(si_highest_education)%>%
  summarize(counts=n())%>%
  ggplot(aes(x=reorder(si_highest_education,-counts),y=counts))+
  geom_bar(stat='identity')+
  labs(x='Final Result',y='Count',
       title='Student Education Background')+
    scale_fill_manual(values=color)+
  theme(axis.text.x = element_text(angle=15,hjust=1,size=12),
        axis.text.y=element_text(size=15),
        axis.text = element_text(size = 15),
        axis.title = element_text(size=20),
        legend.text = element_text(size=12))+
  scale_y_continuous(labels = comma)
```

Student Education Background
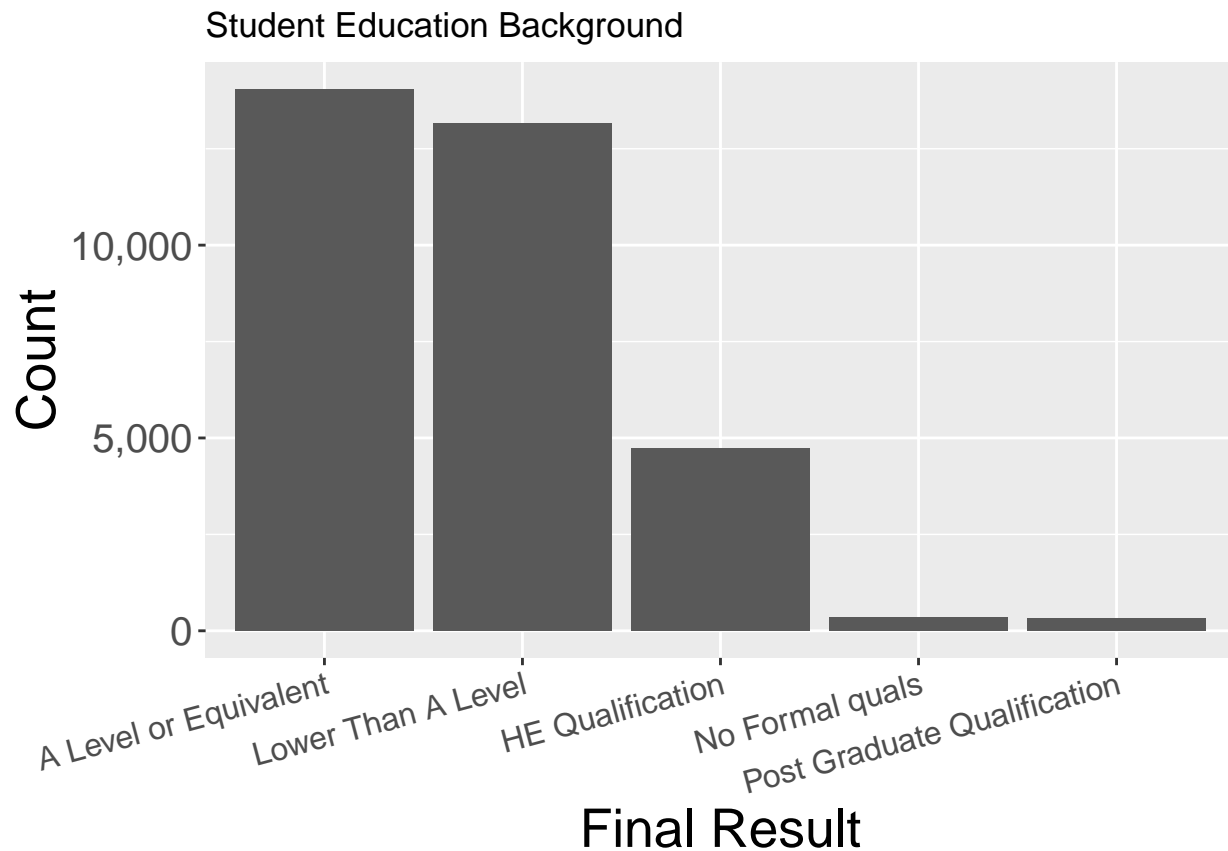

```
## Student Education Background by Student Performance
studentdrop %>%
  mutate(final_result=ifelse(si_final_result %in% c("Pass", "Distinction"), "Pass", as.character(si_fina
  filter(si_highest_education!="No Formal quals")%>%
  group_by(si_highest_education,final_result)%>%
  summarize(counts=n())%>%
  mutate(percent=counts/sum(counts))%>%
  ggplot(aes(x=final_result,y=percent,fill=si_highest_education))+
  geom_bar(stat='identity',position='dodge')+
  labs(x='Final Result',y='Percentage',
       title='Student Education Background by Student Performance')+
  guides(fill=guide_legend(title='Highest Education'))+
    scale_fill_manual(values=color)+
  theme(axis.text.x = element_text(size=12),
        axis.text.y=element_text(size=15),
        axis.text = element_text(size = 15),
        axis.title = element_text(size=20),
        legend.text = element_text(size=12))+
  scale_y_continuous(labels = percent)
```

Student Education Background by Student Performance
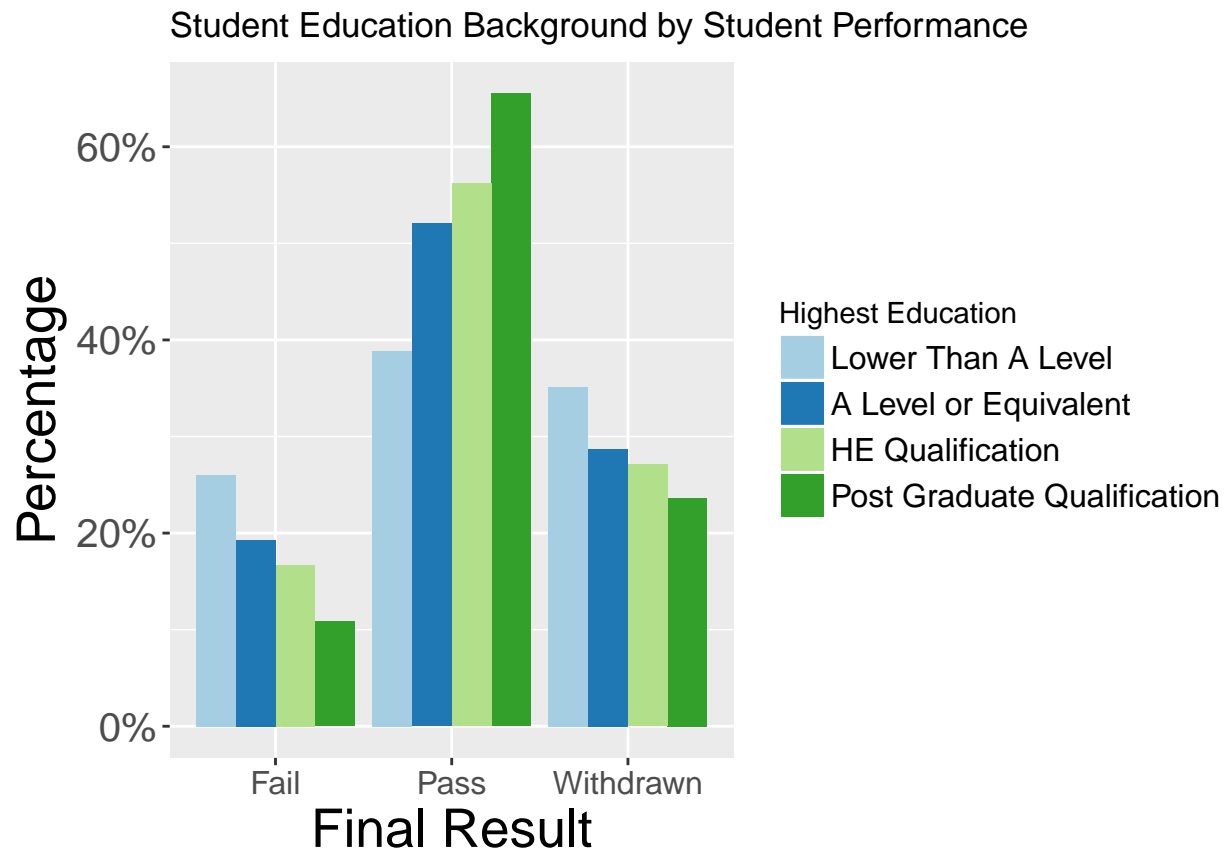

```
student_dailyacti_drop=read.csv("student_dailyacti_drop.csv")
##Average Daily Clicks by Activity Types for Withdrawal
student_dailyacti_drop%>%
  group_by(activity_type)%>%
  summarize(daily_avg=mean(daily_average))%>%
  ggplot(aes(x=reorder(activity_type,-daily_avg),y=daily_avg))+
  geom_bar(stat="identity")+
  theme(axis.text.x = element_text(size=15,angle=45,hjust=1))+
  scale_y_continuous(labels = comma)+
  labs(x='Activity Types',y='Average Daily Clicks',
       title='Average Daily Clicks by Activity Types for Withdrawal')
```
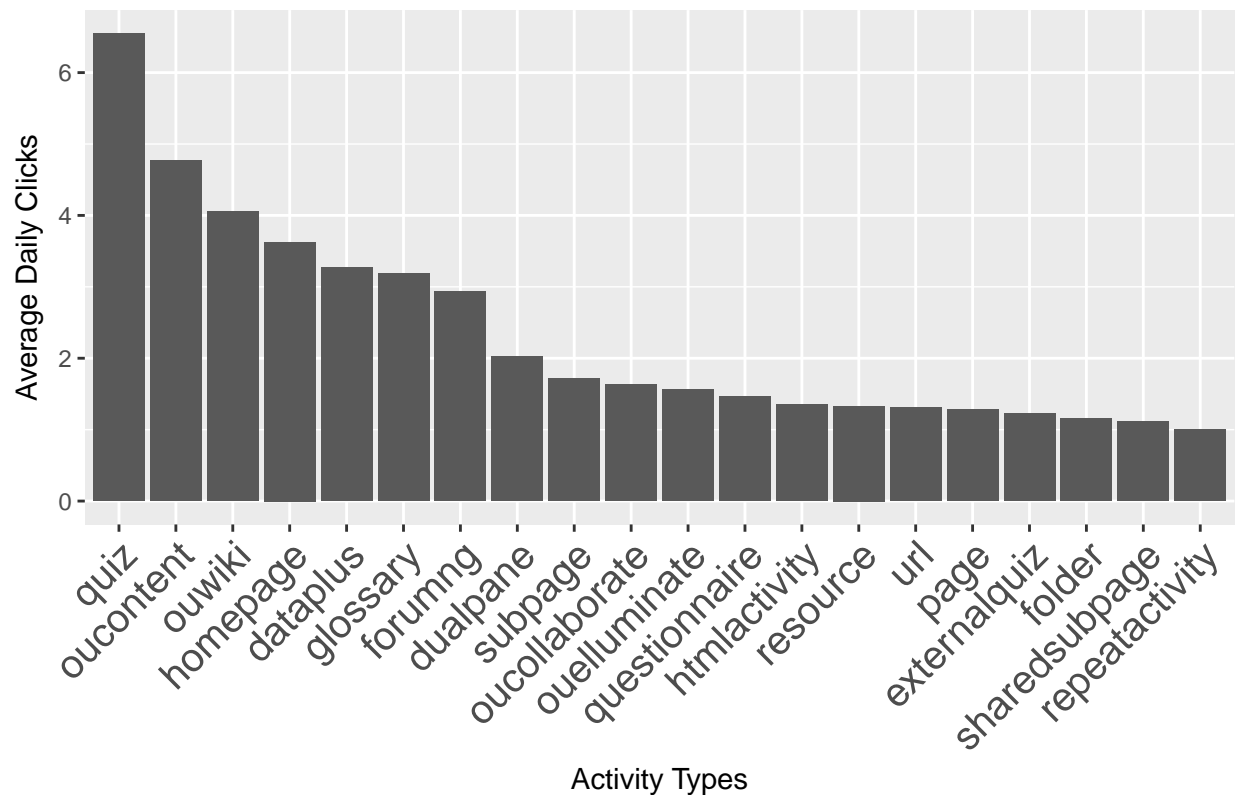
Average Daily Clicks by Activity Types for Withdrawal
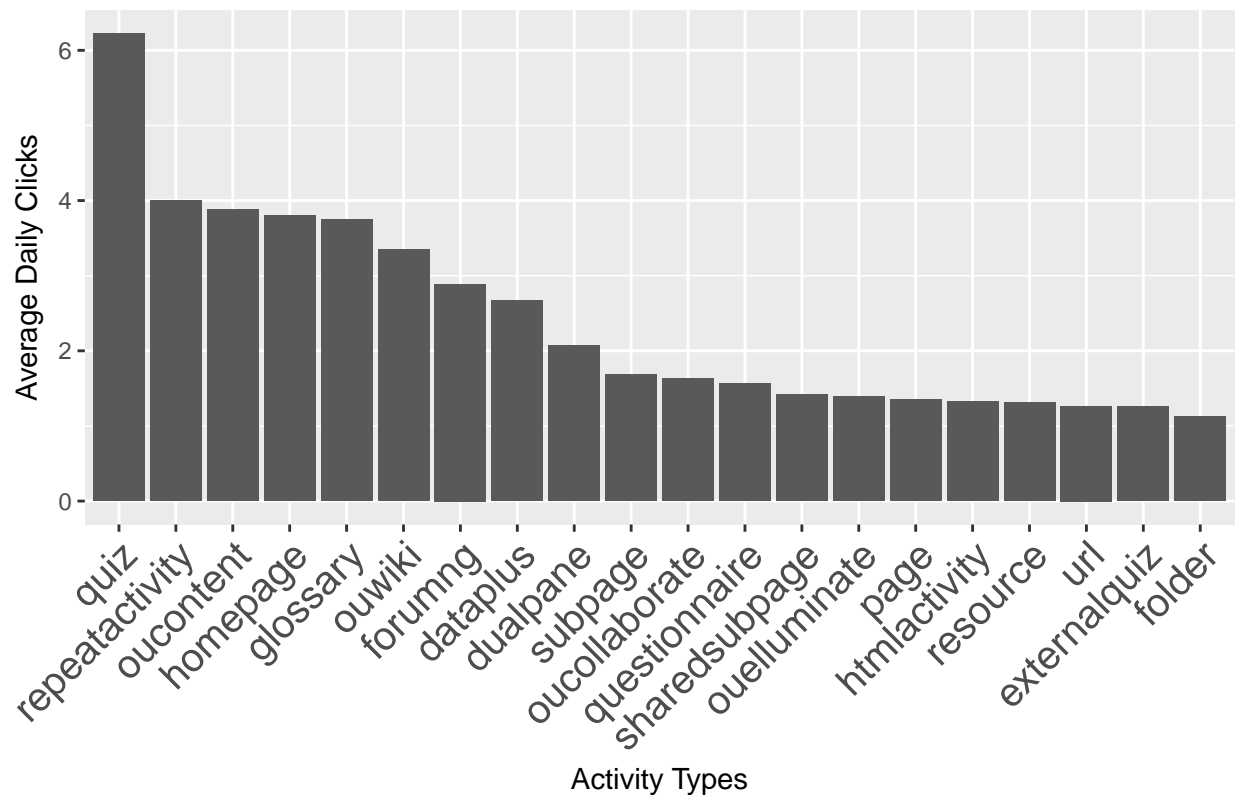

```
student_dailyacti_notdrop=read.csv("student_dailyacti_notdrop.csv")
##Average Daily Clicks by Activity Types for Complete
student_dailyacti_notdrop%>%
  group_by(activity_type)%>%
  summarize(daily_avg=mean(daily_average))%>%
  ggplot(aes(x=reorder(activity_type,-daily_avg),y=daily_avg))+
  geom_bar(stat="identity")+
  theme(axis.text.x = element_text(size=15,angle=45,hjust=1))+
  scale_y_continuous(labels = comma)+
  labs(x='Activity Types',y='Average Daily Clicks',
       title='Average Daily Clicks by Activity Types for Complete')
```

Average Daily Clicks by Activity Types for Complete

Average Daily Clicks (y-axis) vs Activity Types (x-axis): quiz, oucontent, ouwiki, homepage, dataplus, glossary, forumng, dualpane, subpage, oucollaborate, ouelluminate, questionnaire, htmlactivity, resource, url, page, externalquiz, folder, sharedsubpage, repeatactivity

```r
mean(student_dailyacti_drop$daily_average)
```

```
## [1] 2.70025
```

```r
mean(student_dailyacti_notdrop$daily_average)
```

```
## [1] 2.858956
```

```r
student_dailyacti_all=read.csv("student_dailyacti_all.csv")

student_dailyacti_all%>%
  group_by(final_result)%>%
  summarize(daily_avg=mean(daily_average))%>%
  ggplot(aes(x=reorder(final_result,-daily_avg),y=daily_avg))+
  geom_bar(stat="identity")+
  theme(axis.text.x = element_text(size=15))+
  labs(x='Final Result',y='Average Daily Clicks',
       title='Average Daily Clicks by Final Result for Complete')
```

Average Daily Clicks by Final Result for Complete