Socialisation Graphs

Aditya

2018

R Markdown

```
library(ggplot2)
data = read.csv("data/Student_data_standard.csv")
```

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.2.1 ──
```

```
## ✔ tibble  1.4.2     ✔ purrr   0.2.4
## ✔ tidyr   0.8.0     ✔ dplyr   0.7.4
## ✔ readr   1.1.1     ✔ stringr 1.3.0
## ✔ tibble  1.4.2     ✔ forcats 0.3.0
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
summary(data)
```

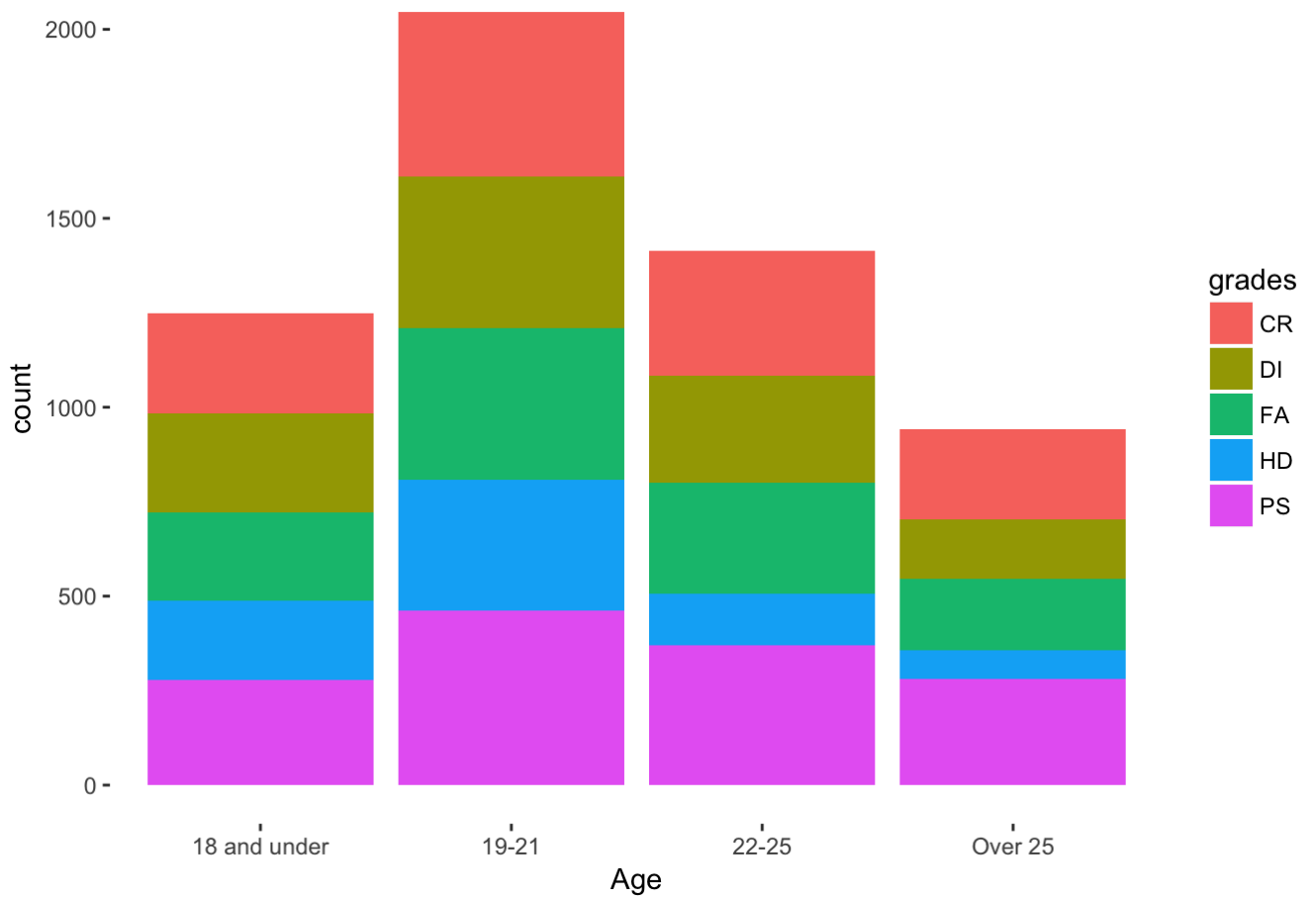
```
##       Year      Domestic.Intl Gender          Mode
##  Min.   :2012   D:3814        F:2665   Full time:4310
##  1st Qu.:2013   I:1838        M:2987   Part time:1342
##  Median :2015
##  Mean   :2015
##  3rd Qu.:2016
##  Max.   :2017
##
##            Age       Unit.of.Study   Unit.of.Study.Level
##  18 and under:1248   Unit E : 575   Advanced   : 893
##  19-21       :2047   Unit B : 548   Fundamental:2195
##  22-25       :1414   Unit A : 521   Mainstream :2564
##  Over 25     : 943   Unit C : 509
##                      Unit H : 495
##                      Unit G : 475
##                      (Other):2529
##  Unit.of.Study.Grade     Count
##  CR:1273             Min.   :  1.00
##  DI:1103             1st Qu.:  1.00
##  FA:1115             Median :  3.00
##  HD: 771             Mean   : 11.41
##  PS:1390             3rd Qu.: 10.00
##                      Max.   :282.00
##
```

```
str(data)
```

```
## 'data.frame':    5652 obs. of  9 variables:
##  $ Year               : int  2013 2012 2013 2013 2012 2014 2012 2012 2016 2014 ...
##  $ Domestic.Intl      : Factor w/ 2 levels "D","I": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Gender             : Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Mode               : Factor w/ 2 levels "Full time","Part time": 1 1 1 1 1 1 1
1 1 1 ...
##  $ Age                : Factor w/ 4 levels "18 and under",..: 2 2 2 2 2 2 2 2 2 2
...
##  $ Unit.of.Study      : Factor w/ 14 levels "Unit A","Unit B",..: 3 5 2 1 1 3 2 3
2 5 ...
##  $ Unit.of.Study.Level: Factor w/ 3 levels "Advanced","Fundamental",..: 3 3 3 3 3
3 3 3 3 3 ...
##  $ Unit.of.Study.Grade: Factor w/ 5 levels "CR","DI","FA",..: 5 5 5 5 5 5 5 5 5 5
...
##  $ Count              : int  282 263 252 233 228 220 219 207 192 186 ...
```

does age influence study grade?

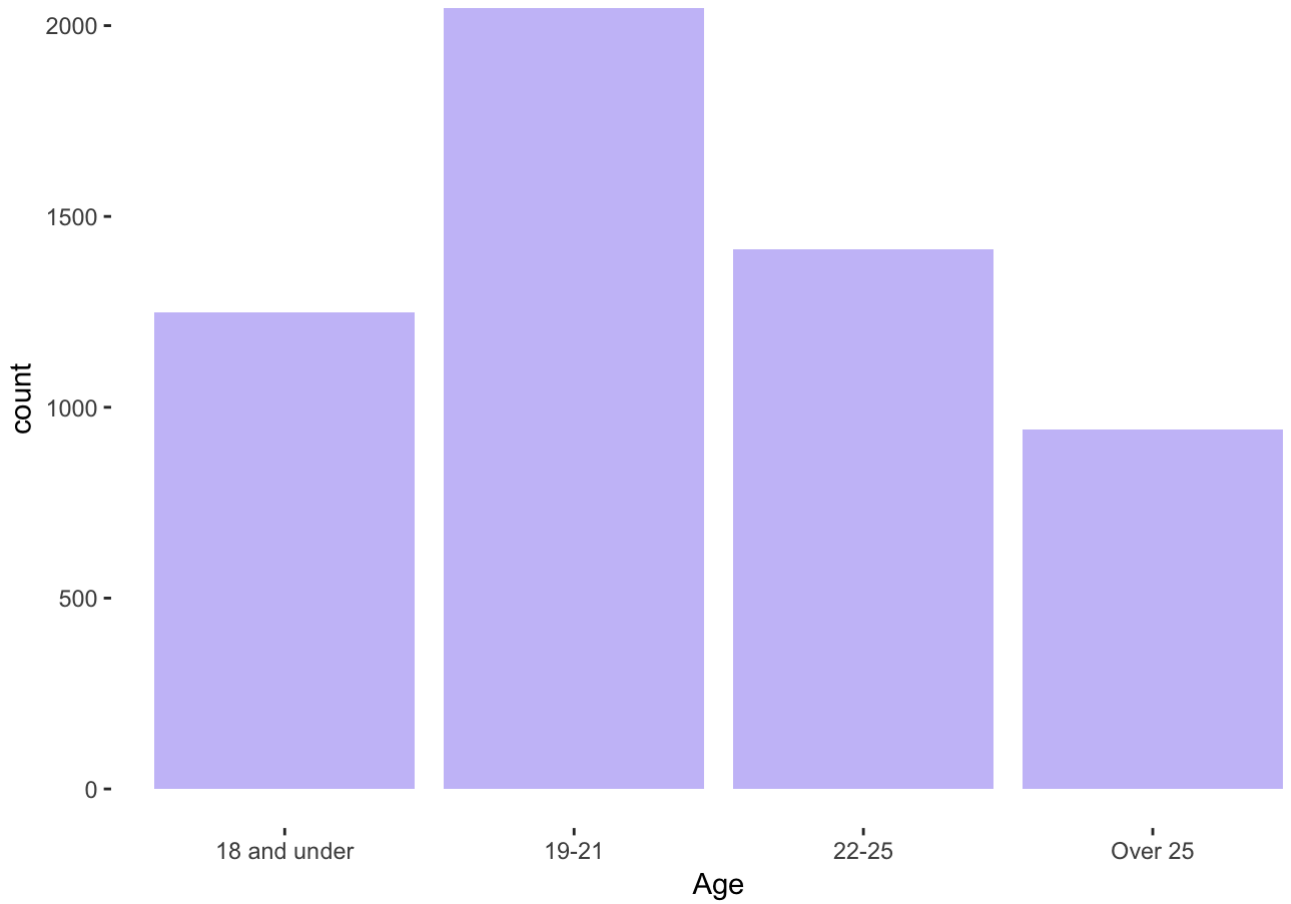
```
grades = data$Unit.of.Study.Grade
ggplot(data=data) +
  geom_bar(mapping=aes(x=data$Age, fill = grades ))+
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Age")
```



```
table( data$Age)
```

```
##
## 18 and under        19-21        22-25      Over 25
##         1248         2047         1414          943
```

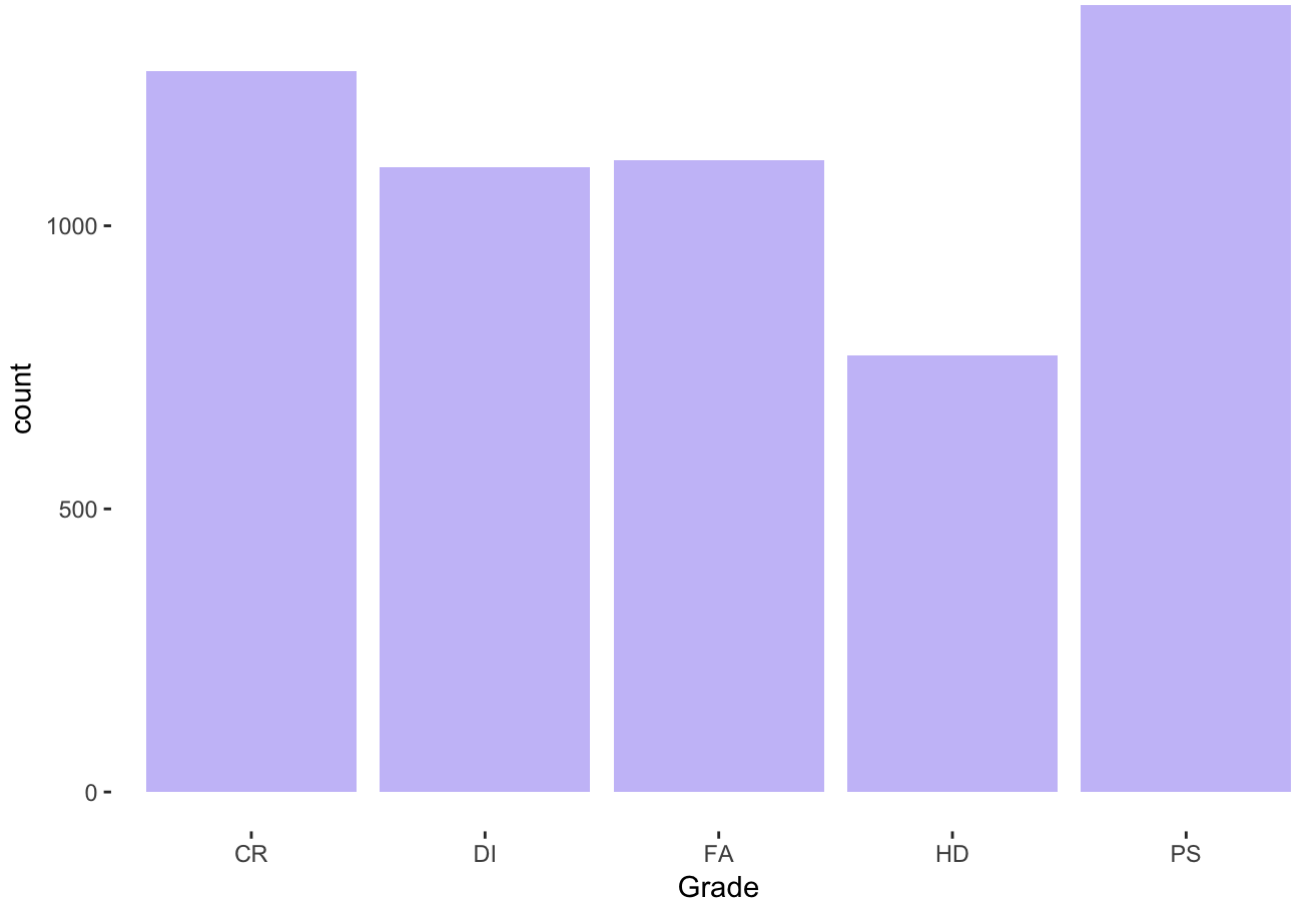
```
ggplot(data=data) +
  geom_bar(mapping=aes(x=data$Age), fill = "slateblue2", alpha = 3/7) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Age")
```




```
table(data$Unit.of.Study.Grade)
```

```
##
##   CR   DI   FA   HD   PS
## 1273 1103 1115  771 1390
```

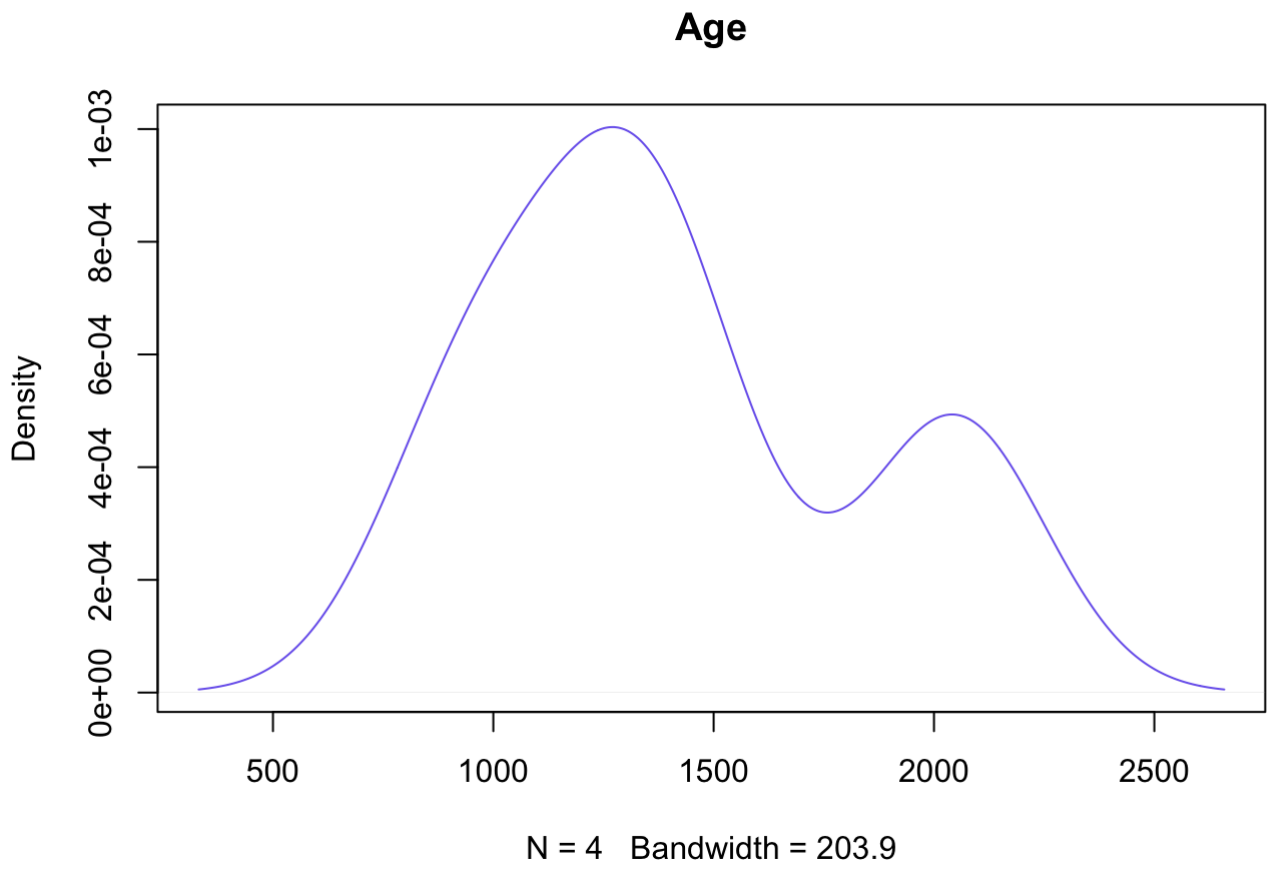
```
ggplot(data=data) +
  geom_bar(mapping=aes(x=data$Unit.of.Study.Grade), fill = "slateblue2", alpha = 3/7)
+
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Grade")
```



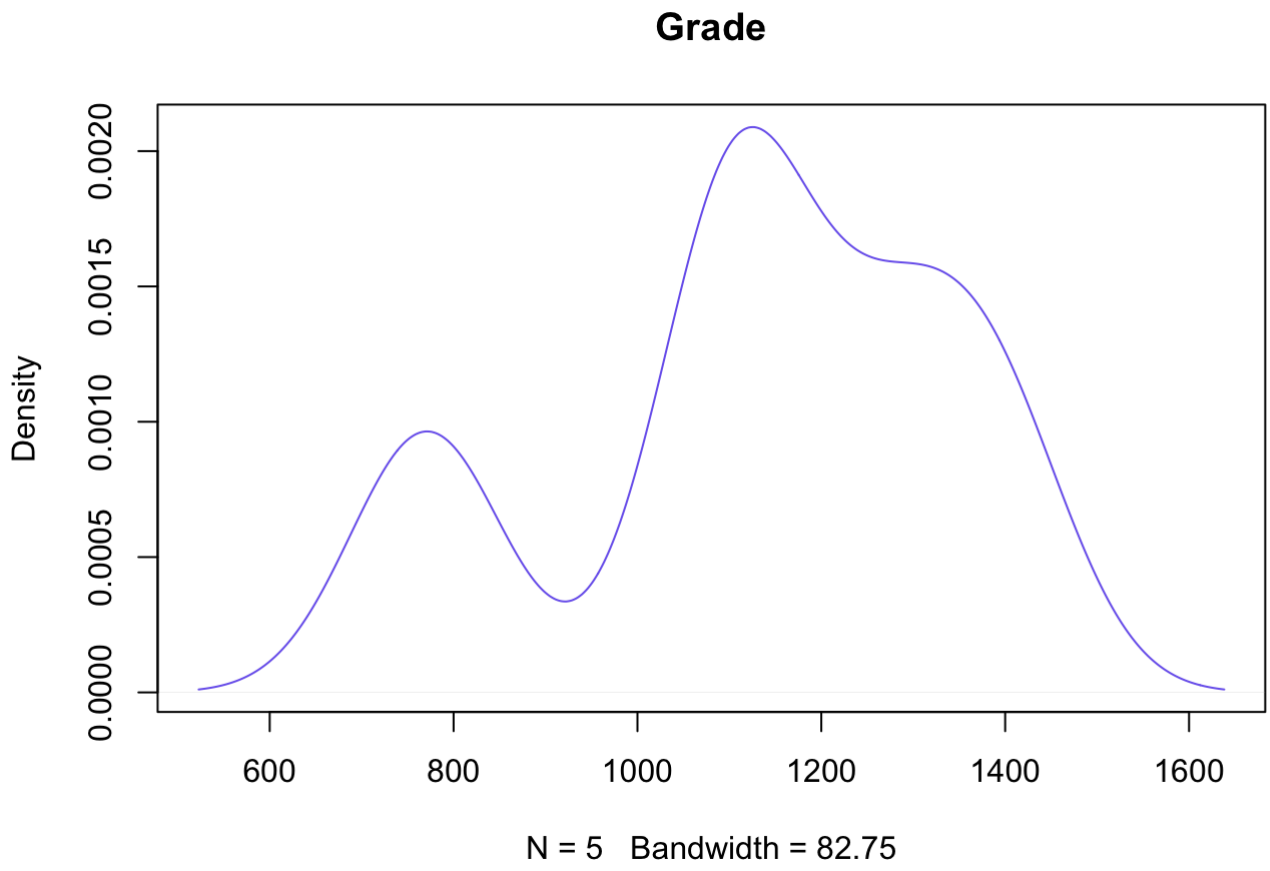

```
table1 = table(data$Age, data$Unit.of.Study.Grade)
table1
```

```
##
##                 CR  DI  FA  HD  PS
##   18 and under 265 262 232 210 279
##   19-21        437 400 401 348 461
##   22-25        330 285 293 136 370
##   Over 25      241 156 189  77 280
```

```
table = table(data$Age)
d <- density(table, na.rm = T)
plot(d, main = "Age", col = "slateblue2")
```




```
table = table(data$Unit.of.Study.Grade)
d <- density(table, na.rm = T)
plot(d, main = "Grade", col = "slateblue2")
```



Is the mode (fulltime/partime) influenced by whether you are a domestic or international student?

```
table1 = table(data$Mode)
table1
```

```
##
## Full time Part time
##      4310      1342
```

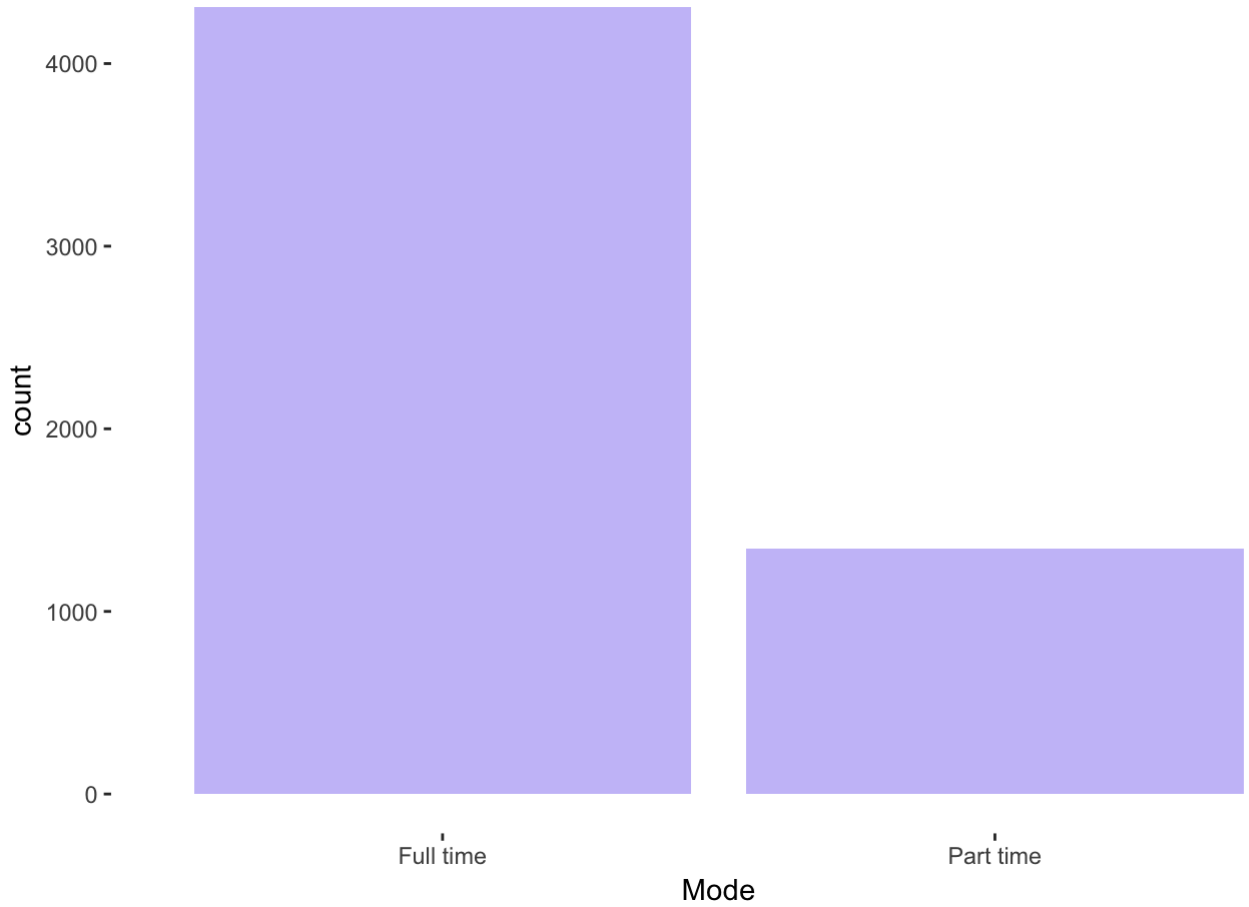
```
table2 = table(data$Domestic.Intl)
table2
```

```
##
##    D    I
## 3814 1838
```

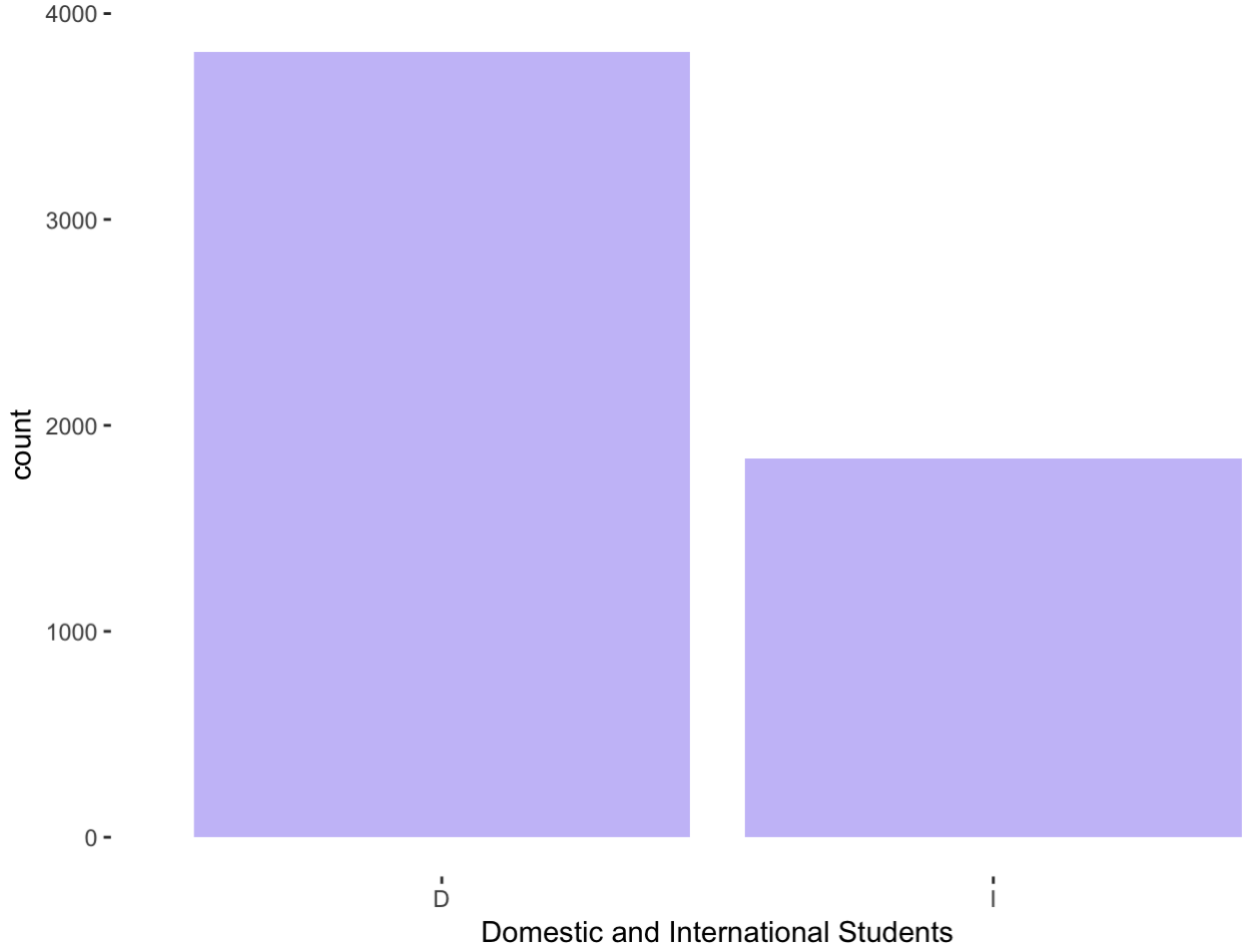
```
table3 = table(data$Domestic.Intl, data$Mode)
table3
```

```
##
##     Full time Part time
##   D      2634      1180
##   I      1676       162
```

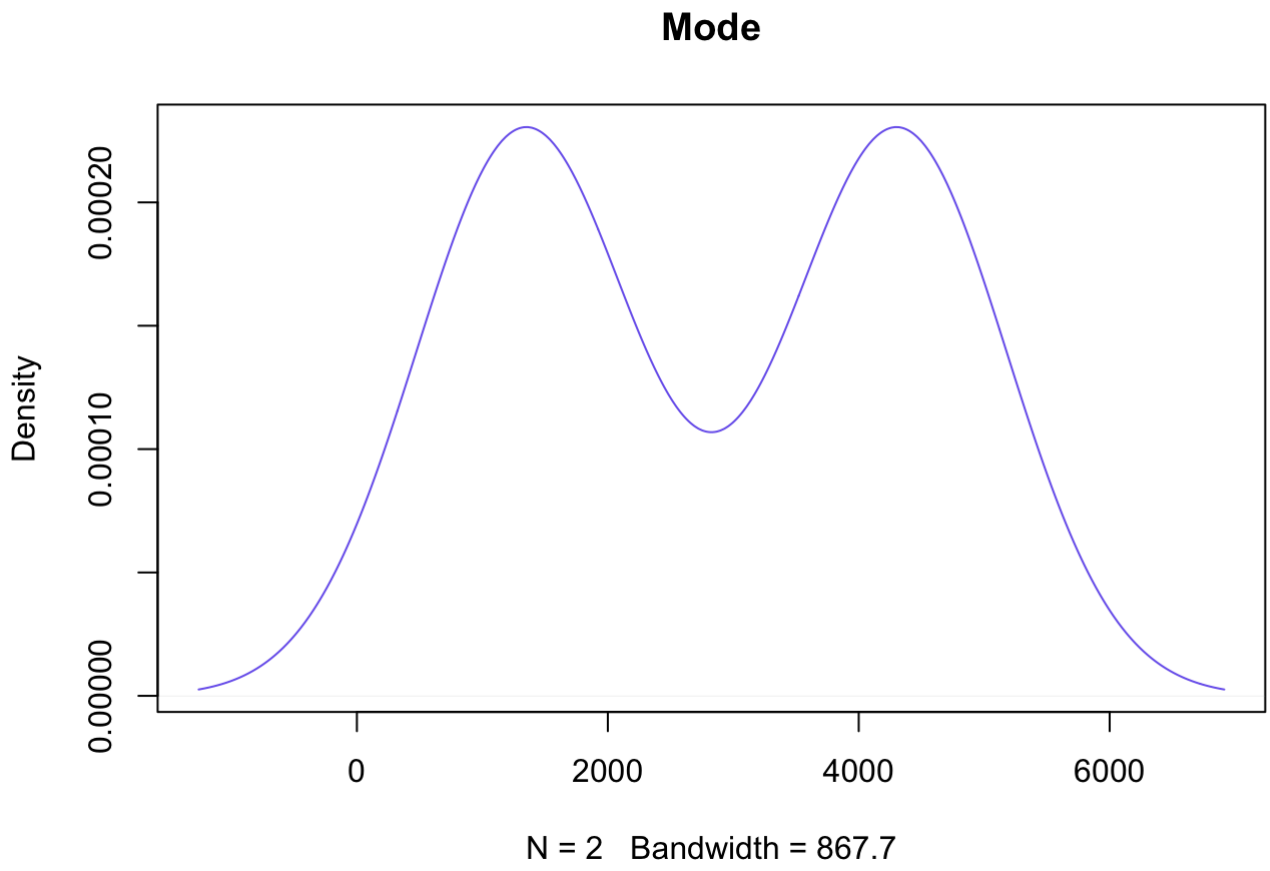
```
ggplot(data=data) +
  geom_bar(mapping=aes(x=data$Mode), fill = "slateblue2", alpha = 3/7) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Mode")
```



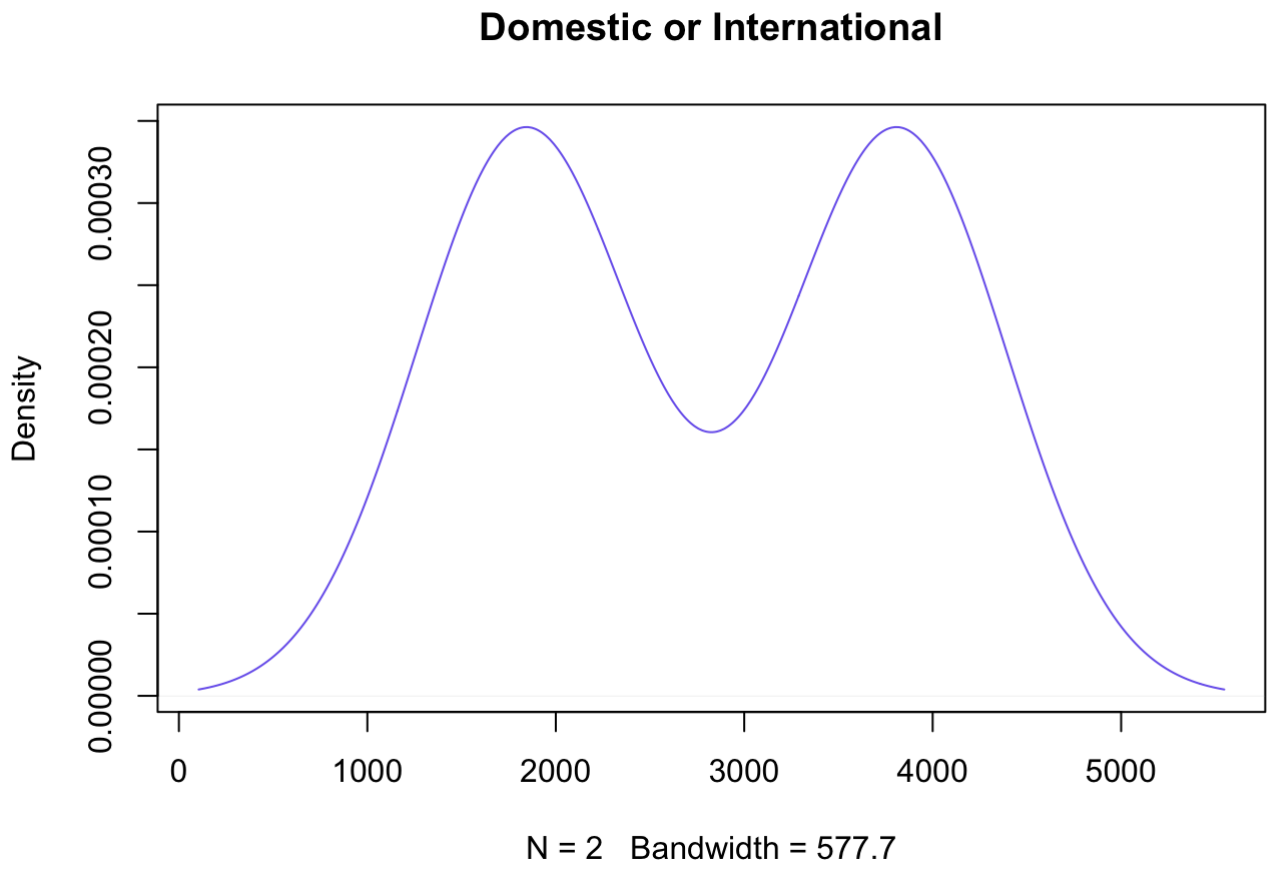

```
ggplot(data=data) +
  geom_bar(mapping=aes(x=data$Domestic.Intl), fill = "slateblue2", alpha = 3/7) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Domestic and International Students")
```



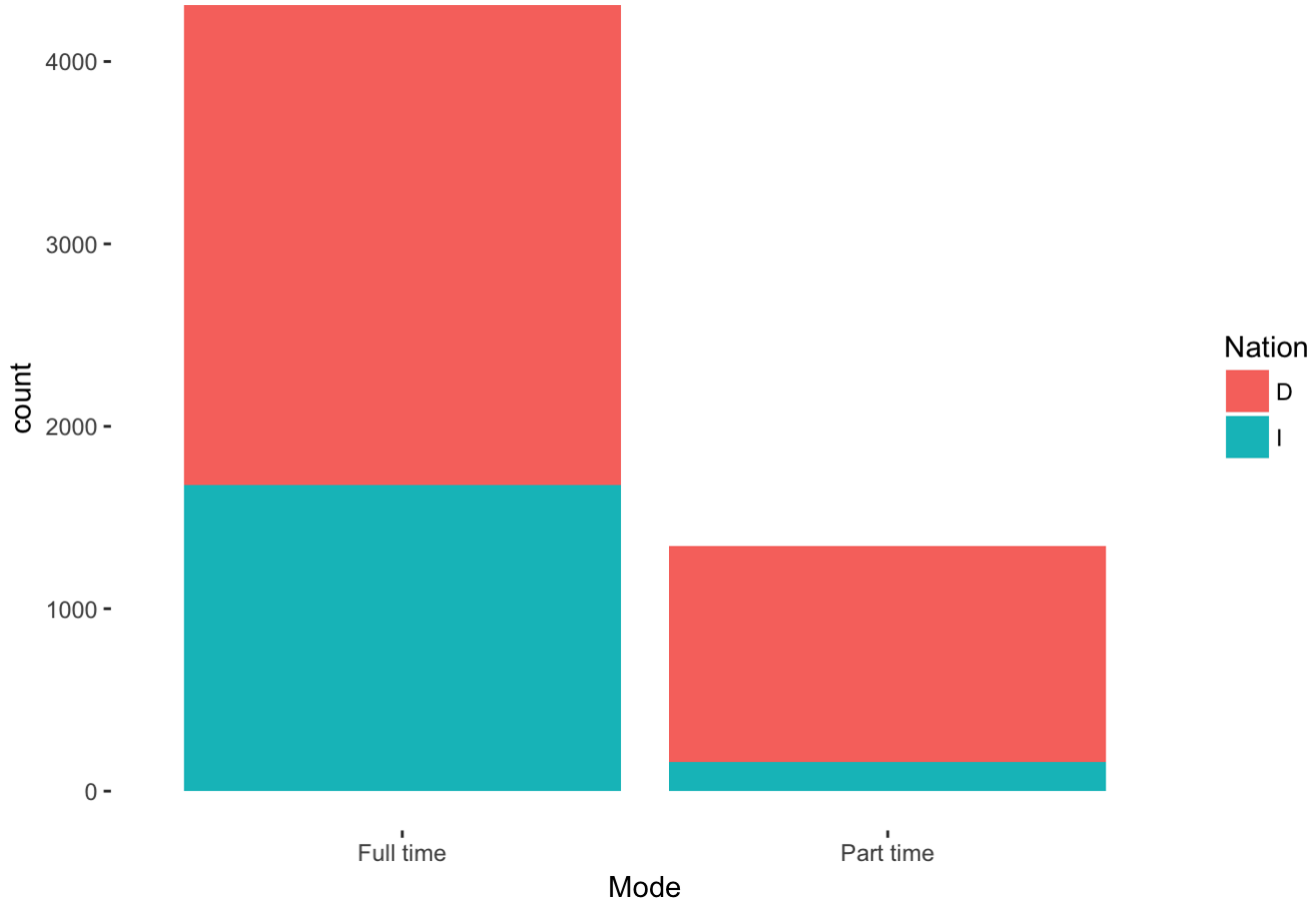

```
table = table(data$Mode)
d <- density(table, na.rm = T)
plot(d, main = "Mode", col = "slateblue2")
```




```
table = table(data$Domestic.Intl)
d <- density(table, na.rm = T)
plot(d, main = "Domestic or International", col = "slateblue2")
```




```
Nation = data$Domestic.Intl
ggplot(data=data) +
  geom_bar(mapping=aes(x=data$Mode, fill = Nation  ))+
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Mode")
```

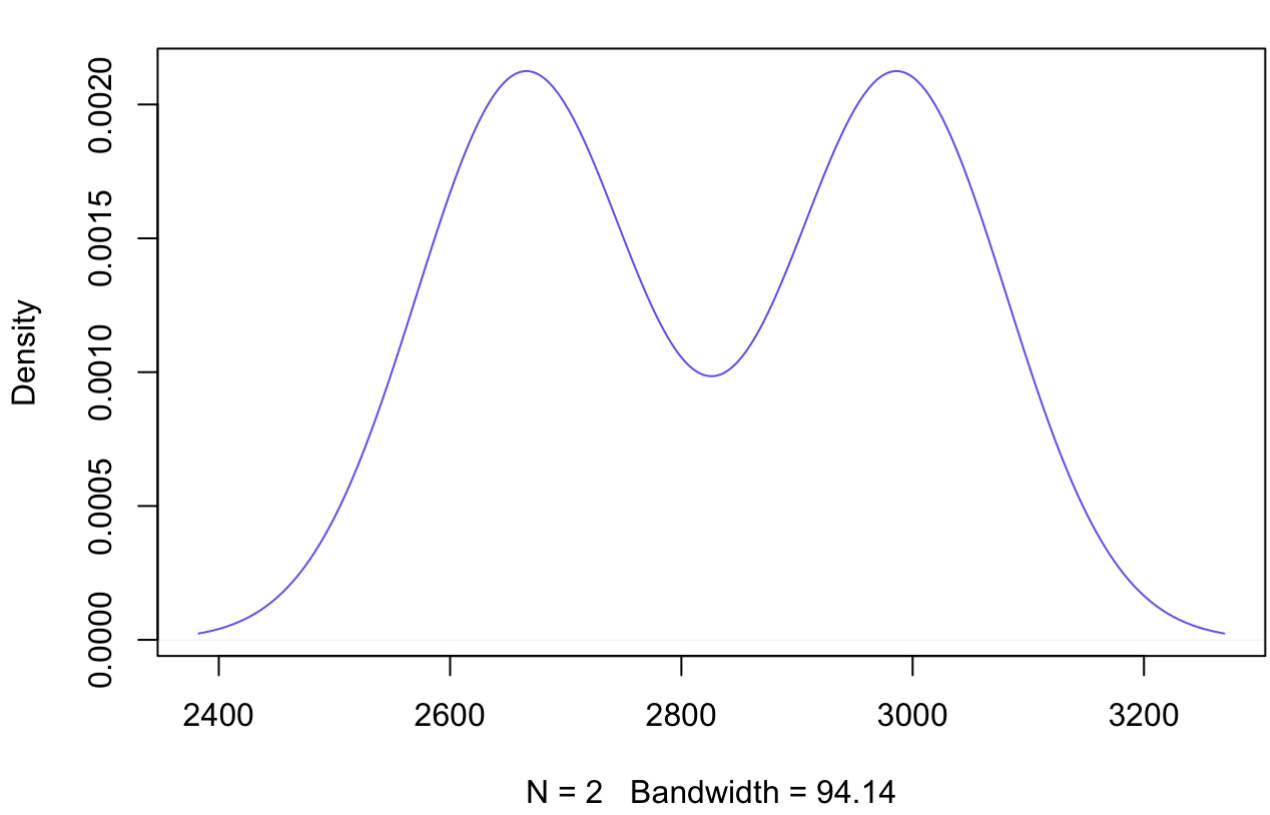


are there correlations between gender and study grade?

```
table1 = table(data$Gender)
table1
```

```
##
##    F    M
## 2665 2987
```

```
table = table(data$Gender)
d <- density(table, na.rm = T)
plot(d, main = "Gender", col = "slateblue2")
```



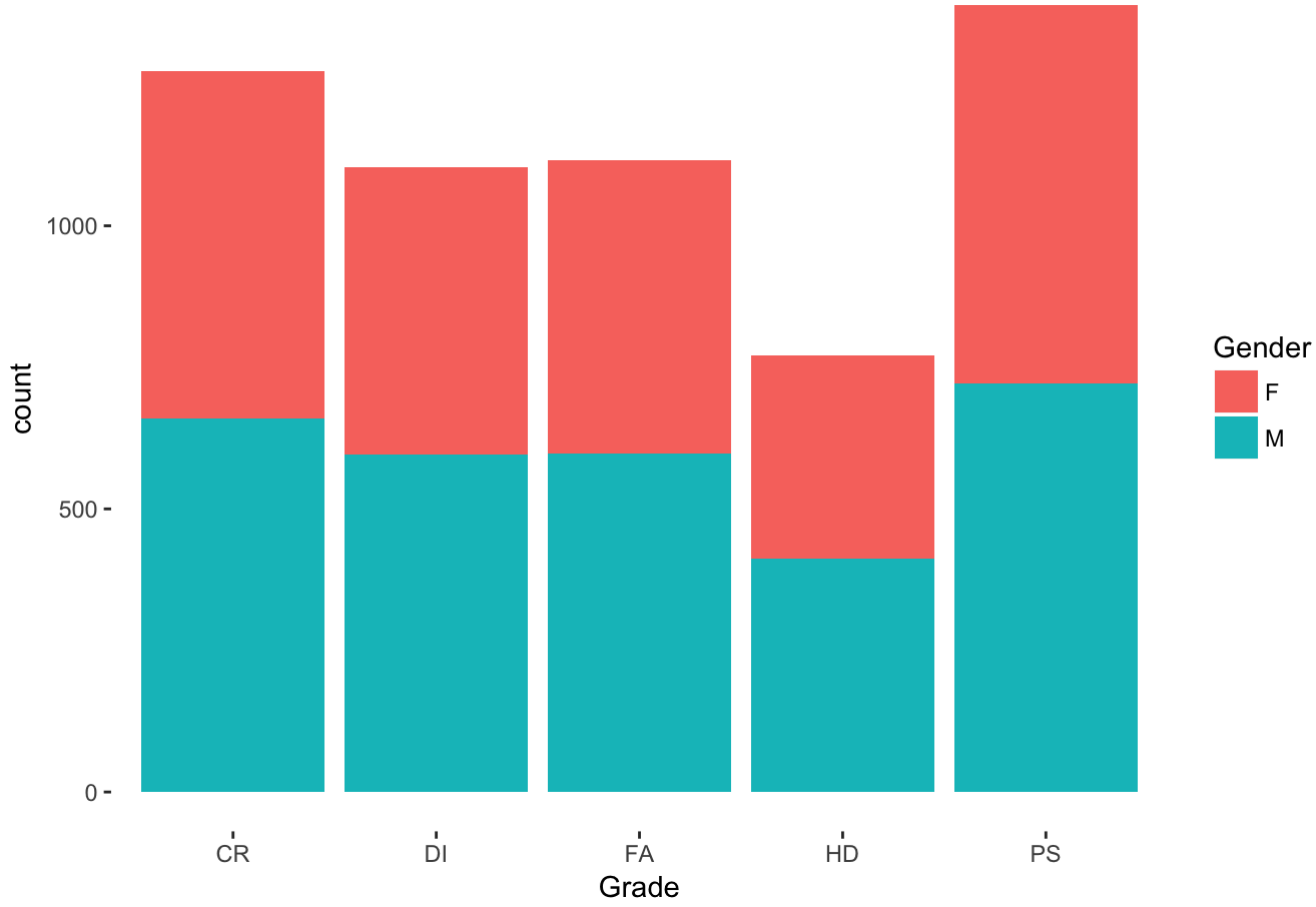

```
table2 = table(data$Unit.of.Study.Grade)
table2
```

```
##
##   CR   DI   FA   HD   PS
## 1273 1103 1115  771 1390
```

```
table(data$Gender, data$Unit.of.Study.Grade)
```

```
##
##      CR  DI  FA  HD  PS
##   F 614 507 518 358 668
##   M 659 596 597 413 722
```

```
Gender = data$Gender
ggplot(data=data) +
  geom_bar(mapping=aes(x=data$Unit.of.Study.Grade, fill = Gender  ))+
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Grade")
```



does unit of study level impact count?

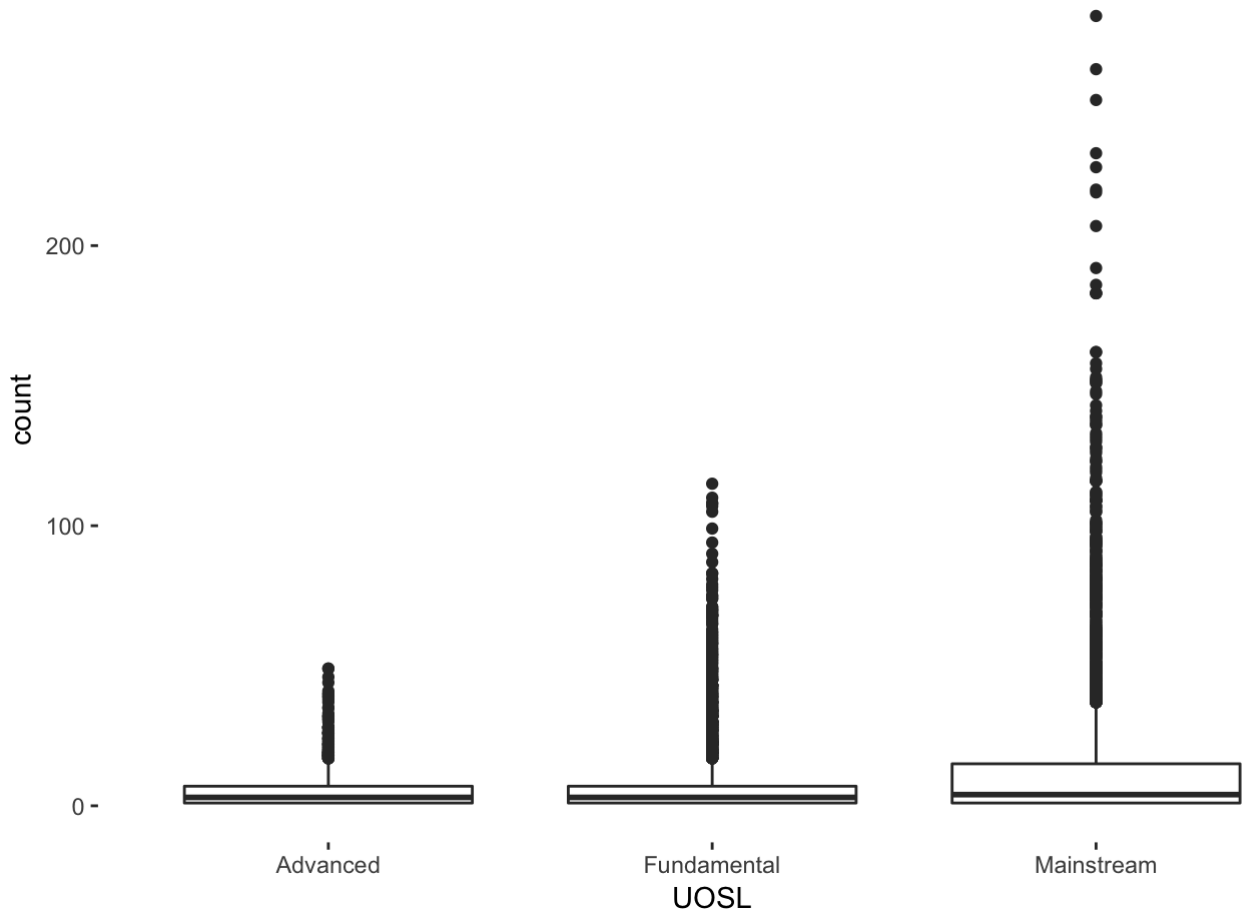
```
table(data$Unit.of.Study.Level)
```

```
##
##    Advanced Fundamental  Mainstream
##         893        2195        2564
```

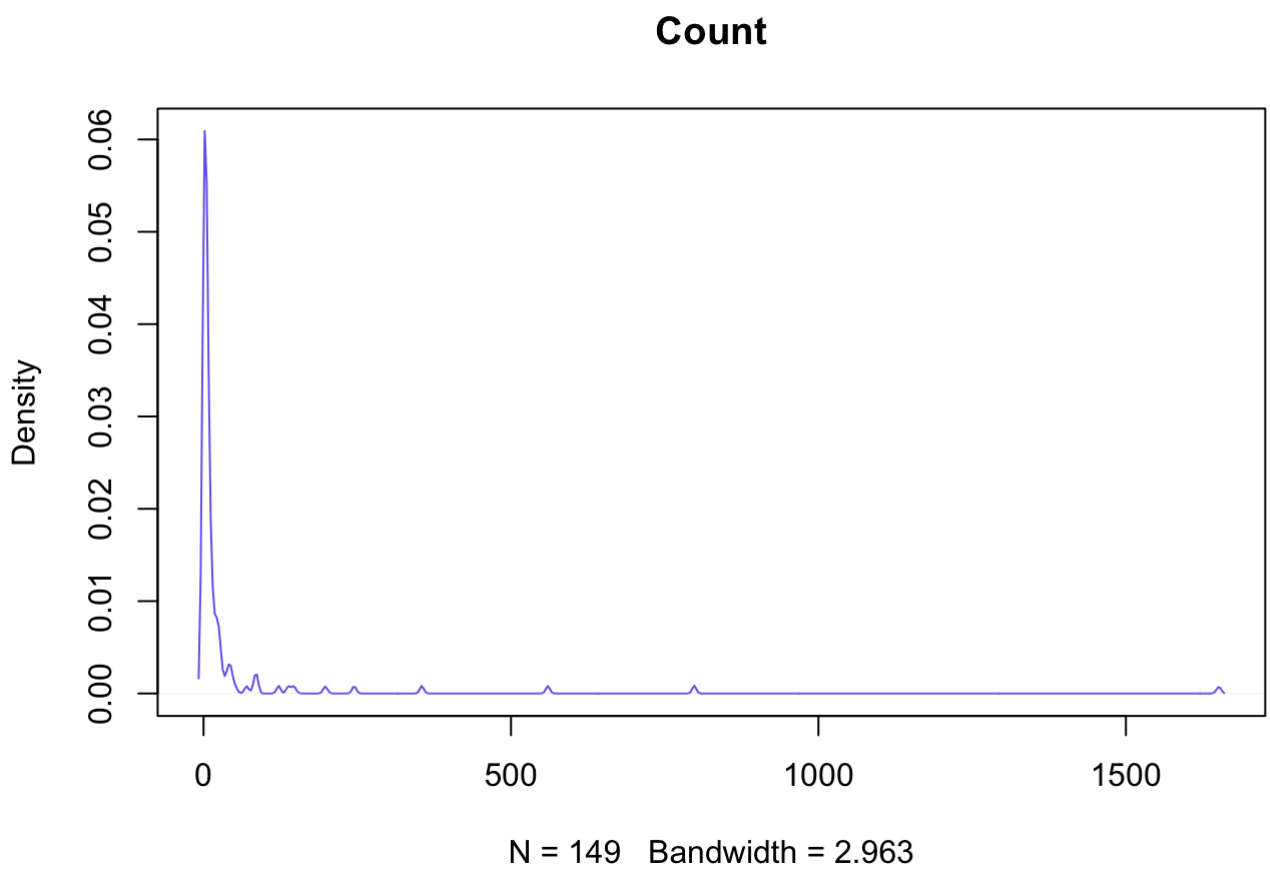
```
summary(data$Count)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    1.00    3.00   11.41   10.00  282.00
```

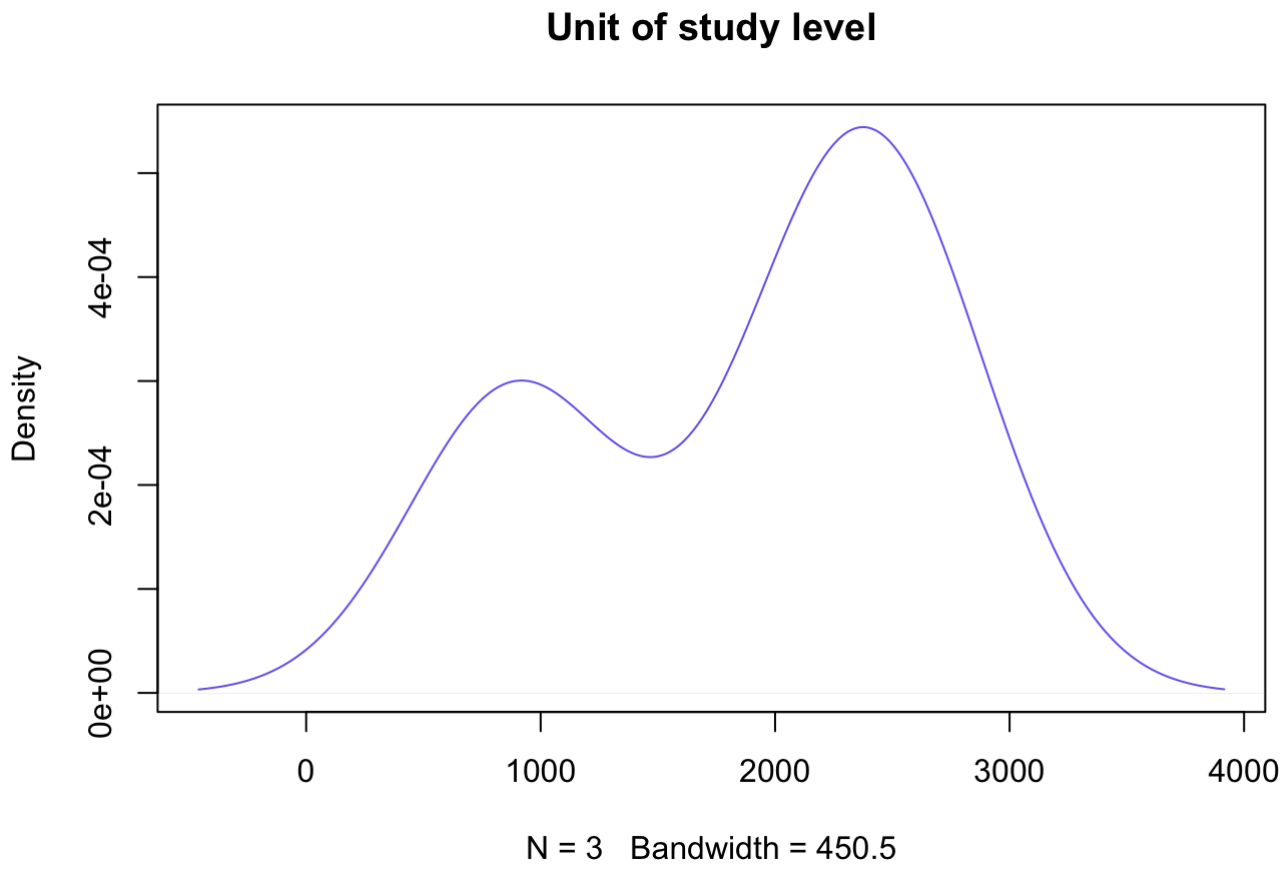
```
count = data$Count
UOSL = data$Unit.of.Study.Level
ggplot(data=data) +
  geom_boxplot(mapping=aes(x=UOSL, y=count))+
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank())
```




```
table = table(data$Count)
d <- density(table, na.rm = T)
plot(d, main = "Count", col = "slateblue2")
```




```
table = table(data$Unit.of.Study.Level)
d <- density(table, na.rm = T)
plot(d, main = "Unit of study level", col = "slateblue2")
```


Unit of study level
Density
0e+00
2e-04
4e-04
0
1000
2000
3000
4000
N = 3 Bandwidth = 450.5