dataAnalysisOCS

OCS Team

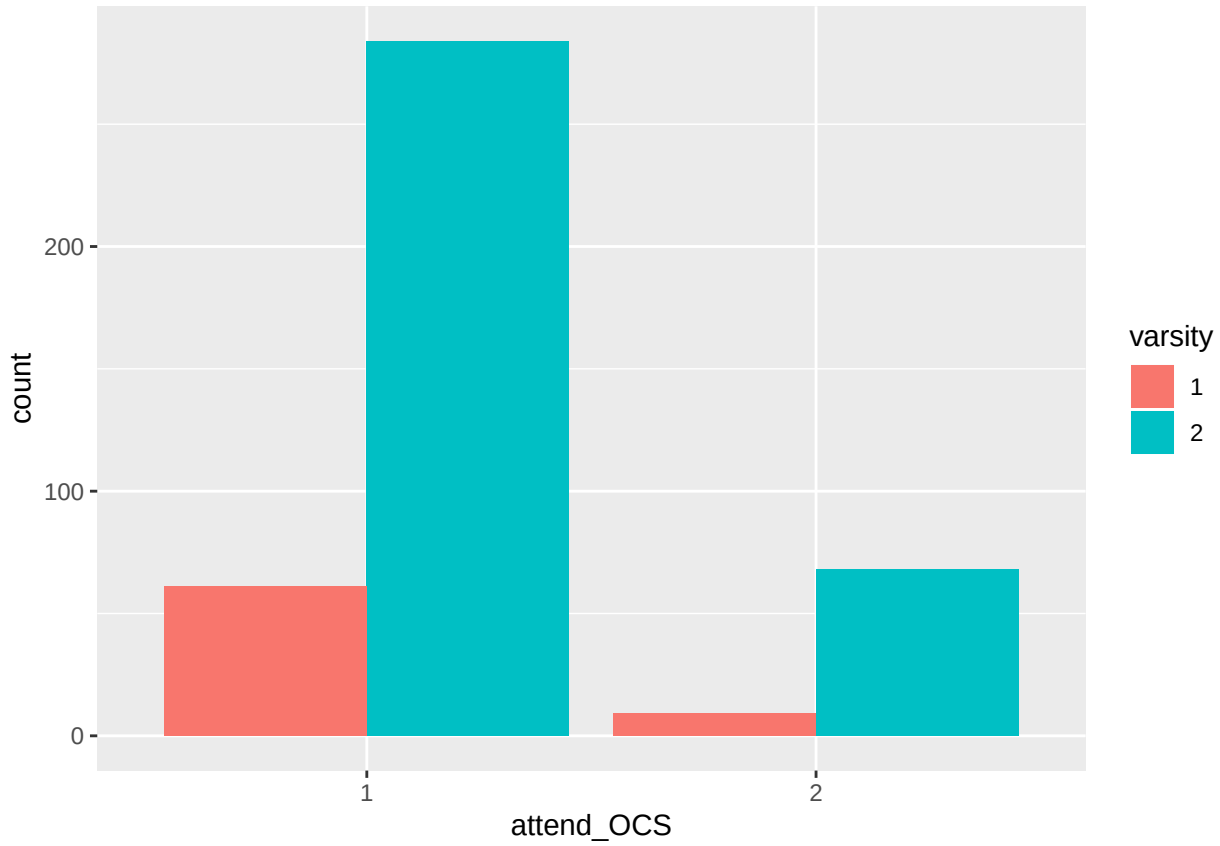
5/22/2020

```
ocs <- read_csv("ocs_data_may22.csv")
```

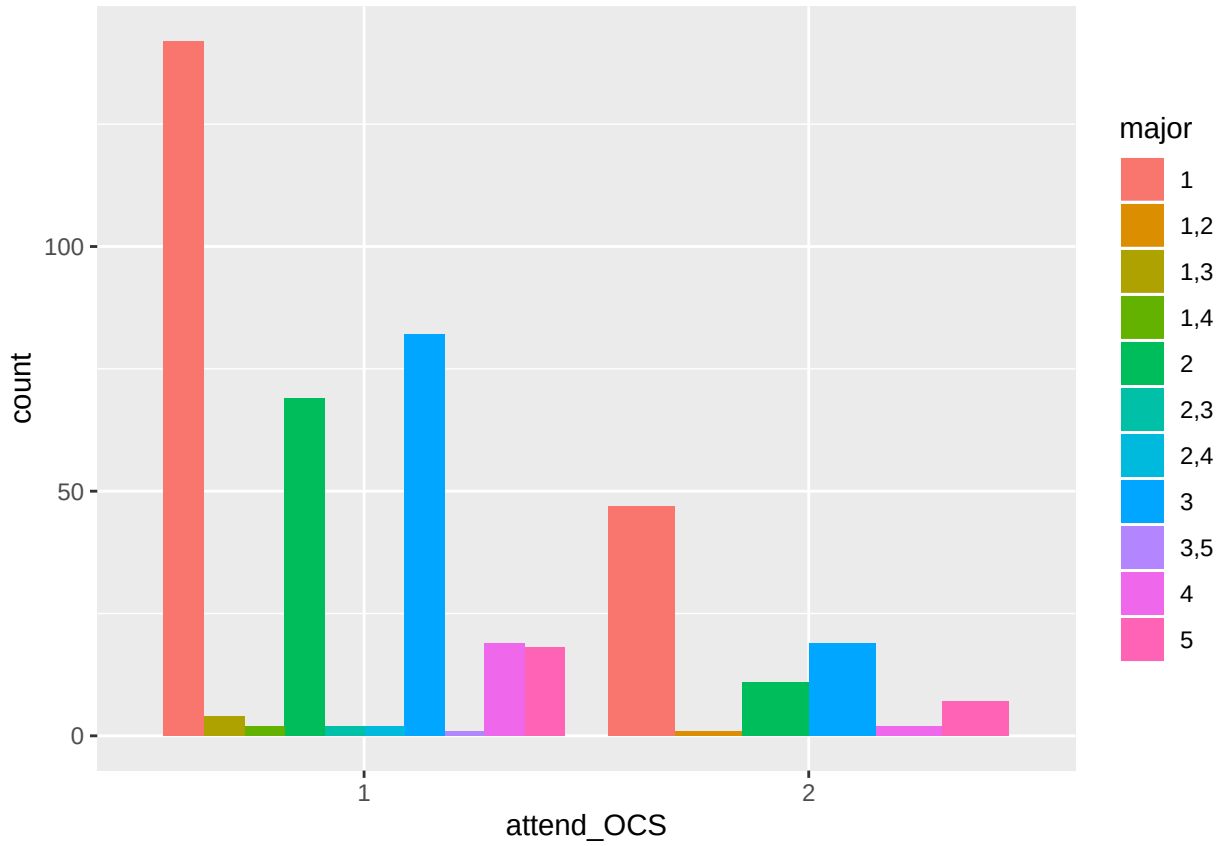
```
## Parsed with column specification:
## cols(
##   .default = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

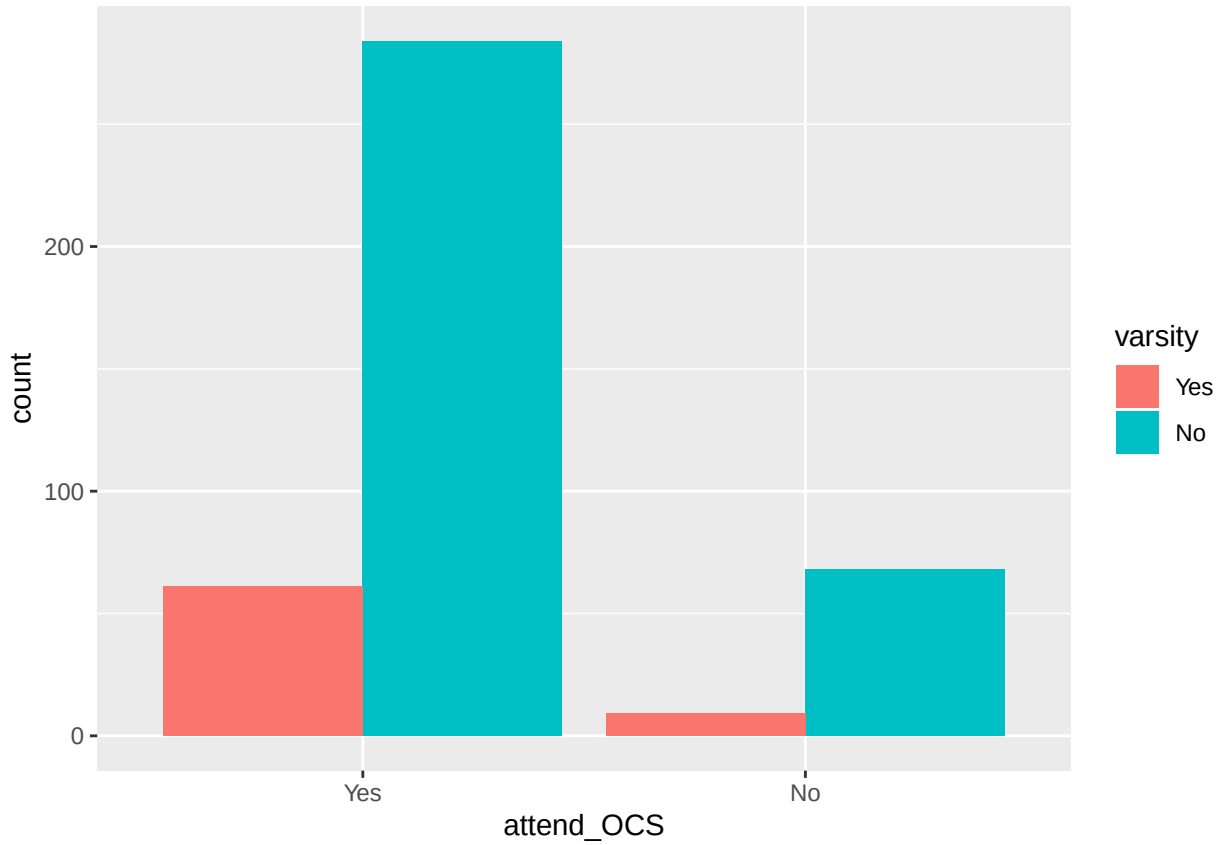
```
#select only the question columns
ocs <- ocs %>%
  select(starts_with("Q"))

#rename column names
names(ocs) <- c("attend_OCS", "reason_not", "reason_not_text", "abroad_classYear", "europe", "reason_eu

#filtering out the first two rows
ocs <- ocs[3:nrow(ocs),]

#find out the percentage of varsity students at Carleton so we can normalize
#https://apps.carleton.edu/voice/?story_id=1836663&section_id=353600&issue_id=1836011
#70% of varsity students study abroad

ocs %>%
  drop_na(varsity) %>%
  group_by(attend_OCS, varsity) %>%
  summarize(count = length(varsity)) %>%
  ggplot(aes(x = attend_OCS, y = count, fill = varsity)) + geom_bar(position = "dodge", stat = "identity
```



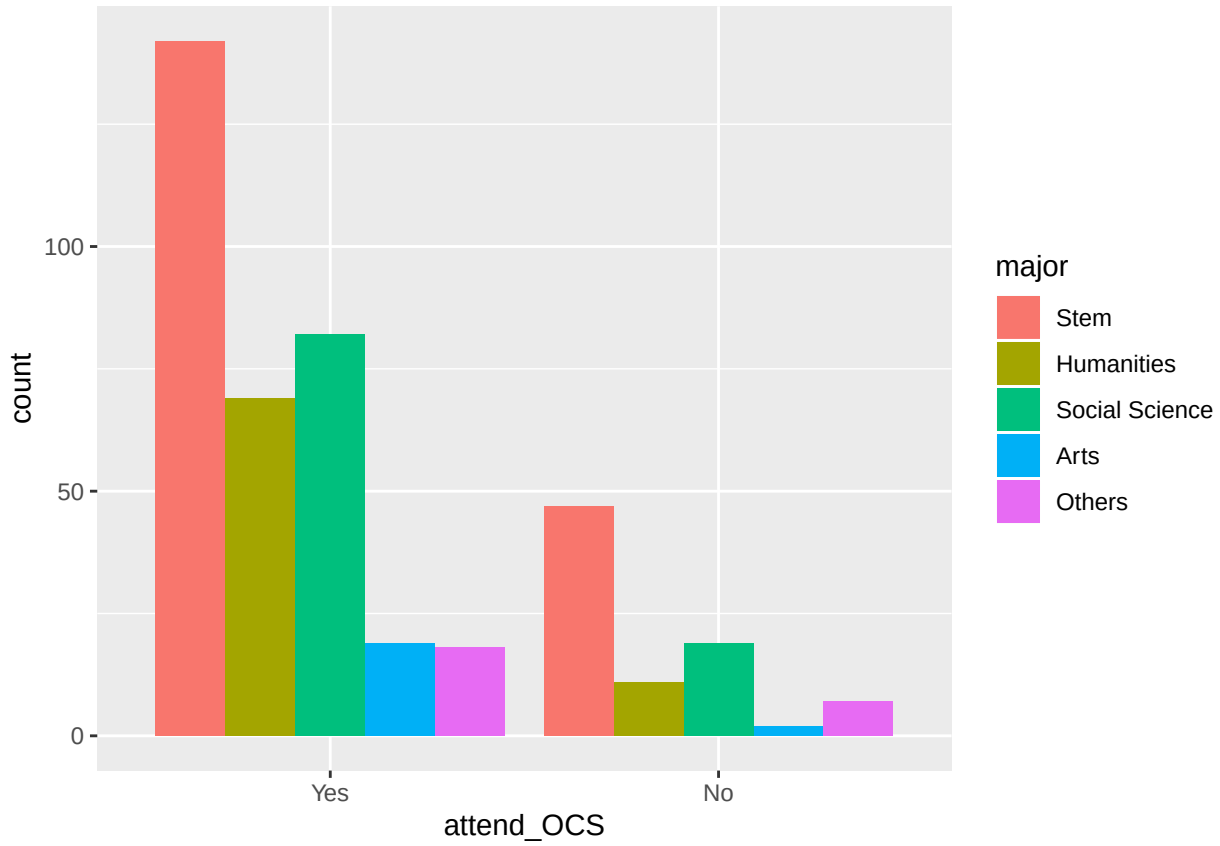

```
ocs %>%
  drop_na(major) %>%
  group_by(attend_OCS, major) %>%
  summarize(count = length(major)) %>%
  ggplot(aes(x = attend_OCS, y = count, fill = major)) + geom_bar(position = "dodge", stat = "identity")
```




```
ocs$attend_OCS <- as.factor(ocs$attend_OCS) %>%
  recode_factor("1" = "Yes", "2" = "No")

ocs$varsity <- as.factor(ocs$varsity) %>%
  recode_factor("1" = "Yes", "2" = "No")

ocs %>%
  drop_na(varsity) %>%
  group_by(attend_OCS, varsity) %>%
  summarize(count = length(varsity)) %>%
  ggplot(aes(x = attend_OCS, y = count, fill = varsity)) + geom_bar(position = "dodge", stat = "identity
```




```
ocs$major <- as.factor(ocs$major) %>%
  recode_factor("1" = "Stem", "2" = "Humanities", "3" = "Social Science", "4" = "Arts", "5" = "Others")

#not including double majors
ocs %>%
  drop_na(major) %>%
  filter(major == "Stem" | major == "Humanities" |major == "Social Science" | major =="Arts"| major ==
  group_by(attend_OCS, major) %>%
  summarize(count = length(major)) %>%
  ggplot(aes(x = attend_OCS, y = count, fill = major)) + geom_bar(position = "dodge", stat = "identity"
```




```
ocs %>%
  group_by(varsity) %>%
  summarize(count = length(varsity))
```

```
## Warning: Factor `varsity` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
## # A tibble: 3 x 2
##   varsity count
##   <fct>   <int>
## 1 Yes        70
## 2 No        352
## 3 <NA>       18
```

```
ocs %>%
  filter(attend_OCS == "No") %>%
  group_by(reason_not) %>%
  summarize(count = length(reason_not))
```

```
## # A tibble: 17 x 2
##    reason_not count
##    <chr>      <int>
##  1 1              5
##  2 1,2            3
##  3 1,2,3          1
##  4 1,2,5          1
##  5 1,4            2
##  6 2             18
```

```
##  7 2,3            2
##  8 2,3,5          1
##  9 2,4            2
## 10 2,5            5
## 11 3              4
## 12 3,4            2
## 13 3,5            7
## 14 4              1
## 15 4,5            1
## 16 5             31
## 17 <NA>           3
```

```
#check the statistical significance between number of people who are in varsity and poeple who didn't g
```

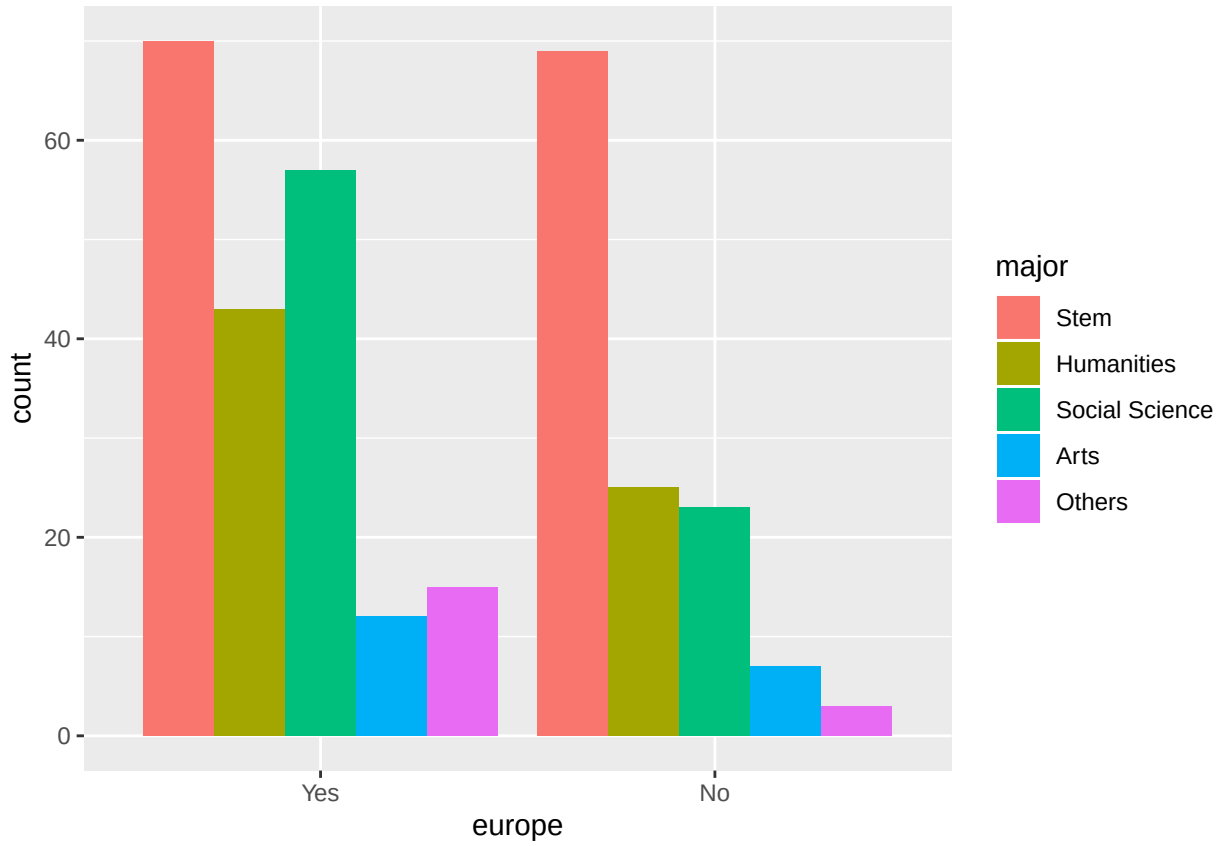
```
ocs$europe <- as.factor(ocs$europe) %>%
  recode_factor("1" = "Yes", "2" = "No")

ocs %>%
  group_by(europe, major) %>%
  summarize(count = length(europe))
```

```
## Warning: Factor `europe` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
## Warning: Factor `major` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
## # A tibble: 28 x 3
## # Groups:   europe [3]
##    europe major          count
##    <fct>  <fct>          <int>
##  1 Yes    Stem              70
##  2 Yes    Humanities        43
##  3 Yes    Social Science    57
##  4 Yes    Arts              12
##  5 Yes    Others            15
##  6 Yes    1,3                2
##  7 Yes    2,3                1
##  8 Yes    2,4                1
##  9 Yes    <NA>               2
## 10 No     Stem              69
## # ... with 18 more rows
```

```
#not including double majors
ocs %>%
  drop_na(major, europe) %>%
  filter(major == "Stem" | major == "Humanities" |major == "Social Science" | major =="Arts"| major ==
  group_by(europe, major) %>%
  summarize(count = length(major)) %>%
  ggplot(aes(x = europe, y = count, fill = major)) + geom_bar(position = "dodge", stat = "identity")
```




```
ocs %>%
  filter(europe == "Yes") %>%
  group_by(reason_europe) %>%
  summarize(count = length(reason_not))
```

```
## # A tibble: 29 x 2
##    reason_europe count
##    <chr>         <int>
##  1 1                 5
##  2 1,2              12
##  3 1,2,3,4           2
##  4 1,2,3,4,5         1
##  5 1,2,4            36
##  6 1,2,4,5          16
##  7 1,2,5             9
##  8 1,2,5,6           1
##  9 1,2,6             2
## 10 1,3,4             1
## # ... with 19 more rows
```