initial-eda

Lillian Clark

11/8/2020

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
## v purrr   0.3.4

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

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

```
library(broom)
library(knitr)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

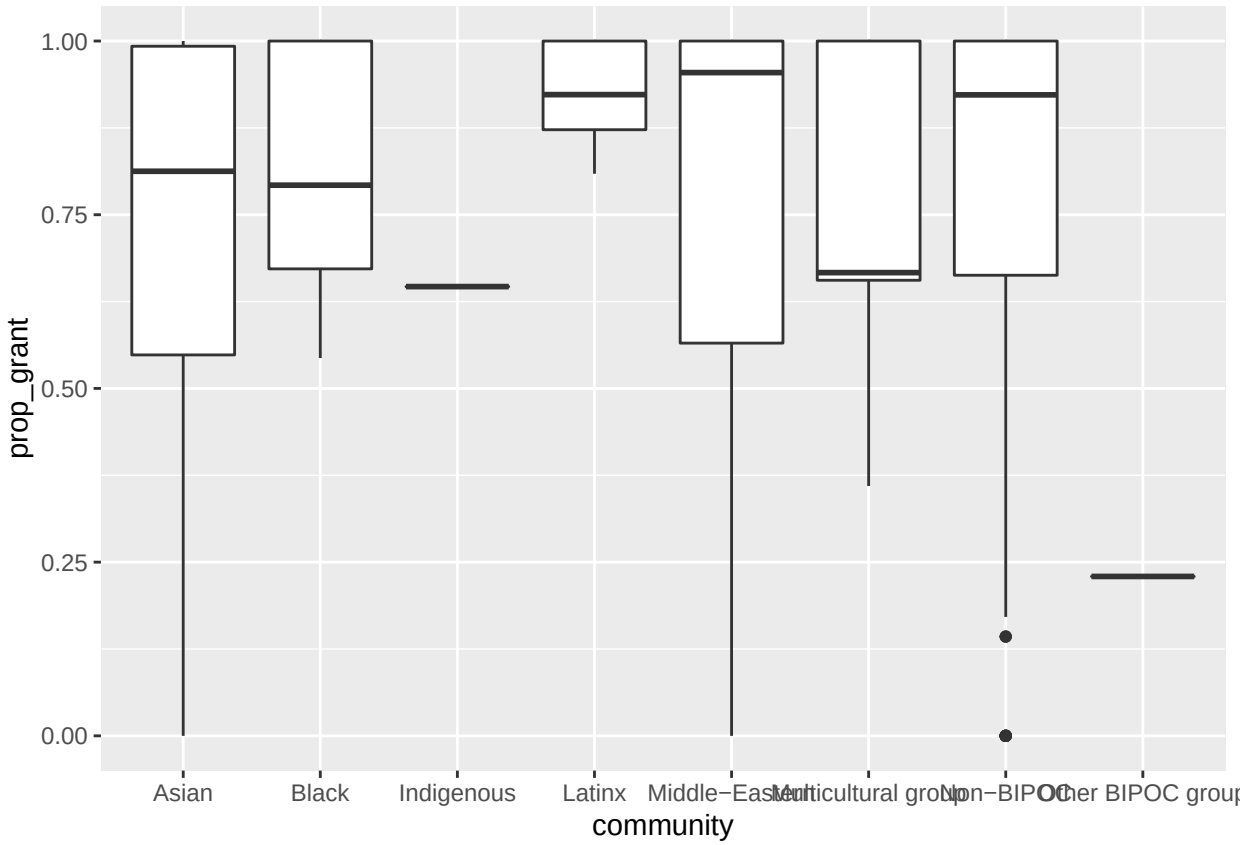
```
prog <- read.csv("data-labeled/programming.csv")
prog <- prog[-1]
budget <- read.csv("data-labeled/budget.csv")
budget <- budget[-1]
sofc <- read.csv("data-labeled/sofc.csv")
sofc <- sofc[-1]
```

What to do with these groups? >> Center for Race Relations

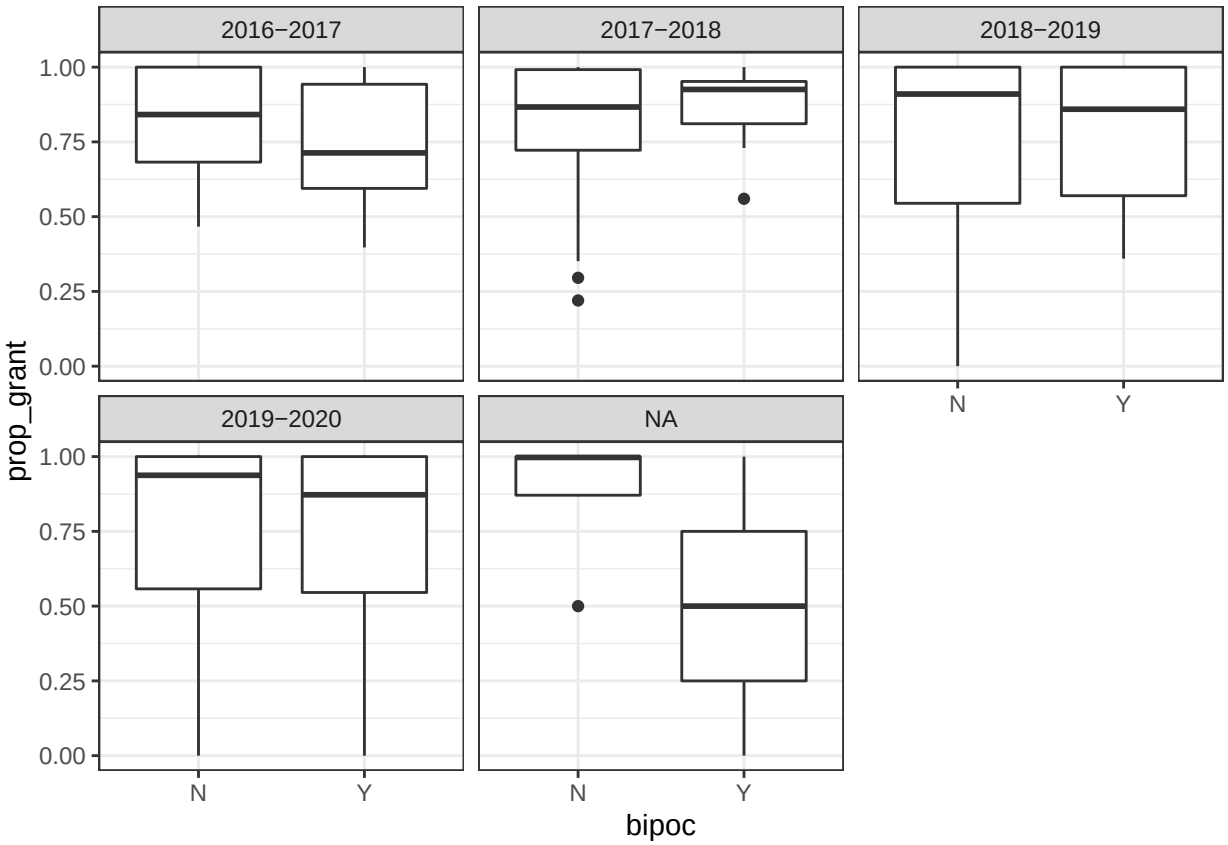
```
make_plots <- function(df) {
  plot1 <- ggplot(df, aes(x = community, y = prop_grant)) +
    geom_boxplot()
    theme_bw()

  plot2 <- ggplot(df, aes(x = bipoc, y = prop_grant)) +
    geom_boxplot() +
    facet_wrap(. ~ schoolyr) +
    theme_bw()

  plot3 <- ggplot(df, aes(x = community, y = prop_grant)) +
    geom_boxplot() +
    coord_flip() +
    facet_wrap(. ~ schoolyr) +
    theme_bw()

  plot4 <- ggplot(df, aes(x = community, y = prop_grant)) +
    geom_point(alpha = 0.3)+
    theme_bw()

  plot5 <- ggplot(df, aes(x = prop_grant)) +
    geom_histogram(aes(fill = factor(community, levels=c("Asian", "Black",
                                                         "Indigenous", "Latinx",
                                                         "Middle-Eastern",
                                                         "Multicultural group",
                                                         "Other BIPOC group",
                                                         "Non-BIPOC"))),
                   position = "stack", color = "white") +
    scale_fill_discrete(name = "community") +
    theme_bw()

  plot6 <- ggplot(df, aes(x = prop_grant)) +
    geom_histogram() +
    facet_wrap(. ~ community) +
    theme_bw()

  return(list(plot1, plot2, plot3, plot4, plot5, plot6))
}
```

```
make_plots(prog)
```

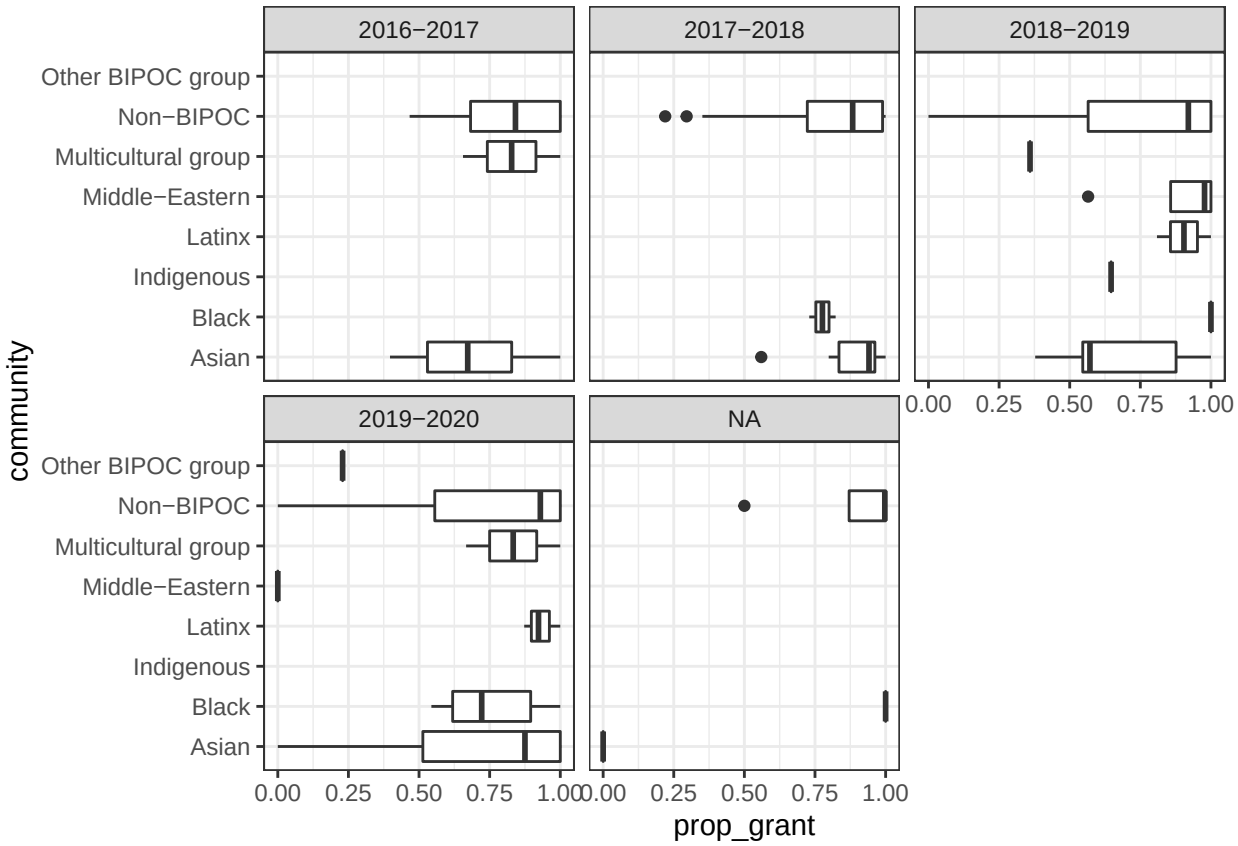
```
## [[1]]
```


1.00
0.75
0.50
0.25
0.00
prop_grant
Asian
Black
Indigenous
Latinx
Middle−Eastern
Multicultural group
Non−BIPOC
Other BIPOC group
community

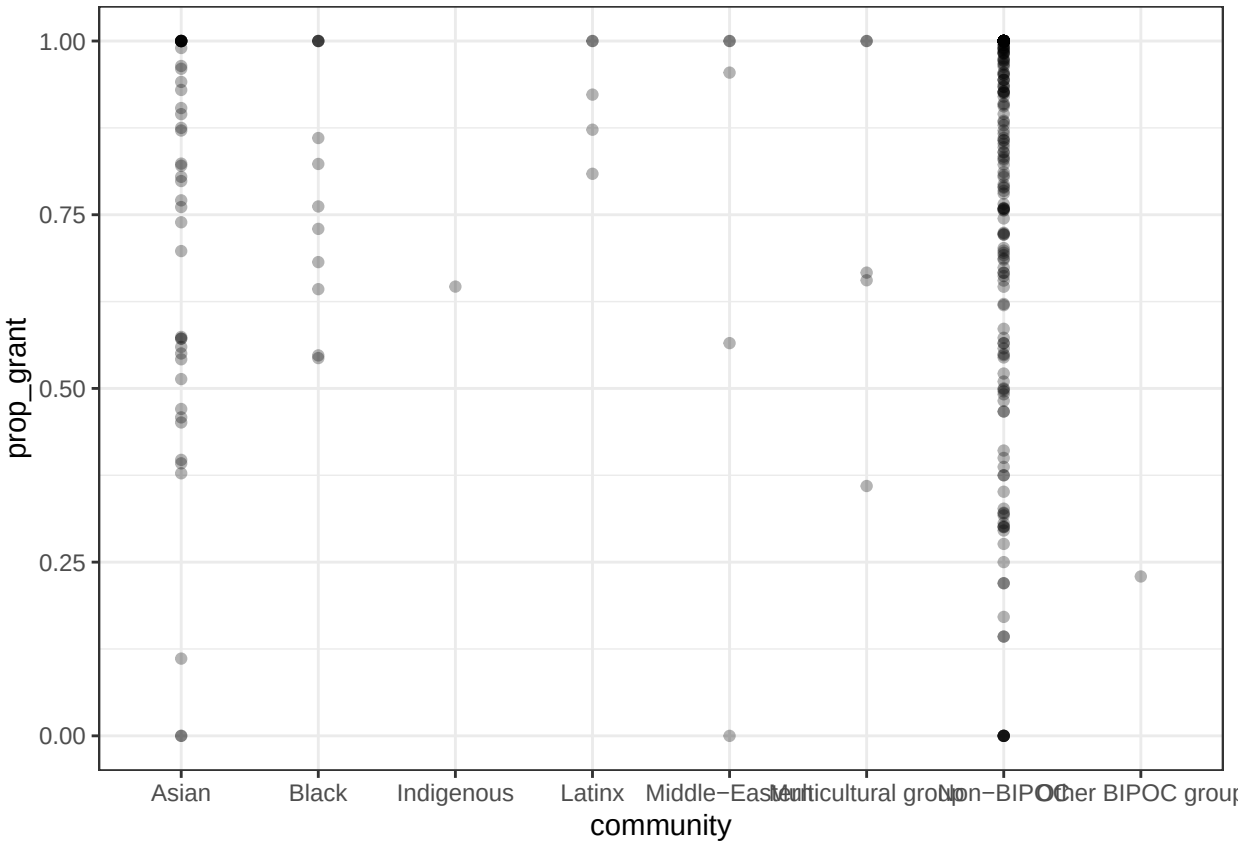
```
##
## [[2]]
```


2016−2017
2017−2018
2018−2019
2019−2020
NA
prop_grant
1.00
0.75
0.50
0.25
0.00
N
Y
bipoc

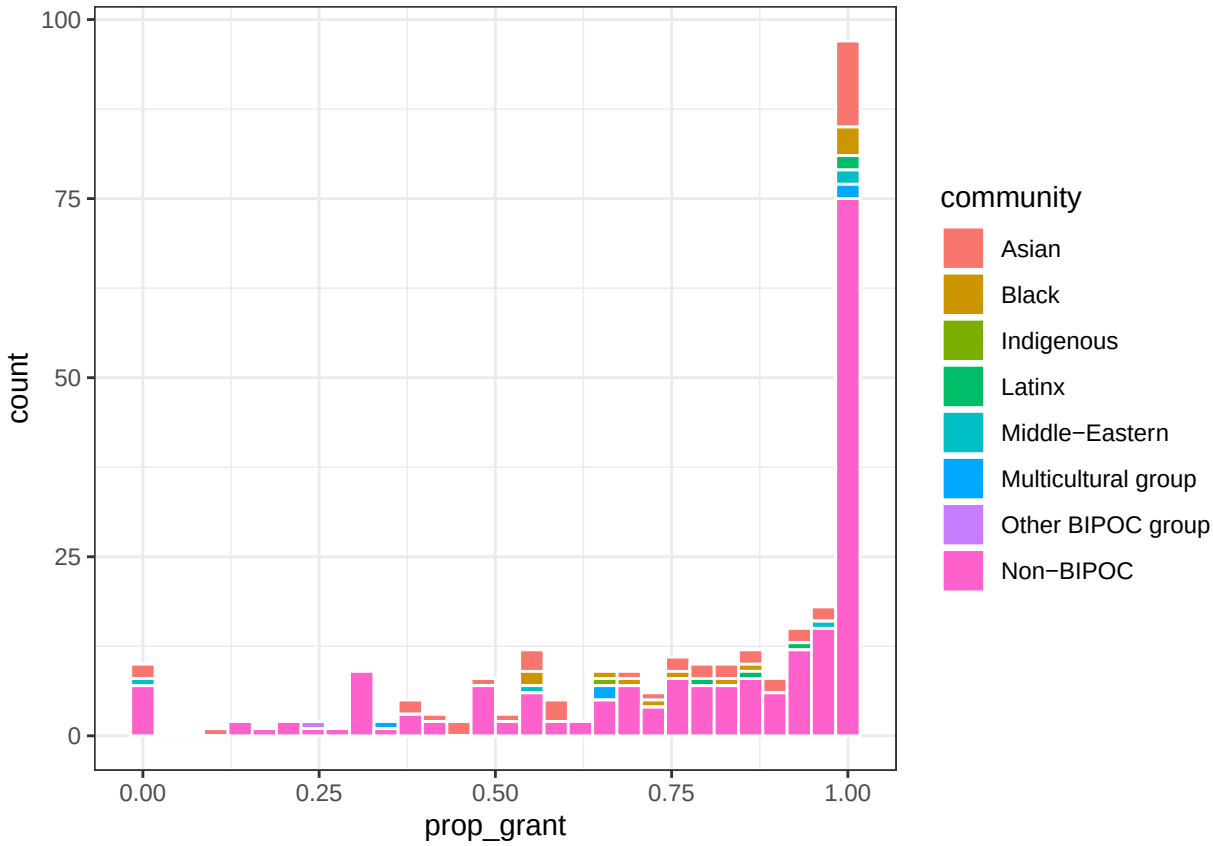
```
##
## [[3]]
```


2016−2017
2017−2018
2018−2019
2019−2020
NA
Other BIPOC group
Non−BIPOC
Multicultural group
Middle−Eastern
Latinx
Indigenous
Black
Asian
community
0.00
0.25
0.50
0.75
1.00
prop_grant

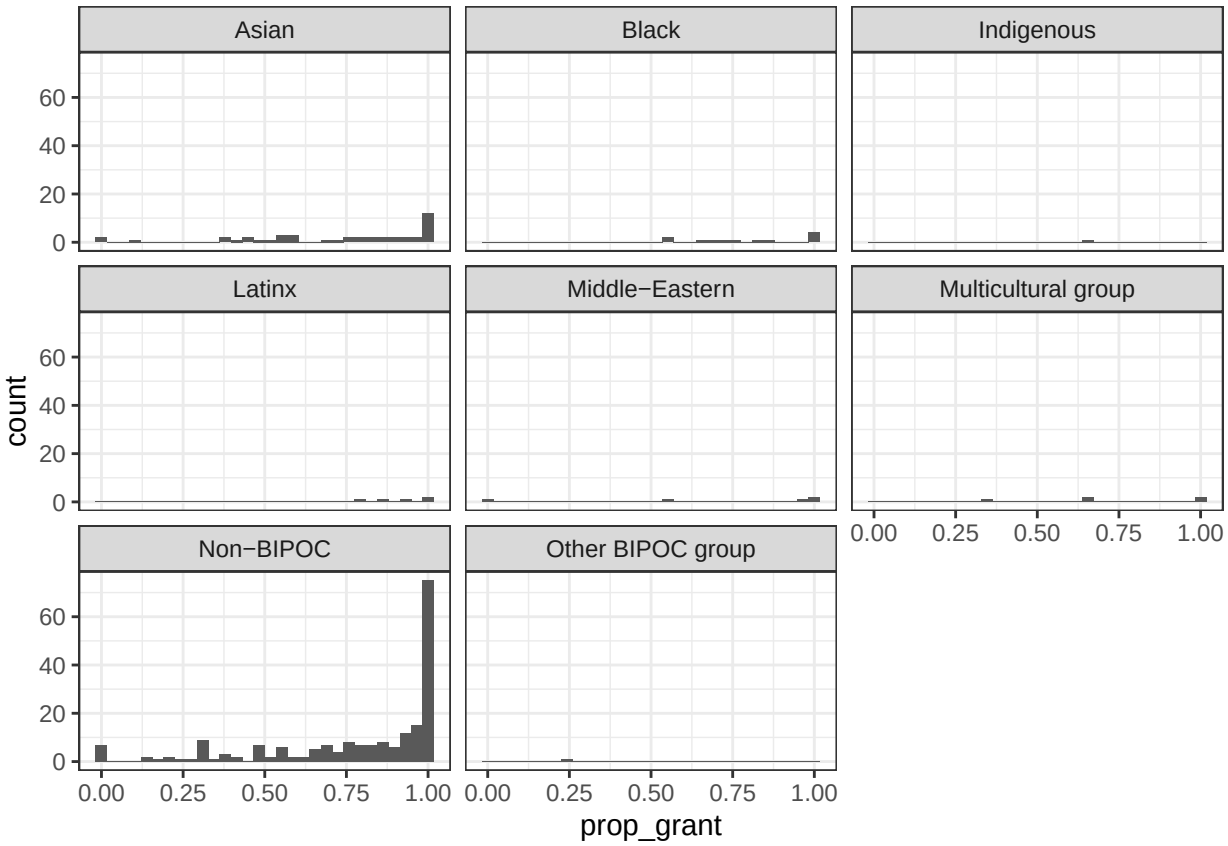
```
##
## [[4]]
```




```
##
## [[5]]

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

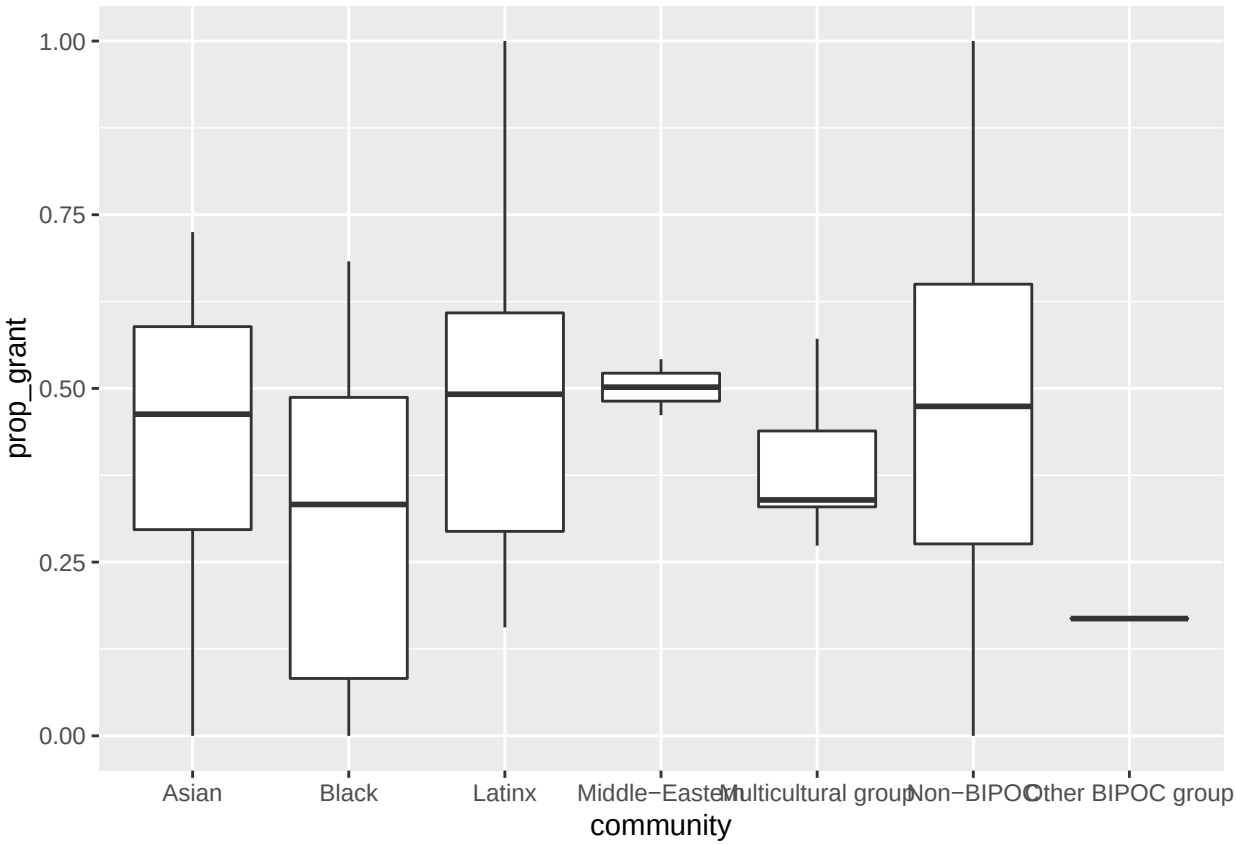



```
##
## [[6]]

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

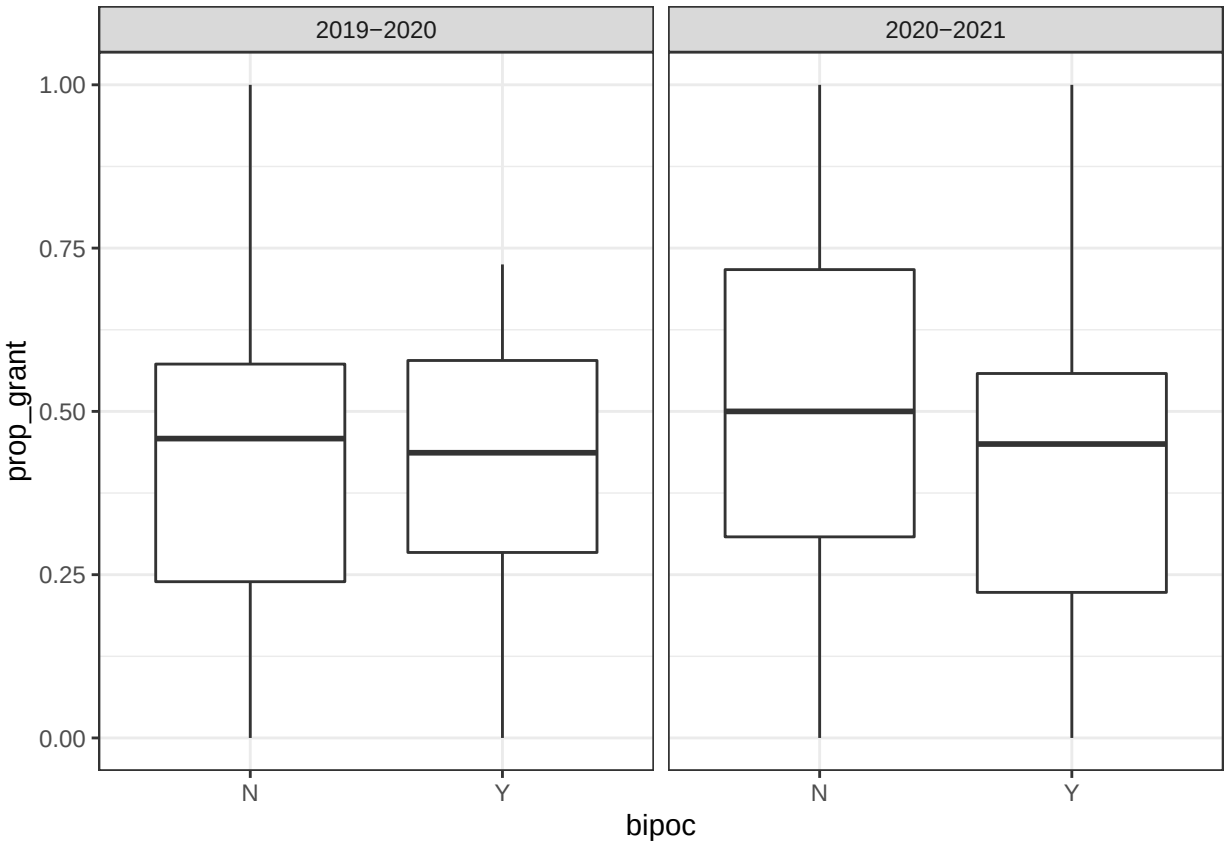



```
make_plots(budget)
```

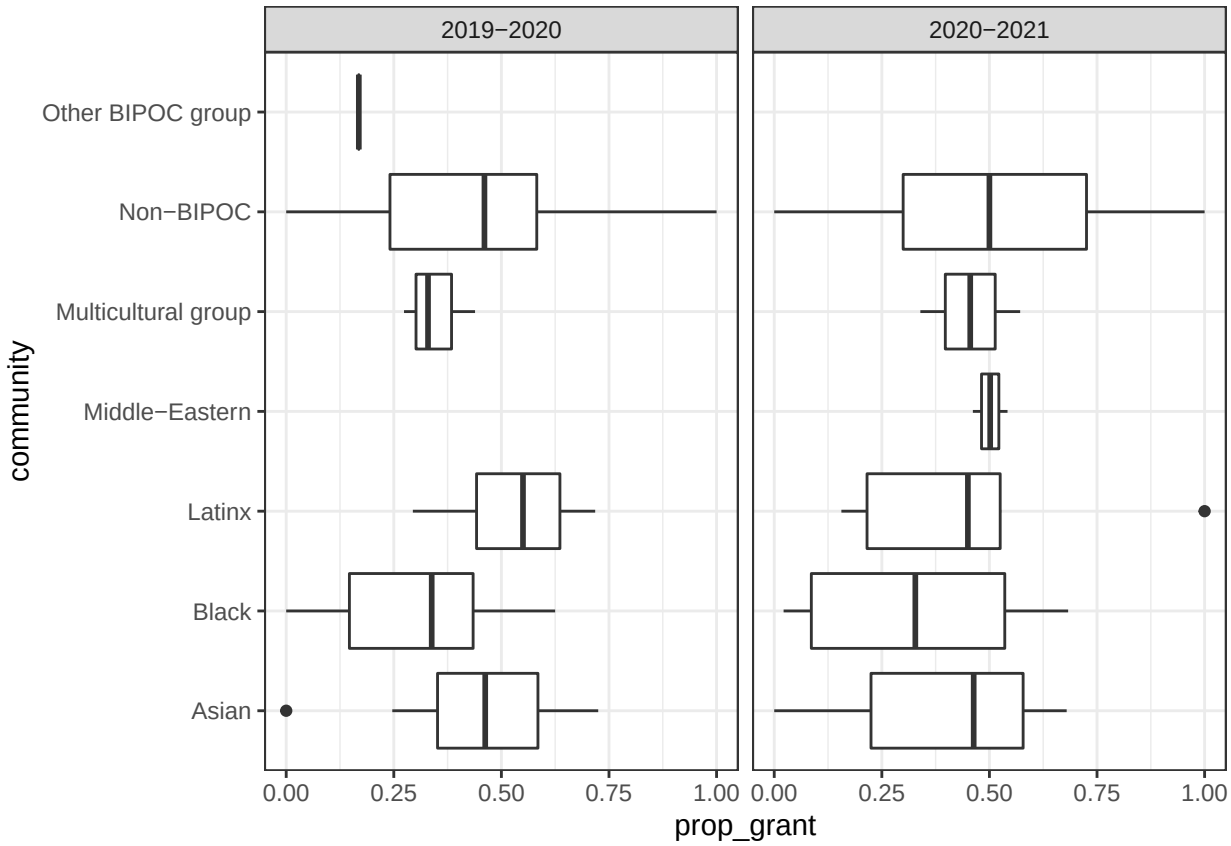
```
## [[1]]
```


1.00
0.75
0.50
0.25
0.00
prop_grant
Asian
Black
Latinx
Middle−Eastern
Multicultural group
Non−BIPOC
Other BIPOC group
community

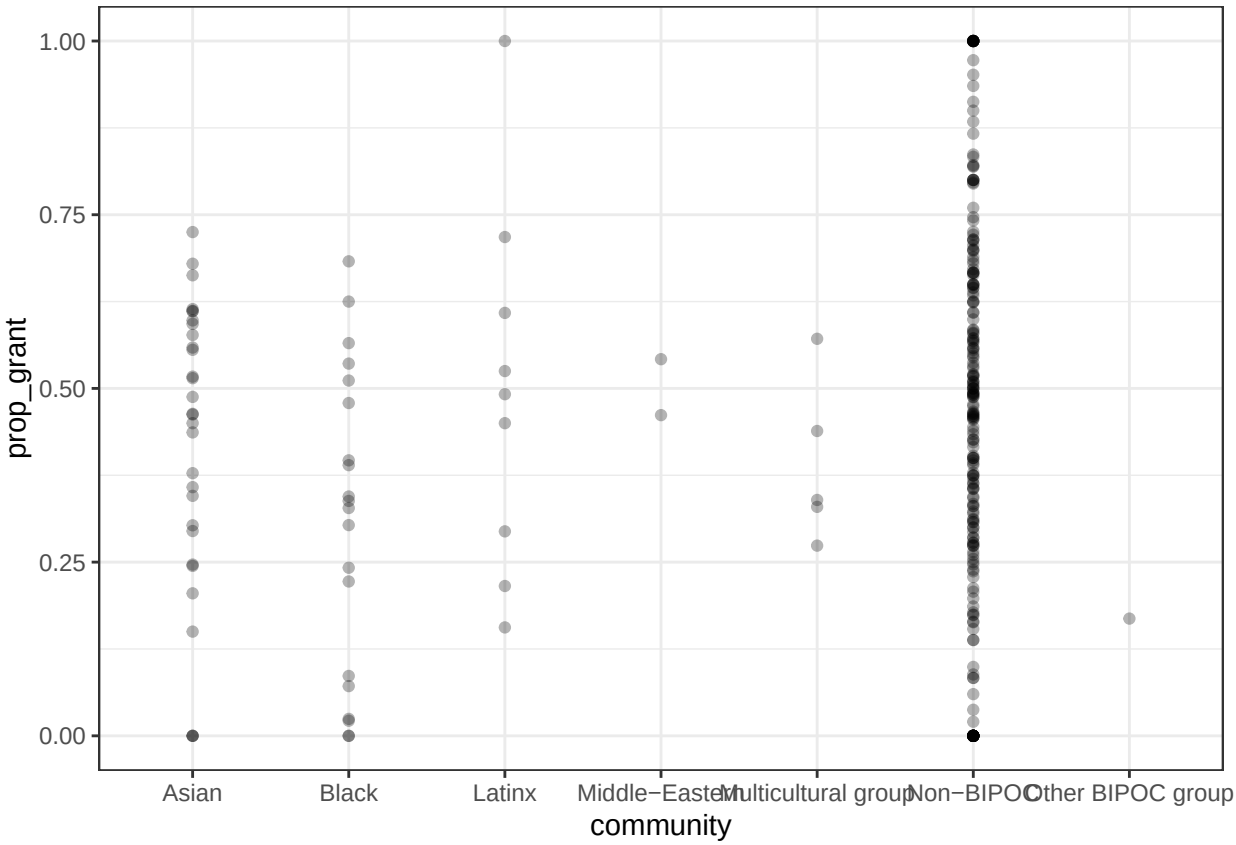
```
##
## [[2]]
```



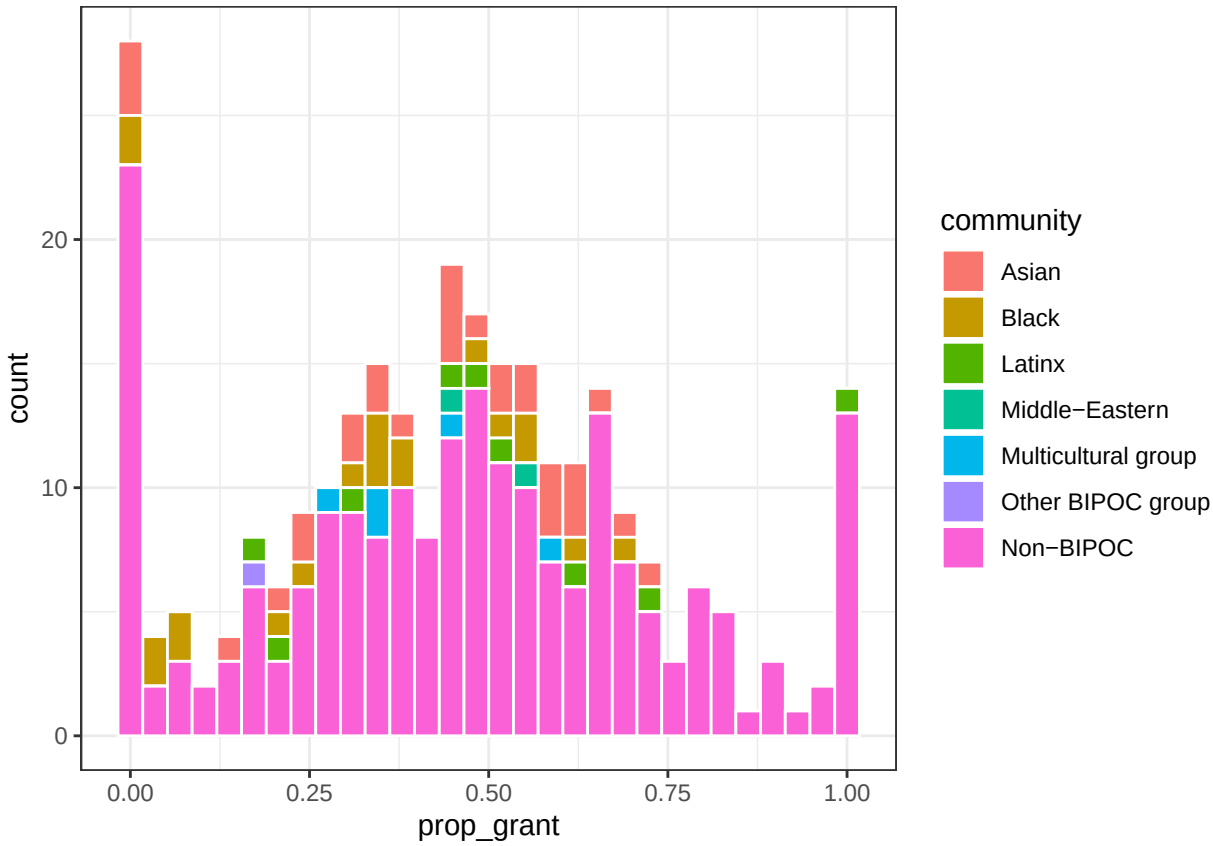

```
##
## [[3]]
```


2019−2020
2020−2021
Other BIPOC group
Non−BIPOC
Multicultural group
Middle−Eastern
Latinx
Black
Asian
community
0.00
0.25
0.50
0.75
1.00
prop_grant

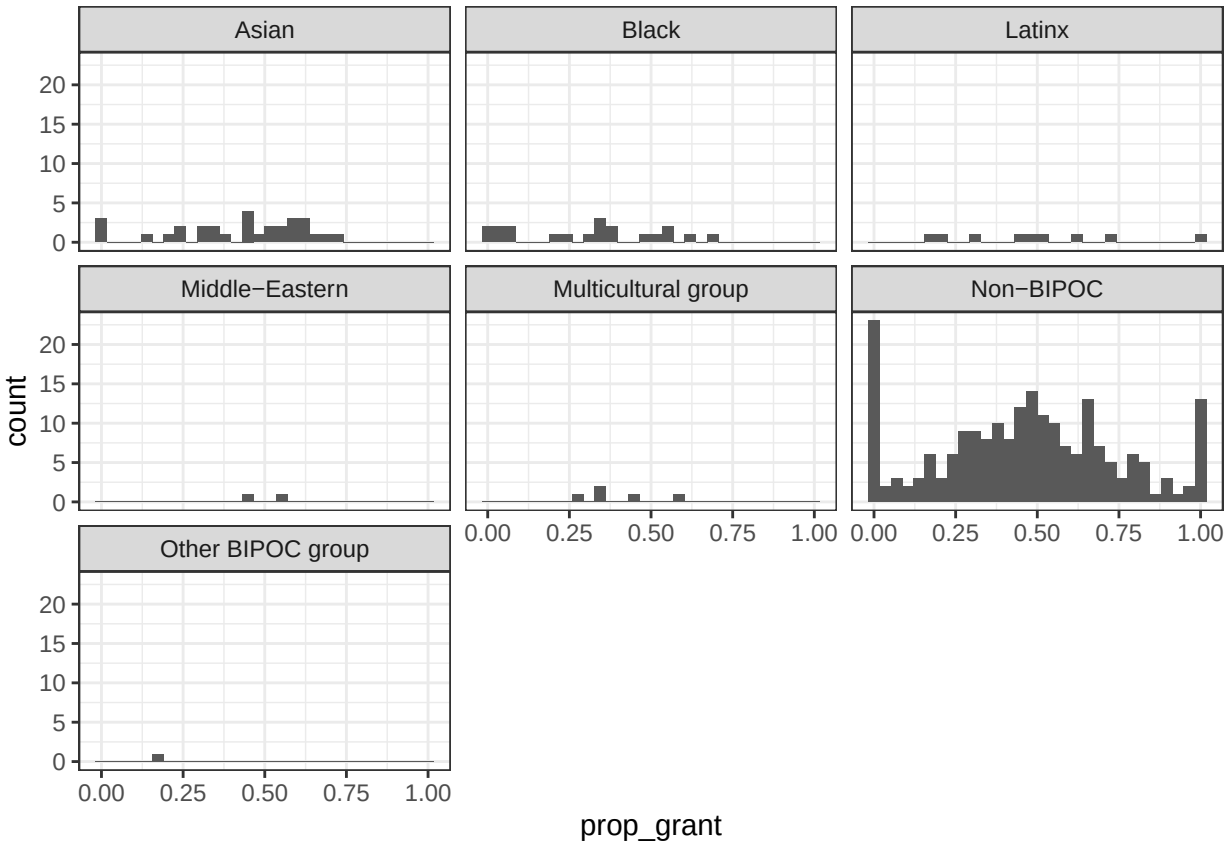
```
##
## [[4]]
```




```
##
## [[5]]

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

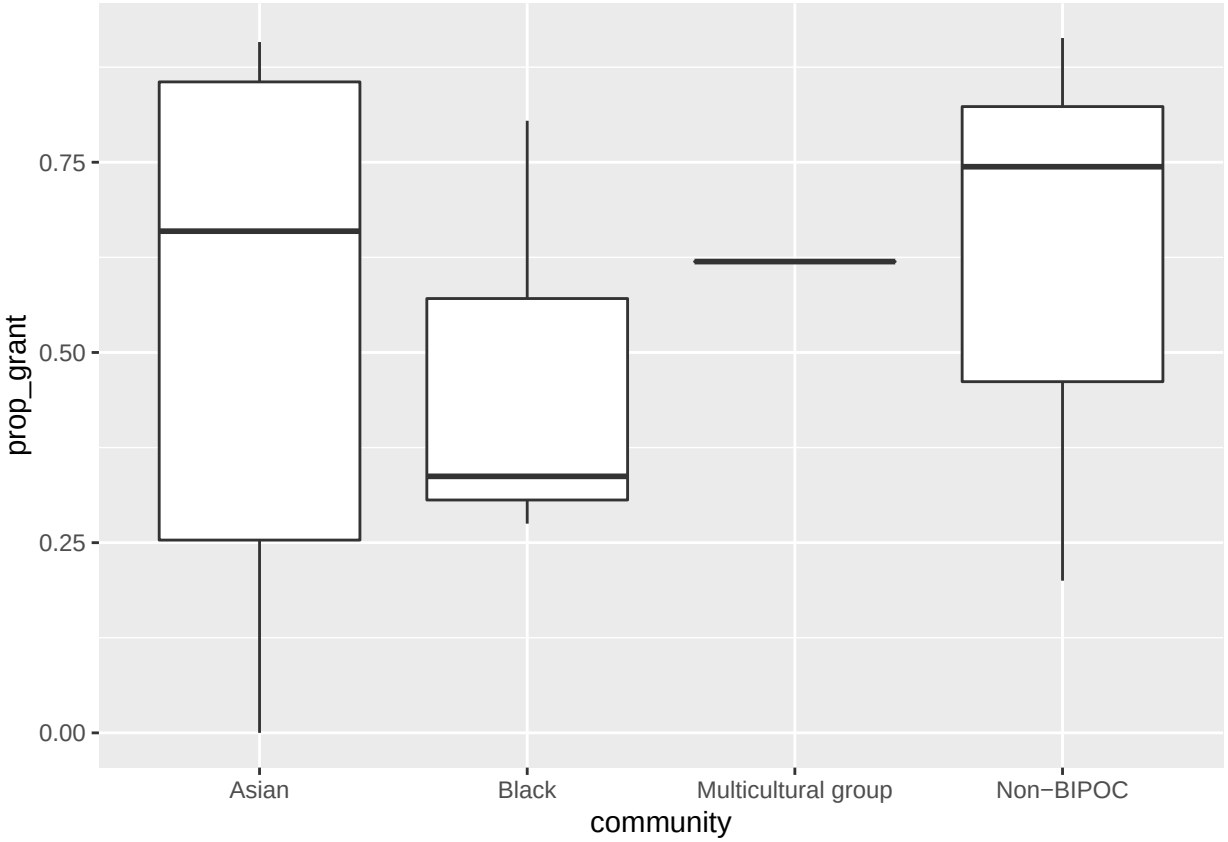



```
##
## [[6]]

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

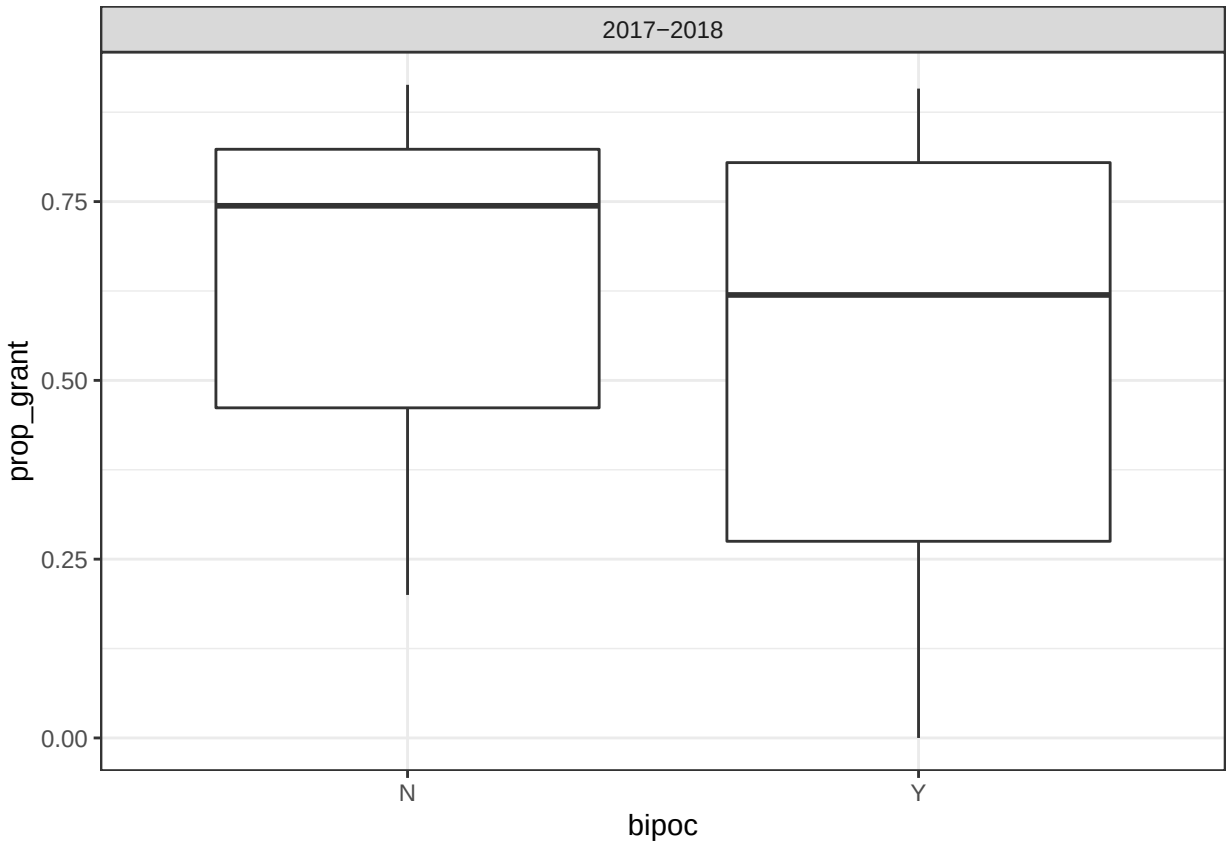



```
make_plots(sofc)
```

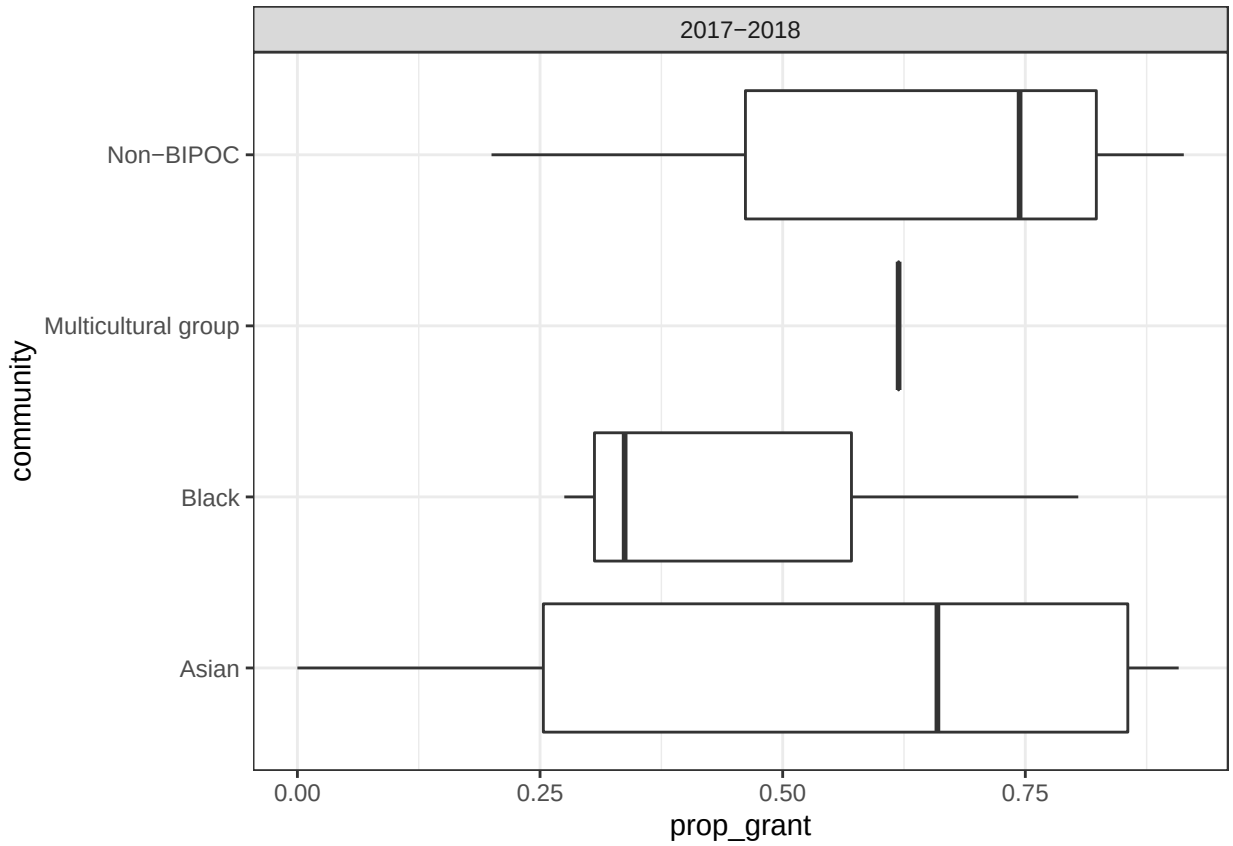
```
## [[1]]
```


0.75
0.50
0.25
0.00
prop_grant
Asian
Black
Multicultural group
Non−BIPOC
community

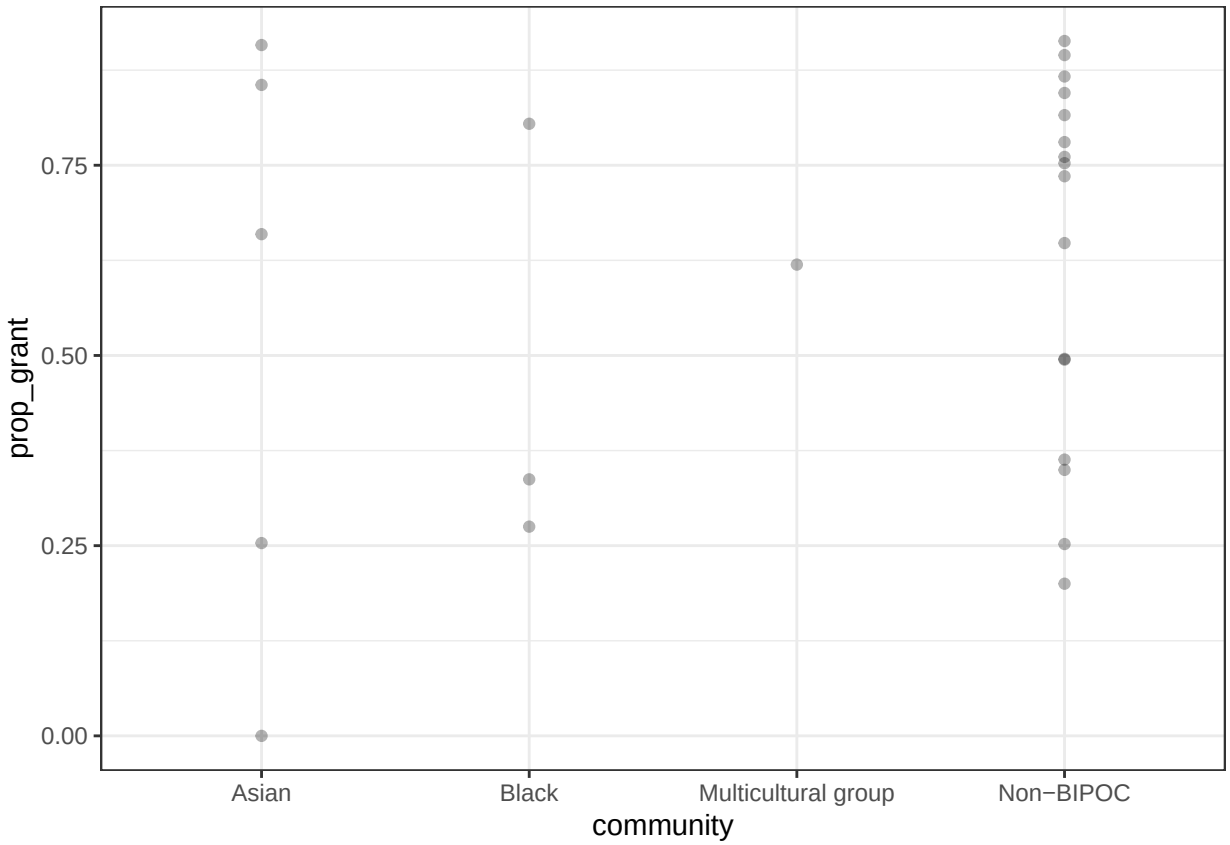
```
##
## [[2]]
```



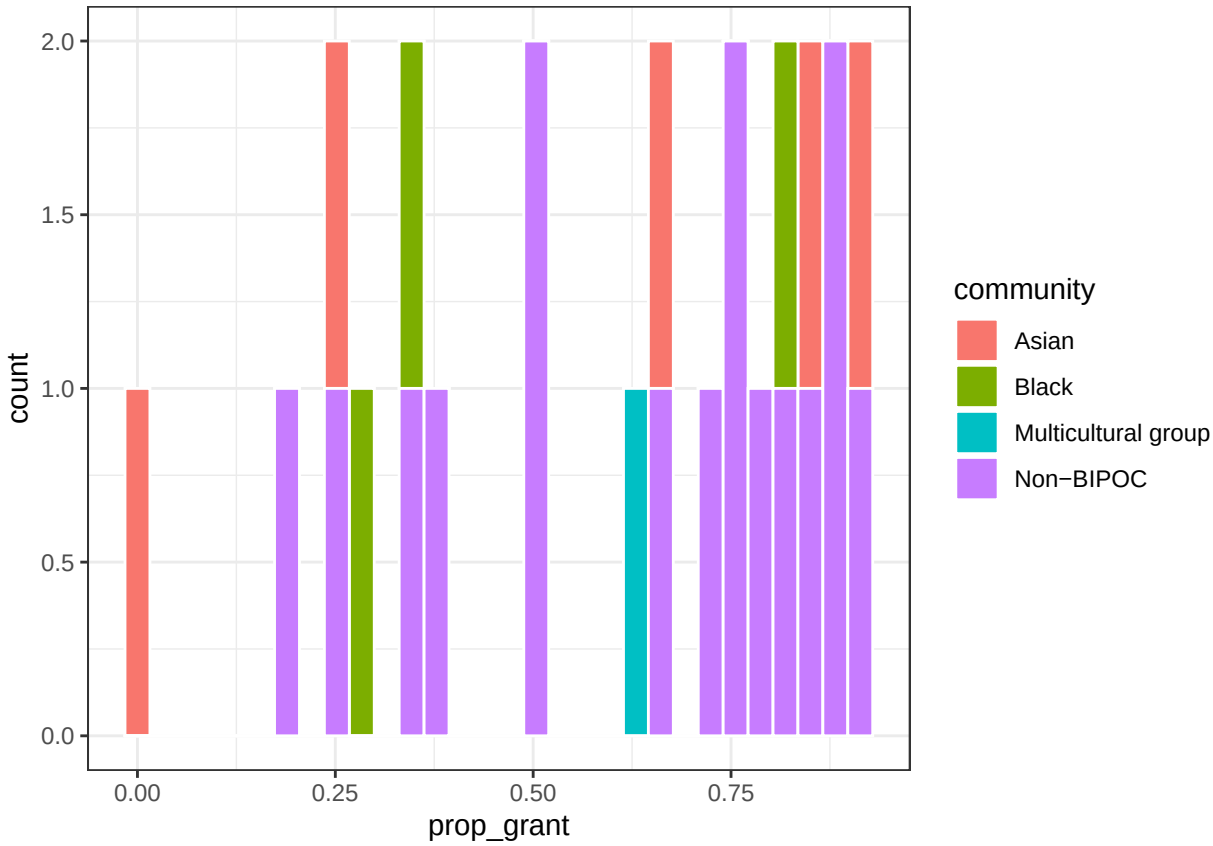

```
##
## [[3]]
```


2017−2018
Non−BIPOC
Multicultural group
Black
Asian
community
0.00
0.25
0.50
0.75
prop_grant

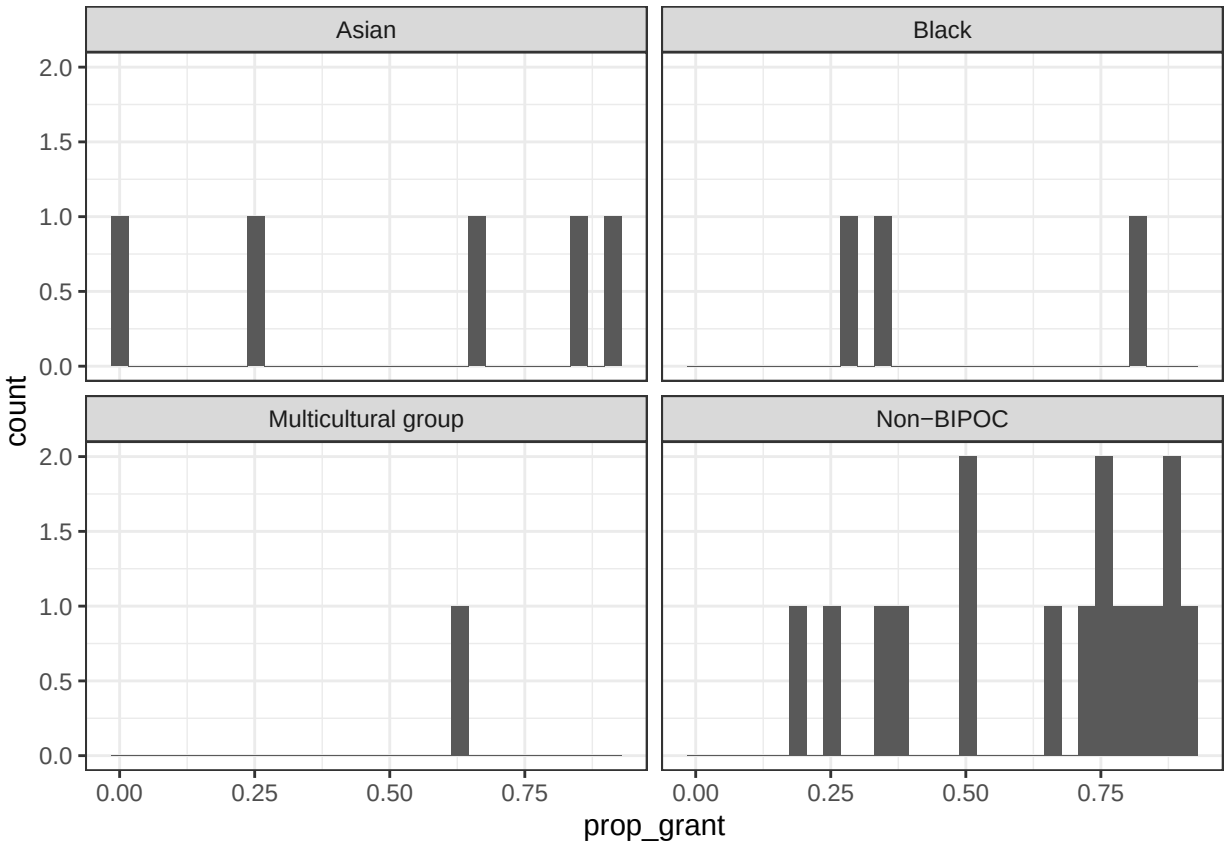
```
##
## [[4]]
```



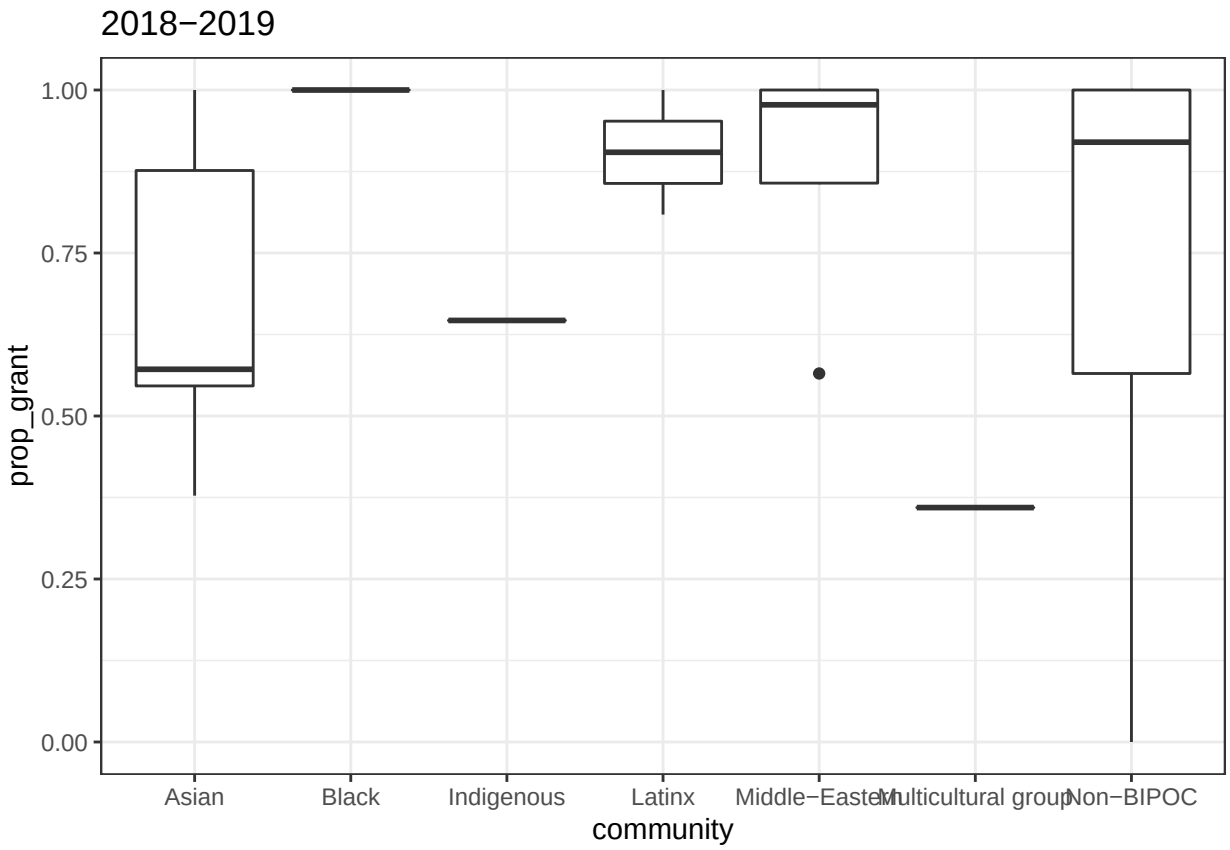

```
##
## [[5]]

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



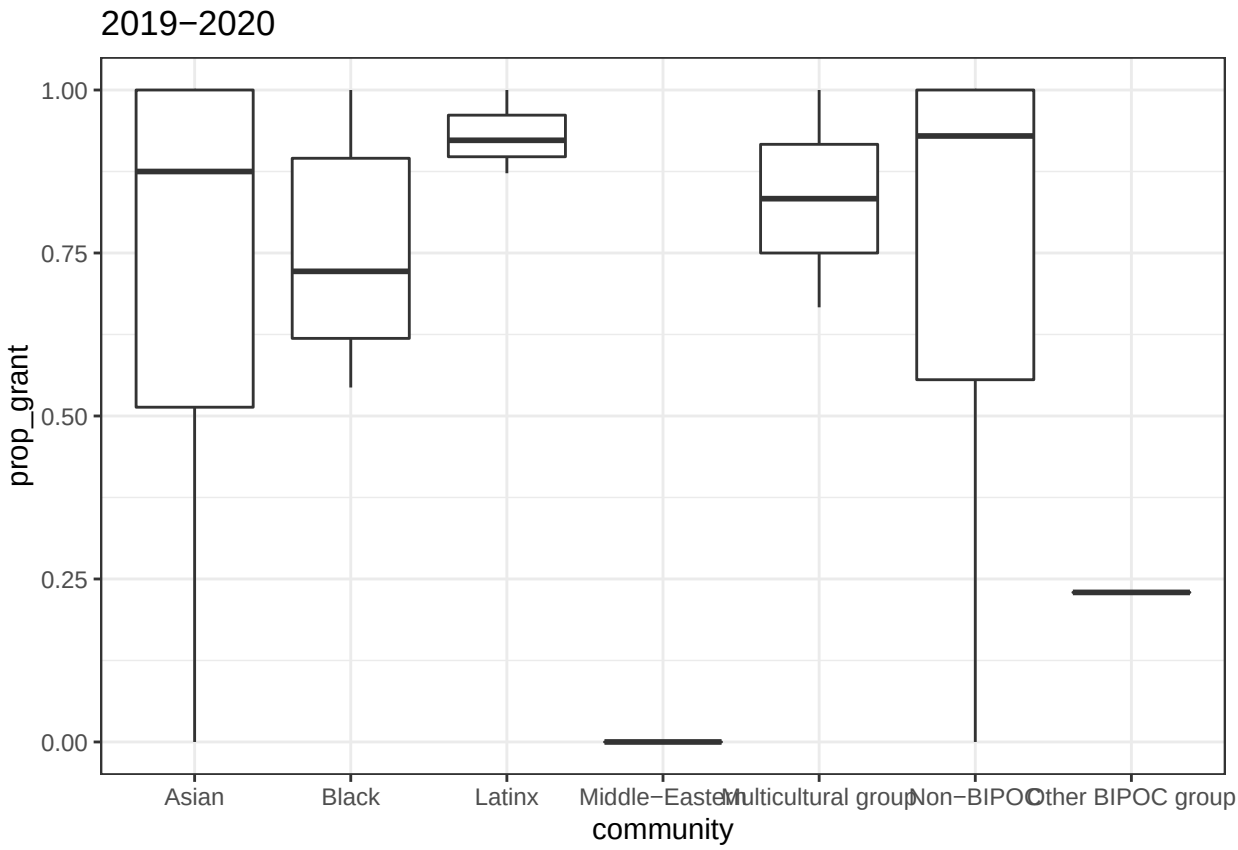

```
##
## [[6]]

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



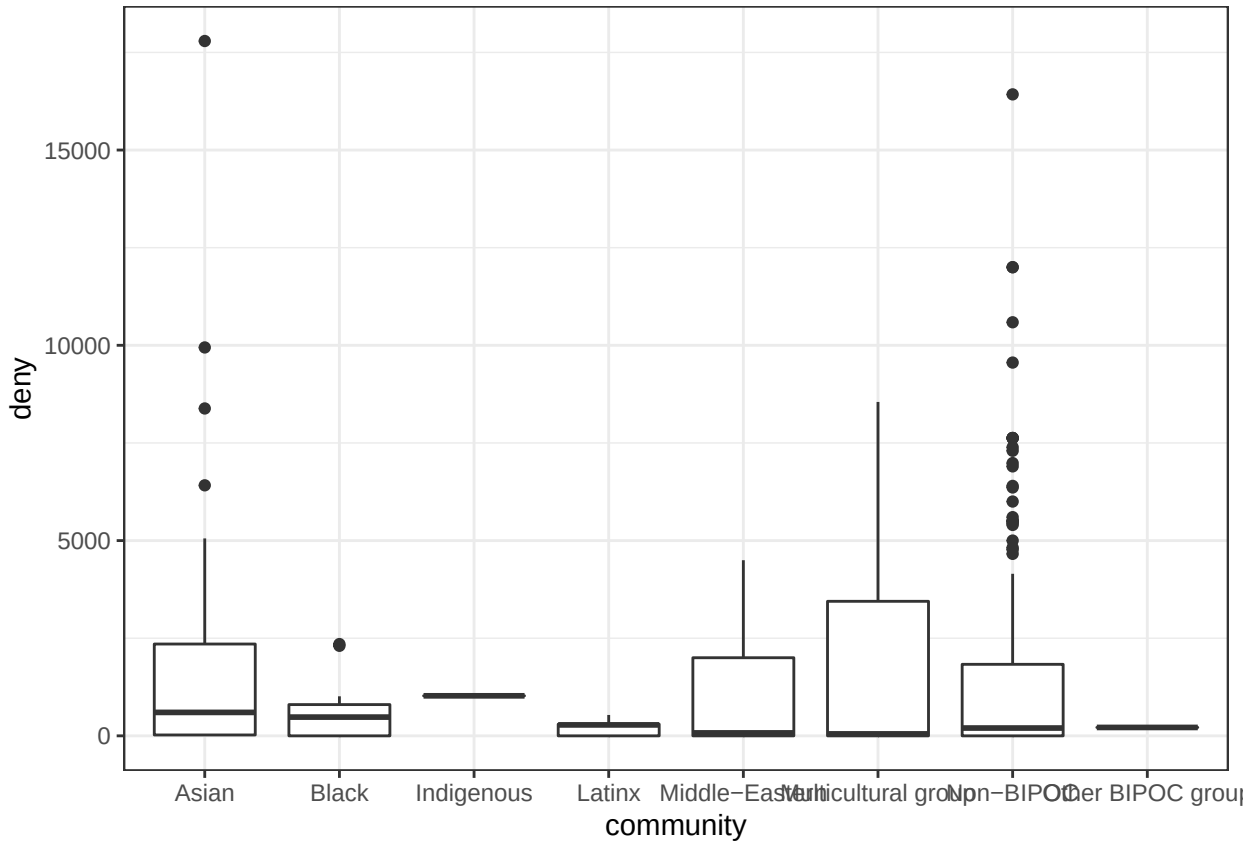

```
prog %>%
  filter(schoolyr == "2018-2019") %>%
  ggplot(aes(x = community, y = prop_grant)) +
  labs(title = "2018-2019") +
  geom_boxplot() +
  theme_bw()
```



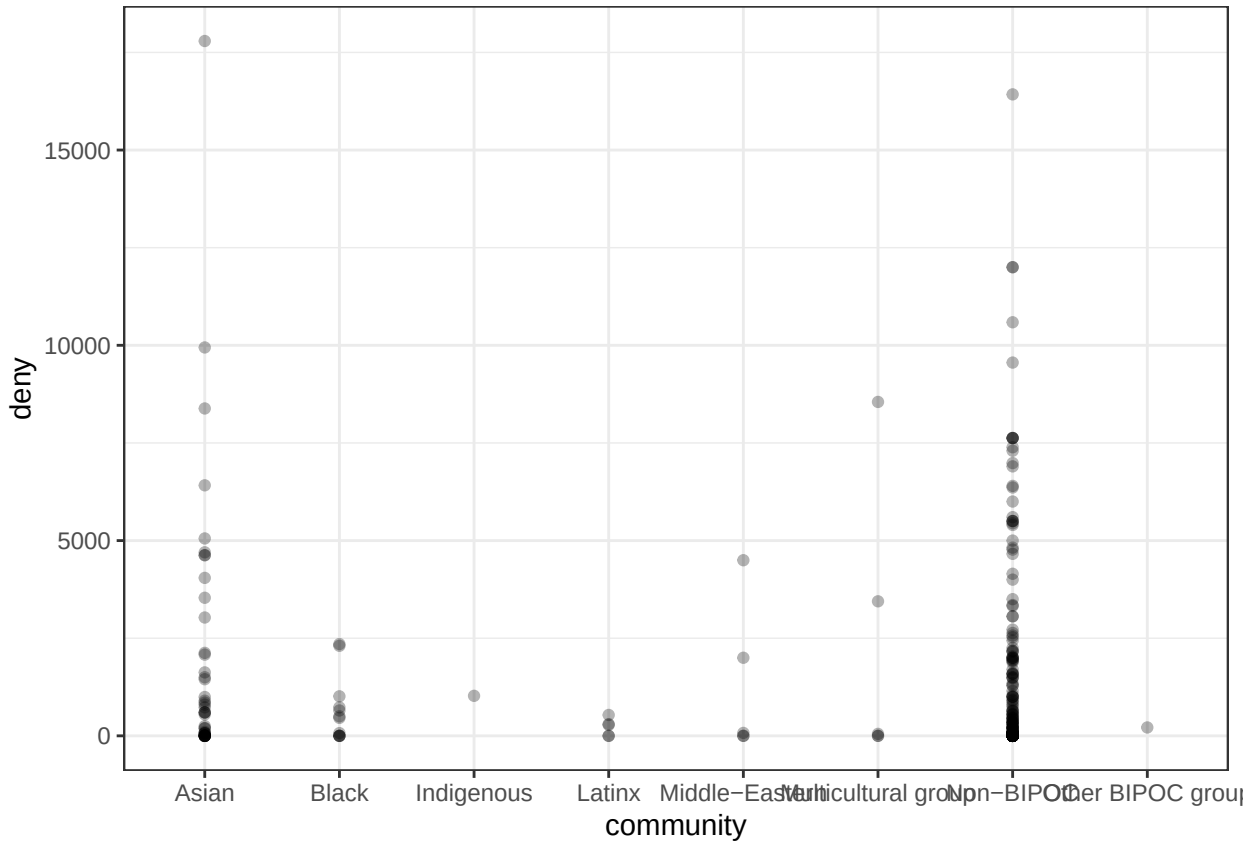

```
prog %>%
  filter(schoolyr == "2019-2020") %>%
  ggplot(aes(x = community, y = prop_grant)) +
  labs(title = "2019-2020") +
  geom_boxplot() +
  theme_bw()
```




```
ggplot(prog, aes(x = community, y = deny)) +
  geom_boxplot() +
  theme_bw()
```

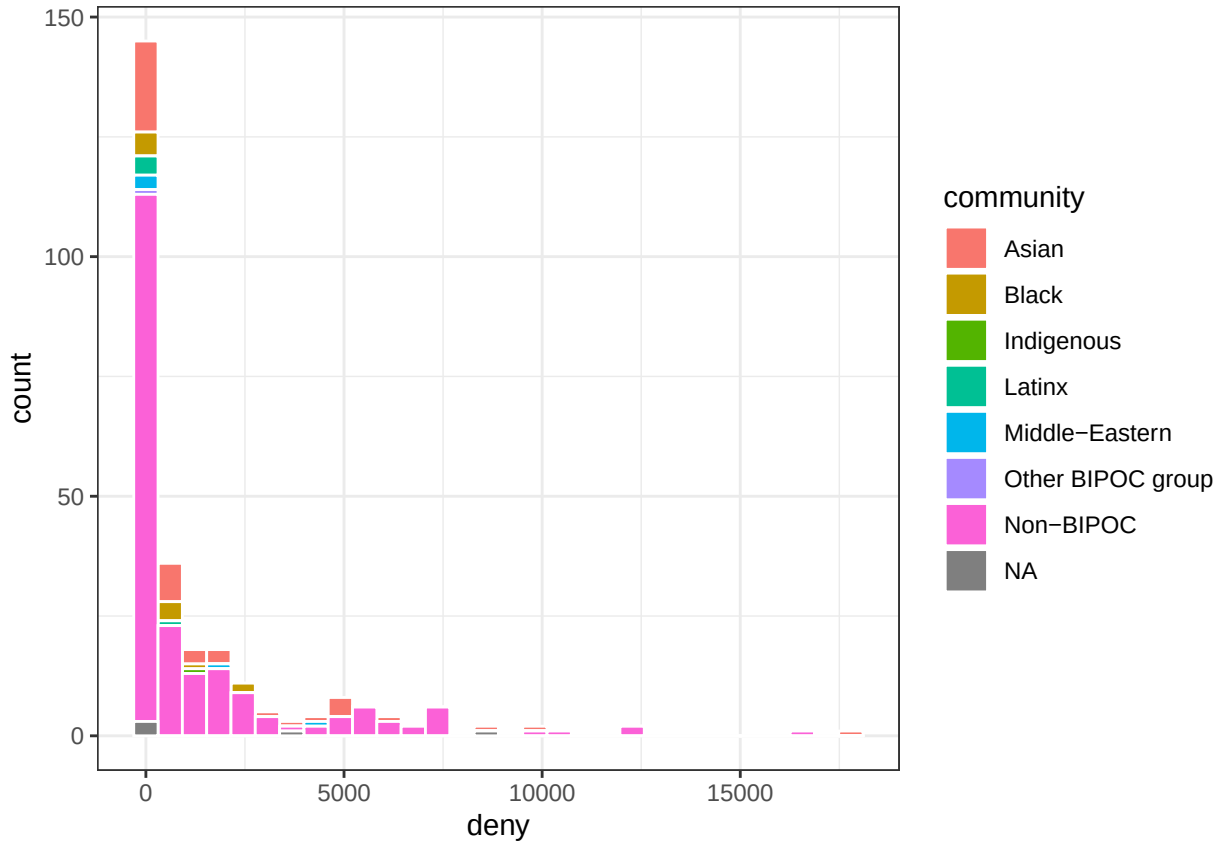



```
ggplot(prog, aes(x = community, y = deny)) +
  geom_point(alpha = 0.3) +
  theme_bw()
```



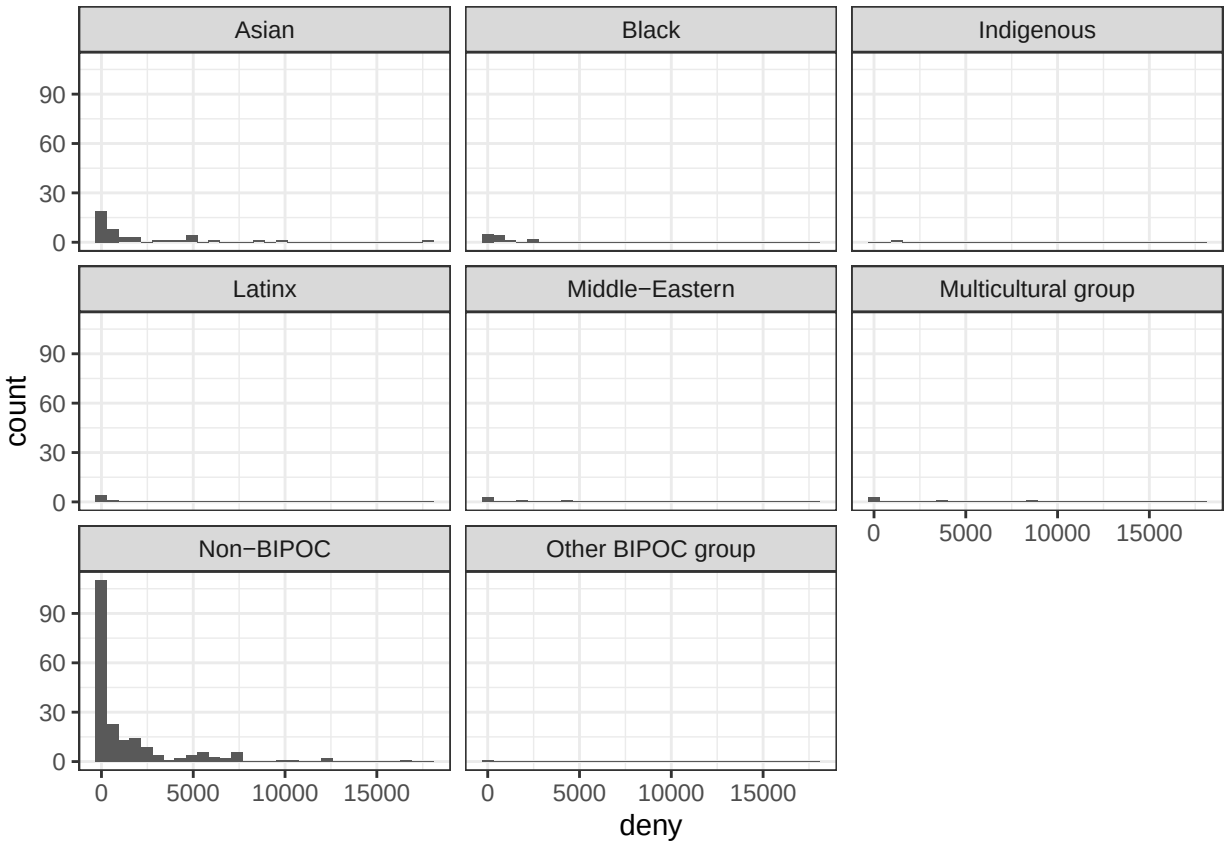

```
ggplot(prog, aes(x = deny)) +
  geom_histogram(aes(fill = factor(community, levels=c("Asian", "Black", "Indigenous", "Latinx", "Middl
                 position = "stack", color = "white") +
  scale_fill_discrete(name = "community") +
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(prog, aes(x = deny)) +
  geom_histogram() +
  facet_wrap(. ~ community) +
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
aggregate(prog$prop_grant, list(prog$org), mean) %>%
  arrange(desc(x)) %>%
  head(10)
```

```
##                                    Group.1 x
## 1                          Acapella Council 1
## 2                       Amnesty International 1
## 3                    Asian American Alliance  1
## 4  Asian Intervarsity Christian Fellowship 1
## 5                                 Brownstone 1
## 6                         CrossFit Blue Devil 1
## 7                               Devilish Keys 1
## 8                         Duke Amandla Chorus 1
## 9                                Duke Archery 1
## 10                        Duke Chinese Dance  1
```

```
aggregate(prog$prop_grant, list(prog$community), mean) %>%
  arrange(desc(x))
```

```
##                Group.1         x
## 1               Latinx 0.9208275
## 2                Black 0.7992408
## 3            Non-BIPOC 0.7822895
## 4 Multicultural group 0.7363779
## 5                Asian 0.7292228
```

```
## 6      Middle-Eastern 0.7039526
## 7          Indigenous 0.6465517
## 8   Other BIPOC group 0.2293907
```

```
aggregate(prog$grant, list(prog$org), sum) %>%
  arrange(desc(x)) %>%
  head(10)
```

```
##                          Group.1        x
## 1             Blue Devils United 53714.07
## 2      Asian Students Association 35721.00
## 3             Blue Devils United  33139.00
## 4           Duke Catholic Center 30565.61
## 5           Duke Chinese Theater 28713.25
## 6         Duke Conservation Tech 25905.50
## 7                       TEDxDuke 25895.00
## 8    National Panhellenic Council 23179.35
## 9                     Duke Diya 21137.75
## 10 Singapore Students Association 20680.00
```

```
aggregate(prog$grant, list(prog$community), sum) %>%
  arrange(desc(x))
```

```
##               Group.1         x
## 1           Non-BIPOC 665802.74
## 2               Asian 147190.64
## 3               Black  41542.10
## 4 Multicultural group  21480.28
## 5      Middle-Eastern  14175.00
## 6              Latinx  12215.00
## 7          Indigenous   1875.00
## 8   Other BIPOC group     64.00
```

```
aggregate(prog$deny, list(prog$community), sum) %>%
  arrange(desc(x))
```

```
##               Group.1         x
## 1           Non-BIPOC 295114.14
## 2               Asian  88525.03
## 3 Multicultural group  12045.00
## 4               Black   8078.00
## 5      Middle-Eastern   6572.00
## 6              Latinx   1114.99
## 7          Indigenous   1025.00
## 8   Other BIPOC group    215.00
```

```
# ANOVA for programming funds
model_bipoc <- lm(prop_grant ~ bipoc, data = prog)
kbl(model_bipoc %>% tidy(conf.int=TRUE),digits=3)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.777	0.020	38.242	0.000	0.737	0.816
bipocY	-0.013	0.035	-0.363	0.717	-0.083	0.057

```
kbl(tidy(aov(model_bipoc)),digits=3)
```

term	df	sumsq	meansq	statistic	p.value
bipoc	1	0.010	0.010	0.132	0.717
Residuals	273	20.823	0.076	NA	NA

```
model_comm <- lm(prop_grant ~ community,data=prog)
kbl(tidy(aov(model_comm)),digits=3)
```

term	df	sumsq	meansq	statistic	p.value
community	7	0.561	0.080	1.056	0.393
Residuals	267	20.272	0.076	NA	NA

```
# ANOVA for budget funds
model_bipoc <- lm(prop_grant ~ bipoc, data = budget)
kbl(model_bipoc %>% tidy(conf.int=TRUE),digits=3)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.462	0.018	25.054	0.000	0.426	0.498
bipocY	-0.058	0.037	-1.569	0.118	-0.131	0.015

```
kbl(tidy(aov(model_bipoc)),digits=3)
```

term	df	sumsq	meansq	statistic	p.value
bipoc	1	0.175	0.175	2.46	0.118
Residuals	276	19.597	0.071	NA	NA

```
model_comm <- lm(prop_grant ~ community,data=budget)
kbl(tidy(aov(model_comm)),digits=3)
```

term	df	sumsq	meansq	statistic	p.value
community	6	0.588	0.098	1.384	0.221
Residuals	271	19.184	0.071	NA	NA

```
# ANOVA for SOFC programming totals (right now this is only 2017-2018)
model_bipoc <- lm(prop_grant ~ bipoc, data = sofc)
kbl(model_bipoc %>% tidy(conf.int=TRUE),digits=3)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.635	0.067	9.419	0.00	0.496	0.775
bipocY	-0.112	0.112	-0.995	0.33	-0.344	0.121

```
kbl(tidy(aov(model_bipoc)),digits=3)
```

term	df	sumsq	meansq	statistic	p.value
bipoc	1	0.072	0.072	0.99	0.33
Residuals	23	1.675	0.073	NA	NA

```
model_comm <- lm(prop_grant ~ community,data=sofc)
kbl(tidy(aov(model_comm)),digits=3)
```

term	df	sumsq	meansq	statistic	p.value
community	3	0.090	0.030	0.38	0.769
Residuals	21	1.657	0.079	NA	NA