

initial-eda

Lillian Clark

11/8/2020

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

```
## -- Attaching packages -----
## v tibble  3.0.3      v purrr  0.3.4
## v tidyr   1.1.1      v dplyr  1.0.1
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts -----
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()       masks base::date()
## x dplyr::filter()         masks stats::filter()
## x readr::guess_encoding() masks rvest::guess_encoding()
## x lubridate::intersect()  masks base::intersect()
## x dplyr::lag()            masks stats::lag()
## x purrr::pluck()          masks rvest::pluck()
## x lubridate::setdiff()    masks base::setdiff()
## x lubridate::union()      masks base::union()
```

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

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

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

```
prog <- read.csv("data/programming.csv")
prog <- prog[2:6]
```

```
prog_cat <- prog %>%
  mutate(bipoc = case_when(
    org %in% c("Asian American Alliance",
              "Alpha Kappa Alpha Sorority, Inc.",
              "Alpha Phi Alpha Fraternity, Inc.",
              "Asian Interversity Christian Fellowship", #preetha added
              "Duke Amandla Chorus",
              "Duke Africa", #preetha added
              "Asian Students Association",
              "Black Student Alliance",
```

```

"Duke Chinese Theater",
"Duke Chinese Dance ", #preetha added
"Duke Sangeet", #preetha added
"Students of the Caribbean Association",
"Duke Chinese Student Association", # I added
"Singapore Students Association", # I added
"Duke Africa", # I added
"Black Men's Union", # I added
"Delta Sigma Theta Sorority, Inc.",
"Duke Dhamaka",
"Pureun", #preetha added
"Duke East Asia Nexus", #preetha added
"Duke East Asian Nexus", #preetha added
"Duke Diya", #preetha added
"Hindu Students Association",
"Kappa Alpha Psi Fraternity, Inc.",
"Lambda Theta Alpha Latin Sorority, Inc.",
"La Unidad Latina Lambda Upsilon Lambda Fraternity, Inc.",
"Mi Gente",
"Multicultural Greek Council", #preetha added
"Minority Association of Premedical Students", #preetha added
"Nakisai African Dance Ensemble",
"National Society of Black Engineers",
"Native American Student Organization", #preetha added
"Duke Rhydhun",
"Taiwanese American Students Association",
"Duke Southeast Asian Students Association", #preetha added
"The Bridge",
"United in Praise",
"Zeta Phi Beta Sorority, Inc.",
"Duke Nepali Student Association",
"Duke Ethiopian/Eritrean Student Transactional Association",
"Desarrolla",
"Sabrosura", #preetha added
"Gente Aprendiendo para Nuevas Oportunidades",
"Project H.E.A.L. (Health Education and Awareness in Latin America)",
"Pakistani Students Association",
"Duke CommuniTEA",
"Duke Association for the Middle East",
"International Association",
"Muslim Students Association",
"Duke Sikh Society",
"Students for Justice in Palestine") ~ "Y",
TRUE ~ "N"))

```

```

prog_cat <- prog_cat %>%
  mutate(community = case_when(
    org %in% c("Asian American Alliance",
              "Asian Students Association",
              "Asian Intersarsity Christian Fellowship",
              "Duke Chinese Theater",
              "Duke Chinese Dance ", #preetha added
              "Duke Dhamaka",
              "Duke Diya",

```

```

    "Hindu Students Association",
    "Asian Intersity Christian Fellowship", #preetha added
    "Duke Rhydhun",
    "Pureun", #preetha added
    "Duke East Asia Nexus", #preetha added
    "Duke East Asian Nexus", #preetha added
    "Taiwanese American Student Association",
    "Duke CommuniTEA",
    "Duke Sikh Society",
    "Duke Nepali Student Association",
    "Pakistani Students Association",
    "Duke Southeast Asian Students Association",
    "Duke Chinese Student Association",
    "Duke Sangeet",
    "Singapore Students Association") ~ "Asian",
  org %in% c("Alpha Kappa Alpha Sorority, Inc.",
    "Alpha Phi Alpha Fraternity, Inc.",
    "Duke Amandla Chorus",
    "Black Student Alliance",
    "Delta Sigma Theta Sorority, Inc.",
    "Kappa Alpha Psi Fraternity, Inc.",
    "Nakisai African Dance Ensemble",
    "National Society of Black Engineers",
    "United in Praise",
    "Zeta Phi Beta Sorority Inc.",
    "Duke Ethiopian/Eritrean Student Transactional Association",
    "Duke Africa",
    "Black Men's Union") ~ "Black",
  org %in% c("Lambda Theta Alpha Latin Sorority, Inc.",
    "La Unidad Latina Lambda Upsilon Lambda Fraternity, Inc.",
    "Mi Gente",
    "Sabrosura",
    "Gente Aprendiendo para Nuevas Oportunidades",
    "Project H.E.A.L. (Health Education and Awareness in Latin America)",
    "Desarrolla") ~ "Latinx",
  org %in% c("Duke Association for the Middle East",
    "Duke Muslim Students Association",
    "Duke Students for Justice in Palestine") ~ "Middle-Eastern",
  org %in% c("Native American Student Organization") ~ "Indigenous",
  org %in% c("Students of the Caribbean Association",
    "Multicultural Greek Council", #preetha added
    "Minority Association of Premedical Students", #preetha added
    "The Bridge",
    "United in Praise",
    "International Association") ~ "Other BIPOC group",
  TRUE ~ "Nonspecific"
))

```

```

prog_cat <- prog_cat %>%
  filter(!is.na(date), deny >= 0) %>%
  mutate(prop_grant = grant / req,
    year = year(date),
    month = month(date),
    sem = case_when(

```

```

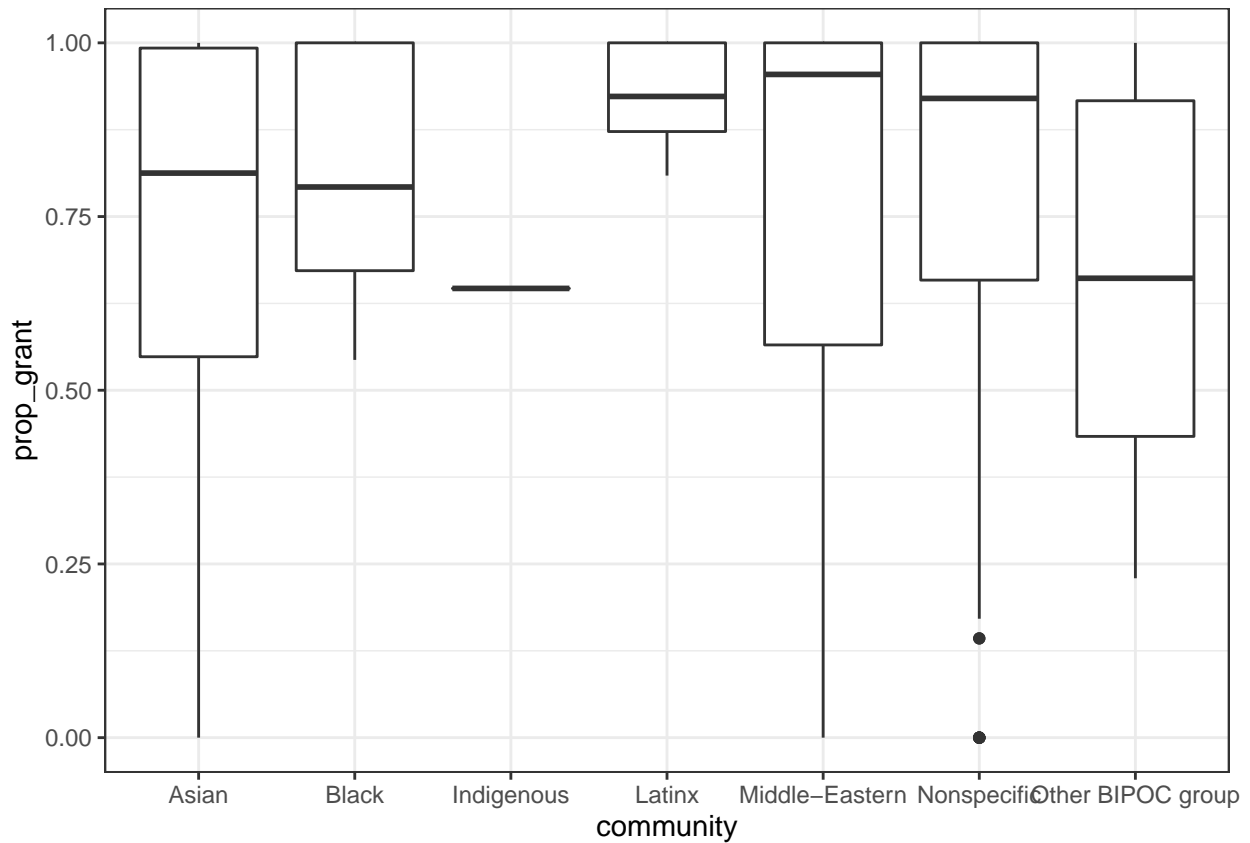
    month %in% c(1, 2, 3, 4, 5, 6) ~ "Spring",
    month %in% c(7, 8, 9, 10, 11, 12) ~ "Fall"),
  schoolyr = case_when(
    year == 2016 & sem == "Fall" ~ "2016-2017",
    year == 2017 & sem == "Spring" ~ "2016-2017",
    year == 2017 & sem == "Fall" ~ "2017-2018",
    year == 2018 & sem == "Spring" ~ "2017-2018",
    year == 2018 & sem == "Fall" ~ "2018-2019",
    year == 2019 & sem == "Spring" ~ "2018-2019",
    year == 2019 & sem == "Fall" ~ "2019-2020",
    year == 2020 & sem == "Spring" ~ "2019-2020"
  ))

```

```

ggplot(prog_cat, aes(x = community, y = prop_grant)) +
  geom_boxplot() +
  theme_bw()

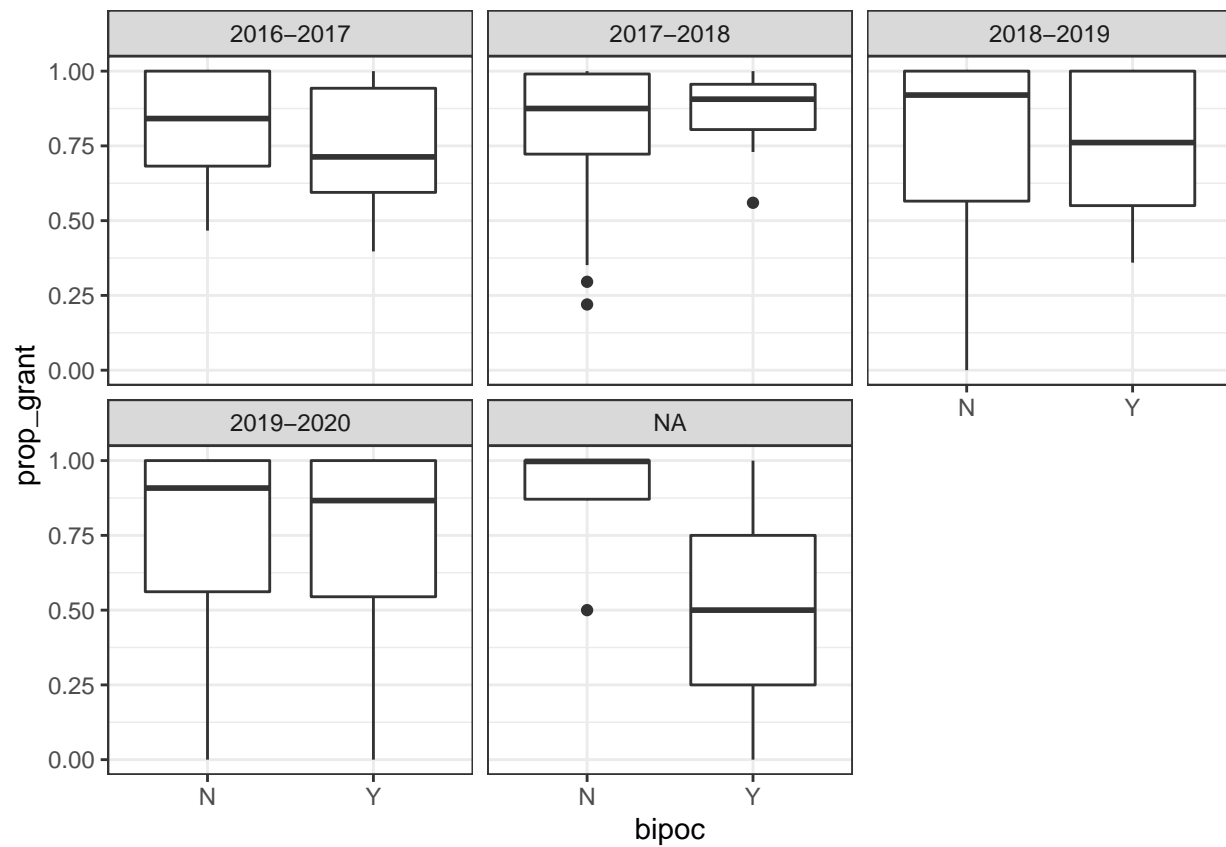
```



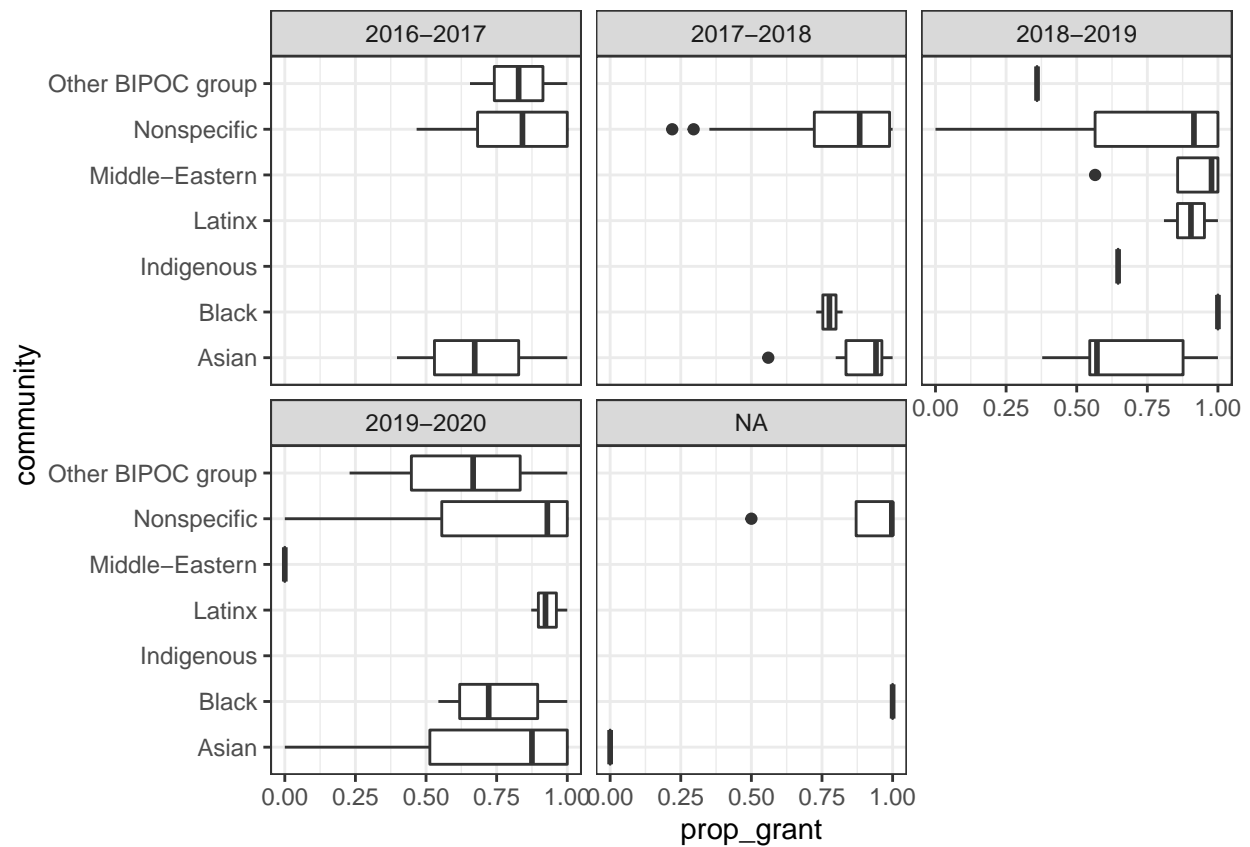
```

ggplot(prog_cat, aes(x = bipoc, y = prop_grant)) +
  geom_boxplot() +
  facet_wrap(. ~ schoolyr) +
  theme_bw()

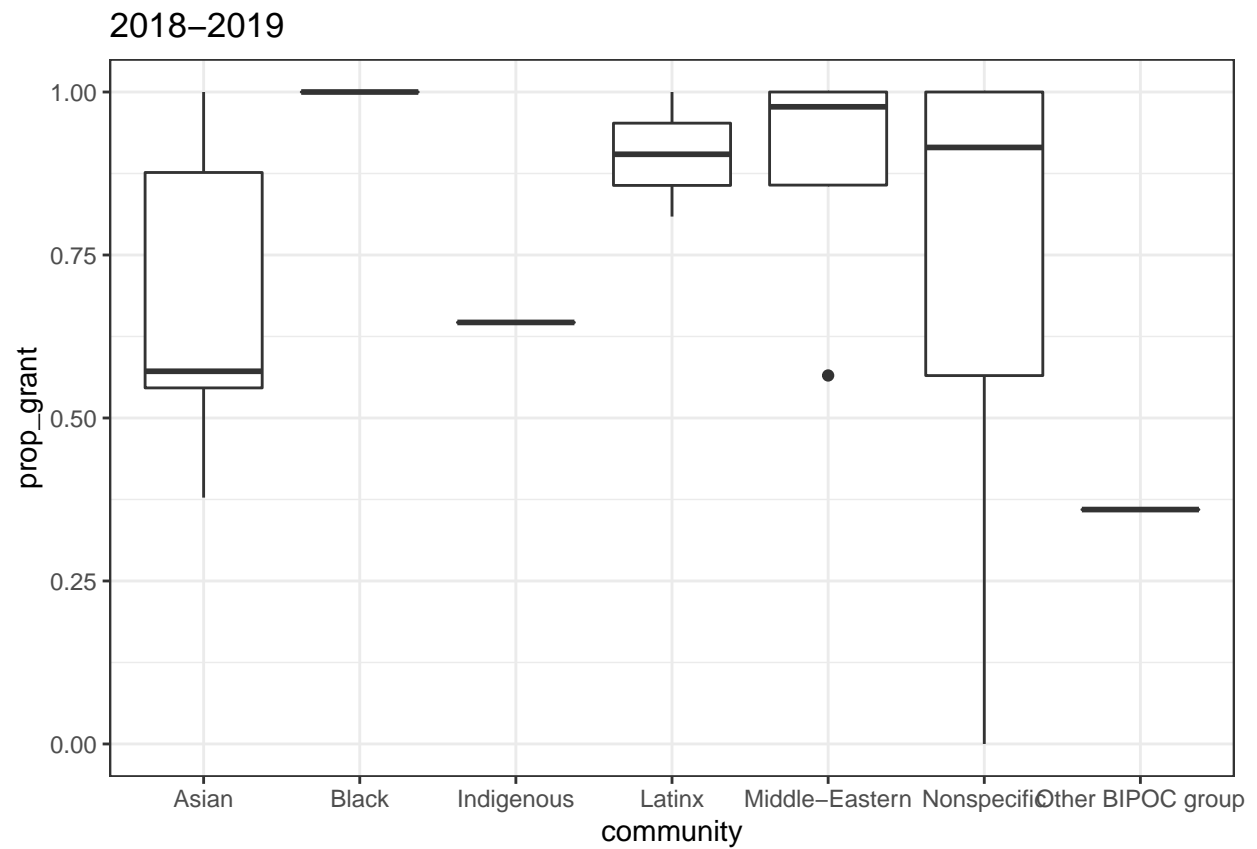
```



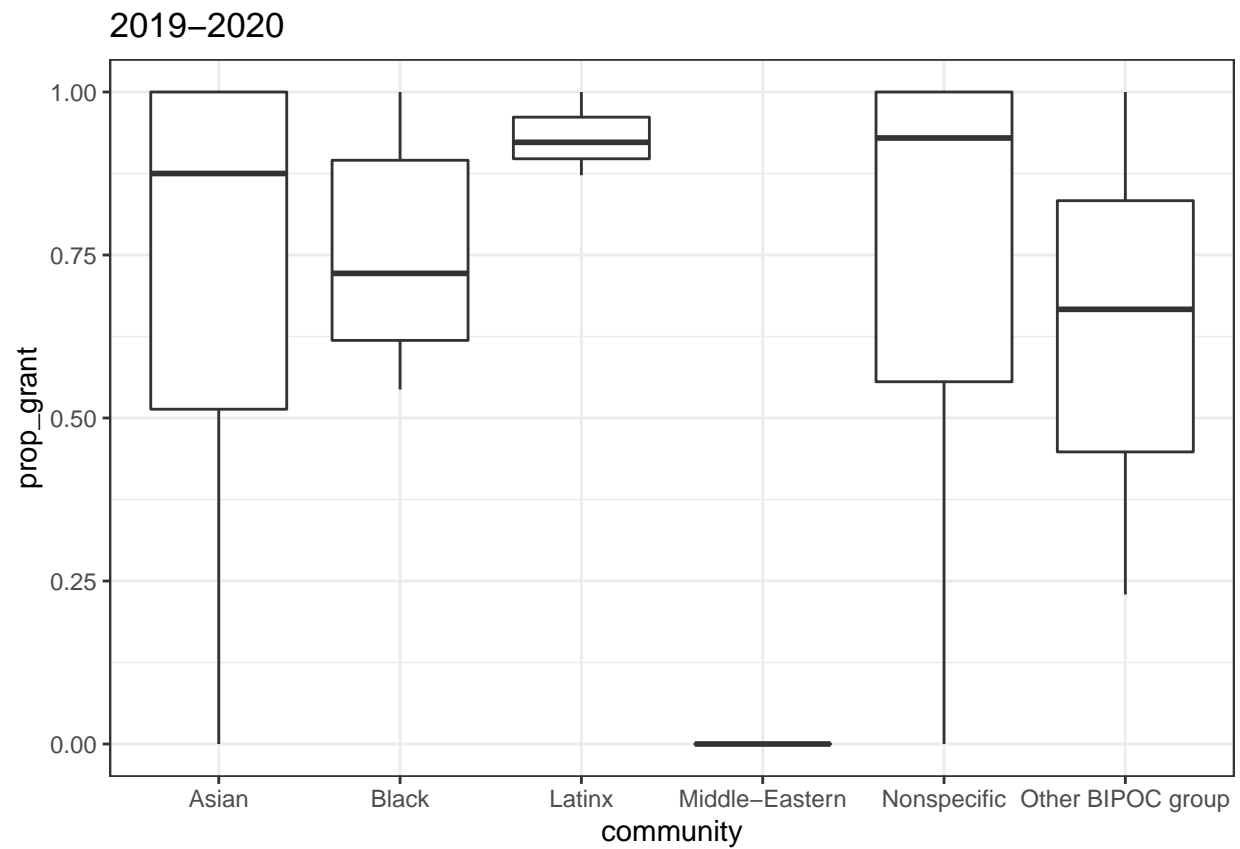
```
ggplot(prog_cat, aes(x = community, y = prop_grant)) +
  geom_boxplot() +
  coord_flip() +
  facet_wrap(. ~ schoolyr) +
  theme_bw()
```



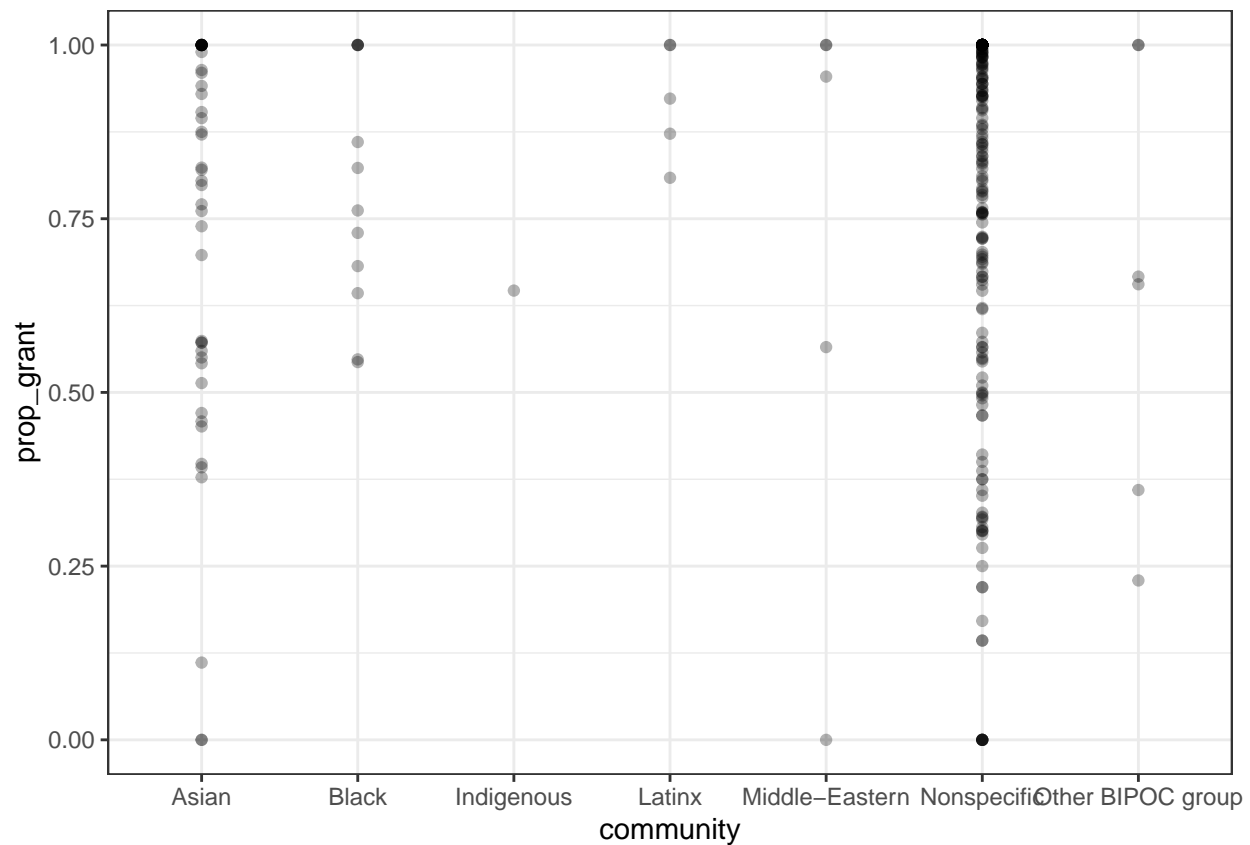
```
prog_cat %>%
  filter(schoolyr == "2018-2019") %>%
  ggplot(aes(x = community, y = prop_grant)) +
  labs(title = "2018-2019") +
  geom_boxplot() +
  theme_bw()
```



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

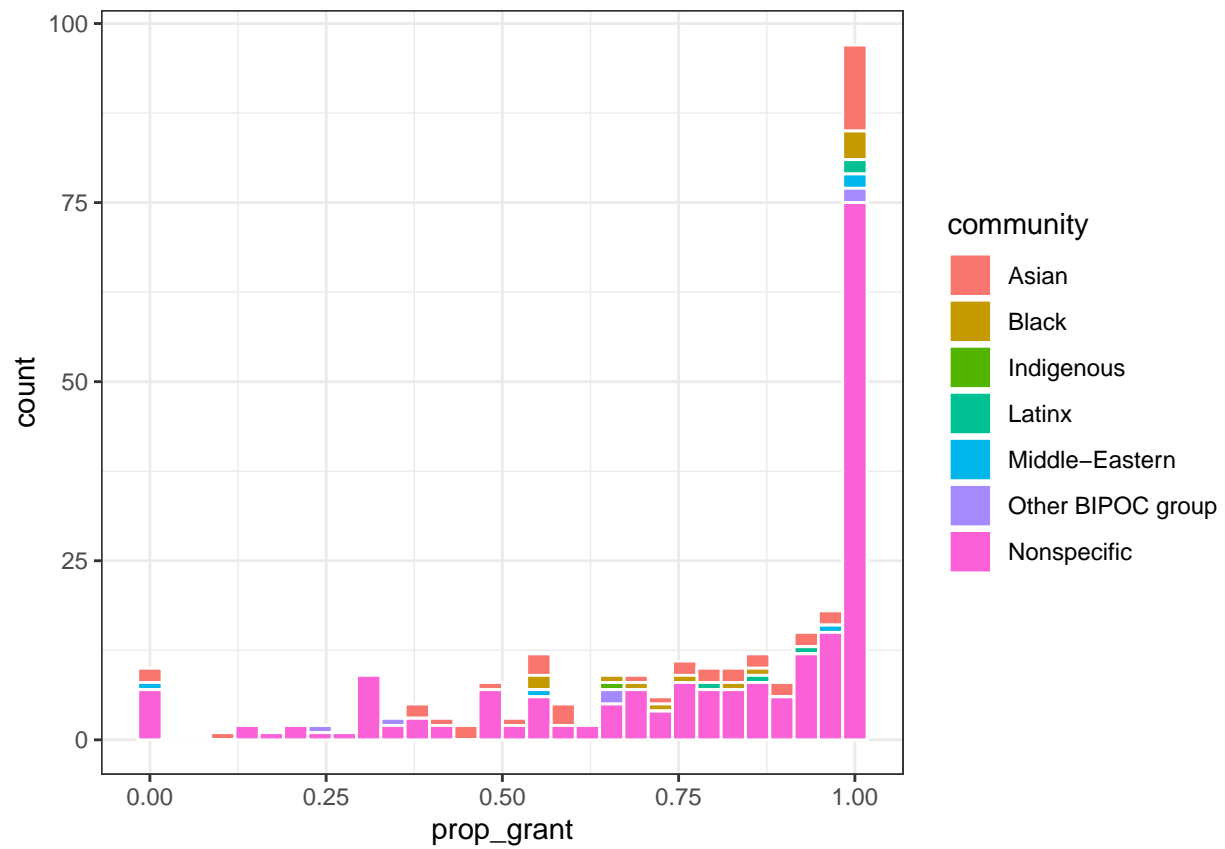


```
ggplot(prog_cat, aes(x = community, y = prop_grant)) +  
  geom_point(alpha = 0.3) +  
  theme_bw()
```

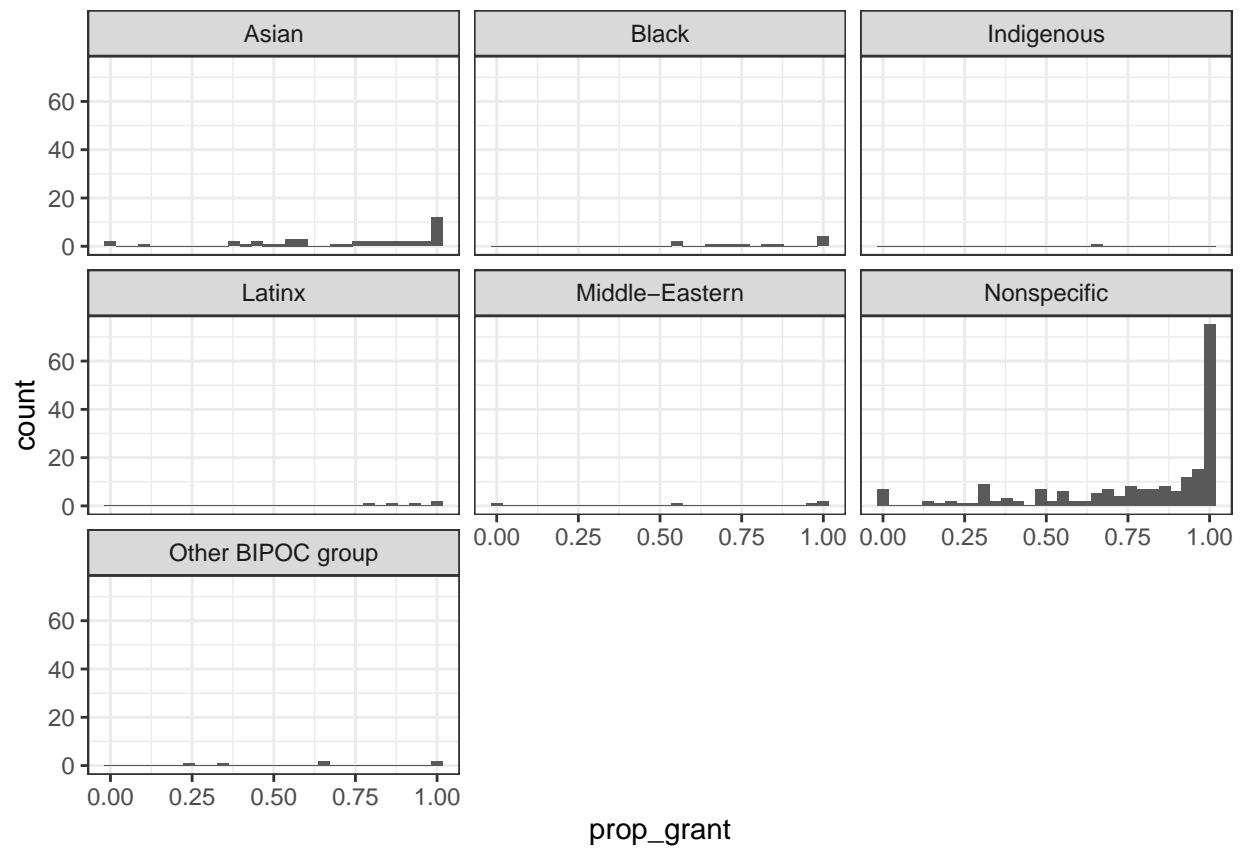
```
ggplot(prog_cat, aes(x = prop_grant)) +
  geom_histogram(aes(fill = factor(community, levels=c("Asian", "Black", "Indigenous", "Latinx", "Middle-Eastern", "Nonspecific", "Other BIPOC group")),
                    position = "stack", color = "white")) +
  scale_fill_discrete(name = "community") +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

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

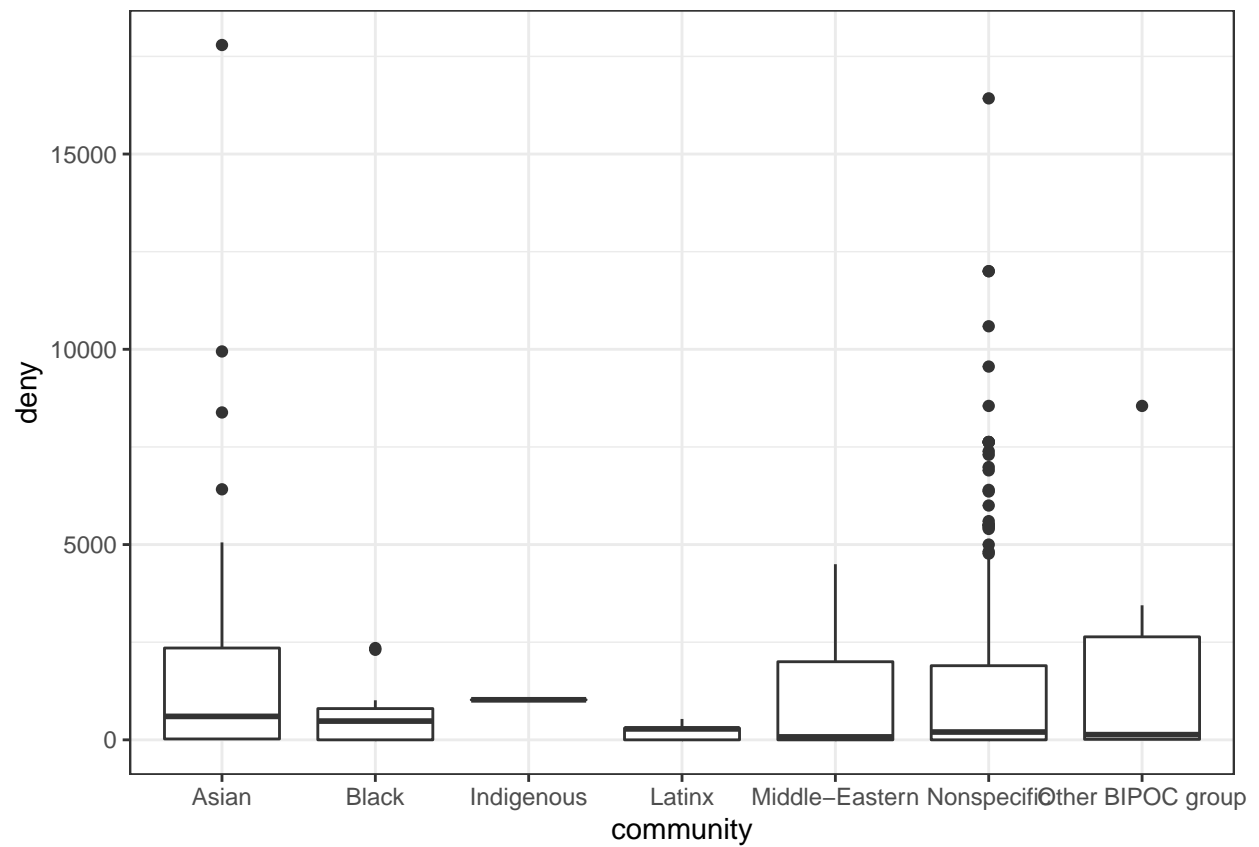


```
ggplot(prog_cat, aes(x = prop_grant)) +
  geom_histogram() +
  facet_wrap(. ~ community) +
  theme_bw()
```

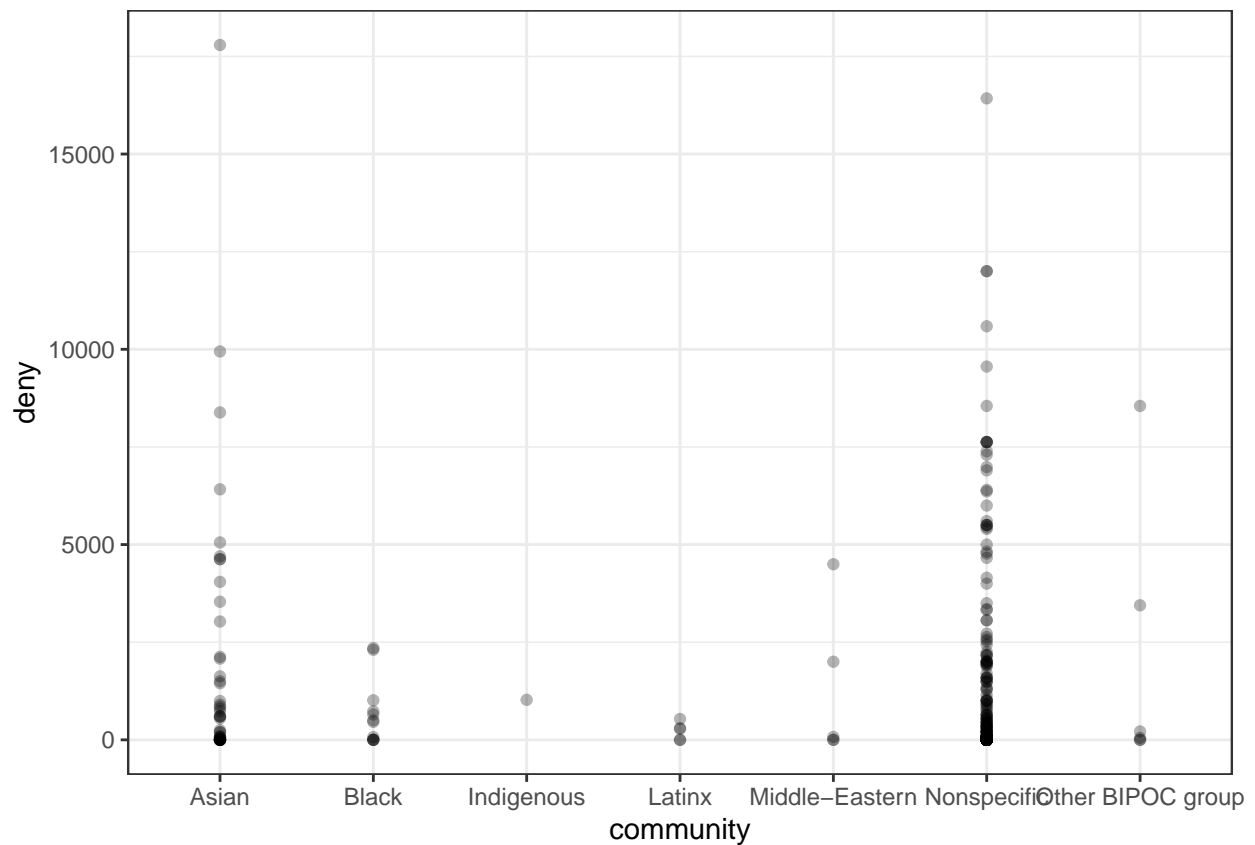
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



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

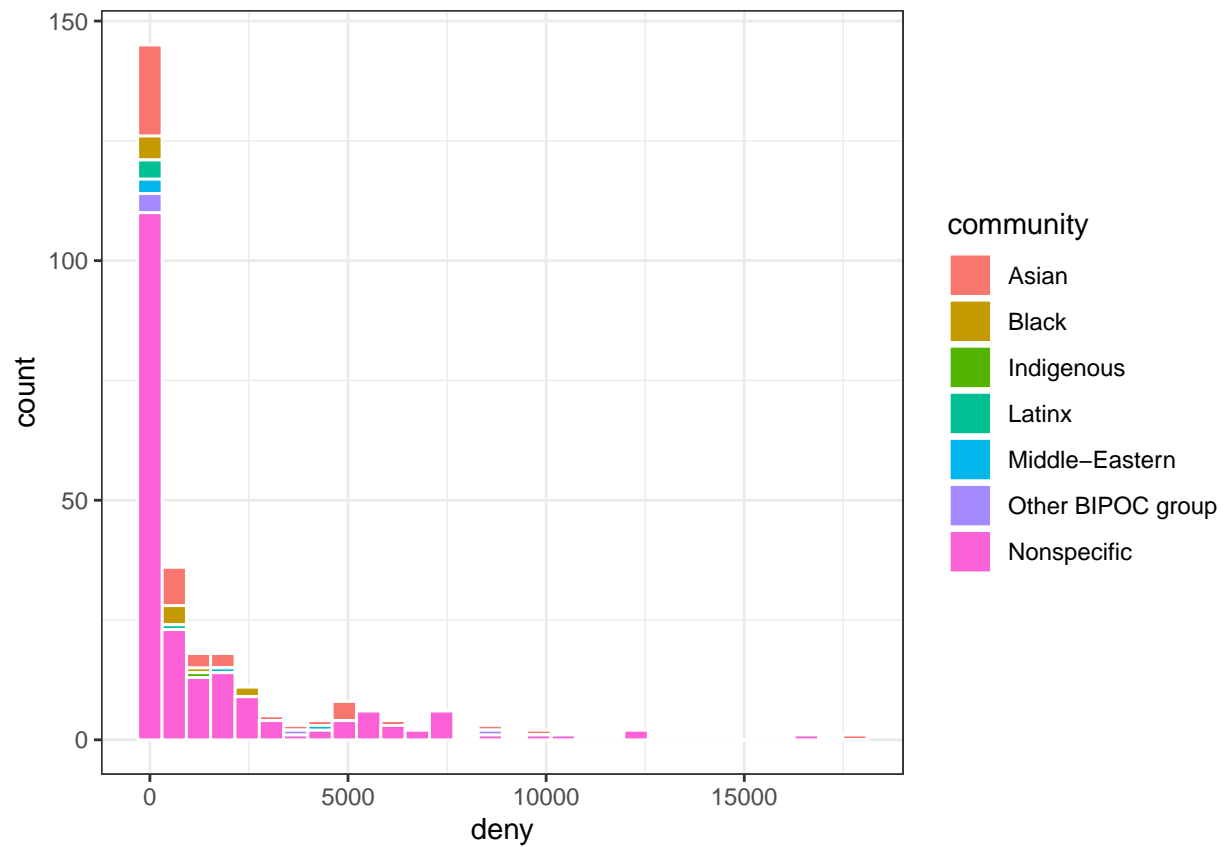


```
ggplot(prog_cat, aes(x = community, y = deny)) +
  geom_point(alpha = 0.3) +
  theme_bw()
```



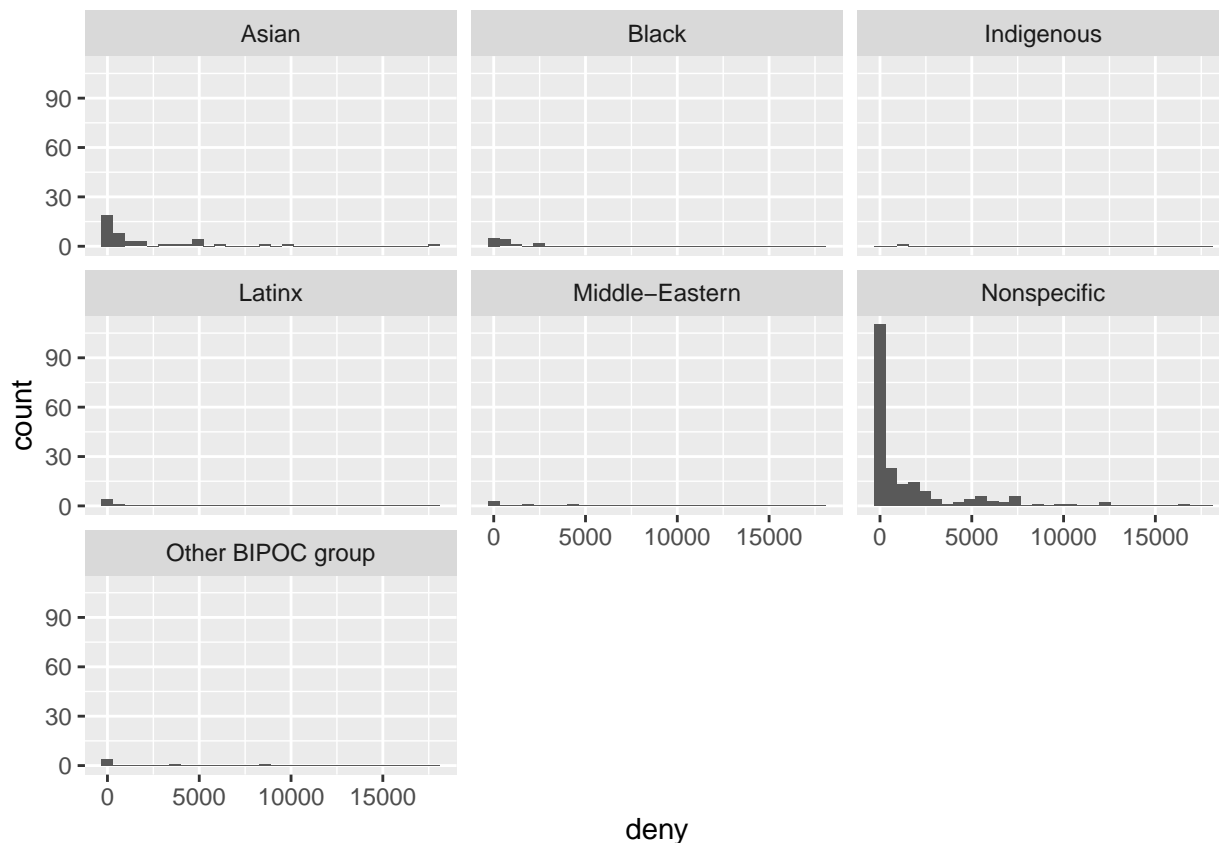
```
ggplot(prog_cat, aes(x = deny)) +
  geom_histogram(aes(fill = factor(community, levels=c("Asian", "Black", "Indigenous", "Latinx", "Middle-Eastern", "Nonspecific", "Other BIPOC group"),
    position = "stack", color = "white")) +
  scale_fill_discrete(name = "community") +
  theme_bw()
```

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



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

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



```
aggregate(prog_cat$prop_grant, list(prog_cat$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 Intersvarsity 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_cat$prop_grant, list(prog_cat$community), mean) %>%
  arrange(desc(x))
```

```
##           Group.1      x
## 1           Latinx 0.9208275
## 2           Black 0.7992408
## 3       Nonspecific 0.7802070
## 4           Asian 0.7292228
## 5 Middle-Eastern 0.7039526
## 6 Other BIPOC group 0.6518800
## 7       Indigenous 0.6465517
```

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

```
##           Group.1      x
## 1      Nonspecific 670602.74
## 2           Asian 147190.64
## 3           Black 41542.10
## 4 Other BIPOC group 21544.28
## 5    Middle-Eastern 14175.00
## 6           Latinx 12215.00
## 7      Indigenous 1875.00
```

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

```
##           Group.1      x
## 1      Nonspecific 303664.14
## 2           Asian 88525.03
## 3 Other BIPOC group 12260.00
## 4           Black 8078.00
## 5    Middle-Eastern 6572.00
## 6           Latinx 1114.99
## 7      Indigenous 1025.00
```

```
model_bipoc <- lm(prop_grant ~ bipoc, data = prog_cat)
kbl(model_bipoc %>% tidy(conf.int=TRUE), digits=3)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.781	0.020	39.853	0.000	0.742	0.819
bipocY	-0.036	0.037	-0.969	0.333	-0.109	0.037

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

term	df	sumsq	meansq	statistic	p.value
bipoc	1	0.072	0.072	0.939	0.333
Residuals	274	20.931	0.076	NA	NA

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


term	df	sumsq	meansq	statistic	p.value
community	6	0.339	0.056	0.735	0.622
Residuals	269	20.664	0.077	NA	NA