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)
```

```
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.",
               "Duke Amandla Chorus",
               "Asian Students Association",
               "Black Student Alliance",
               "Duke Chinese Theater",
```

```
              "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",
              "Hindu Students Association",
              "Kappa Alpha Psi Fraternity, Inc.",
              "Lambda Theta Alpha Latin Sorority, Inc.",
              "La Unidad Latina Lambda Upsilon Lambda Fraternity, Inc.",
              "Mi Gente",
              "Nakisai African Dance Ensemble",
              "National Society of Black Engineers",
              "Duke Rhydhun",
              "Taiwanese American Student Association",
              "The Bridge",
              "United in Praise",
              "Zeta Phi Beta Sorority, Inc.",
              "Duke Nepali Student Association",
              "Duke Ethiopian/Eritrean Student Transactional Association",
              "Desarrolla",
              "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",
              "Duke Muslim Students Association",
              "Duke Sikh Society",
              "Duke Students for Justice in Palestine") ~ "Y",
    TRUE ~ "N"))
```

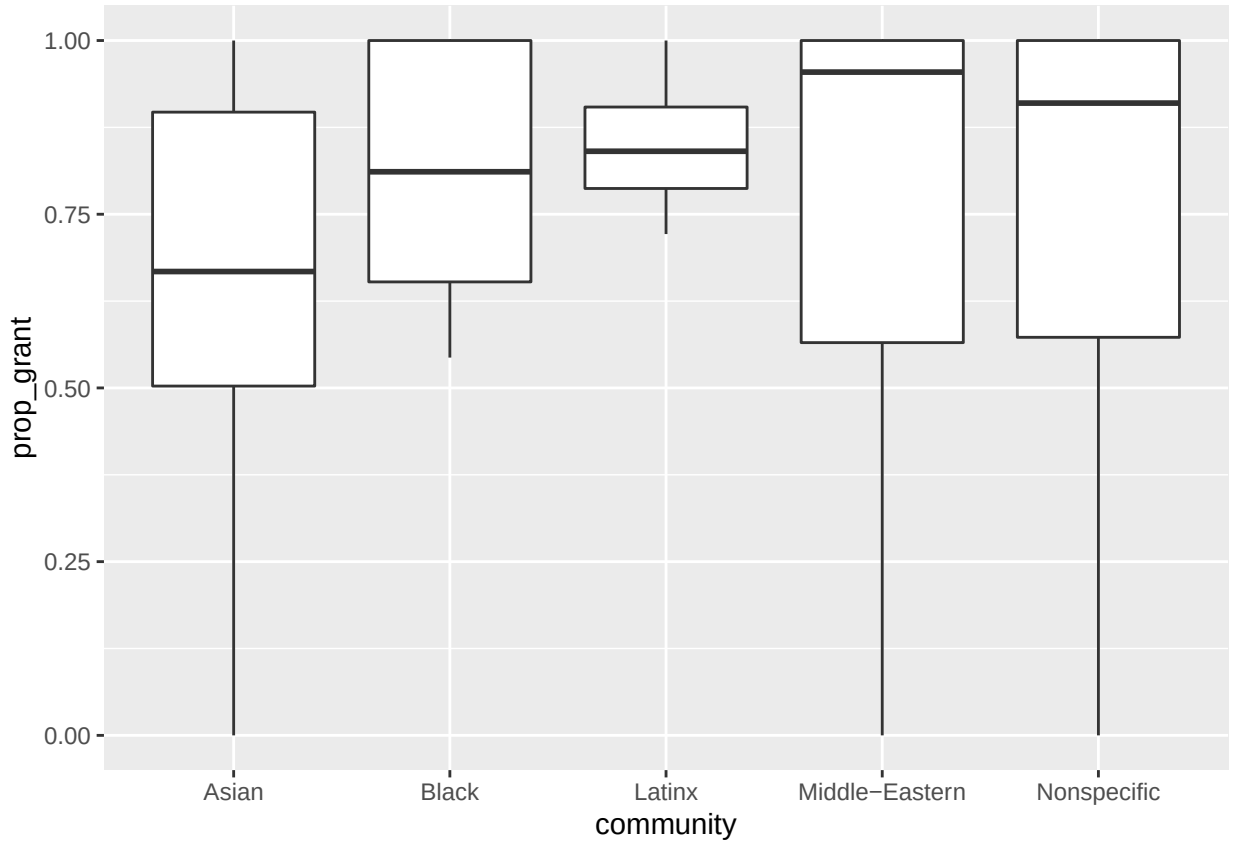
```
prog_cat <- prog_cat %>%
  mutate(community = case_when(
    org %in% c("Asian American Alliance",
               "Asian Students Association",
               "Duke Chinese Theater",
               "Duke Dhamaka",
               "Hindu Students Association",
               "Duke Rhydhun",
               "Taiwanese American Student Association",
               "Duke CommuniTEA",
               "Duke Sikh Society",
               "Duke Nepali Student Association",
               "Pakistani Students Association",
               "Duke Chinese Student Association",
               "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",
              "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",
    TRUE ~ "Nonspecific"
  ))
```

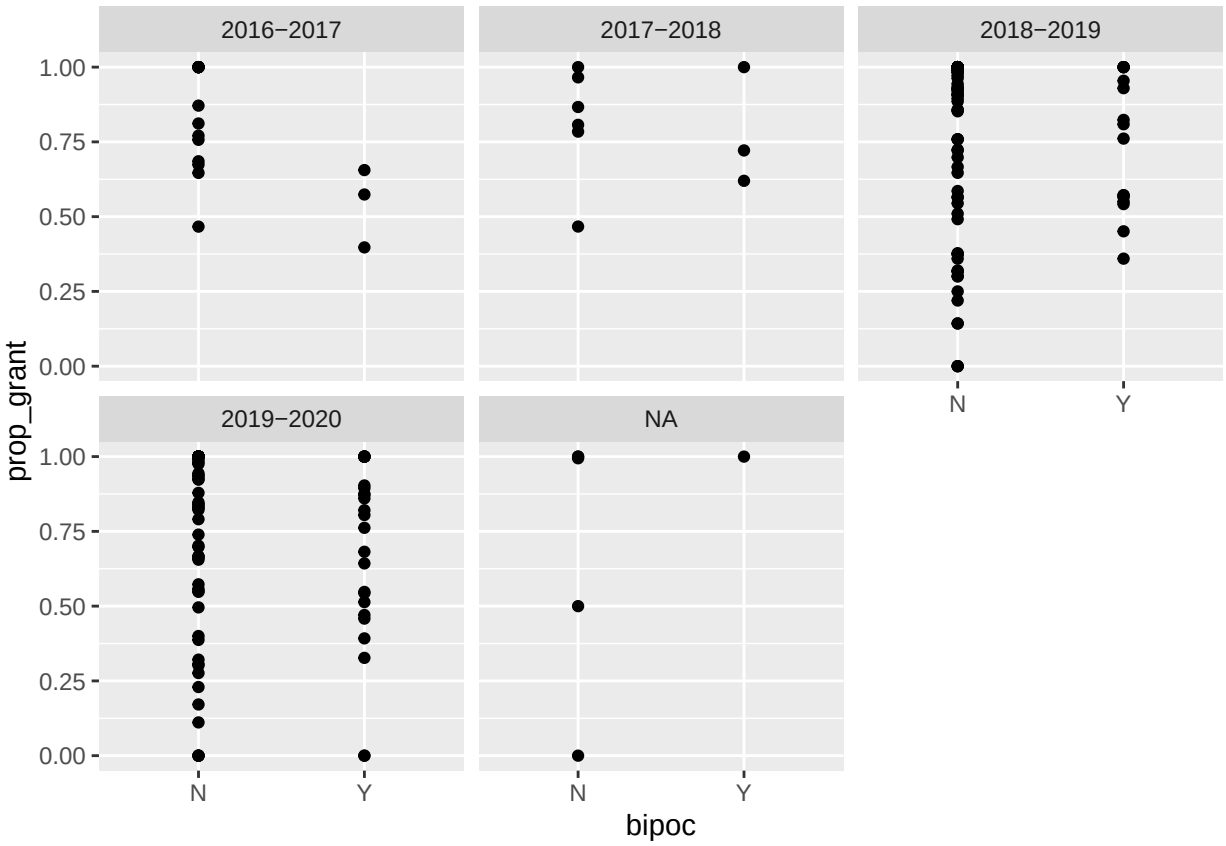
```
prog_cat <- prog_cat %>%
  filter(!is.na(date), deny >= 0) %>%
  mutate(prop_grant = grant / req,
         year = year(date),
         month = month(date),
         sem = case_when(
           month %in% c(1, 2, 3, 4, 5, 6) ~ "Spring",
           month %in% c(7, 8, 9, 10, 11, 12) ~ "Fall"),
         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"
         ))
```

```
## Warning: Problem with `mutate()` input `year`.
## i tz(): Don't know how to compute timezone for object of class factor; returning "UTC". This warning
## i Input `year` is `year(date)`.

## Warning: tz(): Don't know how to compute timezone for object of class factor;
## returning "UTC". This warning will become an error in the next major version of
## lubridate.

## Warning: Problem with `mutate()` input `month`.
## i tz(): Don't know how to compute timezone for object of class factor; returning "UTC". This warning
## i Input `month` is `month(date)`.

## Warning: tz(): Don't know how to compute timezone for object of class factor;
## returning "UTC". This warning will become an error in the next major version of
## lubridate.
```

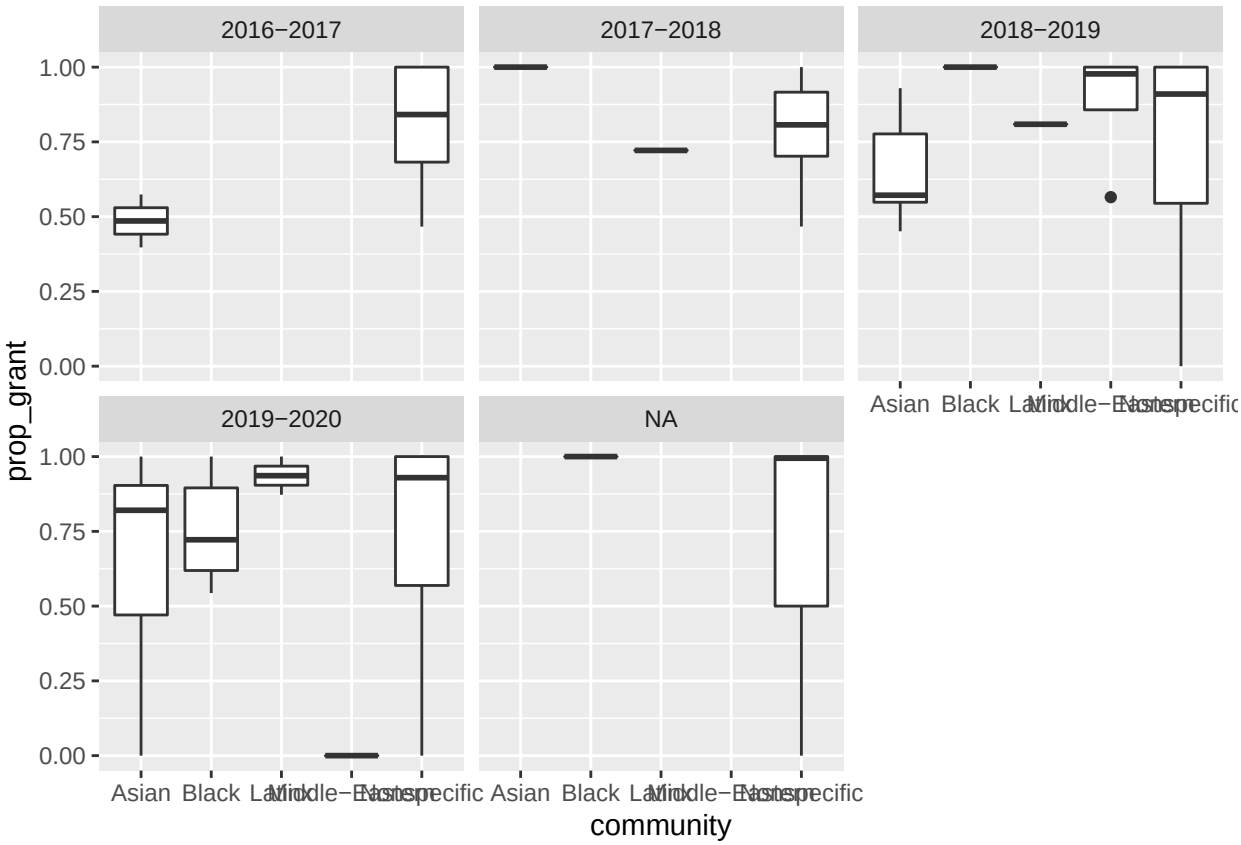
```
ggplot(prog_cat, aes(x = community, y = prop_grant)) +
  geom_boxplot()
```



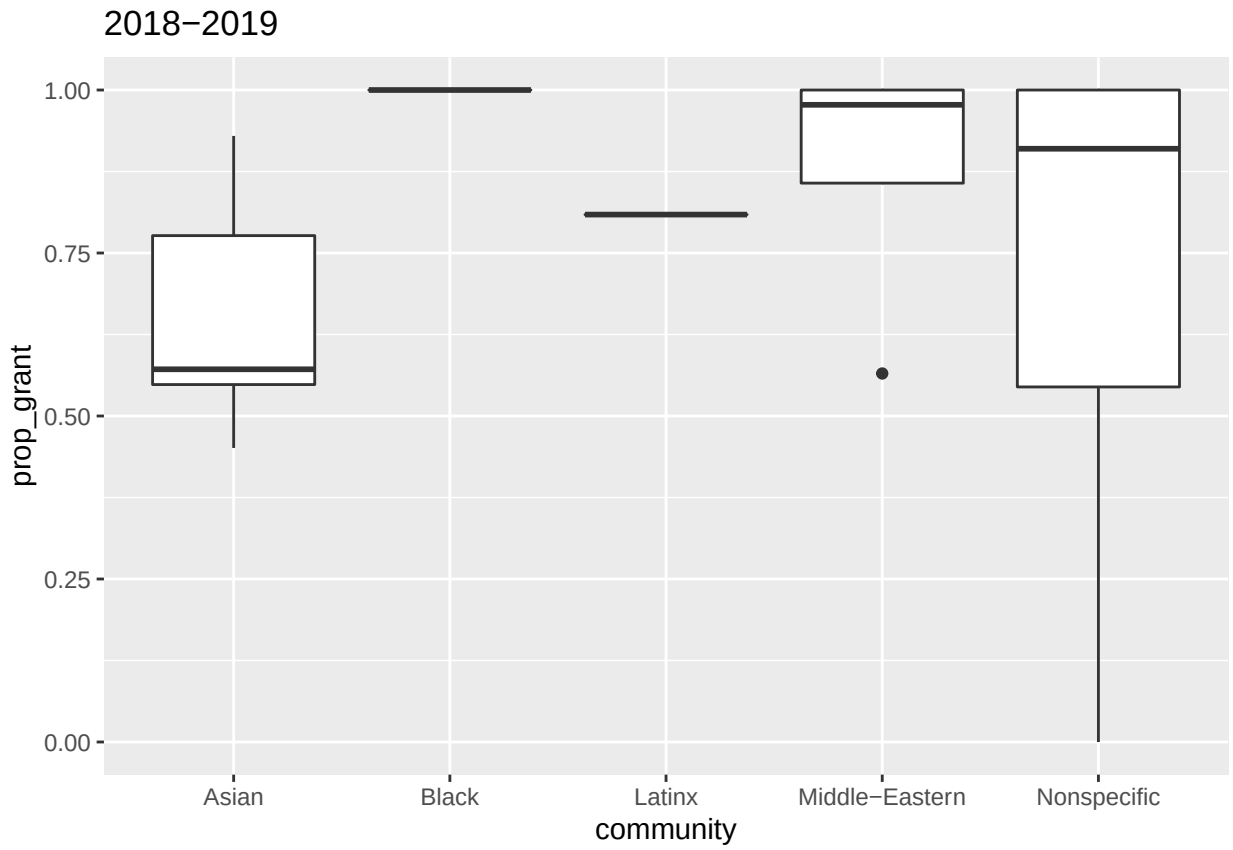

```
ggplot(prog_cat, aes(x = bipoc, y = prop_grant)) +
  geom_point() +
  facet_wrap(. ~ schoolyr)
```



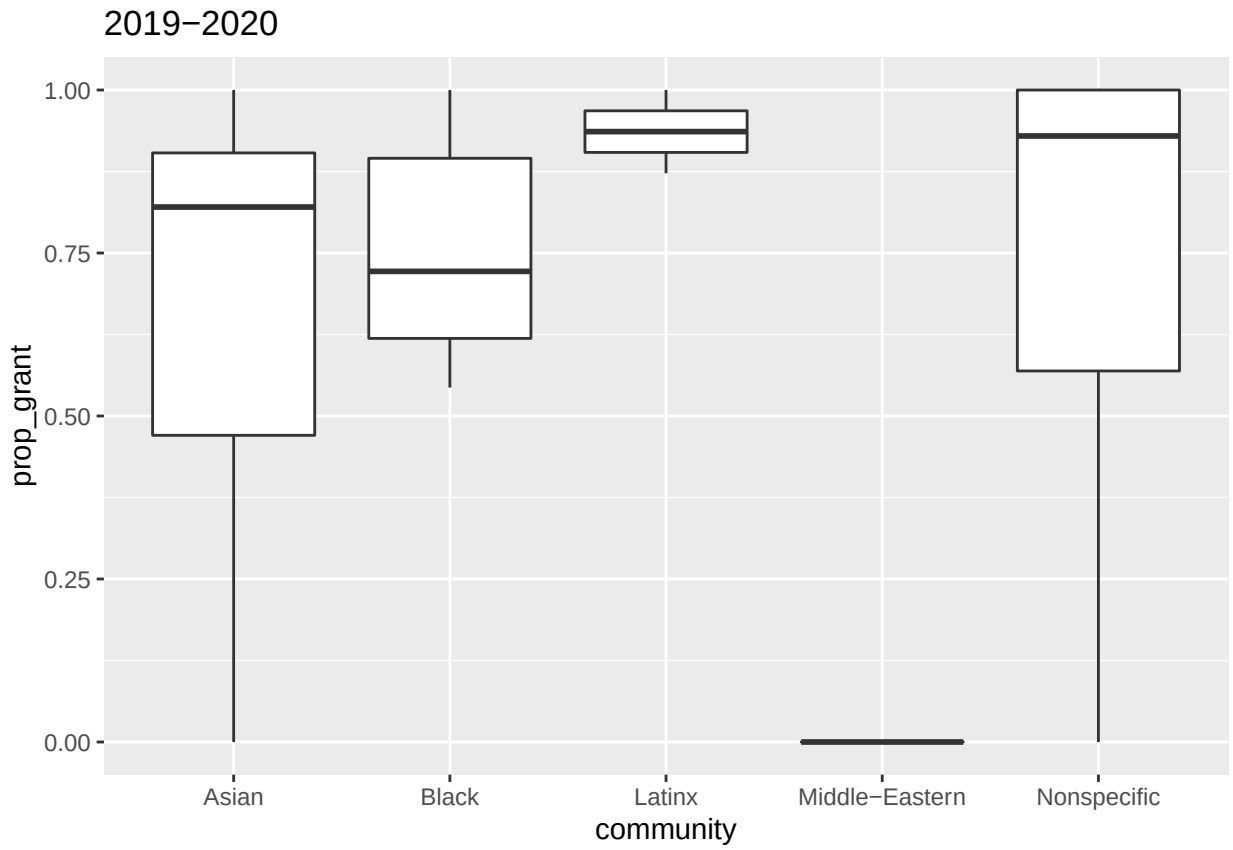

```
ggplot(prog_cat, aes(x = community, y = prop_grant)) +
  geom_boxplot() +
  facet_wrap(. ~ schoolyr)
```



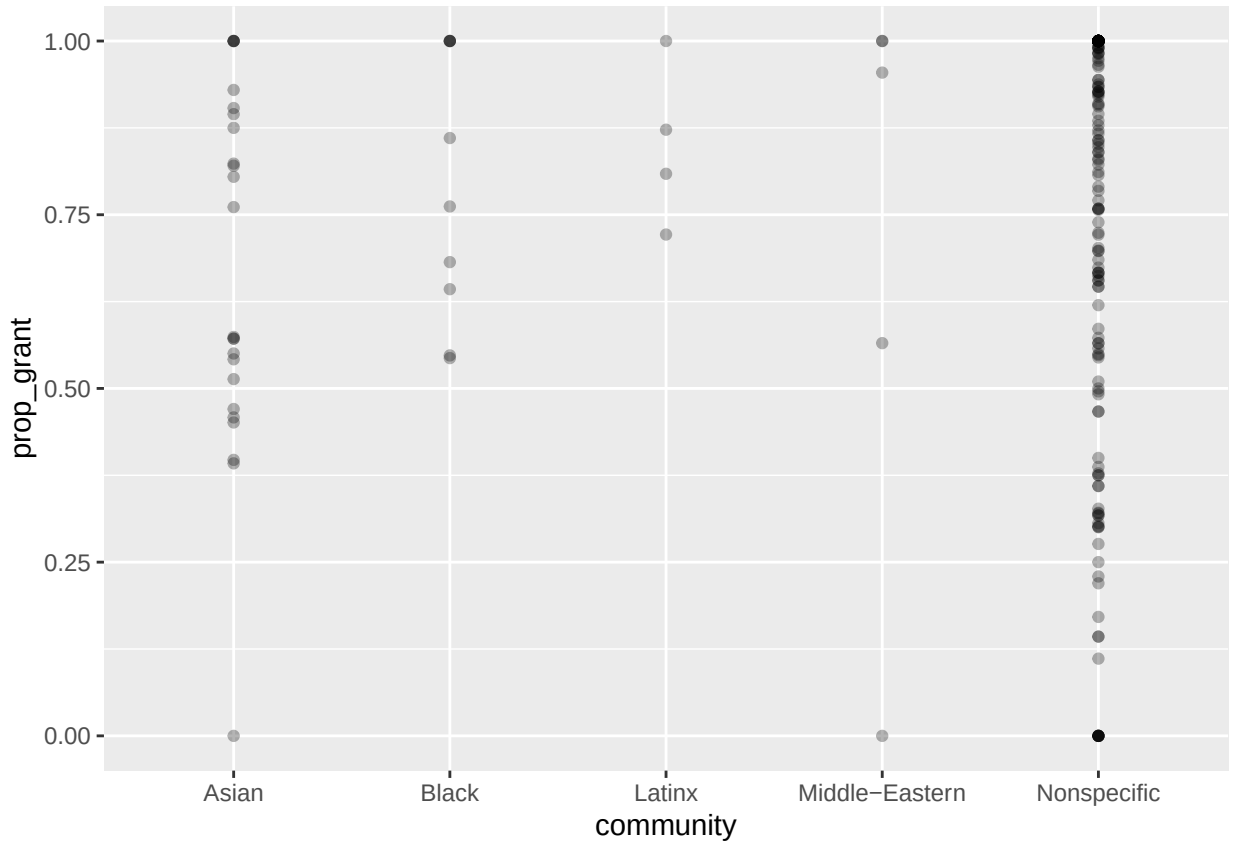

```
prog_cat %>%
  filter(schoolyr == "2018-2019") %>%
  ggplot(aes(x = community, y = prop_grant)) +
  labs(title = "2018-2019") +
  geom_boxplot()
```




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

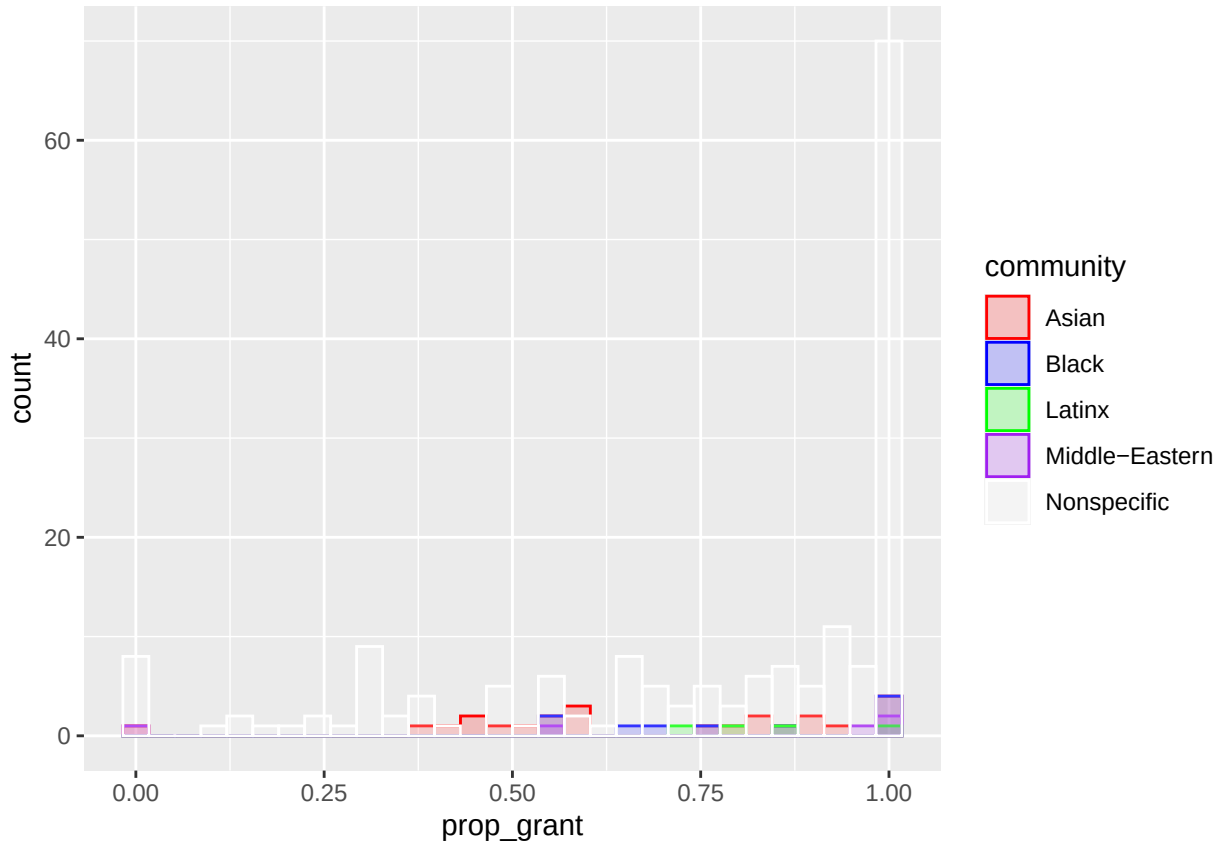



```
ggplot(prog_cat, aes(x = community, y = prop_grant)) +
  geom_point(alpha = 0.3)
```



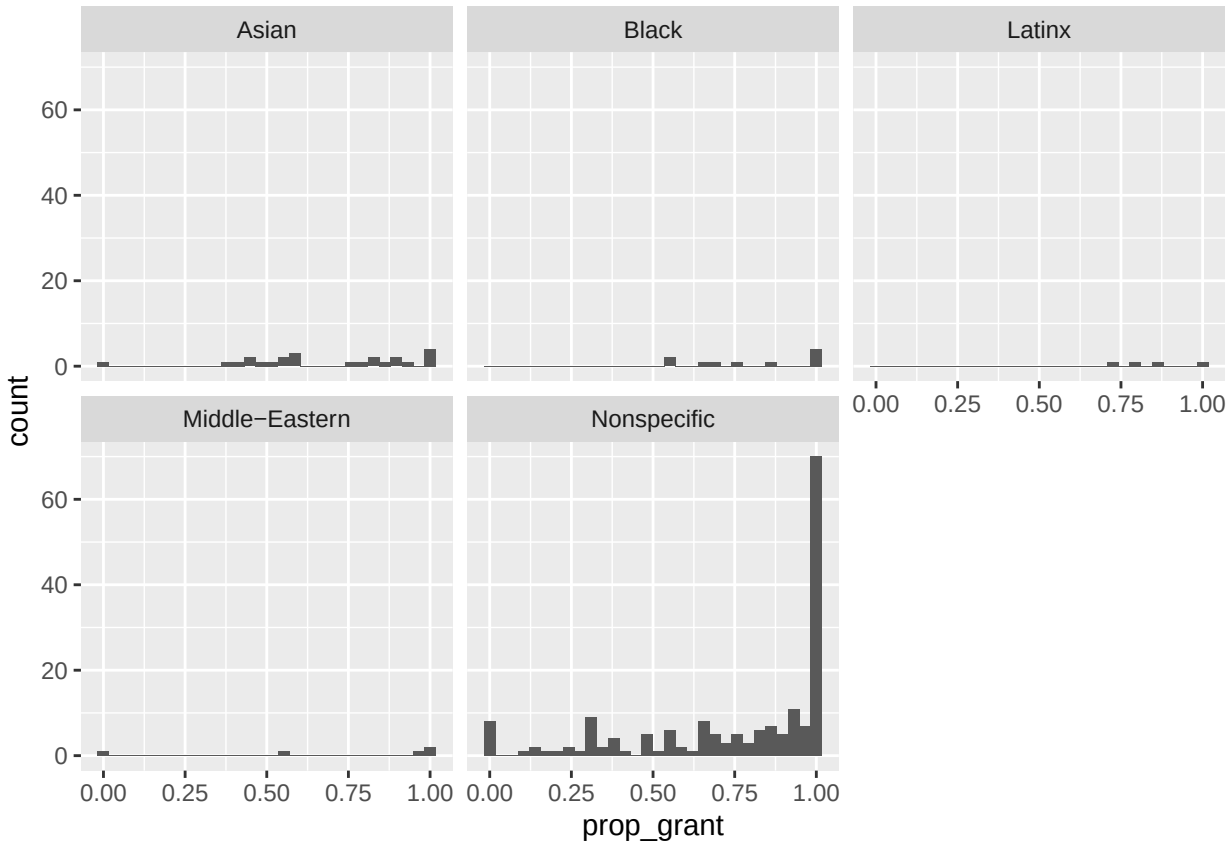

```
ggplot(prog_cat, aes(x = prop_grant)) +
  geom_histogram(aes(color = community, fill = community),
                 position = "identity", alpha = 0.2) +
  scale_color_manual(values = c("red", "blue", "green", "purple", "white")) +
  scale_fill_manual(values = c("red", "blue","green", "purple", "white"))
```

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

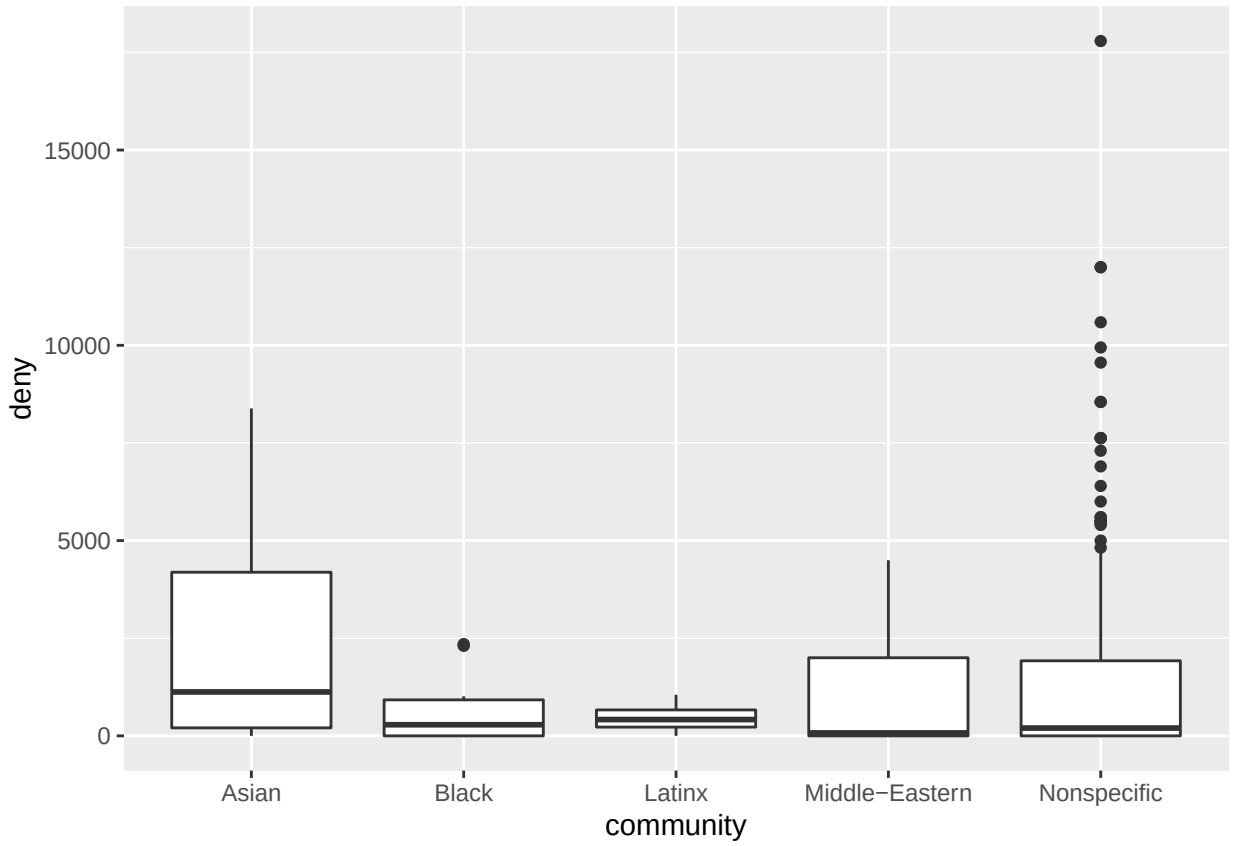



```
ggplot(prog_cat, aes(x = prop_grant)) +
  geom_histogram() +
  facet_wrap(. ~ community)
```

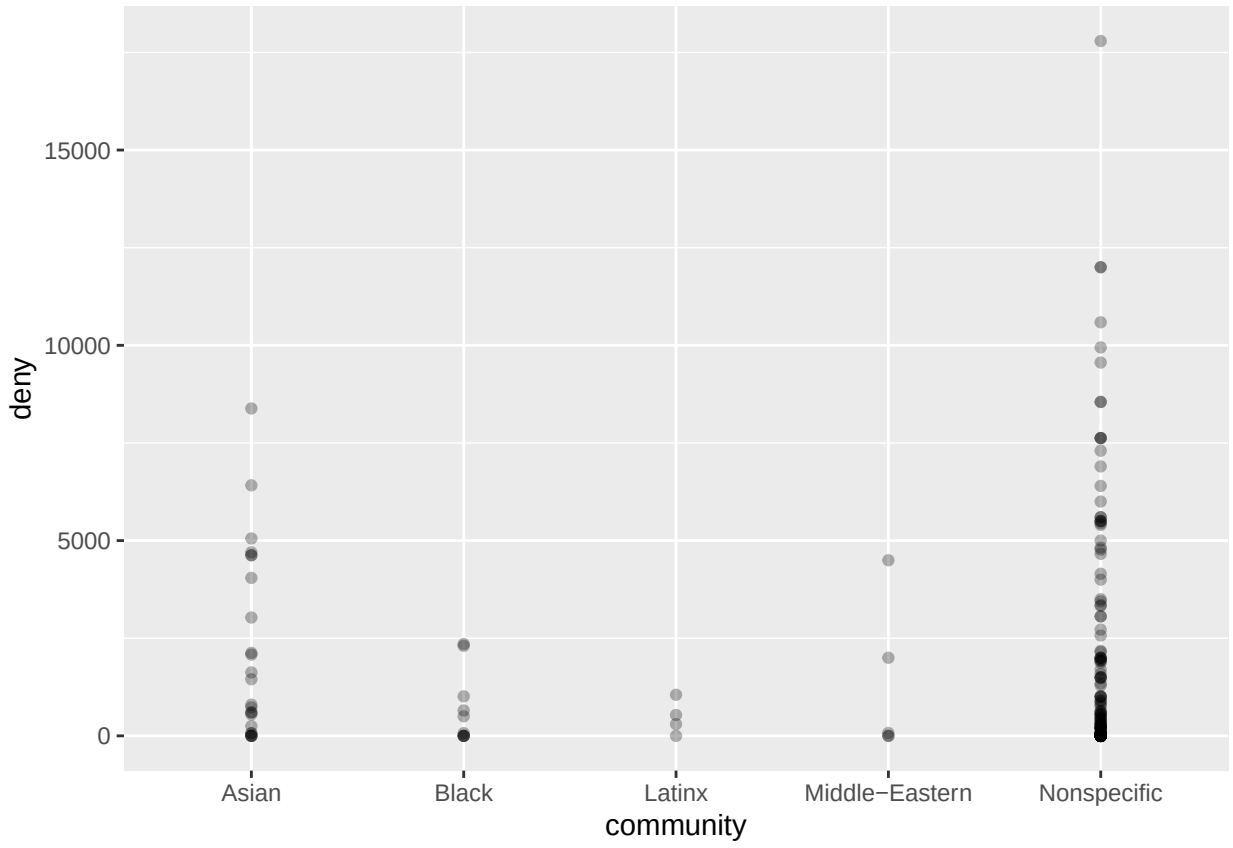
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




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

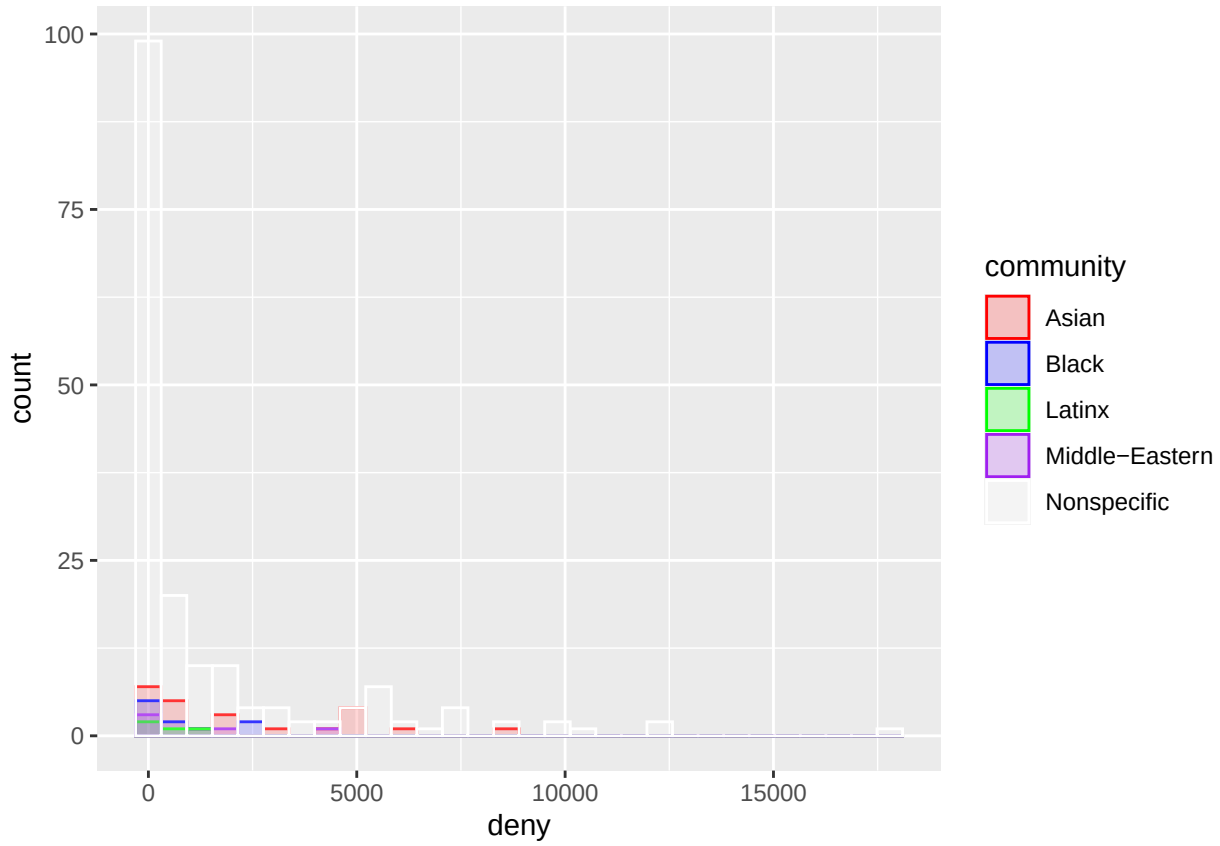



```
ggplot(prog_cat, aes(x = community, y = deny)) +
  geom_point(alpha = 0.3)
```



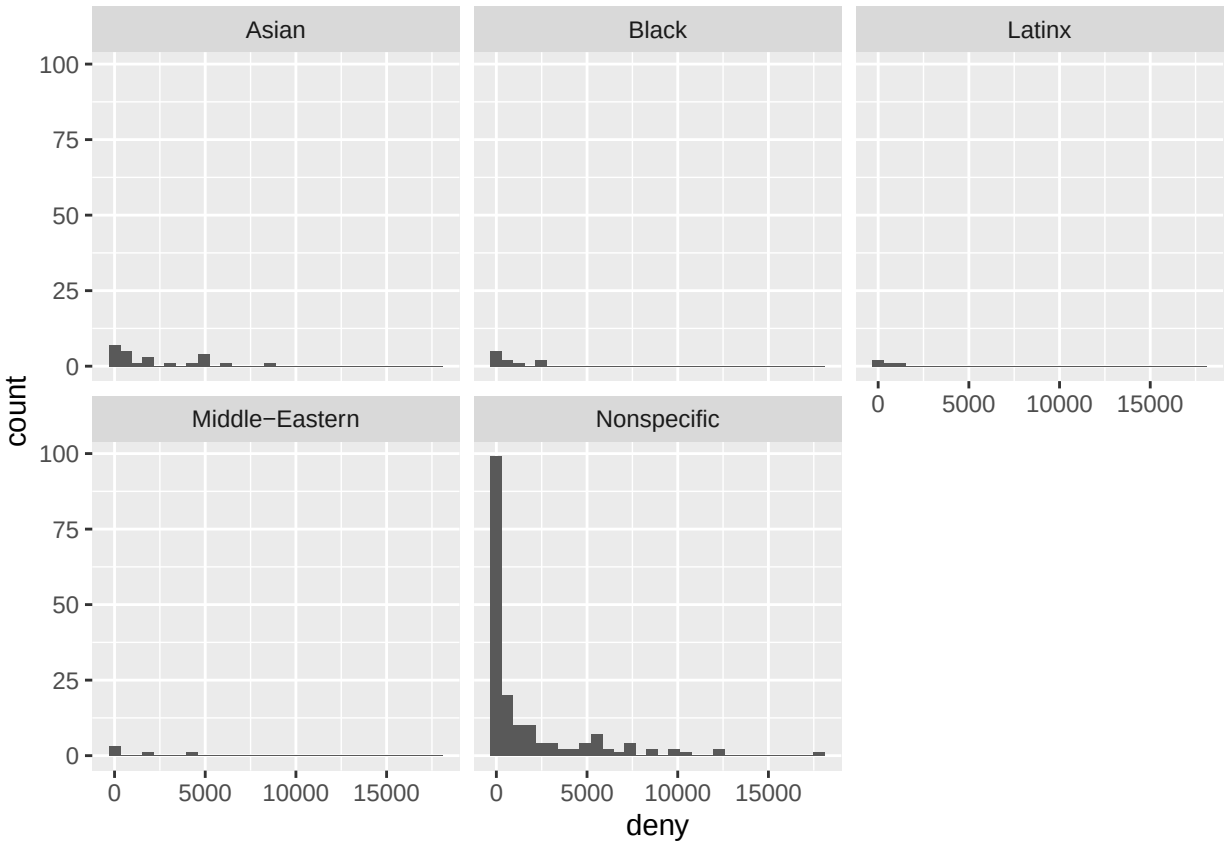

```
ggplot(prog_cat, aes(x = deny)) +
  geom_histogram(aes(color = community, fill = community),
                 position = "identity", alpha = 0.2) +
  scale_color_manual(values = c("red", "blue", "green", "purple", "white")) +
  scale_fill_manual(values = c("red", "blue","green", "purple", "white"))
```

```
## `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                Asian American Alliance  1
## 3  Asian Intervarsity Christian Fellowship 1
## 4                               Brownstone 1
## 5                     CrossFit Blue Devil 1
## 6                        Devils en Pointe 1
## 7                     Duke Amandla Chorus 1
## 8                           Duke Archery 1
## 9                     Duke Chinese Dance  1
## 10               Duke Club Figure Skating 1
```

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

```
##          Group.1         x
## 1         Latinx 0.8506862
## 2          Black 0.8038183
## 3    Nonspecific 0.7607200
## 4 Middle-Eastern 0.7039526
## 5          Asian 0.6793620
```

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

```
##                              Group.1        x
## 1               Blue Devils United  33139.00
## 2        Asian Students Association 24371.00
## 3         International Association 24295.00
## 4      National Panhellenic Council 23179.35
## 5                         Duke Diya 21137.75
## 6    Singapore Students Association 20680.00
## 7               Duke Chinese Theater 19650.10
## 8                           TEDxDuke 19570.00
## 9            Duke Conservation Tech 18295.00
## 10 Delta Sigma Theta Sorority, Inc. 18169.20
```

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

```
##          Group.1        x
## 1    Nonspecific 570650.5
## 2          Asian  85376.1
## 3          Black  37432.1
## 4 Middle-Eastern  14175.0
## 5         Latinx   8740.0
```

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

```
##          Group.1         x
## 1    Nonspecific 281154.53
## 2          Asian  51822.00
## 3          Black   6888.00
## 4 Middle-Eastern   6572.00
## 5         Latinx   1886.99
```

```r
model_bipoc <- lm(prop_grant ~ bipoc, data = prog_cat)
kable(model_bipoc %>% tidy(conf.int=TRUE),format="html",digits=3)
```

term

estimate

std.error

statistic

p.value

conf.low

conf.high

(Intercept)

0.764

0.022

34.204

0.00

0.720

0.808

bipocY

-0.045

0.047

-0.957

0.34

-0.139

0.048

```
kable(tidy(aov(model_bipoc)),format="html",digits=3)
```

term

df

sumsq

meansq

statistic

p.value

bipoc

1

0.078

0.078

0.916

0.34

Residuals

218

18.611

0.085

NA

NA

```
model_comm <- lm(prop_grant ~ community,data=prog_cat)
kable(tidy(aov(model_comm)),format="html",digits=3)
```

term

df

sumsq

meansq

statistic

p.value

community

4

0.216

0.054

0.63

0.642

Residuals

215

18.473

0.086

NA

NA