

Exe. 3

```
library(ggraph)
library(igraph)

library(arrow)
library(tidyverse)
library(gender)
library(wru)
library(lubridate)

library(ggplot2)
library(gridExtra)
library(grid)

library("gender")
```

Data

```
app <- read_parquet('/Users/danystefan/Documents/01 McGill University/01 MMA/01
edges <- read_csv('/Users/danystefan/Documents/01 McGill University/01 MMA/01
```

Add gender part

```
# first name
names <- app %>% distinct(examiner_name_first)
# names and gender
names_gender <- names %>%
  do(results = gender(.$examiner_name_first, method = "ssa")) %>%
  unnest(cols = c(results), keep_empty = TRUE) %>%
  select(
    examiner_name_first = name,
    gender,
    proportion_female
  )
names_gender <- names_gender %>% select(examiner_name_first, gender)
```

```
# join
app <- app %>% left_join(names_gender, by='examiner_name_first')
```

Add race part

```
# last names
sur <- app %>% select(surname = examiner_name_last) %>% distinct()

race <- predict_race(voter.file = sur, surname.only = T) %>% as_tibble()

## [1] "Proceeding with surname-only predictions..."

race <- race %>%
  mutate(max_race_p = pmax(pred.asi, pred.bla, pred.his, pred.oth, pred.whi))
  mutate(race = case_when(
    max_race_p == pred.asi ~ "Asian",
    max_race_p == pred.bla ~ "black",
    max_race_p == pred.his ~ "Hispanic",
    max_race_p == pred.oth ~ "other",
    max_race_p == pred.whi ~ "white",
    TRUE ~ NA_character_
  ))

# cleanup
race <- race %>% select(surname, race)

#join
app <- app %>% left_join(race, by=c("examiner_name_last" = "surname"))
```

Add tenure part

```
# get dates
dates <- app %>% select(examiner_id, filing_date, appl_status_date)
# calculate start and end date
dates <- dates %>% mutate(start_date = ymd(filing_date), end_date = as_date(dr

dates <- dates %>%
  group_by(examiner_id) %>%
  summarise(
    earliest_date = min(start_date, na.rm = TRUE),
```

```

    latest_date = max(end_date, na.rm = TRUE),
    tenure_days = interval(earliest_date, latest_date) %/% days(1)
  ) %>%
  filter(year(latest_date)<2018)

# join
app <- app %>% left_join(dates, by="examiner_id")

```

Set up 2 work groups

```

# get 2 work groups
w1 = app[substr(app$examiner_art_unit, 1, 3) == 163,]
w2 = app[substr(app$examiner_art_unit, 1, 3) == 177,]

```

Summary

```
summary(w1)
```

```

## application_number filing_date examiner_name_last examiner_name_f
## Length:90860      Min.   :2000-01-02 Length:90860      Length:90860
## Class :character  1st Qu.:2003-12-19 Class :character  Class :character
## Mode  :character  Median :2007-12-17 Mode  :character  Mode  :character
##                      Mean    :2008-02-03
##                      3rd Qu.:2011-11-21
##                      Max.    :2017-04-27
##
## examiner_name_middle examiner_id examiner_art_unit uspc_class
## Length:90860      Min.   :59156 Min.   :1631 Length:90860
## Class :character  1st Qu.:67173 1st Qu.:1633 Class :character
## Mode  :character  Median :75340 Median :1635 Mode  :character
##                      Mean    :78698 Mean    :1635
##                      3rd Qu.:93760 3rd Qu.:1637
##                      Max.    :99764 Max.    :1639
##                      NA's    :861
## uspc_subclass patent_number patent_issue_date
## Length:90860      Length:90860 Min.   :2000-12-12
## Class :character  Class :character 1st Qu.:2007-08-28
## Mode  :character  Mode  :character Median :2011-05-31
##                      Mean    :2010-10-24
##                      3rd Qu.:2013-12-17

```

```

##                               Max.      :2017-06-20
##                               NA's       :53499
##   abandon_date      disposal_type      appl_status_code appl_status_date
## Min.      :1990-07-01  Length:90860      Min.      : 1.0      Length:90860
## 1st Qu.:2006-11-13    Class :character  1st Qu.:150.0      Class :character
## Median :2009-10-27    Mode  :character  Median :161.0      Mode  :character
## Mean    :2009-12-02                                Mean    :148.9
## 3rd Qu.:2013-01-23                                3rd Qu.:161.0
## Max.    :2017-05-31                                Max.    :854.0
## NA's    :49524                                    NA's    :134
##           tc           gender           race           earliest_date
## Min.      :1600      Length:90860      Length:90860      Min.      :2000-01-02
## 1st Qu.:1600      Class :character  Class :character  1st Qu.:2000-01-10
## Median :1600      Mode  :character  Mode  :character  Median :2000-02-04
## Mean    :1600                                Mean    :2000-10-02
## 3rd Qu.:1600                                3rd Qu.:2000-11-20
## Max.    :1600                                Max.    :2010-09-10
##                               NA's       :2820
##   latest_date      tenure_days
## Min.      :2000-12-14  Min.      : 251
## 1st Qu.:2017-05-19    1st Qu.:6016
## Median :2017-05-20    Median :6296
## Mean    :2017-04-27    Mean    :6051
## 3rd Qu.:2017-05-22    3rd Qu.:6339
## Max.    :2017-05-23    Max.    :6349
## NA's    :2820          NA's    :2820

```

summary(w2)

```

## application_number  filing_date      examiner_name_last examiner_name_f
## Length:83266        Min.      :2000-01-03  Length:83266      Length:83266
## Class :character    1st Qu.:2005-04-01    Class :character   Class :character
## Mode  :character    Median :2010-05-13    Mode  :character   Mode  :character
##                               Mean    :2009-06-23
##                               3rd Qu.:2013-04-18
##                               Max.    :2017-04-25
##
## examiner_name_middle examiner_id    examiner_art_unit  uspc_class
## Length:83266        Min.      :59201  Min.      :1771    Length:83266
## Class :character    1st Qu.:65934    1st Qu.:1772      Class :character
## Mode  :character    Median :72332    Median :1774      Mode  :character
##                               Mean    :77486    Mean    :1774
##                               3rd Qu.:93496    3rd Qu.:1776
##                               Max.    :99945    Max.    :1779
##                               NA's    :262
## uspc_subclass      patent_number      patent_issue_date
## Length:83266      Length:83266      Min.      :2000-09-12

```

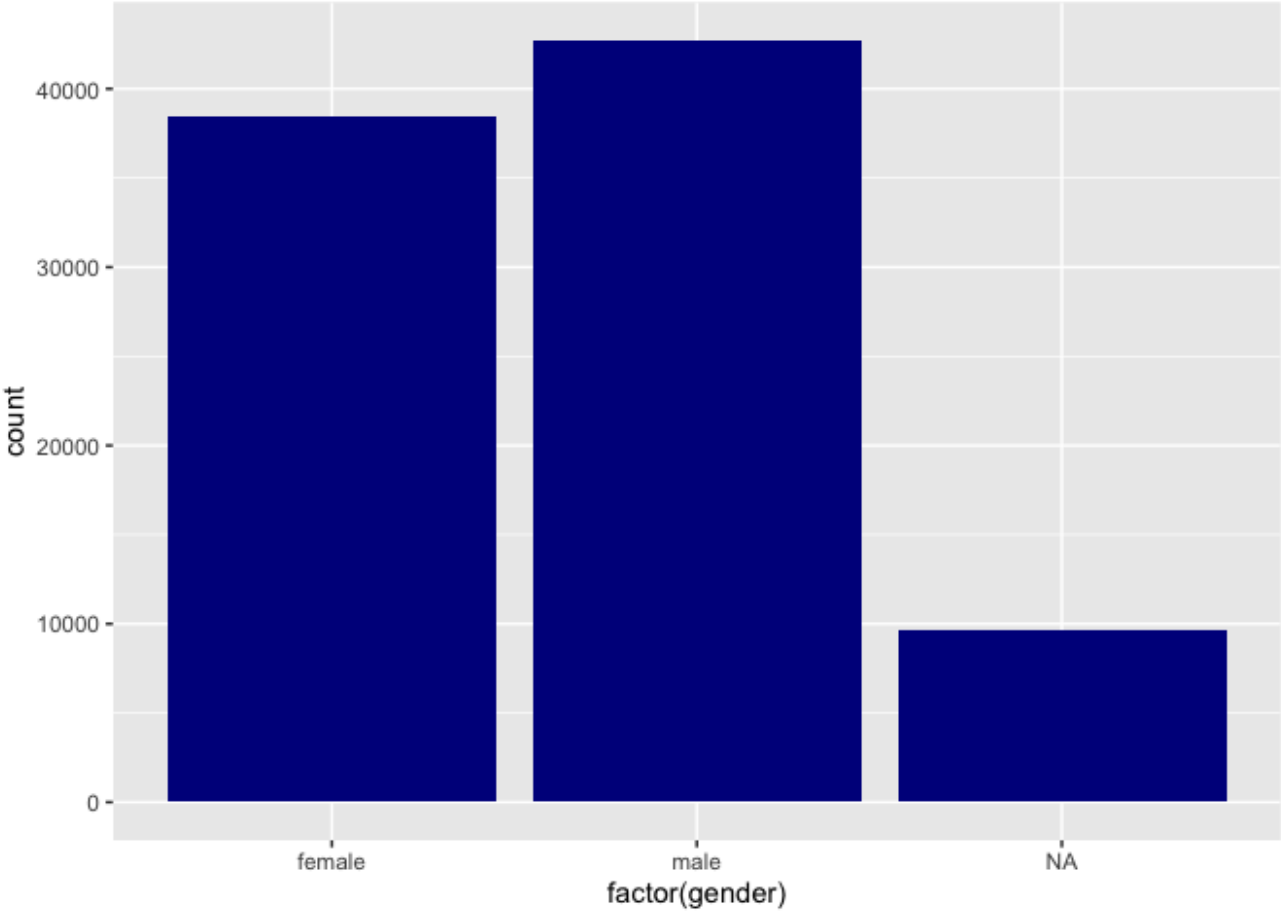
```
## Class :character   Class :character   1st Qu.:2006-12-05
## Mode :character   Mode :character   Median :2012-10-16
##                                     Mean  :2011-05-07
##                                     3rd Qu.:2014-10-28
##                                     Max.   :2017-06-20
##                                     NA's    :37600
##   abandon_date      disposal_type      appl_status_code appl_status_date
## Min.   :2000-06-14   Length:83266      Min.    : 1.0      Length:83266
## 1st Qu.:2006-11-27   Class :character   1st Qu.:150.0      Class :character
## Median :2012-03-22   Mode  :character   Median :150.0      Mode  :character
## Mean   :2011-03-05                      Mean   :146.2
## 3rd Qu.:2014-06-26                      3rd Qu.:161.0
## Max.   :2017-06-01                      Max.   :454.0
## NA's   :58964                      NA's   :141
##      tc              gender              race              earliest_date
## Min.   :1700      Length:83266      Length:83266      Min.   :2000-01-03
## 1st Qu.:1700      Class :character   Class :character   1st Qu.:2000-01-06
## Median :1700      Mode  :character   Mode  :character   Median :2000-02-07
## Mean   :1700                                      Mean   :2002-07-03
## 3rd Qu.:1700                                      3rd Qu.:2004-07-20
## Max.   :1700                                      Max.   :2014-03-27
##                                     NA's    :262
##   latest_date      tenure_days
## Min.   :2000-10-23   Min.    : 216
## 1st Qu.:2017-05-19   1st Qu.:4686
## Median :2017-05-22   Median :6282
## Mean   :2017-05-07   Mean   :5422
## 3rd Qu.:2017-05-23   3rd Qu.:6342
## Max.   :2017-05-23   Max.   :6350
## NA's   :262          NA's   :262
```

Distributions for gender in each group

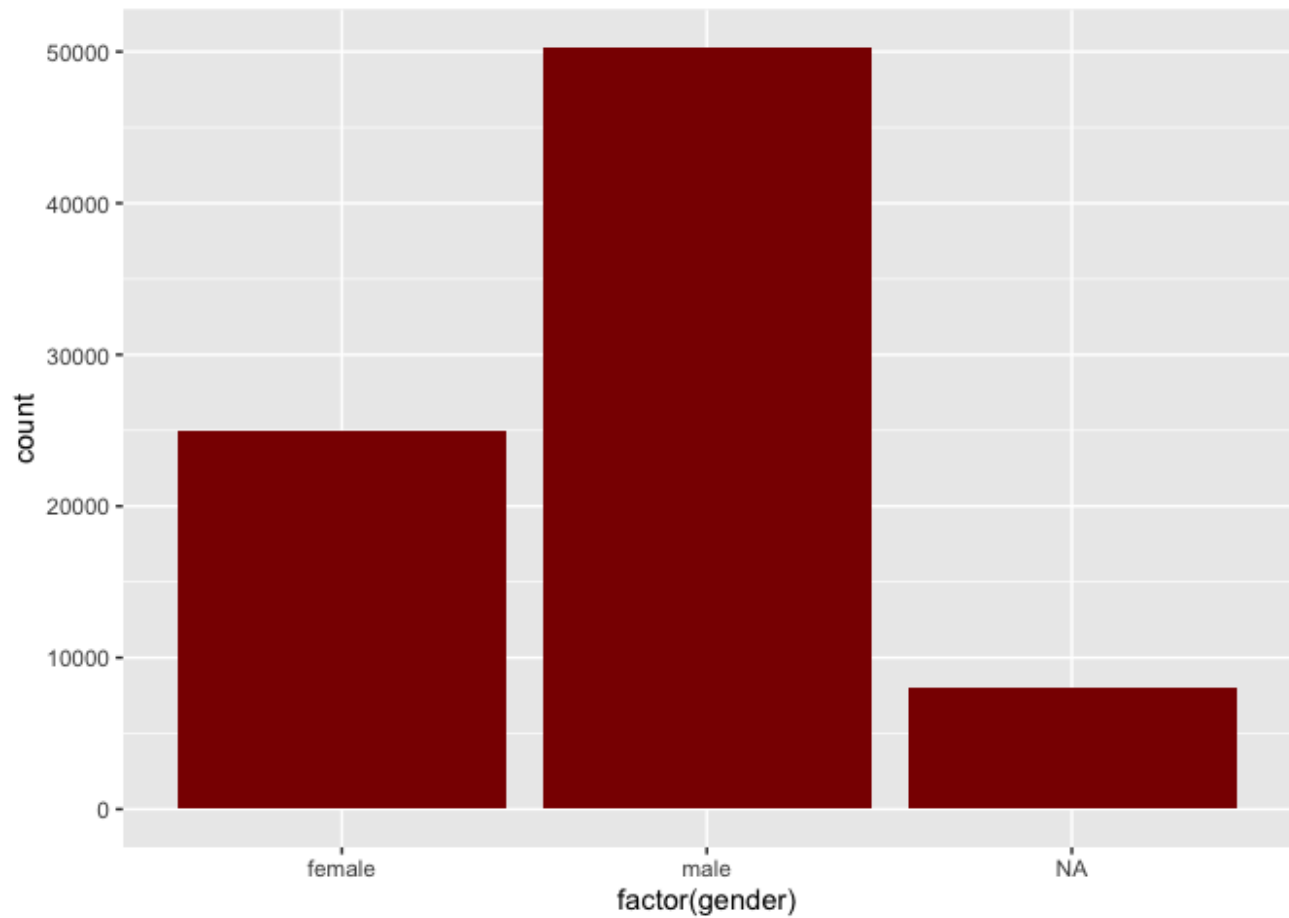
```
a <- ggplot(data=w1, aes(x=factor(gender))) + geom_bar(fill="darkblue")
```

```
b <- ggplot(data=w2, aes(x=factor(gender))) + geom_bar(fill="darkred")
```

a



b



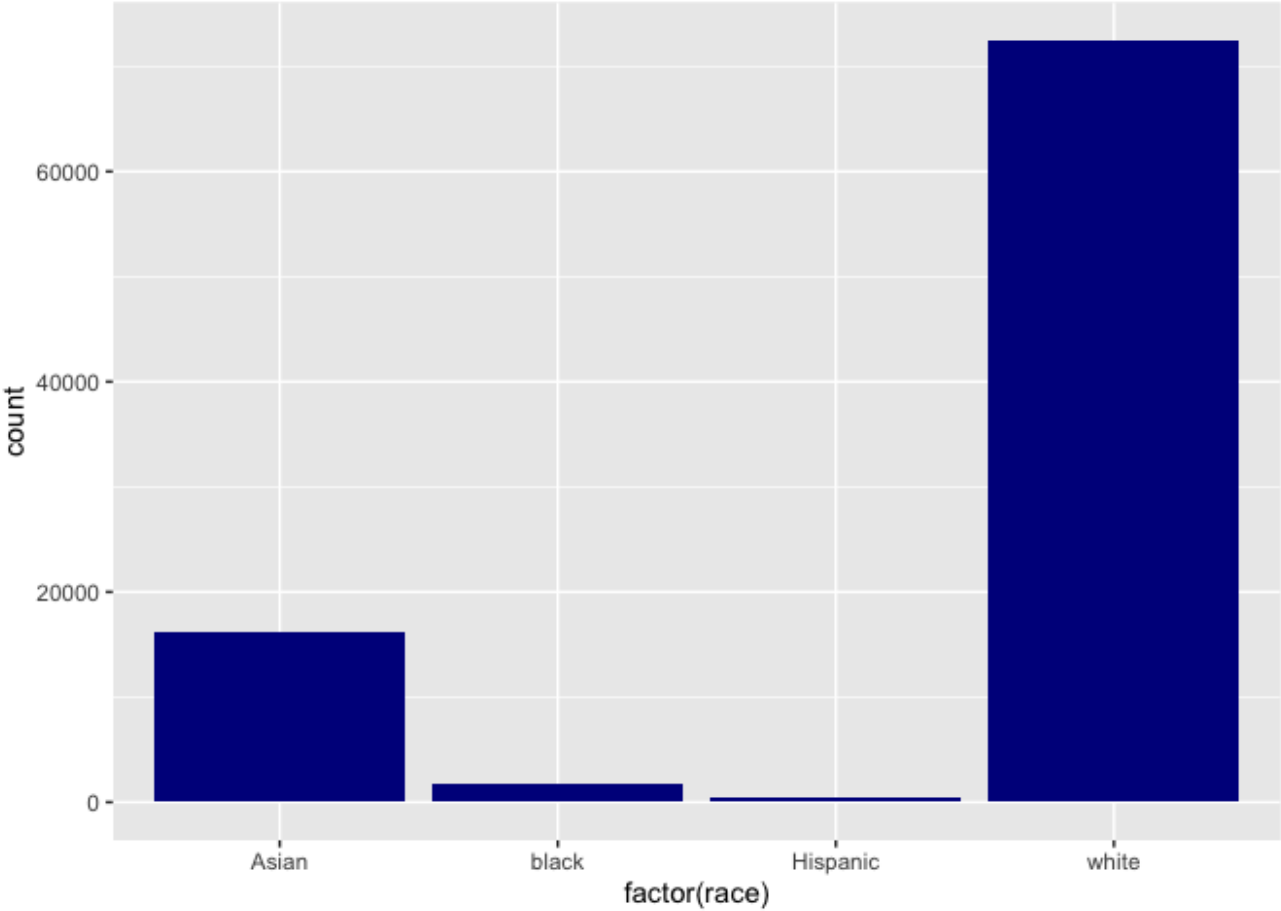
For group w1, the ratio is even. However, for w2, there are definitely more male than females.

Distributions for race in each group

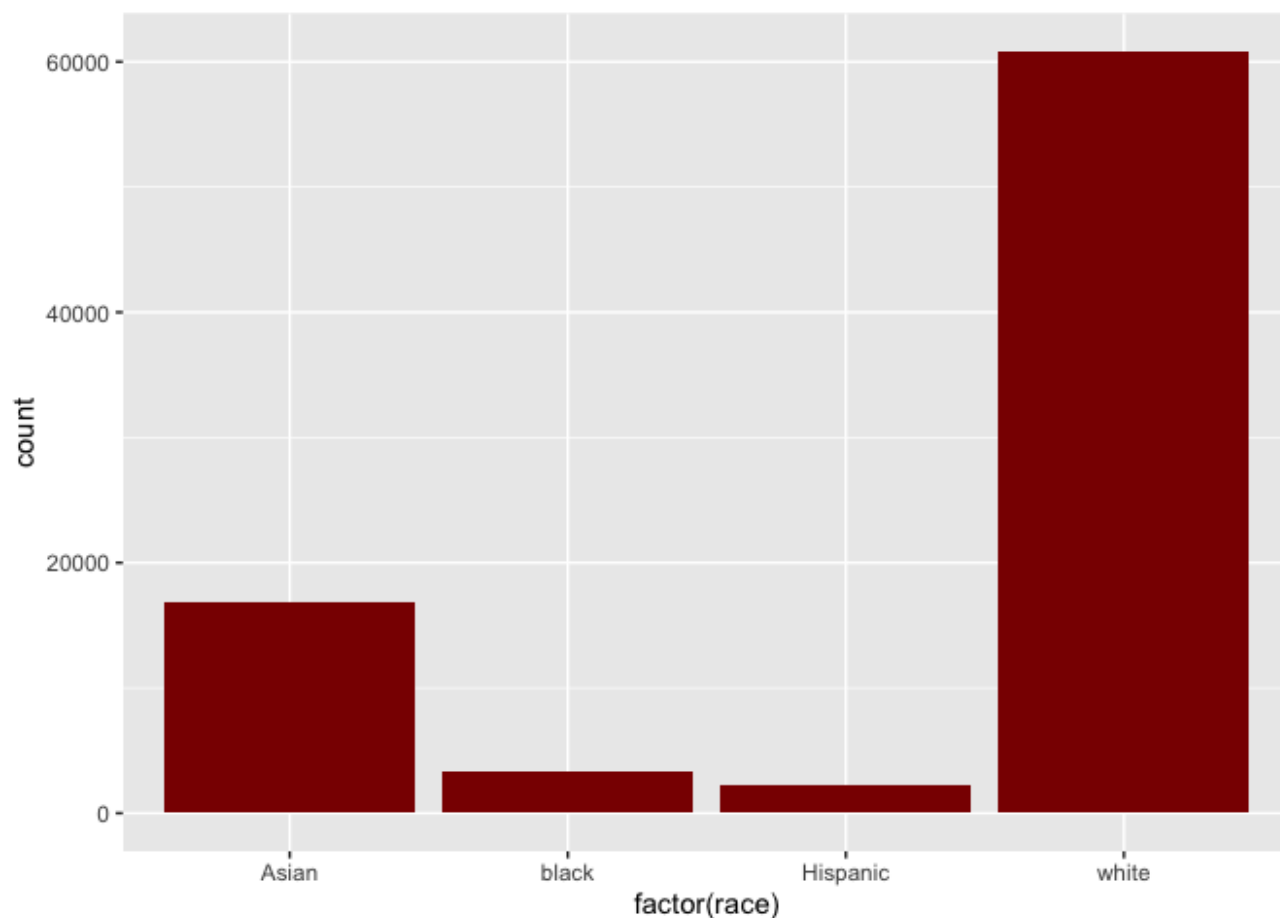
```
a <- ggplot(data=w1, aes(x=factor(race))) + geom_bar(fill="darkblue")
```

```
b <- ggplot(data=w2, aes(x=factor(race))) + geom_bar(fill="darkred")
```

a

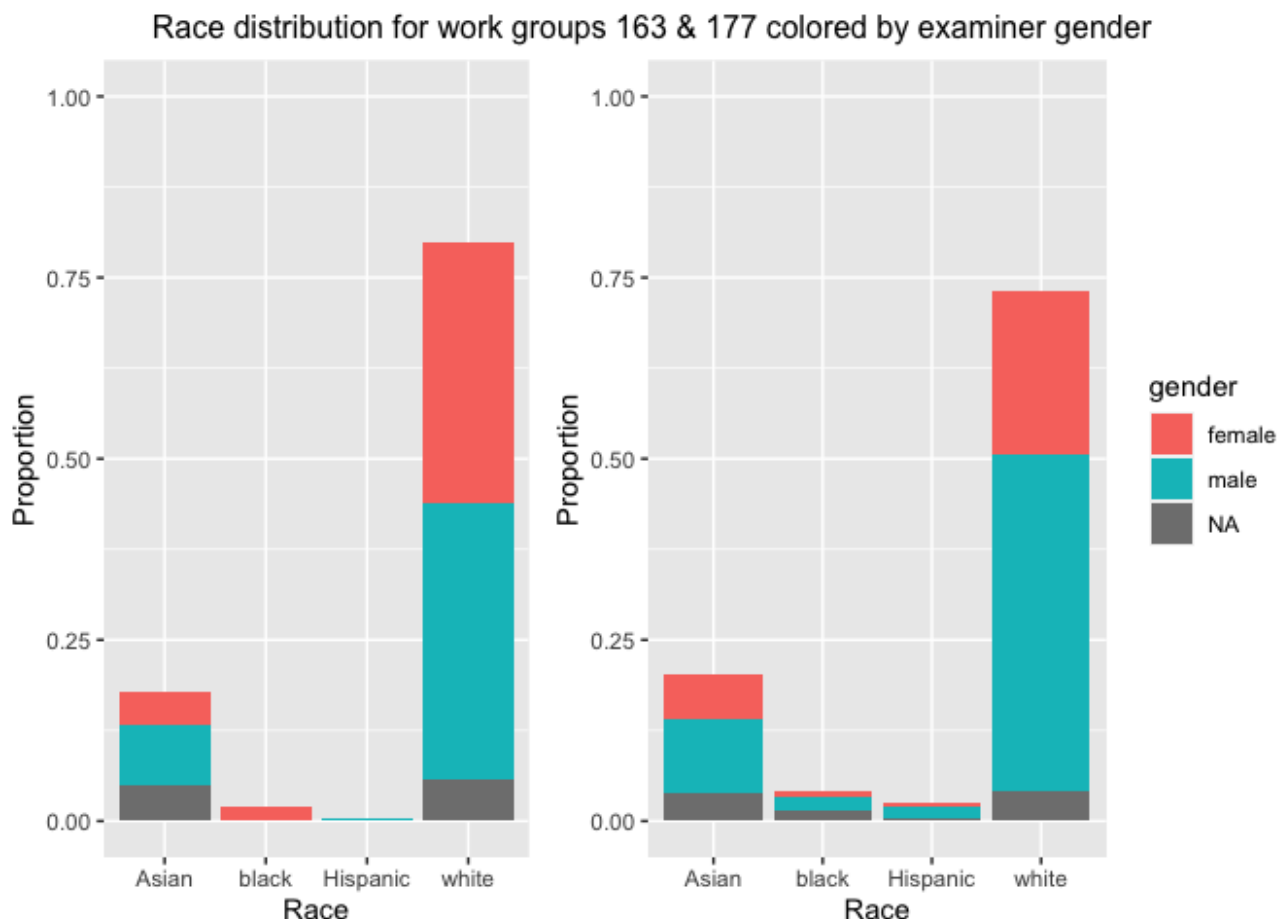


b



In both groups, asians and whites are predominantly the majority compared to other races.

```
a <- ggplot(data=w1, aes(x=race)) +
  geom_bar(aes(y = (..count..)/sum(..count..), fill=gender ), show.legend=FALSE) +
  ylab("Proportion")+
  ylim(0,1)+
  xlab("Race")
b <- ggplot(data=w2, aes(x=race)) +
  geom_bar(aes(y = (..count..)/sum(..count..), fill=gender )) +
  ylab("Proportion")+
  ylim(0,1)+
  xlab("Race")
grid.arrange(a,b,ncol=2, widths=c(1,1.4), top=textGrob("Race distribution for
```



Network

```
# limit to 2 groups
aus = distinct(subset(app, select=c(examiner_art_unit, examiner_id)))
aus$wg = substr(aus$examiner_art_unit, 1,3)
aus = aus[aus$wg == 163 | aus$wg == 177,]

tm = merge(x=edges, y=aus, by.x="ego_examiner_id", by.y="examiner_id", all.x=T)
tm = tm %>% rename(ego_art_unit=examiner_art_unit, ego_wg=wg)
tm = drop_na(tm)

tm = merge(x=tm, y=aus, by.x="alter_examiner_id", by.y="examiner_id", all.x=TF)
tm = tm %>% rename(alter_art_unit=examiner_art_unit, alter_wg=wg)
tm = drop_na(tm)

egoNodes = subset(tm, select=c(ego_examiner_id,ego_art_unit, ego_wg)) %>% rena
alterNodes = subset(tm, select=c(alter_examiner_id,alter_art_unit, alter_wg))%
nodes = rbind(egoNodes, alterNodes)
nodes = distinct(nodes)
nodes = nodes %>% group_by(examiner_id) %>% summarise(examiner_id=first(examir
```

```
net = graph_from_data_frame(d=tm, vertices=nodes, directed=TRUE)
net
```

```
## IGRAPH c68c38e DN-- 173 1142 --
## + attr: name (v/c), art_unit (v/n), wg (v/c), application_number (e/c),
## | advice_date (e/n), ego_art_unit (e/n), ego_wg (e/c), alter_art_unit
## | (e/n), alter_wg (e/c)
## + edges from c68c38e (vertex names):
## [1] 59456->99518 59589->69665 60706->78051 60706->78051 60706->78051
## [6] 60706->78051 60706->78051 60706->78051 60706->78051 60706->65547
## [11] 60706->78051 60706->78051 60706->78051 61519->72253 61519->61519
## [16] 61519->61519 61519->72253 61519->61519 61519->72253 62024->71388
## [21] 62312->61519 62312->98614 62312->98614 62312->86861 62312->61519
## [26] 62312->66971 62312->98614 62312->98614 62312->66971 62312->98614
## + ... omitted several edges
```

```
# centrality
Degree <- degree(net, v=V(net))
Betweenness <- betweenness(net)
Eigenvector <- evcent(net)$vector
```

```
V(net)$size = Degree
V(net)$eig = round(Eigenvector,2)
V(net)$bet = round(Betweenness,2)
```

```
V(net)$color = nodes$art_unit
```

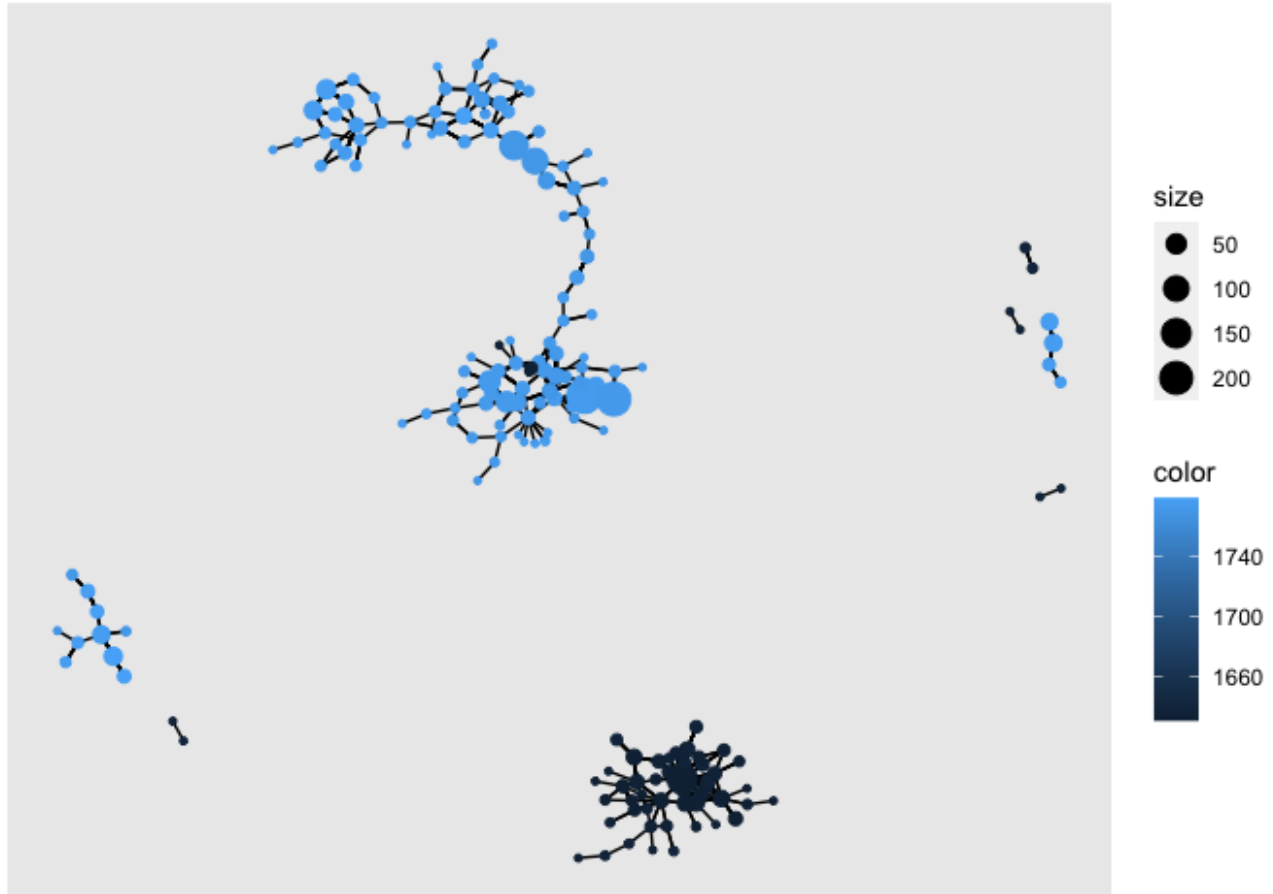
```
centralities <- cbind(Degree, Eigenvector, Betweenness)
centralities = round(centralities,2)
centralities = data.frame(centralities)
```

```
V(net)
```

```
## + 173/173 vertices, named, from c68c38e:
## [1] 59156 59407 59456 59539 59589 60706 60837 60991 61299 61416 61519 620
## [13] 62312 62464 62495 62621 62767 62778 62815 62862 62990 63027 63030 631
## [25] 63219 63234 63244 63713 63977 64073 64823 64992 65121 65131 65179 652
## [37] 65547 65601 65646 65757 65934 66264 66387 66442 66762 66769 66824 668
## [49] 66910 66949 66971 67173 67300 67376 67515 68017 68141 68153 68598 690
## [61] 69098 69464 69539 69665 69780 69909 70032 70423 70458 70799 70993 710
## [73] 71087 71120 71123 71385 71388 71720 71853 71931 72097 72122 72165 722
## [85] 72332 72514 72576 72591 72848 72995 73239 73880 74725 74727 75336 763
## [97] 76516 76927 77184 77772 77958 78019 78051 78406 79289 79495 79538 801
## [109] 80730 80908 81337 82433 83475 84313 85060 85736 86422 86861 86928 882
## + ... omitted several vertices
```

Graph

```
ggraph(net, layout="kk") +
  geom_edge_link()+
  geom_node_point(aes(size=size, color=color), show.legend=T)
```



It looks like there are examiners leaving their group and seeking advice elsewhere (those are the smaller groups formed around the larger two groups).

It also looks like in the blue network, there are some nodes in black corresponding to examiners starting with identifier 16. Let's look at id 64507 and get a summary.

```
exam <- app %>% filter(examiner_id==64507)
summary(exam)
```

```
## application_number filing_date examiner_name_last examiner_name_f
## Length:442 Min. :2000-01-14 Length:442 Length:442
```

```

##      Class :character      1st Qu.:2002-01-28      Class :character      Class :character
##      Mode  :character      Median :2003-11-09      Mode  :character      Mode  :character
##                                     Mean  :2004-05-17
##                                     3rd Qu.:2006-09-06
##                                     Max.   :2012-05-08
##
##      examiner_name_middle  examiner_id      examiner_art_unit  uspc_class
##      Length:442           Min.    :64507      Min.    :1644      Length:442
##      Class :character      1st Qu.:64507      1st Qu.:1644      Class :character
##      Mode  :character      Median :64507      Median :1644      Mode  :character
##                                     Mean  :64507      Mean  :1644
##                                     3rd Qu.:64507      3rd Qu.:1644
##                                     Max.   :64507      Max.   :1644
##
##      uspc_subclass         patent_number      patent_issue_date
##      Length:442           Length:442      Min.    :2001-12-25
##      Class :character      Class :character      1st Qu.:2005-09-16
##      Mode  :character      Mode  :character      Median :2008-03-11
##                                     Mean  :2008-01-31
##                                     3rd Qu.:2010-08-17
##                                     Max.   :2013-03-05
##                                     NA's   :167
##
##      abandon_date         disposal_type      appl_status_code  appl_status_date
##      Min.    :2001-07-24   Length:442      Min.    :150.0     Length:442
##      1st Qu.:2005-11-28   Class :character      1st Qu.:150.0     Class :character
##      Median :2007-05-08   Mode  :character      Median :161.0     Mode  :character
##      Mean    :2007-09-13                                     Mean  :169.9
##      3rd Qu.:2010-01-02                                     3rd Qu.:161.0
##      Max.    :2012-11-01                                     Max.   :250.0
##      NA's     :275
##
##      tc                   gender              race              earliest_date
##      Min.    :1600        Length:442      Length:442      Min.    :2000-01-14
##      1st Qu.:1600        Class :character      Class :character      1st Qu.:2000-01-14
##      Median :1600        Mode  :character      Mode  :character      Median :2000-01-14
##      Mean    :1600                                     Mean  :2000-01-14
##      3rd Qu.:1600                                     3rd Qu.:2000-01-14
##      Max.    :1600                                     Max.   :2000-01-14
##
##      latest_date          tenure_days
##      Min.    :2017-05-12   Min.    :6328
##      1st Qu.:2017-05-12   1st Qu.:6328
##      Median :2017-05-12   Median :6328
##      Mean    :2017-05-12   Mean    :6328
##      3rd Qu.:2017-05-12   3rd Qu.:6328
##      Max.    :2017-05-12   Max.    :6328
##

```

```
## # A tibble: 442 × 21
##   application_number filing_date examiner_name_last examiner_name_first
##   <chr>             <date>      <chr>             <chr>
## 1 090000004         2001-05-02 SAUNDERS           DAVID
## 2 09463470         2000-01-21 SAUNDERS           DAVID
## 3 09483588         2000-01-14 SAUNDERS           DAVID
## 4 09486311         2000-02-25 SAUNDERS           DAVID
## 5 09490520         2000-01-25 SAUNDERS           DAVID
## 6 09492191         2000-01-27 SAUNDERS           DAVID
## 7 09500135         2000-02-08 SAUNDERS           DAVID
## 8 09509734         2000-06-14 SAUNDERS           DAVID
## 9 09513145         2000-02-25 SAUNDERS           DAVID
## 10 09521700        2000-03-09 SAUNDERS           DAVID
## # ... with 432 more rows, and 17 more variables: examiner_name_middle <chr>,
## #   examiner_id <dbl>, examiner_art_unit <dbl>, uspc_class <chr>,
## #   uspc_subclass <chr>, patent_number <chr>, patent_issue_date <date>,
## #   abandon_date <date>, disposal_type <chr>, appl_status_code <dbl>,
## #   appl_status_date <chr>, tc <dbl>, gender <chr>, race <chr>,
## #   earliest_date <date>, latest_date <date>, tenure_days <dbl>
```
