Exercice3

GitHub Documents

This is an R Markdown format used for publishing markdown documents to GitHub. When you click the **Knit** button all R code chunks are run and a markdown file (.md) suitable for publishing to GitHub is generated.

Including Code

You can include R code in the document as follows:

```
#install.packages('wru')
library(wru)
```

```
## Warning: le package 'wru' a été compilé avec la version R 4.1.3
```

```
library(arrow)
```

```
## Warning: le package 'arrow' a été compilé avec la version R 4.1.3
```

```
##
## Attachement du package : 'arrow'
```

```
## L'objet suivant est masqué depuis 'package:utils':
##
##     timestamp
```

```
library(lubridate)
```

```
##
## Attachement du package : 'lubridate'
```

```
## L'objet suivant est masqué depuis 'package:arrow':
##
##     duration
```

```
## Les objets suivants sont masqués depuis 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(tidyverse)
```

```
## Warning: le package 'tidyverse' a été compilé avec la version R 4.1.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
```

```
## Warning: le package 'tidyr' a été compilé avec la version R 4.1.3
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x lubridate::as.difftime() masks base::as.difftime()
```

```
## x lubridate::date()        masks base::date()
## x lubridate::duration()    masks arrow::duration()
## x dplyr::filter()          masks stats::filter()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
```

```
#install.packages('gender')
library(gender)
```

```
## Warning: le package 'gender' a été compilé avec la version R 4.1.3
```

```
library(gridExtra)
```

```
##
## Attachement du package : 'gridExtra'
```

```
## L'objet suivant est masqué depuis 'package:dplyr':
##
##     combine
```

```
library(ggplot2)
library(grid)
library(igraph)
```

```
## Warning: le package 'igraph' a été compilé avec la version R 4.1.3
```

```
##
## Attachement du package : 'igraph'
```

```
## Les objets suivants sont masqués depuis 'package:dplyr':
##
##     as_data_frame, groups, union
```

```
## Les objets suivants sont masqués depuis 'package:purrr':
##
##     compose, simplify
```

```
## L'objet suivant est masqué depuis 'package:tidyr':
##
##     crossing
```

```
## L'objet suivant est masqué depuis 'package:tibble':
##
##     as_data_frame
```

```
## Les objets suivants sont masqués depuis 'package:lubridate':
##
##     %--%, union
```

```
## Les objets suivants sont masqués depuis 'package:stats':
##
##     decompose, spectrum
```

```
## L'objet suivant est masqué depuis 'package:base':
##
##     union
```

```
library(ggraph)
```

```
## Warning: le package 'ggraph' a été compilé avec la version R 4.1.3
```

Including Plots

You can also embed plots, for example:

```
library(readr)

#install.packages("arrow")
library("arrow")

edges <- read_csv("C:/Users/Mehdi/Desktop/2022-ona-assignements/edges_sample.csv")
```

```
## Rows: 32906 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): application_number
## dbl  (2): ego_examiner_id, alter_examiner_id
## date (1): advice_date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
applications <- read_parquet("C:/Users/Mehdi/Desktop/2022-ona-assignements/app_data_sample.parquet")
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
# get list of unique first names
name_of_examiner = applications %>% distinct(examiner_name_first)
name_of_examiner
```

```
## # A tibble: 2,595 x 1
##    examiner_name_first
##    <chr>
##  1 JACQUELINE
##  2 BEKIR
##  3 CYNTHIA
##  4 MARY
##  5 MICHAEL
##  6 LINDA
##  7 KARA
##  8 VANESSA
##  9 TERESA
## 10 SUN
## # ... with 2,585 more rows
```

Get Gender

```
library(dplyr)
gender = name_of_examiner %>% do(outcome = gender(.$examiner_name_first, method = "ssa")) %>% unnest(co
  select(examiner_name_first = name, gender,proportion_female)

gender
```

```
## # A tibble: 1,822 x 3
```

```
##    examiner_name_first gender proportion_female
##    <chr>               <chr>              <dbl>
##  1 AARON               male              0.0082
##  2 ABDEL               male              0
##  3 ABDOU               male              0
##  4 ABDUL               male              0
##  5 ABDULHAKIM          male              0
##  6 ABDULLAH            male              0
##  7 ABDULLAHI           male              0
##  8 ABIGAIL             female            0.998
##  9 ABIMBOLA            female            0.944
## 10 ABRAHAM             male              0.0031
## # ... with 1,812 more rows
```

Joining and Cleaning

```
# keep only the name and the gender columns in the table
gender = gender %>% select(examiner_name_first, gender)

# Adding the gender to the previous data frame
applications = applications %>% left_join(gender, by = "examiner_name_first")

applications
```

```
## # A tibble: 2,018,477 x 17
##    application_number filing_date examiner_name_last examiner_name_first
##    <chr>              <date>      <chr>              <chr>
##  1 08284457           2000-01-26  HOWARD             JACQUELINE
##  2 08413193           2000-10-11  YILDIRIM           BEKIR
##  3 08531853           2000-05-17  HAMILTON           CYNTHIA
##  4 08637752           2001-07-20  MOSHER             MARY
##  5 08682726           2000-04-10  BARR               MICHAEL
##  6 08687412           2000-04-28  GRAY               LINDA
##  7 08716371           2004-01-26  MCMILLIAN          KARA
##  8 08765941           2000-06-23  FORD               VANESSA
##  9 08776818           2000-02-04  STRZELECKA         TERESA
## 10 08809677           2002-02-20  KIM                SUN
## # ... with 2,018,467 more rows, and 13 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

```
surname = applications %>% distinct(surname = examiner_name_last)

surname
```

```
## # A tibble: 3,806 x 1
##    surname
##    <chr>
##  1 HOWARD
```

```
##  2 YILDIRIM
##  3 HAMILTON
##  4 MOSHER
##  5 BARR
##  6 GRAY
##  7 MCMILLIAN
##  8 FORD
##  9 STRZELECKA
## 10 KIM
## # ... with 3,796 more rows
```

```
race = predict_race(voter.file=surname, surname.only=T) %>% as_tibble()
```

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

```
## Warning in merge_surnames(voter.file): Probabilities were imputed for 698
## surnames that could not be matched to Census list.
```

```
# Get Race probability based on surname

race = race %>%
  mutate(max_proba_race = pmax( pred.his,pred.oth,pred.asi, pred.bla, pred.whi))

race = race %>%
  mutate(race = case_when(
    max_proba_race == pred.bla ~ "Black",
    max_proba_race == pred.whi ~ "white",
    max_proba_race == pred.asi ~ "Asian",
    max_proba_race == pred.his ~ "Hispanic",
    max_proba_race == pred.oth ~ "Other",
    TRUE ~ NA_character_
  ))

race
```

```
## # A tibble: 3,806 x 8
##    surname    pred.whi pred.bla pred.his pred.asi pred.oth max_proba_race race
##    <chr>         <dbl>    <dbl>    <dbl>    <dbl>    <dbl>          <dbl> <chr>
##  1 HOWARD       0.643   0.295    0.0237   0.005    0.0333          0.643 white
##  2 YILDIRIM     0.861   0.0271   0.0609   0.0135   0.0372          0.861 white
##  3 HAMILTON     0.702   0.237    0.0245   0.0054   0.0309          0.702 white
##  4 MOSHER       0.947   0.00410  0.0241   0.00640  0.0185          0.947 white
##  5 BARR         0.827   0.117    0.0226   0.00590  0.0271          0.827 white
##  6 GRAY         0.687   0.251    0.0241   0.0054   0.0324          0.687 white
##  7 MCMILLIAN    0.359   0.574    0.0189   0.00260  0.0463          0.574 Black
##  8 FORD         0.620   0.32     0.0237   0.0045   0.0313          0.620 white
##  9 STRZELECKA   0.666   0.0853   0.137    0.0797   0.0318          0.666 white
## 10 KIM          0.0252  0.00390  0.00650  0.945    0.0198          0.945 Asian
## # ... with 3,796 more rows
```

```
# keeping only the race and the surname
race = race %>% select(surname,race)

#Joining to the data frame
applications = applications %>% left_join(race, by = c("examiner_name_last" = "surname"))
```

Tenure

```
# get filling dates, start and end date and calculate the tenure
dates = applications %>% select(examiner_id, filing_date, appl_status_date) %>% mutate(start_date = ymd
  summarise(
    earliest = min(start_date, na.rm = TRUE),
    latest = max(end_date, na.rm = TRUE),
    tenure = interval(earliest, latest) %/% days(1)
    ) %>% filter(year(latest)<2018)

dates
```

```
## # A tibble: 5,625 x 4
##    examiner_id earliest   latest     tenure
##          <dbl> <date>     <date>      <dbl>
##  1       59012 2004-07-28 2015-07-24   4013
##  2       59025 2009-10-26 2017-05-18   2761
##  3       59030 2005-12-12 2017-05-22   4179
##  4       59040 2007-09-11 2017-05-23   3542
##  5       59052 2001-08-21 2007-02-28   2017
##  6       59054 2000-11-10 2016-12-23   5887
##  7       59055 2004-11-02 2007-12-26   1149
##  8       59056 2000-03-24 2017-05-22   6268
##  9       59074 2000-01-31 2017-03-17   6255
## 10       59081 2011-04-21 2017-05-19   2220
## # ... with 5,615 more rows
```

```
# Join to data frame
applications = applications %>% left_join(dates, by = "examiner_id")
applications
```

```
## # A tibble: 2,018,477 x 21
##    application_number filing_date examiner_name_last examiner_name_first
##    <chr>              <date>      <chr>              <chr>
##  1 08284457           2000-01-26  HOWARD             JACQUELINE
##  2 08413193           2000-10-11  YILDIRIM           BEKIR
##  3 08531853           2000-05-17  HAMILTON           CYNTHIA
##  4 08637752           2001-07-20  MOSHER             MARY
##  5 08682726           2000-04-10  BARR               MICHAEL
##  6 08687412           2000-04-28  GRAY               LINDA
##  7 08716371           2004-01-26  MCMILLIAN          KARA
##  8 08765941           2000-06-23  FORD               VANESSA
##  9 08776818           2000-02-04  STRZELECKA         TERESA
## 10 08809677           2002-02-20  KIM                SUN
## # ... with 2,018,467 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>, latest <date>, tenure <dbl>
```

Pick two workgroups you want to focus on (remember that a workgroup is

represented by the first 3 digits of 'examiner_art_unit' value)

```
group_162 = applications[substr(applications$examiner_art_unit, 1,3)==162,]

group_163 = applications[substr(applications$examiner_art_unit, 1,3)==163,]

summary(group_162)
```

```
##  application_number  filing_date         examiner_name_last examiner_name_first
##  Length:141390      Min.   :2000-01-03   Length:141390      Length:141390
##  Class :character   1st Qu.:2005-01-18   Class :character   Class :character
##  Mode  :character   Median :2008-11-25   Mode  :character   Mode  :character
##                     Mean   :2008-10-24
##                     3rd Qu.:2012-08-23
##                     Max.   :2017-05-09
##
##  examiner_name_middle  examiner_id    examiner_art_unit  uspc_class
##  Length:141390        Min.   :59440   Min.   :1620      Length:141390
##  Class :character     1st Qu.:65768   1st Qu.:1624      Class :character
##  Mode  :character     Median :73364   Median :1625      Mode  :character
##                       Mean   :78439   Mean   :1625
##                       3rd Qu.:93677   3rd Qu.:1626
##                       Max.   :99990   Max.   :1629
##                       NA's   :682
##  uspc_subclass      patent_number      patent_issue_date
##  Length:141390      Length:141390      Min.   :2000-08-08
##  Class :character   Class :character   1st Qu.:2006-11-07
##  Mode  :character   Mode  :character   Median :2011-04-19
##                                        Mean   :2010-06-28
##                                        3rd Qu.:2014-02-18
##                                        Max.   :2017-06-20
##                                        NA's   :57816
##   abandon_date        disposal_type      appl_status_code appl_status_date
##  Min.   :2000-06-05   Length:141390      Min.   :  1.0    Length:141390
##  1st Qu.:2009-02-18   Class :character   1st Qu.:150.0    Class :character
##  Median :2011-06-27   Mode  :character   Median :150.0    Mode  :character
##  Mean   :2011-01-30                      Mean   :161.3
##  3rd Qu.:2013-09-09                      3rd Qu.:161.0
##  Max.   :2017-06-05                      Max.   :454.0
##  NA's   :97057                           NA's   :262
##        tc          gender             race              earliest
##  Min.   :1600   Length:141390      Length:141390      Min.   :2000-01-03
##  1st Qu.:1600   Class :character   Class :character   1st Qu.:2000-01-07
##  Median :1600   Mode  :character   Mode  :character   Median :2000-02-22
##  Mean   :1600                                         Mean   :2001-06-10
##  3rd Qu.:1600                                         3rd Qu.:2002-10-22
##  Max.   :1600                                         Max.   :2012-07-25
##                                                       NA's   :4389
##      latest               tenure
##  Min.   :2001-09-23   Min.   : 614
##  1st Qu.:2017-05-19   1st Qu.:5282
```

```
##  Median :2017-05-22   Median :6262
##  Mean   :2017-05-03   Mean   :5806
##  3rd Qu.:2017-05-23   3rd Qu.:6340
##  Max.   :2017-11-08   Max.   :6518
##  NA's   :4389         NA's   :4389
```

```
summary(group_163)
```

```
##  application_number  filing_date         examiner_name_last examiner_name_first
##  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
##  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               tenure
##  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
```

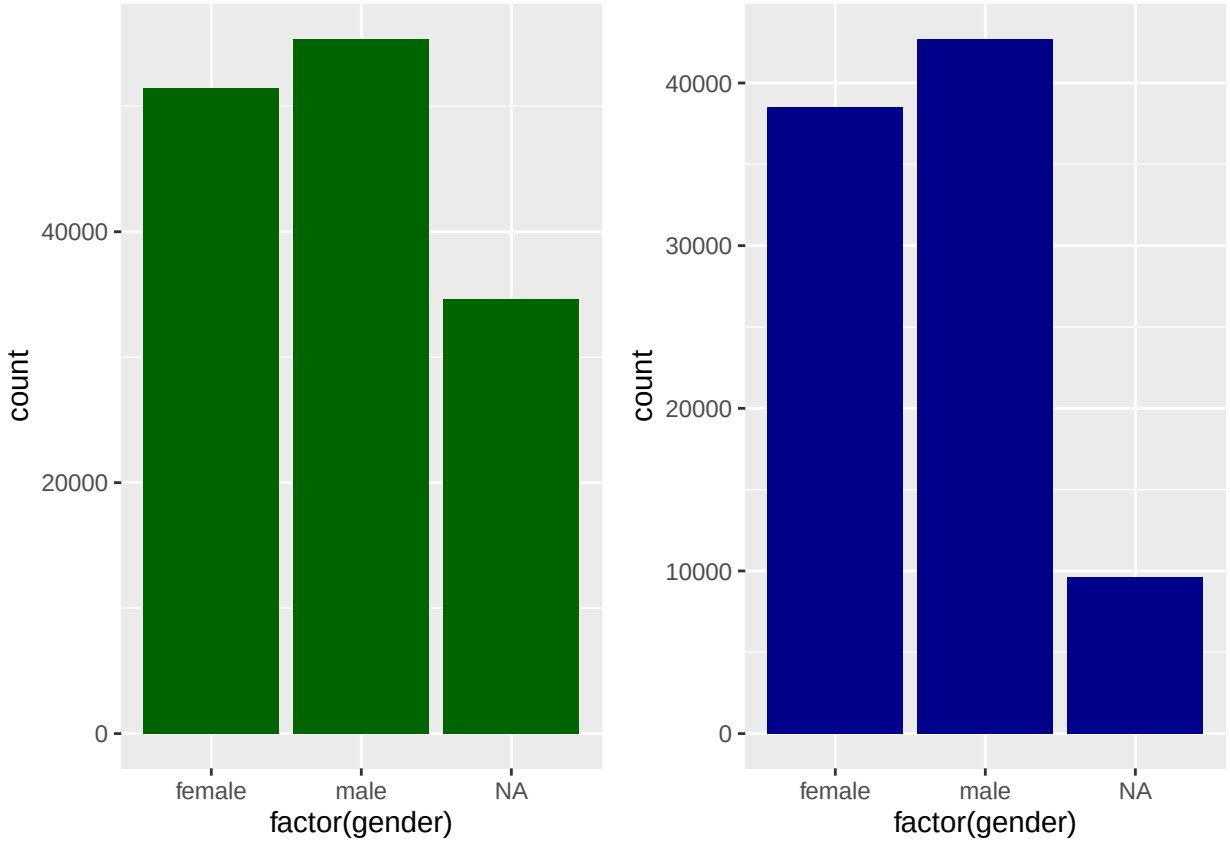
```
library(gridExtra)

par(mfrow=c(1,2))

plot1 = ggplot(group_162, aes(x = factor(gender)))+geom_bar(fill="darkgreen")

plot2 = ggplot(group_163, aes(x = factor(gender)))+geom_bar(fill="darkblue")

grid.arrange(plot1, plot2, ncol=2)
```



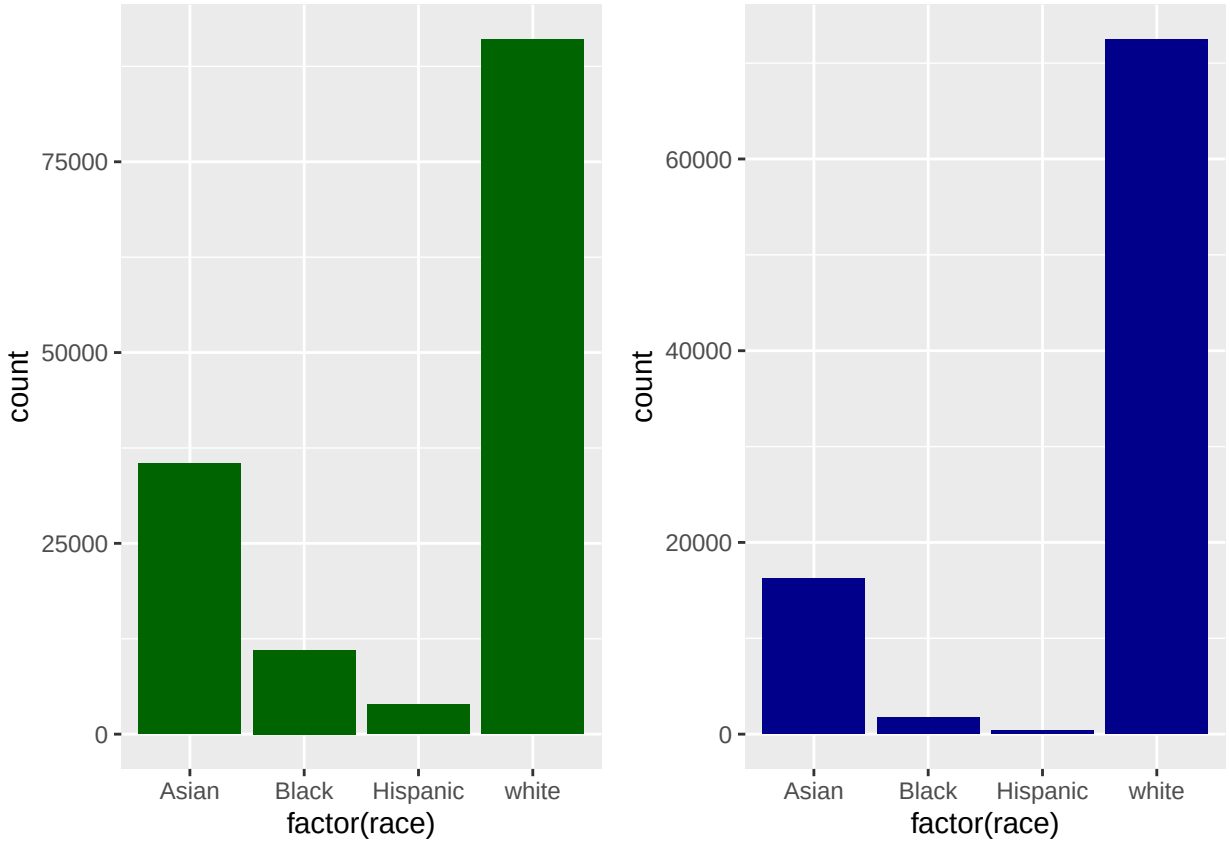
Both groups seem to have more examiners identified as males. However, the proportion of NA data is more important in the group 162 than in the group 163. We also notice that group 162 seems to have more people in it as there is more male, female and NA data than in group 163.

```
par(mfrow=c(1,2))

plot3 = ggplot(group_162, aes(x = factor(race)))+geom_bar(fill="darkgreen")

plot4 = ggplot(group_163, aes(x = factor(race)))+geom_bar(fill="darkblue")

grid.arrange(plot3, plot4, ncol=2)
```



The most frequent race in both groups is White. It represents the majority. Also, in both groups, the second most frequent race is Asian people. The number of Black and Hispanic people in group 163 is very low. In the group 162, there is more Black and Hispanic people even though they represent a minority compared to the white and asian people.

Advice networks

```
# get the work groups of interest
art_unit = distinct(subset(applications, select=c(examiner_art_unit, examiner_id)))

# add work group to use it as an indicator in the graph and select the groups of interest

art_unit$work_group = substr(art_unit$examiner_art_unit, 1,3)

art_unit = art_unit[art_unit$work_group==162 | art_unit$work_group==163,]

# Merging
my_merger = merge(x=edges, y=art_unit, by.x="ego_examiner_id", by.y="examiner_id", all.x=TRUE) %>% rena

# drop the NA values (other groups than 162 or 163)
my_merger = drop_na(my_merger)

# Doing the same for the alter examiners
my_merger = merge(x=my_merger, y=art_unit, by.x="alter_examiner_id", by.y="examiner_id", all.x=TRUE)%>%

my_merger = drop_na(my_merger)
```

We have 592 edges left that represents examiners from groups 162 and 163

```
# Ego Nodes vs alter Nodes
ego_nodes = subset(my_merger, select=c(ego_examiner_id, art_unit_ego, work_group_ego))

ego_nodes = ego_nodes %>% rename(examiner_id=ego_examiner_id,art_unit=art_unit_ego,work_group=work_grou

alter_nodes = subset(my_merger, select=c(alter_examiner_id, art_unit_alter, work_group_alter))

alter_nodes = alter_nodes %>% rename(examiner_id=alter_examiner_id,art_unit=art_unit_alter, work_group=

nodes = distinct(rbind(ego_nodes, alter_nodes)) %>% group_by(examiner_id) %>% summarise(examiner_id = f

network = graph_from_data_frame(d=my_merger, vertices=nodes, directed=TRUE)
network
```

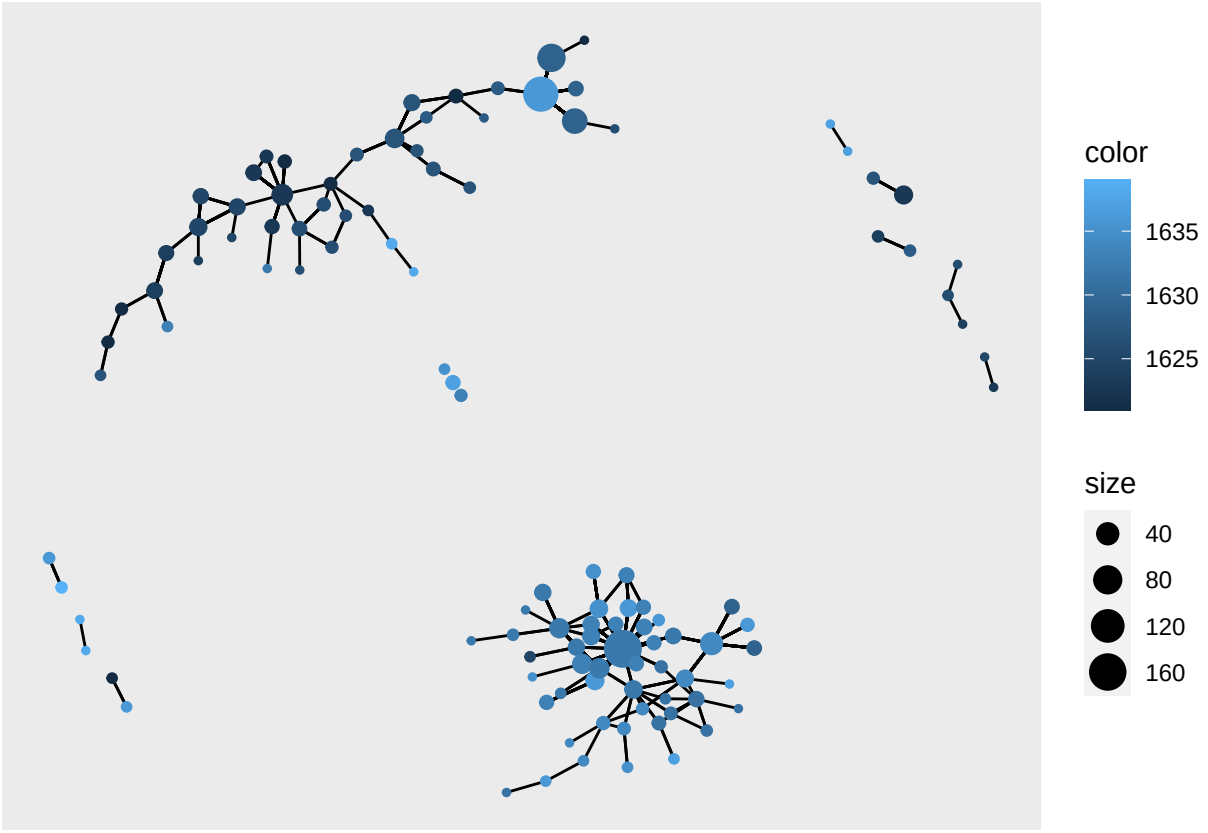
```
## IGRAPH 5360559 DN-- 112 592 --
## + attr: name (v/c), art_unit (v/n), work_group (v/c),
## | application_number (e/c), advice_date (e/n), art_unit_ego (e/n),
## | work_group_ego (e/c), art_unit_alter (e/n), work_group_alter (e/c)
## + edges from 5360559 (vertex names):
##  [1] 61519->72253 61519->61519 61519->72253 61519->61519 61519->72253
##  [6] 61519->61519 62253->67690 62253->67690 62253->67690 62253->67690
## [11] 62312->61519 62312->94257 62312->98614 62312->63030 62312->66971
## [16] 62312->66971 62312->95543 62312->95543 62312->66971 62312->98614
## [21] 62312->86861 62312->59156 62312->61519 62312->63030 62312->66971
## [26] 62312->61519 62312->98614 62312->66971 62312->66971 62312->98614
## + ... omitted several edges
```

```
#Get centrality scores

my_degree <- round(degree(network, v=V(network)),2)
my_betweenness <- round(betweenness(network),2)
my_closeness <- round(closeness(network),2)

V(network)$size = my_degree
V(network)$bet = my_betweenness
V(network)$clo = my_closeness
V(network)$color = nodes$art_unit

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



The visualization shows that the majority of the examiners are getting advised by people from the same art unit. However it seems that some of the examiners are left out of the two big clusters in the graph. Some are are isolated and other are in groups of 2 or 3. Even if the majority of the connections are made within same groups, some of them (as shown in the network at the top) may connect with examiners from other work groups.

```
centralities = data.frame(cbind(my_degree, my_betweenness, my_closeness))


centralities[order(-my_degree, -my_betweenness),]
```

```
##       my_degree my_betweenness my_closeness
## 72253       165             54         0.07
## 72814       135              0         0.25
## 73364        73              0          NaN
## 81959        54              0          NaN
## 71087        39              0         0.20
## 67690        31              0          NaN
## 98614        26              0          NaN
## 62312        25              0         0.02
## 71931        25              0          NaN
## 67256        22              0         0.14
## 64073        20              0          NaN
## 65111        20              0         1.00
## 67515        18              0         0.03
## 94257        18              0          NaN
## 95446        17              0          NaN
```

```
## 95543        16              1         1.00
## 94579        16              0          NaN
## 73880        15              0          NaN
## 61519        14             27         0.03
## 66971        14              0          NaN
## 86861        14              0          NaN
## 62661        13              0          NaN
## 98700        13              0         0.25
## 61299        12              0          NaN
## 63977        12              0          NaN
## 65713        12              0         0.50
## 70767        12              0          NaN
## 68166        11              0         1.00
## 96339        11              0         0.14
## 94070        10              0         0.50
## 97102        10              0         0.33
## 69138         9              0          NaN
## 71445         9              0          NaN
## 81337         9              0          NaN
## 91989         9              0          NaN
## 93403         9              0         0.25
## 73777         8              1         1.00
## 63244         8              0         1.00
## 67173         8              0          NaN
## 68141         8              0          NaN
## 78019         8              0          NaN
## 91374         8              0          NaN
## 95565         8              0         1.00
## 68695         7              1         1.00
## 67581         7              0         0.25
## 97242         7              0          NaN
## 60302         6              0          NaN
## 70993         6              0          NaN
## 81865         6              0         1.00
## 95091         6              0          NaN
## 96643         5              1         1.00
## 59407         5              0          NaN
## 62253         5              0         1.00
## 85216         5              0          NaN
## 87486         5              0         0.14
## 96898         5              0          NaN
## 70206         4              0          NaN
## 71120         4              0         0.50
## 88508         4              0          NaN
## 89882         4              0         0.50
## 94925         4              0          NaN
## 97586         4              0          NaN
## 99047         4              0          NaN
## 99381         4              0          NaN
## 75034         3              1         1.00
## 61529         3              0          NaN
## 63030         3              0          NaN
## 65131         3              0          NaN
## 65537         3              0         1.00
```

```
## 80106         3             0         1.00
## 81211         3             0          NaN
## 91956         3             0          NaN
## 92219         3             0          NaN
## 98182         3             0         1.00
## 88443         2             1         1.00
## 61417         2             0          NaN
## 63234         2             0          NaN
## 67731         2             0         1.00
## 72848         2             0          NaN
## 73239         2             0         1.00
## 79495         2             0          NaN
## 82997         2             0         0.02
## 84925         2             0         1.00
## 91337         2             0         0.33
## 91747         2             0         1.00
## 92902         2             0          NaN
## 93677         2             0          NaN
## 95525         2             0         0.50
## 97603         2             0         1.00
## 59156         1             0          NaN
## 61416         1             0          NaN
## 62621         1             0         1.00
## 63190         1             0          NaN
## 63822         1             0         1.00
## 65536         1             0         1.00
## 65737         1             0         1.00
## 67753         1             0          NaN
## 68339         1             0         1.00
## 69896         1             0          NaN
## 71123         1             0         1.00
## 71385         1             0          NaN
## 71853         1             0         1.00
## 72941         1             0          NaN
## 72995         1             0         1.00
## 77348         1             0         1.00
## 80247         1             0         0.33
## 85736         1             0         1.00
## 88294         1             0         1.00
## 93955         1             0         0.33
## 94915         1             0         1.00
## 97520         1             0         1.00
## 99424         1             0          NaN
```

We see that examiner 72253 has the biggest degree of centrality and the highest associated betweeness. The examiner 72814 seems also interesting because of it's high degree of centrality. Let's explore these two.

```
applications[applications$examiner_id==72253,]
```

```
## # A tibble: 9,628 x 21
##    application_number filing_date examiner_name_last examiner_name_first
##    <chr>              <date>      <chr>              <chr>
##  1 <NA>               NA          <NA>               <NA>
##  2 09242244           2000-02-29  WOITACH            JOSEPH
##  3 09402130           2000-02-02  WOITACH            JOSEPH
```

```
##  4 09402527           2000-01-03  WOITACH            JOSEPH
##  5 09403707           2000-03-17  WOITACH            JOSEPH
##  6 09423935           2000-03-13  WOITACH            JOSEPH
##  7 09446717           2000-04-13  WOITACH            JOSEPH
##  8 <NA>               NA          <NA>               <NA>
##  9 09463276           2000-05-12  WOITACH            JOSEPH
## 10 <NA>               NA          <NA>               <NA>
## # ... with 9,618 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>, latest <date>, tenure <dbl>
```

This examiner has been working for 17 years. He has been in the art unit 1632 during all this time. This may explain his importance in the network as he is very experienced and should have developed some strong relationships and influence. This person seems to correspond to the central node of the network at the bottom of the graph.

```
applications[applications$examiner_id==72814,]
```

```
## # A tibble: 9,550 x 21
##    application_number filing_date examiner_name_last examiner_name_first
##    <chr>              <date>      <chr>              <chr>
##  1 <NA>               NA          <NA>               <NA>
##  2 <NA>               NA          <NA>               <NA>
##  3 <NA>               NA          <NA>               <NA>
##  4 <NA>               NA          <NA>               <NA>
##  5 <NA>               NA          <NA>               <NA>
##  6 <NA>               NA          <NA>               <NA>
##  7 <NA>               NA          <NA>               <NA>
##  8 09486623           2000-07-06  MARSCHEL           ARDIN
##  9 <NA>               NA          <NA>               <NA>
## 10 09488339           2000-01-20  MARSCHEL           ARDIN
## # ... with 9,540 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>, latest <date>, tenure <dbl>
```

This examiner also worked during 17 years. However he started in the unit 1631 before moving to unit 1634. This may explain why he has connections with other work groups. As the previous examiner, his importance in the network may be explained by his experience and the relationships he may have built during this 17 years. This examiner corresponds to the biggest node of the network at the top in the graph.