

MATH 4330 Activity 2

Chi Zhang, Ya Gao, Heajung Nam (Team Bose)

2017-09-17

MATH 4330 Activity 2

Chi Zhang, Ya Gao, Heajung Nam (Team Bose)

September 17, 2017

- Introduction
- Data Biography
- Data Directory
- Interesting Questions And Solving By Using Data Displays
- Conclusions
- References

Last updated on: September 18, 2017 at 22:13

Introduction

The goal of this report is to determine whether the Canadian Refugee-Determination system is fair or not. In order to investigate, we have analyzed the data of Court decisions on refugee-determination to the applications for leave to appeal filed in 1990. Among the whole data, 384 cases were sampled and analyzed. The general information of the data set was derived using the `xqplots`. Based on the simple observations, we came up with three possible questions. In order to answer these questions, we checked the existence of correlation between variables by using bar charts, plots, and graphs.

Data Biography

Data set used in this report could be downloaded from the following directory:

```
fox_data <- "http://socserv.socsci.mcmaster.ca/jfox/Books/Applied-Regression-3E/datasets/"
```

The data set used is called "Greene.txt" and downloaded for reporting.

```
knitr::opts_chunk$set(warning=FALSE,comment=NA)
options(useFancyQuotes = FALSE)
library(car)
library(spida2)
library(lattice)
library(latticeExtra)

## Loading required package: RColorBrewer

library(Hmisc)

## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
```

```

##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:latticeExtra':
##
##   layer

## The following object is masked from 'package:spida2':
##
##   labs

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:spida2':
##
##   fillin, na.include

## The following objects are masked from 'package:base':
##
##   format.pval, round.POSIXt, trunc.POSIXt, units

download.file(paste0(fox_data, 'Greene.txt'), 'Greene.txt') #data
list.files(pattern='Greene')

## [1] "Greene.txt"                "IanGreenePaulShafferLeave.pdf"

appeals <- read.table('Greene.txt')
N<-length(appeals$decision)
appeals0 <- appeals

```

except "rater", I think.

The original data can be found in the paper called, "Leave to Appeal and Leave to Commence Judicial Review in Canada's Refugee-Determination System: Is the Process Fair?" by Ian Greene and Paul Shaffer. All the files relate with Court decisions on refugee-determination to the applications for leave to appeal filed in 1990 were stored in chronological order of files in boxes of the Federal Court of Appeal office in Ottawa.

The sample frame is a list of approximately two thousand applications for leave to appeal filed in 1990. A size of 611 samples was generated by pulling every third file in the total. If encountered a file missing due to lack of material, then a file next to it in the box would be substituted into. Eventually, a size of 608 samples had been generated since information about the disposition was missing from three of the files. This method is called systematic sampling, which is a better method than simple random sample(SRS) in an ordered sample frame. According to the paper, authors did not set any sample weights to adjust possible biases. They do, however, considered the association between judges and decisions under two possible variable, which are the countries that applicants came from and the time period that judges made decisions. They group different region by geographic locations and success rates respectively. Furthermore, they considered the decision made by judges in different time periods. All of them lead to insignificant results. That is, leave to appeal decisions were made by judges are highly dependent to themselves. Two factors, time period and countries that applicants came from, have little contribution to the decision compared to judges' decisions.

Here is the brief summary of the data:

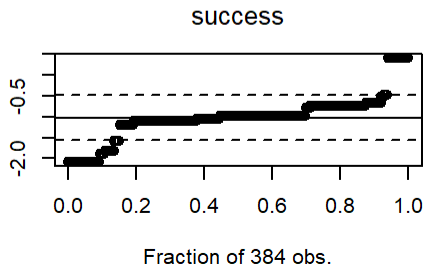
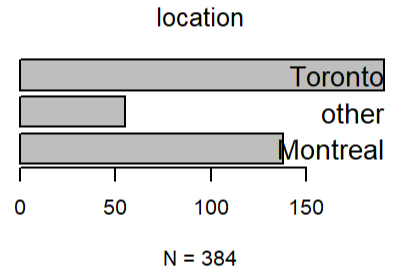
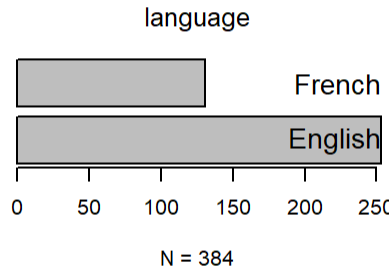
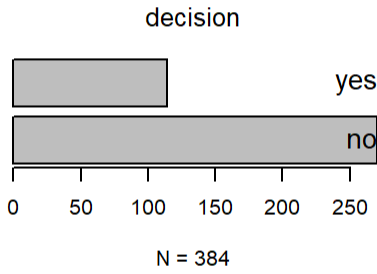
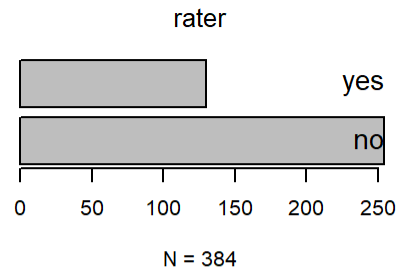
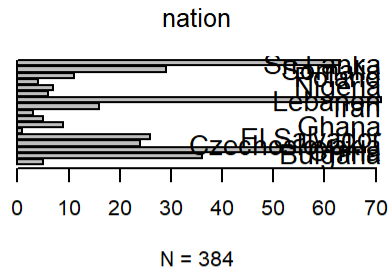
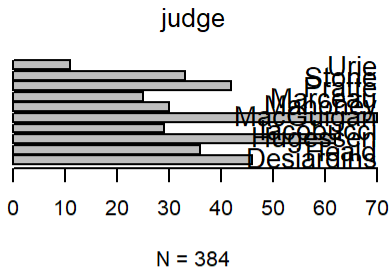
```

xqplot(appeals)

```

small Some grammatical issues here. I'd be happy to meet.

This seems like a very important statement but I don't understand it.



`describe(appeals)`

appeals

7 Variables 384 Observations

judge

n	missing	distinct
384	0	10

Value	Desjardins	Heald	Hugessen	Iacobucci	MacGuigan
Frequency	46	36	62	29	70
Proportion	0.120	0.094	0.161	0.076	0.182

Value	Mahoney	Marceau	Pratte	Stone	Urie
Frequency	30	25	42	33	11
Proportion	0.078	0.065	0.109	0.086	0.029

nation

n	missing	distinct
384	0	17

Argentina (5, 0.013), Bulgaria (36, 0.094), China (68, 0.177), Czechoslovakia (24, 0.062), El.Salvador (26, 0.068), Fiji (1, 0.003), Ghana (9, 0.023), Guatemala (5, 0.013), India (3, 0.008), Iran (16, 0.042), Lebanon (71, 0.185), Nicaragua (6, 0.016), Nigeria (7, 0.018), Pakistan (4, 0.010), Poland (11, 0.029), Somalia (29, 0.076), Sri.Lanka (63, 0.164)

rater

```

      n missing distinct
384      0          2

Value      no  yes
Frequency  254 130
Proportion 0.661 0.339
-----
decision
      n missing distinct
384      0          2

Value      no  yes
Frequency  270 114
Proportion 0.703 0.297
-----
language
      n missing distinct
384      0          2

Value      English  French
Frequency   253    131
Proportion  0.659  0.341
-----
location
      n missing distinct
384      0          3

Value      Montreal  other  Toronto
Frequency   138     55    191
Proportion  0.359  0.143  0.497
-----
success
      n missing distinct  Info  Mean  Gmd  .05  .10
384      0          14  0.972  -1.02  0.5357  -2.0907  -1.9010
.25  .50  .75  .90  .95
-1.0986 -0.9946 -0.7538 -0.6633  0.4055

Value      -2.09074 -1.90096 -1.81529 -1.58563 -1.20831 -1.09861 -1.04597
Frequency   36      5      11      6      16      71      26
Proportion  0.094  0.013  0.029  0.016  0.042  0.185  0.068

Value      -0.99462 -0.80012 -0.75377 -0.66329 -0.53222 -0.48955  0.40547
Frequency   97      5      63      17      3      4      24
Proportion  0.253  0.013  0.164  0.044  0.008  0.010  0.062
-----

```

Data Directory

```
dim(appeals)
```

```
[1] 384 7
```

We have 7 variables in the data set:

1.judge, 2.nation, 3.rater, 4.decision, 5.language, 6.location, 7.success.

and here are the descriptions for the each variable:

1.judge: Name of judge hearing case. (Desjardins, Heald, Hugessen, Iacobucci, MacGuigan, Mahoney, Marceau, Pratte, Stone, Urie)

2.nation: Nation of origin of claimant. (Argentina, Bulgaria, China, Czechoslovakia, El.Salvador, Fiji, Ghana, Guatemala, India, Iran, Lebanon, Nicaragua, Nigeria, Pakistan, Poland, Somalia, Sri.Lanka)

3.rater: Judgement of independent rater. (No, case has no merit; Yes, case has some merit. Leave to appeal should be granted.)

4.decision: Judge's decision. (No, leave to appeal not granted; Yes, leave to appeal granted)

5.language: Language of case. (English, French)

6.location: Location of the original refugee claim. (Montreal, Toronto, Other)

7.success: Logit of success rate for all cases from the applicant's nation.

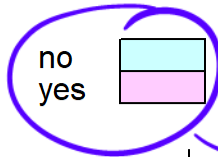
Interesting Questions And Solving By Using Data Displays

By looking at the data set, our group came with three possible questions.

1. Are decisions of judges similar to each other? Was there any specific judge that makes positive decisions on the appeals more frequent than others?

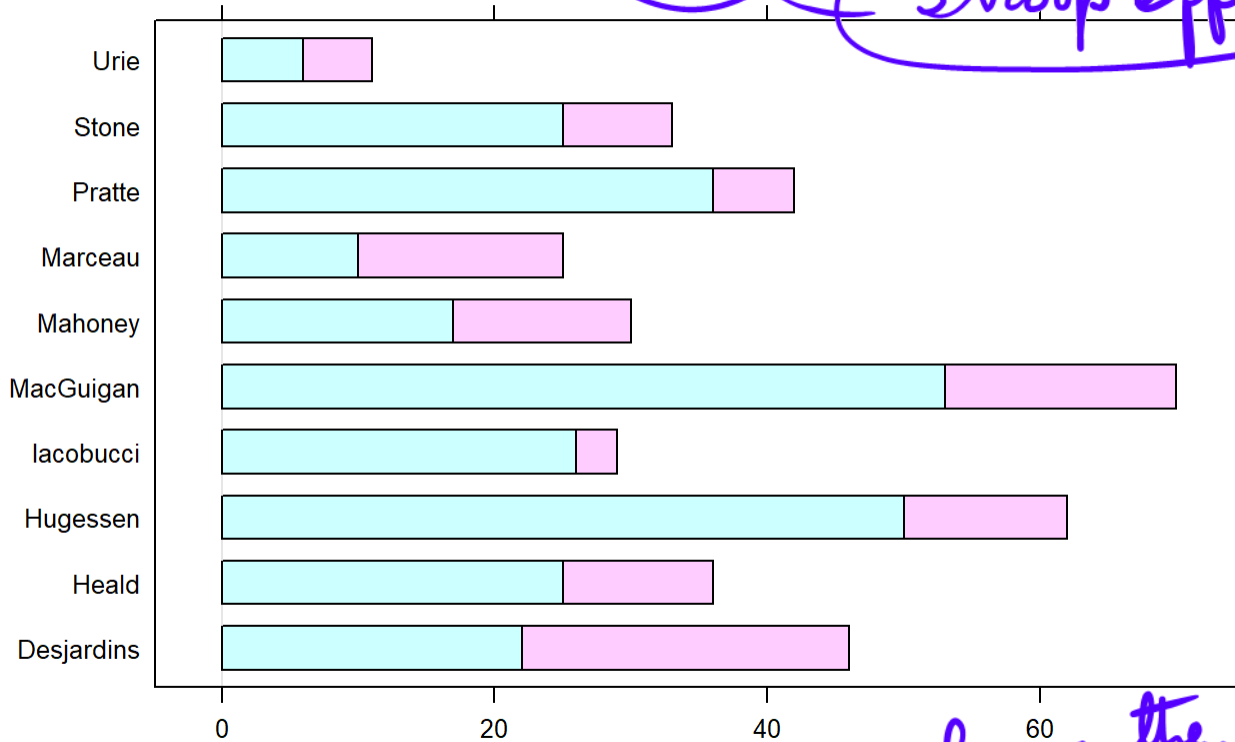
First, the table based on judges and their decisions was observed:

```
tab_(appeals,~judge+decision) %>%  
  barchart(auto.key=T)
```



Good - but is it because of the judge or because of the cases?

Stroop effect!



not really - it shows the frequency. Could you show the proportion?

The above bar chart shows the proportion of granted cases and non-granted cases for each judge. It seems like some judges are more likely to grant a leave of appeal for refugees than others. From the above graph, we can see that the Judge Marceau is most likely to grant an appeal and the judge Iacobucci is the least likely to grant an appeal.

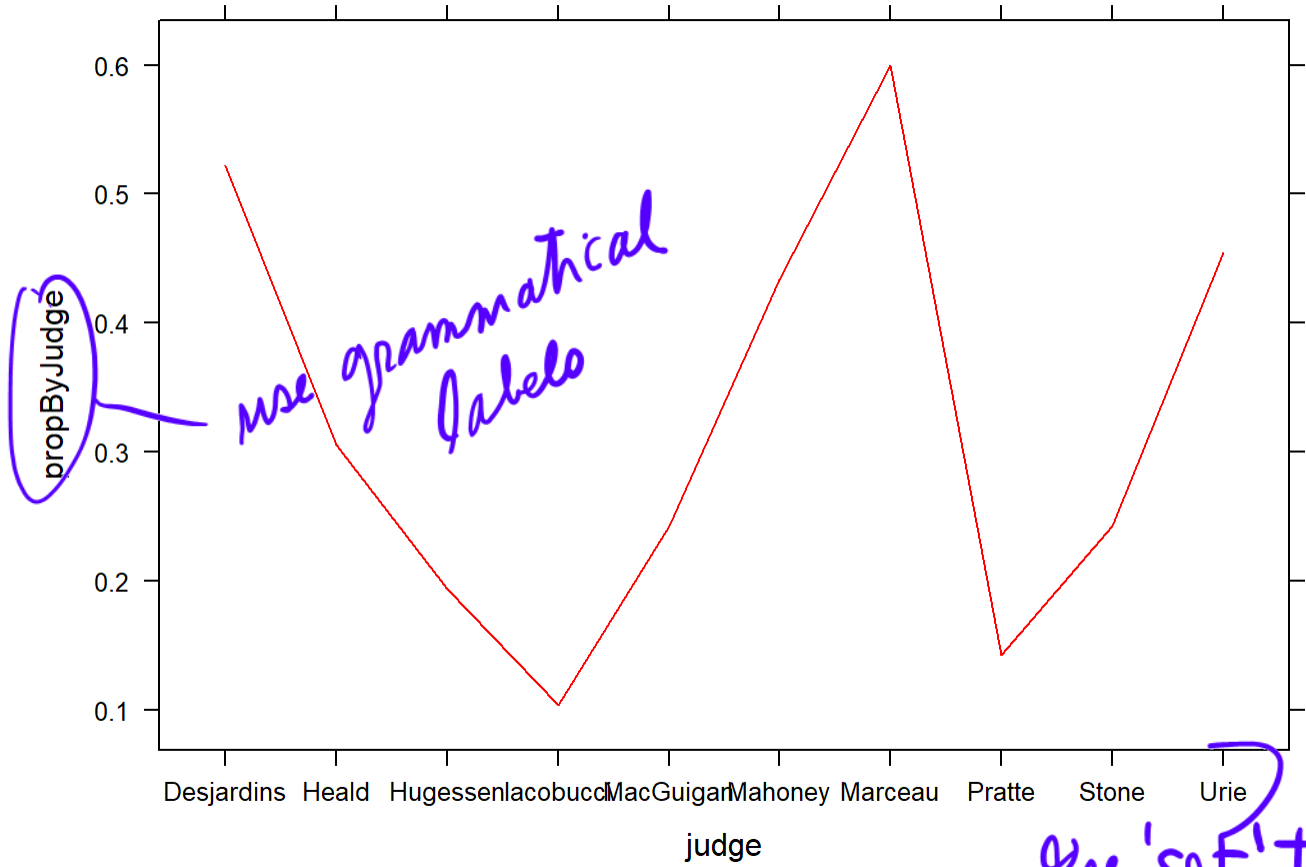
When we plot the proportion of each judge in the total number of grants, we would get the following:

```

appeals0$numD <- appeals0[,"decision"]=="yes"
appeals0$numD <- as.numeric(appeals0$numD)
appeals0$propByJudge <- with(appeals0,capply(appeals0$numD,appeals0$judge,mean))
appeals0 <- sortdf(appeals0,~judge)
xyplot(propByJudge~judge,appeals0,type = 'l', col='red')

```

Wow!!



Please note that the confidence interval of decision data was 95%.

```

pr <- sum(appeals$decision == 'yes')/length(appeals$decision)
dstd <- sqrt(pr*(1-pr)/length(appeals$decision))
left <- pr - 2*dstd
right <- pr + 2*dstd

```

use 'sort' to rotate
very good.

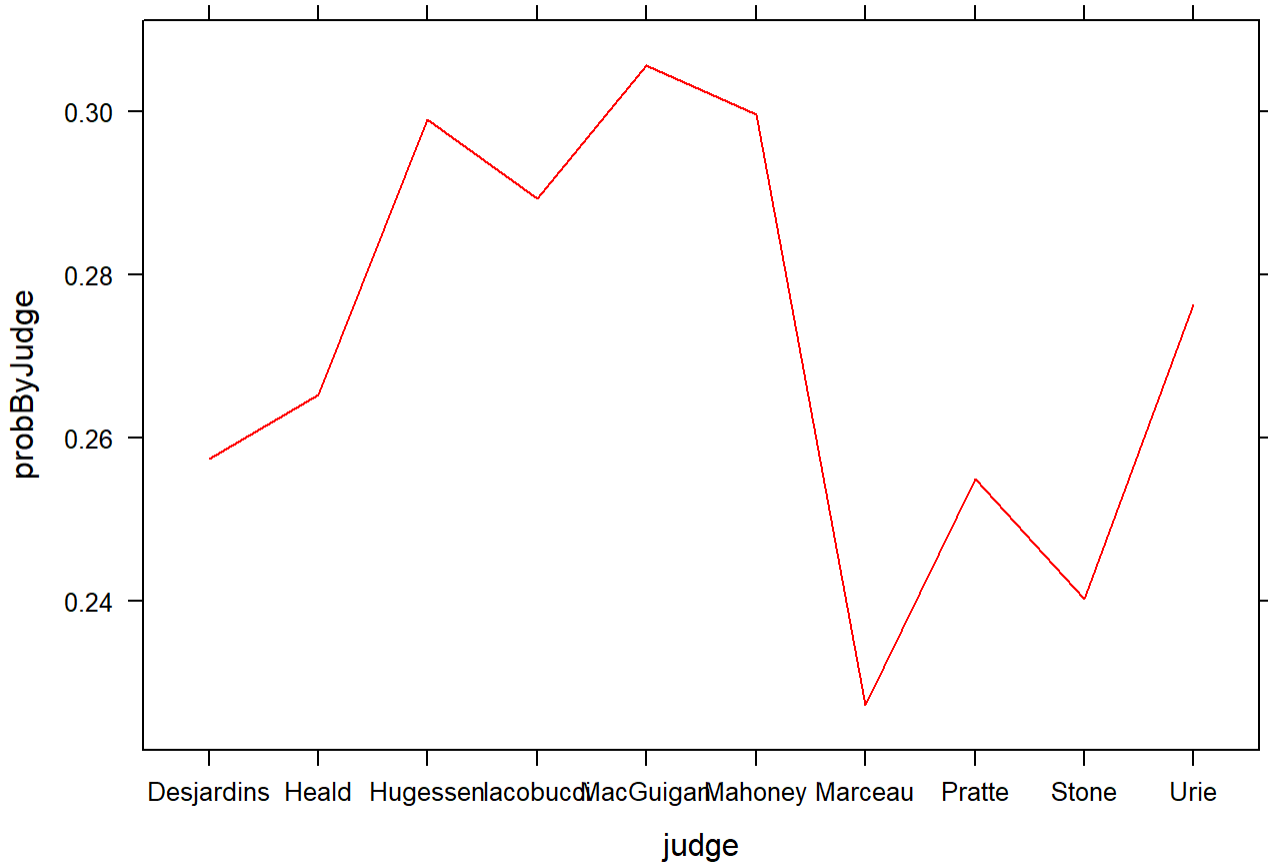
It turns out that left(= 0.250244) and right(= 0.343505). We could have a 95% Confidence interval with 4 decimals ([0.2502, 0.3435]). This implies that the possibility that different judges will make different decision is significant. Moreover, we could obtain much detailed information of judges' decisions by using the given success factor in the data set. We plotted the average success rate of appeal under each judge.

```

prob <- function(x){
  1/(1+exp(-x))
}
dpr <- prob(appeals$success)
appeals0$sucProb <-dpr
appeals0$probByJudge <- with(appeals0,capply(appeals0$sucProb,appeals0$judge,mean))
appeals0 <- sortdf(appeals0,~judge)
xyplot(probByJudge~judge,appeals0,type = 'l', col='red')

```

not sure what's going on here



The above plot shows the actual number of grants given by each judge in the overall number of grants given by all judges. We can conclude that there were judges who would more likely to grant appeals over other judges.

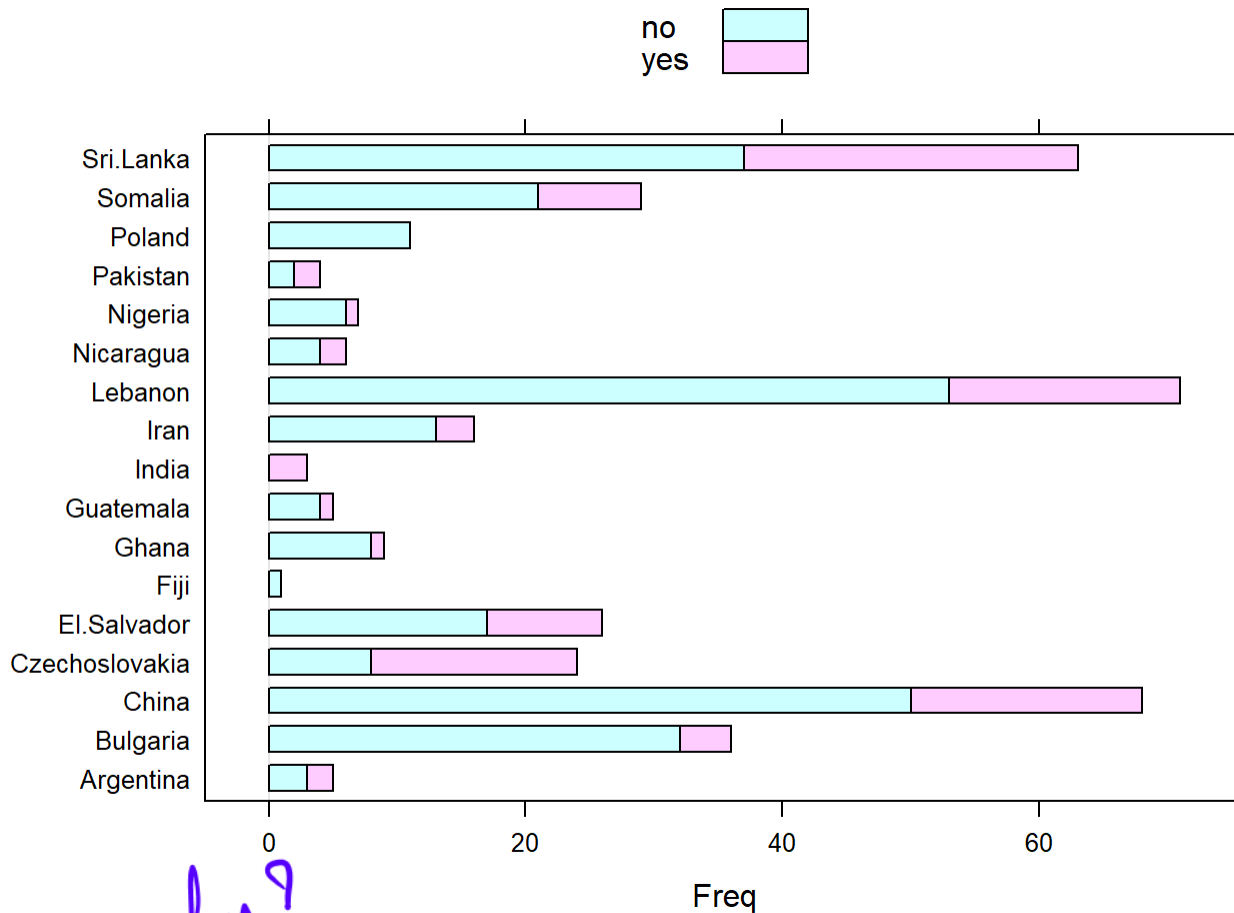
2. Does some judges have certain preference on specific nation?

A basic graph has shown the following:

```
tab_(appeals, ~nation+decision) %>%
  barghant(auto.key=T)
```

Good-

It seems to me that the previous plot addresses the question but not this one.

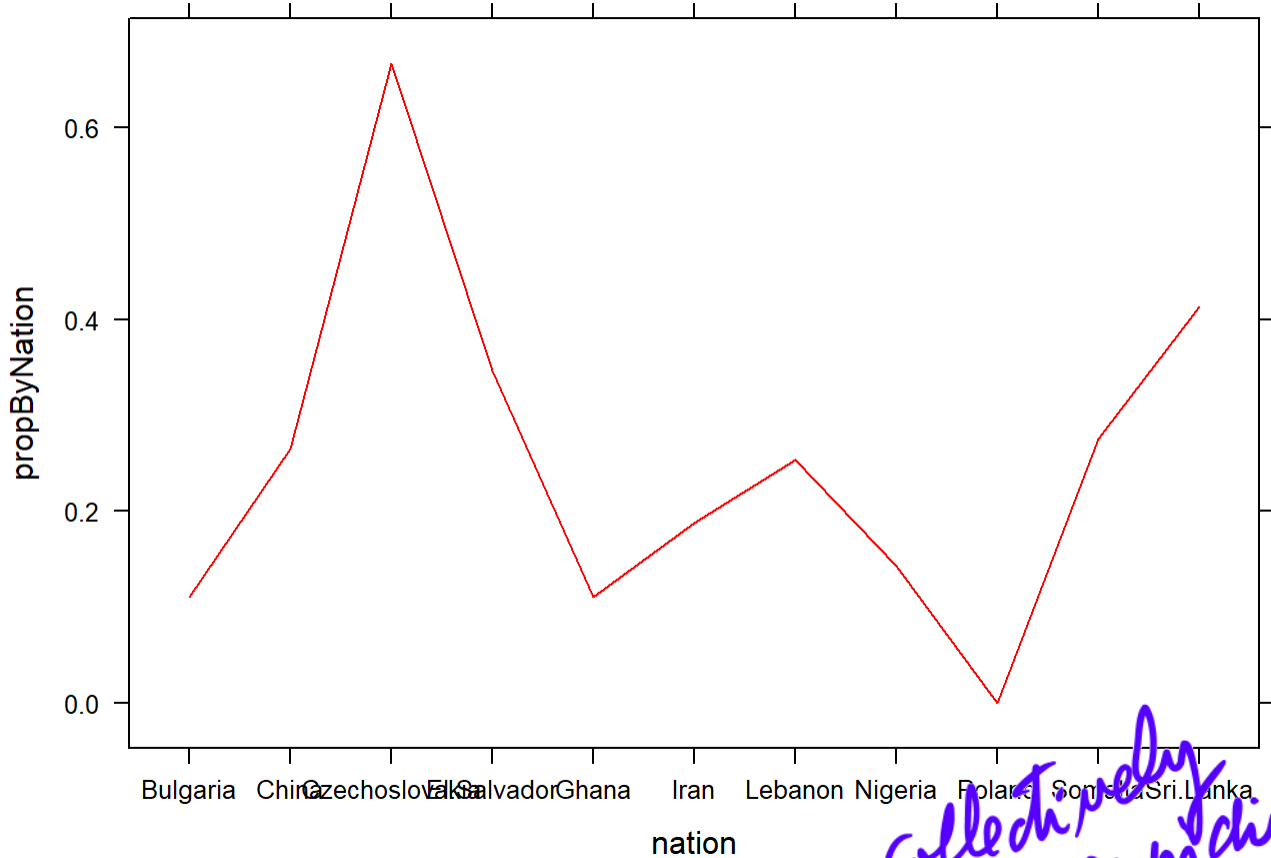


why?
 We simply following the same methodology used in the first question. However, we have noticed some nations occur less than 5 times. So we better subtract those nations from the total number of cases which is 384 to derive the reasonable sample variance.

```

appeals0$propByNation <- with(appeals0, capply(appeals0$numD, appeals0$nation, mean))
appeals0 <- sortdf(appeals0, ~nation)
appeals0$isBig <- appeals0[,"nation"] == "Bulgaria" | appeals0[,"nation"] == "China" | appeals0[,"nation"] == "Cze
  appeals0[,"nation"] == "El.Salvador" | appeals0[,"nation"] == "Ghana" | appeals0[,"nation"] == "Iran" | appeals0[,
  appeals0[,"nation"] == "Nicaraguar" | appeals0[,"nation"] == "Nigeria" | appeals0[,"nation"] == "Somalia" | appeal
  appeals0[,"nation"] == "Poland"
xyplot(propByNation~nation, type = 'l', col='red', subset(appeals0, isBig == TRUE))

```

collected individually

Based on the above plot, it seems reasonable to believe that judges do have some preference in applicant's nationality. Applicants from Czechoslovakia were the most successful of getting the appeals granted and applicants from Poland were the least successful of getting the appeals granted.

3. Was judge's decision correlated with rater's decision on most cases? Do judges always agree with raters?

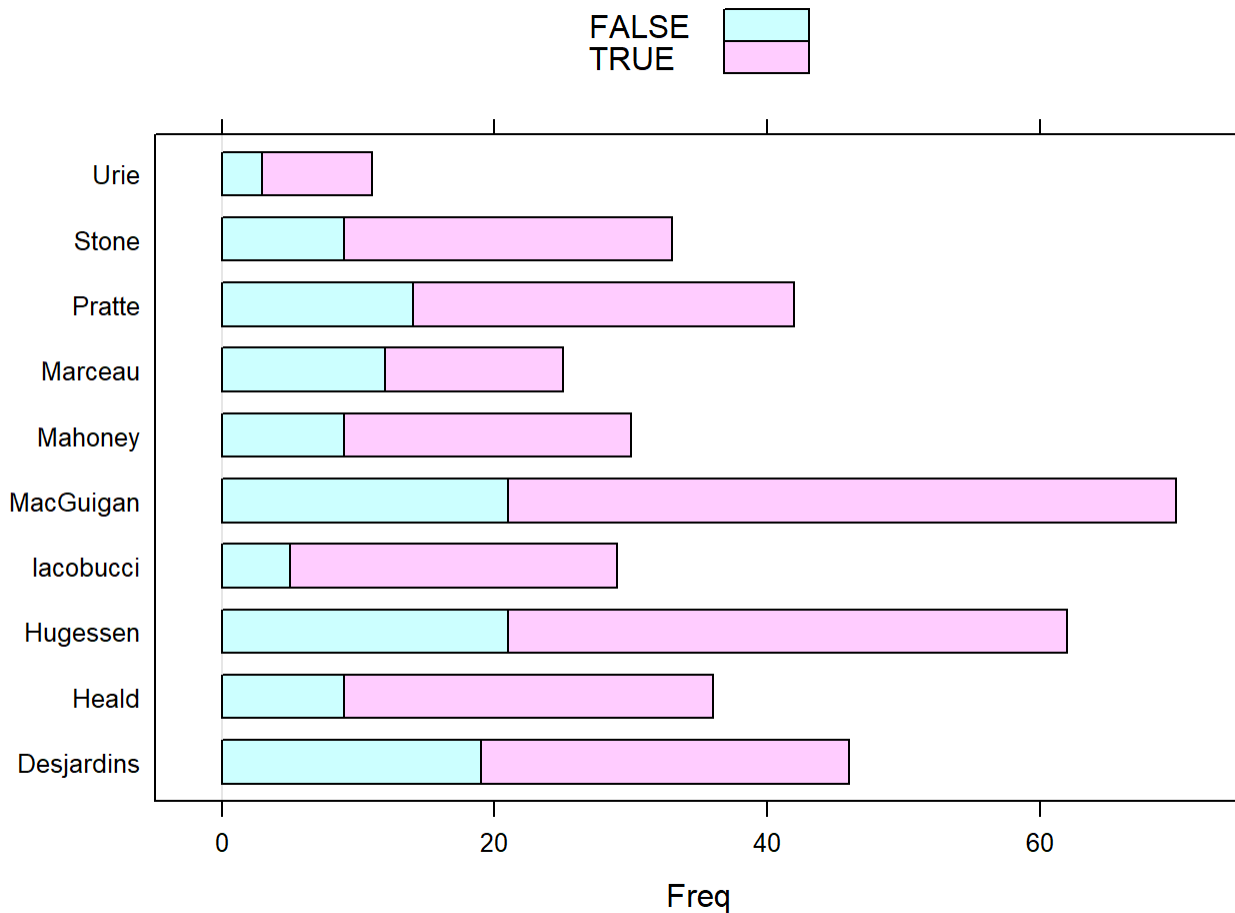


We can come with a basic graph as the following:

```

appeals0$diff <- appeals0["decision"] == appeals0["rater"]
tab_(appeals0, ~judge+diff)%>%
  barchart(auto.key=T)

```



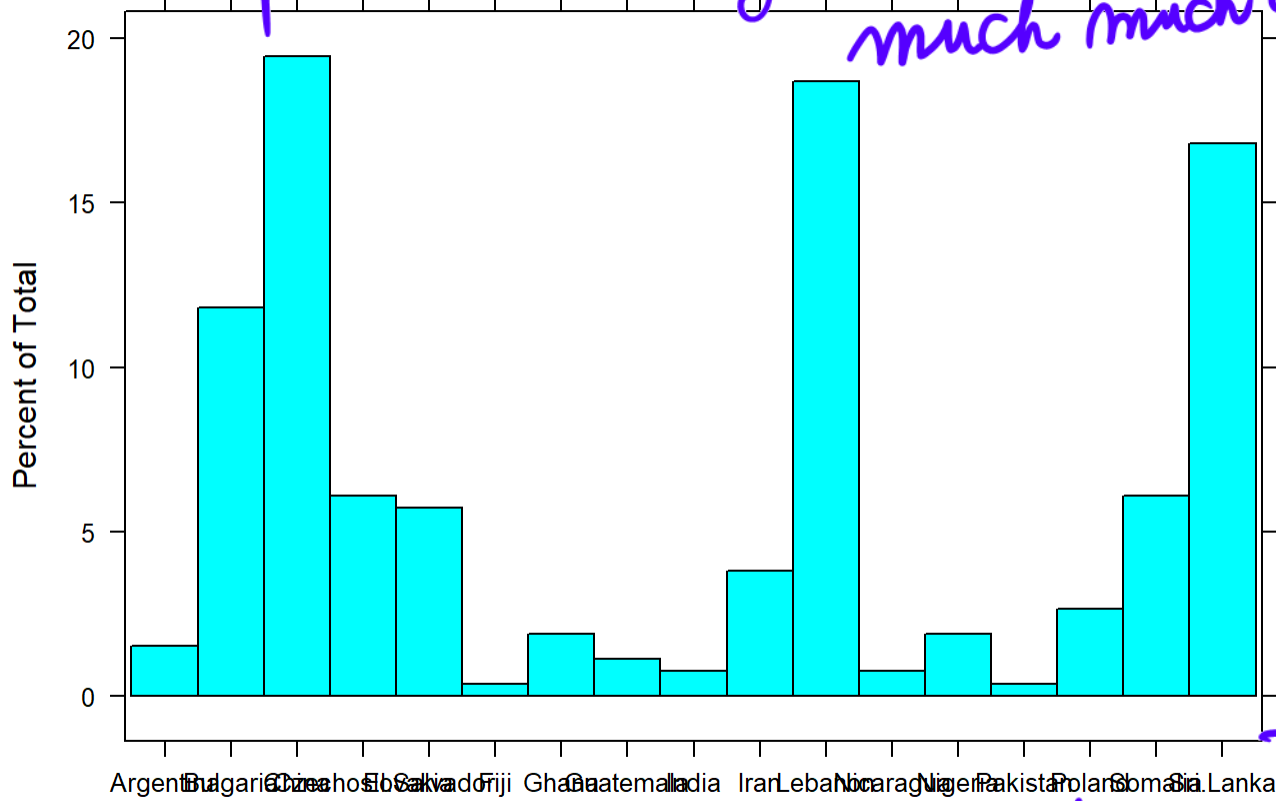
It shows the number of agreements ('yes') and disagreements ('no') between judges and raters for each judge. We can observe that judges agreed with raters in more cases than the cases where they disagreed with raters.

We could also convert the above graph into a histogram. This time, let's take a look at the number of agreements and disagreements between judges and raters for cases of each nation.

```
histogram(appeals0[appeals0[,"diff"]==TRUE,]$nation)
```

But is this a good measure of agreement... this leads to a long story

I think you're trying to do something important here but you need to make it much much clearer.



appeals0[appeals0[, "diff"] == TRUE,]\$nation

use 'not

better labels

According to the histogram, it seems like cases of applicants from China, Lebanon, and Sri Lanka were the most controversial ones. In other means, there were higher number of disagreements between judges and raters for cases of the applicants from the above countries.

To see the proportion of disagreements in the above cases, We can construct a table.

```
a<- sum(as.numeric(appeals0[appeals0[, "diff"]==FALSE,]$nation == "China"))/sum(as.numeric(appeals0[, "nation"] == "China"))
b<- sum(as.numeric(appeals0[appeals0[, "diff"]==FALSE,]$nation == "Lebanon"))/sum(as.numeric(appeals0[, "nation"] == "Lebanon"))
c<- sum(as.numeric(appeals0[appeals0[, "diff"]==FALSE,]$nation == "Sri.Lanka"))/sum(as.numeric(appeals0[, "nation"] == "Sri.Lanka"))
pp<- matrix(c(a,b,c),ncol=3,byrow = TRUE)
colnames(pp) <- c("China", "Lebanon", "Sri.Lanka")
```

The table is shown as the following:

```
pp
      China Lebanon Sri.Lanka
[1,] 0.25 0.3098592 0.3015873
```

We can conclude that judges were likely to agree with raters' decisions in most cases. When there were disagreements, it happend more with the cases of applicants from China, Lebanon, and Sri Lanka. The percentage of agreements between judges and raters was 25% for China, which was the least among nations.

Conclusions

In this report, we have analyzed the sampled data of 384 leave to appeal for refugee cases.

After the simple observation of the data set, our team came up with three possible questions as follows: 1. Are decisions of judges similar to each other? 2. Does some judges have certain preference on specific nation? 3. Was judge's decision correlated with rater's decision on most cases? The questions were answered by checking the relationship between applicable variables. It was found that different judges make different decisions and judge's preference over specific nation do exist. Also, the judge's decision was not always correlated with rater's decision. Based on the analysis, we have determined that Canada's

controlling for anything?

I didn't see this.

Refugee-Determination system might not be a perfectly fair system since the decision is most likely to depend on the judge than other variables.

References

Fox, John. 2016. Applied Regression Analysis and Generalized Linear Models. 3rd ed. Sage Publications.

Ian Greene; Paul Shaffer, Leave to Appeal and Leave to Commerce Judicial Review in Canada's Refugee-Determination System: Is the Process Fair, 4 Int'l J. Refugee L. 71, 83 (1992).

- Very good. Good start experimenting with tools like capply.
- Could have used percentages in addition to raw frequencies.
- You didn't actually address the question of individual judges favouring selected countries.