

Activity 2

Yingyun Guo, Manuel Blain, Chenyu Li, Hanyi Zhang

2017-09-18

Activity 2

MATH 4330: Applied Categorical Data Analysis – Fall 2017

Yingyun Guo, Manuel Blain, Chenyu Li, Hanyi Zhang

September 18, 2017

Last updated on: septembre 18, 2017 at 20:39

Note: When running this file manually, make sure that the working directory is set to the directory in which this file has been saved. Use the RStudio menus: Session > Set Working Directory > To Source File Directory.

interesting how R localization uses language of the computer.

Introduction

The purpose of this study is to determine whether refugees and other persons who are applying for leave to appeal a decision of the Convention Refugee Determination Division of the Canadian Immigration and Refugee Board, a decision of the Appeal Division of the Immigration and Refugee Board, or to commence an action for judicial review have a relatively equal chance to convince the Federal Court of Appeals of the merits of their applications. This material is based on Fox (2016) pp. 1 – 11.

Data Biography

The data used in Fox (2016) can be downloaded from this directory:

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

Here's a link to the Fox data directory

Greene and Shaffer collected this data on Canada's refugee determination process during 1990-1992. In early April 1991, a systematic random sample of 611 cases - a sample more than sufficiently large to meet social science standards for a study of this kind - was selected from the approximately two thousand applications for leave to appeal filed in 1990. Every third file was pulled to generate the sample. This method of generating a representative sample is credible and widely used in the social sciences. So there is no bias on the data. Because the sample is large and representative, inferences can be made about the leave cases filed in 1990 in general with little fear that the sampling procedure may have accidentally resulted in the selection of a higher proportion of weak cases for some of the judges.

You can read the data set directly from the directory on the web: <chrome-extension://ikhdkkncnoglhjlkmcimlnlhkeamad/pdf-viewer/web/viewer.html?file=https%3A%2F%2Fyorkspace.library.yorku.ca%2Fxmlui%2Fbitstream%2Fhandle%2F10315%2F7864%2FGreene-LeavetoAppeal.pdf%3Fsequence%3D1%26isAllowed%3Dy>

```
Greene <- read.table(paste0(fox_data, 'Greene.txt'), header = TRUE)
head(Greene)
dim(Greene)
```

To get information on the file:

or "fair" given particular conditions

Very good but cite the source for this and include in bibliography.

```

library(knitr)
library(car)
library(spida2)
library(lattice)
library(latticeExtra)

| Loading required package: RColorBrewer

list.files(pattern='greene') %>% file.info # example using a magrittr pipe

| [1] size isdir mode mtime ctime atime uid gid uname grname
| <0 rows> (or 0-length row.names)

# Note that the '%>%' sequence of characters
# are produced in RStudio with Control-Shift-M
file.info(list.files(pattern = 'greene')) # same thing using conventional function composition

| [1] size isdir mode mtime ctime atime uid gid uname grname
| <0 rows> (or 0-length row.names)

# We also need to download the bibliography file
download.file('http://blackwell.math.yorku.ca/MATH4330/common/4330.bib', '4330.bib')

```

Data Directory

Open the data ‘directory’ or ‘codebook’ to see the description of the variables.

This data shows the 608 cases, filed in 1990, in which refugee claimants who were turned down by the Immigration and Refugee Board asked the Federal Court of Appeal for leave to appeal the board’s determination.

Variables:

- **judge:** Name of judge hearing case: Desjardins, Heald, Hugessen, Iacobucci, MacGuigan, Mahoney, Marceau, Pratte, Stone, Urie.
- **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.
- **rater:** Judgment of independent rater: no, case has no merit; yes, case has some merit (leave to appeal should be granted).
- **decision:** Judge’s decision: no, leave to appeal not granted; yes, leave to appeal granted.
- **language:** Language of case: English, French.
- **location:** Location of original refugee claim: Montreal, other, Toronto.
- **success:** Logit of success rate, for all cases from the applicant’s nation.

We can read the data file into R locally:

```
Greene <- read.table('Greene.txt')
```

Quick Look

```
summary(Greene)
```

	judge	nation	rater	decision	language
	MacGuigan :70	Lebanon :71	no :254	no :270	English:253
	Hugessen :62	China :68	yes:130	yes:114	French :131
	Desjardins:46	Sri.Lanka :63			
	Pratte :42	Bulgaria :36			
	Heald :36	Somalia :29			
	Stone :33	El.Salvador:26			
	(Other) :95	(Other) :91			
	location	success			

	Montreal:138	Min.	:-2.0907
	other : 55	1st Qu.:	-1.0986
	Toronto :191	Median :	-0.9946
		Mean :	-1.0204
		3rd Qu.:	-0.7538
		Max. :	0.4055

Interesting Questions

Good

This dataset raises a number of interesting questions about refugee appeals, of which we will investigate three:

1. Does there exist a bias in the judges' decision making process? Are the decisions of the judges consistent with each other and with the judgement of the independent rater?
2. Is there a bias against claimants based on their nation of origin?
3. Is there a regional or a language bias? These variables may act as a proxy for the judges political leanings.

Data Displays

Bias in Judges' Decision Making Process

```
Greene.judgeRaterDecision <- tab_(Greene, ~ judge + rater + decision, pct = c(1,2))
Greene.judgeRaterDecision[order(Greene.judgeRaterDecision[,3,2]),,] %>%
  barchart(ylab = 'name of judge hearing case', xlab = 'case has merit',
           horizontal = TRUE,
           xlim = c(0,100),
           auto.key=list(space='top',title='appeal granted'))
```

Ok but is it a bias if the proportion accepted differs because, e.g. human rights climate is different in different countries.

*Interesting!
But some comments apply.*

Saying "is the rate accepted" the same does not suggest a conclusion of unfairness.

I need to qualify this because you do conclusion on the independent rater evaluation - so you are looking at 'judge's behaviour relative to the rater's assessment which could perhaps be called "bias" if you accept that the raters are "unbiased".

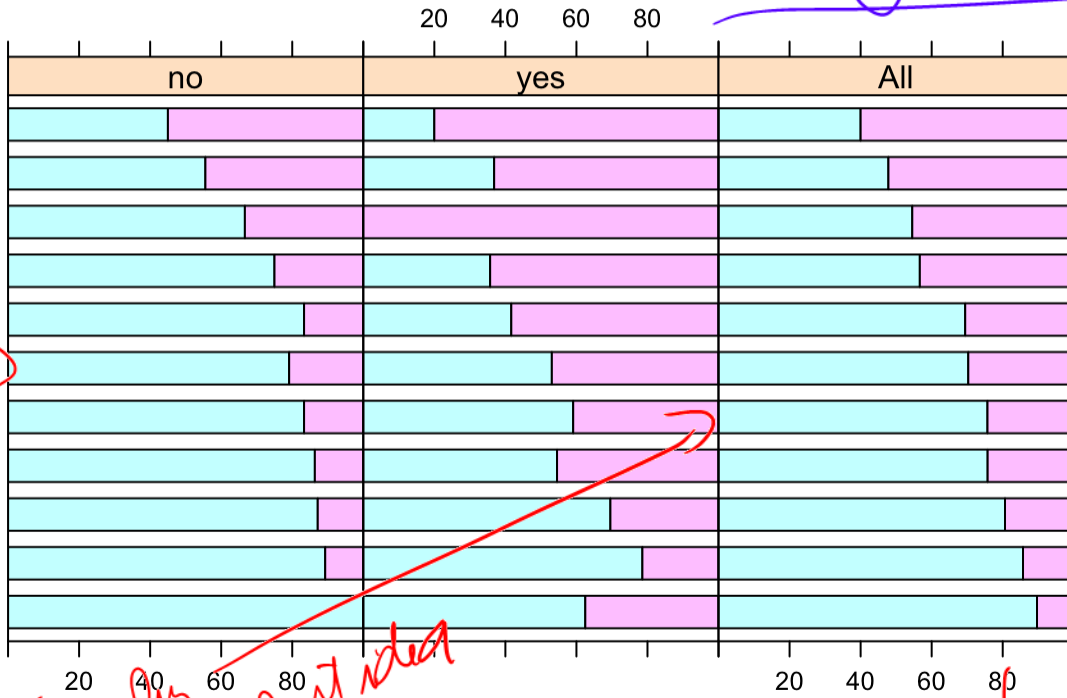
Possible "Stroop effect"

appeal granted



Very well done!

name of judge hearing case



ordering by "All" excellent idea

case has merit

might be hard to link this with this

Here we are comparing appeal success rate by judge against the judgement of the independent rater. We can use this view to verify if the judges tend to follow the judgement of the independent rater and are consistent with each other.

a detailed caption would be good.

Bias Based on Claimant Nation of Origin

```
Greene.byNationRater <- tab_(Greene, ~ nation + rater, pct = 1)
Greene.byNationDecision <- tab_(Greene, ~ nation + decision, pct = 1)
Greene.byNationComp <- merge(Greene.byNationRater[,2],Greene.byNationDecision[,2],by=0,all=TRUE)
colnames(Greene.byNationComp) <- c("nation","hasMerit","appealGranted")
Greene.byNationComp$difference <- Greene.byNationComp$appealGranted - Greene.byNationComp$hasMerit
kable(Greene.byNationComp[order(Greene.byNationComp$difference),], digits=1, row.names = FALSE, caption="Rate
```

but the overall design of the plot looks excellent to me.

Table 1: Rate of cases with merit against rate of successful appeals by nation of origin

nation	hasMerit	appealGranted	difference
Poland	36.4	0.0	-36.4
Nicaragua	66.7	33.3	-33.3
Nigeria	42.9	14.3	-28.6
Pakistan	75.0	50.0	-25.0
Ghana	33.3	11.1	-22.2
Argentina	60.0	40.0	-20.0
El.Salvador	53.8	34.6	-19.2
Iran	31.2	18.8	-12.5

ordering of my differences very good.

grammatical labels better

might be good to

have a column for n. of cases

is very different from 0% of 30

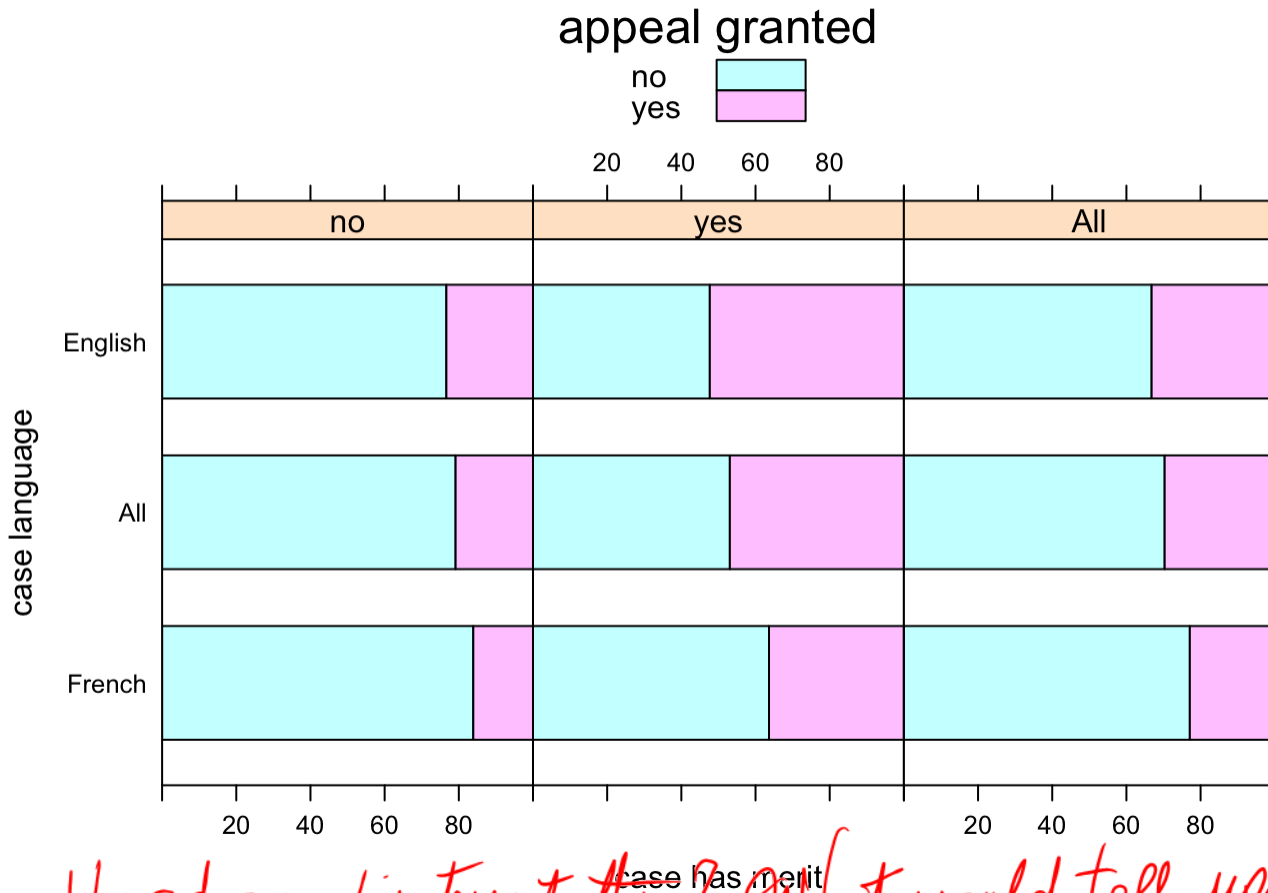
nation	hasMerit	appealGranted	difference
Somalia	37.9	27.6	-10.3
All	33.9	29.7	-4.2
Lebanon	28.2	25.4	-2.8
Bulgaria	13.9	11.1	-2.8
Sri.Lanka	42.9	41.3	-1.6
Fiji	0.0	0.0	0.0
Guatemala	20.0	20.0	0.0
China	22.1	26.5	4.4
Czechoslovakia	41.7	66.7	25.0
India	66.7	100.0	33.3

This table compares the overall rate of cases determined to have merit by the independent rater against the rate of successful appeal by claimant nation of origin. Nations with a significant difference in these rates may be subject to a bias.

by judges relative to raters

Regional and language bias

```
Greene.langRaterDecision <- tab_(Greene, ~ language + rater + decision, pct = c(1,2))
Greene.langRaterDecision[order(Greene.langRaterDecision[,3,2]),,] %>%
  barchart(ylab = 'case language', xlab = 'case has merit',
           horizontal = TRUE,
           xlim = c(0,100),
           auto.key=list(space='top',title='appeal granted'))
```



How do you interpret this? What would tell us that judges favour one language over the other relative to raters

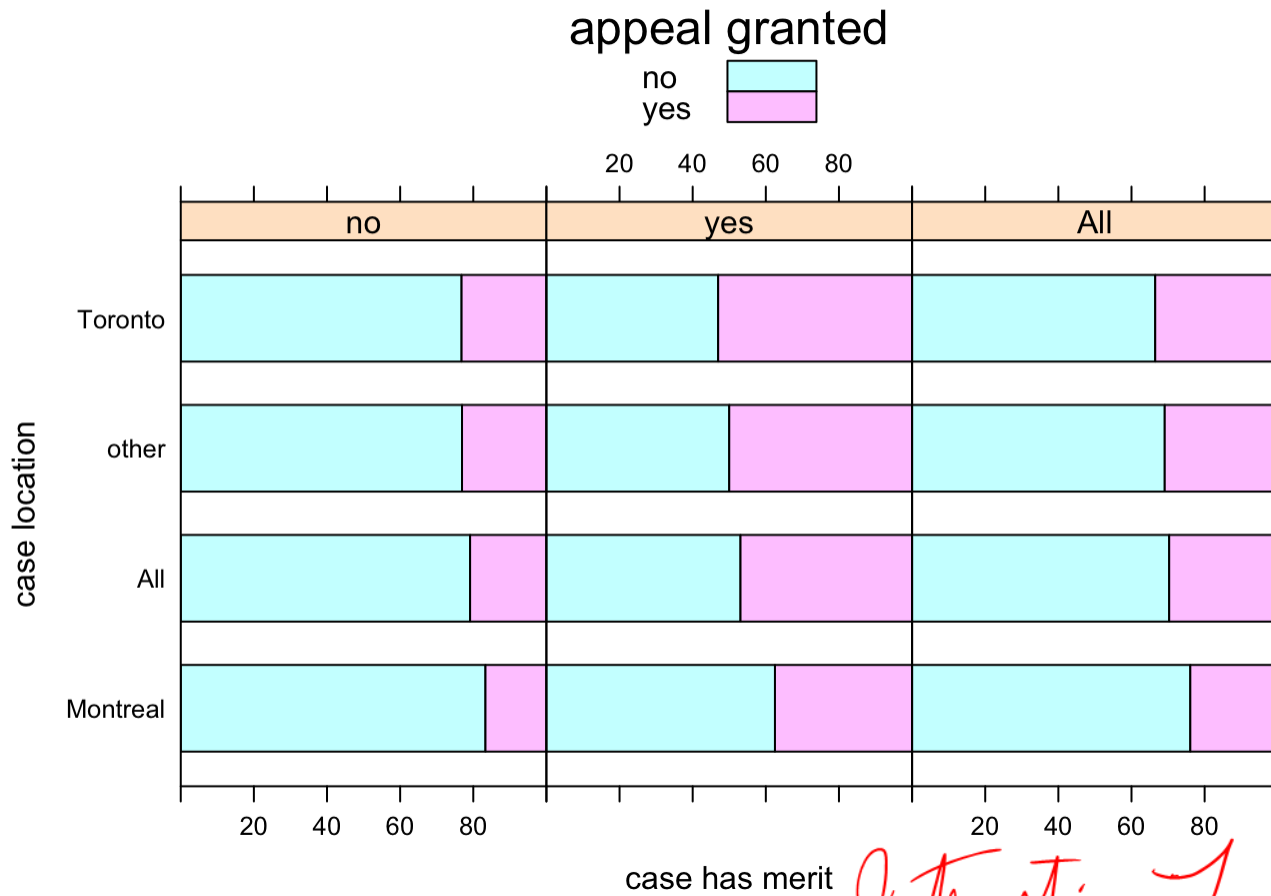
Here we are comparing appeal success rate by language against the judgement of the independent rater.

```
Greene.locRaterDecision <- tab_(Greene, ~ location + rater + decision, pct = c(1,2))
Greene.locRaterDecision[order(Greene.locRaterDecision[,3,2]),,] %>%
```

```

barchart(ylab = 'case location', xlab = 'case has merit',
horizontal = TRUE,
xlim = c(0,100),
auto.key=list(space='top',title='appeal granted'))

```



Interesting, location is probably strongly related to language.

Here we are comparing appeal success rate by location against the judgement of the independent rater

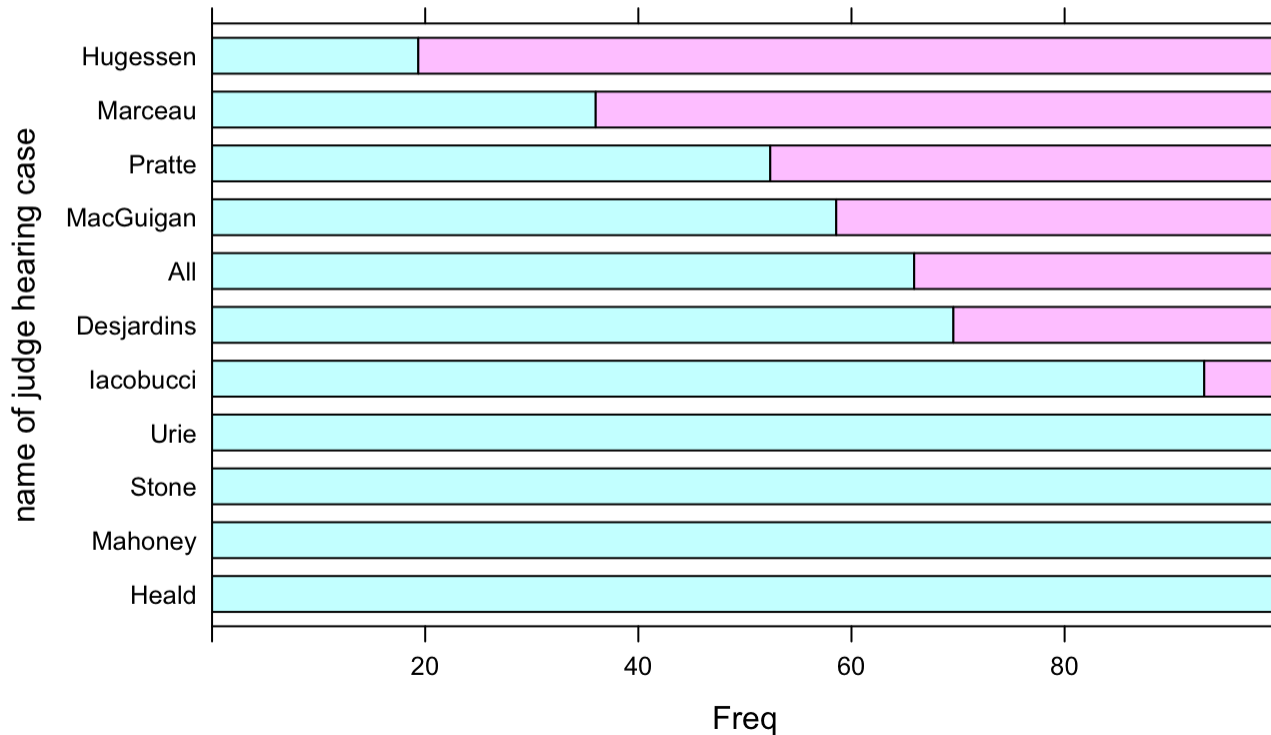
```

Greene.judgeLang <- tab_(Greene, ~ judge + language, pct = 1)
Greene.judgeLang[order(Greene.judgeLang[,2]),] %>%
  barchart(ylab = 'name of judge hearing case',
horizontal = TRUE,
xlim = c(0,100),
auto.key=list(space='top',title='cases by language'))

```

cases by language

English
French

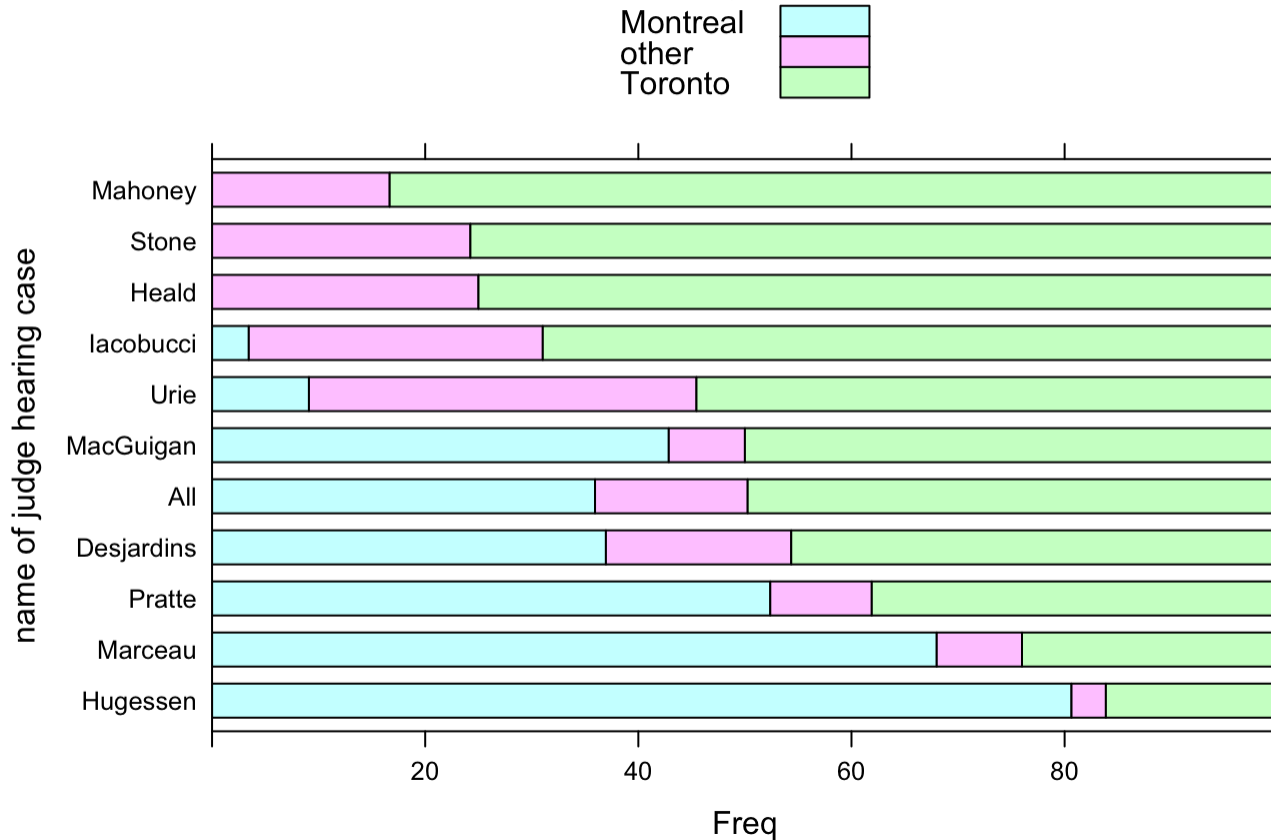


This shows the frequency of cases by language by judge. We can compare against the first table showing the decision by judge to see if there is a greater incidence of judge bias based on language.

```
Greene.judgeLoc <- tab_(Greene, ~ judge + location, pct = 1)
```

```
Greene.judgeLoc[order(Greene.judgeLoc[,3]),] %>%  
  barchart(ylab = 'name of judge hearing case',  
           horizontal = TRUE,  
           xlim = c(0,100),  
           auto.key=list(space='top',title='cases by location'))
```

cases by location



This shows the frequency of cases by location by judge. Similar to the above, we can try to identify if judges handling more cases from a specific location show evidence of bias.

Conclusions

To investigate the data, we identified three interesting questions and use data displays to show the relationships directly. The conclusion is inescapable that a bias exists between individual judges and the rate of success of claimant for leave to appeal. For the first question, we compare the appeal success rate by judge against the judgement of the independent rater. We set term 'All' in the name of judge hearing case as the standard, so that the bar chart shows that when a case does not have merit, for Marceau, Desjardins and Urie it has a larger chance to have leave granted, and for Iacobucci, cases having no merit will absolutely not have leave granted. When cases have merit, Urie grants all appeals and Marceau grants most; on the other hand, Pratte seems more strict while others are all normally distributed. The right part of the table shows us the overall result, regardless of whether the case has merit. We can see Marceau, Desjardins, Urie and Mahoney, all have a larger proportion of appeals granted while Pratte and Iacobucci are stricter.

For the second question, we discuss whether there is a bias against claimants based on their nation of origin. People from Pakistan have the highest percentage of cases with merit (75.0%), however the percentage of their appeals granted is only 50.0%, in contrast to India, who have appeals granted 100% of the time, even though only 66.7% of them have merit. Nicaragua also has 66.7% of cases having merit, but with a lower rate of appeals granted (33.3%). From the data, we can see that Poland, Nicaragua, Nigeria and Pakistan have the largest discrepancies at their disadvantage between case merit and appeal success rate, while India and Czechoslovakia have the largest discrepancies in their favor.

The last question is about whether regional or language bias exist. From the table it is hard to say if language or location bias exist, as only small differences appear when comparing against the 'All' category. There are also discrepancies between judges handling most cases in a particular language and their appeal success rate. For example, both Hugessen and Marceau handle a large proportion of French cases, but Hugessen has a low appeal success rate, while Marceau has the highest appeal success rate. A similar situation exists with the judges handling mostly English cases.

During our research, we encountered many limitations within the dataset. For example, we do not know the reason why an

appeal was not granted and have little information regarding specifics of the case. As such, more datasets would be useful to help validate our interpretations.

Questions for Further Study

We cannot answer the questions deeply, as there are more factors that influence the outcomes.

For the first question, it shows there is bias in the judges' decision making process. We can consider the judges' political leanings as another factor and use it to analyze the topic more deeply. We can also look at specific factors involved in the decision making process to determine if systemic issues exist (i.e. judges failing to grant appeal when cases have merit for specific situations).

For the second question, it shows there is serious bias based on claimant nation of origin. Data on the situation of the claimant's nation, such as political policy, the economy and the environment may provide more clarity on these discrepancies. In recent years, the stability of a region is one of the most significant factors considered when making a decision.

References

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

Include Greene & Shaffer.