

MATH 4330: Activity 2

Marion Fernandes, Salvador Freire, Monica Liu, Lois Li

2017-09-18

MATH 4330: Activity 2

Marion Fernandes, Salvador Freire, Monica Liu, Lois Li

September 18, 2017

- 1) Introduction
- 2) Data Biography
- 3) Data Directory
- 4) Interesting questions
- 5) Data displays using tables and barcharts
 - Package required
 - Download data
 - Summary
 - Tables
 - Bar Charts
- 6) Conclusions

```
knitr::opts_chunk$set(comment=NA) # suppresses '## ' in output
```

1) Introduction

In this report, our group was working on a data story “Greene”, involving 7 variables (judge, nation, rater, decision, language, location and success) and filling up in 1999 in which refugee claimants who were forced to leave by the Canadian Immigration and Refugee Board, asked the Federal Court of Appeal for leave to appeal the negative ruling of the Board. To be more specific, we wanted to find the correlation between the nation and judge with rater, decision and success. According to our analysis, we found a similar pattern between the rater, decision and success when conditioned on the nation, continent and judge.

What does between A and B with C, D and E mean?

Correlation is normally used in reference to linear relationship between numeric variables. Use "relationship among"

2) Data Biography

We downloaded the “greene.txt” and started our analysis. We needed to know more about the detailed information we were looking at so that we could better understand the patterns we were finding. In this case, the first thing we needed to do was find out where the information was coming from. The data reflected about people with judge’s name and their status (decision, rater, and success...) from website “<http://socserv.socsci.mcmaster.ca/jfox/Books/Applied-Regression-3E/datasets/Greene.txt>”. Next, we would like to figure out who collected these data. Unfortunately, we have no idea about that part. From the data provided, it is clear that there is a strong relationship between the rater and decision based on individual nations and judges. We can see that the rater is fairly accurate in determining the decision outcome based on judges preciding over the case, or based on the nation except for a few cases. These discrepancies are observed to be caused by a small number of appeal cases in those nations. For example, only a few people on the file came from Iran and Sri.Lanka.

You need to do some detective work with, for example, Google Scholar.

3) Data Directory

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.

4) Interesting questions

- a. How are the values obtained for success?
- b. Where does the independent rater come from?
- c. Can we state that the judge's decision depends on the nation that the refugees come from according to the data?

?
- good

Do you mean that different judge's have different patterns with respect to the nation of origin? Or just that different origins result in different probabilities of success on appeal?

5) Data displays using tables and barcharts

Package required

```
library(car)
library(spida2)
library(lattice)
library(latticeExtra)
```

Loading required package: RColorBrewer

```
library(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 data

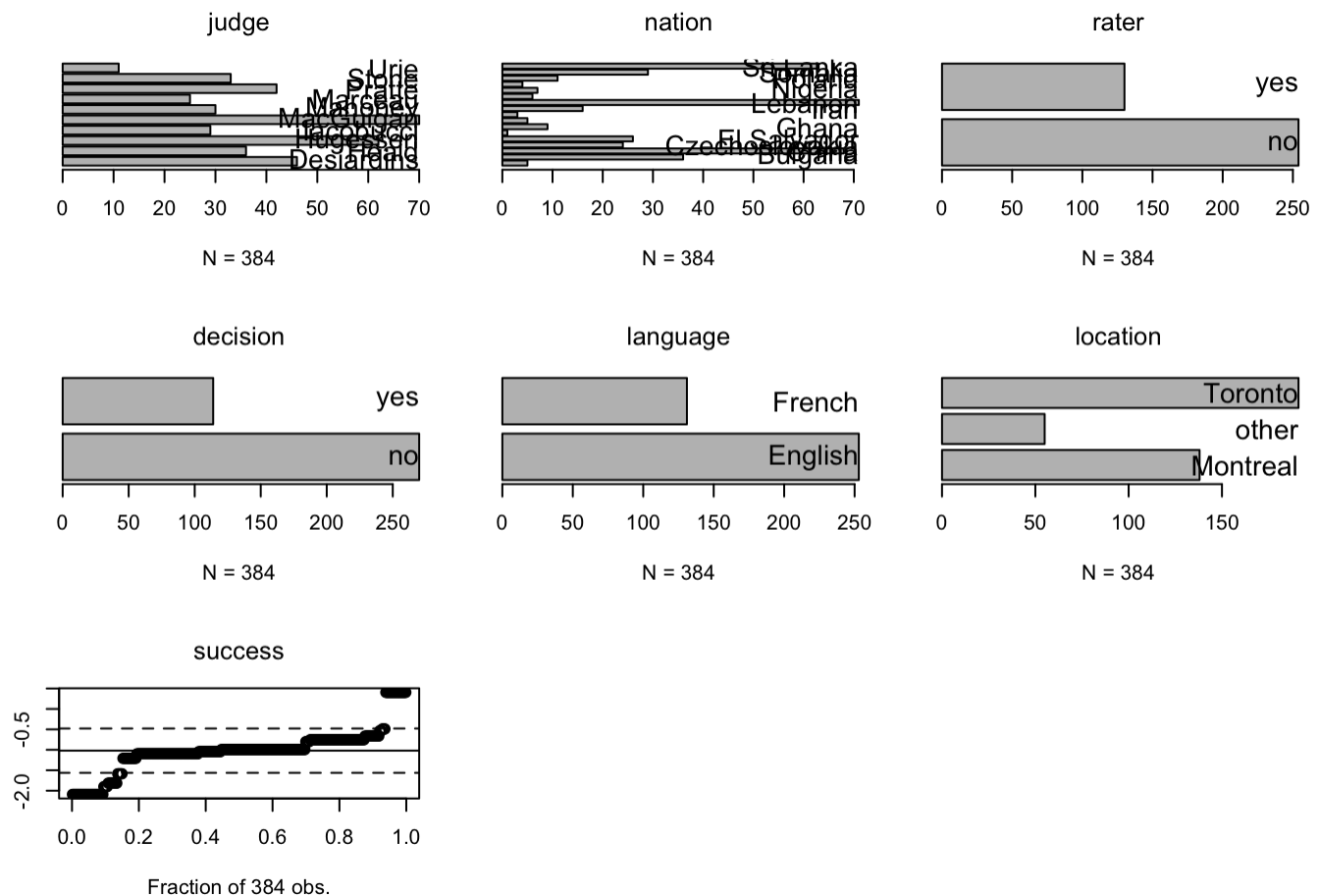
```
download.file("http://socserv.socsci.mcmaster.ca/jfox/Books/Applied-Regression-3E/datasets/Greene.txt",
             "Greene.txt")
greene <- read.table("Greene.txt", header = T)
```

Summary

```
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			

```
xqplot(greene) #raw values
```



Adding a column 'probability' to the dataset using success values

```
probability <- 1/(1+exp(-greene$success))
Greene_new <- cbind(greene, probability)
summary(probability)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.1100 0.2500 0.2700 0.2762 0.3200 0.6000
```

Tables

```
Greene_new %>%
  tab_(~ rater + decision + language, pct = c(1,2)) %>%
  round(1) %>%
  ftable
```

		language	
		English	French
rater	decision		
no	no	63.7	36.3
	yes	73.6	26.4
yes	no	59.4	40.6
	yes	73.8	26.2

```
Greene_new %>%
  tab_(~ rater + decision + location, pct = c(1,2)) %>%
  round(1) %>%
  ftable
```

		location		
		Montreal	other	Toronto
rater	decision			
no	no	37.3	14.9	47.8
	yes	28.3	17.0	54.7
yes	no	43.5	11.6	44.9
	yes	29.5	13.1	57.4

```
tab(Greene_new, ~ rater + decision + nation, pct = 0)
```

```
, , nation = Argentina
```

		decision		Total
rater		no	yes	
no		0.5208333	0.0000000	0.5208333
yes		0.2604167	0.5208333	0.7812500
	Total	0.7812500	0.5208333	1.3020833

```
, , nation = Bulgaria
```

		decision		Total
rater		no	yes	
no		7.5520833	0.5208333	8.0729167
yes		0.7812500	0.5208333	1.3020833
	Total	8.3333333	1.0416667	9.3750000

```
, , nation = China
```

		decision		Total
rater		no	yes	
no		11.1979167	2.6041667	13.8020833
yes		1.8229167	2.0833333	3.9062500
	Total	13.0208333	4.6875000	17.7083333

```
, , nation = Czechoslovakia
```

```
decision
```

this works but more common is

greene \$ probability <- 1/(1+exp(-greene\$success)) i.e. just one command

what does this table say? Is it of interest?

How would you interpret these?

What should I be looking for?

rater	no	yes	Total
no	1.8229167	1.8229167	3.6458333
yes	0.2604167	2.3437500	2.6041667
Total	2.0833333	4.1666667	6.2500000

, , nation = El.Salvador

decision			
rater	no	yes	Total
no	2.3437500	0.7812500	3.1250000
yes	2.0833333	1.5625000	3.6458333
Total	4.4270833	2.3437500	6.7708333

, , nation = Fiji

decision			
rater	no	yes	Total
no	0.2604167	0.0000000	0.2604167
yes	0.0000000	0.0000000	0.0000000
Total	0.2604167	0.0000000	0.2604167

, , nation = Ghana

decision			
rater	no	yes	Total
no	1.3020833	0.2604167	1.5625000
yes	0.7812500	0.0000000	0.7812500
Total	2.0833333	0.2604167	2.3437500

, , nation = Guatemala

decision			
rater	no	yes	Total
no	0.7812500	0.2604167	1.0416667
yes	0.2604167	0.0000000	0.2604167
Total	1.0416667	0.2604167	1.3020833

, , nation = India

decision			
rater	no	yes	Total
no	0.0000000	0.2604167	0.2604167
yes	0.0000000	0.5208333	0.5208333
Total	0.0000000	0.7812500	0.7812500

, , nation = Iran

decision			
rater	no	yes	Total
no	2.3437500	0.5208333	2.8645833
yes	1.0416667	0.2604167	1.3020833
Total	3.3854167	0.7812500	4.1666667

, , nation = Lebanon

decision			
rater	no	yes	Total
no	10.6770833	2.6041667	13.2812500
yes	3.1250000	2.0833333	5.2083333
Total	13.8020833	4.6875000	18.4895833

, , nation = Nicaragua

decision			
rater	no	yes	Total
no	0.2604167	0.2604167	0.5208333
yes	0.7812500	0.2604167	1.0416667
Total	1.0416667	0.5208333	1.5625000

, , nation = Nigeria

decision			
rater	no	yes	Total
no	1.0416667	0.0000000	1.0416667
yes	0.5208333	0.2604167	0.7812500
Total	1.5625000	0.2604167	1.8229167

, , nation = Pakistan

decision			
rater	no	yes	Total
no	0.0000000	0.2604167	0.2604167
yes	0.5208333	0.2604167	0.7812500
Total	0.5208333	0.5208333	1.0416667

, , nation = Poland

decision			
rater	no	yes	Total
no	1.8229167	0.0000000	1.8229167
yes	1.0416667	0.0000000	1.0416667
Total	2.8645833	0.0000000	2.8645833

, , nation = Somalia

decision			
rater	no	yes	Total
no	3.3854167	1.3020833	4.6875000
yes	2.0833333	0.7812500	2.8645833
Total	5.4687500	2.0833333	7.5520833

, , nation = Sri.Lanka

decision			
rater	no	yes	Total
no	7.0312500	2.3437500	9.3750000
yes	2.6041667	4.4270833	7.0312500
Total	9.6354167	6.7708333	16.4062500

, , nation = Total

decision			
rater	no	yes	Total
no	52.3437500	13.8020833	66.1458333
yes	17.9687500	15.8854167	33.8541667
Total	70.3125000	29.6875000	100.0000000

tab(Greene_new, ~ rater + decision + judge, pct = 0)

, , judge = Desjardins

decision			
rater	no	yes	Total
no	3.9062500	3.1250000	7.0312500
yes	1.8229167	3.1250000	4.9479167
Total	5.7291667	6.2500000	11.9791667

, , judge = Heald

decision			
rater	no	yes	Total
no	5.2083333	1.0416667	6.2500000
yes	1.3020833	1.8229167	3.1250000
Total	6.5104167	2.8645833	9.3750000

, , judge = Hugessen

decision			
rater	no	yes	Total
no	8.8541667	1.3020833	10.1562500
yes	4.1666667	1.8229167	5.9895833
Total	13.0208333	3.1250000	16.1458333

, , judge = Iacobucci

decision			
rater	no	yes	Total
no	5.4687500	0.0000000	5.4687500
yes	1.3020833	0.7812500	2.0833333
Total	6.7708333	0.7812500	7.5520833

, , judge = MacGuigan

decision			
rater	no	yes	Total
no	10.4166667	2.0833333	12.5000000
yes	3.3854167	2.3437500	5.7291667
Total	13.8020833	4.4270833	18.2291667

, , judge = Mahoney

decision			
rater	no	yes	Total
no	3.1250000	1.0416667	4.1666667
yes	1.3020833	2.3437500	3.6458333
Total	4.4270833	3.3854167	7.8125000

, , judge = Marceau

decision			
rater	no	yes	Total
no	2.3437500	2.8645833	5.2083333
yes	0.2604167	1.0416667	1.3020833
Total	2.6041667	3.9062500	6.5104167

, , judge = Pratte

decision			
rater	no	yes	Total
no	6.5104167	0.7812500	7.2916667
yes	2.8645833	0.7812500	3.6458333

Total 9.3750000 1.5625000 10.9375000

, , judge = Stone

decision			
rater	no	yes	Total
no	4.9479167	0.7812500	5.7291667
yes	1.5625000	1.3020833	2.8645833
Total	6.5104167	2.0833333	8.5937500

, , judge = Urie

decision			
rater	no	yes	Total
no	1.5625000	0.7812500	2.3437500
yes	0.0000000	0.5208333	0.5208333
Total	1.5625000	1.3020833	2.8645833

, , judge = Total

decision			
rater	no	yes	Total
no	52.3437500	13.8020833	66.1458333
yes	17.9687500	15.8854167	33.8541667
Total	70.3125000	29.6875000	100.0000000

Table to look at % of yes/no from decisions and raters for each nation

```
tab__(Greene_new, ~ decision + nation, pct = 2) %>%
  round(1)
```

nation								
decision	Argentina	Bulgaria	China	Czechoslovakia	El.Salvador	Fiji	Ghana	
no	60.0	88.9	73.5		33.3	65.4	100.0	88.9
yes	40.0	11.1	26.5		66.7	34.6	0.0	11.1

nation								
decision	Guatemala	India	Iran	Lebanon	Nicaragua	Nigeria	Pakistan	Poland
no	80.0	0.0	81.2	74.6	66.7	85.7	50.0	100.0
yes	20.0	100.0	18.8	25.4	33.3	14.3	50.0	0.0

nation		
decision	Somalia	Sri.Lanka
no	72.4	58.7
yes	27.6	41.3

```
tab__(Greene_new, ~ rater + nation, pct = 2) %>%
  round(1)
```

nation								
rater	Argentina	Bulgaria	China	Czechoslovakia	El.Salvador	Fiji	Ghana	
no	40.0	86.1	77.9		58.3	46.2	100.0	66.7
yes	60.0	13.9	22.1		41.7	53.8	0.0	33.3

nation								
rater	Guatemala	India	Iran	Lebanon	Nicaragua	Nigeria	Pakistan	Poland
no	80.0	33.3	68.8	71.8	33.3	57.1	25.0	63.6
yes	20.0	66.7	31.2	28.2	66.7	42.9	75.0	36.4

nation		
rater	Somalia	Sri.Lanka
no	62.1	57.1
yes	37.9	42.9

unique(probability) * 100 # Somalia and China have same probability of success

[1] 25.00004 32.00004 25.99997 60.00012 23.00002 27.00005 14.00000


```
[8] 16.99996 34.00009 12.99999 11.00001 30.99999 36.99993 37.99996
```

Creating a new table to compare decision and probability for each nation

```
NewTable <- data.frame(nation = c("Argentina","Bulgaria","China","Czechoslovakia","El.Salvador","Fiji","Ghana"  
                                decision = c("40.0","11.1","26.5","66.7","34.6","0.0","11.1","20.0","100.0","18.8","25.
```

```
NewTable$probability = Greene_new$probability[match(NewTable$nation, Greene_new$nation)]*100
```

Table to look at % of yes/no from decisions and raters for each judge

```
tab__(Greene_new, ~ decision + judge, pct = 2) %>%  
  round(1)
```

```
      judge  
decision Desjardins Heald Hugessen Iacobucci MacGuigan Mahoney Marceau  
no        47.8  69.4    80.6      89.7      75.7    56.7    40.0  
yes       52.2  30.6    19.4     10.3     24.3    43.3    60.0
```

```
      judge  
decision Pratte Stone Urie  
no        85.7  75.8  54.5  
yes       14.3  24.2  45.5
```

```
tab__(Greene_new, ~ rater + judge, pct = 2) %>%  
  round(1)
```

```
      judge  
rater Desjardins Heald Hugessen Iacobucci MacGuigan Mahoney Marceau Pratte  
no        58.7  66.7    62.9     72.4     68.6    53.3    80.0    66.7  
yes       41.3  33.3    37.1     27.6     31.4    46.7    20.0    33.3
```

```
      judge  
rater Stone Urie  
no        66.7  81.8  
yes       33.3  18.2
```

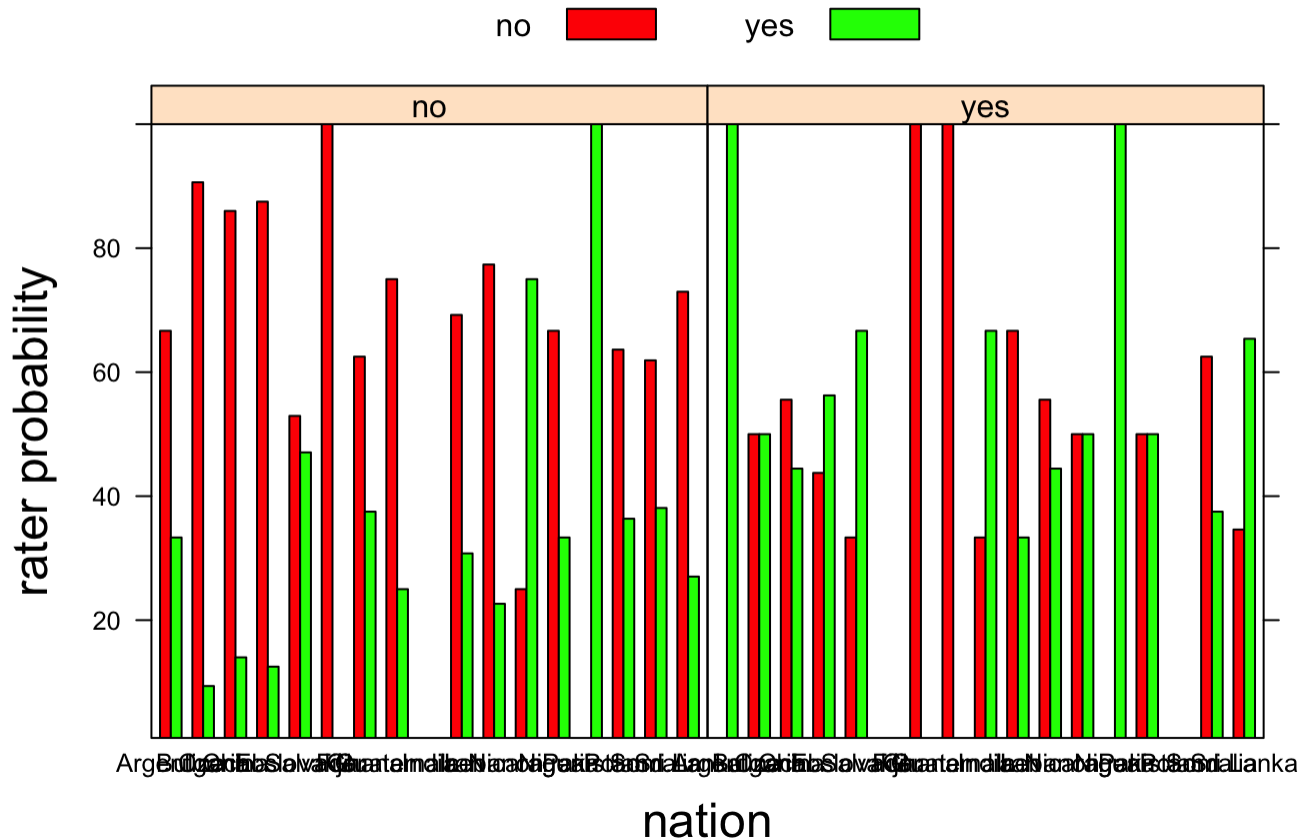
Bar Charts

```
colors = c("red","green")
```

Barchart showing rater outcome vs. decision for each nation

```
tab__(Greene_new, ~ nation + decision + rater, pct = c(1,2)) %>%  
  barchart(ylab = list(label = "rater probability", cex = 1.5),  
           ylim = c(1,100),  
           xlab = list(label = "nation", cex = 1.5),  
           horizontal = FALSE,  
           stack = FALSE,  
           auto.key = list(space = list(space = "top"), title = "Rater vs. Decision for each nation", column =  
                             par.settings = list(superpose.polygon = list(col = colors)))
```

Rater vs. Decision for each nation



Adding column for continents to include in barchart

```
for(i in 1:nrow(greene)){
  if(greene$nation[i] == "Bulgaria"){greene$continent[i] <- "Europe"}
  else if (greene$nation[i] == "Argentina"){greene$continent[i] <- "Asia"}
  else if (greene$nation[i] == "China"){greene$continent[i] <- "Asia"}
  else if (greene$nation[i] == "Czechoslovakia"){greene$continent[i] <- "Europe"}
  else if (greene$nation[i] == "El.Salvador"){greene$continent[i] <- "South America"}
  else if (greene$nation[i] == "Fiji"){greene$continent[i] <- "Oceania"}
  else if (greene$nation[i] == "Ghana"){greene$continent[i] <- "Africa"}
  else if (greene$nation[i] == "Guatemala"){greene$continent[i] <- "South America"}
  else if (greene$nation[i] == "India"){greene$continent[i] <- "Asia"}
  else if (greene$nation[i] == "Iran"){greene$continent[i] <- "Asia"}
  else if (greene$nation[i] == "Lebanon"){greene$continent[i] <- "Asia"}
  else if (greene$nation[i] == "Nicaragua"){greene$continent[i] <- "South America"}
  else if (greene$nation[i] == "Nigeria"){greene$continent[i] <- "Africa"}
  else if (greene$nation[i] == "Pakistan"){greene$continent[i] <- "Asia"}
  else if (greene$nation[i] == "Poland"){greene$continent[i] <- "Europe"}
  else if (greene$nation[i] == "Somalia"){greene$continent[i] <- "Africa"}
  else if (greene$nation[i] == "Sri.Lanka"){greene$continent[i] <- "Asia"}
  else {greene$continent[i] <- "N/A"}
}
Greene_new$continent <- factor(greene$continent)
```

Barchart showing rater outcome vs. decision for each continent

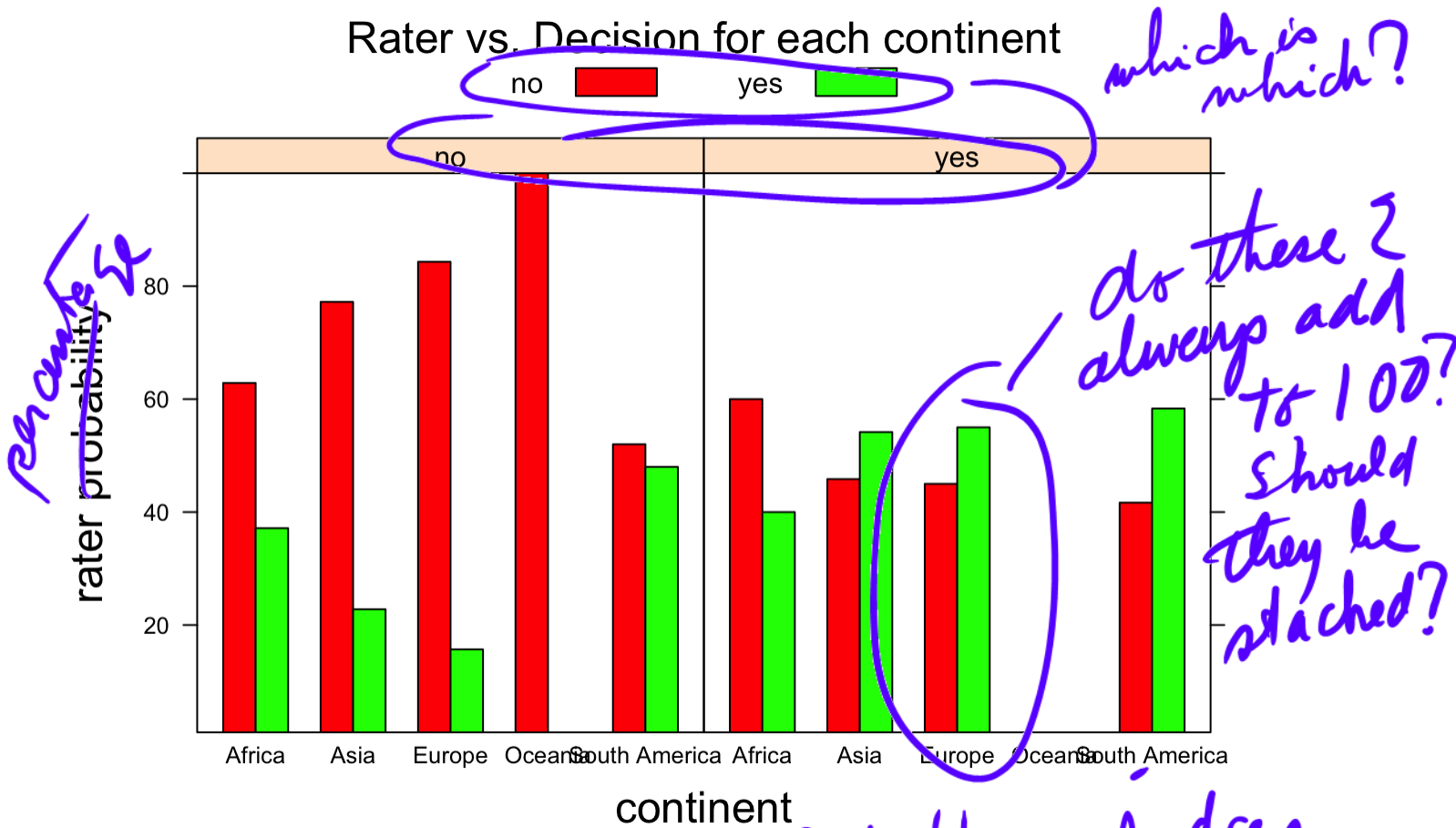
```
tab__(Greene_new, ~ continent + decision + rater, pct = c(1,2)) %>%
  barchart(ylab = list(label = "rater probability", cex = 1.5),
           ylim = c(1,100),
           xlab = list(label = "continent", cex = 1.5),
           horizontal = FALSE,
```

```

stack = FALSE,
auto.key = list(space = list(space = "top"), title = "Rater vs. Decision for each continent", column = 1,
par.settings = list(superpose.polygon = list(col = colors)))

```

Rater vs. Decision for each continent



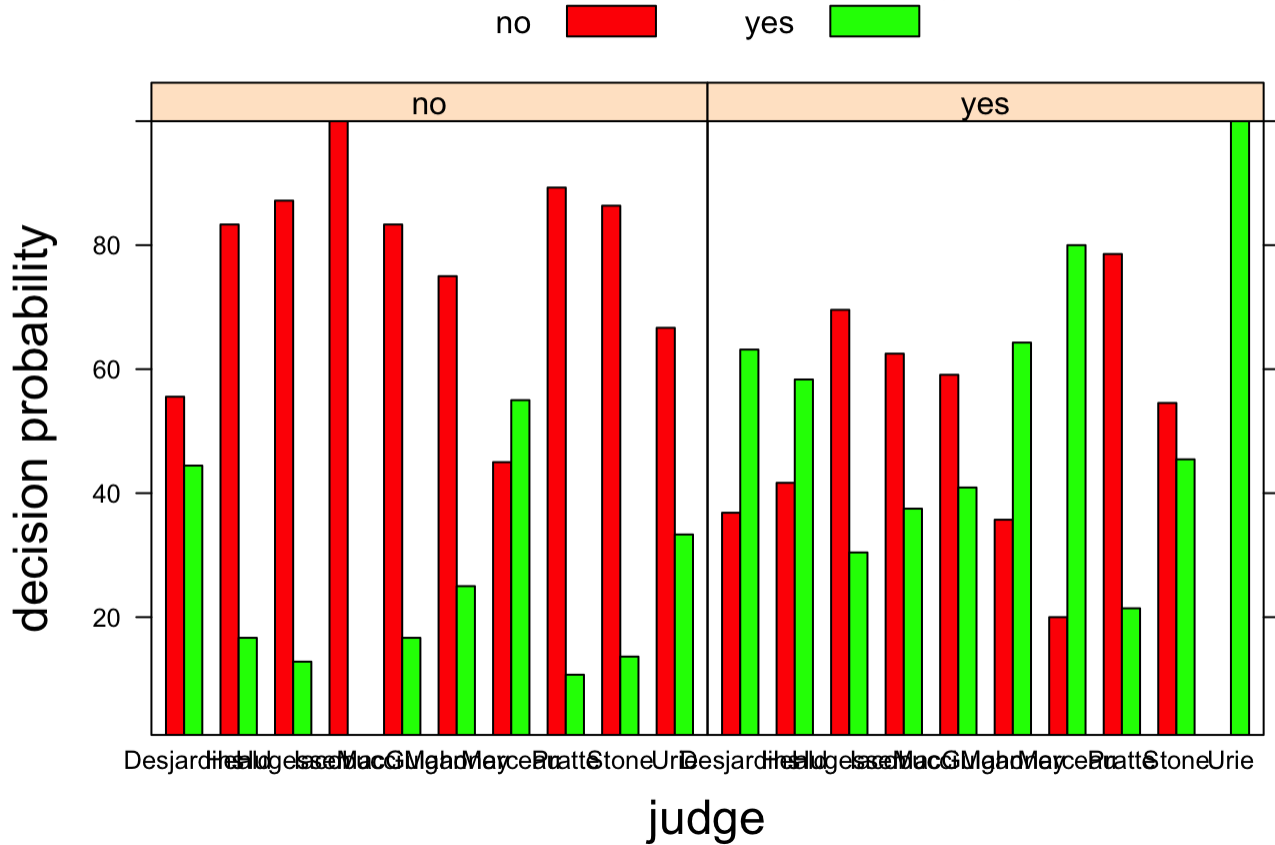
Barchart showing decision outcome vs. rater for each judge

```

tab__(Greene_new, ~ judge + rater + decision, pct = c(1,2)) %>%
  bargchart(ylab = list(label = "decision probability", cex = 1.5),
            ylim = c(1,100),
            xlab = list(label = "judge", cex = 1.5),
            horizontal = FALSE,
            stack = FALSE,
            auto.key = list(space = list(space = "top"), title = "Decision vs. Rater for each judge", column = 1,
            par.settings = list(superpose.polygon = list(col = colors)))

```

Decision vs. Rater for each judge



Axis variables: y-axis: rater probability, x-axis: nation & continent

Panel variables: decision

Grouping variables: Africa, Asia, Europe, North America, Oceania, South America

6) Conclusions

After analysing the data, we would like to do more. Unfortunately, the data set has some limitations, although we have tried to manipulate the data by adding new columns (eg continent) to study some relation between the explanatory variables. At this time, it is hard to get significant conclusions, but it is possible to suspect which direction we can follow. For example we can see strong patterns when the rater is no, the decision is normally no too. The nation/continent seems to not be correlated with the decision. Intuitively, the language and location don't have any relation with the decision.

patterns of what kind.

- You have produced some nice bar charts. They would have benefitted from clear captions to explain what they mean. Always look at your work from the perspective of a reader and ask yourself "if

If were reading this, what would I understand from it. The numbers are of very little interest unless you explain what they mean.

The tables at the beginning aren't focused on a question I can understand, in particular your question about Country of origin.

Imagine that you are reading a newspaper article on this topic, what table would you expect to see.