Problem Set 2, Part I

Alin Ierima

Sources of Empathy in the Circuit Courts

In this exercise, we will analyze the relationship between the gender composition among a judge's children and voting behavior among circuit court judges. In a recent paper, Adam N. Glynn and Maya Sen argue that having a female child causes circuit court judges to make more pro-feminist decisions. The paper can be found at:

Glynn, Adam N., and Maya Sen. (2015). "Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?." *American Journal of Political Science* Vol. 59, No. 1, pp. 37–54.

The dataset `judges.csv` contains the following variables about individual judges:

Name	Description
`name`	The judge's name
`num_kids`	The number of children each judge has.
`circuit`	Which federal circuit the judge serves in.
`girls`	The number of female children the judge has.
`progressive_vote`	The proportion of the judge's votes on women's issues which were decided in a pro-feminist direction.
`race`	The judge's race (`1` = white, `2` = African-American, `3` = Hispanic, `4` = Asian-American).
`religion`	The judge's religion (`1` = Unitarian, `2` = Episcopalian, `3` = Baptist, `4` = Catholic, `5` = Jewish, `7` = Presbyterian, `8` = Protestant, `9` = Congregationalist, `10` = Methodist, `11` = Church of Christ, `16` = Baha'i, `17` = Mormon, `21` = Anglican, `24` = Lutheran, `99` = unknown).
`republican`	Takes a value of `1` if the judge was appointed by a Republican president, `0` otherwise. Used as a proxy for the judge's party.
`sons`	The number of male children the judge has.
`woman`	Takes a value of `1` if the judge is a woman, `0` otherwise.
`X`	Indicator for the observation number.
`yearb`	The year the judge was born.

Question 1 (15 points)

- Load the `judges.csv` file into a data frame called `judges`.
- Create a cross-tab (of proportions, not counts) of judge gender on the rows and whether the appointing president was Republican on the columns. Save this table with the name `gender_rep_table`. (*6 points*)
- Use `knitr::kable()` to create a nicely formatted version of this table. (*6 points*)

In your write-up, answer the following questions in the text:

- How many judges are in this data set? (*1 point*)
- What proportion of the judges are men? (*1 point*)
- Is the party composition different for male and female judges? (*1 point*)

NOTE: to change the row and column labels for the output table using `knitr::kable()`, you can change the row and column names of the table you pass to it. For example, if I am passing a table called `my_table`, then I can use the following:

```
rownames(my_table) <- c("Row 1 Label", "Row 2 Label")
colnames(my_table) <- c("Column 1 Label", "Column 2 Label")
```

Answer 1

```
judges <- read.csv("data/judges.csv")
```

Contingency table

```
gender_rep_table = addmargins(prop.table((table(judges$woman, judges$republican))))
rownames(gender_rep_table) = c("Male", "Female", "Column totals")
colnames(gender_rep_table) = c("Democrats", "Republicans", "Row totals")
gender_rep_table
```

```
##
##                  Democrats Republicans Row totals
##   Male          0.33928571  0.49107143 0.83035714
##   Female        0.12053571  0.04910714 0.16964286
##   Column totals 0.45982143  0.54017857 1.00000000
```

Formatted contingency table

```
knitr::kable(gender_rep_table)
```

	Democrats	Republicans	Row totals
Male	0.3392857	0.4910714	0.8303571
Female	0.1205357	0.0491071	0.1696429
Column totals	0.4598214	0.5401786	1.0000000

Insert your answer here:1. According to nrow(judges), there are 224 judges. Of those, 83% are male and ~17% are female. It seems females are three times more pro-democrat than pro-republican, whereas there are more republican males than democrat males.

Question 2 (20 points)

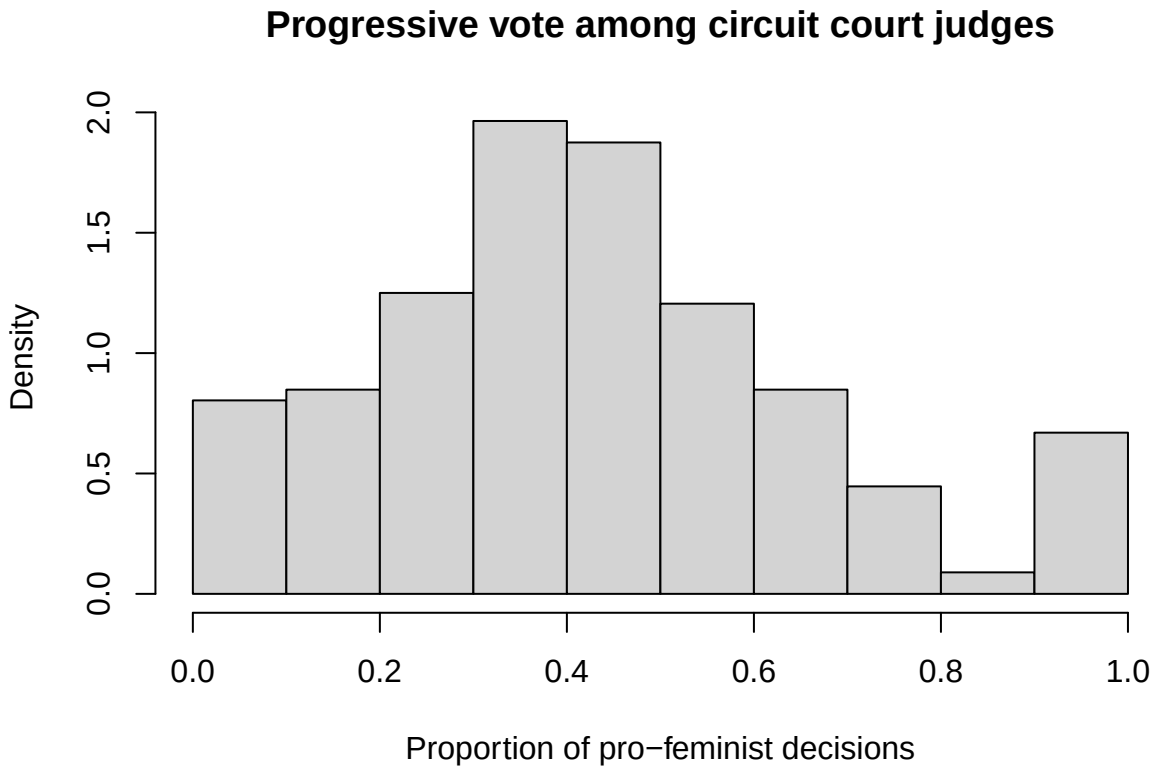
Our outcome in this exercise will be the *proportion of feminist rulings on issues related to gender* - the `progressive_vote` variable.

- Create a nicely formatted histogram of the `progressive_vote` variable. (Remember to set the argument `freq = FALSE`.) Nicely formatted means the histogram has informative title and axis labels and does not contain random R syntax. (*5 points*)
- Provide a written summary of this graph. Roughly speaking, discuss where is the region of highest density of this variable. (*3 points*)

- Notice that the minimum of `progressive_vote` is exactly 0 and the maximum is exactly 1. How can we interpret these values - what does 0 and 1 mean? (*3 points*)
- Investigate how many judges have `progressive_vote` equal to 1 and how many equal to 0. Report these values in the text. What is the party composition of these judges (party composition of those judges who have `progressive_vote` equal to 0 and of those who have `progresive_vote` of 1)? (*6 points*)
- Use `knitr::kable()`to produce nicely formatted tables with this information. Use more informative column labels than "Progressive vote = 1" and "Progressive vote = 0") (*3 points*)

Answer 2

```
hist(judges$progressive_vote, freq = F, main = "Progressive vote among circuit court judges",xlab =  "P
```



Insert your answer here: According to the histogram, the density is the biggest between 0.3 and 0.5, indicating that judges are on average more conservative in their decisions. However the last bar (from 0.9 to 1.0) indicates that there is a sizable portion of very progressive judges, which goes against the trend of diminishing density as one gets further away from 0.5. By the way, a progressive_vote of 0 means that all the judging decisions taken by a judge were conservative, whereas 1 means 100% of the decisions were pro-feminist.

Number of judges with `progressive_vote` == 1

```
progressive_judges = length(judges$progressive_vote[judges$progressive_vote == 1])
progressive_judges
```

```
## [1] 14
```

Number of judges with `progressive_vote == 0`

```
conservative_judges = length(judges$progressive_vote[judges$progressive_vote == 0])
conservative_judges
```

```
## [1] 17
```

```
extremes_jud_subset = judges[judges$progressive_vote == 0 | judges$progressive_vote == 1, ]

table_ideology_votes = addmargins(table(extremes_jud_subset$republican,extremes_jud_subset$progressive_v

colnames(table_ideology_votes) = c("Judges with exclusively conservative decisions", "Judges with exclus
rownames(table_ideology_votes) = c("Democrats", "Republicans", "Sum")

knitr::kable(table_ideology_votes)
```

	Judges with exclusively conservative decisions	Judges with exclusively pro-feminist decisions	Row totals
Democrats	8	8	16
Republicans	9	6	15
Sum	17	14	31

Insert your answer here: (3 points) I expected the party composition of judges voting conservatively to be.. conservative, however this is not the case, as there are 8 democrats voting conservatively and 8 republicans voting conservatively; conversely, there are 6 republicans voting progressively (again, surprising), and 8 democrats voting pro-feminist.

Question 3 (20 points)

Next, we consider differences between some groups.

- Create a new factor variable called `judges$gender_party` that takes the following 4 categories: (3 points)
  - `"Dem.Woman"` for women appointed by Democratic presidents
  - `"Rep.Woman"` for women appointed by Republican presidents
  - `"Dem.Man"` for men appointed by Democratic presidents
  - `"Rep.Man"` for men appointed by Republican presidents
- Get the the number of observations in each of the 4 categories in the variable `gender_party`. (*1 point*)
- Use `tapply` to calculate the mean of `progressive_vote` in each of these groups and save this vector as `gender_party_means`. Plot these means using a **barplot**. (*5 points*)
- In addition to the barplot, do the following. For each of the four groups defined by gender and partisanship (Republican men, Republican women, Democratic men and Democratic women), create a **boxplot** using a single command that illustrates the differences in `progressive_vote` (Hint: section 3.3.3 in QSS can be helpful here). (*5 points*)

*Note: If you want to change the location of the boxplot labels on the x axis you can use the `par(mgp = c(3,1.5,0))` function before running the `boxplot()` function (You can, but do *not have* to do that). The `mgp` argument sets the axis label locations. The first argument specifies the location of the labels (i.e. xlab and ylab in plot), the second argument the tick-mark labels, and the third argument the tick marks. The default is c(3, 1, 0).

- Briefly interpret the results from the barplot and boxplot by answering the following questions.
  - Do any of the results surprise you? (*1 point*)
  - Does it appear that partisanship, gender, or both contribute to progressive voting patterns? (*2 points*)
  - Should we interpret any of these effects causally? Why or why not? (*3 points*)

Answer 3

```
judges$gender_party = ifelse(test = judges$republican == 0 & judges$woman ==1, "Dem.Woman",
        ifelse(judges$republican == 1 & judges$woman == 1, "Rep.Woman",
               ifelse(judges$republican == 0 & judges$woman ==0, "Dem.Man",
                      ifelse(judges$republican == 1 & judges$woman == 0, "Rep.Man", NA))))
```

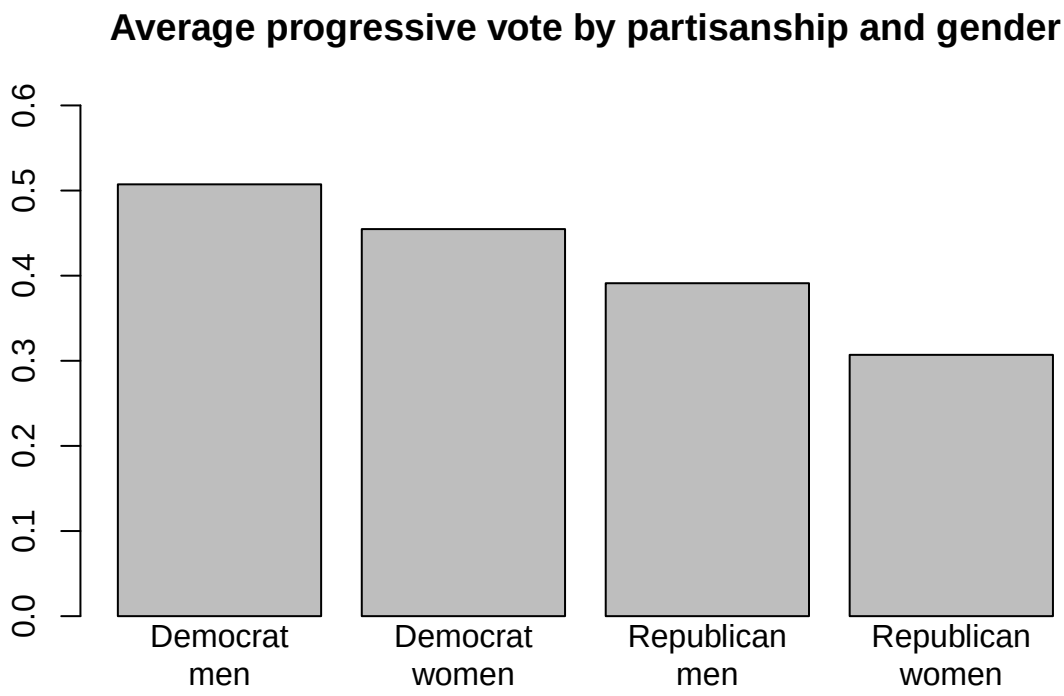
Number of observations in each category of the variable `gender_party`

```
table(judges$gender_party)
```

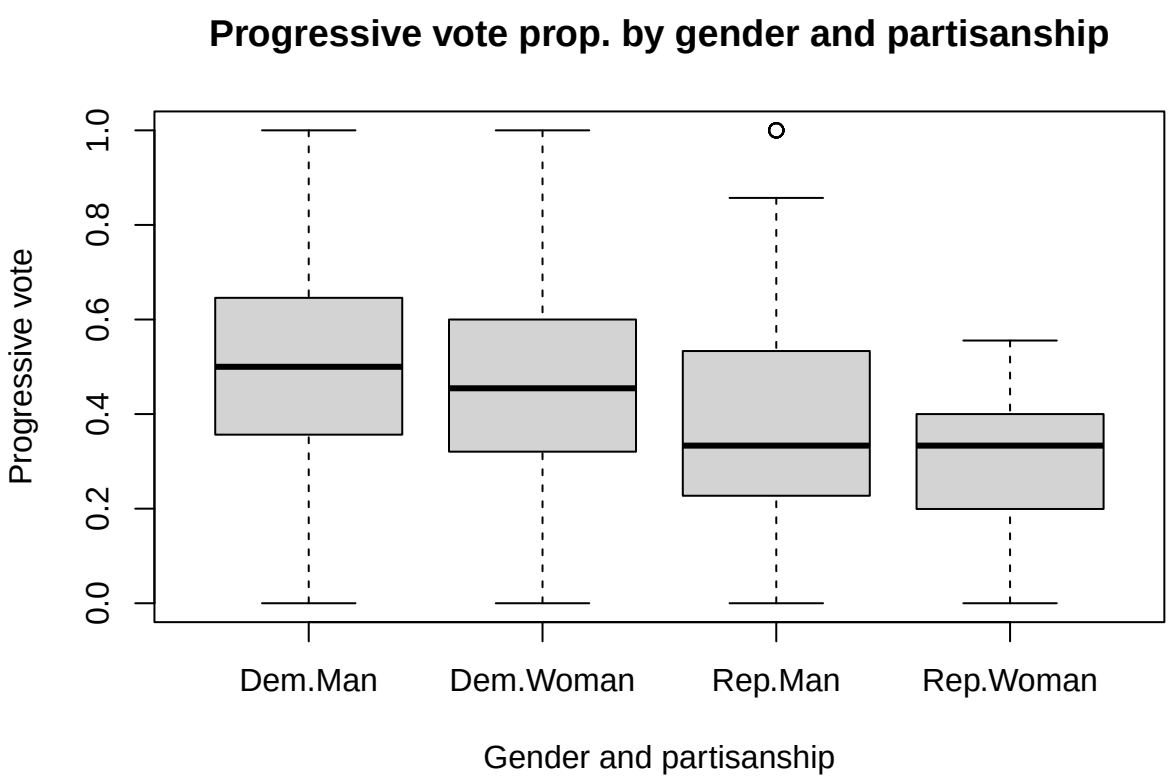
```
##
##   Dem.Man Dem.Woman   Rep.Man Rep.Woman
##        76        27       110        11
```

```
gender_party_means = tapply(judges$progressive_vote, judges$gender_party, mean)

barplot(gender_party_means, main = "Average progressive vote by partisanship and gender",
        names.arg = c("Democrat\nmen", "Democrat\nwomen", "Republican\nmen", "Republican\nwomen")
        ,ylim = c(0, 0.6))
```




```
boxplot(judges$progressive_vote ~  judges$gender_party,
        main = "Progressive vote prop. by gender and partisanship",
        xlab = "Gender and partisanship",
        ylab ="Progressive vote")
```



Insert your answer here: The barplot results are fairly surprising: even though you would expect democrats to be more progressive than republicans, it is more unexpected that democrat/republican men are consistently more progressive than women. In the case of the boxplot, I don't think it supports the idea that gender plays a role in progressive voting patterns, because we can see the median of dem men and dem women OR rep men and rep women are quite close. Rather, the biggest difference should be made by partisanship. However maybe we should refrain from making any causal claims because the median difference between democrats and republicans is not that big, it's just that the range differs.

Question 4 (25 points)

What is the difference in the proportion of pro-feminist decisions between judges who have at least one daughter and those who do not have any daughters?

- To compute this difference, first create a variable called `judges$any_girls` that is 1 when the judge has at least 1 girl and 0 otherwise. (*4 points*)
- Then, create a subset of the data called `parents` that contains judges that have at least one *child* (no matter whether boy or a girl). (*3 points*)
- Create an object called `ate` (for sample average treatment effect) that is the difference in means of `progressive_vote` between judges that have at least *one girl* versus those that have *no girls* among those judges with any children. (*3 points*)
- Provide a concise discussion of your findings in the text. In addition, answer the following questions:
  - What assumptions are required for us to interpret this difference as an estimate of a causal effect? (*5 points*)

- Are these assumptions plausible? Under what circumstances would any of these assumptions be violated? Note: it is sufficient to discuss one assumption. (*10 points*)
  * Think about under what circumstances (or assumptions) can we talk about causal effects?
  * Did the decision to run this analysis only among those judges, who had at least one child and not among the full sample, which includes those with no children at all, makes it more plausible to talk about causal effects?
  * Further, can you think about the existence of any potential confounding factors? Should we, for example, worry about the *number of children* to be a possible confounder? If yes discuss why. If no, discuss why not.

Answer 4

For your information: summary of the variable `any_girls`

```
judges$any_girls = ifelse(judges$girls > 0, 1, 0)
summary(judges$any_girls)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  1.0000  0.7232  1.0000  1.0000
```

```
parents = subset(judges, num_kids >0)
dim(parents)
```

```
## [1] 199  15
```

```
ate = mean(parents$progressive_vote[parents$girls == 1]) - mean(parents$progressive_vote[parents$girls =
ate
```

```
## [1] 0.1071728
```

Insert your answer here: There are some assumptions that allow us to see this difference as an estimate of the causal effect. First, the fact that the two samples (the one with judges with girls and the one without) are similar in all aspects except for whether they have a girl or not // the treatment. Under circumstances in which this would not be true, the results would not be persuasive. Another assumption is that the treatment has produced the difference in the outcomes, making judges with girls on average more pro-feminist in their decisions. This could be wrong in case there is a confounding factor that we did not control for.

I believe the decision to only look at judges with children is logical because we are comparing two samples that are similar on this aspect as well (being a parent), thus those two subsets are even more similar to each other than they would be with the judges that are not parents at all. However, this decreases external validity, because the effect might be different for those who do not have children, so it would also be helpful to test for that.

I think one confounding variable might be that, since judges are not randomly having children or not, there might be a systematic difference between the parent and non-parent judges which is unaccounted for in the analysis. On similar grounds, a bigger number of children is a decision that might be about values (perhaps judges with more children keep conservative values to higher regard).

Question 5 (20 points)

Given that the number of children might be a confounder for the relationship between number of girls and voting, let's estimate the effects using statistical control for the number of children using the following steps:

- Create one subset of the data called `girls_123` that restricts to judges with one, two or three children AND have at least one girl. (*2 points*)
- Create another subset of the data called `nogirls_123` that restricts to judges with one, two or three children and have no girls. (*2 points*)
- Calculate the mean of `progressive_vote` within levels of the variable numbers of kids (`num_kids`) for each of these subsets. Save these vectors as `girls_vote_by_nkids` and `nogirls_vote_by_nkids`. (*4 points*)
- Use these two vectors to estimate the sample average treatment effect within levels, saving this vector as `ate_nkids`. (*5 points*)
- Answer the following questions in the text:
  - Are these estimated effects largely similar or largely different than what you found using all of the data? (*1 point*)
  - What assumption do you need to make to interpret these effects causally? (*2 points*)
  - Do you think it is plausible in this case? (*4 points*)

Answer 5

For your information: dimensions of the dataset `girls_123`

```
girls_123 = subset(parents,num_kids %in% c(1,2,3) & girls > 0)
dim(girls_123)
```

```
## [1] 118  15
```

For your information: dimensions of the dataset `nogirls_123`

```
nogirls_123 = subset(parents,num_kids %in% c(1,2,3) & girls == 0)
dim(nogirls_123)
```

```
## [1] 34 15
```

Sample average treatment effect of having at least one girl by family size (justices with 1, 2, or 3 kids)

```
girls_vote_by_nkids =tapply(girls_123$progressive_vote, girls_123$num_kids, mean)
nogirls_vote_by_nkids= tapply(nogirls_123$progressive_vote, nogirls_123$num_kids, mean)

ate_ntkids = girls_vote_by_nkids - nogirls_vote_by_nkids
names(ate_ntkids) = c("one child", "two children", "three children")
ate_ntkids
```

```
##      one child   two children three children
##      0.2408984      0.0847949      0.1893141
```

Insert your answer here: The results are consistent with my belief that having children + having MANY children can be confounding variables. Again, the assumption in order to be able to causally analyze this is that there are no other confounding variables. Comparing this with the ATE from the previous exercise we can observe that the effect is much bigger on judges with one or three children, where having a girl produces bigger effects on the decisions of parent judges. Overall, I think the conclusion is plausible, as we controlled for two observable variables - gender and the number of children - when estimating the ate. Of course there could be another unobserved variables that might affect the voting behaviour of judges, but I think it is plausible and believable to say that judges with girls (and especially when the girl is an only child) influences judges decisions.

Bonus question 1 (15 points)

Let's consider the design of this study. The original authors assume that, conditional on the number of children a judge has, the number of daughters is random (as we did in question 5).

- Check the validity of this assumption. If this assumption is true, half of a judge's children should be female, on average. A deviation from this proportion could indicate that a gender preference among judges due to a stopping rule such as "have children until we get one girl," which would violate the randomization assumption.
  - One way to approach this task is to analyze the average gender distributions for judges separately by the number of children they have. We can find the average number of girls (or boys, equivalently) for judges with 1, 2, 3 children separately. Hopefully this reminds you of `tapply()`! (*3 points*)
  - Alternatively you can first calculate the proportion of girls a judge has (divide the number of girls by the number of total kids a judge has). Then find the average proportion of girls by family type (the average proportion of girls among judges with 1, 2 or 3 kids). (*3 points*)
  - Create a barplot that plots these proportions on the y-axis with the number of children on the x-axis. This barplot should have: (*4 points*)
    * informative labels on each axis
    * a y-axis range that runs from 0 to 1
    * a horizontal line at 0.5 to compare against.

Answer the following questions in the text:

- Is the assumption that the number of daughters is random reasonable? (*1 points*)
- Is there a scenario under which this assumption can be violated? (*2 points*)

- Do the data support the assumption? (*1 point*)
- Investigating the barplot, does it appear that there is strong gender preference/selection happening among judges? (*1 point*)

Answer bonus question 1

Do not forget to focus on the subset of judges with at least one child!

Average number of girls by family type (number of kids)

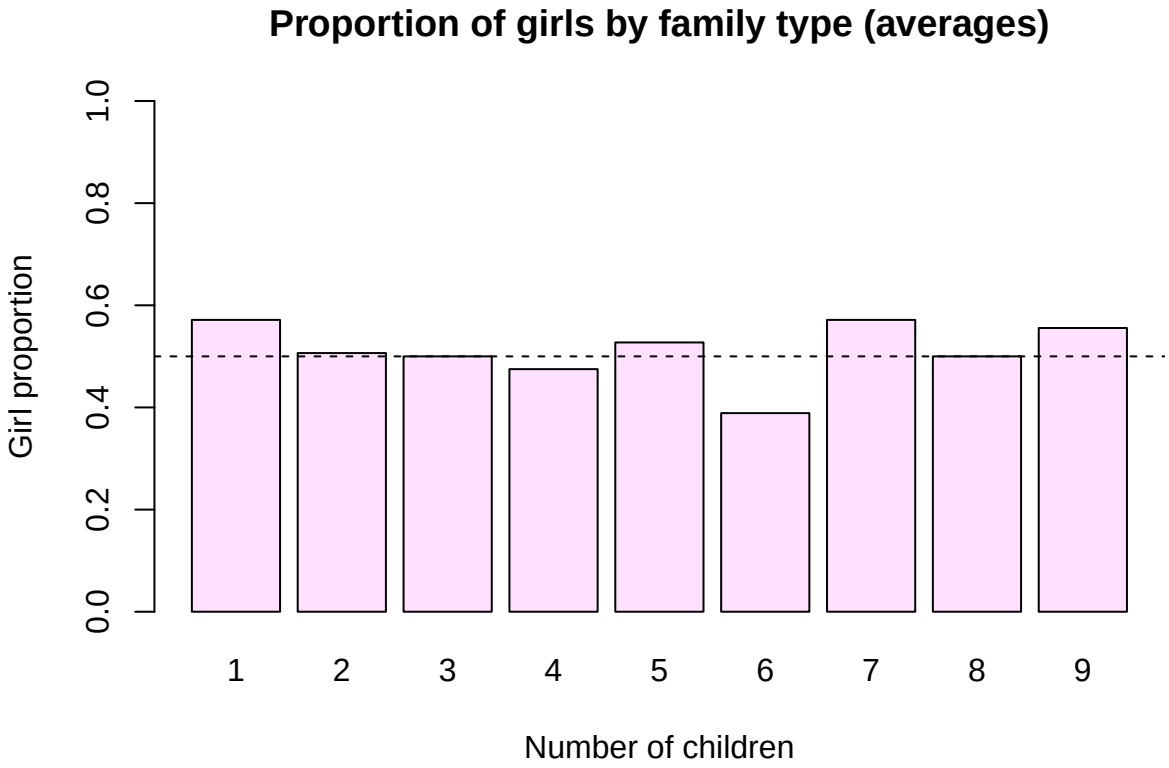
```
tapply (parents$girls, parents$num_kids, mean)
```

```
##         1         2         3         4         5         6         7         8
## 0.5714286 1.0129870 1.5000000 1.9000000 2.6363636 2.3333333 4.0000000 4.0000000
##         9
## 5.0000000
```

Average *proportion* of girls by family type (number of kids)

```
girl_proportion_by_family = tapply(parents$girls / parents$num_kids, parents$num_kids, mean)

barplot(girl_proportion_by_family,ylim = c(0,1),main = "Proportion of girls by family type (averages)",
abline(h=0.5, col=("black"),lty=2)
```



Insert your answer here:The girl proportion is close to 50% even in the cases of multiple children (with one outlier at 6 children). I don't think there is a gender preference, the data does not really support this idea of conventional wisdom.

Evaluation

- 5 questions for a total of 100 points
- 1 bonus question for 15 points