Varallyay Fall 2020
R Lab #1

1. Use the **table**, **summary**, or **describe** command to tabulate one variable; explain your results.

I chose to look at the *parborn* variable from the ARDA's GSS Dataset. This variable represents the respondent's response to the question, "Were both your parents born in this country?".

As Table 1 indicates, valid answers represent the various combinations of mother and father birth location, in terms of whether they were born in the United States. Zero represents that both parents were born in the U.S., and 8 represents that neither parent were born in the U.S. (The other categories were: (2) Mother no, father yes, (3) Mother yes, father don't know, (4) Mother no, father don't know, (5) Mother don't know, father yes, (6) Mother don't know, father no, and (7) Mother don't know, father don't know).

Table 2 shows the frequency of each category of parents' birth place, in terms of whether they were born in the U.S.

Table 3 shows that a large portion of respondents' parents were both born here in the U.S. (79%). The second largest category was *mother was born here, but whose father was no*t (3%), followed by *mother no, father yes* (1.84%), then *mother yes, father don't know* (0.2%), then *mother no, father don't know* (0.16%), *mother don't know, father yes* (0%), *mother don't know, father no* (0%), *mother don't know, father don't know* (0%), *neither born in U.S.* (15.6%). Of the 2992 respondents, 1523 responses were missing.

Table 1: *parborn*, from the Electronic Codebook

54) Were both your parents born in this country? (PARBORN)

	TOTAL	**%**
0) Both born in U.S.	2360	78.9
1) Mother yes, father no	91	3.0
2) Mother no, father yes	55	1.8
3) Mother yes, father don't know	6	0.2
4) Mother no, father don't know	5	0.2
5) Mother don't know, father yes	1	0.0

6) Mother don't know, father no	1	0.0
7) Mother don't know, father don't know	1	0.0
8) Neither born in U.S.	467	15.6
9) No answer	5	0.2
Missing	1518	
TOTAL	2992	100.0

Table 2: Responses for *parborn* (Frequencies)

```
table(gss$parborn)
```

```
   0   1   2   3   4   5   6   7   8
2360  91  55   6   5   1   1   1 467
```

Table 3: Distribution of *parborn* (Percentages)

```
prop.table(table(gss$parborn))
```

```
         0          1          2          3          4          5
0.7900903917 0.0304653498 0.0184131235 0.0020087044 0.0016739203 0.0003347841
         6          7          8
0.0003347841 0.0003347841 0.1563441580
```

2. Using either the **plyr** or **doBy** commands, compare the means and standard deviations for one variable, for at least two sub-groups; explain your results.

I chose to look at the variable, *natenvir*, that investigates if we are spending too much, too little, or about the right amount on improving and protecting the environment, as Table 1 indicates. The second variable I chose to look at, *spkrac,* asked respondents to "consider a person who believes that Blacks are genetically inferior, if such a person wanted to make a speech in your community claiming that Blacks are inferior, should he be allowed to speak, or not?". The choice to investigate these two variables was not based on any theory. I currently

believe the two most important factors in American society revolve around systemic racism and environmental impact/climate change. As such, I was curious whether an individual's parents' birth place (whether it was in the U.S. or not) had any bearing on views towards environmental spending or towards racist public speakers being allowed to present in their community. Moreover, I thought it would be interesting to compare the two. I realize this is not an ideal approach to research, but as an exercise to familiarize myself with R and to start exercising my statistical literacy, I thought it would be okay.

I looked at differences in the mean and median (and standard deviations) responses of spending on environment, by whether a racist individual should be allowed to make a racist speech in the respondent's community. As Table 5 indicates, represented are the descriptive statistics for *spkrac* – (1)should be allowed to speak, or (2) should not be allowed to speak, for each category of environmental spending – (1) too little, (2) about the right amount, or (3) too much. Now, I am not quite sure if this is really interpretable considering the nature of the variables. Essentially, it seems that, on average, individuals who stated we were spending too little on the environment, also tended to believe the racist speaker should be allowed to speak (Mean = 1.34, SD = 0.47). Individuals who reported we were spending about the right amount on the environment also tended to believe the racist speaker should be allowed to speak (Mean = 1.39, SD = 0.49). Individuals who stated we were spending too much on the environment, also tended to believe the racist speaker should be allowed to speak (Mean = 1.34, SD = 0.47). My approach was based on rounding down to 1 (should be allowed to speak) as none of the means were above 1.5. These results and my interpretation could make sense if we consider that these responses are probably uncorrelated and that Americans hold dear the First Amendment of the Bill of Rights, the freedom of speech. Despite this potential rational, as *spkrac* is also a categorical variable, this interpretation does not hold in a statistical sense. After all, what categorical value exists between yes and no, between 1 and 2? At the very least, this was a good exercise in attempting to explain/describe what I found (Table 1-4 show data for *natenvir* and *spkrac* from the Electronic Codebook as well as the descriptive statistics as a function of *parborn,* respectively).

Table 1: *natenvir*, from the Electronic Codebook

80) Are we spending too much, too little, or about the right amount on (ITEM)? B. Improving and protecting the environment (NATENVIR)

	TOTAL	%
1) Too little	992	66.8

2) About right	365	24.6
3) Too much	89	6.0
8) Don't know	38	2.6
Missing	3026	
TOTAL	1484	100.0

Table 2: Mean and SD for responses to *natenvir,* as a function of *parborn*

```
ddply(gss, "parborn", summarise, Mean = mean(natenvir, na.rm = T), SD = sd(natenvir, na.rm = T))
```

	parborn	Mean	SD
1	0	1.372073	0.5961252
2	1	1.439024	0.6726359
3	2	1.400000	0.6746647
4	3	3.000000	NA
5	4	1.000000	NA
6	5	NaN	NA
7	6	1.000000	NA
8	7	NaN	NA
9	8	1.377880	0.5814009
10	NA	1.000000	0.0000000

Table 3: *spkrac*, from the Electronic Codebook

112) Or consider a person who believes that Blacks are genetically inferior. A. If such a person wanted to make a speech in your community claiming that Blacks are inferior, should he be allowed to speak, or not? (SPKRAC)

	TOTAL	**%**
1) Yes, allowed to speak	1236	61.7
2) Not allowed	740	36.9

8) Don't know	24	1.2
9) No answer	3	0.1
Missing	2507	
TOTAL	2003	100.0

Table 4: Mean and SD for responses to *spkrac,* as a function of *parborn*

```
ddply(gss, "parborn", summarise, Mean = mean(spkrac, na.rm = T), SD = sd(spkrac, na.rm = T))
```

	parborn	Mean	SD
1	0	1.346351	0.4759590
2	1	1.296875	0.4604927
3	2	1.414634	0.4987790
4	3	1.500000	0.5773503
5	4	1.750000	0.5000000
6	5	2.000000	NA
7	6	1.000000	NA
8	7	NaN	NA
9	8	1.528814	0.5000173
10	NA	1.250000	0.5000000

Table 5: Comparison of mean and SD for responses to *natenvir* and *spkrac*

```
ddply(gss, "natenvir", summarise, Mean = mean(spkrac, na.rm = T), SD = sd(spkrac, na.rm = T))
```

natenvir <int>	**Mean** <dbl>	**SD** <dbl>
1	1.337481	0.4732184
2	1.394309	0.4896981
3	1.400000	0.4940322
NA	1.391431	0.4883083

4 rows

3. Using the **gmodels** or another command, cross-tabulate two categorical or ordinal variables (getting proportions); and explain your results.

I chose to stick with the variables, *parborn* and *natenvir*. I wanted to cross-tabulate these variables hypothesizing that the "outdated" cultural value of stewardship – ostensibly modeled by grandparents to parents then to offspring (the respondents), would be transmitted. As seen in Table 3, of the total observations (1444 respondents with valid responses), for those who claimed we were spending too little on the environment, 80% stated both their parents were born in the U.S. Interestingly, percentages drop tremendously for all other categories of *parborn* except for those whose parents were both ***not*** born in the U.S. (about 15%). Though, that is also a huge difference. We have a similar pattern for the other categories of *natenvir.* Moreover, the proportions of responses are also similar. For those who responded that we were spending about the right amount on the environment, 79% of respondents stated both parents were born in the U.S. For those who responded that we were spending too much on the environment, about 79% of respondents also stated both parents were born in the U.S. Furthermore, in terms of the other categories of *parborn,* the same pattern appears, there is a tremendous drop in percentages except for those whose parents were both ***not*** born in the U.S. For those who stated we were spending about the right amount, 16% stated both their parents were ***not*** born in the U.S. And finally, for those who stated we were spending too much on the environment, 12% stated both their parents were ***not*** born in the U.S. What I find most interesting is that the pattern, the distribution of proportions, are almost identical for the 3 different categories of environmental spending. As far as we have seen I don't think I can make a claim whether the hypothesis is supported.

Table 1: *parborn*, from the Electronic Codebook

54) Were both your parents born in this country? (PARBORN)

	TOTAL	**%**
0) Both born in U.S.	2360	78.9
1) Mother yes, father no	91	3.0
2) Mother no, father yes	55	1.8
3) Mother yes, father don't know	6	0.2
4) Mother no, father don't know	5	0.2

5) Mother don't know, father yes	1	0.0
6) Mother don't know, father no	1	0.0
7) Mother don't know, father don't know	1	0.0
8) Neither born in U.S.	467	15.6
9) No answer	5	0.2
Missing	1518	
TOTAL	2992	100.0

Table 2: *natenvir*, from the Electronic Codebook

80) Are we spending too much, too little, or about the right amount on (ITEM)? B. Improving and protecting the environment (NATENVIR)

	TOTAL	**%**
1) Too little	992	66.8
2) About right	365	24.6
3) Too much	89	6.0
8) Don't know	38	2.6
Missing	3026	
TOTAL	1484	100.0

Table 3: Cross-Tabulation of respondent's parents' birth place – in terms of whether they were born in the U.S. (*parborn*), with the Respondent's view on our environmental spending (*natenvir*)

```
CrossTable(gss$parborn, gss$natenvir, prop.r=F, prop.c=T, prop.t=F, prop.chisq=F, format="SPSS")
```

```
##
##    Cell Contents
## |-------------------------|
## |                   Count |
## |           Column Percent |
## |-------------------------|
##
## Total Observations in Table:  1444
##
##                | gss$natenvir
##  gss$parborn |         1  |         2  |         3  | Row Total  |
## -------------|------------|------------|------------|------------|
##            0 |       794  |       289  |        70  |      1153  |
##              |    80.202% |    79.178% |    78.652% |            |
## -------------|------------|------------|------------|------------|
##            1 |        27  |        10  |         4  |        41  |
##              |     2.727% |     2.740% |     4.494% |            |
## -------------|------------|------------|------------|------------|
##            2 |        21  |         6  |         3  |        30  |
##              |     2.121% |     1.644% |     3.371% |            |
## -------------|------------|------------|------------|------------|
##            3 |         0  |         0  |         1  |         1  |
##              |     0.000% |     0.000% |     1.124% |            |
## -------------|------------|------------|------------|------------|
##            4 |         1  |         0  |         0  |         1  |
##              |     0.101% |     0.000% |     0.000% |            |
## -------------|------------|------------|------------|------------|
##            5 |         0  |         0  |         0  |         0  |
##              |     0.000% |     0.000% |     0.000% |            |
## -------------|------------|------------|------------|------------|
##            6 |         1  |         0  |         0  |         1  |
##              |     0.101% |     0.000% |     0.000% |            |
## -------------|------------|------------|------------|------------|
##            7 |         0  |         0  |         0  |         0  |
##              |     0.000% |     0.000% |     0.000% |            |
## -------------|------------|------------|------------|------------|
##            8 |       146  |        60  |        11  |       217  |
##              |    14.747% |    16.438% |    12.360% |            |
## -------------|------------|------------|------------|------------|
## Column Total |       990  |       365  |        89  |      1444  |
##              |    68.560% |    25.277% |     6.163% |            |
## -------------|------------|------------|------------|------------|
```