Homosexuality Opinion in the US: Does Country of Origin Matter?

Introduction:

Homosexuality is defined as romantic or sexually attractive behavior towards an individual of the same gender/sex.It is opposite to heterosexual behavior which is considered to be more natural and is better accepted by the human society. Homosexuality has been in existence since the BCE era.Historical records mention it being practiced in most regions of the world.

Science has attributed this behavior to a combination of biological (genetic and hormonal) and environmental (sociological, psychological) factors. Social acceptance of homosexuality has varied over time and regions. While some regions may prohibit such behavior and regard it as a sin/crime others may be neutral or accept it.

Most developed nations such as the US, Europe and Canada have accepted homosexuality and there are laws in place which prohibit workplace and social discrimination on grounds of sexual orientation.However, regions such as Africa and Asia continue to view homosexuality as a social stigma. A recent development in this direction being in India where homosexuality was recognized as a crime by the Supreme Court of India.

United States is considered to be a popular immigrant destination which is reflected in the significant proportion of immigrants in the US population [13% in 2010(US Census Bureau)]. The high number of immigrants implies that people with different cultural and societal backgrounds reside in the US and live together with the US born population.

As mentioned earlier, homosexuality seems to have received greater acceptance in developed countries such as the US with anti-discrimination laws being passed. Hence it would be interesting to know the perception of the general US population which would comprise of both US born and not US born people. Do people born in the US allow for same sex relations ? What about the not US born people? Do their opinions differ from the US born population ? This study uses the US GSS data from 1972-2012 to explore these questions.

Data:

The United States General Social Survey (US GSS) data from 1972-2012 is used in this work. This dataset has 57,061 cases and 114 variables. The dataset has missing values corresponding to no responses by people which are coded as NA. The units of observation are responses of people to questions. Variables of interest for this study are homosex (opinion on homosexual behavior) and born (US or other country).Both are categorical variables. The variable homosex has 5 categories (Always Wrong(AW),Almst Always Wrg (AAW),Sometimes Wrong (SW), Not Wrong At All (NWAA), Other) whereas born has 2 (Yes, No). The population of interest is the US population. Since the data includes a significant portion of the US population and covers a long span of time, this dataset can be assumed to represent the population of interest. This analysis could be used as an initial step towards determining an association between country of birth and homosexuality opinion. In order to confidently report a causal link, analysis of possible confounding factors such as age, gender etc. should also be done which is beyond the scope of this work.

Given below is a summary of the homosex variable in the raw dataset.

```
##     Always Wrong Almst Always Wrg  Sometimes Wrong Not Wrong At All
##            21601             1581             2243             7282
##            Other             NA's
##               82            24272
```

Given below is a summary of the born variable in the raw dataset.

```
##   Yes    No  NA's
## 43705  4099  9257
```

Exploratory data analysis:

The data was processed to remove missing values and archive the responses on homosexuality based on country of birth. It is important to point out here that for this study only that part of the dataset was employed where responses were available for both homosexuality opinion and country of birth. Hence data was filtered according to this criteria.

Provided below is the filtered data.

```
##              Opinion on Homosexuality
## Birth Country    AW  AAW   SW NWAA Other
##         US    17049 1191 1742 6148     0
##         Other  1454  132  159  549     0
```

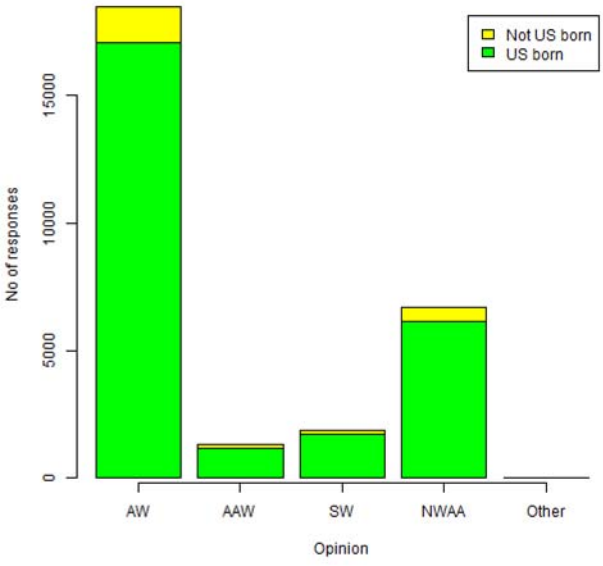


Figure 1: Bar plot showing the people's responses on homosexuality (each bar consists of two stacked bars corresponding to US born (green) and not US born (yellow) respondents)

From this figure, it can be seen that the number of respondents not born in US are much less than those born in the US, which makes sense as the survey was conducted in the US. However, this study will make use of the proportions in these two populations for comparison rather than the absolute numbers.

The proportions in the two groups having a particular opinion look close to each other based on a rough calculation.

For further analysis, the negative opinions (levels AW, AAW and SW) on homosexuality are combined together to form a level Wrong and the level NWAA is represented as Right.

The constructed 2x2 table is shown below.

```
##              Opinion on Homosexuality
## Birth Country Wrong Right
##   US born     19982  6148
##   Not US born  1745   549
```

The 2x2 table in terms of proportions is shown below.

```
##              Opinion on Homosexuality
## Birth Country  Wrong  Right
##   US born     0.7647 0.2353
##   Not US born 0.7607 0.2393
```

Inference:

1.) Test for Equality of Proportions Between the 2 Populations

This test was used to determine whether there is a statistically significant difference between proportion of people in the two groups (US born and not US born) who think that homosexual behavior is wrong. The conditions for application of this test are evaluated below.

a.) Independence: WITHIN GROUPS: The US GSS dataset involves random sampling and the number of observations are definitely less than 10% of the total population. BETWEEN GROUPS: The two groups are non-paired as they have different individuals (born in US and not born in US).

b.) Sample size/skew: The number of successes (homosexuality is wrong) and failures (homosexuality is right) in both groups are greater than 10.

Thus, all conditions are satisfied.

The hypotheses are as follows:

Null Hypothesis: proportion in US born = proportion in not US born

Alternate Hypothesis: proportion in US born is different from proportion in not US born

(A significance level of 0.05 is used here.)

The results of the test are shown below.

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  new_data
## X-squared = 0.1689, df = 1, p-value = 0.6811
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.01440  0.02247
## sample estimates:
## prop 1 prop 2
## 0.7647 0.7607
```

Since the p-value (0.68) is greater than the assumed significance level, we cannot reject the null hypothesis. Therefore, it can be concluded that the proportions in the populations (US born and not US born) are not significantly different.

2.) Odds Ratio

Odds ratio in this study is defined as the ratio of the odds in both groups (US-born and not US born) of believing that homosexuality is wrong. It could be used to determine if the odds of getting a negative opinion on homosexuality are different in the two groups or the same. The null value for an odds ratio is 1 indicating the odds are the same in both groups. A value greater than 1 would indicate more chances of a negative opinion in the US-born group and a value less than 1 would indicate less chances of a negative opinion in the US-born group.

Null Hypothesis: The odds ratio is 1.0 (Null value)

Alternate Hypothesis: The odds ratio is different than 1.0 (Two-sided test)

Test results are shown below.

```
## prop 1 prop 2
## 0.7647 0.7607
```

```
## prop 1
##  1.023
```

```
## [1] 0.9252 1.1302
```

The computed odds ratio is very close to 1 (1.023). However, to check whether it is statistically significant or not, a 95% confidence interval is constructed and if it includes the null value (1.0) then we can conclude that the odds of having a negative opinion are the same in both populations. The obtained confidence interval is (0.9252 1.1302) which contains the null value. Hence the null hypothesis cannot be rejected and it can be stated that the negative response is equally likely in both populations.

3.) Chi-Squared Independence Test

The Chi-Squared Independence test is used here to determine whether there is any association between opinion on homosexuality and the country an individual is born in. For this, all the levels of opinion on homosexuality were considered and compared to the country of birth. Since no responses were there for the Other level it was excluded. The conditions to be satisfied for the test are evaluated below.

a.) Independence: Sampling observations are independent since US GSS employs random sampling. The value of total cases is definitely much smaller than the respective populations (US born and not US born). Each case contributes to one cell in the table.

b.) Sample Size: Each particular scenario has more than 5 expected cases.

Thus, all the conditions are satisfied.

Null Hypothesis: No association between the country of birth and homosexuality opinion

Alternate Hypothesis: There is a relationship between country of birth and homosexuality opinion

(A significance level of 0.05 is used here.)

Test results are shown below.

```
##
##  Pearson's Chi-squared test
##
## data:  opinion1
## X-squared = 7.975, df = 3, p-value = 0.04654
```

The obtained p-value (0.04654) is less than the assumed significance level which indicates that there is an association between the homosexuality opinion and country of birth.

Thus, the third test reports a different result than the first two.

Conclusion:

This study aimed at finding out whether an individual's opinion on homosexuality in the US depends on his/her country of birth. The dataset from US GSS from 1972-2012 was used from which responses were filtered out according to country of birth. 3 statistical tests (equality of proportions between 2 populations, odds ratio and chi-squared independence test) were performed. For the first two tests, the negative responses were combined to give a 2x2 contingency table whereas the original table was used in test 3. Results from the first two tests indicate that in the US, there is no association between opinion on homosexuality and country of birth. However, the third test indicates that an association exists. These contradictory results could be attributed to a phenomenon known as Simpson's paradox due to which trends are different with the contents of the table combined than with the original table. Resolving this is beyond the scope of this work and can be part of future work. However, if a conclusion has to be derived we could rely on the result from the third test where the categories were not modified. Hence there exists an association between the country an individual is born in and opinion on homosexuality. However, further work (e.g. correcting for Simpson's paradox, considering possible confounding factors and accounting for them) would be required in order to provide more conclusive evidence in favor of this association.Also, the data used in this study spanned a long period of time (1972-2012). An interesting analysis could be to consider data for more recent years (eg. 1990s to present). A major result from this analysis is that in both groups a major proportion of people still consider homosexual behavior to be wrong which reflects the fact that in the US, the society still has a long way to go in accepting homosexuality.

Acknowledgements

I would like to thank Dr. Mine Çetinkaya-Rundel for the lectures, guidelines in preparing this project and providing the interesting US GSS dataset.

References

1.) http://bit.ly/dasi_gss_data

2.) Agresti, A. (1996). An Introduction to Categorical Data Analysis, Wiley, New York.

3.) Agresti, A. (2002). Categorical Data Analysis (2nd Ed.), Wiley, New York.

4.) https://www.wikipedia.org/

5.) http://www.openintro.org/

Appendix 1

This appendix contains a part of the dataset

1.) Opinion on homosexuality

```
##   [1] Always Wrong     Always Wrong     Not Wrong At All Always Wrong
##   [5] Always Wrong     Always Wrong     Always Wrong     Always Wrong
##   [9] Always Wrong     Always Wrong     Always Wrong     Always Wrong
##  [13] <NA>             Always Wrong     Almst Always Wrg Not Wrong At All
##  [17] Not Wrong At All <NA>             Always Wrong     Always Wrong
##  [21] Always Wrong     Always Wrong     Always Wrong     Always Wrong
##  [25] Always Wrong     Always Wrong     Sometimes Wrong  Always Wrong
##  [29] Always Wrong     Always Wrong     Always Wrong     Always Wrong
##  [33] Sometimes Wrong  Always Wrong     Always Wrong     Always Wrong
##  [37] Always Wrong     Always Wrong     Always Wrong     Always Wrong
##  [41] Always Wrong     Other            Always Wrong     Almst Always Wrg
##  [45] Always Wrong     Sometimes Wrong  Always Wrong     Always Wrong
##  [49] Always Wrong     Not Wrong At All Almst Always Wrg Almst Always Wrg
##  [53] Always Wrong     Always Wrong     Always Wrong     Not Wrong At All
##  [57] Always Wrong     Always Wrong     Always Wrong     Always Wrong
##  [61] Always Wrong     Always Wrong     Always Wrong     Always Wrong
##  [65] Always Wrong     Always Wrong     Always Wrong     Almst Always Wrg
##  [69] Always Wrong     Not Wrong At All Sometimes Wrong  Always Wrong
##  [73] Always Wrong     Sometimes Wrong  Always Wrong     Always Wrong
##  [77] Almst Always Wrg Not Wrong At All Always Wrong     Not Wrong At All
##  [81] Always Wrong     <NA>             Always Wrong     Always Wrong
##  [85] Always Wrong     Not Wrong At All Almst Always Wrg Always Wrong
##  [89] Always Wrong     Not Wrong At All Almst Always Wrg Always Wrong
##  [93] Always Wrong     Always Wrong     Always Wrong     Not Wrong At All
##  [97] Always Wrong     Almst Always Wrg Not Wrong At All Always Wrong
```

```
##  [97] Always Wrong     Almst Always Wrg Not Wrong At All Always Wrong
## [101] Always Wrong     Always Wrong     Always Wrong     Not Wrong At All
## [105] Always Wrong     Always Wrong     Always Wrong     Not Wrong At All
## [109] Always Wrong     Always Wrong     Always Wrong     Almst Always Wrg
## [113] Always Wrong     Always Wrong     Almst Always Wrg Always Wrong
## [117] Not Wrong At All Always Wrong     Not Wrong At All Always Wrong
## [121] Always Wrong     Always Wrong     Almst Always Wrg Not Wrong At All
## [125] Not Wrong At All Always Wrong     Always Wrong     Always Wrong
## [129] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [133] Always Wrong     Always Wrong     Always Wrong     Almst Always Wrg
## [137] Always Wrong     Not Wrong At All Always Wrong     Always Wrong
## [141] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [145] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [149] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [153] Other            Always Wrong     Always Wrong     Always Wrong
## [157] Always Wrong     Always Wrong     Always Wrong     Not Wrong At All
## [161] Always Wrong     Always Wrong     Always Wrong     Other
## [165] Always Wrong     Almst Always Wrg Always Wrong     Not Wrong At All
## [169] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [173] Not Wrong At All Always Wrong     Always Wrong     Not Wrong At All
## [177] Always Wrong     Sometimes Wrong  Other            Not Wrong At All
## [181] Always Wrong     Sometimes Wrong  Always Wrong     Always Wrong
## [185] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [189] Always Wrong     <NA>             Always Wrong     Always Wrong
## [193] Always Wrong     Not Wrong At All Always Wrong     Always Wrong
## [197] Sometimes Wrong  Always Wrong     Almst Always Wrg Always Wrong
## [201] Not Wrong At All Always Wrong     Always Wrong     Always Wrong
## [205] Not Wrong At All Almst Always Wrg Sometimes Wrong  Always Wrong
## [209] <NA>             Always Wrong     Always Wrong     Always Wrong
## [213] Sometimes Wrong  Always Wrong     Always Wrong     <NA>
## [217] <NA>             Sometimes Wrong  Always Wrong     Always Wrong
## [221] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [225] Sometimes Wrong  Not Wrong At All Always Wrong     Always Wrong
## [229] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [233] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [237] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [241] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [245] Always Wrong     Sometimes Wrong  Sometimes Wrong  Not Wrong At All
## [249] Always Wrong     Always Wrong     Not Wrong At All Always Wrong
## [253] Always Wrong     Sometimes Wrong  Not Wrong At All Not Wrong At All
## [257] Sometimes Wrong  Always Wrong     Always Wrong     Always Wrong
## [261] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [265] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [269] Always Wrong     Not Wrong At All Always Wrong     Other
## [273] Not Wrong At All Not Wrong At All Always Wrong     Not Wrong At All
## [277] <NA>             Almst Always Wrg Always Wrong     Always Wrong
## [281] Always Wrong     Always Wrong     Not Wrong At All Always Wrong
## [285] Always Wrong     Almst Always Wrg Not Wrong At All Always Wrong
## [289] Almst Always Wrg Always Wrong     Always Wrong     Always Wrong
## [293] Always Wrong     Always Wrong     Almst Always Wrg Other
## [297] Always Wrong     Always Wrong     Always Wrong     Almst Always Wrg
## [301] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [305] Almst Always Wrg Always Wrong     Sometimes Wrong  Sometimes Wrong
## [309] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [313] Always Wrong     Always Wrong     Not Wrong At All Always Wrong
## [317] Always Wrong     Sometimes Wrong  Always Wrong     Not Wrong At All
## [321] Always Wrong     Always Wrong     Sometimes Wrong  Not Wrong At All
## [325] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [329] Always Wrong     Not Wrong At All Always Wrong     Always Wrong
## [333] <NA>             Always Wrong     Always Wrong     Always Wrong
## [337] Always Wrong     <NA>             Always Wrong     <NA>
## [341] Always Wrong     Not Wrong At All Sometimes Wrong  Always Wrong
## [345] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [349] Always Wrong     Almst Always Wrg <NA>             Always Wrong
## [353] Always Wrong     Always Wrong     Sometimes Wrong  Always Wrong
## [357] Not Wrong At All Not Wrong At All Always Wrong     Not Wrong At All
## [361] Always Wrong     Not Wrong At All Always Wrong     Always Wrong
## [365] Not Wrong At All Not Wrong At All Always Wrong     Always Wrong
## [369] Always Wrong     Always Wrong     Not Wrong At All Almst Always Wrg
## [373] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [377] Always Wrong     Always Wrong     Sometimes Wrong  Always Wrong
## [381] Sometimes Wrong  Always Wrong     Always Wrong     Always Wrong
## [385] Not Wrong At All Sometimes Wrong  Sometimes Wrong  Almst Always Wrg
## [389] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [393] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [397] Always Wrong     Always Wrong     <NA>             Always Wrong
## [401] Almst Always Wrg Sometimes Wrong  Always Wrong     Always Wrong
## [405] Always Wrong     Sometimes Wrong  Almst Always Wrg Always Wrong
## [409] Sometimes Wrong  <NA>             <NA>             <NA>
## [413] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [417] Always Wrong     Always Wrong     Always Wrong     Not Wrong At All
## [421] Sometimes Wrong  Always Wrong     Always Wrong     Always Wrong
## [425] Always Wrong     Almst Always Wrg Sometimes Wrong  Always Wrong
## [429] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [433] Always Wrong     Sometimes Wrong  Always Wrong     Always Wrong
## [437] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [441] Always Wrong     Always Wrong     Always Wrong     Sometimes Wrong
## [445] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [449] Always Wrong     Sometimes Wrong  Always Wrong     Always Wrong
## [453] Always Wrong     Always Wrong     Sometimes Wrong  Almst Always Wrg
## [457] Almst Always Wrg Always Wrong     Always Wrong     Always Wrong
## [461] Always Wrong     Always Wrong     Always Wrong     Sometimes Wrong
## [465] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [469] Always Wrong     Always Wrong     Always Wrong     Not Wrong At All
## [473] Always Wrong     Almst Always Wrg Always Wrong     Sometimes Wrong
## [477] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [481] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [485] Always Wrong     Always Wrong     Always Wrong     <NA>
## [489] Always Wrong     Always Wrong     Always Wrong     Almst Always Wrg
## [493] Always Wrong     Always Wrong     Always Wrong     Always Wrong
## [497] Always Wrong     Always Wrong     Sometimes Wrong  <NA>
## [501] Not Wrong At All
## 5 Levels: Always Wrong Almst Always Wrg ... Other
```

2.) Birth Country

```
##   [1] Yes  Yes  No   Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
##  [15] Yes  Yes  Yes  Yes  No   Yes  Yes  Yes  Yes  Yes  No   Yes  No   Yes
##  [29] Yes  Yes  Yes  No   Yes  No   Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
##  [43] Yes  Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  Yes  Yes  Yes  Yes  Yes
##  [57] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
##  [71] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
##  [85] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
##  [99] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [113] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [127] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [141] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [155] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  Yes
## [169] No   Yes  Yes  Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  Yes  Yes  Yes
## [183] Yes  No   Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [197] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  Yes  Yes  Yes
## [211] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [225] Yes  Yes  Yes  Yes  Yes  Yes  Yes  No   Yes  <NA> Yes  Yes  Yes  Yes
## [239] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [253] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  Yes  Yes  Yes
## [267] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [281] Yes  Yes  Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  Yes  Yes  Yes  Yes
## [295] Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  No   Yes  Yes  Yes  Yes  Yes
## [309] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [323] Yes  Yes  Yes  No   Yes  Yes  Yes  No   Yes  Yes  No   Yes  Yes  No
## [337] No   No   Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [351] Yes  Yes  Yes  No   Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  Yes  Yes
## [365] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  Yes  Yes  Yes
## [379] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [393] Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [407] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  Yes  Yes
## [421] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [435] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [449] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [463] Yes  Yes  Yes  Yes  Yes  No   Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [477] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## [491] Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
## Levels: Yes No
```

Appendix 2

This appendix contains the R-code and output.

```
### Load the dataset and provide summary of the variables ###

load(url("http://bit.ly/dasi_gss_data"))
summary(gss$homosex)
```

```
##     Always Wrong Almst Always Wrg  Sometimes Wrong Not Wrong At All
##            21601             1581             2243             7282
##            Other             NA's
##               82            24272
```

```
summary(gss$born)
```

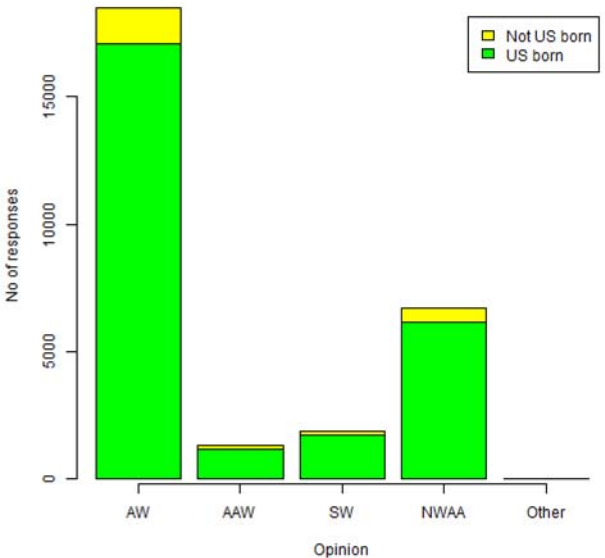
```
##   Yes    No  NA's
## 43705  4099  9257
```

```
### Declare counters to count the number of responses for each category ###
### The categories are listed below ###

# Category 1: Homosex: Always Wrong, Birth: Yes # Category 2: Homosex:
# Always Wrong, Birth: No # Category 3: Homosex: Almst Always Wrg, Birth:
# Yes # Category 4: Homosex: Almst Always Wrg, Birth: No # Category 5:
# Homosex: Sometimes Wrong, Birth: Yes # Category 6: Homosex: Sometimes
# Wrong, Birth: No # Category 7: Homosex: Not Wrong At All, Birth: Yes #
# Category 8: Homosex: Not Wrong At All, Birth: No # Category 9: Homosex:
# Other, Birth: Yes # Category 10: Homosex: Other, Birth: No #

count1 = 0
count2 = 0
count3 = 0
count4 = 0
count5 = 0
count6 = 0
count7 = 0
count8 = 0
count9 = 0
count10 = 0

### Loop to get responses according to birth country

for (i in 1:length(gss$homosex)) {


    if (!is.na(gss$homosex[i]) && !is.na(gss$born[i])) {


        if (gss$homosex[i] == "Always Wrong" && gss$born[i] == "Yes") {
```

```
        count1 = count1 + 1

    } else if (gss$homosex[i] == "Always Wrong" && gss$born[i] == "No") {
        count2 = count2 + 1


    } else if (gss$homosex[i] == "Almst Always Wrg" && gss$born[i] == "Yes") {
        count3 = count3 + 1


    } else if (gss$homosex[i] == "Almst Always Wrg" && gss$born[i] == "No") {
        count4 = count4 + 1


    } else if (gss$homosex[i] == "Sometimes Wrong" && gss$born[i] == "Yes") {

        count5 = count5 + 1


    } else if (gss$homosex[i] == "Sometimes Wrong" && gss$born[i] == "No") {

        count6 = count6 + 1

    } else if (gss$homosex[i] == "Not Wrong At All" && gss$born[i] == "Yes") {

        count7 = count7 + 1


    } else if (gss$homosex[i] == "Not Wrong At All" && gss$born[i] == "No") {

        count8 = count8 + 1


    } else if (gss$homosex[i] == "Other" && gss$born[i] == "Yes") {

        count9 = count9 + 1


    } else if (gss$homosex[i] == "Other" && gss$born[i] == "No") {


        count10 = count10 + 1

    }


}
}


### Bar plot & display filtered data ###
opinion = matrix(c(count1, count2, count3, count4, count5, count6, count7, count8,
    count9, count10), nrow = 2, ncol = 5)
dimnames(opinion) <- list(c("US", "Other"), c("AW", "AAW", "SW", "NWAA", "Other"))
names(dimnames(opinion)) <- c("Birth Country", "Opinion on Homosexuality")
print(opinion)
```

```
##              Opinion on Homosexuality
## Birth Country    AW  AAW   SW NWAA Other
##         US    17049 1191 1742 6148     0
##         Other  1454  132  159  549     0
```

```
op = c("AW", "AAW", "SW", "NWAA", "Other")
country = c("US born", "Not US born")
barplot(opinion, names.arg = op, col = c("Green", "Yellow"), legend = country,
    xlab = "Opinion", ylab = "No of responses", axis.lty = 1)
```



```
### Display the derived 2x2 contingency table ###
new_data = matrix(c(sum(opinion[1, 1:3]), opinion[1, 4], sum(opinion[2, 1:3]),
    opinion[2, 4]), nrow = 2, ncol = 2, byrow = TRUE)
dimnames(new_data) <- list(c("US born", "Not US born"), c("Wrong", "Right"))
names(dimnames(new_data)) <- c("Birth Country", "Opinion on Homosexuality")
print(new_data)
```

```
##              Opinion on Homosexuality
## Birth Country Wrong Right
##   US born     19982  6148
##   Not US born  1745   549
```

```
### Display the derived 2x2 contingency table in terms of proportions ###

row1_sum = sum(opinion[1, 1:5])
row2_sum = sum(opinion[2, 1:5])
prop_tab = matrix(c((sum(opinion[1, 1:3])/row1_sum), (opinion[1, 4]/row1_sum),
    (sum(opinion[2, 1:3])/row2_sum), (opinion[2, 4]/row2_sum)), nrow = 2, ncol = 2,
    byrow = TRUE)
dimnames(prop_tab) <- list(c("US born", "Not US born"), c("Wrong", "Right"))
names(dimnames(prop_tab)) <- c("Birth Country", "Opinion on Homosexuality")
print(prop_tab)
```

```
##              Opinion on Homosexuality
## Birth Country  Wrong  Right
##   US born     0.7647 0.2353
##   Not US born 0.7607 0.2393
```

```
### Statistical test 1 ###

prop.test(new_data)
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  new_data
## X-squared = 0.1689, df = 1, p-value = 0.6811
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.01440  0.02247
## sample estimates:
## prop 1 prop 2
## 0.7647 0.7607
```

```
### Statistical test 2 ###
table.test <- prop.test(new_data)
table.test$estimate
```

```
## prop 1 prop 2
## 0.7647 0.7607
```

```
odds <- table.test$estimate/(1 - table.test$estimate)
oddsratio <- odds[1]/odds[2]
print(oddsratio)
```

```
## prop 1
##  1.023
```

```
SE <- sqrt(sum(1/new_data))
logtheta.CI <- log(oddsratio) + c(-1, 1) * 1.96 * SE
print(exp(logtheta.CI))
```

```
## [1] 0.9252 1.1302
```

```
### Statistical test 3 ###
opinion1 = matrix(c(count1, count2, count3, count4, count5, count6, count7,
    count8), nrow = 2, ncol = 4)

chisq.test(opinion1)
```

```
##
##  Pearson's Chi-squared test
##
## data:  opinion1
## X-squared = 7.975, df = 3, p-value = 0.04654
```