duke_project.R

donso

Mon Apr 20 11:45:19 2015

```
# Part 1: Introduction
# Using the General Social Survey (GSS) data, do views on the country's level o
f investment (Too Much, About Right, Too Little) in Highway/Bridges (natroad) v
ary by Race (race)? How has this change over time?
# Investment in highway/bridges is very important for the safety of the populat
ion using those highway/bridges and it is important for our economy.


# Part 2: Data
# Cumulative file from 1972-2012 of survery responses from Americans on a wide
variety of topics from national spending priorities, marijuana use, crime and p
unishment, race relations, quality of life, confidence in institutions, and sex
ual behavior.
# Note - not all respondents answered all questions in each of the years.
# A case is an individuals (CASEID) response to a question in a given year.
# RACE and NATROAD are the two variables I will be examining. Both datasets ar
e categorical.
# The population of interest is the overall US population. I think the data ca
n be generalized for the full population since the vast majority of the populat
ion is a potential respondent and would have access/ability to take the study.
# Potential sources of bias could be the study is in English and would require
the respondent to be literate.
# This is an observational study and not an experiment. Data is collected throu
ghout the study and the data is collected in a way that doesn't impact with ho
w the data arise. Experiment randomly assigns subjects to groups which does no
t occur here. Causality can not be established.


# load in the data frame
load("/nfs/science/raleys/donso/sandbox/gss.Rdata")

# review the names of the full data frame
names(gss)
```

```
##   [1] "caseid"   "year"     "age"      "sex"      "race"     "hispanic"
##   [7] "uscitzn"  "educ"     "paeduc"   "maeduc"   "speduc"   "degree"
##  [13] "vetyears" "sei"      "wrkstat"  "wrkslf"   "marital"  "spwrksta"
##  [19] "sibs"     "childs"   "agekdbrn" "incom16"  "born"     "parborn"
##  [25] "granborn" "income06" "coninc"   "region"   "partyid"  "polviews"
##  [31] "relig"    "attend"   "natspac"  "natenvir" "natheal"  "natcity"
##  [37] "natcrime" "natdrug"  "nateduc"  "natrace"  "natarms"  "nataid"
##  [43] "natfare"  "natroad"  "natsoc"   "natmass"  "natpark"  "confinan"
##  [49] "conbus"   "conclerg" "coneduc"  "confed"   "conlabor" "conpress"
##  [55] "conmedic" "contv"    "conjudge" "consci"   "conlegis" "conarmy"
##  [61] "joblose"  "jobfind"  "satjob"   "richwork" "jobinc"   "jobsec"
##  [67] "jobhour"  "jobpromo" "jobmeans" "class"    "rank"     "satfin"
##  [73] "finalter" "finrela"  "unemp"    "govaid"   "getaid"   "union"
##  [79] "getahead" "parsol"   "kidssol"  "abdefect" "abnomore" "abhlth"
##  [85] "abpoor"   "abrape"   "absingle" "abany"    "pillok"   "sexeduc"
##  [91] "divlaw"   "premarsx" "teensex"  "xmarsex"  "homosex"  "suicide1"
##  [97] "suicide2" "suicide3" "suicide4" "fear"     "owngun"   "pistol"
## [103] "shotgun"  "rifle"    "news"     "tvhours"  "racdif1"  "racdif2"
## [109] "racdif3"  "racdif4"  "helppoor" "helpnot"  "helpsick" "helpblk"
```

```
# narrow down to just those variables that are needed for the analysis
gss_sub <- gss[,c("natroad","year","race","caseid")]

# review names of the subsetted data frame
names(gss_sub)
```

```
## [1] "natroad" "year"    "race"    "caseid"
```

```
# look at the structure of the data frame
str(gss_sub)
```

```
## 'data.frame':    57061 obs. of  4 variables:
##  $ natroad: Factor w/ 3 levels "Too Little","About Right",..: NA NA NA NA N
A NA NA NA NA NA ...
##  $ year   : int  1972 1972 1972 1972 1972 1972 1972 1972 1972 1972 ...
##  $ race   : Factor w/ 3 levels "White","Black",..: 1 1 1 1 1 1 1 1 2 2 ...
##  $ caseid : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
# quick summary of the variables in the data frame
summary(gss_sub)
```
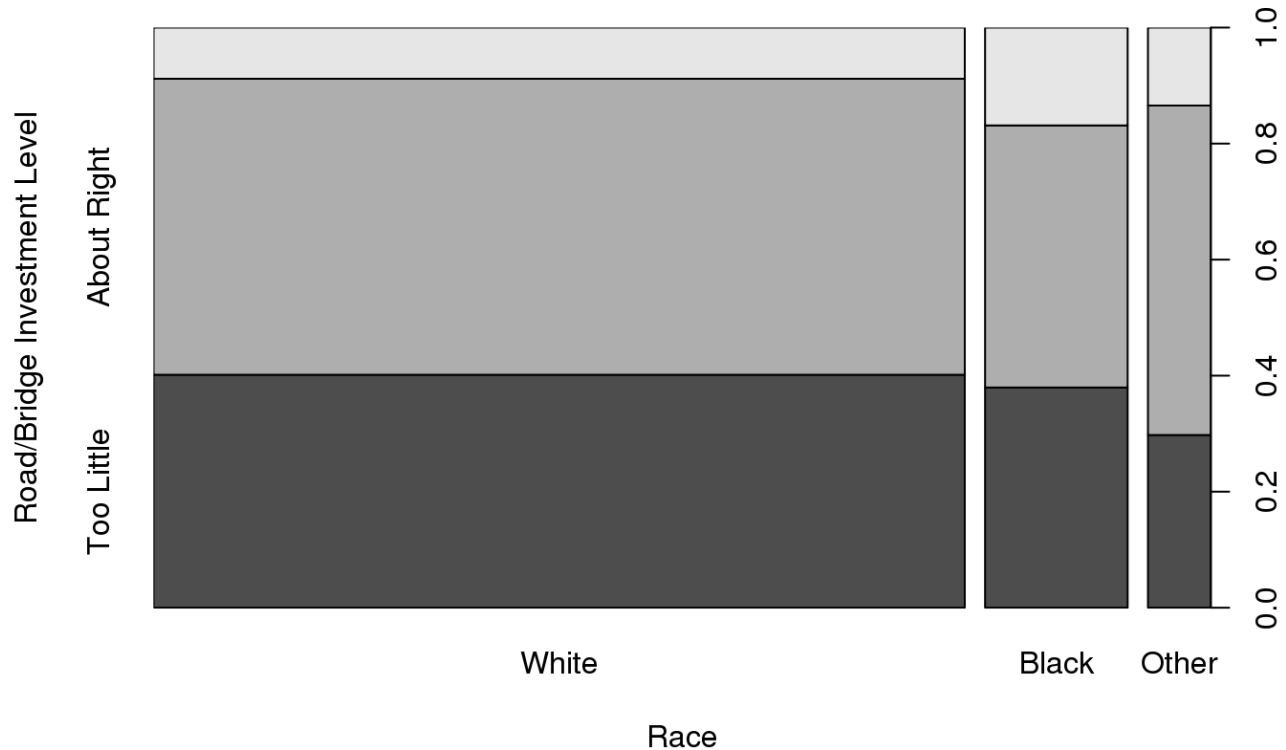
```
##          natroad            year           race           caseid
##   Too Little :14701   Min.   :1972   White:46350   Min.   :    1
##   About Right:18954   1st Qu.:1983   Black: 7926   1st Qu.:14266
##   Too Much   : 3840   Median :1993   Other: 2785   Median :28531
##   NA's       :19566   Mean   :1992                 Mean   :28531
##                       3rd Qu.:2002                 3rd Qu.:42796
##                       Max.   :2012                 Max.   :57061
```

```
# print the first 50 obs
head(gss_sub, n=50)
```

```
##     natroad year  race caseid
## 1     <NA> 1972 White      1
## 2     <NA> 1972 White      2
## 3     <NA> 1972 White      3
## 4     <NA> 1972 White      4
## 5     <NA> 1972 White      5
## 6     <NA> 1972 White      6
## 7     <NA> 1972 White      7
## 8     <NA> 1972 White      8
## 9     <NA> 1972 Black      9
## 10    <NA> 1972 Black     10
## 11    <NA> 1972 Black     11
## 12    <NA> 1972 Black     12
## 13    <NA> 1972 Black     13
## 14    <NA> 1972 Black     14
## 15    <NA> 1972 Black     15
## 16    <NA> 1972 Black     16
## 17    <NA> 1972 White     17
## 18    <NA> 1972 White     18
## 19    <NA> 1972 White     19
## 20    <NA> 1972 White     20
## 21    <NA> 1972 White     21
## 22    <NA> 1972 White     22
## 23    <NA> 1972 White     23
## 24    <NA> 1972 White     24
## 25    <NA> 1972 White     25
## 26    <NA> 1972 White     26
## 27    <NA> 1972 White     27
## 28    <NA> 1972 White     28
## 29    <NA> 1972 White     29
## 30    <NA> 1972 White     30
## 31    <NA> 1972 White     31
## 32    <NA> 1972 White     32
## 33    <NA> 1972 White     33
## 34    <NA> 1972 White     34
## 35    <NA> 1972 White     35
## 36    <NA> 1972 White     36
## 37    <NA> 1972 White     37
## 38    <NA> 1972 White     38
## 39    <NA> 1972 White     39
## 40    <NA> 1972 White     40
## 41    <NA> 1972 White     41
## 42    <NA> 1972 White     42
## 43    <NA> 1972 White     43
## 44    <NA> 1972 White     44
## 45    <NA> 1972 White     45
## 46    <NA> 1972 White     46
## 47    <NA> 1972 White     47
```

```
## 48    <NA> 1972 White    48
## 49    <NA> 1972 White    49
## 50    <NA> 1972 White    50
```

```
plot(gss_sub$race,gss_sub$natroad,xlab="Race",ylab="Road/Bridge Investment Leve
l")
```



```
# References
# Smith, Tom W., Michael Hout, and Peter V. Marsden. General Social Survey, 197
2-2012 [Cumulative File]. ICPSR34802-v1. Storrs, CT: Roper Center for Public Op
inion Research, University of Connecticut /Ann Arbor, MI: Inter-university Cons
ortium for Political and Social Research [distributors], 2013-09-11. doi:10.388
6/ICPSR34802.v1
# Persistent URL: http://doi.org/10.3886/ICPSR34802.v1

# Appendix
# print the first 50 obs
head(gss_sub, n=50)
```

```
##    natroad year  race caseid
## 1     <NA> 1972 White      1
## 2     <NA> 1972 White      2
## 3     <NA> 1972 White      3
## 4     <NA> 1972 White      4
## 5     <NA> 1972 White      5
## 6     <NA> 1972 White      6
## 7     <NA> 1972 White      7
## 8     <NA> 1972 White      8
## 9     <NA> 1972 Black      9
## 10    <NA> 1972 Black     10
## 11    <NA> 1972 Black     11
## 12    <NA> 1972 Black     12
## 13    <NA> 1972 Black     13
## 14    <NA> 1972 Black     14
## 15    <NA> 1972 Black     15
## 16    <NA> 1972 Black     16
## 17    <NA> 1972 White     17
## 18    <NA> 1972 White     18
## 19    <NA> 1972 White     19
## 20    <NA> 1972 White     20
## 21    <NA> 1972 White     21
## 22    <NA> 1972 White     22
## 23    <NA> 1972 White     23
## 24    <NA> 1972 White     24
## 25    <NA> 1972 White     25
## 26    <NA> 1972 White     26
## 27    <NA> 1972 White     27
## 28    <NA> 1972 White     28
## 29    <NA> 1972 White     29
## 30    <NA> 1972 White     30
## 31    <NA> 1972 White     31
## 32    <NA> 1972 White     32
## 33    <NA> 1972 White     33
## 34    <NA> 1972 White     34
## 35    <NA> 1972 White     35
## 36    <NA> 1972 White     36
## 37    <NA> 1972 White     37
## 38    <NA> 1972 White     38
## 39    <NA> 1972 White     39
## 40    <NA> 1972 White     40
## 41    <NA> 1972 White     41
## 42    <NA> 1972 White     42
## 43    <NA> 1972 White     43
## 44    <NA> 1972 White     44
## 45    <NA> 1972 White     45
## 46    <NA> 1972 White     46
## 47    <NA> 1972 White     47
```

```
## 48    <NA> 1972 White      48
## 49    <NA> 1972 White      49
## 50    <NA> 1972 White      50
```