DATA_110_FINAL

Juan J.

May 2019

Set working directory and use bring up haven to read an spss file.

```
setwd("C:\\Users\\Juan Nunez\\Desktop\\DATA_110")
##install.packages("haven")
library(haven)
```

Open, load, and save data.

```
#GSS_2018 <- read_spss("C:\\Users\\Juan Nunez\\Desktop\\DATA_110\\GSS18.sav")
#save(GSS_2018, file = "GSS2018.Rda")
```

Write the data to .csv form.

```
#write.csv(GSS_2018, file = "GSS_18.csv",row.names=FALSE, na="")
```

Bring up packages to analyze data.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
library(ggplot2)
library(tidyr)
```

Use the .csv to bring up the data because haven makes it a strange file type.

```
gss <- read_csv("GSS_18.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   AWAY6 = col_logical(),
##   AWAY7 = col_logical(),
##   COJew = col_logical(),
##   MHP3R2 = col_logical(),
##   WHERE7 = col_logical()
## )
## See spec(...) for full column specifications.
## Warning: 4 parsing failures.
##  row    col           expected actual         file
## 1512 COJew  1/0/T/F/TRUE/FALSE     4  'GSS_18.csv'
```

```
## 1677 MHP3R2 1/0/T/F/TRUE/FALSE     11 'GSS_18.csv'
## 1794 COJew  1/0/T/F/TRUE/FALSE     2  'GSS_18.csv'
## 2318 MHP3R2 1/0/T/F/TRUE/FALSE     13 'GSS_18.csv'
```

Let's see the dimension top of this data.

```r
dim(gss)
```

```
## [1] 2348 1064
```

```r
head(gss)
```

```
## # A tibble: 6 x 1,064
##   ABANY ABDEFECT ABFELEGL ABHELP1 ABHELP2 ABHELP3 ABHELP4 ABHLTH ABINSPAY
##   <dbl>    <dbl>    <dbl>   <dbl>   <dbl>   <dbl>   <dbl>  <dbl>    <dbl>
## 1     2        1       NA       1       1       1       1      1        1
## 2     1        1        3       2       2       2       2      1        2
## 3    NA       NA       NA       1       2       1       1     NA        2
## 4    NA       NA        1       1       1       1       1     NA        1
## 5     2        1       NA       2       2       2       1      1        2
## 6     1        1        1       1       1       1       1      1        1
## # ... with 1,055 more variables: ABMEDGOV1 <dbl>, ABMEDGOV2 <dbl>,
## #   ABMELEGL <dbl>, ABMORAL <dbl>, ABNOMORE <dbl>, ABPOOR <dbl>,
## #   ABPOORW <dbl>, ABRAPE <dbl>, ABSINGLE <dbl>, ABSTATE1 <dbl>,
## #   ABSTATE2 <dbl>, ACQNTSEX <dbl>, ACTSSOC <dbl>, ADULTS <dbl>,
## #   ADVFRONT <dbl>, AFFRMACT <dbl>, AFRAIDOF <dbl>, AFTERLIF <dbl>,
## #   AGE <dbl>, AGED <dbl>, AGEKDBRN <dbl>, ANCESTRS <dbl>, ARTHRTIS <dbl>,
## #   ASTROLGY <dbl>, ASTROSCI <dbl>, ATHEISTS <dbl>, ATTEND <dbl>,
## #   ATTEND12 <dbl>, ATTENDMA <dbl>, ATTENDPA <dbl>, AWAY1 <dbl>,
## #   AWAY11 <dbl>, AWAY2 <dbl>, AWAY3 <dbl>, AWAY4 <dbl>, AWAY5 <dbl>,
## #   AWAY6 <lgl>, AWAY7 <lgl>, BABIES <dbl>, BACKPAIN <dbl>, BALLOT <dbl>,
## #   BALNEG <dbl>, BALPOS <dbl>, BEFAIR <dbl>, BETRLANG <dbl>, BIBLE <dbl>,
## #   BIGBANG <dbl>, BIGBANG1 <dbl>, BIGBANG2 <dbl>, BIRD <dbl>,
## #   BIRDB4 <dbl>, BORN <dbl>, BOYORGRL <dbl>, BREAKDWN <dbl>,
## #   BUDDHSTS <dbl>, BUYESOP <dbl>, BUYVALUE <dbl>, CANTRUST <dbl>,
## #   CAPPUN <dbl>, CAT <dbl>, CATB4 <dbl>, CHARACTR <dbl>, CHEMGEN <dbl>,
## #   CHILDS <dbl>, CHLDIDEL <dbl>, CHRISTNS <dbl>, CHURHPOW <dbl>,
## #   CLASS <dbl>, CLERGVTE <dbl>, CLOSETO1 <dbl>, CLOSETO2 <dbl>,
## #   CLOSETO3 <dbl>, CLOSETO4 <dbl>, CLOSETO5 <dbl>, CNTCTFAM <dbl>,
## #   CNTCTFRD <dbl>, CNTCTKID <dbl>, CNTCTPAR <dbl>, CNTCTSIB <dbl>,
## #   CODEG <dbl>, CODEN <dbl>, COEDUC <dbl>, COEVWORK <dbl>, COFUND <dbl>,
## #   COHORT <dbl>, COHRS1 <dbl>, COHRS2 <dbl>, COIND10 <dbl>,
## #   COISCO08 <dbl>, COJew <lgl>, COLATH <dbl>, COLCOM <dbl>,
## #   COLDEG1 <dbl>, COLHOMO <dbl>, COLMIL <dbl>, COLMSLM <dbl>,
## #   COLRAC <dbl>, COLSCI <dbl>, COLSCINM <dbl>, COMFORT <dbl>, ...
```

Below is a question that asks respondents if they have ever smoked crack cocaine. 1 is yes, 2 is no.

```r
table(gss$EVCRACK)
```

```
##
##    1    2
##   83 1310
```

Below is a question that asks respondents if they think marijuana should be made legal. 1 is should and 2 is should not.

```r
table(gss$GRASS)
```

```
##
##   1   2
## 938 509
```

I create a version of the variable above but with the correct labels.

```r
gss$weed <- gss$GRASS
gss$weed[gss$weed== 1] <- "Should be"
gss$weed[gss$weed== 2] <- "Should not be"
```

Let's see if it worked.

```r
table(gss$weed)
```

```
##
##     Should be Should not be
##           938           509
```

Let's see the class of the variables

```r
class(gss$GRASS)
```

```
## [1] "numeric"
```

```r
class(gss$weed)
```

```
## [1] "character"
```

I use the package psych to look at the data.

```r
library(psych)
```

```
## Warning: package 'psych' was built under R version 3.5.3
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

Let's see the descriptive statistics for the variable POLVIEWS. This variable is a 7 point scale that asks respondents to rate themselves on a scale from 1-7.

1. Extremely liberal
2. Liberal
3. Slightly liberal
4. Moderate, middle of the road
5. Slightly conservative
6. Conservative
7. Extremely conservative

```r
describe(gss$POLVIEWS)
```

```
##    vars    n mean  sd median trimmed  mad min max range skew kurtosis   se
## X1    1 2247 4.05 1.5      4    4.07 1.48   1   7     6 -0.1    -0.48 0.03
```

It's interesting to see that the mean is almost at the middle number. Let's look at the table of this variable.

```
table(gss$POLVIEWS)
```

```
##
##   1   2   3   4   5   6   7
## 122 278 256 855 283 354  99
```

I see tthat most people consider themselves moderate. It's also interesting the second largest category is conservative. Let's look at age. Perhaps age and whether someone is conservative are correlated.

```
class(gss$POLVIEWS)
```

```
## [1] "numeric"
```

We should turn this variable to a factor.

```
gss$pol <- gss$POLVIEWS
gss$pol[gss$pol== 1] <- "Extremely liberal"
gss$pol[gss$pol== 2] <- "Liberal"
gss$pol[gss$pol== 3] <- "Slightly liberal"
gss$pol[gss$pol== 4] <- "Moderate"
gss$pol[gss$pol== 5] <- "Slightly conservative"
gss$pol[gss$pol== 6] <- "Conservative"
gss$pol[gss$pol== 7] <- "Extremely conservative"
```

Let's see if it worked.

```
table(gss$pol)
```

```
##
##           Conservative Extremely conservative      Extremely liberal
##                    354                     99                    122
##                Liberal               Moderate  Slightly conservative
##                    278                    855                    283
##       Slightly liberal
##                    256
```

```
class(gss$pol)
```

```
## [1] "character"
```

It looks like we need to reoder this variable.

```
gss$pol = factor(gss$pol,levels = c("Extremely liberal","Liberal","Slightly liberal","Moderate","Slightl
```

Let's compare it to the original.

```
table(gss$pol)
```

```
##
##      Extremely liberal                Liberal       Slightly liberal
##                    122                    278                    256
##               Moderate  Slightly conservative           Conservative
##                    855                    283                    354
## Extremely conservative
##                     99
```

```
table(gss$POLVIEWS)
```

```
##
```

```
##   1   2   3   4   5   6   7
## 122 278 256 855 283 354  99
```

Let's look at age. It doesn't seem to be skewed.

```
describe(gss$AGE)
```

```
##    vars    n  mean    sd median trimmed   mad min max range skew kurtosis
## X1    1 2341 48.97 18.06     48    48.4 22.24  18  89    71 0.22    -0.91
##      se
## X1 0.37
```

We see the mean age is 48.97. Let's see the gender. 1 is male and 2 is female.

```
table(gss$SEX)
```

```
##
##    1    2
## 1052 1296
```

We see that there are more females than males. Let's recode this variable.

```
gss$gender <- gss$SEX
gss$gender[gss$gender== 1] <- "Male"
gss$gender[gss$gender== 2] <- "Female"

table(gss$gender)
```

```
##
## Female   Male
##   1296   1052
```

Let's now look at education.

```
table(gss$EDUC)
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
##   4   2   4  10   5   3  20   8  35  51  65  95 657 183 313 127 430  97
##  18  19  20
## 119  45  72
```

```
describe(gss$EDUC)
```

```
##    vars    n  mean   sd median trimmed  mad min max range  skew kurtosis
## X1    1 2345 13.73 2.97     14   13.78 2.97   0  20    20 -0.49     1.62
##      se
## X1 0.06
```

The mean years of school are 13. Let's now see marital status.

```
table(gss$MARITAL)
```

```
##
##   1   2   3   4   5
## 998 200 403  75 670
```

We clearly need to recode this variable. I am going to make a dummy where participants will be divided between those that are married vs those that aren't.

Married

Widowed

Divorced

Separated

Never married

```
gss$married <- gss$MARITAL
gss$married[gss$married== 1] <- "Married"
gss$married[gss$married== 2] <- "Not married"
gss$married[gss$married== 3] <- "Not married"
gss$married[gss$married== 4] <- "Not married"
gss$married[gss$married== 5] <- "Not married"

table(gss$married)
```

```
##
##     Married Not married
##         998        1348
```

We see that most are not married even those there are many that are. The following questions asks respondents at what age they had the first kid.

```
table(gss$AGEKDBRN)
```

```
##
##   12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29
##    1   2   6  20  35  70  96 131 126 142 113 101 115 120  63  67  87  59
##   30  31  32  33  34  35  36  37  38  39  40  41  42  43  44  45  46  47
##   83  37  43  38  20  28   9  10  10   7  10   4   3   2   1   1   1   3
##   50  51
##    1   1
```

```
describe(gss$AGEKDBRN)
```

```
##    vars    n mean   sd median trimmed  mad min max range skew kurtosis
## X1    1 1666 24.3 5.74     23   23.82 5.93  12  51    39 0.85     0.85
##      se
## X1 0.14
```

The mean age people had kids was approximately 24 years old. The following question asks respondents whether they think we spend too little, about right, or too much on supporting scientific research.

```
table(gss$NATSCI)
```

```
##
##    1    2    3
##  986 1026  197
```

```
describe(gss$NATSCI)
```

```
##    vars    n mean   sd median trimmed  mad min max range skew kurtosis
## X1    1 2209 1.64 0.64      2    1.57 1.48   1   3     2 0.48    -0.68
##      se
## X1 0.01
```

I guess the mean would be between too little and about right. Now let's see the siblings.

```
table(gss$SIBS)
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
##  95 472 483 384 244 176 147 125  76  44  34  19  13  14   5   2   4   1
##  19  21  23  24  25
##   1   1   1   1   1
```

```
describe(gss$SIBS)
```

```
##    vars    n mean   sd median trimmed  mad min max range skew kurtosis
## X1    1 2343 3.58 2.86      3    3.17 2.97   0  25    25 1.75     5.36
##      se
## X1 0.06
```

Most people, interestingly have more than 3 siblings. Now let's see what the respondents think about how much we spend on developing alternative energy resources.

```
table(gss$NATENRGY)
```

```
##
##    1    2    3
## 1277  788  169
```

Most think we spend too little. Let's see about improving and protecting the environment.

```
table(gss$NATENVIR)
```

```
##
##   1   2   3
## 790 284  83
```

Most think too little. The following variable asks respondents if they are very happy, pretty happy, or not too happy these days.

```
table(gss$HAPPY)
```

```
##
##    1    2    3
##  701 1307  336
```

Most are pretty happy. The following question asks respondents if they have views of nature when they are home.

1. Strongly agree
2. Somewhat agree
3. Somewhat disagree, or
4. Strongly disagree?

```
table(gss$NATVIEWS)
```

```
##
##   1   2   3   4
## 781 275  58  32
```

Now I am going to use dplyr to subset this data. Below I use the pipping and select from the dplyr package.

```
df <- gss %>%
  select(NATVIEWS, HAPPY, NATENVIR, NATENRGY, SIBS, NATSCI, AGEKDBRN, married, EDUC, gender, AGE, pol, w
```

Let's look at the top of our subsetted data.

```r
head(df)
```

```
## # A tibble: 6 x 15
##   NATVIEWS HAPPY NATENVIR NATENRGY  SIBS NATSCI AGEKDBRN married  EDUC
##      <dbl> <dbl>    <dbl>    <dbl> <dbl>  <dbl>    <dbl> <chr>   <dbl>
## 1       NA     2        2        1     4      2       NA Not ma~    14
## 2        1     1       NA        2     4      1       21 Not ma~    10
## 3       NA     1        1        1     2     NA       35 Married    16
## 4        1     1       NA        1     3      1       32 Married    16
## 5        1     2       NA        1     3      1       NA Not ma~    18
## 6       NA     3       NA        2     1      2       27 Not ma~    16
## # ... with 6 more variables: gender <chr>, AGE <dbl>, pol <fct>,
## #   weed <chr>, EVCRACK <dbl>, POLVIEWS <dbl>
```

Everything looks in order. Let's look at our data one more time to see if there is more changing we need to do.

```r
describe(df)
```

```
## Warning in describe(df): NAs introduced by coercion

## Warning in describe(df): NAs introduced by coercion

## Warning in describe(df): NAs introduced by coercion

## Warning in FUN(newX[, i], ...): no non-missing arguments to min; returning
## Inf

## Warning in FUN(newX[, i], ...): no non-missing arguments to min; returning
## Inf

## Warning in FUN(newX[, i], ...): no non-missing arguments to min; returning
## Inf

## Warning in FUN(newX[, i], ...): no non-missing arguments to max; returning
## -Inf

## Warning in FUN(newX[, i], ...): no non-missing arguments to max; returning
## -Inf

## Warning in FUN(newX[, i], ...): no non-missing arguments to max; returning
## -Inf
```

```
##          vars    n  mean    sd median trimmed   mad min  max range  skew
## NATVIEWS    1 1146  1.42  0.72      1    1.27  0.00   1    4     3  1.81
## HAPPY       2 2344  1.84  0.65      2    1.81  0.00   1    3     2  0.16
## NATENVIR    3 1157  1.39  0.62      1    1.27  0.00   1    3     2  1.34
## NATENRGY    4 2234  1.50  0.63      1    1.41  0.00   1    3     2  0.88
## SIBS        5 2343  3.58  2.86      3    3.17  2.97   0   25    25  1.75
## NATSCI      6 2209  1.64  0.64      2    1.57  1.48   1    3     2  0.48
## AGEKDBRN    7 1666 24.30  5.74     23   23.82  5.93  12   51    39  0.85
## married*    8 2346   NaN    NA     NA     NaN    NA Inf -Inf  -Inf    NA
## EDUC        9 2345 13.73  2.97     14   13.78  2.97   0   20    20 -0.49
## gender*    10 2348   NaN    NA     NA     NaN    NA Inf -Inf  -Inf    NA
## AGE        11 2341 48.97 18.06     48   48.40 22.24  18   89    71  0.22
## pol*       12 2247  4.05  1.50      4    4.07  1.48   1    7     6 -0.10
## weed*      13 1447   NaN    NA     NA     NaN    NA Inf -Inf  -Inf    NA
```

```
## EVCRACK    14 1393  1.94  0.24      2    2.00  0.00   1    2     1 -3.72
## POLVIEWS   15 2247  4.05  1.50      4    4.07  1.48   1    7     6 -0.10
##           kurtosis   se
## NATVIEWS      3.02 0.02
## HAPPY        -0.67 0.01
## NATENVIR      0.66 0.02
## NATENRGY     -0.29 0.01
## SIBS          5.36 0.06
## NATSCI       -0.68 0.01
## AGEKDBRN      0.85 0.14
## married*        NA   NA
## EDUC          1.62 0.06
## gender*         NA   NA
## AGE          -0.91 0.37
## pol*         -0.48 0.03
## weed*           NA   NA
## EVCRACK      11.83 0.01
## POLVIEWS     -0.48 0.03
```

The only variable I would change is evcrack.

```
df$EVCRACK[df$EVCRACK == 1] <- "Yes"
df$EVCRACK[df$EVCRACK == 2] <- "No"
table(df$EVCRACK)
```

```
##
##   No  Yes
## 1310   83
```

Let's see if it worked.

```
describe(df)
```

```
## Warning in describe(df): NAs introduced by coercion

## Warning in describe(df): NAs introduced by coercion

## Warning in describe(df): NAs introduced by coercion

## Warning in describe(df): NAs introduced by coercion

## Warning in FUN(newX[, i], ...): no non-missing arguments to min; returning
## Inf

## Warning in FUN(newX[, i], ...): no non-missing arguments to min; returning
## Inf

## Warning in FUN(newX[, i], ...): no non-missing arguments to min; returning
## Inf

## Warning in FUN(newX[, i], ...): no non-missing arguments to min; returning
## Inf

## Warning in FUN(newX[, i], ...): no non-missing arguments to max; returning
## -Inf

## Warning in FUN(newX[, i], ...): no non-missing arguments to max; returning
## -Inf
```

```
## Warning in FUN(newX[, i], ...): no non-missing arguments to max; returning
## -Inf

## Warning in FUN(newX[, i], ...): no non-missing arguments to max; returning
## -Inf
```

```
##           vars    n  mean    sd median trimmed   mad min  max range  skew
## NATVIEWS     1 1146  1.42  0.72      1    1.27  0.00   1    4     3  1.81
## HAPPY        2 2344  1.84  0.65      2    1.81  0.00   1    3     2  0.16
## NATENVIR     3 1157  1.39  0.62      1    1.27  0.00   1    3     2  1.34
## NATENRGY     4 2234  1.50  0.63      1    1.41  0.00   1    3     2  0.88
## SIBS         5 2343  3.58  2.86      3    3.17  2.97   0   25    25  1.75
## NATSCI       6 2209  1.64  0.64      2    1.57  1.48   1    3     2  0.48
## AGEKDBRN     7 1666 24.30  5.74     23   23.82  5.93  12   51    39  0.85
## married*     8 2346   NaN    NA     NA     NaN    NA Inf -Inf  -Inf    NA
## EDUC         9 2345 13.73  2.97     14   13.78  2.97   0   20    20 -0.49
## gender*     10 2348   NaN    NA     NA     NaN    NA Inf -Inf  -Inf    NA
## AGE         11 2341 48.97 18.06     48   48.40 22.24  18   89    71  0.22
## pol*        12 2247  4.05  1.50      4    4.07  1.48   1    7     6 -0.10
## weed*       13 1447   NaN    NA     NA     NaN    NA Inf -Inf  -Inf    NA
## EVCRACK*    14 1393   NaN    NA     NA     NaN    NA Inf -Inf  -Inf    NA
## POLVIEWS    15 2247  4.05  1.50      4    4.07  1.48   1    7     6 -0.10
##           kurtosis   se
## NATVIEWS      3.02 0.02
## HAPPY        -0.67 0.01
## NATENVIR      0.66 0.02
## NATENRGY     -0.29 0.01
## SIBS          5.36 0.06
## NATSCI       -0.68 0.01
## AGEKDBRN      0.85 0.14
## married*        NA   NA
## EDUC          1.62 0.06
## gender*         NA   NA
## AGE          -0.91 0.37
## pol*         -0.48 0.03
## weed*           NA   NA
## EVCRACK*        NA   NA
## POLVIEWS     -0.48 0.03
```

For the numerical variables we get the descriptive statistics. For the factor variables we do not. Let's save the data to csv.

```r
write.csv(df, "gss_sub.csv", row.names = FALSE, na="")
```

Below is the link to a couple of Tableau vizzes.

https://public.tableau.com/profile/juan.n.#!/vizhome/pol_views/mar_pol?publish=yes

Now let's see if there are any meaningfull relationships between our variables.

p <- gapminder %>% filter(year==1977) %>% ggplot(aes(gdpPercap, lifeExp, size = pop, color=continent)) + geom_point() + scale_x_log10() + theme_bw()

ggplotly(p)

```r
library(plotly)
```

```
## Warning: package 'plotly' was built under R version 3.5.3

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout
```

```
#install.packages("gapminder")
library(gapminder)
```

```
## Warning: package 'gapminder' was built under R version 3.5.3
```

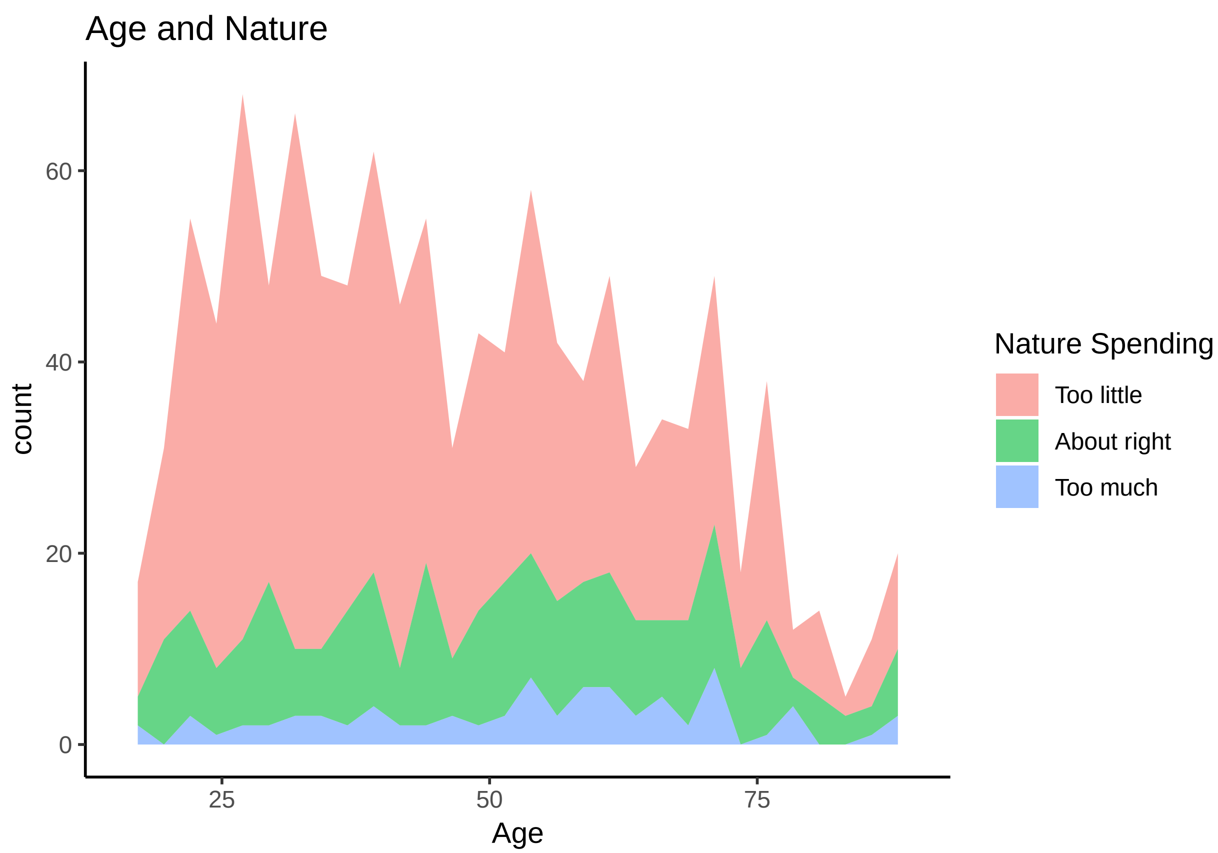
Below is a plot

```
plot <- df %>% filter(!is.na(NATENVIR)) %>%
  ggplot( aes(x = AGE)) + geom_area(aes(fill = as.factor(NATENVIR)),stat = "bin", alpha =0.6) + theme_cl

plot
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 3 rows containing non-finite values (stat_bin).
```



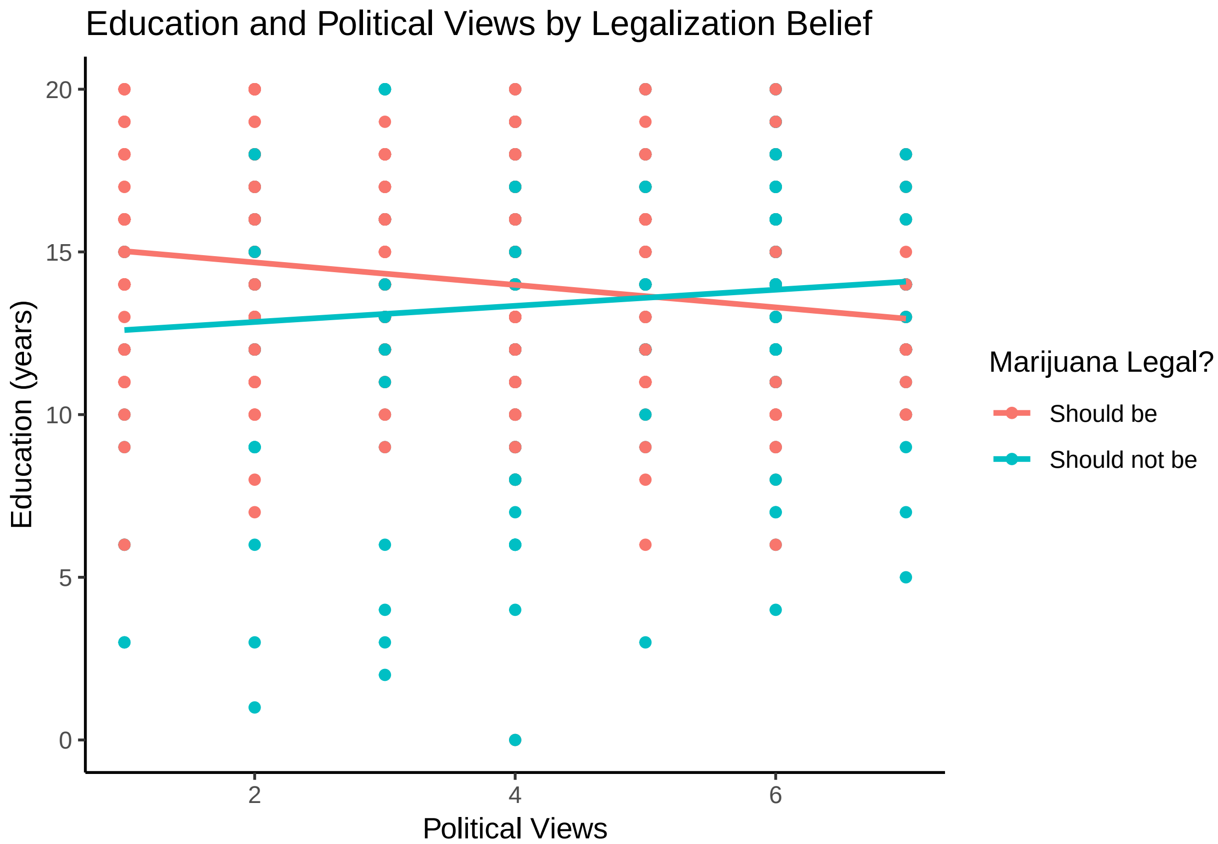

```
#ggplotly(plot)
```

Below is a graph with education and political views. The cases are split between those that believe marijuana should and should not be legal.

```
df %>% filter(!is.na(weed)) %>%
  ggplot() +
  aes(x=POLVIEWS, y=EDUC, color=weed) +
  geom_point() +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)+
  labs(title = "Education and Political Views by Legalization Belief", x = "Political Views", y= "Educa
```

```
## Warning: Removed 61 rows containing non-finite values (stat_smooth).

## Warning: Removed 61 rows containing missing values (geom_point).
```



We found an interaction!! Let's test it using linear regression

```
model1 <- lm(POLVIEWS ~   EDUC + weed + AGE + NATENVIR + SIBS + married + gender + NATVIEWS + HAPPY + E
summary(model1)
```

```
##
## Call:
## lm(formula = POLVIEWS ~ EDUC + weed + AGE + NATENVIR + SIBS +
##     married + gender + NATVIEWS + HAPPY + EVCRACK + NATENRGY +
##     NATSCI, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4899 -0.7855  0.0201  0.8304  3.9642
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.61576    0.63834   4.098 5.44e-05 ***
## EDUC                 -0.04536    0.02802  -1.619  0.10657
## weedShould not be     0.49028    0.17905   2.738  0.00657 **
## AGE                   0.00600    0.00489   1.227  0.22081
## NATENVIR              0.70412    0.13904   5.064 7.36e-07 ***
## SIBS                  0.05964    0.03155   1.890  0.05971 .
## marriedNot married -0.07831    0.17568  -0.446  0.65613
## genderMale           -0.14514    0.16702  -0.869  0.38557
## NATVIEWS              0.01600    0.10898   0.147  0.88337
## HAPPY                -0.03697    0.11875  -0.311  0.75579
```

```
## EVCRACKYes         -0.08721    0.28391  -0.307  0.75894
## NATENRGY            0.31366    0.12886   2.434  0.01554 *
## NATSCI              0.06348    0.12511   0.507  0.61227
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.325 on 286 degrees of freedom
##   (2049 observations deleted due to missingness)
## Multiple R-squared:  0.2433, Adjusted R-squared:  0.2115
## F-statistic: 7.662 on 12 and 286 DF,  p-value: 2.563e-12
```

There were 4 signficant variables. I'll do a smaller version of the model

```r
model2 <- lm(POLVIEWS ~   EDUC + weed + AGE + NATENVIR + NATENRGY , data= df)
summary(model2)
```

```
##
## Call:
## lm(formula = POLVIEWS ~ EDUC + weed + AGE + NATENVIR + NATENRGY,
##     data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1161 -0.8820  0.1170  0.8897  3.6472
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         2.786795   0.329085   8.468  < 2e-16 ***
## EDUC               -0.032079   0.018146  -1.768 0.077559 .
## weedShould not be   0.514832   0.120259   4.281 2.14e-05 ***
## AGE                 0.006743   0.003249   2.075 0.038336 *
## NATENVIR            0.564389   0.096871   5.826 8.91e-09 ***
## NATENRGY            0.310958   0.092378   3.366 0.000807 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.363 on 652 degrees of freedom
##   (1690 observations deleted due to missingness)
## Multiple R-squared:  0.1761, Adjusted R-squared:  0.1698
## F-statistic: 27.88 on 5 and 652 DF,  p-value: < 2.2e-16
```

```r
table(df$POLVIEWS)
```

```
##
##   1   2   3   4   5   6   7
## 122 278 256 855 283 354  99
```

We see that all of the variables are significant.

For 1 year of education increase, a person is -0.03 more liberal.

Compared to those that think marijuana should be legal, those that think it shouldn't be legal were 0.51 more conservative.

For 1 year in age increase, conservatism increases by 0.007.

For a 1 unit increase in belief we spend too much on the environment, conservatism increases by 0.56.

For a 1 unit increase in belief we spend too much on finding sustainable energy resources, conservatism increases by 0.31.

Let's see about our interaction we obeserved earlier.

```r
model3 <- lm(POLVIEWS ~   EDUC * weed +  AGE + NATENVIR + NATENRGY, data= df)
summary(model3)
```

```
##
## Call:
## lm(formula = POLVIEWS ~ EDUC * weed + AGE + NATENVIR + NATENRGY,
##     data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6695 -0.9297  0.0957  0.9136  3.8662
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             3.811625   0.403768   9.440  < 2e-16 ***
## EDUC                   -0.105436   0.024810  -4.250 2.45e-05 ***
## weedShould not be      -1.580083   0.504439  -3.132 0.001812 **
## AGE                     0.006856   0.003207   2.138 0.032899 *
## NATENVIR                0.556777   0.095631   5.822 9.13e-09 ***
## NATENRGY                0.317816   0.091193   3.485 0.000525 ***
## EDUC:weedShould not be  0.152497   0.035689   4.273 2.22e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.345 on 651 degrees of freedom
##   (1690 observations deleted due to missingness)
## Multiple R-squared:  0.1986, Adjusted R-squared:  0.1912
## F-statistic: 26.89 on 6 and 651 DF,  p-value: < 2.2e-16
```

We see that it turns out that it is signficant! The way to explain this interaction is a bit tricky so we'll give it shot.

First we see that the R^2 increased so that means that the interaction adds to the regression.

The effect of education for those that think marijuana should be legal is -0.105436

The ADDITIONAL effect of education for those that think marijuana shouldn't be legal is 0.152497.

This means that the effect of education for those that think marijuana shouldn't be legal is -0.105436+0.152497 = 0.047061

Thus, for those that think marijuana should be legal, a 1 unit increase in education results in -0.11 decrease in conservatism.

For those that think marijuana shouldn't be legal, a 1 unit increase in education results in a 0.05 increase in conservatism.