Job satisfaction and Level of Education.

Introduction:

Are people with advanced degree more satisfied with their job. This is an important question, as it helps one to decide whether to go for that expensive college Degree or high school education is good enough. Is the time, money, effort spent in earning advanced degree pays off in getting a job with higher satisfaction.Job satisfaction sometimes acts as a motivation factor for the student to go for advanced degree, hence it is important to study the relation between the two variable.

This research will help potential students to make that critical decision.

Data:

The data used in this project is obtained from General social survey(GSS). GSS has been conducting surveys from 1972-2012 to monitor the American societal change and to capture the trend changes that happen over the years. Survey consists of specific set of questions that has been repeated over the years.

The survey has been conducted by computer-assisted personal interview (CAPI), face-to-face interview, telephone interview and the respondents for this survey are randomly selected. Since we are merely observing the responses and monitor them over years, it is an observational study. Hence the result of this study is generalizable to US population.

The two categorical variable which are used in this study are satjob(response variable) and degree(explanatory variable). We cannot deduce any causal relationship between these two variables because we are not doing any random assignment and hence correlation between the two variable might due to some third confounding variable which is not recorded.

The data set has 57061 records and 114 variables.

```
dim(gss)
```

```
## [1] 57061   114
```

Exploratory data analysis:

To begin with we will first summarize the two variable.

degree - It captures the level of education of each respondent and it has 5 levels.

```
attach(gss)
summary(degree)
```

```
## Lt High School    High School Junior College       Bachelor       Graduate
##          11822          29287           3070           8002           3870
##           NA's
##           1010
```

satjob - Tells us about the job satisfaction and it has 4 levels.

```
summary(satjob)
```

```
##    Very Satisfied    Mod. Satisfied   A Little Dissat Very Dissatisfied
##             19717             15736              4109              1715
##              NA's
##             15784
```

from the summary of these two variable we can see that there are missing values.First, we will convert these two variable in to data frame and then remove those missing values.

```
data <- data.frame(degree, satjob)
head(data)
```

```
##           degree          satjob
## 1       Bachelor A Little Dissat
## 2 Lt High School            <NA>
## 3    High School  Mod. Satisfied
## 4       Bachelor  Very Satisfied
## 5    High School            <NA>
## 6    High School  Mod. Satisfied
```

```
gooddata <- na.omit(data)
head(gooddata)
```

```
##        degree          satjob
## 1    Bachelor A Little Dissat
## 3 High School  Mod. Satisfied
## 4    Bachelor  Very Satisfied
## 6 High School  Mod. Satisfied
## 7 High School  Very Satisfied
## 8    Bachelor A Little Dissat
```

```
tail(gooddata)
```

```
##               degree          satjob
## 57056 Lt High School  Mod. Satisfied
## 57057       Bachelor  Mod. Satisfied
## 57058    High School  Mod. Satisfied
## 57059    High School  Very Satisfied
## 57060    High School  Mod. Satisfied
## 57061    High School A Little Dissat
```

we can see that the missing values has been removed from the dataframe.
using the summary function to compare the data with NA and without NA's

```
summary(data)
```

```
##             degree                    satjob
##  Lt High School:11822   Very Satisfied   :19717
##  High School   :29287   Mod. Satisfied   :15736
##  Junior College: 3070   A Little Dissat  : 4109
##  Bachelor      : 8002   Very Dissatisfied: 1715
##  Graduate      : 3870   NA's             :15784
##  NA's          : 1010
```

```
nrow(data)
```

```
## [1] 57061
```

```
summary(gooddata)
```

```
##             degree                    satjob
##  Lt High School: 7341   Very Satisfied   :19414
##  High School   :21744   Mod. Satisfied   :15513
##  Junior College: 2367   A Little Dissat  : 4057
##  Bachelor      : 6246   Very Dissatisfied: 1688
##  Graduate      : 2974
```

```
nrow(gooddata)
```

```
## [1] 40672
```

Now we will be creating frequency table, to explore each level of both the variables.

```
frequency <- table(gooddata$degree, gooddata$satjob)
frequency
```

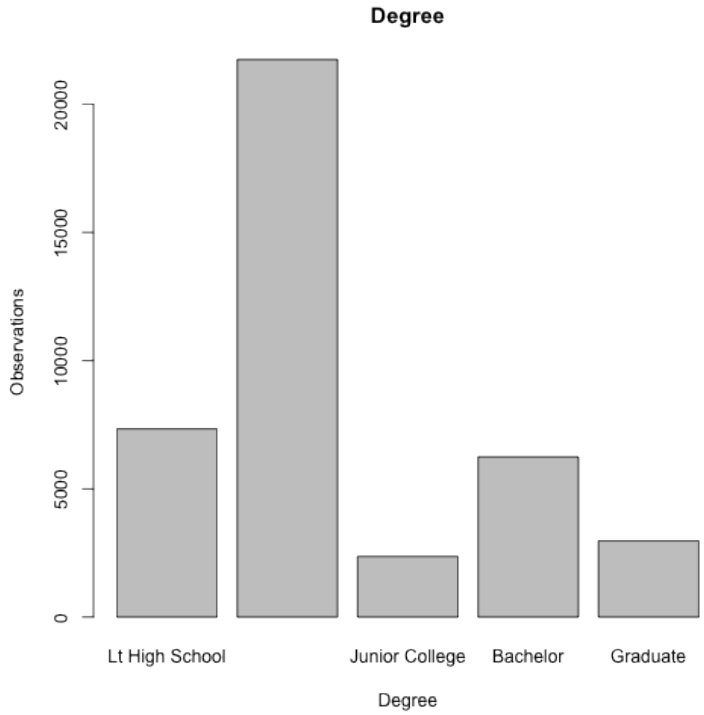
```
##
##                  Very Satisfied Mod. Satisfied A Little Dissat
##   Lt High School            3349           2793              821
##   High School              10005           8497             2281
##   Junior College            1201            883              214
##   Bachelor                  3106           2386              546
##   Graduate                  1753            954              195
##
##                  Very Dissatisfied
##   Lt High School                378
##   High School                   961
##   Junior College                 69
##   Bachelor                      208
##   Graduate                       72
```

```
Distribution of the variables using plots.
Degree
```

```
attach(gooddata)
```

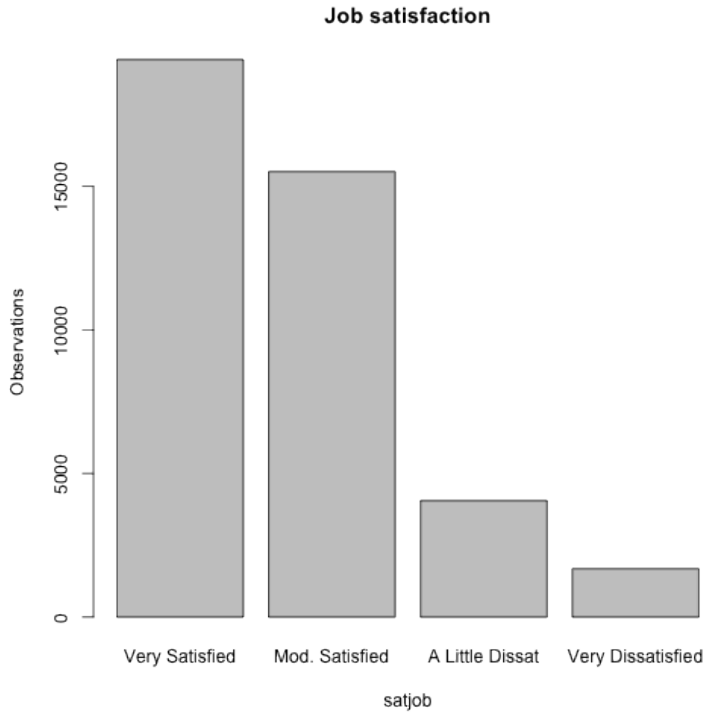
```
## The following object is masked from gss:
##
##     degree, satjob
```

```
plot(gooddata$degree, xlab = "Degree", ylab = "Observations", main = "Degree")
```



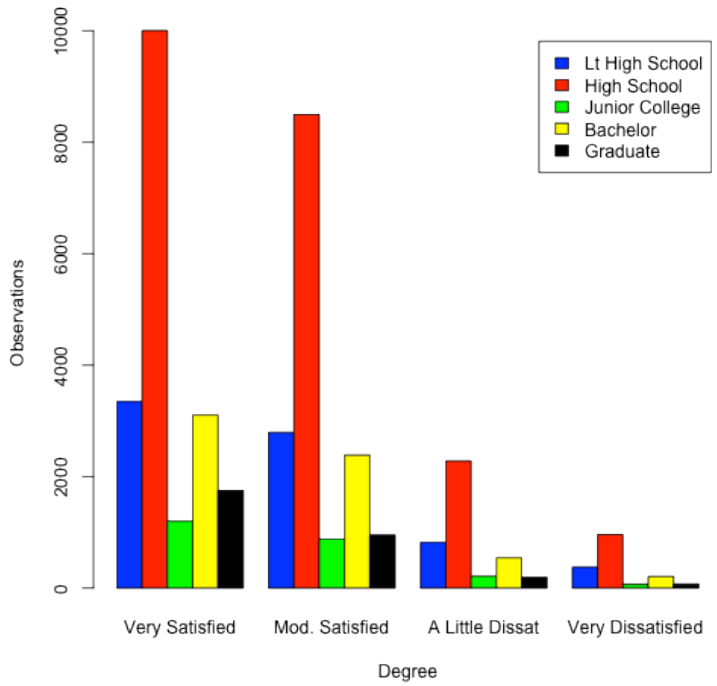

```
satjob
```

```
plot(gooddata$satjob, xlab = "satjob", ylab = "Observations", main = "Job
satisfaction")
```

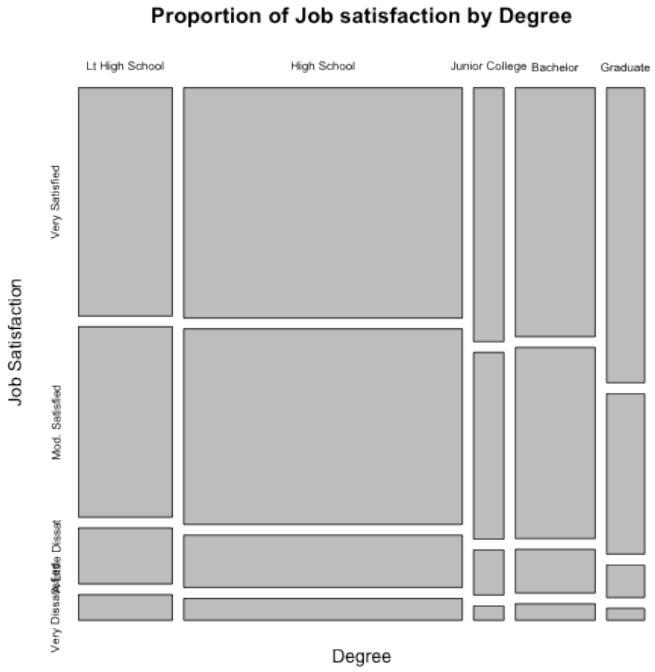



```
frequency plot of degree and satjob
```

```
barplot(frequency, xlab = "Degree", ylab = "Observations", beside = TRUE, main
= "",
    col = c("blue", "red", "green", "yellow", "black"), legend =
rownames(frequency))
```




```
plot(frequency, xlab = "Degree", ylab = "Job Satisfaction", main = "Proportion
of Job satisfaction by Degree")
```




```
Initial observation of the frequency plot suggests that those who hold some
sort of college degree are more satisfied with their job than those who left
High school and graduated High school.
The level of job Disatisfaction is also high among those who do not hold
college degrees.
```

Inference:

```
In this section we will discuss about the relationship between the two
variables Degree vs Satjob.
Before we go further we will set our hypothesis condition.

H_0 - Null hypothesis - Degree and satjob are independent. Job
satisfation(satjob) do not vary by the degree.
H_A - Alternate hypothesis - Degree and satjob are dependent. Job satisfaction
(satjob) do vary by the degree.

Loading the custom inference function that is used in lab.
```

```
load(url("http://s3.amazonaws.com/assets.datacamp.com/course/dasi/inference.Rdata"))
```

Inference function takes two categorical variable satjob and degree as input, we use proportion parameter to calcuate our test statistics.since we have two categorical variable with many levels, we will be using chi-square test of independence. The inference function identifies two categorical variable and runs chi-square test for us.The parameter to estimate for categorical variable is proportion.

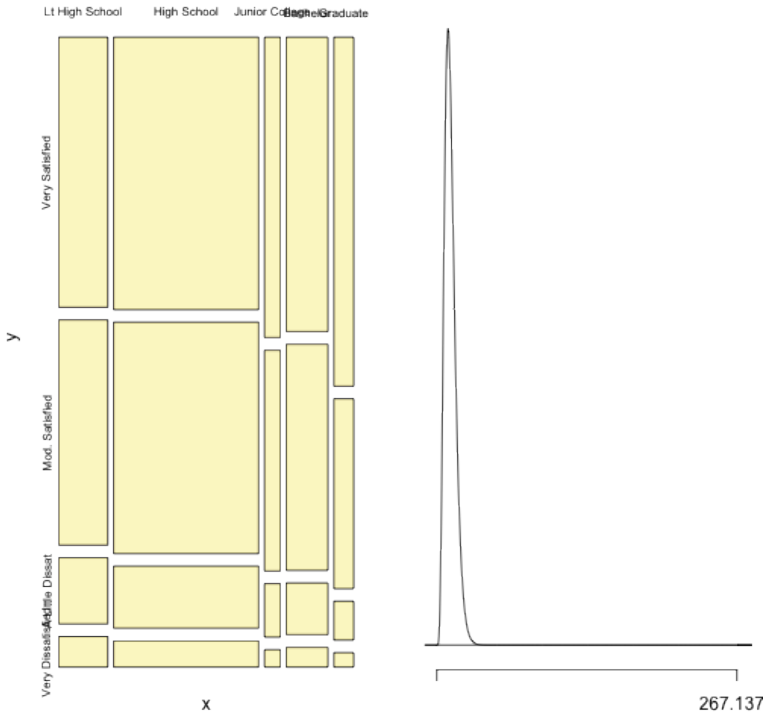
Before we go ahead and run the inference function, we need to check the condition for chi-square test.
  1)sampled observations are independent of each other.
  2)The samples are obtained by random sampling and without replacement.
  3)The total number of observations after removing NA are 40672 < 10% of US population.
  4)each case contributes to one cell in the table(data frame - gooddata).
  5)from the freqeuncy table we can observe that each particular scenario has more than 5 cases/counts

```
inference(gooddata$satjob, gooddata$degree, est = "proportion", type = "ht",
    method = "theoretical", alternative = "greater", inf_lines = TRUE)
```

```
## Warning: package 'lmPerm' was built under R version 3.0.2
## Warning: package 'openintro' was built under R version 3.0.2
## Warning: package 'BHH2' was built under R version 3.0.2
```

```
## Response variable: categorical, Explanatory variable: categorical
## Chi-square test of independence
##
## Summary statistics:
##                    x
## y                   Lt High School High School Junior College Bachelor
##   Very Satisfied              3349       10005           1201     3106
##   Mod. Satisfied              2793        8497            883     2386
##   A Little Dissat              821        2281            214      546
##   Very Dissatisfied            378         961             69      208
##   Sum                         7341       21744           2367     6246
##                    x
## y                   Graduate   Sum
##   Very Satisfied        1753 19414
##   Mod. Satisfied         954 15513
##   A Little Dissat        195  4057
##   Very Dissatisfied       72  1688
##   Sum                   2974 40672
```

```
## H_0: Response and explanatory variable are independent.
## H_A: Response and explanatory variable are dependent.
## Check conditions: expected counts
##                    x
## y                   Lt High School High School Junior College Bachelor
##   Very Satisfied            3504.1     10379.1        1129.84   2981.4
##   Mod. Satisfied            2800.0      8293.5         902.81   2382.3
##   A Little Dissat            732.3      2168.9         236.11    623.0
##   Very Dissatisfied          304.7       902.4          98.24    259.2
##                    x
## y                   Graduate
##   Very Satisfied      1419.6
##   Mod. Satisfied      1134.3
##   A Little Dissat      296.6
##   Very Dissatisfied    123.4
##
##  Pearson's Chi-squared test
##
## data:  y_table
## X-squared = 267.1, df = 12, p-value < 2.2e-16
```



Since,chisq test calculates test statistics (chisq) from the Expected and observed counts.First the observed counts(contingency table) is calculated and expected counts of each level of categorical variable is calculated.The test statistic value is very high around 267 and degree of freedom is 12. The higher test statistic means higher deviation from the null hypothesis and it provides strong evidence for alternate hypothesis. we get p-Value less than 2.2e<-16 though it is not an exact value, but small enough to reject null hypothesis.
Alternatively we can calculate this by using chisq function available in R.

```
chisq.test(frequency)
```

```
##
##  Pearson's Chi-squared test
##
## data:  frequency
## X-squared = 267.1, df = 12, p-value < 2.2e-16
```

we can calculate the exact p value which is less than 2.2e<-16 using the pchisq function in R.

```
pchisq(267.1, 12, lower.tail = FALSE)
```

```
## [1] 3.677e-50
```

P value obtained is very low and since it is less than 5% significance level, we reject null hypothesis that there is no relationship between the satjob and degree. There is a relationship between Degree and satjob. As one obtains college or higher degree he is more satisfed with his job and less dissatisfaction with is job.
From the chisquare graph we can observe the p value. The tail area above the calculated test statistic value of 267.137, it is very small thin dark line.

Conclusion:

Though the findings of this study shows that there is a significance relationship between Degree attained and job satisfaction. we have to be careful on generalizing this study, Educational attainment alone does not translate to job satisfaction. It may also depends on individuals performance. whether the effort they put in matches up to their expectation which results in Job satisfaction.In the future studies we can also include a new variable for effort, and check how it measures up with the job satisfaction and degree attained.
finally, It pays to earn some sort of college degree and satisfaction level only grows more with the advanced degrees.

References:

```
Data Citation:
Smith, Tom W., Michael Hout, and Peter V. Marsden. General Social Survey,
1972-2012 [Cumulative File]. ICPSR34802-v1. Storrs, CT: Roper Center for
Public Opinion Research, University of Connecticut /Ann Arbor, MI: Inter-
university Consortium for Political and Social Research [distributors], 2013-
09-11. doi:10.3886/ICPSR34802.v1

Links:

Research Home page -
http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/34802/version/1

Variable Description -
https://d396qusza40orc.cloudfront.net/statistics%2Fproject%2Fgss1.html

Data - http://bit.ly/dasi_gss_data
```

Appendix:

```
data[1:50, ]
```

```
##            degree          satjob
## 1        Bachelor A Little Dissat
## 2  Lt High School            <NA>
## 3     High School  Mod. Satisfied
## 4        Bachelor  Very Satisfied
## 5     High School            <NA>
## 6     High School  Mod. Satisfied
## 7     High School  Very Satisfied
## 8        Bachelor A Little Dissat
## 9     High School  Mod. Satisfied
## 10    High School  Mod. Satisfied
## 11    High School            <NA>
## 12 Lt High School  Very Satisfied
## 13 Lt High School  Very Satisfied
## 14 Lt High School  Mod. Satisfied
## 15 Lt High School  Very Satisfied
## 16    High School  Mod. Satisfied
## 17    High School            <NA>
## 18 Lt High School            <NA>
## 19       Bachelor  Very Satisfied
## 20    High School            <NA>
## 21    High School  Very Satisfied
## 22    High School  Very Satisfied
## 23    High School  Mod. Satisfied
## 24    High School            <NA>
```

```
## 25       Bachelor            <NA>
## 26    High School  Mod. Satisfied
## 27    High School  Mod. Satisfied
## 28    High School A Little Dissat
## 29    High School  Mod. Satisfied
## 30 Lt High School  Very Satisfied
## 31 Lt High School  Mod. Satisfied
## 32    High School  Very Satisfied
## 33       Bachelor            <NA>
## 34 Lt High School            <NA>
## 35    High School  Mod. Satisfied
## 36    High School  Mod. Satisfied
## 37    High School            <NA>
## 38 Lt High School            <NA>
## 39 Lt High School  Mod. Satisfied
## 40    High School  Mod. Satisfied
## 41 Lt High School            <NA>
## 42    High School  Very Satisfied
## 43 Lt High School  Mod. Satisfied
## 44 Lt High School            <NA>
## 45 Lt High School            <NA>
## 46    High School            <NA>
## 47    High School  Very Satisfied
## 48    High School  Very Satisfied
## 49 Lt High School  Mod. Satisfied
## 50 Lt High School  Mod. Satisfied
```