

Statistical inference with the GSS data

Setup

Load packages

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
library(statsr)
```

```
## Warning: package 'BayesFactor' was built under R version 4.0.5
```

```
## Warning: package 'coda' was built under R version 4.0.5
```

Load data

Make sure your data and R Markdown files are in the same directory. When loaded your data file will be called `gss`. Delete this note when before you submit your work.

```
load("gss.Rdata")
```

Part 1: Data

Since 1972, the General Social Survey (GSS) has been monitoring societal change and studying the growing complexity of American society. The GSS aims to gather data on contemporary American society in order to monitor and explain trends and constants in attitudes, behaviors, and attributes; to examine the structure and functioning of society in general as well as the role played by relevant subgroups; to compare the United States to other societies in order to place American society in comparative perspective and develop cross-national models of human society; and to make high-quality data easily accessible to scholars, students, policy makers, and others, with minimal cost and waiting.

GSS questions cover a diverse range of issues including national spending priorities, marijuana use, crime and punishment, race relations, quality of life, confidence in institutions, and sexual behavior.

Part 2: Research question

My research question is : Do education reflects the income of an individual? Do higher degree people tends to have higher income.

Part 3: Exploratory data analysis

```
class(gss$degree)
```

```
## [1] "factor"
```

```
class(gss$incom16)
```

```
## [1] "factor"
```

```
gss%>%  
  select(degree,incom16) %>%  
  str()
```

```
## 'data.frame':   57061 obs. of  2 variables:  
## $ degree : Factor w/ 5 levels "Lt High School",...: 4 1 2 4 2 2 2 4 2 2 ...  
## $ incom16: Factor w/ 6 levels "Far Below Average",...: 3 4 3 3 2 3 4 3 3 1 ...
```

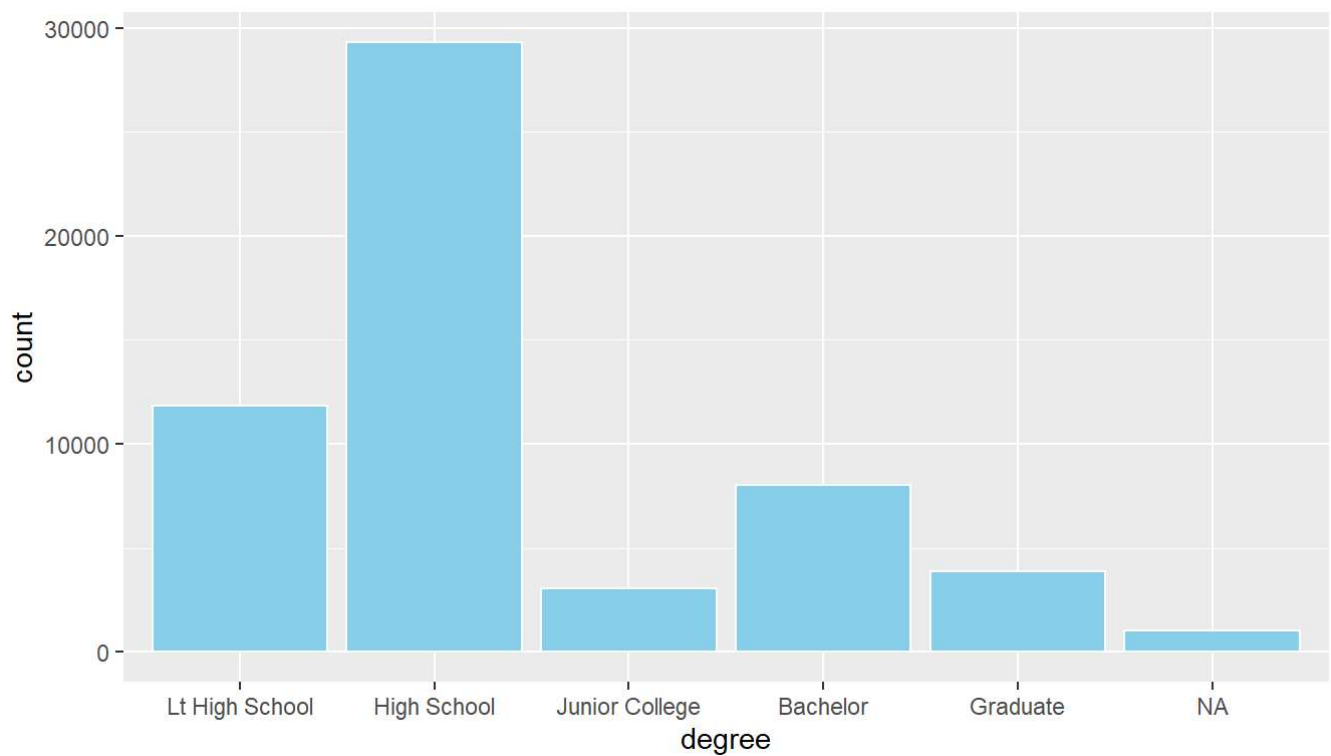
```
gss %>%  
  group_by(degree) %>%  
  summarise(n())
```

```
## # A tibble: 6 x 2  
##   degree      `n()`  
##   <fct>      <int>  
## 1 Lt High School 11822  
## 2 High School   29287  
## 3 Junior College 3070  
## 4 Bachelor      8002  
## 5 Graduate      3870  
## 6 <NA>          1010
```

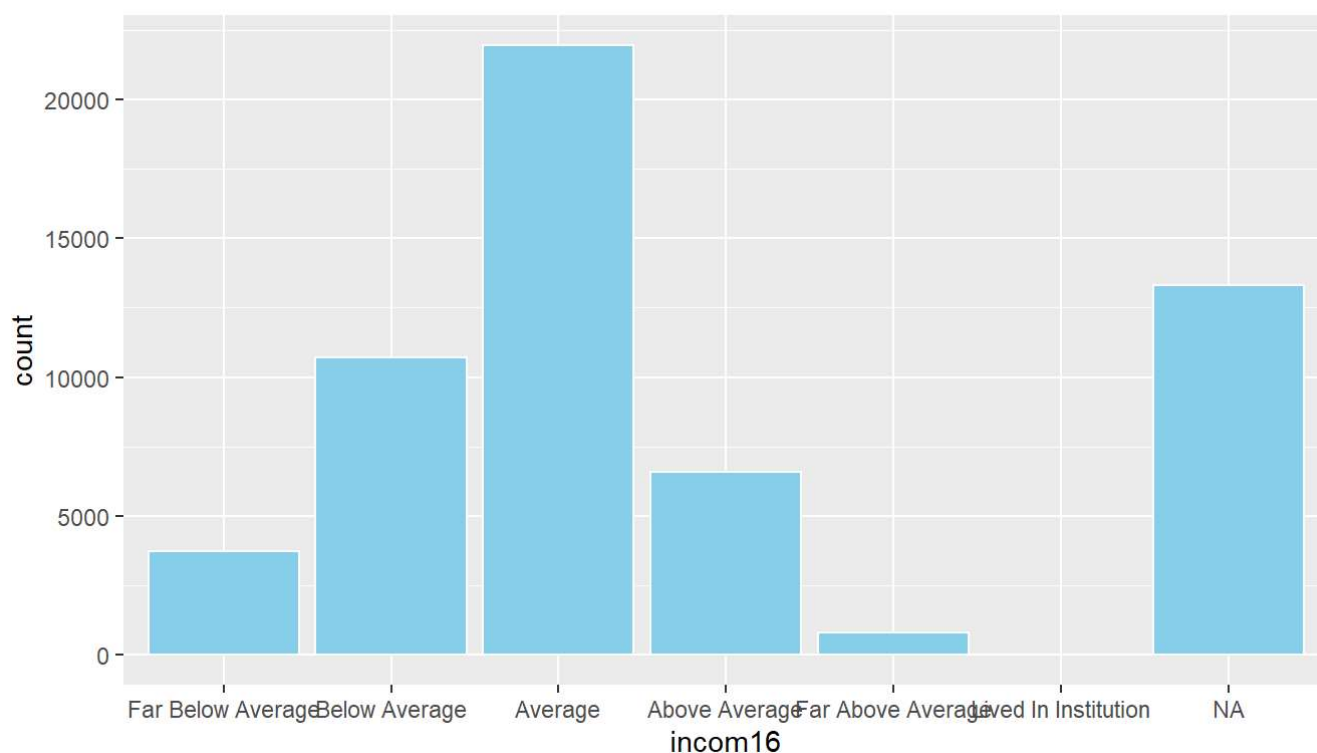
```
gss %>%  
  group_by(incom16) %>%  
  summarise(n())
```

```
## # A tibble: 7 x 2
##   incom16      `n()`
##   <fct>      <int>
## 1 Far Below Average    3725
## 2 Below Average    10692
## 3 Average    21941
## 4 Above Average    6575
## 5 Far Above Average    796
## 6 Lived In Institution    10
## 7 <NA>    13322
```

```
ggplot(data=gss)+
  geom_bar(aes(degree),color="white",fill="skyblue")
```



```
ggplot(data=gss)+
  geom_bar(aes(incom16),color="white",fill="skyblue")
```



The analysis shows that most of the people have completed the high school. And most of the people tends to have an average income.

Part 4: Inference

The null hypothesis

H0 : Degree and Income are independent of each other

HA : Degree and Income are not independent of each other and are related in some way.

Since we are dealing with two multilevel categorical variables we will use Chi square test for independence to see if there is a relation between the two variables.

Conditions for Chi Square Test :

Independence : We have used random sampling and total observations are less than 10% of the population. Also each case contributes to only one cell.

```
chisq.test(gss$degree,gss$incom16)
```

```
## Warning in chisq.test(gss$degree, gss$incom16): Chi-squared approximation may be
## incorrect
```

```
##
## Pearson's Chi-squared test
##
## data: gss$degree and gss$incom16
## X-squared = 2852.5, df = 20, p-value < 2.2e-16
```

The p-value is very close to zero. Hence we reject our null hypothesis and conclude that there is a relation between highest degree obtained and social class. However since no random assignment was used, we can't infer causation from this study.

Conclusion: A reading of this observational study indicates that members who have higher degree tends to have higher income.