Project One

Christine Zlotek and Brad Dale

```
setwd("~/Dropbox/School/MKTG 420 - Data Vis")
survey = load("EdGenderWork.Rda")
survey = EdGenderWork
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(tidyr)
library(ggplot2)
library(car)
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
library(RColorBrewer)
```

Introduction:

The Princeton Survey Research collects data on the respondent's general demographics and opinions on education, careers and equality. We utilized this data set to give us insight to correlations that are present between demographics and opinions. Using R allowed us to display the data in a way that was readable and understandable for a range of audiences. The questions we posed were:

1. Does income affect a respondent's happiness? OR Is there a correlation between happiness and equality?

2. Are people who take out more debt more satisfied with their education?

3. Does a respondents gender influence their view of equality in the workplace?

4. Does the level of education influence one's view of equality?

To get a sense of our sample information, let's explore some basic demographics regarding our sample participants.

```
survey$age=as.numeric(survey$age)
ggplot(survey, aes(age))+geom_density()+ggtitle("Age Density")+scale_x_continuous(breaks=c(10,20,30,40,!
labs(x="Age", y="Density")
```

Age Density

The age density graph displays the how dense each age group is from the individuals who were surveyed. The graph shows that most of the individuals who were interviewed for this survey were between the ages 25 and 35 years old. The density starts to decline after the age 65, which means as the age demographic becomes older there were less people interviewed.

```
survey$gender = Recode(survey$sex, "1='Male'; 2='Female'")
ggplot(survey, aes(factor(gender), age)) + geom_boxplot() +
  ggtitle("Gender and Age")+
  labs(x = "Gender", y = "Age")
```

Gender and Age
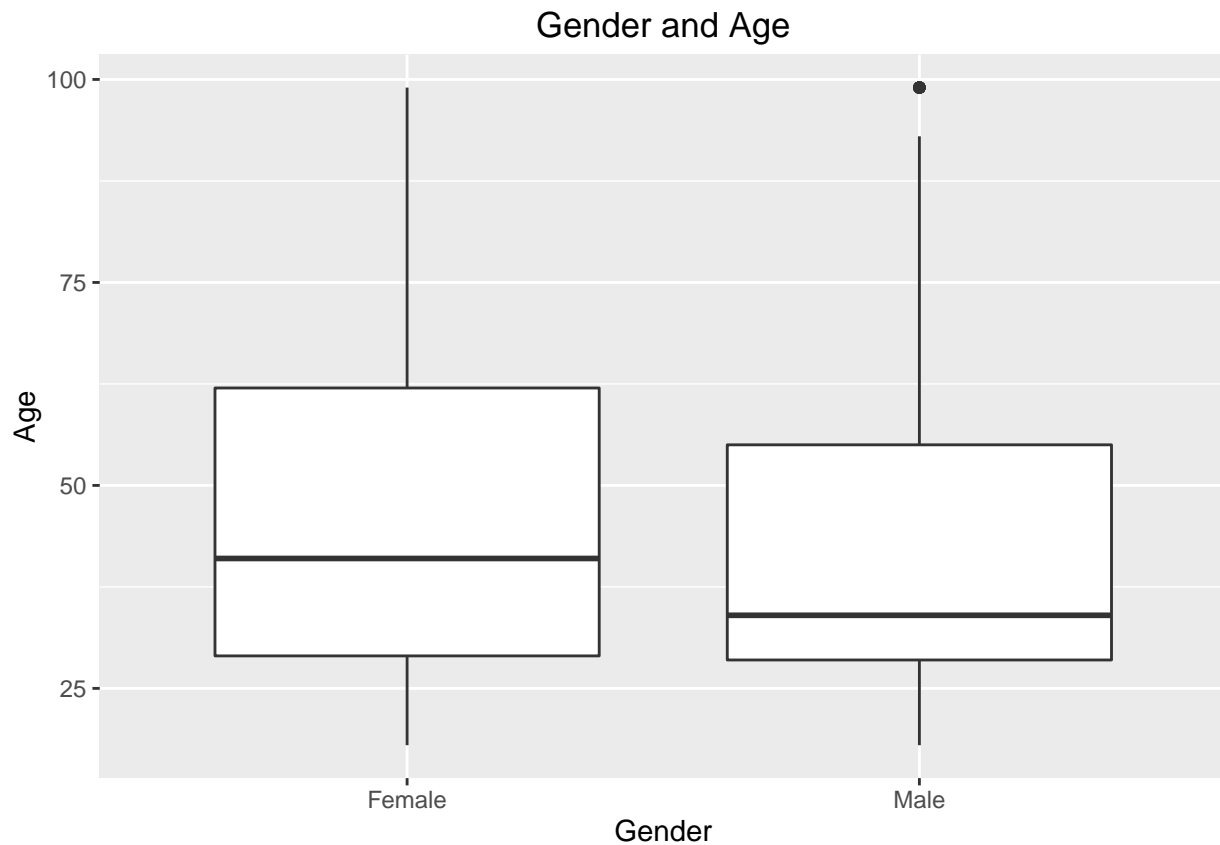
This boxplot shows the dispersion in age between males and females. The females sampled were on average older than the males, with the range of ages also being greater for females.

```
str(survey$state)
```

```
## Class 'labelled'  atomic [1:2002] 29 55 19 37 36 51 48 48 13 12 ...
##   ..- attr(*, "label")= chr "State from FIPS code - 2012 zip merge"
##   ..- attr(*, "labels")= Named num [1:51] 1 2 4 5 6 8 9 10 11 12 ...
##   .. ..- attr(*, "names")= chr [1:51] "Alabama" "Alaska" "Arizona" "Arkansas" ...
```

```
survey$state = as.factor(survey$state)
```

```
ggplot(survey, aes(factor(sstate),age))+geom_boxplot()+scale_x_discrete("US States", labels= c("1" = "A
  theme(axis.text.x=element_text(angle=90))+coord_flip()
```

Ages in US States

This boxplot shows the range of states that people were from and the average age. Whyoming had the greatest range of ages for the middle 50% of those sampled and had an average age of older than 60.

```r
survey$income= as.numeric(survey$income)
survey$income= ifelse(survey$income == 99,NA, survey$income)

survey$income[survey$income==1]<- "<10K"
survey$income[survey$income==2]<- "10K-20K"
survey$income[survey$income==3]<- "20K-30K"
survey$income[survey$income==4]<- "30K-40K"
survey$income[survey$income==5]<- "40K-50K"
survey$income[survey$income==6]<- "50K-75K"
survey$income[survey$income==7]<- "75K-100K"
survey$income[survey$income==8]<- "100K-150K"
survey$income[survey$income==9]<- "150K+"

survey$income= as.factor(survey$income)

survey$income = factor(survey$income,levels(survey$income)[c(1,3,5,6,7,8,9,2,4)])

df1=survey[!is.na(survey$income),]
df1=df1[df1$q1!="NA",]
df1$q1=factor(df1$q1)
summary(df1$q1)
```

```
##   1   2   3   9
## 614 911 223  32
```

4

```
ggplot(df1, aes(x =income))+geom_bar(stat="count")+ggtitle("Income Brackets")+
labs(x= "Income" , y="Count")+
  coord_flip()
```
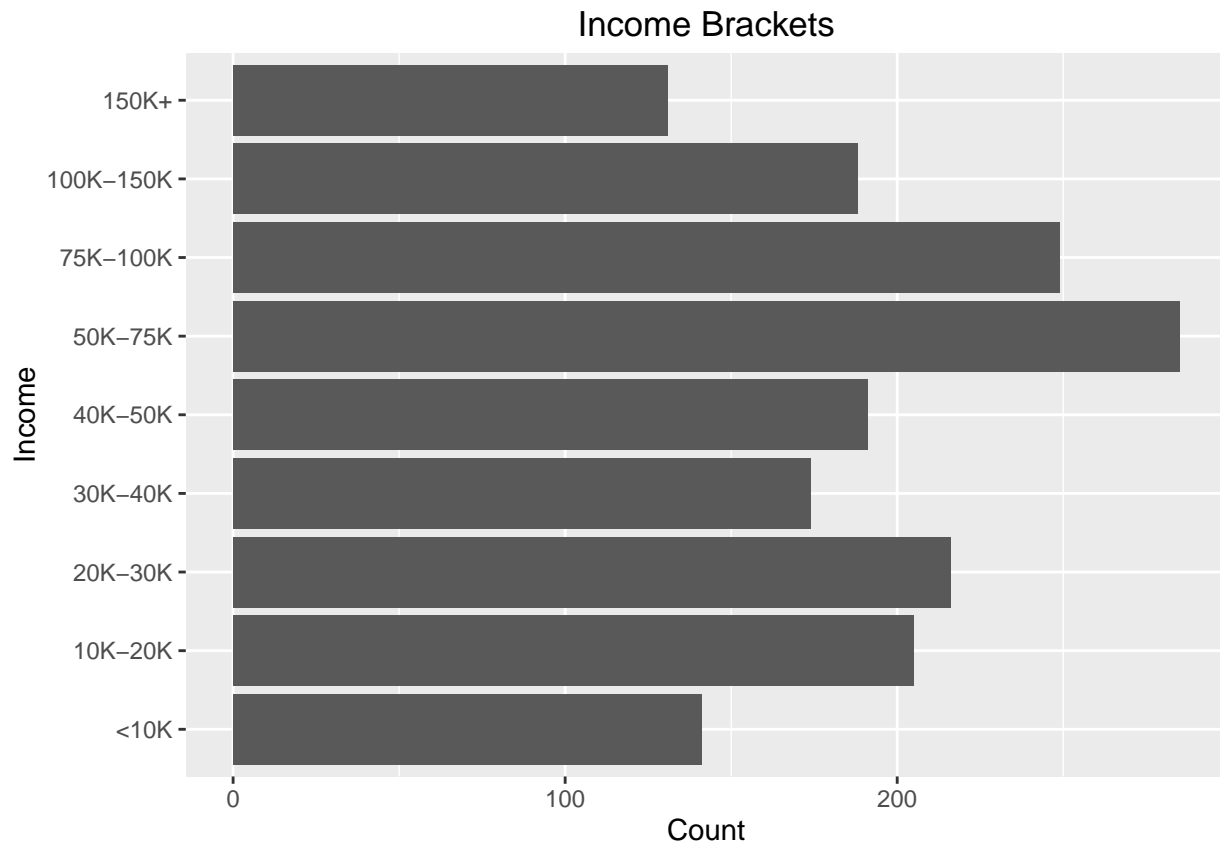
Income Brackets

This graph shows the income brackets that the surveyed individuals fell into. The majority of the individuals were in the $50,000 to $75,000 income range. It is imperative to note that the income distribution did not match that of the U.S. income distribution. The $150,000 or more range was extremely high compared to the U.S. distribution, meaning this survey captured more individuals with wealthier incomes.

```
survey$gender = Recode(survey$sex, "1='Male'; 2='Female'")

survey$q4=as.factor(survey$q4)
survey$gender=as.factor(survey$gender)
survey$q4= ifelse(survey$q4==9,NA, survey$q4)



ggplot(survey[!is.na(survey$q4),], aes(x= as.factor(q4), fill=gender))+geom_bar(stat="count", position =
  scale_x_discrete("Equality between Men and Women", labels= c("1" = "Men and Women Equally" , "2"= "Wor
ggtitle("Views of Equality by Gender")+
  scale_fill_brewer(palette = "Set2", name = "Gender")+
labs(y="Count")
```

Views of Equality by Gender
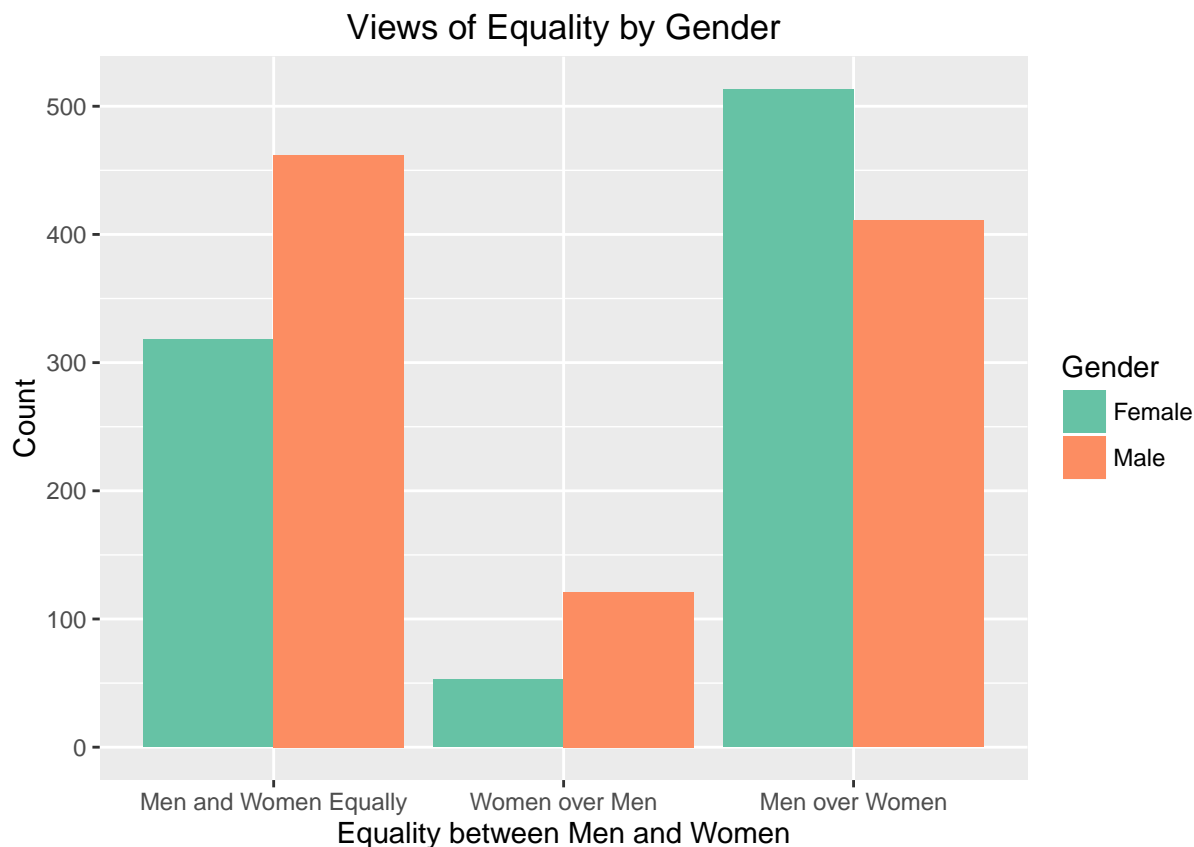
There is a clear difference of opinion by men and women about equality. More often than not, men think men and women are treated equally. Whereas women more often than not, think men are generally treated more favorably than women. Additionally, looking at the middle columns showing "Women over Men" illustrates that nearly twice as many men believe in that form of inequality as women do.

```
survey$income=as.factor(survey$income)
survey$q1=as.factor(survey$q1)

survey$q1 = Recode(survey$q1, "1='Very happy'; 2='Pretty happy'; 3='Not too happy'; 9='NA'")

levels(survey$income) <- gsub(" ", "\n", levels(survey$income))
df1=survey[!is.na(survey$income),]
df1=df1[df1$q1!="NA",]
df1$q1=factor(df1$q1)


ggplot(df1, aes(x=income, fill=q1))+geom_bar(stat = "count", position = "dodge")+ labs(x = "Age", y = "(
  labs(x = "Income", y = "Count")+
  ggtitle("Happiness Based on Income")+
  scale_fill_brewer(palette = "Set2", name = "Happiness")+
  theme(axis.text.x=element_text(angle=17))
```

Happiness Based on Income
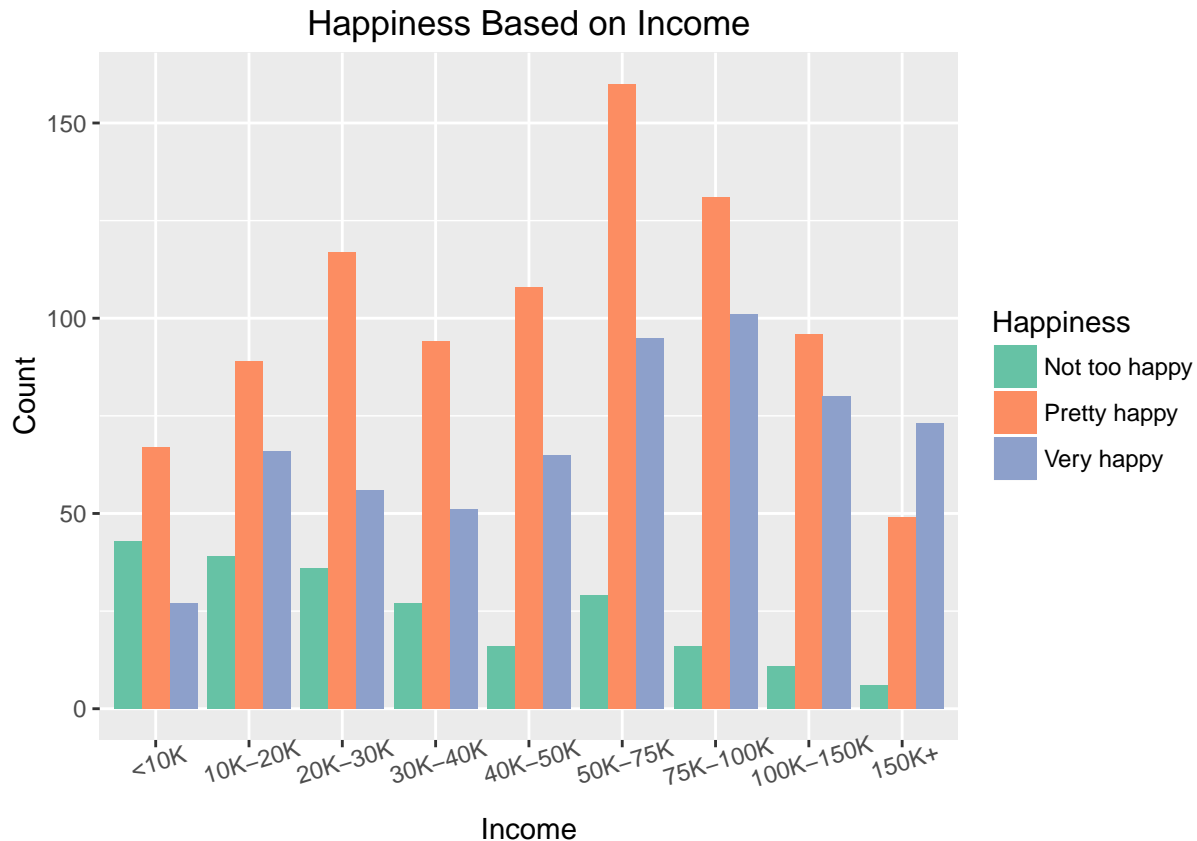
The higher the income the happier individuals reported to be. At the lower income brackets there are significant amounts of "Not too happy" where the higher income brackets reported almost no unhappiness. This trend illustrates that with increased wealth, happiness tends to increase. With more resources in wealth, people are able to partake more in our economy, purchase more goods they desire, and worry less about money.

```r
survey$q4=ifelse(survey$q4==9,NA,survey$q4)
survey$q1=ifelse(survey$q1==9,NA,survey$q1)
survey$q1=as.factor(survey$q1)
survey$q4=as.factor(survey$q4)

survey$q4 = recode(survey$q4, "1='Men and Women Equal'; 2 = 'Women over Men'; 3='Men over Women'")

survey$q1 = factor(survey$q1, levels(survey$q1)[c(4,3,2,1)])

ggplot(survey, aes(x=q1, fill=q4))+geom_bar(stat = "count", position = "dodge")+
  scale_x_discrete("Happiness", labels = c("4" = "Very Happy","3" = "Pretty Happy", "2" = "Not Happy",
  ggtitle("Happiness and Views on Equality")+
  scale_fill_brewer(palette= "Set2", name ="Views on Equality")
```

Happiness and Views on Equality
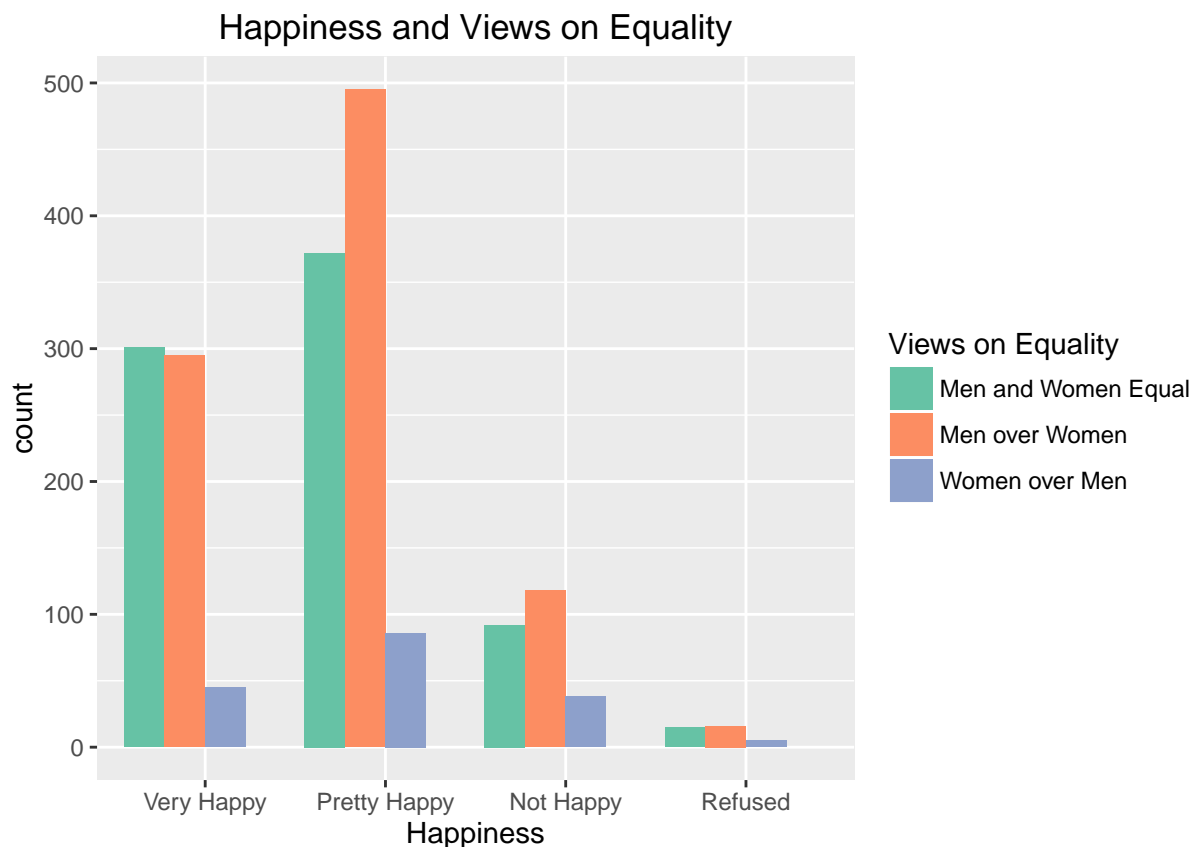
We hypothesized that there would be a correlation between one's views on equality and their level of happiness. Based on the information provided, we could not make a decision as all three views on equality were fairly equally divided up between all three levels of happiness. While there is a peak of "Men over Women" in the "Pretty Happy" group, the levels are nearly flat with "Men and Women Equal" for the other two levels of happiness.

```r
survey$q13 = Recode(survey$q13, "1='Less than $10,000'; 2='Between $10,000 and $20,000'; 3='Between $20

survey$q9 = Recode(survey$q9, "1='Yes, has paid off'; 2='No, has not paid off'; 3='Not yet, but it will

survey$q13 = as.factor(survey$q13)
survey$q9 = as.factor(survey$q9)


survey$q13 = factor(survey$q13, levels(survey$q13)[c(4,1,2,3,5)])


survey$q9 = factor(survey$q9, levels(survey$q9)[c(5,1,2,4,3)])

levels(survey$q13) <- gsub(" ", "\n", levels(survey$q13))

ggplot(survey[!is.na(survey$q13),], aes(x=q13, fill=q9))+geom_bar(stat = "count", position = "dodge")+
  labs(x = "Debt Taken Out", y = "Count")+
  ggtitle("Debt Taken Out to Satisfaction with Education")+
  scale_fill_brewer(palette = "Set2", name = "Satisfaction with Education")
```

Debt Taken Out to Satisfaction with Education
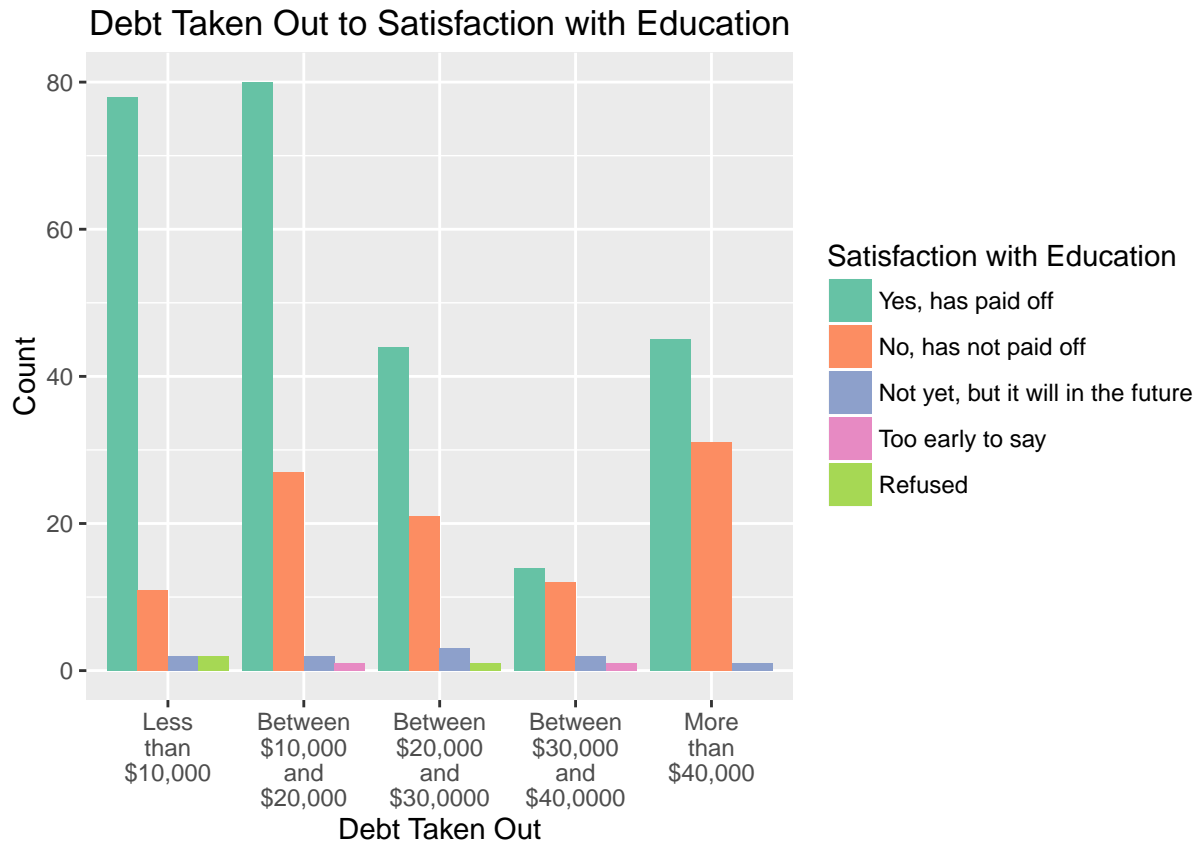
It appears there is a correlation between the amount of debt our surveyed individuals take out versus their satisfaction with their education. The groups that most believe their education has not paid off for themselves are the groups who have taken out the most debt, $30,000 or more. The group who took out the least debt ($10,000 or less), was the most satisfied with their education and were the most likely to say their education has paid off.

```
survey$age = as.numeric(survey$age)


survey$q1 = factor(survey$q1, levels(survey$q1)[c(4,3,2,1)])

ggplot(survey, aes(factor(q1), age)) + geom_boxplot()+
  labs(x = "Happiness", y = "Age")+
  ggtitle("Average Age for Happiness Levels")+
  scale_x_discrete(labels=c("4"="Very Happy", "3"="Pretty Happy", "2"="Not Too Happy", "1"="Refused"))
```

Average Age for Happiness Levels
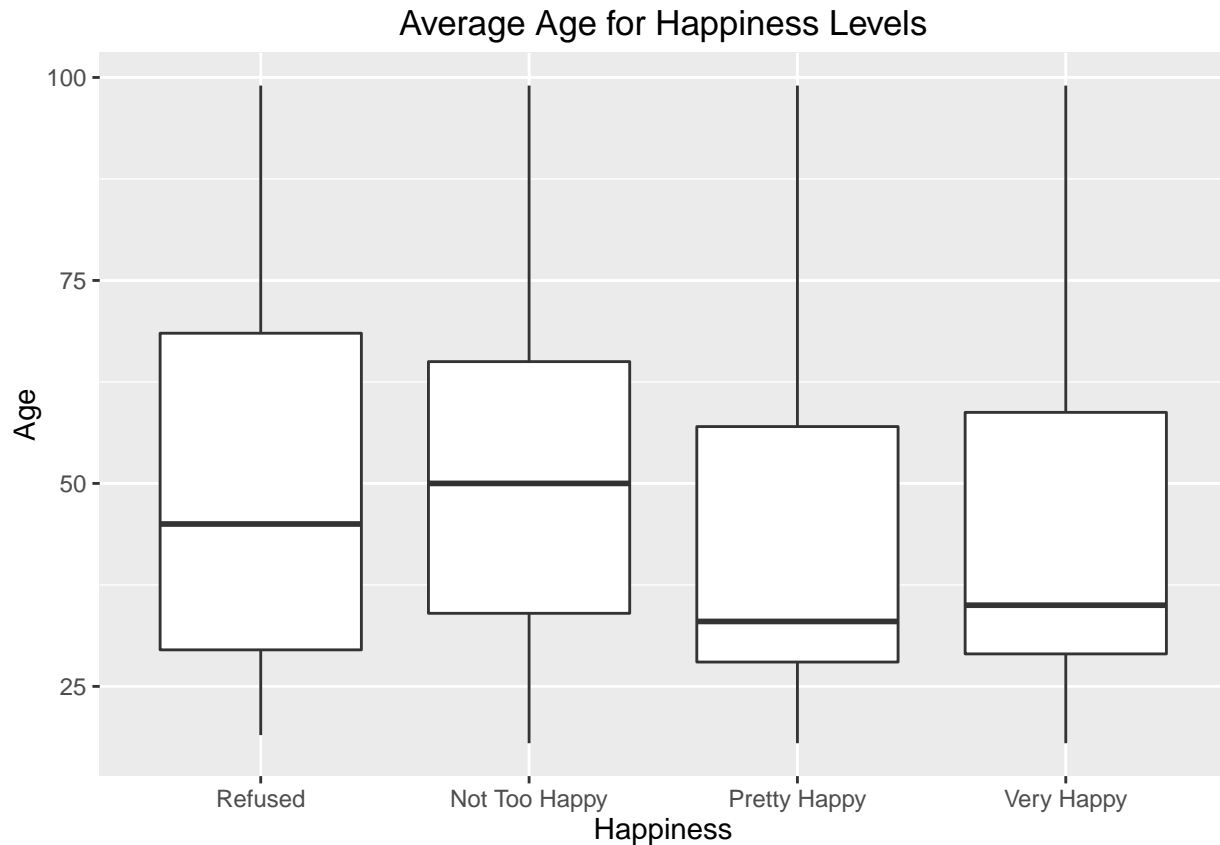
From our sample group, it appears there is a gloomy future for some of us. The "Not Too Happy" group has the oldest age for all quartiles as illustrated by the box plot lines. The mean age for the unhappy group is around 50, while both of the happy groups are around 35. This doesn't mean as you get older you get unhappy, but it does illustrate that the older population of our sample is less happy with their lives.

```r
survey$income = Recode(survey$income, "1='Less than $10,000'; 2='10 to under $20,000'; 3='20 to under $

survey$income = as.factor(survey$income)
survey$age=as.numeric(survey$age)

survey$income = factor(survey$income, levels(survey$income)[c(1,3,5,6,7,8,9,2,4,10)])

levels(survey$income) <- gsub(" ", "\n", levels(survey$income))

ggplot(survey, aes(factor(income), age)) + geom_boxplot(width = .5) +
  labs(x = "Income", y = "Age")+
  ggtitle("Average Age of Income Brackets")+
  theme(axis.text.x=element_text(angle=20))
```

Average Age of Income Brackets

These box plots show that on average, income increases with age. The ages are pretty consistent across the lowest half of incomes, but it starts to diverge more significantly at the middle and upper income levels of 50K-75K and 75K-100K. As one gains experience, moves jobs and gains promotions as they age, they typically move vertically to a higher income bracket.

```
survey$q43 = Recode(survey$q43, "1='Yes'; 2='No'; 9='NA'")

survey$q43 = as.factor(survey$q43)

survey$q43 = factor(survey$q43, levels(survey$q43)[c(3,2,1)])

ggplot(survey[!is.na(survey$q43),], aes(x=q43, fill=gender))+geom_bar(stat = "count", position = "dodge
  labs(x = "Has the US Made the Necessary Changes to Give Men and Women Equality?", y = "Count") +
  ggtitle("Has the US Made Changes to Promote Equality?") +
    scale_fill_brewer(palette= "Set2", name ="Gender")
```

Has the US Made Changes to Promote Equality?

Has the US Made the Necessary Changes to Give Men and Women Equality?
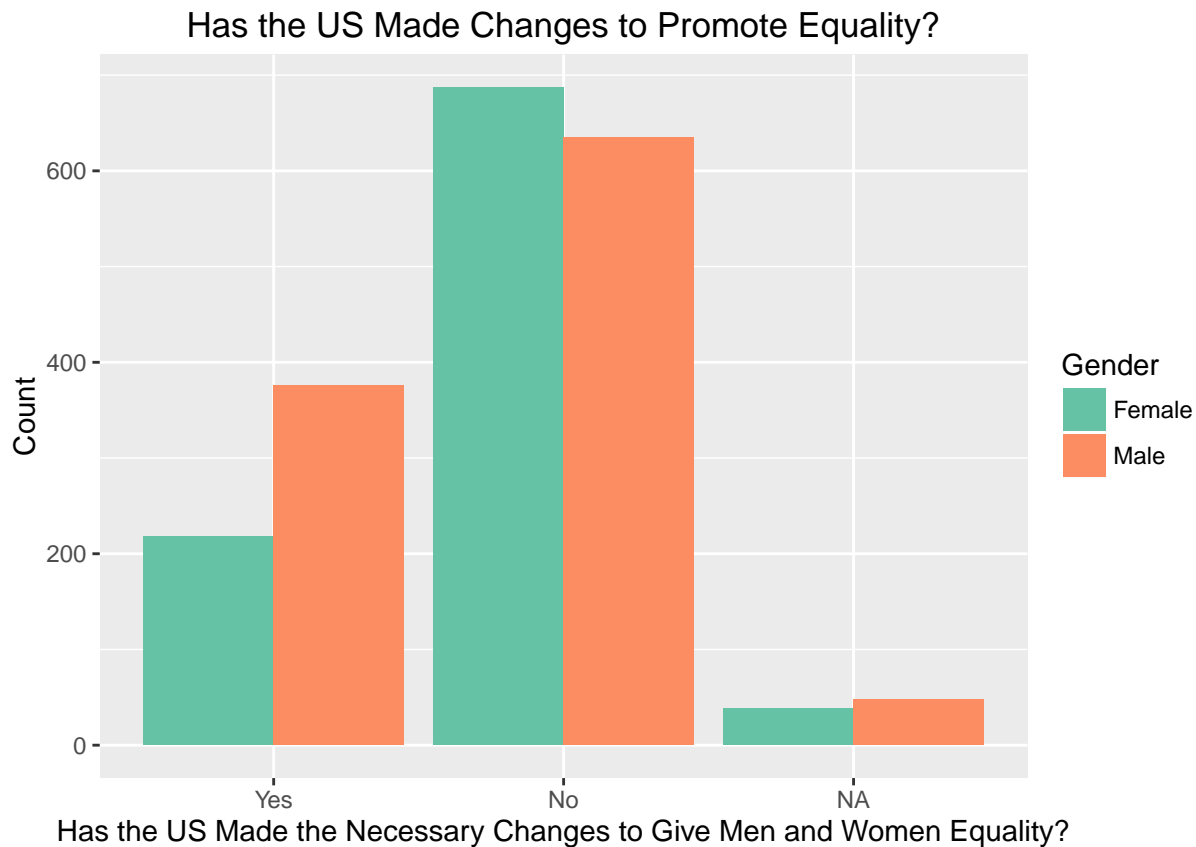
Women are more likely to think the US needs to continue making changes to give men and women equality in the workplace. On the other hand, men are much more likely to think the US has already taken the actions needed to promote equality in the workplace. Women are less likely to be satisfied with the US's actions taken so far and want more from changes to happen. Women more often feel there is inequality in the workplace and this is further reflected in that they want change to continue to take place.

```
survey$gender=as.factor(survey$gender)
survey$qe7=as.factor(survey$qe7)
survey$qe7= ifelse(survey$qe7==9,NA, survey$qe7)

ggplot(survey[!is.na(survey$qe7),], aes(x= as.factor(qe7), fill=gender))+
  geom_bar(stat="count", position = "dodge")+
  scale_x_discrete("Asked for a Pay Raise or Promotion", labels= c("1" = "Yes" , "2"=   "No"))+
  ggtitle("Gender and Asking for a Raise")+
  labs(y="Count", fill= "Gender")+
  scale_fill_brewer(palette= "Set2")
```

Gender and Asking for a Raise
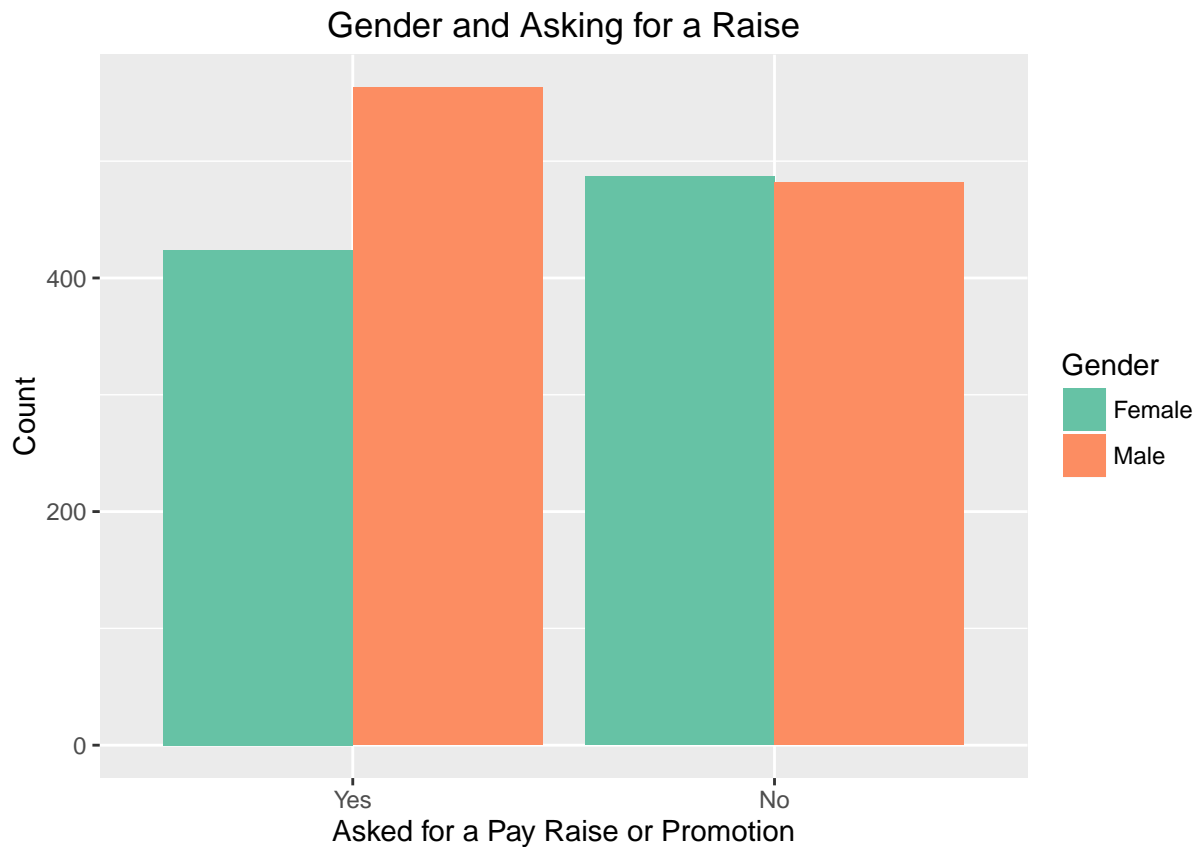
In this display, relationship between those who ask for a raise and respondents gender, there was a stark difference between men who ask for a raise over women. Women are not as likely as men to ask for a raise or promotion in their careers.

```
survey$q48 = Recode(survey$q48, "1='Men More Focused'; 2='Women More Focused'; 3='Equally Focused'; 4='

survey$gender = Recode(survey$sex, "1='Male'; 2='Female'")
survey$q48 = as.factor(survey$q48)
survey$gender=as.factor(survey$gender)


survey$q48 = factor(survey$q48, levels(survey$q48)[c(1,2,4,3)])

levels(survey$q48) <- gsub(" ", "\n", levels(survey$q48))

ggplot(survey, aes(x=q48, fill=gender))+geom_bar(stat = "count", position = "dodge")+
  labs(x = "What Gender is More focused on Their Career?", y = "Count") +
  ggtitle("What Genders Think of Career Focus") +
    scale_fill_brewer(palette= "Set2", name ="Gender")
```

What Genders Think of Career Focus
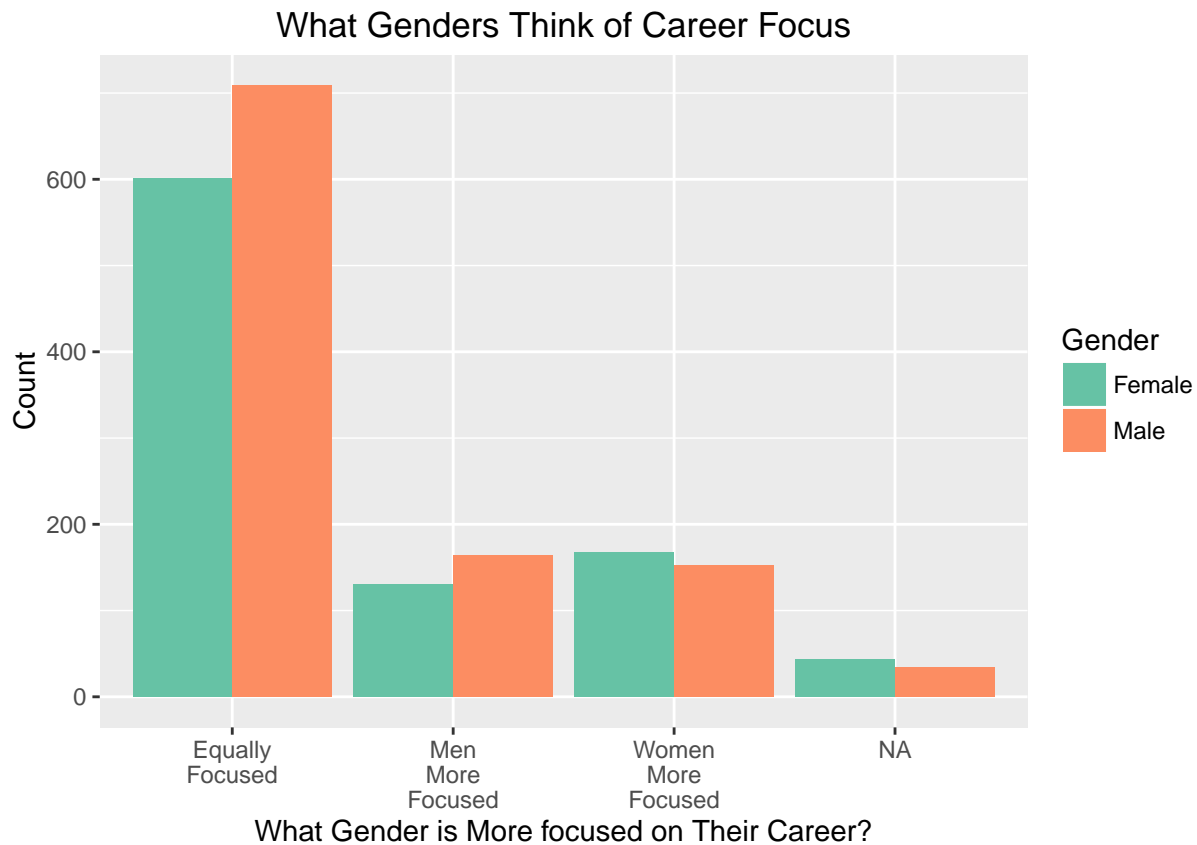
From looking at the results of a question asking which gender is more focused on their career, it appears that women see themselves more focused on their own career as men see themselves more focused on their career. Each gender is illustrating confidence that they are actually the gender that is more focused on their career.

```
survey$q44 = Recode(survey$q44, "1='Easier for Men'; 2='Easier for Women'; 3='No Difference'; 9='NA'")

survey$q44 = as.factor(survey$q44)

survey$q44 = factor(survey$q44, levels(survey$q44)[c(1,2,4,3)])

ggplot(survey[!is.na(survey$q44),], aes(x=q44, fill=gender))+geom_bar(stat = "count", position = "dodge
  labs(x = "Which Gender is it Easier for to Get a Job?", y = "Count")+
    scale_fill_brewer(palette= "Set2", name ="Gender")+
  ggtitle("Who Gets a Job Easier?")
```

Who Gets a Job Easier?
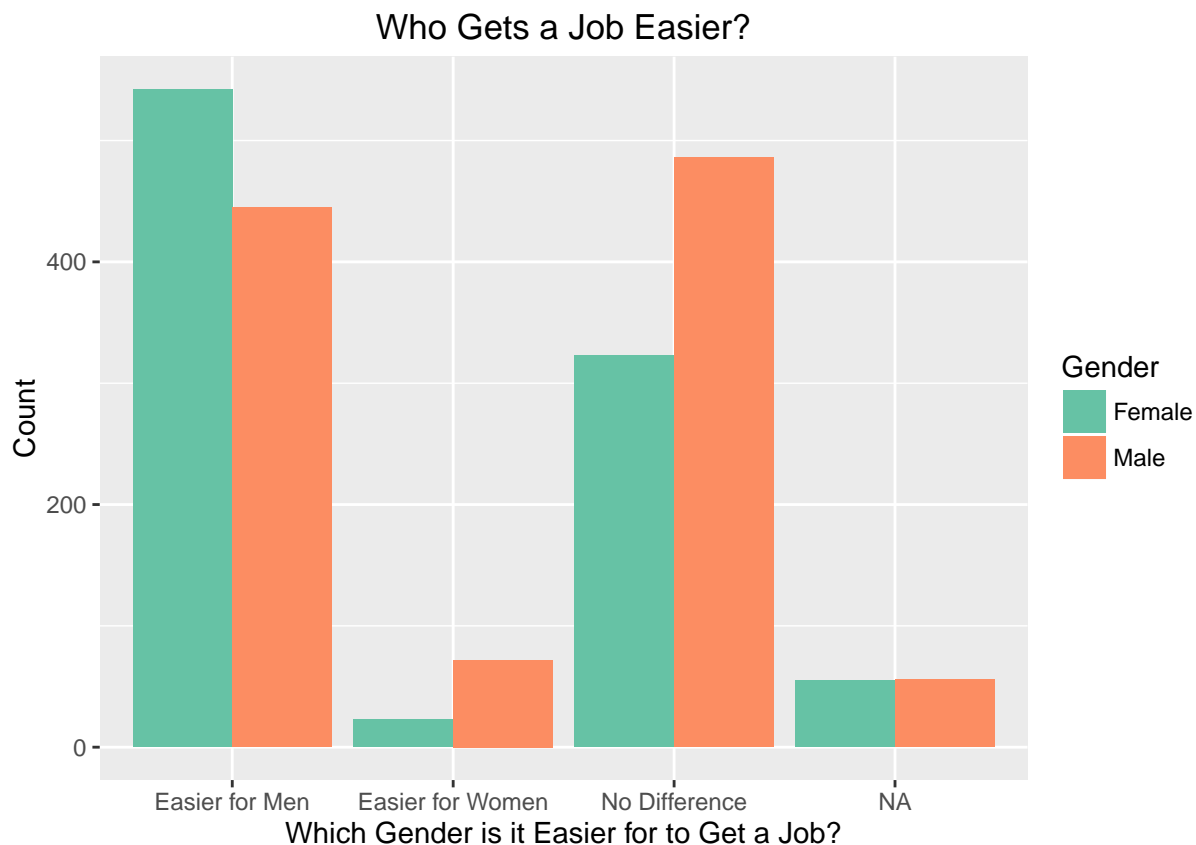
Both men and women think it's easier for the other sex to get a job. Nearly three quarters of all responses that thought it was easier for women to get a job came from male respondents. This illustrates that people also generally think it's easier for others in a different situation than it is for themselves.

```
survey$q45 = Recode(survey$q45, "1='Men Earn More'; 2='Women Earn More'; 3='Both Earn Equal Amounts'; 9=

survey$q45 = as.factor(survey$q45)

survey$q45 = factor(survey$q45, levels(survey$q45)[c(1,2,4,3)])

levels(survey$q45) <- gsub(" ", "\n", levels(survey$q45))

ggplot(survey[!is.na(survey$q45),], aes(x=q45, fill=gender))+geom_bar(stat = "count", position = "dodge
  labs(x = "What Gender Generally Gets Paid More?", y = "Count")+
      scale_fill_brewer(palette= "Set2", name ="Gender")+
  ggtitle("Who Gets Paid More?")
```
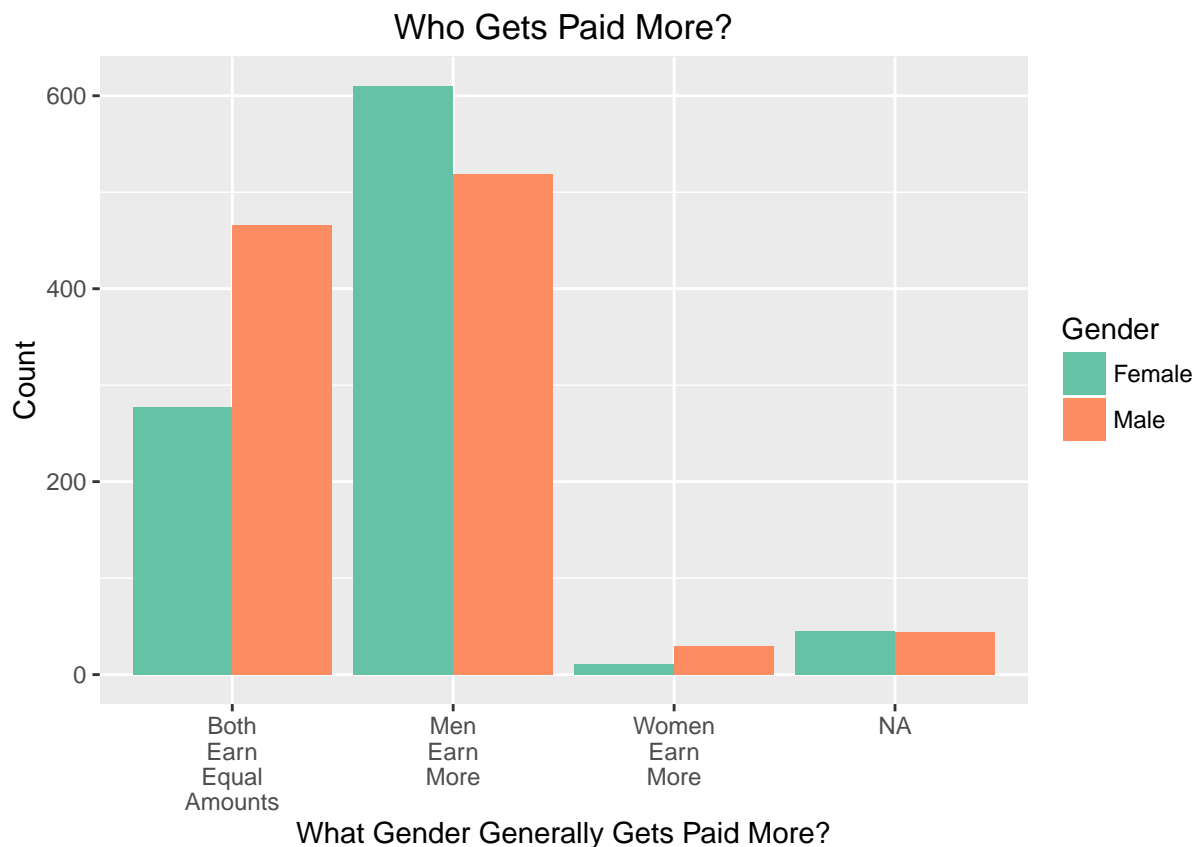
Who Gets Paid More?

What Gender Generally Gets Paid More?

Additionally, it appears that each gender also believes the opposite sex gets paid more. From looking at this graph in combination with others above, it leads one to believe that each gender has confidence in themselves as they think they're more focused on their careers, but at the same time they believe the opposite gender gets a job more easily and is paid more. This could possibly be explained as a type of underdog mentality where people are demonstrating confidence in themselves and seeing others as the more favored group. In contrast, it is possible that people are showing that they're having to be more focused on their careers in response to the inequality that they perceive to exist.

```r
survey$educationlvl = Recode(survey$educ2, "1='Less than High School'; 2='High School Incomplete'; 3='H:

survey$educationlvl = as.factor(survey$educationlvl)

survey$educationlvl = factor(survey$educationlvl, levels(survey$educationlvl)[c(4,3,2,7,9,1,8,6,5)])

survey$educationlvl = as.factor(survey$educationlvl)
survey$q4 = as.factor(survey$q4)

levels(survey$educationlvl) <- gsub(" ", "\n", levels(survey$educationlvl))
levels(survey$q4) <- gsub("\n"," ", levels(survey$q4))

ggplot(survey, aes(x=educationlvl, fill=q4))+geom_bar(stat = "count", position = "dodge") +
  labs(x ="Education", y = "Count") +
  ggtitle("Education in Relation to Views on Equality")+
    scale_fill_brewer(palette= "Set2", name ="Views on Equality")+ coord_flip()
```

Education in Relation to Views on Equality
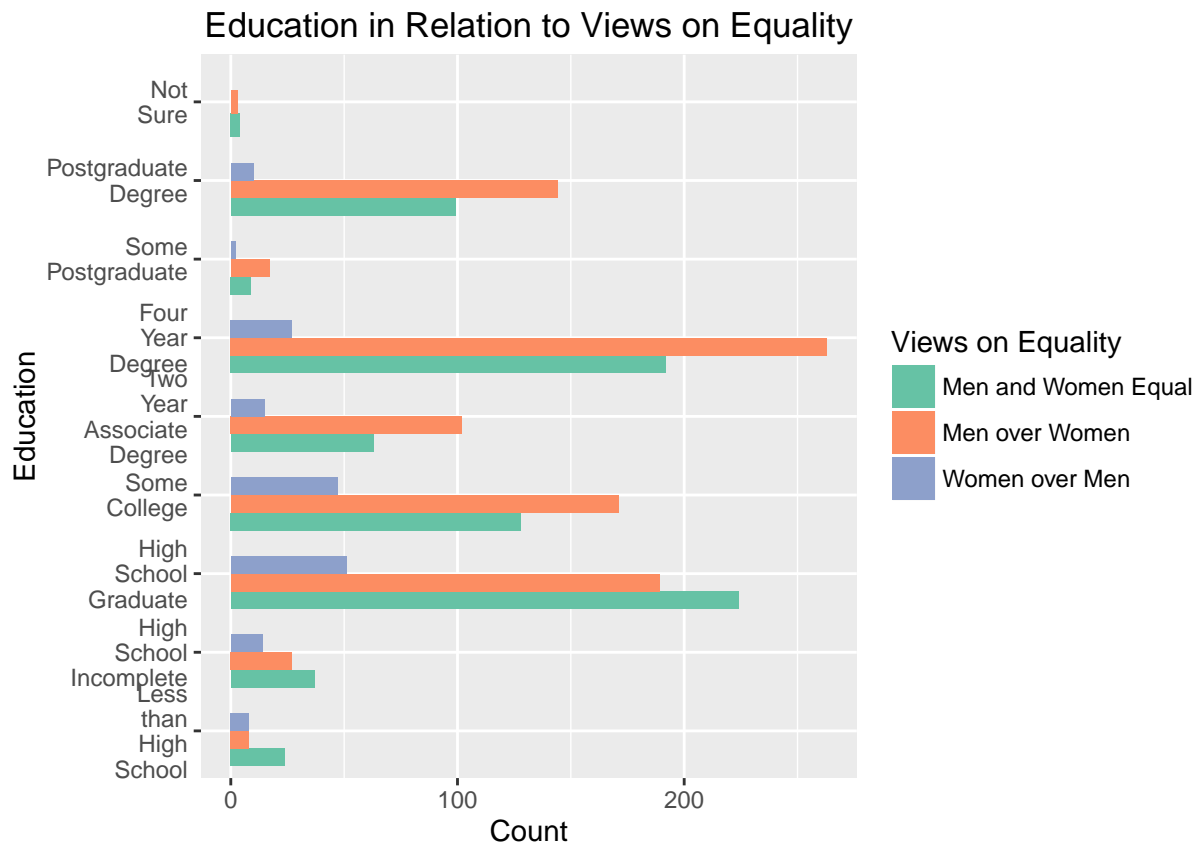
It's also been found that those with lower educational experience are more likely to think men and women are equally treated. Each education level below "Some College" shows that the most prominent view is that men and women are equal. Whereas, that trend flips moving forward with more education as the prominent view becomes "Men over Women". From our sample, it appears that more education leads to a more prominent view that there is inequality in our country.

Limitations

The sample was limiting in that we didn't think it accurately reflected the American public. First off, there were more men in this study then women, but according to the US Census there aer more women then men. Additionally, we believer there were more wealthy individuals sampled than what accurately reflects the US population.

Other limitations of the data was that answers were often subjective and not numeric in meaning. It would have been helpful having more questions where respondents stated their level oof agreement as opposed to choosing discrete options.

Conclusion

From analysing the data it is clear that there is a corrlation between income and happiness. With each increase in income from one bracket to the next, there was a constant trend in increased happiness and decreased unhappiness. We found no meaningful relationship between happiness and views on equality.

We did find meaningful correlation between high levels of debt and unsatisfaction with education. The individuals who felt their education has not paid off were mainly from the group of people who took out high amounts of debt to fund their education.

Gender had significant influence on how respondents viewed the workplace. For starters, they much more likely to ask for a raise if they were male. Also, each gender thinks their own gender is more focused on their career than the other. At the same time, women also felt more change needed to take place to promote equality in the workplace. Finally, each gender also thought it was easier for the opposite gender to get a job and also thought the opposite gender was paid more. This was true for both male and female respondents.

As a last question, we found that in the least educated individuals were the most likely to think there is equality between men and women. There was not significant differences in views of equality based on education beyond "Some College", but for the least educated individuals their most prominent view was that men and women are equal.