Week1 Assignment

Chunjie Nan

28/08/2021

The author of "How Americans Like Their Stake" brings out a question about how does a risk-taking behavior associate with steak in rare. Along with other variables, he found that the the risk-taking behavior is statistically insignificant to steak rareness. https://fivethirtyeight.com/features/how-americans-like-their-steak/

Import data as csv

linked phrase

```
a<-getURL("https://raw.githubusercontent.com/fivethirtyeight/data/master/steak-survey/steak-risk-survey
b<-data.frame(read.csv(text=a, header=T))
head(b)
```

```
##   RespondentID
## 1           NA
## 2   3237565956
## 3   3234982343
## 4   3234973379
## 5   3234972383
## 6   3234958833
##   Consider.the.following.hypothetical.situations...br.In.Lottery.A..you.have.a.50..chance.of.success
## 1
## 2
## 3
## 4
## 5
## 6
##   Do.you.ever.smoke.cigarettes. Do.you.ever.drink.alcohol. Do.you.ever.gamble.
## 1                      Response                  Response            Response
## 2
## 3                            No                       Yes                  No
## 4                            No                       Yes                 Yes
## 5                           Yes                       Yes                 Yes
## 6                            No                       Yes                  No
##   Have.you.ever.been.skydiving. Do.you.ever.drive.above.the.speed.limit.
## 1                      Response                                Response
## 2
## 3                            No                                      No
## 4                            No                                     Yes
## 5                            No                                     Yes
## 6                            No                                     Yes
##   Have.you.ever.cheated.on.your.significant.other. Do.you.eat.steak.
## 1                                         Response          Response
```

```
## 2
## 3                                                 No               Yes
## 4                                                Yes               Yes
## 5                                                Yes               Yes
## 6                                                Yes               Yes
##   How.do.you.like.your.steak.prepared.   Gender      Age  Household.Income
## 1                            Response Response Response          Response
## 2
## 3                         Medium rare     Male     > 60 $50,000 - $99,999
## 4                                Rare     Male     > 60         $150,000+
## 5                              Medium     Male     > 60 $50,000 - $99,999
## 6                              Medium     Male     > 60 $50,000 - $99,999
##                         Education Location..Census.Region.
## 1                        Response                Response
## 2
## 3 Some college or Associate degree       East North Central
## 4                  Graduate degree           South Atlantic
## 5                  Bachelor degree              New England
## 6                  Graduate degree          Middle Atlantic
```

```
b<-b[-1,-2]  # removed the first row because it is an empty row, and removed the 2nd row which is out o
head(b)
```

```
##   RespondentID Do.you.ever.smoke.cigarettes. Do.you.ever.drink.alcohol.
## 2   3237565956
## 3   3234982343                            No                        Yes
## 4   3234973379                            No                        Yes
## 5   3234972383                           Yes                        Yes
## 6   3234958833                            No                        Yes
## 7   3234955240                            No                         No
##   Do.you.ever.gamble. Have.you.ever.been.skydiving.
## 2
## 3                  No                           No
## 4                 Yes                           No
## 5                 Yes                           No
## 6                  No                           No
## 7                  No                           No
##   Do.you.ever.drive.above.the.speed.limit.
## 2
## 3                                       No
## 4                                      Yes
## 5                                      Yes
## 6                                      Yes
## 7                                      Yes
##   Have.you.ever.cheated.on.your.significant.other. Do.you.eat.steak.
## 2
## 3                                                No               Yes
## 4                                               Yes               Yes
## 5                                               Yes               Yes
## 6                                               Yes               Yes
## 7                                                No               Yes
##   How.do.you.like.your.steak.prepared. Gender   Age  Household.Income
## 2
## 3                         Medium rare   Male  > 60 $50,000 - $99,999
## 4                                Rare   Male  > 60         $150,000+
```

```
## 5                              Medium   Male  > 60 $50,000 - $99,999
## 6                              Medium   Male  > 60 $50,000 - $99,999
## 7                         Medium rare   Male 18-29      $0 - $24,999
##                          Education Location..Census.Region.
## 2
## 3 Some college or Associate degree       East North Central
## 4                  Graduate degree           South Atlantic
## 5                  Bachelor degree              New England
## 6                  Graduate degree          Middle Atlantic
## 7 Some college or Associate degree       West South Central
```

```
dim(b)
```

```
## [1] 550  14
```

Rename the column headers

```
data<-rename(b,c("RespondentID"="ID","Do.you.ever.smoke.cigarettes."="Cigarettes","Do.you.ever.drink.al
```

```
## The following `from` values were not present in `x`: VAge
```

```
head(data)
```

```
##           ID Cigarettes Alcohol Gamble Skydiving Drive_limit Cheat Steak
## 2 3237565956
## 3 3234982343         No     Yes     No        No          No    No   Yes
## 4 3234973379         No     Yes    Yes        No         Yes   Yes   Yes
## 5 3234972383        Yes     Yes    Yes        No         Yes   Yes   Yes
## 6 3234958833         No     Yes     No        No         Yes   Yes   Yes
## 7 3234955240         No      No     No        No         Yes    No   Yes
##      prepared Gender   Age      Income_range                          Education
## 2
## 3 Medium rare   Male  > 60 $50,000 - $99,999 Some college or Associate degree
## 4        Rare   Male  > 60         $150,000+                  Graduate degree
## 5      Medium   Male  > 60 $50,000 - $99,999                  Bachelor degree
## 6      Medium   Male  > 60 $50,000 - $99,999                  Graduate degree
## 7 Medium rare   Male 18-29      $0 - $24,999 Some college or Associate degree
##               Region
## 2
## 3 East North Central
## 4     South Atlantic
## 5        New England
## 6    Middle Atlantic
## 7 West South Central
```

```
str(data)
```

```
## 'data.frame':    550 obs. of  14 variables:
##  $ ID          : num  3.24e+09 3.23e+09 3.23e+09 3.23e+09 3.23e+09 ...
##  $ Cigarettes  : chr  "" "No" "No" "Yes" ...
##  $ Alcohol     : chr  "" "Yes" "Yes" "Yes" ...
##  $ Gamble      : chr  "" "No" "Yes" "Yes" ...
##  $ Skydiving   : chr  "" "No" "No" "No" ...
##  $ Drive_limit : chr  "" "No" "Yes" "Yes" ...
##  $ Cheat       : chr  "" "No" "Yes" "Yes" ...
##  $ Steak       : chr  "" "Yes" "Yes" "Yes" ...
##  $ prepared    : chr  "" "Medium rare" "Rare" "Medium" ...
```

```
##  $ Gender      : chr  "" "Male" "Male" "Male" ...
##  $ Age         : chr  "" "> 60" "> 60" "> 60" ...
##  $ Income_range: chr  "" "$50,000 - $99,999" "$150,000+" "$50,000 - $99,999" ...
##  $ Education   : chr  "" "Some college or Associate degree" "Graduate degree" "Bachelor degree" ...
##  $ Region      : chr  "" "East North Central" "South Atlantic" "New England" ...
```

```
data$ID<-as.character(data$ID)   # change the ID to Character, since the number in ID is not meaningful
str(data)
```

```
## 'data.frame':    550 obs. of  14 variables:
##  $ ID          : chr  "3237565956" "3234982343" "3234973379" "3234972383" ...
##  $ Cigarettes  : chr  "" "No" "No" "Yes" ...
##  $ Alcohol     : chr  "" "Yes" "Yes" "Yes" ...
##  $ Gamble      : chr  "" "No" "Yes" "Yes" ...
##  $ Skydiving   : chr  "" "No" "No" "No" ...
##  $ Drive_limit : chr  "" "No" "Yes" "Yes" ...
##  $ Cheat       : chr  "" "No" "Yes" "Yes" ...
##  $ Steak       : chr  "" "Yes" "Yes" "Yes" ...
##  $ prepared    : chr  "" "Medium rare" "Rare" "Medium" ...
##  $ Gender      : chr  "" "Male" "Male" "Male" ...
##  $ Age         : chr  "" "> 60" "> 60" "> 60" ...
##  $ Income_range: chr  "" "$50,000 - $99,999" "$150,000+" "$50,000 - $99,999" ...
##  $ Education   : chr  "" "Some college or Associate degree" "Graduate degree" "Bachelor degree" ...
##  $ Region      : chr  "" "East North Central" "South Atlantic" "New England" ...
```

Data cleaning

```
# I have seen some problems in this data set. There  are empty/blank within the variable of Gender and
# My idea here is convert the blank column to NA, then drop the row that has NA from the data set.


summary(as.factor(data$Gender))  # There are 36 missing in Gender.
```

```
##        Female   Male
##     36    268    246
```

```
summary(as.factor(data$prepared)) # There are 118 missing in Prepared.
```

```
##                  Medium Medium rare Medium Well        Rare        Well
##         118         132         166          75          23          36
```

```
missing_Gender<-data$Gender ==""     # Make a new variable missing in gender.
data$missing_Gender<-ifelse(missing_Gender==TRUE, NA,"Keep")  # convert the empty column to NA
missing_Prepared<-data$prepared =="" # Make a new variable missing in prepared.
data$missing_Prepared<-ifelse(missing_Prepared==TRUE, NA, "Keep")  #convert the empty column to NA


data<-data[!is.na(data$missing_Gender),]
data<-data[!is.na(data$missing_Prepared),]

# now I can see the rows are dropped successfully, and the data observation is dropped from 550 to 412.
```

How many female like the steak in rare?

```
steak_rare_female<- subset(data, Gender=="Female" & prepared =="Rare")
steak_rare_female
```

```
##             ID Cigarettes Alcohol Gamble Skydiving Drive_limit Cheat Steak
## 153 3234870354        Yes      No     No         No          No    No   Yes
## 195 3234854914         No     Yes    Yes         No         Yes   Yes   Yes
## 199 3234853259         No     Yes     No         No         Yes   Yes   Yes
## 258 3234839680        Yes     Yes    Yes         No         Yes    No   Yes
## 262 3234839073         No     Yes    Yes         No         Yes    No   Yes
## 272 3234836496         No     Yes    Yes         No         Yes    No   Yes
## 280 3234834715         No     Yes     No         No         Yes    No   Yes
## 283 3234834323         No     Yes     No         No         Yes   Yes   Yes
## 345 3234817851         No      No    Yes         No         Yes    No   Yes
## 396 3234804474         No      No     No        Yes         Yes    No   Yes
## 447 3234789052         No     Yes     No         No         Yes   Yes   Yes
## 467 3234784673         No     Yes     No         No         Yes    No   Yes
##     prepared Gender   Age        Income_range                        Education
## 153     Rare Female  > 60        $0 - $24,999 Some college or Associate degree
## 195     Rare Female 30-44           $150,000+ Some college or Associate degree
## 199     Rare Female  > 60   $25,000 - $49,999                  Graduate degree
## 258     Rare Female 30-44   $50,000 - $99,999 Some college or Associate degree
## 262     Rare Female  > 60   $50,000 - $99,999 Some college or Associate degree
## 272     Rare Female 45-60 $100,000 - $149,999                  Graduate degree
## 280     Rare Female  > 60 $100,000 - $149,999                  Bachelor degree
## 283     Rare Female 45-60                                      Bachelor degree
## 345     Rare Female 30-44        $0 - $24,999 Some college or Associate degree
## 396     Rare Female 45-60   $50,000 - $99,999                  Bachelor degree
## 447     Rare Female 45-60   $50,000 - $99,999                  Graduate degree
## 467     Rare Female 18-29   $50,000 - $99,999                  Bachelor degree
##                 Region missing_Gender missing_Prepared
## 153    Middle Atlantic           Keep             Keep
## 195            Pacific           Keep             Keep
## 199            Pacific           Keep             Keep
## 258 West North Central           Keep             Keep
## 262            Pacific           Keep             Keep
## 272     South Atlantic           Keep             Keep
## 280    Middle Atlantic           Keep             Keep
## 283    Middle Atlantic           Keep             Keep
## 345     South Atlantic           Keep             Keep
## 396     South Atlantic           Keep             Keep
## 447 West North Central           Keep             Keep
## 467        New England           Keep             Keep
```

```
count(steak_rare_female$Gender) # 12 female like the steak rare
```

```
##        x freq
## 1 Female   12
```

```
count(data$Gender=="Female")# There are 200 female in this data set.
```

```
##       x freq
## 1 FALSE  212
## 2  TRUE  200
```

```
12/200      # about 6 % of female in this data like the steak in rare.
```

```
## [1] 0.06
```

Therefore, there are 12 female like the steak rare which is about 6 % of female in this data like the steak in rare.

How many male like the steak in rare?

```
##             ID Cigarettes Alcohol Gamble Skydiving Drive_limit Cheat Steak
## 4   3234973379         No     Yes    Yes         No         Yes   Yes   Yes
## 15  3234946739        Yes     Yes    Yes         No         Yes    No   Yes
## 77  3234898344         No     Yes     No         No         Yes    No   Yes
## 89  3234894892         No     Yes     No         No         Yes    No   Yes
## 90  3234894723        Yes     Yes    Yes         No         Yes    No   Yes
## 121 3234882580         No     Yes    Yes         No         Yes    No   Yes
## 168 3234861097                Yes     No         No         Yes    No   Yes
## 251 3234840931         No     Yes    Yes         No         Yes    No   Yes
## 462 3234785120        Yes     Yes    Yes         No         Yes    No   Yes
## 488 3234778651         No     Yes     No         No         Yes    No   Yes
##     prepared Gender   Age        Income_range                     Education
## 4       Rare   Male  > 60           $150,000+                 Graduate degree
## 15      Rare   Male 18-29   $50,000 - $99,999                 Graduate degree
## 77      Rare   Male  > 60                                     Graduate degree
## 89      Rare   Male 30-44   $25,000 - $49,999                 Bachelor degree
## 90      Rare   Male  > 60                                     Bachelor degree
## 121     Rare   Male 45-60 $100,000 - $149,999                 Bachelor degree
## 168     Rare   Male 18-29   $50,000 - $99,999              High school degree
## 251     Rare   Male 45-60 $100,000 - $149,999                 Graduate degree
## 462     Rare   Male 18-29   $25,000 - $49,999 Some college or Associate degree
## 488     Rare   Male  > 60           $150,000+                 Graduate degree
##                  Region missing_Gender missing_Prepared
## 4       South Atlantic           Keep             Keep
## 15             Pacific           Keep             Keep
## 77          New England           Keep             Keep
## 89             Pacific           Keep             Keep
## 90     Middle Atlantic           Keep             Keep
## 121    Middle Atlantic           Keep             Keep
## 168 East North Central           Keep             Keep
## 251    Middle Atlantic           Keep             Keep
## 462 West North Central           Keep             Keep
## 488     South Atlantic           Keep             Keep
```

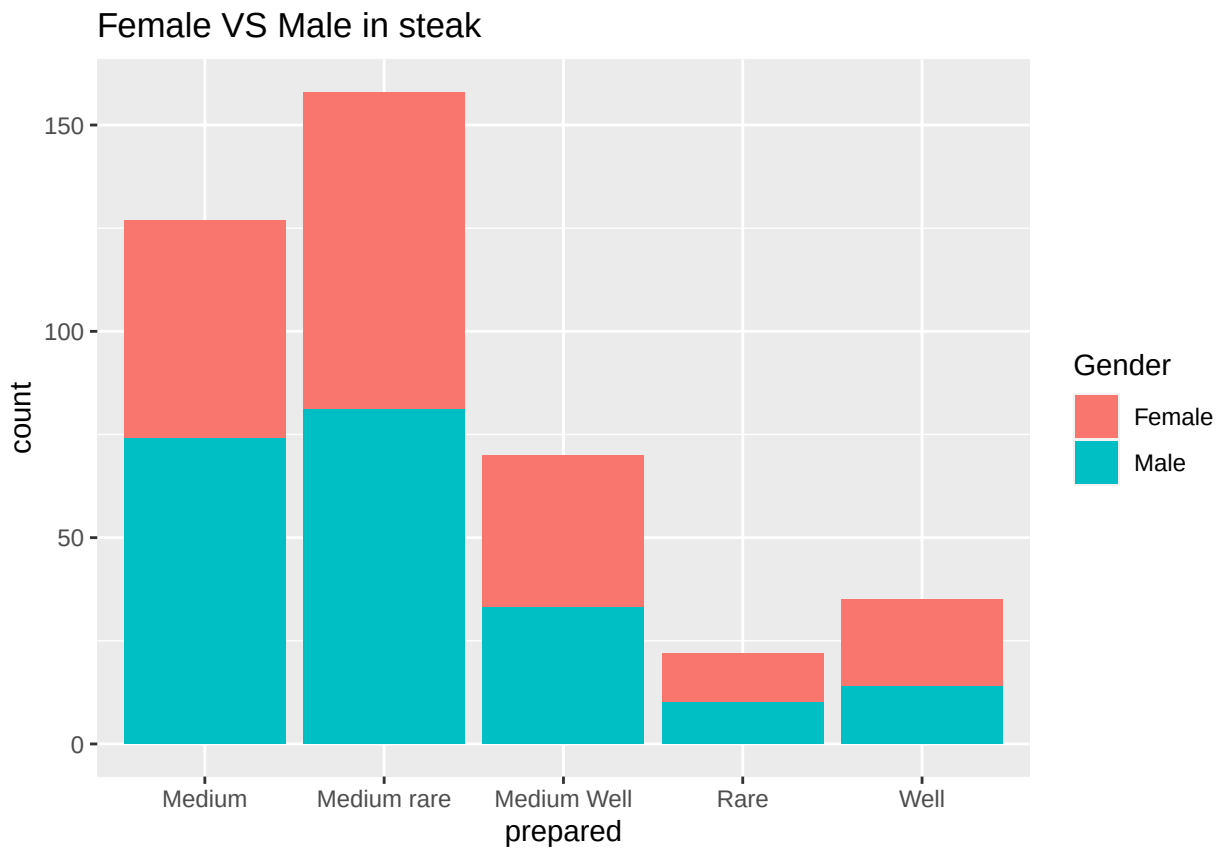
```
##      x freq
## 1 Male   10
```

```
##       x freq
## 1 FALSE  200
## 2  TRUE  212
```

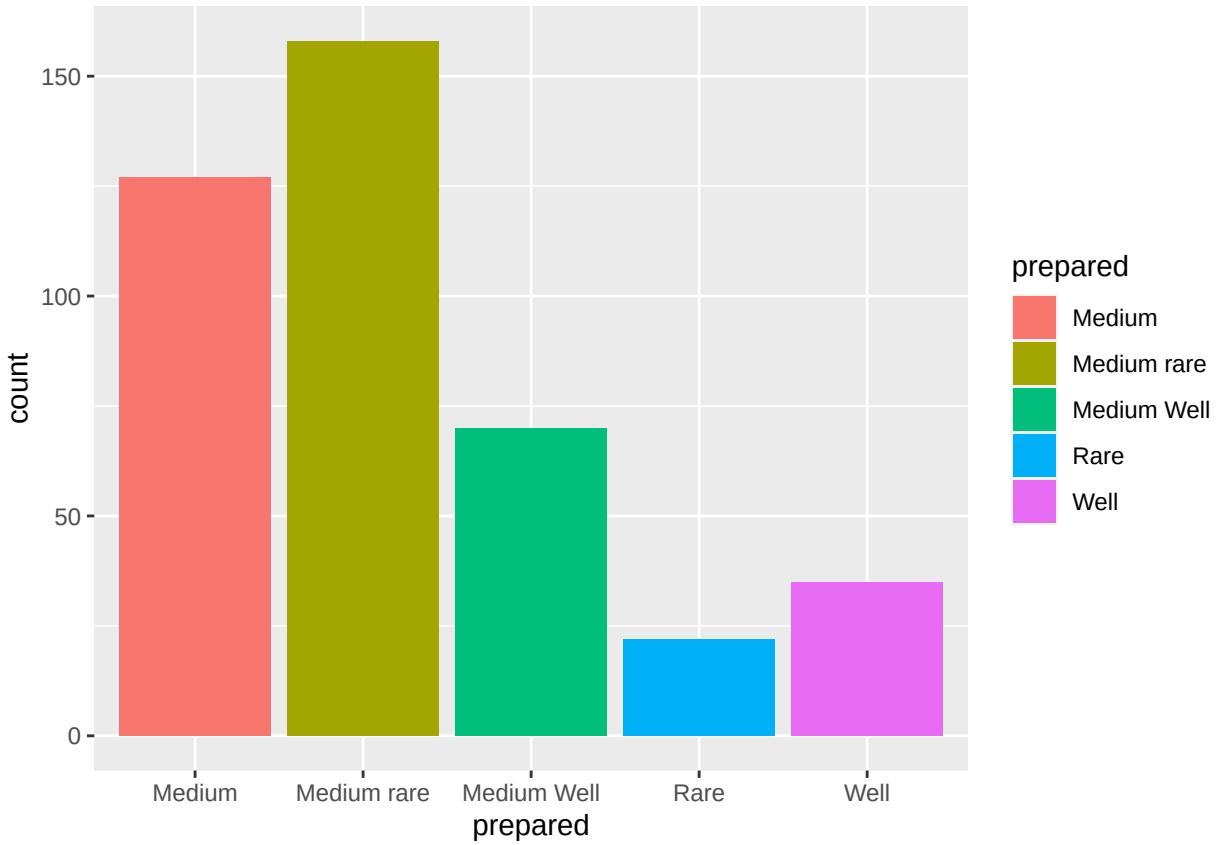
```
## [1] 0.04716981
```

There are 10 male like the steak in rare which means about 4.7 % of the male like the steak in rare in this data set.

```
library(ggplot2)
ggplot(data=data)+
  geom_bar(mapping=aes(x=prepared,fill=Gender))+
  ggtitle("Female VS Male in steak")
```




```
ggplot(data=data)+
  geom_bar(mapping=aes(x=prepared,fill=prepared))
```



From above two plots, we can see that Female and Male have the similar pattern in make their steak prepared. The proportions are very close to each other.

How many steak lovers like the stake in rare?

```
##             ID Cigarettes Alcohol Gamble Skydiving Drive_limit Cheat Steak
## 4   3234973379         No     Yes    Yes         No         Yes   Yes   Yes
## 15  3234946739        Yes     Yes    Yes         No         Yes    No   Yes
## 77  3234898344         No     Yes     No         No         Yes    No   Yes
## 89  3234894892         No     Yes     No         No         Yes    No   Yes
## 90  3234894723        Yes     Yes    Yes         No         Yes    No   Yes
## 121 3234882580         No     Yes    Yes         No         Yes    No   Yes
## 153 3234870354        Yes      No     No         No          No    No   Yes
## 168 3234861097                Yes     No         No         Yes    No   Yes
## 195 3234854914         No     Yes    Yes         No         Yes   Yes   Yes
## 199 3234853259         No     Yes     No         No         Yes   Yes   Yes
## 251 3234840931         No     Yes    Yes         No         Yes    No   Yes
## 258 3234839680        Yes     Yes    Yes         No         Yes    No   Yes
## 262 3234839073         No     Yes    Yes         No         Yes    No   Yes
## 272 3234836496         No     Yes    Yes         No         Yes    No   Yes
## 280 3234834715         No     Yes     No         No         Yes    No   Yes
## 283 3234834323         No     Yes     No         No         Yes   Yes   Yes
## 345 3234817851         No      No    Yes         No         Yes    No   Yes
## 396 3234804474         No      No     No        Yes         Yes    No   Yes
## 447 3234789052         No     Yes     No         No         Yes   Yes   Yes
## 462 3234785120        Yes     Yes    Yes         No         Yes    No   Yes
## 467 3234784673         No     Yes     No         No         Yes    No   Yes
## 488 3234778651         No     Yes     No         No         Yes    No   Yes
```

```
##     prepared Gender   Age        Income_range                    Education
## 4       Rare   Male  > 60           $150,000+                Graduate degree
## 15      Rare   Male 18-29   $50,000 - $99,999                Graduate degree
## 77      Rare   Male  > 60                                    Graduate degree
## 89      Rare   Male 30-44   $25,000 - $49,999                Bachelor degree
## 90      Rare   Male  > 60                                    Bachelor degree
## 121     Rare   Male 45-60 $100,000 - $149,999                Bachelor degree
## 153     Rare Female  > 60        $0 - $24,999 Some college or Associate degree
## 168     Rare   Male 18-29   $50,000 - $99,999               High school degree
## 195     Rare Female 30-44           $150,000+ Some college or Associate degree
## 199     Rare Female  > 60   $25,000 - $49,999                Graduate degree
## 251     Rare   Male 45-60 $100,000 - $149,999                Graduate degree
## 258     Rare Female 30-44   $50,000 - $99,999 Some college or Associate degree
## 262     Rare Female  > 60   $50,000 - $99,999 Some college or Associate degree
## 272     Rare Female 45-60 $100,000 - $149,999                Graduate degree
## 280     Rare Female  > 60 $100,000 - $149,999                Bachelor degree
## 283     Rare Female 45-60                                    Bachelor degree
## 345     Rare Female 30-44        $0 - $24,999 Some college or Associate degree
## 396     Rare Female 45-60   $50,000 - $99,999                Bachelor degree
## 447     Rare Female 45-60   $50,000 - $99,999                Graduate degree
## 462     Rare   Male 18-29   $25,000 - $49,999 Some college or Associate degree
## 467     Rare Female 18-29   $50,000 - $99,999                Bachelor degree
## 488     Rare   Male  > 60           $150,000+                Graduate degree
##                  Region missing_Gender missing_Prepared
## 4       South Atlantic           Keep             Keep
## 15             Pacific           Keep             Keep
## 77          New England           Keep             Keep
## 89             Pacific           Keep             Keep
## 90      Middle Atlantic           Keep             Keep
## 121     Middle Atlantic           Keep             Keep
## 153     Middle Atlantic           Keep             Keep
## 168 East North Central           Keep             Keep
## 195            Pacific           Keep             Keep
## 199            Pacific           Keep             Keep
## 251     Middle Atlantic           Keep             Keep
## 258 West North Central           Keep             Keep
## 262            Pacific           Keep             Keep
## 272      South Atlantic           Keep             Keep
## 280     Middle Atlantic           Keep             Keep
## 283     Middle Atlantic           Keep             Keep
## 345      South Atlantic           Keep             Keep
## 396      South Atlantic           Keep             Keep
## 447 West North Central           Keep             Keep
## 462 West North Central           Keep             Keep
## 467         New England           Keep             Keep
## 488      South Atlantic           Keep             Keep
```

```
##     x freq
## 1 Yes   22
```

I found that there are 22 steak lovers like eating steak in rare. This is exactly the sum of as above, 10 of male love steak in rare and 12 female love steak in rare.

Conclusion

The data explains that there are 12 females, and 10 males like the steak in rare. Also, we found that people who cook the steak in rare are all steak lovers with total number of 22