

steak_survey

60070016, 60070034

April 17, 2019

รายงานเรื่อง ความนิยมสเต็กในแต่ละระดับความสุก

ชื่อ: ไชยพงศ์ นิธิพิรวัฒน์, ชिरดา ชีร์ธรรม

รหัสนักศึกษา: 60070016, 60070034

เวลาเรียน: อังคาร 9.00-12.00 น.

โจทย์ปัญหา :

ปัจจัยต่าง ๆ ส่งผลต่อการเลือกรับประทานสเต็กอย่างไร ?

ชุดข้อมูลที่ใช้จาก fivethirtyeight : steak_survey

Data set from <https://github.com/fivethirtyeight/data/tree/master/steak-survey>
(<https://github.com/fivethirtyeight/data/tree/master/steak-survey>) by Walt Hickey

1. Load libraries

```
library(fivethirtyeight)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.2.1 --
```

```
## v ggplot2 3.1.0      v purrr   0.3.1
## v tibble  2.0.1      v dplyr   0.8.0.1
## v tidyr   0.8.3      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
```

2. Look at the data (steak_survey)

```
# print out the dataframe
steak_survey
```

```
## # A tibble: 550 x 15
##   respondent_id lottery_a smoke alcohol gamble skydiving speed cheated
##   <dbl> <lgl>    <lgl> <lgl> <lgl> <lgl> <lgl> <lgl>
## 1 3237565956 FALSE    NA    NA    NA    NA    NA    NA
## 2 3234982343 TRUE     FALSE TRUE  FALSE FALSE  FALSE FALSE
## 3 3234973379 TRUE     FALSE TRUE  TRUE  FALSE  TRUE  TRUE
## 4 3234972383 FALSE    TRUE  TRUE  TRUE  FALSE  TRUE  TRUE
## 5 3234958833 FALSE    FALSE TRUE  FALSE FALSE  TRUE  TRUE
## 6 3234955240 TRUE     FALSE FALSE FALSE FALSE  TRUE  FALSE
## 7 3234955097 TRUE     FALSE TRUE  FALSE FALSE  TRUE  TRUE
## 8 3234955010 TRUE     FALSE TRUE  TRUE  TRUE  TRUE  FALSE
## 9 3234953052 TRUE     TRUE  TRUE  TRUE  FALSE  TRUE  FALSE
## 10 3234951249 FALSE    FALSE TRUE  TRUE  FALSE  FALSE FALSE
## # ... with 540 more rows, and 7 more variables: steak <lgl>,
## #   steak_prep <ord>, female <lgl>, age <ord>, hhold_income <fct>,
## #   educ <ord>, region <chr>
```

```
# Approximately look
summary(steak_survey)
```

```
## respondent_id      lottery_a      smoke      alcohol
## Min.      :3.235e+09  Mode :logical  Mode :logical  Mode :logical
## 1st Qu.:3.235e+09  FALSE:279    FALSE:453    FALSE:125
## Median :3.235e+09  TRUE :267    TRUE :84     TRUE :416
## Mean    :3.235e+09  NA's :4      NA's :13     NA's :9
## 3rd Qu.:3.235e+09
## Max.     :3.238e+09
##
## gamble      skydiving      speed      cheated
## Mode :logical  Mode :logical  Mode :logical  Mode :logical
## FALSE:280     FALSE:502     FALSE:59      FALSE:447
## TRUE :257     TRUE :36     TRUE :480     TRUE :92
## NA's :13      NA's :12     NA's :11     NA's :11
##
##
##
## steak      steak_prep      female      age
## Mode :logical  Rare      : 23  Mode :logical  18-29:110
## FALSE:109     Medium rare:166  FALSE:246     30-44:133
## TRUE :430     Medium      :132  TRUE :268     45-60:140
## NA's :11      Medium Well: 75  NA's :36      > 60 :131
##              Well      : 36      NA's : 36
##              NA's      :118
##
##
##              hhold_income      educ
## $0 - $24,999      : 51  Less than high school degree      : 2
## $25,000 - $49,999 : 77  High school degree      : 39
## $50,000 - $99,999 :172  Some college or Associate degree:164
## $100,000 - $149,999: 76  Bachelor degree      :174
## $150,000+      : 54  Graduate degree      :133
## TRUE          : 0    NA's      : 38
## NA's          :120
##
## region
## Length:550
## Class :character
## Mode :character
##
##
##
##
```

3. Looking for answers

```
#Find percent of each steak prepared's type
new_df <- steak_survey %>%
  filter(steak_prep != "NA") %>%
  group_by(steak_prep) %>%
  summarise(amount= n())

new_df
```

```
## # A tibble: 5 x 2
##   steak_prep amount
##   <ord>      <int>
## 1 Rare          23
## 2 Medium rare   166
## 3 Medium        132
## 4 Medium Well   75
## 5 Well          36
```

```
#Separate amount of smoker in each steak prepared's type
smoke_df <- steak_survey %>%
  filter(steak_prep != "NA", smoke != "NA") %>%
  group_by(smoke, steak_prep) %>%
  summarise(amount=n())

smoke_df
```

```
## # A tibble: 10 x 3
## # Groups:   smoke [2]
##   smoke steak_prep amount
##   <lg1> <ord>      <int>
## 1 FALSE Rare          16
## 2 FALSE Medium rare   136
## 3 FALSE Medium        111
## 4 FALSE Medium Well    63
## 5 FALSE Well           30
## 6 TRUE  Rare           6
## 7 TRUE  Medium rare    30
## 8 TRUE  Medium         20
## 9 TRUE  Medium Well    11
## 10 TRUE Well           5
```

```
#Separate amount of drinker in each steak prepared's type
alco_df <- steak_survey %>%
  filter(steak_prep != "NA", alcohol != "NA") %>%
  group_by(alcohol, steak_prep) %>%
  summarise(amount=n())

alco_df
```

```
## # A tibble: 10 x 3
## # Groups:   alcohol [2]
##   alcohol steak_prep amount
##   <lg1>   <ord>      <int>
## 1 FALSE   Rare         3
## 2 FALSE   Medium rare   38
## 3 FALSE   Medium       23
## 4 FALSE   Medium Well   17
## 5 FALSE   Well         12
## 6 TRUE    Rare         20
## 7 TRUE    Medium rare  128
## 8 TRUE    Medium      109
## 9 TRUE    Medium Well   58
## 10 TRUE   Well         24
```

```
#Separate sex in each steak prepared's type
sex_df <- steak_survey %>%
  filter(steak_prep != "NA") %>%
  group_by(female, steak_prep) %>%
  summarise(amount=n())
```

sex_df

```
## # A tibble: 15 x 3
## # Groups:   female [3]
##   female steak_prep amount
##   <lg1>   <ord>      <int>
## 1 FALSE   Rare         10
## 2 FALSE   Medium rare   81
## 3 FALSE   Medium       74
## 4 FALSE   Medium Well   33
## 5 FALSE   Well         14
## 6 TRUE    Rare         12
## 7 TRUE    Medium rare   77
## 8 TRUE    Medium       53
## 9 TRUE    Medium Well   37
## 10 TRUE   Well         21
## 11 NA     Rare         1
## 12 NA     Medium rare   8
## 13 NA     Medium       5
## 14 NA     Medium Well   5
## 15 NA     Well         1
```

```
#Separate by age
age_df <- steak_survey %>%
  filter(steak_prep != "NA", age != "NA") %>%
  group_by(steak_prep, age) %>%
  summarise(amount=n())
age_df
```

```
## # A tibble: 20 x 3
## # Groups:   steak_prep [5]
##   steak_prep age amount
##   <ord>      <ord> <int>
## 1 Rare      18-29     4
## 2 Rare      30-44     4
## 3 Rare      45-60     6
## 4 Rare      > 60      8
## 5 Medium rare 18-29    29
## 6 Medium rare 30-44    40
## 7 Medium rare 45-60    42
## 8 Medium rare > 60    47
## 9 Medium     18-29    30
## 10 Medium     30-44    35
## 11 Medium     45-60    32
## 12 Medium     > 60    30
## 13 Medium Well 18-29    16
## 14 Medium Well 30-44    17
## 15 Medium Well 45-60    18
## 16 Medium Well > 60    19
## 17 Well      18-29     6
## 18 Well      30-44    11
## 19 Well      45-60    13
## 20 Well      > 60     5
```

```
#Separate by household income
hhold_df <- steak_survey %>%
  filter(steak_prep != "NA") %>%
  group_by(steak_prep, hhold_income) %>%
  summarise(amount=n())
```

```
## Warning: Factor `hhold_income` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
hhold_df
```

```
## # A tibble: 30 x 3
## # Groups:   steak_prep [5]
##   steak_prep hhold_income amount
##   <ord>      <fct>      <int>
## 1 Rare      $0 - $24,999      2
## 2 Rare      $25,000 - $49,999  3
## 3 Rare      $50,000 - $99,999  7
## 4 Rare      $100,000 - $149,999 4
## 5 Rare      $150,000+        3
## 6 Rare      <NA>           4
## 7 Medium rare $0 - $24,999     12
## 8 Medium rare $25,000 - $49,999 26
## 9 Medium rare $50,000 - $99,999 53
## 10 Medium rare $100,000 - $149,999 29
## # ... with 20 more rows
```

```
#Separate by education
educate_df <- steak_survey %>%
  filter(steak_prep != "NA", educ != "NA") %>%
  group_by(steak_prep, educ) %>%
  summarise(amount=n())

educate_df
```

```
## # A tibble: 21 x 3
## # Groups:   steak_prep [5]
##   steak_prep educ amount
##   <ord>      <ord>      <int>
## 1 Rare      High school degree 1
## 2 Rare      Some college or Associate degree 6
## 3 Rare      Bachelor degree 7
## 4 Rare      Graduate degree 8
## 5 Medium rare High school degree 14
## 6 Medium rare Some college or Associate degree 51
## 7 Medium rare Bachelor degree 47
## 8 Medium rare Graduate degree 46
## 9 Medium      High school degree 9
## 10 Medium      Some college or Associate degree 39
## # ... with 11 more rows
```

```
#Separate by region
region_df <- steak_survey %>%
  filter(steak_prep != "NA", region != "NA") %>%
  group_by(steak_prep, region) %>%
  summarise(amount=n())

region_df
```

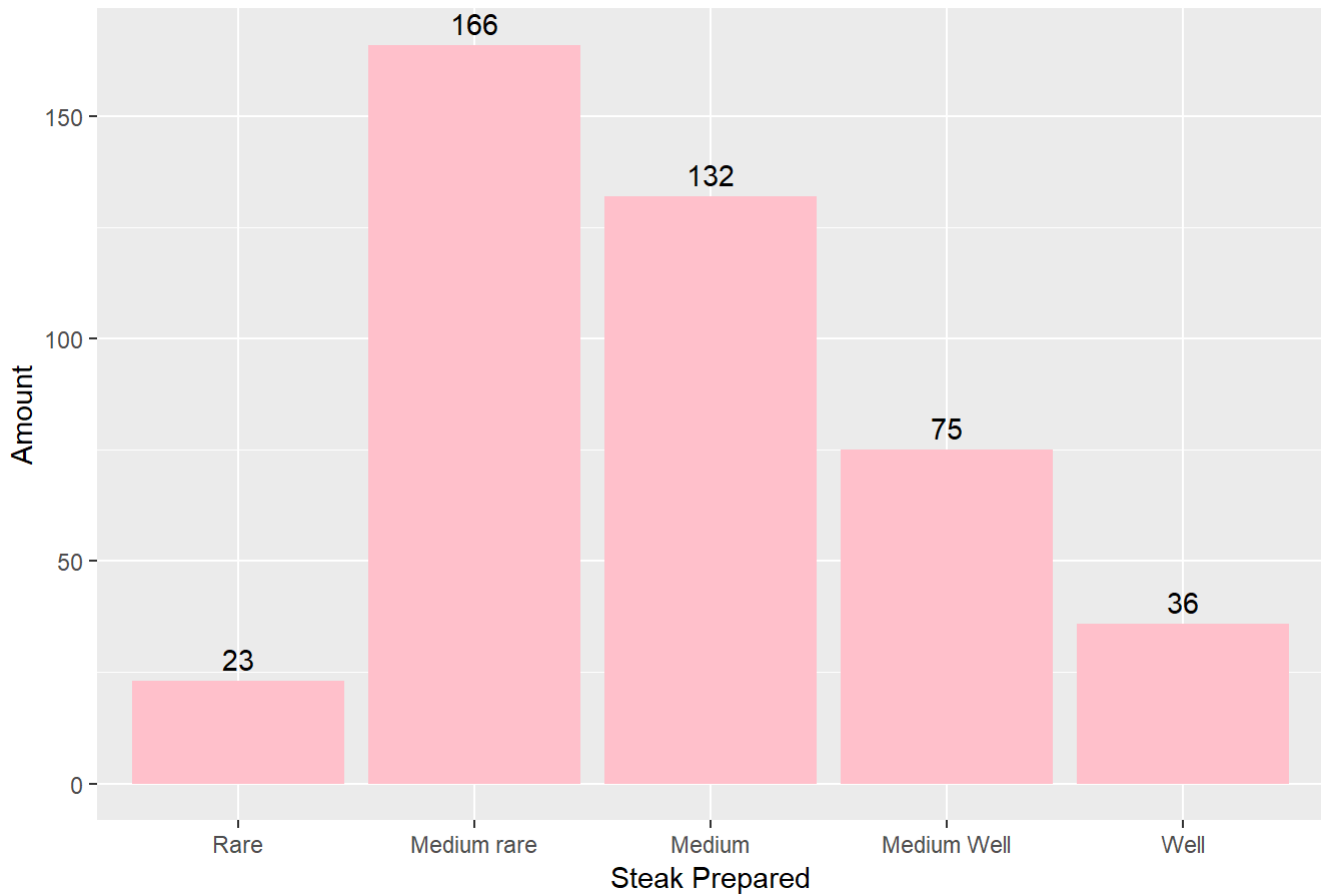
```
## # A tibble: 41 x 3
## # Groups:   steak_prep [5]
##   steak_prep region      amount
##   <ord>      <chr>      <int>
## 1 Rare      East North Central    1
## 2 Rare      Middle Atlantic      6
## 3 Rare      New England          2
## 4 Rare      Pacific              5
## 5 Rare      South Atlantic         5
## 6 Rare      West North Central     3
## 7 Medium rare East North Central   24
## 8 Medium rare East South Central  10
## 9 Medium rare Middle Atlantic    14
## 10 Medium rare Mountain          9
## # ... with 31 more rows
```

4. Visualize data

```
#Plot steak prepared chart
b0 <- ggplot(new_df, aes(x = steak_prep)) +
  # add geometry of the plot (bar chart)
  #colour CR.https://ggplot2.tidyverse.org/reference/aes\_colour\_fill\_alpha.html
  geom_bar(aes(y = amount), stat = "identity", fill = "pink") +
  # add labels
  labs(x="Steak Prepared", y="Amount", title="Steak Prepared Chart") +
  geom_text(aes(label = amount, y = amount), stat= "identity", vjust = -.5)

b0
```

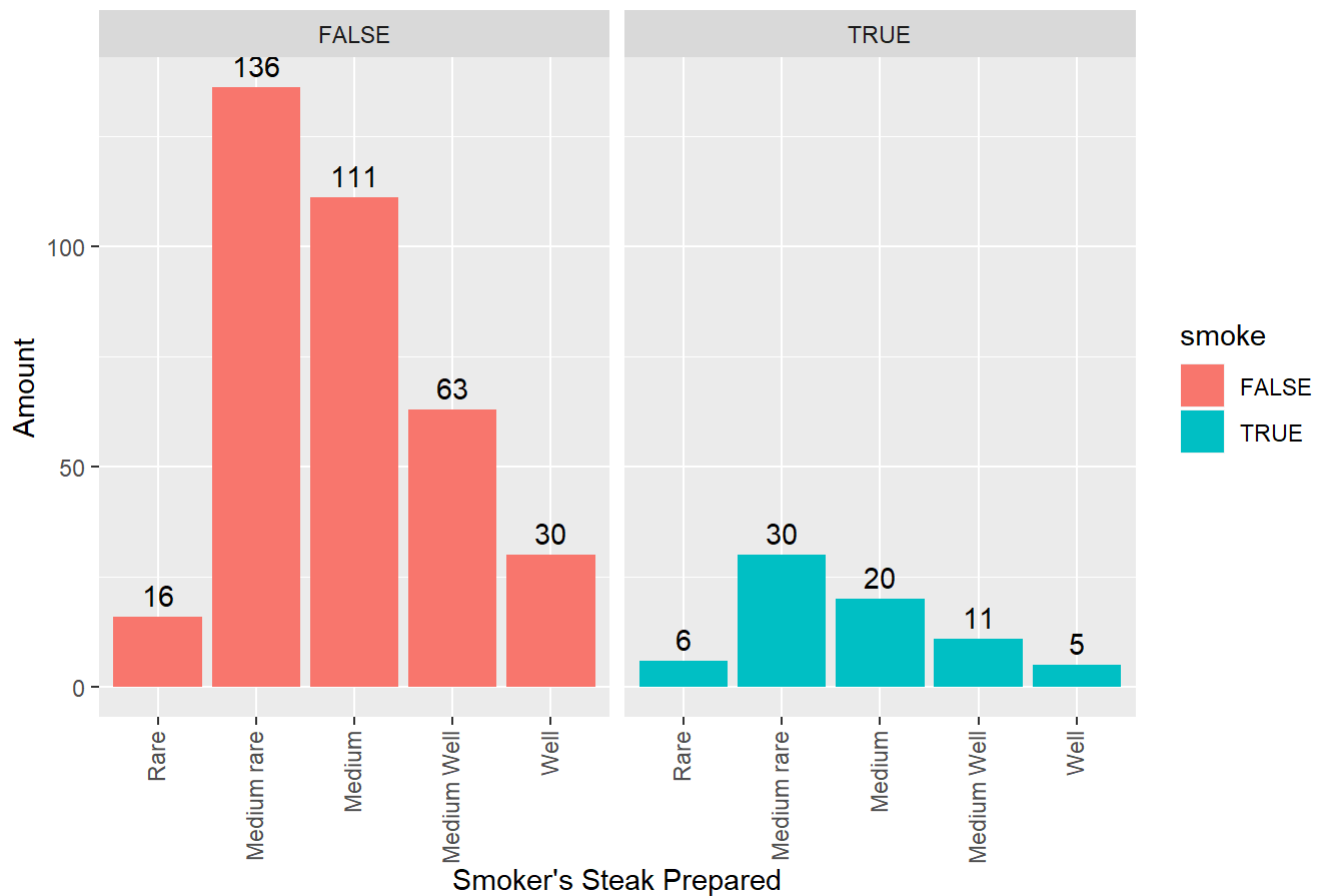

Steak Prepared Chart



```
#Plot bar chart of smoker's steak
b1 <- ggplot(smoke_df, aes(x = steak_prep, fill = smoke)) +
  # add geometry of the plot (bar chart)
  geom_bar(aes(y = amount), stat = "identity") +
  # add labels
  labs(x="Smoker's Steak Prepared", y="Amount", title="Smoker's Steak Prepared Bar Chart")
+
  #Separate charts CR.https://sebastiansauer.github.io/percentage_plot_ggplot2_V2/
  facet_grid(~smoke) +
  geom_text(aes(label = amount, y = amount), stat= "identity", vjust = -.5) +
  #Rotate text under the chart CR.https://stackoverflow.com/questions/1330989/rotating-and
  #spacing-axis-labels-in-ggplot2
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = .35))

b1
```

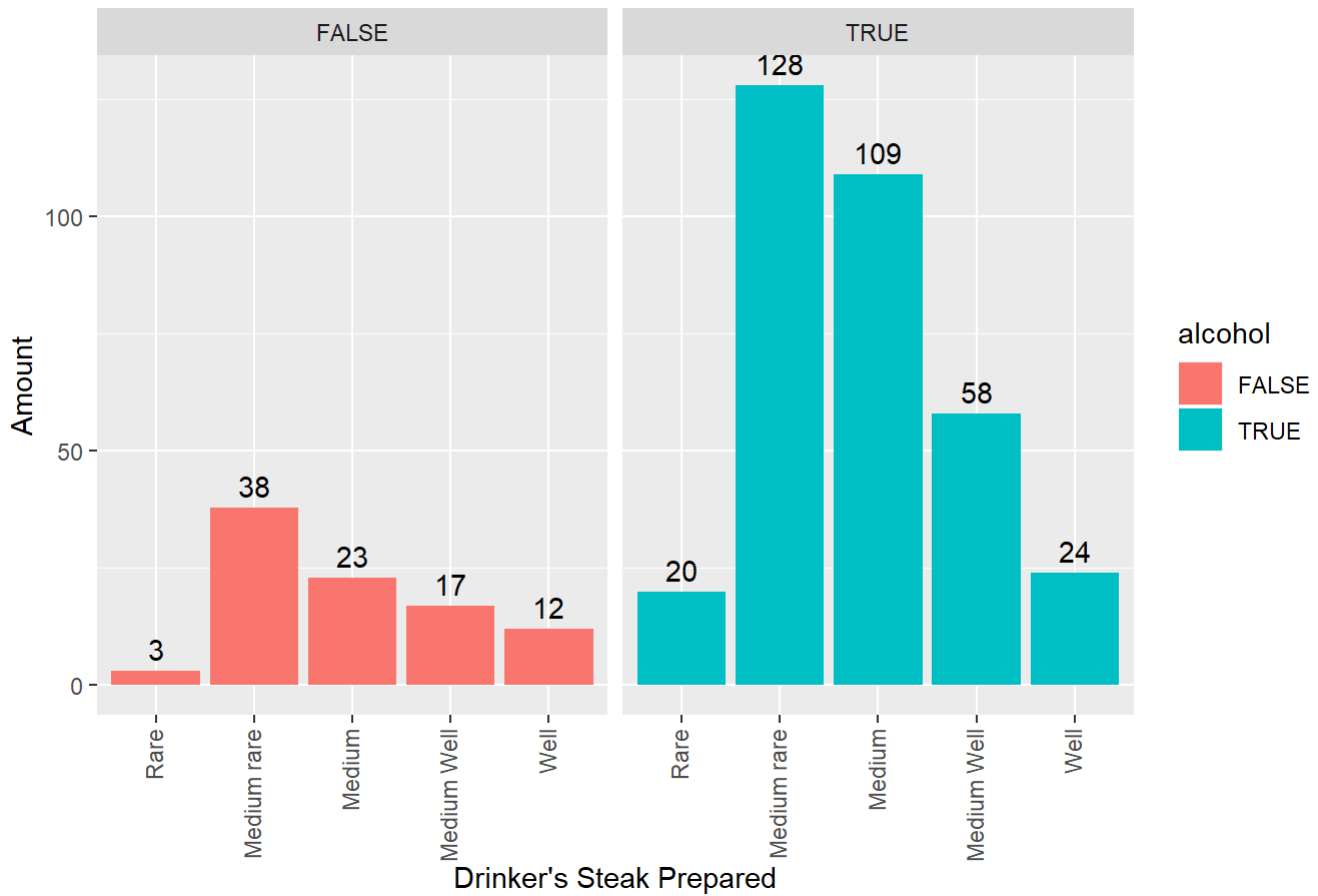
Smoker's Steak Prepared Bar Chart



```
#Plot bar chart of drinker's steak
b2 <- ggplot(alco_df, aes(x = steak_prep, fill = alcohol)) +
  # add geometry of the plot (bar chart)
  geom_bar(aes(y = amount), stat = "identity") +
  # add labels
  labs(x="Drinker's Steak Prepared", y="Amount", title="Drinker's Steak Prepared Bar Chart") +
  facet_grid(~alcohol) +
  geom_text(aes(label = amount, y = amount), stat= "identity", vjust = -.5) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = .35))
```

b2

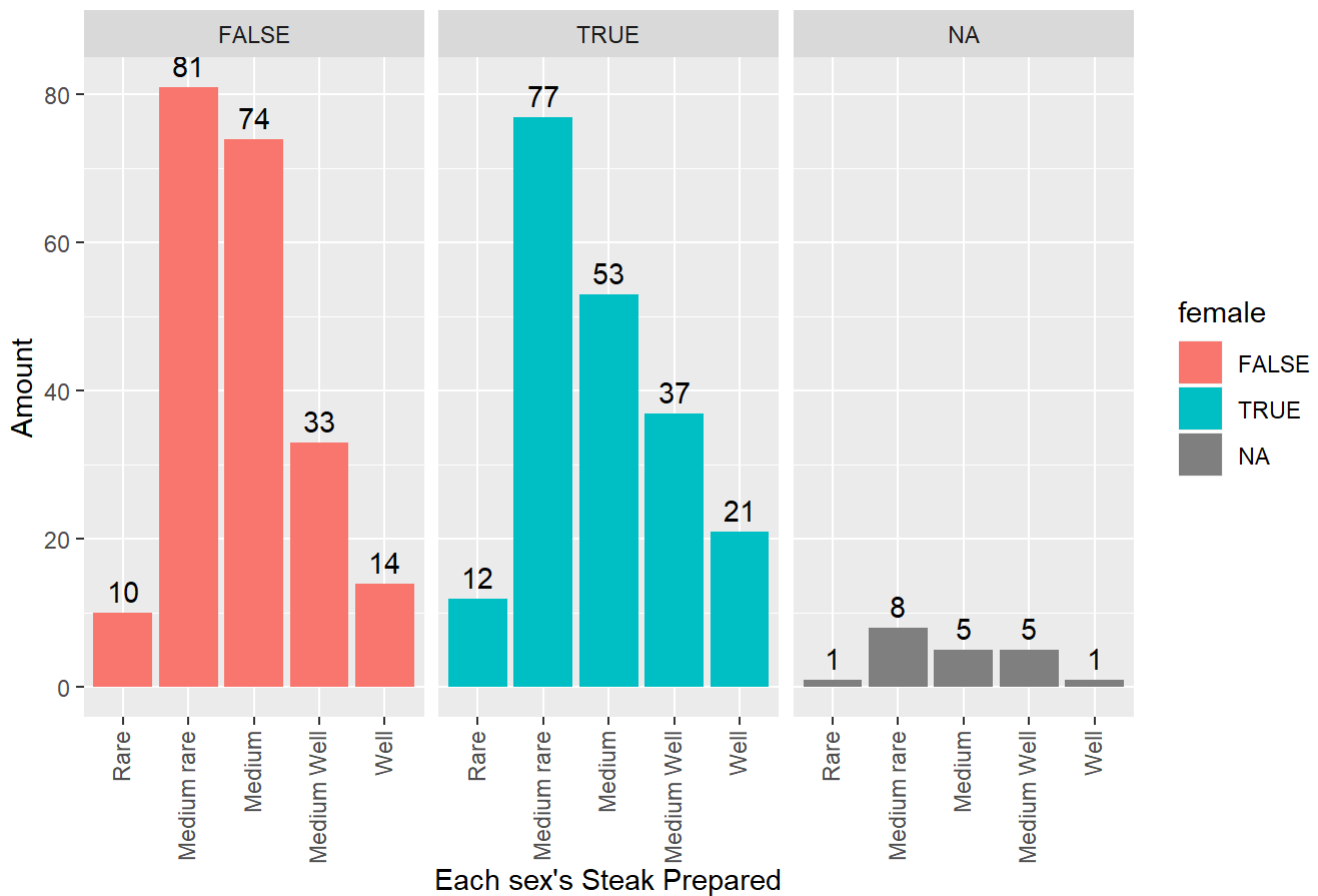
Drinker's Steak Prepared Bar Chart



```
#Plot bar chart of each sex's steak
b3 <- ggplot(sex_df, aes(x = steak_prep, fill = female)) +
  # add geometry of the plot (bar chart)
  geom_bar(aes(y = amount), stat = "identity") +
  # add labels
  labs(x="Each sex's Steak Prepared", y="Amount", title="Each sex's Steak Prepared Bar Chart") +
  facet_grid(~female) +
  geom_text(aes(label = amount, y = amount), stat= "identity", vjust = -.5) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = .35))
```

b3

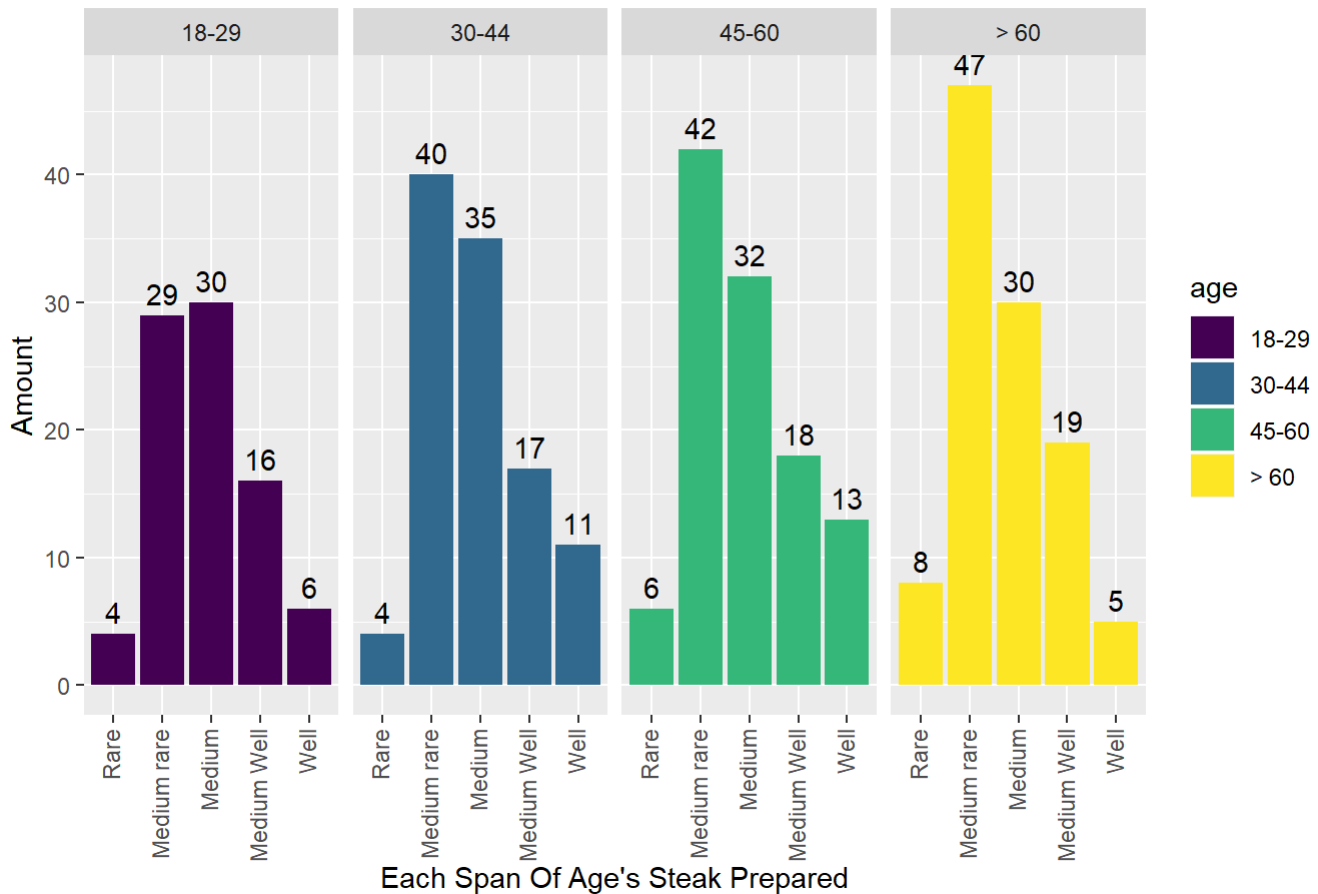
Each sex's Steak Prepared Bar Chart



```
#Plot bar chart of each age's steak
b4 <- ggplot(age_df, aes(x = steak_prep, fill = age)) +
  # add geometry of the plot (bar chart)
  geom_bar(aes(y = amount), stat = "identity") +
  # add Label
  labs(x="Each Span Of Age's Steak Prepared", y="Amount", title="Each Span Of Age's Steak
Prepared Bar Chart")+
  facet_grid(~age) +
  geom_text(aes(label = amount, y = amount), stat= "identity", vjust = -.5) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = .35))
```

b4

Each Span Of Age's Steak Prepared Bar Chart

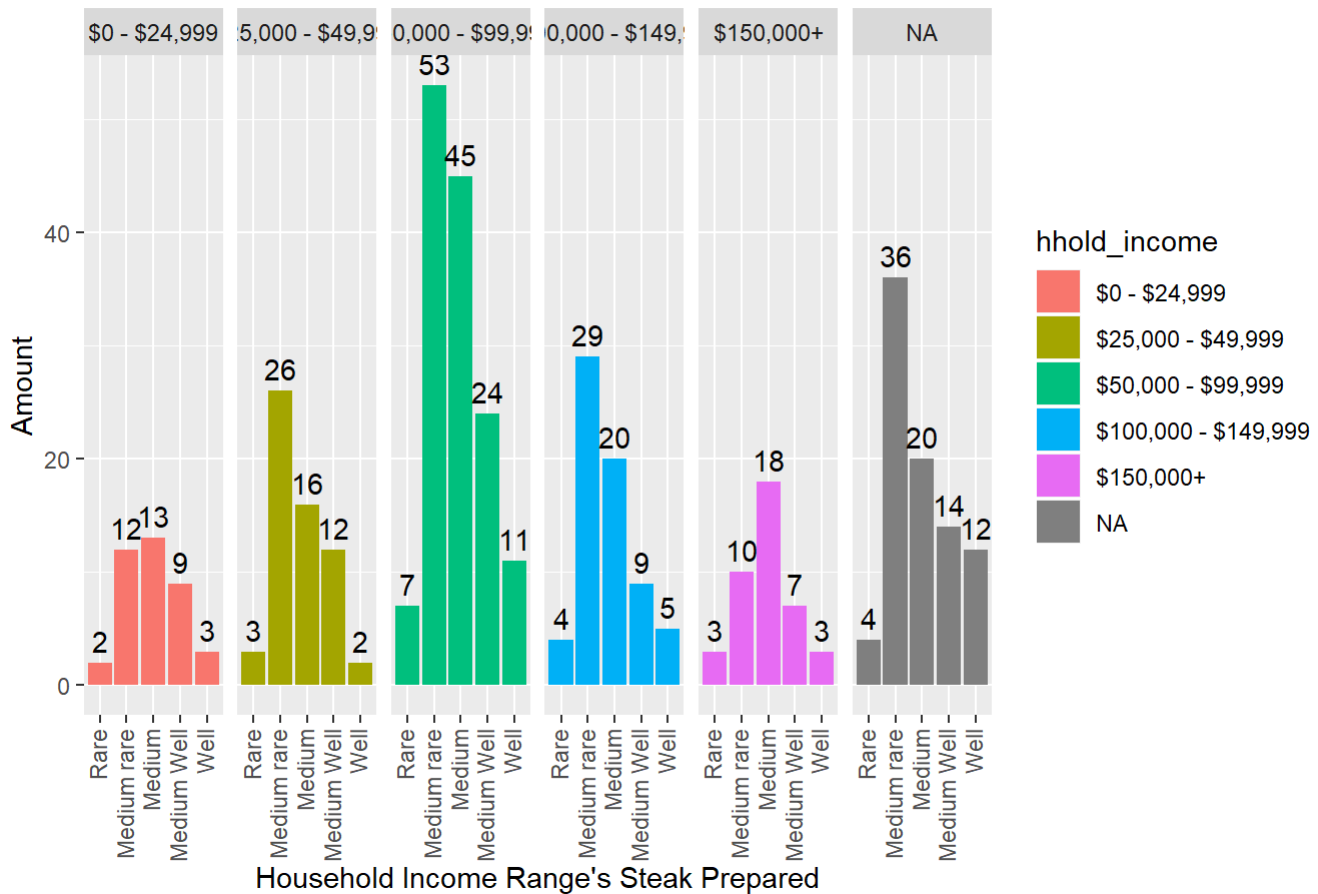


#Plot bar chart of household income's steak

```
b5 <- ggplot(hhold_df, aes(x = steak_prep, fill = hhold_income)) +
  # add geometry of the plot (bar chart)
  geom_bar(aes(y = amount), stat = "identity") +
  # add labels
  labs(x="Household Income Range's Steak Prepared", y="Amount", title="Household Income Range's Steak Prepared Bar Chart") +
  facet_grid(~hhold_income) +
  geom_text(aes(label = amount, y = amount), stat= "identity", vjust = -.5) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = .35))
```

b5

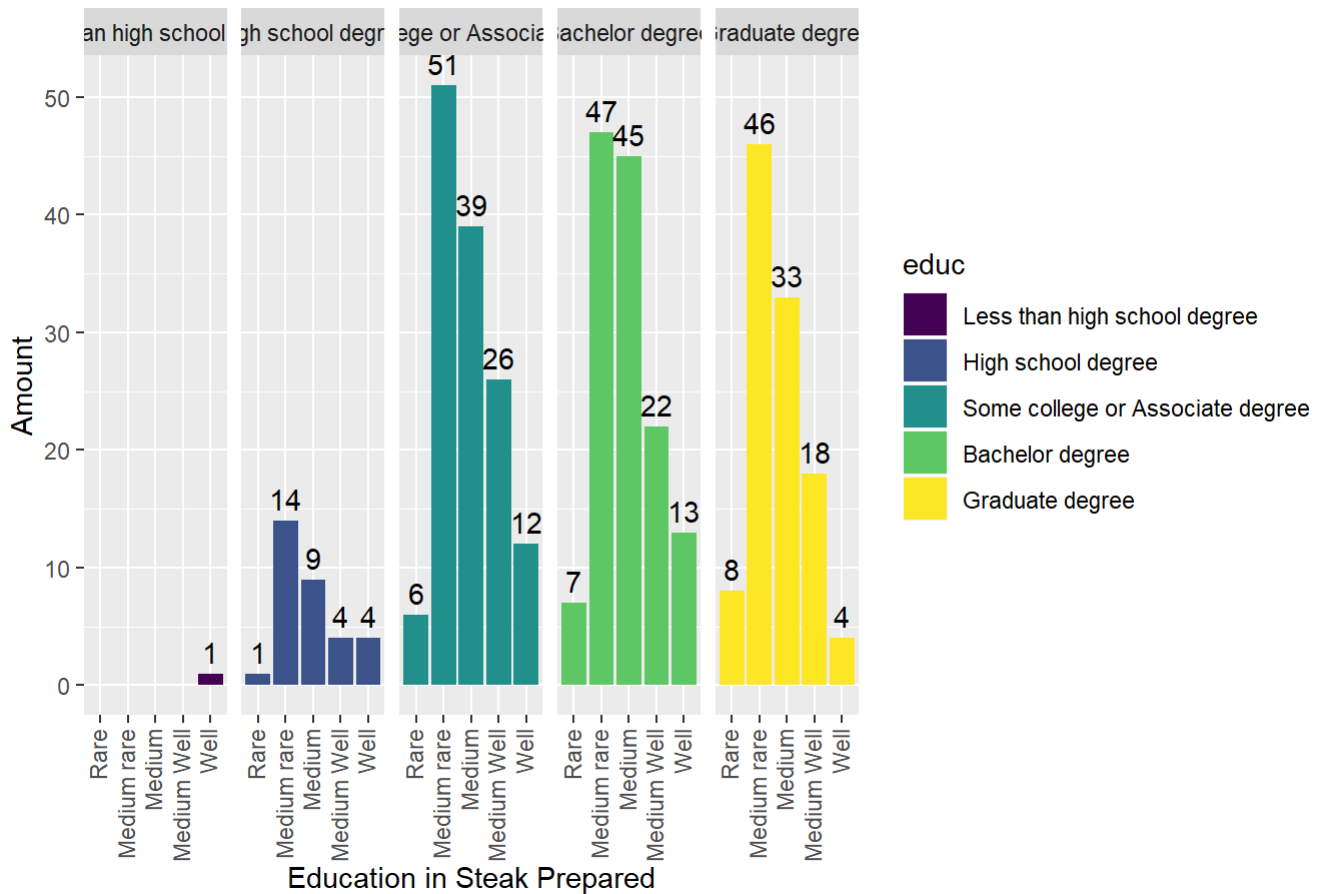
Household Income Range's Steak Prepared Bar Chart



```
#Plot bar chart of education's steak
b6 <- ggplot(educate_df, aes(x = steak_prep, fill = educ)) +
  # add geometry of the plot (bar chart)
  geom_bar(aes(y = amount), stat = "identity") +
  # add labels
  labs(x="Education in Steak Prepared", y="Amount", title="Education's Steak Prepared Bar
Chart") +
  facet_grid(~educ) +
  geom_text(aes(label = amount, y = amount), stat= "identity", vjust = -.5) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = .35))
```

b6

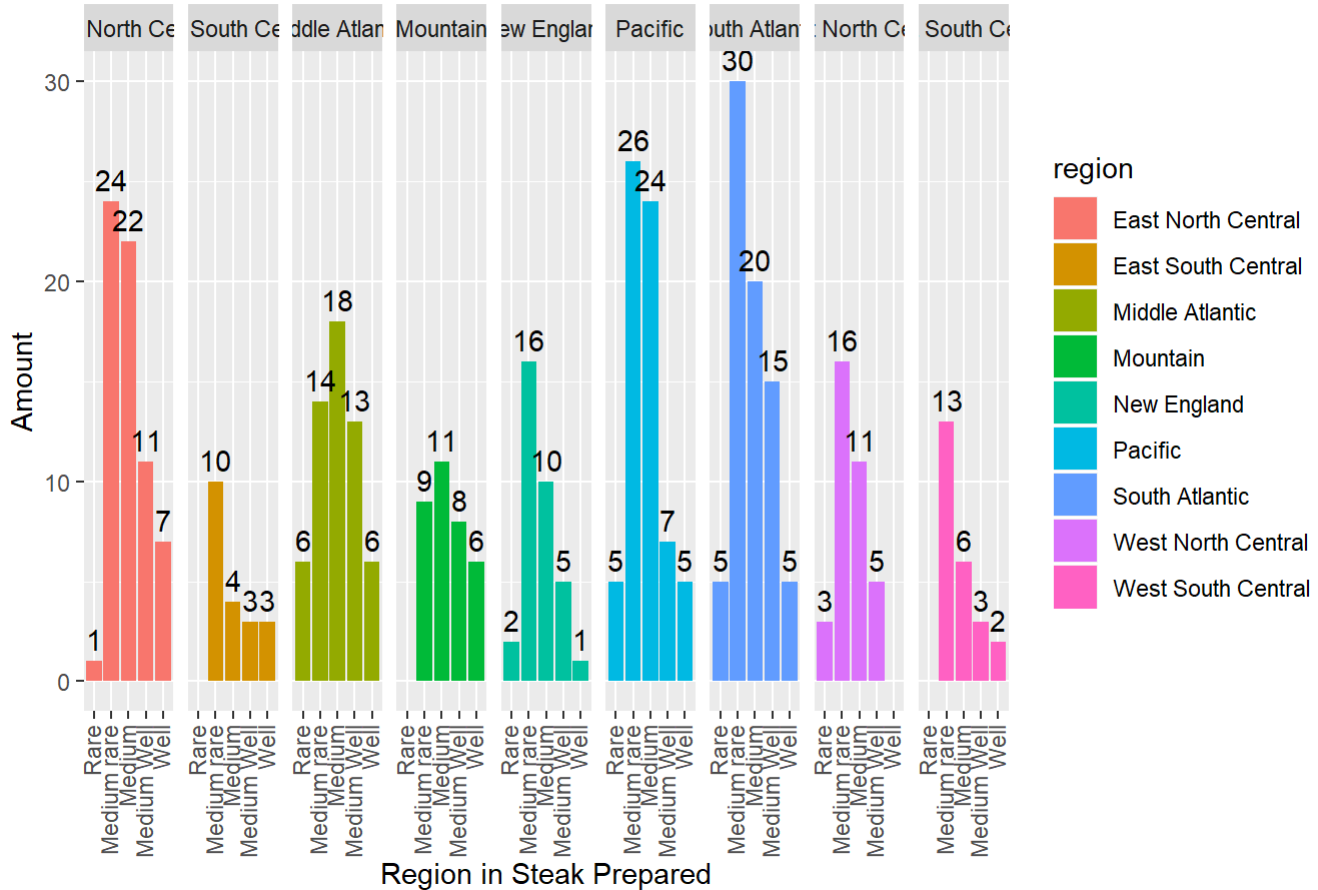
Education's Steak Prepared Bar Chart



```
#Plot bar chart of region's steak
b7 <- ggplot(region_df, aes(x = steak_prep, fill = region)) +
  # add geometry of the plot (bar chart)
  geom_bar(aes(y = amount), stat = "identity") +
  # add labels
  labs(x="Region in Steak Prepared", y="Amount", title="Region's Steak Prepared Bar Chart")
) +
  facet_grid(~region) +
  geom_text(aes(label = amount, y = amount), stat= "identity", vjust = -.5) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = .35))

b7
```

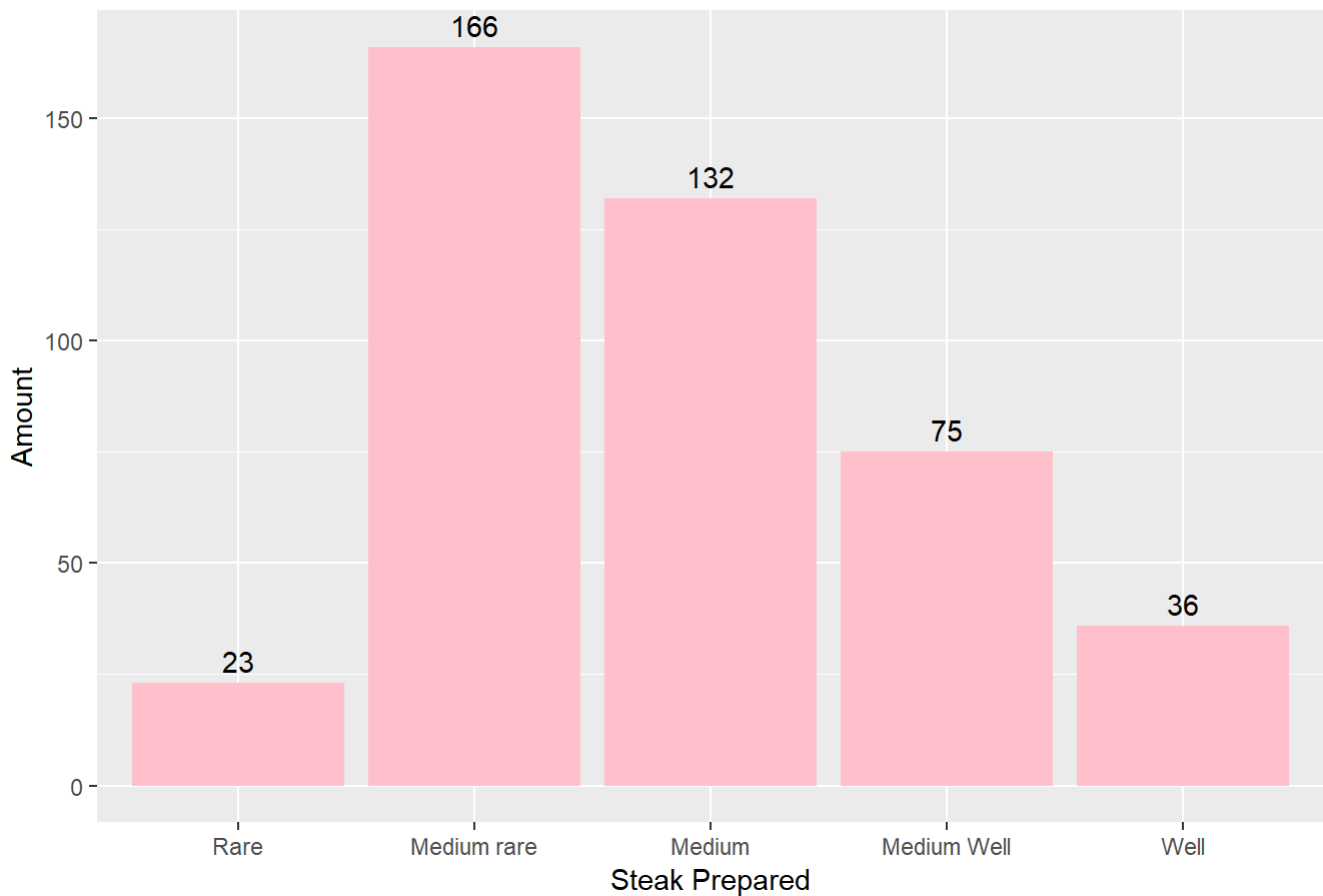
Region's Steak Prepared Bar Chart



สรุปผล:

1. จากกราฟ จะเห็นได้ว่า ลำดับความนิยม ของประเภทของ steak จะเป็น Medium rare, Medium, Medium well, Well, Rare ตามลำดับจากมากไปน้อย

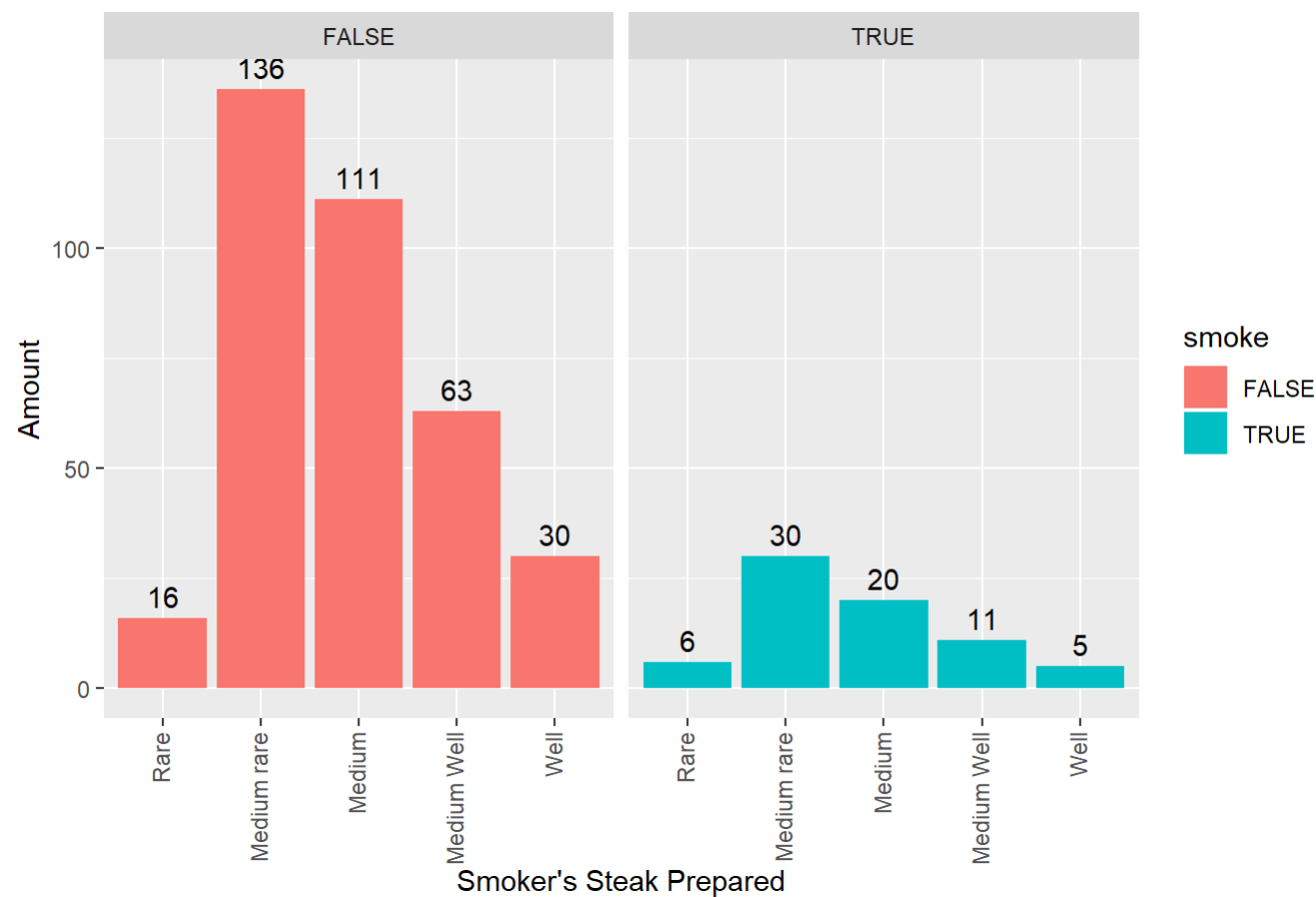
Steak Prepared Chart



2. จะเห็นได้ว่า กราฟผู้ที่สูบบุหรี่และดื่มสุรา มีรูปแบบกราฟที่เหมือนกับกราฟรวม จึงคาดการณ์ว่า การสูบบุหรี่และดื่มสุราไม่มีผล หรือมีผลน้อยต่อการเลือกทาน steak

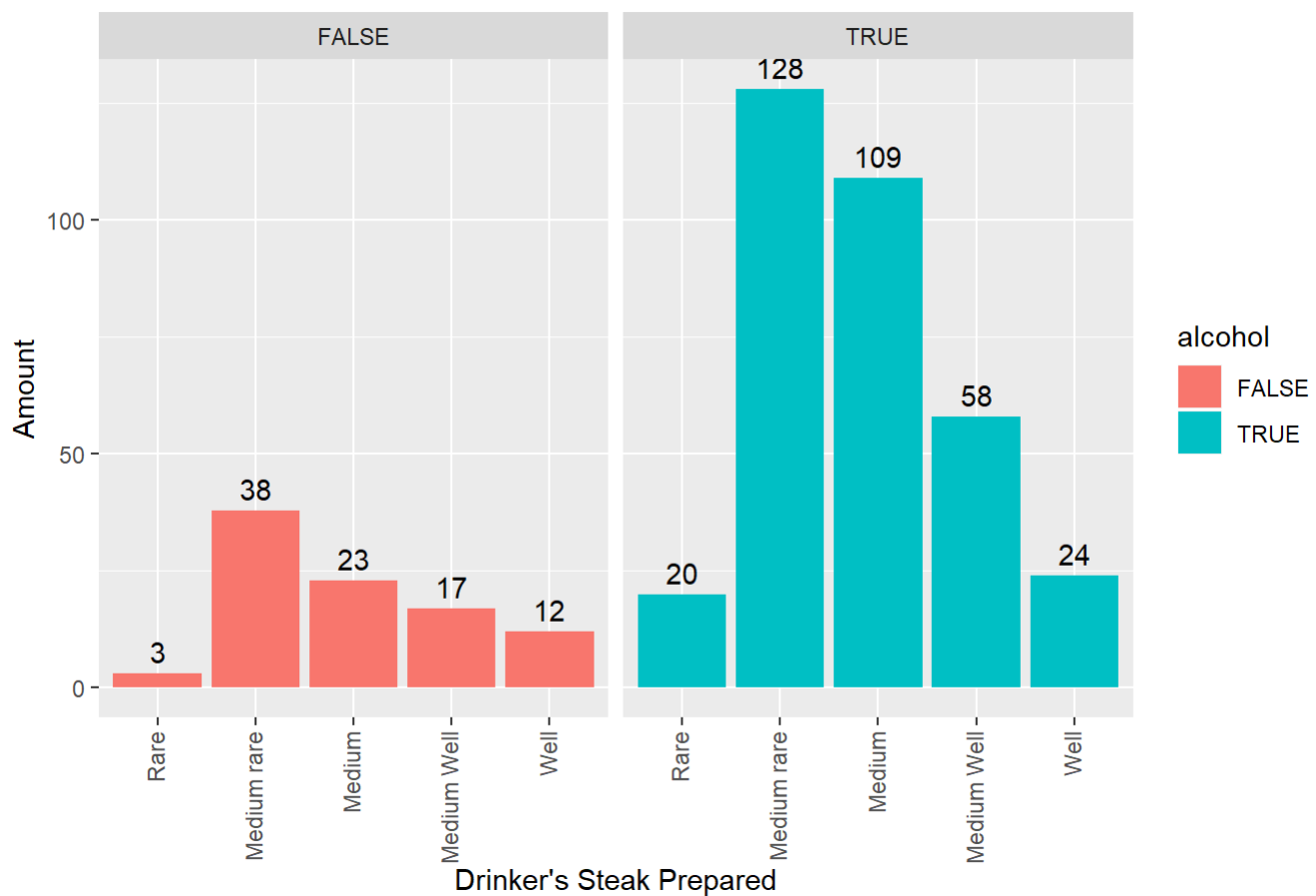
b1

Smoker's Steak Prepared Bar Chart



b2

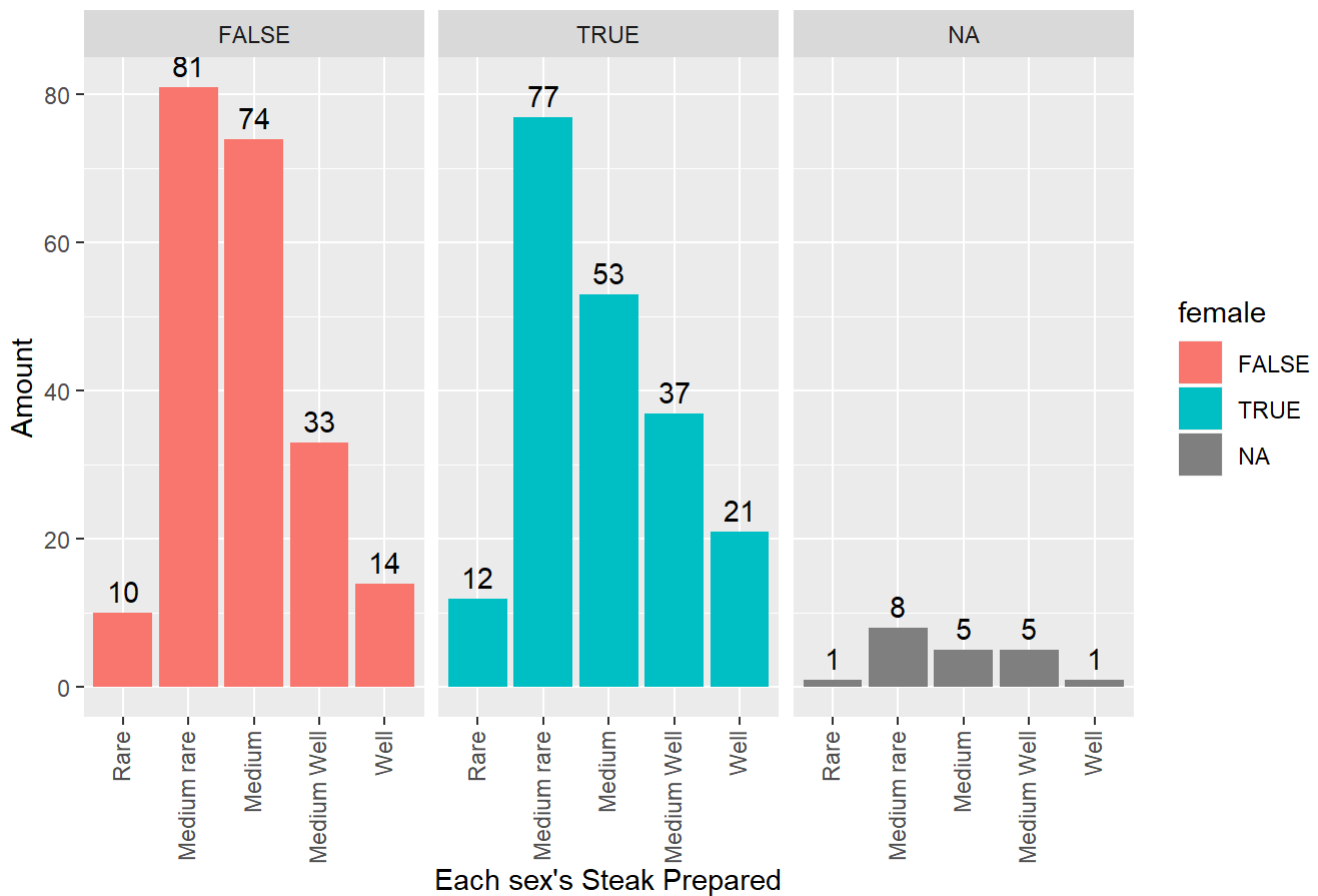
Drinker's Steak Prepared Bar Chart



3. ส่วนเพศนั้นไม่มีผลต่อความนิยม steak ในระดับต่าง ๆ

b3

Each sex's Steak Prepared Bar Chart

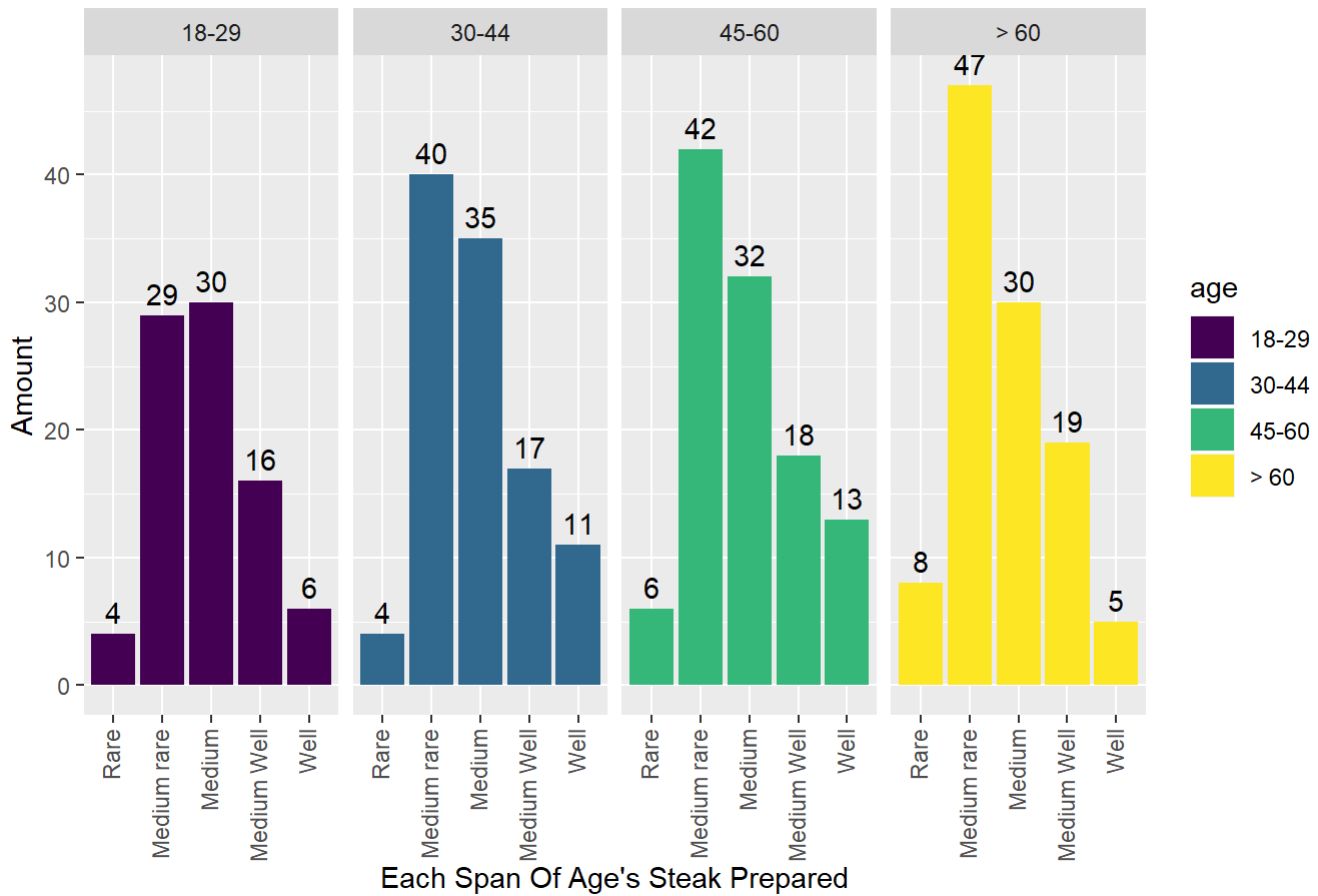


4. ในช่วงอายุ 18 - 29 ปีจะนิยม steak ระดับ Medium มากกว่า Medium rare

ส่วนในช่วงอายุ มากกว่า 60 ปี จะนิยม steak ระดับ Rare มากกว่า Well

b4

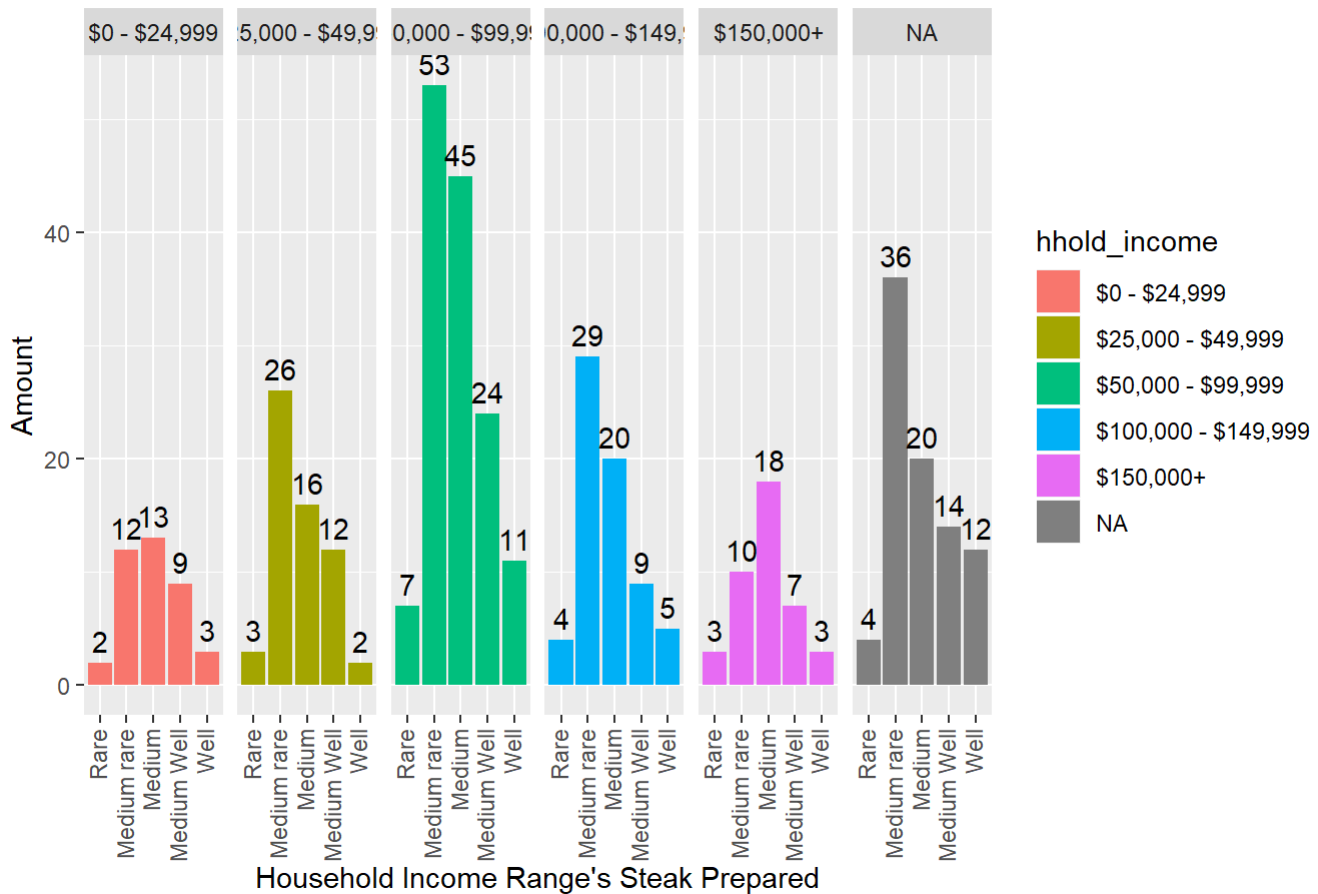
Each Span Of Age's Steak Prepared Bar Chart



5. คนที่มีรายได้ครัวเรือนมากกว่า 150,000 ดอลลาร์ มีแนวโน้มว่านิยม steak ระดับ Medium มากกว่า Medium rare

b5

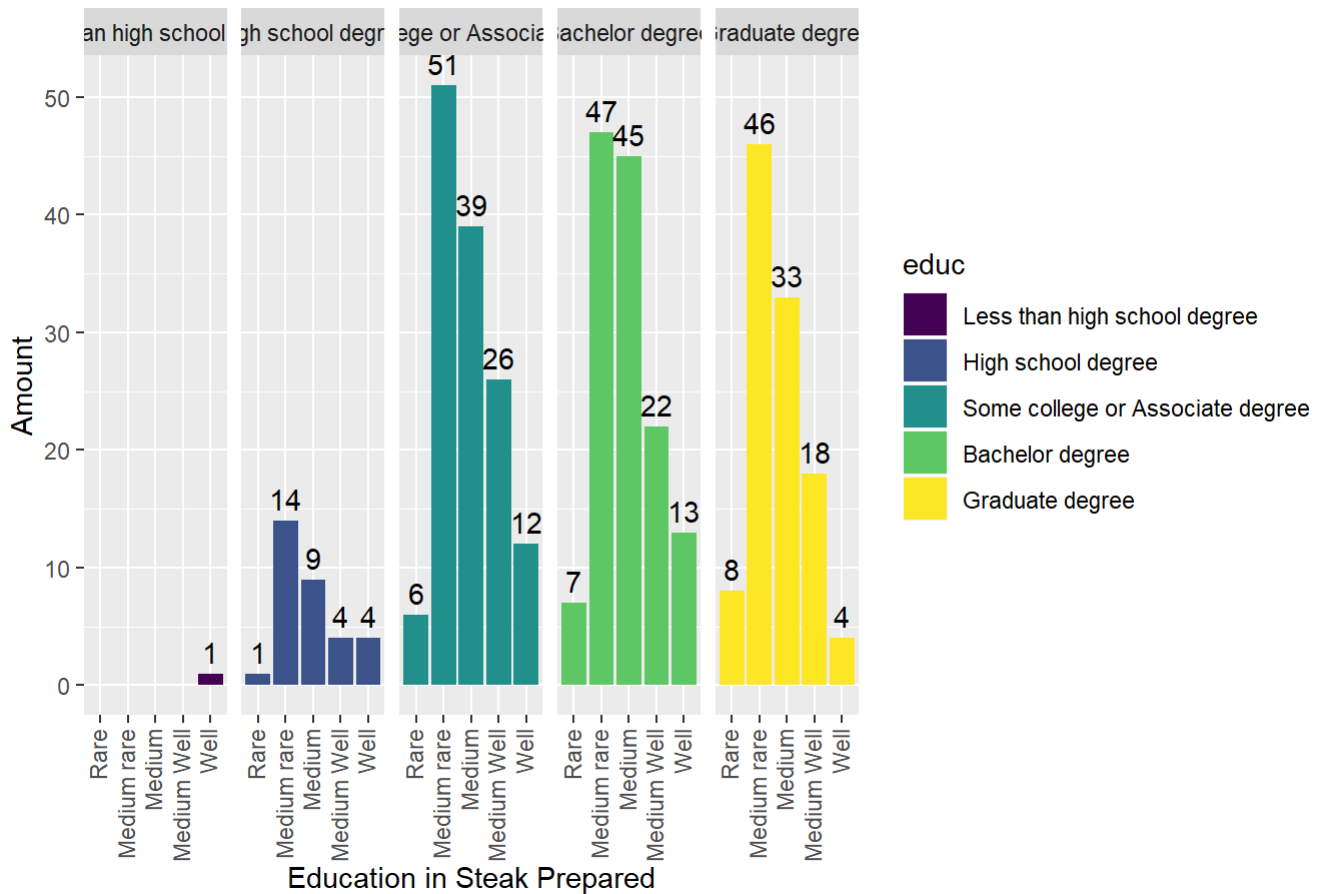
Household Income Range's Steak Prepared Bar Chart



6. คนที่จบการศึกษาระดับบัณฑิตศึกษามีแนวโน้มที่จะนิยม steak ระดับ Rare มากกว่า Well

b6

Education's Steak Prepared Bar Chart



7. คนที่อยู่ ใน Middle Atlantic จะนิยม steak ระดับ Medium มากกว่า Medium rare เล็กน้อย

คนที่อยู่ใน East South Central, Mountain และ West South Central จะไม่นิยมทาน steak ระดับ Rare

Region's Steak Prepared Bar Chart

