R 練習

YuTing

Online Food Dataset From Kaggle

複習資料視覺化

```
anyNA(onlinefood)
```

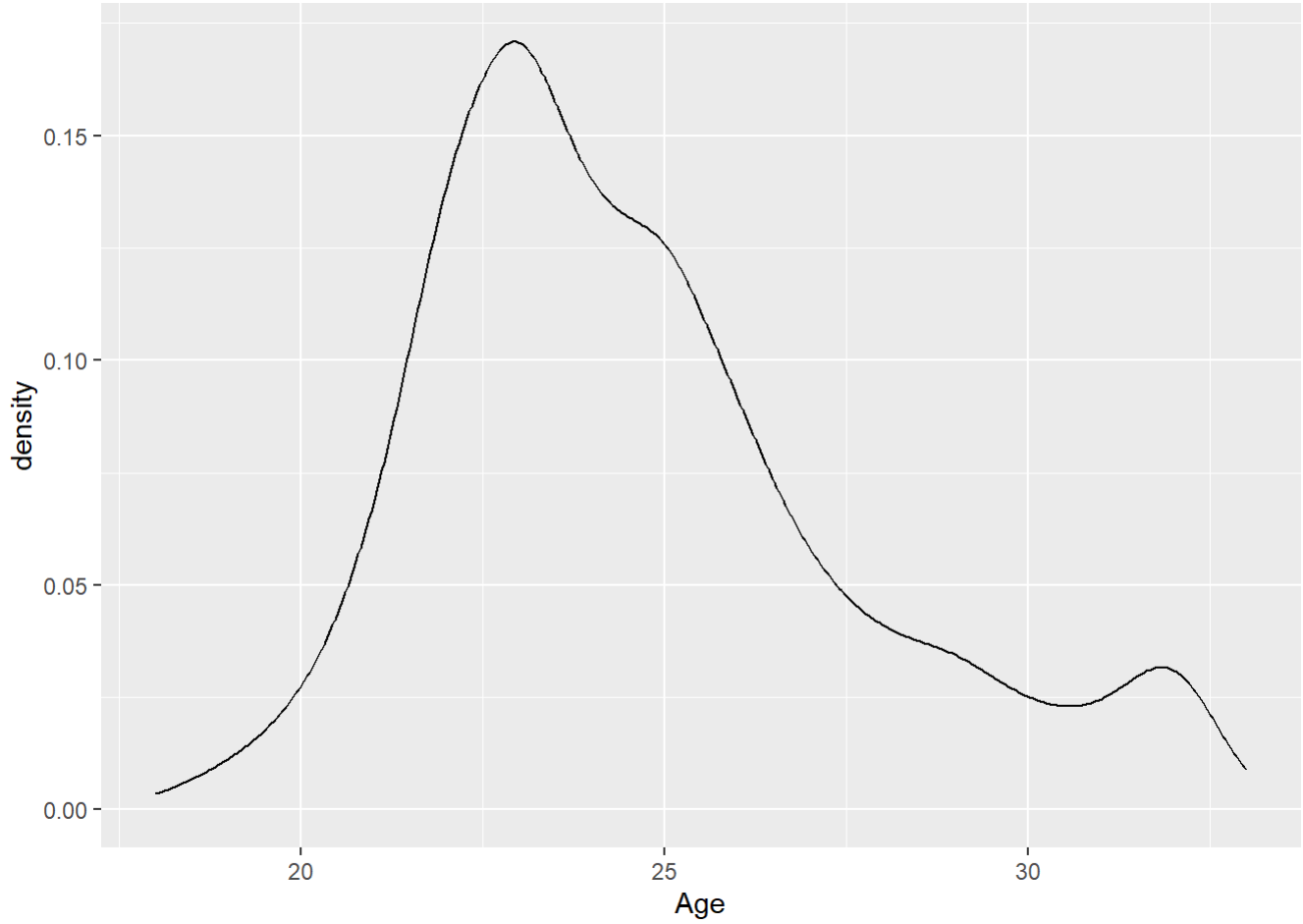
```
## [1] FALSE
```

```
head(onlinefood)
```

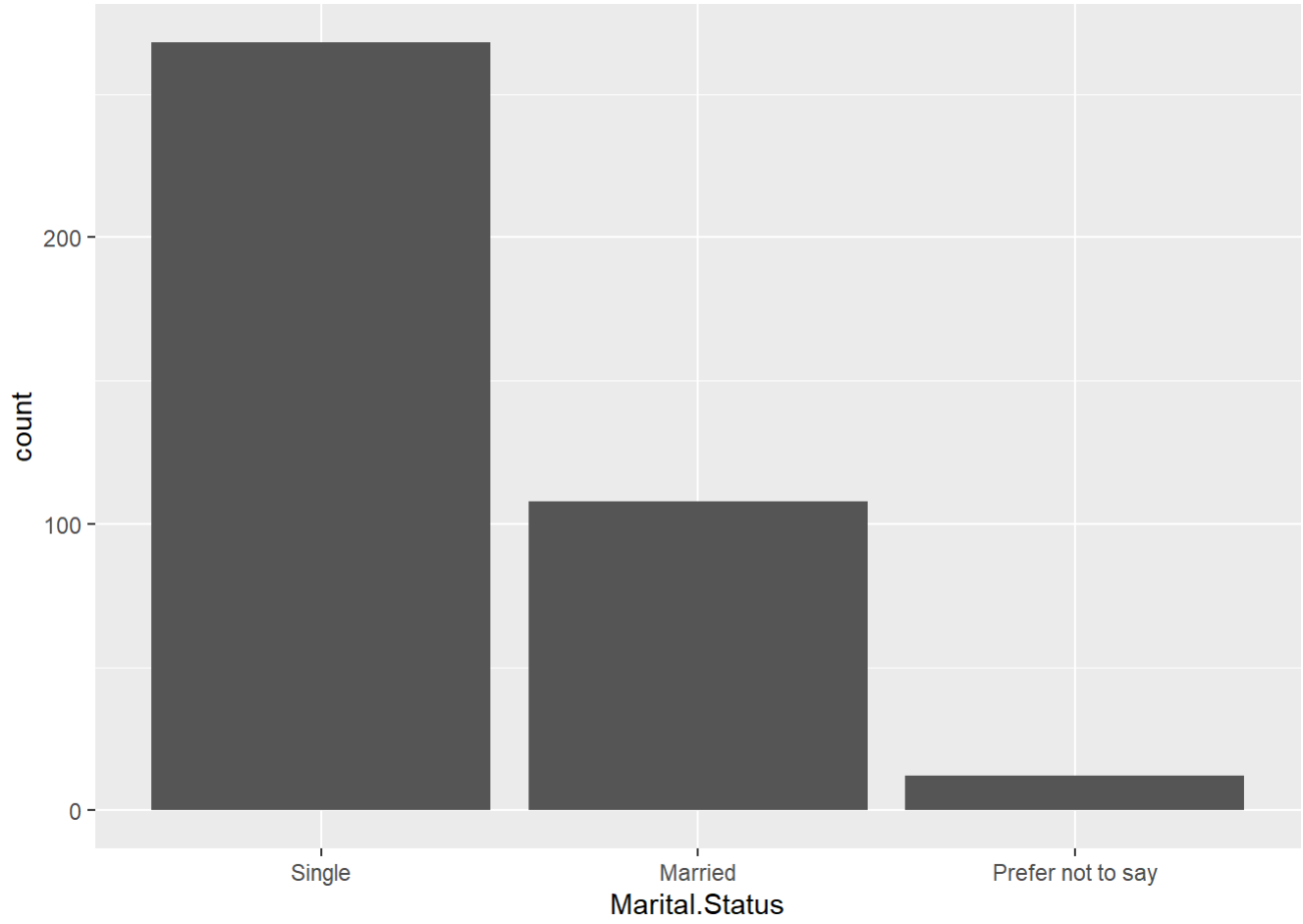
```
##   Age Gender Marital.Status Occupation  Monthly.Income
## 1  20 Female         Single    Student       No Income
## 2  24 Female         Single    Student  Below Rs.10000
## 3  22   Male         Single    Student  Below Rs.10000
## 4  22 Female         Single    Student       No Income
## 5  22   Male         Single    Student  Below Rs.10000
## 6  27 Female        Married   Employee More than 50000
##   Educational.Qualifications Family.size latitude longitude Pin.code Output
## 1              Post Graduate           4  12.9766   77.5993   560001    Yes
## 2                   Graduate           3  12.9770   77.5773   560009    Yes
## 3              Post Graduate           3  12.9551   77.6593   560017    Yes
## 4                   Graduate           6  12.9473   77.5616   560019    Yes
## 5              Post Graduate           4  12.9850   77.5533   560010    Yes
## 6              Post Graduate           2  12.9299   77.6848   560103    Yes
##    Feedback   X
## 1  Positive Yes
## 2  Positive Yes
## 3 Negative  Yes
## 4  Positive Yes
## 5  Positive Yes
## 6  Positive Yes
```

檢視年齡分布

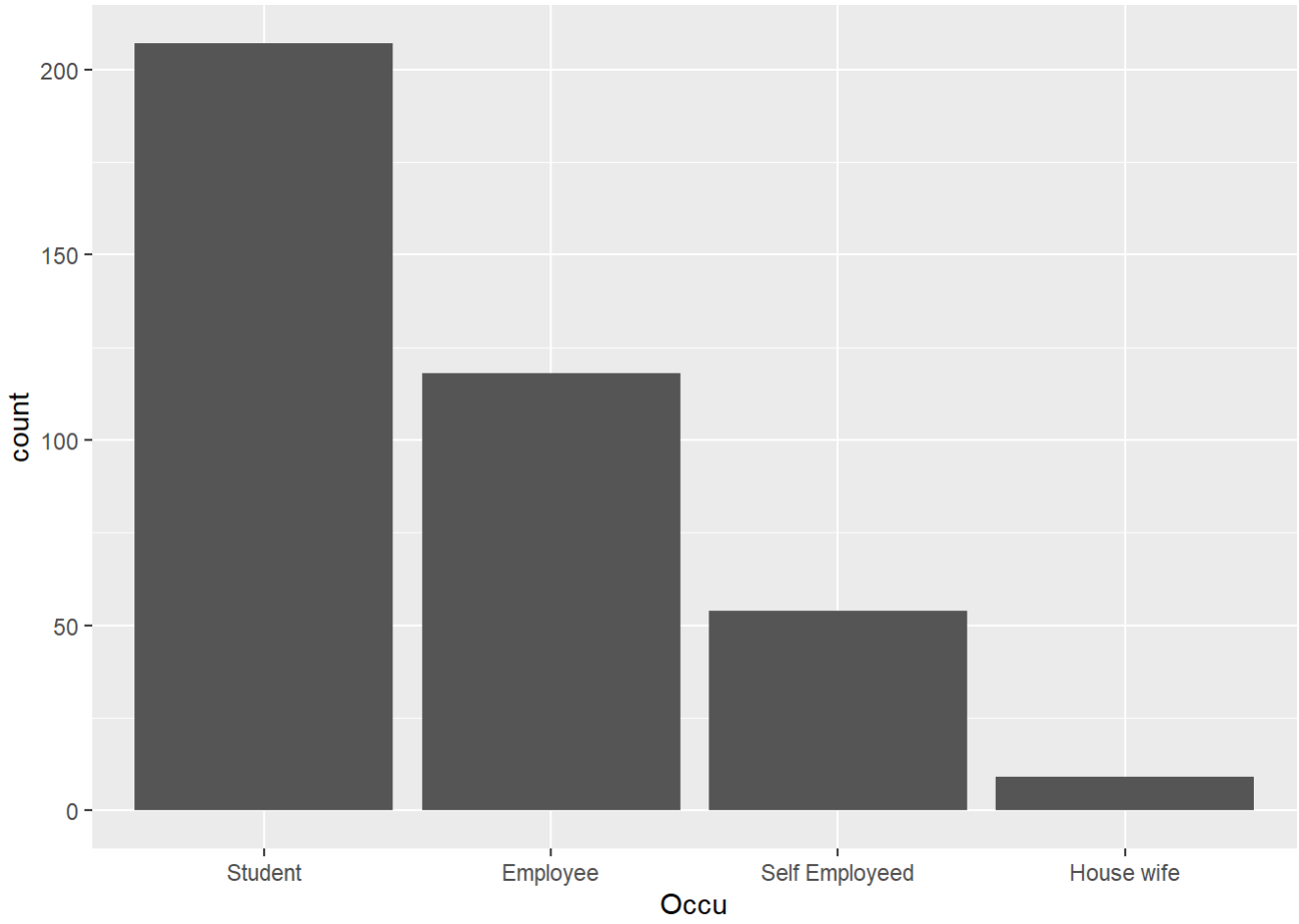
```
onlinefood<-subset(onlinefood,select = -c(X, latitude, longitude,Pin.code,Output))
onlinefood%>%
  ggplot(aes(x=Age))+geom_density()
```



Including Plots



學生點外送比例高，依個數排序職業類別



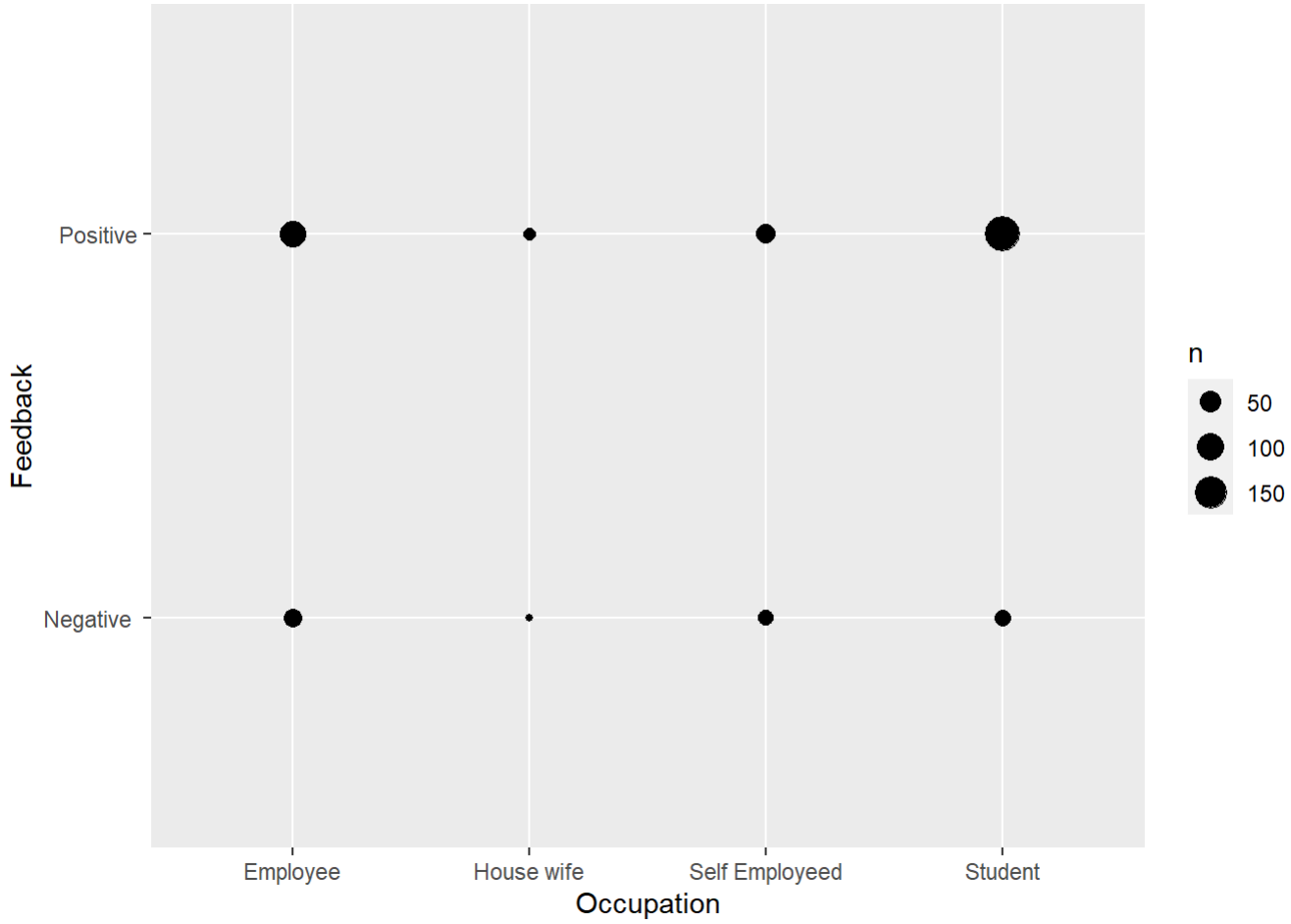
處理收入變數

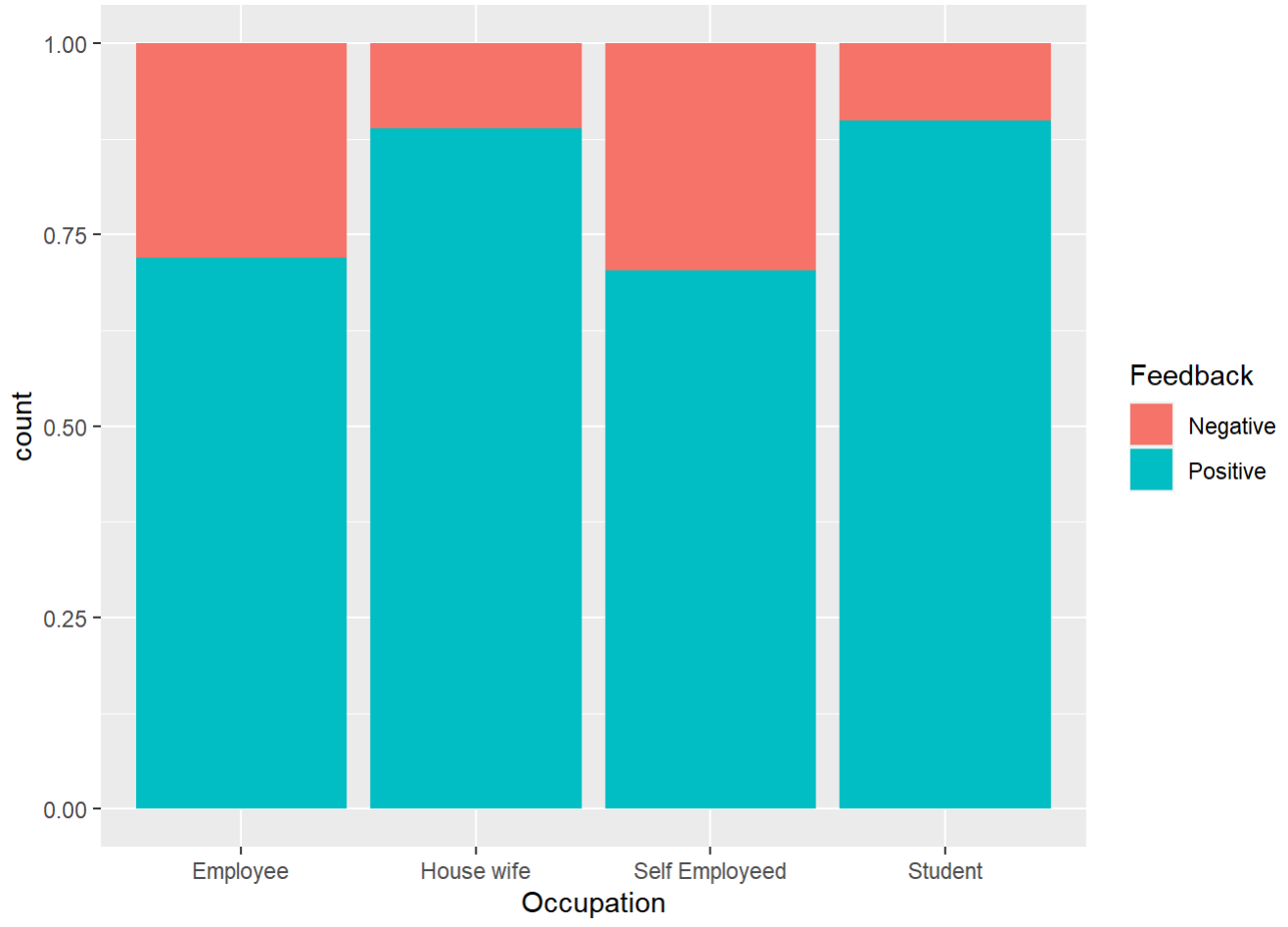
```
table(onlinefood$Monthly.Income)
```

```
##
##  10001 to 25000  25001 to 50000  Below Rs.10000 More than 50000       No Income
##              45              69              25              62             187
```

```
onlinefood<-onlinefood%>%
  mutate(income=case_when(Monthly.Income=="Below Rs.10000"~5000,
                          Monthly.Income=="10001 to 25000"~17500,
                          Monthly.Income=="25001 to 50000"~37500,
                          Monthly.Income=="More than 50000"~50000,
                          TRUE~0))%>%
  mutate(feedback10=ifelse(Feedback=="Positive",10,-10))
```

接著檢視負評與哪項因素較有關

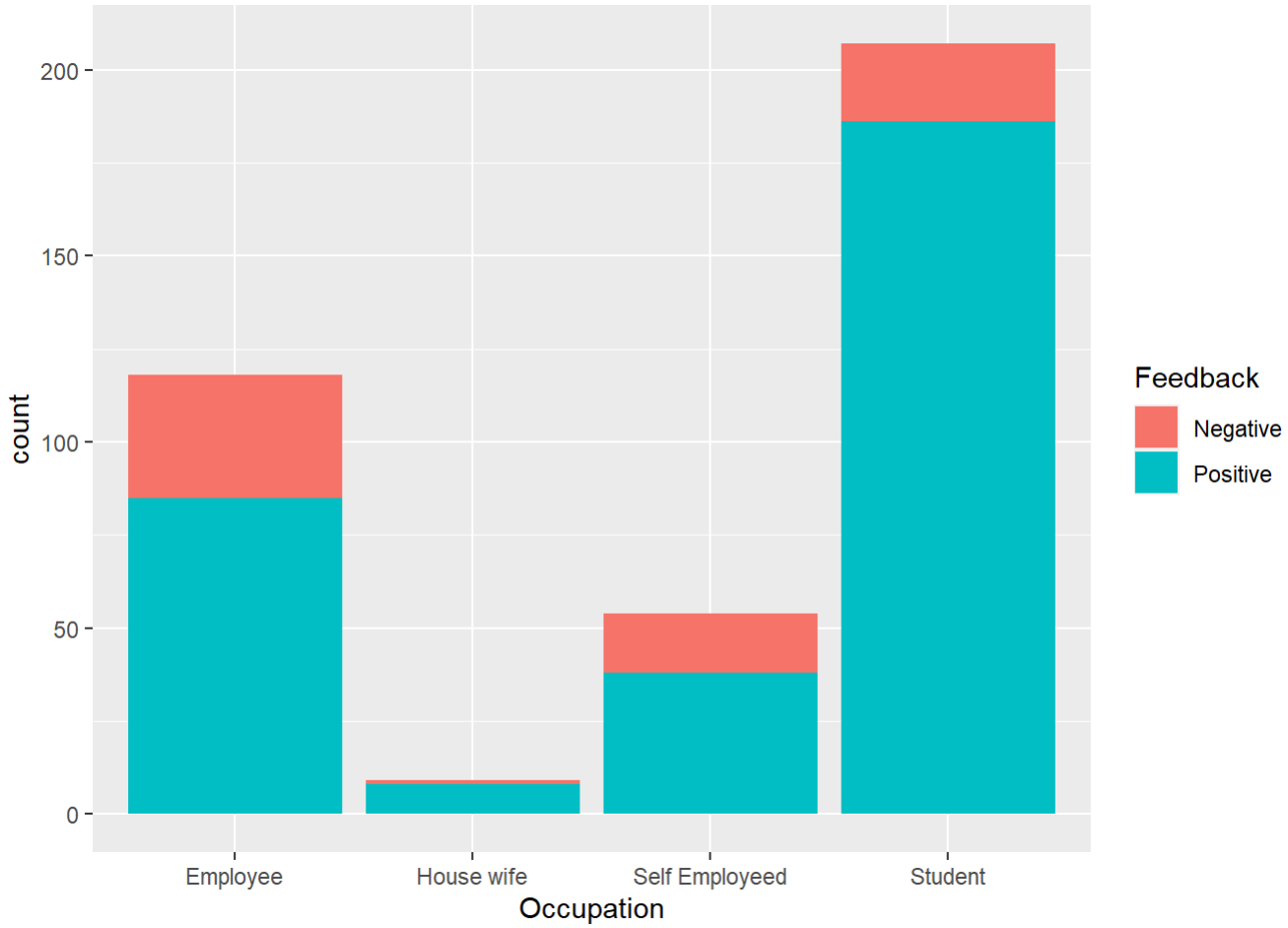




上圖之數值(滿意度比例by就業情況)

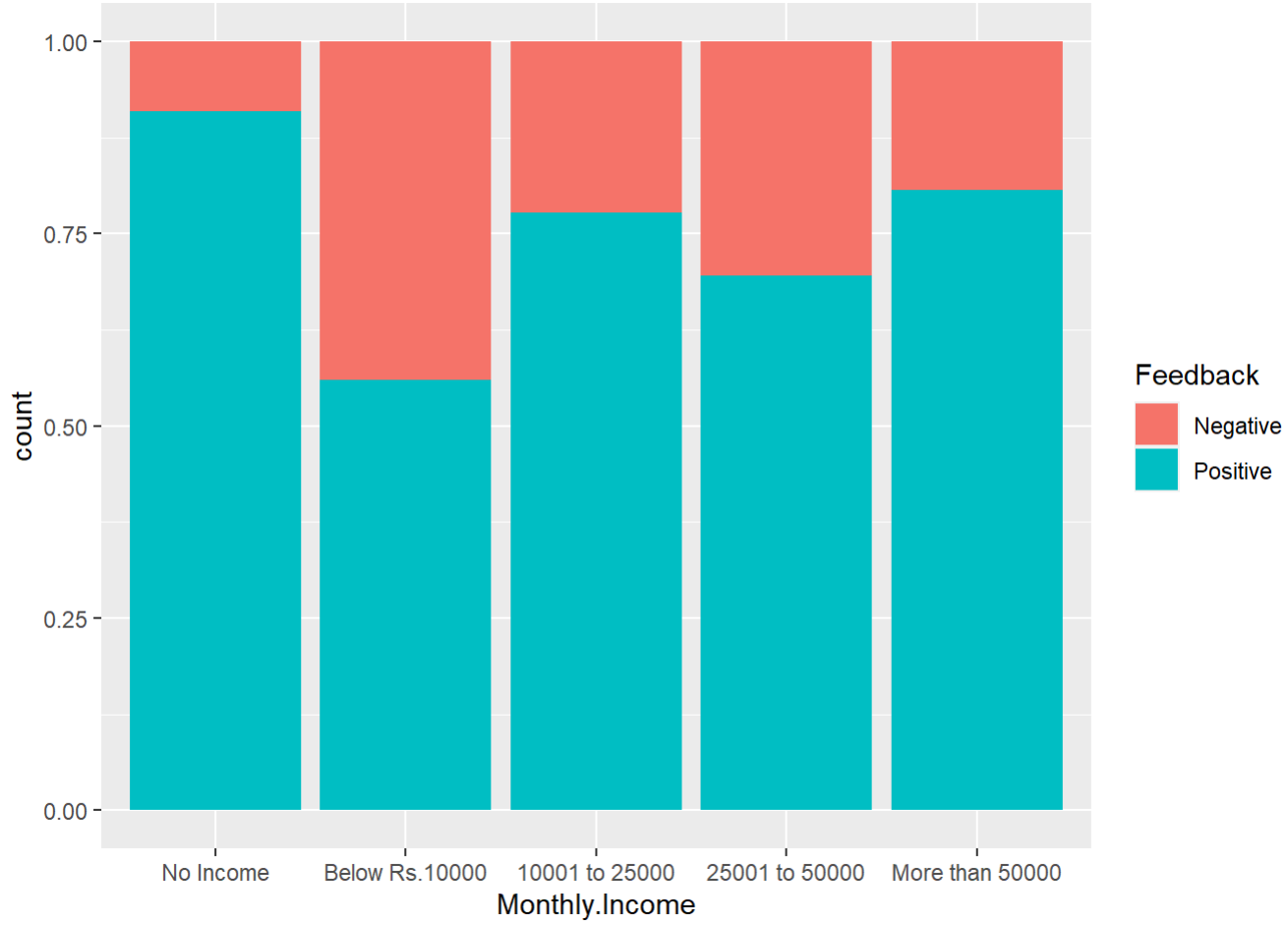
```
## # A tibble: 4 × 4
##   Occupation     Positive_n     n   per
##   <chr>               <int> <int> <dbl>
## 1 Employee               85   118 0.720
## 2 House wife              8     9 0.889
## 3 Self Employeed         38    54 0.704
## 4 Student               186   207 0.899
```

把(position = “fill”)拿掉，從累積百分比，變成個數，以檢視實際個數差距

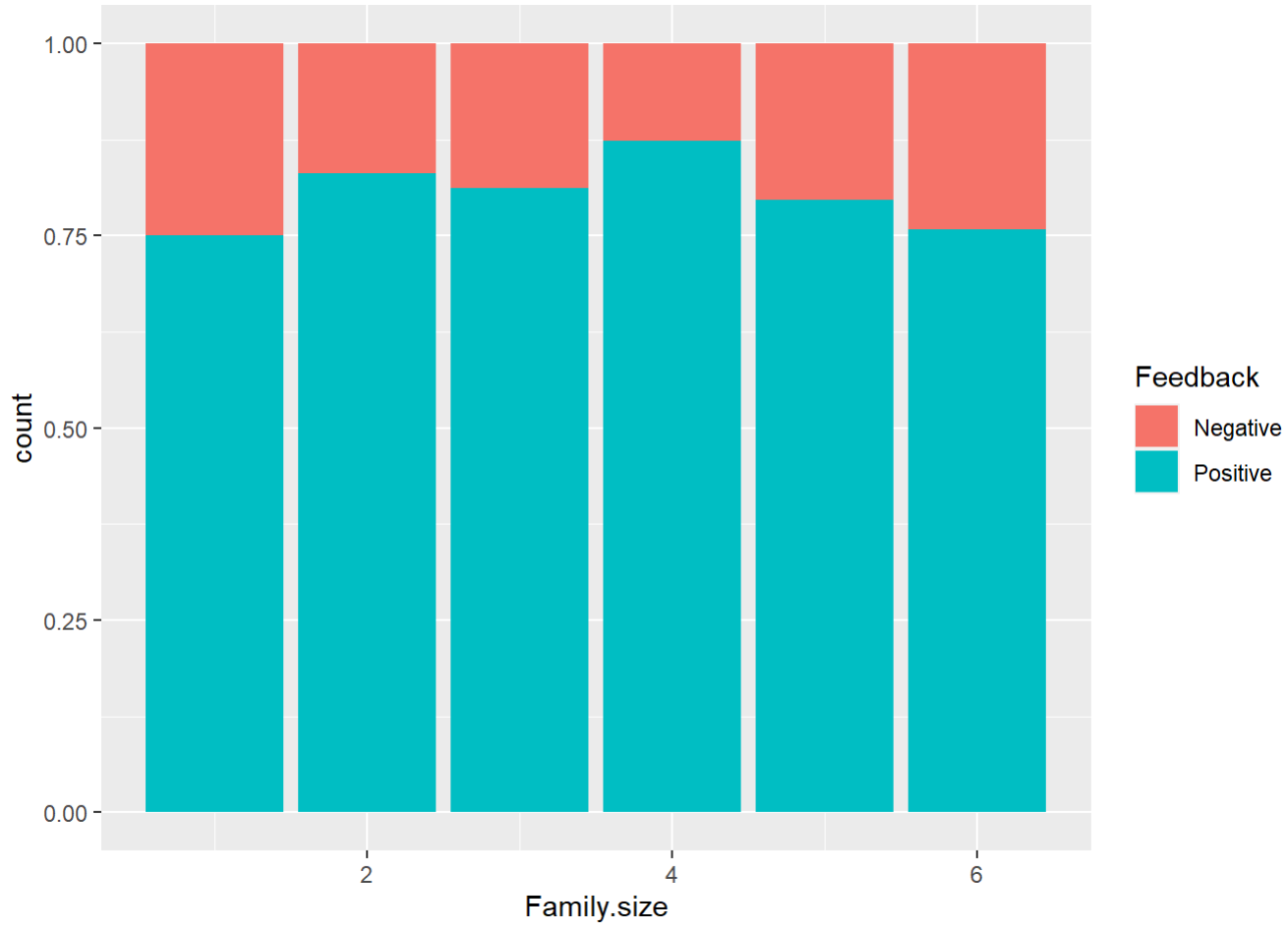


接著嘗試月收入

```
onlinefood$Monthly.Income<-factor(onlinefood$Monthly.Income,level=c("No Income","Below Rs.100
00","10001 to 25000","25001 to 50000","More than 50000"))
ggplot(data = onlinefood, aes(x = Monthly.Income, fill = Feedback)) +
  geom_bar(position = "fill")
```

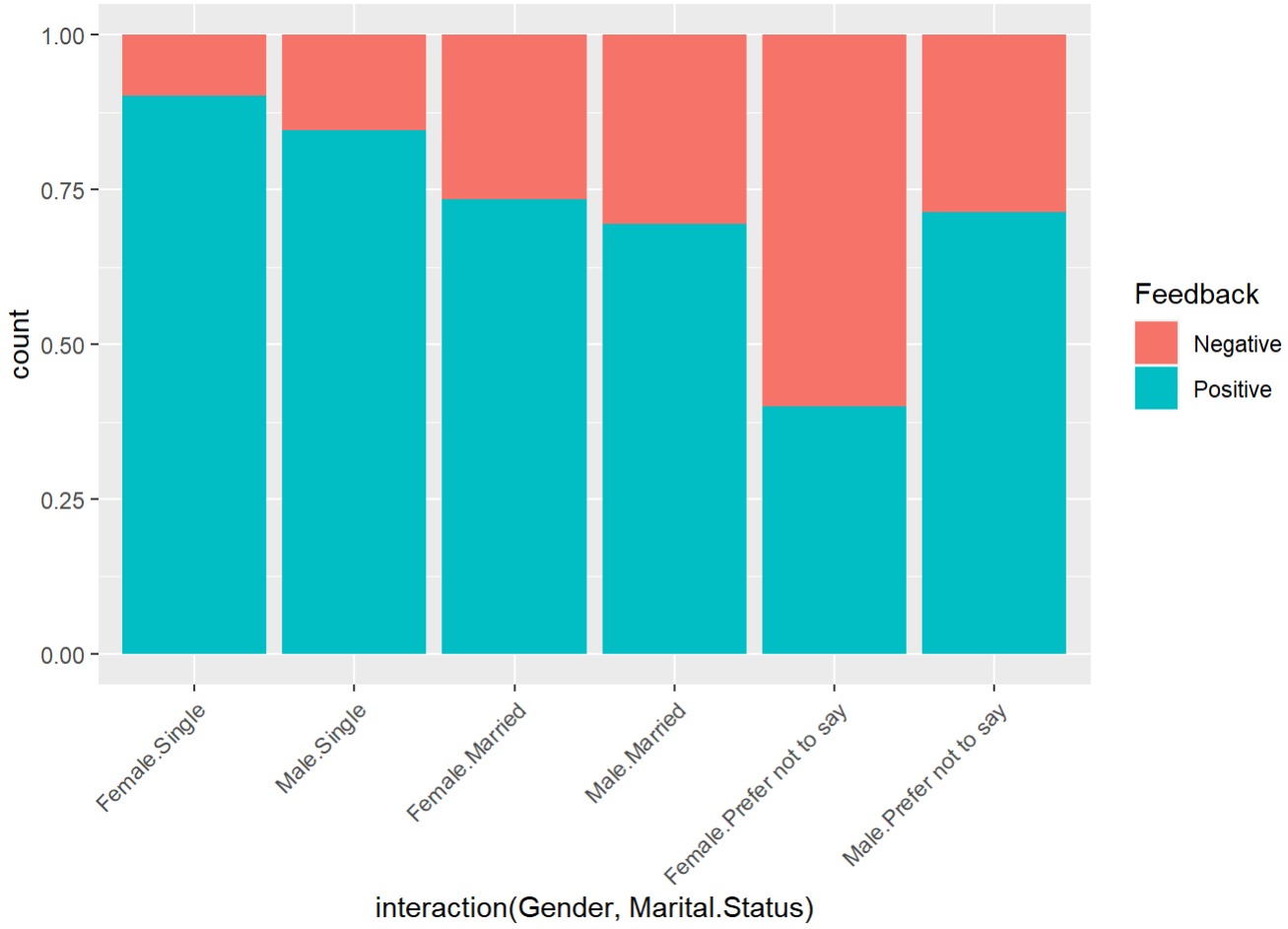


接著嘗試家庭人數



接著嘗試 性別*感情狀況

```
ggplot(data = onlinefood, aes(x = interaction(Gender,Marital.Status), fill = Feedback)) +
  geom_bar(position = "fill")+scale_x_discrete(guide=guide_axis(angle=45))
```



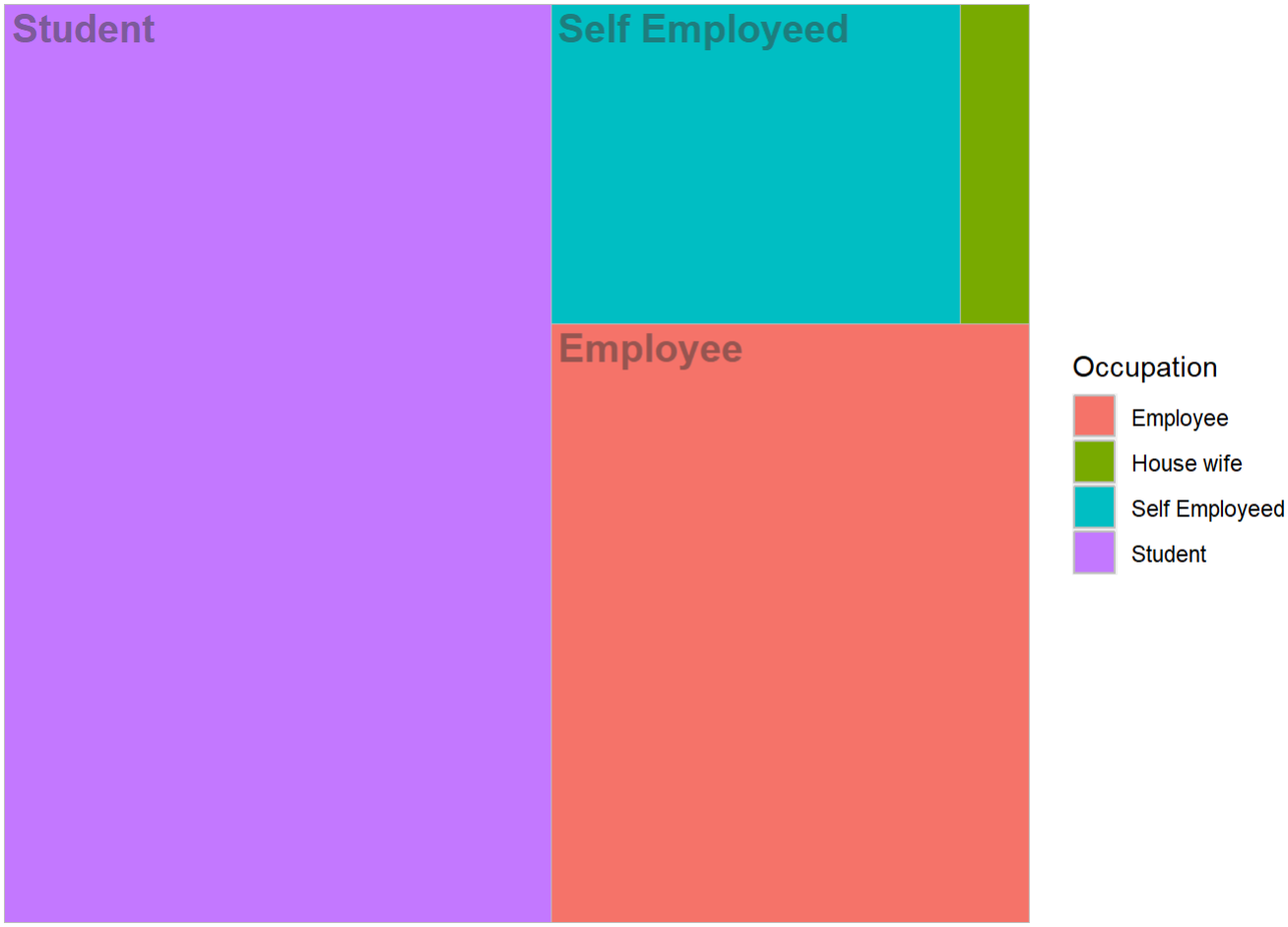
矩形式樹狀圖

安裝treemapify

ggplot(資料, aes(area=, color=, fill=, linetype=, alpha=, subgroup=, subgroup2=, subgroup3=))

函式	功能	引數
`geom_treemap()`	建立樹狀圖	引數設定承襲於 `ggplot()` 函式內各引數
`geom_treemap_subgroup_border()`	設定樹狀圖子類別間線條樣式	
`geom_treemap_text()`	設定樹狀圖文字樣式	`color` :文字顏色 `size` :文字大小 `alpha` :文字透明度 `family` :文字字型 `fontface` :文字樣式 `angle` :文字角度 `place` :文字位置 (`bottom` 、 `topleft` 、 `top` 、 `topright` ...) `grow` :將文字壓縮於矩形內
`geom_treemap_subgroup_text()`	設定樹狀圖子類別文字樣式	

```
library(treemapify)
onlinefood_occ<-onlinefood_occ%>%
  mutate(negative=n-Positive_n)
ggplot(onlinefood_occ,
       aes(area=n, fill=Occupation,label=Occupation, subgroup=Occupation))+
  geom_treemap()+
  geom_treemap_text(size=15, color="#3C3C3C", alpha=0.5, fontface="bold")
```



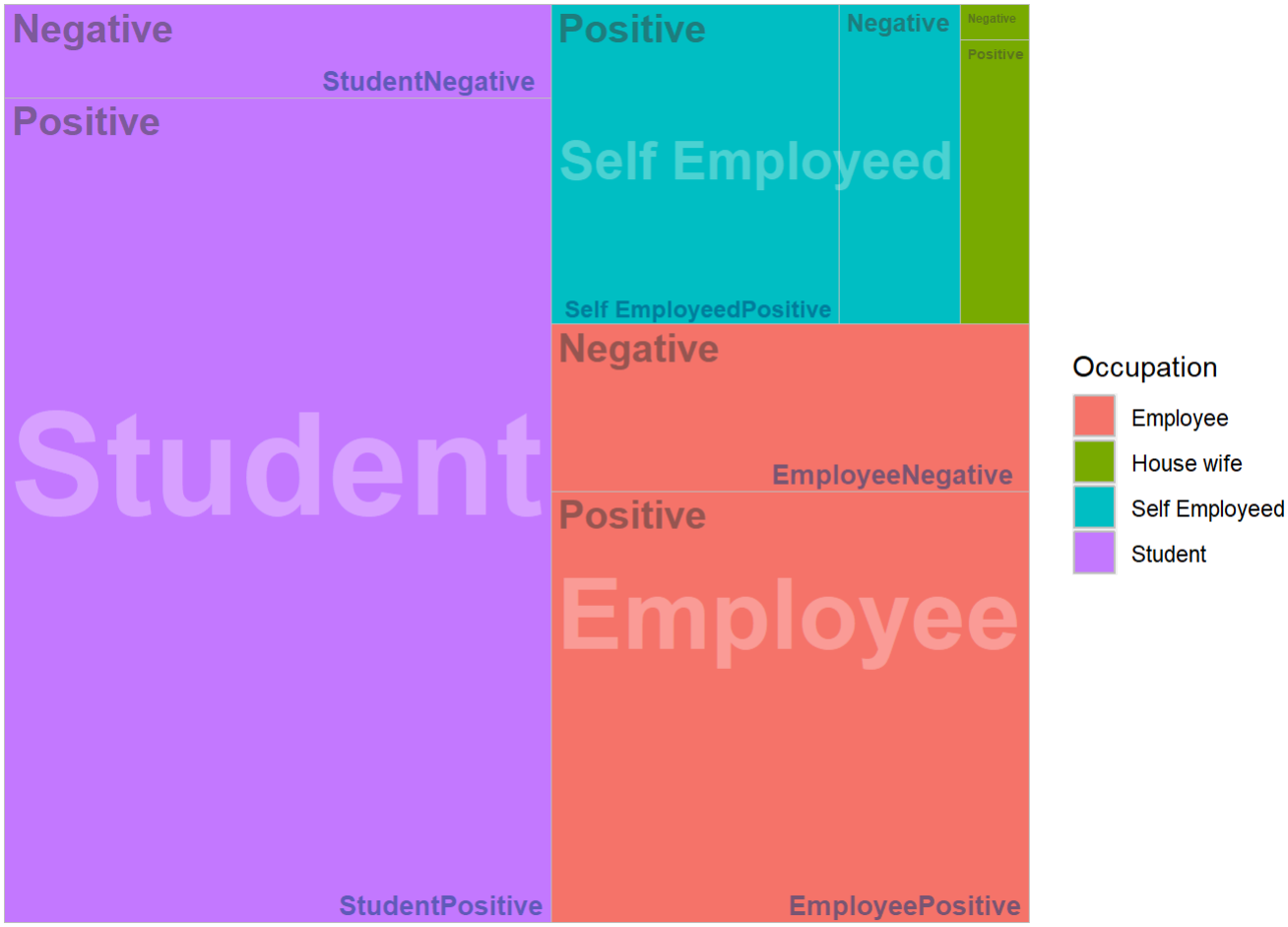

```
onlinefood_tree<-onlinefood%>%
  group_by(Occupation,Feedback)%>%
  dplyr::summarise(n=n())%>%
  mutate(Occ_feedback=paste0(Occupation,Feedback))
```

```
## `summarise()` has grouped output by 'Occupation'. You can override using the
## `.groups` argument.
```

```
onlinefood_tree
```

```
## # A tibble: 8 × 4
## # Groups:   Occupation [4]
##   Occupation     Feedback        n Occ_feedback
##   <chr>          <chr>       <int> <chr>
## 1 Employee       "Negative "    33 "EmployeeNegative "
## 2 Employee       "Positive"     85 "EmployeePositive"
## 3 House wife     "Negative "     1 "House wifeNegative "
## 4 House wife     "Positive"      8 "House wifePositive"
## 5 Self Employeed "Negative "    16 "Self EmployeedNegative "
## 6 Self Employeed "Positive"     38 "Self EmployeedPositive"
## 7 Student        "Negative "    21 "StudentNegative "
## 8 Student        "Positive"    186 "StudentPositive"
```

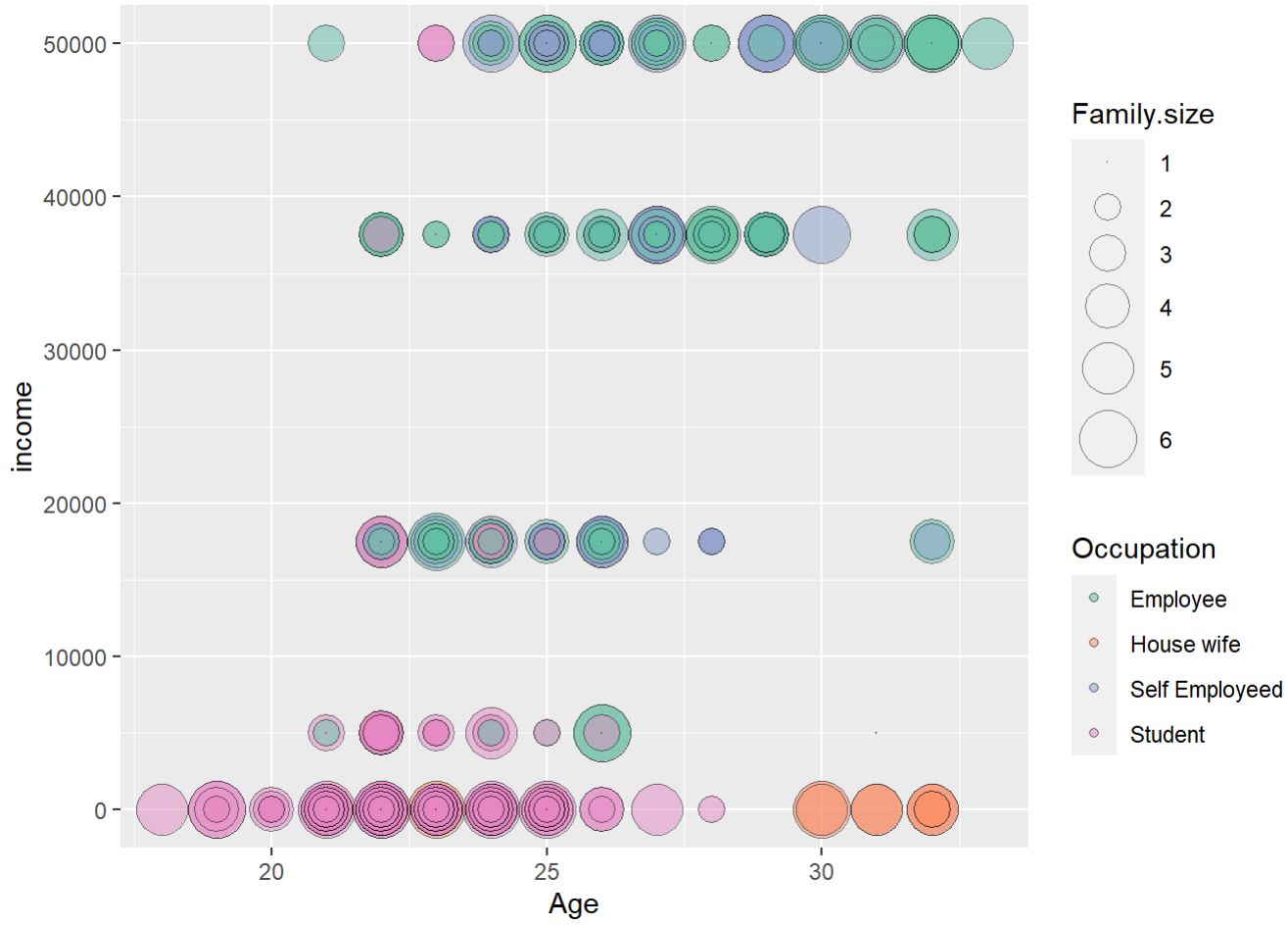
```
ggplot(onlinefood_tree,
       aes(area=n, fill=Occupation,label=Feedback, subgroup=Occupation,subgroup2=Occ_feedbac
k))+
  geom_treemap()+
  geom_treemap_text(size=15, color="#3C3C3C", alpha=0.5, fontface="bold")+
    geom_treemap_subgroup_text(grow=T, color="white", alpha=0.3, fontface="bold", place="cent
er")+
  geom_treemap_subgroup2_text(size=10, color="#003D79", alpha=0.5, fontface="bold", place="bo
ttomright")
```



泡泡圖

```
onlinefood=arrange(onlinefood, desc(Family.size))

ggplot(data=onlinefood, aes(x=Age, y=income))+
  geom_point(aes(size=Family.size, fill=Occupation), alpha=0.5, shape=21)+
  scale_size_continuous(range=c(.1, 10))+
  scale_fill_brewer(palette="Set2")
```



失敗 income不是真正的連續變數，容易重疊