ggplot2

Zhan

2019年3月30日

install ggplot2

```
install.packages("ggplot2")
```

And use library() to load it

```
library("ggplot2")
```

read the data

```
#read the data and ggplot2
library(ggplot2)

data=read.csv("SASdata.csv",fileEncoding = "Big5")
attach(data)
#SASdata 包含中文因此使用big5 encode

#str(data)
#summary(data)
#sapply(data, function(df) sum(is.na(df)))
```

ggplot syntax

Unlike other data visualization package ggplot2 works with dataframe 但還是有不是使用dataframe來畫圖的方式

ggplot2不同於其他資料視覺化套件的地方在於 使用疊加(“+”)的方式來進行繪圖

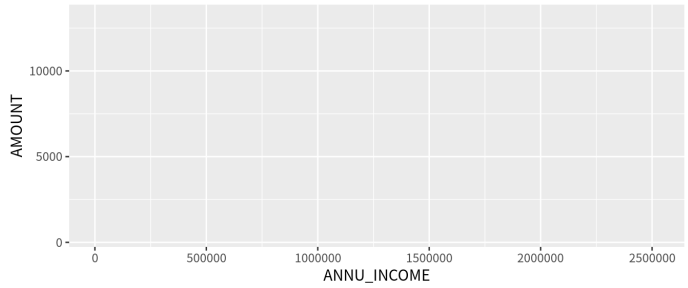
主要可以分為三大部份一層一層疊加上去來完成一張圖表

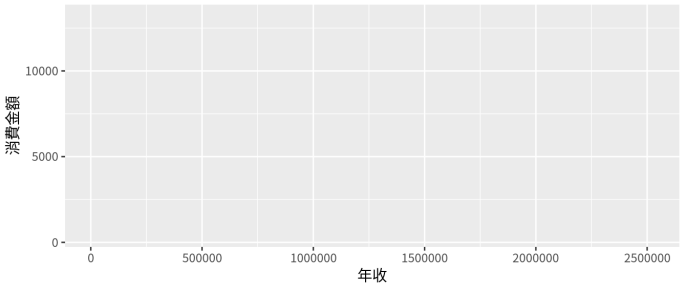
```
data + aesthetics + geometry
```

- data : data frame
- aesthetics:用來設定X,Y軸的變數.圖形的顏色大小…
- geometry:畫圖的類型,如直方圖散佈圖或盒型圖等等

首先先示範如繪製一張ggplot的圖及更改XY軸名稱

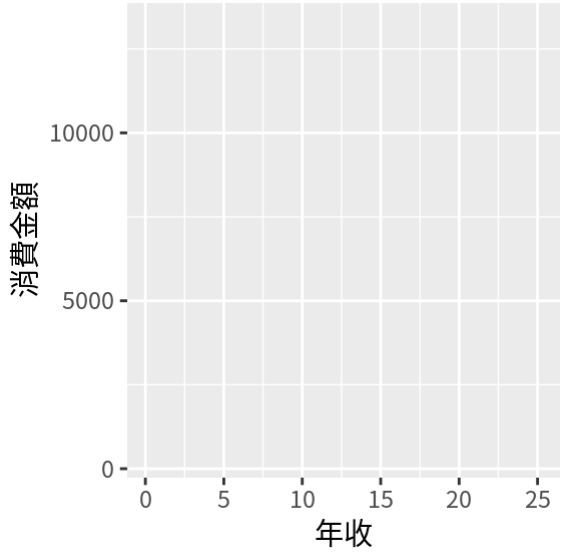
```
gg.scatter=ggplot(data=data,aes(x=ANNU_INCOME,y=AMOUNT))
gg.scatter

gg.scatter=ggplot(data=data,aes(x=ANNU_INCOME,y=AMOUNT))+labs(x="年收",y="消費金額")
gg.scatter
```





改變X軸刻度

```
gg.scatter=ggplot(data=data,aes(x=ANNU_INCOME,y=AMOUNT))+labs(x="年收",y="消費金額")+
  scale_x_continuous(labels=c(0,5,10,15,20,25))
gg.scatter
```

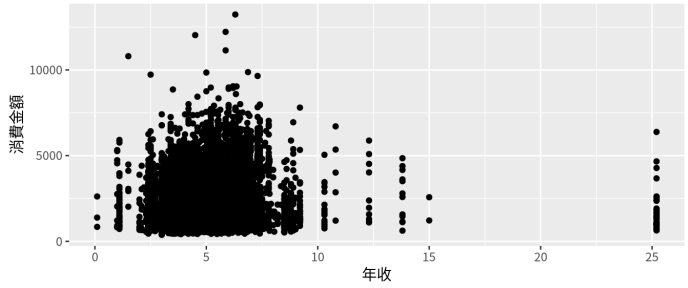


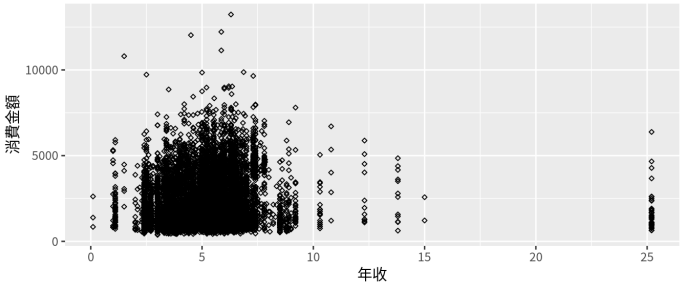
Scatter

geom_point()

Plot scatter by plus geom_point

```
gg.scatter+geom_point()

#change the point size
gg.scatter+geom_point(size=1,shape=23)
```



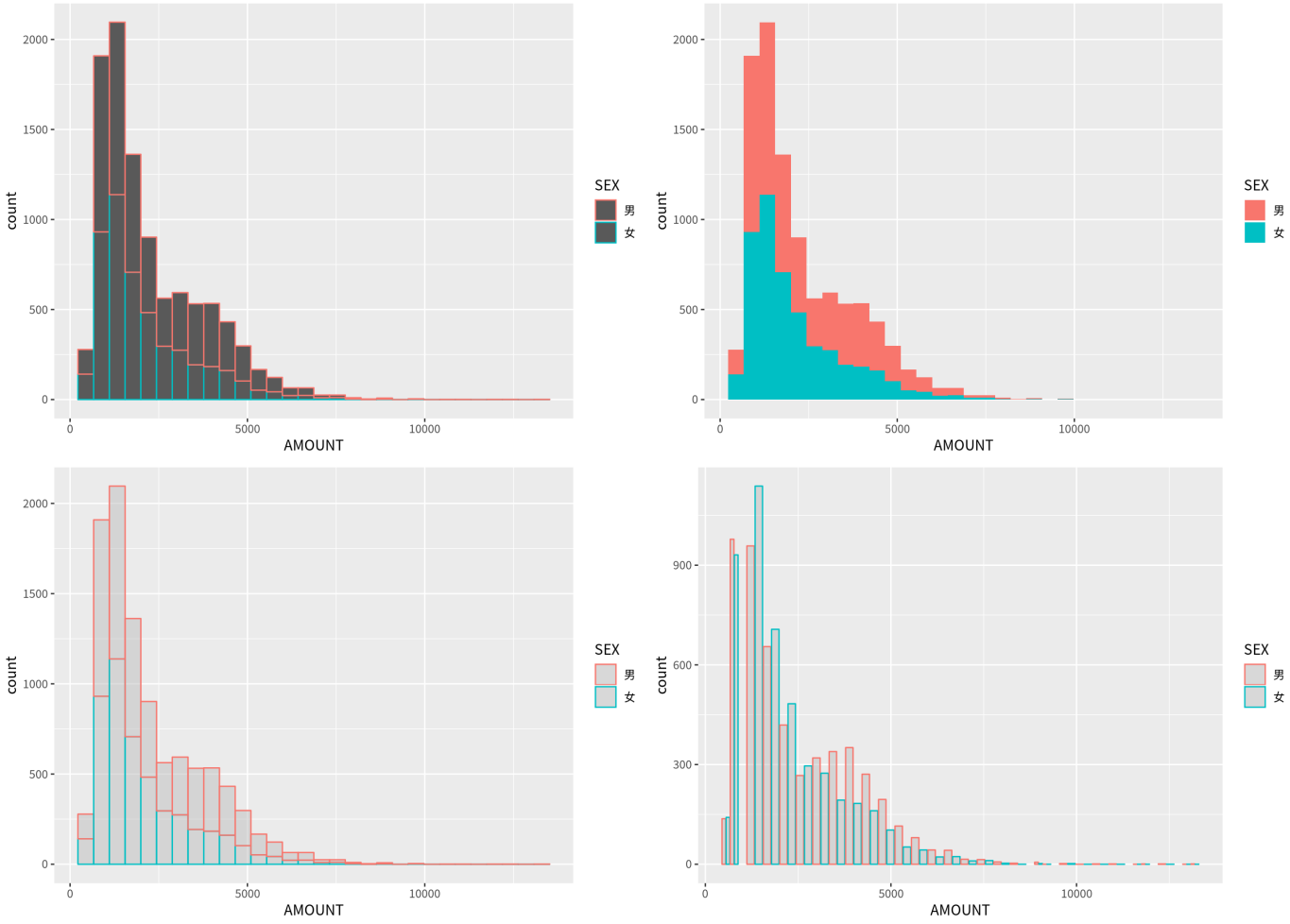


Histogram

geom_histogram()

為了了解該比資料中各變數的分佈使用畫圖來了解是一個相當好的方法

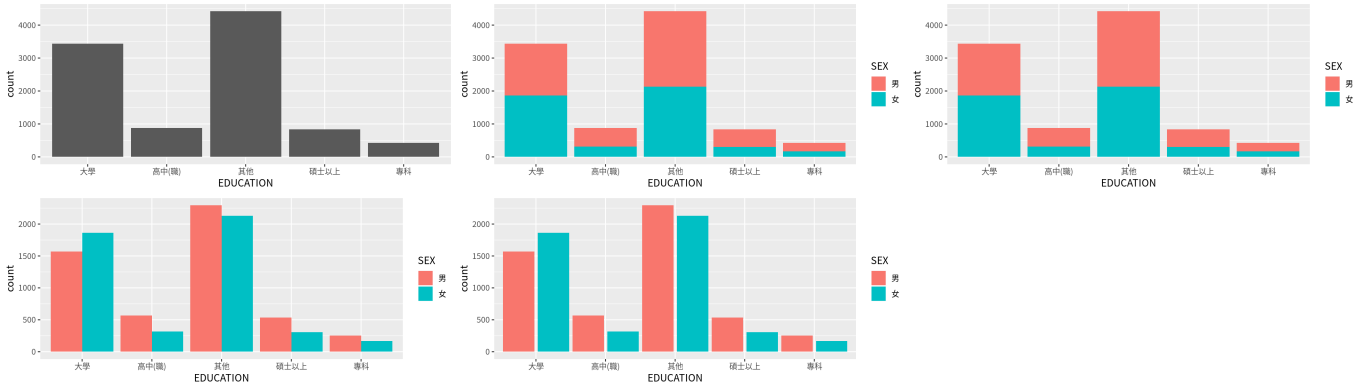
```
a=ggplot(data=data,aes(x=AMOUNT))+labs(xlabel="消費金額")

#split by sex
a+geom_histogram(aes(color=SEX))

a+geom_histogram(aes(fill=SEX))

#調透明度
a+geom_histogram(aes(color=SEX),fill="gray",alpha=0.5)

a+geom_histogram(aes(color=SEX),fill="gray",alpha=0.5,position ="dodge2")
```



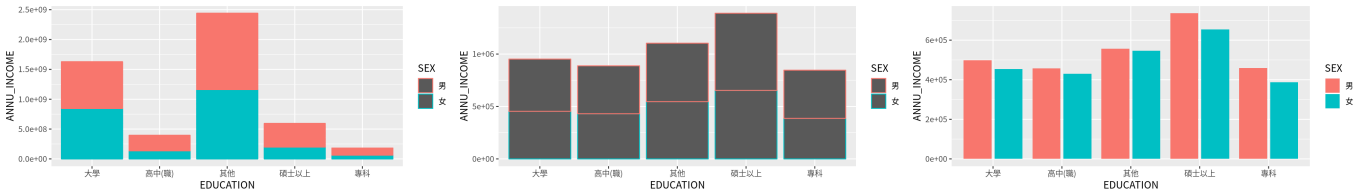
bar plot

geom_bar()

```
b=ggplot(data=data,aes(x=EDUCATION))
b+geom_bar()

b+geom_bar(aes(fill=SEX))

b+geom_bar(stat="count",aes(fill=SEX))

b+geom_bar(stat="count",aes(fill=SEX),position="dodge")

b+geom_bar(stat="count",aes(fill=SEX),position=position_dodge(1))
```



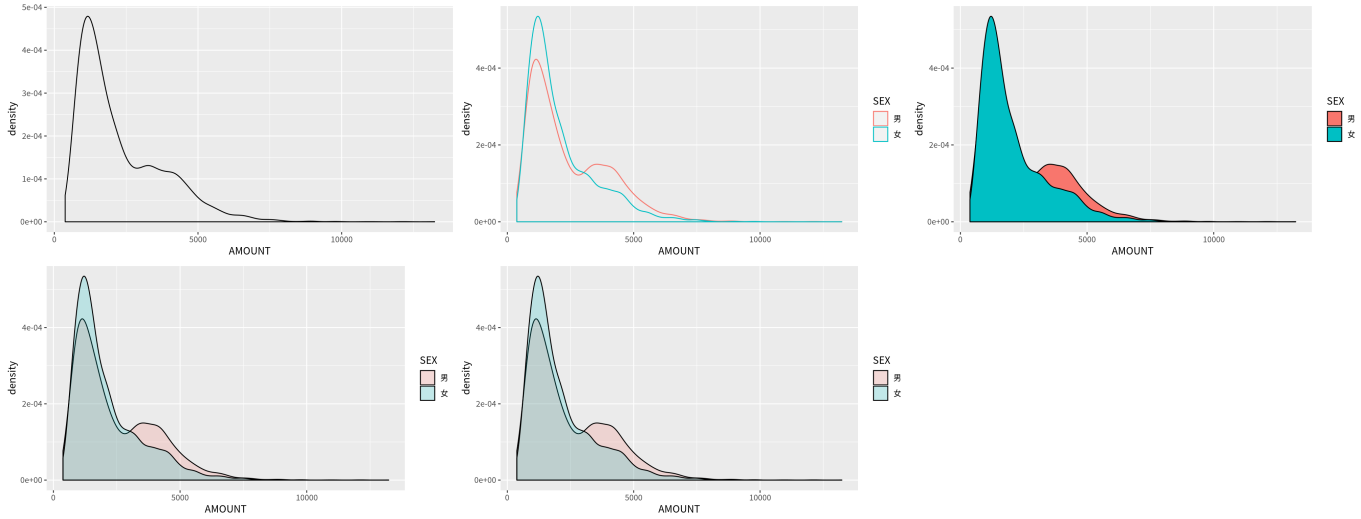

```
c=ggplot(data=data,aes(x=EDUCATION,y=ANNU_INCOME))

c+geom_bar(stat="identity",aes(color=SEX))

c+geom_bar(stat="summary",fun.y = "mean",aes(color=SEX))

c+geom_bar(stat="summary", fun.y = "mean",aes(fill=SEX),position="dodge2")
```



density

geom_density()

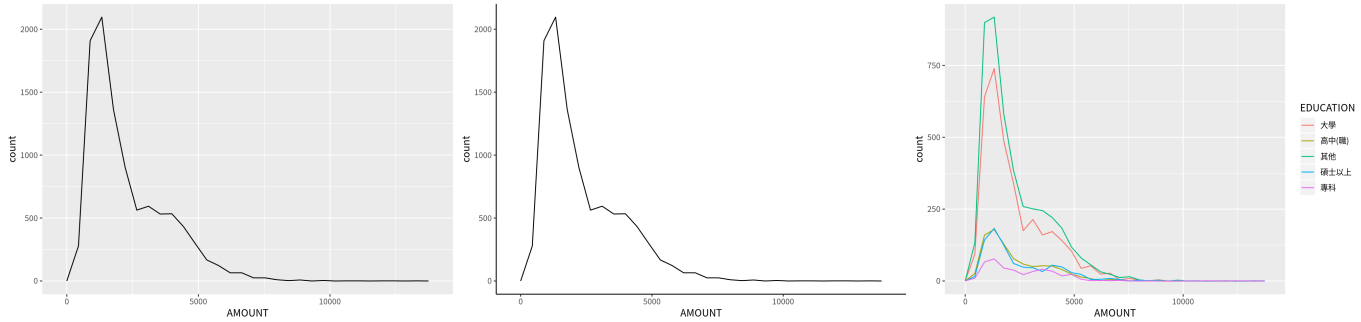
```
a+geom_density()
a+geom_density(aes(color=SEX))
a+geom_density(aes(fill=SEX))

#調透明
a+geom_density(aes(fill=SEX),alpha=0.2)
a+geom_density(aes(fill=SEX),alpha=0.2)
```



frequency

geom_freqpoly()

```
a+geom_freqpoly()

#change the theme
#theme()
a+geom_freqpoly()+theme_classic()
a+geom_freqpoly(aes(color=EDUCATION))
#fill
```

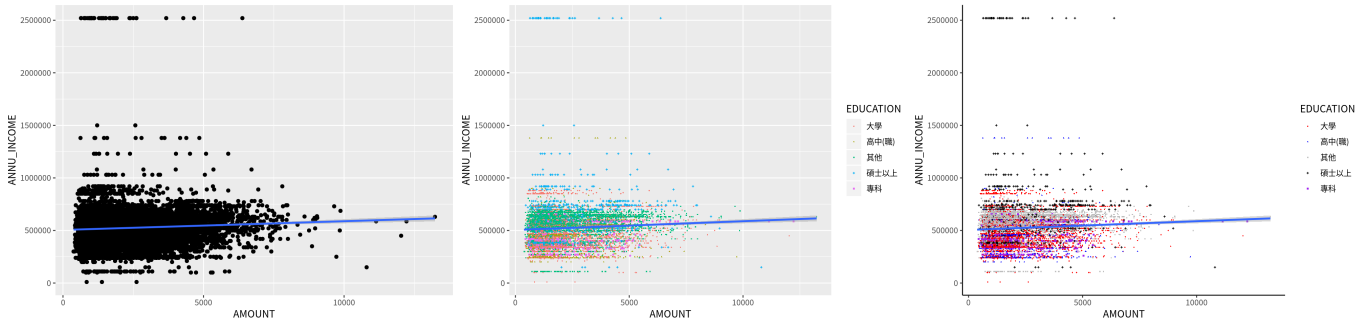


Linear model

geom_smooth()

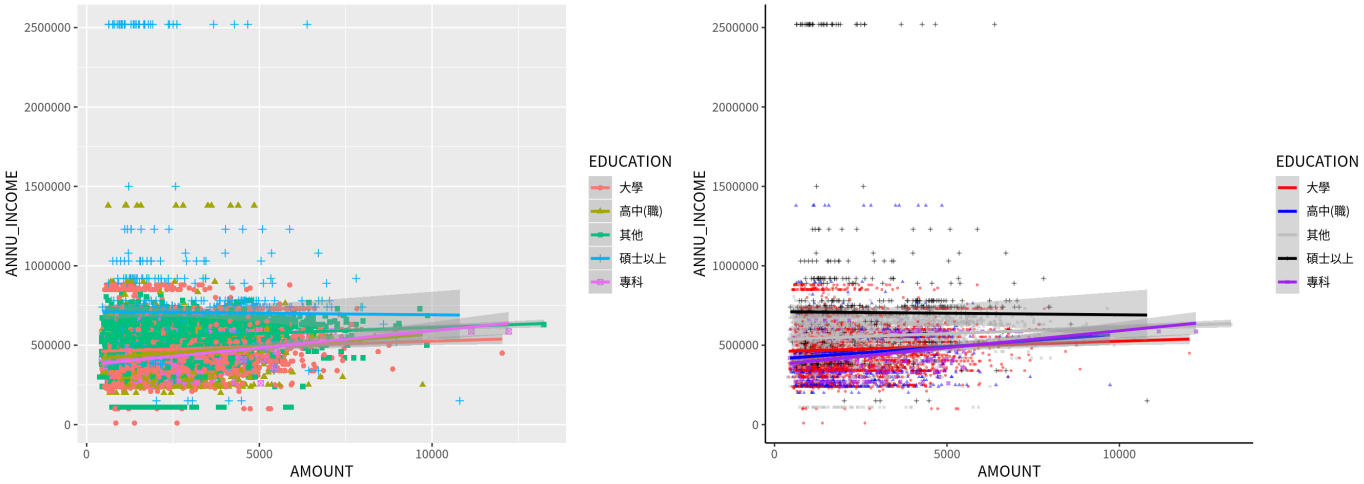
More geom_smooth() methods (https://ggplot2.tidyverse.org/reference/geom_smooth.html)

```
ggplot(data=data,aes(x=AMOUNT,y=ANNU_INCOME))+
  geom_point()+geom_smooth(method="lm")

ggplot(data=data,aes(x=AMOUNT,y=ANNU_INCOME))+
  geom_point(aes(color=EDUCATION,shape=EDUCATION),size=0.3)+
  geom_smooth(method="lm")

ggplot(data=data,aes(x=AMOUNT,y=ANNU_INCOME))+
  geom_point(aes(color=EDUCATION,shape=EDUCATION),size=0.3)+
  geom_smooth(method="lm")+
  scale_color_manual(values=c("red","blue","gray","black","purple"))+theme_classic()

#formula()

#you can see other methods by ?geom_smooth

#gg.scatter+geom_point()+geom_smooth(method="lm")+xlim(c(0,1750000)) if you want to d
elete some data

#gg.scatter+geom_point()+geom_smooth(method="lm")+coord_cartesian(xlim=c(0,1750000))
 if you just want to zoom in
```



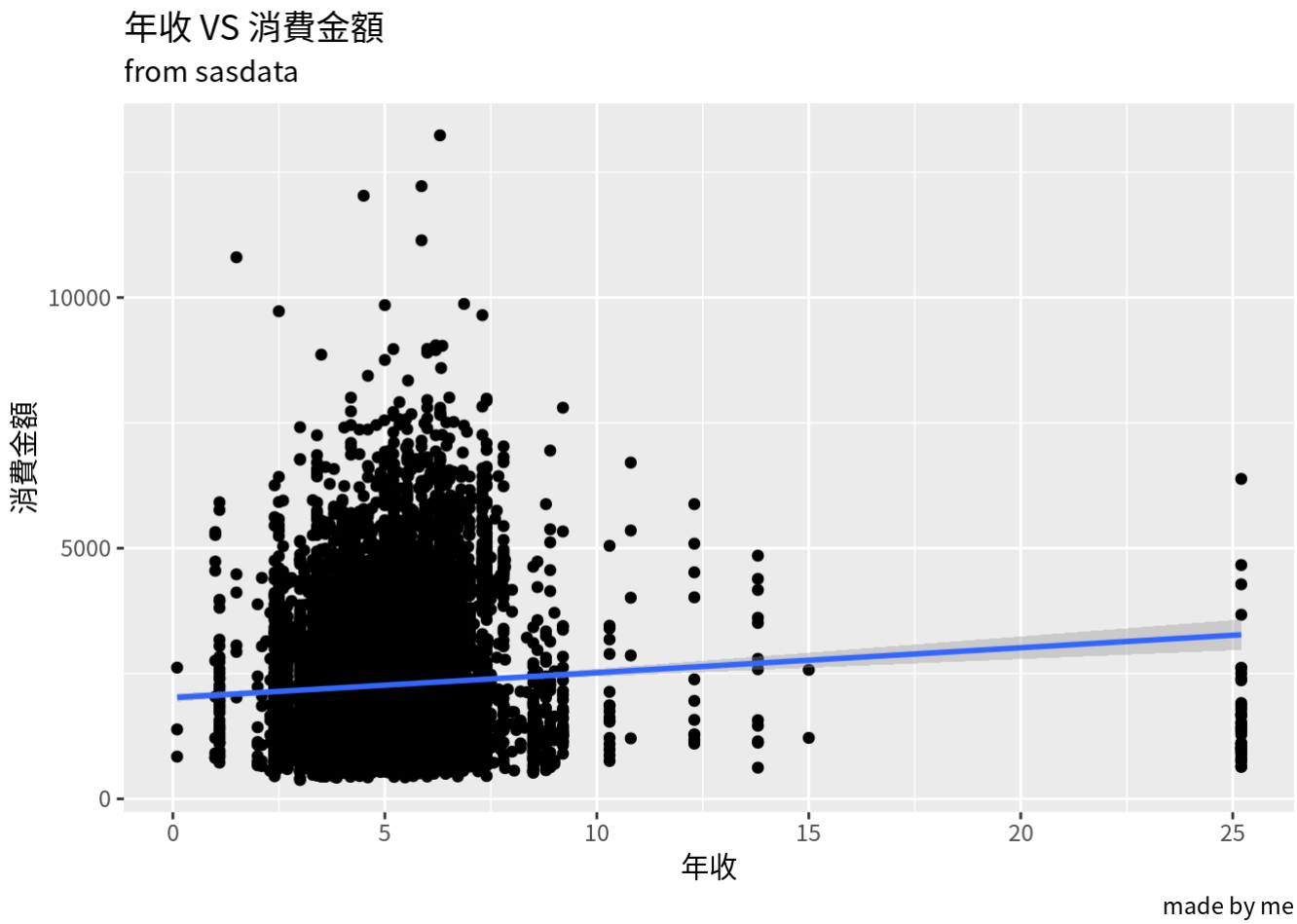
畫出分群的迴歸線！

```
ggplot(data=data,aes(x=AMOUNT,y=ANNU_INCOME))+
  geom_point(aes(color=EDUCATION,shape=EDUCATION))+
  geom_smooth(method="lm",aes(color=EDUCATION))

ggplot(data=data,aes(x=AMOUNT,y=ANNU_INCOME))+
  geom_point(aes(color=EDUCATION,shape=EDUCATION),size=0.7,alpha=0.5)+
  geom_smooth(method="lm",aes(color=EDUCATION))+
  scale_color_manual(values=c("red","blue","gray","black","purple"))+
  theme_classic()
```



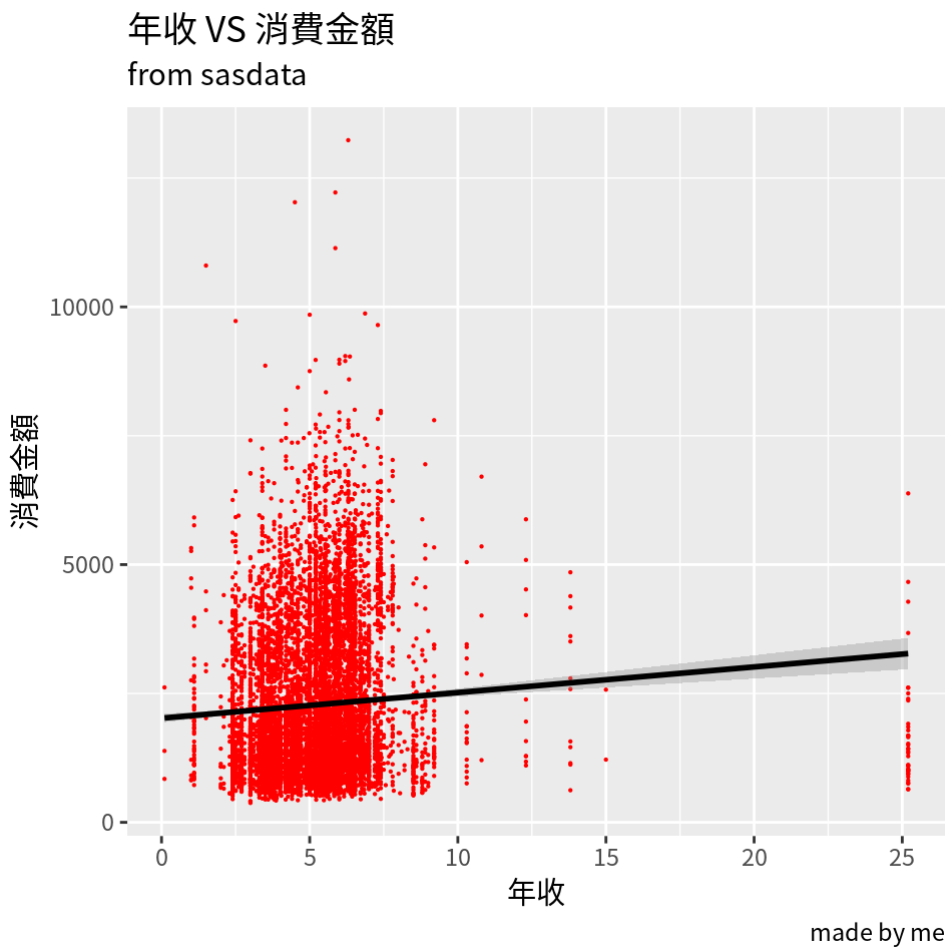
Change title and axis lable

```
gg.scatter+
  geom_point()+
  geom_smooth(method="lm")+
  labs(title="年收 VS 消費金額",subtitle ="from sasdata",x="年收",y="消費金額",caption="m
ade by me")
```



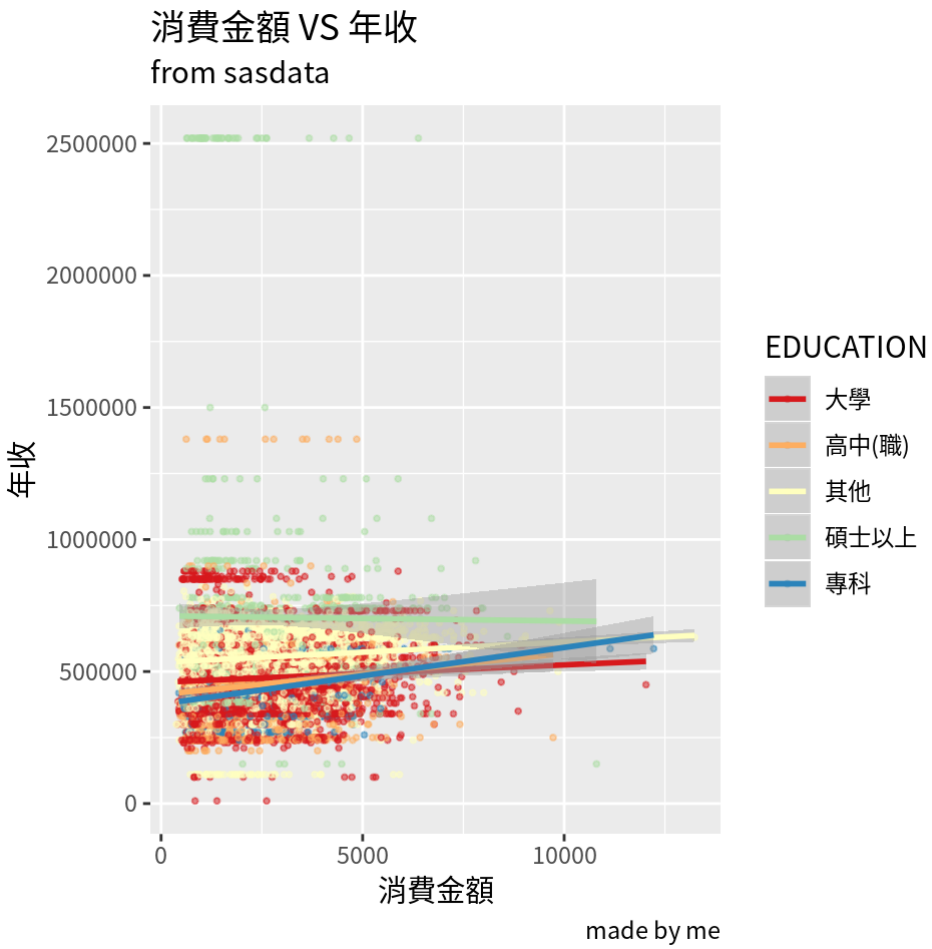
Change the color and size of point

```
gg.scatter+
  geom_point(size=0.1,col="red")+#change point
geom_smooth(method="lm",col="black")+#chnage line
labs(title="年收 VS 消費金額",subtitle ="from sasdata",x="年收",y="消費金額",caption="mad
e by me")
```



palette

```
ggplot(data=data,aes(x=AMOUNT,y=ANNU_INCOME))+
  geom_point(aes(color=EDUCATION),alpha=0.5,size=0.7)+
  geom_smooth(method="lm",aes(color=EDUCATION))+
  scale_color_brewer(palette="Spectral")+
  labs(title="消費金額 VS 年收",subtitle ="from sasdata",x="消費金額",y="年收",caption="m
ade by me")

#google for more palette
```


消費金額 VS 年收
from sasdata
2500000
2000000
1500000
1000000
500000
0
年收
0
5000
10000
消費金額
made by me
EDUCATION
大學
高中(職)
其他
碩士以上
專科