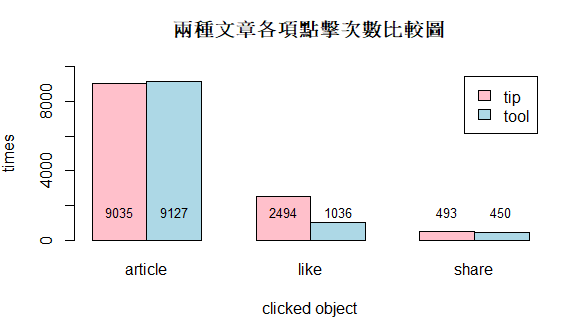
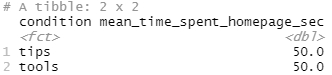
**商業分析R/SAS應用 Homework6**

107304023 統計四 張煜均

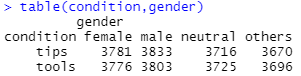
**#不同類型文章的點擊、按讚、分享數、網頁停留時間直方圖**



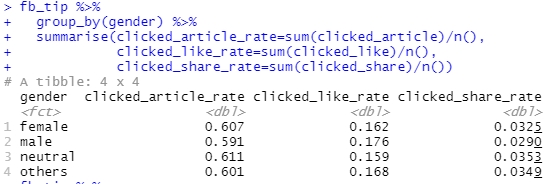
****

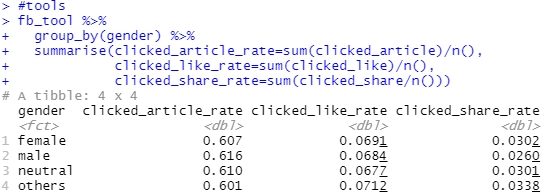
**由上圖顯示兩種文章的點擊次數、網頁停留時間皆差不多，tips類型的文章按讚數較高，文章分享數也稍微多一些，但兩種文章的分享次數差異不大。**

**#以condition分組**

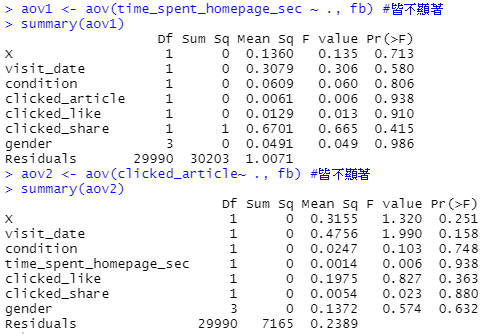


**將兩種文章的粉絲以性別分組，粉絲的性別分布很相近，因此這兩種文章的客群沒有特定性別，為大眾皆會閱讀的文章。**

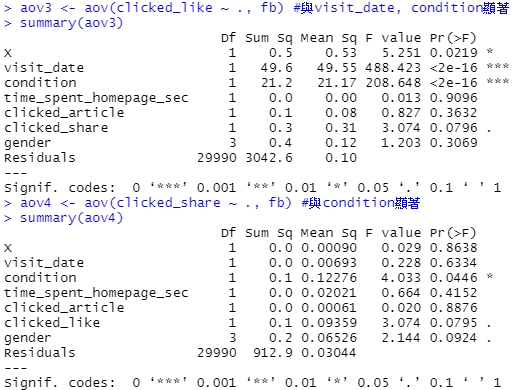
****

****

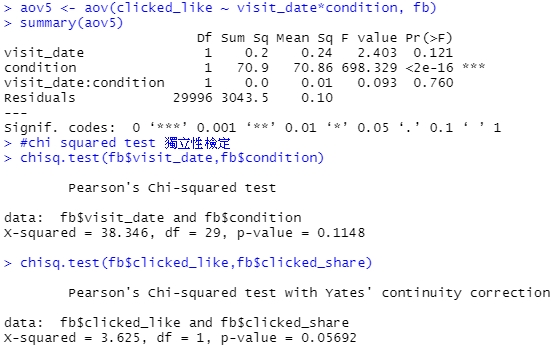
**將兩種文章用性別分組，以肉眼觀察其文章點擊數、按讚數、分享數，發現數據皆沒有很大的差異，之後將以A/B test來檢測各組數據是否有差異。**



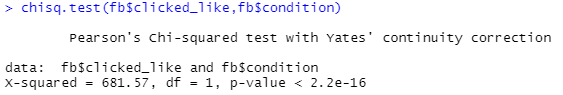
**上述檢定結果顯示，頁面停留時間、文章點擊數皆和其他變數不顯著。**

****

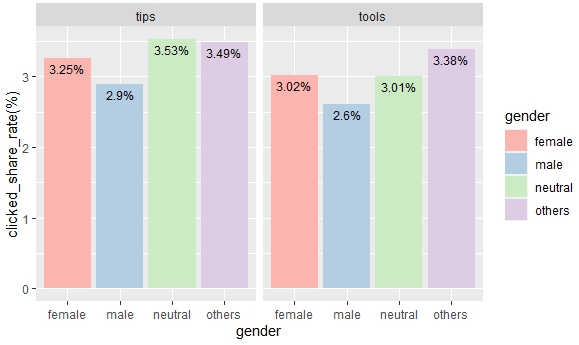
**檢定結果顯示，文章按讚數和造訪日、文章類型有顯著的關聯；而文章分享數則和文章類型有顯著相關。**

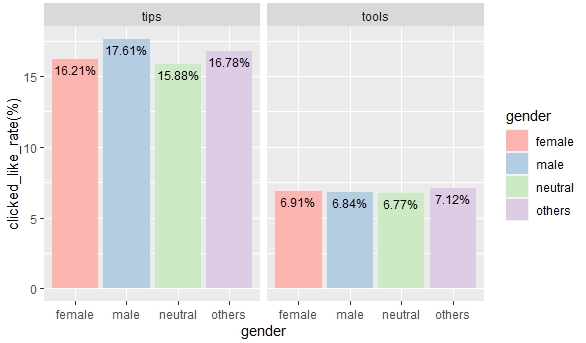
****

**由卡方獨立性檢定，檢測變數之間是否互相獨立，造訪日和文章類型檢定結果不顯著，兩者之間無顯著關聯性。**

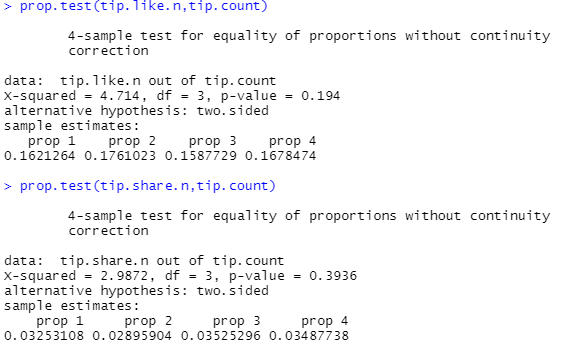


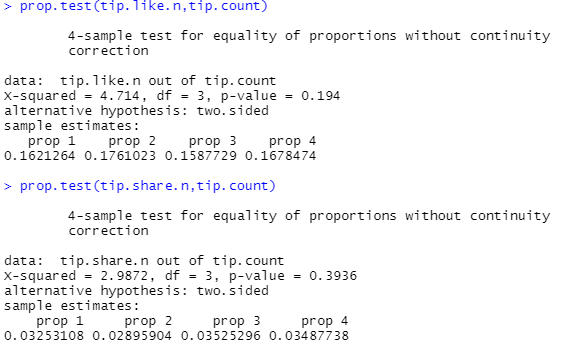
**而文章按讚數跟文章類型顯著，兩者的確有關聯。**

**#依文章和性別分組，計算按讚比例與分享比例並畫圖**

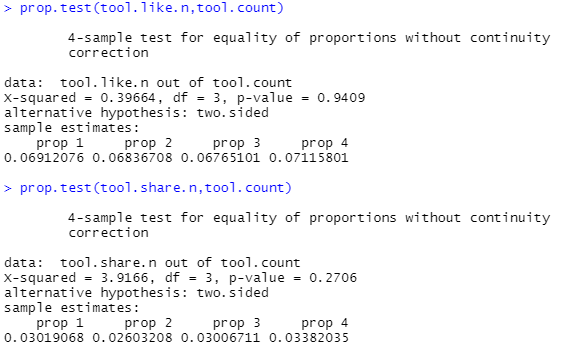
****

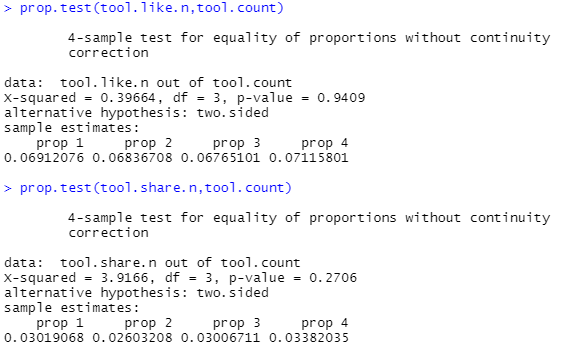
**#tips類型的各性別按讚比例與分享比例皆無顯著差異**

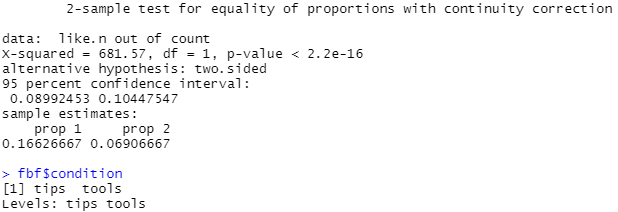




**#tools類型的各性別按讚比例與分享比例皆無顯著差異**

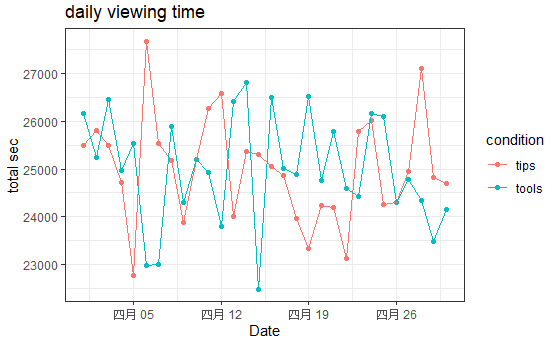






**總結上述結果，文章的按讚數僅和文章類型有顯著相關，且兩文章的按讚比例顯著不同，即tips類型文章的按讚率顯著高於tools類的文章。而文章的分享數和文章類型也有顯著相關，由前方的直方圖可知，tips文章的分享數也高於tools類的文章。**

**#結論**



**上圖為兩種文章每日的總停留時間，tips類的文章在4/6及4/28有較高的觀看時間，而這兩天tools類的文章觀看時間較短，仔細觀察可發現兩種文章的觀看時長大部分呈現負相關，可能是某種文章曝光時，另一種文章容易被忽略，綜合上述分析，推測該網紅的粉絲較喜歡tips類型文章，故建議網紅可專注在此類文章，以增加文章的整體曝光度、按讚數及分享數。**

**#程式碼**

**setwd("C:/Users/USER/Downloads")**

**fb <- read.csv("hw6-fb.csv")**

**library(tidyverse)**

**library(ggplot2)**

**attach(fb)**

**names(fb)**

**str(fb)**

**fb$gender <- as.factor(gender)**

**fb$visit\_date <- as.Date(visit\_date)**

**fb$condition <- as.factor(condition)**

**#資料分析**

**summary(fb)**

**fb1 <- fb %>%**

**group\_by(condition) %>%**

**summarise(mean\_spent\_sec=mean(time\_spent\_homepage\_sec),**

**article.n=sum(clicked\_article),**

**like.n=sum(clicked\_like),**

**share.n=sum(clicked\_share))**

**fb1 = as.matrix(fb1[,3:5])**

**b1 <- barplot(fb1, names.arg=c('article','like','share'), col=c("pink","lightblue"),**

**xlab="clicked object", ylab="times", ylim=c(0,10000),beside=T,**

**legend.text=c('tip','tool'),main="兩種文章各項點擊次數比較圖")**

**text(b1,labels=fb1,y=2,pos=3,offset=1.2,cex=0.8)**

**fb2 <- fb %>%**

**group\_by(visit\_date,condition) %>%**

**mutate(total\_sec=sum(time\_spent\_homepage\_sec))**

**ggplot(fb2, aes(x = visit\_date, y = total\_sec, colour = condition)) +**

**geom\_point() + geom\_line() +**

**xlab("Date") + ylab("total sec") +**

**ggtitle(label="daily viewing time") +**

**theme\_bw()**

**#分組**

**table(condition,gender)**

**fb\_tip <- fb %>% filter(fb$condition=="tips")**

**fb\_tool <- fb %>% filter(fb$condition=="tools")**

**#tips**

**fb\_tip %>%**

**group\_by(gender) %>%**

**summarise(clicked\_article\_rate=sum(clicked\_article)/n(),**

**clicked\_like\_rate=sum(clicked\_like)/n(),**

**clicked\_share\_rate=sum(clicked\_share)/n())**

**fb\_tip %>%**

**group\_by(gender) %>%**

**summarise(mean\_time\_spent\_homepage\_sec=mean(time\_spent\_homepage\_sec))**

**#tools**

**fb\_tool %>%**

**group\_by(gender) %>%**

**summarise(clicked\_article\_rate=sum(clicked\_article)/n(),**

**clicked\_like\_rate=sum(clicked\_like)/n(),**

**clicked\_share\_rate=sum(clicked\_share/n()))**

**fb\_tool %>%**

**group\_by(gender) %>%**

**summarise(mean\_time\_spent\_homepage\_sec=mean(time\_spent\_homepage\_sec))**

**#anova**

**#檢定網頁停留時間與什麼變數顯著**

**aov1 <- aov(time\_spent\_homepage\_sec ~ ., fb) #皆不顯著**

**aov2 <- aov(clicked\_article~ ., fb) #皆不顯著**

**aov3 <- aov(clicked\_like ~ ., fb) #與visit\_date, condition顯著**

**aov4 <- aov(clicked\_share ~ ., fb) #與condition顯著**

**summary(aov1)**

**summary(aov2)**

**summary(aov3)**

**summary(aov4)**

**#chi squared test 獨立性檢定**

**chisq.test(fb$visit\_date,fb$condition)**

**chisq.test(fb$clicked\_like,fb$condition)**

**#依文章和性別分組，計算案讚比例與分享比例**

**fb1 <- fb %>%**

**group\_by(condition,gender) %>%**

**summarise(clicked\_article\_rate=sum(clicked\_article)/n()\*100,**

**clicked\_like\_rate=sum(clicked\_like)/n()\*100,**

**clicked\_share\_rate=sum(clicked\_share)/n()\*100)**

**ggplot(fb1, aes(x=gender,y=clicked\_like\_rate,fill=gender))+**

**geom\_col()+**

**geom\_text(aes(label=paste0(round(clicked\_like\_rate,2),"%")),vjust=1.5,size=3.1)+**

**ylab("clicked\_like\_rate(%)")+**

**facet\_wrap(~condition)+**

**scale\_fill\_brewer(palette = "Pastel1")**

**ggplot(fb1, aes(x=gender,y=clicked\_share\_rate,fill=gender))+**

**geom\_col()+**

**geom\_text(aes(label=paste0(round(clicked\_share\_rate,2),"%")),vjust=1.5,size=3.1)+**

**ylab("clicked\_share\_rate(%)")+**

**facet\_wrap(~condition)+**

**scale\_fill\_brewer(palette = "Pastel1")**

**#分組中各性別按讚比例與分享比例是否有顯著差異**

**fbb <- fb %>%**

**group\_by(condition,gender) %>%**

**summarise(like\_n=sum(clicked\_like),**

**share\_n=sum(clicked\_share), count=n())**

**tip.like.n = as.numeric(fbb[1:4,3]$like\_n)**

**tip.count = as.numeric(fbb[1:4,5]$count)**

**prop.test(tip.like.n,tip.count)**

**tip.share.n = as.numeric(fbb[1:4,4]$share\_n)**

**tip.count = as.numeric(fbb[1:4,5]$count)**

**prop.test(tip.share.n,tip.count)**

**tool.like.n = as.numeric(fbb[5:8,3]$like\_n)**

**tool.count = as.numeric(fbb[5:8,5]$count)**

**prop.test(tool.like.n,tool.count)**

**tool.share.n = as.numeric(fbb[5:8,4]$share\_n)**

**tool.count = as.numeric(fbb[5:8,5]$count)**

**prop.test(tool.share.n,tool.count)**

**#兩種文章按讚比例是否顯著不同**

**fbf <- fb %>%**

**group\_by(condition) %>%**

**summarise(like\_n=sum(clicked\_like), count=n())**

**like.n = as.numeric(fbf$like\_n)**

**count = as.numeric(fbf$count)**

**prop.test(like.n,count)**

**fbf$condition**