R_Finalexam_code and output　107508006 歐西四 陳葳芃

```
###[0-資料清理]
setwd("~/Downloads/1102 R/Final")
library(readr)
data <- read_csv("Data E.csv")
summary(data)
#缺失值疑除
data<-na.omit(data)

##資料觀察
summary(data)
ls(data)
colnames(data) <- gsub(" ", "_", colnames(data))


###[1-入榜頻道類型分析]
#頻道類型走向建議
data %>%
  group_by(category)%>%
  count()

###[2-訂閱數觀看數分析]
##看看subscribers(訂閱數)及video_views(影片觀看數)關係走向
library(tidyverse)
data %>%
  ggplot(aes(x=subscribers, y=video_views)) +
  geom_point()

###[3-找出流量密碼]
##找出每個種類下訂閱數字最多的人
top_subscribers<-data %>%
  group_by(category)%>%
  slice(which.max(subscribers))
top_subscribers
##找出每個種類下觀看次數最多的人
top_video_views<-data %>%
  group_by(category)%>%
  slice(which.max(video_views))
top_video_views

###[4-頻道建立趨勢]
##畫圖來觀看頻道種類頻道建立趨勢
ggplot(data=data, aes(x=started, y=video_views, fill=category)) +
  geom_bar(stat="identity", color="black", position=position_dodge())+
  theme_minimal()
##因為1970值年代太久遠，所以將其視為離群值移除
delete<-which(data$started==1970)
data2 <- data[-delete,]
##畫圖來觀看頻道種類趨勢2.0
ggplot(data=data2, aes(x=started, y=video_views, fill=category)) +
  geom_bar(stat="identity", color="black", position=position_dodge())+
  theme_minimal()
```

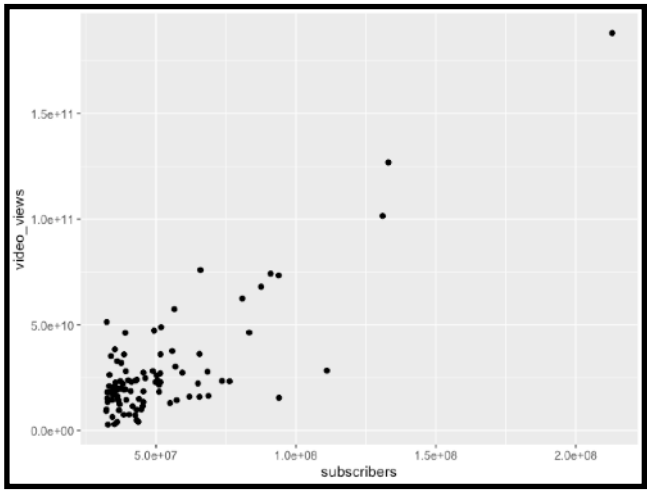
>>程式碼output

```
> ###[0-資料清理]
> setwd("~/Downloads/1102 R/Final")
> library(readr)
> data <- read_csv("Data E.csv")
Rows: 100 Columns: 7
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (2): youtuber, category
dbl (4): rank, subscribers, video count, started

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
> summary(data)
      rank          youtuber          subscribers         video views
 Min.   :  1.00   Length:100         Min.   : 32200000   Min.
:2.648e+09
 1st Qu.: 25.75   Class :character   1st Qu.: 36200000   1st
Qu.:1.486e+10
 Median : 50.50   Mode  :character   Median : 43200000   Median
:2.210e+10
 Mean   : 50.50                      Mean   : 53363000   Mean
:2.993e+10
 3rd Qu.: 75.25                      3rd Qu.: 57100000   3rd
Qu.:3.053e+10
 Max.   :100.00                      Max.   :213000000   Max.
:1.881e+11

  video count         category           started
 Min.   :    45.0   Length:100         Min.   :1970
 1st Qu.:   393.5   Class :character   1st Qu.:2008
 Median :  1139.0   Mode  :character   Median :2012
 Mean   : 15847.2                      Mean   :2011
 3rd Qu.:  4986.0                      3rd Qu.:2014
 Max.   :209351.0                      Max.   :2018
 NA's   :5
> #缺失值疑除
> data<-na.omit(data)
> ##資料觀察
> summary(data)
      rank          youtuber          subscribers         video views
 Min.   :  1.00   Length:95          Min.   : 32200000   Min.
:2.648e+09
 1st Qu.: 28.50   Class :character   1st Qu.: 36200000   1st
Qu.:1.456e+10
 Median : 52.00   Mode  :character   Median : 42700000   Median
:2.088e+10
 Mean   : 52.02                      Mean   : 51232632   Mean
:2.767e+10
 3rd Qu.: 75.50                      3rd Qu.: 55450000   3rd
Qu.:2.815e+10
```

```
 Max.   :100.00                       Max.   :213000000   Max.
:1.881e+11
  video count         category            started
 Min.   :    45.0   Length:95          Min.   :1970
 1st Qu.:   393.5   Class :character   1st Qu.:2007
 Median :  1139.0   Mode  :character   Median :2011
 Mean   : 15847.2                      Mean   :2011
 3rd Qu.:  4986.0                      3rd Qu.:2014
 Max.   :209351.0                      Max.   :2018
> ls(data)
[1] "category"    "rank"        "started"     "subscribers" "video count"
"video views"
[7] "youtuber"
> colnames(data) <- gsub(" ", "_", colnames(data))
> ###[1-入榜頻道類型分析]
> #頻道類型走向建議
> data %>%
+   group_by(category)%>%
+   count()
# A tibble: 13 × 2
# Groups:   category [13]
   category                  n
   <chr>                 <int>
 1 Comedy                    4
 2 Education                 6
 3 Entertainment            19
 4 Film & Animation          4
 5 Gaming                    7
 6 How to & Style            2
 7 Music                    33
 8 News & Politics           2
 9 Nonprofits & Activism     1
10 People & Blogs           10
11 Shows                     4
12 Sports                    2
13 Trailers                  1
> ###[2-訂閱數觀看數分析]
> ##看看subscribers(訂閱數)及video_views(影片觀看數)關係走向
> library(tidyverse)
> data %>%
+   ggplot(aes(x=subscribers, y=video_views)) +
+   geom_point()
```



```
> ###[3-找出流量密碼]
> ##找出每個種類下訂閱數字最多的人
> top_subscribers<-data %>%
+   group_by(category)%>%
+   slice(which.max(subscribers))
> top_subscribers
# A tibble: 13 × 7
# Groups:   category [13]
    rank youtuber            subscribers video_views video_count category
started
   <dbl> <chr>                     <dbl>       <dbl>       <dbl> <chr>
<dbl>
 1    48 Whinderssonnunes       43700000     4.06e 9         517 Comedy
2013
 2     3 Cocomelon - Nurser…   133000000     1.27e11         751
Education          2006
 3     7 MrBeast                 93900000     1.54e10         721
Entertainment      2012
 4    17 Goldmines               68800000     1.63e10        3050 Film &
Animati…    2012
 5     6 PewDiePie              111000000     2.83e10        4472 Gaming
2010
 6    14 5-Minute Crafts         76300000     2.32e10        5281 How to &
Style     2016
 7     1 T-Series               213000000     1.88e11       16708 Music
2006
 8    34 Aaj Tak                 51100000     1.82e10      175877 News &
Politics    2009
 9    87 TEDx Talks              34500000     6.36e 9      179816
Nonprofits & A…    2009
10     8 Kids Diana Show         93800000     7.33e10         977 People &
Blogs     2015
11     4 SET India              131000000     1.02e11       78334 Shows
2006
12    11 WWE                     87600000     6.80e10       60699 Sports
2007
13    72 Ishtar Music            36600000     1.42e10        4496 Trailers
2005
> ##找出每個種類下觀看次數最多的人
> top_video_views<-data %>%
+   group_by(category)%>%
+   slice(which.max(video_views))
> top_video_views
# A tibble: 13 × 7
# Groups:   category [13]
    rank youtuber            subscribers video_views video_count category
started
   <dbl> <chr>                     <dbl>       <dbl>       <dbl> <chr>
<dbl>
 1    88 Jkk Entertainment       34500000     1.44e10         186 Comedy
2017
 2     3 Cocomelon - Nurser…   133000000     1.27e11         751
Education          2006
```

```
 3     13 Vlad and Niki           80900000     6.24e10         443
Entertainment      2018
 4     27 Movieclips              56600000     5.73e10       38746 Film &
Animati…    2006
 5      6 PewDiePie              111000000     2.83e10        4472 Gaming
2010
 6     14 5-Minute Crafts         76300000     2.32e10        5281 How to &
Style     2016
 7      1 T-Series               213000000     1.88e11       16708 Music
2006
 8     34 Aaj Tak                 51100000     1.82e10      175877 News &
Politics    2009
 9     87 TEDx Talks              34500000     6.36e 9      179816
Nonprofits & A…    2009
10     10 Like Nastya             91000000     7.42e10         670 People &
Blogs     2016
11      4 SET India              131000000     1.02e11       78334 Shows
2006
12     11 WWE                     87600000     6.80e10       60699 Sports
2007
13     72 Ishtar Music            36600000     1.42e10        4496 Trailers
2005
> ###[4-頻道建立趨勢]
> ##畫圖來觀看頻道種類頻道建立趨勢
> ggplot(data=data, aes(x=started, y=video_views, fill=category)) +
+   geom_bar(stat="identity", color="black", position=position_dodge())+
+   theme_minimal()
> ##因為1970值年代太久遠，所以將其視為離群值移除
> delete<-which(data$started==1970)
> data2 <- data[-delete,]
> ##畫圖來觀看頻道種類趨勢2.0
> ggplot(data=data2, aes(x=started, y=video_views, fill=category)) +
+   geom_bar(stat="identity", color="black", position=position_dodge())+
+   theme_minimal()
```