TIBAME線上課程
爬蟲分析

組員：郭郁鴻、潘德春、廖威程

目錄

- 爬蟲網站介紹
- 爬蟲程式說明
- 商業意涵分析
- 結論

PART1、爬蟲網站介紹

爬蟲網站

- TibaMe 全方位線上課程與職能學習平台
- https://www.tibame.com/courselibrary (取自2019/4/1資料)

爬蟲網站介紹-課程類別

爬蟲網站介紹-課程標籤

PART2、爬蟲程式說明

爬蟲程式說明-問題

- 發現網站會擋住爬蟲程式
- 將網頁原始碼複製出來放入localhost抓

```
> doc <- read_html("https://www.tibame.com/courselibrary/ai")
> text <- html_nodes(doc, ".course-card__-x2XW .card-title")
> head(text)
{xml_nodeset (0)}
> html_nodes(doc, ".course-card__-x2XW .card-title")
{xml_nodeset (0)}
> html_nodes(doc, ".card-title")
{xml_nodeset (0)}
```

爬蟲程式說明(一)-AI人工智慧

■ 每個陣列第一格對應第一堂課程

```
> ai <- read_html("http://localhost/tibame_ai.html")
> ai_classname <- html_nodes(ai, ".card-title") %>% html_text()
> ai_classprice <- html_nodes(ai, ".current-price___2Zu5l") %>% html_text()
> ai_numofpeople <- html_nodes(ai, ".footer___2mND7 div:first-child span:first-child") %>% html_text()
> ai_classlen <- html_nodes(ai, ".footer___2mND7 div:first-child span:nth-child(2)") %>% html_text()
>
> ai_classname
 [1] "尹相志深度學習實戰　大師跨海教學直播課程"          "Python AI三部曲-精通程式設計到深度學習"
 [3] "【解鎖課程】Python人工智慧入門：機器學習到深..." "Python人工智慧入門：機器學習到深度學習"
 [5] "地表最好懂的Python網路爬蟲 零基礎也能學會！"     "【解鎖課程】地表最好懂的網路爬蟲彩蛋內容！"
 [7] "AI第二部曲-Python機器學習"                        "AI第三部曲-Python深度學習"
 [9] "關聯式資料庫設計SQL Server和MySQL資..."          "AI首部曲-Python從零開始精通程式設計"
[11] "Python 輕鬆上手學"                                "Python 網站擷取與資料分析"
[13] "Hadoop架設與分析技術"                             "Spark首部曲：實務基礎入門篇"
[15] "第一張MPP微軟專業認證 - AI人工智慧實作入門"      "【立即考照】MPP微軟專業認證 - Python資..."
[17] "NoSQL輕鬆學：架構、設計與案例探討"               "R語言/R-Studio入門-統計分析、資料探勘與..."
[19] "Spark第三部曲： MLlib完成資料挖掘與機器..."      "ETL網路爬蟲實戰技巧"
[21] "教你完全掌握 SQL- 從資料庫建立到成為資料處理..." "MongoDB快速建構實戰"
[23] "資料探勘一步到位：原理與分類/聚類演算"           "Spark第二部曲： SQL輕鬆處理半結構化資料技..."
[25] "MySQL資料庫管理技巧"                              "資料庫設計原理"
[27] "Chatbot的趨勢與實務應用"                          "第一次用Python開發Web就上手－Taipei..."
[29] "Python技術者年會－PyCon Taiwan ..."              "AI人工智慧語意分割的技術與應用-微軟Azure獨..."
> ai_classprice
 [1] "NT$12,000" "NT$5,700"  "NT$5,000"  "NT$4,800"  "NT$4,200"  "NT$3,000"  "NT$2,200"  "NT$2,000"  "NT$1,900"  "NT$1,500"
[11] "NT$1,390"  "NT$1,350"  "NT$1,245"  "NT$1,200"  "NT$1,050"  "NT$1,050"  "NT$890"    "NT$813"    "NT$700"    "NT$700"
[21] "NT$699"    "NT$650"    "NT$650"    "NT$650"    "NT$390"    "NT$100"    "免費"      "免費"      "免費"      "免費"
> ai_numofpeople
 [1] " 75人"    " 86人"    " 30人以下" " 54人"    " 142人"   " 30人以下" " 30人以下" " 30人以下" " 94人"    " 36人"
[11] " 920人"   " 36人"    " 379人"   " 92人"    " 277人"   " 567人"   " 280人"   " 567人"   " 65人"    " 258人"
[21] " 30人以下" " 154人"   " 284人"   " 58人"    " 529人"   " 403人"   " 1,389人" " 1,673人" " 1,962人" " 835人"
> ai_classlen
 [1] " 90h17m" " 50h29m" " 02h20m" " 53h37m" " 77h52m" " 08h12m" " 14h39m" " 21h12m" " 19h13m" " 20h38m" " 04h42m" " 19h06m" " 08h36m"
[14] " 03h00m" " 03h24m" " 12h07m" " 02h52m" " 04h02m" " 01h28m" " 02h21m" " 05h01m" " 02h39m" " 03h40m" " 01h13m" " 04h00m" " 02h49m"
[27] " 02h25m" " 01h47m" " 13h54m" " 01h24m"
```

爬蟲程式說明(二)-雲端技術

```
> cloud_computing <- read_html("http://localhost/tibame_cloud_computing.html")
> cloud_computing_classname <- html_nodes(cloud_computing, ".card-title") %>% html_text()
> cloud_computing_classprice <- html_nodes(cloud_computing, ".current-price___2Zu5l") %>% html_text()
> cloud_computing_numofpeople <- html_nodes(cloud_computing, ".footer___2mND7 div:first-child span:first-child") %>% html_text()
> cloud_computing_classlen <- html_nodes(cloud_computing, ".footer___2mND7 div:first-child span:nth-child(2)") %>% html_text()
>
> cloud_computing_classname
 [1] "Java Web_深入Spring Framewo..."          "從實務出發：企業資訊安全架構與方案解析 | 資安線..."
 [3] "時代新主力-3 大面向透視雲端架構"           "MIS網管人必學 | Wireshark操作與分析..."
 [5] "實戰子網路切割-MIS網管工程師必備技能"       "Struts 2 技術實作"
 [7] "吳永志Java Web 技術開發 | JSP + ..."        "Oracle APEX 入門實戰！讓你的 Exce..."
 [9] "從零開始拿高分 | 實戰APCS大學程式設計先修檢..."  "MIS網管人必學 | Wireshark操作與分析..."
[11] "開發人員 3 小時速成學習 – Windows C..."     "【最優惠的早鳥預售開跑囉】一鍵上手 || AWS認..."
[13] "Java 起步走!"                             "Agile 敏捷開發方法"
[15] "軟體版本控管實務"                          "Windows Container入門與實作"
[17] "敏捷開發『隱藏的質量』問題- Agile Taip..."    "Agile Summit 2018敏捷高峰會"
[19] "敏捷，談談那些坑和故事"                      "微軟DevOps轉型之旅"
> cloud_computing_classprice
 [1] "NT$10,000" "NT$5,070"  "NT$3,200"  "NT$3,000"  "NT$2,880"  "NT$2,100"  "NT$1,840"  "NT$1,280"  "NT$1,020"  "NT$1,000"
[11] "NT$990"    "NT$900"    "NT$890"    "NT$590"    "NT$340"    "免費"       "免費"       "免費"       "免費"       "免費"
> cloud_computing_numofpeople
 [1] " 30人以下" " 30人以下" " 441人"    " 30人以下" " 30人以下" " 109人"    " 237人"    " 30人以下" " 33人"     " 30人以下"
[11] " 30人以下" " 30人以下" " 363人"    " 1,376人"  " 607人"    " 848人"    " 223人"    " 775人"    " 262人"    " 195人"
> cloud_computing_classlen
 [1] " 40h27m" " 05h31m" " 09h30m" " 04h17m" " 08h42m" " 06h57m" " 09h27m" " 02h57m" " 03h34m" " 01h54m" " 02h57m" " 01h16m" " 08h44m"
[14] " 02h37m" " 00h48m" " 01h17m" " 01h47m" " 08h51m" " 01h40m" " 01h51m"
```

PART3、商業意涵分析

商業意涵分析-介紹

■ 整體：
　　所有課程免費/付費人數、總長度
■ 概況：
　　各類課程人數、價格、總長度分布
■ 細節：
　　某一類課程人數、價格、總長度分析

所有免費/付費課程人數-資料清理

■ 將各類課程按照免費/付費分類

> free_ai_numofpeople <- ai_numofpeople[27:30]

> cost_ai_numofpeople <- ai_numofpeople[1:26]

> free_blockchain_numofpeople <- blockchain_numofpeople[5:21]

> cost_blockchain_numofpeople <- blockchain_numofpeople[1:4]

> free_e_commerce_numofpeople <- e_commerce_numofpeople[31:32]

> cost_e_commerce_numofpeople <- e_commerce_numofpeople[1:30]

> free_game_design_numofpeople <- game_design_numofpeople[34:35]

> cost_game_design_numofpeople <- game_design_numofpeople[1:33]

> free_digital_workforce_numofpeople <- digital_workforce_numofpeople[34:37]

> cost_digital_workforce_numofpeople <- digital_workforce_numofpeople[1:33]

> free_cloud_computing_numofpeople <- cloud_computing_numofpeople[16:20]

> cost_cloud_computing_numofpeople <- cloud_computing_numofpeople[1:15]

> free_mobile_apps_numofpeople <- mobile_apps_numofpeople[18:19]

> cost_mobile_apps_numofpeople <- mobile_apps_numofpeople[1:17]

> free_iot_numofpeople <- iot_numofpeople[5:12]

> cost_iot_numofpeople <- iot_numofpeople[1:4]

■ 清理完後存入變數

> free_all_numofpeople <- free_ai_numofpeople

> free_all_numofpeople <- append(free_all_numofpeople ,free_blockchain_numofpeople)

> free_all_numofpeople <- append(free_all_numofpeople ,free_e_commerce_numofpeople)

> free_all_numofpeople <- append(free_all_numofpeople ,free_game_design_numofpeople)

> free_all_numofpeople <- append(free_all_numofpeople ,free_digital_workforce_numofpeople)

> free_all_numofpeople <- append(free_all_numofpeople ,free_cloud_computing_numofpeople)

> free_all_numofpeople <- append(free_all_numofpeople ,free_mobile_apps_numofpeople)

> free_all_numofpeople <- append(free_all_numofpeople ,free_iot_numofpeople)

所有免費/付費課程人數-敘述統計

> 免費課程人數 <- summary(free_all_numofpeople)

> 付費課程人數 <- summary(cost_all_numofpeople)

> SUMMARY_all_numofpeople <-rbind(免費課程人數,付費課程人數)

> SUMMARY_all_numofpeople

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
免費課程人數	30	169	308	539.7500	721.00	1962
付費課程人數	30	30	30	180.0185	98.75	2331

所有免費/付費課程人數-圖表

> boxplot(free_all_numofpeople,cost_all_numofpeople,main="免費/付費課程-人數比較", xlab="費用別",ylab="修課人數",col=rainbow(2))

> axis(side=1, las=2, at = 1:2, labels=c("免費課程人數","付費課程人數"))
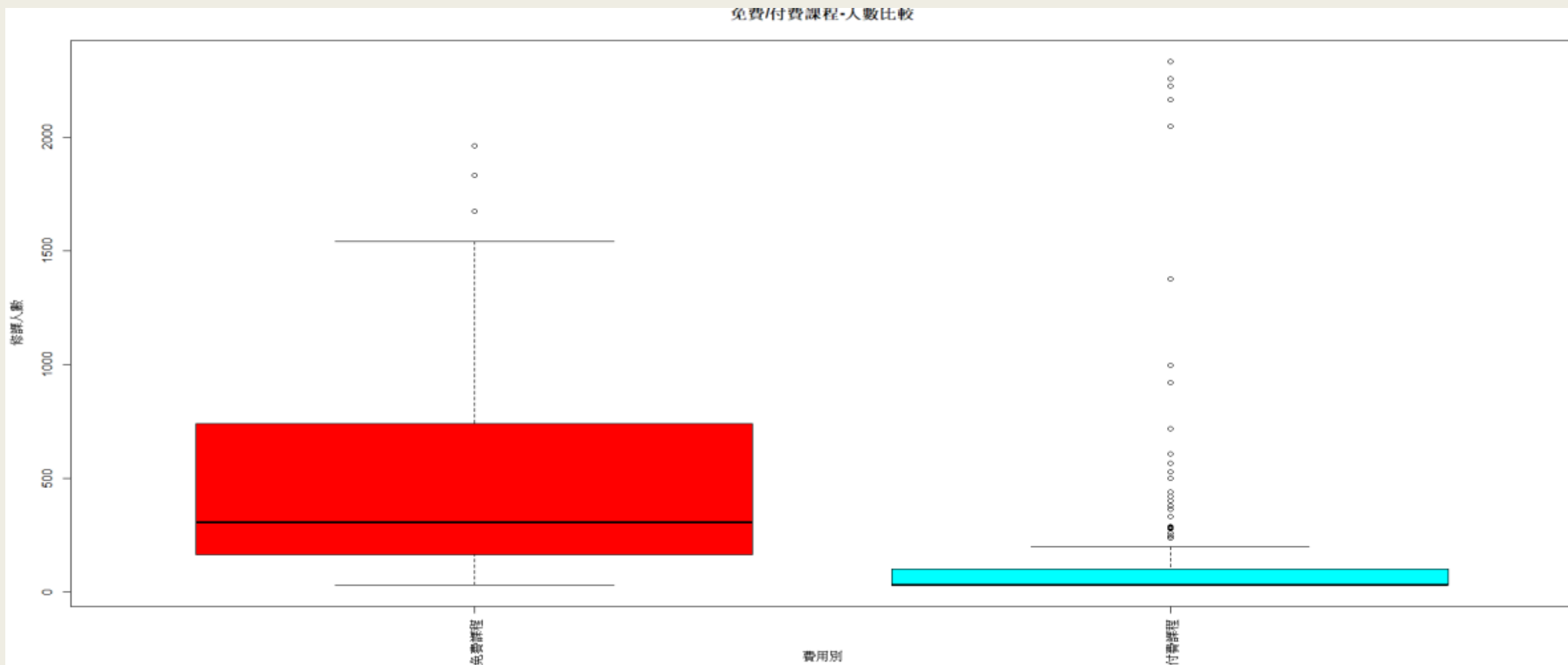
所有免費/付費課程人數-圖表
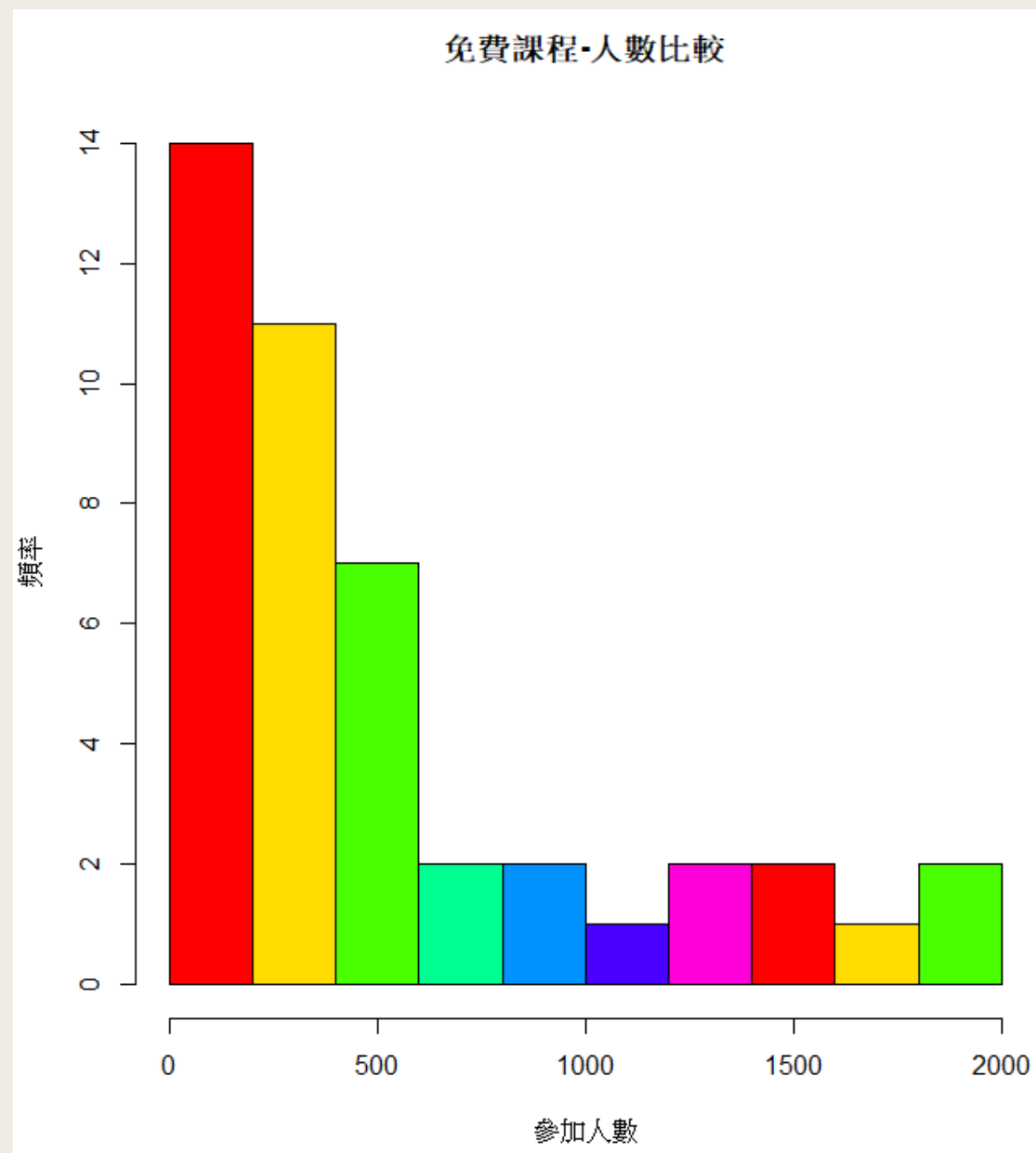
所有免費/付費課程人數-圖表
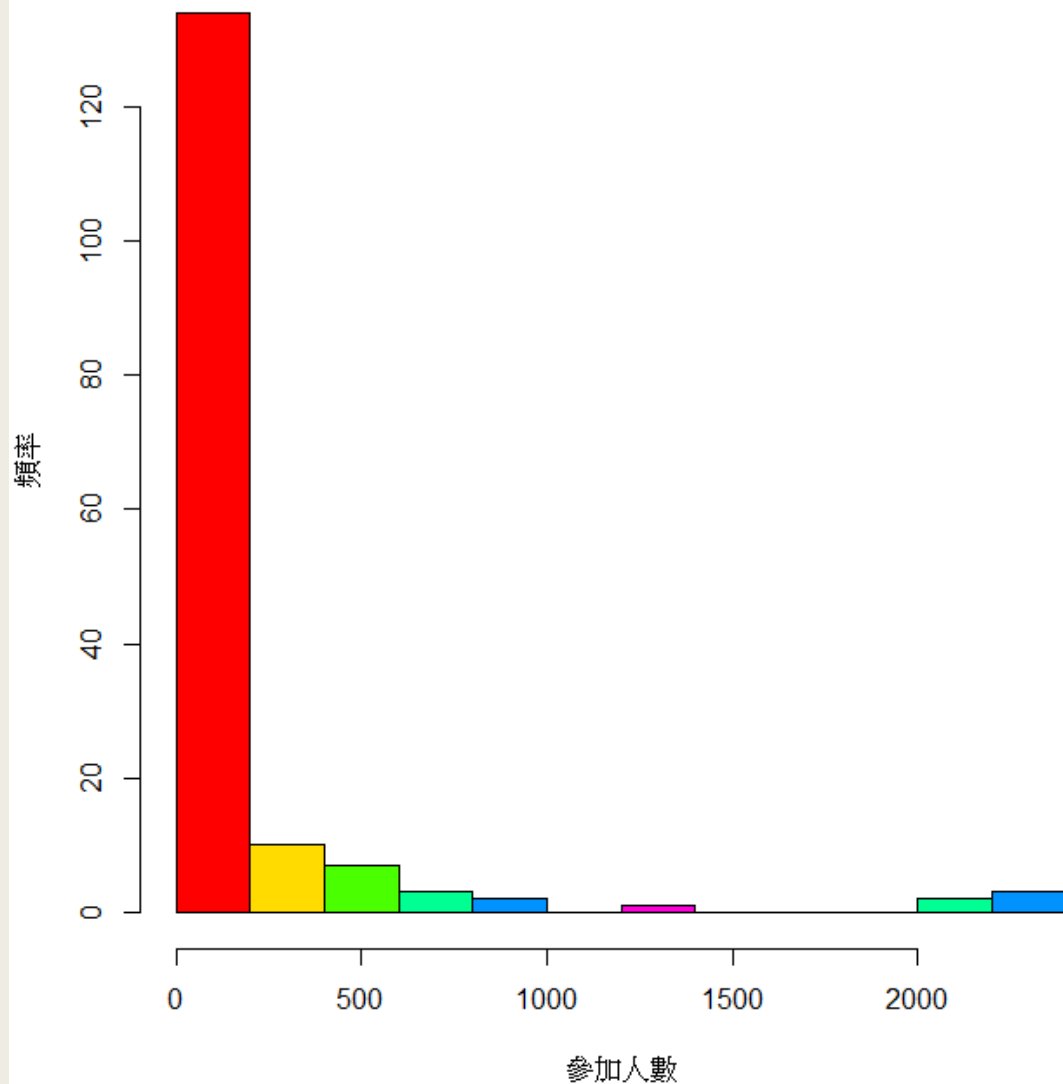
細部觀察
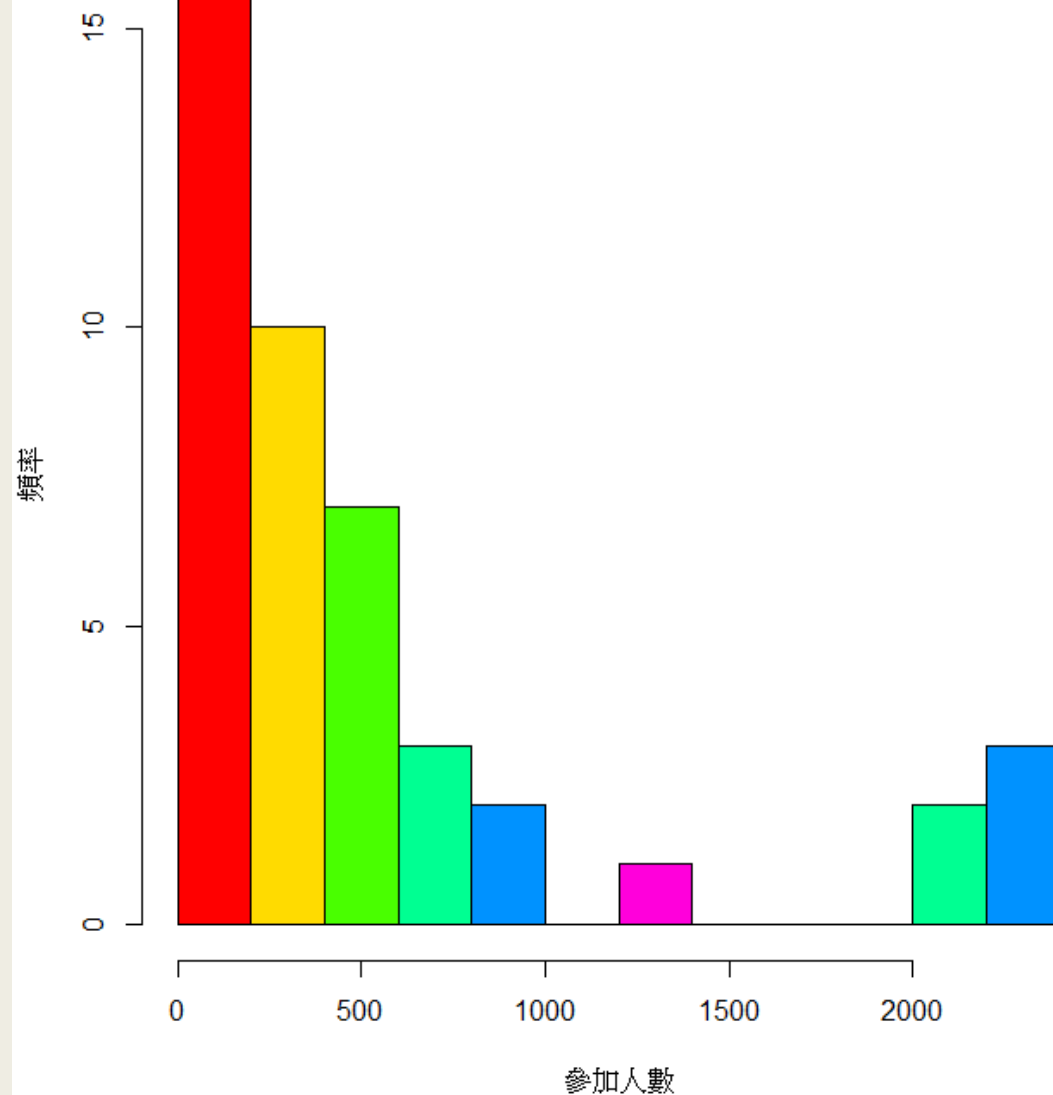
所有免費/付費課程總長度-資料清理

■ 將各類課程按照免費/付費分類

```
> free_ai_classlen <- ai_classlen[27:30]

> cost_ai_classlen <- ai_classlen[1:26]

> free_blockchain_classlen <- blockchain_classlen[5:21]

> cost_blockchain_classlen <- blockchain_classlen[1:4]

> free_e_commerce_classlen <- e_commerce_classlen[31:32]

> cost_e_commerce_classlen <- e_commerce_classlen[1:30]

> free_game_design_classlen <- game_design_classlen[34:35]

> cost_game_design_classlen <- game_design_classlen[1:33]

> free_digital_workforce_classlen <- digital_workforce_classlen[34:37]

> cost_digital_workforce_classlen <- digital_workforce_classlen[1:33]

> free_cloud_computing_classlen <- cloud_computing_classlen[16:20]

> cost_cloud_computing_classlen <- cloud_computing_classlen[1:15]

> free_mobile_apps_classlen <- mobile_apps_classlen[18:19]

> cost_mobile_apps_classlen <- mobile_apps_classlen[1:17]

> free_iot_classlen <- iot_classlen[5:12]

> cost_iot_classlen <- iot_classlen[1:4]
```

■ 清理完後存入變數

```
> free_all_classlen <- free_ai_classlen

> free_all_classlen <-
append(free_all_classlen ,free_blockchain_classlen)

> free_all_classlen <-
append(free_all_classlen ,free_e_commerce_classlen)

> free_all_classlen <-
append(free_all_classlen ,free_game_design_classlen)

> free_all_classlen <-
append(free_all_classlen ,free_digital_workforce_classlen)

> free_all_classlen <-
append(free_all_classlen ,free_cloud_computing_classlen)

> free_all_classlen <-
append(free_all_classlen ,free_mobile_apps_classlen)

> free_all_classlen <-
append(free_all_classlen ,free_iot_classlen)
```

所有免費/付費課程總長度-敘述統計

> 免費課程總長度 <- summary(free_all_classlen)

> 付費課程總長度 <- summary(cost_all_classlen)

> SUMMARY_all_classlen <-rbind(免費課程總長度,付費課程總長度)

> SUMMARY_all_classlen

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
免費課程總長度	0	48.75	61.5	104.5227	106.25	834
付費課程總長度	0	150.25	260.5	438.5494	473.50	5417

所有免費/付費課程總長度-圖表

> boxplot(free_all_classlen,cost_all_classlen,main="免費/付費課程-課程總長度比較", xlab="費用別",ylab="課程總長度",col=rainbow(2))

> axis(side=1, las=2, at = 1:2, labels=c("免費課程","付費課程"))
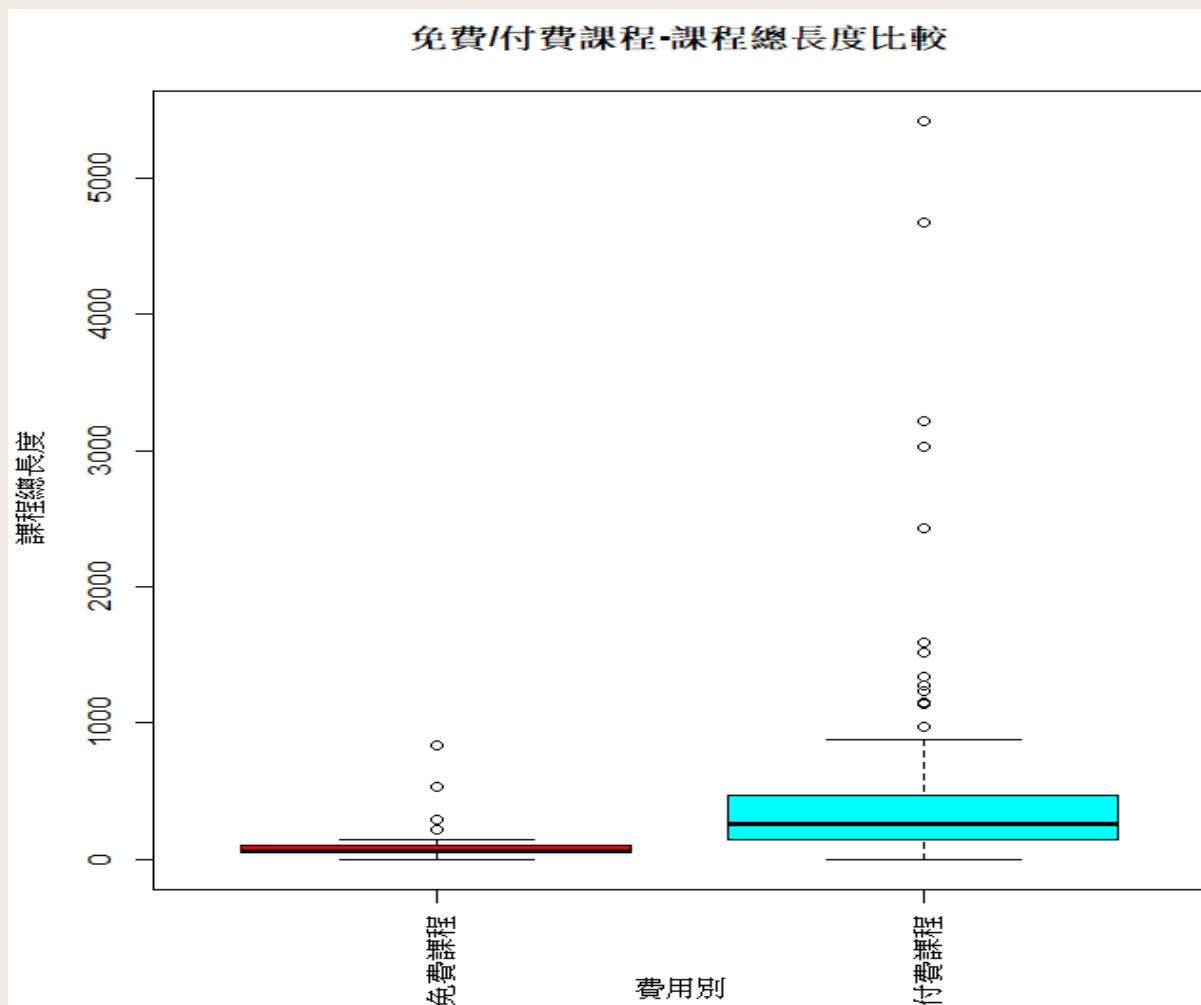
所有免費/付費課程總長度-圖表

細部觀察

各類課程人數分布-敘述統計

■ 資料清理

```
> ai_numofpeople <- as.numeric(gsub(",","",gsub("人","",gsub("以下","",ai_numofpeople))))

> blockchain_numofpeople <- as.numeric(gsub(",","",gsub("人","",gsub("以下","",blockchain_numofpeople))))

> e_commerce_numofpeople <- as.numeric(gsub(",","",gsub("人","",gsub("以下","",e_commerce_numofpeople))))

> game_design_numofpeople <- as.numeric(gsub(",","",gsub("人","",gsub("以下","",game_design_numofpeople))))

> digital_workforce_numofpeople <- as.numeric(gsub(",","",gsub("人","",gsub("以下","",digital_workforce_numofpeople))))

> cloud_computing_numofpeople <- as.numeric(gsub(",","",gsub("人","",gsub("以下","",cloud_computing_numofpeople))))

> mobile_apps_numofpeople <- as.numeric(gsub(",","",gsub("人","",gsub("以下","",mobile_apps_numofpeople))))

> iot_numofpeople <- as.numeric(gsub(",","",gsub("人","",gsub("以下","",iot_numofpeople))))
```
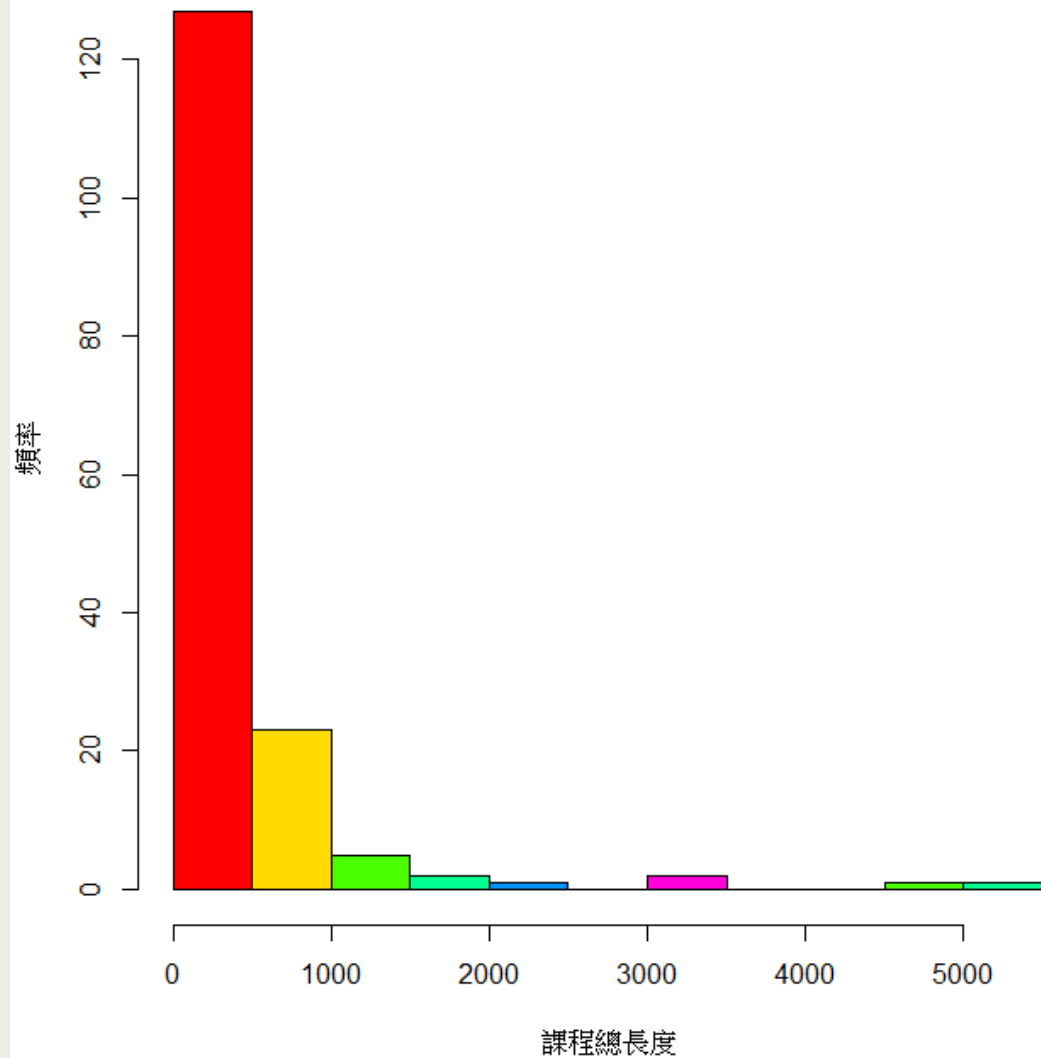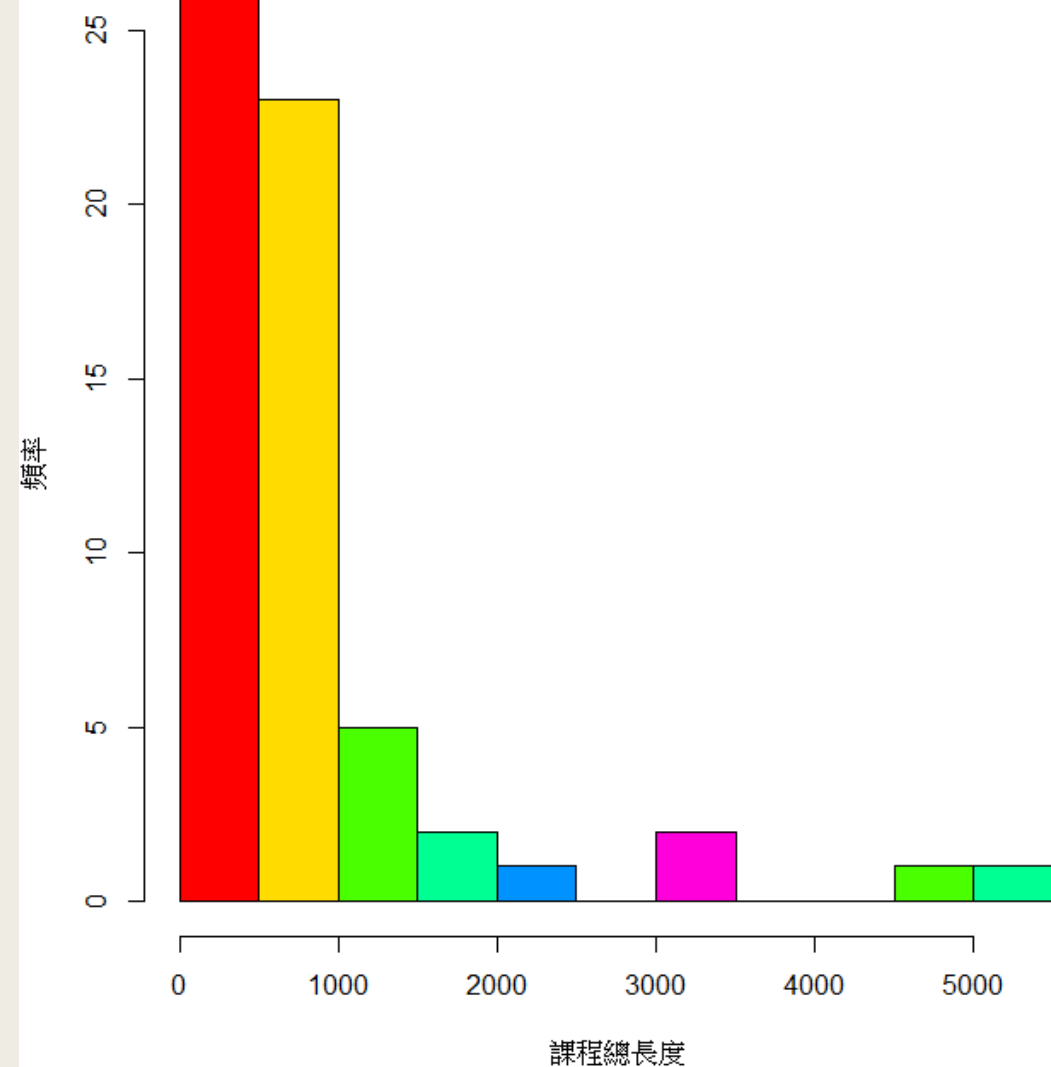
■ 敘述統計

```
>人工智慧 <- summary(ai_numofpeople)

> 區塊鏈 <- summary(blockchain_numofpeople)

> 數位商務 <- summary(e_commerce_numofpeople)

> 遊戲設計 <- summary(game_design_numofpeople)

> 科技管理 <- summary(digital_workforce_numofpeople)

> 雲端技術 <- summary(cloud_computing_numofpeople)

> 行動應用 <- summary(mobile_apps_numofpeople)

> 物聯網 <- summary(iot_numofpeople)

> SUMMARY_numofpeople <-rbind(人工智慧,區塊鏈,數位商務,遊戲設計,科技管理,雲端技術,行動應用,物聯網)
```

> SUMMARY_numofpeople

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
人工智慧	30	55.00	148.0	378.83333	497.50	1962
區塊鏈	30	97.00	150.0	354.61905	501.00	1832
數位商務	30	30.00	30.0	131.84375	50.75	1340
遊戲設計	30	30.00	30.0	75.08571	38.50	518
科技管理	30	30.00	33.0	250.59459	118.00	2165
雲端技術	30	30.00	152.0	285.45000	382.50	1376
行動應用	30	48.50	86.0	507.89474	633.00	2331
物聯網	30	165.25	237.5	218.50000	285.50	421

各類課程人數分布-圖表

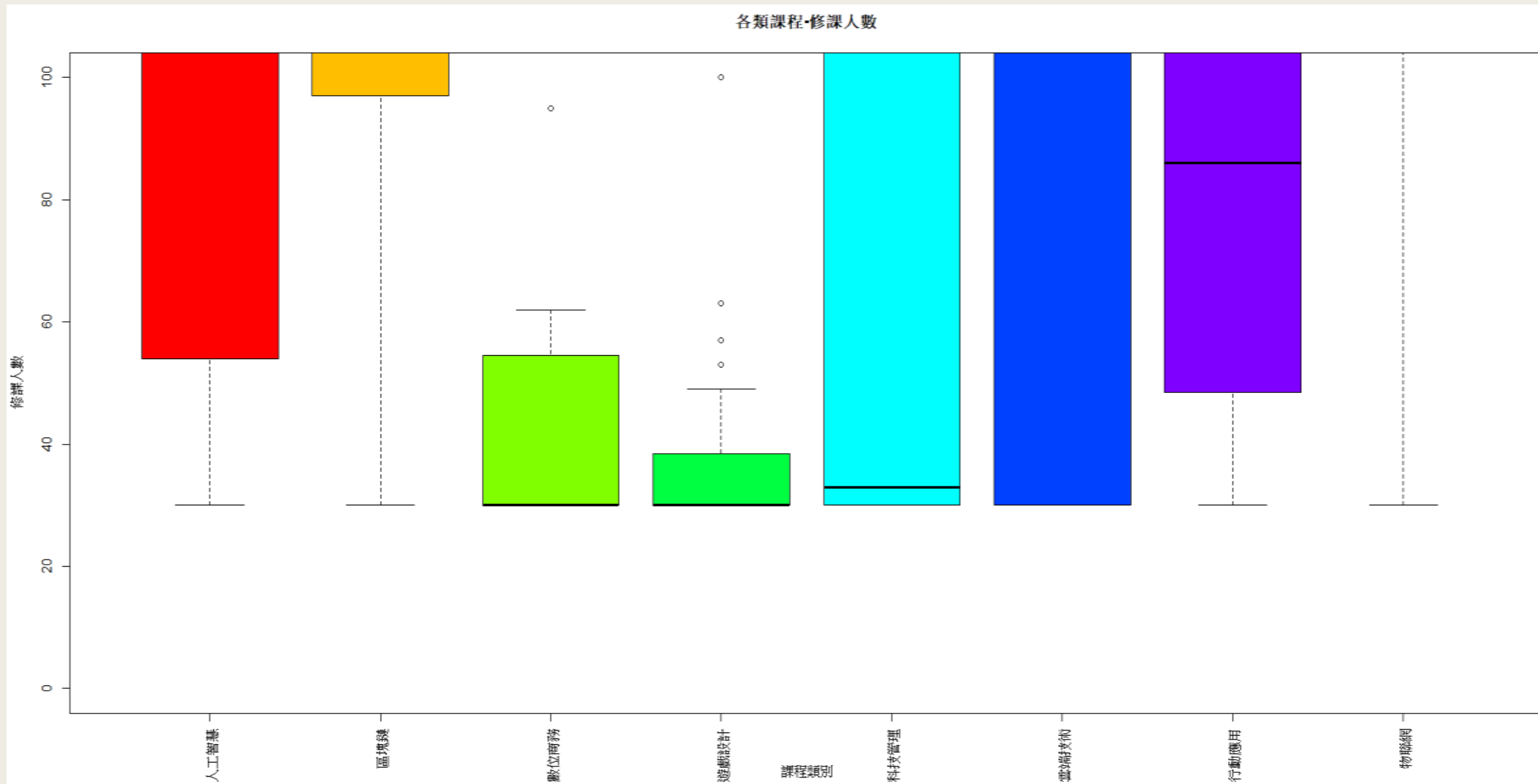>boxplot(ai_numofpeople,blockchain_numofpeople,e_commerce_numofpeople,game_design_numofpeople,digital_workforce_numofpeople,cloud_computing_numofpeople,mobile_apps_numofpeople,iot_numofpeople,main="各類課程-修課人數", xlab="課程類別",ylab="修課人數",col=rainbow(8))

> axis(side=1, las=2, at = 1:8, labels=c("人工智慧","區塊鏈","數位商務","遊戲設計","科技管理","雲端技術","行動應用","物聯網"))
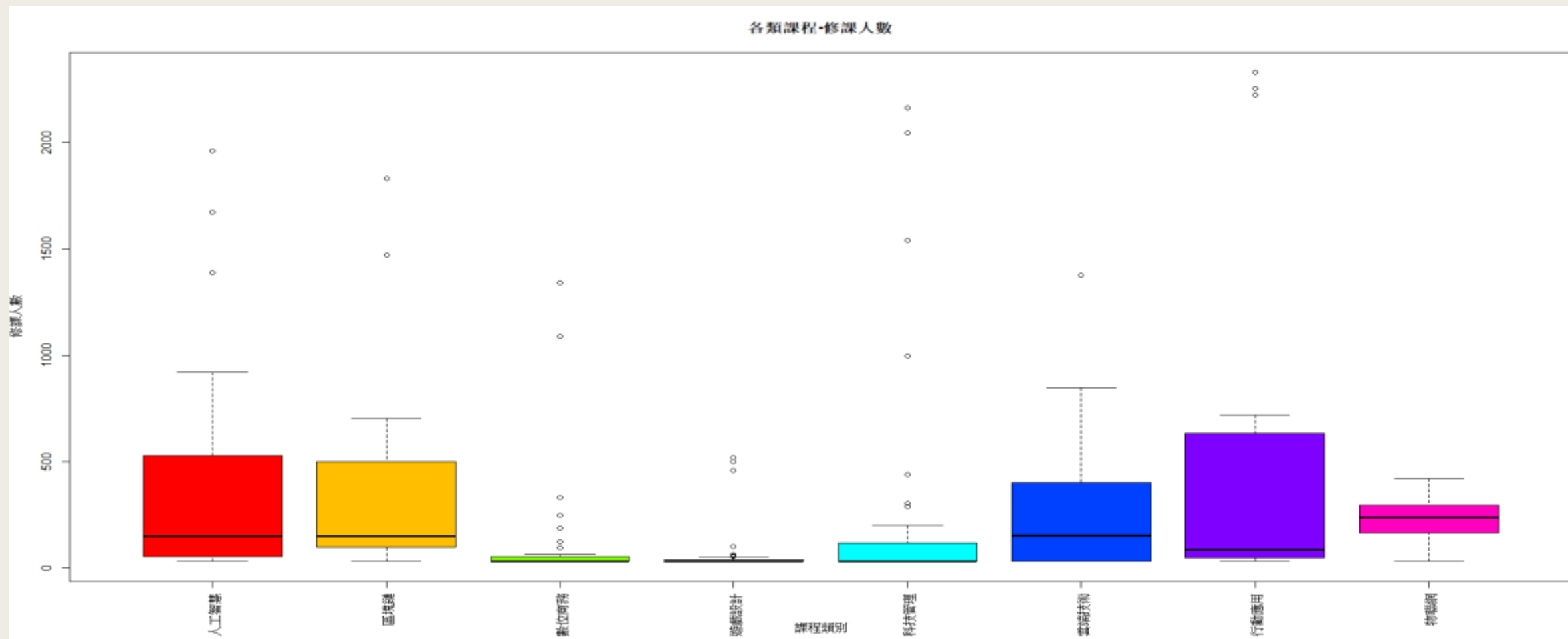
各類課程人數分布-圖表

各類課程價格分布-敘述統計

■ 資料清理

```
> ai_classprice <- as.numeric(gsub(",","",gsub("NT\\$","",gsub("免費","0",ai_classprice))))

> blockchain_classprice <- as.numeric(gsub(",","",gsub("NT\\$","",gsub("免費","0",blockchain_classprice))))

> e_commerce_classprice <- as.numeric(gsub(",","",gsub("NT\\$","",gsub("免費","0",e_commerce_classprice))))

> game_design_classprice <- as.numeric(gsub(",","",gsub("NT\\$","",gsub("免費","0",game_design_classprice))))

> digital_workforce_classprice <- as.numeric(gsub(",","",gsub("NT\\$","",gsub("免費","0",digital_workforce_classprice))))

> cloud_computing_classprice <- as.numeric(gsub(",","",gsub("NT\\$","",gsub("免費","0",cloud_computing_classprice))))

> mobile_apps_classprice <- as.numeric(gsub(",","",gsub("NT\\$","",gsub("免費","0",mobile_apps_classprice))))

> iot_classprice <- as.numeric(gsub(",","",gsub("NT\\$","",gsub("免費","0",iot_classprice))))
```

■ 敘述統計

```
> 人工智慧 <- summary(ai_classprice)

> 區塊鏈 <- summary(blockchain_classprice)

> 數位商務 <- summary(e_commerce_classprice)

> 遊戲設計 <- summary(game_design_classprice)

> 科技管理 <- summary(digital_workforce_classprice)

> 雲端技術 <- summary(cloud_computing_classprice)

> 行動應用 <- summary(mobile_apps_classprice)

> 物聯網 <- summary(iot_classprice)

> SUMMARY_classprice <-rbind(人工智慧,區塊鏈,數位商務,遊戲設計,科技管理,雲端技術,行動應用,物聯網)
```

> SUMMARY_classprice

|	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
人工智慧	0	650.0	1050	1860.9000	1975.0	12000
區塊鏈	0	0.0	0	593.3333	0.0	8000
數位商務	0	949.0	2019	2102.8438	2605.0	6800
遊戲設計	0	979.5	1780	1599.8000	2000.0	4990
科技管理	0	825.0	1100	1280.9189	1500.0	3990
雲端技術	0	255.0	995	1755.0000	2295.0	10000
行動應用	0	545.0	1200	1424.6316	2099.5	3900
物聯網	0	0.0	0	499.1667	917.5	2300

各類課程價格分布-圖表

>boxplot(ai_classprice,blockchain_classprice,e_commerce_classprice,game_design_classprice,digital_workforce_classprice,cloud_computing_classprice,mobile_apps_classprice,iot_classprice,main="各類課程-課程價格", xlab="課程類別",ylab="課程價格",col=rainbow(8))

> axis(side=1, las=2, at = 1:8, labels=c("人工智慧","區塊鏈","數位商務","遊戲設計","科技管理","雲端技術","行動應用","物聯網"))
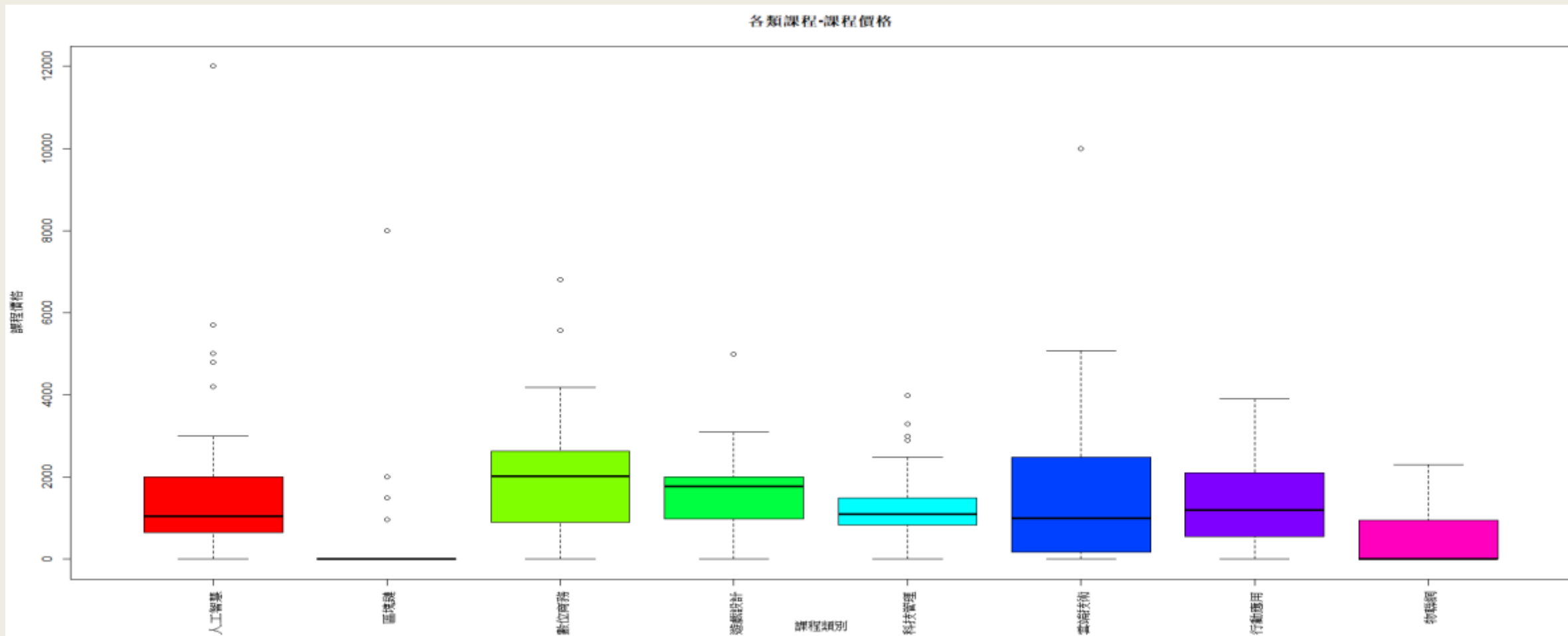
各類課程價格分布-圖表
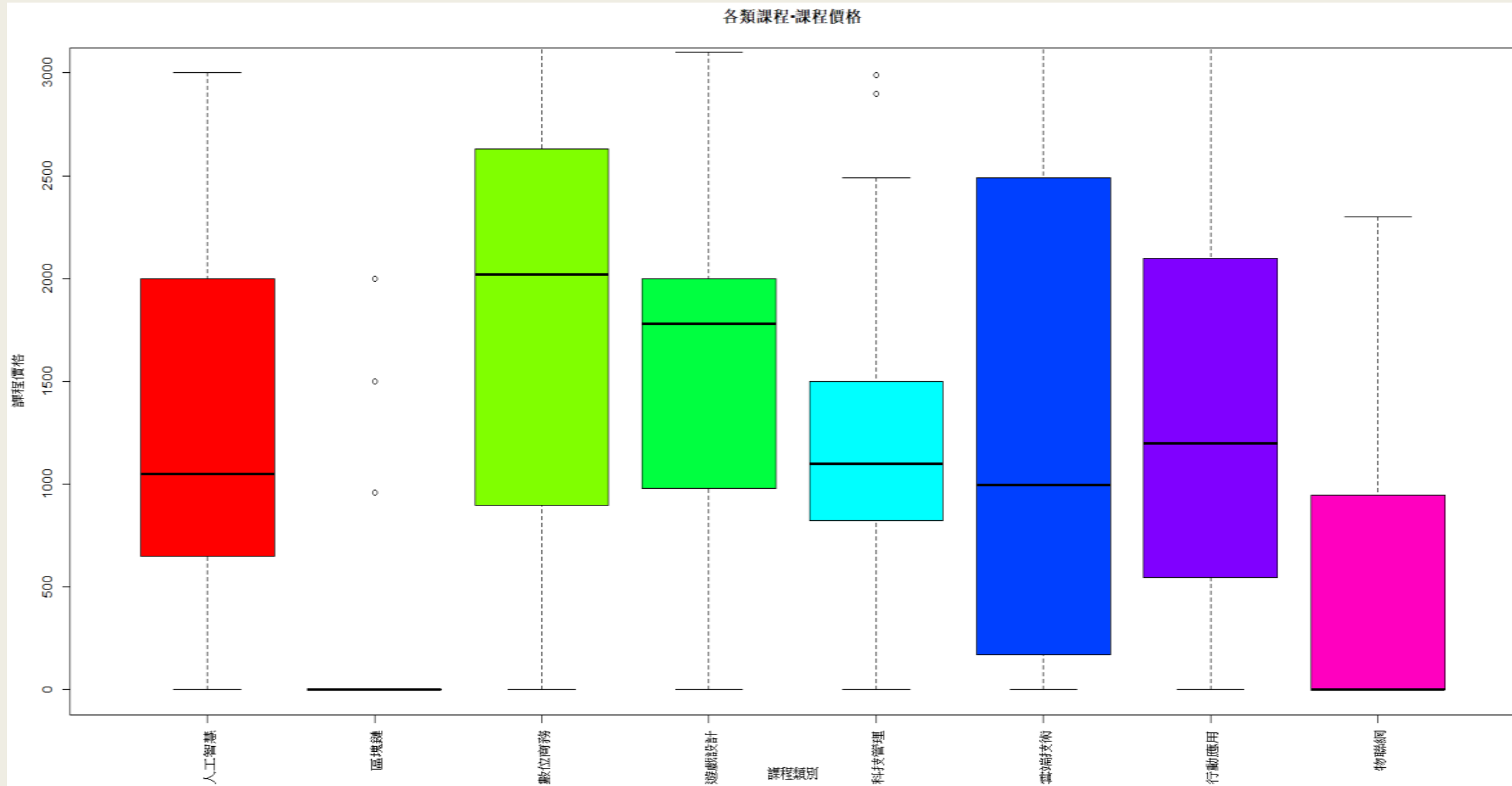
各類課程總長度分布-敘述統計

```
> for(i in 1:length(ai_classlen)){ai_classlen[i] <- eval(parse(text = gsub("h","*60+",gsub("m","",ai_classlen[i]))))}

> for(i in 1:length(blockchain_classlen)){blockchain_classlen[i] <- eval(parse(text = gsub("h","*60+",gsub("m","",blockchain_classlen[i]))))}

> for(i in 1:length(e_commerce_classlen)){e_commerce_classlen[i] <- eval(parse(text = gsub("h","*60+",gsub("m","",e_commerce_classlen[i]))))}

> for(i in 1:length(game_design_classlen)){game_design_classlen[i] <- eval(parse(text = gsub("h","*60+",gsub("m","",game_design_classlen[i]))))}

> for(i in 1:length(digital_workforce_classlen)){digital_workforce_classlen[i] <- eval(parse(text = gsub("h","*60+",gsub("m","",digital_workforce_classlen[i]))))}

> for(i in 1:length(cloud_computing_classlen)){cloud_computing_classlen[i] <- eval(parse(text = gsub("h","*60+",gsub("m","",cloud_computing_classlen[i]))))}

> for(i in 1:length(mobile_apps_classlen)){mobile_apps_classlen[i] <- eval(parse(text = gsub("h","*60+",gsub("m","",mobile_apps_classlen[i]))))}

> for(i in 1:length(iot_classlen)){iot_classlen[i] <- eval(parse(text = gsub("h","*60+",gsub("m","",iot_classlen[i]))))}
```

■ 敘述統計

```
> 人工智慧 <- summary(ai_classlen)

> 區塊鏈 <- summary(blockchain_classlen)

> 數位商務 <- summary(e_commerce_classlen)

> 遊戲設計 <- summary(game_design_classlen)

> 科技管理 <- summary(digital_workforce_classlen)

> 雲端技術 <- summary(cloud_computing_classlen)

> 行動應用 <- summary(mobile_apps_classlen)

> 物聯網 <- summary(iot_classlen)

> SUMMARY_classlen <-rbind(人工智慧,區塊鏈,數位商務,遊戲設計,科技管理,雲端技術,行動應用,物聯網)
```

> SUMMARY_classlen

|	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
人工智慧	73	161.50	262.0	917.9667	1079.25	5417
區塊鏈	22	50.00	56.0	100.7143	109.00	497
數位商務	58	150.75	247.0	272.1250	376.25	972
遊戲設計	28	116.50	270.0	335.3143	375.00	1593
科技管理	0	88.00	179.0	258.0270	324.00	1519
雲端技術	48	110.00	195.5	375.2000	522.50	2427
行動應用	27	79.00	266.0	331.5789	536.00	726
物聯網	34	79.25	100.5	182.9167	253.25	537

各類課程總長度分布-圖表

>boxplot(ai_classlen,blockchain_classlen,e_commerce_classlen,game_design_classlen,digital_workforce_classlen,cloud_computing_classlen,mobile_apps_classlen,iot_classlen,main="各類課程-課程總長度", xlab="課程類別",ylab="課程總長度",col=rainbow(8))

> axis(side=1, las=2, at = 1:8, labels=c("人工智慧","區塊鏈","數位商務","遊戲設計","科技管理","雲端技術","行動應用","物聯網"))

各類課程總長度分布-圖表

各類課程-課程總長度
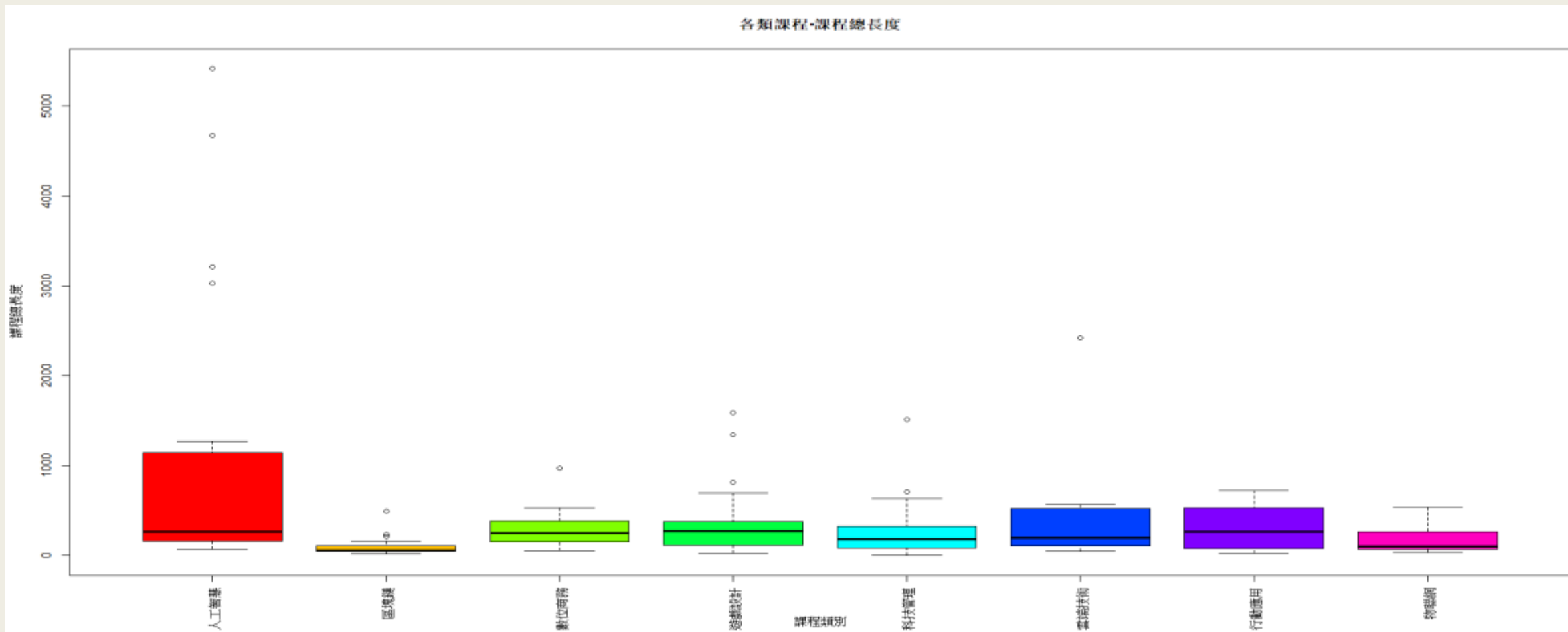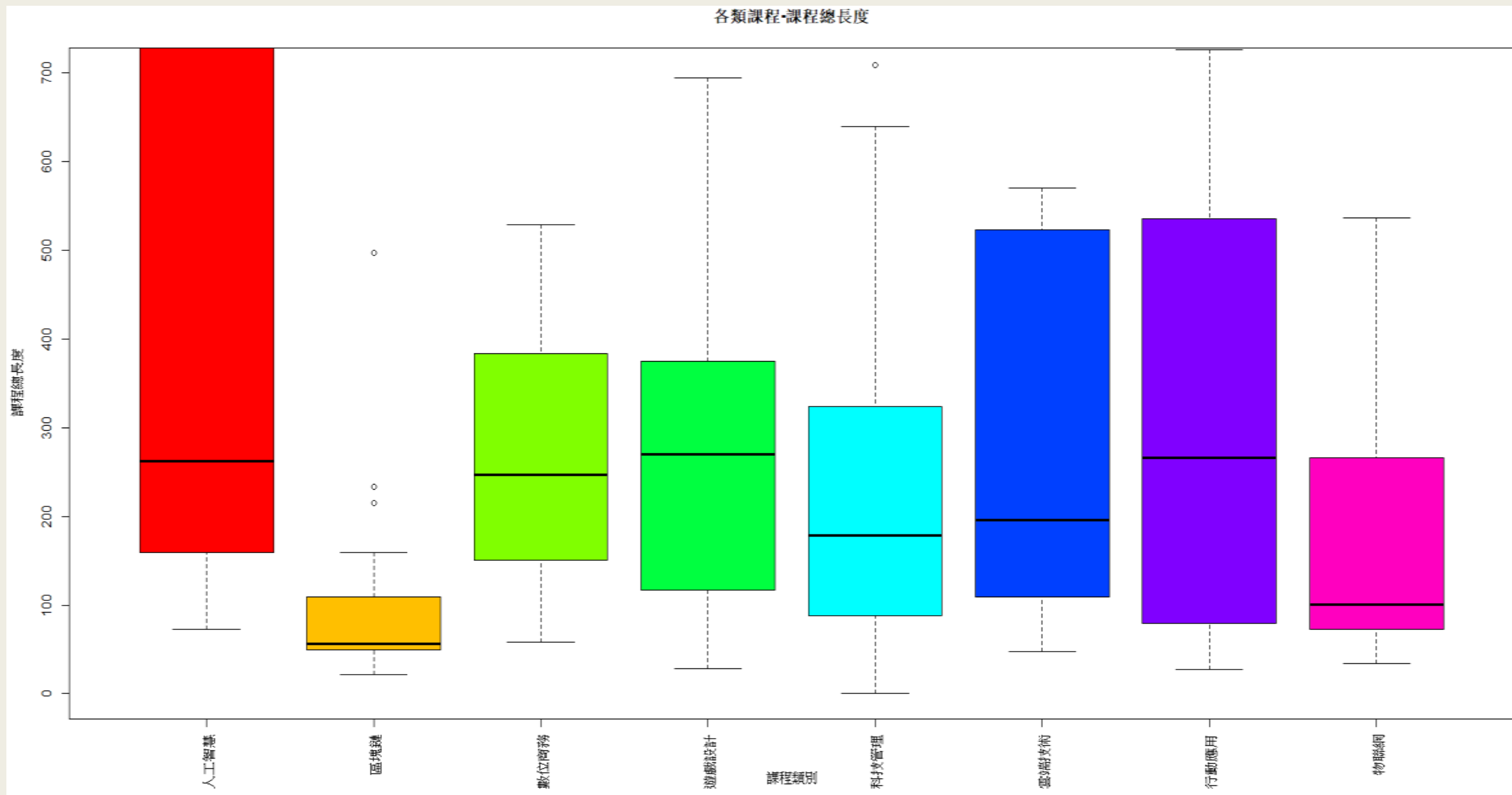
PART4、結論

結論

一、免費比付費課程平均多了三倍人數，但兩者的最大值是付費較高，有2200多人修課

二、行動應用與科技管理是兩個最熱門類別，價位在1200~1400算是中等

三、遊戲設計有超過75%的課不到40人修課，是最冷門類別

四、區塊鏈有超過八成是免費課，少部分課價錢較低，可能剛推出這類別，講師們還在試水溫

五、人工智慧課程總長度是所有類當中最長，平均將近15小時，可能範圍較廣需要較多時間來學會

THANK YOU