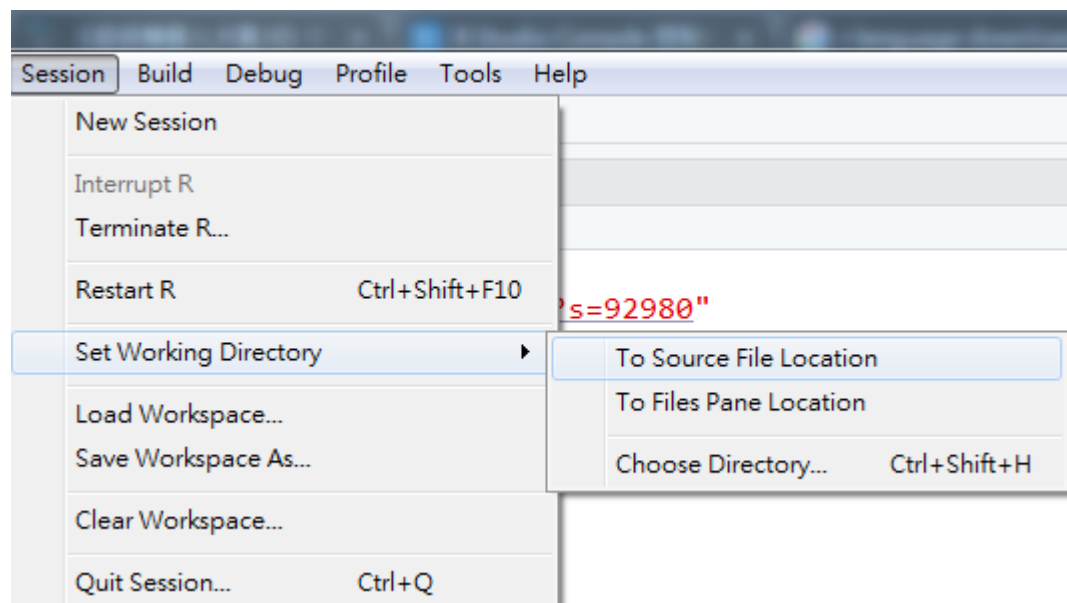


# Chapter 12

---

WRAPPER WEBSITE USING R

# 設定工作目錄



```
> setwd("F:/Course/1062/data mining/code/R/ch12")  
> |
```

# 網頁

https://acg.gamer.com.tw/acgDetail.php?s=92980



PS4

作品資料

圖片影片

攻略百科

公會社團

玩家心得

玩家名單



哈啦板



### 超級機器人大戰 X

スーパーロボット大戦X  
Super Robot Wars X

- ▶ 主機平台：PS4
- ▶ 遊戲類型：策略模擬
- ▶ 發售日期：2018-03-29
- ▶ 遊戲售價：¥ 8600
- ▶ 遊戲人數：1人
- ▶ 作品分級：CERO C
- ▶ 製作廠商：B.B.STUDIO
- ▶ 發行廠商：BANDAI NAMCO Entertainment
- ▶ 代理廠商：台灣萬代南夢宮娛樂
- ▶ 官方網站：<http://srwx.suparobo.jp/>



編輯 

 想玩 | PS4

572

#### 網友評價



0人

輕鬆休閒	0%
重視對戰	0%
重視思考	0%
緊張感	0%
多人同樂	0%

我的評價 ▶     

# 網頁原始碼

<https://acg.gamer.com.tw/acgDetail.php?s=92980>

```
<html xmlns="http://www.w3.org/1999/xhtml" xmlns:og="http://ogp.me/ns#">
  <head>...</head>
  <body> == $0
    <!-- Google Tag Manager -->
    <noscript>...</noscript>
    <script>...</script>
    <!-- End Google Tag Manager -->
    <div id="fb-root" class=" fb_reset">...</div>
    <script>...</script>
    <div class="TOP-bh">...</div>
    <!--全站內頁天結束-->
    <div class="bh-banner">...</div>
    <div id="BH-background">
      <div id="BH-wrapper" style="margin-top: 0px;">
        <!--最外層寬1000包裝-->
        <!--全站主選單-->
        <div id="BH-menu-fill">&nbsp;</div>
        <div id="BH-menu-path" class="BH-menu">...</div>
        <!--全站主選單結束-->
        <div style="width:0;height:0;position:relative;top:0px;left:0px;z-index:100;">...</div>
        <!--搜尋BAR下拉結束-->
        <!--服務選單-->
        <div id="BH-master">
          <!--內容左側區塊-->
          <script language="javascript">...</script>
          <div class="BH-lbox ACG-mster_box1 hreview-aggregate hreview">...</div>
          <!--區塊1結束-->
          <h5></h5>
          <!--圖片影片區塊開始-->
          <h4>...</h4>
          <div class="BH-lbox ACG-list2">...</div>
          <!--圖片影片區塊結束-->
          <h4>...</h4>
          <div class="BH-lbox ACG-mster_box6">...</div>
          <!--區塊6結束-->
        </div>
      </div>
    </div>
  </body>
</html>
```

# 要抓的網頁部份內容

<https://acg.gamer.com.tw/acgDetail.php?s=92980>

div.BH-lbox.ACG-mster\_box1.hreview-aggregate.hreview | 690 × 303

心得 | 玩家名單



## 超級機器人大戰 X スーパーロボット大戦X Super Robot Wars X

- ▶ 主機平台：PS4
- ▶ 遊戲類型：策略模擬
- ▶ 發售日期：2018-03-29
- ▶ 遊戲售價：¥ 8600
- ▶ 遊戲人數：1人
- ▶ 作品分級：CERO C
- ▶ 製作廠商：B.B.STUDIO
- ▶ 發行廠商：BANDAI NAMCO Entertainment
- ▶ 代理廠商：台灣萬代南夢宮娛樂
- ▶ 官方網站：<http://srwx.suparobo.jp/>



編輯

# 原始碼解析

<https://acg.gamer.com.tw/acgDetail.php?s=92980>

```
<div class="BH-lbox ACG-mster_box1 hreview-aggregate hreview"> == $0
  <!--區塊1開始-->
  <div class="ACG-box1-left">
    <div id="ACG-box1pic">
      <span></span>
      
    </div>
  </div>
  <h1>超級機器人大戰 X</h1>
  <h2>スーパーロボット大戦X</h2>
  <h2>Super Robot Wars X</h2>
  <ul class="ACG-box1listA">
    <li>...</li>
    <li>...</li>
    <li>發售日期：2018-03-29</li>
    <li>遊戲售價：¥ 8600</li>
    <li>遊戲人數：1人</li>
    <li>作品分級：CERO C</li>
  </ul>
  <ul class="ACG-box1listB">
    <li>
      製作廠商：
      <a href="//acg.gamer.com.tw/search.php?st=1&kw=B.B.STUDIO" target="_blank">B.B.STUDIO</a>
    </li>
    <li>
      發行廠商：
      <a href="//acg.gamer.com.tw/search.php?st=1&kw=BANDAI+NAMCO+Entertainment" target="_blank">BANDAI NAMCO Entertainment</a>
    </li>
    <li>...</li>
    <li>...</li>
  </ul>
  <p id="ACG-box1mark">
    <a href="createEdit.php?s=92980">...</a>
    
    
  </p>
  <p class="clearfloat">
  </p>
</div>
```

# 讀取網頁資料

```
1 #rm(list=ls())
2 library(rvest)
3 url <- "https://acg.gamer.com.tw/acgDetail.php?s=92980"
4 srw <- read_html(url)
5 print(srw)
6
```

```
{xml_document}
<html xmlns="http://www.w3.org/1999/xhtml" xmlns:og="http://ogp.me/ns#">
[1] <head>\n<meta name="viewport" content="width=1250">\n<meta http-equiv="Content-Type" content="text/html; ch ...
[2] <body>\n<!-- Google Tag Manager -->\n<noscript><iframe src="//www.googletagmanager.com/ns.html?id=GTM-KDKMG ...
> |
```

Install Packages

Install from: [? Configuring Repositories](#)

Repository (CRAN, CRANextra) ▼

Packages (separate multiple with space or comma):

rvest

**rvest** o Library:

# 讀取網頁資料

```
1 #rm(list=ls())
2 library(rvest)
3 url <- "https://acg.gamer.com.tw/acgDetail.php?s=92980"
4 srw <- read_html(url)
5
6 img <- srw %>% html_nodes(css = "img") %>%
7   html_attr("src")
8 img[1]
9
10 ul <- srw %>% html_nodes(css = "ul" ) %>%
11   html_text()
12 ul
13
14 h1 <- srw %>% html_nodes(css = "h1" ) %>%
15   html_text()
16 h1
17
18 h2 <- srw %>% html_nodes(css = "h2" ) %>%
19   html_text()
20 h2
```

```
> img[1]
[1] "https://p2.bahamut.com.tw/B/ACG/c/80/0000092980.JPG"
```

```
> ul
```

```
[1] "主機平台：PS4\n遊戲類型：策略模擬\n發售日期：2018-03-29\n遊戲售價
```

```
[2] "製作廠商：B.B.STUDIO\n發行廠商：BANDAI NAMCO Entertainment\n代  
bo.jp/\n"
```

```
[3] "2018-03-11 《超級機器人大戰 X》釋出第一話「旅程的序章」13 分鐘遊  
機體 釋出更多參戰機體戰鬥畫面2018-02-20 《超級機器人大戰 X》釋出第 2 波  
《超級機器人大戰 X》公布「額外動作」「額外命令」概要 以及舞台背景設定等情
```

```
[4] "輕鬆休閒重視對戰重視思考緊張感多人同樂"
```

```
[5] "0%\n0%\n0%\n0%\n0%\n"
```

```
[6] "Another\n社團成員： 624 人\n話題數：1112\n昨日人氣：5\n招募方式：
```

```
[7] "スーパーロボット大戦\n社團成員： 136 人\n話題數：54\n昨日人氣：5\n
```

```
[8] "PC線上遊戲\n暗黒破壊神 3\n收藏： 21822 人\n評價： 7 分\n"
```

```
[9] "Web遊戲\n信喵之野望\n收藏： 76429 人\n評價： 8.6 分\n"
```

```
[10] "PS3\n第 2 次超級機器人大戰 OG\n收藏： 1671 人\n評價： 9.4 分\n"
```

```
> h1
```

```
[1] "超級機器人大戰 X"
```

```
>
```

```
> h2 <- srw %>% html_nodes(css = "h2" ) %>%
+   html_text()
```

```
> h2
```

```
[1] "スーパーロボット大戦X" "Super Robot Wars X"
```



# 讀取網頁資料

```
1 library(rvest)
2 library(jpeg)
3
4 url <- "https://acg.gamer.com.tw/acgDetail.php?s=92980"
5 srw <- read_html(url)
6
7 get_img <- srw %>% html_nodes(css = "img") %>%
8   html_attr("src")
9
10 download.file(get_img[1], "srw.jpg", mode = 'wb')
11
12 img <- readJPEG("srw.jpg", native=TRUE)
13 plot(NA, xlim=c(0, nrow(img)), ylim=c(0, nrow(img)))
14 rasterImage(img, 0, 0, ncol(img), nrow(img))
15
```

Install Packages

Install from: [? Configuring Repositories](#)

Repository (CRAN, CRANextra) ▼

Packages (separate multiple with space or comma):

jpeg

jpeg to Library:



# 讀取網頁資料

```
16 ul      <- "ul"
17 h1      <- "h1"
18 h2      <- "h2"
19 para    <- c(u1,h1,h2)
20 res     <-list()
21
22 for (i in 1:length(para))
23 {
24   tmp <- srw %>% html_nodes(css = para[i]) %>%
25               html_text()
26   res[[i]] <- tmp
27 }
28 res
```

[9] "Web遊戲\n信喵之野望\n收藏： 76429 人\n評價： 8.6 分\n"

[10] "PS3\n第 2 次超級機器人大戰 OG\n收藏： 1671 人\n評價： 9.4 分\n"

[[2]]

[1] "超級機器人大戰 X"

[[3]]

[1] "スーパーロボット大戦X" "Super Robot Wars X"

# 整理網頁資料

```
29 names(res) <- c("detail","name1","name2")
30 res$detail <- strsplit(res$detail,"\n")
31
32 a <- paste(res$name1,res$name2[1])
33 b <- paste(res$detail[[1]][1:4])
34
35 print(a)
36 print(b)
```

```
> print(a)
```

```
[1] "超級機器人大戰 X スーパーロボット大戦X"
```

```
> print(b)
```

```
[1] "主機平台：PS4"          "遊戲類型：策略模擬"    "發售日期：2018-03-29" "遊戲售價：¥ 8600"
```

# 隨堂練習 1

---

1. 抓出小叮噠電影版：南極大冒險的 wiki 資料

<https://ppt.cc/fDvZpx>

2. 秀出

1. 片名
2. 電影播放時間
3. 所有配音名單 (ex, ドラえもん - 水田わさび)

Any Questions !?