MOVIE EXTRACTION

Kent Hervey D. Gener BSIT2-B

2024-02-06

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

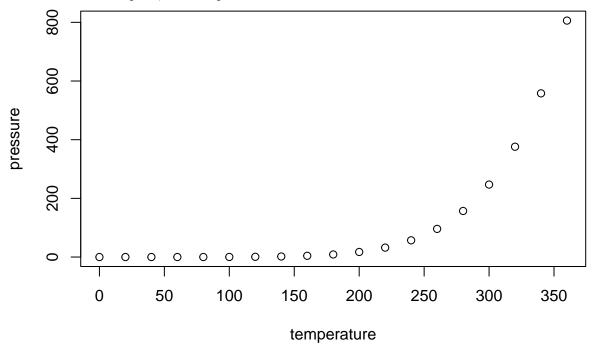
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

```
##
        speed
                         dist
##
    Min.
           : 4.0
                    Min.
                            : 2.00
##
    1st Qu.:12.0
                    1st Qu.: 26.00
##
    Median:15.0
                    Median: 36.00
            :15.4
                            : 42.98
##
    Mean
                    Mean
##
    3rd Qu.:19.0
                    3rd Qu.: 56.00
##
    Max.
            :25.0
                    Max.
                            :120.00
```

Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
library(rvest)
library(dplyr)
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
       filter, lag
##
## The following objects are masked from 'package:base':
##
       intersect, setdiff, setequal, union
##
scrape_reviews <- function(url) {</pre>
  page <- read_html(url)</pre>
  user <- page %>% html_nodes(".display-name-link") %>% html_text()
  date <- page %>% html_nodes(".review-date") %>% html_text()
  rating <- page %>% html_nodes(".rating-other-user-rating") %>% html_text()
  comment_title <- page %>% html_nodes(".title") %>% html_text()
  comment <- page %>% html_nodes(".text.show-more__control") %>% html_text()
  reviews <- data.frame(
    User = head(user, 300),
    Date = head(date, 300),
    Rating = head(rating, 300),
    Title = head(comment title, 300),
    Comment = head(comment, 300)
 return(reviews)
}
scrape_multiple_pages <- function(base_url) {</pre>
  all_reviews <- data.frame()</pre>
  reviews_per_page <- 10
  total_reviews_target <- 300</pre>
  num_pages <- ceiling(total_reviews_target / reviews_per_page)</pre>
  for (page_num in 1:num_pages) {
    url <- pasteO(base_url, "&start=", (page_num - 1) * reviews_per_page)</pre>
    reviews <- scrape_reviews(url)</pre>
    all_reviews <- bind_rows(all_reviews, reviews)</pre>
    if (nrow(all_reviews) >= total_reviews_target) {
      break
    }
  }
 return(all_reviews)
}
```

```
all_reviews <- scrape_multiple_pages(imdb_url)</pre>
print(head(all_reviews))
##
                             Date
             User
                                                             Rating
                                                 5/10\n
## 1 DocteurDream 30 January 2024 \n
         kjproulx 12 January 2024 \n
                                                 7/10\n
## 3 BA_Harrison 12 January 2024 n
                                                 7/10\n
       FeastMode 12 January 2024 \n
                                                 3/10\n
## 4
## 5 sbweightman 12 January 2024 \n
                                                 8/10\n
      Neptune165 2 February 2024 \n
## 6
                                                 3/10\n
                                                                                    Title
##
\#\# 1 Verona Parker is easily one of the worst characters ever written for the screen\n
## 2
                                                           A Cheesy, Cool Action Flick\n
## 3
                                                               Provides a decent buzz.\n
## 4
                                             Awesome fight scenes, bad everything else\n
## 5
                                                                               A keeper\n
## 6
                                    Horrible Movie. Bad acting and even worse writing\n
##
```

2 By now, audiences should realize what they are in for when it comes to Jason Statham leading an ac

imdb_url <- "https://www.imdb.com/title/tt15314262/reviews/?ref_=tt_ov_rt"</pre>

1

3 ## 4 ## 5 ## 6