

ps2

Kehsin Su Esther 3033114294

September 15, 2017

```
knitr::opts_chunk$set(tidy = TRUE, cache = TRUE)
```

(a) save letters in text format:

randomly come out 1e6 samples of letters from a to z and these sample are picked without replacement. write these samples into tmp1.csv view the information of tmp1.csv As each letter is one byte and there will be one byte each ", so totally "a" will contains three bytes. The tmp1.csv file contains 1e6 items, so it will be 3,000,000 bytes.

```
library(pryr)

## Warning: package 'pryr' was built under R version 3.3.3

library(stringr)

## Warning: package 'stringr' was built under R version 3.3.3

chars <- sample(letters, 1e+06, replace = TRUE)
write.table(chars, file = "tmp1.csv", row.names = FALSE, quote = FALSE, col.names = FALSE)
system("ls -l tmp1.csv", intern = TRUE)

## [1] "-rw-r--r-- 1 Esther 197609 3000000 Sep 15 12:32 tmp1.csv"
```

combines samples without delimiters write the combined string into a tmp2.csv view the information of tmp2.csv(the size has changed) the totally byte will become 1e6 chars add one byte for each ', so it is 1e6+2 bytes

```
chars <- paste(chars, collapse = "")
write.table(chars, file = "tmp2.csv", row.names = FALSE, quote = FALSE, col.names = FALSE)
system("ls -l tmp2.csv", intern = TRUE)

## [1] "-rw-r--r-- 1 Esther 197609 1000002 Sep 15 12:32 tmp2.csv"
```

save in binary format: simulate 1e6 numeric data and save them into binary data, each is 8bytes and due to system setting, totally will nearly 8*1e6

```

nums <- rnorm(1e+06)
save(nums, file = "tmp3.Rda")
system("ls -l tmp3.Rda", intern = TRUE)

## [1] "-rw-r--r-- 1 Esther 197609 7677868 Sep 15 12:32 tmp3.Rda"

```

save in text format: it will cost more space when comparing to binary data type the longest length of the item is those negative items with length 18 and consider the delimiter ',', which totally has 19bytes, so roughly, tmp4 will have 19×10^6 and consider some of them are positive, it will be about 18×10^6 to 19×10^6 bytes

```

write.table(nums, file = "tmp4.csv", row.names = FALSE, quote = FALSE, col.names = FALSE,
  sep = ",")
system("ls -l tmp4.csv", intern = TRUE)

## [1] "-rw-r--r-- 1 Esther 197609 19160587 Sep 15 12:32 tmp4.csv"

```

after rounding, length of each item shorter, so totally cost less space same concept as above, each item has length 4 or 5 and consider the delimiter of ',', the totally bytes will between 5×10^6 to 6×10^6

```

write.table(round(nums, 2), file = "tmp5.csv", row.names = FALSE, quote = FALSE,
  col.names = FALSE, sep = ",")
system("ls -l tmp5.csv", intern = TRUE)

## [1] "-rw-r--r-- 1 Esther 197609 6378532 Sep 15 12:32 tmp5.csv"

```

(b) each char is 1byte and consider the system setting it will be about 10^6 bytes

```

chars <- sample(letters, 1e+06, replace = TRUE)
chars <- paste(chars, collapse = "")
save(chars, file = "tmp6.Rda")
system("ls -l tmp6.Rda", intern = TRUE)

## [1] "-rw-r--r-- 1 Esther 197609 635280 Sep 15 12:32 tmp6.Rda"

summary(chars)

##      Length      Class      Mode 
##           1 character character

object_size(chars[1])

## 1 MB

```

use rep function and save into a pointer, so the size will become smaller significantly.

```
chars <- rep("a", 1e+06)
chars <- paste(chars, collapse = "")
save(chars, file = "tmp7.Rda")
system("ls -l tmp7.Rda", intern = TRUE)

## [1] "-rw-r--r-- 1 Esther 197609 1056 Sep 15 12:32 tmp7.Rda"

summary(chars)

##      Length      Class      Mode 
##           1 character character

object_size(chars[1])

## 1 MB
```

2 (a) step1, add scholar's name into the url, and use paste function to combine it to the original url step2, use if statement to rule out the invalid scholar names step3, use readline and getlink to the scholar's citation page step4, get the scholar's id from those links obtain in step3 and add it into the google scholar url to download the html page step5, print out the scholar's id and return the scholar's html page

```
library(stringr)
library(XML)
library(RCurl)
library(curl)
scholarlookup <- function(sch_name) {
  url <- paste("https://scholar.google.com/citations?view_op=search_authors&mauthors=",
    sch_name, "&hl=en&oi=ao")
  Sys.sleep(2)
  if (str_detect(sch_name, "[[:punct:]]") || str_detect(sch_name, "[[:digit:]]")) {
    cat("invalid scholar name")
  } else {
    page <- readLines(url)
    link <- getHTMLLinks(page)
    id <- strsplit(strsplit(link[9], split = "=")[[1]][2], split = "&")[[1]][1]
    htmlpage <- gsub(" ", "", paste("https://scholar.google.com", link[9]),
      fixed = TRUE)

    scholar <- htmlParse(readLines(htmlpage))
    cat(sch_name, "s id is", id, "\n")
    return(scholar)
  }
}
```

```
scholarlookup("Geoffrey Hinton") #test1

## Warning in readLines(url): incomplete final line found on 'http
Geoffrey Hinton &hl=en&oi=ao'
## Warning in readLines(htmlpage): incomplete final line found on

## Geoffrey Hinton 's id is JicYPdAAAAAJ
## <!DOCTYPE html>
## <html>
## <head>
## <meta http-equiv="Content-Type" content="text/html; charset=ISO-
## <meta http-equiv="X-UA-Compatible" content="IE=Edge">
## <meta name="referrer" content="always">
## <meta name="viewport" content="width=device-width, initial-scale=
## <style>@viewport{width:device-width;min-zoom:1;max-zoom:2;}</st
## <meta name="format-detection" content="telephone=no">
## <style>html,body,form,table,div,h1,h2,h3,h4,h5,h6,img,ol,ul,li,
## <script>var gs_ie_ver=100;</script><!--[if lte IE 8]><script>gs
## <link rel="canonical" href="http://scholar.google.co.uk/citatio
## </head>
## <body>
## <div id="gs_top">
## <style>#gs_gb{position:relative;height:30px;background:#2d2d2d;
## <div id="gs_gb" role="navigation">
## <div id="gs_gb_lt">
## <a href="//www.google.com/webhp?hl=en&amp;">Web</a><a href="//w
## </div>
## <a id="gs_gb_lp" aria-label="Web" href="//www.google.com/webhp?
## </div>
## <!--[if lte IE 9]><div class="gs_alrt" style="padding:16px"><di
## <script>function gs_md_opw(d,e,b){var r=document.documentElemen
## </div>";gs_md_opw(d,e,b);gs_ajax(u,p,function(x,t){c.innerHTML=
## <div id="gs_md_w" style="display:none">
## <div id="gsc_md_cbym" style="display:none" class="gs_md_d gs_md
## <a id="gsc_md_cbym-x" href="#" role="button" aria-label="Cancel
## <p>This "Cited by" count includes citations to the following ar
## <div id="gsc_md_cbym_l"></div>
## <div class="gs_md_btns"><button type="button" id="gsc_md_cbym_c
## <script>function gsc_md_show_cbym(x,e){var a=gs_id("gsc_md_cbym
## </div>
## <script>gs_md_ifc("gsc_md_cbym","gsc_md_cbym-x");</script><div
## <a id="gsc_md_cbyd-x" href="#" role="button" aria-label="Cancel
```

```

## <p>The following articles are merged in Scholar. Their <a id="gsc_md_cbyd_c" href="#">con
## <div id="gsc_md_cbyd_l"></div>
## <div class="gs_md_btns"><button type="button" id="gsc_md_cbyd_cancel" class=""><span clas
## <script>function gsc_md_show_cbyd(a,x,u,e){var f=gs_id("gsc_md_cbyd_f"),m=gs_id("gsc_md_c
## </div>
## <script>gs_md_ifc("gsc_md_cbyd","gsc_md_cbyd-x");</script><div id="gsc_md_hist" style="di
## <a id="gsc_md_hist-x" href="#" role="button" aria-label="Cancel" class="gs_ico gs_ico_X g
## <div id="gsc_md_hist_c"></div>
## </div>
## <script>gs_md_ifc("gsc_md_hist","gsc_md_hist-x");</script>
## </div>
## <style>#gs_alrt_w{position:fixed;width:100%;z-index:1050;display:none;opacity:0;transiti
## <div id="gs_alrt_w"><div id="gs_alrt_p">
## <form><input id="gs_alrt_ss"></form>
## <form action="" method="post" id="gs_alrt" class="gs_alrt">
## <span id="gs_alrt_m"></span><span id="gs_alrt_h"></span><span id="gs_alrt_u" style="displ
## </form>
## </div></div>
## <script>var gs_wa_t;function gs_wa_hf(e){gs_uevt(document,"click",gs_wa_hf);clearTimeout
## <div id="gsc_rsb_m" role="search"><div class="gsc_rsb_s_m" id="gsc_rsb_lg_m">
## <a href="/schhp?hl=en&oe=ASCII" id="gsc_rsb_icol_m" aria-label="Scholar Home"><span c
## <input type="hidden" name="hl" value="en"><input type="hidden" name="oe" value="ASCII"><
## </form>
## </div></div>
## <div class="gsc_lcl" role="main" id="gsc_prf_w"><div id="gsc_prf" class="gs_scl">
## <div id="gsc_prf_pu">
## <style>#gsc_prf_pup{width:96px;height:128px;}.gs_el_ta #gsc_prf_pup{width:64px;height:85p
## <a href="/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII" id="gsc_prf_pua">
## <div id="gsc_prf_ib"><div id="gsc_prf_ibi">
## <div style="position:relative" class="gs_ibl">
## <button type="button" id="gsc_fol_btn" aria-controls="gsc_fol_dd" aria-haspopup="true" cl
## <input type="hidden" name="xsrf" value="AMstHGQAAAAAWb19lkmftJqHnG1YY1LusyWbFk5_e740"><in
## <div class="gs_in_txtf"><input type="text" class="gs_in_txt" name="email_for_op" value=""
## <div id="gsc_fol_cb">
## <div class="gsc_fol_cr"><span class="gs_in_cb" onclick="void(0)"><input type="checkbox" r
## <div class="gsc_fol_cr"><span class="gs_in_cb" onclick="void(0)"><input type="checkbox" r
## </div>
## <div>
## <button type="submit" id="gsc_fol_b" class=" gs_btn_cre"><span class="gs_wr"><span class=
## </div>
## </form></div>
## </div>
## <script>gs_md_ifc("gsc_fol_dd","gsc_fol_m");!function(){var t=gs_id("gsc_fol_btn"),d="gs

```

```

## </div></div>
## <div id="gsc_prf_in">Geoffrey Hinton</div>
## <div class="gsc_prf_il">Emeritus Professor of Computer Science, <a href="/citations?view=
## <div class="gsc_prf_il">
## <a href="/citations?view_op=search_authors&hl=en&oe=ASCII&mauthors=label:mach
## </div>
## <div class="gsc_prf_il" id="gsc_prf_ivh">Verified email at cs.toronto.edu - <a href="http
## </div>
## </div>
## </div></div>
## <div id="gsc_rsb" role="navigation">
## <div class="gsc_rsb_s" id="gsc_rsb_lg">
## <a href="/schhp?hl=en&oe=ASCII" id="gsc_rsb_icol" aria-label="Scholar Home"><span cla
## <input type="hidden" name="hl" value="en"><input type="hidden" name="oe" value="ASCII"><
## </form>
## <div id="gsc_rsb_gpl"><button type="button" onclick="window.location='/citations?hl\x3den
## </div>
## <div class="gsc_rsb_s">
## <div id="gsc_rsb_t">Google Scholar</div>
## <table id="gsc_rsb_st">
## <tr>
## <th class="gsc_rsb_sc1"><a href="javascript:void(0)" onclick="gs_md_hist_opn(event);">Cit
## <th class="gsc_rsb_sth">All</th>
## <th class="gsc_rsb_sth">Since 2012</th>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='This is the
## <td class="gsc_rsb_std">182782</td>
## <td class="gsc_rsb_std">92276</td>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='h-index is
## <td class="gsc_rsb_std">132</td>
## <td class="gsc_rsb_std">97</td>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='i10-index i
## <td class="gsc_rsb_std">311</td>
## <td class="gsc_rsb_std">225</td>
## </tr>
## </table>
## <div id="gsc_g">
## <div id="gsc_g_x">
## <span class="gsc_g_t" style="left:3px">2009</span><span class="gsc_g_t" style="left:35px">
## </div>

```

```

## <div id="gsc_gBars">
## <a href="javascript:void(0)" class="gsc_g_a" style="left:8px;height:9px;z-index:9"><span>
## </div>
## </div>
## <script>function gs_md_hist_opn(e){gs_md_ldw("gsc_md_hist",e,0,"gsc_md_hist_c","/citation
## </div>
## <div class="gsc_rsb_s" id="gsc_rsb_co">
## <h3 class="gsc_rsb_h">Co-authors<a class="gsc_rsb_lc" href="/citations?view_op=list_coll
## </h3>
## <ul class="gsc_rsb_a">
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=m1qAiOUAAAAJ&hl=en&oe=ASCII">Terrence
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=S2yYRQIAAAAAJ&hl=en&oe=ASCII">James L
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=tJ_PrzgAAAAJ&hl=en&oe=ASCII">Abdel-ra
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=rr8pZoUAAAAJ&hl=en&oe=ASCII">Radford
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=ghbWy-OAAAAJ&hl=en&oe=ASCII">George B
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=RnoIxUwAAAAJ&hl=en&oe=ASCII">Vinod Na
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=bf7ilxMAAAAAJ&hl=en&oe=ASCII">Sidney B
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=rvKJDbIAAAAAJ&hl=en&oe=ASCII">Chris W
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=Rr847PIAAAAJ&hl=en&oe=ASCII">David C
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=ZpG_cJwAAAAJ&hl=en&oe=ASCII">Robert T
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=pKuBFaQAAAAJ&hl=en&oe=ASCII">Demetri
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=v02T2c0AAAAJ&hl=en&oe=ASCII">Steven I

```

```

## </li>
## <li><a class="gsc_rsb_aa" href="/citations?user=mZxsTCcAAAAJ&hl=en&oe=ASCII">Ram
## <li class="gsc_rsb_fade">
## </ul>
## </div>
## </div>
## <div class="gsc_lcl" role="complementary" id="gsc_art"><form method="post" action="/citat
## <input type="hidden" name="xsrf" value="AMstHGQAAAAWb19lkmftJqHnG1YY1LusyWbFk5_e740"><t
## <thead id="gsc_a_hd">
## <tr id="gsc_a_tr0" aria-hidden="true">
## <th class="gsc_a_t" id="gsc_a_tr0_t"></th>
## <th class="gsc_a_c"></th>
## <th class="gsc_a_y"></th>
## </tr>
## <tr id="gsc_a_trh">
## <th class="gsc_a_t" id="gsc_a_trh_t" scope="col">
## <span id="gsc_a_ta"><a href="/citations?hl=en&oe=ASCII&user=JicYPdAAAAAJ&view
## </th>
## <th class="gsc_a_c" scope="col"><span id="gsc_a_ca">Cited by</span></th>
## <th class="gsc_a_y" scope="col"><span class="gsc_a_h"><a href="/citations?hl=en&oe=AS
## </tr>
## </thead>
## <tbody id="gsc_a_b">
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">Nature 323, 533-536<span class="gs_oph">, 1986</span>
## </div>
## </td>
## <td class="gsc_a_c">
## <a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&cites=1374
## </td>
## <td class="gsc_a_y"><span class="gsc_a_h">1986</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">Parallel Distributed Processing: Explorations in the Microstructure
## </div>
## </td>
## <td class="gsc_a_c">
## <a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&cites=1374
## </td>
## <td class="gsc_a_y"><span class="gsc_a_h">1986</span></td>
## </tr>

```



```

## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">CALIFORNIA UNIV SAN DIEGO LA JOLLA INST FOR<span class="gs_oph">, 19
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">1985</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">MIT press 1, 184<span class="gs_oph">, 1987</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">1987</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">Advances in neural information processing systems, 1097-1105<span cl
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2012</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">Neural computation 18 (7), 1527-1554<span class="gs_oph">, 2006</sp
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2006</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">MIT press<span class="gs_oph">, 1987</span>
## </div>
## </td>
## <td class="gsc_a_c">
## <a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## </td>

```

```

## <td class="gsc_a_y"><span class="gsc_a_h">1987</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an=
## <div class="gs_gray">science 313 (5786), 504-507<span class="gs_oph">, 2006</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2006</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an=
## <div class="gs_gray">Nature 521 (7553), 436-444<span class="gs_oph">, 2015</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2015</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an=
## <div class="gs_gray">Neural computation 3 (1), 79-87<span class="gs_oph">, 1991</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">1991</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an=
## <div class="gs_gray">Journal of machine learning research 15 (1), 1929-1958<span class="gs_oph">, 2014</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2014</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an=
## <div class="gs_gray">Journal of machine learning research 15 (1), 1929-1958<span class="gs_oph">, 1985</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=

```

```

## <td class="gsc_a_y"><span class="gsc_a_h">1985</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">Journal of Machine Learning Research 9 (Nov), 2579-2605<span class="
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2008</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">IEEE Signal Processing Magazine 29 (6), 82-97<span class="gs_oph">,
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2012</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">Neural computation 14 (8), 1771-1800<span class="gs_oph">, 2002</sp
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">Learning in graphical models, 355-368<span class="gs_oph">, 1998</sp
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">1998</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">IEEE transactions on acoustics, speech, and signal processing 37 (3)
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=

```

```

## <td class="gsc_a_y"><span class="gsc_a_h">1989</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">arXiv preprint arXiv:1207.0580<span class="gs_oph">, 2012</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2012</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">Proceedings of the 27th international conference on machine learning
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2010</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">Artificial intelligence 40 (1-3), 185-234<span class="gs_oph">, 1989
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">1989</span></td>
## </tr>
## </tbody>
## </table>
## <div id="gsc_a_sp" style="display:none"></div>
## <div id="gsc_a_err" class="gs_alrt" style="display:none">The system can't perform the op
## <div id="gsc_lwp"><div id="gsc_bpf">
## <button type="button" id="gsc_bpf_prev" aria-label="Previous" disabled class="gs_btnPL gs
## </div></div>
## </form></div>
## <div class="gsc_lcl"><div id="gs_ftr" role="contentinfo">
## <div id="gsc_dscl">Dates and citation counts are estimated and are determined automaticall
## <div id="gsc_ftr_h">
## <a href="/intl/en/scholar/citations.html">Help</a> <a href="//www.google.com/intl/en/poli
## </div>
## </div></div>
## </div>
## <script>var gs_zvb;!function(u){gs_zvb=new Image();gs_zvb.onload=gs_zvb.onerror=function

```

```
## <div id="gs_rdy"></div>
## </body>
## </html>
##
```

```
scholarlookup("Sandrine Dudoit") #test2
```

```
## Warning in readLines(url): incomplete final line found on 'https://scholar.google.com/ci
Sandrine Dudoit &hl=en&oi=ao'
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

## Sandrine Dudoit 's id is hBWARYEAAAAJ
## <!DOCTYPE html>
## <html>
## <head>
## <meta http-equiv="Content-Type" content="text/html; charset=ISO-8859-1">
## <meta http-equiv="X-UA-Compatible" content="IE=Edge">
## <meta name="referrer" content="always">
## <meta name="viewport" content="width=device-width, initial-scale=1, minimum-scale=1, maximum
## <style>@viewport{width:device-width;min-zoom:1;max-zoom:2;}</style>
## <meta name="format-detection" content="telephone=no">
## <style>html,body,form,table,div,h1,h2,h3,h4,h5,h6,img,ol,ul,li,button{margin:0;padding:0;
## <script>var gs_ie_ver=100;</script><!--[if lte IE 8]><script>gs_ie_ver=8;</script><![end
## <link rel="canonical" href="http://scholar.google.com/citations?user=hBWARYEAAAAJ&amp;hl=
## </head>
## <body>
## <div id="gs_top">
## <style>#gs_gb{position:relative;height:30px;background:#2d2d2d;font-size:13px;line-height
## <div id="gs_gb" role="navigation">
## <div id="gs_gb_lt">
## <a href="//www.google.com/webhp?hl=en&amp;">Web</a><a href="//www.google.com/imghp?hl=en&
## </div>
## <a id="gs_gb_lp" aria-label="Web" href="//www.google.com/webhp?hl=en&amp;"></a><div id="g
## </div>
## <!--[if lte IE 9]><div class="gs_alrt" style="padding:16px"><div>Sorry, some features may
## <script>function gs_md_opw(d,e,b){var r=document.documentElement,s=gs_id("gs_md_s").style
## </div>";gs_md_opw(d,e,b);gs_ajax(u,p,function(x,t){c.innerHTML=x===200?t:"<div class="gs_
## <div id="gs_md_w" style="display:none">
## <div id="gsc_md_cbym" style="display:none" class="gs_md_d gs_md_wn gs_ttzi" role="dialog"
## <a id="gsc_md_cbym-x" href="#" role="button" aria-label="Cancel" class="gs_ico gs_ico_X g
## <p>This "Cited by" count includes citations to the following articles in Scholar. The one
## <div id="gsc_md_cbym_l"></div>
## <div class="gs_md_btns"><button type="button" id="gsc_md_cbym_cancel" class=""><span clas
## <script>function gsc_md_show_cbym(x,e){var a=gs_id("gsc_md_cbym_e");a&&(a.href=a.href.rep
## </div>
```

```

## <script>gs_md_ifc("gsc_md_cbyd","gsc_md_cbym-x");</script><div id="gsc_md_cbyd" style="display:none">
## <a id="gsc_md_cbyd-x" href="#" role="button" aria-label="Cancel" class="gs_ico gs_ico_X"></a>
## <p>The following articles are merged in Scholar. Their <a id="gsc_md_cbyd_c" href="#">cancel</a>
## <div id="gsc_md_cbyd_l"></div>
## <div class="gs_md_btns"><button type="button" id="gsc_md_cbyd_cancel" class="gs_md_btn"></button></div>
## <script>function gsc_md_show_cbyd(a,x,u,e){var f=gs_id("gsc_md_cbyd_f"),m=gs_id("gsc_md_cbyd_m");
## </div>
## <script>gs_md_ifc("gsc_md_cbyd","gsc_md_cbyd-x");</script><div id="gsc_md_hist" style="display:none">
## <a id="gsc_md_hist-x" href="#" role="button" aria-label="Cancel" class="gs_ico gs_ico_X"></a>
## <div id="gsc_md_hist_c"></div>
## </div>
## <script>gs_md_ifc("gsc_md_hist","gsc_md_hist-x");</script>
## </div>
## <style>#gs_alrt_w{position:fixed;width:100%;z-index:1050;display:none;opacity:0;transition:opacity 0.5s;
## <div id="gs_alrt_w"><div id="gs_alrt_p">
## <form><input id="gs_alrt_ss"></form>
## <form action="" method="post" id="gs_alrt" class="gs_alrt">
## <span id="gs_alrt_m"></span><span id="gs_alrt_h"></span><span id="gs_alrt_u" style="display:none"></span>
## </form>
## </div></div>
## <script>var gs_wa_t;function gs_wa_hf(e){gs_uevt(document,"click",gs_wa_hf);clearTimeout(gs_wa_t);
## <div id="gsc_rsb_m" role="search"><div class="gsc_rsb_s_m" id="gsc_rsb_lg_m">
## <a href="/schhp?hl=en&amp;oe=ASCII" id="gsc_rsb_icol_m" aria-label="Scholar Home"><span class="gs_ico"></span>
## <input type="hidden" name="hl" value="en"><input type="hidden" name="oe" value="ASCII"></div>
## </form>
## </div></div>
## <div class="gsc_lcl" role="main" id="gsc_prf_w"><div id="gsc_prf" class="gs_scl">
## <div id="gsc_prf_pu">
## <style>#gsc_prf_pup{width:128px;height:128px;}.gs_el_ta #gsc_prf_pup{width:85px;height:85px;}.gs_el_ta
## <a href="/citations?user=hBWARYEAAAAAJ&amp;hl=en&amp;oe=ASCII" id="gsc_prf_pua">
## <div id="gsc_prf_ib"><div id="gsc_prf_ibi">
## <div style="position:relative" class="gs_ibl">
## <button type="button" id="gsc_fol_btn" aria-controls="gsc_fol_dd" aria-haspopup="true" class="gs_btn"></button>
## <input type="hidden" name="xsrif" value="AMstHGQAAAAAWb19mSN2LW5OEKBWZm4SeSew5Kd7OW3w"><input type="text"
## <div class="gs_in_txtf"><input type="text" class="gs_in_txt" name="email_for_op" value=""></div>
## <div id="gsc_fol_cb">
## <div class="gsc_fol_cr"><span class="gs_in_cb" onclick="void(0)"><input type="checkbox" value=""></div>
## <div class="gsc_fol_cr"><span class="gs_in_cb" onclick="void(0)"><input type="checkbox" value=""></div>
## </div>
## <div>
## <button type="submit" id="gsc_fol_b" class="gs_btn_cre"><span class="gs_wr"><span class="gs_wr"></span></span>
## </div>
## </form></div>

```

```

## </div>
## <script>gs_md_ifc("gsc_fol_dd","gsc_fol_m");!function(){var t=gs_id("gsc_fol_btn"),d="gs
## </div></div>
## <div id="gsc_prf_in">Sandrine Dudoit</div>
## <div class="gsc_prf_il">Division of Biostatistics and Department of Statistics, <a href=
## </div>
## <div class="gsc_prf_il">
## <a href="/citations?view_op=search_authors&hl=en&oe=ASCII&mauthors=label:app
## </div>
## <div class="gsc_prf_il" id="gsc_prf_ivh">Verified email at stat.berkeley.edu - <a href="
## </div>
## </div>
## </div></div>
## <div id="gsc_rsb" role="navigation">
## <div class="gsc_rsb_s" id="gsc_rsb_lg">
## <a href="/schhp?hl=en&oe=ASCII" id="gsc_rsb_icol" aria-label="Scholar Home"><span cla
## <input type="hidden" name="hl" value="en"><input type="hidden" name="oe" value="ASCII"><
## </form>
## <div id="gsc_rsb_gpl"><button type="button" onclick="window.location='/citations?hl\x3der
## </div>
## <div class="gsc_rsb_s">
## <div id="gsc_rsb_t">Google Scholar</div>
## <table id="gsc_rsb_st">
## <tr>
## <th class="gsc_rsb_sc1"><a href="javascript:void(0)" onclick="gs_md_hist_opn(event);">Cit
## <th class="gsc_rsb_sth">All</th>
## <th class="gsc_rsb_sth">Since 2012</th>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='This is the
## <td class="gsc_rsb_std">34836</td>
## <td class="gsc_rsb_std">15933</td>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='h-index is
## <td class="gsc_rsb_std">52</td>
## <td class="gsc_rsb_std">38</td>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='i10-index i
## <td class="gsc_rsb_std">74</td>
## <td class="gsc_rsb_std">58</td>
## </tr>
## </table>
## <div id="gsc_g">

```

```

## <div id="gsc_g_x">
## <span class="gsc_g_t" style="left:3px">2009</span><span class="gsc_g_t" style="left:35px">
## </div>
## <div id="gsc_gBars">
## <a href="javascript:void(0)" class="gsc_g_a" style="left:8px;height:45px;z-index:9"><span
## </div>
## </div>
## <script>function gs_md_hist_opn(e){gs_md_ldw("gsc_md_hist",e,0,"gsc_md_hist_c","/citation
## </div>
## </div>
## <div class="gsc_lcl" role="complementary" id="gsc_art"><form method="post" action="/citati
## <input type="hidden" name="xsrf" value="AMstHGQAAAAAwb19mSN2LW50EKBWZm4SeSew5Kd70W3w"><ta
## <thead id="gsc_a_hd">
## <tr id="gsc_a_tr0" aria-hidden="true">
## <th class="gsc_a_t" id="gsc_a_tr0_t"></th>
## <th class="gsc_a_c"></th>
## <th class="gsc_a_y"></th>
## </tr>
## <tr id="gsc_a_trh">
## <th class="gsc_a_t" id="gsc_a_trh_t" scope="col">
## <span id="gsc_a_ta"><a href="/citations?hl=en&amp;oe=ASCII&amp;user=hBWARYEAAAAJ&amp;view
## </th>
## <th class="gsc_a_c" scope="col"><span id="gsc_a_ca">Cited by</span></th>
## <th class="gsc_a_y" scope="col"><span class="gsc_a_h"><a href="/citations?hl=en&amp;oe=AS
## </tr>
## </thead>
## <tbody id="gsc_a_b">
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&amp;hl=en&amp;oe=ASCII&amp;user=hBWARYEAAAAJ&an
## <div class="gs_gray">Genome biology 5 (10), R80<span class="gs_oph">, 2004</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&amp;hl=en&amp;oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2004</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&amp;hl=en&amp;oe=ASCII&amp;user=hBWARYEAAAAJ&an
## <div class="gs_gray">Nucleic acids research 30 (4), e15-e15<span class="gs_oph">, 2002</s
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&amp;hl=en&amp;oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>

```



```

## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&ar
## <div class="gs_gray">Journal of the American statistical association 97 (457), 77-87<span
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&ar
## <div class="gs_gray">Springer<span class="gs_oph">, 2005</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2005</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&ar
## <div class="gs_gray">Statistica sinica 12 (1), 111-140<span class="gs_oph">, 2002</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&ar
## <div class="gs_gray">Statistical Science, 71-103<span class="gs_oph">, 2003</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2003</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&ar
## <div class="gs_gray">BMC bioinformatics 11 (1), 94<span class="gs_oph">, 2010</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2010</span></td>
## </tr>

```

```

## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an">
## <div class="gs_gray">Nature 471 (7339), 473-479<span class="gs_oph">, 2011</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an">
## <td class="gsc_a_y"><span class="gsc_a_h">2011</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an">
## <div class="gs_gray">Proceedings of the National Academy of Sciences 99 (20), 12877-12882
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an">
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an">
## <div class="gs_gray">Science 330 (6012), 1787-1797<span class="gs_oph">, 2010</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an">
## <td class="gsc_a_y"><span class="gsc_a_h">2010</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an">
## <div class="gs_gray">Molecular biology of the cell 13 (6), 1929-1939<span class="gs_oph">
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an">
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an">
## <div class="gs_gray">Genome biology 3 (7), research0036<span class="gs_oph">, 2002</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an">
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>

```

```

## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Journal of computational and graphical statistics 11 (1), 108-136<sp
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">bepress<span class="gs_oph">, 2004</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2004</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Test 12 (1), 1-77<span class="gs_oph">, 2003</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2003</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Springer<span class="gs_oph">, 2008</span>
## </div>
## </td>
## <td class="gsc_a_c">
## <a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&cites=148
## </td>
## <td class="gsc_a_y"><span class="gsc_a_h">2008</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">BiOS 2001 The International Symposium on Biomedical Optics, 141-152<
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=

```

```

## <td class="gsc_a_y"><span class="gsc_a_h">2001</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Bioinformatics 19 (9), 1090-1099<span class="gs_oph">, 2003</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2003</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Nucleic acids research 38 (12), e131-e131<span class="gs_oph">, 2010
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2010</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Proceedings of the National Academy of Sciences 99 (2), 972-977<span
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## </tbody>
## </table>
## <div id="gsc_a_sp" style="display:none"></div>
## <div id="gsc_a_err" class="gs_alrt" style="display:none">The system can't perform the op
## <div id="gsc_lwp"><div id="gsc_bpf">
## <button type="button" id="gsc_bpf_prev" aria-label="Previous" disabled class="gs_btnPL gs
## </div></div>
## </form></div>
## <div class="gsc_lcl"><div id="gs_ftr" role="contentinfo">
## <div id="gsc_dscl">Dates and citation counts are estimated and are determined automaticall
## <div id="gsc_ftr_h">
## <a href="/intl/en/scholar/citations.html">Help</a> <a href="//www.google.com/intl/en/poli
## </div>
## </div></div>
## </div>
## <div id="gs_rdy"></div>

```

```
## </body>
## </html>
##
```

(b) step1, use the above results of 2(a) to get scholar's citation page step2, use getNodeSet function to get different part of html page, and found article, in under 'gsc_a_t' class step3, found authors and journal are both in the 'gs_gray' class step4, divided the authors/journal by di-

```
scholarlookup <- function(sch_name) {
  url <- paste("https://scholar.google.com/citations?view_op=search_authors&mauthors=",
    sch_name, "&hl=en&oi=ao")
  Sys.sleep(2)
  if (str_detect(sch_name, "[[:punct:]]") || str_detect(sch_name, "[[:digit:]]")) {
    cat("invalid scholar name")
  } else {
    page <- readLines(url)
    link <- getHTMLLinks(page)
    id <- strsplit(strsplit(link[9], split = "=")[[1]][2], split = "&")[[1]][1]
    htmlpage <- gsub(" ", "", paste("https://scholar.google.com", link[9]),
      fixed = TRUE)

    scholar <- htmlParse(readLines(htmlpage))
    article_content <- getNodeSet(scholar, "//td[@class = 'gsc_a_t']/a")
    article <- sapply(article_content, xmlValue)
    gray_content <- getNodeSet(scholar, "//div[@class='gs_gray']")
    a_j <- sapply(gray_content, xmlValue)
    authors <- a_j[seq(1, length(a_j), 2)]
    journal <- a_j[seq(2, length(a_j), 2)]
    cit_content <- getNodeSet(scholar, "//td[@class = 'gsc_a_c']/a")
    cit <- sapply(cit_content, xmlValue)
    year_content <- getNodeSet(scholar, "//span[@class = 'gsc_a_h']")
    year <- sapply(year_content, xmlValue)[-1]
    df_sc <- data.frame(article, authors, journal, cit, year)
    return(summary(df_sc))
  }
}
```

```
scholarlookup("Geoffrey Hinton") #test1

## Warning in readLines(url): incomplete final line found on 'https://scholar.google.com/ci
Geoffrey Hinton &hl=en&oi=ao'
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

## Geoffrey Hinton 's id is JicYPdAAAAAJ
```

```

## <!DOCTYPE html>
## <html>
## <head>
## <meta http-equiv="Content-Type" content="text/html; charset=ISO-8859-1">
## <meta http-equiv="X-UA-Compatible" content="IE=Edge">
## <meta name="referrer" content="always">
## <meta name="viewport" content="width=device-width, initial-scale=1, minimum-scale=1, maximum-scale=1">
## <style>@viewport{width:device-width;min-zoom:1;max-zoom:2;}</style>
## <meta name="format-detection" content="telephone=no">
## <style>html,body,form,table,div,h1,h2,h3,h4,h5,h6,img,ol,ul,li,button{margin:0;padding:0;}</style>
## <script>var gs_ie_ver=100;</script><!--[if lte IE 8]><script>gs_ie_ver=8;</script><![endif]>
## <link rel="canonical" href="http://scholar.google.co.uk/citations?user=JicYPdAAAAAJ&amp;hl=en">
## </head>
## <body>
## <div id="gs_top">
## <style>#gs_gb{position:relative;height:30px;background:#2d2d2d;font-size:13px;line-height:1.2}</style>
## <div id="gs_gb" role="navigation">
## <div id="gs_gb_lt">
## <a href="//www.google.com/webhp?hl=en&amp;">Web</a><a href="//www.google.com/imghp?hl=en&amp;">Images</a></div>
## <a id="gs_gb_lp" aria-label="Web" href="//www.google.com/webhp?hl=en&amp;"></a><div id="gs_gb_rp">
## </div>
## <!--[if lte IE 9]><div class="gs_alrt" style="padding:16px"><div>Sorry, some features may not work properly in your browser. Please use a supported browser to get the best experience.</div>
## <script>function gs_md_opw(d,e,b){var r=document.documentElement,s=gs_id("gs_md_s").style.visibility="visible";gs_md_opw(d,e,b);gs_ajax(u,p,function(x,t){c.innerHTML=x===200?t:"<div class="gs_md_w" style="display:none">
## <div id="gs_md_w" style="display:none">
## <div id="gsc_md_cbym" style="display:none" class="gs_md_d gs_md_wn gs_ttzi" role="dialog">
## <a id="gsc_md_cbym-x" href="#" role="button" aria-label="Cancel" class="gs_ico gs_ico_X gs_md_cbym_x"></a>
## <p>This "Cited by" count includes citations to the following articles in Scholar. The one with the most citations is shown first.
## <div id="gsc_md_cbym_l"></div>
## <div class="gs_md_btns"><button type="button" id="gsc_md_cbym_cancel" class=""><span class="gs_md_cbym_cancel"></span></button></div>
## <script>function gsc_md_show_cbym(x,e){var a=gs_id("gsc_md_cbym_e");a&&(a.href=a.href.replace("#",""));gs_md_ifc("gsc_md_cbym","gsc_md_cbym-x");</script><div id="gsc_md_cbyd" style="display:none">
## <a id="gsc_md_cbyd-x" href="#" role="button" aria-label="Cancel" class="gs_ico gs_ico_X gs_md_cbyd_x"></a>
## <p>The following articles are merged in Scholar. Their <a id="gsc_md_cbyd_c" href="#">combined citation counts</a> are shown.
## <div id="gsc_md_cbyd_l"></div>
## <div class="gs_md_btns"><button type="button" id="gsc_md_cbyd_cancel" class=""><span class="gs_md_cbyd_cancel"></span></button></div>
## <script>function gsc_md_show_cbyd(a,x,u,e){var f=gs_id("gsc_md_cbyd_f"),m=gs_id("gsc_md_cbyd_m");f.innerHTML+=a;gs_md_ifc("gsc_md_cbyd","gsc_md_cbyd-x");</script><div id="gsc_md_hist" style="display:none">
## <a id="gsc_md_hist-x" href="#" role="button" aria-label="Cancel" class="gs_ico gs_ico_X gs_md_hist_x"></a>
## <div id="gsc_md_hist_c"></div>
## </div>
## <script>gs_md_ifc("gsc_md_hist","gsc_md_hist-x");</script>

```

```

## </div>
## <style>#gs_alrt_w{position:fixed;width:100%;z-index:1050;display:none;opacity:0;transition:
## <div id="gs_alrt_w"><div id="gs_alrt_p">
## <form><input id="gs_alrt_ss"></form>
## <form action="" method="post" id="gs_alrt" class="gs_alrt">
## <span id="gs_alrt_m"></span><span id="gs_alrt_h"></span><span id="gs_alrt_u" style="display:
## </form>
## </div></div>
## <script>var gs_wa_t;function gs_wa_hf(e){gs_uevt(document,"click",gs_wa_hf);clearTimeout
## <div id="gsc_rsb_m" role="search"><div class="gsc_rsb_s_m" id="gsc_rsb_lg_m">
## <a href="/schhp?hl=en&oe=ASCII" id="gsc_rsb_icol_m" aria-label="Scholar Home"><span c
## <input type="hidden" name="hl" value="en"><input type="hidden" name="oe" value="ASCII"></
## </form>
## </div></div>
## <div class="gsc_lcl" role="main" id="gsc_prf_w"><div id="gsc_prf" class="gs_scl">
## <div id="gsc_prf_pu">
## <style>#gsc_prf_pup{width:96px;height:128px;}.gs_el_ta #gsc_prf_pup{width:64px;height:85px;
## <a href="/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII" id="gsc_prf_pua">
## <div id="gsc_prf_ib"><div id="gsc_prf_ibi">
## <div style="position:relative" class="gs_ibl">
## <button type="button" id="gsc_fol_btn" aria-controls="gsc_fol_dd" aria-haspopup="true" c
## <input type="hidden" name="xsrf" value="AMstHGQAAAAAWb19nW5ISQOQ3whXDrQvq5eHhirTw5-B"><in
## <div class="gs_in_txtf"><input type="text" class="gs_in_txt" name="email_for_op" value="
## <div id="gsc_fol_cb">
## <div class="gsc_fol_cr"><span class="gs_in_cb" onclick="void(0)"><input type="checkbox" r
## <div class="gsc_fol_cr"><span class="gs_in_cb" onclick="void(0)"><input type="checkbox" r
## </div>
## <div>
## <button type="submit" id="gsc_fol_b" class=" gs_btn_cre"><span class="gs_wr"><span class=
## </div>
## </form></div>
## </div>
## <script>gs_md_ifc("gsc_fol_dd","gsc_fol_m");!function(){var t=gs_id("gsc_fol_btn"),d="gs
## </div></div>
## <div id="gsc_prf_in">Geoffrey Hinton</div>
## <div class="gsc_prf_il">Emeritus Professor of Computer Science, <a href="/citations?view
## <div class="gsc_prf_il">
## <a href="/citations?view_op=search_authors&hl=en&oe=ASCII&mauthors=label:mach
## </div>
## <div class="gsc_prf_il" id="gsc_prf_ivh">Verified email at cs.toronto.edu - <a href="http
## </div>
## </div>
## </div>
## </div></div>

```

```

## <div id="gsc_rsb" role="navigation">
## <div class="gsc_rsb_s" id="gsc_rsb_lg">
## <a href="/schhp?hl=en&oe=ASCII" id="gsc_rsb_icol" aria-label="Scholar Home"><span cla
## <input type="hidden" name="hl" value="en"><input type="hidden" name="oe" value="ASCII"></div>
## </div>
## <div id="gsc_rsb_gpl"><button type="button" onclick="window.location='/citations?hl\x3den'"
## </div>
## <div class="gsc_rsb_s">
## <div id="gsc_rsb_t">Google Scholar</div>
## <table id="gsc_rsb_st">
## <tr>
## <th class="gsc_rsb_sc1"><a href="javascript:void(0)" onclick="gs_md_hist_opn(event);">Cit
## <th class="gsc_rsb_sth">All</th>
## <th class="gsc_rsb_sth">Since 2012</th>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='This is the
## <td class="gsc_rsb_std">182782</td>
## <td class="gsc_rsb_std">92276</td>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='h-index is
## <td class="gsc_rsb_std">132</td>
## <td class="gsc_rsb_std">97</td>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='i10-index i
## <td class="gsc_rsb_std">311</td>
## <td class="gsc_rsb_std">225</td>
## </tr>
## </table>
## <div id="gsc_g">
## <div id="gsc_g_x">
## <span class="gsc_g_t" style="left:3px">2009</span><span class="gsc_g_t" style="left:35px"
## </div>
## <div id="gsc_gBars">
## <a href="javascript:void(0)" class="gsc_g_a" style="left:8px;height:9px;z-index:9"><span
## </div>
## </div>
## <script>function gs_md_hist_opn(e){gs_md_ldw("gsc_md_hist",e,0,"gsc_md_hist_c","/citation
## </div>
## <div class="gsc_rsb_s" id="gsc_rsb_co">
## <h3 class="gsc_rsb_h">Co-authors<a class="gsc_rsb_lc" href="/citations?view_op=list_coll
## </h3>
## <ul class="gsc_rsb_a">

```



```

## <li>
## <a class="gsc_rsb_aa" href="/citations?user=m1qAiOUAAAAJ&hl=en&oe=ASCII">Terrence
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=S2yYRQIAAAAAJ&hl=en&oe=ASCII">James L
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=tJ_PrzgAAAAJ&hl=en&oe=ASCII">Abdel-ra
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=rr8pZoUAAAAJ&hl=en&oe=ASCII">Radford
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=ghbWy-OAAAAJ&hl=en&oe=ASCII">George F
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=RnoIxUwAAAAJ&hl=en&oe=ASCII">Vinod Na
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=bf7ilxMAAAAAJ&hl=en&oe=ASCII">Sidney F
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=rvKJDbIAAAAAJ&hl=en&oe=ASCII">Chris W
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=Rr847PIAAAAJ&hl=en&oe=ASCII">David C
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=ZpG_cJwAAAAJ&hl=en&oe=ASCII">Robert T
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=pKuBFaQAAAAJ&hl=en&oe=ASCII">Demetri
## </li>
## <li>
## <a class="gsc_rsb_aa" href="/citations?user=v02T2c0AAAAJ&hl=en&oe=ASCII">Steven I
## </li>
## <li><a class="gsc_rsb_aa" href="/citations?user=mZxsTCcAAAAJ&hl=en&oe=ASCII">Rama
## <li class="gsc_rsb_fade">
## </ul>
## </div>
## </div>
## <div class="gsc_lcl" role="complementary" id="gsc_art"><form method="post" action="/citat
## <input type="hidden" name="xsrf" value="AMstHGQAAAAAwb19nW5ISQOQ3whXDrQvq5eHhirTw5-B"><ta
## <thead id="gsc_a_hd">
## <tr id="gsc_a_tr0" aria-hidden="true">

```

```

## <th class="gsc_a_t" id="gsc_a_tr0_t"></th>
## <th class="gsc_a_c"></th>
## <th class="gsc_a_y"></th>
## </tr>
## <tr id="gsc_a_trh">
## <th class="gsc_a_t" id="gsc_a_trh_t" scope="col">
## <span id="gsc_a_ta"><a href="/citations?hl=en&oe=ASCII&user=JicYPdAAAAAJ&view=ASCI">
## </th>
## <th class="gsc_a_c" scope="col"><span id="gsc_a_ca">Cited by</span></th>
## <th class="gsc_a_y" scope="col"><span class="gsc_a_h"><a href="/citations?hl=en&oe=ASCI">
## </tr>
## </thead>
## <tbody id="gsc_a_b">
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&view=ASCI">
## <div class="gs_gray">Nature 323, 533-536<span class="gs_oph">, 1986</span></div>
## </td>
## <td class="gsc_a_c">
## <a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&cites=1374">
## </td>
## <td class="gsc_a_y"><span class="gsc_a_h">1986</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&view=ASCI">
## <div class="gs_gray">Parallel Distributed Processing: Explorations in the Microstructure
## </div>
## </td>
## <td class="gsc_a_c">
## <a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&cites=1374">
## </td>
## <td class="gsc_a_y"><span class="gsc_a_h">1986</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&view=ASCI">
## <div class="gs_gray">CALIFORNIA UNIV SAN DIEGO LA JOLLA INST FOR<span class="gs_oph">, 19
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCI">
## <td class="gsc_a_y"><span class="gsc_a_h">1985</span></td>
## </tr>
## <tr class="gsc_a_tr">

```

```

## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">MIT press 1, 184<span class="gs_oph">, 1987</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">1987</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">Advances in neural information processing systems, 1097-1105<span c
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2012</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">Neural computation 18 (7), 1527-1554<span class="gs_oph">, 2006</spa
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2006</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">MIT press<span class="gs_oph">, 1987</span>
## </div>
## </td>
## <td class="gsc_a_c">
## <a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&cites=4574
## </td>
## <td class="gsc_a_y"><span class="gsc_a_h">1987</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">science 313 (5786), 504-507<span class="gs_oph">, 2006</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2006</span></td>

```

```

## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">Nature 521 (7553), 436-444<span class="gs_oph">, 2015</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2015</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">Neural computation 3 (1), 79-87<span class="gs_oph">, 1991</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">1991</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">Journal of machine learning research 15 (1), 1929-1958<span class="g
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2014</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">Cognitive science 9 (1), 147-169<span class="gs_oph">, 1985</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">1985</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&an
## <div class="gs_gray">Journal of Machine Learning Research 9 (Nov), 2579-2605<span class="
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2008</span></td>

```

```

## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">IEEE Signal Processing Magazine 29 (6), 82-97<span class="gs_oph">,
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2012</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">Neural computation 14 (8), 1771-1800<span class="gs_oph">, 2002</sp
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">Learning in graphical models, 355-368<span class="gs_oph">, 1998</sp
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">1998</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">IEEE transactions on acoustics, speech, and signal processing 37 (3)
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">1989</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">arXiv preprint arXiv:1207.0580<span class="gs_oph">, 2012</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2012</span></td>

```

```

## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">Proceedings of the 27th international conference on machine learning
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2010</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=JicYPdAAAAAJ&ar
## <div class="gs_gray">Artificial intelligence 40 (1-3), 185-234<span class="gs_oph">, 1989
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">1989</span></td>
## </tr>
## </tbody>
## </table>
## <div id="gsc_a_sp" style="display:none"></div>
## <div id="gsc_a_err" class="gs_alrt" style="display:none">The system can't perform the op
## <div id="gsc_lwp"><div id="gsc_bpf">
## <button type="button" id="gsc_bpf_prev" aria-label="Previous" disabled class="gs_btnPL gs
## </div></div>
## </form></div>
## <div class="gsc_lcl"><div id="gs_ftr" role="contentinfo">
## <div id="gsc_dscl">Dates and citation counts are estimated and are determined automaticall
## <div id="gsc_ftr_h">
## <a href="/intl/en/scholar/citations.html">Help</a> <a href="//www.google.com/intl/en/poli
## </div>
## </div></div>
## </div>
## <script>var gs_zvb;!function(u){gs_zvb=new Image();gs_zvb.onload=gs_zvb.onerror=function
## <div id="gs_rdy"></div>
## </body>
## </html>
##

```

```

scholarlookup("Sandrine Dudoit") #test2

```

```

## Warning in readLines(url): incomplete final line found on 'https://scholar.google.com/ci
Sandrine Dudoit &hl=en&oi=ao'
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

```

```

## Sandrine Dudoit 's id is hBWARYEAAAAJ
## <!DOCTYPE html>
## <html>
## <head>
## <meta http-equiv="Content-Type" content="text/html; charset=ISO-8859-1">
## <meta http-equiv="X-UA-Compatible" content="IE=Edge">
## <meta name="referrer" content="always">
## <meta name="viewport" content="width=device-width, initial-scale=1, minimum-scale=1, maximum-scale=1">
## <style>@viewport{width:device-width;min-zoom:1;max-zoom:2;}</style>
## <meta name="format-detection" content="telephone=no">
## <style>html,body,form,table,div,h1,h2,h3,h4,h5,h6,img,ol,ul,li,button{margin:0;padding:0;}</style>
## <script>var gs_ie_ver=100;</script><!--[if lte IE 8]><script>gs_ie_ver=8;</script><![endif]>
## <link rel="canonical" href="http://scholar.google.com/citations?user=hBWARYEAAAAJ&hl=en">
## </head>
## <body>
## <div id="gs_top">
## <style>#gs_gb{position:relative;height:30px;background:#2d2d2d;font-size:13px;line-height:30px;}</style>
## <div id="gs_gb" role="navigation">
## <div id="gs_gb_lt">
## <a href="//www.google.com/webhp?hl=en&";>Web</a><a href="//www.google.com/imghp?hl=en&";>Images</a></div>
## <a id="gs_gb_lp" aria-label="Web" href="//www.google.com/webhp?hl=en&";></a><div id="gs_gb_rp">
## </div>
## <!--[if lte IE 9]><div class="gs_alrt" style="padding:16px"><div>Sorry, some features may not work properly in your browser. Please use a supported browser to get the best experience.</div>
## <script>function gs_md_opw(d,e,b){var r=document.documentElement,s=gs_id("gs_md_s").style.display="block";gs_md_opw(d,e,b);gs_ajax(u,p,function(x,t){c.innerHTML=x===200?t:"<div class="gs_md_w" style="display:none">
## <div id="gs_md_w" style="display:none">
## <div id="gsc_md_cbym" style="display:none" class="gs_md_d gs_md_wn gs_ttzi" role="dialog">
## <a id="gsc_md_cbym-x" href="#" role="button" aria-label="Cancel" class="gs_ico gs_ico_X gs_md_cbym_x"></a>
## <p>This "Cited by" count includes citations to the following articles in Scholar. The one with the most citations is shown first.
## <div id="gsc_md_cbym_l"></div>
## <div class="gs_md_btns"><button type="button" id="gsc_md_cbym_cancel" class=""><span class="gs_md_cbym_cancel"></span></button></div>
## <script>function gsc_md_show_cbym(x,e){var a=gs_id("gsc_md_cbym_e");a&&(a.href=a.href.replace("#",""));gs_md_ifc("gsc_md_cbym","gsc_md_cbym-x");</script><div id="gsc_md_cbyd" style="display:none">
## <a id="gsc_md_cbyd-x" href="#" role="button" aria-label="Cancel" class="gs_ico gs_ico_X gs_md_cbyd_x"></a>
## <p>The following articles are merged in Scholar. Their <a id="gsc_md_cbyd_c" href="#">combined</a> citation count is shown.
## <div id="gsc_md_cbyd_l"></div>
## <div class="gs_md_btns"><button type="button" id="gsc_md_cbyd_cancel" class=""><span class="gs_md_cbyd_cancel"></span></button></div>
## <script>function gsc_md_show_cbyd(a,x,u,e){var f=gs_id("gsc_md_cbyd_f"),m=gs_id("gsc_md_cbyd_m");f.innerHTML+=a;gs_md_ifc("gsc_md_cbyd","gsc_md_cbyd-x");</script><div id="gsc_md_hist" style="display:none">
## <a id="gsc_md_hist-x" href="#" role="button" aria-label="Cancel" class="gs_ico gs_ico_X gs_md_hist_x"></a>
## <div id="gsc_md_hist_c"></div>
## </div>

```

```

## <script>gs_md_ifc("gsc_md_hist","gsc_md_hist-x");</script>
## </div>
## <style>#gs_alrt_w{position:fixed;width:100%;z-index:1050;display:none;opacity:0;transiti
## <div id="gs_alrt_w"><div id="gs_alrt_p">
## <form><input id="gs_alrt_ss"></form>
## <form action="" method="post" id="gs_alrt" class="gs_alrt">
## <span id="gs_alrt_m"></span><span id="gs_alrt_h"></span><span id="gs_alrt_u" style="displ
## </form>
## </div></div>
## <script>var gs_wa_t;function gs_wa_hf(e){gs_uevt(document,"click",gs_wa_hf);clearTimeout
## <div id="gsc_rsb_m" role="search"><div class="gsc_rsb_s_m" id="gsc_rsb_lg_m">
## <a href="/schhp?hl=en&oe=ASCII" id="gsc_rsb_icol_m" aria-label="Scholar Home"><span c
## <input type="hidden" name="hl" value="en"><input type="hidden" name="oe" value="ASCII"><
## </form>
## </div></div>
## <div class="gsc_lcl" role="main" id="gsc_prf_w"><div id="gsc_prf" class="gs_scl">
## <div id="gsc_prf_pu">
## <style>#gsc_prf_pup{width:128px;height:128px;}.gs_el_ta #gsc_prf_pup{width:85px;height:8
## <a href="/citations?user=hBWARYEAAAAJ&hl=en&oe=ASCII" id="gsc_prf_pua"><img src=
## </div>
## <div id="gsc_prf_i">
## <div id="gsc_prf_ib"><div id="gsc_prf_ibi">
## <div style="position:relative" class="gs_ibl">
## <button type="button" id="gsc_fol_btn" aria-controls="gsc_fol_dd" aria-haspopup="true" c
## <input type="hidden" name="xsrf" value="AMstHGQAAAAAwB19oN7FMG2wG4i0PCrDrSLB1Gv7sv4M"><in
## <div class="gs_in_txtf"><input type="text" class="gs_in_txt" name="email_for_op" value="
## <div id="gsc_fol_cb">
## <div class="gsc_fol_cr"><span class="gs_in_cb" onclick="void(0)"><input type="checkbox" r
## <div class="gsc_fol_cr"><span class="gs_in_cb" onclick="void(0)"><input type="checkbox" r
## </div>
## <div>
## <button type="submit" id="gsc_fol_b" class=" gs_btn_cre"><span class="gs_wr"><span class=
## </div>
## </form></div>
## </div>
## <script>gs_md_ifc("gsc_fol_dd","gsc_fol_m");!function(){var t=gs_id("gsc_fol_btn"),d="gs
## </div></div>
## <div id="gsc_prf_in">Sandrine Dudoit</div>
## <div class="gsc_prf_il">Division of Biostatistics and Department of Statistics, <a href=
## </div>
## <div class="gsc_prf_il">
## <a href="/citations?view_op=search_authors&hl=en&oe=ASCII&mauthors=label:appl
## </div>
## <div class="gsc_prf_il" id="gsc_prf_ivh">Verified email at stat.berkeley.edu - <a href="
## </div>

```



```

## </div>
## </div></div>
## <div id="gsc_rsb" role="navigation">
## <div class="gsc_rsb_s" id="gsc_rsb_lg">
## <a href="/schhp?hl=en&oe=ASCII" id="gsc_rsb_icol" aria-label="Scholar Home"><span cla
## <input type="hidden" name="hl" value="en"><input type="hidden" name="oe" value="ASCII"></div>
## </form>
## <div id="gsc_rsb_gpl"><button type="button" onclick="window.location='/citations?hl\x3den'"
## </div>
## <div class="gsc_rsb_s">
## <div id="gsc_rsb_t">Google Scholar</div>
## <table id="gsc_rsb_st">
## <tr>
## <th class="gsc_rsb_sc1"><a href="javascript:void(0)" onclick="gs_md_hist_opn(event);">Cit
## <th class="gsc_rsb_sth">All</th>
## <th class="gsc_rsb_sth">Since 2012</th>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='This is the
## <td class="gsc_rsb_std">34836</td>
## <td class="gsc_rsb_std">15933</td>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='h-index is
## <td class="gsc_rsb_std">52</td>
## <td class="gsc_rsb_std">38</td>
## </tr>
## <tr>
## <td class="gsc_rsb_sc1"><a href="javascript:void(0)" class="gsc_rsb_f" title='i10-index i
## <td class="gsc_rsb_std">74</td>
## <td class="gsc_rsb_std">58</td>
## </tr>
## </table>
## <div id="gsc_g">
## <div id="gsc_g_x">
## <span class="gsc_g_t" style="left:3px">2009</span><span class="gsc_g_t" style="left:35px"
## </div>
## <div id="gsc_gBars">
## <a href="javascript:void(0)" class="gsc_g_a" style="left:8px;height:45px;z-index:9"><span
## </div>
## </div>
## <script>function gs_md_hist_opn(e){gs_md_ldw("gsc_md_hist",e,0,"gsc_md_hist_c","/citation
## </div>
## </div>
## <div class="gsc_lcl" role="complementary" id="gsc_art"><form method="post" action="/citat

```

```

## <input type="hidden" name="xsrf" value="AMstHGQAAAAAWb19oN7FMG2wG4i0PCrDrSLB1Gv7sv4M"><input type="hidden" name="xsrf" value="AMstHGQAAAAAWb19oN7FMG2wG4i0PCrDrSLB1Gv7sv4M"></input>
## <thead id="gsc_a_hd">
## <tr id="gsc_a_tr0" aria-hidden="true">
## <th class="gsc_a_t" id="gsc_a_tr0_t"></th>
## <th class="gsc_a_c"></th>
## <th class="gsc_a_y"></th>
## </tr>
## <tr id="gsc_a_trh">
## <th class="gsc_a_t" id="gsc_a_trh_t" scope="col">
## <span id="gsc_a_ta"><a href="/citations?hl=en&oe=ASCII&user=hBWARYEAAAAJ&view_op=view_citation" data-bbox="211 298 1000 313">Cited by</a></span></th>
## </th>
## <th class="gsc_a_c" scope="col"><span id="gsc_a_ca">Cited by</span></th>
## <th class="gsc_a_y" scope="col"><span class="gsc_a_h"><a href="/citations?hl=en&oe=ASCII&user=hBWARYEAAAAJ&view_op=view_citation" data-bbox="211 343 1000 358"></a></span></th>
## </tr>
## </thead>
## <tbody id="gsc_a_b">
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&view_op=view_citation" data-bbox="211 433 1000 448"></a>
## <div class="gs_gray">Genome biology 5 (10), R80<span class="gs_oph">, 2004</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&user=hBWARYEAAAAJ&view_op=view_citation" data-bbox="211 493 1000 508"></a>
## <td class="gsc_a_y"><span class="gsc_a_h">2004</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&view_op=view_citation" data-bbox="211 568 1000 583"></a>
## <div class="gs_gray">Nucleic acids research 30 (4), e15-e15<span class="gs_oph">, 2002</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&user=hBWARYEAAAAJ&view_op=view_citation" data-bbox="211 628 1000 643"></a>
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&view_op=view_citation" data-bbox="211 703 1000 718"></a>
## <div class="gs_gray">Journal of the American statistical association 97 (457), 77-87<span class="gs_oph">, 2002</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&user=hBWARYEAAAAJ&view_op=view_citation" data-bbox="211 763 1000 778"></a>
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">

```

```

## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Springer<span class="gs_oph">, 2005</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2005</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Statistica sinica 12 (1), 111-140<span class="gs_oph">, 2002</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Statistical Science, 71-103<span class="gs_oph">, 2003</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2003</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">BMC bioinformatics 11 (1), 94<span class="gs_oph">, 2010</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2010</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Nature 471 (7339), 473-479<span class="gs_oph">, 2011</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2011</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">

```

```

## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Proceedings of the National Academy of Sciences 99 (20), 12877-12882
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Science 330 (6012), 1787-1797<span class="gs_oph">, 2010</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2010</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Molecular biology of the cell 13 (6), 1929-1939<span class="gs_oph">
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Genome biology 3 (7), research0036<span class="gs_oph">, 2002</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Journal of computational and graphical statistics 11 (1), 108-136<sp
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">

```

```

## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">bepress<span class="gs_oph">, 2004</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2004</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Test 12 (1), 1-77<span class="gs_oph">, 2003</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2003</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Springer<span class="gs_oph">, 2008</span>
## </div>
## </td>
## <td class="gsc_a_c">
## <a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=ASCII&cites=1488
## </td>
## <td class="gsc_a_y"><span class="gsc_a_h">2008</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">BiOS 2001 The International Symposium on Biomedical Optics, 141-152
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2001</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&an
## <div class="gs_gray">Bioinformatics 19 (9), 1090-1099<span class="gs_oph">, 2003</span>
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2003</span></td>
## </tr>

```

```

## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&ar
## <div class="gs_gray">Nucleic acids research 38 (12), e131-e131<span class="gs_oph">, 2010
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2010</span></td>
## </tr>
## <tr class="gsc_a_tr">
## <td class="gsc_a_t">
## <a href="/citations?view_op=view_citation&hl=en&oe=ASCII&user=hBWARYEAAAAJ&ar
## <div class="gs_gray">Proceedings of the National Academy of Sciences 99 (2), 972-977<span
## </div>
## </td>
## <td class="gsc_a_c"><a href="https://scholar.google.com/scholar?oi=bibs&hl=en&oe=
## <td class="gsc_a_y"><span class="gsc_a_h">2002</span></td>
## </tr>
## </tbody>
## </table>
## <div id="gsc_a_sp" style="display:none"></div>
## <div id="gsc_a_err" class="gs_alrt" style="display:none">The system can't perform the op
## <div id="gsc_lwp"><div id="gsc_bpf">
## <button type="button" id="gsc_bpf_prev" aria-label="Previous" disabled class="gs_btnPL gs
## </div></div>
## </form></div>
## <div class="gsc_lcl"><div id="gs_ftr" role="contentinfo">
## <div id="gsc_dscl">Dates and citation counts are estimated and are determined automaticall
## <div id="gsc_ftr_h">
## <a href="/intl/en/scholar/citations.html">Help</a> <a href="//www.google.com/intl/en/poli
## </div>
## </div></div>
## </div>
## <script>var gs_zvb;!function(u){gs_zvb=new Image();gs_zvb.onload=gs_zvb.onerror=function
## <div id="gs_rdy"></div>
## </body>
## </html>
##

```

(C) use test that function to test whether the function is correct. first, try the invalid name, and it return no erro, which means the function works for invalid scholar's name. second, try to make sure the data frame is in correct dimation with 5 columns

```

library(testthat)
test_that("content of scholar is correct?", {
  expect_error(scholar_df("abc"))
  expect_equal(dim(df_sc)[2], 5)
})

## Error: Test failed: 'content of scholar is correct?'
## * scholar_df("abc") did not throw an error.
##
## * object 'df_sc' not found
## 1: expect_equal(dim(df_sc)[2], 5) at <text>:4
## 2: compare(object, expected, ...)

```

(d) I found that if I choose load more page, the url will add 20 each time, so I put it into while loop. I assume that if the page is full(with 20 items), it might be another public next page, so I will continuously download data for the next page until the page that contains less than 20 items.

```

scholar_dff <- function(sch_name) {
  i = 0
  url <- paste("https://scholar.google.com/citations?view_op=search_authors&mauthors=",
    sch_name, "&hl=en&oi=ao")
  Sys.sleep(2)
  if (str_detect(sch_name, "[[:punct:]]") || str_detect(sch_name, "[[:digit:]]")) {
    cat("invalid scholar name")
  } else {
    page <- readLines(url)
    link <- getHTMLLinks(page)
    id <- strsplit(strsplit(link[9], split = "=")[[1]][2], split = "&")[[1]][1]
    htmlpage <- gsub(" ", "", paste("https://scholar.google.com", link[9],
      "&cstart=", i * 20, "&pagesize=20"), fixed = TRUE)

    scholar <- htmlParse(readLines(htmlpage))
    article_content <- getNodeSet(scholar, "//td[@class = 'gsc_a_t']/a")
    article <- sapply(article_content, xmlValue)
    gray_content <- getNodeSet(scholar, "//div[@class='gs_gray']")
    a_j <- sapply(gray_content, xmlValue)
    authors <- a_j[seq(1, length(a_j), 2)]
    journal <- a_j[seq(2, length(a_j), 2)]
    cit_content <- getNodeSet(scholar, "//td[@class = 'gsc_a_c']/a")
    cit <- sapply(cit_content, xmlValue)
    year_content <- getNodeSet(scholar, "//span[@class = 'gsc_a_h']")
    year <- sapply(year_content, xmlValue)[-1]
    df_sc <- data.frame(article, authors, journal, cit, year)
    while (dim(df_sc)[1] == 20 * (i + 1)) {
      i = i + 1
    }
  }
}

```

```

    print(i)
    htmlpage <- gsub(" ", "", paste("https://scholar.google.com", link[9],
    "&cstart=", i * 20, "&pagesize=20"), fixed = TRUE)
    print(htmlpage)
    scholar <- htmlParse(readLines(htmlpage))
    article_content <- getNodeSet(scholar, "//td[@class = 'gsc_a_t']/a")
    article <- sapply(article_content, xmlValue)
    gray_content <- getNodeSet(scholar, "//div[@class='gs_gray']")
    a_j <- sapply(gray_content, xmlValue)
    authors <- a_j[seq(1, length(a_j), 2)]
    journal <- a_j[seq(2, length(a_j), 2)]
    cit_content <- getNodeSet(scholar, "//td[@class = 'gsc_a_c']/a")
    cit <- sapply(cit_content, xmlValue)
    year_content <- getNodeSet(scholar, "//span[@class = 'gsc_a_h']")
    year <- sapply(year_content, xmlValue)[-1]
    dfn <- data.frame(article, authors, journal, cit, year)
    df_sc <- rbind(df_sc, dfn)
  }
  return(df_sc)
}

df_sc1 <- scholardff("Geoffrey Hinton")

## Warning in readLines(url): incomplete final line found on 'https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=20&pagesize=20'
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=20&pagesize=20'
## [1] 1
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=20&pagesize=20"
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=40&pagesize=20'
## [1] 2
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=40&pagesize=20"
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=60&pagesize=20'
## [1] 3
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=60&pagesize=20"
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=80&pagesize=20'
## [1] 4
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=80&pagesize=20"
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=100&pagesize=20'
## [1] 5
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=100&pagesize=20"

```



```

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 6
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=120&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 7
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=140&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 8
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=160&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 9
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=180&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 10
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=200&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 11
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=220&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 12
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=240&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 13
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=260&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 14
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=280&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 15
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=300&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 16
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=320&pag

```

```

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 17
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=340&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 18
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=360&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 19
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=380&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 20
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=400&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 21
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=420&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 22
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=440&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 23
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=460&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 24
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=480&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 25
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=500&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 26
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=520&pag
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
## [1] 27
## [1] "https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en&oe=ASCII&cstart=540&pag

```

```

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c
df_sc2 <- scholardff("Sandrine Dudoit")

## Warning in readLines(url): incomplete final line found on 'https://scholar.google.com/ci
Sandrine Dudoit &hl=en&oi=ao'
## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

## [1] 1
## [1] "https://scholar.google.com/citations?user=hBWARYEAAAAJ&hl=en&oe=ASCII&cstart=20&pag

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

## [1] 2
## [1] "https://scholar.google.com/citations?user=hBWARYEAAAAJ&hl=en&oe=ASCII&cstart=40&pag

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

## [1] 3
## [1] "https://scholar.google.com/citations?user=hBWARYEAAAAJ&hl=en&oe=ASCII&cstart=60&pag

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

## [1] 4
## [1] "https://scholar.google.com/citations?user=hBWARYEAAAAJ&hl=en&oe=ASCII&cstart=80&pag

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

## [1] 5
## [1] "https://scholar.google.com/citations?user=hBWARYEAAAAJ&hl=en&oe=ASCII&cstart=100&pag

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

## [1] 6
## [1] "https://scholar.google.com/citations?user=hBWARYEAAAAJ&hl=en&oe=ASCII&cstart=120&pag

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

## [1] 7
## [1] "https://scholar.google.com/citations?user=hBWARYEAAAAJ&hl=en&oe=ASCII&cstart=140&pag

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

## [1] 8
## [1] "https://scholar.google.com/citations?user=hBWARYEAAAAJ&hl=en&oe=ASCII&cstart=160&pag

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

## [1] 9
## [1] "https://scholar.google.com/citations?user=hBWARYEAAAAJ&hl=en&oe=ASCII&cstart=180&pag

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

```

```

## [1] 10
## [1] "https://scholar.google.com/citations?user=hBWARYEAAAAJ&hl=en&oe=ASCII&cstart=200&pag

## Warning in readLines(htmlpage): incomplete final line found on 'https://scholar.google.c

summary(df_sc1)

##                                     article
## System and method for parallelizing convolutional neural networks:  3
## WJ CLANCEY                                                            :  3
## Deep learning                                                         :  2
## Learning internal representations by error propagation               :  2
## Learning representations by back-propagating errors                 :  2
## Parallel distributed processing                                       :  2
## (Other)                                                              :539
##
##                                     authors
## GE Hinton                                                             : 47
## DE Rumelhart, GE Hinton, RJ Williams: 29
## G Hinton                                                             : 19
## Z Ghahramani, GE Hinton                                              : 10
## GE Hinton, RR Salakhutdinov                                          :  6
## A Krizhevsky, I Sutskever, GE Hinton:  5
## (Other)                                                              :437
##
##                                     journal
##                                     : 78
## Parallel distributed processing: Explorations in the microstructure of ..., 1986 :  6
## MIT Press, Cambridge, MA, 1986                                       :  6
## MIT Press, Cambridge, 1986                                           :  6
## MIT Press, 1986                                                       :  6
## Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International ..., 2013:  3
## (Other)                                                              :448
##
##      cit      year
##      :112      : 80
## 5      : 23   1986   : 54
## 3      : 23   2009   : 23
## 6      : 15   2010   : 20
## 2      : 15   2012   : 18
## 4      : 12   1996   : 18
## (Other):353   (Other):340

dim(df_sc2)

## [1] 207  5

```