Problem 2

After reading in the character vector text from the online traffic logs *IPs.RData*, the code below could extract out all the IP numbers. In the meanwhile, it can determine how many IP addresses are in each element of the vector with function *getIPnum*.

```
patIP <- "((\\d{1,3}\\.){3}\\d{1,3})"  #Perl style pattern for IP
getIPnum <- function(t) {
    if (length(grep(patIP, t, perl = TRUE)) == 0) {
        return(0)
    } else {
        return(length(gregexpr(patIP, t, perl = TRUE)[[1]]))
    }
}  # get number of IPs per text element
getIPidx <- function(t) {
    return(gregexpr(patIP, t, perl = TRUE)[[1]])
}  # get index of IPs in text element
getIPsub <- function(i) {
    return(substring(text[i], ipIdx[[i]], ipIdx[[i]] + attr(ipIdx[[i]], "match.length") -
        1))
}  # use idx to get IP strings
ipNum <- sapply(text, getIPnum, USE.NAMES = FALSE)  # apply to the character vector
ipIdx <- lapply(text, getIPidx)  # maintain list structure of index for each IP log
ipStr <- sapply(1:length(text), getIPsub)
# obtain IPs in list; multiple IPs per element stored in list items
```

The results of getting the number of IPs within one element of *text* is one integer vector of the length *length(text)*. All the extracted IPs are stored in the list *ipStr*, with each element of a character vector in length *ipNum[i]* of all the IPs from line *text[i]*. NA or no-IP results are treated as empty string.

```
## The results of # of IPs in each text element (1:100):

##   [1] 1 0 1 0 2 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 1 0 1 1 1 0 0 1 1
##  [36] 2 1 1 1 1 1 0 0 0 1 2 0 0 1 0 1 1 1 1 1 1 1 1 1 1 2 0 1 0 0 1 0 0 0 0
##  [71] 0 2 1 1 0 1 1 1 0 0 1 0 1 1 0 1 2 1 1 2 1 0 1 0 1 1 1 1 1 1

## The first 10 lines of results:

## $`from munnari.OZ.AU (localhost [127.0.0.1]) by delta.cs.mu.OZ.AU
## (8.11.6/8.11.6) with ESMTP id g7MBQPW13260; Thu, 22 Aug 2002 18:26:25
## +0700 (ICT)`
## [1] "127.0.0.1"
##
## $`from SpoolDir by EMS-SRV0 (Mercury 1.44); 22 Aug 02 14:50:31 +0000`
## [1] ""
##
## $`from ee.ed.ac.uk (sxs@dunblane [129.215.34.86]) by
## postbox.ee.ed.ac.uk (8.11.0/8.11.0) with ESMTP id g7ME1Li02942 for
## <forteana@yahoogroups.com>; Thu, 22 Aug 2002 15:01:21 +0100 (BST)`
## [1] "129.215.34.86"
##
## $`from SpoolDir by EMS-SRV0 (Mercury 1.44); 22 Aug 02 15:01:34 +0000`
## [1] ""
##
```

```
## $`from [192.168.0.4] (chello062178142216.4.14.vie.surfer.at
## [62.178.142.216]) (authenticated bits=0) by mail.uptime.at (8.12.5/8.12.5)
## with ESMTP id g7MEI7Vp022036 for
## <spamassassin-devel@lists.sourceforge.net>; Thu, 22 Aug 2002 16:18:07
## +0200`
## [1] "192.168.0.4"    "62.178.142.216"
##
## $`from m206-56.dsl.tsoft.com ([198.144.206.56] helo=perkel.com) by
## darwin.ctyme.com with asmtp (TLSv1:RC4-MD5:128) (Exim 3.35 #1) id
## 17htgP-0004te-00; Thu, 22 Aug 2002 08:15:37 -0700`
## [1] "198.144.206.56"
##
## $`by jlooney.jinny.ie (Postfix, from userid 500) id 4F57189D;
## Thu, 22 Aug 2002 16:25:45 +0100 (IST)`
## [1] ""
##
## $`from dcu.ie (136.206.21.115) by hawk.dcu.ie (6.0.040) id
## 3D6203D3000136AD for iiu@taint.org; Thu, 22 Aug 2002 16:59:17 +0100`
## [1] "136.206.21.115"
##
## $`from [66.218.67.174] by n19.grp.scd.yahoo.com with NNFMP;
## 22 Aug 2002 16:11:27 -0000`
## [1] "66.218.67.174"
##
## $`from [66.218.67.189] by n10.grp.scd.yahoo.com with NNFMP;
## 22 Aug 2002 16:17:40 -0000`
## [1] "66.218.67.189"
```