Geocaching in Israel

Elisa Lerner

25 September 2015

Introduction

Have you ever been geocaching? Have you even heard of geocaching? Well, over twenty million people over the world spend part of their leisure time hiding caches and looking for caches that other players have hidden. Once a cache has been hidden, the longitude and latitude coordinates are uploaded to a geocaching site and happy seekers use GPS applications on their mobile phones to see how close they are. For further details - Wikipedia
I was curious to find out what is the distribution of geocaches around Israel.
It is possible to find free online information about many geocaches, though some are reserved for 'premium' paying members. Files containing information about a number of geocaches may be downloaded only by premium members. The files are GPX files, which is a form of XML file containing GPS data and can be read into R using the package "plotKML".
I ran a search for all geocaches hidden within Israel up to 25/09/2015 and downloaded the search results. I actually ran two searches, because each search is limited to a maximum of 1000 caches. The archive 16648504.zip contains GPX files with information about caches hidden up to the end of 2013, and the archive 16648512.zip, from the beginning of 2014 until 25/09/2015.

Loading the data into R

```
library(plotKML)
```

```
## Warning: package 'plotKML' was built under R version 3.2.2
```

```
## plotKML version 0.5-3 (2015-07-07)
## URL: http://plotkml.r-forge.r-project.org/
```

```
unzip("16648504.zip")
unzip("16648512.zip")
gc2013 <- readGPX("16648504.gpx")
gc2015 <- readGPX("16648512.gpx")
```

Each gpx file contains a list. The first element in the list is metadata, for example:

```
gc2013[[1]]
```

```
##   name                                  desc      author
## 1  gc1 Geocache file generated by Groundspeak Groundspeak
##                      email                              url
## 1 contact@groundspeak.com 2015-09-26T19:53:41.6524676Z
```

The second element shows the maximum and minimum GPS coordinates of all the caches found.

```
gc2013[[2]]
```

```
##          min     max
## lat 34.40827 35.8325
## lon 29.51052 33.2754
```

```
gc2015[[2]]
```

```
##          min      max
## lat 34.45238 35.83715
## lon 29.54442 33.19480
```

The third element is a data frame containing the details of the caches.

Processing data before analysis

Since the two files were separated because of the download constraint, the details from the two files were merged into one data frame.

```
gc <- merge(gc2013[[3]],gc2015[[3]],by=names(gc2013[[3]]),all=TRUE)
```

The number of rows in the new data frame is 1313 - the total number of geocaches in Israel, and the number of columns 10 - the number of variables describing the caches.

Variables

The first column 'lon' contains the longitudes of the caches; the second column 'lat' contains the latitudes. The third column is called 'time'. This is actually the date on which the cache was placed. The dates need to be transformed into an appropriate form.

```
l <- nrow(gc)
date <- gc$time
date <- gsub("T(.*)","",date)
date <-as.Date(date)
gc[,3] <- date
names(gc)[3] <- "date"
```

The fourth variable 'name' is the code name of the cache. The fifth variable 'desc' is a short description of the name of the cache, who it was placed by, the type of cache , the difficulty level of finding it and the level of difficulty of the terrain. These parameters need to be separated into separate variables, but since some of this information is duplicated in other variables, only new information is retrieved and joined to the data frame.

```
props <- gc$desc
by <- vector("character",l)
for (i in 1:l){ temp1 <- strsplit(props[i],"by");
                temp2<-strsplit(temp1[[1]][2],",");
                by[i] <- temp2[[1]][1]
                }
```

```
d <- vector("character",l)
t <- vector("character",l)

level <- gsub("(.*),([a-z A-Z]*)\\(","",props)
level <- gsub("\\)","",level)

for (i in 1:l){ dt <- strsplit(level[i],"/");
                d[i]<- dt[[1]][1];t[i]<-dt[[1]][2]
                }
difficulty <- as.numeric(d)
terrain <- as.numeric(t)

gc$by <- by
gc$difficulty <- difficulty
gc$terrain <- terrain
```

The "by" variable, holds names of the cachers who placed the geocaches. I thought it might be interesting to look at the level of activity of each of these "hiders",ie who hid the most caches, etc, but looking at the contents of the variable I found that sometimes they cooperated and hid caches together. The following chunk of code separates them all out and determines the frequency of caches hidden by each cacher. Obviously the total number hidden is greater than the total number of caches because some are counted more than once.

```
cachers <- gc$by
cachers <- gsub("-"," ",cachers)
cachers <- gsub("maintained","",cachers)
cachers <- gsub("the help of","",cachers)
cachers <- gsub("\\(","",cachers)
cachers <- gsub("\\)","",cachers)
cachers <- gsub("in memory of somloci","",cachers)
cachers <- gsub("Team","",cachers)
cachers <- gsub("freinds","friends",cachers)
cachers <- gsub("2 from","",cachers)
cachers<- strsplit(cachers,"\\& | and | \\+ | with")
cachers <- toupper(unlist(cachers))
cachers <- trimws(cachers)
counts <- table(cachers)
topCachers <- counts[counts>10]
```

The next two variables 'url' and 'urlname' are self explanatory. The variable 'sym' has the same entry Geocache for each cache, and the variable 'type' descibes the type of cache as one of the following types:

```
unique(gc$type)
```

```
## [1] "Geocache|Traditional Cache"        "Geocache|Unknown Cache"
## [3] "Geocache|Multi-cache"              "Geocache|Letterbox Hybrid"
## [5] "Geocache|Wherigo Cache"            "Geocache|Virtual Cache"
## [7] "Geocache|Earthcache"               "Geocache|Cache In Trash Out Event"
```

This webpage characterizes the different types of cache.

The tenth column, 'cache', duplicates several of the descriptive variables already mentioned, but also contains information about the size of the cache, which needs to be extracted, as well as alot of text containing

messages left by geocachers who seeked the cache.Definitions of the size categories in this link.
Although the messages are useful for seekers, I was interested in counting the number of seekers who actually looked for the cache, as a measure of popularity, and the proportion who succeeded in finding it. This is assuming that every seeker did actually leave a message! However, after examining the file I discovered that a maximum of 5 messages (logs) can be downloaded in the gpx file. These appear to be the last 5 logs, ie the last 5 seekers of the particular cache, so that the counts cannot represent a level of popularity, but the number of finds out of the recent seeks can be assumed to indicate a sort of recent success rate.

```
size <- vector("character",l)
micro <- grep("CacheMicro",gc[,10])
small <- grep("CacheSmall",gc[,10])
regular <- grep("CacheRegular",gc[,10])
large <- grep("CacheLarge",gc[,10])
other <- grep("CacheOther",gc[,10])
size[small] <- "small"
size[micro] <- "micro"
size[regular] <- "regular"
size[large] <- "large"
size[other] <- "other"

gc$size <- size

#The number of seekers: each log by a seeker is preceded by a date (in character format).
#Splitting the text at each date results in a list where the first entry is
#general data and the rest are logs made by seekers. The number of times
#"found  it" or "TFTC" appears in the logs is the number of seekers who found the cache
#(or at least reported they found it).
seeked <- vector("numeric",l)
found <- rep(0,l)
for (i in 1:l){
        infoSplit <- strsplit(gc[i,10],"20[0-9][0-9]-[0-9][0-9]-[0-9][0-9]T")
        logs <- infoSplit[[1]][2:length(infoSplit[[1]])] #ignore first element
        seeked[i] <- length(logs)
        found[i] <- sum(grepl("TFTC",logs) | grepl("found it",logs,ignore.case=TRUE))
}

gc$seeked <- seeked
gc$found <- found
```

Exploratory Results

1. Most of the caches to be found in Israel are Traditional Caches.

```
table(gc$type)
```

```
##
## Geocache|Cache In Trash Out Event              Geocache|Earthcache
##                                 1                                12
##         Geocache|Letterbox Hybrid             Geocache|Multi-cache
##                                 5                                68
##        Geocache|Traditional Cache           Geocache|Unknown Cache
##                              1122                               102
```

```
##            Geocache|Virtual Cache            Geocache|Wherigo Cache
##                                 1                                 2
```

2. The oldest/newest caches in Israel are:

```
gc[which.min(gc$date),1:5]
```

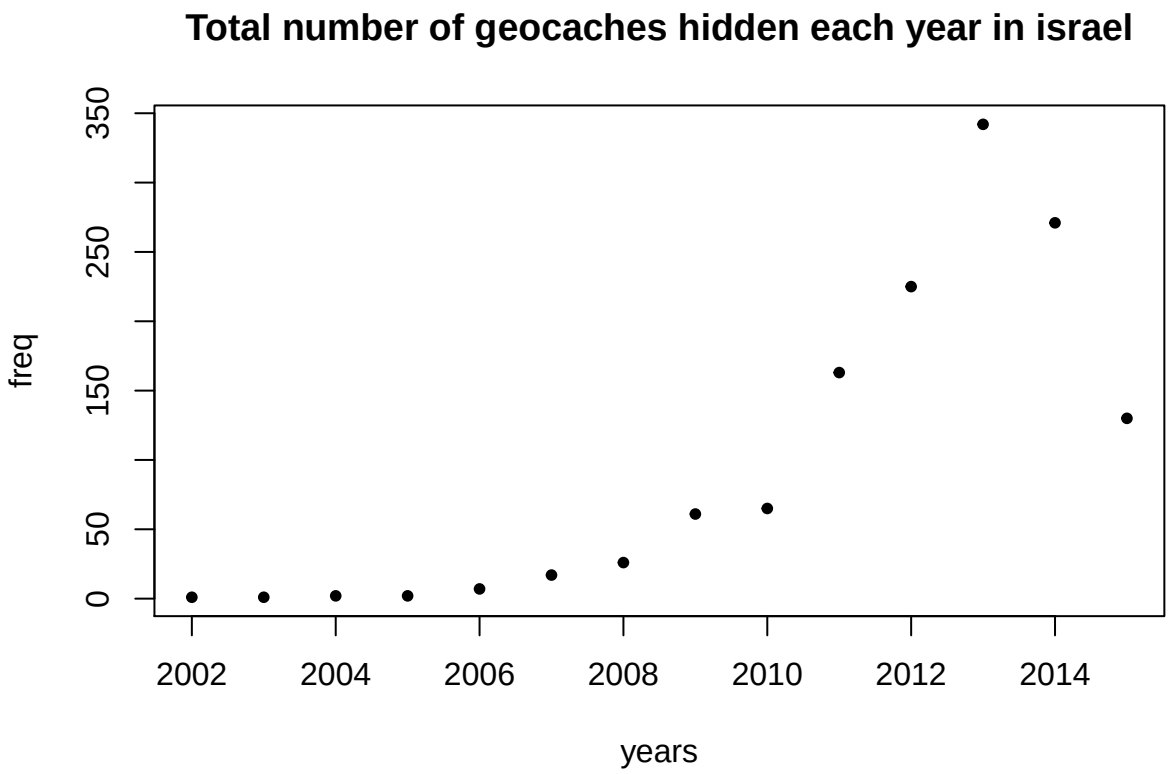
```
##          lon      lat       date   name
## 194 34.80343 32.17005 2002-04-14 GC4CEC
##                                                    desc
## 194 20 Years (Isreal) by GekoTeam, Virtual Cache (1/1)
```

```
gc[which.max(gc$date),1:5]
```

```
##          lon      lat       date    name
## 540 34.96327 32.65205 2015-09-15 GC63H1J
##                                                              desc
## 540 Ido Cohen Lookout by tifferets, Traditional Cache (2.5/2.5)
```

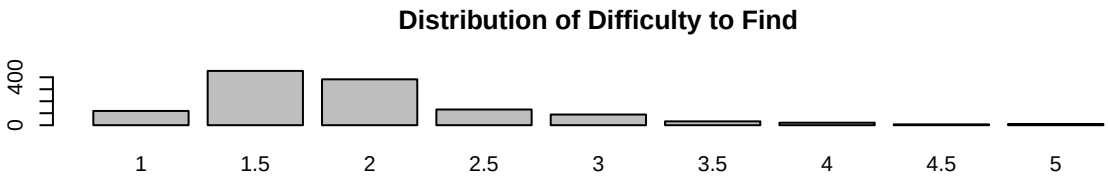
3. The number of caches hidden each year increased from 2002 until 2013, and has since decreased.

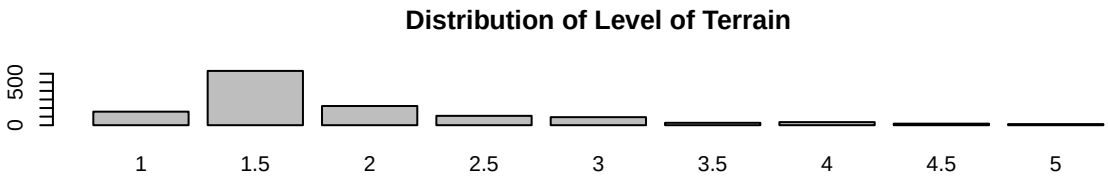
```
years <- 2002:2015
freq<- vector("numeric",14)
for (i in seq_along(years)){
        freq[i]<- sum(grepl(as.character(2001+i),gc$date))
}
plot(years,freq,main="Total number of geocaches hidden each year in israel",pch=20)
```

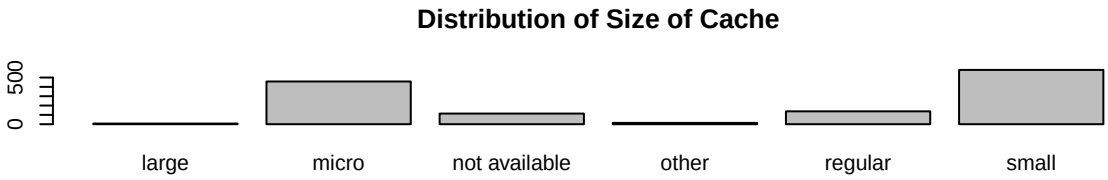


(It should be noted that 2015 is not yet over and the frequency observed is not the final count for this year.)
4. The distributions of the level of difficulty to find, the level of terrain, and cache size are given in the following plots:

```
par(mfrow=c(3,1))
plot(as.factor(gc$difficulty),main="Distribution of Difficulty to Find")
plot(as.factor(gc$terrain),main="Distribution of Level of Terrain")
for (i in 1:l){if (gc$size[i] =="")gc$size[i]<-"not available"}
plot(as.factor(gc$size),main="Distribution of Size of Cache")
```







It is clear that most difficulty and terrian levels are low (not difficult) and most caches are micro or small.

5. The distributions of recent seeks and finds are:

```
table(gc$seeked)
```

```
##
##    1    2    3    4    5
##   10   29   44   55 1175
```

```
table(gc$found)
```

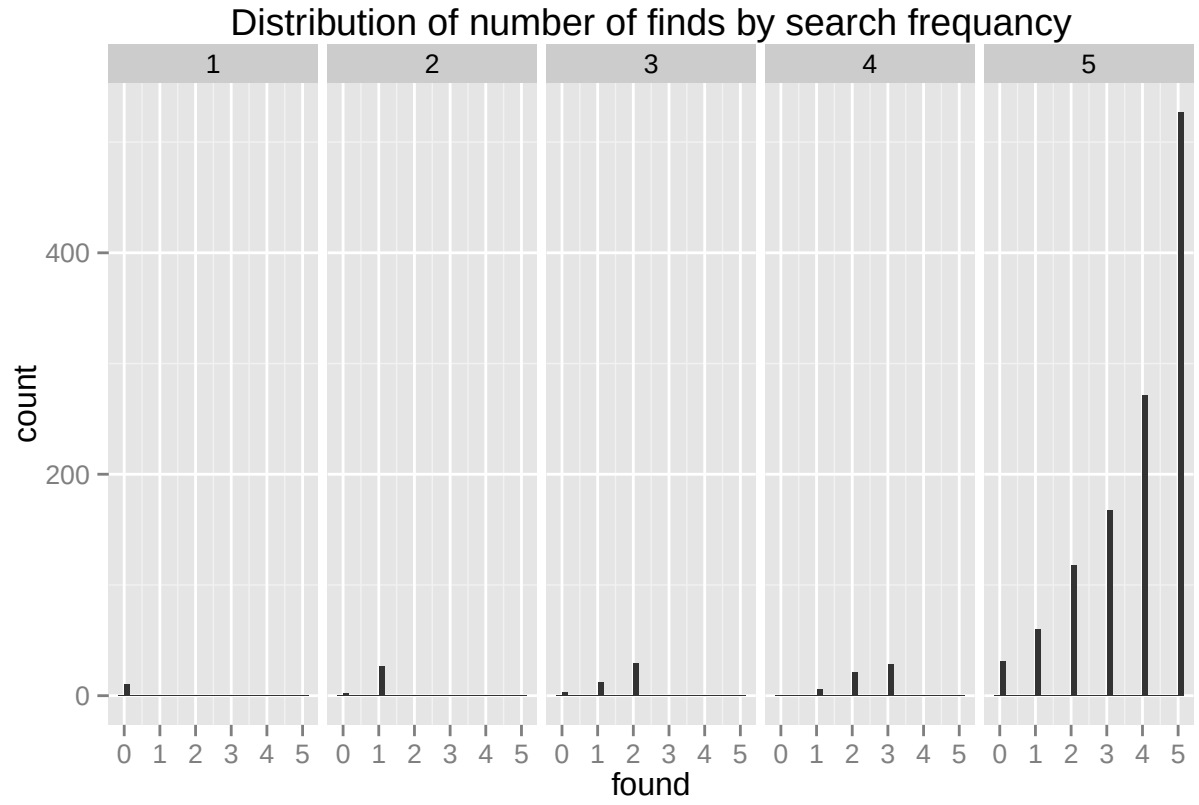
```
##
##   0   1   2   3   4   5
##  46 105 168 196 271 527
```

There are apparently no caches that have never been searched for, and 89.5% of the caches have been searched for at least 5 times.
The number of finds for each frequency of searches (out of the last 5 searches) is shown as follows:

```
library(ggplot2)
qplot(found,data=gc,facets=.~seeked,main="Distribution of number of finds by search frequancy")
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



In summary, caches that were sought once were never found, caches that were sought twice were almost all found once. Caches that were sought three times were mostly found twice, sometimes once and rarely not found at all. Caches that were sought four times were usually found by two or three of the seekers, and rarely by one seeker. Many of the caches that were sought by five or more seekers were found five or more times, and with decreasing frequencies were found four or three times or twice or once, or not at all. In general, the number of successful searches increases with the number of searches made.

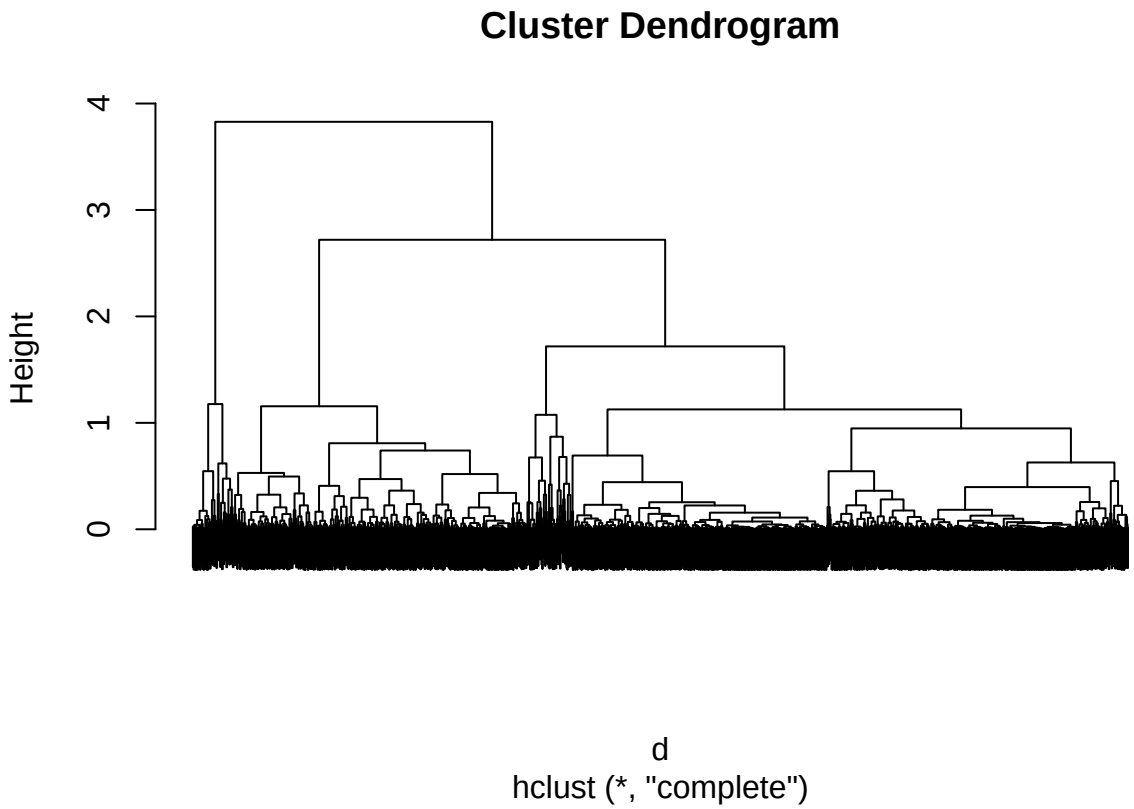
6.The twenty most active cache hiders are:

```
sort(topCachers, decreasing =TRUE)
```

```
## cachers
##      SHAISHOO         AVI.N       YARIVES  MAVERICK1234    ISRACACHER
##           130           125           118            91            82
##        JDEVOR  THEDAVIDSONS         SSABW     TIFFERETS      LIBIKLIL
##            76            57            45            37            35
##    AMNONYMOUS        IRISVO       YANIVRA        WINDOW        TAMARS
##            32            24            19            17            16
##      AVIRAM_C     YISHAYSHO       DUDIG19     RICKISMOM WALDLÄUFER123
##            15            14            12            12            11
```

7. A look at the map of Geocaches in Israel shows that there are regions with many geocaches and other regions with none. Hierarchical clustering using Euclidean distances between caches produced the following dendogram.

```
d <- dist(gc[,1:2])
par(mfrow=c(1,1))
hc <- hclust(d)
plot(hc,labels=FALSE)
```



Cutting off at 16 clusters yields the following clusters (listing 6 caches in each cluster):

```
groups <- cutree(hc,16)
gc$groups <- groups
table(gc$groups)
```

```
##
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16
##  12   8 146  23  13 355  35 132  17 251 117  13  26 111   3  51
```

```
for (i in 1:16){
        print(i)
        print(head(gc$desc[gc$groups==i]))
        print("")
}
```

```
## [1] 1
## [1] "Peace Sculpture by Isracacher, Traditional Cache (1.5/1.5)"
## [2] "Caving to the Cache by Isracacher, Traditional Cache (2/2)"
## [3] "Horev Operation Memorial by Isracacher, Traditional Cache (1.5/1)"
## [4] "Small and Colorful by lessbink, Traditional Cache (1.5/1.5)"
## [5] "Ein Avdat by avi.n, Traditional Cache (1.5/1.5)"
## [6] "Ben Gurion Memorial by Isracacher, Traditional Cache (1/1.5)"
## [1] ""
## [1] 2
## [1] "Mizpe Gvoulot by jokotik, Traditional Cache (1.5/1.5)"
## [2] "Old Be'eri - Nahabir by a_nave, Traditional Cache (2/2)"
## [3] "Ilan's Bar Mitzvah Cache by Isracacher, Traditional Cache (3/3)"
## [4] "Suspension bridge by yanivra, Traditional Cache (1/1.5)"
## [5] "Lookout! by SheepOnTrip, Traditional Cache (1/1.5)"
## [6] "Mefalsim by susgor, Traditional Cache (1.5/1.5)"
## [1] ""
## [1] 3
## [1] "Tel Ashkelon by Isracacher, Unknown Cache (2.5/1.5)"
## [2] "Abba Hillel Silver - farm by libiklil, Traditional Cache (3/1.5)"
## [3] "Ashdod Park Yam by borntocache, Traditional Cache (2/1.5)"
## [4] "Park Ad-Halom by Alan_Yana_Lin_Geela, Traditional Cache (2/1.5)"
## [5] "zohar memorial Cache by yariminal, Traditional Cache (3/1.5)"
## [6] "Palmachim South Sunset by waldläufer123, Traditional Cache (2/2)"
## [1] ""
## [1] 4
## [1] "My First Stargazing by Nagasaki45, Traditional Cache (2.5/2.5)"
## [2] "White Canyon by itzix, Traditional Cache (1.5/4.5)"
## [3] "Mount Karkom by yanivra, Traditional Cache (1.5/4)"
## [4] "Alpaca Farm by avi.n, Traditional Cache (1.5/1.5)"
## [5] "Ramon Crater Rim Trail by TheDavidsons, Traditional Cache (2.5/3)"
## [6] "Haminsara - The Carpentry by TheDavidsons, Traditional Cache (1.5/2)"
## [1] ""
## [1] 5
## [1] "Sculpture Garden by Isracacher, Traditional Cache (1.5/1.5)"
## [2] "Geographical midpoint of Israel by ssabw, Traditional Cache (2/2.5)"
## [3] "Fighters by AlexeyG, Traditional Cache (1.5/1.5)"
## [4] "West of Pura - Shikma River by Nehora Family, Traditional Cache (2/2)"
## [5] "Lone Scouts at Lake Shiqma by Lone Scouts Jerusalem, Traditional Cache (1/2.5)"
## [6] "First Beer Sheva Cache by TelAvivGhost & ssabw, Traditional Cache (1.5/1.5)"
## [1] ""
## [1] 6
## [1] "Jaffa Port Voyages by TheDavidsons, Unknown Cache (3/1.5)"
## [2] "Three Wishes by Irisvo, Traditional Cache (1.5/1.5)"
## [3] "The Dutch Conundrum by shaishoo, Unknown Cache (2.5/1)"
## [4] "Puzzled with Hebrew by TheDavidsons, Unknown Cache (3/2)"
## [5] "Some Hebrew Facts by TheDavidsons, Unknown Cache (3.5/1.5)"
## [6] "APC St. in Tel-Aviv(<U+05E8><U+05D7><U+05D5><U+05D1> <U+05D0><U+05E4><U+05E7> <U+05E4><U+05D9><U
## [1] ""
## [1] 7
## [1] "Mt. Neshef, Eilat Mountains by Isracacher, Traditional Cache (1.5/3.5)"
## [2] "Netafim Spring, Eilat Mountains by Isracacher, Traditional Cache (2/3)"
## [3] "Red Canyon by ssabw, Traditional Cache (2.5/3)"
## [4] "Shvil Israel S-2-N. Day #1. Send me a Sign! by Yarives, Traditional Cache (3/5)"
## [5] "Mt.Yehoram Look-Out, Eilat Mountains by Isracacher, Traditional Cache (1.5/3)"
```

```
## [6] "Mt. Shlomo, Eilat Mountains by Isracacher, Multi-cache (3/4)"
## [1] ""
## [1] 8
## [1] "Feinberg Memorial by tifferets, Traditional Cache (1/2)"
## [2] "Ogen Ha-Pal Yam Memorial Site by fb158, Traditional Cache (1/1.5)"
## [3] "Flavios Strategios Road by tifferets, Traditional Cache (2/1.5)"
## [4] "Caesarea Aqueduct by tifferets, Traditional Cache (2/2)"
## [5] "Heftziba Farm by avi.n, Traditional Cache (1.5/1.5)"
## [6] "Bird Mosaic by tifferets, Traditional Cache (3/2.5)"
## [1] ""
## [1] 9
## [1] "Yeruham park view by shaishoo, Traditional Cache (2.5/2.5)"
## [2] "Beer Yeruham by ssabw, Traditional Cache (1.5/1.5)"
## [3] "Shvil Israel S-2-N. Day #13 Electric Company by Yarives, Traditional Cache (3/4)"
## [4] "Shvil Israel S-2-N. Day #14 The Karbolet Mountain by Yarives, Traditional Cache (5/5)"
## [5] "Colored Sands, Large Crater by Isracacher, Traditional Cache (2/1)"
## [6] "Mount Zin by yanivra, Traditional Cache (2/4)"
## [1] ""
## [1] 10
## [1] "Bet Guvrin: Bell Caves by klippdachs, Earthcache (2/1.5)"
## [2] "Goded Hill by sleepless_knight & Avishai, Traditional Cache (2/3)"
## [3] "crawling at Tel Goded by tamush, Traditional Cache (1/1)"
## [4] "Mitzpe Massua by Isracacher, Traditional Cache (1.5/1.5)"
## [5] "Only On Saturday  by Isracacher, Traditional Cache (2/2)"
## [6] "Leora and Uri's Birthday cache by shaishoo, Traditional Cache (2/1.5)"
## [1] ""
## [1] 11
## [1] "cache med by maverick1234, Traditional Cache (3.5/5)"
## [2] "Siah Wadi 3/3 - The Siah Spring by idanpl, Traditional Cache (3/3)"
## [3] "Siah Wadi 2/3 - The Monastery  by idanpl, Traditional Cache (1.5/3)"
## [4] "sewage-less sewerage by Arbel's, Traditional Cache (2.5/4)"
## [5] "Hills of Ramat Begin by mnahon, Traditional Cache (2.5/3.5)"
## [6] "Helsinki Rock Star by Gamburger, Traditional Cache (3.5/1.5)"
## [1] ""
## [1] 12
## [1] "Shvil Israel S-2-N. Day #18 A Tree @ Gev Kina by Yarives, Traditional Cache (4/5)"
## [2] "Shvil Israel S-2-N. Day #17 A Tree @ Nahal Kina by Yarives, Traditional Cache (3/4)"
## [3] "Masada view by D@ewo, Traditional Cache (1.5/3)"
## [4] "Iron at the Dead Sea by ebagpsil, Traditional Cache (1/1)"
## [5] "Dead Sea Resorts by G-2--maintained by ssabw, Traditional Cache (1.5/1.5)"
## [6] "THE DEAD SEA - EARTH'S LOWEST ELEVATION ON LAND by Team Ferreira, Earthcache (1.5/1.5)"
## [1] ""
## [1] 13
## [1] "03 Lookout by Yarives, Traditional Cache (1/1.5)"
## [2] "Tip of the Lake Park by Yarives, Traditional Cache (1.5/1.5)"
## [3] "Israel Wineries - #7 by OKSYcachers & jillycash, Traditional Cache (1.5/1.5)"
## [4] "Maale Adumim Water Tower by Yarives, Traditional Cache (1.5/1.5)"
## [5] "Ednas Garden by Yarives, Traditional Cache (2/2)"
## [6] "Chagai Lookout by Yarives, Traditional Cache (1/1)"
## [1] ""
## [1] 14
## [1] "Ein Yarket by yanivra, Traditional Cache (1/2.5)"
## [2] "Yarom's Huta by Yarives, Traditional Cache (2/2)"
## [3] "Mokas' Cache by Yarives, Traditional Cache (1/2.5)"
```

```
## [4] "Touching The Clouds @ Adir Mountain by Yarives, Traditional Cache (2/2)"
## [5] "Eviatar's Cache Day 3 . Sea 2 Sea Trek by Yarives, Traditional Cache (1.5/1.5)"
## [6] "Ein Neria by Yarives, Traditional Cache (2.5/2.5)"
## [1] ""
## [1] 15
## [1] "Three-Seas Lookout by Hamavreg, Traditional Cache (1.5/2)"
## [2] "The 4 Elements \"Challenge\" by magals90, Unknown Cache (2.5/1.5)"
## [3] "The WRECK TANK by magals90, Traditional Cache (1.5/1.5)"
## [1] ""
## [1] 16
## [1] "Mount Turan by aviram_c, Traditional Cache (1.5/2)"
## [2] "Day # 10 - In Tavor's Shadow by The Shulem Family, Traditional Cache (4/2)"
## [3] "Mt. Tabor by Isracacher, Traditional Cache (2/2)"
## [4] "Nahal Harod number 5 by Avantland, Traditional Cache (1.5/1.5)"
## [5] "The Irisim by maverick1234, Traditional Cache (2/1.5)"
## [6] "yitspor canyon entrance by maverick1234, Traditional Cache (3.5/1.5)"
## [1] ""
```