PreCheck Donkey data - Judas efficiency

Carlo Pacioni

Load libraries

```
# Load libraries
options(java.parameters = "-Xmx8024m")
library(XLConnect, quietly = T)
```

```
## XLConnect 0.2-15 by Mirai Solutions GmbH [aut],
##   Martin Studer [cre],
##   The Apache Software Foundation [ctb, cph] (Apache POI),
##   Graph Builder [ctb, cph] (Curvesapi Java library)
```

```
## http://www.mirai-solutions.com
## https://github.com/miraisolutions/xlconnect
```

```
library(ggplot2, quietly = T)
library(data.table, quietly = T)
```

Read data

```
data.path <- "../Data/"
judas.master <- data.table(readWorksheetFromFile(
  file = file.path(data.path, "Tracking_History_Judas_with habitat_MZ.xlsx"),
                       sheet="Judas_Tracking_History"))
names(judas.master)
```

```
##  [1] "Tracking_ID"  "REGION"       "SHIRE"        "AREA"
##  [5] "JUDAS_ID"     "EVENT_DATE"   "LAT"          "LONG"
##  [9] "SEARCHED"     "EVENT_ID"     "EVENT"        "N_FERALS"
## [13] "N_JUDAS"      "ACTION"       "Habitat.Type" "LAT_ORG"
## [17] "LONG_ORG"
```

Check formatting

```
sapply(judas.master, class)
```

```
## $Tracking_ID
## [1] "numeric"
##
## $REGION
## [1] "character"
##
## $SHIRE
## [1] "character"
##
## $AREA
## [1] "character"
##
## $JUDAS_ID
## [1] "character"
##
## $EVENT_DATE
## [1] "POSIXct" "POSIXt"
##
```

```
## $LAT
## [1] "numeric"
##
## $LONG
## [1] "character"
##
## $SEARCHED
## [1] "character"
##
## $EVENT_ID
## [1] "numeric"
##
## $EVENT
## [1] "character"
##
## $N_FERALS
## [1] "numeric"
##
## $N_JUDAS
## [1] "numeric"
##
## $ACTION
## [1] "character"
##
## $Habitat.Type
## [1] "character"
##
## $LAT_ORG
## [1] "numeric"
##
## $LONG_ORG
## [1] "character"
```

LONG is character while it should be numeric. Checked what values are not numeric.

```
judas.master[is.na(sapply(judas.master[, LONG], as.numeric)), ]
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
##    Tracking_ID
## 1:       20149
## 2:       24575
##
## 1: KIMBERLEY
## 2: KIMBERLEY
##    SHIRE        AREA JUDAS_ID EVENT_DATE       LAT LONG SEARCHED EVENT_ID
## 1:    HC Mabel Downs     MD03 2003-08-15 -17.20017 NULL      YES    16788
## 2:    HC   Nicholson    NIC06 2003-11-08 -17.01800 NULL      YES    20025
##    EVENT N_FERALS N_JUDAS ACTION Habitat.Type   LAT_ORG LONG_ORG
## 1: FERAL        8       1  FREED         NULL -17.20017     NULL
## 2: ALONE        0       1  FREED         NULL -17.01800     NULL
```

```
# Number dead
judas.master[, sum(ACTION == "DEAD")]
```

```
## [1] 7
```

```
# Number collared
judas.master[, sum(ACTION == "COLLARED")]
```

```
## [1] 1721
```

```
# Number of judas in DB
judas.master[, length(unique(JUDAS_ID))]
```

```
## [1] 1721
```

```
# Rm useless spaces in REGION
judas.master[, REGION := sub(pattern = " +", replacement = "", x = REGION)]
# Number by region/shire
judas.master[, .(njudas=length(unique(JUDAS_ID)),
                 start.date=min(EVENT_DATE), end.date=max(EVENT_DATE)), by=REGION]
```

```
##       REGION njudas start.date   end.date
## 1: KIMBERLEY   1385 1994-07-12 2017-11-14
## 2:   PILBARA    336 1998-05-20 2017-11-03
```

```
judas.master[, .(njudas=length(unique(JUDAS_ID)),
                 start.date=min(EVENT_DATE), end.date=max(EVENT_DATE)),
                 by=c("REGION", "SHIRE")]
```

```
##       REGION SHIRE njudas start.date   end.date
## 1: KIMBERLEY    EK    513 1997-10-21 2017-11-14
## 2: KIMBERLEY    WK    451 1994-07-12 2017-06-20
## 3:   PILBARA    PB    336 1998-05-20 2017-11-03
## 4: KIMBERLEY    HC    421 1995-10-26 2017-11-14
```

Check whether there are animals that moved between shires during the study (it seems none)

```
# build a function where IDbyLoc is a data.table with one col reporting locations
   # (in this case shire), and the second, named 'IDs', is the judas' ids for each location
   # location is a character vector of with the heading of the column reporting the location
   # (here: location<- "SHIRE")
check.migration <- function(IDbyLoc, location) {
  areas <- unique(IDbyLoc[[location]])
  nareas <- length(areas)
  un.IDs <- IDbyLoc[, unique(IDs)]
  setkeyv(IDbyLoc, location)
  l2 <- vector(mode = "list", length = nareas)
  for (a in areas) {
    l2[[a]] <- vector(length = length(un.IDs))
    names(l2[[a]]) <- un.IDs
    for (u in un.IDs) {
      l2[[a]][u] <- sum(IDbyLoc[a, IDs] %in% un.IDs[u]) > 0
    }
  }
  names(l2) <- areas
  d <- do.call(cbind, args = l2)
  keep.row <- apply(d, MARGIN = 1, sum)
  d <- cbind(un.IDs, data.table(d))
  keep.row <- keep.row > 1
  return(d[keep.row,])
}
```

```
# Create IDbyLoc
IDbyShire <- judas.master[, .(IDs=unique(JUDAS_ID)), by=SHIRE]
check.migration(IDbyShire, location = "SHIRE")
```

```
## Empty data.table (0 rows) of 5 cols: un.IDs,NA,NA,NA,NA
```

Clean up and generating variables

```
# Rm found dead because they are out of the program
judas.cleaned <- judas.master[ACTION != "DEAD", ]

# Rm not found because they do not contribute
judas.cleaned <- judas.cleaned[ACTION != "NONE", ]

# Cross check totals are still the same
judas.cleaned[, .(njudas=length(unique(JUDAS_ID)),
                  start.date=min(EVENT_DATE), end.date=max(EVENT_DATE)),
              by=c("REGION", "SHIRE")]
```

```
##       REGION SHIRE njudas start.date   end.date
## 1: KIMBERLEY    EK    513 1997-10-21 2017-11-14
## 2: KIMBERLEY    WK    451 1994-07-12 2017-06-20
## 3:   PILBARA    PB    336 1998-05-20 2017-11-03
## 4: KIMBERLEY    HC    421 1995-10-26 2017-11-14
```

```
# Rm judas with no long
keep <- !is.na(sapply(judas.cleaned[, LONG], as.numeric))
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
judas.cleaned <- judas.cleaned[keep, ]
judas.cleaned[is.na(sapply(judas.cleaned[, LONG], as.numeric)), ]
```

```
## Empty data.table (0 rows) of 17 cols: Tracking_ID,REGION,SHIRE,AREA,JUDAS_ID,EVENT_DATE...
```

```
judas.cleaned[, LONG := as.numeric(LONG)]

# Set start and date
judas.cleaned[, start.date := min(EVENT_DATE), by=JUDAS_ID]
judas.cleaned[, end.date := max(EVENT_DATE), by=JUDAS_ID]

# Cross check start.date matches collared
judas.cleaned[ACTION == "COLLARED", date.coll := EVENT_DATE, by=JUDAS_ID]
judas.cleaned[ACTION == "COLLARED", sum(start.date != date.coll, na.rm = T)]
```

```
## [1] 3
```

```
judas.cleaned[start.date != date.coll, ]
```

```
##    Tracking_ID    REGION SHIRE        AREA JUDAS_ID EVENT_DATE       LAT
## 1:       18898 KIMBERLEY    WK     Leopold     LP11 1996-07-16 -17.58333
## 2:       19091 KIMBERLEY    WK     Leopold     LP22 2006-05-12 -17.81228
## 3:       19862 KIMBERLEY    WK Mt Anderson     MA10 1996-10-22 -17.99817
##        LONG SEARCHED EVENT_ID EVENT N_FERALS N_JUDAS   ACTION Habitat.Type
## 1: 125.2833      YES    15886 ALONE        0       1 COLLARED       HLLEWS
## 2: 125.4543      YES    16024 ALONE        0       1 COLLARED       UPLEWG
```

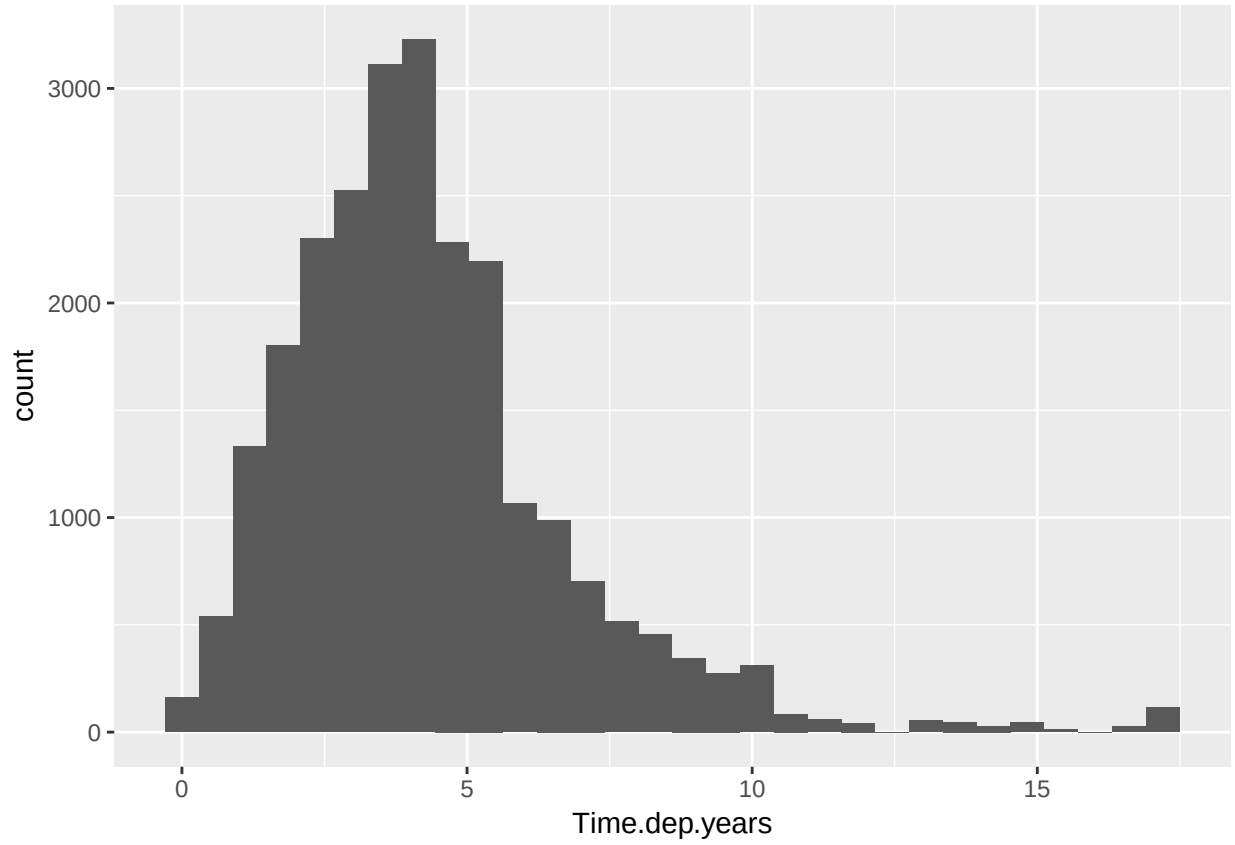
```
## 3: 123.8610      YES    16550 ALONE        0       1 COLLARED       SADPWG
##      LAT_ORG    LONG_ORG start.date   end.date  date.coll
## 1: -17.58333 125.2833333 1996-03-18 2000-12-12 1996-07-16
## 2: -17.81228   125.45425 1996-08-30 2010-09-16 2006-05-12
## 3: -17.99817     123.861 1996-07-17 1997-11-24 1996-10-22
```

```
# Rm date.coll
judas.cleaned[, date.coll := NULL]

# Length in the program
judas.cleaned[, Time.deployment := difftime(end.date, start.date, units="weeks")]
judas.cleaned[, Time.dep.years := round(as.numeric(Time.deployment) / 52, 2)]

ggplot(judas.cleaned, aes(Time.dep.years)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
summary(judas.cleaned[, Time.dep.years])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   2.680   3.940   4.313   5.270  17.200
```

Some animals have end.date set on the day they have been collared (listed below). I'm assumed that these were animals that were collared on start.date and searched and not found, is this correct? Some other aniamls were tracked for several years, it is correct?

```
# List of animals that have been deployed for 0
judas.cleaned[Time.deployment == 0, .SD,
              .SDcols=c("SHIRE", "JUDAS_ID", "EVENT_DATE", "EVENT_ID", "EVENT",
```

```
                        "ACTION")]
```

```
##     SHIRE JUDAS_ID EVENT_DATE EVENT_ID       EVENT   ACTION
##  1:    WK    ANA06 2011-08-26      133 FERAL/JUDAS COLLARED
##  2:    PB    COR17 2017-08-24     4707       ALONE COLLARED
##  3:    EK     CR25 2004-08-04     5007       ALONE COLLARED
##  4:    WK    DEF25 2017-06-20     6289       ALONE COLLARED
##  5:    EK    DNP75 2013-10-07     7664       FERAL COLLARED
##  6:    EK    ELL17 2001-12-06     9048       ALONE COLLARED
##  7:    WK    ELZ14 2005-07-12     9863       FERAL COLLARED
##  8:    WK    ELZ19 2007-09-07     9923       FERAL COLLARED
##  9:    WK    GIB25 2005-07-12    11154       FERAL COLLARED
## 10:    WK    GIB36 2014-06-23    11267       FERAL COLLARED
## 11:    WK    GIB38 2017-06-20    11272       FERAL COLLARED
## 12:    PB  H-ETH01 2017-05-26    12256       ALONE COLLARED
## 13:    PB    HIL33 2005-05-10    12674       FERAL COLLARED
## 14:    WK     KM18 1996-09-24    13807       ALONE COLLARED
## 15:    WK     LA03 1996-10-30    14522       FERAL COLLARED
## 16:    PB    MEE04 2005-10-28    17165       ALONE COLLARED
## 17:    HC    ORD03 1998-11-12    20396       FERAL COLLARED
## 18:    HC   ORD06C 1998-11-14    20405       ALONE COLLARED
## 19:    HC    ORD18 1999-03-17    20575       FERAL COLLARED
## 20:    HC    ORD29 1999-10-09    20705       FERAL COLLARED
## 21:    HC   ORD39C 2005-05-18    20829       ALONE COLLARED
## 22:    HC    ORD46 2006-07-31    20878       FERAL COLLARED
## 23:    WK    PNT19 2017-06-20    21191       ALONE COLLARED
## 24:    WK    PNT20 2017-06-20    21192       FERAL COLLARED
## 25:    PB    ROY02 2016-11-04    21383       ALONE COLLARED
## 26:    HC     SO10 1997-05-08    22299       FERAL COLLARED
## 27:    HC     SO25 2003-05-08    22500       ALONE COLLARED
## 28:    HC    SPC13 2012-06-04    22677       FERAL COLLARED
## 29:    HC     TA17 2005-05-18    23058       ALONE COLLARED
## 30:    HC     TA21 2006-07-31    23080       FERAL COLLARED
## 31:    PB    VCL30 2017-08-24    23653       ALONE COLLARED
## 32:    PB    WAN07 2008-11-28    23890       FERAL COLLARED
## 33:    HC     WR07 2000-05-24    24109       FERAL COLLARED
## 34:    HC     WR28 2005-05-18    24366       ALONE COLLARED
## 35:    HC     WR32 2005-05-18    24395       ALONE COLLARED
## 36:    WK    XCK01 2003-08-06    24776       FERAL COLLARED
##     SHIRE JUDAS_ID EVENT_DATE EVENT_ID       EVENT   ACTION
```

```
judas.cleaned <- judas.cleaned[Time.deployment>0,]

# List of animals that have been deployed for > 10 yrs
judas.cleaned[Time.dep.years > 10,  unique(JUDAS_ID)]
```

```
##  [1] "BEV11" "HAR17" "HAR22" "BEV18" "COR12" "GIB17" "HAR23" "MHO02"
##  [9] "WAN03" "MX12"  "THE01" "GIB11" "MHO26" "LP22"  "MA01"  "MX03"
## [17] "MX10"  "NIC06" "NIC13" "ORD11" "DNP01" "DNP10" "MX04"  "NIC12"
## [25] "ORD37" "SPC06"
```

Check whether there are judas with < 5 data points and rm

```
locs <- judas.cleaned[, .N, by=JUDAS_ID]
locs[, sum(N<6)]
```

```
## [1] 295
```

```
IDs.rm <- locs[N<6, JUDAS_ID]
judas.cleaned <- judas.cleaned[!JUDAS_ID %in% IDs.rm, ]
```

Save final dataset and csv with sample size

```
dir.create(file.path(data.path, "Analysis"))
```

```
## Warning in dir.create(file.path(data.path, "Analysis")): '..\Data\
## \Analysis' already exists
```

```
save(judas.cleaned, file = file.path(data.path, "Analysis", "judas.cleaned.rda"))
descr.fin <- judas.cleaned[, .(njudas=length(unique(JUDAS_ID)),
                 start.date=min(EVENT_DATE), end.date=max(EVENT_DATE)),
                 by=c("REGION", "SHIRE")]
descr.fin
```

```
##       REGION SHIRE njudas start.date   end.date
## 1: KIMBERLEY    EK    433 1997-10-21 2017-11-14
## 2: KIMBERLEY    WK    379 1994-07-12 2017-06-20
## 3:   PILBARA    PB    254 1998-05-20 2017-11-01
## 4: KIMBERLEY    HC    324 1995-10-26 2017-06-20
```

```
descr.fin[, sum(njudas)]
```

```
## [1] 1390
```

```
write.csv(descr.fin, file = file.path(data.path, "Analysis", "descr.fin.csv"),
          row.names = F)

ntrack.events <- judas.cleaned[, .N, by=JUDAS_ID]
ntrack.events[, summary(N)]
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.00   11.00   16.00   16.97   22.00   61.00
```