# Metrics for Recorder behaviour

*Tom August*

*19 September 2016*

## Metrics

We are going to split metric into three broad groups: *Engagement profile*, *Spatial*, and *Taxanomic*

## Engagement Profile Metrics

## Spatial Meterics

## Taxanomic Metrics

These metric relate the the species that people record

### Taxanomic Breadth

This is simply a measure of the proportion of taxa a person has recorded. Note this is going to be correlated to the number of records.

```r
taxa_breadth <- function(data, recorder_name,
                         sp_col = 'default_common_name',
                         recorder_col = 'recorders'){

  data_rec <- data[data[,recorder_col] == recorder_name, c(sp_col, recorder_col)]

  return(data.frame(recorder = recorder_name,
                    taxa_breadth = length(unique(data_rec[ ,sp_col])),
                    taxa_prop = length(unique(data_rec[ ,sp_col]))/length(unique(data[ ,sp_col])),
                    n = nrow(data_rec)))
}

taxa_breadth <- do.call(rbind, lapply(unique(iRB$recorders), FUN = taxa_breadth, data = iRB))

temp <- taxa_breadth[taxa_breadth$n > 400, ]

# Lets have a look at some people who have recorded a lot
temp[order(temp$taxa_prop, decreasing = TRUE),]
```

```
##                  recorder taxa_breadth taxa_prop    n
## 2              Roy, David           67 0.9710145  868
## 9           Warren, Martin          51 0.7391304 1666
## 4             Fox, Richard          48 0.6956522  864
## 1402        Saville, Simon          48 0.6956522  441
## 729             Cox, Steve          45 0.6521739  687
## 16             Allan, David          44 0.6376812 2081
## 1765        Steele, Andrew           42 0.6086957  499
```

```
## 12                Kilbey, Dave        41 0.5942029  794
## 574               fenn, paul          41 0.5942029 2239
## 659               Atkin, Paul         41 0.5942029  613
## 788     Pennington, Robert            40 0.5797101  432
## 2041            Shersby, Megan         38 0.5507246  478
## 2193            Gillie, Tony           38 0.5507246  660
## 1176            Hill, Brian            37 0.5362319  444
## 64              shilland, ewan         36 0.5217391 1292
## 183             Newbould, John         32 0.4637681  704
## 1067 Partridge, Francesca              32 0.4637681 1062
## 350               Sell, Claire         31 0.4492754  461
## 169             Hunter, Amands         30 0.4347826  729
## 356               Stewart, Tam         29 0.4202899 1444
## 334               Jones, Dave          23 0.3333333 1194
## 181       Checkley, Graham             22 0.3188406 1813
## 2920             Leaver, Kim           21 0.3043478  416
```

**Species Rarity**

We want to capture the rarity of the species that people record. For example are they just recording the common species or are they only recording the rare ones, or perhaps they are recording everything. Since we dont know the real frequency distribution we can only compare people to the global average in the dataset. We can look to see what the distribution of species rank for each recorder is and how this compares to all records. A recorder only interested in rare species will have a median rank higher than the average. A recorder only recording common species will have a value lower than the average.

```r
# Lets look at a recorder
species_rank <- function(data, recorder_name,
                         sp_col = 'default_common_name',
                         recorder_col = 'recorders'){

  data <- data[,c(sp_col, recorder_col)]
  rank_species <- rank(abs(table(data[,sp_col])-max(table(data[,sp_col]))))
  sp_counts <- table(data[,sp_col])

  rank_reps <- rep(rank_species, sp_counts)
  grand_median <- median(rank_reps)
  grand_sd <- sd(rank_reps)

  recorder_data <- data[data[,recorder_col] == recorder_name,]
  recorder_data$rank <- rank_species[recorder_data$default_common_name]

  return(data.frame(recorder = as.character(recorder_name),
                    median = median(recorder_data$rank),
                    median_diff = median(recorder_data$rank) - grand_median,
                    stdev = sd(recorder_data$rank),
                    n = nrow(recorder_data)))
}

rarity_preference <- do.call(rbind, lapply(unique(iRB$recorders), FUN = species_rank, data = iRB))

temp <- rarity_preference[rarity_preference$n > 400, ]
```

```r
# Lets have a look at some people who have recorded a lot
temp[order(temp$median_diff, decreasing = TRUE),]
```

```
##                    recorder median median_diff    stdev    n
## 1402         Saville, Simon     13           5 12.374296  441
## 2               Roy, David     12           4 15.894358  868
## 356            Stewart, Tam     12           4 10.221537 1444
## 9            Warren, Martin     11           3 10.898222 1666
## 659             Atkin, Paul     11           3  9.668459  613
## 181        Checkley, Graham     10           2  7.082567 1813
## 64           shilland, ewan      9           1  8.128562 1292
## 169          Hunter, Amands      9           1  7.166763  729
## 183          Newbould, John      9           1  8.056680  704
## 574              fenn, paul      9           1  8.826334 2239
## 1176            Hill, Brian      9           1  9.552423  444
## 2041         Shersby, Megan      9           1  8.531861  478
## 2193           Gillie, Tony      9           1  8.672914  660
## 4               Fox, Richard      8           0  9.941570  864
## 12             Kilbey, Dave      8           0  9.123507  794
## 350             Sell, Claire      8           0  8.710295  461
## 729               Cox, Steve      8           0  9.393764  687
## 788       Pennington, Robert      8           0  8.709649  432
## 1765         Steele, Andrew      8           0  9.155960  499
## 16             Allan, David      6          -2  8.076737 2081
## 334              Jones, Dave      6          -2  5.009584 1194
## 1067   Partridge, Francesca      6          -2  7.018191 1062
## 2920            Leaver, Kim      5          -3  4.625817  416
```

Here `median_diff` gives the difference between the grand median for all records and the recorders median. This suggests `Roy, David` prefers to record rare species and `Leaver, Kim` prefers to record common species.