Data Wrangling EMu Reports

Michelle Thompson

2/26/2022

Background

Sometimes reports from EMu are not in the format we want them to be. This tutorial presents code and steps of how to data wrangles those reports into more useful forms.

We will use RI-12 MAMMALS collection event report as an example. Once that report is generated from EMu open the file on your computer and resave as csv file (this gets rid of problematic formatting embedded in EMu csv file report). File and full code at {https://github.com/MichelleThompson86/EMu_reports_RIdata}

***Note, I have this broken into steps but once you have the code how you want it you can run the code all at once for one, easy, quick job.

Load libraries

```
#load libraries
library(dplyr)
library(splitstackshape)
library(tidyr)
```

Read the report into R, call it 'CEdata' and check the first few rows of the datafile

```
#read data
CEdata <- read.csv(file="Group1.csv")

#check first dew rows of data
head(CEdata)
```

```
##   Group1_key ecollectionevents_key     irn ColCollectionEventCode
## 1          1                     1 3190484          RI-12 MAMMALS
## 2          2                     2 3165876    RI-12.B-1-2 MAMMALS
## 3          3                     2 3165876    RI-12.B-1-2 MAMMALS
## 4          4                     3 3550165    RI-12.B-1-7 MAMMALS
## 5          5                     3 3550165    RI-12.B-1-7 MAMMALS
## 6          6                     4 3550166    RI-12.B-1-8 MAMMALS
##   ColCollectionType ColCollectionMethods ColSpecifics           NamFullName
## 1       Terrestrial
## 2 Human Observation                 Trap     Mist net    Olga L. Montenegro
## 3 Human Observation                                    Mário Escobedo Torres
## 4 Human Observation   Direct Observation       Burrow    Olga L. Montenegro
## 5 Human Observation   Direct Observation       Tracks Mário Escobedo Torres
## 6 Human Observation   Direct Observation       Tracks    Olga L. Montenegro
##   ColParticipantRole_tab SitSiteNumber MapASLQualifierAccuracy ColDateVisitedTo
```

```
## 1                               RI-12                     NA
## 2   Collector\nCollector     RI-12.B-1                     NA        8/21/2003
## 3   Collector\nCollector     RI-12.B-1                     NA        8/21/2003
## 4   Collector\nCollector     RI-12.B-1                     NA        8/21/2003
## 5   Collector\nCollector     RI-12.B-1                     NA        8/21/2003
## 6   Collector\nCollector     RI-12.B-1                     NA        8/21/2003
##   ColDateVisitedFrom SecDepartment_tab HabHabitat AdmPublishWebNoPassword
## 1                             Action         NA                     Yes
## 2           8/3/2003   Action\nMammals         NA                     Yes
## 3           8/3/2003   Action\nMammals         NA                     Yes
## 4           8/3/2003   Action\nMammals         NA                     Yes
## 5           8/3/2003   Action\nMammals         NA                     Yes
## 6           8/3/2003   Action\nMammals         NA                     Yes
##   AdmPublishWebPassword
## 1                   Yes
## 2                   Yes
## 3                   Yes
## 4                   Yes
## 5                   Yes
## 6                   Yes
```

1. Split and wrangle first part of data

```
#split and wrangle first part of data
#delete key and any columns that printed out in multiple rows: collector name,
#collection methods, collection specifics by using 'select'
T1 <- CEdata %>%
  select(-Group1_key, -NamFullName, -ColCollectionMethods,-ColSpecifics )%>%
  distinct()  #get distinct rows (collapse data)

#check firt 3 rows
head(T1, 3)
```

```
##   ecollectionevents_key     irn ColCollectionEventCode ColCollectionType
## 1                     1 3190484          RI-12 MAMMALS       Terrestrial
## 2                     2 3165876    RI-12.B-1-2 MAMMALS Human Observation
## 3                     3 3550165    RI-12.B-1-7 MAMMALS Human Observation
##   ColParticipantRole_tab SitSiteNumber MapASLQualifierAccuracy ColDateVisitedTo
## 1                               RI-12                     NA
## 2   Collector\nCollector     RI-12.B-1                     NA        8/21/2003
## 3   Collector\nCollector     RI-12.B-1                     NA        8/21/2003
##   ColDateVisitedFrom SecDepartment_tab HabHabitat AdmPublishWebNoPassword
## 1                             Action         NA                     Yes
## 2           8/3/2003   Action\nMammals         NA                     Yes
## 3           8/3/2003   Action\nMammals         NA                     Yes
##   AdmPublishWebPassword
## 1                   Yes
## 2                   Yes
## 3                   Yes
```

2. Split and wrangle second part of data

```
#split and wrangle second part of data
#select only collections event key and collector name column
T2 <- CEdata %>%
  select(ecollectionevents_key, NamFullName)

T2<- mutate(getanID(data =T2, id.vars = "ecollectionevents_key")) %>%
  spread(.id, NamFullName)

#rename columns
T2<- T2 %>%
  rename(ColName_1 = 2, ColName_2 = 3, ColName_3 = 4)

#check first 3 rows
head(T2, 3)
```

```
##    ecollectionevents_key          ColName_1             ColName_2 ColName_3
## 1:                     1                                     <NA>      <NA>
## 2:                     2 Olga L. Montenegro Mário Escobedo Torres      <NA>
## 3:                     3 Olga L. Montenegro Mário Escobedo Torres      <NA>
```

3. Split and wrangle third part of data

```
#split and wrangle third part of data for collection methods
#select only collections event key and col methods
T3 <- CEdata %>%
  select(ecollectionevents_key, ColCollectionMethods)

T3<- mutate(getanID(data = T3, id.vars = "ecollectionevents_key")) %>%
  spread(.id, ColCollectionMethods)

#rename columns
T3<- T3 %>%
  rename(ColCollectionMethods_1 = 2, ColCollectionMethods_2 = 3, ColCollectionMethods_3 = 4)

#check first 3 rows
head(T3, 3)
```

```
##    ecollectionevents_key ColCollectionMethods_1 ColCollectionMethods_2
## 1:                     1                                         <NA>
## 2:                     2                  Trap
## 3:                     3     Direct Observation     Direct Observation
##    ColCollectionMethods_3
## 1:                   <NA>
## 2:                   <NA>
## 3:                   <NA>
```

4. Split and wrangle forth part of data

```
#split and wrangle third part of data for collection methods
#select only collections event key and col specifics
T4 <- CEdata %>%
  select(ecollectionevents_key, ColSpecifics)
```

```
T4<- mutate(getanID(data = T4, id.vars = "ecollectionevents_key")) %>%
  spread(.id, ColSpecifics)

#rename columns
T4<- T4 %>%
  rename(ColSpecifics_1 = 2, ColSpecifics_2 = 3, ColSpecifics_3 = 4)

#check firt 3 rows
head(T4, 3)
```

```
##    ecollectionevents_key ColSpecifics_1 ColSpecifics_2 ColSpecifics_3
## 1:                     1                          <NA>          <NA>
## 2:                     2       Mist net                         <NA>
## 3:                     3         Burrow         Tracks          <NA>
```

5. Merge together

```
#merge datasets together and last manipulation of separating
#"participant roles" into separate columns
T_merge1 <- full_join(T1, T2, by = 'ecollectionevents_key')
T_merge2 <- full_join(T_merge1, T3, by = 'ecollectionevents_key')
T_merge3 <- full_join(T_merge2, T4, by = 'ecollectionevents_key')

#you will need to create enough rows for the maximum number of collectors in dataset
T_mergeFinal<- T_merge3 %>%
  separate(ColParticipantRole_tab, c('Role1', 'Role2', 'Role3'))%>%
  separate(SecDepartment_tab, c('Dep1', 'Dep2')) #sep departments into columns
```

6. Write to new csv file

```
#write csv
write.csv(T_mergeFinal, file="Group1_wide.csv")
```