

# Motus R Book

*Tara L. Crewe, Zoe Crysler, and Philip Taylor*

# Contents

<b>A walk through the use of R for Motus automated radio-telemetry data</b>	<b>4</b>
<b>1 Introduction</b>	<b>5</b>
1.1 What this book does not cover . . . . .	6
1.2 Prerequisites . . . . .	6
1.3 Sample dataset . . . . .	6
1.4 Acknowledgements . . . . .	7
<b>2 Loading R Packages</b>	<b>8</b>
2.1 Internal data processing . . . . .	9
<b>3 Accessing and understanding detections data</b>	<b>10</b>
3.1 Data structure . . . . .	10
3.2 Database types . . . . .	11
3.3 Load relevant R packages . . . . .	11
3.4 Set system environment . . . . .	11
3.5 Downloading tag detections . . . . .	12
3.6 Export your ‘flat’ dataframe to CSV or RDS file . . . . .	16
3.7 Update and/or open an existing database . . . . .	17
3.8 Check if new data are available . . . . .	17
3.9 Force an update/re-import of tag and receiver deployment metadata . . . . .	18
3.10 Import full tag and receiver metadata . . . . .	18
3.11 Ensure that you have the correct database version . . . . .	18
3.12 R object naming convention . . . . .	19
3.13 Summary: download, access, and export data . . . . .	19
<b>4 Tag and Receiver Deployments</b>	<b>21</b>
4.1 Load relevant R packages and set working environment . . . . .	21
4.2 Load .motus file . . . . .	22
4.3 Tag Deployments . . . . .	22
4.4 Check Receiver Metadata . . . . .	27
<b>5 Data Cleaning</b>	<b>39</b>
5.1 Load required packages . . . . .	40
5.2 Load detections data . . . . .	40
5.3 Preliminary data checks . . . . .	41
5.4 Examining ambiguous detections . . . . .	46
5.5 Checking validity of run lengths of 2 . . . . .	59
5.6 Filtering the data . . . . .	59
<b>6 Exploring data with the Motus R package</b>	<b>61</b>
6.1 Load required packages . . . . .	61
6.2 Load data . . . . .	61

6.3 Summarizing your data . . . . .	62
6.4 Plotting your data . . . . .	65
6.5 Mapping your data . . . . .	72
<b>A Appendix - alltags structure</b>	<b>78</b>
<b>B Appendix - Troubleshooting</b>	<b>80</b>
B.1 Logging out of motus . . . . .	80
B.2 Resume data download . . . . .	80
B.3 Common error messages and solutions: . . . . .	80
<b>C Appendix - The Motus R Package</b>	<b>83</b>
C.1 checkVersion . . . . .	83
C.2 sunRiseSet . . . . .	84
C.3 plotAllTagsCoord . . . . .	86
C.4 plotAllTagsSite . . . . .	87
C.5 plotDailySiteSum . . . . .	88
C.6 plotRouteMap . . . . .	89
C.7 plotSite . . . . .	90
C.8 plotSiteSig . . . . .	91
C.9 plotTagSig . . . . .	92
C.10 simSiteDet . . . . .	93
C.11 siteSum . . . . .	95
C.12 siteSumDaily . . . . .	97
C.13 siteTrans . . . . .	97
C.14 tagSum . . . . .	98
C.15 tagSumSite . . . . .	99
C.16 timeToSunriset . . . . .	100
<b>D Appendix - The motusClient Package - Data filtering functions</b>	<b>103</b>
D.1 listRunsFilters . . . . .	103
D.2 Dependencies . . . . .	103
D.3 Example . . . . .	103
D.4 createRunsFilter . . . . .	104
D.5 getRunsFilters . . . . .	104
D.6 writeRunsFilter . . . . .	105

# A walk through the use of R for Motus automated radio-telemetry data

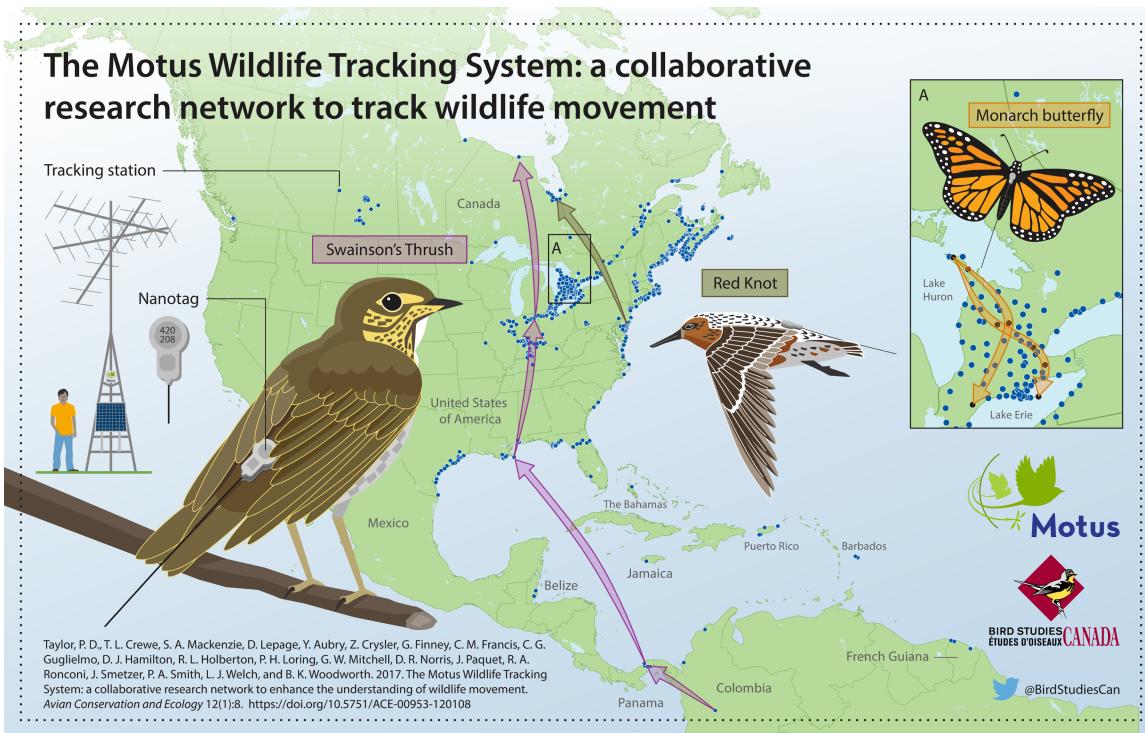


Our goal with this online ‘handbook’ is to show Motus (<https://motus.org>) users how to use the R statistical programming language (<https://www.r-project.org/>) to import tag detections data for a project or receiver; clean data and filter false positives; explore detections data through visualizations and summaries; transform the data, e.g., by determining time since sunrise/sunset or magnetic declination; and run various analytical procedures. We hope the contents will be of use, and if you have suggestions for additional examples, please let us know by emailing [motus@birdscanada.org](mailto:motus@birdscanada.org).

Version 1.0 Published January 2018

# Chapter 1

## Introduction



The Motus Wildlife Tracking System ('Motus'; Taylor et al. 2017; <https://www.motus.org>) is an international, collaborative automated radio-telemetry network to track the movement and behaviour of flying organisms affixed with digitally encoded radio-transmitters. Motus was developed at Acadia University in 2012-2013. In 2014, a major infrastructure expansion was made possible through a Canada Foundation for Innovation grant to Western University, The University of Guelph, and Acadia University. Since then, Motus has grown through the collaboration of independent researchers and organizations (see <https://motus.org/about/>). It is now managed as a program of Bird Studies Canada (<https://www.birdscanada.org>) in partnership with Acadia University.

Motus is unique among automated telemetry arrays in that all researchers in a geographic region (e.g., the Americas or Europe) use a shared radio frequency. This allows tagged animals to be detected by any receiving station across the network, greatly broadening the spatial scope of potential research questions. Motus users also use a shared data infrastructure and web portal: all data collected from across the network are centrally stored and archived, which allows users to access detections of their tags by anyone's receiver in the network, and individuals that maintain receivers have access to all detections of anyone's tags on those receivers.

Having a shared data infrastructure also means that users can benefit from R functions written specifically for Motus data by any and all users. The Motus R package described in this book is in continual development, and the intent of this online ‘handbook’ is to help users learn the various functionalities of the package, and potentially contribute to it. We also show how additional R packages such as ggplot can be used to explore, visualize, transform, and analyze Motus data.

The content of the handbook will continue to evolve and grow along with the analytical needs of the network. Those interested in contributing code to the Motus R package or this handbook can send proposed additions to [motus@birdscanada.org](mailto:motus@birdscanada.org).

Taylor, P. D., T. L. Crewe, S. A. Mackenzie, D. Lepage, Y. Aubry, Z. Crysler, G. Finney, C. M. Francis, C. G. Guglielmo, D. J. Hamilton, R. L. Holberton, P. H. Loring, G. W. Mitchell, D. R. Noriis, J. Paquet, R. A. Ronconi, J. Smetzer, P. A. Smith, L. J. Welch, and B. K. Woodworth. 2017. The Motus Wildlife Tracking System: a collaborative research network to enhance the understanding of wildlife movement. *Avian Conservation and Ecology* 12(1):8. <https://doi.org/10.5751/ACE-00953-120108>.

## 1.1 What this book does not cover

This book does not cover how to register radio tags with Motus, manage tags and station deployments, or upload raw detections data for processing. Information to guide you through those tasks can be found under the ‘resources’ tab on the Motus website at <https://motus.org/resources/>. Please remember to register your tags **prior to deployment**, and enter tag and station metadata online in a timely manner. Please also review the Motus collaboration policy and tag registration and fee schedule at <https://motus.org/policy/>.

## 1.2 Prerequisites

This book assumes that you have a basic understanding of R. Regardless of whether you are new to R or not, we highly recommend that you become familiar with ‘R for Data Science’ by Garrett Grolemund and Hadley Wickham (<http://r4ds.had.co.nz/>). Their book covers how to import, visualize, and summarize data in R using the tidyverse collection of R packages (<https://www.tidyverse.org/>). It also provides an invaluable framework for organizing your workflow to create clean, reproducible code (<http://r4ds.had.co.nz/workflow-projects.html>). We follow their lead by, wherever possible, using the tidyverse framework throughout this book.

## 1.3 Sample dataset

Throughout this book we use a subset of data from the James Bay Shorebird Project to illustrate how to access, manage, and analyze Motus data in R. We recommend that you run through the sample code in each chapter with the sample dataset **before** running through with your own data, because you will undoubtedly need to modify the code we provide in order to deal most effectively with your own data (every situation is different).

The James Bay Shorebird Project conducts monitoring and research on shorebirds staging along the James Bay coast, and is a collaborative effort among the Ontario Ministry of Natural Resources and Forestry, Bird Studies Canada, Trent University, and Environment and Climate Change Canada’s Canadian Wildlife Service, in conjunction with a larger conservation initiative involving James Bay First Nations and Nature Canada. The Royal Ontario Museum was a contributing partner until 2016. The goals of the project are to 1) improve the ability to estimate indices of abundance and population trends for shorebird species staging along the western James Bay coast, 2) understand movement patterns and their causes, and 3) identify the relative importance of shorebird staging sites and their habitats. Collectively, this information will aid in the development of conservation measures for Red Knot and other shorebird species through habitat protection like Western Hemisphere Shorebird Reserve Network (WHSRN) designation. More information can be viewed

on the James Bay Shorebird Project website at <https://www.jamesbayshorebirdproject.com/>, on Facebook <https://www.facebook.com/jamesbayshorebirdproject/>, or by contacting their project lead:

Christian Friis Wildlife Biologist Canadian Wildlife Service Environment and Climate Change Canada / Government of Canada [christian.friis@canada.ca](mailto:christian.friis@canada.ca) / Tel: 416.739.4908

Biogiste de la Faune

Service Canadien de la Faune Environnement et Changement Climatique Canada / Gouvernement du Canada [christian.friis@canada.ca](mailto:christian.friis@canada.ca) / Tél. : 416.739.4908

## 1.4 Acknowledgements

Some of the text included in this book was adapted from John Brzustowski's github repository for the Motus R package at: <https://github.com/jbrzusto/motus>.

Motus was conceived as the SensorGnome network by Philip Taylor and John Brzustowski at Acadia University. Initial expansion of the network was supported by a Canada Foundation for Innovation Grant to Western University (Dr. Christopher Guglielmo), The University of Guelph (Dr. Ryan Norris), and Acadia University (Dr. Philip Taylor). The development of the Motus web interface, R package, and accompanying handbook were made possible through a Canarie grant to Bird Studies Canada (<https://www.canarie.ca/>). Motus continues to grow as a program of Bird Studies Canada, through the collaboration of numerous independent researchers, organizations, and individuals. A non-exhaustive list of Motus partners and collaborators can be found at <https://motus.org/data/partners.jsp>. If your organization is not listed, please contact [motus@birdscanada.org](mailto:motus@birdscanada.org).

Many people have worked together to bring Motus technology, the web interface, and the R-package together. The core 'Motus Team' includes John Brzustowski, Tara Crewe, Zoe Crysler, Jeremy Hussell, Catherine Jardine, Denis Lepage, Stuart Mackenzie, Paul Morrill, and Philip Taylor.

# Chapter 2

## Loading R Packages

Two R packages have been developed for Motus users:

1. motus: provides functions to output summary plots, and to transform (add sun rise/sun set times) and analyze Motus data.
2. motusClient: provides functions to download and update detections data and tag and receiver deployment metadata from the Motus server.

Motus **users** can install the latest stable versions of the R packages using the following code. As with all R packages, you only need to install the packages once; after installation, you need to load each package (using `library()` or `require()`) each time you open a new R session.

Please note that some functionalities of the `devtools` package may require updated versions of R and RStudio. To avoid errors, please ensure you are using the most recent releases of R and RStudio, and update your R packages using `update.packages()` in the R console.

To update your existing packages:

```
update.packages()
```

Begin by installing the required packages, if not already installed. Note that the `motusClient` package, which is required to access detection data from the Motus server, is a dependency of the broader `motus` package, i.e., you should only need to run the code to install the `motus` package, and the `motusClient` package will be automatically loaded. The code to install the `motusClient` R package independently is included below, but you should not need to run it.

```
install.packages("devtools")
library(devtools)

# install motus for data download, data
# manipulation, visualization and analysis
install_github("MotusWTS/motus")

# install motusClient for data download
install_github("MotusWTS/motusClient")

library(motus)
```

If you need to update an existing `motus` or `motusClient` package, you need to specify ‘`force = TRUE`’:

```
# force a re-installation of motus package in case
# of required updates
```

```
install_github("MotusWTS/motus", force = TRUE)

# force a re-installation of motusClient package in
# case of required updates
install_github("MotusWTS/motusClient", force = TRUE)

library(motus)
```

Throughout the book, we use tidyverse, which is a collection of R packages for data science, including tidyr, dplyr, ggplot2, and lubridate for managing and manipulating dates. More information on tidyverse can be found at <https://www.tidyverse.org/>, or by browsing (or better still, thoroughly reading) ‘R for Data Science’ by Garrett Grolemund and Hadley Wickham (<http://r4ds.had.co.nz/>). For mapping we also use the rworldmap, and ggmap packages. These can be installed from CRAN, as follows:

```
library(maps)

install.packages("tidyverse")
library(tidyverse)

install.packages("tidyrr")
library(tidyrr)

install.packages("rworldmap")
library(rworldmap)

install.packages("ggmap")
library(ggmap)
```

We also install but do not load the plyr package; we use it directly for the handy round\_any function, but loading it can cause problems with the dplyr functions:

```
install.packages("plyr")
```

## 2.1 Internal data processing

As an animal moves within the detection range of a Motus station, radio transmissions, or ‘bursts’, are detected by antenna(s) and recorded by a receiver. These raw detection data are either uploaded to the Motus database instantaneously via internet connection, or downloaded from the receiver and uploaded to Motus manually. Behind the scenes, various functions read and process the raw detections data to produce the tag detections file that users access using the R package (see Chapter 3). While most users will not need to call on the internal data processing functions, a complete list of functions within the Motus server R package can be found on GitHub (<https://github.com/jbrzusto/motusServer>). The code behind each function can be viewed on GitHub, or by typing the following in the R console after loading the R package, replacing ‘function.name’ with the name of the R function of interest:

```
function.name()
```

In the next chapter we will examine and load some data.

# Chapter 3

## Accessing and understanding detections data

Before downloading your detection data, please ensure that you have no pending metadata issues through the online Data Issues page

This chapter will begin with an introduction to the structure of the detections database, followed by instructions on how to download and access the data. At the end, a summary section that includes a script to download, select variables, clean data, and export is provided (see 3.13)

### 3.1 Data structure

Each tag detection database is stored as an SQLite ('dplyr::src\_sqlite') file with the extension '.motus'. The sqlite format was chosen because:

1. it is **flexible**, allowing for many data formats.
2. it is **accessible** from many software platforms (not just R).
3. it is **appendable**: the database can be created and updated on disk without having to read in and resave the entire contents. This will save time and computer memory when searching to see if any new detections are available for your project or receiver.

The .motus file contains a series of interrelated tables where data are stored in a condensed format to save memory. The following tables are included in your .motus file;

1. **antDeps**: metadata related to antenna deployments, e.g., deployment height, angle, antenna type.
2. **batchRuns**: metadata for runIDs and associated batchIDs
3. **batches**: detection data for a given receiver and boot number.
4. **filters**: metadata related to user created filters associated with the specified receiver.
5. **gps**: metadata related to Geographic Positioning System (GPS) position of receiver.
6. **hits**: detection data at the level of individual hits.
7. **meta**: metadata related to the project and datatype (tags vs. receivers) that are included in the .motus file
8. **projAmbig**: metadata related to what projects have ambiguous tag detections
9. **projBatch**: metadata for the number of detections contained in each batch
10. **projs**: metadata related to projects, e.g., project name, principal investigator.
11. **recvDeps**: metadata related to receiver deployments, e.g., deployment date, location, receiver characteristics.
12. **recvs**: metadata related to receiver serial number and associated Motus deviceID

13. **runs**: detection data associated with a run (continuous detections of a unique tag on a given receiver).
14. **runsFilters**: a list of runIDs associated with user created filters and assigned probabilities.
15. **species**: metadata related to species, e.g., unique identifier, scientific name, common name.
16. **tagAmbig**: metadata related to ambiguous tags, e.g., ambigID and associated motusTagID
17. **tagDeps**: metadata related to tag deployments, e.g., deployment date, location, and species.
18. **tags**: metadata related to tags, e.g., unique identifier, tag characteristics (e.g., burst interval).

In addition to these tables, there are also ‘virtual’ tables or ‘views’, which have been created through queries that merge data from the various tables into a single convenient ‘view’ that contains all of the fields you are likely to need. The following views are currently included in each .motus file:

1. **allambigs**: lists in long-data format each motusTagID (up to 6) associated with each negative ambigID.
2. **alltags**: provides the full detection data for all tags, and all ambiguous (duplicate) tags, associated with your project. Ambiguous detections are repeated for each motusTagID represented by each ambigID.

Because the file is a dplyr::src\_sqlite file, all of the dplyr functions can be used to filter and summarize the .motus database, without needing to first save the data as a *flat* file (a typical two-dimensional dataframe). The SQL format is very advantageous when you have a large file – the queries using SQL will be substantially faster than those done on a flat dataframe.

## 3.2 Database types

There are two types of tag detection databases available for download:

1. **receiver database**: includes all detections of any registered tags from a single receiver. A receiver database has a name like SG-1234BBBBK5678.motus, where the name is the serial number of the receiver.
2. **project database**: includes all detections of your registered tags from across the Motus network. A tag project database has a name like project-123.motus, where the number is the Motus project ID.

These two databases correspond to the basic model of data sharing:

1. you get all detections of anyone’s tags by *your* receivers (i.e., one receiver tag database for each receiver you deploy).
2. you get all detections of *your* tags by *anyone*’s receivers (i.e., one project tag database for each of your Motus projects).

## 3.3 Load relevant R packages

Before we begin working with data, we need to load the required packages for this chapter. If you have not yet *installed* these packages (from github and CRAN) then please return to Chapter 2 and do so.

```
# load required 'motus' package from github
require(motus)
```

## 3.4 Set system environment

Set the system environment time zone to Greenwich Mean Time (GMT), to ensure that you are always working in GMT. This is a very important step, and should be part of every working session. If you fail to do this, then two problems can arise. Times are stored in the Motus database in GMT, and if you do not keep your environment in GMT, then they can be inadvertently changed during import. Second, if tags have been detected across multiple time zones, then they can also inadvertently be changed.

```
Sys.setenv(TZ = "GMT")
```

## 3.5 Downloading tag detections

To import tag detections for your project or receiver, you need a numerical project id or character scalar receiver serial number.

The success of the Motus network is dependent on the timely upload of detection data from receivers, and on the maintenance of accurate and up to date tag and receiver metadata by collaborators. After downloading your data from the Motus server, users are encouraged to check for updated detection data (see sections 3.7 and 3.8) and metadata (see section 3.9) each time they run an analysis, because collaborators can add detection data and metadata at any time, and these could influence the completeness of your own detections data.

Be warned that large datasets can take some time (sometimes a few hours) to download from the Motus server when downloading for the first time. After the initial download, loading a .motus file into R and updating for any new data will be near instantaneous.

### 3.5.1 Download data for a project for the *first time*

All data downloads are completed using the tagme() function in the motus R package. This function will save an SQLite database to your computer with the extension “.motus”; further details on data structure are in section 3.1. The following parameters are available for the tagme() function:

- **projRecv**: integer project number OR a character scalar receiver serial number.
- **new**: if set to “TRUE”, it will create a new empty .motus file in your local directory. Do not use this parameter or set it to “FALSE” if you already have a .motus file.
- **update**: if set to “TRUE”, will download all available data to your existing .motus file. Must be set to “TRUE” on your first data download and any subsequent downloads if you wish to check for new data. Set to “FALSE” if you do not wish to check for new data (e.g., if working offline).
- **dir**: Your .motus data is automatically saved to your working directory, unless you specify a different location using this parameter.
- **forceMeta**: if set to “TRUE”, it will force an update of metadata to an existing .motus file.

Throughout this book we use sample data (see section 1.3) which has been assigned to project 176.

Let’s get started by downloading data by project - this will include all detections of your tags on any receiver.

Note that when downloading data from the Motus server for the **first time**, you must specify ‘new = TRUE’ and ‘update = TRUE’. You will also be prompted to login (see 3.5.2 below).

**Unless the directory that you want your data saved in is stated explicitly within the function call, data will be downloaded to the current working directory.**

```
# lets start by determining what our working
# directory is so we know where our file will be
# saved you can change the working directory with
# setwd()
getwd()
# specify the project number you wish to download
# detections for, in this case the sample data
# project;
proj.num <- 176
# As this is the first time you are downloading
# data for project 176, set new = TRUE and update =
```

```

# TRUE. This will create a .motus file in your
# current working directory, which was shown above
# using getwd() This will also create an SQL object
# in your R environment called 'sql.motus'
sql.motus <- tagme(projRecv = proj.num, new = TRUE,
                    update = TRUE)
# OR you can specify a different location to save
# the data by entering your preferred filepath. In
# this example we save to our data folder using the
# dir parameter:
sql.motus <- tagme(projRecv = proj.num, new = TRUE,
                    update = TRUE, dir = "C:/Users/guest/Documents/data/")

```

Using tagme() as shown above will download a file to your working or specified directory called “project-176.motus” for the sample data (the number in the file name corresponds to the project number). The progress of the download process should print on the console; if you are not seeing it, try scrolling down your screen while tagme is running.

In the event that your connection to the Motus server fails prior to a complete download (e.g., due to a poor internet connection), use ‘tagme(proj.num, update = TRUE)’ to continue the download from where it left off, ensuring to specify a directory if it is saved outside the working directory.

### 3.5.2 User Authentication

#### 3.5.2.1 Login

The first time you call a function using the Motus R package, you will be asked to enter your motus.org username and password in the R console to authenticate your access to project data. This will only happen once per R session. If you do not have a Motus username and password, you can sign up at <https://motus.org/data/user/new>. Permission to access project data will then be granted by Motus staff or the project principal investigator.

Throughout this book we use sample data (see section 1.3) which has been assigned to project 176. When accessing this data you will need to login using username and password ‘motus.sample’ in the R console when prompted by the tagme() function (see section 3.5.1). It will look like this:

```

> tagme(176, update = TRUE, dir = "./data/")
Please enter a value for login name at motus.org
==>
motus.sample
Please enter a value for password at motus.org
==>
motus.sample
Checking for new data in project 176
src: sqlite 3.19.3 [F:\Motus\Motus_RPackage\MotusRBoo
tbls: admInfo, allambigs, alltags, antDeps, batches, b
, recvs,
  runs, runsFilters, species, tagAmbig, tagDeps, tags
> |

```

To download data for one of your own projects, you simply need to change the project number to that of your own project in the tagme() call, and enter your own Motus login/password in the R console when prompted.

If you are already logged in as the sample data user, you will need to first logout to download your own data (see 3.5.2.2).

### 3.5.2.2 Logging out

Once you are logged in under one user account, you will not be able to access data from another account. If you need to logout of the current account to access other data, you can run the code below.

```
motusLogout()
```

### 3.5.3 Download data for a receiver for the *first time*

We could also download data by receiver through the same process as described above. This will provide you with all detections of any tags on the specified receiver. As there are no receivers registered to sample project 176, **this call will not work**. If you have a receiver registered to your own project, replace the receiver serial number in the tagme call below with the serial number for your own receiver, ensuring that you are logged in using your own credentials (see section 3.5.2.2).

```
proj.num = "SG-123BBBK1234"
sql.motus <- tagme(projRecv = proj.num, new = TRUE,
                    update = TRUE)
```

This will download a file to your working directory named ‘SG-123BBBK1234.motus’.

Users may wish to work directly with the .motus SQLite file. However, since many users are more familiar with a ‘flat’ dataframe format, instructions to view the the data as a flat dataframe within R, and on how to export the flat file to .csv or .rds format, are included below. Throughout the majority of this book, we use a flat dataframe format.

### 3.5.4 Accessing downloaded detection data

Now that we’ve downloaded our data as an SQLite database and loaded it into an R object called ‘sql.motus’, we want to access the tables stored within. Detailed descriptions of all the tables stored in the .motus file can be found in section 3.1.

You can also view the list of tables, and variables contained within those tables, using the code below:

```
# specify the filepath where your .motus file is
# saved, and the file name.
file.name <- dbConnect(SQLite(), "./data/project-176.motus")

# get a list of tables in the .motus file specified
# above.
dbListTables(file.name)

# get a list of variables in the 'species' table in
# the .motus file.
dbListFields(file.name, "species")
```

The *virtual* table ‘alltags’ contains the detection data, along with all metadata variables that most users will ever need from the various underlying .motus tables. We access the tables using the `tbl()` function from the `dplyr` package which we installed in Chapter 2:

```
# this retrieves the 'alltags' table from the
# 'sql.motus' SQLite file we read in earlier
tbl.alltags <- tbl(sql.motus, "alltags") # virtual table
```

We now have a new ‘tbl.alltags’ object in R. The underlying structure of these tables is a list of length 2:

```
str(tbl.alltags)
```

The first part of the list, ‘src’, is a list that provides details of the `SQLiteConnection`, including the directory where the database is stored. The second part is a list that includes the underlying table. Thus, the R object ‘alltags’ is a *virtual* table that stores the database structure and information required to connect to the underlying data in the .motus file. As stated above, the advantage of storing the data in this way is that it saves memory when accessing very large databases, and functions within the `dplyr` package can be used to manipulate and summarize the tables before collecting the results into a typical ‘flat’ format dataframe.

If you want to use familiar functions to get access to components of the underlying data frame, then use the ‘collect’ function. For example, to look at the names of the variables in the alltags table:

```
tbl.alltags %>% collect() %>% names() # list the variable names in the table
```

```
## [1] "hitID"                 "runID"                  "batchID"
## [4] "ts"                     "sig"                    "sigsd"
## [7] "noise"                  "freq"                   "freqsd"
## [10] "slop"                   "burstSlop"              "done"
## [13] "motusTagID"             "ambigID"                "port"
## [16] "runLen"                 "bootnum"                "tagProjID"
## [19] "mfgID"                  "tagType"                "codeSet"
## [22] "mfg"                    "tagModel"               "tagLifespan"
## [25] "nomFreq"                "tagBI"                  "pulseLen"
## [28] "tagDeployID"            "speciesID"              "markerNumber"
## [31] "markerType"              "tagDeployStart"         "tagDeployEnd"
## [34] "tagDeployLat"            "tagDeployLon"           "tagDeployAlt"
## [37] "tagDeployComments"       "fullID"                 "deviceID"
## [40] "recvDeployID"            "recvDeployLat"           "recvDeployLon"
## [43] "recvDeployAlt"           "recv"                   "recvDeployName"
## [46] "recvSiteName"            "isRecvMobile"           "recvProjID"
## [49] "antType"                 "antBearing"              "antHeight"
## [52] "speciesEN"               "speciesFR"               "speciesSci"
## [55] "speciesGroup"            "tagProjName"             "recvProjName"
## [58] "gpsLat"                  "gpsLon"                 "gpsAlt"
```

To convert the ‘alltags’ view or other table in the .motus file into a typical ‘flat’ format, i.e., with every record for each field filled in, use the ‘`collect()`’ and ‘`as.data.frame()`’ functions. The output can then be further manipulated, or used to generate a RDS file of your data for archiving or export.

```
df.alltags <- tbl.alltags %>% collect() %>% as.data.frame()
```

Now we have a flat dataframe of the alltags table called ‘df.alltags’. We can look at some metrics of the file:

```
names(df.alltags) # field names
str(df.alltags) # structure of your data fields
head(df.alltags) # prints the first 6 rows of your df to the console
summary(df.alltags) # summary of each column in your df
```

Note that the format of the time stamp (ts) field is numeric and represents seconds since January 1 1970. We recommend that when you transform your tables into flat dataframes, that you format the time stamp using the `lubridate` package at that time, e.g.:

```
df.alltags <- tbl.alltags %>%
  collect() %>%
  as.data.frame() %>%      # for all fields in the df (data frame)
  mutate(ts = as_datetime(ts, tz = "UTC", origin = "1970-01-01"))

# the tz = "UTC" is not necessary here, provided you have set your system time to UTC/GMT
# ... but it serves as a useful reminder!
```

Note that time stamps can only be manipulated in this way *after* collecting the data into a flat dataframe.

If you want to load only part of your entire virtual table (e.g. certain fields, certain tags, or all tags from a specified project or species), you can use dplyr functions to filter the data before collecting into a dataframe. Some examples are below:

1. To select certain variables:

```
# to grab a subset of variables, in this case a
# unique list of Motus tag IDs at each receiver and
# antenna.
df.alltagsSub <- select(tbl.alltags, recv, port, motusTagID) %>%
  distinct() %>% collect() %>% as.data.frame()
```

2. To select certain tag IDs:

```
# filter to include only motusTagIDs 16011, 23316
df.alltagsSub <- filter(tbl.alltags, motusTagID %in%
  c(16011, 23316)) %>% collect() %>% as.data.frame() %>%
  mutate(ts = as_datetime(ts, tz = "UTC", origin = "1970-01-01"))
```

3. To select a specified species:

```
# filter to only Red Knot (using speciesID)
df.4670 <- filter(tbl.alltags, speciesID == 4670) %>%
  collect() %>% as.data.frame() %>% mutate(ts = as_datetime(ts,
  tz = "UTC", origin = "1970-01-01"))

# filter to only Red Knot (using English name)
df.redKnot <- filter(tbl.alltags, speciesEN == "Red Knot") %>%
  collect() %>% as.data.frame() %>% mutate(ts = as_datetime(ts,
  tz = "UTC", origin = "1970-01-01"))
```

Using dplyr(), your virtual table can also be summarized before converting to a flat file. For example, to find the number of different detections for each tag at each receiver:

```
df.detectSum <- tbl.alltags %>% group_by(motusTagID,
  recv) %>% tally() %>% collect() %>% as.data.frame()
```

In later chapter(s) we will show you additional ways of summarizing and working with your data.

## 3.6 Export your ‘flat’ dataframe to CSV or RDS file

A good workflow is to create a script that deals with all your data issues (as described in later chapters), and then saves the resulting dataframe for re-use. If you do this, you can quickly start an analysis or visualization session from a known (and consistent) starting point. We use an .rds file, which preserves all of the associated R data structures (such as time stamps).

```
# save the df.alltags file as an RDS file in your
# data folder
saveRDS(df.alltags, "./data/df.alltags.RDS")
```

Some users may also want to export the flat dataframe into a .csv file for analysis in other programs. This can easily be done with the following code. Note that it **does not** preserve time stamps:

```
# save the df.alltags file as a .csv file in your
# data folder
write.csv(df.alltags, "./data/df.alltags.CSV")
```

## 3.7 Update and/or open an existing database

As you or other users upload data to our server, you may have additional tag detections that weren't present in your initial data download. Since the .motus file is a SQLite database, you can update your existing file with any newly available data, rather than doing a complete new download of the entire database. To open and update a detections database that already exists (has been downloaded previously), we use the tagme() function but set new = FALSE:

```
sql.motus <- tagme(projRecv = proj.num, new = FALSE,
                    update = TRUE, dir = "./data/") # use dir = to specify a directory
```

If you are working offline, and simply want to open an already downloaded database without connecting to the server to update, use new = FALSE and update = FALSE:

```
sql.motus <- tagme(projRecv = proj.num, new = FALSE,
                    update = FALSE) # use dir = to specify a directory
```

## 3.8 Check if new data are available

To check if new data are available for your project or receiver without downloading the data, you can use the tellme() function, which returns a list with:

- **numHits**: number of new tag detections.
- **numBytes**: approximate uncompressed size of data transfer required, in megabytes.
- **numRuns**: number of runs of new tag detections, where a run is a series of continuous detections for a tag on a given antenna.
- **numBatches**: number of batches of new data.
- **numGPS**: number of GPS records of new data.

The following assumes that a local copy of the database already exists:

```
tellme(projRecv = proj.num) # If db is in the working directory
tellme(projRecv = proj.num, dir = "./data/") # To specify a different directory
```

To check how much data is available for a project but you *do not* have a database for it, use the 'new' parameter:

```
tellme(projRecv = proj.num, new = TRUE)
```

### 3.9 Force an update/re-import of tag and receiver deployment metadata

Tag and receiver metadata are automatically merged with tag detections when data are downloaded. However, if metadata have been updated since your initial download, you can force re-import of the metadata when updating a database by running:

```
sql.motus <- tagme(projRecv = proj.num, forceMeta = TRUE)
```

### 3.10 Import full tag and receiver metadata

When you use tagme() to download or update your .motus file, you are provided with the metadata for:

1. any tags registered to your project which have detections;
2. tags from other projects which are associated with ambiguous detections (see Chapter 5) in your data;
3. receivers that your tags and any ambiguous tags were detected on.

In many instances, you will want access to the full metadata for all tags and receivers across the network, e.g., to determine how many of your deployed tags were not detected, or to plot the location of stations with and without detections. The metadata() function can be used to add the complete Motus metadata to your .motus file. The metadata function only needs to be run once, but we suggest that you re-import the metadata occasionally to ensure that you have the most recent and up-to-date information.

Running the metadata function as follows will add the appropriate metadata from across the network (all tags and all receivers) to the ‘recvDeps’ and ‘tagDeps’ tables in your .motus file:

```
# access all tag and receiver metadata for all
# projects in the network.
metadata(sql.motus)
```

Alternatively, you can load metadata for a specific project(s) using:

```
# access tag and receiver metadata associated with
# project 176
metadata(sql.motus, projectIDs = 176)

# access tag and receiver metadata associated with
# projects 176 and 1
metadata(sql.motus, projectIDs = c(176, 1))
```

### 3.11 Ensure that you have the correct database version

When you call the tagme function to load the sqlite database, the version of the R package used to download the data is stored in an admInfo table. Over time, changes will be made to the functionality of the R package that may require adding new tables, views or fields to the database. If your version of the database does not match the version of the R package, some of the examples contained in this book may not work. The following call will check that your database has been updated to the version matching the current version of the motus R package. If your database does not match the most current version of the R package, see Chapter 2 for instructions on updating the motus and motusClient R packages, and then use tagme() with update = TRUE to update your database to the correct format. Refer to Appendix B if the checkVersion() call returns a warning.

```
checkVersion(sql.motus)
```

## 3.12 R object naming convention

Throughout this chapter and the rest of the book, we name R objects according to their structure and the source of the data contained in the object. So, SQLite objects will be prefixed with ‘sql.’, virtual table objects will be prefixed with ‘tbl.’, and dataframe objects will be prefixed with ‘df.’; the rest of the name will include the name of the .motus table that the data originates from. Throughout the rest of the book we will be relying on and referencing the naming formats below; please ensure that you are familiar with these before continuing to the next chapter. The following code assumes you have already downloaded the sample data and do not need to update it; if you have not downloaded the data, see section 3.5.1 for instructions on initial download:

```
# SQLite R object, which links to the .motus file:
sql.motus <- tagme(176, update = TRUE, dir = "./data")

# virtual table object of the alltags table in the sample.motus file:
tbl.alltags <- tbl(sql.motus, "alltags")
df.alltags <-tbl.alltags %>%
  collect() %>%
  as.data.frame() %>% # dataframe ("flat") object of alltags table
  mutate(ts = as_datetime(ts, tz = "UTC", origin = "1970-01-01"))
```

## 3.13 Summary: download, access, and export data

Throughout the book we will predominantly be working with the ‘alltags’ table in flat format. Although described in detail above, here we include a quick summary of how you download, access, and convert the sample data for use in the rest of the book. In later sections of the book, we include additional recommended modifications and filtering of unnecessary variables, more information is available in section 5.1 in Chapter 5.

```
# set proj.num to 176 for sample data
proj.num <- 176
# download and load detection data from the sample
# project and save in data folder login and
# password for sample data is 'motus.sample' if you
# are accessing already-downloaded data, use new =
# FALSE; if you don't want to update your data,
# also set update = FALSE
sql.motus <- tagme(proj.num, new = TRUE, update = TRUE,
  dir = "./data/")
# access the 'alltags' table within the SQLite file
tbl.alltags <- tbl(sql.motus, "alltags")
# convert 'tbl.alltags' to a flat dataframe and
# change numeric time to a datetime object
df.alltags <-tbl.alltags %>% collect() %>% as.data.frame() %>%
  mutate(ts = as_datetime(ts, tz = "UTC", origin = "1970-01-01"))
```

For your own data we suggest creating a script with the following workflow:

1. download and/or update your data
2. select variables of interest for the table you are working with (typically ‘alltags’)
3. include any initial cleaning

4. save the resulting data as an .rds file as described in section 3.6. We suggest using RDS instead of CSV, because the RDS format preserves the underlying structure of the data (e.g. POSIX times stay as POSIX times). If you want to export your data to another program, then a CSV format might be preferred.

We caution that producing a flat file using the full suite of fields can use a lot of memory, and can slow R down considerably when dealing with large datasets. For some combinations of data sets and computers, it may be impossible to directly use data frames in R. If that is the case, then this is the point in your workflow where you should carefully consider the information you need from within your data set (for example, how it is aggregated) and simplify it. You can always return to this script and creating a new RDS file with different variables, or aggregated at a different scale.

To read in a saved RDS file you can run:

```
df.alltags.saved <- readRDS("./data/df.alltags.rds") # reads in your file 'df.alltags.rds' saved in the
```

In the next chapter we will check for missing metadata.

# Chapter 4

# Tag and Receiver Deployments

Before working with your detection data, a first step is to summarize and visualize the metadata for tag and receiver deployments registered to your project. Summarizing and plotting your deployments can be an effective way to find any errors in tag or receiver deployment metadata, which can in turn influence the completeness of the detections data for your project and the projects of others with detections of their own tags on your receivers.

This chapter is a complement to the online Data Issues page, which provides each project with a list of metadata issues (missing or outlying values) to be accepted or ignored. As such, please address any and all errors associated with your project on the Data Issues page **before** importing your data through R. This chapter does not provide a full check of your deployment metadata, but will help uncover errors that have been missed by the automatic queries on the Data Issues page.

We use the James Bay Shorebird Project sample dataset throughout this chapter (see Section 1.3). As you run through the code to look at your own deployments, **please fix any errors or omissions in your metadata by signing in to <https://motus.org/>**, and under the ‘Manage Data’ tab, select either ‘Manage Tags’ to fix tag deployment metadata or ‘Manage Receivers’ to fix receiver deployment metadata. It is important to fix metadata errors online, so that errors are fixed at the source and archived on the Motus Server, ensuring all users have access to the correct tag and receiver metadata. Metadata corrected online will automatically be corrected in your detection files. If you have already downloaded your detection data, you can update the existing file to include new metadata and detections (see sections 3.9, 3.7).

## 4.1 Load relevant R packages and set working environment

Before we begin working with data, we need to load the required packages for this chapter. If you have not yet installed these packages (from github and CRAN) then please return to Chapter 2 and do so.

```
library(tidyverse)
library(tidyr)
library(motus)

# Set the system environment time zone to GMT (to
# ensure that you are always working in GMT)
Sys.setenv(TZ = "GMT")
```

## 4.2 Load .motus file

This chapter assumes that the .motus file has already been downloaded, if you have not done so please return to Chapter 3 for instructions on how to do so. To update and load the existing file into R, use tagme(), you may have to login as described in the previous chapter with username **and** password ‘motus.sample’

```
proj.num <- 176

sql.motus <- tagme(proj.num, update = TRUE, dir = "./data")
```

## 4.3 Tag Deployments

In your .motus file, when using the tagme function, you are only provided with the metadata for any tags from your project with detections along with metadata for associated ambiguous tags from other projects, and receiver metadata for stations where your tags were detected. Here we will:

1. download full tag metadata for our project only;
2. determine how many tags are registered to your project;
3. determine how many of those registered tags were deployed;
4. determine location of tag deployments;
5. determine completeness and accuracy of tag deployment metadata.

We will run through each of these in sequence.

### 4.3.1 Download full project tag metadata

Incomplete metadata or missing tag registrations can result in missing detection data. We therefore want to assess the completeness of all tags registered to our projects - not just tags for which we have detections. In order to do this we will use the metadata() function for project 176, described in more detail in section 3.10.

```
metadata(sql.motus, projectIDs = proj.num)
```

### 4.3.2 Number of registered tags

Now that we have complete tag metadata for our project, we can check the number of tags registered by loading the ‘tags’ table in the .motus file. The ‘tags’ table contains the metadata of each registered tag, including a unique tagID and information on manufacturer, model, nominal and offset frequency, burst interval, and pulse length. The ‘tags’ table does not include deployment information. We select the metadata specific to the James Bay Shorebird Project, and ignore tag metadata associated with any duplicate tags belonging to other projects:

```
tbl.tags <- tbl(sql.motus, "tags")
df.tags <-tbl.tags %>% filter(projectID == proj.num) %>%
  collect() %>% as.data.frame()
```

The number of rows in the ‘df.tags’ database is equivalent to the number of tags registered to the James Bay Shorebird Project in the sample dataset (i.e., 18 tags):

```
nrow(df.tags) # number of registered tags in the database
```

```
## [1] 18
```

You can view the motusTagIDs:

```
unique(df.tags$tagID)
```

```
## [1] 16011 16035 16036 16037 16038 16039 16044 16047 16048 16052 17357
## [12] 19129 22867 22897 22902 22905 23316 23319
```

If you are missing registered tags, please follow the instructions at <https://motus.org/tag-registration/>.

### 4.3.3 Number of registered tags that were deployed

The tag deployment metadata table ('tagDeps') in the .motus file is required to check which registered tags have deployments. This file includes the date, time, species, and location of tag deployment. The database is subset to project '176', and we use the anti\_join function to determine which registered tags have (or do not have) corresponding deployment information.

```
tbl.tagDeps <- tbl(sql.motus, "tagDeps")
df.tagDeps <-tbl.tagDeps %>%
  filter(projectID == proj.num) %>%
  collect() %>%
  as.data.frame() %>% # once in df format, can format dates with lubridate
  mutate(tsStart = as_datetime(tsStart, tz = "UTC", origin = "1970-01-01"),
         tsEnd = as_datetime(tsEnd, tz = "UTC", origin = "1970-01-01"))

anti_join(df.tags, df.tagDeps, by = "tagID")

## [1] tagID      projectID    mfgID      type       codeSet
## [6] manufacturer model      lifeSpan    nomFreq    offsetFreq
## [11] bi          pulseLen
```

In the sample data, there are no registered tags without deployment metadata, which suggests that all tags were deployed. If you have undeployed tags in your own files, please check your records to ensure this is the case; without deployment metadata, detections for registered but 'undeployed' tags will be missing from your detections database.

### 4.3.4 Number of deployments per tag

A tag might be deployed more than once, for example, if a previously deployed tag was recovered, and then later re-deployed on another individual. When tags are deployed more than once, the detections data should be considered independently for each deployment.

Throughout this book we use the 'motusTagID' as a unique identifier for a deployment. However, when a tag is deployed more than once, the 'motusTagID' will remain consistent between deployments, and we instead need to use the 'tagDeployID', or combination of 'motusTagID' and 'tagDeployID' to distinguish which detections belong to which deployment.

Here, we check whether there are any tags with more than one deployment in the sample data (there are none), and then show you how to make a combined tagID/deployID variable to use in place of the 'motusTagID' if you have multiple deployments of a tag in your own data:

```

df.alltags %>%
  select(motusTagID, tagDeployID) %>%
  filter(!is.na(tagDeployID)) %>% # remove NA tagDeployIDs
  distinct() %>%
  group_by(motusTagID) %>%
  mutate(n = n()) %>%
  filter(n > 1)

## # A tibble: 0 x 3
## # Groups:   motusTagID [0]
## # ... with 3 variables: motusTagID <int>, tagDeployID <int>, n <int>

If you do have multiple deployments for a tag, we recommend creating a ‘motusTagDepID’ to use in place of the ‘motusTagID’ to define unique deployments of a tag. Moving forward, you would use ‘motusTagDepID’ in place of ‘motusTagID’ as you work through the rest of the book:

df.alltags <- df.alltags %>% mutate(motusTagDepID = paste(motusTagID,
  tagDeployID, sep = "."))

# and do the same for the tag metadata

df.tagDeps <- df.tagDeps %>% mutate(motusTagDepID = paste(tagID,
  deployID, sep = "."))

```

### 4.3.5 Location of tag deployments

Creating a map of your tag deployments can point out any obvious errors in the tag deployment latitude or longitude that weren’t captured by the online metadata message center queries.

#### a. Load base map files

Load base map files from the rworldmap package:

```

na.lakes <- map_data(map = "lakes")
na.lakes <- mutate(na.lakes, long = long - 360)

# Include all of the Americas to begin
na.map <- map_data(map = "world2")
na.map <- filter(na.map, region %in% c("Canada", "USA"))

na.map <- mutate(na.map, long = long - 360)

# Others countries in the Americas that you may
# want to plot, depending on your location:
# 'Mexico', 'lakes', 'Belize', 'Costa Rica',
# 'Panama', 'Guatemala', 'Honduras', 'Nicaragua',
# 'El Salvador', 'Colombia', 'Venezuela',
# 'Ecuador', 'Peru', 'Brazil', 'Guyana', 'Suriname',
# 'Bolivia', 'French Guiana', 'Jamaica', 'Cuba',
# 'Haiti', 'Dominican Republic', 'The Bahamas',
# 'Turks and Caicos Islands', 'Puerto Rico',
# 'British Virgin Islands', 'Montserrat',
# 'Dominica', 'Saint Lucia', 'Barbados', 'Grenada',
# 'Trinidad and Tobago', 'Chile', 'Argentina',
# 'Uruguay'

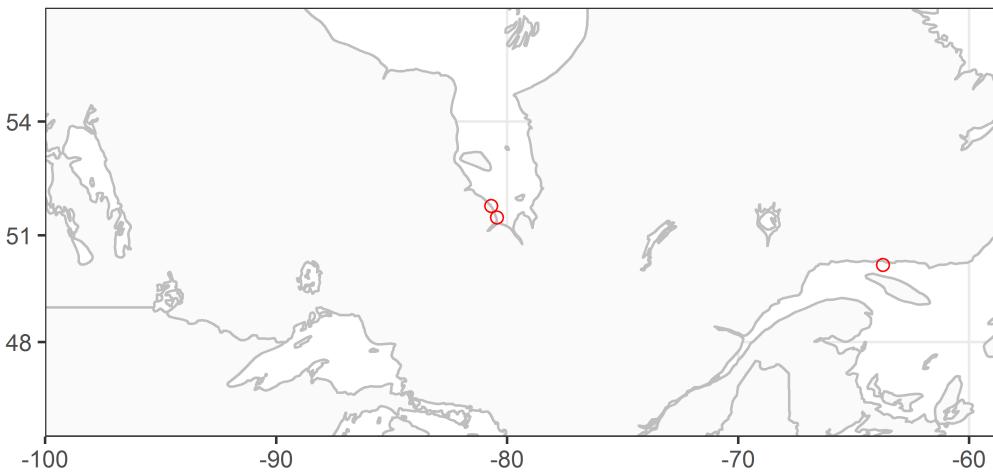
```

### b. Map the locations of tag deployments

Map the location of tag deployments for the sample data:

```
# set limits to map based on locations of
# detections, ensuring they include the deployment
# locations
xmin <- -100 #min(df.tagDeps$longitude, na.rm = TRUE) - 5
xmax <- max(df.tagDeps$longitude, na.rm = TRUE) + 5
ymin <- min(df.tagDeps$latitude, na.rm = TRUE) - 5
ymax <- max(df.tagDeps$latitude, na.rm = TRUE) + 5

# map using ggplot
ggplot(na.lakes, aes(long, lat)) + geom_polygon(data = na.map,
  aes(long, lat, group = group), colour = "grey",
  fill = "grey98") + geom_polygon(aes(group = group),
  colour = "grey", fill = "white") + coord_map(projection = "mercator",
  xlim = c(xmin, xmax), ylim = c(ymin, ymax)) + xlab("") +
  ylab("") + theme_bw() + geom_point(data = filter(df.tagDeps,
  projectId == 176), aes(longitude, latitude), cex = 2,
  pch = 1, colour = "red")
```



If there are any errors in tag deployment location, please correct these online at <https://motus.org/data/>.

#### 4.3.6 Check completeness and accuracy of tag deployment metadata

Required tag metadata includes deployment start date/time, end date/time (if applicable), deployment latitude, deployment longitude, and species. Lack of information on deployment date, time, and location in particular can influence the estimated lifespan of your tag, and therefore whether the tagFinder will ‘look’ for your tag at the appropriate time(s). It can also increase the potential for ambiguities with duplicate tags in the system.

##### a. Look at range of metadata values

As a first step, use `summary(df.tagDeps)` to get an idea of the range of each variable, and whether any variables have missing (NA) or odd values. The following summarizes a subset of the variables in the `df.tagDeps` database. There are several things to consider: are the range of start and end dates reasonable

for your deployments, or are there obvious errors in the timing of deployments? Is the range in deployment latitude and longitude reasonable? Are the values for species IDs correct?

```
df.tagDeps %>% select(tagID, projectID, tsStart, tsEnd,
                        speciesID, latitude, longitude) %>% summary()

## #> #> #> #> #>
```

	tagID	projectID	tsStart
## Min.	16011	Min.	2015-08-02 11:40:00
## 1st Qu.	16038	1st Qu.	2015-08-13 15:25:00
## Median	16050	Median	2015-09-10 17:50:30
## Mean	18616	Mean	2016-01-24 12:49:36
## 3rd Qu.	22890	3rd Qu.	2016-09-25 15:34:15
## Max.	23319	Max.	2016-10-15 16:00:00
	tsEnd	speciesID	latitude
## Min.	2015-12-17 11:40:00	Min.	50.19
## 1st Qu.	2015-12-28 15:25:00	1st Qu.	50.52
## Median	2016-03-10 17:50:30	Median	51.48
## Mean	2016-07-28 18:09:36	Mean	51.18
## 3rd Qu.	2017-06-06 09:53:45	3rd Qu.	51.48
## Max.	2017-06-26 16:00:00	Max.	51.80
	longitude		
## Min.	-80.69		
## 1st Qu.	-80.45		
## Median	-80.45		
## Mean	-75.85		
## 3rd Qu.	-67.92		
## Max.	-63.75		

There are no missing start dates (tsStart), and deployment start dates range from 2015 to 2016, which is reasonable for this project.

The species IDs are numeric, and somewhat meaningless without an ability to assign an actual species name to the numeric ID, which we do next, however there are no missing values.

### b. Check that species IDs are appropriate for your data

The ‘species’ table in the .motus file associates each numeric species ID with an English, French, and scientific name. We load that table, and subset to the suite of numeric speciesIDs in the tag metadata:

```
# generate list of species IDs in project 176
# metadata
sp.list <- unique(df.tagDeps$speciesID)

# Species metadata
tbl.species <- tbl(sql.motus, "species")
tbl.species %>% filter(id %in% sp.list) %>% collect() %>%
  as.data.frame()
```

	id	english	french
## 1	4180	Semipalmated Plover	Pluvier semipalmé
## 2	4670	Red Knot	Bécasseau maubèche
## 3	4680	Sanderling	Bécasseau sanderling
## 4	4690	Semipalmated Sandpiper	Bécasseau semipalmé
## 5	4760	White-rumped Sandpiper	Bécasseau à croupion blanc
## 6	4780	Pectoral Sandpiper	Bécasseau à poitrine cendrée
## 7	4820	Dunlin	Bécasseau variable
		scientific group sort	
## 1	Charadrius semipalmatus	BIRDS	NA

```
## 2      Calidris canutus BIRDS    NA
## 3      Calidris alba BIRDS     NA
## 4      Calidris pusilla BIRDS   NA
## 5      Calidris fuscicollis BIRDS NA
## 6      Calidris melanotos BIRDS NA
## 7      Calidris alpina BIRDS    NA
```

This lists all species that are included in the tag deployment metadata for the project. If there are species that do not make sense, this is likely due to a data entry error when assigning a deployment to a species. You can look for records in your tag metadata that are associated with a particular speciesID using the following code; you would then use the deployID associated with the entry/entries to find and update the deployment record in your project metadata online:

```
filter(df.tagDeps, speciesID == 4780)
```

```
##   deployID tagID projectID status          tsStart          tsEnd
## 1  10517 22867       176 <NA> 2016-09-06 15:35:00 2017-05-18 15:35:00
##   deferSec speciesID bandNumber markerNumber markerType latitude longitude
## 1        NA      4780        <NA> 2641-20877 metal band 51.79861 -80.69139
##   elevation
## 1        NA
##
## 1 Sex:F, Age:HY, Bill28, Tarsus:26.2, Wing Chord:123, Wing Flat:129, Mass:57.7, Flag: (FEW)7P6, Blood
##   id bi tsStartCode tsEndCode           fullID
## 1 NA NA      1L       3L SampleData#272.1:5.3@166.38(M.22867)
##   motusTagDepID
## 1  22867.10517
```

Please remember, any metadata corrections need to be made online

## 4.4 Check Receiver Metadata

There are two sources of receiver metadata in Motus detection data: receivers registered to your own project, and receivers registered to the projects of others. You can access metadata for all receivers in the network, because negative data (i.e., my tag was *not* detected at station x even though station x was active) is often as important as positive data. It also allows you to map where your tags were detected relative to the distribution of receivers throughout the Motus network.

Receiver metadata errors or omissions that you find in your .motus file can only be fixed for receivers registered to your own project.

All users are encouraged to enter complete and accurate receiver metadata for the benefit of the entire network. If you anticipate needing specific information on receiver or antenna deployments for stations deployed by others, please consider using the Motus discussion group (<https://motus.org/discussion/>) to request that other registered users record the receiver deployment details you will need; be specific about the exact receiver deployment details you are interested in, and when and where in the network your tags will be deployed and potentially detected. In the following steps we will:

1. download full receiver metadata across the network;
2. determine number of project receiver deployments;
3. determine timing of project receiver deployments;
4. determine location of network-wide and project receiver deployments;
5. determine completeness and accuracy of receiver metadata.

#### 4.4.1 Download full receiver metadata

Later on in this chapter we will want to map all receivers in the network, so we will now load metadata from all projects, as opposed to simply project 176 as we did above. The `metadata()` function is described in more detail in section 3.10.

```
metadata(sql.motus)
```

#### 4.4.2 Number of project receiver deployments

To see which (if any) receiver deployments are registered to your project, import, subset and summarize the receiver deployment data:

```
tbl.recvDeps <- tbl(sql.motus, "recvDeps")
df.projRecvs <-tbl.recvDeps %>% filter(projectID ==
  proj.num) %>% collect() %>% as.data.frame() %>%
  mutate(tsStart = as_datetime(tsStart, tz = "UTC",
    origin = "1970-01-01"), tsEnd = as_datetime(tsEnd,
    tz = "UTC", origin = "1970-01-01"))

summary(df.projRecvs)

##      deployID      serno      receiverType      deviceID
##  Min.   :1134  Length:29      Length:29      Min.   : 74.0
##  1st Qu.:2289  Class :character  Class :character  1st Qu.:138.5
##  Median :3101  Mode  :character  Mode  :character  Median :280.0
##  Mean   :3007
##  3rd Qu.:4048
##  Max.   :4222
##                               NA's   :2
##      macAddress      status      name
##  Length:29      Length:29      Length:29
##  Class :character  Class :character  Class :character
##  Mode  :character  Mode  :character  Mode  :character
##
##      siteName      fixtureType      latitude      longitude
##  Length:29      Length:29      Min.   :51.15  Min.   :-80.80
##  Class :character  Class :character  1st Qu.:51.29  1st Qu.:-80.57
##  Mode  :character  Mode  :character  Median :51.48  Median :-80.45
##                               Mean   :51.51  Mean   :-80.35
##                               3rd Qu.:51.66  3rd Qu.:-80.12
##                               Max.   :51.88  Max.   :-79.69
##                               NA's   :4     NA's   :4
##      isMobile      tsStart
##  Min.   :0.0000  Min.   :2014-07-12 00:00:00
##  1st Qu.:0.0000  1st Qu.:2015-05-25 00:00:00
##  Median :0.0000  Median :2016-05-18 00:00:00
##  Mean   :0.1379  Mean   :2016-02-20 17:19:14
##  3rd Qu.:0.0000  3rd Qu.:2017-05-16 15:55:00
##  Max.   :1.0000  Max.   :2017-08-20 23:30:00
##
```

```

##      tsEnd                  projectID    elevation
##  Min.   :2014-11-06 00:00:00   Min.   :176     Min.   :-7
##  1st Qu.:2015-09-09 12:00:00  1st Qu.:176     1st Qu.:-4
##  Median :2015-11-03 00:00:00  Median :176     Median :-4
##  Mean   :2016-02-22 05:38:15  Mean   :176     Mean   : 3
##  3rd Qu.:2016-11-29 12:00:00  3rd Qu.:176     3rd Qu.: 0
##  Max.   :2017-08-20 23:30:00  Max.   :176     Max.   :30
##  NA's    :6                   NA's    :24

```

There are 29 receiver deployments registered to the sample project. Four deployments are missing latitude and longitude, and six deployments are missing end dates, which suggests that those receivers are still deployed.

The following code keeps only variables of interest (by removing those we do not need), and arranges the remaining records by receiver ID, latitude, and start date:

```

df.projRecvs %>% mutate(dateStart = date(tsStart)) %>%
  select(-serno, -fixtureType, -macAddress, -tsStart,
         -tsEnd, -elevation, -projectID, -status, -receiverType,
         -siteName) %>% arrange(deviceID, latitude,
                                   dateStart)

```

	deployID	deviceID		name	latitude	longitude	isMobile
## 1	3100	74		Washkaugou	51.1540	-79.8144	0
## 2	2291	75		North Bluff	51.4839	-80.4500	0
## 3	3102	75		North Bluff	51.4839	-80.4501	0
## 4	4051	75		North Bluff	51.4839	-80.4501	0
## 5	4221	75		North Bluff	51.4839	-80.4501	0
## 6	3103	78		Piskwamish	51.6579	-80.5678	0
## 7	4050	78		Piskwamish	51.6580	-80.5679	0
## 8	3101	199		Netitishi	51.2912	-80.1167	0
## 9	4052	199		Netitishi	51.2913	-80.1167	0
## 10	4222	261		North Bluff	51.4839	-80.4501	0
## 11	1140	261		North Bluff	51.4840	-80.4500	0
## 12	1134	280		Longridge	51.8230	-80.6911	0
## 13	2285	280		Longridge	51.8231	-80.6912	0
## 14	3097	280		Longridge	51.8244	-80.6909	0
## 15	4048	280	Halfway Point		51.8753	-80.7973	0
## 16	1135	285		Netitishi	51.2913	-80.1167	0
## 17	2289	285		Netitishi	51.2913	-80.1168	0
## 18	2290	324	Washiskougau Creek		51.1542	-79.8145	0
## 19	1136	324	Washiskougau Creek		51.1545	-79.8137	0
## 20	2286	349		Piskwamish	51.6578	-80.5676	0
## 21	1137	349		Piskwamish	51.6582	-80.5669	0
## 22	3814	527	LR mobile		NA	NA	1
## 23	3813	528	NP mobile		NA	NA	1
## 24	4001	528	BurntPointAerial		NA	NA	1
## 25	4002	528	JamesBayAerial		NA	NA	1
## 26	3099	609	East Point		51.2301	-79.7124	0
## 27	3098	647	Halfway Point		51.8752	-80.7973	0
## 28	4053	NA	East Point		51.3818	-79.6857	0
## 29	4049	NA	Longridge		51.8246	-80.6909	0
			dateStart				
## 1			2016-05-18				
## 2			2015-05-25				
## 3			2016-05-18				
## 4			2017-05-17				

```
## 5 2017-08-20
## 6 2016-05-18
## 7 2017-05-17
## 8 2016-05-18
## 9 2017-05-17
## 10 2017-08-20
## 11 2014-08-01
## 12 2014-07-16
## 13 2015-05-24
## 14 2016-05-17
## 15 2017-05-16
## 16 2014-07-12
## 17 2015-05-25
## 18 2015-05-25
## 19 2014-07-12
## 20 2015-05-24
## 21 2014-07-16
## 22 2015-07-07
## 23 2015-07-06
## 24 2016-07-19
## 25 2016-08-09
## 26 2016-05-18
## 27 2016-05-17
## 28 2017-05-18
## 29 2017-05-16
```

The number of receiver deployments in the metadata should correspond with the number of field deployments.

Looking at the ‘isMobile’ column for the four receiver deployments that are missing latitude and longitude information, it is evident that these are mobile receivers that do not have a fixed position (ie. they have a value of 1). Because they are mobile, coordinates of the deployment aren’t expected, and in this case will remain NA. Receiver deployment coordinates for mobile receivers, when present, are meant to represent the starting point for the deployment.

#### 4.4.3 Timing of project receiver deployments

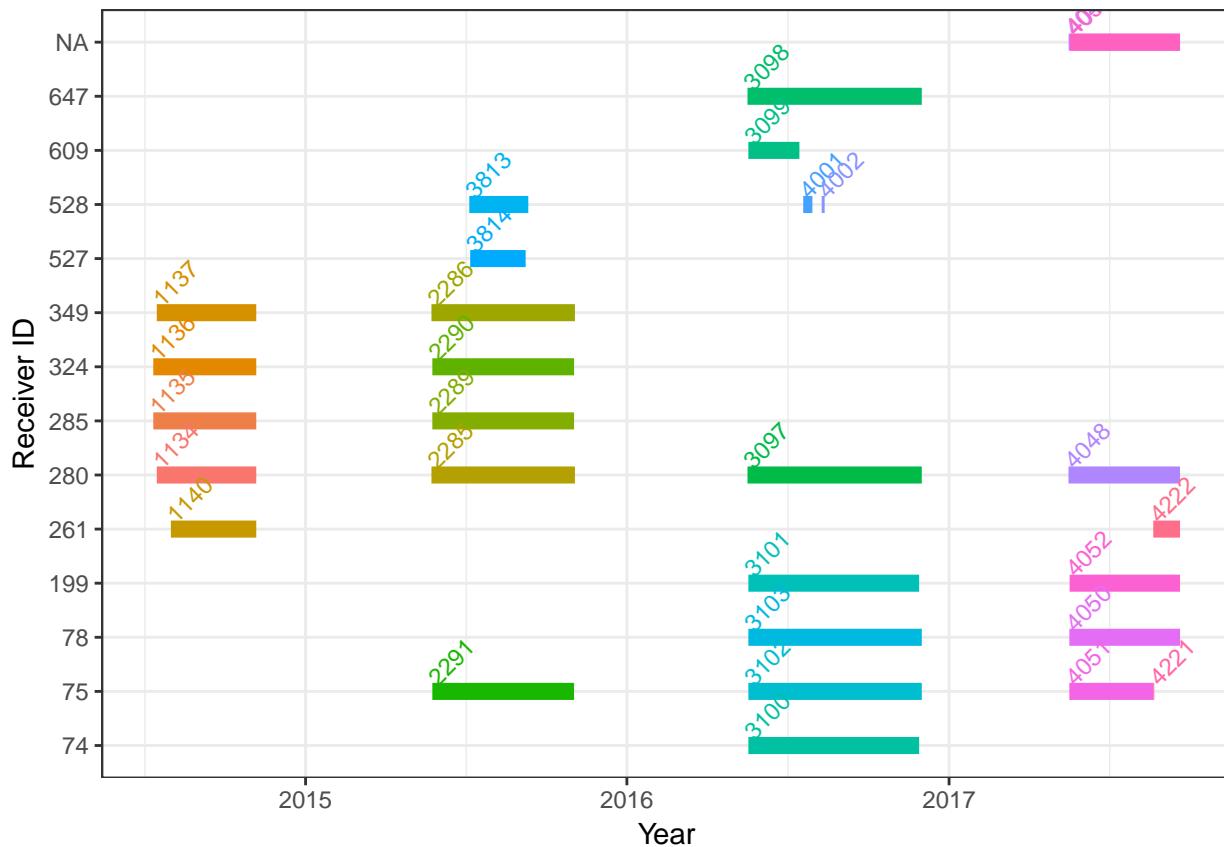
The timing of deployments can be displayed graphically; horizontal line(s) in the following plot show the time span for each receiver (deviceID) deployment registered to the James Bay Shorebird Project. Note that for the two receivers without deployment end dates, the code assigns an arbitrary end date based on the maximum end date of the other receivers plus one month - without this fix, deployments without end dates do not get displayed. Different deployments of the same receiver should not overlap in time:

```
# put data in long format to simplify plotting (or
# use geom_segment)

df.projRecvs.long <- select(df.projRecvs, deviceID,
  deployID, tsStart, tsEnd) %>% tidyrr::gather(when,
  ts, c(tsStart, tsEnd)) %>% mutate(ts = if_else(is.na(ts),
  max(ts, na.rm = TRUE) + duration(1, "month"), ts)) # fake end date

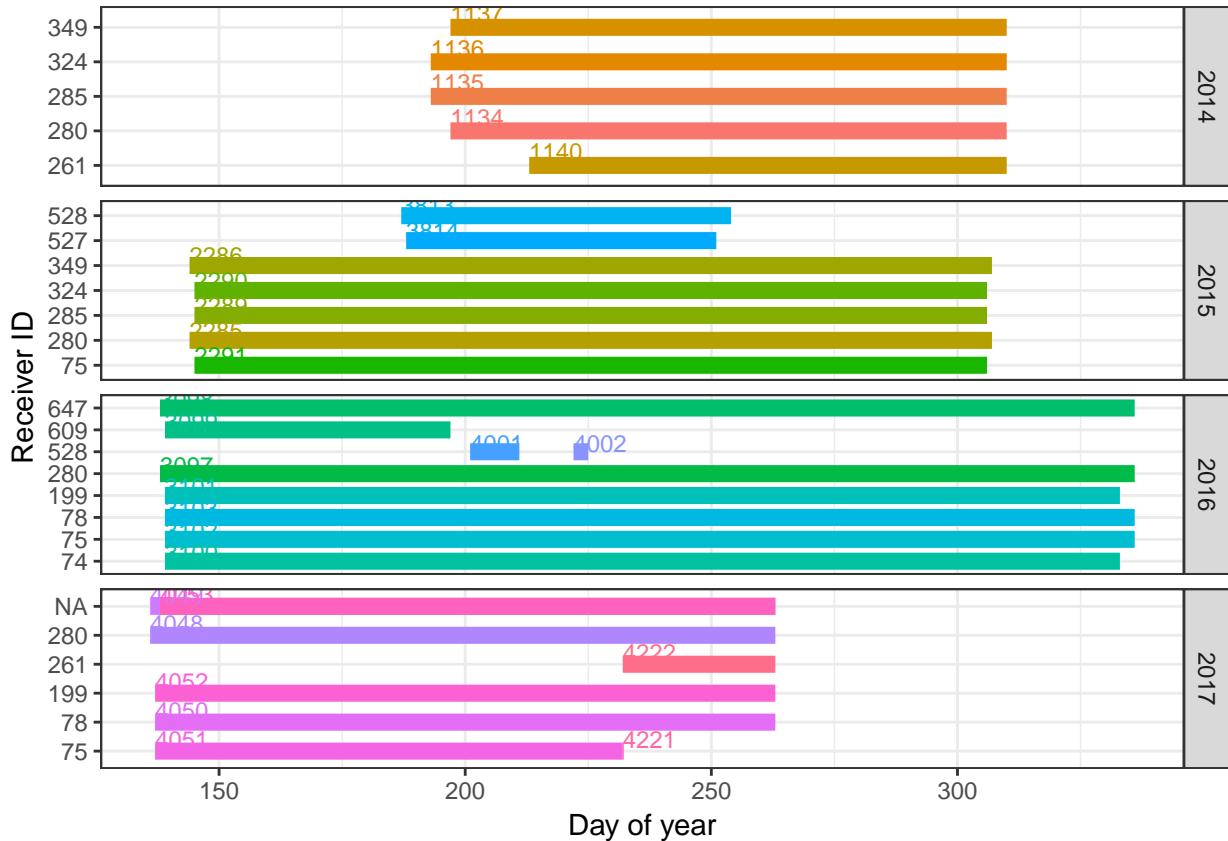
ggplot(df.projRecvs.long, aes(y = as.factor(deviceID),
  x = ts, colour = as.factor(deployID))) + geom_line(lwd = 3) +
  # instead, center to the right
```

```
geom_text(data = filter(df.projRecvs.long, when == "tsStart"), aes(label = deployID), hjust = "left",
nudge_y = 0.2, size = 3, angle = 45) + theme_bw() +
ylab("Receiver ID") + xlab("Year") + theme(legend.position = "none")
```



If you want more detail for a given year (or all years) you can either subset and re-plot, or use the day of year on the x-axis, and facet\_wrap by year.

```
ggplot(df.projRecvs.long, aes(y = as.factor(deviceID),
x = yday(ts), colour = as.factor(deployID))) +
geom_line(lwd = 3) +
# center labels to the left
geom_text(data = filter(df.projRecvs.long, when == "tsStart"), aes(label = deployID), hjust = "left",
nudge_y = 0.4, size = 3) + theme_bw() + ylab("Receiver ID") +
xlab("Day of year") + theme(legend.position = "none") +
facet_grid(year(ts) ~ ., scales = "free")
```



#### 4.4.4 Location of receiver deployments

Maps provide better spatial context than simple plots; the following steps plot the location of Motus receivers on a map of North America, with receivers deployed by the sample project displayed in red.

##### a. Load all receiver metadata

```
df.recvDeps <- tbl.recvDeps %>% collect() %>% as.data.frame() %>%
  mutate(tsStart = as_datetime(tsStart, tz = "UTC",
    origin = "1970-01-01"), tsEnd = as_datetime(tsEnd,
    tz = "UTC", origin = "1970-01-01"))
```

##### b. Load base map files

```
na.lakes <- map_data(map = "lakes")
na.lakes <- mutate(na.lakes, long = long - 360)

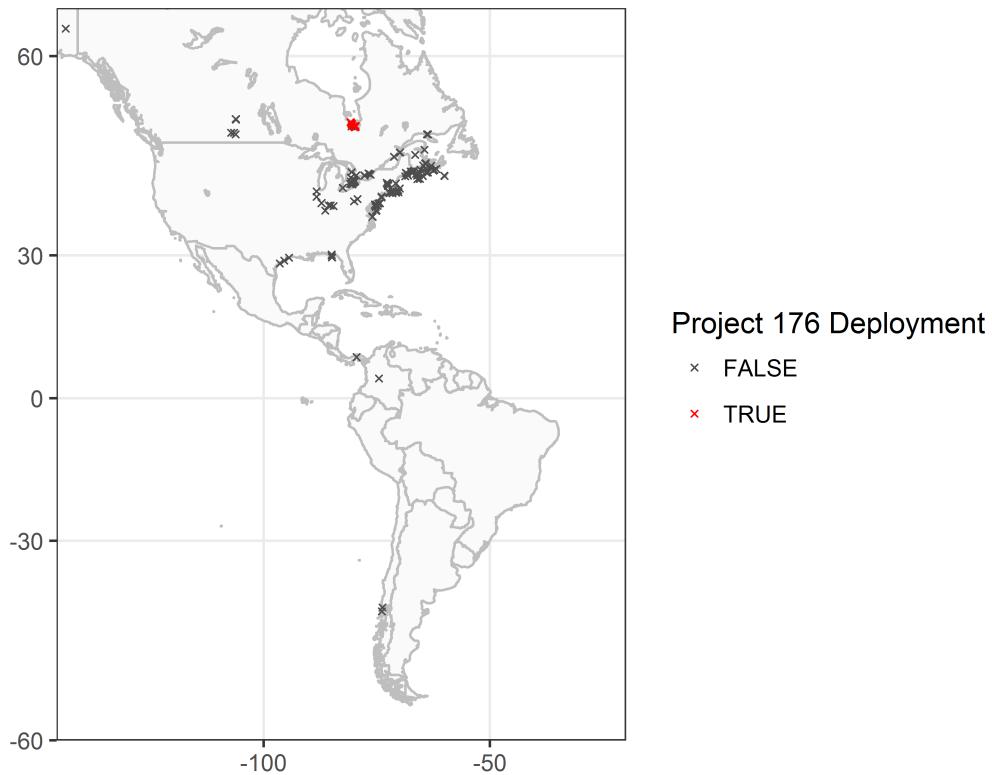
# Include all of the Americas to begin
na.map <- map_data(map = "world2")
na.map <- filter(na.map, region %in% c("Canada", "USA",
  "Mexico", "lakes", "Belize", "Costa Rica", "Panama",
  "Guatemala", "Honduras", "Nicaragua", "El Salvador",
  "Colombia", "Venezuela", "Ecuador", "Peru", "Brazil",
  "Guyana", "Suriname", "Bolivia", "French Guiana",
  "Jamaica", "Cuba", "Haiti", "Dominican Republic",
  "The Bahamas", "Turks and Caicos Islands", "Puerto Rico",
```

```
"British Virgin Islands", "Montserrat", "Dominica",
"Saint Lucia", "Barbados", "Grenada", "Trinidad and Tobago",
"Chile", "Argentina", "Uruguay", "Paraguay")) %>%
mutate(long = long - 360)
```

**c. Map the location of receivers in the Americas** Map showing the location of network-wide receivers (dark grey ‘x’) and receivers deployed by the James Bay Shorebird Project (project 176; red ‘x’).

```
# set map limits using detection locations; ensure
# they include the deployment locations
xmin <- min(df.recvDeps$longitude, na.rm = TRUE) -
  2
xmax <- -20 # restrict to the Americas (excluding a few points in Europe)
ymin <- -60 #min(df.recvDeps$longitude, na.rm = TRUE) - 2
ymax <- max(df.recvDeps$latitude, na.rm = TRUE) + 2

# map
ggplot(na.lakes, aes(long, lat)) + geom_polygon(data = na.map,
  aes(long, lat, group = group), colour = "grey",
  fill = "grey98") + geom_polygon(aes(group = group),
  colour = "grey", fill = "white") + coord_map(projection = "mercator",
  xlim = c(xmin, xmax), ylim = c(ymin, ymax)) + xlab("") +
  ylab("") + theme_bw() + geom_point(data = df.recvDeps,
  aes(longitude, latitude, colour = as.logical(projectID ==
  176)), cex = 0.8, pch = 4) + scale_colour_manual(values = c("grey30",
  "red"), name = "Project 176 Deployment")
```



#### d. Map the location of project specific receivers only

Map of project-specific receivers, created by setting the x-axis (longitude) and y-axis (latitude) map limits using the ‘df.projRecvs’ dataframme created above. Deployments are restricted to those that were active at in 2016.

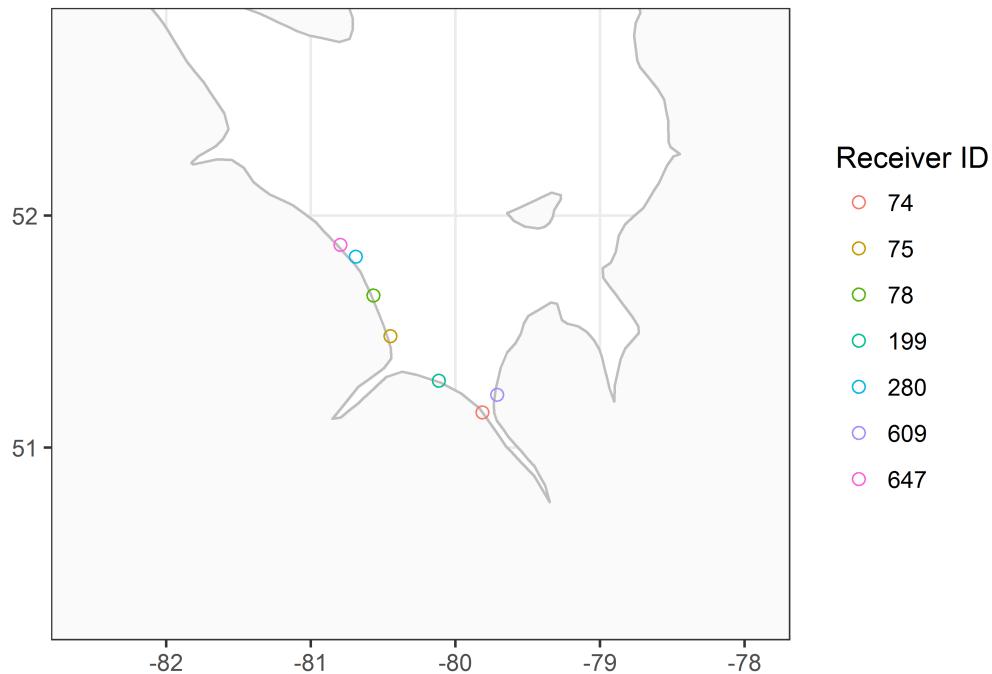
```
# set map limits using detection locations;
# ensure they include the deployment locations
xmin <- min(df.projRecvs$longitude, na.rm = TRUE) - 2
xmax <- max(df.projRecvs$longitude, na.rm = TRUE) + 2
ymin <- min(df.projRecvs$latitude, na.rm = TRUE) - 1
ymax <- max(df.projRecvs$latitude, na.rm = TRUE) + 1

# map
ggplot(na.lakes, aes(long, lat)) +
  geom_polygon(data = na.map,
               aes(long, lat, group=group), colour = "grey", fill="grey98") +
  geom_polygon(aes(group = group), colour = "grey", fill = "white") +
  coord_map(projection="mercator", xlim = c(xmin, xmax), ylim = c(ymin, ymax)) +
#  coord_map(projection="mercator", xlim = c(xmin, xmax), ylim = c(50.154, 52.875)) +
  xlab("") + ylab("") +
  theme_bw() +
  geom_point(data = filter(df.projRecvs,
                           year(tsStart) == 2016,
```

```

!is.na(latitude)), # remove mobile receivers
aes(longitude, latitude, colour = as.factor(deviceID)), cex = 2, pch = 1) +
scale_colour_discrete(name = "Receiver ID")

```



#### 4.4.5 Completeness and accuracy of receiver metadata

Motus users will be concerned primarily with the completeness of metadata for receiver deployments with detection(s) of their tags, because these can directly influence the interpretation of those detections. For example, missing deployment latitude or longitude will result in an unknown location for the tag detection, and missing information on antenna type and/or orientation can impede the estimation of flight or departure orientation.

In many cases, however, metadata for receiver deployments *without* tag detections can also be useful, for example to estimate probability of detecting an animal that passes within range of a station.

In this section, the focus is on metadata for receivers registered to a particular project. Depending on your interests, these summaries can be applied to a larger group of receivers, e.g., all receivers with detections or all receivers within certain geographic limits (with or without detections).

##### a. Load receiver and antenna metadata

```

# antenna metadata for ALL Motus antenna
# deployments; to simplify, keep only the variables
# of interest.
tbl.antDeps <- tbl(sql.motus, "antDeps")
df.antDeps <-tbl.antDeps %>% select(deployID, port,
antennaType, bearing, heightMeters) %>% collect() %>%
as.data.frame()

# receiver deployments; select variables of

```

```
# interest
df.recvDeps <- df.recvDeps %>% select(deployID, receiverType,
  deviceID, name, latitude, longitude, isMobile,
  tsStart, tsEnd, projectID, elevation)

df.stationDeps <- left_join(df.recvDeps, df.antDeps,
  by = "deployID")
```

Subset these to receivers registered to a project:

```
df.stationDeps <- filter(df.stationDeps, projectID ==
  proj.num)
```

### b. Look at range of metadata values

Use summary() to get a general idea of the distribution of the variables in the data.

```
summary(df.stationDeps)
```

```
##      deployID    receiverType        deviceID       name
##  Min.   :1134    Length:91        Min.   : 74.0  Length:91
##  1st Qu.:2289    Class :character  1st Qu.: 78.0  Class :character
##  Median :3100    Mode  :character  Median :280.0  Mode  :character
##  Mean   :2996                Mean   :258.7
##  3rd Qu.:4049                3rd Qu.:324.0
##  Max.   :4222                Max.   :647.0
##                    NA's   :6
##      latitude     longitude      isMobile
##  Min.   :51.15    Min.   :-80.80    Min.   :0.00000
##  1st Qu.:51.29    1st Qu.:-80.57   1st Qu.:0.00000
##  Median :51.48    Median :-80.45   Median :0.00000
##  Mean   :51.52    Mean   :-80.36   Mean   :0.04396
##  3rd Qu.:51.66    3rd Qu.:-80.12   3rd Qu.:0.00000
##  Max.   :51.88    Max.   :-79.69   Max.   :1.00000
##  NA's   :4        NA's   :4
##      tsStart          tsEnd        projectID
##  Min.   :2014-07-12 00:00:00  Min.   :2014-11-06 00:00:00  Min.   :176
##  1st Qu.:2015-05-25 00:00:00  1st Qu.:2015-11-02 00:00:00  1st Qu.:176
##  Median :2016-05-18 00:00:00  Median :2016-07-15 00:00:00  Median :176
##  Mean   :2016-03-28 08:40:07  Mean   :2016-04-04 01:07:25  Mean   :176
##  3rd Qu.:2017-05-16 19:42:00  3rd Qu.:2016-12-01 00:00:00  3rd Qu.:176
##  Max.   :2017-08-20 23:30:00  Max.   :2017-08-20 23:30:00  Max.   :176
##  NA's   :21
##      elevation       port      antennaType      bearing
##  Min.   :-7.000    Min.   :1.000  Length:91        Min.   : 0.0
##  1st Qu.:-4.000    1st Qu.:1.000  Class :character  1st Qu.: 50.0
##  Median :-4.000    Median :2.000  Mode  :character  Median :145.0
##  Mean   : 4.235    Mean   :2.231                Mean   :152.6
##  3rd Qu.: 0.000    3rd Qu.:3.000                3rd Qu.:290.0
##  Max.   :30.000    Max.   :4.000                Max.   :357.5
##  NA's   :74
##      heightMeters
##  Min.   :4.800
##  1st Qu.:5.600
##  Median :5.800
##  Mean   :5.832
```

```
## 3rd Qu.:6.000
## Max. :6.200
## NA's :4
```

There are the 4 deployments with missing latitude and longitude associated with the four deployments of mobile receivers that we saw earlier.

Elevation is missing from 74 of 91 records, but elevation is not a required field, and can be estimated from other sources, or directly in R (for example, see <https://stackoverflow.com/questions/8973695/conversion-for-latitude-longitude-to-altitude-in-r>).

Antenna bearing is missing from 18 of 91 records, and height of the antenna(s) is missing for 4 of 91 records. Subset the records with missing antenna bearing to see if these can be fixed:

```
filter(df.stationDeps, is.na(bearing)) %>% select(-elevation,
  -deviceID, -tsEnd)
```

	deployID	receiverType	name	latitude	longitude	isMobile
## 1	3097	SENSORGNAME	Longridge	51.8244	-80.6909	0
## 2	3098	SENSORGNAME	Halfway Point	51.8752	-80.7973	0
## 3	3099	SENSORGNAME	East Point	51.2301	-79.7124	0
## 4	3100	SENSORGNAME	Washkaugou	51.1540	-79.8144	0
## 5	3101	SENSORGNAME	Netitishi	51.2912	-80.1167	0
## 6	3102	SENSORGNAME	North Bluff	51.4839	-80.4501	0
## 7	3103	SENSORGNAME	Piskwamish	51.6579	-80.5678	0
## 8	3813	LOTEKSRX800	NP mobile	NA	NA	1
## 9	3814	LOTEKSRX800	LR mobile	NA	NA	1
## 10	4001	LOTEKSRX800	BurntPointAerial	NA	NA	1
## 11	4002	LOTEKSRX800	JamesBayAerial	NA	NA	1
## 12	4048	SENSORGNAME	Halfway Point	51.8753	-80.7973	0
## 13	4049	<NA>	Longridge	51.8246	-80.6909	0
## 14	4050	SENSORGNAME	Piskwamish	51.6580	-80.5679	0
## 15	4051	SENSORGNAME	North Bluff	51.4839	-80.4501	0
## 16	4052	SENSORGNAME	Netitishi	51.2913	-80.1167	0
## 17	4221	SENSORGNAME	North Bluff	51.4839	-80.4501	0
## 18	4222	SENSORGNAME	North Bluff	51.4839	-80.4501	0
	tsStart	projectID	port	antennaType	bearing	heightMeters
## 1	2016-05-17 00:00:00	176	4	omni-whip	NA	6.0
## 2	2016-05-17 00:00:00	176	4	omni-whip	NA	6.0
## 3	2016-05-18 00:00:00	176	4	omni-whip	NA	6.0
## 4	2016-05-18 00:00:00	176	4	omni-whip	NA	6.0
## 5	2016-05-18 00:00:00	176	4	omni-whip	NA	6.0
## 6	2016-05-18 00:00:00	176	4	omni-whip	NA	6.0
## 7	2016-05-18 00:00:00	176	4	omni-whip	NA	6.0
## 8	2015-07-06 00:00:00	176	1	yagi-3	NA	NA
## 9	2015-07-07 00:00:00	176	1	yagi-3	NA	NA
## 10	2016-07-19 08:00:00	176	1	yagi-3	NA	NA
## 11	2016-08-09 07:15:00	176	1	yagi-3	NA	NA
## 12	2017-05-16 15:55:00	176	3	omni-whip	NA	6.2
## 13	2017-05-16 19:42:00	176	4	omni-whip	NA	6.2
## 14	2017-05-17 15:19:00	176	4	omni-whip	NA	6.2
## 15	2017-05-17 15:00:00	176	4	omni-whip	NA	6.2
## 16	2017-05-17 22:47:00	176	4	omni-whip	NA	6.2
## 17	2017-08-20 23:30:00	176	4	omni-whip	NA	6.2
## 18	2017-08-20 23:30:00	176	4	omni-whip	NA	6.2

Receiver deployments with missing antenna bearing(s) are restricted to deployments of omni-directional

antennas or mobile receivers, and so the missing values make sense. These records also show that the four records with missing antenna height are also associated with the four mobile receivers, and so again the missing values make sense and do not need to be fixed.

Remember that any missing metadata needs to be corrected online. Metadata corrected online will automatically be corrected in your detection files. If you have already downloaded your detection data, you can update the existing file to include new metadata and detections (see sections 3.9, 3.7).

In the next chapter we will examine our data for false positives, and remove detections of ambiguous tags.

# Chapter 5

## Data Cleaning

There are three sources of ‘error’ that can result in tag detections appearing in your database that are incorrect.

First, random radio noise (‘static’) can be detected and interpreted to be the transmission of a tag. These are called ‘false positives’.

Second, despite our best efforts to avoid it, duplicate tags are sometimes transmitting in the network at the same time. When two tags are deployed at the same time that have the same ID code, burst interval, and nominal transmit frequency, it results in situations where the detections may belong to either tag. If that happens, we must rely on contextual information to separate them (if we can). We term these ‘Ambiguous tags’.

Third, a tag can appear to be present when two tags are transmitting at the same time that by chance produce a signal that looks like a third tag that is not in fact present. Such tags are most common at roosting sites or breeding colonies, where many tags are transmitting simultaneously. We term these ‘Aliased tags’. We do not deal explicitly with Aliased tags in this chapter; we are working on a way to globally identify them and eliminate them from the data. We mention them here because you may encounter situations with what appear to be highly plausible detections that don’t make biological sense. Please contact us if you think you have some of these Aliased tag detections in your database.

The *goal of this chapter* is to provide you with the tools you need to check your data for false detections, and remove them from your data. We do so by providing example workflows that deal with ‘false positives’ and ‘ambiguous tags’ in the following steps:

- 1) Run a preliminary filter to remove all detections with runLen of 2. A run is a group of consecutive detections of a tag detected on a single antenna at a single receiver. In general, a detection with a run length of 2 (i.e., 2 bursts) has a high probability of being a false positive detection. With the exception of a few ‘quiet’ stations with little noise, we generally recommend that you filter out all detections with a run length of 2. However, because you will likely lose some true detections in the process, we also recommend that after a full analysis of your data, you return to these detections and examine them individually, to determine (usually contextually) if they can be considered real.
- 2) Determine how many of your tag detections may be ambiguous detections.
- 3) Provide a workflow for examining individual tags, and determine if runs in those tags are errors.
- 4) Filter errors from your data.

## 5.1 Load required packages

Follow the instructions in Chapter 2 to install the following packages before loading, if they are not already installed.

```
Sys.setenv(tz = "GMT")

# library(devtools)
library(motus)
library(tidyverse)
library(lubridate)
# library(rworldmap) # for mapping
```

## 5.2 Load detections data

Recall from Chapter 3 that when accessing the sample database, you will need to input ‘motus.sample’ in the R console as both username and password when prompted by the tagme() user authentication process. This section assumes you have already completed the initial sample data download.

When accessing the alltags table, we remove some unnecessary variables to reduce the overall size of the data set and make it easier to work with. **This is particularly important for large, unwieldy projects**; details on how to view the variables in a tbl, and how to filter and subset prior to collecting data into a dataframe can be found in Chapter ???. We then create receiver latitude and longitude variables (‘recvLat’, ‘recvLon’, ‘recvAlt’) based on the coordinates recorded by the receiver GPS (‘gpsLat’, ‘gpsLon’, ‘gpdAlt’), and where those are not available, infilled with coordinates from the receiver deployment metadata (‘recvDeployLat’, ‘recvDeployLon’, ‘recvDeployAlt’). We use the ‘collect()’ and ‘as.data.frame()’ statements to transform the dataframe into a ‘flat’ file, and then transform all time stamp variables from seconds since January 1 1970 to datetime (POSIXct) format. Finally, we create ‘receiver names’ from the latitude and longitude variables for those receivers in the database that do not have these values filled in.

```
proj.num <- 176

# Load detection data, select variables, create latitude variables, and transform to
# flat file. We also fix up some sites that are missing receiver deployment data, or
# do not have names. As more users explore (and fix!) their metadata, these missing
# values should begin to disappear.
sql.motus <- tagme(proj.num, update = TRUE, dir = "./data/")
tbl.alltags <- tbl(sql.motus, "alltags")

df.alltags <- tbl.alltags %>%
  mutate(recvLat = if_else((is.na(gpsLat)|gpsLat == 0),
                           recvDeployLat, gpsLat),
         recvLon = if_else((is.na(gpsLon)|gpsLon == 0),
                           recvDeployLon, gpsLon),
         recvAlt = if_else(is.na(gpsAlt), recvDeployAlt, gpsAlt)) %>%
  select(-noise, -slop, -burstSlop, -done, -bootnum, -mfgID,
         -codeSet, -mfg, -nomFreq, -markerNumber, -markerType,
         -tagDeployComments, -fullID, -deviceID, -recvDeployLat,
         -recvDeployLon, -recvDeployAlt, -speciesGroup, -gpsLat,
         -gpsLon, -recvAlt, -recvSiteName) %>%
  collect() %>%
  as.data.frame() %>%
  mutate(ts = as_datetime(ts), # work with dates AFTER transforming to flat file
```

```

tagDeployStart = as_datetime(tagDeployStart),
tagDeployEnd = as_datetime(tagDeployEnd),
recvLat = plyr::round_any(recvLat, 0.05),
recvLon = plyr::round_any(recvLon, 0.05),
recvDeployName = ifelse(is.na(recvDeployName),
                        paste(recvLat, recvLon, sep=":"),  

                        recvDeployName)

# Note that in the select statement, you can just select the variables you need
# e.g.: select(runID, ts, sig, freqsd, motusTagID, ambigID, runLen, tagProjID,
#              tagDeployStart, tagDeployEnd, etc.)

```

## 5.3 Preliminary data checks

Prior to filtering the data, we do a few summaries and plots of the data.

### 5.3.1 Summarize tag detections

First, determine which project tags have detections, and determine how many are of run length 2. There are several reasons why deployed tags might not be detected, including:

- 1) The tag was not properly activated on deployment. To avoid this, always check that a tag is active using a hand-held receiver before attaching the tag to your study animal and releasing it.
- 2) An animal with a properly activated tag might not have passed within range of a receiving station. Study designs that incorporate strategic placement of receivers to meet project goals can improve the probability of a tag being detected.
- 3) Missing or incorrect tag deployment metadata in the Motus database can result in the data processing algorithm not ‘looking’ for your tag at the time the tag was deployed, or at all. Please ensure your tag metadata are entered correctly.

Before going further, **please check whether any of your tags were deployed more than once**, as described in section 4.3.4. If so, you will need to use ‘tagDeployID’ or a combination of ‘motusTagID’ and ‘tagDeployID’ to uniquely define detections associated with a tag deployment (either will do, but combining the two fields will let you know which tagID is associated with each deployment).

In the sample data, all tags were deployed only once, and so we use the ‘motusTagID’ as a unique identifier for a tag deployment in all R code throughout the book.

Using the following code, we see there are detections for 18 tags deployed by the sample project, and that many have run lengths of 2 (TRUE):

```

df.alltags %>%
  filter(tagProjID == proj.num) %>% # subset to include only tags registered to project
  mutate(rl.gt.2 = runLen == 2) %>%
  group_by(motusTagID, rl.gt.2) %>%
  tally() %>%
  spread(key = rl.gt.2, value=n)

## # A tibble: 18 x 3
## # Groups:   motusTagID [18]
##   motusTagID `FALSE` `TRUE`
##   *      <int>    <int>   <int>

```

```
## 1    16011    125     2
## 2    16035    454     2
## 3    16036    106    12
## 4    16037   1307    46
## 5    16038     84    78
## 6    16039   1098    28
## 7    16044    289    16
## 8    16047    773    66
## 9    16048     84    14
## 10   16052    133    26
## 11   17357    277    12
## 12   19129    568   720
## 13   22867   5545   222
## 14   22897  34308   488
## 15   22902    2815   108
## 16   22905   25684   326
## 17   23316    5518   216
## 18   23319   22471   288
```

Although some of these may be valid detections, we have found it simpler to just remove them from our analysis, and possibly revisit them at a later stage. We therefore filter on runLen ( $> 2$ ) for most subsequent operations. We save these in a block to add to our other filters later.

```
df.alltags.sub <- filter(df.alltags, runLen > 2)

df.block.0 <- filter(df.alltags, runLen == 2) %>% select(motusTagID,
  runID) %>% distinct()
```

An initial view of the data is best achieved by plotting. We will show you later how to plot detections on a map, but we prefer a simpler approach first; plotting detections through time by both latitude and longitude. First however, we should simplify the data. If we don't, we risk trying to plot thousands or millions of points on a plot (which can take a long time). We'll do this by creating a little function here, since we will use this operation again in future steps.

Note that we need to remove about 150 detections, because there is no geographic data associated with the receiver metadata, and so no way to determine the location of those detections. Do a simple check to see if these receivers belong to you, and if so, please **fix the metadata online!**

```
filter(df.alltags.sub, is.na(recvLat)) %>% select(recvLat,
  recvLon, recvDeployName, recvDeployID, recv, recvProjID,
  recvProjName) %>% distinct()
```

```
##   recvLat recvLon recvDeployName recvDeployID          recv recvProjID
## 1      NA      NA        NP mobile       3813  Lotek-280      176
## 2      NA      NA        NA:NA        NA SG-1415BBBK0382      NA
## 3      NA      NA        NA:NA        NA SG-2814BBBK0547      NA
##   recvProjName
## 1 SampleData
## 2 <NA>
## 3 <NA>
```

### Simplify the data for plotting

```
# simplify the data by summarizing by the runID.
# If you want to summarize at a finer/coarser scale, you can also create other groups.
# The simplest alternative is a rounded timestamp variable; for example by using
# mutate(ts.h = plyr::round_any(ts, 3600) function call.
```

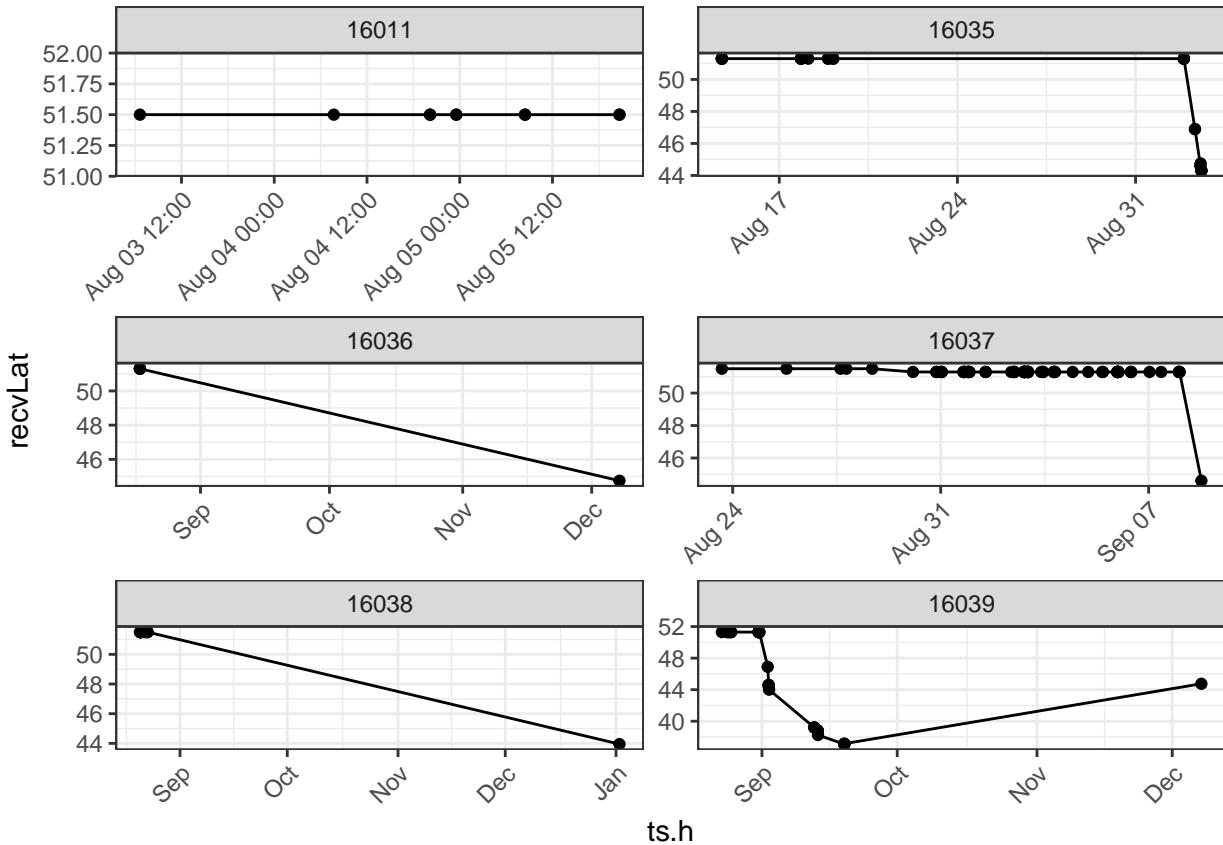
```
# Other options are to just use date (e.g date = as_date(ts))

#
fun.getpath <- function(df)
{
  df %>%
    filter(tagProjID == proj.num, # keep only tags registered to the sample project
          !is.na(recvLat) | !(recvLat == 0)) %>% # drops data without lon/lat
    group_by(motusTagID, runID, recvDeployName, ambigID,
              tagDeployLon, tagDeployLat, recvLat, recvLon) %>%
    #summarizing by runID to get max run length and mean time stamp:
    summarize(max.runLen = max(runLen), ts.h = mean(ts)) %>%
    arrange(motusTagID, ts.h)
} # end of function call

df.alltags.path <- fun.getpath(df.alltags.sub)
```

We would initially plot a subset of tags by either latitude or longitude, to get an overview of where there might be issues. Here, to simplify the example, we plot only six tags. We avoid examining the ambiguous tags for now.

```
p <- ggplot(data = filter(df.alltags.path, motusTagID %in%
                           c(16011, 16035, 16036, 16037, 16038, 16039)), aes(ts.h,
                           recvLat))
p + geom_point() + geom_path() + theme_bw() + facet_wrap(~motusTagID,
               scales = "free", ncol = 2) + theme(axis.text.x = element_text(angle = 45,
               vjust = 1, hjust = 1))
```



It is immediately apparent that there may be an issue with some tags showing up around 44 degrees during in the winter, which is possible but unlikely for the sample project's shorebirds. Let's examine these tags in more detail by examining the runs in the data frame that are associated with detections in December and January.

```
filter(df.alltags.sub, month(ts) %in% c(12, 1), motusTagID %in%
  c(16036, 16038, 16039)) %>% group_by(recvDeployName,
  month(ts), runLen) %>% summarize(n = length(ts),
  n.tags = length(unique(motusTagID)))
```

```
## # A tibble: 2 x 5
## # Groups:   recvDeployName, month(ts) [?]
##   recvDeployName `month(ts)` runLen     n n.tags
##   <chr>           <dbl>    <int> <int> <int>
## 1 Sable West Light 2      1.00      3     3     1
## 2 Swallowtail        12.0       3     6     2
```

These detections were at sites around the Canadian Maritimes (Sable Island, Nova Scotia; Grand Manan, New Brunswick) and were run lengths of 3. These are indicators of likely false positives. We'll now start a tally of the particular runs involved, so that we can collate them in to a filter later.

If you are interested, you can re-run the code above, but on the full data frame (df.alltags) containing run lengths of 2. You will see that there are additional false positive detections at these sites, that are already eliminated by filtering on runLen > 2. These additional detections provide further evidence that these sites experienced some radio noise during these particular months, resulting in some false positive detections.

You may also be interested more generally in exploring which data have only short run lengths. For example, the following code shows the maximum run length at all sites by month (for those runs with runLen > 2).

```
df.alltags.sub %>% mutate(month = month(ts)) %>% group_by(recvDeployName,
month) %>% summarize(max.rl = max(runLen)) %>%
spread(key = month, value = max.rl)
```

```
## # A tibble: 49 x 9
## # Groups:   recvDeployName [49]
##   recvDeployName `1` `3` `4` `5` `8` `9` `10` `12`
##   * <chr>        <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Assateague State Park     NA    NA    NA NA      NA  6.00 NA     NA
## 2 BennettMeadow            NA    NA    NA NA      NA NA  11.0  NA
## 3 Binbrook_Conservation_~  NA    NA    NA  3.00    NA NA    NA    NA
## 4 BISE                      NA    NA    NA NA      NA NA    6.00  NA
## 5 Bombay Hook               NA    NA    NA NA      NA 53.0 NA     NA
## 6 Brier2                    NA    NA    NA NA      NA 29.0 NA     NA
## 7 BSC HQ                     NA    NA    NA 21.0    NA NA    NA    NA
## 8 BULL                      NA    NA    NA NA      NA 38.0  5.00  NA
## 9 Comeau (Marshalltown)    NA    NA    NA NA      NA  4.00 NA     NA
## 10 CONY                     NA   NA    NA NA      NA NA    7.00  NA
## # ... with 39 more rows
```

Alternatively, you can produce a list of sites where the maximum run length of detections was never greater than (say) 4, which may sometimes (but not always!) indicate they are simply false detections.

```
df.alltags.sub %>% mutate(month = month(ts)) %>% group_by(recvDeployName,
month) %>% summarize(max.rl = max(runLen)) %>%
filter(max.rl < 5) %>% spread(key = month, value = max.rl)
```

```
## # A tibble: 12 x 8
## # Groups:   recvDeployName [12]
##   recvDeployName `1` `3` `4` `5` `9` `10` `12`
##   * <chr>        <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Binbrook_Conservation_Area NA    NA    NA  3.00 NA    NA    NA
## 2 Comeau (Marshalltown)    NA    NA    NA    NA    4.00 NA    NA
## 3 Koffler                   NA    NA    3.00 NA    NA    NA    NA
## 4 LLICALDAD                NA    4.00 NA    NA    NA    NA    NA
## 5 MountToby                 NA    NA    NA    NA    NA    3.00 NA
## 6 NP mobile                  NA    NA    NA    NA    3.00 NA    NA
## 7 Old Cut                    NA    NA    4.00 NA    NA    NA    NA
## 8 OSCT                      NA    NA    NA    NA    NA    3.00 NA
## 9 Quempillen (Chile)       NA    3.00 NA    NA    NA    NA    NA
## 10 Sable West Light 2       3.00 NA    NA    NA    NA    NA    NA
## 11 Swallowtail              NA    NA    NA    NA    NA    NA    3.00
## 12 TRUS                      NA    NA    NA    NA    NA    4.00 NA
```

It is impossible to go through every possible issue that you may encounter here. Users are strongly encouraged to explore their data fully, and make reasoned decisions on which detections are unlikely or indeterminate. Through the rest of this chapter we will show you how to collect these runs, and apply them to your data prior to analysis.

To start, we'll create a data frame that contains the motusTagIDs and runIDs for the false positives identified above.

We will then re-create the plot With the newly filtered data.

```
# create the filter
df.block.1 <- filter(df.alltags.sub, month(ts) %in%
c(12, 1), motusTagID %in% c(16036, 16038, 16039)) %>%
```

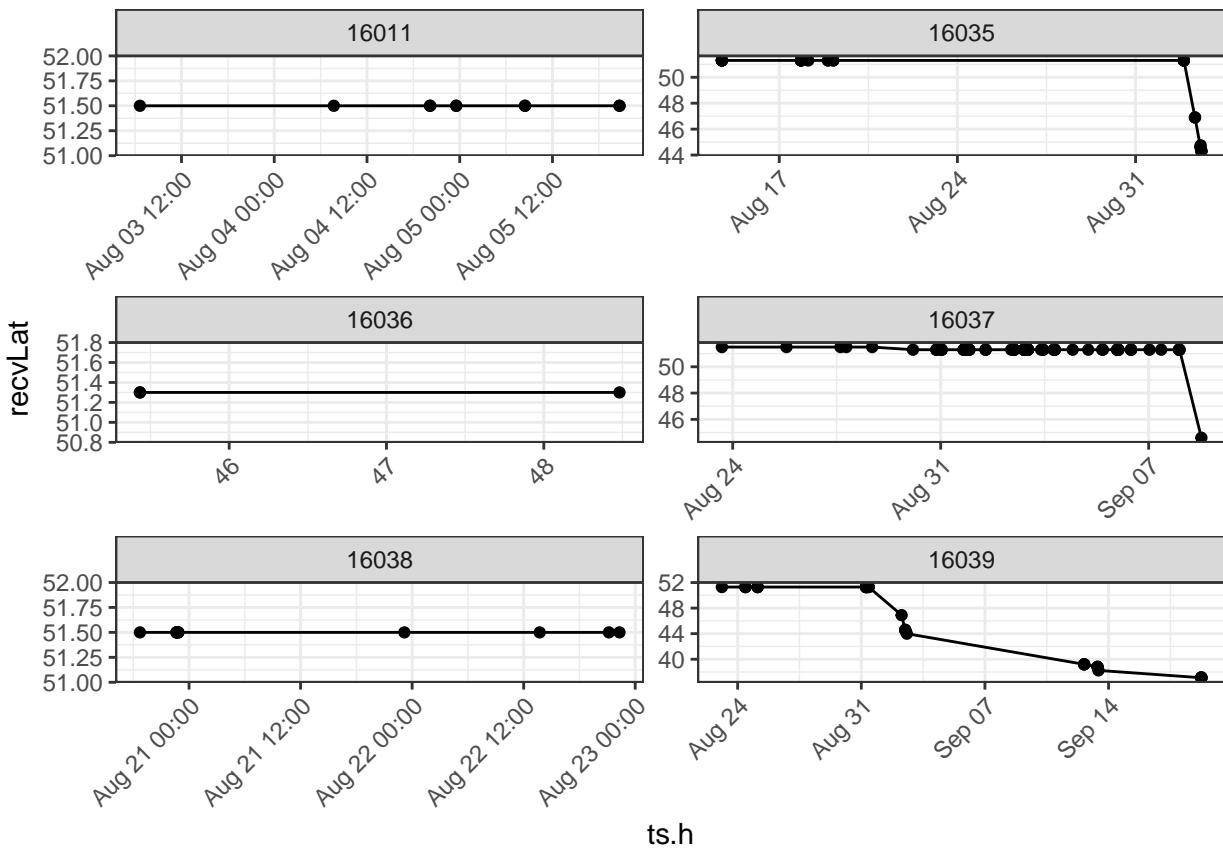
```

  select(motusTagID, runID) %>% distinct()

# use the function we created earlier to make a new
# 'path' data frame for plotting
df.alltags.path <- fun.getpath(filter(df.alltags.sub,
  motusTagID %in% c(16011, 16035, 16036, 16037, 16038,
  16039), !(runID %in% df.block.1$runID)))

p <- ggplot(data = df.alltags.path, aes(ts.h, recvLat))
p + geom_point() + geom_path() + theme_bw() + facet_wrap(~motusTagID,
  scales = "free", ncol = 2) + theme(axis.text.x = element_text(angle = 45,
  vjust = 1, hjust = 1))

```



We can see that most of the remaining detections now appear to make more sense, with tags 16035, 16037 and 16039 having been detected during migration, in what appears to be a reasonable latitudinal progression with time, and the other three tags which were not detected very far away from their deployment location.

The reader is encouraged to explore the rest of the tags within this group, to determine if there are additional false positives.

## 5.4 Examining ambiguous detections

Before we go further, we need to check to see if any tags have ambiguous detections. If there are, we will need to explore them, and create additional filters to remove detections from our database.

### Are any of your tags associated with ambiguous detections?

The ‘clarify()’ function in the motusClient R package provides a summary of ambiguities in the detections data. Each ambigID refers to a selection of detections that could belong to one or more (up to 6) motusTagIDs, which are listed in the id1 to id6 fields:

```
clarify(sql.motus)
```

```
##   ambigID numHits   id1                      fullID1   id2
## 1      -56     5734 22867 SampleData#272.1:5.3@166.38(M.22867) 23316
## 2     -106      279 17021           Selva#172:6.1@166.38(M.17021) 17357
## 3     -114       86 22897 SampleData#303.1:5.3@166.38(M.22897) 24298
## 4     -134     22749 22905   SampleData#301:5.3@166.38(M.22905) 23319
## 5     -171      2074 22778 RBrownAMWD#308:5.3@166.38(M.22778) 22902
## 6     -337        4 10811      Niles#152:6.1@166.38(M.10811) 16011
##                               fullID2   id3
## 1 SampleData#272:5.3@166.38(M.23316) NA
## 2 SampleData#172:6.1@166.38(M.17357) NA
## 3 NEONICS#303:5.3@166.38(M.24298) NA
## 4 SampleData#301.1:5.3@166.38(M.23319) NA
## 5 SampleData#308.1:5.3@166.38(M.22902) 24303
## 6 SampleData#152:6.1@166.38(M.16011) NA
##                               fullID3 id4 fullID4 id5 fullID5 id6 fullID6
## 1 <NA>    NA <NA>    NA <NA>    NA <NA>
## 2 <NA>    NA <NA>    NA <NA>    NA <NA>
## 3 <NA>    NA <NA>    NA <NA>    NA <NA>
## 4 <NA>    NA <NA>    NA <NA>    NA <NA>
## 5 NEONICS#308:5.3@166.38(M.24303) NA <NA>    NA <NA>    NA <NA>
## 6 <NA>    NA <NA>    NA <NA>    NA <NA>
##   motusTagID tsStart tsEnd
## 1        NA      NA     NA
## 2        NA      NA     NA
## 3        NA      NA     NA
## 4        NA      NA     NA
## 5        NA      NA     NA
## 6        NA      NA     NA
```

We can see that there are six tags with ambiguous detections within this data set. Detections associated with five of the six ambigIDs could belong to one of two tags, and detections associated with one ambigID (-171) could belong to one of three tags. The fullID fields list the project names associated with the duplicate tags (e.g., “SampleData”, “Selva”, “Niles”), along with features of the tags (manufacturer tag ID, burst, and transmit frequency).

Let’s get a vector of these, and do some plots to see where there may be issues.

```
df.ambigTags <- select(df.alltags.sub, ambigID, motusTagID) %>%
  filter(!is.na(ambigID)) %>% distinct()
```

Using our getpath function, we’ll create paths and then plot these detections. We’ll add some information to the plot, showing where (in time) the tags are actually ambiguous. We can then inspect the overall plots (or portions of them) to determine if we can contextually unambiguously assign a detection of an ambiguous tag to a single deployment.

```
df.alltags.path <- fun.getpath(filter(df.alltags.sub,
  motusTagID %in% df.ambigTags$motusTagID, tagProjID ==
  proj.num)) %>% # create a boolean variable for ambiguous
# detections:
```

```

mutate(Ambiguous = !(is.na(ambigID)))

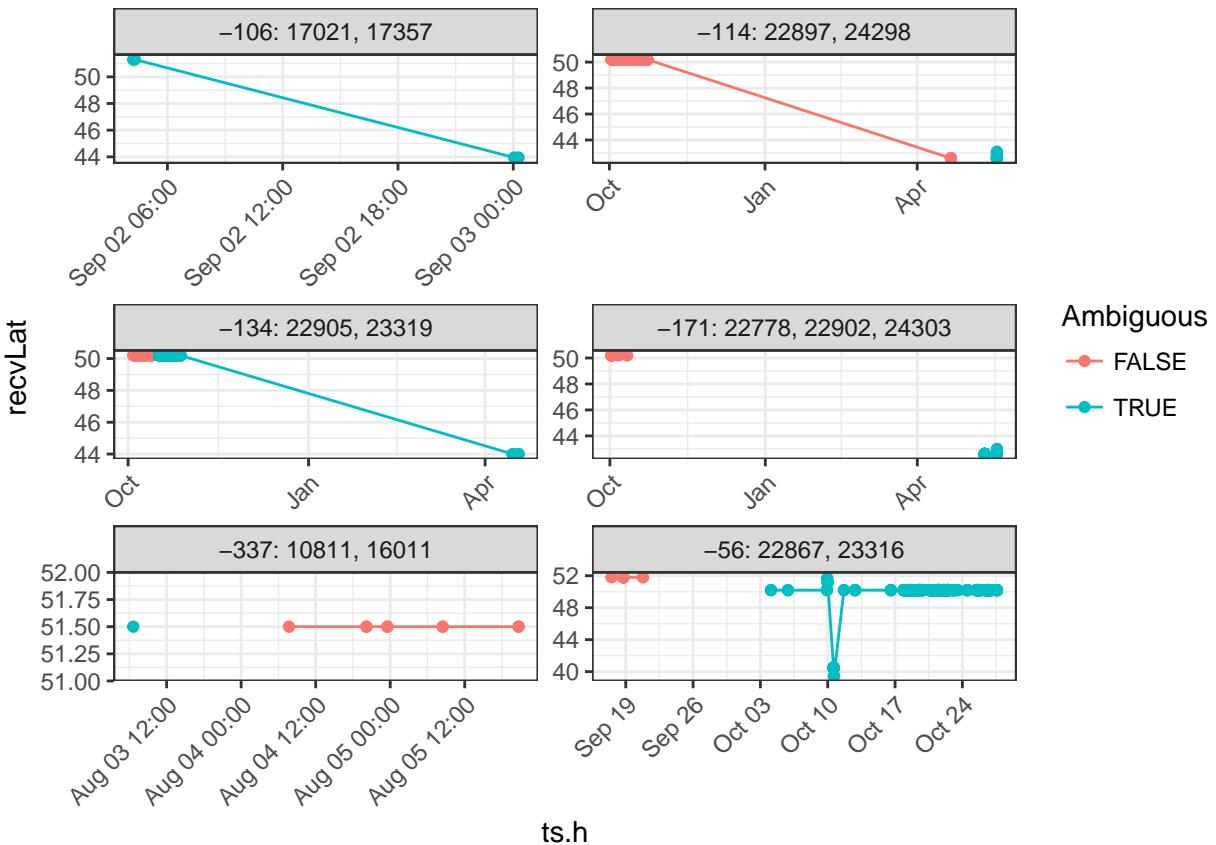
# to put all ambiguous tags from the same project
# on the same plot together, we need to create a
# new 'ambig tag' variable we call 'newID'.

ambigTags.2 <- filter(df.alltags.sub) %>% select(ambigID,
  motusTagID) %>% filter(!is.na(ambigID)) %>% distinct() %>%
  group_by(ambigID) %>% summarize(newID = paste(unique(ambigID),
  toString(motusTagID), sep = ": ")) %>% left_join(df.ambigTags,
  by = "ambigID")

# and merge that with df.alltags.path
df.alltags.path <- left_join(df.alltags.path, ambigTags.2,
  by = "motusTagID") %>% arrange(ts.h)

p <- ggplot(data = df.alltags.path, aes(ts.h, recvLat,
  group = Ambiguous, colour = Ambiguous))
p + geom_point() + geom_path() + theme_bw() + facet_wrap(~newID,
  scales = "free", ncol = 2) + theme(axis.text.x = element_text(angle = 45,
  vjust = 1, hjust = 1))

```



Let's deal with the easy ones first.

**ambigID -337: motusTagIDs 10811 and 16011**

```

filter(df.alltags.sub, ambigID == -337) %>% group_by(motusTagID,
  tagDeployStart, tagDeployEnd, tagDeployLat, tagDeployLon) %>%
  tally()

## # A tibble: 2 x 6
## # Groups:   motusTagID, tagDeployStart, tagDeployEnd, tagDeployLat [?]
##   motusTagID tagDeployStart      tagDeployEnd      tagDeployLat
##   <int>     <dttm>           <dttm>           <dbl>
## 1       10811 2014-10-28 07:00:00 2015-08-03 07:00:00    39.1
## 2       16011 2015-08-02 11:39:59 2015-12-17 11:39:59    51.5
## # ... with 2 more variables: tagDeployLon <dbl>, n <int>

```

We can see from the plot that ambiguous tag -337 is ambiguous only at the beginning of the deployment.

We can see from the summary of the tag deployment data that there were only 4 detections, at the exact latitude of deployment of tag 16011, and just before the non-ambiguous detections of motusTagID 16011. So the issue here is simply that the tail end of the deployment of tag 10811 slightly overlaps with the deployment of tag 16011. We can confidently claim these detections as belonging to motusTagID 16011, and remove the ambiguous detections assigned to the other tag.

We'll create another data frame to keep track of these runs.

```

# we want the detections associated with the
# motusTagID that we want to ultimately REMOVE from
# the data frame
df.block.2 <- filter(df.alltags.sub, ambigID == -337,
  motusTagID == 10811) %>% select(motusTagID, runID) %>%
  distinct()

```

#### ambigID -134: motusTagIDs 22905 and 23319

```

filter(df.alltags.sub, ambigID == -134) %>% group_by(motusTagID,
  tagDeployStart, tagDeployEnd, tagDeployLat, tagDeployLon,
  month(ts)) %>% tally()

```

```

## # A tibble: 4 x 7
## # Groups:   motusTagID, tagDeployStart, tagDeployEnd, tagDeployLat,
## #   tagDeployLon [?]
##   motusTagID tagDeployStart      tagDeployEnd      tagDeployLat
##   <int>     <dttm>           <dttm>           <dbl>
## 1       22905 2016-10-01 16:00:00 2017-06-12 16:00:00    50.2
## 2       22905 2016-10-01 16:00:00 2017-06-12 16:00:00    50.2
## 3       23319 2016-10-15 16:00:00 2017-06-26 16:00:00    50.2
## 4       23319 2016-10-15 16:00:00 2017-06-26 16:00:00    50.2
## # ... with 3 more variables: tagDeployLon <dbl>, `month(ts)` <dbl>,
## #   n <int>

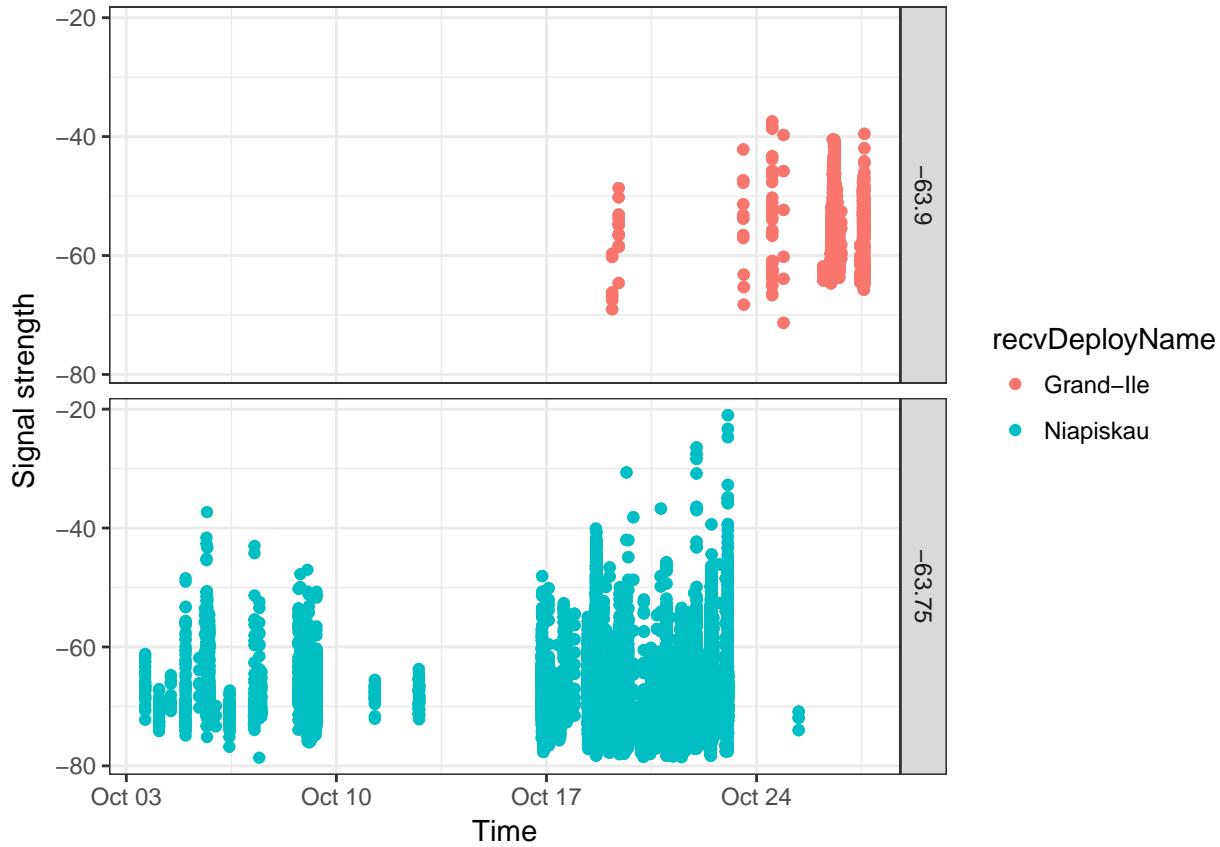
```

Here we have a similar situation, but one that is a bit more complex. Two identical tags were deployed at the same location, shortly after one another. Let's examine a simple plot.

```

filter(df.alltags.sub, motusTagID %in% c(22905, 23319),
  month(ts) == 10) %>% ggplot(aes(ts, sig, group = recvDeployName,
  colour = recvDeployName)) + geom_point() + theme_bw() +
  xlab("Time") + ylab("Signal strength") + facet_grid(recvLon ~
  .)

```



It appears that these are overlapping detections, at two sites in proximity to one another. Additional information from the field researchers may enable us to disentangle them, but it is not clear from the data.

We also examine the non-ambiguous detections of tag -134 that occur in mid-April. These are very early for a Red Knot to be flying through southern Ontario, so should be questioned.

We see from the following filter that there are two separate runs of length 3 each, separated by 3 days. If we inspect the rest of this batch (that is, if we also look at run lengths of 2 from the original data frame) ...

```
filter(df.alltags, batchID == 79646) %>% select(runLen,
  recvDeployName) %>% group_by(runLen, recvDeployName) %>%
  tally()
```

```
## # A tibble: 2 x 3
## # Groups:   runLen [?]
##   runLen recvDeployName     n
##   <int> <chr>           <int>
## 1      2 Koffler          110
## 2      3 Koffler          12
```

... we can see that there are many false positives at this tower around the same time (within the same batch) and so the run lengths of 3 are likely false positives. We will therefore remove all detections of this ambiguous tag from the database.

```
# we want the detections associated with the
# motusTagID that we want to ultimately REMOVE from
# the data frame

df.block.3 <- filter(df.alltags.sub, ambigID == -134) %>%
```

```
select(motusTagID, runID) %>% distinct()
```

**ambigID -171: motusTagIDs 22778, 22902 and 22403**

The ambiguous detections for this tag, which occur in the Great Lakes region, could also belong to motusTagID 22778 from the RBrownAMWO project or motusTagID 24303 from the Neonics project. Let's take a closer look at these detections.

First, find the deployment dates and locations for each tag.

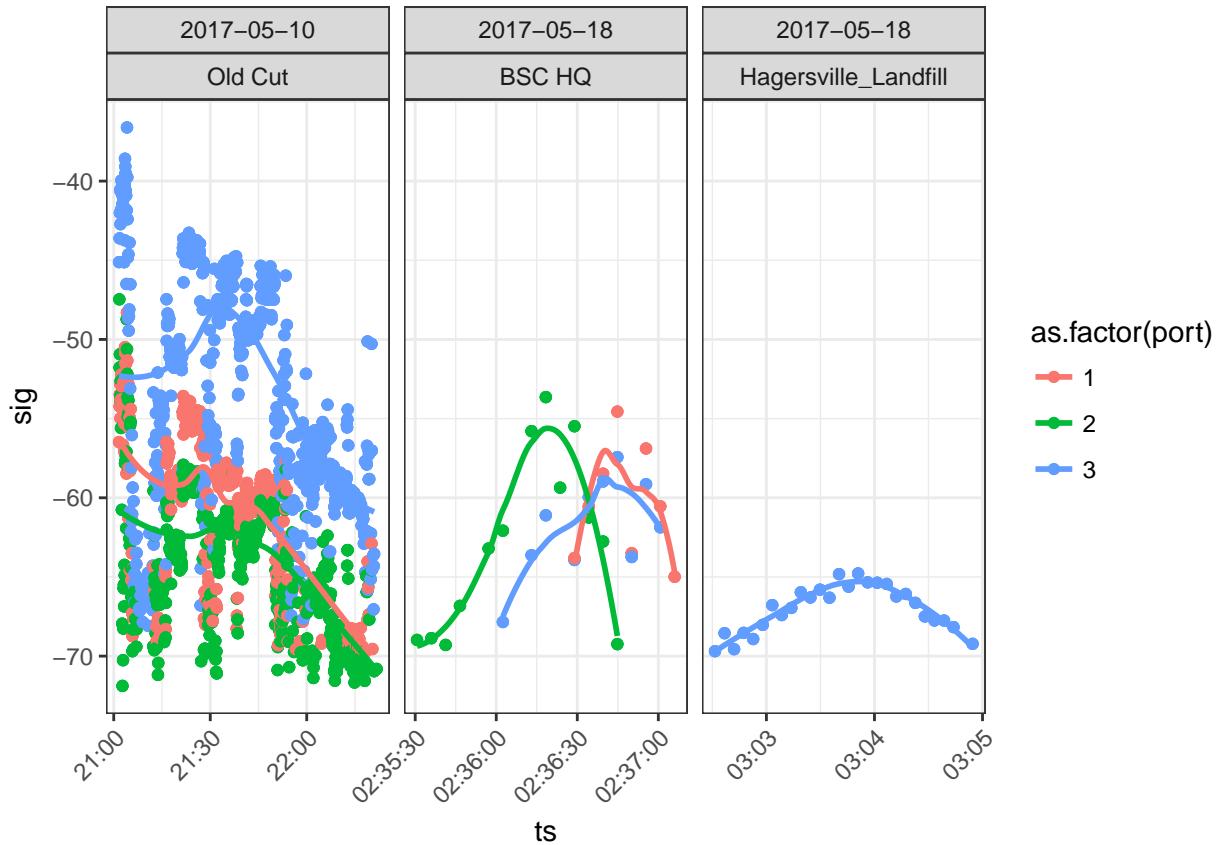
```
filter(df.alltags, ambigID == -171) %>% filter(!is.na(tagDeployStart)) %>%
  select(motusTagID, tagProjID, start = tagDeployStart,
         end = tagDeployEnd, lat = tagDeployLat, lon = tagDeployLon,
         species = speciesEN) %>% distinct() %>% arrange(start)
```

	motusTagID	tagProjID	start	end	lat
## 1	22902	176	2016-10-01 16:00:00	2017-06-12 16:00:00	50.19278
## 2	22778	82	2016-10-21 00:00:00	2018-09-09 00:00:00	45.13535
## 3	24303	146	2017-05-10 22:30:59	2017-06-30 22:30:59	42.60600
	lon	species			
## 1	-63.74528	Red Knot			
## 2	-67.29323	American Woodcock			
## 3	-80.46900	White-crowned Sparrow			

And plot the ambiguous detections.

```
df.ambig.171 <- filter(df.alltags.sub, ambigID == -171)

p <- ggplot(data = df.ambig.171, aes(ts, sig, colour = as.factor(port)))
p + geom_point() + geom_smooth(method = "loess", se = FALSE) +
  theme_bw() + facet_wrap(as_date(ts) ~ recvDeployName,
  scales = "free_x") + theme(axis.text.x = element_text(angle = 45,
  vjust = 1, hjust = 1))
```



We see that there are a large number of ambiguous detections on 10 May 2017 at Old Cut (Long Point, Lake Erie, Ontario), consistent with a bird ‘hanging around’. These are almost certainly detections of motusTagID ‘24303’ which was deployed at Old Cut on 10 May 2017. Subsequent detections on the 18th of May are near Old Cut (Bird Studies Canada HQ, Port Rowan, Ontario), and then a location to the North of Old Cut (Hagersville, Ontario). These detections are consistent with a bird departing on migration. Note in particular the pattern in the latter two panels of increasing then decreasing signal strength which indicates a bird is flying through the beam of an antenna.

These detections belong to another project, so we simply remove all detections of that ambiguous tag from our database.

```
# we want the detections associated with the
# motusTagID that we want to ultimately REMOVE from
# the data frame

df.block.4 <- filter(df.alltags.sub, ambigID == -171) %>%
  select(motusTagID, runID) %>% distinct()
```

#### ambigID -114: motusTagIDs 22897 and 24298

Next we look at the ambiguities for ambiguous tag -114.

```
filter(df.alltags, ambigID == -114) %>% filter(!is.na(tagDeployStart)) %>%
  select(motusTagID, tagProjID, start = tagDeployStart,
         end = tagDeployEnd, lat = tagDeployLat, lon = tagDeployLon,
         species = speciesEN) %>% distinct() %>% arrange(start)
```

	motusTagID	tagProjID	start	end	lat
## 1	22897	176	2016-10-01 16:00:00	2017-06-12 16:00:00	50.19278

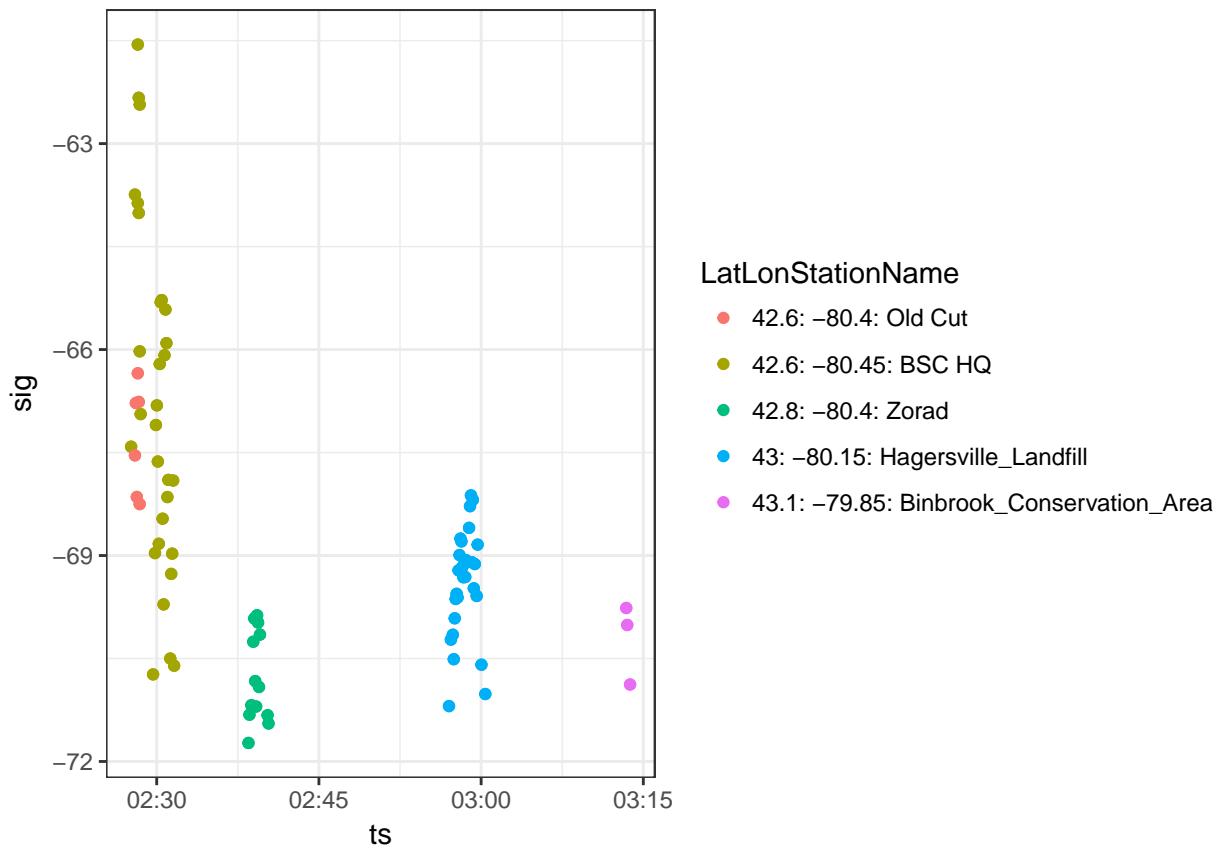
```
## 2      24298      146 2017-05-10 03:00:00 2017-06-30 03:00:00 42.60690
##   lon          species
## 1 -63.74528      Red Knot
## 2 -80.46900 White-crowned Sparrow
```

We again subset these and plot them. An initial plot suggested that all of the detections are of a migratory flight, so we construct a somewhat different plot from the one above, that emphasizes this behaviour better.

```
df.ambig.114 <- filter(df.alltags.sub, ambigID == -114) %>%
  mutate(LatLonStationName = paste(recvLat, recvLon,
    recvDeployName, sep = " : "))

p <- ggplot(data = df.ambig.114, aes(ts, sig, colour = LatLonStationName))

p + geom_point() + theme_bw()
```



Notice that the detections are consistent with a migratory departure from the Long Point area (Old Cut Field Station, Lake Erie, Ontario) about a week after the ambiguous tag 24298 was deployed at the same location. This again suggests that these ambiguous detections can be removed from our data because they belong to another project.

```
df.block.5 <- filter(df.alltags.sub, ambigID == -114) %>%
  select(motusTagID, runID) %>% distinct()
```

#### ambigID -106: motusTagIDs 17021 and 17357

These two tags pose an interesting problem. There is only a short period of overlap, between mid August 2015 and mid September. One individual is a Grey-cheeked Thrush, tagged in Colombia, the other a White-rumped Sandpiper, associated with the sample project.

```

filter(df.alltags, ambigID == -106) %>% filter(!is.na(tagDeployStart)) %>%
  select(motusTagID, tagProjID, start = tagDeployStart,
         end = tagDeployEnd, lat = tagDeployLat, lon = tagDeployLon,
         species = speciesEN) %>% distinct() %>% arrange(start)

##   motusTagID tagProjID           start           end      lat
## 1      17021        57 2015-04-30 05:00:00 2015-09-14 05:00:00 11.12265
## 2      17357       176 2015-08-11 07:20:00 2015-12-26 07:20:00 51.48390
##   lon      species
## 1 -74.08735 Gray-cheeked Thrush
## 2 -80.45000 White-rumped Sandpiper

```

We plot the ambiguous detections to examine the period of overlap.

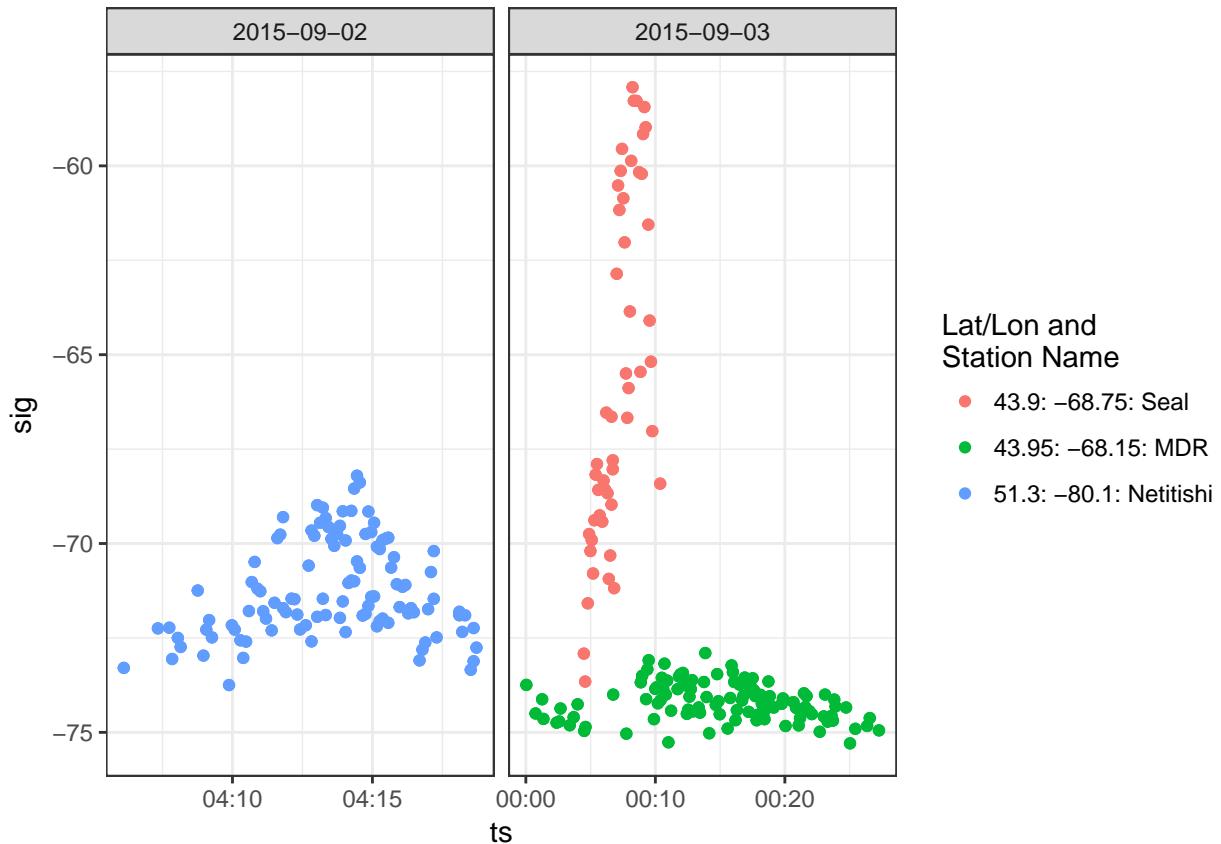
```

df.ambig.106 <- filter(df.alltags.sub, ambigID == -106)

p <- ggplot(data = df.ambig.106, aes(ts, sig, colour = paste(recvLat,
  recvLon, recvDeployName, sep = " : ")))

p + geom_point() + scale_colour_discrete(name = "Lat/Lon and\nStation Name") +
  theme_bw() + facet_wrap(~as_date(ts), scales = "free_x")

```



Both sets of detections are long run lengths, and look valid (increasing then decreasing signal strength). They are about a day apart, and so it is possible they represent two different birds, or the departure flight of the white-rumped sandpiper from its staging ground. Let's use the siteTrans function (in the motus package, see section C.13) to examine the flight from Netitishi to MDR/Seal (in the Gulf of Maine)

```
df.ambig.106 %>% filter(motusTagID == 17021) %>% # just pick one of the two ambiguous IDs
siteTrans(latCoord = "recvLat", lonCoord = "recvLon") %>%
  ungroup() %>%
  filter(rate < 60) %>% # remove the simultaneous detections from Seal and MDR
  mutate(total.time = as.numeric(round(seconds_to_period(tot_ts)))) %>%
  select(start=recvDeployName.x, end=recvDeployName.y, date=ts.x, "rate(m/s)" = rate,
         dist, total.time = total.time, bearing)
```

```
## # A tibble: 1 x 7
##   start     end   date           `rate(m/s)`    dist total.time bearing
##   <chr>    <chr> <dttm>          <dbl>    <dbl>      <dbl>    <dbl>
## 1 Netiti~ MDR_4~ 2015-09-02 04:18:42       17.1  1.21e6     70879     128
```

These detections are >1200 km distant from one another, but the flight speed (17 m/s) is consistent with a white-rumped Sandpiper. Given that the Gray-cheeked Thrush tag was near the end of its expected lifetime, we can reasonably claim these detections for our project, and remove the ambiguous detections associated with motusTagID 17021.

```
df.block.6 <- filter(df.alltags.sub, ambigID == -106,
                      motusTagID == 17021) %>% select(motusTagID, runID) %>%
  distinct()
```

### ambigID -56: motusTagIDs 22867 and 23316

These two tags were also both deployed by the same project.

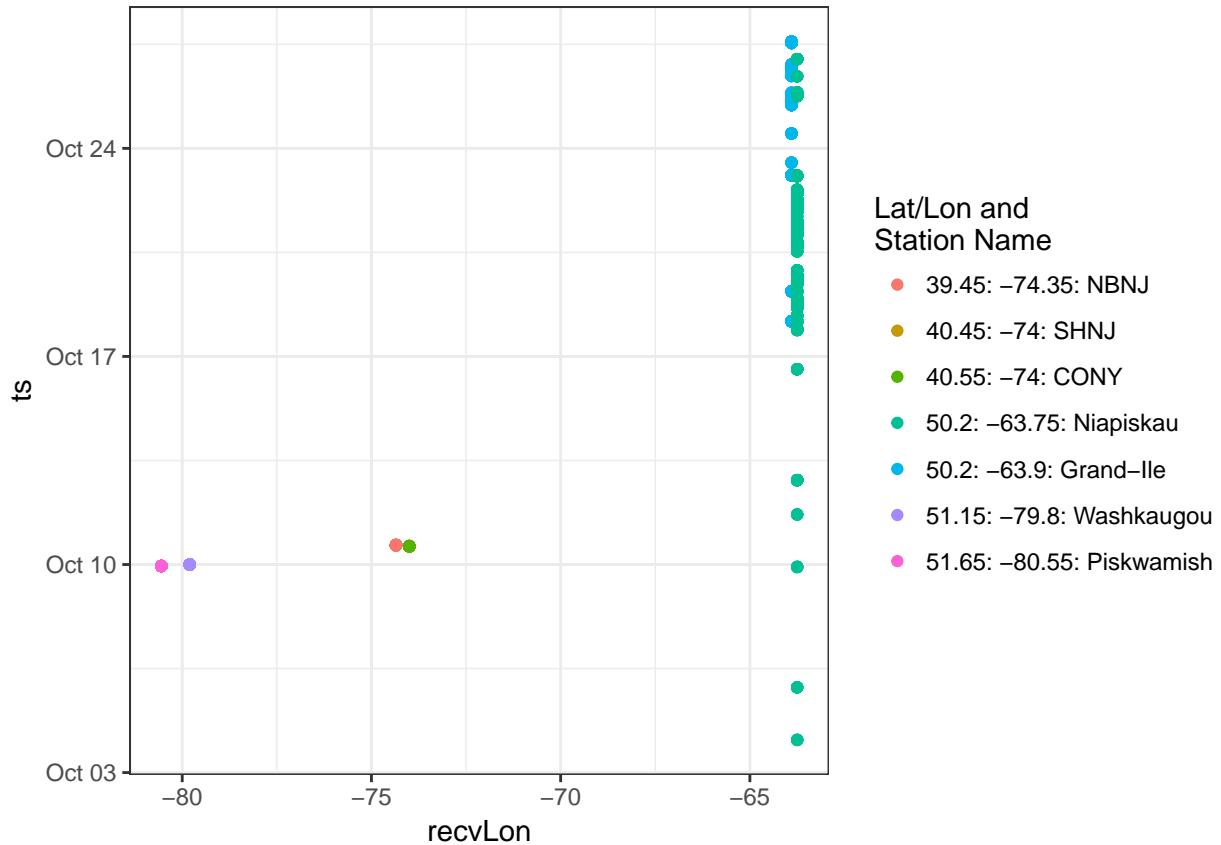
```
filter(df.alltags, ambigID == -56) %>% filter(!is.na(tagDeployStart)) %>%
  select(motusTagID, tagProjID, start = tagDeployStart,
         end = tagDeployEnd, lat = tagDeployLat, lon = tagDeployLon,
         species = speciesEN) %>% distinct() %>% arrange(start)
```

```
##   motusTagID tagProjID           start           end      lat
## 1      22867        176 2016-09-06 15:35:00 2017-05-18 15:35:00 51.79861
## 2      23316        176 2016-10-02 16:00:00 2017-06-13 16:00:00 50.19278
##   lon      species
## 1 -80.69139 Pectoral Sandpiper
## 2 -63.74528      Red Knot
```

Tag 23316 was deployed by the James Bay Shorebird Project (sample project) about three weeks after tag 22867, which was deployed from a location far to the west.

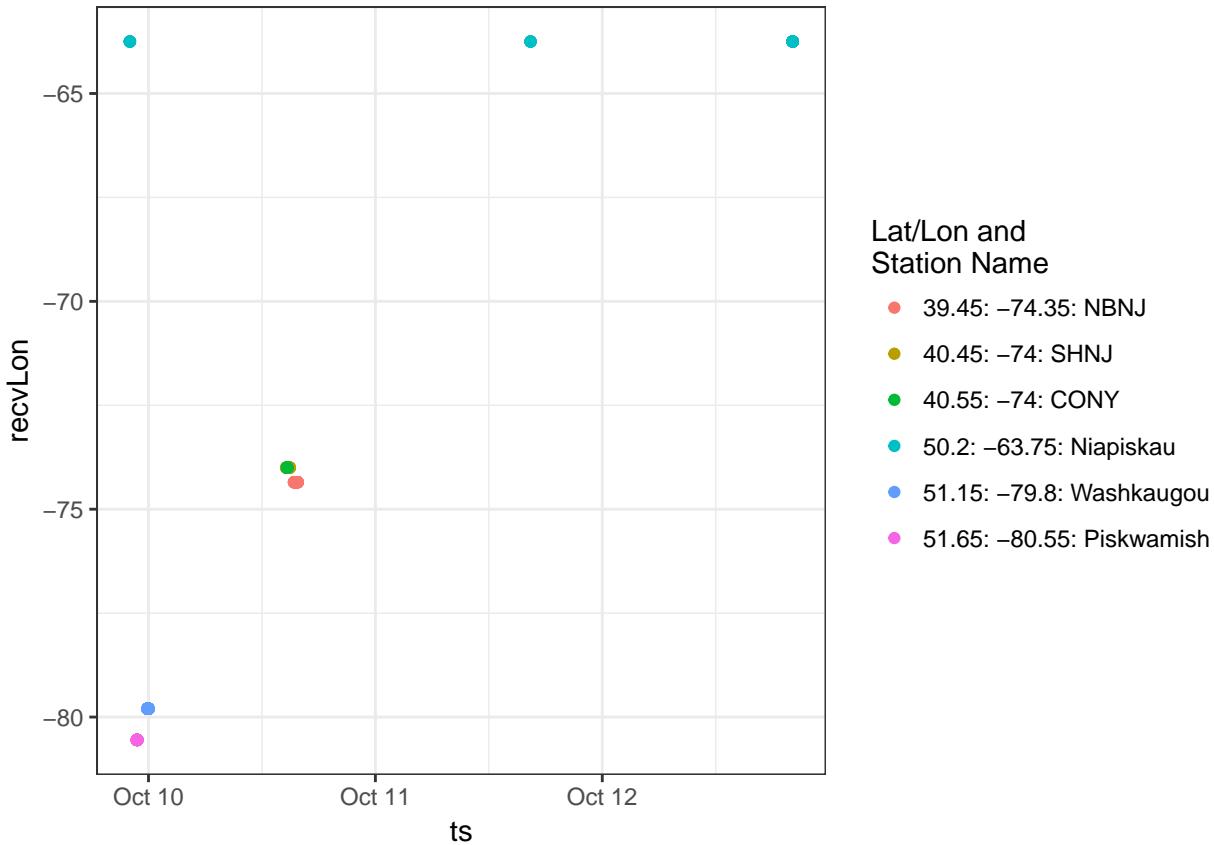
```
df.ambig.56 <- filter(df.alltags.sub, ambigID == -56) %>%
  mutate(sig = ifelse(sig > 0, sig * -1, sig))

p <- ggplot(data = df.ambig.56, aes(recvLon, ts, colour = paste(recvLat,
  recvLon, recvDeployName, sep = ": ")))
p + geom_point() + theme_bw() + scale_colour_discrete(name = "Lat/Lon and\nStation Name")
```



We can see from the plot that a tag is detected consistently near longitude -65, which is near the deployment location for motusTagID 23316 and after it's deployment start date, it was also present at -65 during and after detections far to the west. It's likely all the detections at -65 belong to motusTagID 23316, but it is also clear that anything informative about this ambiguity occurs between about 9-11 October, so let's zoom in on that part of the data set.

```
ts.begin <- ymd_hms("2016-10-06 00:00:00")
ts.end <- ymd_hms("2016-10-12 23:00:00")
p <- ggplot(data = filter(df.ambig.56, ts > ts.begin,
                           ts < ts.end), aes(ts, recvLon, colour = paste(recvLat,
                           recvLon, recvDeployName, sep = ":")))
p + geom_point() + theme_bw() + scale_colour_discrete(name = "Lat/Lon and\nStation Name")
```



We can see that the ambiguous tag was detected consistently at Niapiskau and Grand Ile before and after the period when it was also detected to the north and west (at Washkaugou and Piskwamish) and then to the south (NBNJ, SHNJ, and CONY). We can look at this transition by filtering out the portion of the data not near Niapiskau, and again using the siteTrans function from the motus package.

```
# other tag is a duplicate
df.56.tmp <- filter(df.ambig.56, !(recvLat == 50.2), motusTagID == 22867)

siteTrans(df.56.tmp, latCoord = "recvLat", lonCoord = "recvLon") %>%
  ungroup() %>%
  filter(rate < 60) %>% # get rid of simultaneous detections
  mutate(total.time = as.numeric(round(seconds_to_period(tot_ts)))) %>%
  select(start=recvDeployName.x,
         end=recvDeployName.y,
         date=ts.x, "rate(m/s)" = rate,
         dist, total.time = total.time, bearing)

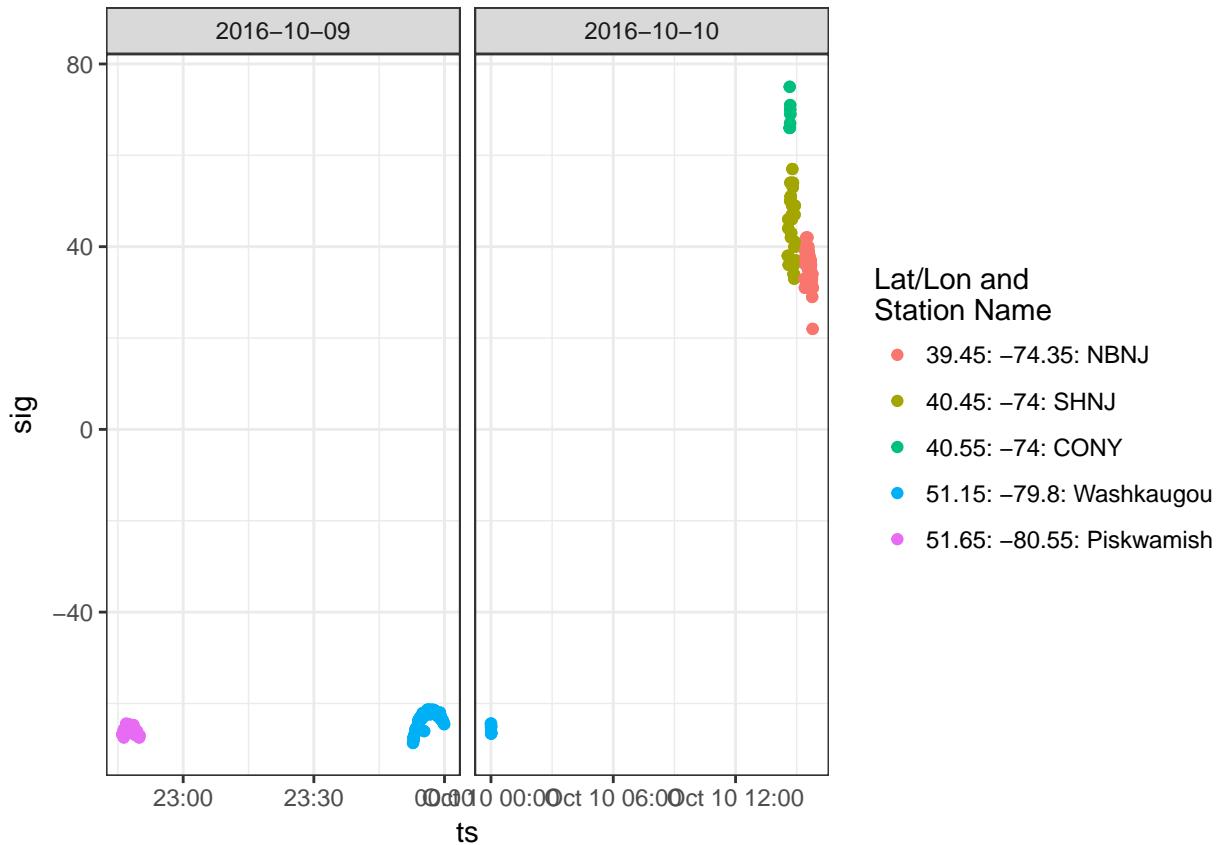
## # A tibble: 2 x 7
##   start   end     date      `rate(m/s)`    dist total.time bearing
##   <chr>   <chr>   <dttm>        <dbl>    <dbl>       <dbl>    <dbl>
## 1 Piskwa~ Washk~ 2016-10-09 22:49:59    20.3  7.63e4      3767     137
## 2 Washka~ SHNJ_~ 2016-10-10 00:00:42    24.3  1.27e6      52386     157
```

The bird made a 14.5 hour flight between Washkaugou and SHNJ at a rate of 24 m/s, which is plausible. The researchers involved may have other data to support or refute the inference (e.g. an actual sighting of the Red Knot still in Niapiskau after this flight was recorded) but it seems likely that while one tag remained at sites around longitude -65, another tag made the above migratory flights. We can make another more

detailed plot of signal strength to examine these potential migratory flights more closely:

```
df.56.tmp <- filter(df.alltags.sub, ambigID == -56,
  recvLon < -70)

p <- ggplot(data = df.56.tmp, aes(ts, sig, colour = paste(recvLat,
  recvLon, recvDeployName, sep = ": ")))
p + geom_point() + theme_bw() + scale_colour_discrete(name = "Lat/Lon and\nStation Name") +
  facet_wrap(~as_date(ts), scales = "free_x")
```



These look like typical fly-by patterns of increasing and then decreasing signal strength. This, coupled with overall detection patterns and knowledge of the species, leads us to believe that the ambiguous detections can be reasonably divided between the two individuals; one detected consistently around longitude -65 (23316), and the other migrating SW during the same period (22867).

To address this problem, we need to create two filters - one that excludes ambiguous detections of tag 22867, and one that excludes some detections of 23316. In this instance, we can do this most easily by filtering on motusTagID and recvDeployName.

```
# tag 23316 was only ever at 'Grand-Ile',
# 'Niapiskau', and tag 22867 was never detected at
# those sites. So we exclude all detections not at
# 'Grand-Ile', 'Niapiskau' for motusTag 23316, and
# do the opposite for tag 22867.

df.block.7 <- filter(df.alltags.sub, ambigID == -56,
  motusTagID == 23316, !(recvDeployName %in% c("Grand-Ile",
  "Niapiskau")) ) %>% select(motusTagID, runID) %>%
```

```

distinct()

df.block.8 <- filter(df.alltags.sub, ambigID == -56,
  motusTagID == 22867, recvDeployName %in% c("Grand-Ile",
  "Niapiskau")) %>% select(motusTagID, runID) %>%
distinct()

```

## 5.5 Checking validity of run lengths of 2

At the beginning of this chapter, we dropped all detections with a run length of 2 because they are considered to have a high probability of being false positive. Now that we've cleaned the data, and are confident in the detections that remain, you might at this point decide to go back and take a closer look at those detections with a run length of 2. You could do this, for example, by re-running the various plots described in this chapter (begin with lat/lon by time plots), to see if any of those detections make sense in the context of where the true detections lie. It is up to the user to decide which detections are reasonable in terms of the biology and behaviour of each tagged individual.

## 5.6 Filtering the data

### 5.6.1 Filter and save to RDS

To filter the data, we can simply join the df.block data frames back to the original data using a left\_join(), and filter those from the data:

```

# combine our df.block data frames into a single
# dataframe, and add probability = 0 for filtered
# records.
df.block.all <- bind_rows(df.block.0, df.block.1, df.block.2,
  df.block.3, df.block.4, df.block.5, df.block.6,
  df.block.7, df.block.8) %>% mutate(probability = 0)

df.alltags.sub <- left_join(df.alltags, df.block.all,
  by = c("runID", "motusTagID")) %>% # assign a probability of 1 to the records that
# will not be filtered
mutate(probability = ifelse(is.na(probability), 1,
  probability)) %>% filter(probability > 0)

```

Now save the local data frame as an RDS file, for use in the next chapter. Recall from Section 3.6 that the RDS format preserves the R data structure, including time stamps. The other benefit of saving to RDS is that you have the output from a given workflow saved as a flat file, which you can access again with a simple readRDS statement.

```
saveRDS(df.alltags.sub, file = "./data/dfAlltagsSub.rds")
```

And to read the data in again:

```
df.alltags.sub <- readRDS("./data/dfAlltagsSub.rds")
```

### 5.6.2 Save a custom filter in the motus database, and apply it to the data

As an alternative to saving your data as an RDS file, the Motus R package offers functionalities to save your filters directly within your .motus file. Once they are saved in your database, you can do the type of `left_join()` as above without having to rely on dataframes or an RDS file to store your data. To learn more about the functions available to work with Motus filters, refer to Appendix D for more details.

```
# combine our df.block data frames into a single
# dataframe, and add probability = 0 for filtered
# records.
df.block.all <- bind_rows(df.block.0, df.block.1, df.block.2,
  df.block.3, df.block.4, df.block.5, df.block.6,
  df.block.7, df.block.8) %>% mutate(probability = 0)

# create a new filter with name filtAmbigFalsePos
# and populate it with df.block.all
tbl.filter = writeRunsFilter(sql.motus, "filtAmbigFalsePos",
  df = df.block.all, delete = TRUE)

# obtain a table object where the filtered records
# from tbl.filter.1 have been removed
tbl.alltags.sub <- left_join(tbl.alltags, tbl.filter,
  by = c("runID", "motusTagID")) %>% mutate(probability = ifelse(is.na(probability),
  1, probability)) %>% filter(probability > 0)
```

# Chapter 6

## Exploring data with the Motus R package

Once you have clarified any possible ambiguous tags, and removed false positives, you are ready to start analyzing your clean data set. This chapter will walk you through some simple procedures to start working with and visualizing the clean sample data set; you can modify these scripts to work with your own data. For a more in-depth R tutorial we strongly recommend working through R for Data Science by Garrett Grolemund and Hadley Wickham (<http://r4ds.had.co.nz/>).

### 6.1 Load required packages

Follow the instructions in Chapter 2 to install the following packages before loading, if you haven't already done so.

```
library(motus)
library(tidyverse)
library(ggmap)

Sys.setenv(TZ = "GMT")
```

### 6.2 Load data

If you followed along with the previous Chapter (Chapter 5) and are working with the cleaned 'df.alltags.sub' file, you can skip this step and move to section 6.3.

Otherwise, if you saved your data as an RDS file, you can load it using:

```
df.alltags.sub <- readRDS("./data/dfAlltagsSub.rds") # change dir to local directory
```

Or, if you've applied a custom filter to your .motus file, you can load the previously downloaded sample .motus data (see Chapter 3 and clean it now). Currently the main benefit of using the custom filter is that you apply the filter to the .motus file, which allows you more flexibility in applying dplyr functions to manage and filter the data (e.g., you can select different variables to include in the data than we included in the RDS file in Chapter 5). This approach also allows you to more readily integrate new data added to your database with the tagme function. Because we are selecting the same variables and filtering the same records, the following gives you the same dataset as the readRDS statement above:

```

# load the .motus file
proj.num = 176
sql.motus <- tagme(proj.num, update = TRUE, dir = "./data/")
tbl.alltags <- tbl(sql.motus, "alltags")

# obtain a table object of the filter
tbl.filter = getRunsFilters(sql.motus, "filtAmbigFalsePos")

# filter and convert the table into a dataframe, with a few modications
df.alltags.sub <- left_join(tbl.alltags, tbl.filter, by = c("runID", "motusTagID")) %>%
  mutate(probability = ifelse(is.na(probability), 1, probability),
         recvLat = if_else((is.na(gpsLat) | gpsLat == 0),
                           recvDeployLat,
                           gpsLat),
         recvLon = if_else((is.na(gpsLon) | gpsLon == 0),
                           recvDeployLon,
                           gpsLon),
         recvAlt = if_else(is.na(gpsAlt),
                           recvDeployAlt,
                           gpsAlt)) %>%
  filter(probability > 0) %>%
  select(-noise, -slop, -burstSlop, -done, -bootnum, -codeSet,
         -mfg, -nomFreq, -markerNumber, -markerType, -tagDeployComments,
         -fullID, -deviceID, -recvDeployLat, -recvDeployLon, -recvDeployAlt,
         -speciesGroup, -gpsLat, -gpsLon, -recvAlt, -recvSiteName) %>%
  collect() %>%
  as.data.frame() %>%
  mutate(ts = as_datetime(ts), # work with dates AFTER transforming to flat file
         tagDeployStart = as_datetime(tagDeployStart),
         tagDeployEnd = as_datetime(tagDeployEnd))

```

Note that if your project is very large, you may want to convert only a portion of it to the dataframe, to avoid memory issues. Details on filtering the `tbl` prior to collecting as a dataframe are available in section ??.

Here we do so by adding a filter to the above command, in this case, only creating a dataframe for motusTagID 16047, but you can decide how to best subset your data based on your need (e.g. by species or year):

```

# create a subset for a single tag, to keep the
# dataframe small
df.alltags.16047 <- df.alltags.sub %>% filter(motusTagID ==
  16047)

```

### 6.3 Summarizing your data

Here we will run through some basic commands, starting with the `summary()` function to view a selection of variables in a data frame:

```

sql.motus %>% tbl("alltags") %>% select(ts, motusTagID,
  runLen, speciesEN, tagDeployLat, tagDeployLon,
  recvDeployLat, recvDeployLon) %>% collect() %>%
  summary()

```

##	ts	motusTagID	runLen	speciesEN
----	----	------------	--------	-----------

```

## Min.    :1.438e+09   Min.    :10811   Min.    : 2.0   Length:108826
## 1st Qu.:1.476e+09   1st Qu.:22897   1st Qu.: 30.0   Class :character
## Median :1.477e+09   Median :22905   Median :122.0   Mode  :character
## Mean   :1.476e+09   Mean   :22660   Mean   :355.9
## 3rd Qu.:1.477e+09   3rd Qu.:23316   3rd Qu.:404.0
## Max.   :1.498e+09   Max.   :24303   Max.   :2474.0
##
##   tagDeployLat   tagDeployLon   recvDeployLat   recvDeployLon
## Min.    :11.12     Min.    :-80.69     Min.    :-42.50     Min.    :-143.68
## 1st Qu.:50.19     1st Qu.:-63.75     1st Qu.: 50.20     1st Qu.: -63.75
## Median :50.19     Median :-63.75     Median : 50.20     Median : -63.75
## Mean   :50.14     Mean   :-65.77     Mean   : 49.05     Mean   : -65.64
## 3rd Qu.:50.19     3rd Qu.:-63.75     3rd Qu.: 50.20     3rd Qu.: -63.75
## Max.   :51.80     Max.   :-63.75     Max.   : 62.89     Max.   : -60.02
## NA's   :2025     NA's   :2025     NA's   :173      NA's   :173
# same summary for the filtered sql data
df.alltags.sub %>% select(ts, motusTagID, runLen, speciesEN,
  tagDeployLat, tagDeployLon, recvLat, recvLon) %>%
  summary()

```

```

##       ts                  motusTagID      runLen
## Min.  :2015-08-03 06:37:11   Min.  :16011   Min.  :  3.0
## 1st Qu.:2016-10-06 10:30:26   1st Qu.:22897   1st Qu.: 25.0
## Median :2016-10-09 21:49:41   Median :22897   Median : 92.0
## Mean   :2016-09-04 20:32:07   Mean   :22247   Mean   :230.7
## 3rd Qu.:2016-10-19 10:41:54   3rd Qu.:22897   3rd Qu.:286.0
## Max.   :2017-04-20 22:33:19   Max.   :23316   Max.   :1371.0
##
##   speciesEN      tagDeployLat   tagDeployLon      recvLat
## Length:49076      Min.  :50.19     Min.  :-80.69     Min.  :-42.50
## Class :character  1st Qu.:50.19     1st Qu.:-63.75     1st Qu.: 50.20
## Mode  :character  Median :50.19     Median :-63.75     Median : 50.20
##                   Mean   :50.34     Mean   :-65.63     Mean   : 49.97
##                   3rd Qu.:50.19     3rd Qu.:-63.75     3rd Qu.: 50.20
##                   Max.   :51.80     Max.   :-63.75     Max.   : 51.82
##                   NA's   :167
##
##      recvLon
## Min.  :-80.69
## 1st Qu.:-63.75
## Median :-63.75
## Mean   :-65.32
## 3rd Qu.:-63.75
## Max.   :-62.99
## NA's   :167

```

The dplyr package allows you to easily summarize data by group, manipulate variables, or create new variables based on your data.

We can manipulate existing variables or create new ones with dplyr's mutate function, here we'll convert ts to a POSIXct format, then make a new variable for year and day of year (doy).

We'll also remove the set of points with missing receiver latitude and longitudes. These may be useful in some contexts (for example if the approximate location of the receiver is known) but can cause warnings or errors when plotting.

```

df.alltags.sub <- df.alltags.sub %>%
  mutate(ts = as_datetime(ts, tz = "UTC"), # convert ts to POSIXct format
         year = year(ts), # extract year from ts
         doy = yday(ts)) %>% # extract numeric day of year from ts
  filter(!is.na(recvLat))
head(df.alltags.sub)

##   hitID runID batchID                  ts sig sigsd freq freqsd
## 1 45107  8886      53 2015-10-26 11:19:49  52     0     4     0
## 2 45108  8886      53 2015-10-26 11:20:28  54     0     4     0
## 3 45109  8886      53 2015-10-26 11:21:17  55     0     4     0
## 4 45110  8886      53 2015-10-26 11:21:55  52     0     4     0
## 5 45111  8886      53 2015-10-26 11:22:44  49     0     4     0
## 6 199885 23305     64 2015-10-26 11:12:04  33     0     4     0
##   motusTagID ambigID port runLen tagProjID mfgID tagType tagModel
## 1       16047     NA    3      5      176    378      ID NTQB-3-2
## 2       16047     NA    3      5      176    378      ID NTQB-3-2
## 3       16047     NA    3      5      176    378      ID NTQB-3-2
## 4       16047     NA    3      5      176    378      ID NTQB-3-2
## 5       16047     NA    3      5      176    378      ID NTQB-3-2
## 6       16047     NA    1     11      176    378      ID NTQB-3-2
##   tagLifespan tagBI pulseLen tagDeployID speciesID      tagDeployStart
## 1          NA 9.6971     2.5      1839      4670 2015-09-10 18:00:00
## 2          NA 9.6971     2.5      1839      4670 2015-09-10 18:00:00
## 3          NA 9.6971     2.5      1839      4670 2015-09-10 18:00:00
## 4          NA 9.6971     2.5      1839      4670 2015-09-10 18:00:00
## 5          NA 9.6971     2.5      1839      4670 2015-09-10 18:00:00
## 6          NA 9.6971     2.5      1839      4670 2015-09-10 18:00:00
##   tagDeployEnd tagDeployLat tagDeployLon tagDeployAlt recvDeployID
## 1 2016-03-10 18:00:00      51.4839     -80.45        NA      2510
## 2 2016-03-10 18:00:00      51.4839     -80.45        NA      2510
## 3 2016-03-10 18:00:00      51.4839     -80.45        NA      2510
## 4 2016-03-10 18:00:00      51.4839     -80.45        NA      2510
## 5 2016-03-10 18:00:00      51.4839     -80.45        NA      2510
## 6 2016-03-10 18:00:00      51.4839     -80.45        NA      2512
##   recv recvDeployName isRecvMobile recvProjID antType antBearing
## 1 Lotek-159      Shelburne          0      74  yagi-9      127
## 2 Lotek-159      Shelburne          0      74  yagi-9      127
## 3 Lotek-159      Shelburne          0      74  yagi-9      127
## 4 Lotek-159      Shelburne          0      74  yagi-9      127
## 5 Lotek-159      Shelburne          0      74  yagi-9      127
## 6 Lotek-164  BennettMeadow          0      74  yagi-9      243
##   antHeight speciesEN      speciesFR      speciesSci tagProjName
## 1          NA Red Knot Bécasseau maubèche Calidris canutus SampleData
## 2          NA Red Knot Bécasseau maubèche Calidris canutus SampleData
## 3          NA Red Knot Bécasseau maubèche Calidris canutus SampleData
## 4          NA Red Knot Bécasseau maubèche Calidris canutus SampleData
## 5          NA Red Knot Bécasseau maubèche Calidris canutus SampleData
## 6          NA Red Knot Bécasseau maubèche Calidris canutus SampleData
##   recvProjName gpsAlt filterID probability recvLat   recvLon year doy
## 1 <NA>        NA      NA           1 42.60699 -72.71657 2015 299
## 2 <NA>        NA      NA           1 42.60699 -72.71657 2015 299
## 3 <NA>        NA      NA           1 42.60699 -72.71657 2015 299
## 4 <NA>        NA      NA           1 42.60699 -72.71657 2015 299

```

```
## 5      <NA>     NA     NA      1 42.60699 -72.71657 2015 299
## 6      <NA>     NA     NA      1 42.68067 -72.47392 2015 299
```

We can also summarize information by group, in this case motusTagID, and apply various functions to these groups such as getting the total number of detections (n) for each tag, the number of receivers each tag was detected on, the first and last detection date, and the total number of days there was at least one detection:

```
tagSummary <- df.alltags.sub %>% group_by(motusTagID) %>%
  summarize(nDet = n(), nRecv = length(unique(recvDeployName)),
            tsMin = min(ts), tsMax = max(ts), totDay = length(unique(doy)))

head(tagSummary)
```

```
## # A tibble: 6 x 6
##   motusTagID nDet nRecv tsMin           tsMax      totDay
##       <int> <int> <int> <dttm>       <dttm>      <int>
## 1     16011    122     1 2015-08-03 06:37:11 2015-08-05 20:41:12 3
## 2     16035    430     5 2015-08-14 17:53:49 2015-09-02 14:06:09 6
## 3     16036     62     1 2015-08-17 21:56:44 2015-08-17 21:58:52 1
## 4     16037   1296     3 2015-08-23 15:13:57 2015-09-08 18:37:16 14
## 5     16038     73     1 2015-08-20 18:42:33 2015-08-22 22:19:37 3
## 6     16039   1050    10 2015-08-23 02:28:45 2015-09-19 06:08:31 8
```

We can also group by multiple variables; applying the same function as above but now grouping by motusTagID and recvDeployName, we will get information for each tag detected on each receiver. Since we are grouping by recvDeployName, there will be by default only one recvDeployName in each group, thus the variable nRecv will be 1 for each row. This is not very informative, however we include this to help illustrate how grouping works:

```
tagRecvSummary <- df.alltags.sub %>% group_by(motusTagID,
  recvDeployName) %>% summarize(nDet = n(), nRecv = length(unique(recvDeployName)),
  tsMin = min(ts), tsMax = max(ts), totDay = length(unique(doy)))
```

```
head(tagRecvSummary)
```

```
## # A tibble: 6 x 7
## # Groups:   motusTagID [2]
##   motusTagID recvDeployName nDet nRecv tsMin           tsMax      totDay
##       <int> <chr>     <int> <int> <dttm>
## 1     16011 North Bluff    122     1 2015-08-03 06:37:11
## 2     16035 Brier2        41     1 2015-09-02 14:03:19
## 3     16035 D'Estimauville  32     1 2015-09-02 07:58:43
## 4     16035 Netitishi      286     1 2015-08-14 17:53:49
## 5     16035 Southwest Head   65     1 2015-09-02 13:06:13
## 6     16035 Swallowtail      6     1 2015-09-02 13:21:27
## # ... with 2 more variables: tsMax <dttm>, totDay <int>
```

## 6.4 Plotting your data

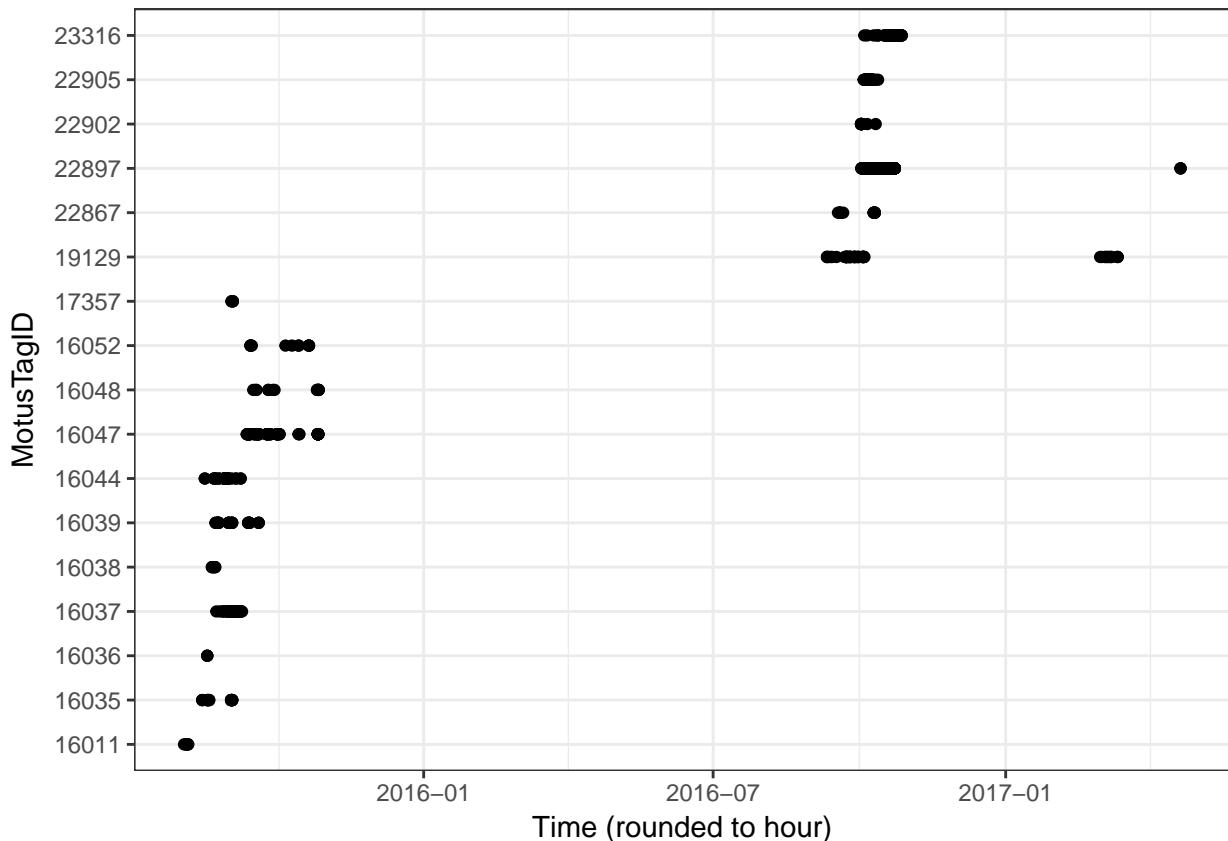
Plotting your data is a powerful way to visualize broad and fine-scale detection patterns. This section will give you a brief introduction to plotting using ggplot2. For more in depth information on the uses of ggplot2, we recommend the Cookbook for R, and the rstudio ggplot2 cheatsheet.

To make coarse-scale plots with large files, we suggest first rounding the detection time to the nearest hour or day so that processing time is faster. Here we round detection times to the nearest hour, then make a basic

plot of hourly detections by motusTagID:

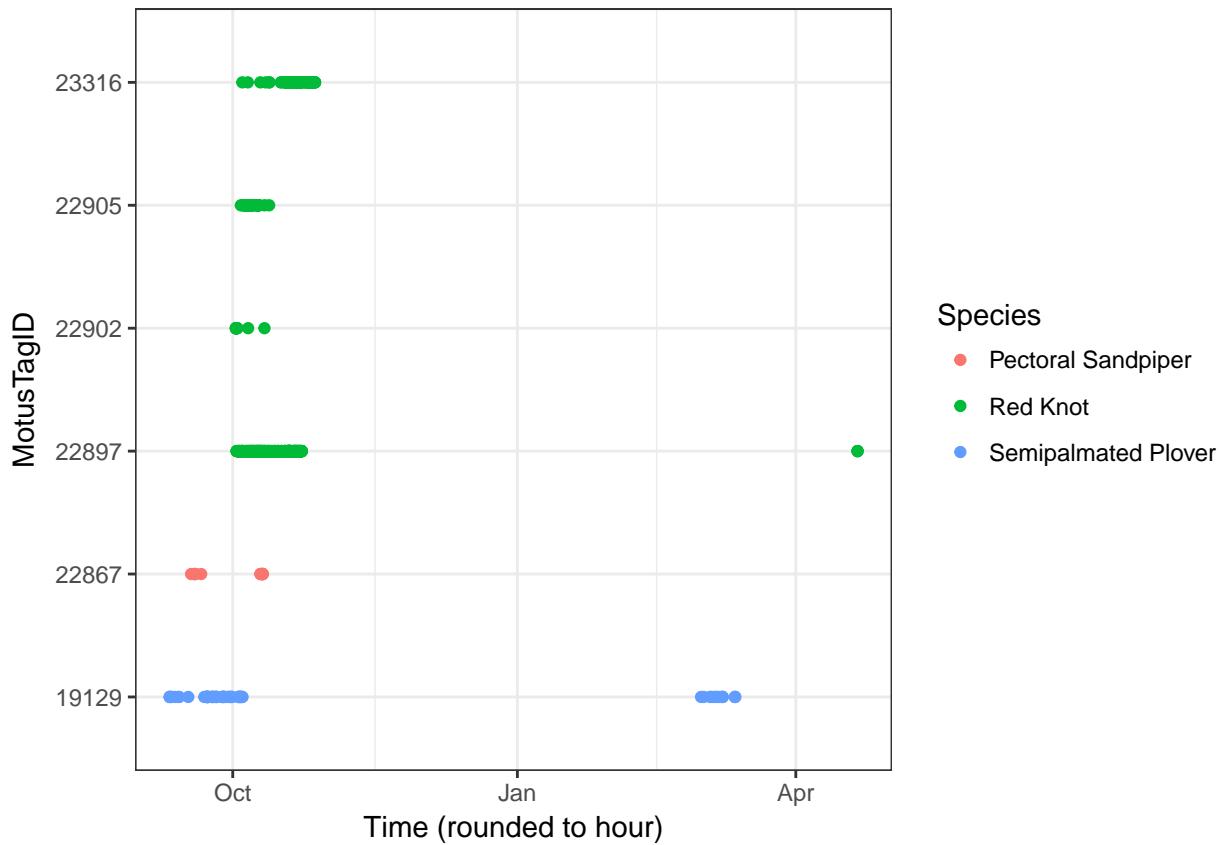
```
df.alltags.sub.2 <- mutate(df.alltags.sub, hour = as.POSIXct(round(ts,
  "hour")))) %>% select(motusTagID, port, tagDeployStart,
  tagDeployLat, tagDeployLon, recvLat, recvLon, recvDeployName,
  antBearing, speciesEN, year, doy, hour) %>% distinct()

p <- ggplot(data = df.alltags.sub.2, aes(hour, as.factor(motusTagID)))
p + geom_point() + ylab("MotusTagID") + xlab("Time (rounded to hour)") +
  theme_bw()
```



Let's focus only on tags deployed in 2016, and we can colour the tags by species:

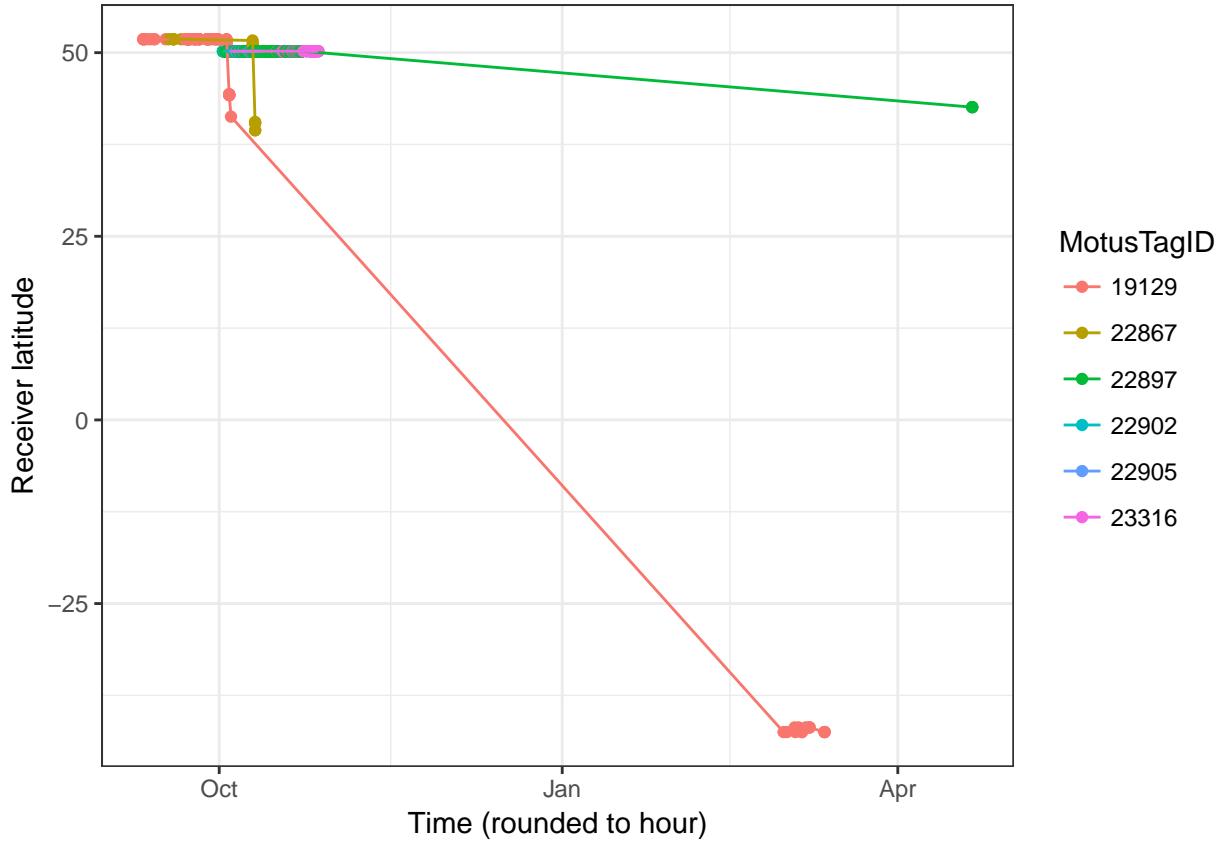
```
p <- ggplot(data = filter(df.alltags.sub.2, year(tagDeployStart) == 2016),
  aes(hour, as.factor(motusTagID), col = speciesEN))
p + geom_point() + ylab("MotusTagID") + xlab("Time (rounded to hour)") +
  scale_colour_discrete(name = "Species") + theme_bw()
```



We can see how tags moved latitudinally by first ordering by hour, and colouring by motusTagID:

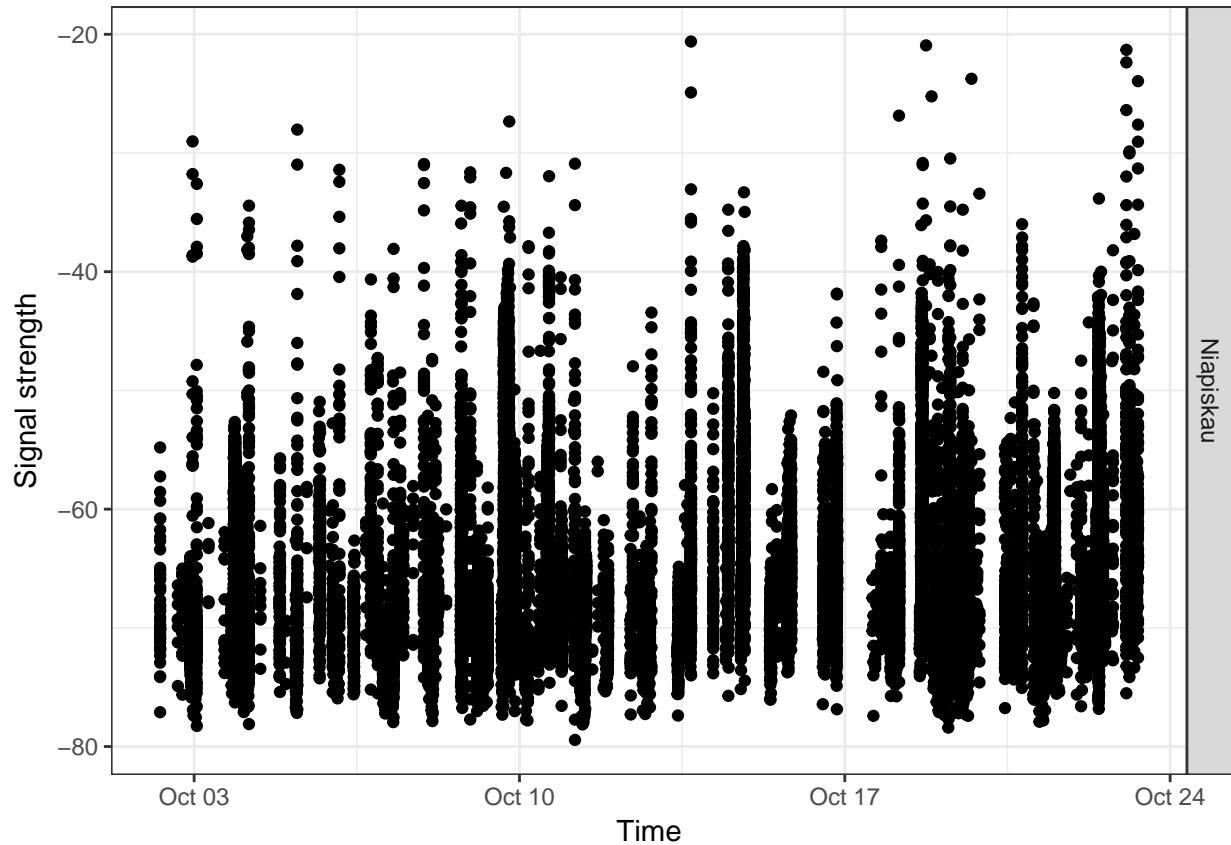
```
df.alltags.sub.2 <- arrange(df.alltags.sub.2, hour)

p <- ggplot(data = filter(df.alltags.sub.2, year(tagDeployStart) ==
  2016), aes(hour, recvLat, col = as.factor(motusTagID),
  group = as.factor(motusTagID)))
p + geom_point() + geom_path() + theme_bw() + xlab("Time (rounded to hour)") +
  ylab("Receiver latitude") + scale_colour_discrete(name = "MotusTagID")
```



Now lets look at more detailed plots of signal variation. We use the full df.alltags.sub dataframe so that we can get signal strength for each detection of a specific tag. Let's examine fall 2016 detections of tag 22897 at Niapiskau; we facet the plot by deployment name, ordered by decreasing latitude:

```
p <- ggplot(filter(df.alltags.sub, motusTagID == 22897,
  recvDeployName == "Niapiskau"), aes(ts, sig))
p + theme_bw() + geom_point() + xlab("Time") + ylab("Signal strength") +
  facet_grid(recvDeployName ~ .)
```

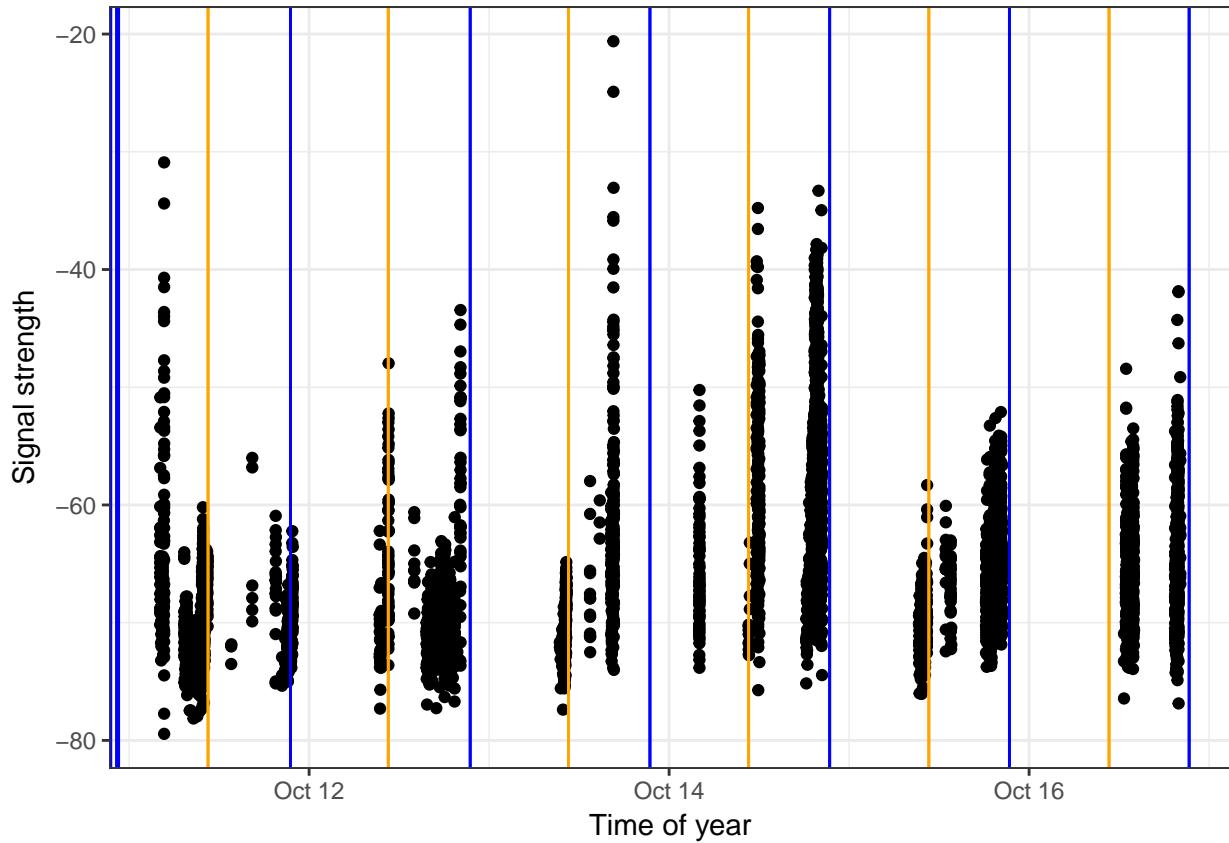


We use the sunRiseSet function available in the motus R package (see C.2) to get sunrise and sunset times for all detections. We then zoom in on a certain timeframe and add that information to the above plot by adding a geom\_vline() statement to the code, which adds a yellow line for sunrise time, and a blue line for sunset time:

```
# add sunrise and sunset times to the dataframe
df.alltags.sub <- sunRiseSet(df.alltags.sub, lat = "recvLat",
                             lon = "recvLon")

p <- ggplot(filter(df.alltags.sub, motusTagID == 22897 &
                     ts > ymd("2016-10-11") & ts < ymd("2016-10-17") &
                     recvDeployName == "Niapiskau"), aes(ts, sig))

p + theme_bw() + geom_point() + xlab("Time of year") +
  ylab("Signal strength") + geom_vline(xintercept = df.alltags.sub$sunrise,
                                       col = "orange") + geom_vline(xintercept = df.alltags.sub$setsun,
                                       col = "blue")
```



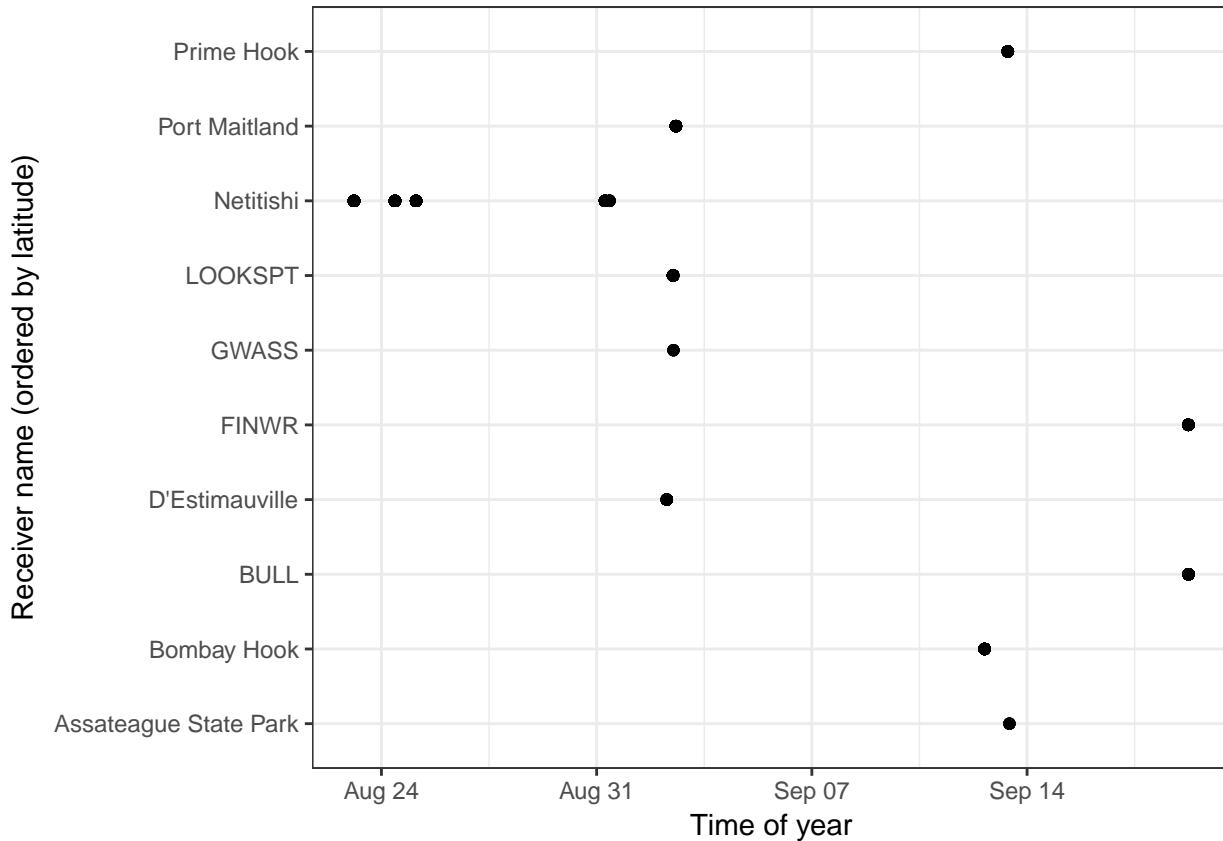
We can see that during this period, the tag was most often detected during the day, suggesting it may be actively foraging in this area during this time.

The same plots can provide valuable movement information when the receivers are ordered geographically. We do this for motusTagID 16039:

```
# We'll first order sitelat by latitude (for plots)
df.alltags.sub <- mutate(df.alltags.sub, recvDeployName = as.factor(as.character(reorder(recvDeployName
recvLat))))
```

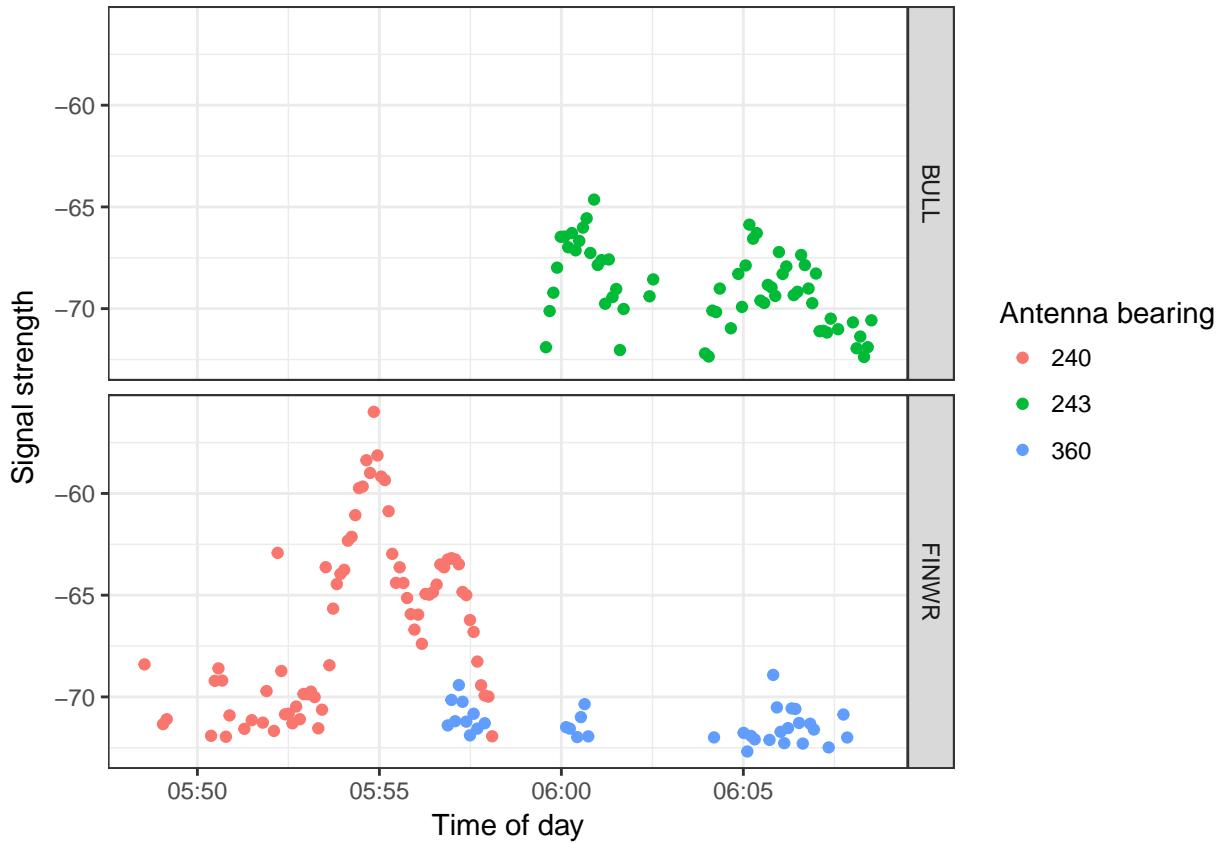
```
p <- ggplot(filter(df.alltags.sub, motusTagID == 16039 &
ts < ymd("2015-10-01")), aes(ts, recvDeployName))
```

```
p + theme_bw() + geom_point() + xlab("Time of year") +
ylab("Receiver name (ordered by latitude)")
```



We zoom in on a section of this plot and look at antenna bearings to see directional movement past stations:

```
p <- ggplot(filter(df.alltags.sub, motusTagID == 16039,
  ts > ymd("2015-09-14"), ts < ymd("2015-10-01")),
  aes(ts, sig, col = as.factor(antBearing)))
p + theme_bw() + geom_point() + xlab("Time of day") +
  ylab("Signal strength") + scale_color_discrete(name = "Antenna bearing") +
  facet_grid(recvDeployName ~ .)
```



This plot shows the typical flyby pattern of a migrating animal, with signal strength increasing and then decreasing as the tag moves through the beams of the antennas.

## 6.5 Mapping your data

To generate maps of tag paths, we will once again use summarized data so we can work with a much smaller database for faster processing. Here we'll summarize detections by day. As we did in Chapter 5, we create a simple function to summarize the data, since we will likely want to do this type of summary over and over again.

```
# simplify the data by summarizing by the runID.
# if you want to summarize at a finer (or coarser) scale, you can also create other groups.
# The simplest alternative is a rounded timestamp variable; for example by using
# mutate(ts.h = plyr::round_any(ts, 3600)) function call.
# Other options are to just use date (e.g date = as_date(ts))

#
fun.getpath <- function(df)
{
  df %>%
    filter(tagProjID == proj.num, # keep only tags registered to the sample project
          !is.na(recvLat) | !(recvLat == 0)) %>%
    group_by(motusTagID, runID, recvDeployName, ambigID,
             tagDeployLon, tagDeployLat, recvLat, recvLon) %>%
    summarize(max.runLen = max(runLen), ts.h = mean(ts)) %>%
```

```

arrange(motusTagID, ts.h) %>%
  data.frame()
} # end of function call

df.alltags.path <- fun.getpath(df.alltags.sub)

df.alltags.sub.path <- df.alltags.sub %>%
  filter(tagProjID == proj.num) %>% # keep to tags registered to the sample project
  arrange(motusTagID, ts) %>%       # order data by time stamp for each motus tag ID
  mutate(date = as_date(ts)) %>%    # create date variable
  group_by(motusTagID, date, recvDeployName, ambigID,
           tagDeployLon, tagDeployLat, recvLat, recvLon)

df.alltags.path <- fun.getpath(df.alltags.sub.path)

```

### 6.5.1 Mapping with Google Maps

Mapping with Google Maps can be a fast way to view flight paths and allows you to select from multiple base layers.

The first step is to create a map with a specified map centre, maptype (“terrain”, “roadmap”, “satellite”, or “hybrid”), and level of zoom (integer for zoom 3-21, 3 being continent level, 10 being city-scale). We then add points for receivers and lines connecting consecutive detections by motusTagID. We can also add points for all receivers that were active during a certain time period if we have already downloaded all metadata.

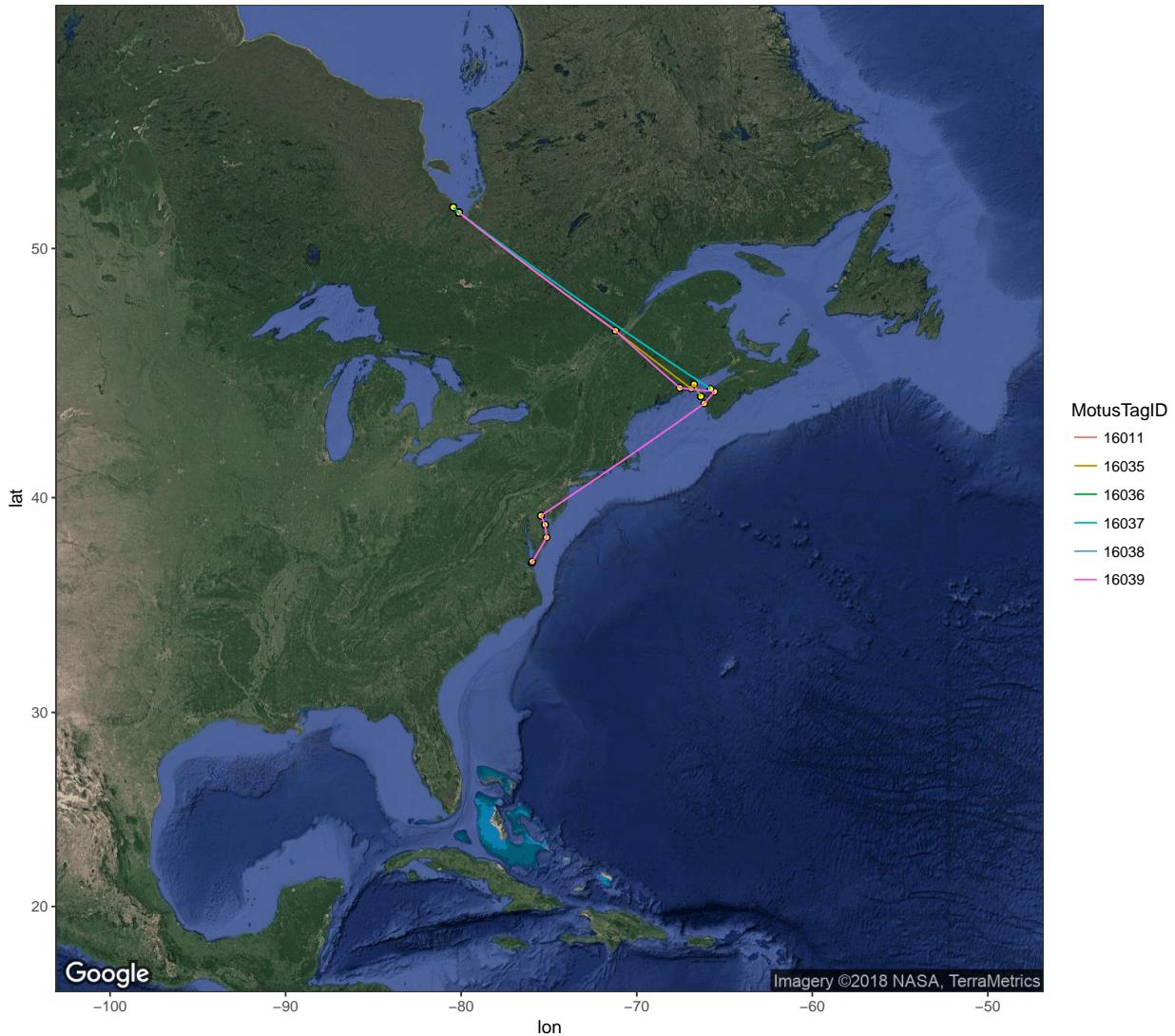
```

gmap <- get_map(location = c(lon = -75, lat = 40), # lon/lat to centre map over
                 maptype = "satellite", # select maptype
                 source = "google",
                 zoom = 4) # zoom, must be a whole number

# just use the tags that we have examined carefully and filtered (in the previous chapter)
df.tmp <- filter(df.alltags.path,
                  motusTagID %in% c(16011, 16035, 16036, 16037, 16038, 16039))
df.tmp <- arrange(df.tmp, ts.h) # arrange by hour
df.tmp <- as.data.frame(df.tmp)

p <- ggmap(gmap)
p + geom_point(data=df.tmp,
                aes(recvLon, recvLat), pch=21, colour = "black", fill = "yellow") +
  geom_path(data=df.tmp,
            aes(recvLon, recvLat, group=motusTagID, col = as.factor(motusTagID))) +
  theme_bw() +
  scale_color_discrete(name="MotusTagID")

```

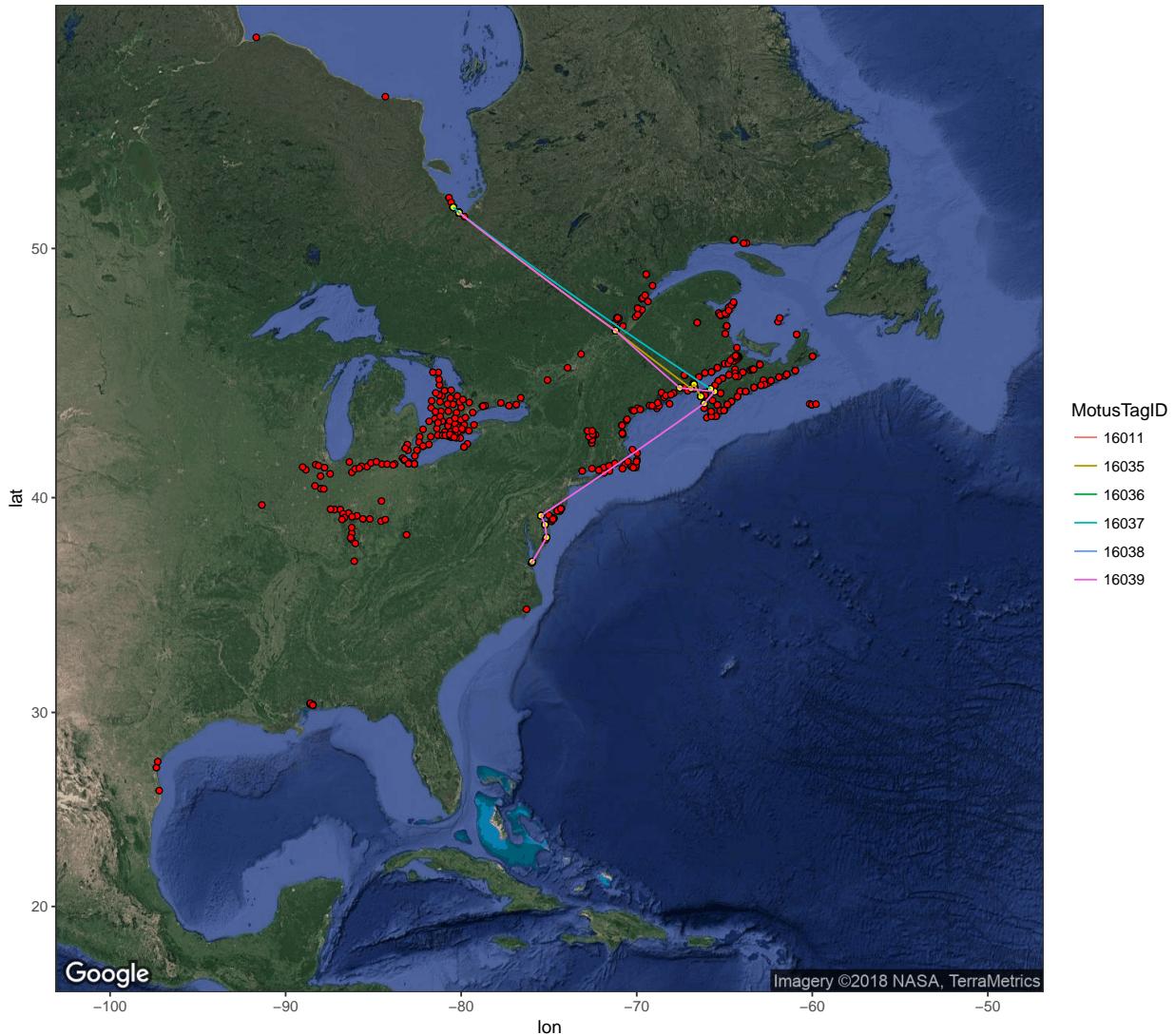


We make the same plot, with additional points for all receivers that were active during a specified time:

```
# get receiver metadata
tbl.recvDeps <- tbl(sql.motus, "recvDeps")
df.recvDeps <-tbl.recvDeps %>% collect %>% as.data.frame() %>%
  mutate(tsStart = as_datetime(tsStart, tz = "UTC",
    origin = "1970-01-01"), tsEnd = as_datetime(tsEnd,
    tz = "UTC", origin = "1970-01-01"))
# for deployments with no end dates, make an end
# date a year from now
df.recvDeps$tsEnd <- as.POSIXct(ifelse(is.na(df.recvDeps$tsEnd),
  as.POSIXct(format(Sys.time(), "%Y-%m-%d %H:%M:%S")) +
    lubridate::dyears(1), df.recvDeps$tsEnd), tz = "UTC",
  origin = "1970-01-01")
# get running intervals for all receiver
```

```
# deployments
site0p <- with(df.recvDeps, lubridate::interval(tsStart,
  tsEnd)) # get running intervals for each deployment
# set the date range you're interested in
dateRange <- lubridate::interval(as.POSIXct("2015-08-01"),
  as.POSIXct("2016-01-01"))
# create new variable 'active' which will be set to
# TRUE if the receiver was active at some point
# during your specified date range, and FALSE if
# not
df.recvDeps$active <- lubridate::int_overlaps(site0p,
  dateRange)

# create map with receivers active during specified
# date range as red, and receivers with detections
# as yellow
p <- ggmap(gmap)
p + geom_point(data = subset(df.recvDeps, active ==
  TRUE), ggplot2::aes(longitude, latitude), pch = 21,
  colour = "black", fill = "red") + geom_point(data = df.tmp,
  aes(recvLon, recvLat), pch = 21, colour = "black",
  fill = "yellow") + geom_path(data = df.tmp, aes(recvLon,
  recvLat, group = motusTagID, col = as.factor(motusTagID))) +
  theme_bw() + scale_color_discrete(name = "MotusTagID")
```



### 6.5.2 Creating simple outline maps

We load the base maps.

```
na.lakes <- map_data(map = "lakes")
na.lakes <- na.lakes %>% mutate(long = long - 360)

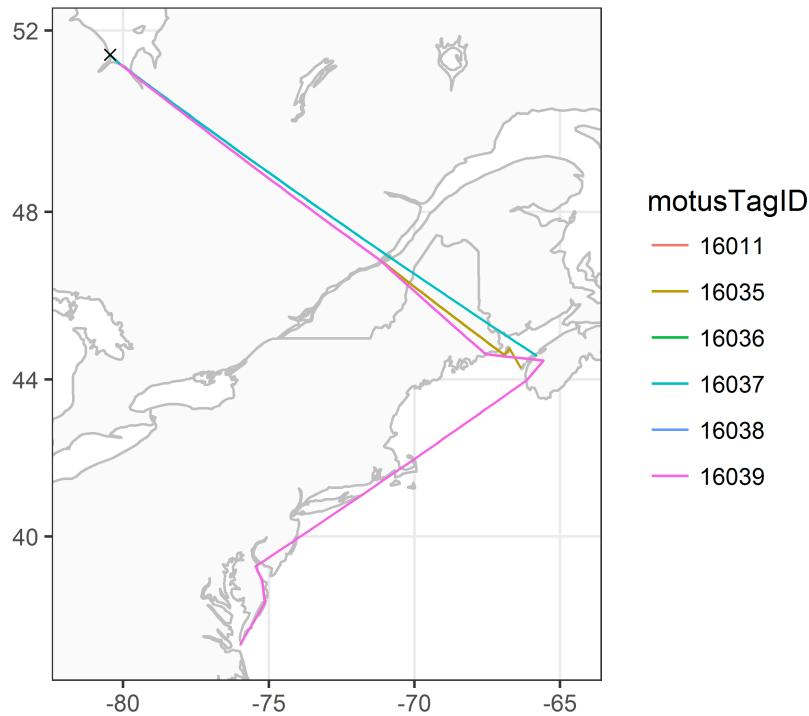
# Include all of the Americas to begin
na.map <- map_data(map = "world2")
na.map <- na.map %>% filter(region %in% c("Canada",
    "USA")) %>% mutate(long = long - 360)
```

Then, to map the paths, we set the x-axis and y-axis limits based on the location of receivers with detections.

Depending on your data, these might need to be modified to encompass the deployment location of the tags, if tags were not deployed near towers with detections. We then use ggplot to plot the map and tag paths. Here we use the Mercator projection and are colouring the paths by motusTagID, including a point for where the tag was deployed:

```
# set limits to map based on locations of
# detections, ensuring they include the deployment
# locations
xmin <- min(df.tmp$recvLon, na.rm = TRUE) - 2
xmax <- max(df.tmp$recvLon, na.rm = TRUE) + 2
ymin <- min(df.tmp$recvLat, na.rm = TRUE) - 1
ymax <- max(df.tmp$recvLat, na.rm = TRUE) + 1

# map
ggplot(na.lakes, aes(long, lat)) + geom_polygon(data = na.map,
  aes(long, lat, group = group), colour = "grey",
  fill = "grey98") + geom_polygon(aes(group = group),
  colour = "grey", fill = "white") + coord_map(projection = "mercator",
  xlim = c(xmin, xmax), ylim = c(ymin, ymax)) + xlab("") +
  ylab("") + theme_bw() + geom_path(data = df.tmp,
  aes(recvLon, recvLat, group = as.factor(motusTagID),
  colour = as.factor(motusTagID))) + geom_point(data = df.tmp,
  aes(tagDeployLon, tagDeployLat), colour = "black",
  shape = 4) + scale_colour_discrete("motusTagID")
```



The functions above provide examples of how you can begin exploring your data and are by no means exhaustive. The next chapter will cover some common errors and troubleshooting you may encounter while trying to download and use the .motus sql files.

# Appendix A

## Appendix - alltags structure

The following variables are included in each ‘alltags’ view in the SQLite file:

Field	Description
hitID	unique Motus ID for this tag detection
runID	unique Motus ID for the run this detection belongs to
batchID	unique Motus ID for the processing batch this detection came from
ts	timestamp, in seconds since 1 Jan, 1970 GMT; precision: 0.1 ms (SG); 2.5 ms (Lotek).
sig	signal strength in “native units”; for SG: dB (max) (logarithmic, 0 = max possible, -10 = 0.1 * max, etc.); for Lotek: raw value (0...255)
sigsd	std. dev. in signal strength among pulses in this burst. SG Only; NA for Lotek
noise	estimate of background radio noise when tag detected, in dB (max) for SG; NA for Lotek
freq	frequency offset from antenna listening frequency, in kHz for SG only; NA for Lotek
freqsd	std. dev. of freq offset among pulses in this burst, in kHz. Values larger than 0.1 kHz suggest a bogus detection. SG only; NA for Lotek.
slop	total absolute difference (milliseconds) in inter-pulse intervals for this burst between registration and detection. SG only; NA for Lotek
burstSlop	signed difference (seconds) between detection and registration burst intervals. This is always 0 for the first burst in a run (see posInRun)
done	logical: is run finished?
motusTagID	Motus tag ID - unique to each individual tag registered
ambigID	ambiguous ID assigned to ambiguous tags
port	the port number that the detection occurred on
runLen	number of tag bursts in the current run; constant for all records having the same runID
bootnum	boot session of receiver for SG; NA for Lotek
tagProjectID	ID of the project that manages this tag.
mfgID	manufacturer ID
tagType	for coded tags, the name of the codeset (e.g. ‘Lotek-3’)
codeSet	tag manufacturer
mfg	manufacturer’s model name for a tag (e.g. ‘NTQB-3-2’)
tagModel	estimated lifespan of tag (days)
tagLifespan	

Field	Description
nomFreq	nominal tag frequency (MOTUS: Nominal frequency receiver was tuned to, in Mhz. This really only applies to SG, where we usually tune 4 kHz below the nominal tag frequency. So in that case, antFreq = 166.376 while nomFreq = 166.380)
tagBI	burst interval of tag, in seconds (e.g., 5.8984)
pulseLen	tag pulse length (milliseconds), if applicable. This value is only assigned based on the sample recording of the tag.
tagDeployID	Motus tag deployment ID
speciesID	unique numeric Motus ID (integer) for the species on which the tag was deployed
markerNumber	number for any additional marker placed on organism (e.g., bird band #)
markerType	type of additional marker (e.g. metal band)
tagDeployStart	tag deployment start date
tagDeployEnd	tag deployment end date
tagDeployLat	latitude of tag deployment, in decimal degrees N - negative values for Southern hemisphere
tagDeployLon	longitude of tag deployment, in decimal degrees E - negative values for Western hemisphere
tagDeployAlt	altitude of tag deployment, in meters ASL
tagDeployComments	additional comments or unclassified metadata for tag (often in JSON format)
fullID	full tag ID as PROJECT#MFGID:BI@NOMFREQ. Not necessarily unique over time. See motusTagID for a unique tag
deviceID	Motus device ID associated with the receiver serial number
recvDeployID	Receiver deployment ID
recvDeployLat	latitude of receiver deployment, in decimal degrees N - negative values for Southern hemisphere
recvDeployLon	longitude of receiver deployment, in decimal degrees E - negative values for Western hemisphere
recvDeployAlt	altitude of receiver deployment, in meters ASL
recv	serial number of the receiver; e.g., SG-1234BBBK5678 or Lotek-12345
recvDeployName	name assigned to the receiver deployment by the project manager
recvSiteName	name assigned to a site by the project manager (e.g. location name). The same label can be used for multiple deployments.
isRecvMobile	logical; whether the sensor is deployed on a mobile platform (eg. a ship)
recvProjID	unique (numeric) ID of the project that deployed this receiver (e.g., 8)
antType	character; antenna type, e.g. "9-element Yagi", "Omni"
antBearing	numeric; compass direction antenna main axis is pointing at (degrees clockwise from local magnetic North 0-360)
antHeight	numeric; height (meters) of antenna main axis above ground
speciesEN	species English (common) name
speciesFR	species French (common) name
speciesSci	species scientific name
speciesGroup	species group, e.g., BIRDS, BATS
tagProjName	short label of project that deployed the tag, e.g., "HolbSESA"
recvProjName	short label of project that deployed the receiver e.g., "HolbSESA"
gpsLat	latitude of receiver GPS location at time of writing the hourly detections file (degrees North)
gpsLon	longitude of receiver GPS location at time of writing the hourly detections file (degrees East)
gpsAlt	altitude of receiver GPS at time of writing the hourly detections file (meters)

## Appendix B

# Appendix - Troubleshooting

As a first step, always ensure you are using the latest version of the motus package (see section @ref(checkVersion.B)), and you have all required packages installed, loaded, and up to date (see Chapter 2).

While attempting to download data with the motus package, you may encounter errors, many of which are likely due to an interrupted connection. **Always ensure you are connected to the internet when using the tagme() function with ‘update = TRUE’.** Most issues can be solved by either logging out of the motus package, or by restarting R and resuming the download using tagme(). If errors persist and you are unable to download your data, the server may be temporarily offline. Please contact Motus with any concerns at motus@birdscanada.org.

### B.1 Logging out of motus

```
motusLogout()
```

### B.2 Resume data download

To resume your data download, run tagme() again, but do not include ‘new = TRUE’:

```
tagme(project.num, update = TRUE, dir = ...)
```

### B.3 Common error messages and solutions:

#### B.3.1 I get the message “Auto-disconnecting SQLiteConnection” one or multiple times after using tagme()

If this occurs after data download has finished, this message can be ignored. If it occurs during an active download, the connection will usually be maintained and the download will continue. However if the download stops, simply run tagme() again. If that does not work, we suggest logging out of the motus package or restarting R (see sections B.1 and B.2).

### B.3.2 I get an “Internal Server Error” message when using `tagme(..., update = TRUE)`

If you get this message while updating your .motus file, use `tagme()` again to continue the download.

### B.3.3 I get an “Error: Forbidden” message when using `tagme()`

This error may occur if you are attempting to download multiple projects simultaneously from the same user account. If you get this error, please logout of the motus package, and try `tagme()` again (see sections B.1 and B.2).

### B.3.4 I get an error “Object ‘xxxx’ not found”, referring to a table or field name, or some of your examples in the book do not work.

Be sure to start the steps from the top of the chapter and run them in sequential order. Another possibility is that your .motus database hasn’t been updated to support the latest version of the motusClient or the motus package. If the `checkVersion` function returns a warning, this may indicate that the internal function used to update your database has not been triggered by `tagme()`. This can happen, for example, if you load the motusClient package without also loading the motus package. Loading the motus package will also load motusClient, so you should only ever need load motus into your R library.

To ensure that your .motus file is up-to-date with the motus package:

```
sql.motus <- tagme(project.num, dir = ...)  
checkVersion(sql.motus)
```

To correct any warnings, you should follow these steps:

1. download the latest versions of the motusClient and then the motus package (refer to Chapter 2).
2. terminate and restart your R session.
3. load the motus library using ‘require(Motus)’ in your R console.
4. load your sqlite file. Look for notes on the console indicating that your database is being updated.
5. check the version again.

```
library(motus)  
sql <- tagme(project.num, dir = ...)  
checkVersion(sql)
```

### B.3.5 I get an error “Error in rssqlite\_connect(dbname, loadable.extensions, flags, vfs) : Could not connect to database: unable to open database file” when attempting to run `tagme()`

If you get this message, it’s likely that you’re attempting a new download or update to a non-existent directory. The directory is specified in the `dir = “”` command of the `tagme()` function. If the directory is not specified, files will be saved to your working directory. Use `getwd()` to determine your current working directory. Use `setwd()` to set a new working directory. To specify a location to save files from your working directory use “./” followed by the file path.

```
getwd() # show working directory, in this case it's 'C:/Documents'  
tagme(proj.num, new = TRUE, update = TRUE) # downloads data to your working directory  
tagme(proj.num, new = TRUE, update = TRUE, dir = "./data/") # downloads data to the data folder within  
tagme(proj.num, new = TRUE, update = TRUE, dir = "C:/Downloads") # downloads data to the file path C:/
```

Of course, there is always the possibility that the book contains errors! If this does not work, please contact [motus@birdscanada.org](mailto:motus@birdscanada.org).

## Appendix C

# Appendix - The Motus R Package

The motus R package offers functions that work with .motus data to do common computations, summaries and plots. This appendix outlines these functions and provides examples on function use. Many of these functions work with both `tbl` and `data.frame` formats, however some require the data to be in `sql` format as specified below. Detailed instructions on accessing and formatting data are available in Chapter 3. The examples throughout this chapter work with the sample data which can be accessed and converted to various formats through the following code:

```
# download and access sample data in sql format,
# username: motus.sample, password: motus.sample
sql.motus <- tagme(176, new = TRUE, update = TRUE,
  dir = "./data")

# extract 'alltags' table from sql file
# 'sql.motus'
tbl.alltags <- tbl(sql.motus, "alltags")

## convert the tbl 'tbl.alltags' to a data.frame
## called 'df.alltags'
df.alltags <- tbl.alltags %>% collect() %>% as.data.frame()
```

You can access the function help pages using ‘`??sunRiseSet`’ in the R console. Or view the underlying function code like this:

```
sunRiseSet
```

### C.1 checkVersion

#### C.1.1 Description

When you call the `tagme()` function to load the sqlite database, there is a process that will verify that your database has the version matching the most current version of the motus package and store the version in a new table called `admInfo`. Over time, changes will be made that require adding new tables, views or fields to the database. The following call will check that your database has been updated to the version matching the current version of the motus package. Refer to Appendix B if this call returns a warning; if you do not have the most recent version, see Chapter 2 to update `motus` and `motusClient`.

### C.1.2 Dependencies

`sql.motus` an sqlite database of .motus data downloaded using `tagme()`

### C.1.3 Example

```
checkVersion(sql.motus)
```

## C.2 sunRiseSet

### C.2.1 Description

Creates and adds a sunrise and sunset variable to a data.frame containing latitude, longitude, and a date/time as POSIXct or numeric.

### C.2.2 Dependencies

`data` can be either a selected table from .motus detection data e.g. “alltags”, or a data.frame of detection data including at a minimum variables for date/time, latitude, and longitude

`lat` variable with latitude values, defaults to `recvDeployLat`

`lon` variable with longitude values, defaults to `recvDeployLon`

`ts` variable with time in UTC as numeric or POSIXct, defaults to `ts`

### C.2.3 Example

Add sunrise/sunset variables to the alltags data.frame

```
alltags.df.sun <- sunRiseSet(df.alltags)
head(alltags.df.sun)
```

```
##   hitID runID batchID          ts sig sigsd noise freq freqsd
## 1  45107  8886      53 2015-10-26 11:19:49  52     0  -96     4      0
## 2  45108  8886      53 2015-10-26 11:20:28  54     0  -96     4      0
## 3  45109  8886      53 2015-10-26 11:21:17  55     0  -96     4      0
## 4  45110  8886      53 2015-10-26 11:21:55  52     0  -96     4      0
## 5  45111  8886      53 2015-10-26 11:22:44  49     0  -96     4      0
## 6 199885 23305     64 2015-10-26 11:12:04  33     0  -96     4      0
##   slop burstSlop done motusTagID ambigID port runLen bootnum tagProjID
## 1 1e-04    0.0000    1     16047     NA    3      5     11     176
## 2 1e-04   -0.0021    1     16047     NA    3      5     11     176
## 3 1e-04    0.0001    1     16047     NA    3      5     11     176
## 4 1e-04   -0.0010    1     16047     NA    3      5     11     176
## 5 1e-04    0.0001    1     16047     NA    3      5     11     176
## 6 1e-04    0.0000    1     16047     NA    1     11      4     176
##   mfgID tagType codeSet    mfg tagModel tagLifespan nomFreq tagBI pulseLen
## 1    378      ID  Lotek4  Lotek  NTQB-3-2        NA 166.38 9.6971     2.5
## 2    378      ID  Lotek4  Lotek  NTQB-3-2        NA 166.38 9.6971     2.5
## 3    378      ID  Lotek4  Lotek  NTQB-3-2        NA 166.38 9.6971     2.5
## 4    378      ID  Lotek4  Lotek  NTQB-3-2        NA 166.38 9.6971     2.5
```

```

## 5 378 ID Lotek4 Lotek NTQB-3-2 NA 166.38 9.6971 2.5
## 6 378 ID Lotek4 Lotek NTQB-3-2 NA 166.38 9.6971 2.5
## tagDeployID speciesID markerNumber markerType tagDeployStart
## 1 1839 4670 135268103 metal band 1441908000
## 2 1839 4670 135268103 metal band 1441908000
## 3 1839 4670 135268103 metal band 1441908000
## 4 1839 4670 135268103 metal band 1441908000
## 5 1839 4670 135268103 metal band 1441908000
## 6 1839 4670 135268103 metal band 1441908000
## tagDeployEnd tagDeployLat tagDeployLon tagDeployAlt
## 1 1457632800 51.4839 -80.45 NA
## 2 1457632800 51.4839 -80.45 NA
## 3 1457632800 51.4839 -80.45 NA
## 4 1457632800 51.4839 -80.45 NA
## 5 1457632800 51.4839 -80.45 NA
## 6 1457632800 51.4839 -80.45 NA
##
## 1 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fullID": "M.16047", "deviceID": "Lotek-159", "recvDeployID": "9.7@166.38", "recvDeployLat": 51.4839, "recvDeployLon": -80.45, "recvDeployAlt": 2510, "recvName": "Shelburne", "recvSiteName": "<NA>"}
## 2 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fullID": "M.16047", "deviceID": "Lotek-159", "recvDeployID": "9.7@166.38", "recvDeployLat": 51.4839, "recvDeployLon": -80.45, "recvDeployAlt": 2510, "recvName": "Shelburne", "recvSiteName": "<NA>"}
## 3 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fullID": "M.16047", "deviceID": "Lotek-159", "recvDeployID": "9.7@166.38", "recvDeployLat": 51.4839, "recvDeployLon": -80.45, "recvDeployAlt": 2510, "recvName": "Shelburne", "recvSiteName": "<NA>"}
## 4 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fullID": "M.16047", "deviceID": "Lotek-159", "recvDeployID": "9.7@166.38", "recvDeployLat": 51.4839, "recvDeployLon": -80.45, "recvDeployAlt": 2510, "recvName": "Shelburne", "recvSiteName": "<NA>"}
## 5 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fullID": "M.16047", "deviceID": "Lotek-159", "recvDeployID": "9.7@166.38", "recvDeployLat": 51.4839, "recvDeployLon": -80.45, "recvDeployAlt": 2510, "recvName": "Shelburne", "recvSiteName": "<NA>"}
## 6 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fullID": "M.16047", "deviceID": "Lotek-159", "recvDeployID": "9.7@166.38", "recvDeployLat": 51.4839, "recvDeployLon": -80.45, "recvDeployAlt": 2510, "recvName": "Shelburne", "recvSiteName": "<NA>"}
## recvDeployLon recvDeployAlt recvName recvSiteName
## 1 -72.71657 NA Lotek-159 Shelburne <NA>
## 2 -72.71657 NA Lotek-159 Shelburne <NA>
## 3 -72.71657 NA Lotek-159 Shelburne <NA>
## 4 -72.71657 NA Lotek-159 Shelburne <NA>
## 5 -72.71657 NA Lotek-159 Shelburne <NA>
## 6 -72.47392 NA Lotek-164 BennettMeadow <NA>
## isRecvMobile recvProjID antType antBearing antHeight speciesEN
## 1 0 74 yagi-9 127 NA Red Knot
## 2 0 74 yagi-9 127 NA Red Knot
## 3 0 74 yagi-9 127 NA Red Knot
## 4 0 74 yagi-9 127 NA Red Knot
## 5 0 74 yagi-9 127 NA Red Knot
## 6 0 74 yagi-9 243 NA Red Knot
## speciesFR speciesSci speciesGroup tagProjName
## 1 Bécasseau maubèche Calidris canutus BIRDS SampleData
## 2 Bécasseau maubèche Calidris canutus BIRDS SampleData
## 3 Bécasseau maubèche Calidris canutus BIRDS SampleData
## 4 Bécasseau maubèche Calidris canutus BIRDS SampleData
## 5 Bécasseau maubèche Calidris canutus BIRDS SampleData
## 6 Bécasseau maubèche Calidris canutus BIRDS SampleData
## recvProjName gpsLat gpsLon gpsAlt sunrise
## 1 <NA> NA NA NA 2015-10-26 11:16:49
## 2 <NA> NA NA NA 2015-10-26 11:16:49

```

```

## 3      <NA>     NA     NA    NA 2015-10-26 11:16:49
## 4      <NA>     NA     NA    NA 2015-10-26 11:16:49
## 5      <NA>     NA     NA    NA 2015-10-26 11:16:49
## 6      <NA>     NA     NA    NA 2015-10-26 11:15:58
##          sunset
## 1 2015-10-26 21:52:11
## 2 2015-10-26 21:52:11
## 3 2015-10-26 21:52:11
## 4 2015-10-26 21:52:11
## 5 2015-10-26 21:52:11
## 6 2015-10-26 21:51:06

```

## C.3 plotAllTagsCoord

### C.3.1 Description

Plot latitude/longitude vs time (UTC rounded to the hour) for each tag using .motus detection data. Coordinate is by default taken from a receivers GPS latitude recordings.

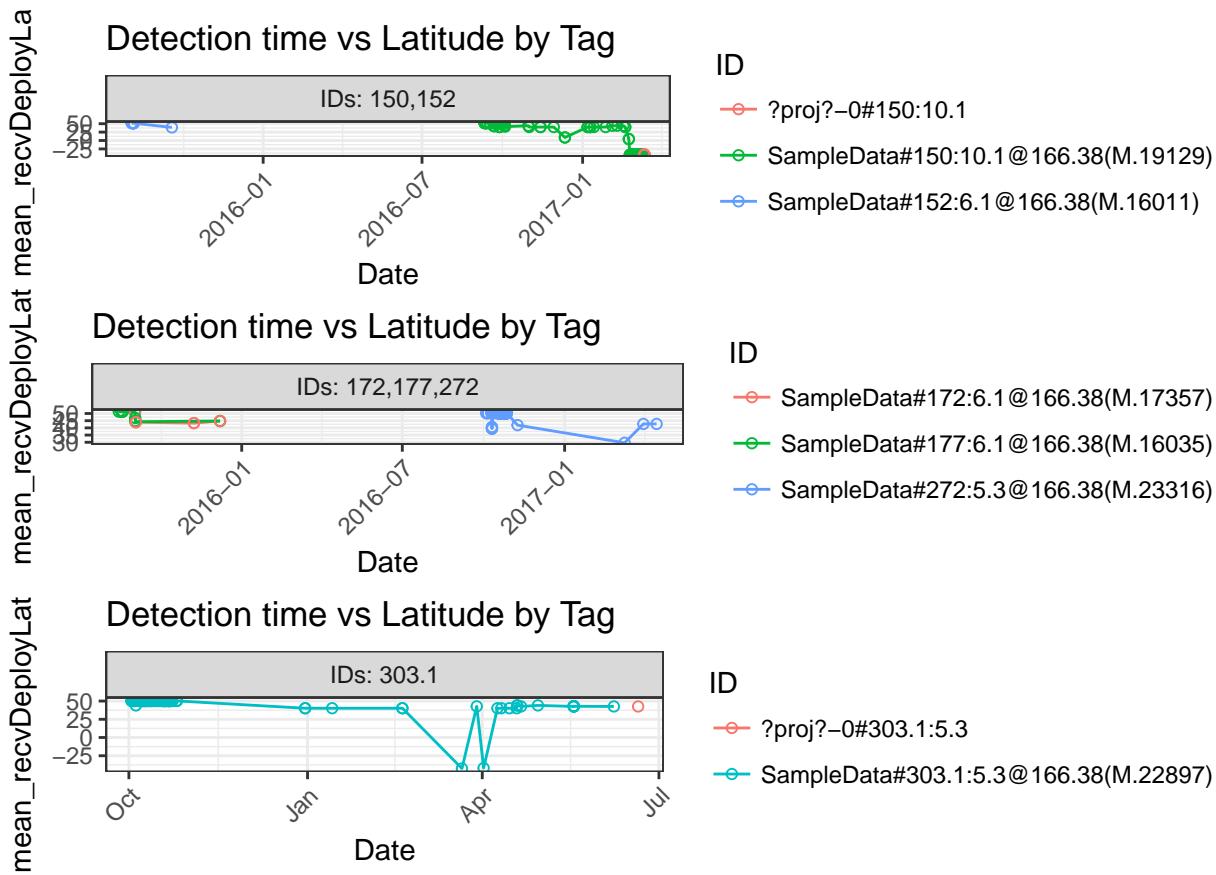
### C.3.2 Dependencies

**data** a selected table from .motus detection data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for date/time, and either latitude or longitude  
**tagsPerPanel** number of tags in each panel of the plot, by default this is 5  
**coordinate** variable name from which to obtain location values, by default it is set to recvDeployLat  
**ts** variable for a date/time object as numeric or POSIXct, defaults to ts  
**recvDepName** variable consisting of receiver deployment name  
**fullID** variable consisting of a tag fullID  
**mfgID** variable consisting of a tags manufacturer ID

### C.3.3 Example

Plot select tags from tbl.alltags with 3 tags per panel

```
plotAllTagsCoord(filter(tbl.alltags, motusTagID %in%
c(19129, 16011, 17357, 16035, 22897, 23316)), tagsPerPanel = 3)
```



## C.4 plotAllTagsSite

### C.4.1 Description

Plot latitude/longitude vs time (UTC rounded to the hour) for each tag using .motus detection data. Coordinate is by default taken from a receivers GPS latitude recordings.

### C.4.2 Dependencies

**data** a selected table from .motus detection data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for date/time, and either latitude or longitude  
**tagsPerPanel** number of tags in each panel of the plot, by default this is 5  
**coordinate** variable name from which to obtain location values, by default it is set to recvDeployLat  
**ts** variable for a date/time object as numeric or POSIXct, defaults to ts  
**recvDepName** variable consisting of receiver deployment name  
**fullID** variable consisting of a tag fullID  
**mfgID** variable consisting of a tags manufacturer ID

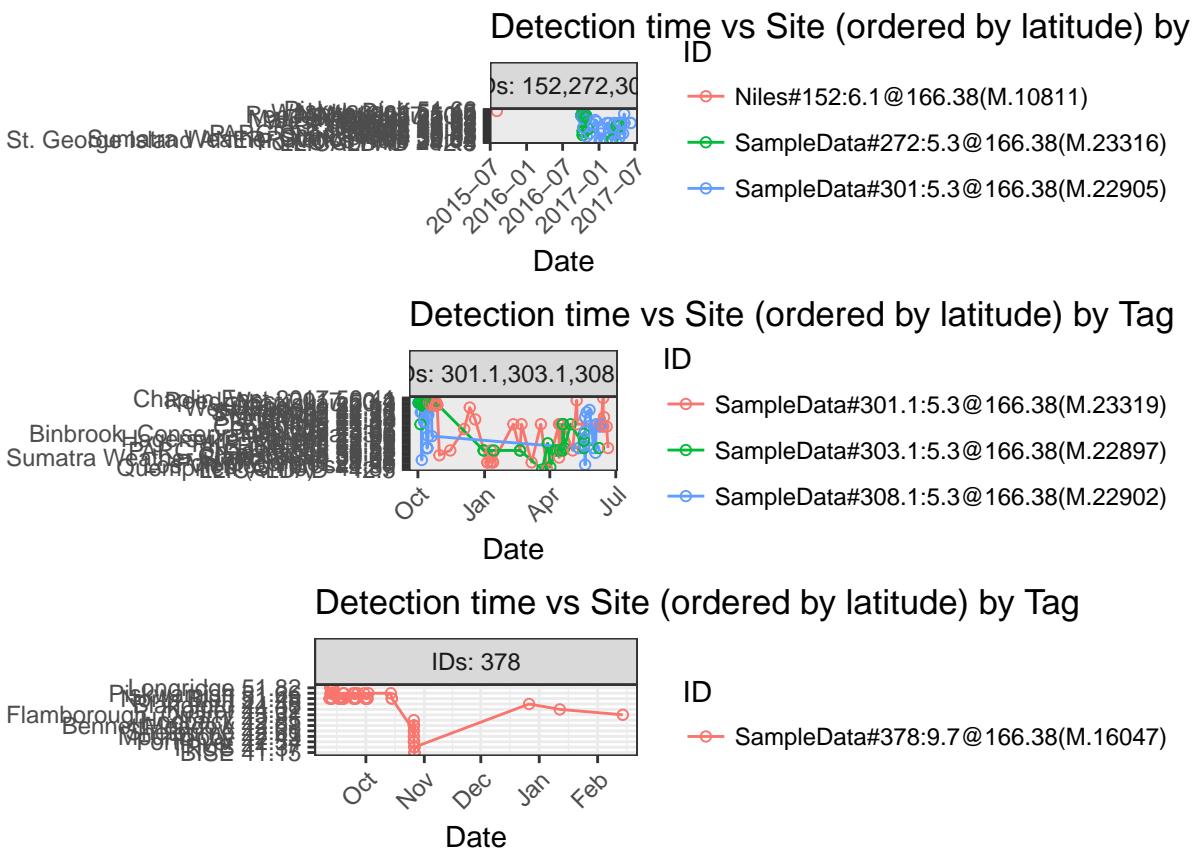
### C.4.3 Example

Plot tbl file tbl.alltags using gpsLat and 3 tags per panel for select species Red Knot

```
plotAllTagsSite(filter(tbl.alltags, speciesEN == "Red Knot"),
  coordinate = "recvDeployLat", tagsPerPanel = 3)
```

```
## Warning: Missing values are always removed in SQL.  
## Use `AVG(x, na.rm = TRUE)` to silence this warning
```

ordered by recvDeployLat ordered by recvDeployLat ordered by recvDeployLat



## C.5 plotDailySiteSum

### C.5.1 Description

Plots total number of detections across all tags, and total number of tags detected per day for a specified site. Depends on siteSumDaily function.

### C.5.2 Dependencies

**data** a selected table from .motus data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for motusTagID, sig, recvDepName, ts

**motusTagID** variable consisting of a motus tag ID

**sig** variable consisting a signal strength variable

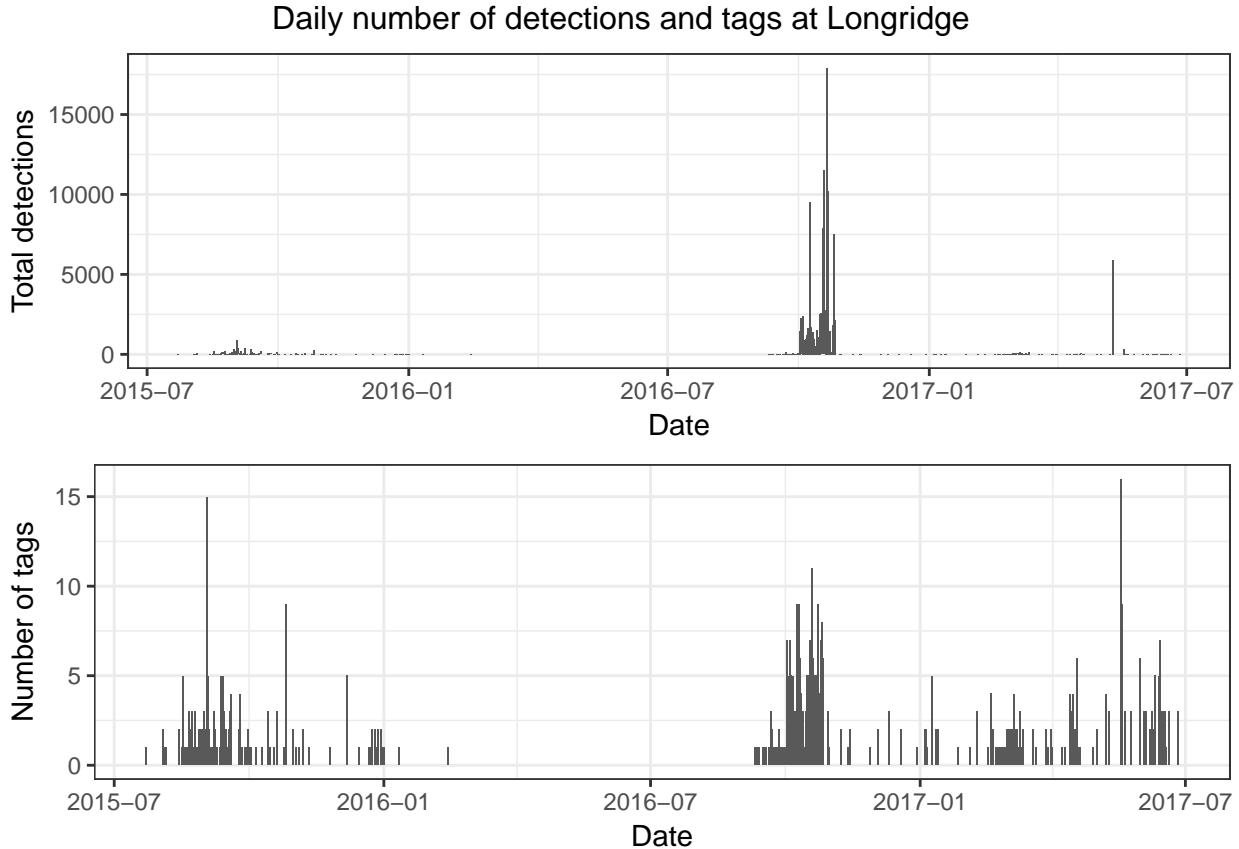
**recvDepName** variable consisting of receiver deployment name

**ts** variable for a date/time object as numeric or POSIXct, defaults to ts

### C.5.3 Example

Plot of all tag detections at site Longridge using data.frame df.alltags

```
plotDailySiteSum(df.alltags, recvDeployName = "Longridge")
```



## C.6 plotRouteMap

### C.6.1 Description

Google map of routes of Motus tag detections coloured by motusTagID. User defines a date range to show points for receivers that were operational at some point during specified date range. ### Dependencies  
data a .motus sql file

**maptype** google map type to display, can be: “terrain”, “roadmap”, “satellite”, or “hybrid”

**latCentre** latitude to centre map around

**lonCentre** longitude to centre map around

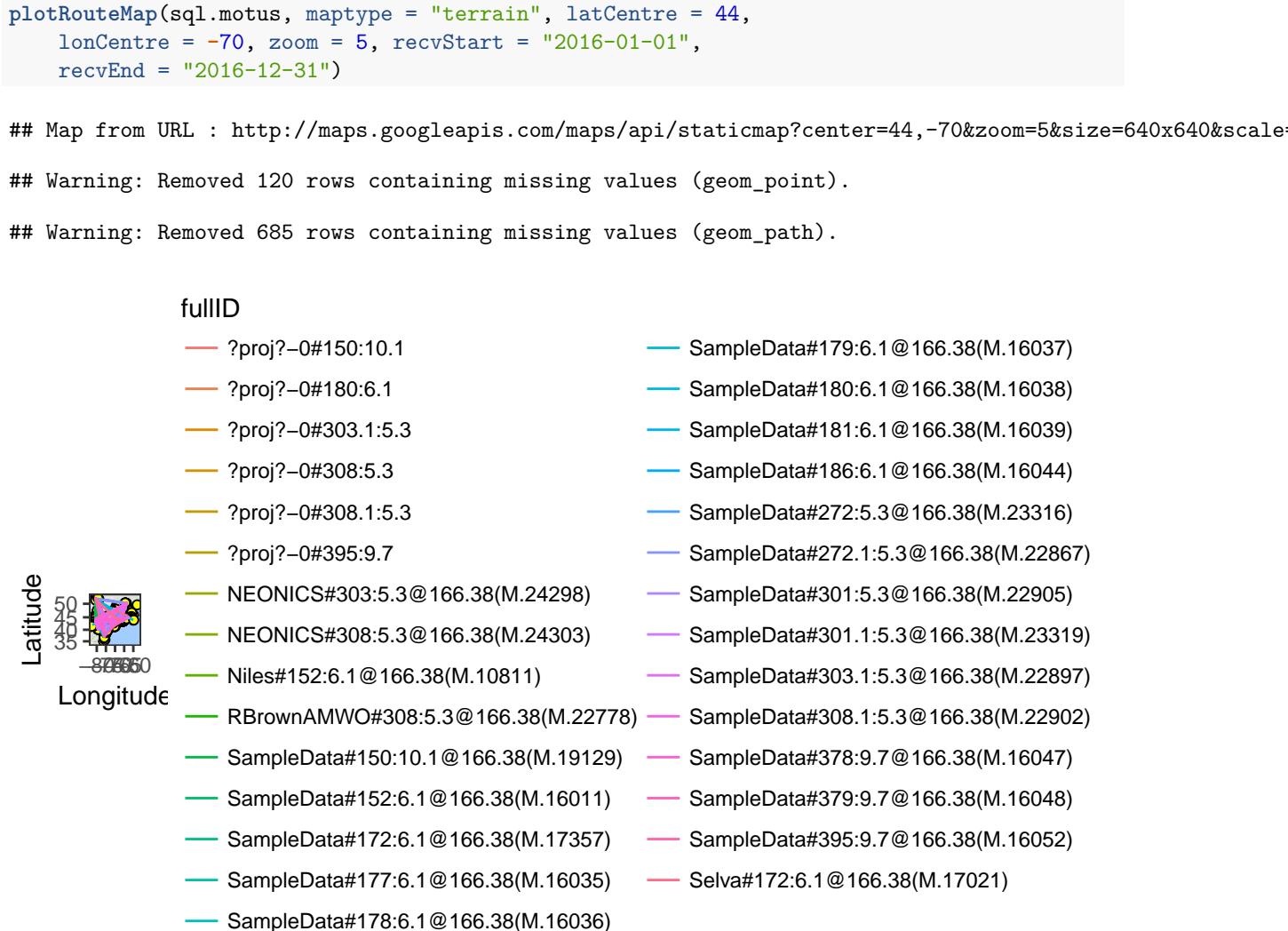
**zoom** integer for zoom 3-21, 3 being continent level, 10 being city-scale

**recvStart** start date for date range of active receivers

**recvEnd** end date for date range of active receivers

### C.6.2 Example

Plot routemap of all detection data, with “terrain” maptype, and receivers active between 2016-01-01 and 2017-01-01



## C.7 plotSite

### C.7.1 Description

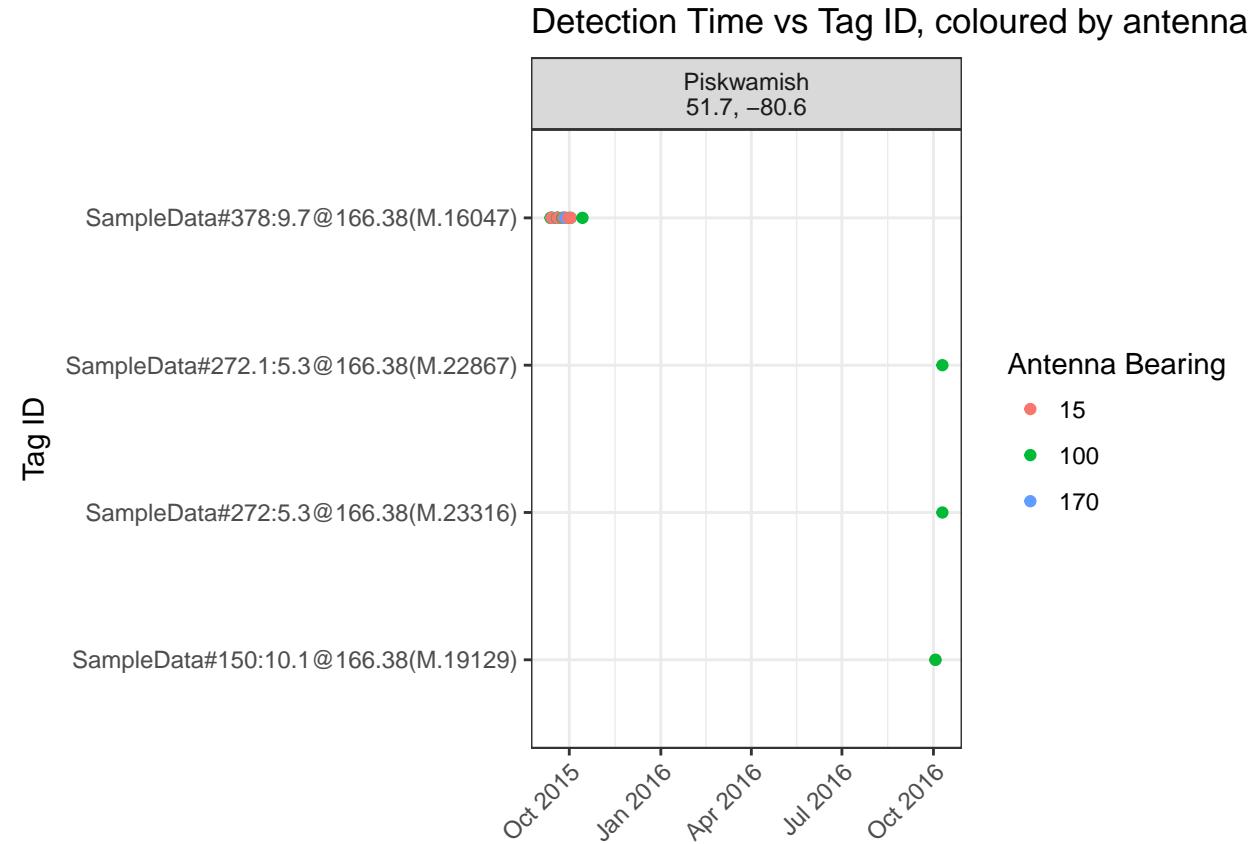
### C.7.2 Dependencies

**data** a selected table from .motus data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for ts, antBearing, fullID, recvDepName  
**ts** variable for a date/time object as numeric or POSIXct, defaults to ts  
**antBearing** variable consisting antenna bearing variable  
**fullID** variable consisting of a tag fullID  
**recvDepName** variable consisting of receiver deployment name

### C.7.3 Example

Plot only detections at a specific site; Piskwamish for data.frame df.alltags

```
plotSite(filter(df.alltags, recvDeployName == "Piskwamish"))
```



## C.8 plotSiteSig

### C.8.1 Description

Plot signal strength vs time for all tags detected at a specified site, coloured by antenna

### C.8.2 Dependencies

**data** a selected table from .motus data, e.g. "alltags", or a data.frame of detection data including at a minimum variables for antBearing, ts, lat, sig, fullID, recvDepName

**antBearing** variable consisting antenna bearing variable

**ts** variable for a date/time object as numeric or POSIXct, defaults to ts

**recvDeployLat** variable consisting of receiver deployment latitude

**sig** variable consisting a signal strength variable

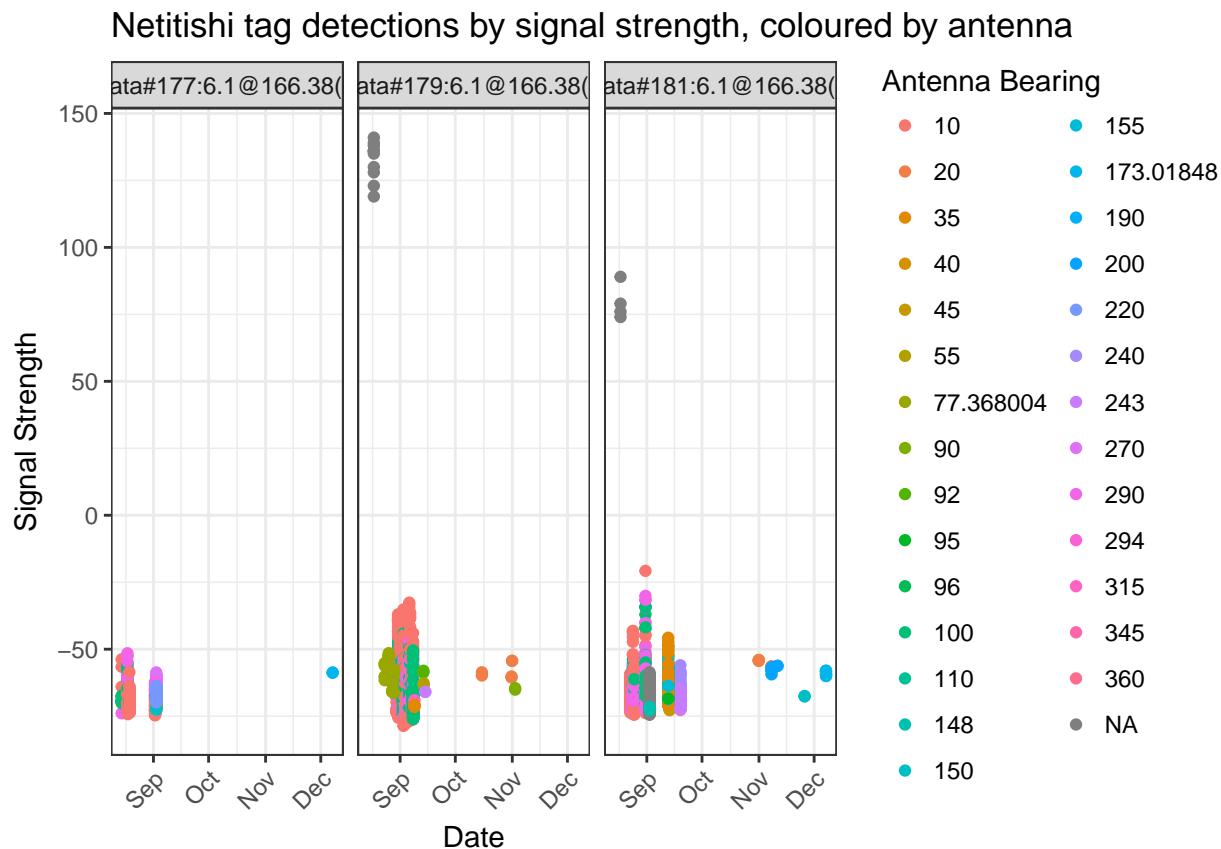
**fullID** variable consisting of a tag fullID

**recvDepName** variable consisting of receiver deployment name

### C.8.3 Example

Plot select tags for site Piskwamish

```
plotSiteSig(filter(df.alltags, motusTagID %in% c(16037,
16039, 16035)), recvDeployName = "Netitishi")
```



## C.9 plotTagSig

### C.9.1 Description

Plot signal strength vs time for specified tag, faceted by site (ordered by latitude) and coloured by antenna

### C.9.2 Dependencies

**data** a selected table from .motus data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for motusTagID, sig, ts, antBearing, recvDeployLat, fullID, recvDepName

**motusTagID** variable consisting of a motus tag ID

**antBearing** variable consisting antenna bearing variable

**ts** variable for a date/time object as numeric or POSIXct, defaults to ts

**recvDeployLat** variable consisting of receiver deployment latitude

**sig** variable consisting a signal strength variable

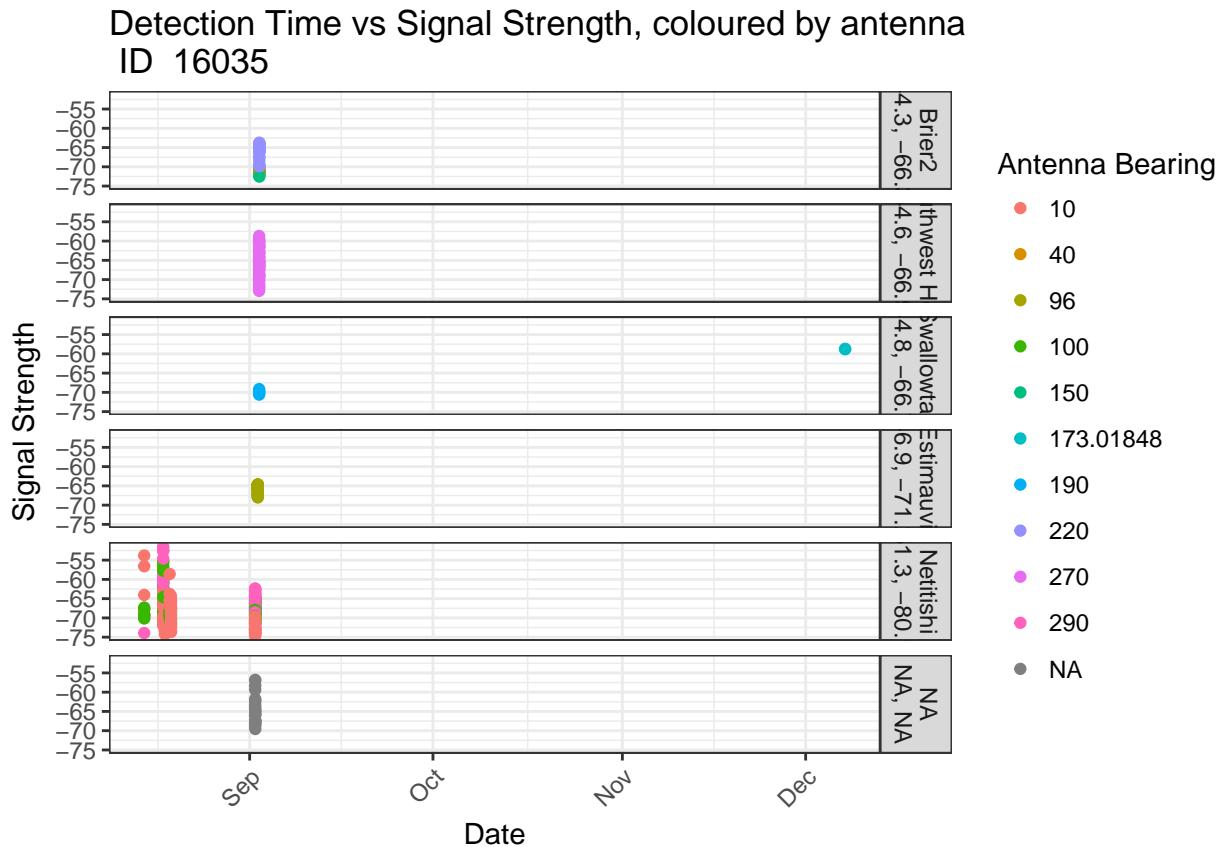
**fullID** variable consisting of a tag fullID

**recvDepName** variable consisting of receiver deployment name

### C.9.3 Example

Plot signal strength of a specified tag using tbl file tbl.alltags

```
plotTagSig(tbl.alltags, motusTagID = 16035)
```



## C.10 simSiteDet

### C.10.1 Description

Creates a data.frame consisting of only detections of tags that are detected at two or more receivers at the same time.

### C.10.2 Dependencies

**data** a selected table from .motus data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for ts, motusTagID, recvDepName

**ts** variable for a date/time object as numeric or POSIXct, defaults to ts

**motusTagID** variable consisting of a motus tag ID

**recvDepName** variable consisting of receiver deployment name

### C.10.3 Example

To get a data.frame called "simSites" of just simultaneous detections from a data.frame df.alltags

```
simSites <- simSiteDet(df.alltags)  
head(simSites)
```

```

## motusTagID          ts num.dup    hitID runID batchID
## 1      16047 2015-10-26 11:20:28    2    45108    8886    53
## 2      16047 2015-10-26 11:20:28    2   199896   23305    64
## 3      16047 2015-10-26 11:22:44    2   199902   23306    64
## 4      16047 2015-10-26 11:22:44    2    45111    8886    53
## 5      22897 2017-05-18 02:27:59    2 159449264  9005196  64491
## 6      22897 2017-05-18 02:27:59    2 201832550 10191953  66308
##           sig   sigsd noise freq freqsd  slop burstSlop done ambigID
## 1  54.0000  0.0000 -96.0000 4.0000 0.0000 1e-04 -0.0021    1     NA
## 2  20.0000  0.0000 -96.0000 4.0000 0.0000 1e-04  0.0012    1     NA
## 3  28.0000  0.0000 -96.0000 4.0000 0.0000 1e-04 -0.0006    1     NA
## 4  49.0000  0.0000 -96.0000 4.0000 0.0000 1e-04  0.0001    1     NA
## 5 -63.7422 14.6880 -76.3434 2.7883 0.0196 1e-04  0.0003    1   -114
## 6 -67.5413 14.6268 -75.6038 2.6942 0.0061 1e-04  0.0000    1   -114
## port runLen bootnum tagProjID mfgID tagType codeSet   mfg tagModel
## 1      3      5     11     176    378     ID Lotek4 Lotek NTQB-3-2
## 2      1     11     4     176    378     ID Lotek4 Lotek NTQB-3-2
## 3      3      6     4     176    378     ID Lotek4 Lotek NTQB-3-2
## 4      3      5     11     176    378     ID Lotek4 Lotek NTQB-3-2
## 5      3      5    145     176  303.1     ID Lotek4 Lotek NTQB-4-2
## 6      1      6    374     176  303.1     ID Lotek4 Lotek NTQB-4-2
## tagLifespan nomFreq tagBI pulseLen tagDeployID speciesID markerNumber
## 1        NA 166.38  9.6971     2.5     1839    4670 135268103
## 2        NA 166.38  9.6971     2.5     1839    4670 135268103
## 3        NA 166.38  9.6971     2.5     1839    4670 135268103
## 4        NA 166.38  9.6971     2.5     1839    4670 135268103
## 5        NA 166.38  5.2978     2.5    10485    4670 9822-53123
## 6        NA 166.38  5.2978     2.5    10485    4670 9822-53123
## markerType tagDeployStart tagDeployEnd tagDeployLat tagDeployLon
## 1 metal band    1441908000 1457632800  51.48390 -80.45000
## 2 metal band    1441908000 1457632800  51.48390 -80.45000
## 3 metal band    1441908000 1457632800  51.48390 -80.45000
## 4 metal band    1441908000 1457632800  51.48390 -80.45000
## 5 metal band   1475337600 1497283200  50.19278 -63.74528
## 6 metal band   1475337600 1497283200  50.19278 -63.74528
## tagDeployAlt
## 1        NA
## 2        NA
## 3        NA
## 4        NA
## 5        NA
## 6        NA
##
## 1 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fatScore": 3, "locationID": 1}
## 2 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fatScore": 3, "locationID": 1}
## 3 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fatScore": 3, "locationID": 1}
## 4 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fatScore": 3, "locationID": 1}
## 5

```

```

## 6
##                                     fullID deviceID recvDeployID recvDeployLat
## 1  SampleData#378:9.7@166.38(M.16047)      486      2510  42.60699
## 2  SampleData#378:9.7@166.38(M.16047)      515      2512  42.68067
## 3  SampleData#378:9.7@166.38(M.16047)      515      2512  42.68067
## 4  SampleData#378:9.7@166.38(M.16047)      486      2510  42.60699
## 5 SampleData#303.1:5.3@166.38(M.22897)     297      2146  42.61540
## 6 SampleData#303.1:5.3@166.38(M.22897)     304      2549  42.58290
##   recvDeployLon recvDeployAlt          recv recvDeployName recvSiteName
## 1    -72.71657           NA  Lotek-159  Shelburne      <NA>
## 2    -72.47392           NA  Lotek-164 BennettMeadow  <NA>
## 3    -72.47392           NA  Lotek-164 BennettMeadow  <NA>
## 4    -72.71657           NA  Lotek-159  Shelburne      <NA>
## 5    -80.45810           NA SG-5113BBBBK2853      BSC HQ      <NA>
## 6    -80.39840           NA SG-5113BBBBK2972      Old Cut      <NA>
##   isRecvMobile recvProjID      antType antBearing antHeight speciesEN
## 1          0        74      yagi-9       127      NA Red Knot
## 2          0        74      yagi-9       243      NA Red Knot
## 3          0        74      yagi-9       120      NA Red Knot
## 4          0        74      yagi-9       127      NA Red Knot
## 5          0        1 yagi-9-laird      180      NA Red Knot
## 6          0        1 yagi-9-laird      290       25 Red Knot
##   speciesFR      speciesSci speciesGroup tagProjName
## 1 Bécasseau maubèche Calidris canutus      BIRDS SampleData
## 2 Bécasseau maubèche Calidris canutus      BIRDS SampleData
## 3 Bécasseau maubèche Calidris canutus      BIRDS SampleData
## 4 Bécasseau maubèche Calidris canutus      BIRDS SampleData
## 5 Bécasseau maubèche Calidris canutus      BIRDS SampleData
## 6 Bécasseau maubèche Calidris canutus      BIRDS SampleData
##   recvProjName gpsLat gpsLon gpsAlt
## 1      <NA>    NA    NA    NA
## 2      <NA>    NA    NA    NA
## 3      <NA>    NA    NA    NA
## 4      <NA>    NA    NA    NA
## 5      <NA>    NA    NA    NA
## 6      <NA>    NA    NA    NA

```

## C.11 siteSum

### C.11.1 Description

Creates a summary of the first and last detection at a site, the length of time between first and last detection, the number of tags, and the total number of detections at a site. Plots total number of detections across all tags, and total number of tags detected at each site.

### C.11.2 Dependencies

**data** a selected table from .motus data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for motusTagID, sig, recvDeployLat, recvDepName, and ts

**motusTagID** variable consisting of a motus tag ID

**sig** variable consisting a signal strength variable

**recvDeployLat** variable consisting of receiver deployment latitude  
**recvDepName** variable consisting of receiver deployment name  
**ts** variable for a date/time object as numeric or POSIXct, defaults to ts  
**units** units to display time difference, defaults to “hours”, options include “secs”, “mins”, “hours”, “days”, “weeks”

### C.11.3 Example

Create site summaries for select sites with time in minutes

```
site_summary <- siteSum(filter(df.alltags, recvDeployName %in%
  c("Niapiskau", "Netitishi", "Old Cur", "Washkaugou")),
  units = "mins")
```



```
head(site_summary)
```

```
## # A tibble: 3 x 6
##   recvDeployName    first_ts           last_ts          tot_ts num.tags
##   <fct>            <dttm>            <dttm>          <time>    <int>
## 1 Niapiskau_50.2~ 2016-10-01 23:47:57 2016-10-27 00:03:21 36015.~      6
## 2 Washkaugou_51.~ 2016-10-09 23:52:45 2016-10-10 00:00:42 7.9468~      2
## 3 Netitishi_51.3~ 2015-08-14 17:53:49 2015-09-08 01:10:13 34996.~      7
## # ... with 1 more variable: num.det <int>
```

## C.12 siteSumDaily

### C.12.1 Description

Creates a summary of the first and last daily detection at a site, the length of time between first and last detection, the number of tags, and the total number of detections at a site for each day. Same as siteSum, but daily by site.

### C.12.2 Dependencies

**data** a selected table from .motus data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for motusTagID, sig, recvDepName, ts  
**motusTagID** variable consisting of a motus tag ID  
**sig** variable consisting a signal strength variable  
**recvDepName** variable consisting of receiver deployment name  
**ts** variable for a date/time object as numeric or POSIXct, defaults to ts  
**units** units to display time difference, defaults to “hours”, options include “secs”, “mins”, “hours”, “days”, “weeks”

### C.12.3 Example

Create site summaries for all sites within detection data with time in minutes using tbl file tbl.alltags

```
daily_site_summary <- siteSumDaily(tbl.alltags, units = "mins")
head(daily_site_summary)
```

```
##                                     recvDeployName      date      first_ts
## 1           Assateague State Park\n38.2, -75.1 2015-09-13 2015-09-13 10:12:50
## 2                   Baccaro\n43.5, -65.5 2017-05-19 2017-05-19 16:01:21
## 3           BennettMeadow\n42.7, -72.5 2015-10-26 2015-10-26 11:12:04
## 4 Binbrook_Conservation_Area\n43.1, -79.8 2017-05-18 2017-05-18 03:13:25
## 5                  BISE\n41.2, -71.6 2015-10-26 2015-10-26 17:55:47
## 6            Blandford\n44.5, -64.1 2015-12-26 2015-12-26 14:58:27
##               last_ts      tot_ts num_tags num_det
## 1 2015-09-13 10:14:40 1.828847 mins     1       6
## 2 2017-05-19 16:02:40 1.323297 mins     3       6
## 3 2015-10-26 11:30:49 18.747692 mins     1      27
## 4 2017-05-18 03:13:46 0.353195 mins     2       6
## 5 2015-10-26 19:16:55 81.132212 mins     1      44
## 6 2015-12-26 14:58:47 0.323150 mins     1       2
```

## C.13 siteTrans

### C.13.1 Description

Creates a data.frame of transitions between sites; detections are ordered by detection time, then “transitions” are identified as the period between the final detection at site x (possible “departure”), and the first detection (possible “arrival”) at site y (ordered chronologically). Each row contains the last detection time and lat/lon of site x, first detection time and lat/lon of site y, distance between the site pair, time between detections, rate of movement between detections, and bearing between site pairs.

### C.13.2 Dependencies

**data** a selected table from .motus data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for ts, motusTagID, tagDeployID, recvDeployLat, recvDeployLon, recvDepName  
**ts** variable for a date/time object as numeric or POSIXct, defaults to ts  
**motusTagID** variable consisting of a motus tag ID  
**tagDeployID** variable consisting of Motus tag deployment ID  
**recvDeployLat** variable consisting of receiver deployment latitude  
**recvDeployLon** variable consisting of receiver deployment longitude  
**recvDepName** variable consisting of receiver deployment name

### C.13.3 Example

View site transitions for only tag 16037 from data.frame df.alltags

```
transitions <- siteTrans(filter(df.alltags, motusTagID ==  

  16037), latCoord = "recvDeployLat", lonCoord = "recvDeployLon")  

head(transitions)
```

```
## # A tibble: 6 x 14  

## # Groups:   motusTagID, tagDeployID [1]  

##   motusTagID tagDeployID ts.x           lat.x lon.x recvDeployName.x  

##       <int>      <int> <dttm>          <dbl> <dbl> <chr>  

## 1     16037      1825 2015-08-17 17:02:39  NA    NA  NP mobile_NA, NA  

## 2     16037      1825 2015-08-28 16:40:18  51.5  -80.4 North Bluff_51.5~  

## 3     16037      1825 2015-09-08 01:10:13  51.3  -80.1 Netitishi_51.3, ~  

## 4     16037      1825 2015-09-08 18:37:16  44.6  -65.8 Comeau (Marshall~  

## 5     16037      1825 2015-09-13 19:46:27  39.0  -74.8 NWW_39, -74.8  

## 6     16037      1825 2015-09-14 15:56:49  37.1  -76.0 BULL_37.1, -76  

## # ... with 8 more variables: ts.y <dttm>, lat.y <dbl>, lon.y <dbl>,  

## #   recvDeployName.y <chr>, tot_ts <time>, dist <dbl>, rate <dbl>,  

## #   bearing <dbl>
```

## C.14 tagSum

### C.14.1 Description

Creates a summary for each tag of it’s first and last detection time, first and last detection site, length of time between first and last detection, straight line distance between first and last detection site, rate of movement, and bearing

### C.14.2 Dependencies

**data** a selected table from .motus data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for motusTagID, fullID, recvDeployLat, recvDeployLon, recvDepName, ts  
**motusTagID** variable consisting of a motus tag ID  
**fullID** variable consisting of a tag fullID  
**recvDeployLat** variable consisting of receiver deployment latitude  
**recvDeployLon** variable consisting of receiver deployment longitude  
**recvDepName** variable consisting of receiver deployment name  
**ts** variable for a date/time object as numeric or POSIXct, defaults to ts

### C.14.3 Example

Create tag summary for all tags within detection data using tbl file tbl.alltags

```
tag_summary <- tagSum(tbl.alltags)
head(tag_summary)

##                                     fullID      first_ts
## 1      Niles#152:6.1@166.38(M.10811) 2015-08-03 06:37:11
## 2                  ?proj?-0#395:9.7 2015-07-23 10:10:54
## 3      Selva#172:6.1@166.38(M.17021) 2015-09-02 04:06:07
## 4 SampleData#152:6.1@166.38(M.16011) 2015-08-03 06:37:11
## 5 SampleData#395:9.7@166.38(M.16052) 2015-09-12 17:38:04
## 6 SampleData#179:6.1@166.38(M.16037) 2015-08-17 17:01:38
##           last_ts      first_site      last_site
## 1 2015-08-03 06:37:35 North Bluff_51.5, -80.5 North Bluff_51.5, -80.5
## 2 2015-09-02 20:06:13    Machias_44.5, -67.1    Ruby's_62.9, -143.7
## 3 2015-09-03 00:27:16   Netitishi_51.3, -80.1       MDR_44, -68.1
## 4 2015-09-18 09:37:39 North Bluff_51.5, -80.5       NWW_39, -74.8
## 5 2015-10-20 20:43:43   Longridge_51.8, -80.7 Mount Thom_45.6, -63
## 6 2015-11-02 13:21:42          NP mobile_NA, NA Hillman_Marsh_42, -82.5
##   recvLat.x recvLon.x recvLat.y recvLon.y      tot_ts      dist
## 1     51.4839   -80.4500    51.48390   -80.45000    24.386 secs      0
## 2     44.5023   -67.1018   62.89072  -143.68170 3578118.674 secs 5087724
## 3     51.2913   -80.1168   43.96893   -68.12830   73268.958 secs 1211577
## 4     51.4839   -80.4500    39.02827   -74.81004 3985228.228 secs 1452237
## 5     51.8231   -80.6912   45.55480   -62.98590 3294338.784 secs 1472844
## 6          NA          NA    42.04270   -82.51440 6639604.044 secs      NA
##   rate      bearing num_det
## 1 0.0000000 -180.00000      4
## 2 1.4218993  -38.36268     12
## 3 16.5360200  127.53144    279
## 4 0.3644051  160.20437     127
## 5 0.4470832  111.25011     147
## 6          NA          NA    1353
```

## C.15 tagSumSite

### C.15.1 Description

Creates a summary for each tag of it's first and last detection time at each site, length of time between first and last detection of each site, and total number of detections at each site.

### C.15.2 Dependencies

**data** a selected table from .motus data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for motusTagID, fullID, recvDepName, ts

**motusTagID** variable consisting of a motus tag ID

**fullID** variable consisting of a tag fullID

**recvDepName** variable consisting of receiver deployment name

**ts** variable for a date/time object as numeric or POSIXct, defaults to ts

### C.15.3 Example

Create tag summaries for only select tags with time in default hours with data.frame df.alltags

```
tag_site_summary <- tagSumSite(filter(df.alltags, motusTagID %in%
  c(16047, 16037, 16039)))
head(tag_site_summary)

##                                     fullID          recvDeployName
## 1 SampleData#179:6.1@166.38(M.16037)      BULL\n37.1, -76
## 2 SampleData#179:6.1@166.38(M.16037) Comeau (Marshalltown)\n44.6, -65.8
## 3 SampleData#179:6.1@166.38(M.16037)      Hillman_Marsh\n42, -82.5
## 4 SampleData#179:6.1@166.38(M.16037)      Netitishi\n51.3, -80.1
## 5 SampleData#179:6.1@166.38(M.16037)      North Bluff\n51.5, -80.5
## 6 SampleData#179:6.1@166.38(M.16037)      NP mobile\nNA, NA
##           first_ts        last_ts      tot_ts num_det
## 1 2015-09-14 15:55:48 2015-09-14 15:56:49 0.01693000 hours     2
## 2 2015-09-08 18:29:57 2015-09-08 18:37:16 0.12192881 hours     6
## 3 2015-11-02 13:20:47 2015-11-02 13:21:42 0.01525056 hours     2
## 4 2015-08-30 01:36:08 2015-09-08 01:10:13 215.56819269 hours 1166
## 5 2015-08-23 15:13:57 2015-08-28 16:40:18 121.43902892 hours     26
## 6 2015-08-17 17:01:38 2015-08-17 17:02:39 0.01693344 hours     11
```

## C.16 timeToSunriset

### C.16.1 Description

Creates and adds variables for time to, and time from sunrise/sunset based on a variable of POSIXct dates/times data.frame must contain latitude, longitude, and a date/time variable

### C.16.2 Dependencies

**data** a selected table from .motus data, e.g. “alltags”, or a data.frame of detection data including at a minimum variables for date/time, latitude, and longitude

**lat** variable with latitude values, defaults to recvDeployLat

**lon** variable with longitude values, defaults to recvDeployLon

**ts** variable for a date/time object as numeric or POSIXct, defaults to ts

**units** units to display time difference, defaults to “hours”, options include “secs”, “mins”, “hours”, “days”, “weeks”

### C.16.3 Example

Get sunrise and sunset information with units in minutes using tbl file tbl.alltags

```
sunrise <- timeToSunriset(tbl.alltags, units = "mins")
head(sunrise)
```

```
##      hitID runID batchID          ts sig sigsd noise freq freqsd
## 1  45107  8886      53 2015-10-26 11:19:49  52     0   -96     4      0
## 2  45108  8886      53 2015-10-26 11:20:28  54     0   -96     4      0
## 3  45109  8886      53 2015-10-26 11:21:17  55     0   -96     4      0
## 4  45110  8886      53 2015-10-26 11:21:55  52     0   -96     4      0
```

```

## 5 45111 8886      53 2015-10-26 11:22:44 49      0   -96    4    0
## 6 199885 23305     64 2015-10-26 11:12:04 33      0   -96    4    0
##   slop burstSlop done motusTagID ambigID port runLen bootnum tagProjID
## 1 1e-04   0.0000   1   16047    NA   3     5    11    176
## 2 1e-04   -0.0021   1   16047    NA   3     5    11    176
## 3 1e-04    0.0001   1   16047    NA   3     5    11    176
## 4 1e-04   -0.0010   1   16047    NA   3     5    11    176
## 5 1e-04    0.0001   1   16047    NA   3     5    11    176
## 6 1e-04   0.0000   1   16047    NA   1    11     4    176
##   mfgID tagType codeSet   mfg tagModel tagLifespan nomFreq tagBI pulseLen
## 1   378      ID Lotek4 Lotek NTQB-3-2        NA 166.38 9.6971    2.5
## 2   378      ID Lotek4 Lotek NTQB-3-2        NA 166.38 9.6971    2.5
## 3   378      ID Lotek4 Lotek NTQB-3-2        NA 166.38 9.6971    2.5
## 4   378      ID Lotek4 Lotek NTQB-3-2        NA 166.38 9.6971    2.5
## 5   378      ID Lotek4 Lotek NTQB-3-2        NA 166.38 9.6971    2.5
## 6   378      ID Lotek4 Lotek NTQB-3-2        NA 166.38 9.6971    2.5
##   tagDeployID speciesID markerNumber markerType tagDeployStart
## 1     1839     4670 135268103 metal band 1441908000
## 2     1839     4670 135268103 metal band 1441908000
## 3     1839     4670 135268103 metal band 1441908000
## 4     1839     4670 135268103 metal band 1441908000
## 5     1839     4670 135268103 metal band 1441908000
## 6     1839     4670 135268103 metal band 1441908000
##   tagDeployEnd tagDeployLat tagDeployLon tagDeployAlt
## 1   1457632800    51.4839   -80.45    NA
## 2   1457632800    51.4839   -80.45    NA
## 3   1457632800    51.4839   -80.45    NA
## 4   1457632800    51.4839   -80.45    NA
## 5   1457632800    51.4839   -80.45    NA
## 6   1457632800    51.4839   -80.45    NA
##
## 1 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fatScore": 3, "locationID": "M.16047", "recvDeployLat": 42.60699}
## 2 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fatScore": 3, "locationID": "M.16047", "recvDeployLat": 42.60699}
## 3 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fatScore": 3, "locationID": "M.16047", "recvDeployLat": 42.60699}
## 4 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fatScore": 3, "locationID": "M.16047", "recvDeployLat": 42.60699}
## 5 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fatScore": 3, "locationID": "M.16047", "recvDeployLat": 42.60699}
## 6 {"ageID": "HY", "bill": 36.5, "blood": "Y", "country": "Canada", "culmen": 36.5, "fatScore": 3, "locationID": "M.16047", "recvDeployLat": 42.60699}
##   fullID deviceID recvDeployID recvDeployLat
## 1 SampleData#378:9.7@166.38(M.16047)     486     2510 42.60699
## 2 SampleData#378:9.7@166.38(M.16047)     486     2510 42.60699
## 3 SampleData#378:9.7@166.38(M.16047)     486     2510 42.60699
## 4 SampleData#378:9.7@166.38(M.16047)     486     2510 42.60699
## 5 SampleData#378:9.7@166.38(M.16047)     486     2510 42.60699
## 6 SampleData#378:9.7@166.38(M.16047)     515     2512 42.68067
##   recvDeployLon recvDeployAlt      recv recvDeployName recvSiteName
## 1   -72.71657          NA Lotek-159  Shelburne <NA>
## 2   -72.71657          NA Lotek-159  Shelburne <NA>
## 3   -72.71657          NA Lotek-159  Shelburne <NA>
## 4   -72.71657          NA Lotek-159  Shelburne <NA>
## 5   -72.71657          NA Lotek-159  Shelburne <NA>
## 6   -72.47392          NA Lotek-164 BennettMeadow <NA>
##   isRecvMobile recvProjID antType antBearing antHeight speciesEN
## 1         0       74  yagi-9      127    NA Red Knot
## 2         0       74  yagi-9      127    NA Red Knot

```

```

## 3      0    74 yagi-9      127      NA Red Knot
## 4      0    74 yagi-9      127      NA Red Knot
## 5      0    74 yagi-9      127      NA Red Knot
## 6      0    74 yagi-9     243      NA Red Knot
##   speciesFR   speciesSci speciesGroup tagProjName
## 1 Bécasseau maubèche Calidris canutus      BIRDS SampleData
## 2 Bécasseau maubèche Calidris canutus      BIRDS SampleData
## 3 Bécasseau maubèche Calidris canutus      BIRDS SampleData
## 4 Bécasseau maubèche Calidris canutus      BIRDS SampleData
## 5 Bécasseau maubèche Calidris canutus      BIRDS SampleData
## 6 Bécasseau maubèche Calidris canutus      BIRDS SampleData
##   recvProjName gpsLat gpsLon gpsAlt       sunrise
## 1 <NA>      NA      NA      NA 2015-10-26 11:16:49
## 2 <NA>      NA      NA      NA 2015-10-26 11:16:49
## 3 <NA>      NA      NA      NA 2015-10-26 11:16:49
## 4 <NA>      NA      NA      NA 2015-10-26 11:16:49
## 5 <NA>      NA      NA      NA 2015-10-26 11:16:49
## 6 <NA>      NA      NA      NA 2015-10-26 11:15:58
##   sunset ts_to_set ts_since_set ts_to_rise ts_since_rise
## 1 2015-10-26 21:52:11  632.3533    806.2056 1438.215974    3.013824
## 2 2015-10-26 21:52:11  631.7104    806.8485 1437.573097    3.656701
## 3 2015-10-26 21:52:11  630.9109    807.6480 1436.773577    4.456221
## 4 2015-10-26 21:52:11  630.2513    808.3076 1436.113997    5.115801
## 5 2015-10-26 21:52:11  629.4518    809.1071 1435.314477    5.915321
## 6 2015-10-26 21:51:06  639.0310    799.5242   3.896393 1437.332308

```

## Appendix D

# Appendix - The motusClient Package - Data filtering functions

The motusClient R package offers functions that can be used to assign probabilities to tag detections, and to filter detections based on those probabilities. For example, as you work through your data to clean false positive and ambiguous detections (see Chapter 5), you may determine that some detections do not belong to your tag(s). Instead of simply using an R script to filter out those detections, you can use these filter functions to create and save a custom filter in your .motus file, which assigns a probability value between 0 and 1 to the runIDs supplied in the filter.

The data filtering functions in the R package work at the level of a run. A run is a group of consecutive detections of a tag detected on a receiver. In general, a detection with a run length of 2 has a high probability of being a false positive detection. The probabilities associated with each runID can be generated in a number of possible ways, including at the simplest level, generating a list of 0's and 1's for records that you would like to exclude or include. Alternatively, you might develop a model that assigns a probability to each runID in your data.

### D.1 listRunsFilters

#### D.1.1 Description

Returns a dataframe containing the filterIDs, logins, names, projectIDs and descriptions for a given tag or receiver projectID available in the local database.

### D.2 Dependencies

`src` is the SQLite object that you get from loading a .motus file into R, e.g., ‘sql.motus’ file in Chapter 3.

### D.3 Example

```
filt.df <- listRunsFilters(src = sql.motus)
```

## D.4 createRunsFilter

### D.4.1 Description

This function can be used mostly by users to modify properties of existing filters (e.g., filter description or projectID), but it is also being called internally by ‘writeRunsFilter’ (section D.6) to generate a new filterID. To save the actual filter records, you must use ‘writeRunsFilter’ (section D.6). The function returns the filterID (integer) in the local database that matches the new or existing filter with the provided filterName. If a filter with the same name already exists, the function generates a warning and returns the ID of the existing filter.

### D.4.2 Dependencies

**src** is the SQLite object that you get from loading a .motus file into R, e.g., ‘sql.motus’ file in Chapter 3. **filterName** the name you would like to assign to the filter. The function only creates a new filter if the name does not already exist locally.

**motusProjID** the numeric ID associated with a project, e.g., 176 for the sample data used throughout this book. The function defaults to motusProjID = ‘NA’ when project ID is not supplied, which is the recommended value for now. The project ID assigned to a filter will mostly be useful for future synchronization of filters with the Motus server. The detection records contained in the filter do not have to be assigned to the projectID assigned to the filter. **descr** default ‘NA’. Optional description of the filter. **update** boolean (default = FALSE). If the filter already exists, determines if the properties (e.g. descr are preserved or updated)

### D.4.3 Example

Create a new filter called “myfilter” for the sql.motus database which is not attached to a specific project:

```
createRunsFilter(sql.motus, "myfilter")

# OR add assignment to project

createRunsFilter(sql.motus, "myfilter", motusProjID = 176)

# OR add project and description, possibly updating
# any previous version called myfilter.

createRunsFilter(sql.motus, "myfilter", motusProjID = 176,
  descr = "assign probability of 0 to false positives",
  update = TRUE)
```

## D.5 getRunsFilters

### D.5.1 Description

Returns a sqlite table reference to the runsFilters records saved in the database (runID, motusTagID, and probability) associated with a specific name (and optionally project) from the local database. For examples on how you can use the returned table to merge with your detection data, refer to section ?? in chapter 5.

### D.5.2 Dependencies

**src** is the SQLite object that you get from loading a .motus file into R, e.g., ‘sql.motus’ file in Chapter 3.  
**filterName** the name you used when you created or saved your filter. Function returns a warning if the filterName doesn’t exist.

**motusProjID** the numeric ID associated with a project, e.g., 176 for the sample data used throughout this book. The function defaults to motusProjID = ‘NA’ when project ID is not supplied.

### D.5.3 Example

```
tbl.filt <- getRunsFilters(src = sql.motus, filterName = "myfilter")
tbl.filt2 <- getRunsFilters(sql.motus, "myfilter2")

# filter records from df that are in tbl.filt
df <- left_join(df, tbl.filt, by = c("runID", "motusTagID")) %>%
  mutate(probability = ifelse(is.na(probability),
    1, probability)) %>% filter(probability > 0)

# you can apply a second filter, tbl.filt2, to the
# result of the previous filter
df <- left_join(df, tbl.filt2, by = c("runID", "motusTagID")) %>%
  mutate(probability = ifelse(is.na(probability),
    1, probability)) %>% filter(probability > 0)
```

## D.6 writeRunsFilter

### D.6.1 Description

Writes to the local database (SQLite file) the content of a dataframe containing runID, motusTagID, and assigned probability. If the filterName provided does not exist, the function will call ‘createRunsFilter’ (section D.4) to create one in your database. The default behaviour of the function is that any new records from the dataframe are appended to the existing or new filter called filterName, those that already are present (same runID and motusTagID) are replaced (overwrite=TRUE), but those that are not included in the dataframe are retained in the existing filter table (delete=FALSE). To entirely replace the existing filter values with those of the new dataframe, use delete=TRUE. The function returns a sqlite table reference to the filter, similarly to ‘getRunsFilter’ (section D.5).

### D.6.2 Dependencies

**src** is the SQLite object that you get from loading a .motus file into R, e.g., ‘sql.motus’ file in Chapter 3.  
**filterName** the name of the filter you would like to assign the database to.

**motusProjID** the numeric ID associated with a project, e.g., 176 for the sample data used throughout this book. Default = ‘NA’ when project ID is not supplied.

**df** dataframe which contains the runID (integer), motusTagID (integer), and probability (float) of detections you would like to assign a filter to. MotusTagID should be the actual tag ID, and not the negative ambigID associated with ambiguous detections. **overwrite** Default = “TRUE”. When TRUE, ensures that existing records (same runID and motusTagID) matching the same filterName and runID get replaced in the local database. **delete** Default = “FALSE”. When TRUE, removes all existing filter records associated with the filterName and re-inserts the ones contained in df. This option should be used if df contains the entire set of filters you want to save.

### D.6.3 Examples

```
# write a dataframe containing filter records
# (runID, motusTagID and probability) to "myfilter"
writeRunsFilter(src = sql.motus, filterName = "myfilter",
                 df = filter.df)

# write a dataframe containing filter records
# (runID, motusTagID and probability) to
# "myfilter", overwriting a previous version
# entirely
writeRunsFilter(src = sql.motus, fileName = "myfilter",
                 df = filter.df, delete = TRUE)

# write a dataframe containing filter records
# (runID, motusTagID and probability) to
# 'myfilter', but only append new records, leaving
# previously created ones intact
writeRunsFilter(src = sql.motus, "myfilter", df = filter.df,
                 overwrite = FALSE)
```