review

Azul Carrillo

2022-11-29

#1 Write a conditional statement that checks if surveys.csv exists in the working directory, if it doesn't then downloads it from https://ndownloader.figshare.com/files/2292172 using download.file(), and finally loads the file into a data frame and displays the first few rows using the head() function. The url needs to be in quotes since it is character data.

```
getwd()
```

```
## [1] "/Users/atziri/Bio 195-197/Data Science/documents"
```

```
"surveys.csv" == c(list.files("/Users/atziri/Bio 195-197/Data Science/raw-data"))
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25]  TRUE FALSE FALSE
```

```
surveys <- list.files("/Users/atziri/Bio 195-197/Data Science/raw-data") == "surveys.csv"
```

```
is.element("surveys.csv", list.files("/Users/atziri/Bio 195-197/Data Science/raw-data") )
```

```
## [1] TRUE
```

```
is.element("surveys.csv", list.files("/Users/atziri/Bio 195-197/Data Science/raw-data") )
```

```
## [1] TRUE
```

```
surveys <- list.files("../raw-data")

if (is.element("surveys.csv", surveys)){
  print("file is downloaded")
} else {
   print("file is not downloaded")
  download.file("https://ndownloader.figshare.com/files/2292172",
               "../raw-data/surveys-download.csv")
surveys_data <- read.csv("../raw-data/surveys-download.csv")
 head(surveys_data)
}
```

```
## [1] "file is downloaded"
```

#2 Make a version of this conditional statement that is a function, where the name of the file is the first argument and the link for downloading the file is the second argumen

```r
#This function tests if a file is in the raw-data directory and if not it download it and read it as da

reading_csv <- function(file_name, file_link) {

# 1. test if file_name is in the raw-data folder
# file_name <- "species.csv"

test <- !is.element(file_name, list.files(path = "../raw-data"))

# 2. if test is FALSE, download the file

if (test) {
  # Option 1: save it with a random name:
  # download.file(url = file_link, destfile = "../raw-data/temporary.csv")

  # result <- read.csv(file = "../raw-data/temporary.csv")

# Option 2: save it with the name given in file nome:
  destination_file <- stringr::str_c("../raw-data/", file_name)
  download.file(url = file_link, destfile = destination_file)
  result <- read.csv(file = destination_file)
}
  return(result)
}
```

```r
reading_csv <- function(file_name, file_link) {
  test <- is.element(file_name, list.files(path = "../raw-data"))# removed the exclamation mark for it
  if (test) {
    destination_file <- stringr:: str_c("../raw-data/", file_name)
    download.file(url = file_link, destfile = destination_file)

    result <- read.csv(file = destination_file)
  }
   return(result)
}
```

```r
reading_csv(file_name = "species.csv",
            file_link = "https://ndownloade.figshare.com/files/3299483")
```

```
##    species_id             genus            species    taxa
## 1          AB        Amphispiza          bilineata    Bird
## 2          AH Ammospermophilus            harrisi  Rodent
## 3          AS        Ammodramus         savannarum    Bird
## 4          BA           Baiomys            taylori  Rodent
## 5          CB   Campylorhynchus  brunneicapillus    Bird
## 6          CM       Calamospiza        melanocorys    Bird
## 7          CQ        Callipepla           squamata    Bird
## 8          CS          Crotalus         scutalatus Reptile
## 9          CT     Cnemidophorus             tigris Reptile
## 10         CU     Cnemidophorus          uniparens Reptile
```

```
## 11    CV     Crotalus        viridis Reptile
## 12    DM    Dipodomys       merriami  Rodent
## 13    DO    Dipodomys         ordii  Rodent
## 14    DS    Dipodomys    spectabilis  Rodent
## 15    DX    Dipodomys           sp.   Rodent
## 16    EO     Eumeces       obsoletus Reptile
## 17    GS     Gambelia          silus Reptile
## 18    NL     Neotoma        albigula  Rodent
## 19    NX     Neotoma            sp.   Rodent
## 20    OL    Onychomys    leucogaster  Rodent
## 21    OT    Onychomys       torridus  Rodent
## 22    OX    Onychomys           sp.   Rodent
## 23    PB    Chaetodipus      baileyi  Rodent
## 24    PC       Pipilo      chlorurus    Bird
## 25    PE   Peromyscus       eremicus  Rodent
## 26    PF   Perognathus        flavus  Rodent
## 27    PG    Pooecetes       gramineus    Bird
## 28    PH   Perognathus      hispidus  Rodent
## 29    PI    Chaetodipus   intermedius  Rodent
## 30    PL   Peromyscus       leucopus  Rodent
## 31    PM   Peromyscus     maniculatus  Rodent
## 32    PP    Chaetodipus   penicillatus  Rodent
## 33    PU       Pipilo        fuscus    Bird
## 34    PX    Chaetodipus          sp.   Rodent
## 35    RF Reithrodontomys   fulvescens  Rodent
## 36    RM Reithrodontomys   megalotis  Rodent
## 37    RO Reithrodontomys    montanus  Rodent
## 38    RX Reithrodontomys         sp.   Rodent
## 39    SA    Sylvilagus      audubonii  Rabbit
## 40    SB     Spizella        breweri    Bird
## 41    SC    Sceloporus         clarki Reptile
## 42    SF     Sigmodon    fulviventer  Rodent
## 43    SH     Sigmodon       hispidus  Rodent
## 44    SO     Sigmodon    ochrognathus  Rodent
## 45    SS  Spermophilus     spilosoma  Rodent
## 46    ST  Spermophilus  tereticaudus  Rodent
## 47    SU    Sceloporus     undulatus Reptile
## 48    SX     Sigmodon           sp.   Rodent
## 49    UL       Lizard           sp. Reptile
## 50    UP       Pipilo           sp.    Bird
## 51    UR       Rodent           sp.   Rodent
## 52    US      Sparrow           sp.    Bird
## 53    ZL  Zonotrichia    leucophrys    Bird
## 54    ZM      Zenaida       macroura    Bird
```

#Exercise 2: Multi-file Analysis

##1. If individual_collar_data.zip is not already in your working directory download the zip file using download.file()

##2. Unzip it using unzip()

##3. Obtain a list of all of the files with file names matching the pattern "collar-data-.*.txt" (using list.files())

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
download.file("http://www.datacarpentry.org/semester-biology/data/individual_collar_data.zip", "collar.:
unzip("collar.zip", exdir = "../raw-data")
collar_data_files <- list.files(pattern = "collar-")
collar_data_files
```

```
##  [1] "collar-data-A1-2016-02-26.txt"  "collar-data-B2-2016-02-26.txt"
##  [3] "collar-data-C3-2016-02-26.txt"  "collar-data-D4-2016-02-26.txt"
##  [5] "collar-data-E5-2016-02-26.txt"  "collar-data-F6-2016-02-26.txt"
##  [7] "collar-data-G7-2016-02-26.txt"  "collar-data-H8-2016-02-26.txt"
##  [9] "collar-data-I9-2016-02-26.txt"  "collar-data-J10-2016-02-26.txt"
```

##4. Use a loop to load each of these files into R and make a line plot (using geom_path()) for each file with long on the x axis and lat on the y axis. Graphs, like other types of output, won't display inside a loop unless you explicitly display them, so you need put your ggplot() command inside a print() statement. Include the name of the file in the graph as the graph title using labs().

##5. Add code to the loop to calculate the minimum and maximum latitude in the file, and store these values, along with the name of the file, in a data frame. Show the data frame as output.

```
min_results <- vector(mode = "integer", length(collar_data_files))
max_results <- vector(mode = "integer", length(collar_data_files))


for(i in 1:length(collar_data_files)){
  filename <- collar_data_files[i]
  data <- read.csv(filename)
  data %>%
    ggplot(aes(x = long, y = lat )) +
    labs(title = filename) +
  geom_path() -> plots
    print(plots)
   min_results[i] <- min(data$lat)
   max_results[i] <- max(data$lat)
}
```
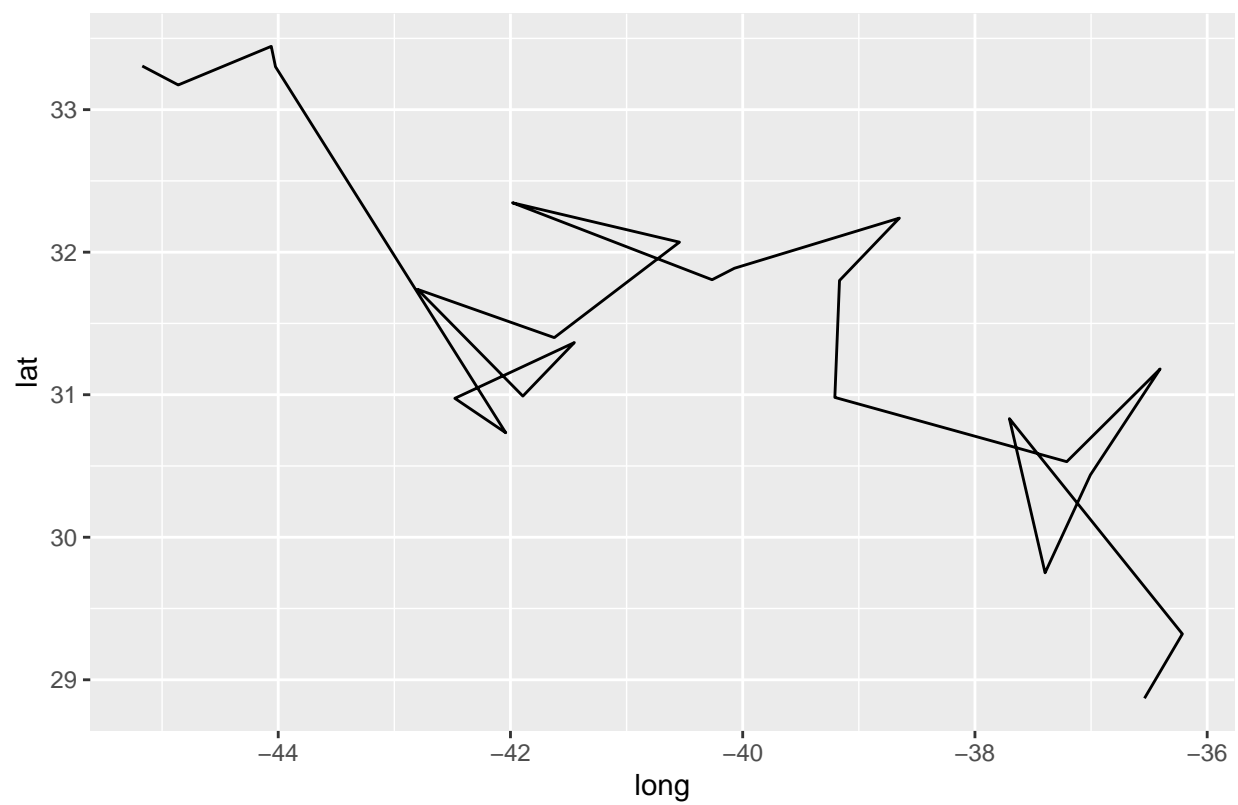
collar−data−A1−2016−02−26.txt

collar–data–B2–2016–02–26.txt

collar−data−C3−2016−02−26.txt
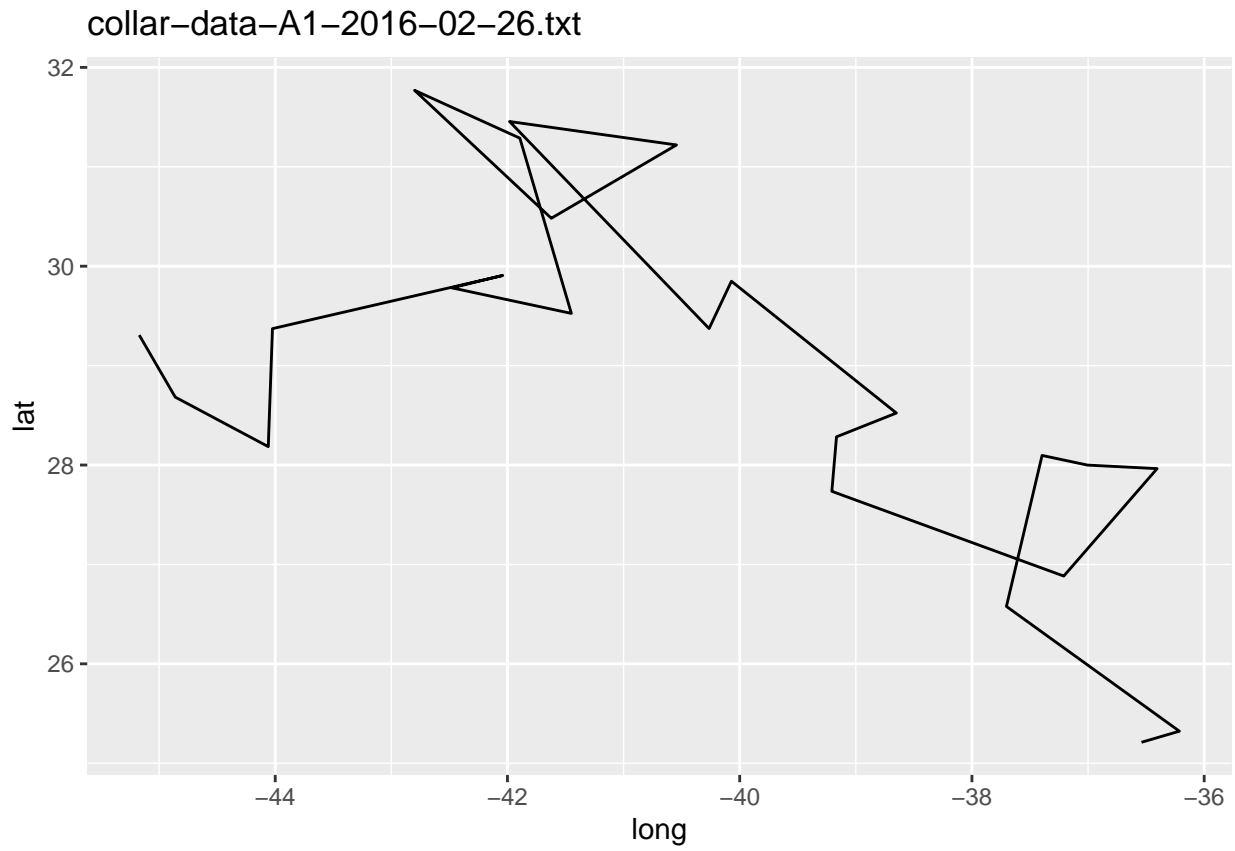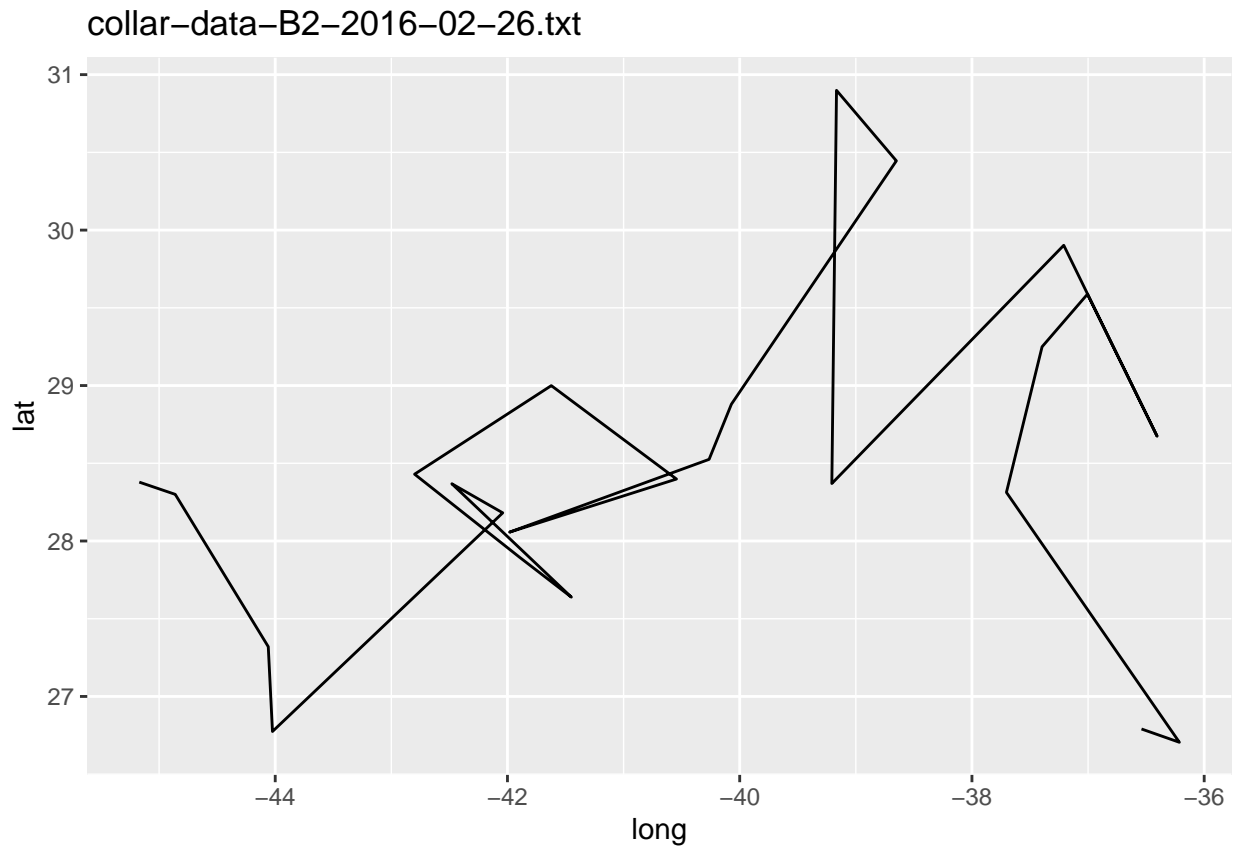
collar–data–D4–2016–02–26.txt

collar–data–E5–2016–02–26.txt
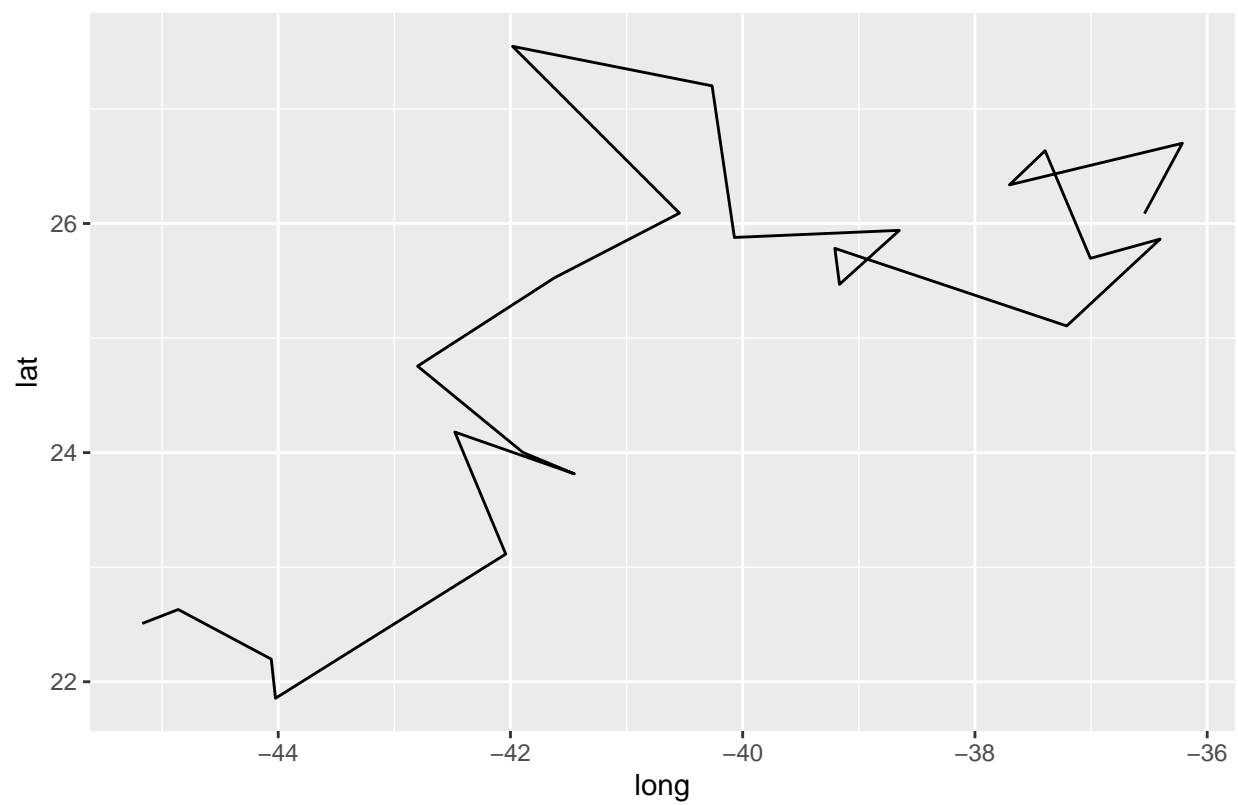
collar–data–F6–2016–02–26.txt
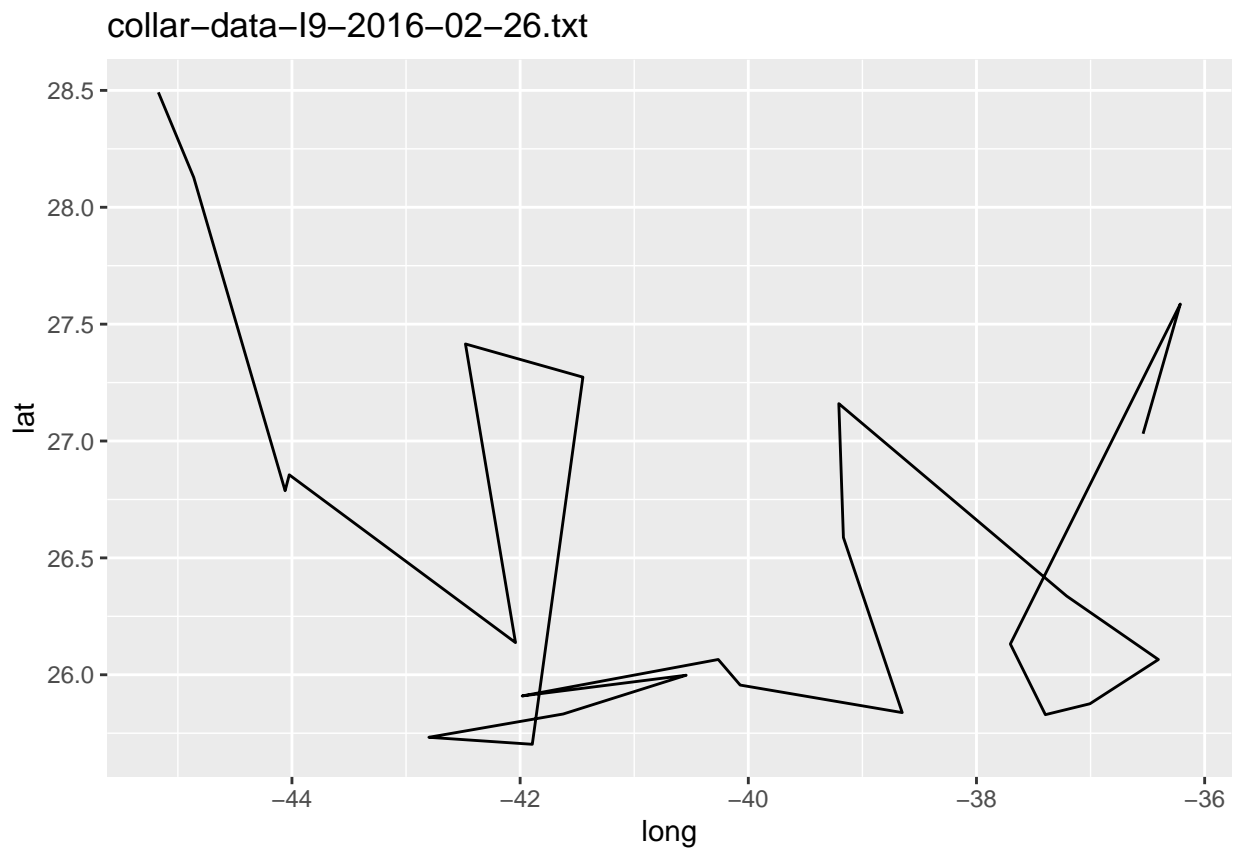
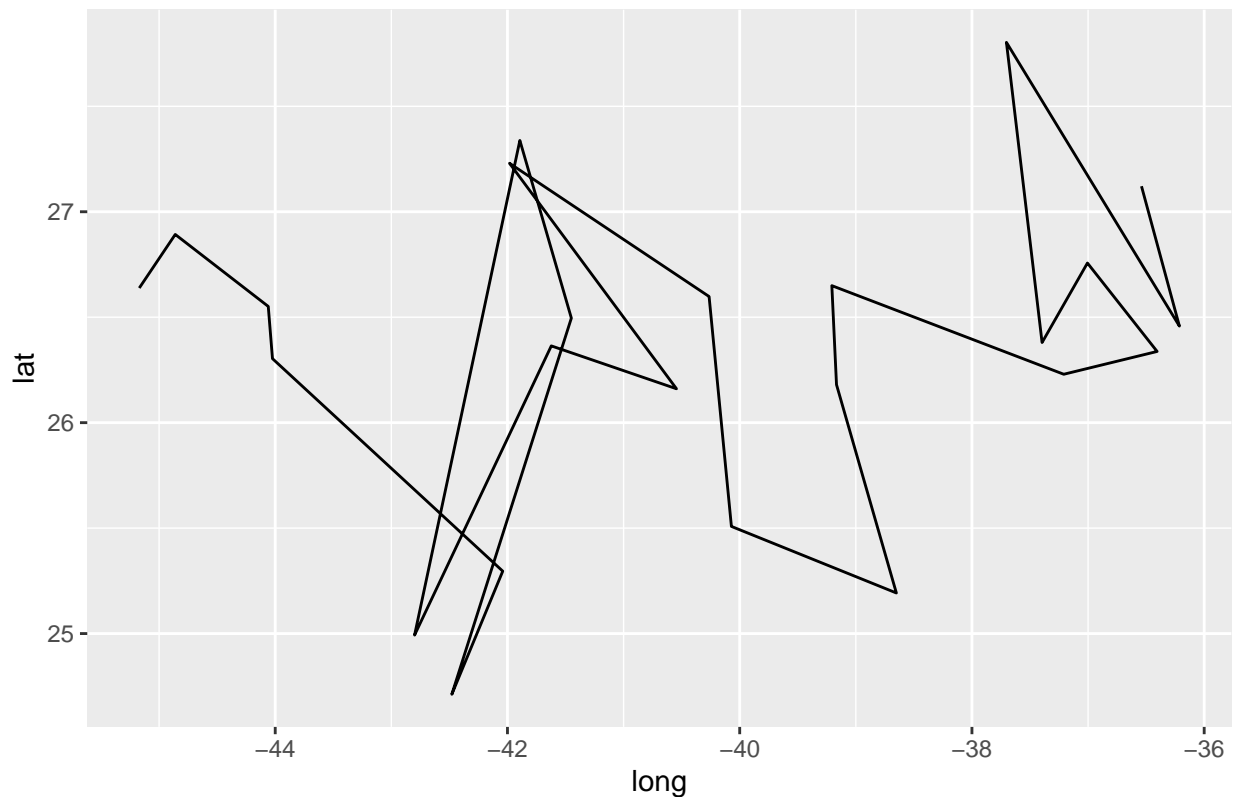collar−data−G7−2016−02−26.txt
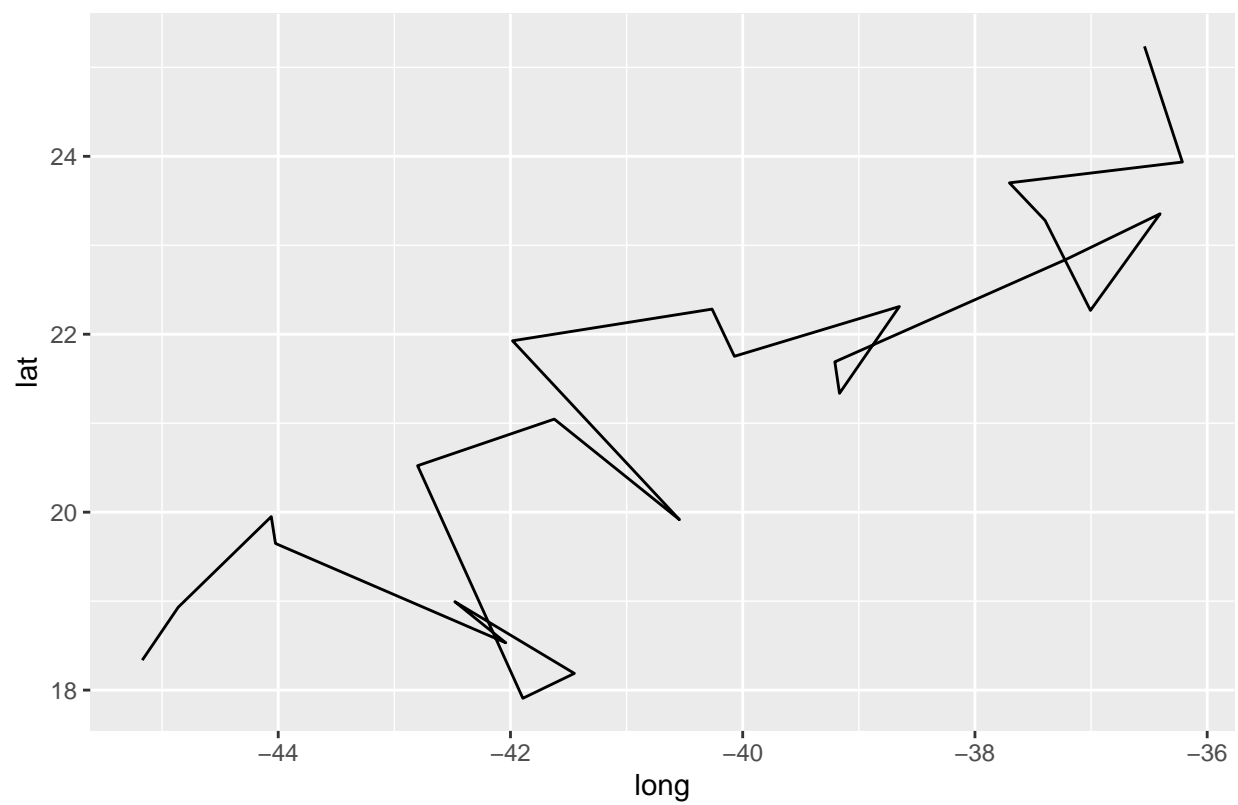
collar−data−H8−2016−02−26.txt
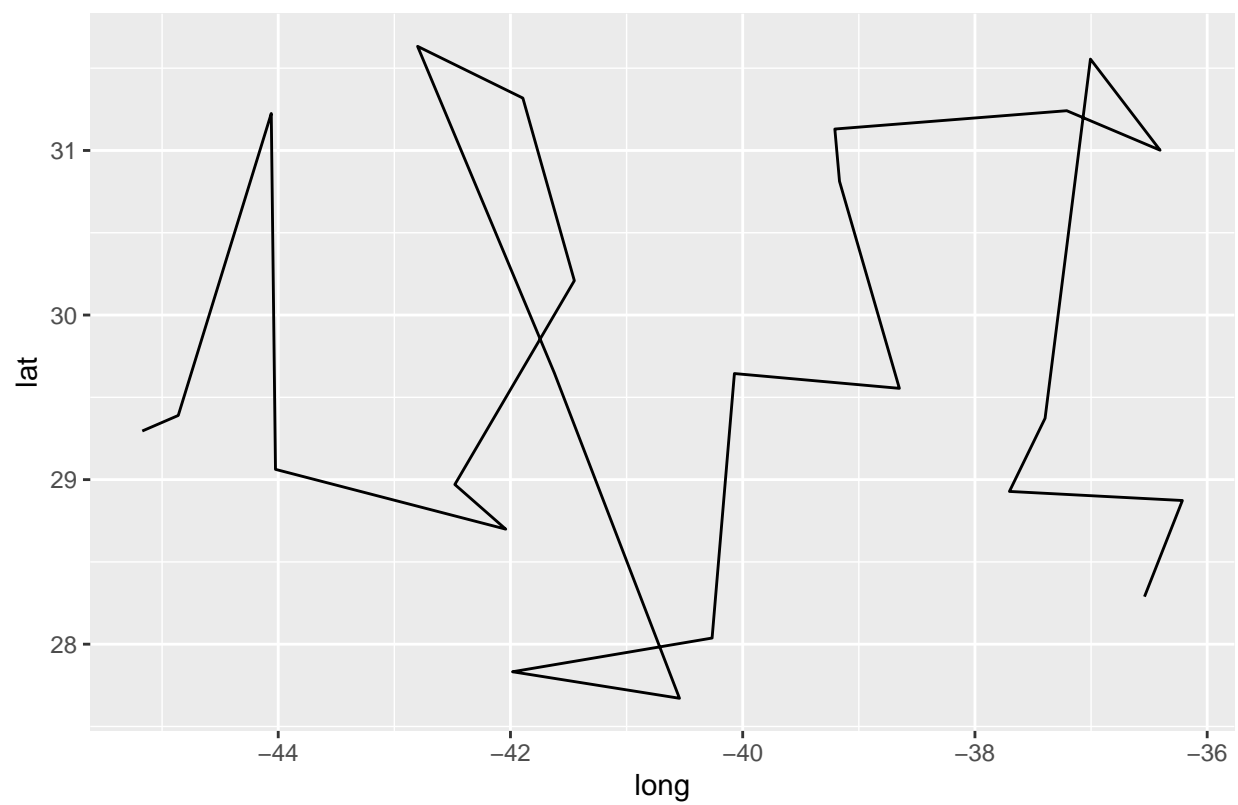
collar-data-I9-2016-02-26.txt
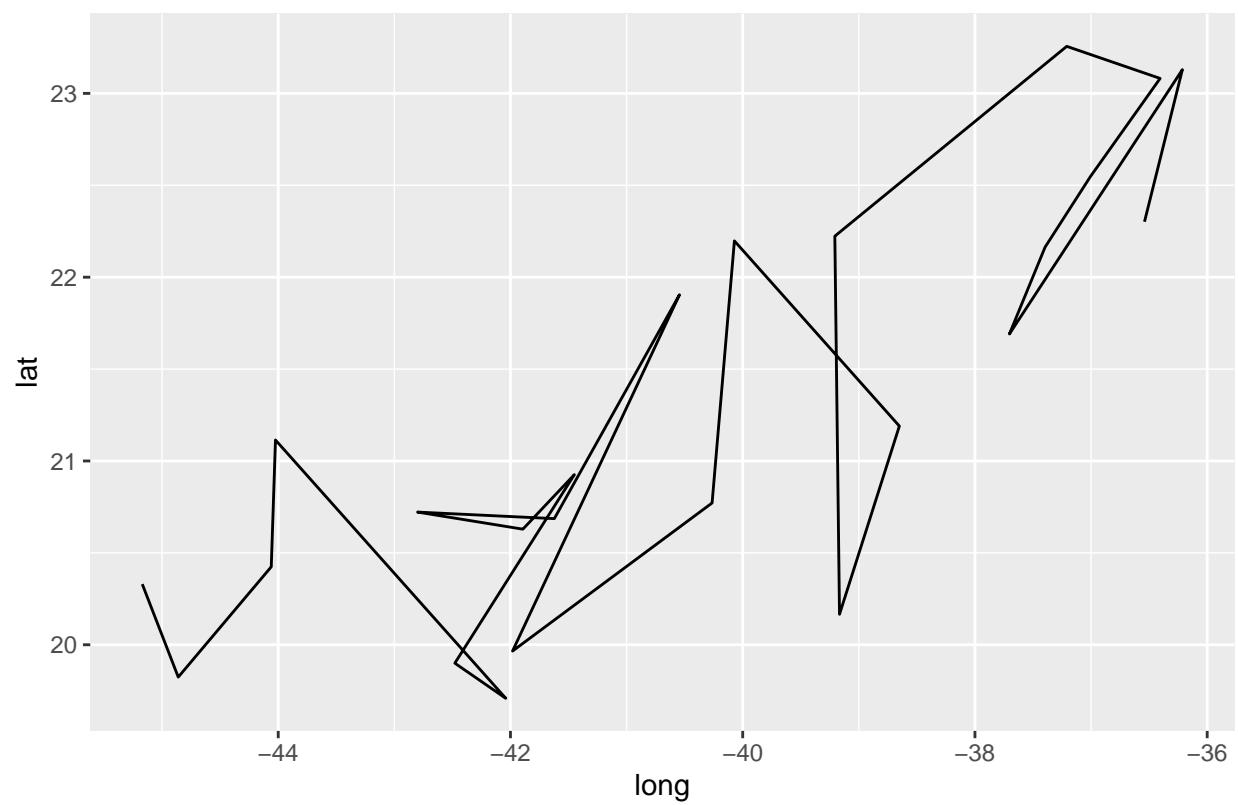
collar–data–J10–2016–02–26.txt


```
min_max_data <- data.frame(collar_data_files, min_results, max_results)

min_max_data
```

```
##               collar_data_files min_results max_results
## 1   collar-data-A1-2016-02-26.txt    25.21080    31.76912
## 2   collar-data-B2-2016-02-26.txt    26.70509    30.89907
## 3   collar-data-C3-2016-02-26.txt    28.86998    33.44421
## 4   collar-data-D4-2016-02-26.txt    21.34315    24.66598
## 5   collar-data-E5-2016-02-26.txt    21.85565    27.54663
## 6   collar-data-F6-2016-02-26.txt    17.90788    25.23623
## 7   collar-data-G7-2016-02-26.txt    27.67120    31.63272
## 8   collar-data-H8-2016-02-26.txt    19.70875    23.25601
## 9   collar-data-I9-2016-02-26.txt    25.70252    28.49172
## 10 collar-data-J10-2016-02-26.txt    24.71200    27.80325
```

##Solution 2

```
all_min <-  vector()
all_max <- vector()
length(all_min)
```

```
## [1] 0
```

14

```r
file_name <- vector()

for(i in collar_data_files){
  print(i)
  print(getwd())
  file_name <- c(file_name,i)
  collar_data_table <- read.csv(file = i )
  min_lat <- min (collar_data_table[,"lat"])
  all_min <- c(all_min,min_lat)
  max_lat <- max(collar_data_table[,"lat"])
  all_max <- c(all_max, max_lat)
  print(all_min)
  print(all_max)
}
```

```
## [1] "collar-data-A1-2016-02-26.txt"
## [1] "/Users/atziri/Bio 195-197/Data Science/documents"
## [1] 25.2108
## [1] 31.76912
## [1] "collar-data-B2-2016-02-26.txt"
## [1] "/Users/atziri/Bio 195-197/Data Science/documents"
## [1] 25.21080 26.70509
## [1] 31.76912 30.89907
## [1] "collar-data-C3-2016-02-26.txt"
## [1] "/Users/atziri/Bio 195-197/Data Science/documents"
## [1] 25.21080 26.70509 28.86998
## [1] 31.76912 30.89907 33.44421
## [1] "collar-data-D4-2016-02-26.txt"
## [1] "/Users/atziri/Bio 195-197/Data Science/documents"
## [1] 25.21080 26.70509 28.86998 21.34315
## [1] 31.76912 30.89907 33.44421 24.66598
## [1] "collar-data-E5-2016-02-26.txt"
## [1] "/Users/atziri/Bio 195-197/Data Science/documents"
## [1] 25.21080 26.70509 28.86998 21.34315 21.85565
## [1] 31.76912 30.89907 33.44421 24.66598 27.54663
## [1] "collar-data-F6-2016-02-26.txt"
## [1] "/Users/atziri/Bio 195-197/Data Science/documents"
## [1] 25.21080 26.70509 28.86998 21.34315 21.85565 17.90788
## [1] 31.76912 30.89907 33.44421 24.66598 27.54663 25.23623
## [1] "collar-data-G7-2016-02-26.txt"
## [1] "/Users/atziri/Bio 195-197/Data Science/documents"
## [1] 25.21080 26.70509 28.86998 21.34315 21.85565 17.90788 27.67120
## [1] 31.76912 30.89907 33.44421 24.66598 27.54663 25.23623 31.63272
## [1] "collar-data-H8-2016-02-26.txt"
## [1] "/Users/atziri/Bio 195-197/Data Science/documents"
## [1] 25.21080 26.70509 28.86998 21.34315 21.85565 17.90788 27.67120 19.70875
## [1] 31.76912 30.89907 33.44421 24.66598 27.54663 25.23623 31.63272 23.25601
## [1] "collar-data-I9-2016-02-26.txt"
## [1] "/Users/atziri/Bio 195-197/Data Science/documents"
## [1] 25.21080 26.70509 28.86998 21.34315 21.85565 17.90788 27.67120 19.70875
## [9] 25.70252
## [1] 31.76912 30.89907 33.44421 24.66598 27.54663 25.23623 31.63272 23.25601
## [9] 28.49172
```

```
## [1] "collar-data-J10-2016-02-26.txt"
## [1] "/Users/atziri/Bio 195-197/Data Science/documents"
##  [1] 25.21080 26.70509 28.86998 21.34315 21.85565 17.90788 27.67120 19.70875
##  [9] 25.70252 24.71200
##  [1] 31.76912 30.89907 33.44421 24.66598 27.54663 25.23623 31.63272 23.25601
##  [9] 28.49172 27.80325
```

```r
min_max_dataframe <- data_frame(file_name, all_min, all_max)
```

```
## Warning: `data_frame()` was deprecated in tibble 1.1.0.
## i Please use `tibble()` instead.
```

```r
min_max_dataframe
```

```
## # A tibble: 10 x 3
##    file_name                   all_min all_max
##    <chr>                         <dbl>   <dbl>
##  1 collar-data-A1-2016-02-26.txt   25.2    31.8
##  2 collar-data-B2-2016-02-26.txt   26.7    30.9
##  3 collar-data-C3-2016-02-26.txt   28.9    33.4
##  4 collar-data-D4-2016-02-26.txt   21.3    24.7
##  5 collar-data-E5-2016-02-26.txt   21.9    27.5
##  6 collar-data-F6-2016-02-26.txt   17.9    25.2
##  7 collar-data-G7-2016-02-26.txt   27.7    31.6
##  8 collar-data-H8-2016-02-26.txt   19.7    23.3
##  9 collar-data-I9-2016-02-26.txt   25.7    28.5
## 10 collar-data-J10-2016-02-26.txt  24.7    27.8
```

```r
#solution 3
library(stringr)
all_min_lat <- vector(mode = "integer", length = length(collar_data_files))
all_max_lat <- all_min_lat
all_file_names <- all_min_lat
length(all_min)
```

```
## [1] 10
```

```r
length(all_max)
```

```
## [1] 10
```

```r
all_file_names
```

```
##  [1] 0 0 0 0 0 0 0 0 0 0
```

```r
for (i in 1:length(collar_data_files)){
  file_name_and_path <- str_c("../raw-data/",collar_data_files[i])
  all_file_names[i] <- file_name_and_path
  print(file_name_and_path)
  collar_data_table <- read.csv(file = file_name_and_path)
```

```
  min_lat <- min(collar_data_table$lat)
  all_min_lat[i] <- min_lat
  max_lat <- max(collar_data_table$lat)
  all_max_lat[i] <- max_lat
  print(all_min_lat)
  print(all_max_lat)
}
```

```
## [1] "../raw-data/collar-data-A1-2016-02-26.txt"
##  [1] 25.2108  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
## [10]  0.0000
##  [1] 31.76912  0.00000  0.00000  0.00000  0.00000  0.00000  0.00000
##  [9]  0.00000  0.00000
## [1] "../raw-data/collar-data-B2-2016-02-26.txt"
##  [1] 25.21080 26.70509  0.00000  0.00000  0.00000  0.00000  0.00000  0.00000
##  [9]  0.00000  0.00000
##  [1] 31.76912 30.89907  0.00000  0.00000  0.00000  0.00000  0.00000  0.00000
##  [9]  0.00000  0.00000
## [1] "../raw-data/collar-data-C3-2016-02-26.txt"
##  [1] 25.21080 26.70509 28.86998  0.00000  0.00000  0.00000  0.00000  0.00000
##  [9]  0.00000  0.00000
##  [1] 31.76912 30.89907 33.44421  0.00000  0.00000  0.00000  0.00000  0.00000
##  [9]  0.00000  0.00000
## [1] "../raw-data/collar-data-D4-2016-02-26.txt"
##  [1] 25.21080 26.70509 28.86998 21.34315  0.00000  0.00000  0.00000  0.00000
##  [9]  0.00000  0.00000
##  [1] 31.76912 30.89907 33.44421 24.66598  0.00000  0.00000  0.00000  0.00000
##  [9]  0.00000  0.00000
## [1] "../raw-data/collar-data-E5-2016-02-26.txt"
##  [1] 25.21080 26.70509 28.86998 21.34315 21.85565  0.00000  0.00000  0.00000
##  [9]  0.00000  0.00000
##  [1] 31.76912 30.89907 33.44421 24.66598 27.54663  0.00000  0.00000  0.00000
##  [9]  0.00000  0.00000
## [1] "../raw-data/collar-data-F6-2016-02-26.txt"
##  [1] 25.21080 26.70509 28.86998 21.34315 21.85565 17.90788  0.00000  0.00000
##  [9]  0.00000  0.00000
##  [1] 31.76912 30.89907 33.44421 24.66598 27.54663 25.23623  0.00000  0.00000
##  [9]  0.00000  0.00000
## [1] "../raw-data/collar-data-G7-2016-02-26.txt"
##  [1] 25.21080 26.70509 28.86998 21.34315 21.85565 17.90788 27.67120  0.00000
##  [9]  0.00000  0.00000
##  [1] 31.76912 30.89907 33.44421 24.66598 27.54663 25.23623 31.63272  0.00000
##  [9]  0.00000  0.00000
## [1] "../raw-data/collar-data-H8-2016-02-26.txt"
##  [1] 25.21080 26.70509 28.86998 21.34315 21.85565 17.90788 27.67120 19.70875
##  [9]  0.00000  0.00000
##  [1] 31.76912 30.89907 33.44421 24.66598 27.54663 25.23623 31.63272 23.25601
##  [9]  0.00000  0.00000
## [1] "../raw-data/collar-data-I9-2016-02-26.txt"
##  [1] 25.21080 26.70509 28.86998 21.34315 21.85565 17.90788 27.67120 19.70875
##  [9] 25.70252  0.00000
##  [1] 31.76912 30.89907 33.44421 24.66598 27.54663 25.23623 31.63272 23.25601
##  [9] 28.49172  0.00000
```

```
## [1] "../raw-data/collar-data-J10-2016-02-26.txt"
##  [1] 25.21080 26.70509 28.86998 21.34315 21.85565 17.90788 27.67120 19.70875
##  [9] 25.70252 24.71200
##  [1] 31.76912 30.89907 33.44421 24.66598 27.54663 25.23623 31.63272 23.25601
##  [9] 28.49172 27.80325
```

```r
min_max_dataframee <- data_frame(all_file_names, all_min_lat, all_max_lat)
min_max_dataframee
```

```
## # A tibble: 10 x 3
##    all_file_names                          all_min_lat all_max_lat
##    <chr>                                         <dbl>       <dbl>
##  1 ../raw-data/collar-data-A1-2016-02-26.txt      25.2        31.8
##  2 ../raw-data/collar-data-B2-2016-02-26.txt      26.7        30.9
##  3 ../raw-data/collar-data-C3-2016-02-26.txt      28.9        33.4
##  4 ../raw-data/collar-data-D4-2016-02-26.txt      21.3        24.7
##  5 ../raw-data/collar-data-E5-2016-02-26.txt      21.9        27.5
##  6 ../raw-data/collar-data-F6-2016-02-26.txt      17.9        25.2
##  7 ../raw-data/collar-data-G7-2016-02-26.txt      27.7        31.6
##  8 ../raw-data/collar-data-H8-2016-02-26.txt      19.7        23.3
##  9 ../raw-data/collar-data-I9-2016-02-26.txt      25.7        28.5
## 10 ../raw-data/collar-data-J10-2016-02-26.txt     24.7        27.8
```

#Exercise 3: A function for the UHURU data set

##1. Explain what each line of code in the body of the function is doing. Add the explanations to your Rmd file as comments, before each line of code.

```r
report_rsquared <- function(data, species, formula){
  subset <- dplyr::filter(data, ANT == species)
  test <- lm(formula, data = subset)
  rsquared <- round(summary(test)$r.squared, 3)
  output <- data.frame(species = species, r2 = rsquared)
  return(output)
}
```

##2. Execute the function using the UHURU data and specifying species = "CM" and formula = "AXIS1~CIRC".

```r
UHURU_data <- read.csv(file = "../raw-data/ACACIA_DREPANOLOBIUM_SURVEY.txt", sep = "\t")

report_rsquared(UHURU_data,"CM","AXIS1~CIRC")
```

```
##   species    r2
## 1      CM 0.866
```

##3. Modify the function so that it also determines if() the rsquared is significant based on a given threshold. The modified function should return() the species, rsquared and a significance value of "S" for a relationship with an rsquared > threshold or "NS" for an rsquared < threshold.

18

```
report_rsquared_modified <- function(data, species, formula){
  subset <- dplyr::filter(data, ANT == species)
  test <- lm(formula, data = subset)
  rsquared <- round(summary(test)$r.squared, 3)
  output <- data.frame(species = species, r2 = rsquared)
  return(output)
}
```

##4. Execute your modified function for species of "CM", "CS", and "TP" given a threshold = 0.667.

```
file_name <-  "individual_collar_data.zip"
#if (!file_present){
  #download.file(url = "http://www.datacarpentry.org/semester-biology/data/individual_collar_data.zip",
             #  destfile = "../raw-data/individual_collar_data.zip ")
#}
```