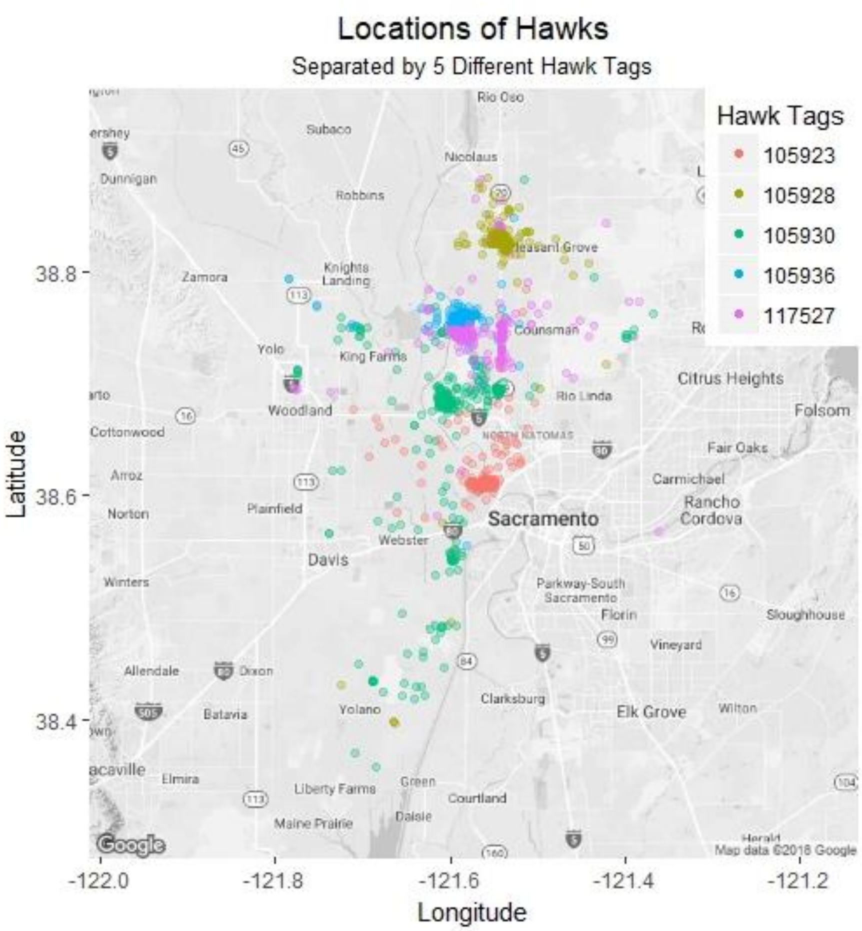
Jared Yu

914640019

STA 141A

Homework Assignment 4

1) The method for creating the map was to utilize the get_center() function that was covered in the TA's discussion. Some slight modification was done to take the mean of the range of the longitude and latitude rather than just the overall mean. This helped to give a more centered plot that included all the data. The map type that was utilized is from "google" and is called "terrain" in black and white. The purpose of this over a more natural map was that it had an uncolored background which made it simple to differentiate all the colored points from the rest of the data.

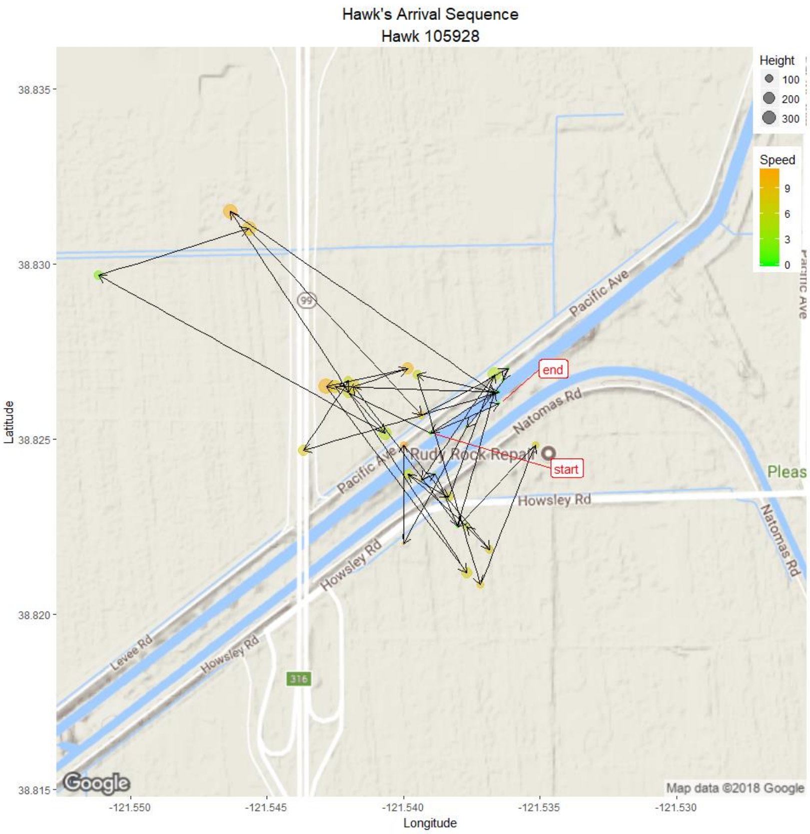


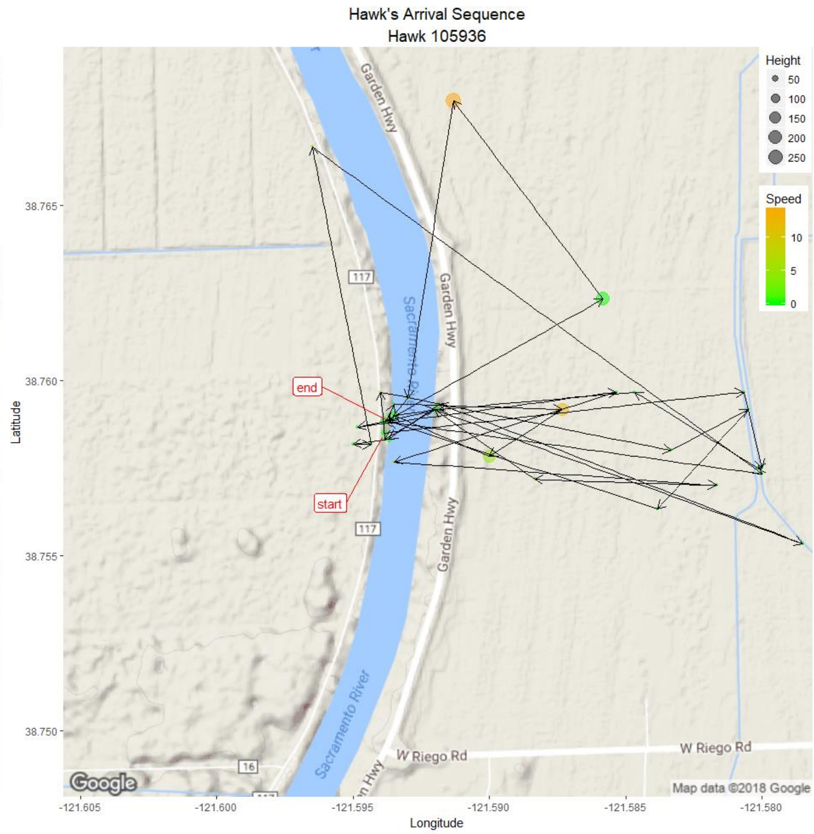
The plot shows that there are 5 different hawks separated by their identification tags. From the geospatial map it's possible to see where the hawks are located. In the upper region, it's mostly just hawk 105928 located around Pleasant Grove. There's indication that other hawks such as 105930, 117527 and 105936 have been to this area also, but their time here is quite brief.

Further south there's a mix of hawks 105936 and 117527 in an area between King Farms and Counsman. The geological point of interest is that the Sacramento River cuts through roughly this area, between the city of Woodland and the capital Sacramento. It is possible that the hawks found that there's sufficient wildlife in this vicinity such that several of them can coexist in this area without fear of competition for prey.

Another hawk in the area, hawk 105930, is distinct from the other hawks in that this hawk will be seen flying throughout the area, spreading out more to the west, east, and south than the others. The last hawk 105923 seems to spend the most time near the capital Sacramento. It seems to have found a place to hunt near where the Sacramento River intersects with the capital Sacramento. Perhaps this is another area where a hawk has found sufficient prey to hunt.

2) In the second problem, the goal is to distinguish the hawks which contain an arrival sequence. From the data it is possible to first find out which of the 5 hawks contain an 'arrival' level of the 'stage' variable. After taking the subset of the data that is related to the arrival sequence, separate maps were made for the two hawks 105928 and 105936. Also, a different get_center2() was used which finds the median of the coordinates over the mean. The reason being that prevalent outliers in this instance could shift the map center further out than necessary. The 'arrival' stage is described as the time when the hawk arrives in the Natomas Basin, up until the day when they're first observed to be at their nest.



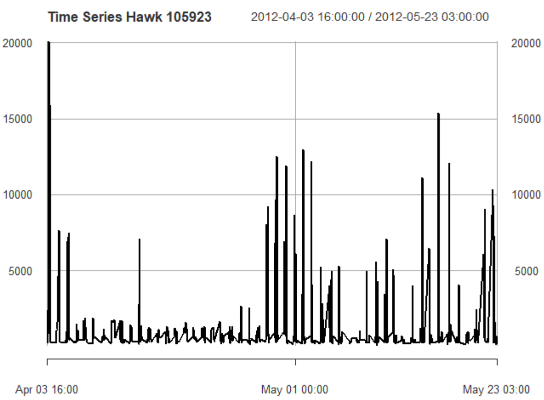


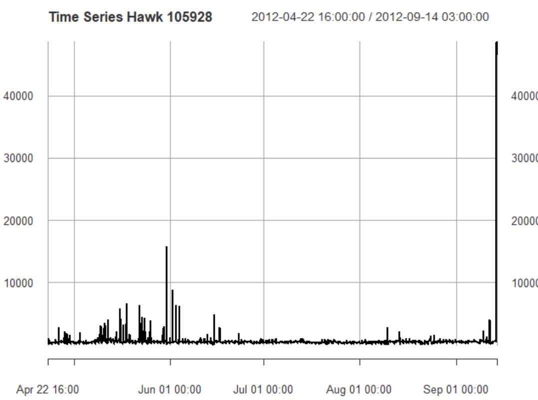
For the left hawk, hawk 105928, the hawk seems to have arrived in an area where a body of water begins to join with the Sacramento River (not shown). The hawk perhaps found that this area was highly fertile due to the nearby Sacramento River, and probably having prey which it may hunt in the vicinity. Therefore, the hawk decided to stay in this local area for quite some time before making a nest. Within the 'ggplot2' package, the function geom_segment() connects each point with a line in the order that they appear in the data, which in this case is by time. Additionally, the function can include arrows from one point to the next in sequence. Labels were also added from the 'ggrepel' package to indicate the start and end of the sequence of points. The pattern of the hawk seems to be that it flew around over the body of water for several hours. The height would vary, sometimes being high, other times possibly being on the ground. Usually the indication is that the hawk is flying due to the hawk being at a high speed rather than a still one.

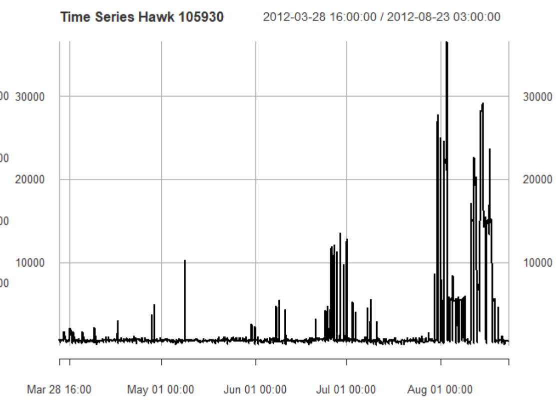
The right hawk, hawk 105936 seems to have decided to build a nest rather close to the previous hawk. The location is only somewhat further down the Sacramento River, and is also quite near to the Fremont Weir State Wildlife Area. This hawk likely also determined that the Sacramento River would make for fertile land that would allow for a comfortable climate to live nearby to avoid excessive heat, along with being able to find other wildlife which may serve as prey for the hawk to hunt. The data for this hawk during the 'arrival' stage seems to show a different pattern than the previous hawk. This hawk seems to have spent significantly more time at very low speeds and low heights, possibly roaming around on the ground instead of flying. Also, checking Google Map's satellite imagery shows that these locations are also where farms are located. There could possibly be numerous other wildlife such as rabbits or mice which the hawk may have decided are suitable prey in this area.

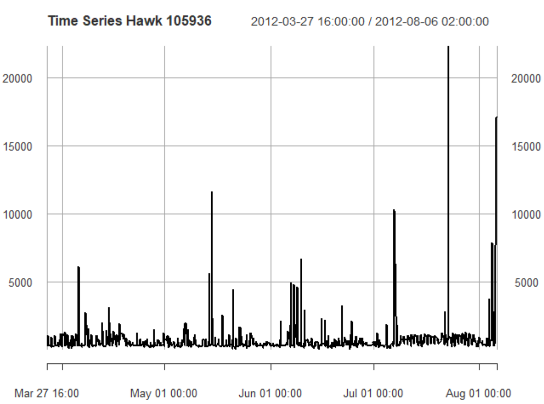
3) The first step was to create a variable for each of the factor levels of the hawk's 'tag' variable. This lead to there being 5 variables for the 5 hawks. After having seen the plot for the different hawks in problem 1, it's possible to determine that each of the hawks frequented a specific area most often, the reason is that there was the greatest density of points in certain areas. Since the data is ordered by time, it makes sense to conclude that the hawks therefore spent the most time in a local area. The data that we're observing is for the travel patterns of hawks, so it's also logical to conclude that each of the 5 hawks were spending time in a specific location for each of their nests. To determine exactly which coordinates are the locations of the nest for each hawk is not too simple. A location can be found simply by eye-balling where the longitude and latitude seem most dense based on the points. However, this method proves to be rather imprecise, so an alternative method was used. To find the densest cell in the grid of coordinates, a 2-D Kernel Density Estimation was used. This was accomplished with the nest_kde() function. It utilizes the kde2d() function that's part of the 'MASS' package. This location is the mode of the distribution of coordinate points. After determining the coordinates of the nests for each of the 5 hawks, a second function was created to determine the distance that each of the hawks were from their respective nest locations at any given hour. The function is called nest_distance(), and utilizes the distGeo() function that's part of the 'geosphere' package. Each of the hawks were then matched with their time series data using time_plot().

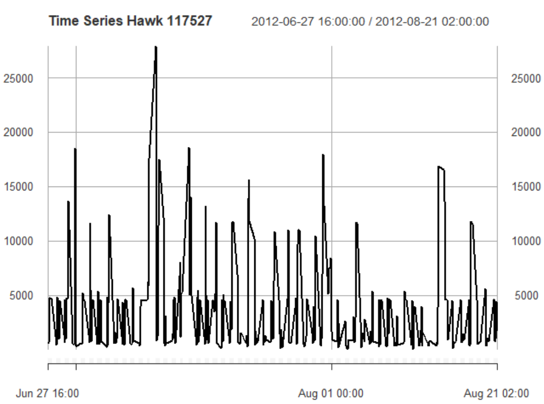
   The following plot shows time on the x-axis and distance in meters on the y-axis from the nest. The data was created using the 'xts' package, and the function time_plot() was created to smoothly create data frames where the indices are sequences of the time-related data, and the only column describes the distance that each hawk is from the nest at each interval. The following series of plots shows then that for each hawk tag, the hawks would often stay within a local range of their predicted nest locations, while occasionally making trips where they would travel further than typical from the nest site. The behavior of hawk 105923 on the top-left shows that the hawk seems to have at first traveled from a far distance before deciding on a nest location. In the beginning the hawk would stay within a short range of the nest. As time passed, the hawk began to take further trips away, and this became a more normal activity. The hawk would fly relatively far from the nest but would eventually return. Therefore, it isn't conclusive in anyway to say that the hawk ever left, since it's a typical habit for that hawk to fly far distances before returning later. Similarly, hawks 105930, 105936, and 117527 have habits of flying far distances from the nest but are shown to return within some period. So, there is only one hawk, 105928, which seems to both in the end not return to the nest along with not having a pattern of flying far away and coming back. Therefore, it seems that hawk 105928 is the only hawk to have left the nest completely.









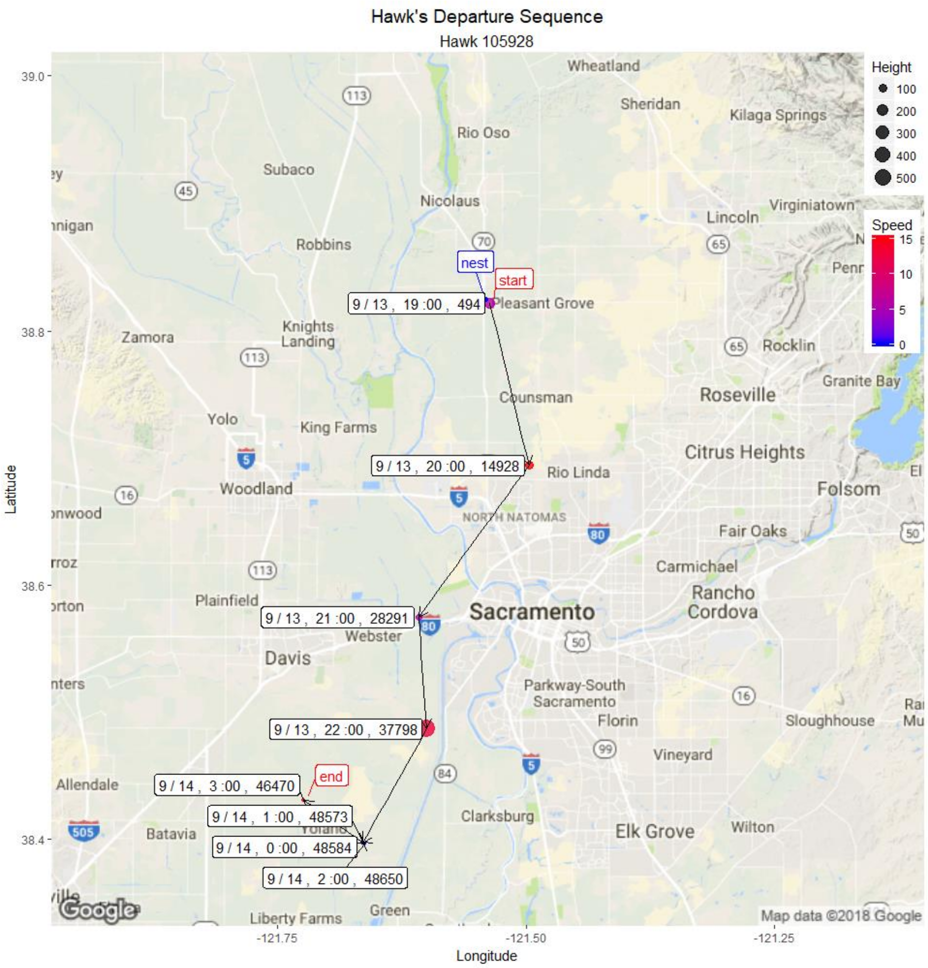


Using View() to examine the data at the end of the time intervals for hawk 105928, it’s evident that the hawk on 9/13/2012 at 20:00 has begun to take off on a long distance journey, and doesn’t show any sign of returning. Below are the data towards the end where the hawk begins to leave and continues further and further without significantly retracing its path.

Time	Distance
2012-09-13 19:00	494
2012-09-13 20:00	14928
2012-09-13 21:00	28291
2012-09-13 22:00	37798
2012-09-14 00:00	48584
2012-09-14 01:00	48573
2012-09-14 02:00	48650
2012-09-14 03:00	46470

It is worth noting, that like the data for the other hawks, it is not completely conclusive that the hawk has left the nest. It is possible that the hawk in a later date returns to the nest. For example, it could also be determined for other hawks such as 105936 that the hawk had left the nest if there were sufficient data to support that claim. However, the data does not go on continuously, so the belief that the hawk has left is based only on the available data.

4) The previous table which shows hawk 105928 and the dates from 9/13 until the end of the data for this hawk on 9/14 are plotted below. The plot shows that the hawk began rather near to the nest itself and began to move further away by the hour after 19:00. By midnight, the hawk had already traveled over 2 miles from the original nest location. Each of the 8 points are plotted along with the month, day, time, and distance that the hawk is from its nest. Previously, in the time-series plot, the hawk never had traveled this far from the nest before, helping with the notion that this hawk had departed from the nest and chose not to return. This is contrary to the other hawks, who are known to leave for certain lengths of time before eventually returning to the nest. To echo what was said previously, had there been more information regarding this hawk's whereabouts, it could be disproven that the hawk truly did depart from the nest. Therefore, this conclusion is still relative to the data that is provided in the original hawk dataset.



The movement of the hawk suddenly is rather interesting, since on one day the hawk seems to have spontaneously decided to fly from its normal area and head south past Sacramento. It's interesting to note however, that the hawk does seem to have followed the body of water downwards, going along with the Sacramento River. It is unlikely that the hawk is merely looking for prey, since such an extensive amount of ground was covered which could've meant that the hawk was leaving the nest to perhaps move somewhere warmer. During the Fall season (such as during the time period of when the hawk leaves), it makes sense that the hawk instinctively began to search for warmer areas to rest. Such migration patterns are known to exist in the natural world, so it's not necessarily an implausible assumption that the hawk left due to climate reasons.

Resources: Huong Vu, Minh Truong, Cameron Chen, Ben Millam, Tiffany Chen, Emily Watkins

TA Discussion code

Piazza: @559, @561, @565, @574

Online References:

http://ggplot2.tidyverse.org/reference/geom_path.html

https://www.littlemissdata.com/blog/maps

http://www.sthda.com/english/wiki/ggplot2-colors-how-to-change-colors-automatically-and-manually

http://www.sthda.com/english/wiki/ggplot2-legend-easy-steps-to-change-the-position-and-the-appearance-of-a-graph-legend-in-r-software

http://www.sthda.com/english/wiki/ggplot2-colors-how-to-change-colors-automatically-and-manually

http://www.sthda.com/english/wiki/ggplot2-legend-easy-steps-to-change-the-position-and-the-appearance-of-a-graph-legend-in-r-software

https://stackoverflow.com/questions/3421331/example-needed-using-arrow-with-ggplot2

https://stackoverflow.com/questions/30900745/plot-2d-kernel-density-from-a-dataframe-set-number-of-grid-positions-bandwith

https://stackoverflow.com/questions/39148984/distance-between-two-places-on-earth-in-r

http://blog.revolutionanalytics.com/2016/01/avoid-overlapping-labels-in-ggplot2-charts.html

https://stackoverflow.com/questions/36541086/adding-a-extra-point-in-a-ggplot2-graph

https://stackoverflow.com/questions/15624656/label-points-in-geom-point

https://stackoverflow.com/questions/38505412/r-geom-path-lines-closing-sometimes-how-to-keep-them-open

https://stackoverflow.com/questions/11724311/how-to-add-a-ggplot2-subtitle-with-different-size-and-colour

http://ggplot2.tidyverse.org/reference/geom_text.html

https://stackoverflow.com/questions/2288485/how-to-convert-a-data-frame-column-to-numeric-type

https://stackoverflow.com/questions/5290003/how-to-set-legend-alpha-with-ggplot2

Code Appendix:

```
library(ggmap); library(ggplot2); library(lubridate); library(geosphere); lib
rary(gridExtra) # Libraries
library(MASS); library(xts); library(forecast); library(tseries); library(ggr
epel)

hawks = read.csv("the_hawks.csv") # Load csv

### 1
get_center = function(x) {
  sapply(x[c("long", "lat")], function(x) mean(range((x))))
} # Find median of map to begin plot.

loc = get_center(hawks) # Save coordinates

hawk_map = get_map(loc, zoom = 10, scale = 1, source = "google", maptype = "t
errain", col = "bw") # Create map type
ggmap(hawk_map, # Plot hawks by tag
      base_layer = ggplot(hawks, aes(long, lat, color = factor(tag)))) +
  geom_point(alpha = 0.3) + labs(x = "Longitude", y = "Latitude", # Labels
                                 title = "Locations of Hawks", subtitle = "Se
parated by 5 Different Hawk Tags") +
  theme(plot.title = element_text(hjust = 0.5), plot.subtitle = element_text
(hjust = 0.5), # Title setting
        legend.justification = c(1, 1), legend.position = c(1, 1)) +
  scale_color_discrete("Hawk Tags") + # Legend title
  guides(colour = guide_legend(override.aes = list(alpha = 1))) # Change lege
nd alpha levels

### 2
unique(hawks$tag[hawks$stage == "arrival"]) # 105936, 105928
arrival1 = subset(hawks, (stage == "arrival" & tag == 105928)) # Subset the t
wo arrival hawks
arrival2 = subset(hawks, (stage == "arrival" & tag == 105936))

get_center2 = function(x) {
  sapply(x[c("long", "lat")], function(x) median((x)))
} # Find median of map to begin plot.

arrival1_loc = get_center2(arrival1) # Create maps for the two arrival hawks
arrival1_map = get_map(arrival1_loc, zoom = 15, scale = 1, source = "google",
 maptype = "terrain")
arrival2_loc = get_center2(arrival2)
arrival2_map = get_map(arrival2_loc, zoom = 15, scale = 1, source = "google",
 maptype = "terrain")

arrival1_plot = ggmap(arrival1_map, # Plot arrival1
                      base_layer = ggplot(arrival1, aes(long, lat))) +
  geom_point(aes(color = speed, size = height), alpha = 0.5) +
```

```
  scale_colour_gradient2(high = 'orange', mid = 'green') +
  geom_segment(aes(xend = c(tail(long, n = -1), NA), yend = c(tail(lat, n = -
1), NA)),
               arrow = arrow(length = unit(0.3, 'cm'))) +
  geom_label_repel(data = tail(arrival1, 1), aes(label = 'end'),
                   colour = 'red', point.padding =0.5, segment.color = 'red',
                   nudge_x = 0.002, nudge_y = 0.001) +
  geom_label_repel(data = head(arrival1, 1), aes(label = 'start'),
                   colour = 'red', point.padding = 0.5, segment.color = 'red
',
                   nudge_x = 0.005, nudge_y = -0.001) +
  labs(x = "Longitude", y = "Latitude", size = 'Height', color = 'Speed',
       title = "Hawk's Arrival Sequence", subtitle = "Hawk 105928") +
  theme(plot.title = element_text(hjust = 0.5, size = 14),
        plot.subtitle = element_text(hjust = 0.5, size = 14),
        legend.justification = c(1, 1), legend.position = c(1, 1)) +
  guides(size = guide_legend(order = 1))

arrival2_plot = ggmap(arrival2_map, # Plot arrival2
                     base_layer = ggplot(arrival2, aes(long, lat))) +
  geom_point(aes(color = speed, size = height), alpha = 0.5) +
  scale_colour_gradient2(high = 'orange', mid = 'green') +
  geom_segment(aes(xend = c(tail(long, n = -1), NA), yend = c(tail(lat, n = -
1), NA)),
               arrow = arrow(length = unit(0.3, 'cm'))) +
  geom_label_repel(data = tail(arrival2, 1), aes(label = 'end'),
                   colour = 'red', point.padding =0.5, segment.color = 'red',
                   nudge_x = -0.003, nudge_y = 0.001) +
  geom_label_repel(data = head(arrival2, 1), aes(label = 'start'),
                   colour = 'red', point.padding = 0.5, segment.color = 'red
',
                   nudge_x = -0.002, nudge_y = -0.002) +
  labs(x = "Longitude", y = "Latitude", size = 'Height', color = 'Speed',
       title = "Hawk's Arrival Sequence", subtitle = "Hawk 105936") +
  theme(plot.title = element_text(hjust = 0.5, size = 14),
        plot.subtitle = element_text(hjust = 0.5, size = 14),
        legend.justification = c(1, 1), legend.position = c(1, 1))

grid.arrange(arrival1_plot, arrival2_plot, nrow = 1) # Combine plots using gr
idExtra

### 3
levels(as.factor(hawks$tag)) # Find the different hawk tags

hawk1 = subset(hawks, tag == "105923"); hawk2 = subset(hawks, tag == "105928
")
hawk3 = subset(hawks, tag == "105930"); hawk4 = subset(hawks, tag == "105936
")
hawk5 = subset(hawks, tag == "117527") # Subset each of the hawks
```

```r
nest_kde = function(hawk) {
  dens = kde2d(hawk$long, hawk$lat, n = 100)
  long = dens$x[which.max(dens$z) %% length(dens$x)]
  lat = dens$y[floor(which.max(dens$z)/length(dens$y))]
  coords = c(long, lat)
  return(coords)
} # Determine the KDE coordinate

nest1 = nest_kde(hawk1); nest2 = nest_kde(hawk2); nest3 = nest_kde(hawk3)
nest4 = nest_kde(hawk4); nest5 = nest_kde(hawk5) # Find out where the nest li
kely is per hawk

nest_distance = function(nest, hawk) {
  distance_vec = rep(0, nrow(hawk))
  for (i in 1:nrow(hawk)) {
    distance_vec[i] = round(distGeo(nest, hawk[i,3:4]))
  }
  return(distance_vec)
} # Determine the distance from the nest by hour

flying1 = nest_distance(nest1, hawk1); flying2 = nest_distance(nest2, hawk2)
flying3 = nest_distance(nest3, hawk3); flying4 = nest_distance(nest4, hawk4)
flying5 = nest_distance(nest5, hawk5) # Find out when the bird is likely flyi
ng

time_plot = function(hawk_num, nest_vector) {
  hawk_time = as.POSIXct(hawk_num$time)
  hawk = xts(nest_vector, order.by = hawk_time)
  return(hawk)
} # Time series version of distances

hwk1 = time_plot(hawk1, flying1); hwk2 = time_plot(hawk2, flying2)
hwk3 = time_plot(hawk3, flying3); hwk4 = time_plot(hawk4, flying4)
hwk5 = time_plot(hawk5, flying5) # Match the time with each hawks' flying vec
tor

plot1 = plot(hwk1, main = "Time Series Hawk 105923"); plot2 = plot(hwk2, main
 = "Time Series Hawk 105928")
plot3 = plot(hwk3, main = "Time Series Hawk 105930"); plot4 = plot(hwk4, main
 = "Time Series Hawk 105936")
plot5 = plot(hwk5, main = "Time Series Hawk 117527"); par(mfrow=c(2,3)); plot
1; plot2; plot3; plot4; plot5
dev.off() # Plot the time series for each of the hawks in a single chart

### 4
# Adjust nest data into a dataframe that fits with ggplot
nest2a = as.data.frame(t(nest2)); colnames(nest2a) = c("long", "lat")

departure2 = subset(hawks, tag == 105928) # Subset the hawk type
departure2$Distance = flying2 # Add distance column
```

```r
View(departure2) # Examine the latest dates to see from which indices did the
 hawk begin to depart

departure2_loc = get_center2(departure2) # Obtain median coordinates
departure2_loc[2] = departure2_loc[2]-0.15 # Get adjusted coordinates
departure2 = departure2[1699:1706,] # Subset the time of departure

departure2_map = get_map(departure2_loc, zoom = 10, scale = 1, source = "goog
le", maptype = "terrain")
departure2_plot = ggmap(departure2_map, # Plot the departure
                        base_layer = ggplot(departure2, aes(long, lat))) +
  geom_point(aes(color = speed, size = height), alpha = 0.8) +
  scale_colour_gradient2(high = 'red', mid = 'blue') +
  geom_segment(aes(xend = c(tail(long, n = -1), NA), yend = c(tail(lat, n = -
1), NA)),
               arrow = arrow(length = unit(0.3, 'cm'))) +
  geom_label_repel(data = tail(departure2, 1), aes(label = 'end'),
                   colour = 'red', point.padding =0.5, segment.color = 'red',

                   nudge_x = 0.02, nudge_y = 0.01) +
  geom_label_repel(data = head(departure2, 1), aes(label = 'start'),
                   colour = 'red', point.padding = 0.5, segment.color = 'red
',
                   nudge_x = 0.01, nudge_y = 0.01) +
  labs(x = "Longitude", y = "Latitude", size = 'Height', color = 'Speed',
       title = "Hawk's Departure Sequence", subtitle = "Hawk 105928") +
  theme(plot.title = element_text(hjust = 0.5, size = 14),
        plot.subtitle = element_text(hjust = 0.5, size = 12),
        legend.justification = c(1, 1), legend.position = c(1, 1)) +
  geom_point(x = nest2[1], y = nest2[2], color = "blue") +
  geom_label_repel(data = nest2a, aes(label = 'nest'),
                   colour = 'blue', point.padding = 0.5, segment.color = 'blu
e',
                   nudge_x = -0.01, nudge_y = 0.03) +
  geom_label_repel(aes(label = paste(month(departure2$time), "/",
                                     day(departure2$time), ", ", hour(departu
re2$time), ":00",
                                     ", ", departure2$Distance)), nudge_x = -
0.05)
```