Essentials of Spatial Ecology: GIS Analysis in R

Week 5 Assignment - Introduction to Animal Movement Analyses

Due: Saturday, April 17 2021 at 11:59 pm EDT

# Description

In this week’s assignment, you’ll follow instructions provided in class to conduct a Resource Selection Function (RSF) analysis on a single white-bearded wildebeest (*Connochaetes taurinus*) tracked via Lotek WildCell GPS collars from 2010-2013. We will focus on animals monitored across the Athi-Kaputiei Plains in southern Kenya. This area is located directly adjacent to Kenya’s capital city, Nairobi. Human population density across the study area is ~45 people/km. Nairobi National Park, established in 1946, represents the dry season range of the species and is located at the northernmost section of the landscape. The park is fenced along its western, northern, and eastern border.

Objectives of this exercise are to improve your skills working with spatial objects. Similar to analyses with Addax (*Addax nasomaculatus*) and dorcas gazelle (*Gazelle dorcas*), you’ll fit a logistic regression model to the data and provide a raster prediction of habitat suitability using a Use vs Availability design. We will also use the output prediction for animal 30077 to provide an average prediction between the two animals.

Once again, please do not hesitate to ask questions or to offer help to others on the [Blackboard](https://mymasonportal.gmu.edu/) discussion board. As per usual, we expect everyone to submit an independent project, but there is no problem with providing assistance to others. Please upload a WORD or html document answering the question below, along with your annotated R script. Both documents should be posted to the relevant assignment in [Blackboard](https://mymasonportal.gmu.edu/).

# Datasets

All datasets required to complete this homework are included in the files provided during lecture. You will need to load two .Rdata files to access the wildebeest movement dataset and the SpatialPolygonsDataFrame of the study area boundary (wild.Rdata) and the spatial data layers (Wild\_SpatialData.Rdata) to relate to the “Use” dataset. Both files are included in the “.zip” file on Blackboard, within the Week4/Data directory. The files included are:

* A dataframe, named wild.Athi, which includes the “Use” locations of 12 wildebeest
* A SpatialPolygonsDataFrame, named Athi\_Bound, which is the boundary of the study area
* Raster files (7 files total) of anthropogenic risk (anth\_risk), distance to fences (fence\_dist), distance to primary roads (prirds\_dist), distance to secondary roads (secrds\_dist), distance to permanent rivers (river\_dist), distance to water wells (waterpts\_dist), and distance to woody vegetation (woody\_dist).

You’ll need the following libraries to complete this assignment:

library(amt)   
library(dplyr)  
library(ggplot2)  
library(jtools)  
library(lme4)  
library(lubridate)  
library(raster)  
library(sp)  
library(sjPlot)  
library(tmap)  
library(usdm)  
library(visreg)

# Assignment

Load the wildMara.Rdata and Wild\_SpatialData.Rdata file into R.

load(file = "./Data/wild.Rdata")  
load(file = "./Data/Wild\_SpatialData.Rdata")

You will now see the movement datasets (you can ignore the wild.Mara and the wb.sim files for this exercise) in your working environment:

* wildAthi - Dataframe (not spatial)
* Athi\_Bound - SpatialPolygonsDataFrame
* anth\_risk - Raster Layer
* fence\_dist - Raster Layer
* prirds\_dist - Raster Layer
* secrds\_dist - Raster Layer
* river\_dist - Raster Layer
* waterpts\_dist - Raster Layer
* woody\_dist - Raster Layer

**Question 1**: Are all the files spatial and in the same coordinate reference system? If no, make all files spatial and make sure their projections match. Use the same projection information provided in class (Albers Equal Area).

# Visually look at all the files  
# wild Athi file is a dataframe and most be projected  
# All other files are fine  
head(wild.Athi)

## id timestamp location.long location.lat external.temperature  
## 1 2840 2010-10-16 16:00:00 36.84518 -1.357289 31  
## 2 2840 2010-10-16 17:00:00 36.84518 -1.357294 28  
## 3 2840 2010-10-16 18:00:00 36.84676 -1.358796 25  
## 4 2840 2010-10-16 21:01:00 36.84691 -1.360435 24  
## 5 2840 2010-10-17 00:00:00 36.84826 -1.362778 20  
## 6 2840 2010-10-17 03:00:00 36.84964 -1.363394 18  
## gps.dop gps.fix.type.raw height.above.ellipsoid sensor.type  
## 1 1.6 3D 1627.62 gps  
## 2 3.2 3D 1642.64 gps  
## 3 4.4 3D 1639.90 gps  
## 4 2.8 3D 1630.99 gps  
## 5 4.6 3D 1627.89 gps  
## 6 1.6 3D 1624.66 gps  
## tag.local.identifier individual.local.identifier animal.life.stage animal.sex  
## 1 2840 Sotua 6 years m  
## 2 2840 Sotua 6 years m  
## 3 2840 Sotua 6 years m  
## 4 2840 Sotua 6 years m  
## 5 2840 Sotua 6 years m  
## 6 2840 Sotua 6 years m  
## study.site  
## 1 Athi-Kaputiei Plains  
## 2 Athi-Kaputiei Plains  
## 3 Athi-Kaputiei Plains  
## 4 Athi-Kaputiei Plains  
## 5 Athi-Kaputiei Plains  
## 6 Athi-Kaputiei Plains

Athi\_Bound

## class : SpatialPolygonsDataFrame   
## features : 1   
## extent : 1203934, 1256094, -230250.7, -160721.6 (xmin, xmax, ymin, ymax)  
## crs : +proj=aea +lat\_0=0 +lon\_0=25 +lat\_1=20 +lat\_2=-23 +x\_0=0 +y\_0=0 +datum=WGS84 +units=m +no\_defs   
## variables : 1  
## names : Id   
## value : 1

anth\_risk

## class : RasterLayer   
## dimensions : 413, 678, 280014 (nrow, ncol, ncell)  
## resolution : 250, 250 (x, y)  
## extent : 1147557, 1317057, -250317.1, -147067.1 (xmin, xmax, ymin, ymax)  
## crs : +proj=aea +lat\_0=0 +lon\_0=25 +lat\_1=20 +lat\_2=-23 +x\_0=0 +y\_0=0 +datum=WGS84 +units=m +no\_defs   
## source : memory  
## names : anth\_risk   
## values : 0, 101.4775 (min, max)

fence\_dist

## class : RasterLayer   
## dimensions : 413, 678, 280014 (nrow, ncol, ncell)  
## resolution : 250, 250 (x, y)  
## extent : 1147557, 1317057, -250317.1, -147067.1 (xmin, xmax, ymin, ymax)  
## crs : +proj=aea +lat\_0=0 +lon\_0=25 +lat\_1=20 +lat\_2=-23 +x\_0=0 +y\_0=0 +datum=WGS84 +units=m +no\_defs   
## source : memory  
## names : fence\_dist   
## values : 0, 79517.91 (min, max)

prirds\_dist

## class : RasterLayer   
## dimensions : 413, 678, 280014 (nrow, ncol, ncell)  
## resolution : 250, 250 (x, y)  
## extent : 1147557, 1317057, -250317.1, -147067.1 (xmin, xmax, ymin, ymax)  
## crs : +proj=aea +lat\_0=0 +lon\_0=25 +lat\_1=20 +lat\_2=-23 +x\_0=0 +y\_0=0 +datum=WGS84 +units=m +no\_defs   
## source : memory  
## names : prirds\_dist   
## values : 27.93375, 53381.27 (min, max)

secrds\_dist

## class : RasterLayer   
## dimensions : 413, 678, 280014 (nrow, ncol, ncell)  
## resolution : 250, 250 (x, y)  
## extent : 1147557, 1317057, -250317.1, -147067.1 (xmin, xmax, ymin, ymax)  
## crs : +proj=aea +lat\_0=0 +lon\_0=25 +lat\_1=20 +lat\_2=-23 +x\_0=0 +y\_0=0 +datum=WGS84 +units=m +no\_defs   
## source : memory  
## names : secrds\_dist   
## values : 26.13304, 36864.55 (min, max)

river\_dist

## class : RasterLayer   
## dimensions : 413, 678, 280014 (nrow, ncol, ncell)  
## resolution : 250, 250 (x, y)  
## extent : 1147557, 1317057, -250317.1, -147067.1 (xmin, xmax, ymin, ymax)  
## crs : +proj=aea +lat\_0=0 +lon\_0=25 +lat\_1=20 +lat\_2=-23 +x\_0=0 +y\_0=0 +datum=WGS84 +units=m +no\_defs   
## source : memory  
## names : river\_dist   
## values : 22.67398, 17605.76 (min, max)

waterpts\_dist

## class : RasterLayer   
## dimensions : 413, 678, 280014 (nrow, ncol, ncell)  
## resolution : 250, 250 (x, y)  
## extent : 1147557, 1317057, -250317.1, -147067.1 (xmin, xmax, ymin, ymax)  
## crs : +proj=aea +lat\_0=0 +lon\_0=25 +lat\_1=20 +lat\_2=-23 +x\_0=0 +y\_0=0 +datum=WGS84 +units=m +no\_defs   
## source : memory  
## names : waterpts\_dist   
## values : 68.64046, 85374.01 (min, max)

woody\_dist

## class : RasterLayer   
## dimensions : 413, 678, 280014 (nrow, ncol, ncell)  
## resolution : 250, 250 (x, y)  
## extent : 1147557, 1317057, -250317.1, -147067.1 (xmin, xmax, ymin, ymax)  
## crs : +proj=aea +lat\_0=0 +lon\_0=25 +lat\_1=20 +lat\_2=-23 +x\_0=0 +y\_0=0 +datum=WGS84 +units=m +no\_defs   
## source : memory  
## names : woody\_dist   
## values : -14.0625, 28972.48 (min, max)

# Spatial reference information  
LatLong.proj <- "+proj=longlat +ellps=WGS84 +datum=WGS84 +no\_defs" # EPSG:4326  
AEA.Africa.proj <- "+proj=aea +lat\_1=20 +lat\_2=-23 +lat\_0=0 +lon\_0=25 +x\_0=0 +y\_0=0 +ellps=WGS84 +datum=WGS84 +units=m +no\_defs" #ESRI:102022  
  
# File has Lat/Long coordinates. Coordinates must be defined first and then projected to Albers  
wild.Athi <- SpatialPointsDataFrame(coords = wild.Athi[,c("location.long","location.lat")],  
 data = wild.Athi,  
 proj4string = CRS(LatLong.proj))   
  
# Transform to Albers Equal Area Project  
wild.Athi <- spTransform(wild.Athi,  
 CRS = CRS(AEA.Africa.proj))

For the next part of the exercises, please convert the wild.Athi back to a dataframe (we only made the file spatial to get the needed coordinates) and subset to animal 2840 (Sotua). Remove unnecessary fields (keeping only id, animal, sex, timestamp, easting, and northing) and name this new file wb2 using the following steps. Please set the seed in your script to set.seed(99) to assure we get the same results.

# Set Seed  
set.seed(99)  
  
# Convert to dataframe for statistical analysis  
wild.Athi <- as.data.frame(wild.Athi)  
  
# Remove columns and rename  
wild.Athi <- wild.Athi[,c(1,11,13,2,15:16)] # Also good to assure yourself that these columns are correct.  
colnames(wild.Athi) <- c("id","animal","sex","timestamp","easting","northing")  
  
wb2 <- wild.Athi[wild.Athi$id==2840,]

Now, convert this dataframe subset (“wb2”) to a track using the mk\_track function in the amt packages. Please order the data based on the timestamp. Continue to name this file “wb2”.

**Question 2**: How many data observations were records for “Sotua”? What is the average step length? What is the average sampling interval?

# How many observations?  
nrow(wb2)

## [1] 4777

# 4777 observations  
  
# Calculate track  
wb2 <- mk\_track(wb2, .x = easting, .y = northing,  
 .t = timestamp, crs = CRS(AEA.Africa.proj),  
 order\_by\_ts = T)

## .t found, creating `track\_xyt`.

# Add steplengths to object  
wb2 <- wb2 %>% mutate(  
 sl = step\_lengths(.))  
  
# Average step length?  
summary(wb2$sl)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.19 39.72 165.91 381.29 435.78 10892.81 1

# 381.29 meters  
  
# Summarize the distribution of time intervals between successive locations to get a general impression for the sampling rate.  
summarize\_sampling\_rate(wb2)

## # A tibble: 1 x 9  
## min q1 median mean q3 max sd n unit   
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <int> <chr>  
## 1 0.95 1 1 1.67 2.97 201. 3.19 4776 hour

# 1.67 hours

Please resample the track to a 3 hour interval and thin the dataset, selecting a random subset of 20% of the “Use” points (same as what we did for animal 30077). Then, generate 50 “Available” points per “Use” point, extract the covariates from the raster layers, and fit a Binomial GLM using the same model structure and covariates provided in class. Name the model M.2840. Make sure to scale and center the covariates included in the model. Plot the animals’ response to Anthropogenic Risk (anth\_risk). Create a predictive surface of the result based on your model coefficients.

**Question 3**: Are there any issues of collinearity that we should be worried about? Is the response of animal 2840 (Sotua) to Anthropogenic risk similar or different to animal 30077 (Ntishya)? Note, I am only looking for a simple visual comparison to the response curve plots from each of these models (i.e., is the effect the same between animals?).

wb2 <- wb2 %>% track\_resample(rate = hours(3), tolerance = minutes(20))  
wb2

## # A tibble: 2,370 x 5  
## x\_ y\_ t\_ sl burst\_  
## \* <dbl> <dbl> <dttm> <dbl> <dbl>  
## 1 1227064. -164301. 2010-10-16 16:00:00 0.689 1  
## 2 1227242. -164676. 2010-10-16 21:01:00 311. 2  
## 3 1227380. -164955. 2010-10-17 00:00:00 161. 2  
## 4 1227522. -165029. 2010-10-17 03:00:00 343. 2  
## 5 1227381. -164716. 2010-10-17 06:00:00 316. 2  
## 6 1227014. -164273. 2010-10-17 09:00:00 196. 2  
## 7 1227158. -164513. 2010-10-17 12:00:00 48.4 2  
## 8 1227255. -164690. 2010-10-17 15:00:00 13.3 2  
## 9 1227196. -164626. 2010-10-17 18:00:00 87.5 2  
## 10 1227267. -164575. 2010-10-17 21:00:00 330. 2  
## # ... with 2,360 more rows

# Ad-hoc thinning of Use locations - 20% removed  
rcd.amt <- ceiling(nrow(wb2)\*0.20)  
wb2 <- wb2[sample(1:nrow(wb2),size = rcd.amt),c("x\_","y\_","t\_")]  
  
# Generating 50 available locations per each used location within animal homerange  
wb2\_RSF50 <- random\_points(wb2,n=nrow(wb2)\*50,typ="random")  
  
# Stack rasters  
rsf.stack <- stack(anth\_risk,fence\_dist,prirds\_dist,secrds\_dist,river\_dist,waterpts\_dist,woody\_dist)  
  
# Extracting the covariates  
wb2\_RSF50 <- wb2\_RSF50 %>% extract\_covariates(rsf.stack)  
  
# Assess collinearity/Correlation  
test.corr <- as.data.frame(wb2\_RSF50)  
vifstep(test.corr[,4:10]) # VIF indicates no collinearity problem, but there is a high level of correlation (max: 0.84)

## No variable from the 7 input variables has collinearity problem.   
##   
## The linear correlation coefficients ranges between:   
## min correlation ( woody\_dist ~ anth\_risk ): 0.008195284   
## max correlation ( waterpts\_dist ~ fence\_dist ): 0.8018695   
##   
## ---------- VIFs of the remained variables --------   
## Variables VIF  
## 1 anth\_risk 1.759089  
## 2 fence\_dist 5.439820  
## 3 prirds\_dist 1.547280  
## 4 secrds\_dist 1.899896  
## 5 river\_dist 1.322528  
## 6 waterpts\_dist 3.413991  
## 7 woody\_dist 2.168809

cor(test.corr[,4:10])

## anth\_risk fence\_dist prirds\_dist secrds\_dist river\_dist  
## anth\_risk 1.00000000 -0.55930708 0.05176007 -0.07113847 0.11846092  
## fence\_dist -0.55930708 1.00000000 -0.48295162 -0.18803621 0.03657647  
## prirds\_dist 0.05176007 -0.48295162 1.00000000 0.12154978 -0.01505541  
## secrds\_dist -0.07113847 -0.18803621 0.12154978 1.00000000 -0.19908661  
## river\_dist 0.11846092 0.03657647 -0.01505541 -0.19908661 1.00000000  
## waterpts\_dist -0.50989051 0.80374634 -0.31514049 0.02045209 -0.07206098  
## woody\_dist 0.01515987 -0.26472044 0.05939937 0.60417141 0.17217415  
## waterpts\_dist woody\_dist  
## anth\_risk -0.50989051 0.01515987  
## fence\_dist 0.80374634 -0.26472044  
## prirds\_dist -0.31514049 0.05939937  
## secrds\_dist 0.02045209 0.60417141  
## river\_dist -0.07206098 0.17217415  
## waterpts\_dist 1.00000000 -0.03457125  
## woody\_dist -0.03457125 1.00000000

# Once again, no collinearity issues.  
  
# Create Model  
M.2840 <- glm(case\_ ~ scale(anth\_risk) + scale(fence\_dist) + scale(waterpts\_dist) + scale(prirds\_dist) + scale(river\_dist) + scale(secrds\_dist) + + scale(woody\_dist), family = binomial(link="logit"),   
 data = wb2\_RSF50)  
summary(M.2840)

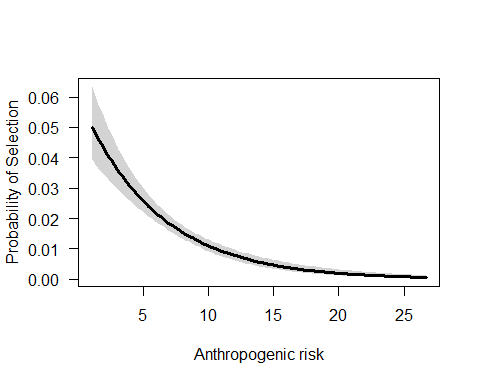
##   
## Call:  
## glm(formula = case\_ ~ scale(anth\_risk) + scale(fence\_dist) +   
## scale(waterpts\_dist) + scale(prirds\_dist) + scale(river\_dist) +   
## scale(secrds\_dist) + +scale(woody\_dist), family = binomial(link = "logit"),   
## data = wb2\_RSF50)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.5304 -0.2353 -0.1544 -0.0793 3.7434   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -4.68532 0.09263 -50.581 < 2e-16 \*\*\*  
## scale(anth\_risk) -0.69062 0.07291 -9.472 < 2e-16 \*\*\*  
## scale(fence\_dist) -0.36943 0.13964 -2.646 0.00815 \*\*   
## scale(waterpts\_dist) -0.62658 0.11727 -5.343 9.15e-08 \*\*\*  
## scale(prirds\_dist) -0.36859 0.05397 -6.829 8.52e-12 \*\*\*  
## scale(river\_dist) 0.44651 0.05095 8.764 < 2e-16 \*\*\*  
## scale(secrds\_dist) -0.87012 0.13163 -6.610 3.83e-11 \*\*\*  
## scale(woody\_dist) -0.62909 0.08548 -7.360 1.84e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 4666.0 on 24173 degrees of freedom  
## Residual deviance: 4180.5 on 24166 degrees of freedom  
## AIC: 4196.5  
##   
## Number of Fisher Scoring iterations: 8

confint(M.2840)

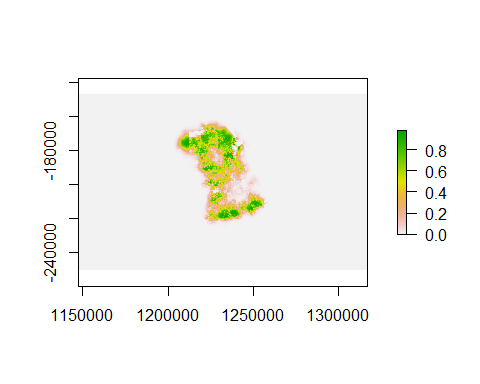
## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) -4.8743863 -4.51090612  
## scale(anth\_risk) -0.8362391 -0.55035192  
## scale(fence\_dist) -0.6419133 -0.09416285  
## scale(waterpts\_dist) -0.8602897 -0.40047102  
## scale(prirds\_dist) -0.4745850 -0.26295498  
## scale(river\_dist) 0.3464590 0.54624996  
## scale(secrds\_dist) -1.1359578 -0.61973619  
## scale(woody\_dist) -0.8015519 -0.46612678

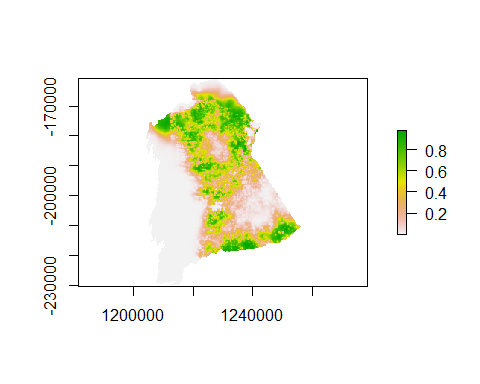
# Plot the response  
visreg(M.2840,"anth\_risk",  
 scale="response",   
 ylab="Probability of Selection",  
 xlab="Anthropogenic risk",  
 partial=F,  
 rug=F,  
 line=list(col="black"),   
 fill=list(col="light gray"))



# The trend of anthropogenic disturbance is the same with animal 2840 as it is with 30077. As anthropogenic risk increases, the probability of selection decreases.  
  
# Create prediction - Extract the coefficient of each predictor from the model summary  
coeff <- M.2840$coefficients  
  
# Remember that we must scale the raster layers too:  
anth.scale <- (anth\_risk - mean(wb2\_RSF50$anth\_risk)) / sd(wb2\_RSF50$anth\_risk)  
fence.scale <- (fence\_dist - mean(wb2\_RSF50$fence\_dist)) / sd(wb2\_RSF50$fence\_dist)  
water.scale <- (prirds\_dist - mean(wb2\_RSF50$prirds\_dist)) / sd(wb2\_RSF50$prirds\_dist)  
prirds.scale <- (secrds\_dist - mean(wb2\_RSF50$secrds\_dist)) / sd(wb2\_RSF50$secrds\_dist)  
river.scale <- (river\_dist - mean(wb2\_RSF50$river\_dist)) / sd(wb2\_RSF50$river\_dist)  
secrds.scale <- (waterpts\_dist - mean(wb2\_RSF50$waterpts\_dist)) / sd(wb2\_RSF50$waterpts\_dist)   
woody.scale <- (woody\_dist - mean(wb2\_RSF50$woody\_dist)) / sd(wb2\_RSF50$woody\_dist)   
  
# Prediction  
pred <- exp(anth.scale\*coeff[[2]]+ fence.scale\*coeff[[3]] + water.scale\*coeff[[4]] + prirds.scale\*coeff[[5]] + river.scale\*coeff[[6]] + secrds.scale\*coeff[[7]] + woody.scale\*coeff[[8]])/(1+exp(anth.scale\*coeff[[2]]+ fence.scale\*coeff[[3]] + water.scale\*coeff[[4]] + prirds.scale\*coeff[[5]] + river.scale\*coeff[[6]] + secrds.scale\*coeff[[7]] + woody.scale\*coeff[[8]]))  
  
# Provide Spatial Prediction - Based off of the coefficients from this single animal  
plot(pred)



# Create new layer and crop it to the analysis extent  
pred.2840 <- mask(pred, Athi\_Bound)  
pred.2840 <- crop(pred.2840,y=extent(Athi\_Bound))  
plot(pred.2840)



Lastly, using the predictive surface created in class for animal 30077, combine the result from animal 2840 by taking the mean of the two predictions (Mn.Predict <- mean(pred.30077, pred.2840)), making sure that the extents exactly match. To save the prediction that we completed in class, run all the code and then save the prediction to a .Rdata file. You can then load the saved file in your current R session. As a final product, use tmap to generate the mean output prediction between the two animals. See the Addax exercises for instructions on creating a tmap output. Use “OpenStreetMap” as your basemap. Use the Export to Image to include the tmap output in your homework file.

# At the end of script to fit 30077  
# We called this file pred.30077  
save(pred.30077, file = "./Data/Prediction30077.Rdata")  
  
# In your current script to fit 2840  
load(file = "./Data/Prediction30077.Rdata")  
  
# You will then see the pred.mask file in your environment settings.

**Question 4**: Does the output change noticeably when the two animals are combined together?

# Load prediction  
load(file = "./Data/Prediction30077.Rdata")  
  
# Create mean prediction  
Mn.Predict <- mean(pred.2840, pred.30077)  
plot(Mn.Predict)  
  
# Create tmap\_model  
tmap\_mode("view")  
tm\_basemap("OpenStreetMap") +  
 tm\_shape(Mn.Predict, name = "Habitat suitability") +  
 tm\_raster(palette="-inferno", n=8, alpha=0.6,   
 title = "Wildebeest Habitat Suitability")  
  
# Yes, the additional information from two animals, highlights differences in habitat suitability across the region.