

Spatial Analysis in R

M. Jordan and TA Gelmi-Candusso

2024-05-15

Spatial data

Spatial data is any type of data that directly or indirectly references a specific geographical area or location. These can be, for example, geographic features on the landscape or environmental properties of an area such as temperature or air quality.

Spatial data can be continuous or discrete just like regular data, and in both cases it can be represented as a vector or a raster. The main difference being vector data uses points and lines to represent spatial data, while raster data represents data uses pixelled or gridded, where each pixel/cell represents a specific geographic location and the information therein. Raster data will be heavily influenced by the size of the pixels/cells, i.e. resolution.

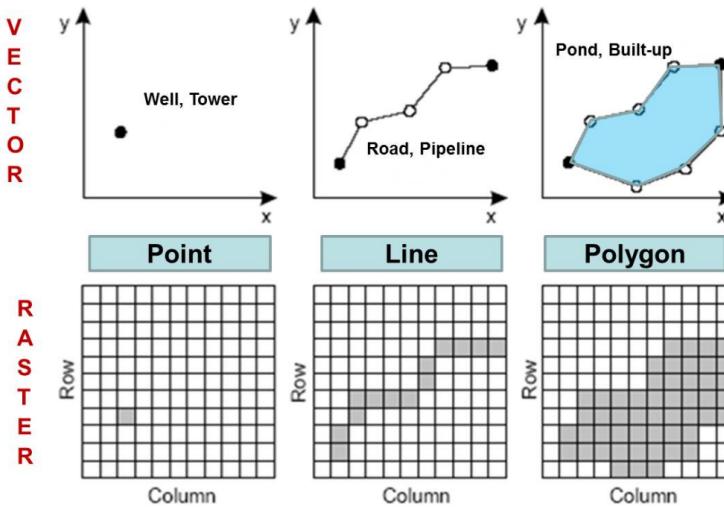
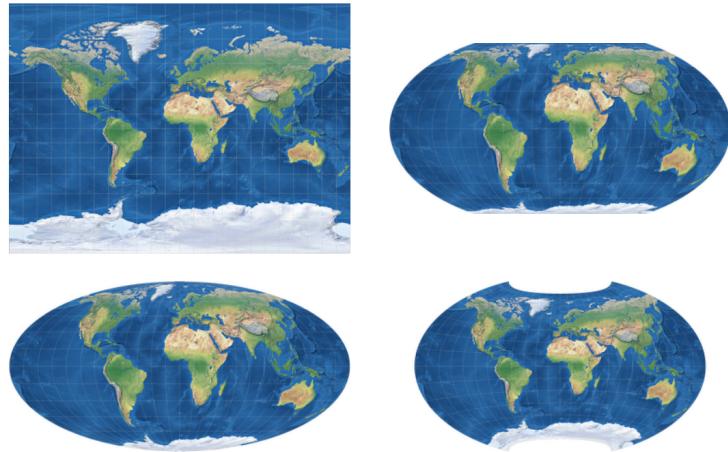


Figure 1: Figure showing difference between vectors and rasters

Both vector and raster data are planar representations of the world, a 3-dimensional sphere, and as such are not perfect copies. Depending on how the planar representation is created it will distort more or less certain areas of the world, therefore many representations exist. These are called projections, as the representations project the 3 dimensional spheric image into a planar, 2-dimensional image.

Maps with different projections are not comparable and cannot be overlaid. Therefore, we need to make sure we work always on the same projection when using more maps. In addition, projections can have different coordinate systems and therefore when extracting distance information from maps, some projections (i.e. metric based projections e.g. Mercator projection or the Albers equal-area projection) will give a more accurate representation of the distance than others. To work around this, we can transform our maps between projections, in R we use EPSG codes to do this.



Left to right, top to bottom: Miller, Wagner IV, Mollweide, Cantrs W14.

Figure 2: Figure showing difference between vectors and rasters

Vector data

In this section we will read and manipulate vector data in R. * Vector data represents real world features within the GIS environment. A feature is anything you can see on the landscape. * Vector data is commonly stored as a shapefile and can contain either point data or polygon data. * Data contains information attached to each feature, we call these attributes.

Features can be points (red) representing specific x,y locations, such as a trees or camera sites; polygons (white) representing areas, such as forests or residential areas; and lines (yellow/green and blue) representing continuous linear features, such as roads or rivers

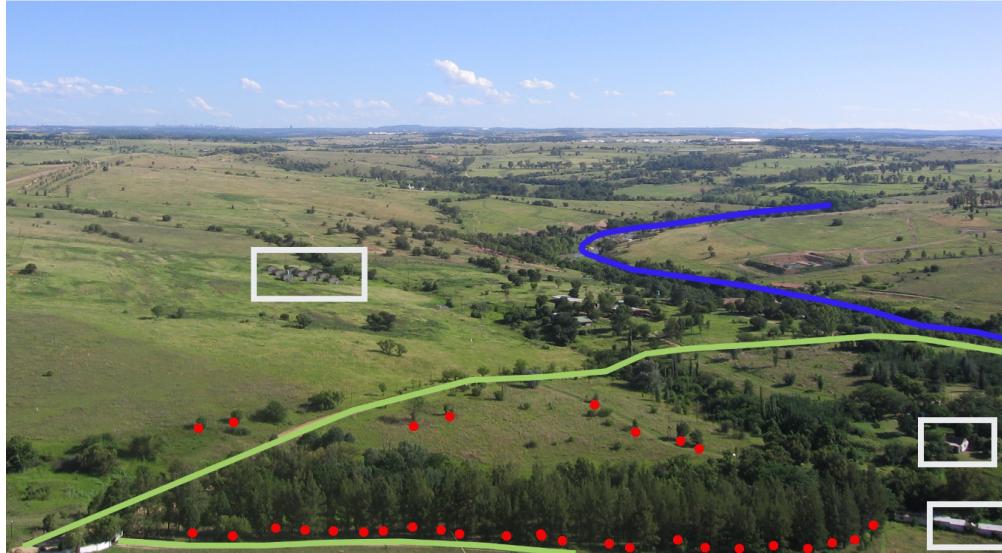


Figure 3: Figure showing polygons, points and lines in the landscape

Vector data reads as a data frame would, each row is a feature and each column is an attribute, and contains usually a geometry column where the xy coordinates for the shapes are stored. Plotting these data will plot the points or shapes in the map using the xy coordinates stored for each feature.

```

## Simple feature collection with 6 features and 3 fields
## Geometry type: POINT
## Dimension: XY
## Bounding box: xmin: -122.3607 ymin: 47.64339 xmax: -122.3607 ymax: 47.64339
## Geodetic CRS: WGS 84
##   speciesname    locationid           date      geometry
## 1 Procyon lotor SEWA_N01_DRP2 2019-07-03 05:02:21 POINT (-122.3607 47.64339)
## 2 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:04:45 POINT (-122.3607 47.64339)
## 3 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:12:10 POINT (-122.3607 47.64339)
## 4 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:13:16 POINT (-122.3607 47.64339)
## 5 Procyon lotor SEWA_N01_DRP2 2019-07-05 03:55:53 POINT (-122.3607 47.64339)
## 6 Procyon lotor SEWA_N01_DRP2 2019-07-05 04:05:21 POINT (-122.3607 47.64339)

```

Packages used to read and manipulate data include the sf package, that reads the shapefile as a spatial data frame, and the terra package that reads the shapefiles as a Spatvector, previously there was also the raster package, but we will try to avoid it as it has been deprecated.

```

library(sf)
library(terra)

```

Vector data: points

Point data can be obtained directly from a shapefile or a csv file where each row is a feature. In this case we will work with camera trap site data and the information collected at each site, i.e. point.

The camera trap sites here are located in Seattle, and have captured coyote and raccoon presence and absence from the 2019 spring season to the 2021 winter season.

The data is stored as a data frame in a csv.

```

captures.table <- read.csv("data/captures.csv")
print(head(captures.table))

```

```

##   speciesname    locationid           date latitude longitude
## 1 Procyon lotor SEWA_N01_DRP2 2019-07-03 05:02:21 47.64339 -122.3607
## 2 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:04:45 47.64339 -122.3607
## 3 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:12:10 47.64339 -122.3607
## 4 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:13:16 47.64339 -122.3607
## 5 Procyon lotor SEWA_N01_DRP2 2019-07-05 03:55:53 47.64339 -122.3607
## 6 Procyon lotor SEWA_N01_DRP2 2019-07-05 04:05:21 47.64339 -122.3607

```

The coordinates are stored in the latitude and longitude, to be able to observe these points in the map, and extract environmental information based on their location, we will have to convert it to a spatial data frame object. We will use the st_as_sf() function from the sf package and we will specify the projection (crs). How do we know which projection our data is in?

<>(This section will also introduce the idea of a CRS. Specifically introduce 4326 and 26910. Tell people how to find their UTM and state plane (others?))

```

captures.spatial <- st_as_sf(captures.table,
                           coords = c("longitude", "latitude"),
                           crs = 4326)
print(head(captures.spatial))

```

```

## Simple feature collection with 6 features and 3 fields
## Geometry type: POINT
## Dimension: XY
## Bounding box: xmin: -122.3607 ymin: 47.64339 xmax: -122.3607 ymax: 47.64339
## Geodetic CRS: WGS 84
##   speciesname    locationid           date      geometry
## 1 Procyon lotor SEWA_N01_DRP2 2019-07-03 05:02:21 POINT (-122.3607 47.64339)
## 2 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:04:45 POINT (-122.3607 47.64339)
## 3 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:12:10 POINT (-122.3607 47.64339)
## 4 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:13:16 POINT (-122.3607 47.64339)
## 5 Procyon lotor SEWA_N01_DRP2 2019-07-05 03:55:53 POINT (-122.3607 47.64339)
## 6 Procyon lotor SEWA_N01_DRP2 2019-07-05 04:05:21 POINT (-122.3607 47.64339)

```

We want our data to be in the NAD83 projection, because we need our data in the UTM coordinate system to be compatible with google map <>(I dont know if I understood correctly the comment below but wrote a draft anyways.) <>(Transform to UTM. Will only plot on google map if it's in lat/lon, so we need to think about where we introduce this.)

```

captures.utm <- st_transform(captures.spatial, crs = 26910)
print(head(captures.utm))

```

```

## Simple feature collection with 6 features and 3 fields
## Geometry type: POINT
## Dimension: XY
## Bounding box: xmin: 548015.5 ymin: 5276863 xmax: 548015.5 ymax: 5276863
## Projected CRS: NAD83 / UTM zone 10N
##   speciesname    locationid           date      geometry
## 1 Procyon lotor SEWA_N01_DRP2 2019-07-03 05:02:21 POINT (548015.5 5276863)
## 2 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:04:45 POINT (548015.5 5276863)
## 3 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:12:10 POINT (548015.5 5276863)
## 4 Procyon lotor SEWA_N01_DRP2 2019-07-03 06:13:16 POINT (548015.5 5276863)
## 5 Procyon lotor SEWA_N01_DRP2 2019-07-05 03:55:53 POINT (548015.5 5276863)
## 6 Procyon lotor SEWA_N01_DRP2 2019-07-05 04:05:21 POINT (548015.5 5276863)

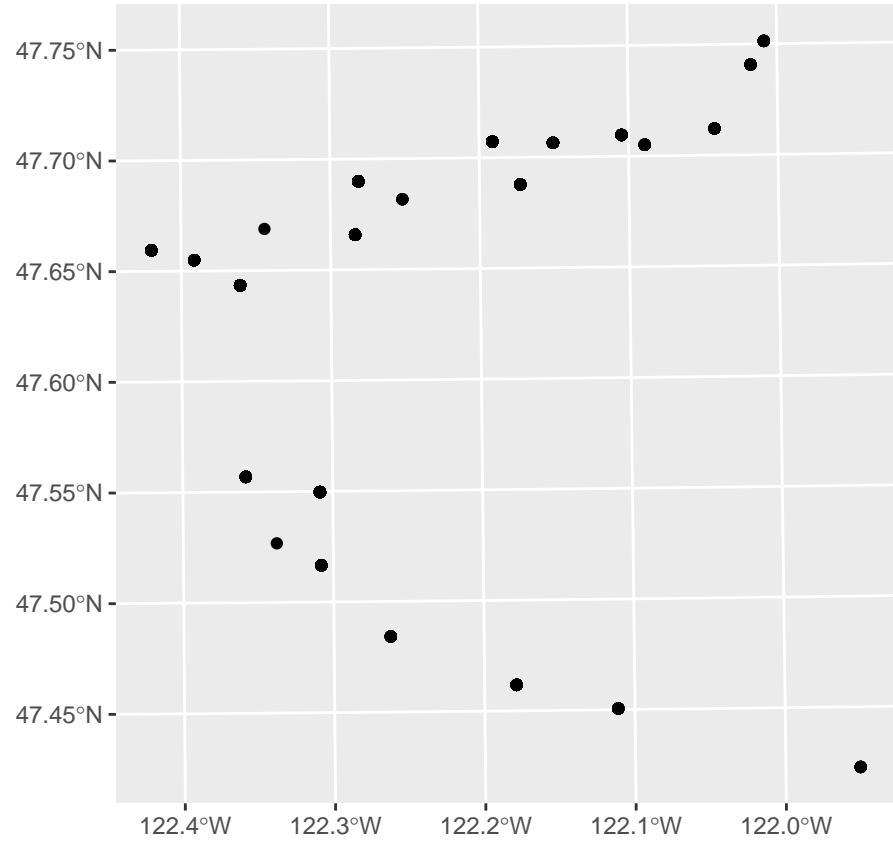
```

Let's observe the spatial distribution of the points by plotting them using the ggplot2 package. The geom_sf() function will allow us to plot the spatial data frame object.

```

library(ggplot2)
ggplot(captures.utm) + geom_sf()

```

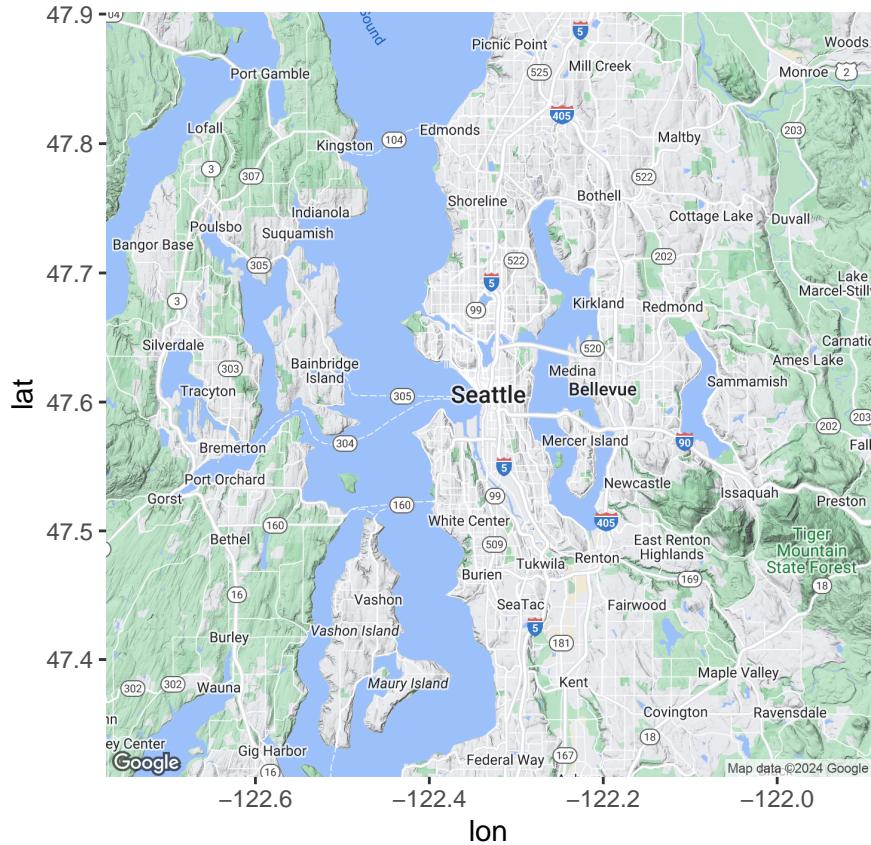


There is no basemap in this plot, we want to add a reference so we can easily distinguish between locations. We will use google maps for this, first we load the ggmap package and register an api from google. <>(Use an API key for the ‘uwin-mapping’ project that I created for this. <>(Describe in the Rmd how to get your own setup API key to use)

```
library(ggmap)
my_api <- 'AIzaSyBt73bzxdv1S6ioit4OTCaIE6SrZJ9aWnA'
register_google(key = my_api)
```

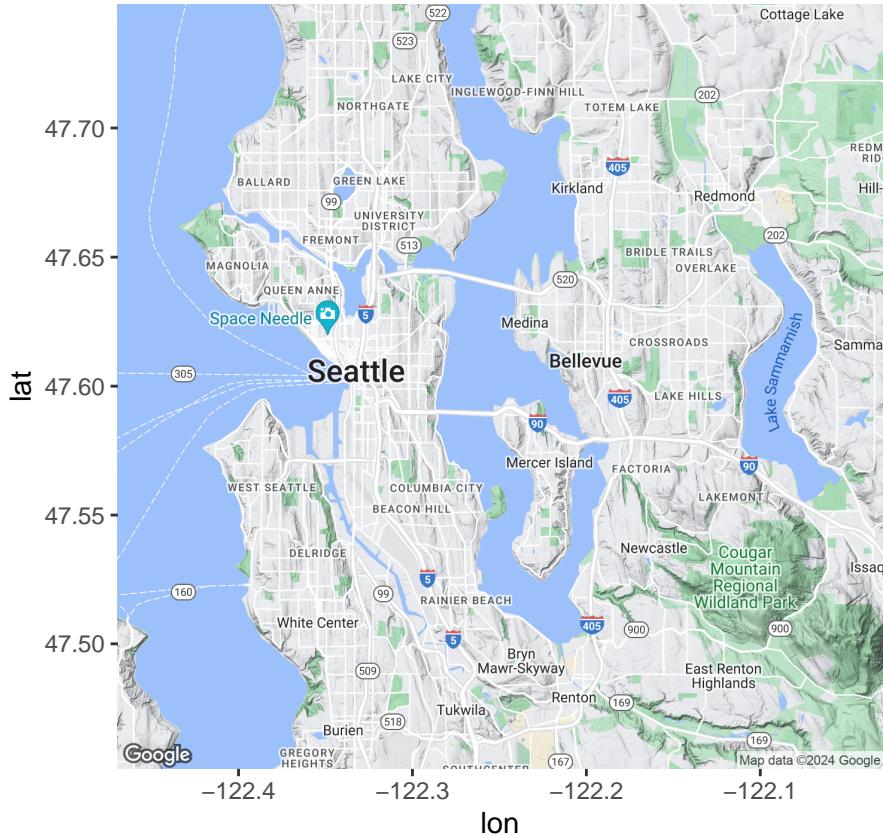
We then get the map relevant to our region using the get_map() function. This can be done both using a bounding box with coordinate information if we want a specific study area, or just the city’s name.

```
seattle <- get_map("seattle", source= "google", api_key = my_api)
ggmap(seattle)
```



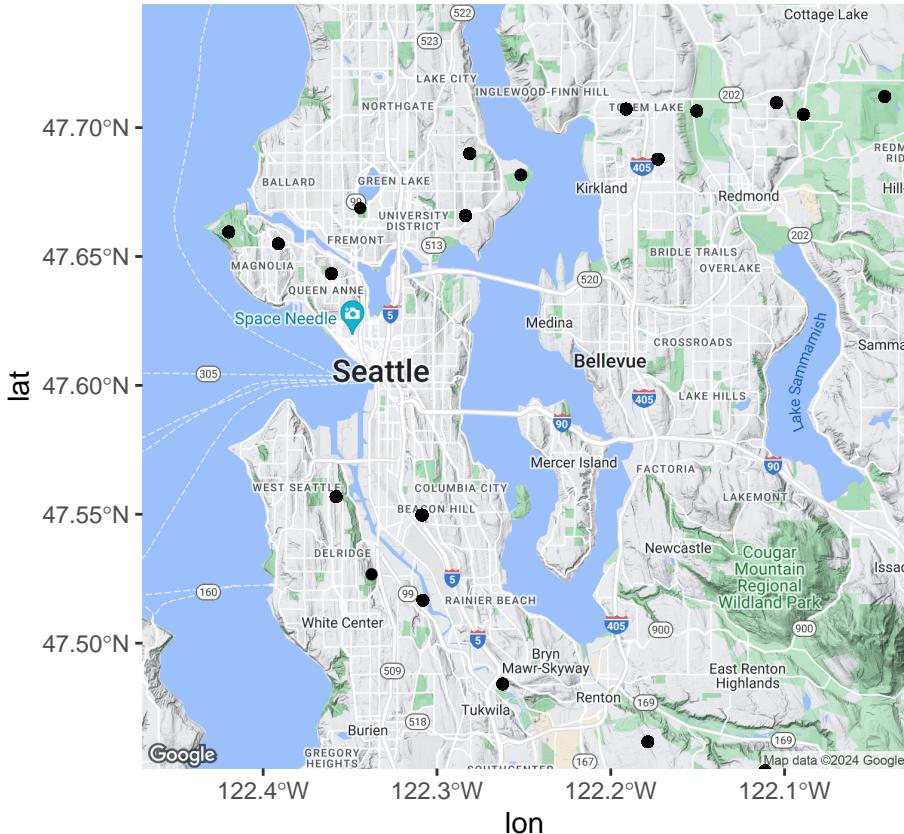
If we use a bounding box, the code will look like this:

```
seattle <- get_map(location = c(left = -122.5, bottom = 47.4,
                               right = -122.0, top = 47.8),
                     source = "google", api_key = my_api)
ggmap(seattle)
```



Now we can plot our camera site locations on the Seattle map <>(note the original crs works with google, not the utm.)

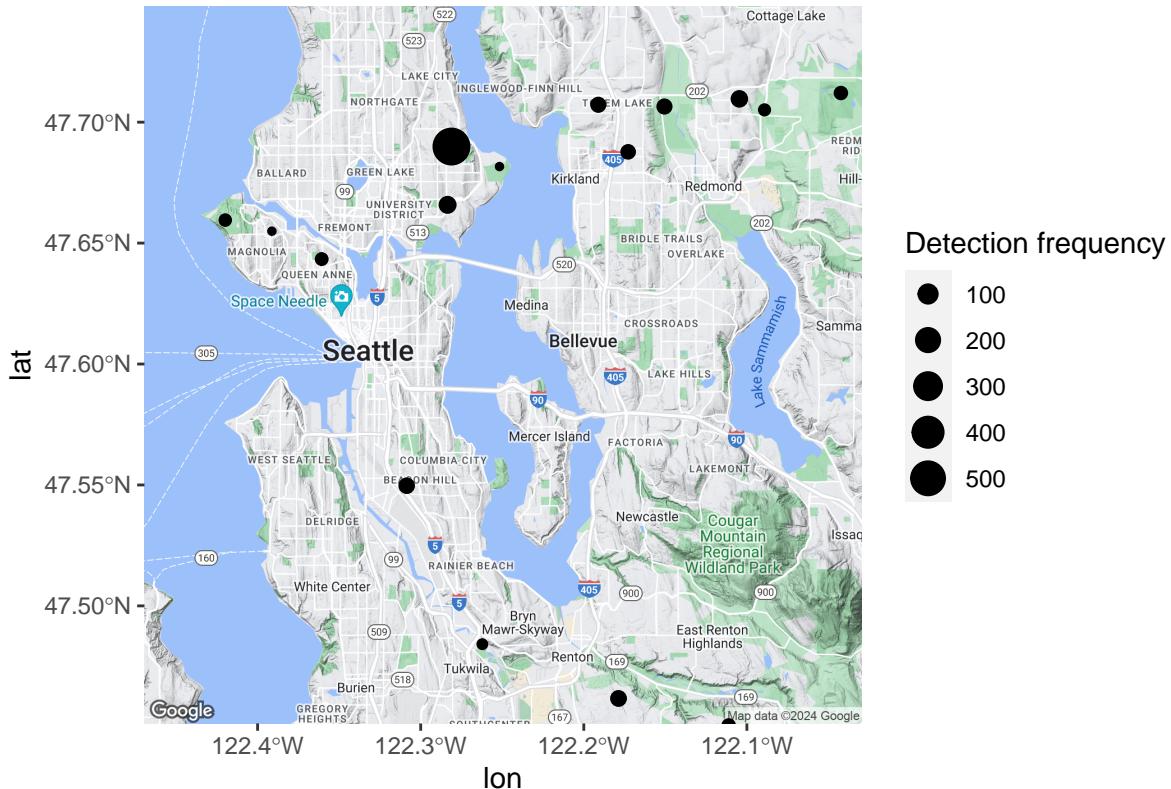
```
ggmap(seattle) +
  geom_sf(data=captures.spatial, inherit.aes = FALSE)
```



Now lets plot on a map the coyotes captured at each the camera trap sites. We will filter the data based on species name, using the dplyr package, and count detections at each site. We will then plot using the function seen above, but setting point size based on the number of detections at each site.

```
library(dplyr)
coyotes <- filter(captures.spatial, speciesname == "Canis latrans") %>%
  group_by(locationid) %>%
  summarize(detections = n())
ggmap(seattle) +
  geom_sf(data = coyotes, inherit.aes = FALSE, aes(size = detections)) +
  ggtitle("Coyote detections") +
  labs(size = "Detection frequency") +
  scale_size_continuous(breaks=seq(100, 500, by=100))
```

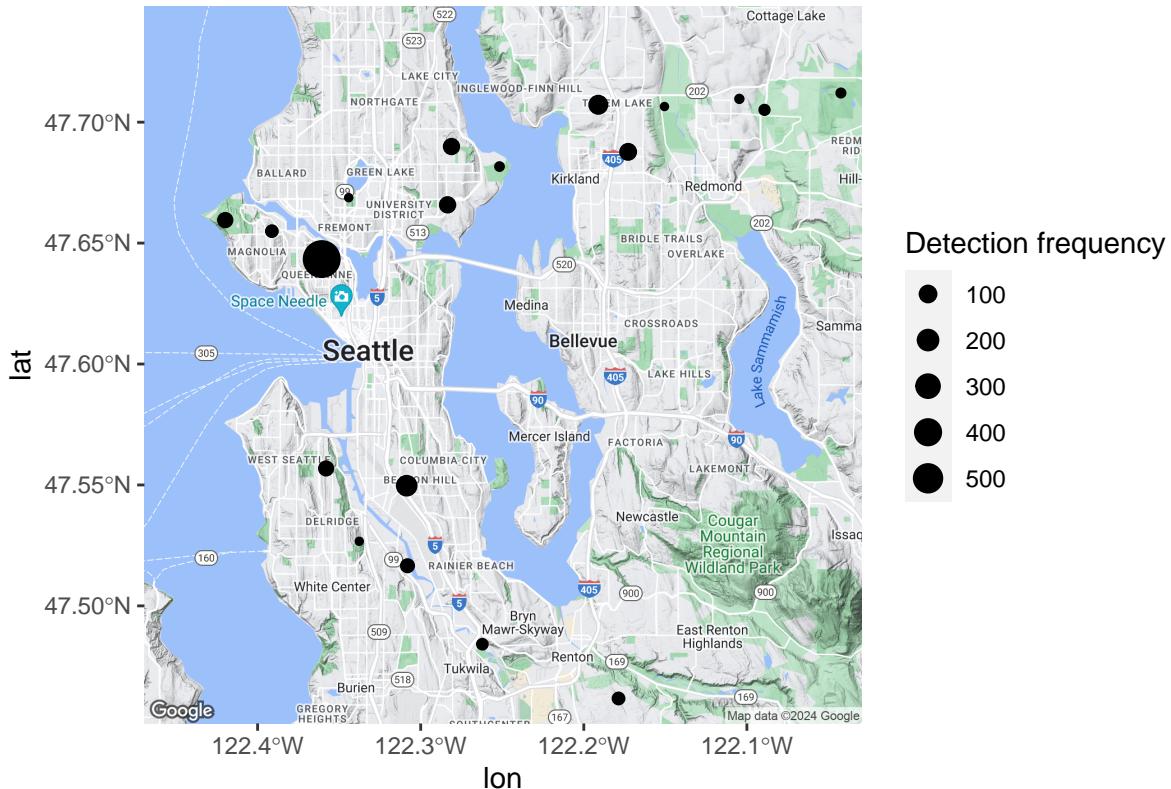
Coyote detections



Now try to do the same for raccoons. <>(we can hide code below using echo=FALSE, but depends on how we knit this document, maybe not a good idea for the pdf)

```
raccoons <- filter(captures.spatial, speciesname == "Procyon lotor") %>%
  group_by(locationid) %>%
  summarize(detections = n())
ggmap(seattle) +
  geom_sf(data = raccoons, inherit.aes = FALSE, aes(size = detections)) +
  ggtitle("Coyote detections") +
  labs(size = "Detection frequency") +
  scale_size_continuous(breaks=seq(100, 500, by=100))
```

Coyote detections



Vector data: lines

We will look into vector data in the form of lines using the TIGER database for Washington, composed of primary and secondary roads. The spatial object will be read in the same way as we did the points, but in this case we will load directly the shapefile containing the features, downloaded from here

The dataset contains 6 attributes (fields) for each feature.

```
roads <- st_read("roads/tl_2019_53_prisecroads.shp")
```

```
## Reading layer 'tl_2019_53_prisecroads' from data source
##   'C:\Users\tizge\Documents\Spatial Analysis in R Workshop\spatial_analysis_workshop_UWIN\roads\tl_2019_53_prisecroads.shp'
##   using driver 'ESRI Shapefile'
## Simple feature collection with 2912 features and 4 fields
## Geometry type: LINESTRING
## Dimension:     XY
## Bounding box:  xmin: -124.6384 ymin: 45.55945 xmax: -117.0359 ymax: 49.00241
## Geodetic CRS:  NAD83
```

```
print(head(roads))
```

```
## Simple feature collection with 6 features and 4 fields
## Geometry type: LINESTRING
## Dimension:     XY
```

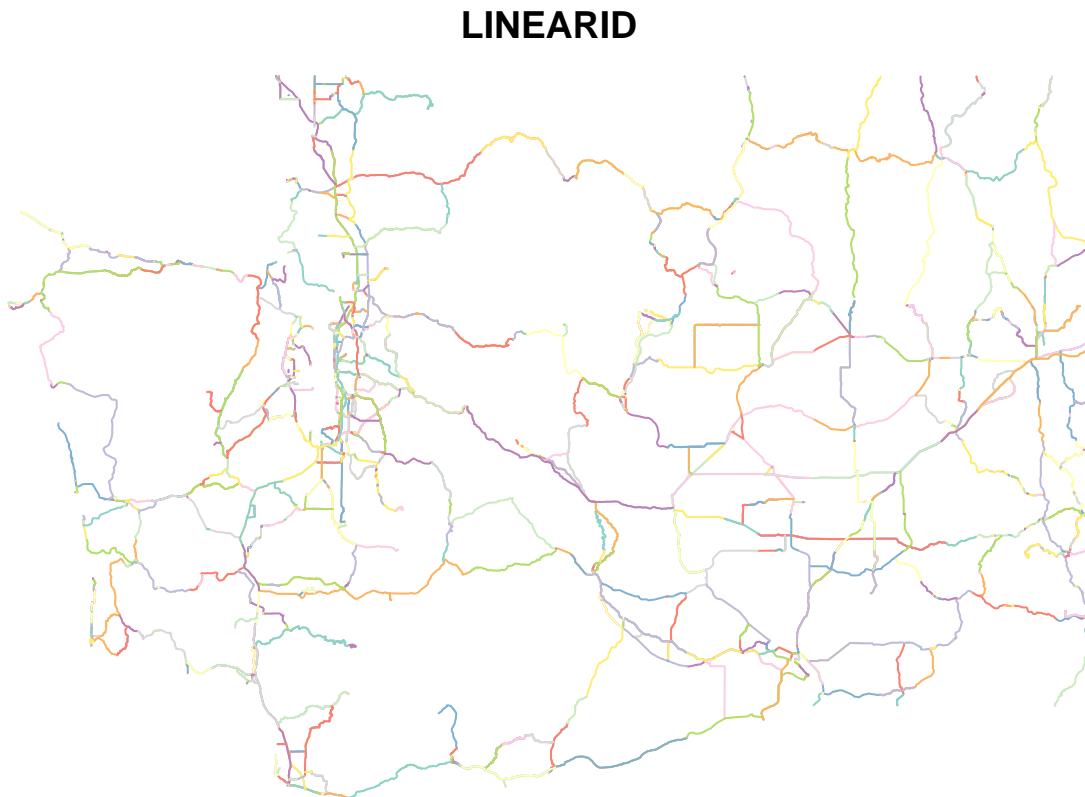
```

## Bounding box: xmin: -124.0058 ymin: 46.32205 xmax: -119.9736 ymax: 47.845
## Geodetic CRS: NAD83
##   LINEARID      FULLNAME RTTYP MTFCC           geometry
## 1 110569183105 State Rte 109 Byp      S S1200 LINESTRING (-123.9242 46.98...
## 2 1102219128828    US Hwy 97 Alt     U S1200 LINESTRING (-120.0899 47.83...
## 3 1102219128829    US Hwy 97 Alt     U S1200 LINESTRING (-120.3242 47.47...
## 4 1105007792960    US Hwy 101 Alt    U S1200 LINESTRING (-124.0055 46.33...
## 5 1103730584015    US Hwy 97 Alt     U S1200 LINESTRING (-119.9736 47.84...
## 6 1105008001942    US Hwy 101 Alt    U S1200 LINESTRING (-124.0058 46.32...

```

Let's plot the dataset to see how it looks, we will only plot one of the attributes, otherwise it will plot one map for each attribute

```
plot(roads[,1])
```



Again, this dataset can be converted to a data frame, this is useful when dealing with large vector data that may be slow to manage.

```

roads.df <- as.data.frame(roads)
print(head(roads.df))

```

```

##   LINEARID      FULLNAME RTTYP MTFCC           geometry
## 1 110569183105 State Rte 109 Byp      S S1200 LINESTRING (-123.9242 46.98...
## 2 1102219128828    US Hwy 97 Alt     U S1200 LINESTRING (-120.0899 47.83...
## 3 1102219128829    US Hwy 97 Alt     U S1200 LINESTRING (-120.3242 47.47...

```

```

## 4 1105007792960    US Hwy 101 Alt      U S1200 LINESTRING (-124.0055 46.33...
## 5 1103730584015    US Hwy 97 Alt      U S1200 LINESTRING (-119.9736 47.84...
## 6 1105008001942    US Hwy 101 Alt      U S1200 LINESTRING (-124.0058 46.32...

```

We can estimate the length of these roads, which will come in handy when estimating road density in a certain area. First we will transform to a distance-friendly projection, and then I will use the `st_length()` function from the `sf` package to estimate the length of each road.

```

roads <- st_transform(roads, crs="EPSG:5070")
roads$length <- sf::st_length(roads)
print(head(roads[,6]))

## Simple feature collection with 6 features and 1 field
## Geometry type: LINESTRING
## Dimension:      XY
## Bounding box:  xmin: -2130827 ymin: 2908853 xmax: -1789090 ymax: 2989999
## Projected CRS: NAD83 / Conus Albers
##           length      geometry
## 1  2948.97485 [m] LINESTRING (-2103609 297704...
## 2   35.07614 [m] LINESTRING (-1797725 298956...
## 3  48720.08804 [m] LINESTRING (-1825003 295515...
## 4   74.62914 [m] LINESTRING (-2130537 290974...
## 5  9526.23736 [m] LINESTRING (-1789090 298793...
## 6  934.32284 [m] LINESTRING (-2130827 290885...

```

The unit will be automatically in meters, we can convert to numeric if we dont want the unit directly in the column using `as.numeric()`

```

roads$length <- as.numeric(roads$length)
print(head(roads[,6]))

## Simple feature collection with 6 features and 1 field
## Geometry type: LINESTRING
## Dimension:      XY
## Bounding box:  xmin: -2130827 ymin: 2908853 xmax: -1789090 ymax: 2989999
## Projected CRS: NAD83 / Conus Albers
##           length      geometry
## 1  2948.97485 LINESTRING (-2103609 297704...
## 2   35.07614 LINESTRING (-1797725 298956...
## 3  48720.08804 LINESTRING (-1825003 295515...
## 4   74.62914 LINESTRING (-2130537 290974...
## 5  9526.23736 LINESTRING (-1789090 298793...
## 6  934.32284 LINESTRING (-2130827 290885...

```

I can estimate total length for each road type, or following any other attribute, for example all roads within a certain county, or any polygon, such as camera trap buffer area.

```

road_lengths <- aggregate(length ~ RTTYP, data=roads, FUN="sum")
print(road_lengths)

```

```

##   RTTYP      length

```

```

## 1      C  12589.73
## 2      I 2471012.92
## 3      M 5794528.48
## 4      O 11446.29
## 5      S 9159988.42
## 6      U 3811870.92

```

With this information I can estimate the road density of each road type within Washington state.

```

road_lengths$road_density <- ((road_lengths$length)/1e+6)/184827
print(head(road_lengths))

```

```

##   RTTYP      length road_density
## 1      C  12589.73 6.811629e-08
## 2      I 2471012.92 1.336933e-05
## 3      M 5794528.48 3.135109e-05
## 4      O 11446.29 6.192975e-08
## 5      S 9159988.42 4.955980e-05
## 6      U 3811870.92 2.062399e-05

```

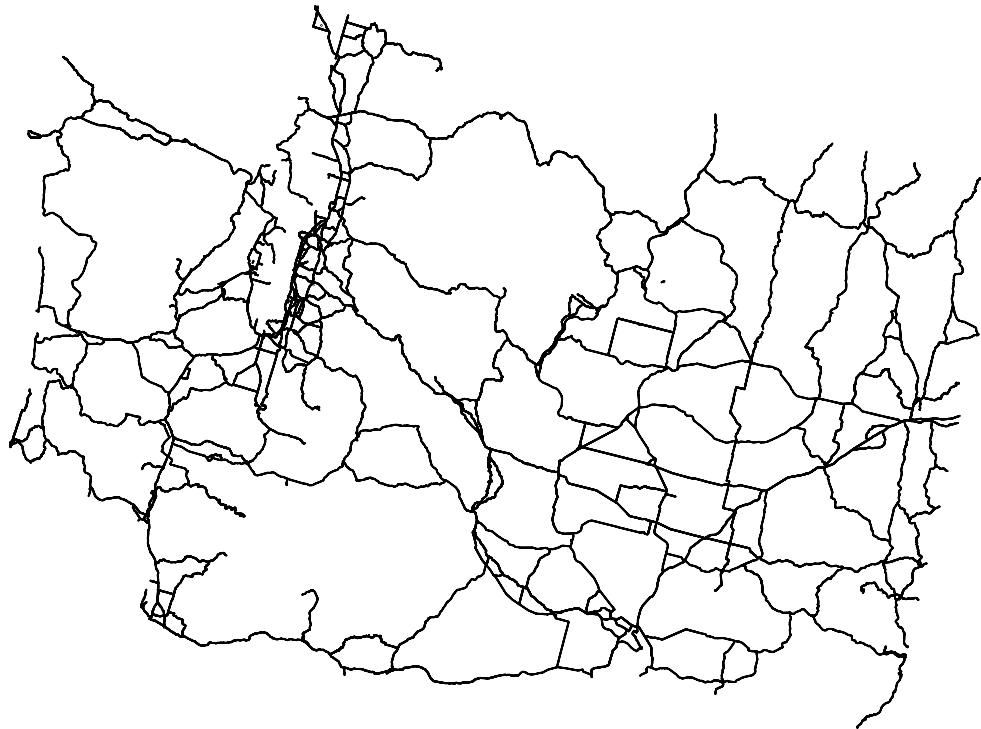
Sometimes it is useful to convert lines to polygons, for example when we want a better representation of the area a linear feature occupies. This might be good for connectivity analysis as road width might define crossing probability, or for considering impervious surface generated by roads. For this we use the `st_buffer()` function and decide a buffer size we will use for the linear feature expansion.

```

roads_p <- sf:::st_buffer(roads, dist=12) #buffer to 12 meters
plot(roads_p[,1])

```

LINEARID



```
print(head(roads_p)) #now our roads are a polygon map layer
```

```
## Simple feature collection with 6 features and 5 fields
## Geometry type: POLYGON
## Dimension: XY
## Bounding box: xmin: -2130839 ymin: 2908841 xmax: -1789078 ymax: 2990011
## Projected CRS: NAD83 / Conus Albers
##          LINEARID      FULLNAME RTTYP MTFCC           geometry
## 1  110569183105 State Rte 109 Byp      S S1200 POLYGON ((-2103573 2977164, ...
## 2  1102219128828    US Hwy 97 Alt      U S1200 POLYGON ((-1797762 2989559, ...
## 3  1102219128829    US Hwy 97 Alt      U S1200 POLYGON ((-1824999 2955186, ...
## 4  1105007792960    US Hwy 101 Alt      U S1200 POLYGON ((-2130535 2909763, ...
## 5  1103730584015    US Hwy 97 Alt      U S1200 POLYGON ((-1792234 2988143, ...
## 6  1105008001942    US Hwy 101 Alt      U S1200 POLYGON ((-2130824 2908883, ...
##          length
## 1  2948.97485
## 2   35.07614
## 3  48720.08804
## 4   74.62914
## 5  9526.23736
## 6   934.32284
```

Vector data: polygon