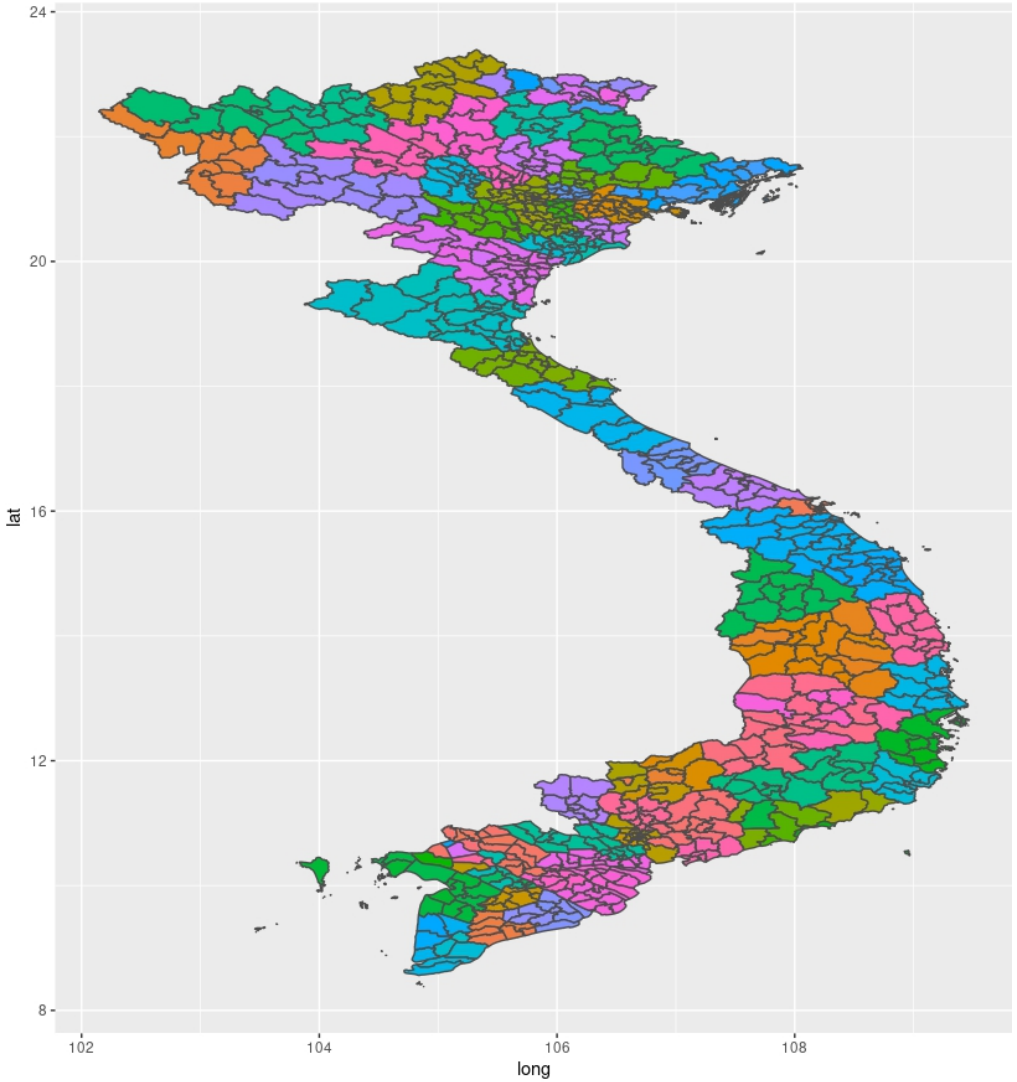
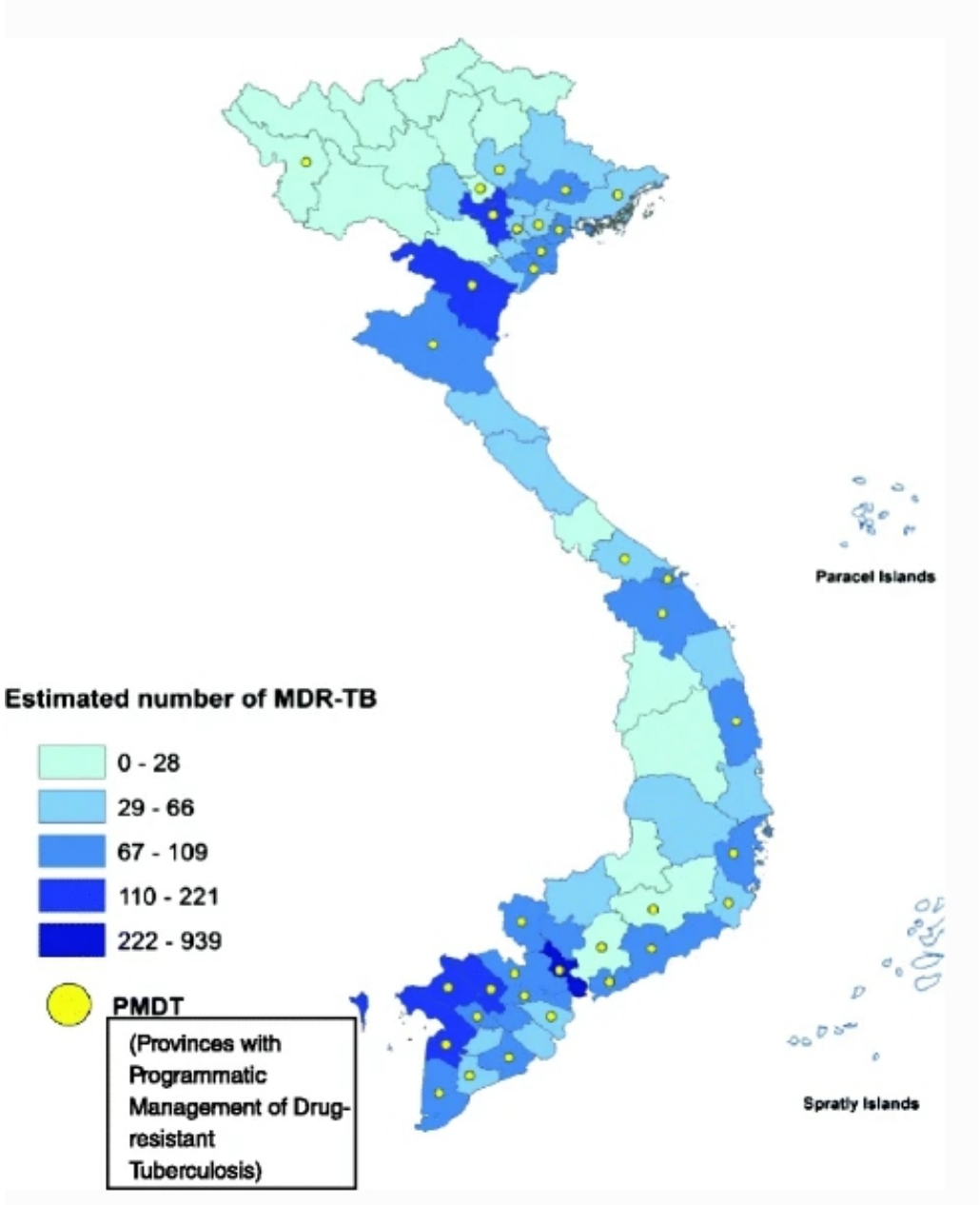
MAP AND LOCATION in R

DATA VISUALIZATION
june, 2020

Ho Thi Kim Cuong





https://link.springer.com/article/10.1186/s12889-015-2338-5

Outline

- Introduction to map data
- R packages
- World map
- Practice: Vietnam map

Introduction to map data

R is an great tool for geospatial data analysis. Heaps of dedicated packages exist. Building a map follows those 2 steps:

- Find data, load it in R: region boundaries can be stored in shapefiles or geoJSON files. Some R libraries also provide the data for the most common places.
- Manipulate and plot it: with packages like sp or ggplot2

https://r-graph-gallery.com/168-load-a-shape-file-into-r.html

https://r-graph-gallery.com/325-background-map-from-geojson-format-in-r.html

Introduction to map data

- Map visualization is used to analyze and display the geographically related data and present it in the form of maps. This kind of data expression is clearer and more intuitive. We can visually see the distribution or proportion of data in each region.
- Every loation is identified by 2 numbers: Longitude (x-axis) and latitude (y-axis)
- Polygon can be applied to visualize the location.

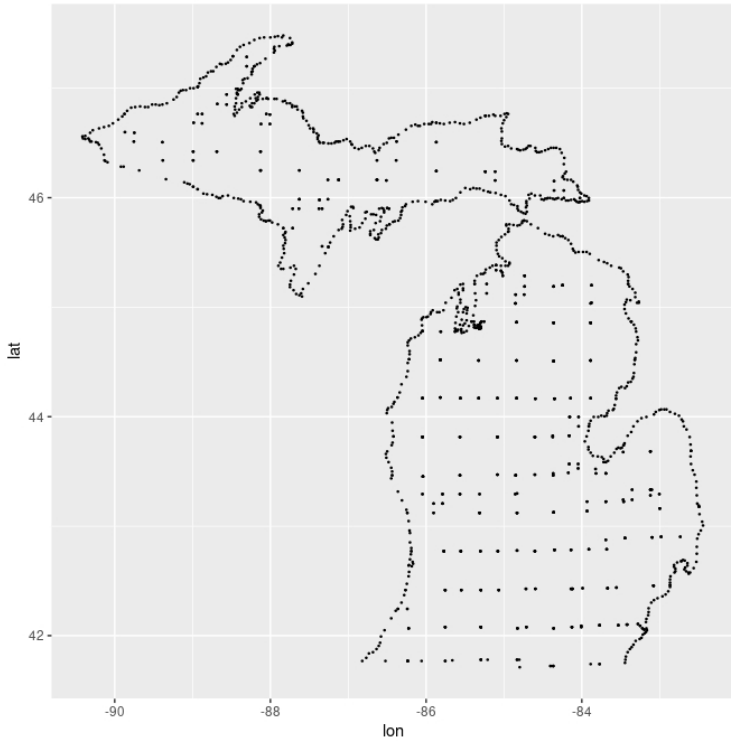
R packages

Packages	Function
sp	Defining spatial data classes
rgeos	Spatial data anipulation: area, perimeter, centroid, etc
rgdal	Basic functions (rarely uesd)
maps	Maping data
raster	Reading, writing, manipulating, analyzing and modeling of spatial data
ggplot2	Data visualization
viridis	Coloring data

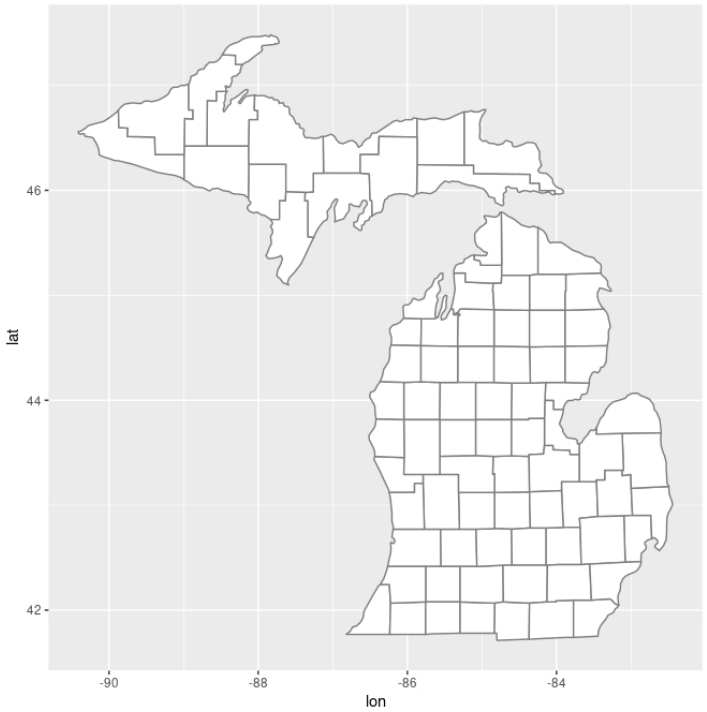
Polygon maps

- Perhaps the simplest approach to drawing maps is to use geom_polygon() to draw boundaries for different regions

```
> mi_counties <- map_data("county", "michigan")
> head(mi_counties)
       long      lat group order   region subregion
1 -83.88675 44.85686     1     1 michigan    alcona
2 -83.36536 44.86832     1     2 michigan    alcona
3 -83.36536 44.86832     1     3 michigan    alcona
4 -83.33098 44.83968     1     4 michigan    alcona
5 -83.30806 44.80530     1     5 michigan    alcona
6 -83.30233 44.77665     1     6 michigan    alcona
>
```



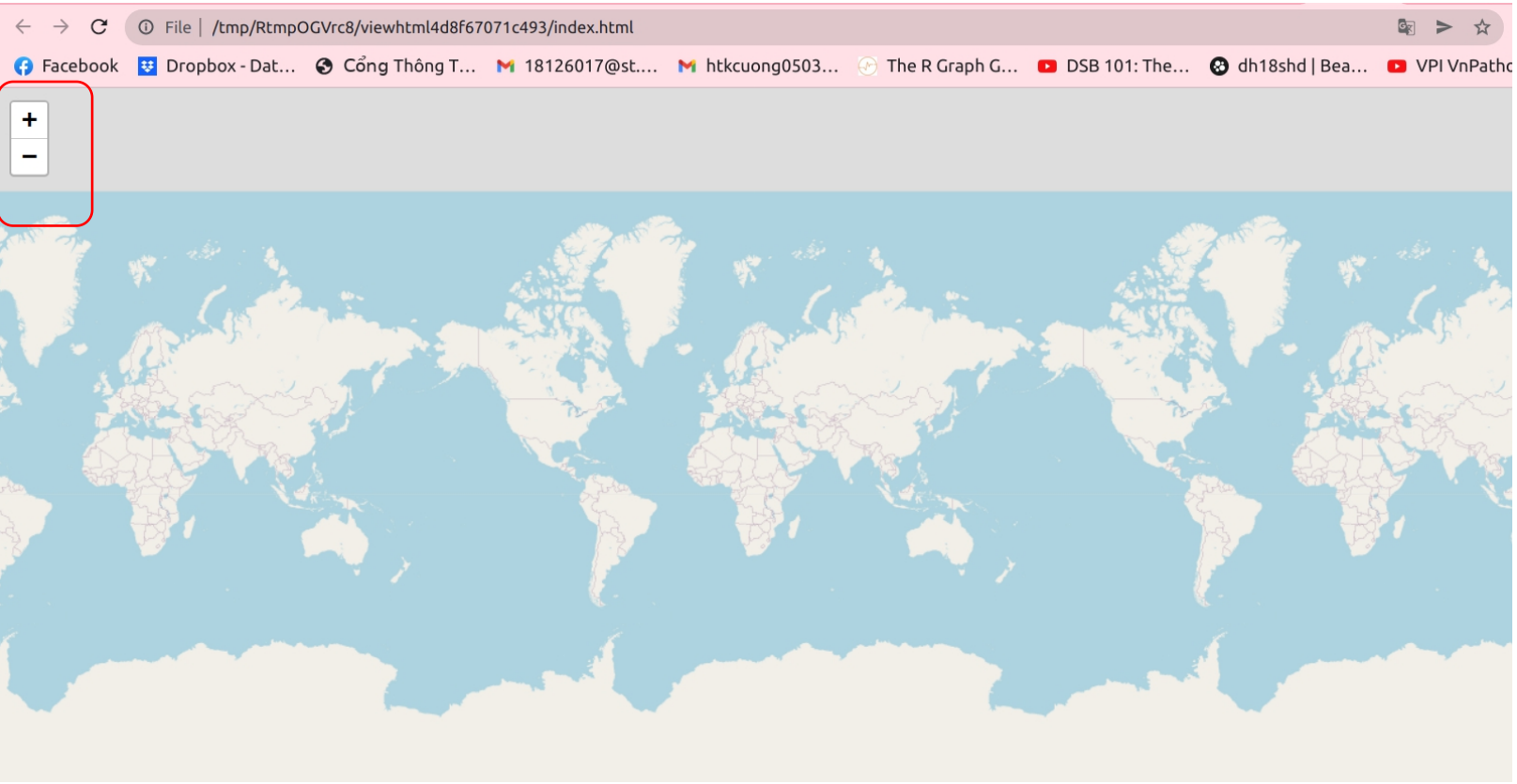
geom_point

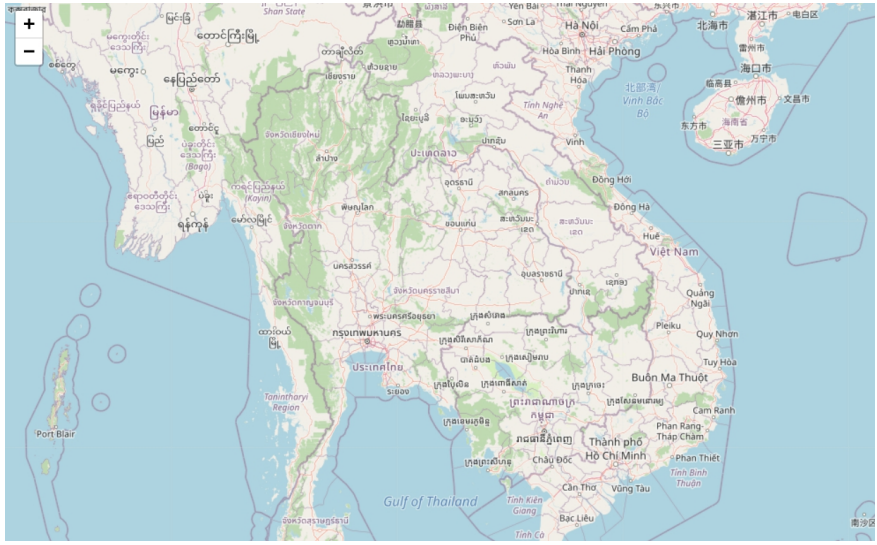


geom_polygon

THE LEAFLET PACKAGE FOR INTERACTIVE MAPS

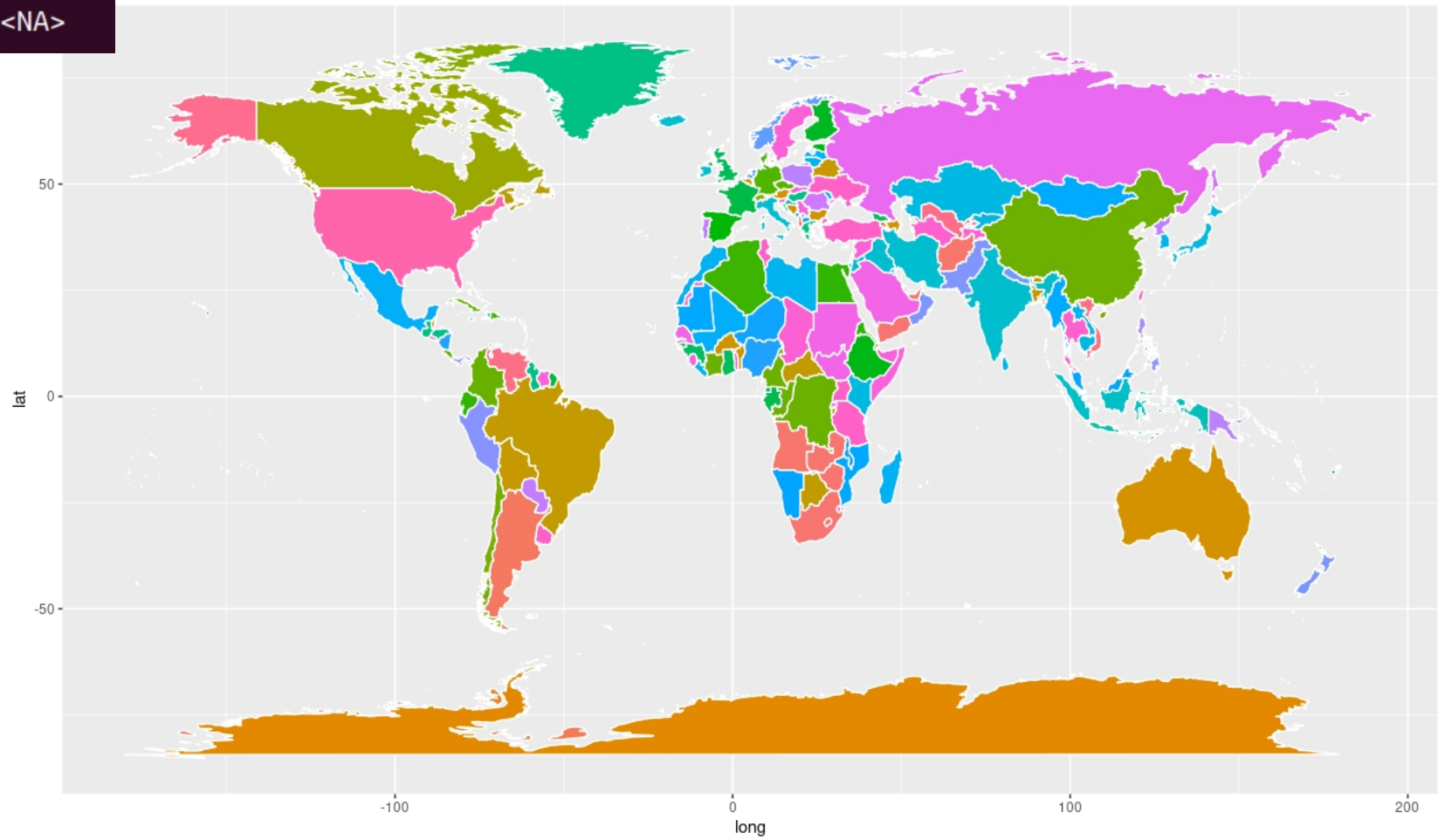
- library(leaflet)
- First initiate the map with the leaflet() function.
- Then add tiles with addTiles()
- By default, you get the map beside. See next charts to learn how to zoom on a zone, change background style.





World map

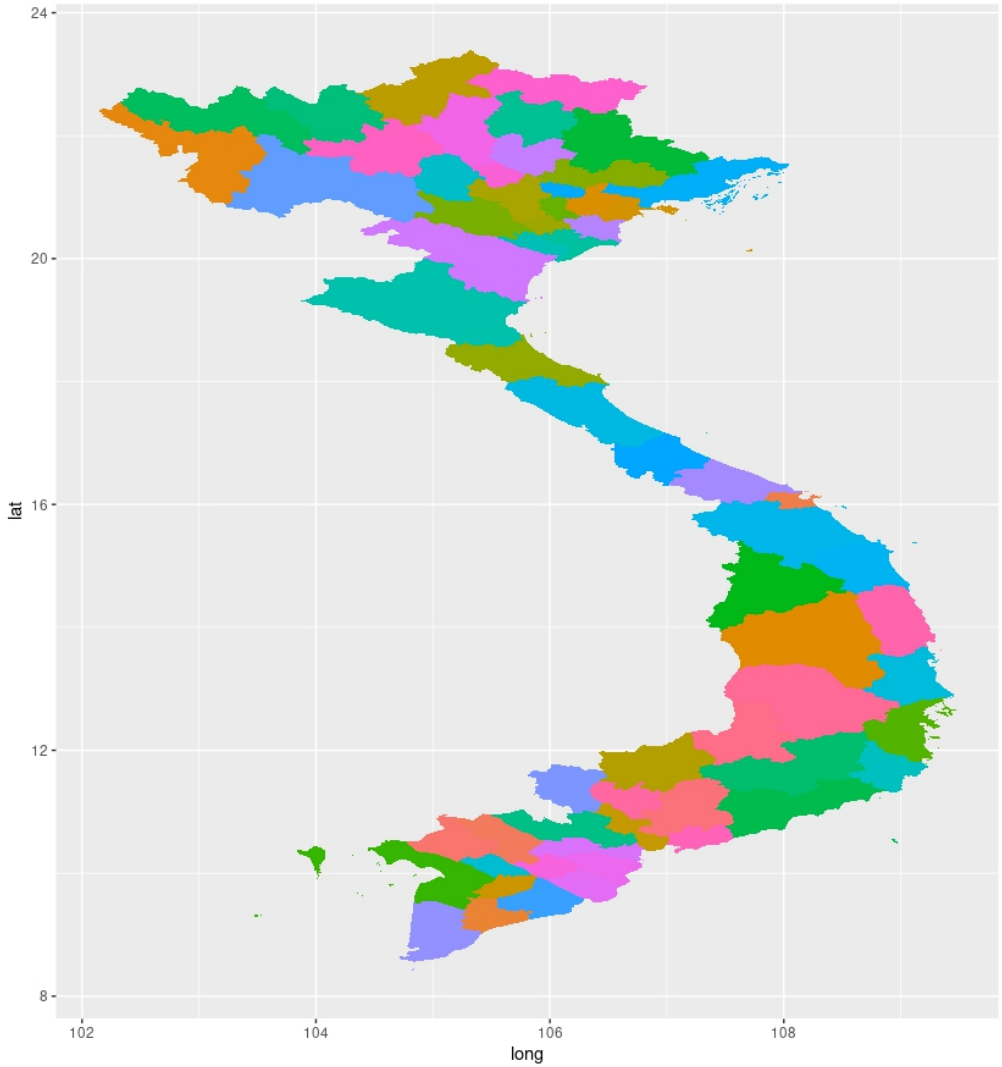
```
> library(ggplot2)
> worldmap <- map_data("world")
> head(worldmap)
      long      lat group order region subregion
1 -69.89912 12.45200     1     1  Aruba      <NA>
2 -69.89571 12.42300     1     2  Aruba      <NA>
3 -69.94219 12.43853     1     3  Aruba      <NA>
4 -70.00415 12.50049     1     4  Aruba      <NA>
5 -70.06612 12.54697     1     5  Aruba      <NA>
6 -70.05088 12.59707     1     6  Aruba      <NA>
```



Vietnam map - level 1 (province)

```
> vnm <- getData("GADM",country="Vietnam", level=1) #level=1:province
>
> head(vnm)
   GID_0  NAME_0   GID_1    NAME_1 VARNAME_1 NL_NAME_1 TYPE_1 ENGTYPE_1 CC_1
1    VNM Vietnam VNM.1_1  An Giang  An Giang       <NA>   Tỉnh  Province <NA>
12   VNM Vietnam VNM.2_1  Bạc Liêu  Bac Lieu       <NA>   Tỉnh  Province <NA>
23   VNM Vietnam VNM.3_1 Bắc Giang Bac Giang       <NA>   Tỉnh  Province <NA>
34   VNM Vietnam VNM.4_1   Bắc Kạn   Bac Kan       <NA>   Tỉnh  Province <NA>
45   VNM Vietnam VNM.5_1  Bắc Ninh  Bac Ninh       <NA>   Tỉnh  Province <NA>
56   VNM Vietnam VNM.6_1   Bến Tre   Ben Tre       <NA>   Tỉnh  Province <NA>
```

```
> head(vn)
      long      lat order  hole piece id group
1 105.3745 10.24604     1 FALSE     1  1   1.1
2 105.3362 10.23442     2 FALSE     1  1   1.1
3 105.3154 10.26842     3 FALSE     1  1   1.1
4 105.3120 10.27393     4 FALSE     1  1   1.1
5 105.3065 10.26894     5 FALSE     1  1   1.1
6 105.3013 10.27698     6 FALSE     1  1   1.1
>
```

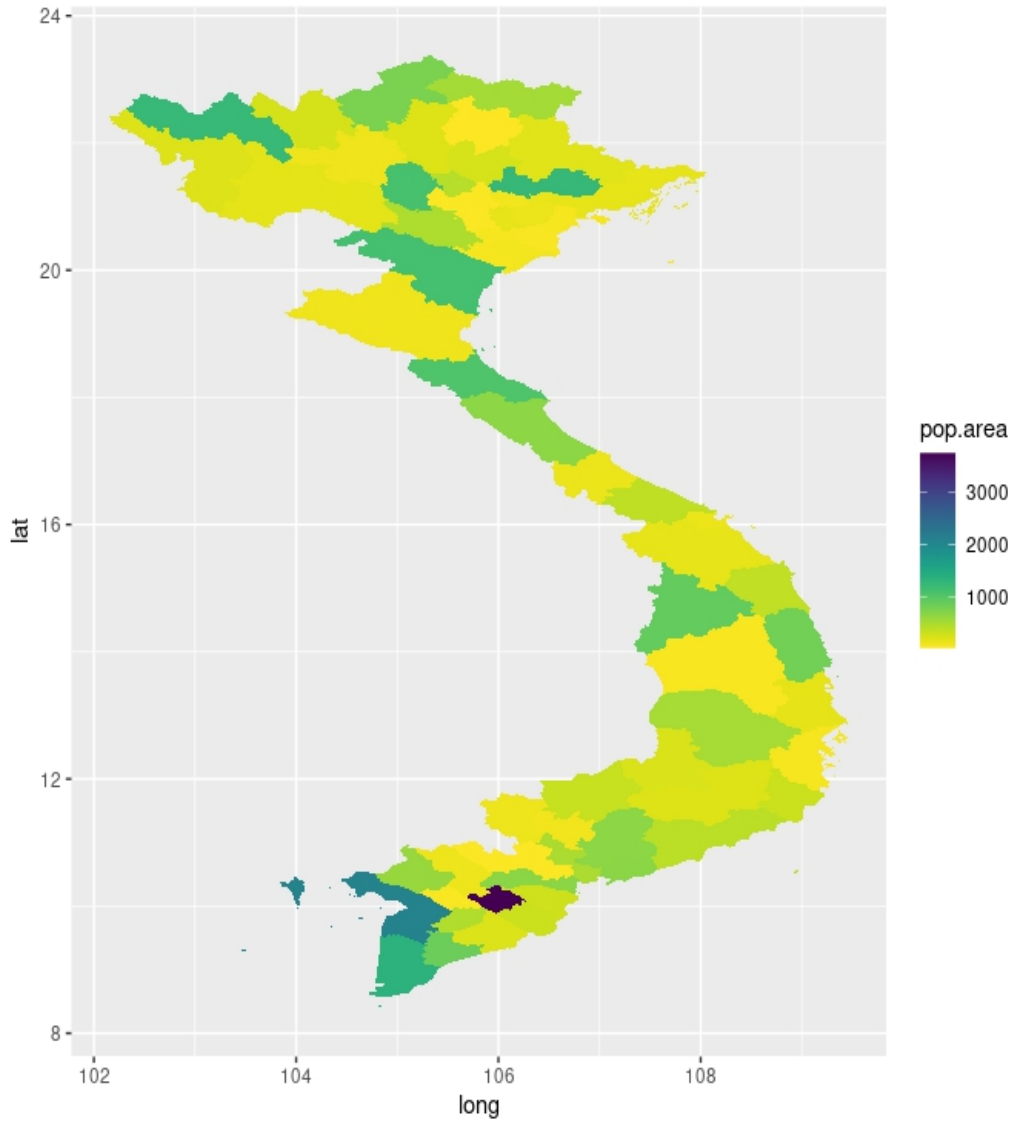


Vietnam map - with popolation data

```
> head(vn)
      long      lat order  hole piece id group
1 105.3745 10.24604     1 FALSE     1  1   1.1
2 105.3362 10.23442     2 FALSE     1  1   1.1
3 105.3154 10.26842     3 FALSE     1  1   1.1
4 105.3120 10.27393     4 FALSE     1  1   1.1
5 105.3065 10.26894     5 FALSE     1  1   1.1
6 105.3013 10.27698     6 FALSE     1  1   1.1
```

```
> head(pop)
# A tibble: 6 × 6
  province     id value  area   pop pop.area
  <chr>     <dbl> <dbl> <dbl> <dbl>    <dbl>
1 An Giang      1  29.8 3537. 2155.      609
2 Bạc Liêu      2  57.4 2469.  877.      355
3 Bắc Giang     3  53.1 3850. 1593.      414
4 Bắc Kạn       4  24.0 4859.  303.       62
5 Bắc Ninh      5  24.0  823. 1114      1354
6 Bến Tre       6  46.2 2360. 1262       535
>
```

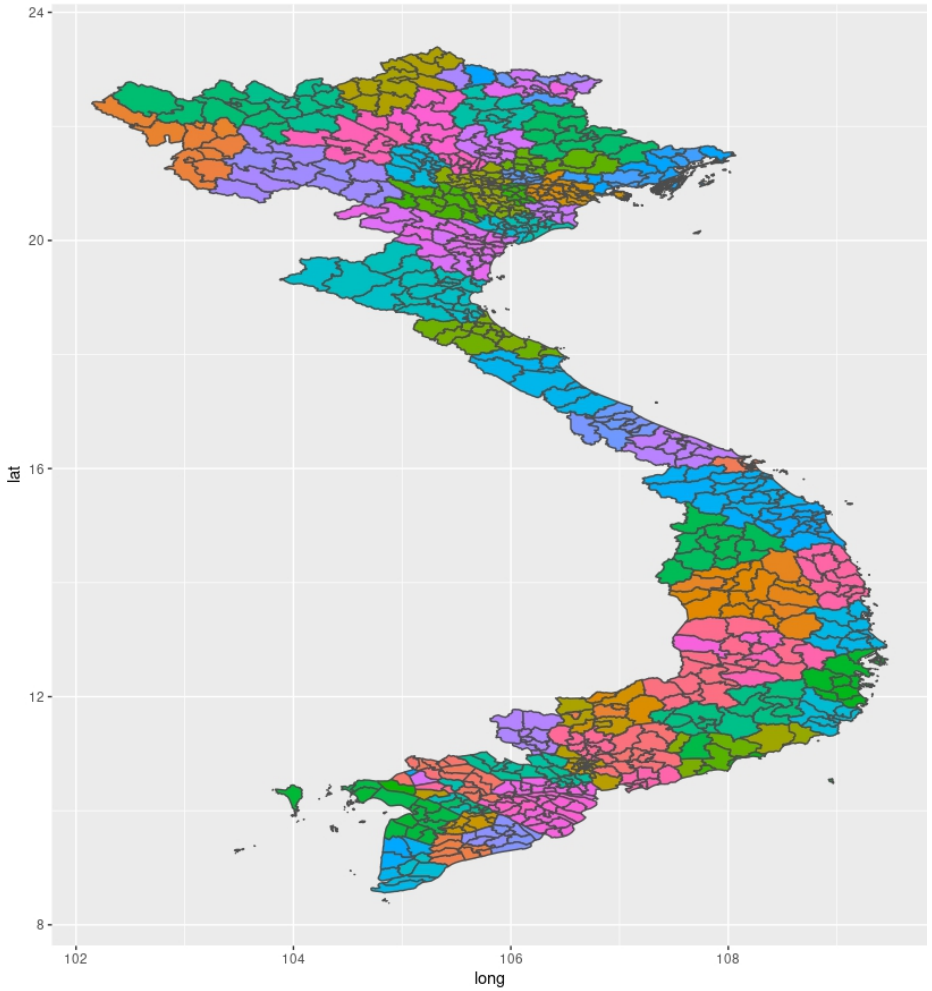
```
vnn <- merge(vn, pop ,by="id")
head (vnn)
id     long      lat order  hole piece group province    value   area    pop
 1 105.1566 10.82866    99 FALSE     1   1.1 An Giang 29.75787 3536.7 2155.3
 1 105.1395 10.77125   102 FALSE     1   1.1 An Giang 29.75787 3536.7 2155.3
 1 105.0755 10.77720    12 FALSE     1   1.1 An Giang 29.75787 3536.7 2155.3
 1 105.1583 10.80932   100 FALSE     1   1.1 An Giang 29.75787 3536.7 2155.3
 1 105.1601 10.78144   101 FALSE     1   1.1 An Giang 29.75787 3536.7 2155.3
 1 105.0545 10.82912    27 FALSE     1   1.1 An Giang 29.75787 3536.7 2155.3
pop.area
     609
     609
     609
     609
     609
     609
head(pop)
```



Vietnam map - level 2 (city/town)

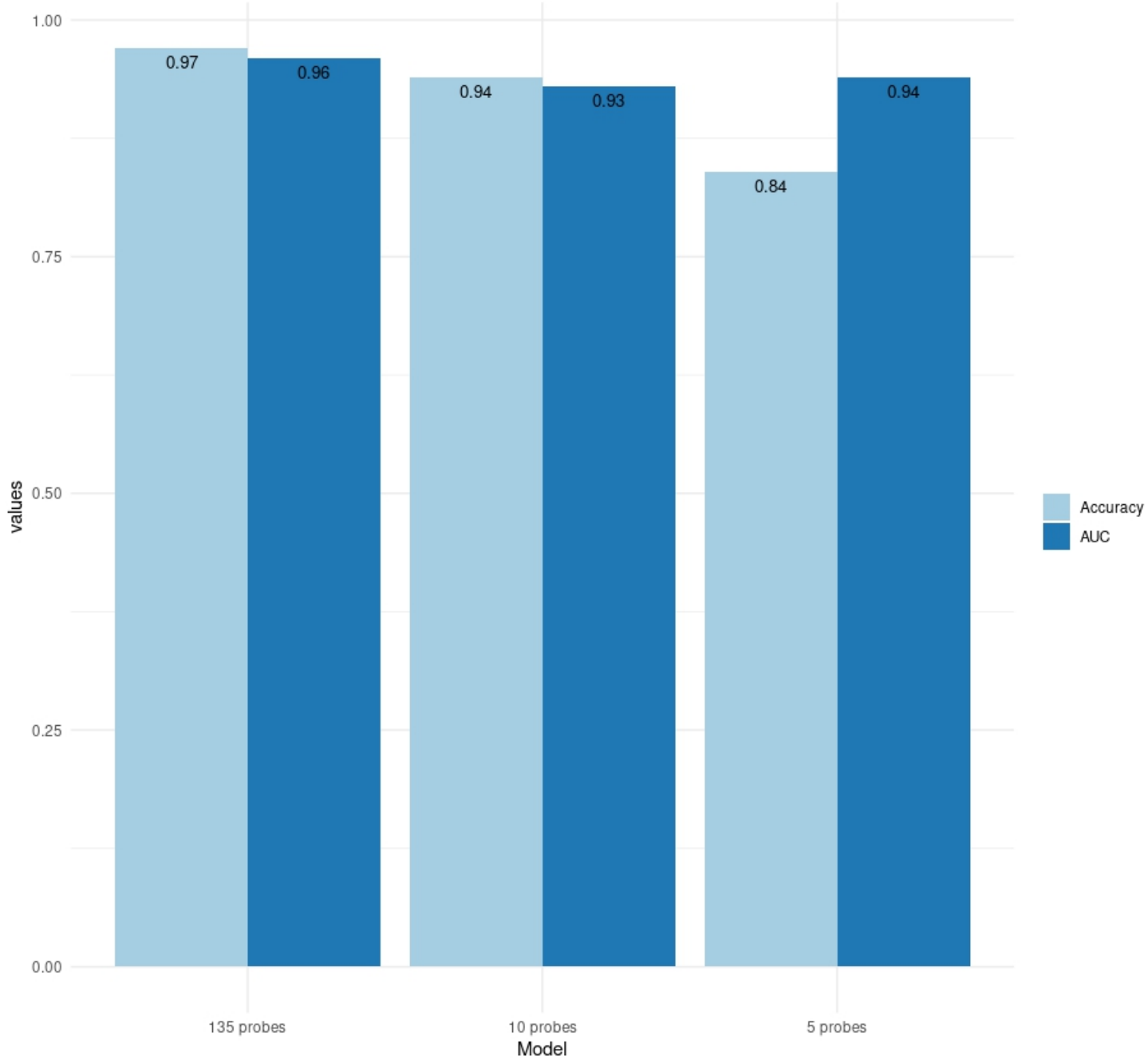
```
> vnm <- getData("GADM",country="Vietnam", level=2) #level=2:city/town
> head(vnm)
  GID_0  NAME_0   GID_1   NAME_1 NL_NAME_1     GID_2     NAME_2  VARNAME_2
1   VNM Vietnam VNM.1_1 An Giang       <NA> VNM.1.1_1     An Phú     An Phu
4   VNM Vietnam VNM.1_1 An Giang       <NA> VNM.1.2_1    Chợ Mới    Cho Moi
5   VNM Vietnam VNM.1_1 An Giang       <NA> VNM.1.3_1   Châu Đốc   Chau Doc
6   VNM Vietnam VNM.1_1 An Giang       <NA> VNM.1.4_1   Châu Phú   Chau Phu
7   VNM Vietnam VNM.1_1 An Giang       <NA> VNM.1.5_1 Châu Thành Chau Thanh
8   VNM Vietnam VNM.1_1 An Giang       <NA> VNM.1.6_1 Long Xuyên Long Xuyen
```

```
      long      lat order  hole piece id group
1 105.1216 10.71159     1 FALSE     1  1   1.1
2 105.1147 10.71974     2 FALSE     1  1   1.1
3 105.1012 10.74415     3 FALSE     1  1   1.1
4 105.0979 10.76532     4 FALSE     1  1   1.1
5 105.0883 10.76043     5 FALSE     1  1   1.1
6 105.0882 10.76061     6 FALSE     1  1   1.1
```



Conclusion

- R is a powerful program for map visualization
- Key variables: longitude and latitude (packages raster)
- packages for ploting: ggplot2


1.00
0.75
0.50
0.25
0.00
values
0.97
0.96
0.94
0.93
0.84
0.94
135 probes
10 probes
5 probes
Model
Accuracy
AUC