Rlab: continuous spatial index

Kristine Dinh

I. Gathering and exploring the data

1. [2 pts] What type of R object is `sd_city`? What coordinate reference system is used?

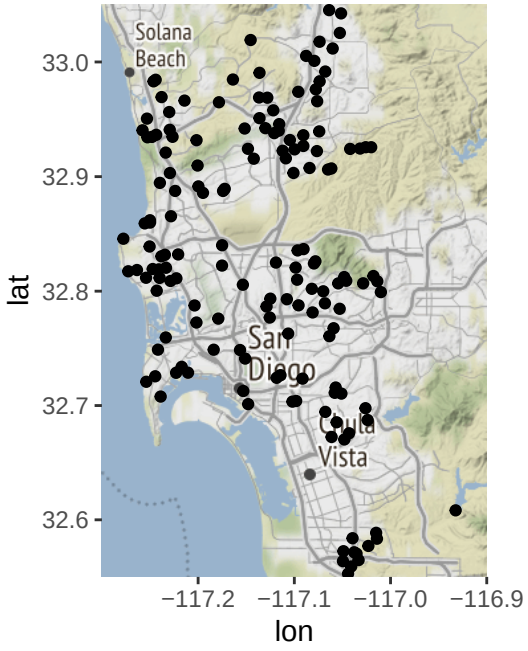
- `sd_city` is SpatialPolygonsDataFrame
- Coordinate reference system:

```
## CRS arguments:
##  +proj=lcc +lat_1=33.88333333333333 +lat_2=32.78333333333333
## +lat_0=32.16666666666666 +lon_0=-116.25 +x_0=2000000.000101601
## +y_0=500000.0001016002 +datum=NAD83 +units=us-ft +no_defs +ellps=GRS80
## +towgs84=0,0,0
```

2. [1 pts] What does line 7 do?

- Transform for map projection and datum transformation. `sites_11` is the object to be transfer. And `proj4string(sd_city)` is the coordinate references CRS object.

3. [4 pts] Make a plot that shows: (i) the locations of the measurement sites, (ii) the boundary of the city, and (iii) basic contextual features of San Diego, such as major roadways.



4. [2 pts] There are 9 NAs in `sites$cl2_total`. Why?

- Because there're 9 sites that exist in `sites_ll$site_ID` but not `water_testing$source`

5. [2 pts] Write your own code that creates a new variable called `sites_sub` that is just like `sites`, but has the 9 sample sites with `NA` values for `cl2_total` removed.

```
sites_sub <- sites[!is.na(sites$cl2_total),]
```

6. [1 pts] What is the coordinate reference system for the tract data?

```
## CRS arguments:
##  +proj=lcc +lat_1=33.88333333333333 +lat_2=32.78333333333333
## +lat_0=32.16666666666666 +lon_0=-116.25 +x_0=2000000.000101601
## +y_0=500000.0001016002 +datum=NAD83 +units=us-ft +no_defs +ellps=GRS80
## +towgs84=0,0,0
```

7. [1 pts] What is the name of the census tract with the lowest median household income? Highest?

- Name of census tract with lowest median income: 33.05
- Name of census tract with highest median income: 215

8. [4 pts] Provide executable code that adds a new column to `sites_sub` called `median_income` with the appropriate values of median household income. Check that your first few values match the ones below.

```
sites_sub %<>% spTransform(., proj4string(sd_tracts))

sites_sub_new <- over(sites_sub, sd_tracts)

sites_sub@data %<>%
  cbind(median_income = sites_sub_new$median_income)

sites_sub$median_income %>% head
```

```
## [1]  72992 166383  72992  45329  52080  47043
```

9. [4 pts] Fit a non-spatial linear model using `lm()` with median household income as the sole predictor of chlorine levels. Give a point estimate and 95% confidence interval for the effect of median household income on chlorine levels. How does the estimate compare to your expectations?

- CI and p value shows that median income has an effect on chlorine level. I wouldn't expect this.

term	estimate	std.error	statistic	p.value	2.5 %	97.5 %
(Intercept)	2.4136060	0.1201074	20.095392	0.00e+00	2.1762322	2.6509798
median_income	-0.0000045	0.0000011	-4.228018	4.14e-05	-0.0000066	-0.0000024

II. Spatial dependence + interpolation

12. [3 pts] Use the krige function to perform universal kriging of chlorine levels for all locations in `sd_grid`, taking into account the effect of household income. Use whichever variogram model you think fit best. Provide the code you used to make your predictions, and report estimates of parameters in the variogram model.

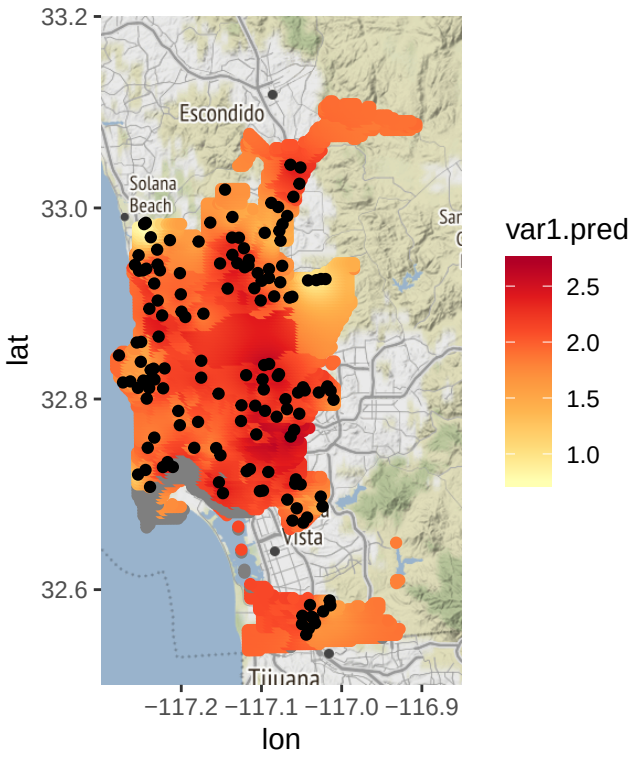
- Table of estiamtes of parameters in variogram model

model	psill	range	kappa
Nug	0.0857314	0.00	0.0
Mat	0.1626253	10002.13	0.5

- Universal kriging

```
fit.k <- krige(formula = cl2_total ~ median_income, locations = sites_sub,
               newdata = sd_grid, model = fit.v)
```

13. [3 pts] Make a map of the predictions, again with contextual information included (hint: if you use ggmap(), you'll need to transform to longitude/latitude coordinates.)



14. [3 pts] Below is a map I made based on my model. How does it compare to yours? What does your map show that mine doesn't?

- Both map looks similiar. However, my map has the emperical points on top and has color.

15. [2 pts] If you look closely, you can see some interesting shapes showing up in my figure. What do they correspond to? Why do you think they're visible in the predicted surface?

- They correspond to the predicted values. They are visible in the predicted surface because the predicted values seems to be dependent.

16. [2 pts] Below is a map I made of the variance in the predicted values. What stands out to you? Why?

- Some part have really high variance, and some low. Looks like there are clusters in the variance plot.