SUMMARY OF EPIDEMIOLOGICAL DATA ANALYSIS

Data Exploration & Aggregation

Loaded the PennLC data from the SpatialEpi package and discovered 67 counties in the state of Pennsylvania

```
Console  Terminal  Background Jobs
R 4.4.1 · ~/Desktop/R Project/
> data("pennLC")
>
> #-------------------------------------------------------
> ##Exercise 1
> ##data exploration
> class(pennLC)
[1] "list"
> names(pennLC)
[1] "geo"             "data"            "smoking"         "spatial.polygon"
>
>
> ##exploring data structure
> head(pennLC$data)
  county cases population race gender      age
1 adams     0       1492    o      f Under.40
2 adams     0        365    o      f    40.59
3 adams     1         68    o      f    60.69
4 adams     0         73    o      f      70+
5 adams     0      23351    w      f Under.40
6 adams     5      12136    w      f    40.59
> head(pennLC$smoking)
      county smoking
1      adams   0.234
2 allegheny   0.245
3 armstrong   0.250
```

Data was grouped by county to compute the total observed cases (Y) and expected cases (E) using stratified rates.

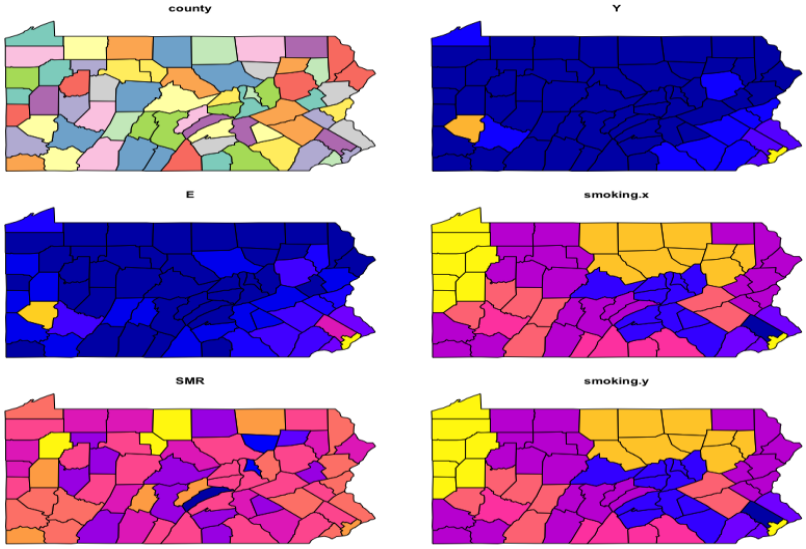
```
26  ###data preparation
27  ##Group the data for each county
28  d <- group_by(pennLC$data, county) %>% summarize(Y = sum(cases))
29  head(d)
30
31  ##arranging all data into a dataframe
32  pennLC$data <- pennLC$data[order(pennLC$data$county, pennLC$data$race, pennLC$data$gender, pennLC$data$age), ]
33
34  ## computing expected values
35  E <- expected(population = pennLC$data$population, cases = pennLC$data$cases, n.strata = 16)
36
37  d$E <- E
38
39
40  ##smokers percentage
41  d <- merge(d, pennLC$smoking, by = "county")
42
43
```

Standardized Mortality Ratio (SMR):

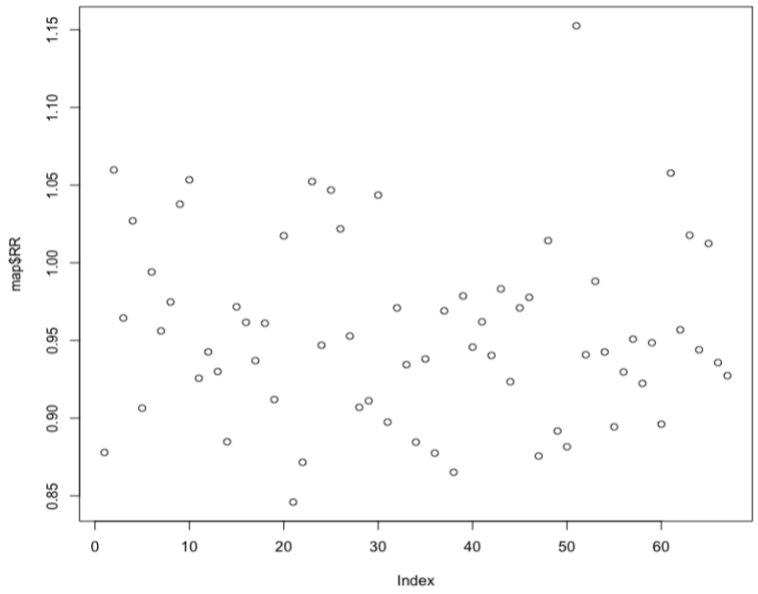
This is given by:

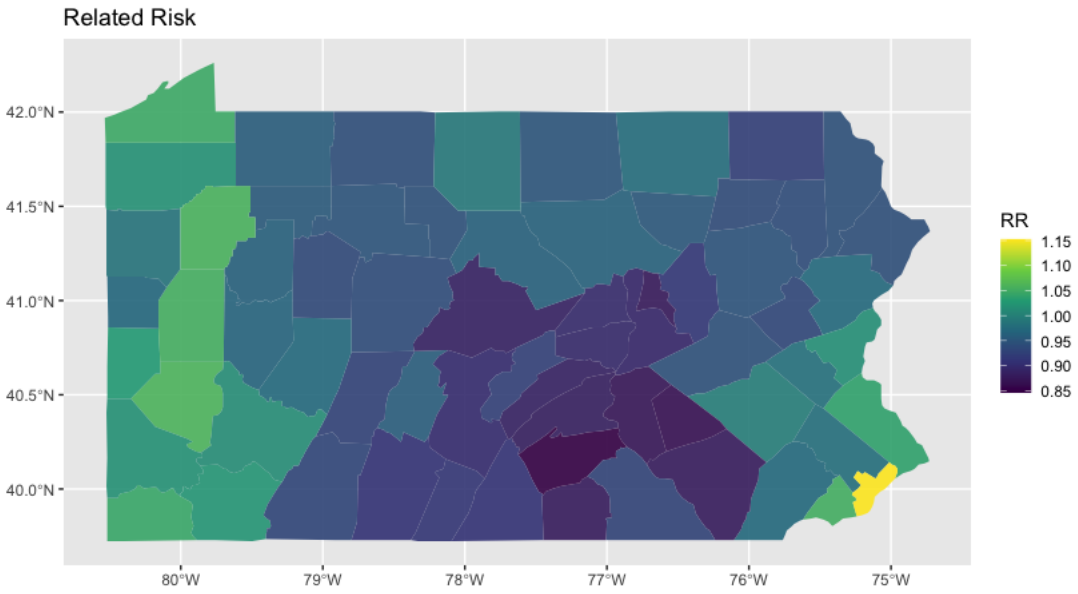
d$SMR = d$Y / d$E

where **Y** is the observed cases, and **E** is the expected cases.



Bayesian Spatial Modeling:
Using INLA to generate neigbourhood structure, estimate Relative Risk, random effect and covariate smoking for fixed effect, this was the results:





Smoking had a posterior mean of β_1=**1.156** suggesting a positive association with RR.

Credible interval **[−0.081,2.384]** crosses zero, indicating uncertainty.

Difference between relative risk from the spatial model with the SMR:

SMR	**RR**
SMR tends to vary a lot and can show extreme values in areas with low population or very few cases	RR is smoother and more consistent because it accounts for spatial relationships and other influencing factors.
The map of RR shows clearer patterns	The map of SMR, which can appear scattered or noisy.