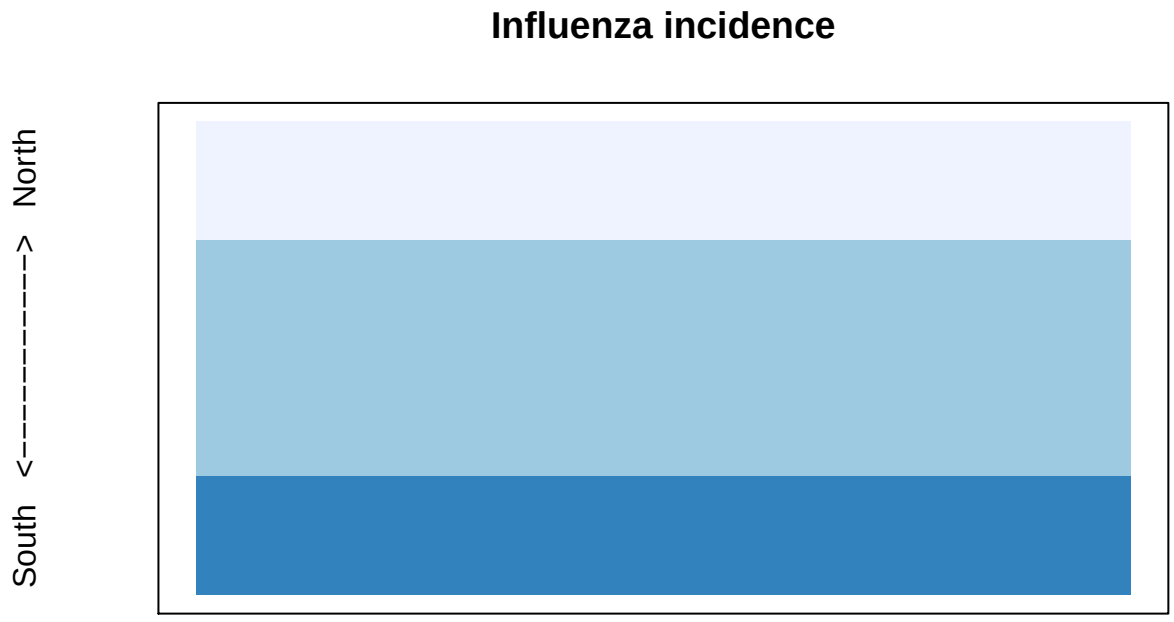
Areal aggregation on maps

18 marks

a. **(2 marks)** Construct the choropleth for the ratio `CasesPerHouse1`. What does it indicate about the geographic distribution of influenza incidence?

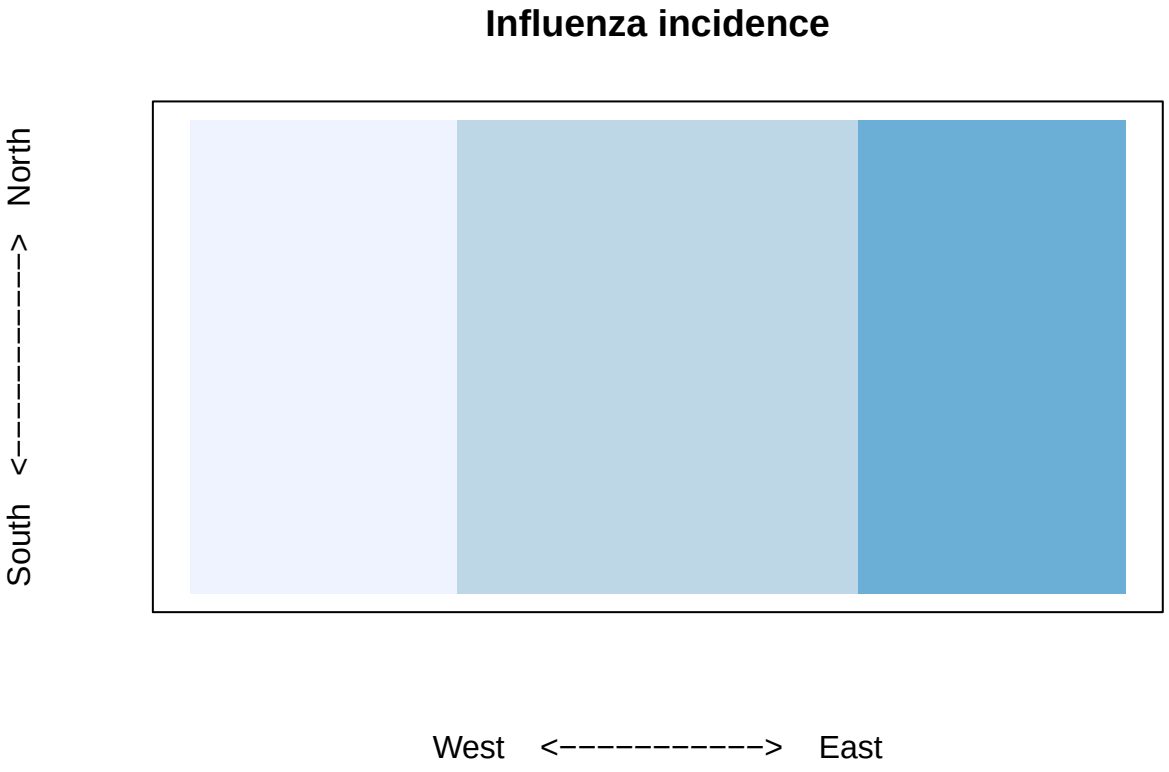
```
NumHouses1 = AggregateByID(NumHouses,RegionID1)
NumCases1 = AggregateByID(NumCases, RegionID1)

CasesPerHouse1 = NumCases1/ NumHouses1

col_areas(CasesPerHouse1,
                  main="Influenza incidence")
```



this is indicating that there is more cases per house on the south than on the north

b. **(3 marks)** A different definition of regions is given by `RegionID2`. Construct the appropriate choropleth for these regions and summarize what it indicates about the geographic distribution of influenza incidence?

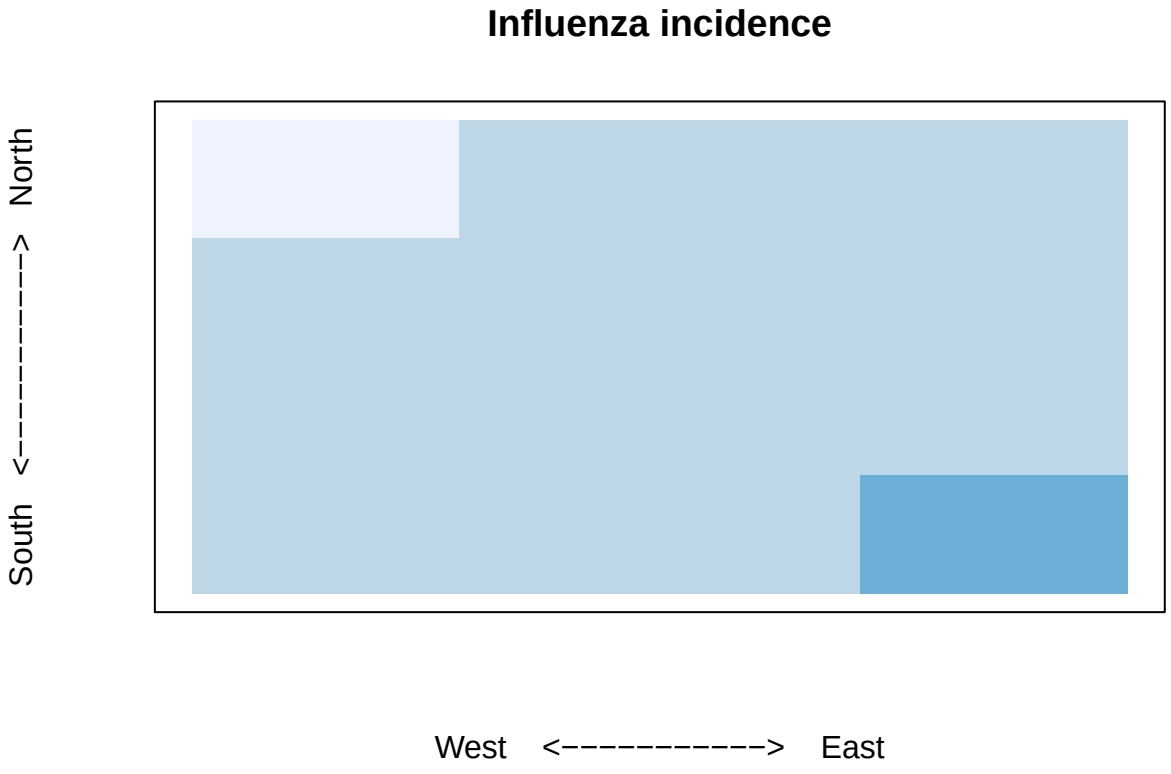
```
NumHouses2 = AggregateByID(NumHouses,RegionID2)
NumCases2 = AggregateByID(NumCases, RegionID2)
CasesPerHouse2 = NumCases2/ NumHouses2
col_areas(CasesPerHouse2,
                main="Influenza incidence")
```



this is indicating that there is more cases per house on the east than on the west

c. **(3 marks)** A third definition of regions is given by `RegionID3`. This definition distinguishes population centres from the rural background. Construct the appropriate choropleth for these regions and summarize what it indicates about the geographic distribution of influenza incidence?

```
NumHouses3 = AggregateByID(NumHouses,RegionID3)
NumCases3 = AggregateByID(NumCases, RegionID3)
CasesPerHouse3 = NumCases3/ NumHouses3
col_areas(CasesPerHouse3,
              main="Influenza incidence")
```



this is indicating that there is more cases per house in south-east than north-west

d. **(2 marks)** What do you conclude about aggregations by political regions as a means to understand the spatial distribution of influenza incidence? Be clear in your reasoning.

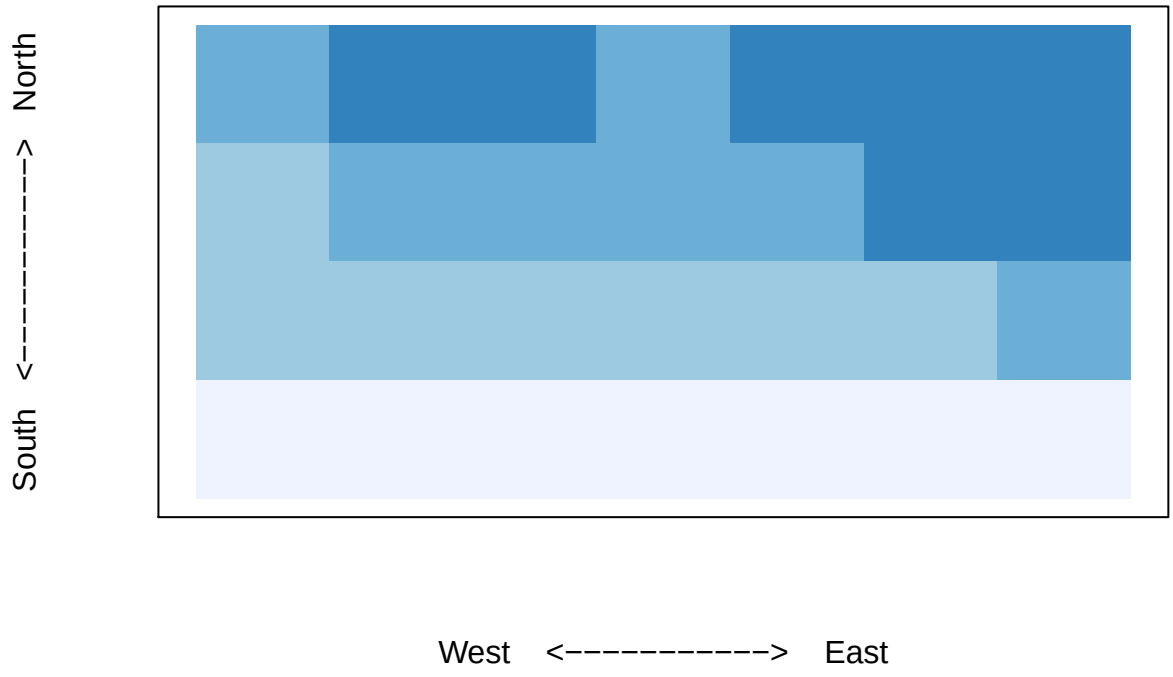
it's not reasonable as political region is not a good explanatory variable for density of influenza. They might not even be correlated, which may result in a uniform distribution of inferred density.

Load the data from the file `Data_b.R` to yield two new data matrices `NumHouses` and `NumCases`.

```
source("Data_b.R")
```

e. **(3 marks)** Construct the choropleth for the influenza incidence based on the townships/towns as regions. Summarize what this display indicates about the geographic distribution of influenza incidence?

```
CasesPerHouse = NumCases/NumHouses
col_areas(CasesPerHouse,
                  main="Influenza incidence")
```

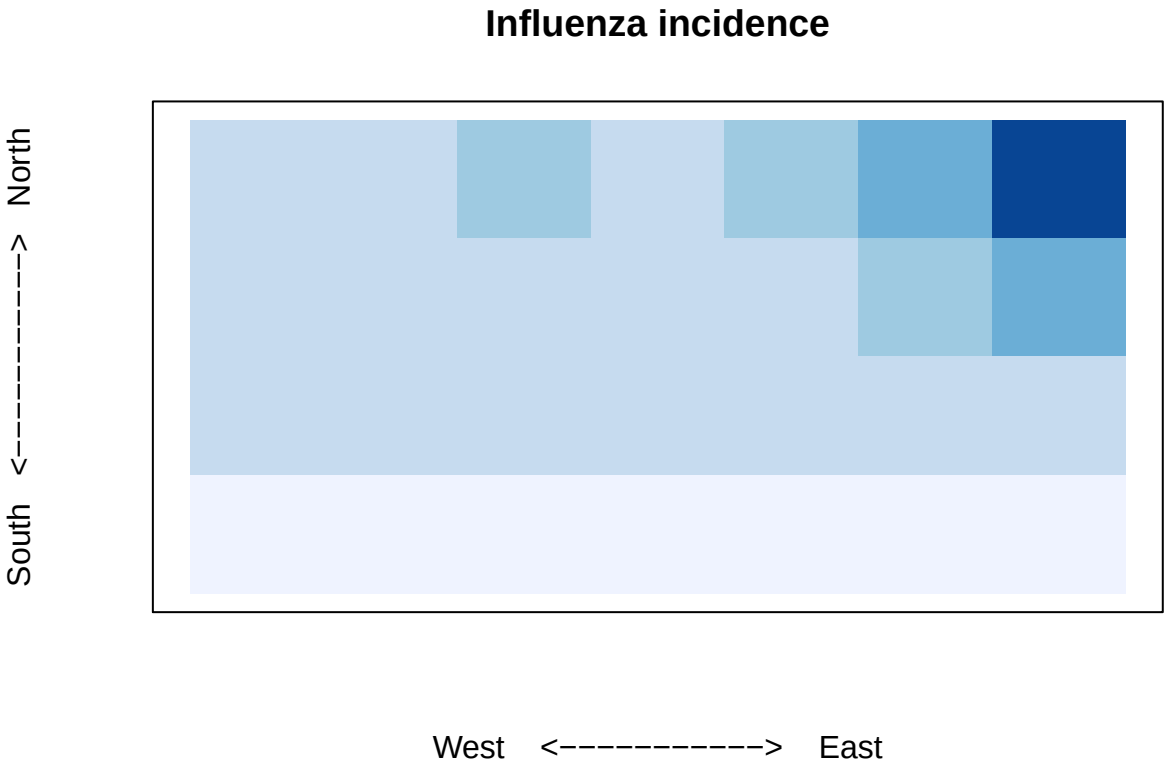


The influenza is most severe on the north-east. In general, it is more severe on the north than on the south.

f. **(3 marks)** The choropleth just constructed uses five-number summary to determine the colours of the townships/towns. In this way, each category has 25% of the data (i.e. boundaries are at the quartiles).

The function `col_areas` also takes an argument `breaks` and so the same choropleth would be produced by Rather than use the quartiles as boundaries, construct a choropleth for these data where the five break points (including the minimum and the maximum) are equally spaced from the minimum to the maximum. Summarize what this display indicates about the geographic distribution of influenza incidence?

```
minden = min(NumCases/NumHouses)
maxden = max(NumCases/NumHouses)
breaks = seq( minden, maxden, by=(maxden-minden)/5 )
col_areas(NumCases/NumHouses, breaks=breaks,
              main="Influenza incidence")
```




```
NumCases/NumHouses
```

```
##           [,1]       [,2]       [,3]  [,4]       [,5] [,6] [,7]
## [1,] 0.1900000 0.19333333 0.20000000 0.175 0.35000000 0.42 0.89
## [2,] 0.1314286 0.15000000 0.15571429 0.160 0.18000000 0.26 0.42
## [3,] 0.0720000 0.07111111 0.07571429 0.080 0.08571429 0.09 0.14
## [4,] 0.0200000 0.02000000 0.02000000 0.020 0.02000000 0.02 0.02
```

```
fivenum(NumCases/NumHouses)
```

```
## [1] 0.02000000 0.04555556 0.13571429 0.19166667 0.89000000
```

this is indicating that only the most north-east town has severe influenza, towns closer to it, in general, has more severe influenza than towns far from it. Most towns that are three blocks away from north-east town has little influenza.

g. **(2 marks)** What do you conclude about basing conclusions about the spatial distribution of influenza incidence based on data aggregated by the value of the incidence itself? Be clear in your reasoning.

this is more helpful in visualizing severity of influenza. As severity, intuitively, is propotional to number of cases/house instead of the quantile of cases/house in the town. The readers are usually assuming the severity is propotional to the satuation of blocks in the choropleth