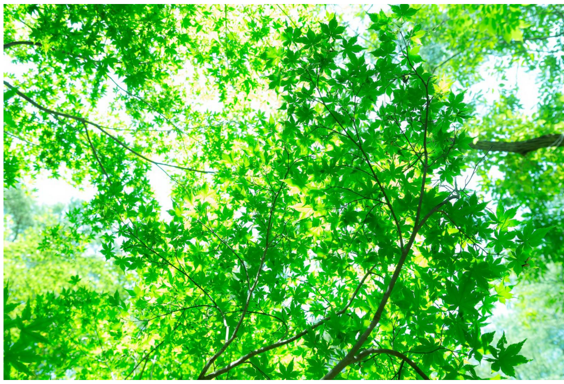


Question 1

1.

$$
\begin{aligned}
E[\overline{y}] &= E[n^{-1}\sum_{i=1}^{n} y_i] \\
&= n^{-1}E[\sum_{i=1}^{n} y_i] \\
&= n^{-1}n\mu \\
&= \mu
\end{aligned}
$$

So $E[\overline{y}]$ is an unbised estimator of μ.

2.

$$
\begin{aligned}
Var(\overline{y}) &= Var(n^{-1}\sum_{i=1}^{n} y_i) \\
&= n^{-2}Var(\sum_{i=1}^{n} y_i) \\
&= n^{-2}(\sum_{i=1}^{n} Var(y_i) + \sum_{i\neq j} Cov(y_i, y_j)) \\
&= n^{-2}(n + (n^2 - n)\rho) \\
&= n^{-1}(1 + (n-1)\rho)
\end{aligned}
$$

3.

$\overline{y}$ is a consistent estimate of μ if and only if $\lim_{n\to\infty} P(|\overline{y} - \mu| > \epsilon) = 0$ for any $\epsilon > 0$

$$\overline{y} \sim N(\mu, n^{-1}(1+(n-1)\rho))$$

$$\begin{aligned}
P(|\overline{y}-\mu| > \epsilon) &= P(\frac{\sqrt{n|\overline{y}-\mu|}}{\sqrt{1+(n-1)\rho}} > \frac{\sqrt{n}\epsilon}{\sqrt{1+(n-1)\rho}}) \\
&= 2(1-\Phi(\frac{\sqrt{n}\epsilon}{\sqrt{1+(n-1)\rho}})) \\
&= 2(1-\Phi(\frac{\epsilon}{\sqrt{\frac{1+(n-1)\rho}{n}}})) \\
&= 2(1-\Phi(\frac{\epsilon}{\sqrt{1/n+(1-1/n)\rho}})) \\
&\to 2(1-\Phi(\frac{\epsilon}{\sqrt{\rho}}))
\end{aligned}$$

as $n \to \infty$. This is not equal to 0 for $\epsilon > 0$ so $\overline{y}$ is not a consistent estimate of μ

Question 2

Answer:Geostatistial data

Tempareture can be measured at any location of US so it is fixed. If you pick two location in US, you can find infinite number of locations in US with tempetature so it is continuous. Therefore, it is a geostatistical data.

Question 3

```r
library(rgdal)
```

```
## Loading required package: sp
```

```
## rgdal: version: 1.4-8, (SVN revision 845)
##  Geospatial Data Abstraction Library extensions to R successfully loaded
##  Loaded GDAL runtime: GDAL 2.2.3, released 2017/11/20
##  Path to GDAL shared files: C:/Users/Bun Hagiwara/Documents/R/win-library/3.6/rgdal/gdal
##  GDAL binary built with GEOS: TRUE
##  Loaded PROJ.4 runtime: Rel. 4.9.3, 15 August 2016, [PJ_VERSION: 493]
##  Path to PROJ.4 shared files: C:/Users/Bun Hagiwara/Documents/R/win-library/3.6/rgdal/proj
##  Linking to sp version: 1.4-1
```

```r
library(spdep)
```

```
## Loading required package: spData
```

```
## To access larger datasets in this package, install the spDataLarge
## package with: 'install.packages('spDataLarge',
## repos='https://nowosad.github.io/drat/', type='source')'
```

```
## Loading required package: sf

## Linking to GEOS 3.6.1, GDAL 2.2.3, PROJ 4.9.3
```

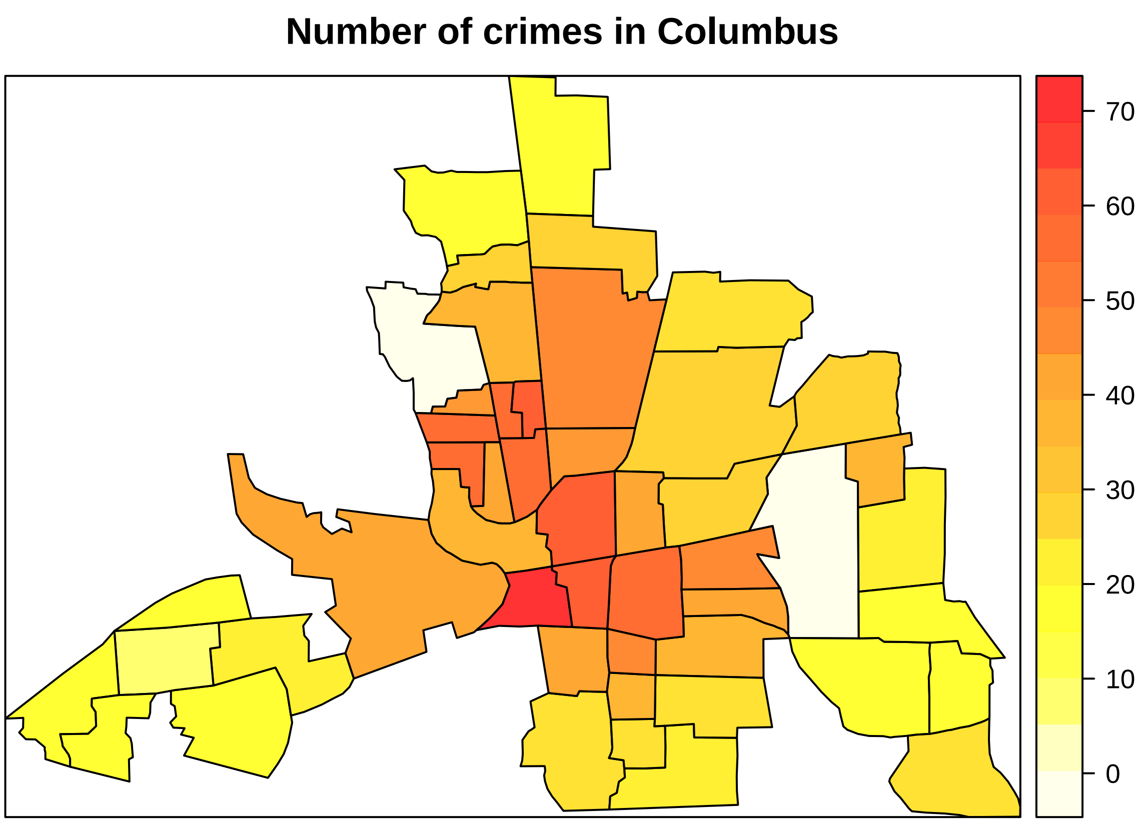
```r
oldc<-data(oldcol)
oldc
```

```
## [1] "oldcol"
```

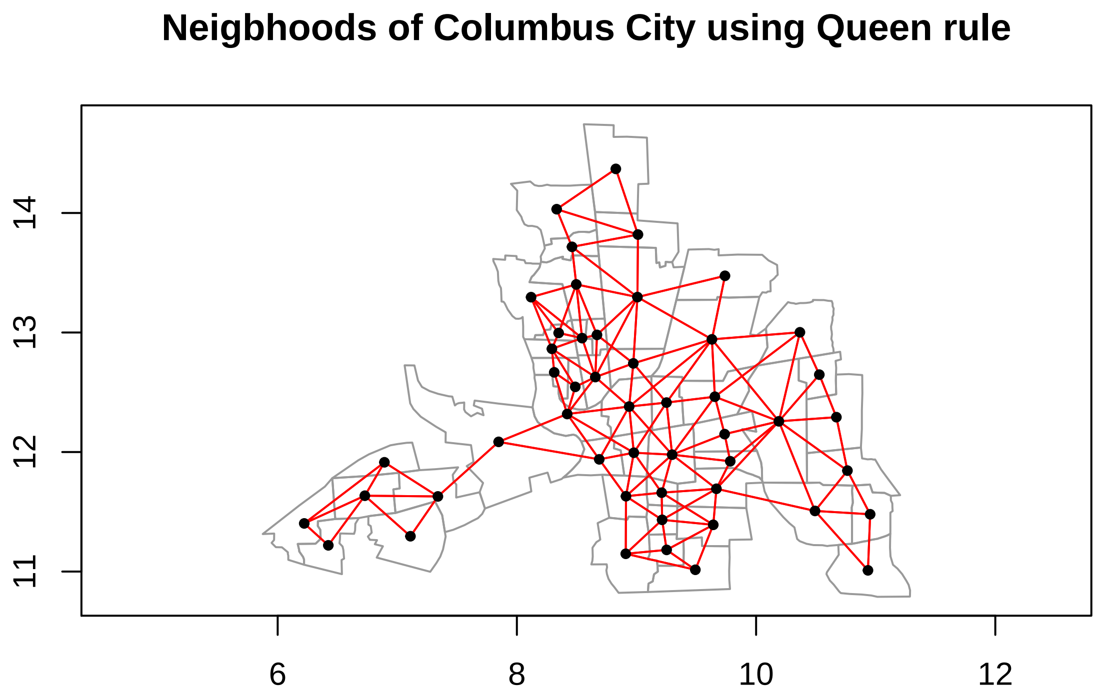
```r
columbus <- readOGR(system.file("shapes/columbus.shp", package="spData")[1])
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\Bun Hagiwara\Documents\R\win-library\3.6\spData\shapes\columbus.shp", layer: "colu
## with 49 features
## It has 20 fields
## Integer64 fields read as strings:  COLUMBUS_ COLUMBUS_I POLYID
```

```r
spplot(columbus, "CRIME", col.regions = rev(heat.colors(20,alpha=0.8)),main="Number of crimes in Columbu
```



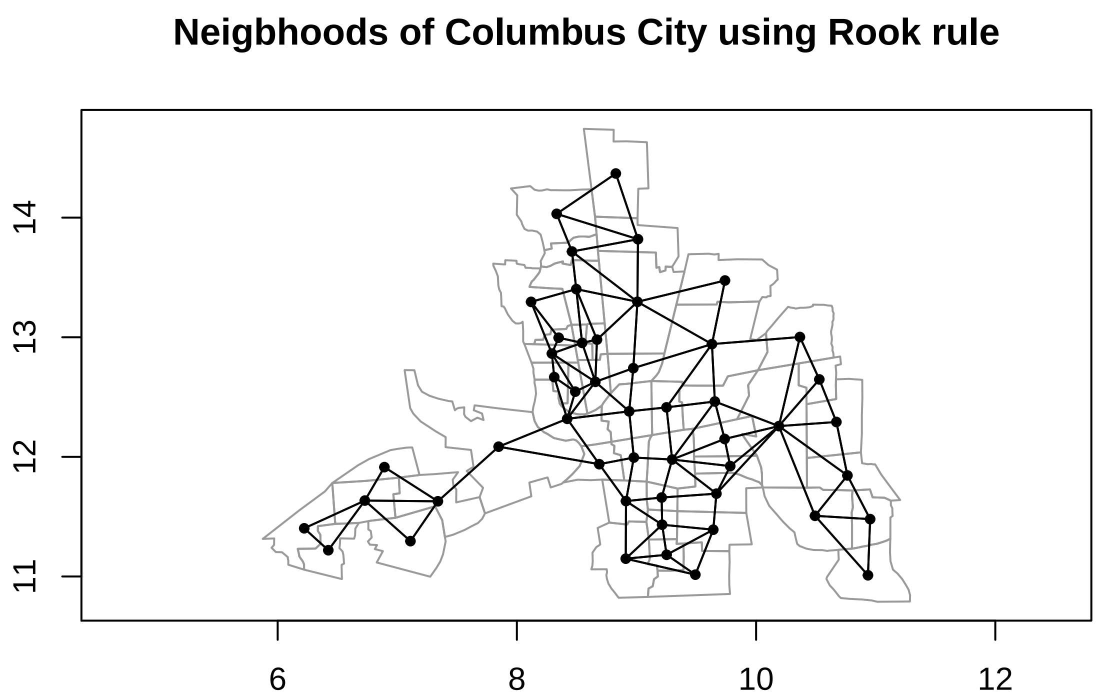

```r
Wqueen = poly2nb(columbus)
Wrook = poly2nb(columbus, queen=FALSE)
#Use queen rule
plot(columbus,border="grey60", axes=TRUE,main="Neigbhoods of Columbus City using Queen rule")
plot(Wqueen, coordinates(columbus), pch=19, cex=0.6, add=TRUE,col="red")
```




```
moran.test(x=columbus$CRIME, nb2listw(Wqueen), alternative="two.sided")
```

```
##
##  Moran I test under randomisation
##
## data:  columbus$CRIME
## weights: nb2listw(Wqueen)
##
## Moran I statistic standard deviate = 5.5894, p-value = 2.279e-08
## alternative hypothesis: two.sided
## sample estimates:
## Moran I statistic       Expectation          Variance
##       0.500188557      -0.020833333       0.008689289
```

```
#Use rook rule
plot(columbus,border="grey60", axes=TRUE,main="Neigbhoods of Columbus City using Rook rule")
plot(Wrook, coordinates(columbus), pch=19, cex=0.6, add=TRUE)
```




```r
moran.test(x=columbus$CRIME, nb2listw(Wrook), alternative="two.sided")
```

```
##
##  Moran I test under randomisation
##
## data:  columbus$CRIME
## weights: nb2listw(Wrook)
##
## Moran I statistic standard deviate = 5.4579, p-value = 4.819e-08
## alternative hypothesis: two.sided
## sample estimates:
## Moran I statistic       Expectation          Variance
##       0.523670213      -0.020833333       0.009952987
```

Null hypothesis $H_0 : I = -1/(n-1) = -0.020833333$, Alternative hypothesis $H_1 : I \neq -1/(n-1) = -0.020833333$

When queen rule is used for spatial weight matrix,

Moran's I = 0.500188557, p-value = 2.279e-08

When rook rule is used for spatial weight matrix,

Moran's I = 0.523670213, p-value = 4.819e-08

Either way, p-value < 0.05 so H_0 is redjected. Moran's I is postive so we can know there is auto-correlation for the number of crimes in US (high-high or low-low for close location).

Question 4

for liner regression $\mathbf{Y} = \beta\mathbf{u} + \epsilon$, normally slope β represents how variable u_i effects on Y_i. But specifically for this case, β can be seen as a weight of $\mathbf{u}$. β is unstandardised estimator of W so weight matrix can be measured using the values.

Question 5

$Y(t) = Y(t-1) + e(t), t = 0, 1, 2...\ Y(0) = 0, \{e(t)\} \overset{iid}{\sim} N(0, \sigma^2)$

Secnd order stationarity: $E(Z(s)) = \mu$ for all $s \in D$ $Cov(Z(s), Z(s+h)) = C(h)$ for all $s, s+h \in D$

Consider about $Y(t)$

$$\begin{aligned}
E[Y(t)] &= E[Y(t-1) + e(t)] \\
&= E[Y(t-2) + e(t-1) + e(t)] \\
&= E[Y(0) + e(1) + ... + e(t)] \\
&= 0 \text{ for all t}
\end{aligned}$$

$$\begin{aligned}
&\text{let } Cov(Y(t), Y(t+1)) = C(1), \\
&\text{Now, } Cov(Y(t+1), Y(t+2)) = E[Y(t+1)Y(t+2)] - E[Y(t+1)]E[Y(t+2)] \\
&= E[(Y(t) + e(t+1))(Y(t) + e(t+2) + e(t+1))] \\
&= C(1) + E[(Y(t) + e(t+1))(e(t+2) + e(t+1))] \\
&= C(1) + E[Y(t)(e(t+2) + e(t+1)) + e(t+1)e(t+2) + e(t+1)^2] \\
&= C(1) + E[(e(t) + ... + e(1))(e(t+2) + e(t+1)) + e(t+1)^2] \\
&\text{since } e(t)\text{'s are i.i.d. with mean 0,} \\
&= C(1) + E[e(t+1)^2] \\
&= C(1) + Var(e(t+1)) \\
&= C(1) + \sigma^2 \\
&\neq C(1)
\end{aligned}$$

Therefore, $Y(t)$ is not second order stationary.

Consider about $D(t) = Y(t) - Y(t-1)$ now.

$$\begin{aligned}
E[D(t)] &= E[Y(t) - Y(t-1)] \\
&= E[e(t)] \\
&= 0 \text{ for all t}
\end{aligned}$$

$$\begin{aligned}
Cov(D(t), D(t+h)) &= E[D(t)D(t+h)] - E[D(t)]E[D(t+h)] \\
&= E[(Y(t) - Y(t-1))(Y(t+h) - Y(t+h-1))] \\
&= E[e(t)e(t+h)] \\
&= 0 \text{ for all t, t+h}
\end{aligned}$$

Therefore, $D(t)$ is second order stationary.