HW3

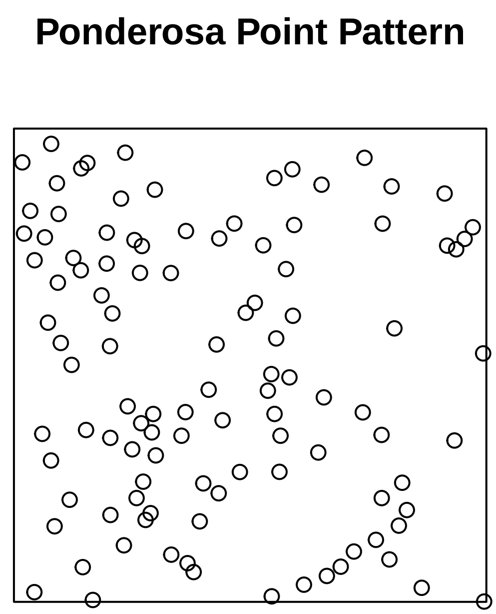
Haojie Liu

2025-05-06

Open RStudio and install the packages “spatstat” and “maptools”. Then load the libraries for these two packages. Also load the library “sf” that you have previously installed.

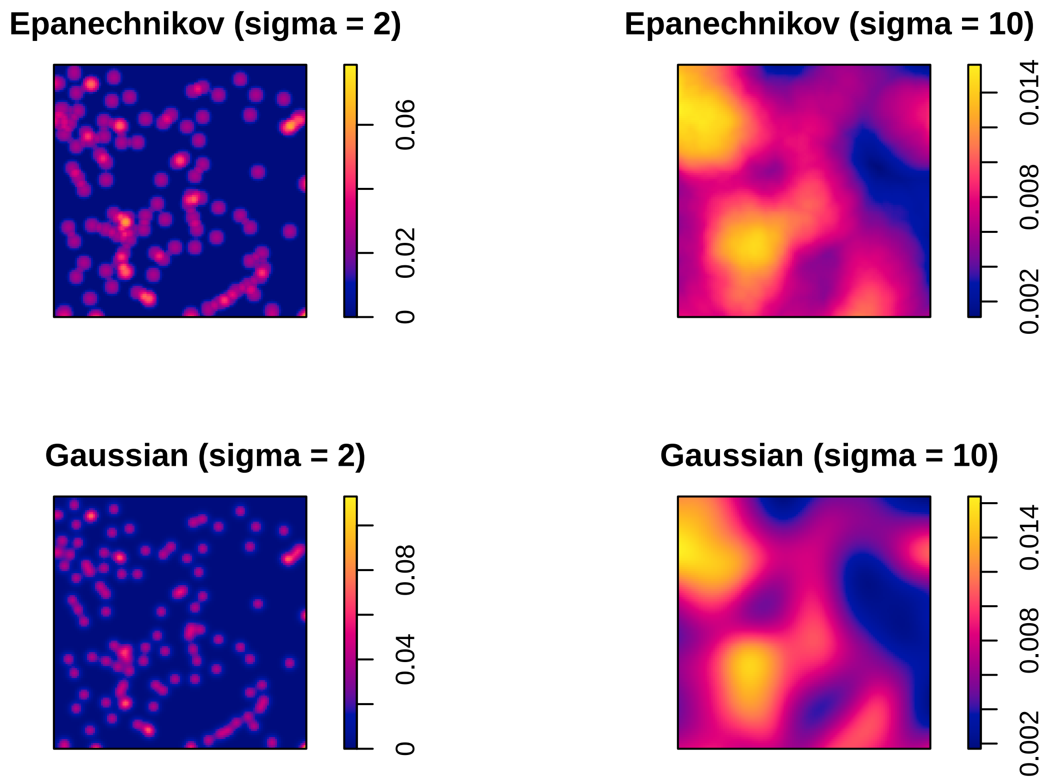
1. Open the spatstat data file “ponderosa”.

A. Plot the point pattern and interpret it.



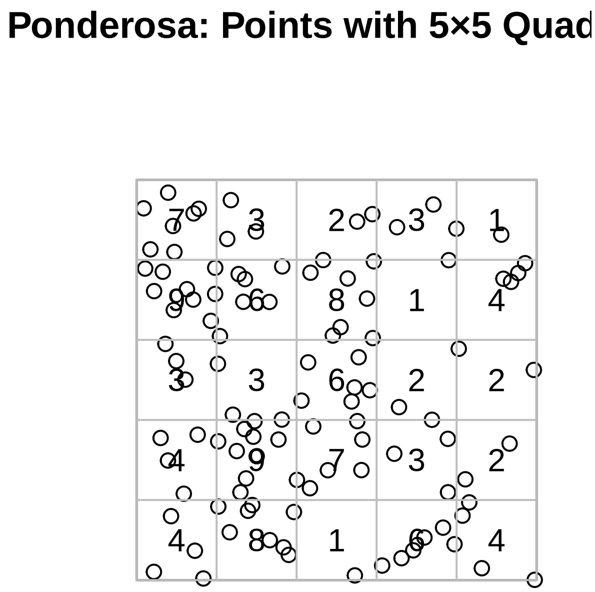
The point pattern of ponderosa pine trees shows an uneven spatial distribution with visible clusters, particularly in the left regions of the study area. This suggests that the trees are not randomly distributed but may exhibit a tendency to aggregate.

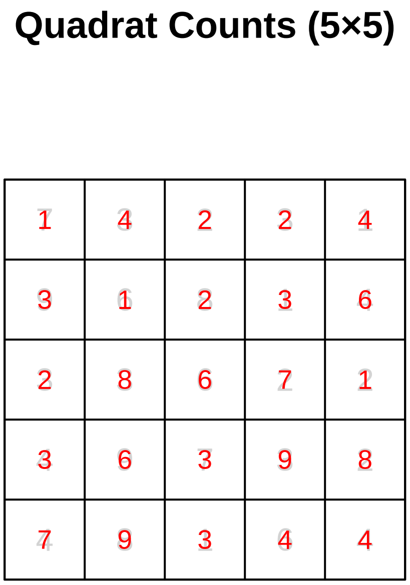
B. Generate kernel density estimates for this point pattern. Show and discuss the effects of choosing two different types of kernel and two different smoothing bandwidths.



The four maps make two things clear. First, how much you smooth the data changes what you see: with tight smoothing you get lots of small, bright spots showing the little clumps of trees, but with loose smoothing those spots blur together into wide hills that show only the big, overall patterns. Second, the smoothing "shape" you pick also matters: the Epanechnikov shape only looks at trees within a fixed distance, so its hills have flat tops and sharp edges, while the Gaussian shape takes every tree into account—just giving farther ones less weight—so its hills are rounder and fade away more gently.

C. Overlay a grid of 5x5 quadrats over the ponderosa point pattern and plot the points and counts in the quadrats.





D. State your hypotheses, perform a quadrat test (the chi-squared test in spatstat is OK) and report whether you would reject or not reject the null hypothesis that the point pattern is consistent with an IRP at the significance level of 0.05.

```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:  ponderosa
## X2 = 36.444, df = 24, p-value = 0.09934
## alternative hypothesis: two.sided
##
## Quadrats: 5 by 5 grid of tiles
```

Since $0.09934 > 0.05$, we failed to reject H_0 There is no statistically significant evidence at the 5% level to reject complete spatial randomness, so the pattern is consistent with an (homogeneous) Poisson process.

E. Now perform a nearest neighbor test of the ponderosa pine data to determine if the point pattern is consistent with an independent random process. This is the test outlined in the lecture video for Module 3 Lecture 4 and in Exercise 10. Use R as much as possible. Move formally through the hypothesis test and report your decision for a chosen significance level.

```
## Observed mean nn-dist: 6.8373
```

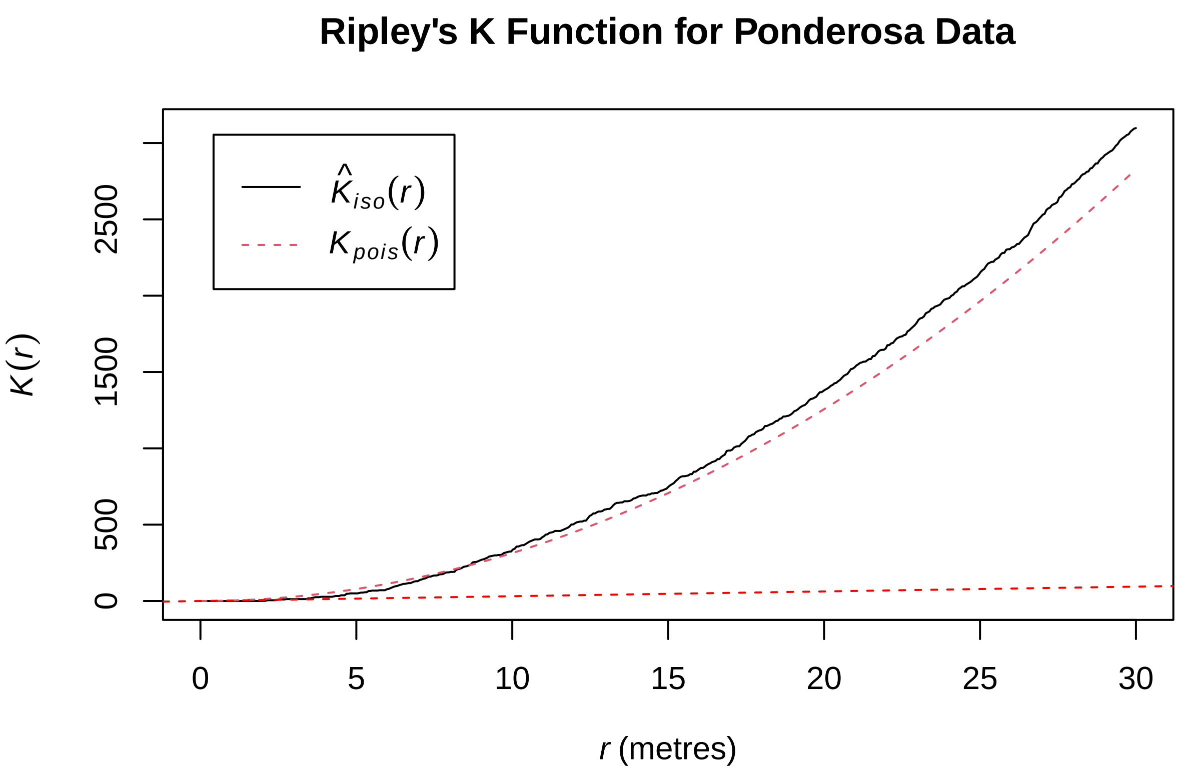
```
## Expected under CSR:  5.7735
```

```
## Z = 3.663,  p = 0.0002

##
##  Clark-Evans test
##  No edge correction
##  Z-test
##
## data:  ponderosa
## R = 1.1842, p-value = 0.0002492
## alternative hypothesis: two-sided
```

The observed mean nearest-neighbor distance (6.8373) exceeded the CSR expectation (5.7735), giving Z = 3.663 (p = 0.0002) and Clark–Evans R = 1.1842 (p = 0.00025). Since $p < 0.05$ and $|Z| > 1.96$, we reject CSR, indicating the trees are significantly more regularly spaced than random.

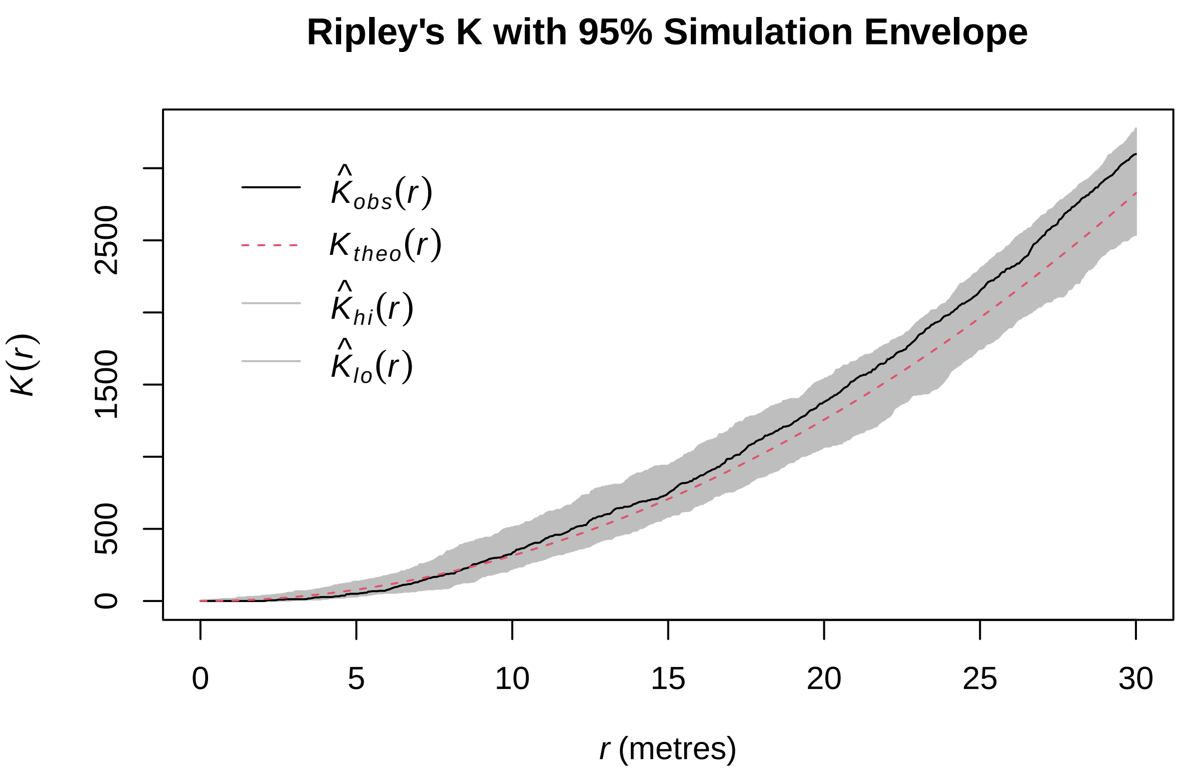
F. Plot and interpret the K function for the ponderosa data using the edge correction='Ripley')



The Ripley-corrected K-function (solid black) lies below the CSR expectation (dashed red) for a little, indicating that trees repel each other at small scales (regular spacing). Beyond that range the empirical K(r) converges on—and even slightly exceeds—the theoretical curve, showing that at larger distances the pattern behaves much like randomness (or very weak clustering), so overall the pines exhibit inhibition at short ranges but no strong departure from CSR at broader scales.

G. Fit a confidence envelope around the expected K-function, plot the envelope function and interpret what you see.

```
## Generating 1000 simulations of CSR  ...
## 1, 2, 3, ......10.........20.........30.........40.........50.........60..
## .......70.........80.........90.........100.........110.........120.........130
## .........140.........150.........160.........170.........180.........190........
## .200.........210.........220.........230.........240.........250.........260......
## ...270.........280.........290.........300.........310.........320.........330....
## .....340.........350.........360.........370.........380.........390.........400..
## .......410.........420.........430.........440.........450.........460.........470
## .........480.........490.........500.........510.........520.........530........
## .540.........550.........560.........570.........580.........590.........600......
## ...610.........620.........630.........640.........650.........660.........670....
## .....680.........690.........700.........710.........720.........730.........740..
## .......750.........760.........770.........780.........790.........800.........810
## .........820.........830.........840.........850.........860.........870........
## .880.........890.........900.........910.........920.........930.........940......
## ...950.........960.........970.........980.........990.........
## 1000.
##
## Done.
```



The simulation envelope shows that the observed K-curve lies below the lower grey bound at small scales—strong evidence of inhibition (regular spacing)—and then remains within the envelope at broader scales, indicating no statistically significant deviation from CSR beyond that initial repulsion.

H. Perform a MAD test of the K function with the ponderosa point pattern. Show your results and interpret them.

```
## Simultaneous critical envelopes for K(r)
## and observed value for 'ponderosa'
## Edge correction: "iso"
## Obtained from 1000 simulations of CSR
## Envelope based on maximum deviation of K(r) from null value for CSR (known
## exactly)
## Alternative: two.sided
## Significance level of simultaneous Monte Carlo test: 1/1001 = 0.000999
## .......................................................
##      Math.label     Description
## r    r              distance argument r
## obs  hat(K)[obs](r) observed value of K(r) for data pattern
## theo K[theo](r)     theoretical value of K(r) for CSR
## lo   hat(K)[lo](r)  lower critical boundary for K(r)
## hi   hat(K)[hi](r)  upper critical boundary for K(r)
## .......................................................
## Default plot formula:  .~r
## where "." stands for 'obs', 'theo', 'hi', 'lo'
## Columns 'lo' and 'hi' will be plotted as shading (by default)
## Recommended range of argument r: [0, 30]
## Available range of argument r: [0, 30]
## Unit of length: 1 metre
```

The simultaneous envelope test at a very stringent significance level (p=0.000999) shows the observed K-curve dipping below the lower critical boundary at small distances, demonstrating significant inhibition among the trees at those scales. Beyond this initial range, the observed values lie entirely within the [lo, hi] band, indicating no significant departure from CSR at larger distances. Overall, the global MAD test confirms strong small-scale regularity but otherwise supports a random Poisson pattern.

2. There is a zipped folder "Calairsites" in the Module 3 datasets folder. The airmonitoringstations.shp file contains point locations for air quality monitoring stations in California (and a few in Mexico). The arb_california_airdistricts_aligned_03.shp file contains air resources board districts (polygons) for California. So here you have point and polygon data for California.

A. Coerce the point and polygon data into spatstat (see Exercise 11). (NB– the point file has some point locations in Mexico. These will cause problems, Exclude them after you first read in the points shapefile by deleting observations where ZIPCODE=="xxxxx".)

```
## Reading layer `airmonitoringstations' from data source
##   `/Users/liuhaojie/Desktop/STATSM222/Week6/calairsites/airmonitoringstations.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 296 features and 12 fields
## Geometry type: POINT
## Dimension:     XY
## Bounding box:  xmin: -351221.1 ymin: -624485.5 xmax: 501013.9 ymax: 423840.9
## Projected CRS: NAD83 / California Albers
```

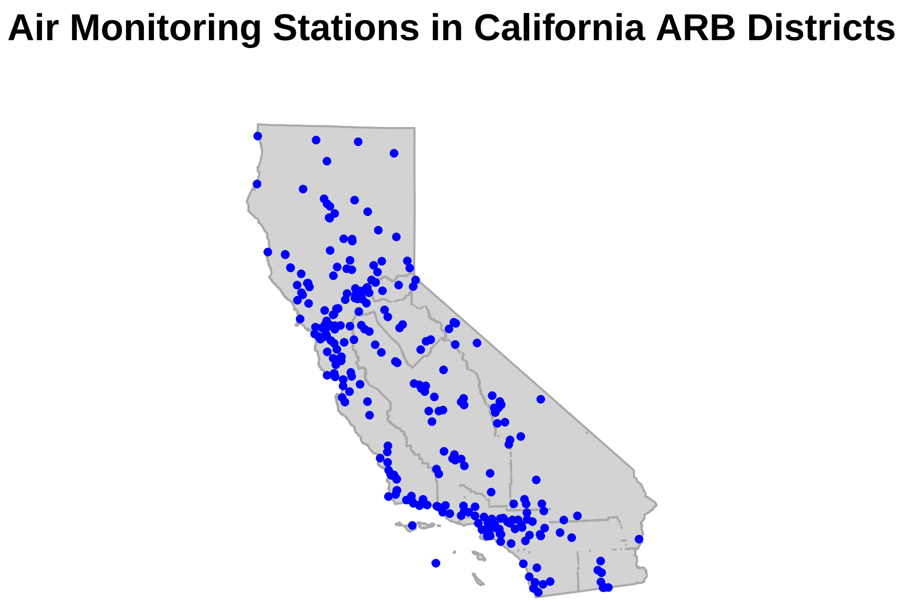
```
## Reading layer `arb_california_airdistricts_aligned_03' from data source
##   `/Users/liuhaojie/Desktop/STATSM222/Week6/calairsites/arb_california_airdistricts_aligned_03.shp'
```

```
##   using driver `ESRI Shapefile'
## Simple feature collection with 47 features and 8 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: -373976.6 ymin: -604526.1 xmax: 540015.4 ymax: 450070.9
## Projected CRS: NAD83 / California Albers
```

```
## List of 10
##  $ xrange      : num [1:2] -373977 540015
##  $ yrange      : num [1:2] -604526 450071
##  $ type        : chr "polygonal"
##  $ area        : num 4.11e+11
##  $ units       :List of 3
##   ..$ singular  : chr "unit"
##   ..$ plural    : chr "units"
##   ..$ multiplier: num 1
##   ..- attr(*, "class")= chr "unitname"
##  $ areafraction: num 0.426
##  $ npoly       : int 5342
##  $ areas       : num [1:5342] 4.10e+11 -1.10e+01 -3.53 -1.66 -9.27e-01 ...
##  $ nvertices   : int [1:5342] 41788 22 9 4 6 4 15 54 20 6 ...
##  $ nhole       : int 5329
##  - attr(*, "class")= chr "summary.owin"
```

```
## List of 7
##  $ is.marked  : logi FALSE
##  $ n          : int 281
##  $ window     :List of 10
##   ..$ xrange      : num [1:2] -373977 540015
##   ..$ yrange      : num [1:2] -604526 450071
##   ..$ type        : chr "polygonal"
##   ..$ area        : num 4.11e+11
##   ..$ units       :List of 3
##   .. ..$ singular  : chr "unit"
##   .. ..$ plural    : chr "units"
##   .. ..$ multiplier: num 1
##   .. ..- attr(*, "class")= chr "unitname"
##   ..$ areafraction: num 0.426
##   ..$ npoly       : int 5342
##   ..$ areas       : num [1:5342] 4.10e+11 -1.10e+01 -3.53 -1.66 -9.27e-01 ...
##   ..$ nvertices   : int [1:5342] 41788 22 9 4 6 4 15 54 20 6 ...
##   ..$ nhole       : int 5329
##   ..- attr(*, "class")= chr "summary.owin"
##  $ intensity  : num 6.84e-10
##  $ nduplicated: int 0
##  $ rounding   : num 14
##  $ rejects    : int 15
##  - attr(*, "class")= chr "summary.ppp"
```

B. Plot the point data over your study region and interpret the plot.

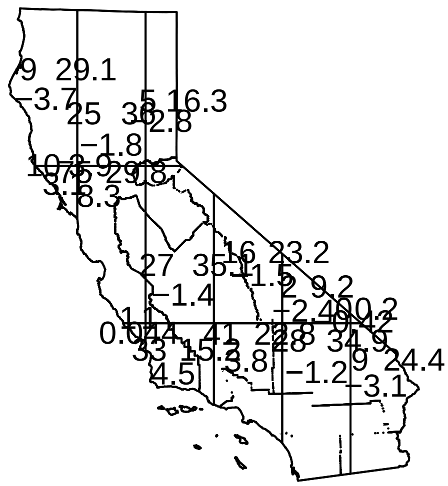


The monitoring stations are clearly concentrated in California's population and pollution centers—particularly the Los Angeles Basin, the San Francisco Bay Area and the Central Valley—while vast mountain, desert and coastal regions have very few sites. This pattern reflects a strategic focus on areas of high human exposure and regulatory concern, leaving more remote or sparsely populated districts with minimal observational coverage.

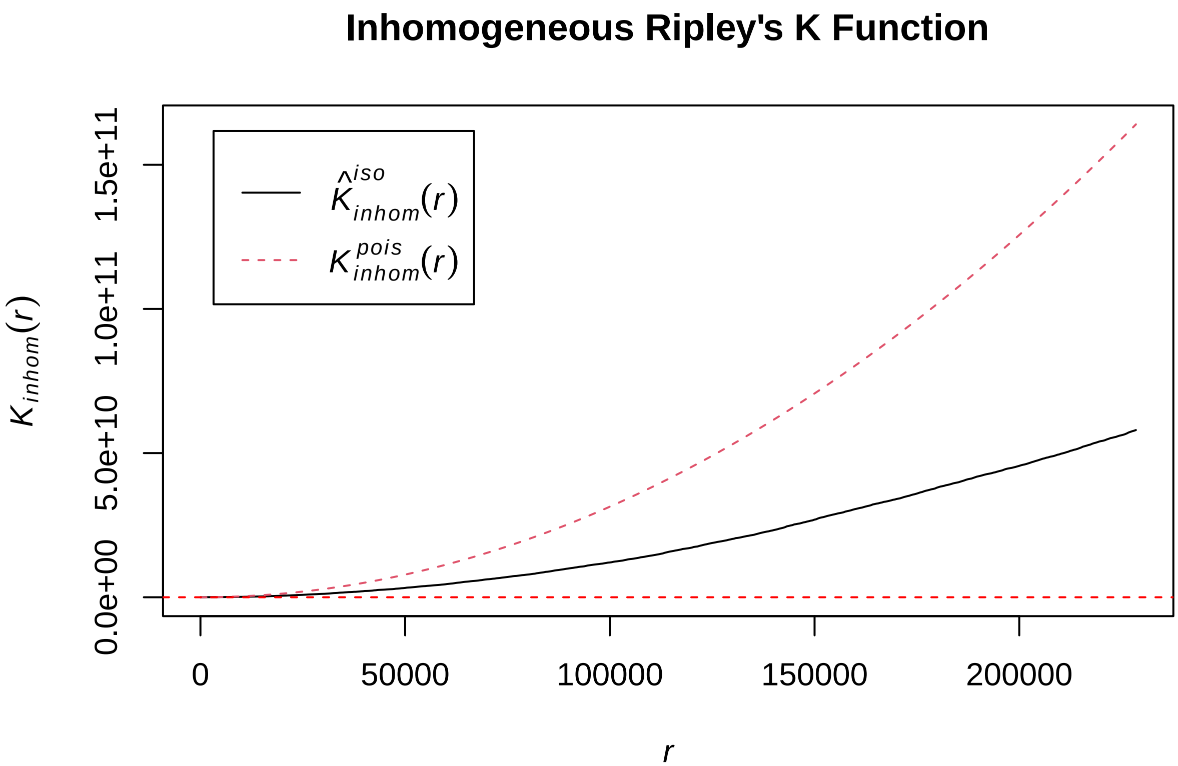
C. Use your knowledge of spatstat to perform first-order and second-order tests of stationarity in your point pattern data. Show all your code, plots & interpretation.

```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:  sites_ppp
## X2 = 159.41, df = 13, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 14 tiles (irregular windows)
```

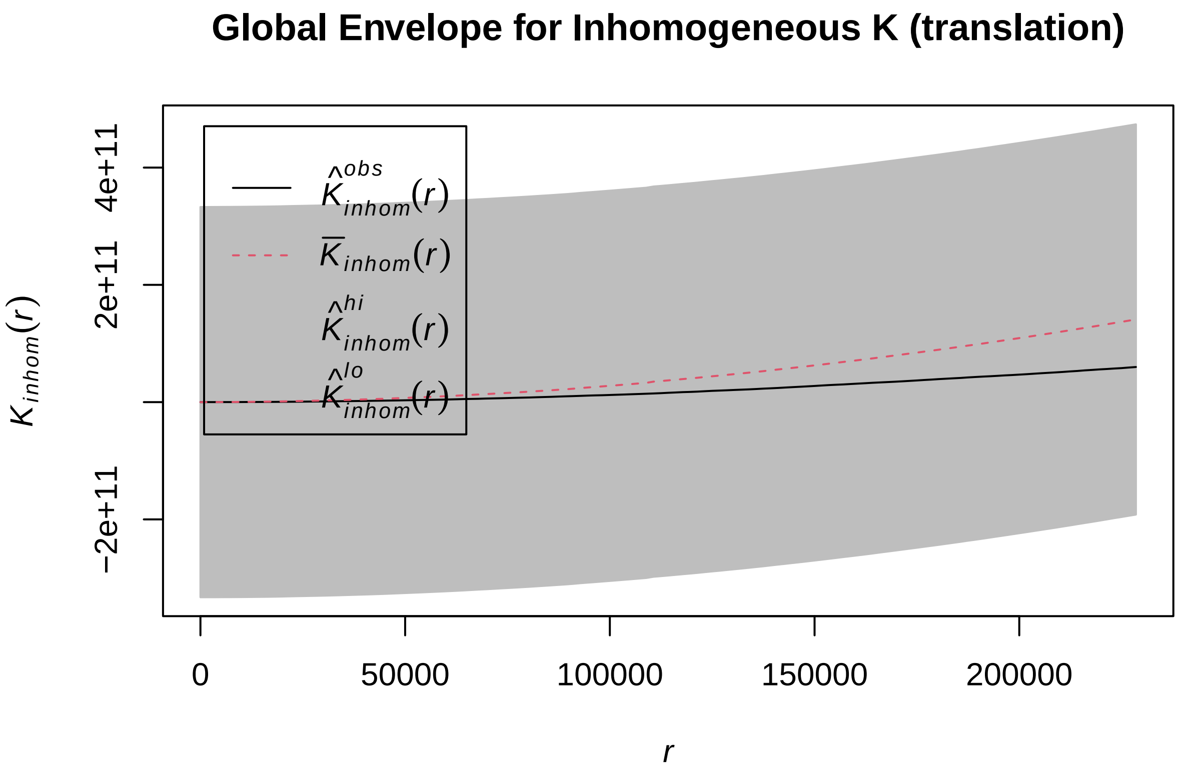
Quadrat Test for First.Order Stationarity



```
## Analysis of Deviance Table
##
## Model 1: ~1   Poisson
## Model 2: ~x + y   Poisson
##   Npar Df Deviance  Pr(>Chi)
## 1    1
## 2    3  2   143.98 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Inhomogeneous Ripley's K Function
$\hat{K}^{iso}_{inhom}(r)$
$K^{pois}_{inhom}(r)$
$K_{inhom}(r)$
0.0e+00
5.0e+10
1.0e+11
1.5e+11
0
50000
100000
150000
200000
r



The quadrat-count χ^2-test yielded $\chi^2 = 159.41$ on 13 degrees of freedom ($p < 2.2\times 10^{-16}$), and the likelihood-ratio test comparing the homogeneous Poisson model (~ 1) to the trend model ($\sim x + y$) gave a deviance of 143.98 ($p < 2.2 \times 10^{-16}$), both decisively rejecting first-order stationarity. After fitting a smooth intensity surface via Diggle's edge-corrected kernel (bandwidth selected by Diggle's criterion) and computing the inhomogeneous Ripley's K-function with translation correction, the observed $\hat{K}_{\mathrm{inhom}}(r)$ remains entirely within the 95% global Monte Carlo envelope based on 1 000 simulations of the fitted inhomogeneous Poisson process. Thus, once the large-scale intensity trend is accounted for, there is no significant residual interaction, and the point pattern is consistent with an inhomogeneous Poisson process (second-order stationarity holds).

D. Conclude by telling me whether your data are consistent with an IRP. The point pattern of air-monitoring stations clearly rejects complete spatial randomness, but after fitting and removing the smooth intensity trend the inhomogeneous Ripley's K-function remains within its global Monte Carlo envelope; hence, the data are consistent with an IRP.