GEOG 413 Unit 3

Aaron Goodman

2025-02-10

Libraries

```r
library(sf)
```

```
## Linking to GEOS 3.12.2, GDAL 3.9.3, PROJ 9.4.1; sf_use_s2() is TRUE
```

```r
library(spatstat)
```

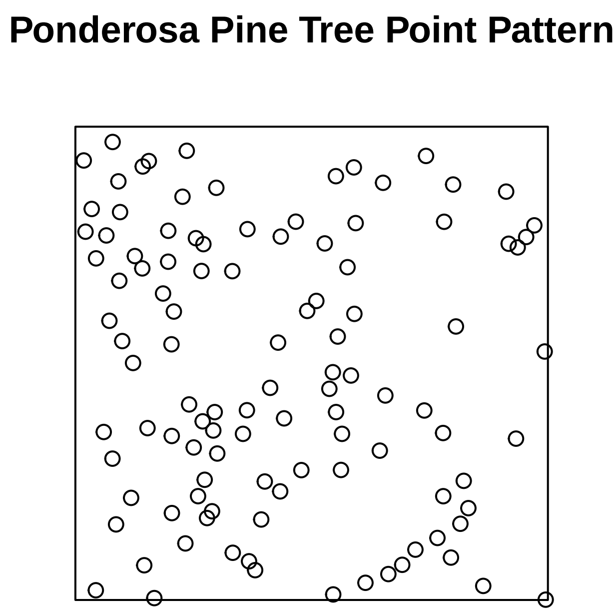
```
## Loading required package: spatstat.data

## Loading required package: spatstat.univar

## spatstat.univar 3.1-1

## Loading required package: spatstat.geom

## spatstat.geom 3.3-5

## Loading required package: spatstat.random

## spatstat.random 3.3-2

## Loading required package: spatstat.explore

## Loading required package: nlme

## spatstat.explore 3.3-4

## Loading required package: spatstat.model

## Loading required package: rpart

## spatstat.model 3.3-4

## Loading required package: spatstat.linnet

## spatstat.linnet 3.2-5

##
## spatstat 3.3-1
## For an introduction to spatstat, type 'beginner'
```

[1]. Open the spatstat data file "ponderosa".

```
data(package="spatstat.data")
data(ponderosa)
```

[A]. Plot the point pattern and interpret it.

```
plot(ponderosa, main="Ponderosa Pine Tree Point Pattern")
```

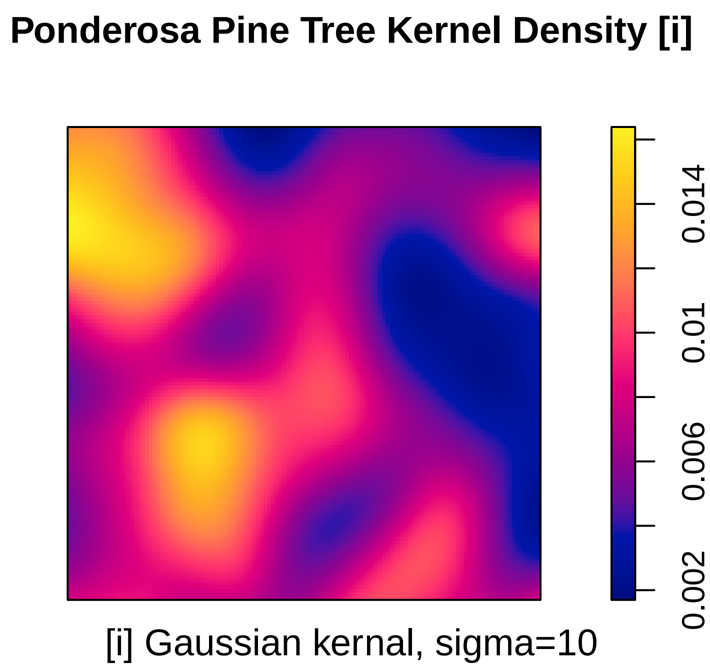


Interpretation The plot of Ponderosa pine tree locations across the given study plot shows some notable clusters of events (trees), and a considerable amount of empty gaps between clusters.

[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.

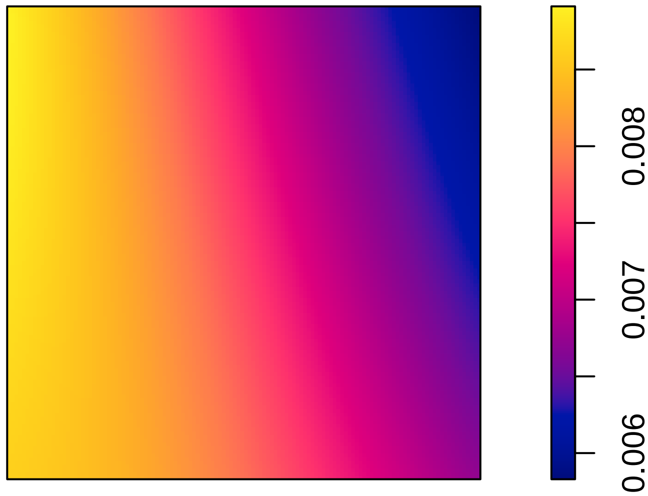
```
ponderosaDensityGauss10 <- density(ponderosa, sigma=10)
ponderosaDensityGauss50 <- density(ponderosa, sigma=50)
ponderosaDensityEpan10 <- density(ponderosa, sigma=10, kernel=c("epanechnikov"))
ponderosaDensityDisc10 <- density(ponderosa, sigma=10, kernel=c("disc"))
```

```r
plot(ponderosaDensityGauss10, main="Ponderosa Pine Tree Kernel Density [i]")
mtext("[i] Gaussian kernal, sigma=10",side=1,line=0.15,cex=1.2)
```




```r
plot(ponderosaDensityGauss50, main="Ponderosa Pine Tree Kernel Density [ii]")
mtext("[ii] Gaussian kernal, sigma=50",side=1,line=0.15,cex=1.2)
```

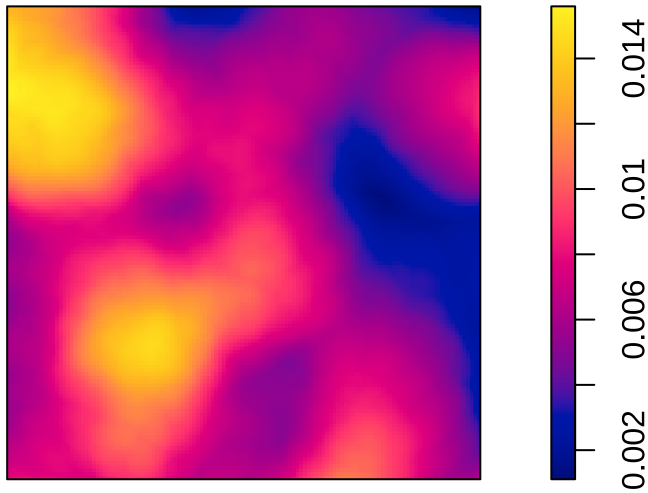
Ponderosa Pine Tree Kernel Density [ii]



[ii] Gaussian kernal, sigma=50

```r
plot(ponderosaDensityEpan10, main="Ponderosa Pine Tree Kernel Density [iii]")
mtext("[iii] Epanechnikov kernel, sigma=10",side=1,line=0.15,cex=1.2)
```

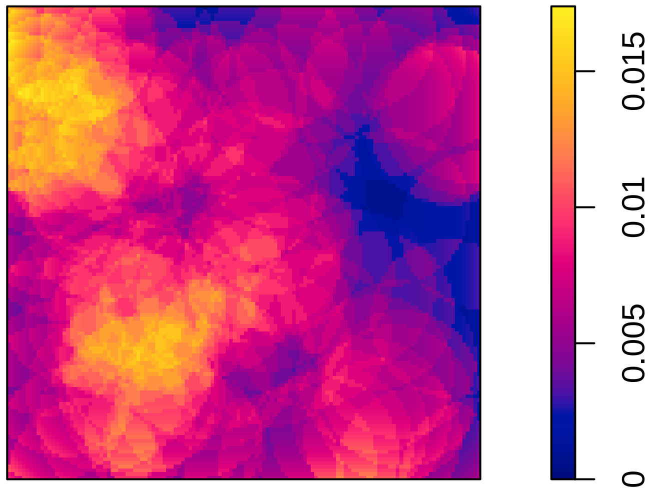
Ponderosa Pine Tree Kernel Density [iii]



[iii] Epanechnikov kernel, sigma=10

```
plot(ponderosaDensityDisc10, main="Ponderosa Pine Tree Kernel Density [iv]")
mtext("[iv] Disc kernel, sigma=10",side=1,line=0.15,cex=1.2)
```

Ponderosa Pine Tree Kernel Density [iv]



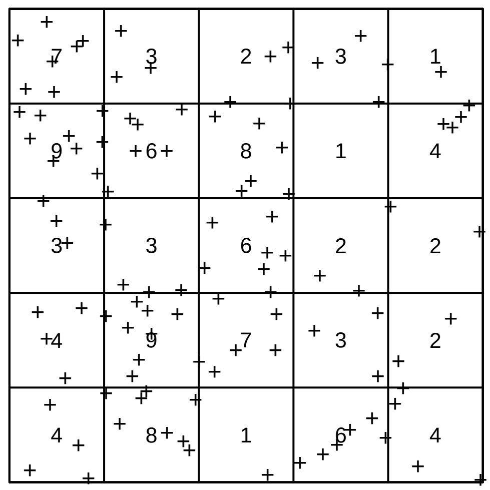
[iv] Disc kernel, sigma=10

Interpretation The Gaussian kernel density plots illustrate the effect of varying bandwidths (sigma=10, sigma=50); larger bandwidths contribute to greater smoothing at the loss of local detail. The Epanechnikov, and especially the Disc, kernel density plots show much more abrupt transitions between shaded areas of varying density.

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

```r
ponderosaQ <- quadratcount(ponderosa, nx=5, ny=5)
plot(ponderosa, cex=0.75, pch="+", main="Ponderosa Pine Tree Quadrat Count")
plot(ponderosaQ, add=TRUE, cex=0.67)
```

Ponderosa Pine Tree Quadrat Count



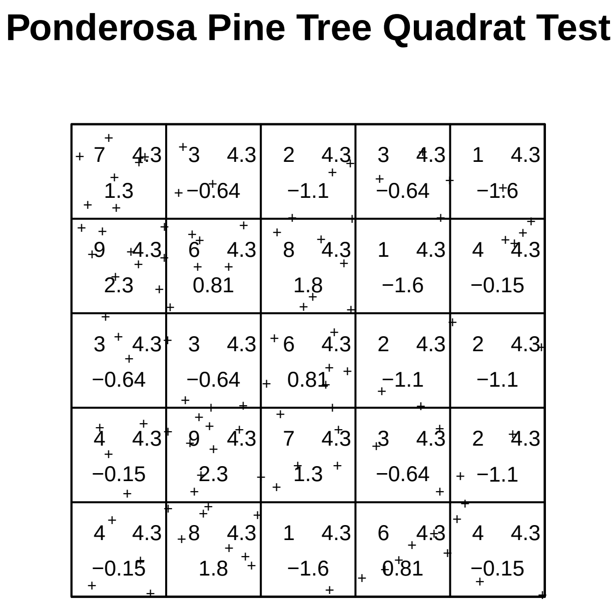
[D]. State your hypotheses, perform a quadrat test (the chi-squared test in spatstat is OK) and report whether you would reject or not reject the null hypothesis that the point pattern is consistent with an IRP at the significance level of 0.05.

Hypothesis: We assume a null hypothesis that the point pattern was produced by a homogeneous Poisson process that is consistent with Complete Spatial Randomness (CSR), ie that is IRP.

```
plot(ponderosa, cex=0.5, pch="+", main="Ponderosa Pine Tree Quadrat Test")
ponderosaTest <- quadrat.test(ponderosa, nx=5, ny=5)
```

```
## Warning: Some expected counts are small; chi^2 approximation may be inaccurate
```

```
plot(ponderosaTest, add=TRUE, cex=0.67)
```




```r
ponderosaTest
```

```
##
##  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
```

Interpretation The quadrat counts in the 5x5 grid show considerable variation in tree density (many with 1-2, some with 7-9). However, Chi-squared test results show p-value = 0.10, which is greater than 0.05, so we fail to reject the null hypothesis (CSR) at 5% significant level.

[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.

```r
# summary(ponderosa)
# 14400 square meters, n=108 points
ponderosaNNDist <- nndist(ponderosa)
```

```r
dbarobs <- sum(ponderosaNNDist)/length(ponderosaNNDist)
# dbarobs
dbarexp <- 0.5/sqrt(108/14400)
# dbarexp
sddbarexp <- 0.26136/sqrt(108^2/14400)
# sddbarexp
indexNN <- dbarobs/dbarexp
indexNN
```

```
## [1] 1.184248
```

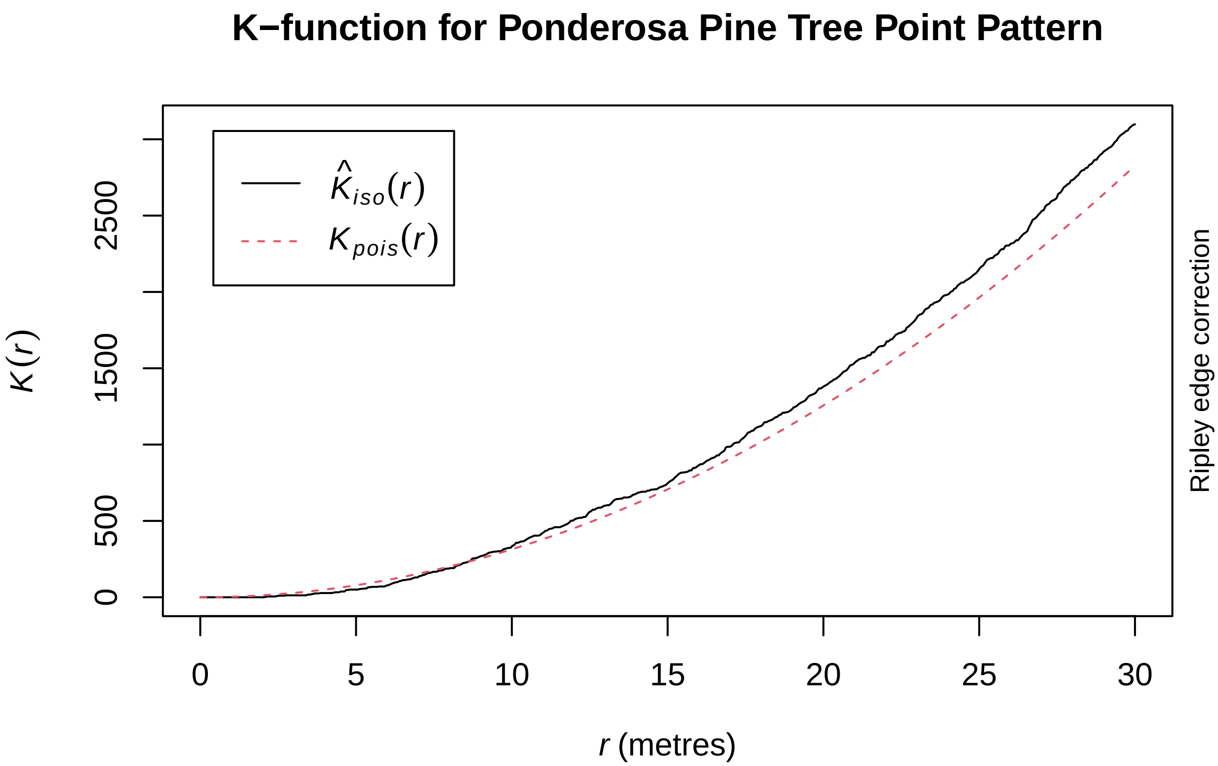
```r
zTest <- (dbarobs - dbarexp)/sddbarexp
zTest
```

```
## [1] 3.663073
```

Interpretation Nearest neighbor index value $R = 1.18$ suggests only a slight tendency towards uniformity ($R > 1$). However, high z-score of 3.663 indicates that the departure from randomness (towards uniformity) is statistically significant at 5% significance level ($p < 0.05$). Thus, there is sufficient evidence to reject the null hypothesis.

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

```r
kf <- Kest(ponderosa, correction='Ripley')
plot(kf, main="K-function for Ponderosa Pine Tree Point Pattern")
mtext("Ripley edge correction",side=4,line=0.15,cex=0.8)
```



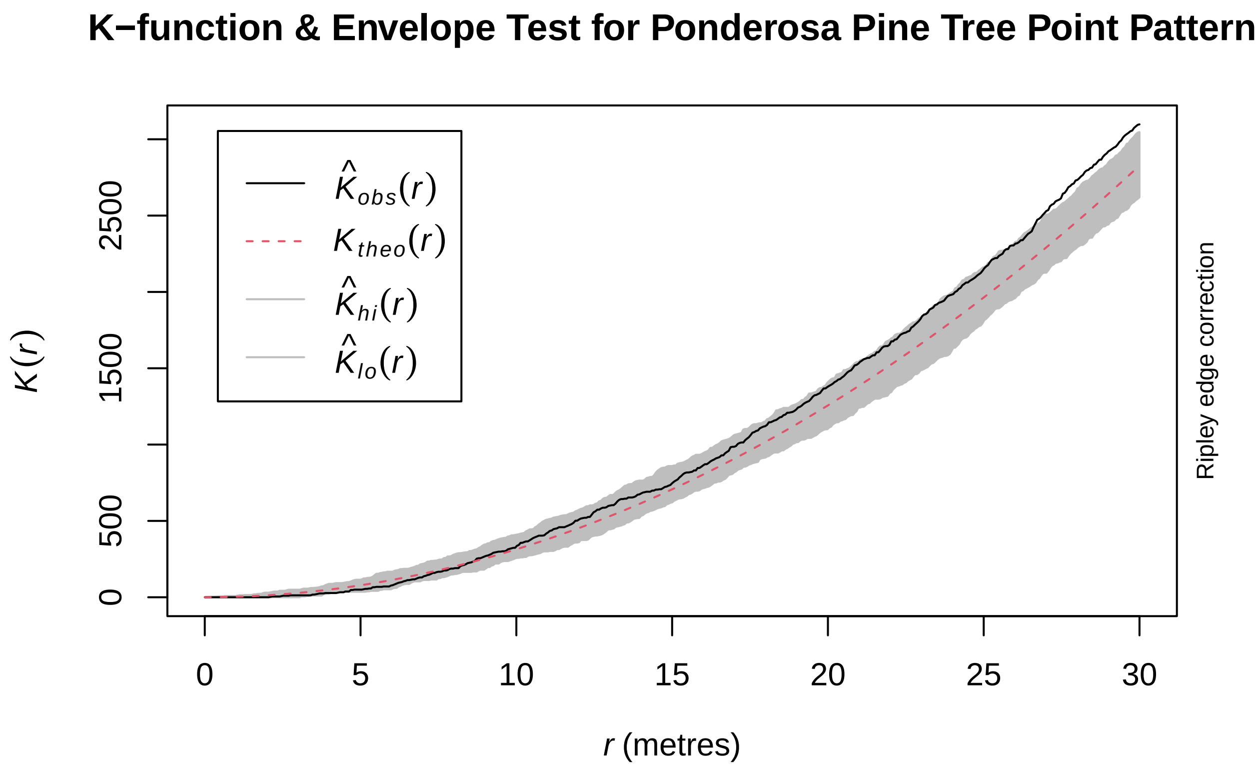
Interpretation Up until around r = 17 metres, K_iso (K_observed) lies very close to K_pois (K_theoretical). Beyond 17 metres, K_iso exhibits higher values than expected under CSR (K_pois). This suggests clustering in the point pattern at larger scales.

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

```
kfenv <- envelope(ponderosa, Kest, correction='Ripley')
```

```
## Generating 99 simulations of CSR  ...
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
## 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40,
## 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
## 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80,
## 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98,
## 99.
##
## Done.
```

```
plot(kfenv, main="K-function & Envelope Test for Ponderosa Pine Tree Point Pattern")
mtext("Ripley edge correction",side=4,line=0.15,cex=0.72)
```



Interpretation The plot of the confidence envelope for the point pattern's K-function shows that K_obs falls beyond the grey envelope (K_hi, K_lo) at several r values above 17 metres. This departure from the confidence envelope indicates statistically significant deviation from CSR.

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

```r
mad.test(ponderosa, Kest)
```

```
## Generating 99 simulations of CSR  ...
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
## 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40,
## 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
## 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80,
## 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98,
## 99.
##
## Done.
```

```
##
##  Maximum absolute deviation test of CSR
##  Monte Carlo test based on 99 simulations
##  Summary function: K(r)
```

```
##  Reference function: theoretical
##  Alternative: two.sided
##  Interval of distance values: [0, 30] metres
##  Test statistic: Maximum absolute deviation
##  Deviation = observed minus theoretical
##
## data:  ponderosa
## mad = 284.95, rank = 3, p-value = 0.03
```

Interpretation With a p-value of 0.02 (less than 0.05), we reject the null hypothesis of CSR at 5% significance level ($p < 0.05$). The MAD test's rank of 2 indicates that only one simulated point pattern exhibited a larger deviation from Complete Spatial Randomness than our observed 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.

```
stations <- st_read("../data/calairsites/airmonitoringstations.shp")
```

```
## Reading layer 'airmonitoringstations' from data source
##   'F:\MAGIST\23_25w_GEOG_413_AppliedGeospatialStats\U3_Density-and-Distance\data\calairsites\airmoni
##   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
```

```
stationsUS <- stations[!(stations$ZIPCODE %in% c("xxxxx")),]
districts <- st_read("../data/calairsites/arb_california_airdistricts_aligned_03.shp")
```

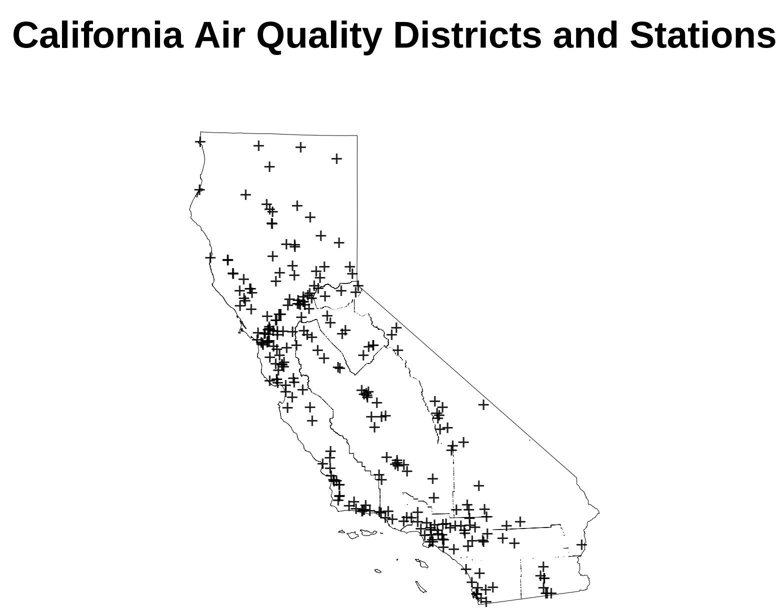
```
## Reading layer 'arb_california_airdistricts_aligned_03' from data source
##   'F:\MAGIST\23_25w_GEOG_413_AppliedGeospatialStats\U3_Density-and-Distance\data\calairsites\arb_cali
##   using driver 'ESRI Shapefile'
## Simple feature collection with 47 features and 8 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: -373976.6 ymin: -604526.1 xmax: 540015.4 ymax: 450070.9
## Projected CRS: NAD83 / California Albers
```

[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".)

```r
districts.owin <- as.owin(districts)
stationsUS.ppp <- as.ppp(stationsUS)
stationsUS.ppp <- unmark(stationsUS.ppp)
```

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

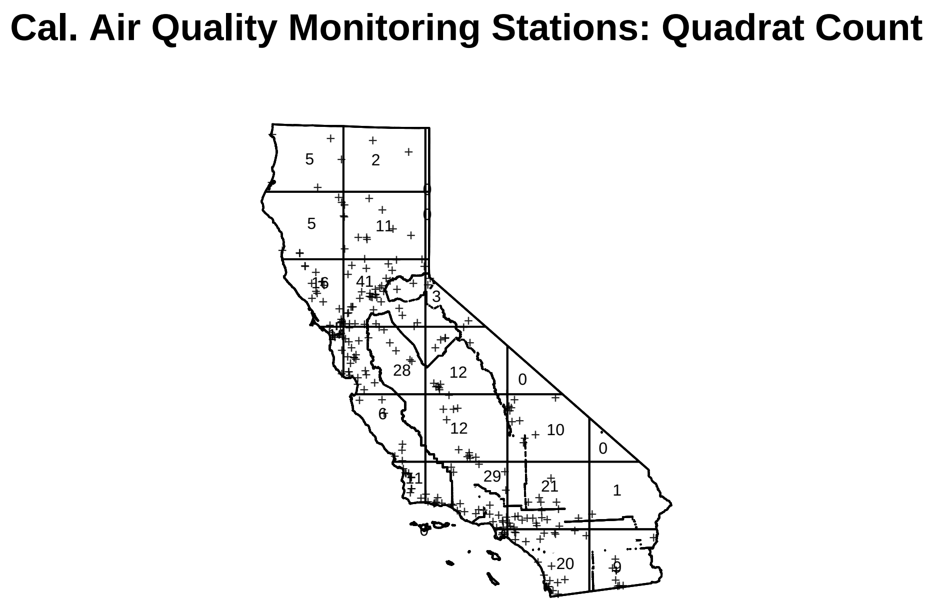
```r
stationsUS.ppp$window <- districts.owin
plot(stationsUS.ppp, cex=0.67, pch="+", lwd="0.125", main="California Air Quality Districts and Stations
```



Interpretation The plot of air quality monitoring stations across Air Quality Monitoring Districts (AQMD) of California shows high concentrations of monitoring stations in urban centers like Los Angeles and the San Francisco Bay.

[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.

```r
# quadrat count
plot(stationsUS.ppp, cex=0.5, pch="+", main="Cal. Air Quality Monitoring Stations: Quadrat Count")
calairQCount <- quadratcount(stationsUS.ppp, nx=5, ny=7)
plot(calairQCount, add=TRUE, cex=0.5)
```

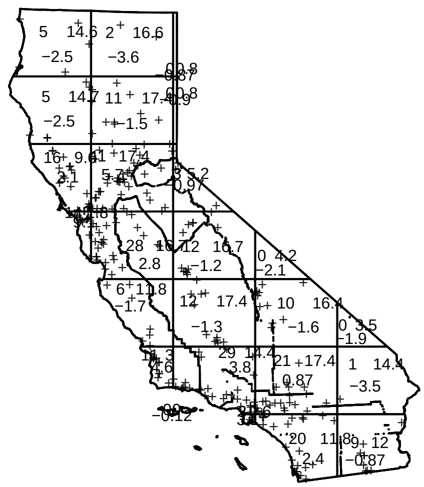



```r
# quadrat test
plot(stationsUS.ppp, cex=0.5, pch="+", main="Cal. Air Quality Monitoring Stations: Quadrat Test")
calairQTest <- quadrat.test(stationsUS.ppp, nx=5, ny=7)
```

```
## Warning: Some expected counts are small; chi^2 approximation may be inaccurate
```

```r
plot(calairQTest, add=TRUE, cex=0.5)
```

Cal. Air Quality Monitoring Stations: Quadrat Test

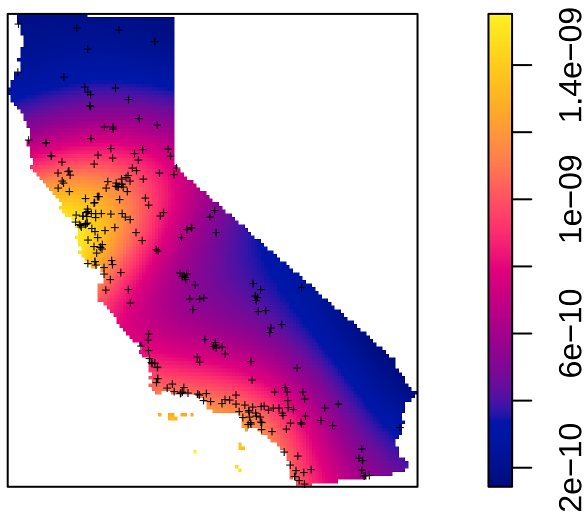


```r
calairQTest
```

```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:  stationsUS.ppp
## X2 = 238.88, df = 24, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 25 tiles (irregular windows)
```

```r
# kernel density estimate
calairKDE <- density(stationsUS.ppp)
plot(calairKDE, main="Cal. Air Quality Monitoring Stations: Kernel Density")
plot(stationsUS.ppp, cex=0.5, pch="+", add=TRUE)
```

Cal. Air Quality Monitoring Stations: Kernel Density




```r
# nearest neighbor analysis
## summary(stationsUS.ppp)
## 4.10598e+11 square units (410598000000), n=259 points
calairNNDist <- nndist(stationsUS.ppp)
dbarobsc <- sum(calairNNDist)/length(calairNNDist)
## dbarobsc
dbarexpc <- 0.5/sqrt(259/410598000000)
## dbarexpc
sddbarexpc <- 0.26136/sqrt(259^2/410598000000)
## sddbarexpc
calIndexNN <- dbarobsc/dbarexpc
calIndexNN
```

```
## [1] 0.7434767
```

```r
calZTest <- (dbarobsc - dbarexpc)/sddbarexpc
calZTest
```

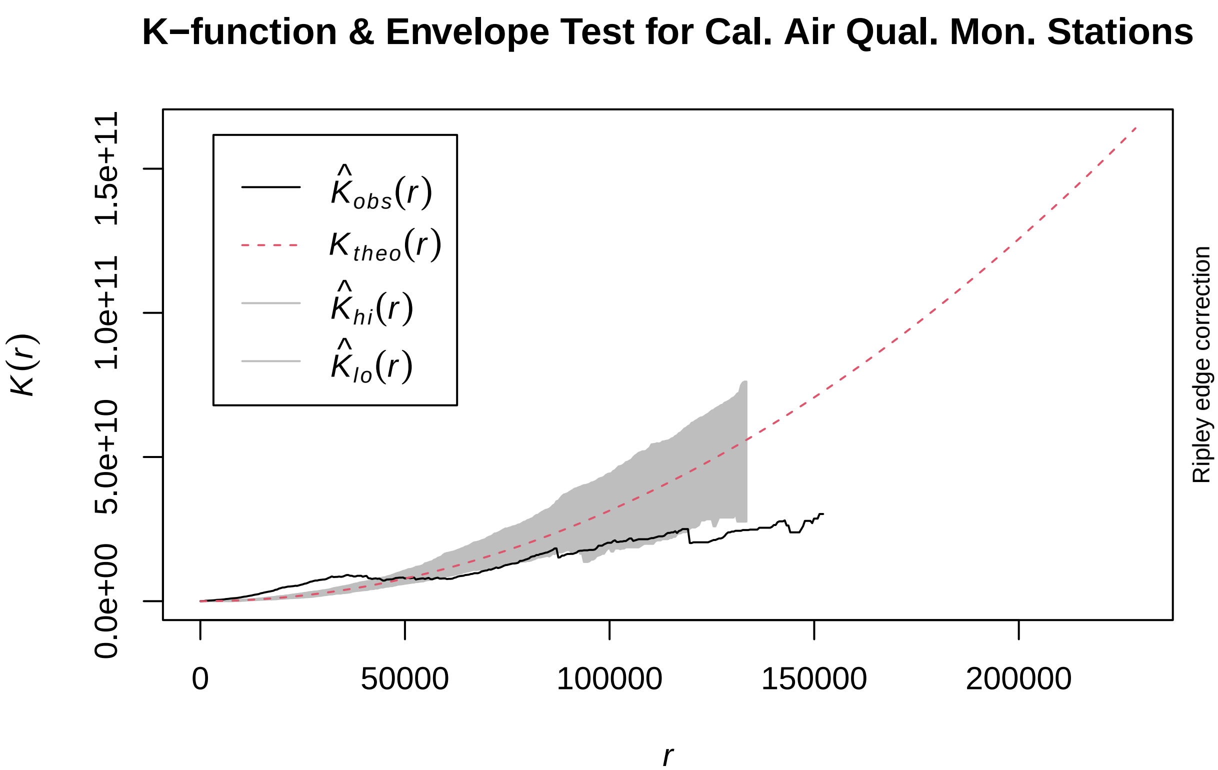
```
## [1] -7.897826
```

```r
# K-function, confidence envelope
calairKenv <- envelope(stationsUS.ppp, Kest, correction='border')
```

```
## Generating 99 simulations of CSR  ...
```

```
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
## 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40,
## 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
## 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80,
## 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98,
## 99.
##
## Done.
```

```r
plot(calairKenv, main="K-function & Envelope Test for Cal. Air Qual. Mon. Stations")
mtext("Ripley edge correction",side=4,line=0.15,cex=0.72)
```



[D]. Conclude by telling me whether your data are consistent with an IRP.

Interpretation

Quadrat Count / Test The quadrat counts in the 5x7 grid show considerable variation in station density. Additionally, Chi-squared quadrat test results show p-value $< 2.2e\text{-}16$, which is far less than 0.05, so we reject the null hypothesis (CSR) at 5% significant level.

Nearest Neighbor Analysis Nearest neighbor index value $R = 0.74$ suggests considerable tendency towards clsutering ($R < 1$). The extremely low z-score of -7.898 indicates that the departure from randomness (towards clustering) is statistically significant at 5% significance level ($p < 0.05$). Thus, there is sufficient evidence to reject the null hypothesis.

K-function & Envelope Test The plot of the confidence envelope for the point pattern's K-function shows that K_obs exceeds K_theo for lower r values, falling above the grey confidence envelope. K_obs intersects with K_theo around r = 50000, after which K_obs tracks towards the bottom of the grey confidence envelope, before finally deviating from it completely around r = 125000. This departure indicates statistically significant deviation from CSR.