Assignment 6: Point Pattern Analysis

Your Name

Date

Instructions

Download the assignment06.Rmd file from Canvas and open it in RStudio. Complete this assignment by filling in the answers below in the R Markdown Notebook document.

Data
In this assignment, you will apply point pattern analysis to the PUF Lands drilling pad measurements that you collected for Assignment #4. To remove the effects of double counting, pads that are within 500 meters of any other pads have been removed. To remove any spatial distortion, latitude/longitude coordinates have been converted to UTM Zone 14 coordinates (given in meters). If you are not familiar with the UTM coordinate system, read about it here.

The drilling pad data are contained within two CSV files called "andrews.csv" and "CRIS.csv", one for each survey area. They are located at the following URLs:
http://people.tamu.edu/~geoallen/courses/312/andrews.csv
http://people.tamu.edu/~geoallen/courses/312/CRIS.csv
Each CSV contains the following variables: (1) approach; (2) x; (3) y; (4) area_m2 (the area of the pad).

Additionally, quadrat count information from the CRIS survey can be found at the following URL: http://people.tamu.edu/~geoallen/courses/312/quadCounts_CRIS.csv

Deliverables

Please submit to Canvas the following items:
1. An HTML (or Word or PDF) file knitted from the .Rmd file
2. A completed .Rmd file

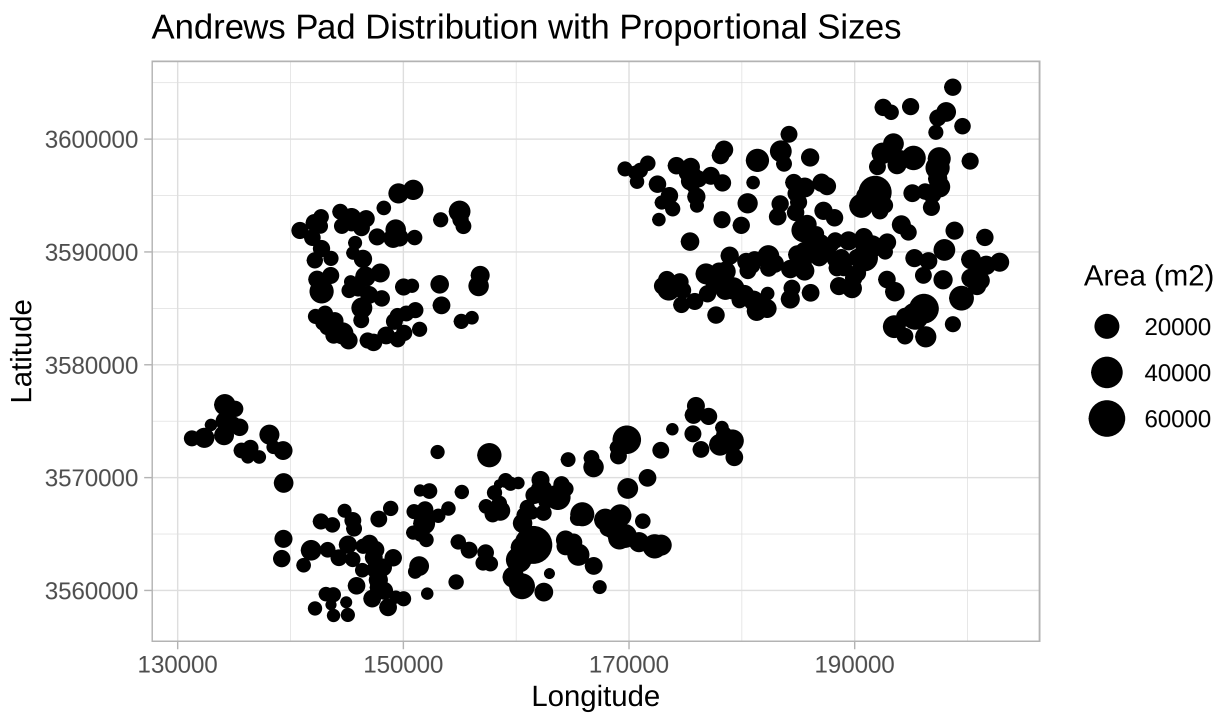
Questions

1. Read in the andrews.csv file into RStudio directly from the URL address provided above and assign it to the variable "andrews". In a single figure, plot the spatial distribution of drilling pads. Scale the graphical symbols by the area of the pad (e.g., larger pads should have a larger point). As usual, all graphics created in this assignment should be done using the ggplot2 R package and will be graded based on the highest standards of data visualization. Tip: use the ggplot function coord_fixed() to plot the map without distortion.

```
path = "http://people.tamu.edu/~geoallen/courses/312/andrews.csv"
andrews = read.csv(path, header=T)
library(ggplot2)
colorGood <- c("#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7", "#999999")

ggplot() +
  geom_point(data=andrews, aes(x=x, y=y, size=area_m2)) +
  coord_fixed() +
  theme_light() +
```

```
  scale_fill_manual(values = colorGood) +
  labs(x = "Longitude",
       y = "Latitude",
       size = "Area (m2)",
       title = "Andrews Pad Distribution with Proportional Sizes")
```

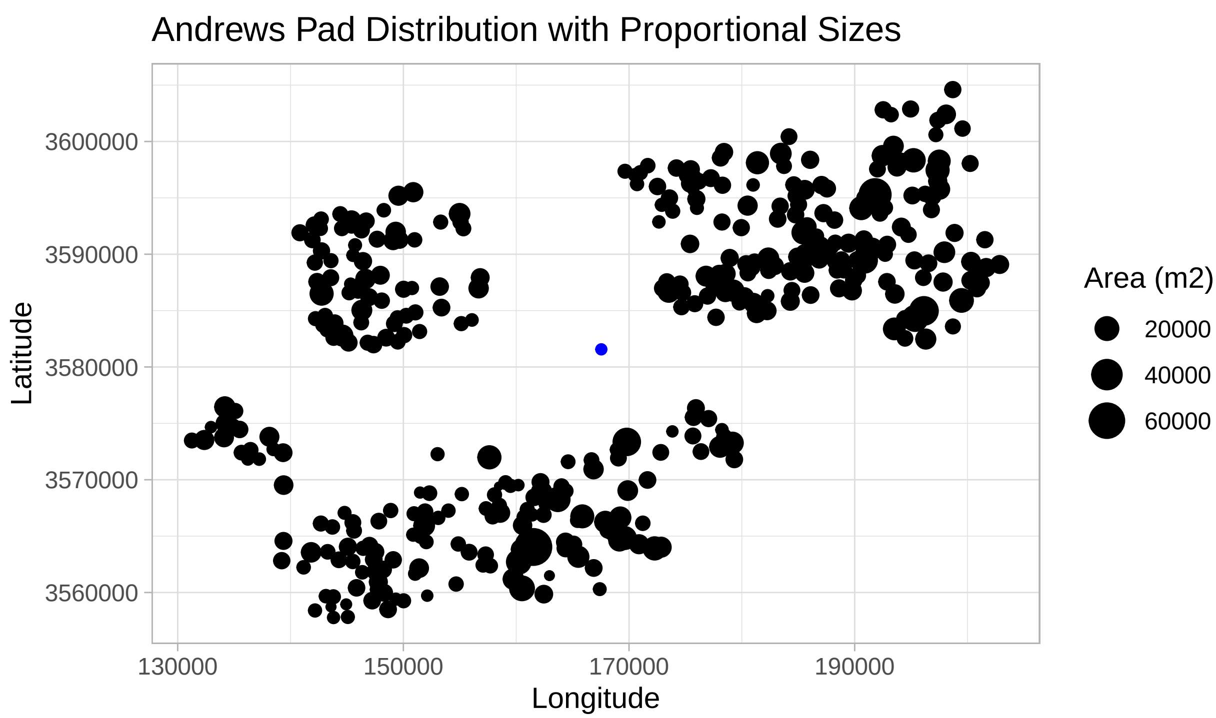


2. Calculate the location of the mean center of the pads. Add the mean center to your map of Andrews pads as a *blue* point. Hint: to plot data in ggplot, remember that your data must be in a data frame (see lecture R code for an example).

```
mu_x = mean(andrews$x, na.rm=T)
mu_y = mean(andrews$y, na.rm=T)

center = data.frame(mu_x = mu_x, mu_y = mu_y)

ggplot() +
  geom_point(data=andrews, aes(x=x, y=y, size=area_m2)) +
  geom_point(data=center, aes(x=mu_x, y=mu_y), color="blue") +
  coord_fixed() +
  theme_light() +
  scale_fill_manual(values = colorGood) +
  labs(x = "Longitude",
       y = "Latitude",
       size = "Area (m2)",
       title = "Andrews Pad Distribution with Proportional Sizes")
```

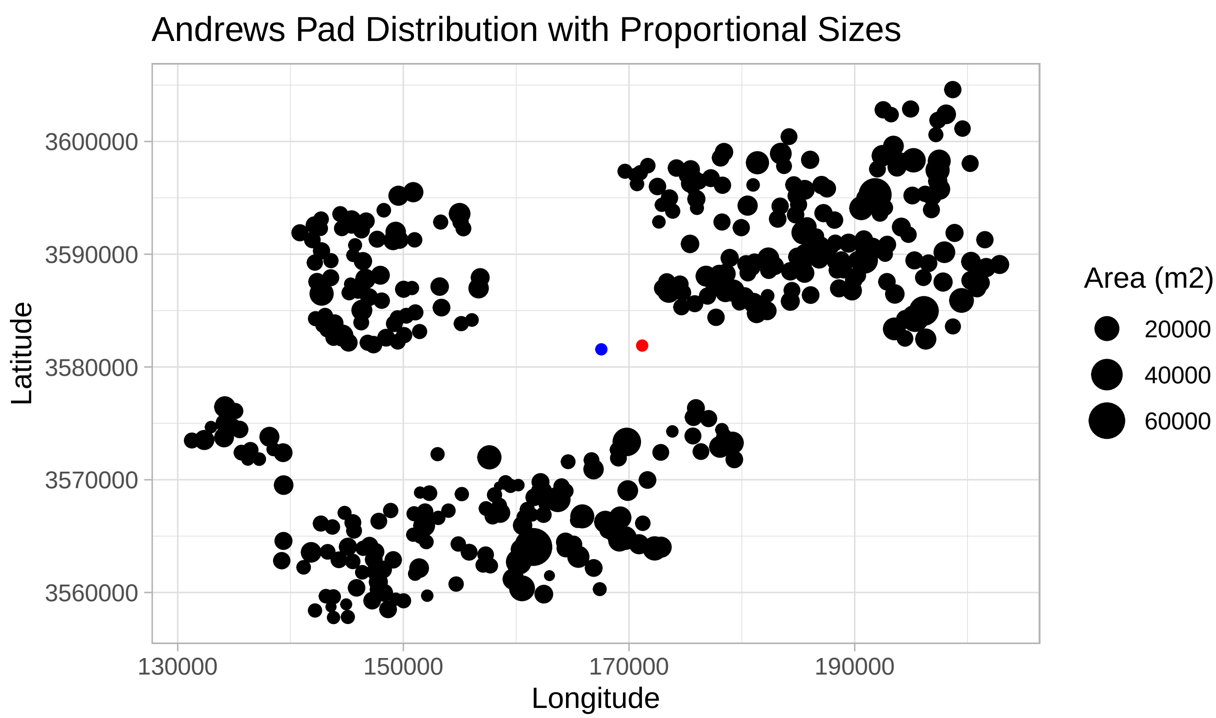


3. Calculate the location of the weighted mean center of the pads and add the weighted mean center to your map of Andrews pads as a *red* point. Use pad area as the weight. Your map should have the location of pads, the mean center, and the weighted mean center. What does the relative position of the mean center and relative mean center tell you about pads in the Andrews survey area?

```
mw_x = weighted.mean(andrews$x, w = andrews$area_m2, na.rm=T)
mw_y = weighted.mean(andrews$y, w = andrews$area_m2, na.rm=T)

weight_center = data.frame(mw_x = mw_x, mw_y = mw_y)


ggplot() +
  geom_point(data=andrews, aes(x=x, y=y, size=area_m2)) +
  geom_point(data=center, aes(x=mu_x, y=mu_y), color="blue") +
  geom_point(data=weight_center, aes(x=mw_x, y=mw_y), color="red") +
  coord_fixed() +
  theme_light() +
  scale_fill_manual(values = colorGood) +
  labs(x = "Longitude",
       y = "Latitude",
       size = "Area (m2)",
       title = "Andrews Pad Distribution with Proportional Sizes")
```



4. Calculate the standard distance of the Andrew pads and assign it to the variable d. Print d and plot it as a summary circle with radius equal to d centered on the mean center (blue point). Note: use the circleFun function below to plot a circle in ggplot (use the geom_path() ggplot function). Your circle should not be an ellipse.

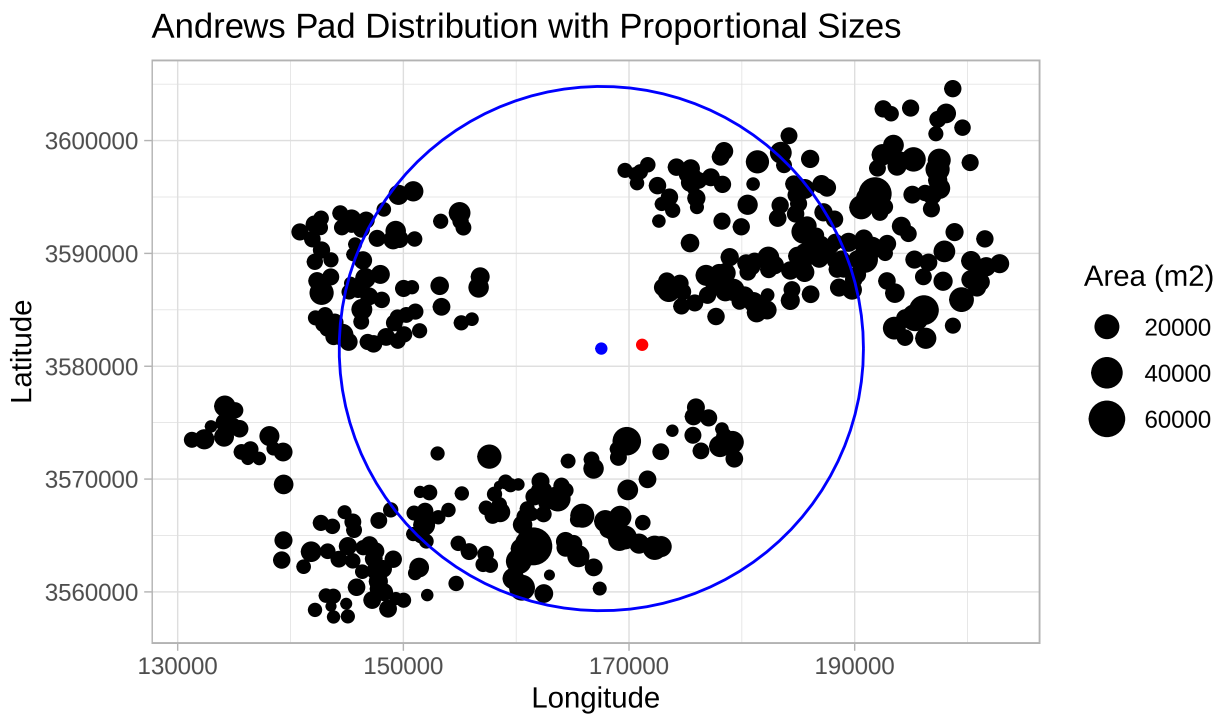
```
circleFun = function(center_x=0, center_y=0, radius=1){
  circle_t = seq(0, 2*pi, length.out=100)
  circle_x = center_x + radius * cos(circle_t)
  circle_y = center_y + radius * sin(circle_t)
  return(data.frame(x = circle_x, y = circle_y))
}

d = sqrt(sum(((andrews$x-mu_x)^2 + (andrews$y-mu_y)^2)/(nrow(andrews)), na.rm=T))

circleCoordinates = circleFun(center_x=mu_x, center_y=mu_y, radius=d)

ggplot() +
  geom_point(data=andrews, aes(x=x, y=y, size=area_m2)) +
  geom_point(data=center, aes(x=mu_x, y=mu_y), color="blue") +
  geom_point(data=weight_center, aes(x=mw_x, y=mw_y), color="red") +
  geom_path(data=circleCoordinates, aes(x=x, y=y), color="blue") +
  coord_fixed() +
  theme_light() +
  scale_fill_manual(values = colorGood) +
  labs(x = "Longitude",
       y = "Latitude",
```

```
       size = "Area (m2)",
       title = "Andrews Pad Distribution with Proportional Sizes")
```



5. (Bonus 5 pts) The study area of the Andrews is 3352 km2 and the study area of CRIS is 1788 km2. Calculate and print the relative distance of Andrews and CRIS pads. What does the difference between the relative distance between the two study areas tell us about the spatial distribution of pads?

```
path1 = "http://people.tamu.edu/~geoallen/courses/312/CRIS.csv"
cris5 = read.csv(path, header=T)

dCris = sqrt(sum(((cris5$x-mu_x)^2 + (cris5$y-mu_y)^2)/(nrow(cris5)), na.rm=T))

#Rearranging circle area equation: A = pi * r^2 -> sqrt(A) = sqrt(pi) * r -> sqrt(A) / sqrt(pi) = r
andrRadius = sqrt(3552) / sqrt(pi)
crisRadius = sqrt(1788) / sqrt(pi)

andrDist = d / andrRadius
crisDist = dCris / crisRadius

print(paste0("Andrews Relative Distance: ", andrDist))
```

```
## [1] "Andrews Relative Distance: 690.945077094169"
```

```
print(paste0("CRIS Relative Distance: ", crisDist))
```

```
## [1] "CRIS Relative Distance: 973.859372281019"
```

As we have the relative distances to remove the difference in study areas as a factor, we can now take a look at how the standard differences actually vary. Despite having a much smaller area, the distance of the CRIS point-set is much higher, indicating a relatively much more spread and sparse distribution despitethe smaller area.

6. Splitting the CRIS study area into 100 quadrats, we can count the number of pads in each quadrat. The quadrant counts are contained in a CSV file located at the following URL: http://people.tamu.edu/~geoallen/courses/312/quadCounts_CRIS.csv Using the data in quadCounts_CRIS.csv, demonstrate whether the pads are spatially dispersed, randomly distributed, or clustered at the 95% confidence interval. Print the Variance-to-Mean Ratio (VMR) and the corresponding Chi-Squared value. Hint: You can refer to the Chi-Squared table in the lecture notes or use R's built in Chi-Squared test.

(http://people.tamu.edu/~geoallen/courses/312/quadrats.png) Note: Image embed removed after all questions completed due to issues with PDF knitting.

```
path2 = "http://people.tamu.edu/~geoallen/courses/312/quadCounts_CRIS.csv"
CRIS = read.csv(path2, header=T)

CRIS_mean = sum(CRIS$count) / nrow(CRIS)
CRIS_var = var(x=CRIS$count)
CRIS_VMR = CRIS_var / CRIS_mean
print(paste0("VMR of CRIS Study = ", CRIS_VMR))
```

```
## [1] "VMR of CRIS Study = 2.19142645971914"
```

```
print("As the VMR > 1, this indicates that the points are clustered")
```

```
## [1] "As the VMR > 1, this indicates that the points are clustered"
```

```
ctest = chisq.test(CRIS$count)
```

```
## Warning in chisq.test(CRIS$count): Chi-squared approximation may be incorrect
```

```
ctest
```

```
##
##  Chi-squared test for given probabilities
##
## data:  CRIS$count
## X-squared = 216.95, df = 99, p-value = 8.264e-11
```

```
print("As the p-value for this test is substantially smaller than 0.05, we can ")
```

```
## [1] "As the p-value for this test is substantially smaller than 0.05, we can "
```

```
print("reject the null hypothesis that the points are not clustered.")
```

```
## [1] "reject the null hypothesis that the points are not clustered."
```

7. Examining the figure provided in the question above, does the point pattern of pads appear to be mainly controlled by first-order effects or second-order effects? Why? Please provide your answer in one sentence.

I would interpret that the clustering of the pads is based on relatively higher efficiency of extraction in the clustered areas for whatever reason, but the pads do not interact or care much about eachother, therefore the distribution is much more likely to be primarily affected by First-Order Effects.