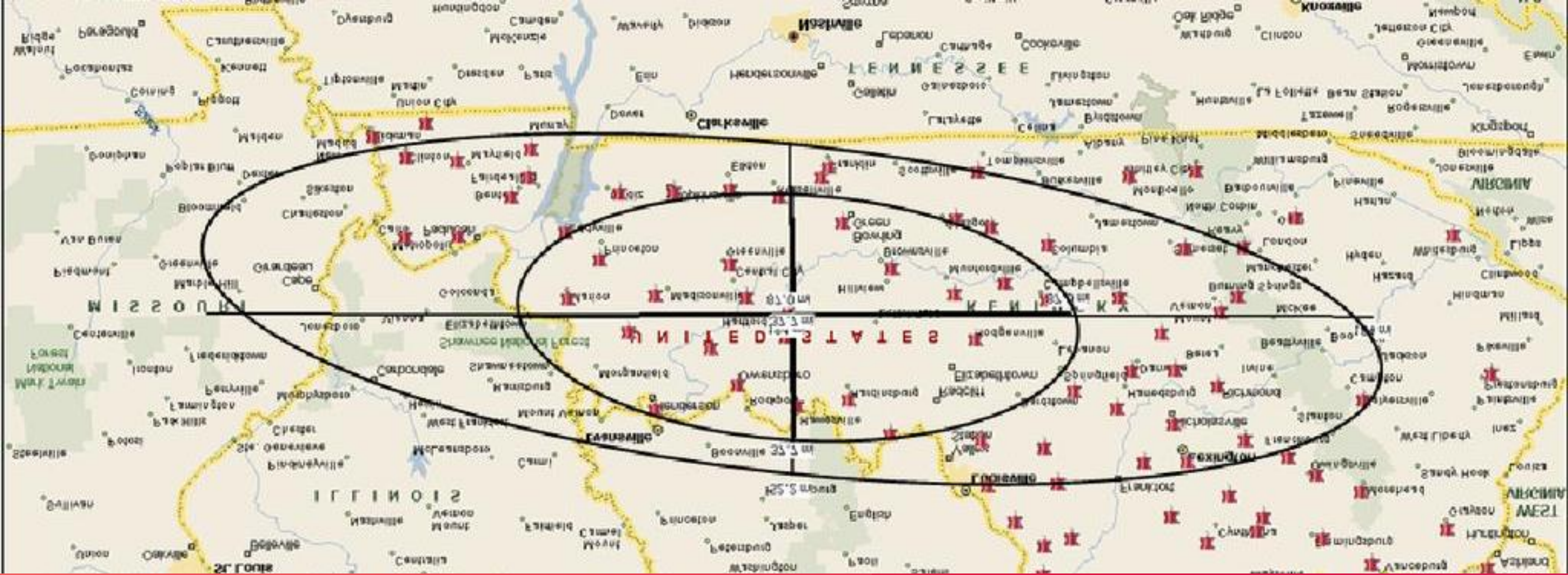


ECON6027 4A

SPATIAL DESCRIPTIVE SUMMARY MEASURES: THEORY

USING STATISTICAL MEASURES TO ANALYSE DATA DISTRIBUTIONS

This lesson is an overview of the statistical and spatial statistical methods that are commonly used in summarising data.

Understanding statistical distributions of a dataset helps us to gain fundamental knowledge to move the analysis forward.

All statistical approaches typically begin with a comprehensive evaluation of the spectrum of data values obtained for each variable in the dataset.

WHY DESCRIPTIVE MEASURES?

1. Organise and visualise data

2. Offer suggestive clues about patterns and trends in data

3. Help generate new research hypothesis

EXERCISE A

1. EXPLAIN THE DIFFERENCE BETWEEN DESCRIPTIVE AND INFERENTIAL STATISTICS.
2. WHAT ARE THE TWO MAIN METHODS OF RUNNING STATISTICAL INFERENCES?

NUMERICAL DESCRIPTIVE SUMMARY MEASURES

Central Tendency

- Mean
- Median
- Mode
- Weighted mean
- Geometric mean

Non-central tendency

- Quartiles
- Deciles
- Percentiles

Dispersion

- Range
- Variance
- Standard deviation
- Coefficient of variation
- Interquartile range

Shape

- Skewness
- Kurtosis
- Higher order moments

"SPATIAL" DESCRIPTIVE SUMMARY MEASURES

Variants of traditional procedures with modifications to cope with unique properties of spatial data.

Objective: to provide numerical summary measures of the "geo-reference system" used (usually the X- and Y-coordinates) to explain any spatial distribution of the events.

Spatial characteristics are commonly explained using

Spatial measures of central tendency	*Spatial* measures of dispersion

SPATIAL MEASURES OF CENTRAL TENDENCY

- Spatial mean (mean center): center of gravity
- Weighted spatial mean (weighted mean center)
- Spatial median (median center)

SPATIAL MEAN

- Average value of observed points for each X- and Y- coordinates.
- It shows the central point **(centre of gravity)** of spatial distribution of events and are sensitive to outlying observations.
- This tool **requires projected data** to accurately measure distances (distances must be in metric units).
- For polygon and line data we can use st_centroid() to find the centre of gravity point. For point data we will use functions in the "aspace" package
- All the values for X and Y are separately summed up and divided by the total number of events/observations as:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \text{ and } \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

WEIGHTED SPATIAL MEAN

- For spatial data, the **weights** represent the **frequency and magnitude** of the events observed at a given location.
- We will weigh each of the location coordinates by the frequency values (or any other variable that measures the magnitude) of the observations in those locations.
- The weighted spatial mean can capture the spatial variations and is pulled towards the weighted points with the highest quantity (hence, less sensitive to outlying observations compared to spatial mean).

$$\bar{X} = \frac{\sum_{i=1}^{n} W_i X_i}{\sum_{i=1}^{n} W_i} \text{ and } \bar{Y} = \frac{\sum_{i=1}^{n} W_i Y_i}{\sum_{i=1}^{n} W_i}$$

SPATIAL MEDIAN

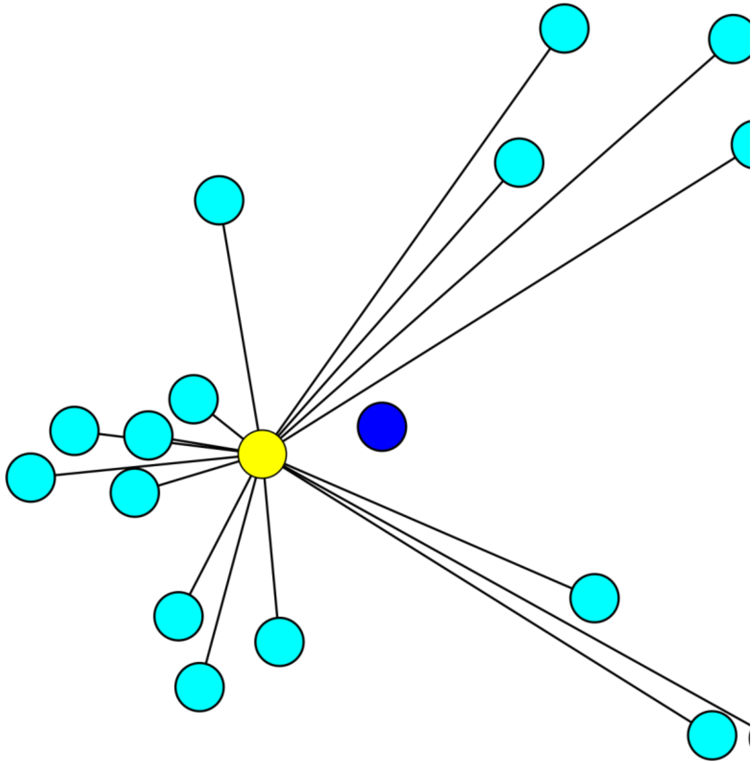
The spatial median measure will **minimise the sum of absolute distances** toward the same points.

- **E.g.: Geometric median** (in yellow) of a series of points. In blue the spatial mean.

The **data must be projected** to accurately measure distances.

Spatial median is an efficient way to estimate the central location parameter of a statistical population.

Less influenced by outliers (resistant measure).



SPATIAL MEASURES OF DISPERSION

- Standard distance
- Weighted standard distance
- Standard deviational ellipse

STANDARD DISTANCE

- Measures the extent to which observation points are dispersed around the spatial mean.
- The spatial counterpart of the standard deviation.
- Valuable statistic to understanding how compact observation points are distributed around their spatial mean.
- Sensitive to outlying observations.

$$SDD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2 + \sum_{i=1}^{n}(y_i - \bar{Y})^2}{n}}$$

WEIGHTED STANDARD DISTANCE

This measure is produced by weighting the sum of squared differences of x- and y- coordinates:

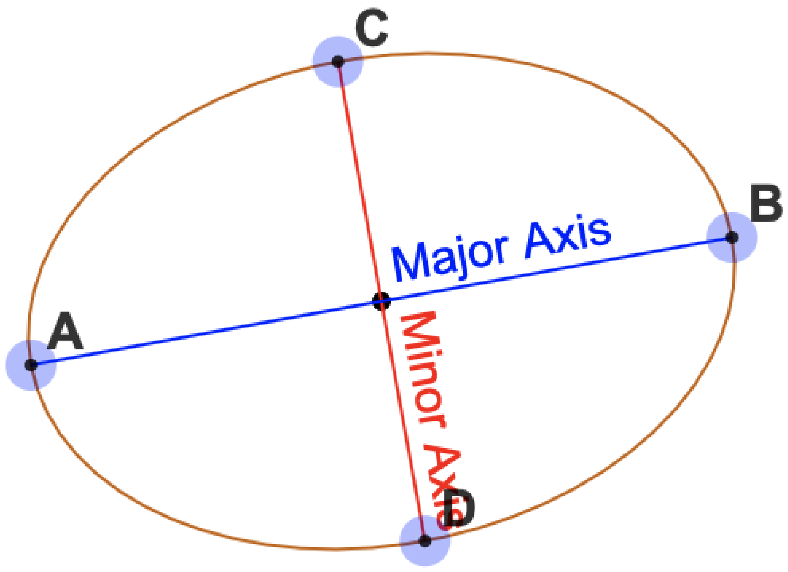
$$SDD_w = \sqrt{\frac{\sum_{i=1}^{n} w_i(x_i - \bar{X})^2 + \sum_{i=1}^{n} w_i(y_i - \bar{Y})^2}{\sum_{i=1}^{n} W_i}}$$

In a typical dataset where weights are given by frequency,

$$\sum_{i=1}^{n} W_i = n$$

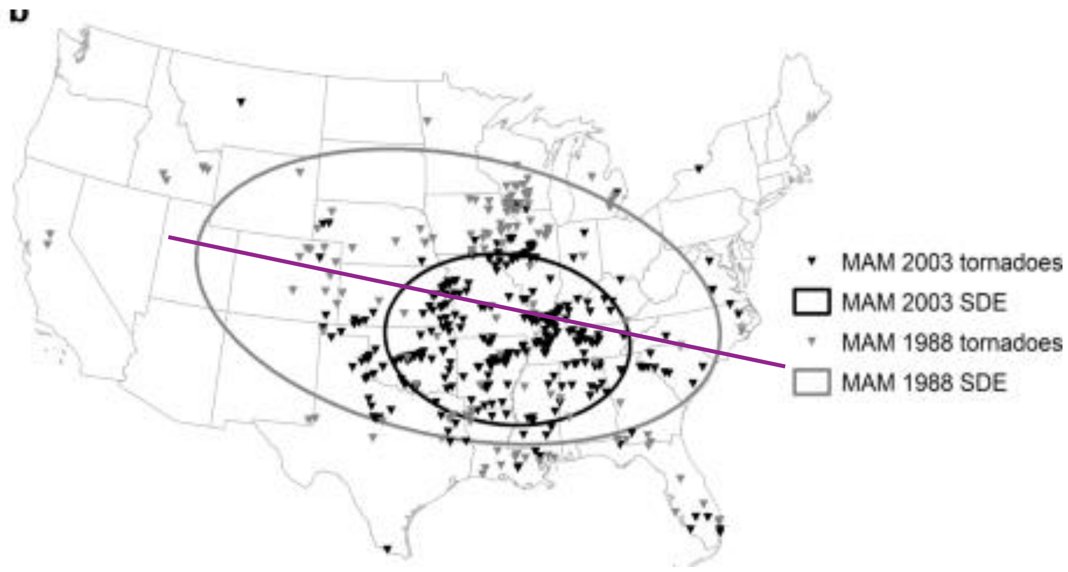
STANDARD DEVIATIONAL ELLIPSE

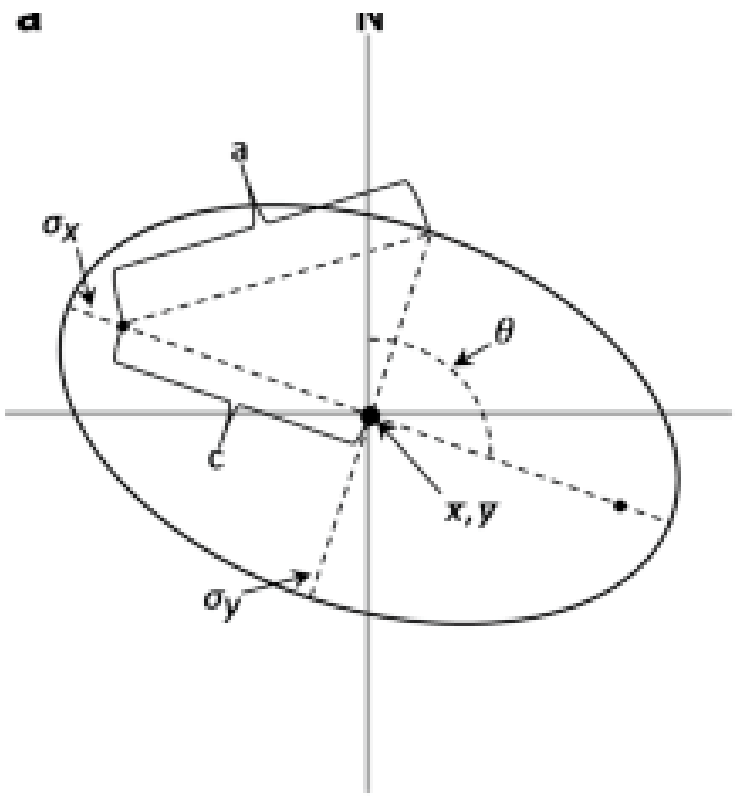
- Another valuable measure of dispersion of spatial events around the spatial mean.
- It gives the dispersion of observations along major and minor axes **using the standard distance separately** in the x and y directions.
- Able to account for both "**distance** and **orientation**/directionality: The ellipse allows you to see if the distribution of features is elongated and hence has a particular orientation.
- Useful for,
    - Summarising data with **distributional directional bias**.
    - Identifying **distributional trends** for geographical phenomena.



STANDARD DEVIATIONAL ELLIPSE

- While you can get a sense of the orientation by drawing the features on a map, calculating the standard deviational ellipse makes the trend clear.
    - Eg: whether a given set of crime locations are along a particular club street?
- You can calculate the standard deviational ellipse using either the locations of the features or using the locations influenced by an attribute value associated with the features. The latter is termed a *weighted standard deviational ellipse*.



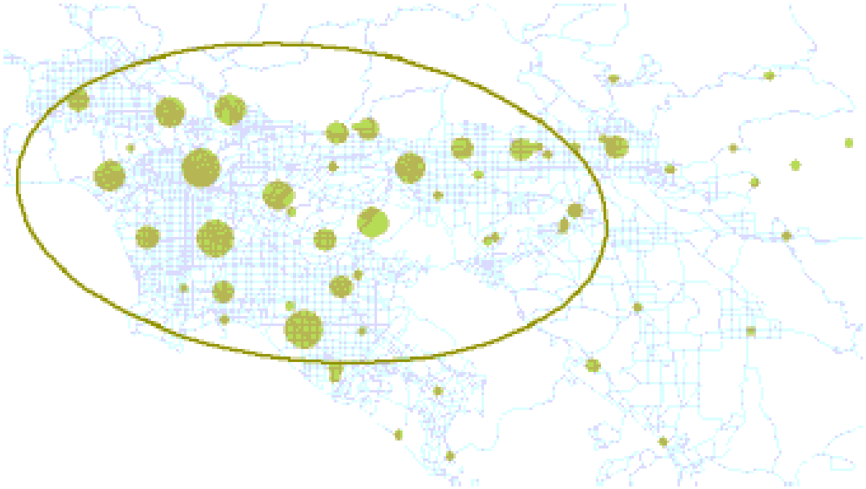


TO DERIVE THE STANDARD DEVIATIONAL ELLIPSE…

We need

1. Spatial mean
2. Angle of rotation from the point of origin (spatial mean)
3. Standard deviations along x- and y- coordinates

STANDARD DEVIATIONS ALONG X- AND Y- COORDINATES



$$SD_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n}}$$

$$SD_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n}}$$

THE ANGLE OF ROTATION

$$\tan\theta = \frac{\sum_{i=1}^{n} {x_i^*}^2 - \sum_{i=1}^{n} {y_i^*}^2 + \sqrt{\sum_{i=1}^{n} {x_i^*}^2 - \sum_{i=1}^{n} {y_i^*}^2 + 4\left(\sum_{i=1}^{n} x_i^* y_i^*\right)^2}}{2\sum_{i=1}^{n} x_i^* y_i^*}$$

where,

$$x_i^* = x_i - \bar{X} \text{ and } y_i^* = y_i - \bar{Y}$$

- The coordinates of the observations are transformed towards the mean centre.
- By rotating the coordinates clockwise about their new origin (mean centre) by a certain "angle", we can determine the standard deviations along the x- and y- coordinates → axes of the ellipse.

STANDARD DEVIATIONS ALONG X- AND Y- AXIS…

…after transformation towards the mean center, the axes are respectively given by,

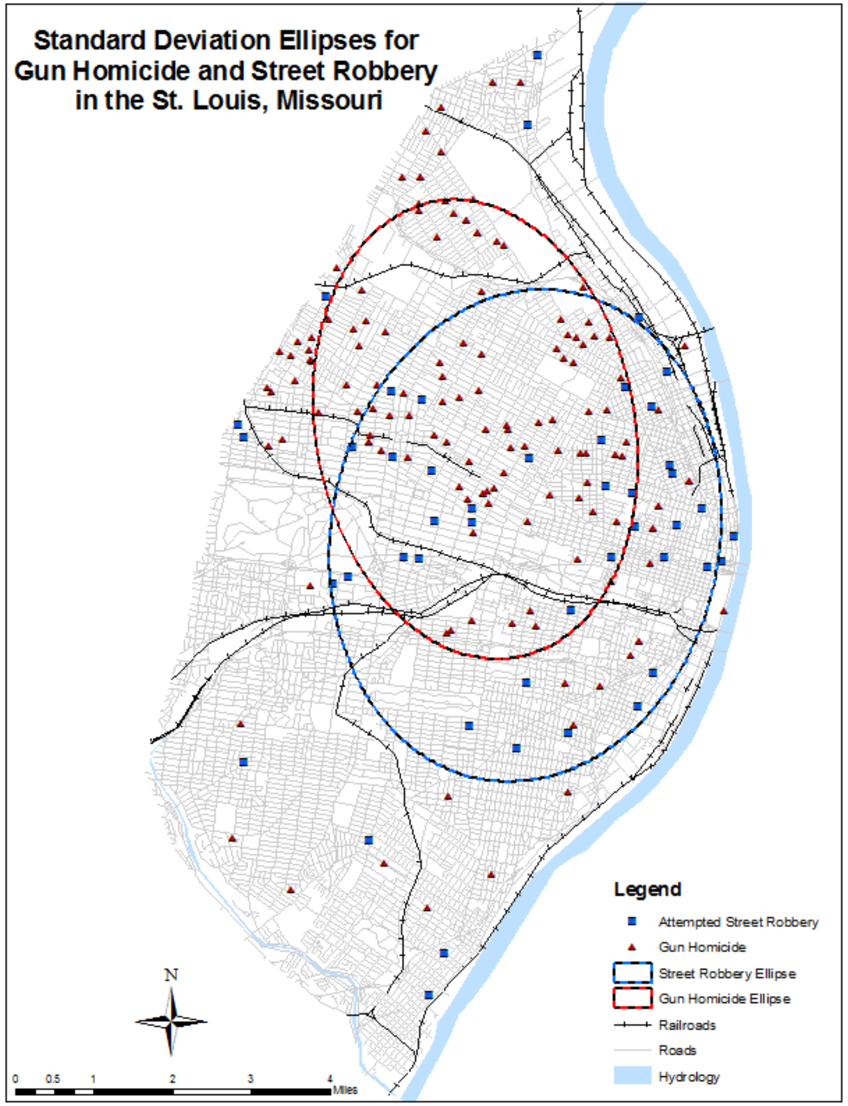
$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i^* \cos\theta - y_i^* \sin\theta\right)^2}{n}}$$

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i^* \sin\theta - y_i^* \cos\theta\right)^2}{n}}$$

These will give the major and minor axes of the ellipse.

POTENTIAL APPLICATIONS (OF SDE)

- Mapping the distributional **trend for a set of crimes** might identify a relationship to physical features (a string of bars or restaurants, neighbourhood, and so on).
- **Mapping groundwater** well samples for a contaminant might indicate how the toxin is spreading and, consequently, may be useful in deploying mitigation strategies.
- Comparing the size, shape, and **overlap of ellipses** for various racial or ethnic groups may provide insights regarding racial or ethnic segregation.
- Plotting ellipses for a **disease outbreak** over time may be used to model its spread.



TAKE HOME POINTS

- Traditional vs spatial measures of numerical descriptive measures
- Common measures of spatial numerical descriptive measures
  - Spatial measures of central tendency
  - Spatial measures of dispersion
- Standard deviational ellipse

REFERENCES

- *Spatial Analysis* by Tonny Oyana, 2nd edition, Chapter 3.
- https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-directional-distribution-standard-deviationa.htm