# LECTURE 16: AREAL DATA OVERVIEW

CLASS INTRO

INTRO QUESTIONS

- We recently wrapped up the section on geostatistical or point-reference data. Assume you have two minutes with a recruiter, explain what you have learned thus far.

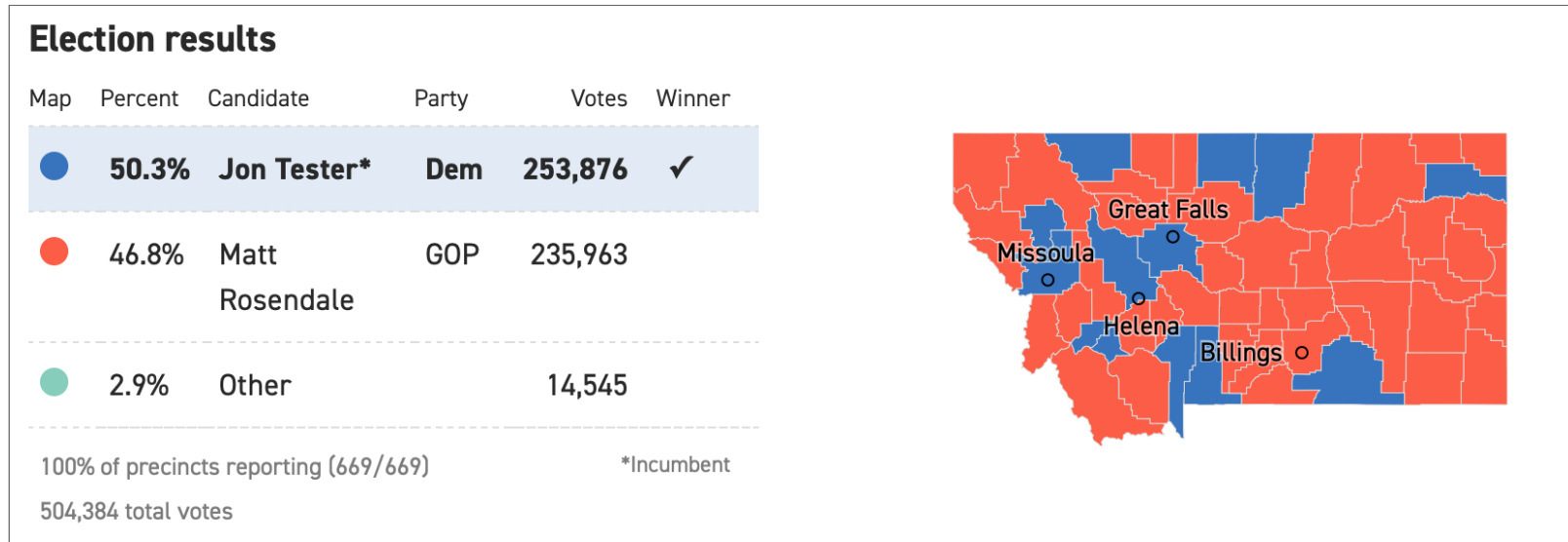- For Today:
  - Areal Data

AREAL DATA OVERVIEW

AREAL DATA

Defining features: random observation measured at well defined subsets, such as a city or state.

source: **https://www.politico.com/election-results/2018/montana/**
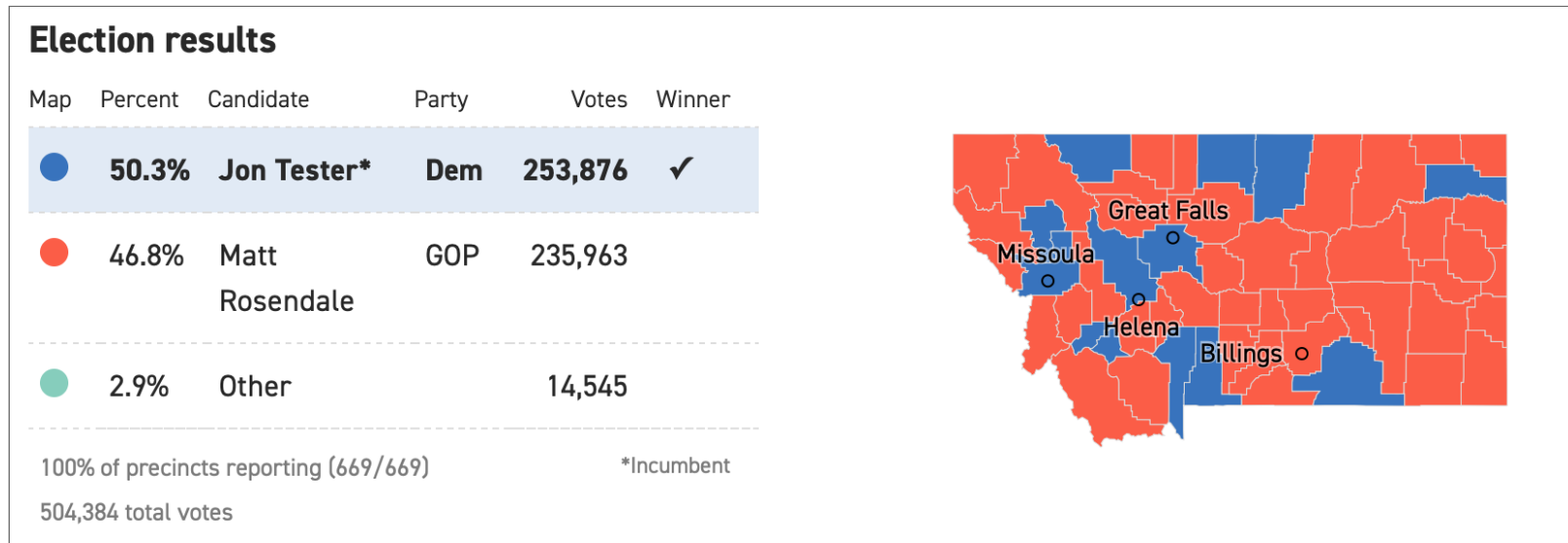
AREAL DATA

Recall: How can spatial information be incorporated with this data structure?

source: **https://www.politico.com/election-results/2018/montana/**

AREAL DATA MODEL OVERVIEW

- Data, typically averages or totals, are captured for geographic units or blocks
- Spatial correlation is incorporated with a *neighbor* structure.
- Autoregressive models on the neighbor structure capture spatial similarities.
- Model based approaches will incorporate covariates and introduce spatial structure with random effects.

AREAL DATA INFERENTIAL QUESTIONS

1. Is there a spatial pattern?
2. In presenting a map of expected responses, should the raw values or a smoothed response be presented?
3. What values would be expected for new set of areal units?

EXPLORATORY METHODS FOR AREAL DATA

DATA VIZ OPTIONS

Creating choropleth maps:

- **<u>Poverty in Nepal with ggplot</u>**
- **<u>Plotly</u>**
- **<u>Crime in Philly</u>**
- **<u>State and County Population</u>**

LEAFLET TUTORIAL

Follow the **Leaflet tutorial** for creating choropleths.

- what is the object `states`?
- based on my drives through South Dakota, I'm convinced there is a data entry error with the density for South Dakota. Fix this.

CHOROPLETH TUTORIAL # 2

- What are the objects `urbnmapr::states` and `urbnmapr::counties`?
- What is `urbnmapr::countydata`?
- Why do you think there only 6 colors for the state?

CHOROPLETH TUTORIAL # 3

- Now download the election results and create two maps by county.
1. Total number of votes for Tester
2. Proportion of votes for Tester

EXPLORATORY APPROACHES FOR AREAL DATA

PROXIMITY MATRIX

- Similar to the distance matrix with point-reference data, a proximity matrix W is used to model areal data.
- Given measurements $Y_i, \ldots, Y_n$ associated with areal units $1, \ldots, n$, the elements of W, w_{ij} connect units i and j
- Common values for w_{ij} are

$$w_{ij} = \begin{cases} 1 & \text{if i and j are adjacent} \\ 0 & \text{otherwise (or if i=j)} \end{cases}$$

GRID EXAMPLE

- Create an adjacency matrix with diagonal neigbors
- Create an adjacency matrix without diagonal neigbors
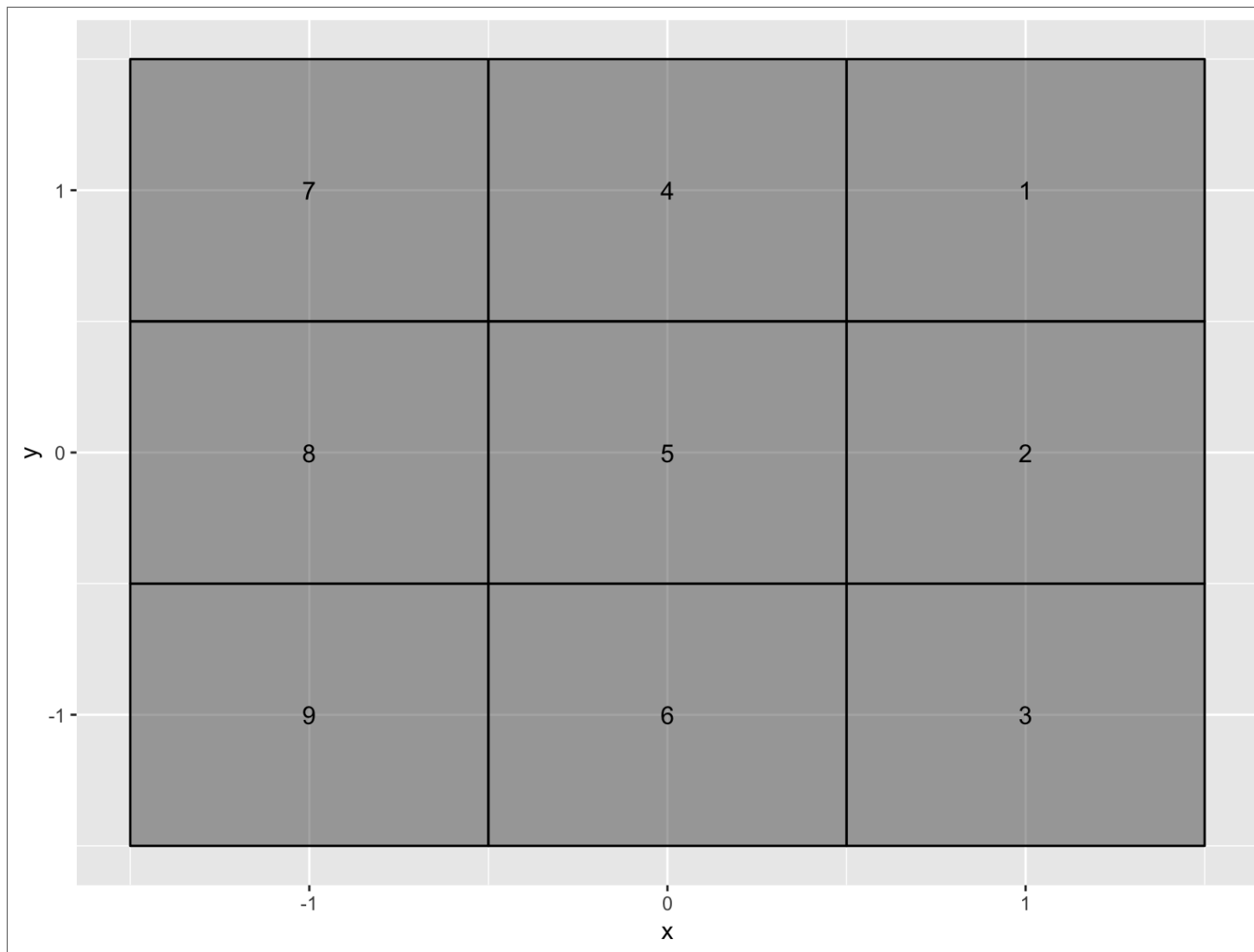
SPATIAL ASSOCIATION

There are two common statistics used for assessing spatial association:
Moran's I and Geary's C.

- Moran's I

$$I = \frac{n \sum_i \sum_j w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i \neq j} w_{ij}) \sum_i (Y_i - \bar{Y})^2}$$

- Geary's C

$$C = \frac{(n-1) \sum_i \sum_j w_{ij}(Y_i - Y_j)^2}{2(\sum_{i \neq j} w_{ij}) \sum_i (Y_i - \bar{Y})^2}$$