Lecture 7: Point Level Models - EDA and Simulation

Class Intro

Intro Questions

- What are some basic guidelines for EDA for point referenced data?
- For Today:
  - More EDA
  - Data Simulation

Spatial EDA Overview

1. Map of locations
2. Histogram or other distributional figure
3. 3D scatterplot
4. General Regression EDA
5. Variograms and variogram clouds
6. Anistopic diagnostics

Scallops Data Example

1. Map of Locations

+
−

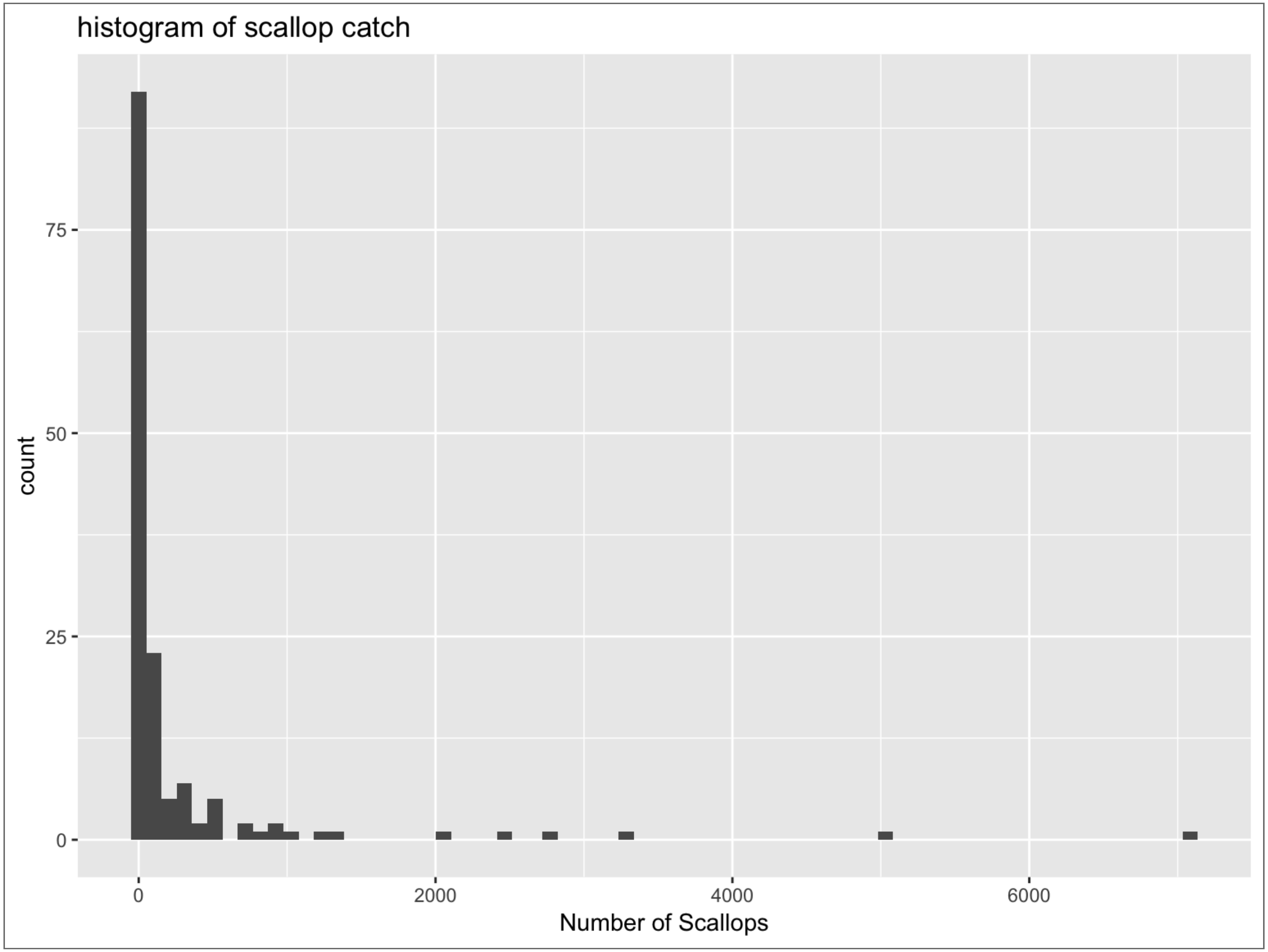
Leaflet

1. Map of Locations - Takeaways

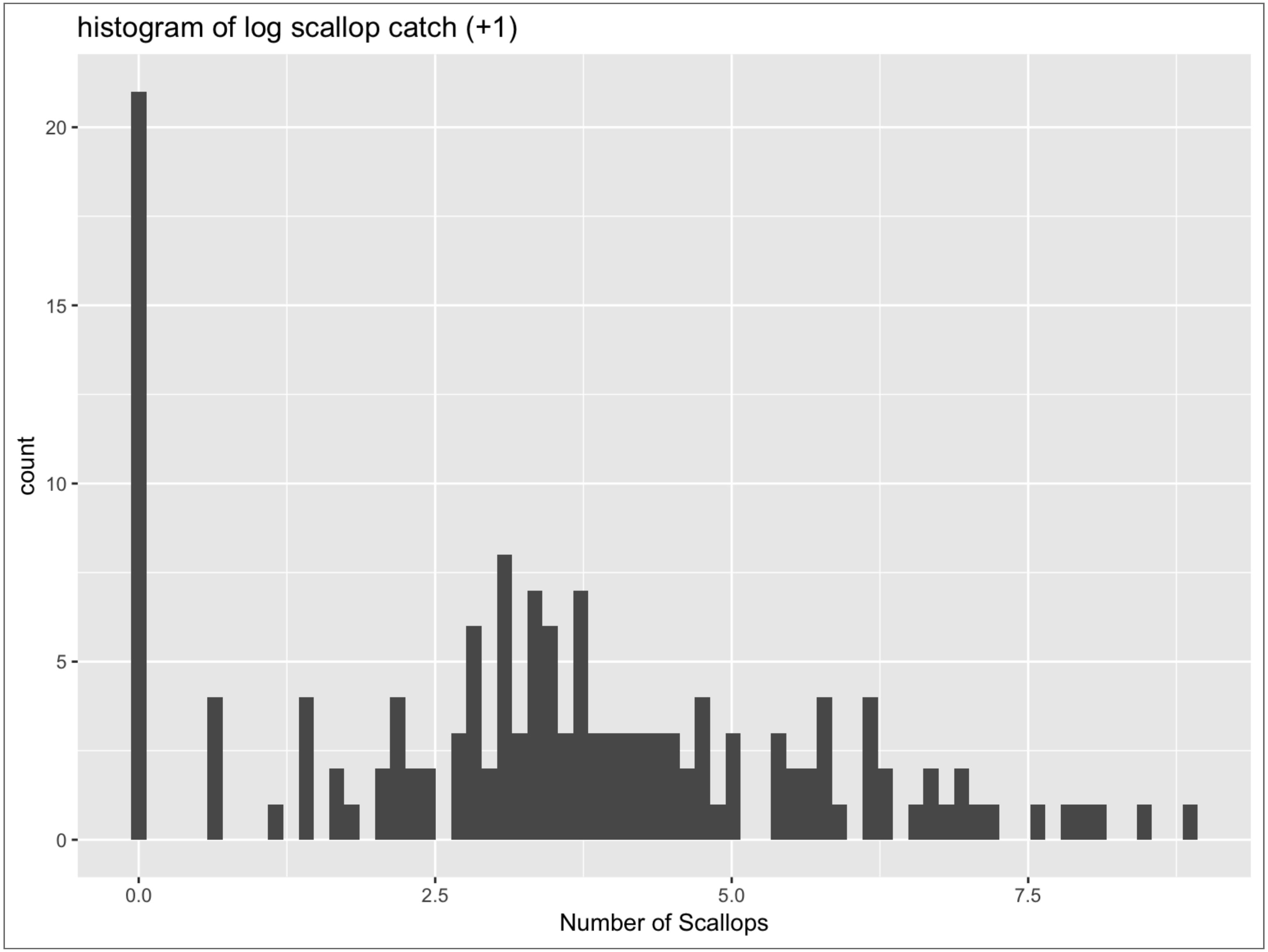
Goal: Understand the sampling approach

- Is this a grid?
- Are there directions that have larger distances?
- How large is the spatial extent?

2. Histogram


histogram of scallop catch
count
75
50
25
0
0
2000
4000
6000
Number of Scallops

2. Histogram



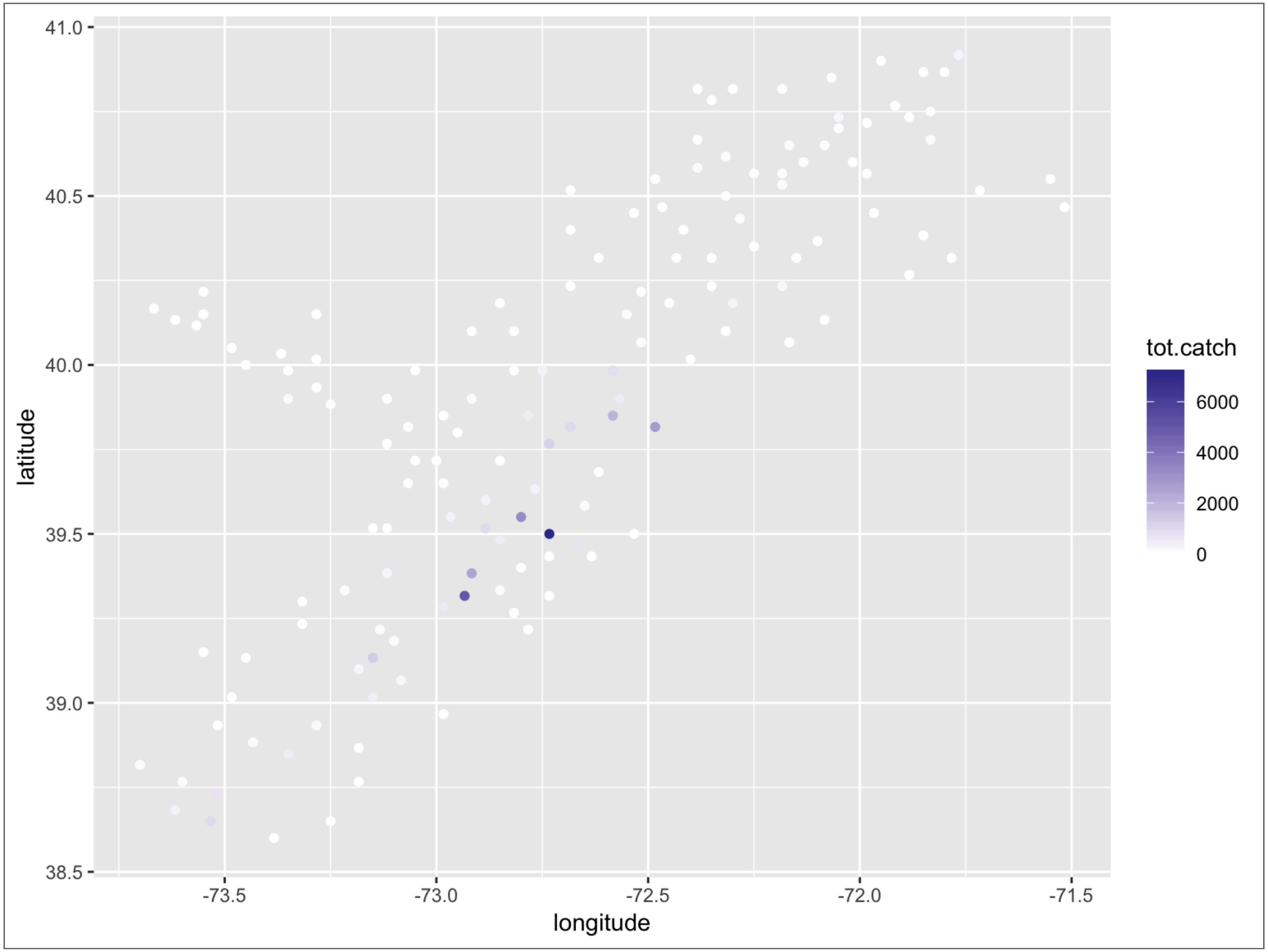
histogram of log scallop catch (+1)
count
20
15
10
5
0
0.0
2.5
5.0
7.5
Number of Scallops

2. Histogram - Takeaways

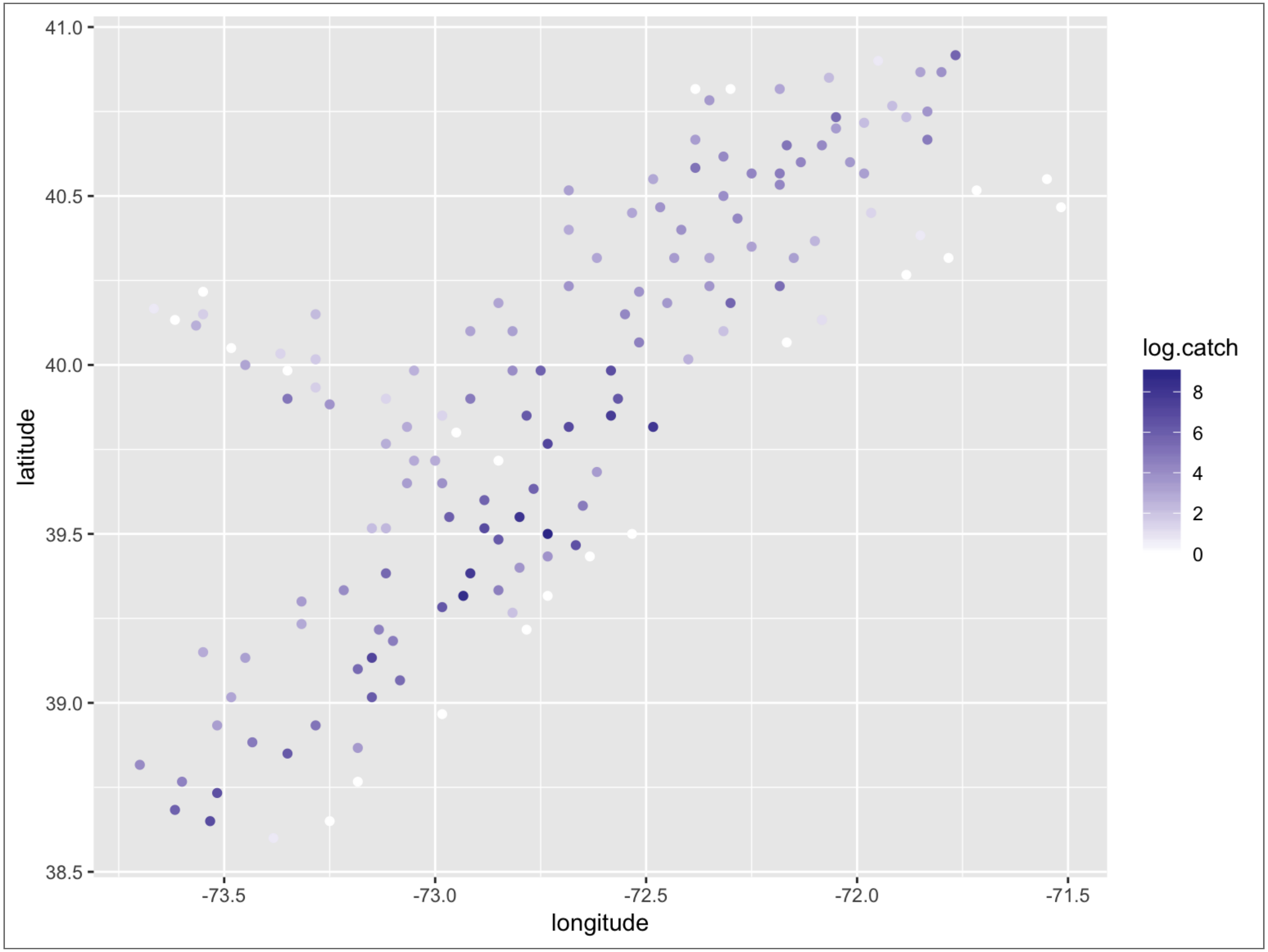
Goal: Identify a sampling distribution for the data

- Continuous or discrete data
- A linear model approach will be used for the response
- Spatial structure can also be included in generalized linear models
- Outliers are worth investigating, but a data point that does not fit the assumed model should not automatically be eliminated

3. 3D scatterplot



3. 3D scatterplot



3. 3D scatterplot - Takeaways

Goal: Examine the spatial pattern of the response

- Again, this is the response not the residual
- Can also think about a contour plot (using some interpolation method)

4. General Regression EDA

- Assessing relationship between variable of interest and covariate information
- No covariates are present in the scallops data

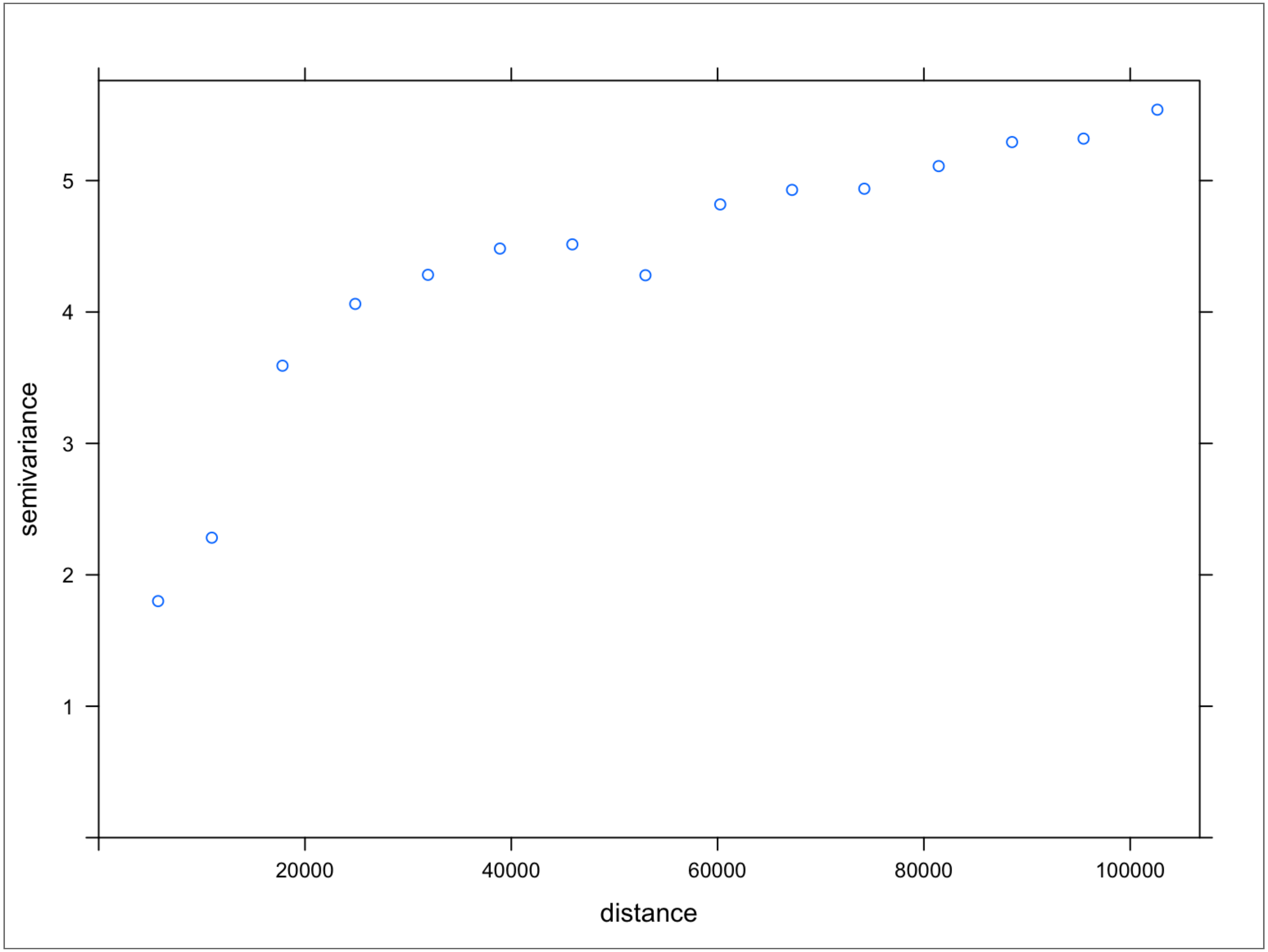
5. Variograms and variogram clouds: Exercise

Explore the code below: what are the differences in the three variograms?

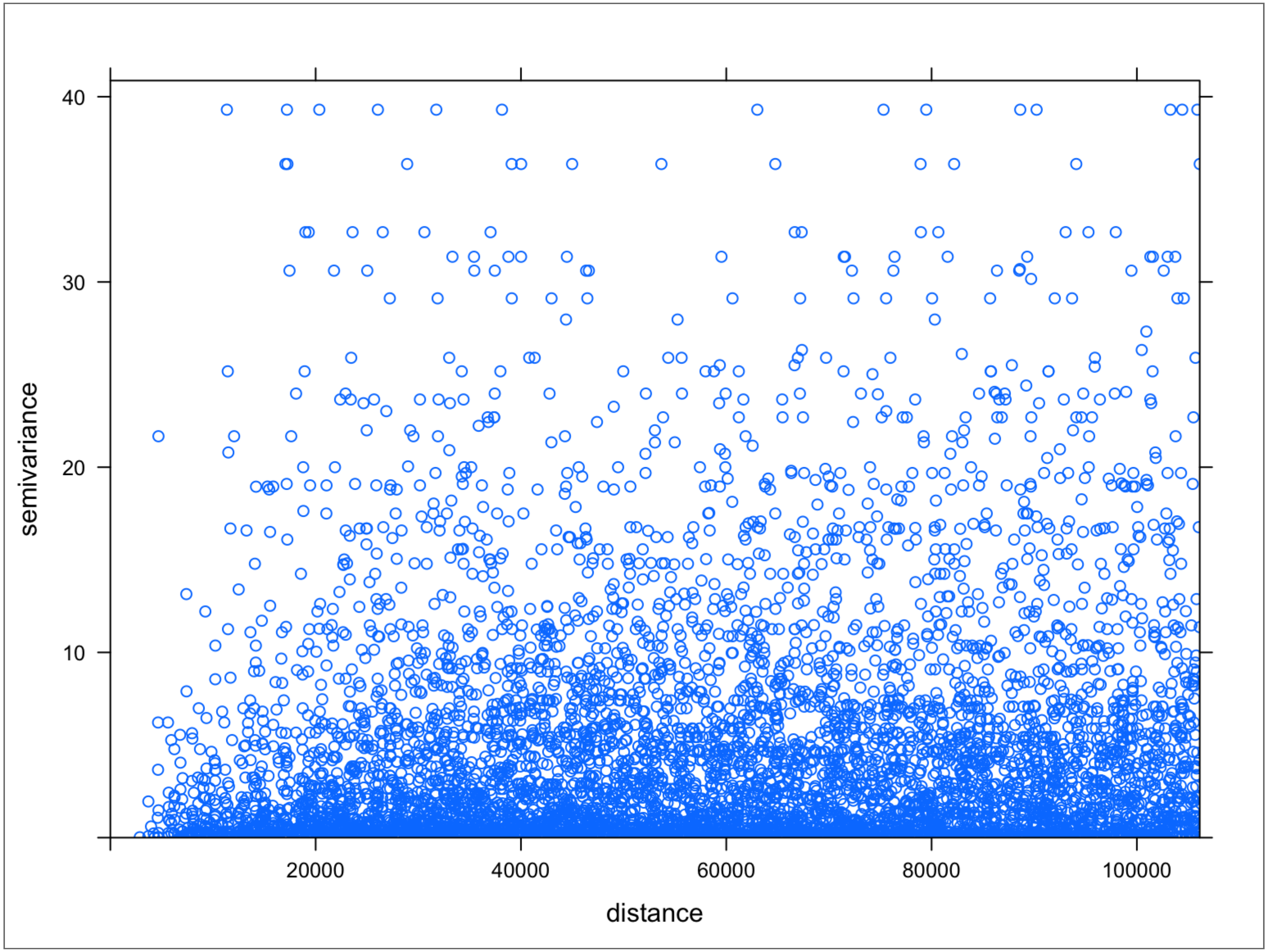
```
coordinates(scallop) = ~longitude+latitude
class(scallop)
scallop.sp <- scallop
proj4string(scallop.sp) <- CRS("+proj=longlat +datum=WGS84")
scallop.utm <- spTransform(scallop.sp, CRS("+proj=utm +zone=18 ellps=WGS84"))

plot(variogram(log.catch~1, scallop))
plot(variogram(log.catch~1, scallop.sp))
plot(variogram(log.catch~1, scallop.utm))
```

5. Variograms



5. Variogram Cloud



5. Variograms and variogram clouds: Takeaways

Goal: Visually diagnose spatial structure

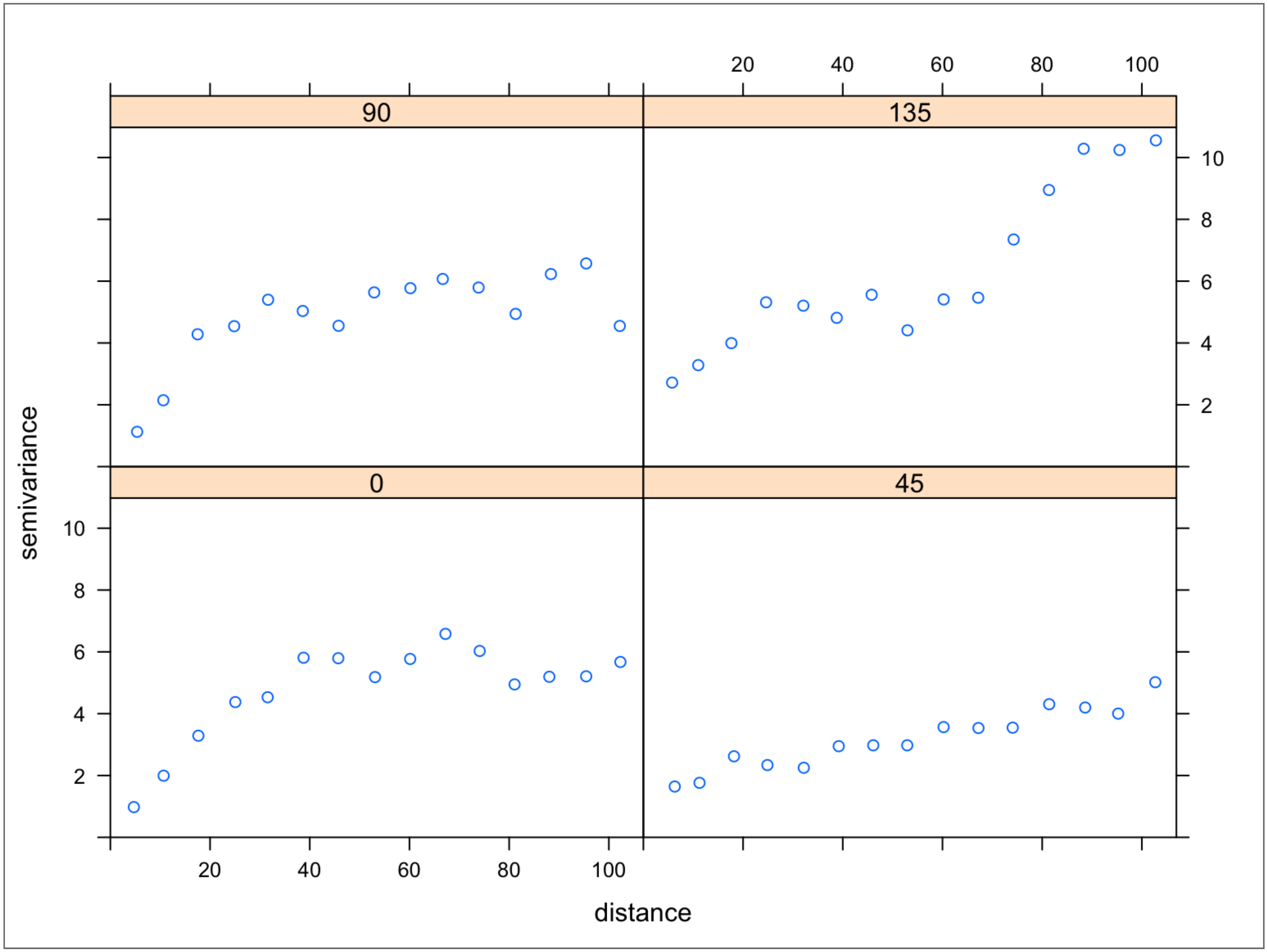
6. Anisotropy

Goal: Determine if direction influencs spatial structure

Anisotropy

Directional Variogram

- All of the variograms we have looked at are isotropic



90
135
0
45
semivariance
distance
2
4
6
8
10
20
40
60
80
100

Separable Correlations Functions

- If the differences in spatial structure are directly related to two coordinate sets, we can create a stationary, anistropic covariance function
- Let
  $cor(Y(\boldsymbol{s} + \boldsymbol{h}), Y(\boldsymbol{s})) = \rho_1(h_y)\rho_2(h_x)$,
  where $\rho_1()$ and $\rho_2()$ are proper correlation functions.
- A scaling factor, σ^2, can be used to create covariance.

Geometric Anistropy

- Another solution is the class of geometric anisotropic covariance functions with
  $C(\mathbf{s} - \mathbf{s}') = \sigma^2 \rho((\mathbf{s} - \mathbf{s}')^T B(\mathbf{s} - \mathbf{s}'))$,
  where B is positive definite matrix and ρ is a valid correlation function
- B is often referred to as a transformation matrix which rotates and scales the coordinates, such that the resulting transformation can be simplified to a distance.

Sill, Nugget, and Range Anisotropy

- Recall the sill is defined as $\lim_{d \to \infty} \gamma(d)$
- Let $\boldsymbol{h}$ be an arbitrary separation vector, that can be normalized as $\frac{\boldsymbol{h}}{||\boldsymbol{h}||}$
- If $\lim_{a \to \infty} \gamma(a \times \frac{\boldsymbol{h}}{||\boldsymbol{h}||})$ depends on $\boldsymbol{h}$, this is referred to as sill anisotropy.
- Similarly the nugget and range can depend on $\boldsymbol{h}$ and give nugget anisotropy and range anisotropy

Model Simulation

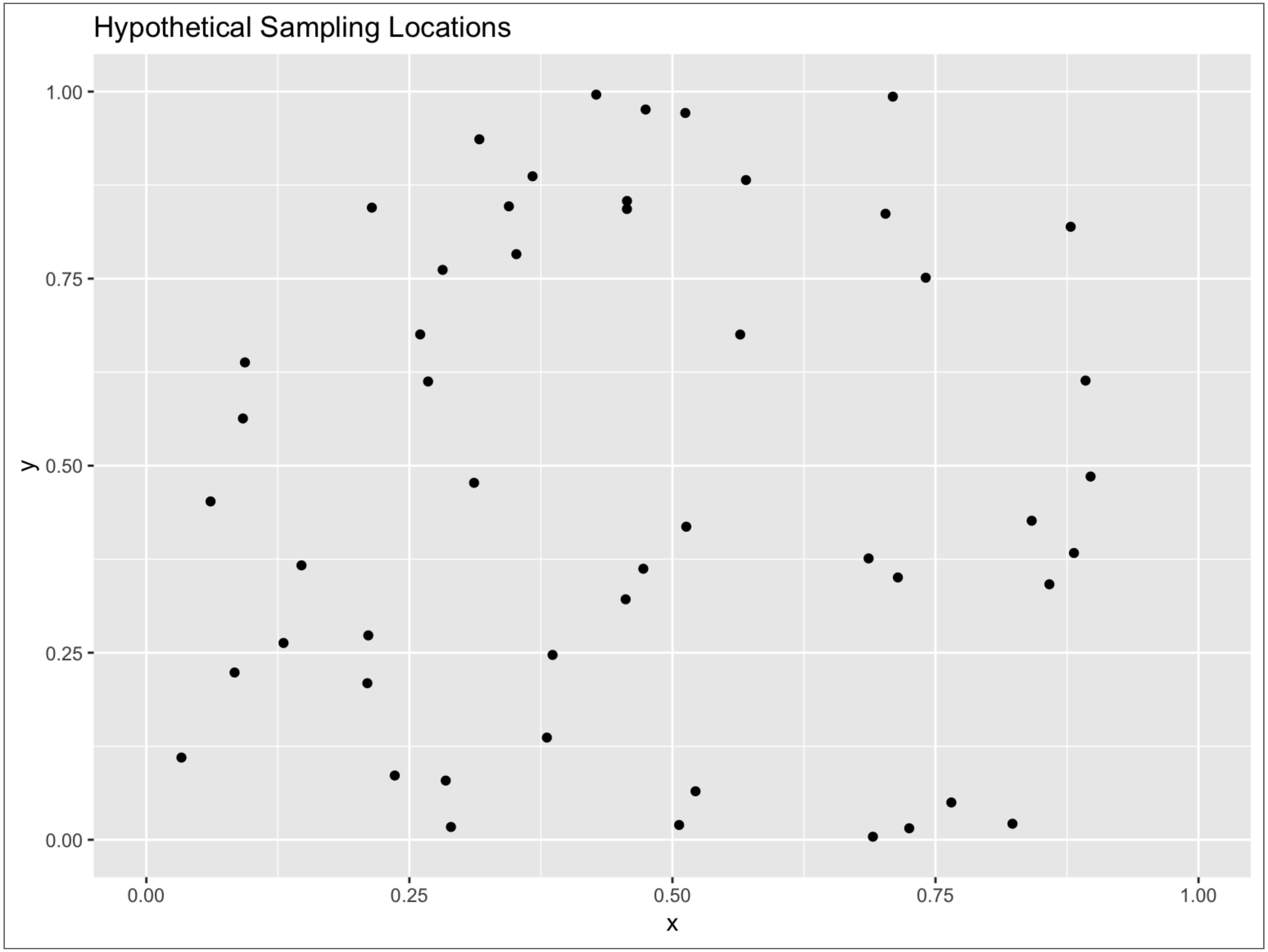
Simulating Spatial Process

- Soon we will look at fitting models for spatial point data
- Simulating data gives a deeper understanding of the model fitting process
- Simulate a mean-zero, isotropic spatial process with a spherical covariance function

Simulating Spatial Process

1. Construct spatial locations
2. Calculate distances
3. Define covariance function and set parameters
4. Sample realization of the process
5. Visualize realization of spatial process

1. Construct Spatial Locations



2. Calculate Distances

```
dist.mat <- dist(coords, diag = T, upper = T) %>% as.matrix()
```

3. Define Covariance Function and Set Parameters

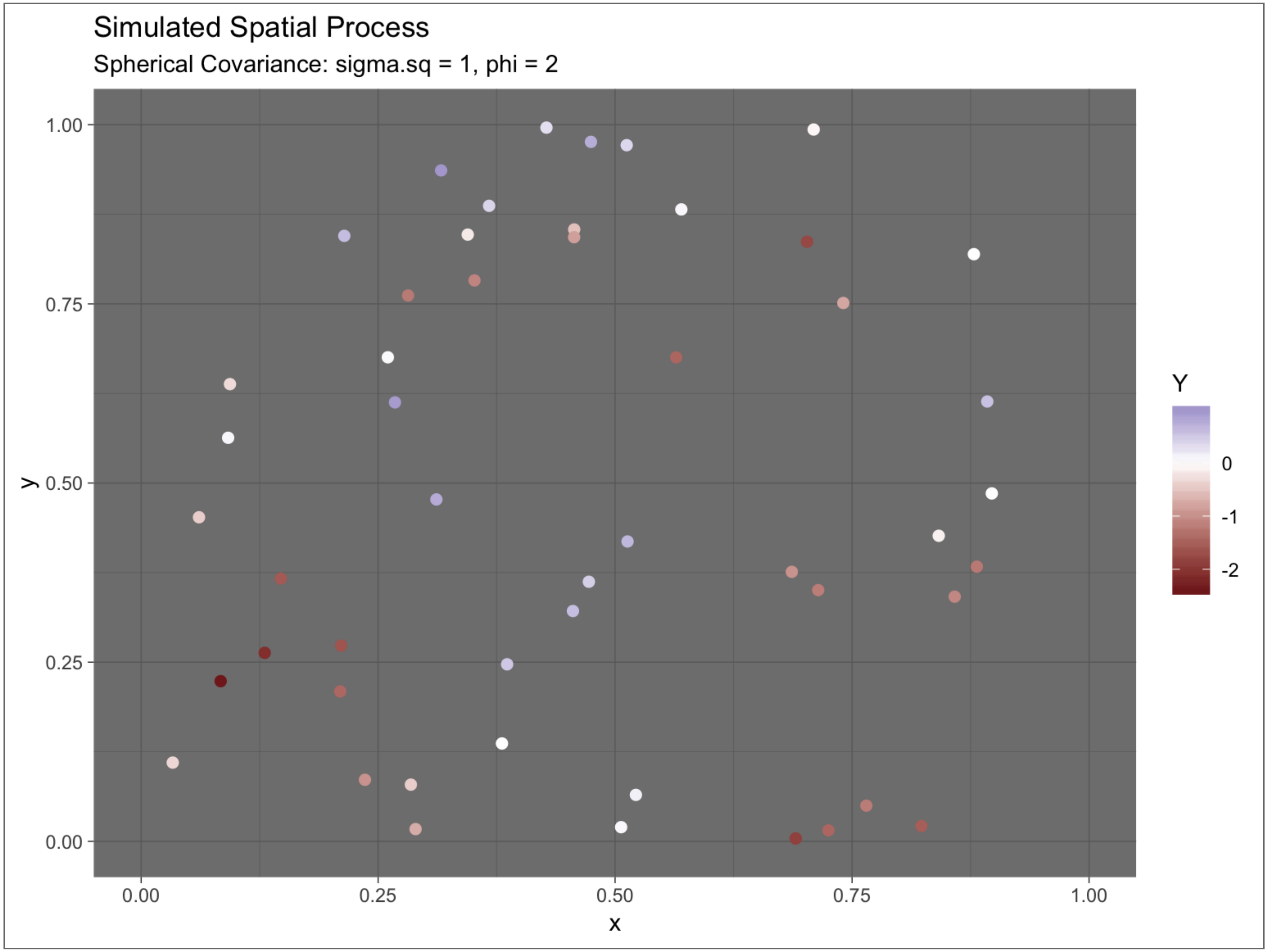
- Use spherical covariance with no nugget:

$$C(d) = \begin{cases} 0 \text{ if } d \geq \frac{1}{\phi} \\ \sigma^2 \left[1 - \frac{3}{2}\phi d + \frac{1}{2}(\phi d)^3\right] & \text{if } 0 \leq d \leq \frac{1}{\phi} \end{cases}$$

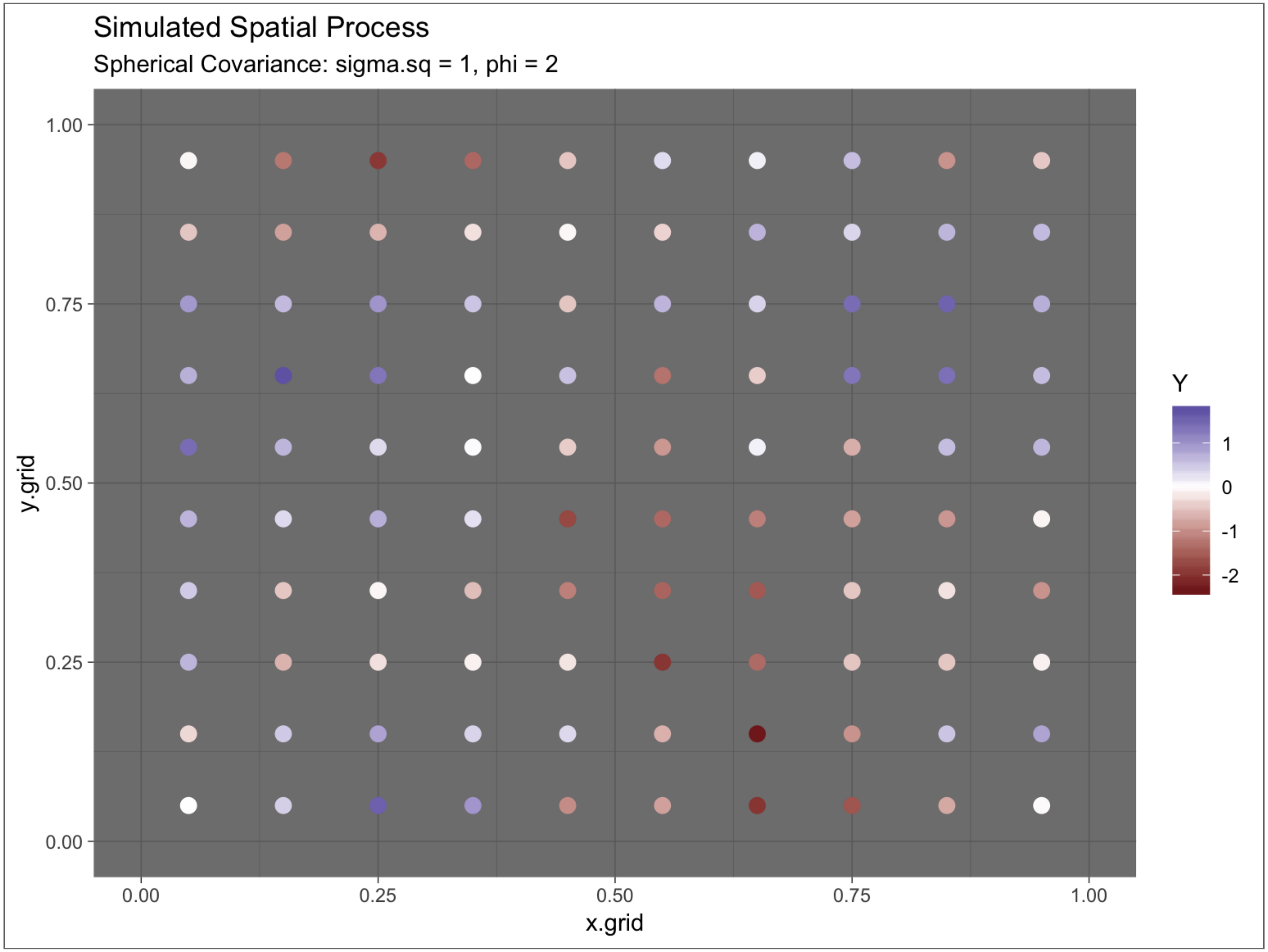
4. Sample realization of the process

- This requires a distributional assumption, we will use the Gaussian distribution
- What about the rest of the locations on the map?

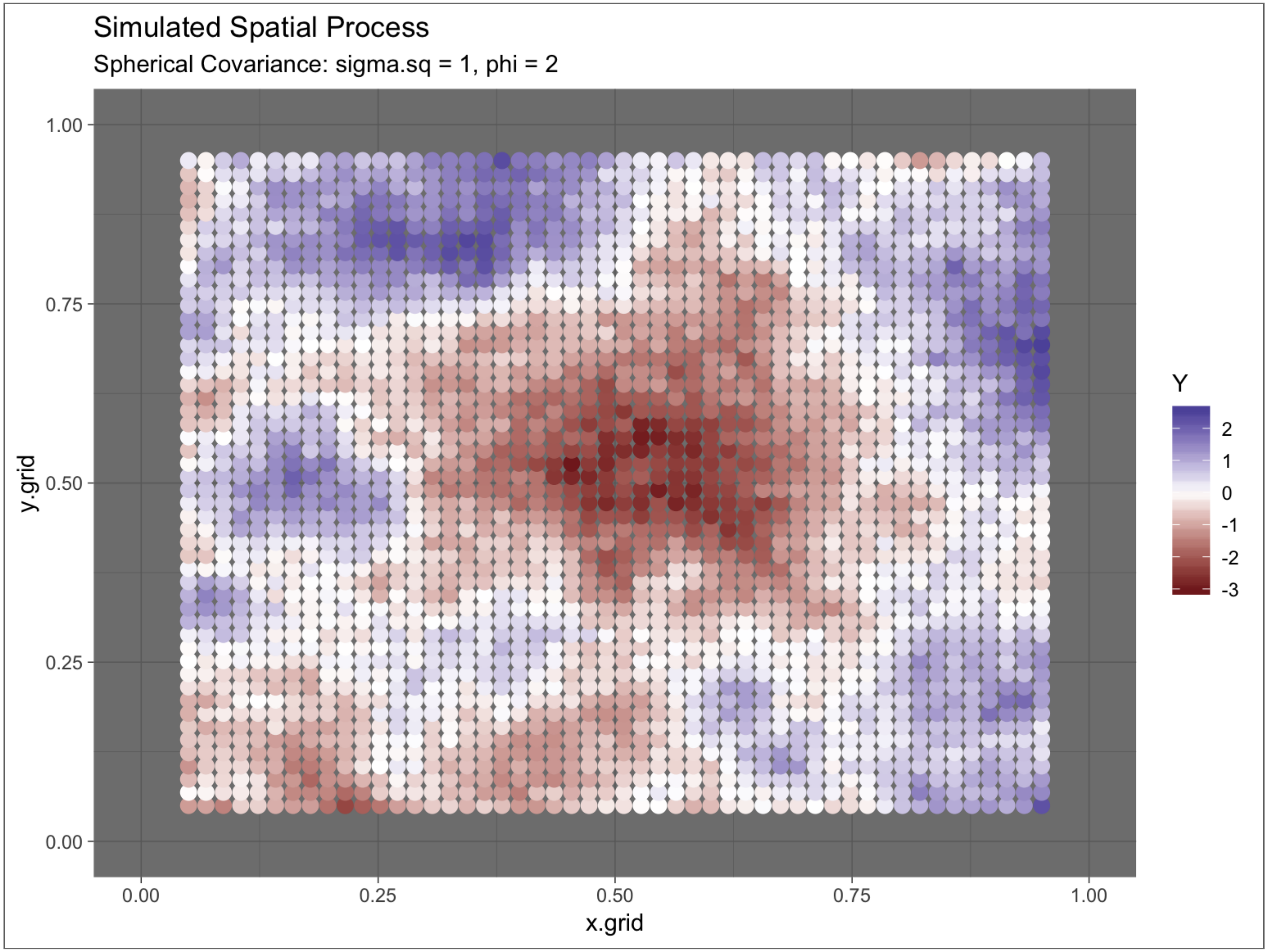
5. Vizualize Spatial Process



Extra: Gridded Spatial Process



Extra: Gridded Spatial Process



Simulated Spatial Process: Exercise

How does the spatial process change with:

- another draw with same parameters?
- a different value of ϕ
- a different value of σ^2
- adding a nugget term, τ^2

Model Fitting

Classical Model Fitting

- The classical approach to spatial prediction is rooted in the minimum the mean-squared error.
- This approach is often referred to as *Kriging* in honor of D.G. Krige a South African mining engineer.
- As a result of Krige's work (along with others), point-level spatial analysis and geostatistical analysis are used interchangeably.

Additional Resources

- **Meuse Data Tutorial**
- **Textbook Data Sets**