Spatial Survey Designs

Anthony (Tony) R. Olsen

USEPA NHEERL

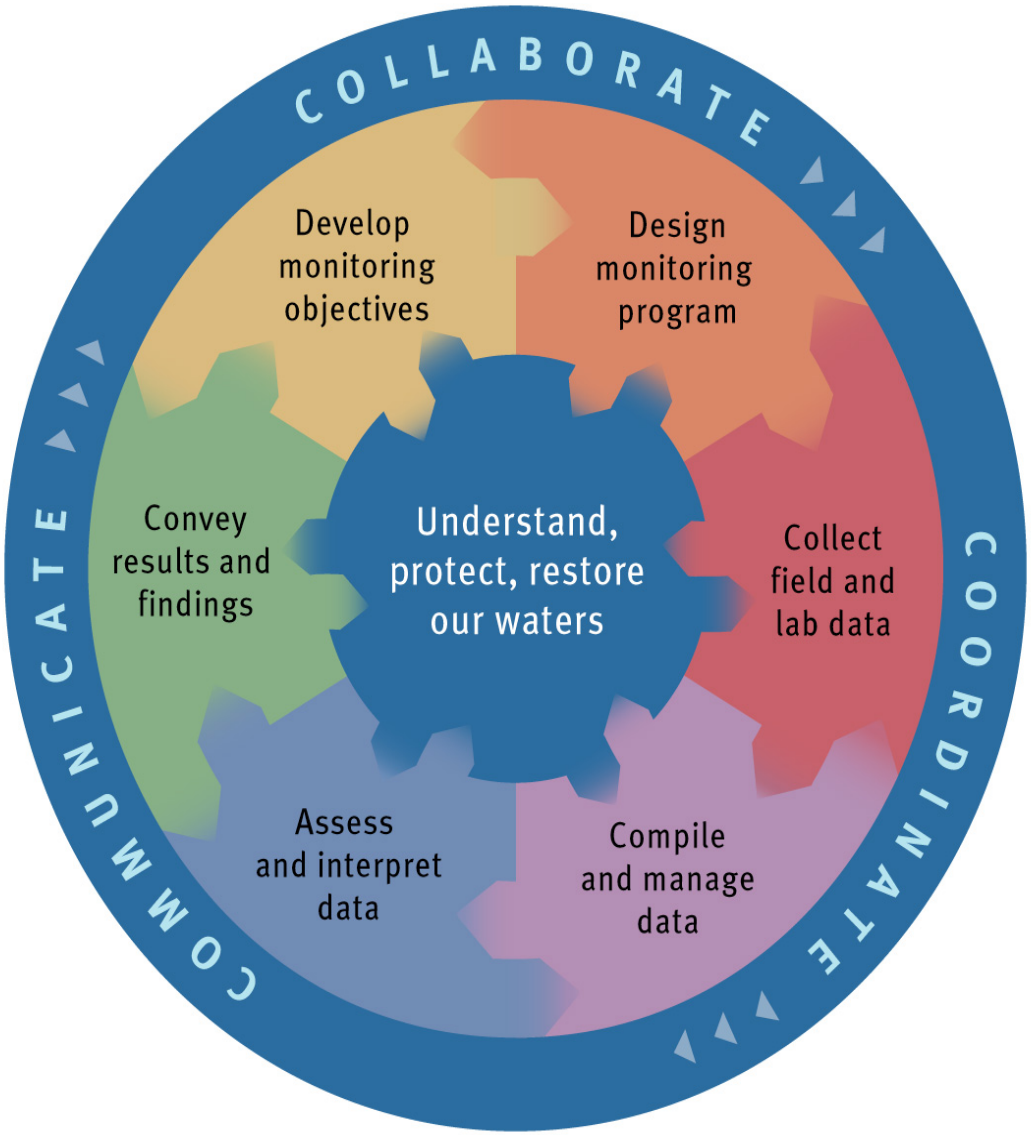
Western Ecology Division

Corvallis, Oregon

(541) 754-4790

Olsen.Tony@epa.gov

Web Page: http://www.epa.gov/nheerl/arm



National Water Quality Monitoring Council: Monitoring Framework

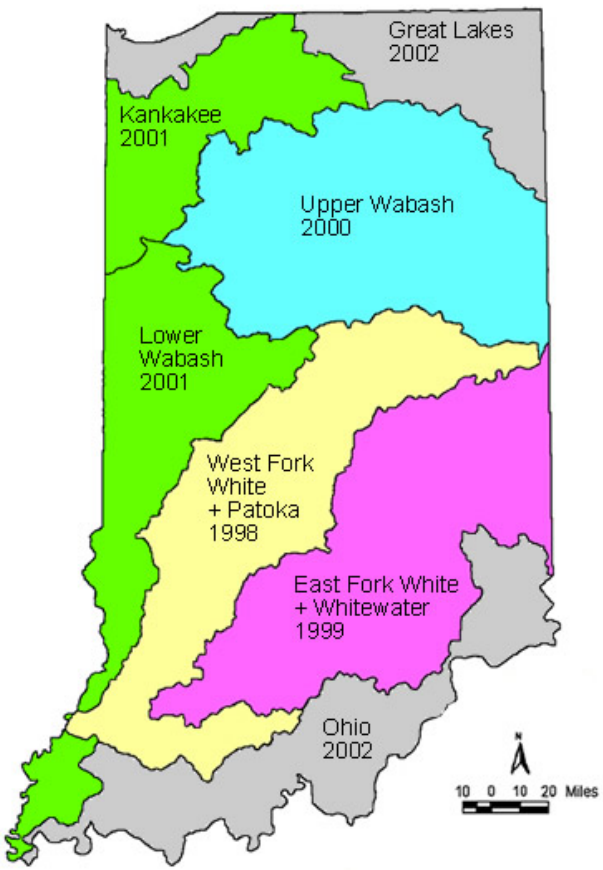
- View any study as information system
- Study pieces must be designed and implemented to fit together

- Reference: Water Resources IMPACT, September 2003 issue

Survey Components
Objectives-Design-Analysis-Report

- Survey objectives
- Institutional constraints
- Target population
- Sample frame
- Indicators and response design
- Design requirements
- Specification of survey design
- Site selection (psurvey.design)
- Site evaluation
- Conduct field and lab measurements
- Indicator results database
- Sample frame summary
- Target population estimation (psurvey.analysis)
- Report results

Indiana Stream Example



- Objective: Within Upper Wabash basin for streams with flowing water estimate km that are impaired and non-impaired for aquatic life use
- Target population: All flowing waters during summer index period within Upper Wabash
- Sample frame: NHD perennial coded streams
- Rotating basins across state

Indiana Stream Example

- Sample size: 50 sites with equal number in 1st, 2nd, 3rd, and 4th+ Strahler order
- Survey Design: GRTS for a linear network (spatially-balanced random sample)
- Oversample: 50 sites

- Site Evaluation
  - TS: Target Sampled
  - NT: Non-target
  - LD: Landowner denied access
  - PB: Physically barrier
- Adjust weights
- Estimate perennial stream extent
- Estimate stream km impaired

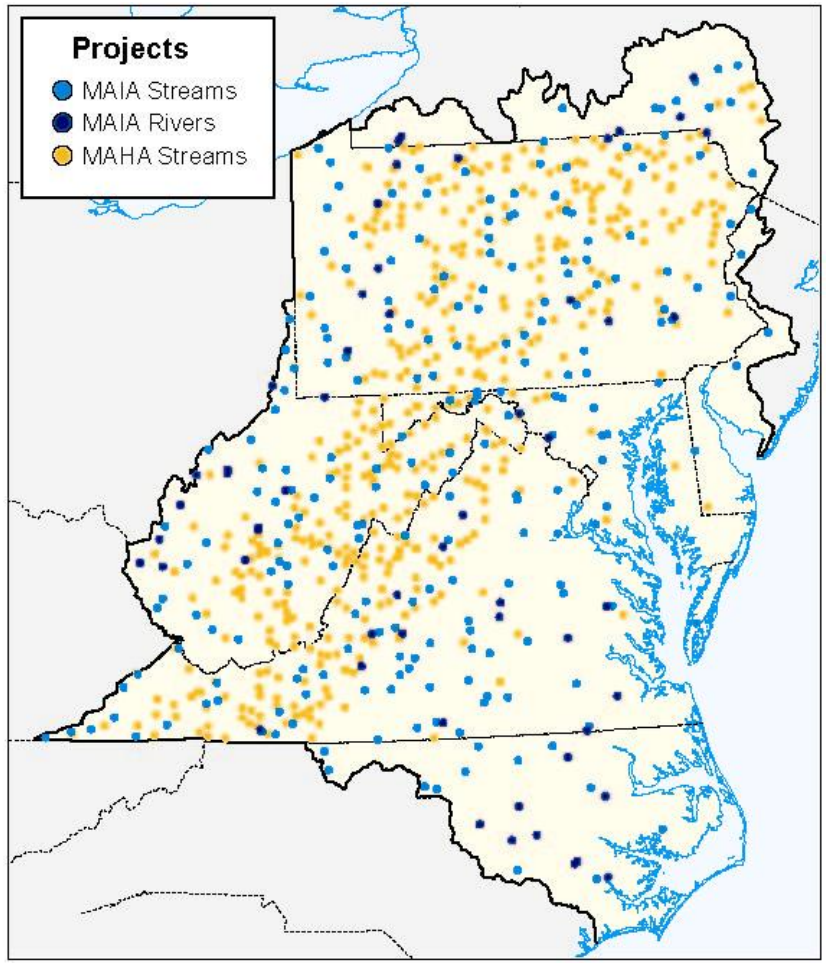
Indiana Upper Wabash Basin: Perennial Stream Length (km)

Stream Length Category	Length (km)	Percent Total
NHD GIS coverage	7358	100
Estimated perennial	5707 ± 724	77.6 ± 9.8
Estimated sampled	3414 ± 729	46.4 ± 9.9

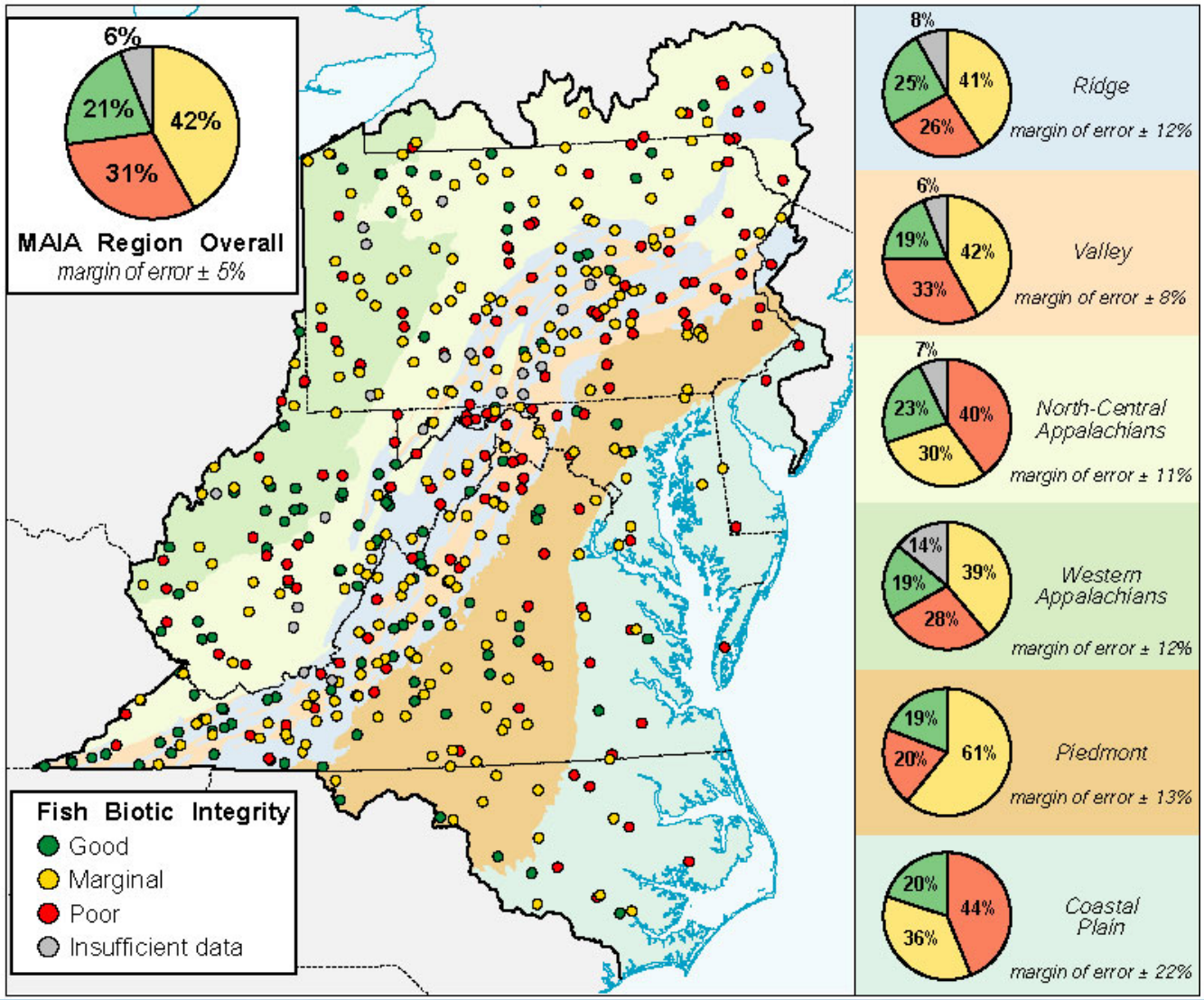
Indiana Upper Wabash Basin: Biological and Habitat Assessment

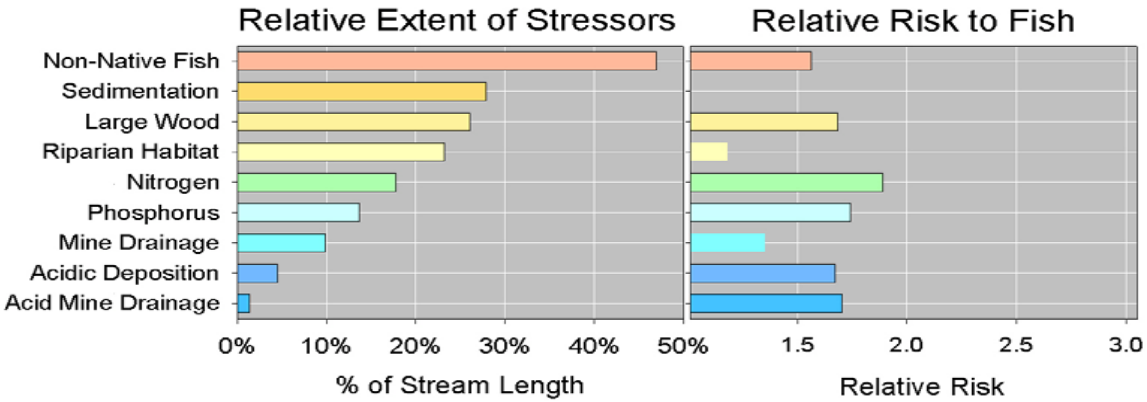
Indicator	Status	Length (km)	Percent Total
IBI	Not Impaired	4128 ± 1134	72.3 ± 13.0
	Impaired	1579 ± 810	27.7 ± 13.0
	Total	5707	100
QHEI	Not Impaired	3373 ± 990	59.1 ± 16.3
	Impaired	2334 ± 1168	40.9 ± 16.3
	Total	5707	100

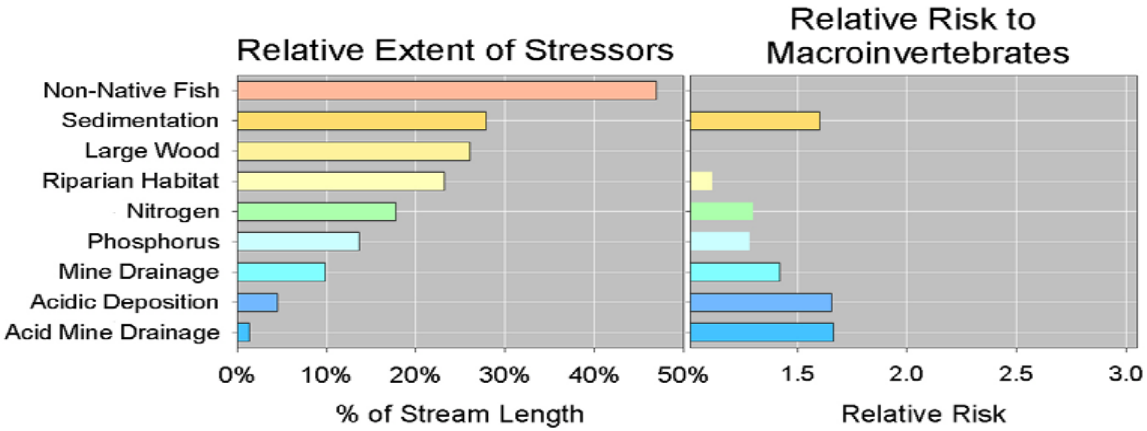
Mid-Atlantic Integrated Assessment (MAIA) Streams and Rivers

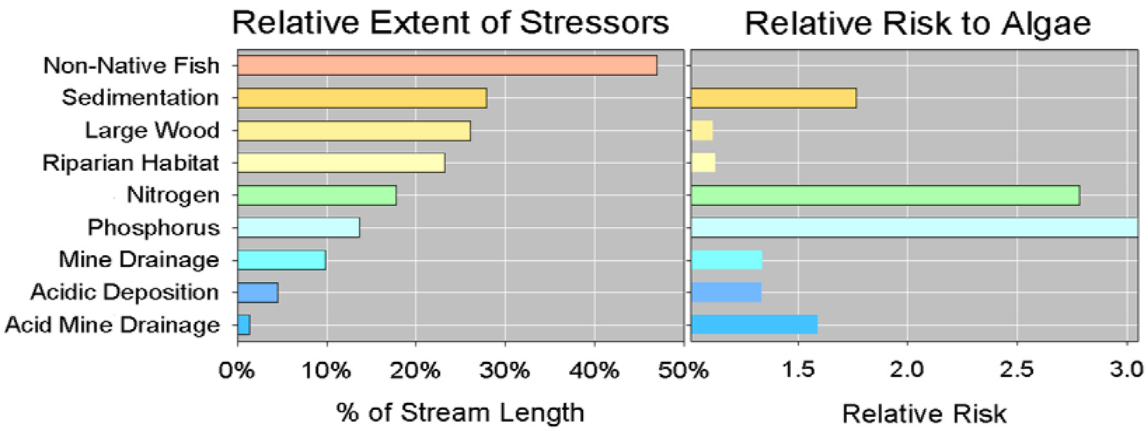


MAIA: Fish Assemblage Condition









MAIA: Relative Risk Assessment

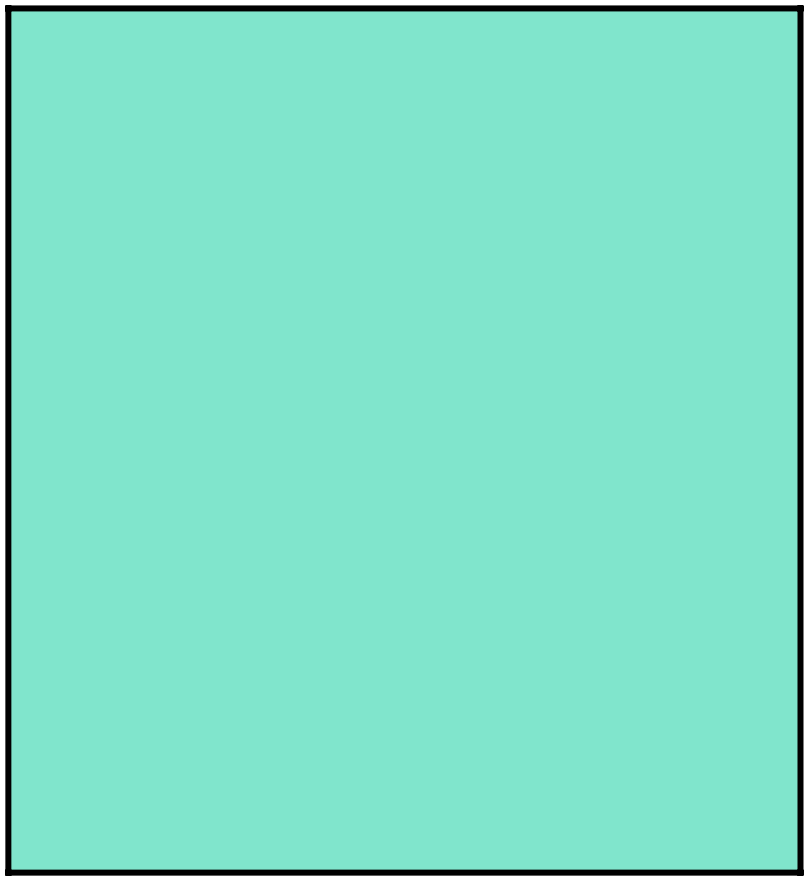
$$RR = \frac{\Pr(\text{Poor BMI, given Poor SED})}{\Pr(\text{Poor BMI, given OK SED})}$$

"The risk of Poor BMI is 1.6 times
greater in streams with Poor SED
than in streams with OK SED."

Features in Space as GIS objects

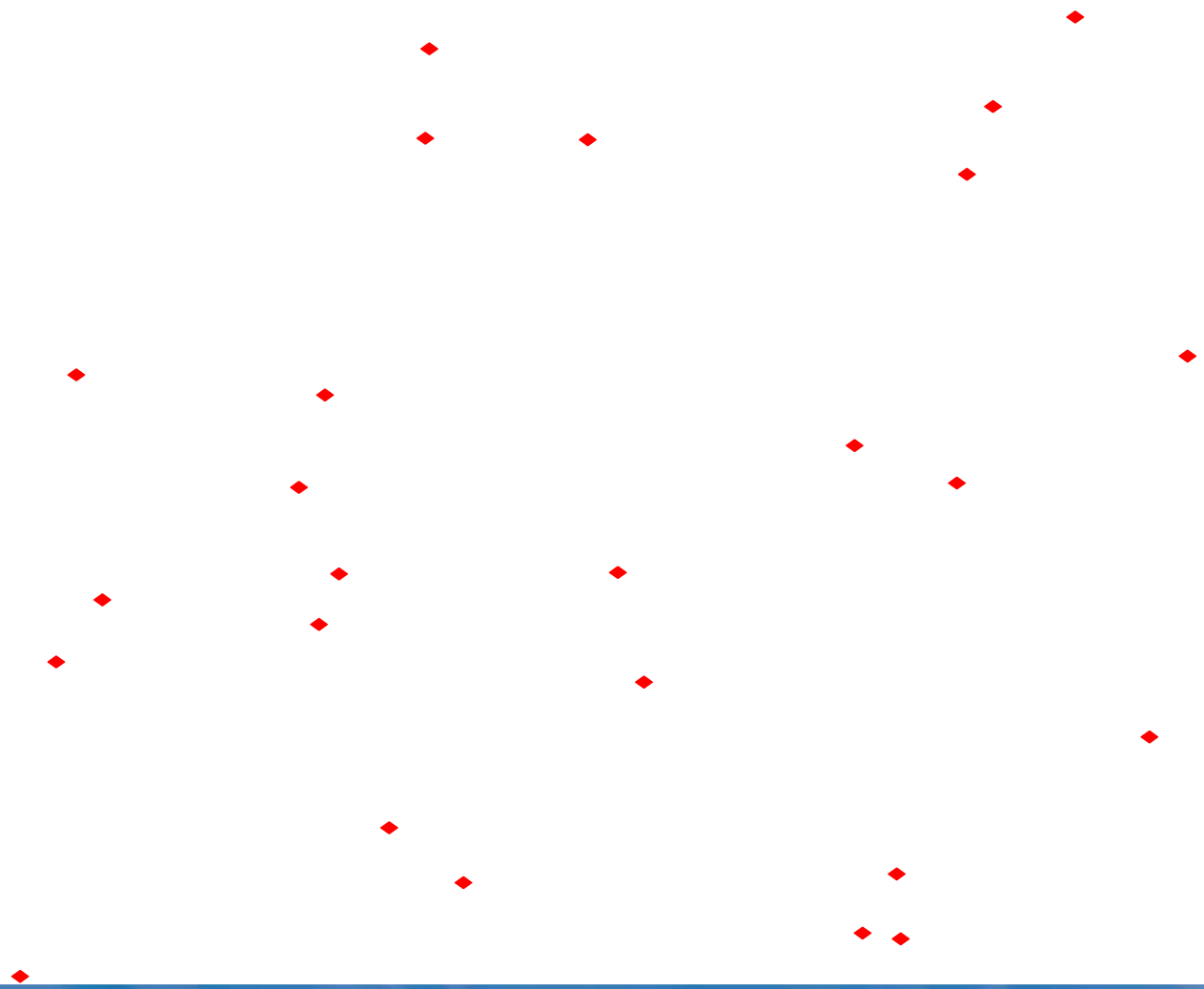
Feature	Points	Lines	Polygons
Lakes	Individual lakes		Lake area
Streams	Segments	Linear network	
Estuaries			Estuarine area
Wetlands	Depressional wetlands		Wetland area
Marine waters			areas
Hydrologic Units	Yes		Yes
Buildings	Yes		

Sampling an area

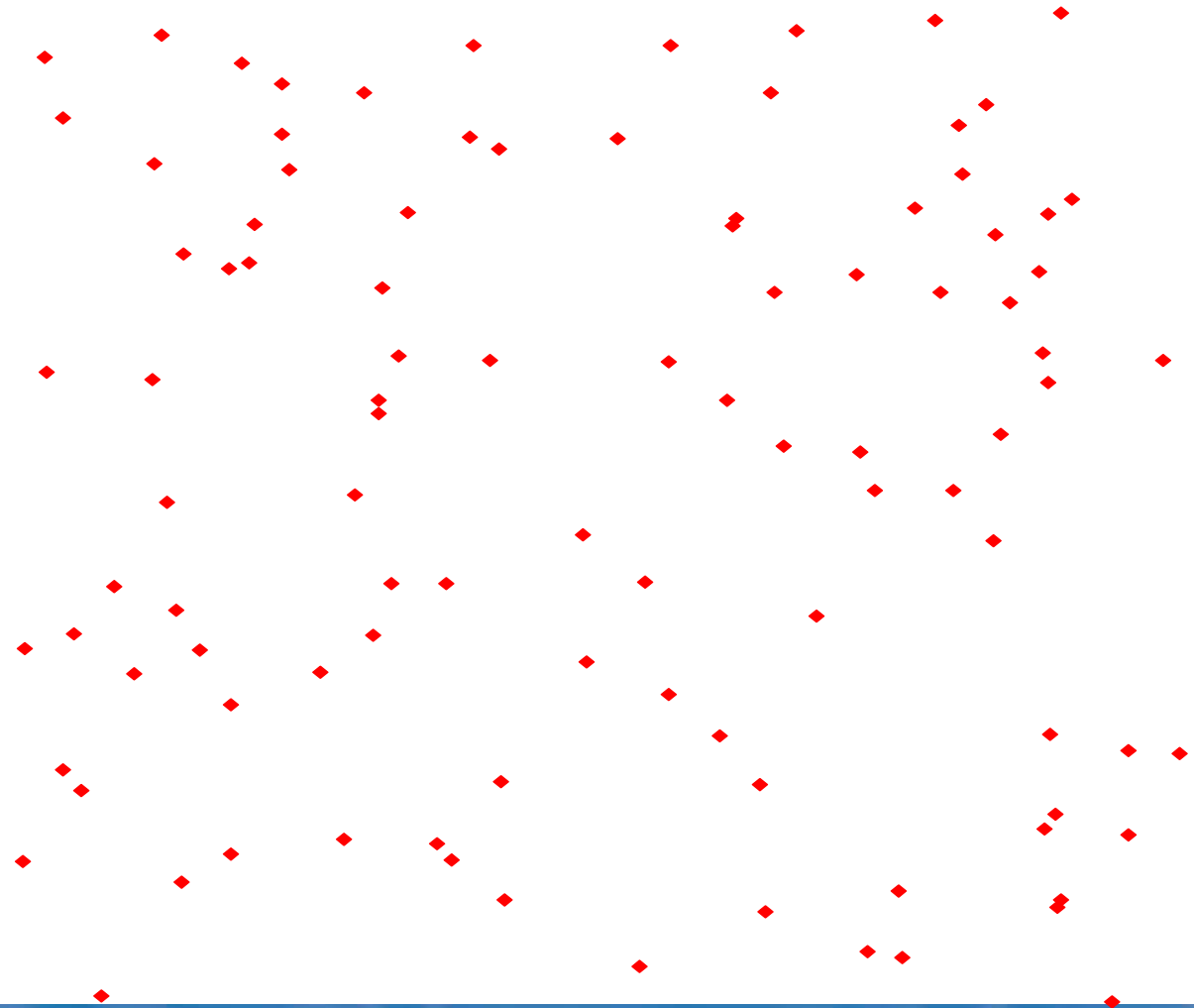


- Assume know nothing about the region
- How would select a set of sites to characterize the region?
- Sites should be “representative” of region
- Systematic point grid is appealing
- Triangular point grid is optimal when have a known isotropic semi-variogram in spatial statistics

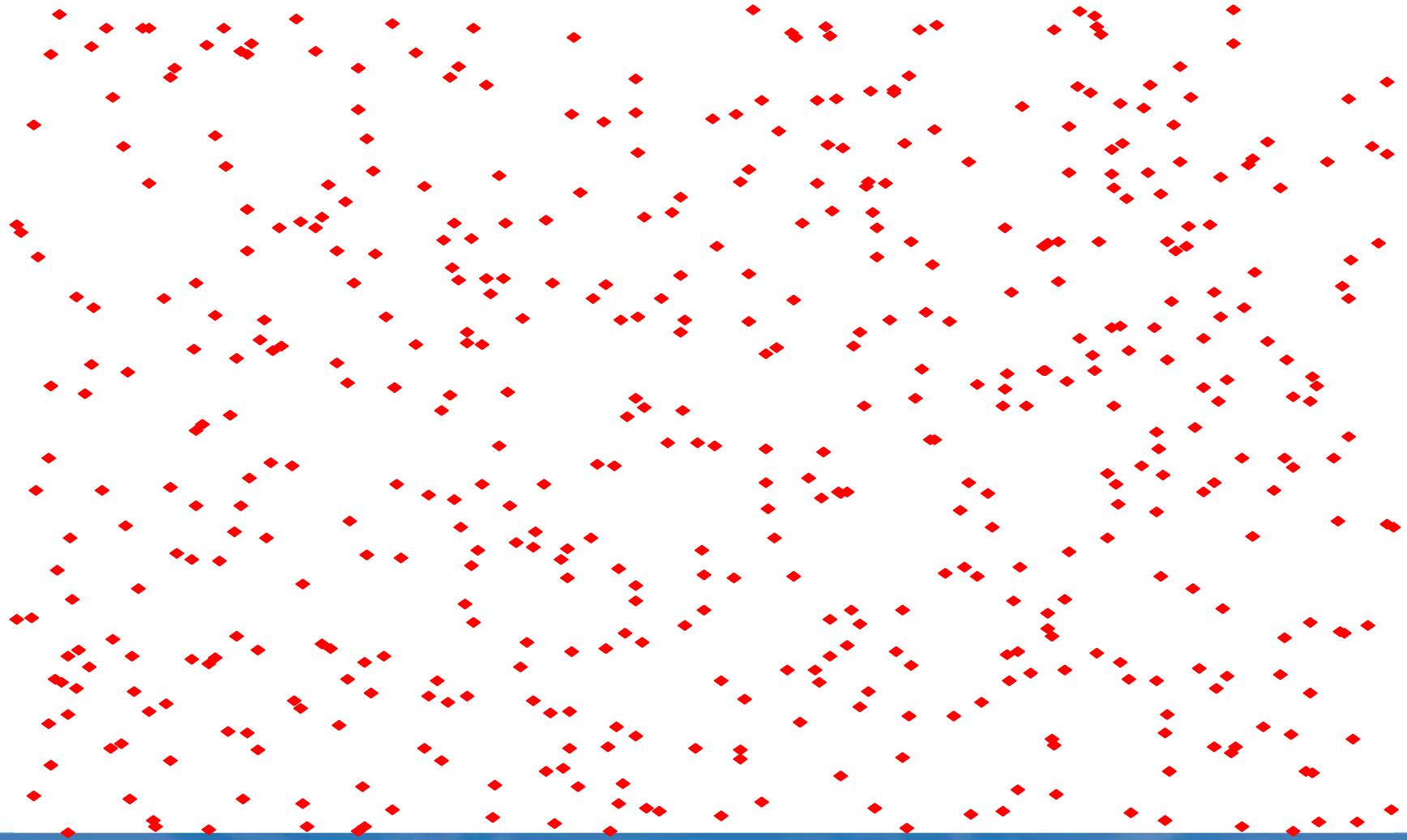
TWO DIMENSIONAL RANDOM POINTS, n = 25



TWO DIMENSIONAL RANDOM POINTS, n = 100



TWO DIMENSIONAL RANDOM POINTS, n = 500



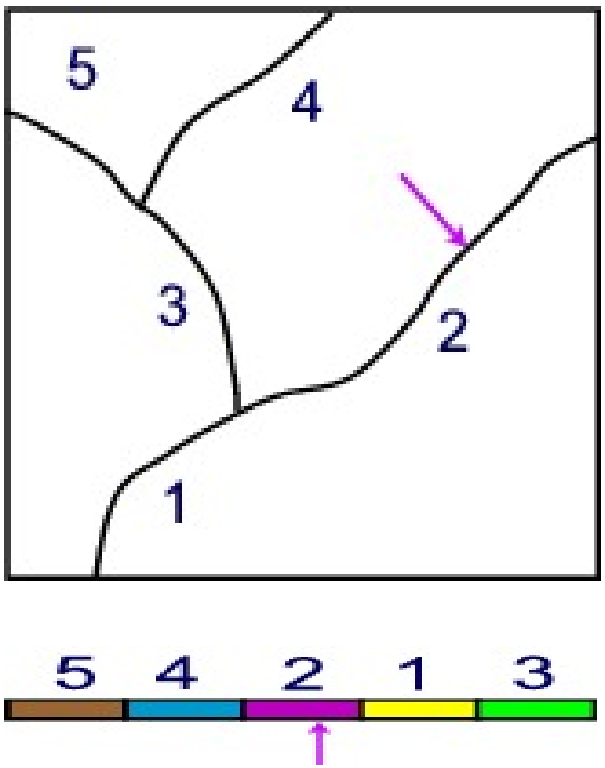
Desirable Properties of Environmental Resource Samples

- (1) Accommodate varying spatial sample intensity
- (2) Spread the sample points evenly and regularly over the domain, subject to (1)
- (3) Allow augmentation of the sample after-the-fact, while maintaining (2)
- (4) Accommodate varying population spatial density for finite & linear populations, subject to (1) & (2).
- (2) + (4) $\Rightarrow$ Sample spatial pattern should reflect the (finite or linear) population spatial pattern

GRTS Implementation Steps

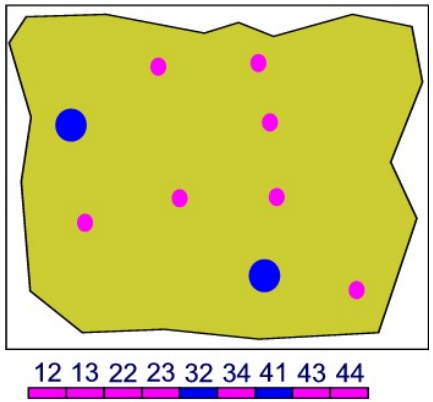
- Concept of selecting a probability sample from a sampling line for the resource
- Create a hierarchical grid with hierarchical addressing
- Randomize hierarchical addresses
- Construct sampling line using randomized hierarchical addresses
- Select a systematic sample with a random start from sampling line
- Place sample in reverse hierarchical address order

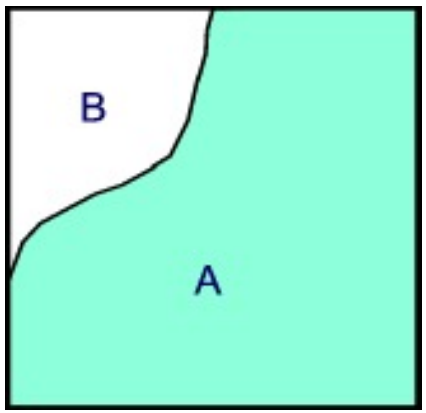
Selecting a Probability Sample from a Sampling Line: Linear Network Case



- Place all stream segments in frame on a linear line
  - Preserve segment length
  - Identify segments by ID
- In what order do place segments on line?
  - Randomly
  - Systematically (minimal spanning tree)
  - Randomized hierarchical grid
- Systematic sample with random start
  - k=L/n, L=length of line, n=sample size
  - Random start d between [0,k)
  - Sample: d + (i-1)*k for i=1,…,n

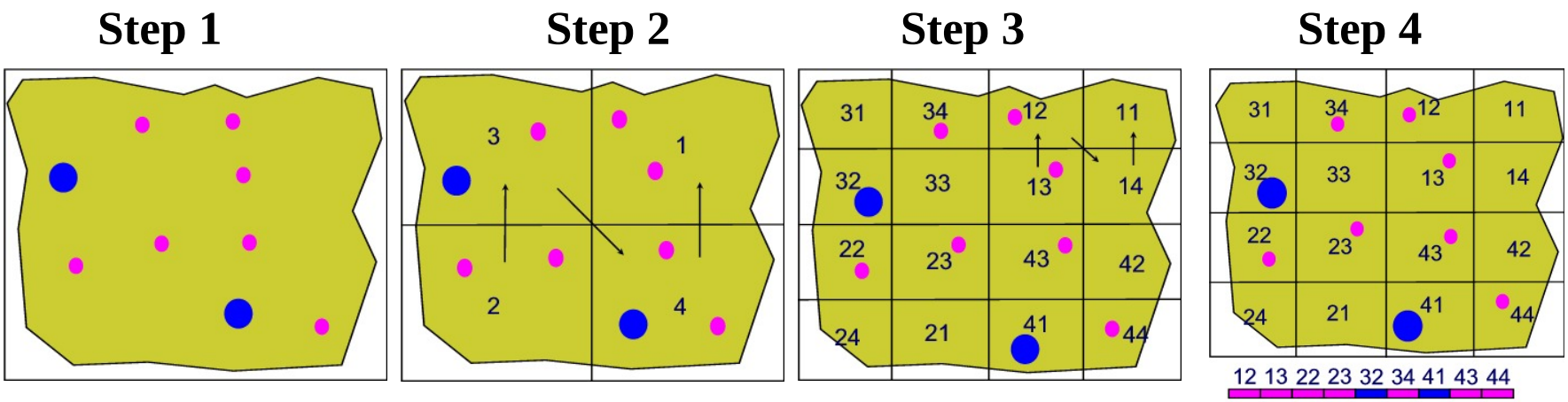
Selecting a Probability Sample from a Sampling Line: Point and Area Cases





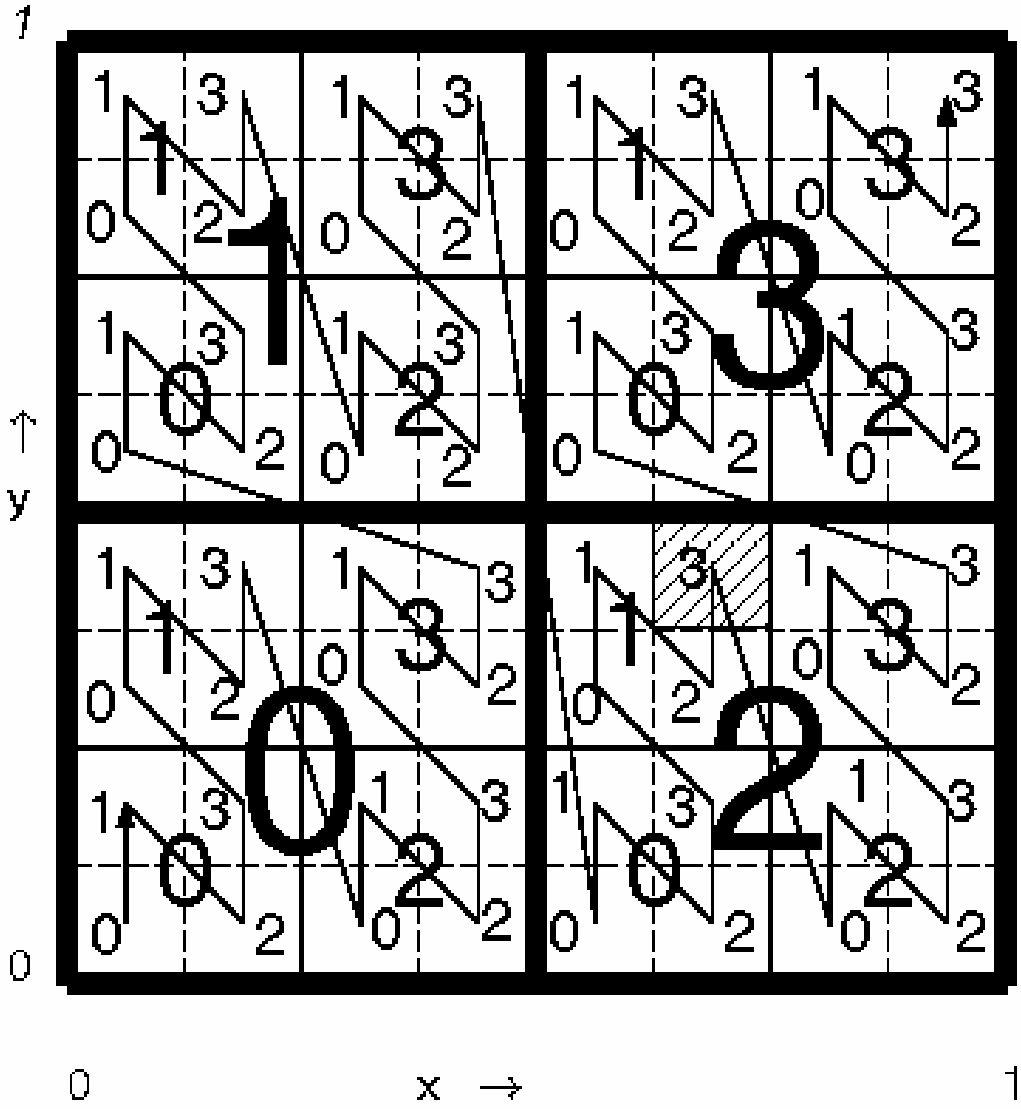
- Point Case:
  - Identify all points in frame
  - Assign each point unit length
  - Place on sample line
- Area Case:
  - Create grid covering region of interest
  - Generate random points within each grid cell
  - Keep random points within resource (A)
  - Assign each point unit length
  - Place on sample line

Randomized Hierarchical Grid



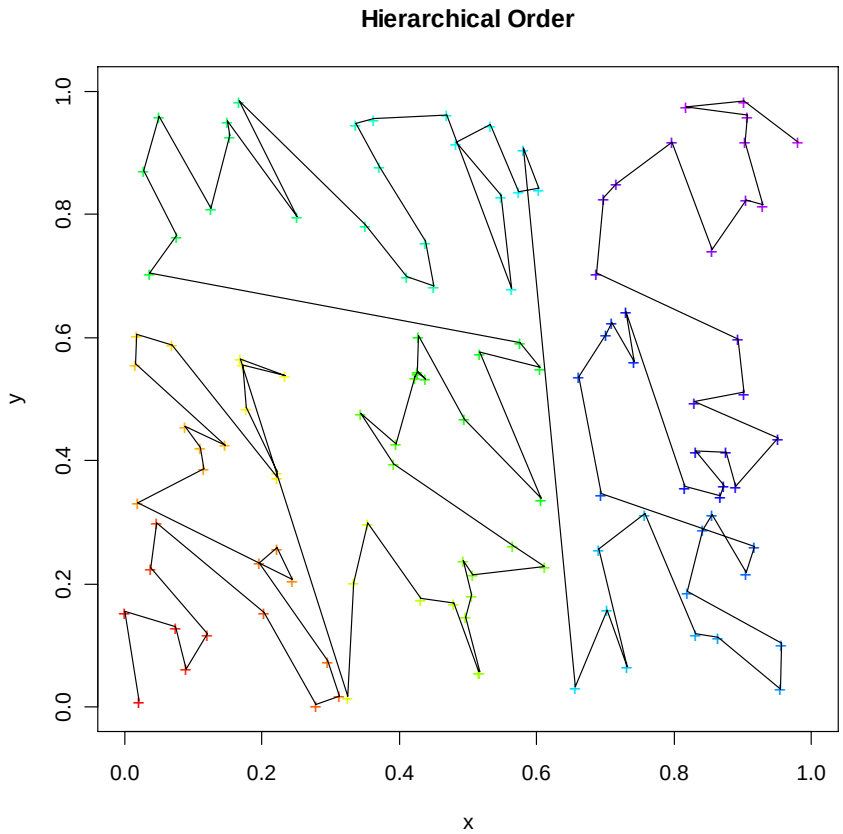
- Step 1: Frame: Large lakes: blue; Small lakes: pink; Randomly place grid over the region
- Step 2: Sub-divide region and randomly assign numbers to sub-regions
- Step 3: Sub-divide sub-regions; randomly assign numbers independently to each new sub-region; create hierarchical address. Continue sub-dividing until only one lake per cell.
- Step 4: Identify each lake with cell address; assign each lake length 1; place lakes on line in numerical cell address order.

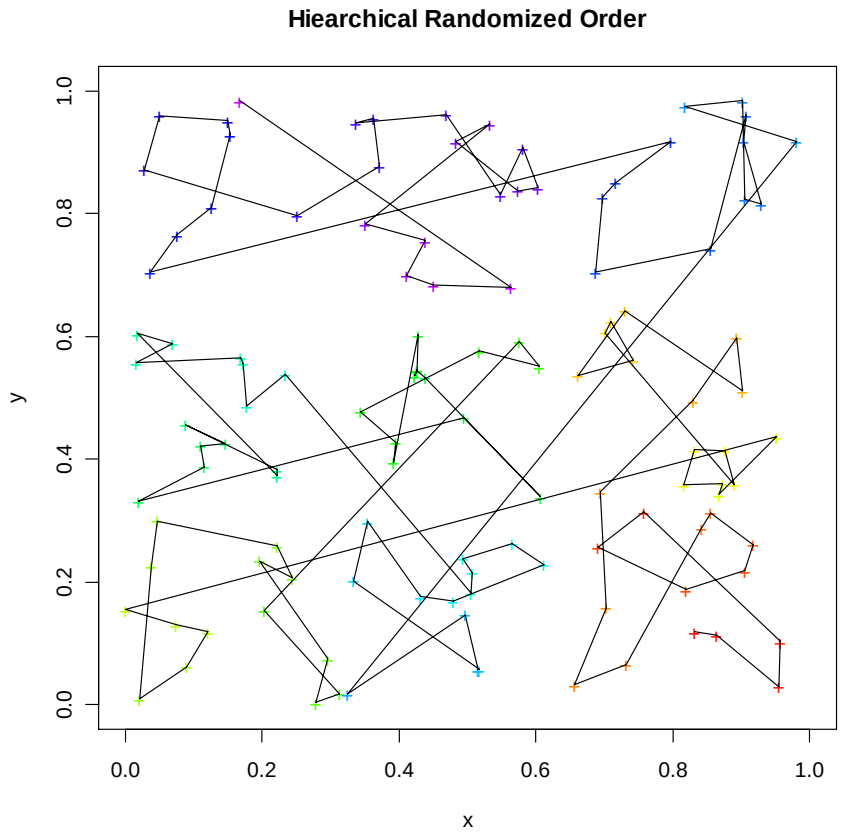
Hierarchical Grid Addressing



213: hierarchical address

Population of 120 points



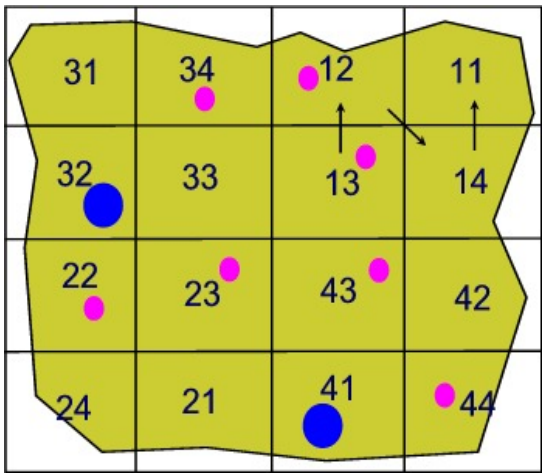


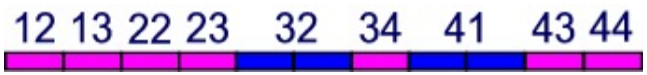
RHO	Reverse Base4	Base4	Original Order
1	00	00	1
2	01	10	5
3	02	20	9
4	03	30	13
5	10	01	2
6	11	11	6
7	12	21	10
8	13	31	14
9	20	02	3
10	21	12	7
11	22	22	11
12	23	32	15
13	30	03	4
14	31	13	8
15	32	23	12
16	33	33	16

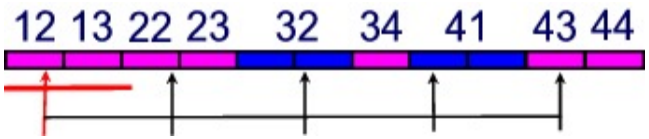
Reverse Hierarchical Order

- Construct reverse hierarchical order
  - Order the sites from 1 to n
  - Create base 4 address for numbers
  - Reverse base 4 address
  - Sort by reverse base 4 address
  - Renumber sites in RHO
- Why use reverse hierarchical order?
  - Results in any contiguous set of sample sites being spatially-balanced
  - Consequence: can begin at the beginning of list and continue using sites until have required number of sites sampled in field

Unequal Probability of Selection





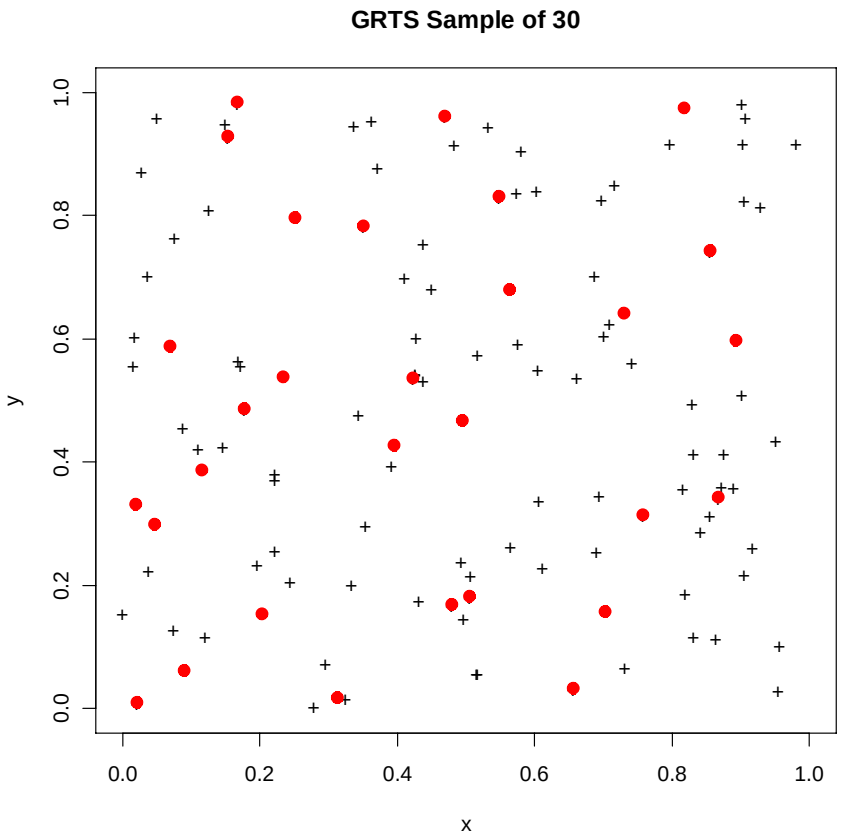


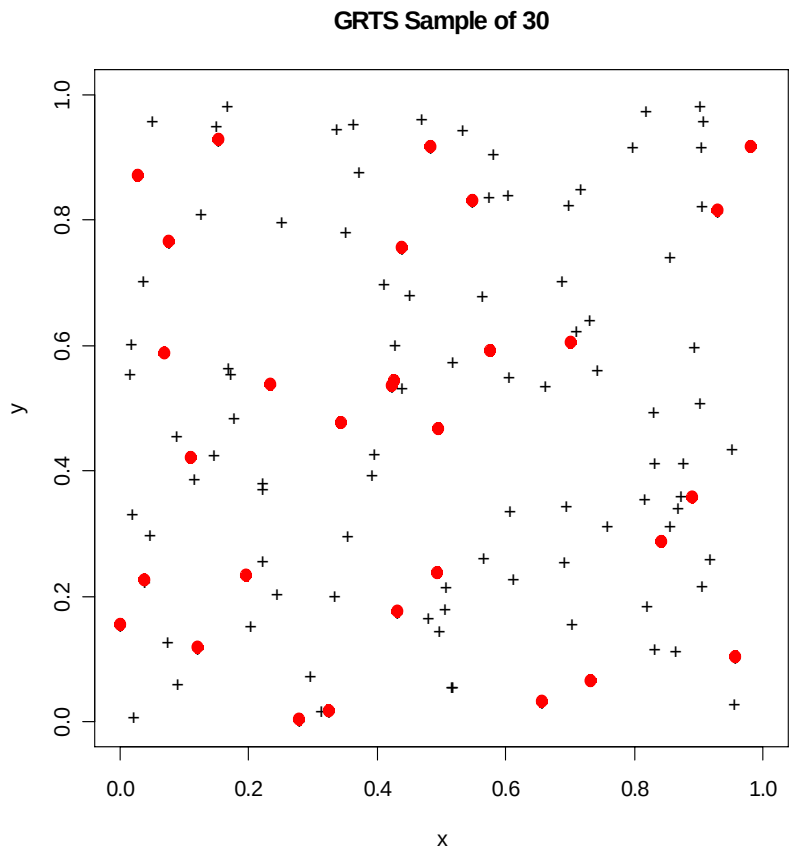
- Assume want large lakes to be twice as likely to be selected as small lakes
- Instead of giving all lakes same unit length, give large lakes twice unit length of small lakes
- To select 5 sites divide line length by 5 (11/5 units); randomly select a starting point within first interval; select 4 additional sites at intervals of 11/5 units
- Same process is used for points and areas (using random points in area)

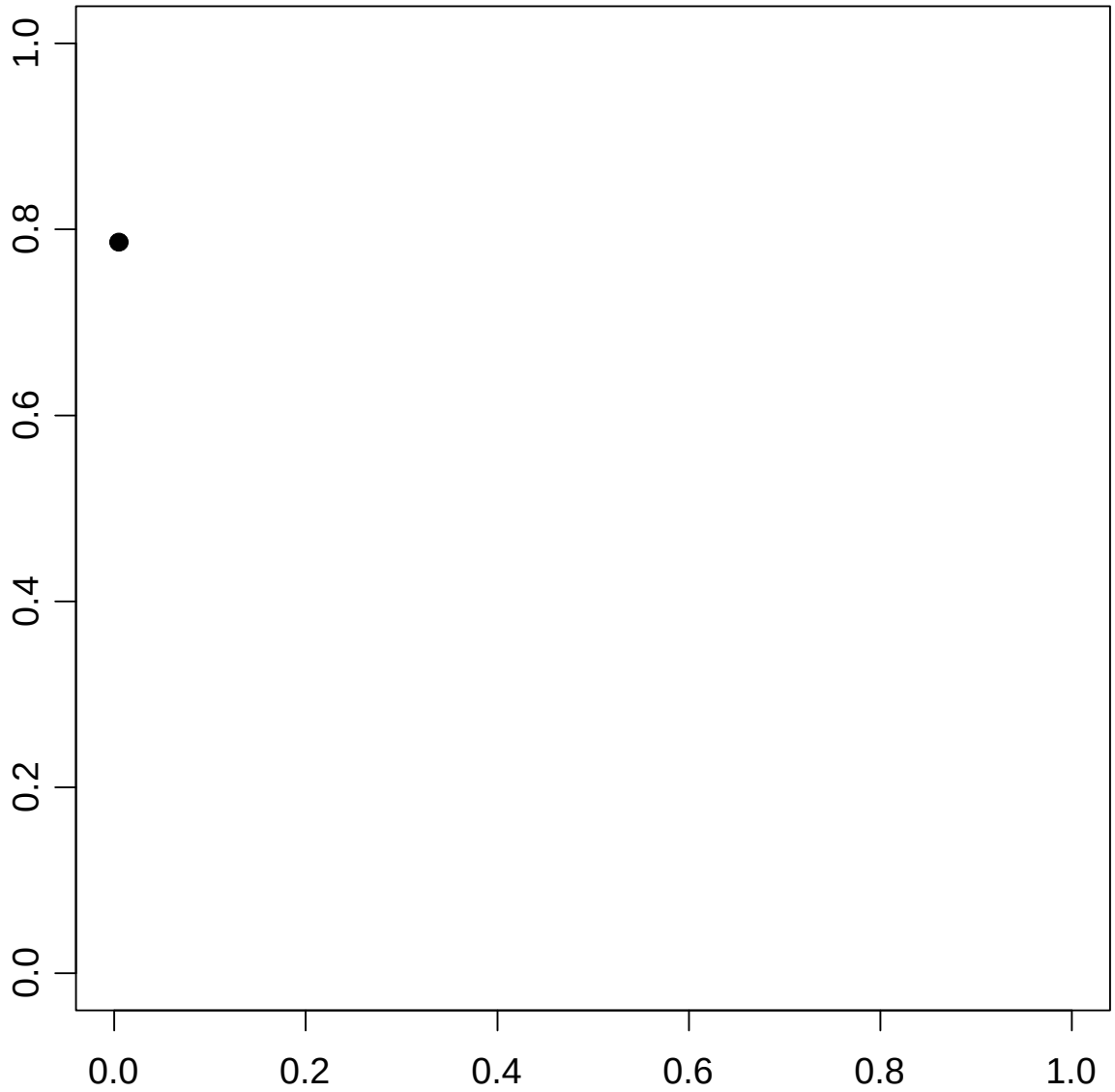
Complex Survey Designs based on GRTS

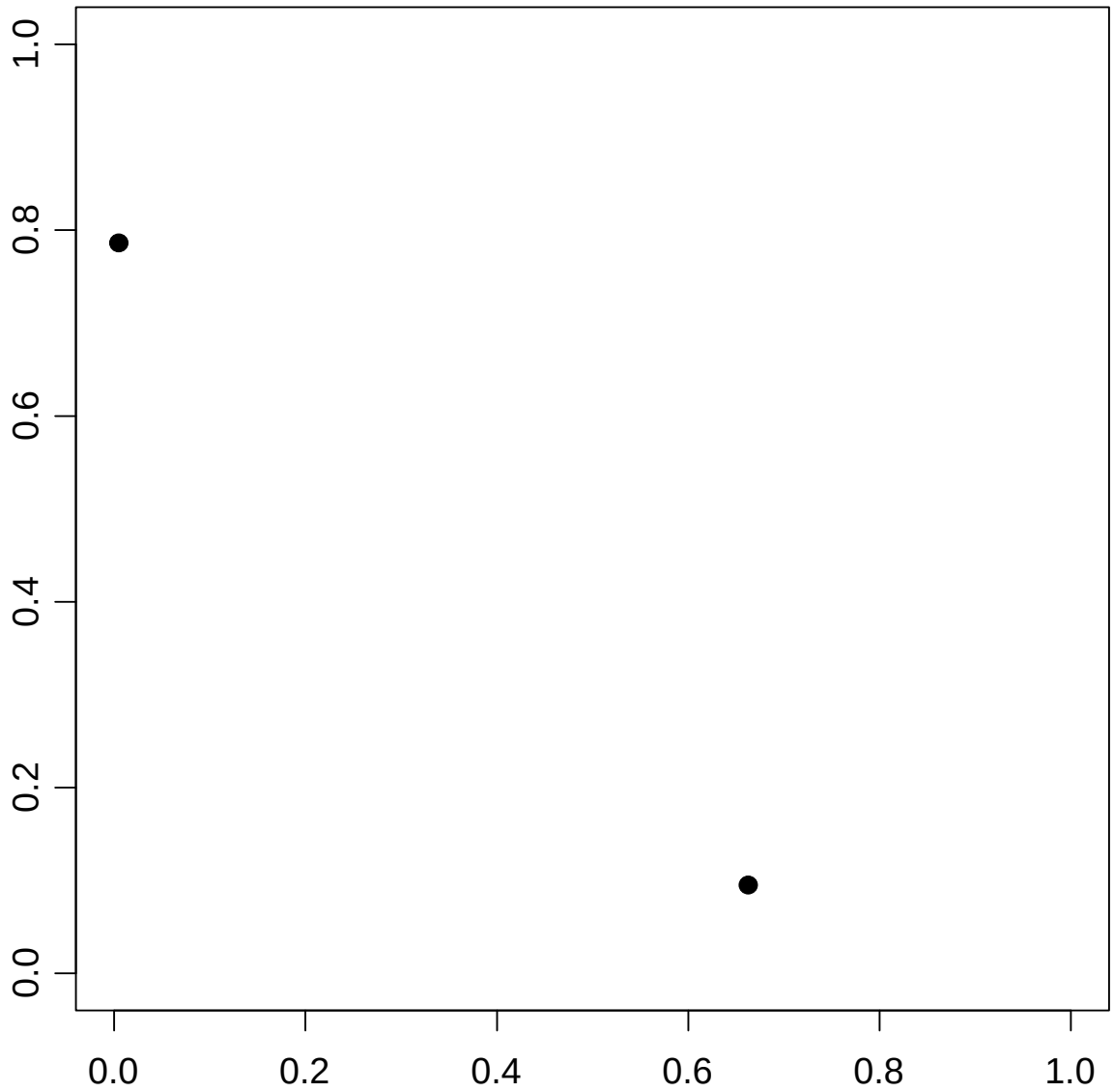
- Stratified GRTS: apply GRTS to each stratum
- Unequal probability GRTS: adjust unit length based on auxiliary information (eg lake area, strahler order, basin, ecoregion)
- Oversample GRTS:
  - Design calls for n sites; some expected non-target, landowner denial, etc; select additional sites to guarantee n field sampled
  - Apply GRTS for sample size 2n; place sites in RHO; use sites in RHO
- Panels for surveys over time
- Nested subsampling
- Two-stage sampling using GRTS at each stage
  - Example: Select USGS 4th field Hucs; then stream sites within Hucs

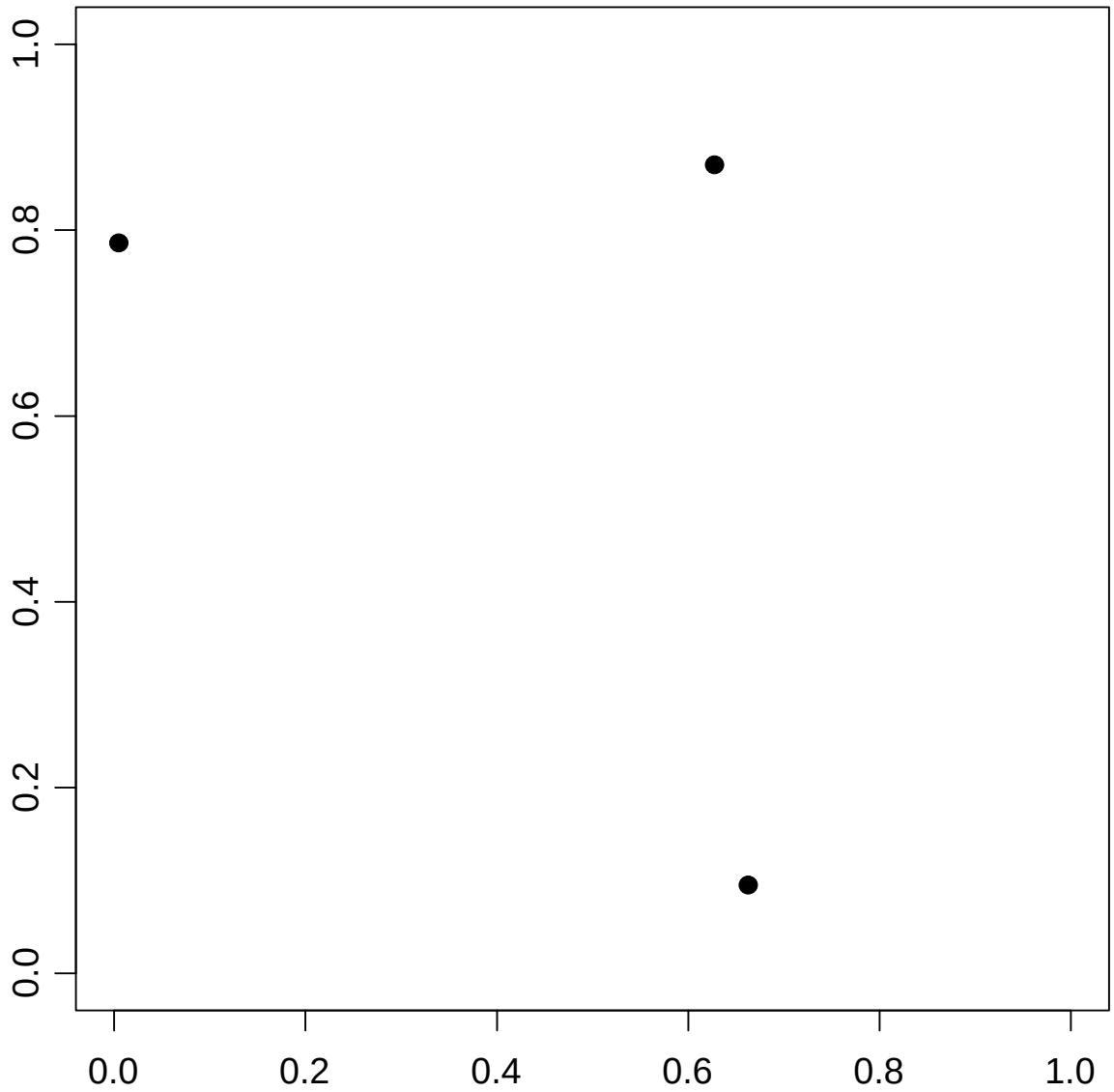
Two GRTS samples: Size 30

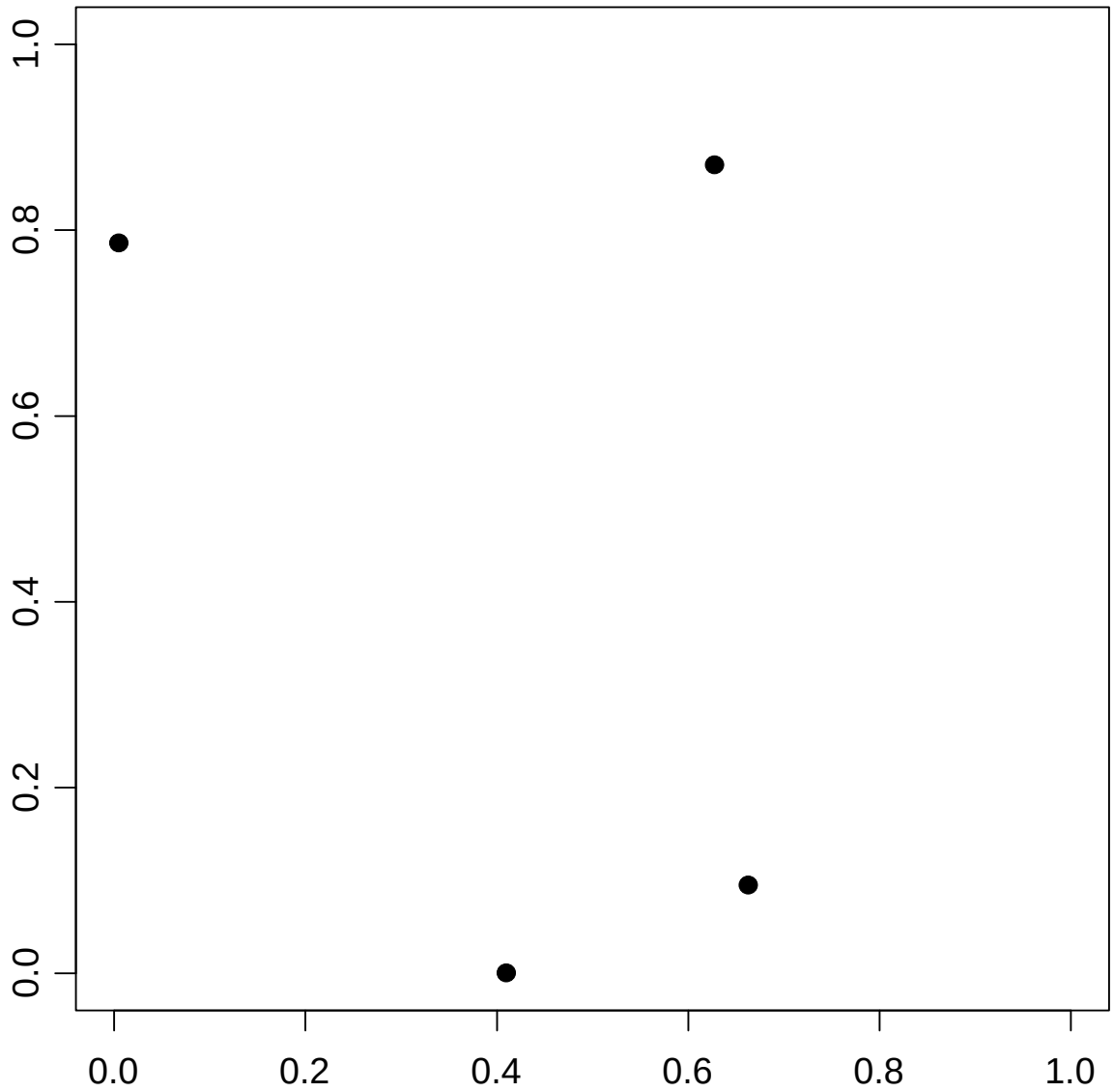


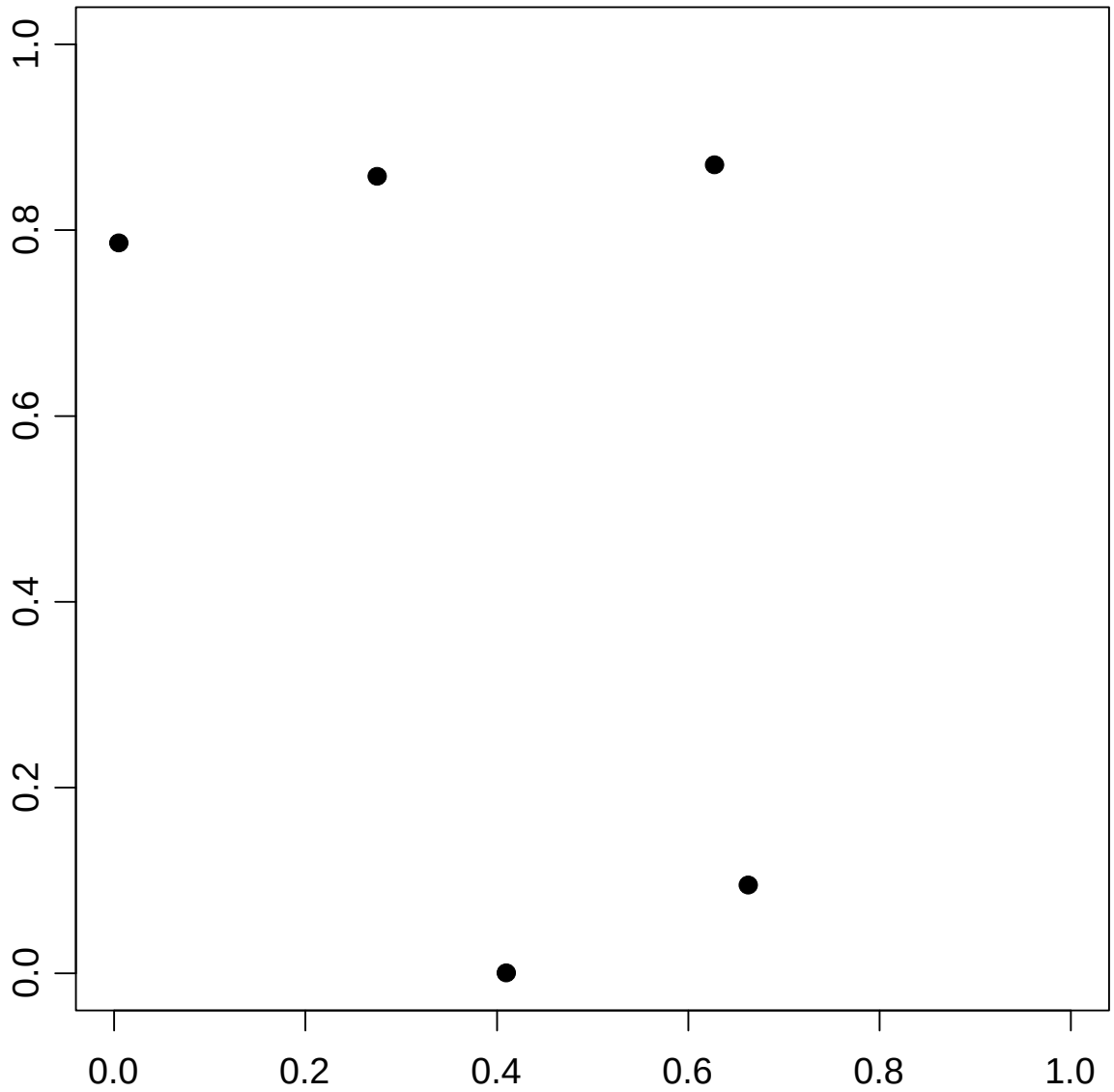


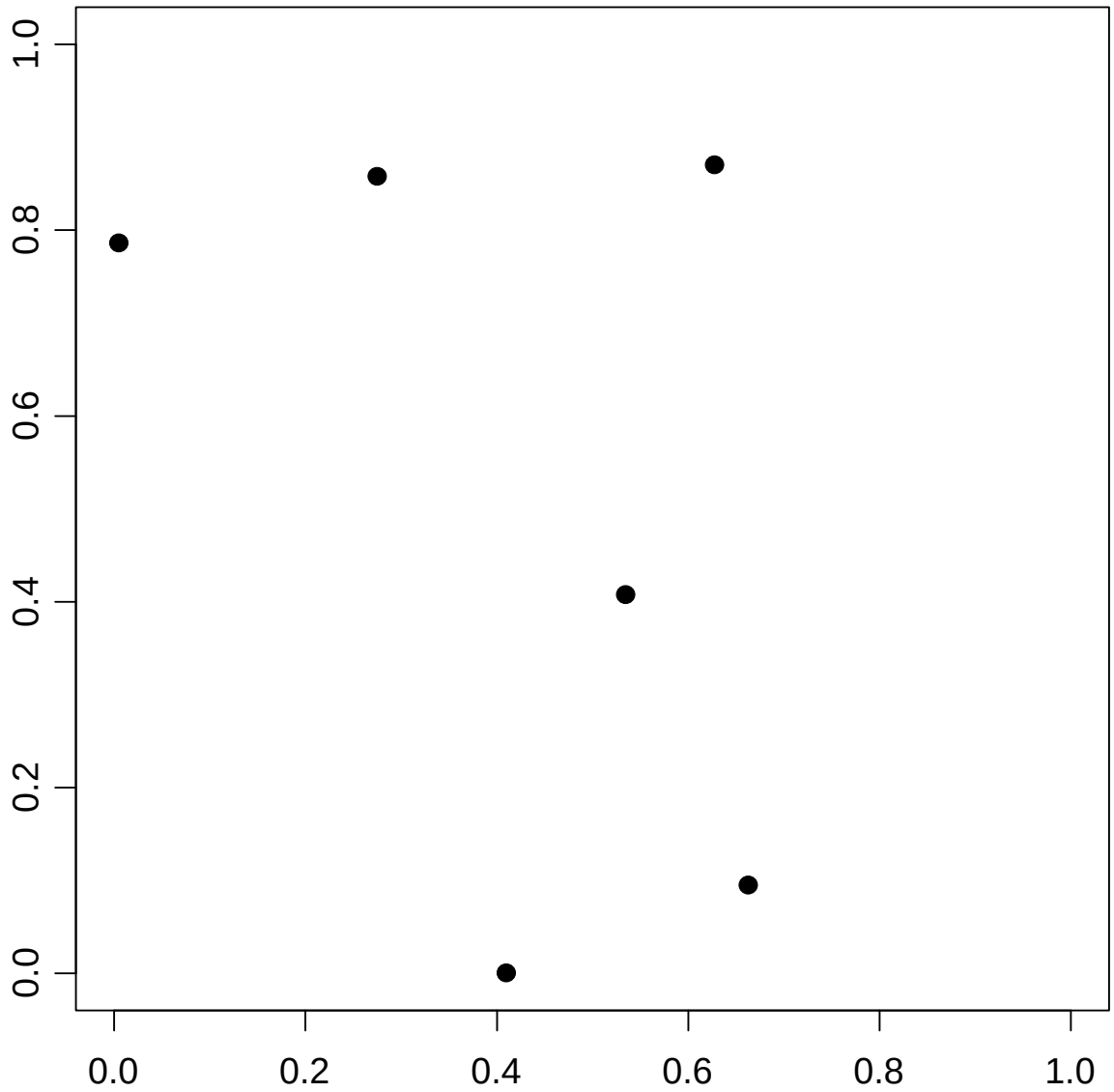

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

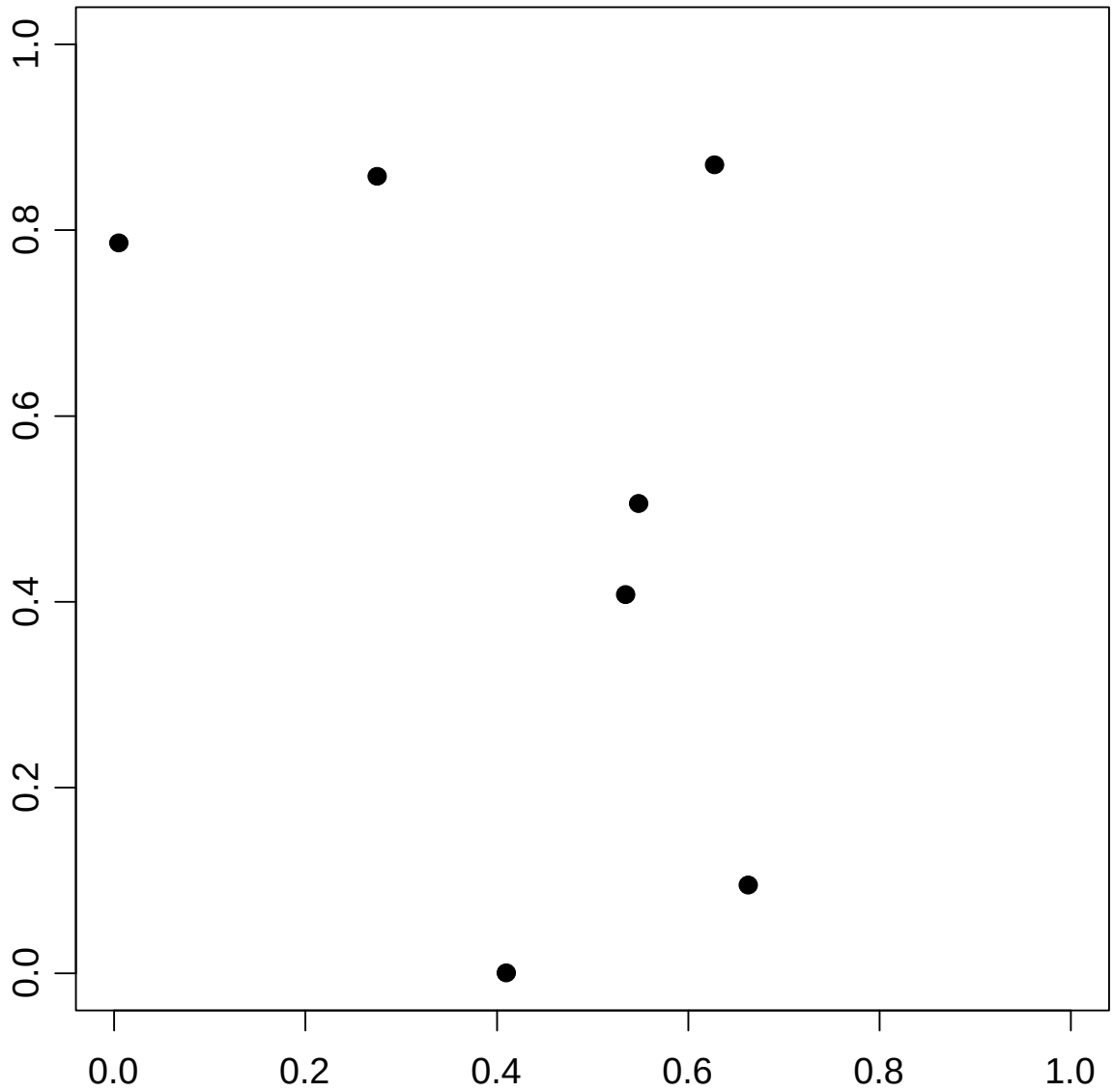

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

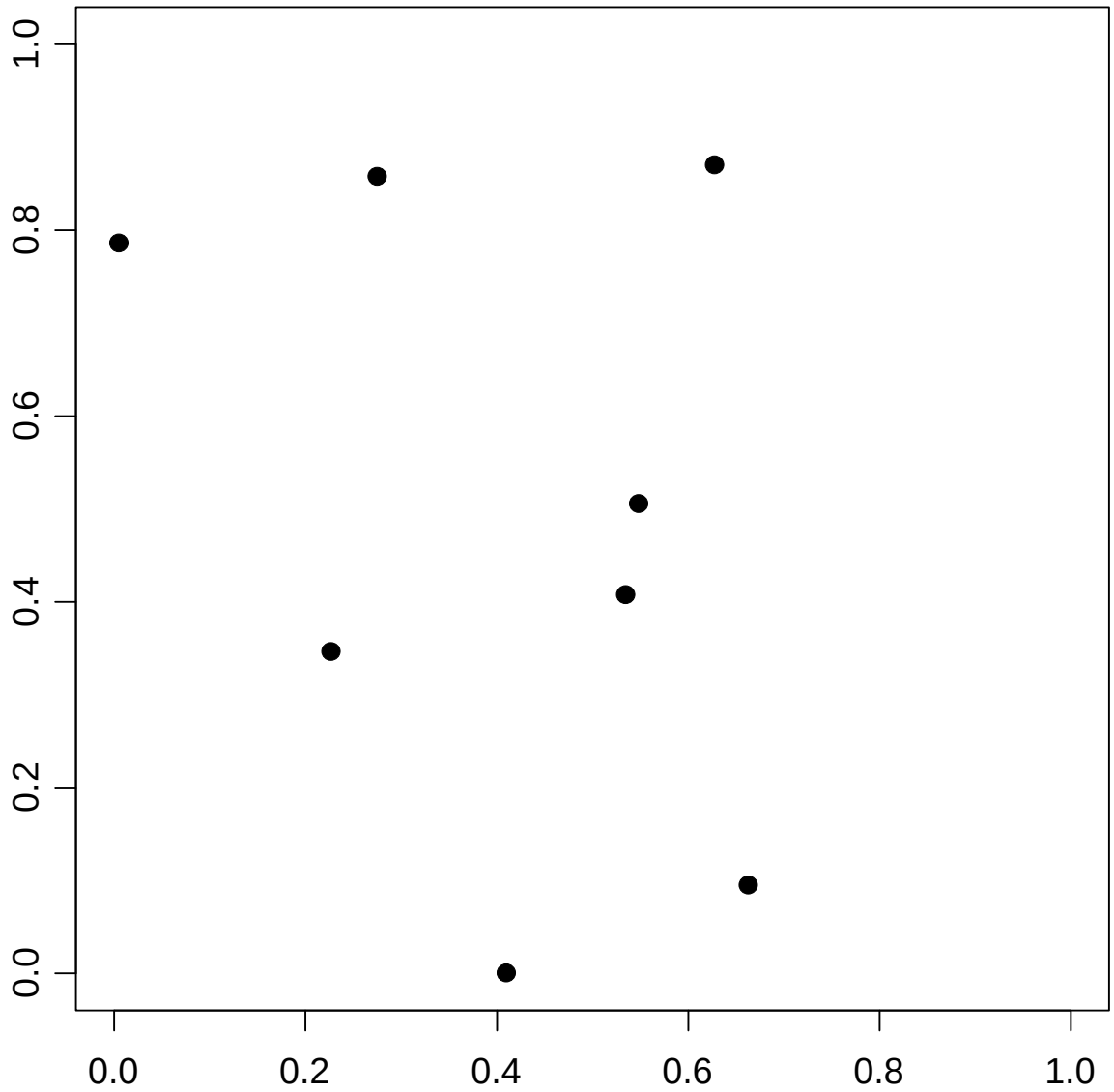

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

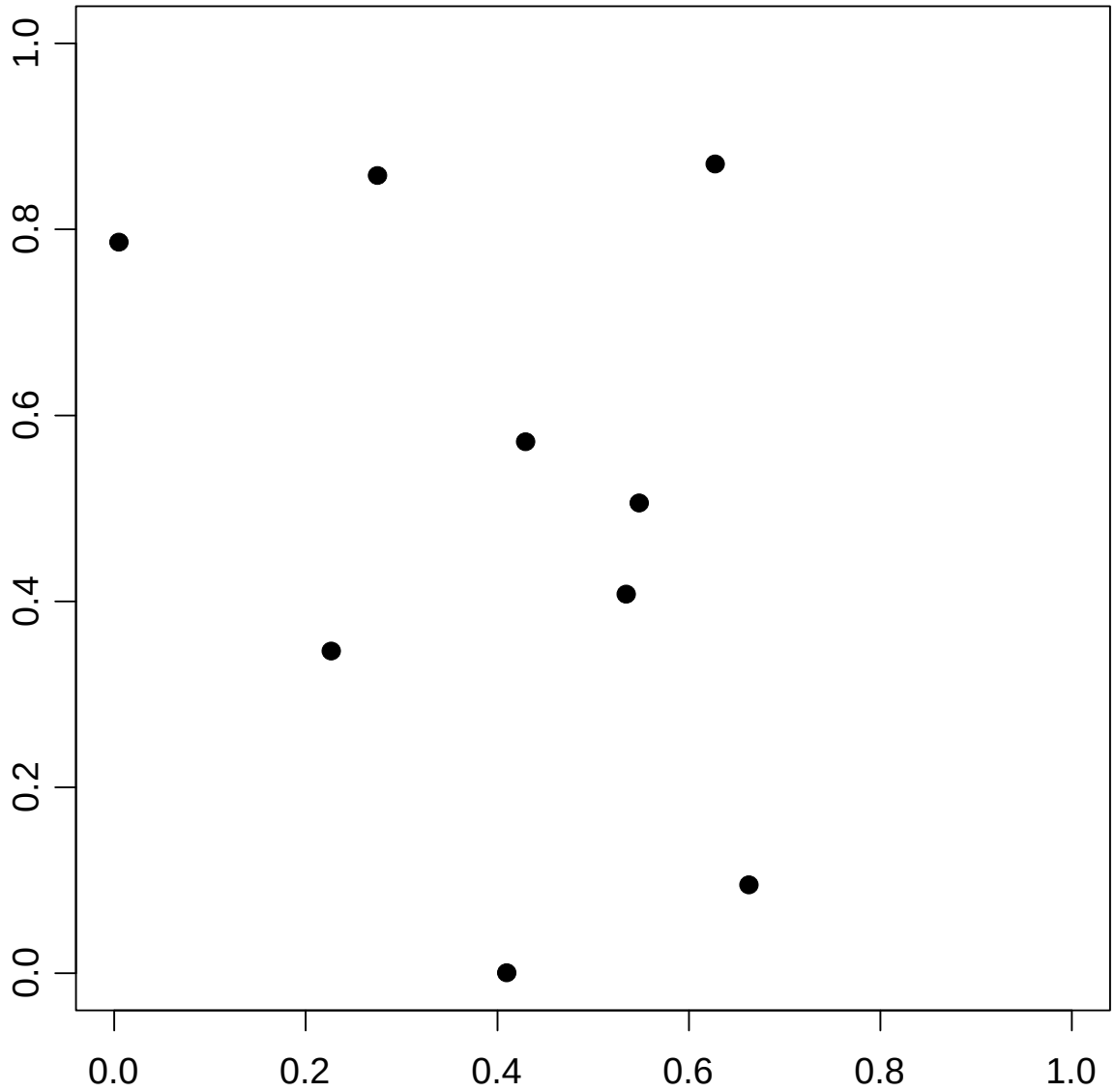

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

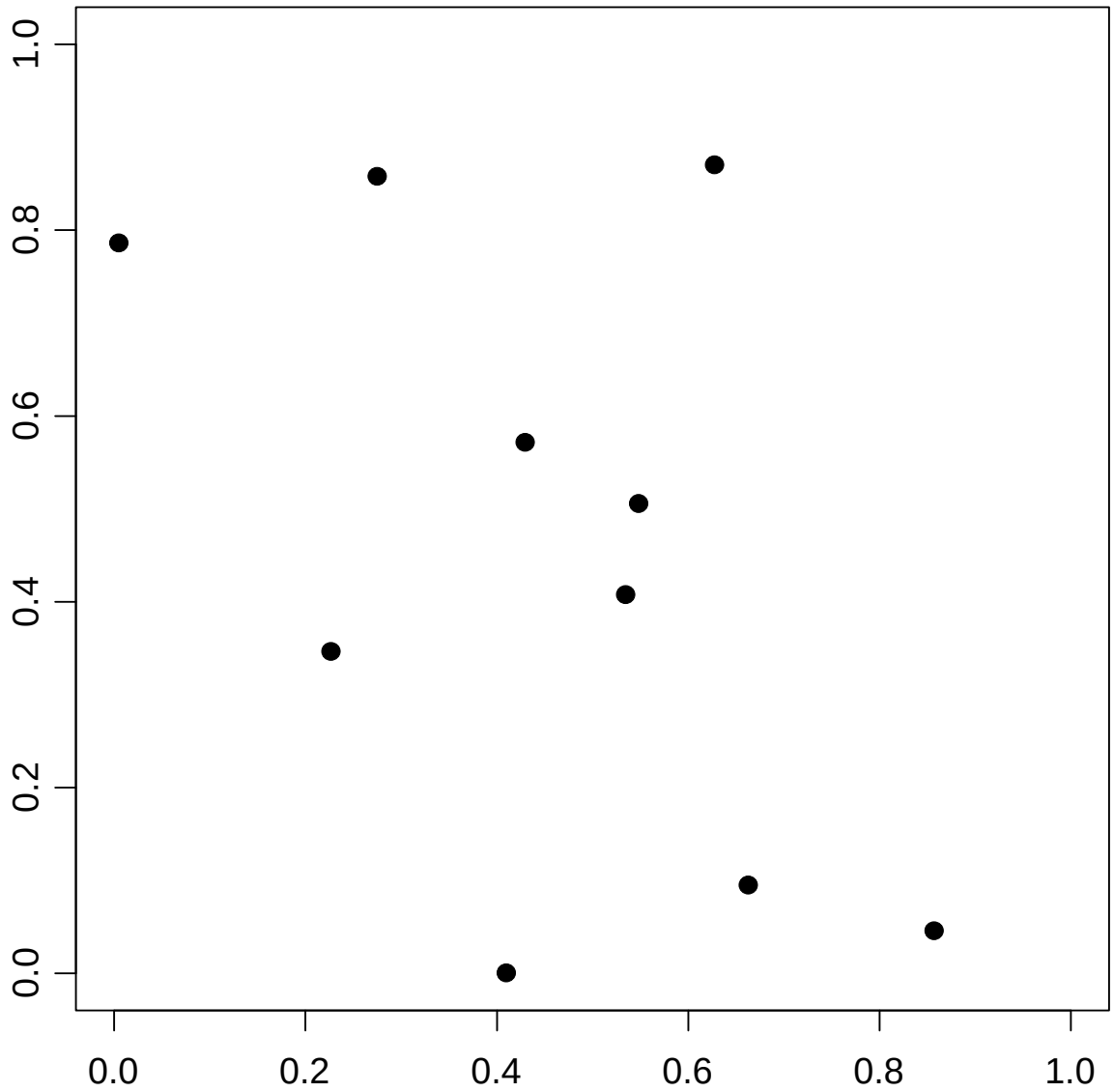

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

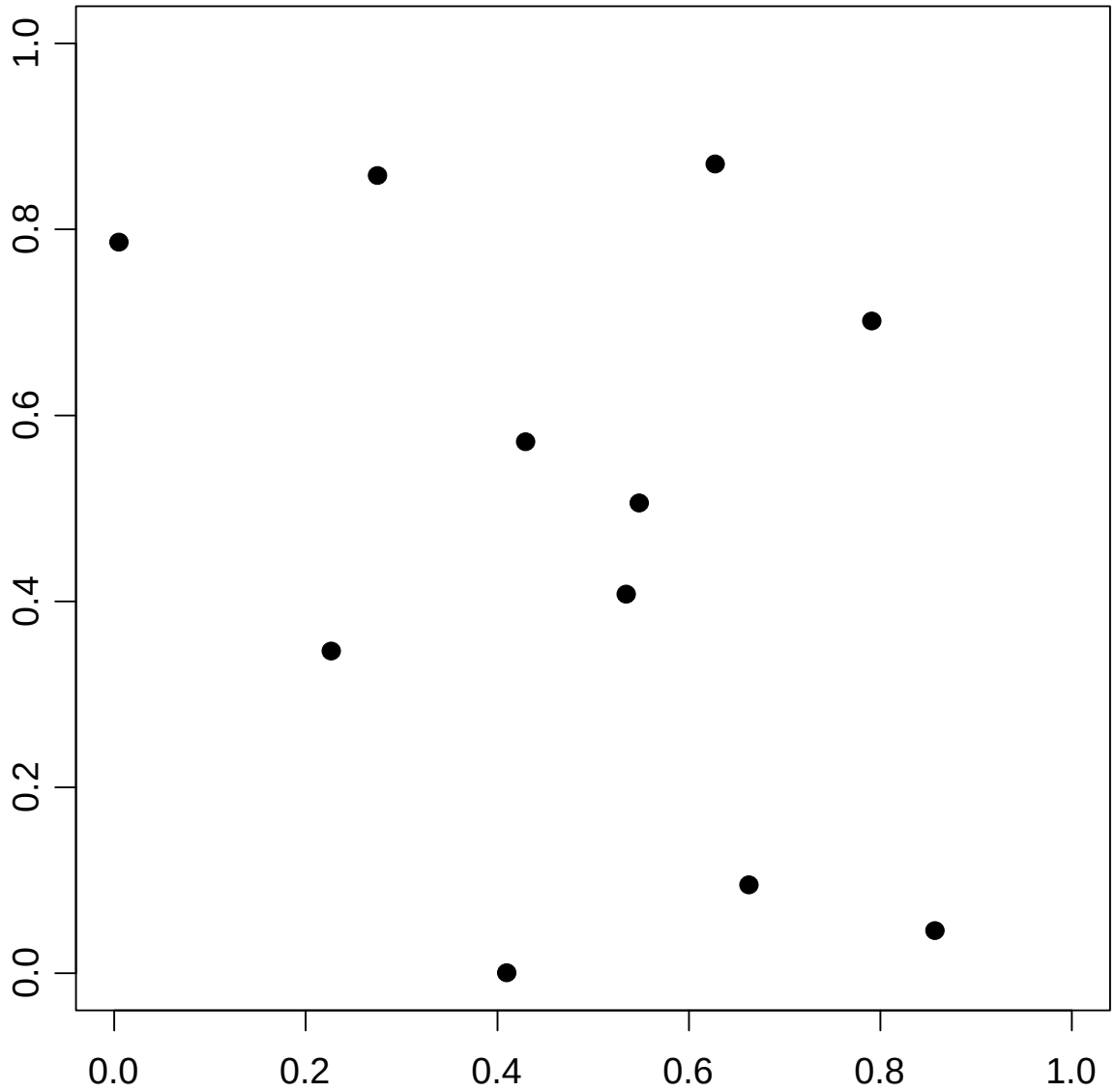

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

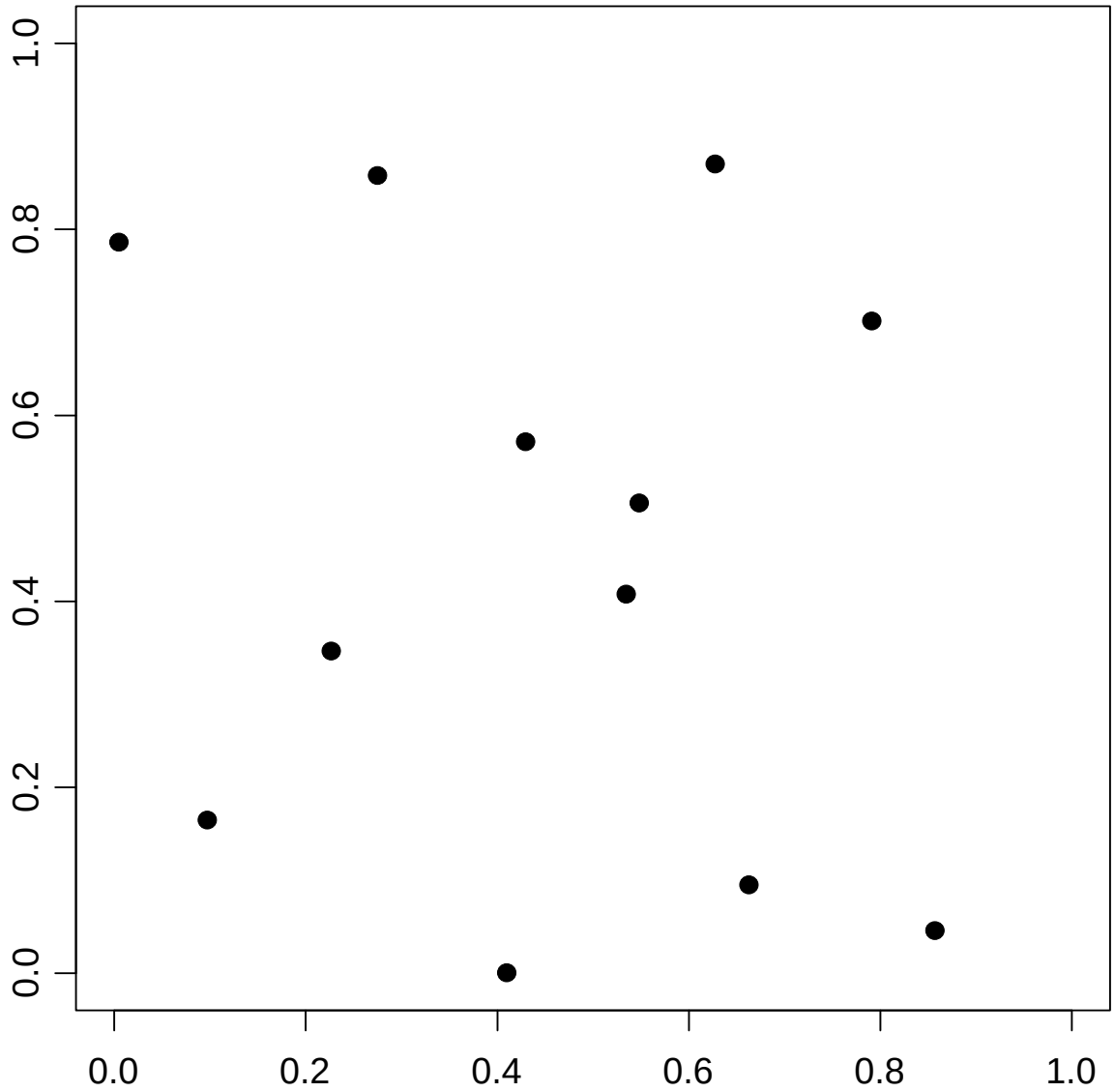

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

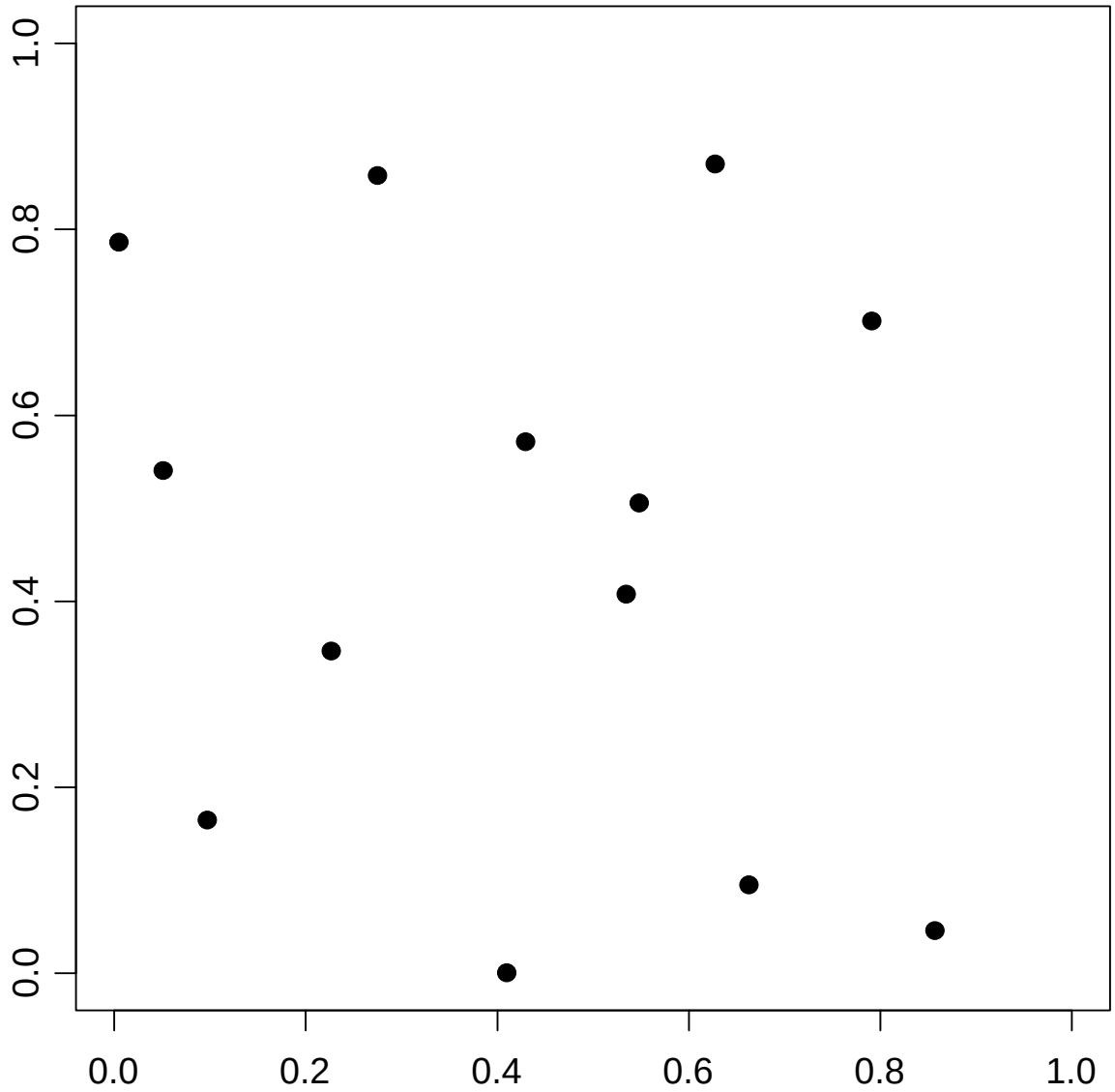

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

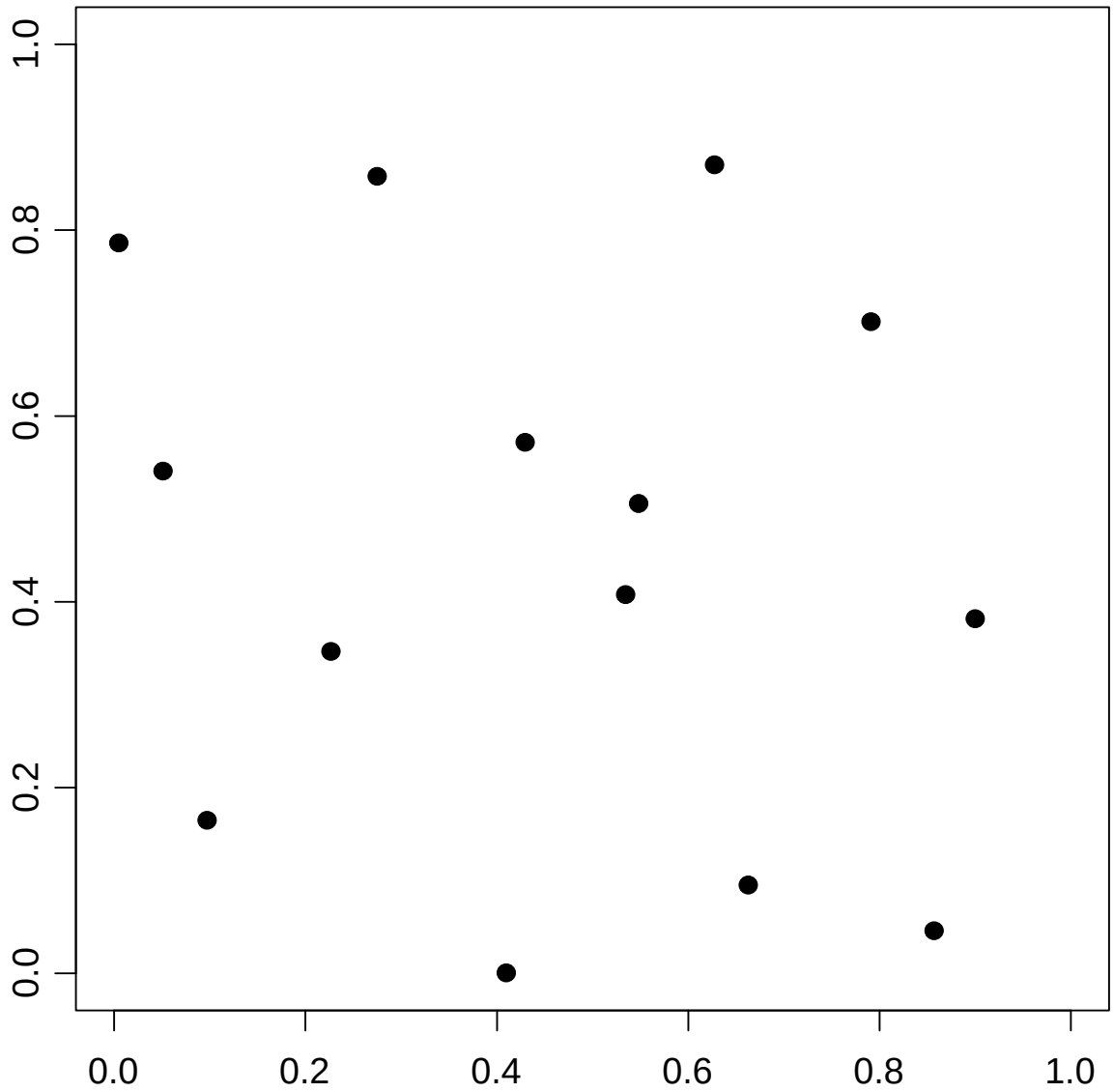

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

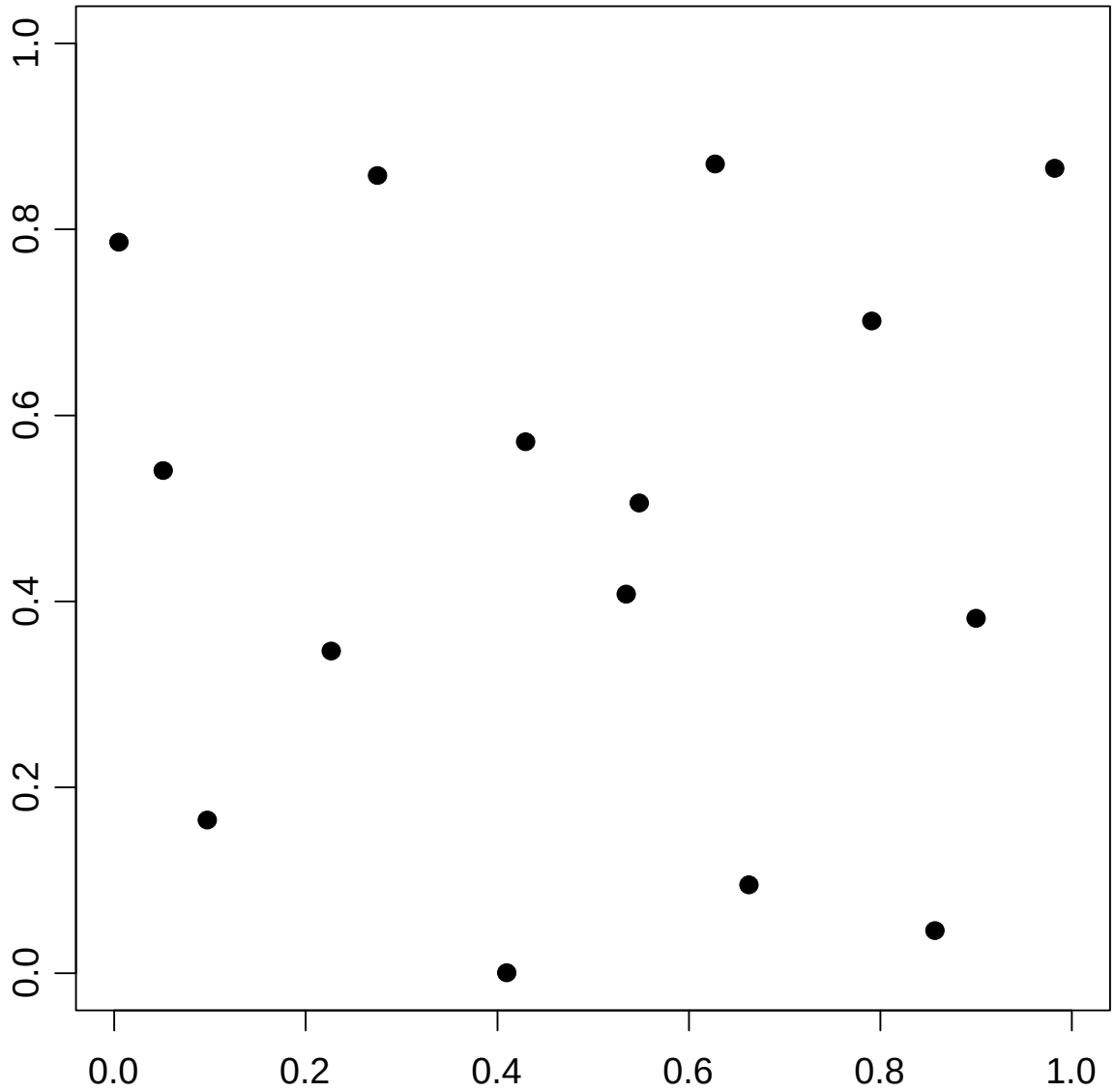

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

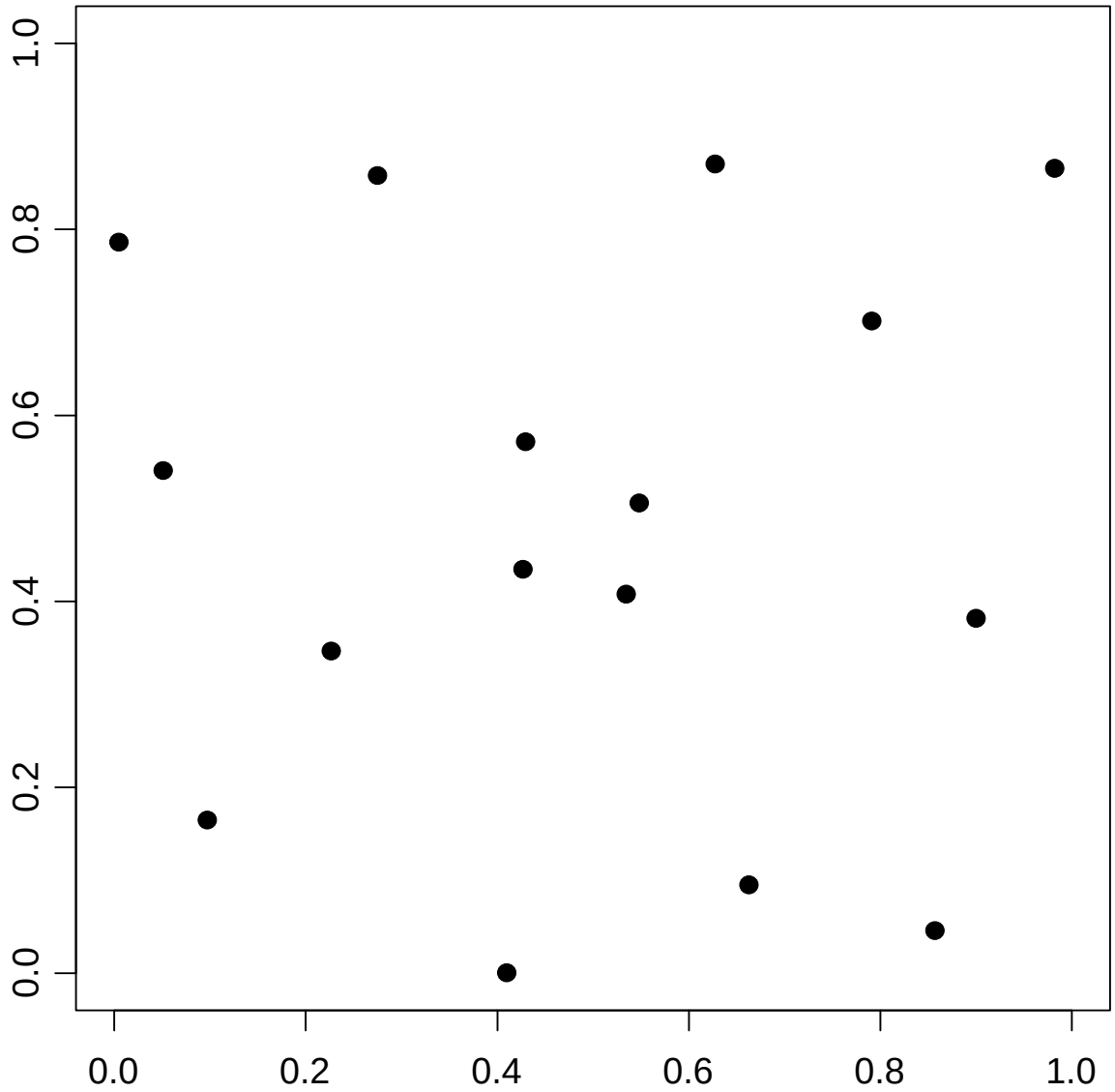

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

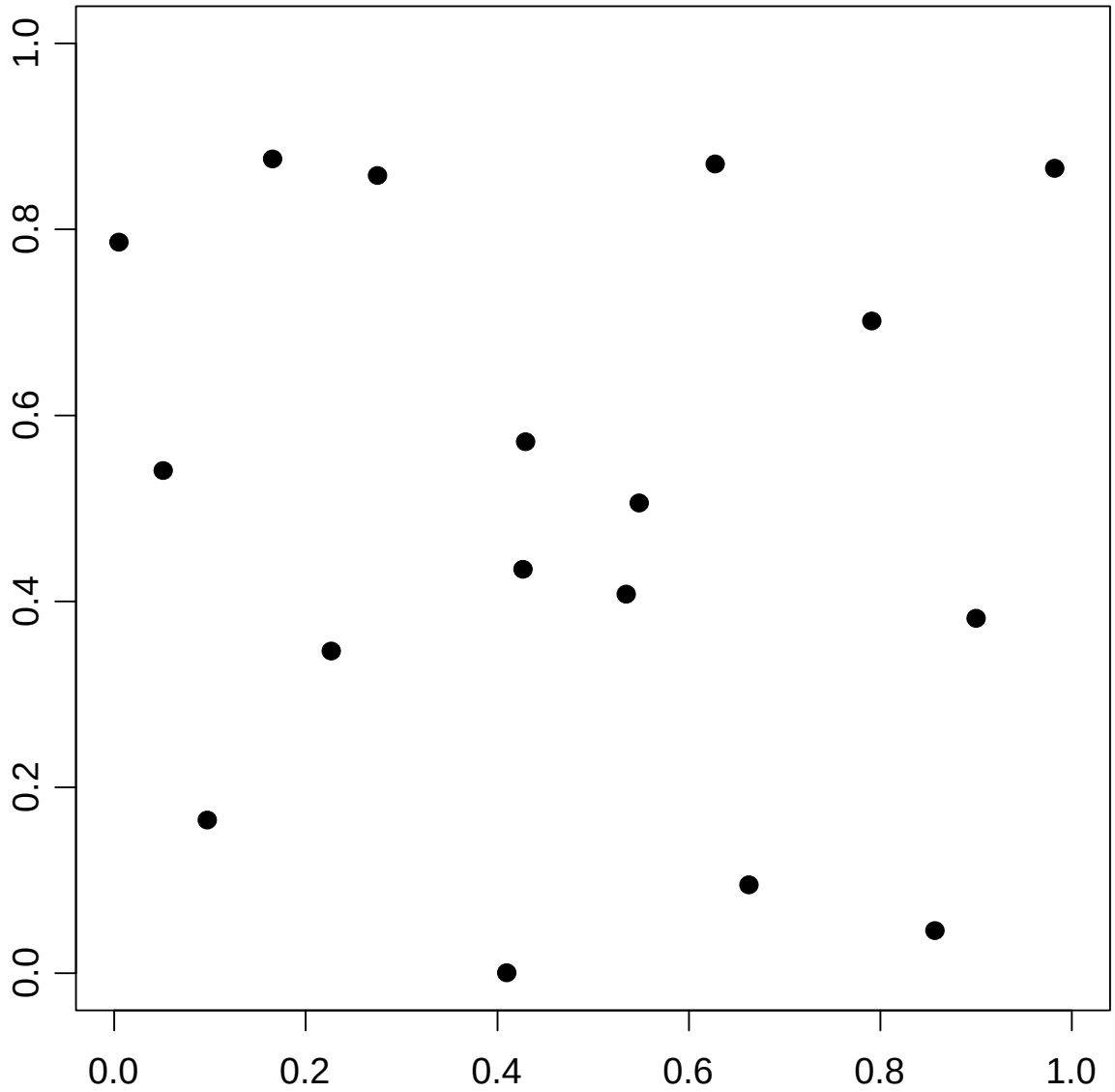

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

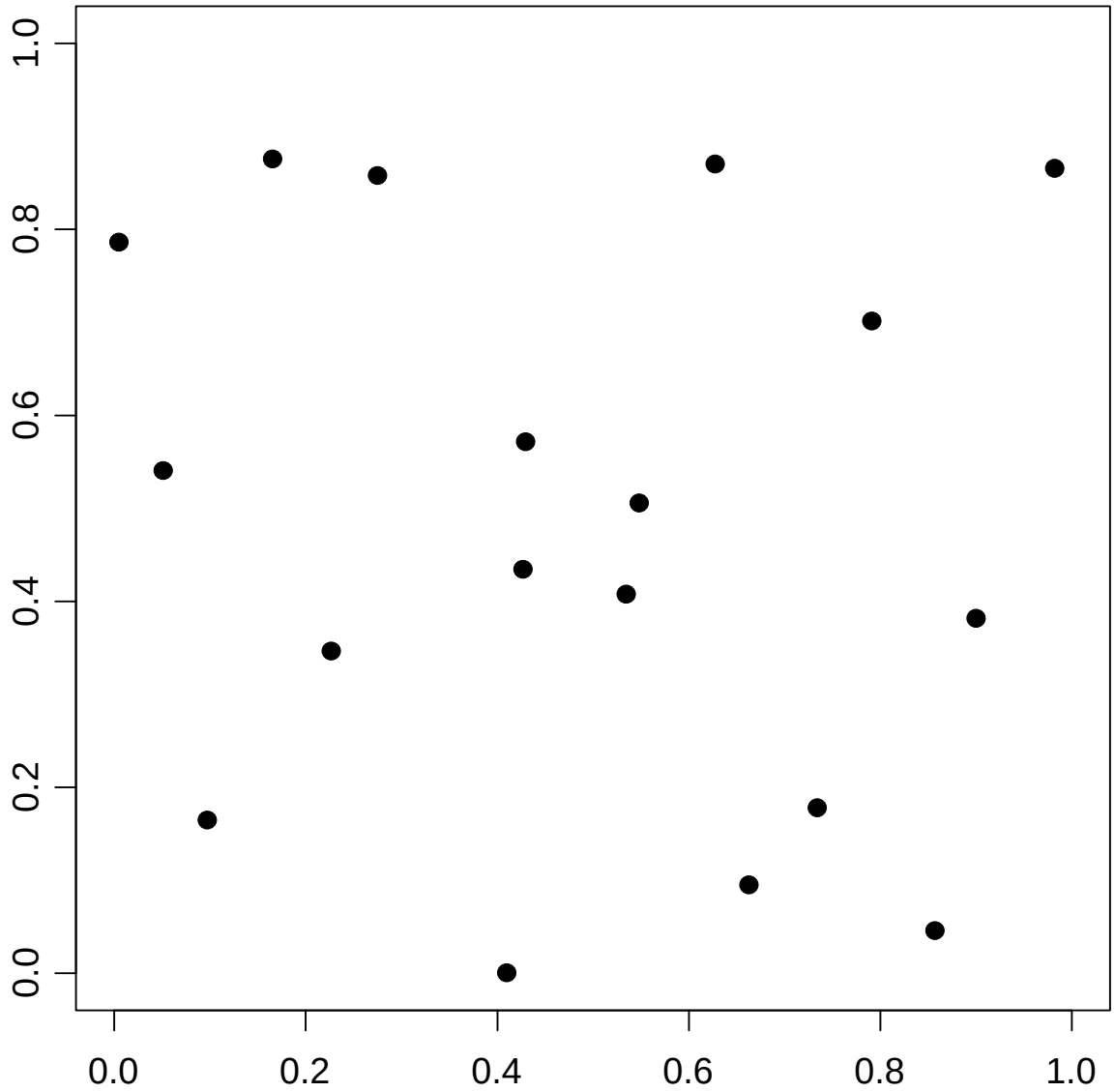

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

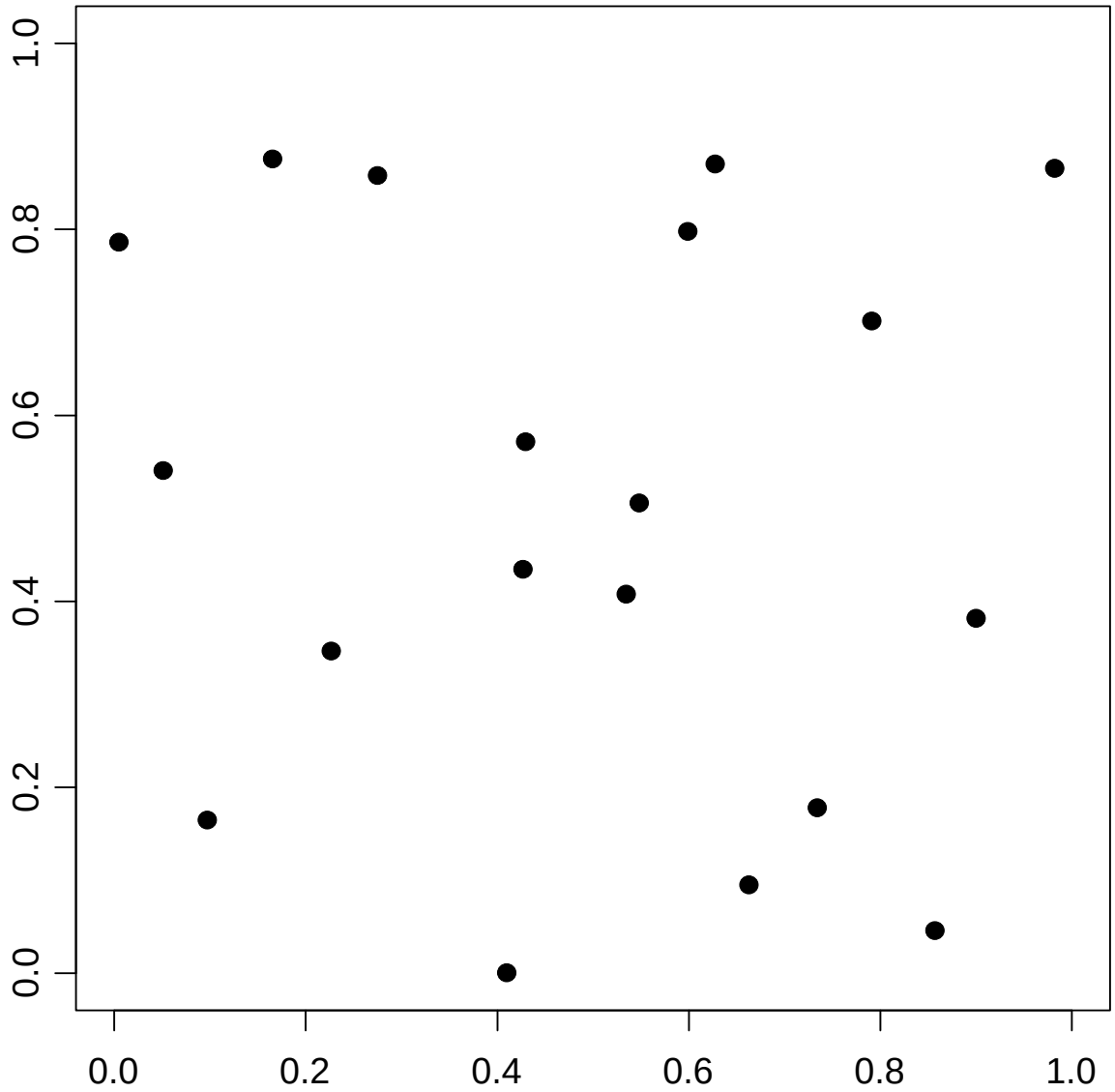

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

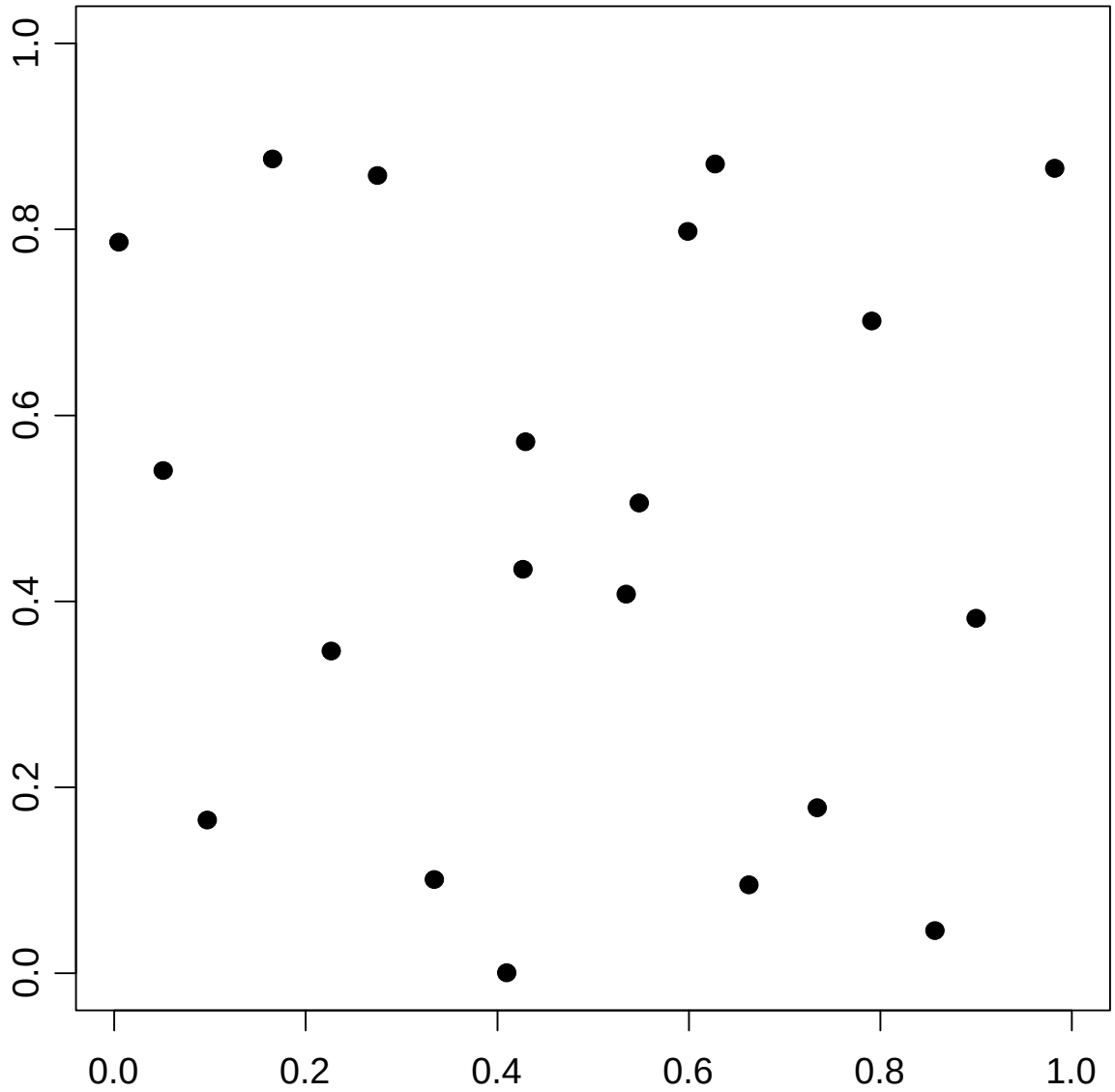

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

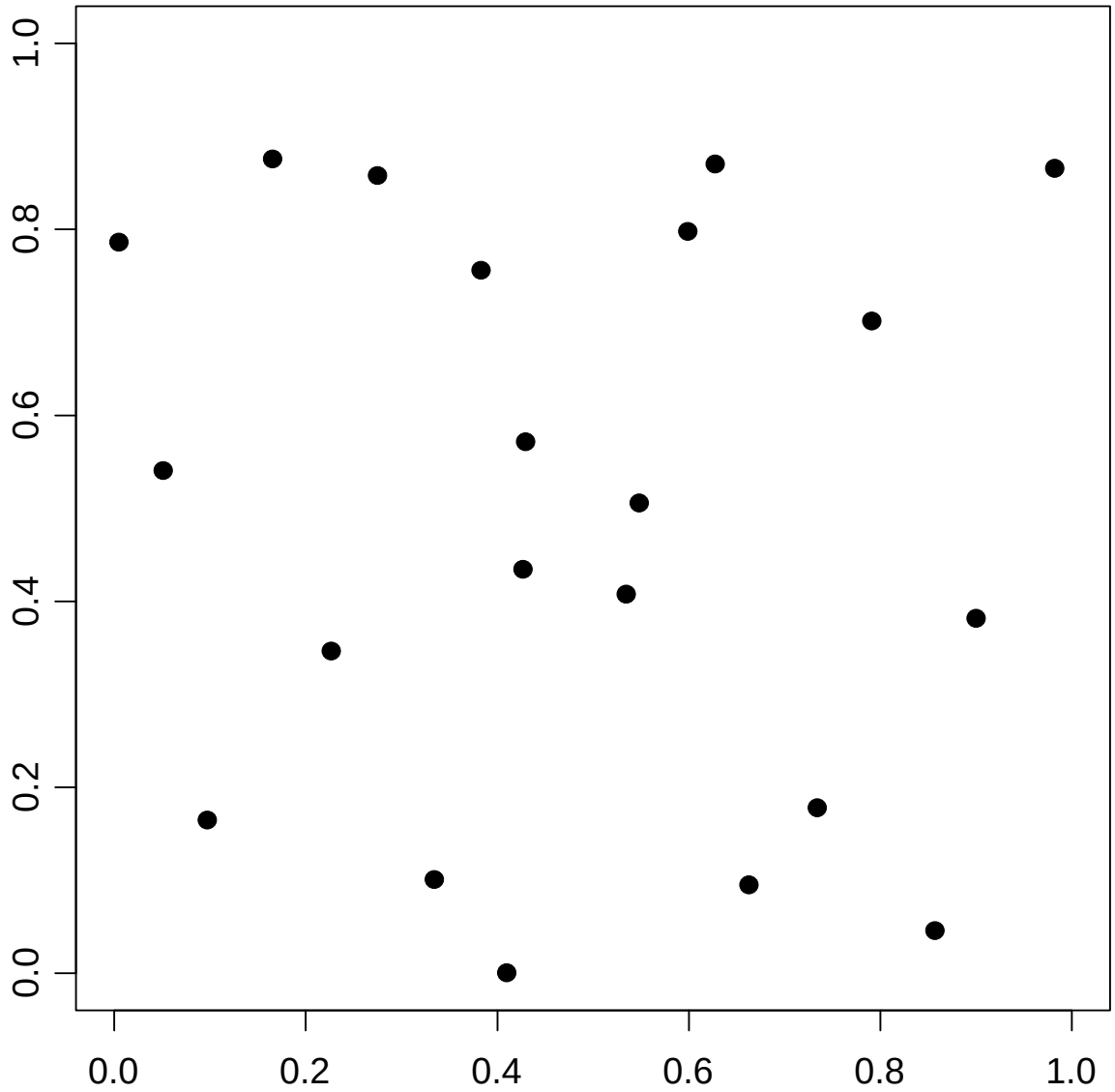

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

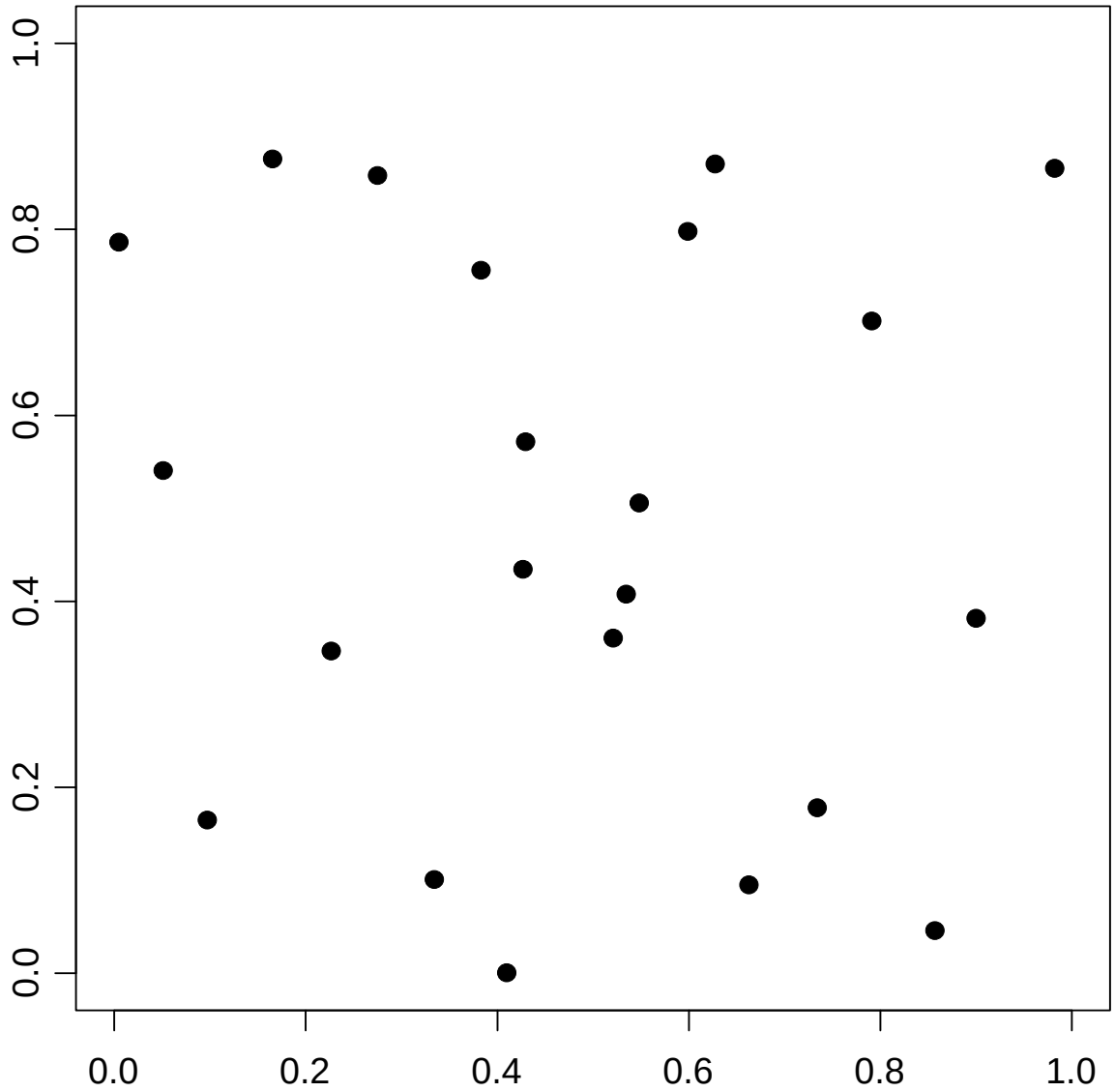

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

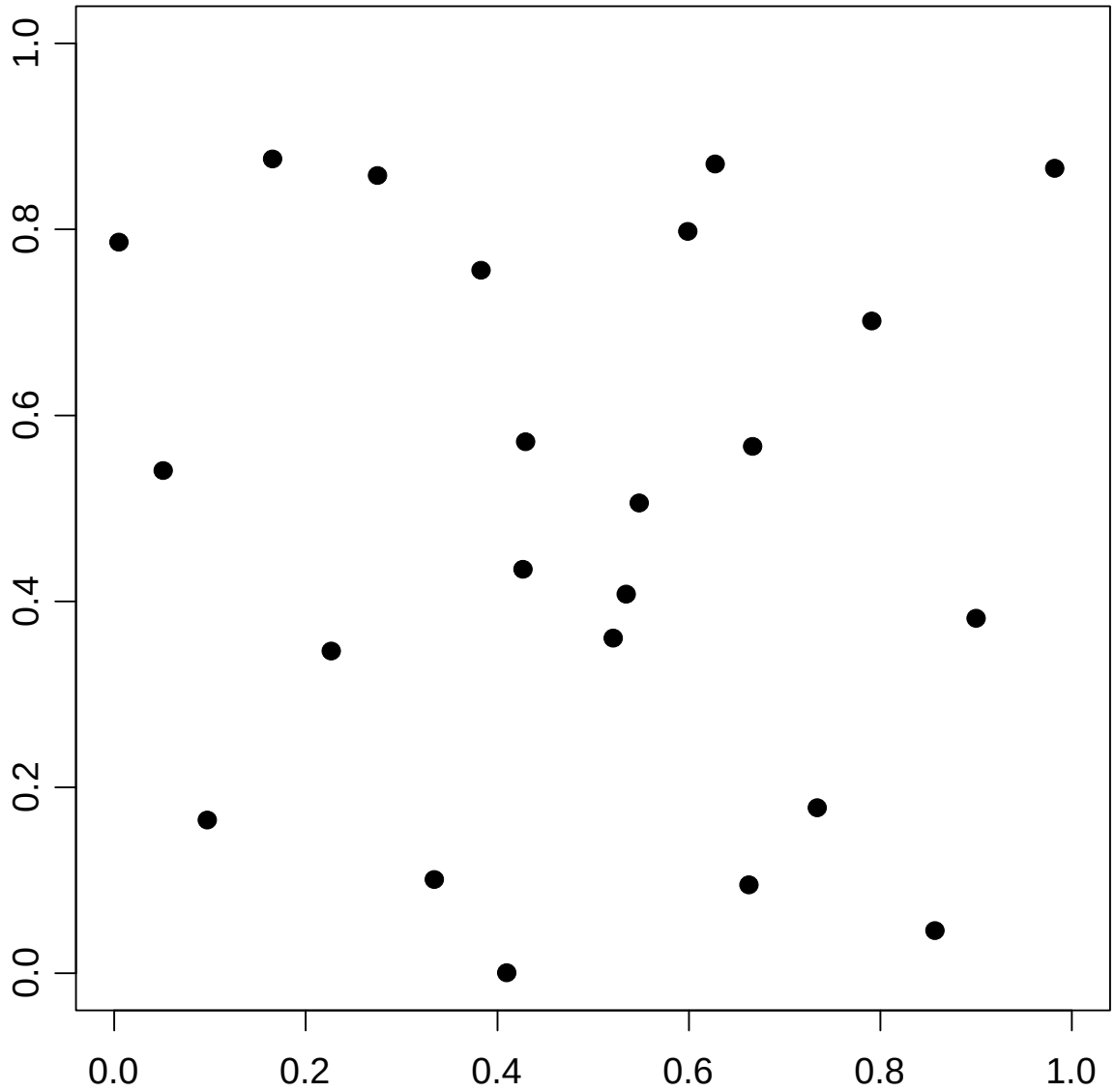

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

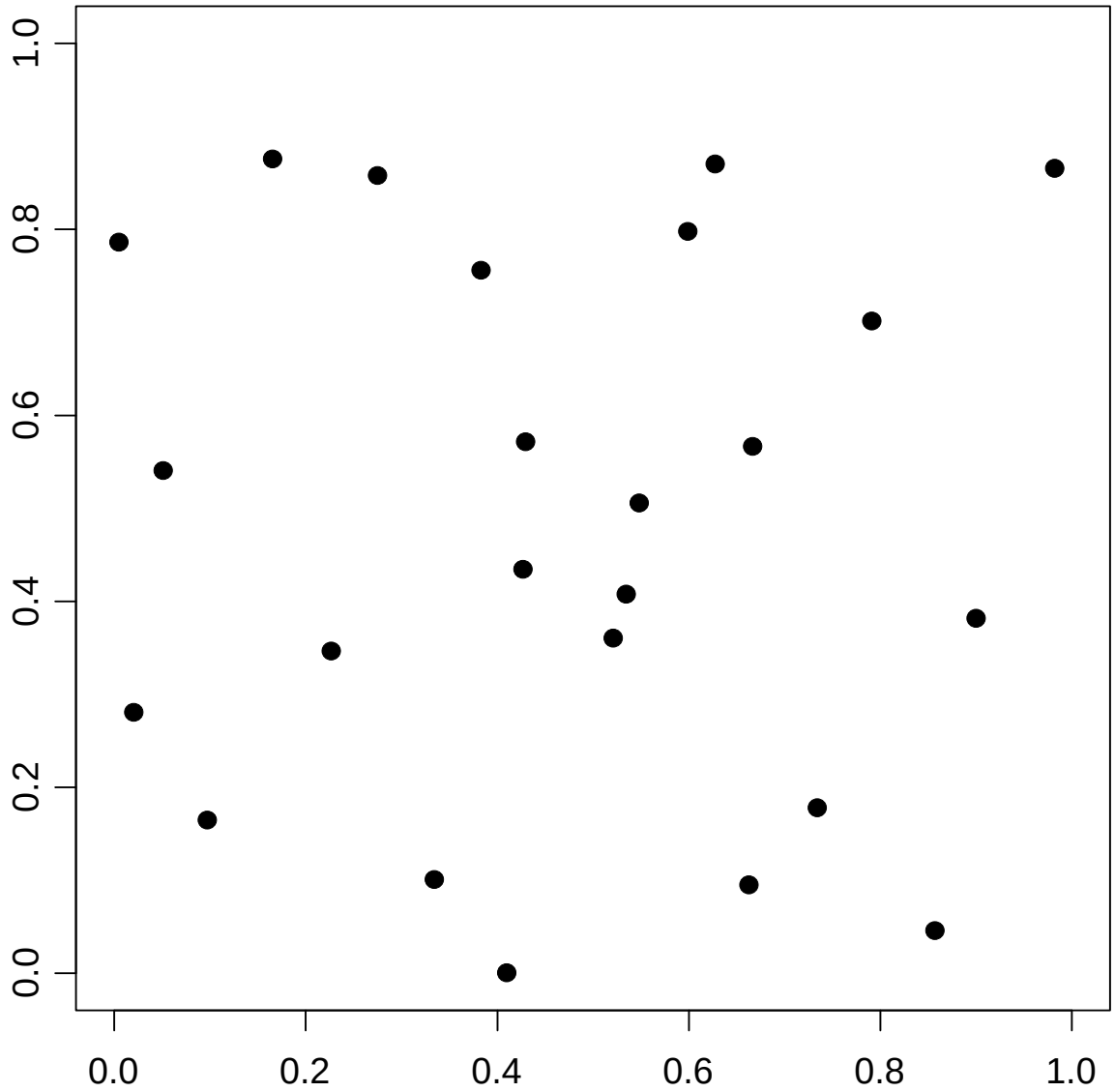

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

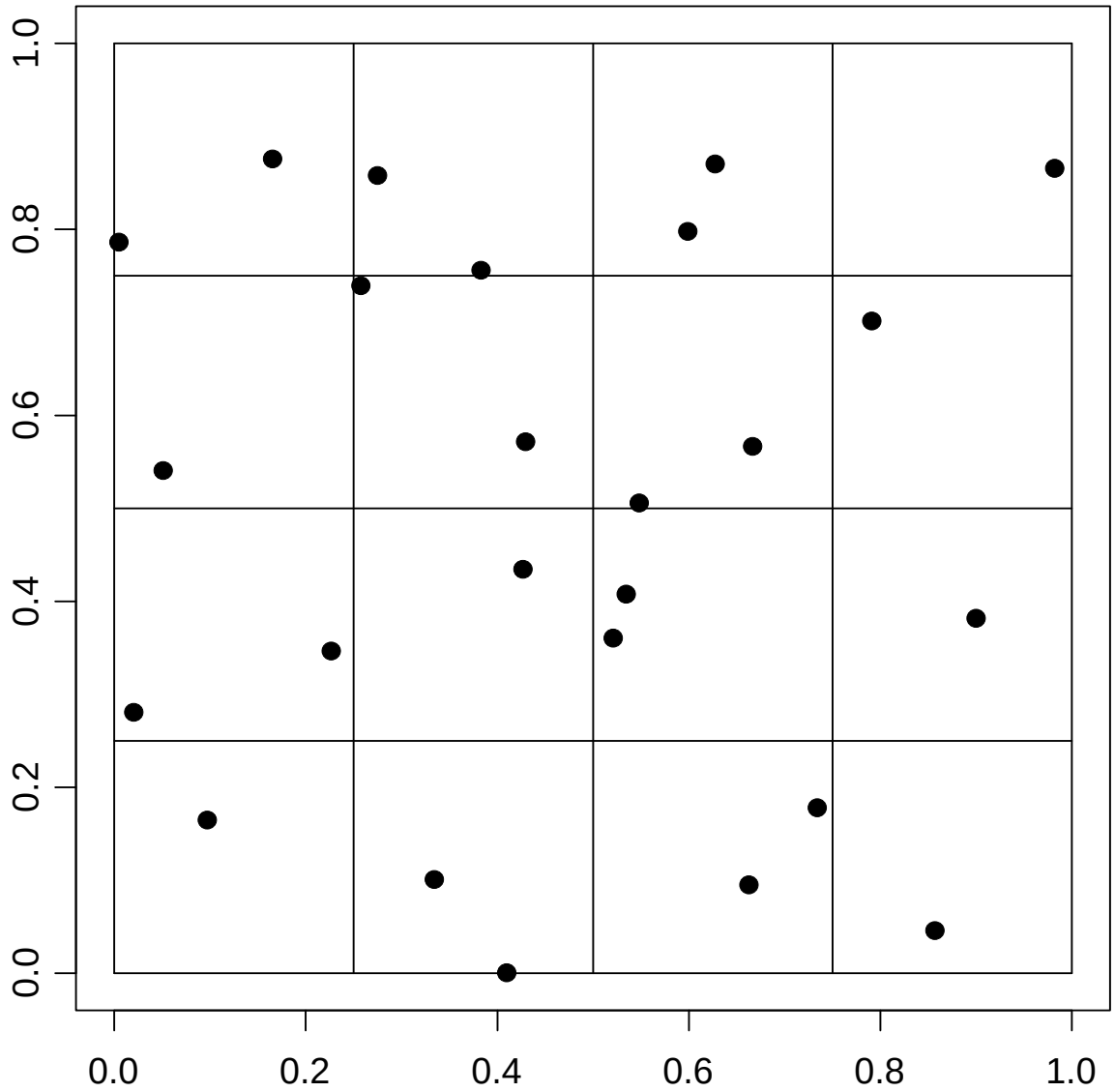

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

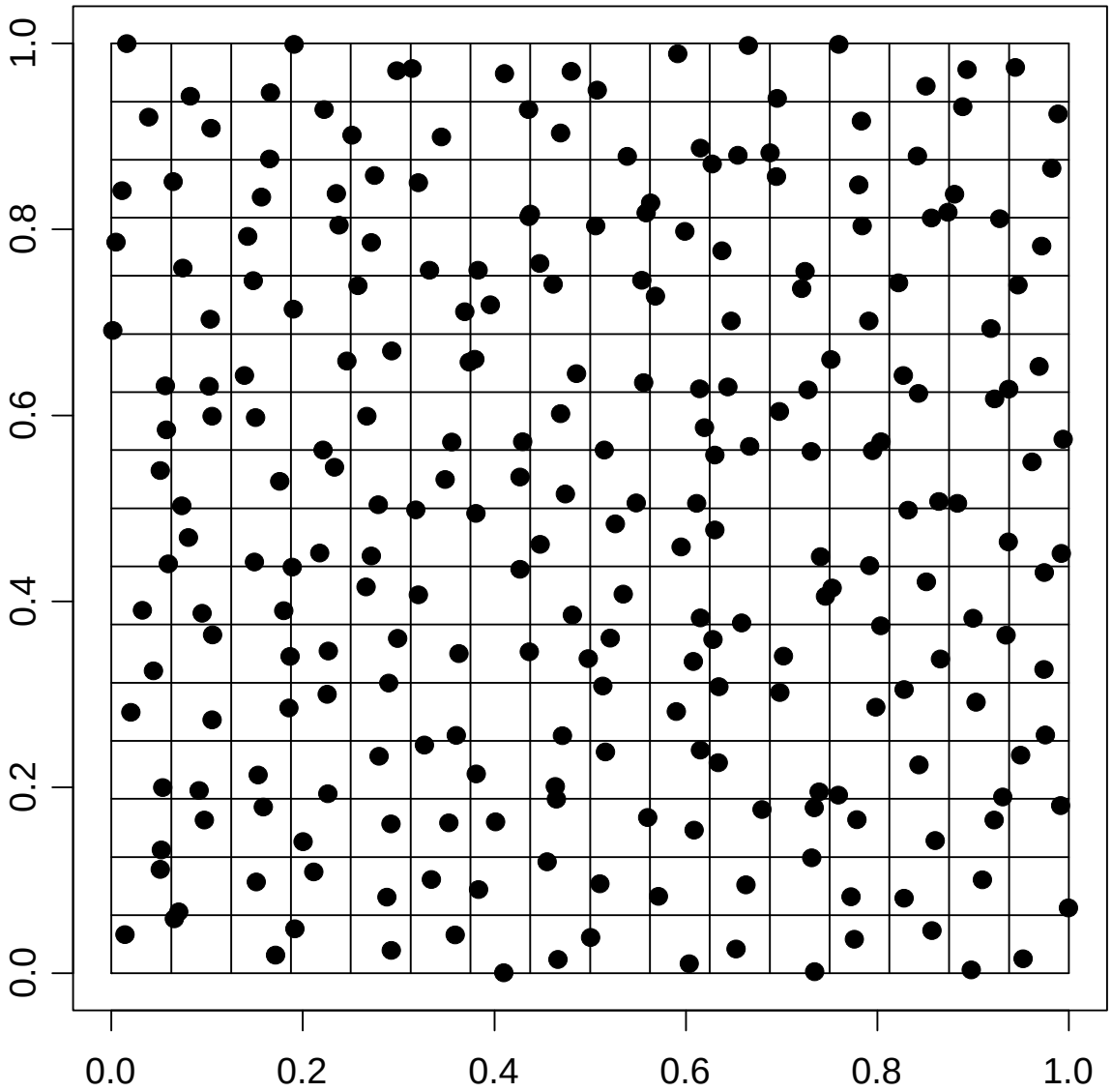

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

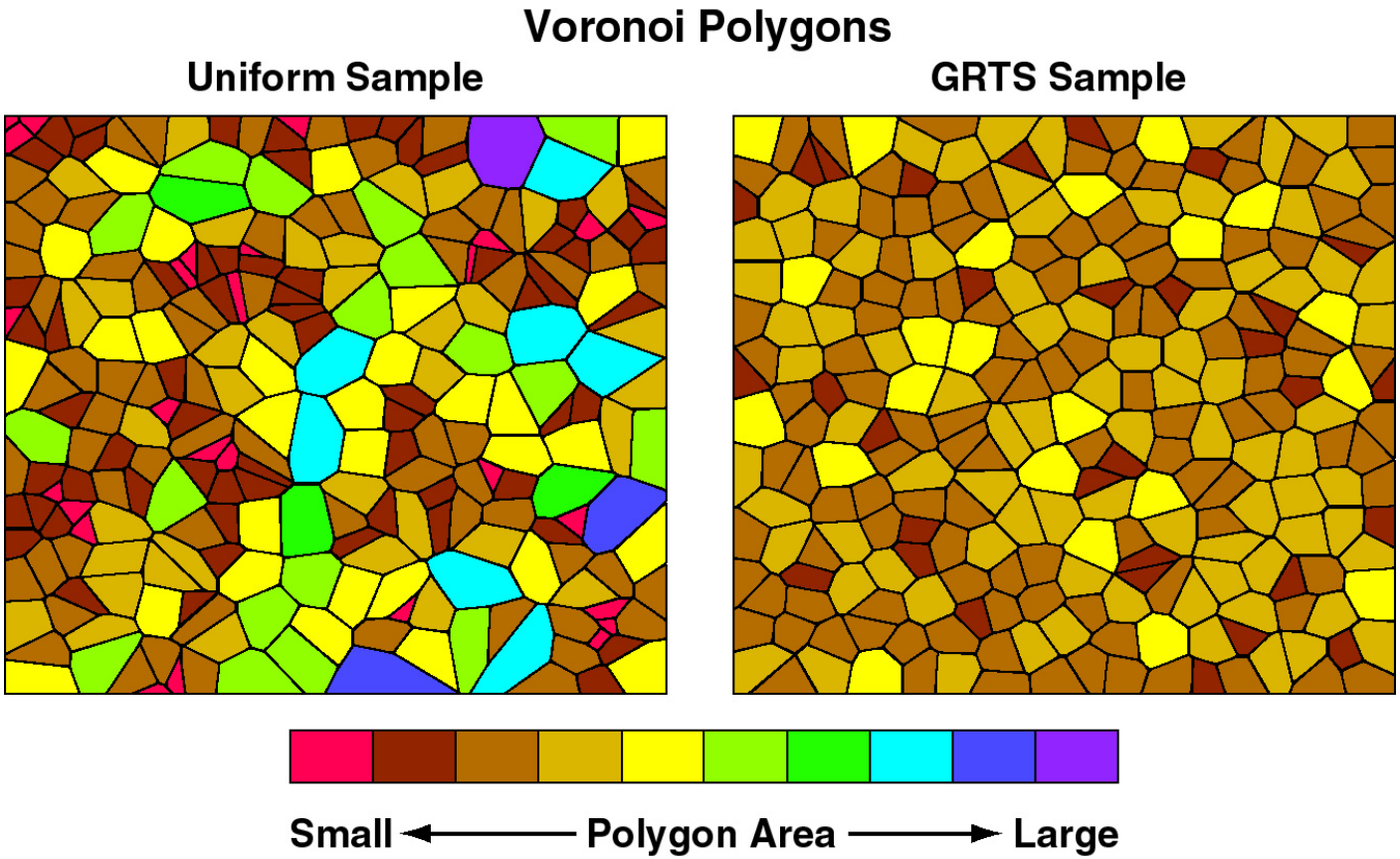

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

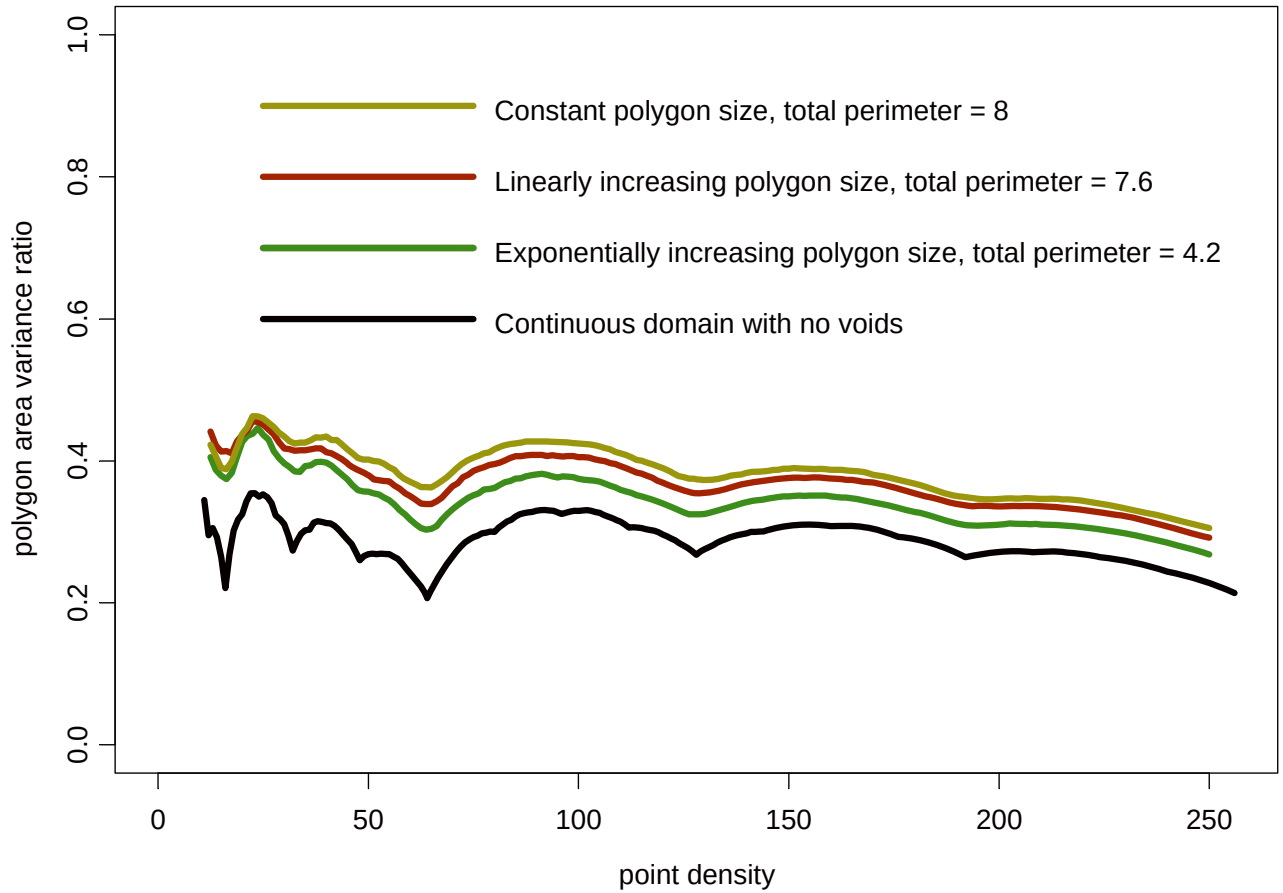

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

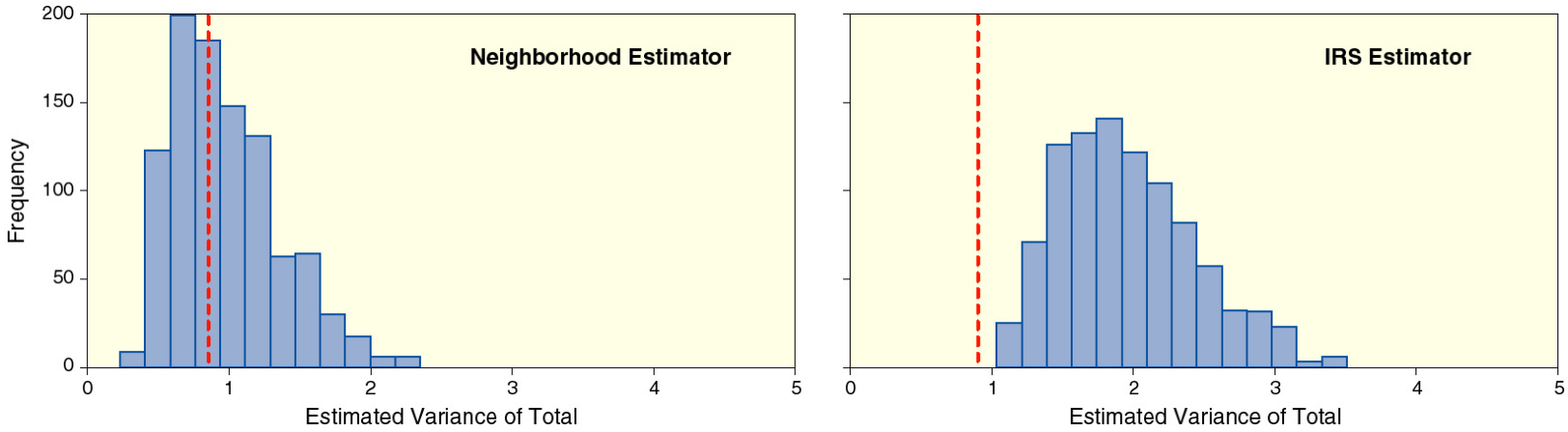
Spatial Balance: 256 points



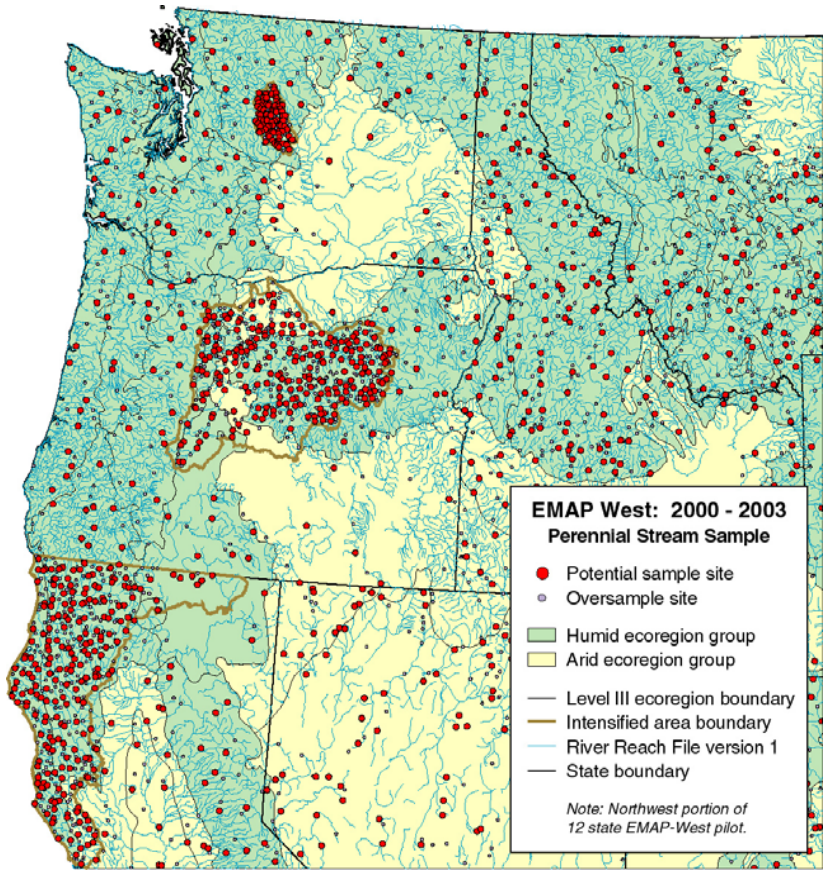
Ratio of GRTS to IRS Voronoi polygon size variance

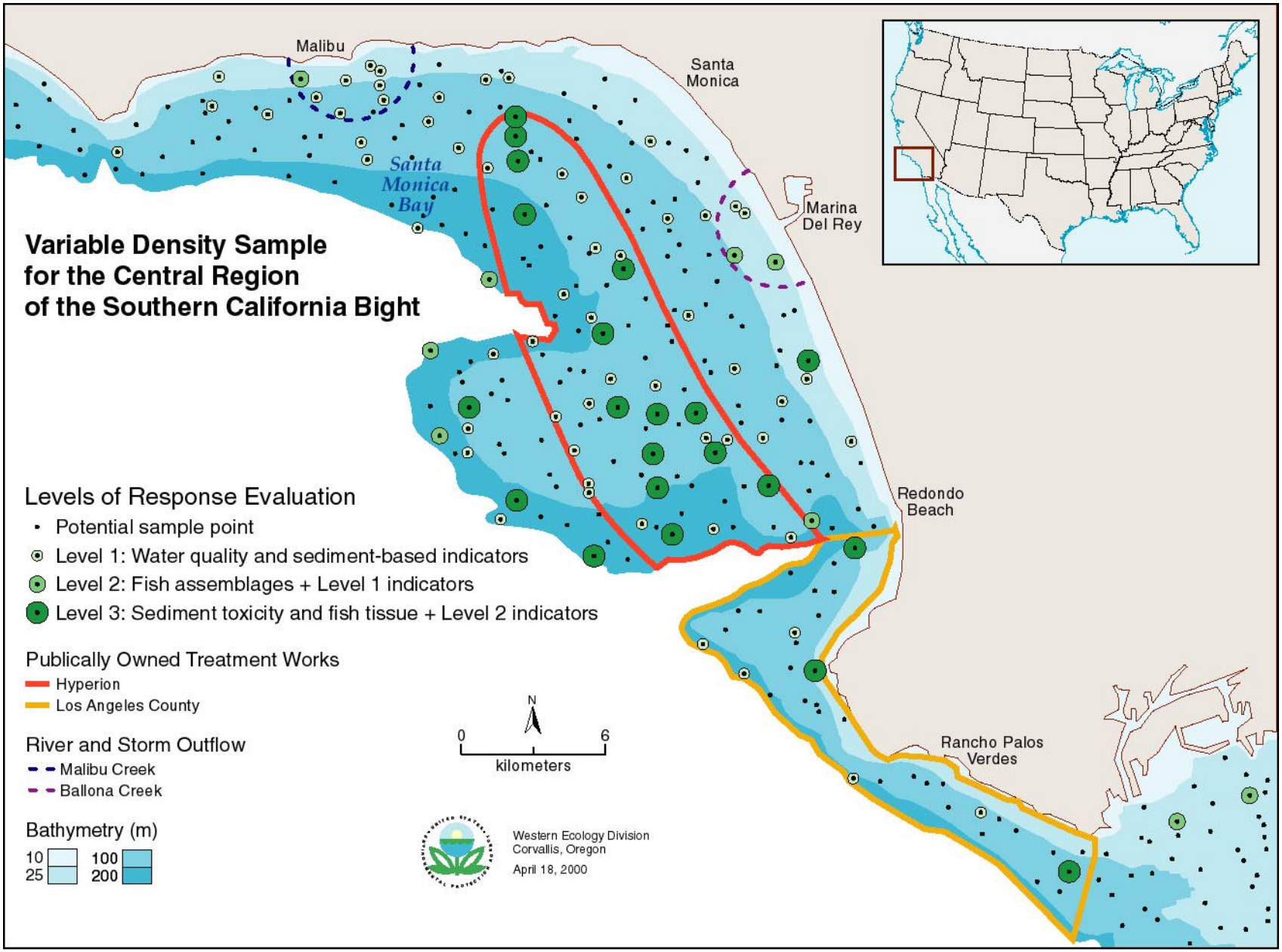


Impact on Variance Estimators of Totals



Perennial Streams GRTS sample



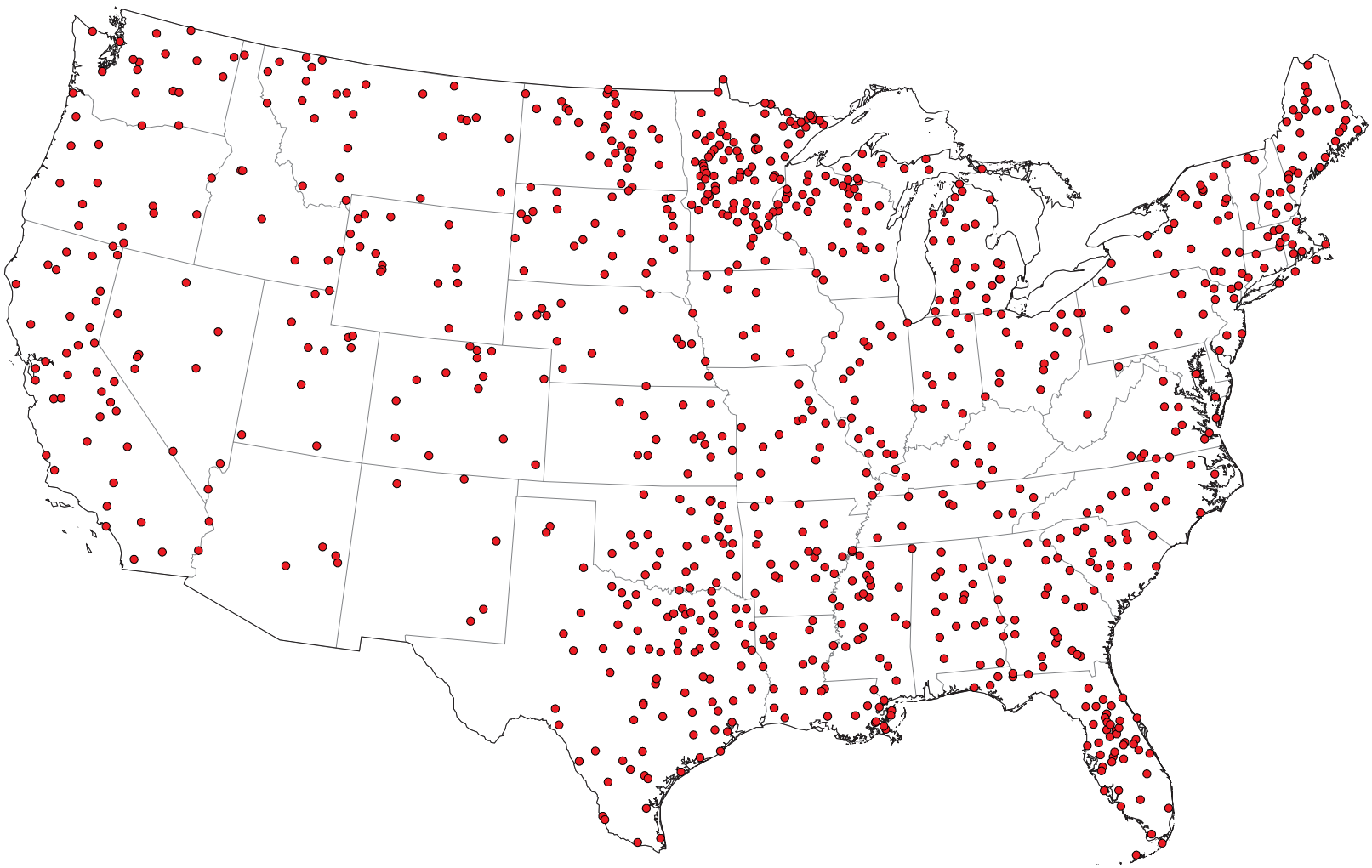


projects/emapgis/urquhart/california bight/figure12-4.ai
4/18/00 smp

RF3 Sample Frame: Lakes

Lake Area (ha)	Number of Lakes	Percent	Cumulative Number of Lakes	Cumulative Percent
1–5	172,747	63.8	172,747	63.8
5–10	44,996	16.6	217,743	80.4
10–50	40,016	14.8	257,759	95.2
50–500	11,228	4.1	268,987	99.3
500–5000	1,500	0.6	270,387	99.9
>5000	274	0.1	270,761	100.0

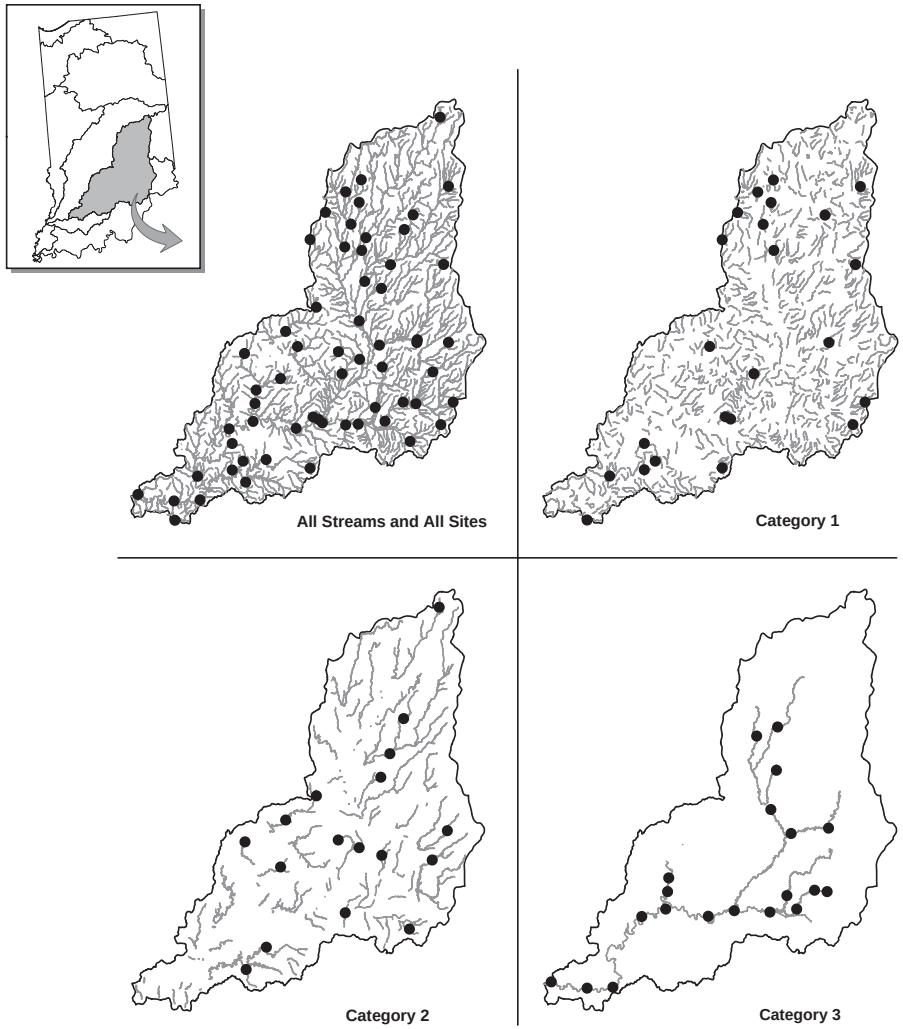
National Fish Tissue Contaminant Lake Survey



Sample Selected: Lakes

Lake Area (ha)	1999	2000	2001	2002	All Years	Expected Weight
1-5	39	41	47	47	174	938.84
5-10	44	40	47	46	177	261.61
10-50	32	47	46	25	150	256.51
50-500	34	37	29	34	134	85.06
500-5000	36	30	31	41	138	11.36
>5000	40	30	25	32	127	2.21
Total	225	225	225	225	900	

GRTS Sample of Streams



Software Implementation

- Create GIS shapefile of sampling frame
  - Must be a point, line, or polygon shapefile (one type only)
  - Attribute file must contain information used to define strata or unequal probability categories
- R statistical software program
  - R is free
  - Load psurvey.design library
  - Specify survey design
  - Select sites using GRTS
  - Output is point shapefile with all required design information