MAKING SENSE OF THE EPIDEMIOLOGICAL LITERATURE: PROBLEMS WITH EPIDEMIOLOGICAL STUDIES OF ECOLOGICAL DESIGN

D. Wayne Berman, Ph.D.
Aeolus, Inc.
Albany, CA

And

Louis Anthony Cox, Ph.D.
Cox Associates
Denver, CO

Acknowledgments/Declarations

- I would like to thank NSSGA for funding the original research.

The Focus

- A Class of Epidemological Studies of ecological design that use regression to evaluate associations
- The Problem:
  - In such studies, proximity to a source or sources of a particular toxin is used as a surrogate for direct estimates of exposure
  - Thus, such studies involve evaluation of ***spatial*** distributions
  - Regression analyses are ***non-spatial*** statistical procedures.
- Approximately 100 studies of similar design addressing a variety of toxins and disease end points have been published and continue to appear

Talk Outline

- Illustrate the problem with using regression in epidemiological studies of ecological design
- Summarize the characteristics of real population distributions
- Summarize findings and conclusions from the Berman, Cox, and Popken papers
- Define some useful criteria for detecting these problems
- Identify some useful references describing how to conduct these types of analyses properly

Pan, XL; Day, HW; Wang, W; Beckett, LA; Schenker, MB. Residential proximity to naturally occurring asbestos and mesothelioma risk in California. *Am J Respir Crit Care Med*. 15;172(8):1019-25. 2005.

Pan, XL; Day, HW; Wang, W; Beckett, LA; Schenker, MB. Residential proximity to naturally occurring asbestos and mesothelioma risk in California. *Am J Respir Crit Care Med*. 15;172(8):1019-25. 2005.

- Mesothelioma risk is positively associated with proximity to ultramafic rocks in California
- Risk decreases 6% with each 10 km increase in distance

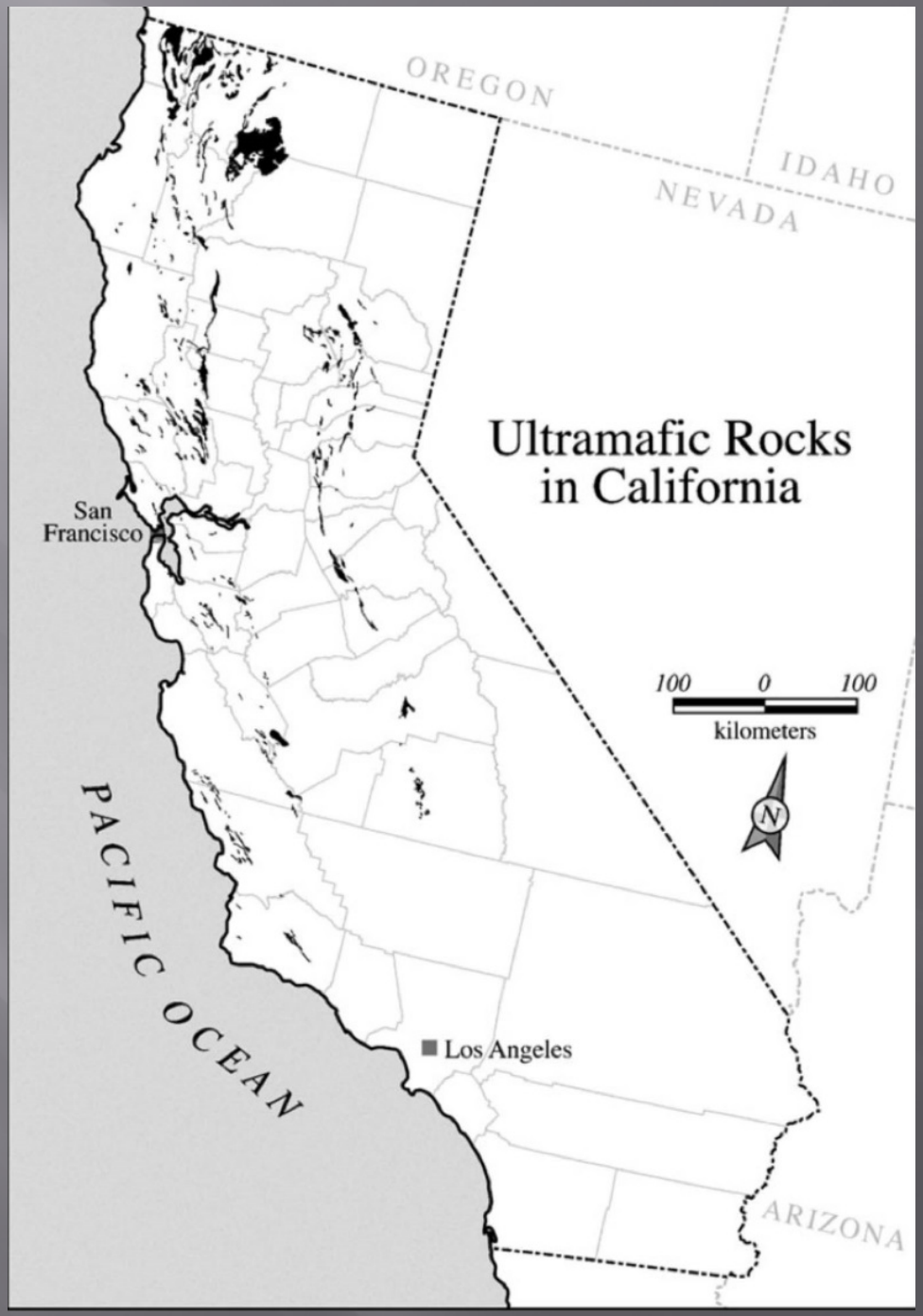
Pan, XL; Day, HW; Wang, W; Beckett, LA; Schenker, MB. Residential proximity to naturally occurring asbestos and mesothelioma risk in California. *Am J Respir Crit Care Med*. 15;172(8):1019-25. 2005.

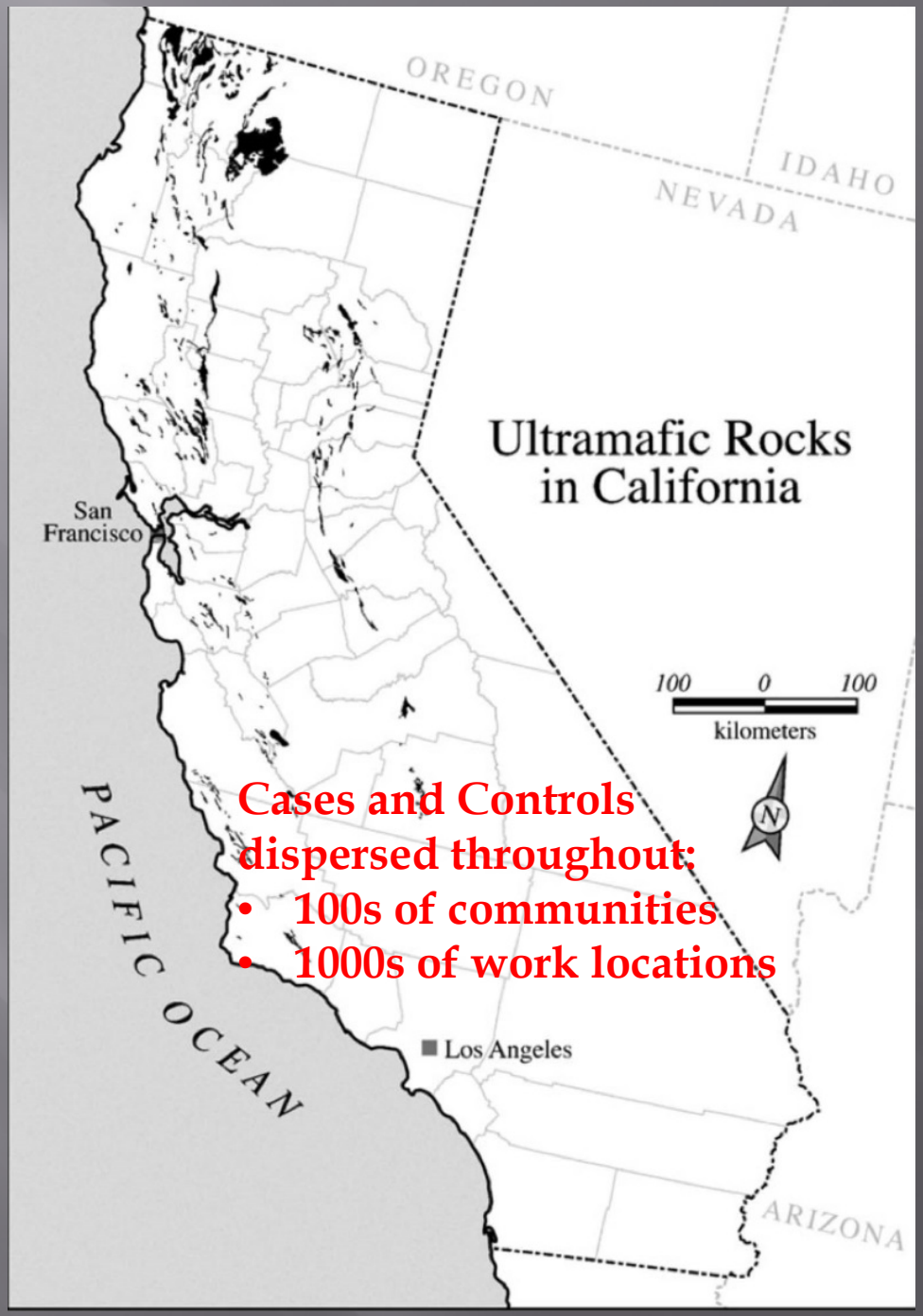
- Mesothelioma risk is positively associated with proximity to ultramafic rocks in California
- Risk decreases 6% with each 10 km increase in distance
- Unfortunately, the associations observed in this study have nothing to do with causality…..

Pan, XL; Day, HW; Wang, W; Beckett, LA; Schenker, MB. Residential proximity to naturally occurring asbestos and mesothelioma risk in California. *Am J Respir Crit Care Med.* 15;172(8):1019-25. 2005.

- Mesothelioma risk is positively associated with proximity to ultramafic rocks in California
- Risk decreases 6% with each 10 km increase in distance
- Unfortunately, the associations observed in these studies have nothing to do with causality…..

…….by the end of this talk, I hope this will be intuitively obvious


OREGON
IDAHO
NEVADA
Ultramafic Rocks
in California
San
Francisco
100 0 100
kilometers
N
PACIFIC OCEAN
Los Angeles
ARIZONA

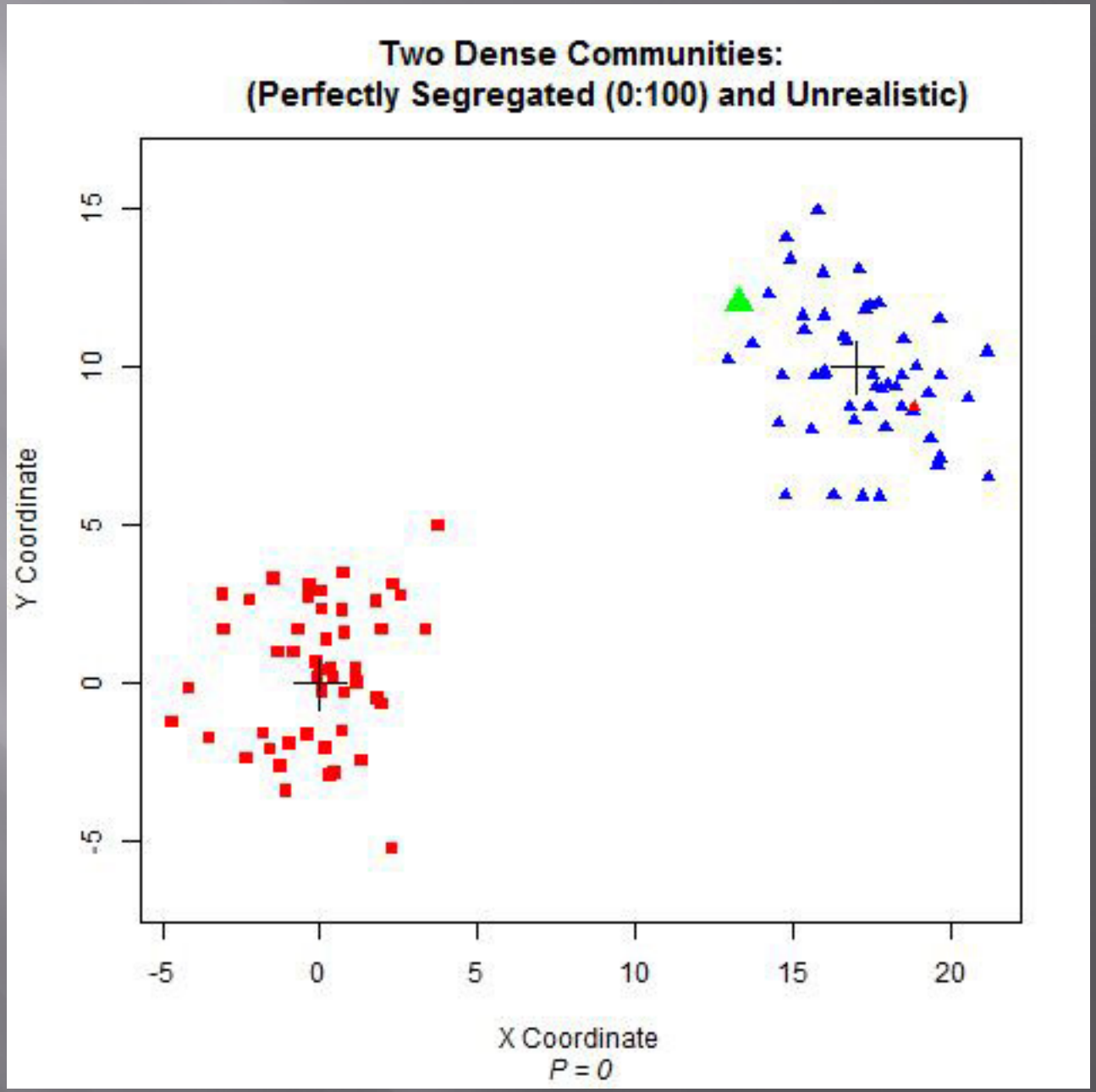

OREGON
IDAHO
NEVADA
Ultramafic Rocks
in California
San
Francisco
100 0 100
kilometers
N
Cases and Controls
dispersed throughout:
• 100s of communities
• 1000s of work locations
PACIFIC OCEAN
Los Angeles
ARIZONA

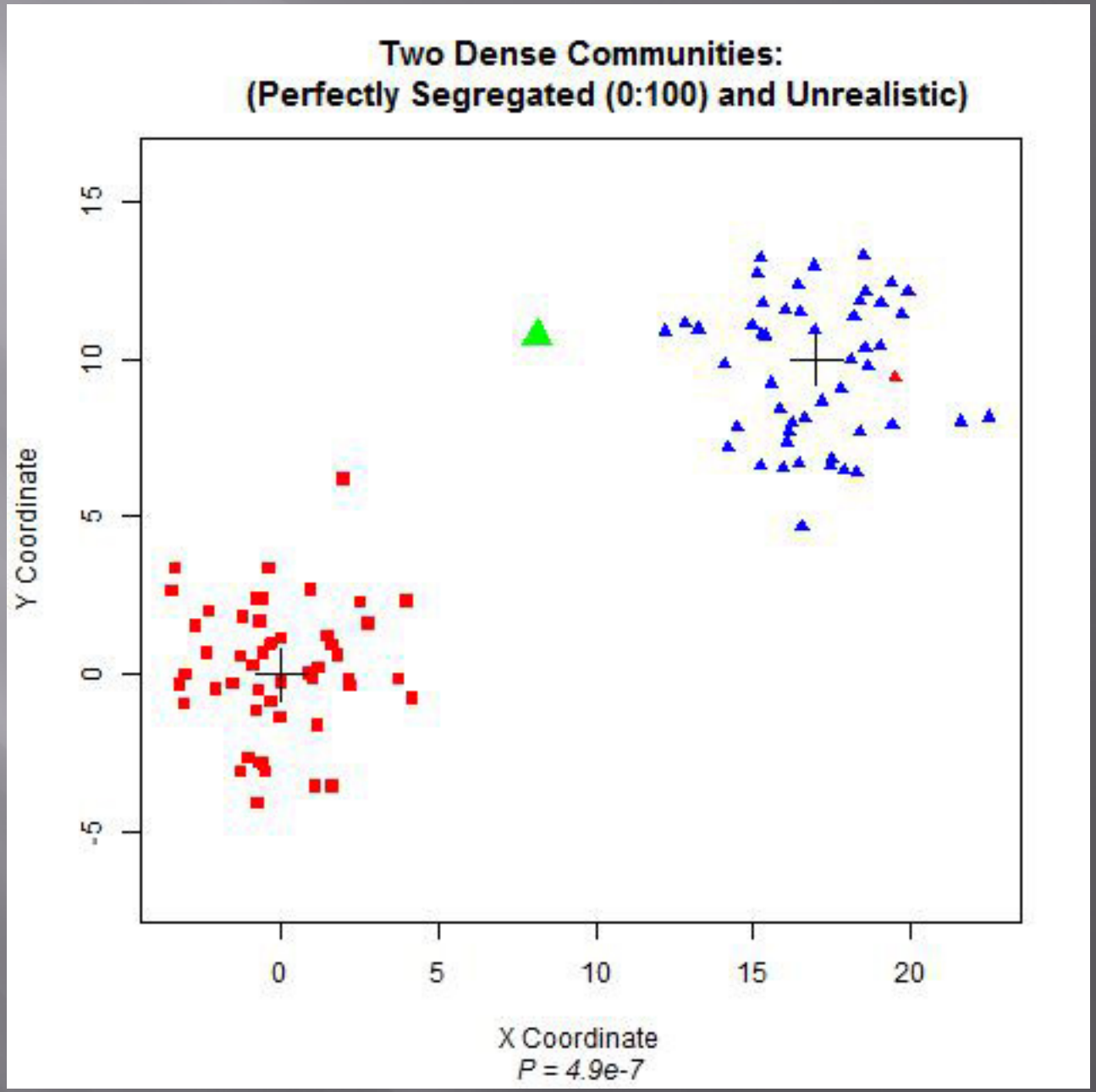
Critical Considerations

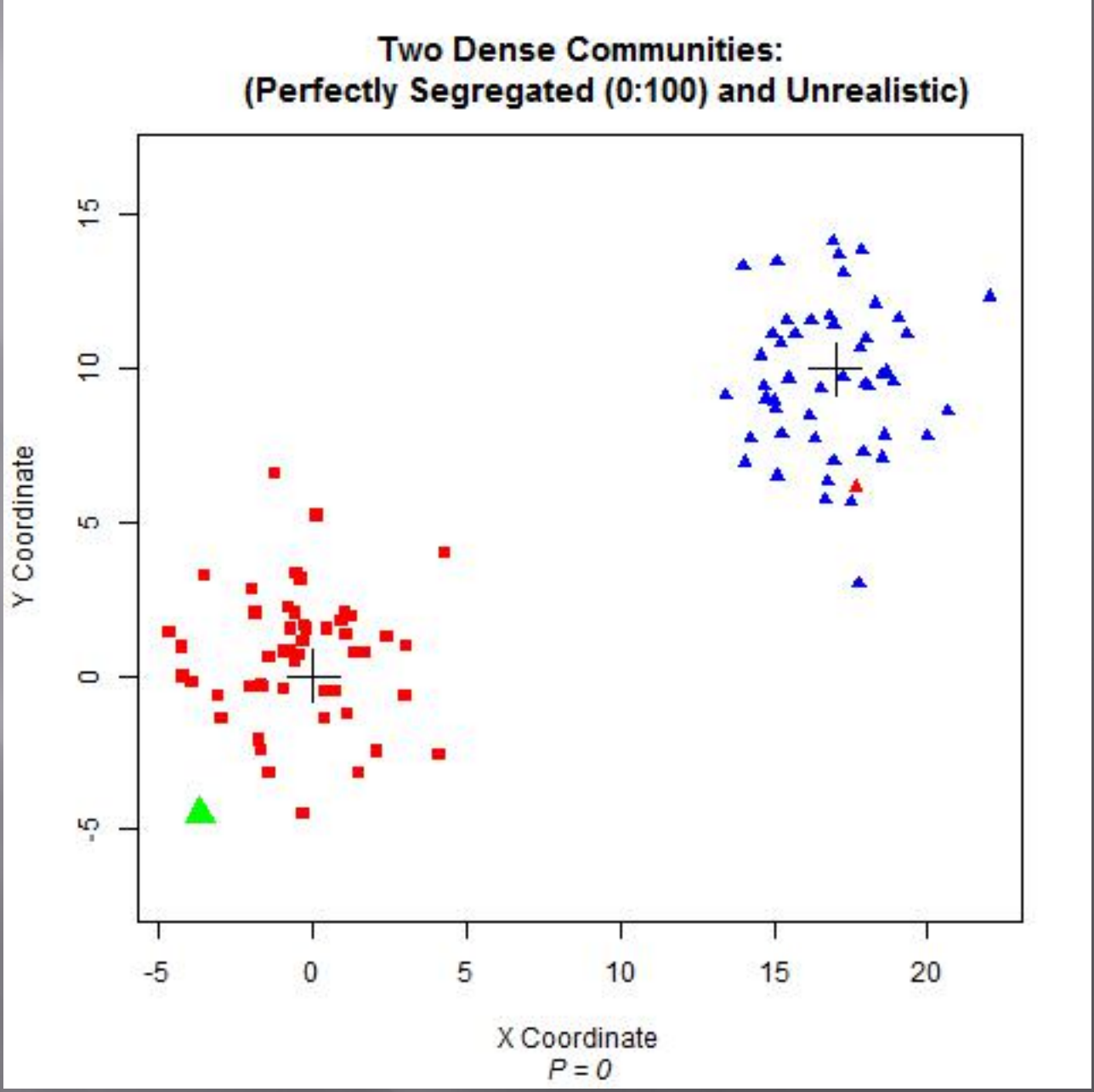
- *By definition,* valid statistical tests show positive results due to chance no more than 5% of the time (meaning of 5% significance).

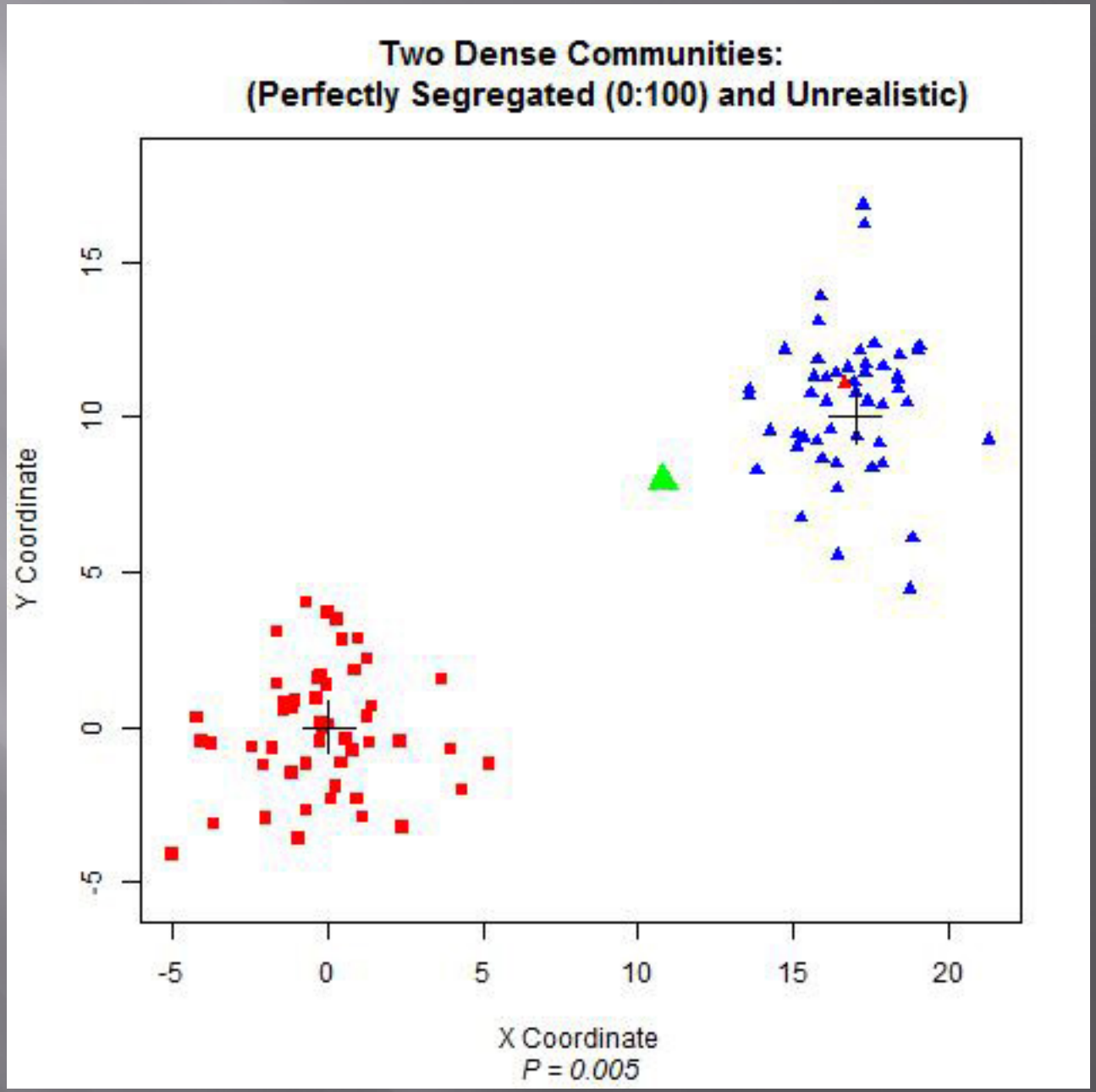
Critical Considerations

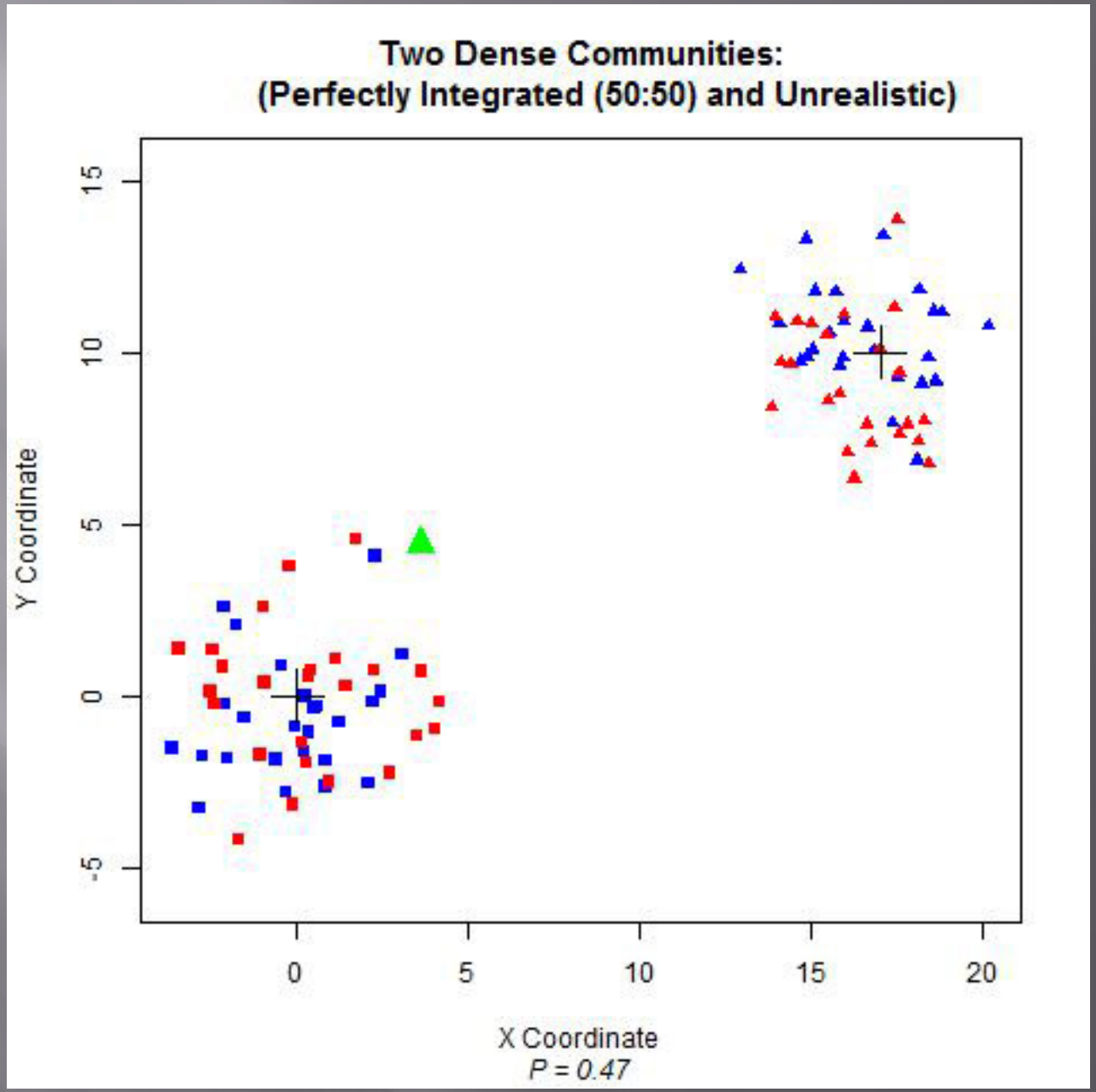
- *By definition,* valid statistical tests show positive results due to chance no more than 5% of the time (meaning of 5% significance).
- Correspondingly, to infer causality, tests for associations cannot detect non-causal (random) associations more than 5% of the time

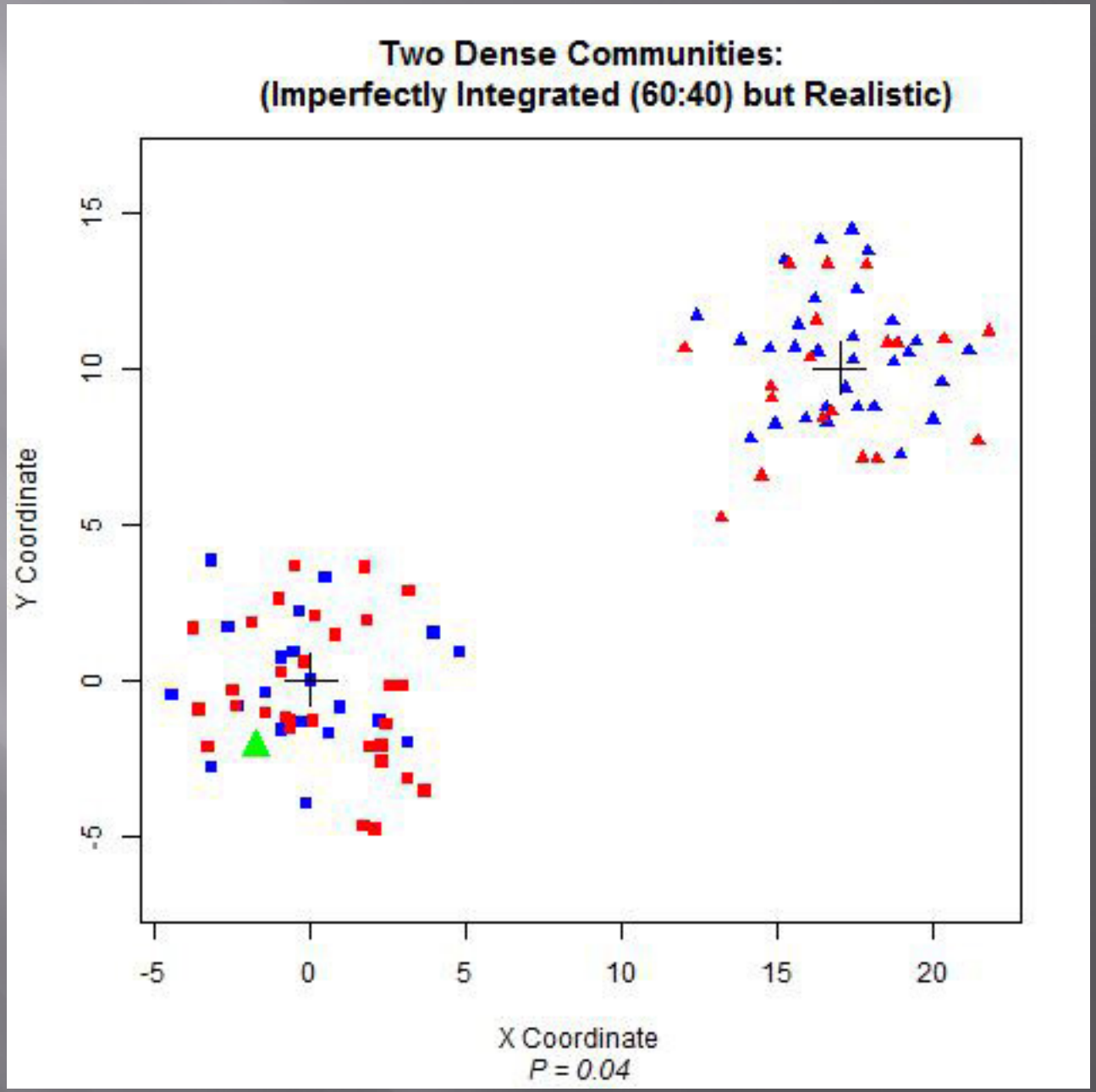

Two Dense Communities:
(Perfectly Segregated (0:100) and Unrealistic)
Y Coordinate
15
10
5
0
-5
-5
0
5
10
15
20
X Coordinate
P = 0

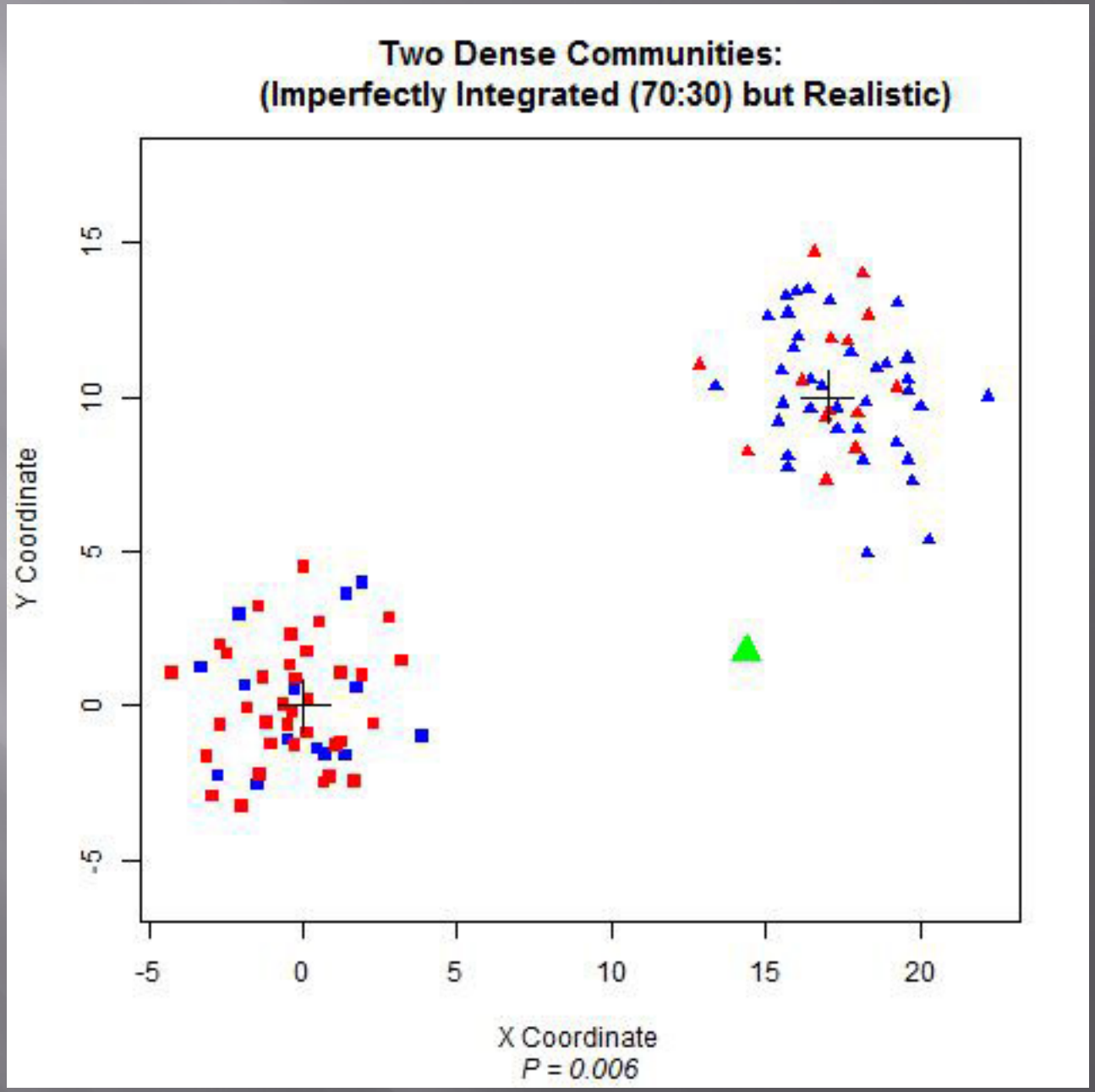

Two Dense Communities:
(Perfectly Segregated (0:100) and Unrealistic)
Y Coordinate
15
10
5
0
-5
0
5
10
15
20
X Coordinate
P = 4.9e-7

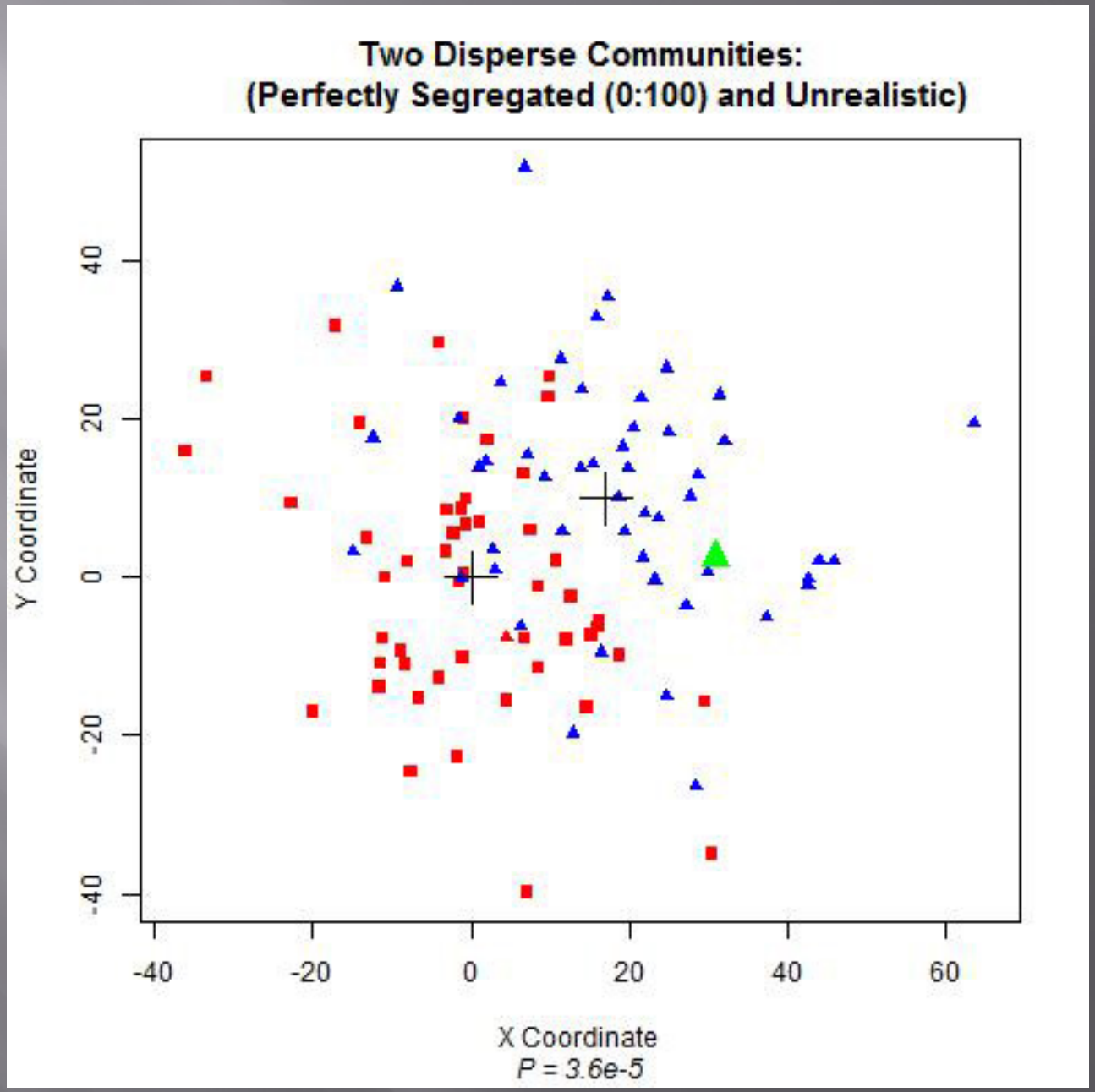

Two Dense Communities:
(Perfectly Segregated (0:100) and Unrealistic)
Y Coordinate
15
10
5
0
-5
-5
0
5
10
15
20
X Coordinate
P = 0

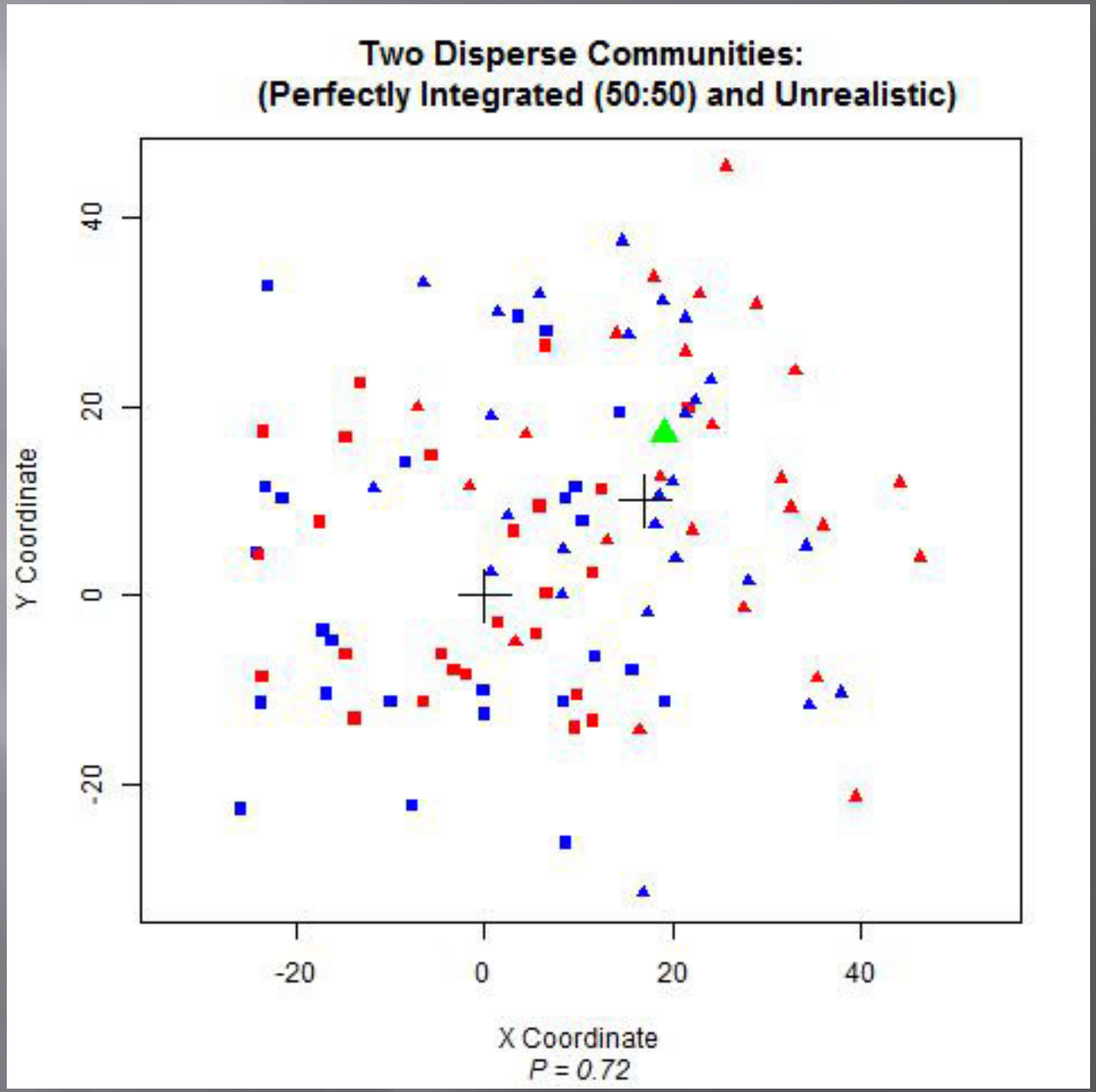

Two Dense Communities:
(Perfectly Segregated (0:100) and Unrealistic)
Y Coordinate
15
10
5
0
-5
-5
0
5
10
15
20
X Coordinate
P = 0.005

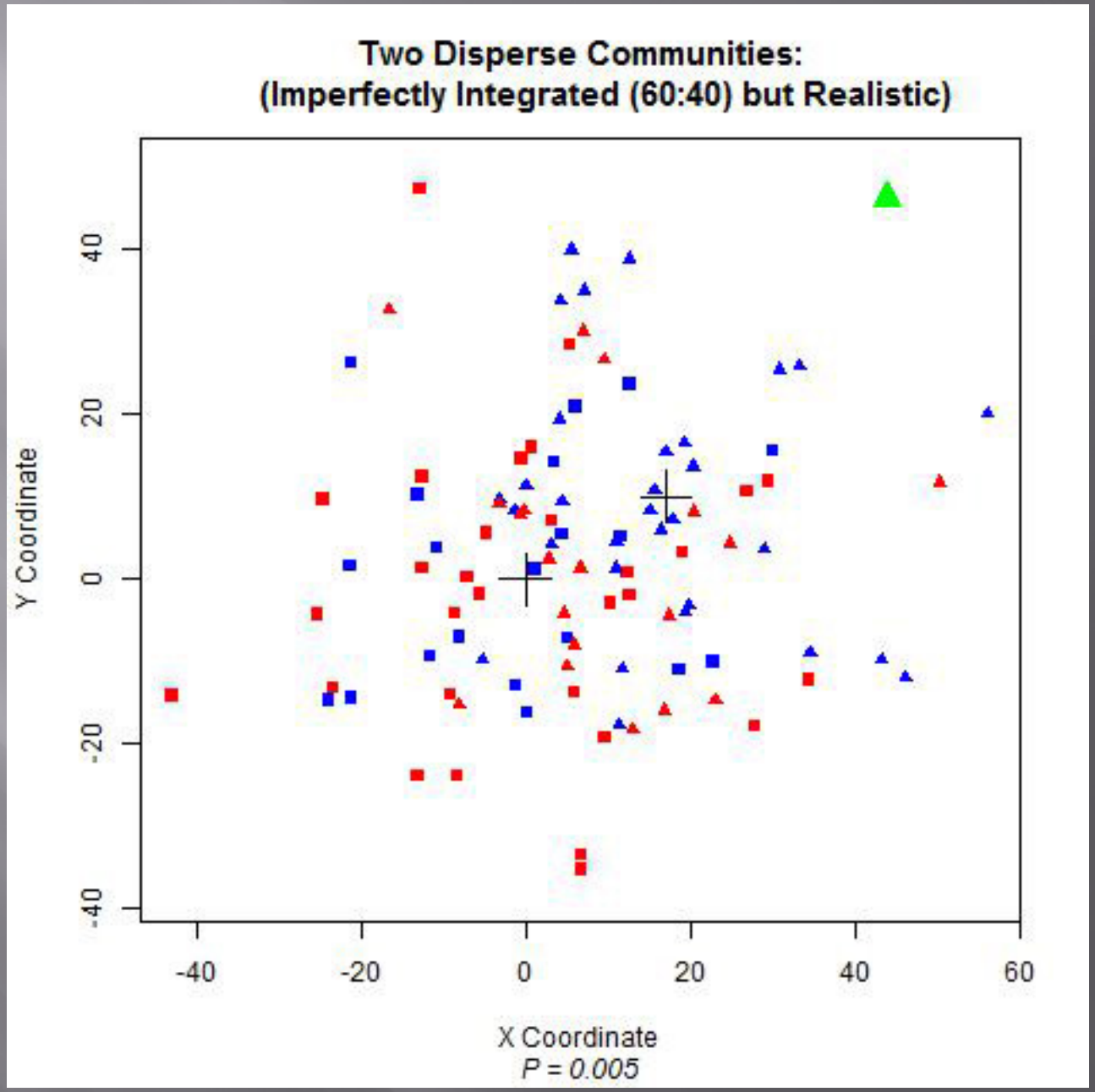

Two Dense Communities:
(Perfectly Integrated (50:50) and Unrealistic)
Y Coordinate
15
10
5
0
-5
0
5
10
15
20
X Coordinate
P = 0.47

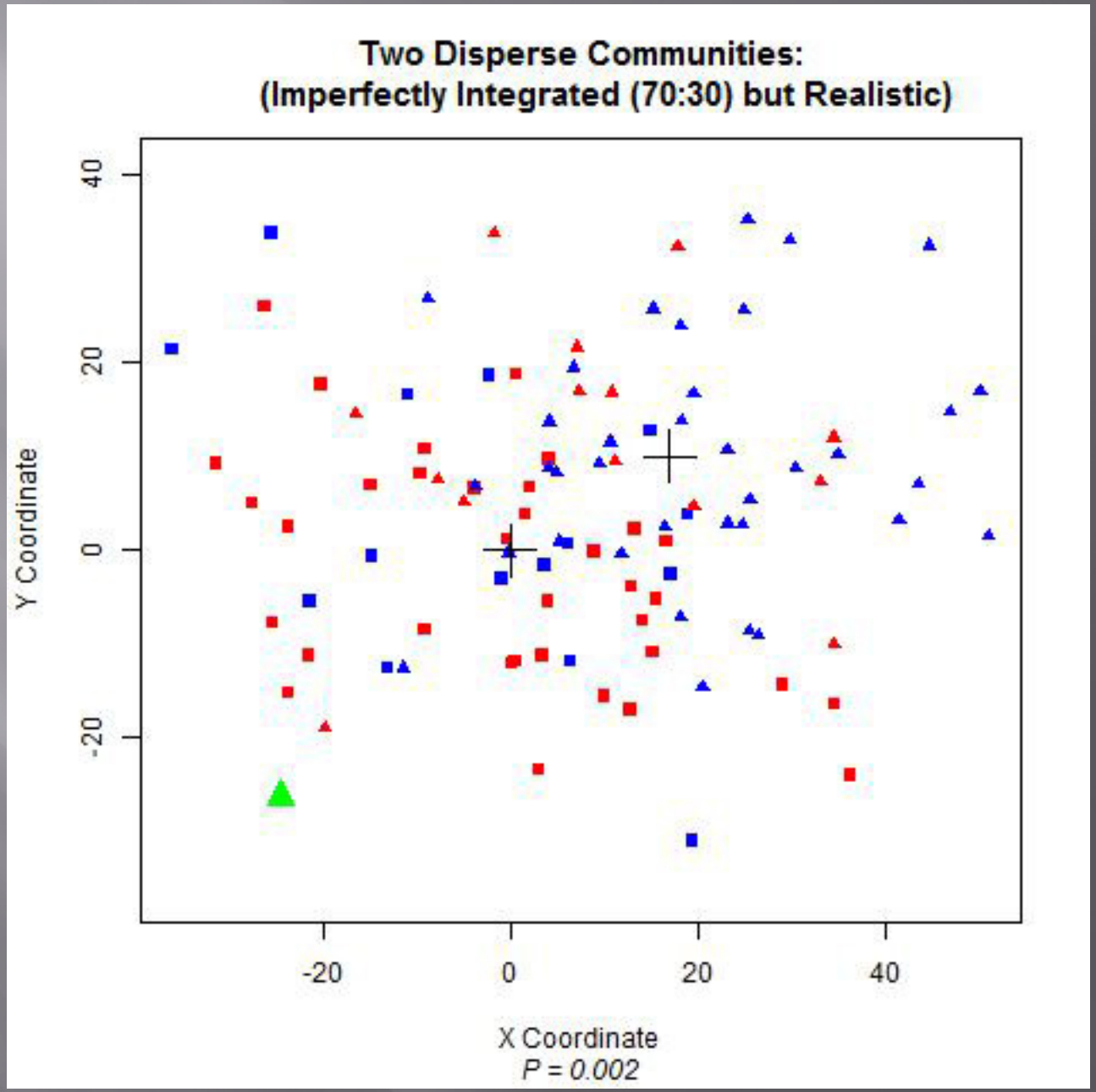

Two Dense Communities:
(Imperfectly Integrated (60:40) but Realistic)
Y Coordinate
15
10
5
0
-5
-5
0
5
10
15
20
X Coordinate
P = 0.04


Two Dense Communities:
(Imperfectly Integrated (70:30) but Realistic)
Y Coordinate
15
10
5
0
-5
-5
0
5
10
15
20
X Coordinate
P = 0.006


Two Disperse Communities:
(Perfectly Segregated (0:100) and Unrealistic)
Y Coordinate
40
20
0
-20
-40
-40
-20
0
20
40
60
X Coordinate
P = 3.6e-5


Two Disperse Communities:
(Perfectly Integrated (50:50) and Unrealistic)
Y Coordinate
40
20
0
-20
-20
0
20
40
X Coordinate
P = 0.72

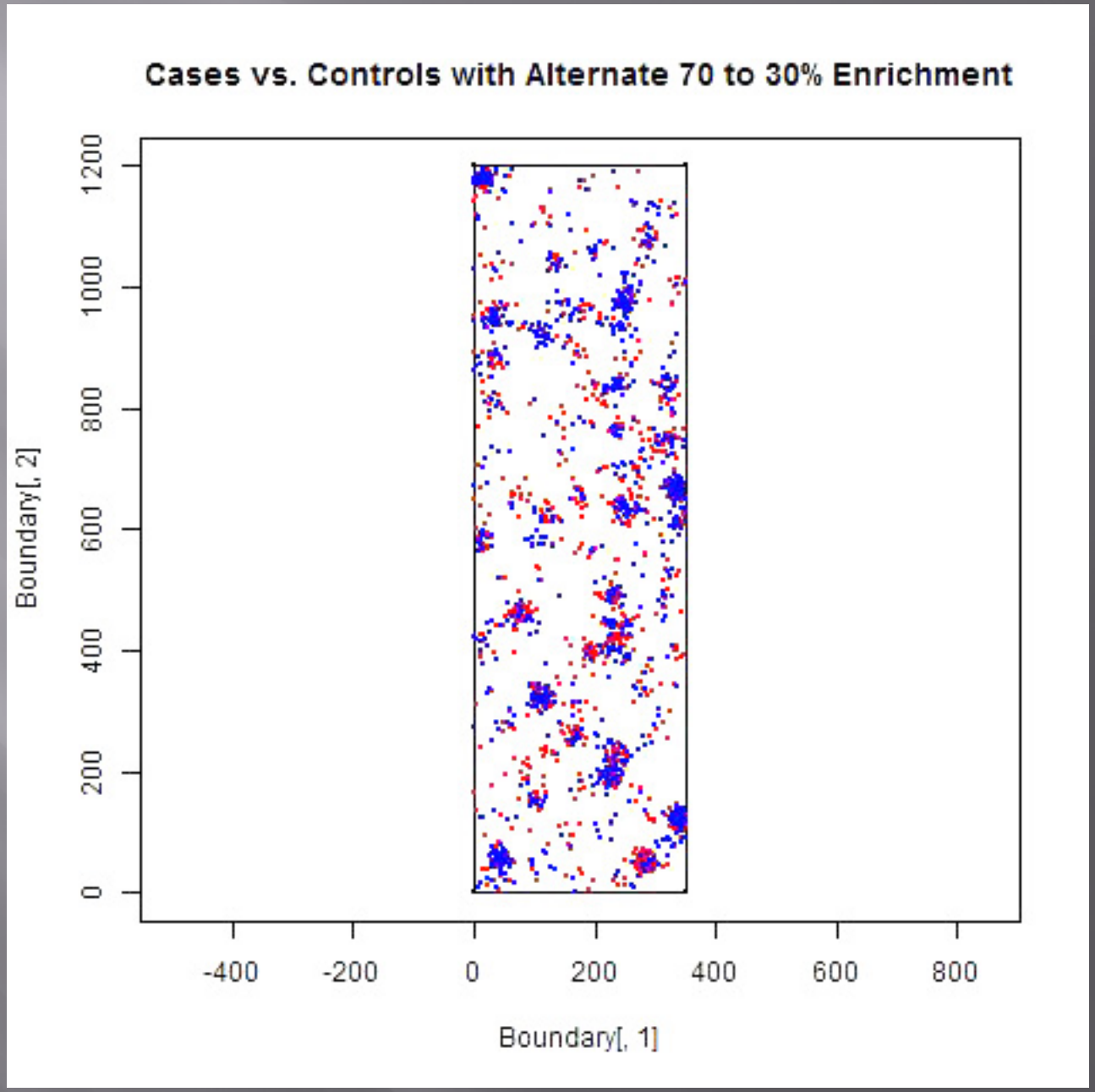

Two Disperse Communities:
(Imperfectly Integrated (60:40) but Realistic)
Y Coordinate
40
20
0
-20
-40
-40
-20
0
20
40
60
X Coordinate
P = 0.005

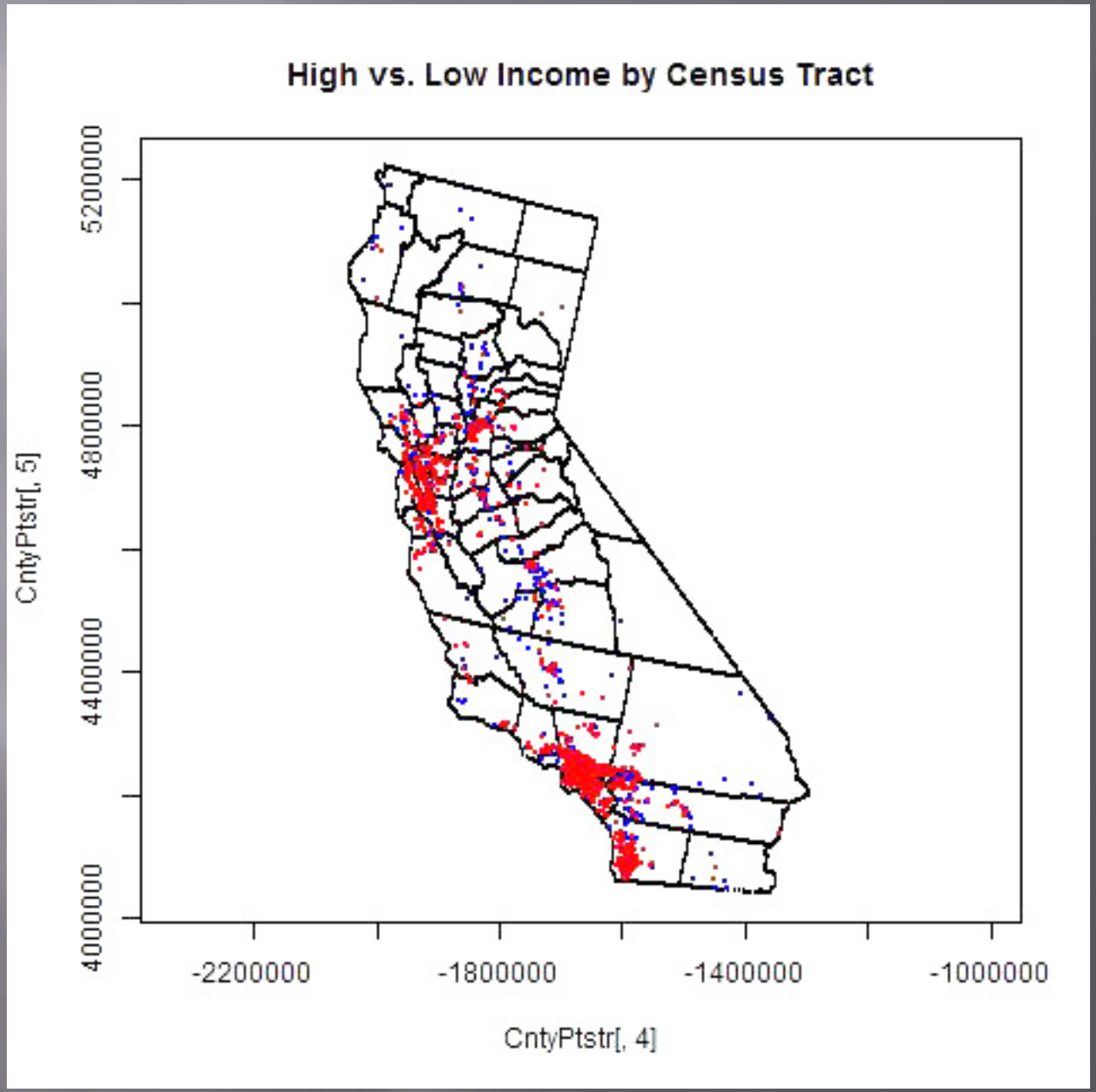

Two Disperse Communities:
(Imperfectly Integrated (70:30) but Realistic)
Y Coordinate
40
20
0
-20
-20
0
20
40
X Coordinate
P = 0.002

OBSERVED RATIOS OF "CASES" VS. "CONTROLS" FOR INDICATED CHARACTERISTICS AMONG THE 7,049 CENSUS TRACTS AND 58 COUNTIES IN CALIFORNIA

Ratios:	Fraction of Observed Ratios at Least as Extreme as Indicated Ratios							
	Higher vs. Lower Income	College Degree vs. None	White Collar vs. Blue Collar	Caucasians vs. Hispanics	Caucasians vs. Blacks	Mesothelioma vs. Pancreatic	Kaposi's vs. Pancreatic	Traffic Deaths vs. Pancreatic
	Census Tract Data							
80/20	23%	26%	22%	49%	**52%**			0%
75/25	35%	37%	31%	**59%**	**61%**			1%
70/30	48%	**50%**	41%	**68%**	**69%**			2%
60/40	**73%**	**75%**	**67%**	**85%**	**85%**			47%
50/50	**99%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
	County Data							
80/20	0%	0%	0%	34%	**55%**	3%	17%	3%
75/25	0%	7%	5%	43%	**62%**	3%	29%	12%
70/30	5%	24%	10%	**50%**	**66%**	7%	38%	29%
60/40	47%	**59%**	33%	**76%**	**79%**	16%	**52%**	**72%**
50/50	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

What's going on?


Cases vs. Controls with Alternate 70 to 30% Enrichment
Boundary[, 2]
0
200
400
600
800
1000
1200
-400
-200
0
200
400
600
800
Boundary[, 1]


High vs. Low Income by Census Tract
CntyPtstr[, 5]
5200000
4800000
4400000
4000000
-2200000
-1800000
-1400000
-1000000
CntyPtstr[, 4]

Conclusions

Epidemiological studies with spatial ecological designs typically produce invalid conclusions if regression models are used to interpret exposure-response associations; causality ***cannot*** be reasonably inferred from these studies.

Note: at least 100 studies of this design have been published and virtually all inappropriately suggest causality.

Useful Criteria: What to Look For in a Study

- Does the study link outcome with exposure explicitly?
- If not, does the study employ appropriate methods of spatial statistics?
- If not, does the study incorporate appropriate negative and positive controls?
- If not, does the study at least attempt to control for all reasonable factors that affect where individuals choose to live?

References

- Berman, D.W.; Cox, L.A.; Popken, D. "A Cautionary Tale: The Characteristics of Two-Dimensional Distributions and their Effects on Epidemiological Studies Employing an Ecological Design." *Critical Reviews in Toxicology*. 2013; 43(S1): 1–25. (doi:10.3109/10408444.2013.777688). Available at: http://informahealthcare.com/toc/txc/43/S1.
- Cox, L.A.; Berman, D.W.; Popken, D. "Causal vs. Spurious Spatial Exposure-Response Associations in Health Risk Analysis." *Critical Reviews in Toxicology*.. 2013;43(S1): 26-38 (doi:10.3109/10408444.2013.777689). Available at: http://informahealthcare.com/toc/txc/43/S1.
- Pan, XL; Day, HW; Wang, W; Beckett, LA; Schenker, MB. 2005. Residential proximity to naturally occurring asbestos and mesothelioma risk in California. *Am J Respir Crit Care Med*. 15;172(8):1019-25. October.
- R Development Core Team. 2011. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. Retrieved: 2/21/12.

Questions