Geographical Science for Beginner GIS Users in Cancer Registries

Gerard Rushton, Ph.D.

Department of Geography and The College of Public Health

The University of Iowa

Gerard-rushton@uiowa.edu

Presented at the NAACCR Symposium on "Cancer Informatics: Essential Technologies and Methodologies for Registries"

Toronto, Canada, June 9, 2002

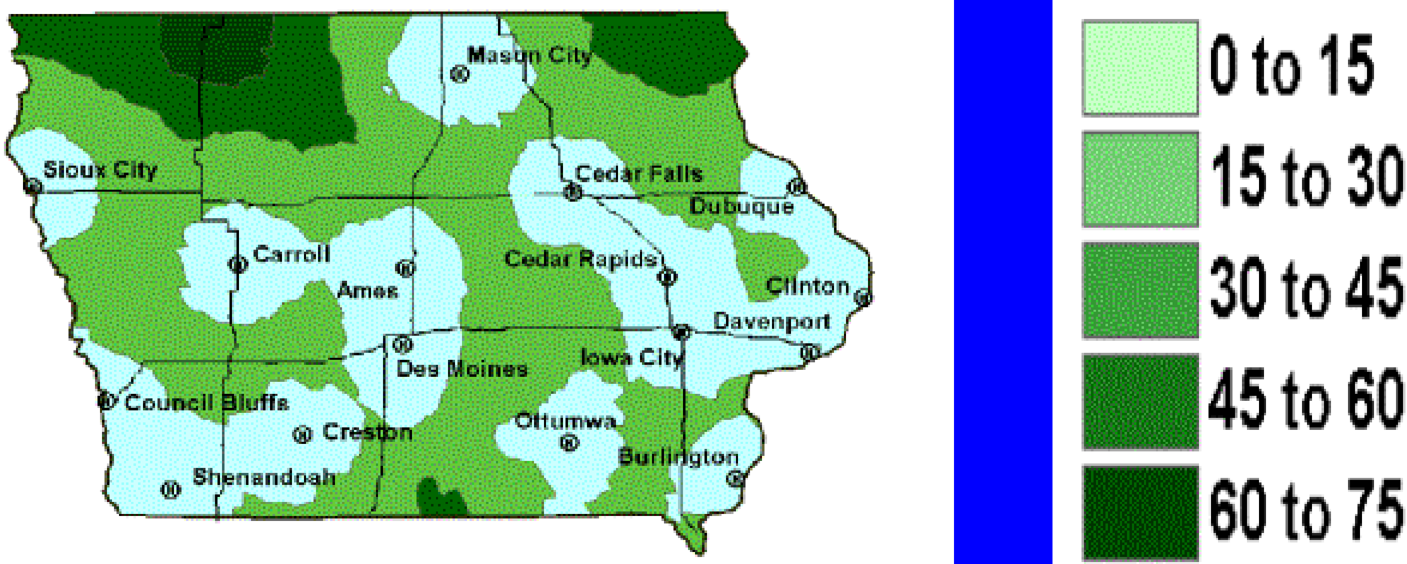
Topics

- Mapping geographic access to cancer treatment facilities
- Mapping changing spatial patterns of use of cancer facilities
- What can be learned by mapping spatial patterns of residuals from statistical models of stage at diagnosis for colorectal cancer and breast cancer?
- Introduction to web site www.uiowa.edu/~giscancr

Geographic access to cancer treatment

- Two basic sets of information for maps of access to treatment:
  - Distances from grid points to treatment facilities.
    - Example: access to radiation treatment facilities
  - Distances from individuals with known characteristics to treatment facilities.
    - Example: percent with localized breast cancer who choose lumpectomy with radiation treatment rather than mastectomy.

Distance to Closest Radiation Treatment Facility (miles)



Cases of Breast Cancer in Iowa 1993-1997

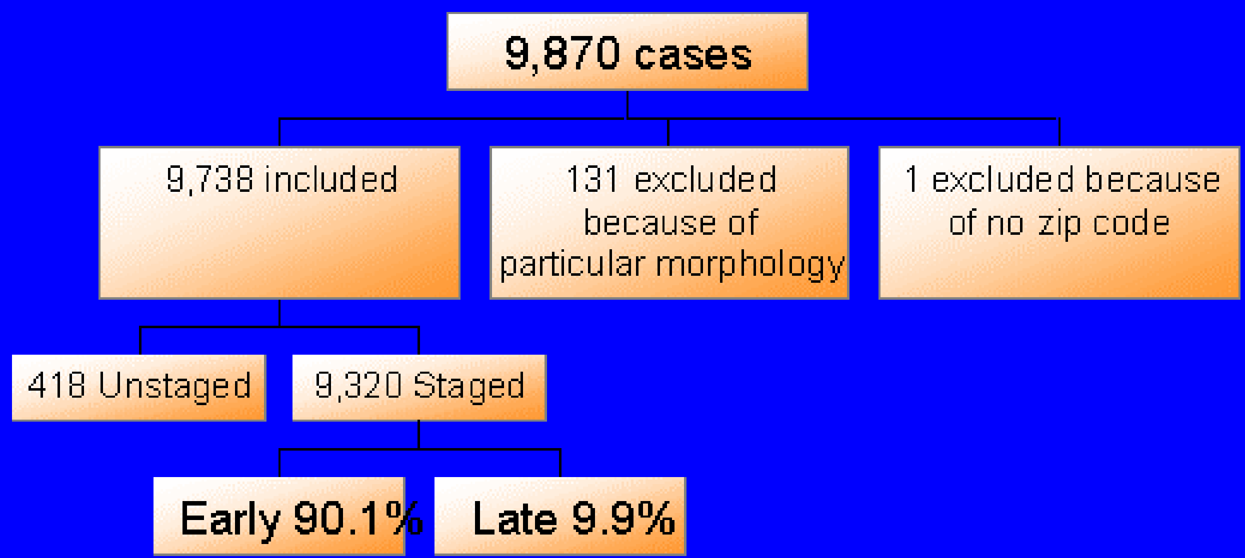
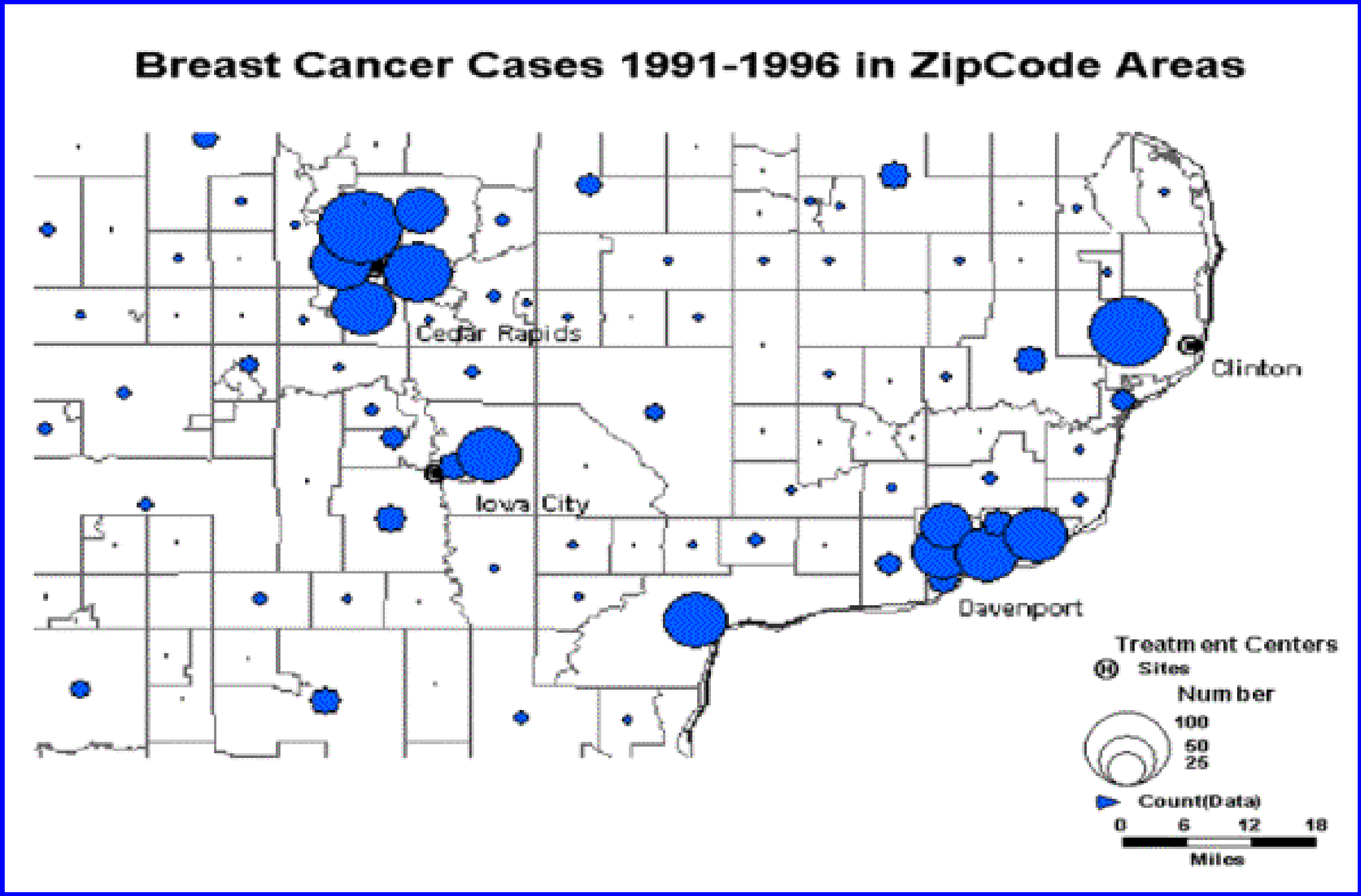


Illustration of Spatial Data: south-east Iowa

- Basic spatial unit is the five digit zip code
- There are approximately 940 such areas in Iowa
- The breast cancer cases are geo-coded to the centroid of the zip code areas
- Radiation facilities are geo-coded to their street address

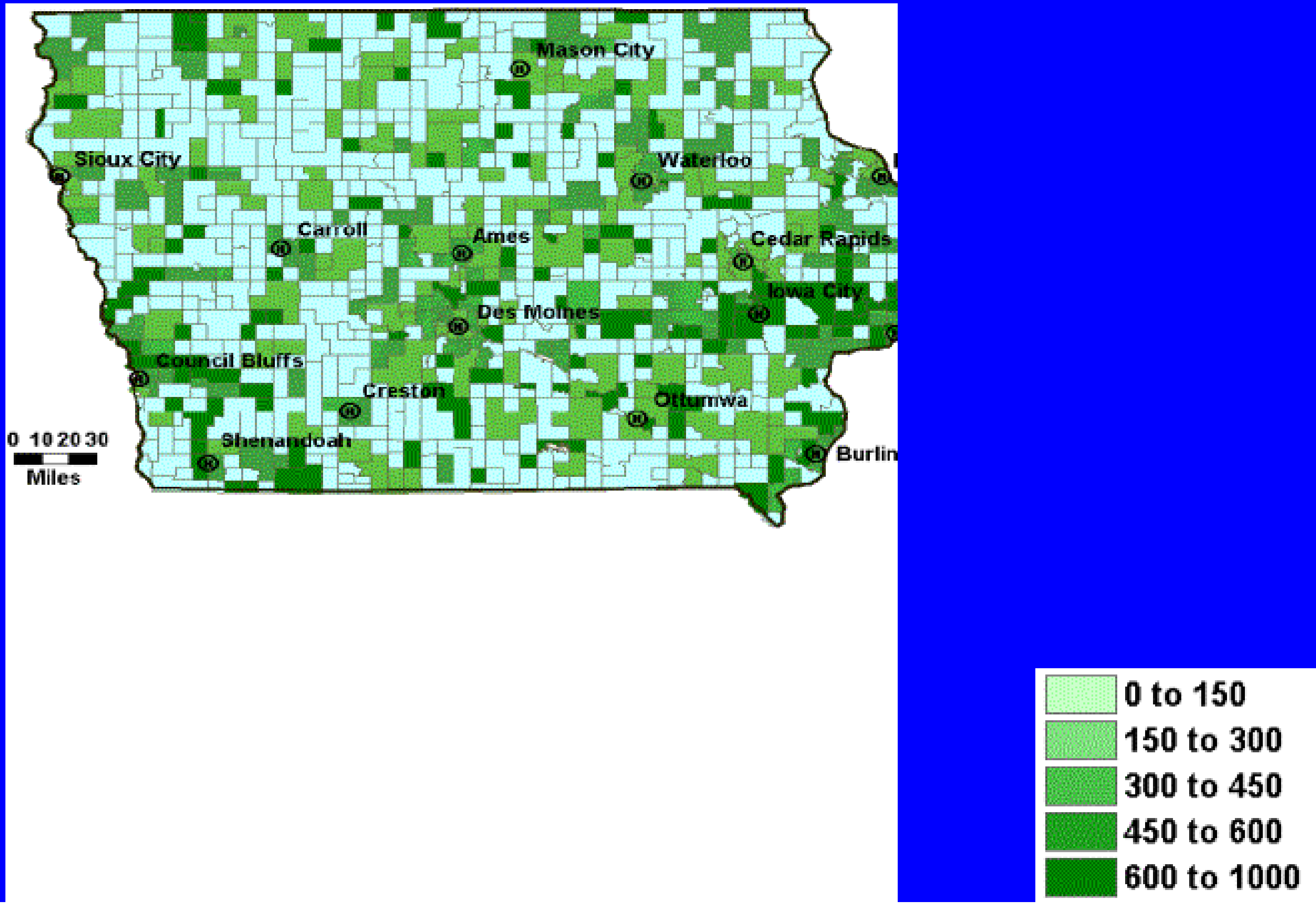
Illustration of the raw data: number of breast cancer cases in the six-year period: south east Iowa.



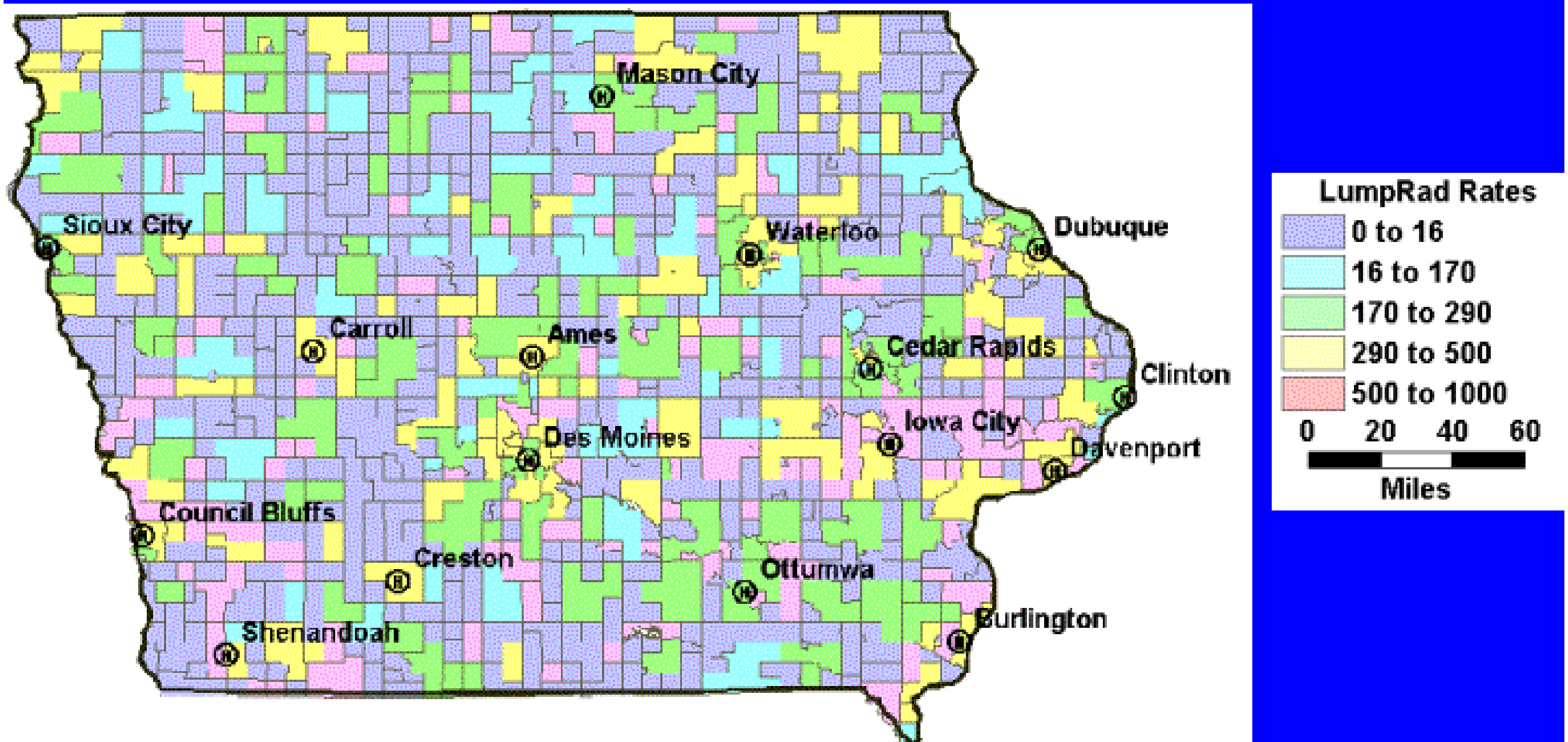
Don't make maps like these next two!

- Why not?

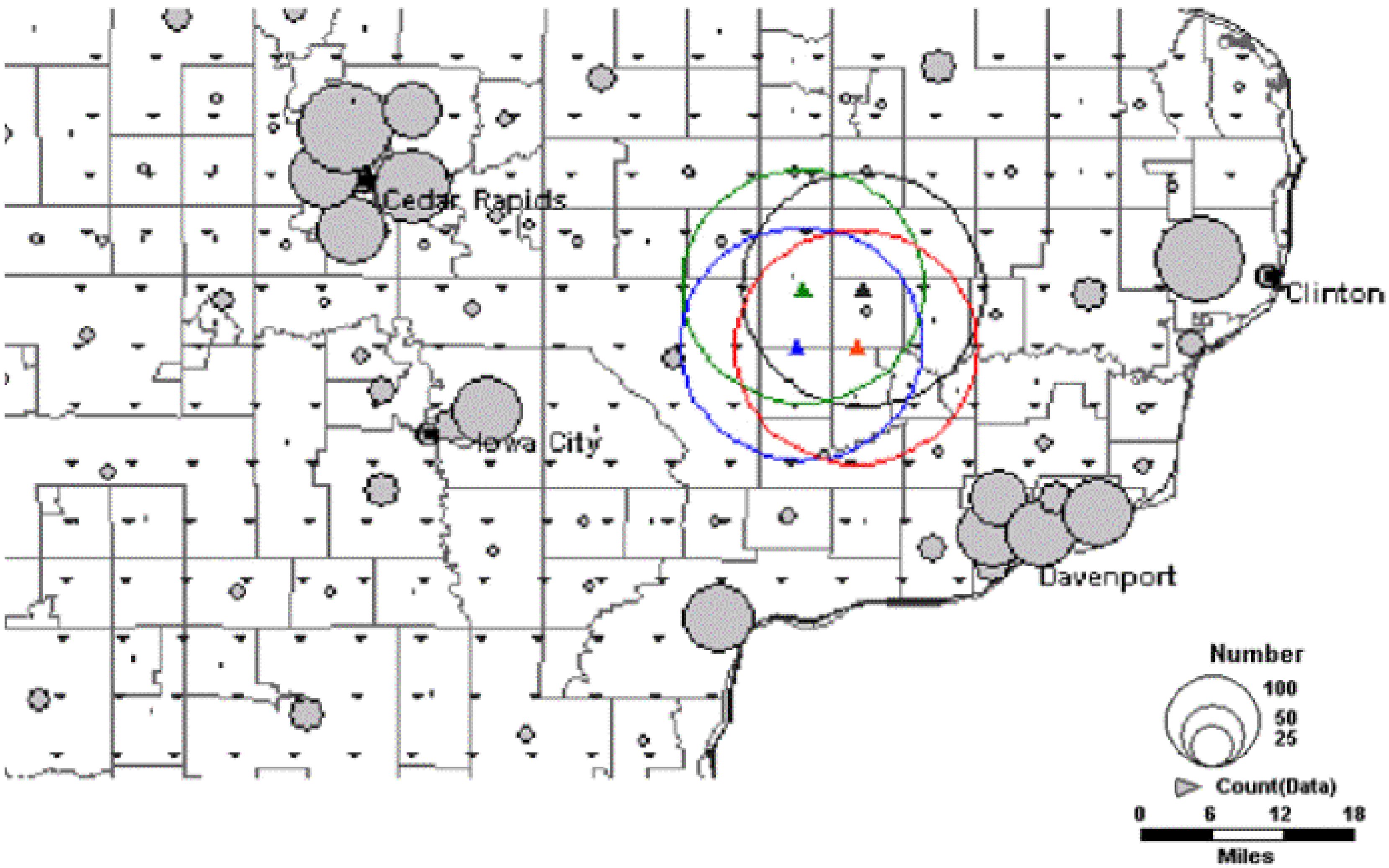
Proportion of Women with Localized Breast Cancer who Selected Lumpectomy with Radiation (per thousand)



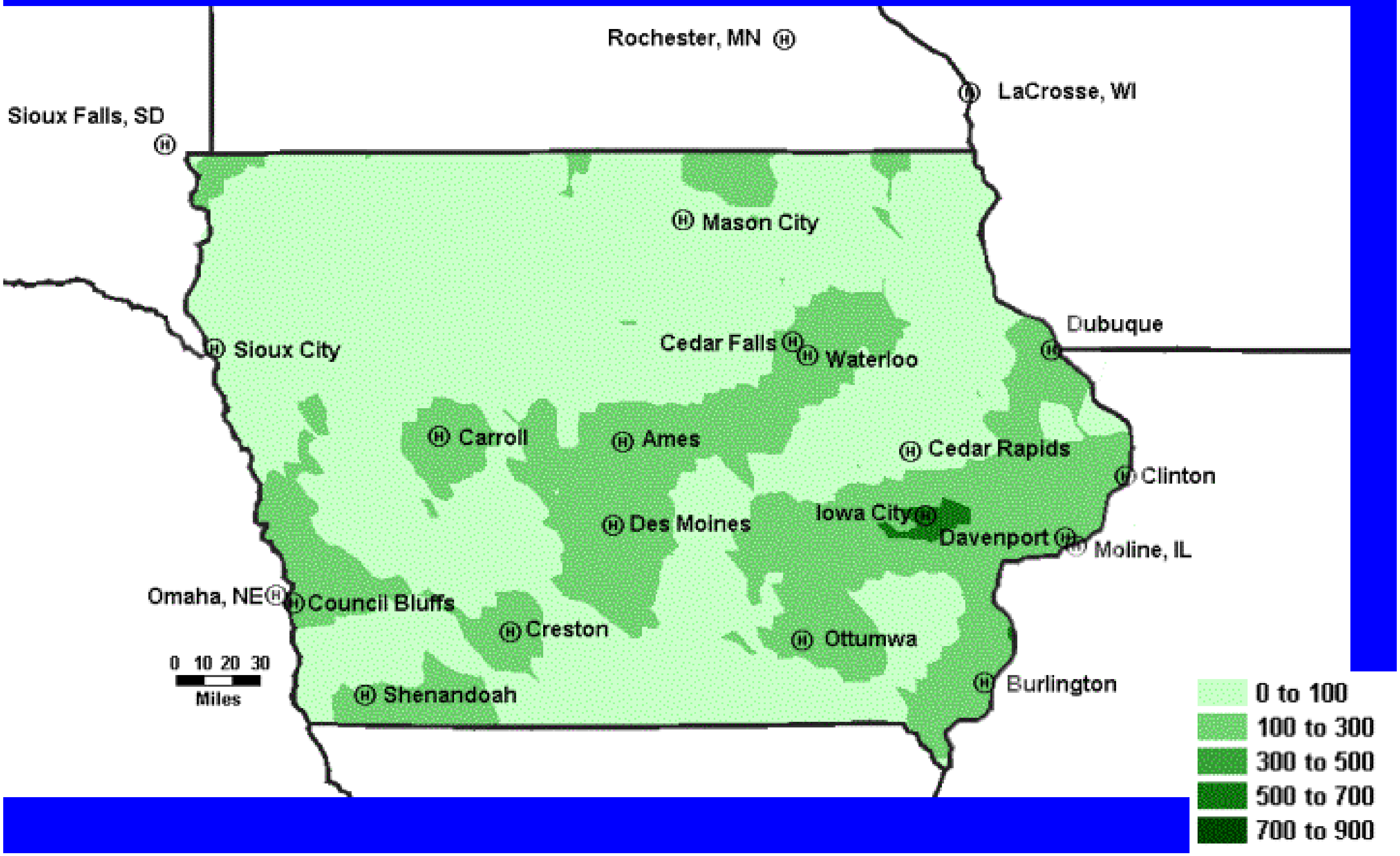
Changing Colors doesn't help!



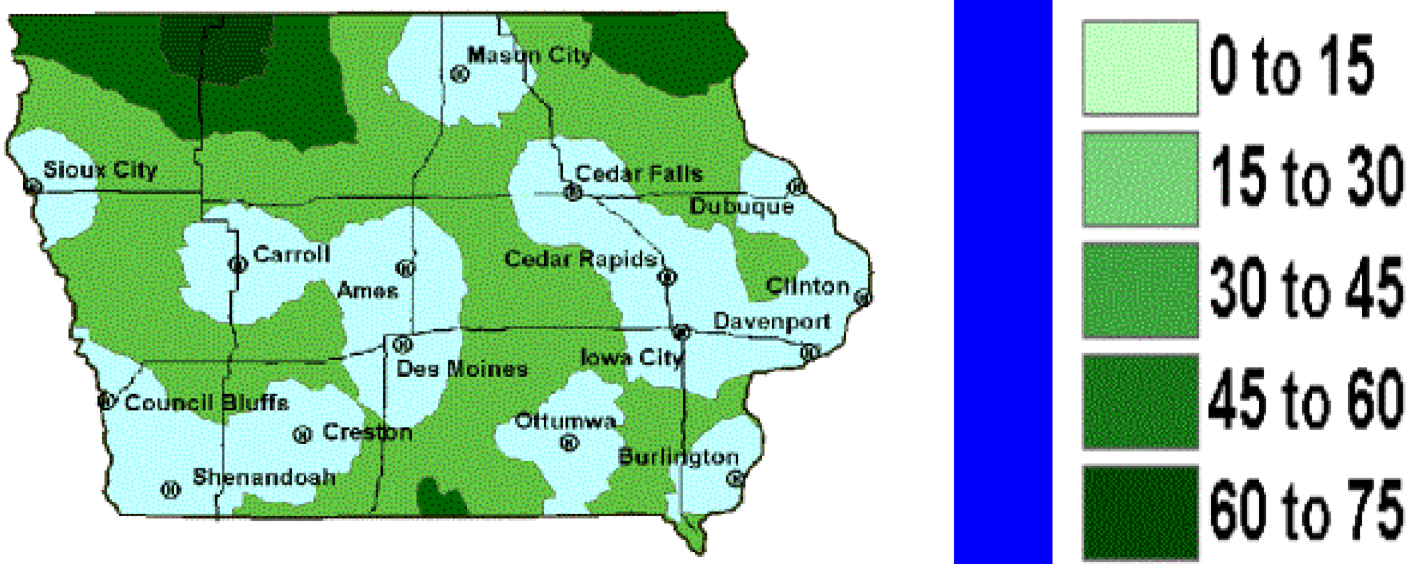
Five Mile Grid and 10 Mile Filters



Compare areas of high rates of lumpectomy/radiation with locations of radiation treatment facilities
Choices per 1,000 cases of localized breast cancer, 15 mile filter



Distance to Closest Radiation Treatment Facility (miles)



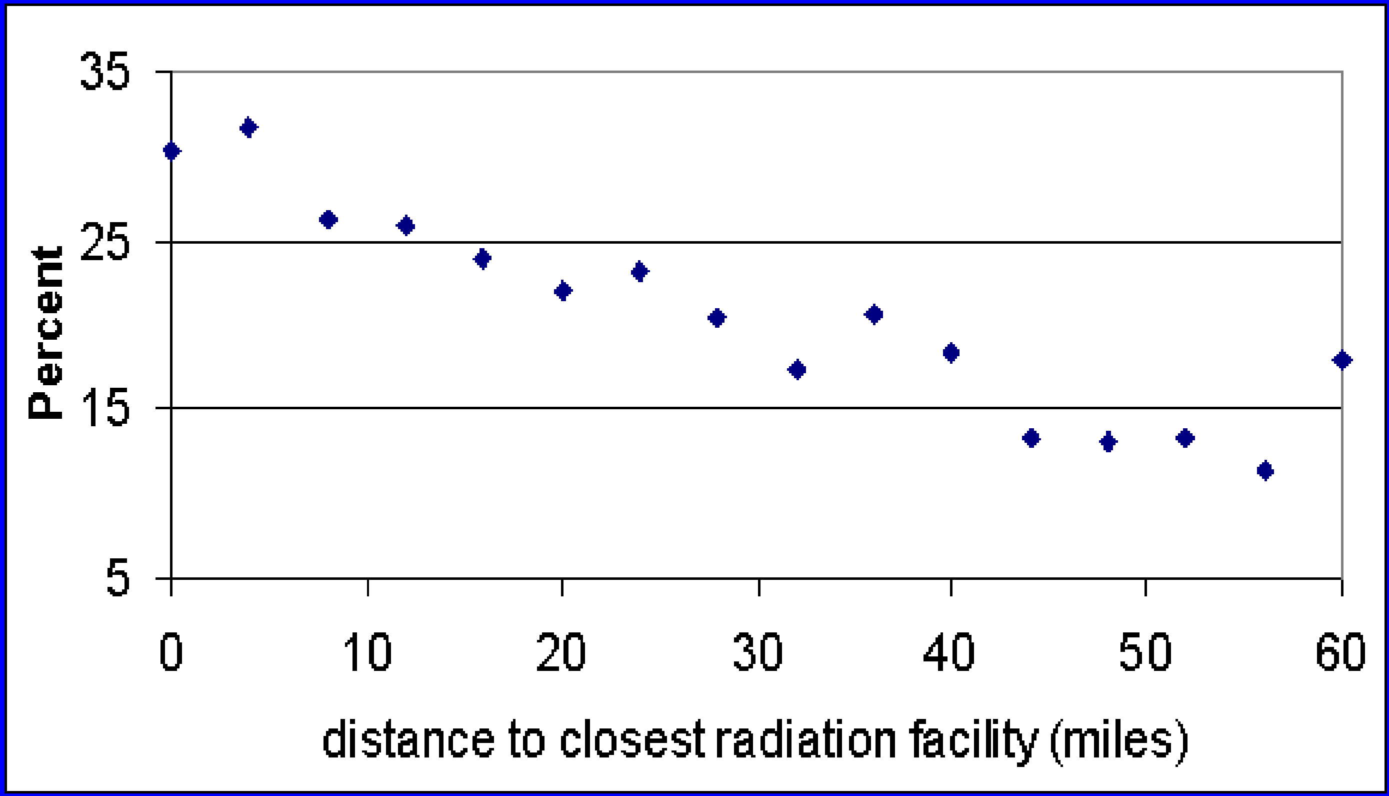
Comment on difference between the two maps

- The two spatial patterns are not obviously similar are they?
- So, is there a relationship between access to a radiation treatment center and choice of treatment type?
- Consider people at the same distance from their closest radiation facility. Are they more likely to use radiation treatment when they are close to a radiation facility?

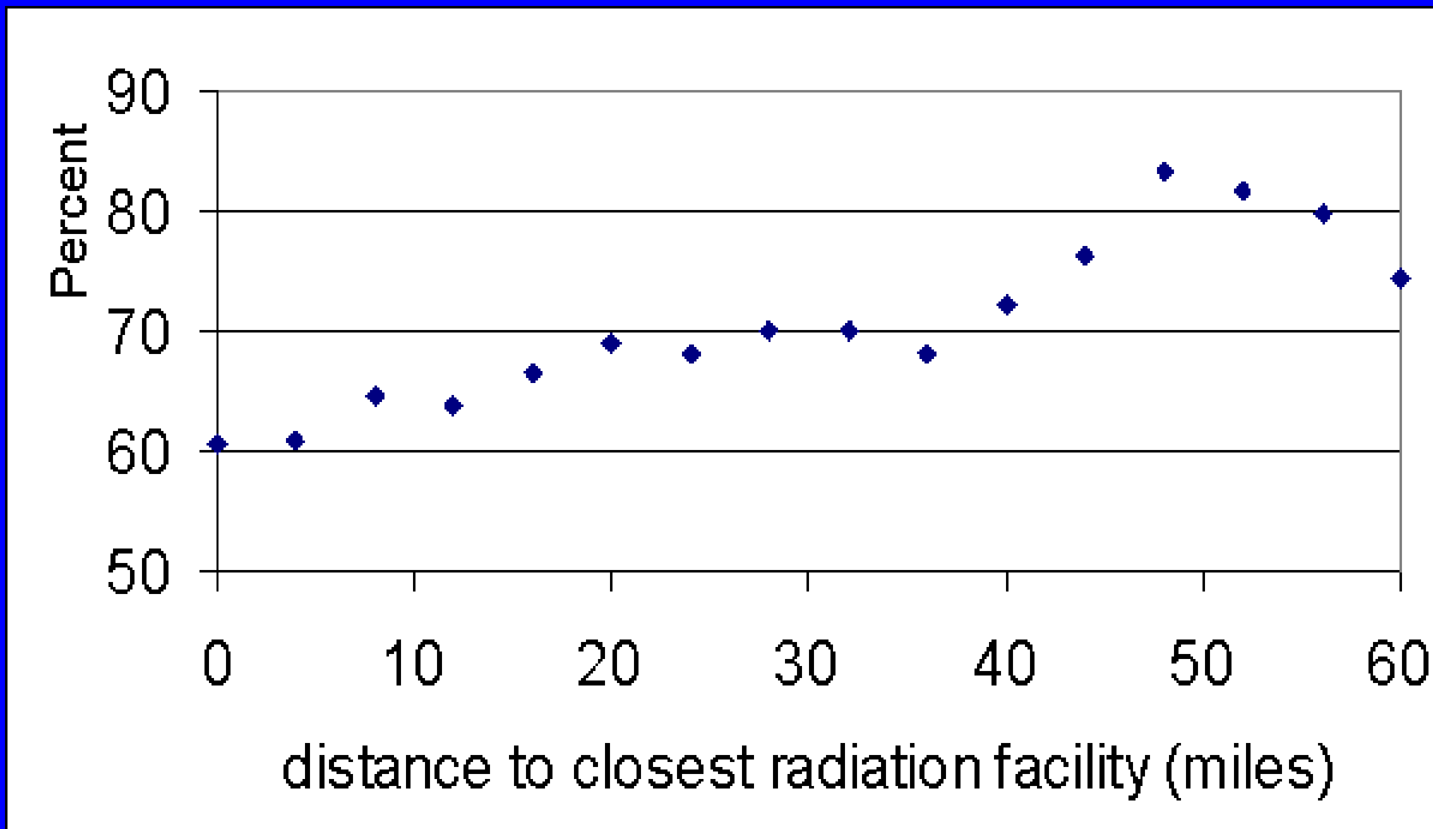
Distance to Closest Radiation Facility and Treatment Choice

Distance	All Cases	# Mast.	%Mast.	# LumpRad.	% LumpRad
0-4	1628	985	60.5	494	30.3
4-8	1141	693	60.7	362	31.7
8-12	832	537	64.5	219	26.3
12-16	584	373	63.9	151	25.9
16-20	687	456	66.4	165	24.0
20-24	819	564	68.9	181	22.1
24-28	977	666	68.2	226	23.1
28-32	936	654	69.9	192	20.5
32-36	753	528	70.1	131	17.4
36-40	522	355	68.0	108	20.7
40-44	451	325	72.1	83	18.4
44-48	331	252	76.1	44	13.3
48-52	160	133	83.1	21	13.1
52-56	173	141	81.5	23	13.3
56-60	114	91	79.8	13	11.4
60+	272	202	74.3	49	18.0
Total	10380	6955	67.0	2462	23.7

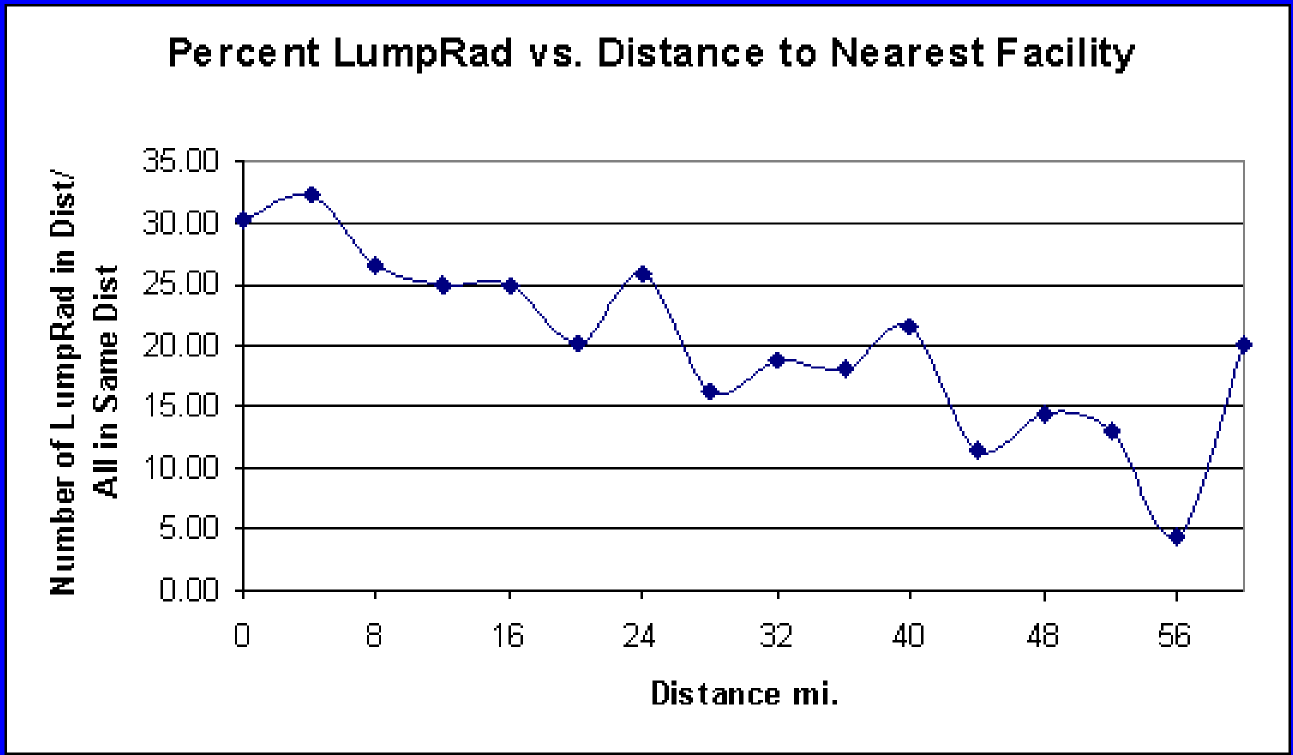
Percent Selecting Lumpectomy/Radiation Treatment



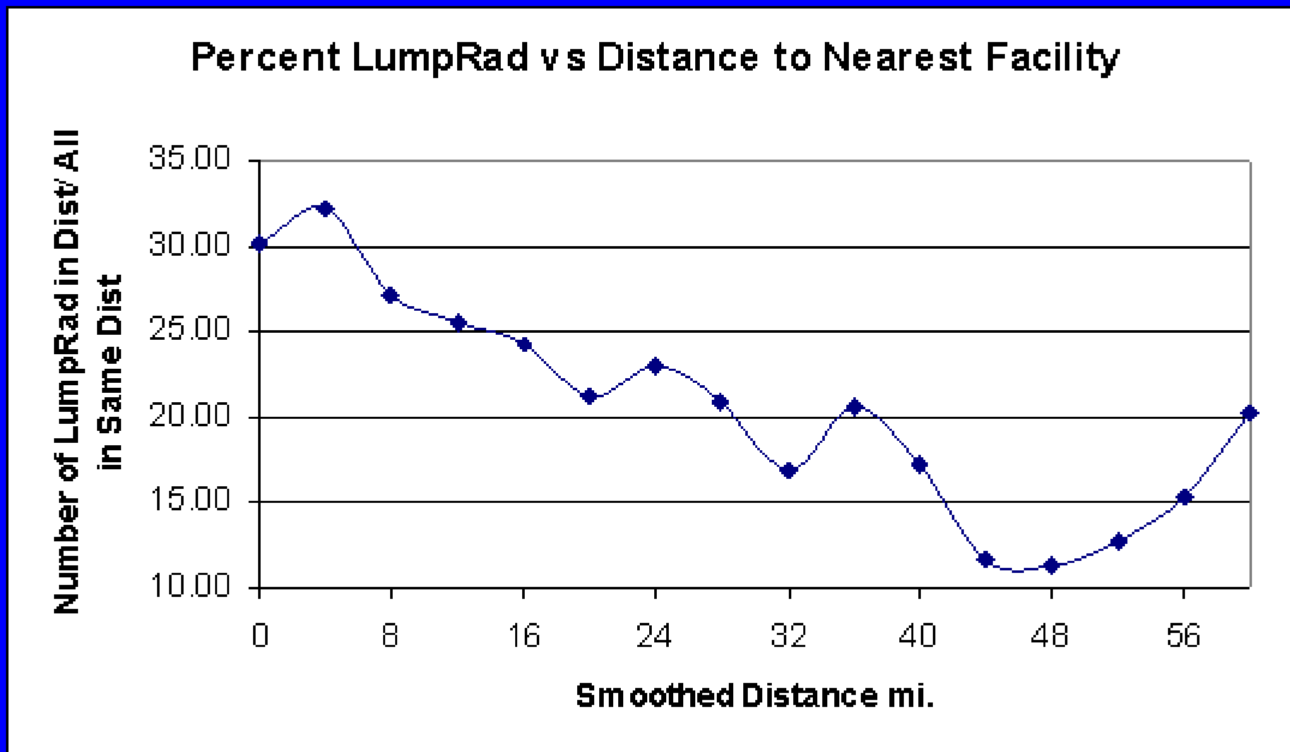
Percent Selecting Mastectomy Treatment



LumpRad - UnSmoothed - Graph Set 1

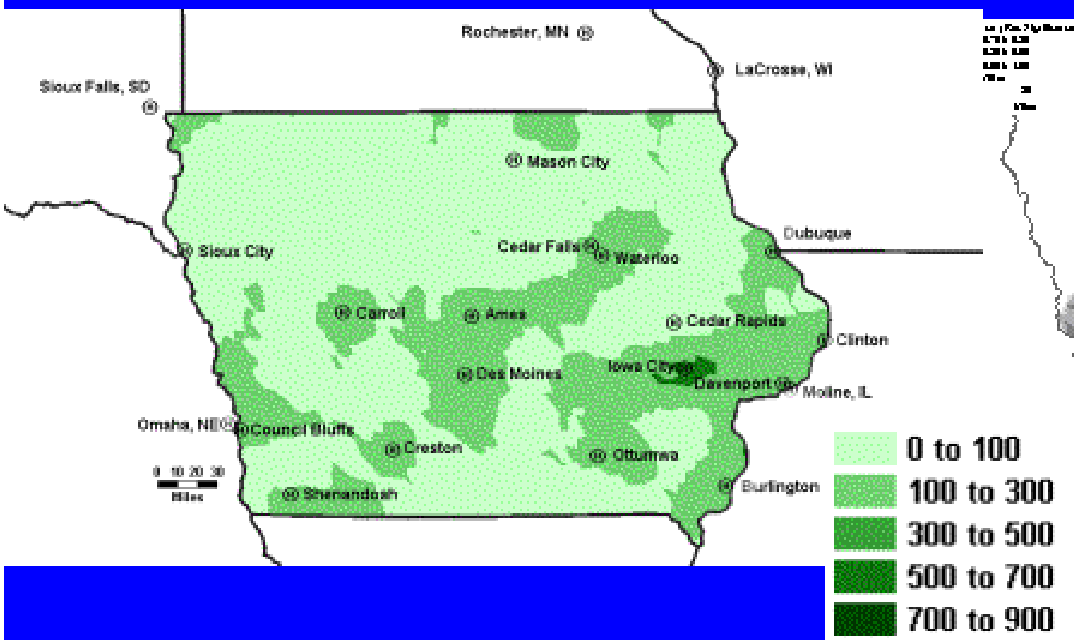


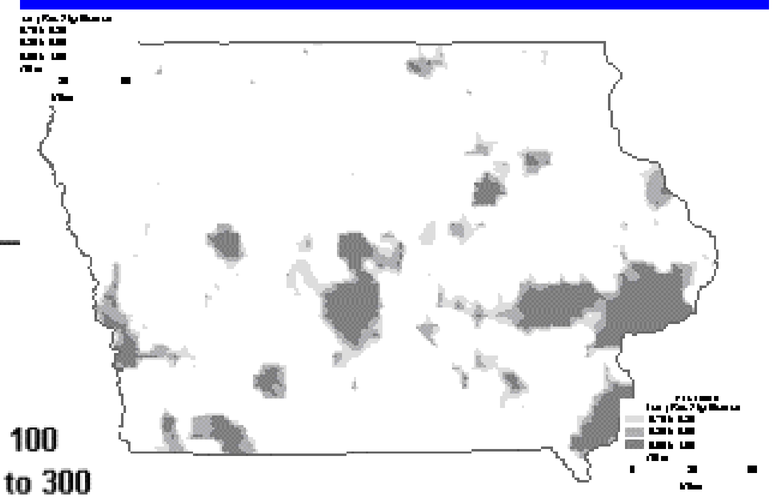
LumpRad - Smoothed - Graph Set 1



Adjust the map of areas of high rates of lumpectomy/radiation for the different numbers of cases within the 15 mile spatial filter

In the dark areas, 90 percent of the simulated maps had lumpectomy/radiation choice rates less than the observed rate at that location



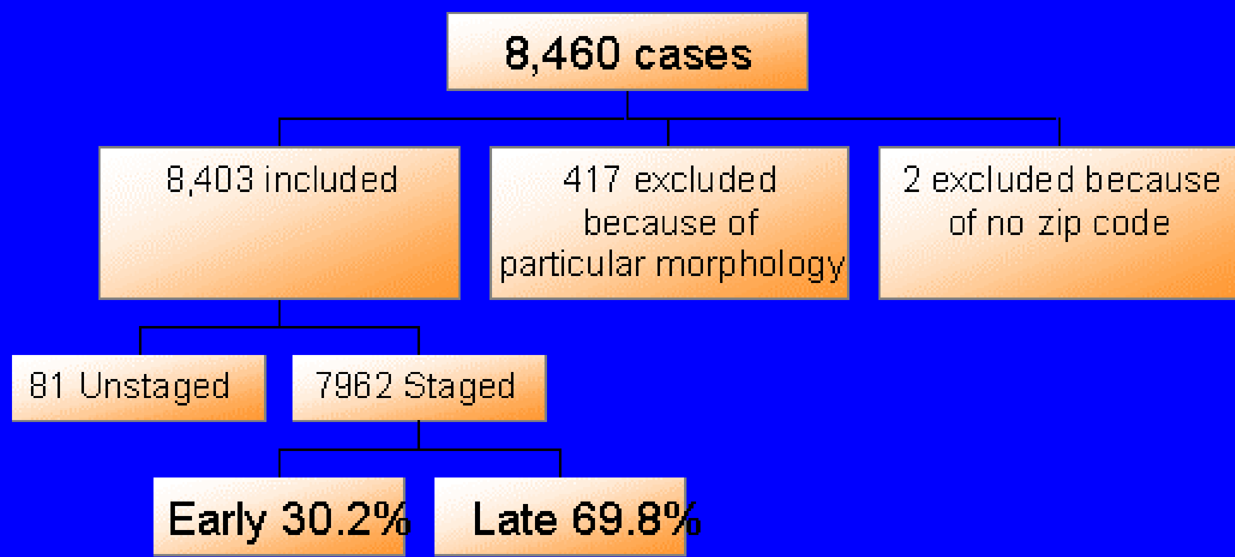


Topics

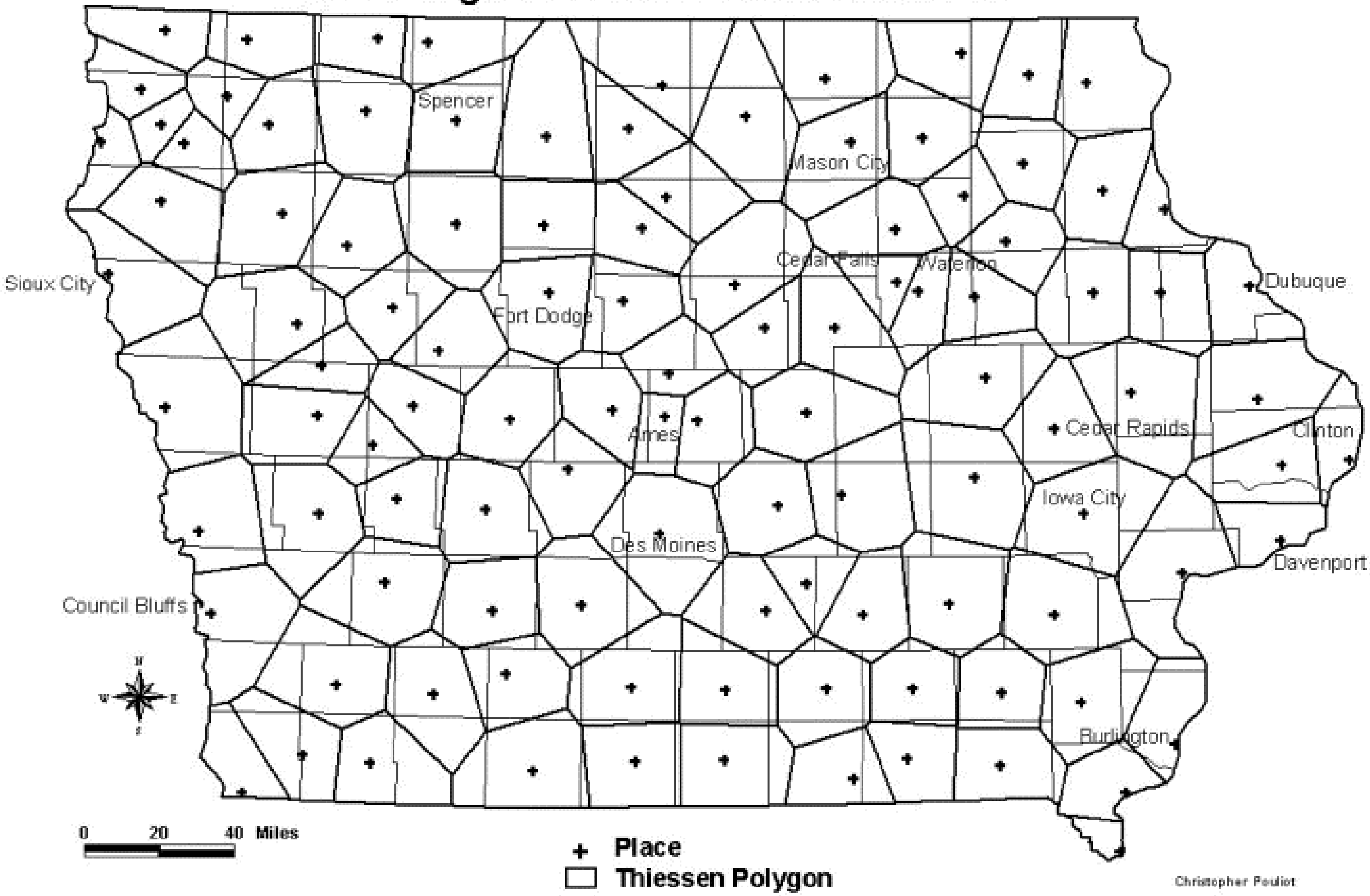
- Mapping geographic access to cancer treatment facilities
- **Mapping changing spatial patterns of use of cancer facilities**
- What can be learned by mapping spatial patterns of residuals from statistical models of stage at diagnosis for colorectal cancer and breast cancer?
- Introduction to web site www.uiowa.edu/~giscancr

Figure 2

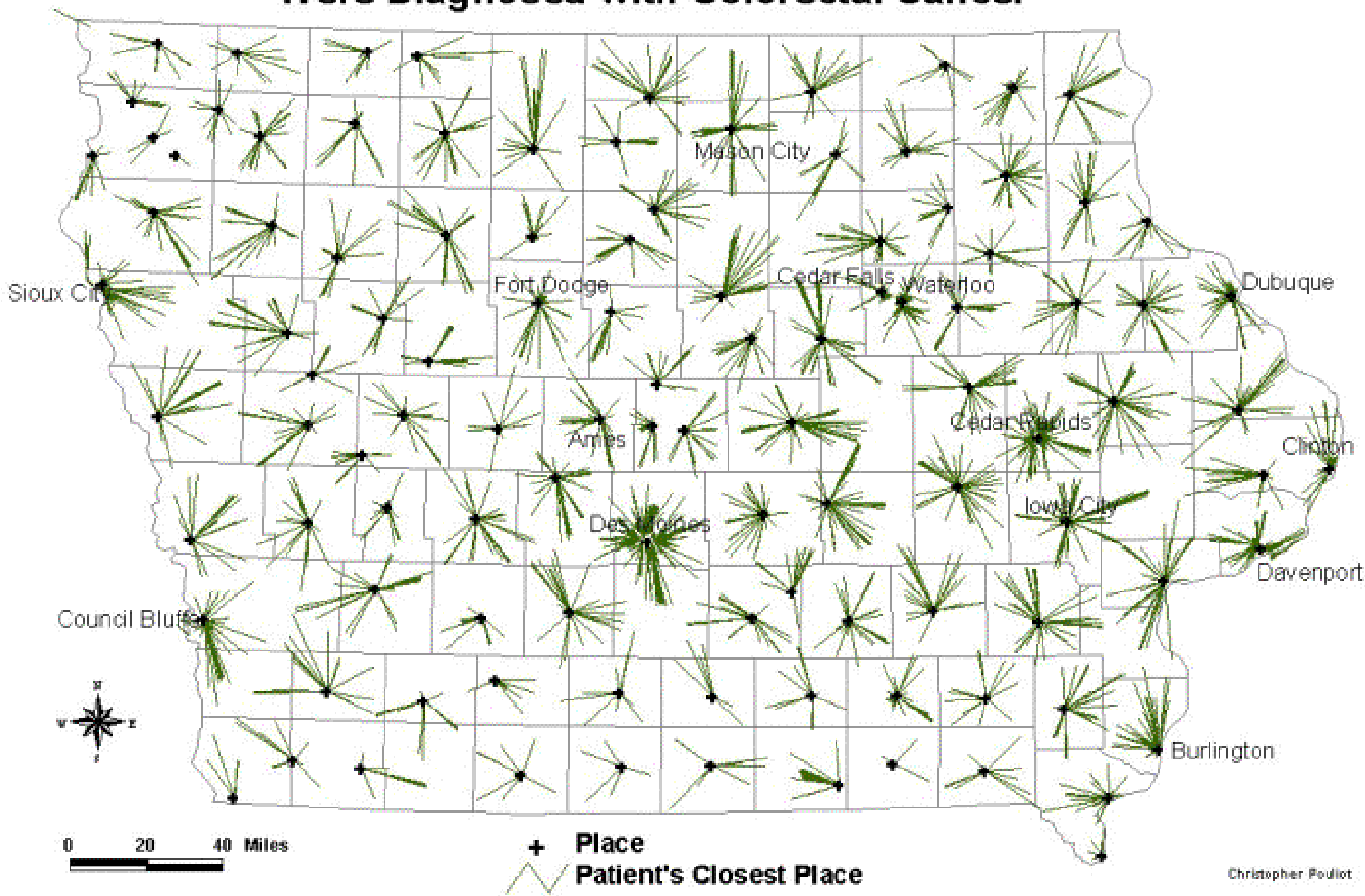
Cases of Colorectal Cancer in Iowa 1993-1997



Closest Place Where One or More Persons Were Diagnosed with Colorectal Cancer



Closest Place Where One or More Persons Were Diagnosed with Colorectal Cancer



Place of First Diagnosis for Colorectal Cancer, Iowa 1993-97

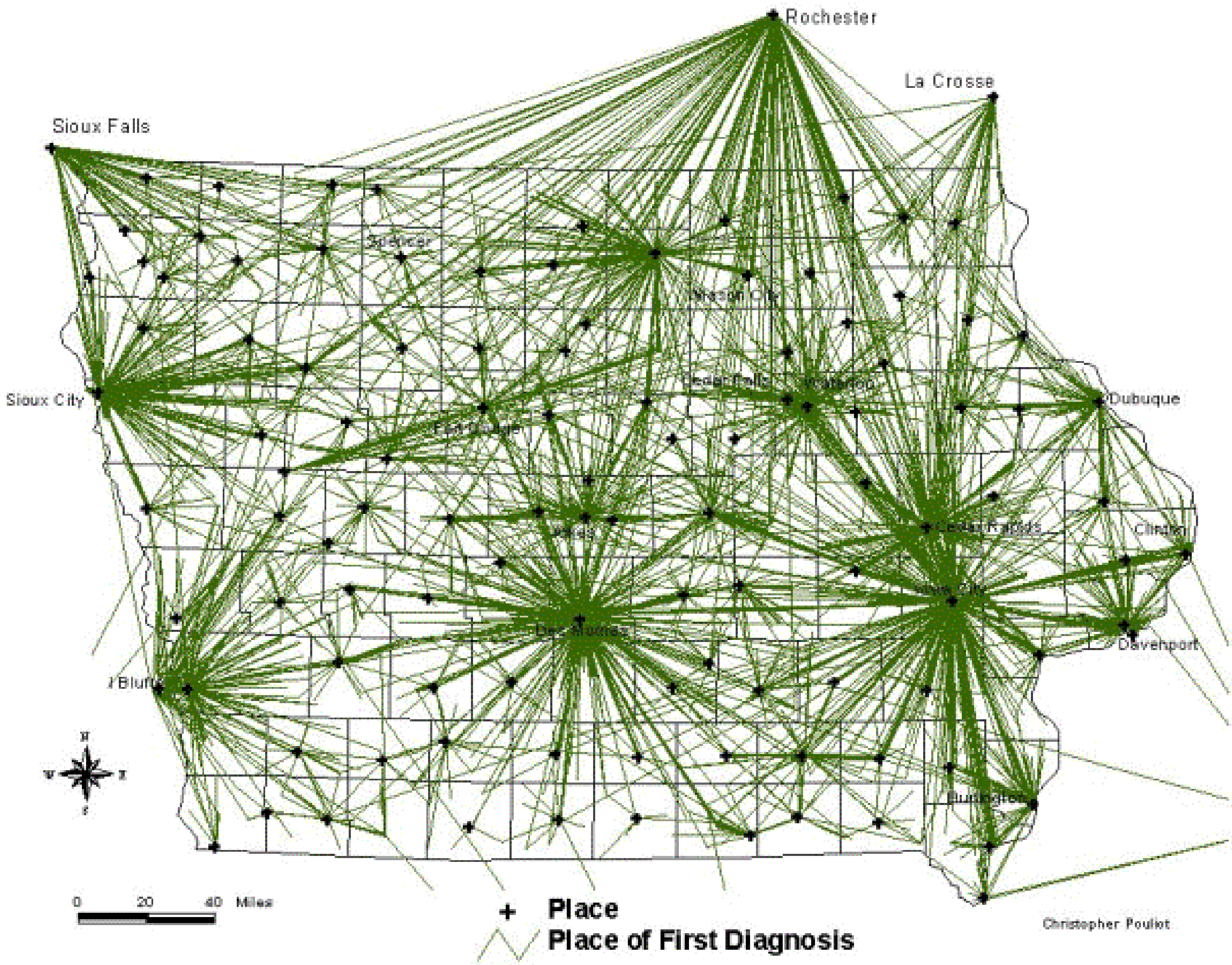
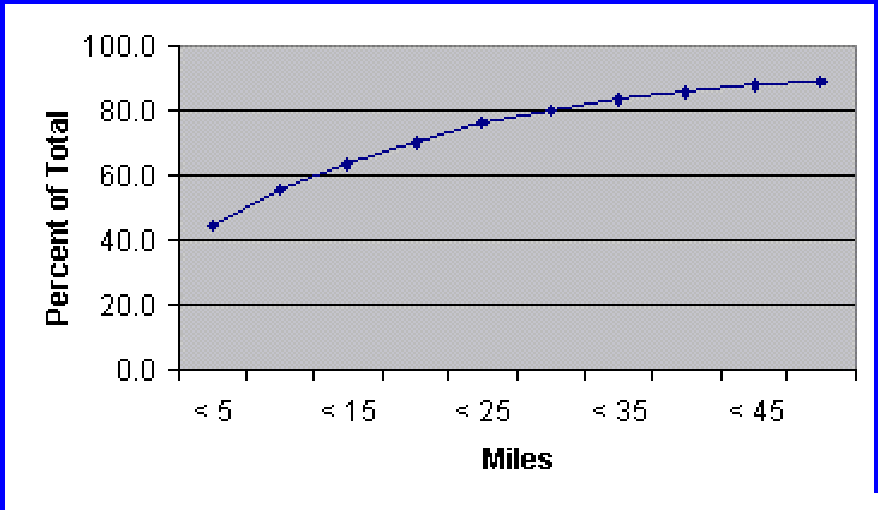


Figure 14

Distance to Place of First Diagnosis

Miles	Count	% of Total
< 5	3663	44.6
< 10	4579	55.7
< 15	5225	63.6
< 20	5736	69.8
< 25	6264	76.2
< 30	6597	80.3
< 35	6864	83.5
< 40	7044	85.7
< 45	7208	87.7
< 50	7321	89.1



Miles	Count	% of Total
<10	4579	55.7
<20	5736	69.8
<30	6597	80.3
<40	7044	85.7
<50	7321	89.1
<60	7532	91.6
<70	7666	93.3
<80	7778	94.6
<90	7894	96.0
<100	7951	96.7

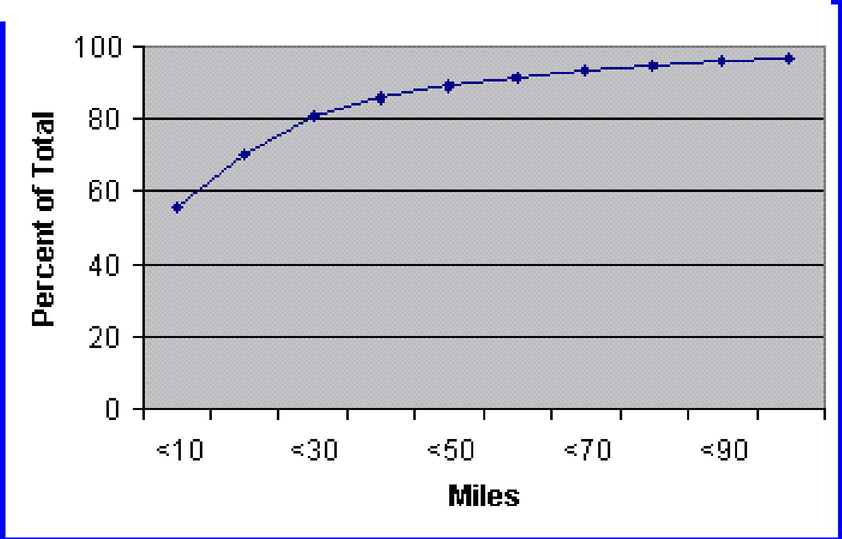
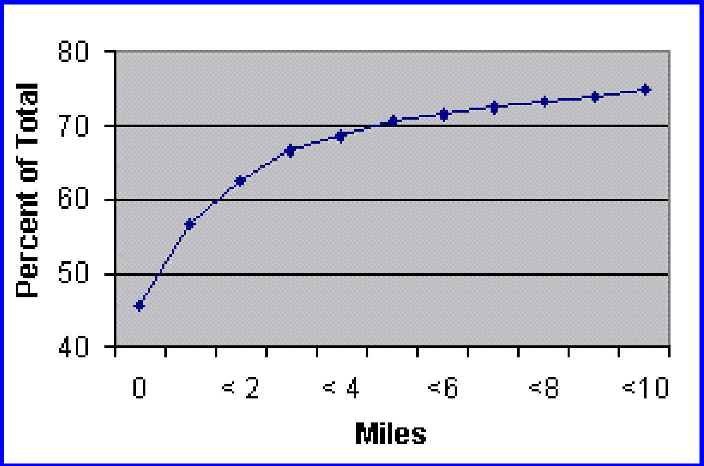


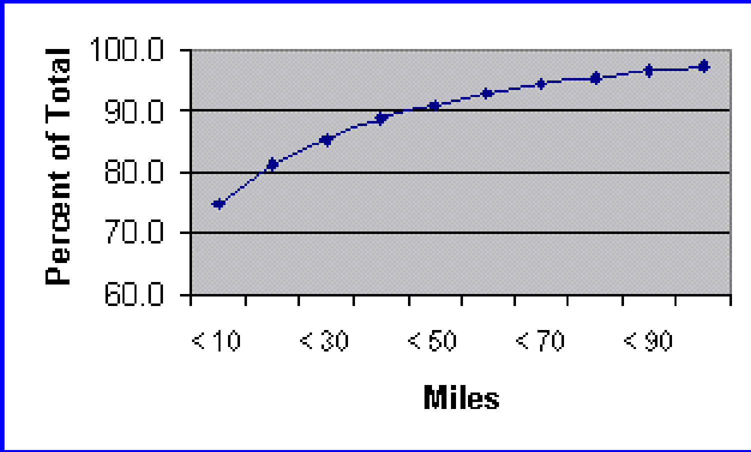
Figure 15

Difference Between Distance to Place of First Diagnosis and Distance to Closest Place Where Such Diagnoses are Made

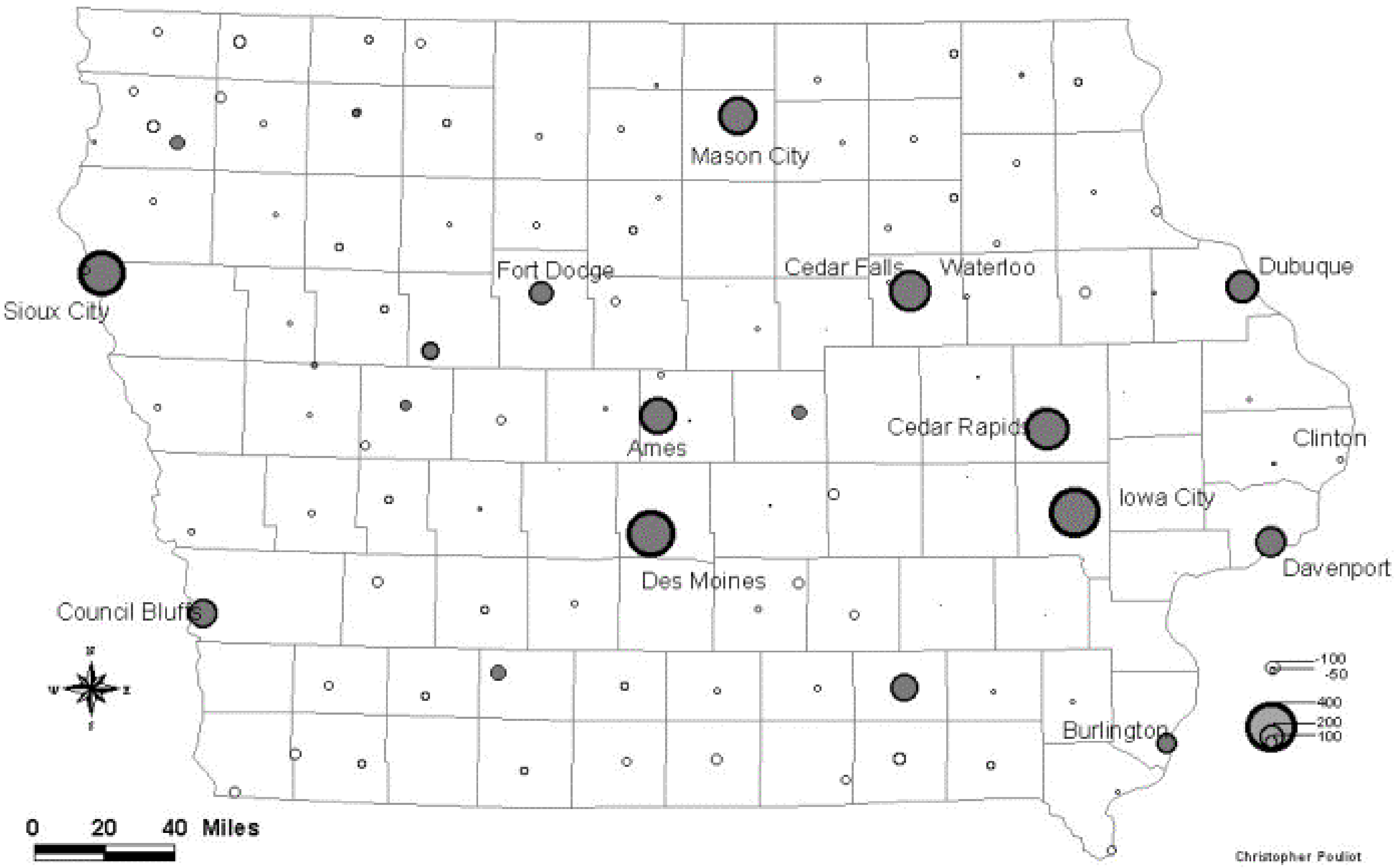
Miles	Count	% of Total
0	3757	45.7
< 1	4651	56.6
< 2	5127	62.4
< 3	5471	66.6
< 4	5632	68.5
< 5	5804	70.6
< 6	5883	71.6
< 7	5946	72.3
< 8	6015	73.2
< 9	6082	74.0
< 10	6151	74.8



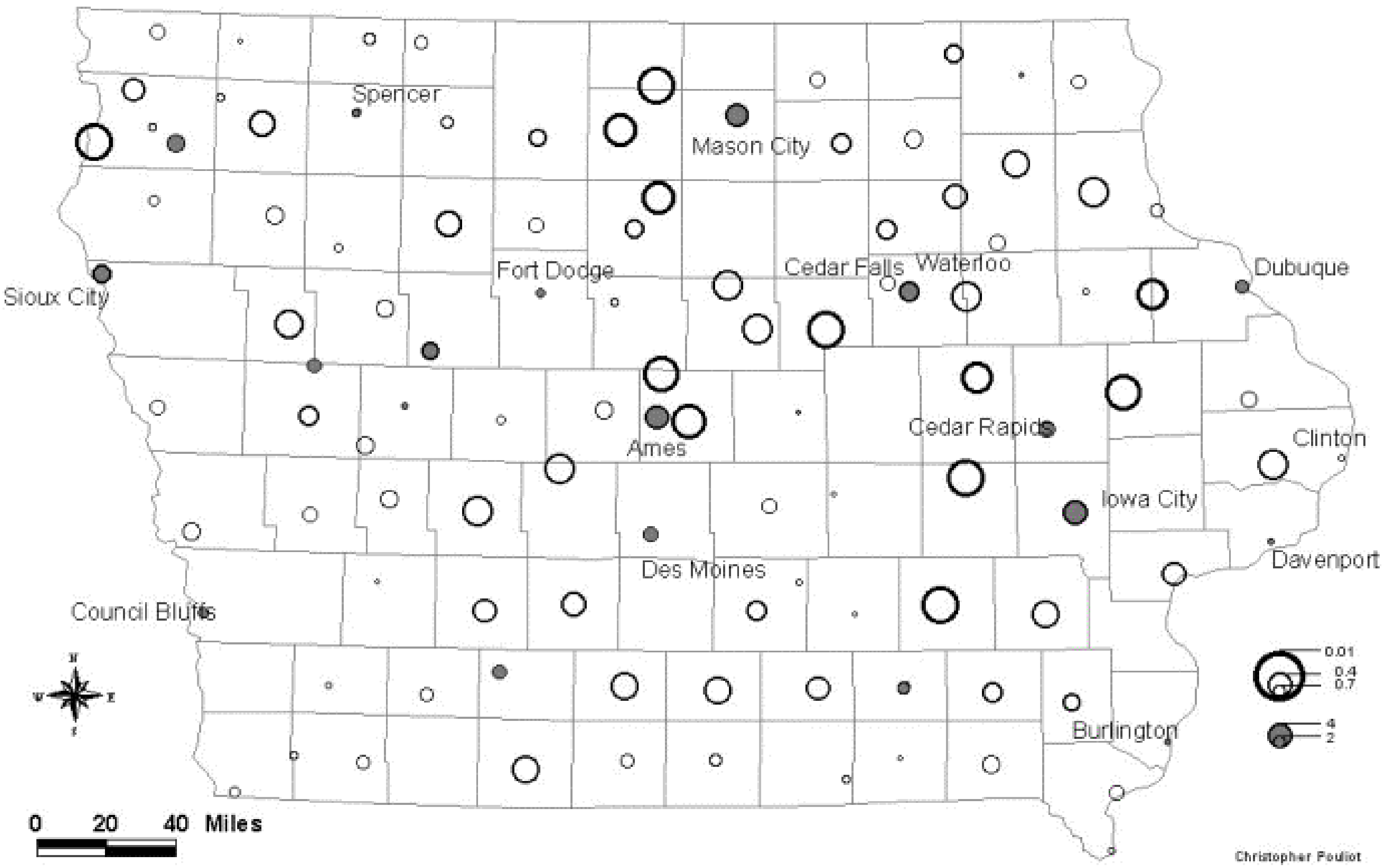
Miles	Count	% of Total
< 10	6151	74.8
< 20	6665	81.1
< 30	7004	85.2
< 40	7282	88.6
< 50	7471	90.9
< 60	7622	92.7
< 70	7745	94.2
< 80	7826	95.2
< 90	7935	96.5
< 100	7977	97.1



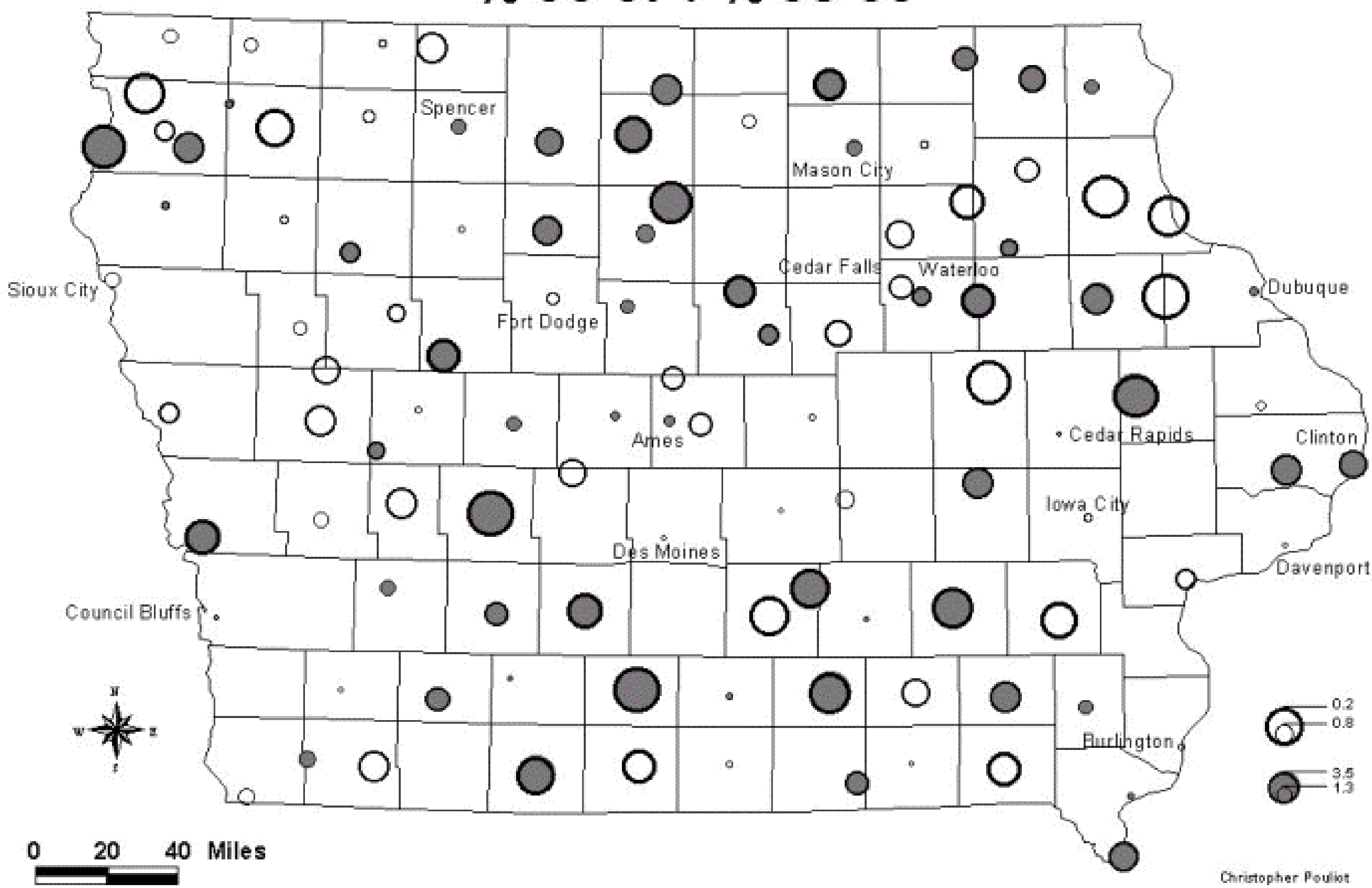
Colorectal Cancer Diagnoses for Iowa Towns, 1993-97
Actual # - Closest Allocation



Colorectal Cancer Diagnoses for Iowa Towns, 1993-97
Actual # / Closest Allocation



Change in Rate of Local Colorectal Cancer Diagnosis
% 96-97 / % 93-95



Topics

- Mapping geographic access to cancer treatment facilities
- Mapping changing spatial patterns of use of cancer facilities
- **What can be learned by mapping spatial patterns of residuals from statistical models of stage at diagnosis for colorectal cancer and breast cancer?**
- Introduction to web site www.uiowa.edu/~giscancr

Choice of Lumpectomy with Radiation

Individual-level model: Logistic regression:

Generalized Linear Model

Choice of Lump/Rad (U) ~ f (AGE + RACE + HISTOLOGY + ESTROGEN RECEPTOR + SIZE OF TUMOR)

Coefficients:	Value	Std. Error	t value
(Intercept)	-2.30	0.28	-8.16
AGE	0.04	0.00	12.44
RACE	0.09	0.19	0.46
histology1	0.15	0.11	1.42
histology2	0.07	0.05	1.43
estrogen.receptor	0.05	0.05	1.13
size.of.tumor	0.05	0.00	11.04

All coefficients are significant at $p < .05$

- Note:The binary logit model is the simplest form of the logit choice model family and predicts the probability of choosing one of two alternatives (here lumprad over mastectomy). In preparing the data for a binary logit model estimation, you must decide on one of the two alternatives as the ‘important’ one and then create a field with an indicator of which of the two alternatives was chosen, by coding the value 1 for the important alternative and 0 for the other. The model has
- the form

- $U = a + b_1x_1 + b_2x_2 + \ldots$

Notes on Logistic Regression Results

The coefficients are used to compute for each individual the expected value of choice of lumpectomy with radiation.

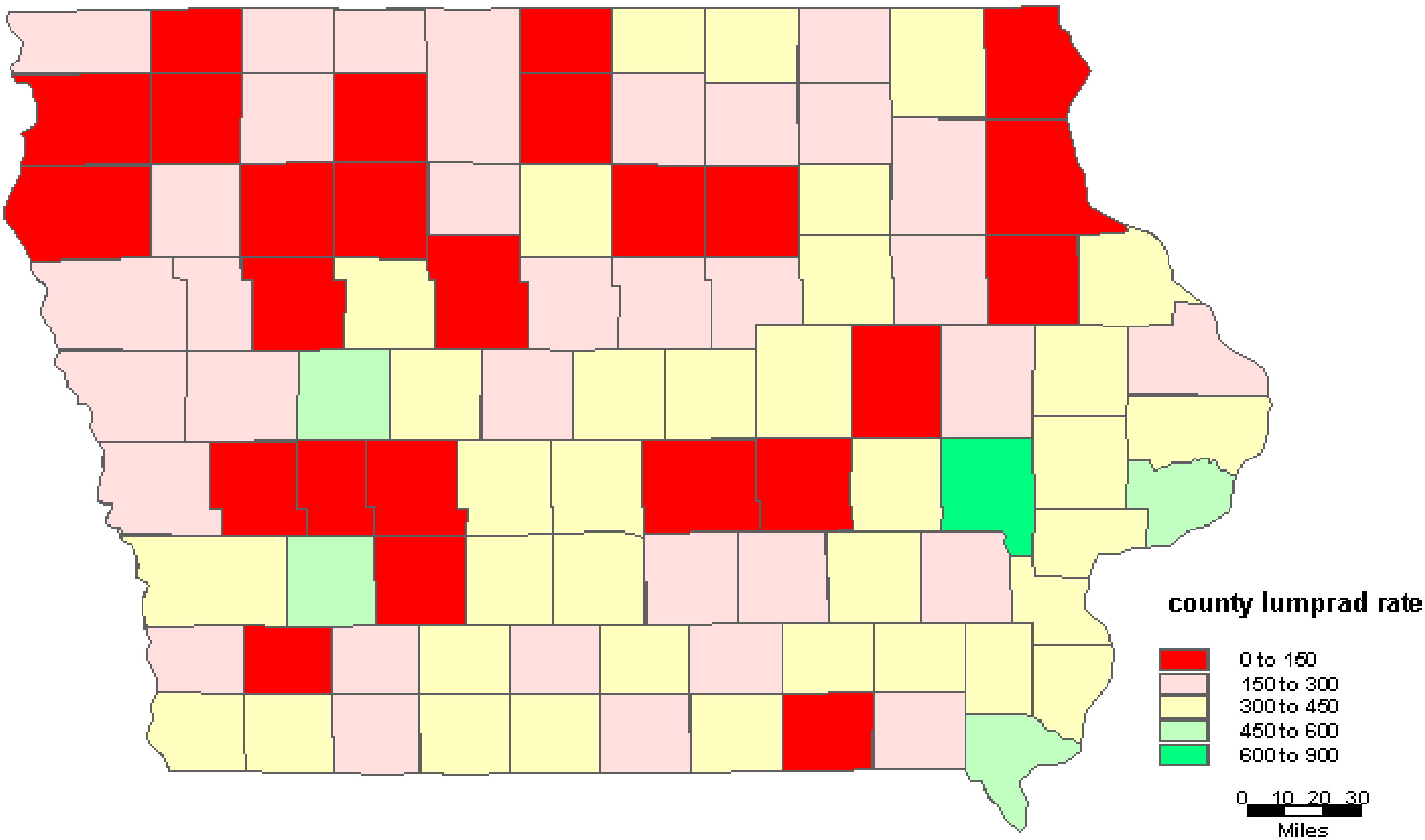
Choice of Lump/Rad (U) ~ *f* (AGE + RACE + HISTOLOGY + ESTROGEN RECEPTOR + SIZE OF TUMOR)

$$P = \frac{e^{U}}{(1+e^{U})}$$

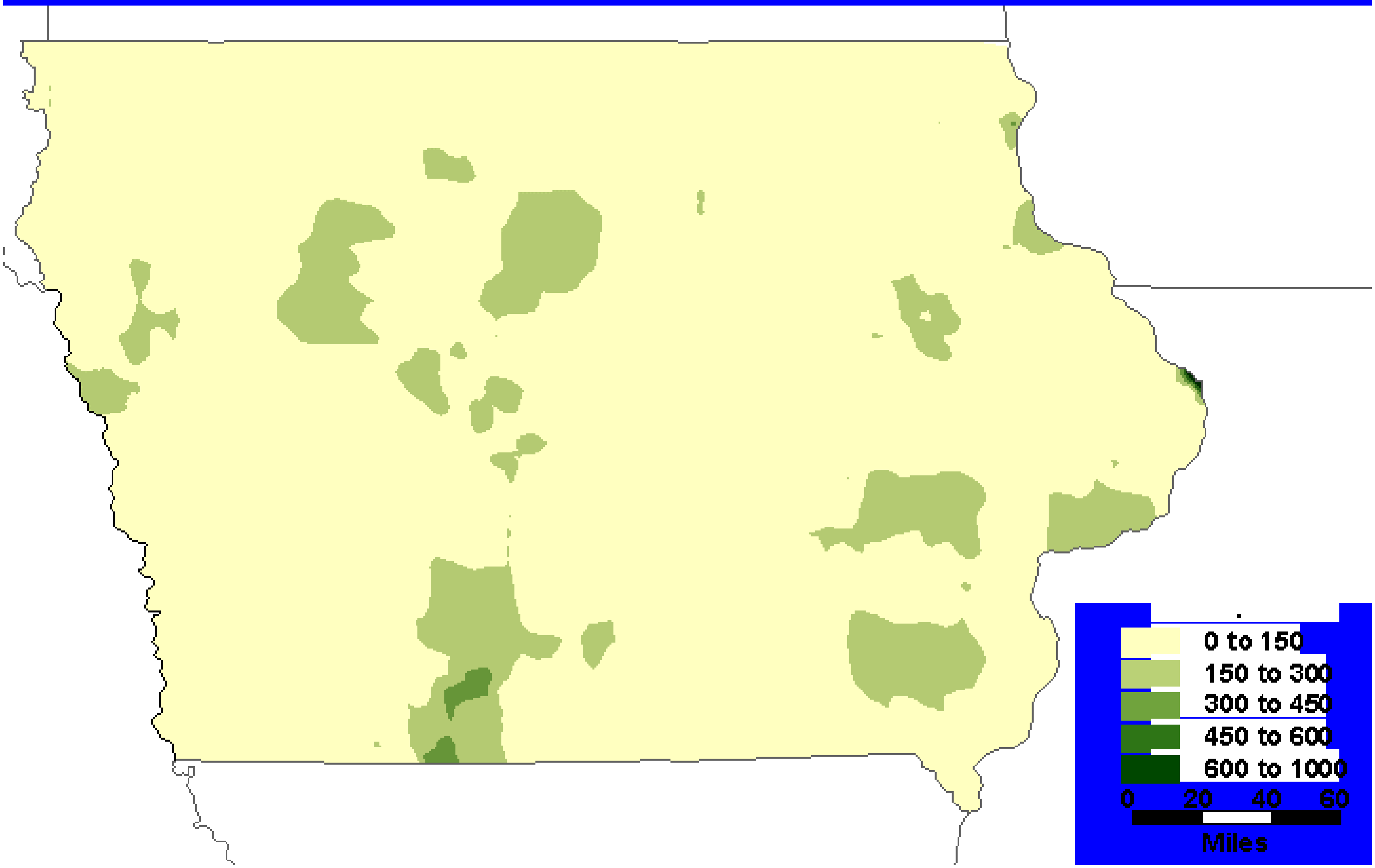
P is the probability of choosing the alternative coded as 1 and U is the relative utility of alternative 1. Note that if U is large (alternative 1 is much better) then the probability of choosing alternative 1 approaches 100 percent.

If U is very negative (alternative 1 is much worse), then the probability of choosing alternative 1 approaches 0 percent. Finally, if U = 0 (no difference between the alternatives) then the probability of choosing alternative 1 is 50 percent.

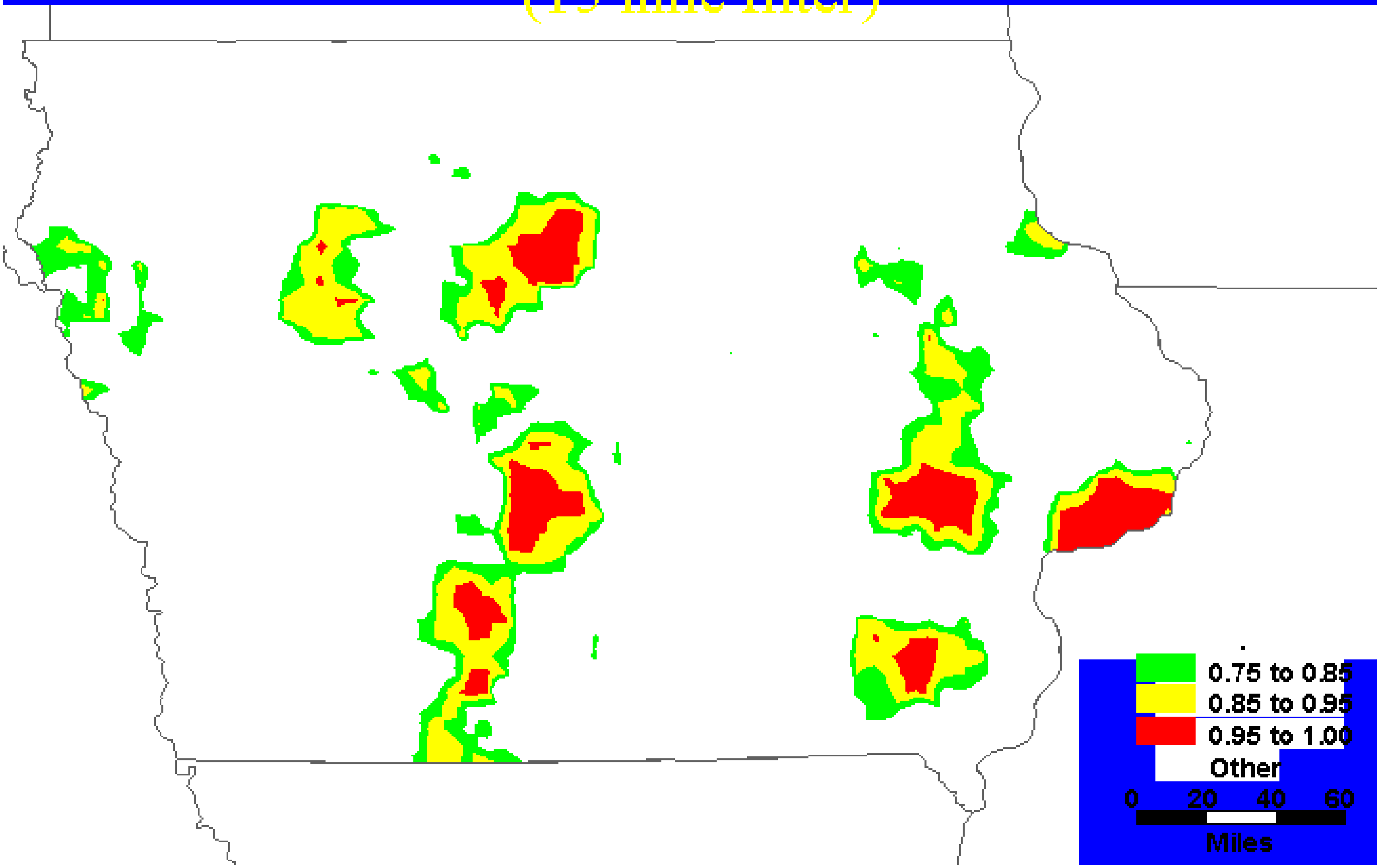
Predicted choice rate for lump/rad



Predicted choice rate for lump/rad (15mi filter)



Significance of Predicted Choice of lump/rad
(15 mile filter)



Now distance is added

Individual-level model: Logistic regression:

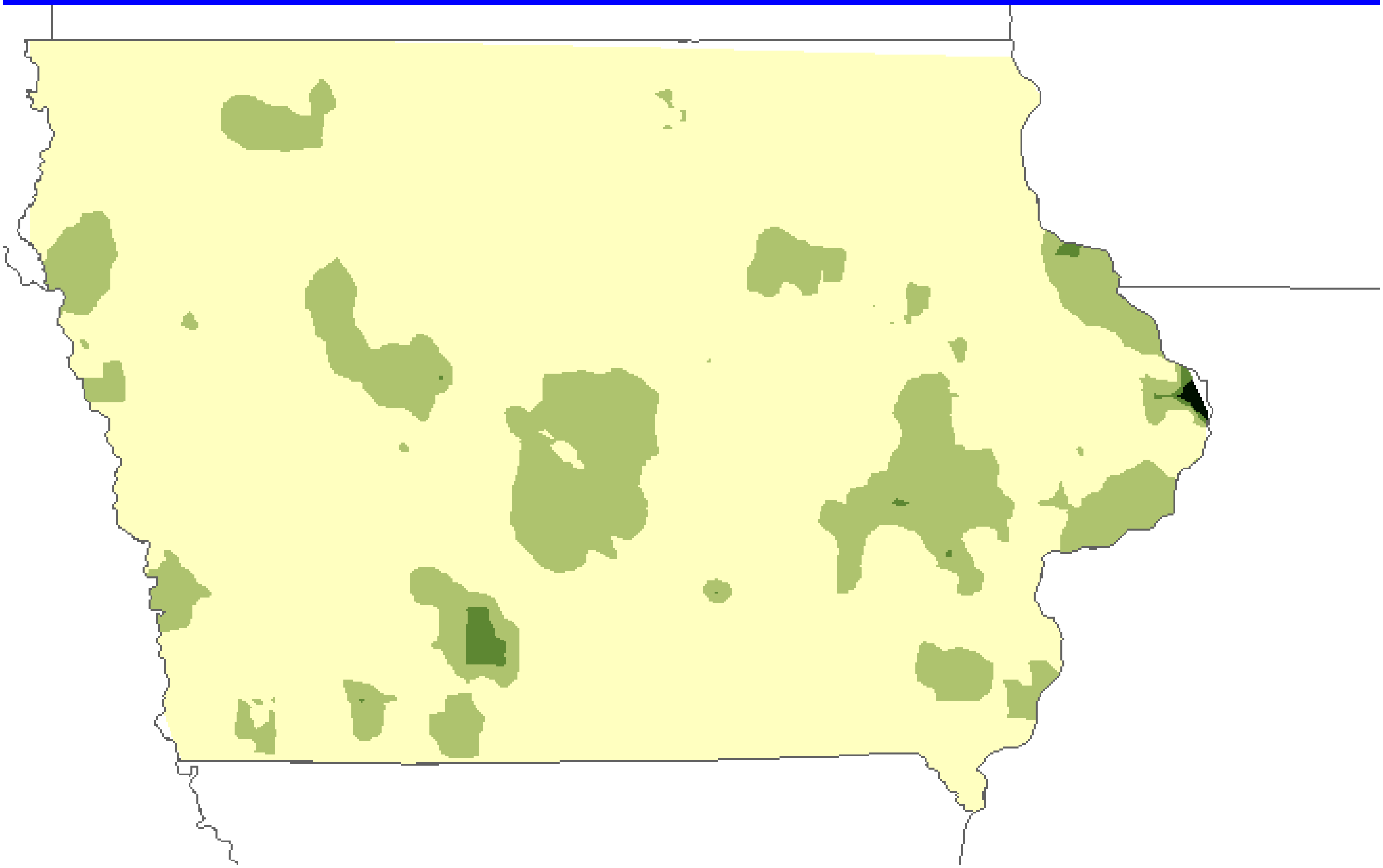
Generalized Linear Model

choice of treatment (U) ~ *f*(distance to nearest facility + age + grade of tumor + estrogen receptor + size of tumor)

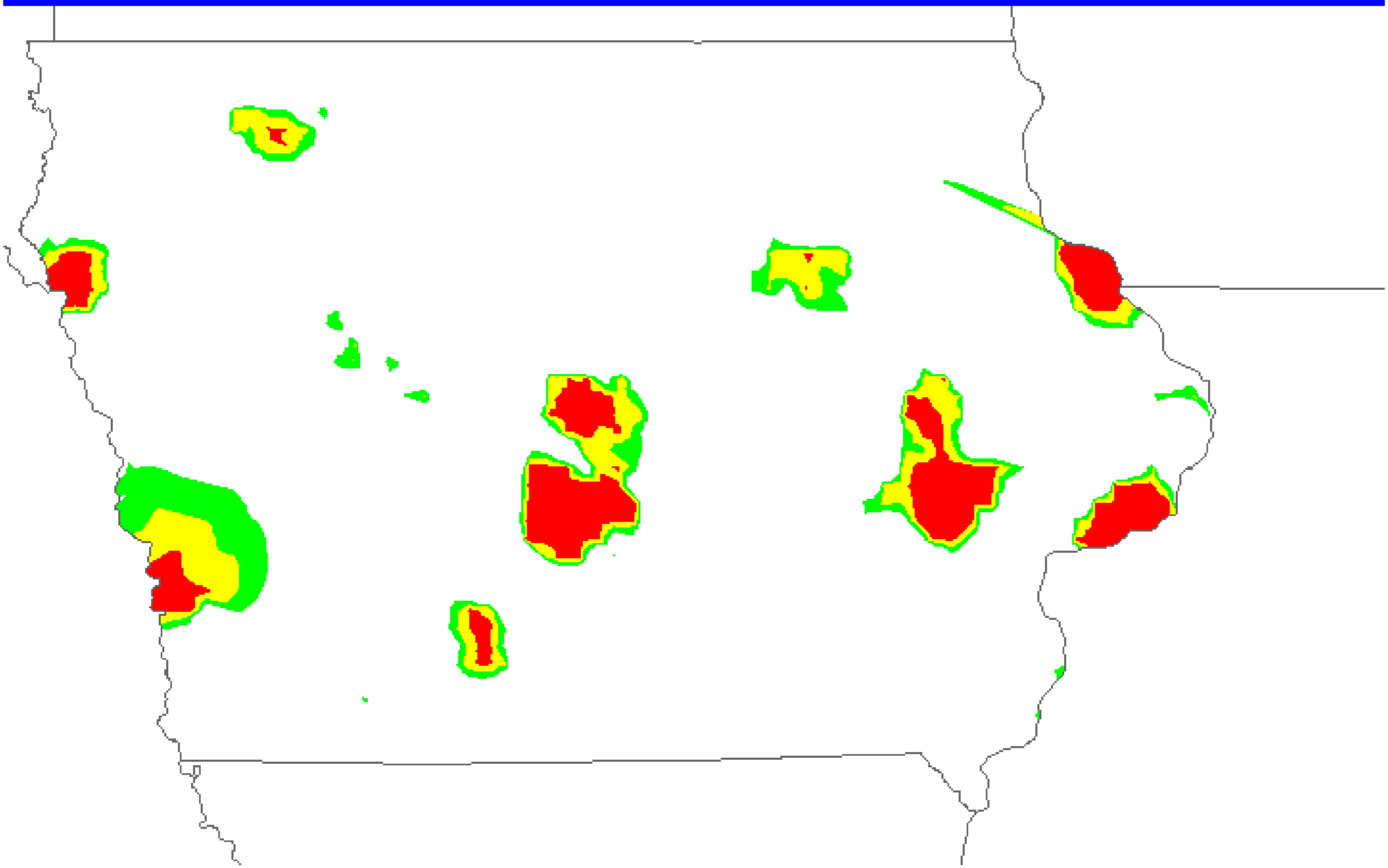
coefficients	Value	Std. Error	t value
(Intercept)	3.11	0.25	12.66
distance to nearest facility	-0.02	0.00	-6.90
age	-0.04	0.00	-12.31
estrogen receptor	-0.10	0.05	-2.04
size of tumor	-0.05	0.00	-10.48
grade of tumor	-0.20	0.05	-3.94

All coefficients are significant at $p < .05$

glm: lumprad rate w/distance (10mi filter)



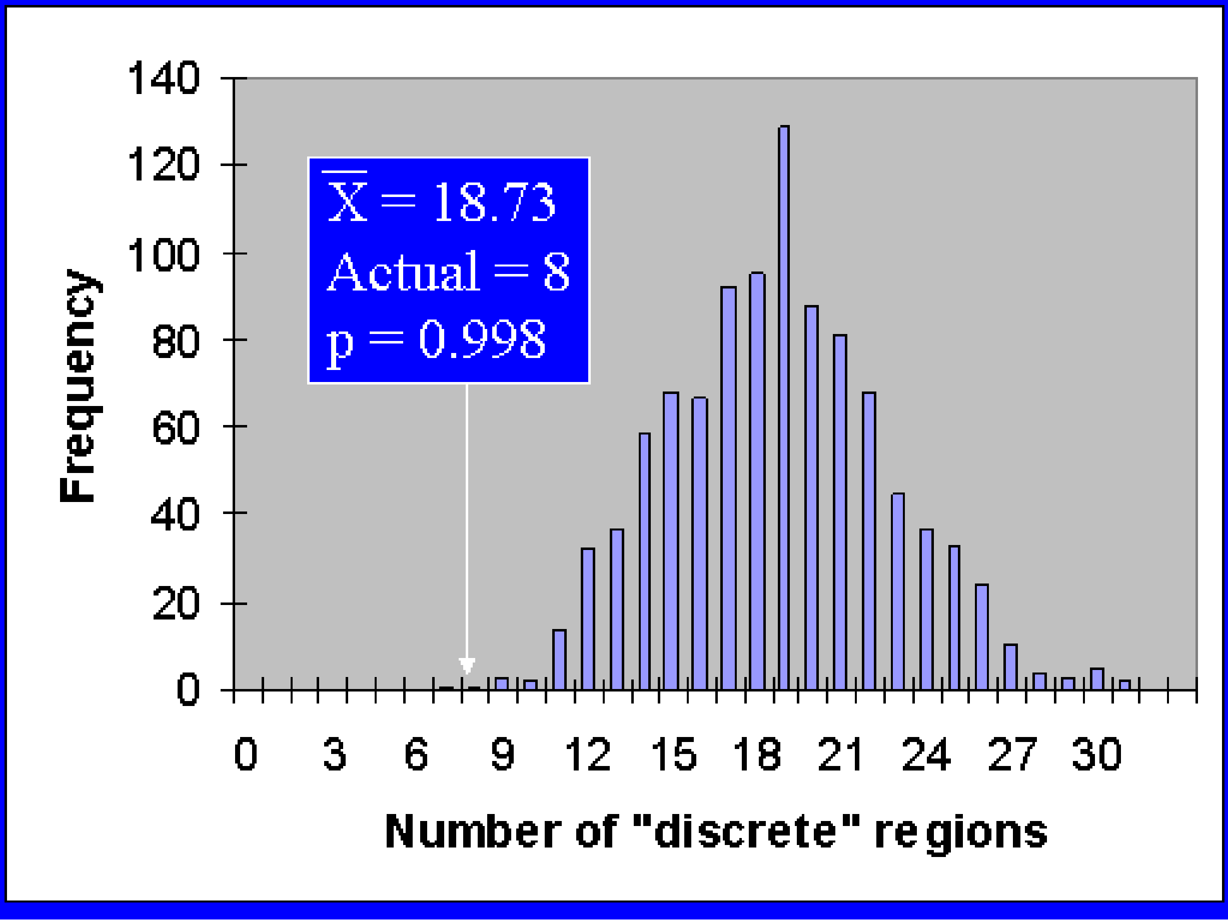
glm: lumprad significance w/ glm probability (10mi filter)

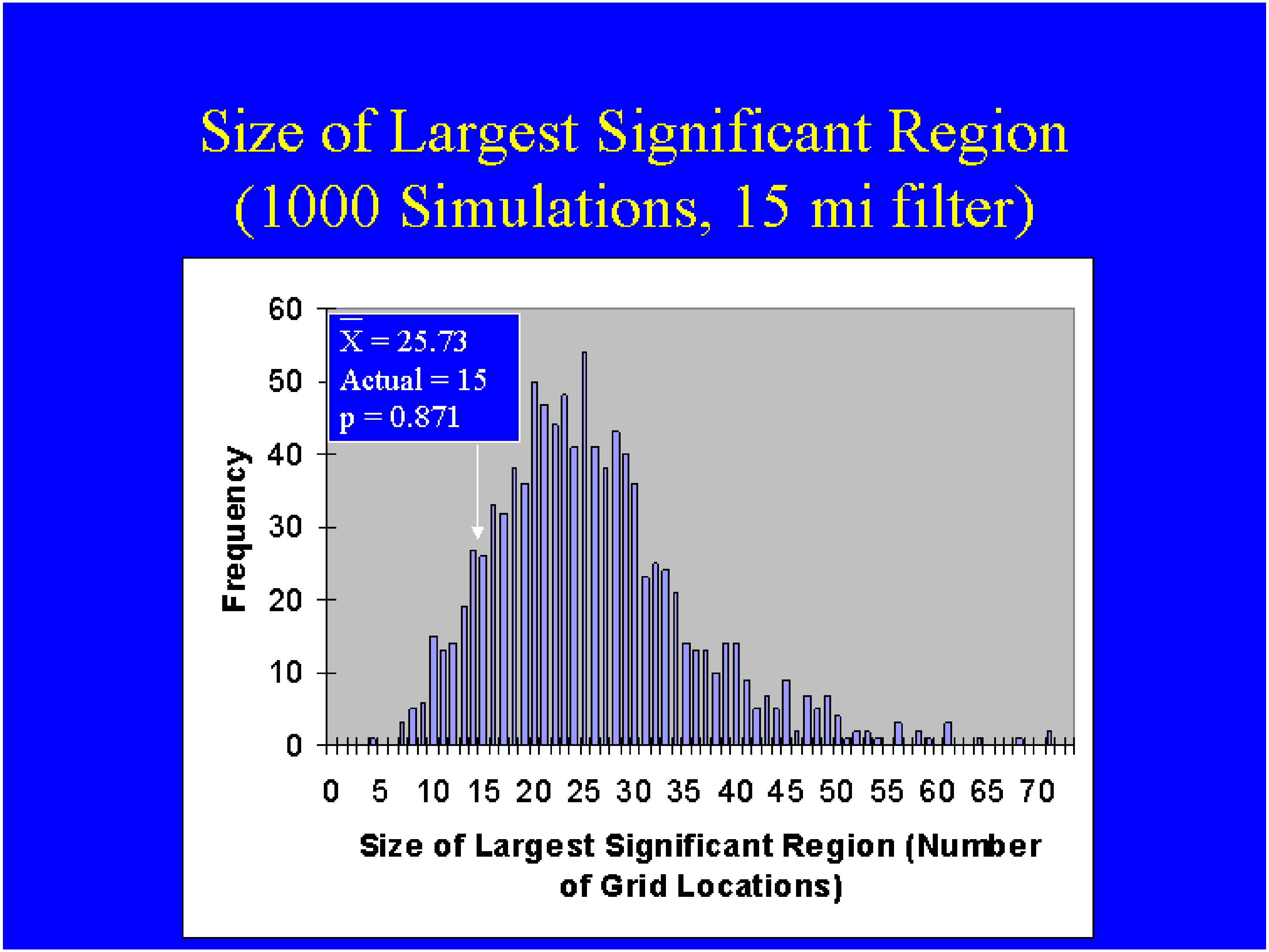


Even beginners should have a sense of where geographic information science is going

- One direction is “geographic feature identification”
- For example, if, because of “the small number problem” you can expect to see many “spatial clusters” of high cancer rates even when you have simulated the geographic pattern of cancer incidences, how many such “clusters” should YOU expect to see in YOUR area?
- We counted the number of discrete clusters in 1,000 simulated maps of breast cancer in Iowa… .

The observed number of clusters (defined as $p < .05$—not adjusted for multiple comparisons) was eight. 99.8 % of the simulated maps had more than 8 clusters, most considerably more.




Size of Largest Significant Region
(1000 Simulations, 15 mi filter)
X = 25.73
Actual = 15
p = 0.871
Frequency
60
50
40
30
20
10
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70
Size of Largest Significant Region (Number of Grid Locations)

Methods for making such GIS-based analyses are described on this web site

- www.uiowa.edu/~giscancr
- For the moment, the site is password protected but will soon be available to all.
- The site uses SIMULATED spatial data for breast cancer in Iowa for the web-based exercises, although it shows the results of the analyses on real Iowa Cancer Registry data.
- The exercises with data have been tested in a class at the University of Iowa but have not yet been tested in a distance education setting.

References on GIS and Health

Bailey, T. C. and A. C. Gatrell. 1995. *Interactive Spatial Data Analysis*. Longman, Harlow.

Briggs, D.J. and P. Elliott. 1995. "The use of geographical information systems in studies on environment and health." World Health Statistics Quarterly 48 85-94.

Cliff, A. D. and P. Haggett. 1996. "The Impact of GIS on Epidemiological Mapping and Modelling." Ch. 17 in P. Longley and M. Batty, *Spatial Analysis: Modelling in a GIS Environment*. Cambridge, Geoinformation International, pp. 321-343.

References

- Gelman, A., P.N. Price and C-Y Lin. 2000. “A method for quantifying artefacts in mapping methods illustrated by application to headbanging.” Statistics in Medicine 19:2309-2320.

Acknowledgments

- Research assistants at the university of Iowa: Matthew Airola, Geoffrey Smith, Chris Poulliot and Aniruddha Banerjee.
- Co-investigators for NCI supported projects: Michele west, Ph.D. And Ika Peleg, MD.
- The Breast Cancer Project was supported by NCI Contract No. N01-PC-67008
- Iowa Cancer (SEER) Registry.

Questions?

This presentation will soon be available at

www.uiowa.edu/~giscancr