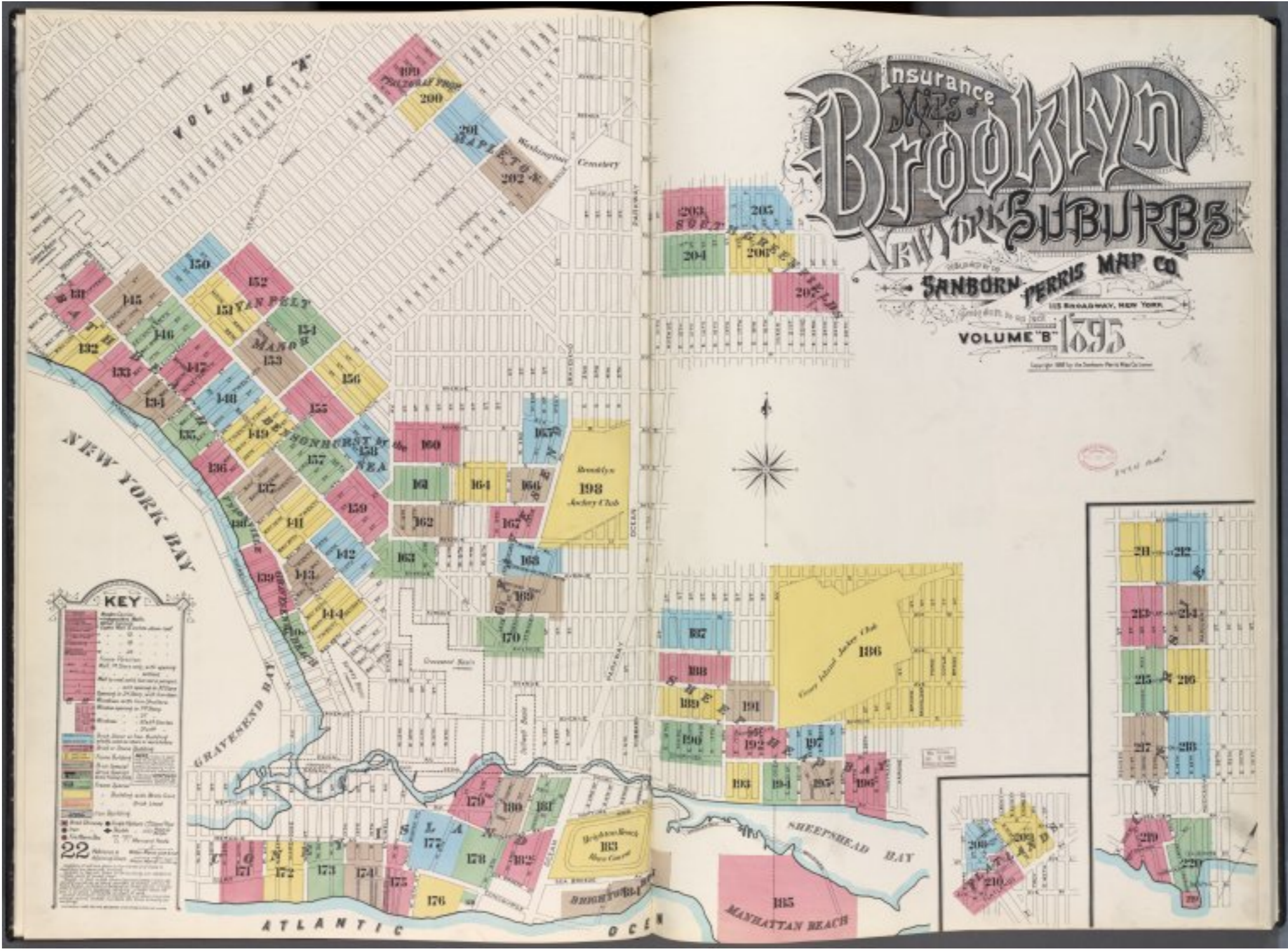


historical map polygon and feature extractor

mauricio giraldo arteaga
NYPL Labs
@mgiraldo

NYGeoCon 2013

background

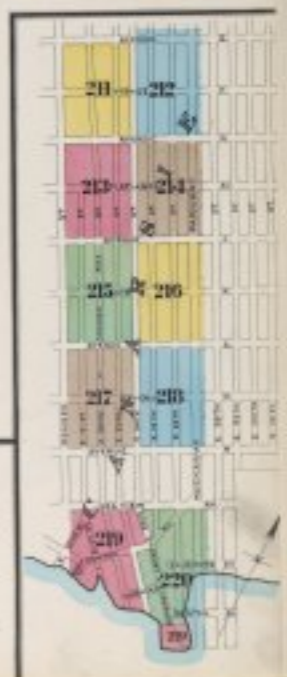


KEY

Lot numbers are shown in black figures. Street names are shown in red letters. The number 22 is shown in a red circle.

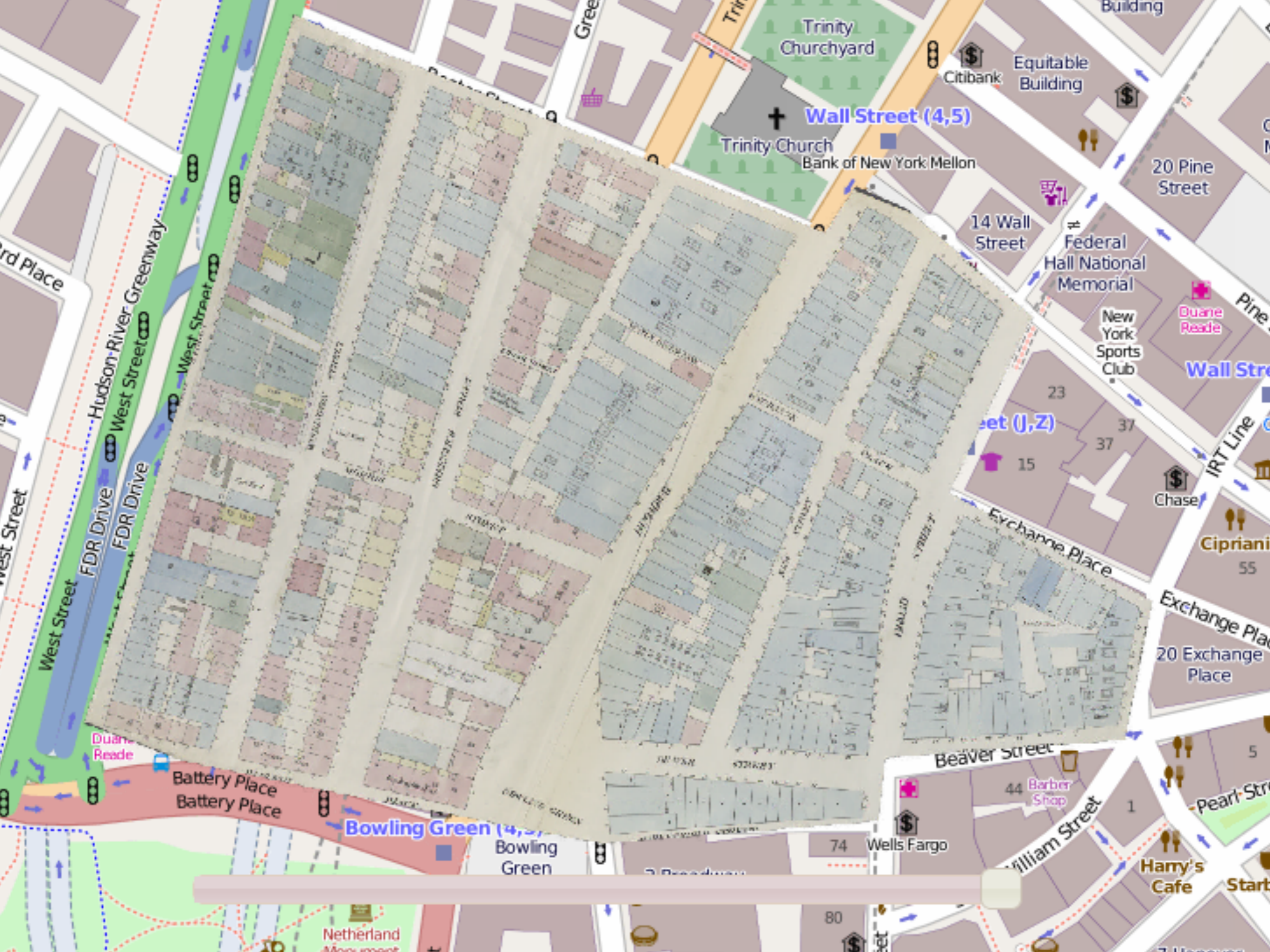
Other symbols include a red square for a lot, a red line for a street, and a red circle for a lot number.

Insurance
Map of Brooklyn
 NEW YORK SUBURBS
 SANBORN PERRIS MAP CO.
 115 BROADWAY, NEW YORK
 VOLUME "B" 1895









Trinity Churchyard

Wall Street (4,5)

Trinity Church
Bank of New York Mellon

Equitable Building

Citibank

20 Pine Street

14 Wall Street

Federal Hall National Memorial

New York Sports Club

Wall Street

Street (J,Z)

IRT Line

Chase

Cipriani

Exchange Place

Exchange Place

20 Exchange Place

Beaver Street

44 Barber Shop

74 Wells Fargo

William Street

Harry's Cafe

Start

Bowling Green (4,5)

Bowling Green

Battery Place
Battery Place

Netherland Monument



WASHINGTON STREET

STREET

WICH STREET

STREET

EDGAR STREET

Potash

Inspection

Bonded Warehouse

Coal Yard

United States
Bonded Warehouse

Adams Express

16

18

40

42

44

46

48

50

52

54

56

58

60

62

64

66

68

70

72

74

76

78

80

82

84

86

88

90

92

94

96

98

100

16

18

40

42

44

46

48

50

52

54

56

58

60

62

64

66

68

70

72

74

76

78

80

82

84

86

88

90

92

94

96

98

100

16

18

40

42

44

46

48

50

52

54

56

58

60

62

64

66

68

70

72

74

76

78

80

82

84

86

88

90

92

94

96

98

100

16

18

40

42

44

46

48

50

52

54

56

58

60

62

64

66

68

70

72

74

76

78

80

82

84

86

88

90

92

94

96

98

100

16

18

40

42

44

46

48

50

52

54

56

58

60

62

64

66

68

70

72

74

76

78

80

82

84

86

88

90

92

94

96

98

100

16

18

40

42

44

46

48

50

52

54

56

58

60

62

64

66

68

70

72

74

76

78

80

82

84

86

88

90

92

94

96

98

100

16

18

40

42

44

46

48

50

52

54

56

58

60

62

64

66

68

70

72

74

76

78

80

82

84

86

88

90

92

94

96

98

100

16

18

40

42

44

46

48

50

52

54

56

58

60

62

64

66

68

70

72

74

76

78

80

82

84

86

88

90

92

94

96

98

100

16

18

40

42

44

46

48

50

52

54

56

58

60

62

64

66

68

70

72

74

76

78

80

82

84

86

88

90

92

94

96

98

100

16

18

40

42

44

46

48

50

52

54

56

58

60

62

64

66

68

70

72

74

76

78

80

82

84

86

88

90

92

94

96

98

100

16

18

40

42

44

46

48

50

5



WASHINGTON

Coal Yard

Coal

M

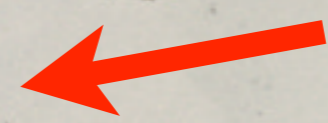
24 23 26 24 27 25 23 21 19 18 16 40 42 44 46 48 49 41 38 36 34 32 30 28 29 31 33 35 35 1/2 37 39



WASHINGTON

Coal Yard

Coal



26

24

22

20

18

16

40

42

44

46

48

49

41

38

36

34

32

30

28

39

37

35

33

31

29

24

23

27

25

23

21

19

35 1/2

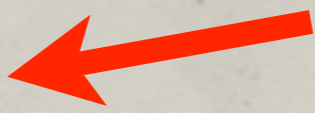
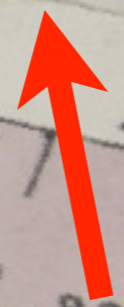
M



WASHINGTON

Coal Yard

Coal



26

24

22

20

18

16

40

42

44

46

48

49

41

38

36

34

32

30

28

39

37

35

33

31

29

24

23

27

25

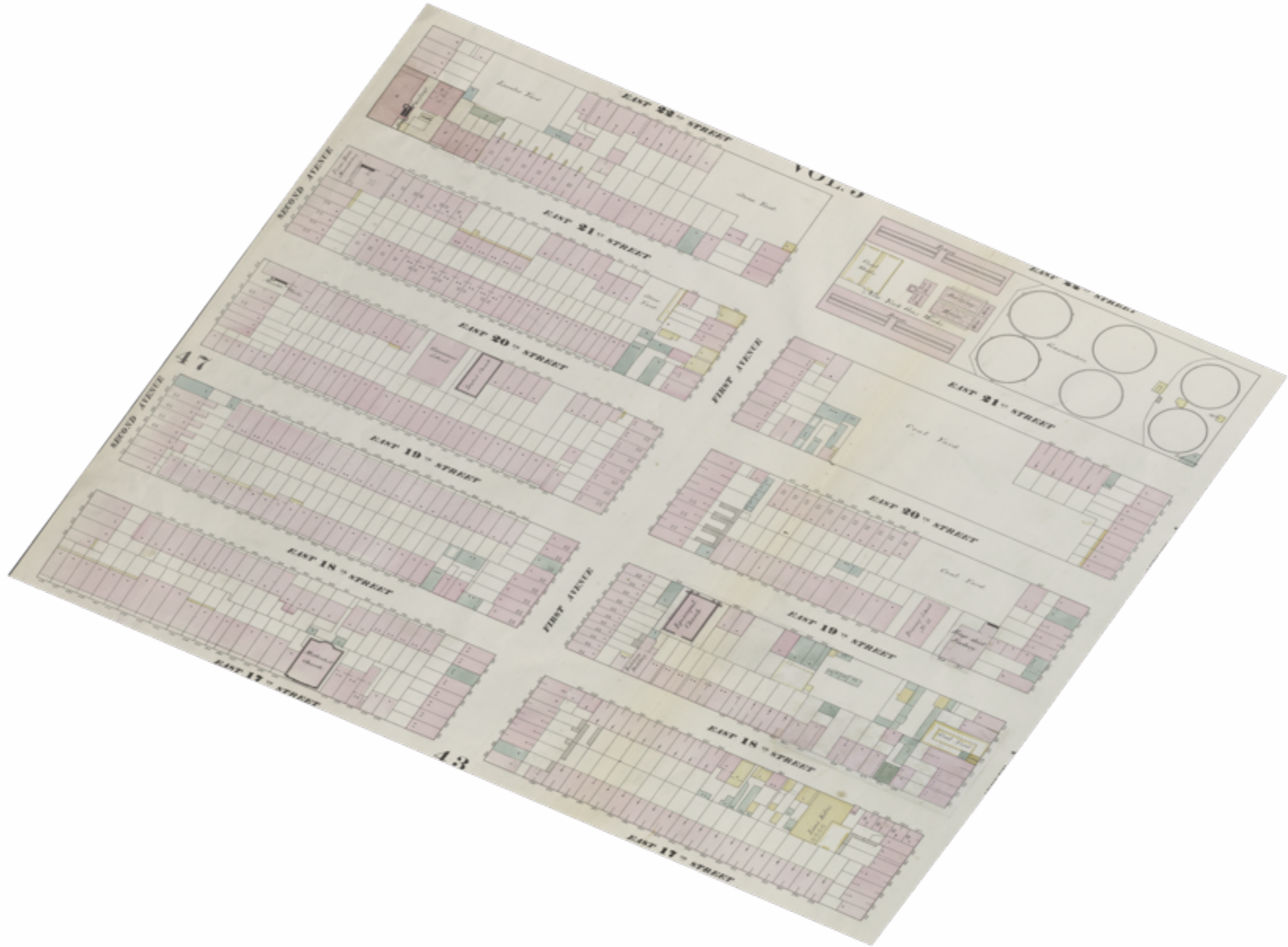
23

21

19

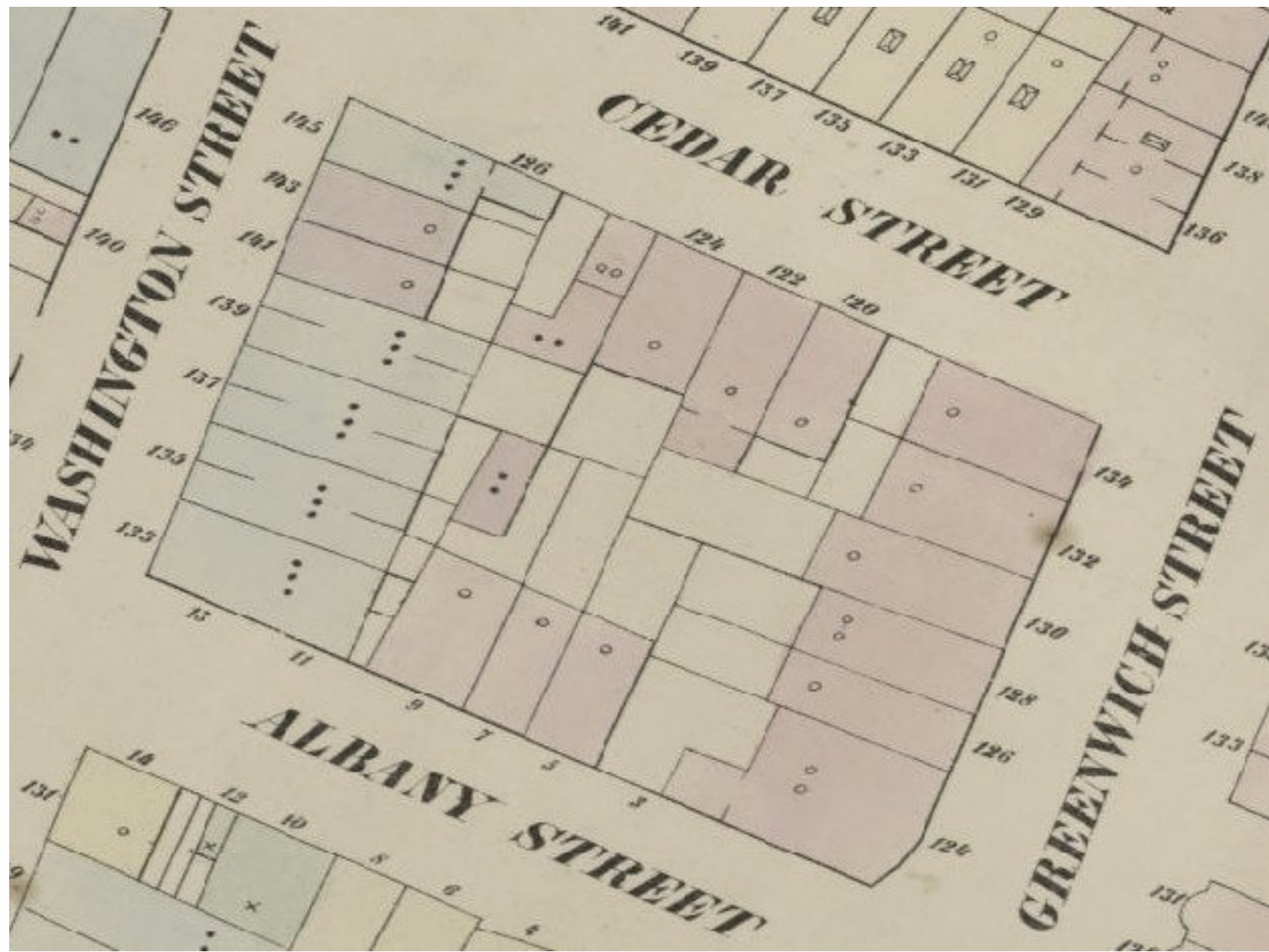
35 1/2

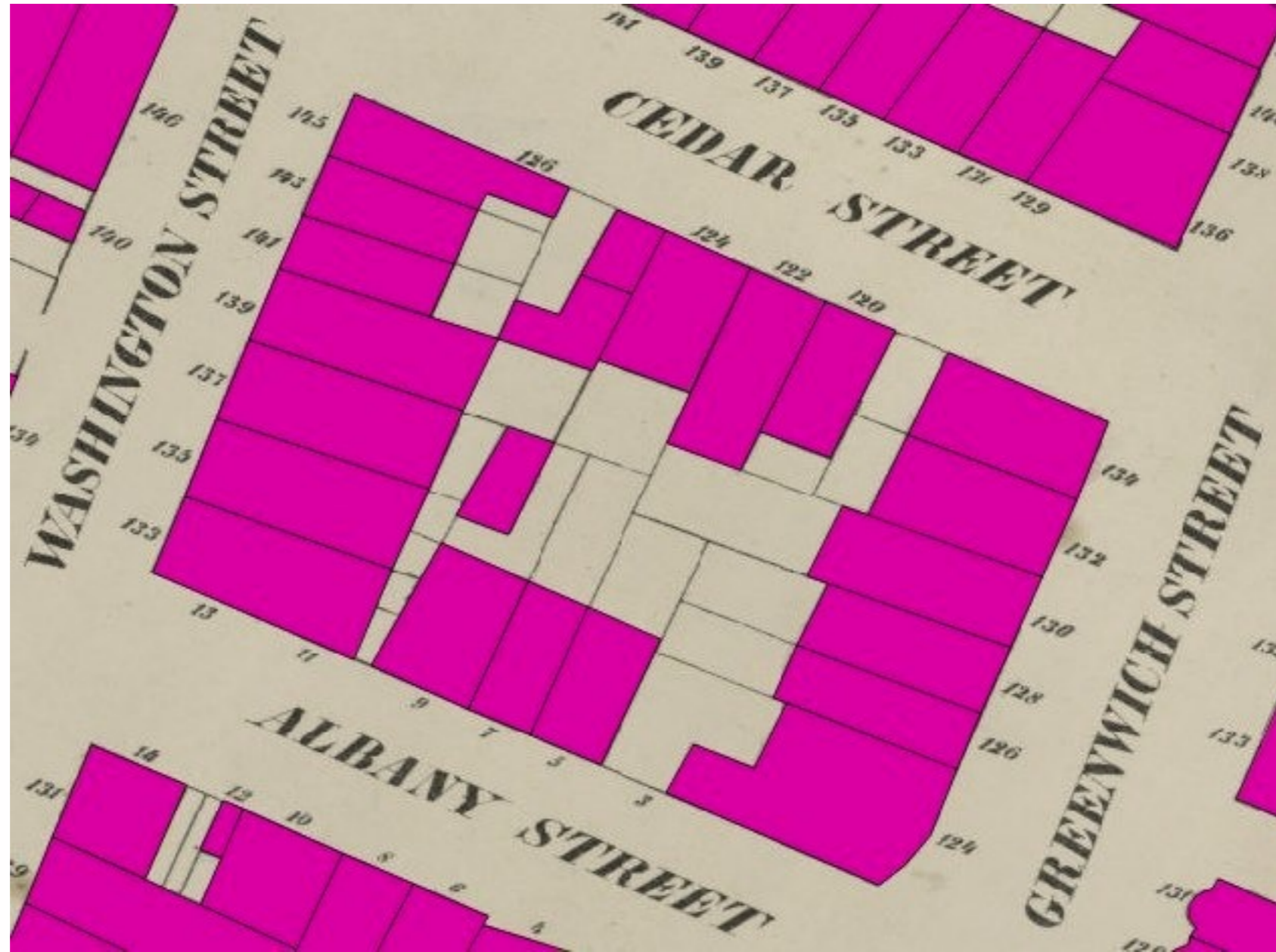




47

43





~120k polygons produced in **three years**
by staff and volunteers

(NYPL  volunteers)

building =



building =

not paper-colored



building =

not paper-colored

completely enclosed by black lines



building =

not paper-colored

completely enclosed by black lines

dashed lines are not walls



building =



not paper-colored

completely enclosed by black lines

dashed lines are not walls

> 20m² (~180ft²)

building =



not paper-colored

completely enclosed by black lines

dashed lines are not walls

> 20m² (~180ft²)

< 3,000m² (~27,000ft²)

building =



not paper-colored

completely enclosed by black lines

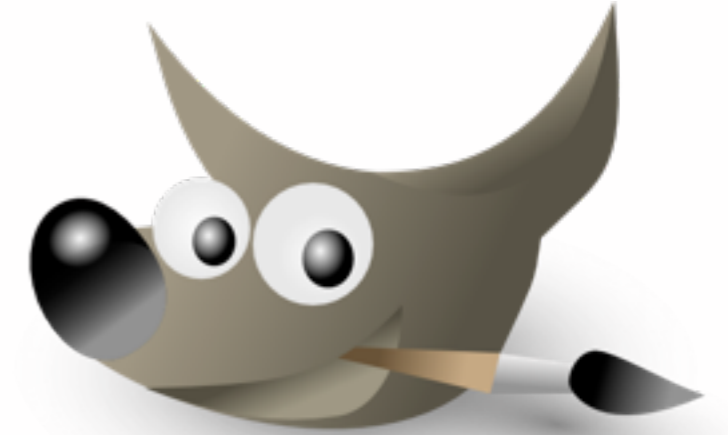
dashed lines are not walls

> 20m² (~180ft²)

< 3,000m² (~27,000ft²)

+ attributes (color, dots, crosses...)

process





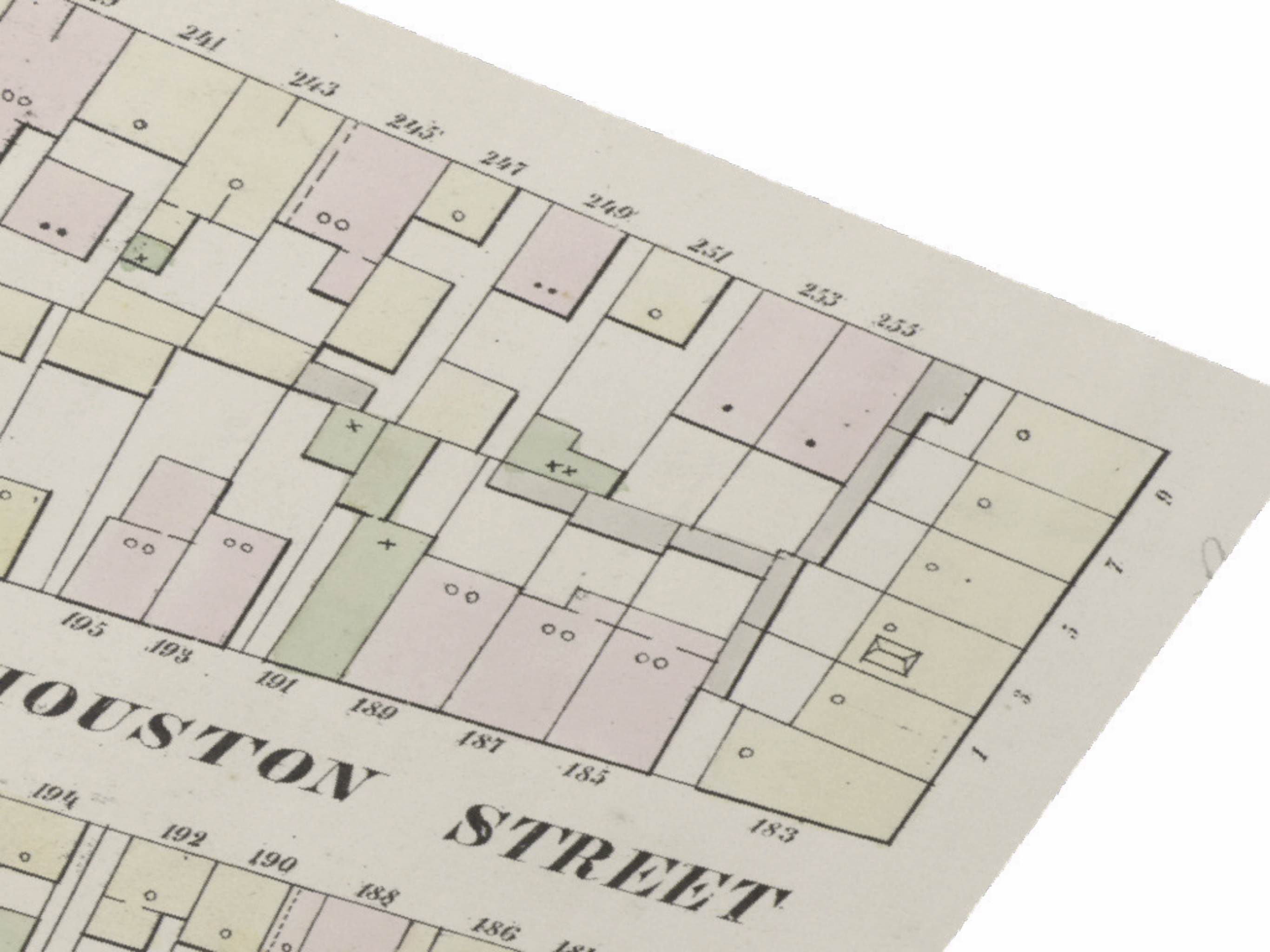
<https://github.com/NYPL/map-vectorizer>

try it!



`gdal_polygonize.py`

generates polygons automagically!



241

243

245

247

249

251

253

255

195

193

191

189

187

185

183

HOUSTON

STREET

194

192

190

188

186

184

1
3
5
7
9

```
$ gdal_polygonize.py test.tif -f "ESRI Shapefile" test.shp test
```



```
$ gdal_polygonize.py test.tif -f "ESRI Shapefile" test.shp test
```



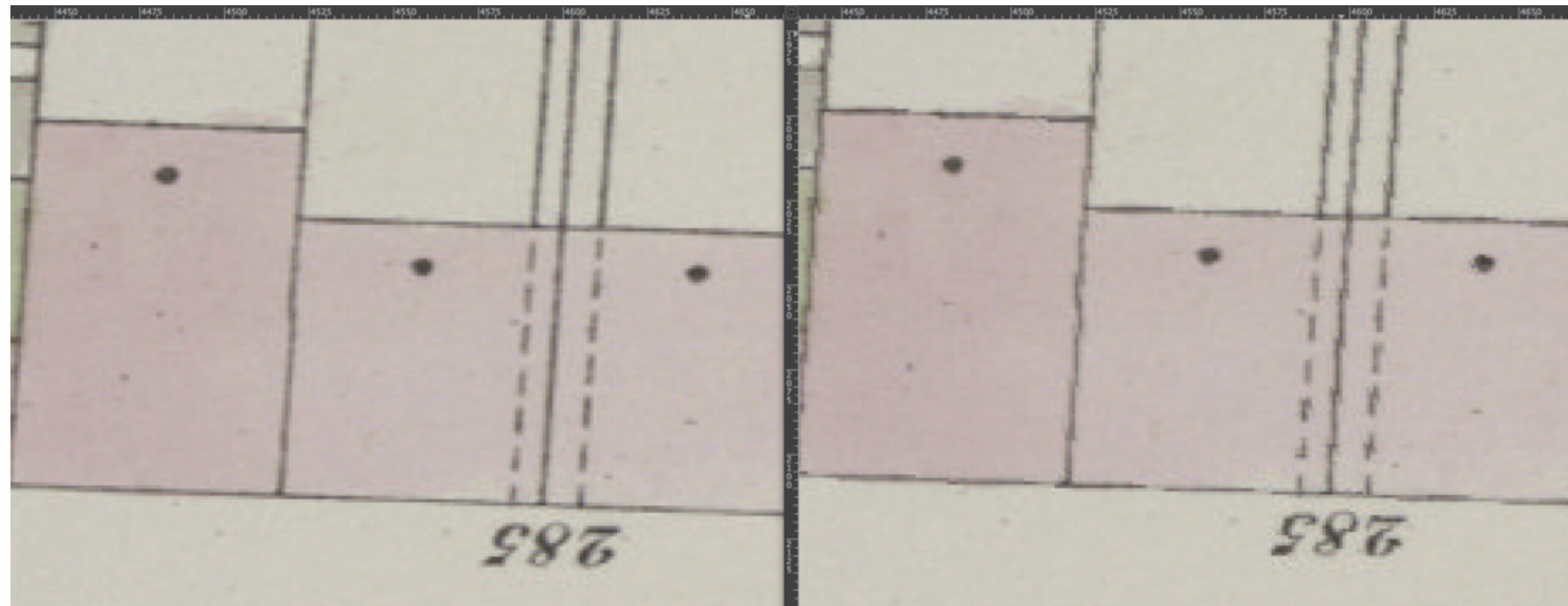
gdal_polygonize.py

generates polygons automagically!

(not really)

we need to **optimize** the **input**

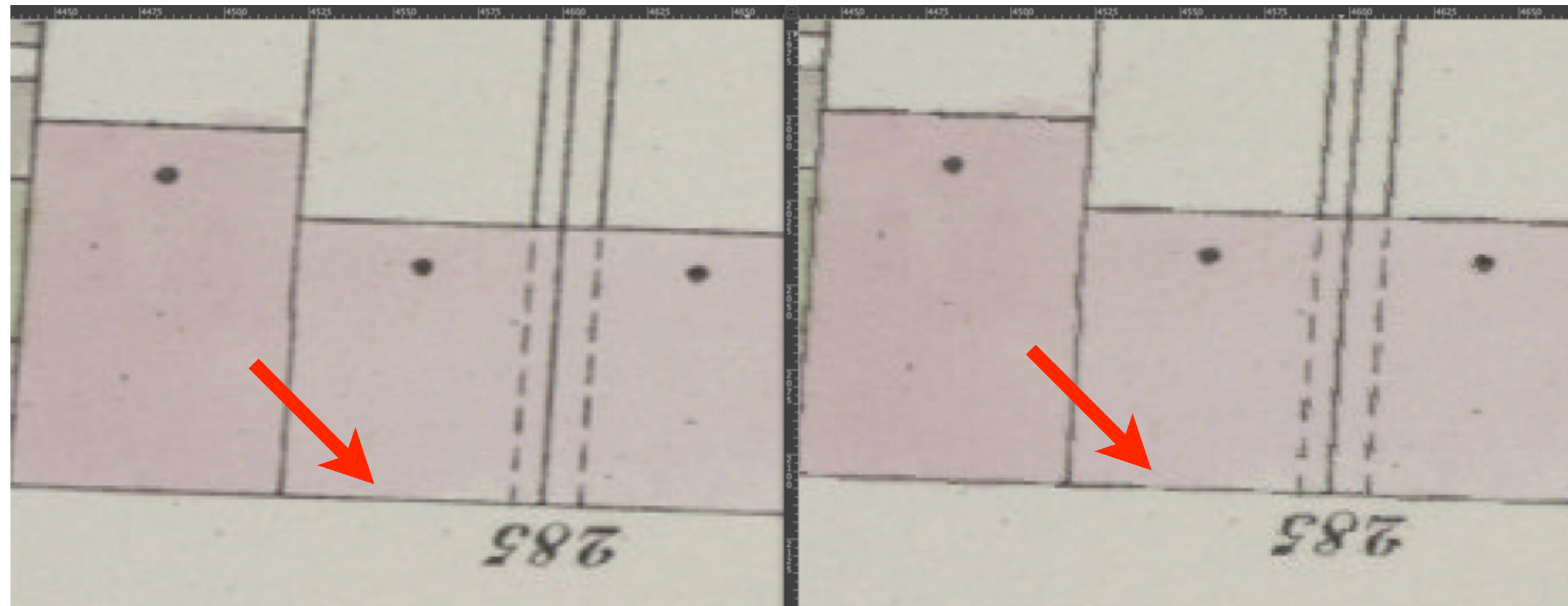
differences in resampling



cubic

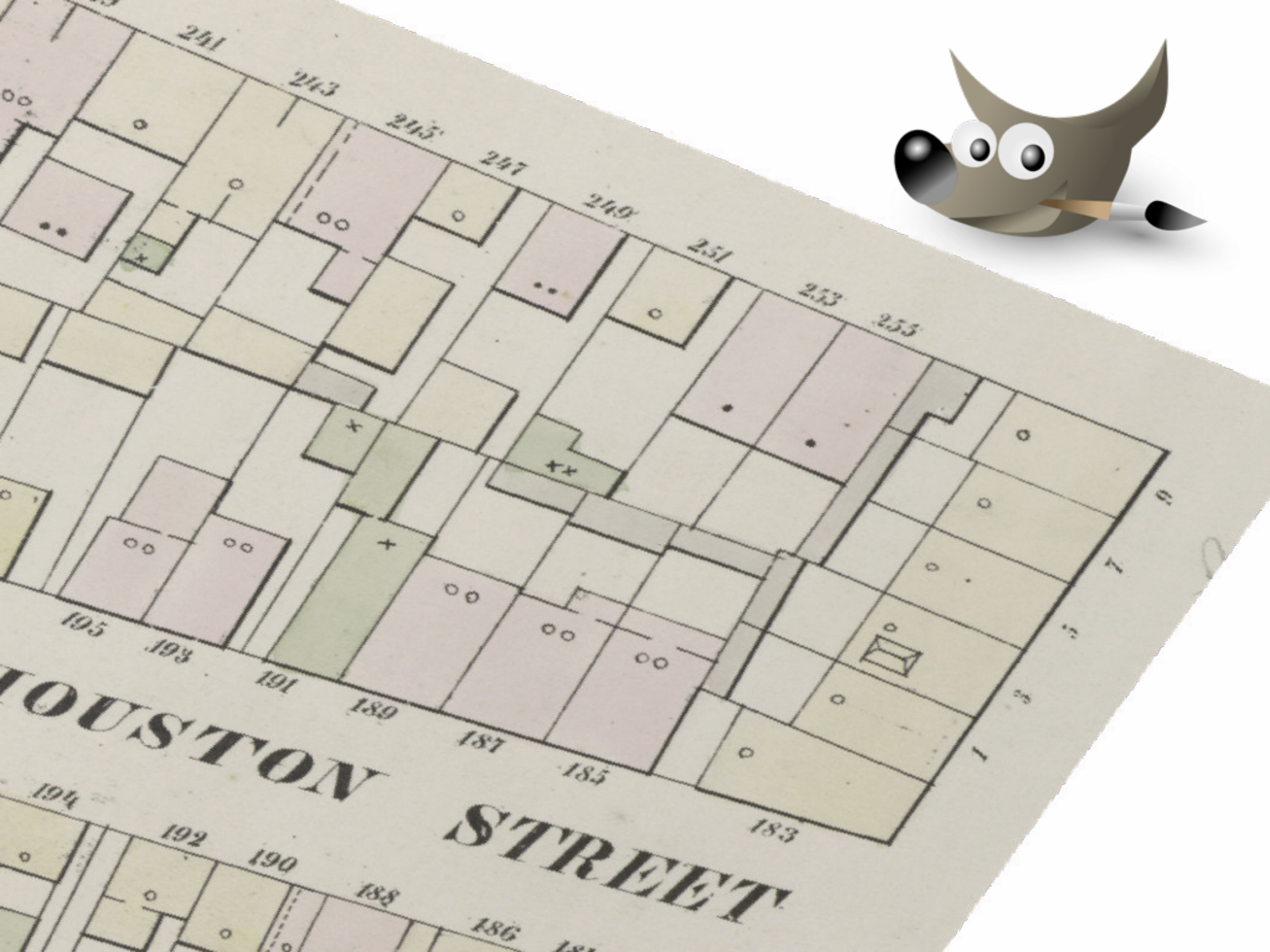
nearest neighbor

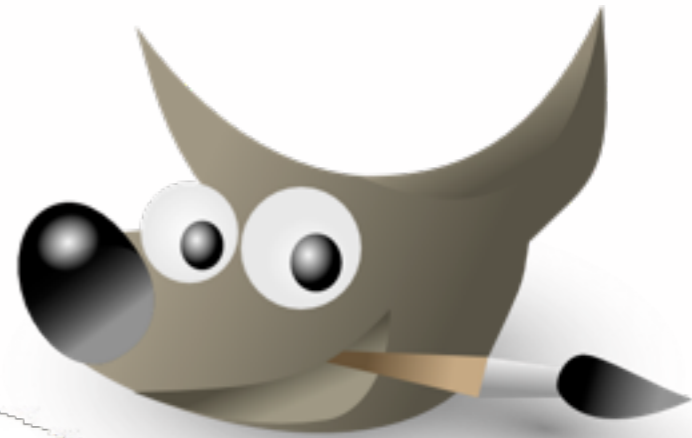
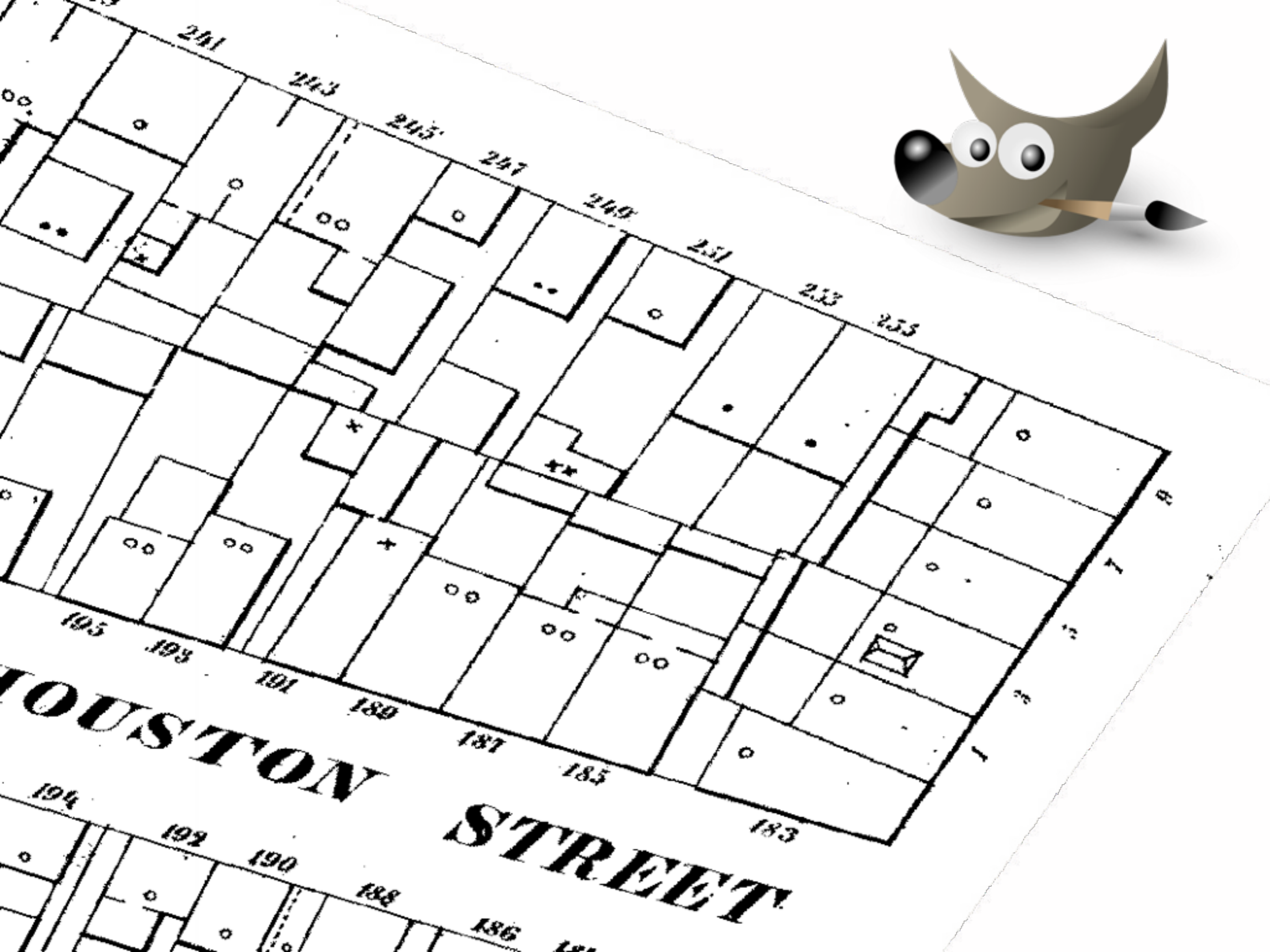
differences in resampling

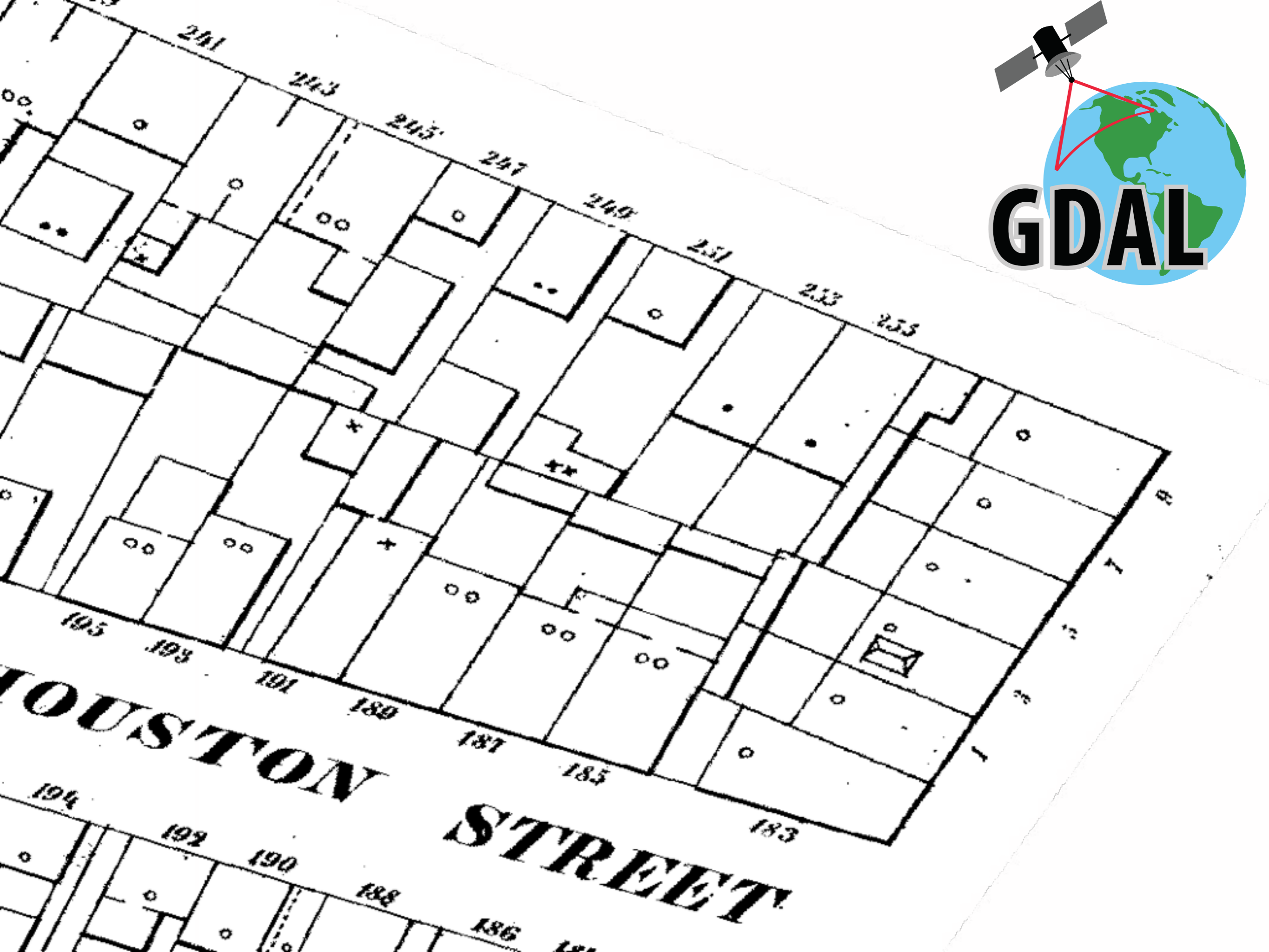


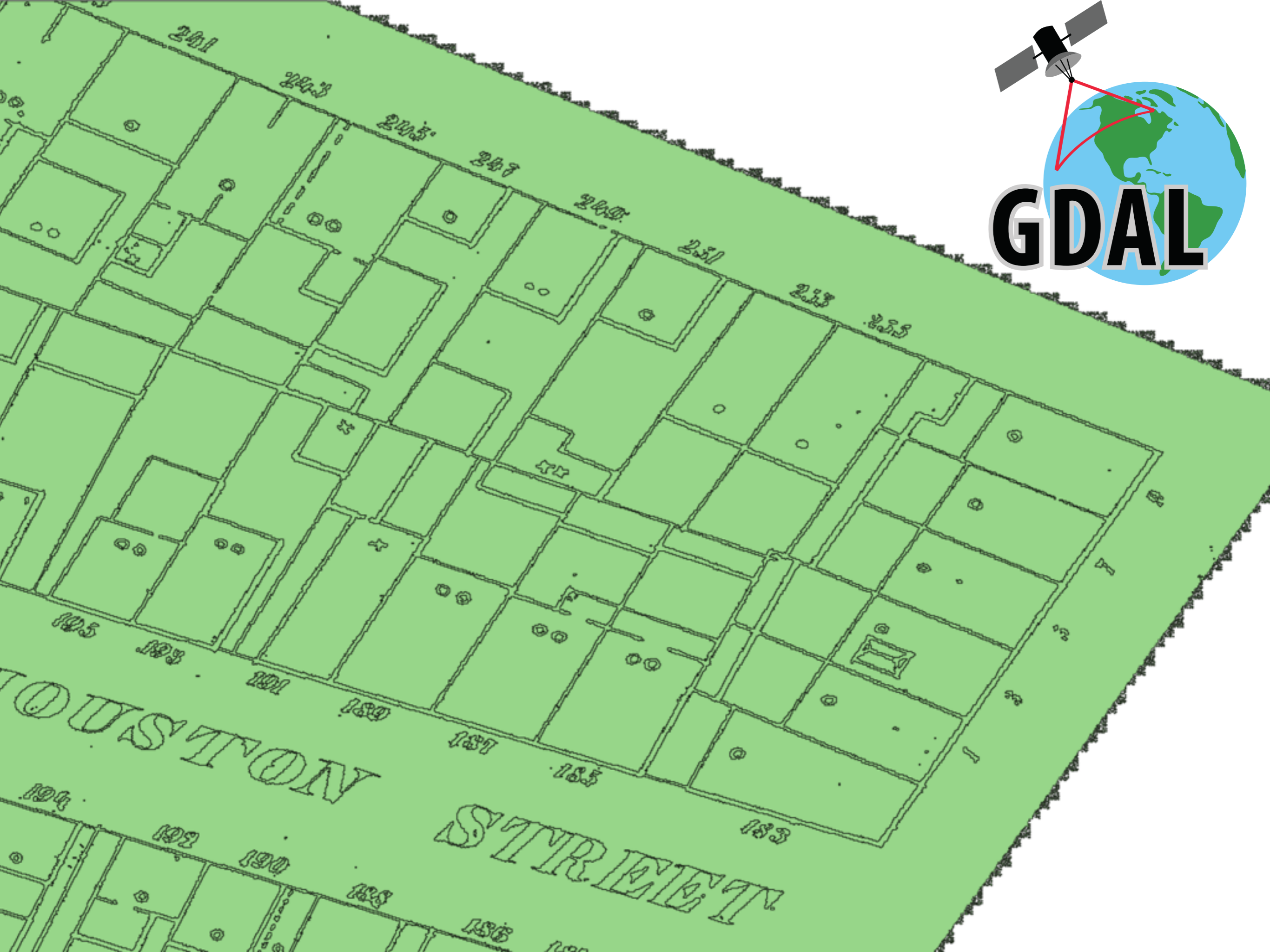
cubic

nearest neighbor







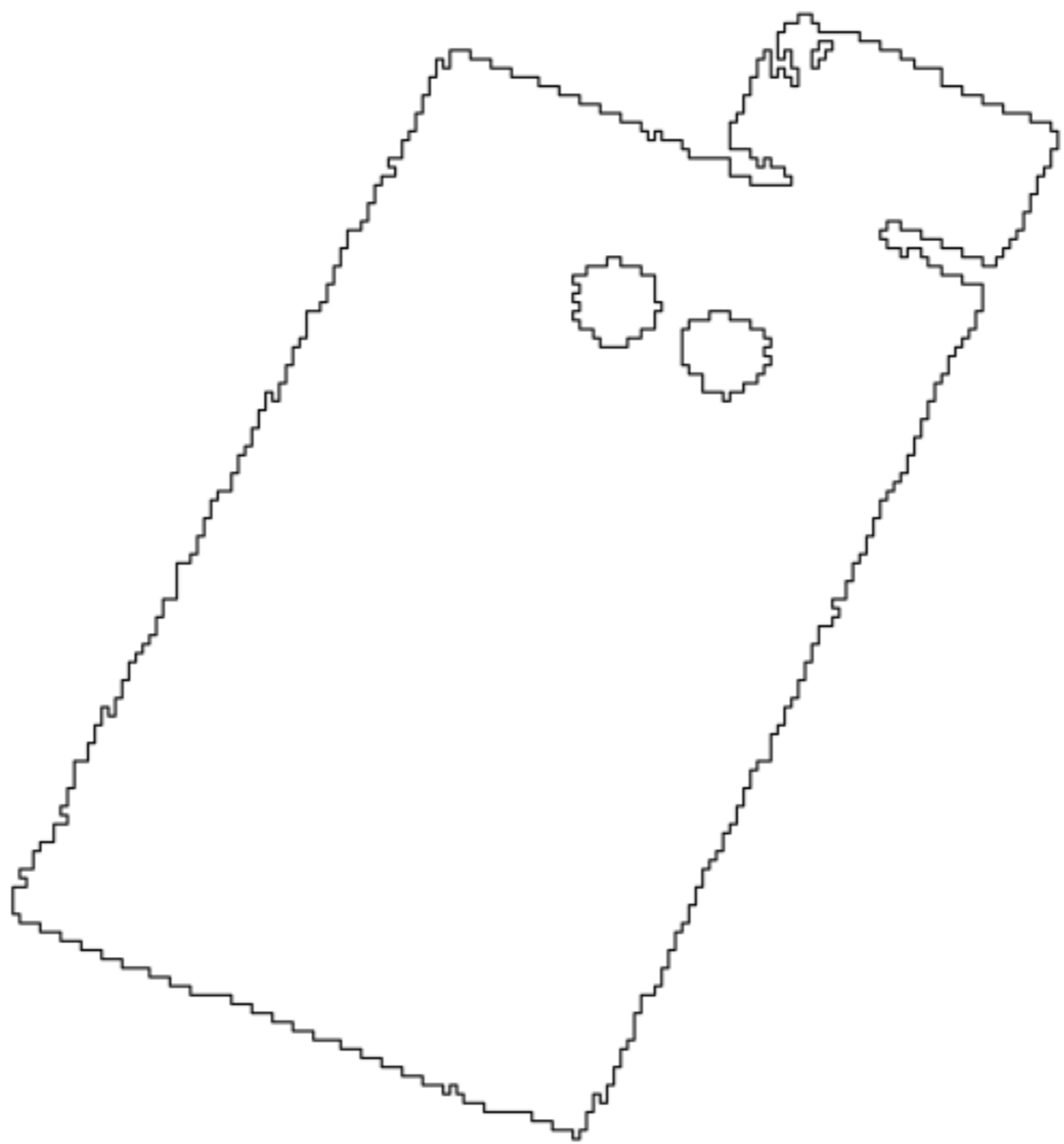




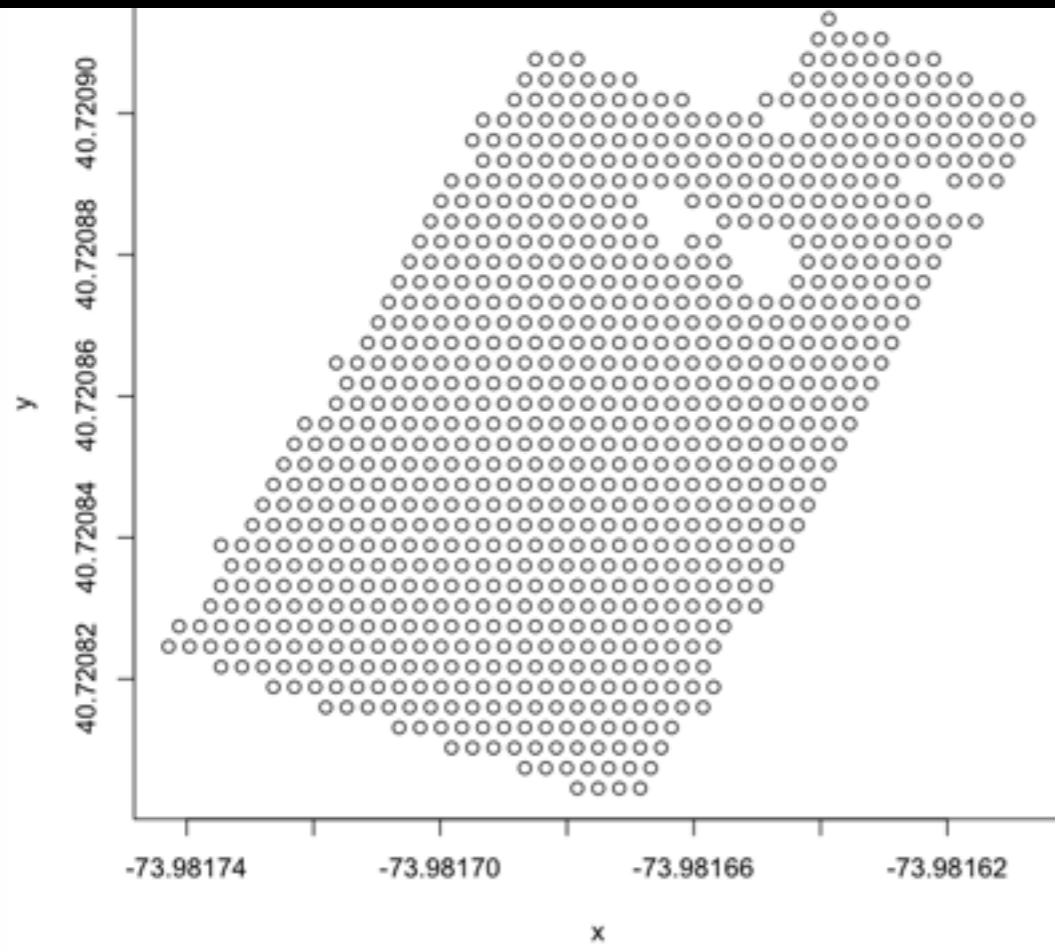
we need to **simplify** the **output**

(for those polygons that we care about)

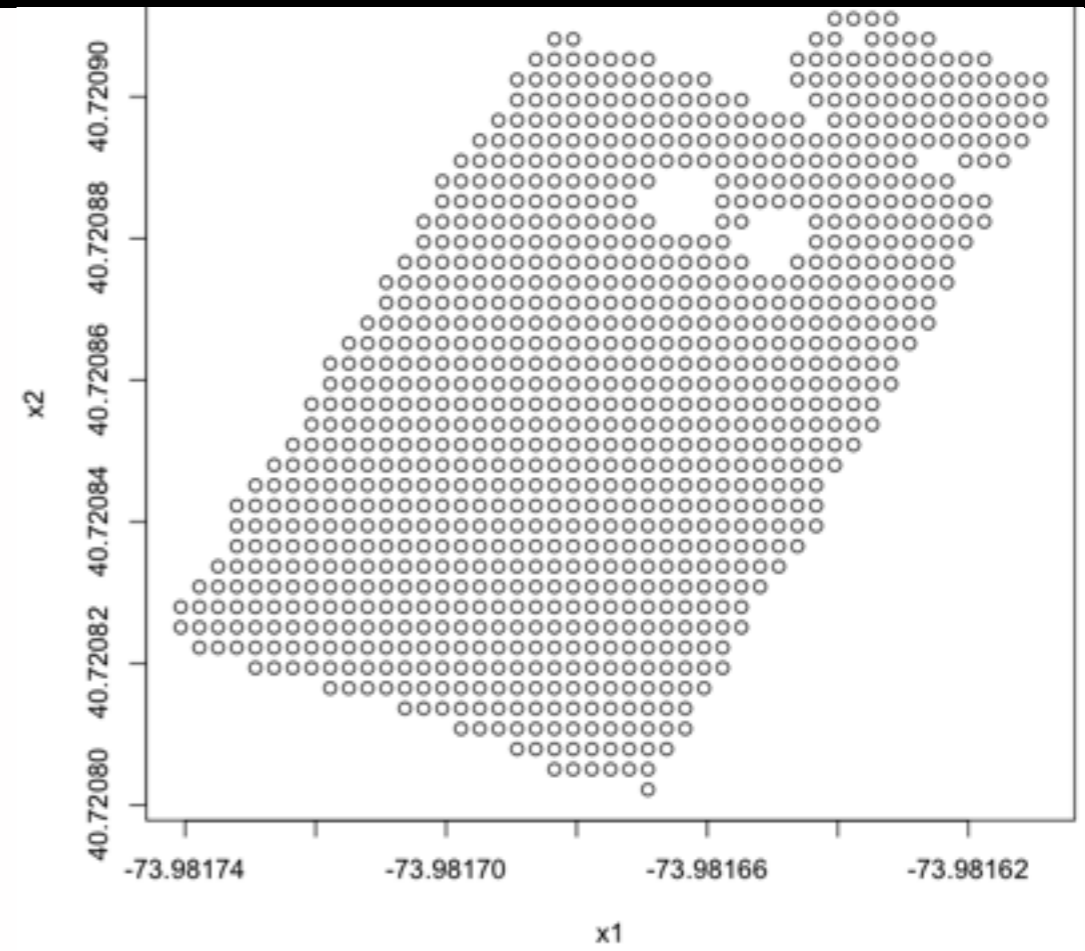
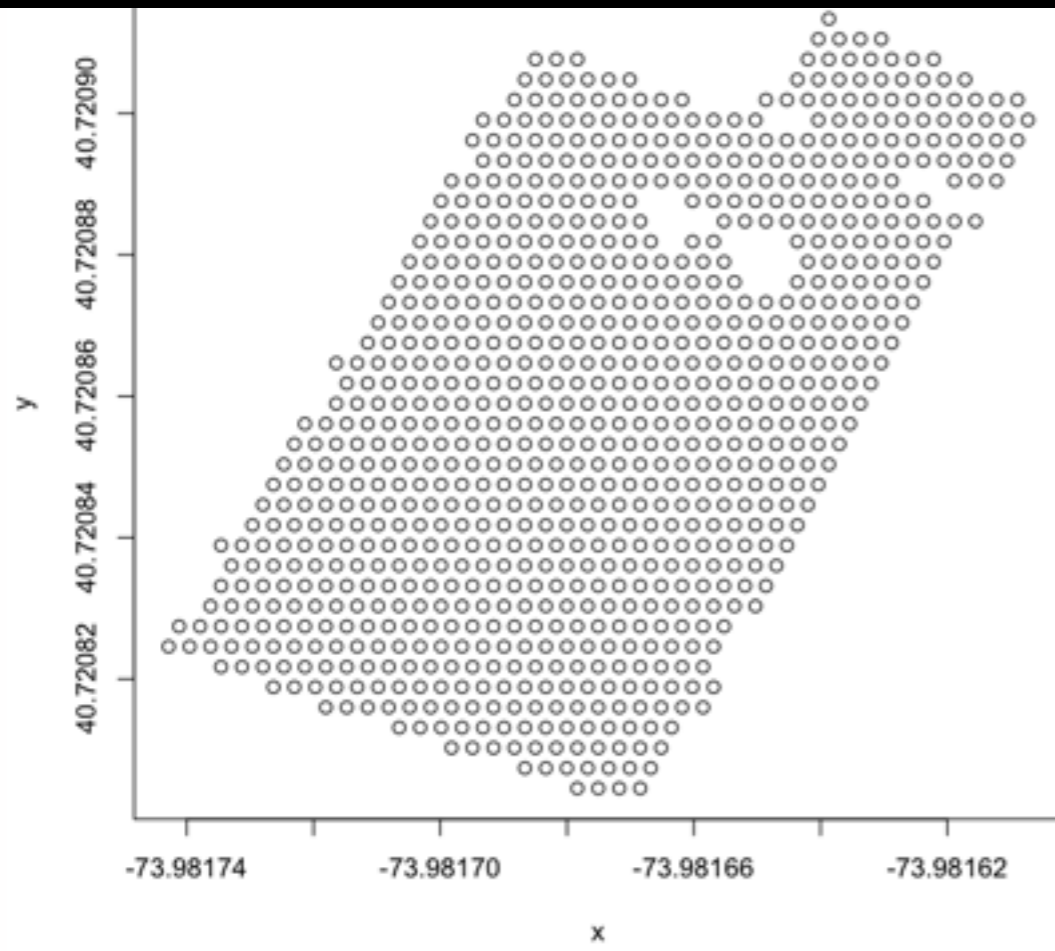




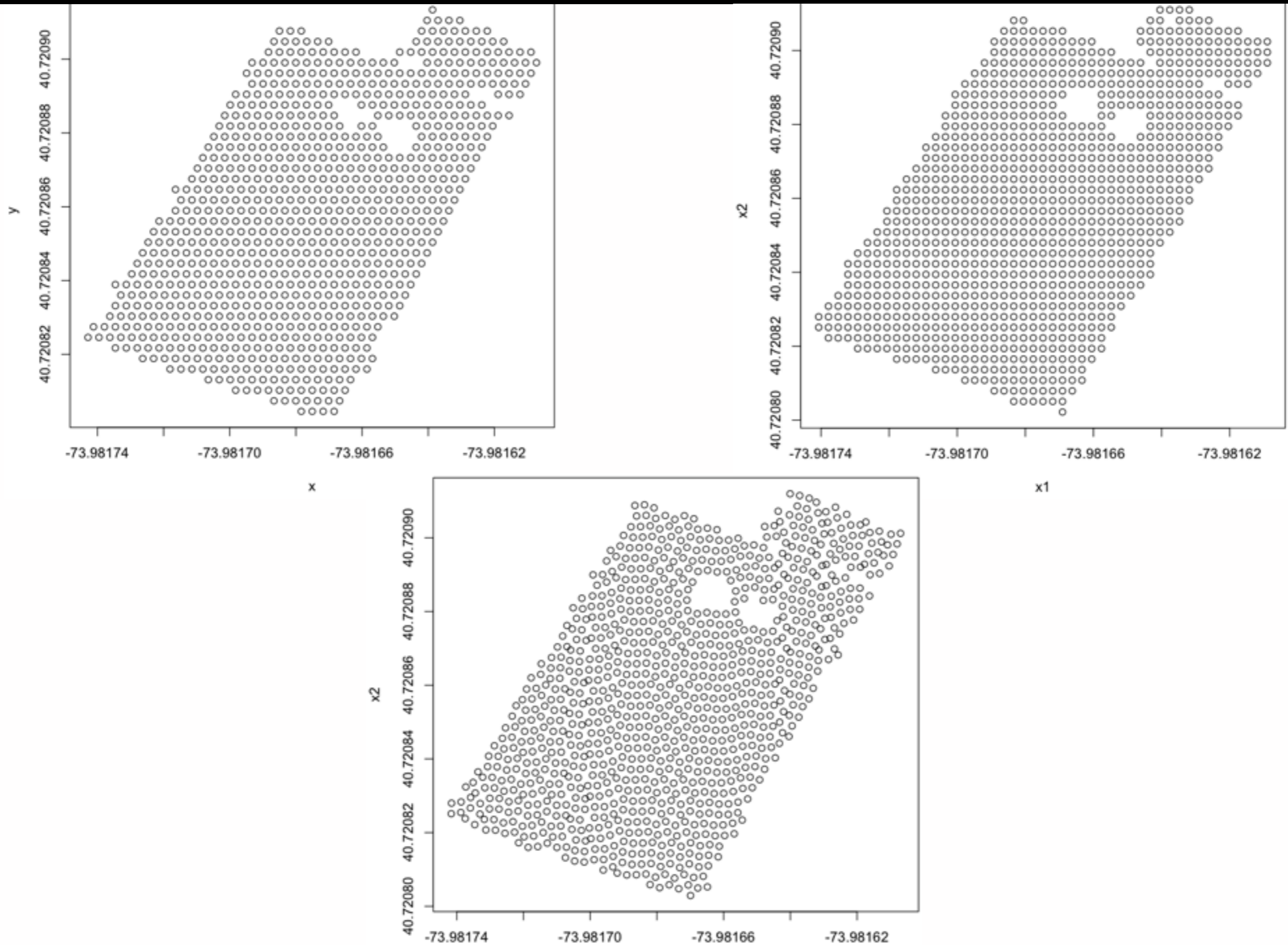
```
pts = spsample(polygon, n=1000, type="hexagonal")
```



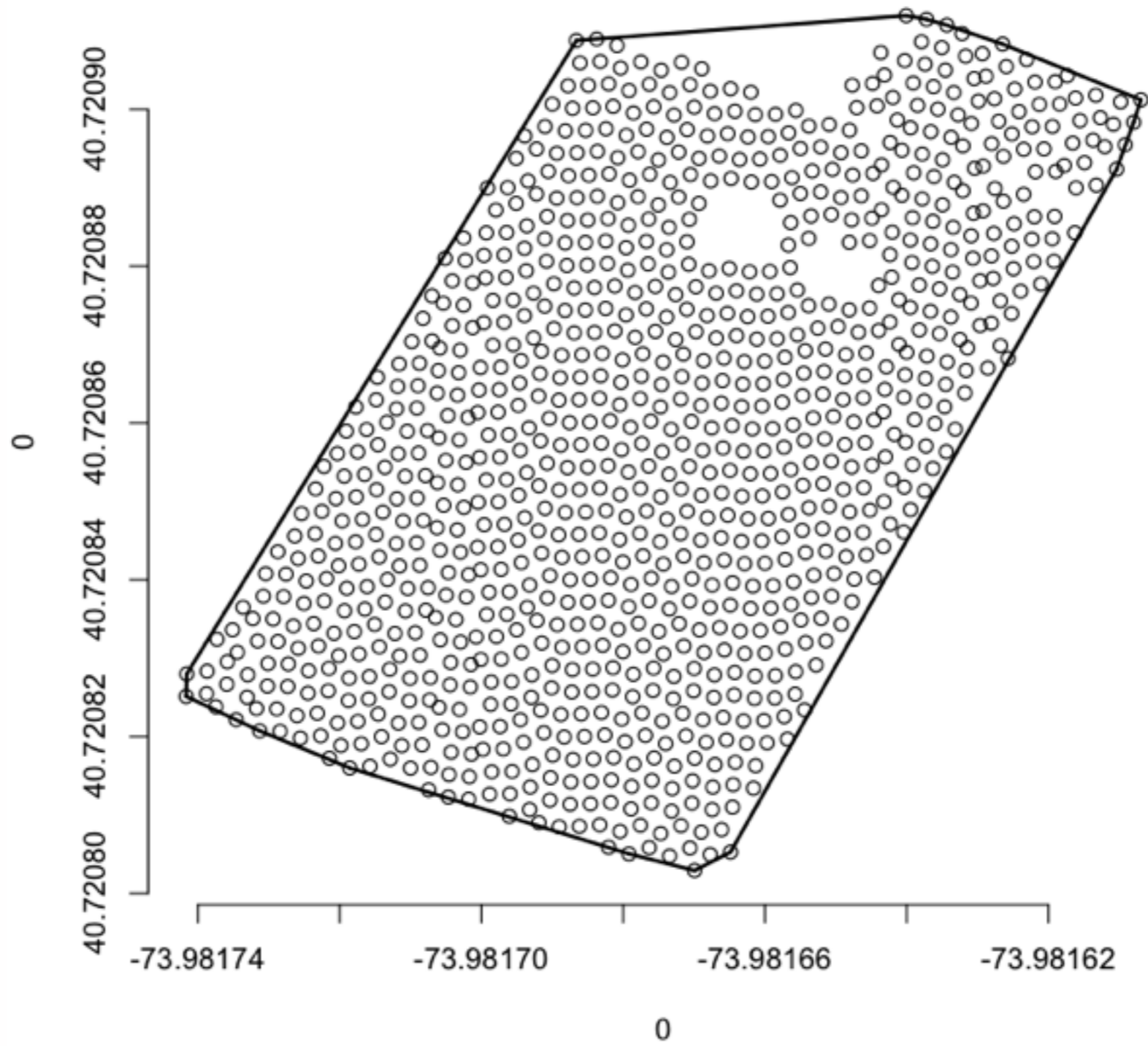
```
pts = spsample(polygon, n=1000, type="regular")
```



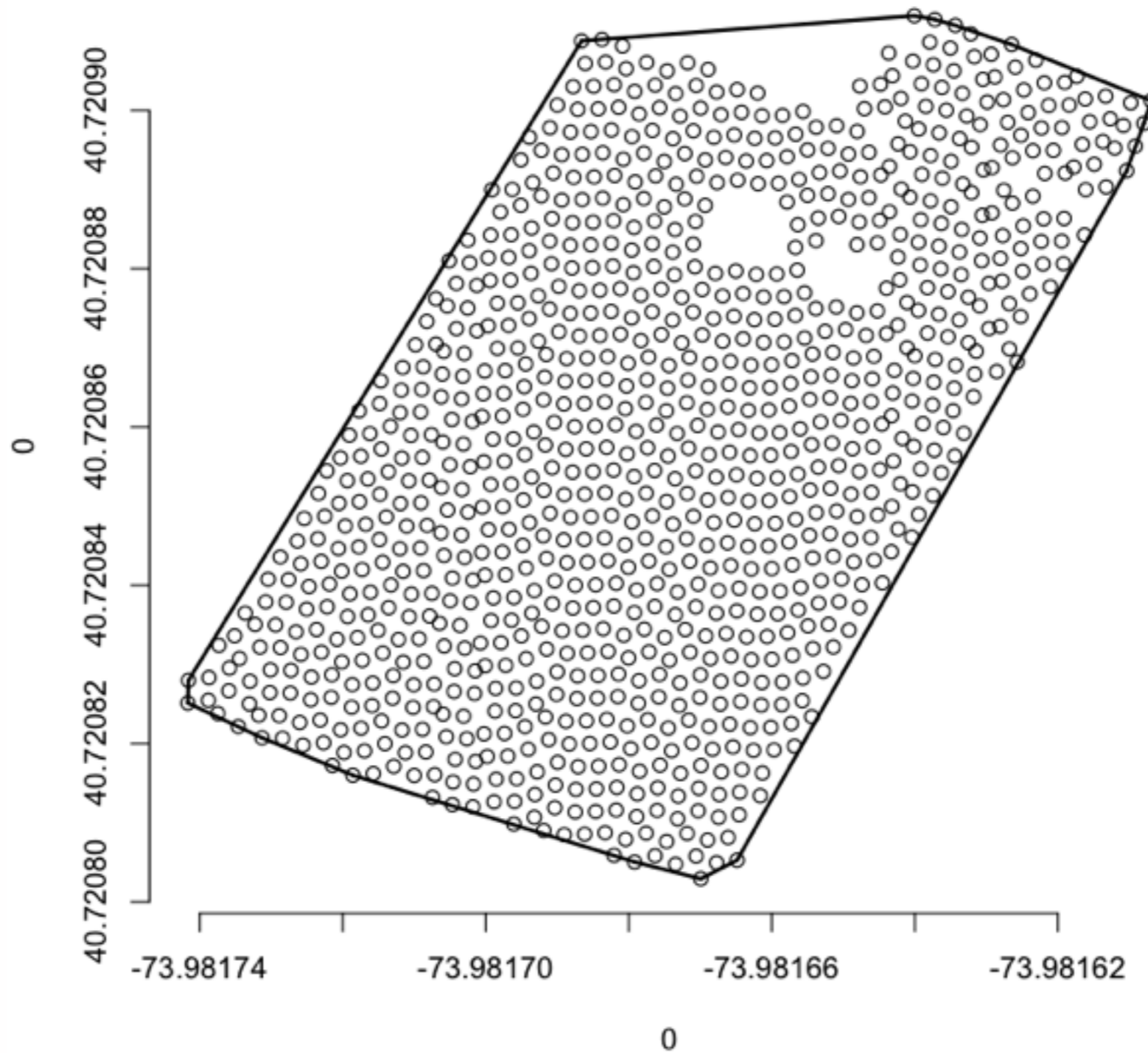
```
pts = spsample(polygon, n=1000, type="random")
```



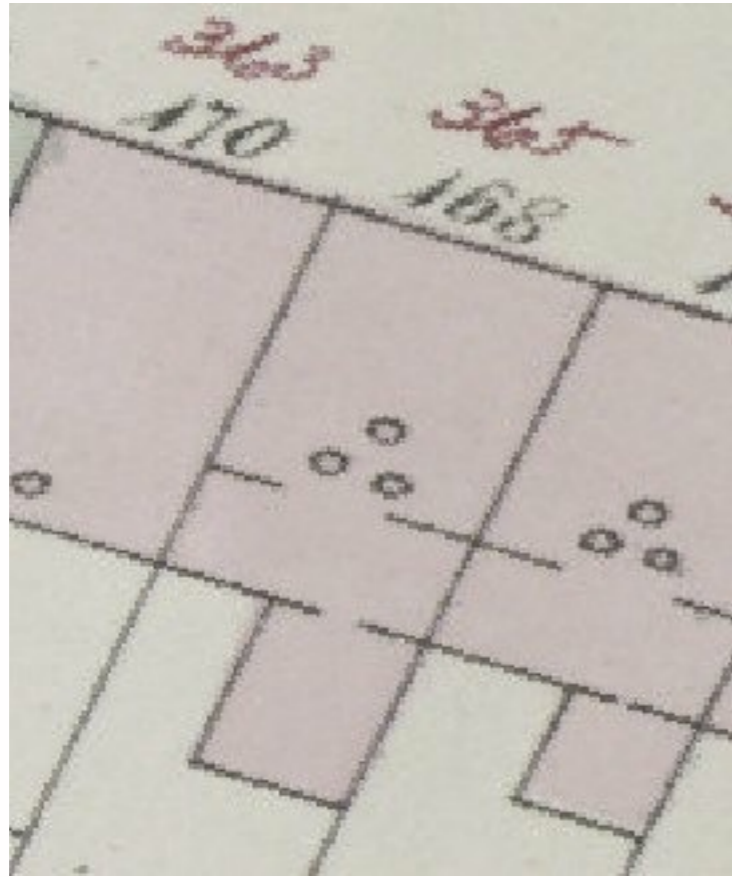
```
x.as = ashape(pts@coords, alpha=2.0)
```

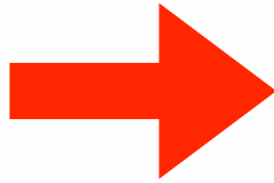
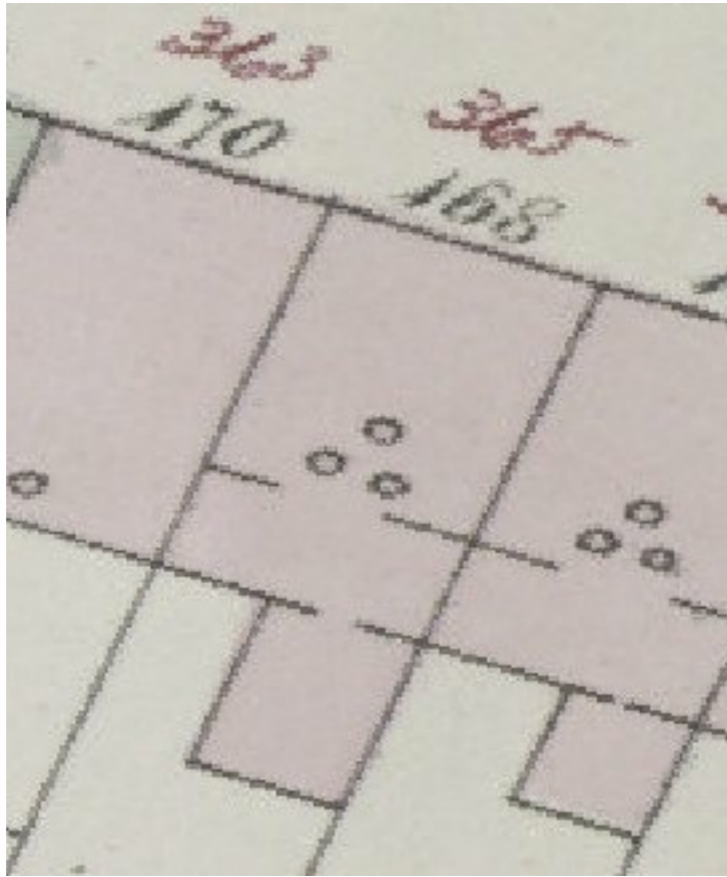


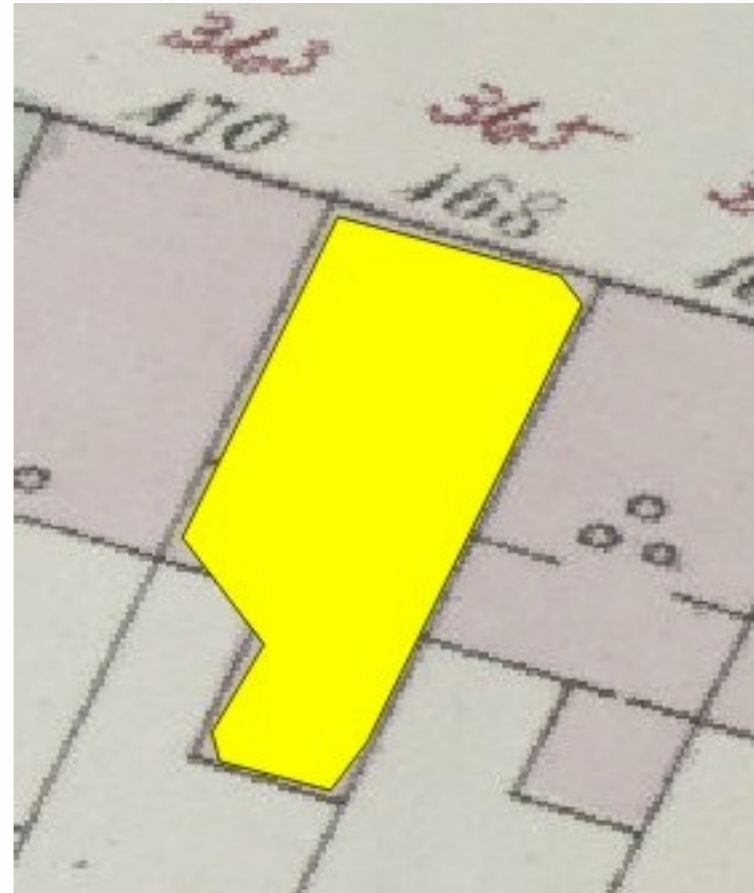
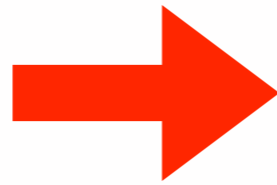
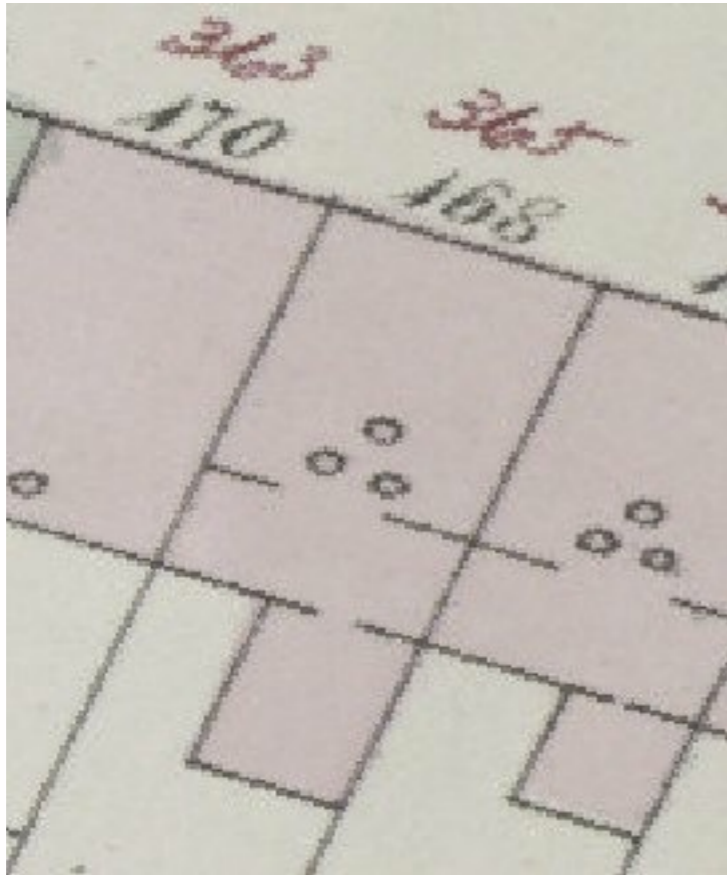
```
x.as = ashape(pts@coords, alpha=2.0)
```



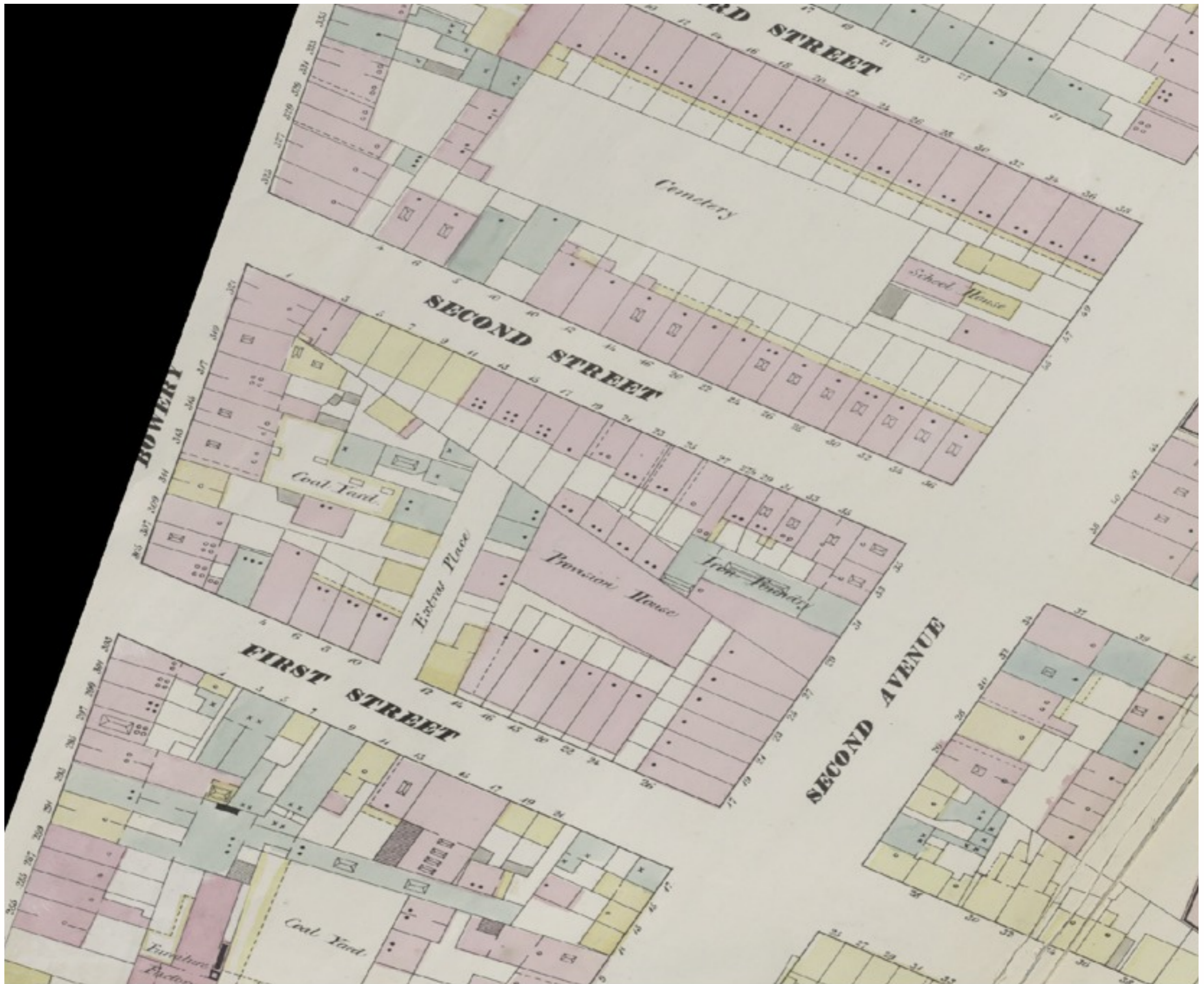
lower alpha produces more concave shapes (good)
but holes may start appearing (bad)



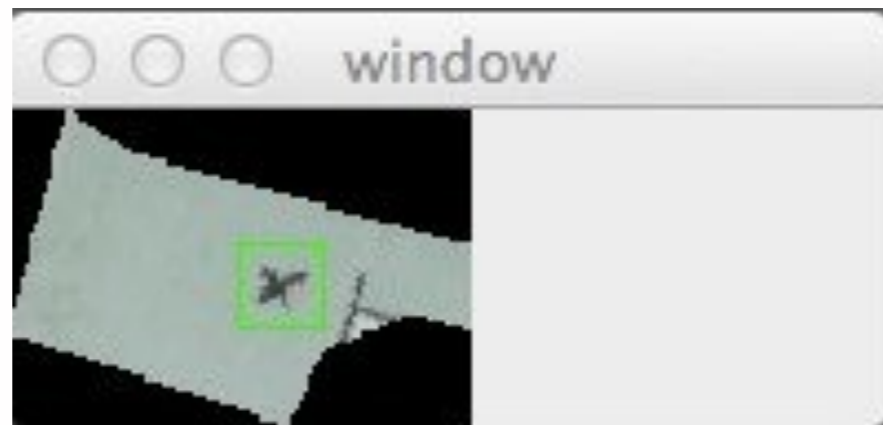
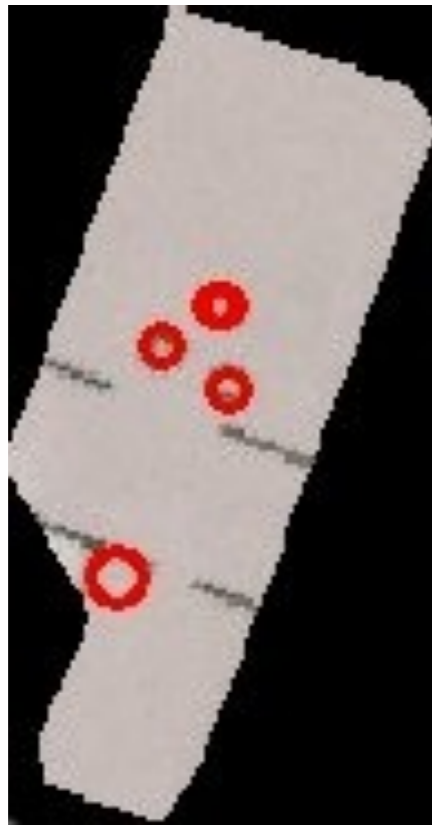




Ramer–Douglas–Peucker
and other point reduction algorithms
can be considered









bash

mga\$ _



66,056 polygons produced in **one day**

(as opposed to **years**)

but:

adjacency is not being enforced
false positives/negatives
buildings may also overlap

we need to **validate** the output

<http://buildinginspector.nypl.org>

*not included in the paper

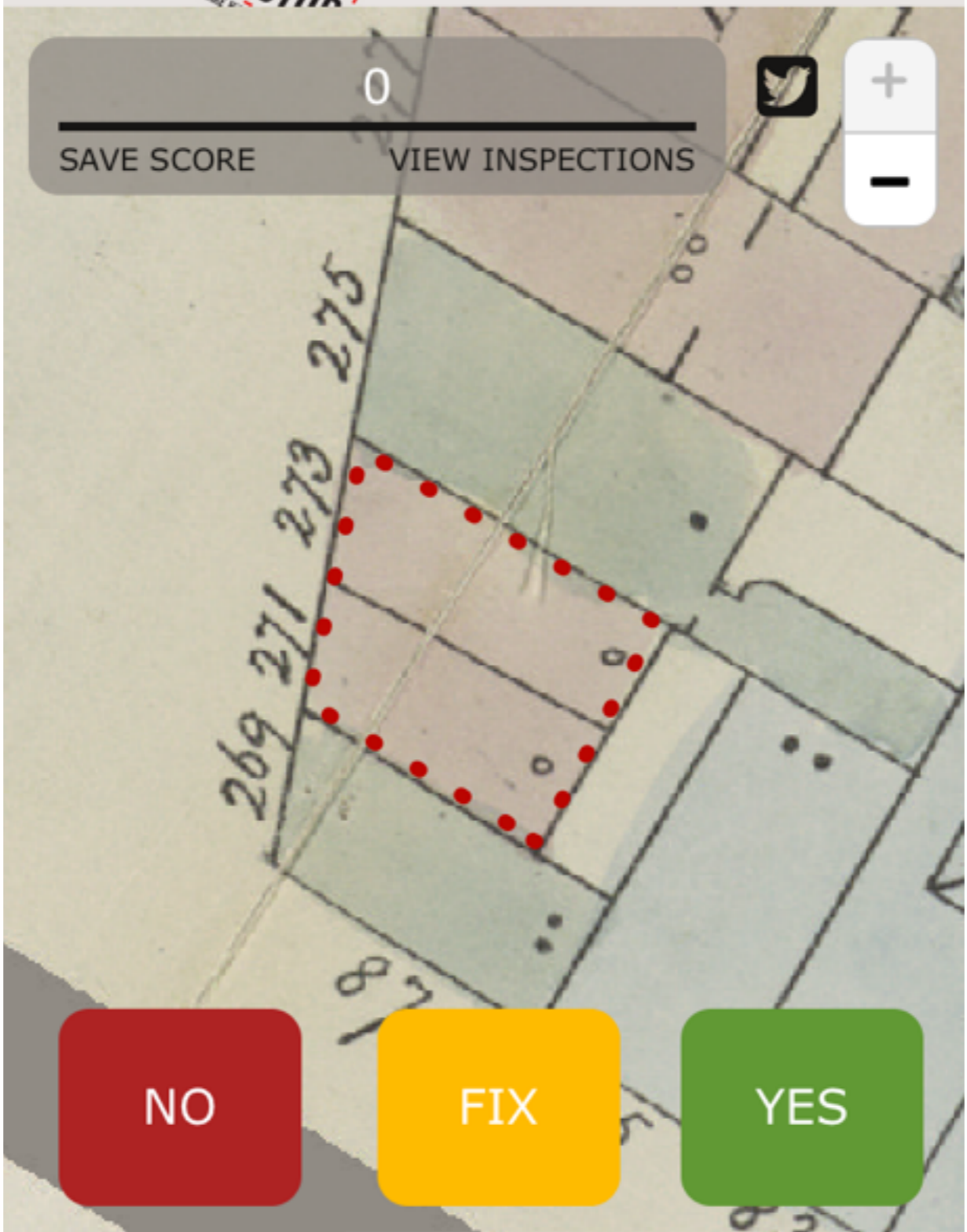


0

SAVE SCORE VIEW INSPECTIONS

Twitter icon

+ -



NO

FIX

YES





2 weeks later..

341,005 flags for **66,055** unique polygons

62,402 polygons with consensus

Yes 84.2%

Fix 6.4%

No 9.4%

“consensus” = 75%+ agreement of 3+ flags

no sleep till Brooklyn

14k+ more polygons

thank you

mauricio giraldo arteaga
NYPL Labs
@mgiraldo

NYGeoCon 2013