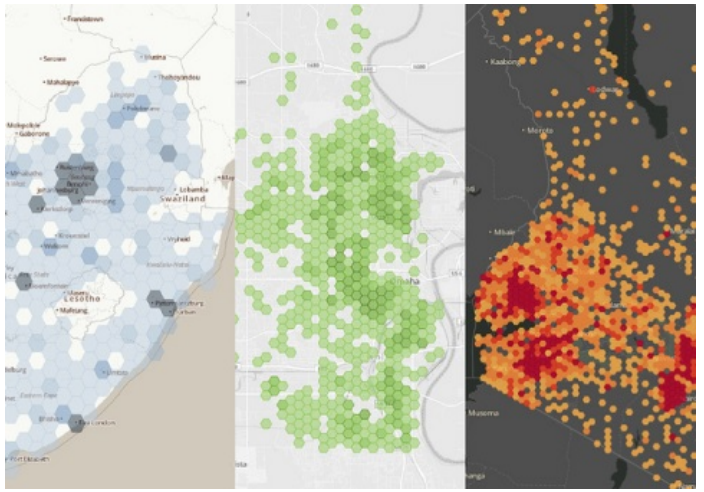
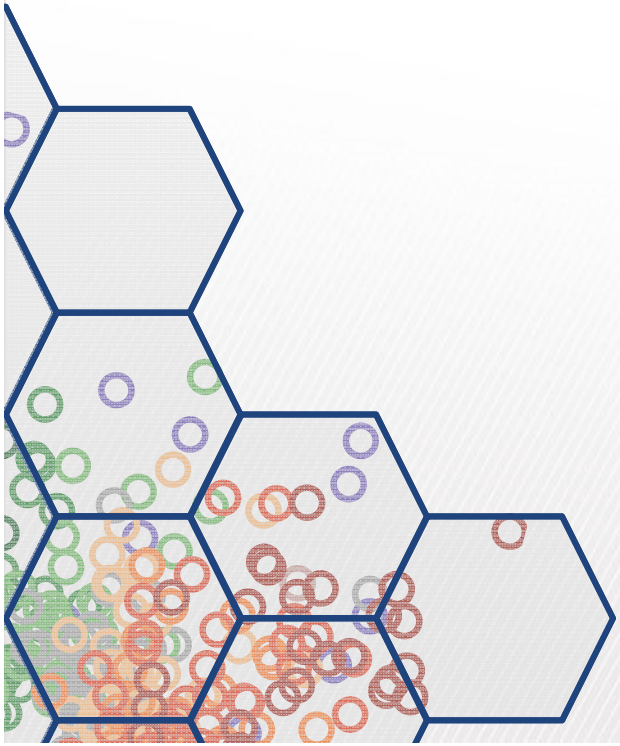


Spatial Binning



Sarah Battersby
Tableau Research
Spatial Gloom-And-Doom Spreader

Why do we bin?

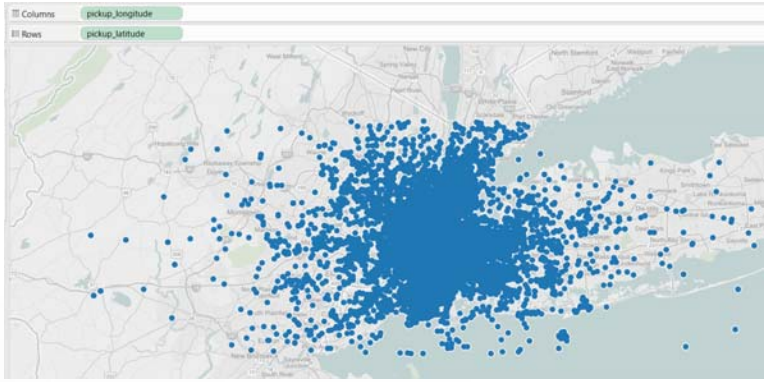
We have lots of data

- Overprinting
- Stacked points make density estimates difficult
- Easier comparisons across locations on different maps (bins give structure)

Because hexbins look awesome

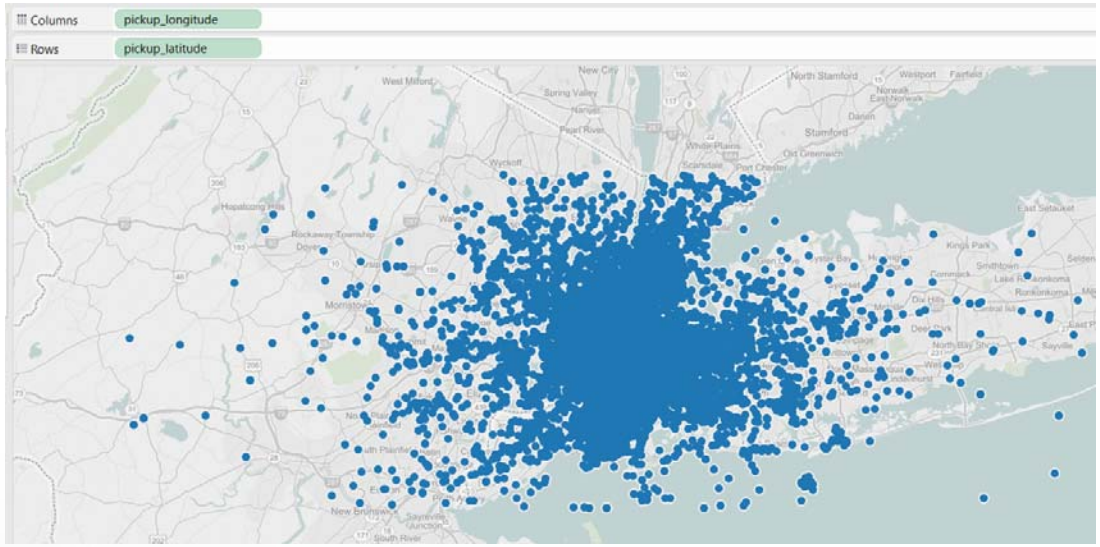
Why not other polygons?

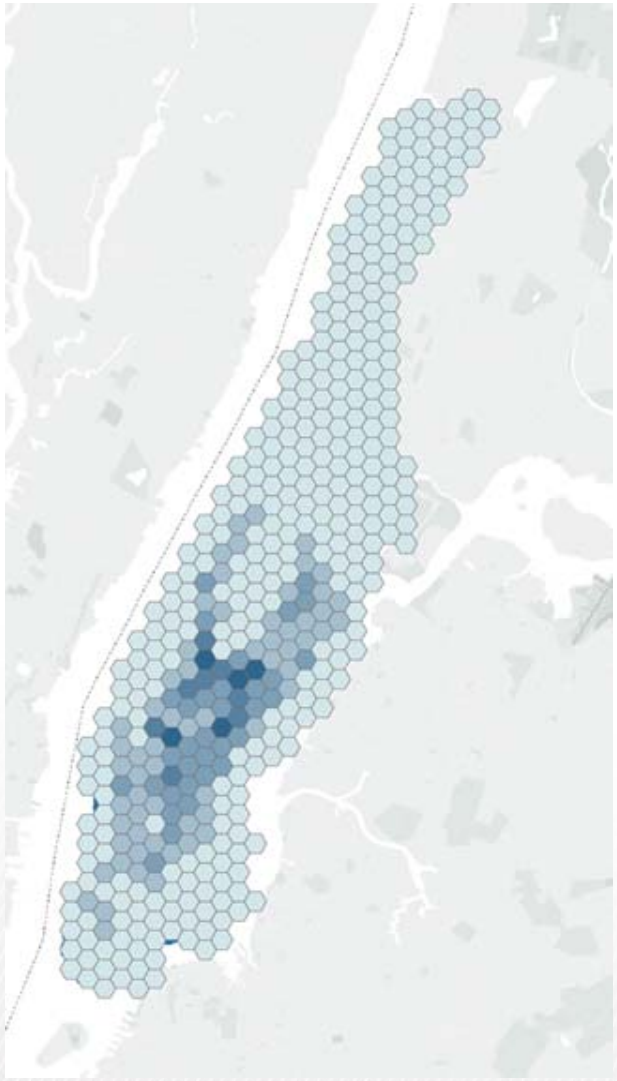
- Might not make sense to aggregate to Census tracts, etc. (show smoother patterns of change, not abrupt)
- Want same size unit for aggregation



Taxi pick up locations, NYC
173m+ records

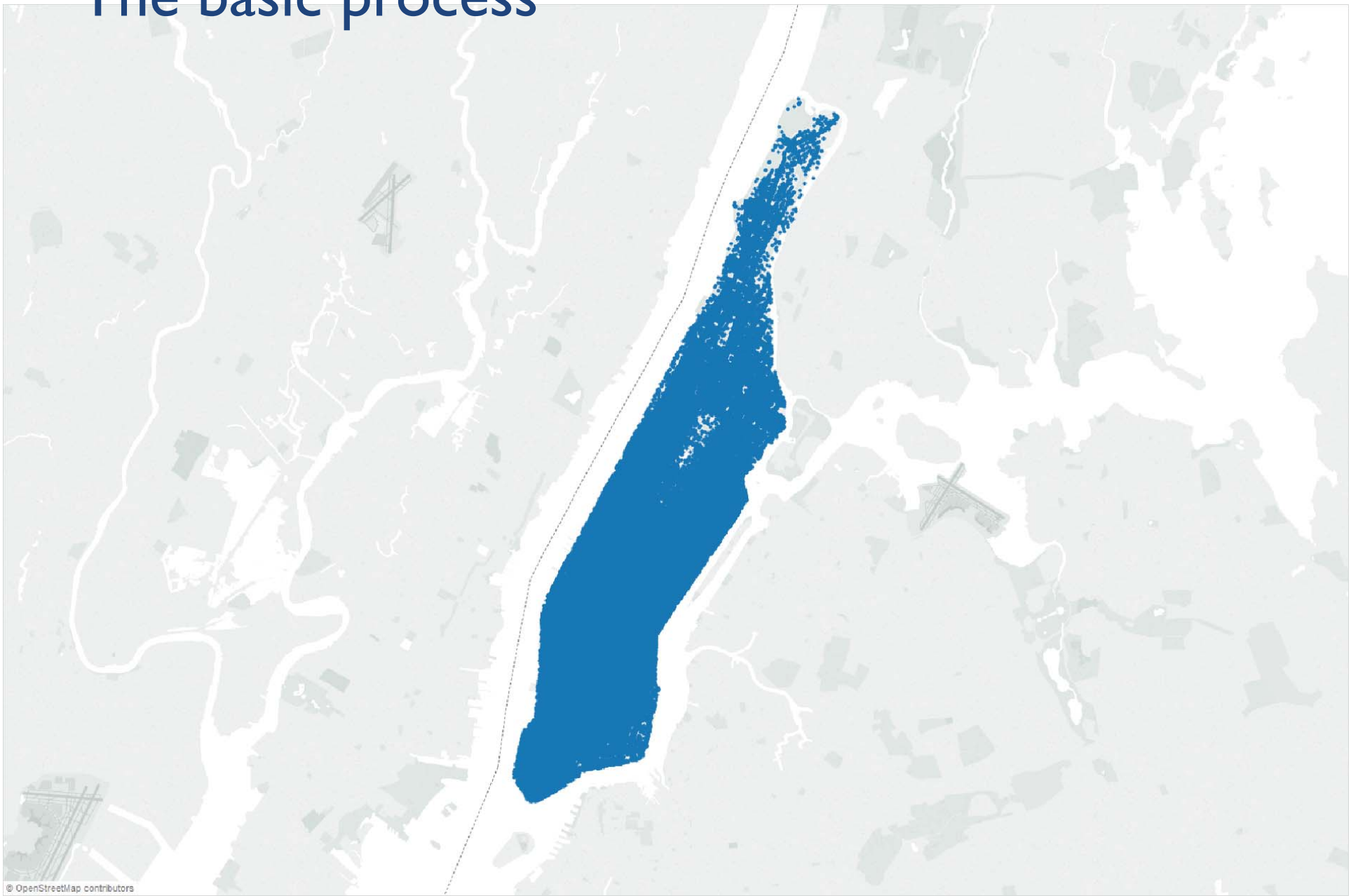
Because it makes the spatial patterns clearer



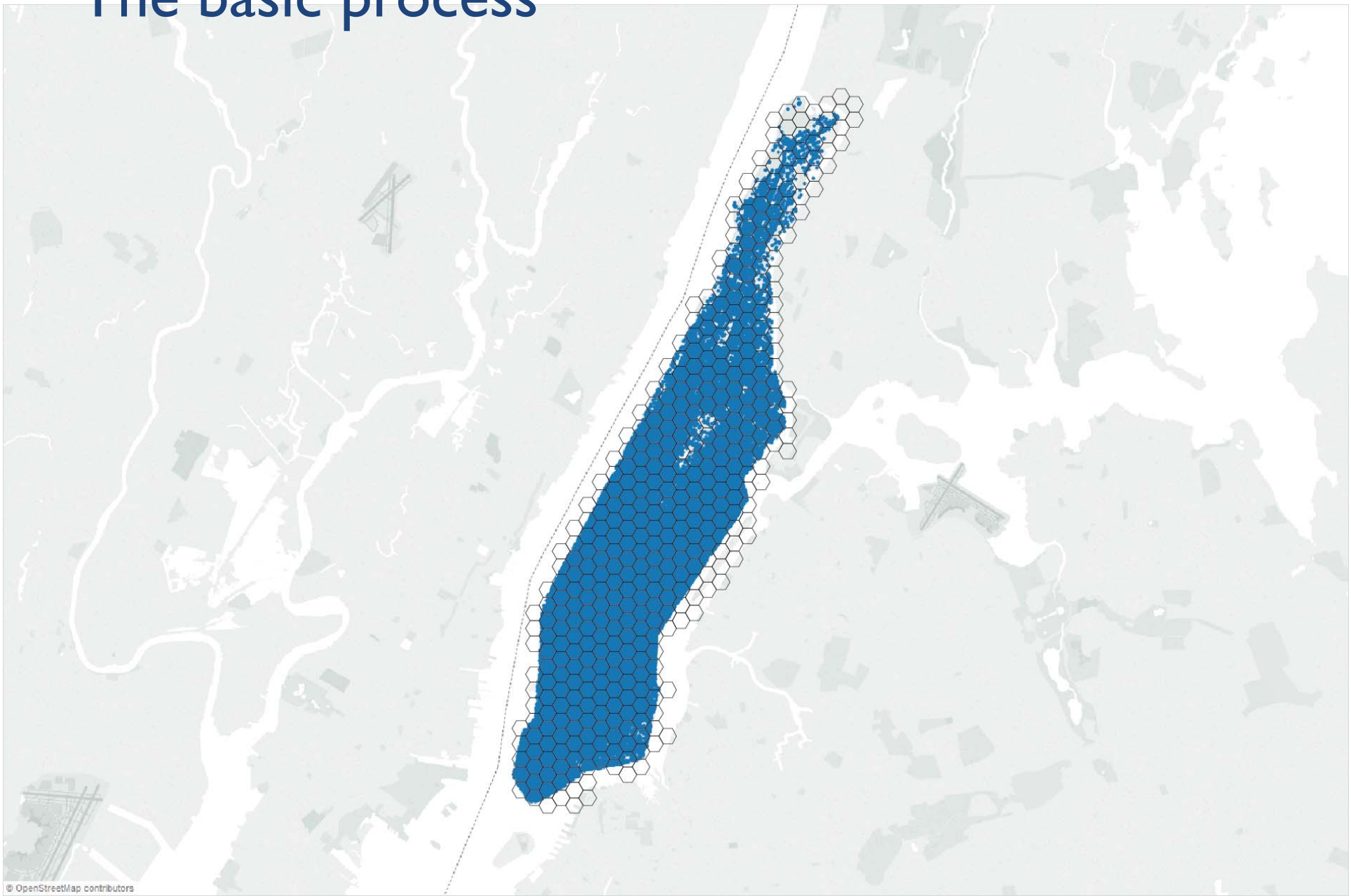


The *idea* is that we can use spatial bins to simplify the process of making density estimates for complex point datasets

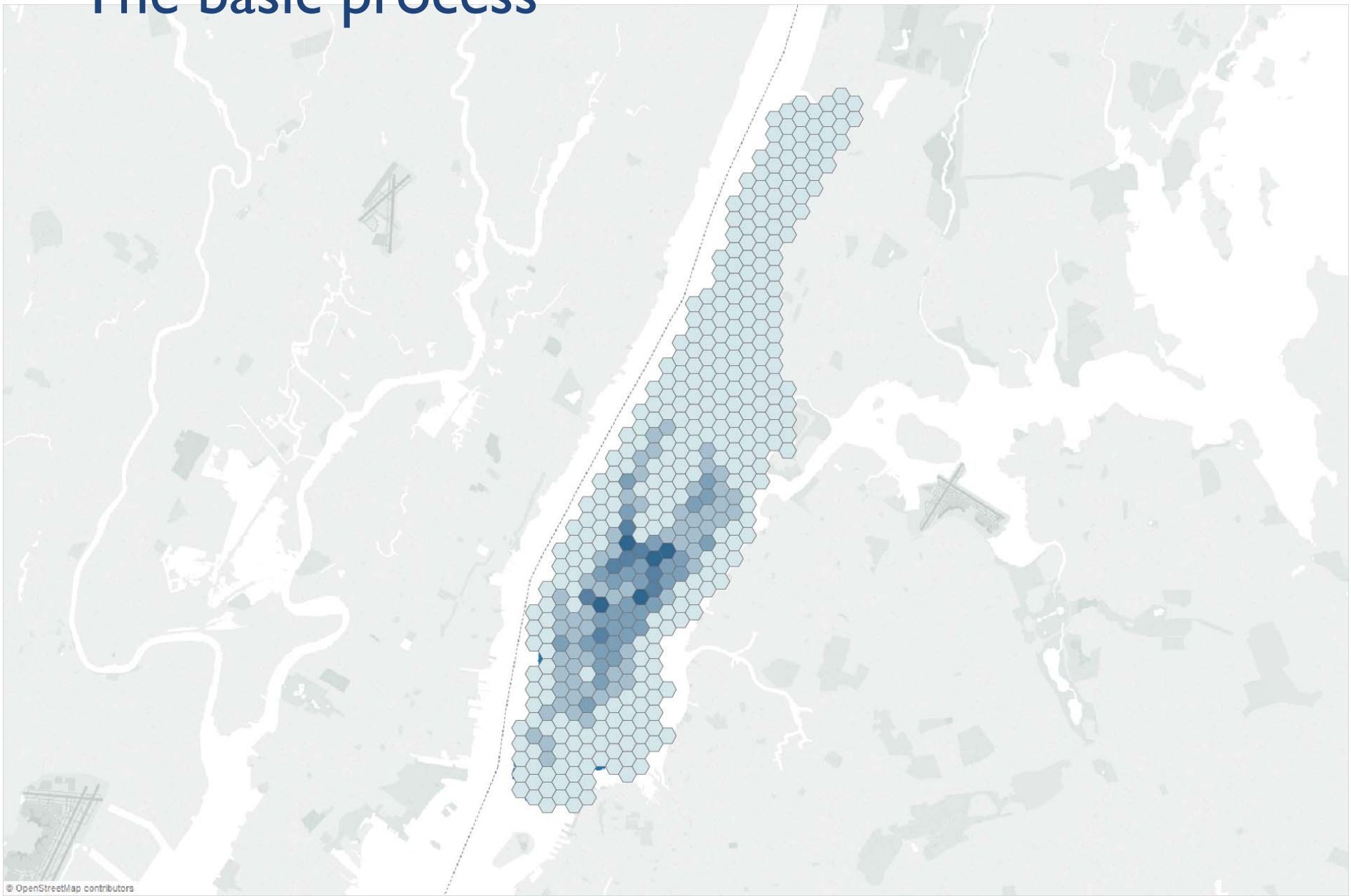
The basic process



The basic process



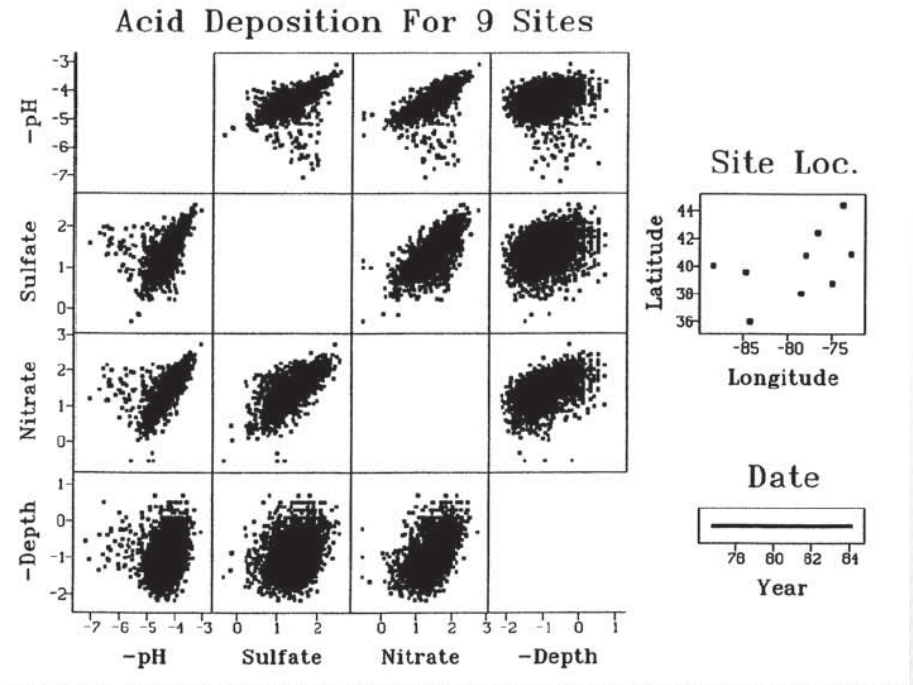
The basic process

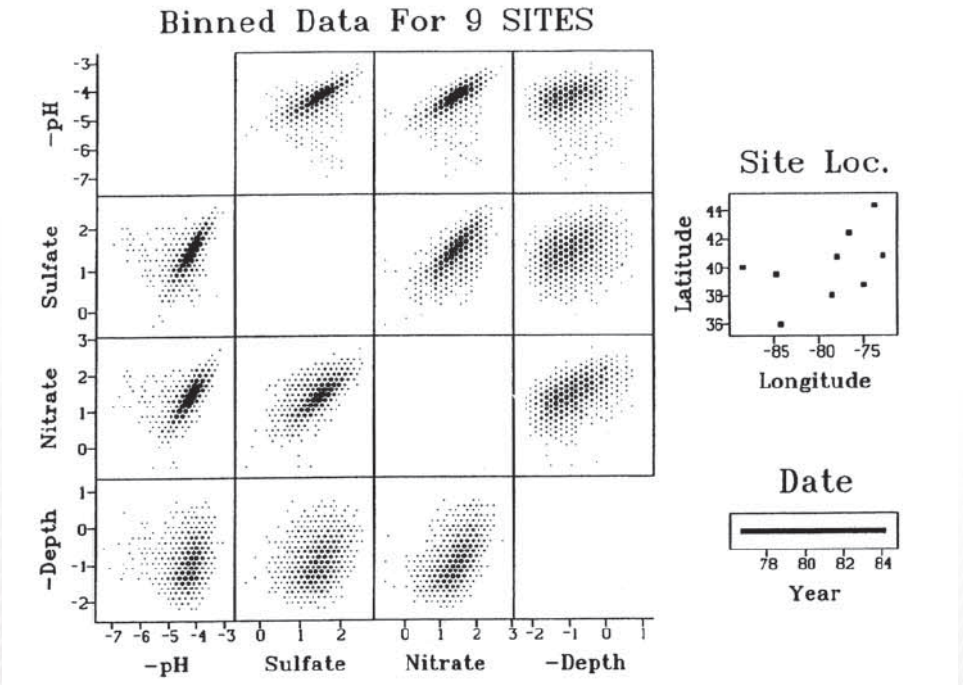


Side note: why *hexagons* for the bins?

Carr et al. 1987

- Scatterplot matrix techniques for large N
- Hexagons instead of rectangles





Rectangular vs. hexagonal binning

Visual appeal

"Rectangular bins lead the eye to align bin centers and to see regularity where there is none." (Wilkinson, 142)

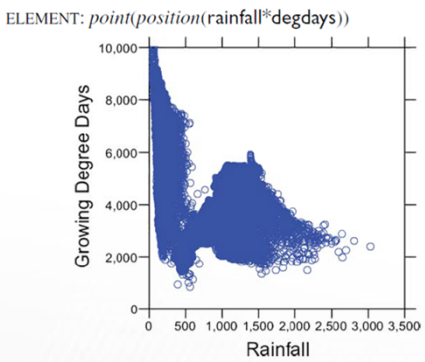


Figure 7.21 Simple scatterplot

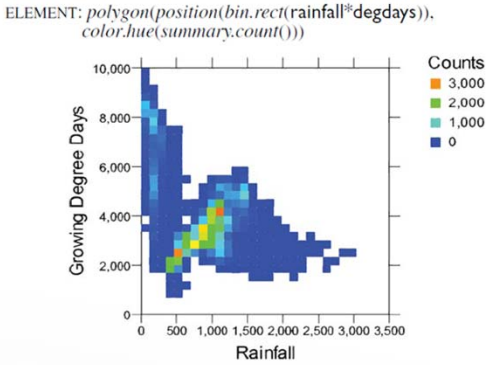


Figure 7.22 Rectangular binning with color representing counts

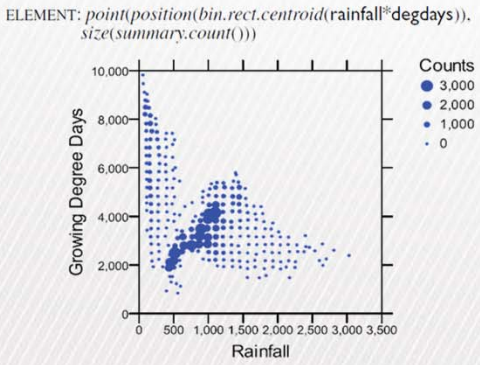


Figure 7.23 Rectangular binning with symbols positioned at bin centroids and size of symbols representing counts

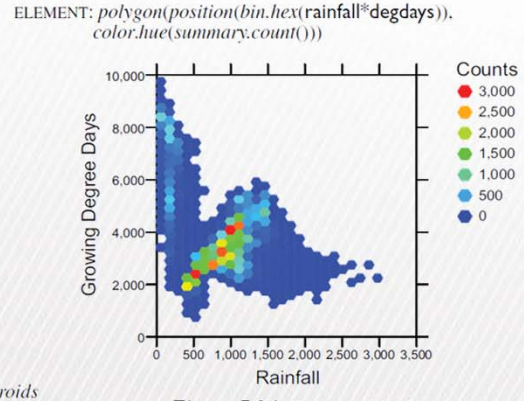
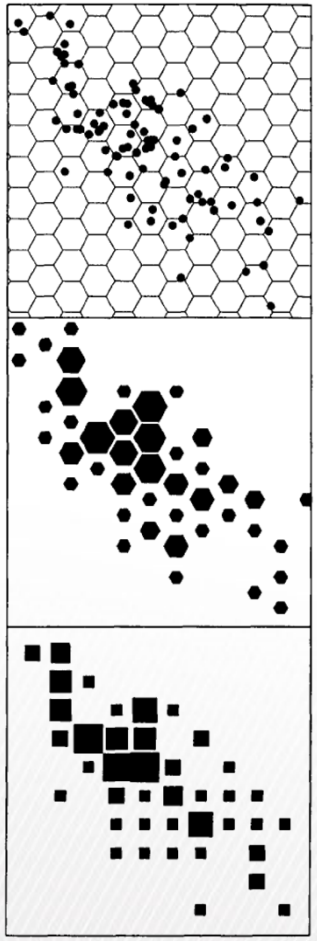


Figure 7.24 Hexagon binning

Source: Wilkinson



Source: Carr et al, 1992

Rectangular vs. hexagonal binning

Representational accuracy

- Scott (1985) – hexagon-based density estimates have somewhat smaller integrated mean square error than square-based estimates
- More compact shapes are preferred
- Carr et al. 1992 suggest the visual argument in favor is stronger than the representational argument

Some other assorted reasons…

- E.g., on maps don't tend to line up with regularly-spaced cultural features

User's mental model for how this works…

Assumption based on *my* mental model:

- Regular grid of same size and same shape
- Bins are definable in understandable units
  - based on axes or real-world distances
- All features assigned to the bin they fall into (point in polygon problem)
- Bins are consistent size/location across viz (e.g., map comparisons)

- Bin contents are a good reflection of what is really *here* in the visualization

Three regular polygons tessellate the plane:

- Equilateral triangles, squares, hexagons

Source: Wolfram.com

What's going on?

Sphere -> plane = Distort areas and/or angles

- You either don't have bins that are the same ground area
- Or you don't have bins that are the same shape
- Or both!

Compounding problems

- Most people don't understand (or can mentally translate) for distortion
- A "strength" of binning is that we have a regular grid to organize our data
- Disconnect between defining bin size (degrees, meters, etc.) and coordinates on map helps black box what is happening

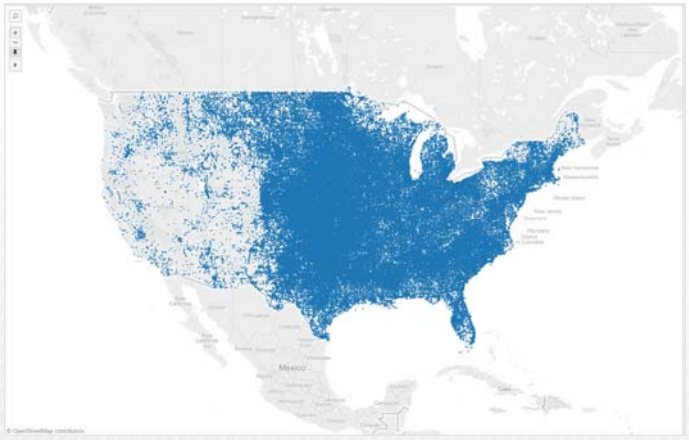
Let's step back…

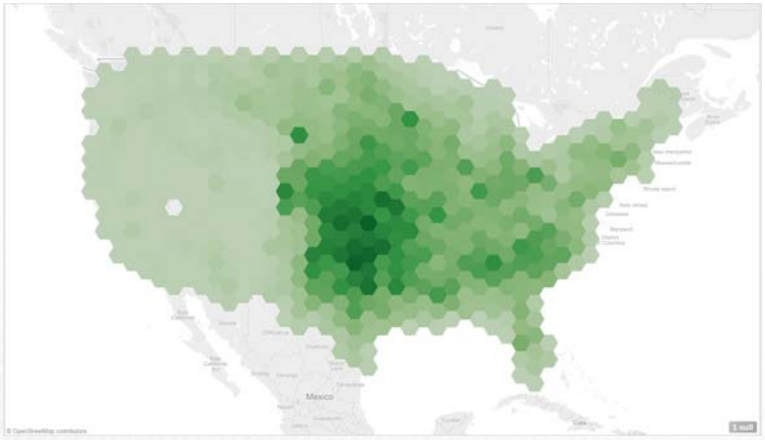
Regular grid of same size / same shape?

What does that *mean* on maps?

- Option 1: Regular grid in Web Mercator coordinates*

Hailstorms 1955-2013 (NOAA)





* I largely live in the world of web maps… Web Mercator? We hang out *all the time!*

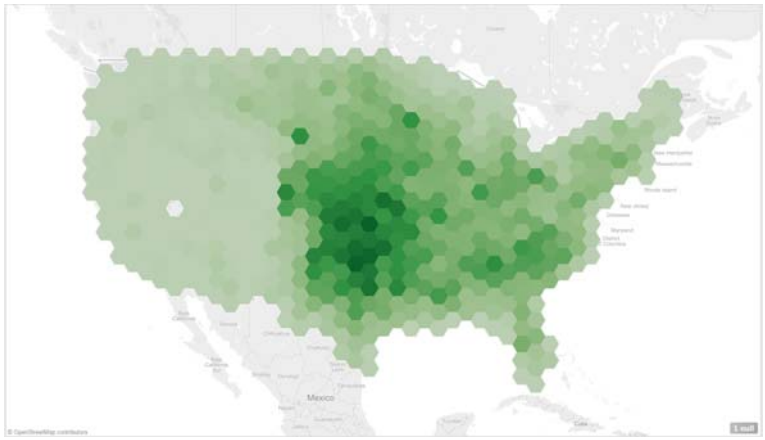
Regular bins in Web Mercator space

Shape –

- Consistent on the plane
- Irregular on the sphere

Size –

- Consistent on the plane
- Map area ≠ Ground area



* Size for visual analysis is constant
* Size for spatial analysis becomes smaller and smaller approaching the poles

Spatial aggregation –

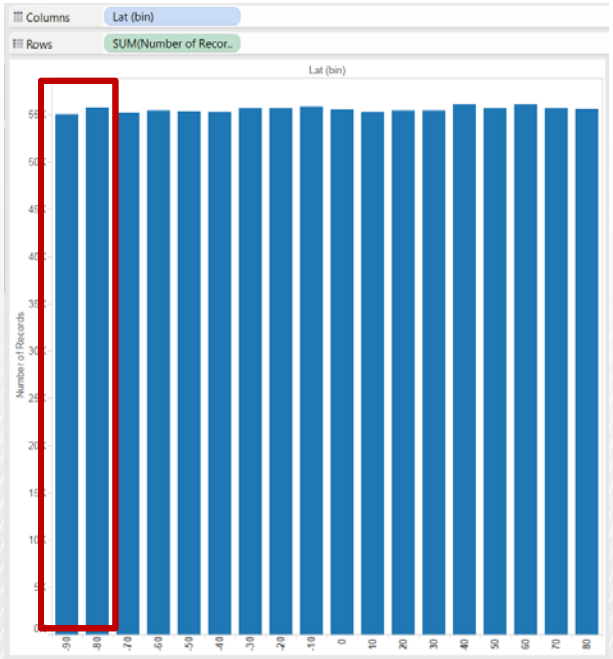
- If binned on the plane, density will most likely decrease approaching poles
- Spherically smaller bins collecting points that are becoming more distant

Regular bins in Web Mercator space

What happens to point distributions in the bins?

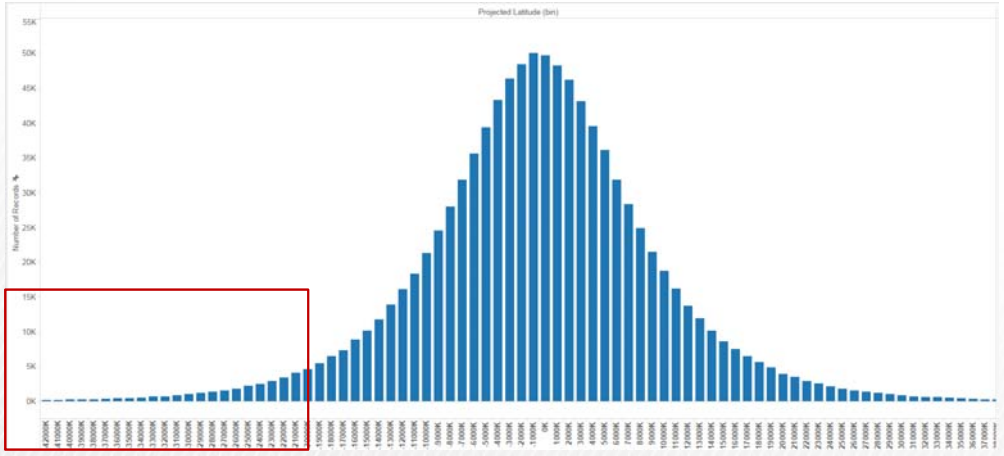
Density goes lower and lower and lower at high latitudes

Totally because there's just less stuff there, right?
Or because the bins cover smaller and smaller and smaller ground areas…



Actual latitude

Web Mercator

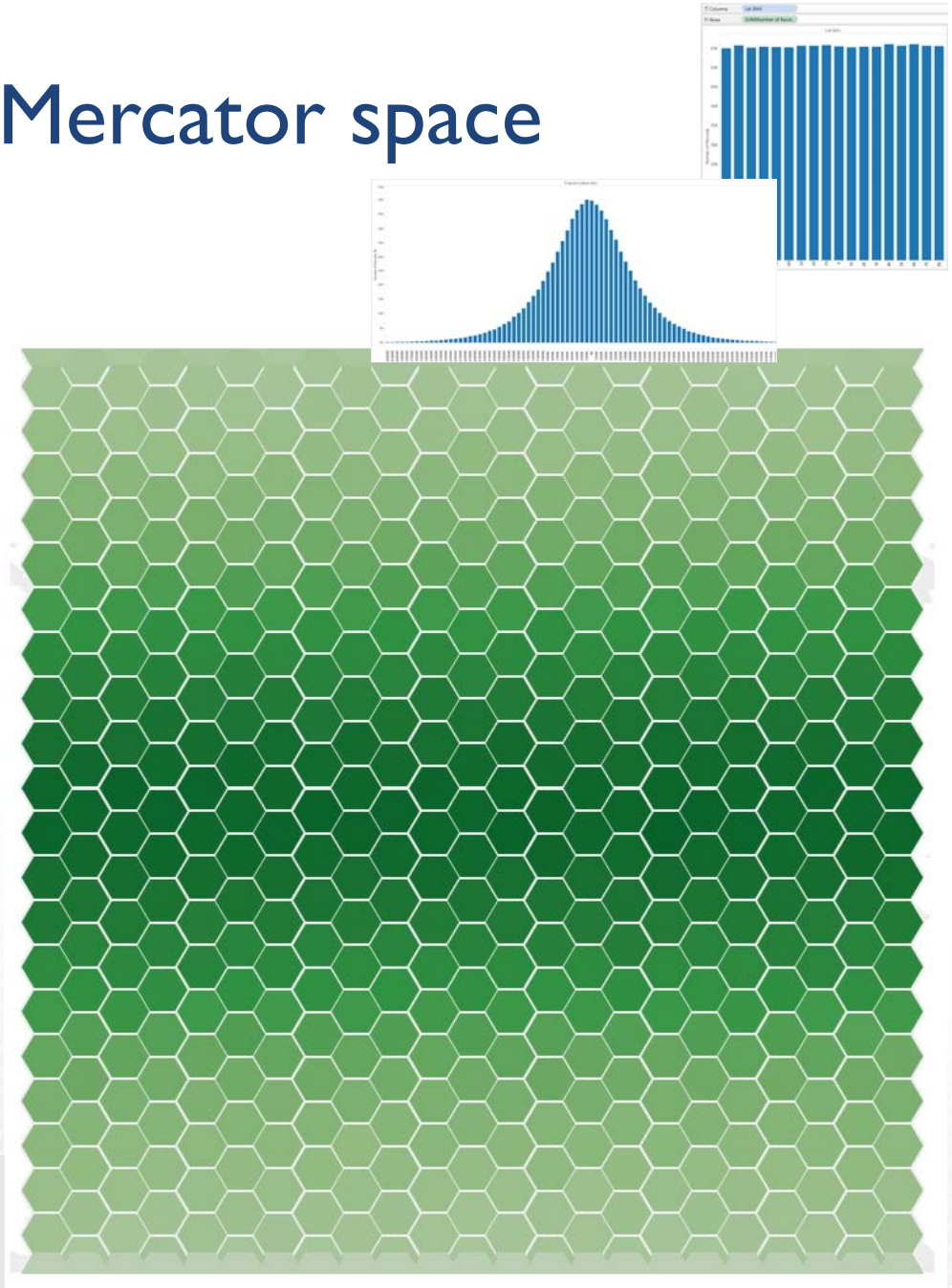


Regular bins in Web Mercator space

Hmm, it's amazing how the visual "density" changes so much when I just show raw counts.

Those bins are the same area, they totally give us accurate density measures.

Hahahahahahhahaha. No.



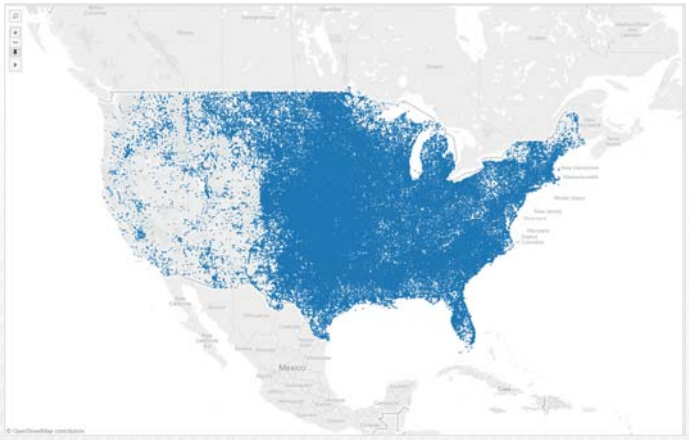
Let's step back…

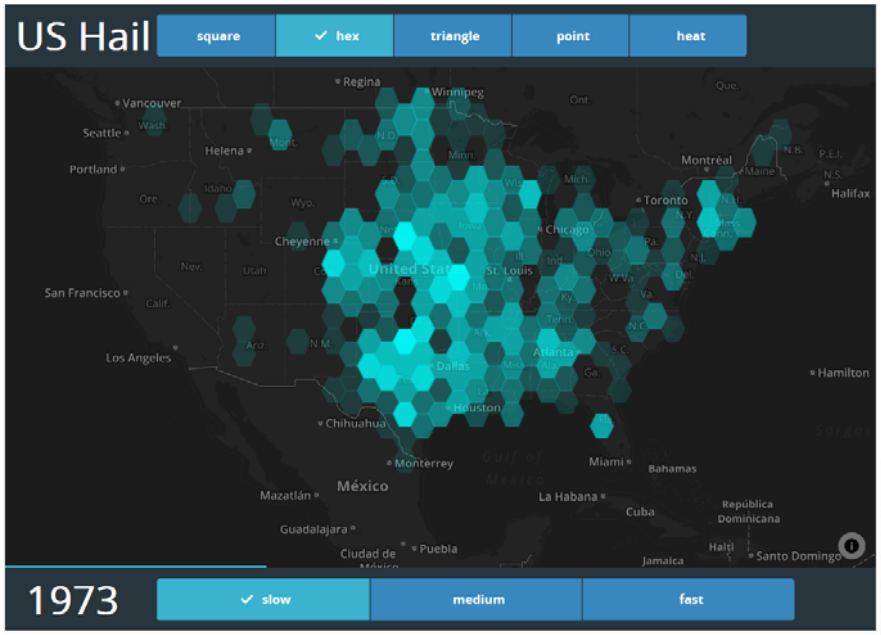
Regular grid of same size / same shape?

What does that *mean* on web maps?

- Option 2: Regular grid in latitude and longitude*

Hailstorms 1955-2013 (NOAA)



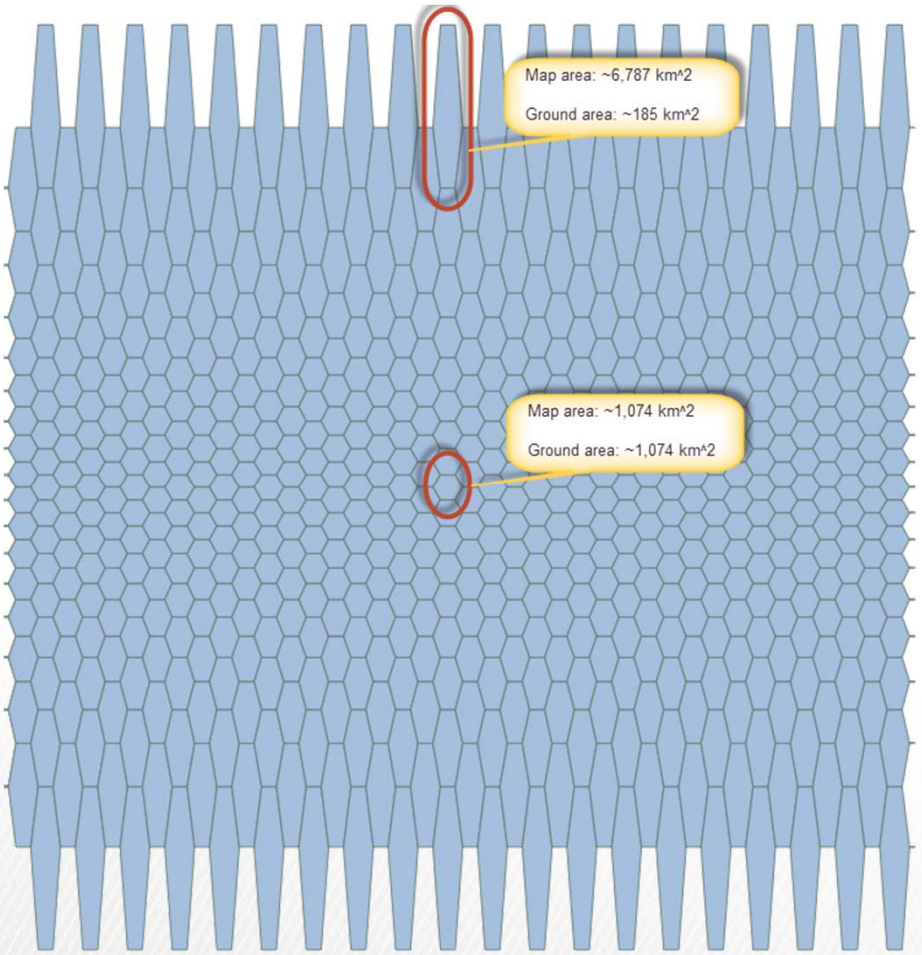


Source: https://www.mapbox.com/blog/heatmaps-and-grids-with-turf/

Regular bins in units of degree

Why would anyone do that?

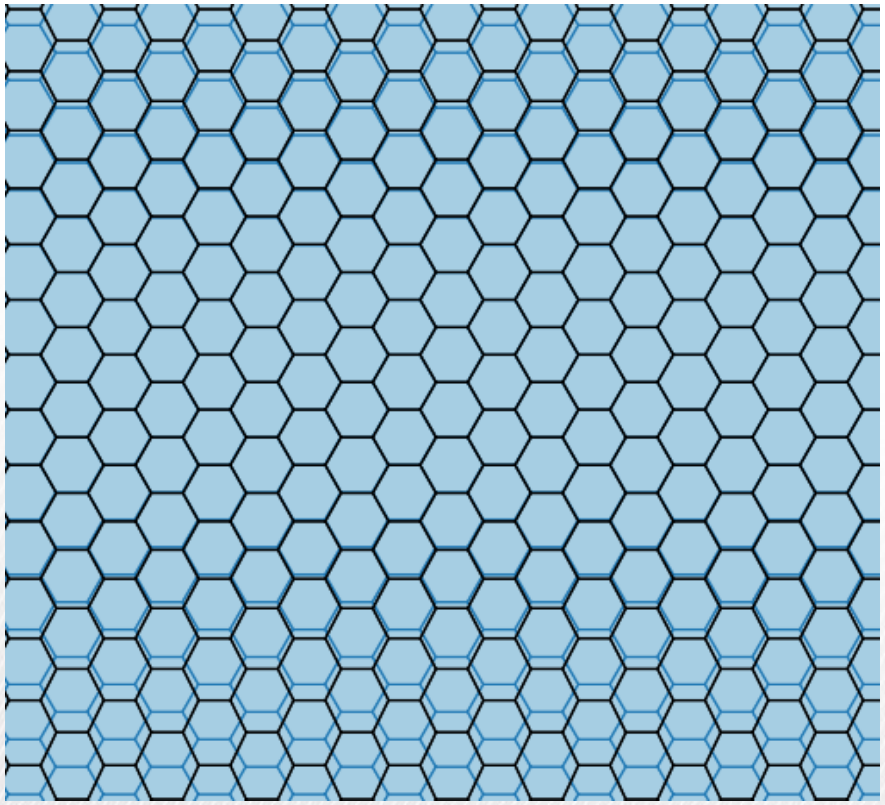
Because we live in a flat map world and people just don't know any better?

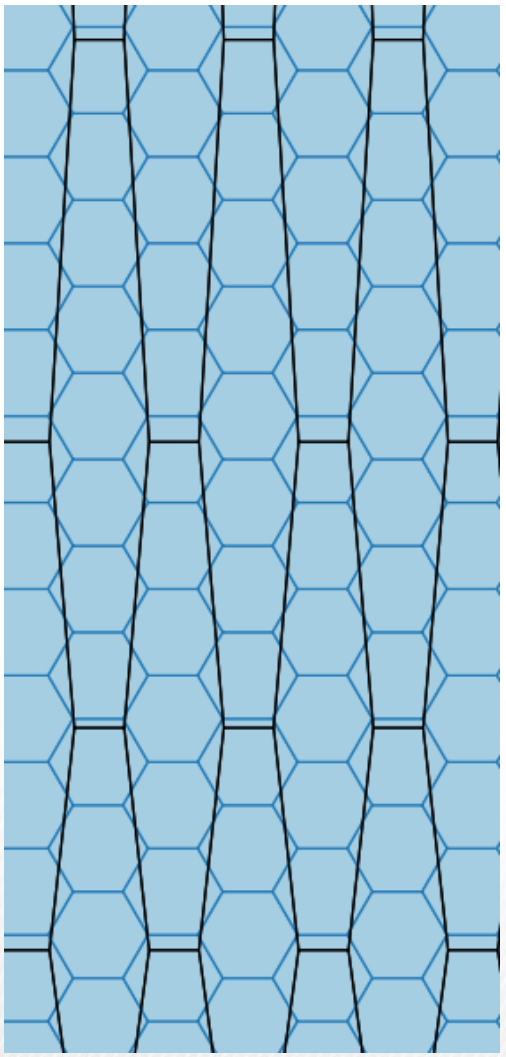


For comparison's sake…

Regular Web Mercator bins vs. lat/long bins

Totally same edge length at the equator





Of these choices, I go with the regular bins in map units

…though you have to accept, understand, acknowledge the errors

Moral of the story…your bins are probably wrong, but try to make them more right-er if you can

Think carefully about what you are trying to map

Think carefully about the size and location of the area you are trying to map

Equal area projection:

- You can preserve area, but the bins will not be regular on sphere

Non-equal area projection:

- You can make regular bins in projected space, but…
  - Do you want to present comparable measure for density?*
    - Reader will interpret as identical quantities that are *less and less* as you move away from equator
  - Or comparable measures for quantity?
    - Reader will interpret as identical densities that are *greater and greater* as you move away from the equator

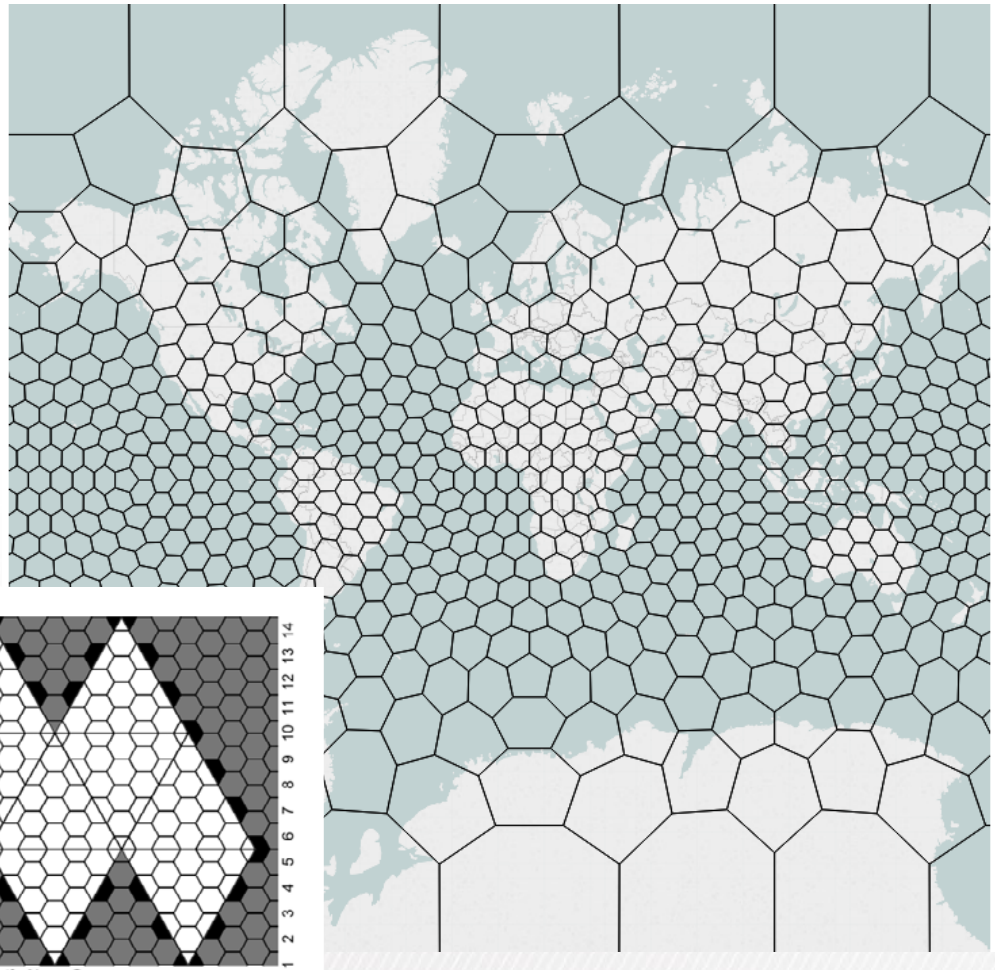
*With equal size bins, damned if you do and damned if you don't – people are probably going to assume density

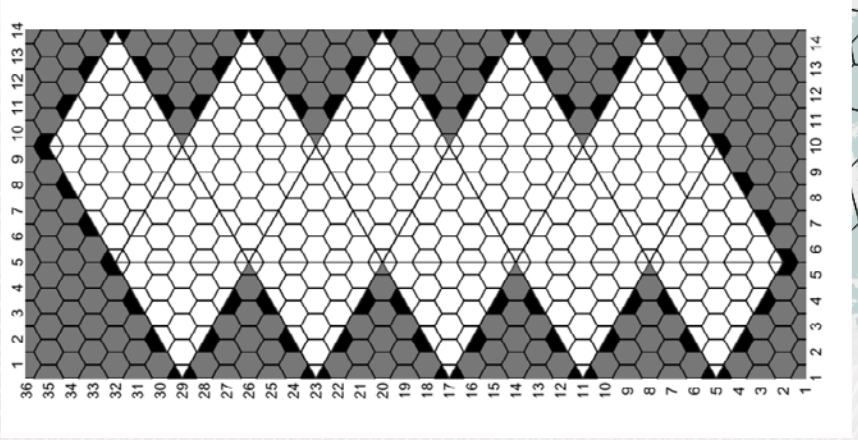
And there are other ways to bin if you really want to get into the dirty details

For instance, ISEA3H (Sahr, et al.)

But you lose the nice benefit of regular bins on Web Mercator

Different story if you want to use the Snyder Icosahedral Projection





Source: Carr et al. 1997