Week 4: Distributions

Fall 2017

Matthew Turk

Broadcasting

go.ischool.illinois.edu/meet2

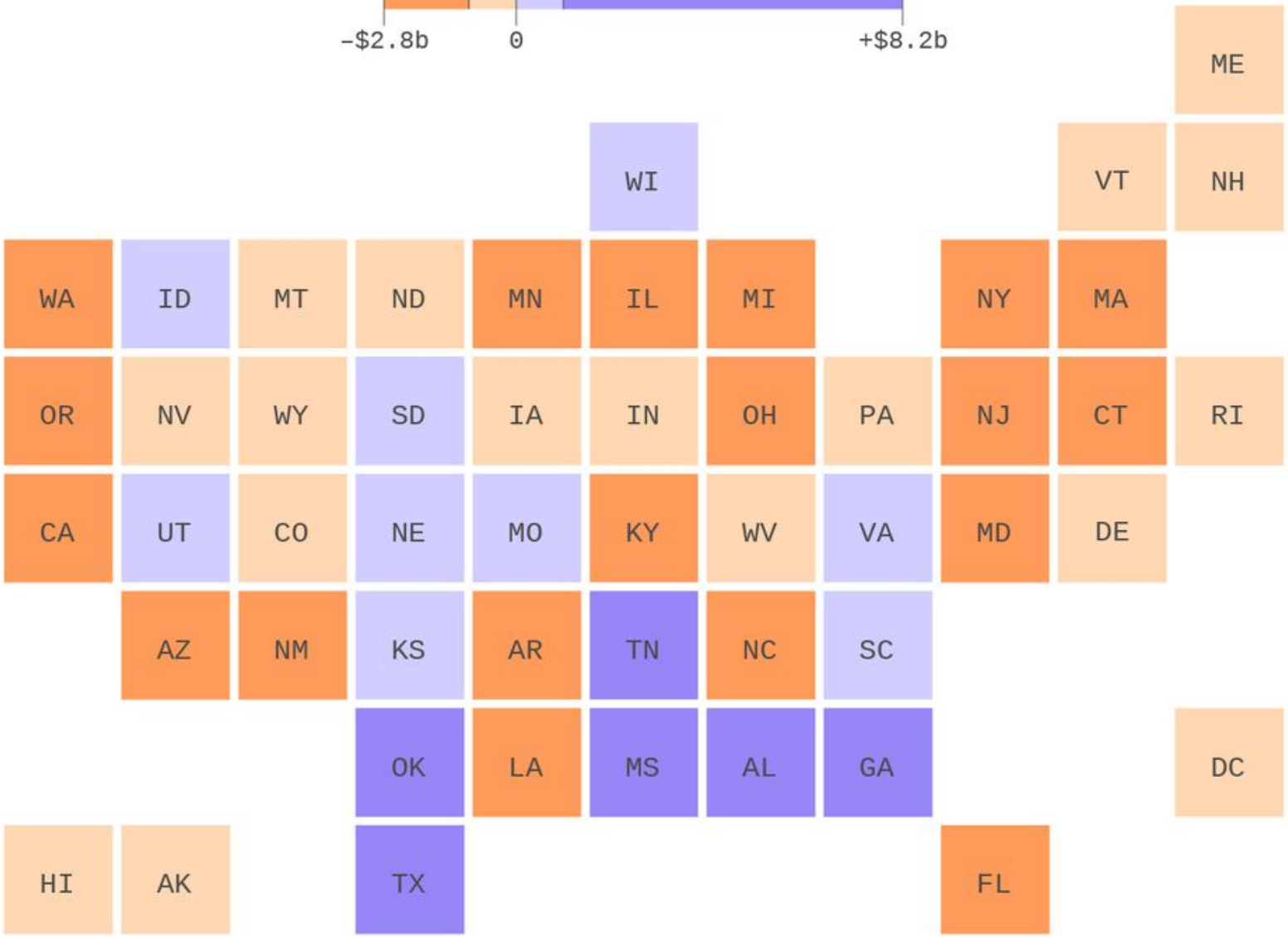
This Week

- Guest Discussion: Jill Naiman
- Aggregating Distributions
  - Principles
  - Weighting and Averaging
  -

Warm-Up Activity

1. What is the visualization trying to show?
2. What are its methods?
3. What are the strengths / weaknesses?
4. (Bonus) How was the data collected?

Estimated federal funding change in 2026



Andrew Witherspoon, Axios

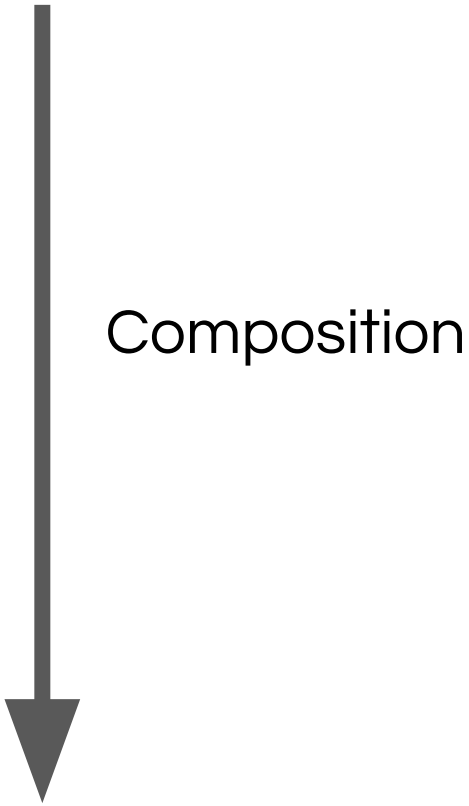
https://www.axios.com/vitals-2486505861.html

GitHub

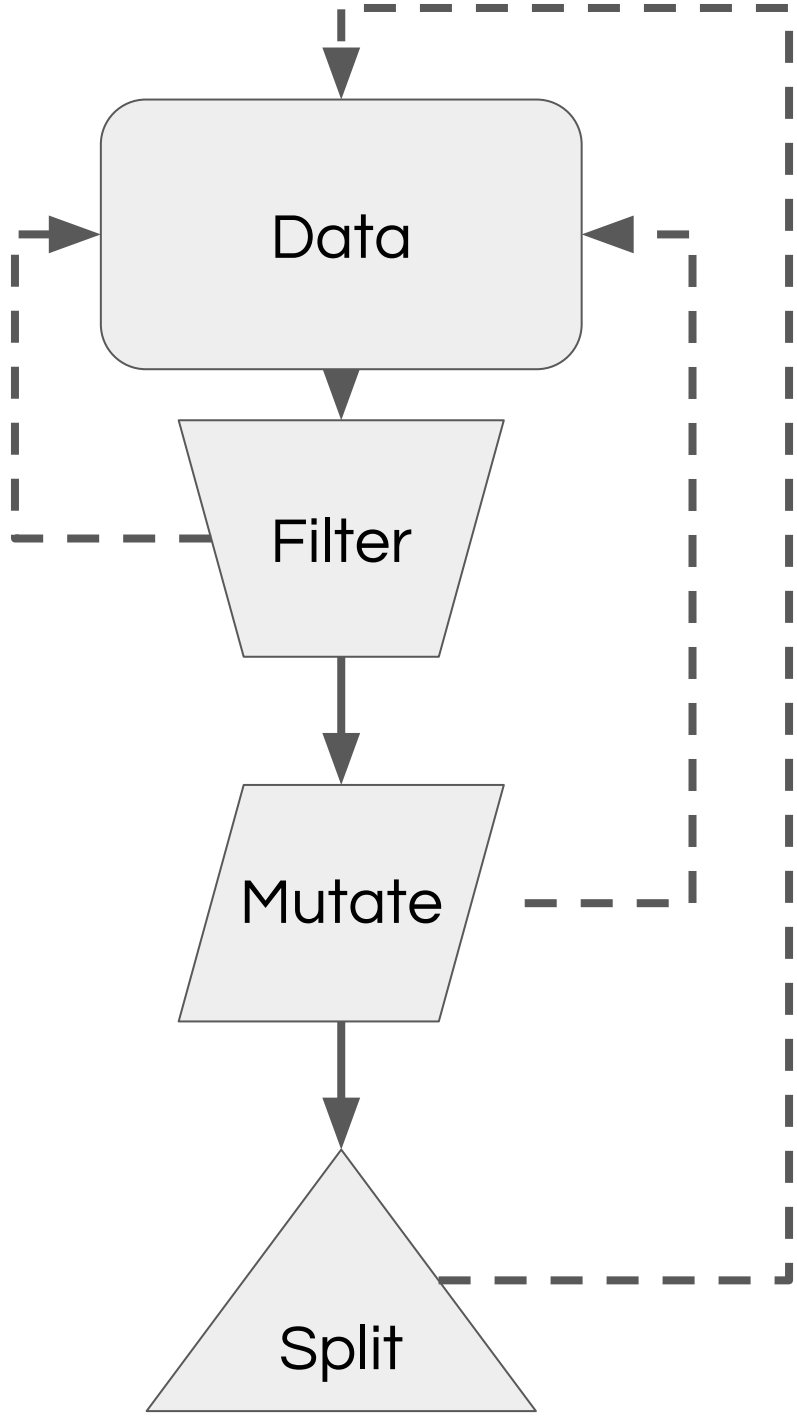
- Reminder: http://github.com/UIUC-iSchool-DataViz/fall2017/
- Lecture notes will be placed there, and available in your JupyterHub instances in data-readonly/fall2017/weekXX .
- Copy the notebooks to your directory before using them.

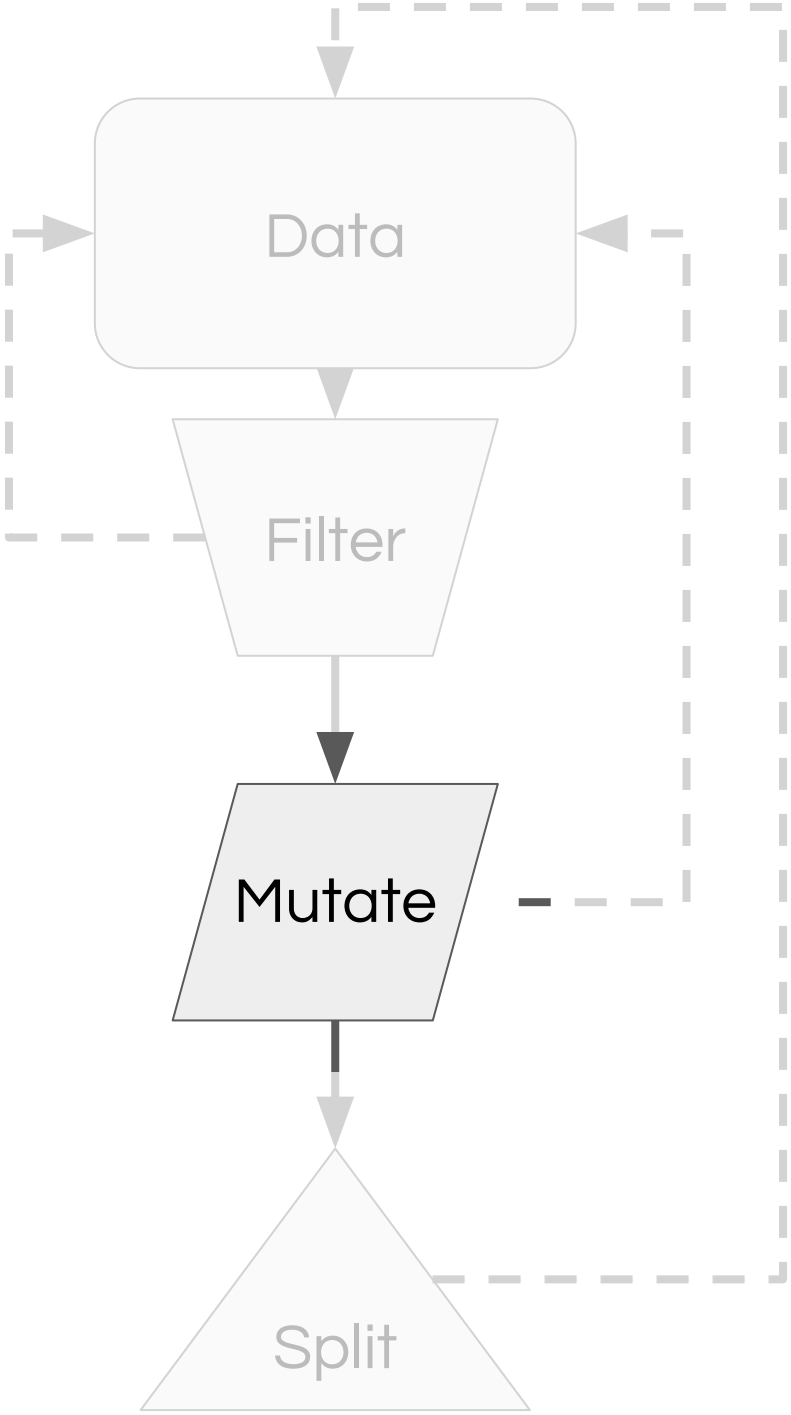
Concepts of Visualization

Concepts of Visualization



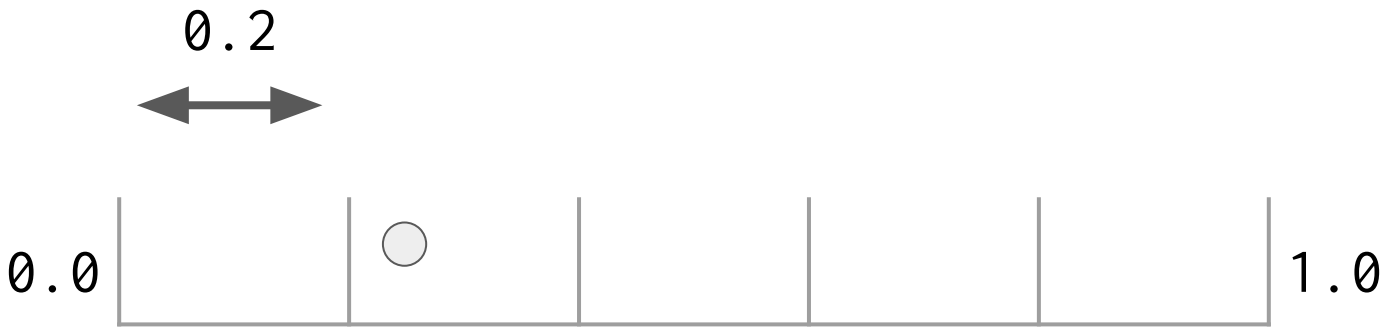
Concepts of Visualization


Data
Filter
Mutate
Split


Data
Filter
Mutate
Split


0.0
1.0


0.0
1.0

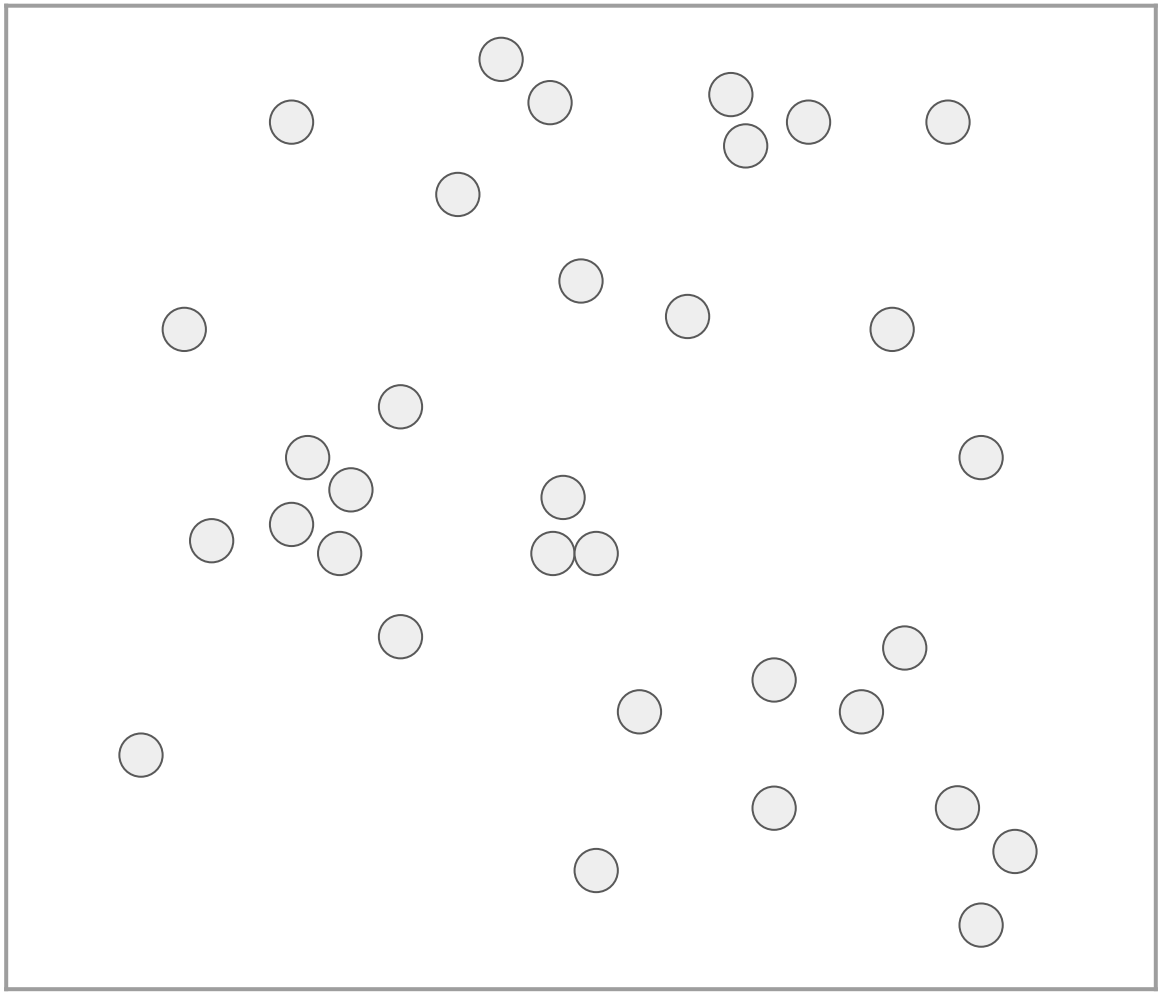

0.2
0.0
1.0

Uniform-width bins:

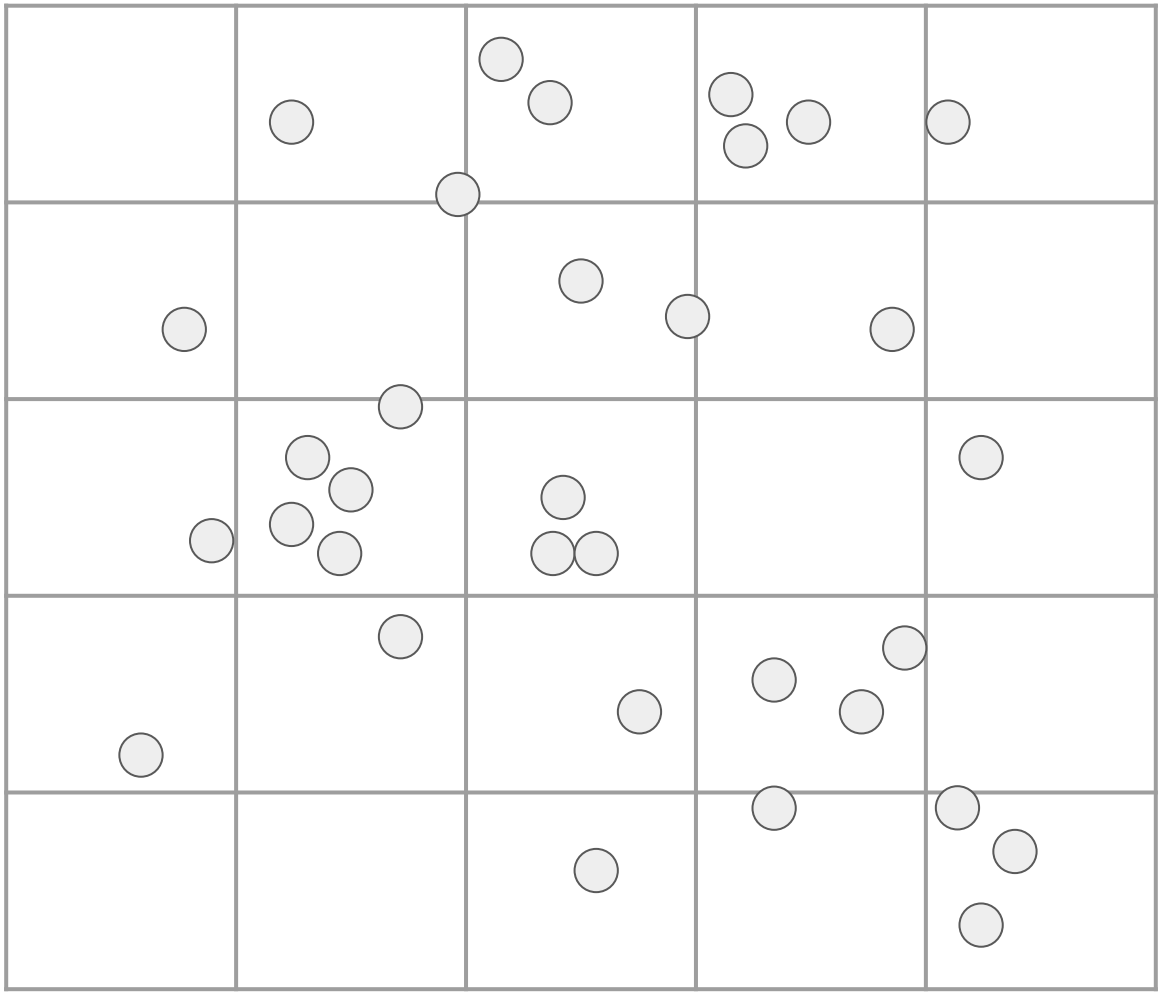
```
bin_id = floor( (value - left_edge) / bin_width )
```

Variable-width bins require searching.

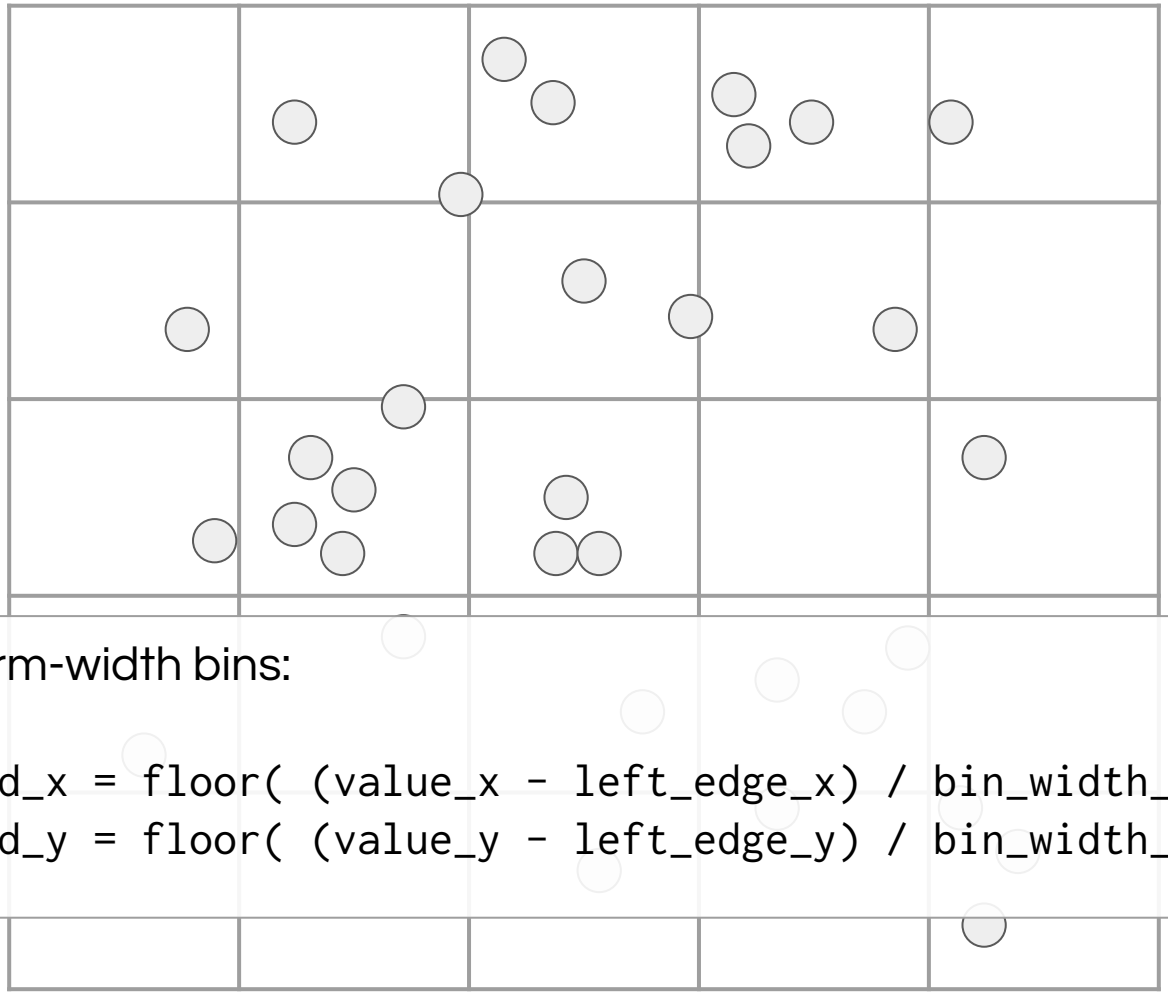
Binning and histograms



Binning and histograms



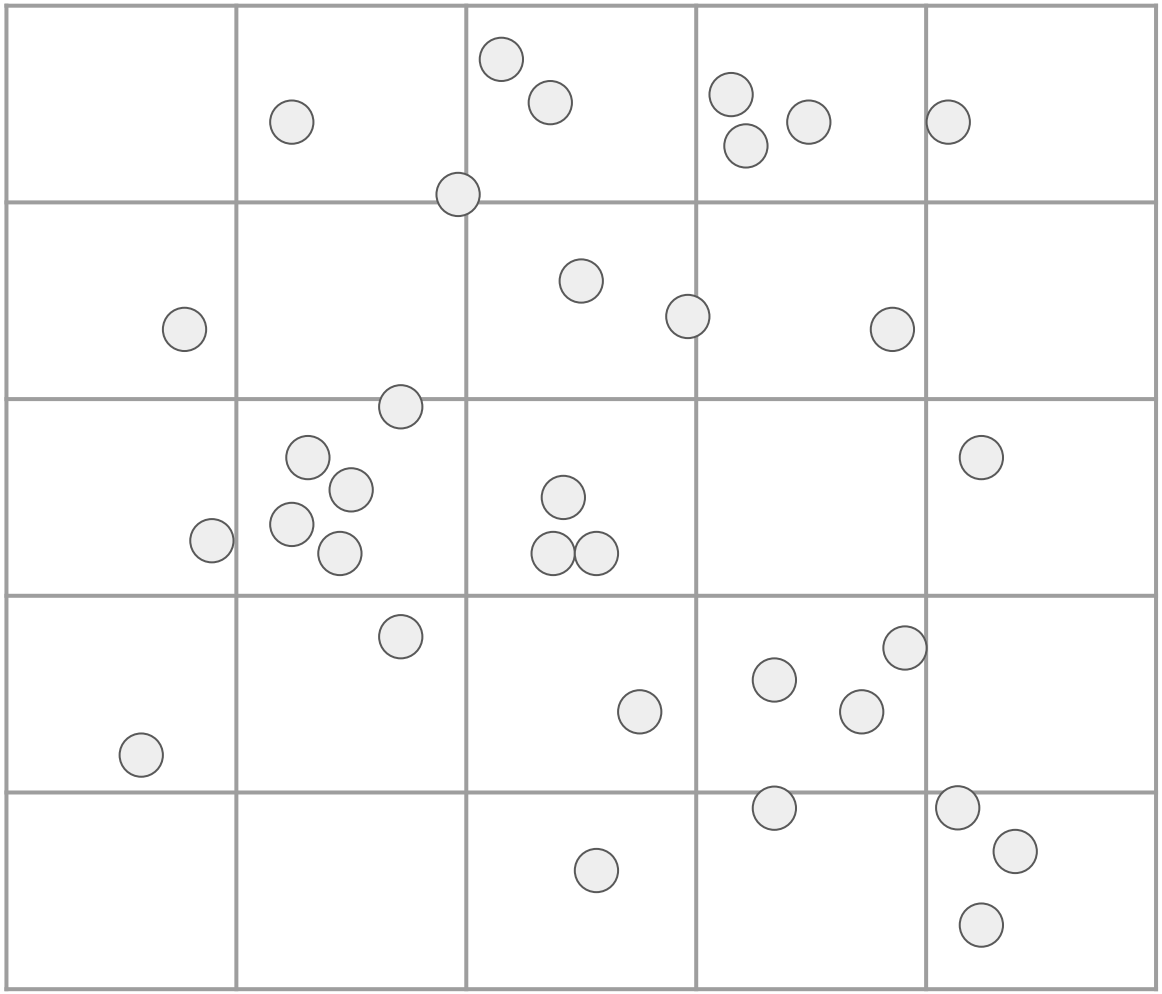
Binning and histograms



Uniform-width bins:

```
bin_id_x = floor( (value_x - left_edge_x) / bin_width_x )
bin_id_y = floor( (value_y - left_edge_y) / bin_width_y )
```

Binning and histograms



- Counts

  sum(1)
- Sum

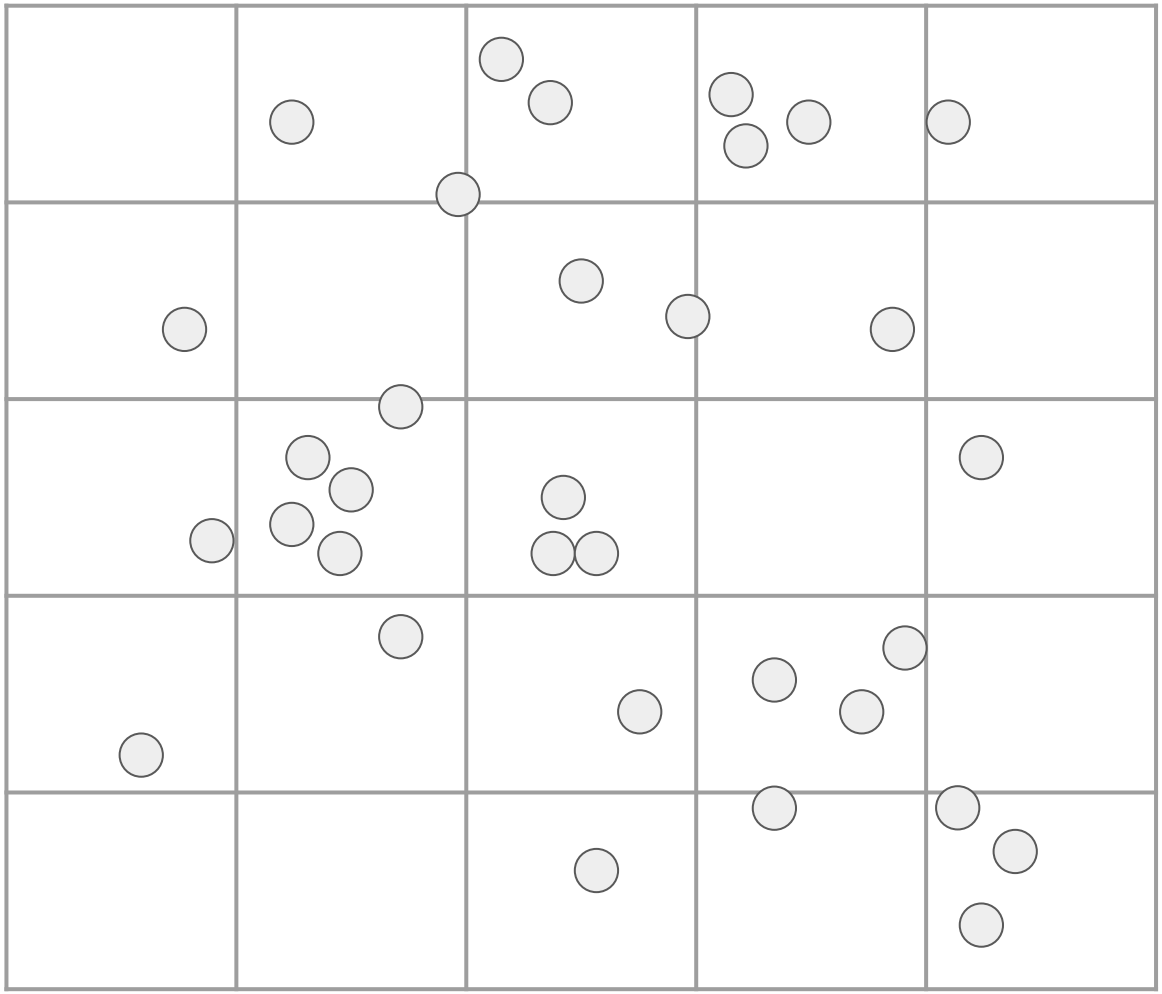
  sum(v_i)
- Average

  sum(v_i)/sum(1)
- Weighted Average

  sum(v_i * w_i) / sum(w_i)

Counts

`sum(1)`

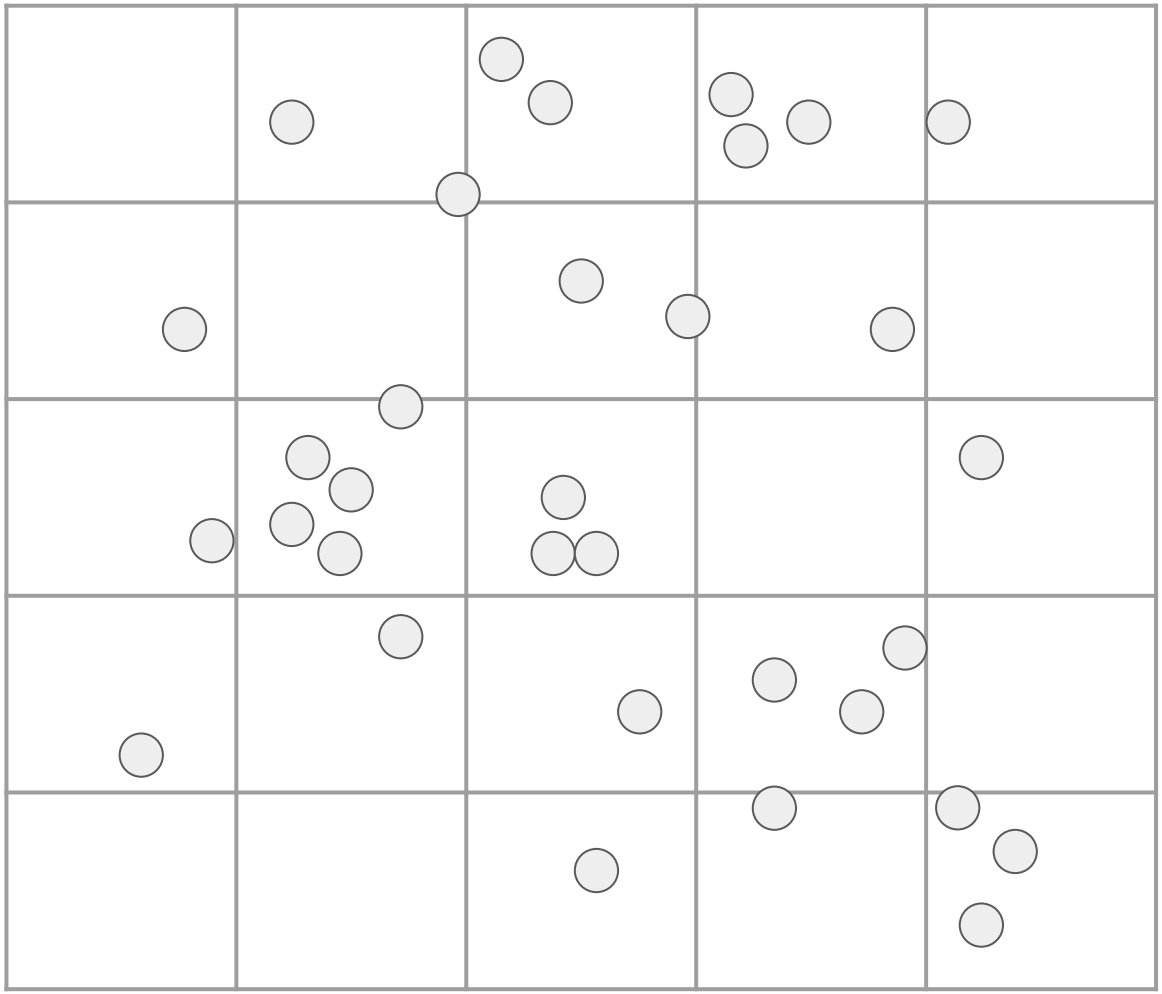


Useful for describing distribution of samples.

- “Number of times the weather station recorded the weather that day”
- “How many times was a UFO seen in the area?”
- “How many voting precincts in this state?”

Sum

$$\texttt{sum}(v_i)$$

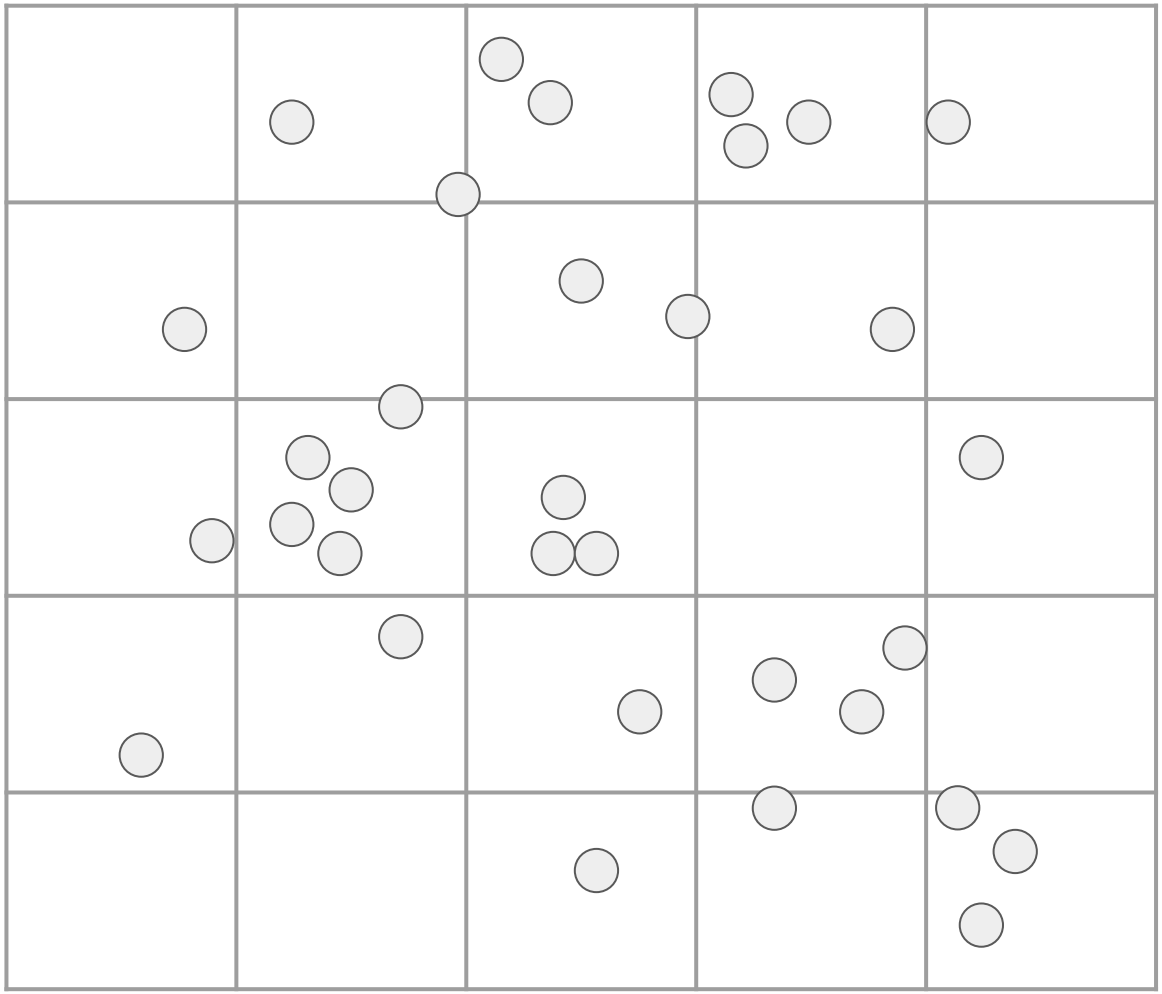


Useful for describing total quantity measured.

- “Inches of rain.”
- “Total time of recorded UFO sitings in the area.”
- “How many votes were cast?”

Average

$$\texttt{sum(v\_i)/sum(1)}$$

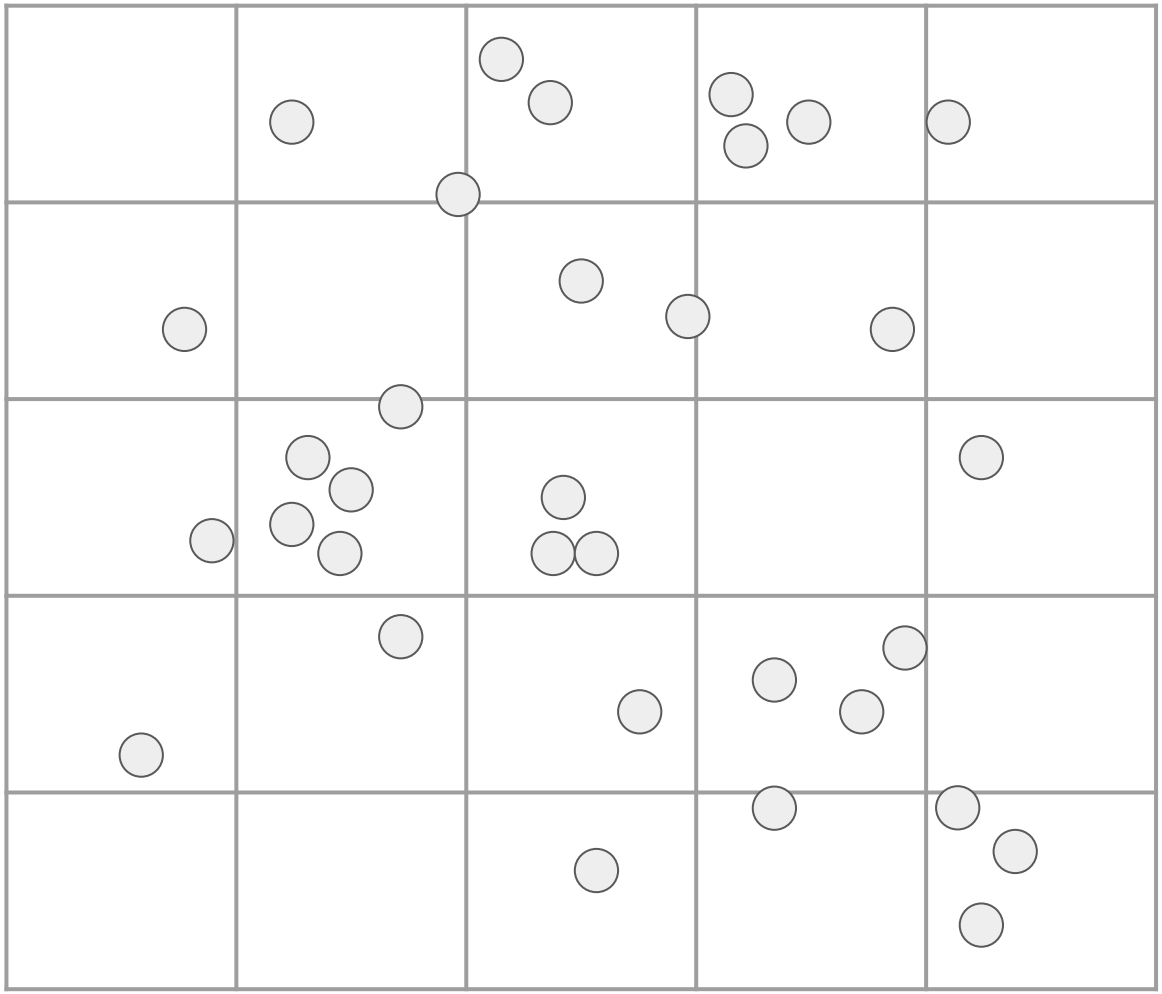


Useful for describing average or mean quantity.

- “Average rainfall in the area.”
- “Average time of UFO sighting.”
- “Who was the average candidate?”

Weighted Average

`sum(v_i*w_i)/sum(w_i)`



Useful for describing mean, but not strict arithmetic mean.

- “Average rainfall, weighted by how humid it was that day.”
- “What was the most commonly seen UFO type, as a function of the time it was seen?”
- “What’s the mean age of a voter, as a function of the total years of experience in the election?”

https://lis590.ncsa.illinois.edu/