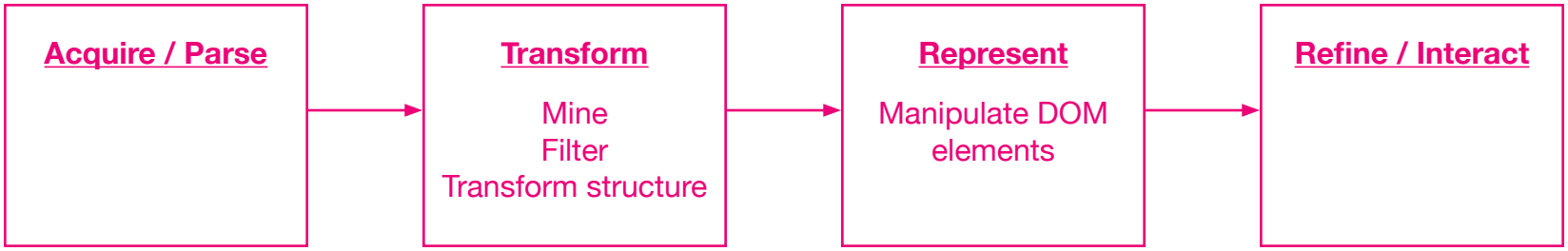
Week 6

First Visualization: Scatterplot

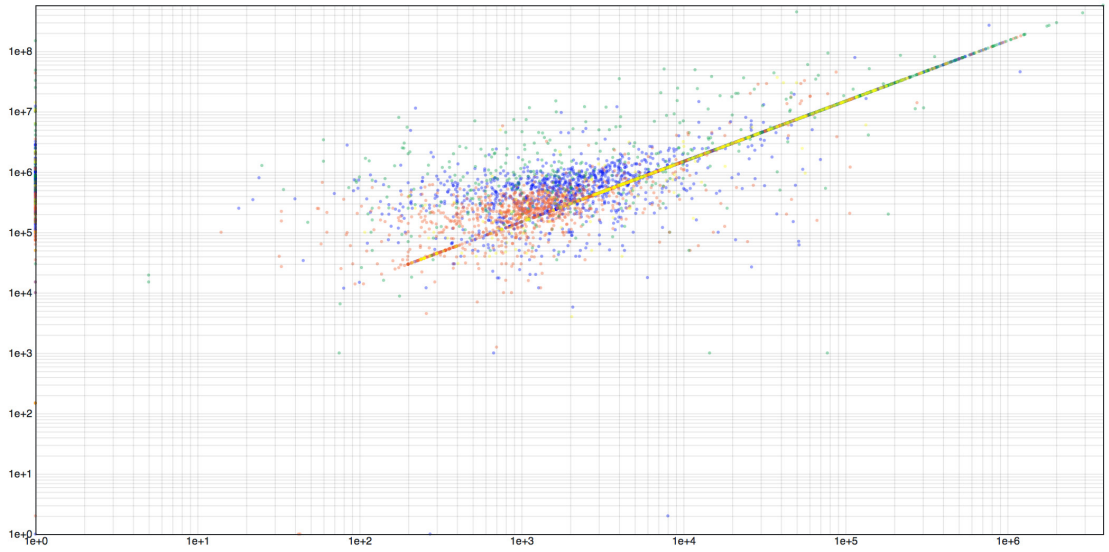
A general “algorithm” for data visualization



The goal today is to work through this problem-solving pipeline in the context of a real-world dataset.

At each step, focus on not only the how (JavaScript implementation), but also why.

Project planning: the scatterplot



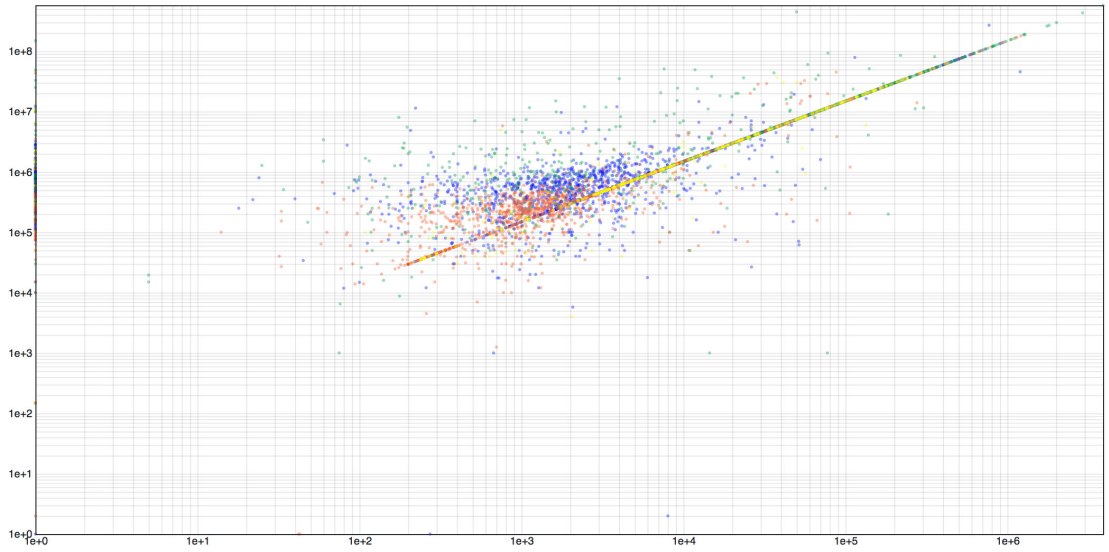
Why

The scatterplot allows us to visually explore correlation between variables.

An important exploratory data visualization tool

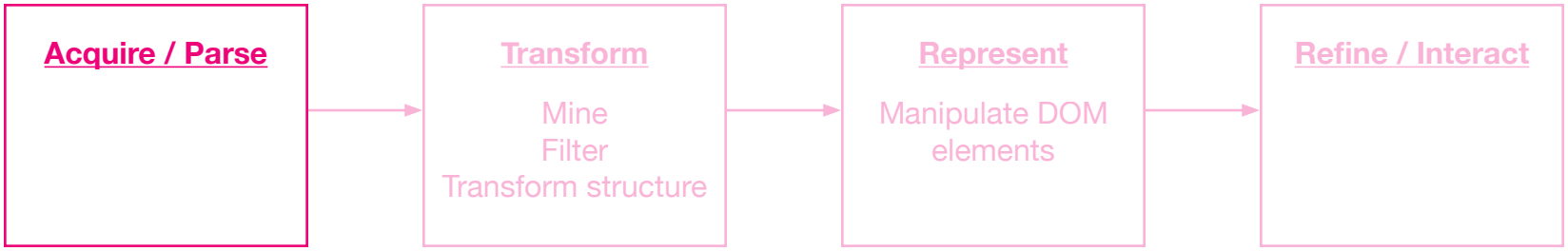
How does it apply to our dataset (NYC construction permits)?

Project planning: the scatterplot



Which variables do we choose to represent on the two axes, and why?

Looking at the example to the left, how do you think we can implement it? Can you draw up a basic algorithm for this?



Goal

Import data into the browser environment

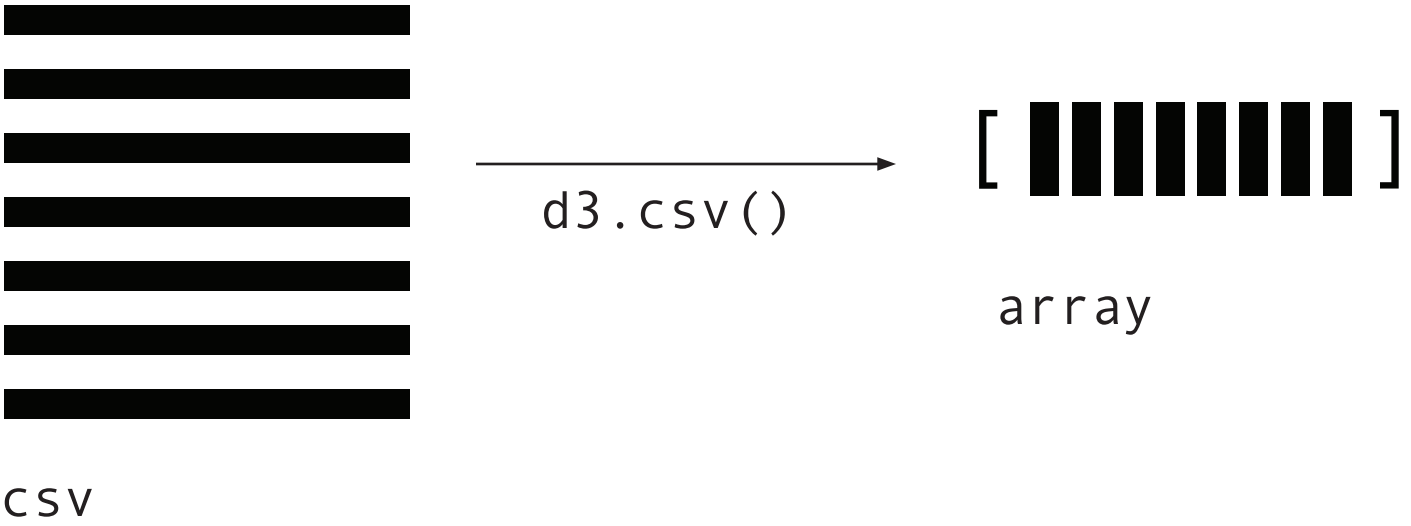
Clean up / parse data:
- Missing values
- Type conversion
- Renaming

Tool

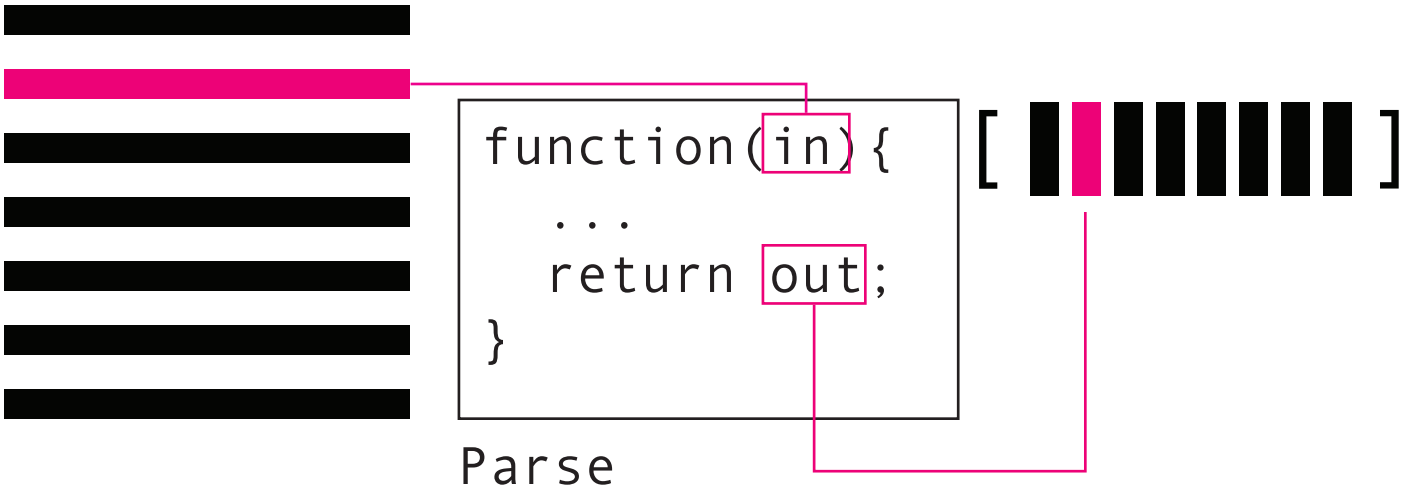
`d3-import` combined with a `Parse` function

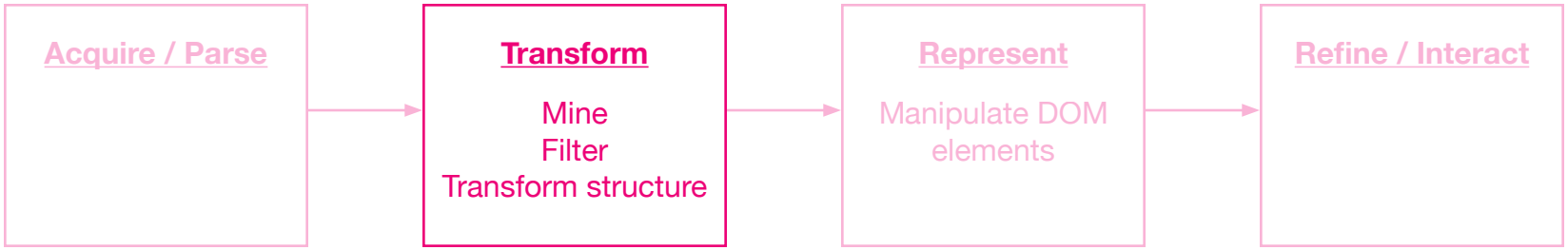
```
d3
.csv(url, parse)
.then(...)
```

d3.csv and parse function explained



d3.csv and parse function explained





Goal

Data mining, discovery, filtering, and transformation

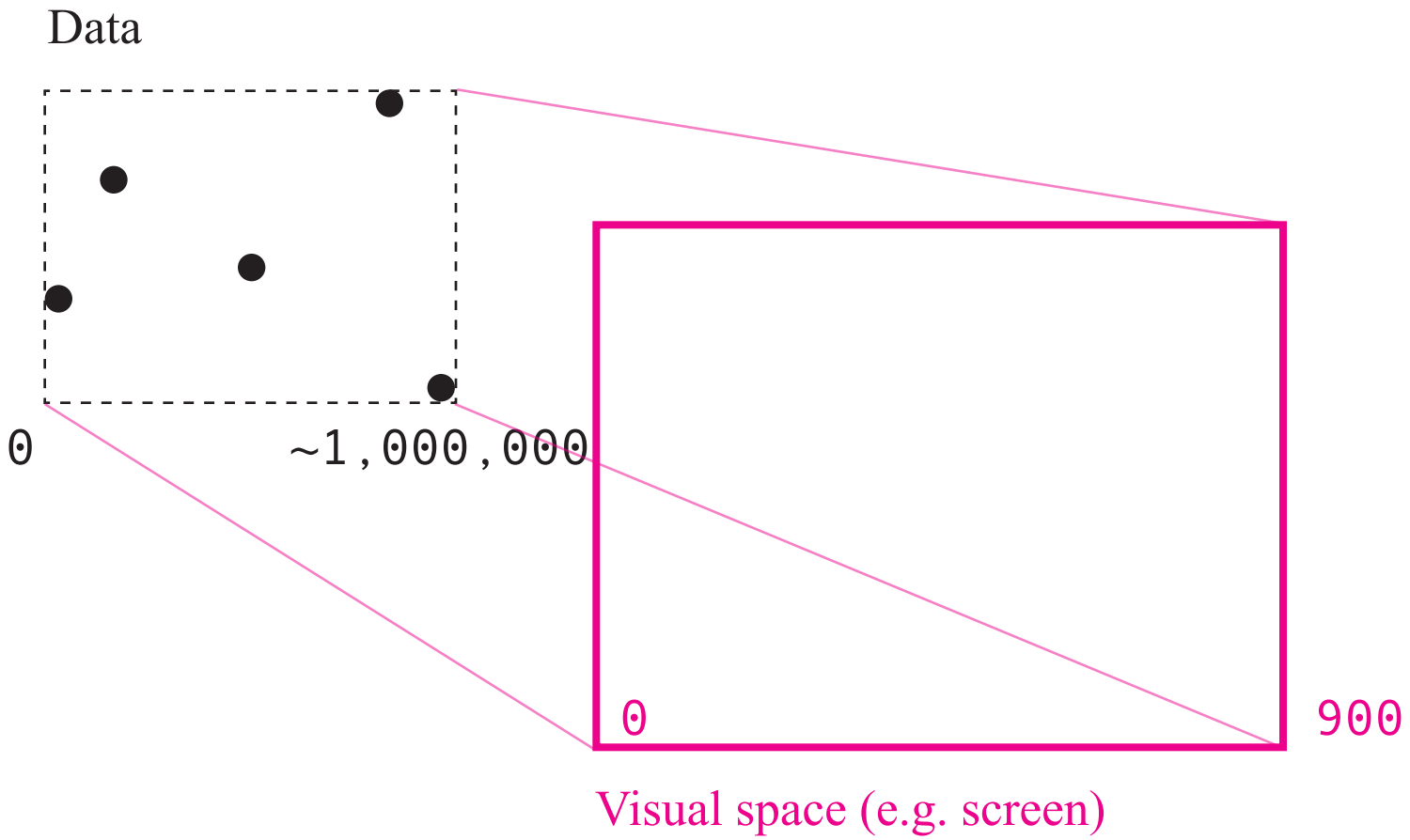
One example is “scaling” data

Tool

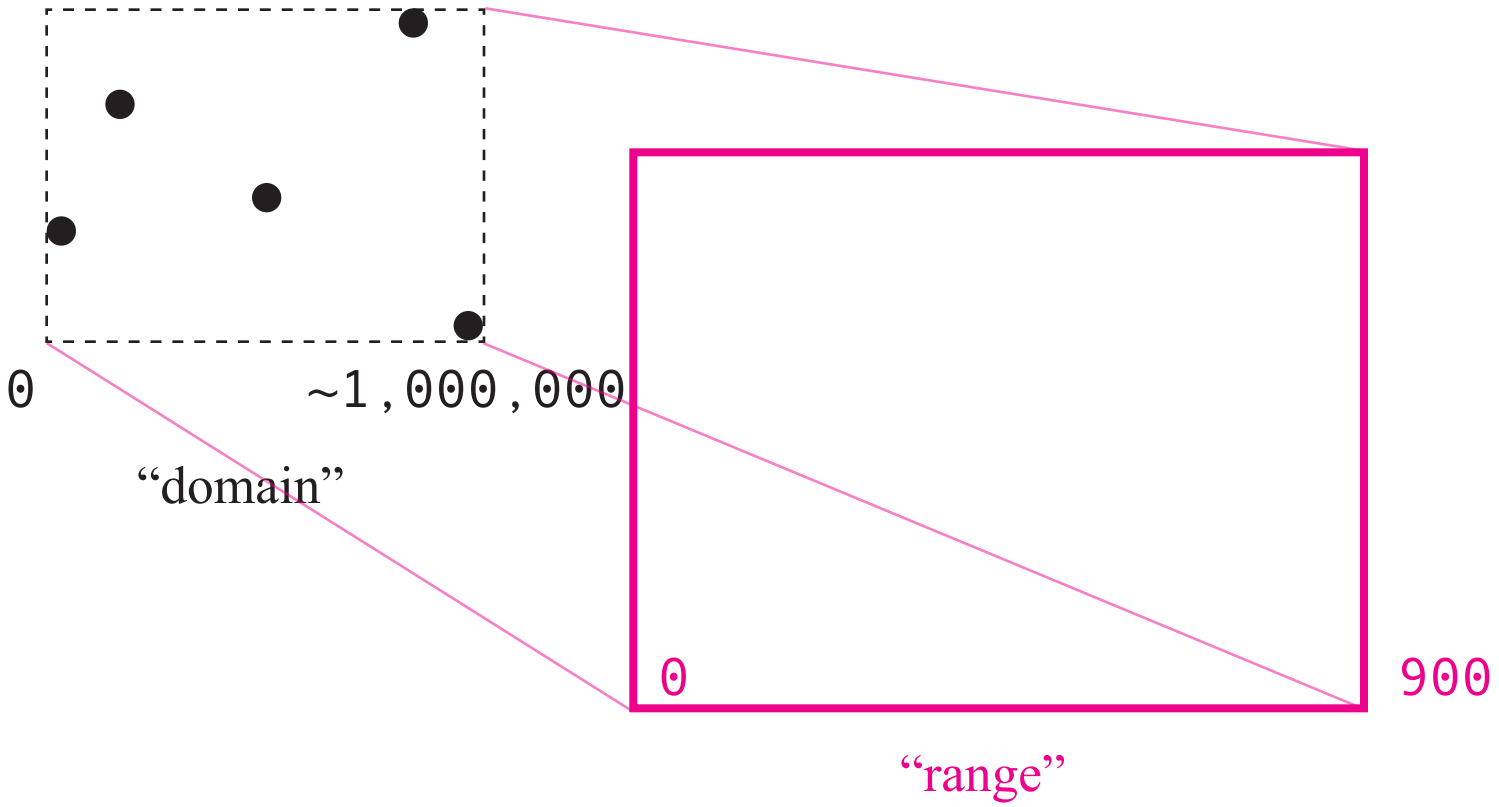
```
d3.min/max/extent
etc.

d3-scale
```

Why scaling data?



Why scaling data?



d3-scale

Setting up a scale

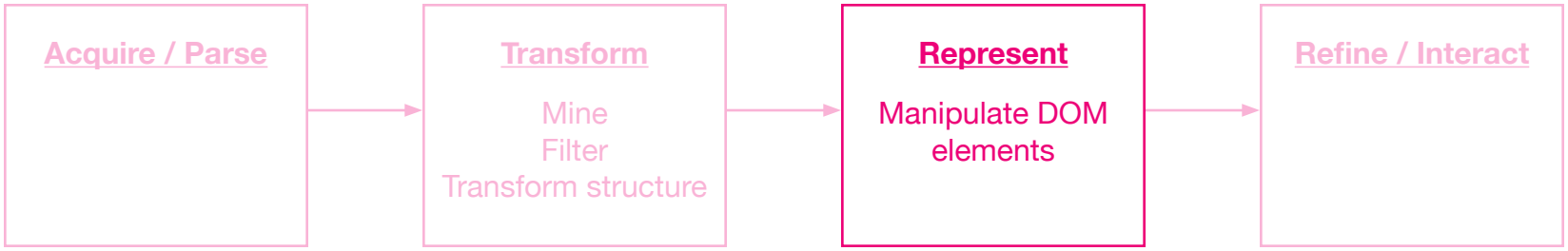
```
scale
    .domain([domain_min, domain_max])
    .range([range_min, range_max])
```

Transforming data from domain to range

```
x_range = scale(x_domain)
```

Transforming data from range to domain

```
x_domain = scale.invert(x_range)
```



Goal

Representing the transformed data using visual symbols

Tool

DOM manipulation using d3

- selection
- selection.append
- selection.attr

Review

1. What is a scatterplot? What is it useful for?
2. Review `d3.csv` and how the `parse` function works
3. Why do we need to parse data when we import it?
4. What are the potential kinds of data transformation we need to perform in a visualization problem?
5. Why do we need to scale data?
6. How does `d3.scale` work? What is domain, and what is range?
7. What's the difference between `d3.scaleLinear` and `d3.scaleOrdinal`?
8. How do we append a collection of visual symbols that correspond to a data array?