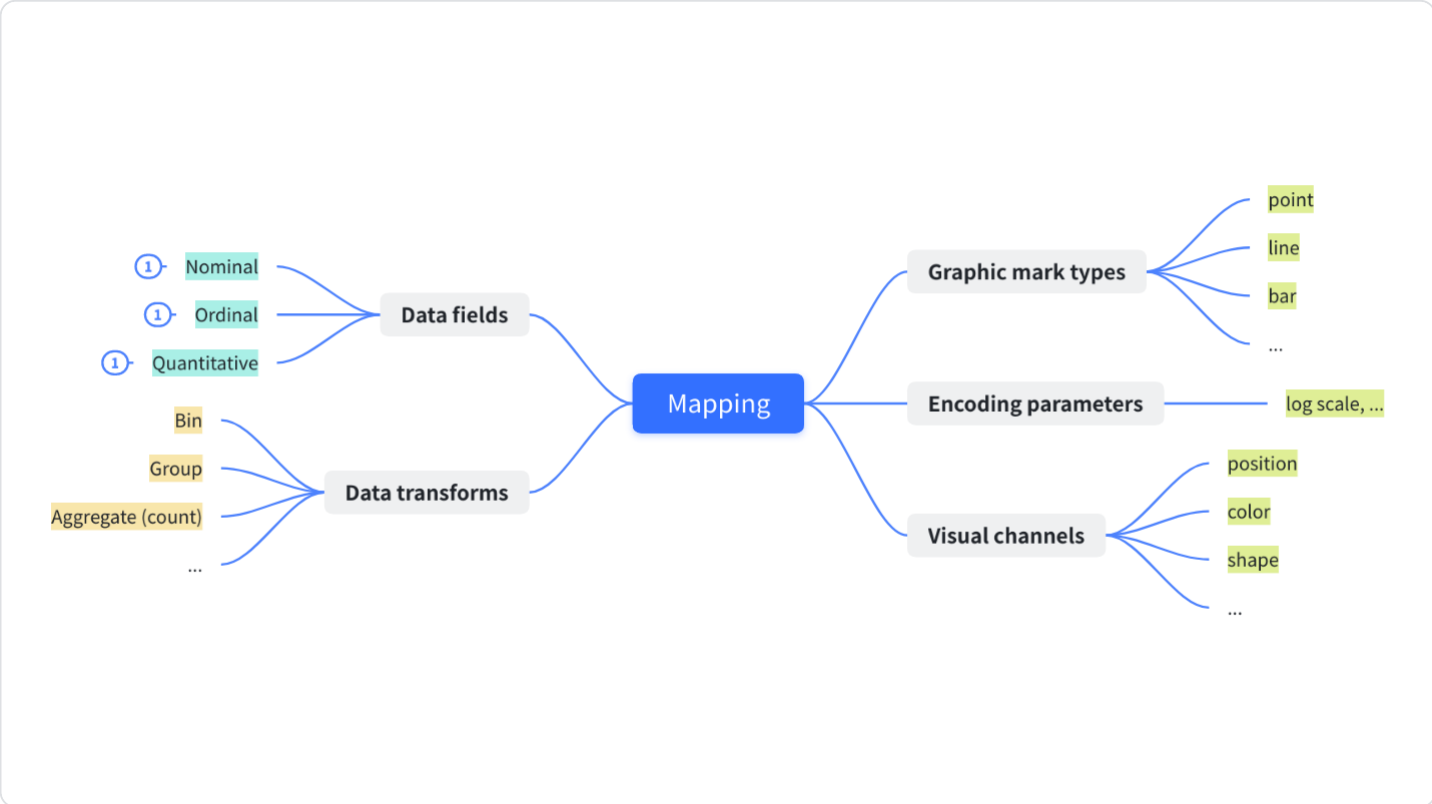
DS363 Visual Encoding Design

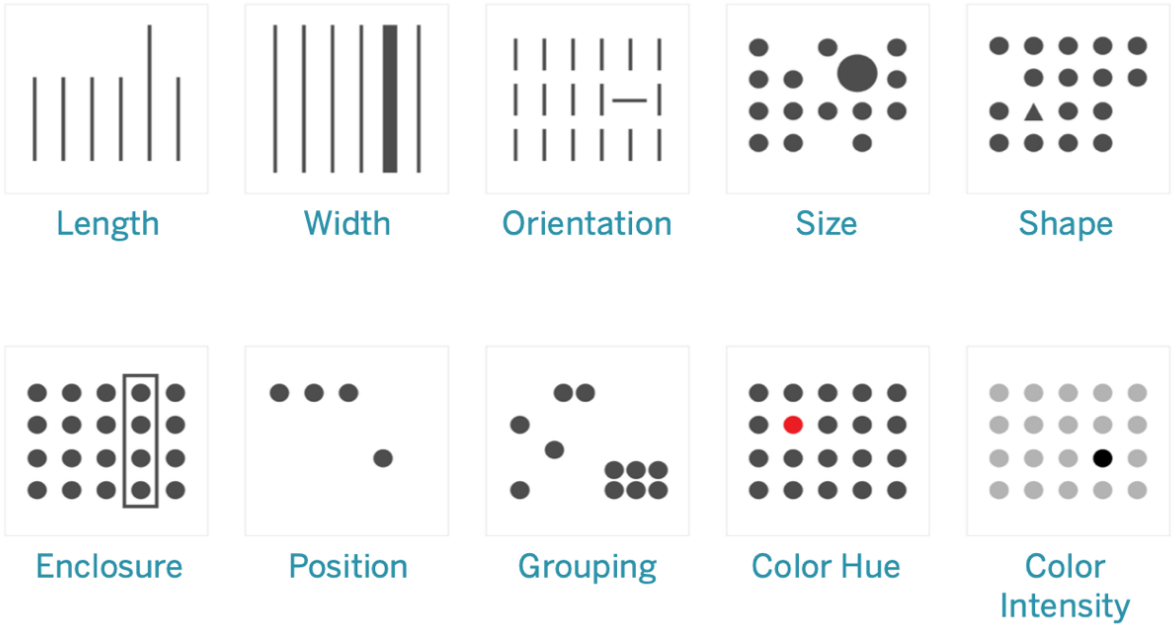
1. A design space of visual encoding

1.1 Mapping Data to Visual Variables

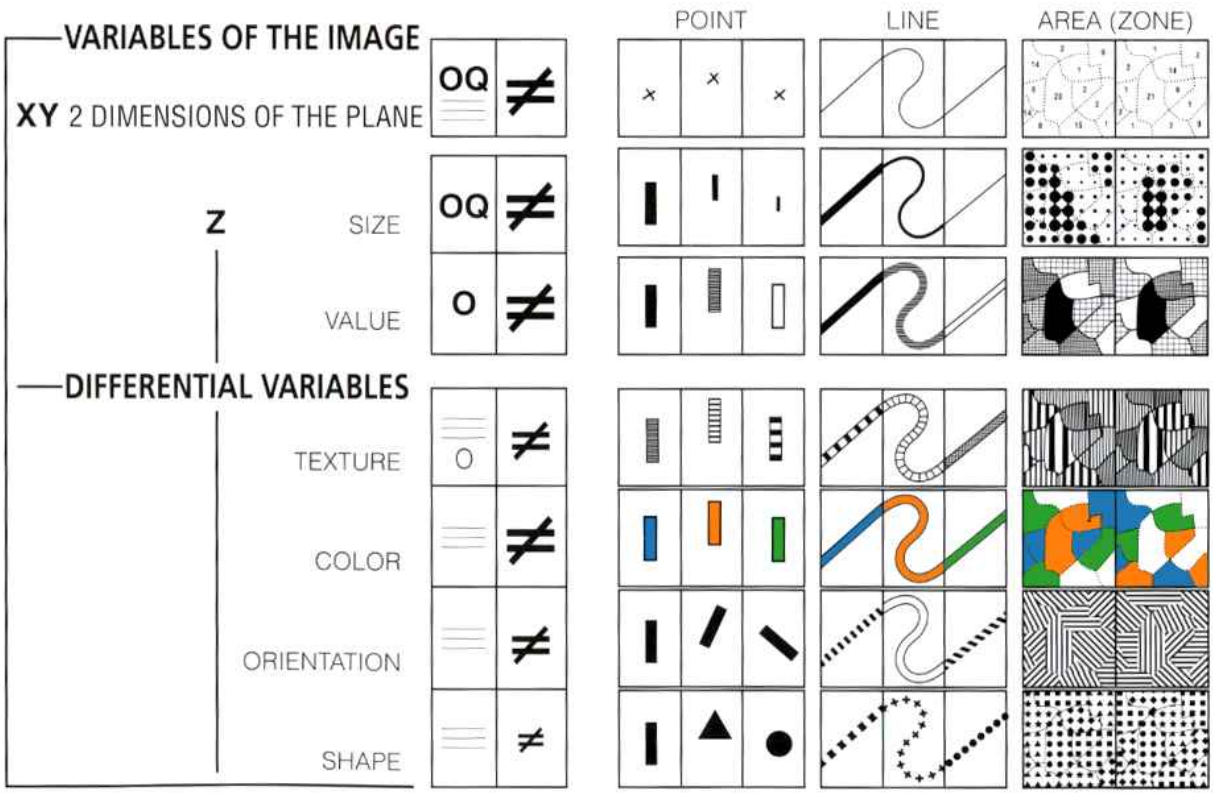
- Assign **data fields** to **visual channels** for a chosen **graphic mark type**
- Additional concerns include choosing appropriate encoding parameters and **data transform**
- These options define a large combinatorial space, containing both **useful and questionable** charts!



1.2 Visual Encoding Channels



1.3 Visual Channels with Graphic Mark Types

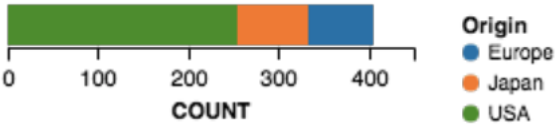


1.4 Practice on the website

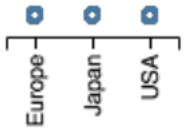
✏️ vega.github.io/voyager/

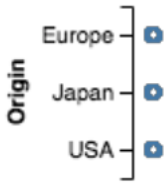
- Reproduce the following plots.
- Are they all effective?

1D: Nominal

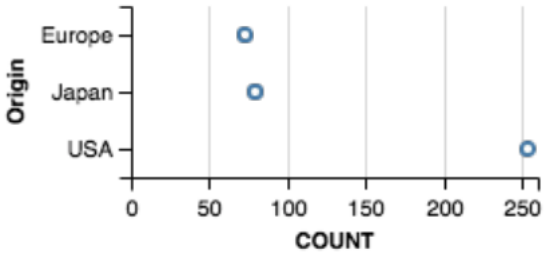


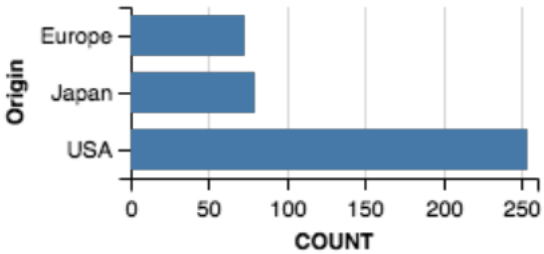
Raw

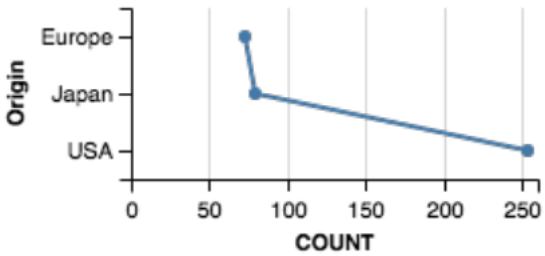


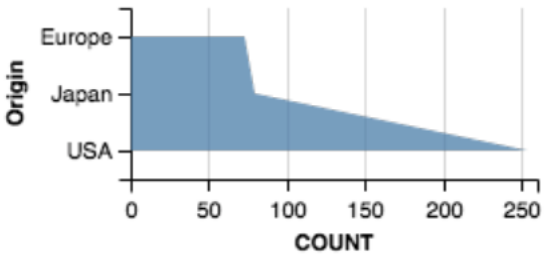


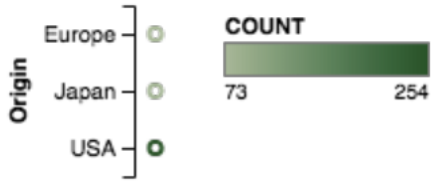
Aggregate (Count)

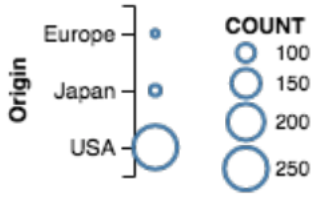






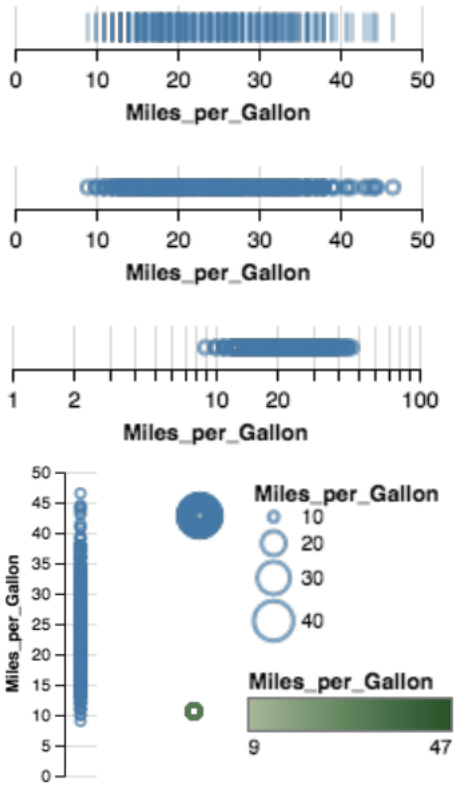




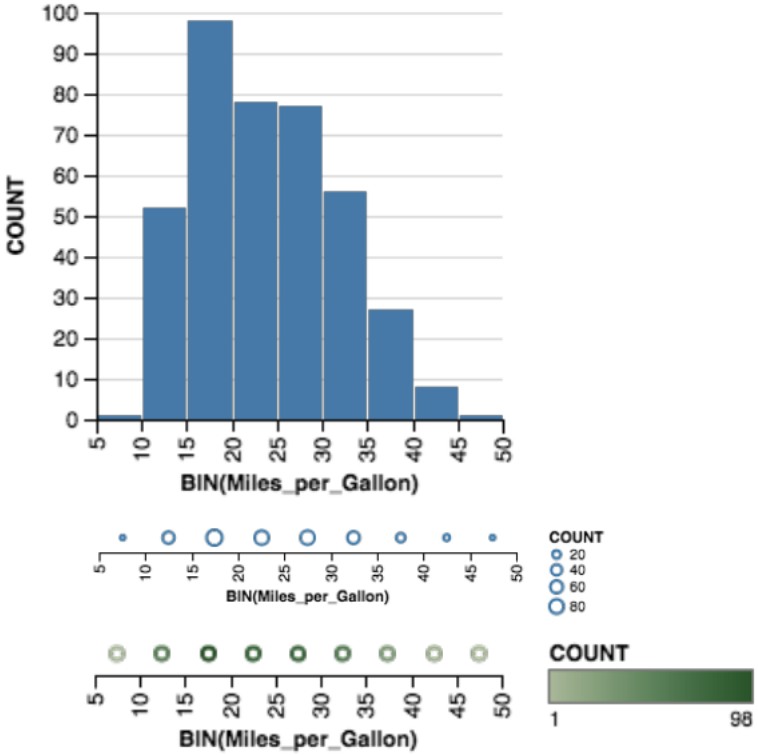


1D: Quantitative

Raw

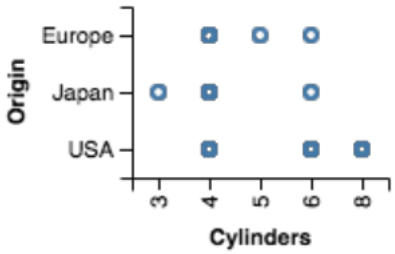


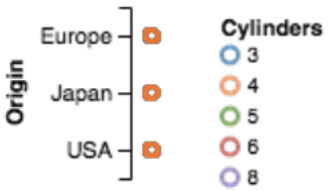
Aggregate (Count)

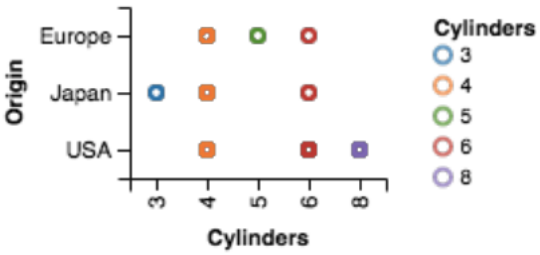


2D: Nominal x Nominal

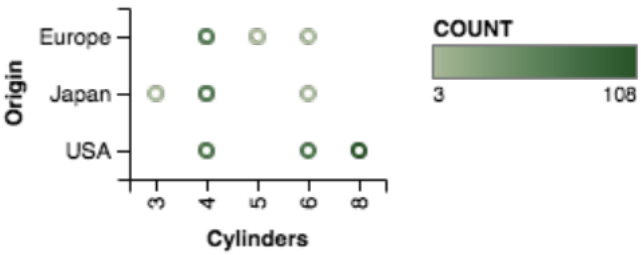
Raw

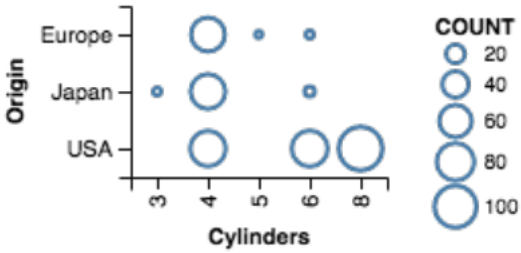


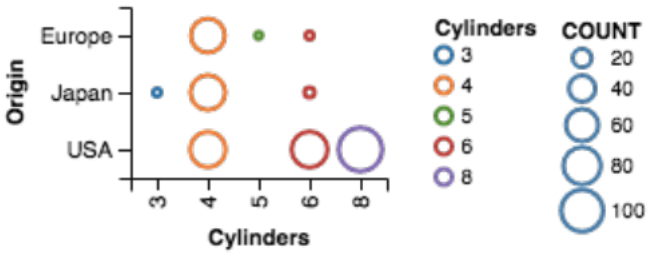




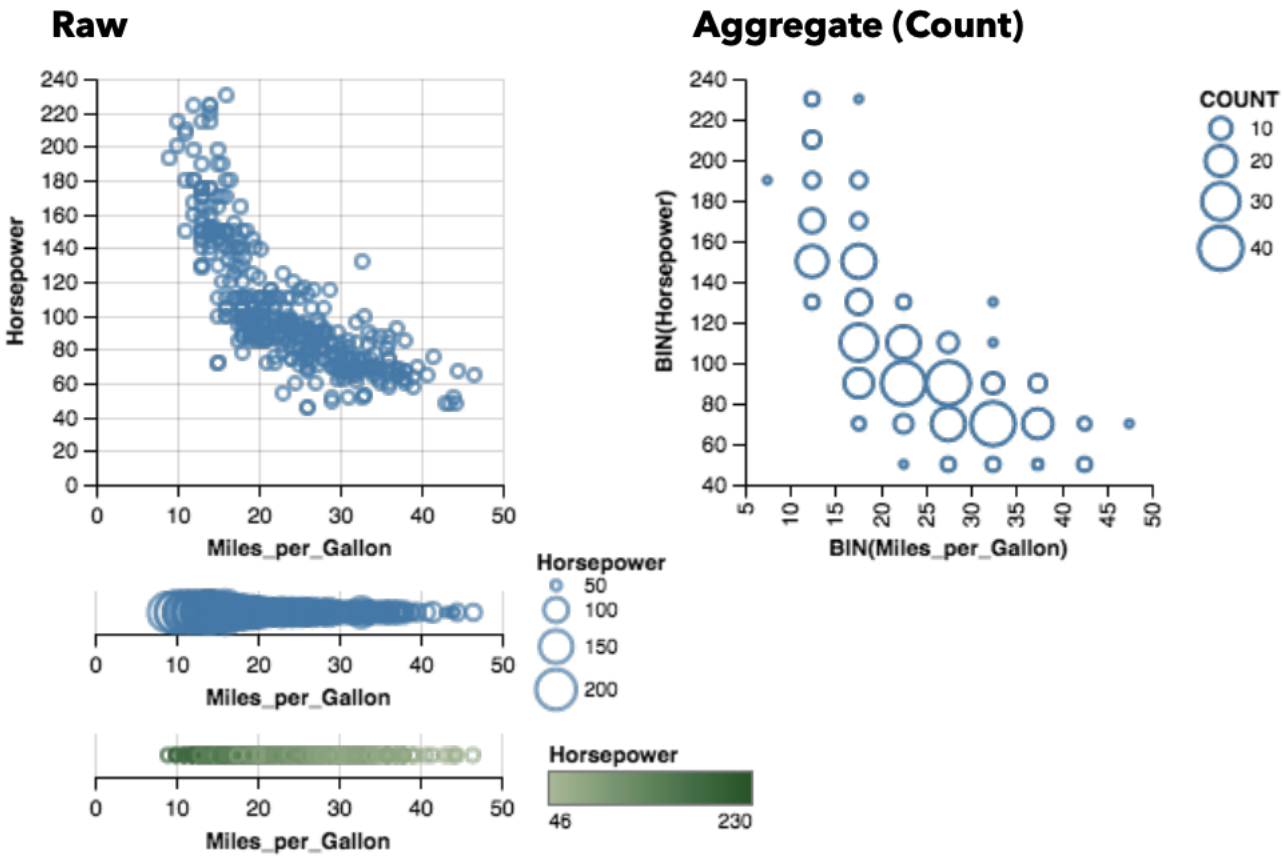
Aggregate (Count)

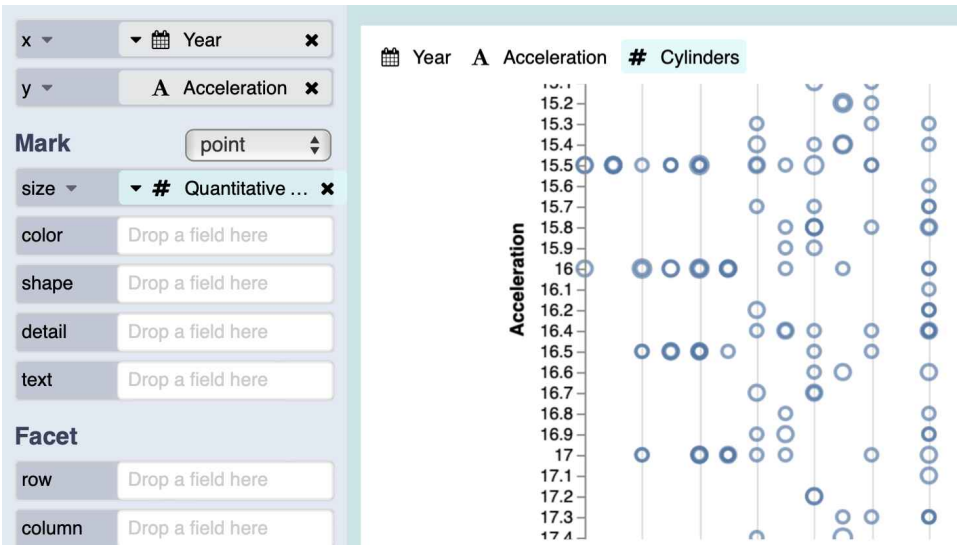


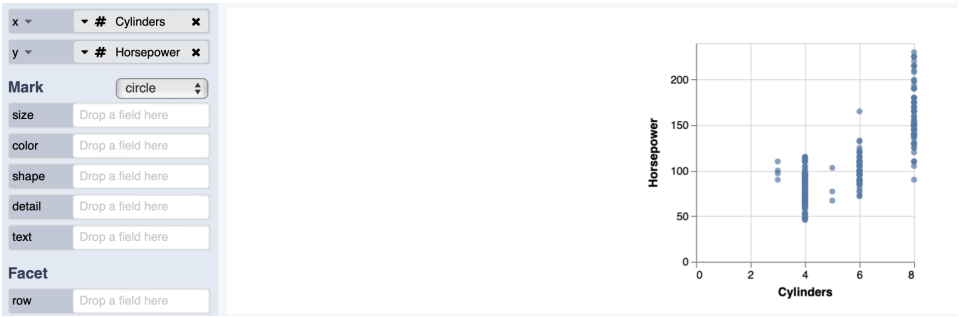


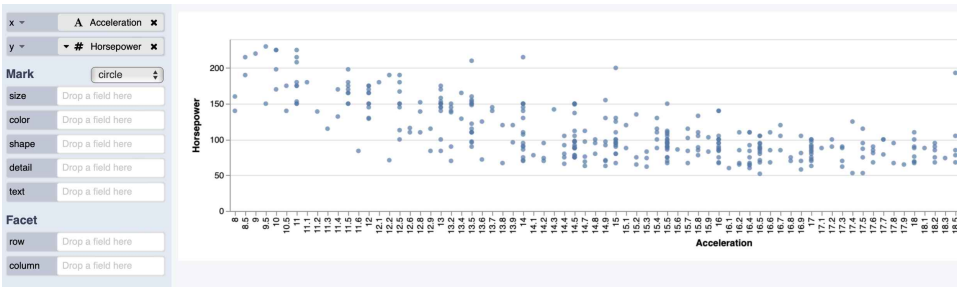


2D: Quantitative x Quantitative

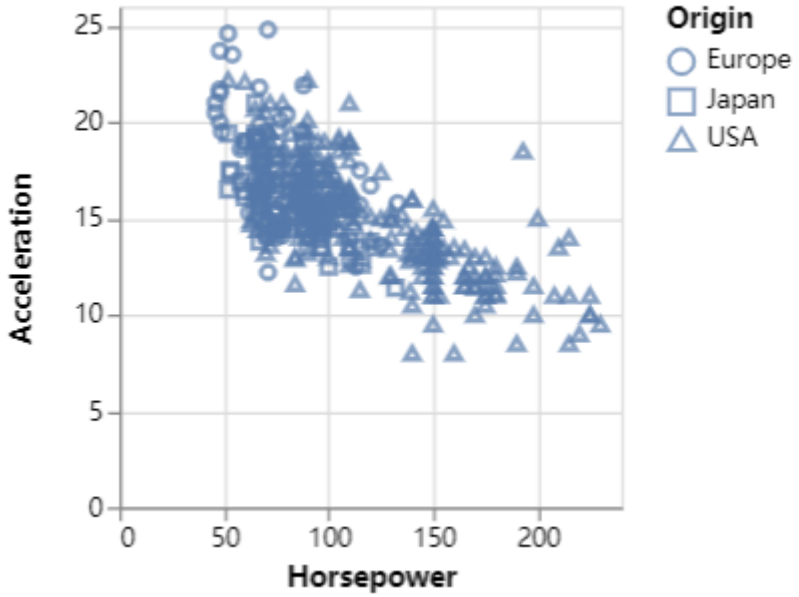


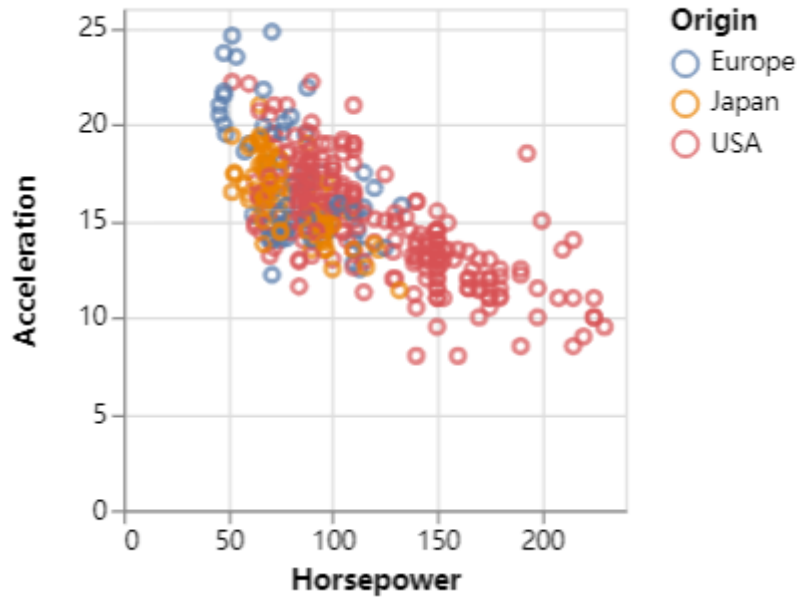
Name	Insert your plot in this column	
蒋含熙		





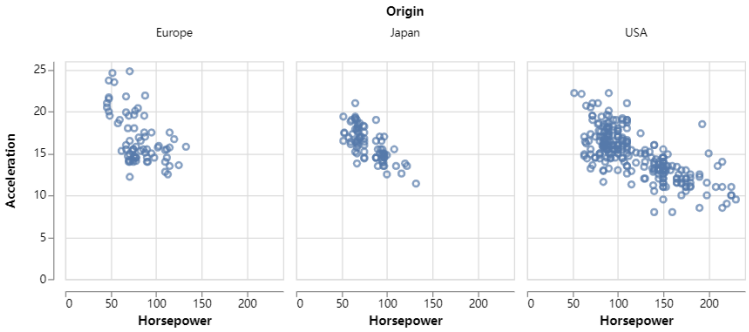
闫思源

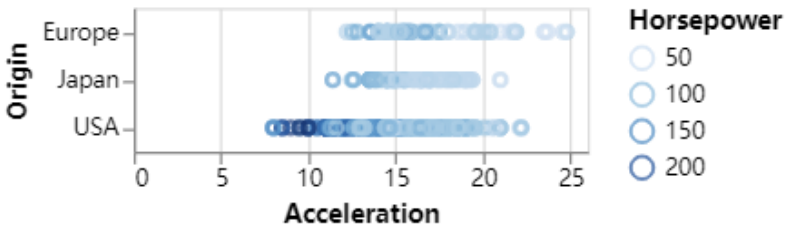


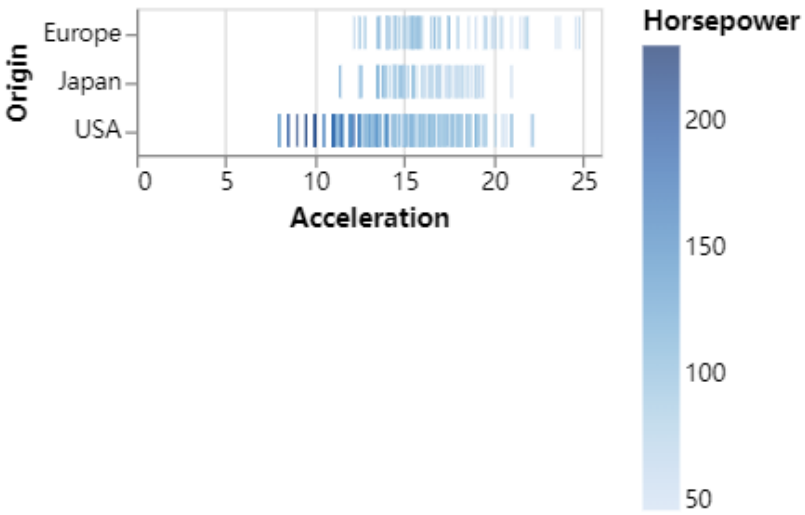


Quantitative × Quantitative × Nominal

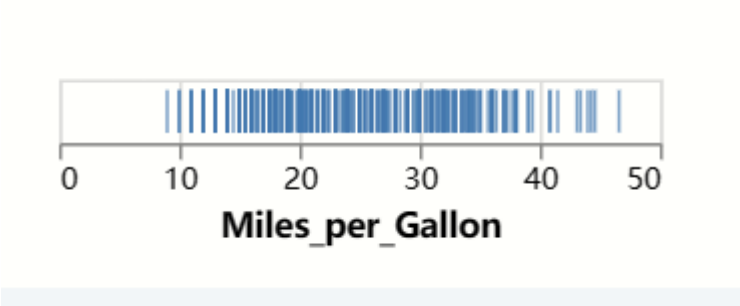
三维数据太密集的时候，效果：Facet > Color > Shape, Size



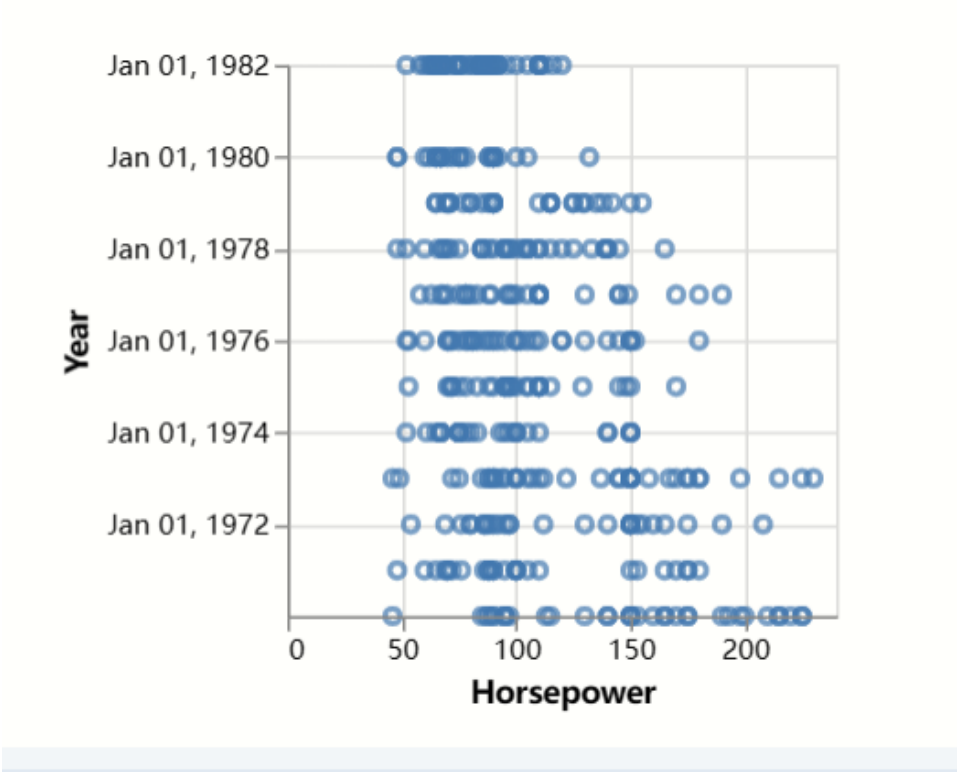




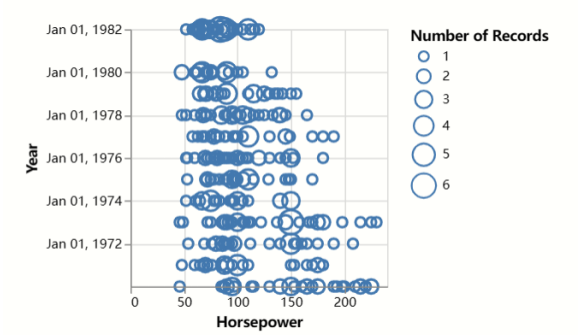
唐子麓

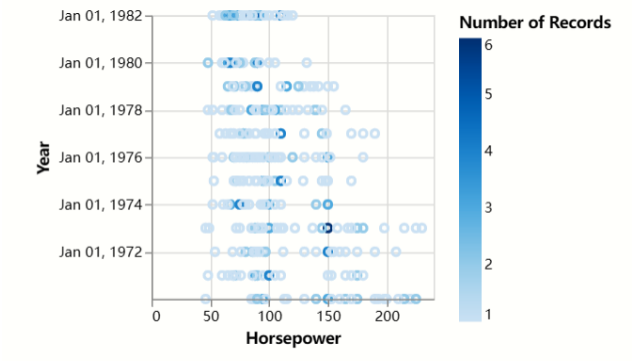


通过密集程度可判断Miles_per_Gallon集中的量（即一般来说生产出的汽车的miles_per_gallon值的大小集中在的区域）

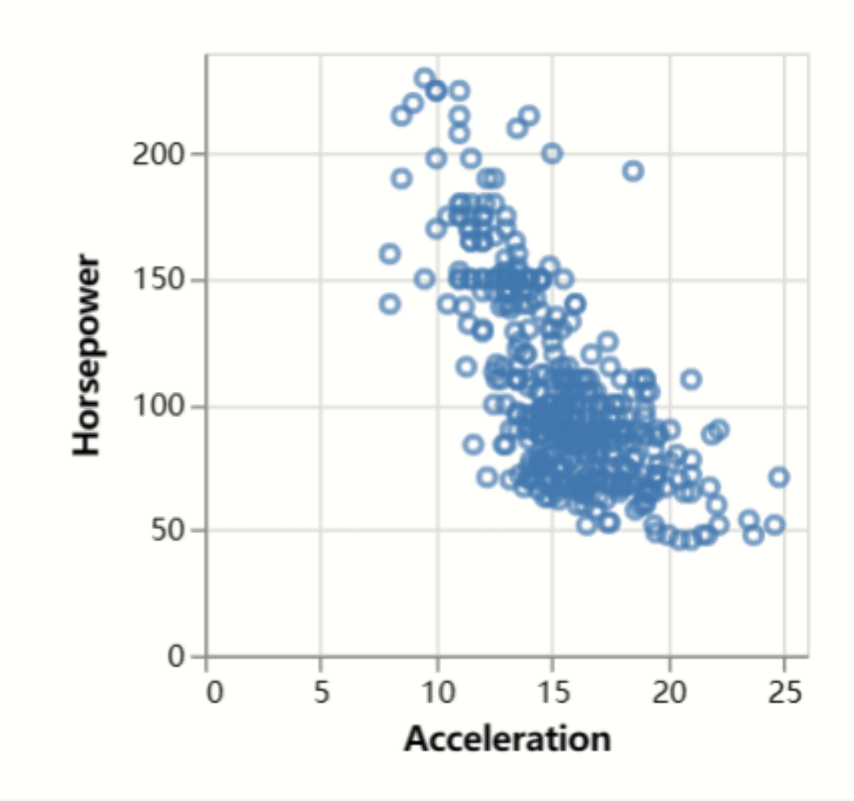


感觉按时间来说，是有一个最大马力在减小的趋势的。

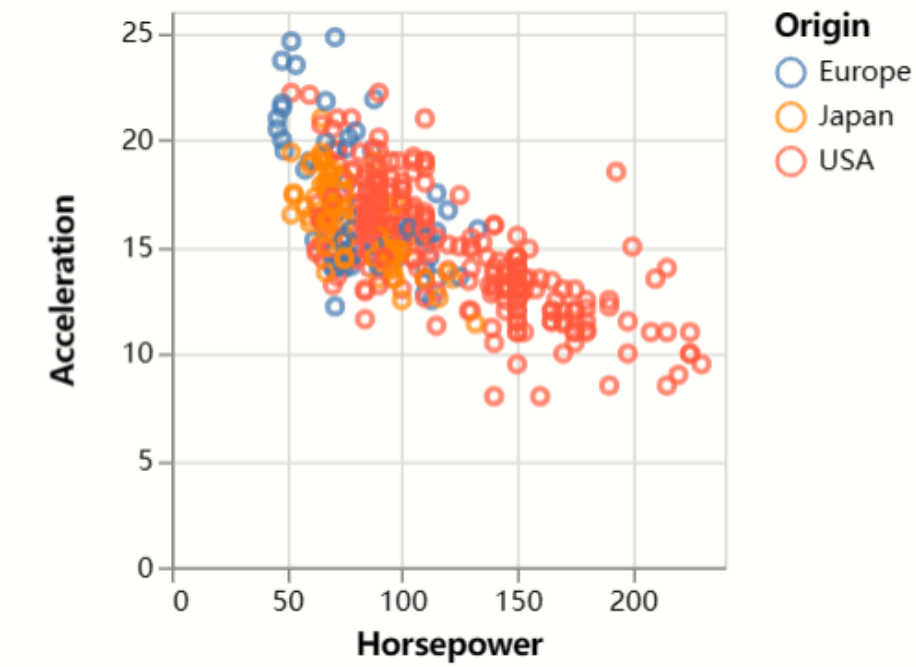




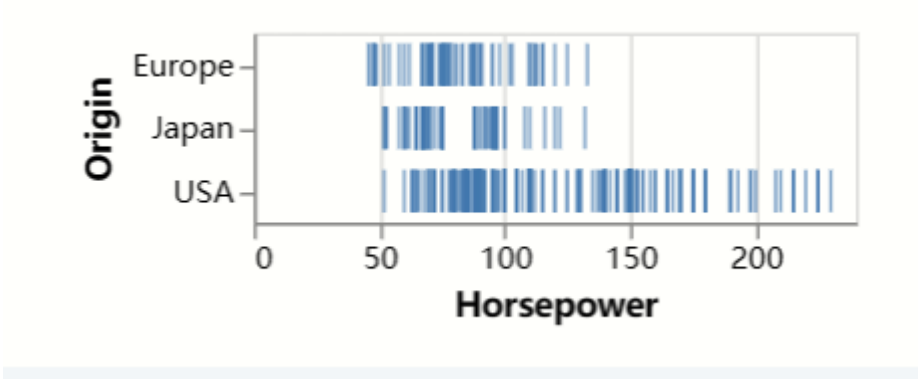
有时候数据很密集的话感觉颜色会更加直观））



通过Horsepower和Acceleration散点的分布状况，可以直观地看出这二者之间大概呈一个负相关关系。



在上图基础上再以color区分出origin，可以发现欧洲生产的车辆相比日本和美国的，大体上马力更小，加速度更快。而美国车更多拥有更高的马力。

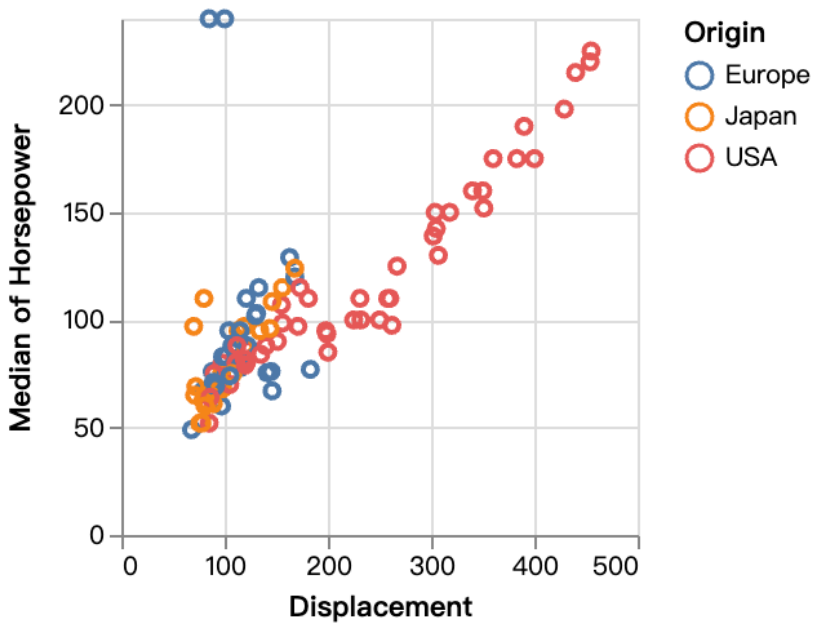


果然如此

btw日本在马力80那一块有很宽的一段断开的，但是美国车在那里一段还蛮密集的。

胡城玮

李衔

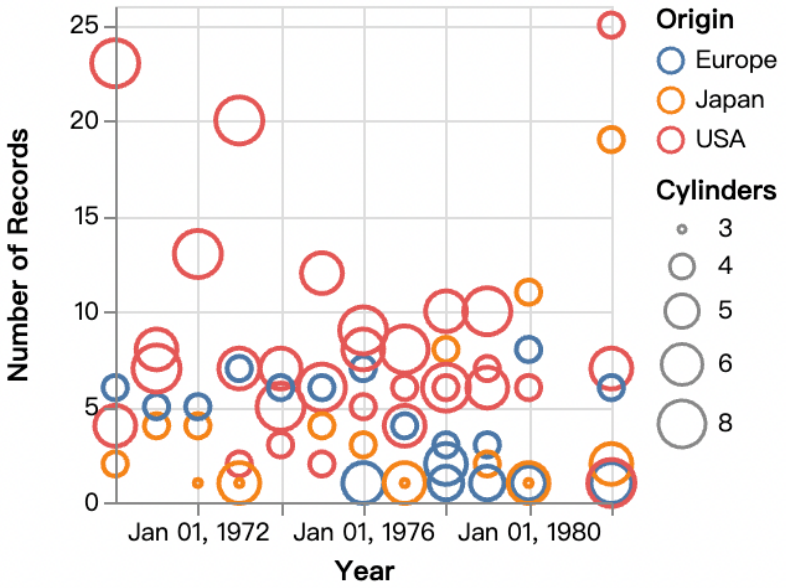


能移动的距离越高，马力也会越大

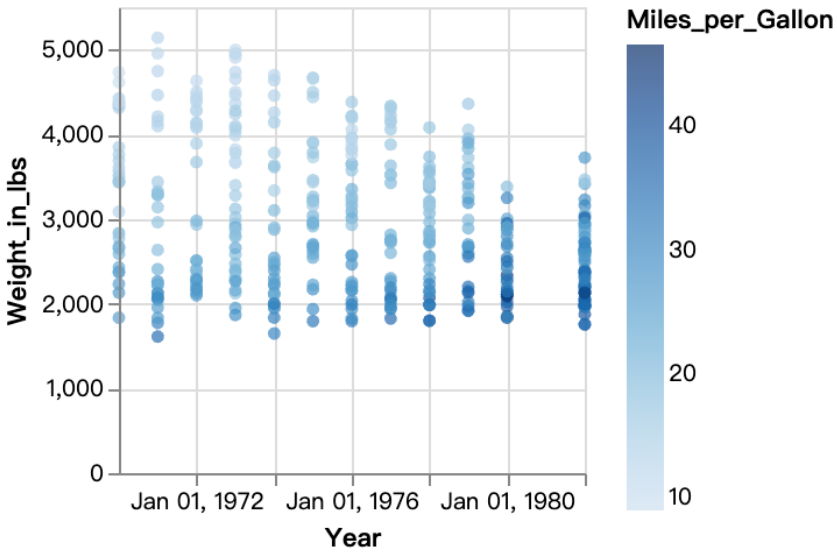
颜色

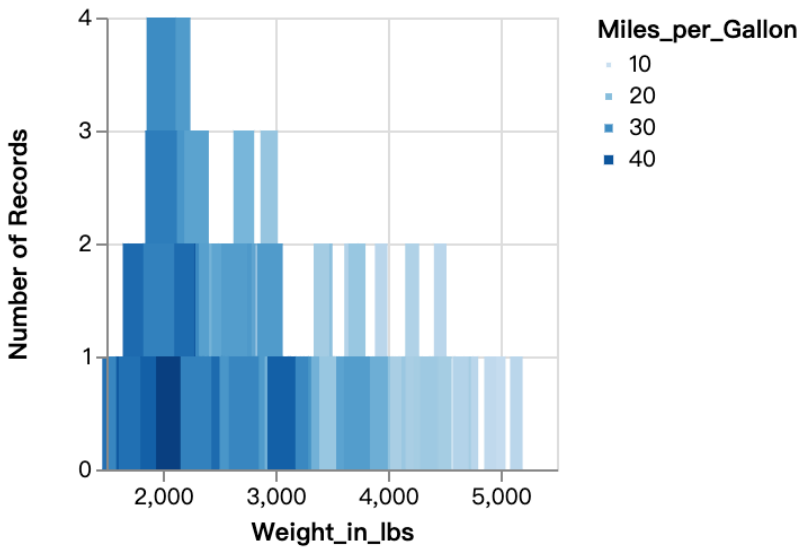
用于标注比较少的数据：用作分类显眼

比较多而且成坐标轴的数据：深浅看起来能够直观感受到差异

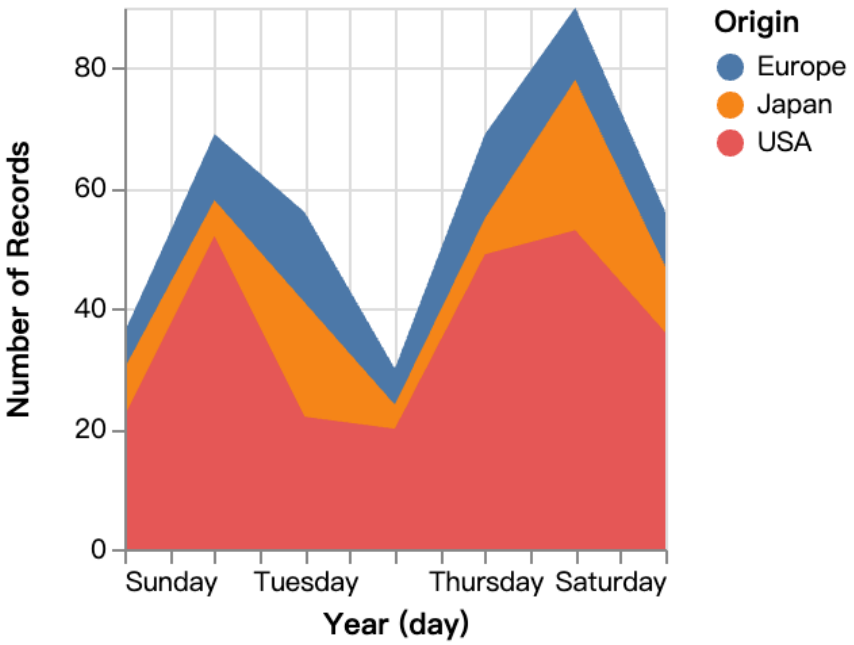


80年代美国产量大马力大，但是一段时间后大排量的车慢慢减少



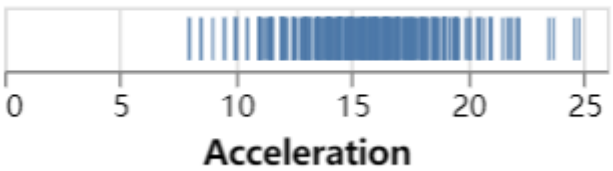


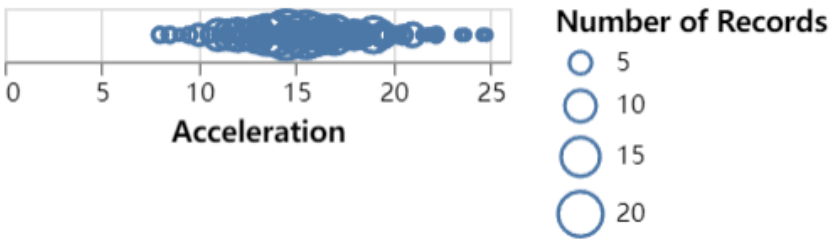
随着年份推移，燃油效率越来越高，同时重量也越来越轻



不知道是什么情况

张奕静





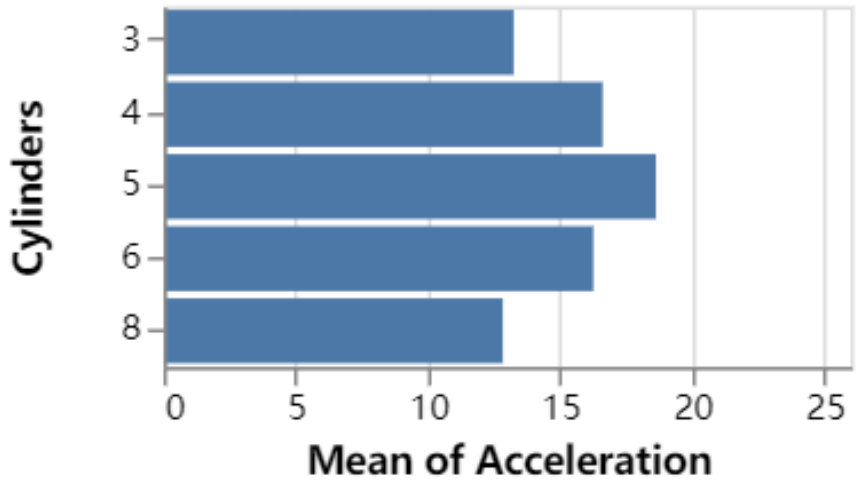
Acceleration:

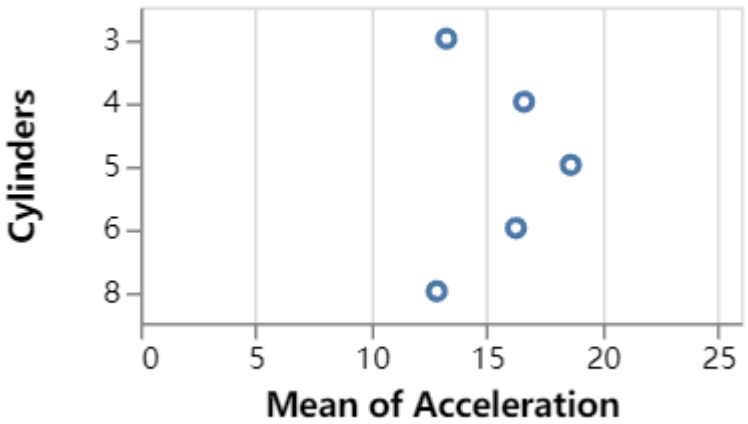
1. 直接通过tick去绘制，通过密集程度去判断加速度

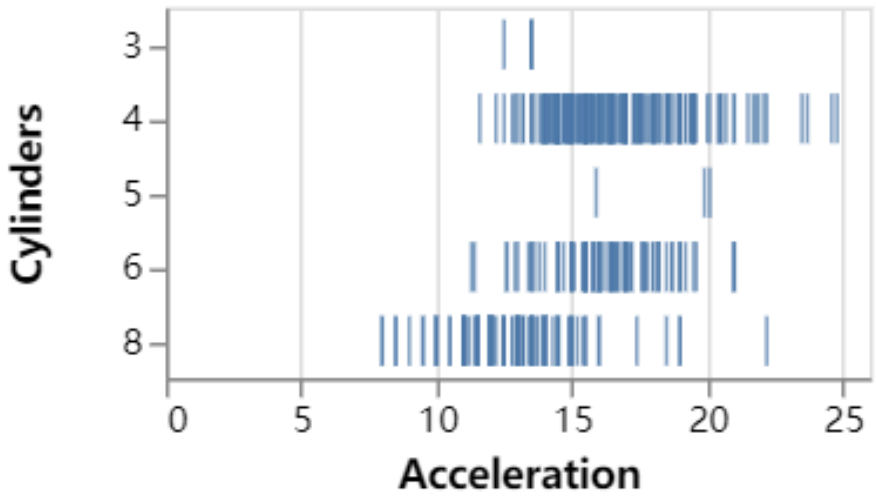
   比较容易判断加速度的分布（集中在哪里）和加速度遍布的范围；比较难以获取数量的数据，更多的是去关注分布

2. 加入Count的数据，但是看不清，更多感觉还是在看分布

3. 通过结合Bin的方式，统计区间内的相应加速度的车辆数，对于数量会有更直观的体验

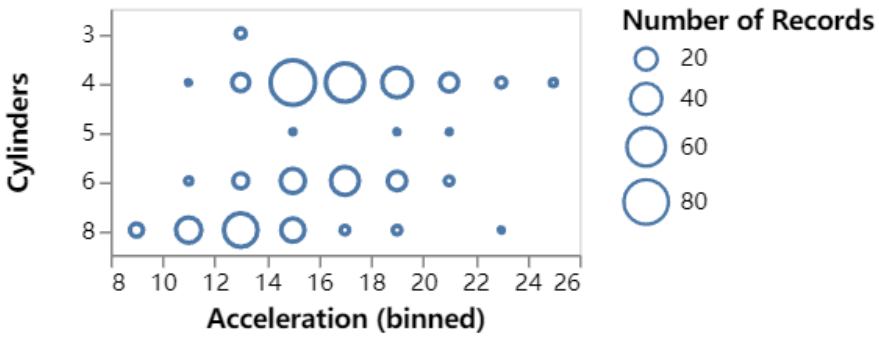


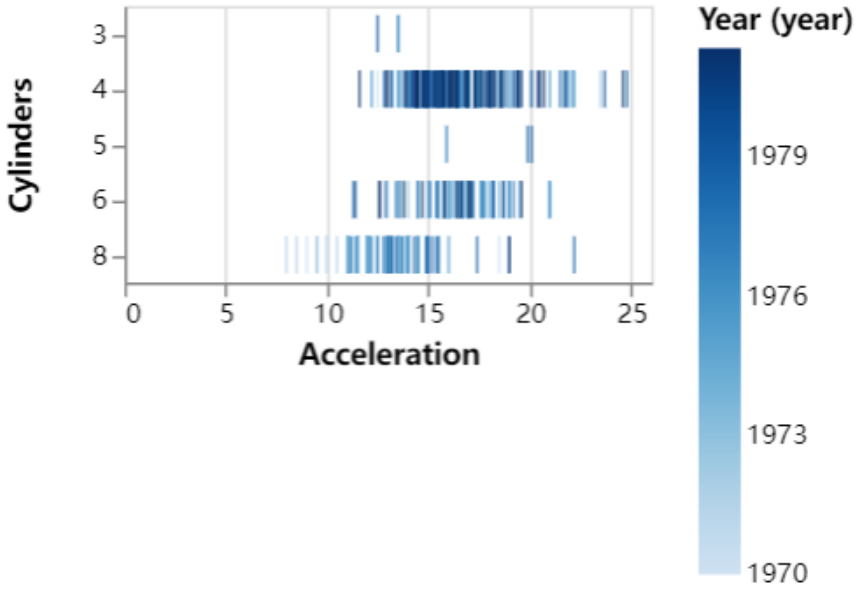


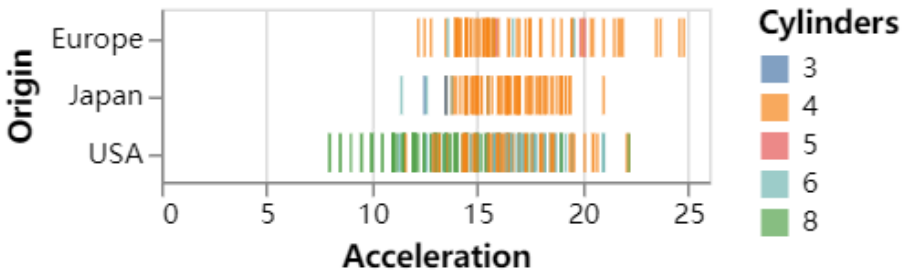


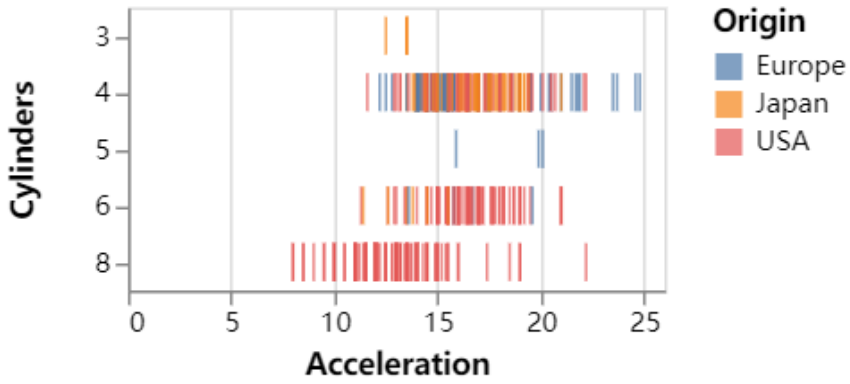
Mean of Acceleration & Cylinders：

1. 可以比较直观的感受到不同类型的车辆之间加速度平均值的区别（长度/面积）
2. 使用点的话感觉会更符合对平均值概念的一个想法（？）



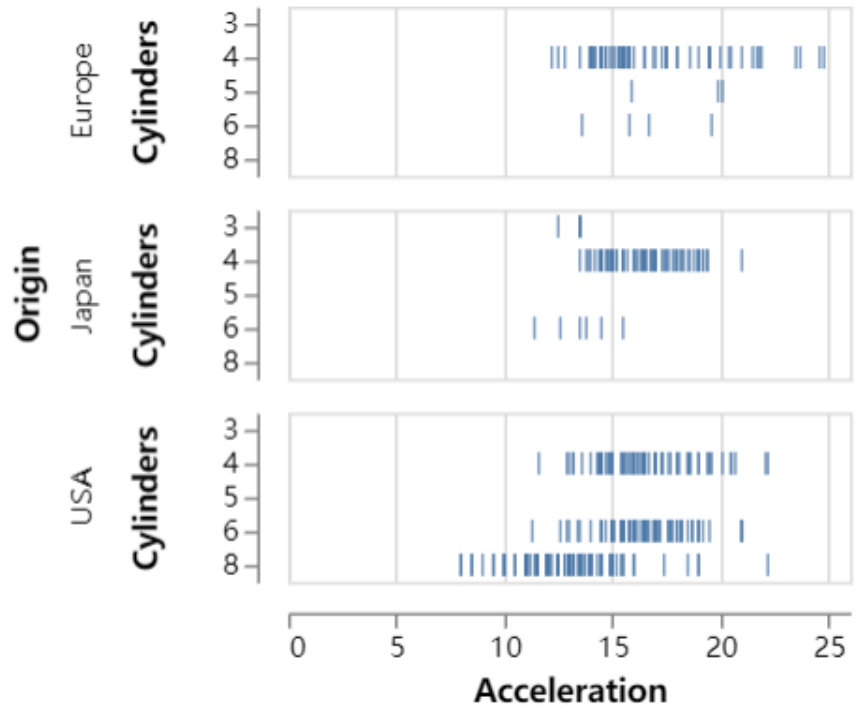






Acceleration & Cylinders：

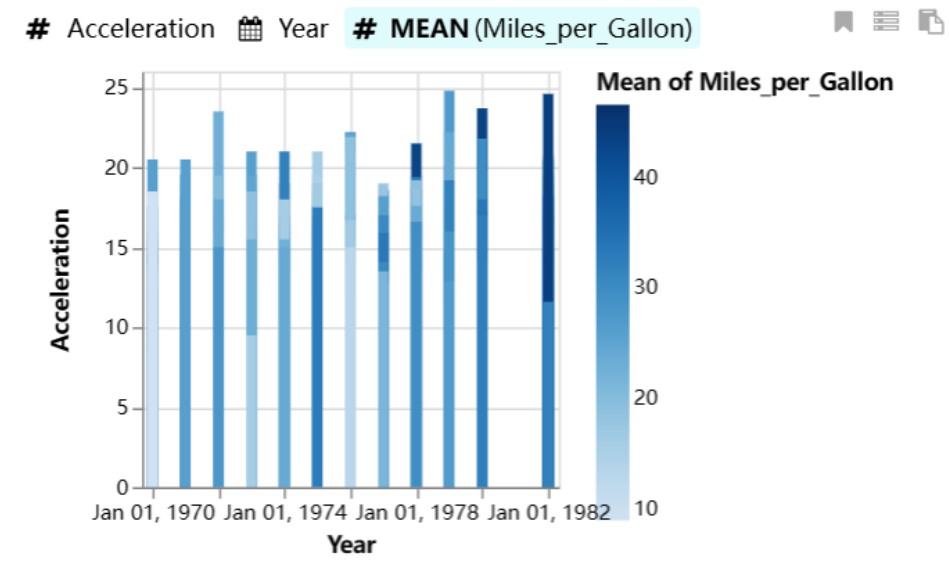
1. 可以较直观的看出每种类型车辆的大致加速度分布，能够相互之间进行比较
2. 通过Bin的方式，让数量更加直观
3. 加入Year的数据，使用颜色的话会受到线条密集程度的影响

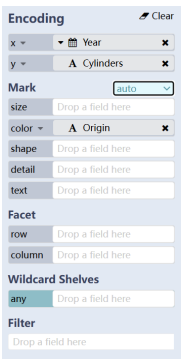


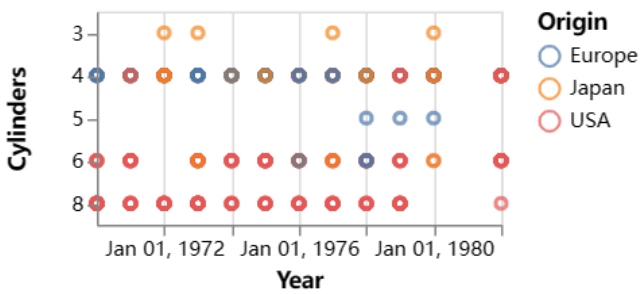
Acceleration & Origin & Cylinders：

1. 将Cylinders通过颜色表示，可以更强烈的感受到各国生产汽车所对应车辆类型；同时能够对于各国加速度范围有一个大致概念
2. 每种类型是哪国生产的，哪种车辆各国都生产
3. 减少了数据密度和颜色的互相干扰，分开看更清楚，但没那么好比较

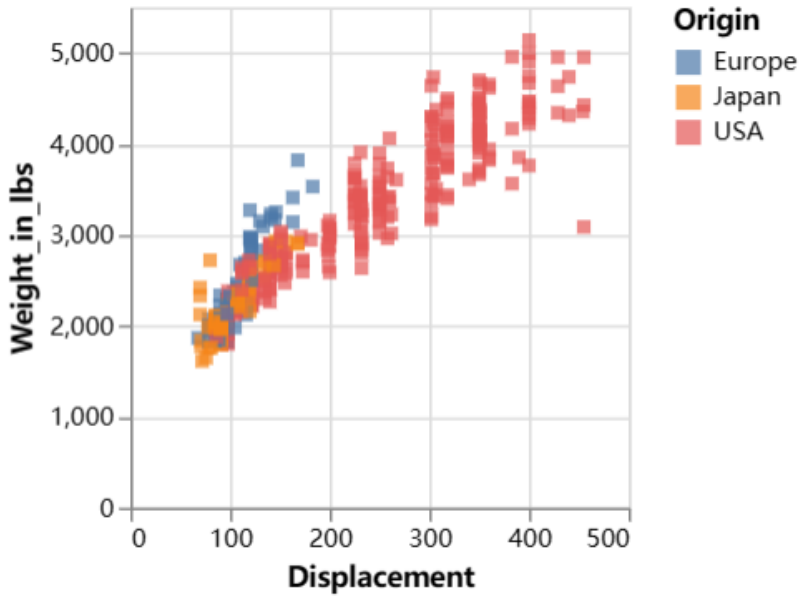
邓詠心

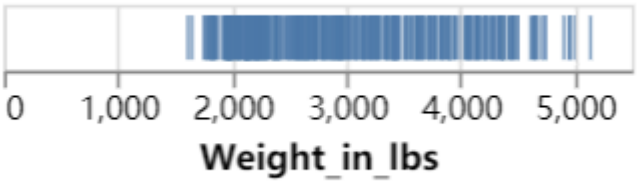




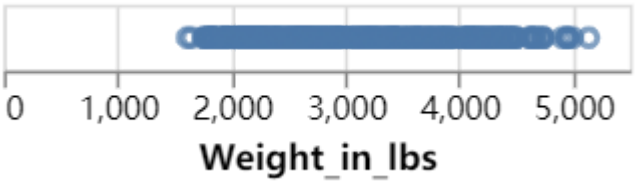


1. BAR: Year （Ordinal） × Acceleration （quantitative），同时以颜色区分Miles_per_Gallon。二维的图像同时显示三项相关数据，美观度较高。但是从信息获取层面来讲，颜色过渡不能让人有效获取Miles_per_Gallon的数据大小（不过可以知道数据大致变化），并且对Year来说，同一时间下变化颜色的Miles_per_Gallon并没有太大意义。

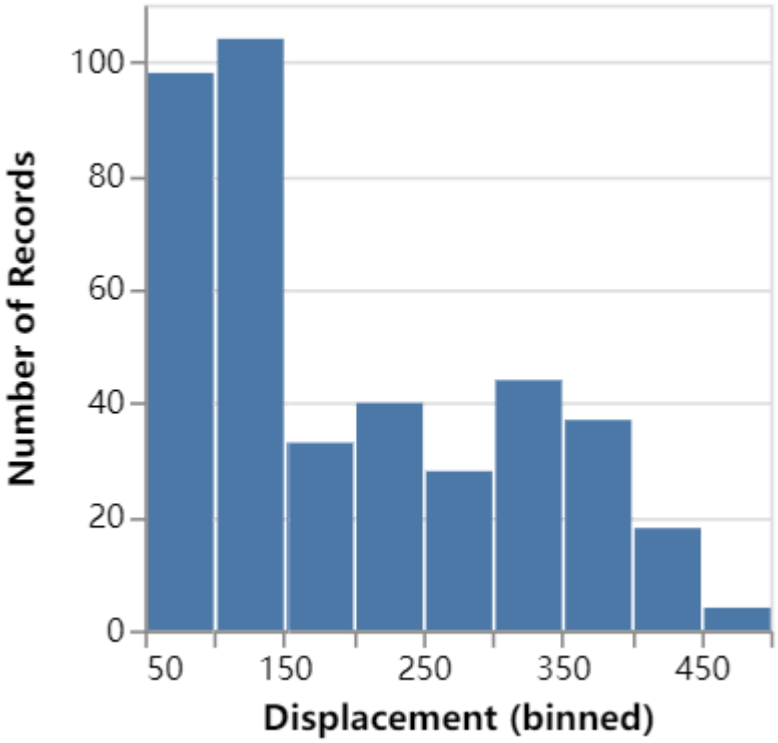
	# Weight_in_lbs A Origin # Displacement  	2. POINT: Year（Ordinal）×Cylinders（Ordinal）,同时以颜色区分Origin（Nominal）。二维的图像同时显示三项相关数据，且图表数据没有堆叠，信息获取较容易。由于Origin类别仅三项，颜色区分清晰并且不存在模糊具体信息的问题。但在该表中出现的颜色不仅仅只有图例里的三个颜色，这一点让人有点困惑（难到生成出错了吗？）。 3. SQUARE: Displacement（quantitative）×Weight_in_lbs（quantitative），同时以颜色区分Origin（Nominal）。这个图中可以比较明显的看出Displacement与Weight_in_lbs的线性相关性。比较有意思的是，通过不同颜色色块的位置分布可以大概感知到不同地区车辆在Displacement与Weight_in_lbs两个方面的差异。
薛仁超		1. 1D quantitative; Good



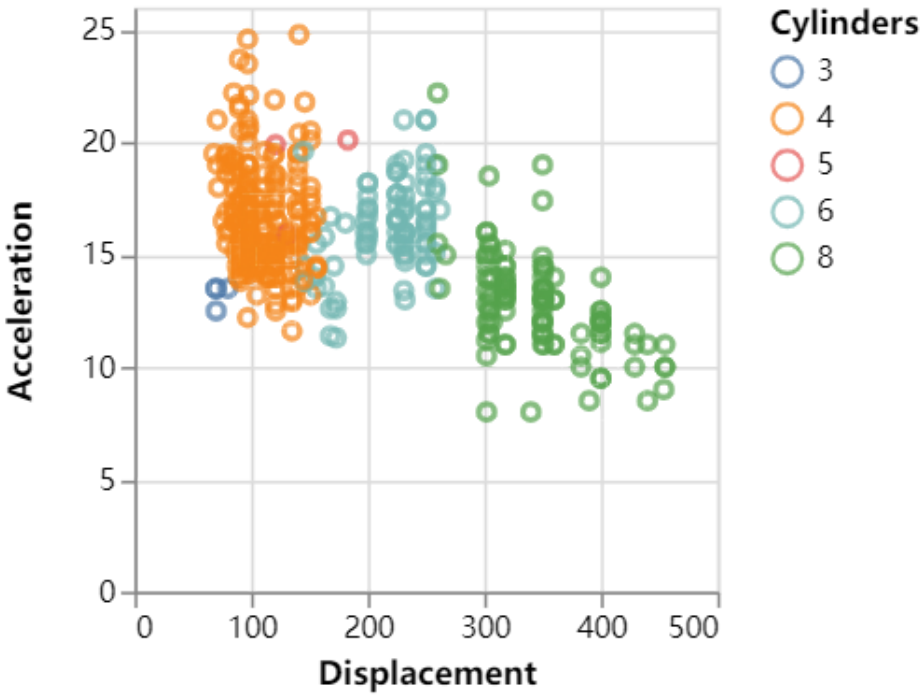
2. 1D quantitative; Not so good



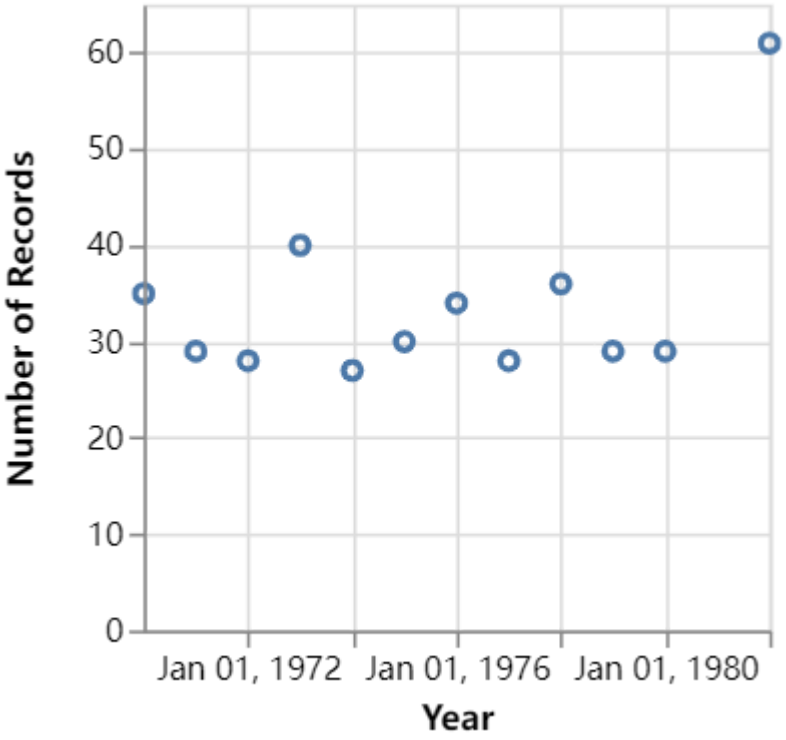
3. 1D quantitative, binned; Good



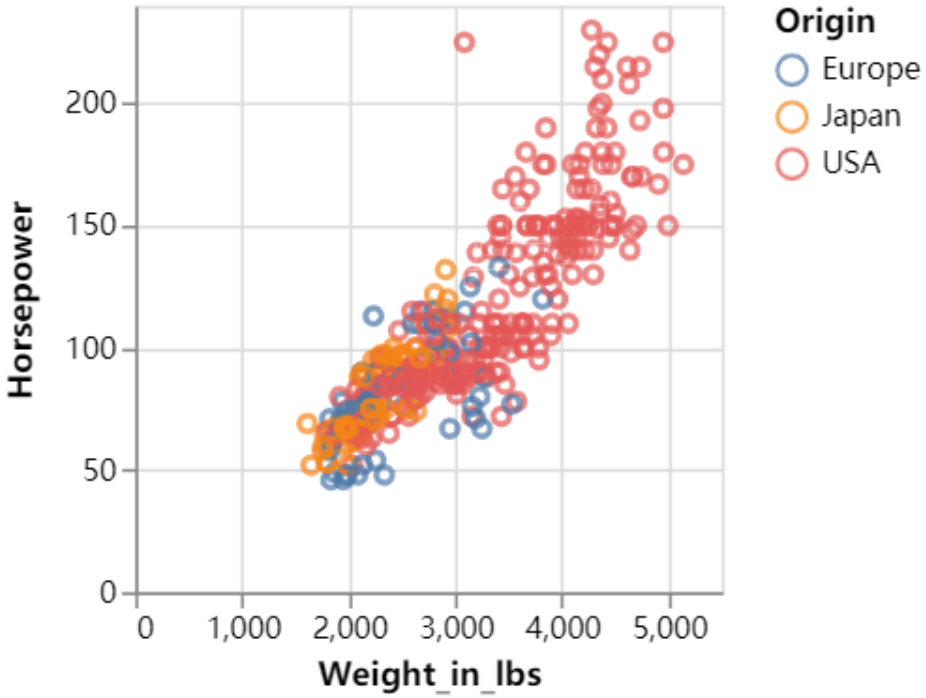
4. 2D quantitative x quantitative, colored; Good



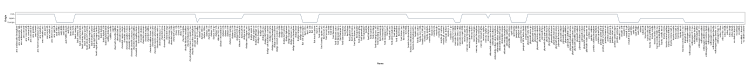
5. 2D temporal x quant; Average



6. 2D quant x quant, colored, clear distribution, useful information; Perfect



7. (too small need to enhance) 2D nom x nom, can't get the ones with same origin together, but still can be seen by length of line. Average



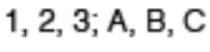
8.

张程浩		

2. Encoding Effectiveness

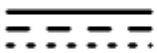
2.1 Properties of Visual Encodings

- Different visual properties vary or may be modified in different ways, which makes them good for encoding different types of data.
- **Two key factors:**
  a. Whether a visual property is naturally ordered
  b. How many distinct values

	Encoding	**Is it Naturally Ordered?** (insert 0/1 below)	**how many distinct values?** (few, medium, many, infinite)
	position		
	text		
	length		
	size,area		
	angle		
	density		
	color saturation		
	color hue		
	shape		

	enclosure/connection		

2.2 Choosing Appropriate Visual Encodings

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (<20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

2.3 Choose

FT4schools_RGS.pdf

3. Useful website

- http://tuzhidian.com/
- https://datavizproject.com
- HSV Color Model:

https://math.hws.edu/graphicsbook/demos/c2/rgb-hsv.html

RGB/HSV Color Demo

RGB and HSV Color Model Demo