# Jan 20 IIAM

## Options for highly dispersed data

1. Logs
2. Add labels for small values (if there are a few)
3. rangewise subplotting
   large country subplot | small country subplot

1,000,000
1,000
1

```
alt.Y("value:Q",
      axis=alt.Axis(format=",.0f")
)
```

Fake data (Horror)

- 20%
- 80%
- 60%
  50%
  30%
- 10%
  50%
- 60%
  30%
- 20%

$\frac{41}{10}$ =

average 41%

q1 (25th percentile) 20%

q3 (75th percentile) 60%

## DPS on Aggregation

| | method | lesson |
|---|---|---|
| X: | month(Release_Date)<br>Release_Date | |
| Y: | average (US_Gross) | summer & Dec high |
| | max (US_Gross) | Dec high |
| | iqr (" ) | high in Jul & Jun |
| | median (" ) | similar to avg... |
| | sum (" ) | Dec 25 max |