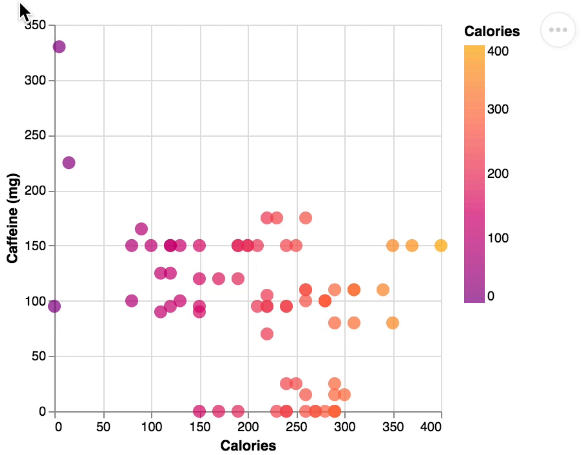
ResBaz Data Visualization Workshop

Agenda:

- Data Vis basics & terminology
- Web Charting with Vega-Lite
- Free experimentation time



Template/Data for Code-Along:

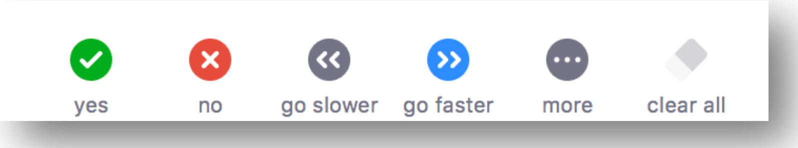
- https://bit.ly/ResBazVisWorkshop

Pre-Survey (Google Form):

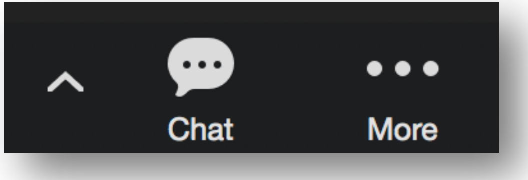
- https://tinyurl.com/VisWorkshopPreSurvey

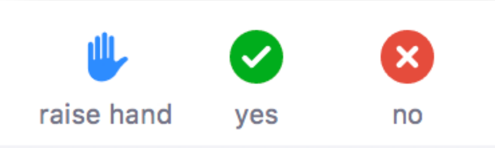
How we’re using Zoom

- Declare you’re finished with activities with ”yes” notice in participant list.



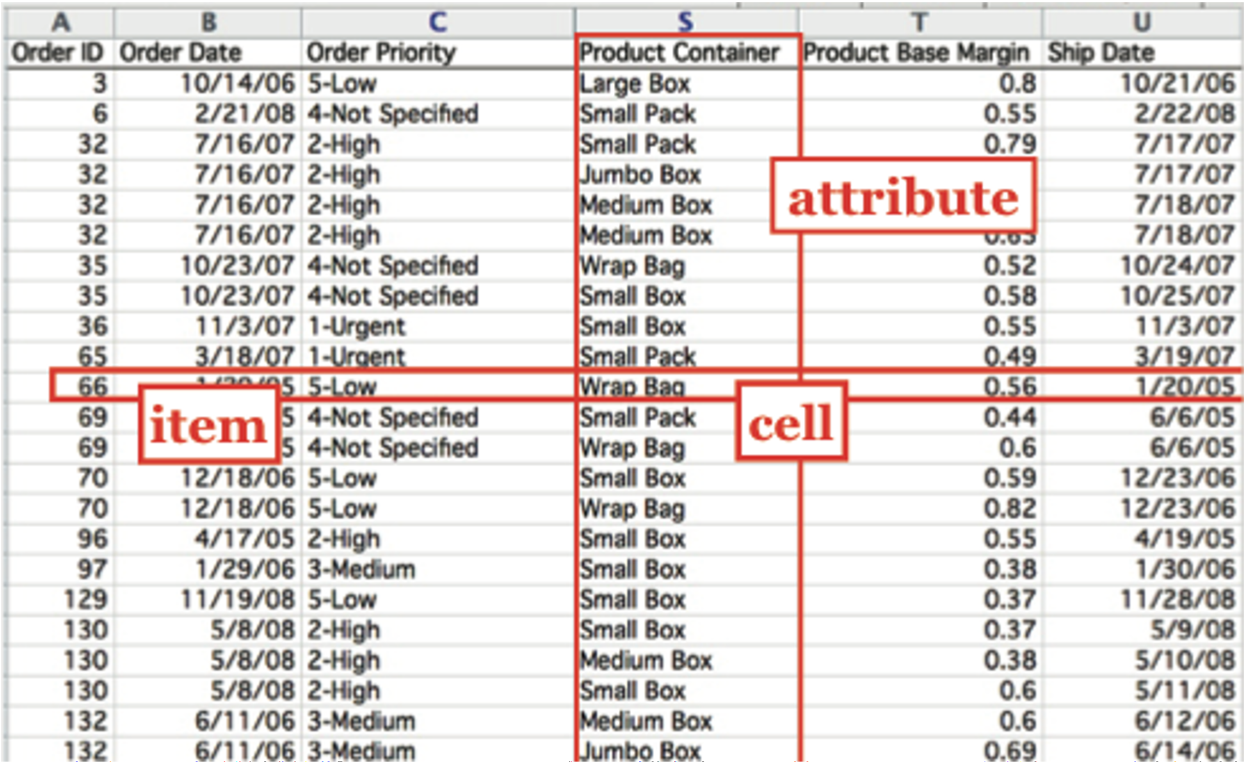
- When you have a question or answer, either:
  - Write it in the chat
  - Use the “raise hand” feature
  - Direct chat helper Alex Bigelow
  - Add to HackMD file





Some Data Terminology

Data Tables



A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	[illegible]	7/17/07
32	7/16/07	2-High	Medium Box	[illegible]	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	[illegible]	4-Not Specified	Small Pack	0.44	6/6/05
69	[illegible]	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06

Each data point is an **item** (or *records*), usually represented as a row.

Columns contain values of a particular **attribute** (or *field*).

The value of an attribute for a particular item is a **cell** (or *attribute value*).

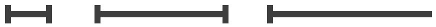
Types of Attributes

Quantitative data has order and allows mathematical operations

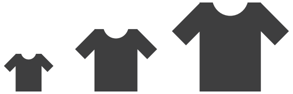
Ordinal data has order but not mathematical relationships

Nominal (a.k.a. Categorical) data has neither order nor mathematical relationships

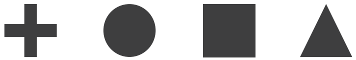
→ *Quantitative*



→ *Ordinal*



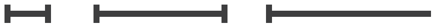
→ *Nominal*



Images from Munzner, Visualization Analysis and Design

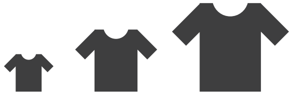
Examples

➔ *Quantitative*



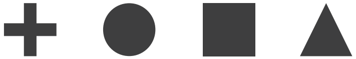
- *Lengths*
- *Counts*
- *Pressure*
- *Temperature*
- *Weights*
- *Distances*
- *Dates*
- *Coordinates*

➔ *Ordinal*



- *S, M, L sizes*
- *Letter grades*
- *Rankings*
- *Likert scales (e.g., rate from very satisfied to very dissatisfied)*

➔ *Nominal*



- *Shapes*
- *Colors*
- *Names*
- *Blood types*
- *Countries*
- *Event types*

Images from Munzner, Visualization Analysis and Design

What operations can you do?

➔ *Quantitative*

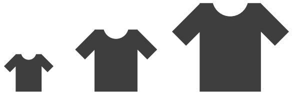
Compare:

=, ≠, <, >, +, −

Ratio Only:

×, ÷, *ratios, proportions*

➔ *Ordinal*



Compare:

=, ≠, <, >

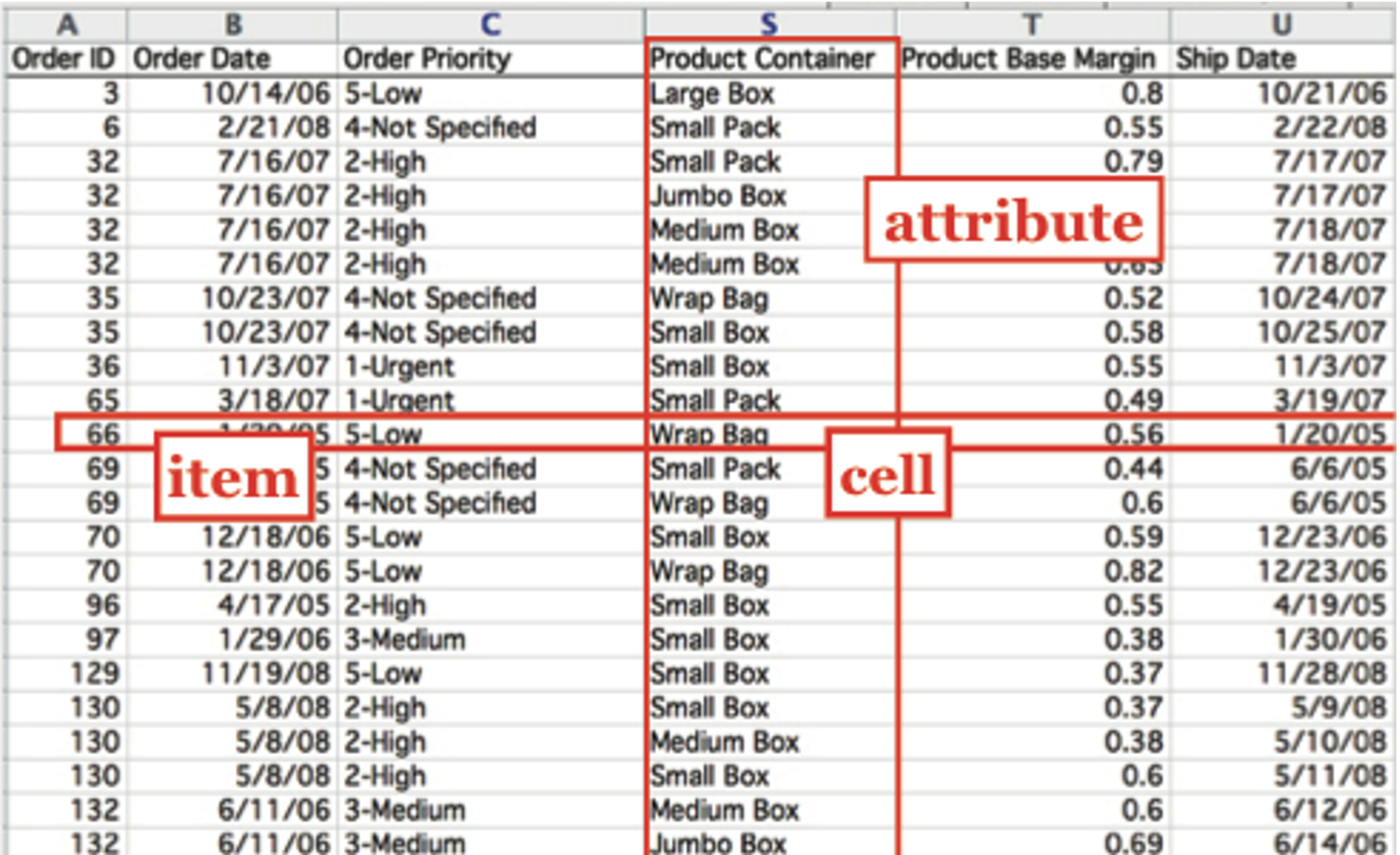
➔ *Nominal*

Compare:

=, ≠

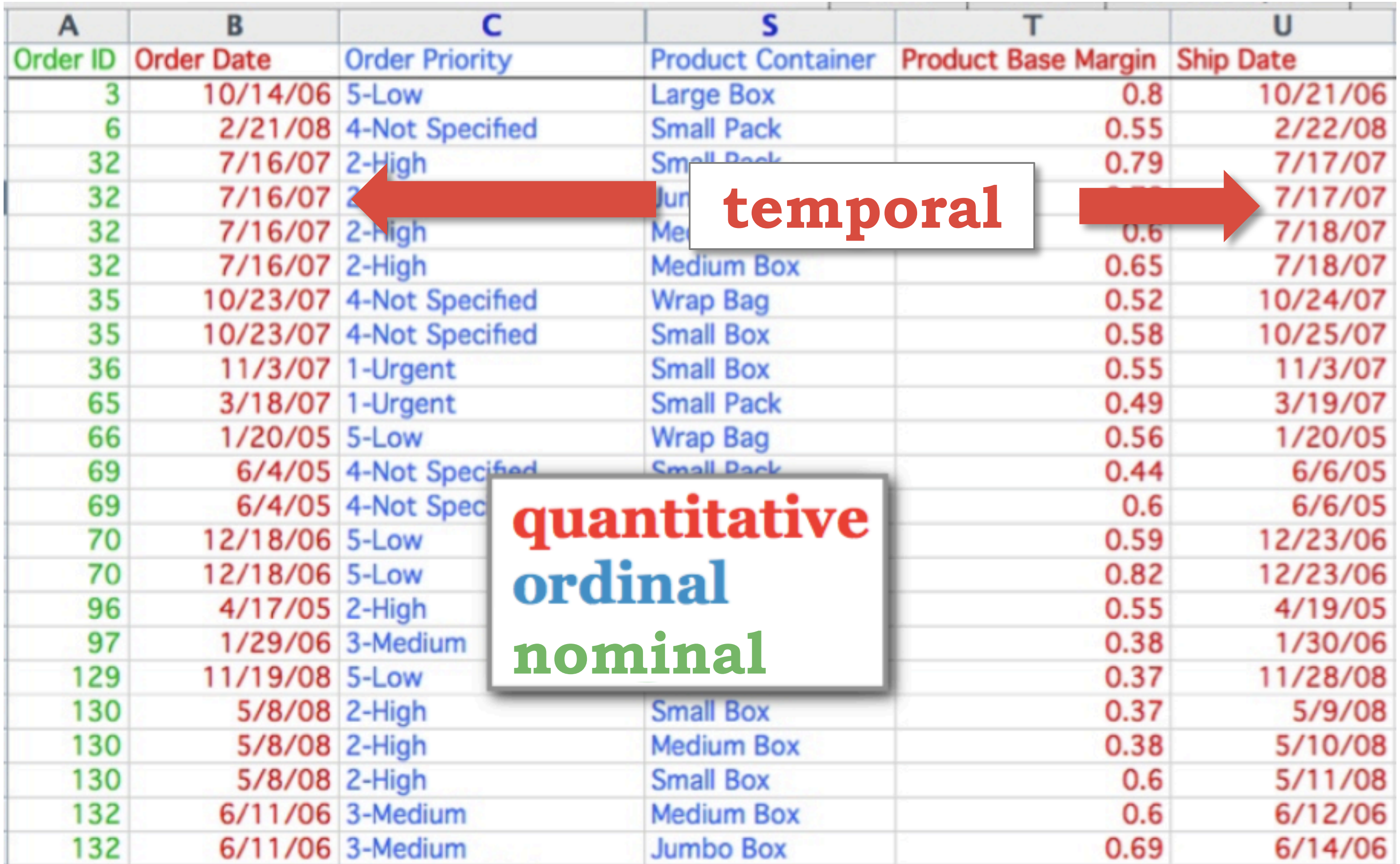
Images from Munzner, Visualization Analysis and Design

Quantitative, Ordinal, or Nominal?



A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	[illegible]	7/17/07
32	7/16/07	2-High	Medium Box	[illegible]	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	[illegible]	4-Not Specified	Small Pack	0.44	6/6/05
69	[illegible]	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06

Quantitative, Ordinal, or Nominal?



A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2	Jun		7/17/07
32	7/16/07	2-High	Me	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Spec		0.6	6/6/05
70	12/18/06	5-Low		0.59	12/23/06
70	12/18/06	5-Low		0.82	12/23/06
96	4/17/05	2-High		0.55	4/19/05
97	1/29/06	3-Medium		0.38	1/30/06
129	11/19/08	5-Low		0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06

temporal

quantitative

ordinal

nominal

Encoding: Mapping Data to Visualization

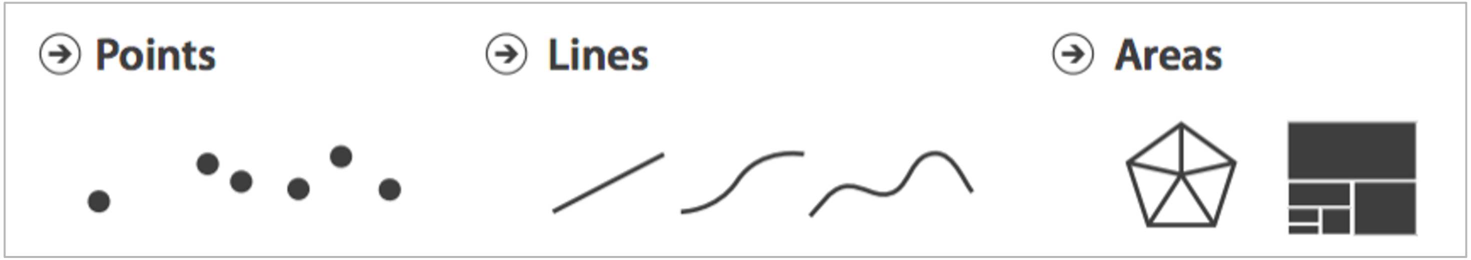
Marks, Channels, & Encoding

Encoding: Map data to visual structure

Marks: Graphical primitives that encode items / entities

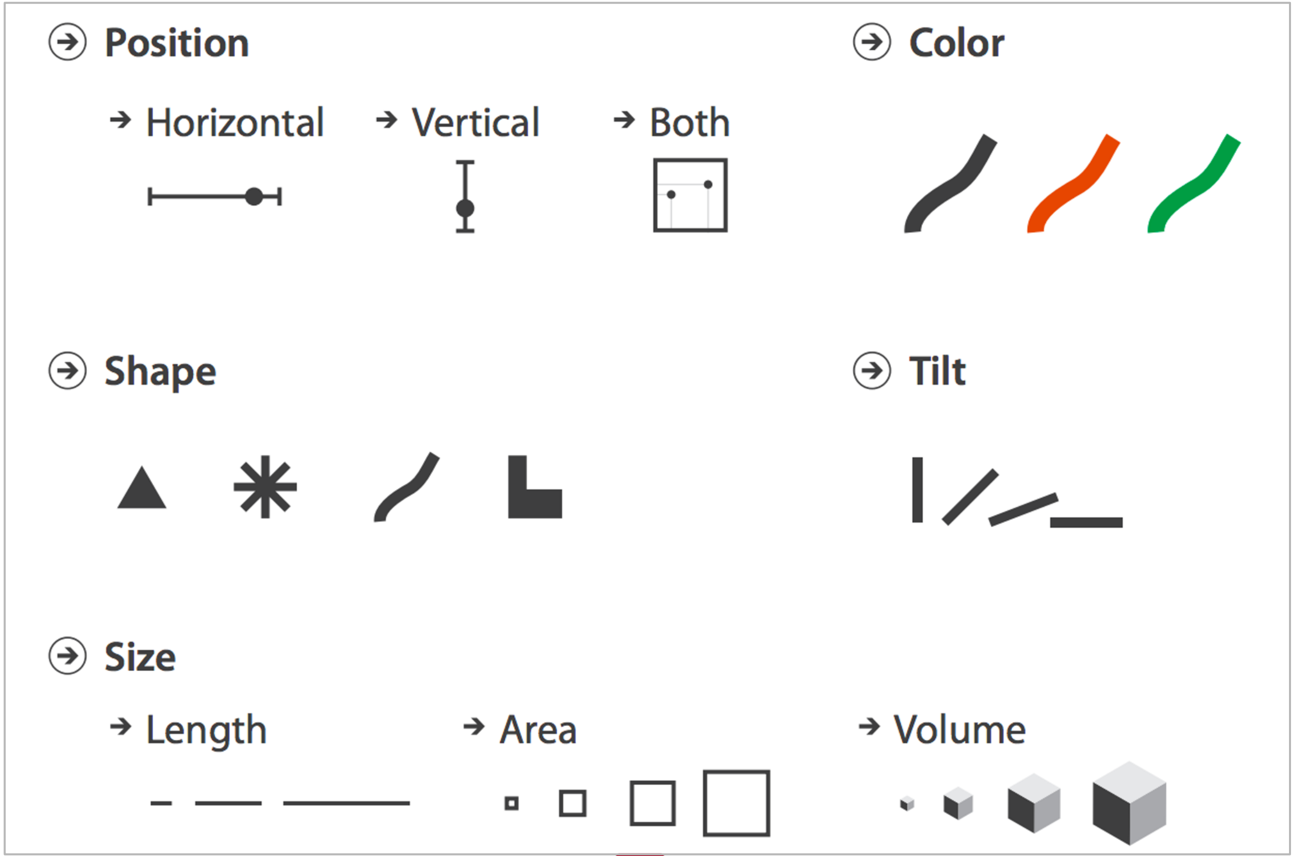
Channels: Properties of mark appearance, often used to encode attributes or other information

Marks: Graphical primitives that encode items or entities

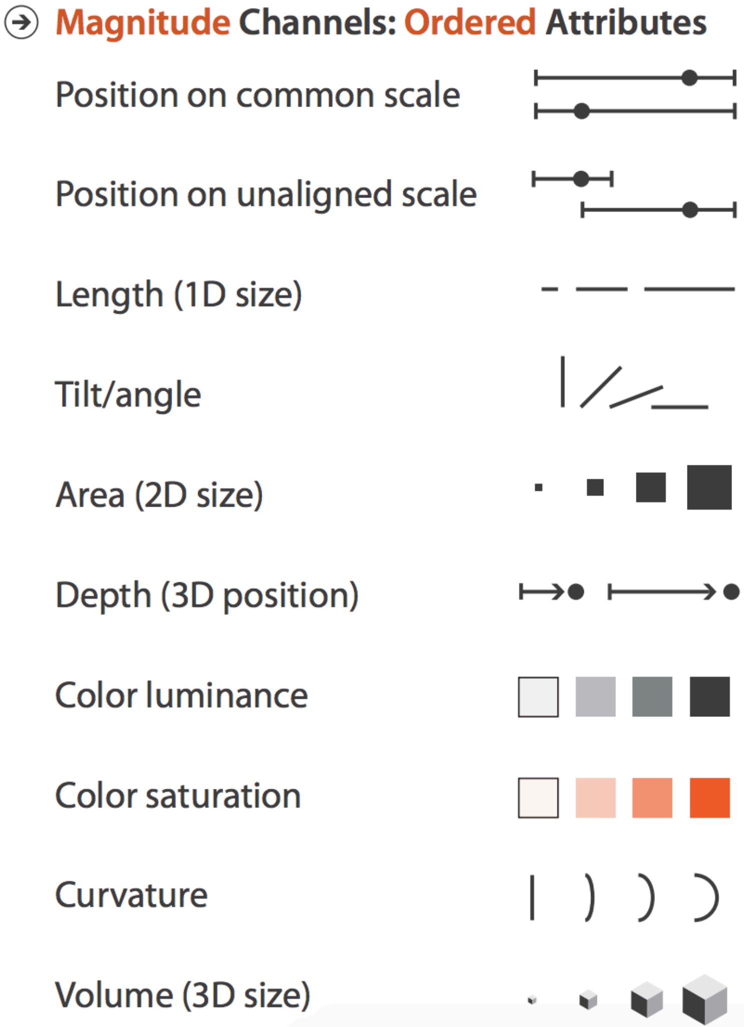


Images from Munzner, Visualization Analysis and Design

Channels: Properties of mark appearance, often used to encode attributes or other information



Images from Munzner, Visualization Analysis and Design



➔ Magnitude Channels: Ordered Attributes

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

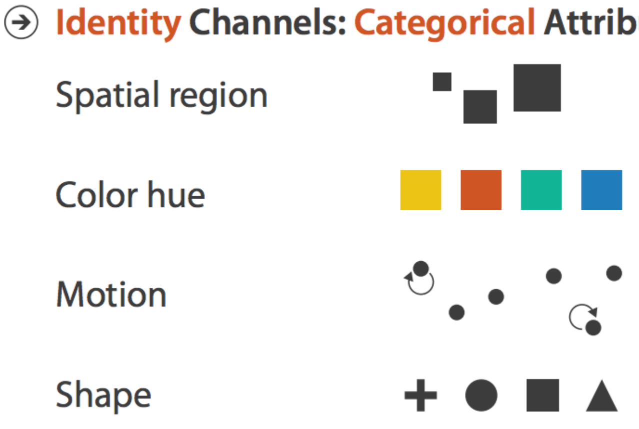
Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)



➔ Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

Shape

We can Construct a Mapping of Data Values to Perceptual Channels

Images from Munzner, Visual Analysis and Design

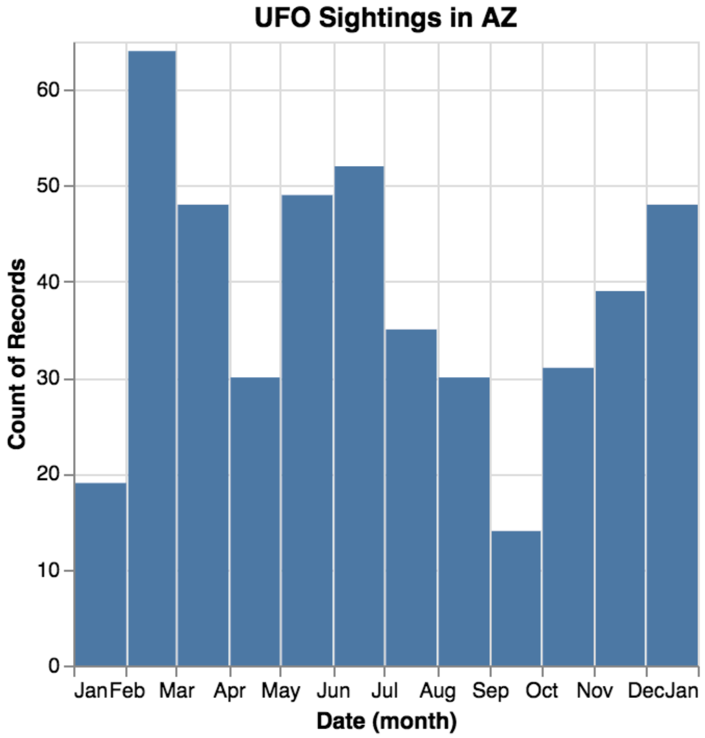
Encodings of Common Charts

Bar Chart: Show relative counts

Marks: rectangles

Encoding: quantitative value is mapped to height of rectangle on a common scale

Nominal value is mapped to x-position

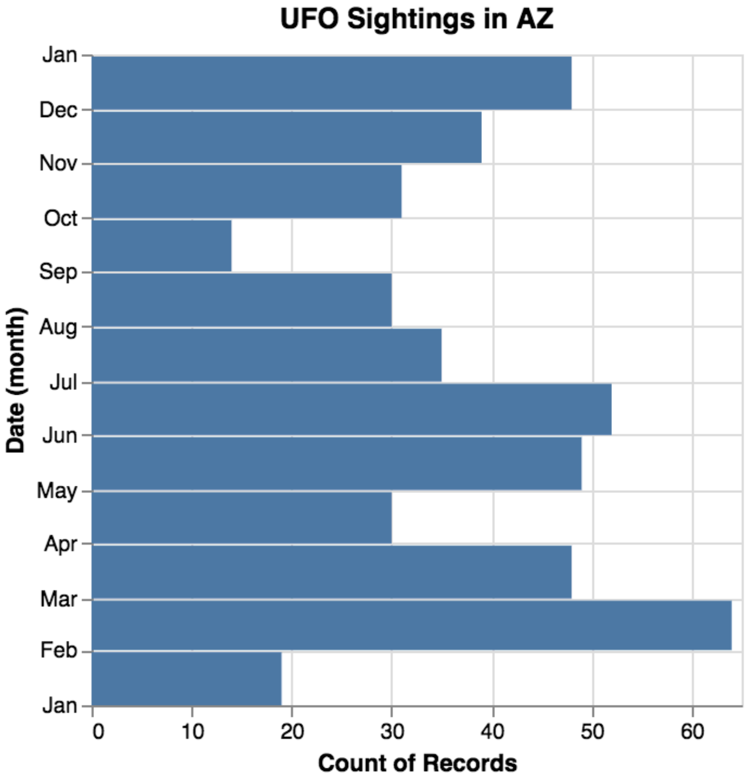


Consider rotating for text readability

Marks: rectangles

Encoding: quantitative value is mapped to width of rectangle on a common scale

Nominal value is mapped to y-position

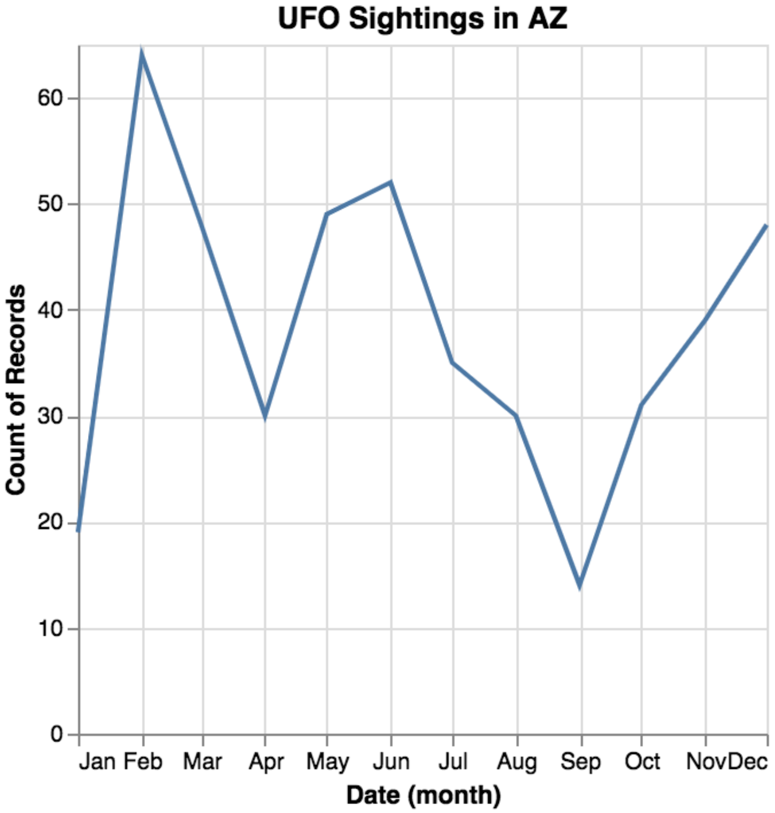


Line Charts: Show trends

Marks: lines

Encoding: quantitative value is mapped to y-position of line endpoint.

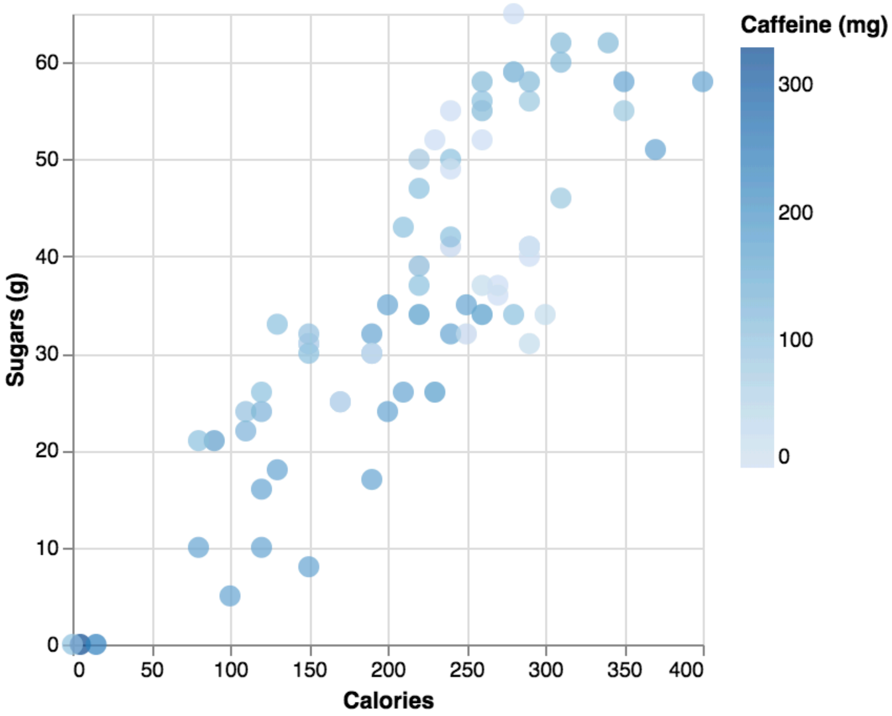
Temporal value is mapped to x-position



Scatter Plots: show correlation

Marks: points

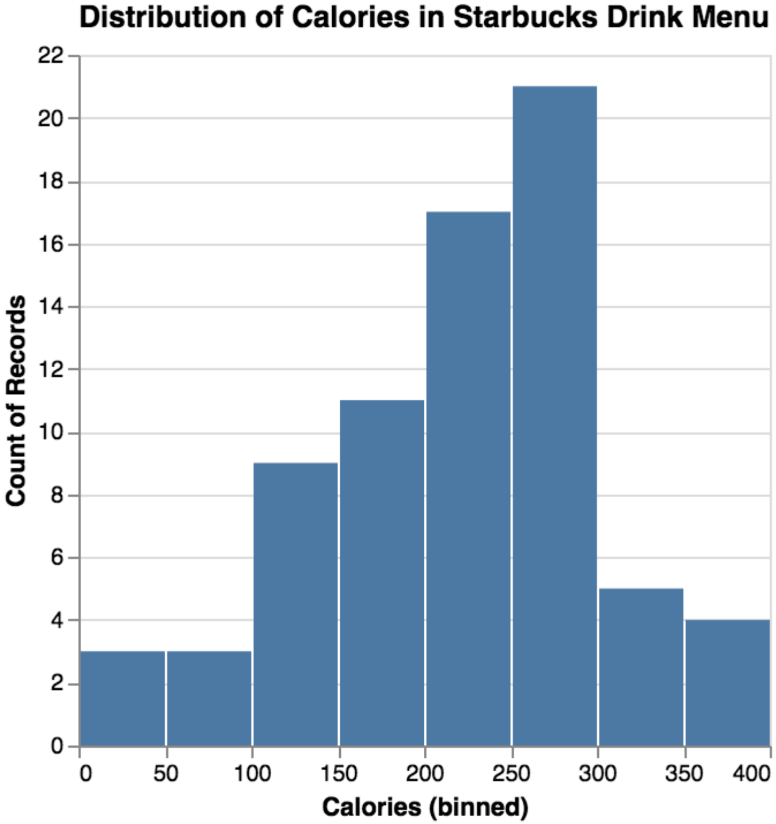
Encoding: two quantitative value is mapped to x and y position respectively



Histograms: show distribution

Marks: bars

Encoding: x position denotes range of calories, y position denotes number of drinks in that calorie range



Vega-Lite

Why Vega-Lite?

At Hackathons, I noticed most projects with visualization used basic charts and some projects had streaming data.

Vega-Lite is a lightweight, robust library when it comes to quickly creating basic charts from data.

Vega-Lite has support for streaming data (not covered in this workshop)

Let’s go through this together!

If you have not already, download the workshop files:
https://bit.ly/ResBazVisWorkshop

Unzip the file and open “template.html” in a web browser

Veg—Lite can be embedded in a webpage

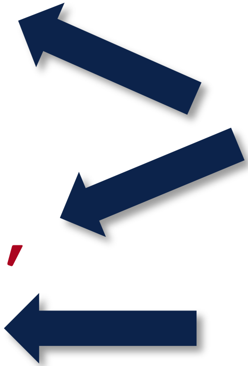
```
<!DOCTYPE html>
<html>
  <head>…</head>
  <body>
    <div id=“vis”></div>
    <script>
      var spec = { …JSON specification here… };
      vegaEmbed(’#vis’, spec);
    </script>
  </body>
</html>
```

General JSON Syntax: Lists

JSON has two structures, an unordered **object {}** of key-value pairs and an ordered **list []** of items, both are comma separated

List Example

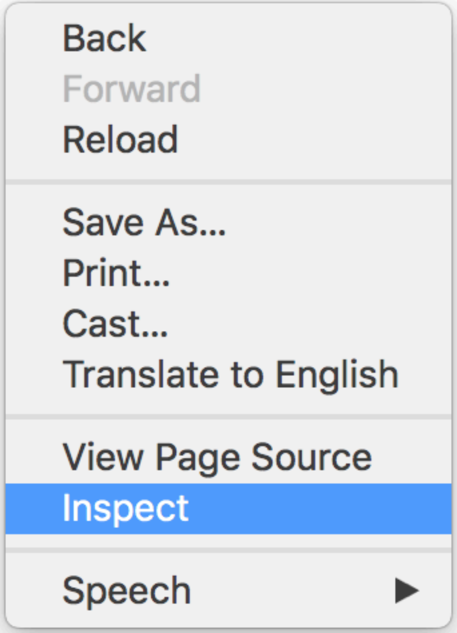
```
[
   "zero",
   "one",
   "two",
   "three",
   "four"
]
```

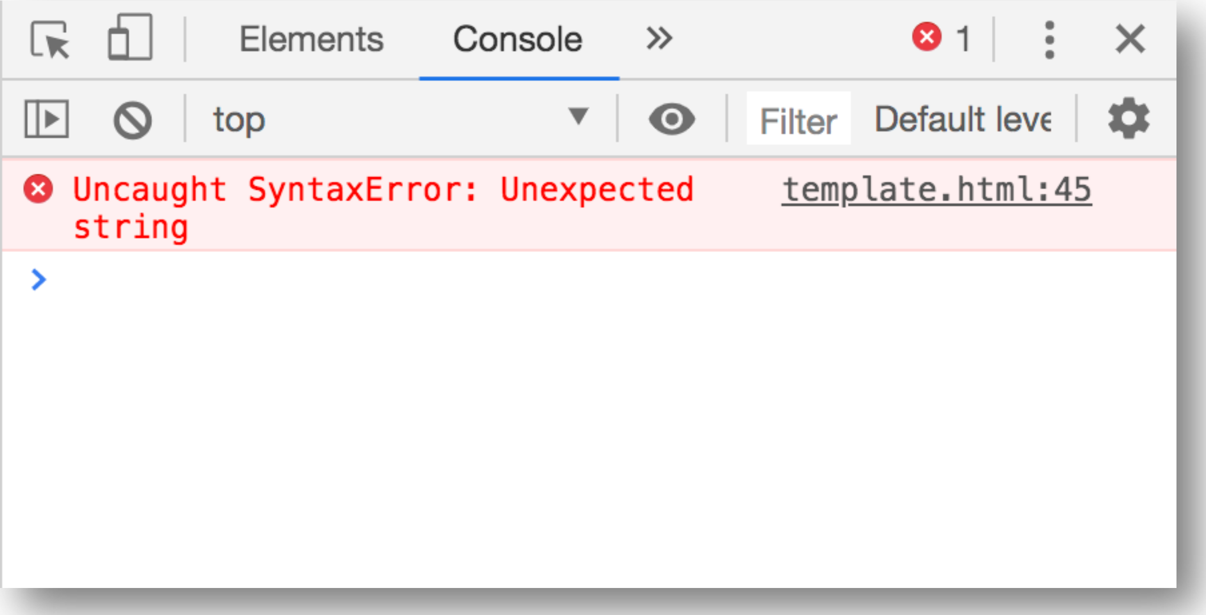


commas

No comma

Missing commas often lead to strange error messages



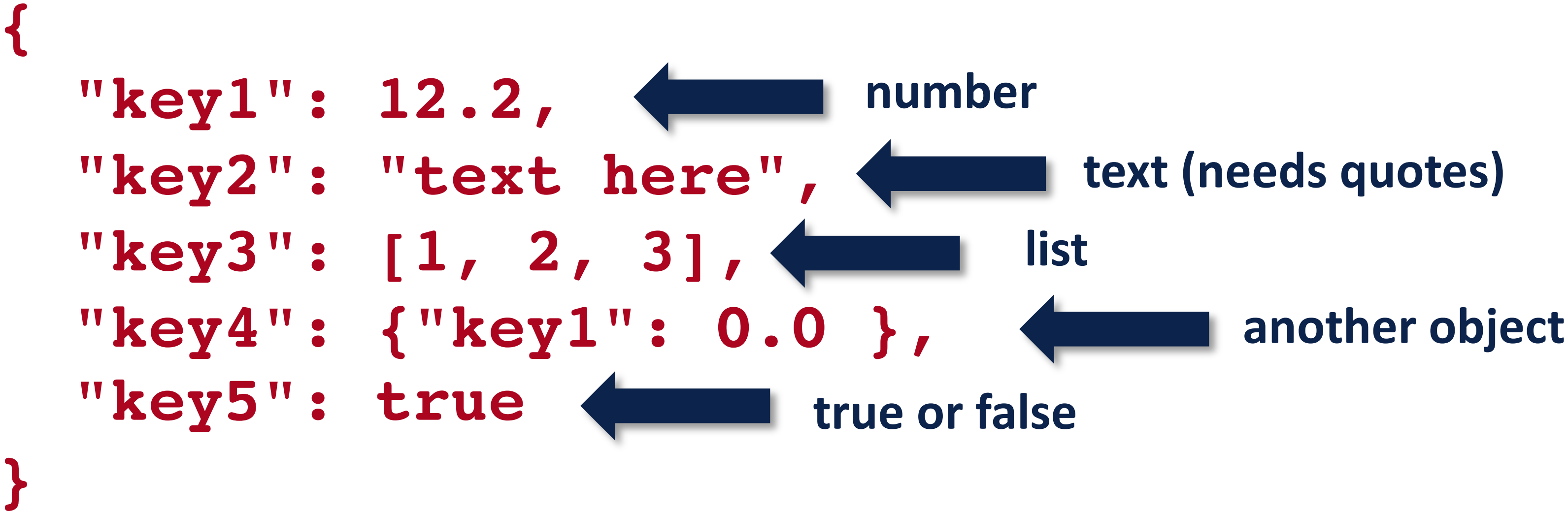


```
Uncaught SyntaxError: Unexpected      template.html:45
string
>
```

General JSON Syntax: Objects

JSON has two structures, an unordered **object {}** of key-value pairs and an ordered **list []** of items, both are comma separated

Object Example



```
{
  "key1": 12.2,
  "key2": "text here",
  "key3": [1, 2, 3],
  "key4": {"key1": 0.0 },
  "key5": true
}
```

number

text (needs quotes)

list

another object

true or false

General JSON Syntax

JSON has two structures, an unordered **object {}** of key-value pairs and an ordered **list []** of items, both are comma separated

Object Example

```
{
  "key1": 12.2,
  "key2": "text here",
  "key3": [1, 2, 3],
  "key4": { "key1": 0.0 },
  "key5": true
}
```

List of Objects Example

```
[
  { "id": 0,
    "name": "foo"
  },
  { "id": 1,
    "name": "bar"
  }
]
```

comma

Anatomy of a Vega-Lite specification

data

mark

encodings

```
{
    "data": { "url": "dir/data.csv" },
    "mark": "point",
    "encoding": {
          "x": {
              "field": "column_name",
              "type": "quantitative"
          }
    }
}
```

Data can be a URL/file, variable name, or inline

```
"data": { "url": "data/mydata.json" }

"data": { "values": variable_name }

"data": {
  "values": [
     { "id": 0, "foo": 7, "bar": "peas" },
     { "id": 1, "foo": 3, "bar": "carrots" },
     { "id": 2, "foo": 6, "bar": "carrots" },
     { "id": 3, "foo": 5.5, "bar": "peas" }
  ]
}
```

Several marks available

```
{
    “mark”: "point",
}
```

area
bar
circle
line
point
rule
square
tick
rect
text
geoshape
boxplot
errorbar
errorband

Tooltips

From encodings:

```
{
 "mark": { "type": "point", "tooltip": true }
}
```

From data:

```
{
 "mark": { "type": "point",
           "tooltip": { "content": "data" }
  }
}
```

Small Example

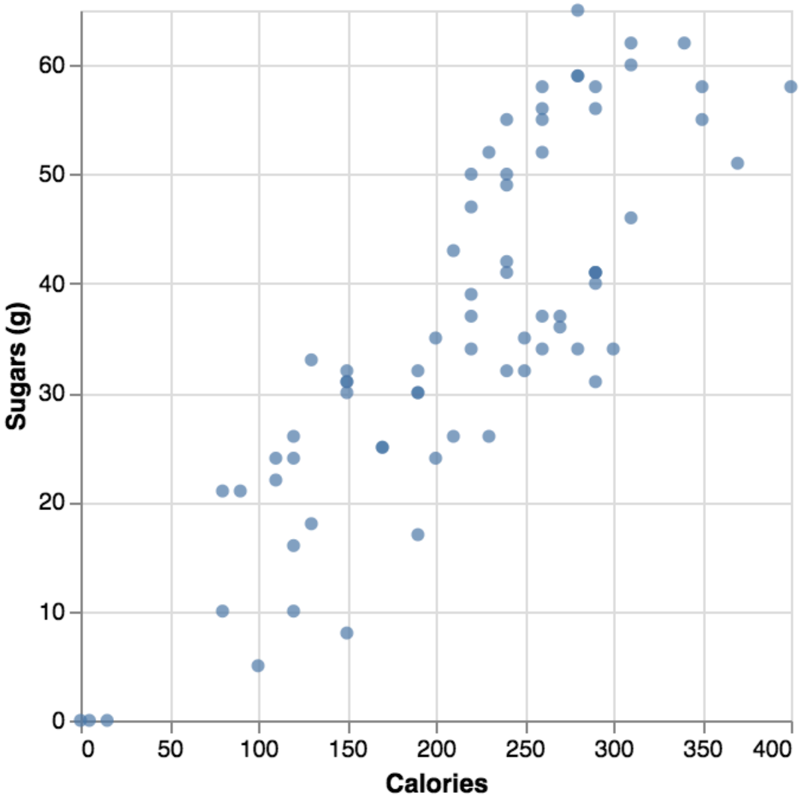
```
var small = [
  { "weather": "sunny", "temp": 35 },
  { "weather": "sunny", "temp": 38 },
  { "weather": "sunny", "temp": 41 },
  { "weather": "partially sunny", "temp": 29 },
  { "weather": "partially sunny", "temp": 34 },
  { "weather": "rainy", "temp": 30 },
];
```

This data is in `resbaz_az.js`

Exercise: Now that we’ve seen the small dataset, try a larger one

Replicate this plot with the Kaggle Starbucks nutritional information data. Don’t forget to add a tooltip!

```
"data": {
 "values": drinks
}
```



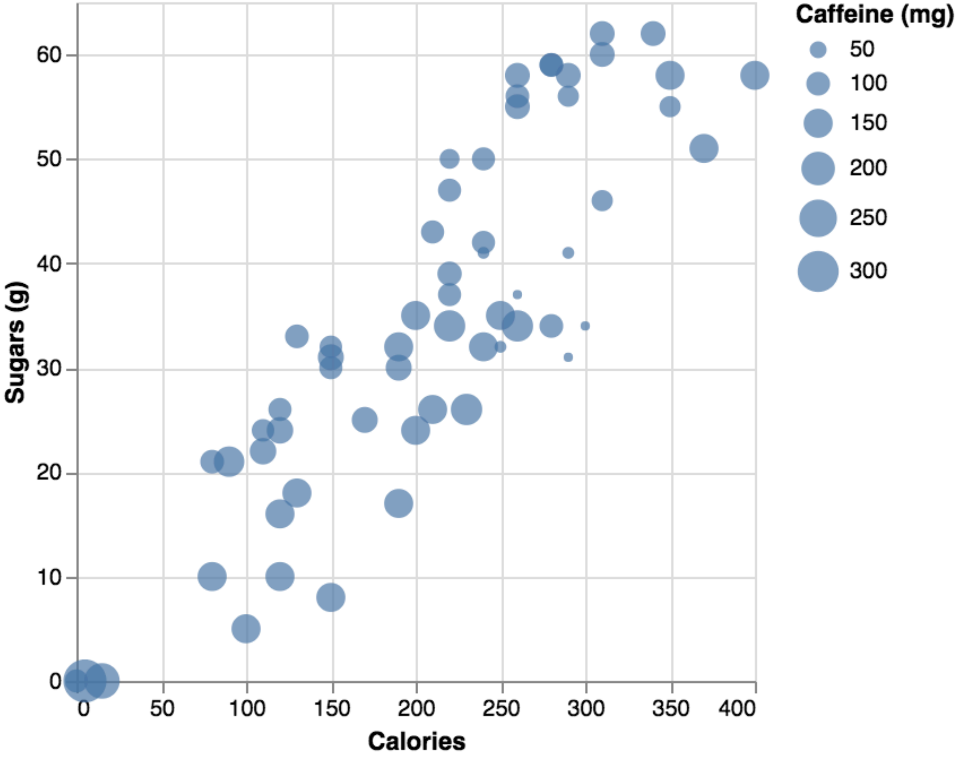
Encoding: Mapping Data to Channels

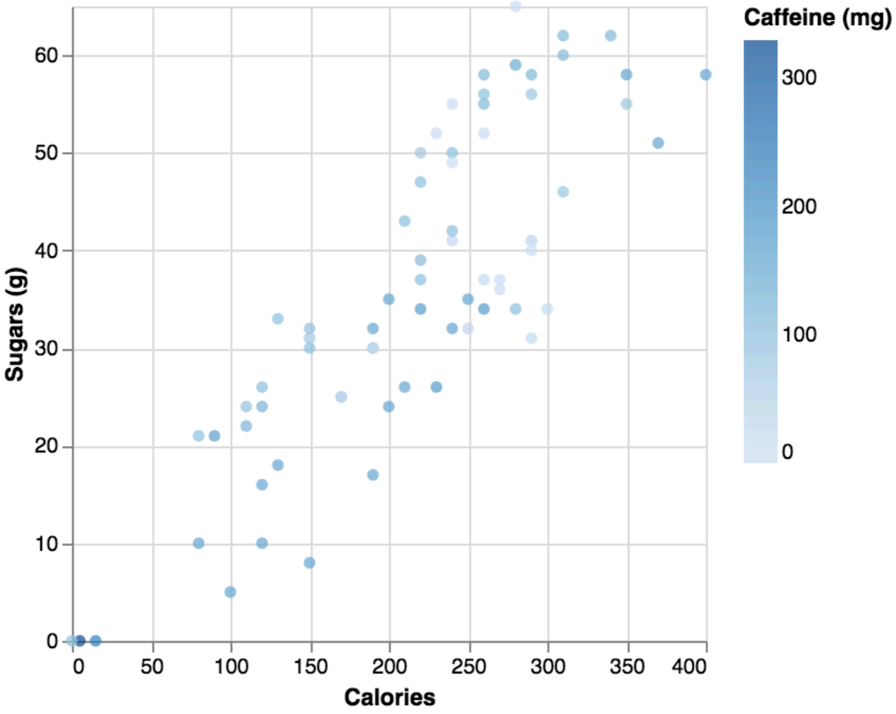
- x
- y
- x2
- y2
- xError
- yError
- xError2
- yError2

- color
- opacity
- fillOpacity
- strokeOpacity
- strokeWidth
- size
- shape

- text
- tooltip
- href
- …more…

Exercise: Let’s encode Caffeine (mg) with size or color





Aggregation of Data

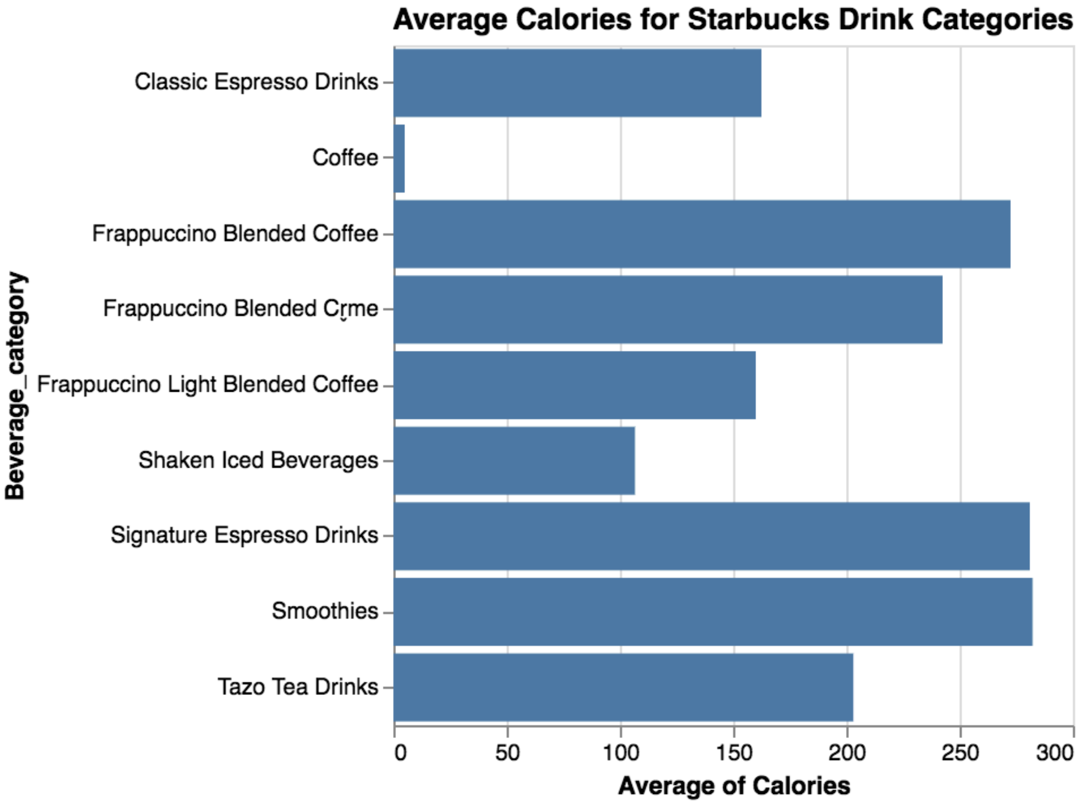
count
sum
mean
average
median
variance
stdev
stderr
min
max
valid
missing
distinct
…more…

```
"encoding": {
  "x": {
    "field": "column_name",
    "type": "quantitative",
    "aggregate": "average"
  }
}
```

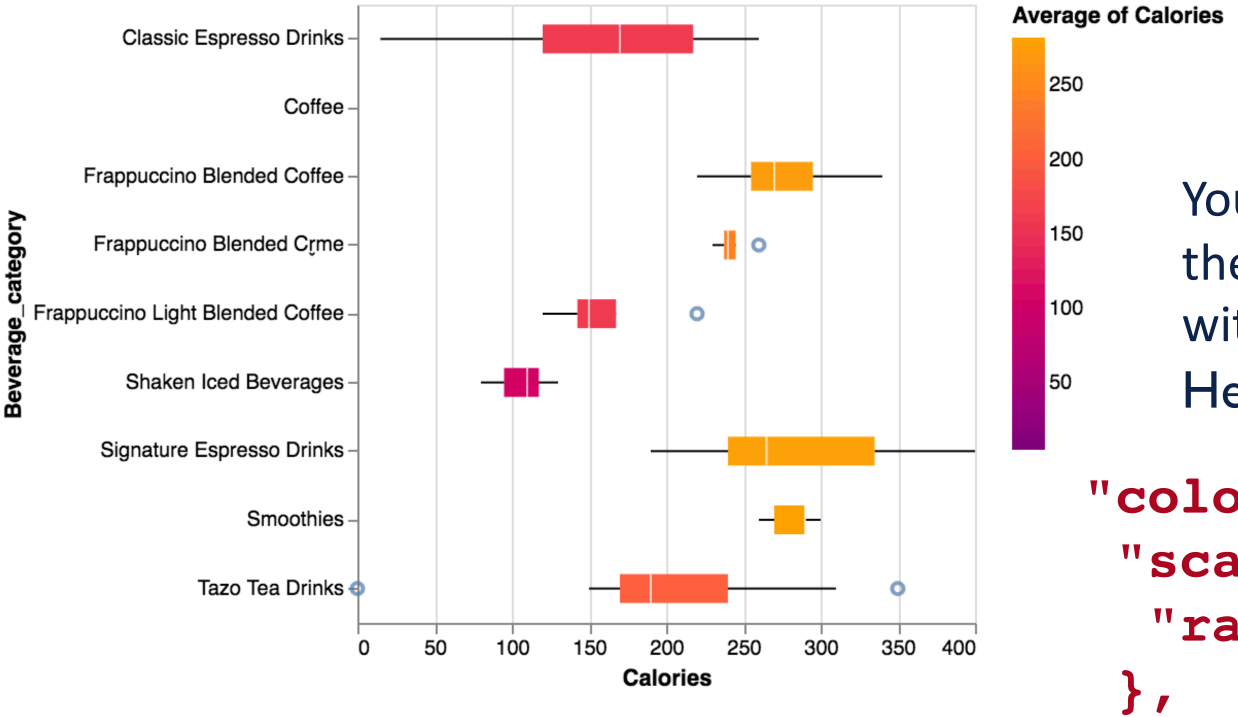
See also binning (histograms) and other transforms…

https://vega.github.io/vega-lite/docs/encoding.html

Exercise: Can you replicate the chart with the Starbucks Data?



Exercise: Replicate this chart



You can specify details on the scale of an encoding with the **scale** parameter. Here’s a color example:

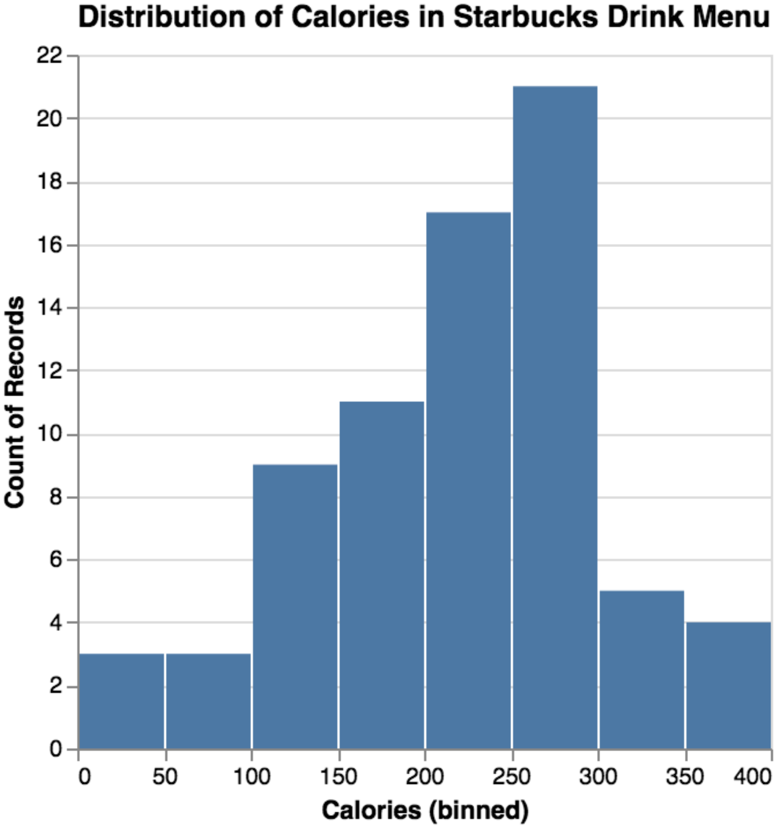
```
"color": {
 "scale": {
  "range": ["red", "blue"]
 },
 "field": "column_name",
 ”type": "quantitative"
}
```

Aggregation of Data - Histograms

count
sum
mean
average
median
variance
stdev
stderr
min
max
valid
missing
distinct
…more…

```
"encoding": {
  "x": {
    "field": "column_name",
    "type": "quantitative",
    "bin": true
  },
 ”y": {
    "type": "quantitative",
    "aggregate": ”count"
  },
}
```

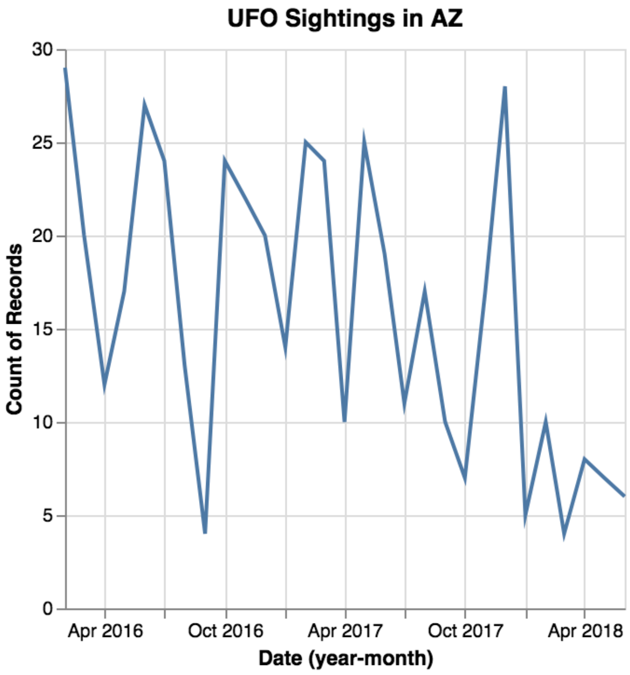
Exercise: Can you replicate this histogram with the Starbucks Data?



Temporal Data

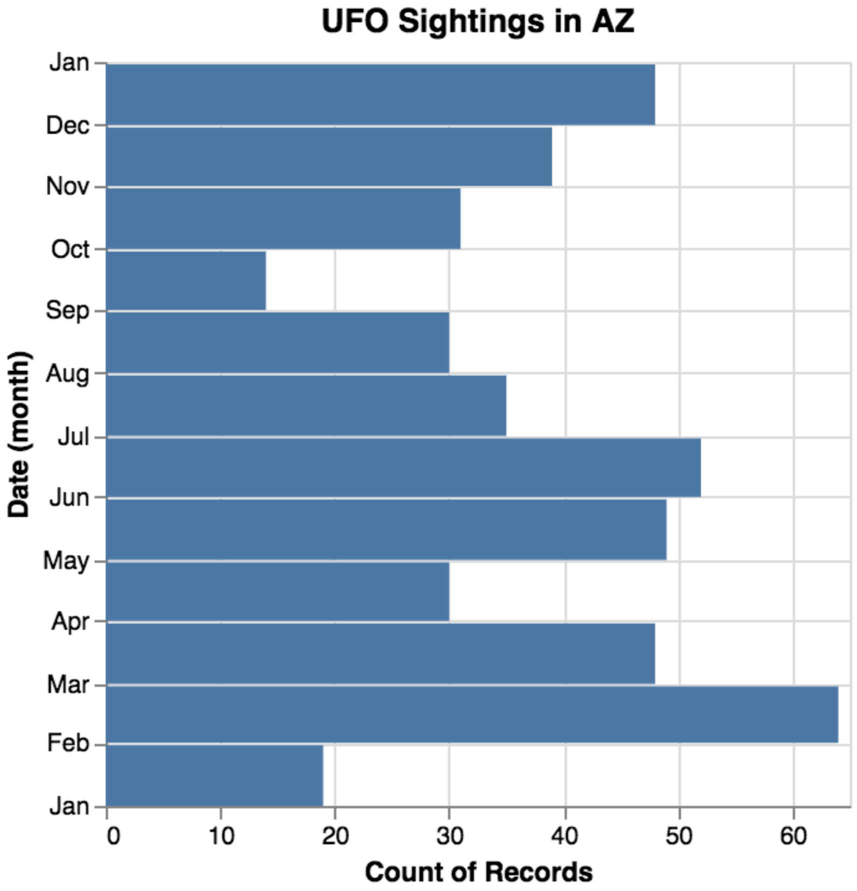
We can set a timeUnit in the encoding to group data and then represent its aggregate:

```
"x": {
    "field": "Date",
    "timeUnit": "yearmonth",
    "type": "temporal"
}
"y": {
    "aggregate": "count",
    "type": "quantative"
}
```



Exercise: Create this chart with the UFO Data

```
"data": {
 "values" : ufos
}
```



Acknowledgements

This workshop is based on the tutorials and documentation at https://vega.github.io

Data Visualization basics are based on Visualization Analysis and Design, by Tamara Munzner

This workshop was funded by the National Science Foundation, under project NSF IIS-1844573