CSCI 1377

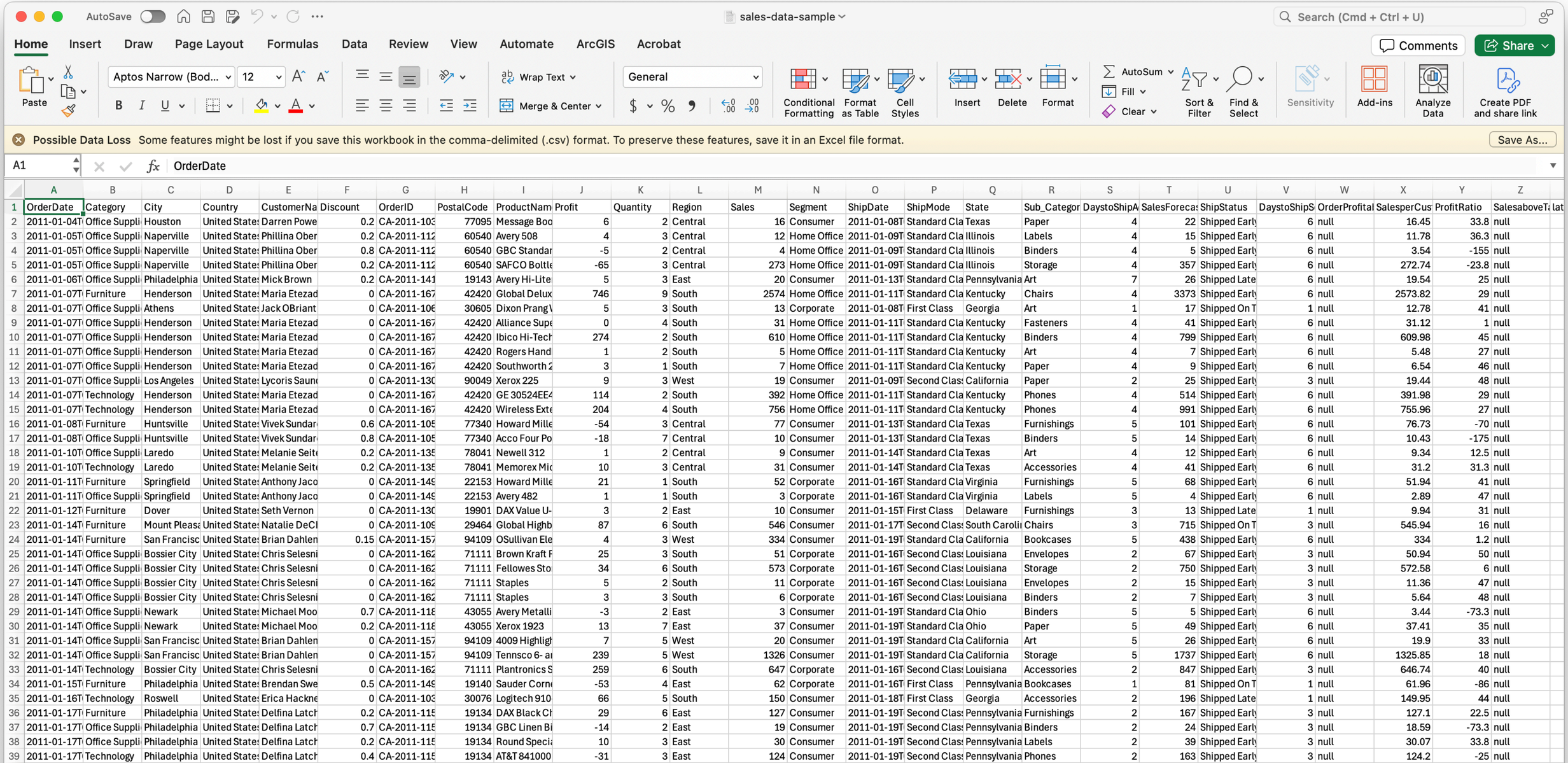
Tools for Thought

Visualization II
Data Visualization Systems

A man who has carefully investigated a printed table finds, when done, that he as only a very faint and partial idea of what he has read, and that like a figure imprinted on sand, it is soon totally erased and defaced.

— William Playfair, *The Commercial and Political Atlas* (1787)

How do cities compare per segment and quarter?



AutoSave sales-data-sample Search (Cmd + Ctrl + U)

Home Insert Draw Page Layout Formulas Data Review View Automate ArcGIS Acrobat Comments Share

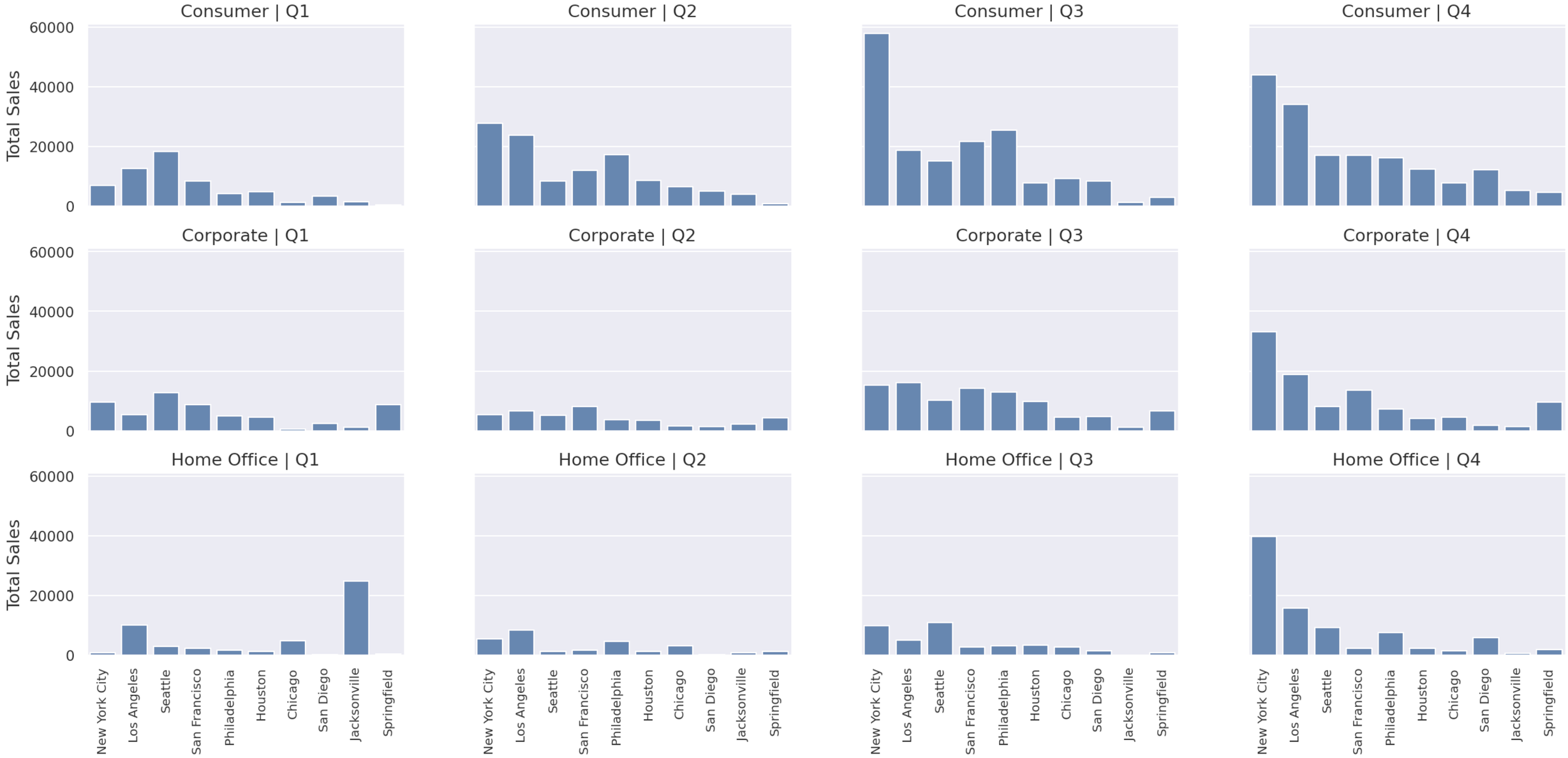
Paste Aptos Narrow (Bod... 12 Wrap Text General Conditional Formatting Format as Table Cell Styles Insert Delete Format AutoSum Fill Clear Sort & Filter Find & Select Sensitivity Add-ins Analyze Data Create PDF and share link

Possible Data Loss Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format. Save As...

A1 OrderDate

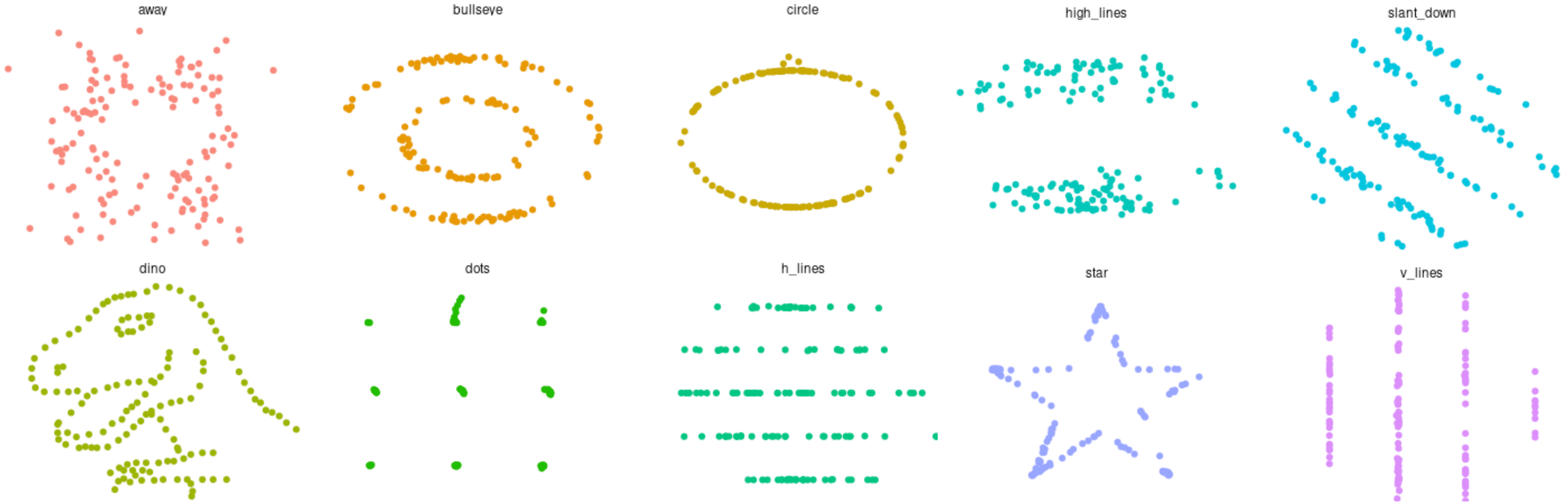
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	OrderDate	Category	City	Country	CustomerNa	Discount	OrderID	PostalCode	ProductNam	Profit	Quantity	Region	Sales	Segment	ShipDate	ShipMode	State	Sub_Categor	DaystoShipA	SalesForeca	ShipStatus	DaystoShipS	OrderProfital	SalesperCus	ProfitRatio	SalesaboveT
2	2011-01-04T	Office Suppli	Houston	United State	Darren Powe	0.2	CA-2011-103	77095	Message Boo	6	2	Central	16	Consumer	2011-01-08T	Standard Cla	Texas	Paper	4	22	Shipped Early	6	null	16.45	33.8	null
3	2011-01-05T	Office Suppli	Naperville	United State	Phillina Ober	0.2	CA-2011-112	60540	Avery 508	4	3	Central	12	Home Office	2011-01-09T	Standard Cla	Illinois	Labels	4	15	Shipped Early	6	null	11.78	36.3	null
4	2011-01-05T	Office Suppli	Naperville	United State	Phillina Ober	0.8	CA-2011-112	60540	GBC Standar	-5	2	Central	4	Home Office	2011-01-09T	Standard Cla	Illinois	Binders	4	5	Shipped Early	6	null	3.54	-155	null
5	2011-01-05T	Office Suppli	Naperville	United State	Phillina Ober	0.2	CA-2011-112	60540	SAFCO Boltle	-65	3	Central	273	Home Office	2011-01-09T	Standard Cla	Illinois	Storage	4	357	Shipped Early	6	null	272.74	-23.8	null
6	2011-01-06T	Office Suppli	Philadelphia	United State	Mick Brown	0.2	CA-2011-141	19143	Avery Hi-Lite	5	3	East	20	Consumer	2011-01-13T	Standard Cla	Pennsylvania	Art	7	26	Shipped Late	6	null	19.54	25	null
7	2011-01-07T	Furniture	Henderson	United State	Maria Etezad	0	CA-2011-167	42420	Global Delux	746	9	South	2574	Home Office	2011-01-11T	Standard Cla	Kentucky	Chairs	4	3373	Shipped Early	6	null	2573.82	29	null
8	2011-01-07T	Office Suppli	Athens	United State	Jack OBriant	0	CA-2011-106	30605	Dixon Prang	5	3	South	13	Corporate	2011-01-08T	First Class	Georgia	Art	1	17	Shipped On T	1	null	12.78	41	null
9	2011-01-07T	Office Suppli	Henderson	United State	Maria Etezad	0	CA-2011-167	42420	Alliance Supe	0	4	South	31	Home Office	2011-01-11T	Standard Cla	Kentucky	Fasteners	4	41	Shipped Early	6	null	31.12	1	null
10	2011-01-07T	Office Suppli	Henderson	United State	Maria Etezad	0	CA-2011-167	42420	Ibico Hi-Tech	274	2	South	610	Home Office	2011-01-11T	Standard Cla	Kentucky	Binders	4	799	Shipped Early	6	null	609.98	45	null
11	2011-01-07T	Office Suppli	Henderson	United State	Maria Etezad	0	CA-2011-167	42420	Rogers Hand	1	2	South	5	Home Office	2011-01-11T	Standard Cla	Kentucky	Art	4	7	Shipped Early	6	null	5.48	27	null
12	2011-01-07T	Office Suppli	Henderson	United State	Maria Etezad	0	CA-2011-167	42420	Southworth 2	3	1	South	7	Home Office	2011-01-11T	Standard Cla	Kentucky	Paper	4	9	Shipped Early	6	null	6.54	46	null
13	2011-01-07T	Office Suppli	Los Angeles	United State	Lycoris Saun	0	CA-2011-130	90049	Xerox 225	9	3	West	19	Consumer	2011-01-09T	Second Class	California	Paper	2	25	Shipped Early	3	null	19.44	48	null
14	2011-01-07T	Technology	Henderson	United State	Maria Etezad	0	CA-2011-167	42420	GE 30524EE4	114	2	South	392	Home Office	2011-01-11T	Standard Cla	Kentucky	Phones	4	514	Shipped Early	6	null	391.98	29	null
15	2011-01-07T	Technology	Henderson	United State	Maria Etezad	0	CA-2011-167	42420	Wireless Exte	204	4	South	756	Home Office	2011-01-11T	Standard Cla	Kentucky	Phones	4	991	Shipped Early	6	null	755.96	27	null
16	2011-01-08T	Furniture	Huntsville	United State	Vivek Sundar	0.6	CA-2011-105	77340	Howard Mille	-54	3	Central	77	Consumer	2011-01-13T	Standard Cla	Texas	Furnishings	5	101	Shipped Early	6	null	76.73	-70	null
17	2011-01-08T	Office Suppli	Huntsville	United State	Vivek Sundar	0.8	CA-2011-105	77340	Acco Four Po	-18	7	Central	10	Consumer	2011-01-13T	Standard Cla	Texas	Binders	5	14	Shipped Early	6	null	10.43	-175	null
18	2011-01-10T	Office Suppli	Laredo	United State	Melanie Seit	0.2	CA-2011-135	78041	Newell 312	1	2	Central	9	Consumer	2011-01-14T	Standard Cla	Texas	Art	4	12	Shipped Early	6	null	9.34	12.5	null
19	2011-01-10T	Technology	Laredo	United State	Melanie Seit	0.2	CA-2011-135	78041	Memorex Mic	10	3	Central	31	Consumer	2011-01-14T	Standard Cla	Texas	Accessories	4	41	Shipped Early	6	null	31.2	31.3	null
20	2011-01-11T	Furniture	Springfield	United State	Anthony Jaco	0	CA-2011-149	22153	Howard Mille	21	1	South	52	Corporate	2011-01-16T	Standard Cla	Virginia	Furnishings	5	68	Shipped Early	6	null	51.94	41	null
21	2011-01-11T	Office Suppli	Springfield	United State	Anthony Jaco	0	CA-2011-149	22153	Avery 482	1	1	South	3	Corporate	2011-01-16T	Standard Cla	Virginia	Labels	5	4	Shipped Early	6	null	2.89	47	null
22	2011-01-12T	Furniture	Dover	United State	Seth Vernon	0	CA-2011-130	19901	DAX Value U-	3	2	East	10	Consumer	2011-01-15T	First Class	Delaware	Furnishings	3	13	Shipped Late	1	null	9.94	31	null
23	2011-01-14T	Furniture	Mount Pleasa	United State	Natalie DeCl	0	CA-2011-109	29464	Global Highb	87	6	South	546	Consumer	2011-01-17T	Second Class	South Caroli	Chairs	3	715	Shipped On T	3	null	545.94	16	null
24	2011-01-14T	Furniture	San Francisc	United State	Brian Dahlen	0.15	CA-2011-157	94109	OSullivan Ele	4	3	West	334	Consumer	2011-01-19T	Standard Cla	California	Bookcases	5	438	Shipped Early	6	null	334	1.2	null
25	2011-01-14T	Office Suppli	Bossier City	United State	Chris Selesni	0	CA-2011-162	71111	Brown Kraft F	25	3	South	51	Corporate	2011-01-16T	Second Class	Louisiana	Envelopes	2	67	Shipped Early	3	null	50.94	50	null
26	2011-01-14T	Office Suppli	Bossier City	United State	Chris Selesni	0	CA-2011-162	71111	Fellowes Sto	34	6	South	573	Corporate	2011-01-16T	Second Class	Louisiana	Storage	2	750	Shipped Early	3	null	572.58	6	null
27	2011-01-14T	Office Suppli	Bossier City	United State	Chris Selesni	0	CA-2011-162	71111	Staples	5	2	South	11	Corporate	2011-01-16T	Second Class	Louisiana	Envelopes	2	15	Shipped Early	3	null	11.36	47	null
28	2011-01-14T	Office Suppli	Bossier City	United State	Chris Selesni	0	CA-2011-162	71111	Staples	3	3	South	6	Corporate	2011-01-16T	Second Class	Louisiana	Binders	2	7	Shipped Early	3	null	5.64	48	null
29	2011-01-14T	Office Suppli	Newark	United State	Michael Moo	0.7	CA-2011-118	43055	Avery Metalli	-3	2	East	3	Consumer	2011-01-19T	Standard Cla	Ohio	Binders	5	5	Shipped Early	6	null	3.44	-73.3	null
30	2011-01-14T	Office Suppli	Newark	United State	Michael Moo	0.2	CA-2011-118	43055	Xerox 1923	13	7	East	37	Consumer	2011-01-19T	Standard Cla	Ohio	Paper	5	49	Shipped Early	6	null	37.41	35	null
31	2011-01-14T	Office Suppli	San Francisc	United State	Brian Dahlen	0	CA-2011-157	94109	4009 Highligh	7	5	West	20	Consumer	2011-01-19T	Standard Cla	California	Art	5	26	Shipped Early	6	null	19.9	33	null
32	2011-01-14T	Office Suppli	San Francisc	United State	Brian Dahlen	0	CA-2011-157	94109	Tennsco 6- a	239	5	West	1326	Consumer	2011-01-19T	Standard Cla	California	Storage	5	1737	Shipped Early	6	null	1325.85	18	null
33	2011-01-14T	Technology	Bossier City	United State	Chris Selesni	0	CA-2011-162	71111	Plantronics S	259	6	South	647	Corporate	2011-01-16T	Second Class	Louisiana	Accessories	2	847	Shipped Early	3	null	646.74	40	null
34	2011-01-15T	Furniture	Philadelphia	United State	Brendan Swe	0.5	CA-2011-149	19140	Sauder Corne	-53	4	East	62	Corporate	2011-01-16T	First Class	Pennsylvania	Bookcases	1	81	Shipped On T	1	null	61.96	-86	null
35	2011-01-16T	Technology	Roswell	United State	Erica Hackne	0	CA-2011-103	30076	Logitech 910	66	5	South	150	Consumer	2011-01-18T	First Class	Georgia	Accessories	2	196	Shipped Late	1	null	149.95	44	null
36	2011-01-17T	Furniture	Philadelphia	United State	Delfina Latch	0.2	CA-2011-115	19134	DAX Black Ch	29	6	East	127	Consumer	2011-01-19T	Second Class	Pennsylvania	Furnishings	2	167	Shipped Early	3	null	127.1	22.5	null
37	2011-01-17T	Office Suppli	Philadelphia	United State	Delfina Latch	0.7	CA-2011-115	19134	GBC Linen Bi	-14	2	East	19	Consumer	2011-01-19T	Second Class	Pennsylvania	Binders	2	24	Shipped Early	3	null	18.59	-73.3	null
38	2011-01-17T	Office Suppli	Philadelphia	United State	Delfina Latch	0.2	CA-2011-115	19134	Round Specia	10	3	East	30	Consumer	2011-01-19T	Second Class	Pennsylvania	Labels	2	39	Shipped Early	3	null	30.07	33.8	null
39	2011-01-17T	Technology	Philadelphia	United State	Delfina Latch	0.4	CA-2011-115	19134	AT&T 841000	-31	3	East	124	Consumer	2011-01-19T	Second Class	Pennsylvania	Phones	2	163	Shipped Early	3	null	124.2	-25	null

How do cities compare per segment and quarter?



What do these datasets have in common?

$$\mu_x \approx 54,\ \mu_y \approx 47,\ \sigma_x \approx 16,\ \sigma_y \approx 26,\ r \approx -.06$$



https://jumpingrivers.github.io/datasauRus/

“About 85% of my “thinking” time was spent getting into a position to think [...] Hours went into the plotting of graphs, and other hours into instructing an assistant how to plot. When the graphs were finished, the relations were obvious at once.”

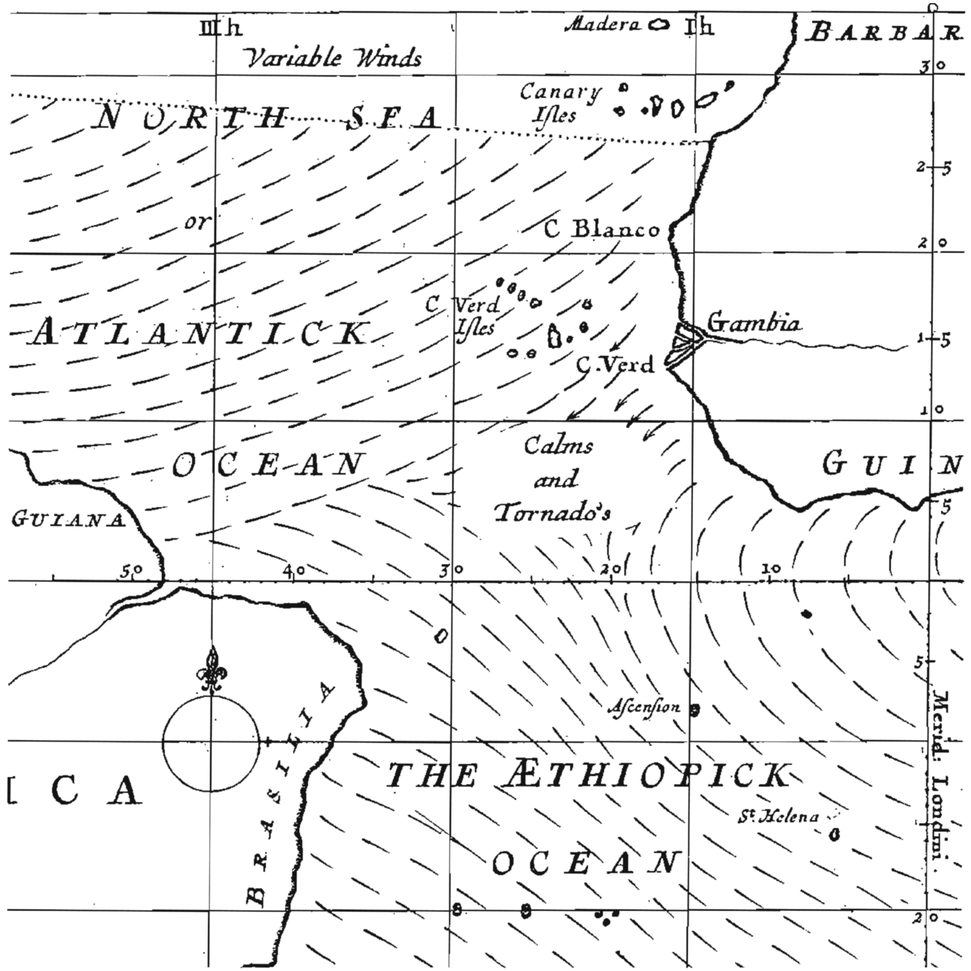
— J. C. R. Licklider, “Man-Computer Symbiosis” (1960)

Last time: how does the human visual system work?

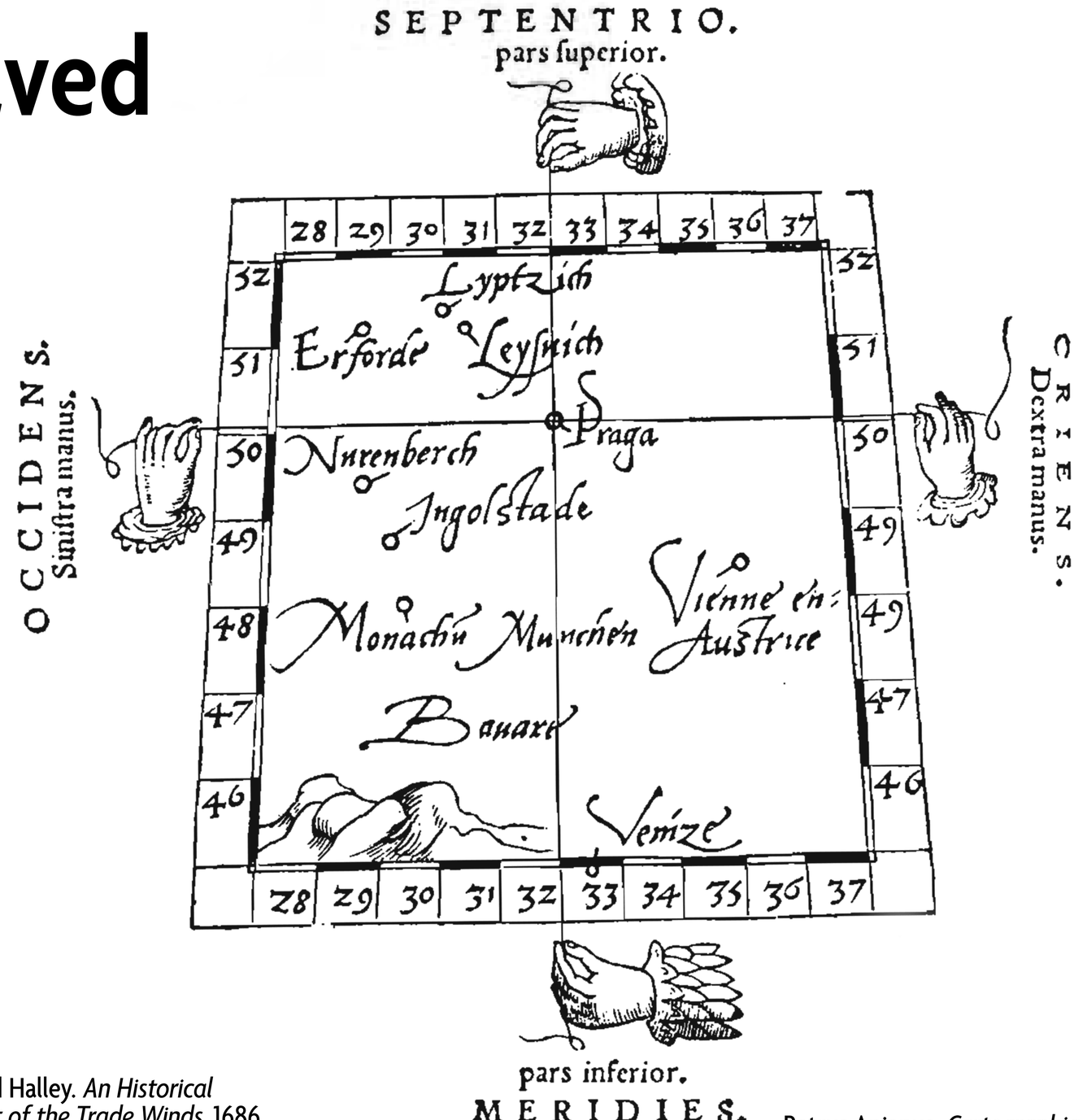
Today: what is the modern way to conceptualize data visualization?

Next time: what are the best practices for visualization design?

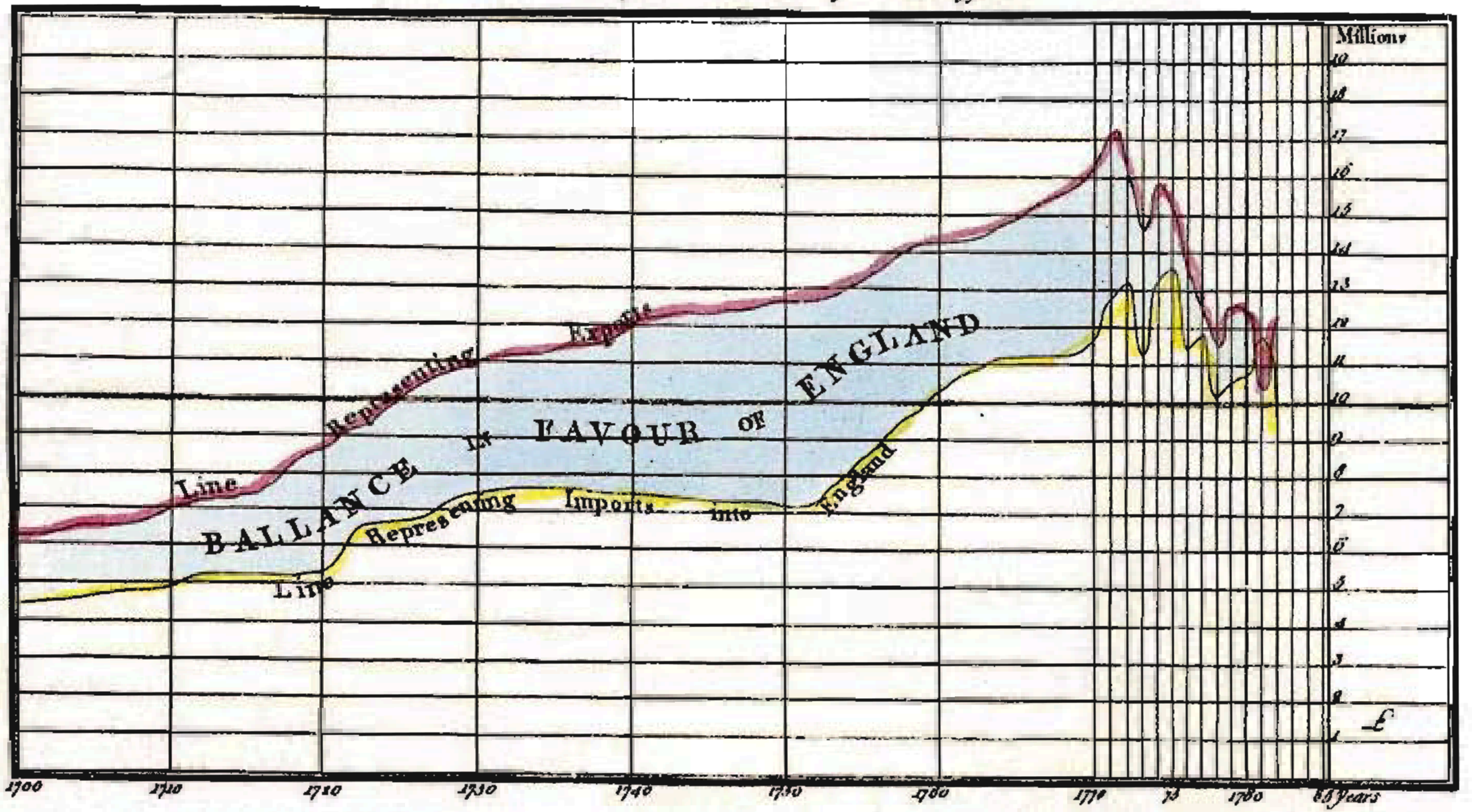
Data visualization evolved from cartography



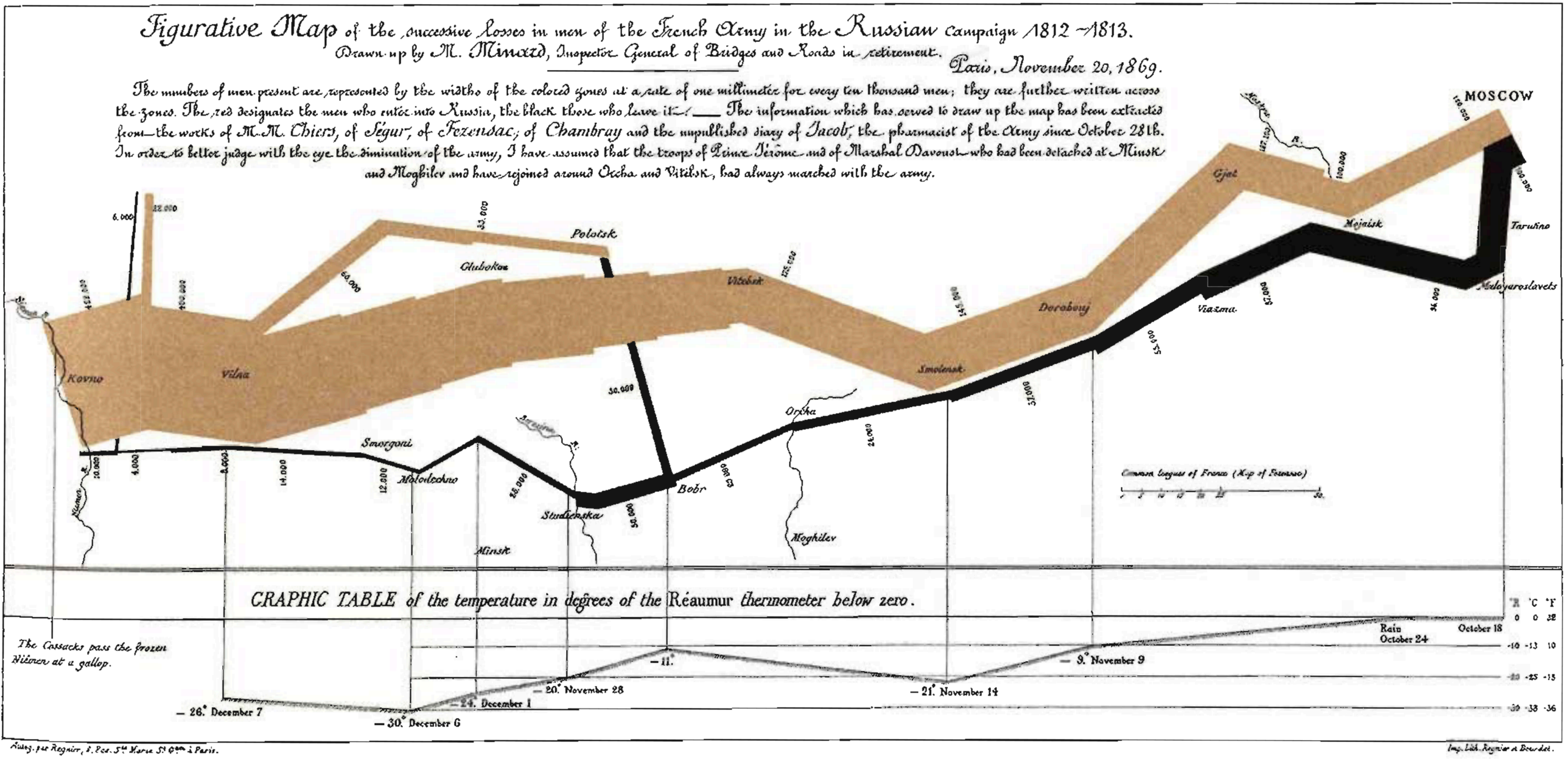
Edmond Halley. *An Historical Account of the Trade Winds.* 1686



Petrus Apianus. *Cartographia.* 1546



William Playfair. *The Commercial and Political Atlas.* 1787



"It may well be the best statistical graphic ever drawn". — Edward Tufte

English translation by Dawn Finley, redrawing by Elaine Morse

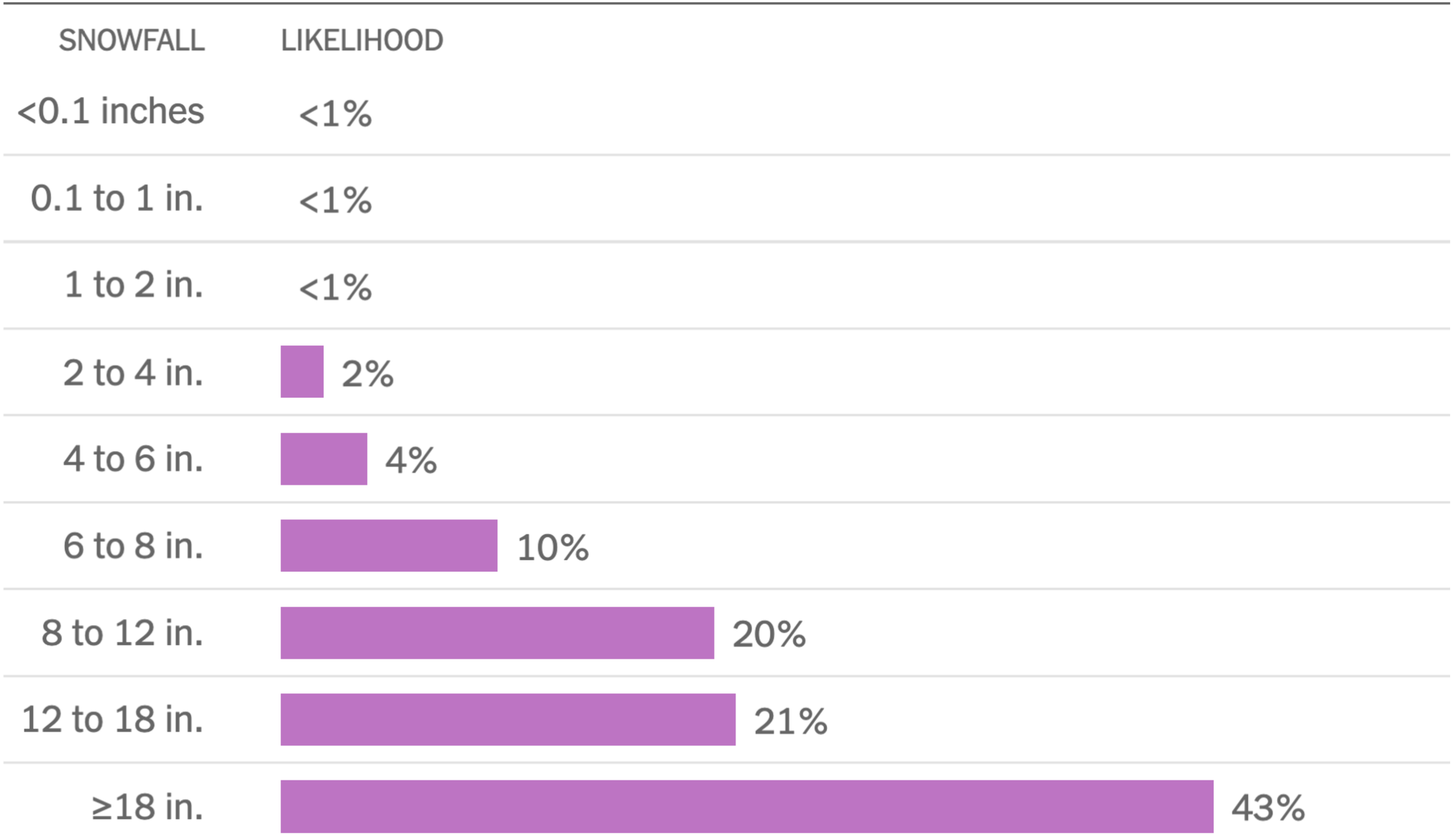
Charles Minard. *Des tableaux graphiques et des cartes figuratives.* 1861

John Snow. *On the Mode of Communication of Cholera.* 1855

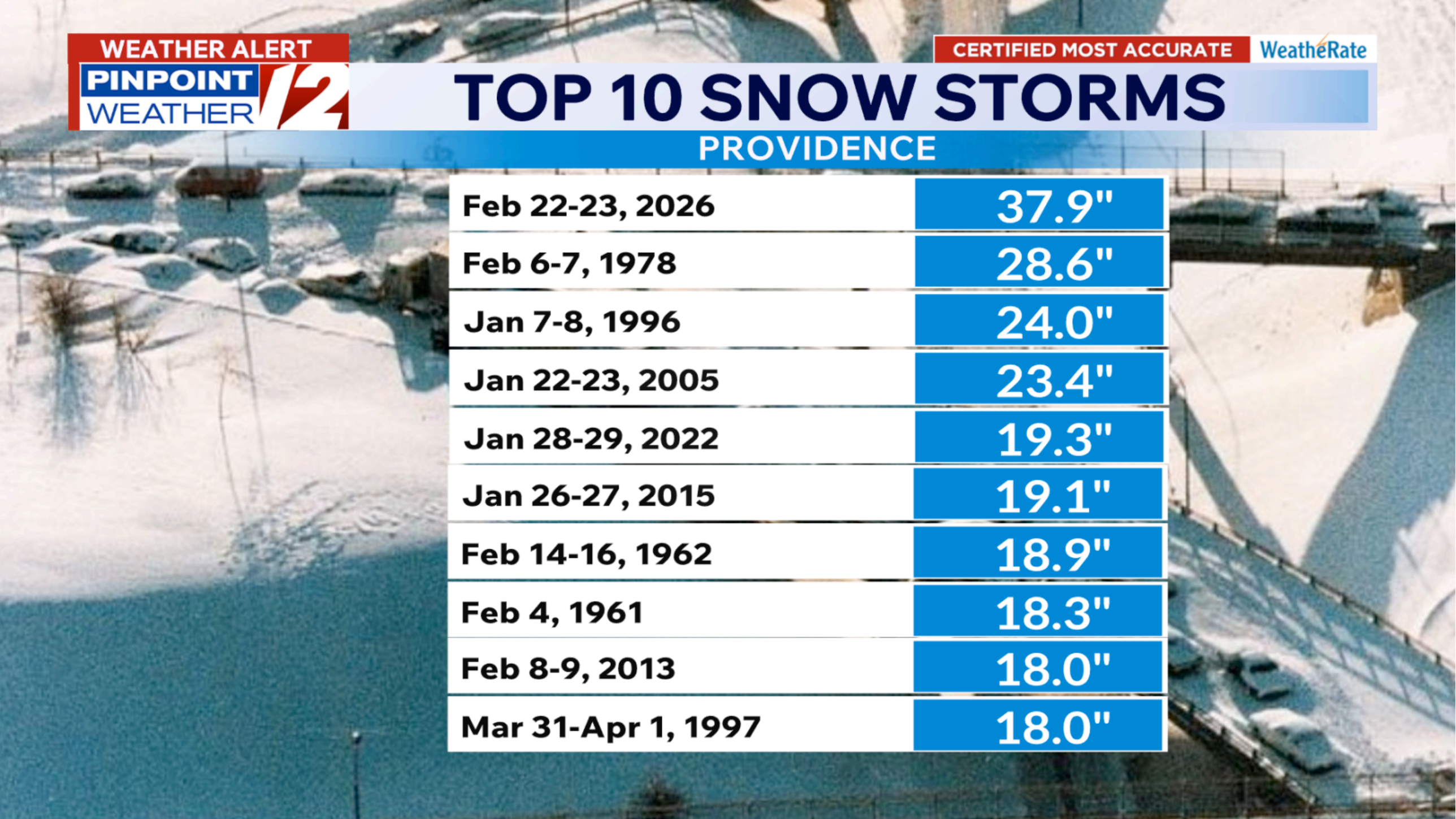
Fast forward 150 years...

How Much More Snow to Expect in Providence, R.I.

Through Tuesday at 7 a.m. Eastern. More snow may fall after this time.

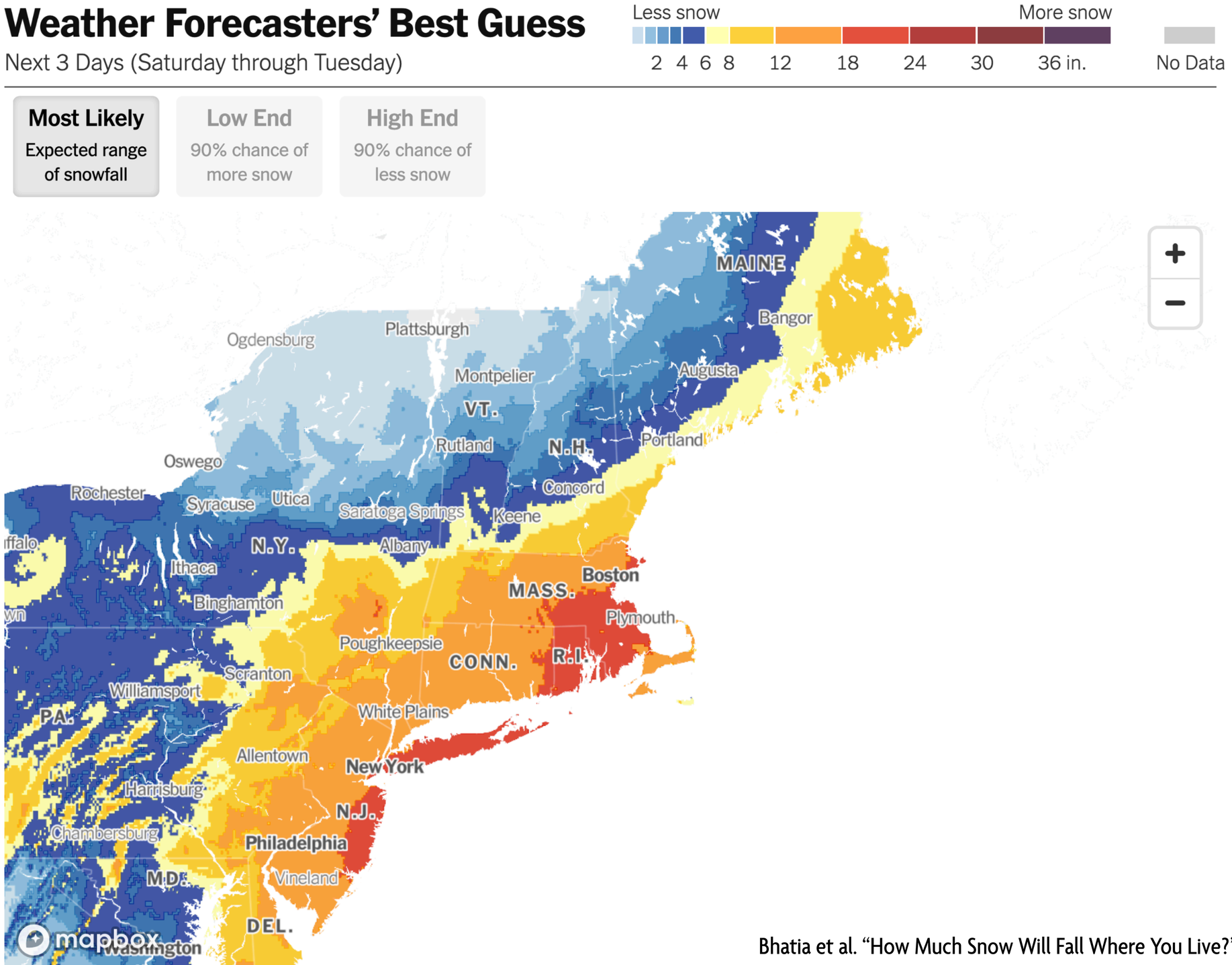


Bhatia et al. "How Much Snow Will Fall Where You Live?" *New York Times*, 2/21/2026

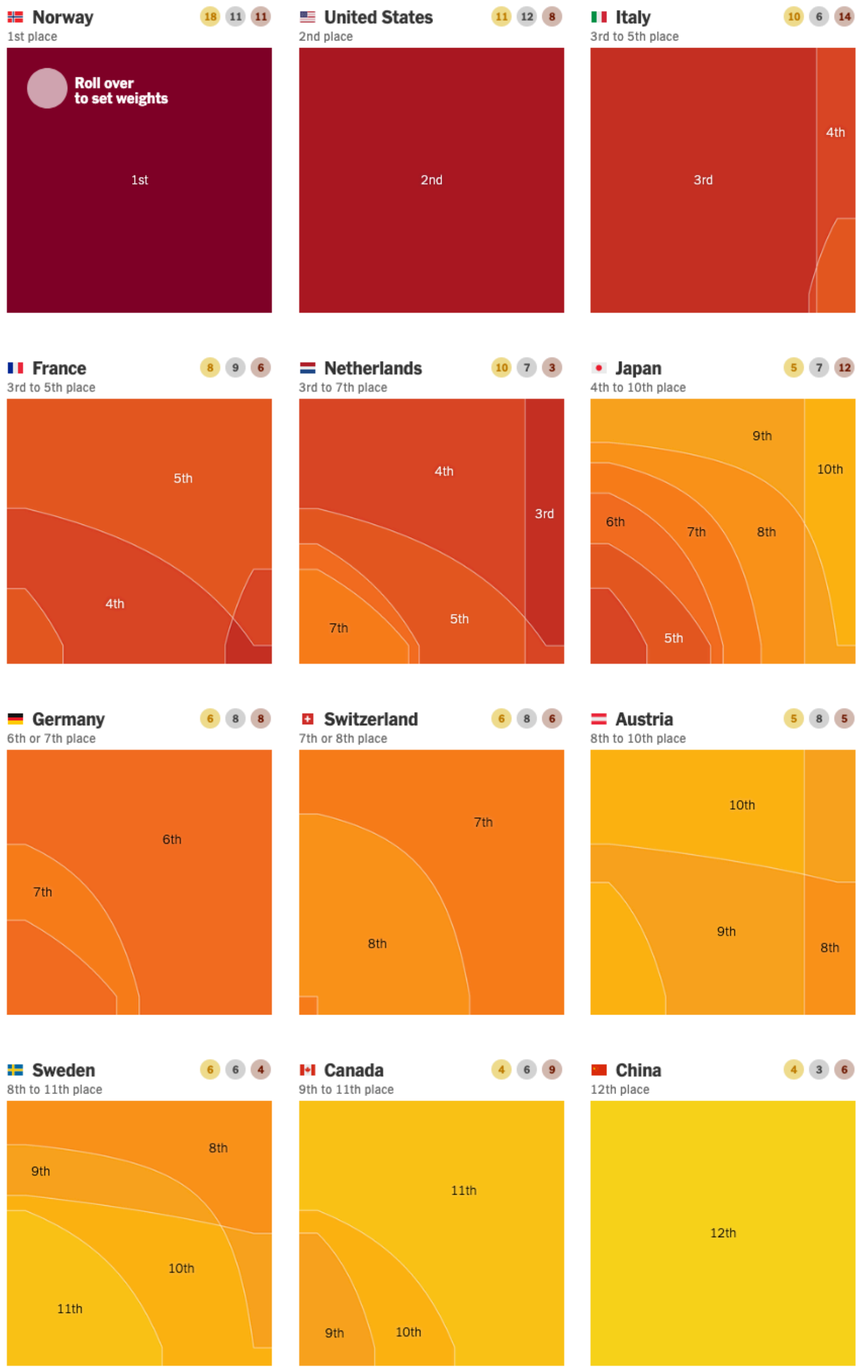

WEATHER ALERT
PINPOINT WEATHER 12
CERTIFIED MOST ACCURATE
WeatheRate
TOP 10 SNOW STORMS
PROVIDENCE
Feb 22-23, 2026 37.9"
Feb 6-7, 1978 28.6"
Jan 7-8, 1996 24.0"
Jan 22-23, 2005 23.4"
Jan 28-29, 2022 19.3"
Jan 26-27, 2015 19.1"
Feb 14-16, 1962 18.9"
Feb 4, 1961 18.3"
Feb 8-9, 2013 18.0"
Mar 31-Apr 1, 1997 18.0"

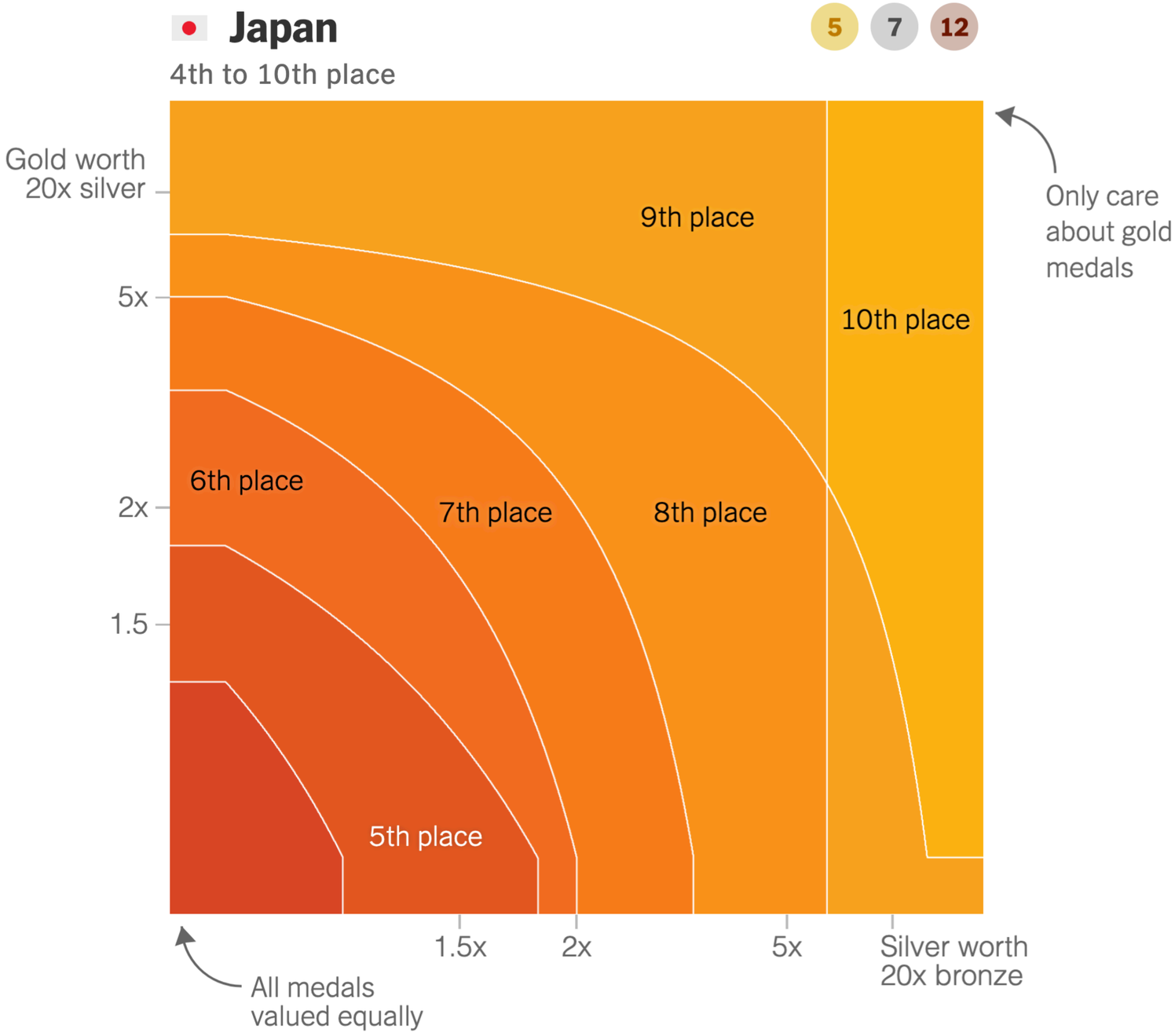
Weather Forecasters' Best Guess

Next 3 Days (Saturday through Tuesday)



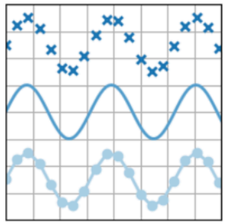
Bhatia et al. "How Much Snow Will Fall Where You Live?" *New York Times*, 2/21/2026



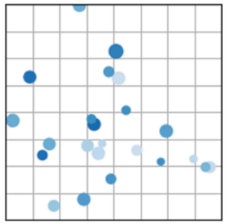


Josh Katz. "Milan-Cortina Olympics: Who Leads the Medal Count?" *New York Times*, 2/21/2026

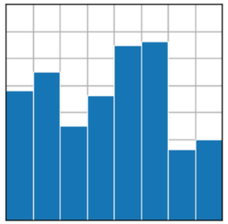
It’s hard to keep up with all these plot types!



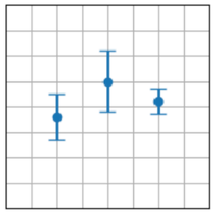
plot(x, y)



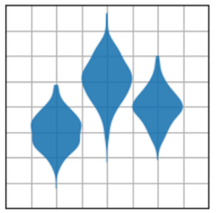
scatter(x, y)



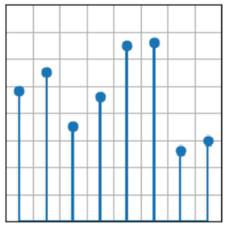
bar(x, height)



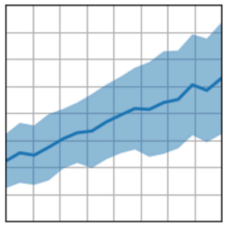
errorbar(x, y, yerr, xerr)



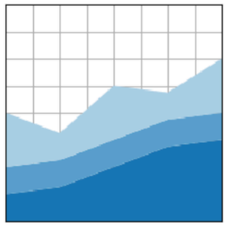
violinplot(D)



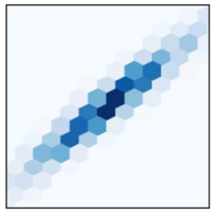
stem(x, y)



fill_between(x, y1, y2)

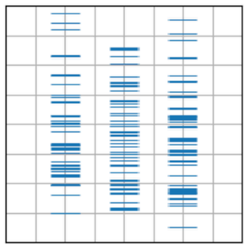


stackplot(x, y)

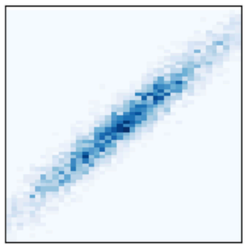


hexbin(x, y, C)

pie(x)

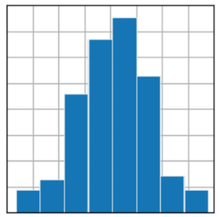


eventplot(D)

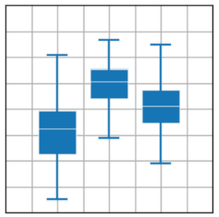


hist2d(x, y)

contour(X, Y, Z)



hist(x)



boxplot(X)

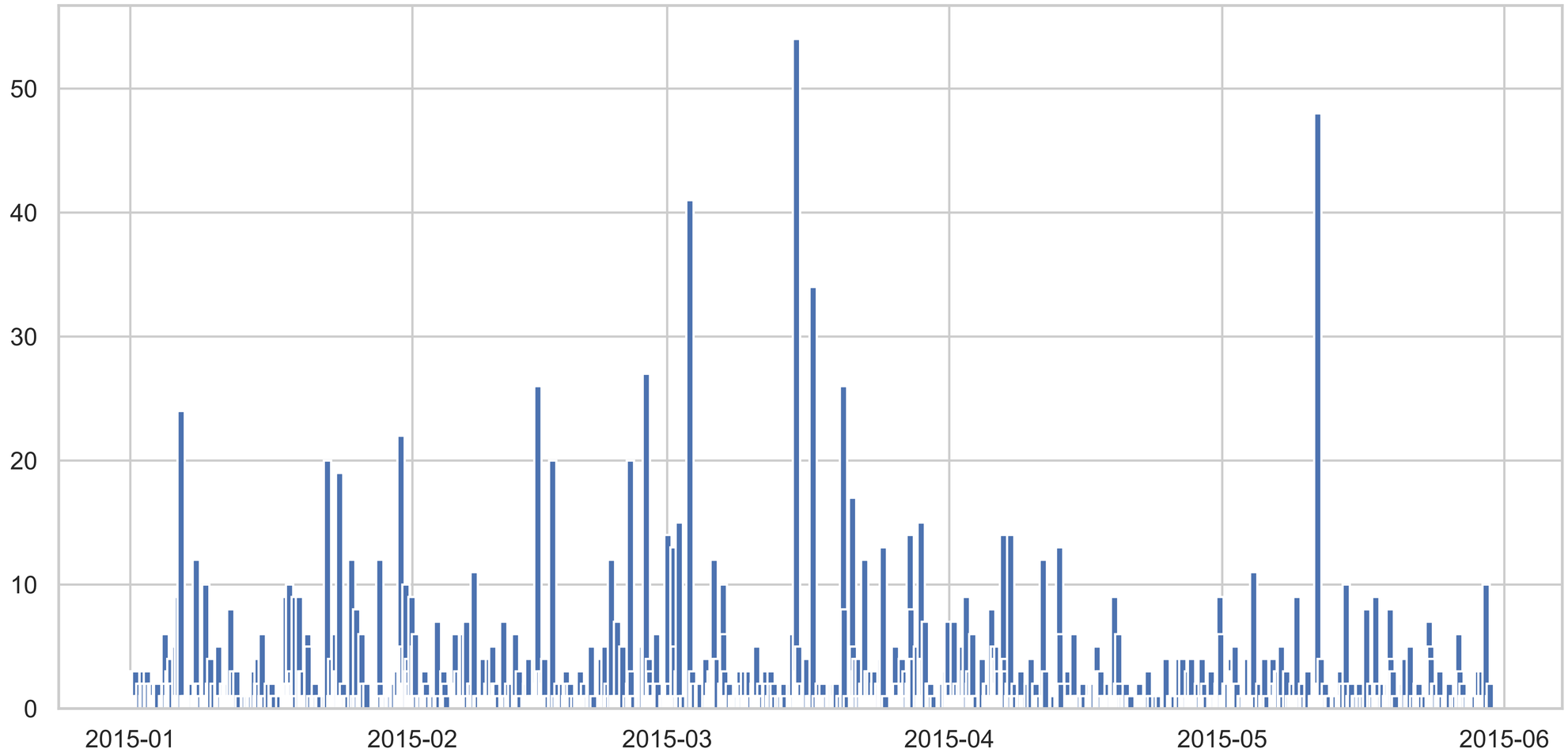
https://matplotlib.org/stable/plot_types/index.html

Modern approaches to data visualization

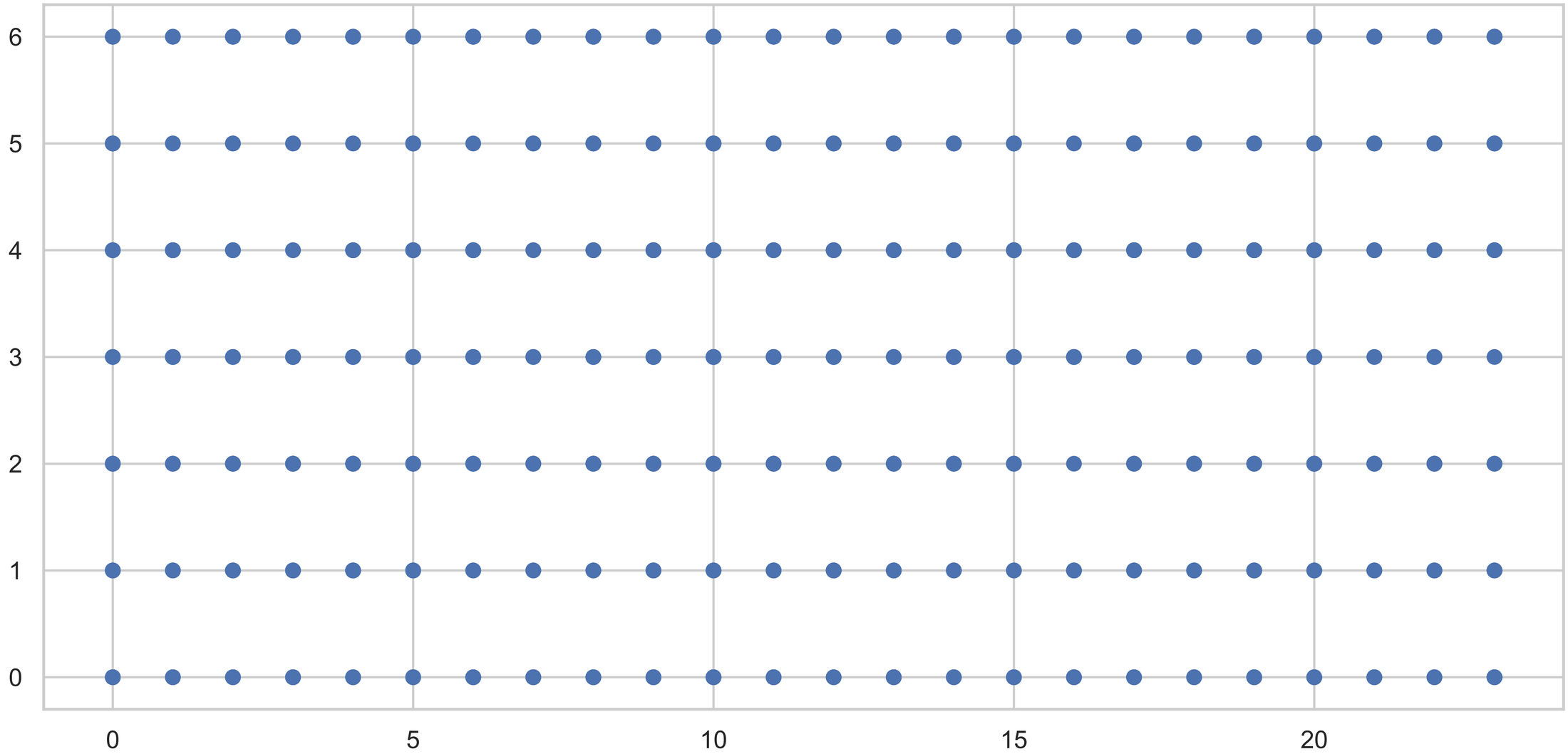
	time	commits
0	2015-01-01 01:00:00	2
1	2015-01-01 04:00:00	3
2	2015-01-01 05:00:00	1
3	2015-01-01 08:00:00	1
4	2015-01-01 09:00:00	3

How many commits did I make each day?

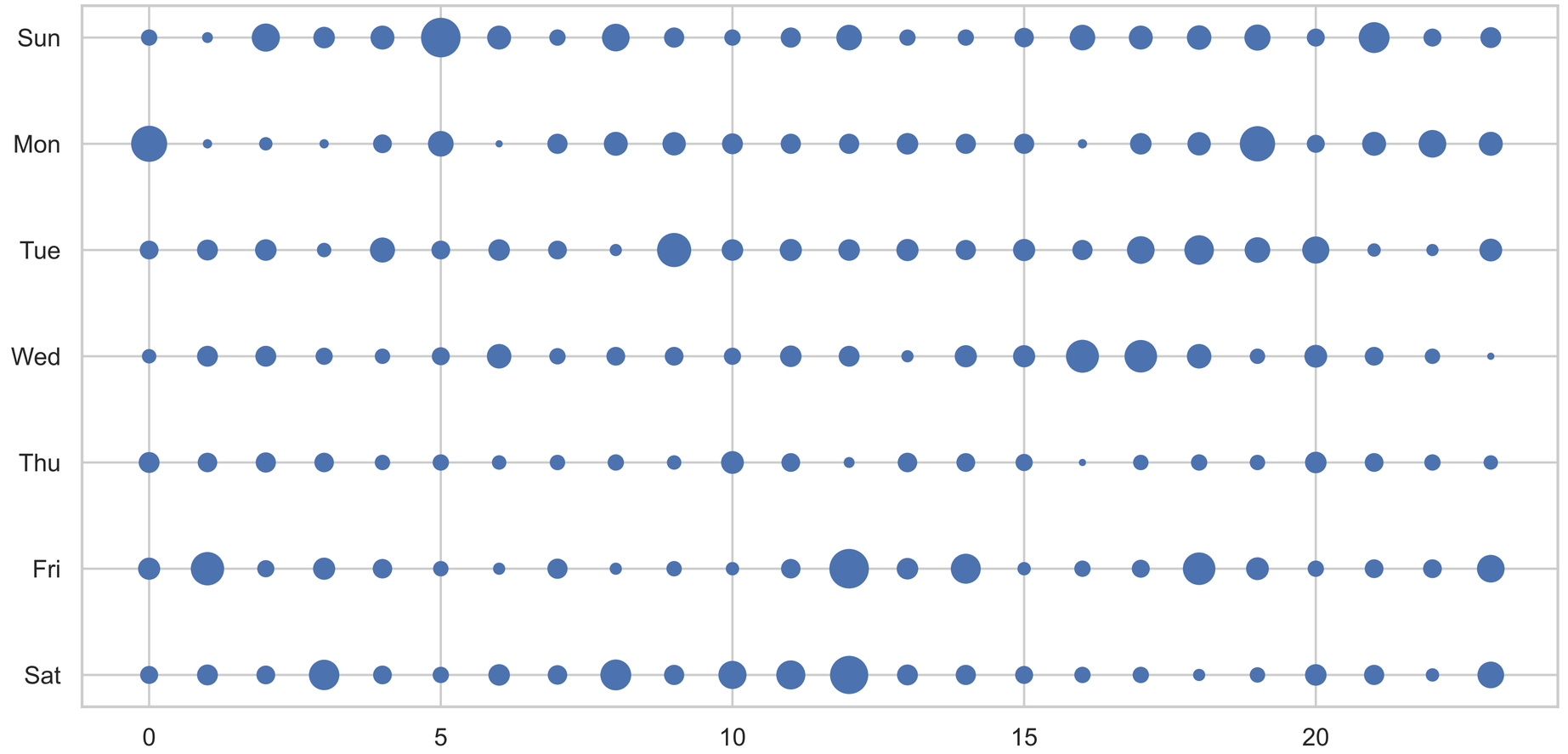
```
import matplotlib.pyplot as plt
plt.bar(github.time, github.commits)
```



```
plt.scatter(github.time.dt.hour, github.time.dt.dayofweek)
```



```
df = github.assign(hour=github.time.dt.hour, day=github.time.dt.dayofweek) \
            .groupby(['day', 'hour']).commits.sum().reset_index()
ax = plt.gca()
ax.scatter(df.hour, df.day, s=df.commits)
ax.set_yticks([6, 5, 4, 3, 2, 1, 0])
ax.set_yticklabels(['Sun','Mon','Tue','Wed','Thu','Fri','Sat'])
```



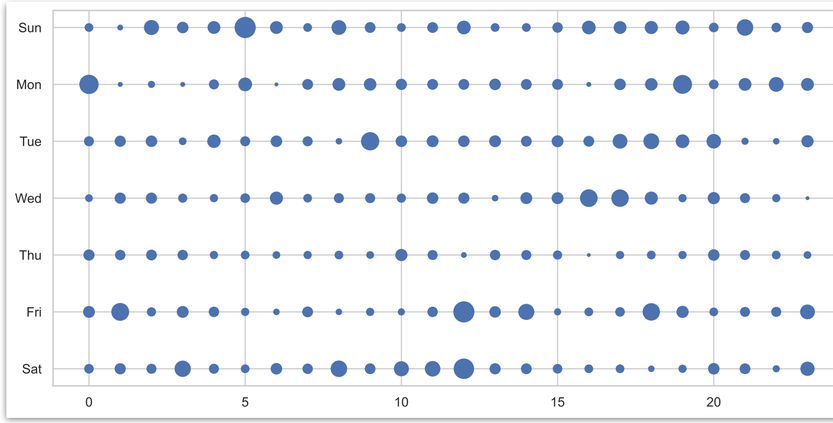
Compute binned sum

```
df = github.assign(
  hour=github.time.dt.hour,
  day=github.time.dt.dayofweek
).groupby(
  ['day', 'hour'],
  as_index=False
).commits
 .sum()

ax = plt.gca()
```

	day	hour	commits
0	0	0	10
1	0	1	14
2	0	2	11
3	0	3	32
4	0	4	11

Create basic plot structure

```
ax.scatter(df.hour, df.day, s=df.commits)
```



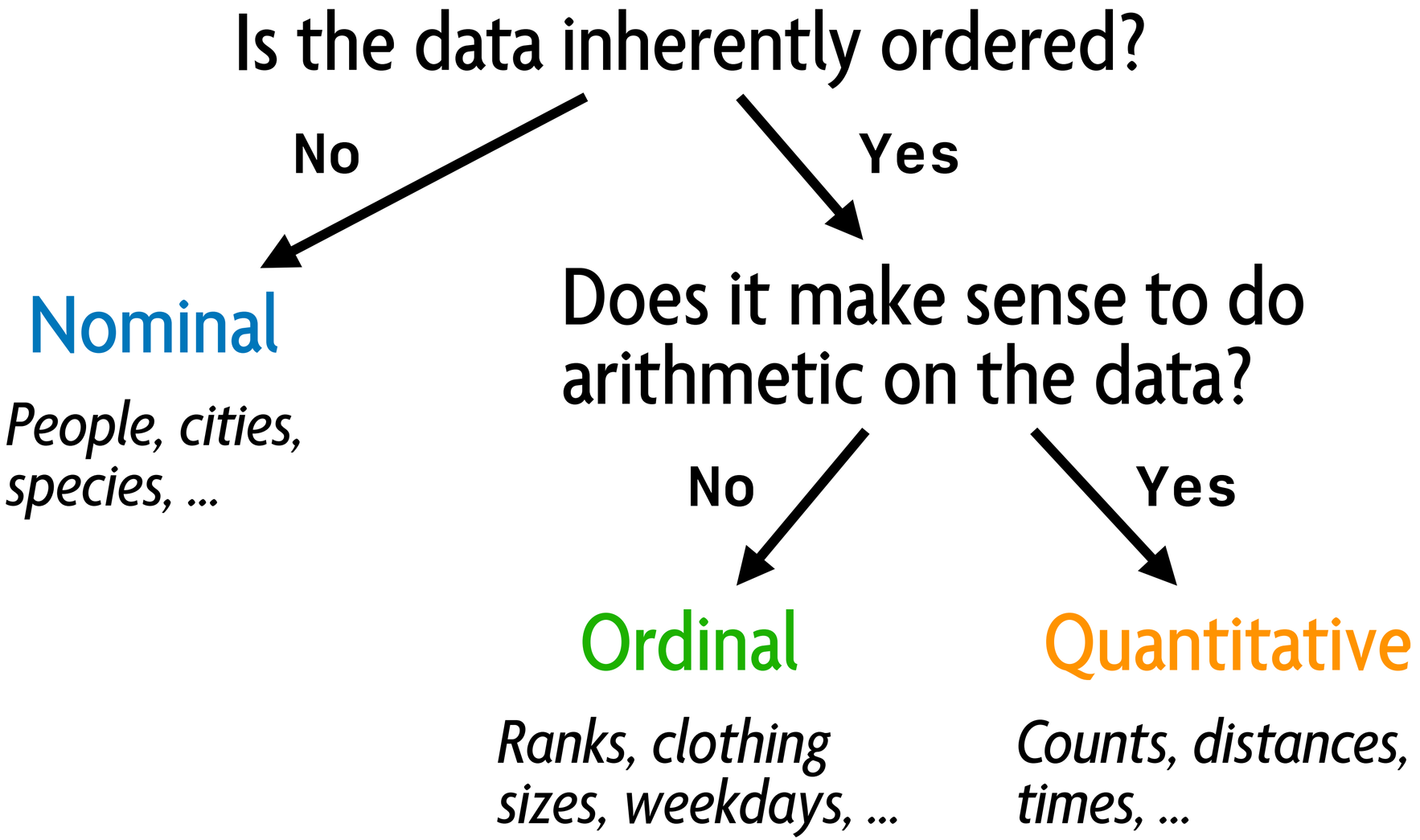
Re-order axis to match top-down reading order

```
ax.set_yticks([6, 5, 4, 3, 2, 1, 0])
```

Add natural language labels matching numeric representation of day

```
ax.set_yticklabels(['Sun','Mon','Tue','Wed','Thu','Fri','Sat'])
```

Data can be classified into “levels of measurement”

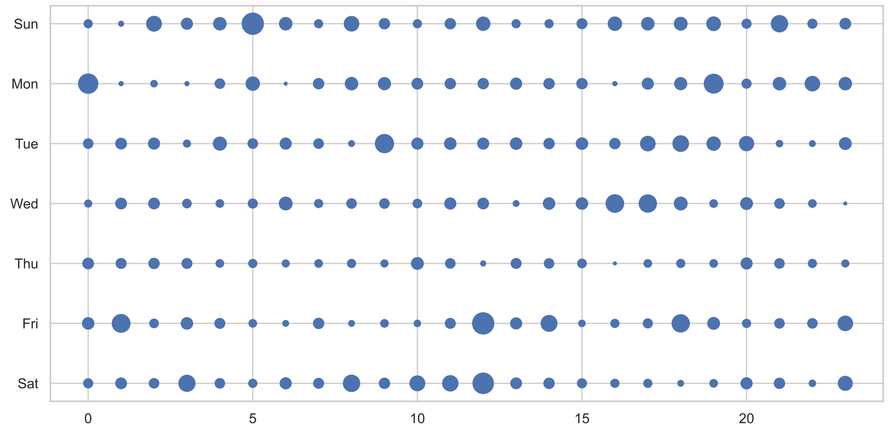


Nominal Ordinal Quantitative

		Q	Q
		time	commits
0	2015-01-01 01:00:00		2
1	2015-01-01 04:00:00		3
2	2015-01-01 05:00:00		1
3	2015-01-01 08:00:00		1
4	2015-01-01 09:00:00		3

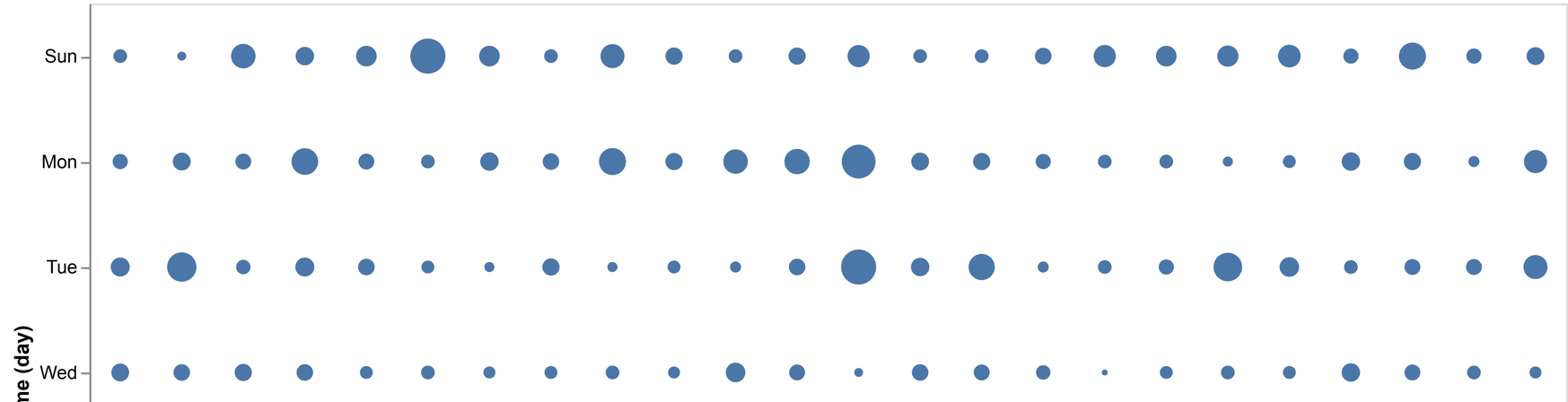
	O	O	Q
	day	hour	commits
0	0	0	10
1	0	1	14
2	0	2	11
3	0	3	32
4	0	4	11

```
github['hour'] = github.time.dt.hour
github['day'] = pd.Categorical(
    github['time'].dt.day_name(),
    categories=[‘Monday','Tuesday','Wednesday',
        'Thursday','Friday','Saturday','Sunday'],
    ordered=True
)

df = github.groupby(
  ['day', ‘hour'], as_index=False
).commits.sum()

ax = plt.gca()
ax.scatter(df.hour, df.day, s=df.commits)
ax.yaxis.set_inverted(True)

sns.scatterplot(data=df, x=‘hour', y=‘day', size=‘commits')
```



Altair is a “Grammar-of-Graphics” Python library

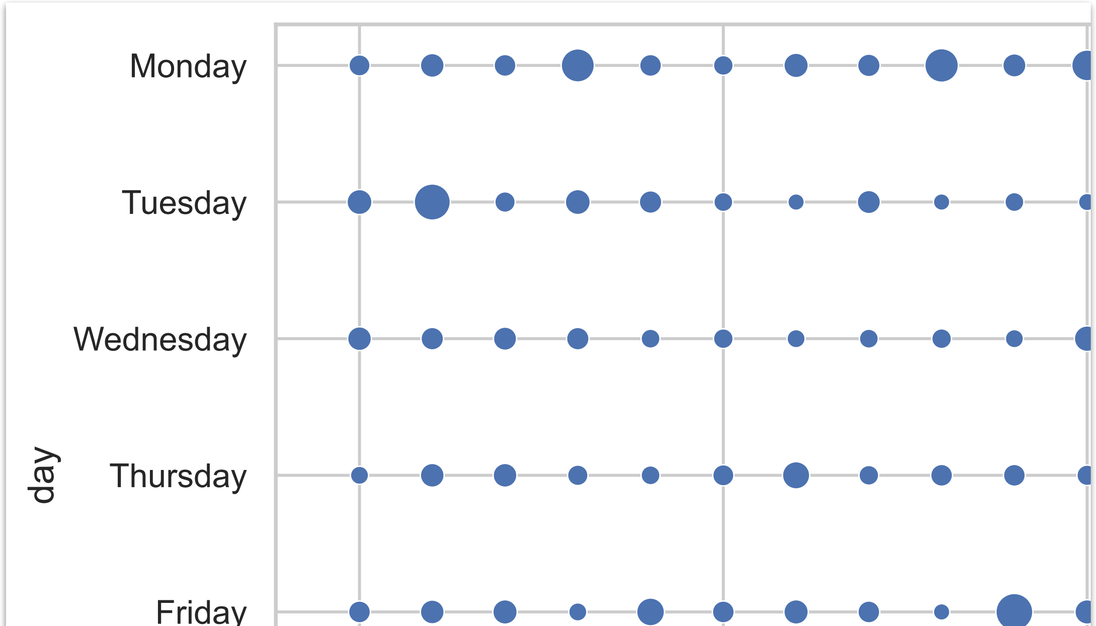
```
alt.Chart(github) \
    .mark_circle() \
    .encode(
            x = 'hours(time):O',
            y = 'day(time):O',
         size = 'sum(commits):Q'
    )
```

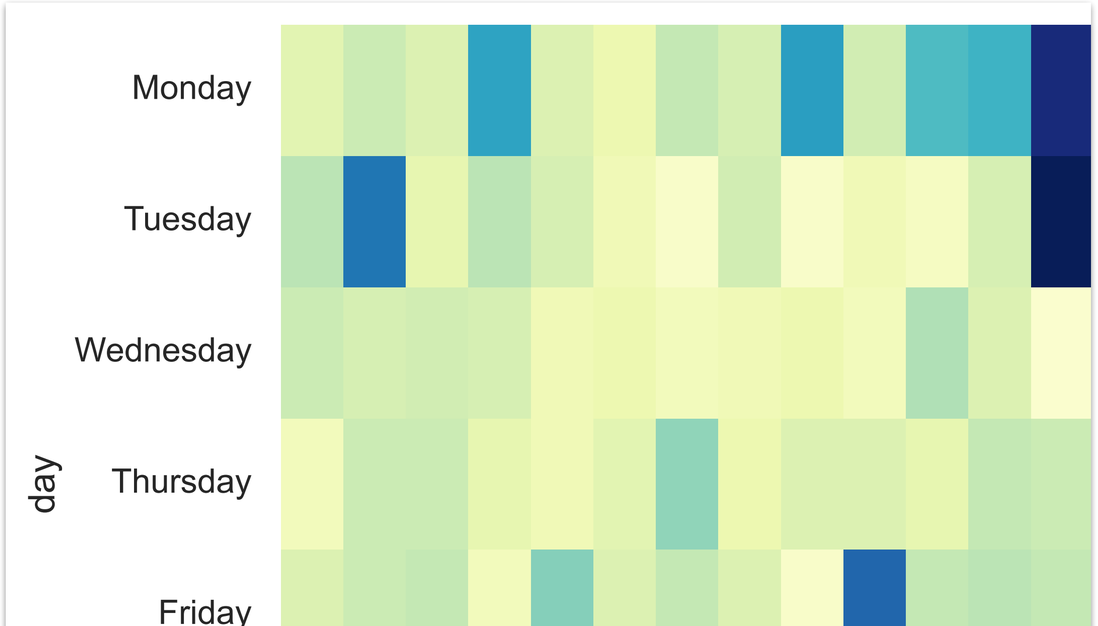


With Seaborn, to get a heatmap, change the plot type

```
df = github.groupby(
  ['day', 'hour'], as_index=False
).commits.sum()

sns.scatterplot(
  data=df, x='hour', y='day',
  size='commits')
```

```
df = github.groupby(
  ['day', 'hour']
).commits.sum().unstack()

sns.heatmap(df)
```

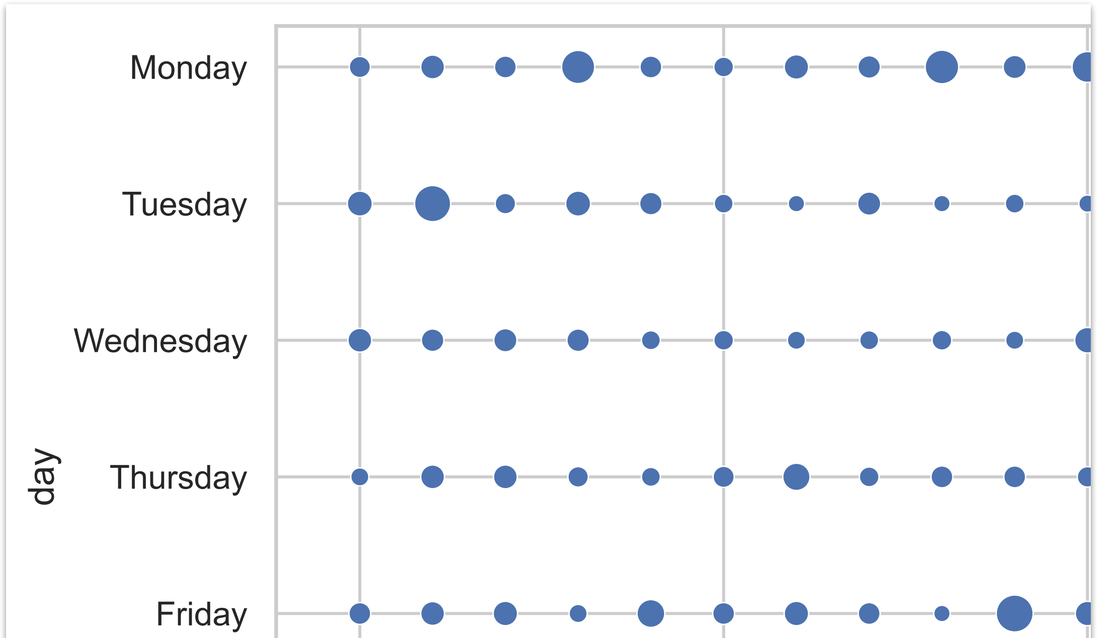


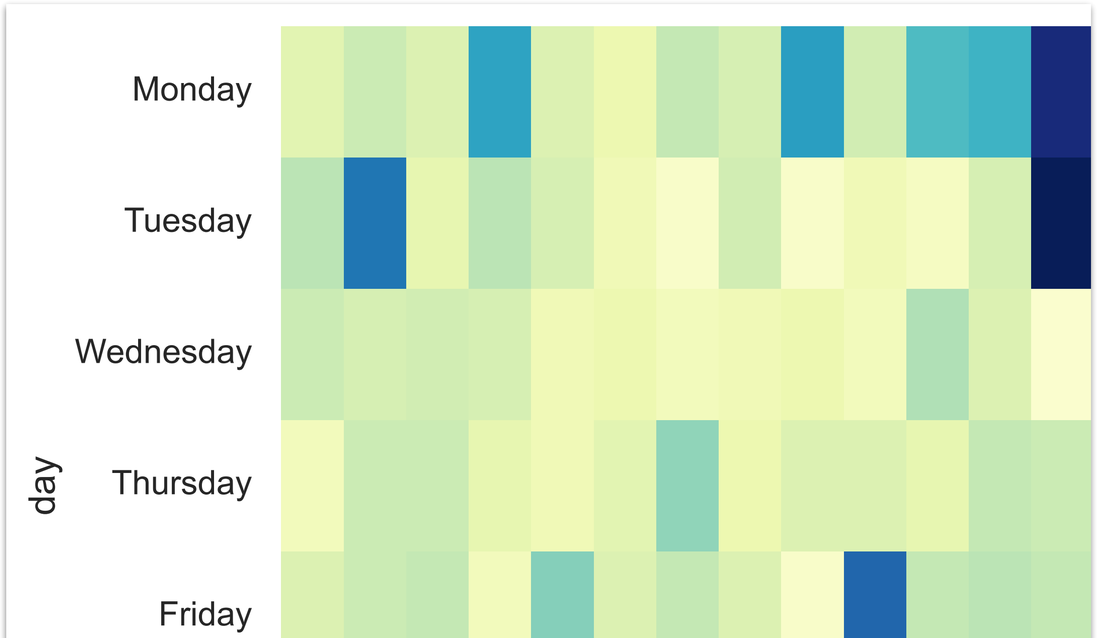


With Altair, to get a heatmap, change the encoding

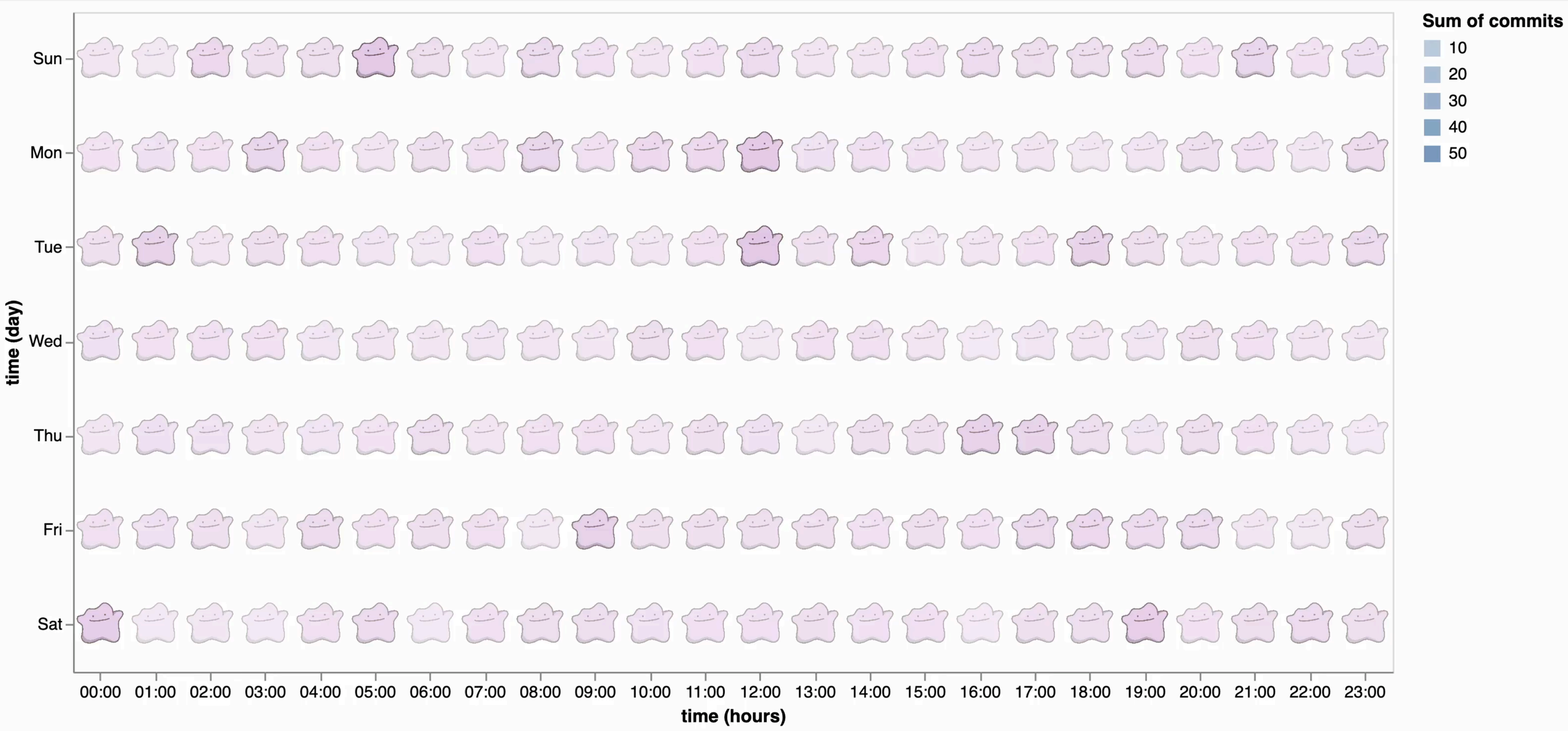
```
alt.Chart(github) \
    .mark_circle() \
    .encode(
            x = 'hours(time):O',
            y = 'day(time):O',
         size = 'sum(commits):Q'
    )
```

```
alt.Chart(github) \
    .mark_rect() \
    .encode(
            x = 'hours(time):O',
            y = 'day(time):O',
        color = 'sum(commits):Q'
    )
```



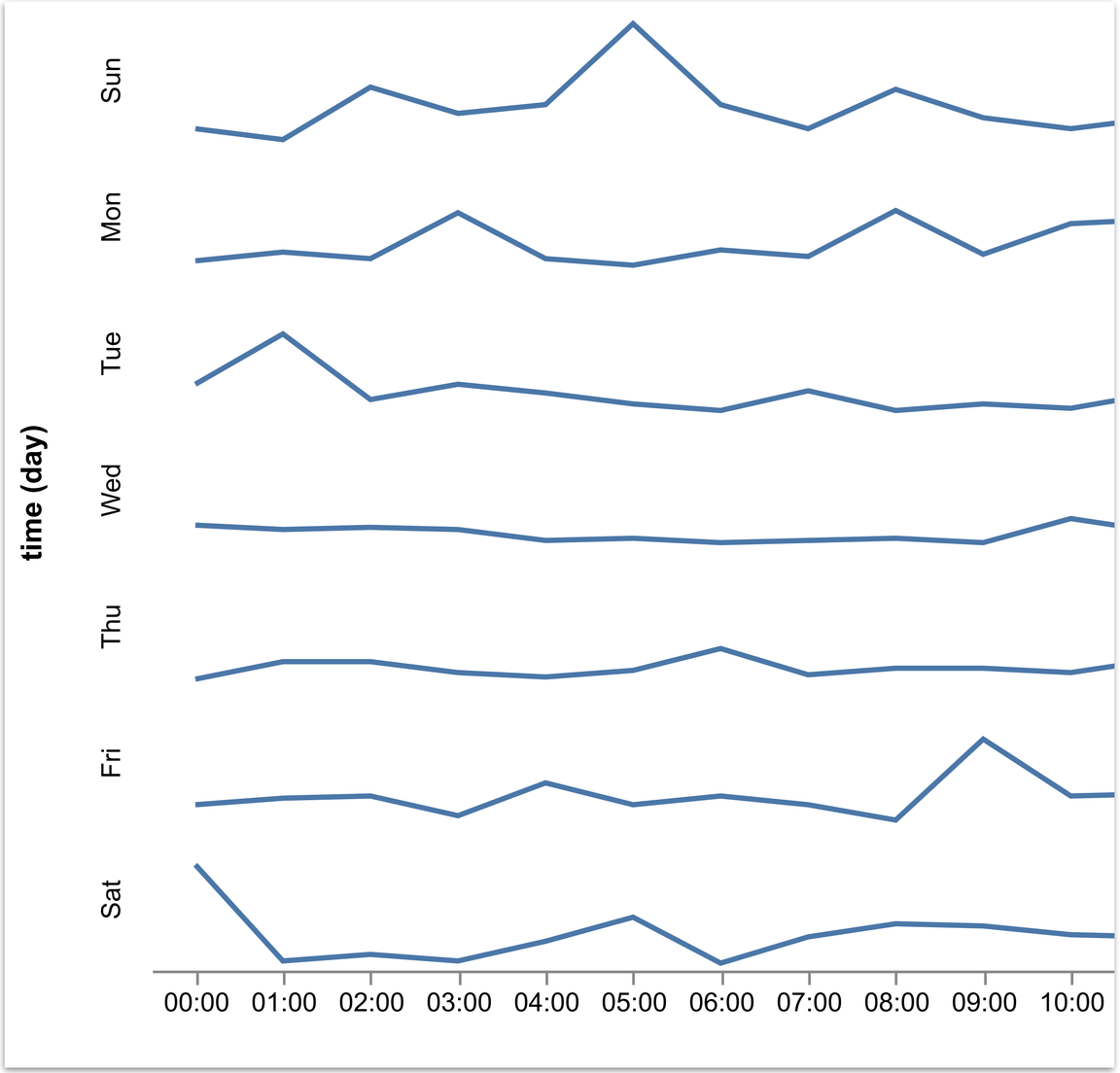


```
base.mark_image(url="ditto.gif").encode(opacity='sum(commits):Q')
```



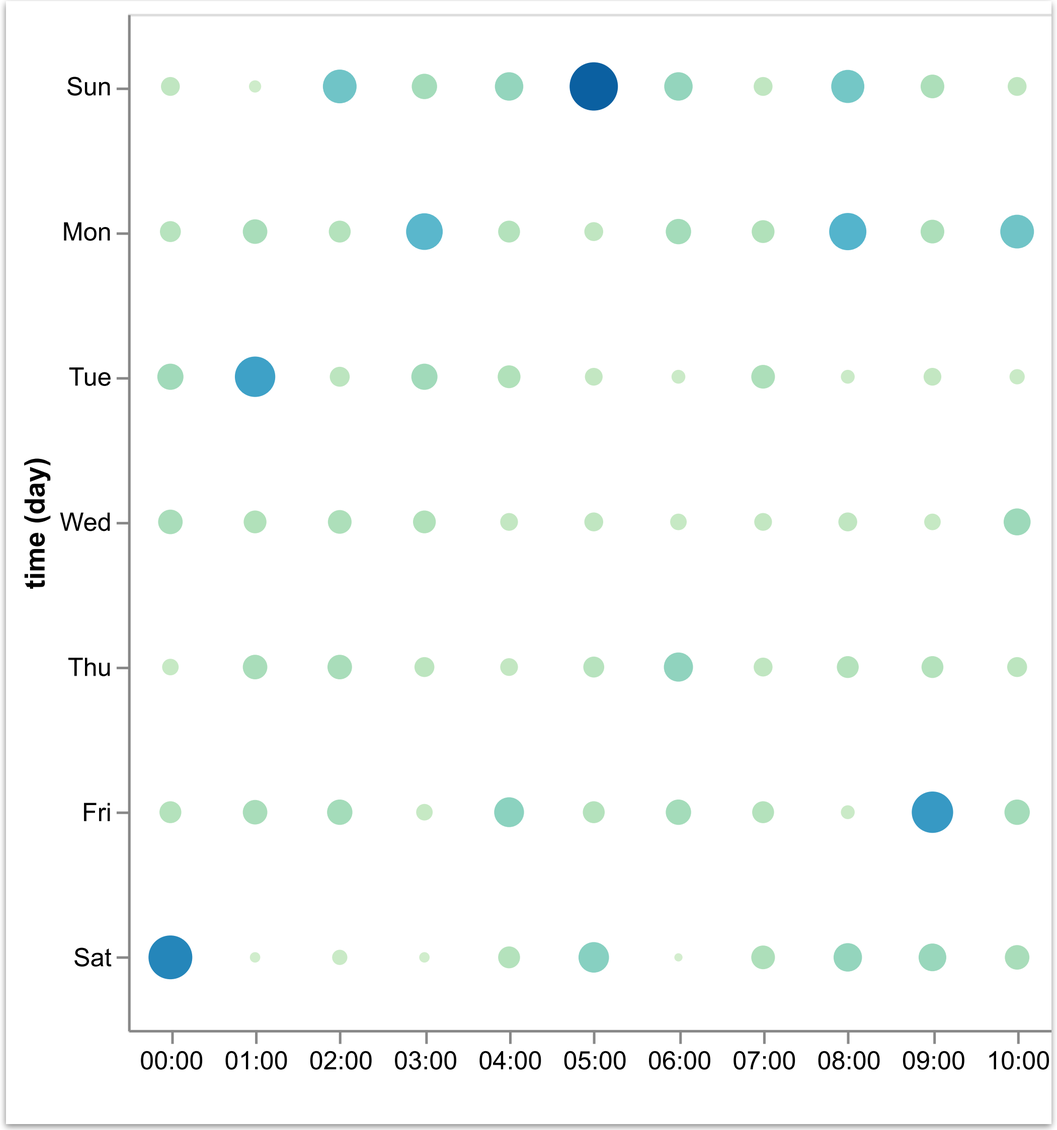
Faceting creates sub-plots

```
alt.Chart(github) \
    .mark_line() \
    .encode(
        x='hours(time):O',
        y='sum(commits):Q'
        row='day(time):O'
    )
```



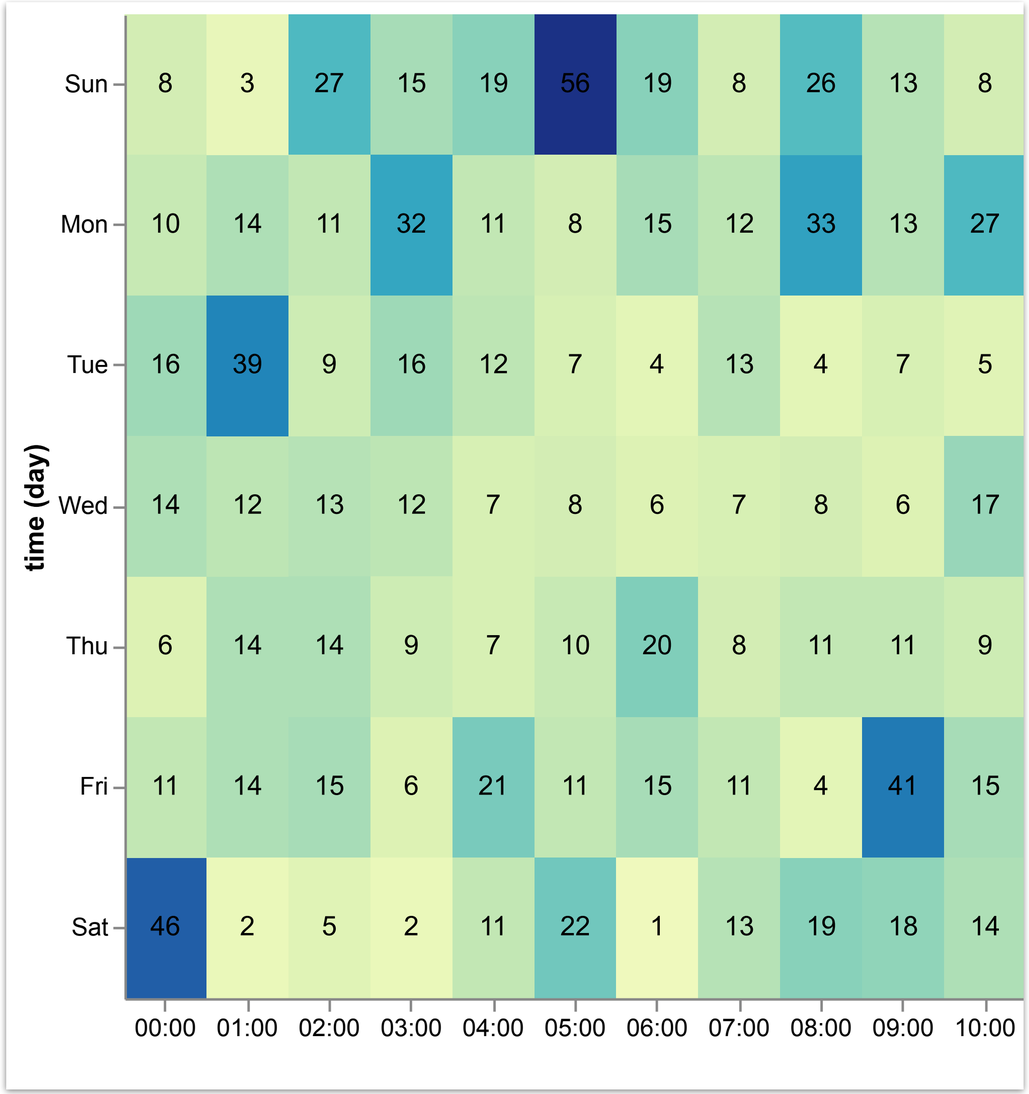
Add multiple encodings of the same mark

```
alt.Chart(github) \
    .mark_rect() \
    .encode(
            x = 'hours(time):O',
            y = 'day(time):O',
        color = 'sum(commits):Q',
         size = 'sum(commits):Q'
    )
```

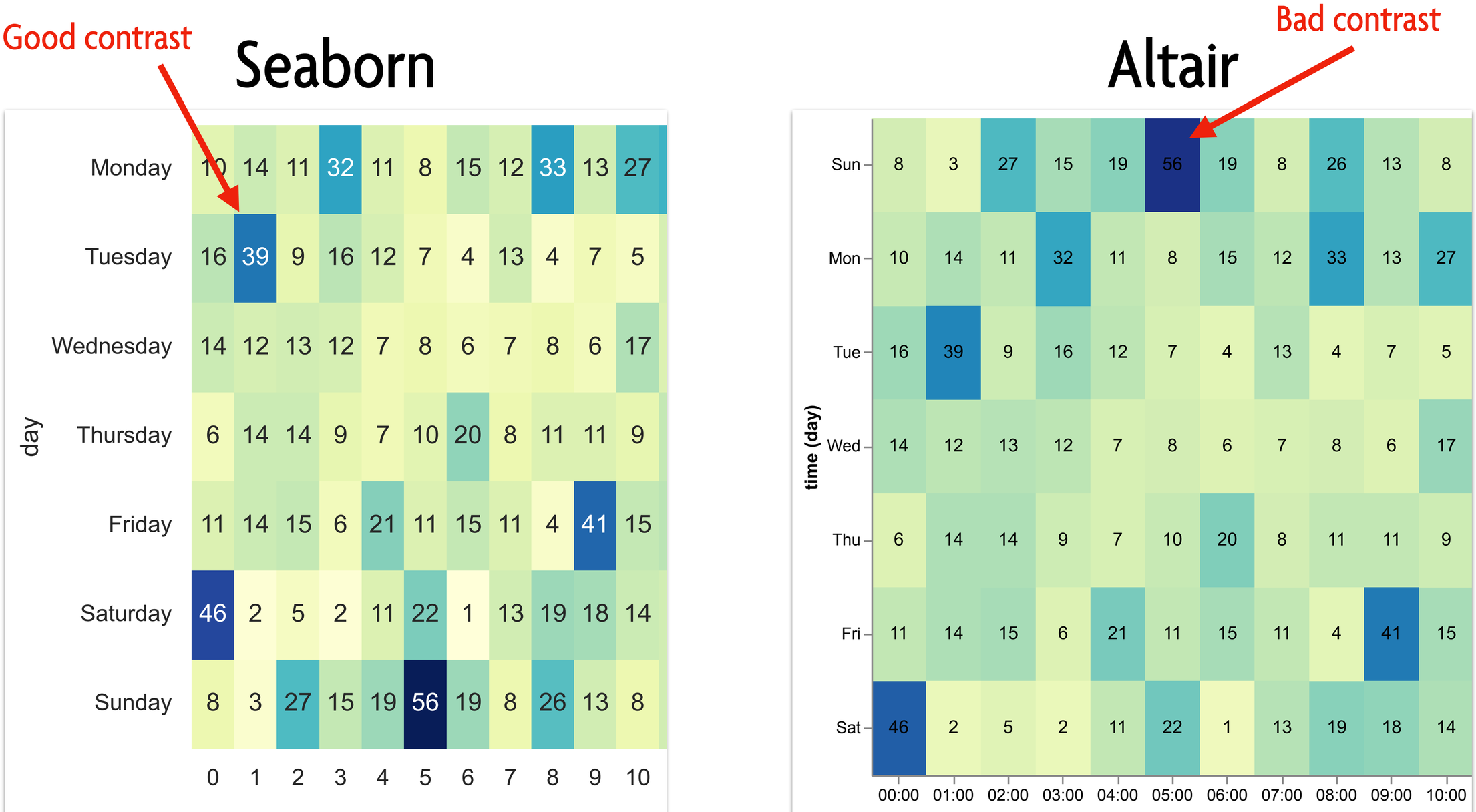


Layer multiple marks with redundant encodings

```
base = alt.Chart(github) \
 .encode(
    x='hours(time):O',
    y='day(time):O'
  )

base.mark_rect().encode(
    color='sum(commits):Q'
) + \
base.mark_text().encode(
    text='sum(commits):Q'
)
```

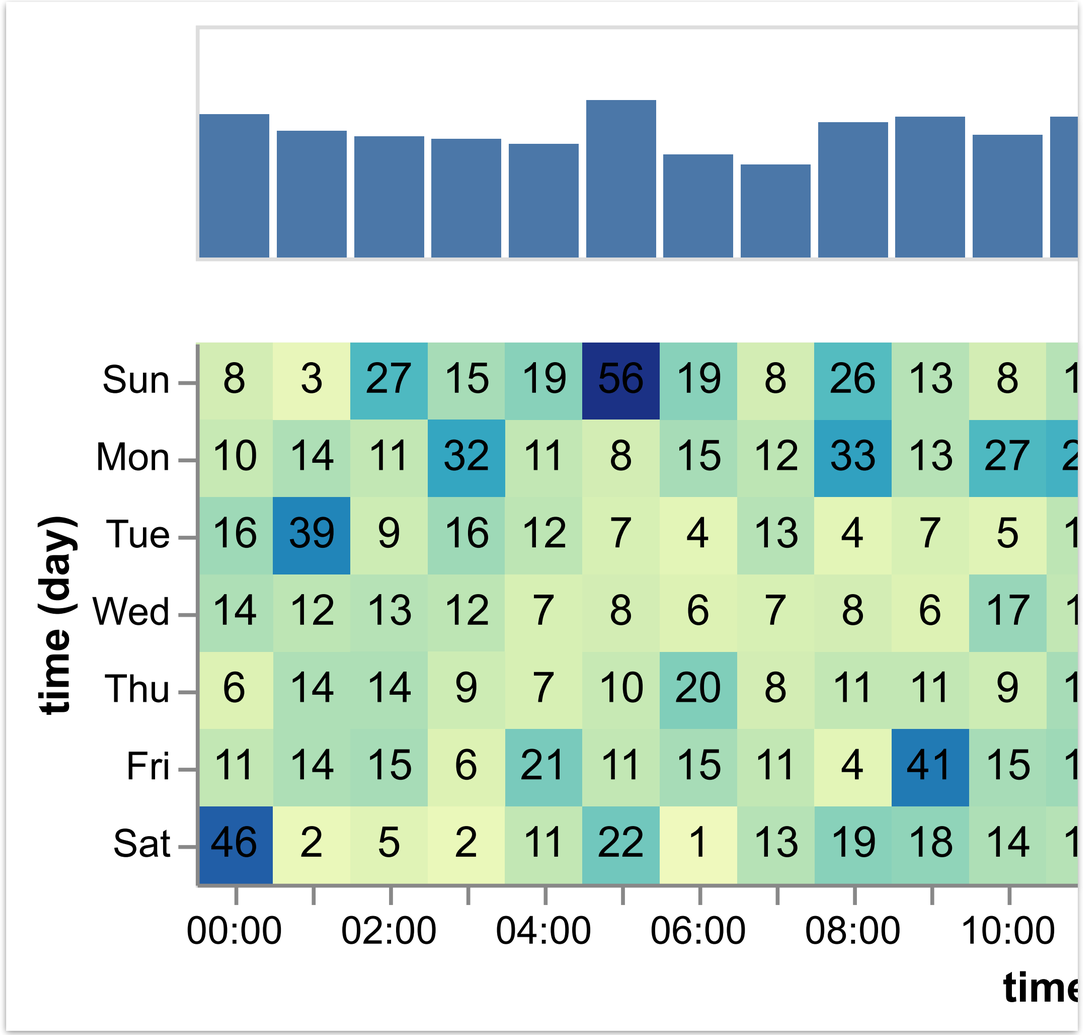


The price of compositionality



Combine multiple charts on a shared axis

```
chart = alt.Chart(github)

hist = chart.mark_bar().encode(
  x='hours(time):O',
  y='sum(commits):Q'
).properties(height=60)

main = chart.encode(
  x='hours(time):O',
  y='day(time):O')

heat = main.mark_rect().encode(
  color='sum(commits):Q') \
+ main.mark_text().encode(
  text='sum(commits):Q')

hist & heat
```

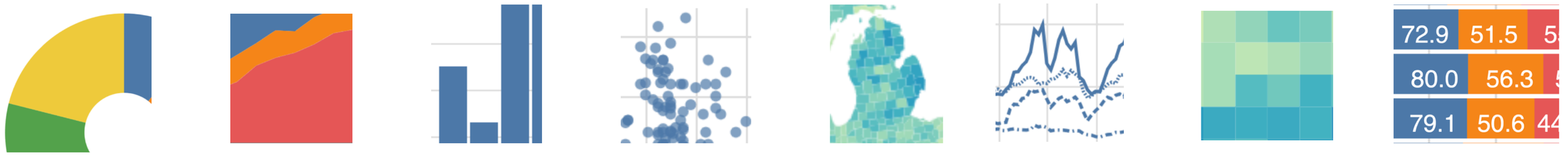


The grammar of graphics is a compositional approach to statistical chart design

A **chart** is composed of **marks** whose **attributes** are mapped from **data**

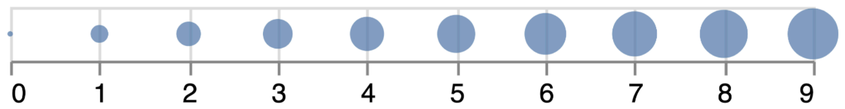
Marks

Arc Area Bar Circle Geo Line Square Text

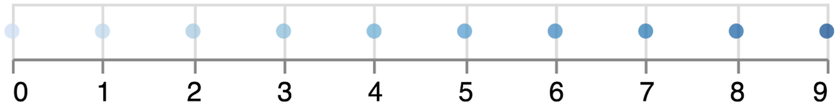


Attributes (Circle)

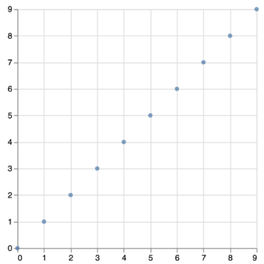
Size

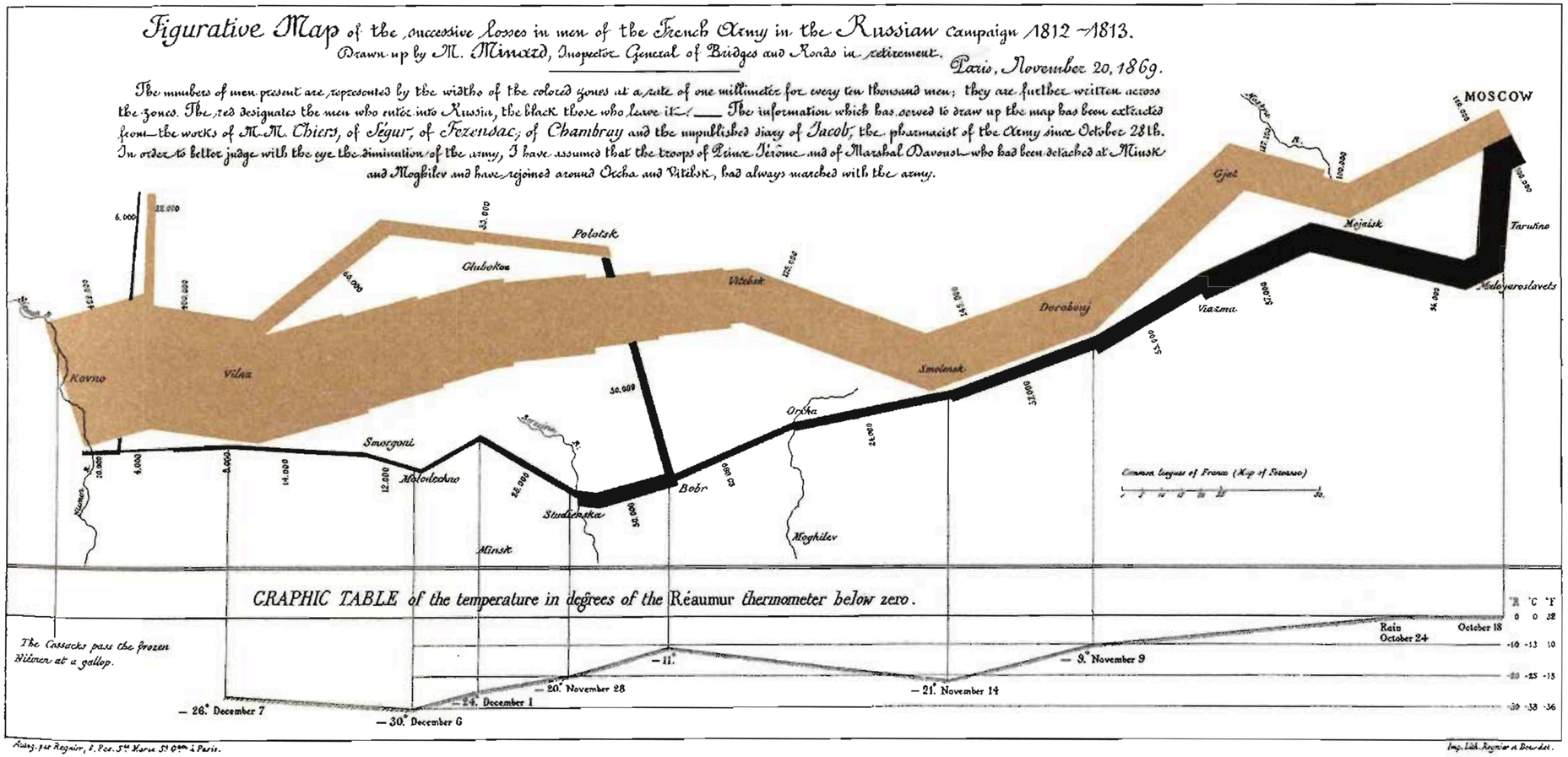


Color



XY-pos





Mark: area
X: longitude (Q)
Y: latitude (Q)
Width: army size (Q)
Color: direction (N)

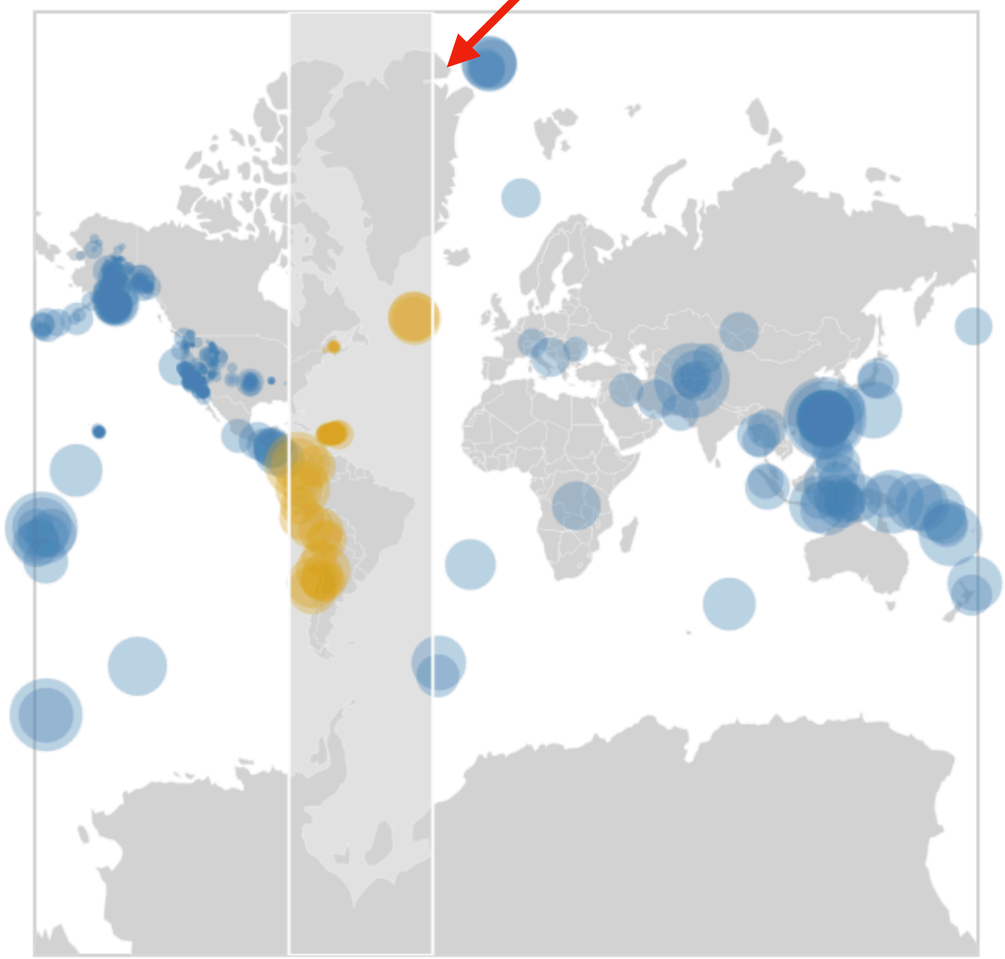
Mark: line
X: longitude (Q)
Y: temperature(Q)

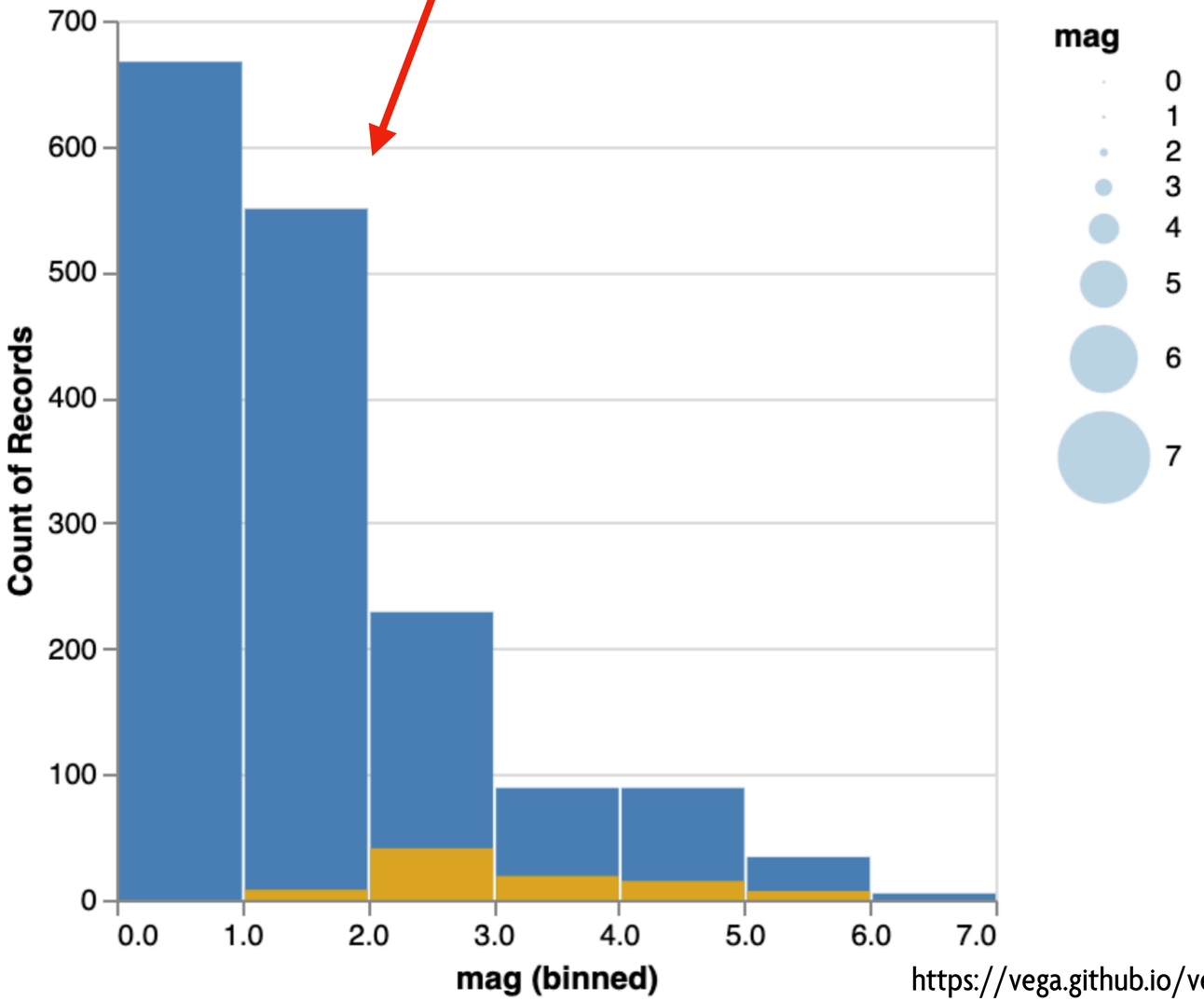
Charles Minard. *Des tableaux graphiques et des cartes figuratives.* 1861

Beyond the basics

Vega-Lite provides a grammar of interactive graphics

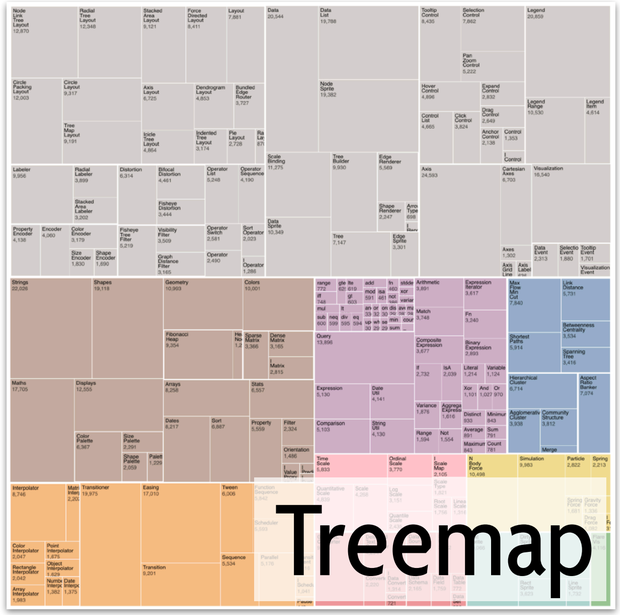
selection controls data for histogram

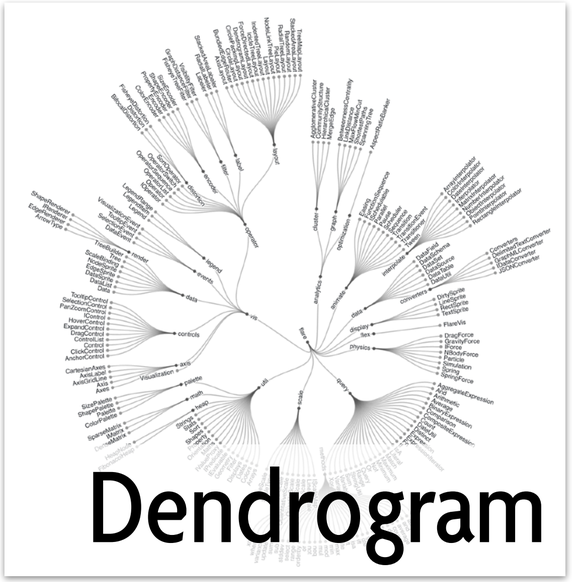


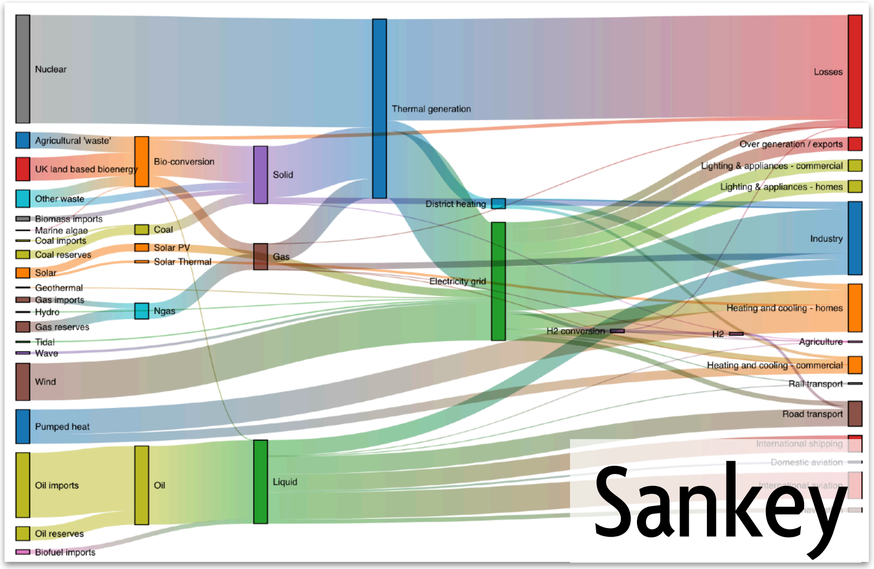


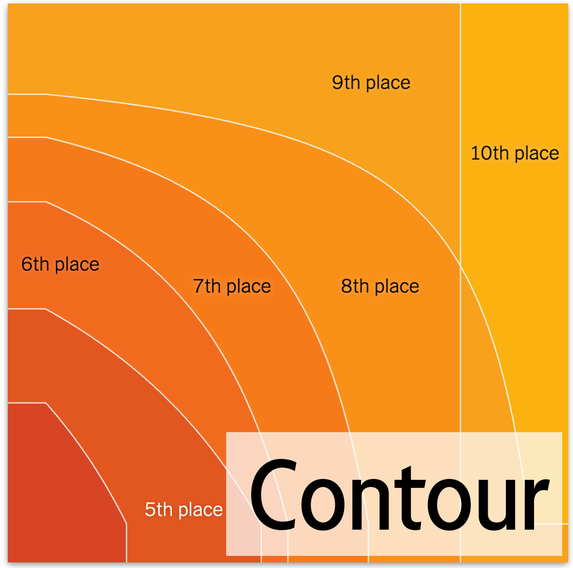
https://vega.github.io/vega-lite/

D3 provides a more expressive 2D graphics API





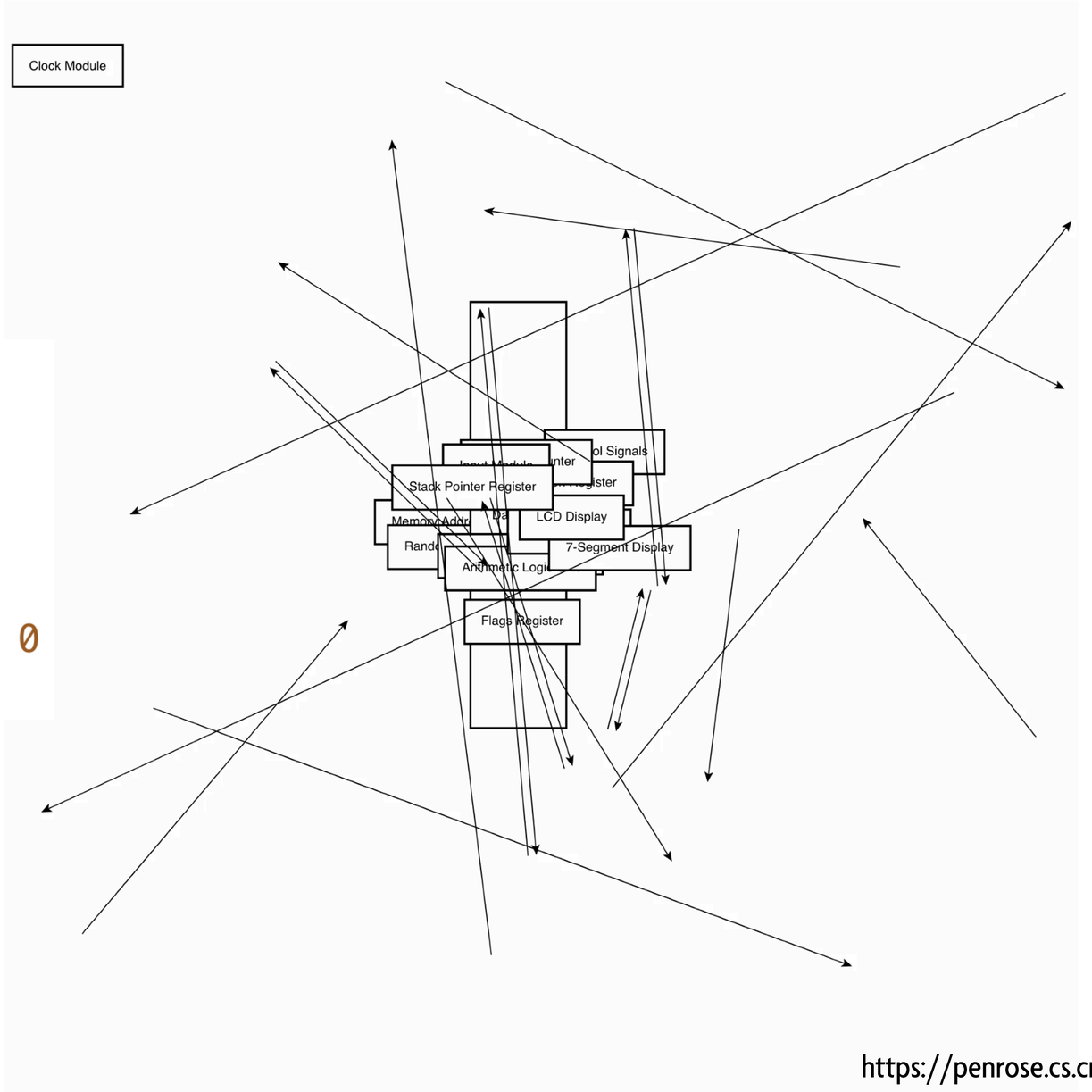




https://d3js.org/

Penrose generates diagrams using force-directed layout

```
ensure signedDistance(a.icon, s.start) == 5
  in arrow
ensure signedDistance(b.icon, s.end) == 5
  in arrow
strength = 100
encourage strength * norm(s.end - s.start) == 0
  in arrow
```



https://penrose.cs.cmu.edu/

Bluefish generates diagrams using local propagation

```
Distribute({ direction: "horizontal", spacing: -r },
  Ref({ select: "A" }), Ref({ select: "B" })),
Distribute({ direction: "horizontal", spacing: 0 },
  Ref({ select: "B" }), Ref({ select: "C" })),
Distribute({ direction: "vertical", spacing: 40 },
  Ref({ select: "rect" }), Ref({ select: "B" })),
Distribute({ direction: "vertical", spacing: 30 },
  Ref({ select: "B" }), Ref({ select: "A" })),
Distribute({ direction: "vertical", spacing: 50 },
  Ref({ select: "B" }), Ref({ select: "C" })),
```