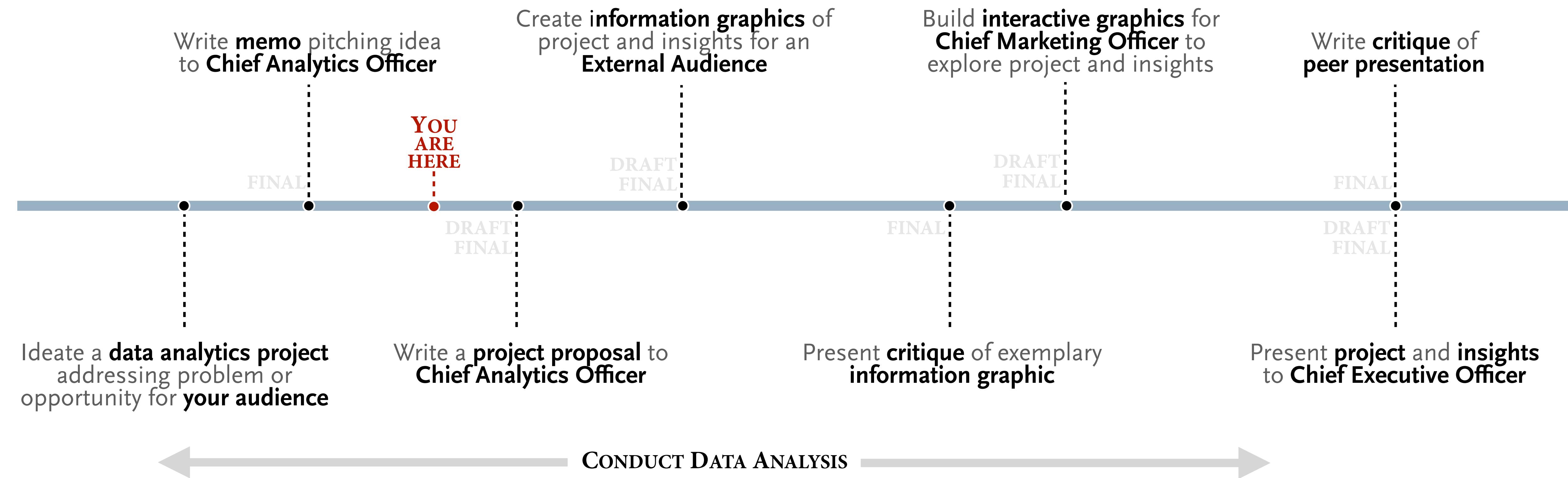


# Storytelling with data

**05 | visual design, data encodings, perceptual psychology**

# course overview | main course deliverables



**Why show data graphically?**

# why data graphics, graphic of a datum — effective? Conveys meaning?

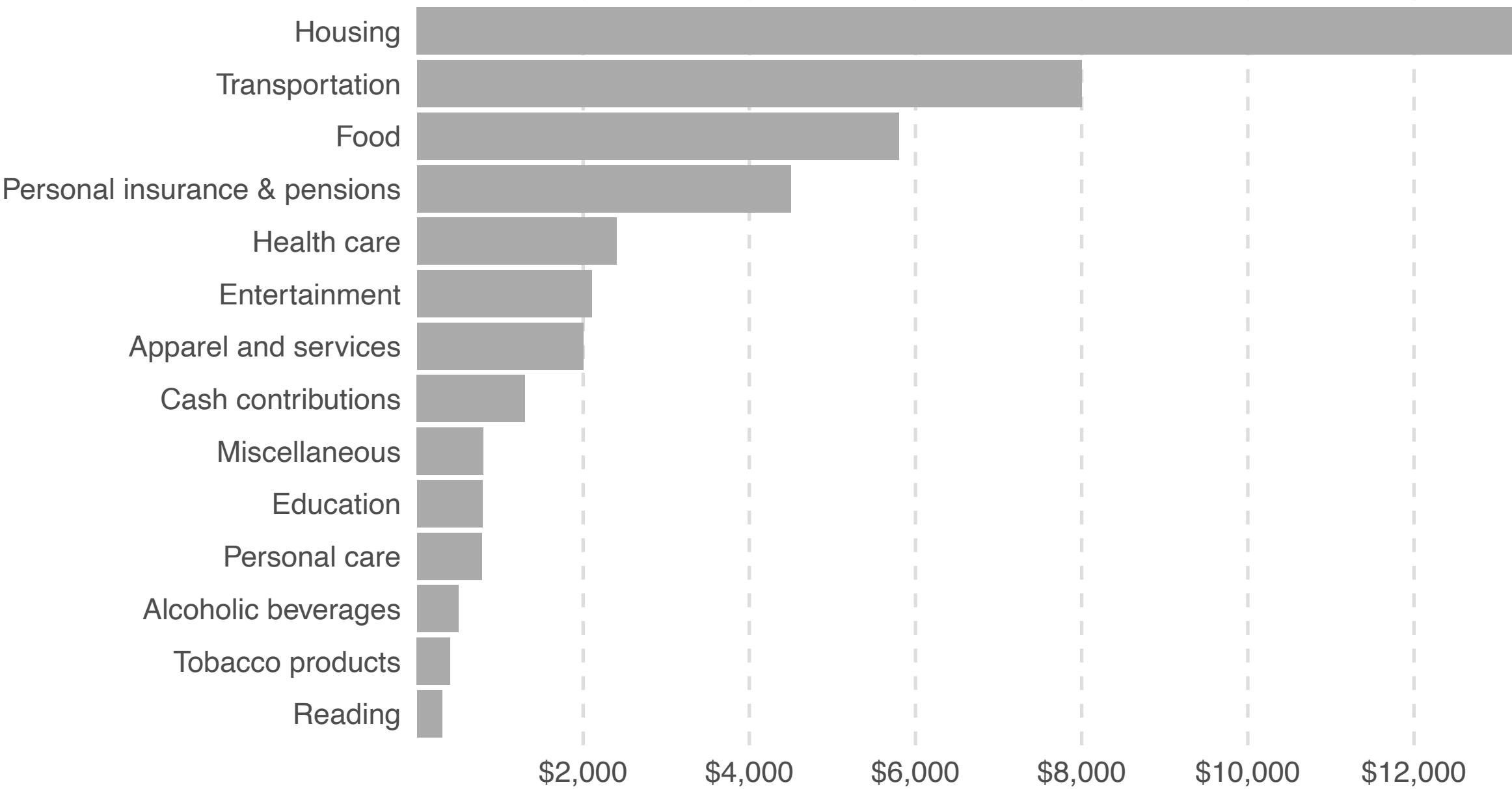


Fig. 3. Major categories of expenditures, descending dollar value, 2002 U.S. Consumer Expenditure Survey

While text can use different types of content structures, an abstract visualization just presents relationships between data points.

Thus, a single bar, map symbol or shape does not convey information. It only becomes meaningful by its relationship with other elements in the image—in other words, it is *polysemic*: **A data graphic acquires its meaning from comparison.**

— Koponen & Hildén, *The Data Visualization Handbook*

# why data graphics, graphic of a datum — effective? Conveys meaning?

Housing

While text can use different types of content structures, an abstract visualization just presents relationships between data points.

Thus, a single bar, map symbol or shape does not convey information. It only becomes meaningful by its relationship with other elements in the image—in other words, it is *polysemic*: **A data graphic acquires its meaning from comparison.**

— Koponen & Hildén, *The Data Visualization Handbook*

# why data graphics, example data from Anscombe

1		2		3		4	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

## why data graphics, example data from Anscombe

1		2		3		4	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

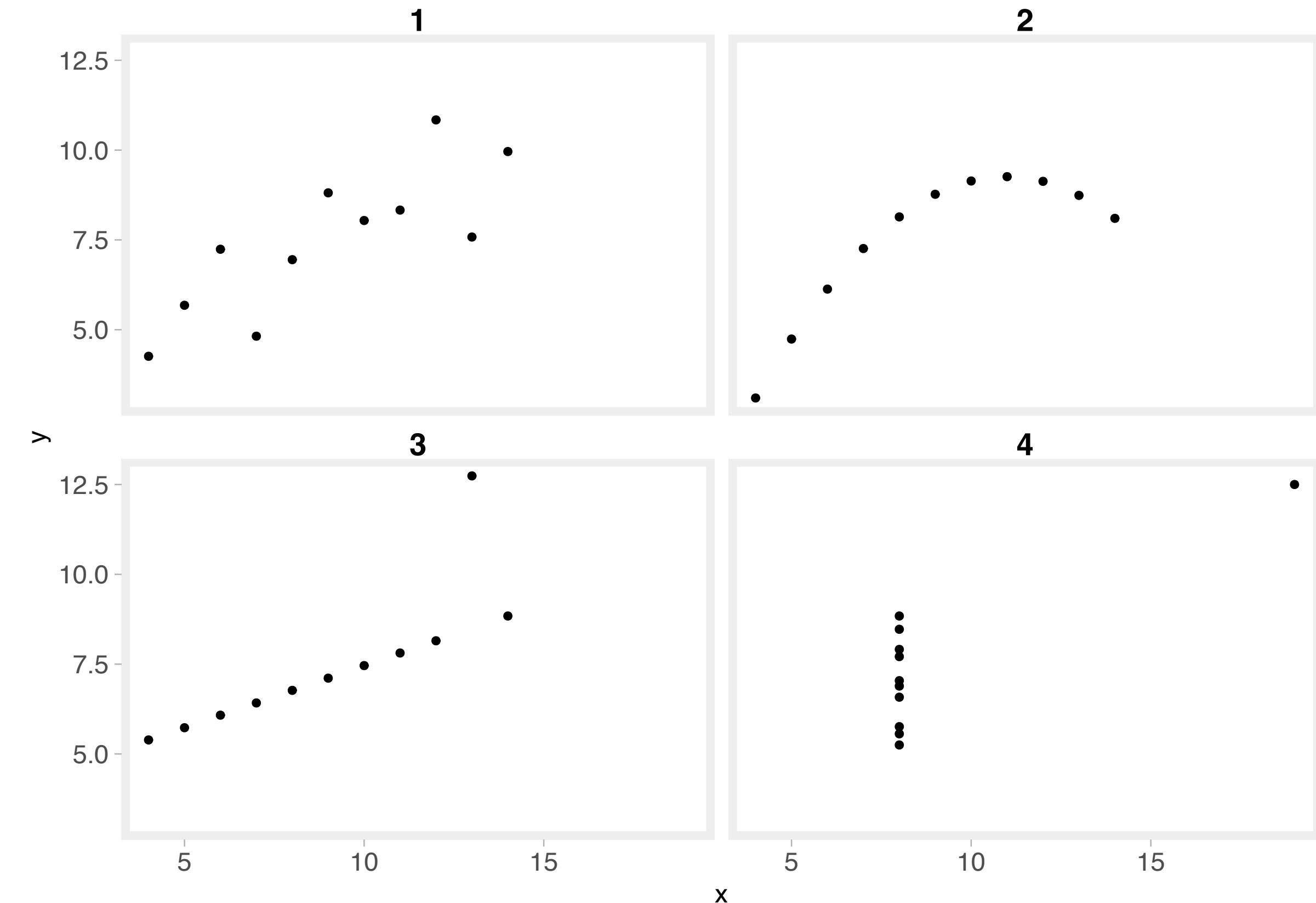
## summary statistics: *are the 4 data sets the same?*

1		2		3		4		
x	y	x	y	x	y	x	y	
mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
sd	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
Parameter		Mean		Std Err		t-val		
<b>Dataset 1</b>								
(Intercept)		3.000		1.125		2.667		
x		0.500		0.118		4.241		
<b>Dataset 2</b>								
(Intercept)		3.001		1.125		2.667		
x		0.500		0.118		4.239		
<b>Dataset 3</b>								
(Intercept)		3.002		1.124		2.670		
x		0.500		0.118		4.239		
<b>Dataset 4</b>								
(Intercept)		3.002		1.124		2.671		
x		0.500		0.118		4.243		

## why data graphics, example data from Anscombe

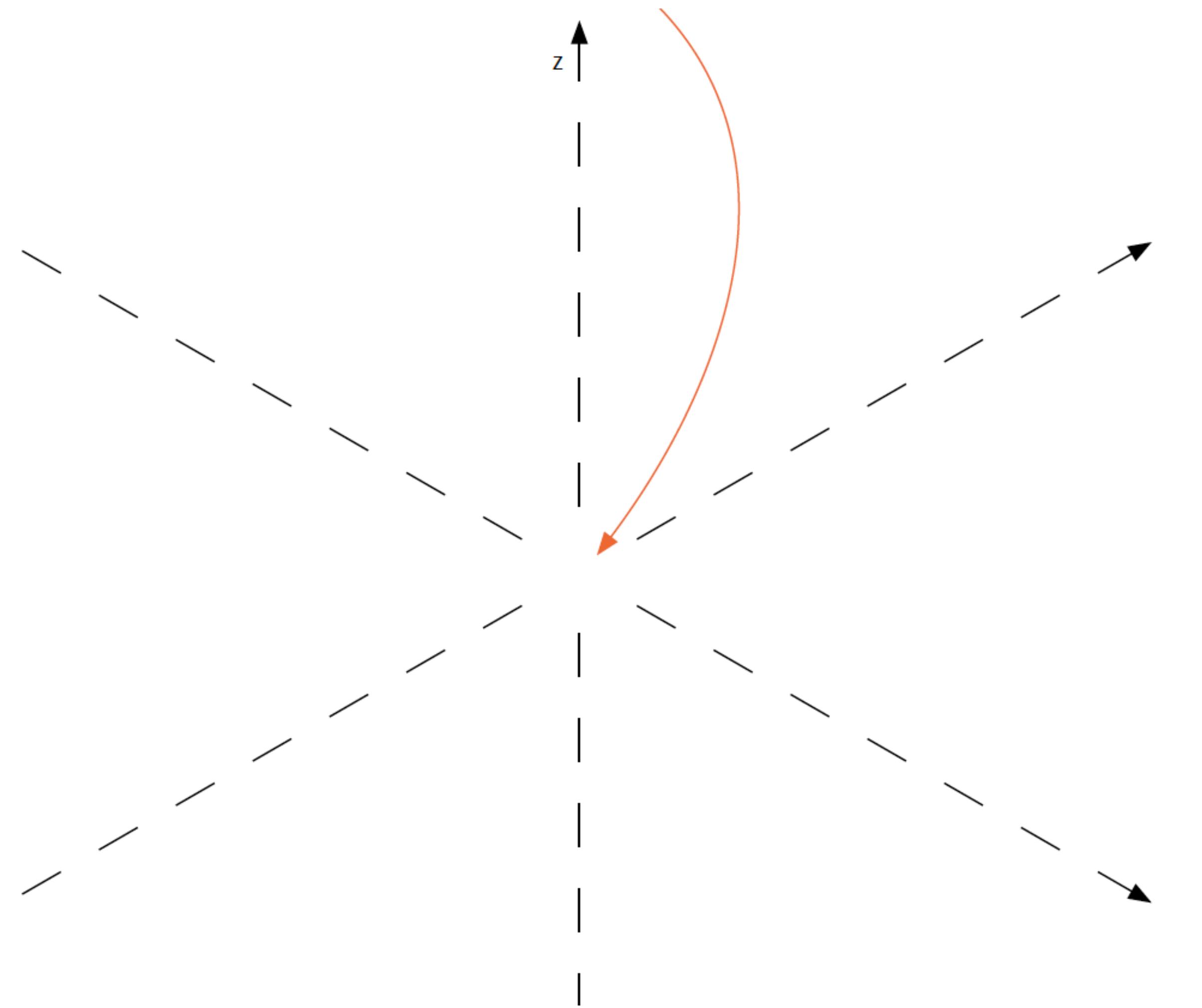
With graphics we can use our natural ability  
to see patterns through visual comparison

1		2		3		4	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

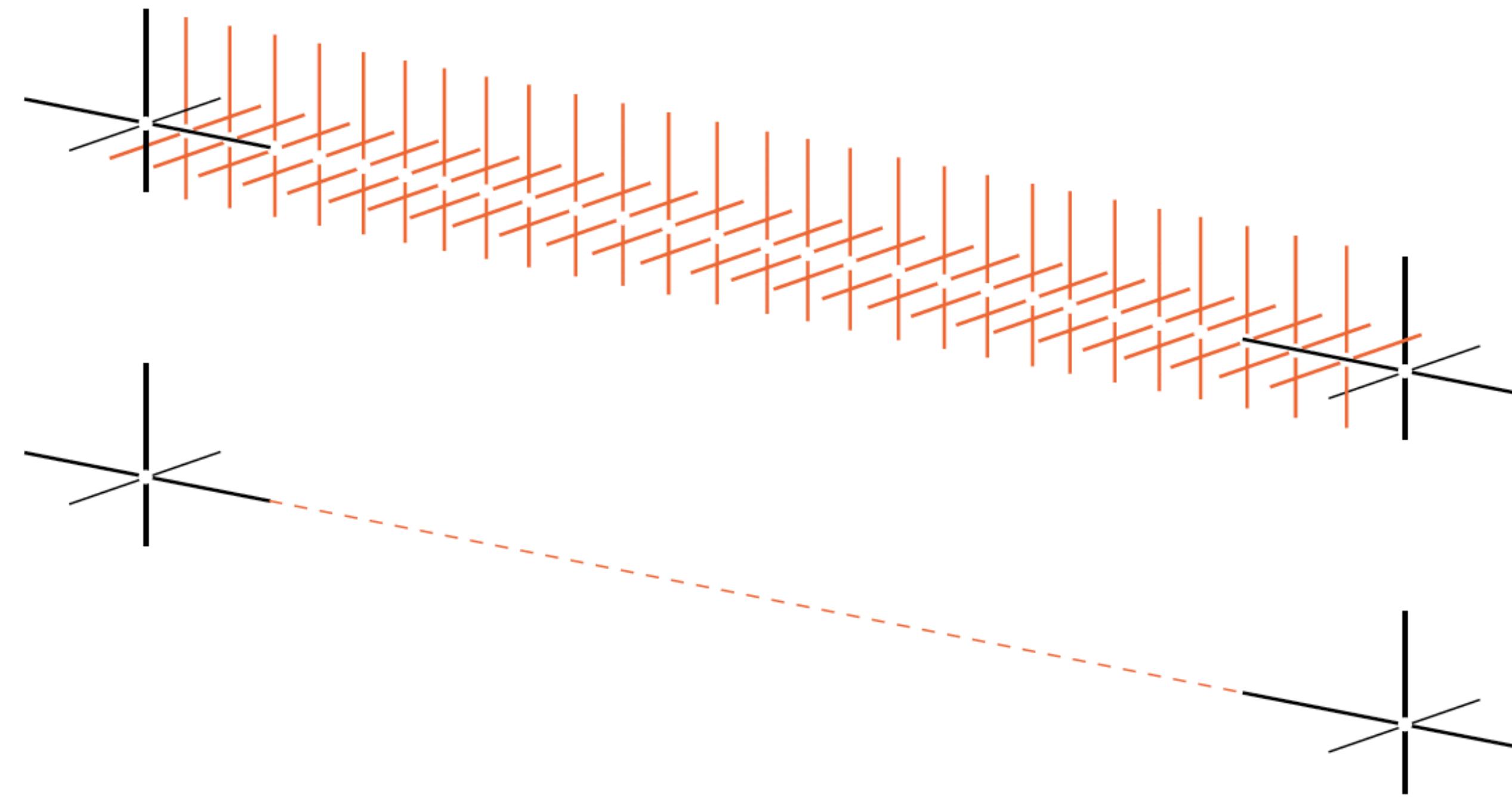


**graphic design concepts**

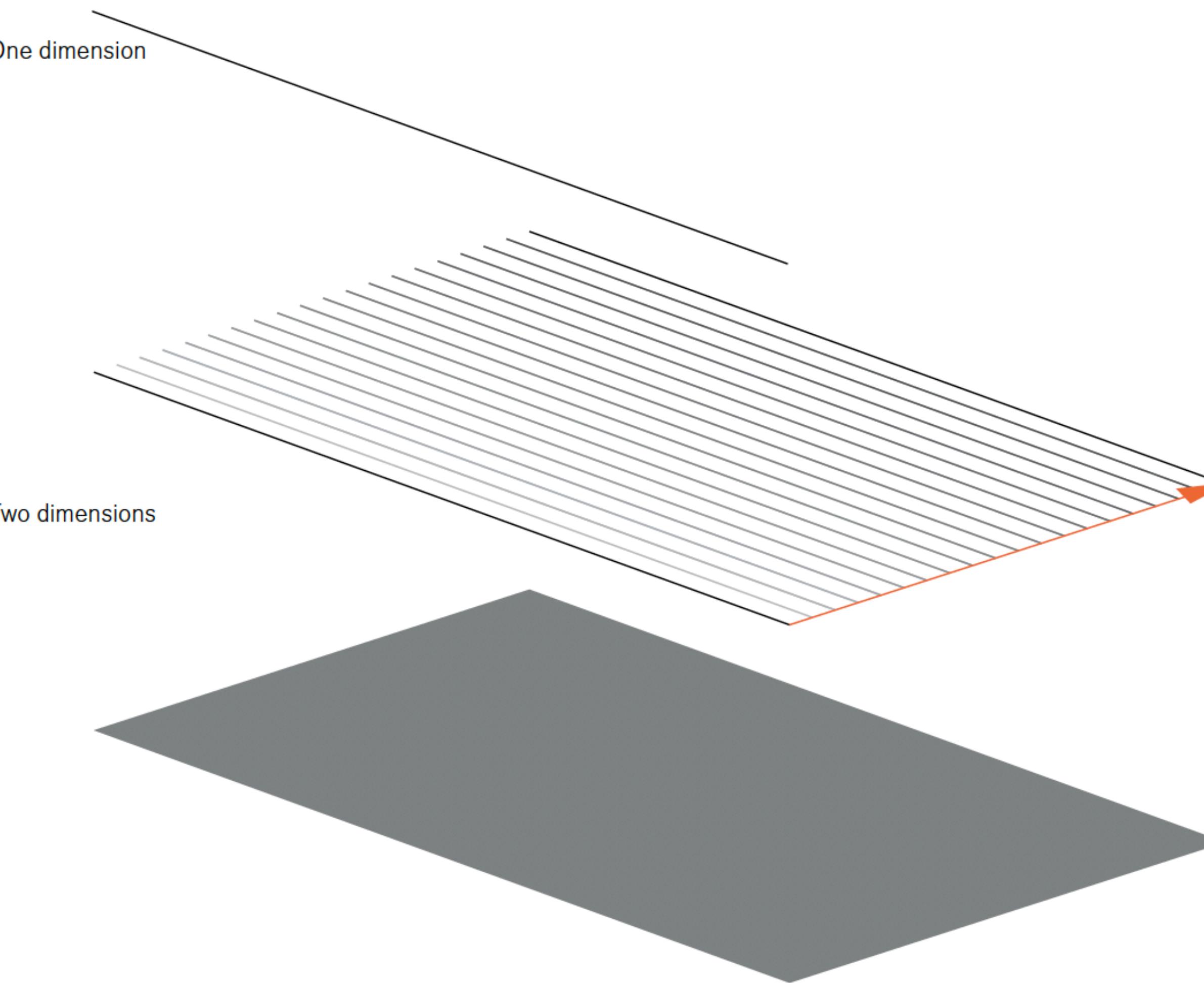
## data encodings, geometry of graphical elements — point



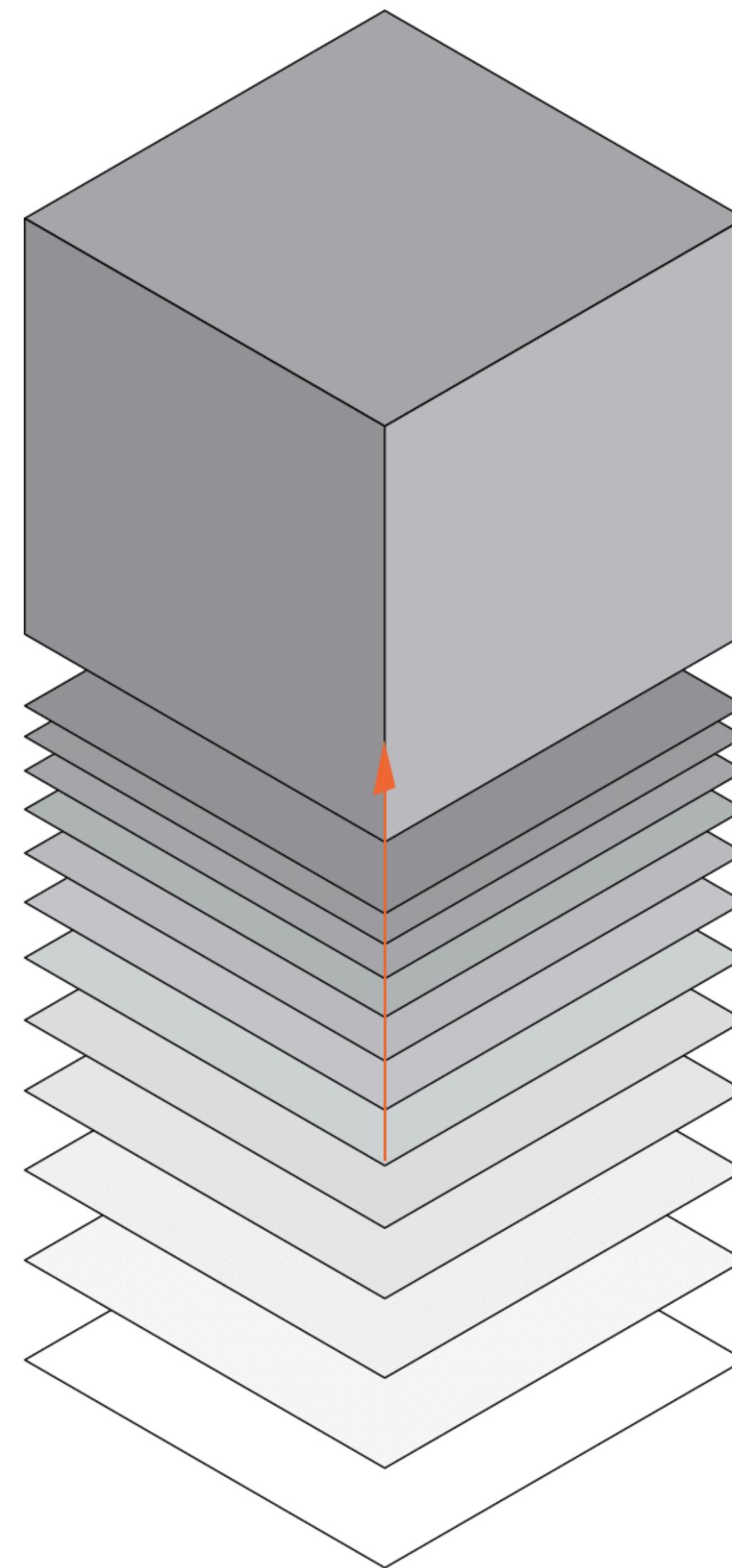
## data encodings, geometry of graphical elements — line



## data encodings, geometry of graphical elements — surface

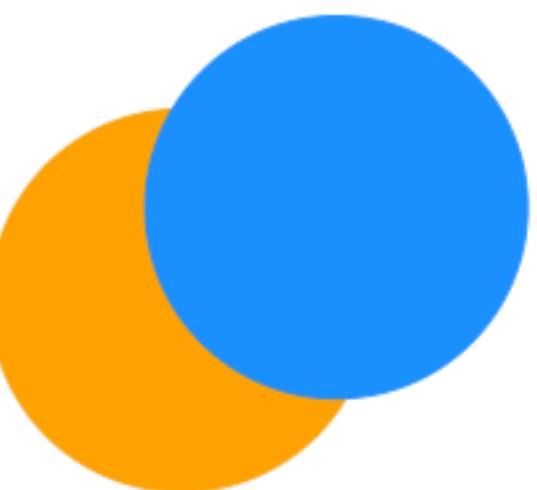


# data encodings, geometry of graphical elements — volume

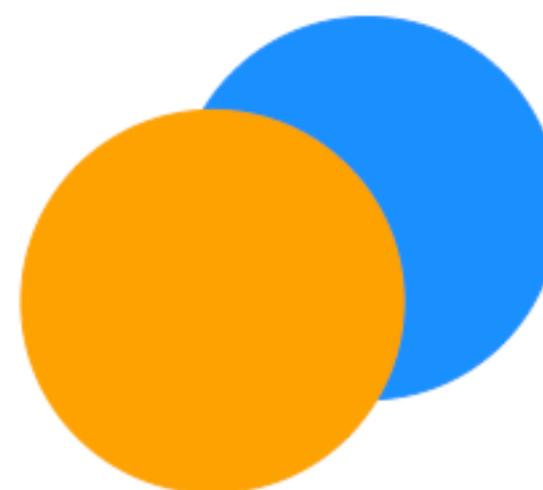


# data encodings, layering — order of elements determines position towards reader and when overlapping, occlude

```
ggplot() +  
  theme_void() +  
  scale_x_continuous(limits = c(-5, 5)) +  
  scale_y_continuous(limits = c(-5, 5)) +  
  geom_point(  
    mapping = aes(  
      x = 0,  
      y = 0),  
    size = 50,  
    color = "orange") +  
  geom_point(  
    mapping = aes(  
      x = 1,  
      y = 1),  
    size = 50,  
    color = "dodgerblue")
```

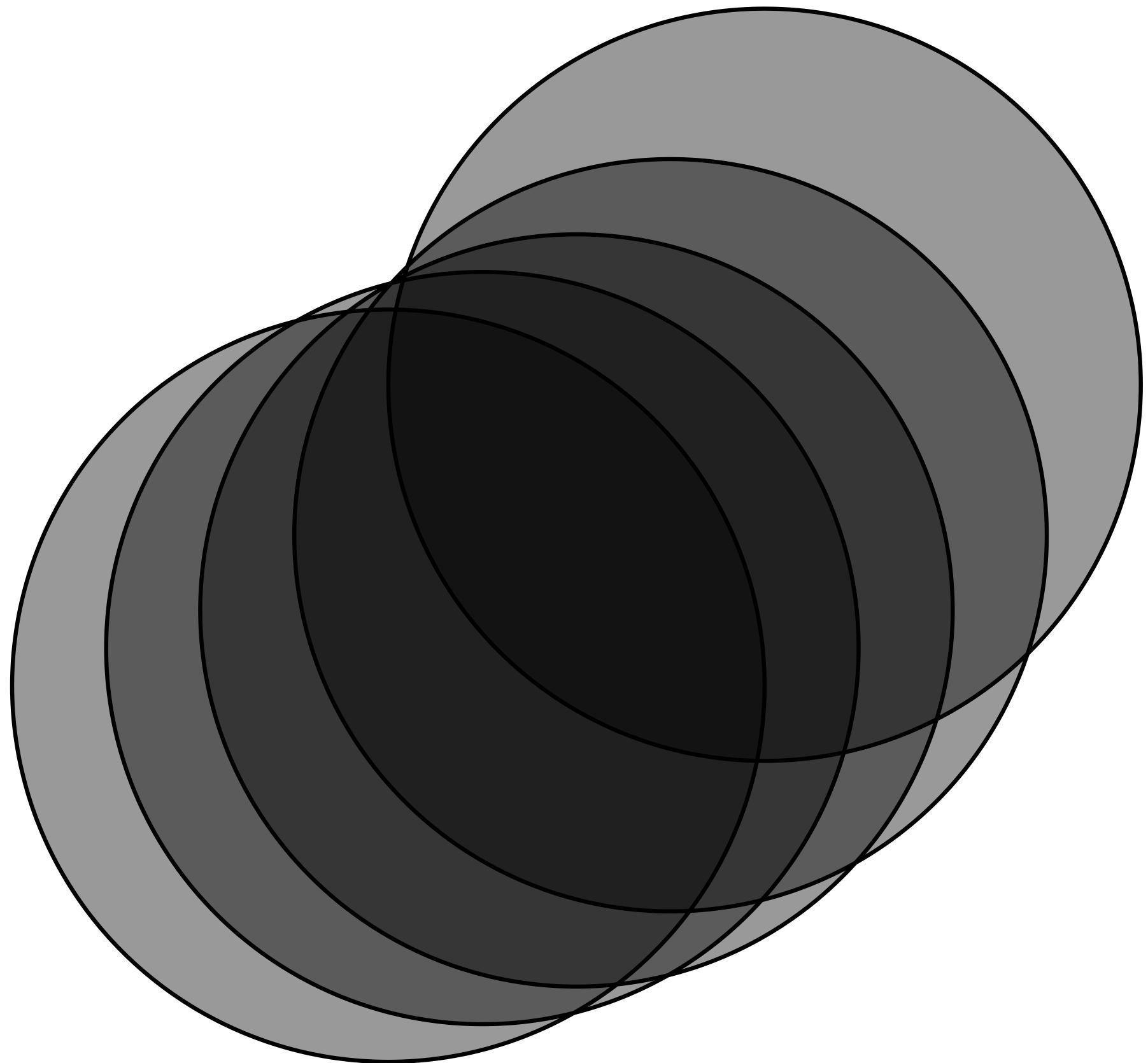


```
ggplot() +  
  theme_void() +  
  scale_x_continuous(limits = c(-5, 5)) +  
  scale_y_continuous(limits = c(-5, 5)) +  
  geom_point(  
    mapping = aes(  
      x = 1,  
      y = 1),  
    size = 50,  
    color = "dodgerblue") +  
  geom_point(  
    mapping = aes(  
      x = 0,  
      y = 0),  
    size = 50,  
    color = "orange")
```



data encodings, layering — transparency (alpha) of monochromes can help us reason about the density of overlapping shapes

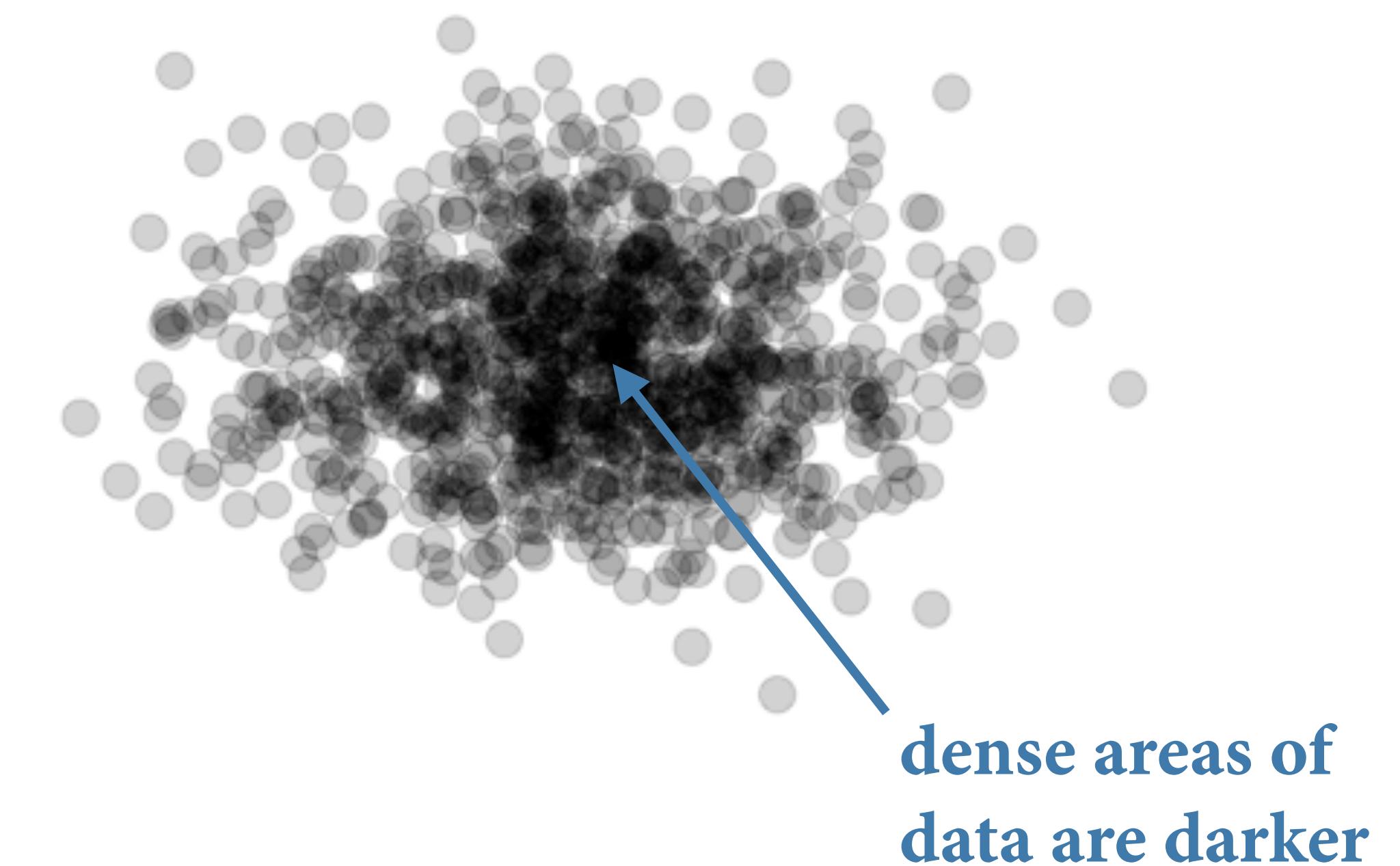
```
ggplot() +  
  theme_void() +  
  coord_equal() +  
  ggforce::geom_circle(  
    mapping = aes(  
      x0 = seq(from = 0, to = 1, length.out = 5),  
      y0 = c(0, .1, .2, .4, .8),  
      r = 1),  
    fill = "#000000",  
    alpha = 0.4)
```



data encodings, layering — transparency (alpha) of monochromes can help us reason about the density of overlapping shapes

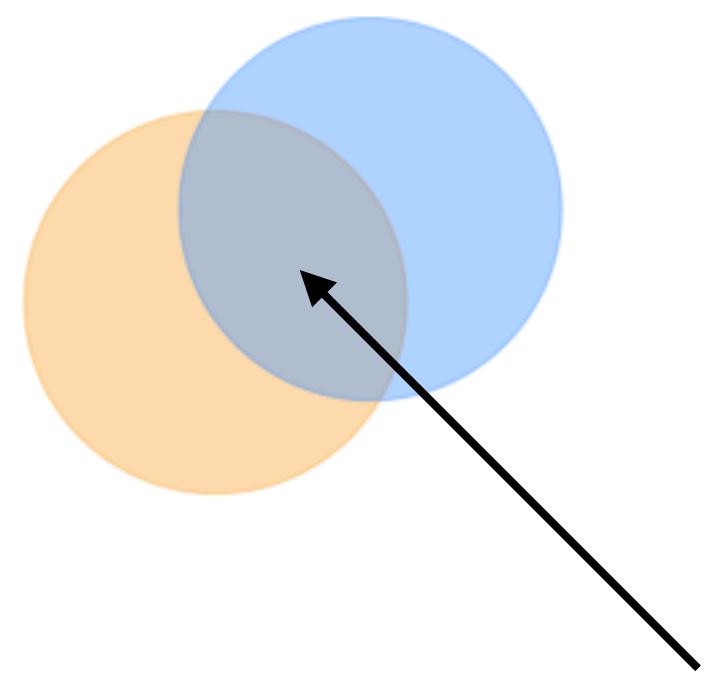
```
x <- rnorm(1000)
y <- rnorm(1000)

ggplot() +
  theme_void() +
  scale_x_continuous(limits = c(-5, 5)) +
  scale_y_continuous(limits = c(-5, 5)) +
  geom_point(
    mapping = aes(
      x = x,
      y = y),
    size = 4,
    color = "black",
    alpha = 0.2)
```



# data encodings, layering — data encoded in *semi-transparent hues*, if overlapping, are affected by transparency!

```
ggplot() +  
  theme_void() +  
  scale_x_continuous(limits = c(-5, 5)) +  
  scale_y_continuous(limits = c(-5, 5)) +  
  geom_point(  
    mapping = aes(  
      x = 0,  
      y = 0),  
    size = 50,  
    color = "orange",  
    alpha = 0.4) +  
  geom_point(  
    mapping = aes(  
      x = 1,  
      y = 1),  
    size = 50,  
    color = "dodgerblue",  
    alpha = 0.4)
```

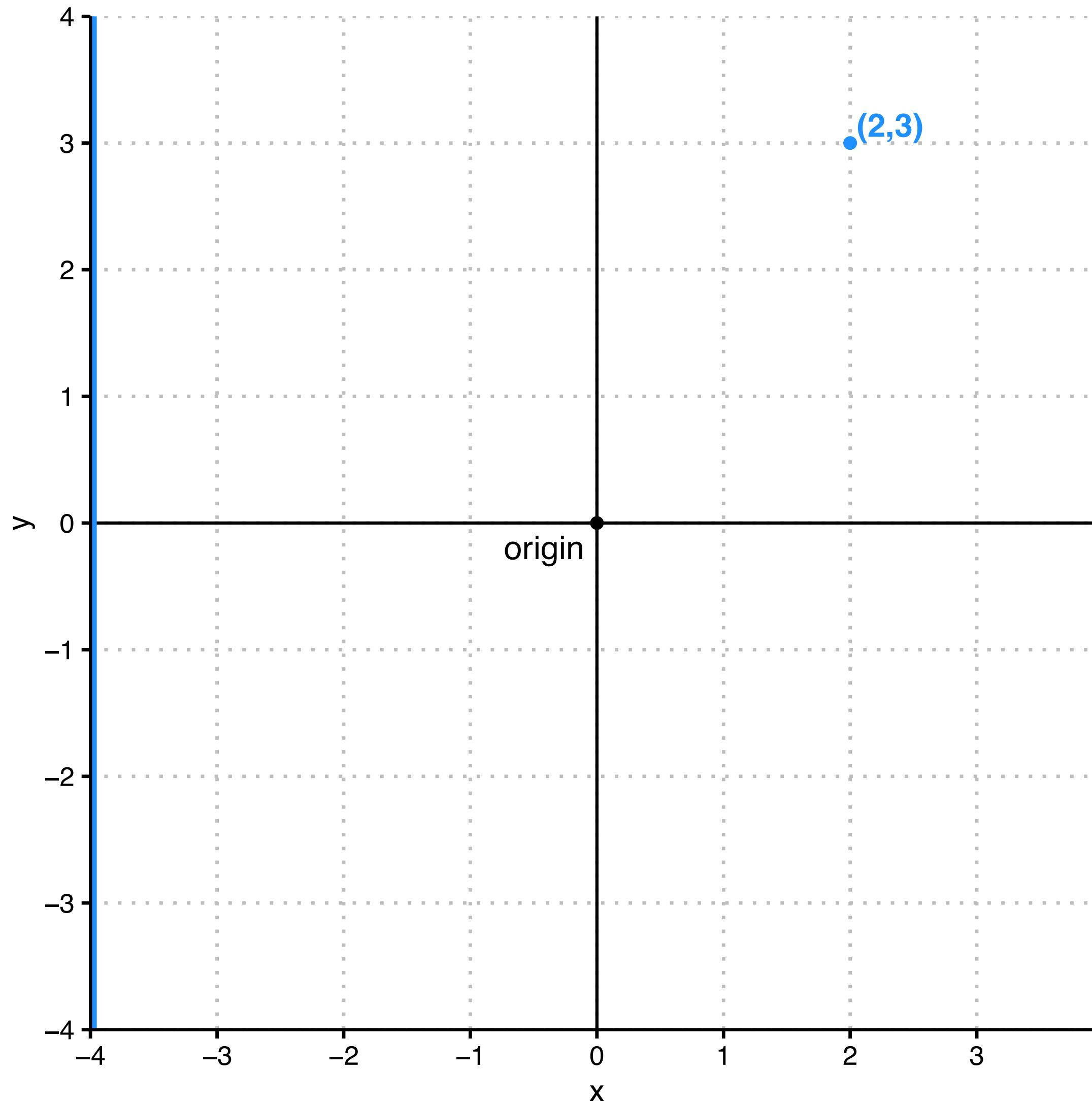


I didn't encode  
data with *this color*!?

## **graphs — coordinate systems and scales**

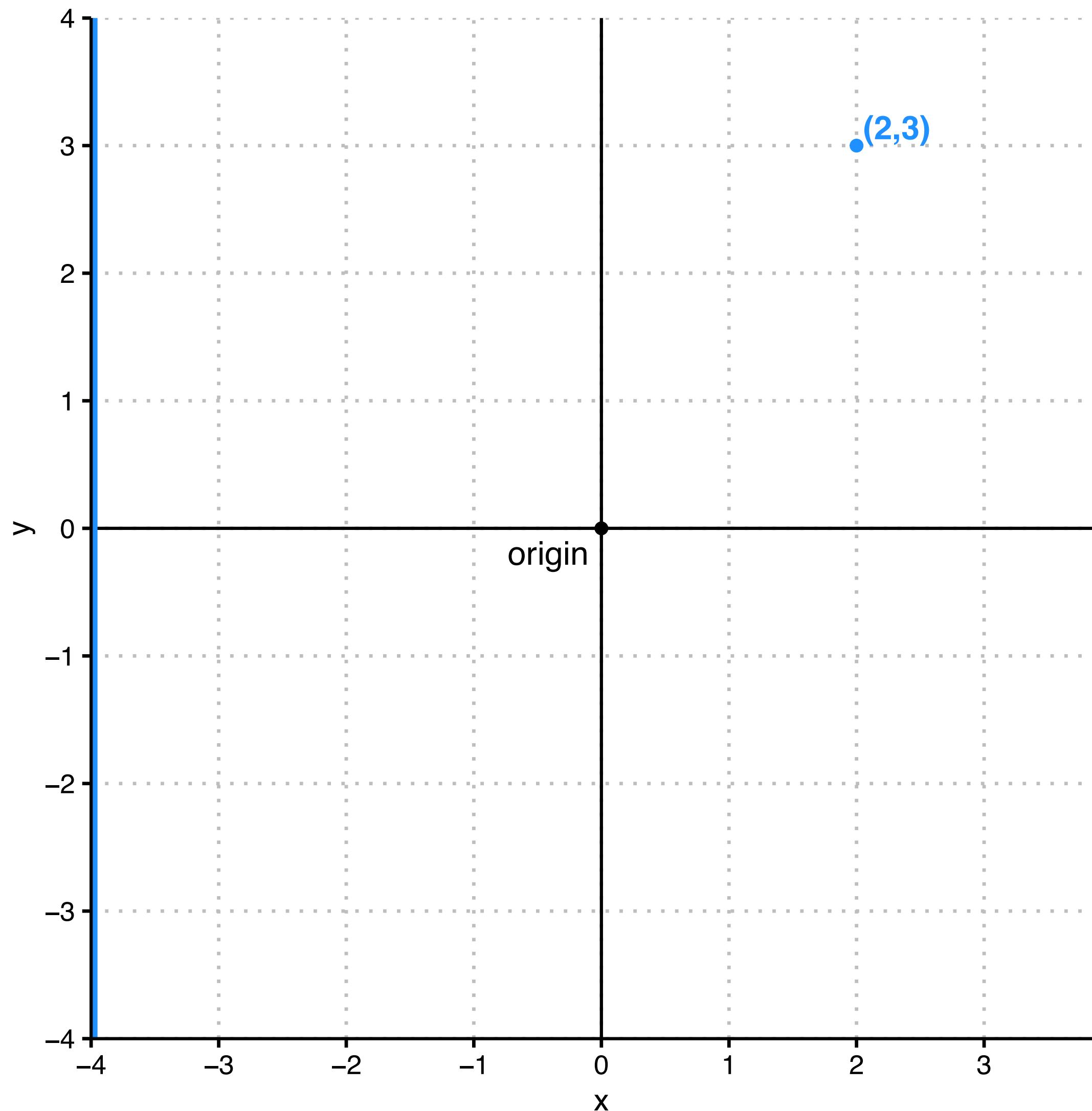
coordinates and scales, *two-dimensional Cartesian* coordinates — x and y axes run orthogonally to each other, and data values placed along linear axes

## cartesian coordinates

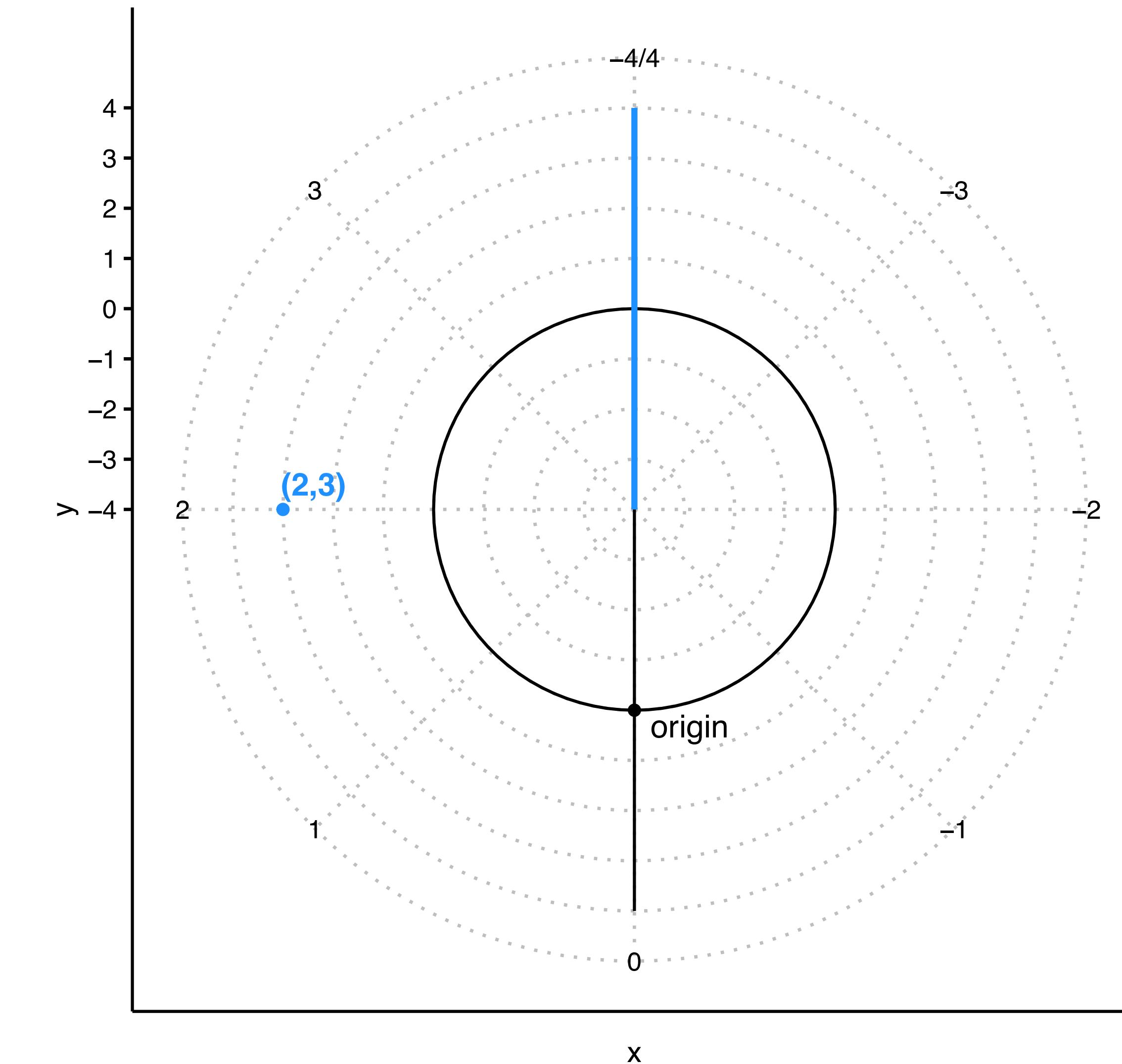


coordinates and scales, *other* coordinate systems are sometimes more effective in conveying information

cartesian coordinates

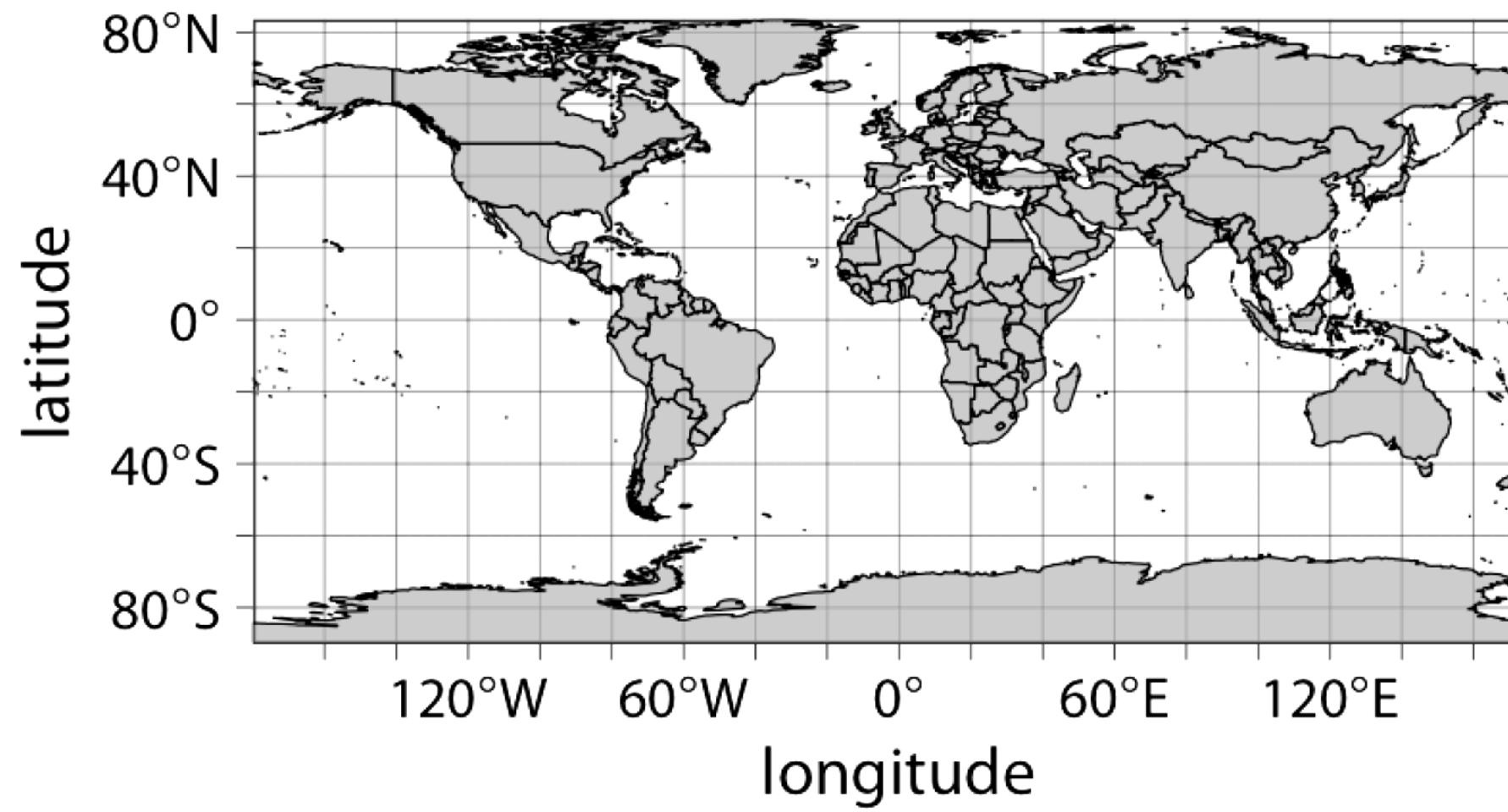


polar coordinates

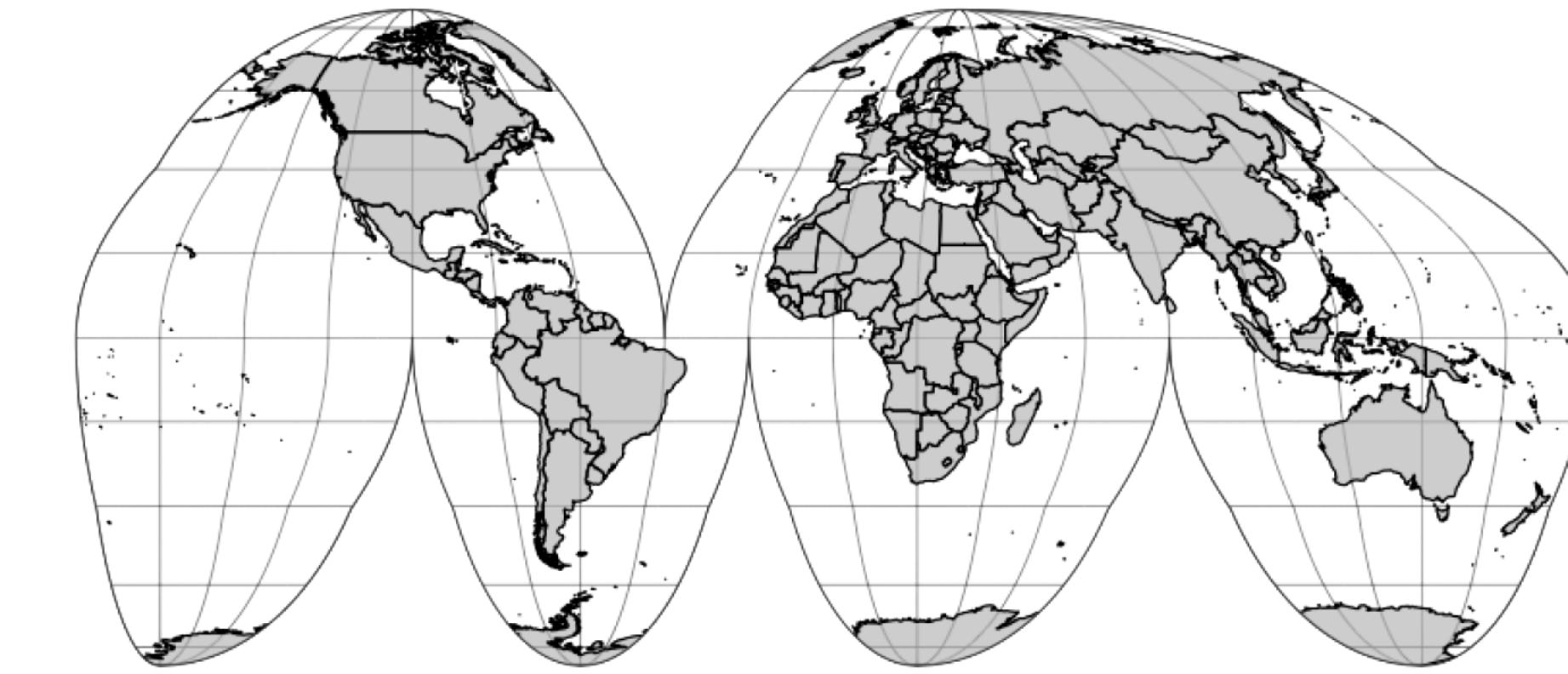


coordinates and scales, *another example*, projecting spherical surface to a plane

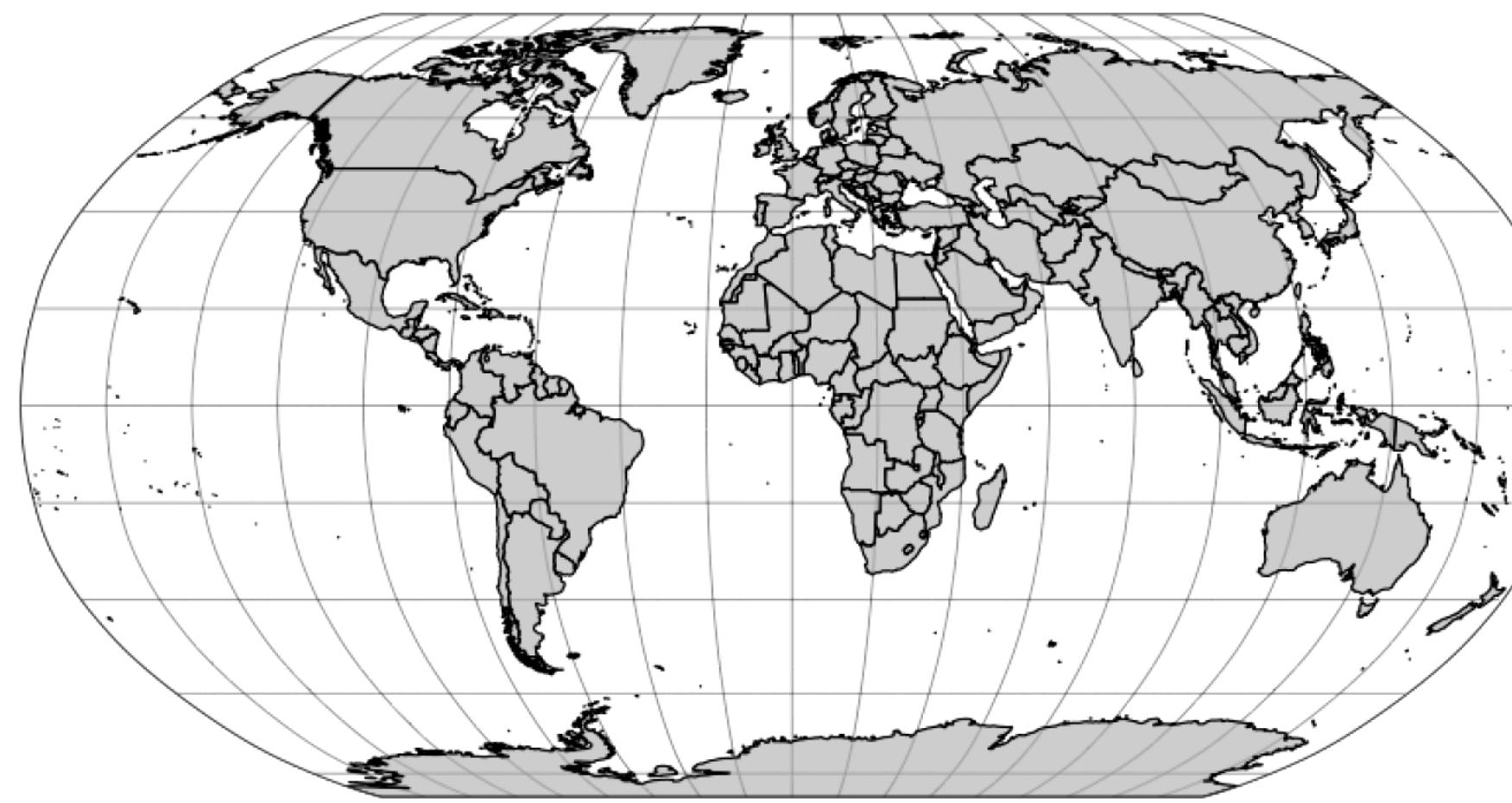
Cartesian longitude and latitude



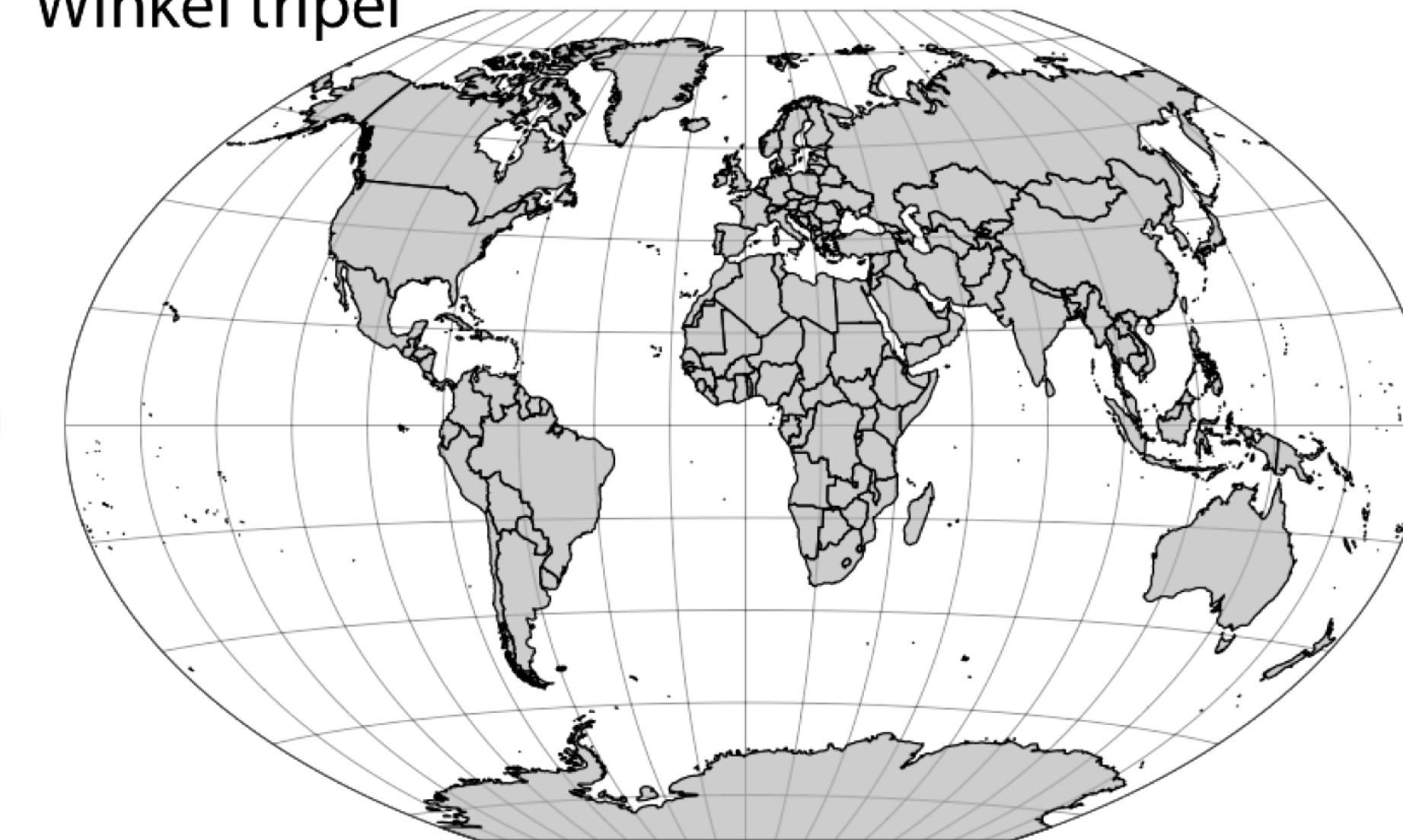
Interrupted Goode homolosine



Robinson

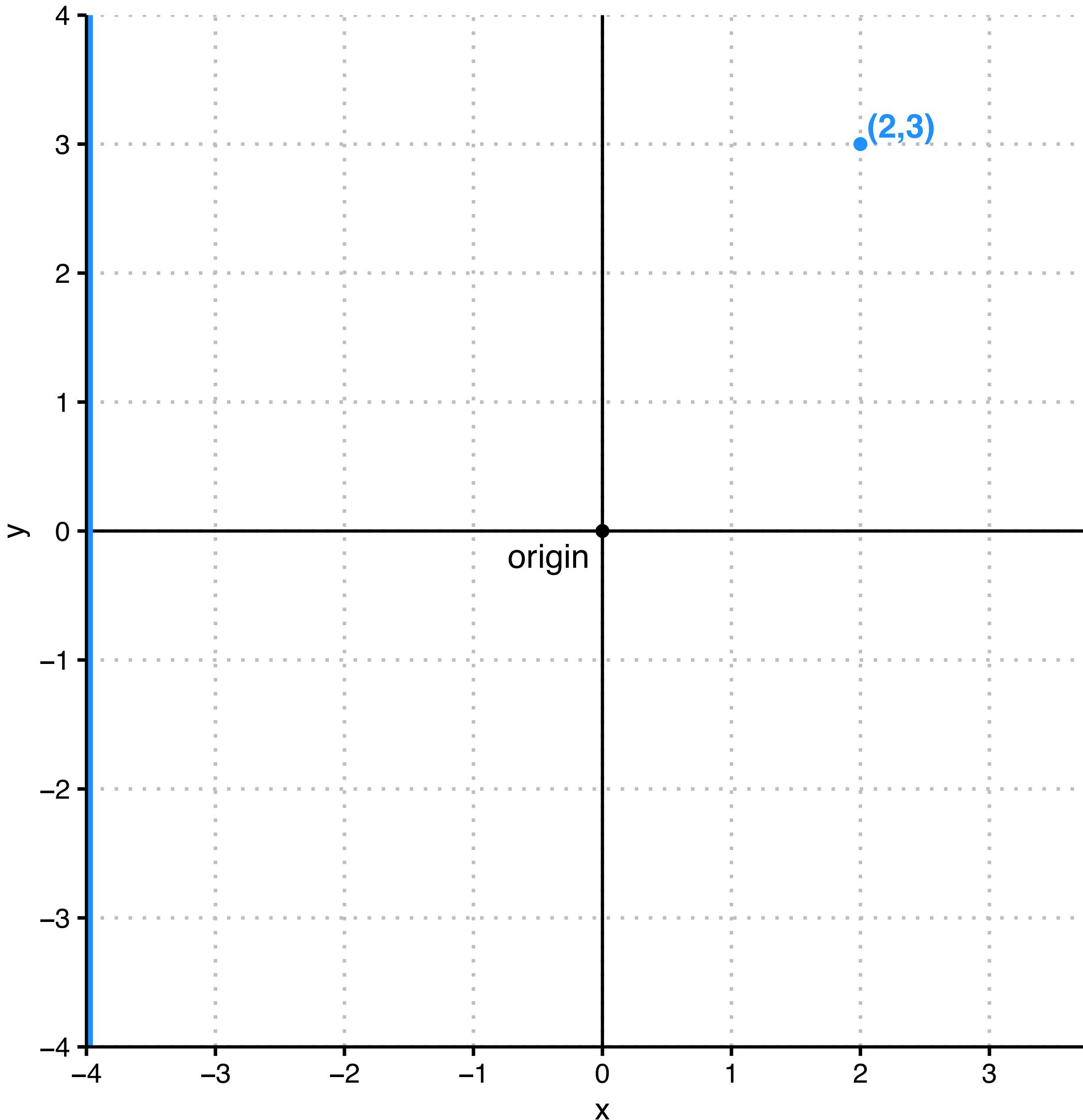


Winkel tripel

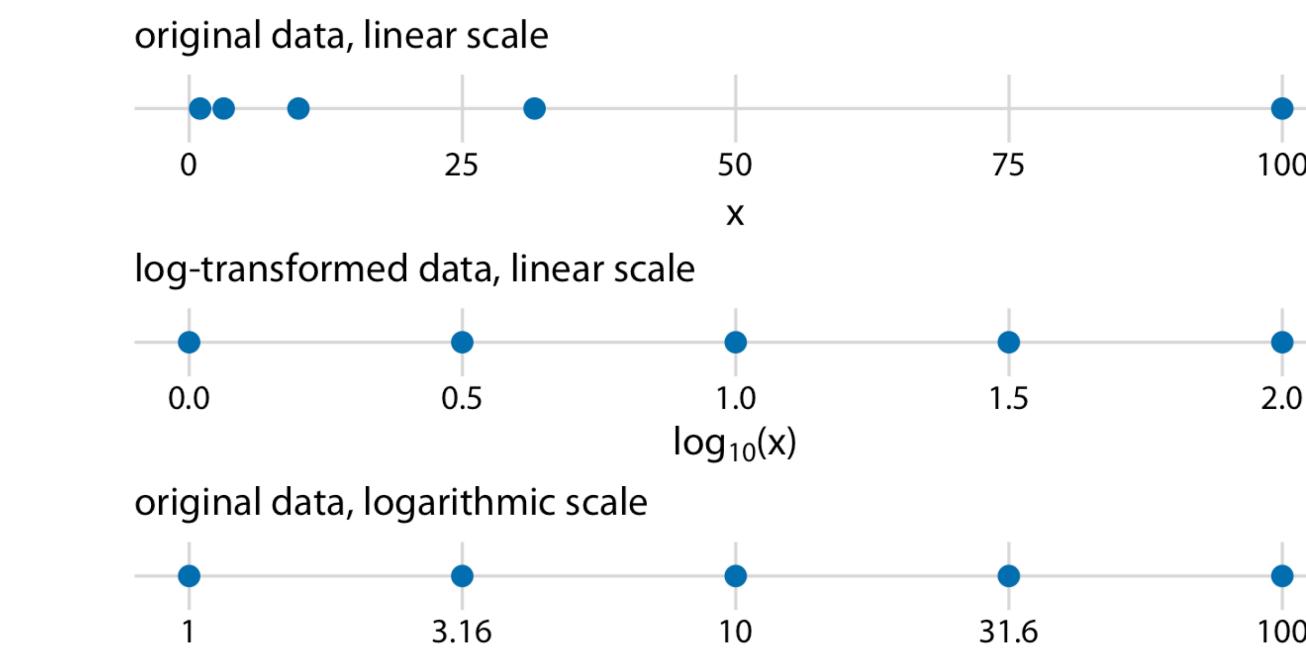


coordinates and scales, as with choosing coordinates, we can *transform scales for data or axes* for better understanding

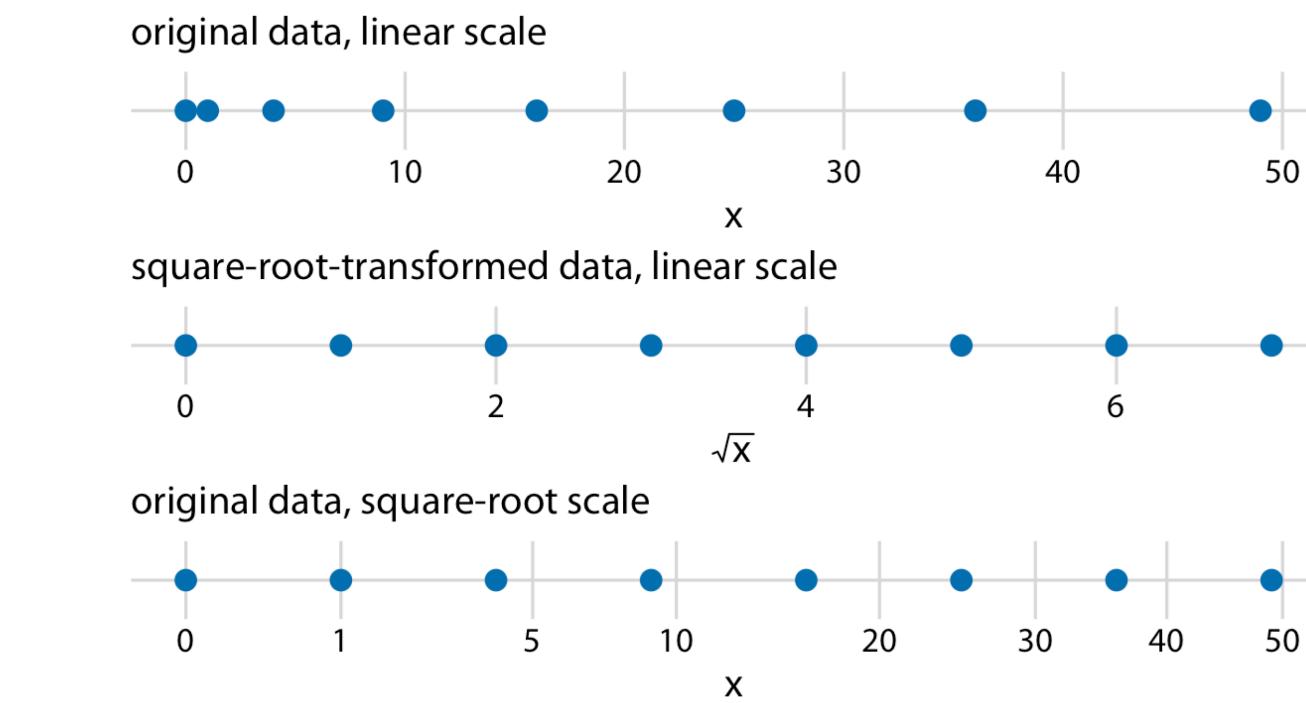
### *linear scales on cartesian coordinates*



### *example — log transforms*

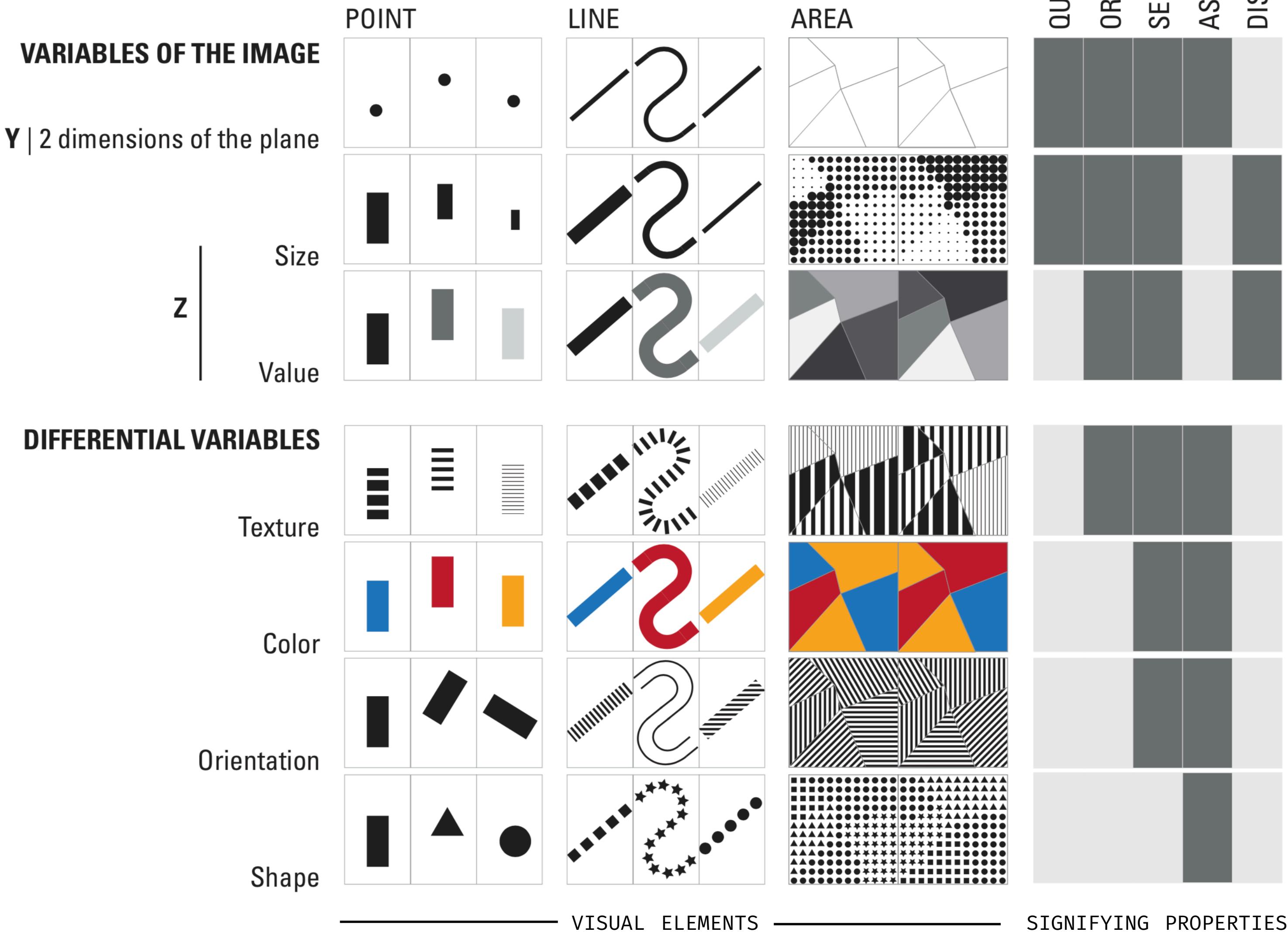


### *example — square-root transforms*



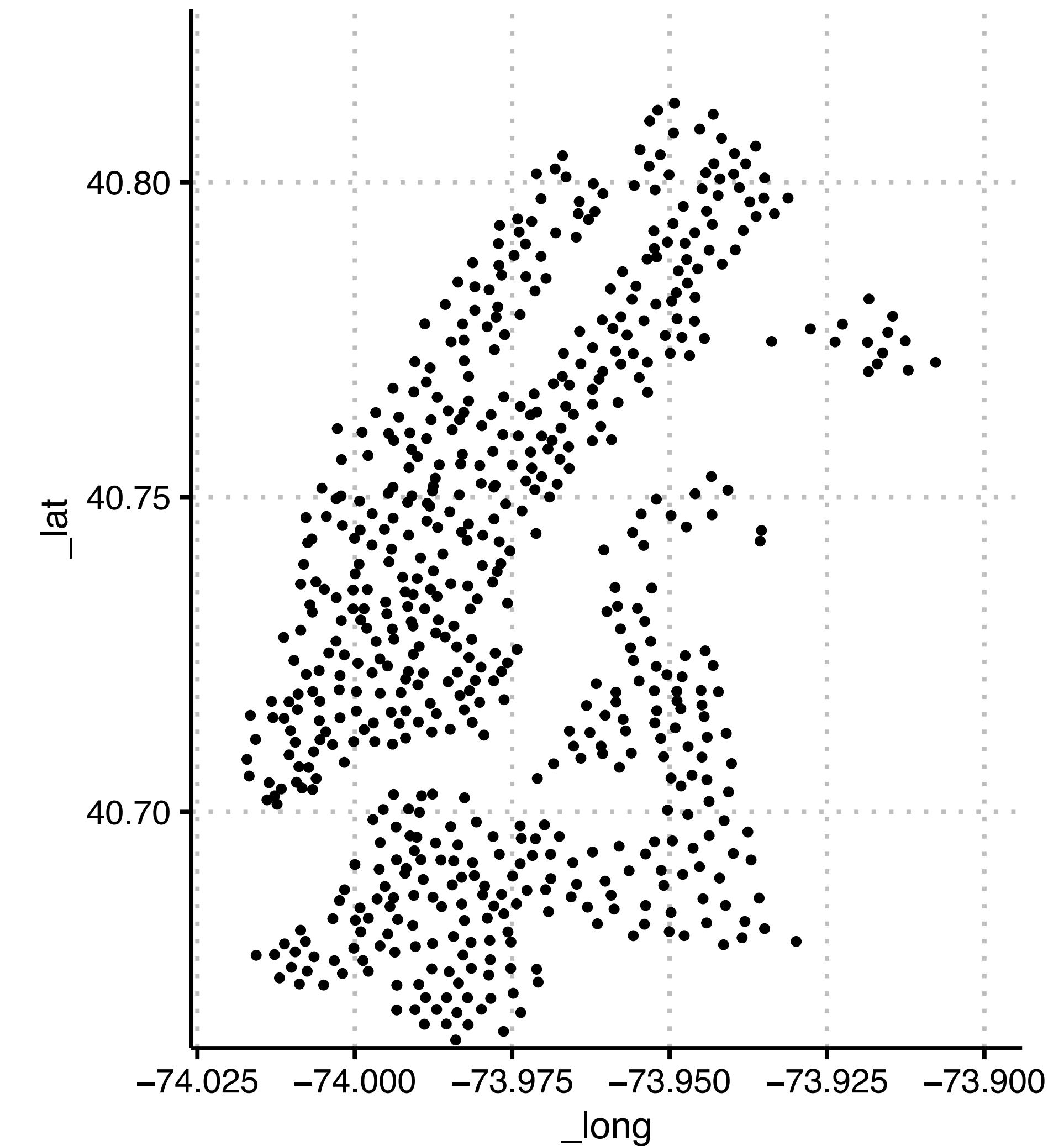
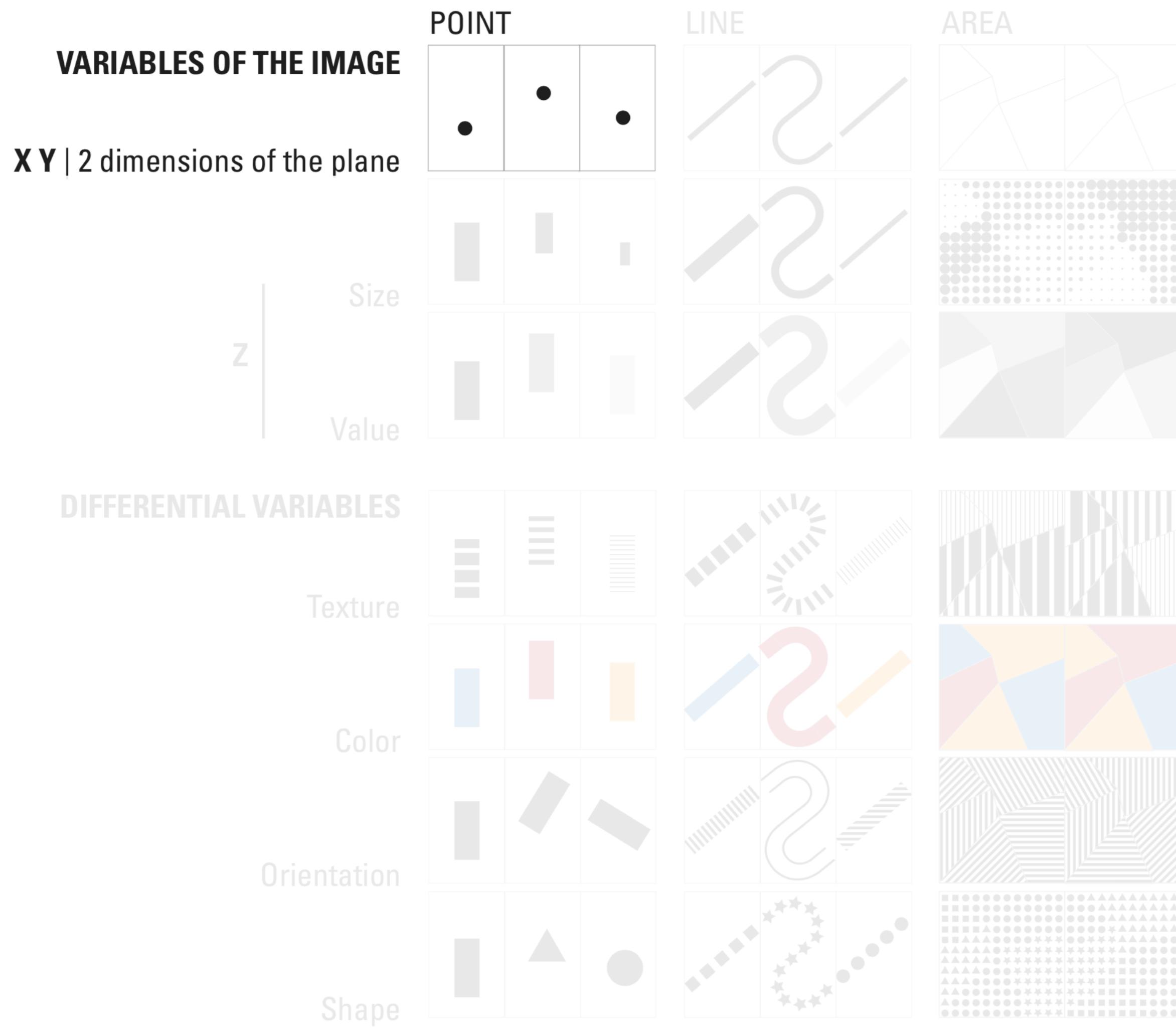
**data encodings for visual comparison**

# data encodings, visual channels for encoding data



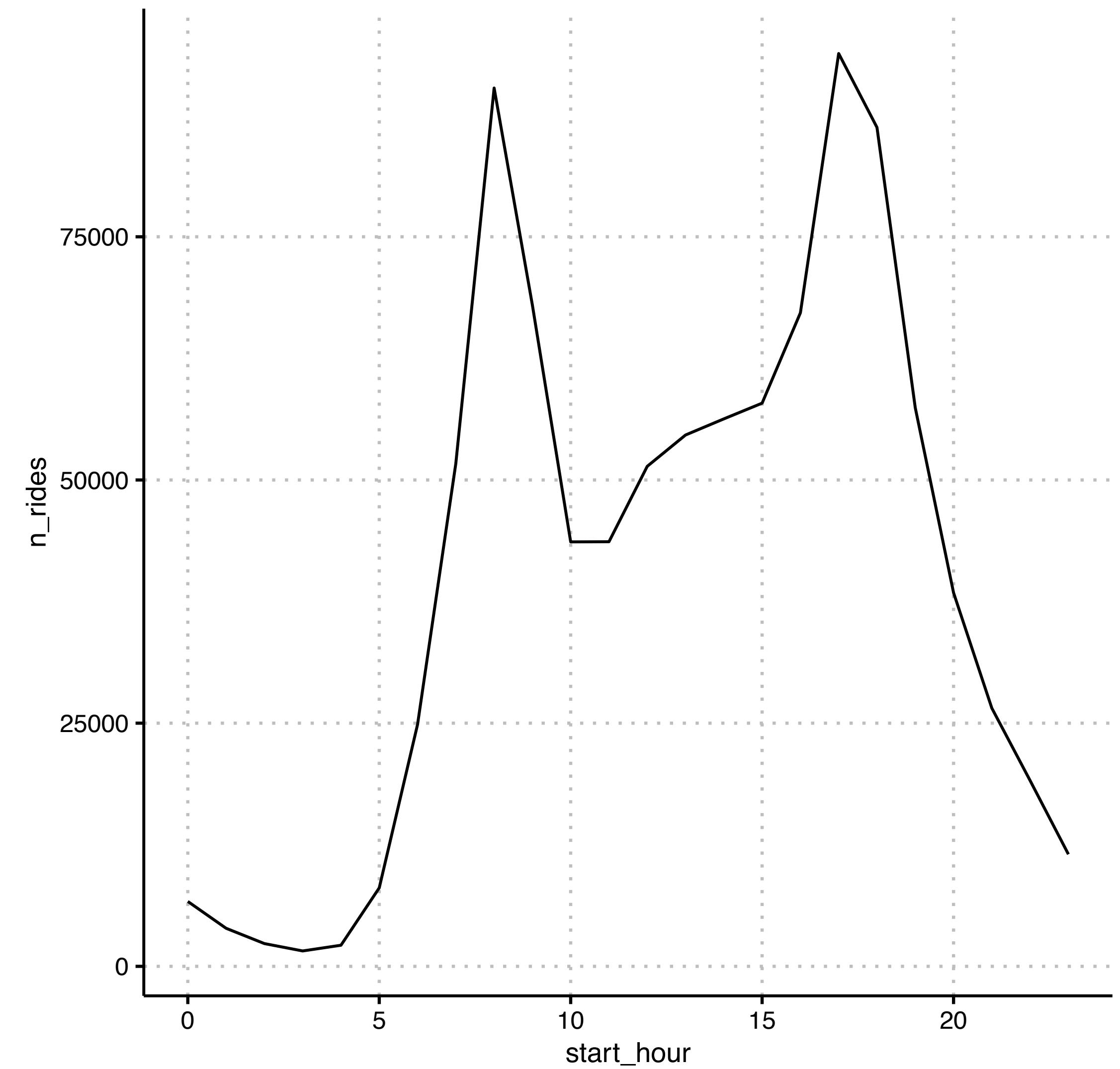
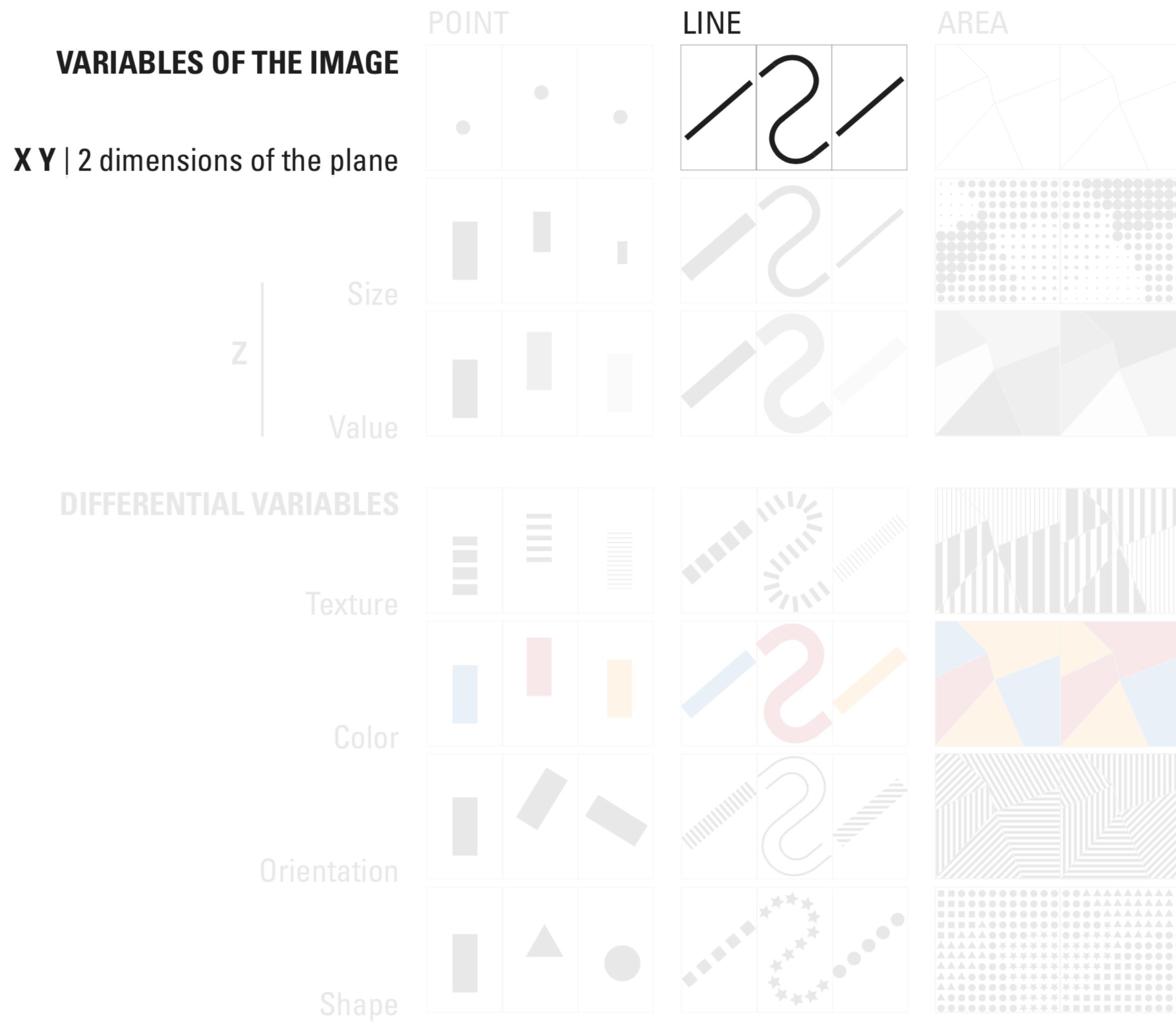
# data encodings, visual channels for encoding data

## Citi Bike example — *exploratory data analysis*



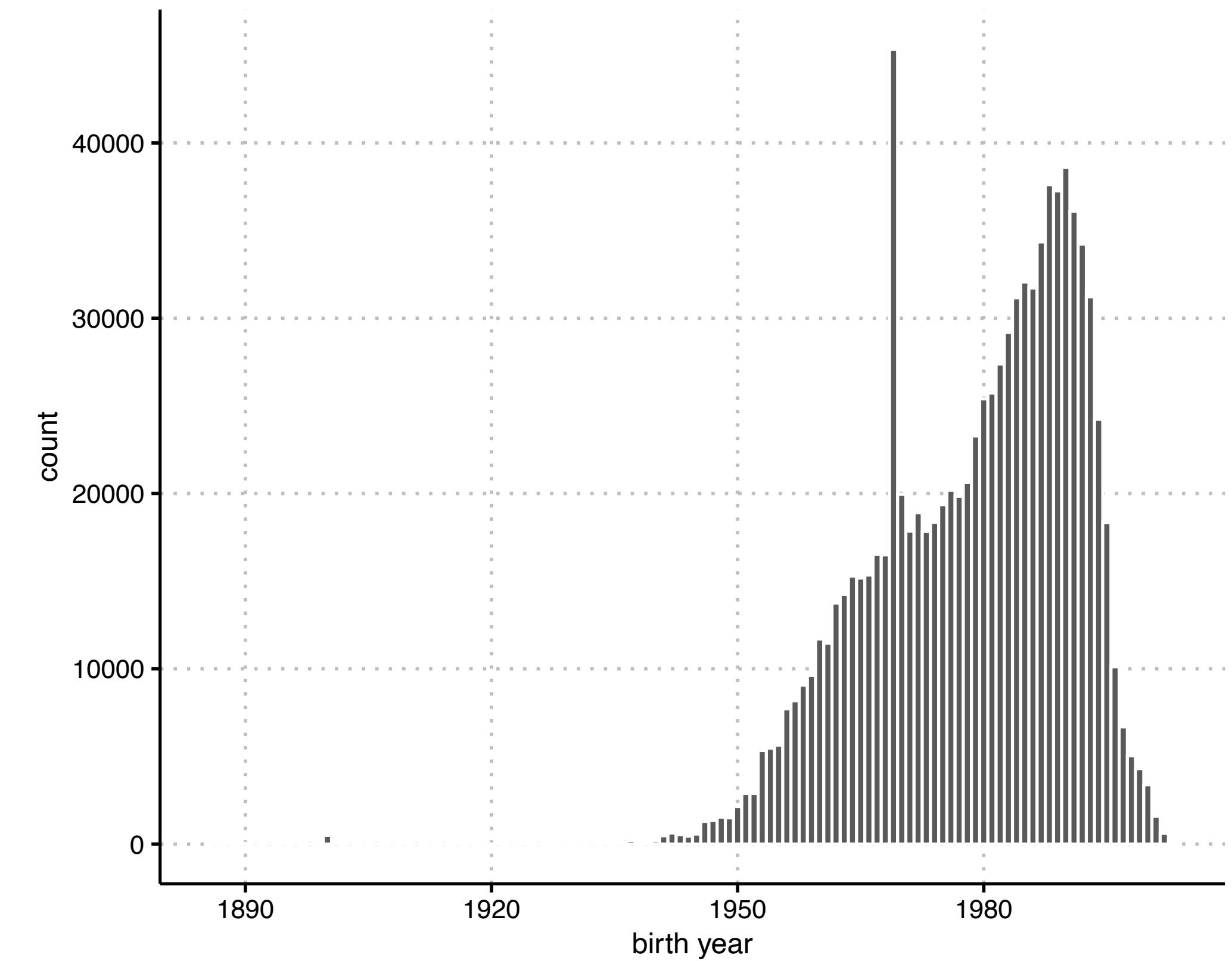
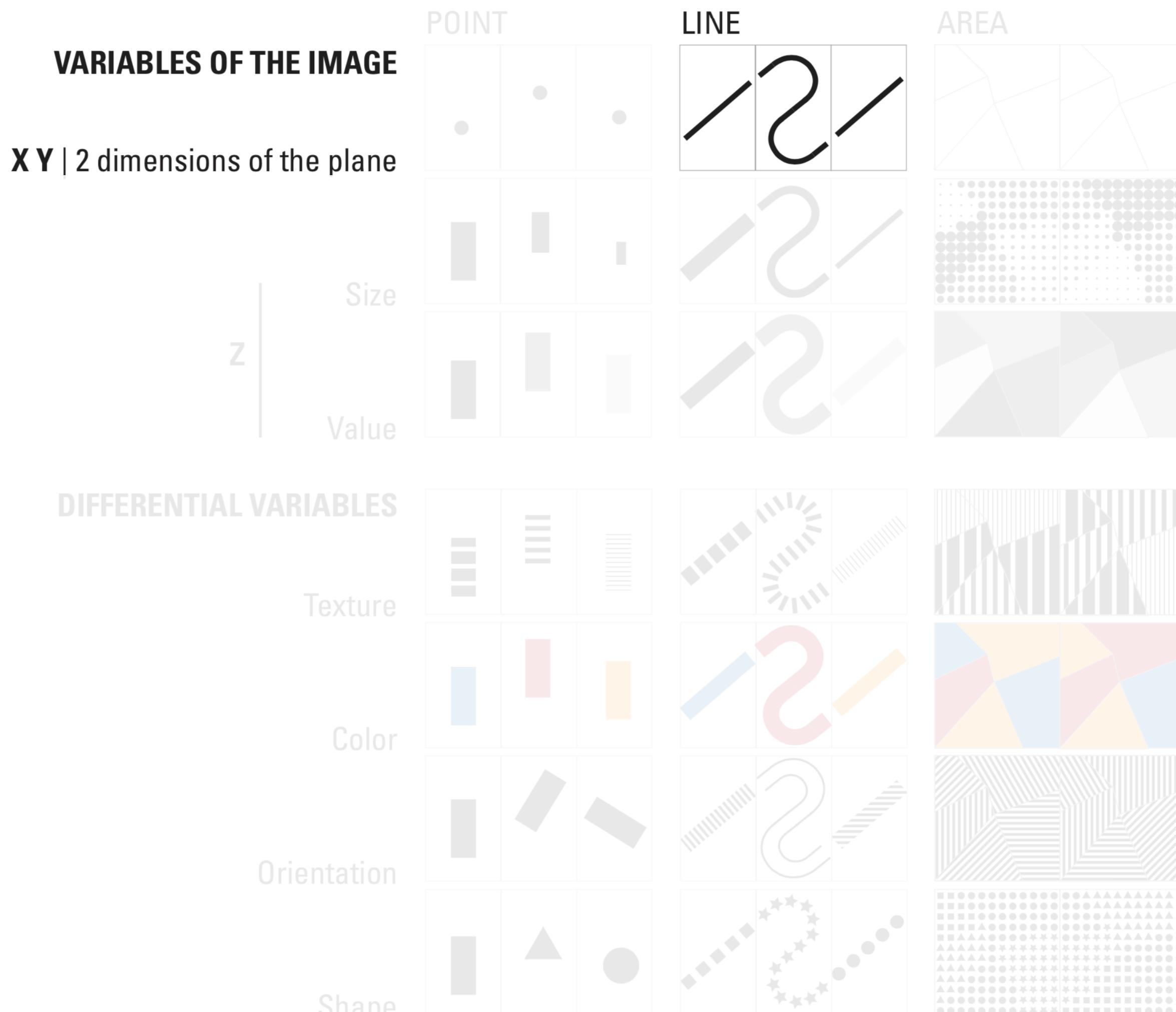
# data encodings, visual channels for encoding data

## Citi Bike example — *exploratory data analysis*



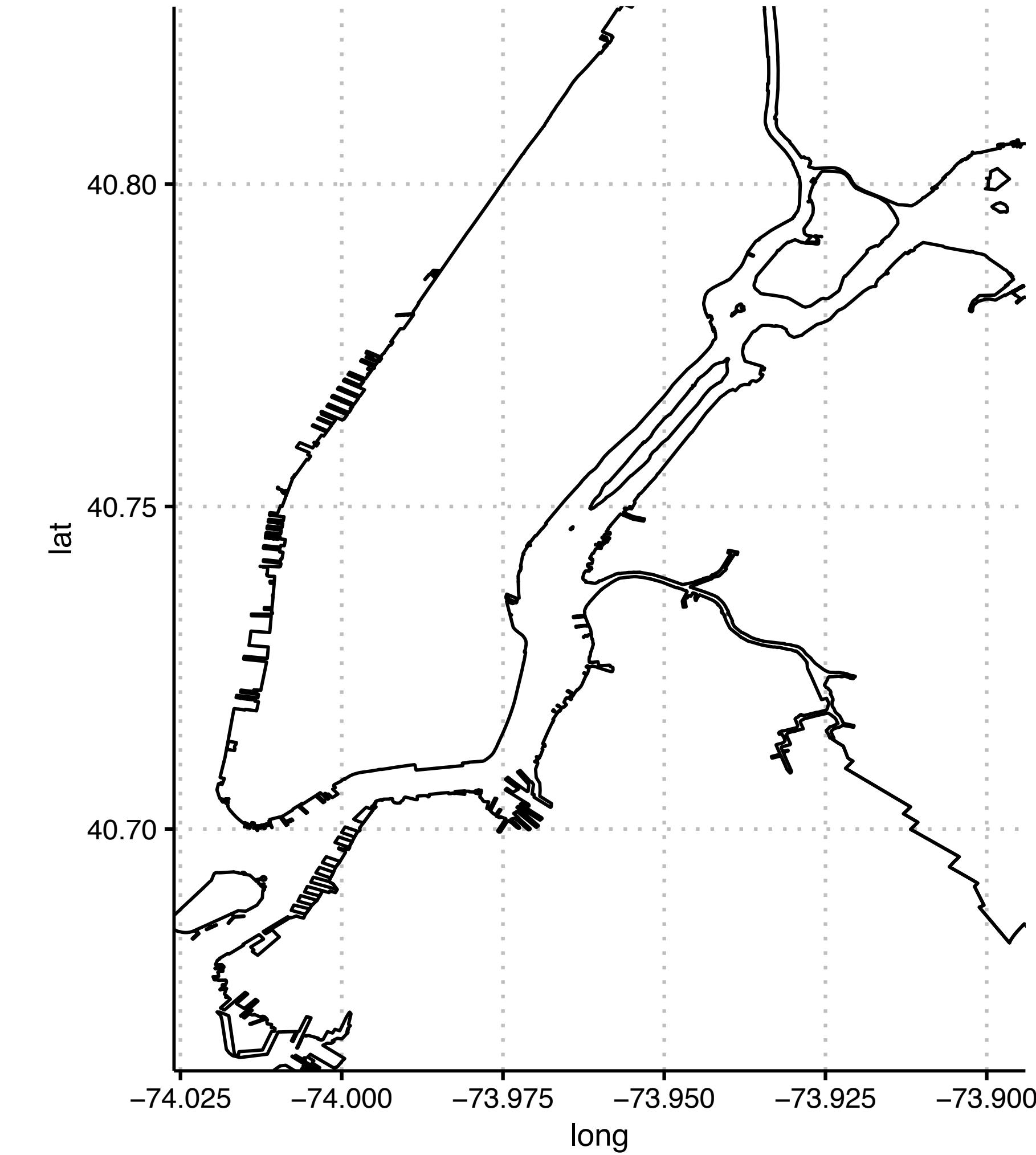
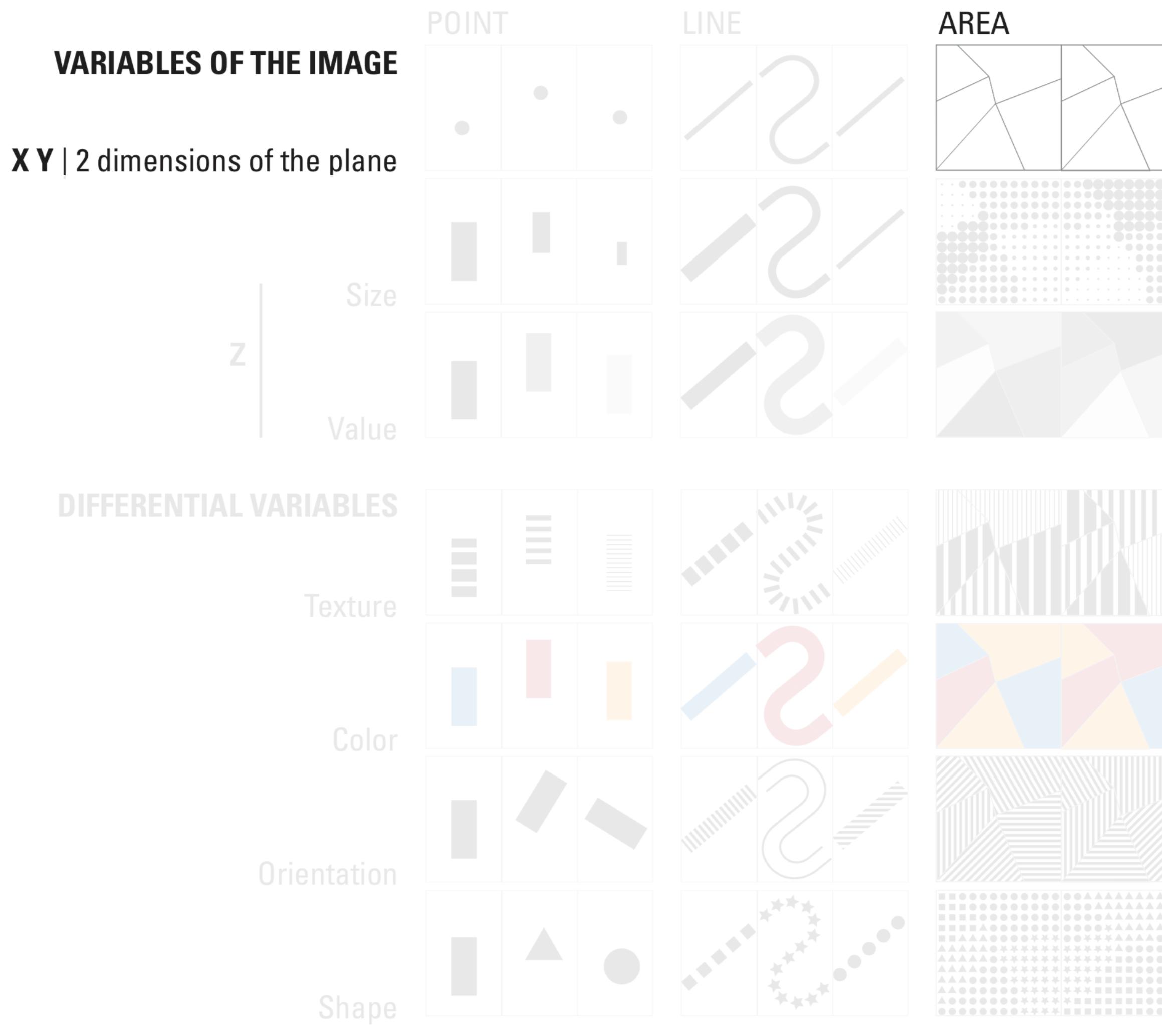
# data encodings, visual channels for encoding data

## Citi Bike example — *exploratory data analysis*



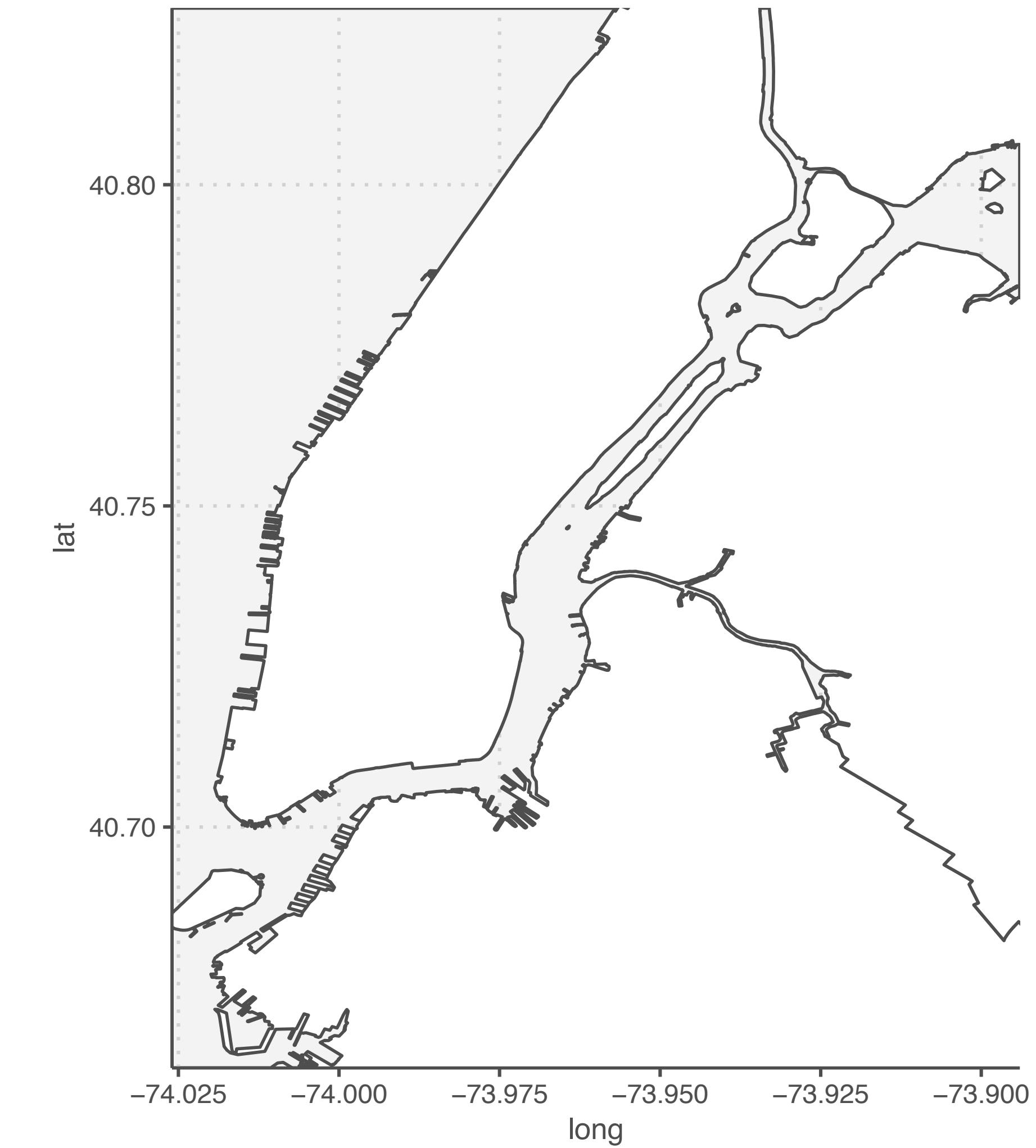
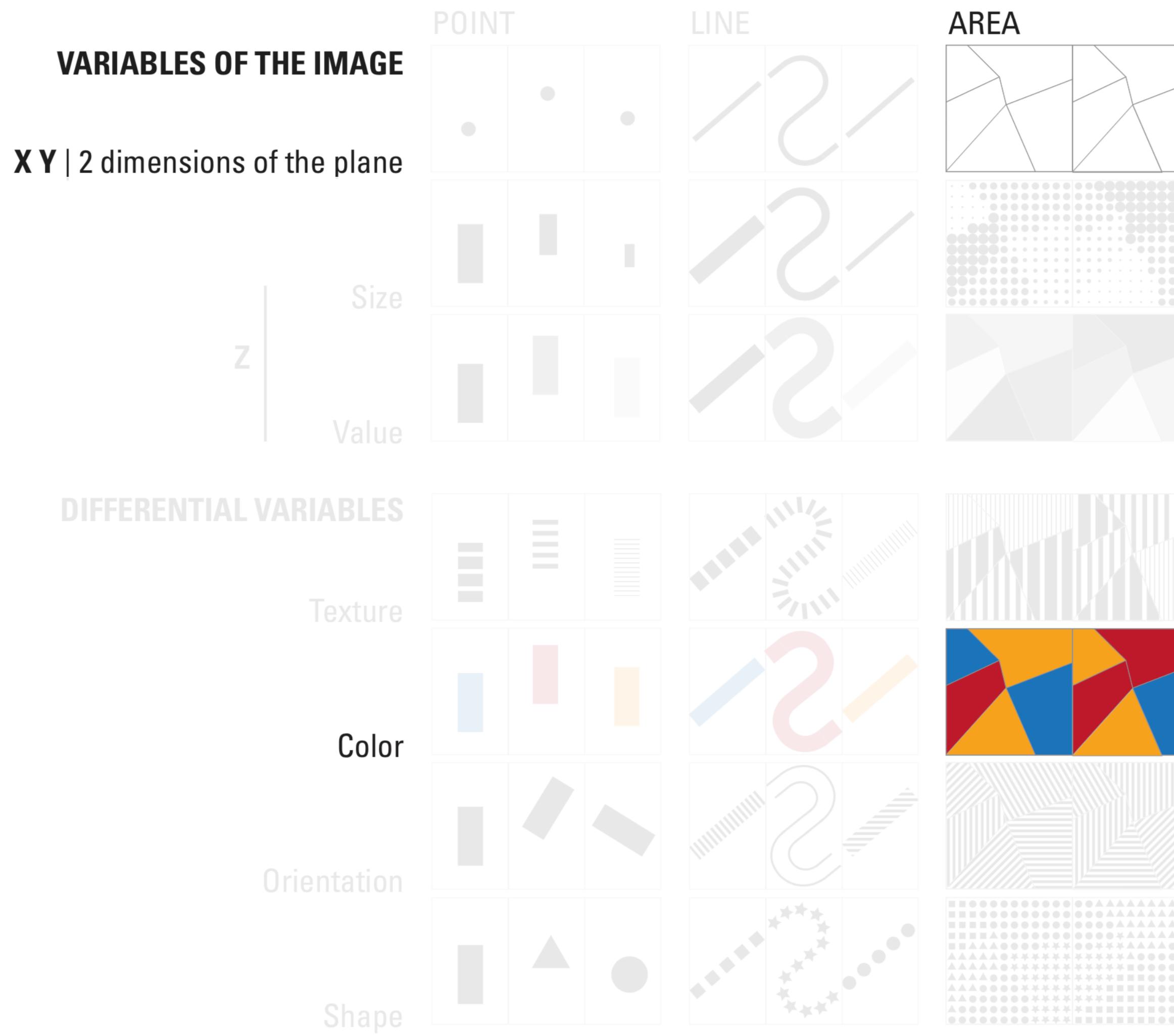
# data encodings, visual channels for encoding data

## Citi Bike example — *exploratory data analysis*



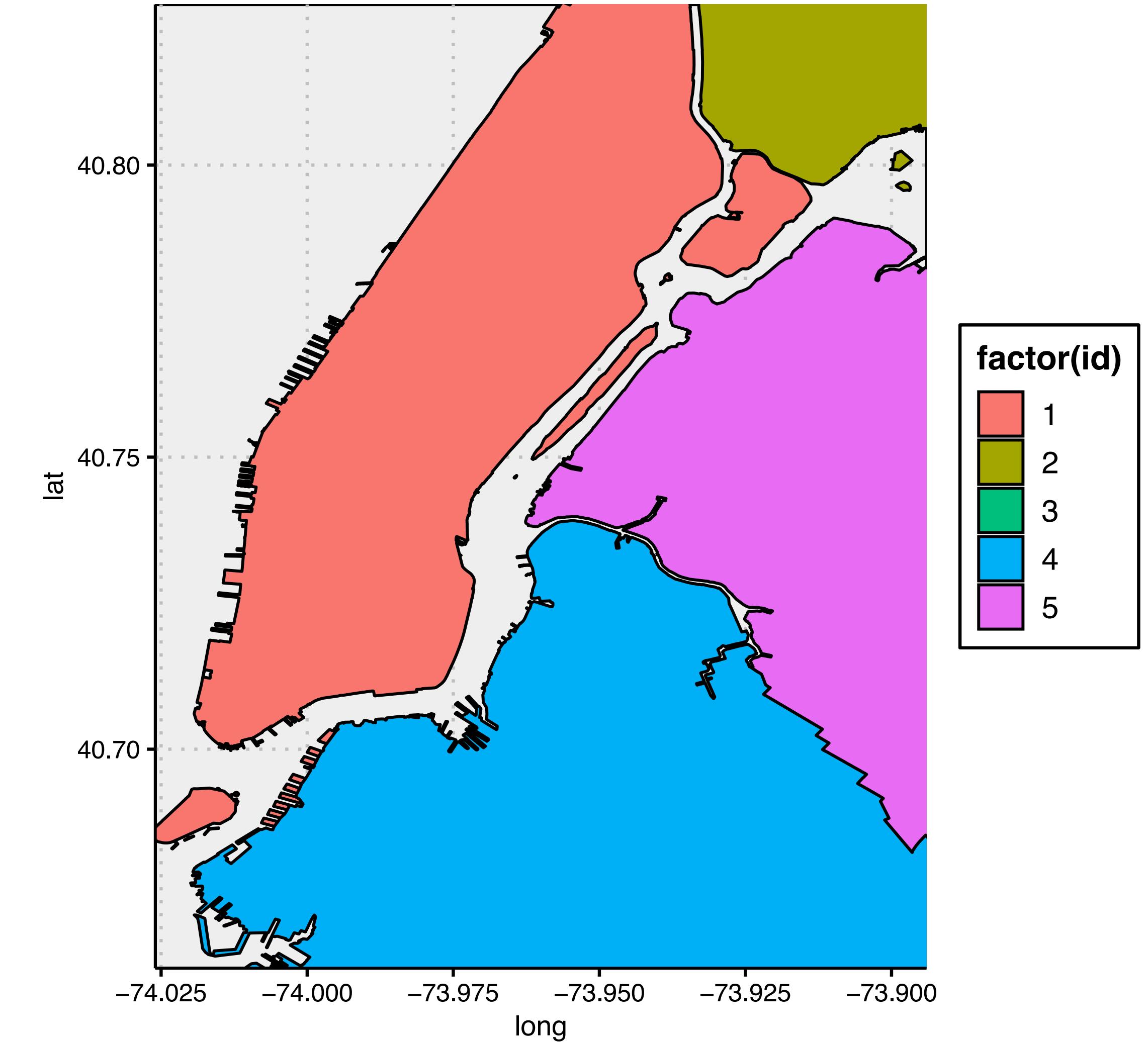
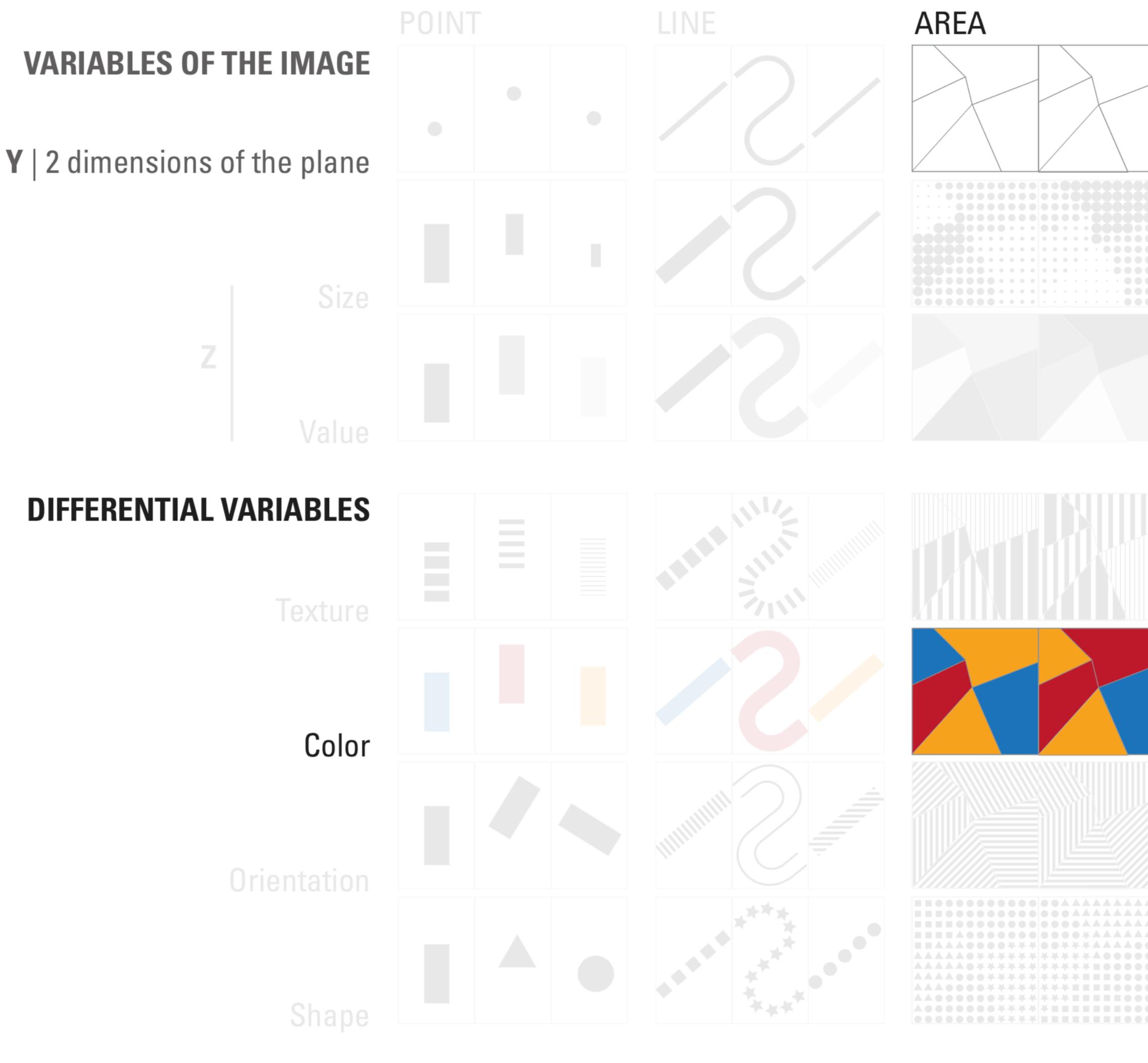
# data encodings, visual channels for encoding data

## Citi Bike example — *exploratory data analysis*



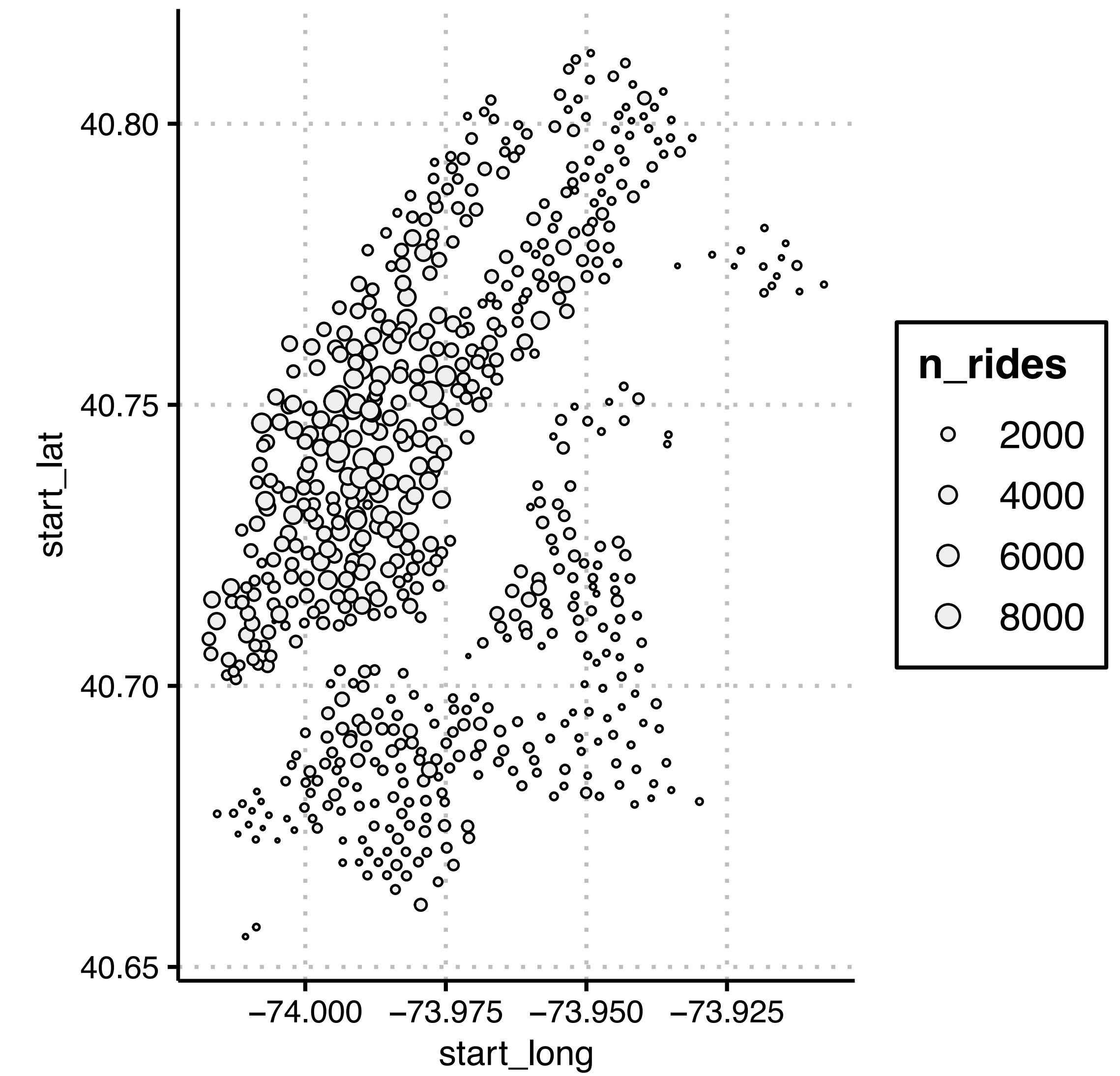
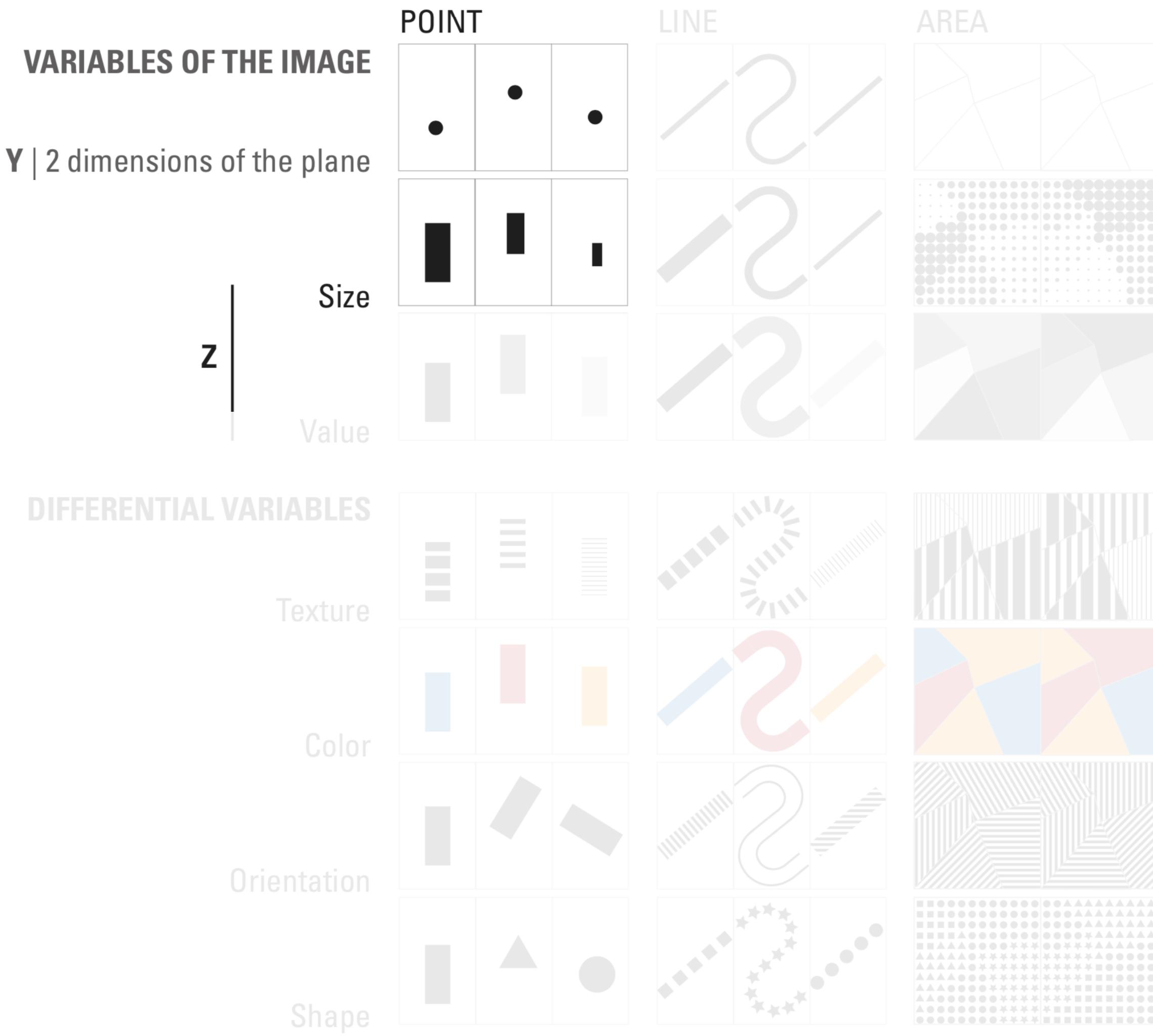
# data encodings, visual channels for encoding data

## Citi Bike example — *exploratory data analysis*



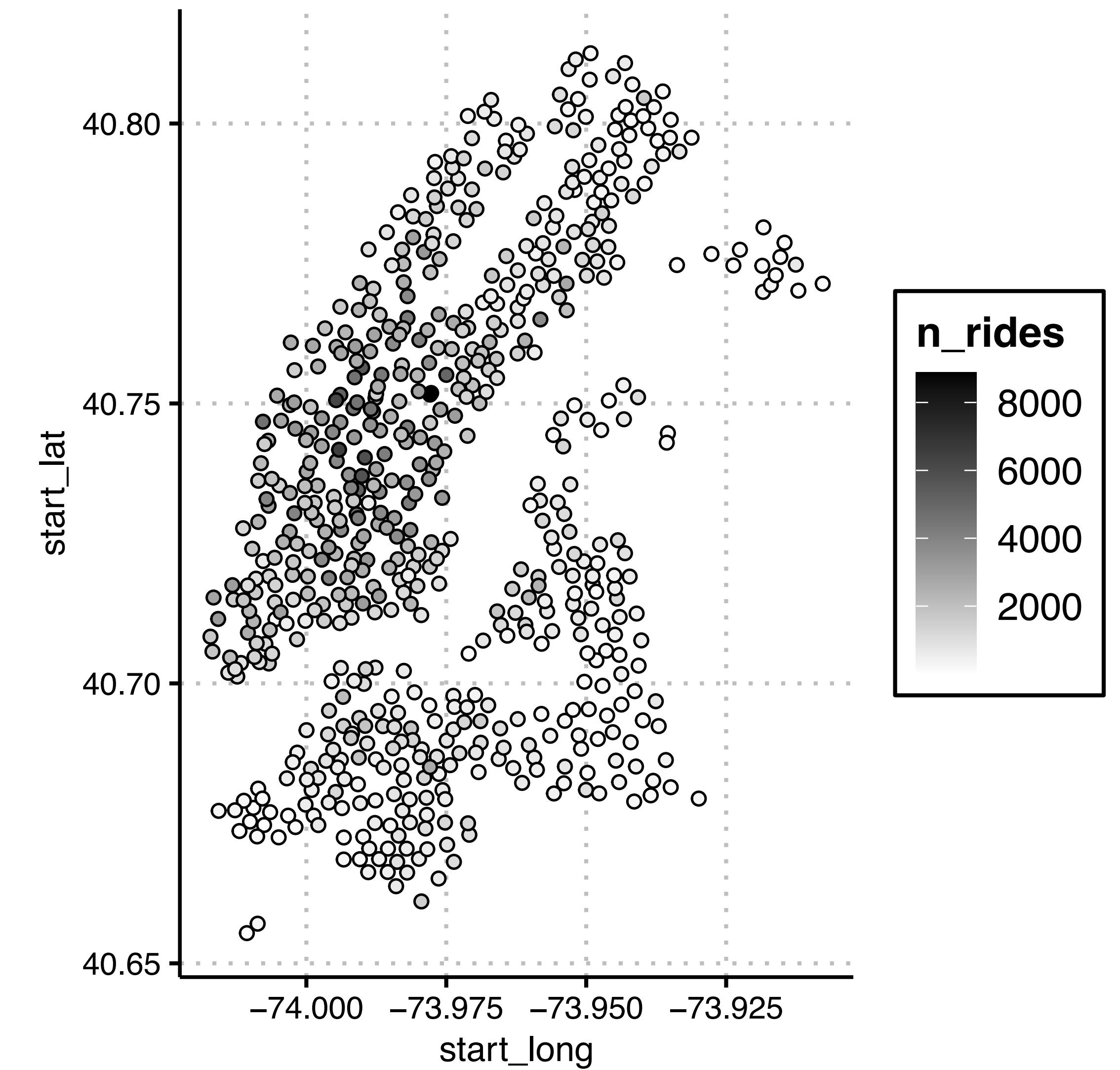
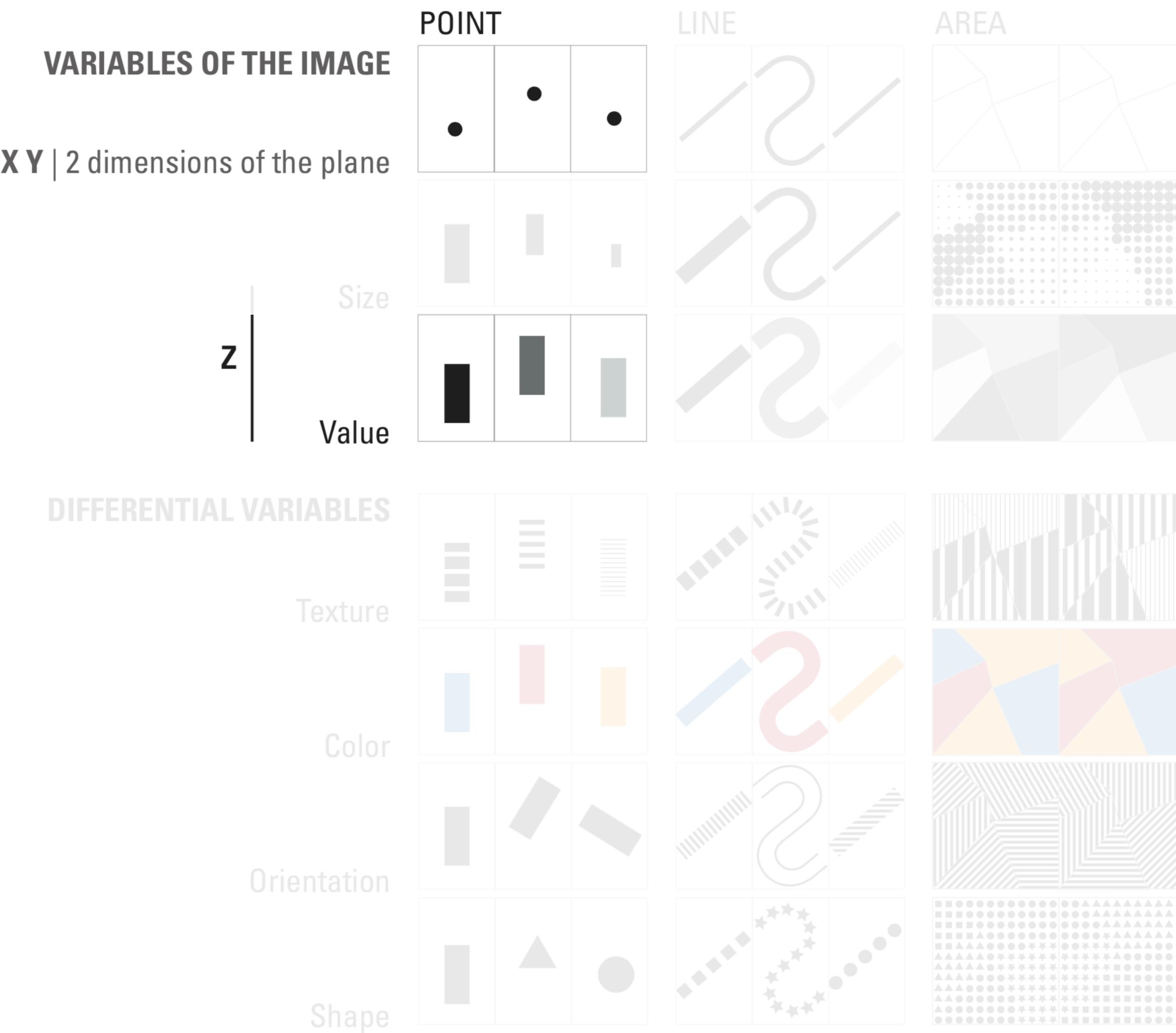
# data encodings, visual channels for encoding data

# Citi Bike example — *exploratory data analysis*



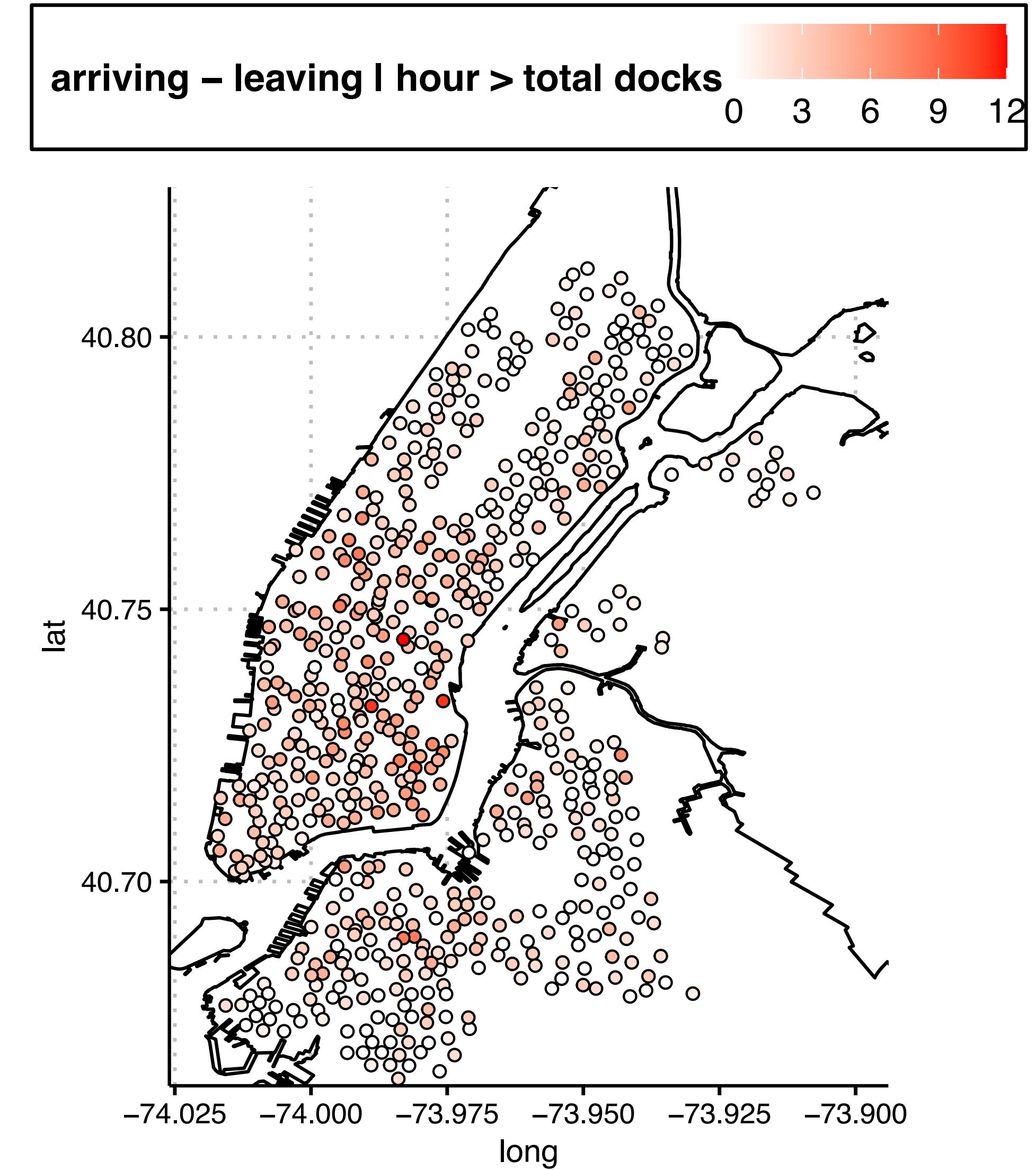
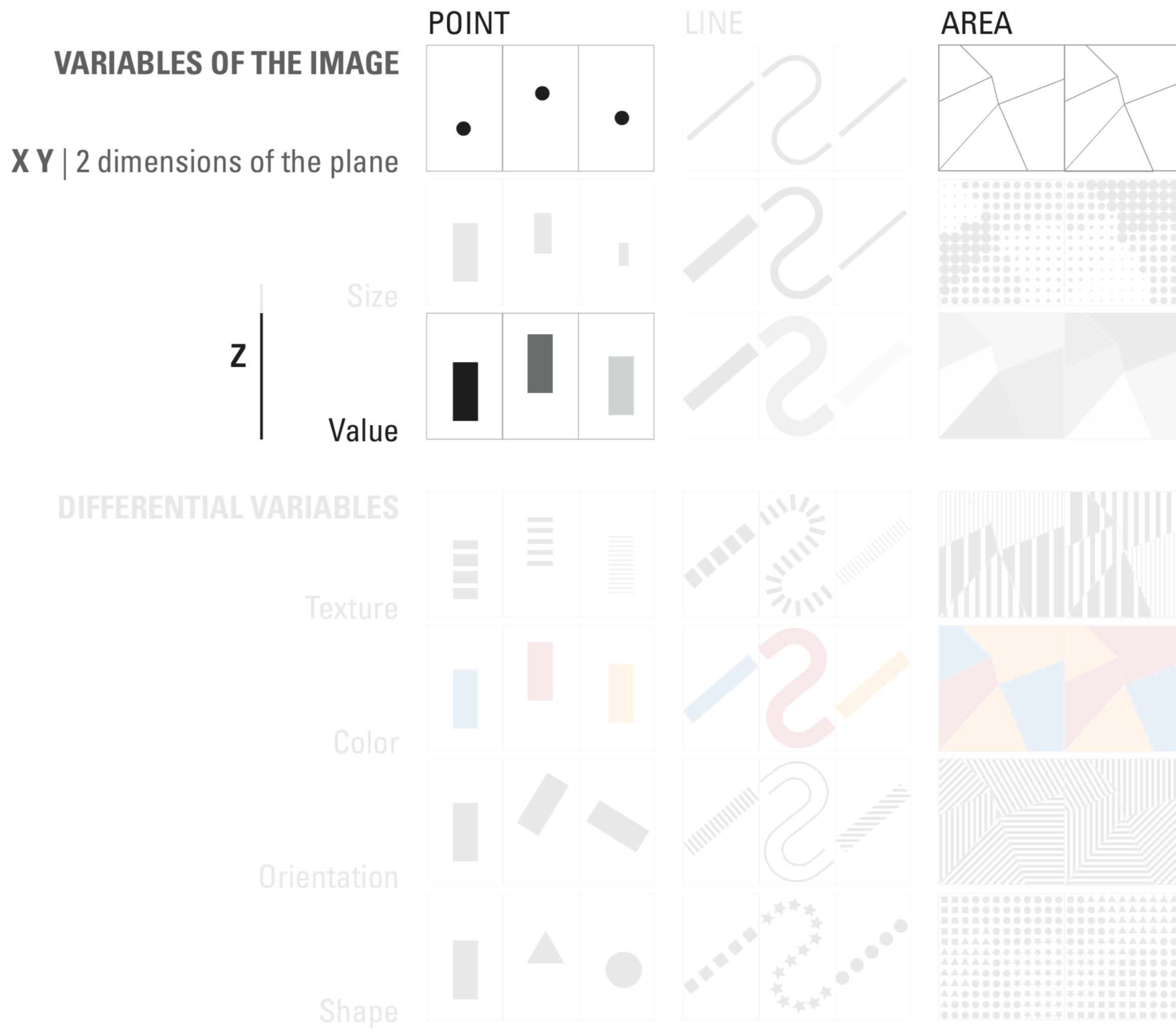
# data encodings, visual channels for encoding data

# Citi Bike example — *exploratory data analysis*



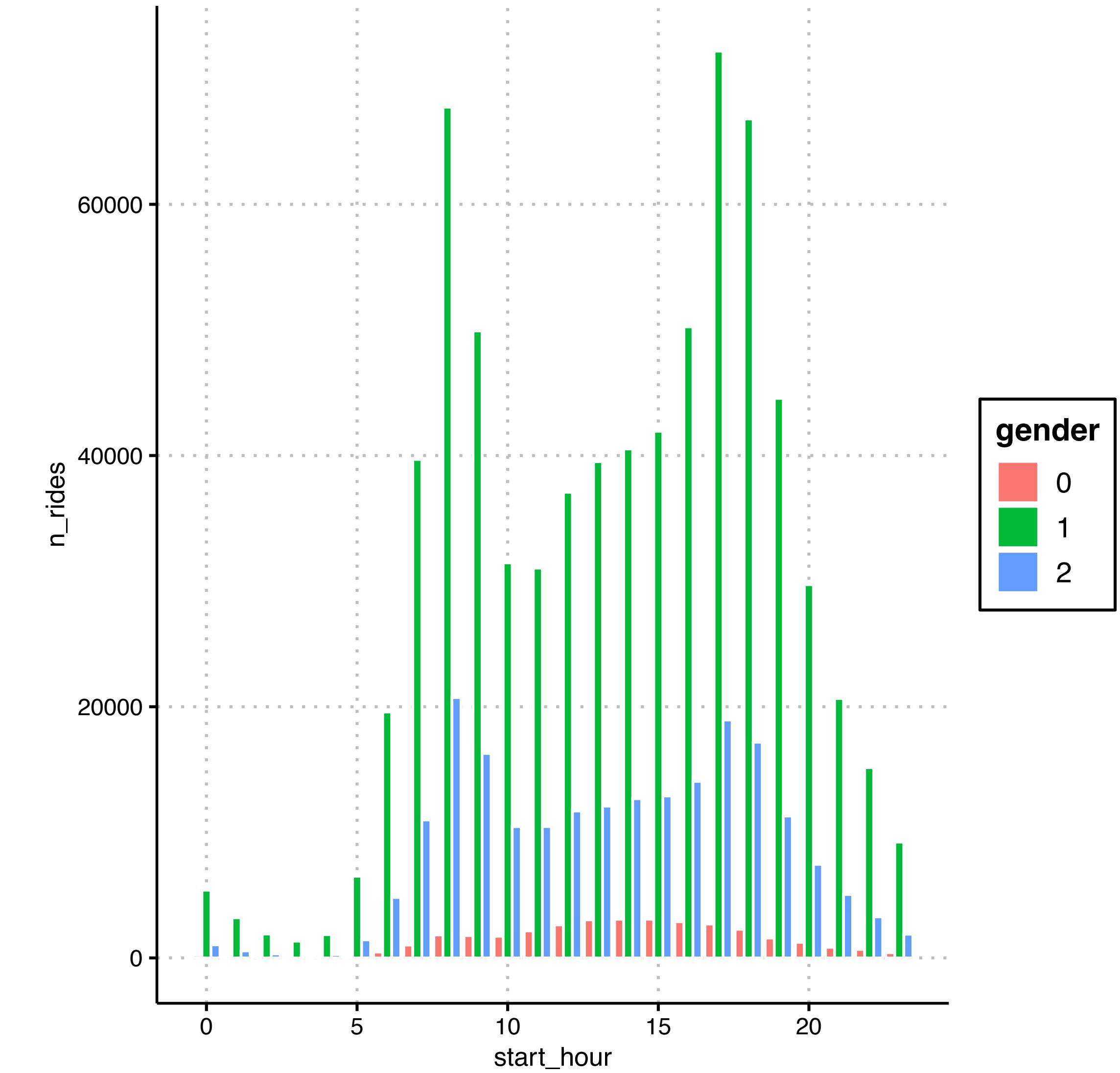
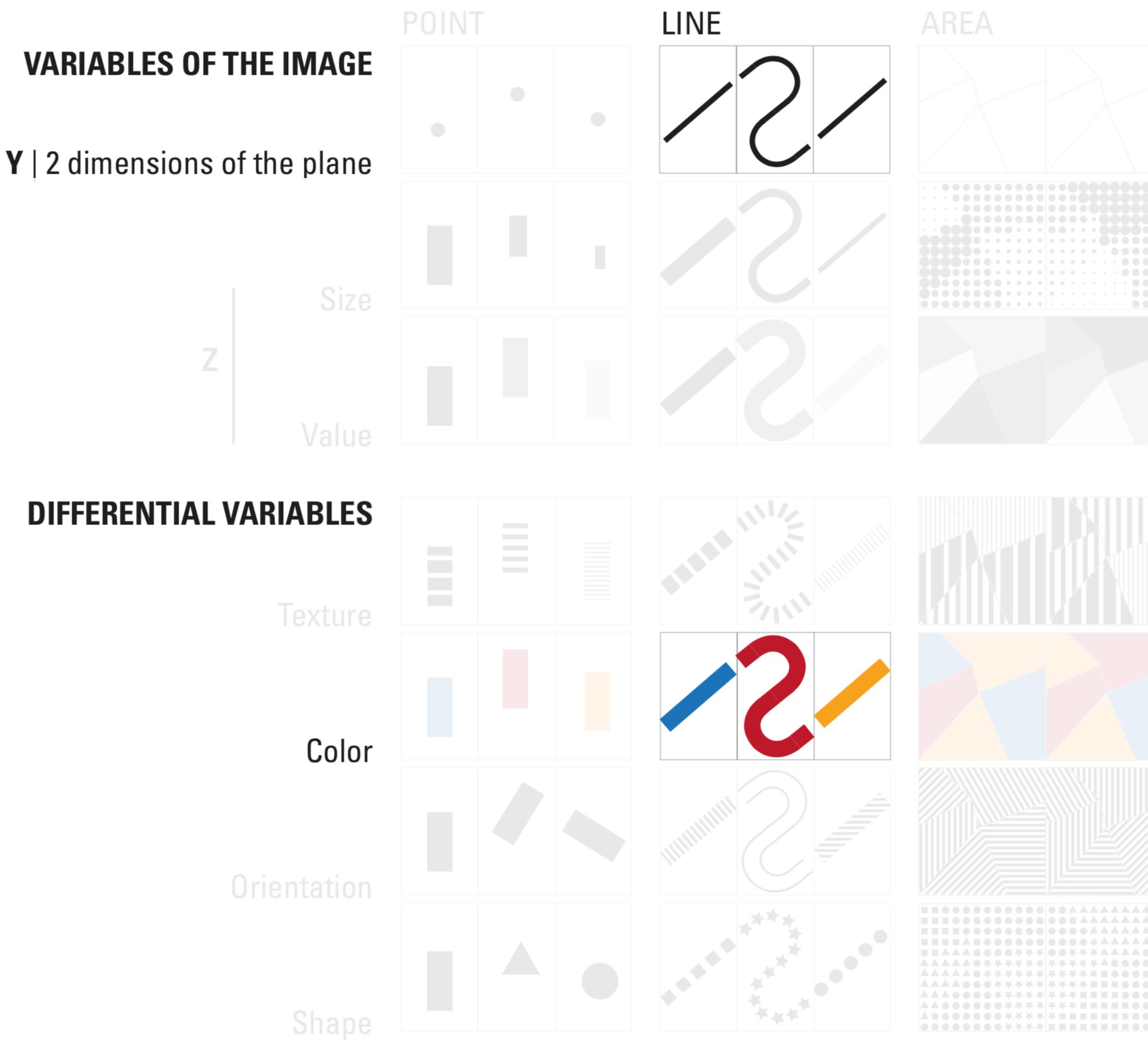
# data encodings, visual channels for encoding data

## Citi Bike example — *exploratory data analysis*



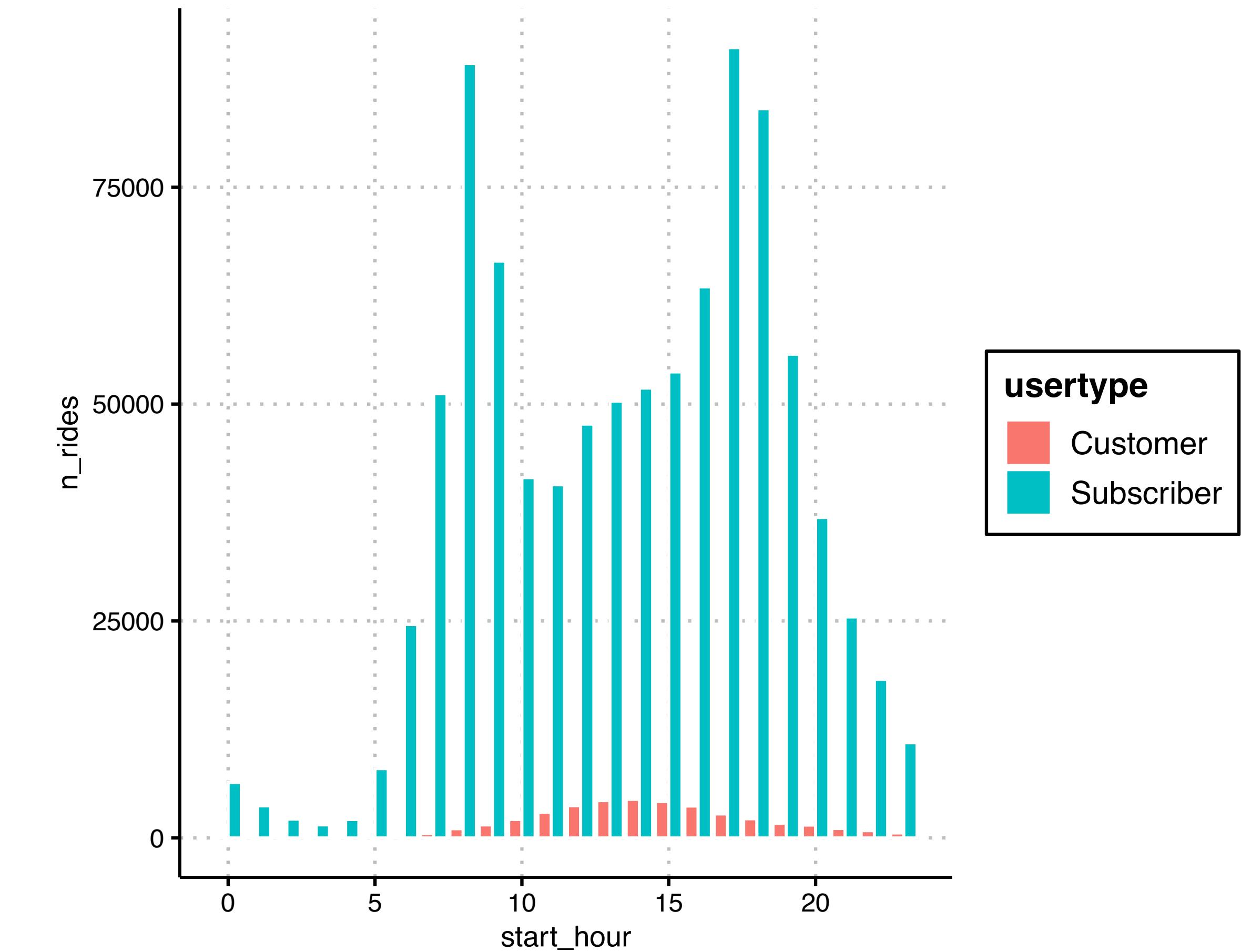
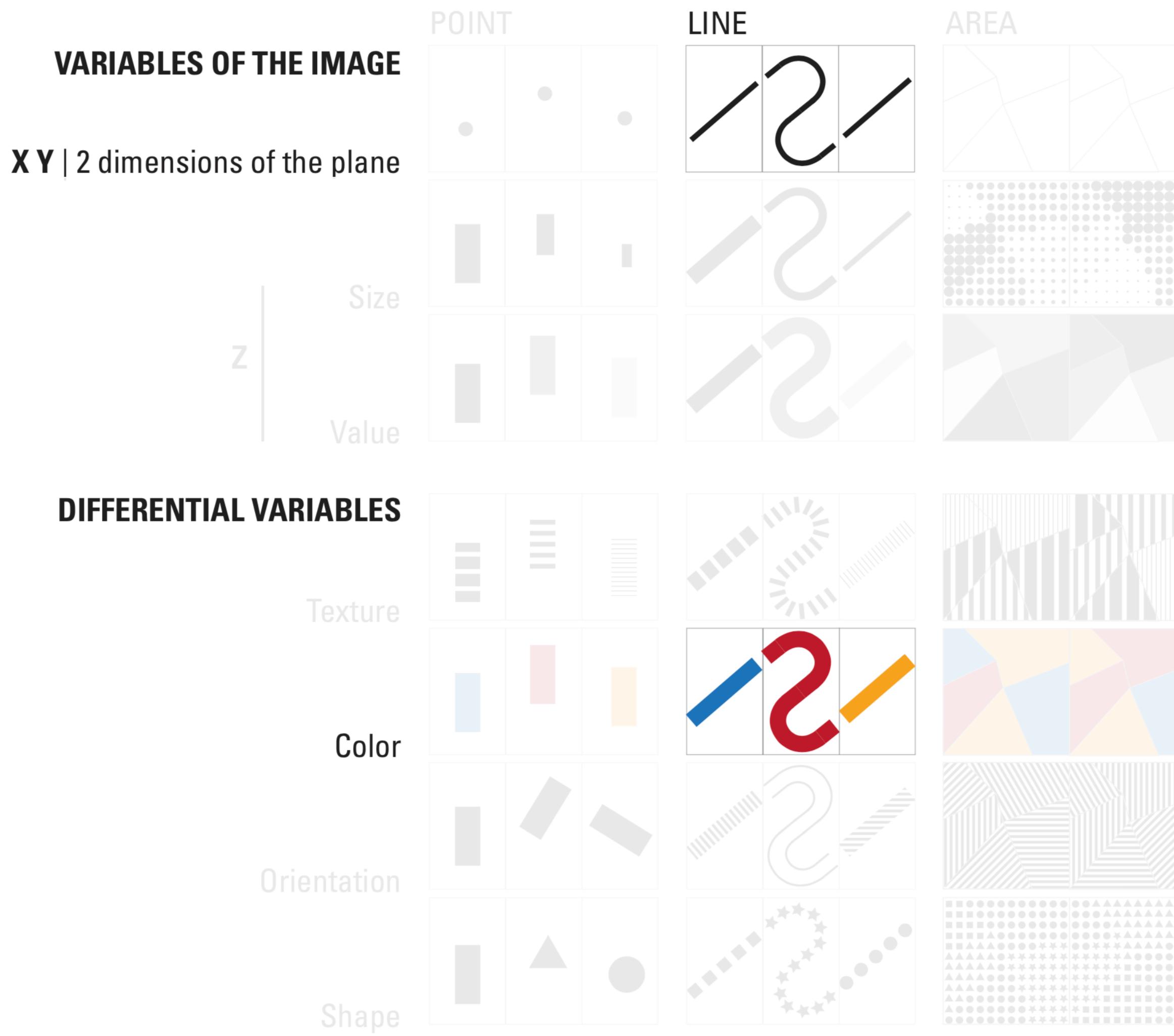
# data encodings, visual channels for encoding data

## Citi Bike example — *exploratory data analysis*



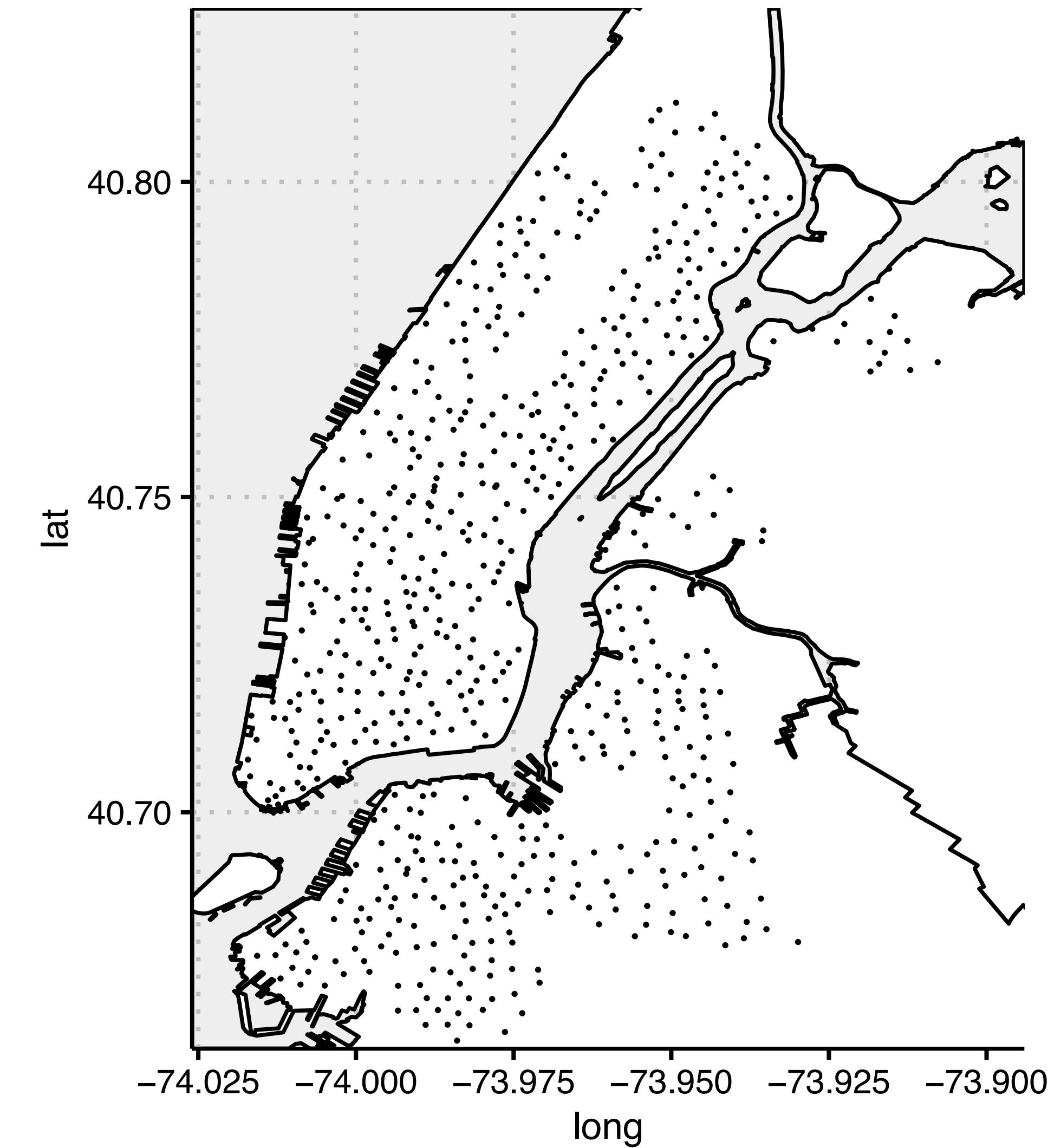
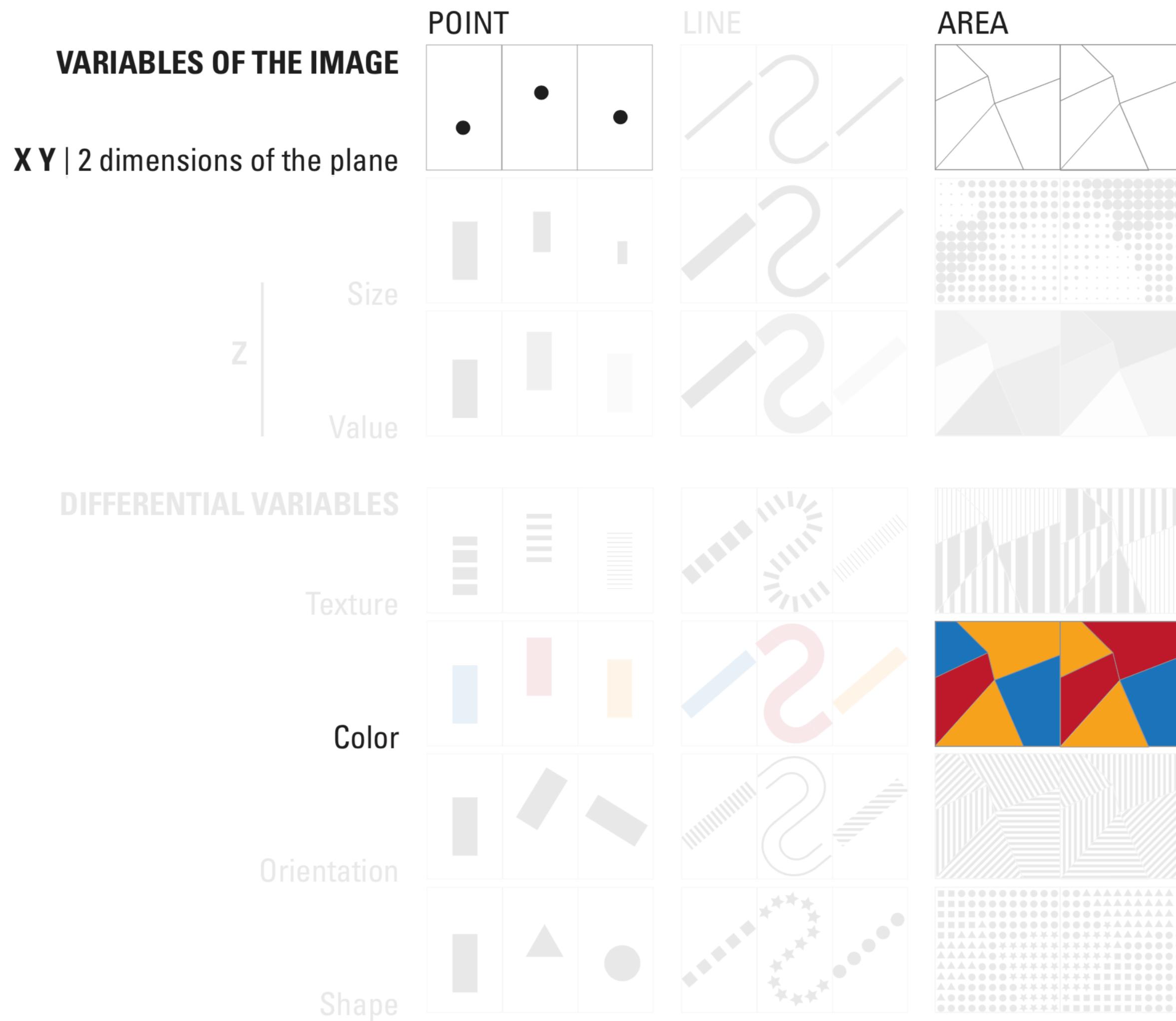
# data encodings, visual channels for encoding data

## Citi Bike example — *exploratory data analysis*



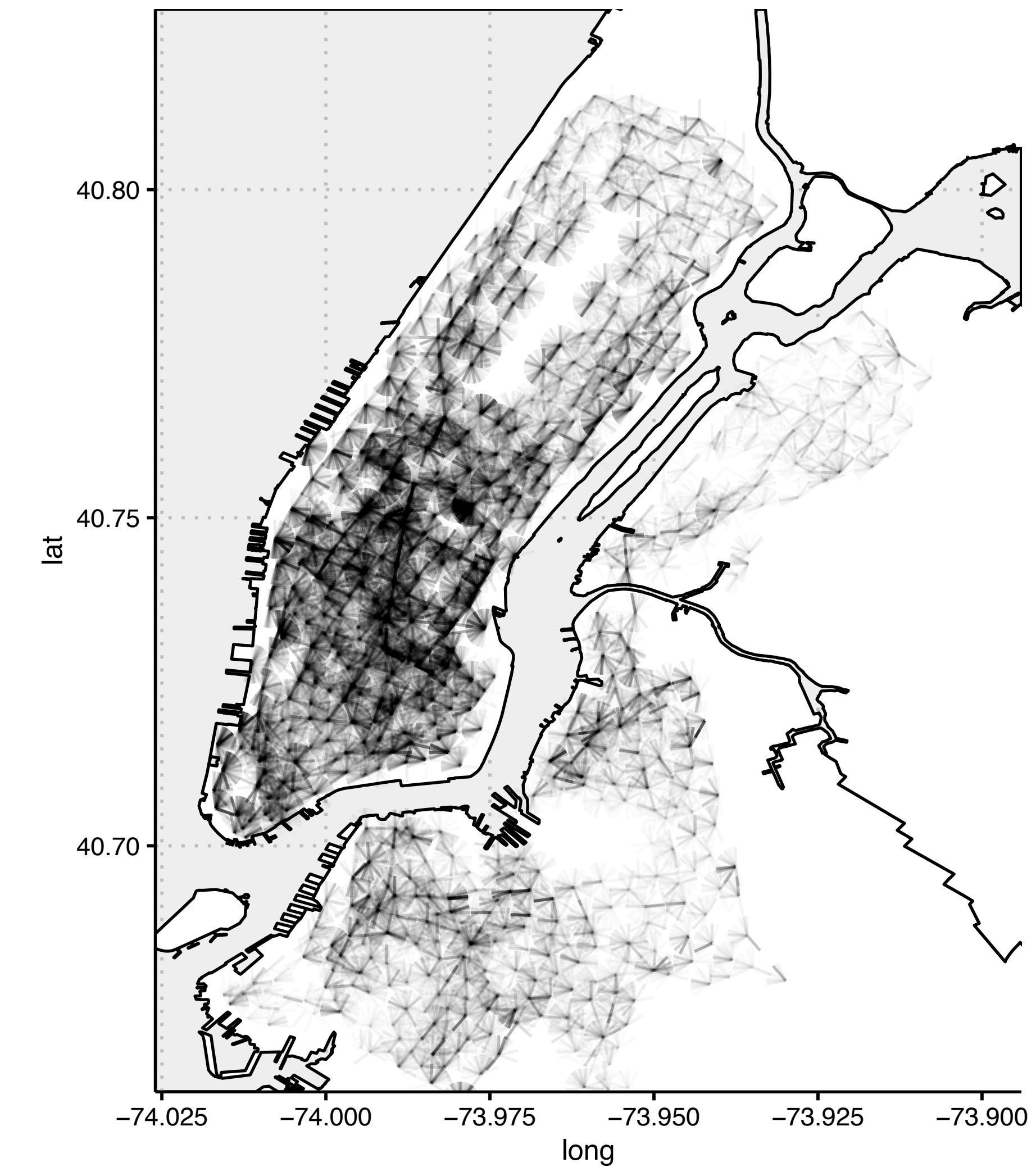
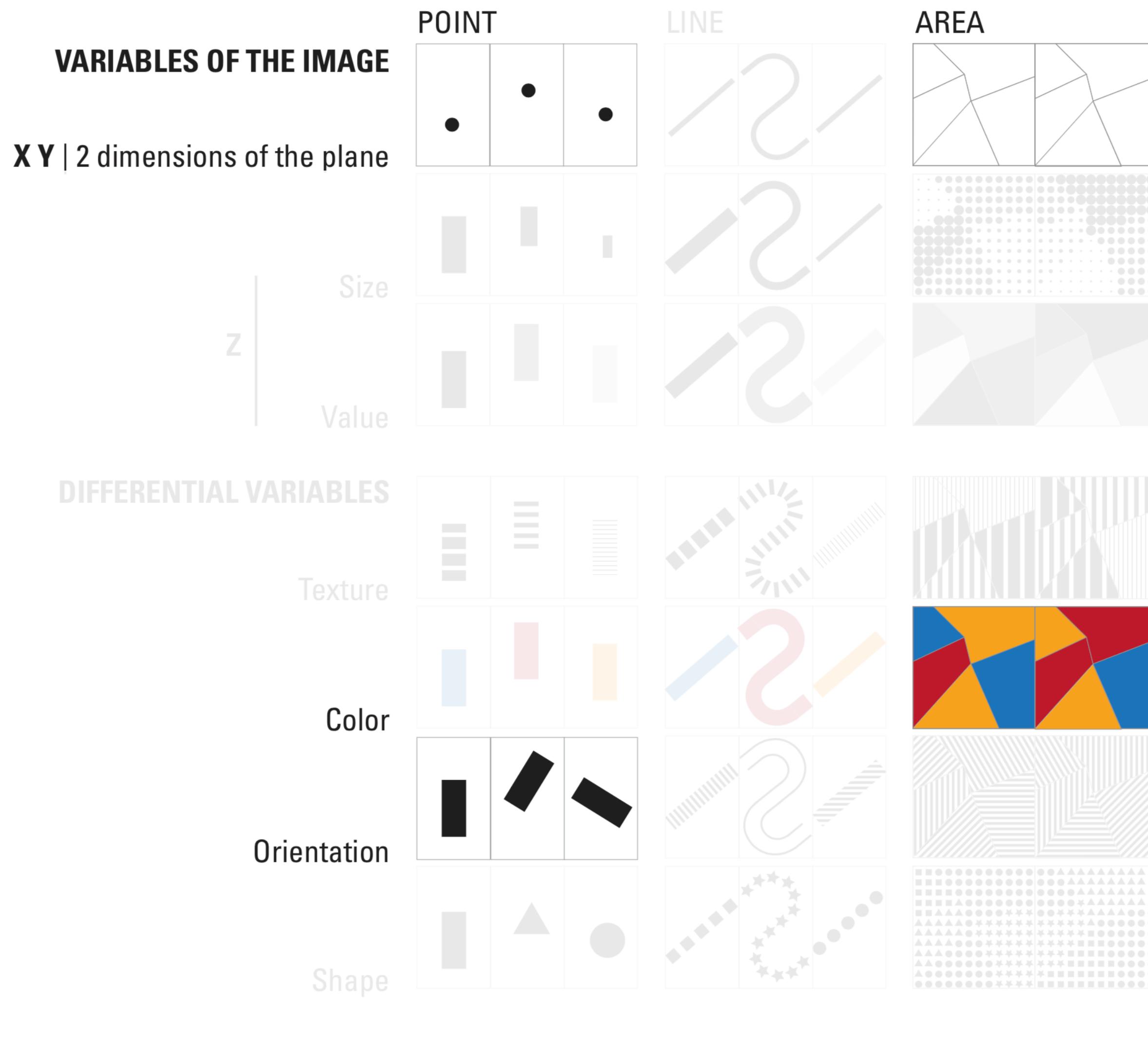
# data encodings, visual channels for encoding data

# Citi Bike example — *exploratory data analysis*



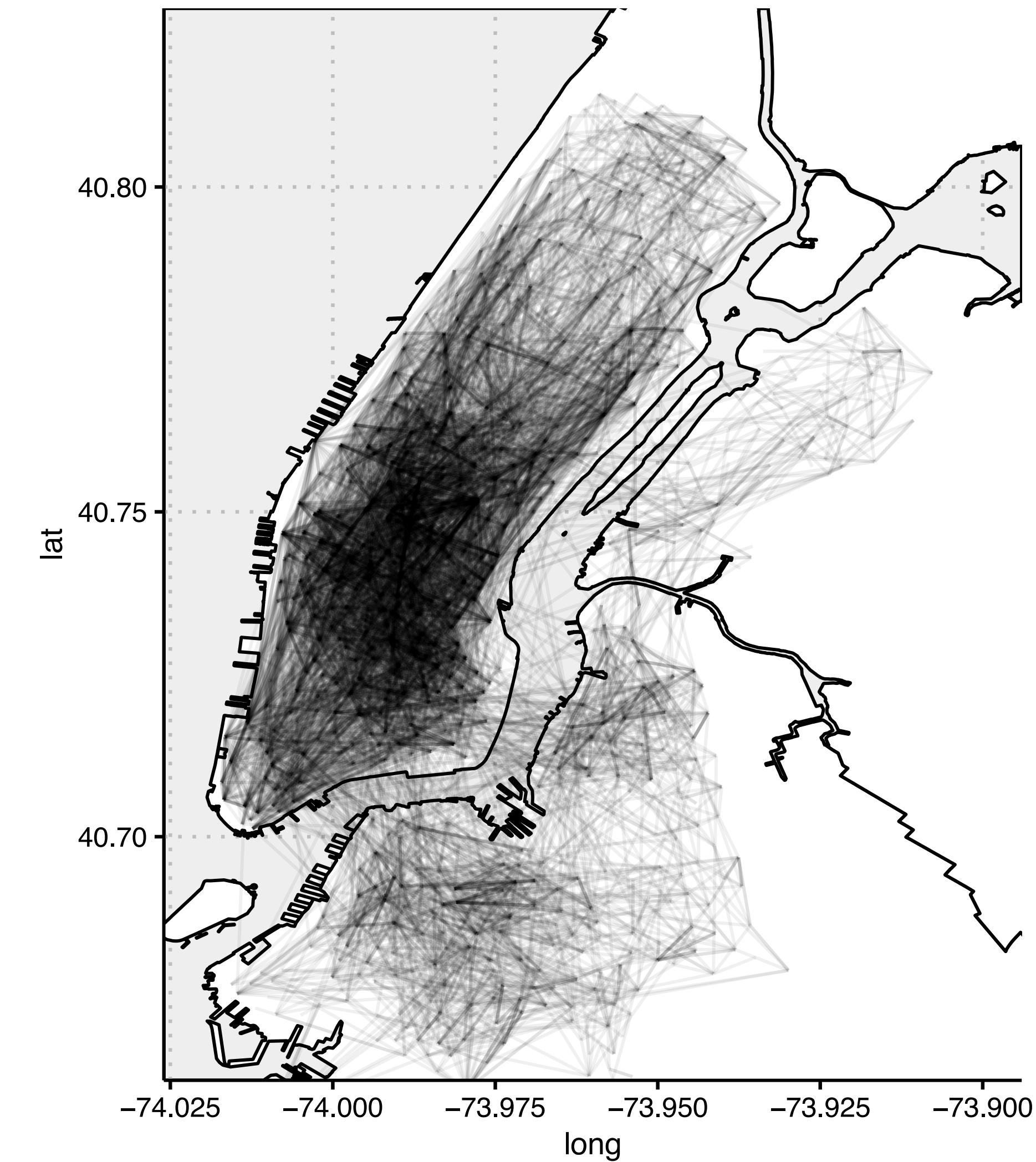
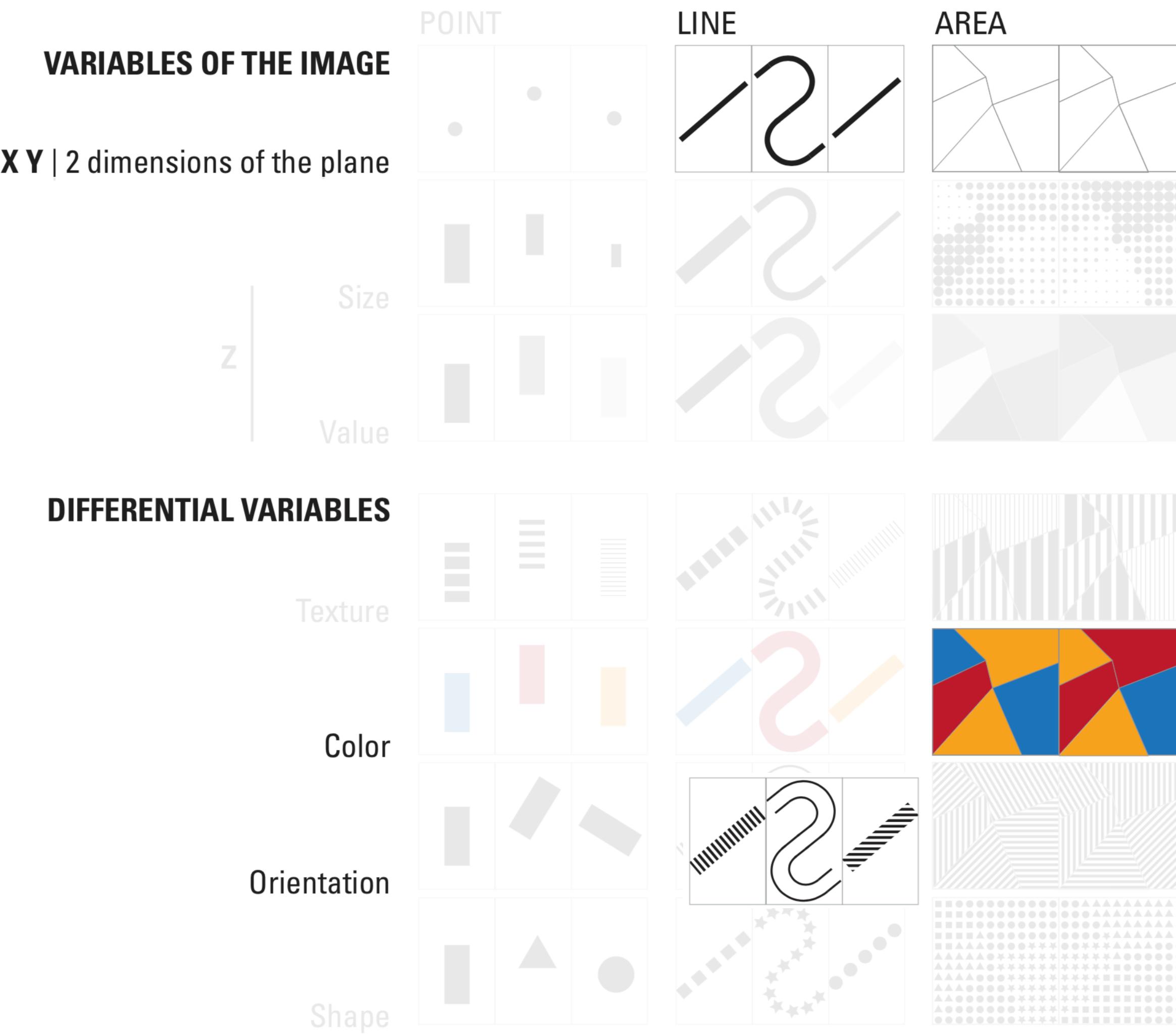
# data encodings, visual channels for encoding data

## Citi Bike example — *exploratory data analysis*



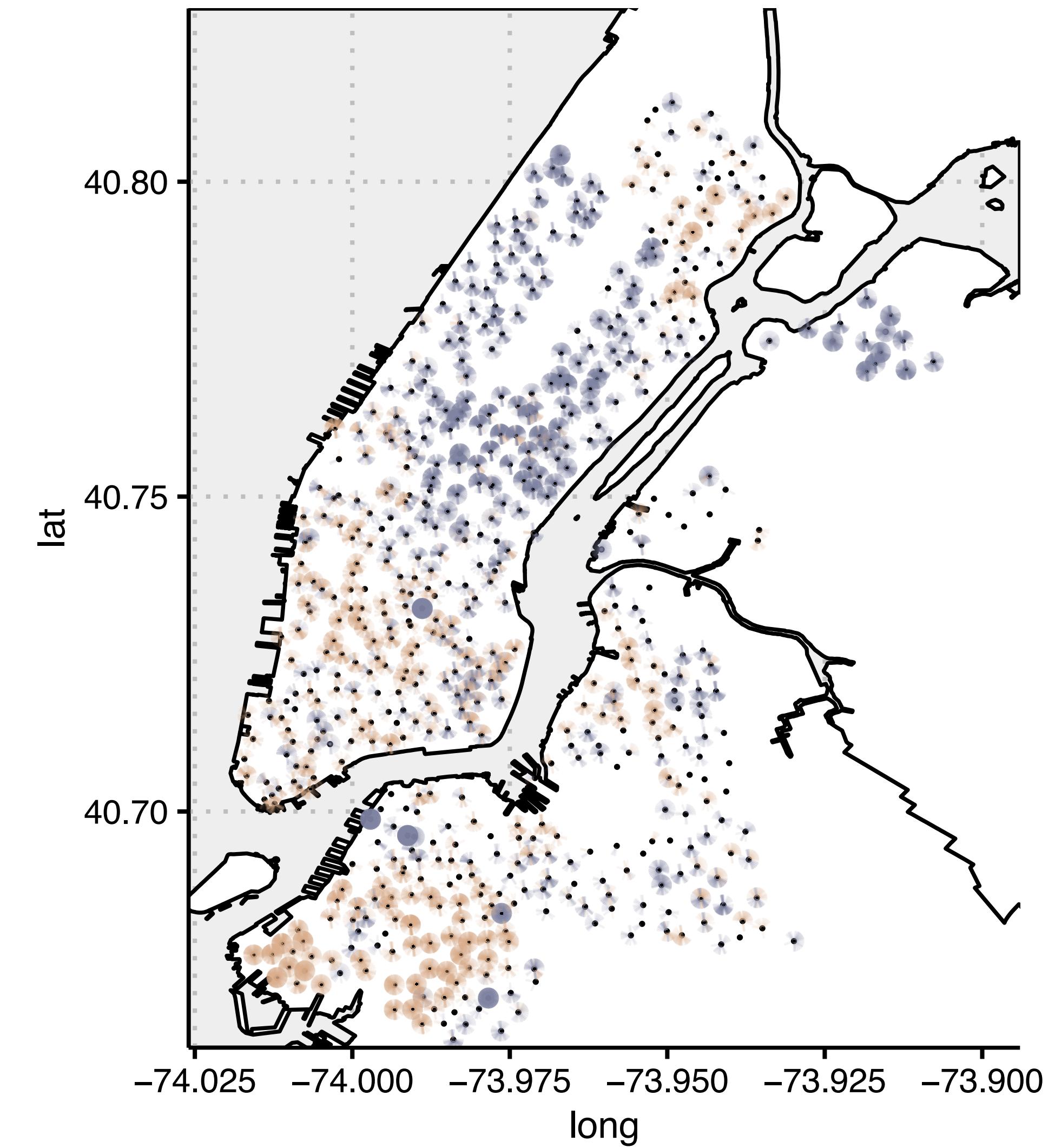
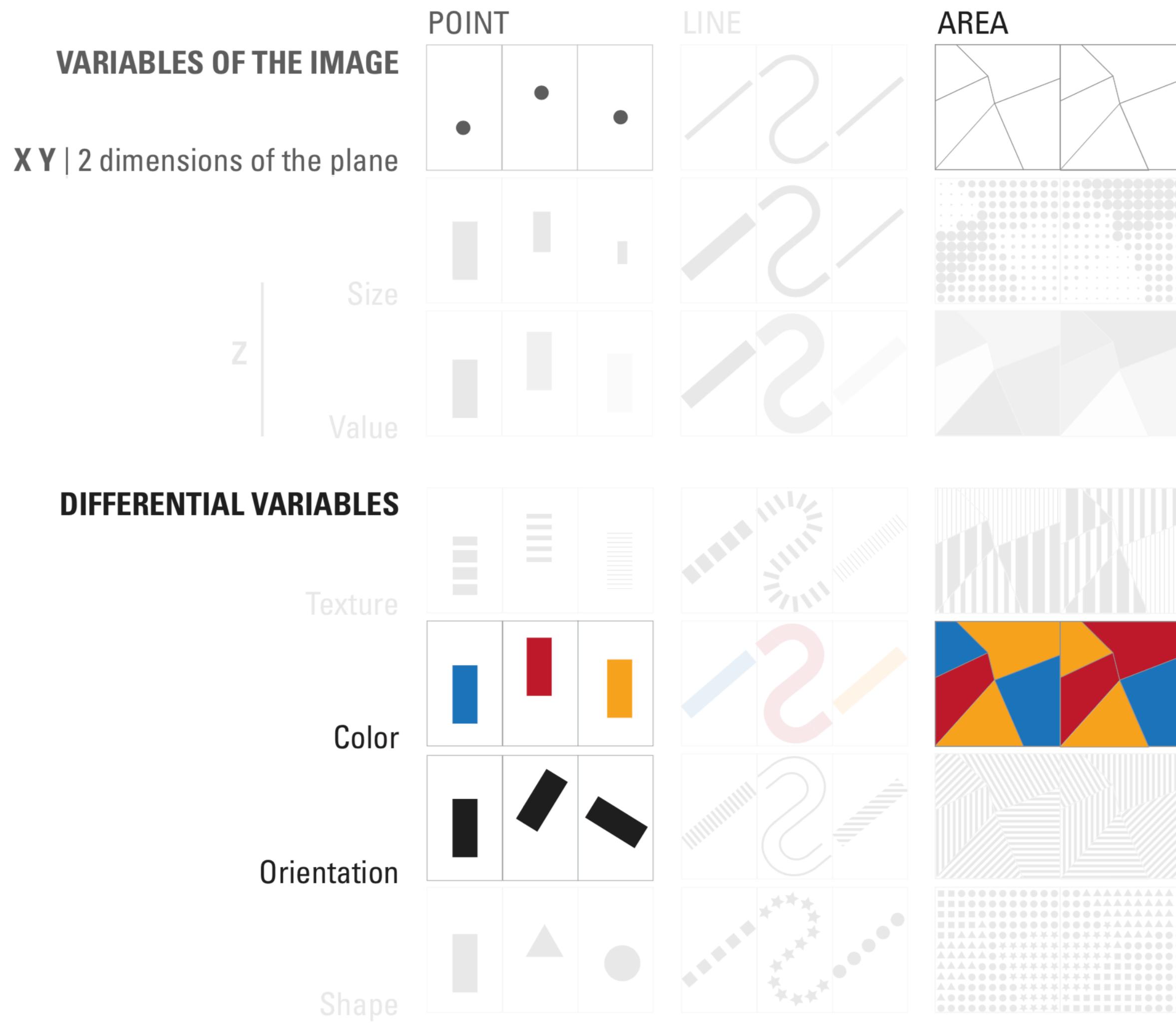
# data encodings, visual channels for encoding data

# Citi Bike example — *exploratory data analysis*



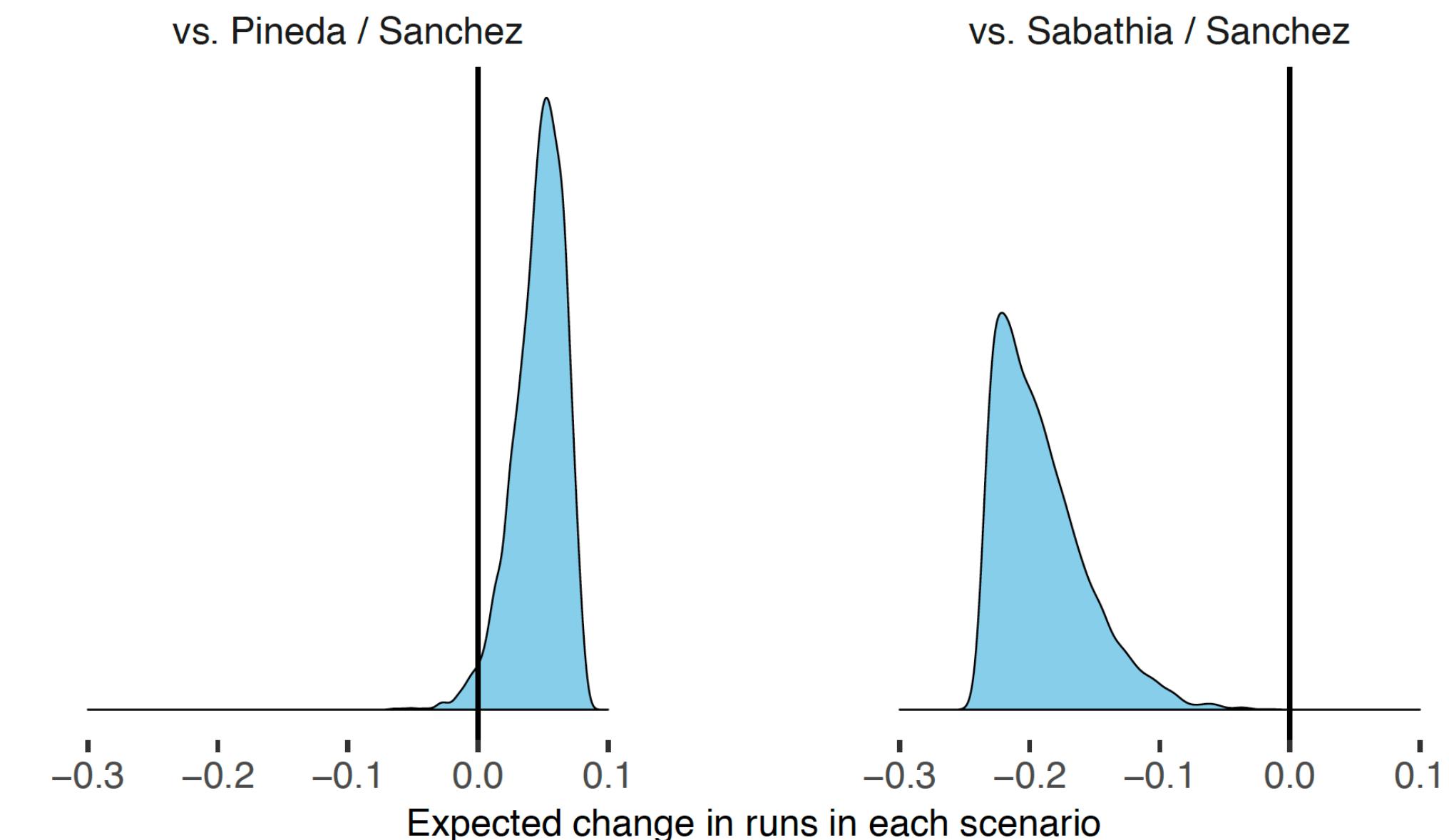
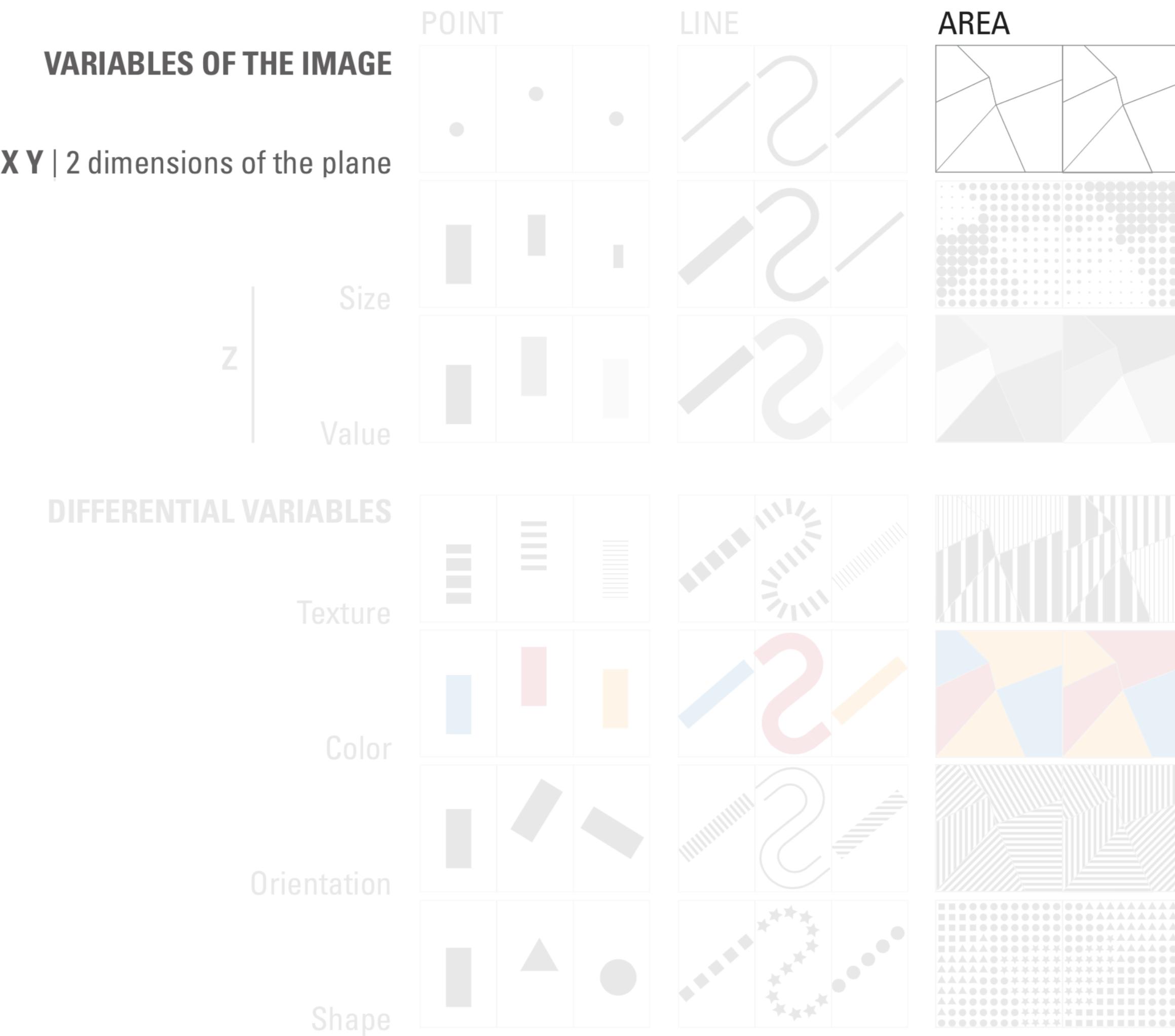
# data encodings, visual channels for encoding data

# Citi Bike example — *exploratory data analysis*



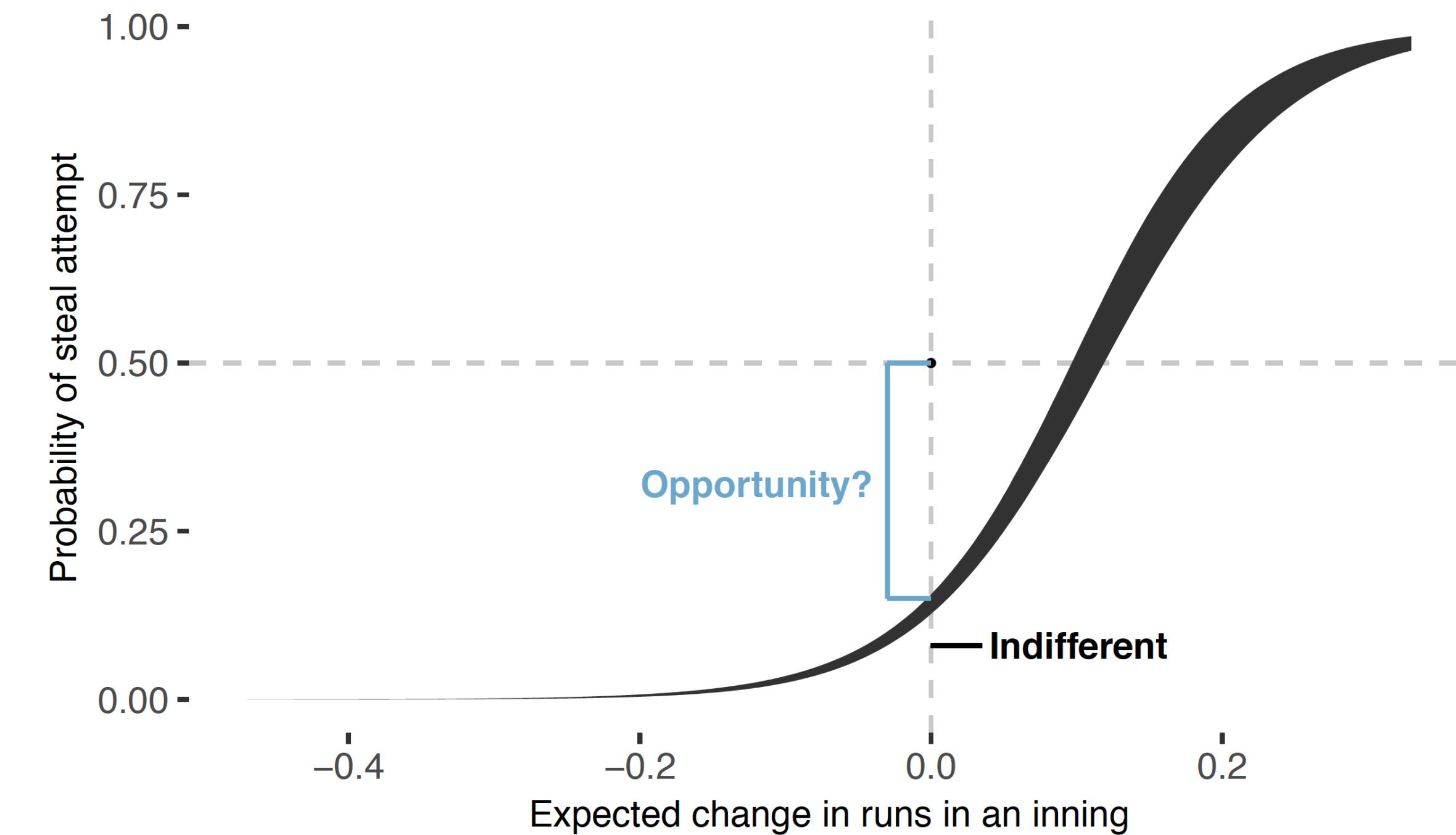
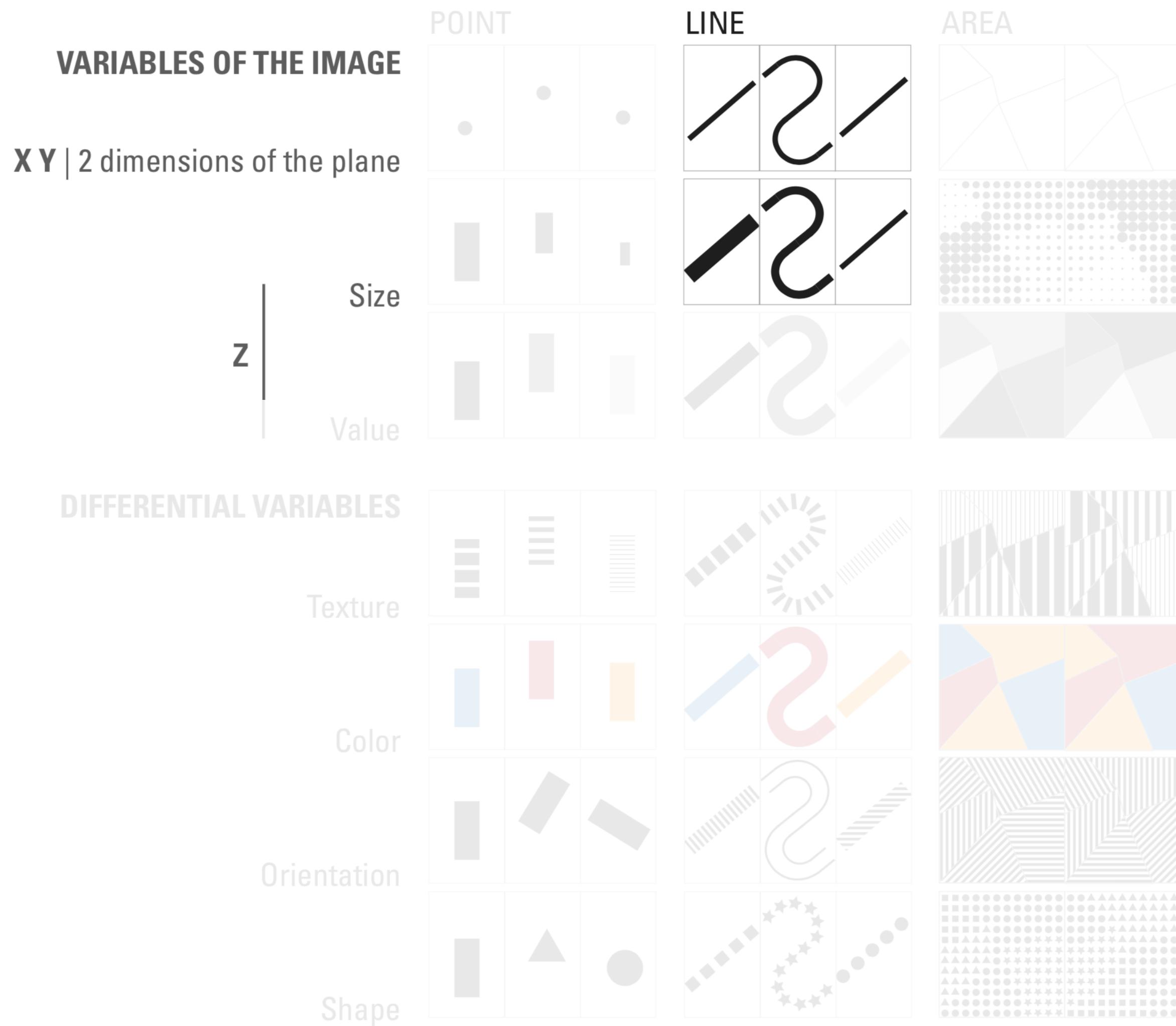
# data encodings, visual channels for encoding data

# Dodgers draft proposal example



# data encodings, visual channels for encoding data

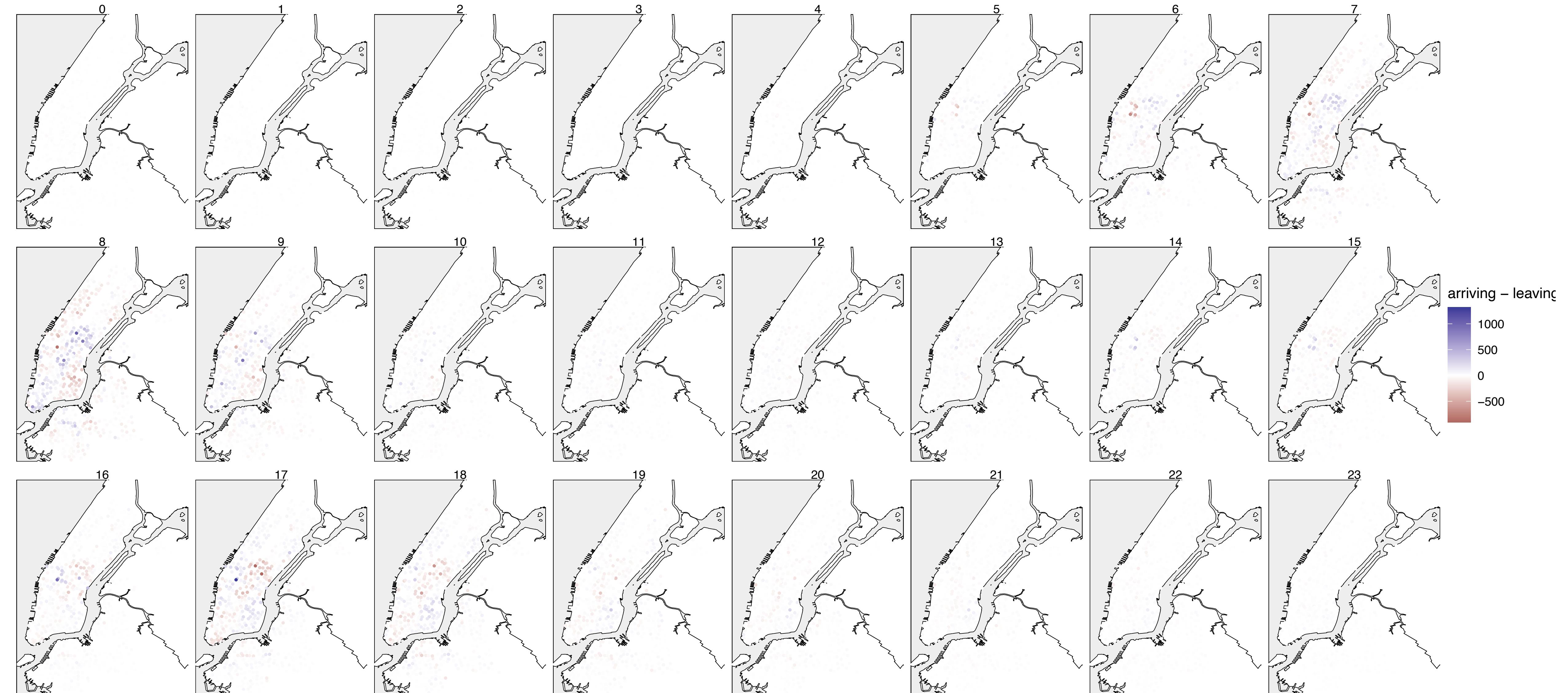
# Dodgers *draft* proposal example



**adding dimensions through small multiples**

# data encodings, small multiples (of area + color + point + value)

# Citi Bike example — *exploratory data analysis*



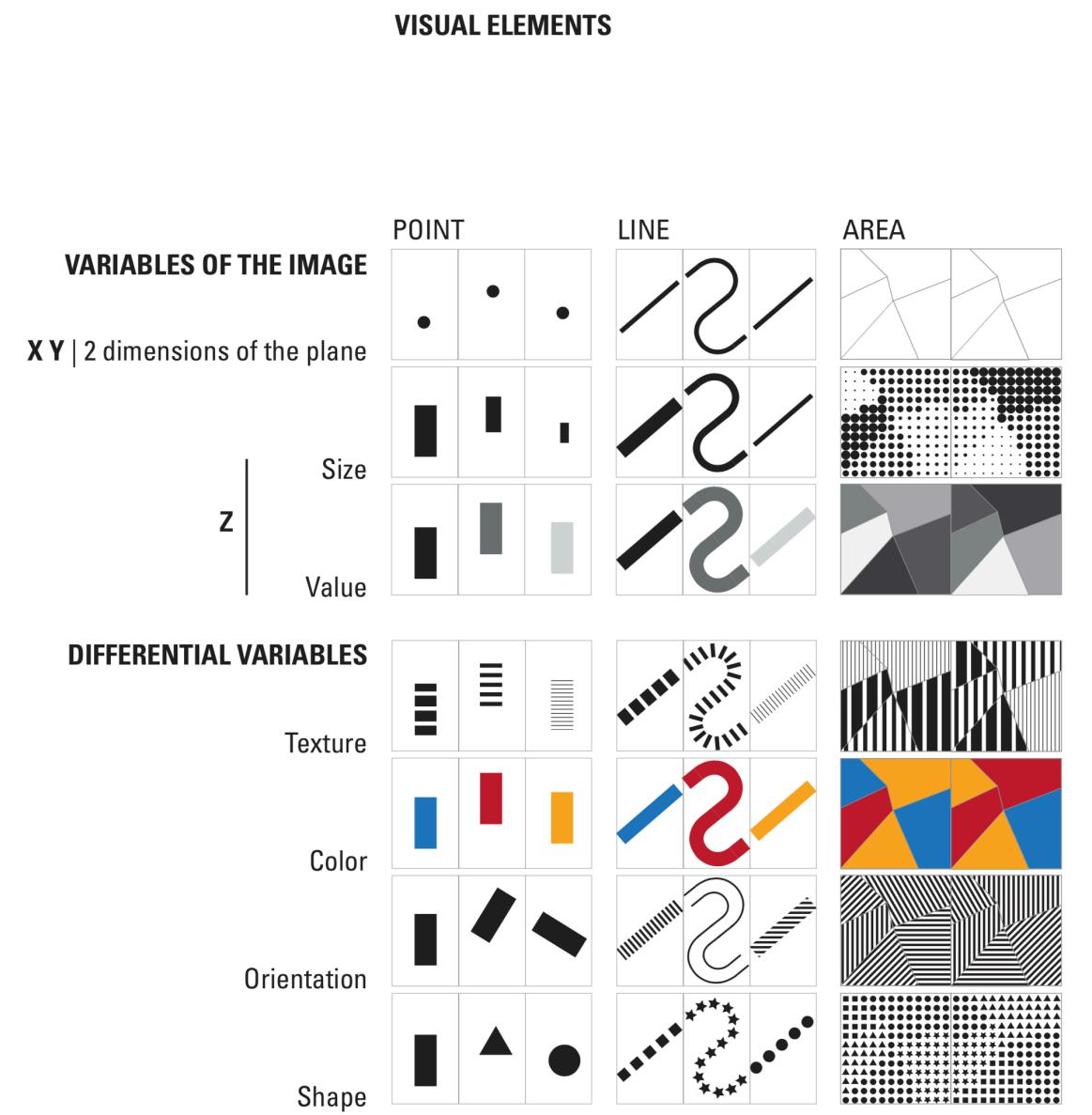
## **class exercises**

# exercise, identify data encodings in visual channels

# Please approve the hire of 2 FTEs

to backfill those who quit in the past year

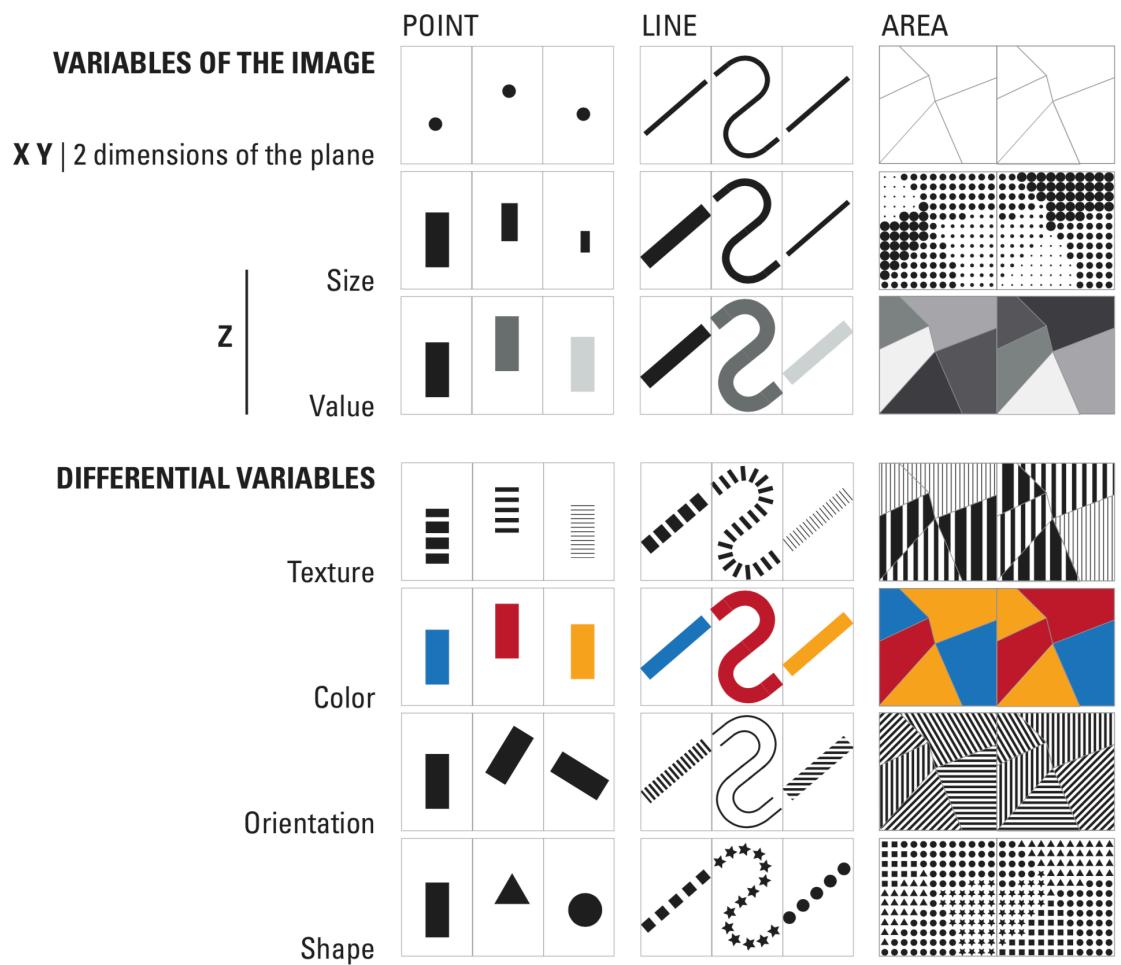
## Ticket volume over time



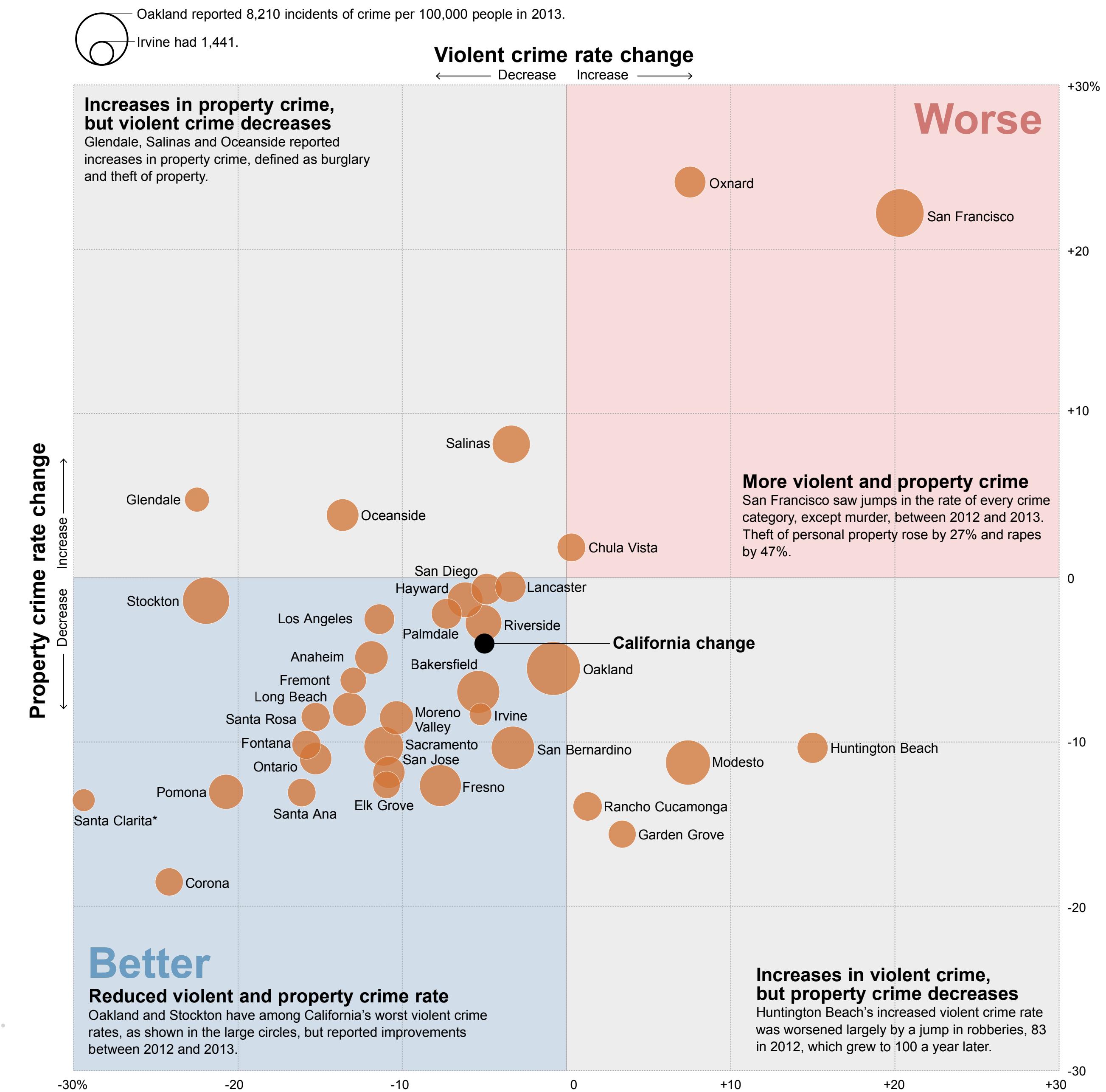
Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

# exercise, identify data encodings in visual channels

## VISUAL ELEMENTS

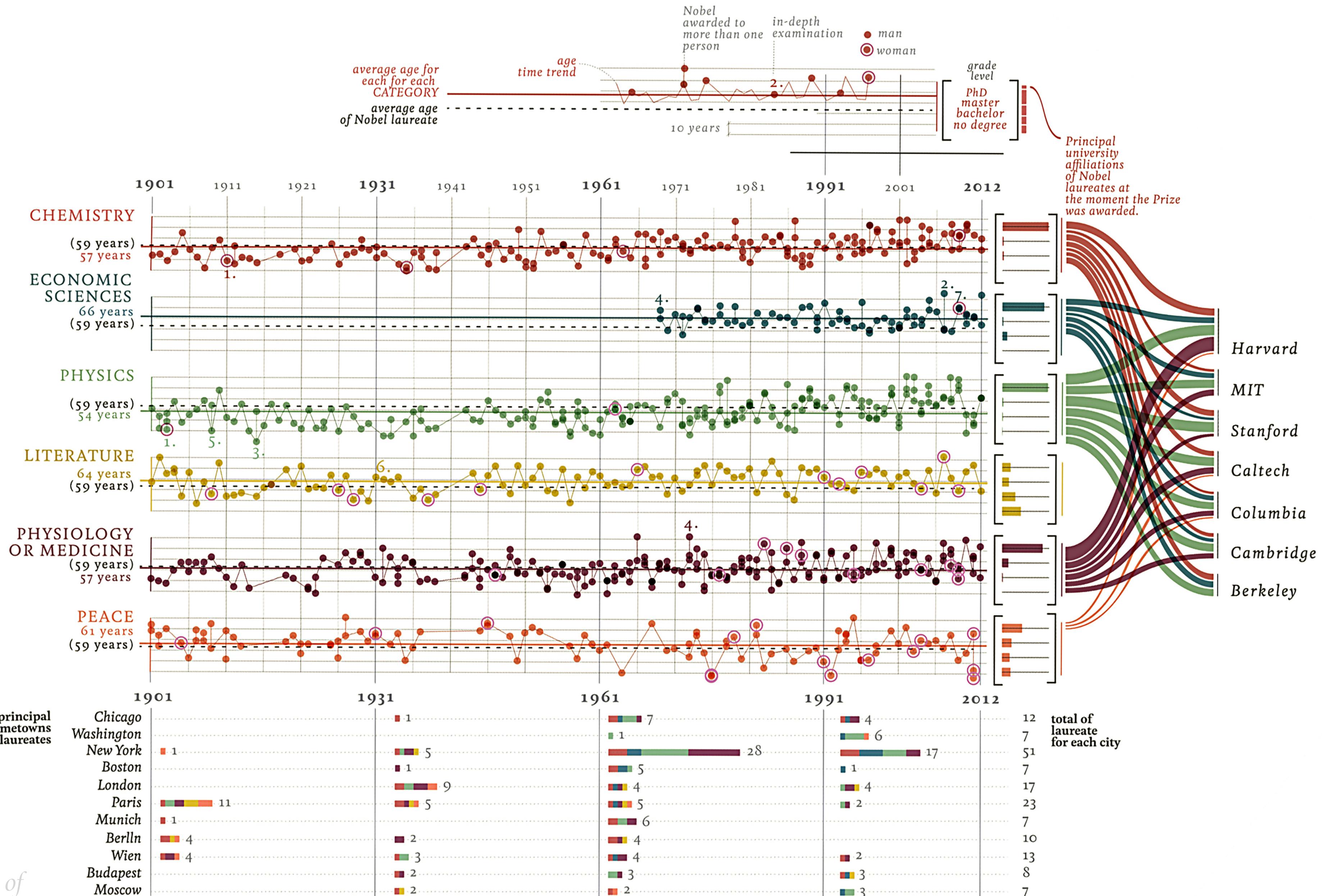
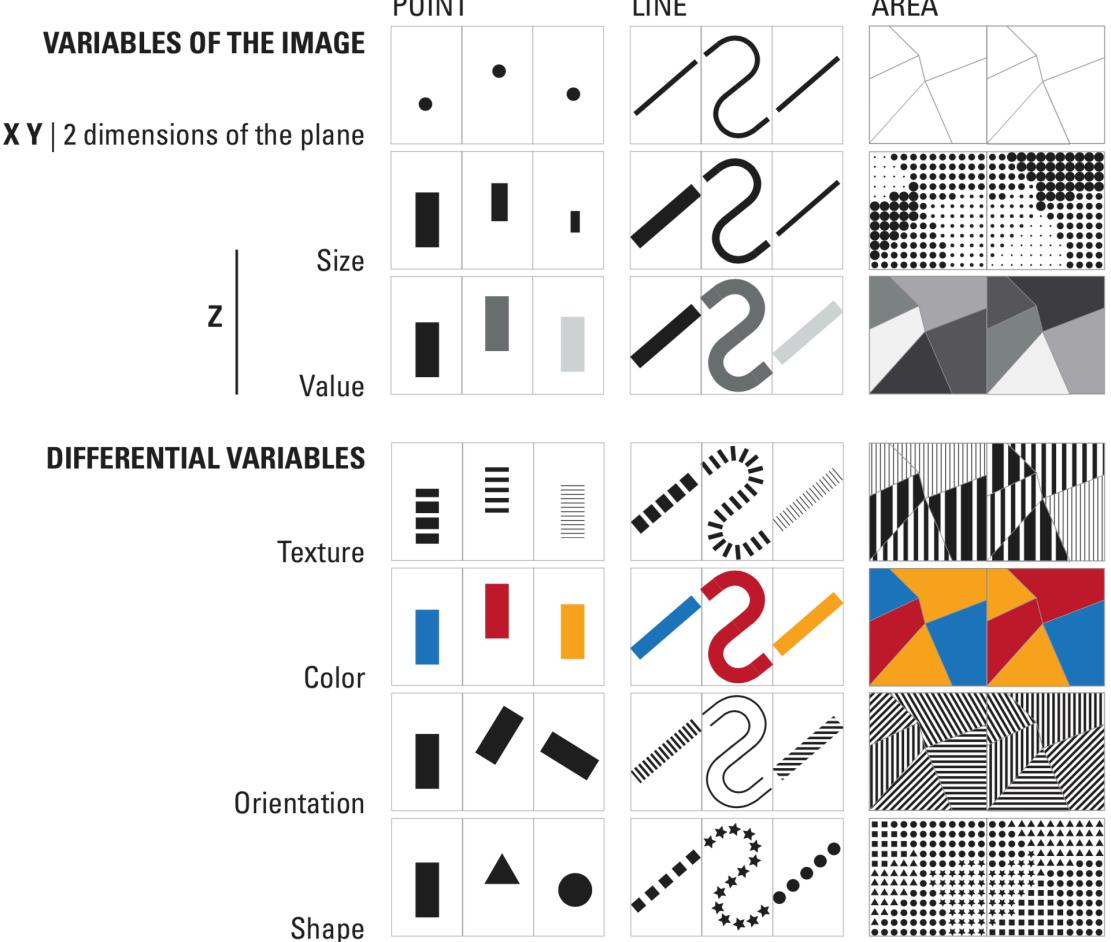


Schleuss, Jon, and Rong-Cong Lin II. 2013.  
“California Crime 2013.” Los Angeles Times.



# exercise, identify data encodings in visual channels

## VISUAL ELEMENTS

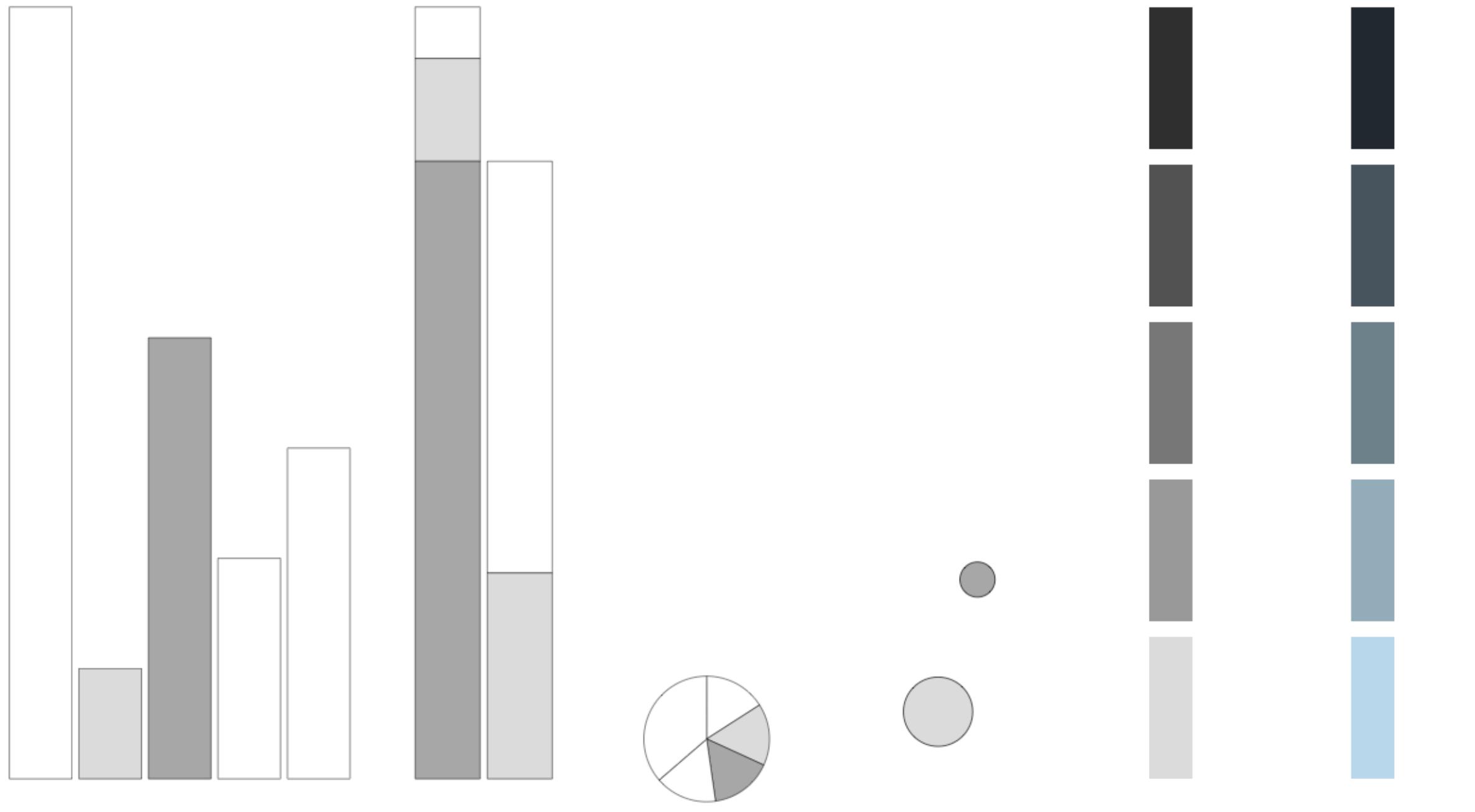


Spencer, Scott. *Approximating the Components of Lupi's Nobels, No Degrees*. March 15, 2019. <https://ssp3nc3r.github.io/post/approximating-the-components-of-lupi-s-nobel-no-degrees/>.

**channel effectiveness for encoding data**

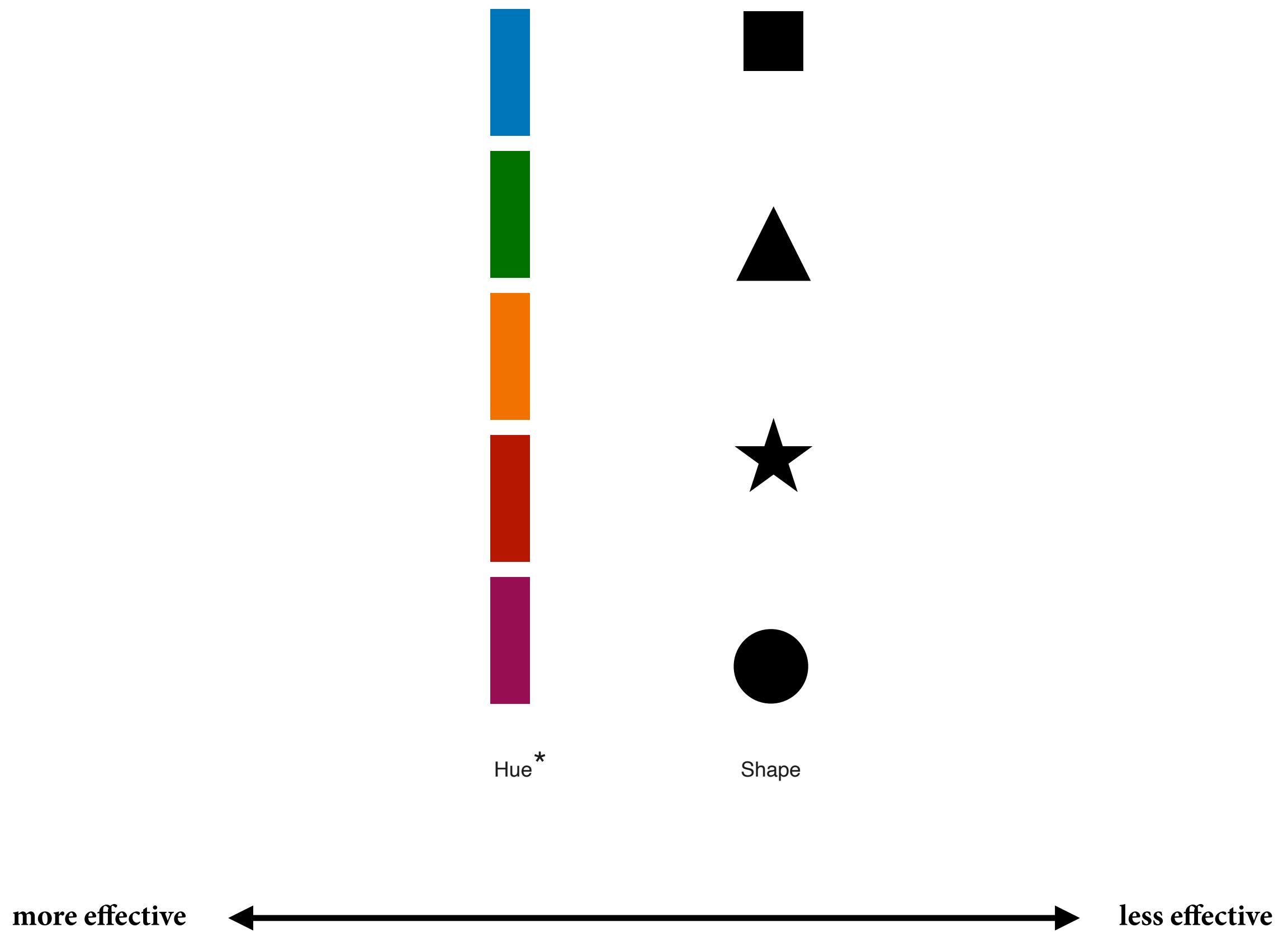
# general channel effectiveness, encoding data

## ratio, interval, and ordered



more effective ← → less effective

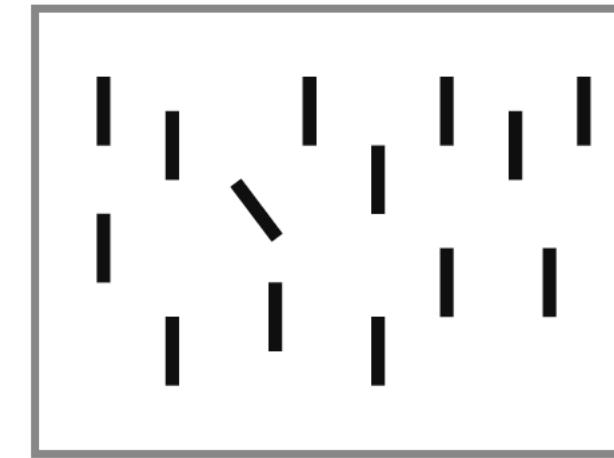
## categorical



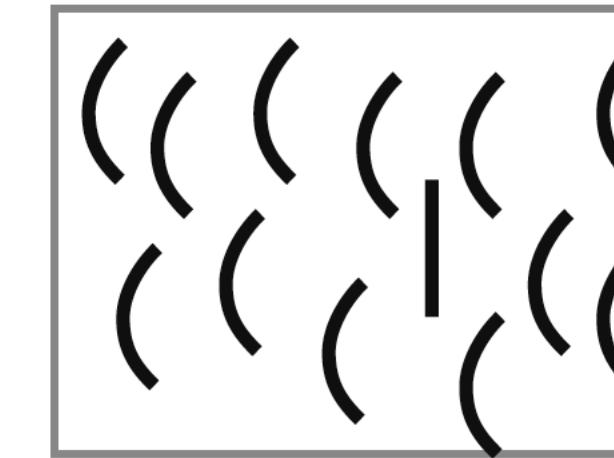
**perceptual psychology**

# perceptual psychology, *pre-attentive attributes*

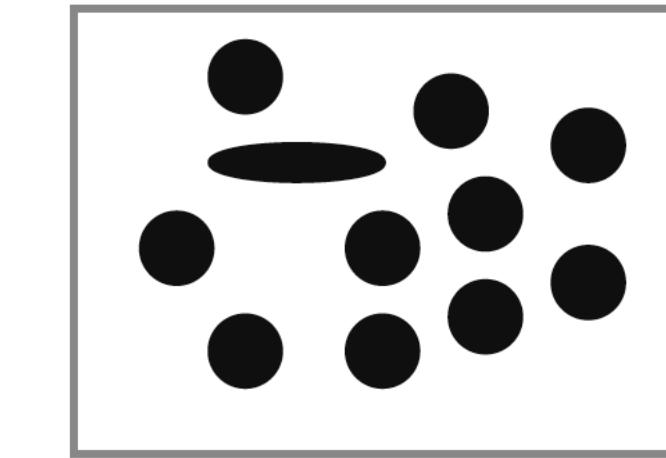
Orientation



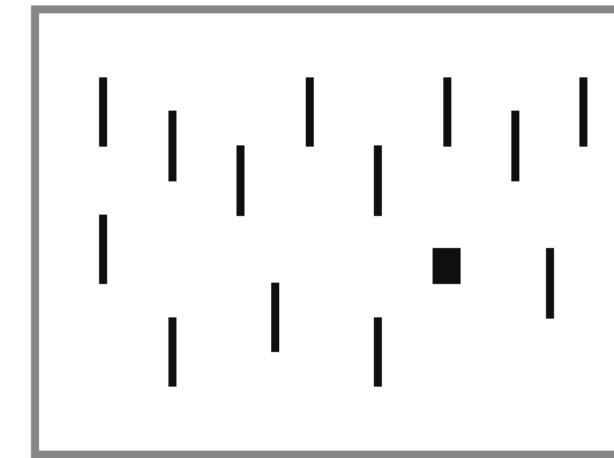
Curved straight



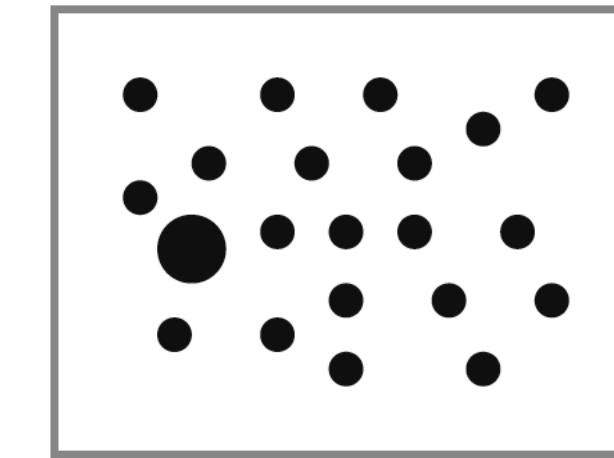
Shape



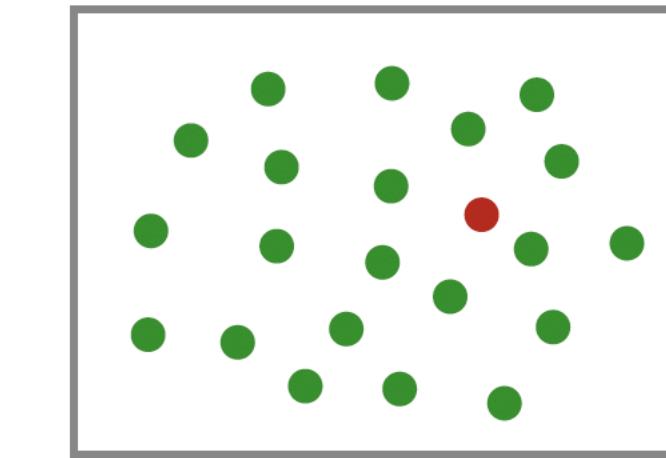
Shape



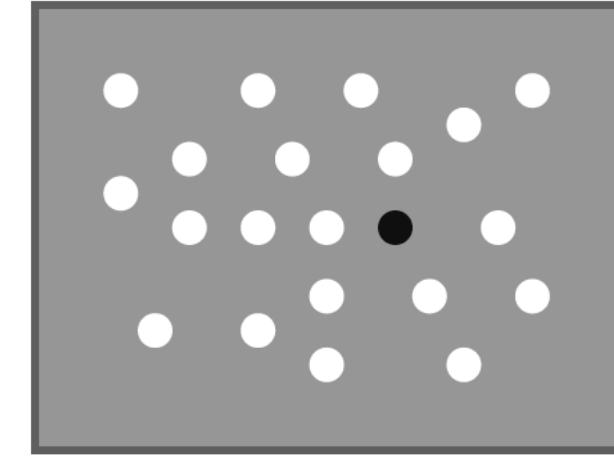
Size



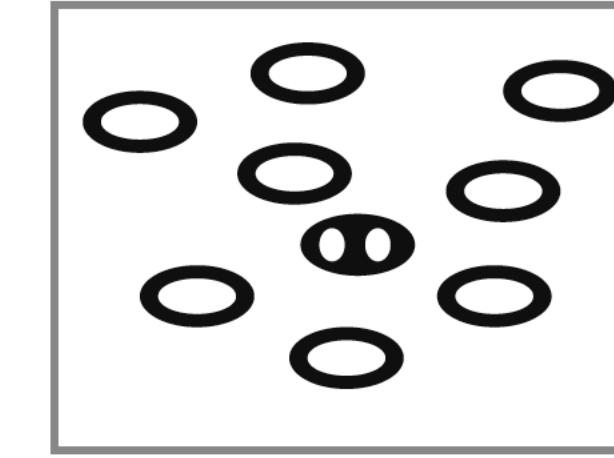
Color



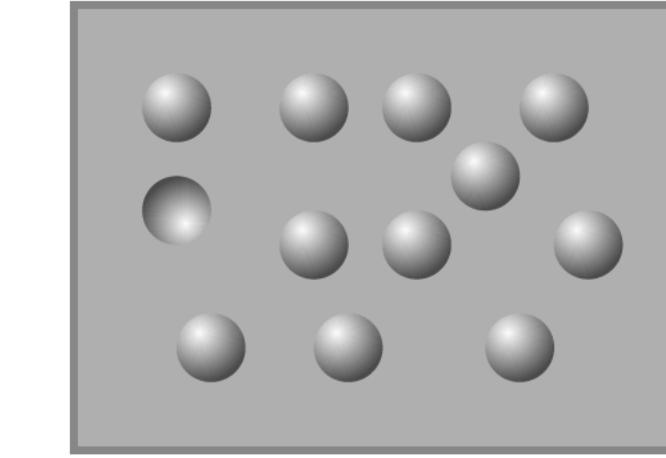
Light/dark



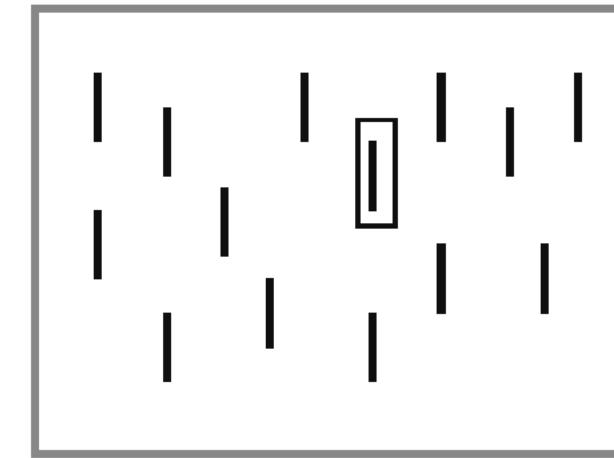
Topology (or count)



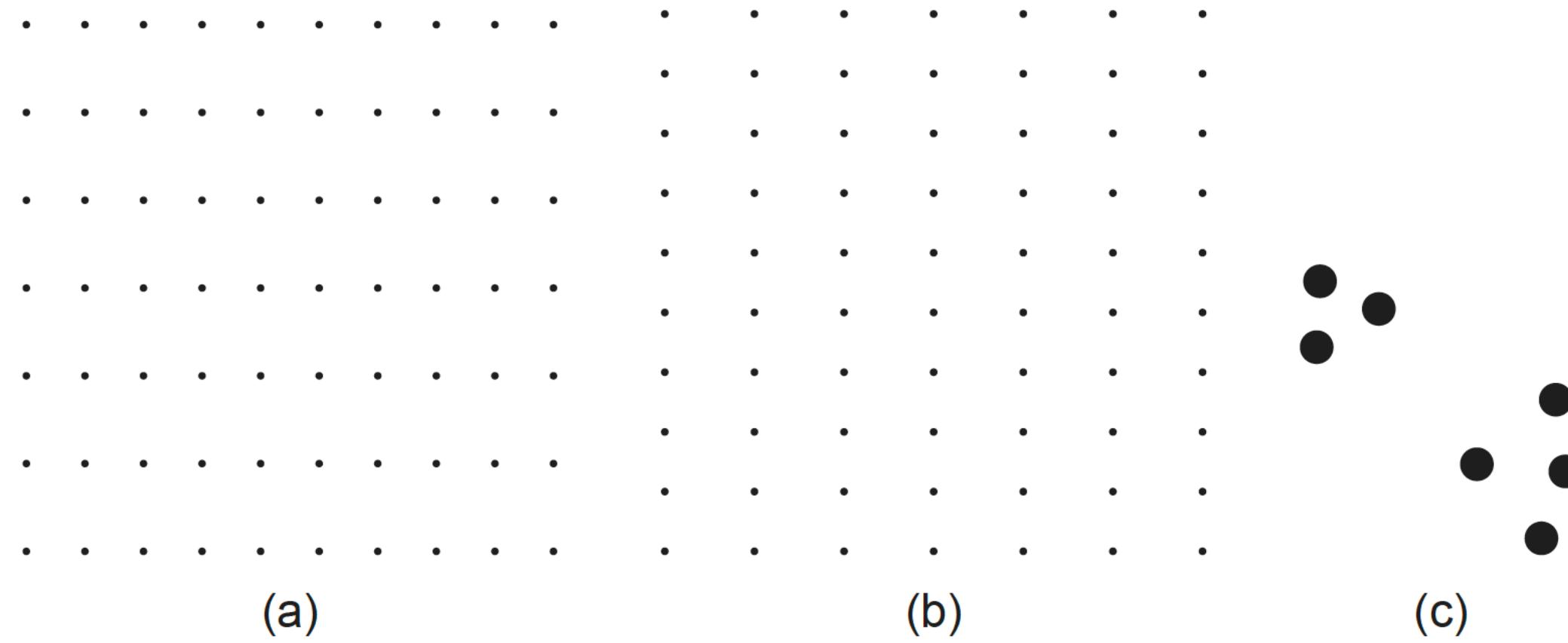
Convex/concave



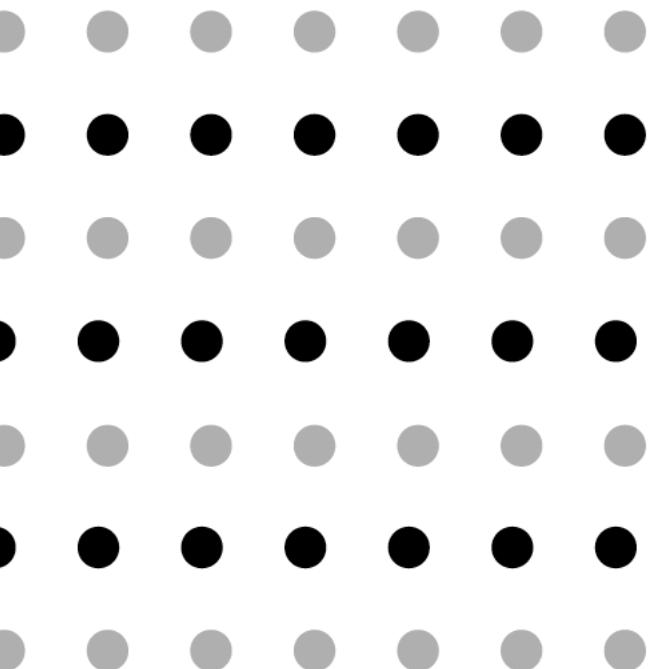
Addition



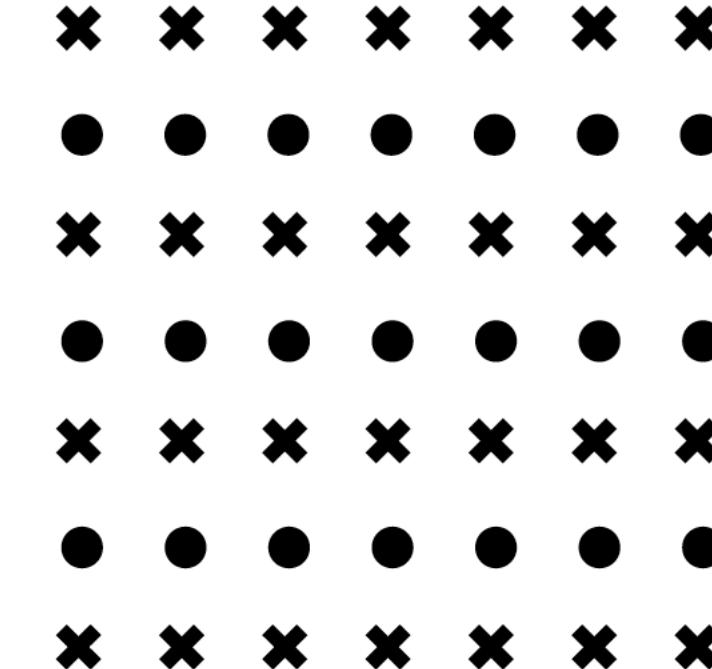
# perceptual psychology, Gestalt principles, *proximity*



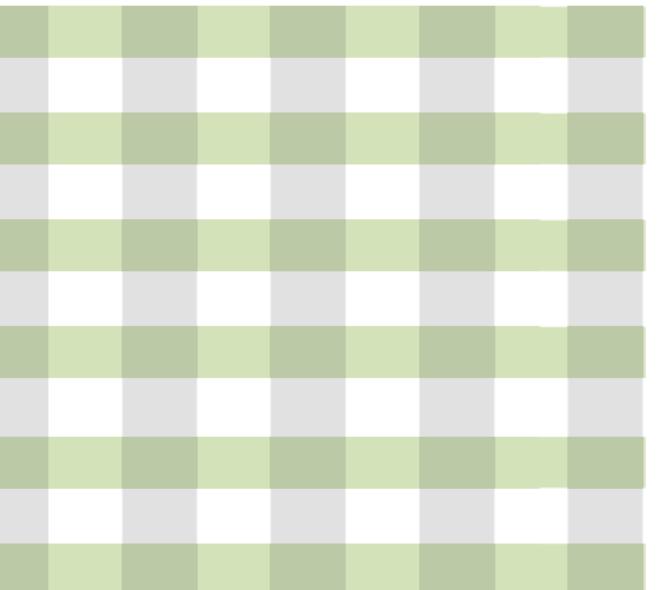
# perceptual psychology, Gestalt principles, *similarity*



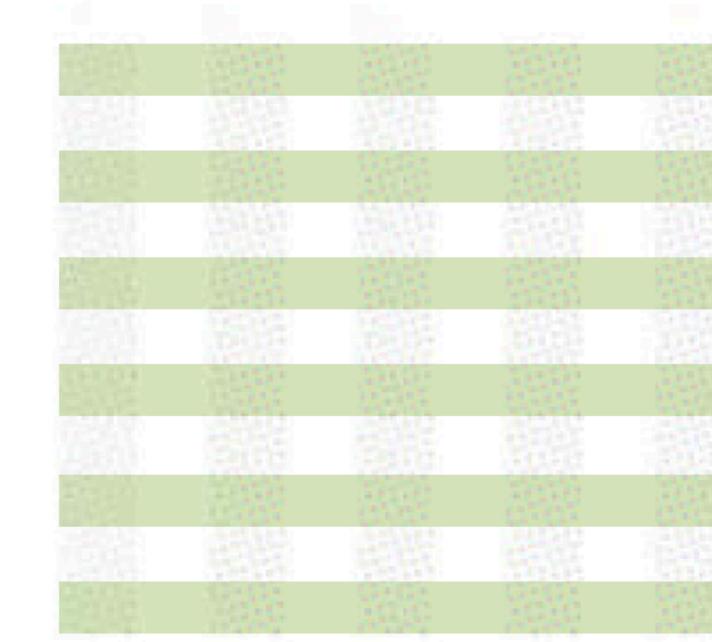
(a)



(b)

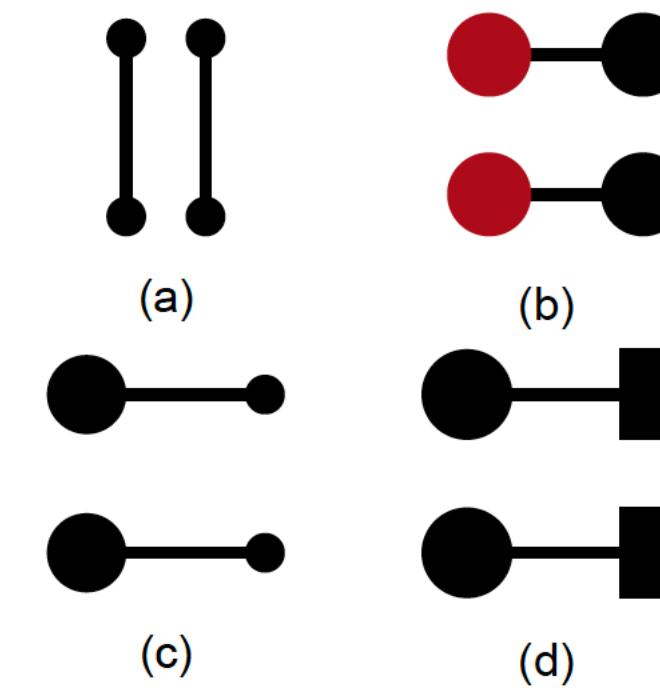


(c)

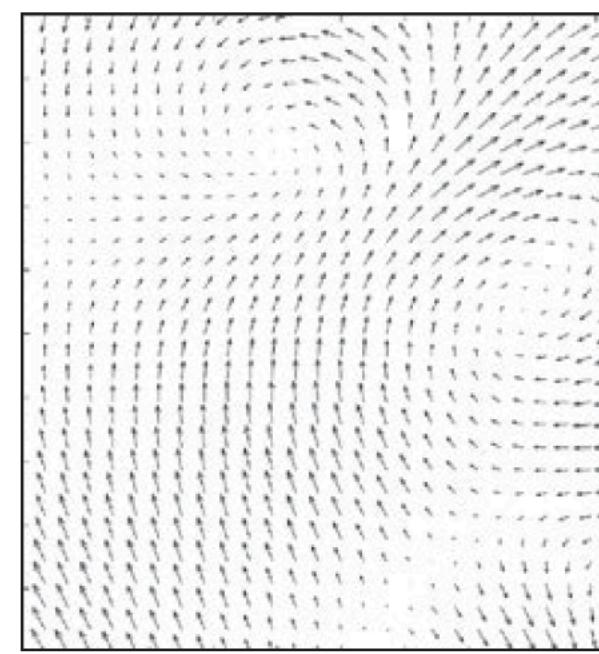


(d)

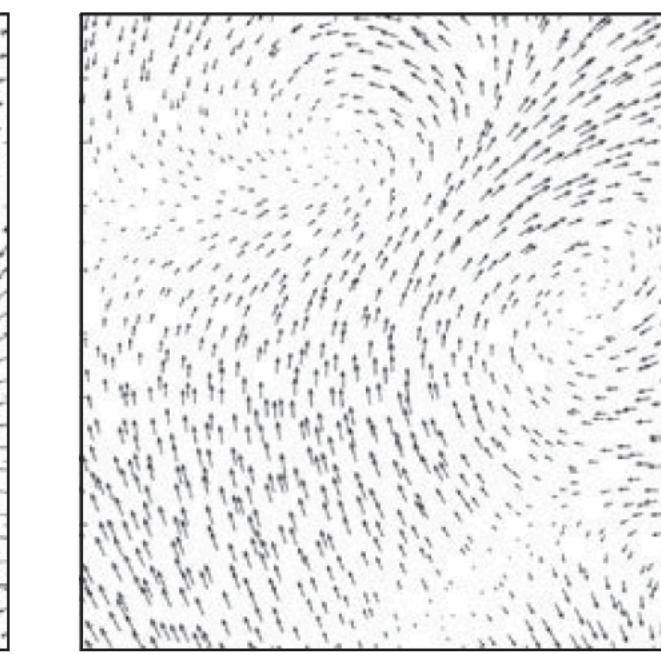
# perceptual psychology, Gestalt principles, *connectedness*



# perceptual psychology, Gestalt principles, *orientation*, *magnitude*, *direction*



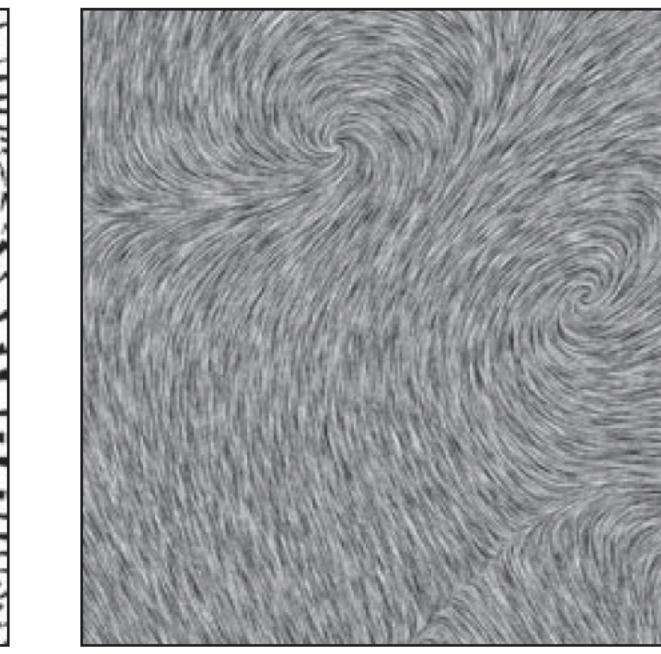
(a)



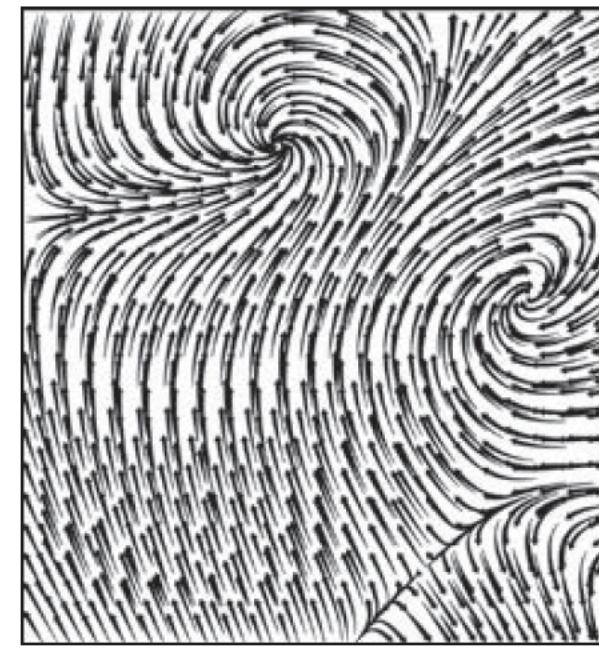
(b)



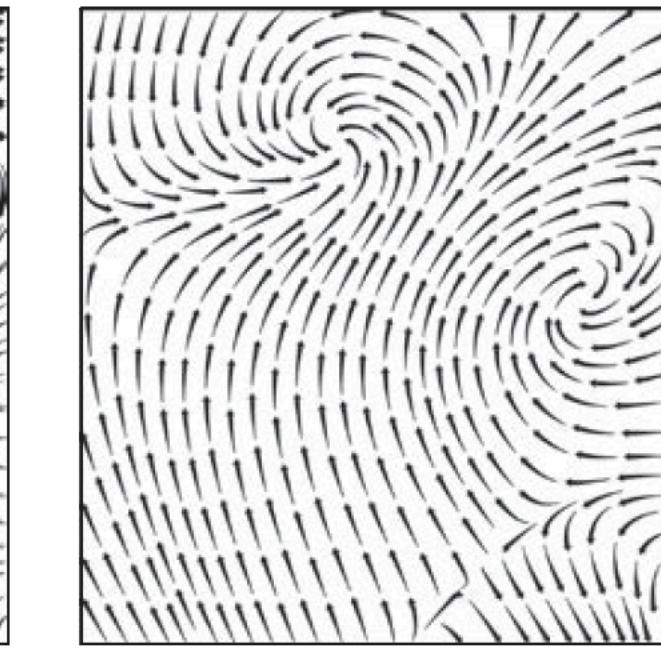
(c)



(d)



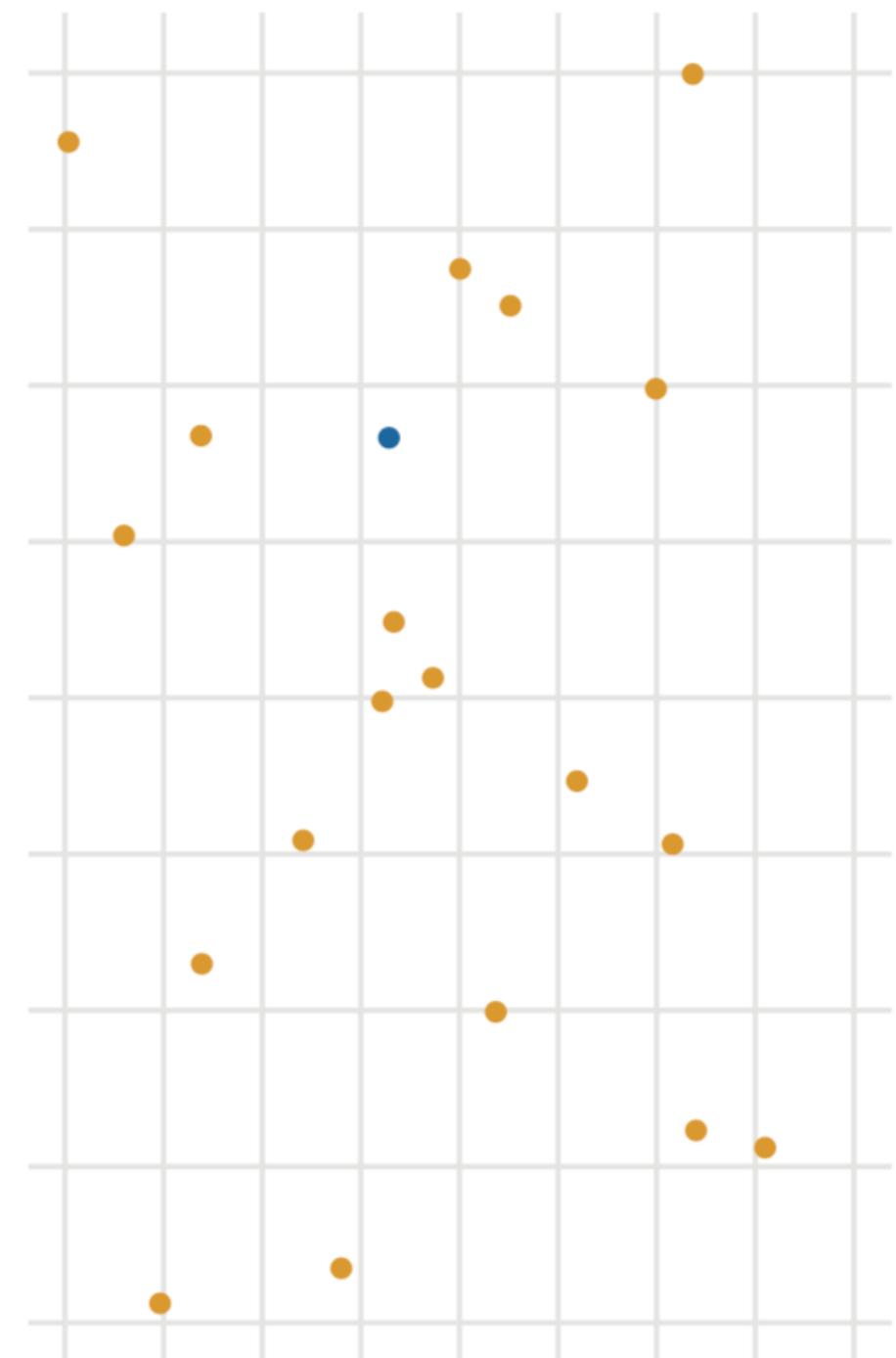
(e)



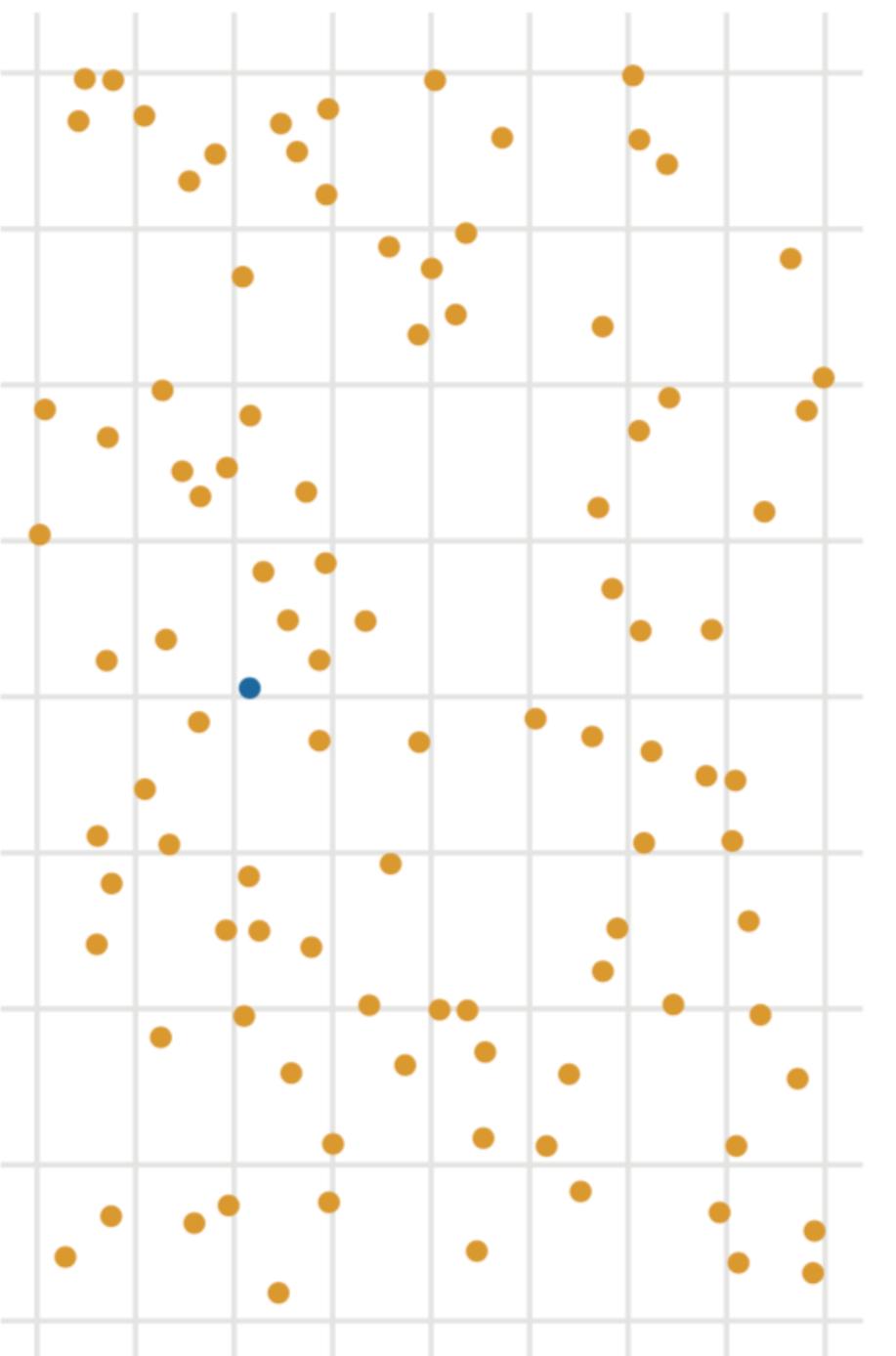
(f)

# perceptual psychology, example — *focusing visual attention*

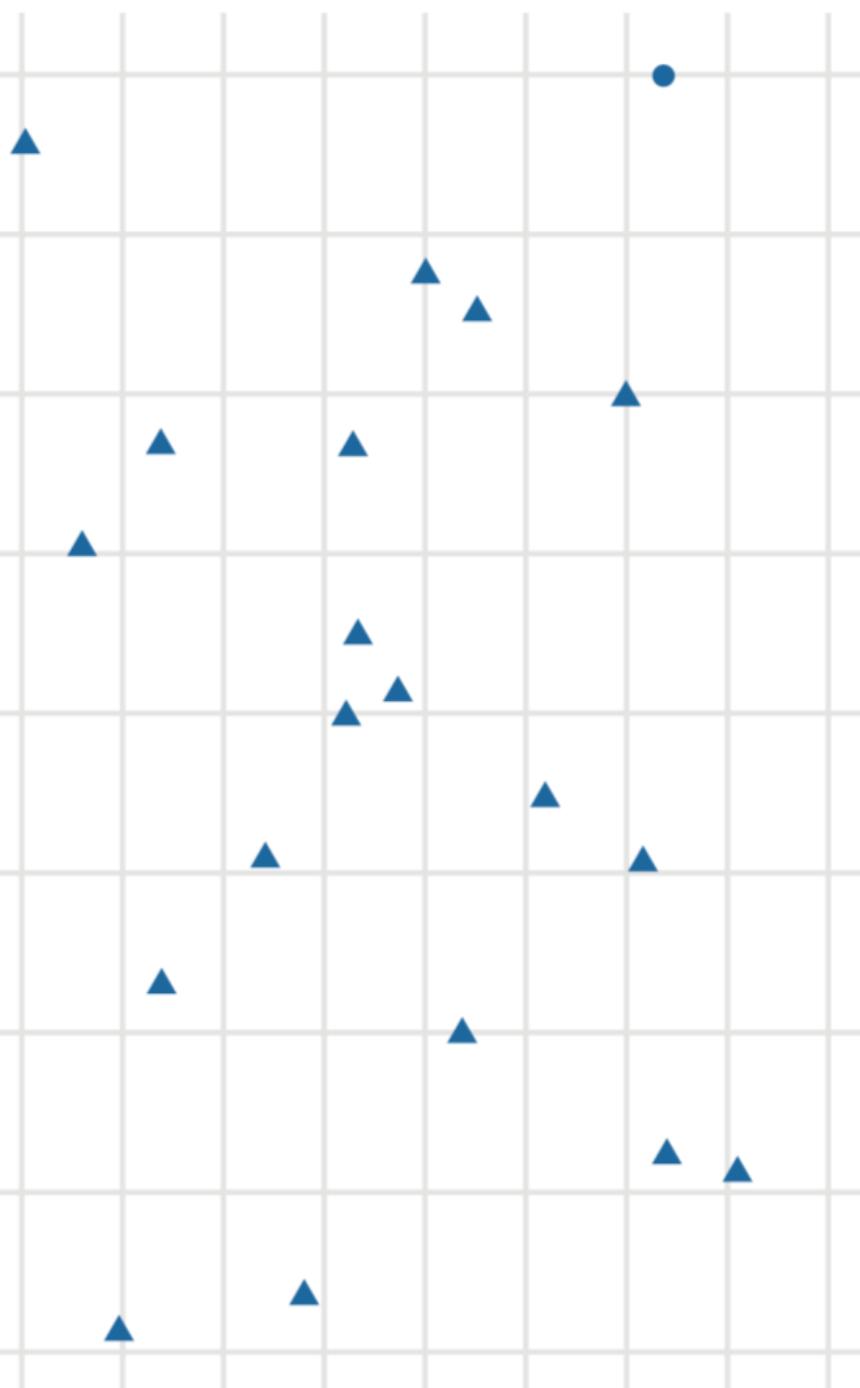
Color only,  $N = 20$



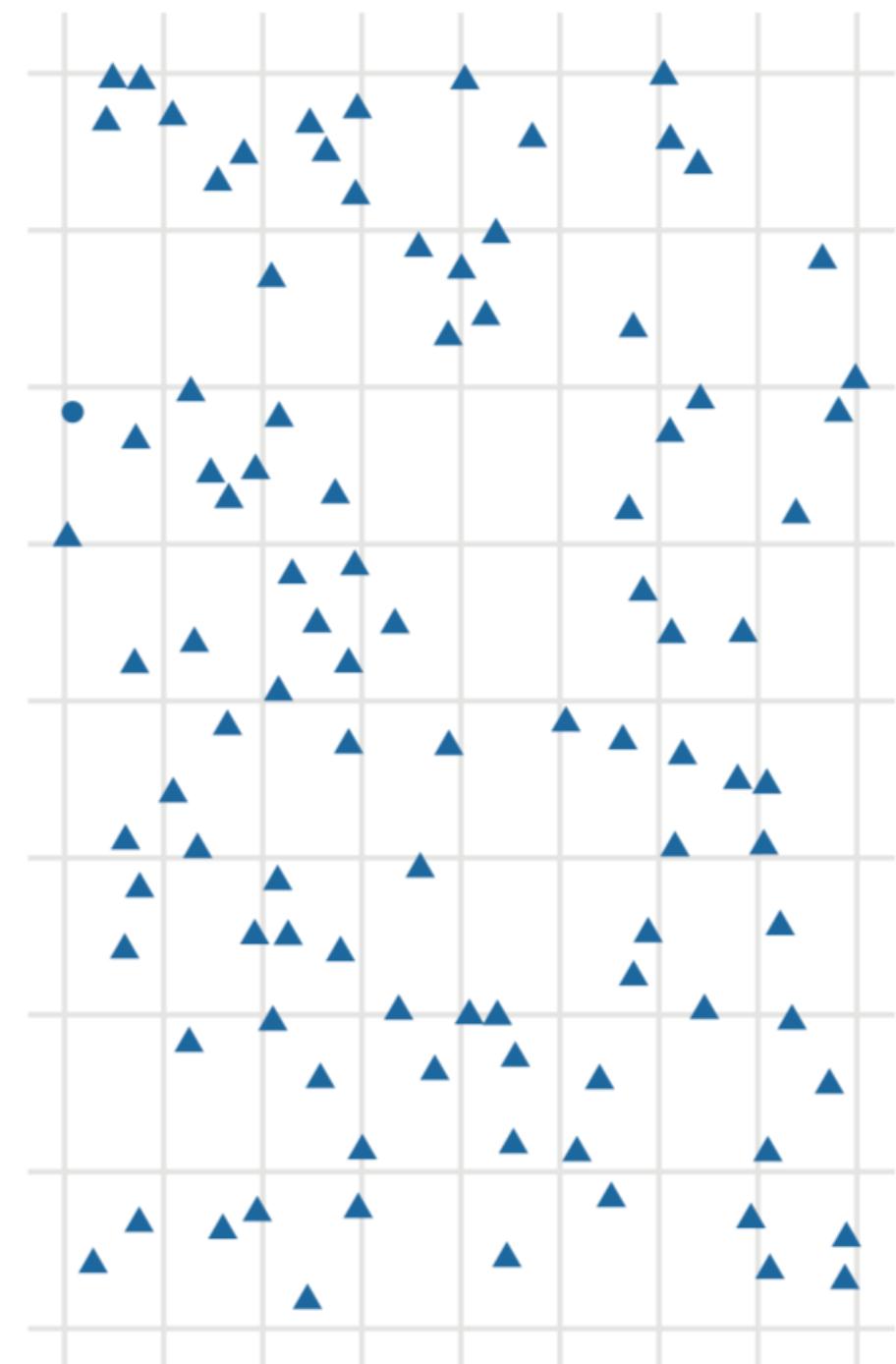
Color only,  $N = 100$



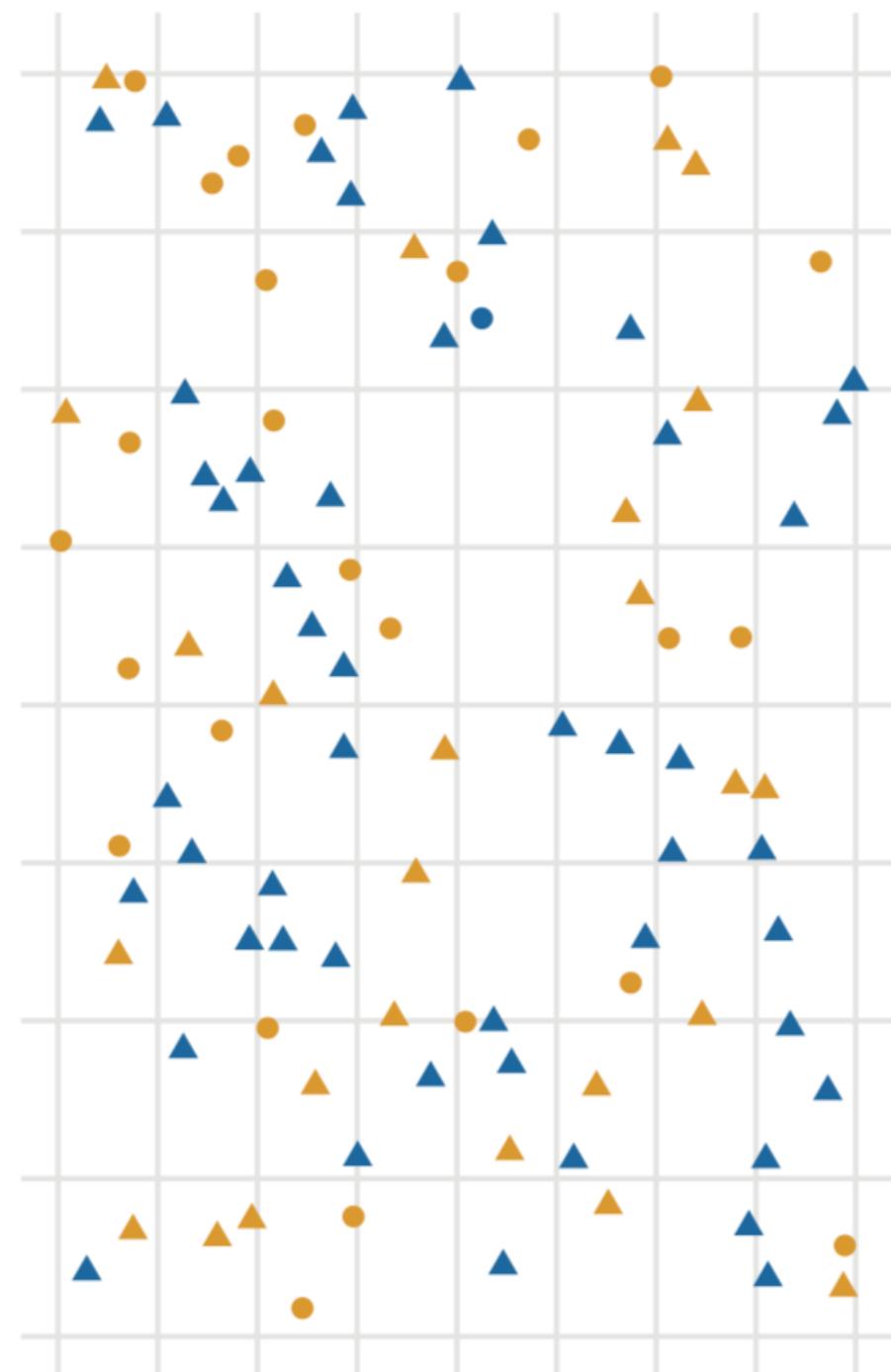
Shape only,  $N = 20$



Shape only,  $N = 100$

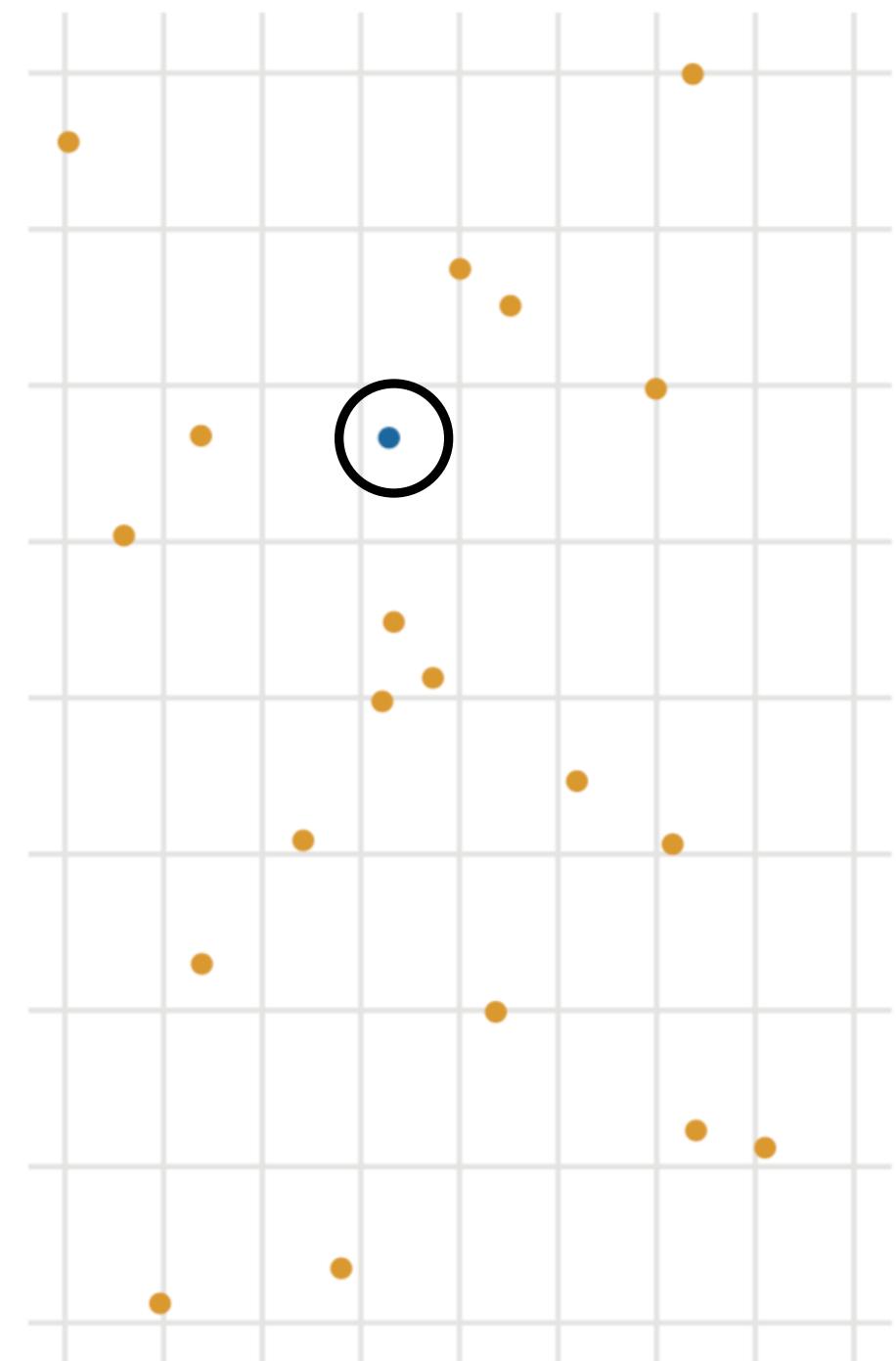


Color & shape,  $N = 100$

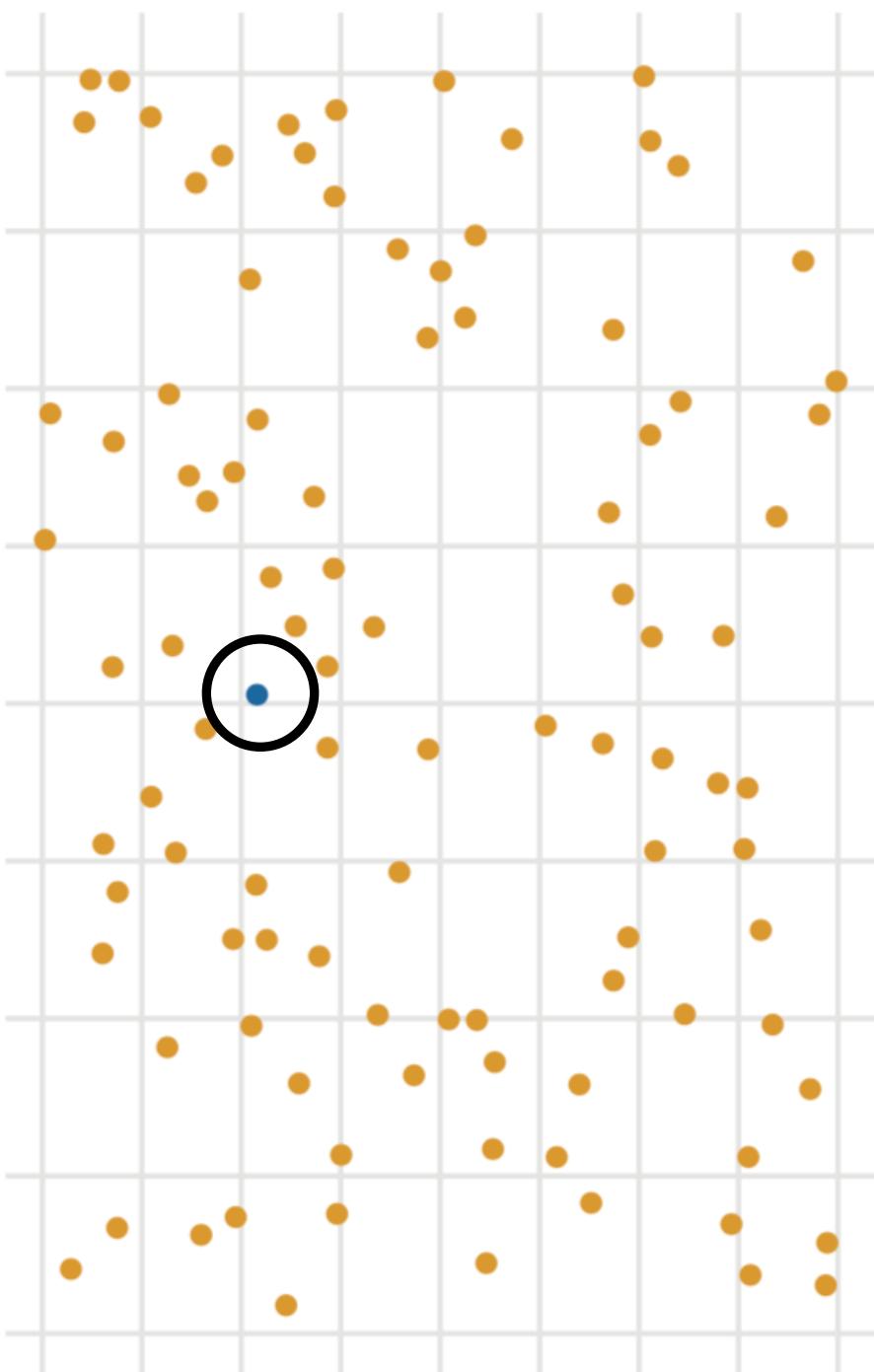


# perceptual psychology, example — *focusing visual attention*

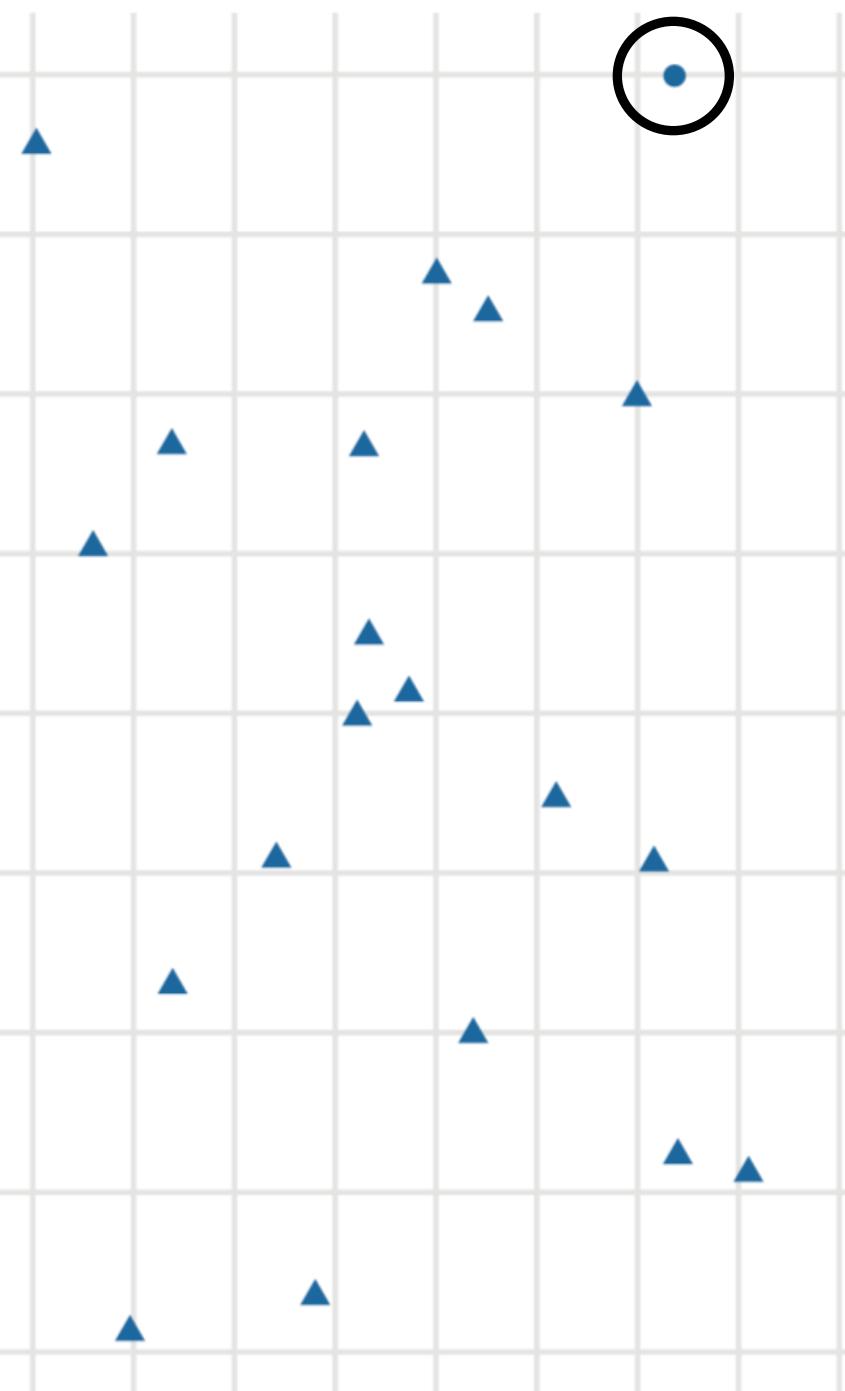
Color only,  $N = 20$



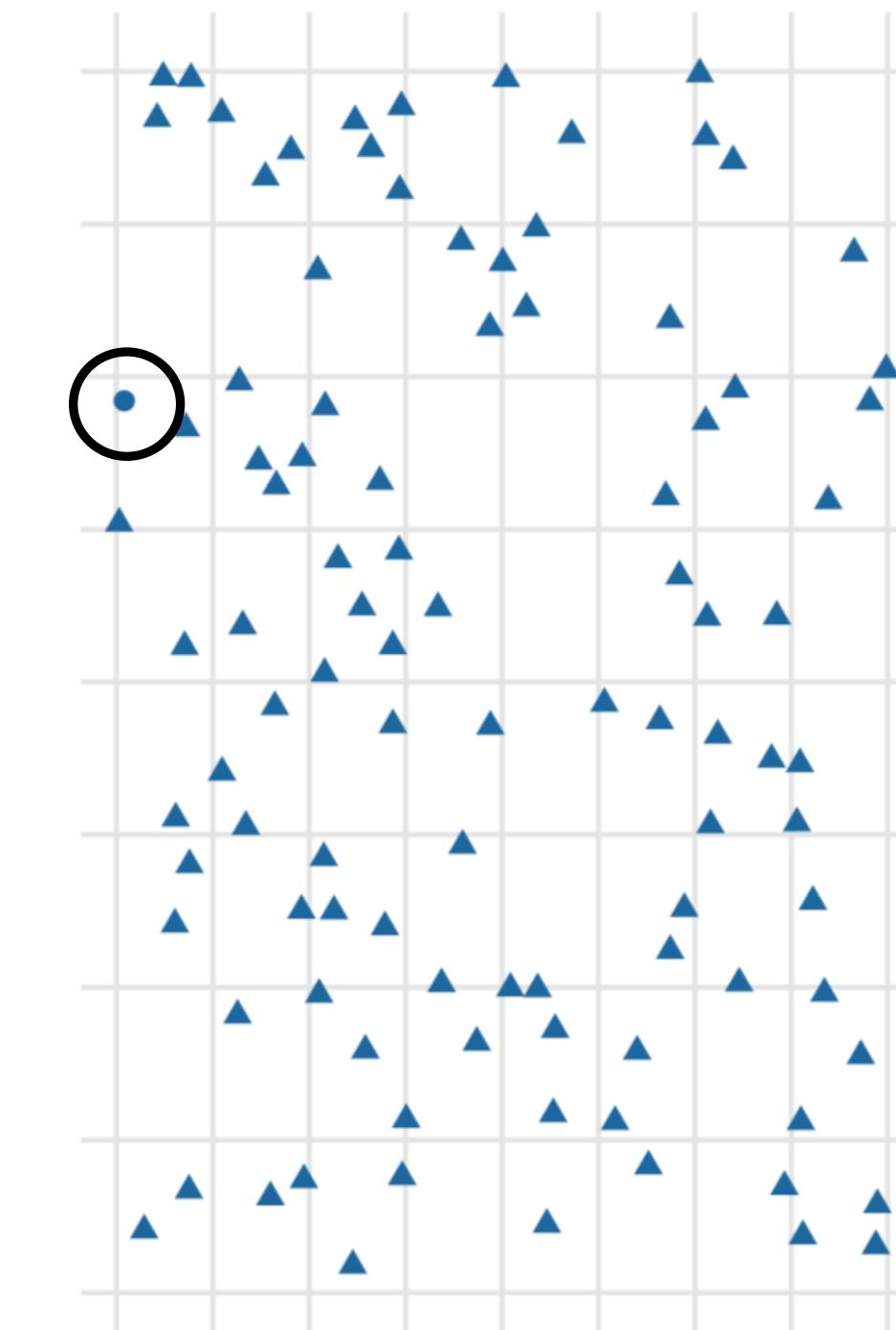
Color only,  $N = 100$



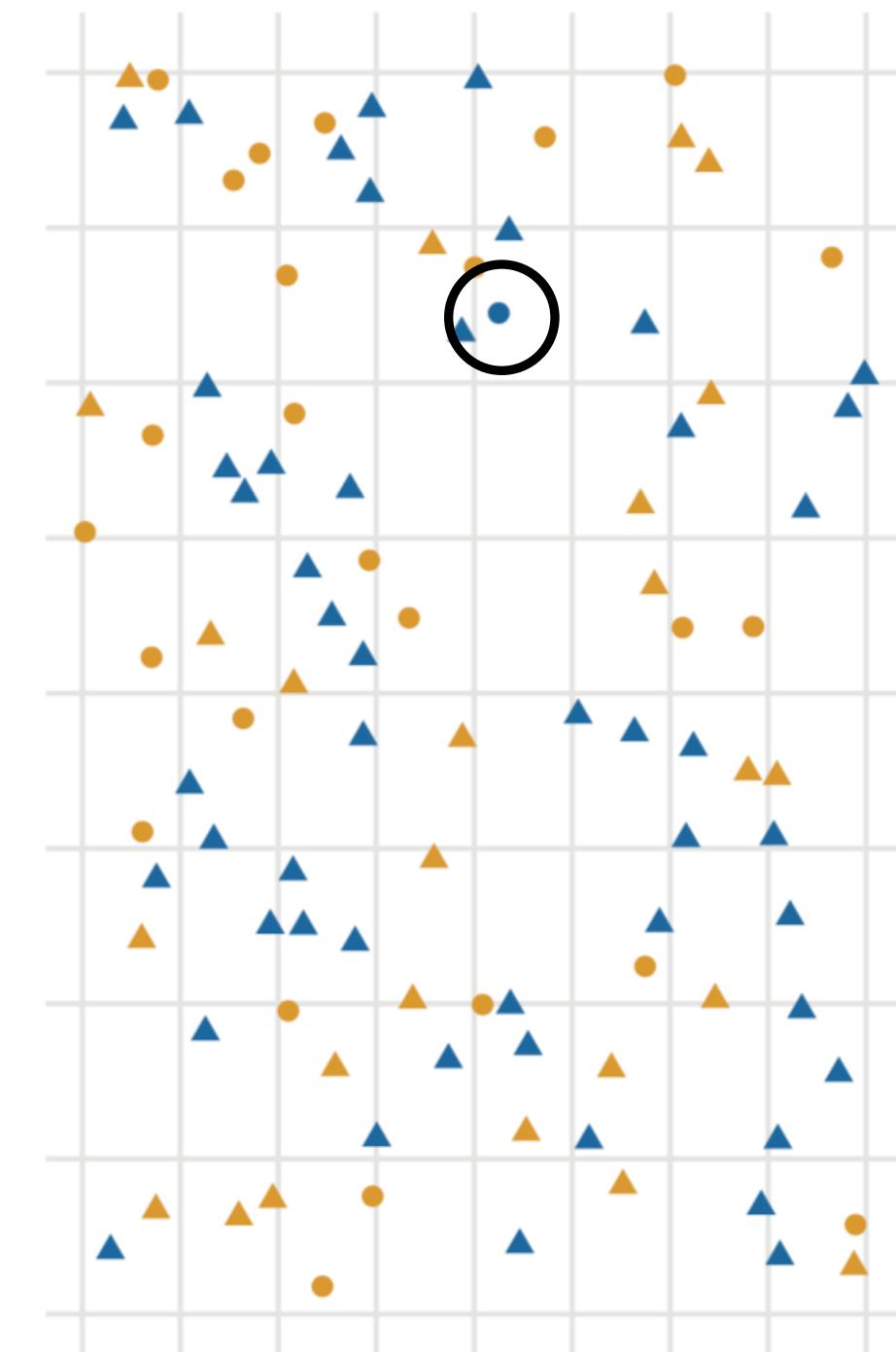
Shape only,  $N = 20$



Shape only,  $N = 100$



Color & shape,  $N = 100$



**resources**

# References

- Spencer**, Scott. "Visual, Sec. 2-2.1.3" In *Data in Wonderland*. 2021. [https://ssp3nc3r.github.io/data\\_in\\_wonderland](https://ssp3nc3r.github.io/data_in_wonderland).
- 
- Anderson**, E. W., K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva. "A User Study of Visualization Effectiveness Using EEG and Cognitive Load." *Computer Graphics Forum* 30, no. 3 (June 2011): 791–800.
- Anscombe**, F J. "Graphs in Statistical Analysis." *The American Statistician* 27, no. 1 (February 1973): 17–21.
- Bertin**, Jacques. *Semiology of Graphics: Diagrams Networks Maps*. Redlands: ESRI Press, 2010.
- Cleveland**, William S, and Robert McGill. "Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data." *Journal of the Royal Statistical Society. Series A* 150, no. 3 (1987): 192–229.
- . "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* 79, no. 387 (September 1984): 531–54.
- Harris**, Robert L. *Information Graphics: A Comprehensive Illustrated Reference*. New York: Oxford University Press, 1999.
- Heer**, Jeffrey, and Michael Bostock. "Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design." In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems*, 203–12, 2010.

- Koponen**, Juuso, and Jonatan Hildén. *Data Visualization Handbook*. First. Finland: Aalto Art Books, 2019.
- Leborg**, Christian. *Visual Grammar*. Princeton Architectural Press, 2004.
- Spencer**, Scott. (Draft) Proposal to Scott Powers. "Proposal for Exploring Game Decisions Informed by Expectations of Joint Probability Distributions." February 14, 2019.
- Tufte**, Edward R. *The Visual Display of Quantitative Information*. Second. Graphics Press, 2001.
- . *Visual Explanations. Images and Quantities, Evidence and Narrative*. Graphics Press, 1997.
- Ware**, Colin. *Information Visualization: Perception for Design*. Fourth. Philadelphia: Elsevier, Inc, 2020.
- Wickham**, Hadley. "A Layered Grammar of Graphics." *Journal of Computational and Graphical Statistics* 19, no. 1 (January 2010): 3–28.
- Wilke**, C. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. First edition. Sebastopol, CA: O'Reilly Media, 2019.
- Wilkinson**, Leland. *The Grammar of Graphics*. Second. Springer, 2005.