07

Visualization Techniques Multivariate Data

Notice

- **Author**
  - **João Moura Pires (jmp@fct.unl.pt)**

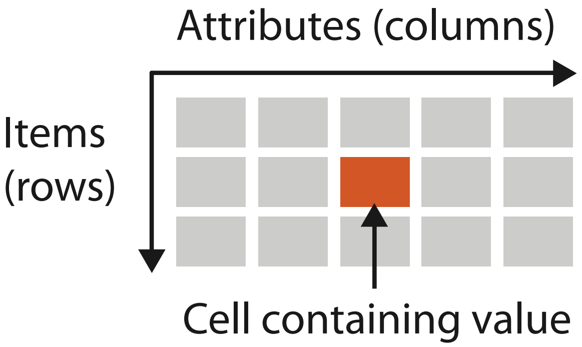
Table of Contents

- **Introduction**
- **Point-Based Techniques**
- **Line-Based Techniques**
- **Region-Based Techniques**
- **Combinations of Techniques**

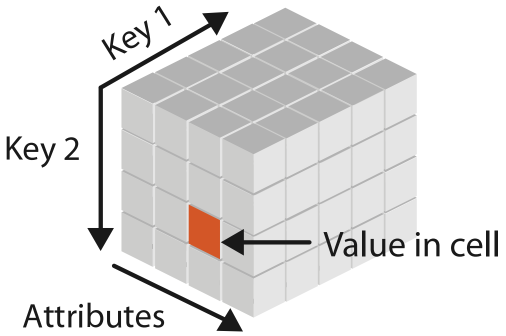
Introduction

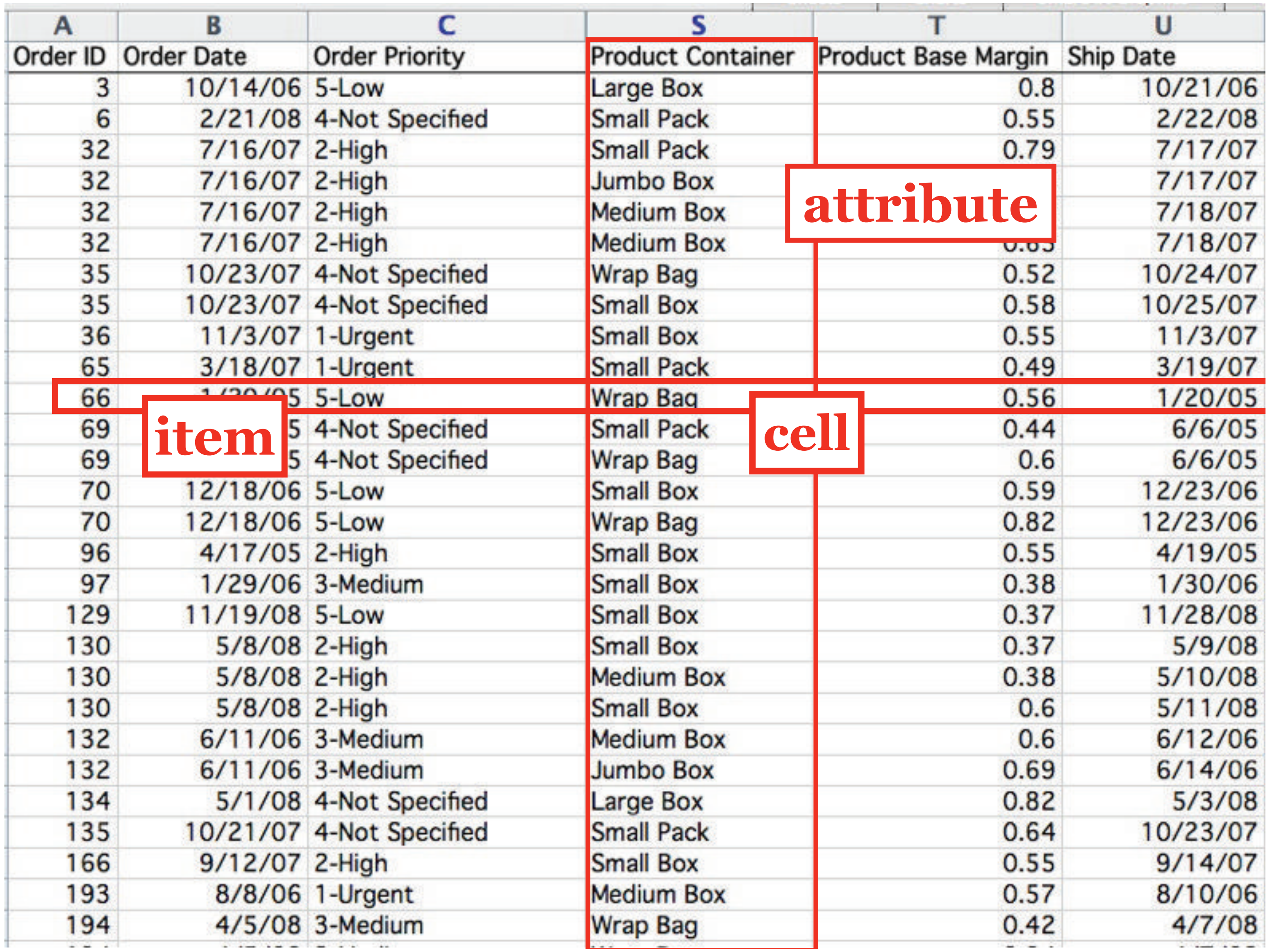
Dataset Types: **Table**

- ➔ Tables



- ➔ *Multidimensional Table*





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	[illegible]	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
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

A **multidimensional table** has a more complex structure for indexing into a cell, with multiple keys.

Tamara Munzner

Multivariate Data

- Data that does not generally have an explicit spatial attribute
- Point-Based Techniques
  - Project records from an n-dimensional data space to an arbitrary k-dimensional display space, such that data records map to k-dimensional points. (e.g. Scatterplots)
- Line-Based Techniques
  - Points corresponding to a particular record or dimension are linked together with straight or curved lines. (e.g. Line Graphs, Parallel Coordinates)
- Region-Based Techniques
  - Filled polygons are used to convey values, based on their size, shape, color, or other attributes. (e.g. Bar Charts/Histograms)

Point-Based Techniques

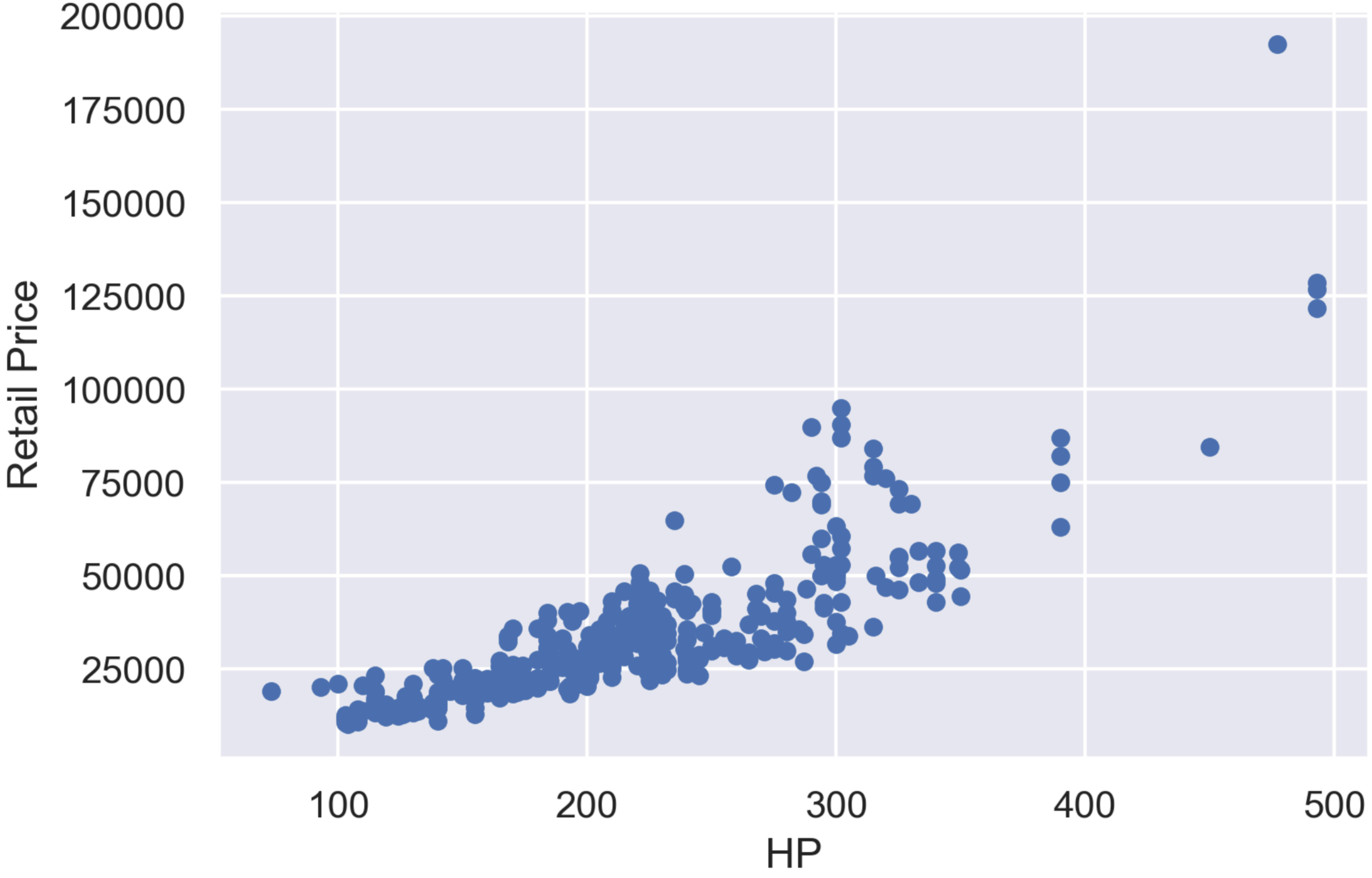
Multivariate Data: Point-Based Techniques

- **Scatterplots and Scatterplot Matrices**
  - **Their success stems from our innate abilities to judge relative position within a bounded space**

- **As the dimensionality of the data increases, the choices for visual analysis consist of:**
  - **dimension subsetting** (user selection or algorithm based suggestion)**;**
  - **dimension embedding** (mapping dimensions to other graphical attributes besides position, such as color, size, and shape)**;**
  - **multiple displays** (either superimposed or juxtaposed - e. g. scatterplot matrix);
  - **dimension reduction** (to transform the high-dimensional data to data of lower dimension).

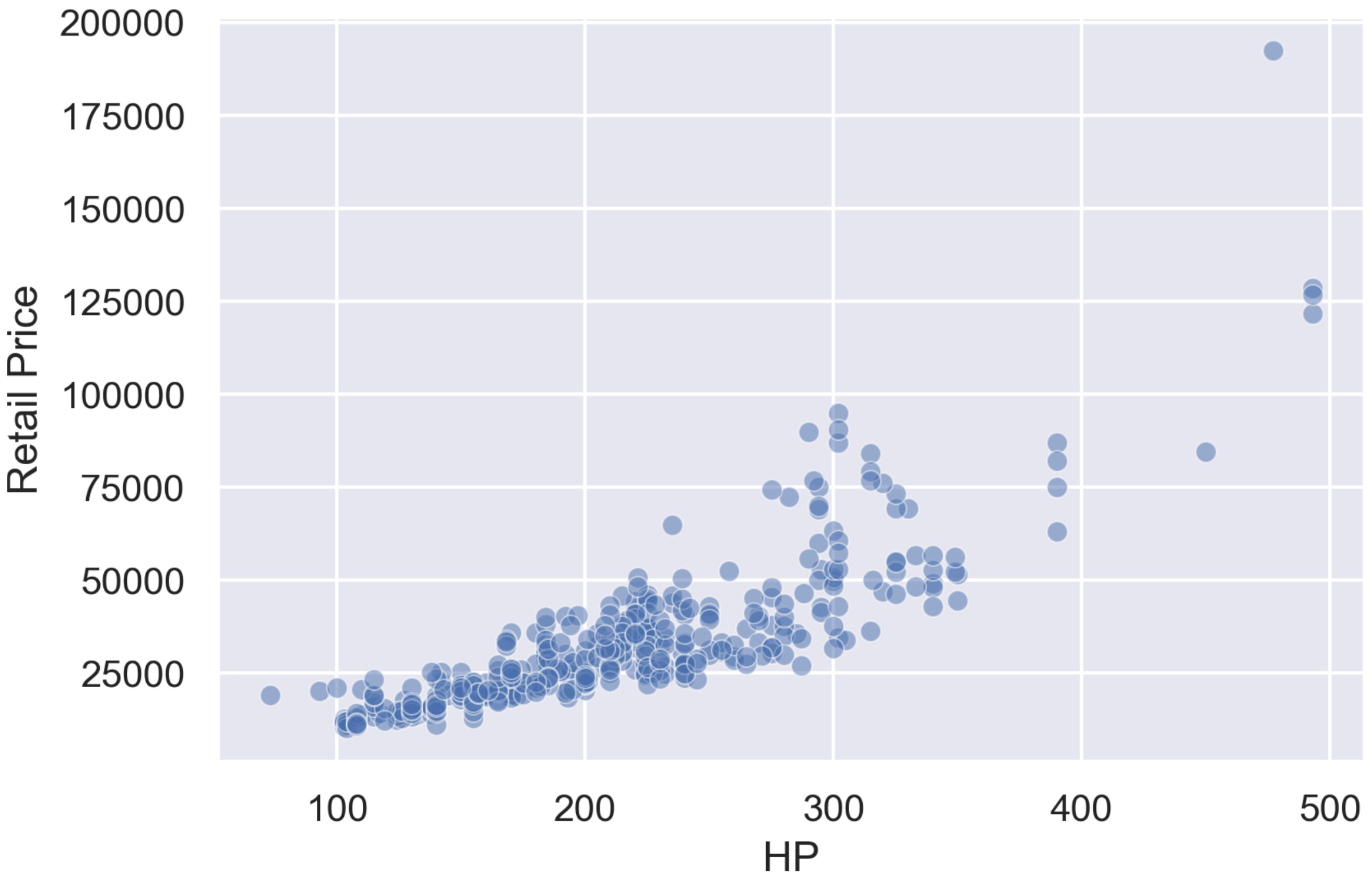
Multivariate Data: Point-based Techniques

- **Scatterplots**



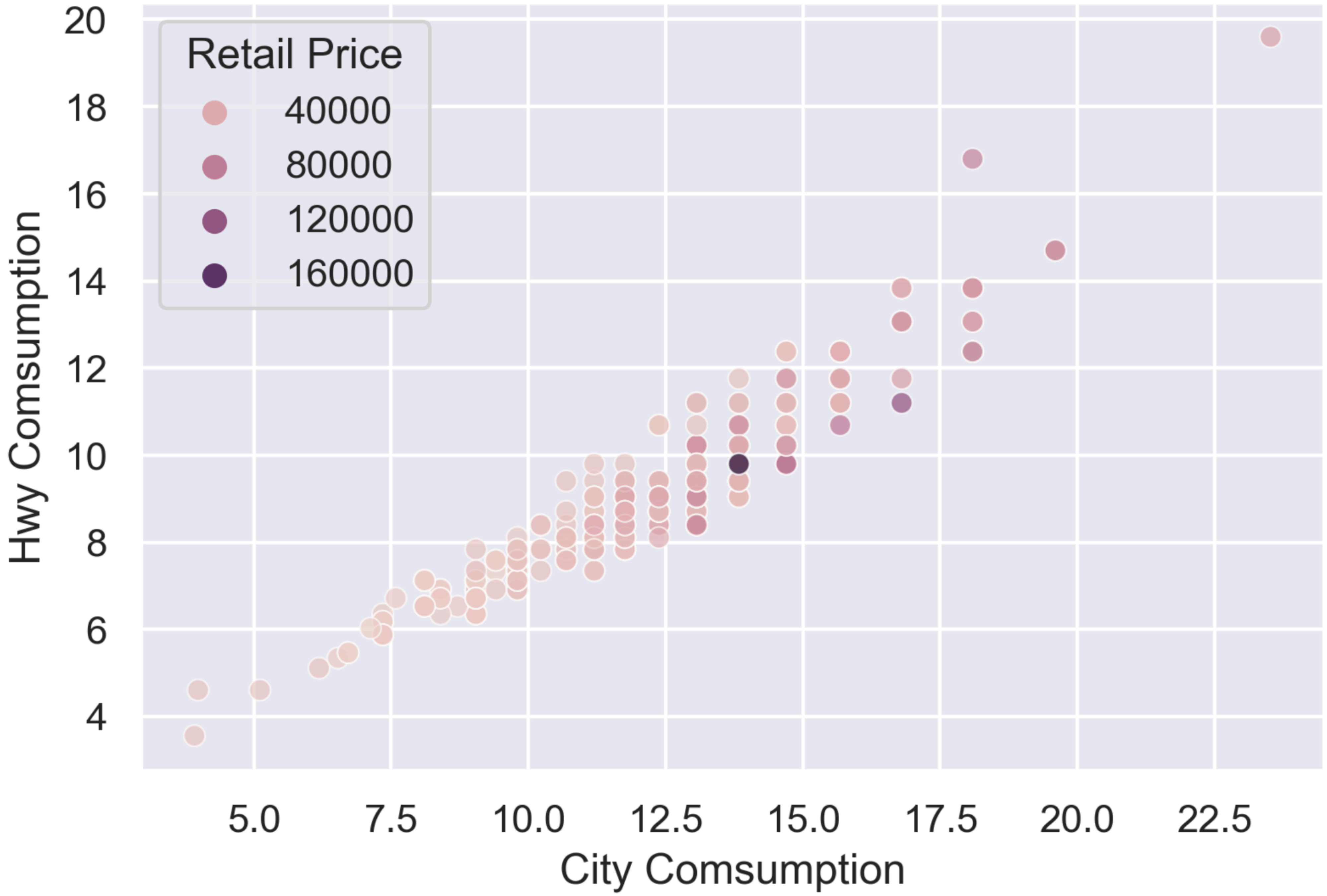
Multivariate Data: Point-based Techniques

- **Scatterplots**



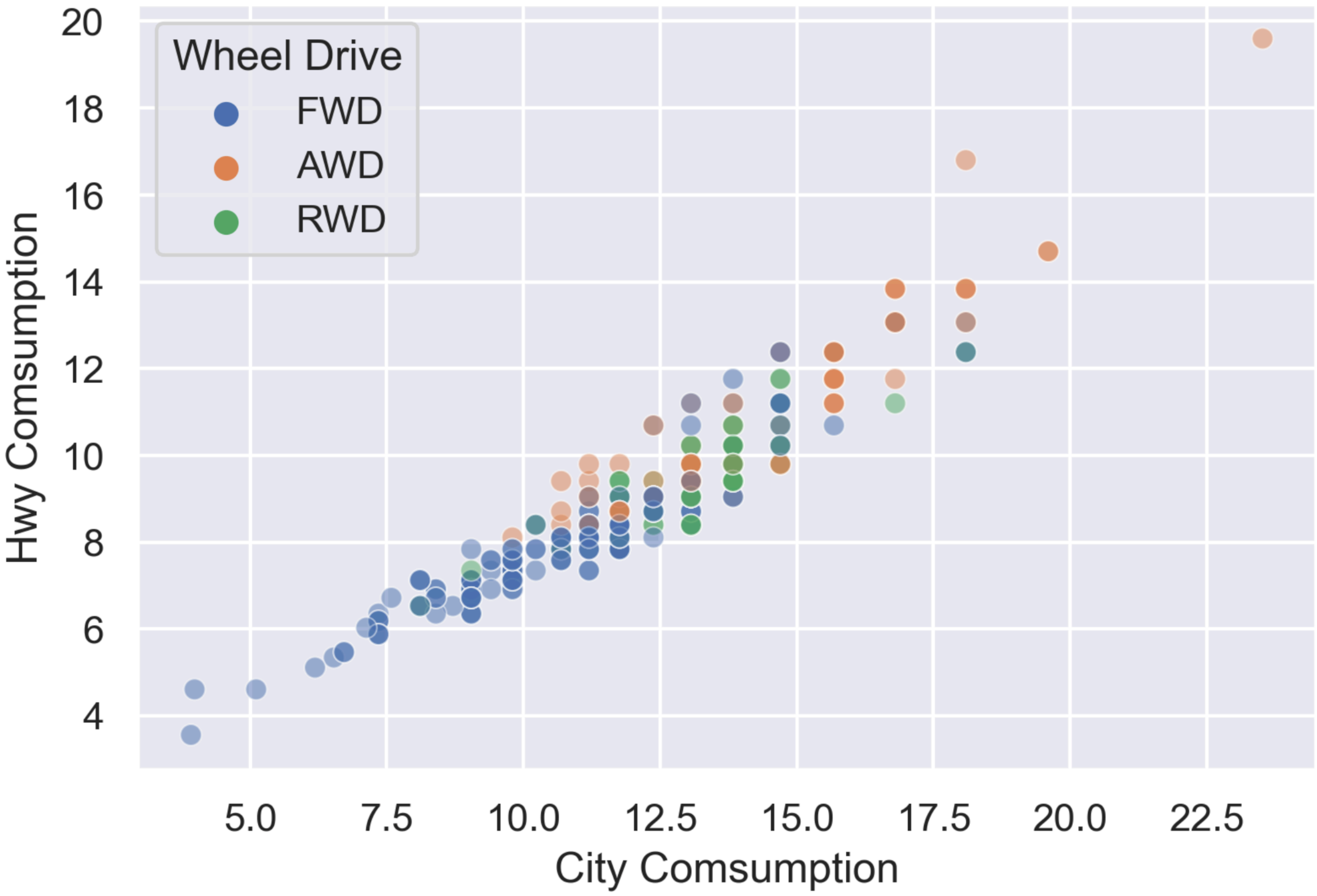
Multivariate Data: Point-based Techniques

- **Scatterplots**



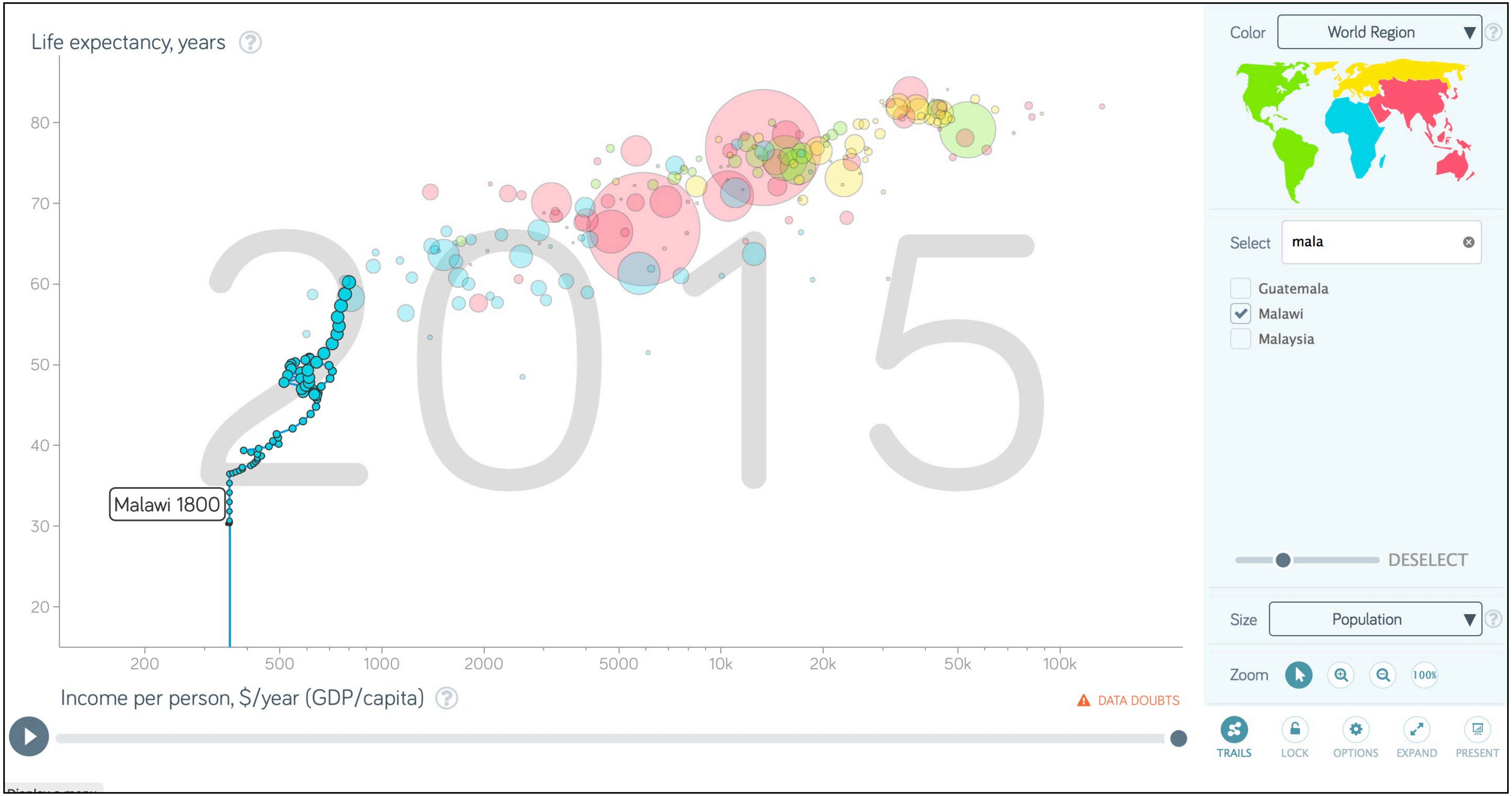
Multivariate Data: Point-based Techniques

- **Scatterplots**



Multivariate Data: Point-based Techniques

Scatterplots



Color: World Region

Select: mala

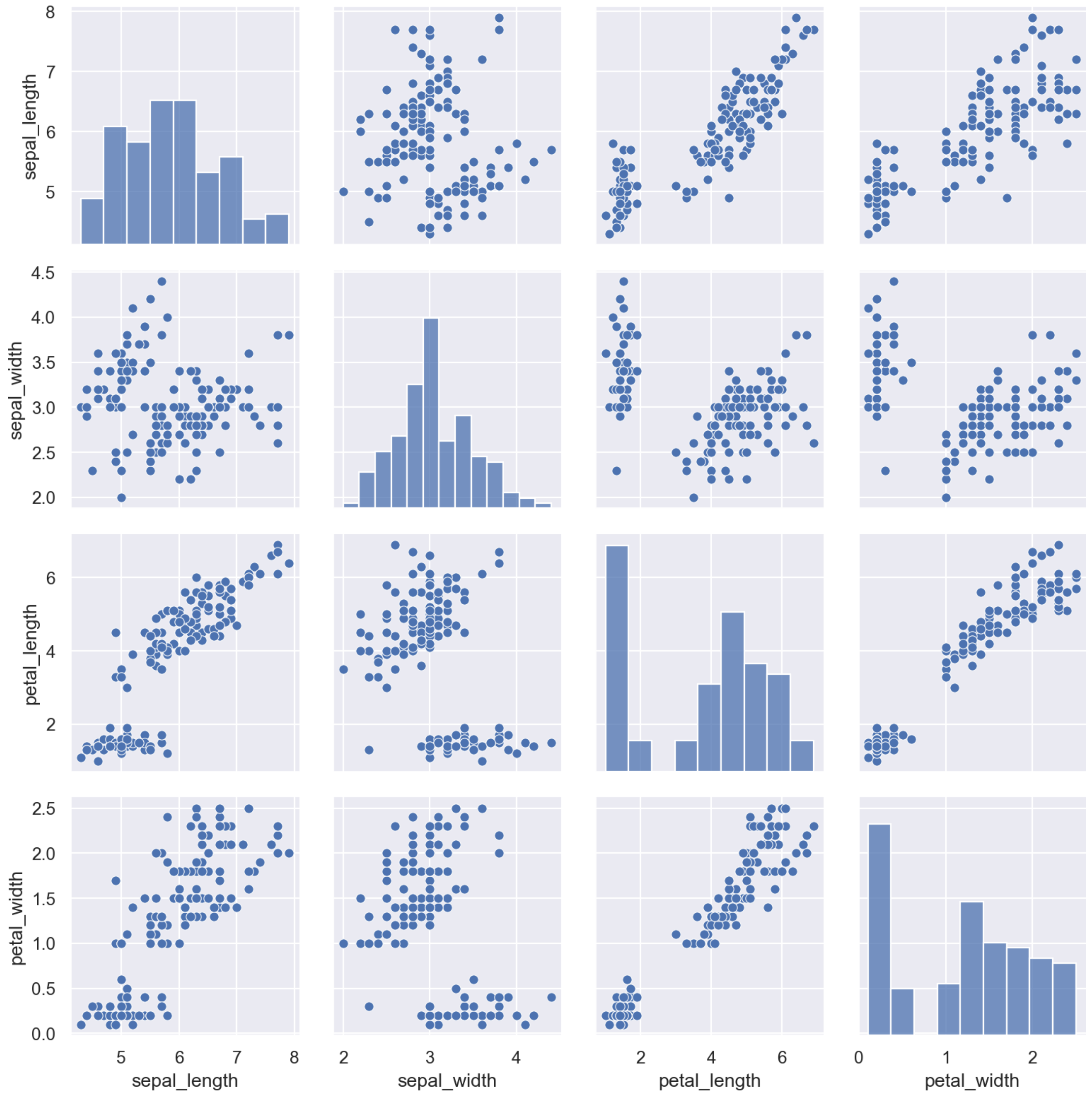
- □ Guatemala
- ☑ Malawi
- □ Malaysia

Size: Population

Multivariate Data: Point-based Techniques

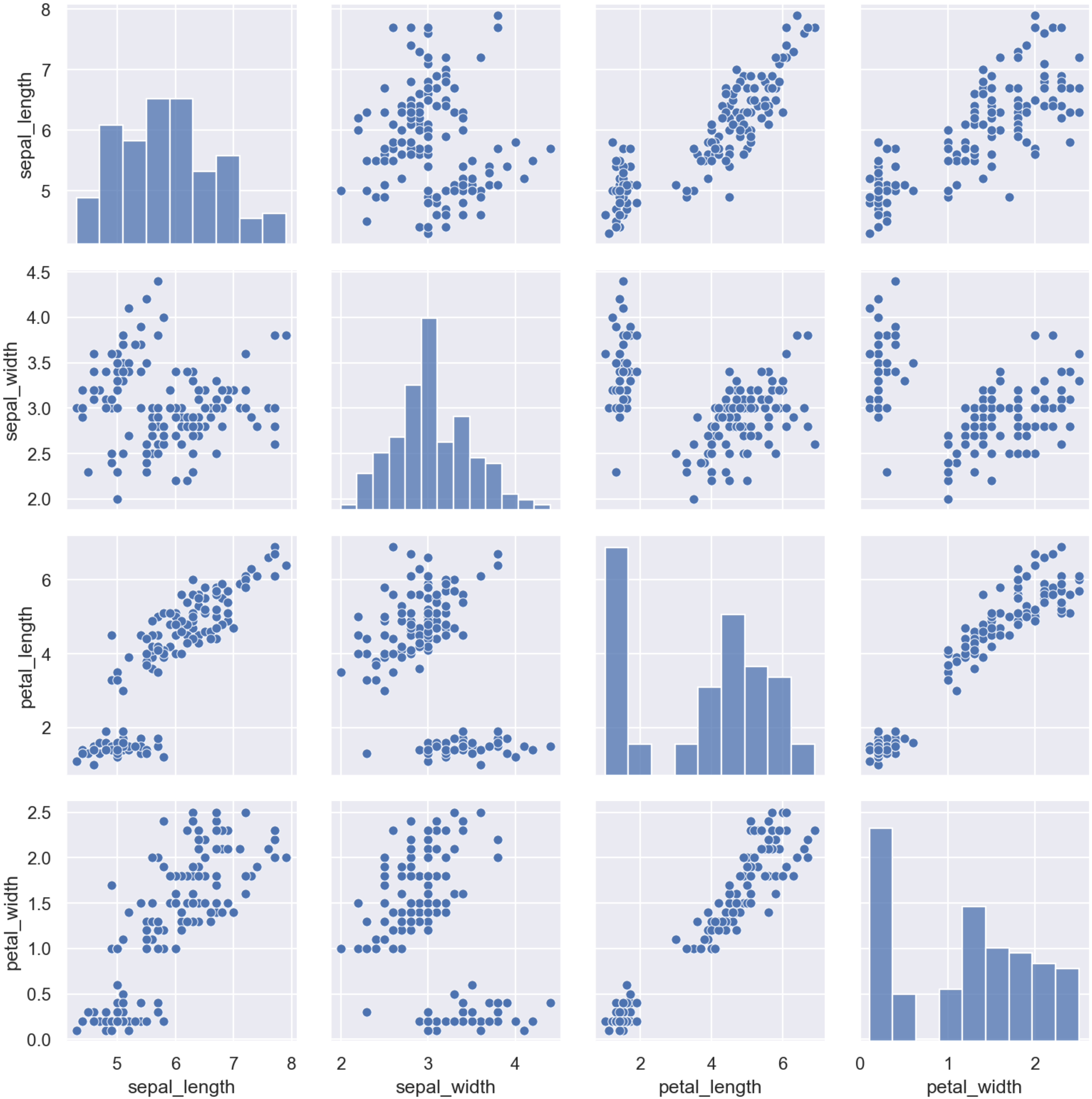
- **Scatterplots**

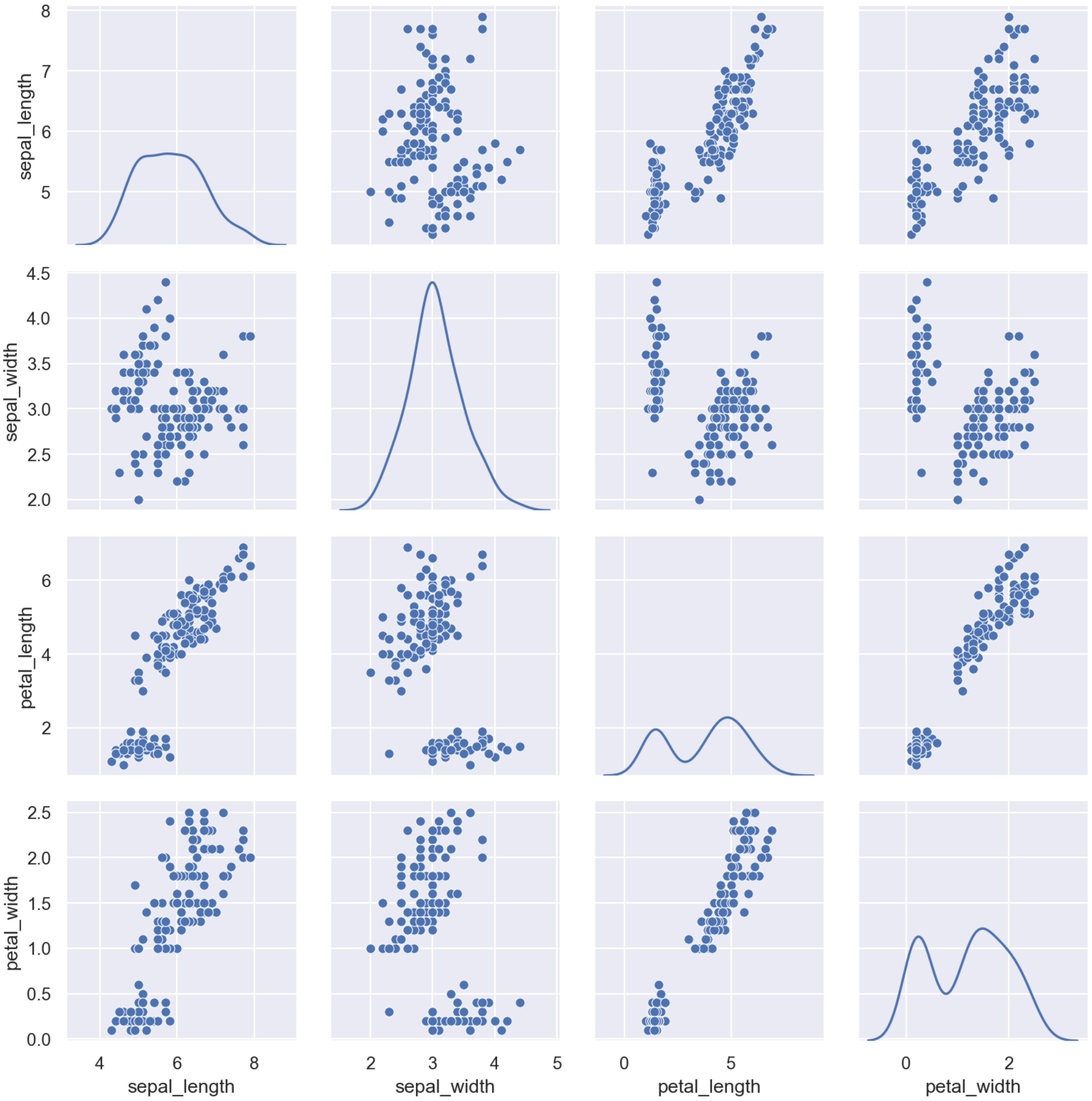
  Matrix

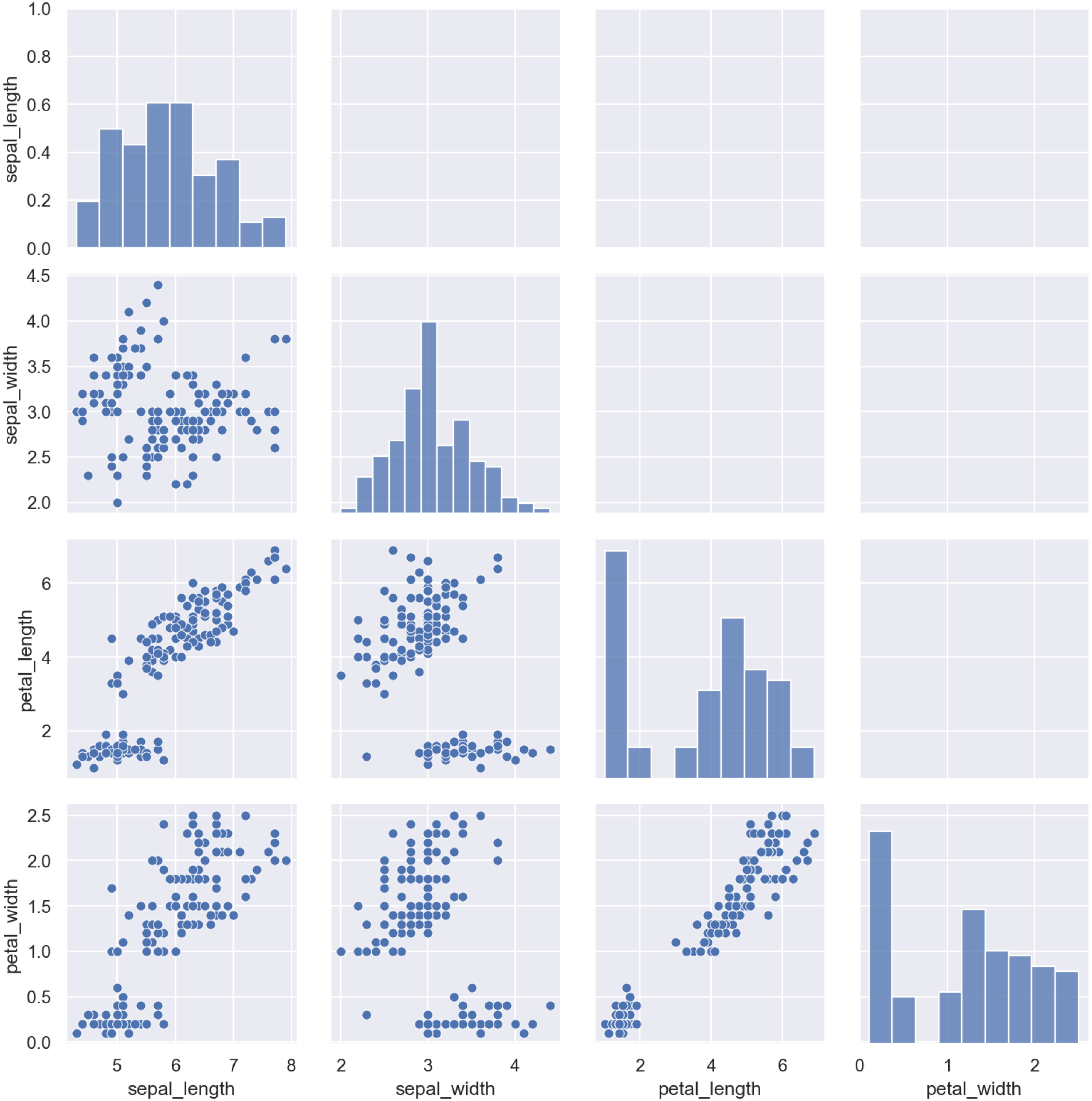


https://archive.ics.uci.edu/ml/datasets/iris

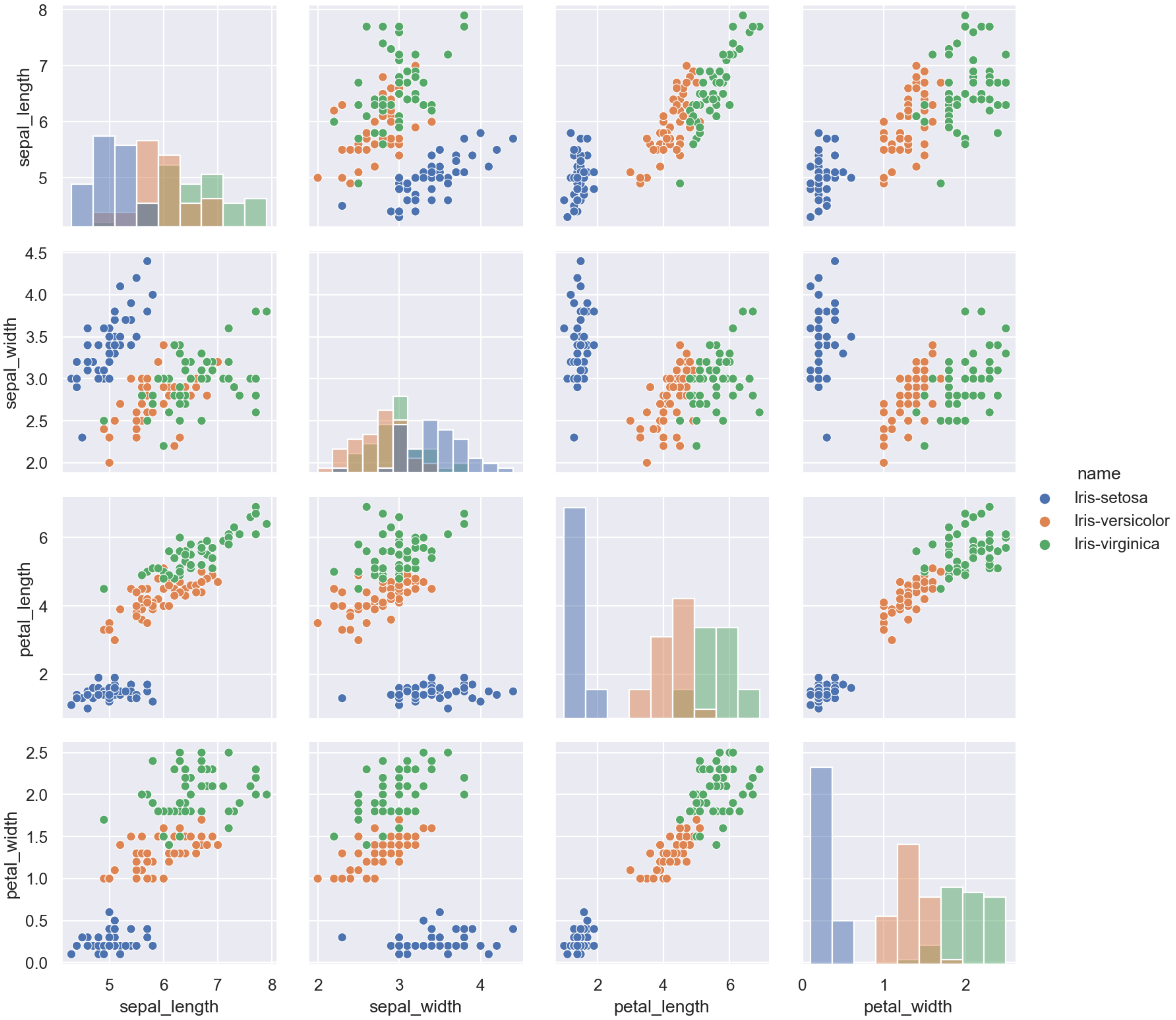
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA


sepal_length
sepal_width
petal_length
petal_width
8
7
6
5
4.5
4.0
3.5
3.0
2.5
2.0
6
4
2
2.5
2.0
1.5
1.0
0.5
0.0
5
6
7
8
2
3
4
2
4
6
0
1
2
sepal_length
sepal_width
petal_length
petal_width

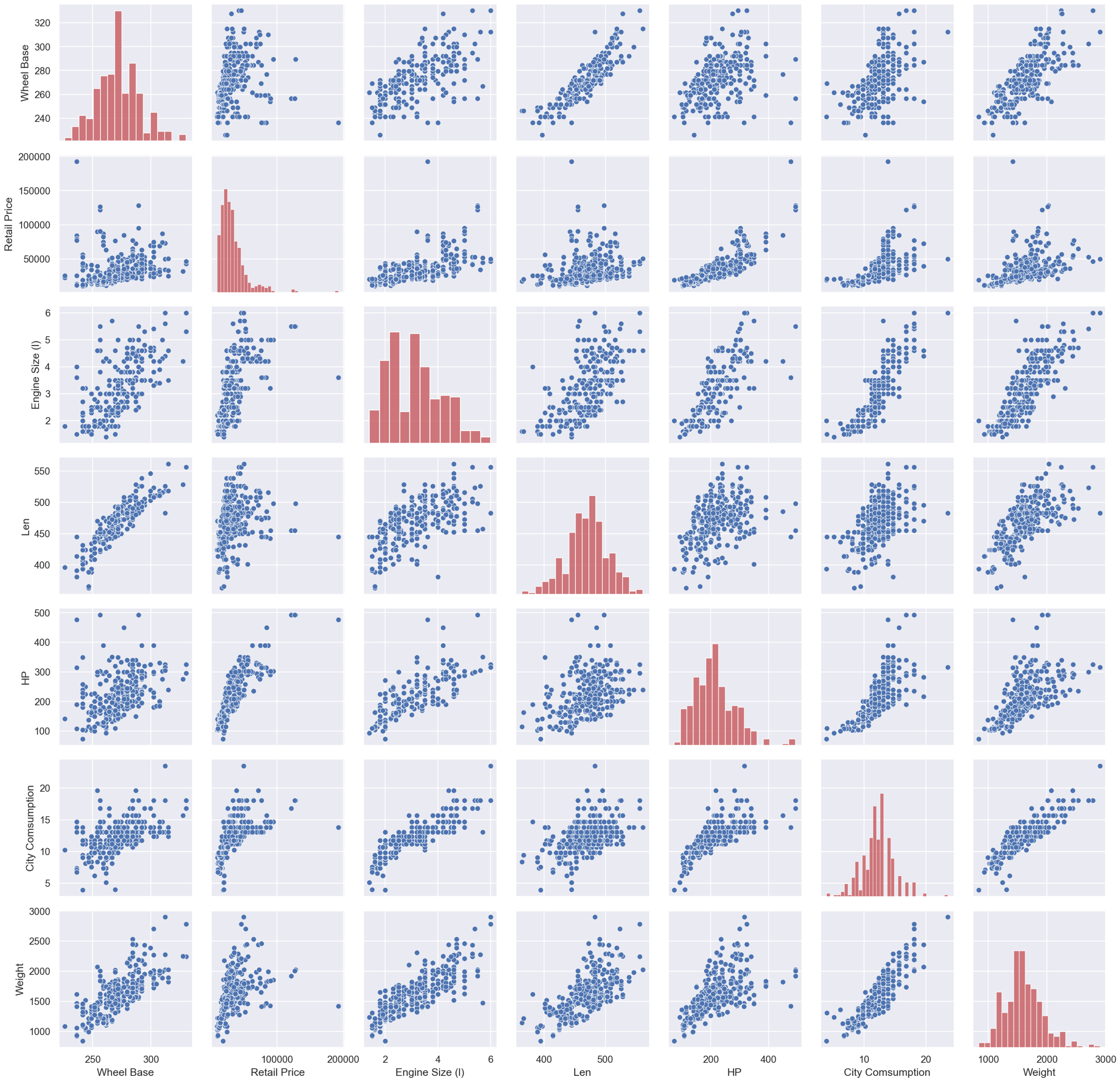

sepal_length
sepal_width
petal_length
petal_width
8
7
6
5
4.5
4.0
3.5
3.0
2.5
2.0
6
4
2
2.5
2.0
1.5
1.0
0.5
0.0
4
6
8
2
3
4
5
0
5
0
2
sepal_length
sepal_width
petal_length
petal_width

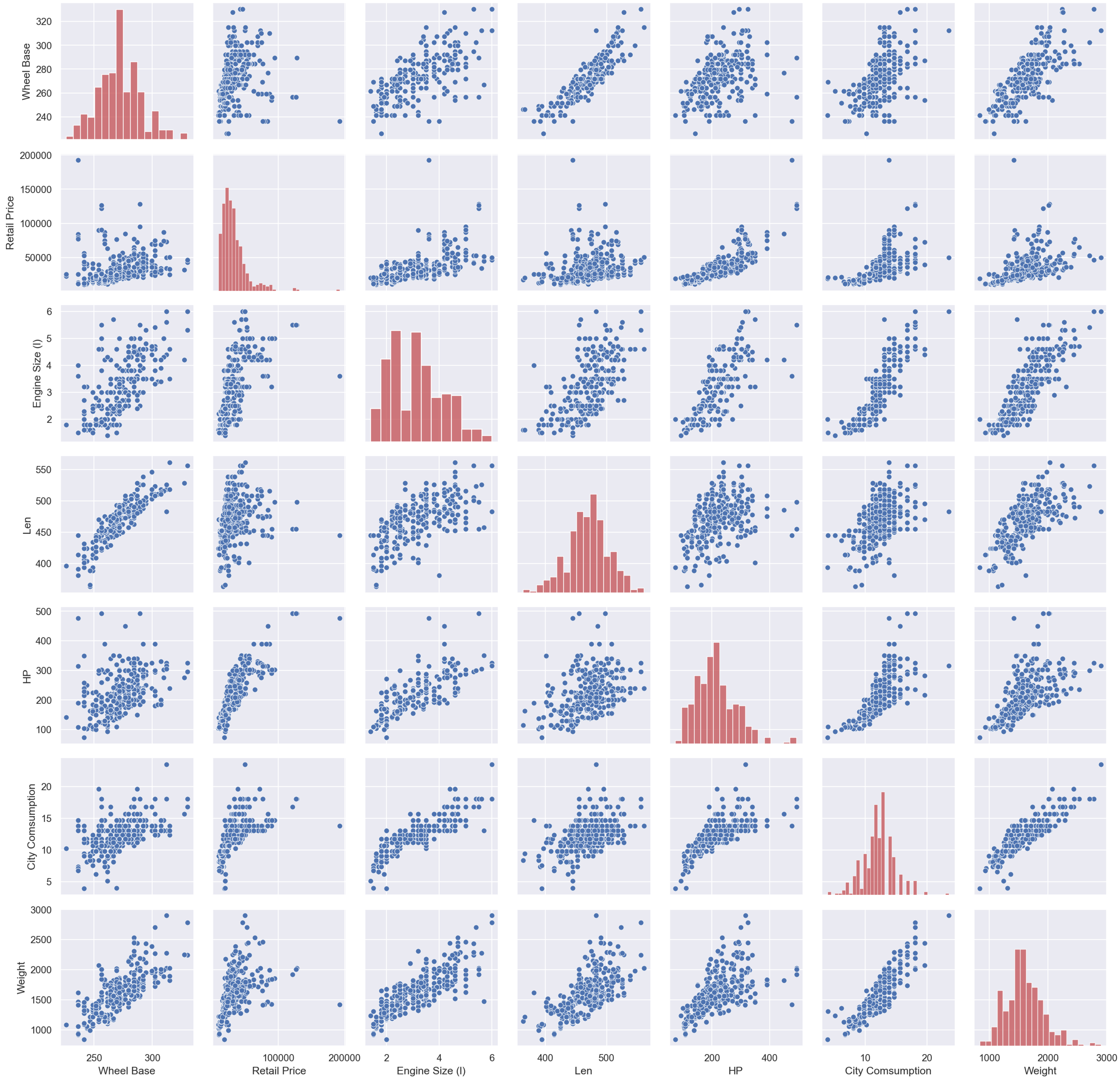


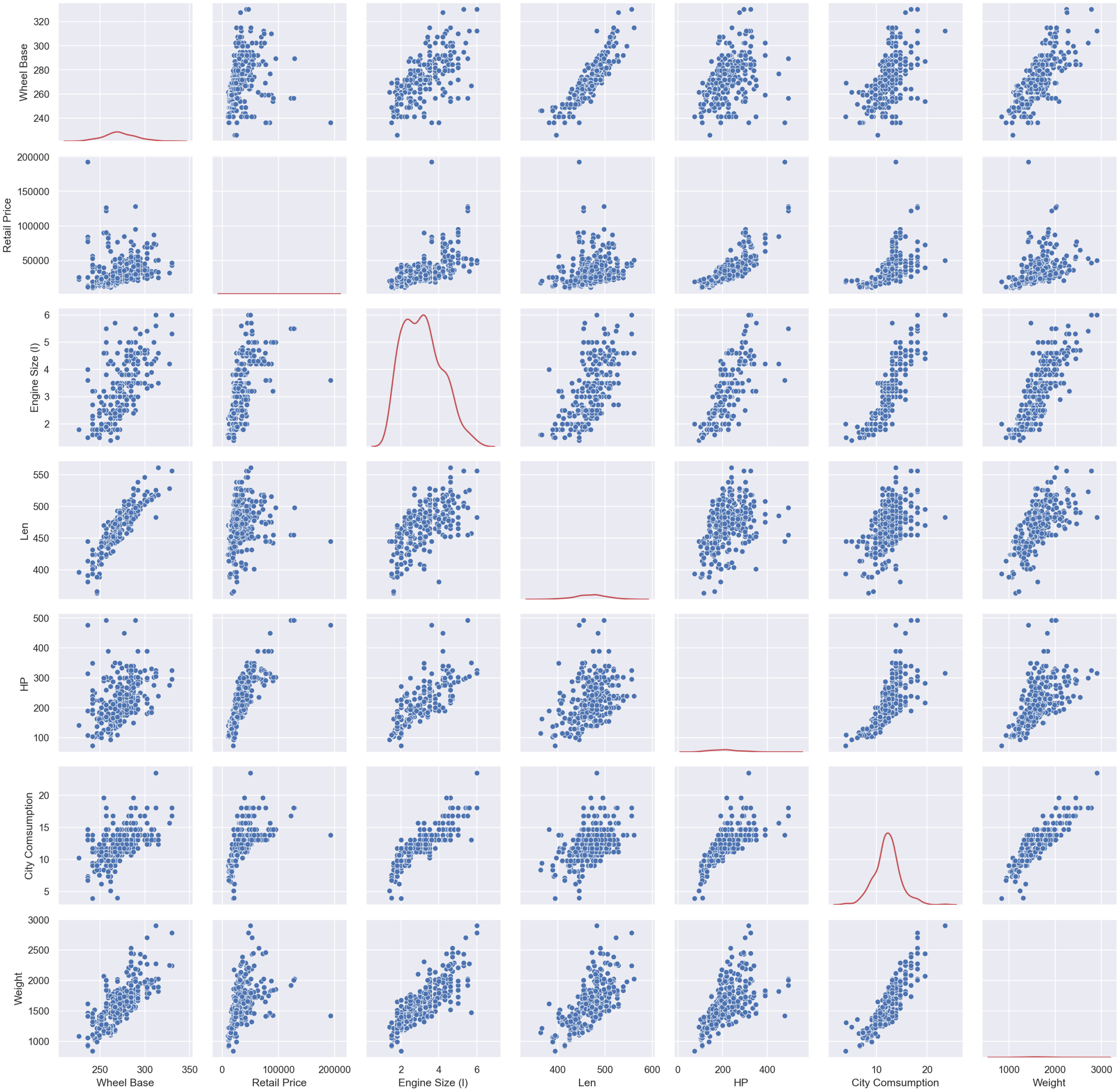
sepal_length
sepal_width
petal_length
petal_width
1.0
0.8
0.6
0.4
0.2
0.0
4.5
4.0
3.5
3.0
2.5
2.0
6
4
2
2.5
2.0
1.5
1.0
0.5
0.0
5
6
7
8
2
3
4
2
4
6
0
1
2
sepal_length
sepal_width
petal_length
petal_width

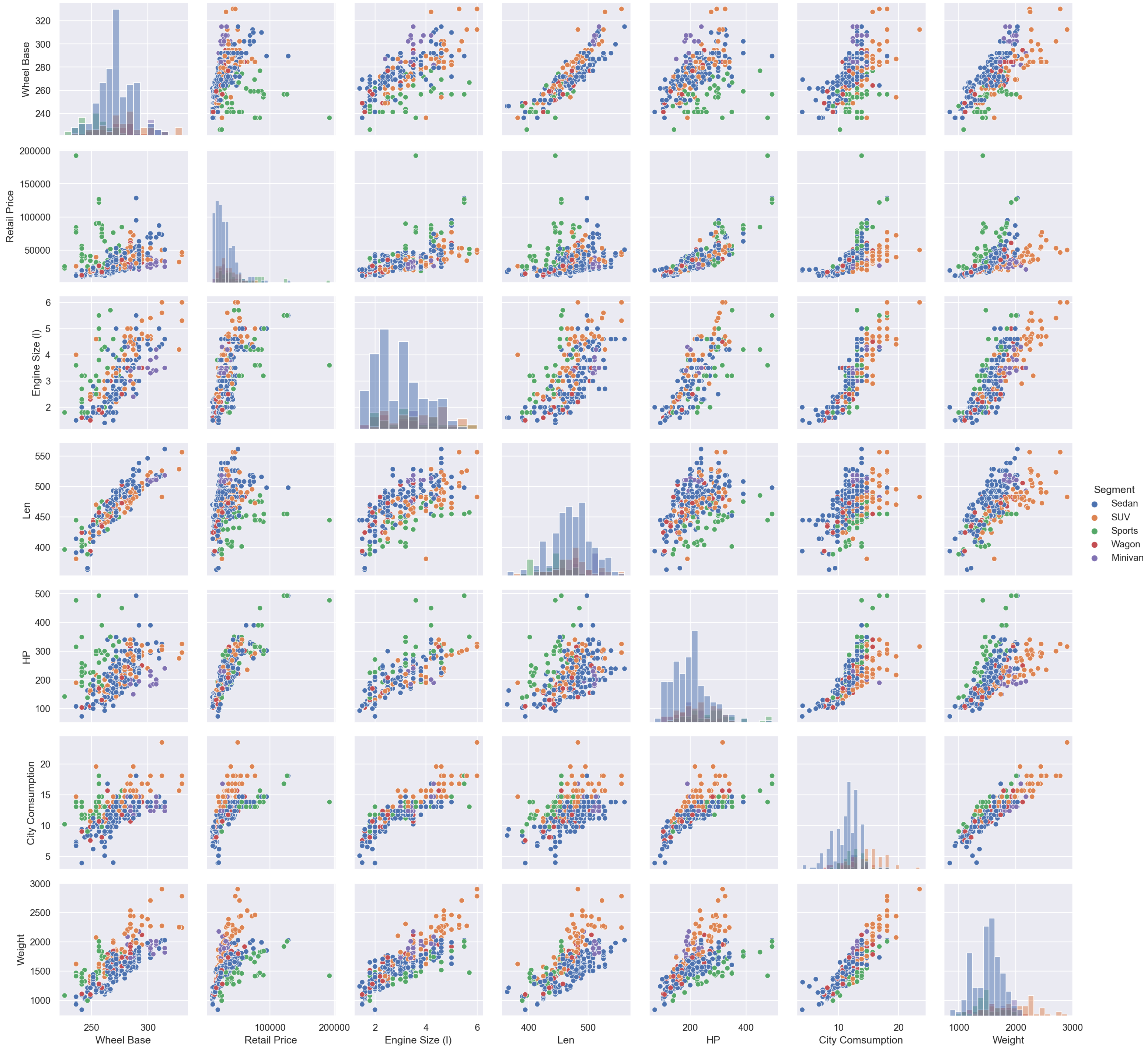

sepal_length
sepal_width
petal_length
petal_width
8
7
6
5
4.5
4.0
3.5
3.0
2.5
2.0
6
4
2
2.5
2.0
1.5
1.0
0.5
0.0
5
6
7
8
2
3
4
2
4
6
0
1
2
name
Iris-setosa
Iris-versicolor
Iris-virginica

Scatter matrix (in Python)

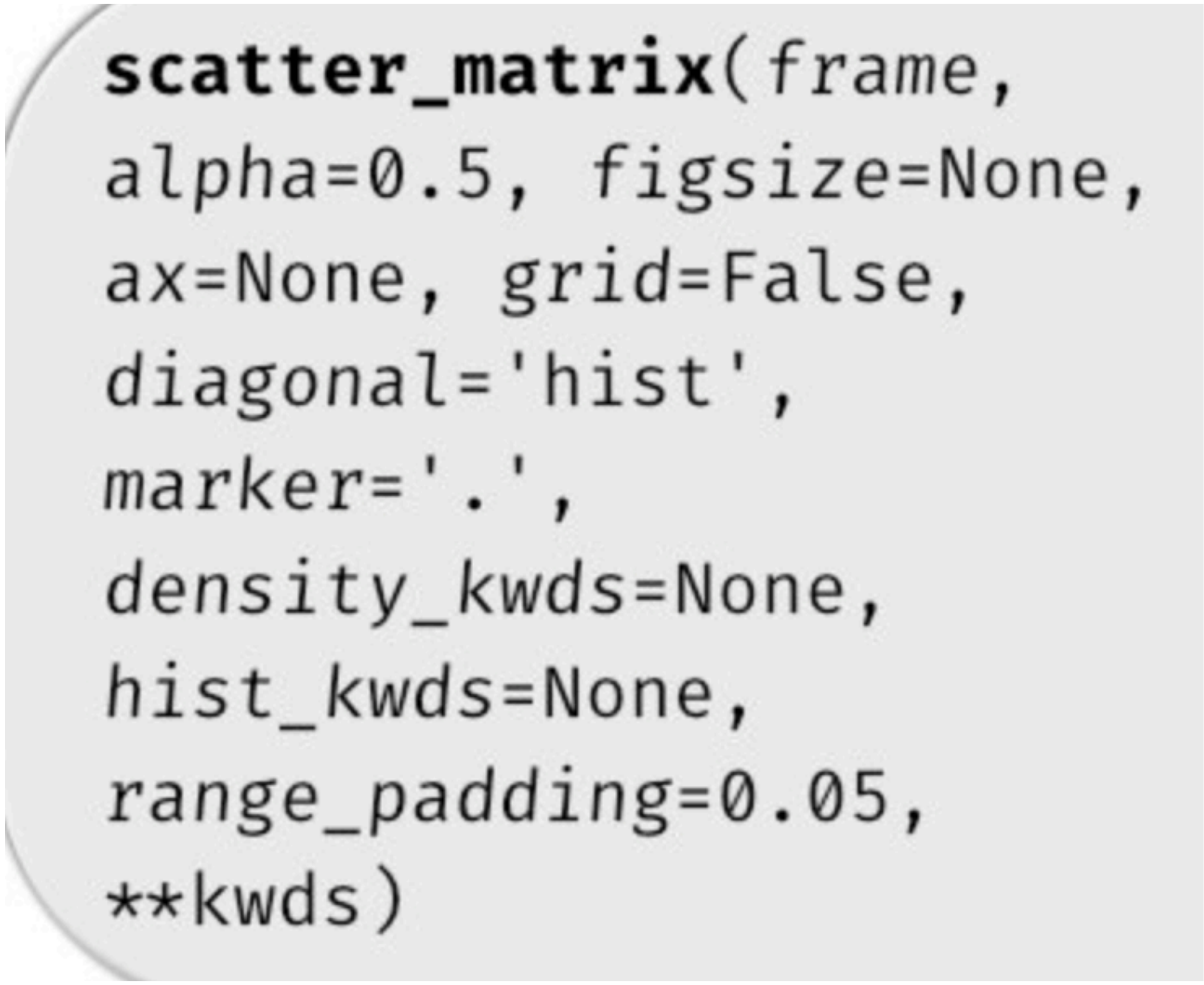



Wheel Base
320
300
280
260
240
Retail Price
200000
150000
100000
50000
Engine Size (l)
6
5
4
3
2
Len
550
500
450
400
HP
500
400
300
200
100
City Comsumption
20
15
10
5
Weight
3000
2500
2000
1500
1000
250
300
Wheel Base
100000
200000
Retail Price
2
4
6
Engine Size (l)
400
500
Len
200
400
HP
10
20
City Comsumption
1000
2000
3000
Weight

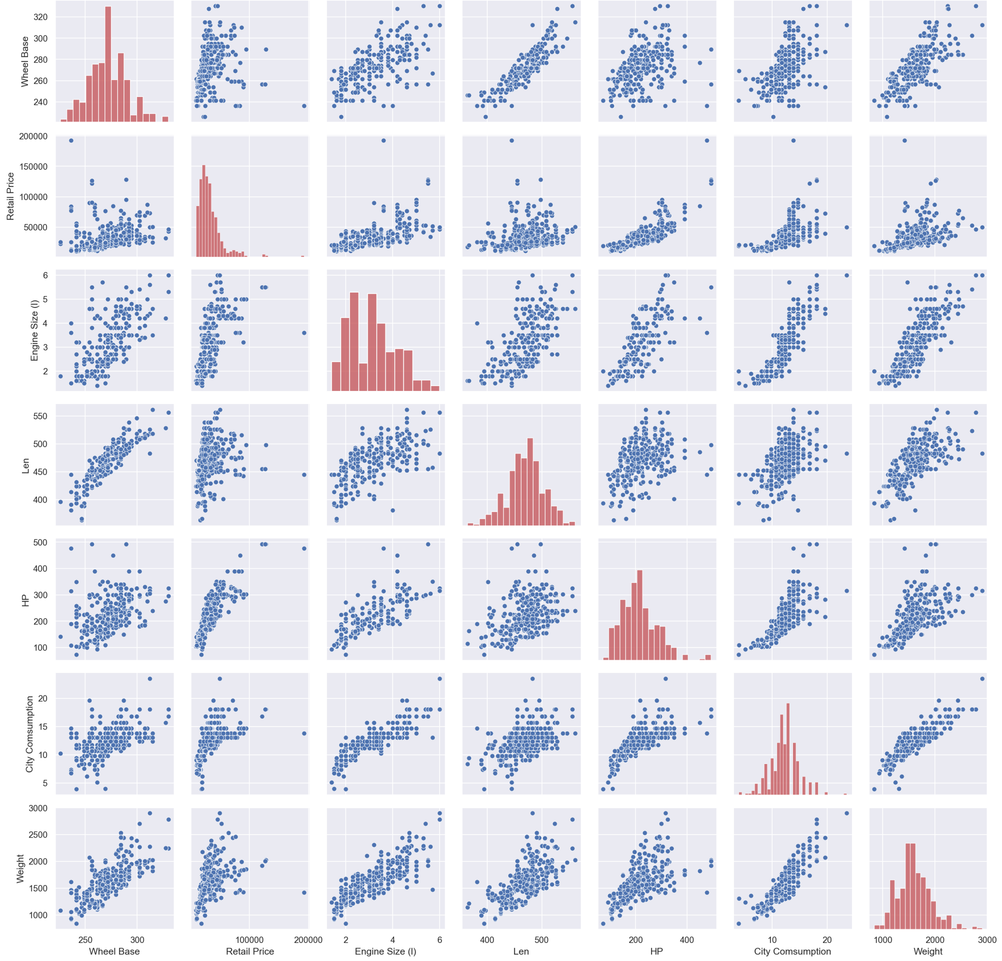

Wheel Base
240
260
280
300
320
Retail Price
50000
100000
150000
200000
Engine Size (l)
2
3
4
5
6
Len
400
450
500
550
HP
100
200
300
400
500
City Comsumption
5
10
15
20
Weight
1000
1500
2000
2500
3000
250
300
350
0
100000
200000
2
4
6
400
500
600
0
200
400
10
20
1000
2000
3000


Wheel Base
240
260
280
300
320
Retail Price
50000
100000
150000
200000
Engine Size (l)
2
3
4
5
6
Len
400
450
500
550
HP
100
200
300
400
500
City Comsumption
5
10
15
20
Weight
1000
1500
2000
2500
3000
250
300
100000
200000
2
4
6
400
500
200
400
10
20
1000
2000
3000
Segment
Sedan
SUV
Sports
Wagon
Minivan

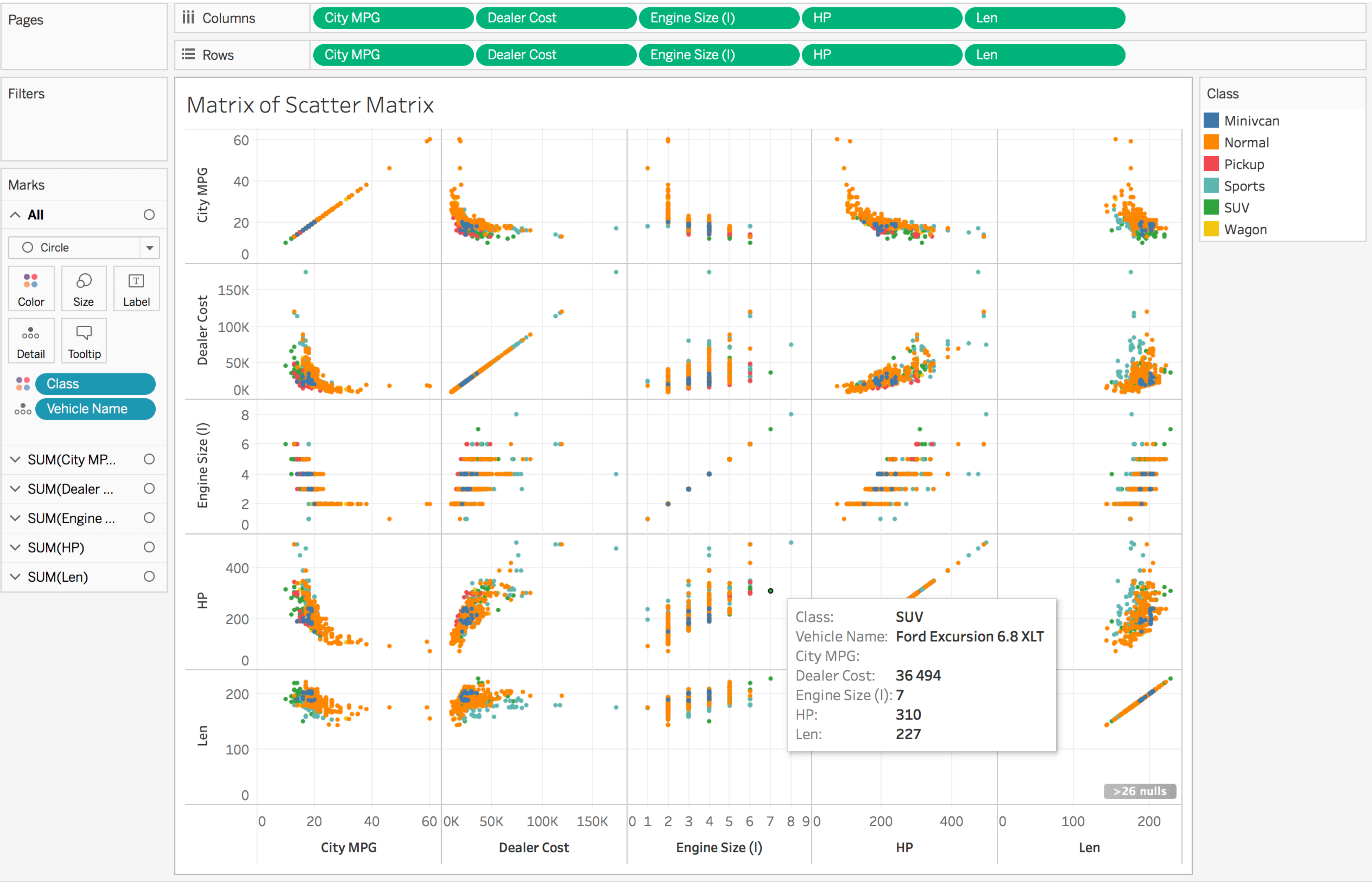
Scatter matrix (in Python)



```
scatter_matrix(frame,
alpha=0.5, figsize=None,
ax=None, grid=False,
diagonal='hist',
marker='.',
density_kwds=None,
hist_kwds=None,
range_padding=0.05,
**kwds)
```



Scatter Matrix (in Tableau)



Columns: City MPG | Dealer Cost | Engine Size (l) | HP | Len

Rows: City MPG | Dealer Cost | Engine Size (l) | HP | Len

Marks: All — Circle

Color: Class

Detail: Vehicle Name

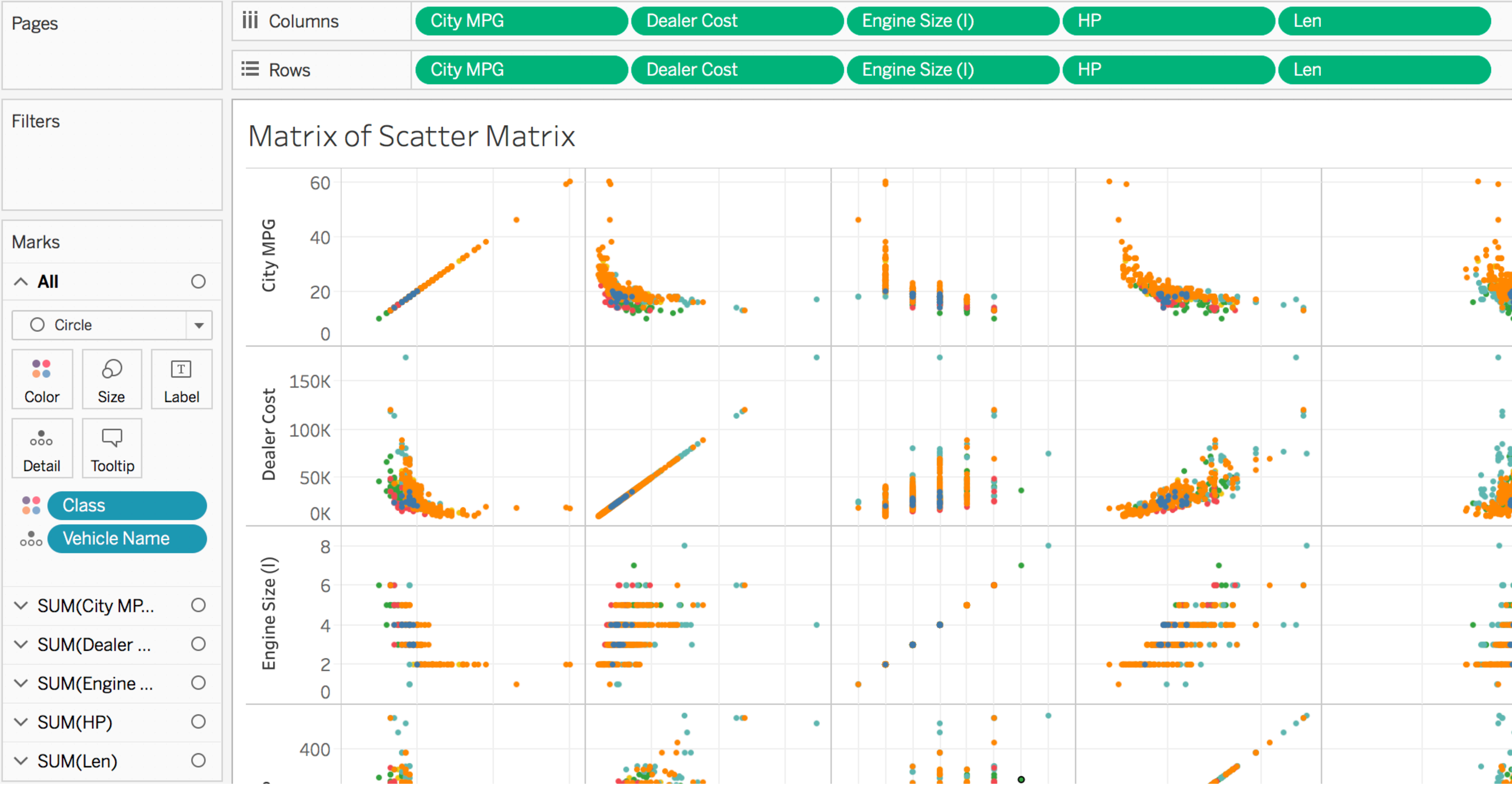
SUM(City MP...
SUM(Dealer ...
SUM(Engine ...
SUM(HP)
SUM(Len)

Matrix of Scatter Matrix

Class
- 🟦 Minivcan
- 🟧 Normal
- 🟥 Pickup
- 🟦 Sports
- 🟩 SUV
- 🟨 Wagon

Class: SUV
Vehicle Name: Ford Excursion 6.8 XLT
City MPG:
Dealer Cost: 36 494
Engine Size (l): 7
HP: 310
Len: 227

Scatter Matrix (in Tableau)



Pages

Columns: City MPG | Dealer Cost | Engine Size (l) | HP | Len

Rows: City MPG | Dealer Cost | Engine Size (l) | HP | Len

Filters

Marks

All ○

○ Circle

Color | Size | Label

Detail | Tooltip

Color: Class

Detail: Vehicle Name

SUM(City MP... ○

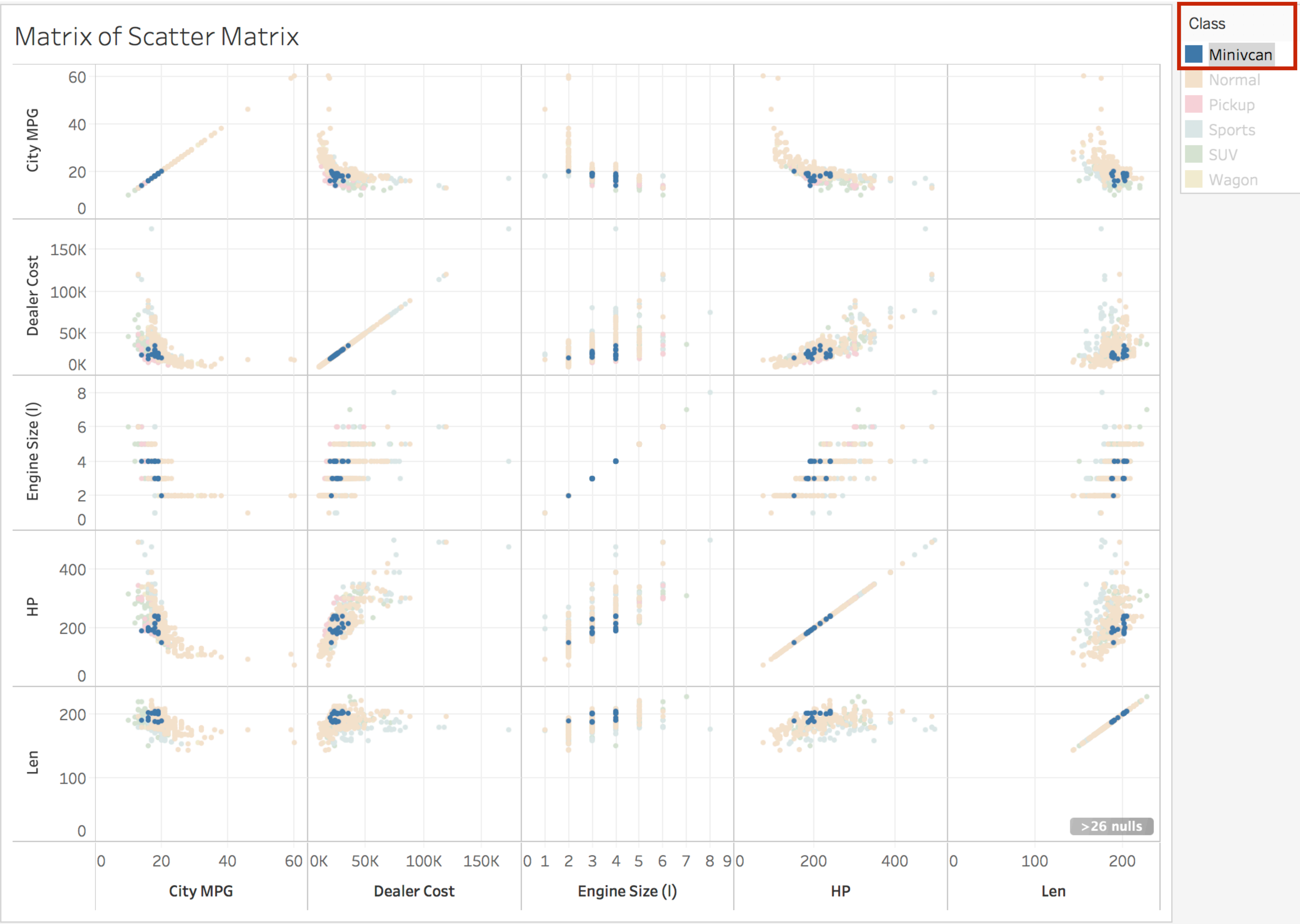
SUM(Dealer ... ○

SUM(Engine ... ○

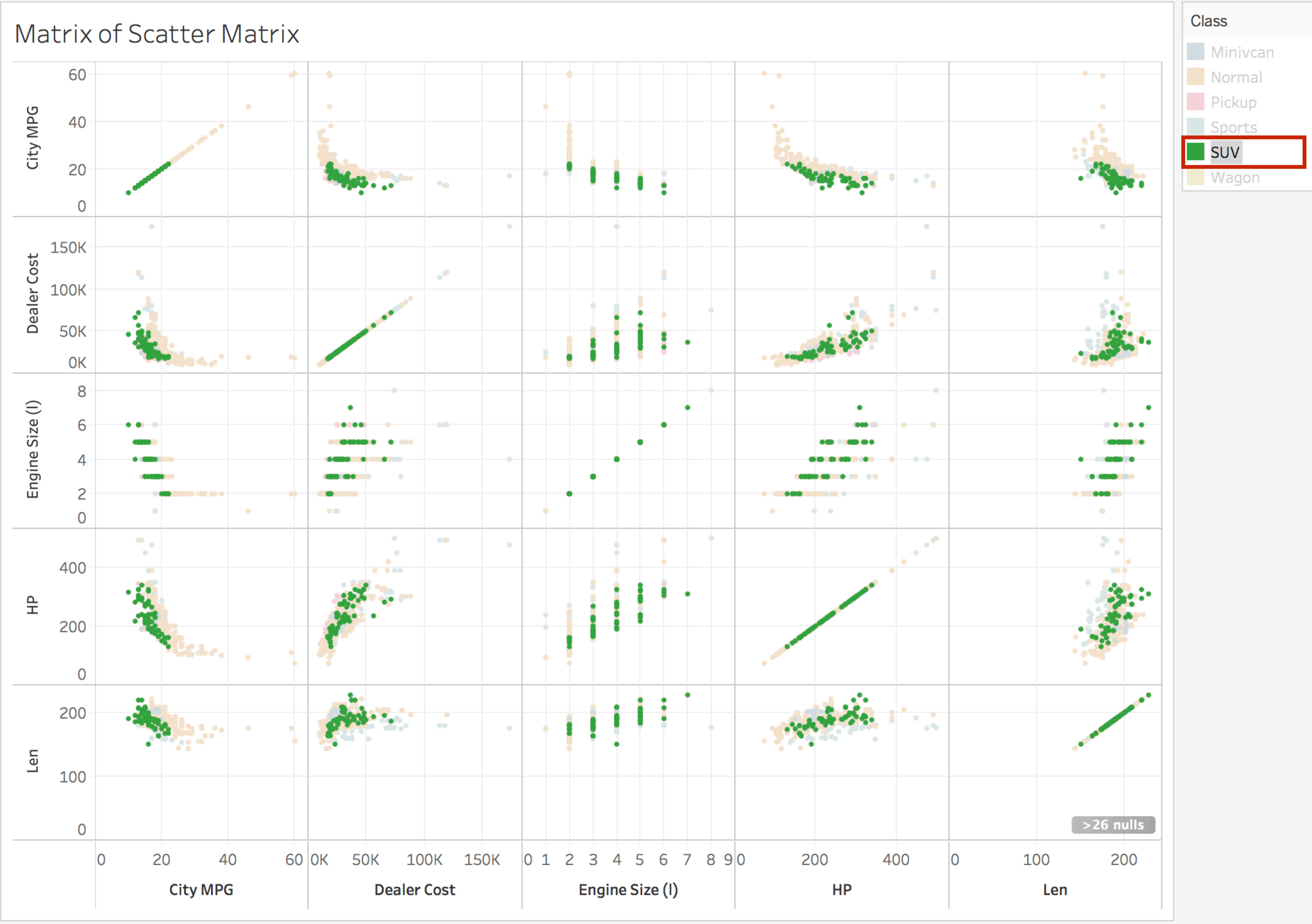
SUM(HP) ○

SUM(Len) ○

Matrix of Scatter Matrix



Matrix of Scatter Matrix
Class
Minivcan
Normal
Pickup
Sports
SUV
Wagon
City MPG
Dealer Cost
Engine Size (l)
HP
Len
60
40
20
0
150K
100K
50K
0K
8
6
4
2
0
400
200
0
200
100
0
0 20 40 60
0K 50K 100K 150K
0 1 2 3 4 5 6 7 8 9
0 200 400
0 100 200
>26 nulls

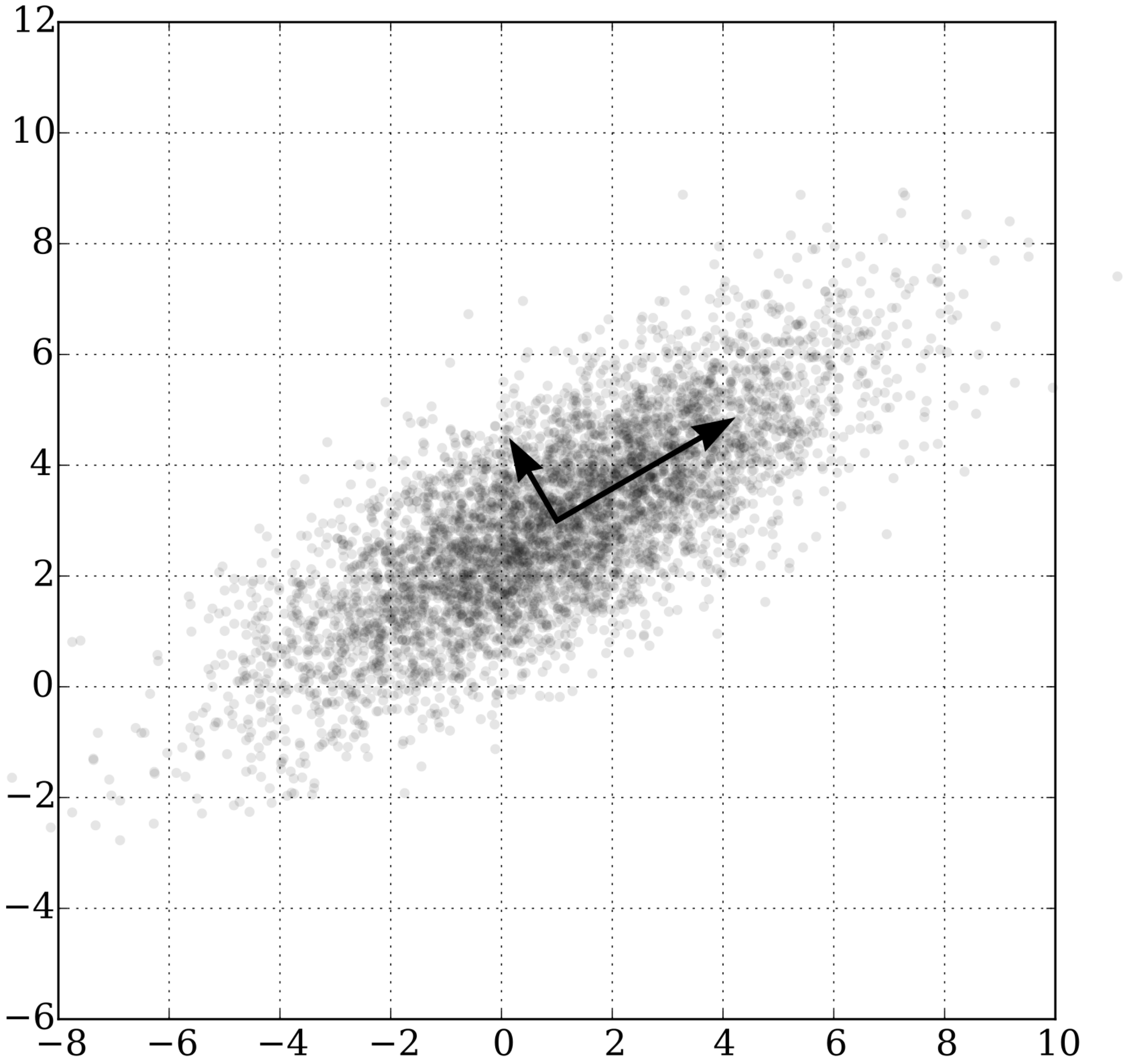


Matrix of Scatter Matrix

Multivariate Data: Point-Based Techniques

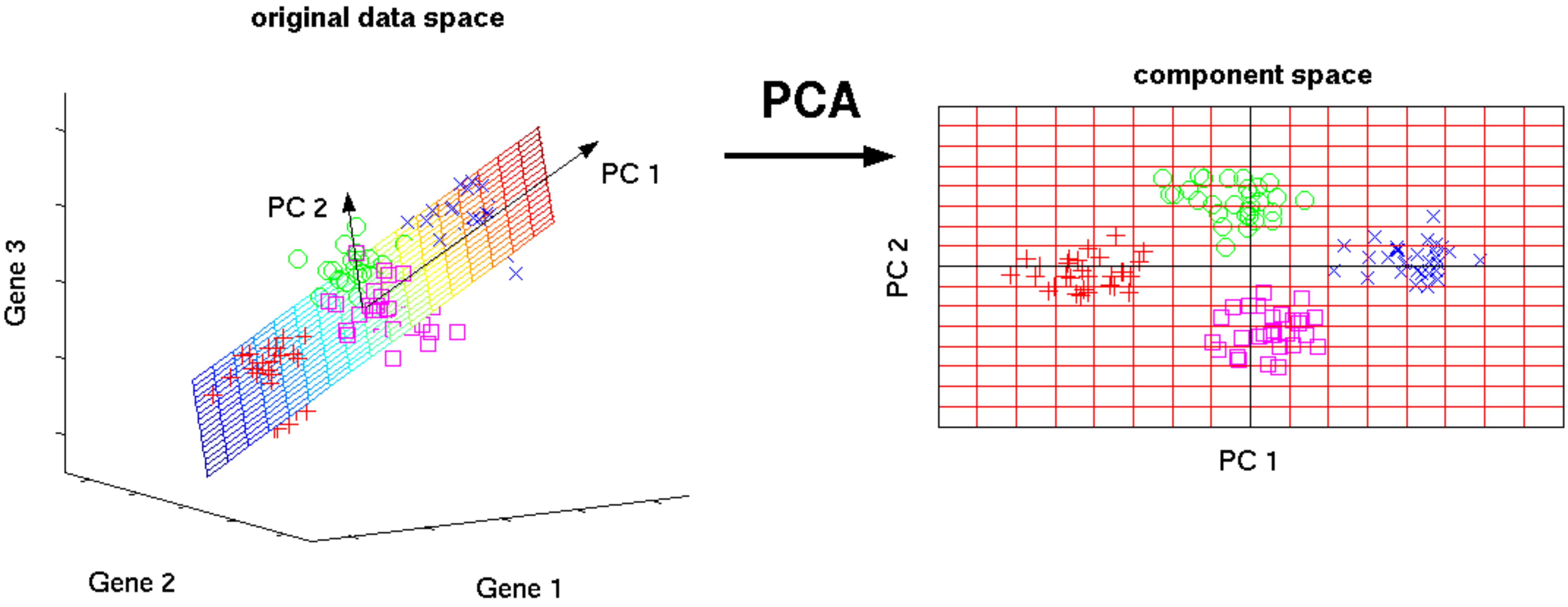
- **In situations where the dimensionality of the data exceeds the capabilities of the visualization technique. It is necessary to investigate ways to reduce the data dimensionality, while at the same time preserving, as much as possible, the information contained within.**
- **Principal Component Analysis (PCA) -** read more **and see this** implementation
- **Multidimensional Scaling (MDS) -** read more **and** more
- **Non-linear dimension reduction techniques:**
  - **Self-organizing Maps (SOMs) -** read more
  - **Local Linear Embeddings (LLE) -** read more
  - **t-distributed Stochastic Neighbor Embedding (t-SNE) -** read more

Principal Component Analysis (PCA)



https://en.wikipedia.org/wiki/Principal_component_analysis

Principal Component Analysis (PCA)



http://www.nlpca.org/pca_principal_component_analysis.html

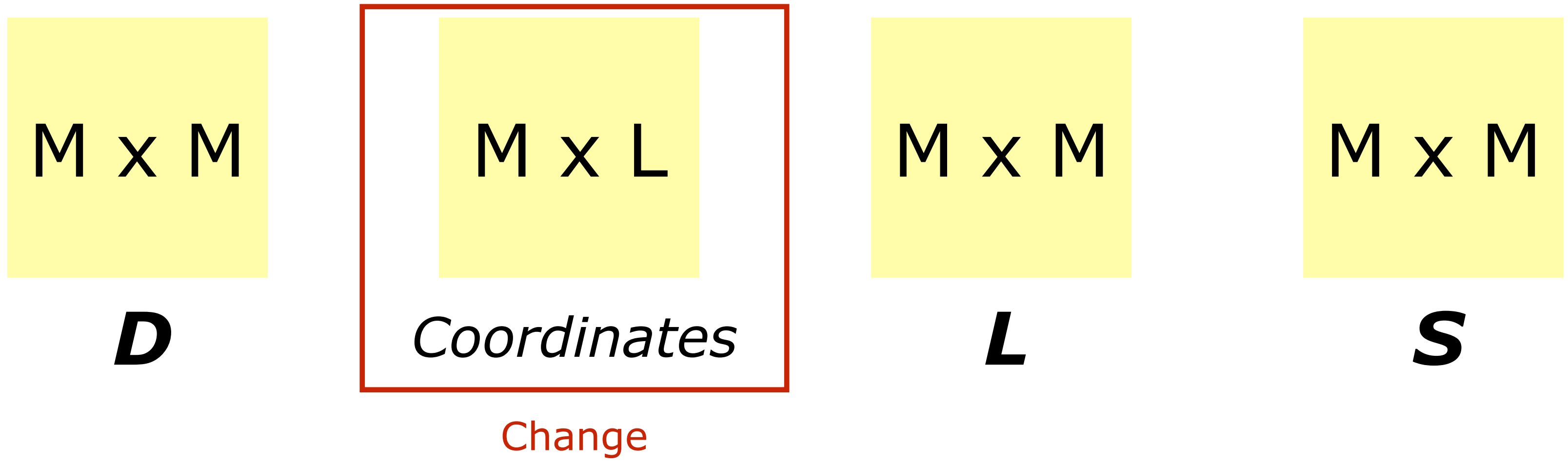
Multidimensional scaling (MDS)

- Projecting M points in N dimensions into L dimensions (L=2 or 3) display space.
- The key goal is to attempt to maintain the N-dimensional features and characteristics of the data through the projection process, e.g., relationships that exist in the original data must also exist after projection.
  - The projection may also unintentionally introduce artifacts that may appear in the visualization and are not present in the data.

1. Create a Similarity M x M Matrix (*D*) (could be distance)
2. Create a coordinates Matrix M x L and fill randomly or other method (ex: PCA)
3. Create an M x M matrix (*L*) based on L coordinates. Also a similarity matrix.
4. Repeat
    1. Compute the stress matrix S (an M x M Matrix), as the difference between D and L.
    2. Shift the positions of points in L-dimensional space (their L coordinates) in a direction that will reduce their individual stress levels.
5. Until S is small or has not changed significantly

Multidimensional scaling (MDS)

- **Projecting M points in N dimensions into L dimensions (L=2 or 3) display space.**



1. **Create a Similarity M x M Matrix (*D*) (could be distance)**
2. **Create a coordinates Matrix M x L and fill randomly or other method (ex: PCA)**
3. **Create an M x M matrix (*L*) based on L coordinates. Also a similarity matrix.**
4. **Repeat**
    1. **Compute the stress matrix *S* (an M x M Matrix), as the difference between *D* and *L*.**
    2. **Shift the positions of points in L-dimensional space (their L coordinates) in a direction that will reduce their individual stress levels.**
5. **Until S is small or has not changed significantly**

Multidimensional scaling (MDS)

- There are many possible variants on this algorithm, including:
  - Different similarity and stress measures
  - Different initial and termination conditions
  - Different position update strategies
- As in any optimization process, there is the potential to fall into a local minimal configuration that still has a high level of stress.
  - Common strategies to alleviate this include occasionally adding a random jump in the position of a point to see if it will converge to a different location
- Obviously, the results are not unique: minor changes in the starting conditions can lead to dramatically different results.

Multivariate Data: Point-Based Techniques

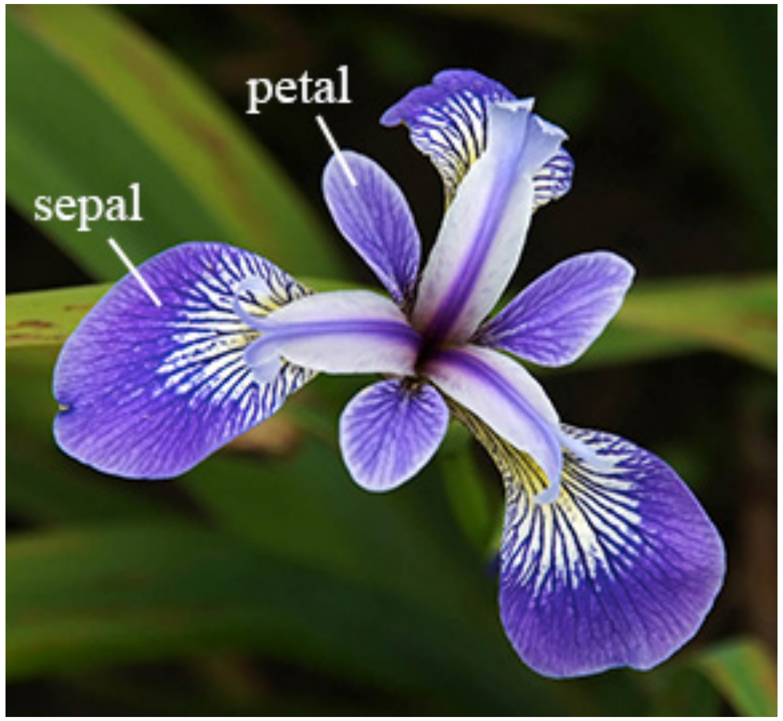
- **Iris flower data set**

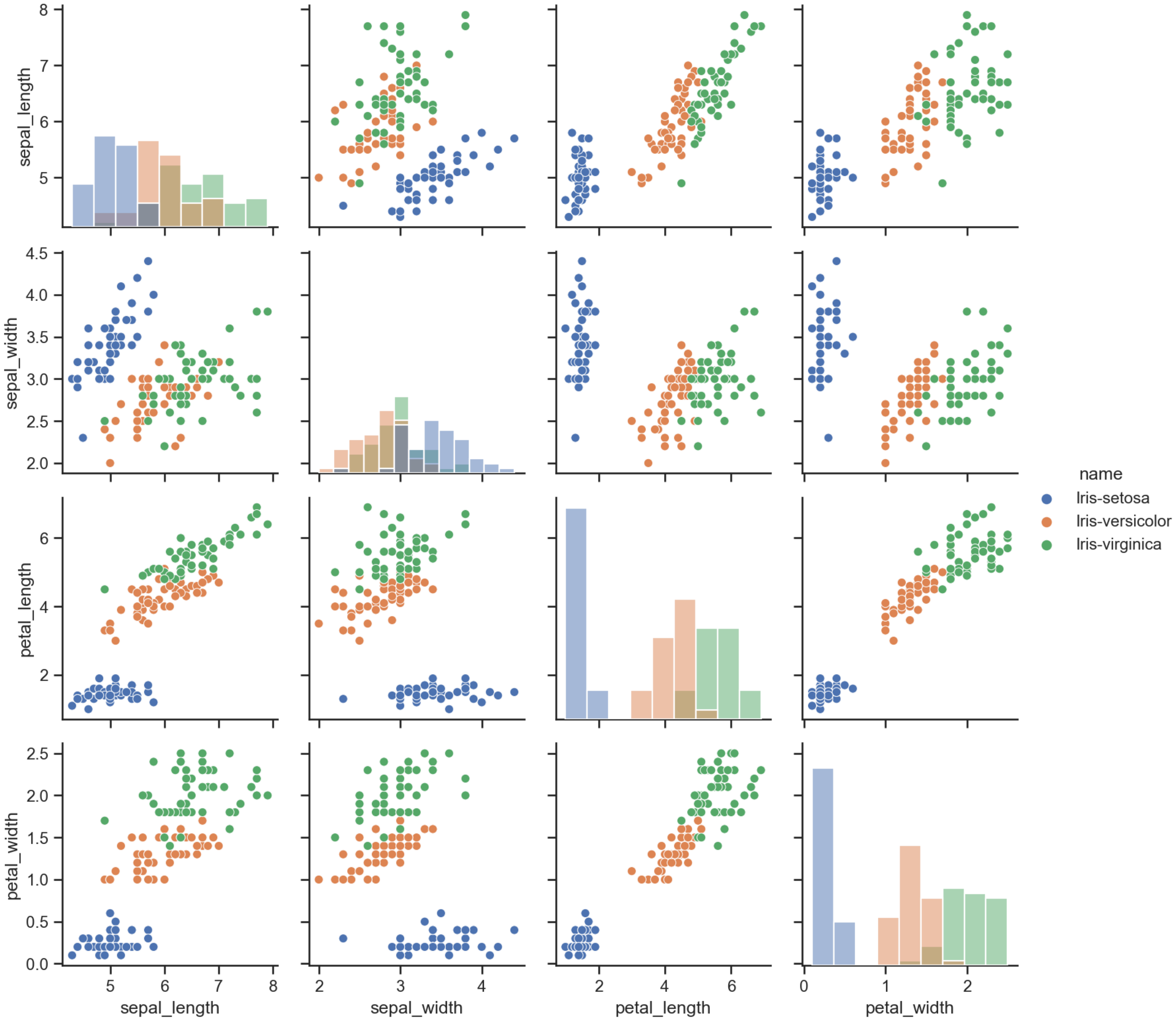
Iris setosa

Iris versicolor

Iris virginica

Multivariate Data: Point-based Techniques

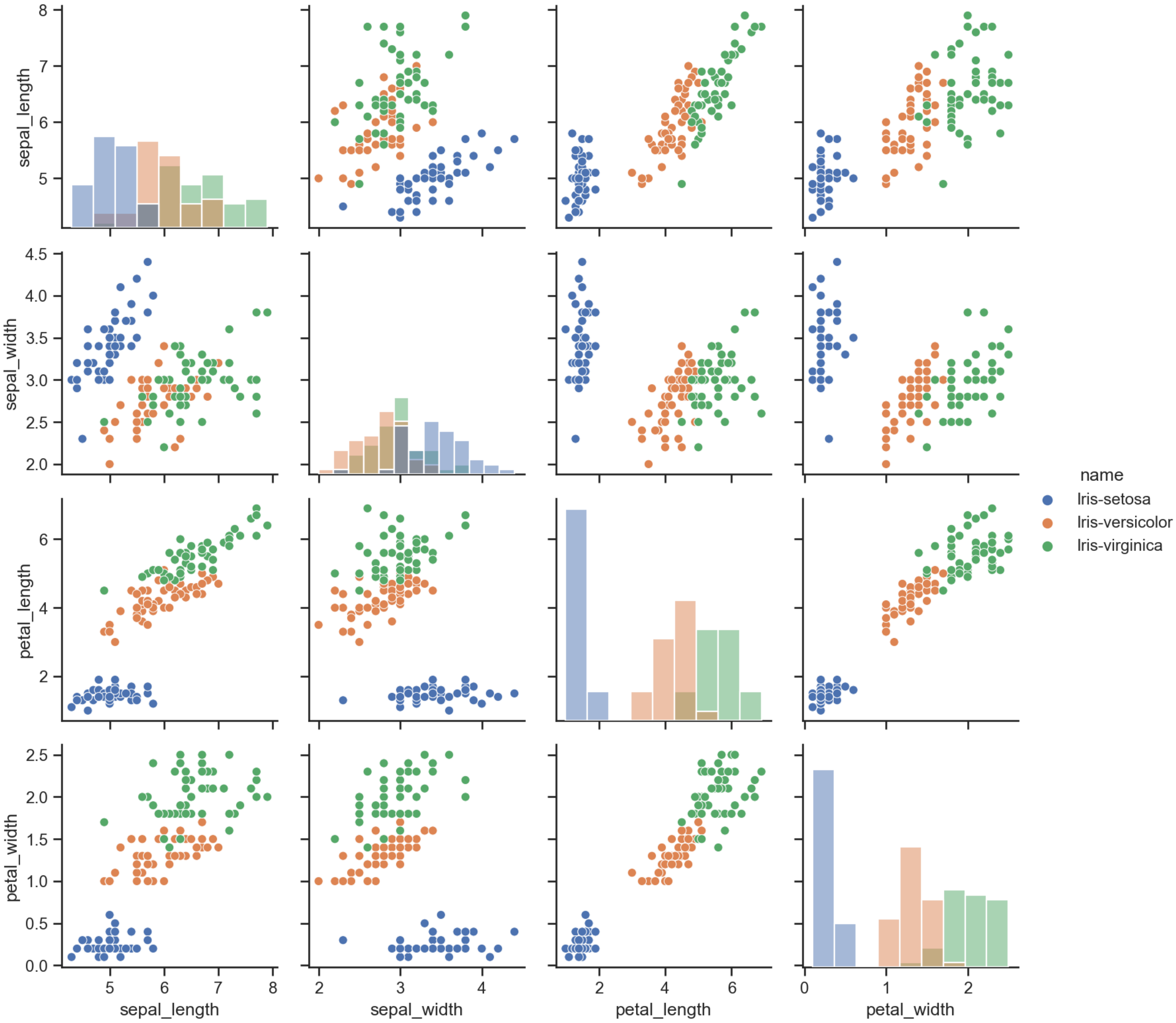




Iris setosa

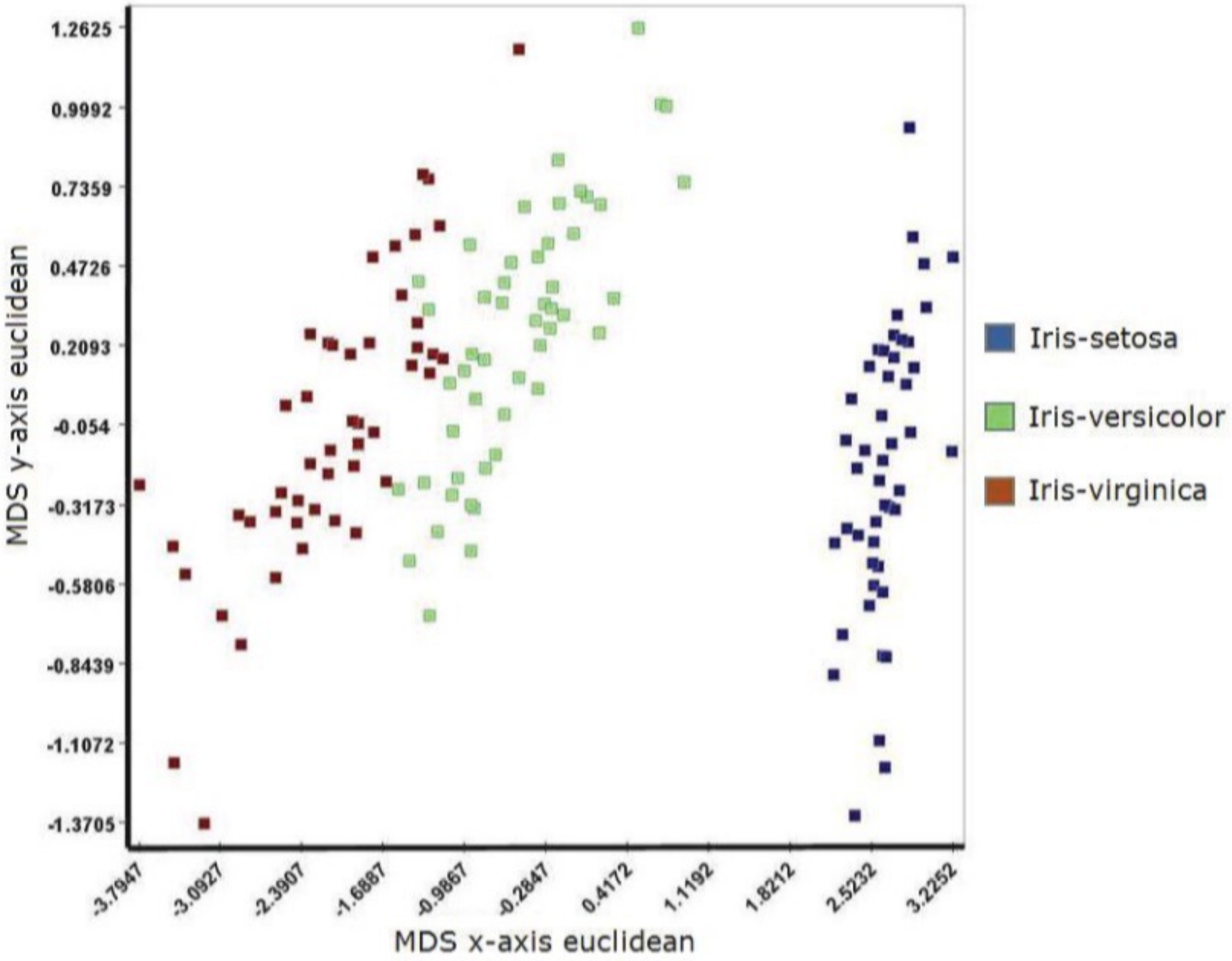
Iris versicolor

Iris virginica


sepal_length
sepal_width
petal_length
petal_width
8
7
6
5
4.5
4.0
3.5
3.0
2.5
2.0
6
4
2
2.5
2.0
1.5
1.0
0.5
0.0
5
6
7
8
2
3
4
2
4
6
0
1
2
sepal_length
sepal_width
petal_length
petal_width
name
Iris-setosa
Iris-versicolor
Iris-virginica

Multivariate Data: Point-Based Techniques

- **Iris data set projected using MDS**



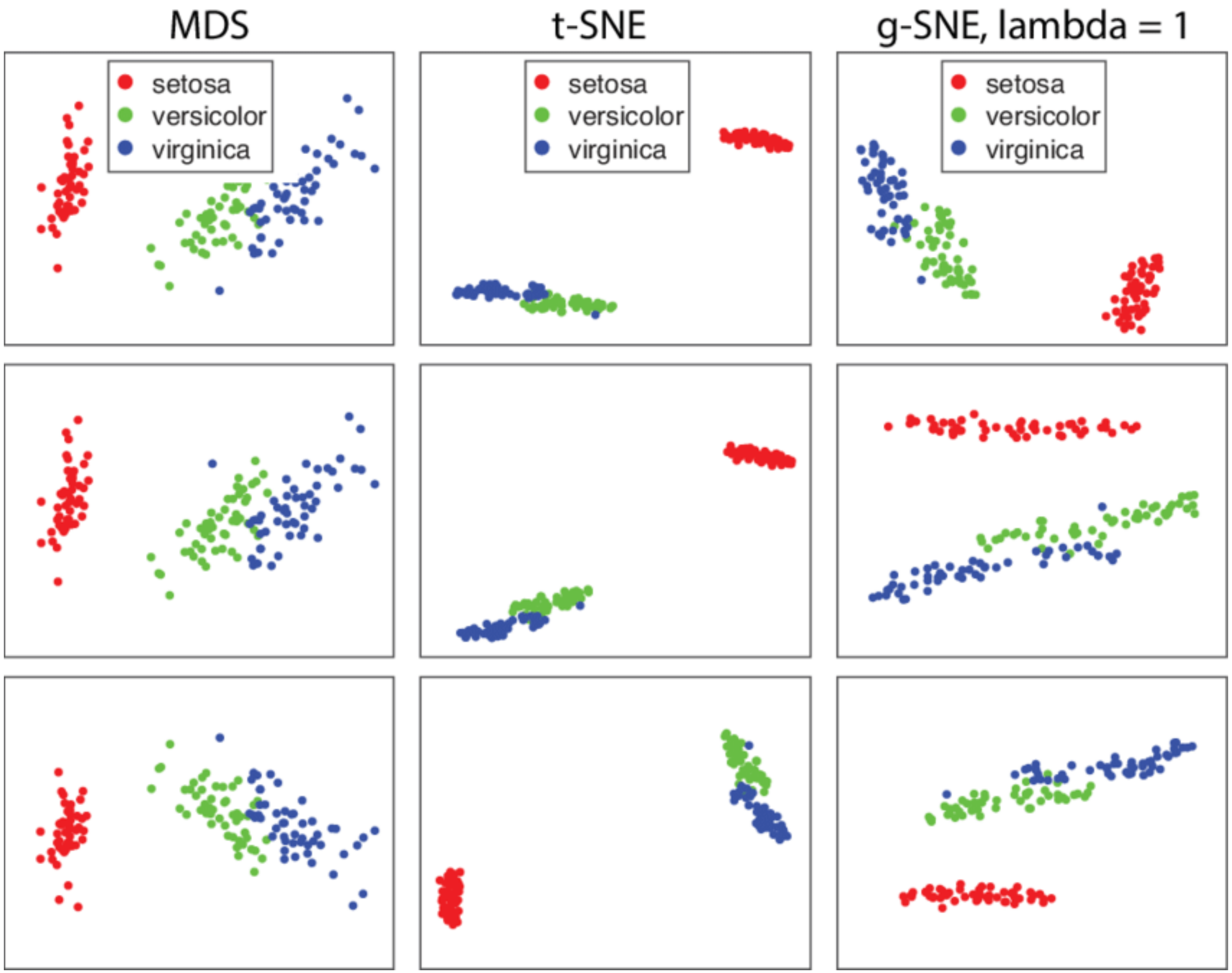
Multivariate Data: Point-Based Techniques

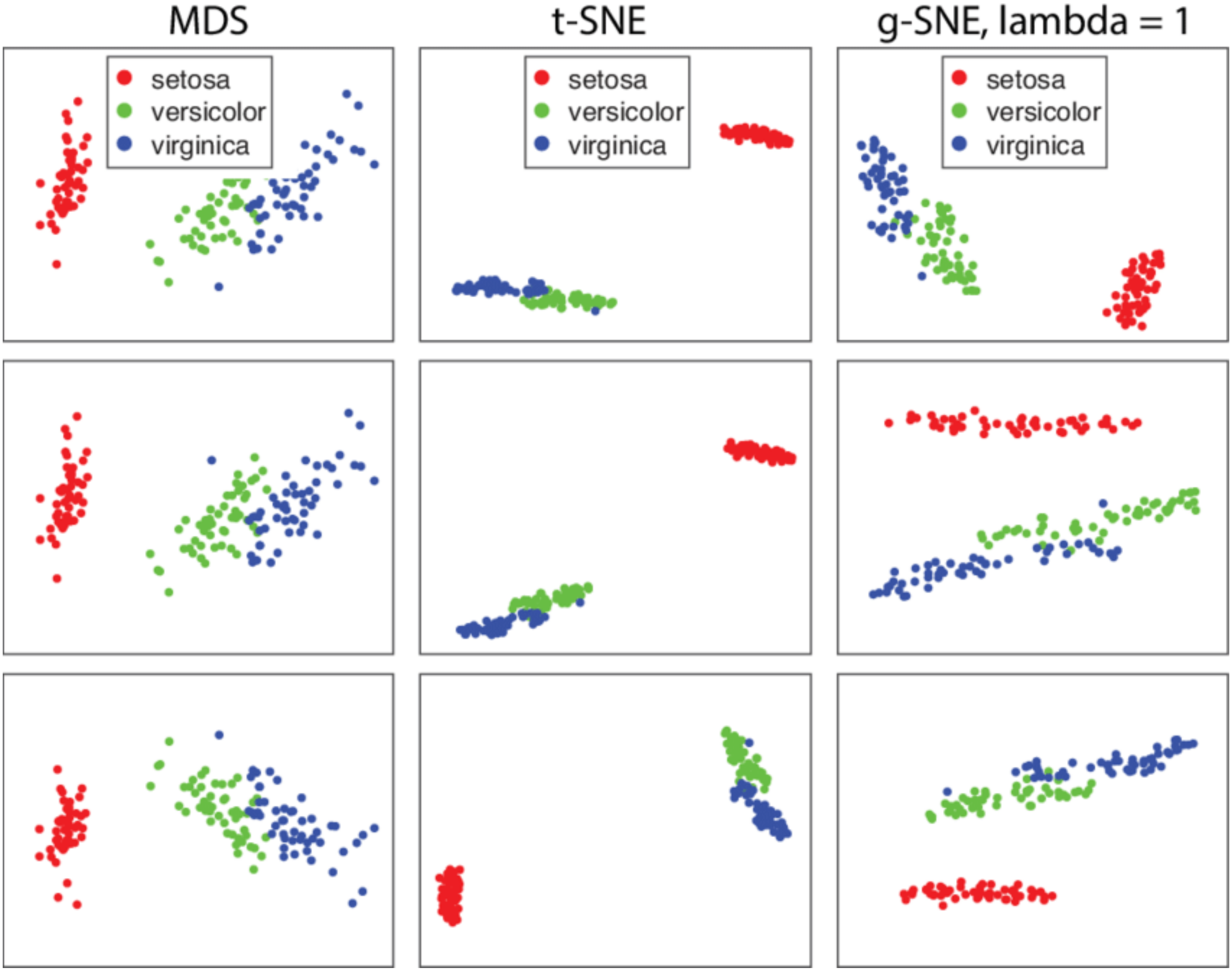
- **In situations where the dimensionality of the data exceeds the capabilities of the visualization technique. It is necessary to investigate ways to reduce the data dimensionality, while at the same time preserving, as much as possible, the information contained within.**
- **Principal Component Analysis (PCA) -** read more **and see this** implementation
- **Multidimensional Scaling (MDS) -** read more **and** more
- **Non-linear dimension reduction techniques:**
  - **Self-organizing Maps (SOMs) -** read more
  - **Local Linear Embeddings (LLE) -** read more
  - **t-distributed Stochastic Neighbor Embedding (t-SNE) -** read more

Applying to iris data set

Link

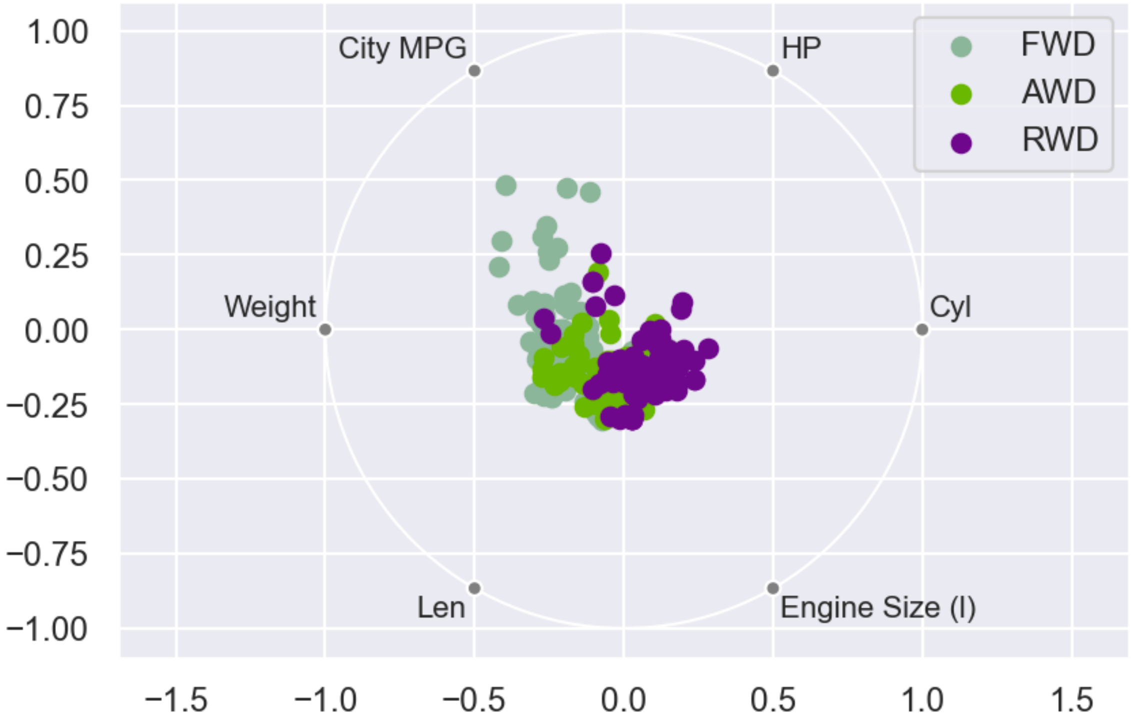
Using global t-SNE to preserve inter-cluster data structure




MDS
t-SNE
g-SNE, lambda = 1
setosa
versicolor
virginica

Multivariate Data: Point-based Techniques

- RadViz: is a force-driven point layout technique that is based on Hooke-s Law for equilibrium.
- For an N-dimensional data set, N anchor points are placed on the circumference of the circle to represent the fixed ends of the N springs attached to each data point.
- Different placement and ordering of the anchors will give different results, and points that are quite distinct in N dimensions may map to the same location in 2D.

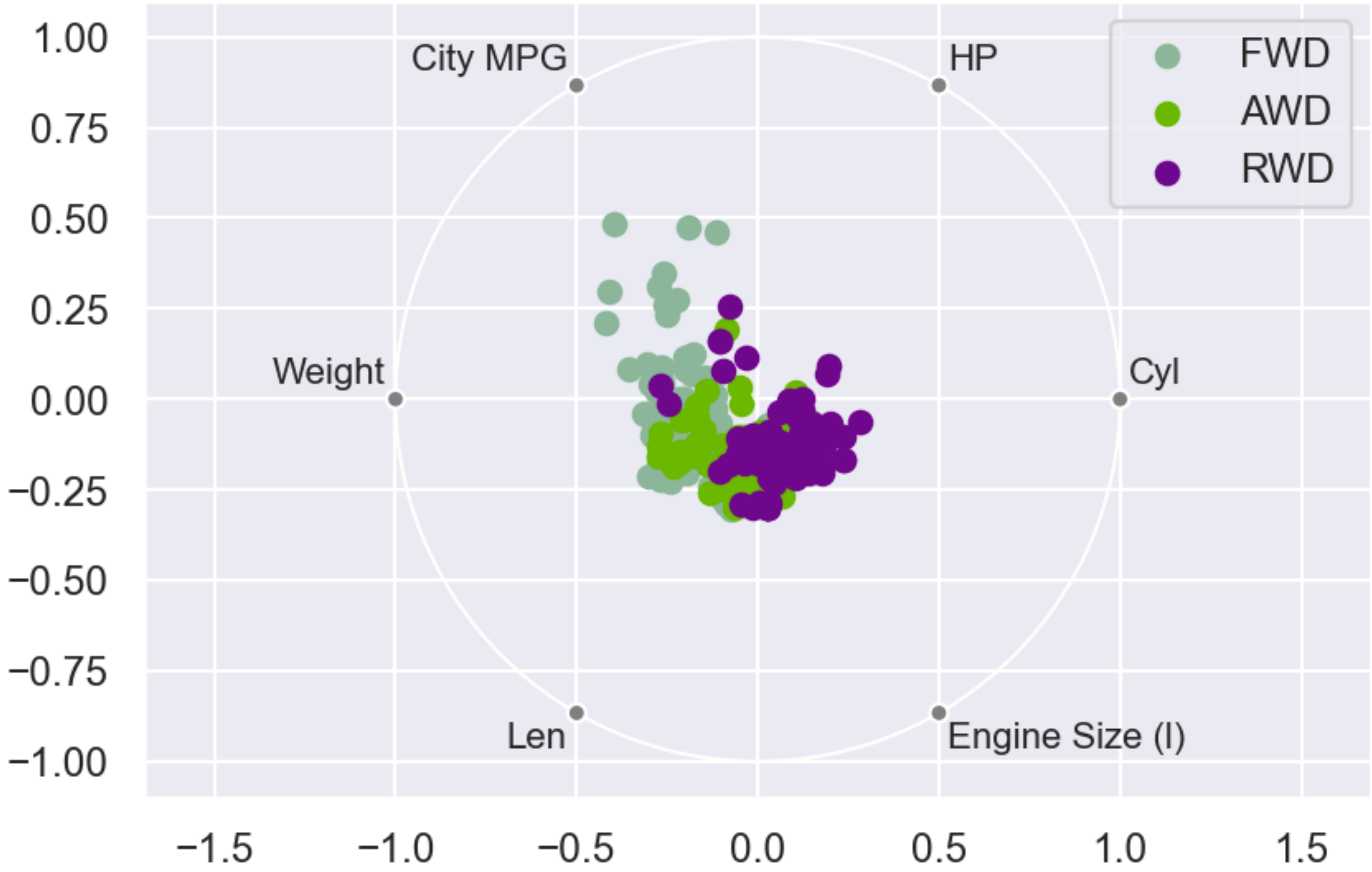


DIMENSIONAL ANCHORS: A GRAPHIC PRIMITIVE FOR MULTIDIMENSIONAL MULTIVARIATE INFORMATION VISUALIZATIONS, Patrick Hoffman, Georges G. Grinstein

Visualizing Multivariate Data with Radviz

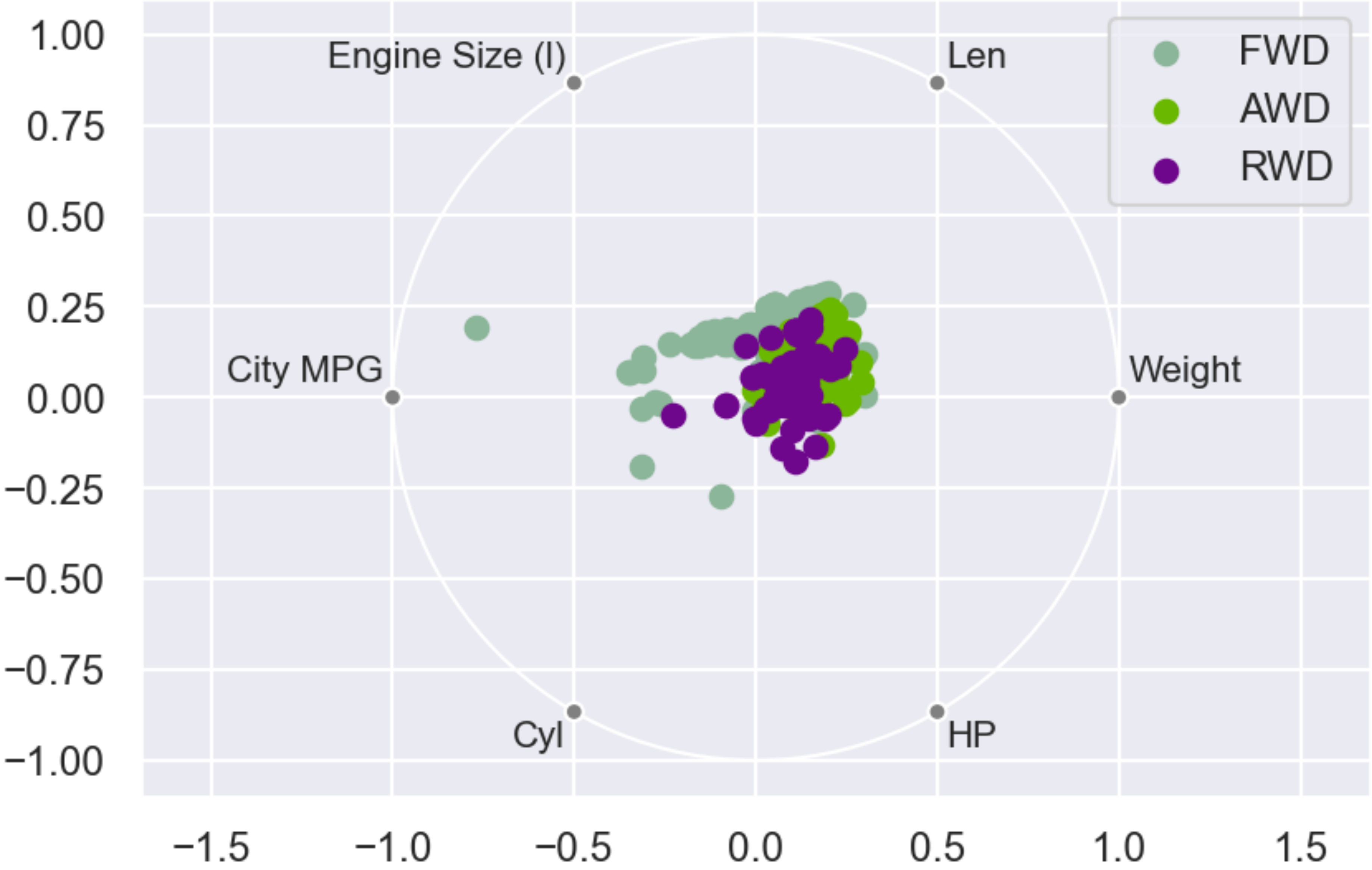
Multivariate Data: Point-based Techniques

- **RadViz: different views of the same data set in RadViz, using manual reordering of dimensions.**



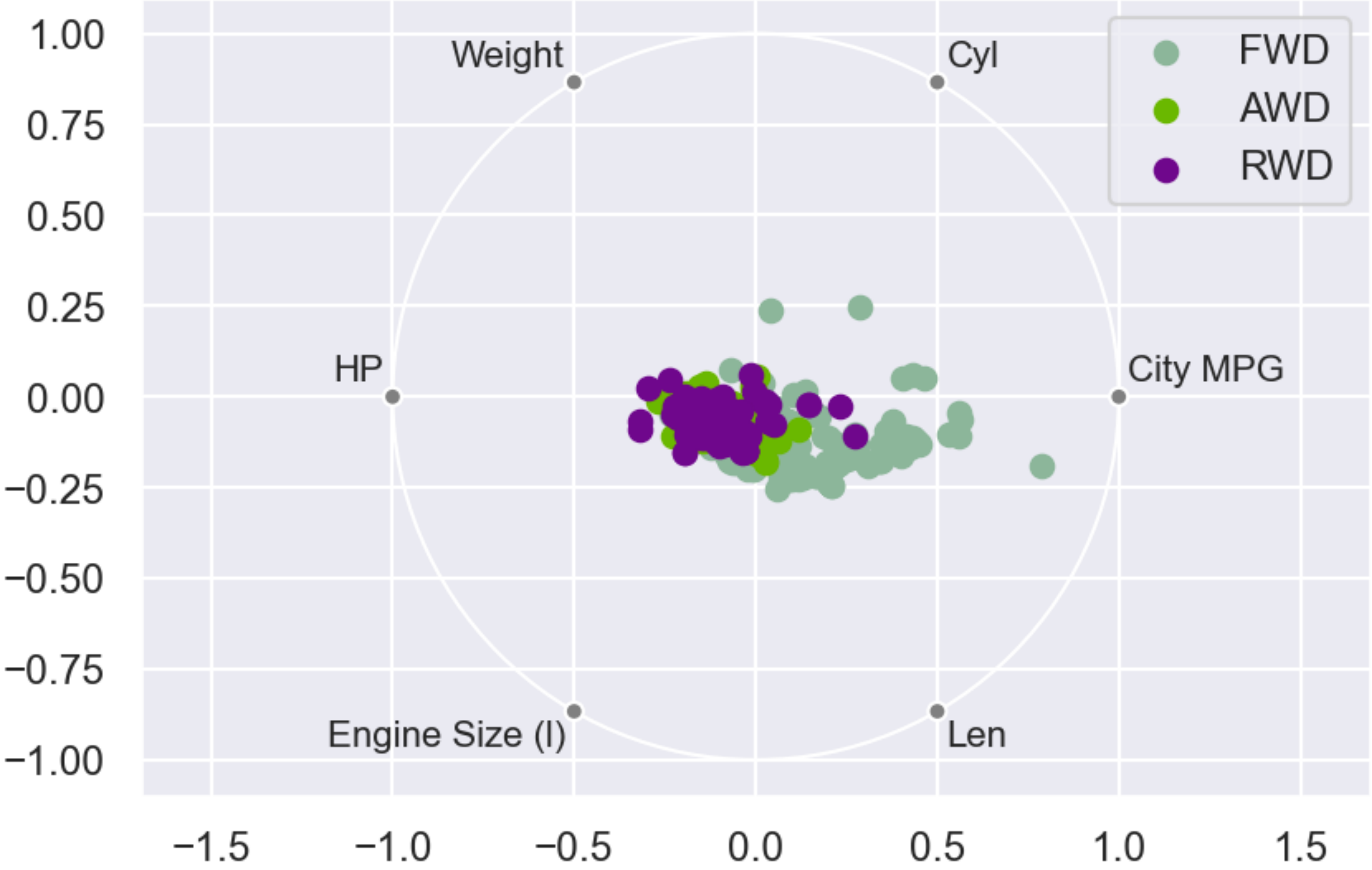
Multivariate Data: Point-based Techniques

- **RadViz: different views of the same data set in RadViz, using manual reordering of dimensions.**



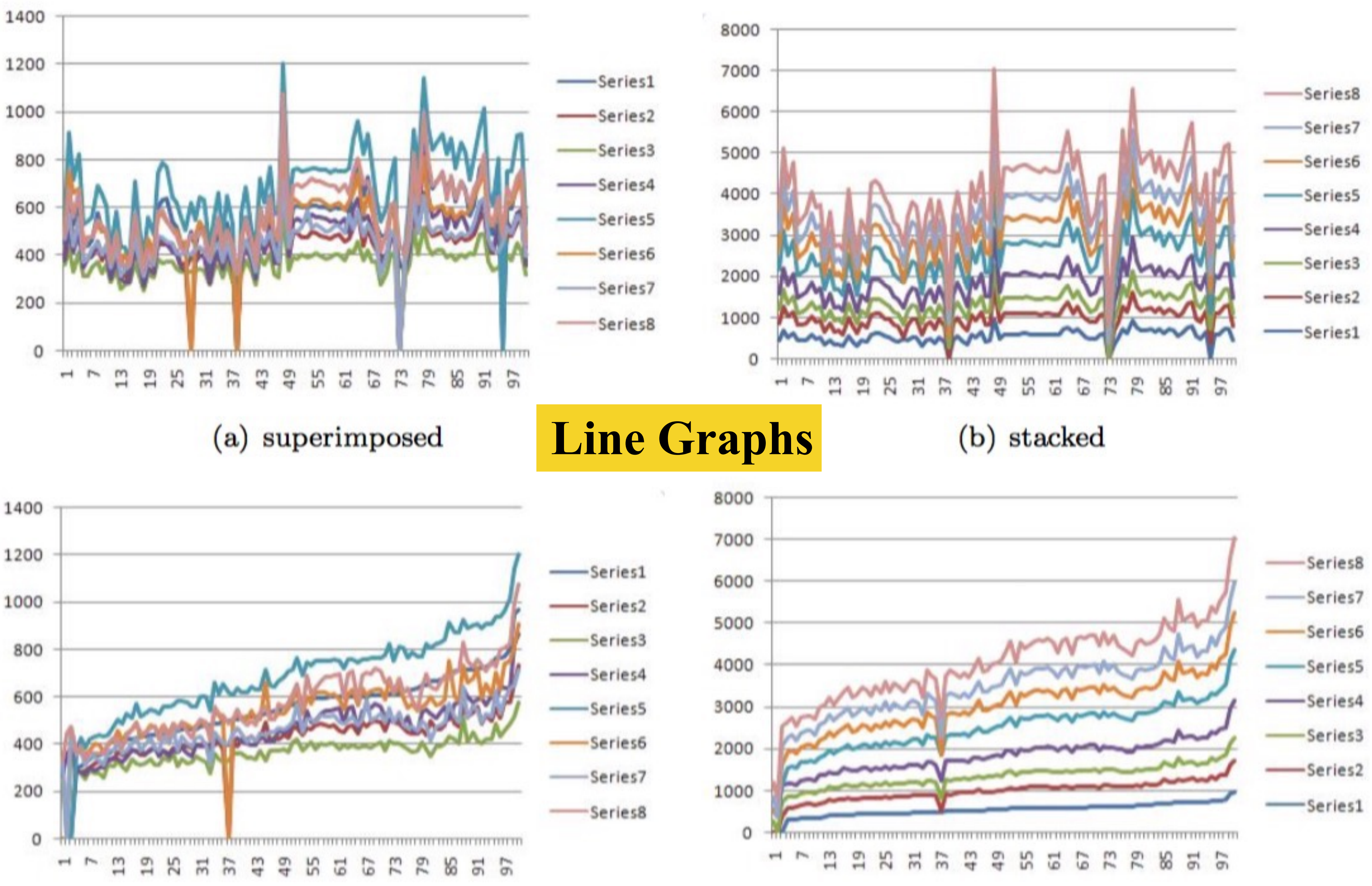
Multivariate Data: Point-based Techniques

- **RadViz: different views of the same data set in RadViz, using manual reordering of dimensions.**



Line-Based Techniques

Multivariate Data: Line-Based Techniques



(a) superimposed

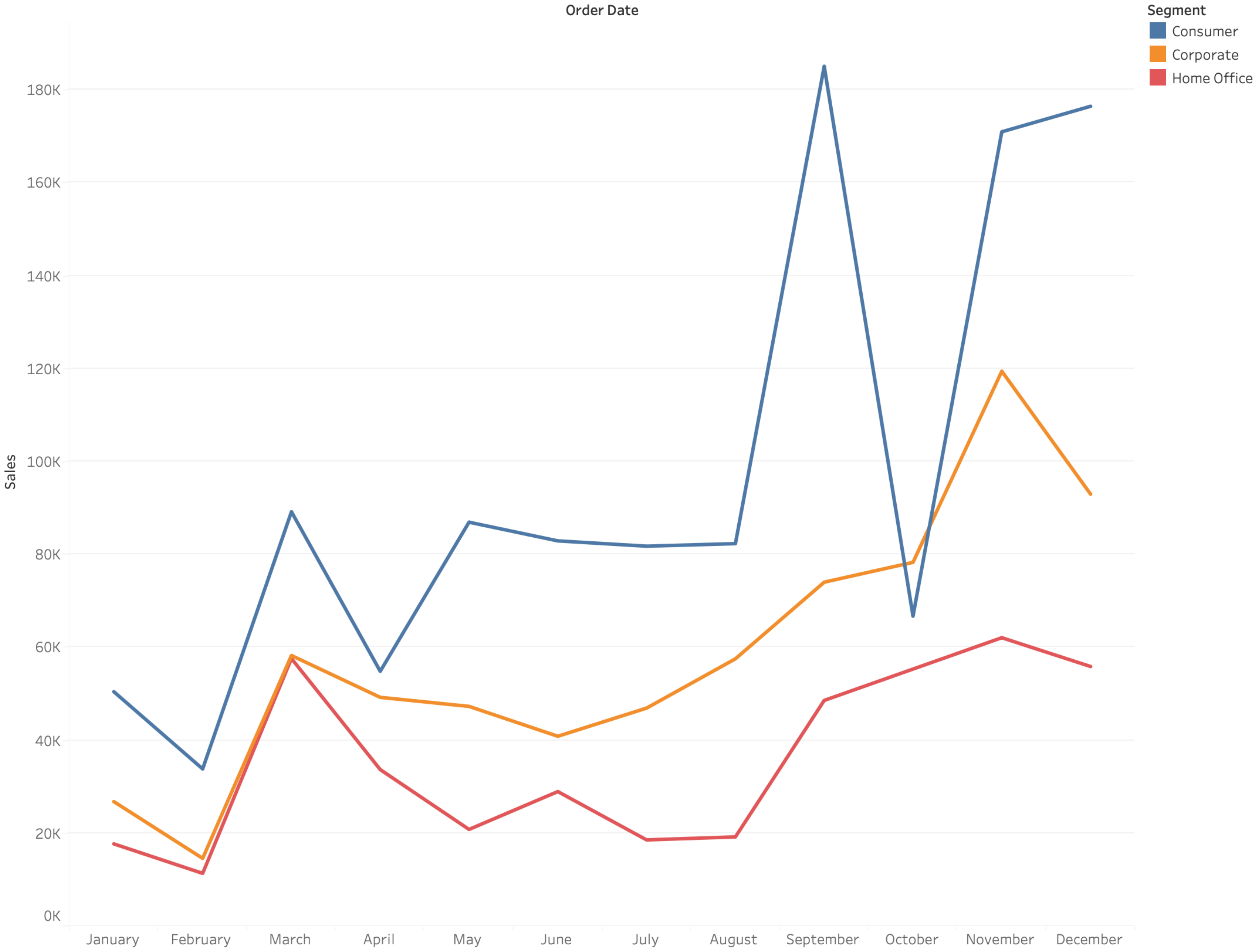
Line Graphs

(b) stacked

(c) ordered superimposed

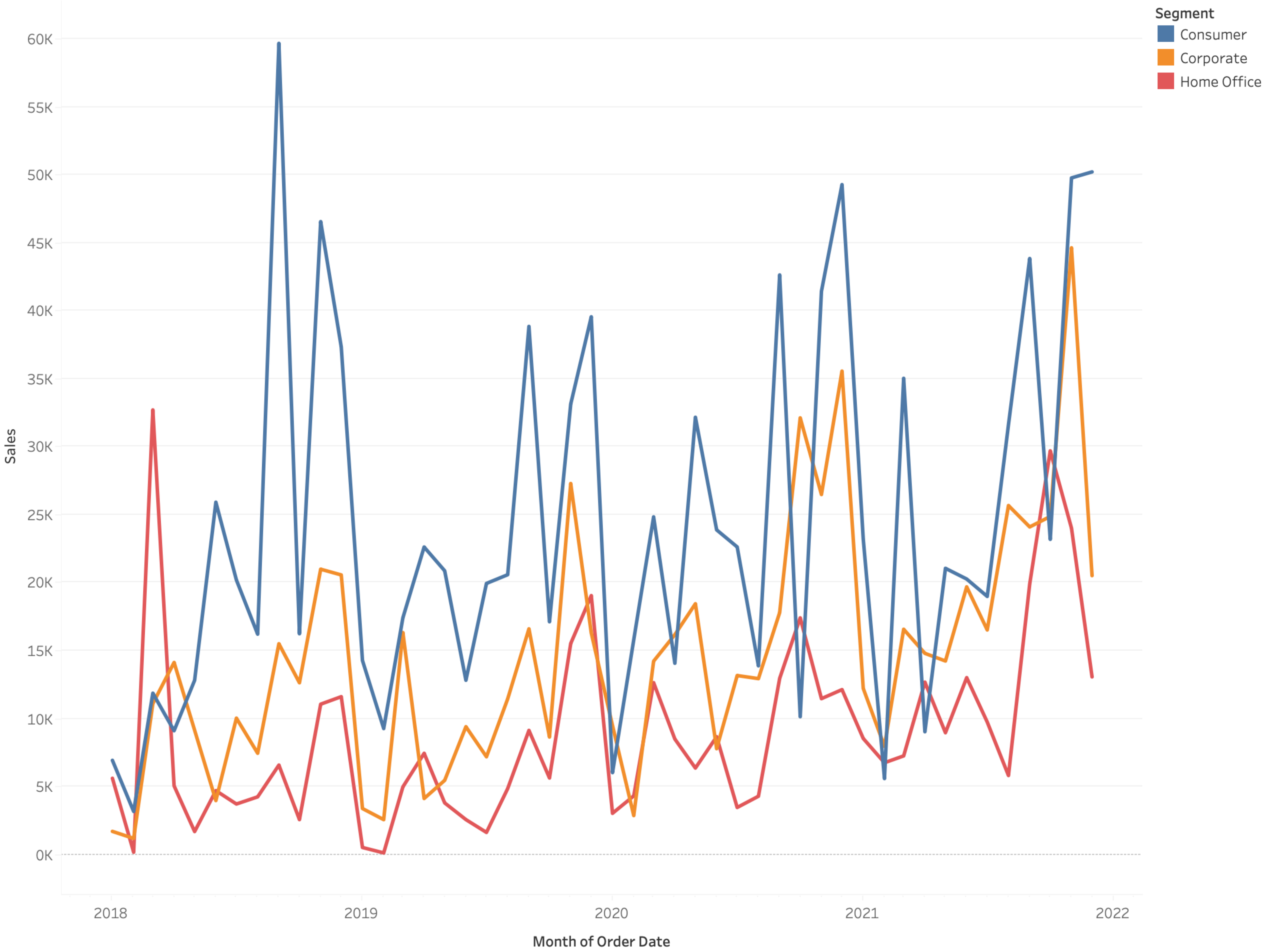
(d) ordered stacked

Superimposed - Discrete



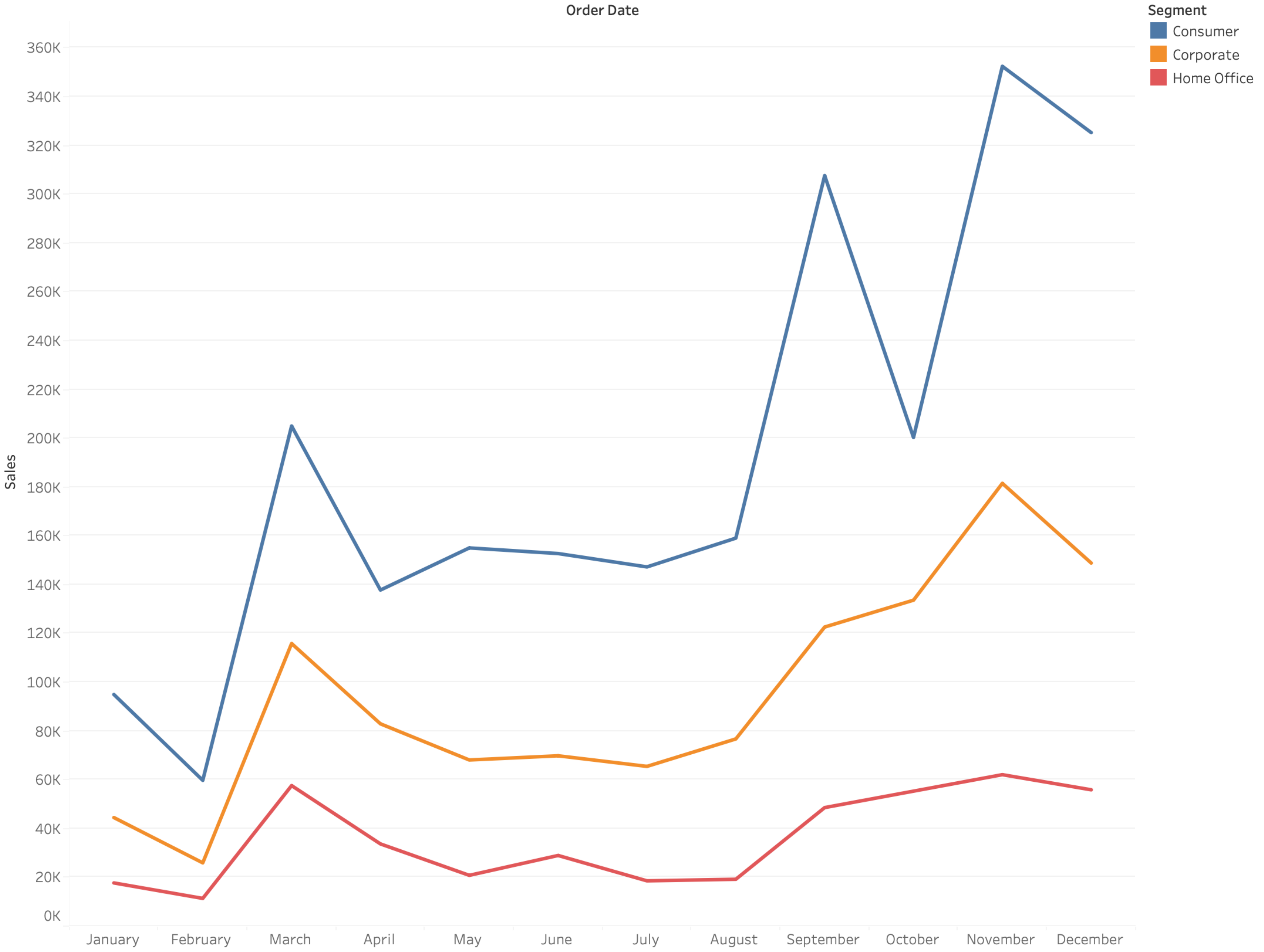
The trend of sum of Sales for Order Date Month. Color shows details about Segment.

Superimposed - Continuous



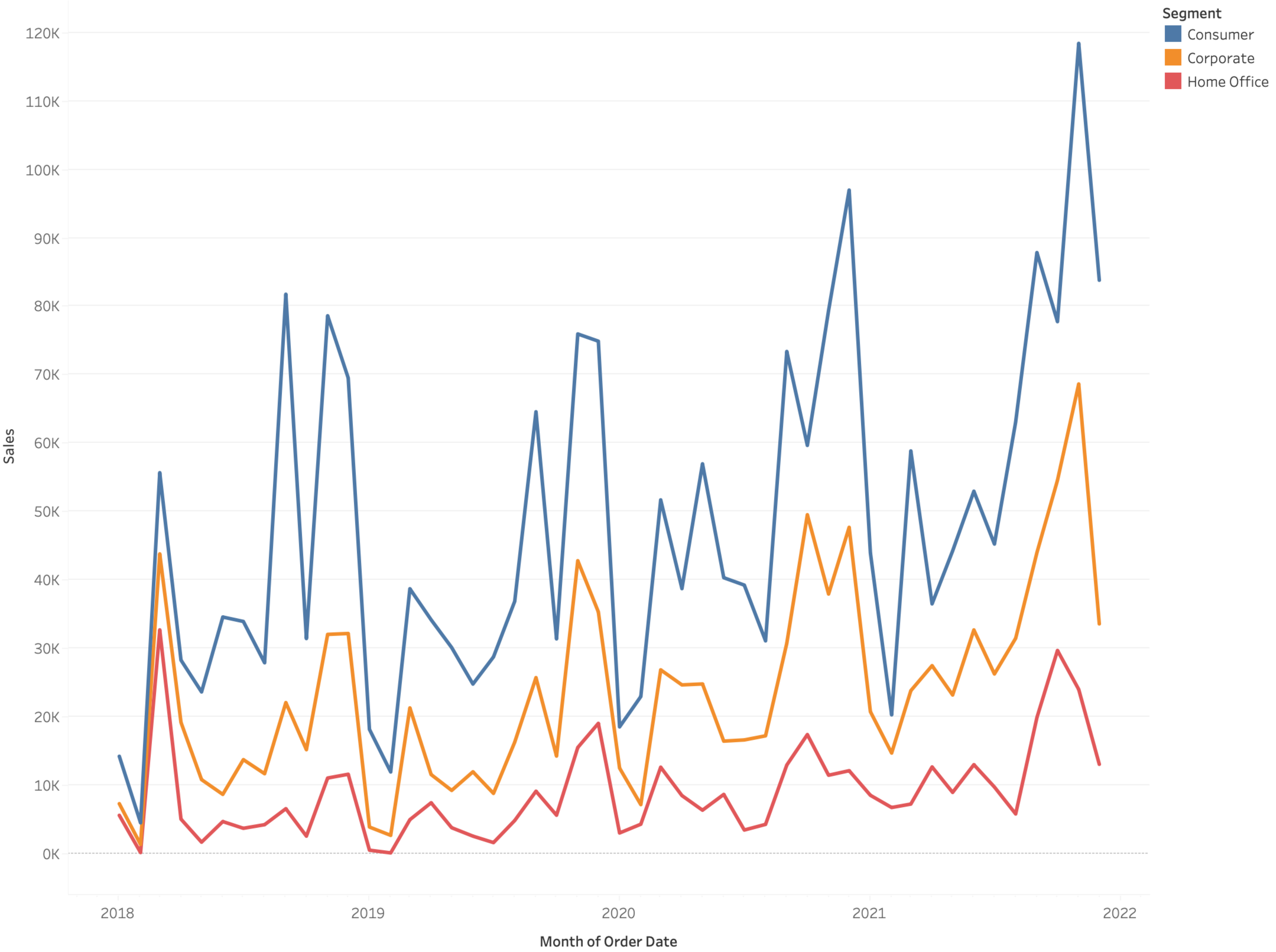
The trend of sum of Sales for Order Date Month. Color shows details about Segment.

Stacked - Discrete



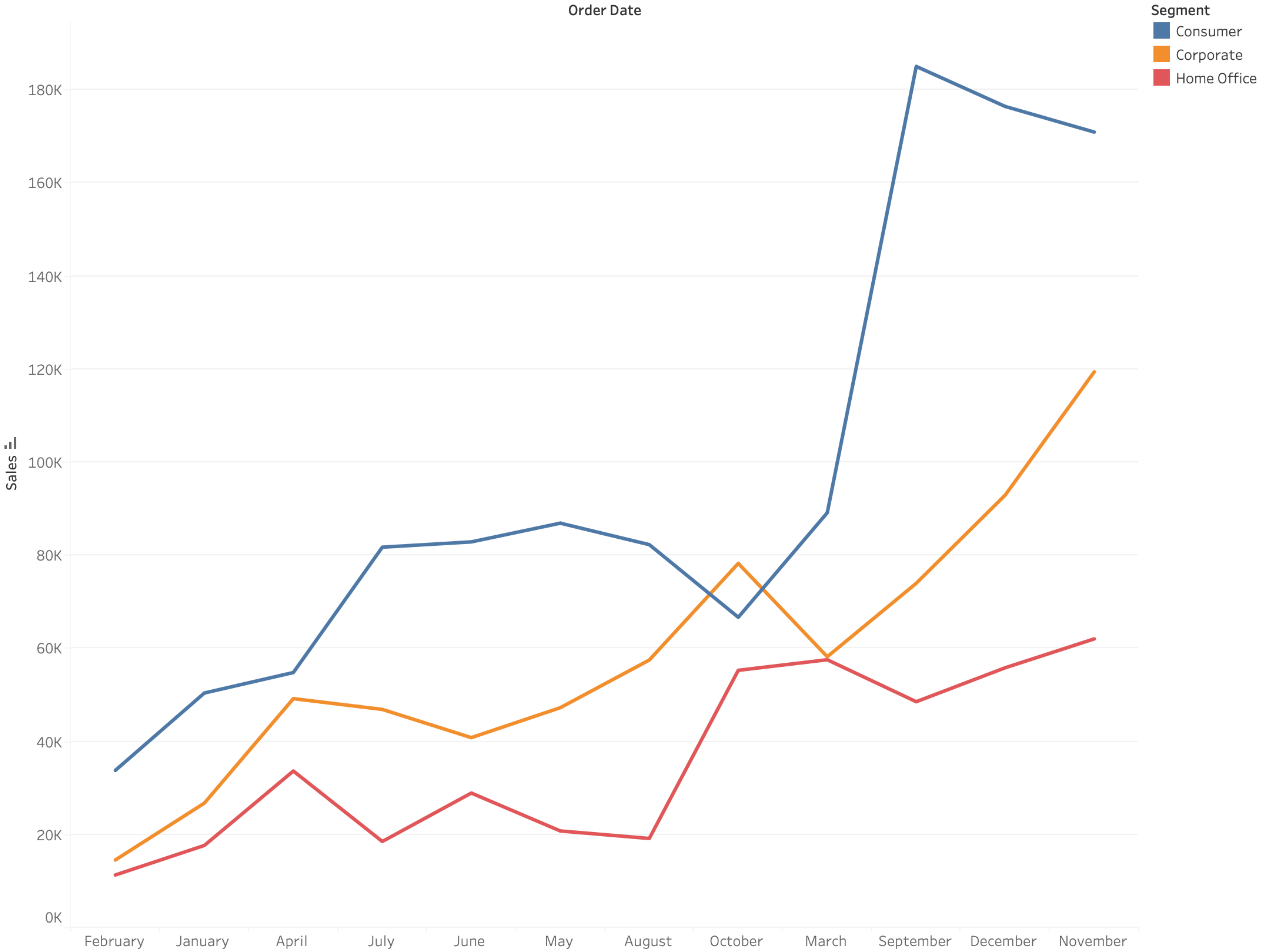
The trend of sum of Sales for Order Date Month. Color shows details about Segment.

Stacked - Continuous



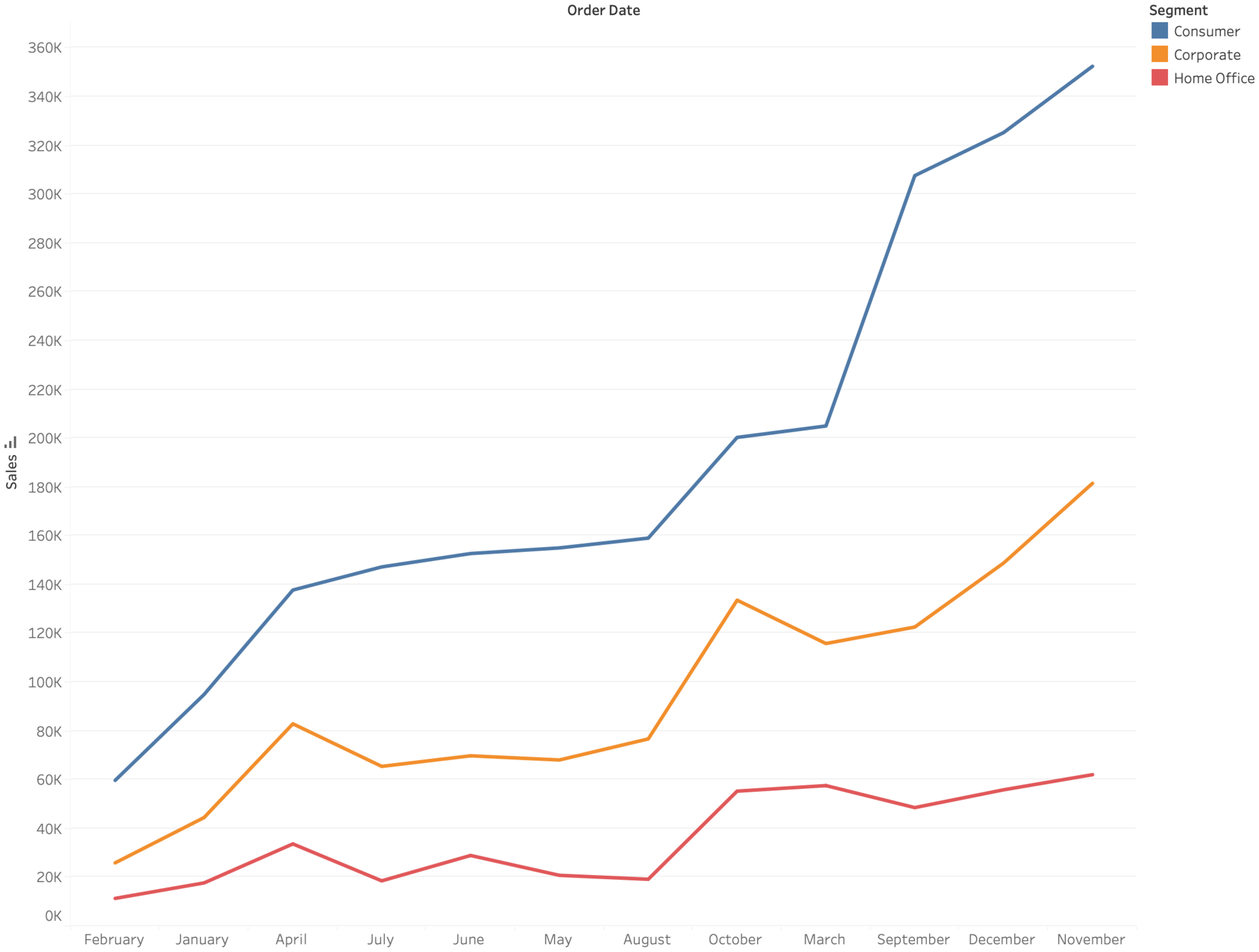
The trend of sum of Sales for Order Date Month. Color shows details about Segment.

Superimposed - Discrete - Ordered



The trend of sum of Sales for Order Date Month. Color shows details about Segment.

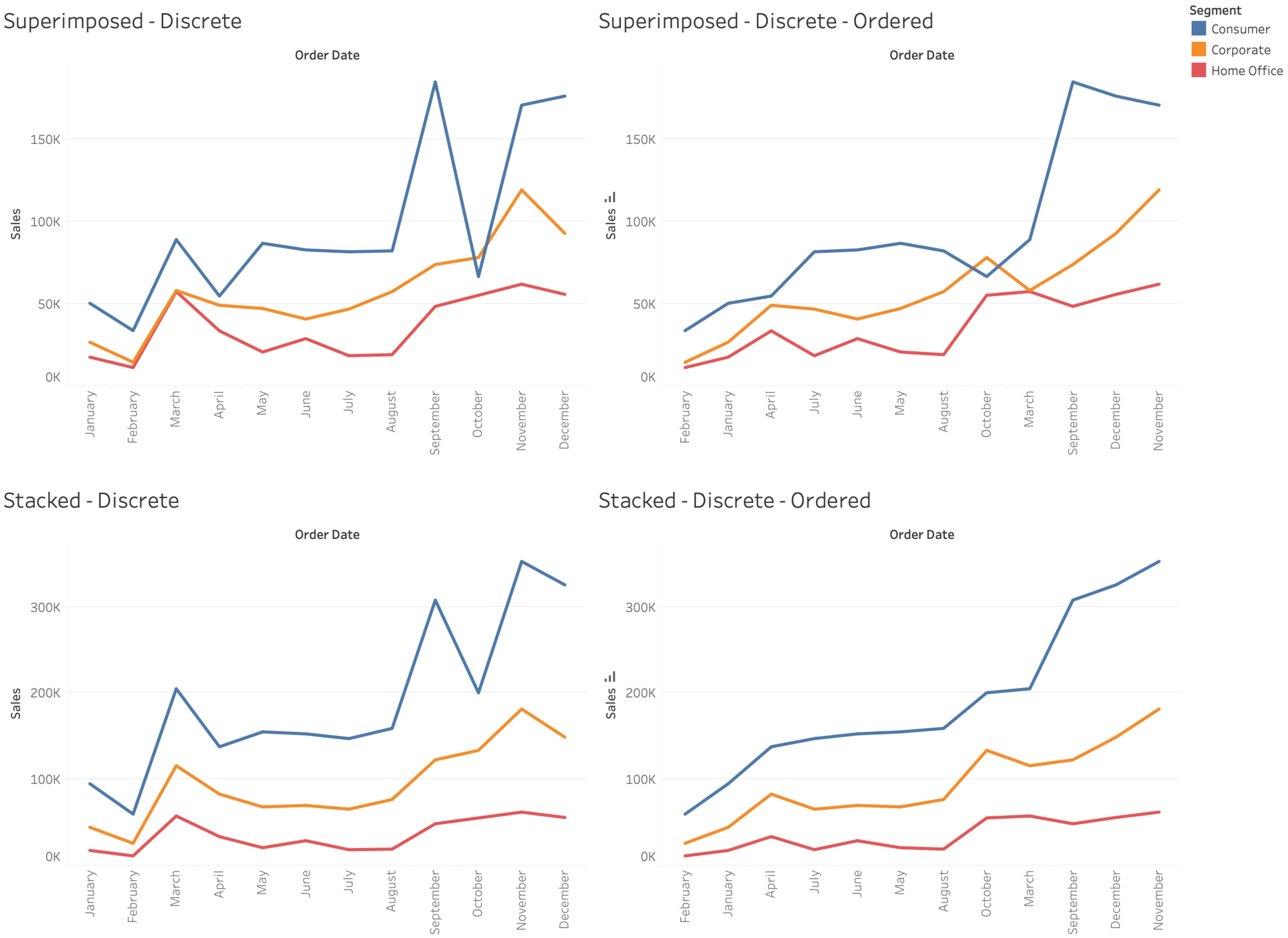
Stacked - Discrete - Ordered



The trend of sum of Sales for Order Date Month. Color shows details about Segment.

Multivariate Data: Line-Based Techniques

- **When Superimpose?**


Superimposed - Discrete

Superimposed - Discrete - Ordered

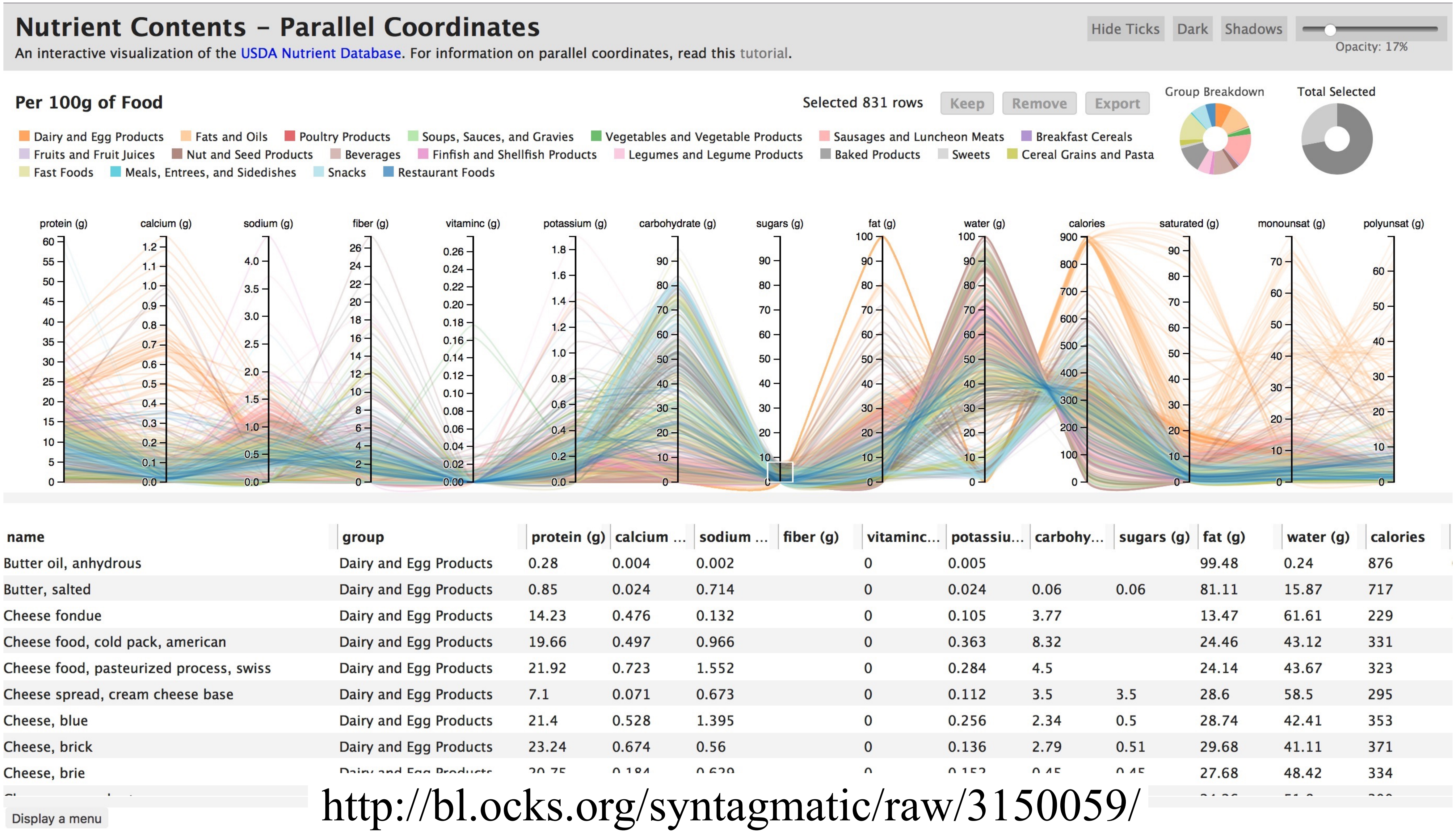
Stacked - Discrete

Stacked - Discrete - Ordered

- **When Order?**
  - **Not possible**
  - **Possible but ….**

Multivariate Data: Line-based Techniques

- **Parallel Coordinates**



Nutrient Contents – Parallel Coordinates

An interactive visualization of the USDA Nutrient Database. For information on parallel coordinates, read this tutorial.

name	group	protein (g)	calcium ...	sodium ...	fiber (g)	vitaminc...	potassiu...	carbohy...	sugars (g)	fat (g)	water (g)	calories
Butter oil, anhydrous	Dairy and Egg Products	0.28	0.004	0.002		0	0.005			99.48	0.24	876
Butter, salted	Dairy and Egg Products	0.85	0.024	0.714		0	0.024	0.06	0.06	81.11	15.87	717
Cheese fondue	Dairy and Egg Products	14.23	0.476	0.132		0	0.105	3.77		13.47	61.61	229
Cheese food, cold pack, american	Dairy and Egg Products	19.66	0.497	0.966		0	0.363	8.32		24.46	43.12	331
Cheese food, pasteurized process, swiss	Dairy and Egg Products	21.92	0.723	1.552		0	0.284	4.5		24.14	43.67	323
Cheese spread, cream cheese base	Dairy and Egg Products	7.1	0.071	0.673		0	0.112	3.5	3.5	28.6	58.5	295
Cheese, blue	Dairy and Egg Products	21.4	0.528	1.395		0	0.256	2.34	0.5	28.74	42.41	353
Cheese, brick	Dairy and Egg Products	23.24	0.674	0.56		0	0.136	2.79	0.51	29.68	41.11	371
Cheese, brie	Dairy and Egg Products	20.75	0.184	0.629		0	0.152	0.45	0.45	27.68	48.42	334

http://bl.ocks.org/syntagmatic/raw/3150059/

Parallel Coordinates (||-coords or PCP)

- **Inselberg in 1985**

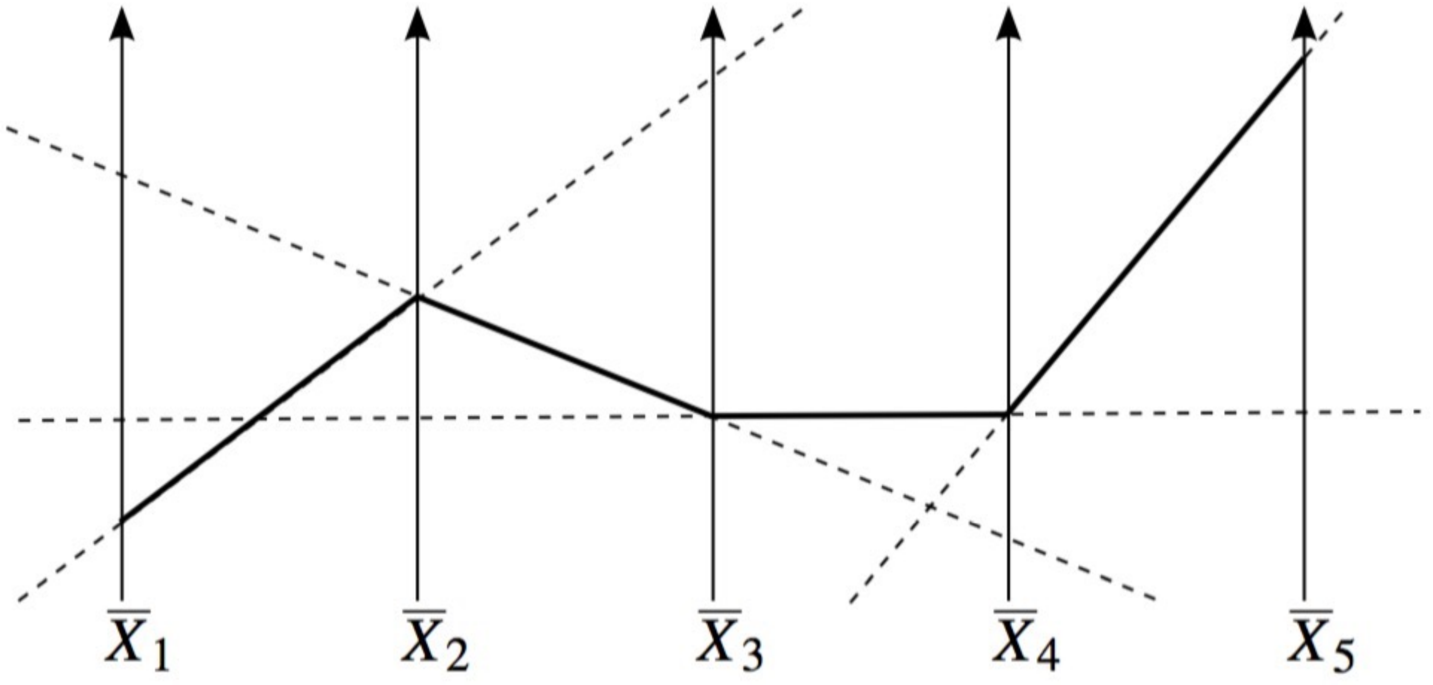


Figure 3: Constructing parallel coordinates with five dimensions represented by $N = 5$ vertical lines. Points in the plane are represented by lines joining the corresponding coordinates at the respective axes. Typically, only the line segments between the axes are drawn (represented by the bold polyline).

State of the Art of Parallel Coordinates
J. Heinrich and D. Weiskopf

Parallel Coordinates (||-coords or PCP)

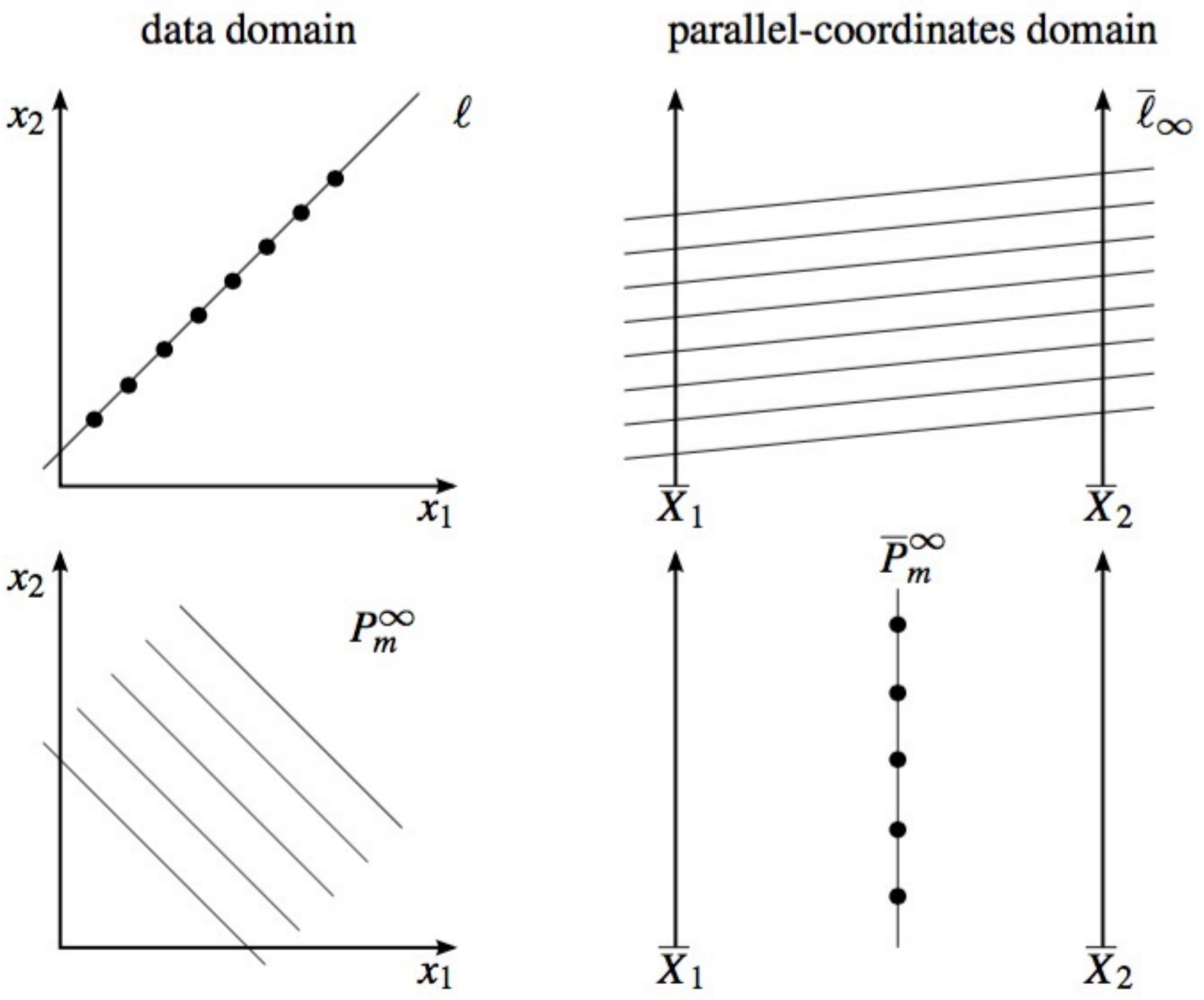


Figure 4: The line with slope $m = 1$ in the data domain is mapped to the ideal point $\bar{\ell}_\infty$ in parallel coordinates (top). The vertical line $\bar{P}_m^\infty : x = \frac{d}{1-m}$ in parallel coordinates is represented by the ideal point P_m^∞ with slope m in the data domain. Both domains are considered projective planes.

State of the Art of Parallel Coordinates
J. Heinrich and D. Weiskopf

Parallel Coordinates (||-coords or PCP)

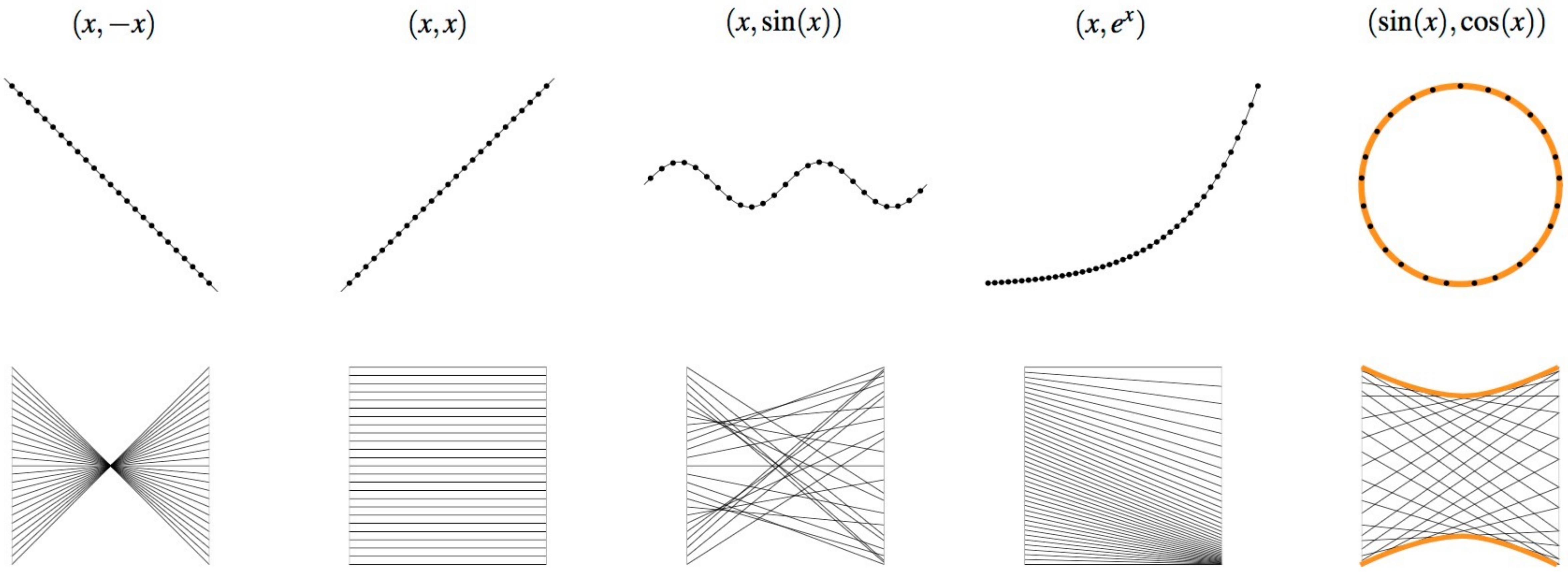
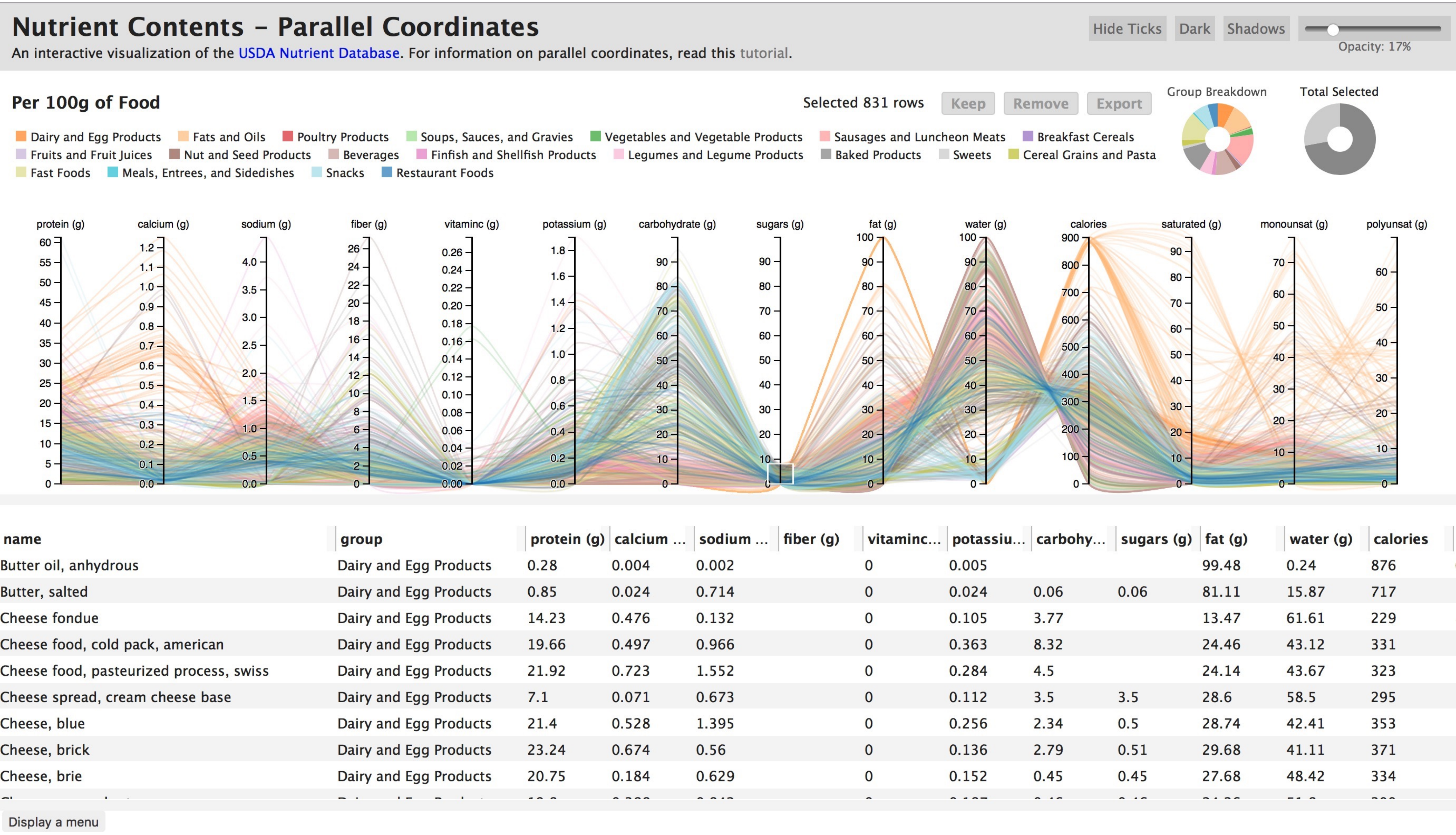


Figure 5: Common patterns in Cartesian coordinates (top) and their dual representation in parallel coordinates (bottom). The envelope of lines is highlighted for the ellipse–hyperbola duality.

State of the Art of Parallel Coordinates
J. Heinrich and D. Weiskopf

http://bl.ocks.org/syntagmatic/raw/3150059/



Nutrient Contents – Parallel Coordinates

An interactive visualization of the USDA Nutrient Database. For information on parallel coordinates, read this tutorial.

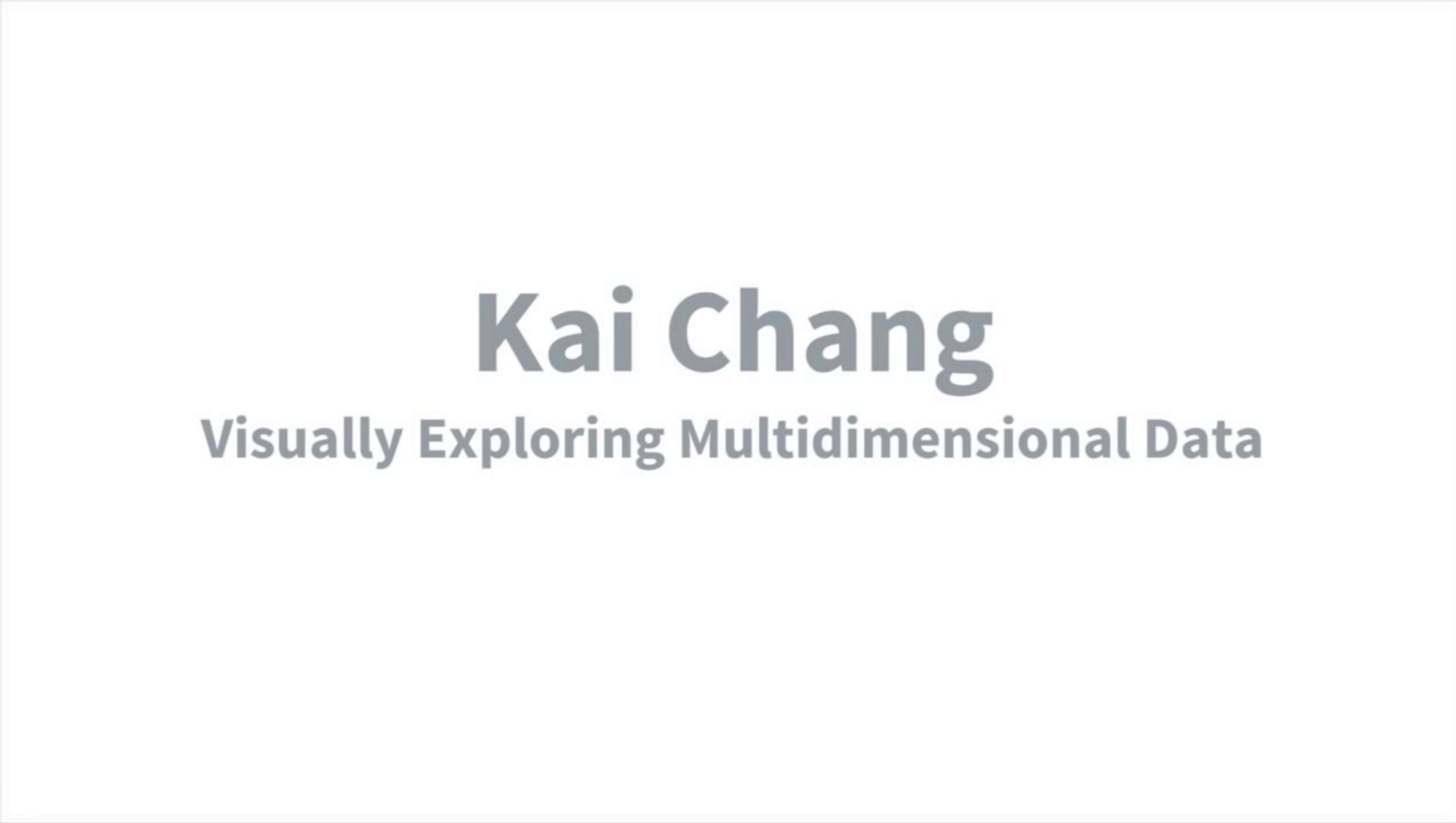
Per 100g of Food

Selected 831 rows

■ Dairy and Egg Products ■ Fats and Oils ■ Poultry Products ■ Soups, Sauces, and Gravies ■ Vegetables and Vegetable Products ■ Sausages and Luncheon Meats ■ Breakfast Cereals
■ Fruits and Fruit Juices ■ Nut and Seed Products ■ Beverages ■ Finfish and Shellfish Products ■ Legumes and Legume Products ■ Baked Products ■ Sweets ■ Cereal Grains and Pasta
■ Fast Foods ■ Meals, Entrees, and Sidedishes ■ Snacks ■ Restaurant Foods

name	group	protein (g)	calcium ...	sodium ...	fiber (g)	vitaminc...	potassiu...	carbohy...	sugars (g)	fat (g)	water (g)	calories
Butter oil, anhydrous	Dairy and Egg Products	0.28	0.004	0.002		0	0.005			99.48	0.24	876
Butter, salted	Dairy and Egg Products	0.85	0.024	0.714		0	0.024	0.06	0.06	81.11	15.87	717
Cheese fondue	Dairy and Egg Products	14.23	0.476	0.132		0	0.105	3.77		13.47	61.61	229
Cheese food, cold pack, american	Dairy and Egg Products	19.66	0.497	0.966		0	0.363	8.32		24.46	43.12	331
Cheese food, pasteurized process, swiss	Dairy and Egg Products	21.92	0.723	1.552		0	0.284	4.5		24.14	43.67	323
Cheese spread, cream cheese base	Dairy and Egg Products	7.1	0.071	0.673		0	0.112	3.5	3.5	28.6	58.5	295
Cheese, blue	Dairy and Egg Products	21.4	0.528	1.395		0	0.256	2.34	0.5	28.74	42.41	353
Cheese, brick	Dairy and Egg Products	23.24	0.674	0.56		0	0.136	2.79	0.51	29.68	41.11	371
Cheese, brie	Dairy and Egg Products	20.75	0.184	0.629		0	0.152	0.45	0.45	27.68	48.42	334

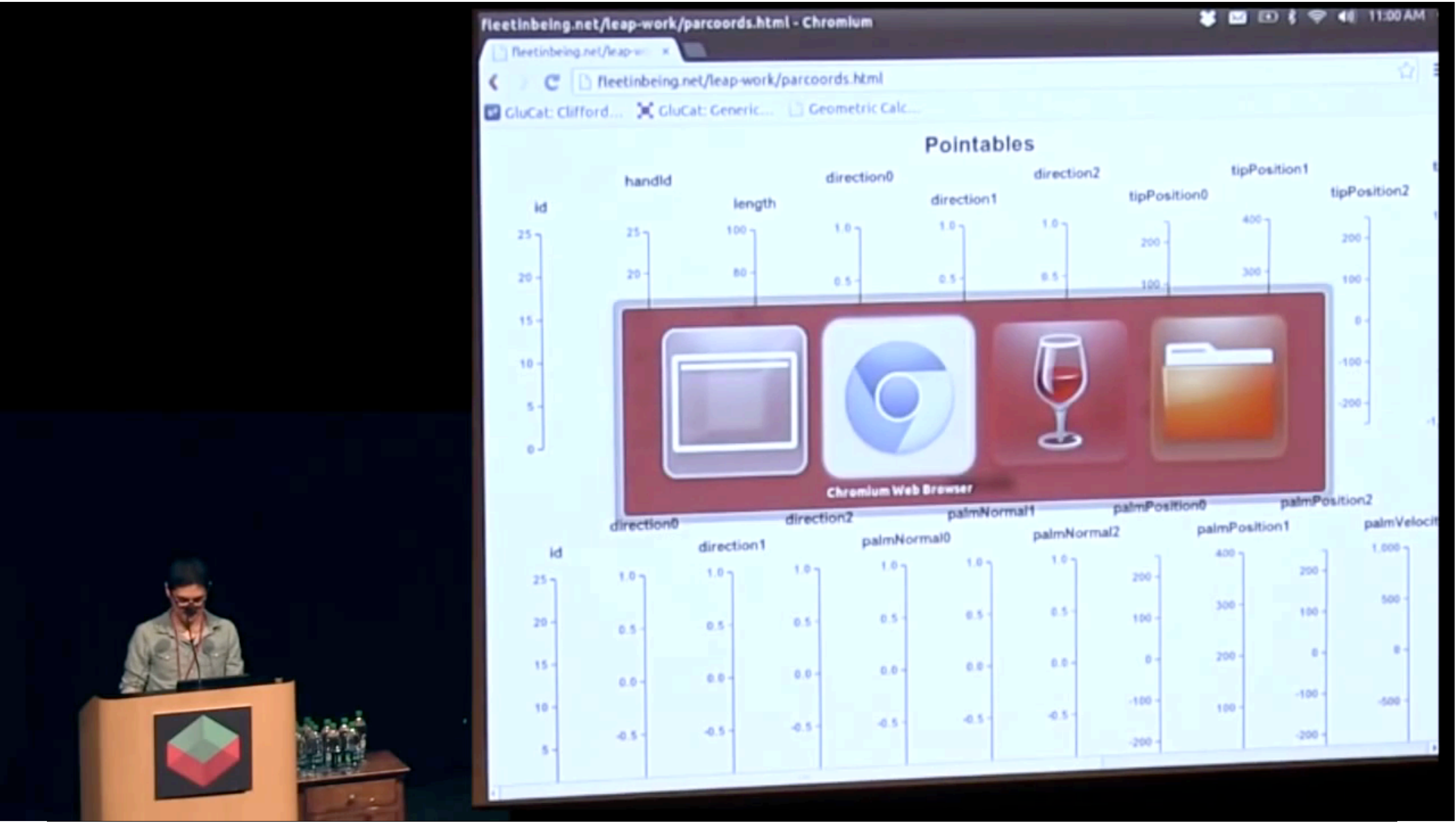
Parallel Coordinates (||-coords or PCP)



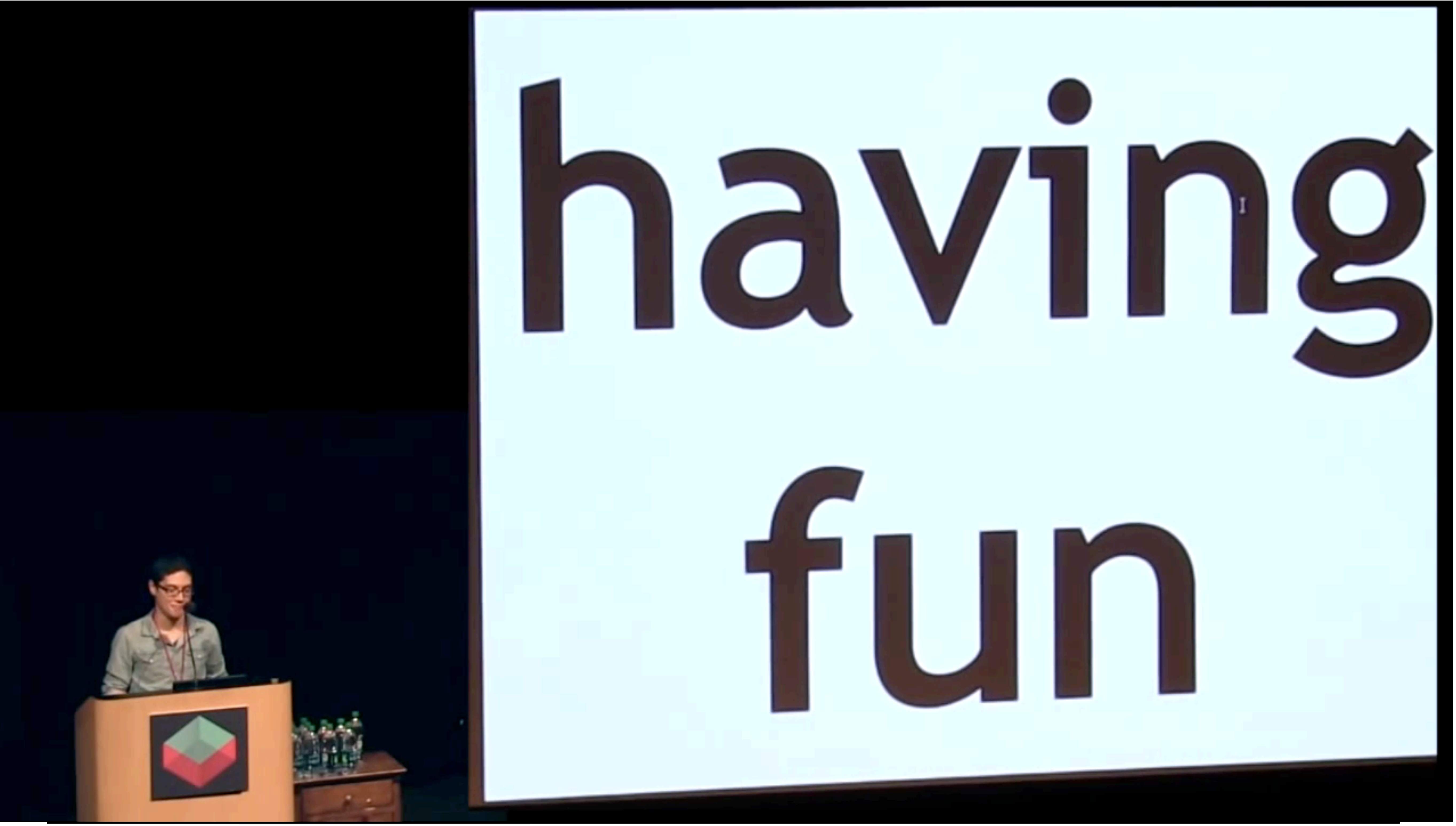
Kai Chang

Visually Exploring Multidimensional Data

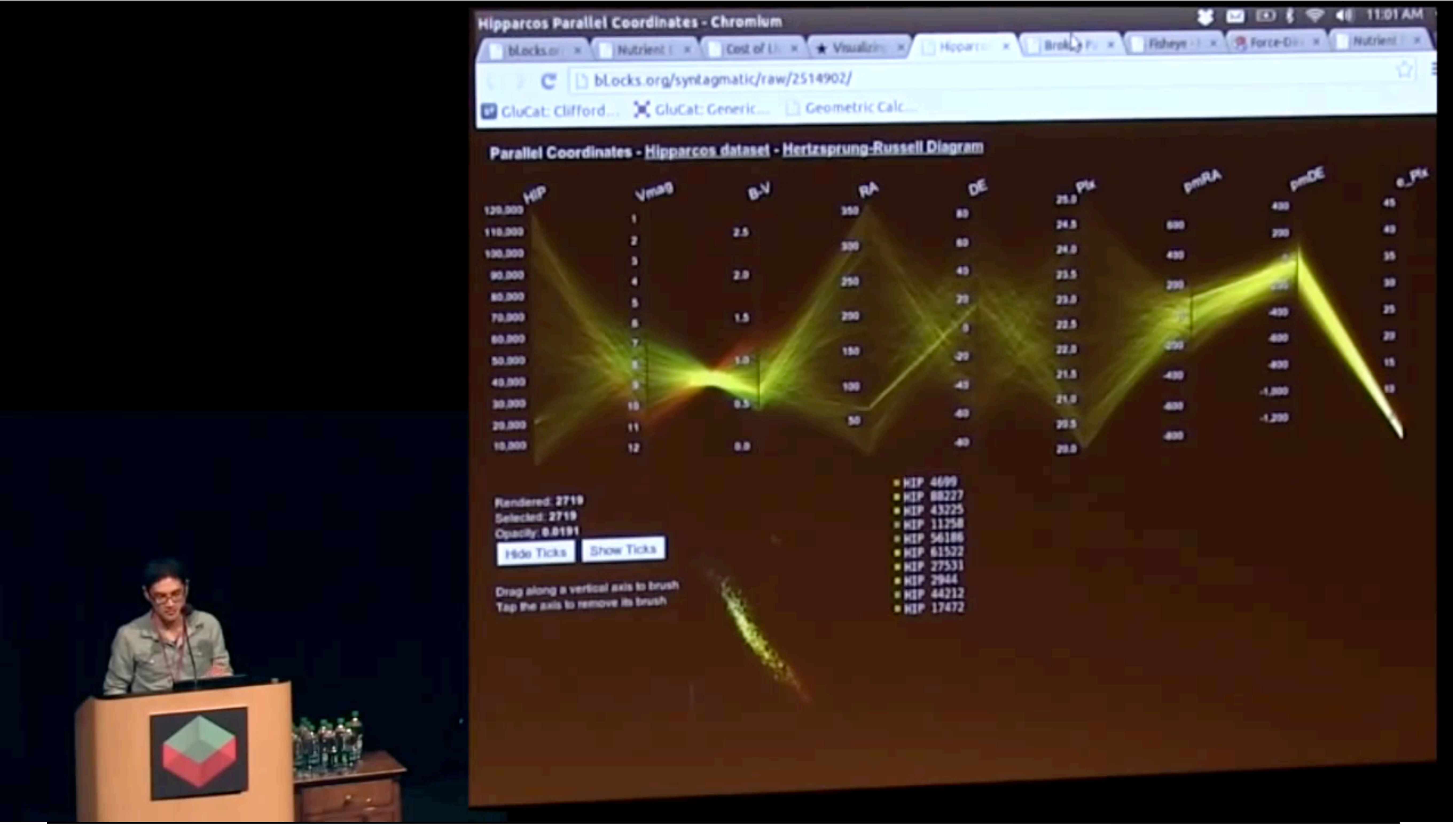
Parallel Coordinates (||-coords or PCP)



Parallel Coordinates (||-coords or PCP)



Parallel Coordinates (||-coords or PCP)



Hipparcos Parallel Coordinates - Chromium

bl.ocks.org/syntagmatic/raw/2514902/

Parallel Coordinates - Hipparcos dataset - Hertzsprung-Russell Diagram

Rendered: 2719
Selected: 2719
Opacity: 0.0191

Hide Ticks | Show Ticks

Drag along a vertical axis to brush
Tap the axis to remove its brush

- HIP 4699
- HIP 88227
- HIP 43225
- HIP 11258
- HIP 56186
- HIP 61522
- HIP 27531
- HIP 2944
- HIP 44212
- HIP 17472

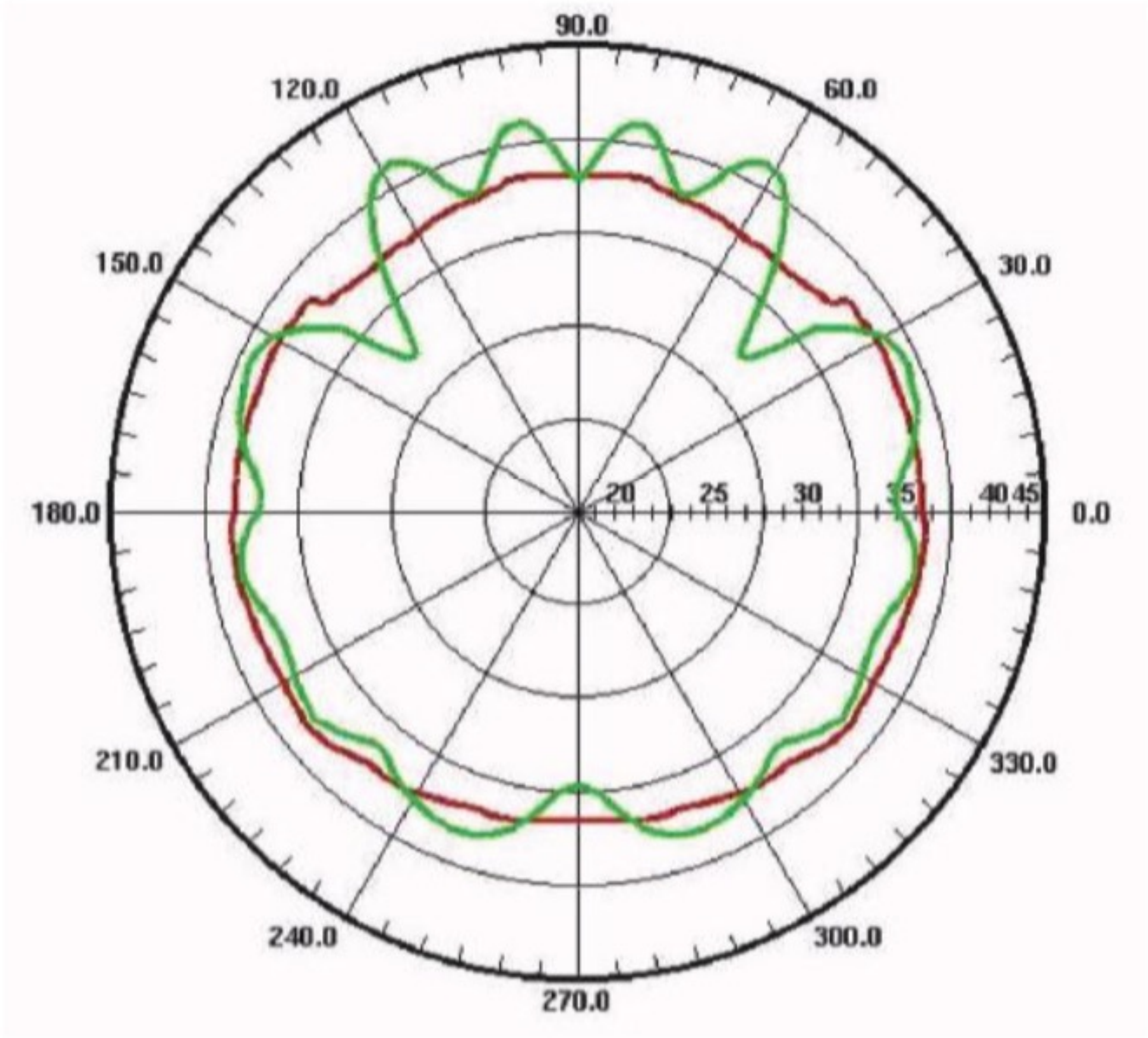
Parallel Coordinates (||-coords or PCP)

- **A brief tutorial on ||-coords (https://eagereyes.org/techniques/parallel-coordinates)**
- **A D3 library (https://syntagmatic.github.io/parallel-coordinates/)**
- **Some very special videos, from Alfred Inselberg's tutorial at iV 2016, at Lisbon (FB and Twitter):**
  - **Part1**
  - **Part 2**
  - **Part 3**

Multivariate Data: Line-Based Techniques

- **Radial Axis Techniques**
  - **circular line graph;**
  - **polar graphs: point plots using polar coordinates;**
  - **circular bar charts: like circular line graphs, but plotting bars on the base line;**
  - **circular area graphs: like a line graph, but with the area under line filled in with a color or texture;**
  - **circular bar graphs: with bars that are circular arcs with a common center point and base line.**

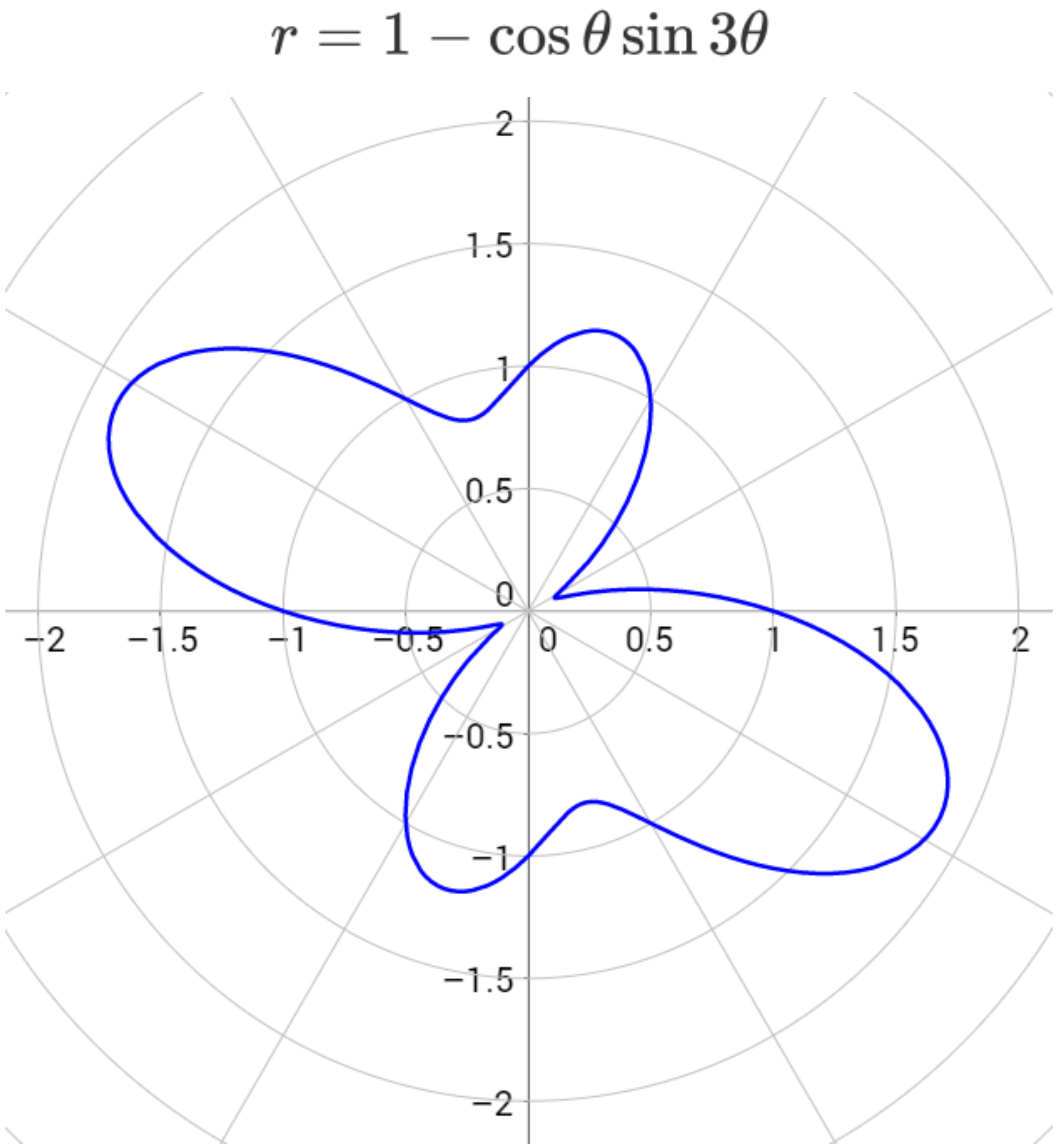
Multivariate Data: Line-Based Techniques



An example of a circular line graph. (Image courtesy http://www.cemframework .com/img/PolarPlot1.png.)

Multivariate Data: Line-Based Techniques

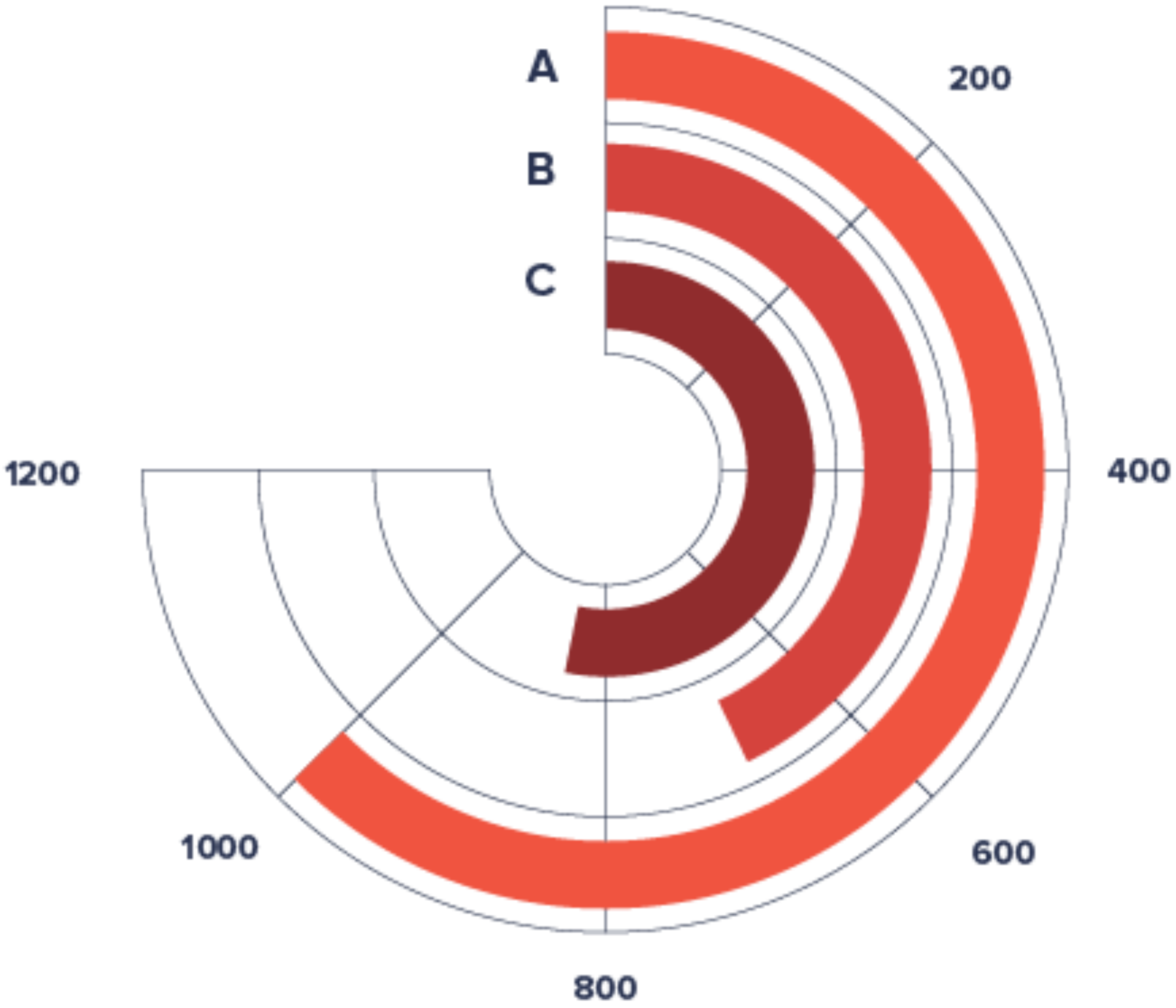
polar graphs - point plots using **polar coordinates**



https://brilliant.org/wiki/polar-curves/

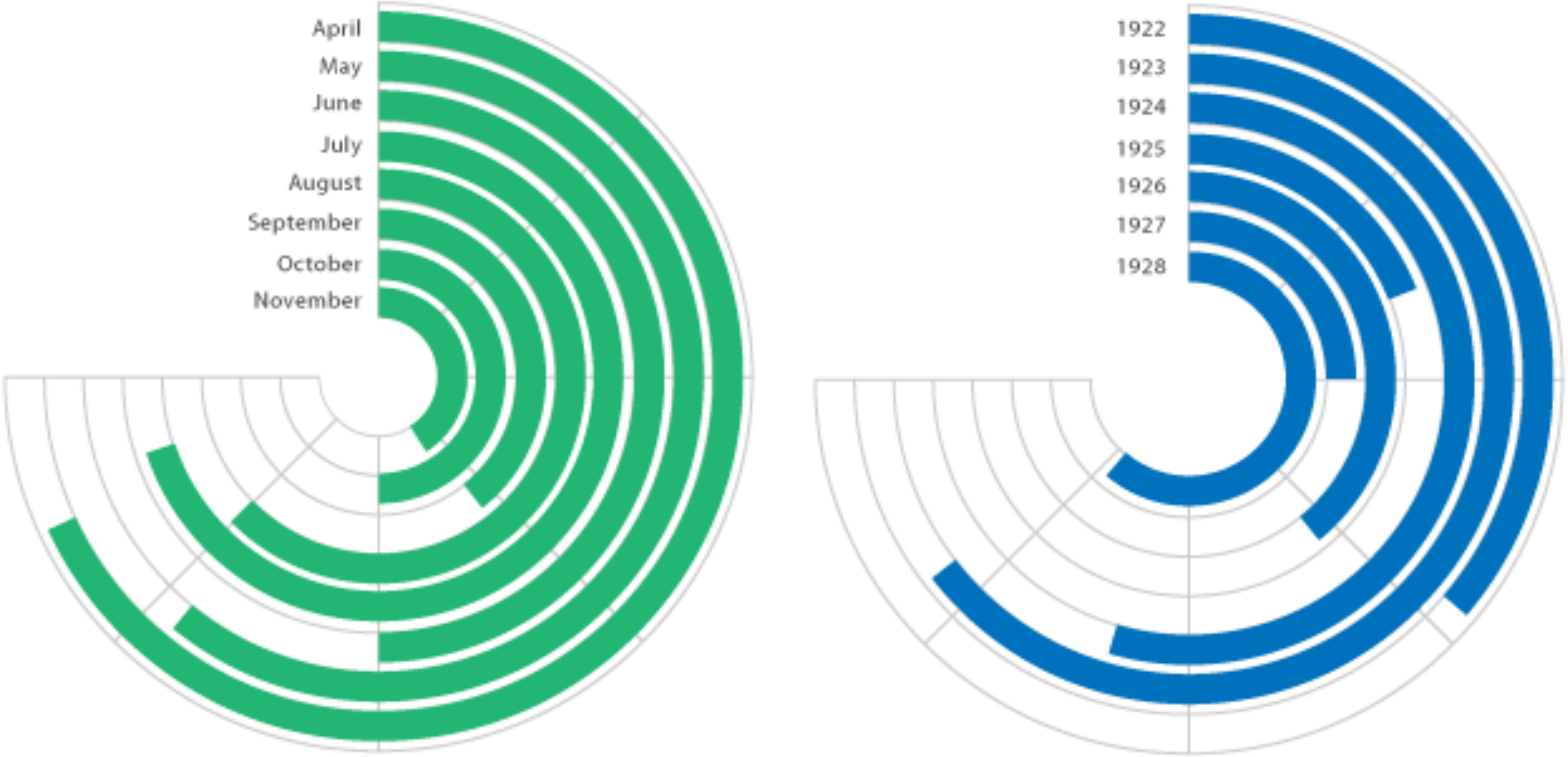
Multivariate Data: Line-Based Techniques

circular bar charts: like circular line graphs, but plotting bars on the base line



Multivariate Data: Line-Based Techniques

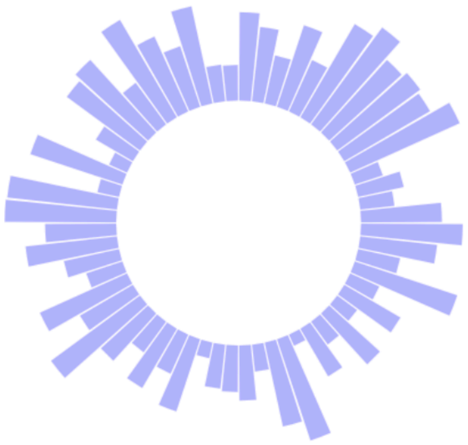
circular bar charts: like circular line graphs, but plotting bars on the base line



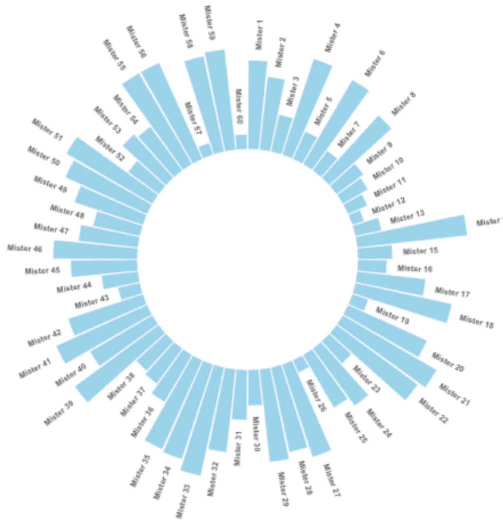
https://datavizcatalogue.com/methods/radial_bar_chart.html

Multivariate Data: Line-Based Techniques

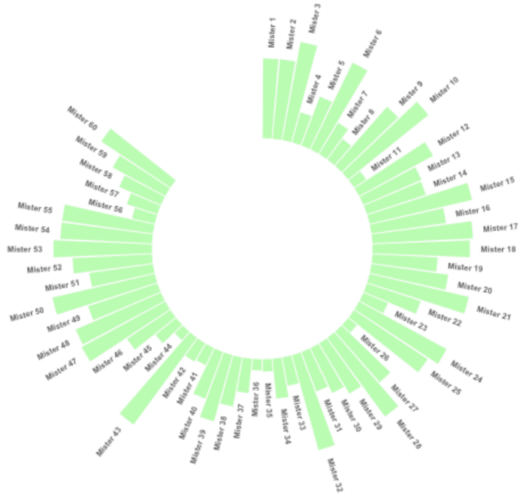
circular bar graphs: with bars that are circular arcs with a common center point and base line.



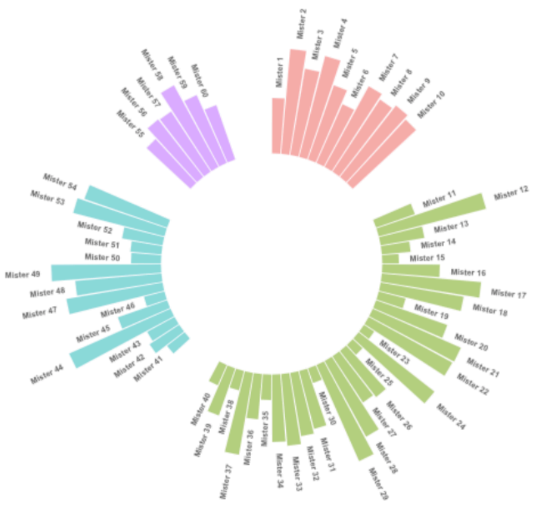
#295 Basic Circular Barplot



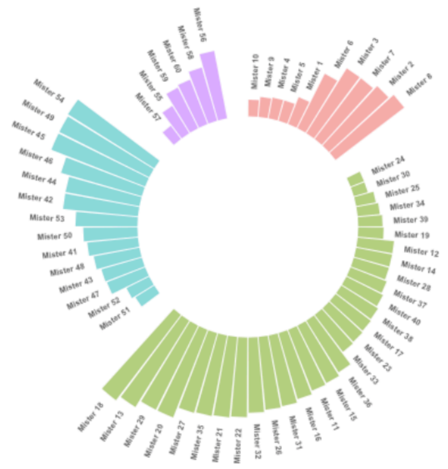
#296 Add labels on circular barplot



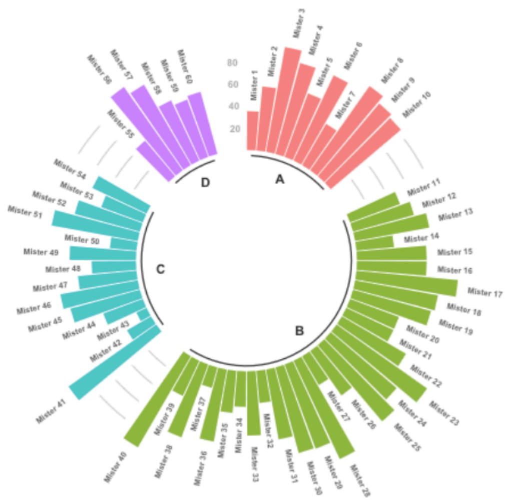
#297 Circular marplot with break



#297 Add breaks between groups

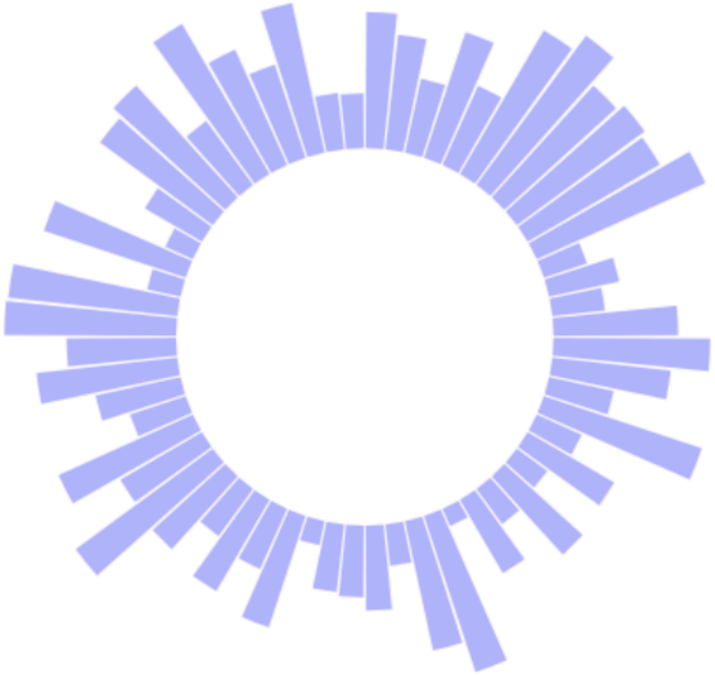


#297 Order each group

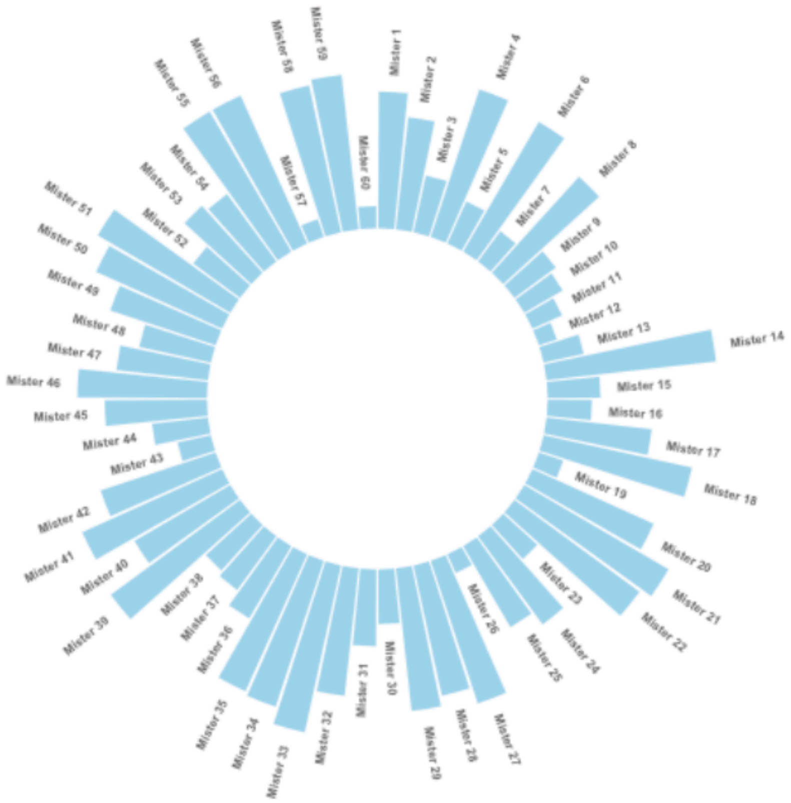


#297 Grouped circular barplot

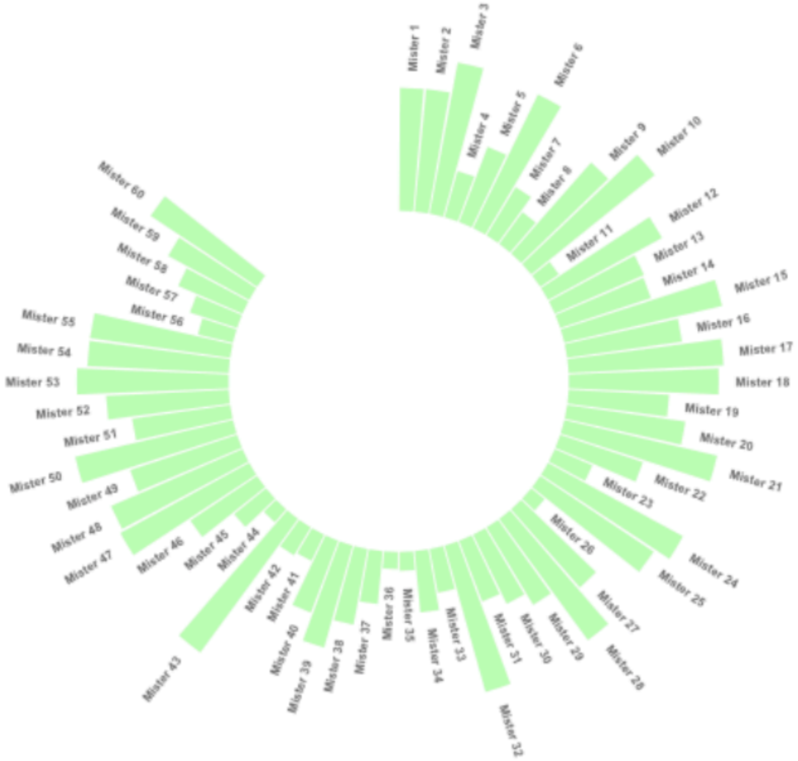
https://www.r-graph-gallery.com/circular-barplot/



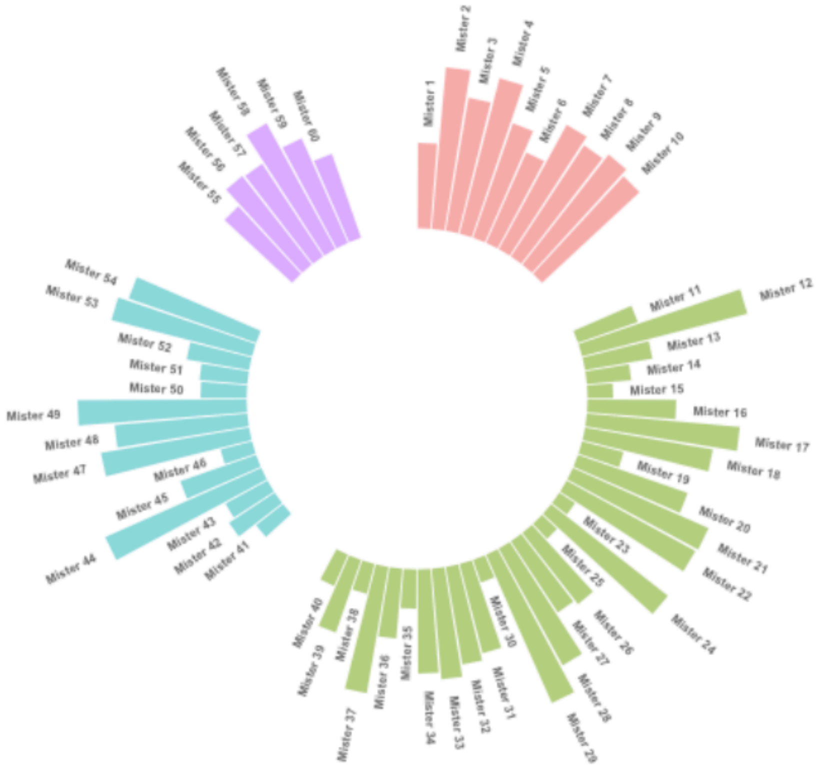
#295 Basic Circular Barplot



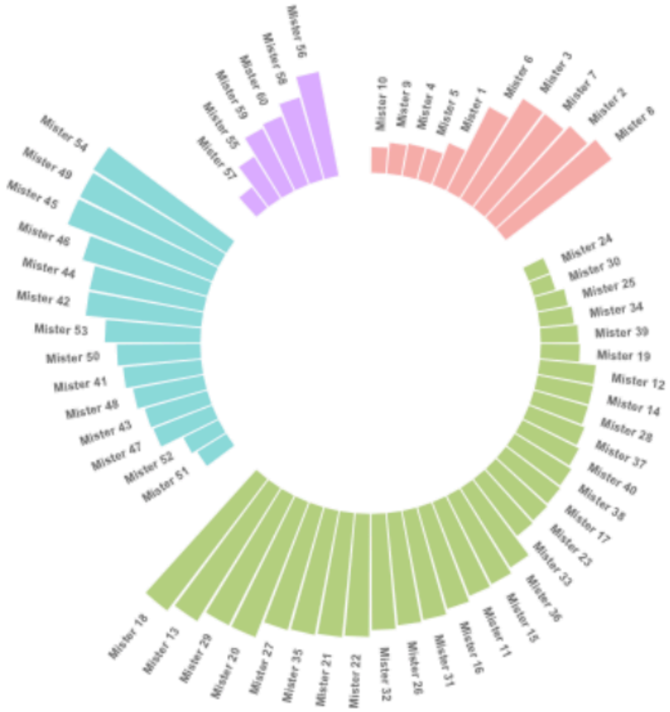
#296 Add labels on circular barplot



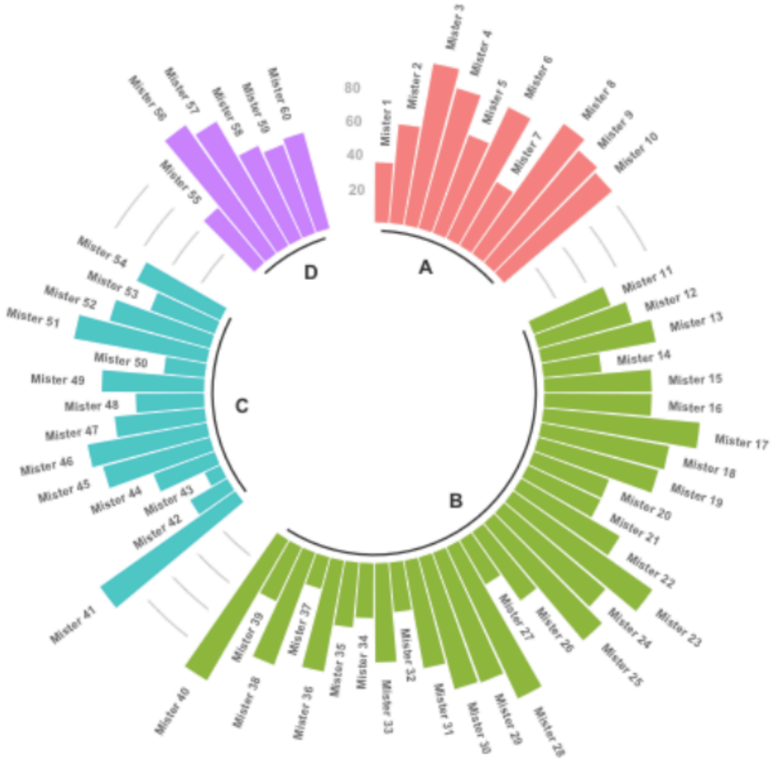
#297 Circular marplot with break



#297 Add breaks between groups



#297 Order each group

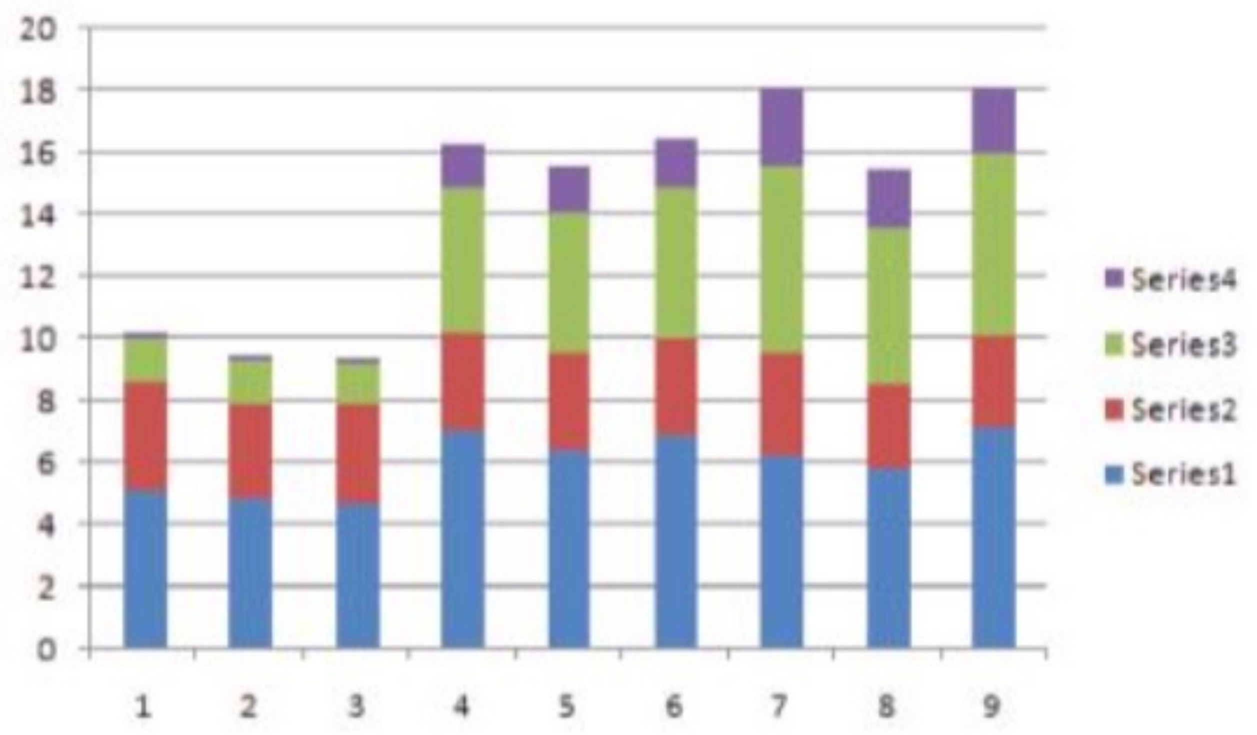


#297 Grouped circular barplot

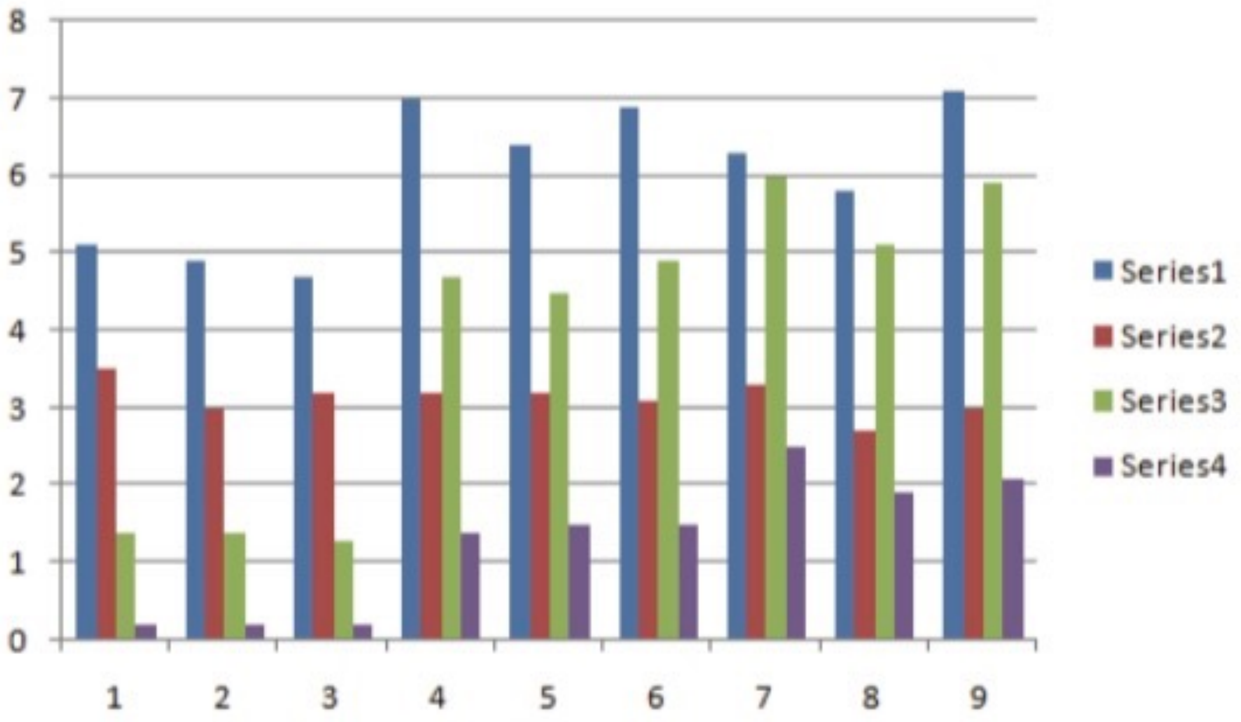
Region-Based Techniques

Multivariate Data: Region-Based Techniques

- **Bar Charts and Area Charts**

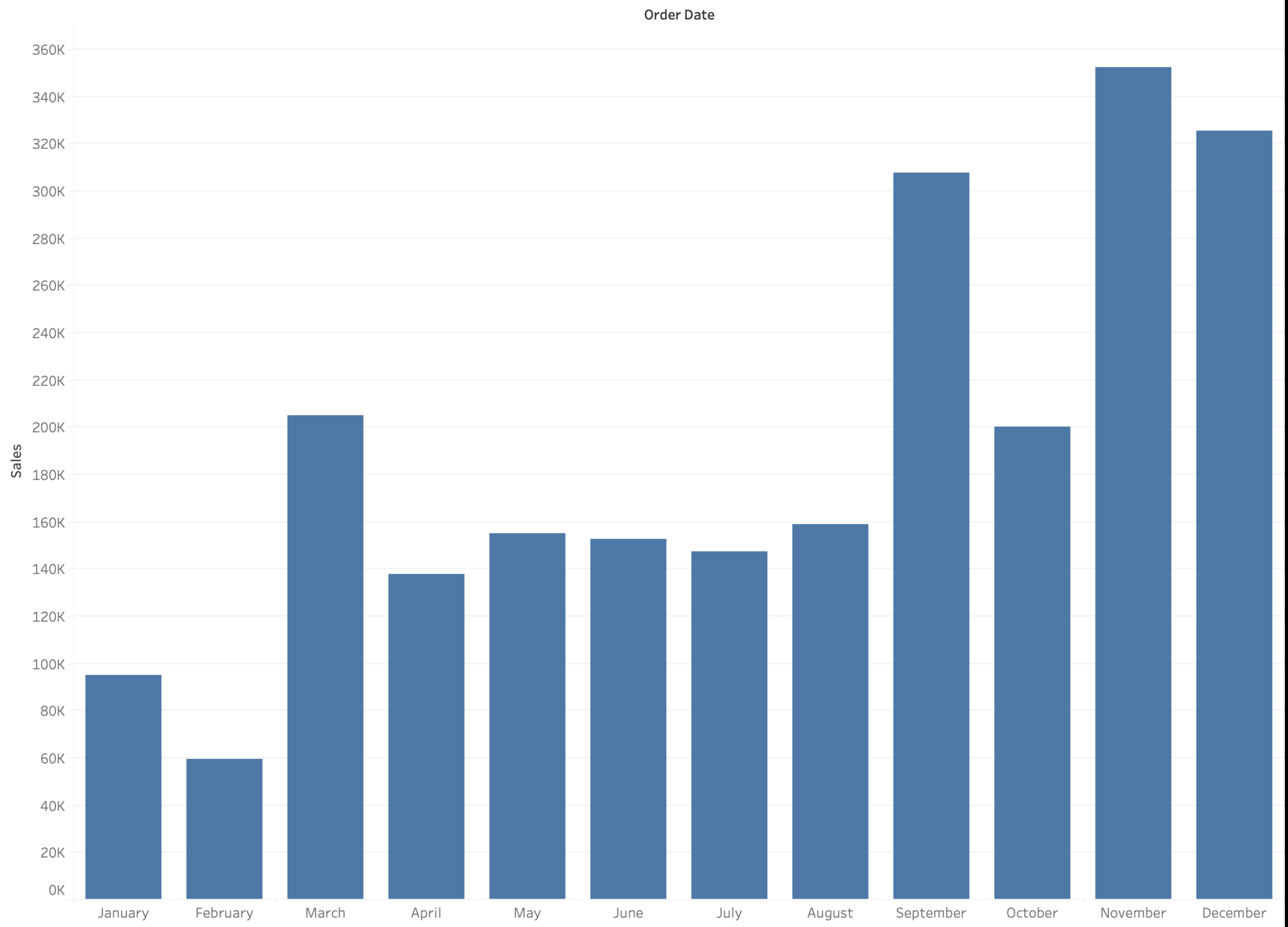


(a) Stacked bar chart.



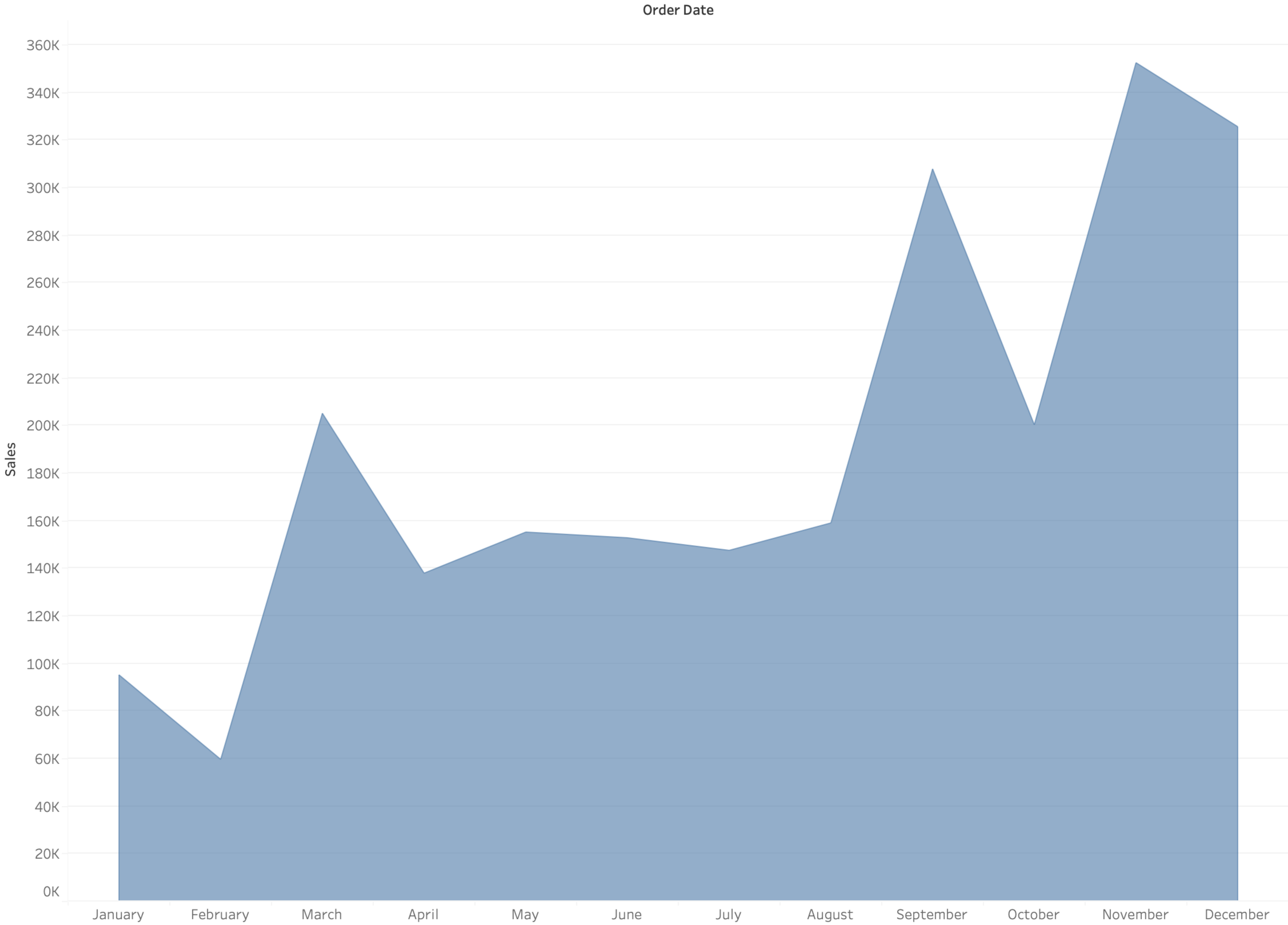
(b) Clustered bar chart.

Simple Bar Chart



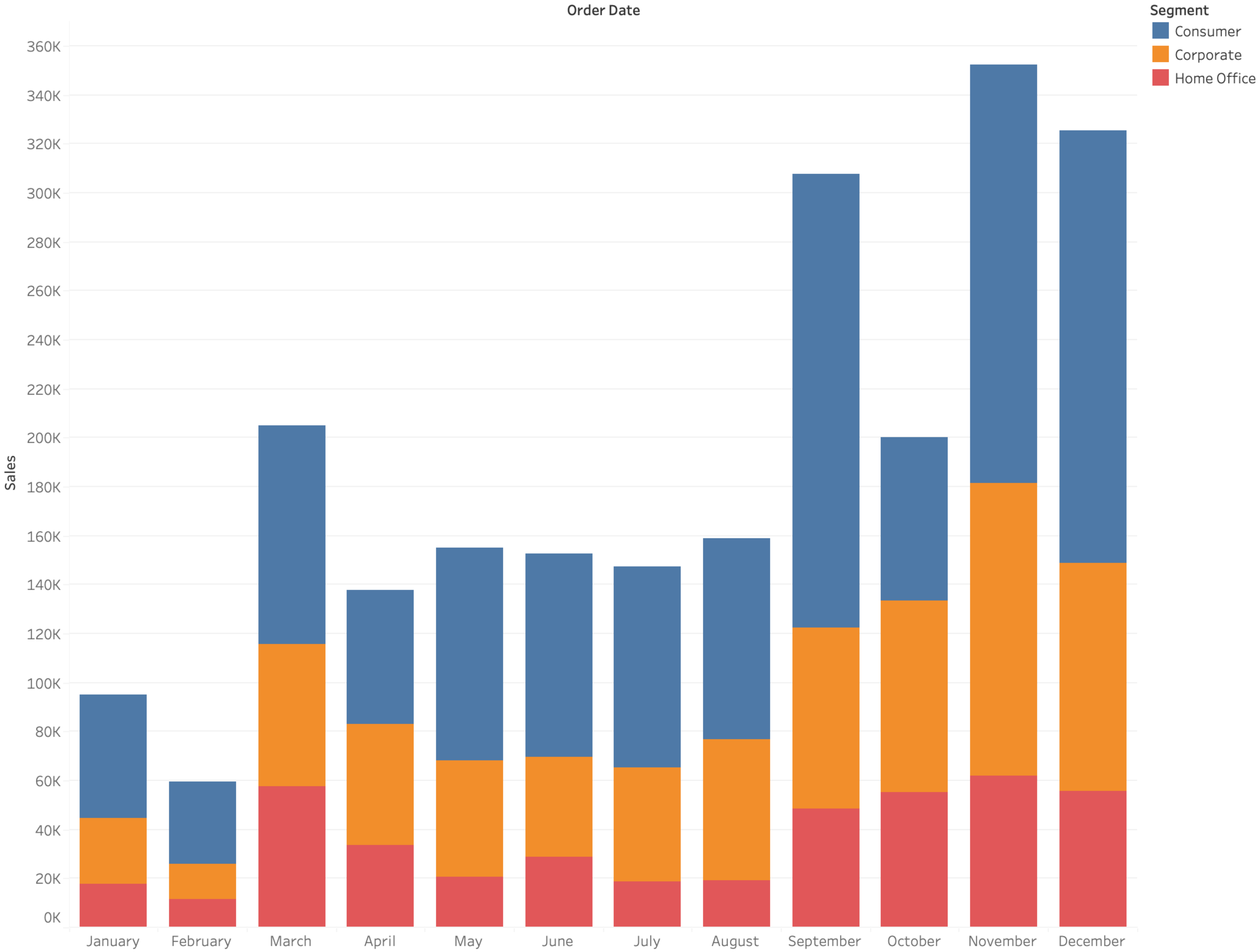
Sum of Sales for each Order Date Month.

Simple Area Chart



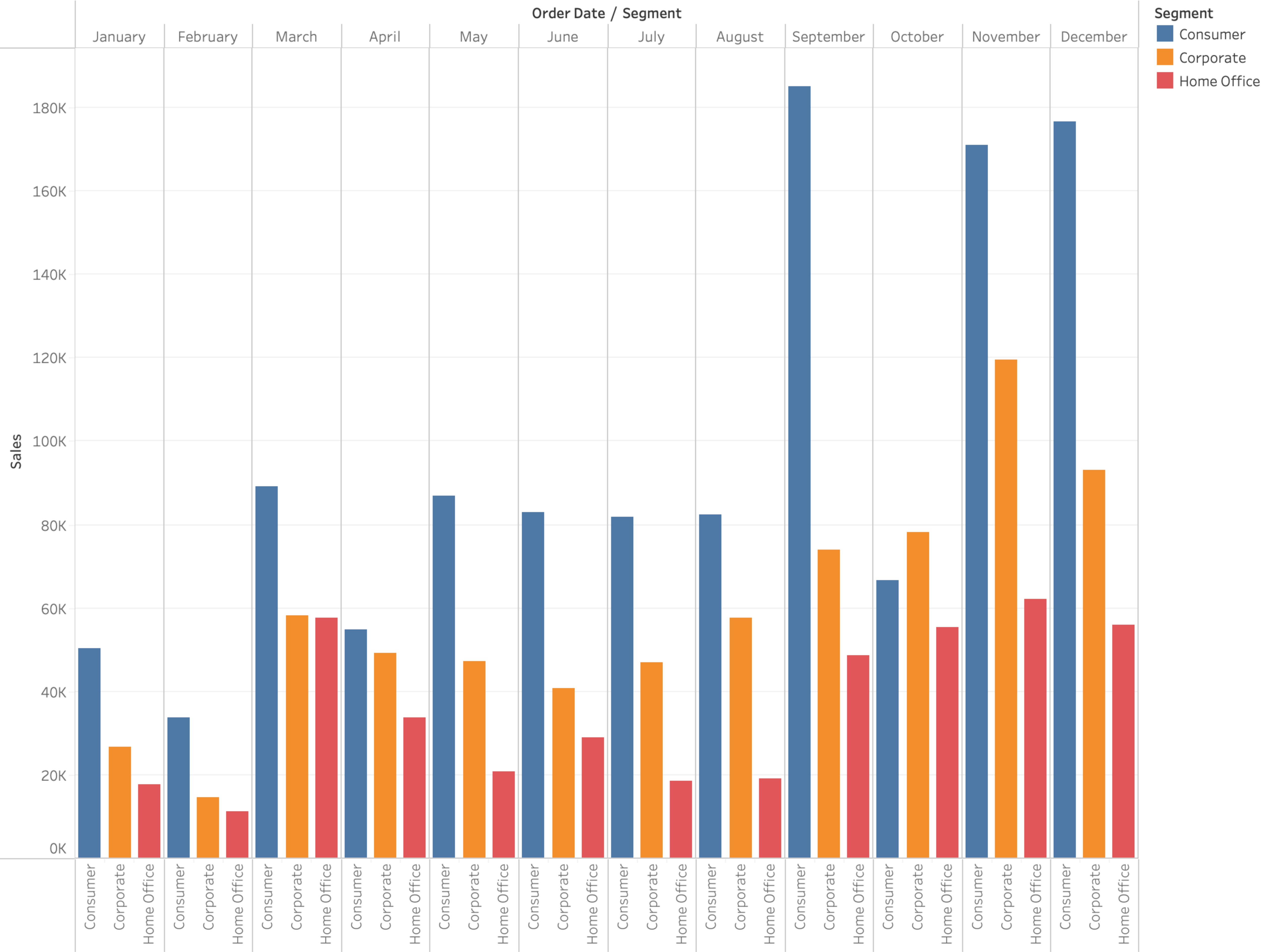
Sum of Sales for each Order Date Month.

Bar Chart - Stacked



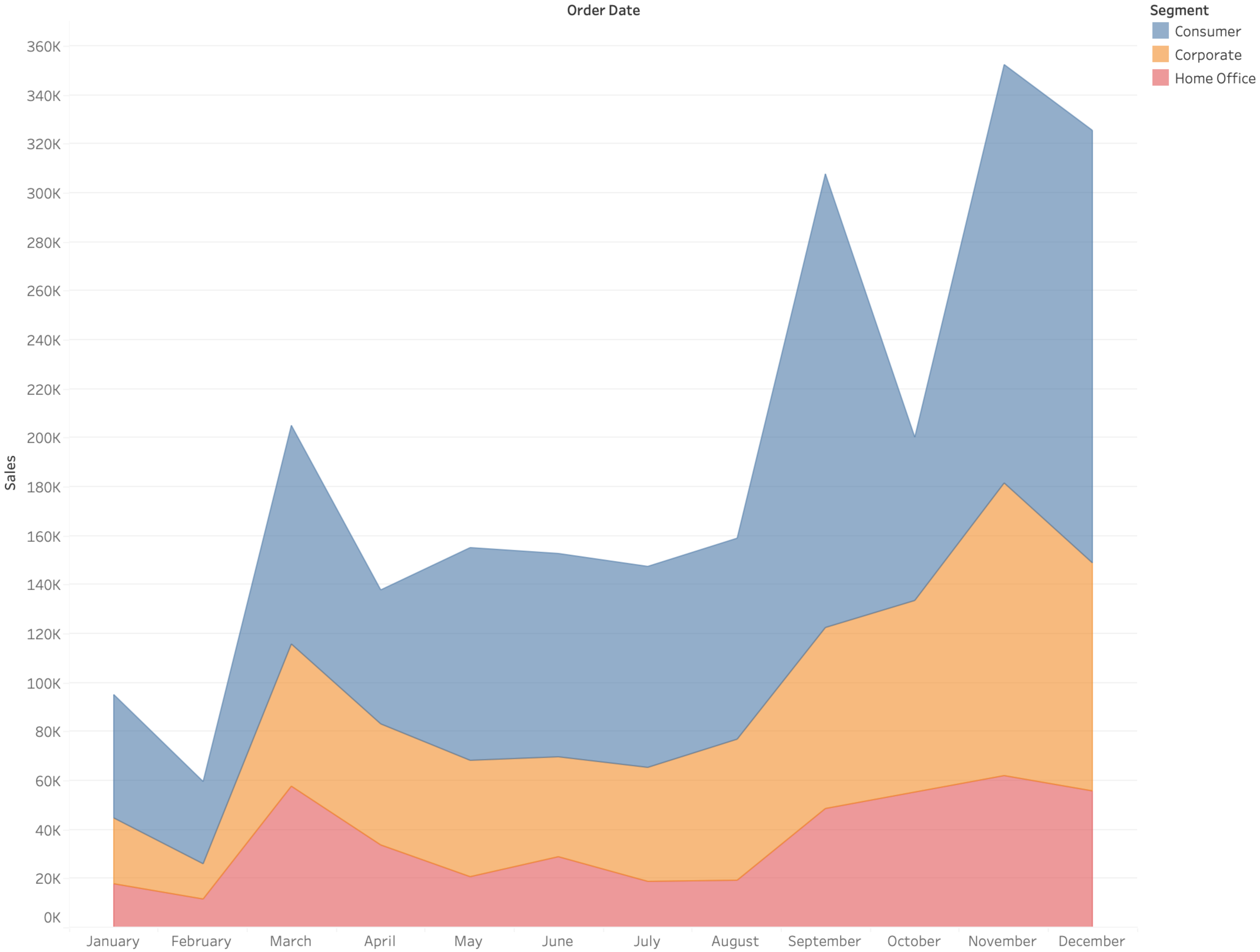
Sum of Sales for each Order Date Month. Color shows details about Segment.

Bar Chart - Clustered



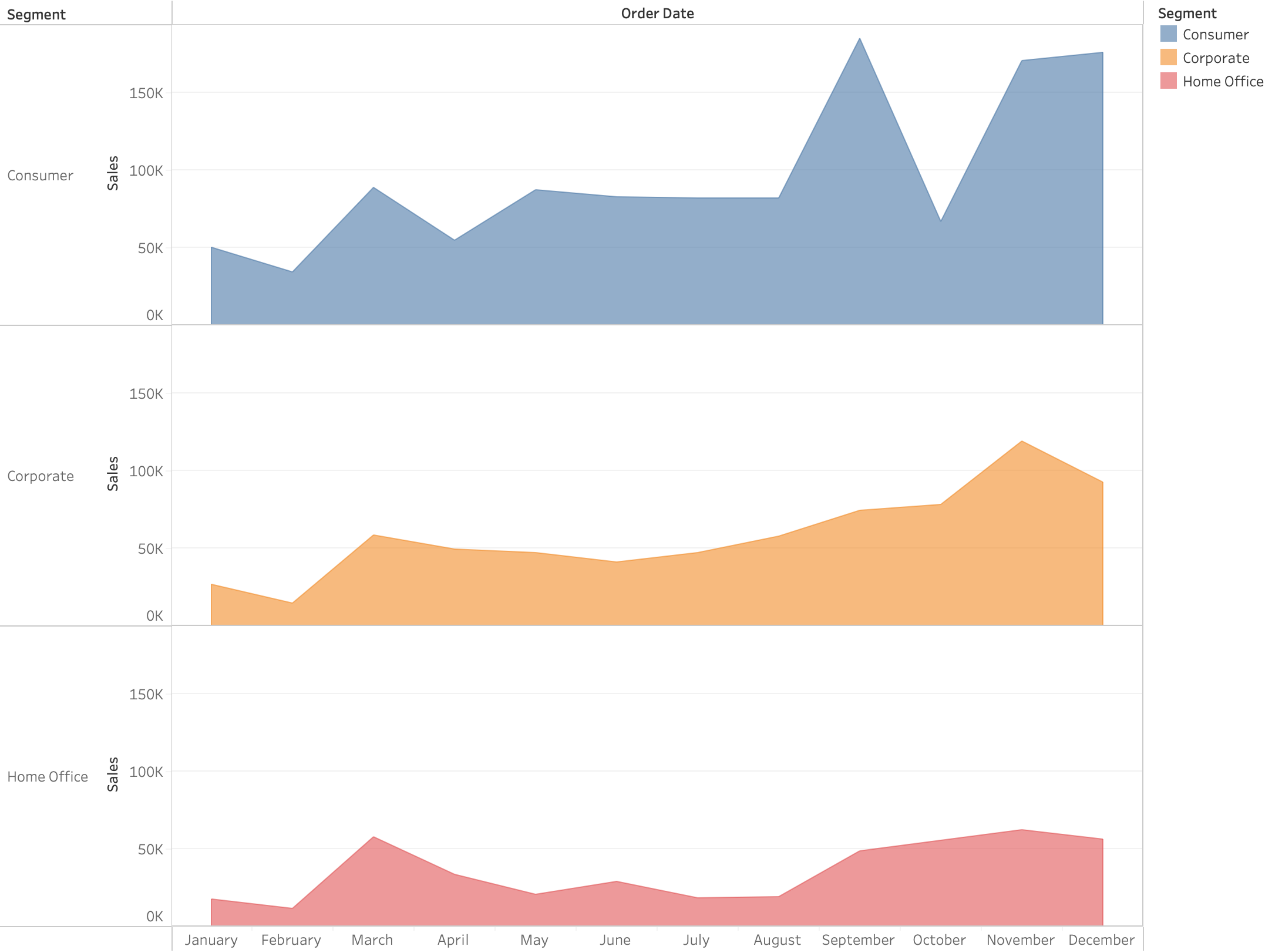
Sum of Sales for each Segment broken down by Order Date Month. Color shows details about Segment.

Area Chart - Stacked



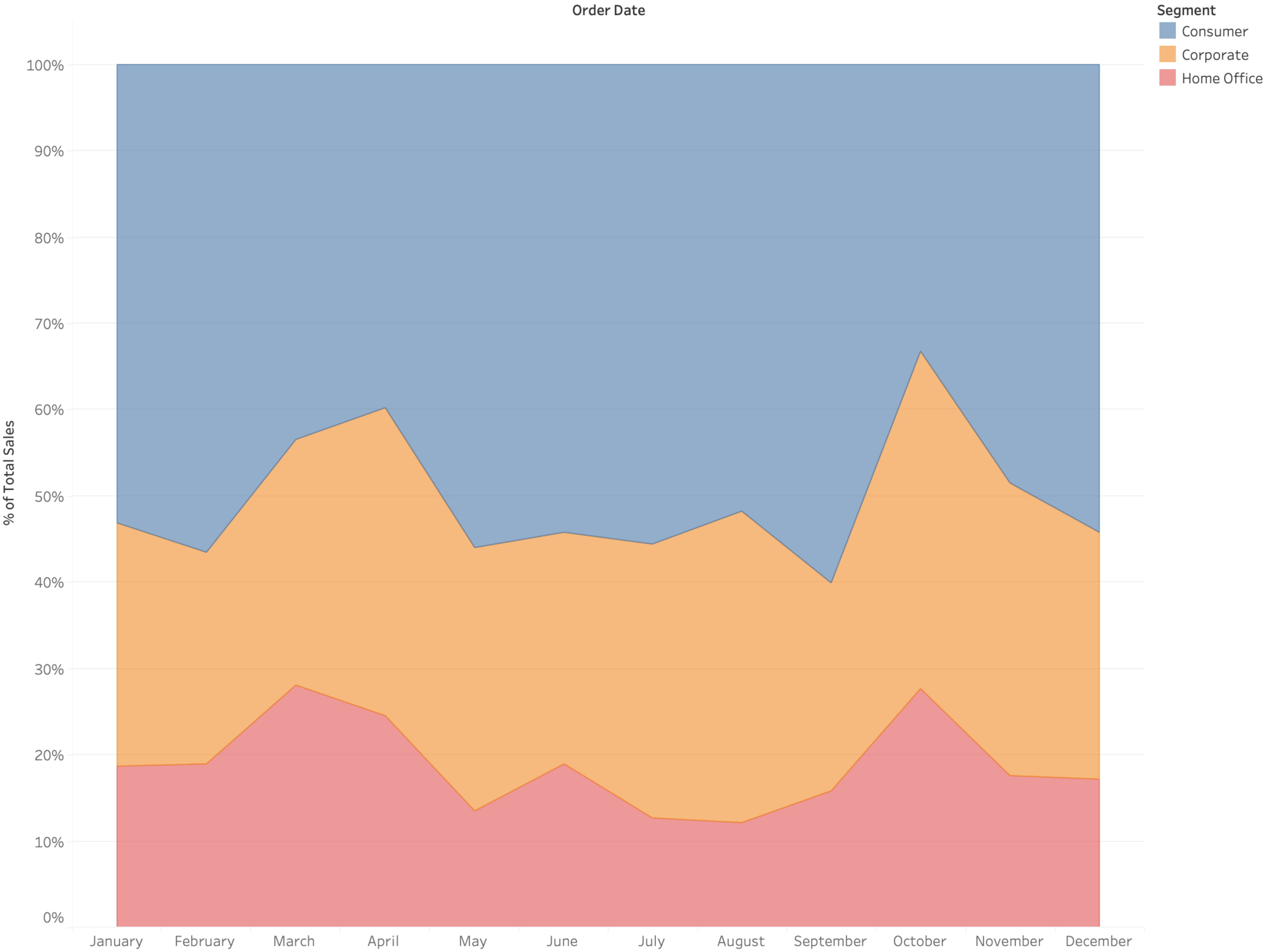
Sum of Sales for each Order Date Month. Color shows details about Segment.

Area Chart - Stacked (2)



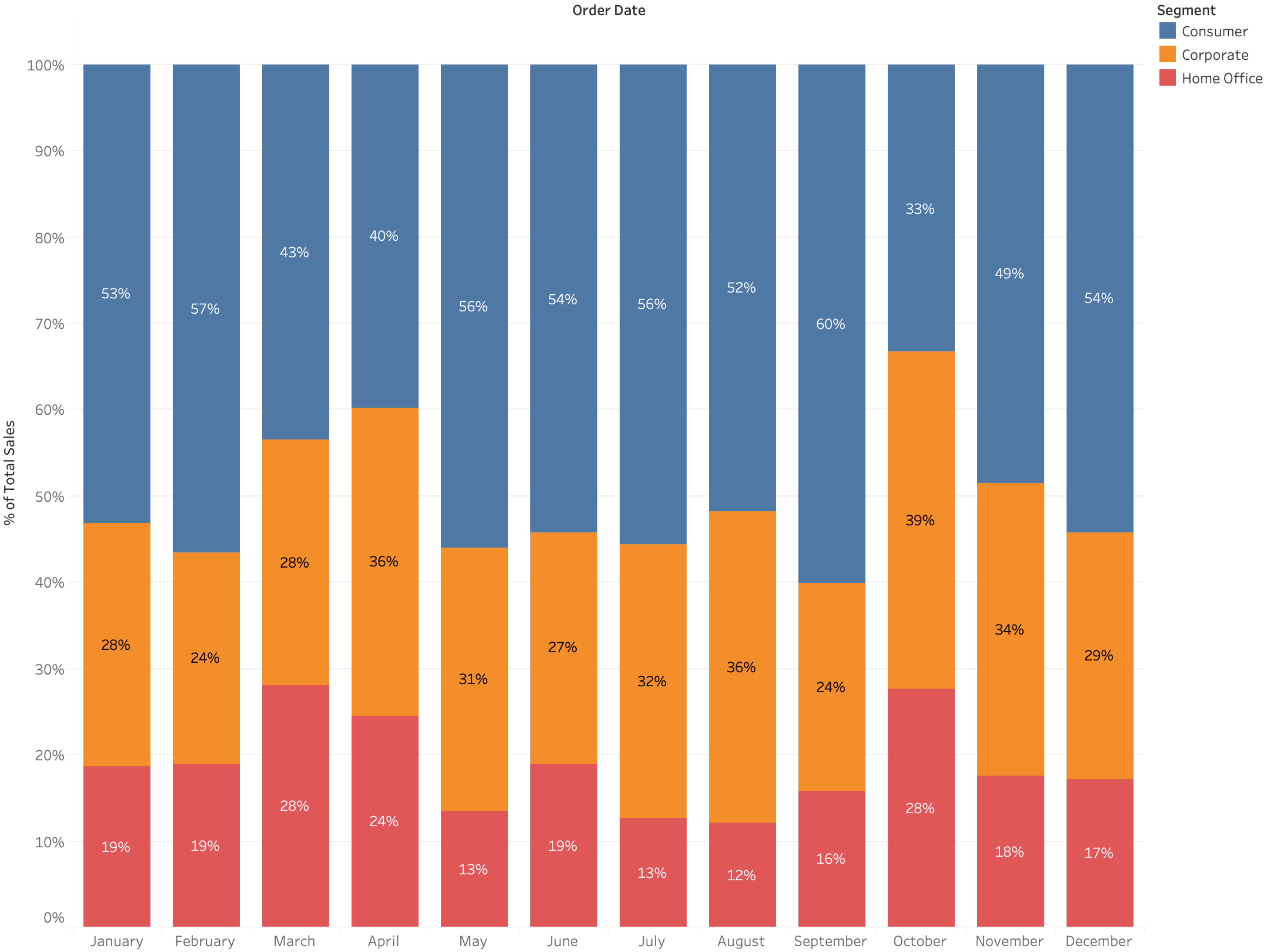
Sum of Sales for each Order Date Month broken down by Segment. Color shows details about Segment.

Area Chart - Stacked - 100%



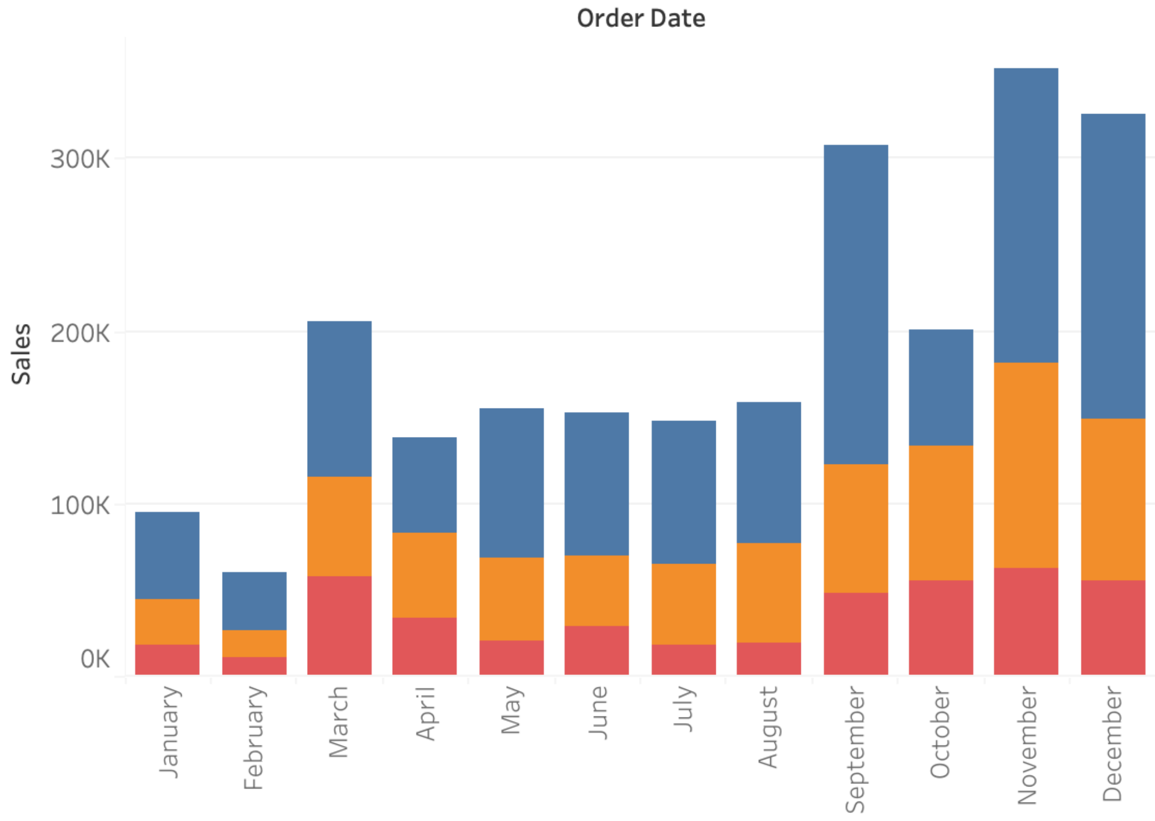
% of Total Sales for each Order Date Month. Color shows details about Segment.

Bar Chart - Stacked - 100%

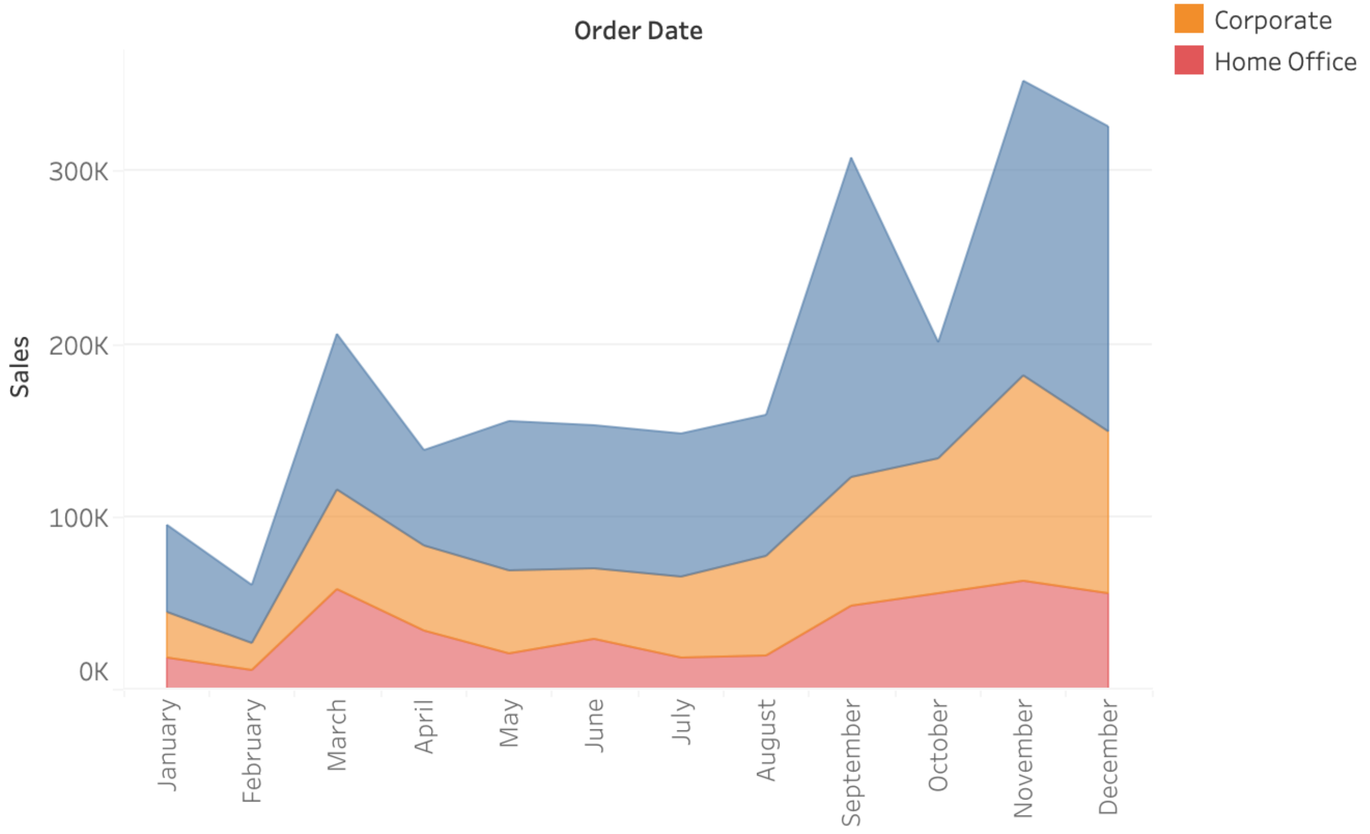


% of Total Sales for each Order Date Month. Color shows details about Segment.

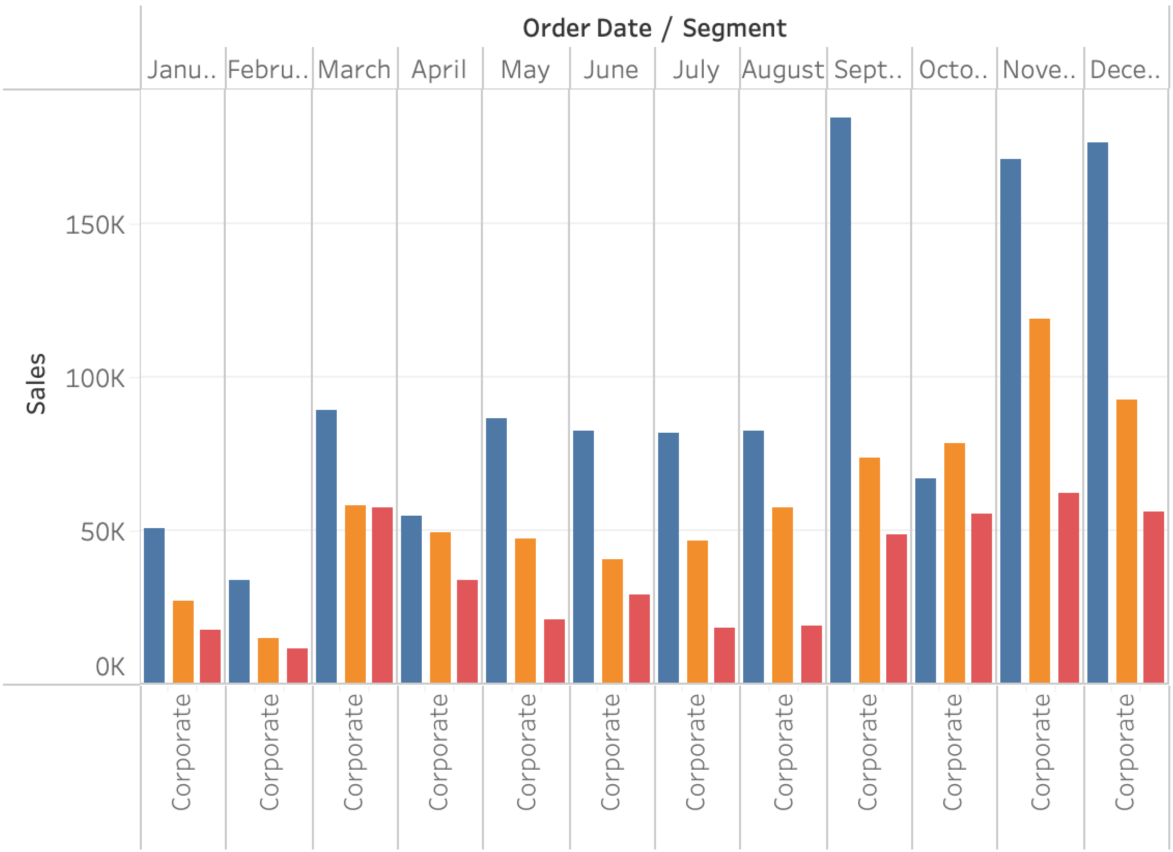
Bar Chart - Stacked



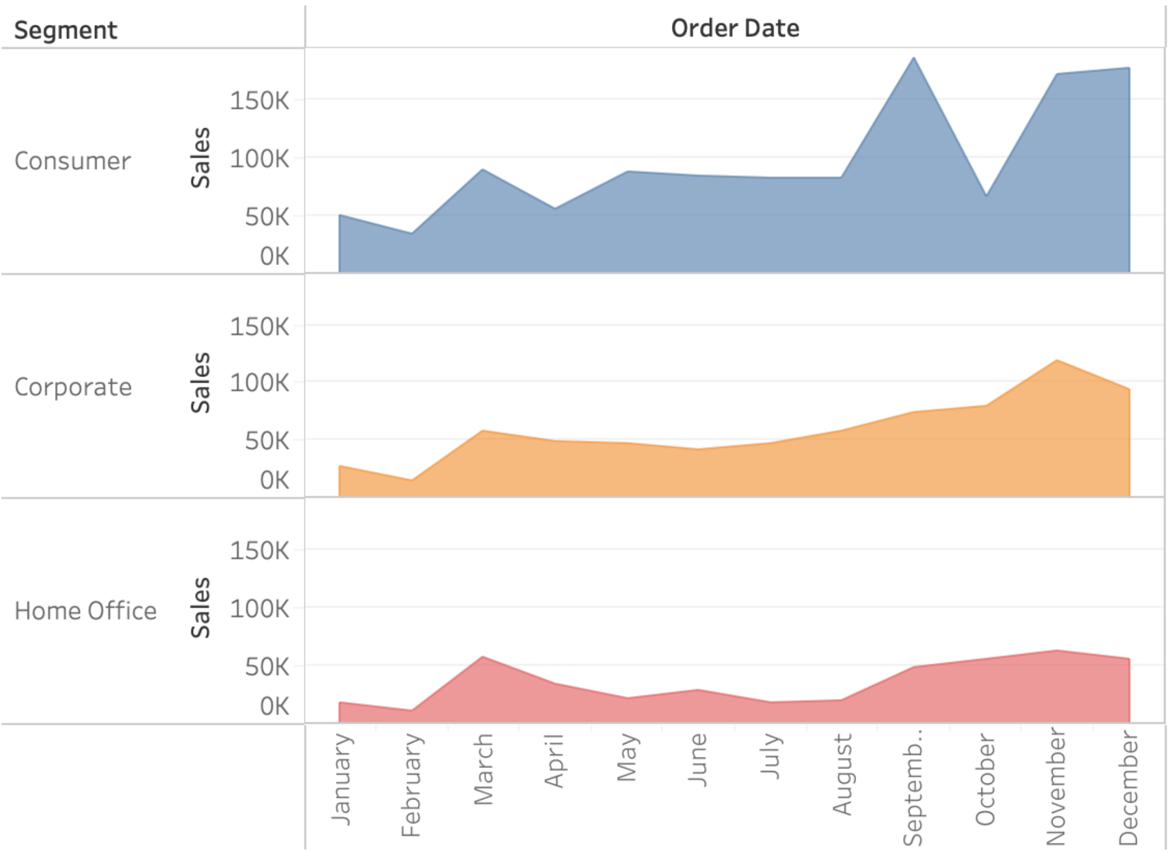
Area Chart - Stacked



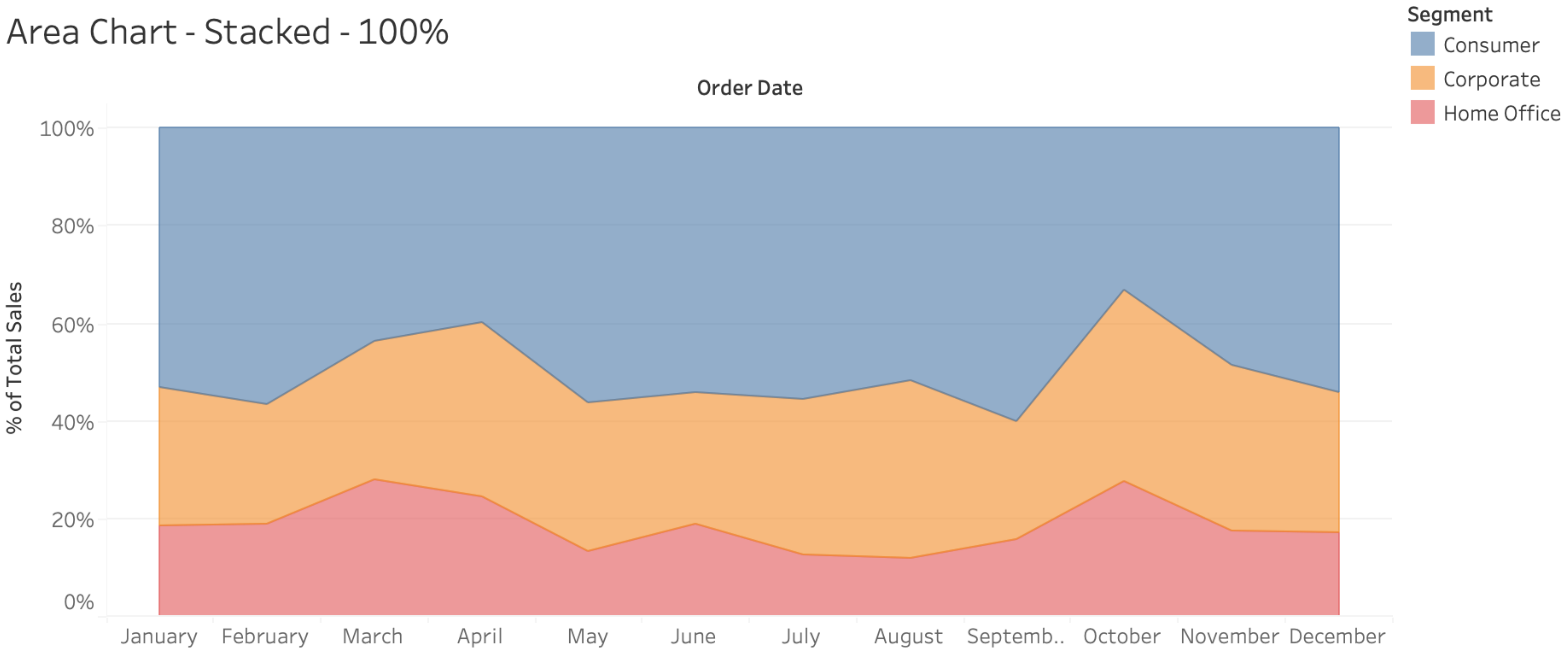
Bar Chart - Clustered



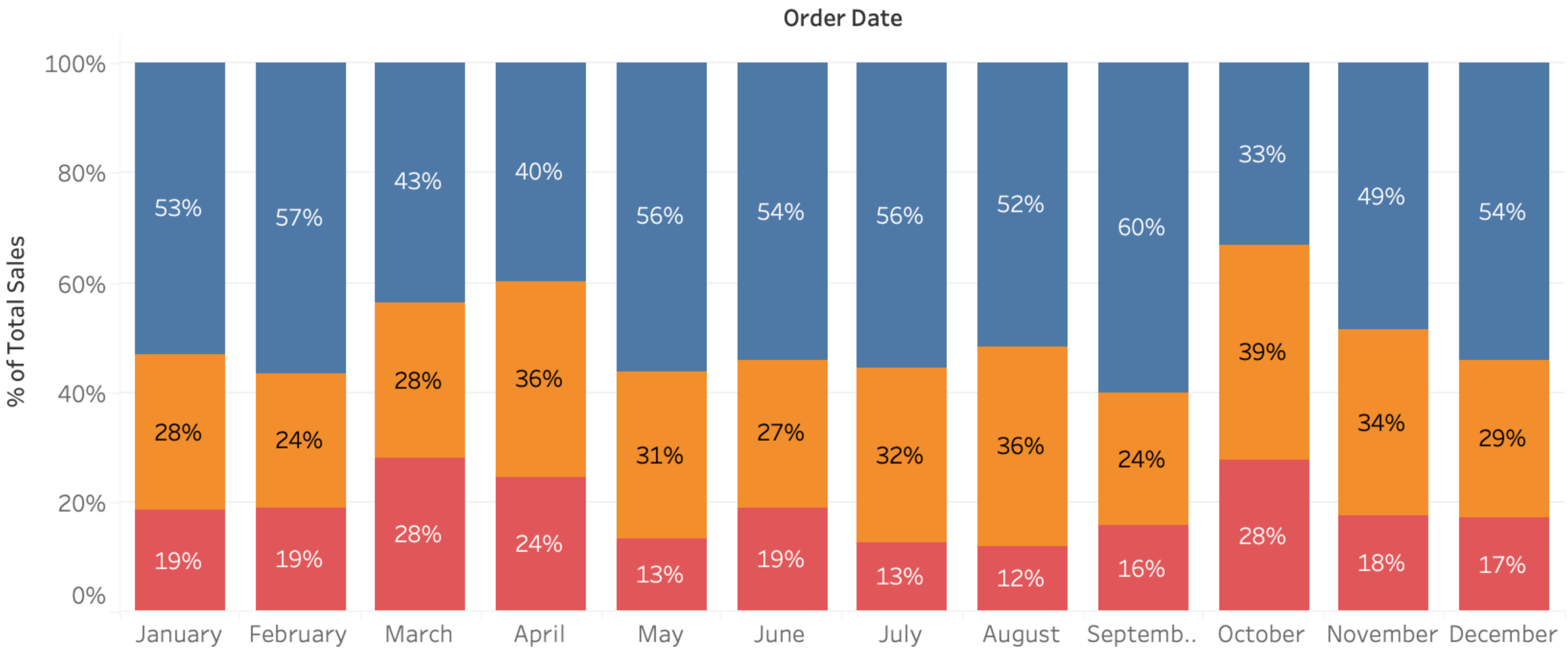
Area Chart - Aligned



Area Chart - Stacked - 100%

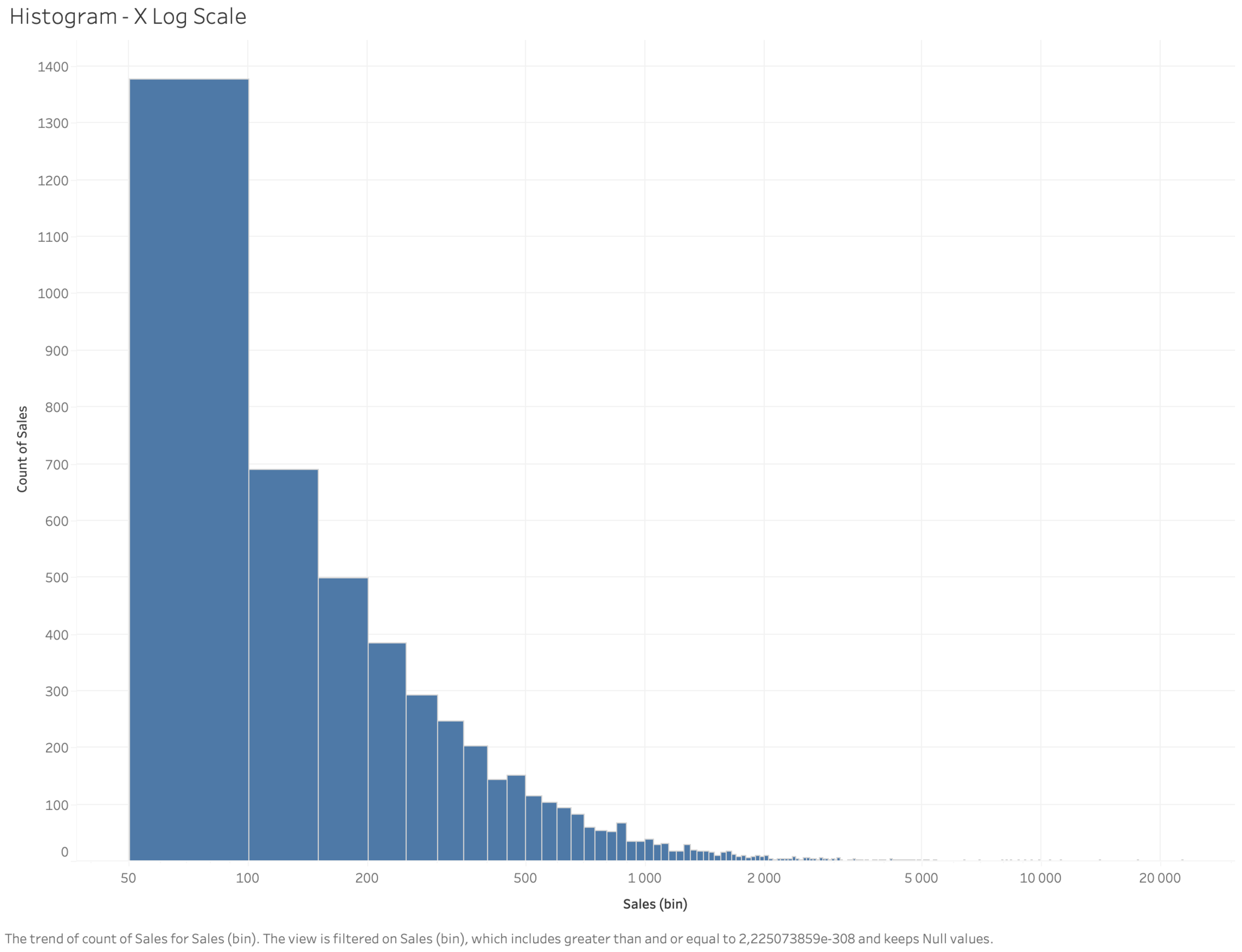


Bar Chart - Stacked - 100%

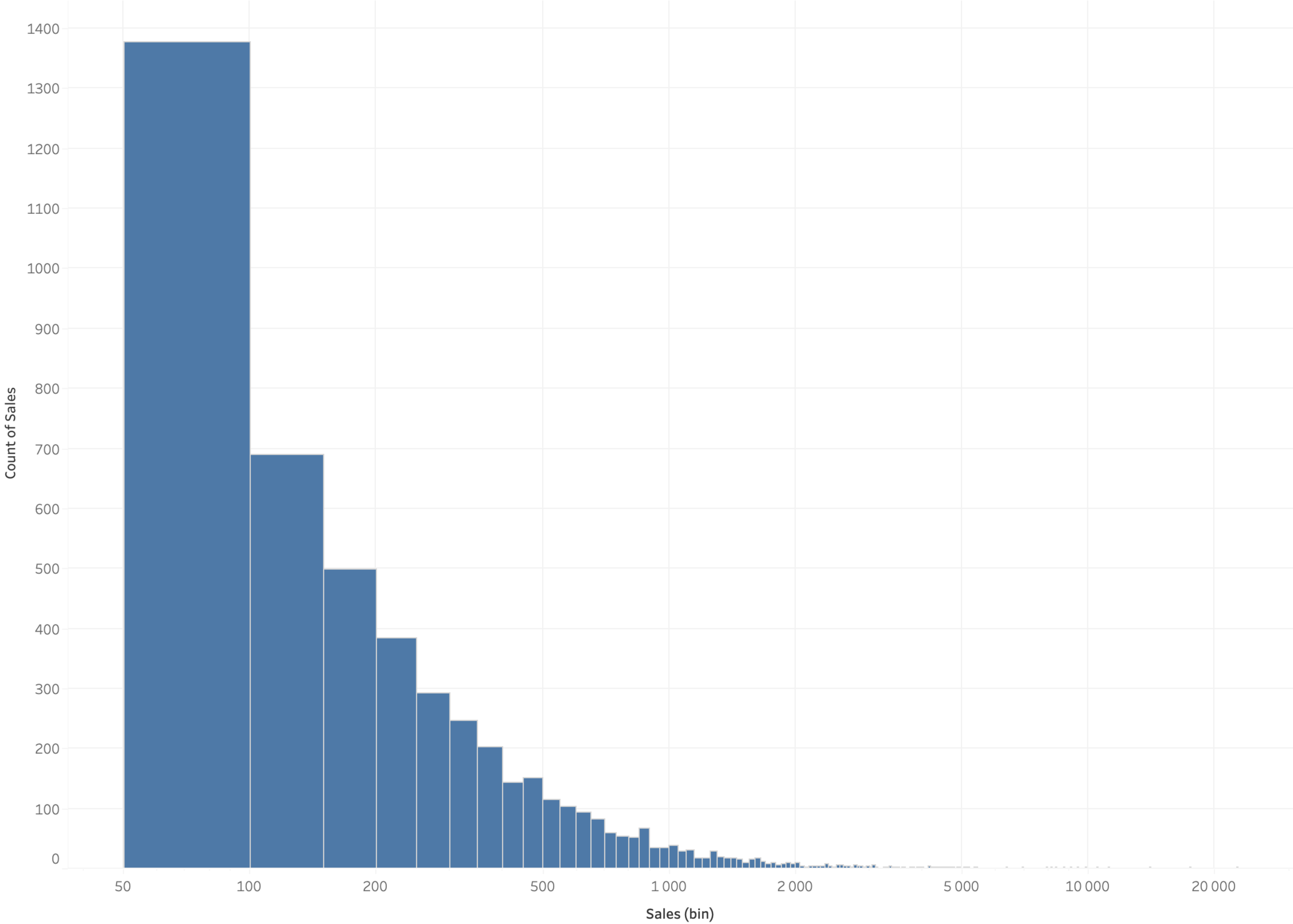


Multivariate Data: Region-Based Techniques

- **Histogram for continuous variables**

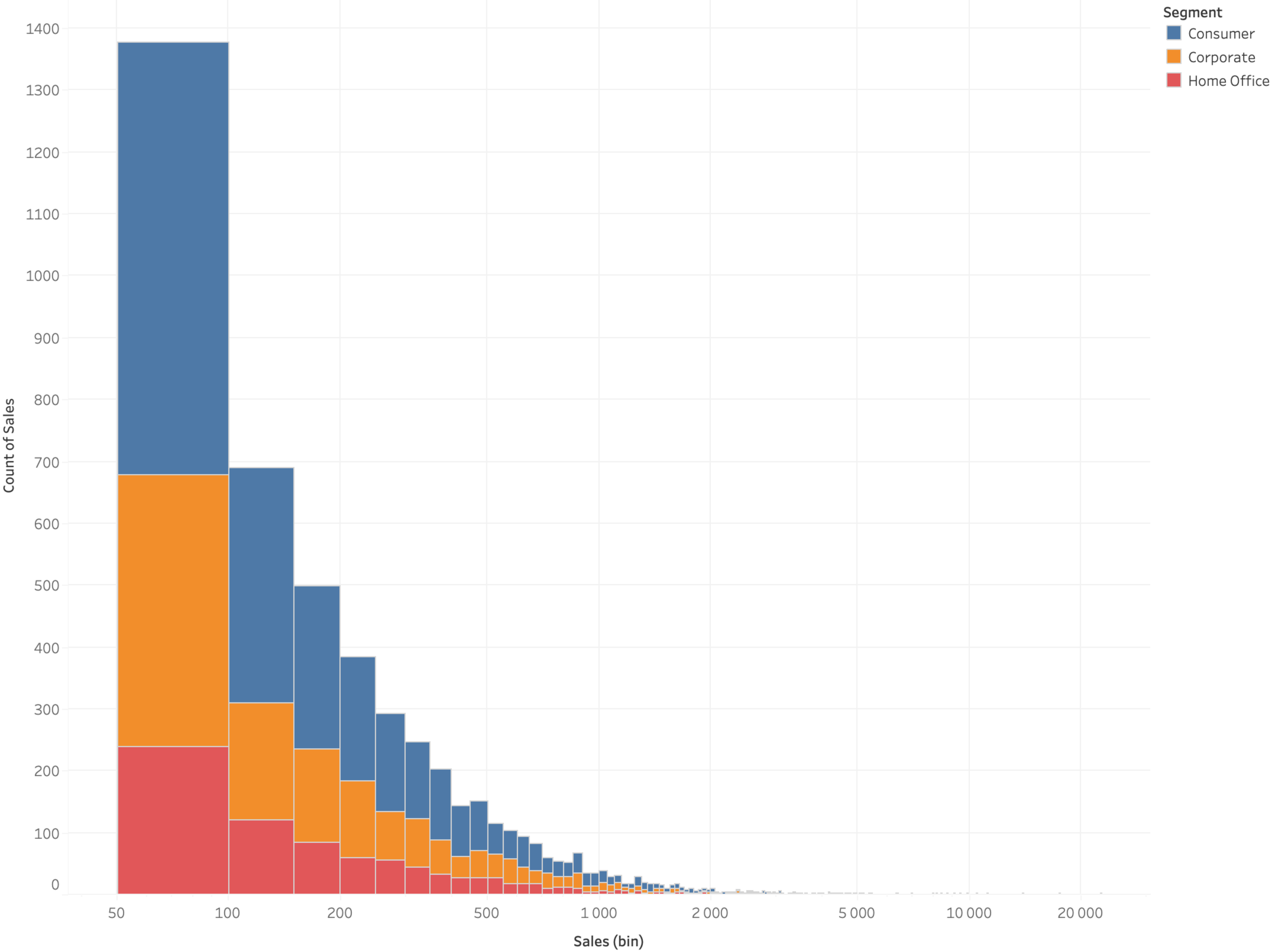


Histogram - X Log Scale



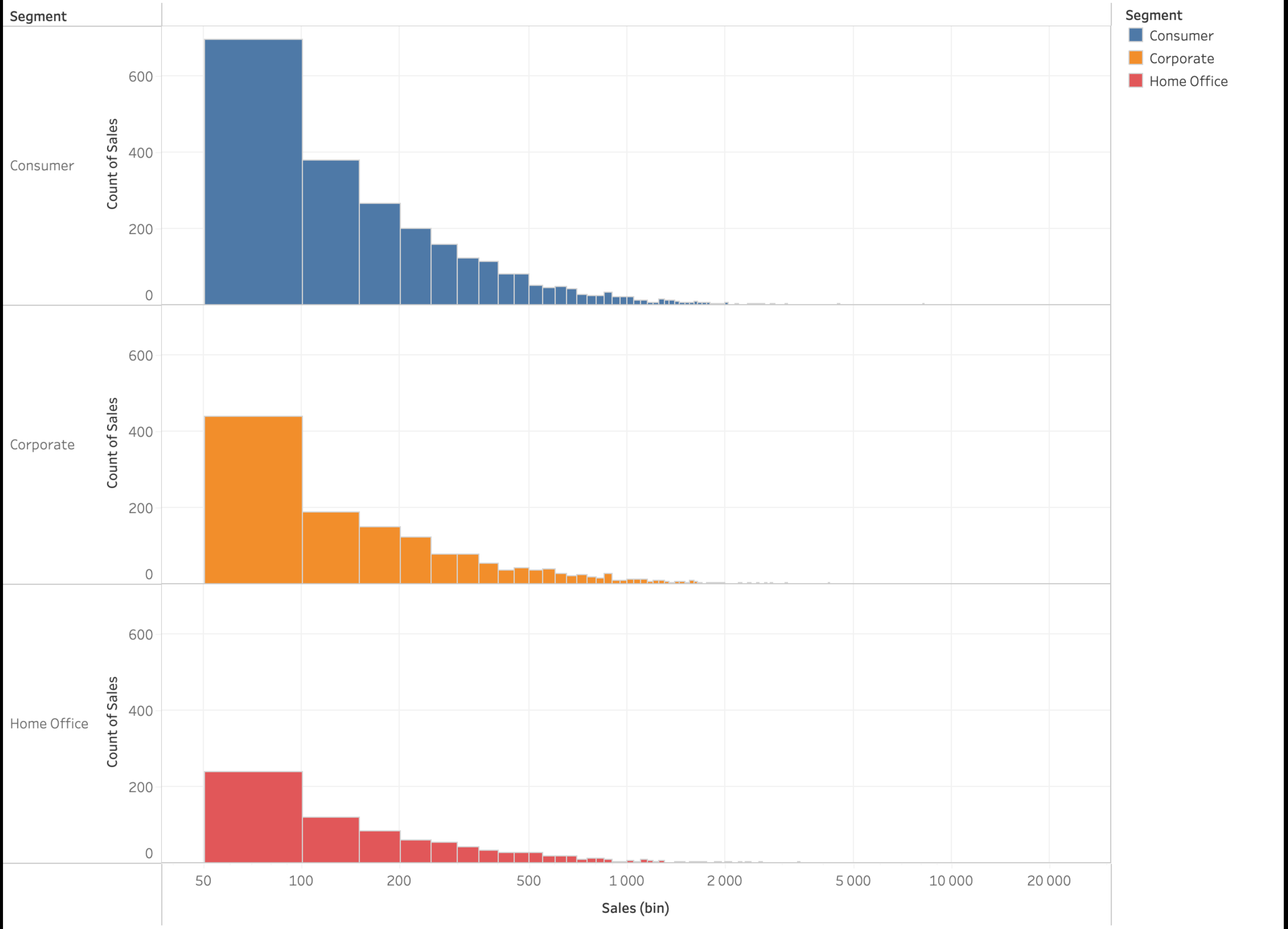
The trend of count of Sales for Sales (bin). The view is filtered on Sales (bin), which includes greater than and or equal to 2,225073859e-308 and keeps Null values.

Histogram - X Log Scale



The trend of count of Sales for Sales (bin). Color shows details about Segment. The view is filtered on Sales (bin), which includes greater than and or equal to 2,225073859e-308 and keeps Null values.

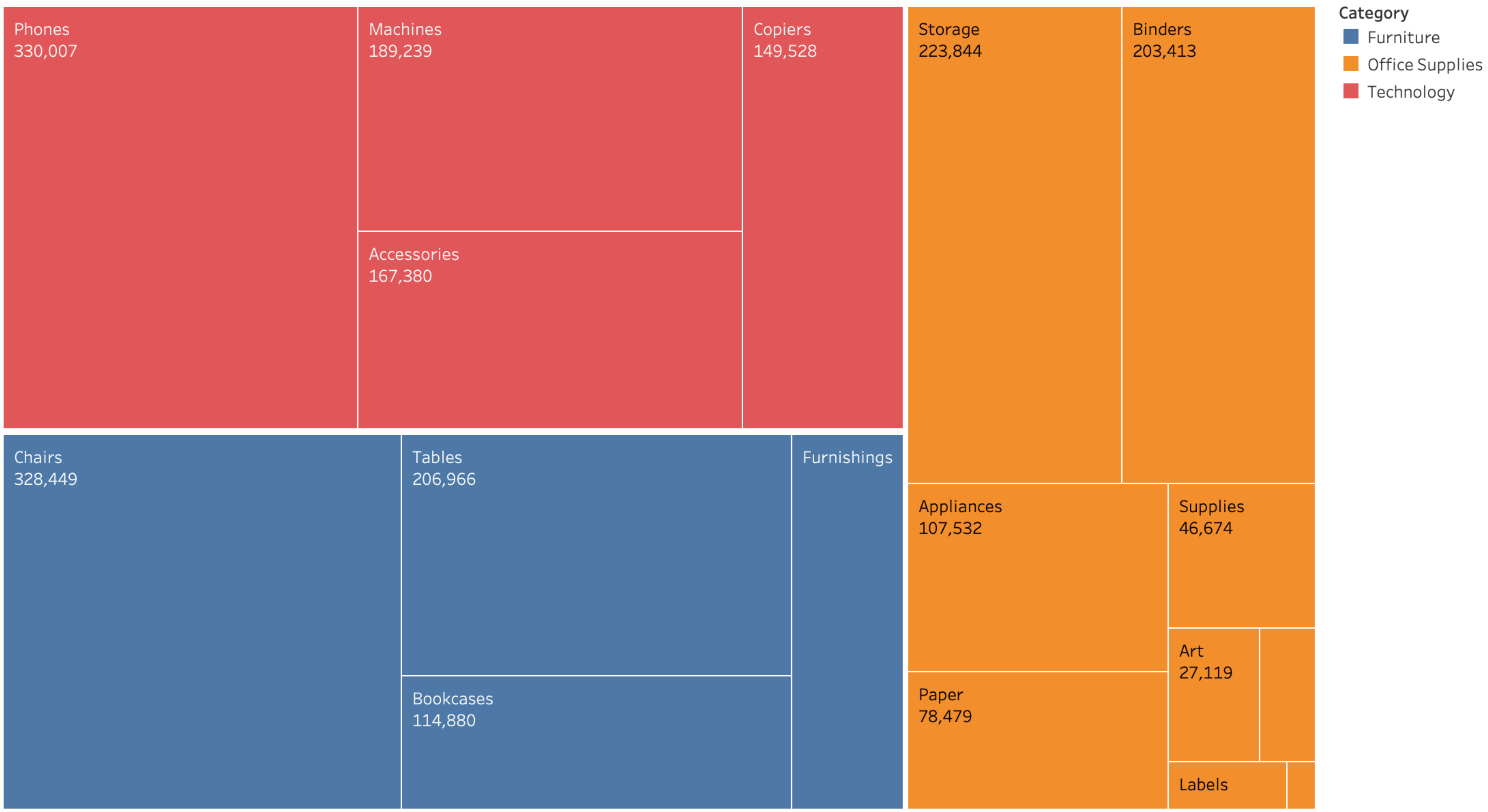
Histogram - X Log Scale



The trend of count of Sales for Sales (bin) broken down by Segment. Color shows details about Segment. The view is filtered on Sales (bin), which includes greater than and or equal to 2,225073859e-308 and keeps Null values.

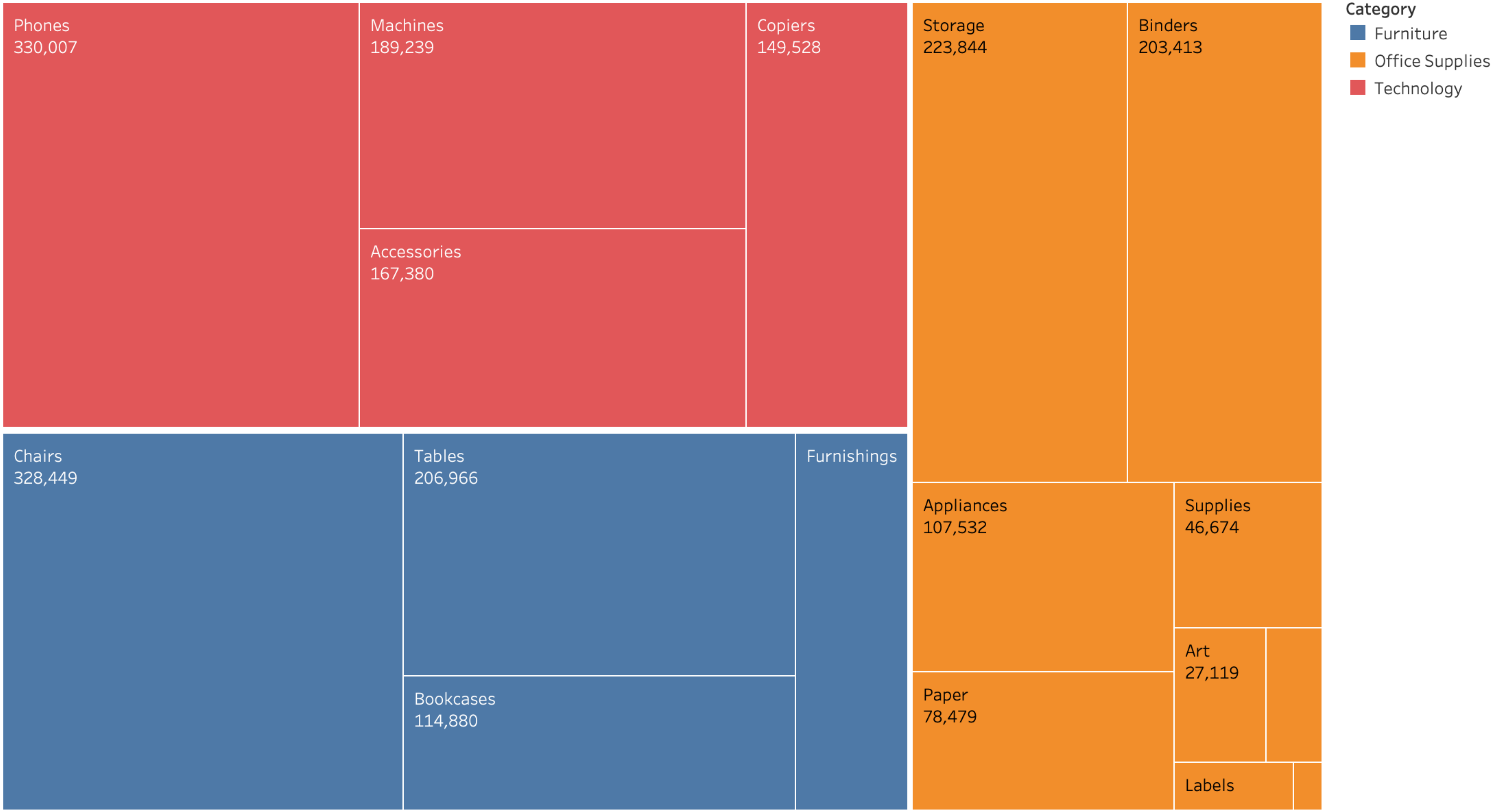
Multivariate Data: Region-Based Techniques

- **Tree Maps**

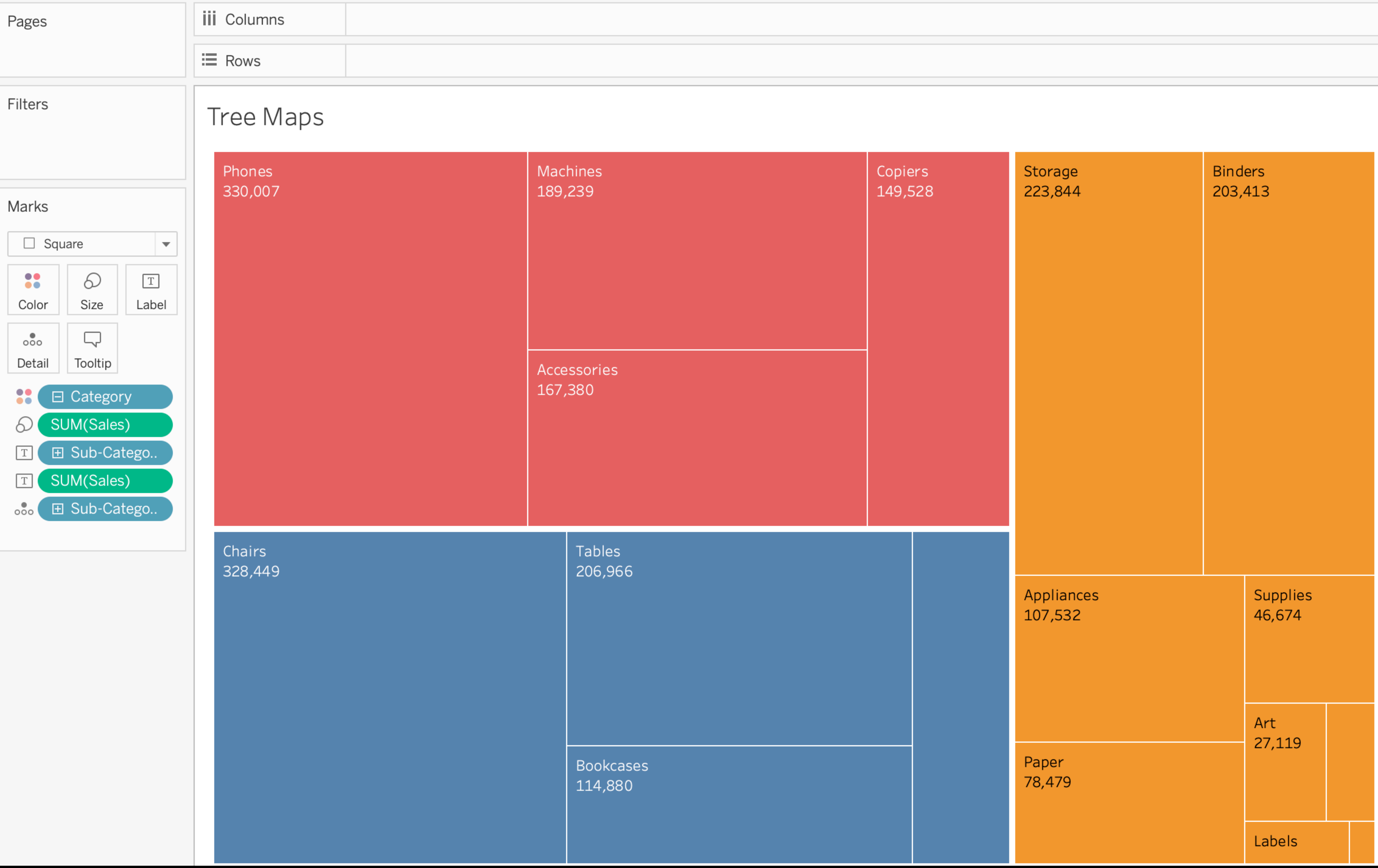


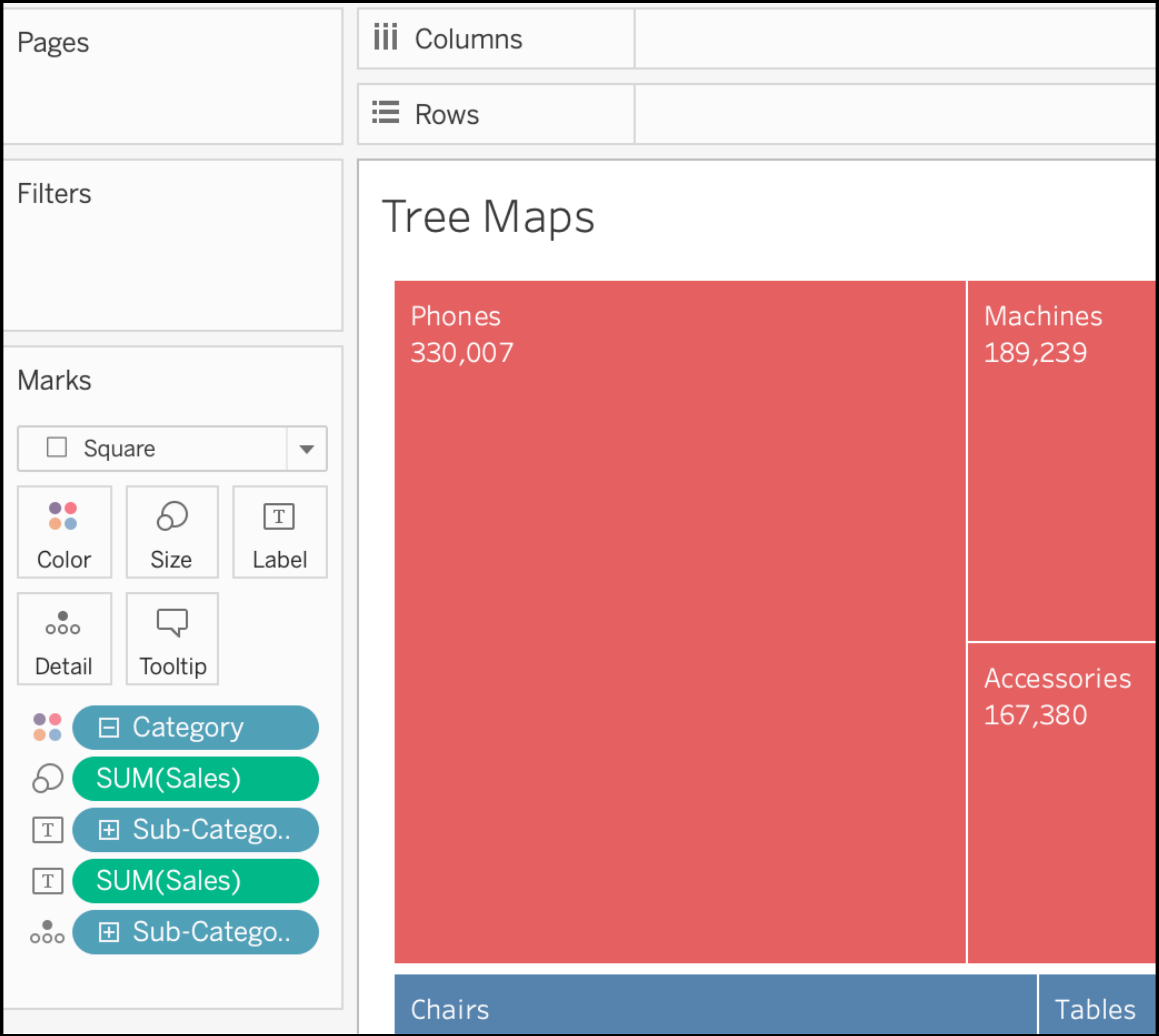
Sub-Category and sum of Sales. Color shows details about Category. Size shows sum of Sales. The marks are labeled by Sub-Category and sum of Sales. Details are shown for Sub-Category.

Tree Maps



Sub-Category and sum of Sales. Color shows details about Category. Size shows sum of Sales. The marks are labeled by Sub-Category and sum of Sales. Details are shown for Sub-Category.


Pages
Columns
Rows
Filters
Marks
Square
Color
Size
Label
Detail
Tooltip
Category
SUM(Sales)
Sub-Catego..
SUM(Sales)
Sub-Catego..
Tree Maps
Phones 330,007
Machines 189,239
Copiers 149,528
Accessories 167,380
Storage 223,844
Binders 203,413
Chairs 328,449
Tables 206,966
Bookcases 114,880
Appliances 107,532
Supplies 46,674
Art 27,119
Paper 78,479
Labels



Pages

Filters

Marks

□ Square

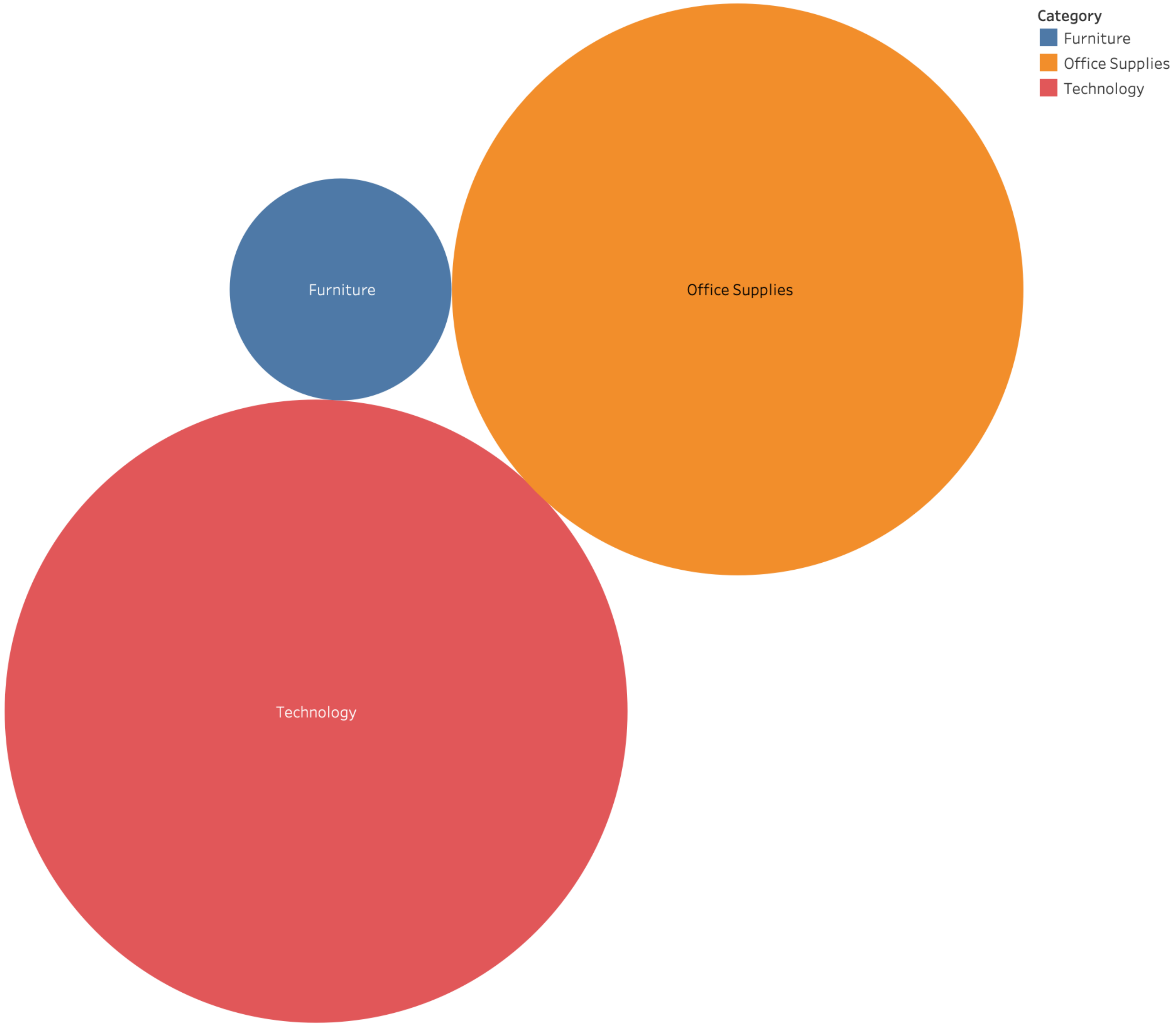
Color / Size / Label / Detail / Tooltip

- Color: Category
- Size: SUM(Sales)
- Label: Sub-Catego..
- Label: SUM(Sales)
- Detail: Sub-Catego..

Multivariate Data: Region-Based Techniques

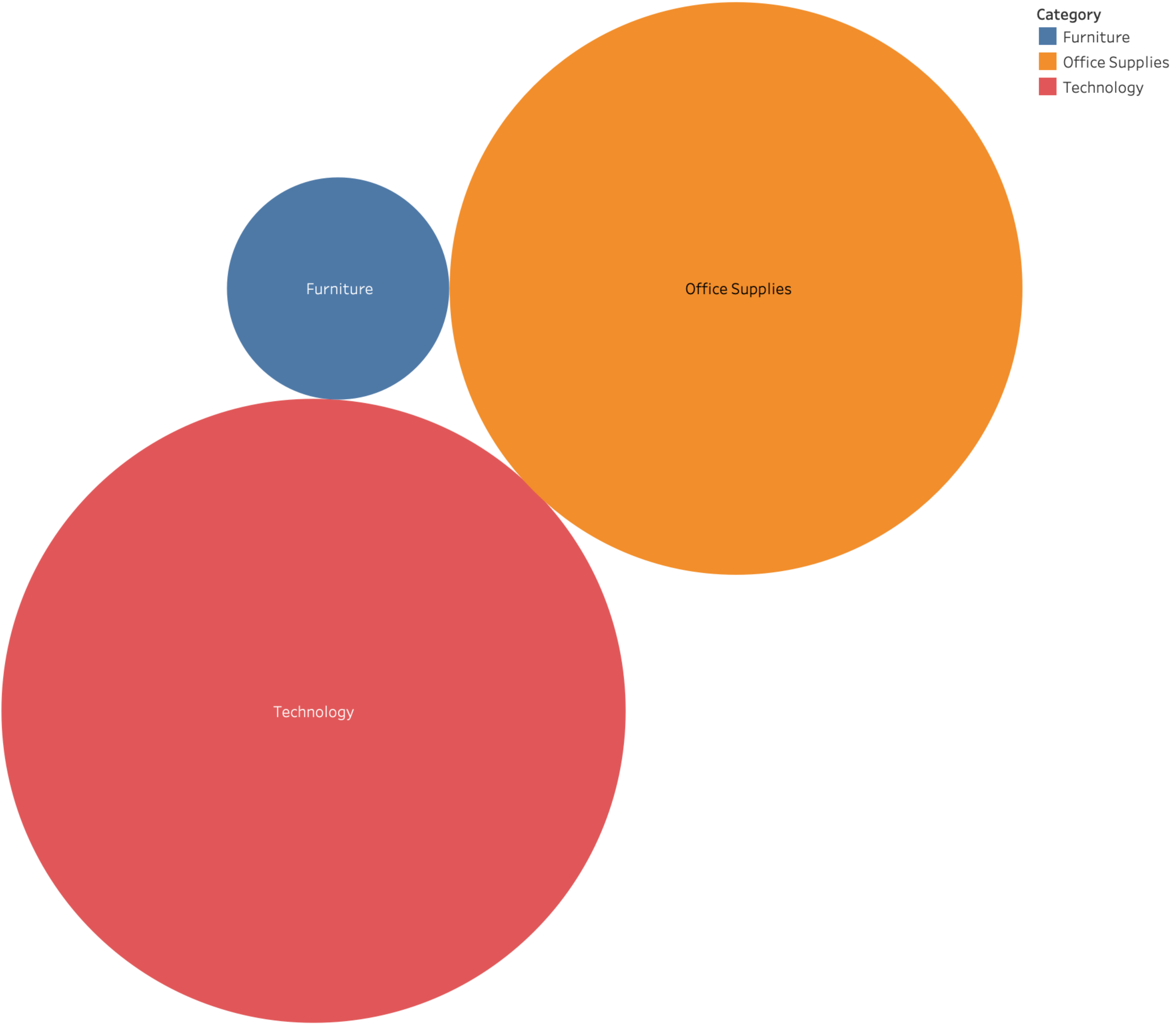
- **Stacked Bubbles**

Packed Bubbles



Category. Color shows details about Category. Size shows sum of Profit. The marks are labeled by Category.

Packed Bubbles

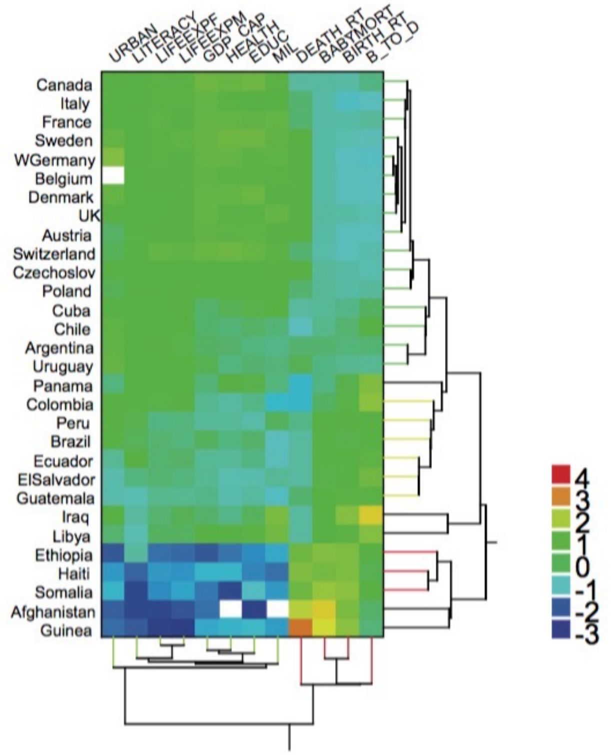


Category. Color shows details about Category. Size shows sum of Profit. The marks are labeled by Category.

Multivariate Data: Region-Based Techniques

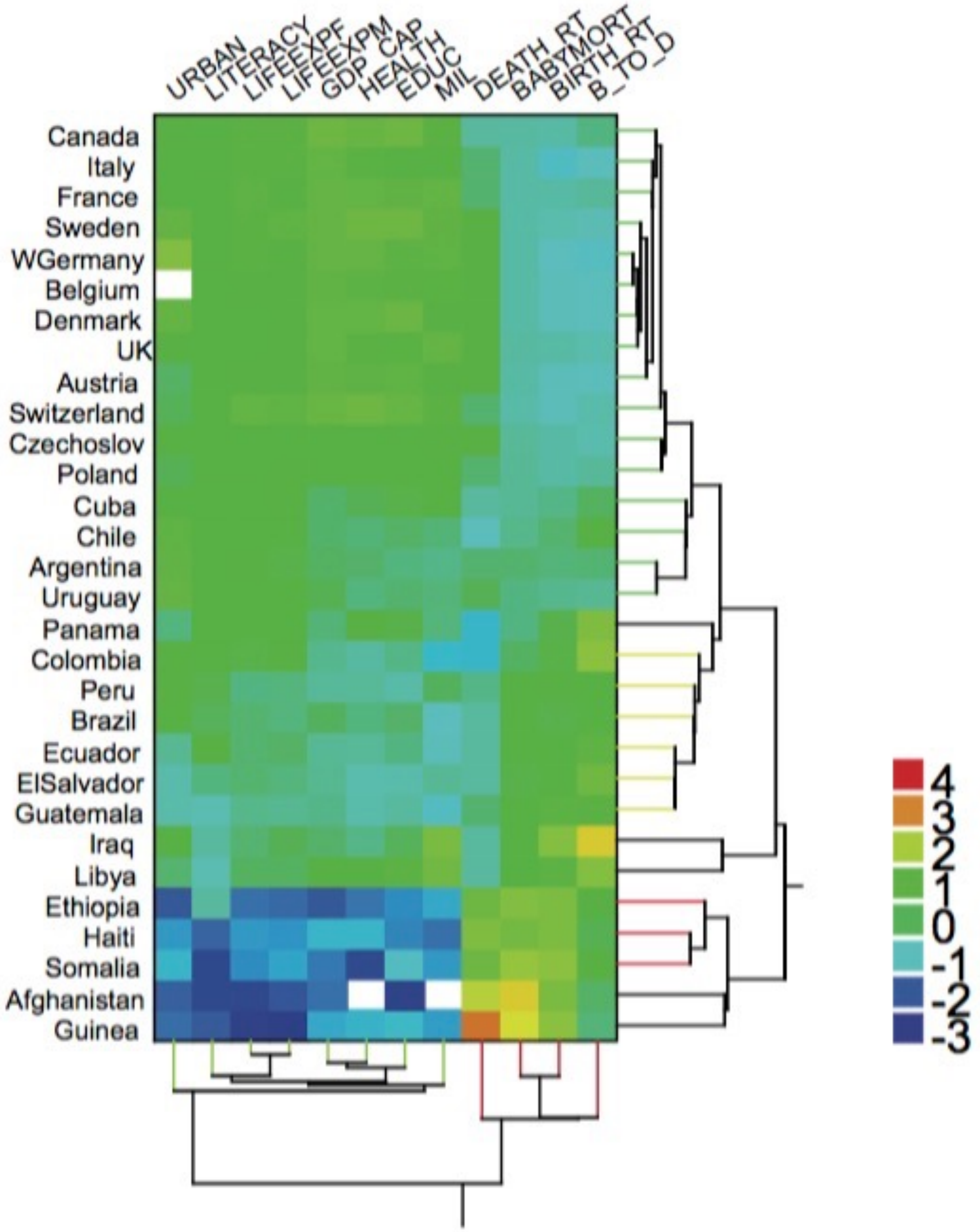
- **Tabular Displays**

  - **Heatmaps** are created by displaying the table of record values **using color rather than text**. All **data values are mapped to the same normalized color space**, and each is rendered as a colored square or rectangle.

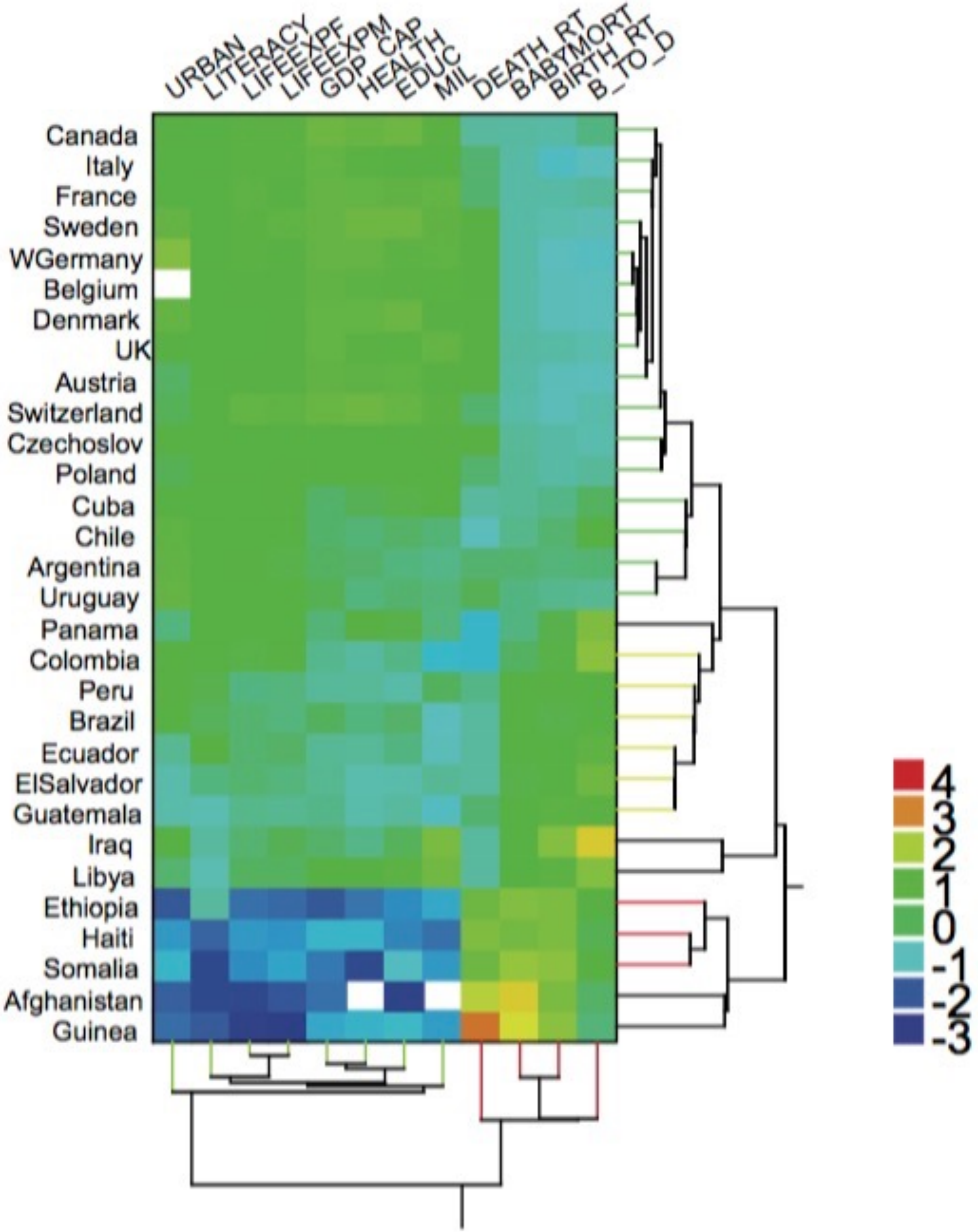


A heatmap showing social statistics for several countries from a U.N. survey. Rows and columns have been reordered via clustering. (Image courtesy Leland Wilkinson [459].)

Multivariate Data: Region-Based Techniques



A heatmap showing social statistics for several countries from a U.N. survey. Rows and columns have been reordered via clustering. (Image courtesy Leland Wilkinson [459].)


URBAN
LITERACY
LIFEEXPF
LIFEEXPM
GDP_CAP
HEALTH
EDUC
MIL
DEATH_RT
BABYMORT
BIRTH_RT
B_TO_D
Canada
Italy
France
Sweden
WGermany
Belgium
Denmark
UK
Austria
Switzerland
Czechoslov
Poland
Cuba
Chile
Argentina
Uruguay
Panama
Colombia
Peru
Brazil
Ecuador
ElSalvador
Guatemala
Iraq
Libya
Ethiopia
Haiti
Somalia
Afghanistan
Guinea
4
3
2
1
0
-1
-2
-3

Heat Table

	Segment		
Sub-Category	Consumer	Corporate	Home Office
Phones	169,933	91,153	68,921
Chairs	172,863	99,141	56,445
Storage	100,492	79,791	43,560
Tables	99,934	70,872	36,160
Binders	118,161	51,560	33,691
Machines	79,543	60,277	49,419
Accessories	87,105	48,191	32,085
Copiers	69,819	46,829	32,880
Bookcases	68,633	34,006	12,241
Appliances	52,820	36,589	18,124
Furnishings	49,620	25,001	17,084
Paper	36,324	23,883	18,272
Supplies	25,741	19,435	1,497
Art	14,252	8,590	4,276
Envelopes	7,771	5,943	2,763
Labels	6,709	4,102	1,675
Fasteners	1,681	783	560

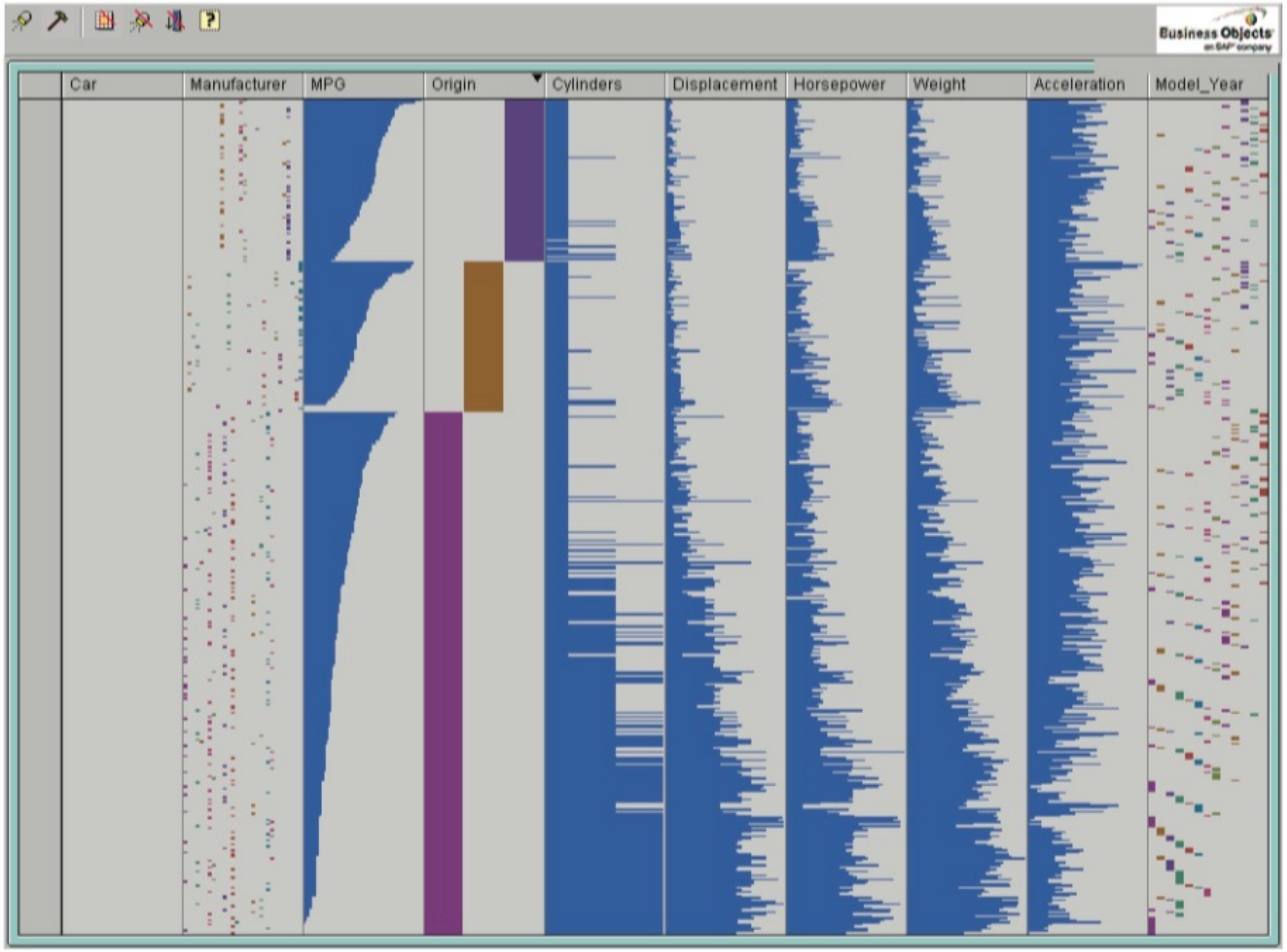
Sales

560 172,863

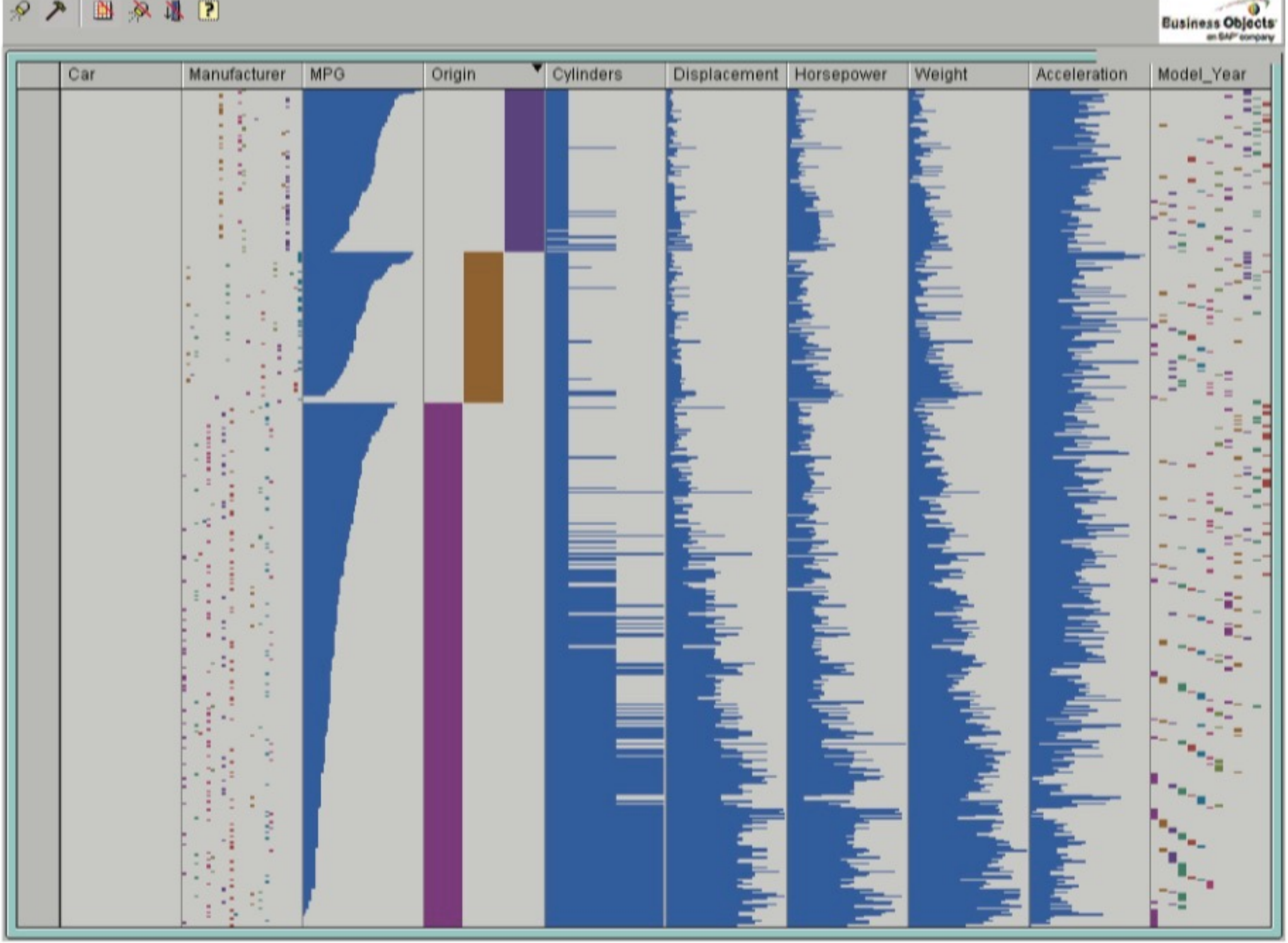
Sum of Sales (color) broken down by Segment vs. Sub-Category.

Multivariate Data: Region-Based Techniques

- **table lens** combines all these ideas and includes a **level-of-detail mechanism** for providing panning and zooming capabilities to display whole table views, while still providing some detail through local table lenses



An example of Inxight Table Lens showing the cars data set sorted first by car origin and then by MPG.



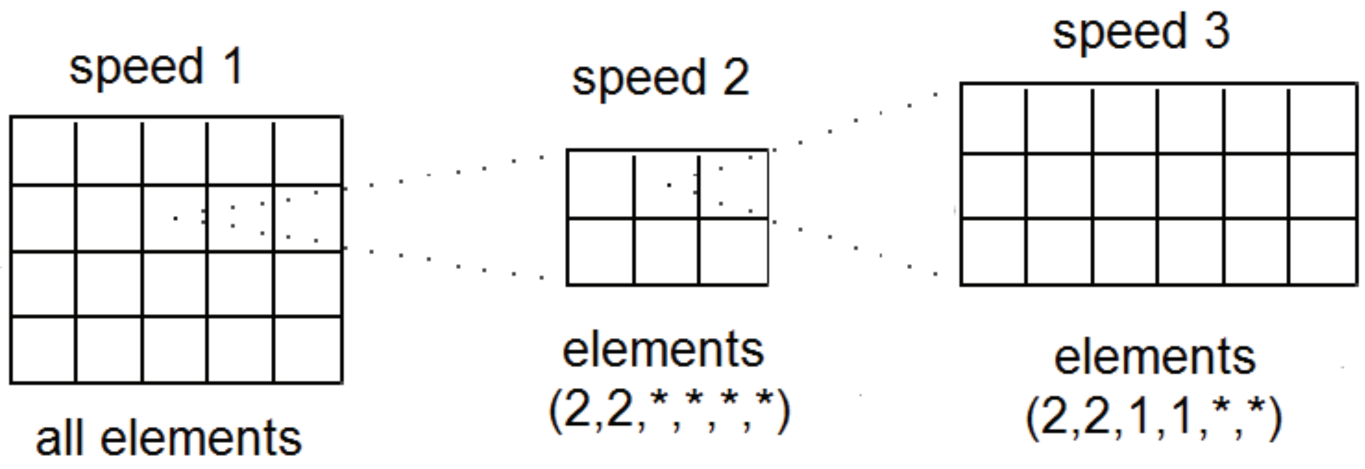
An example of Inxight Table Lens showing the cars data set sorted first by car origin and then by MPG.

Multivariate Data: Region-Based Techniques

- **Dimensional Stacking**
  - Begin with data of dimension **2N** + 1 (for an even number of dimensions there would be an additional implicit dimension of cardinality one).
  - Select a **finite cardinality/discretization** for each dimension.
  - Choose **one** of the dimensions **to be the dependent variable**. The rest will be considered independent
  - Create ordered pairs of the independent dimensions (**N pairs**) and assign to each pair a unique value (speed) from 1 to N.
  - The pair corresponding to speed 1 will create a virtual image whose size coincides with the cardinality of the dimensions (the first dimension in the pair is oriented horizontally, the second vertically).

Multivariate Data: Region-Based Techniques

- **Dimensional Stacking**
  - Create ordered pairs of the independent dimensions (**N pairs**) and assign to each pair a unique value (speed) from 1 to N.
  - The pair corresponding to speed 1 will create a virtual image whose size coincides with the cardinality of the dimensions (the first dimension in the pair is oriented horizontally, the second vertically).

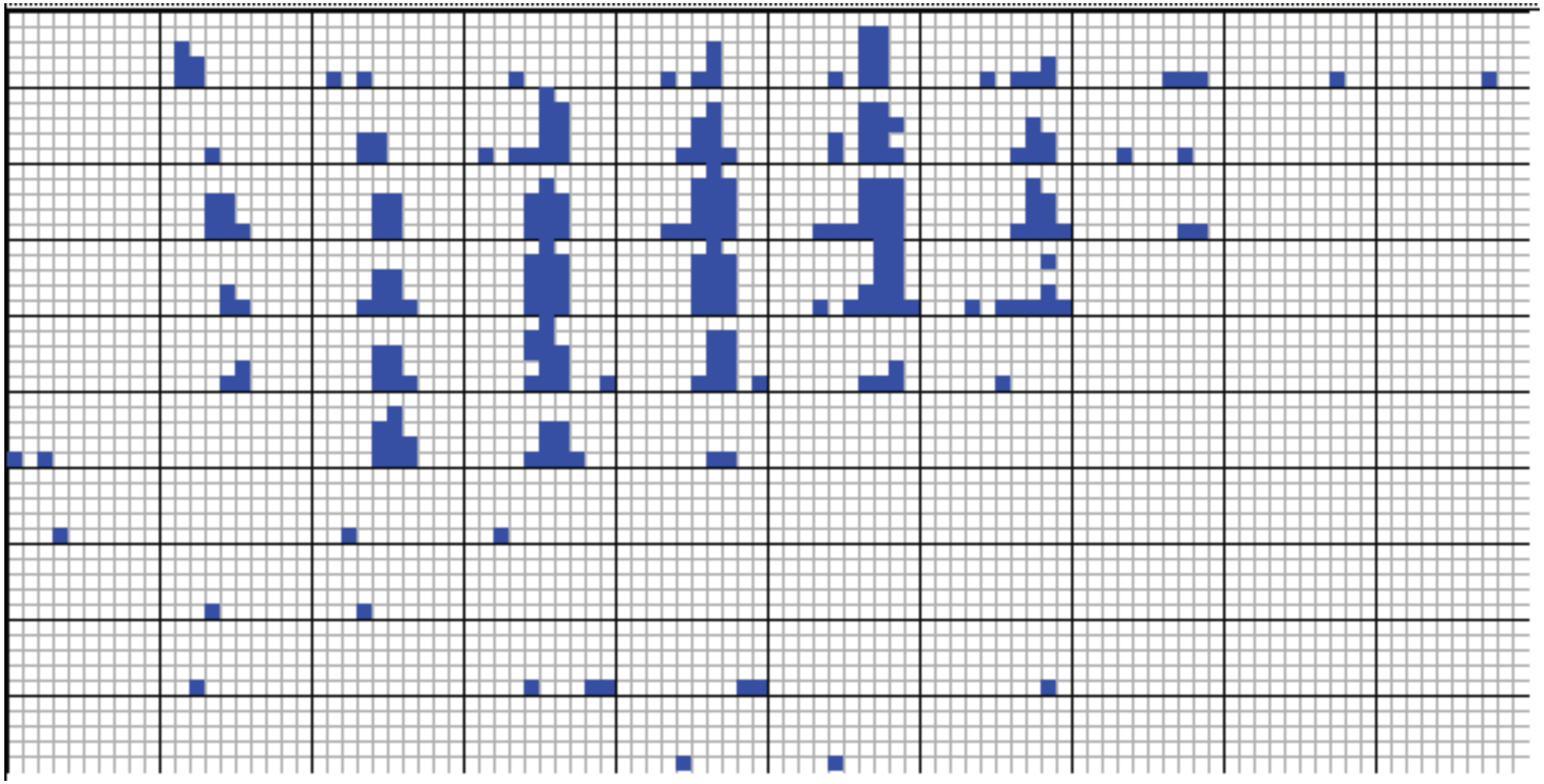


Conceptualization of dimensional stacking; collapsing six dimensions into two dimensions.

d1,. . . , d6 have cardinalities 4, 5, 2, 3, 3, and 6, respectively

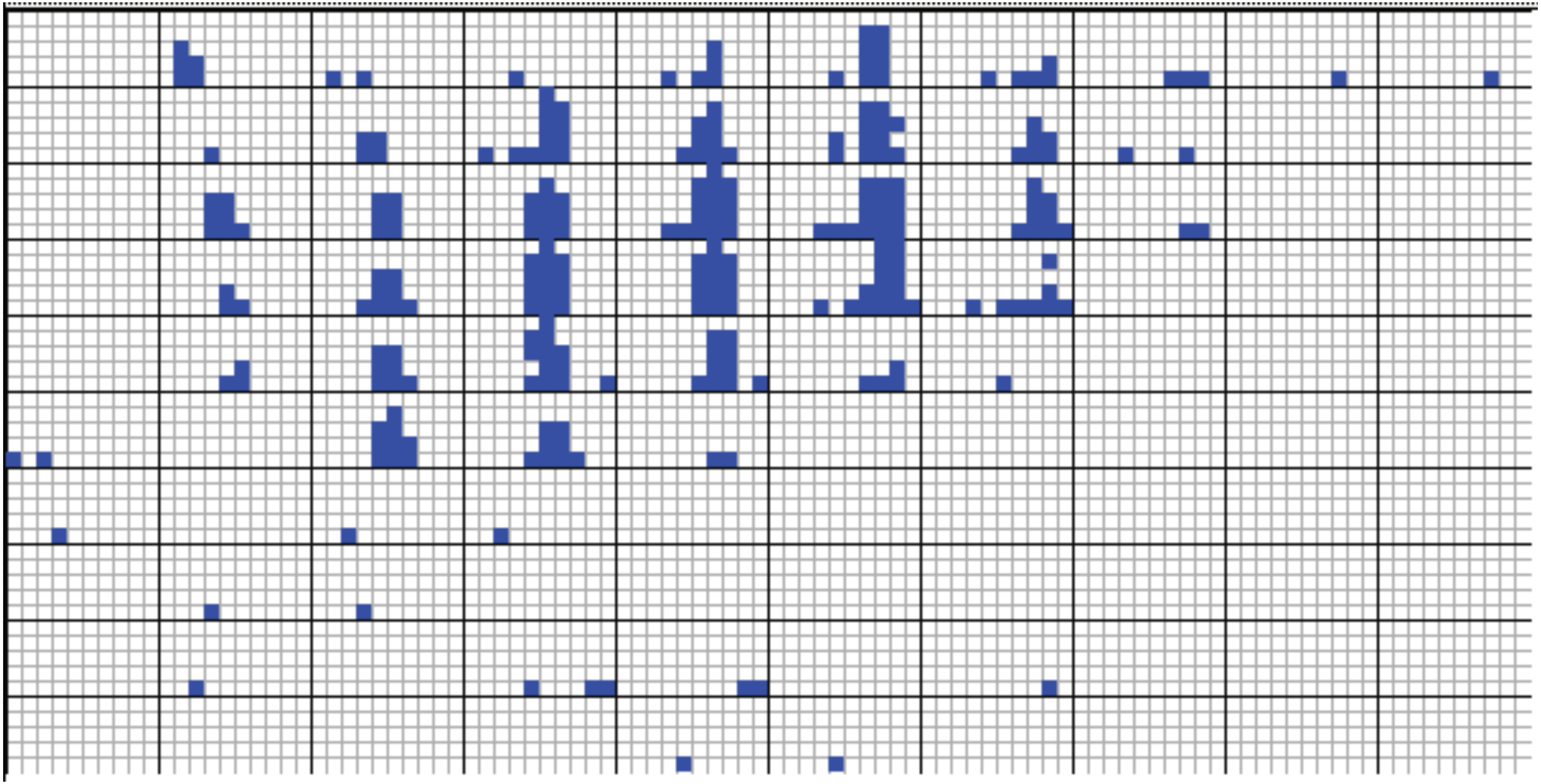
Multivariate Data: Region-Based Techniques

- **Dimensional Stacking**

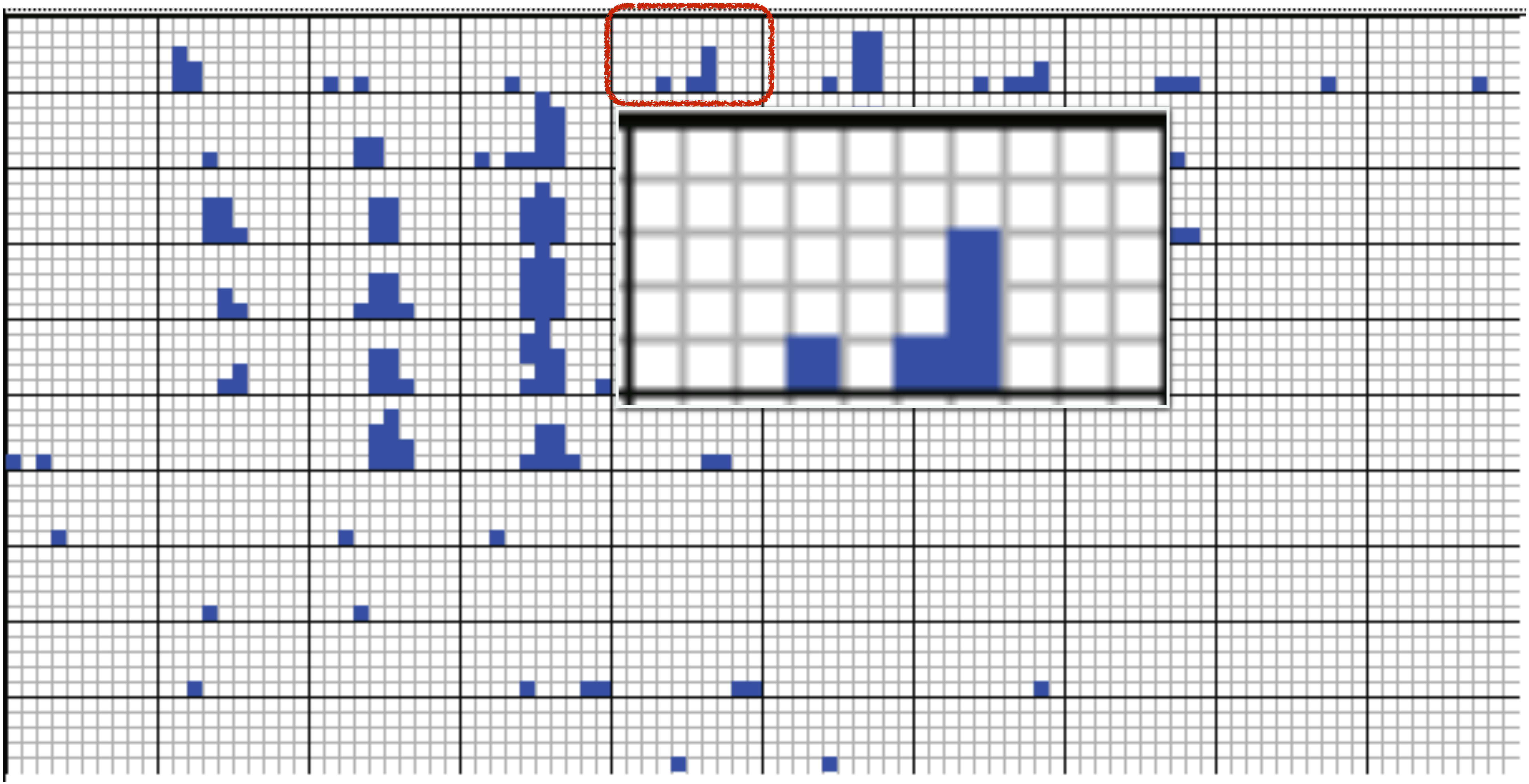


An example of 4D data visualized using dimensional stacking. The data consists of drill-hole data, with three spatial dimensions, and the ore grade as the fourth dimension.

Multivariate Data: Region-Based Techniques



Multivariate Data: Region-Based Techniques



Combinations of Techniques

Multivariate Data: Combinations of Techniques

- **Glyphs and Icons**
- **Dense Pixel Displays**
- **Many others**

Multivariate Data: Combinations of Techniques

- **Glyphs and Icons**

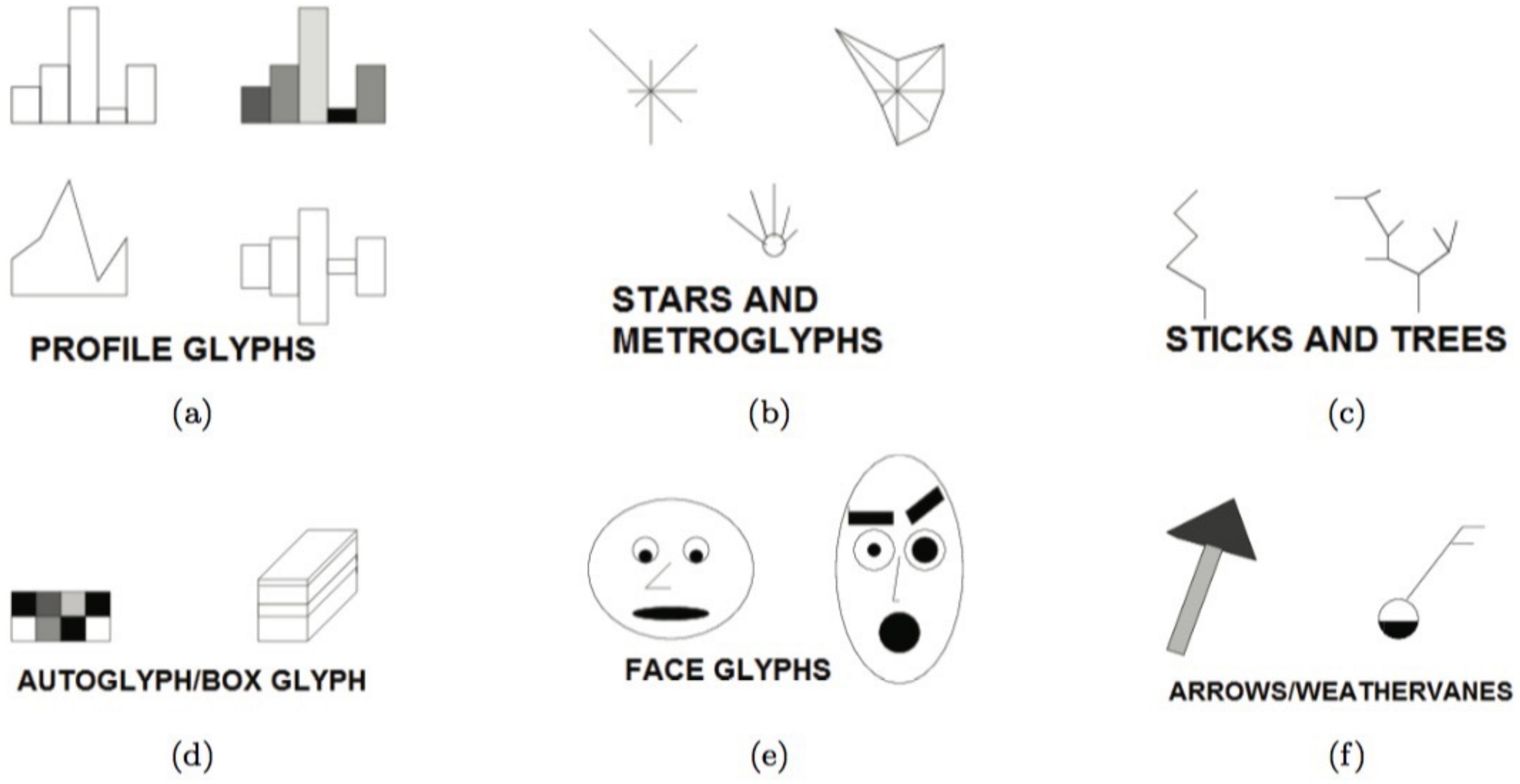


Figure 8.20. Examples of multivariate glyphs (from [445]).

Further Reading and Summary

Further Reading

- **Recommend Readings**
  - Interactive Data Visualization: Foundations, Techniques, and Applications, Matthew O. Ward et all, 2015, pages 285-314.
- **Supplemental readings:**
  - Visualization Analysis & Design , Tamara Munzner, Chapter 7

What you should know

- **Point based techniques**
  - Classical point base techniques have a limited dimensionality - Scatter based
  - Dimension reduction or selection for data viz
- **Line based**
  - Classical line based
  - Radial Axis Techniques
  - Parallel coordinates techniques and related stuff
- **Region based**
  - Reordering the data in graphical tables
- **Combination Techniques**
  - Dense
  - Glyphs