

Knowing Your Data (Data Visualization)

Prepared by Dr. Amjad Qtaish

DS601
Data Mining

Lecture 03



Lecture: #03

Knowing Your Data (Data Visualization)



Dr. Amjad Qtaish

Faculty of Informatics
Midocean University



Data Mining (DS601)



Knowing Your Data (Data Visualization)

- **Data Objects and Attribute Types** 
- Basic Statistical Descriptions of Data
- Data Visualization
- Measuring Data Similarity and Dissimilarity
- Summary





- **Record**
 - Relational records
 - Data matrix, e.g., numerical matrix, crosstabs
 - Document data: text documents: term-frequency vector
 - Transaction data
- **Graph and network**
 - World Wide Web
 - Social or information networks
 - Molecular Structures
- **Ordered**
 - Video data: sequence of images
 - Temporal data: time-series
 - Sequential Data: transaction sequences
 - Genetic sequence data
- **Spatial, image and multimedia**
 - Spatial data: maps
 - Image data:
 - Video data:

	team	coach	play	ball	score	game	win	lost	timeout	season
Document 1	3	0	5	0	2	6	0	2	0	2
Document 2	0	7	0	2	1	0	0	3	0	0
Document 3	0	1	0	0	1	2	2	0	3	0

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk



- **Dimensionality**
 - Curse of dimensionality
- **Sparsity**
 - Only presence counts
- **Resolution**
 - Patterns depend on the scale
- **Distribution**
 - Centrality and dispersion



- Data sets are made up of data objects.
- A **data object** represents an entity.
- Examples:
 - sales database: customers, store items, sales
 - medical database: patients, treatments
 - university database: students, professors, courses
- Also called *samples* , *examples*, *instances*, *data points*, *objects*, *tuples*.
- Data objects are described by **attributes**.
- Database rows -> data objects; columns -> attributes.



- **Attribute (or dimensions, features, variables):** a data field, representing a characteristic or feature of a data object.
 - *E.g., customer_ID, name, address*
- **Types:**
 - Nominal
 - Binary
 - Numeric: quantitative
 - Interval-scaled
 - Ratio-scaled



- **Nominal:** categories, states, or “names of things”
 - $Hair_color = \{auburn, black, blond, brown, grey, red, white\}$
 - marital status, occupation, ID numbers, zip codes
- **Binary**
 - Nominal attribute with only 2 states (0 and 1)
 - Symmetric binary: both outcomes equally important
 - e.g., gender
 - Asymmetric binary: outcomes not equally important.
 - e.g., medical test (positive vs. negative)
 - Convention: assign 1 to most important outcome (e.g., HIV positive)
- **Ordinal**
 - Values have a meaningful order (ranking) but magnitude between successive values is not known.
 - $Size = \{small, medium, large\}$, grades, army rankings



- Quantity (integer or real-valued)
- **Interval**
 - Measured on a scale of **equal-sized units**
 - Values have order
 - E.g., *temperature in C° or F°, calendar dates*
 - No true zero-point
- **Ratio**
 - Inherent **zero-point**
 - We can speak of values as being an order of magnitude larger than the unit of measurement (10 K° is twice as high as 5 K°).
 - e.g., *temperature in Kelvin, length, counts, monetary quantities*



- **Discrete Attribute**

- Has only a finite or countably infinite set of values
 - E.g., zip codes, profession, or the set of words in a collection of documents
- Sometimes, represented as integer variables
- Note: Binary attributes are a special case of discrete attributes

- **Continuous Attribute**

- Has real numbers as attribute values
 - E.g., temperature, height, or weight
- Practically, real values can only be measured and represented using a finite number of digits
- Continuous attributes are typically represented as floating-point variables



Knowing Your Data (Data Visualization)

- Data Objects and Attribute Types
- **Basic Statistical Descriptions of Data** 
- Data Visualization
- Measuring Data Similarity and Dissimilarity
- Summary





- Motivation
 - To better understand the data: central tendency, variation and spread
- Data dispersion characteristics
 - median, max, min, quantiles, outliers, variance, etc.
- Numerical dimensions correspond to sorted intervals
 - Data dispersion: analyzed with multiple granularities of precision
 - Boxplot or quantile analysis on sorted intervals
- Dispersion analysis on computed measures
 - Folding measures into numerical dimensions
 - Boxplot or quantile analysis on the transformed cube



- Mean (algebraic measure) (sample vs. population):

Note: n is sample size and N is population size.

- Weighted arithmetic mean:
- Trimmed mean: chopping extreme values

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad \mu = \frac{\sum x}{N}$$



- Median:
 - Middle value if odd number of values, or average of the middle two values otherwise
 - Estimated by interpolation (for grouped data):
 - where L_1 is the lower boundary of the median interval, 21
 - n is the number of values in the entire data set, (200+..+44)
 - $Sum(freq)_l$ is the sum of the frequencies of all of the intervals that are lower than the median interval.
 - $200+450+300 < 1597 < 200+450+300$ implies that the median interval is [21,50], So $L_1=21$,
 - $freq_median=1500$, $(sum\ of\ freq)_l=200+450+300=950$, width =30. The width should be 30, not 29, because there is no gap between intervals.

<i>age</i>	<i>frequency</i>
1–5	200
6–15	450
16–20	300
21–50	1500
51–80	700
81–110	44

$$median = L_1 + \left(\frac{n/2 - (\sum freq)_l}{freq_{median}} \right) width$$

$$L_1 = 20, n = 3194, (\sum freq)_l = 950, freq_{median} = 1500, width = 30, median = 30.94 \text{ years.}$$

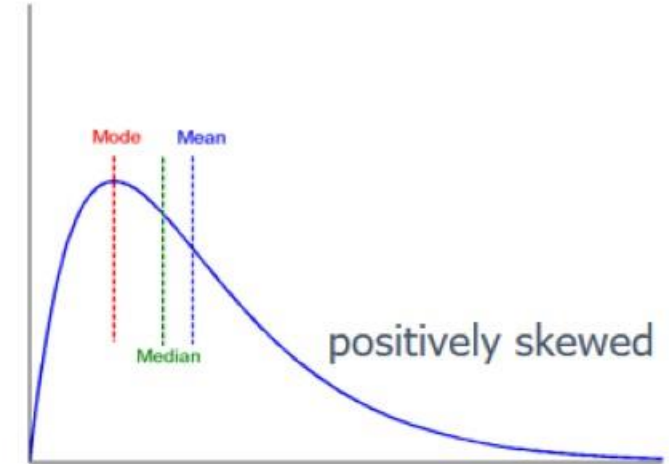
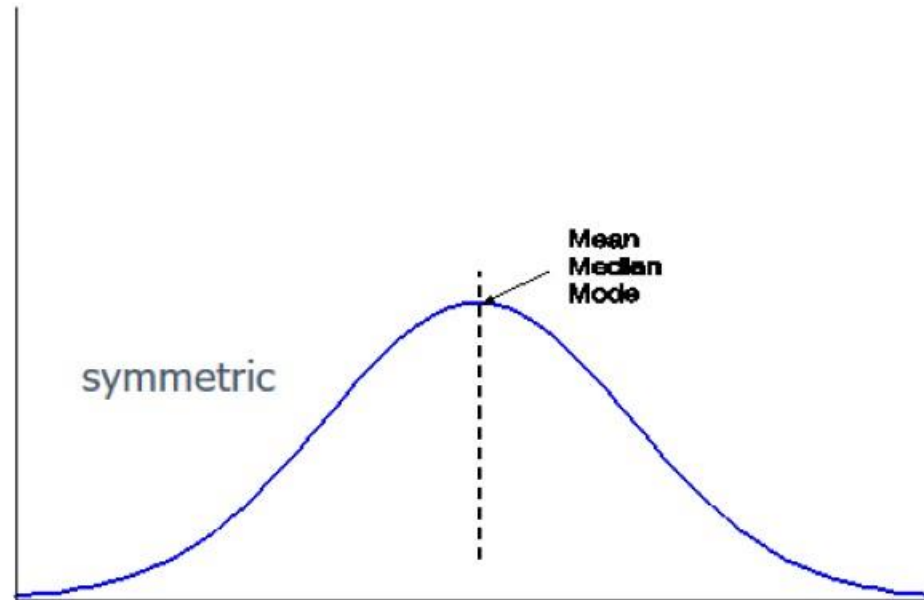


- Mode

- Value that occurs most frequently in the data
- Unimodal, bimodal, trimodal
- Empirical formula: $mean - mode \approx 3 \times (mean - median)$

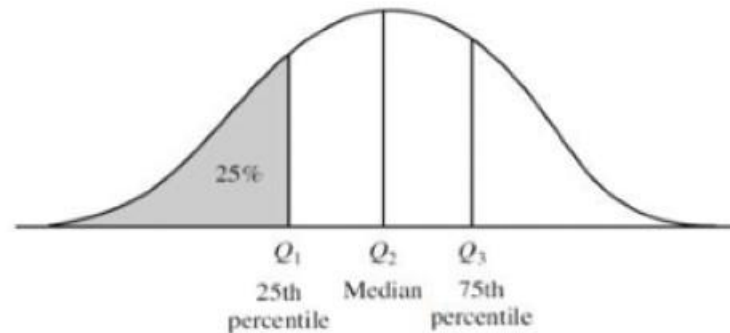


- Median, mean and mode of symmetric, positively and negatively skewed data.





- **Quantiles** are points taken at regular intervals of a data distribution, dividing it into essentially equal-size consecutive sets.
- The k th q -quantile for a given data distribution is the value x such that at most $k=q$ of the data values are less than x and at most .



A plot of the data distribution for some attribute X . The quantiles plotted are quartiles. The three quartiles divide the distribution into four equal-size consecutive subsets. The second quartile corresponds to the median.

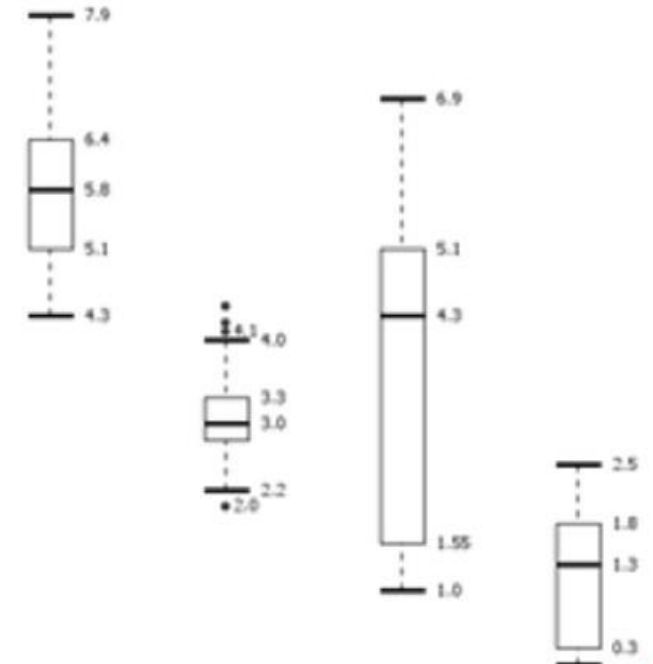
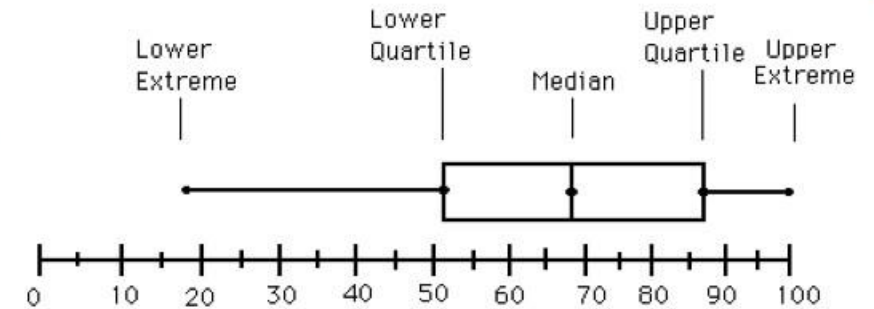


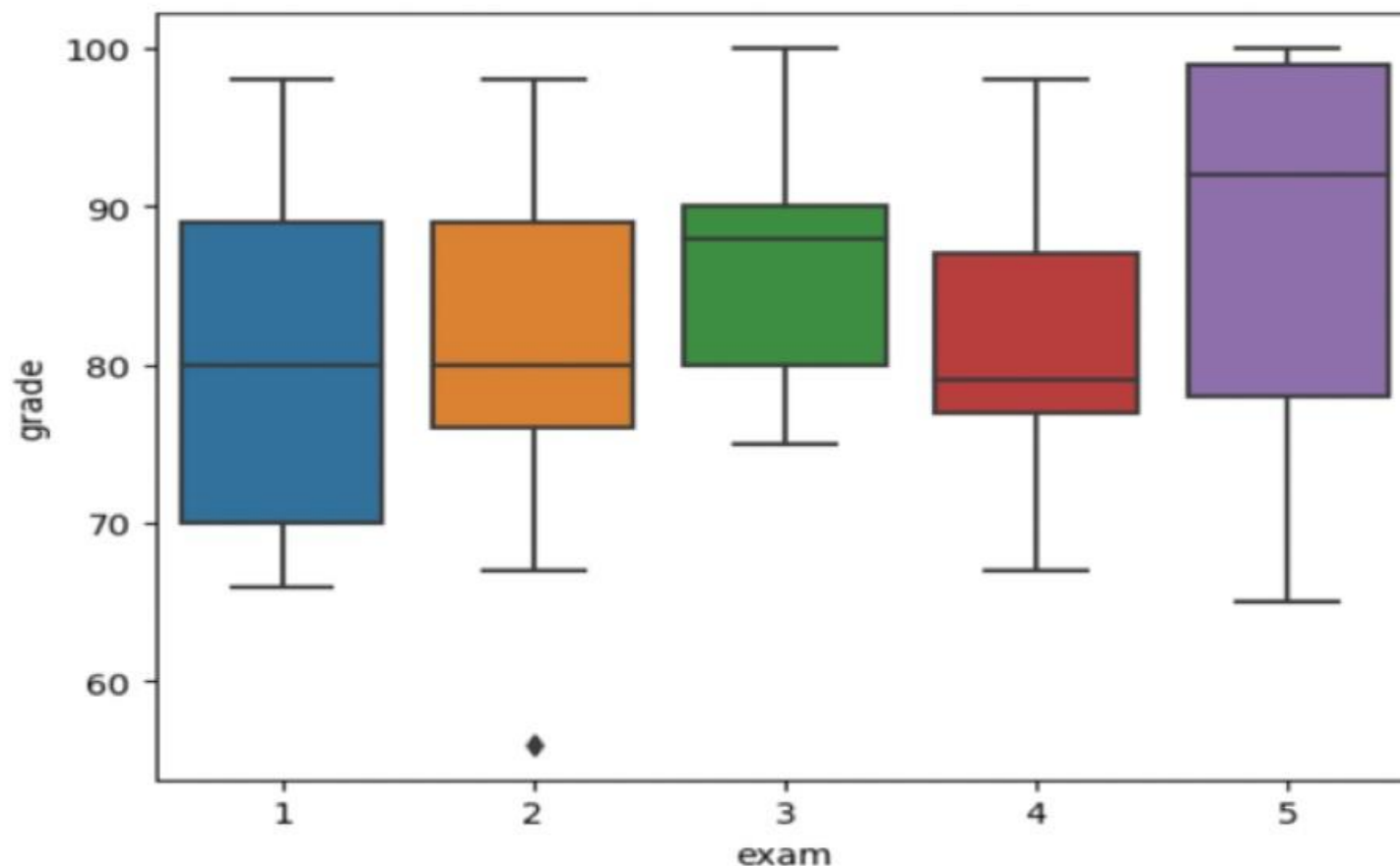
- Quartiles, outliers and boxplots
 - **Quartiles:** Q_1 (25th percentile), Q_3 (75th percentile)
 - **Inter-quartile range:** $IQR = Q_3 - Q_1$
 - **Five number summary:** min, Q_1 , median, Q_3 , max
 - **Boxplot:** ends of the box are the quartiles; median is marked; add whiskers, and plot outliers individually
 - **Outlier:** usually, a value higher/lower than $1.5 \times IQR$
- Variance and standard deviation (*sample: s , population: σ*)
 - **Variance:** (algebraic, scalable computation)

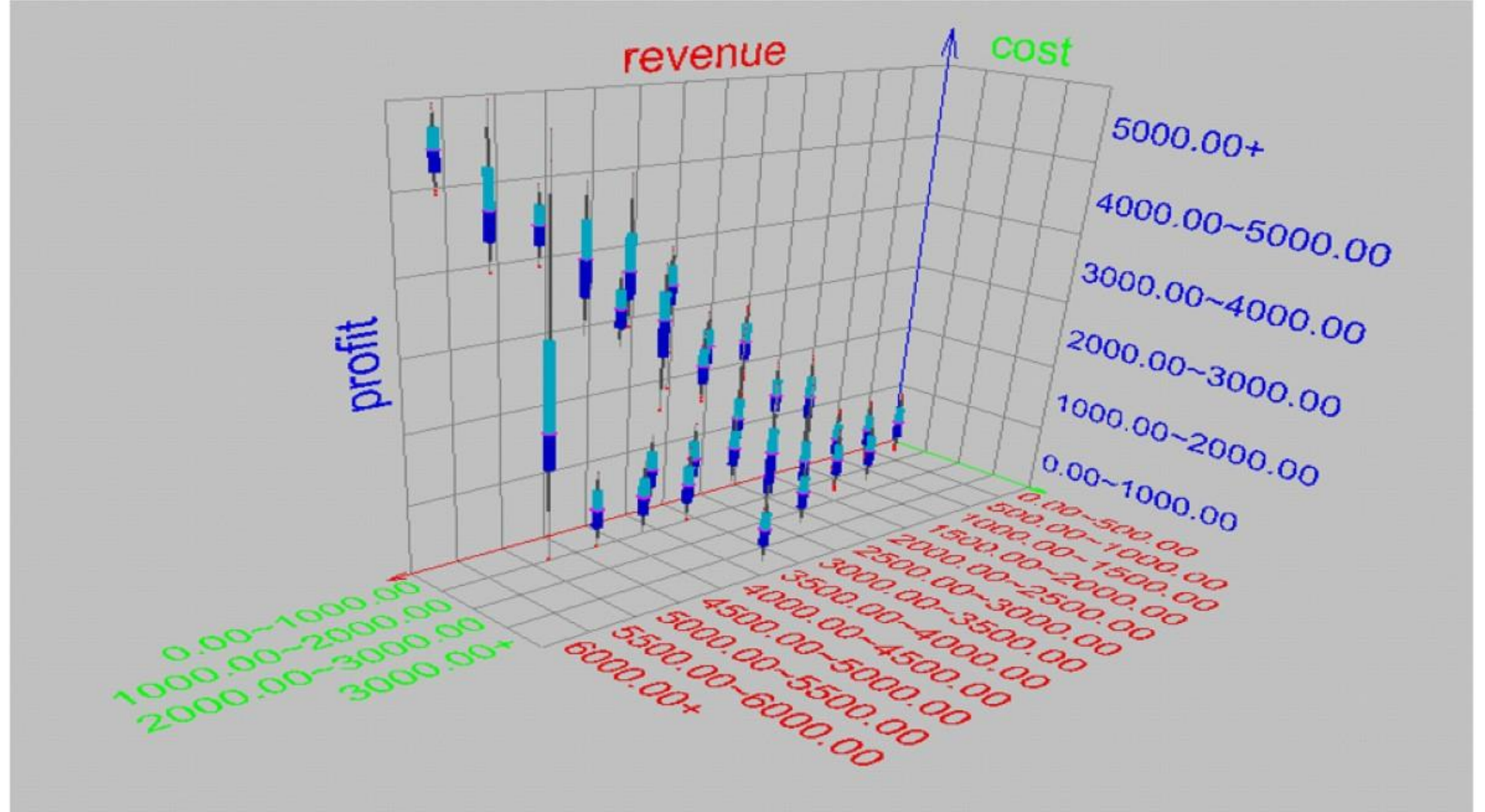
$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right] \quad \sigma^2 = \frac{1}{N} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{N} \sum_{i=1}^n x_i^2 - \mu^2$$
 - **Standard deviation s (or σ)** is the square root of variance s^2 (or σ^2)



- **Five-number summary** of a distribution
 - Minimum, Q1, Median, Q3, Maximum
- **Boxplot**
 - Data is represented with a box
 - The ends of the box are at the first and third quartiles, i.e., the height of the box is IQR
 - The median is marked by a line within the box
 - Whiskers: two lines outside the box extended to Minimum and Maximum
 - Outliers: points beyond a specified outlier threshold, plotted individually

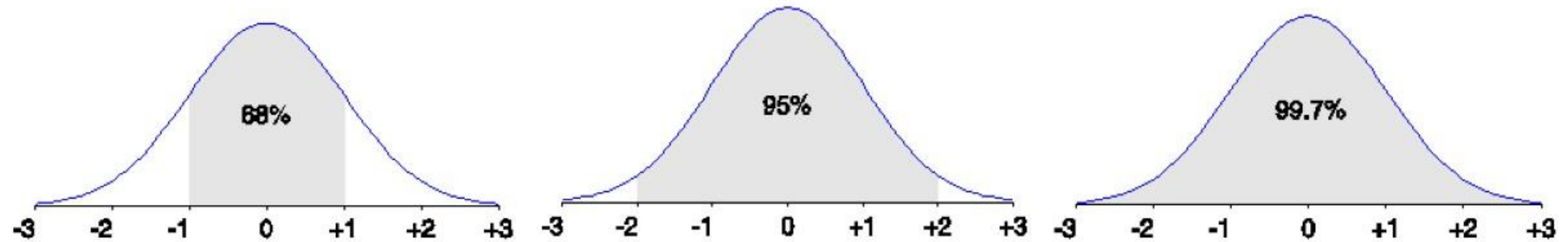








- The normal (distribution) curve
 - From $\mu - \sigma$ to $\mu + \sigma$: contains about 68% of the measurements (μ : mean, σ : standard deviation)
 - From $\mu - 2\sigma$ to $\mu + 2\sigma$: contains about 95% of it
 - From $\mu - 3\sigma$ to $\mu + 3\sigma$: contains about 99.7% of it

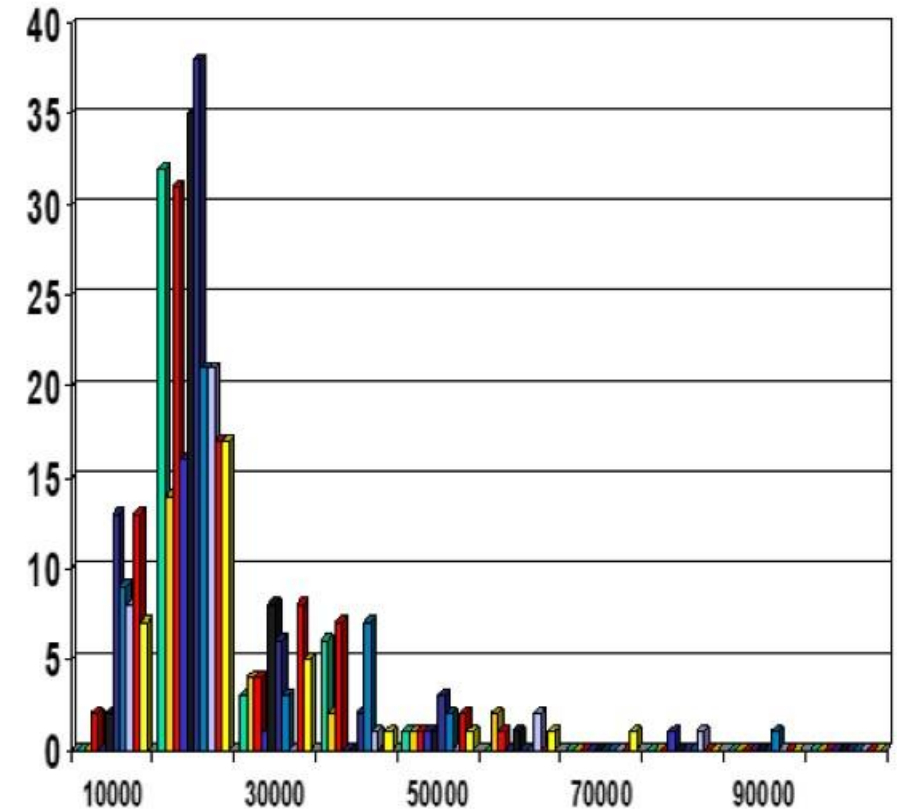




- **Boxplot:** graphic display of five-number summary
- **Histogram:** x-axis are values, y-axis represent frequencies
- **Quantile plot:** each value x_i is paired with f_i indicating that approximately $100 f_i \%$ of data are $\leq x_i$
- **Quantile-quantile (q-q) plot:** graphs the quantiles of one univariate distribution against the corresponding quantiles of another
- **Scatter plot:** each pair of values is a pair of coordinates and plotted as points in the plane

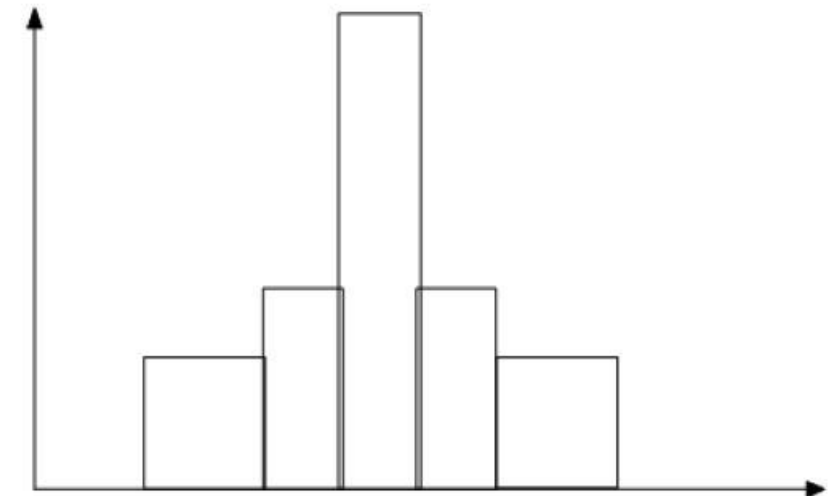
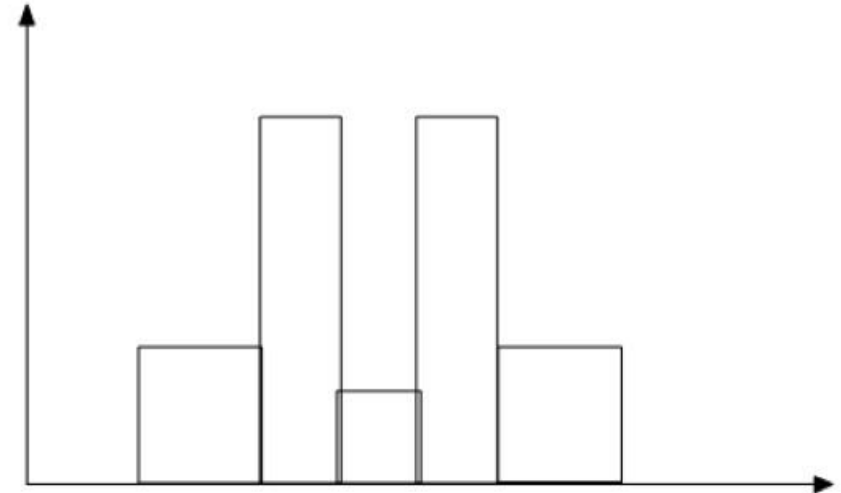


- **Histogram:** Graph display of tabulated frequencies, shown as bars
- It shows what proportion of cases fall into each of several categories
- Differs from a bar chart in that it is the *area* of the bar that denotes the value, not the height as in bar charts, a crucial distinction when the categories are not of uniform width
- The categories are usually specified as non-overlapping intervals of some variable. The categories (bars) must be adjacent.



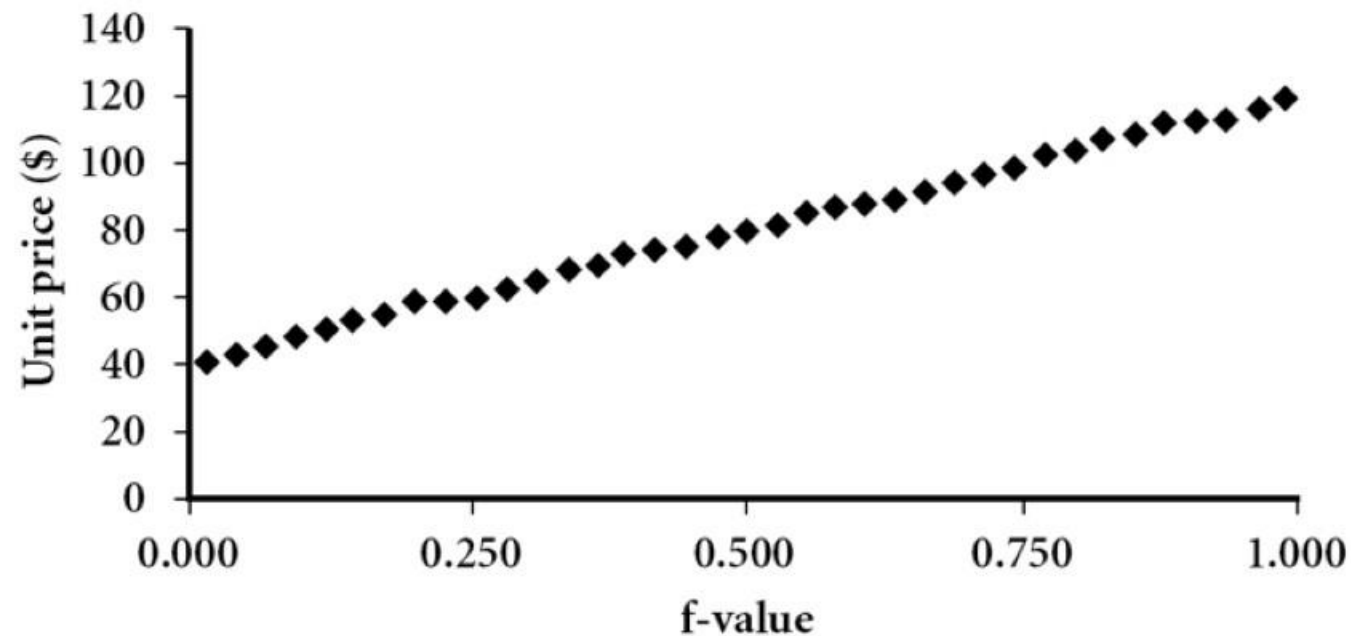


- The two histograms shown in the left may have the same boxplot representation
 - The same values for: min, Q1, median, Q3, max
- But they have rather different data distributions





- Displays all of the data (allowing the user to assess both the overall behavior and unusual occurrences)
- Plots **quantile** information
 - For a data x_i data sorted in increasing order, f_i indicates that approximately $100 f_i\%$ of the data are below or equal to the value x_i

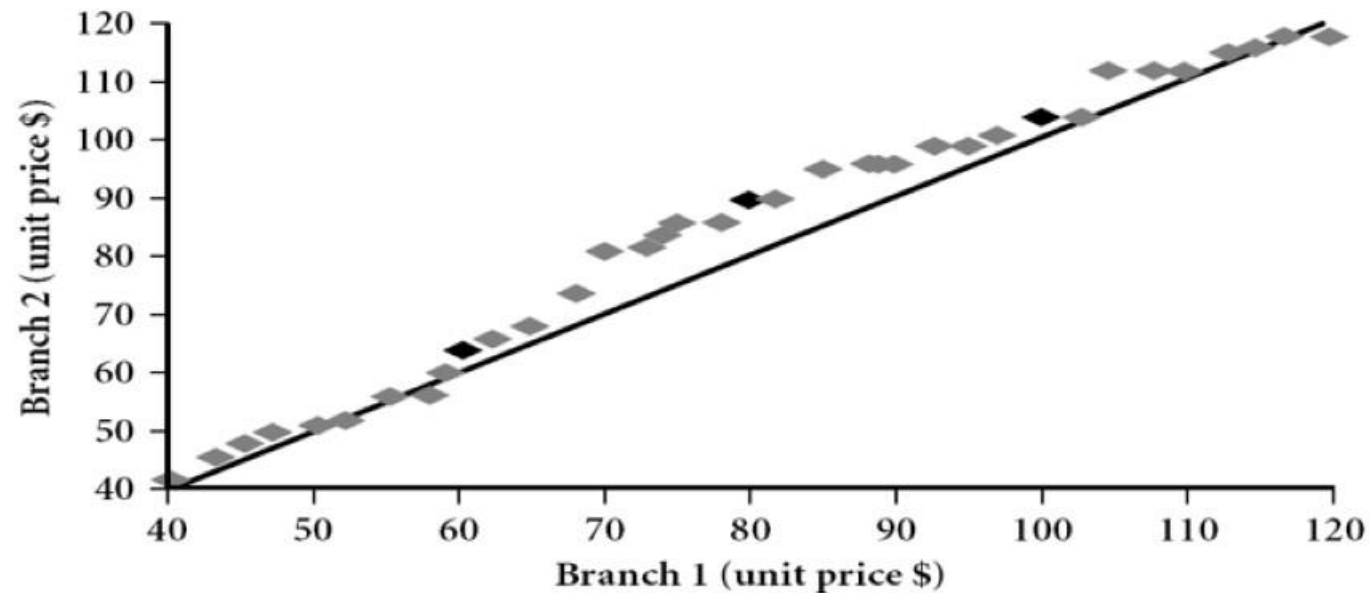




Quantile-Quantile (Q-Q) Plot

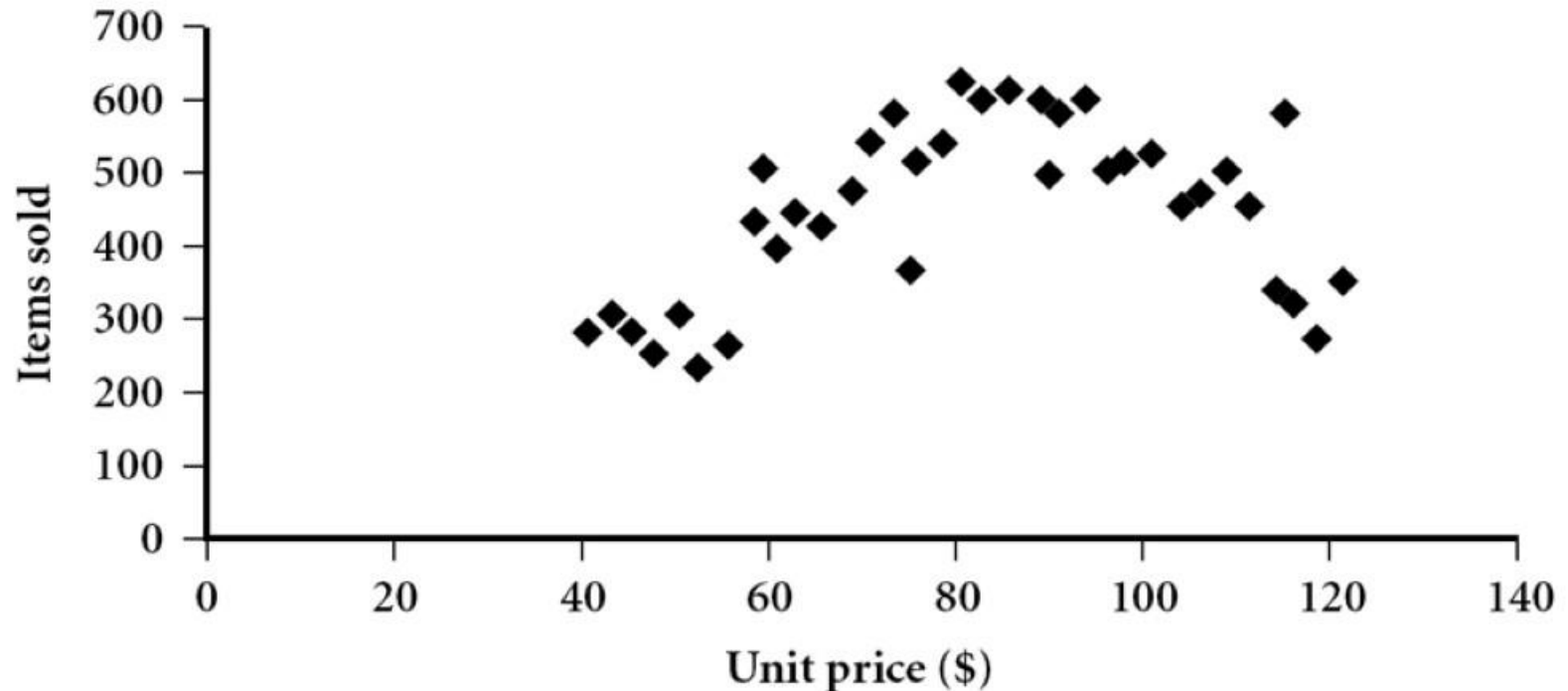


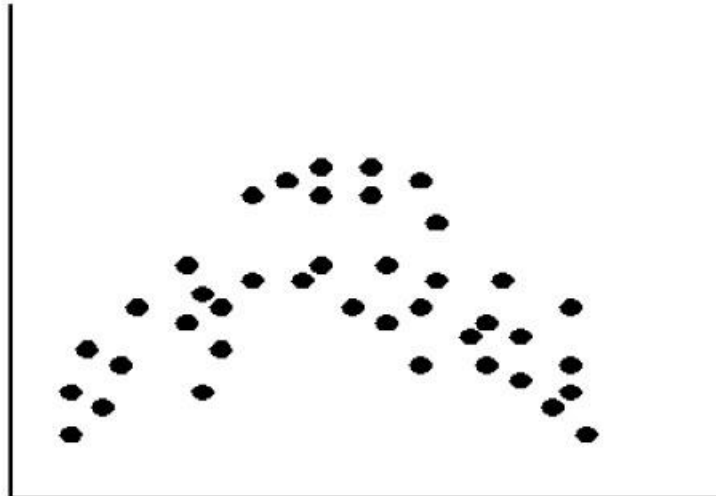
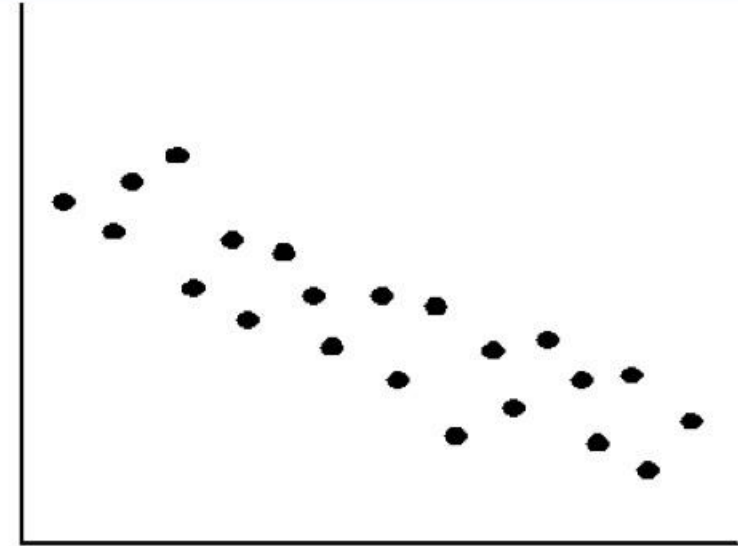
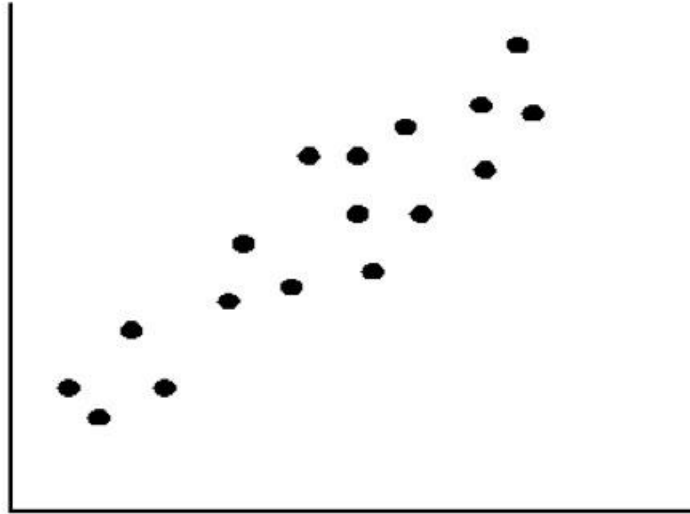
- Graphs the quantiles of one univariate distribution against the corresponding quantiles of another
- View: Is there is a shift in going from one distribution to another?
- Example shows unit price of items sold at Branch 1 vs. Branch 2 for each quantile. Unit prices of items sold at Branch 1 tend to be lower than those at Branch 2





- Provides a first look at bivariate data to see clusters of points, outliers, etc
- Each pair of values is treated as a pair of coordinates and plotted as points in the plane





- The left half fragment is positively correlated
- The right half is negative correlated





Knowing Your Data (Data Visualization)

- Data Objects and Attribute Types
- Basic Statistical Descriptions of Data
- **Data Visualization** 
- Measuring Data Similarity and Dissimilarity
- Summary



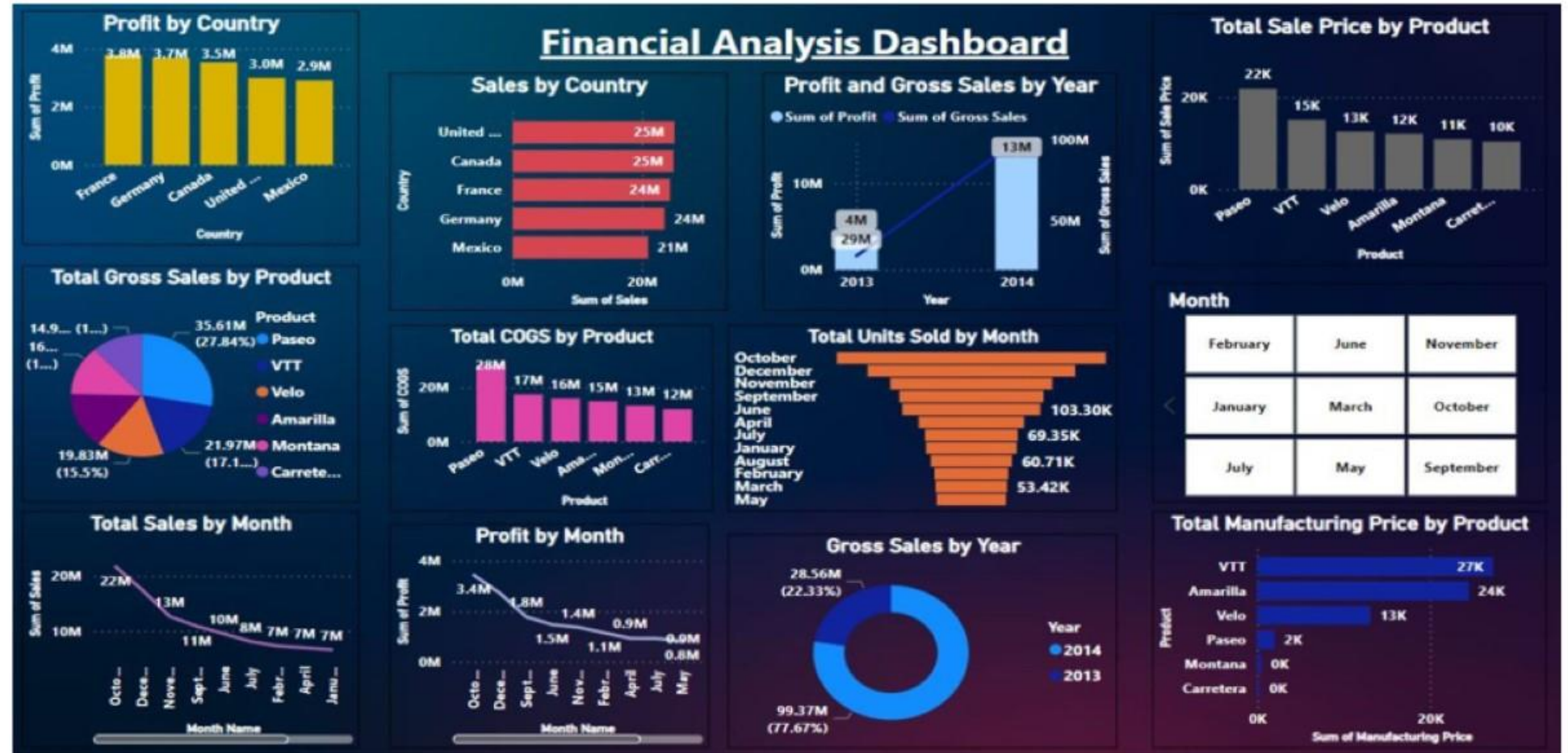


- **Why data visualization?**

- Gain insight into an information space by mapping data onto graphical primitives
- Provide qualitative overview of large data sets
- Search for patterns, trends, structure, irregularities, relationships among data
- Help find interesting regions and suitable parameters for further quantitative analysis
- Provide a visual proof of computer representations derived



- Dashboard





Knowing Your Data (Data Visualization)

- Data Objects and Attribute Types
- Basic Statistical Descriptions of Data
- Data Visualization
- **Measuring Data Similarity and Dissimilarity**
- Summary





- **Similarity**
 - Numerical measure of how alike two data objects are
 - Value is higher when objects are more alike
 - Often falls in the range $[0,1]$
- **Dissimilarity** (e.g., distance)
 - Numerical measure of how different two data objects are
 - Lower when objects are more alike
 - Minimum dissimilarity is often 0
 - Upper limit varies
- **Proximity** refers to a similarity or dissimilarity



- Data matrix

- n data points with p dimensions
- Two modes

$$\begin{bmatrix} x_{11} & \dots & x_{1f} & \dots & x_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ x_{i1} & \dots & x_{if} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{nf} & \dots & x_{np} \end{bmatrix}$$

- Dissimilarity matrix

- n data points, but registers only the distance
- A triangular matrix
- Single mode

$$\begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & & \\ d(n,1) & d(n,2) & \dots & \dots & 0 \end{bmatrix}$$



- Can take 2 or more states, e.g., red, yellow, blue, green (generalization of a binary attribute)

- Method 1: Simple matching

- m : # of matches, p : total # of variables

$$d(i, j) = \frac{p - m}{p}$$

- Method 2: Use a large number of binary attributes

- creating a new binary attribute for each of the M nominal states



- A contingency table for binary data



		Object j		sum
		1	0	
Object i	1	q	r	$q + r$
	0	s	t	$s + t$
sum		$q + s$	$r + t$	p

- Distance measure for symmetric binary variables:

$$d(i, j) = \frac{r + s}{q + r + s + t}$$





- A Distance measure for asymmetric binary variables:

$$d(i, j) = \frac{r + s}{q + r + s}$$

- Jaccard coefficient (*similarity* measure for *asymmetric* binary variables):

$$\text{sim}_{\text{Jaccard}}(i, j) = \frac{q}{q + r + s}$$

- Note: Jaccard coefficient is the same as “coherence”:

$$\text{coherence}(i, j) = \frac{\text{sup}(i, j)}{\text{sup}(i) + \text{sup}(j) - \text{sup}(i, j)} = \frac{q}{(q + r) + (q + s) - q}$$





• Example

Name	Gender	Fever	Cough	Test-1	Test-2	Test-3	Test-4
Jack	M	Y	N	P	N	N	N
Mary	F	Y	N	P	N	P	N
Jim	M	Y	P	N	N	N	N

- Gender is a symmetric attribute
- The remaining attributes are asymmetric binary
- Let the values Y and P be 1, and the value N 0

$$d(\text{jack}, \text{mary}) = \frac{0 + 1}{2 + 0 + 1} = 0.33$$

$$d(\text{jack}, \text{jim}) = \frac{1 + 1}{1 + 1 + 1} = 0.67$$

$$d(\text{jim}, \text{mary}) = \frac{1 + 2}{1 + 1 + 2} = 0.75$$





- Z-score: $z = \frac{x - \mu}{\sigma}$
 - X: raw score to be standardized, μ : mean of the population, σ : standard deviation
 - the distance between the raw score and the population mean in units of the standard deviation
 - negative when the raw score is below the mean, "+" when above
- An alternative way: Calculate the mean absolute deviation

$$s_f = \frac{1}{n} (|x_{1f} - m_f| + |x_{2f} - m_f| + \dots + |x_{nf} - m_f|)$$

where

$$m_f = \frac{1}{n} (x_{1f} + x_{2f} + \dots + x_{nf}).$$

- standardized measure (z-score): $z_{if} = \frac{x_{if} - m_f}{s_f}$
- Using mean absolute deviation is more robust than using standard deviation

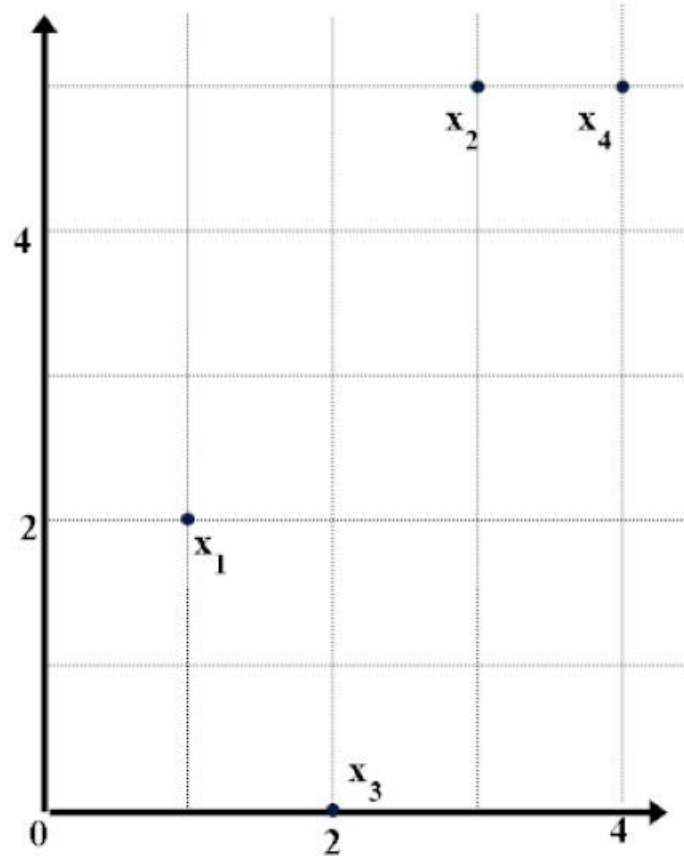


- **Z-score Example:**
- Suppose that the mean and standard deviation of the values for the attribute income (income) are \$54,000 and \$16,000, respectively. **What is z-score of income \$73,600.**
- **Solution:** with z-score normalization, a value of \$73,600 for income is transformed to

$$\frac{73,600 - 54,000}{16,000} = 1.225$$



Example: Data Matrix and Dissimilarity Matrix

**Data Matrix**

point	attribute1	attribute2
$x1$	1	2
$x2$	3	5
$x3$	2	0
$x4$	4	5

Dissimilarity Matrix
(with **Euclidean Distance**)

	$x1$	$x2$	$x3$	$x4$
$x1$	0			
$x2$	3.61	0		
$x3$	5.1	5.1	0	
$x4$	4.24	1	5.39	0



- *Minkowski distance*: A popular distance measure

$$d(i, j) = \sqrt[h]{|x_{i1} - x_{j1}|^h + |x_{i2} - x_{j2}|^h + \dots + |x_{ip} - x_{jp}|^h}$$

where $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ and $j = (x_{j1}, x_{j2}, \dots, x_{jp})$ are two p -dimensional data objects, and h is the order (the distance so defined is also called L- h norm)

- Properties
 - $d(i, j) > 0$ if $i \neq j$, and $d(i, i) = 0$ (Positive definiteness)
 - $d(i, j) = d(j, i)$ (Symmetry)
 - $d(i, j) \leq d(i, k) + d(k, j)$ (Triangle Inequality)
- A distance that satisfies these properties is a **metric**



- $h = 1$: **Manhattan** (city block, L_1 norm) distance
 - E.g., the Hamming distance: the number of bits that are different between two binary vectors

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}|$$

- $h = 2$: (L_2 norm) **Euclidean** distance

$$d(i, j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2)}$$

- $h \rightarrow \infty$. **“supremum”** ($L_{\max}$ norm, L_{∞} norm) distance.
 - This is the maximum difference between any component (attribute) of the vectors

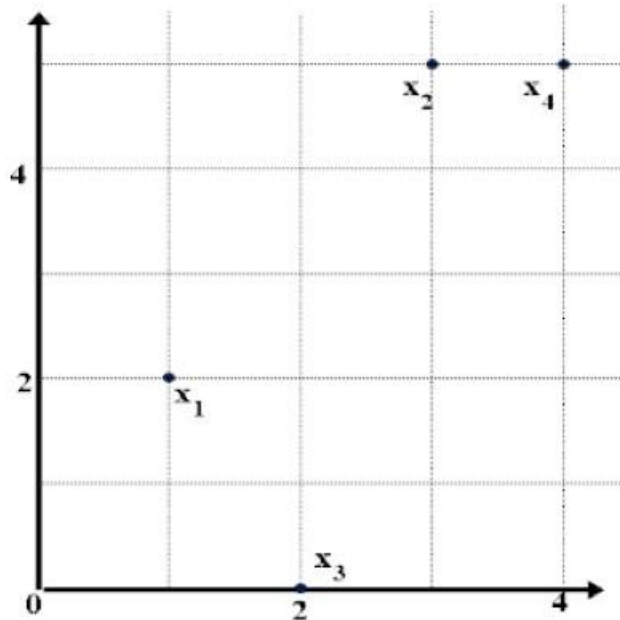
$$d(i, j) = \lim_{h \rightarrow \infty} \left(\sum_{f=1}^p |x_{if} - x_{jf}|^h \right)^{\frac{1}{h}} = \max_f |x_{if} - x_{jf}|$$



Example: Minkowski Distance



point	attribute 1	attribute 2
x1	1	2
x2	3	5
x3	2	0
x4	4	5



Dissimilarity Matrices

Manhattan (L_1)

L	x1	x2	x3	x4
x1	0			
x2	5	0		
x3	3	6	0	
x4	6	1	7	0

Euclidean (L_2)

L2	x1	x2	x3	x4
x1	0			
x2	3.61	0		
x3	2.24	5.1	0	
x4	4.24	1	5.39	0

Supremum

L_∞	x1	x2	x3	x4
x1	0			
x2	3	0		
x3	2	5	0	
x4	3	1	5	0



- An ordinal variable can be discrete or continuous
- Order is important, e.g., rank
- Can be treated like interval-scaled
 - replace x_{if} by their rank $r_{if} \in \{1, \dots, M_f\}$
 - map the range of each variable onto $[0, 1]$ by replacing i -th object in the f -th variable by

$$z_{if} = \frac{r_{if} - 1}{M_f - 1}$$

- compute the dissimilarity using methods for interval-scaled variables



- A database may contain all attribute types
 - Nominal, symmetric binary, asymmetric binary, numeric, ordinal
- One may use a weighted formula to combine their effects

$$d(i, j) = \frac{\sum_{f=1}^p \delta_{ij}^{(f)} d_{ij}^{(f)}}{\sum_{f=1}^p \delta_{ij}^{(f)}}$$

- f is binary or nominal:
 $d_{ij}^{(f)} = 0$ if $x_{if} = x_{jf}$, or $d_{ij}^{(f)} = 1$ otherwise
- f is numeric: use the normalized distance
- f is ordinal
 - Compute ranks r_{if} and
 - Treat z_{if} as interval-scaled

$$z_{if} = \frac{r_{if} - 1}{M_f - 1}$$



- A **document** can be represented by thousands of attributes, each recording the *frequency* of a particular word (such as keywords) or phrase in the document.

Document	team	coach	hockey	baseball	soccer	penalty	score	win	loss	season
Document1	5	0	3	0	2	0	0	2	0	0
Document2	3	0	2	0	1	1	0	1	0	1
Document3	0	7	0	2	1	0	0	3	0	0
Document4	0	1	0	0	1	2	2	0	3	0

- Other vector objects: gene features in micro-arrays, ...
- Applications: information retrieval, biologic taxonomy, gene feature mapping, ...
- Cosine measure: If d_1 and d_2 are two vectors (e.g., term-frequency vectors), then

$$\cos(d_1, d_2) = (d_1 \bullet d_2) / (||d_1|| \cdot ||d_2||),$$

where $\bullet$ indicates vector dot product, $||d||$: the length of vector d



Example: Cosine Similarity



- $\cos(d_1, d_2) = (d_1 \bullet d_2) / (||d_1|| ||d_2||)$,
where $\bullet$ indicates vector dot product, $||d||$: the length of vector d
- Ex: Find the **similarity** between documents 1 and 2.

$$d_1 = (5, 0, 3, 0, 2, 0, 0, 2, 0, 0)$$

$$d_2 = (3, 0, 2, 0, 1, 1, 0, 1, 0, 1)$$

$$d_1 \bullet d_2 = 5*3 + 0*0 + 3*2 + 0*0 + 2*1 + 0*1 + 0*1 + 2*1 + 0*0 + 0*1 = 25$$

$$||d_1|| = (5*5 + 0*0 + 3*3 + 0*0 + 2*2 + 0*0 + 0*0 + 2*2 + 0*0 + 0*0)^{0.5} = (42)^{0.5} = 6.481$$

$$||d_2|| = (3*3 + 0*0 + 2*2 + 0*0 + 1*1 + 1*1 + 0*0 + 1*1 + 0*0 + 1*1)^{0.5} = (17)^{0.5} = 4.12$$

$$\cos(d_1, d_2) = 0.94$$



Knowing Your Data (Data Visualization)

- Data Objects and Attribute Types
- Basic Statistical Descriptions of Data
- Data Visualization
- Measuring Data Similarity and Dissimilarity
- **Summary** 





- Data attribute types: nominal, binary, ordinal, interval-scaled, ratio-scaled
- Many types of data sets, e.g., numerical, text, graph, Web, image.
- Gain insight into the data by:
 - Basic statistical data description: central tendency, dispersion, graphical displays
 - Data visualization: map data onto graphical primitives
 - Measure data similarity
- Many methods have been developed but still an active area of research.



Thank You

Data Visualization - Practice

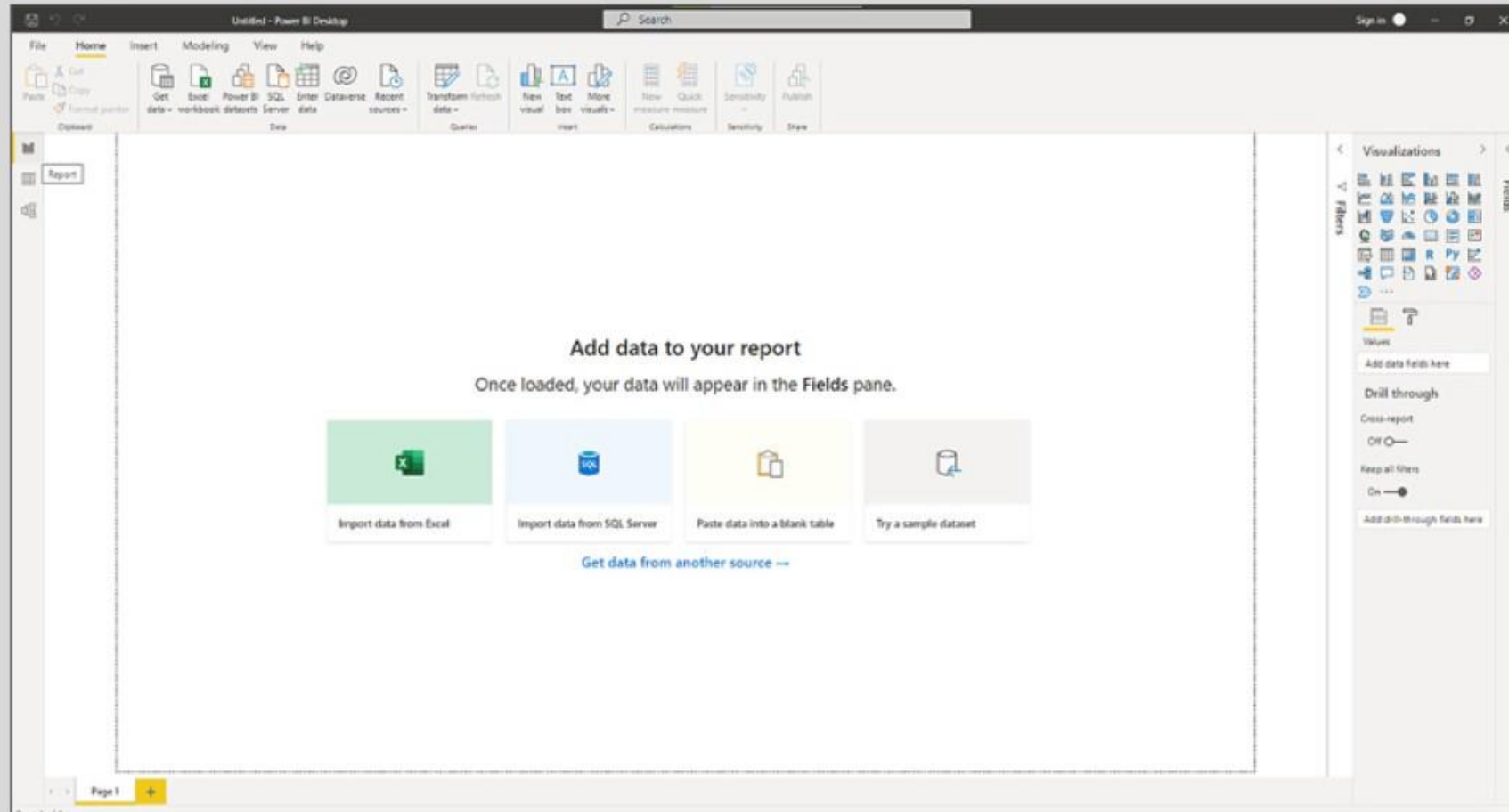


Power BI



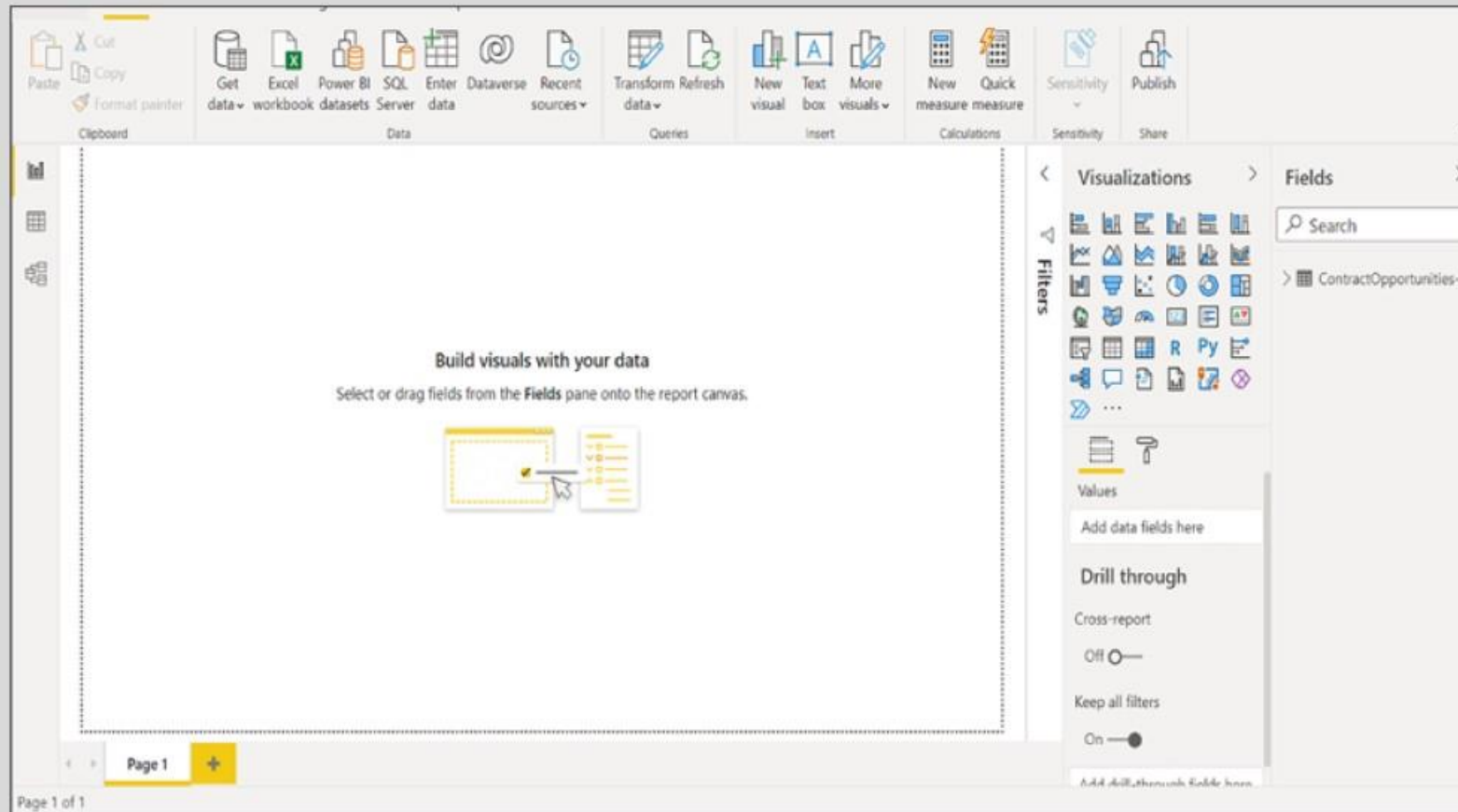
Data Visualization

- Creating visualizations:



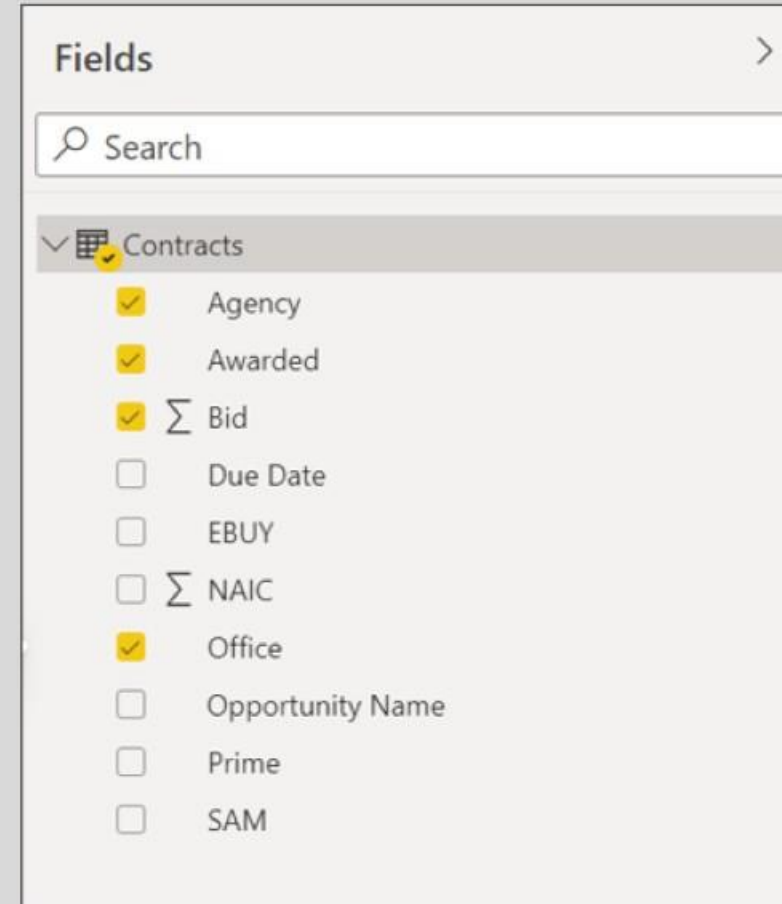
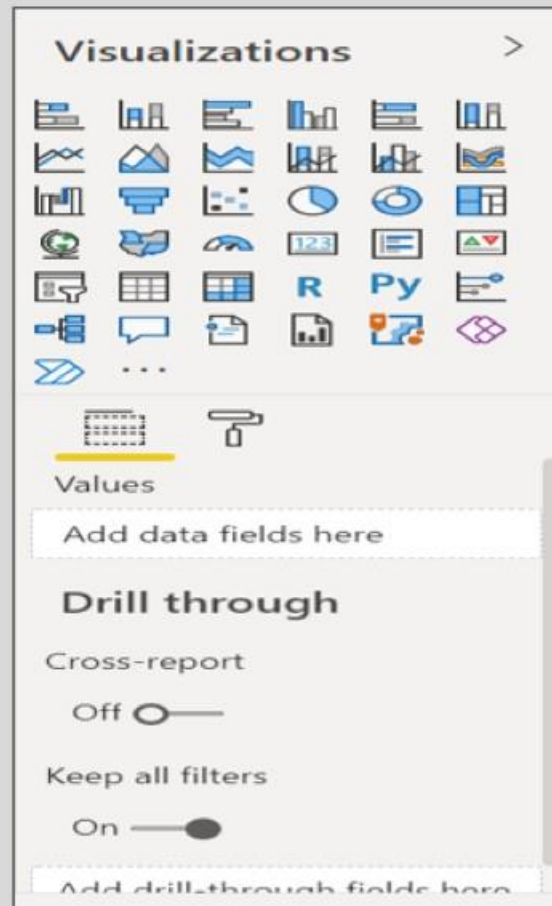
Data Visualization

- Report View:



Data Visualization

◦ Choosing a Visualization:



Data Visualization

◦ Filtering Data:

Filters

Search

Filters on this visual

Awarded is (All)

Filter type ⓘ

Basic filtering

Search

☐ Select all

☐ Awarded 2

☐ In-Progress 3

☐ No 7

☐ Pending 3

Filters

Search

Filters on this visual

Awarded is (All)

Filter type ⓘ

Basic filtering

Search

☒ Select all

☒ Awarded 2

☒ In-Progress 3

☒ No 7

☒ Pending 3

☐ Require single selection

Bid is greater than 100000 and is less than or equal to...

Show items when the value

is greater than

100000

☒ And ☐ Or

is less than or equal to

500000

Apply filter

Add data fields here

Filters on this page

Add data fields here

Visualizations

Search

Legend

Awarded

Details

Add data fields here

Values

Bid

Tooltips

Add data fields here

Drill through

Cross-report

Off

Keep all filters

On

Add drill-through fields here

Fields

Search

Awards

☐ Agency

☒ Awarded

☒ Bid

☐ EBUY

☐ NAIC

☐ Office

☐ Opportunity Name

☐ Prime

☐ SAM

Data Visualization

◦ Filtering Data:

Filters

Search

Basic filtering

Search

☒

Select all

☒

Awarded2

☒

In-Progress3

☒

No7

☒

Pending3

☐

Require single selection

Bid

is greater than 100000 and is less than or equal to...

Show items when the value

is greater than

100000

☒ And ☐ Or

is less than or equal to

500000

Apply filter

Add data fields here

Filters on this page

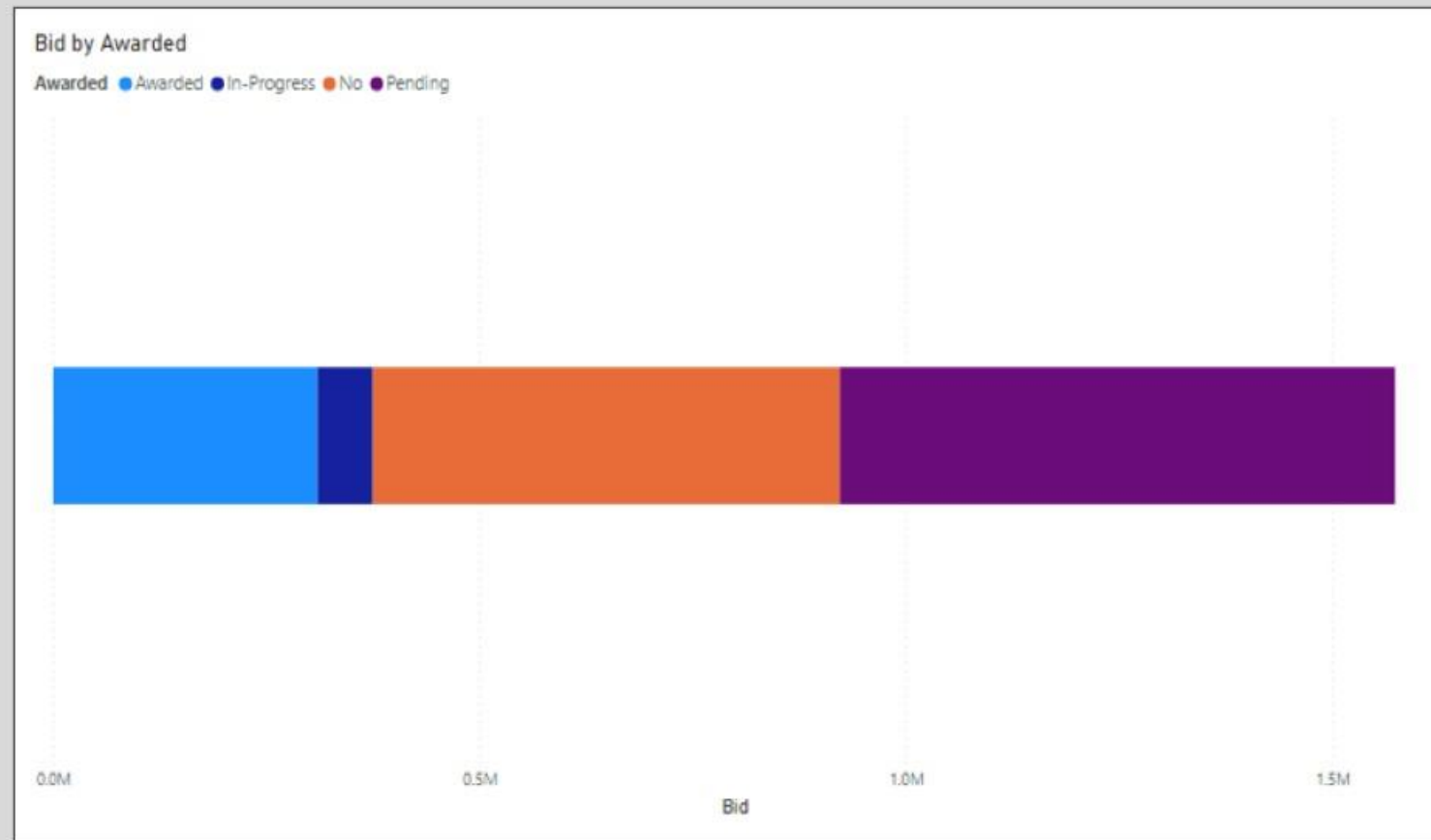
Add data fields here

Filters on all pages

Add data fields here

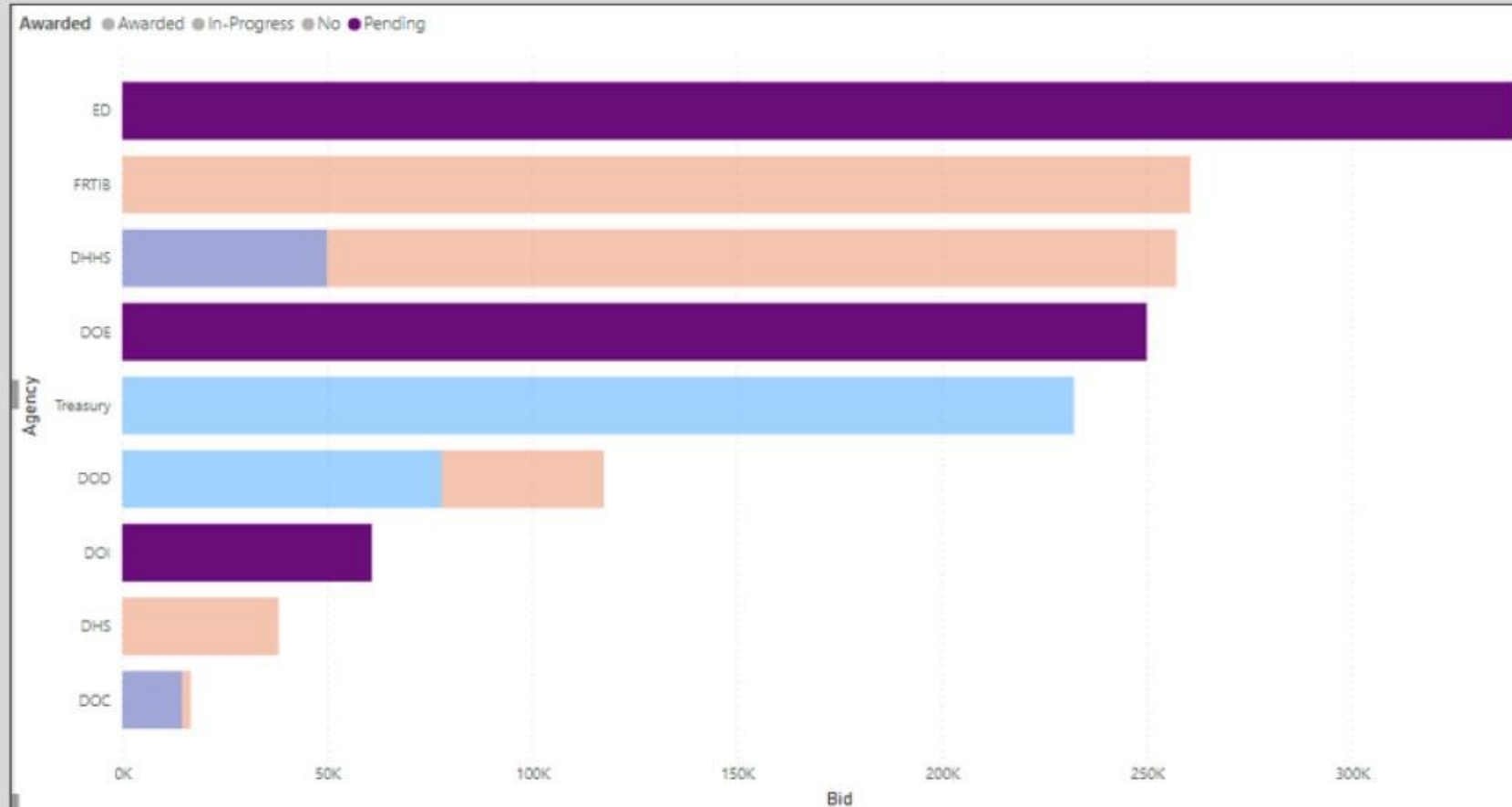
Data Visualization

- Stacked Bar chart:



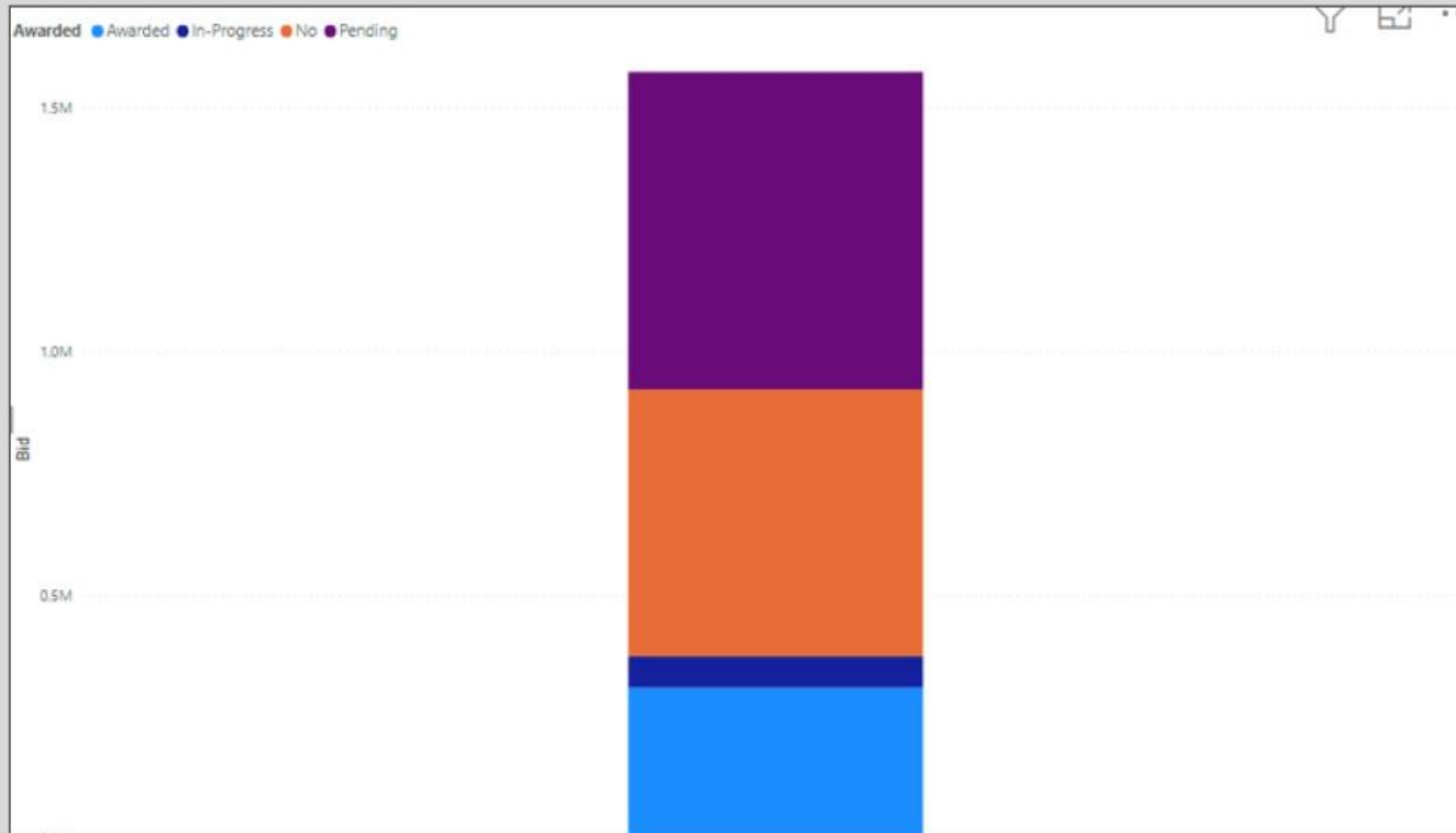
Data Visualization

- Stacked Bar chart (Multiple Dimensions):



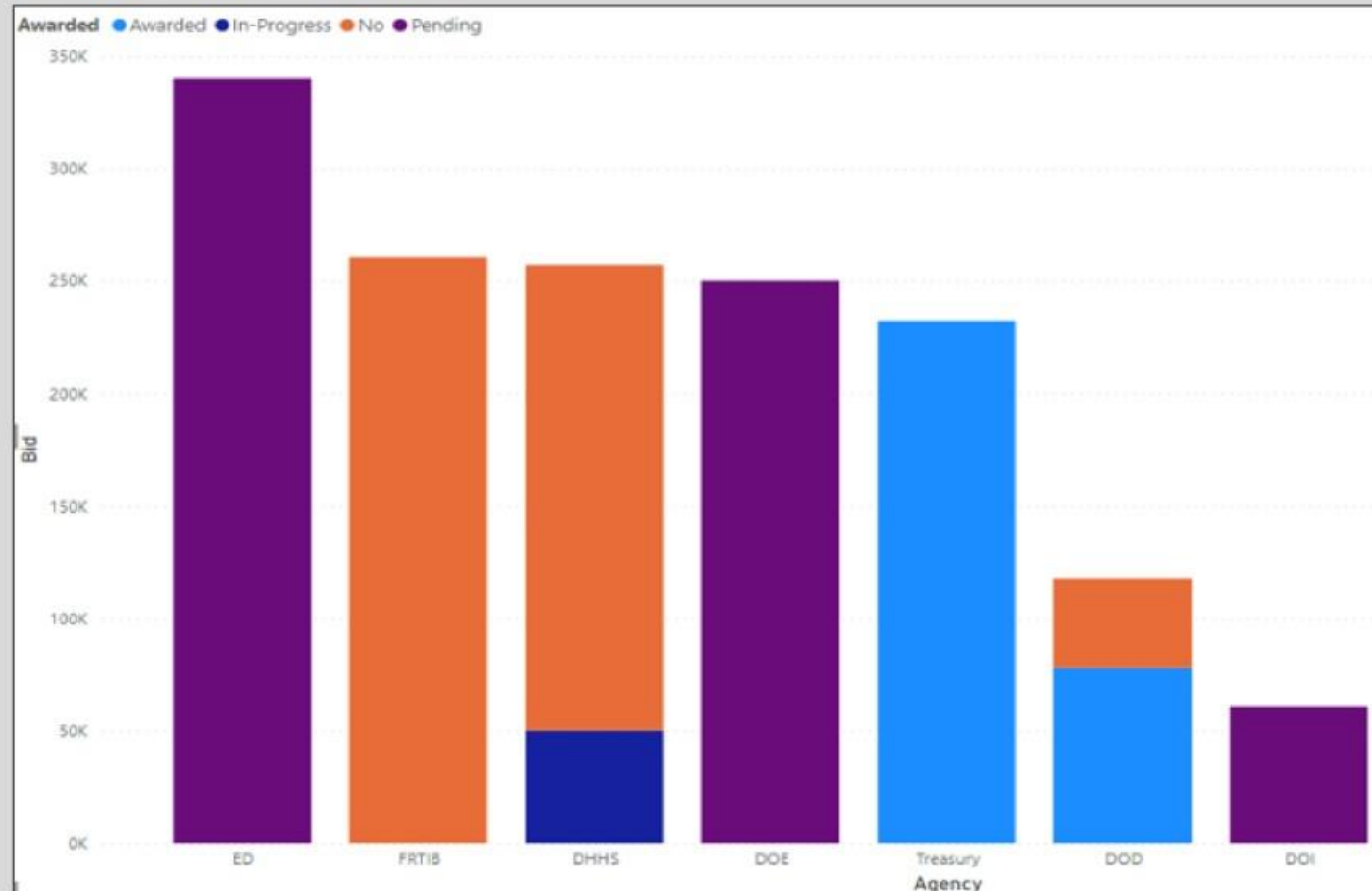
Data Visualization

- Stacked Column chart:



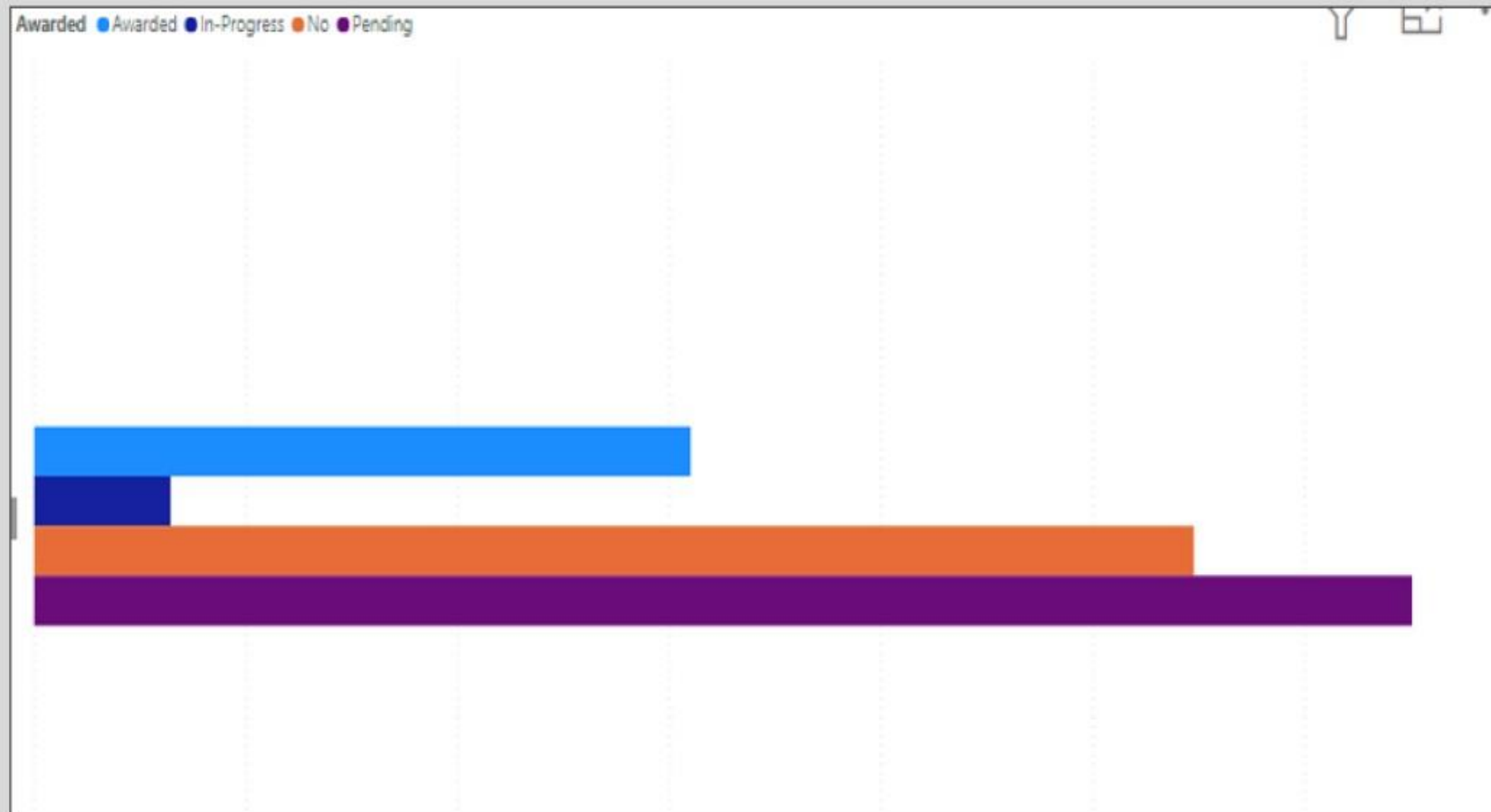
Data Visualization

- Stacked Column chart (Multiple Dimensions):



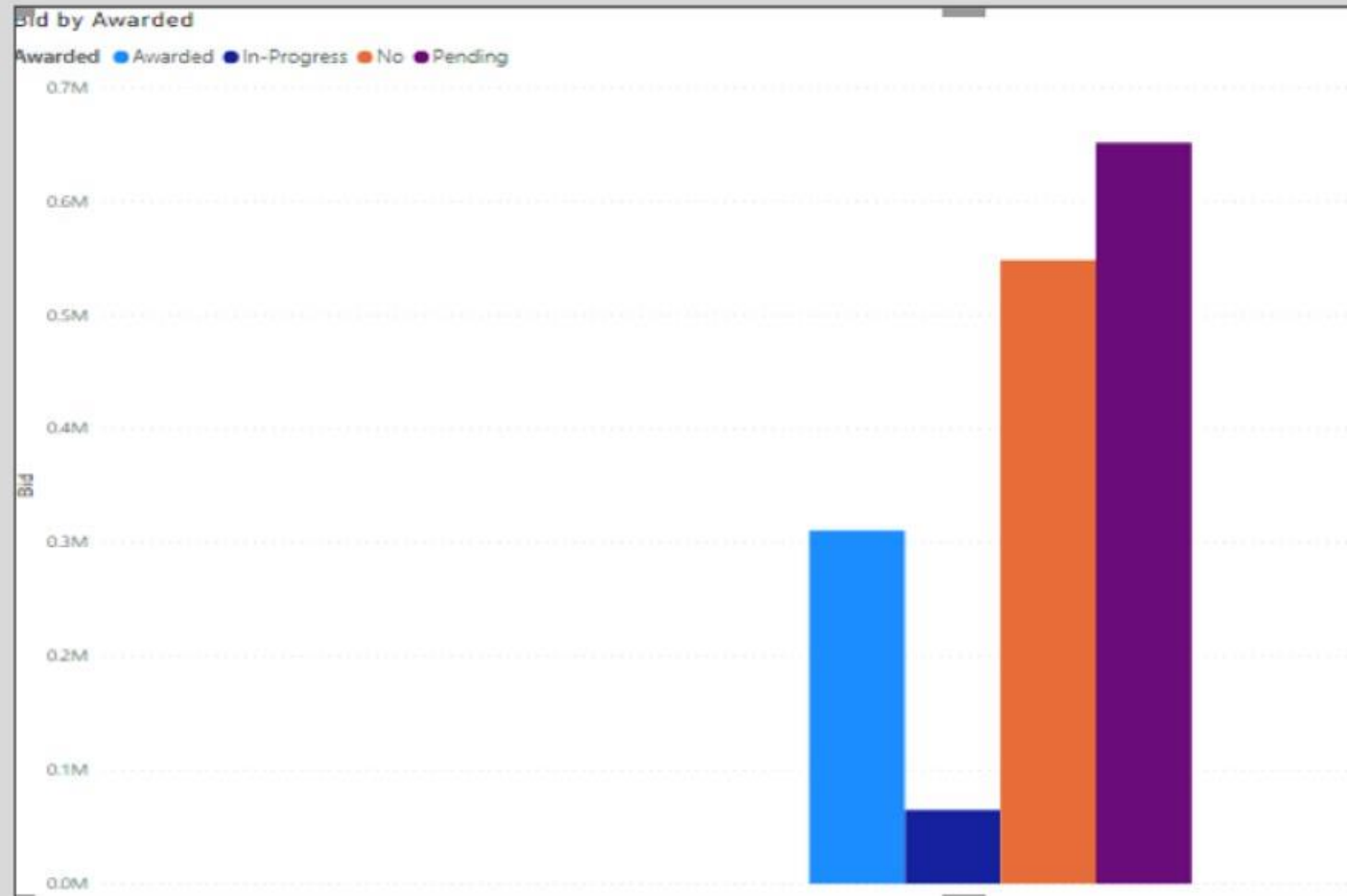
Data Visualization

- Clustered Bar chart:



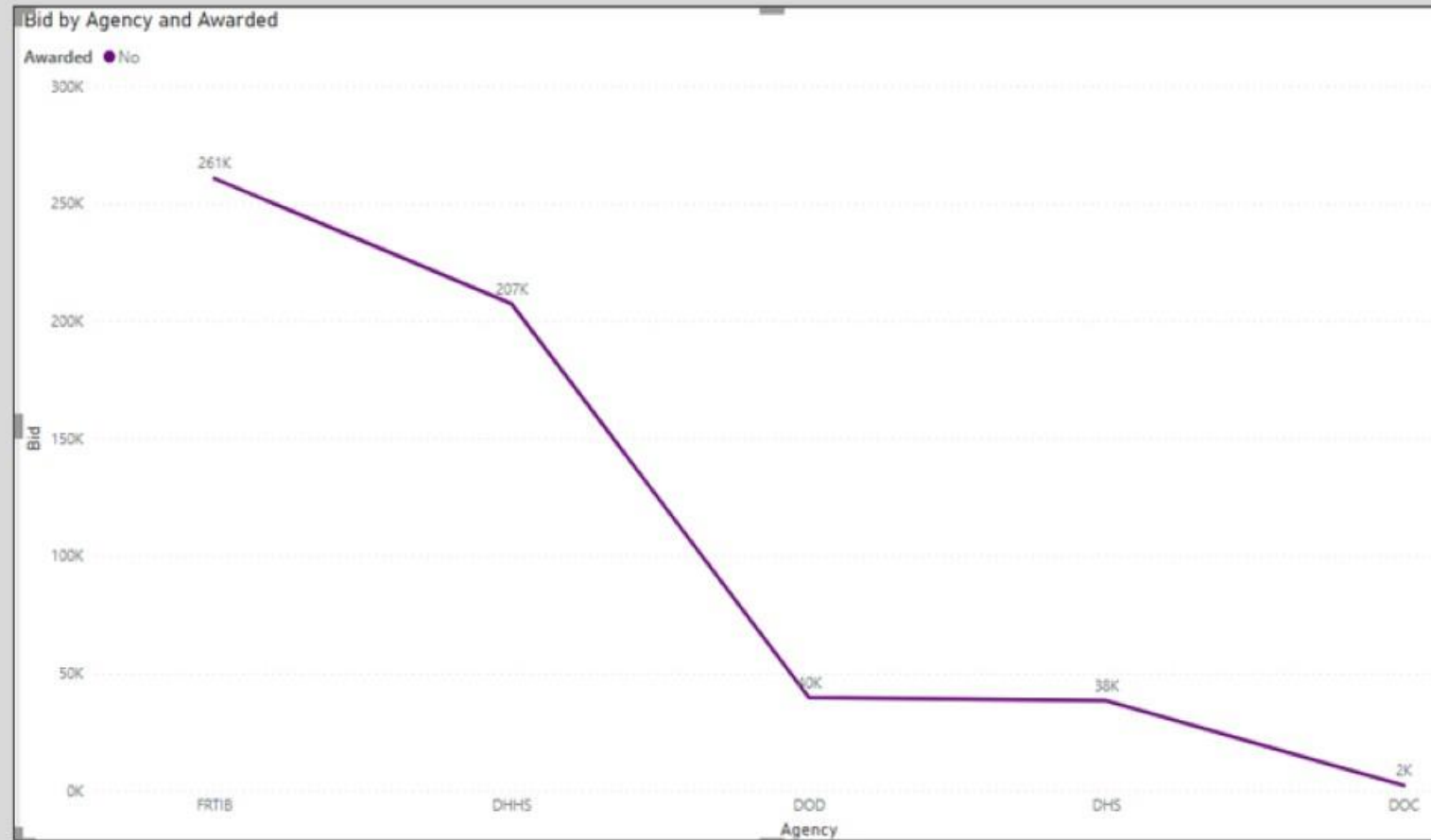
Data Visualization

- Clustered Column chart:



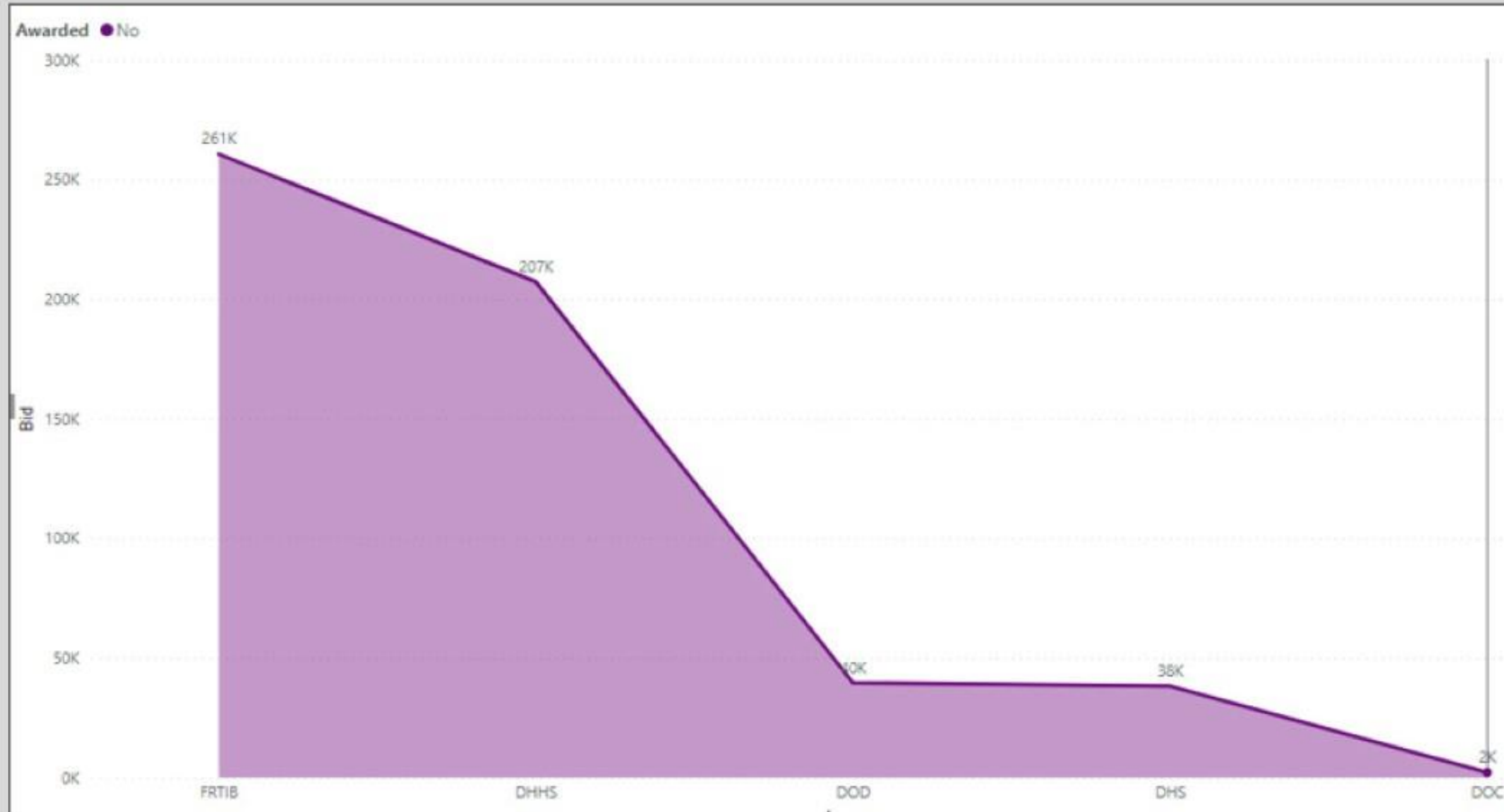
Data Visualization

- Line chart:



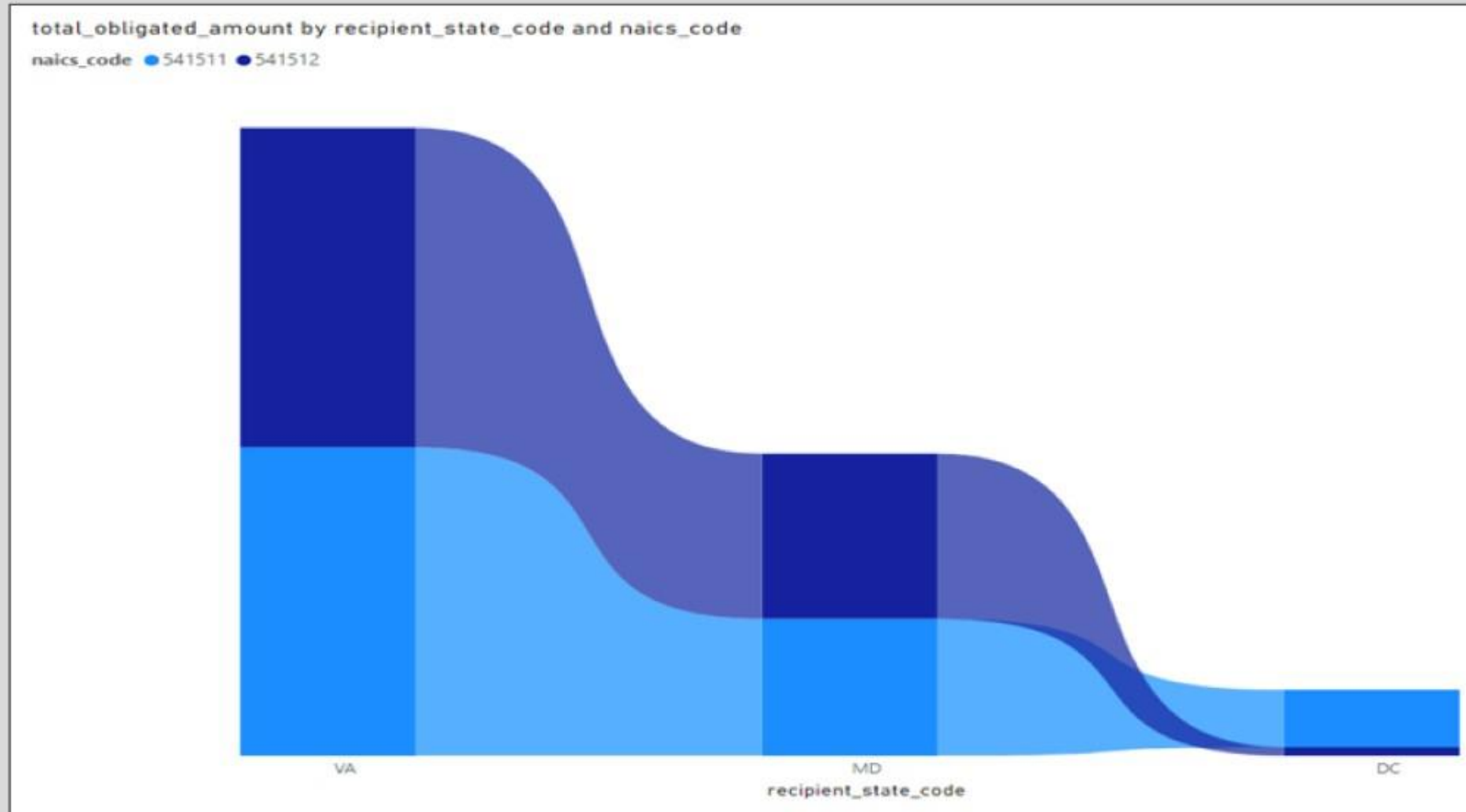
Data Visualization

- Area chart:



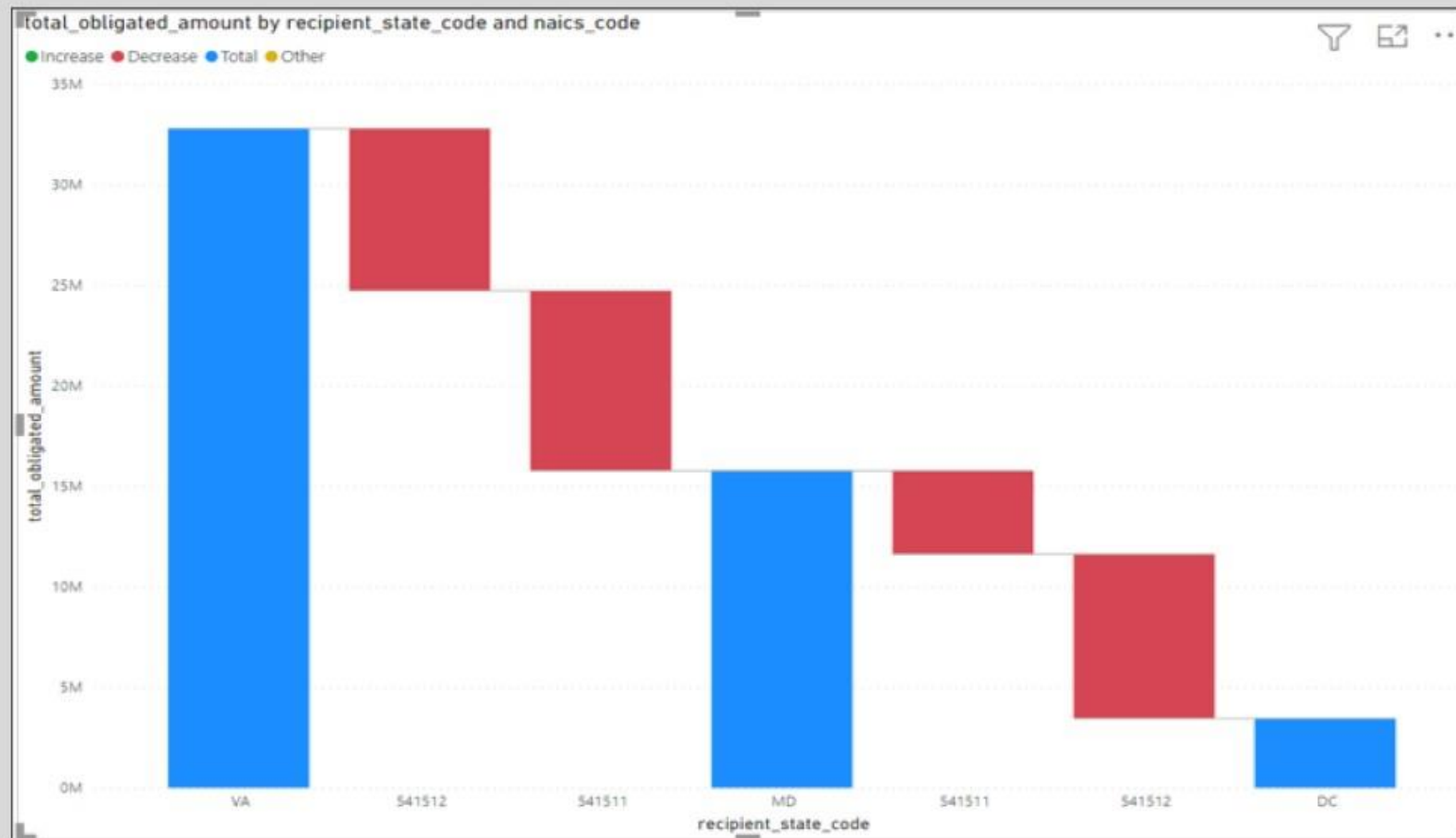
Data Visualization

- Ribbon chart:



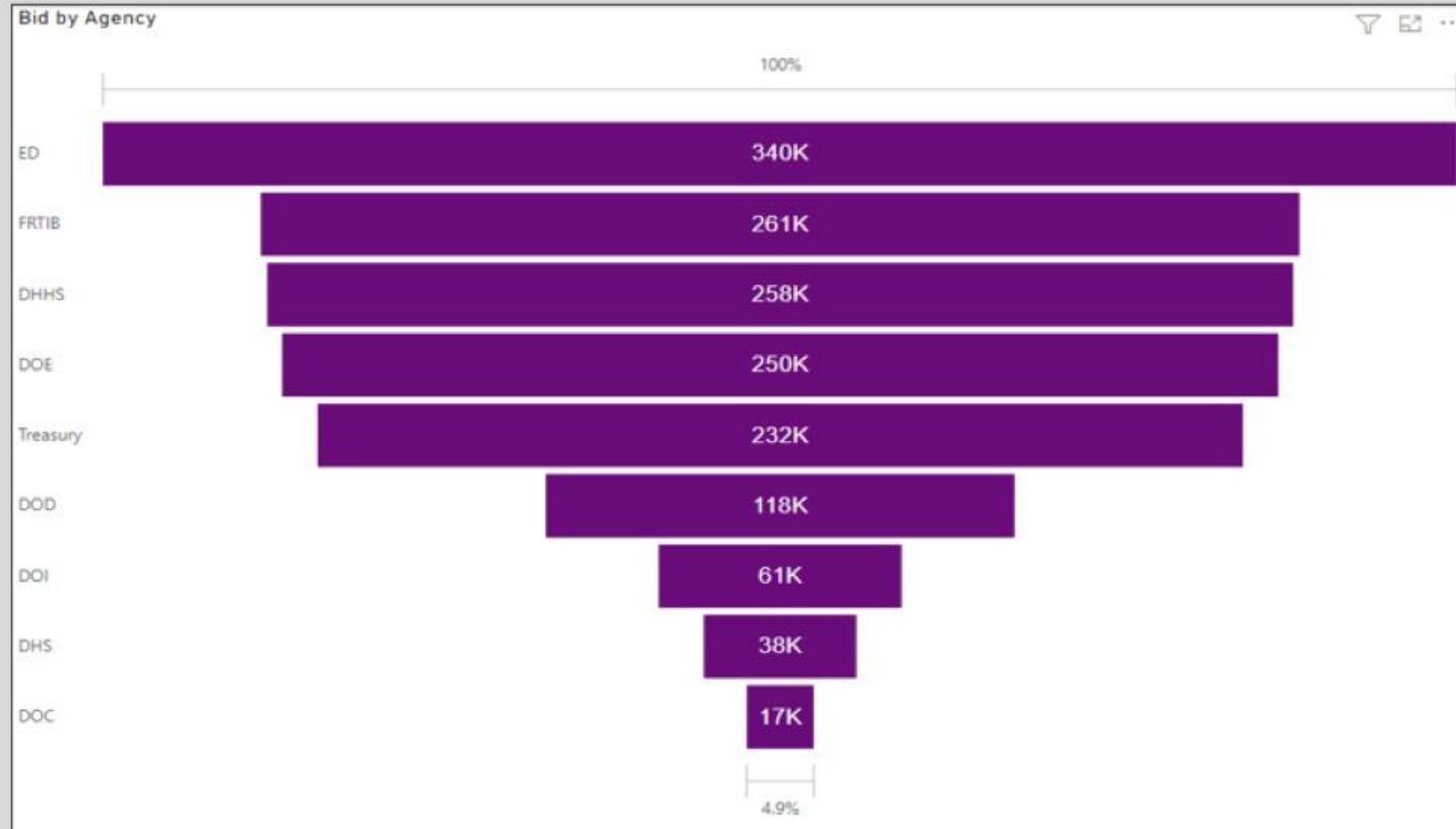
Data Visualization

- Waterfall chart:



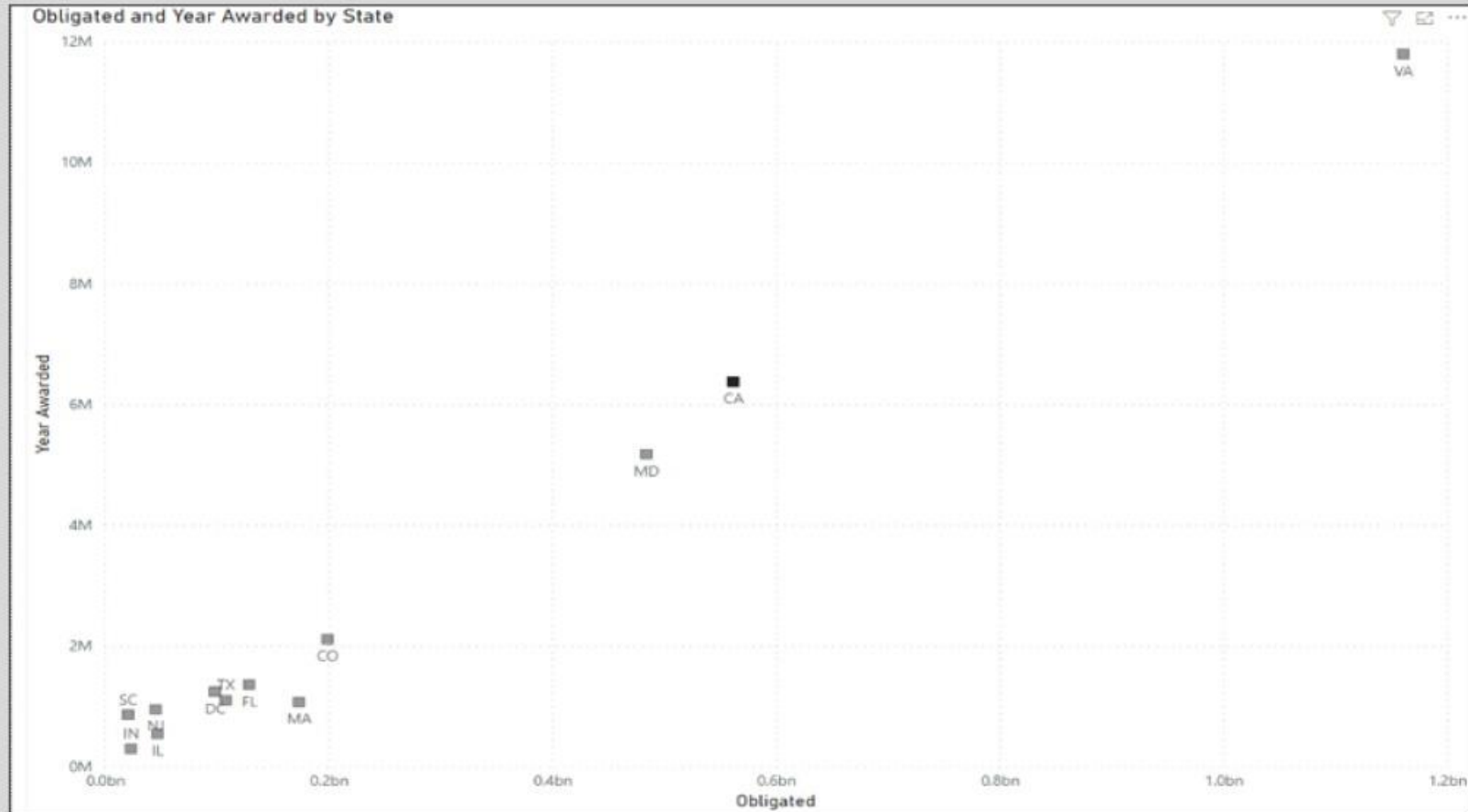
Data Visualization

- Funnel chart:



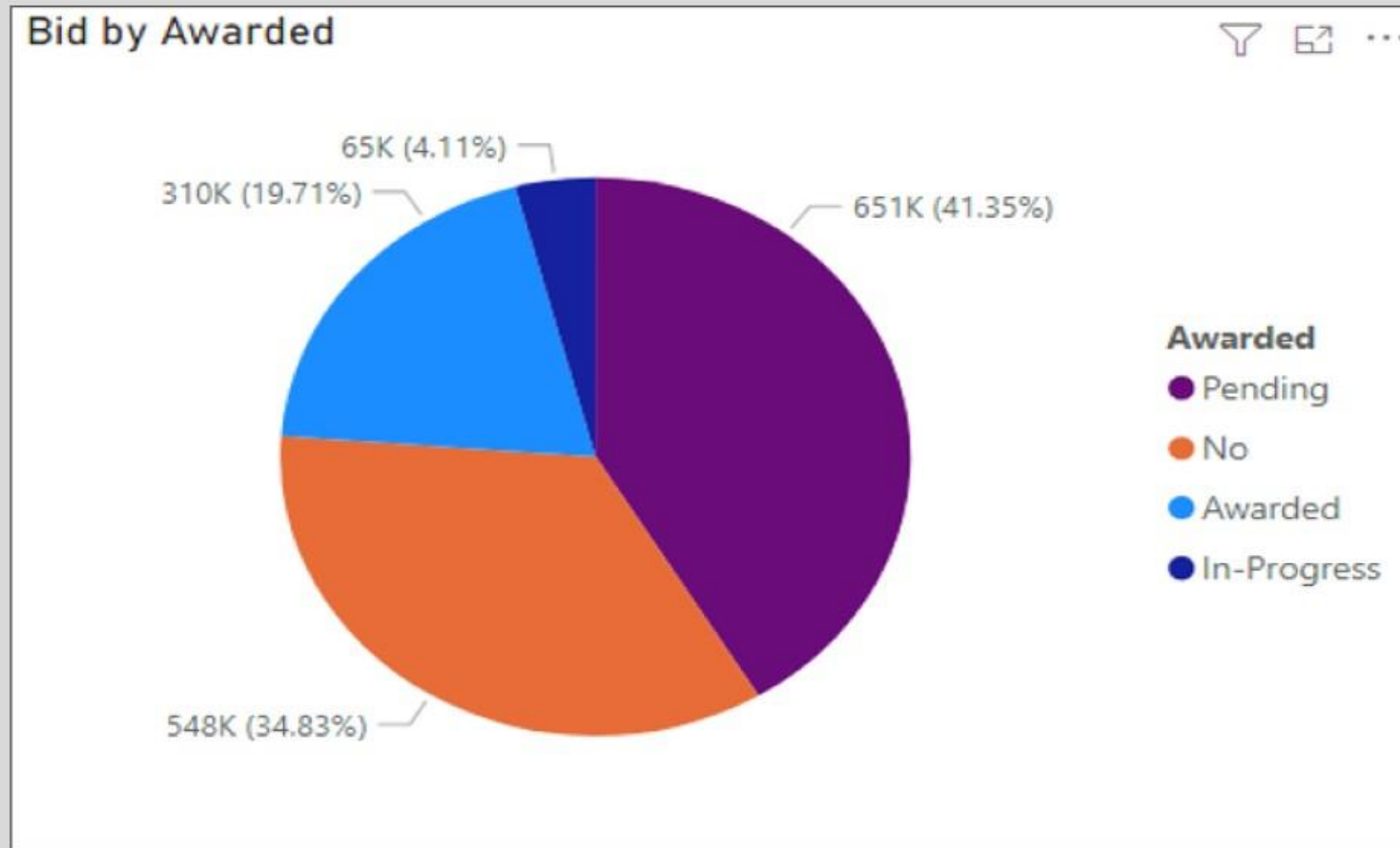
Data Visualization

- Scatter Plot:



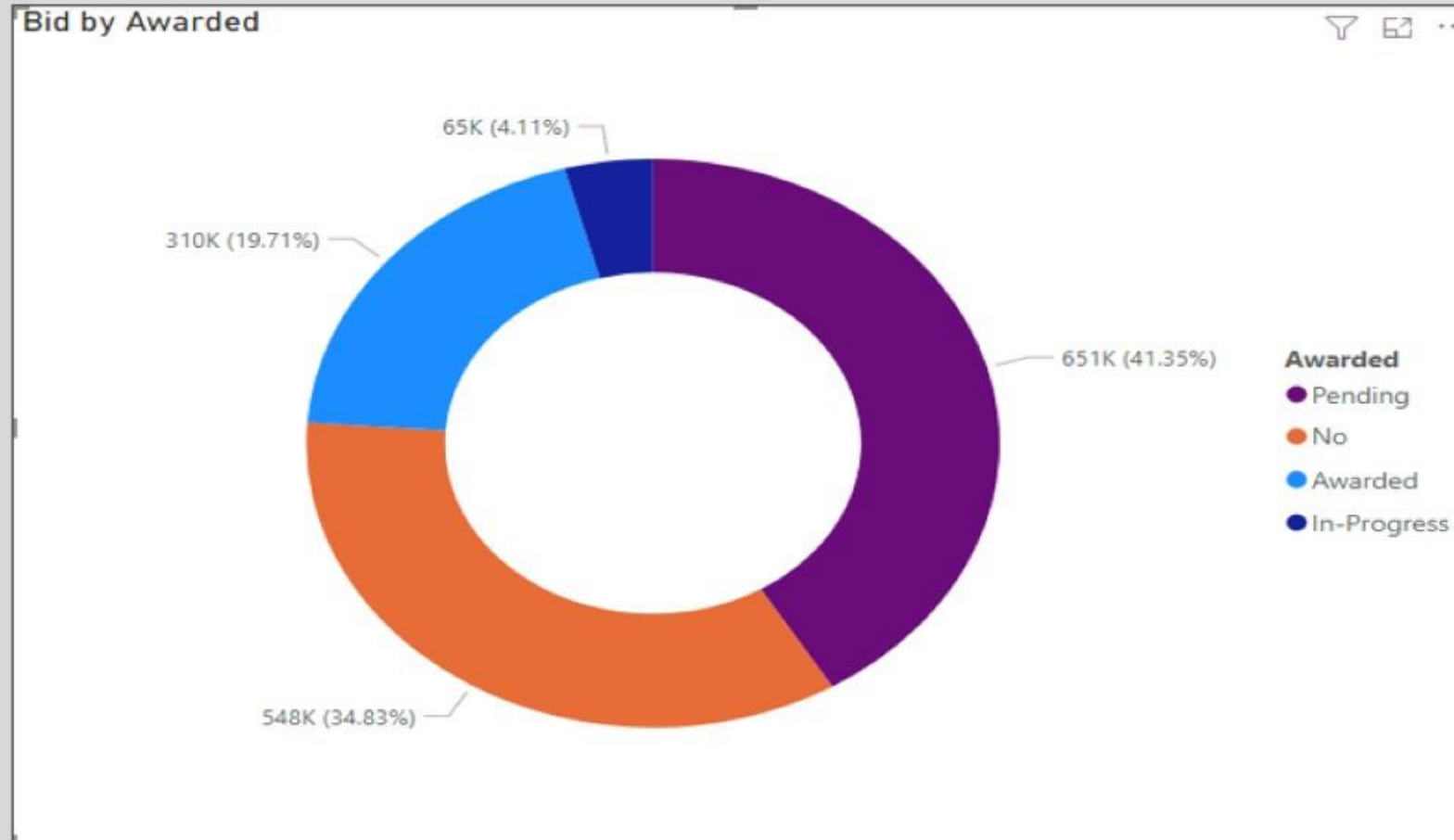
Data Visualization

- Pie chart:



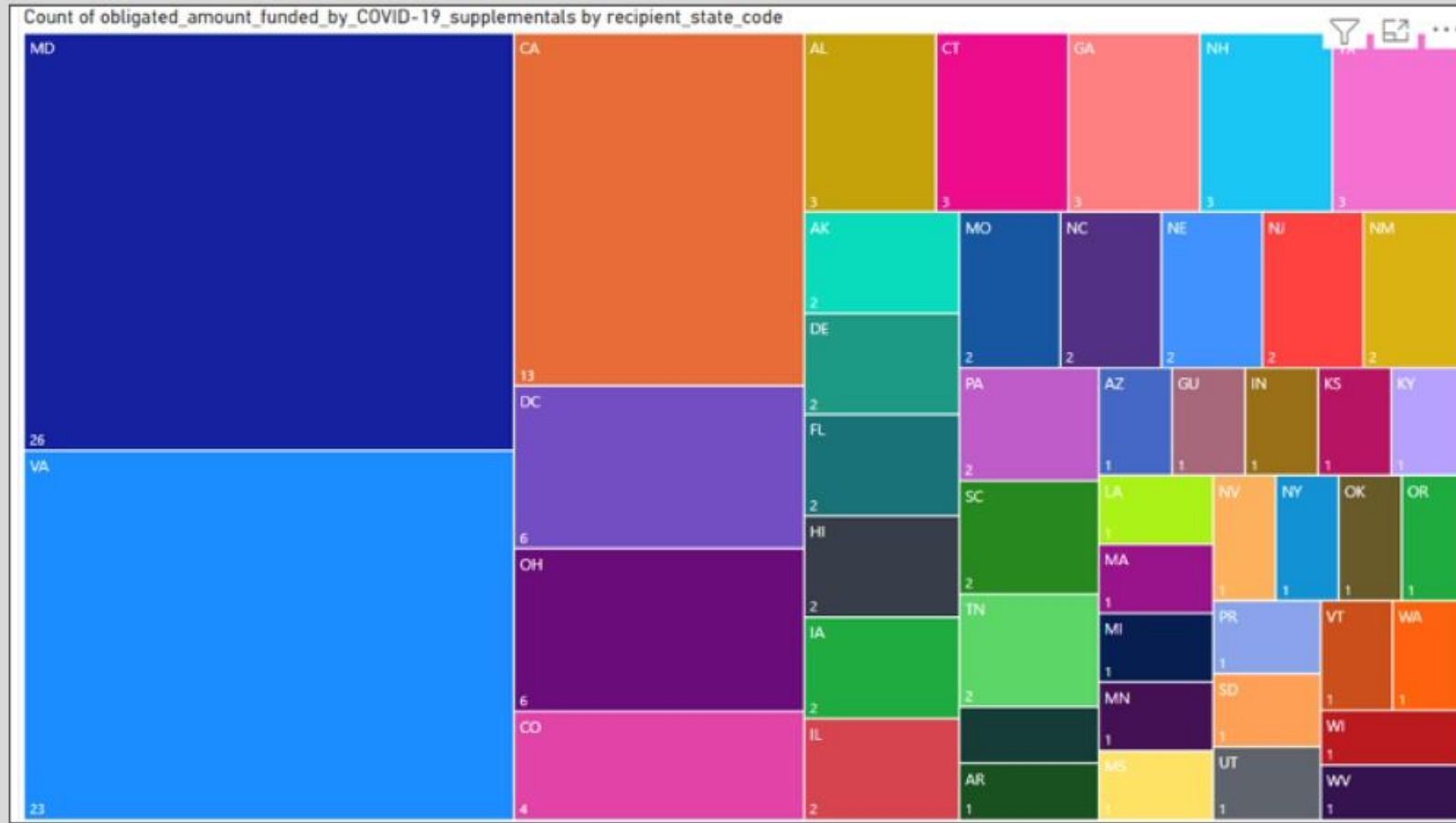
Data Visualization

- Donut chart:



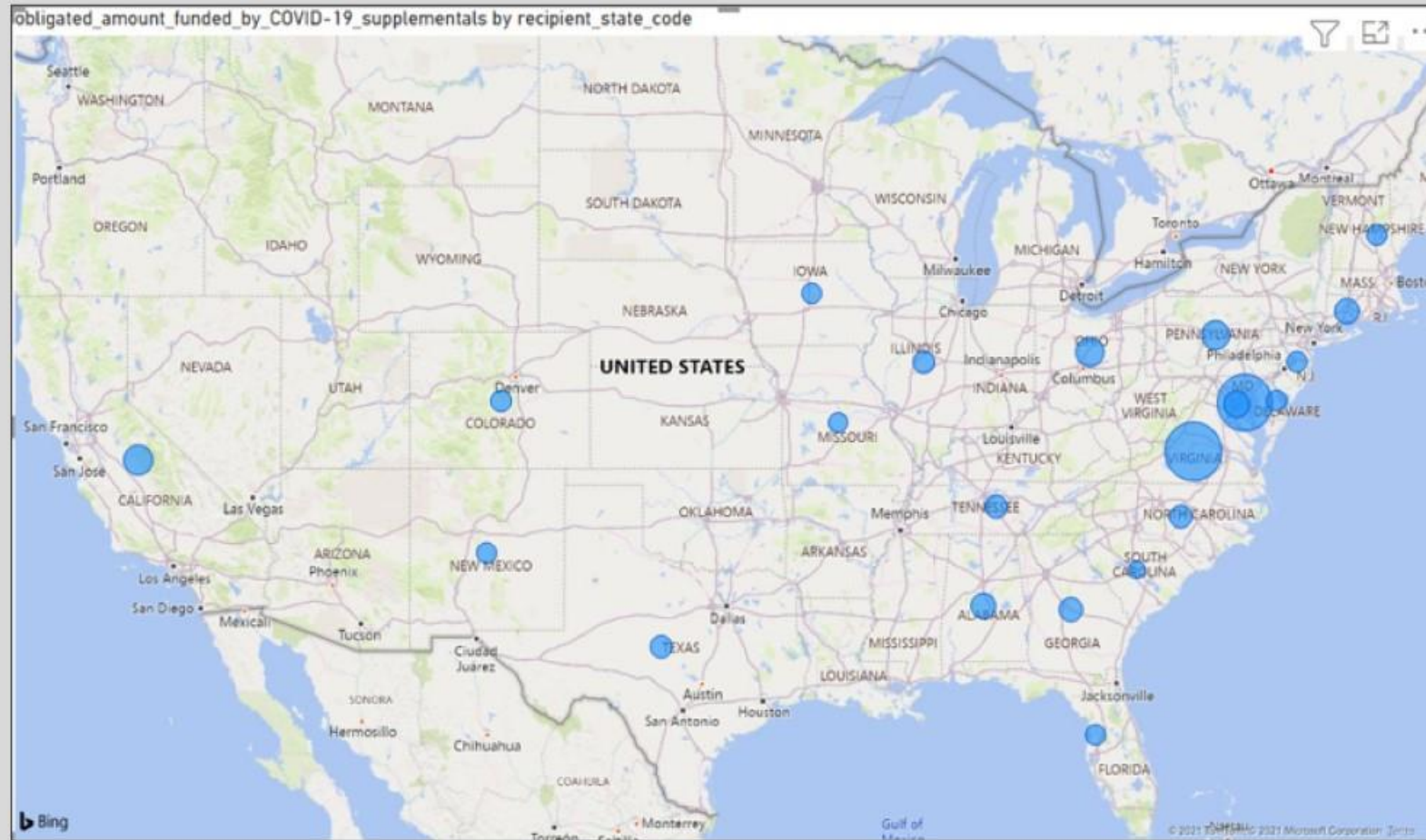
Data Visualization

- Treemap:



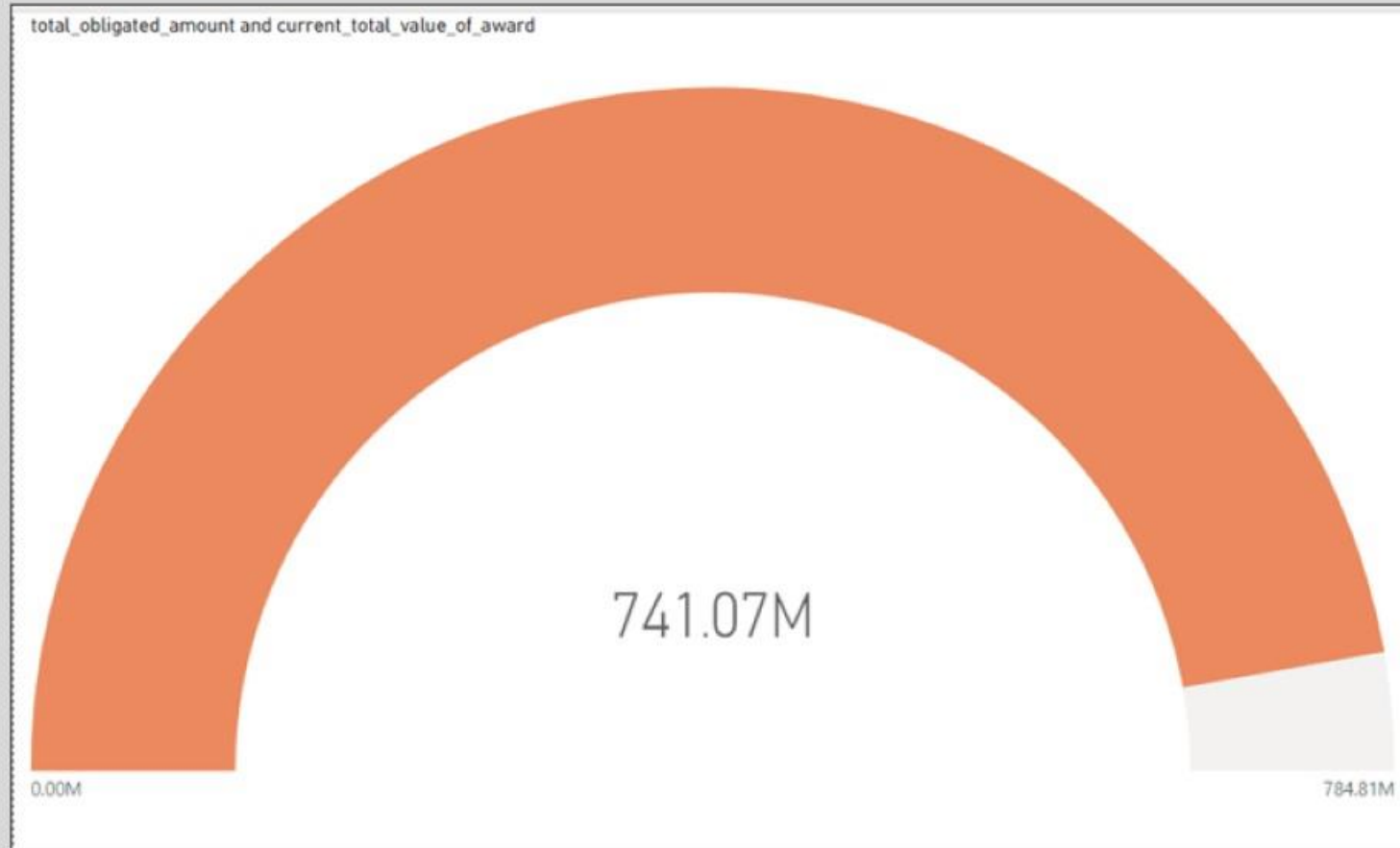
Data Visualization

- **Maps:**



Data Visualization

- Gauge:



Data Visualization

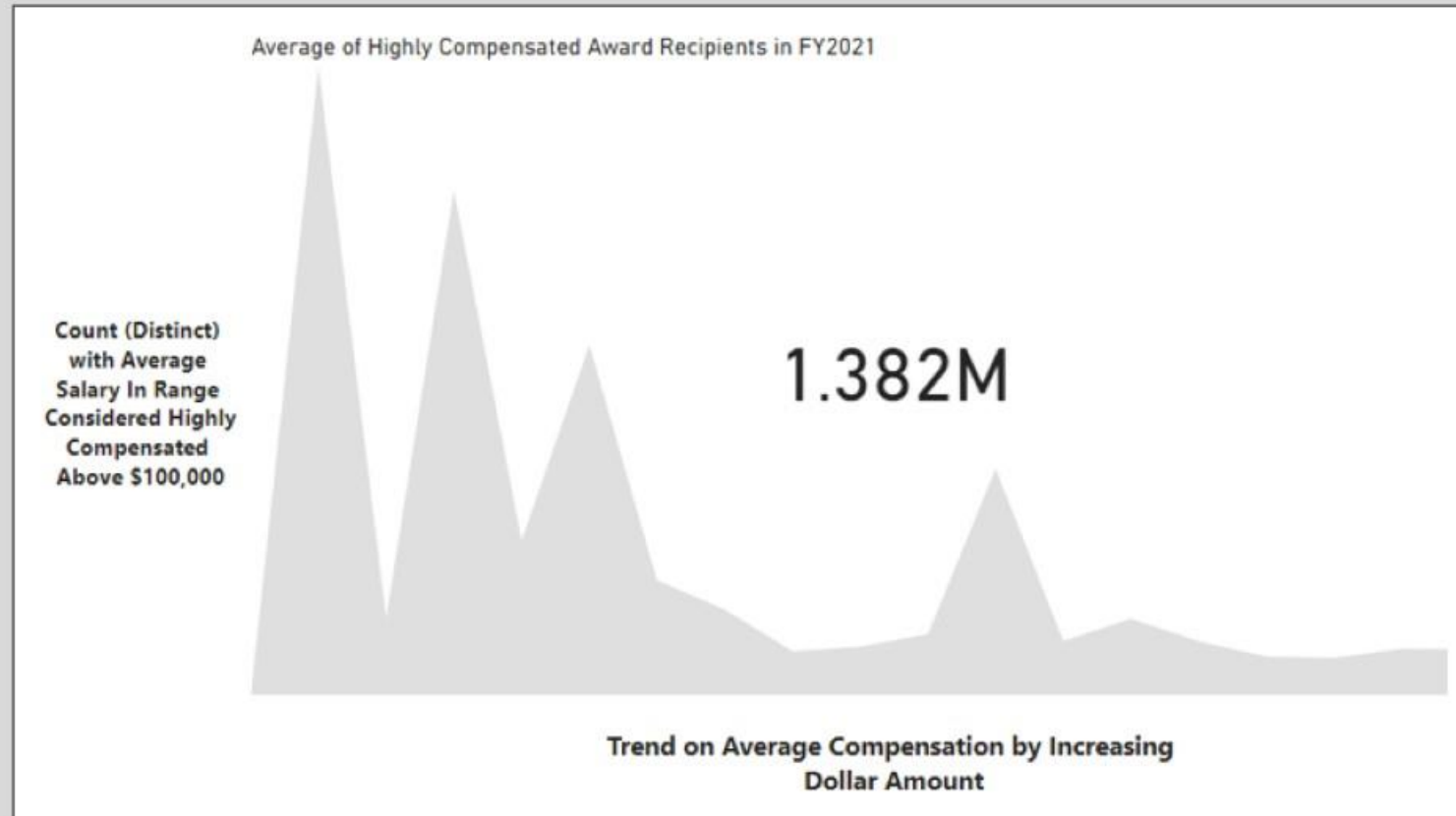
- Card:

5317

Count of Small Disadvantaged Business

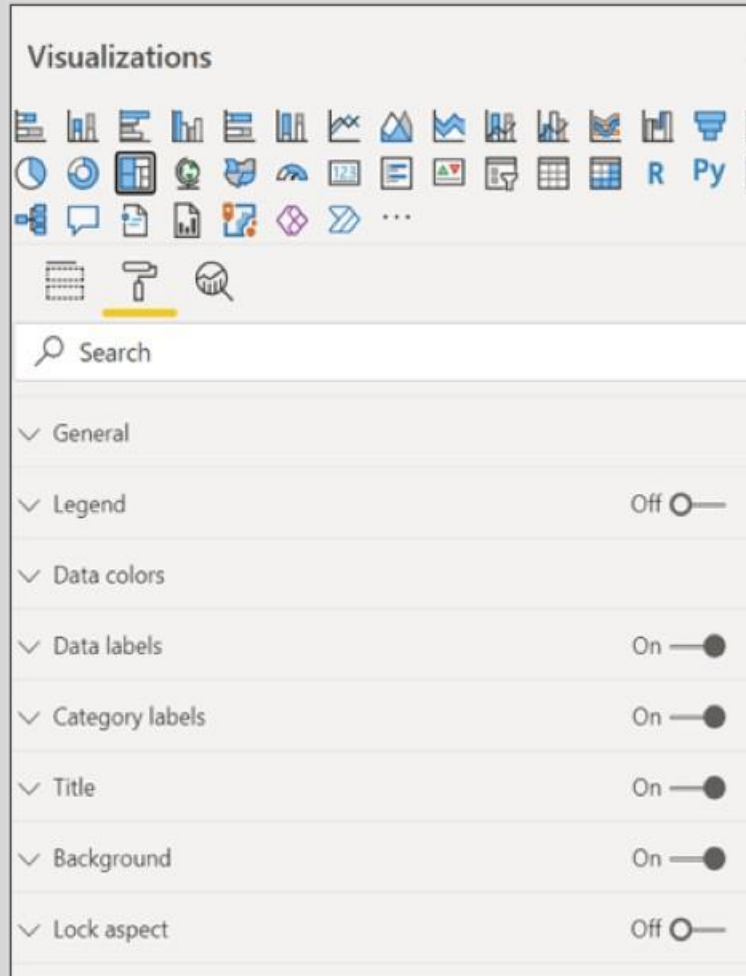
Data Visualization

- Key performance indicators (KPI):



Data Visualization

- **Formatting Features:**



Data Visualization

- **Formatting Features:**

^ Data colors

In-Progress

Pending

Revert to default

^ Data labels

On

Color

fx

Display units

Auto

Value decimal places

Auto

^ Title

On

Title text

Bid by Awarded and Awarded

fx

Title heading

Heading 3

Word wrap
On

Font color

fx

Background color

fx

Alignment

^ Background

On

Color

fx

Transparency

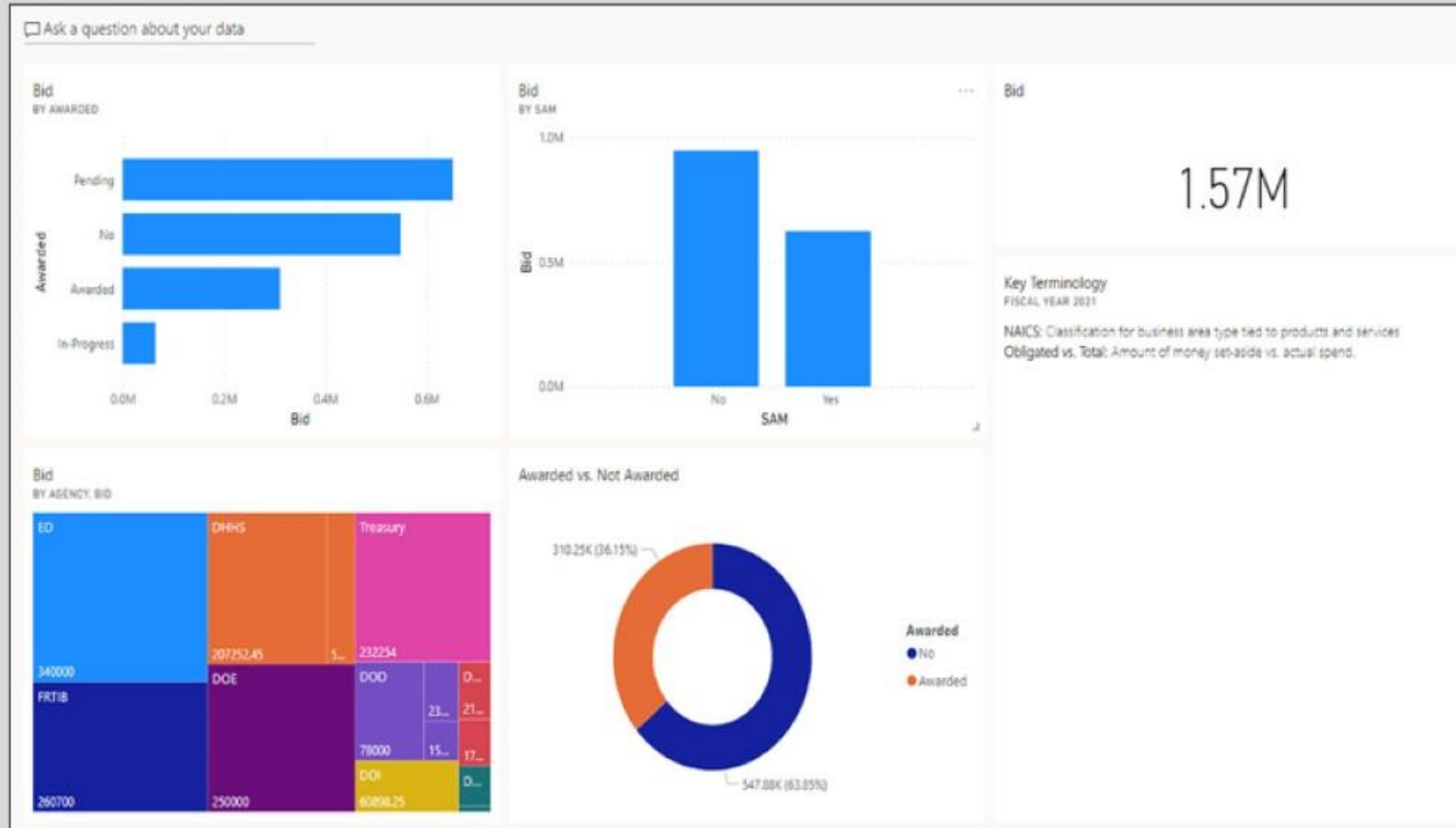
0 %

Revert to default

Dr. Amjad Qtaish

Data Visualization

Dashboard:



Data Visualization

