Insights

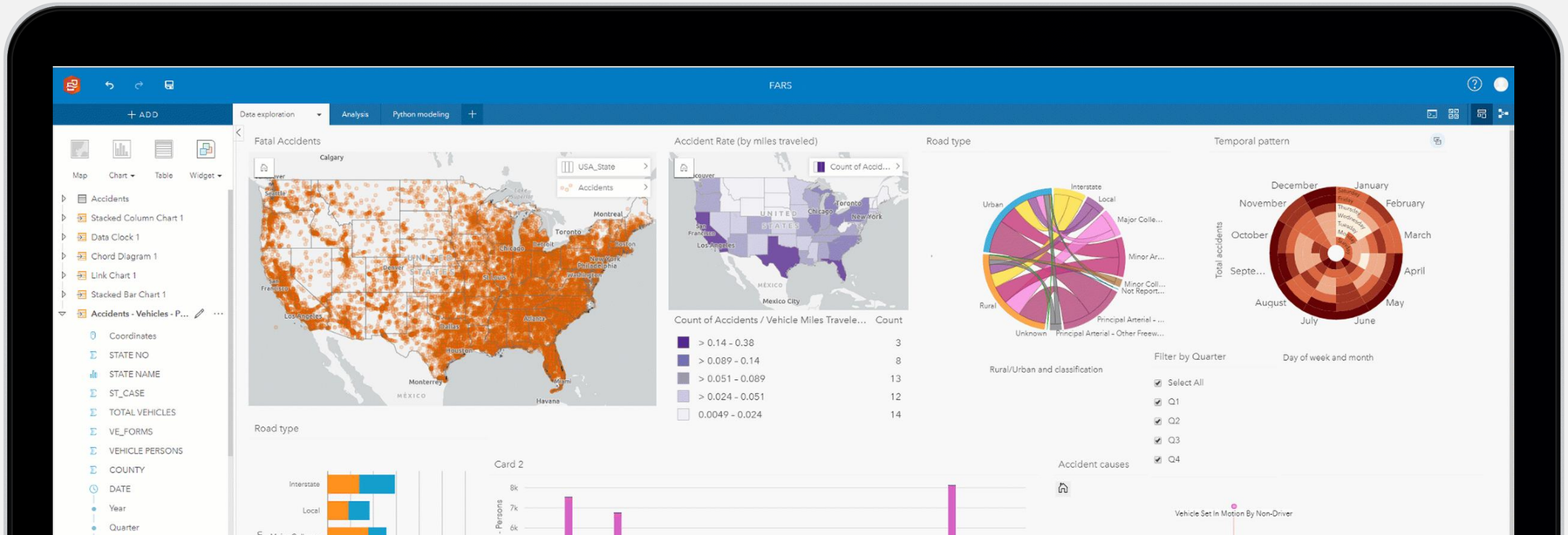
- Provides self-service visual analytics for multiple data types and sources
- Visualizations can be quantified using the accompanying descriptive statistics
- Exploratory analysis can be extended with graph analysis, predictive modeling and open data science

Self-service analytics

Whenever possible, we send the computation to the data. In some cases the data may be ‘live’. Interim results are saved in the workbook.

Data sets can be joined, and we do not flatten tables, which allows for 1-n and n-n joins. There are four attribute join types, compound joins and spatial joins.

Insights works with multiple shape fields, reducing data storage demands. Additional geocoding can be done with using Enable Location.



Key features

01
DATA CONNECTIONS

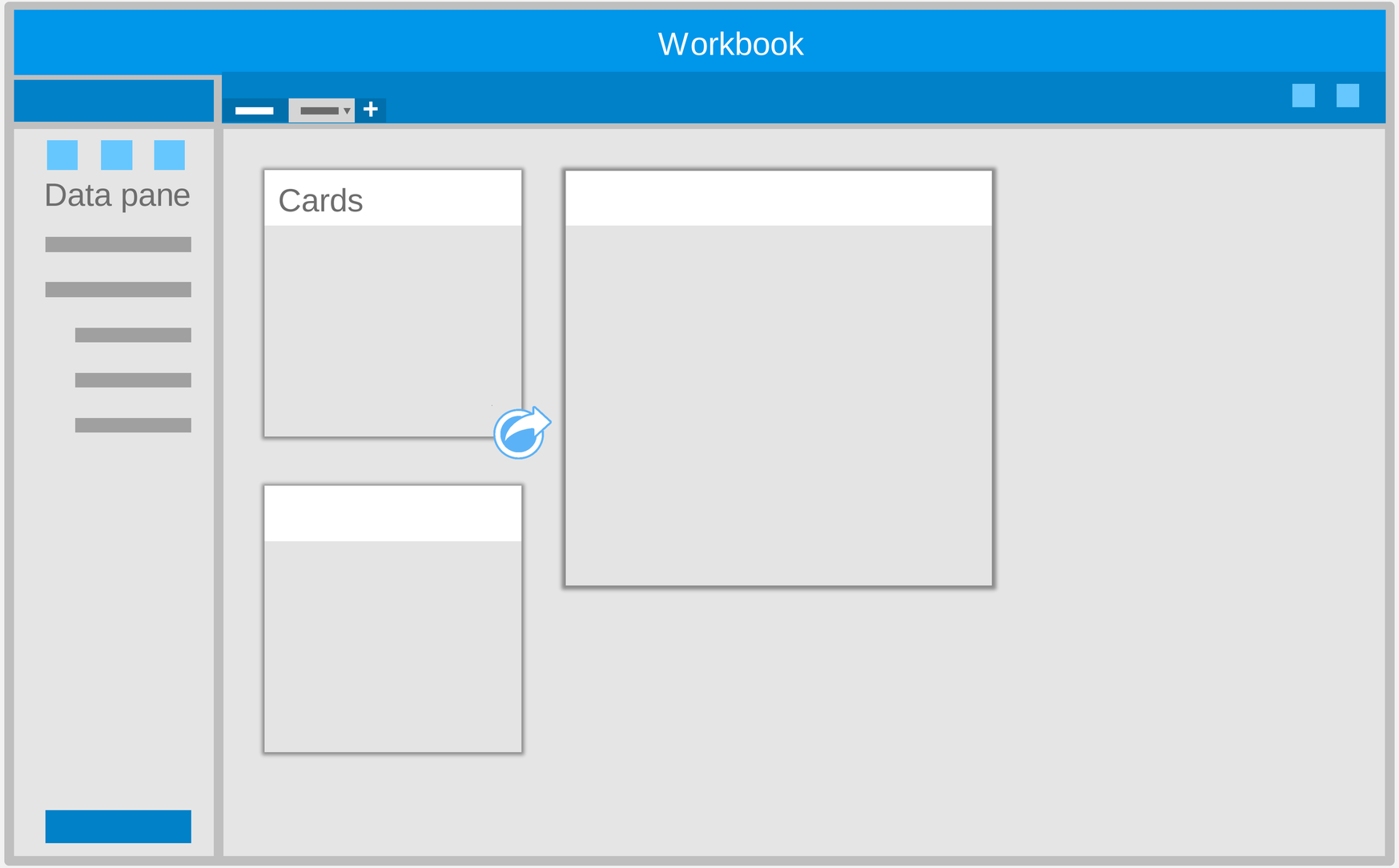
Work with data of multiple types, from numerous sources

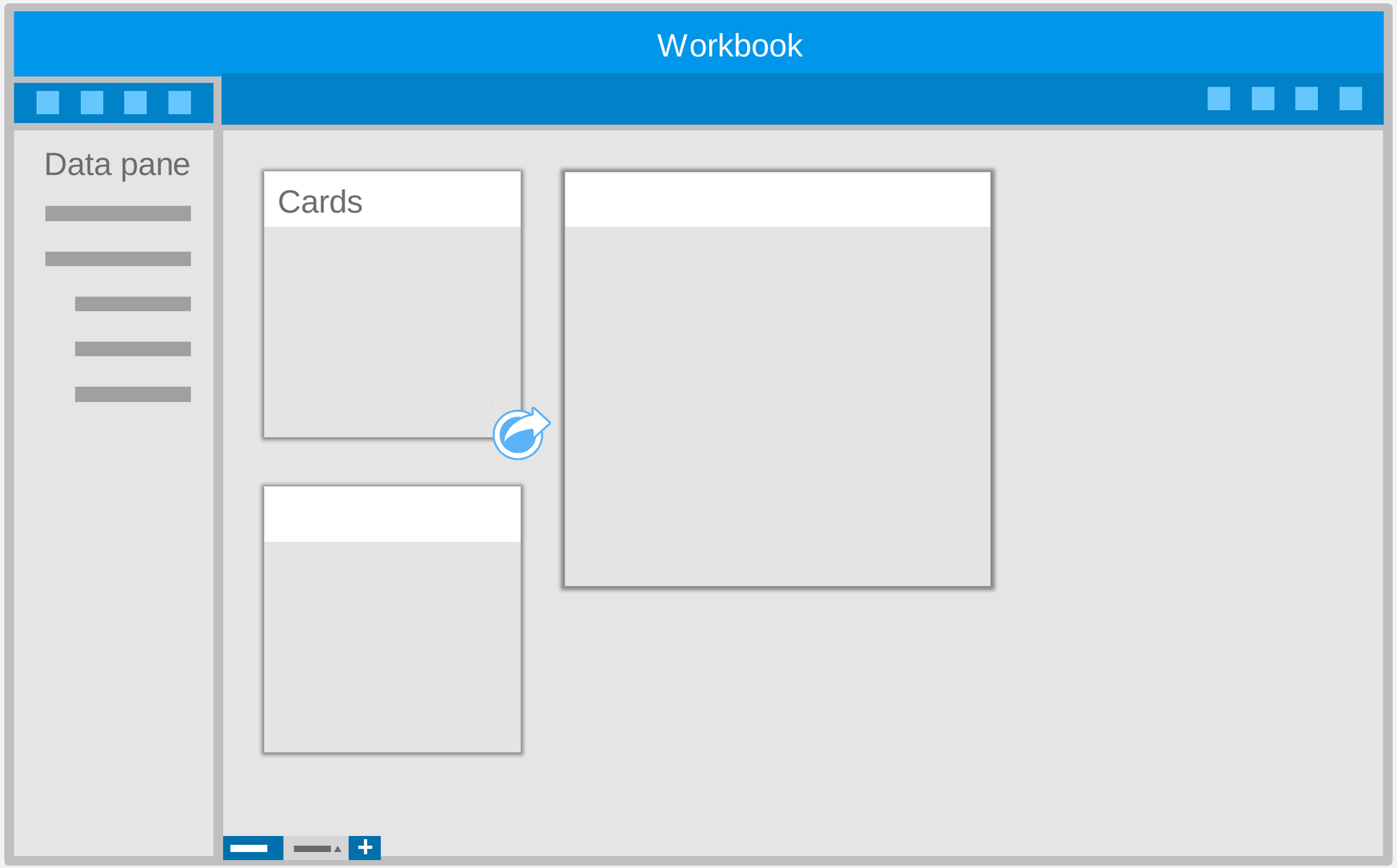
02
VISUAL ANALYTICS

Understand and analyze data easily and quickly share results

03
DATA SCIENCE

Extending analysis & visualizations using Python and or R in Insights


Workbook
Data pane
Cards


Workbook
Data pane
Cards

Data connections

01
DATA ACCESS

Work with data of multiple types, from numerous sources

02
PROCESSING AND STORAGE

Understand where data is analyzed, what is created and stored

03
EXPORT AND SHARING

When sharing results, what data is shared

Deployment and data options

ENTERPRISE

ONLINE

DESKTOP

DATA

- Feature layers
- GeoDatabases
- Shapefiles
- GeoJson
- Excel
- CSV

DATABASES (spatial or non-spatial tables)

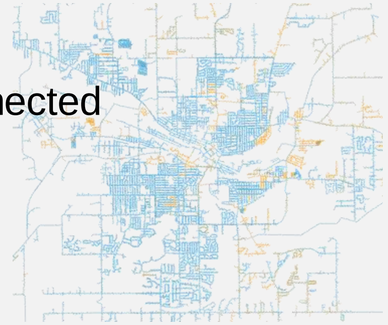
Microsoft SQL Server · ORACLE · PostgreSQL · PostGIS · SAP HANA

OPEN SOURCE

- Python scripts
- Python notebooks
- R scripts

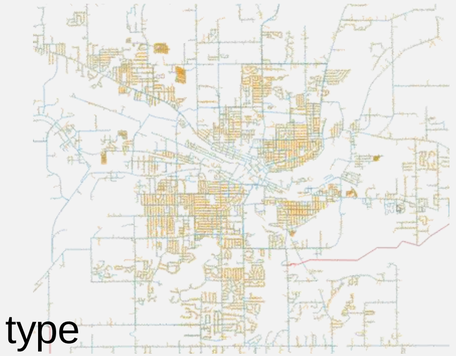
Domains and subtypes

Pipelines



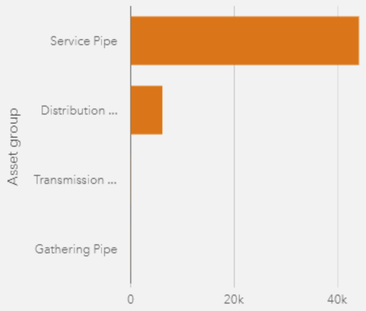
Coded value domain: Is connected

Value	Description
1	True
2	False



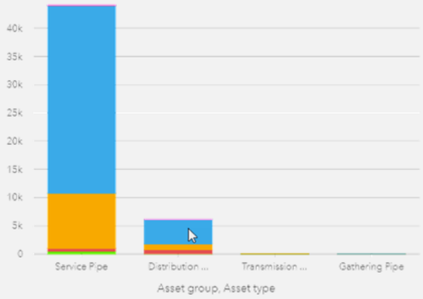
Subtype: Asset group

Value	Description
1	Steel
2	Cast iron
3	Steel
4	Plastic
5	Iron



Coded value domain: Asset type

Value	Description
1	Service
2	Distribution
3	Transmission
4	Gathering



Visual analytics

Combines automated analysis with interactive visualization for ease of understanding, interpretation and decision making.

01
SELECT FIELDS

Start with the variable you need to understand

02
DRAG N DROP

Select what information you want to evaluate

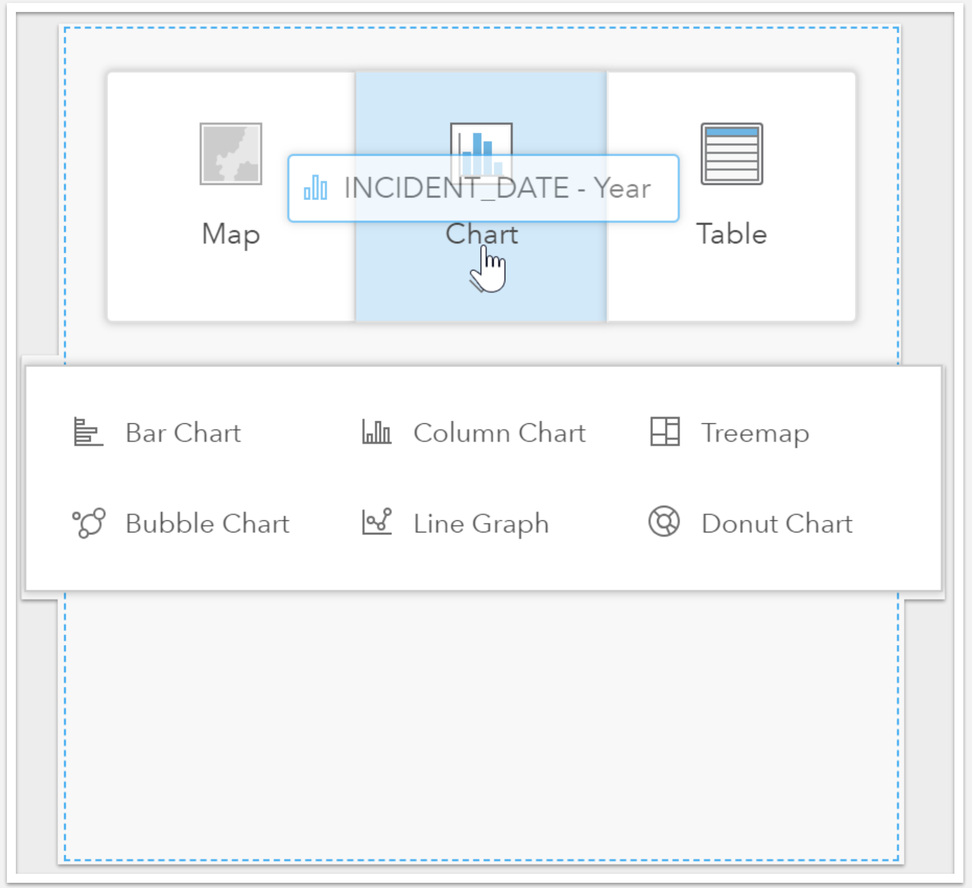
03
INTERPRET

Explore and interpret the results or, change the view or question

Data driven analytics

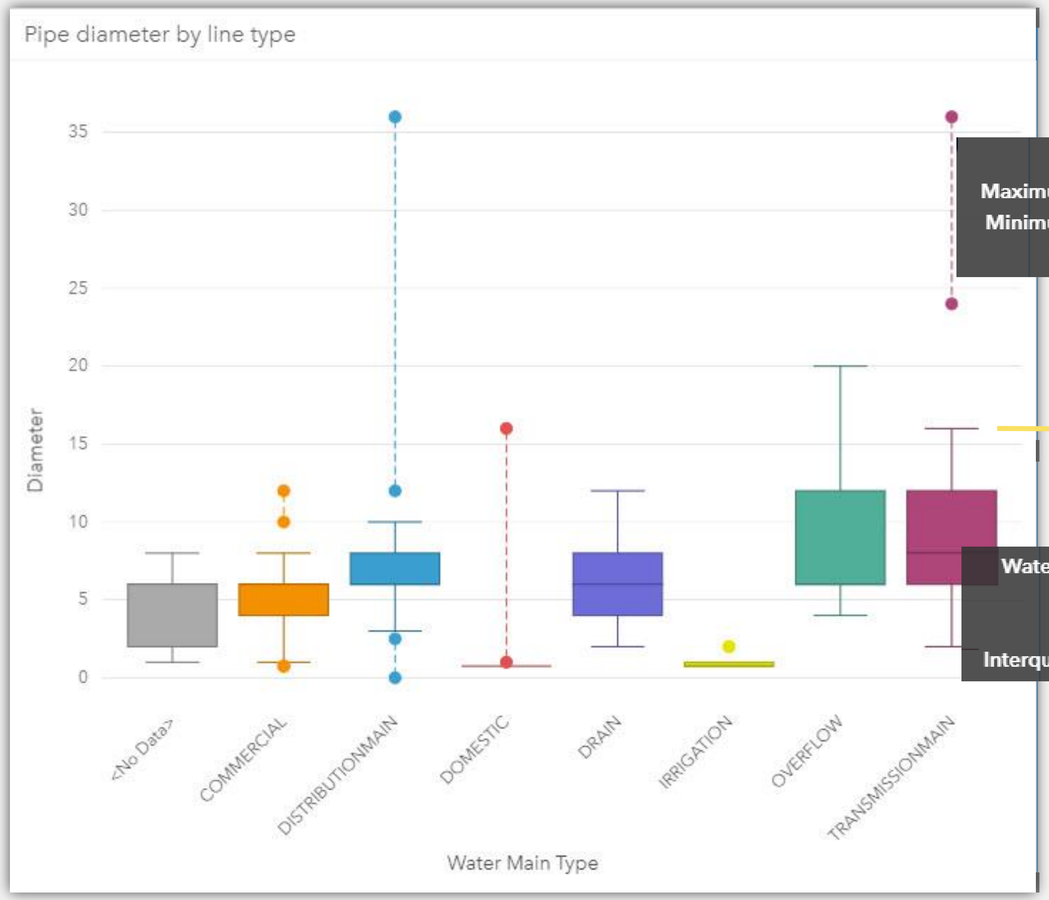
DYNAMIC SEMANTIC MODELING

Strings (Qualitative)
Numbers (Quantitative)
Data/Time (Temporal)
Spatial (Location)



Translating data to answers

Boxplots: show distribution of values - median, upper & lower quartiles, min & max values and, any outliers in the dataset.



OUTLIERS

MAXIMUM VALUE

EDIAN VALUE

HISKER (25% OF THE DATA)

Result dataset

If the analysis results in any changes to the source data, a new result dataset will be created and added to the data pane.

Line type	Pipe diameter
DistributionMain	10
Distribution	4
Transmission	6
DistributionMain	8
Drain	6
Overflow	8
Transmission	6
Transmission	10
Overflow	12
Drain	4
Drain	6
…	…

Line type	Min	Max	Count of lower outliers	Lower whisker	Upper whisker	Q1	Median	Q3	Min upper outlier values	Max upper outlier values	Count of upper outliers
Commercial	0.75	0.75	17	1	8	4	6	6	10	12	8
DistributionMain	0	2.5	872	3	10	6	6	8	12	36	632
Domestic			0	0.75	0.75	0.75	0.75	0.75	1	16	582
Drain			0	2	12	4	6	8			0
Irrigation			0	0.75	1	0.75	0.75	1	2	2	8
…	…										

This data can be shared and the interactivity can be used to select subsets e.g. outliers.

Insights for ArcGIS®

Data type: Qualitative Quantitative Temporal

Measure: ascertain the size, amount, or degree of (something)


A bar graph uses either horizontal or vertical bars to show comparisons among categories. They are valuable to identify broad differences between categories at a glance.


A treemap shows both the hierarchical data as a proportion of a whole and, the structure of data. The proportion of categories can easily be compared by their size.


Bubble charts represent numerical values of variables by area. With two variables (category and numeric), the circles placed so they are packed together.


A heat chart shows total frequency in a matrix. Values in each cell of the rectangular grid are symbolized into classes.

Relationship: a connection or similarity between two or more things or, the state of being related to something else


A choropleth map allows quantitative values to be mapped by area. They should show normalized values not counts collected over unequal areas or populations.


A chord diagram visualizes the inter-relationships between categories and allows comparison of similarities within a dataset or, between different groups of data.


Scatterplots allow you to look at relationships between two numeric variables with both scales showing quantitative variables. The level of correlation can also be quantified.


Link analysis is used to investigate relationships between entities where and an entity is an object, person, place or event. Links connect two or more entities.


Spider lines , also termed desire lines, show paths between origins and destinations. They show connections between places.

Change: process through which something becomes different, often over time


A bar graph uses either horizontal or vertical bars to show comparisons among categories. They are valuable to identify broad differences between categories at a glance.


A heat chart shows total frequency in a matrix. Using a temporal axis values, each cell of the rectangular grid are symbolized into classes over time.


Bubble charts with three numeric variables are multivariate charts that show the relationship between two values while a third value is shown by the circle area.


Graduated symbol maps show a quantitative difference between mapped features by varying symbol size. Data are classified with a symbol assigned to each range.


A Density/heat map calculates spatial concentrations of events or values enabling the distribution to be visualized as a continuous surface.


A Data clock creates a circular chart of temporal data, commonly used to see the number of events at different periods of time.


Line graphs visualize a sequence of continuous numeric values and are used primarily for trends over time. They show overall trends and changes from one value to the next.


A combo chart combines two graphs where they share common information on the x-axis. They allow relationships between two datasets to be shown.

Interaction: flow of information, products or goods between places


A chord diagram visualizes the inter-relationships between categories and allows comparison of similarities within a dataset or, between different groups of data.


Spider lines, also termed desire lines, show paths between origins and destinations. Flow maps show directional connections and flow between places.

Distribution: the arrangement of phenomena, could be numerically or spatially


Histograms show the distribution of a numeric variable. The bar represents the range of the class bin with the height showing the number of data points in the class bin.


A box plot displays data distribution showing the median, upper and lower quartiles, min and max values and, outliers. Distributions between many groups can be compared.


A choropleth map allows quantitative values to be mapped by area. They should show normalized values not counts collected over unequal areas or populations.


Graduated symbol maps show a quantitative difference between mapped features by varying symbol size. Data are classified with a symbol assigned to each range.


A Density/heat map calculates spatial concentrations of events or values enabling the distribution to be visualized as a continuous surface.


A unique symbol map (areas or points) allows descriptive (qualitative) information to be shown by location. Areas have different fills and points can be geometric or pictorial.

Part-to-whole: relative proportions or percentages of categories, showing the relationship between parts and whole


Donut charts are used to show the proportions of categorical data, with the size of each piece representing the proportion of each category.

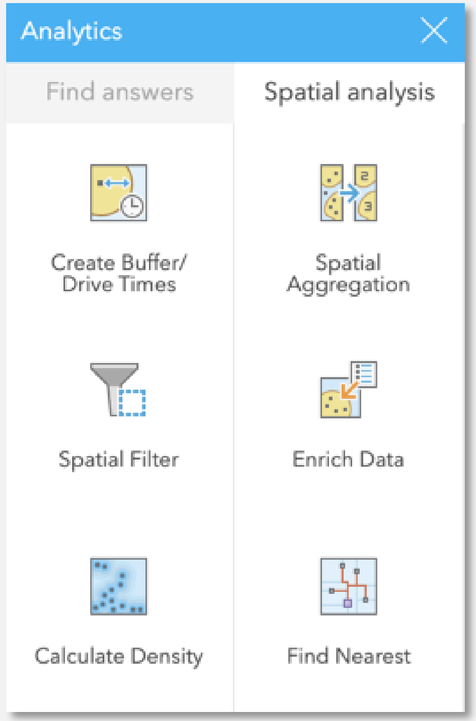

A treemap shows both the hierarchical data as a proportion of a whole and, the structure of data. The proportion of categories can easily be compared by their size.

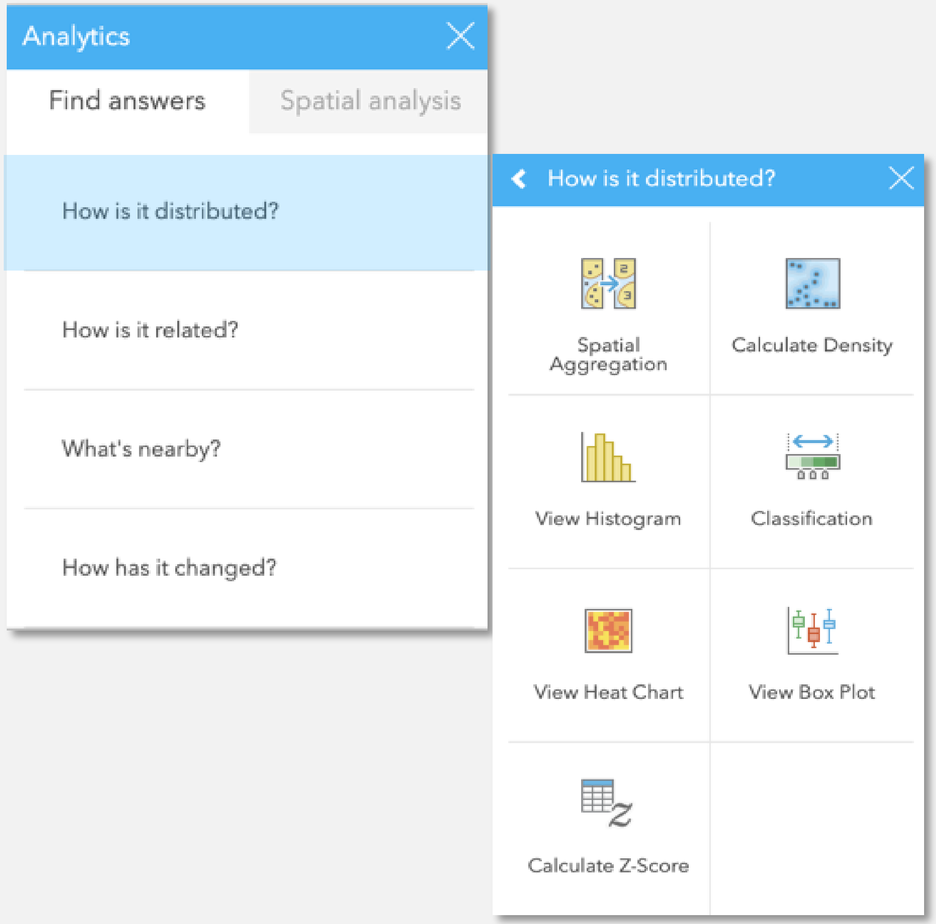
Acknowledgement
Inspired by work by Jon Schwabish and Severino Ribecca, The Graphic Continuum, 2014 and, Alan Smith et al. Visual Vocabulary, The Financial Times, 2016

esri®

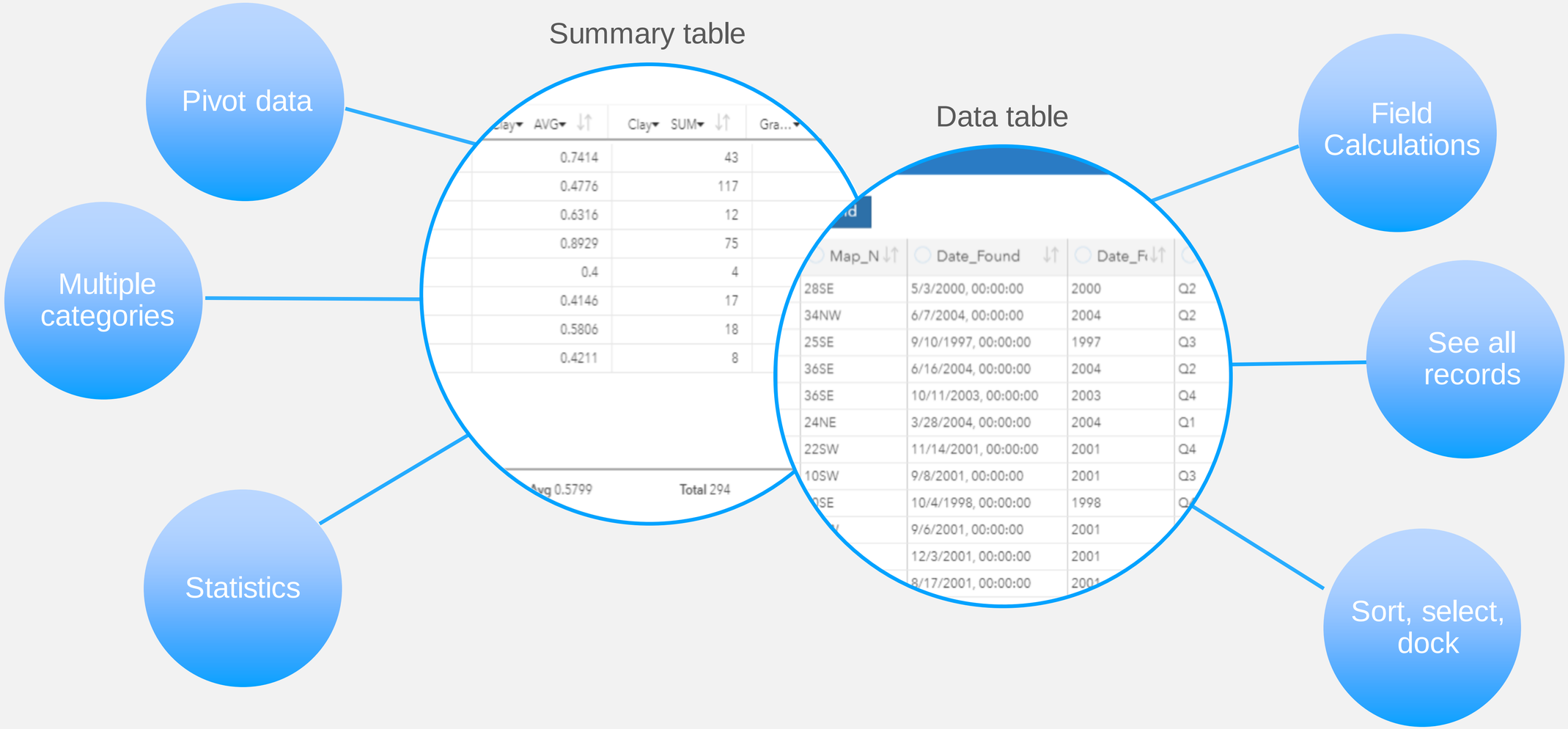
Linda Beale PhD, 2017

Action button

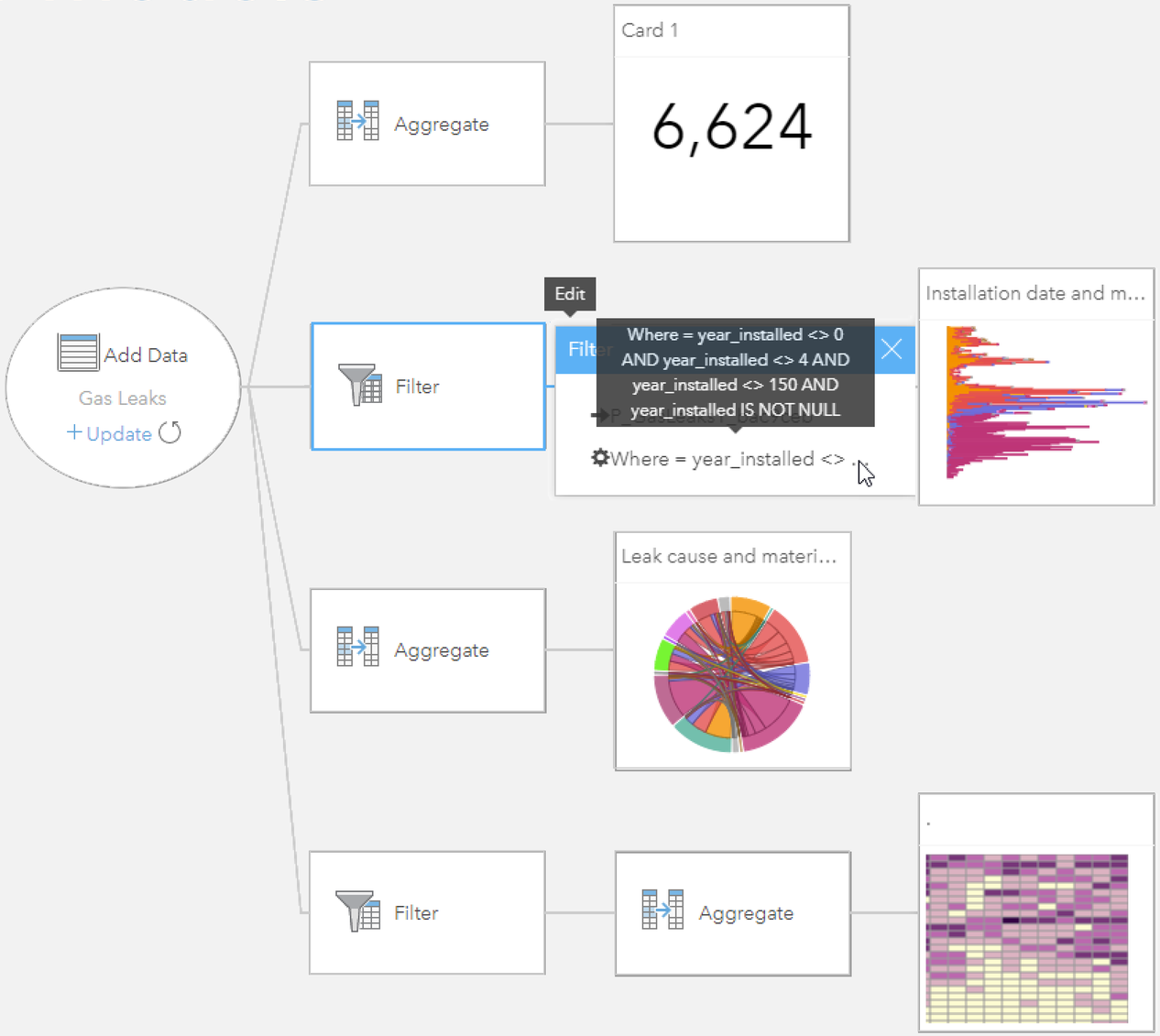




Tables

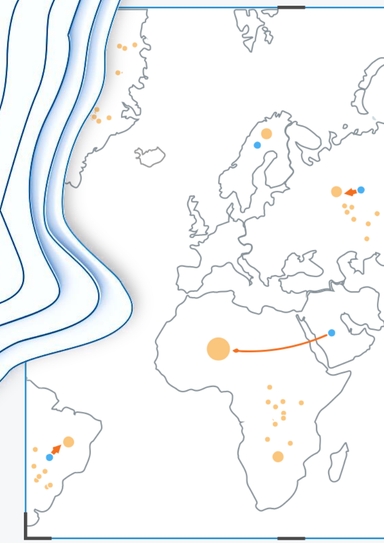

Pivot data
Multiple categories
Statistics
Summary table
Clay AVG
Clay SUM
Gra...
0.7414 43
0.4776 117
0.6316 12
0.8929 75
0.4 4
0.4146 17
0.5806 18
0.4211 8
Avg 0.5799
Total 294
Data table
Map_N
Date_Found
Date_F
28SE 5/3/2000, 00:00:00 2000 Q2
34NW 6/7/2004, 00:00:00 2004 Q2
25SE 9/10/1997, 00:00:00 1997 Q3
36SE 6/16/2004, 00:00:00 2004 Q2
36SE 10/11/2003, 00:00:00 2003 Q4
24NE 3/28/2004, 00:00:00 2004 Q1
22SW 11/14/2001, 00:00:00 2001 Q4
10SW 9/8/2001, 00:00:00 2001 Q3
10/4/1998, 00:00:00 1998
9/6/2001, 00:00:00 2001
12/3/2001, 00:00:00 2001
8/17/2001, 00:00:00 2001
Field Calculations
See all records
Sort, select, dock

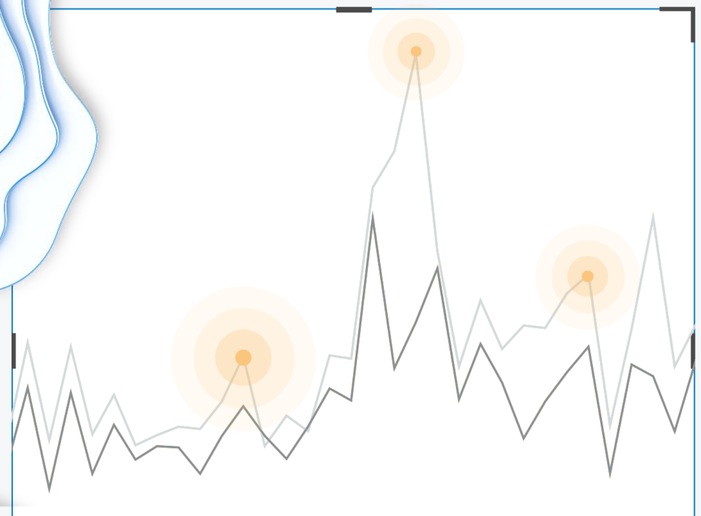
Workflow models



Demo

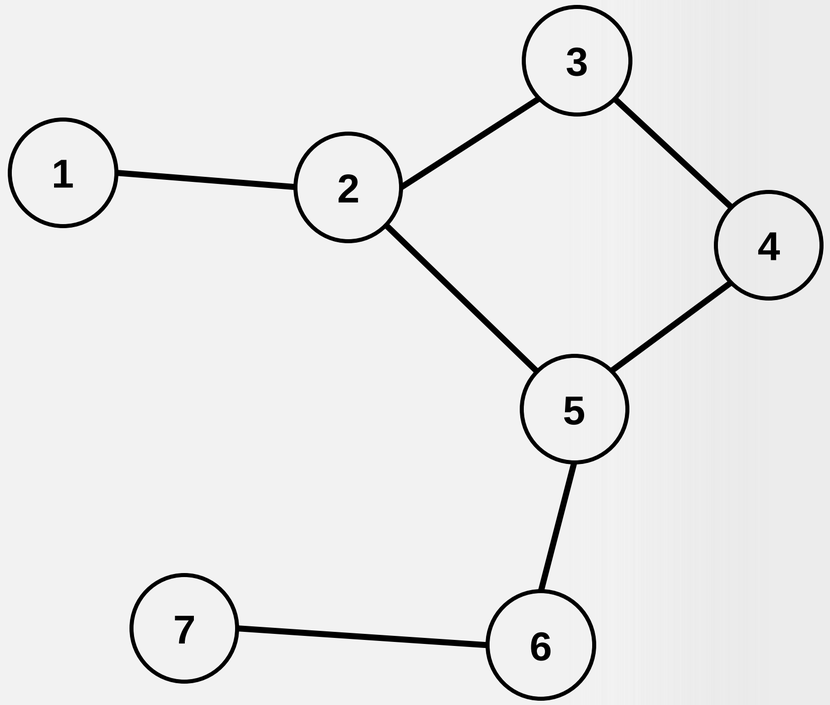
TITLE OF ANALYSIS





Link analysis

Link, network or graph analysis focuses on evaluating relationships, interactions and connections.



ON CHARTS AND MAPS

- Centrality metrics
- Network layouts

Network layout

Different layouts allow you to better visualize relationships in the data.

FORCE-DIRECTED

Nodes and edges are evenly distributed. Links have a consistent length.

Effective for identifying patterns.

ALL LAYOUT TYPES

Hide/show leaf nodes

Change direction

Select root node

HIERARCHICAL

Traditional organizational layout. All links will flow down from parent nodes.

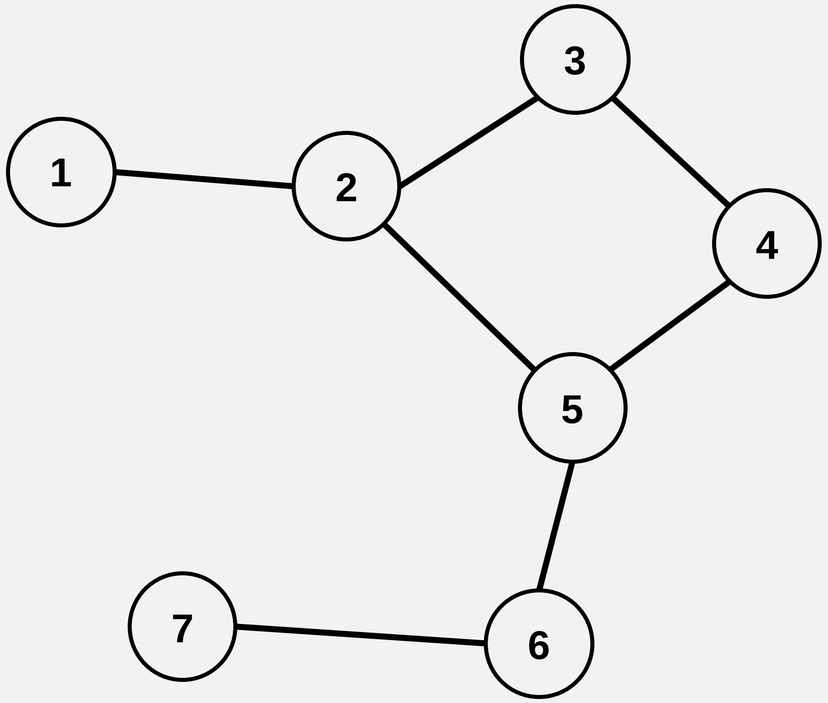
RADIAL

Places nodes in concentric circles around parent nodes.

Good for networks where parent nodes have many child nodes.

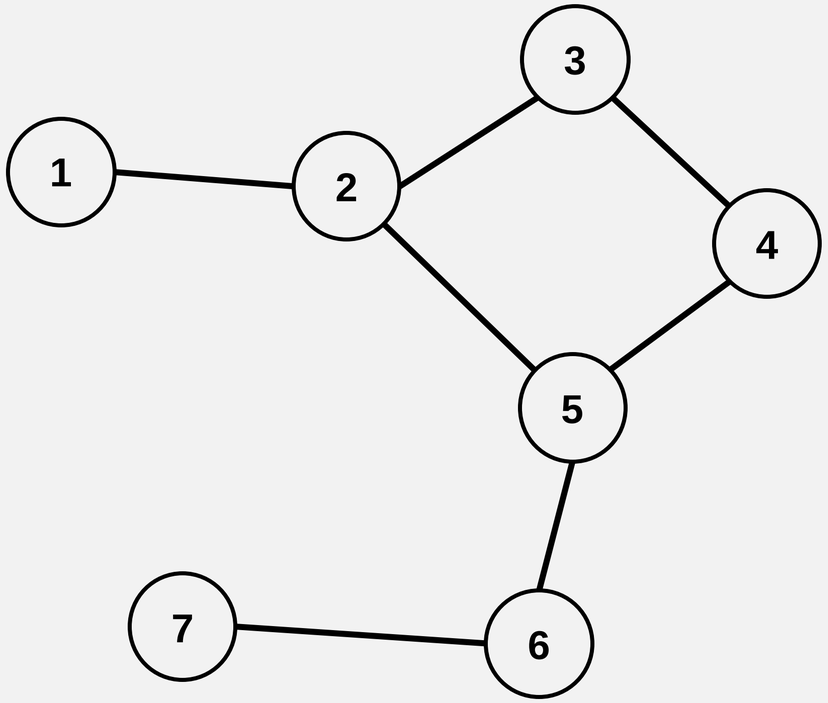
Set central node

Degree centrality



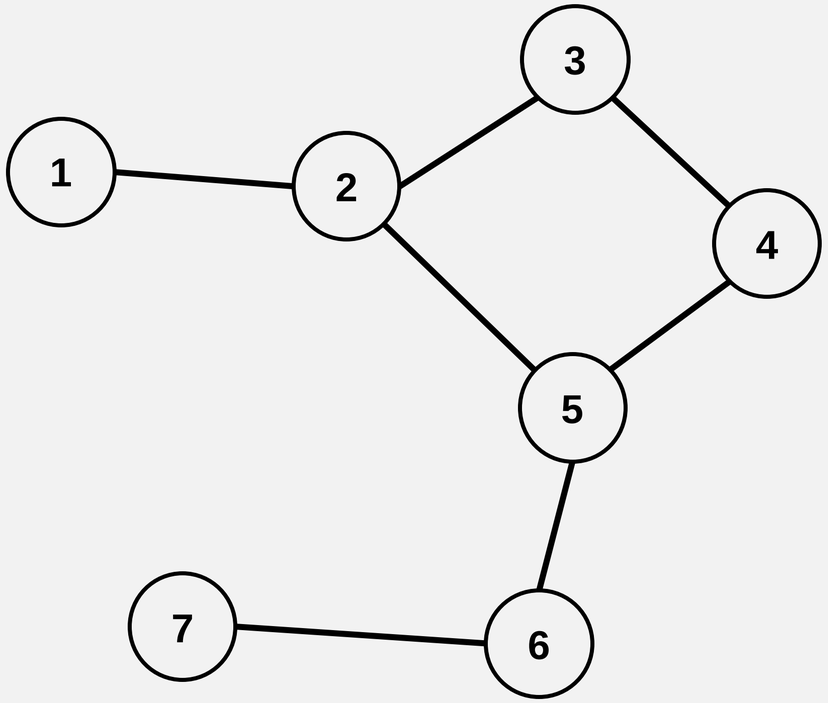
Node	Score	Normalize	Normalized score
1	1	1 / (7-1)	0.1667
2	3	2 / (7-1)	0.5
3	2	2 / (7-1)	0.3333
4	2	2 / (7-1)	0.3333
5	3	3 / (7-1)	0.5
6	2	2 / (7-1)	0.333
7	1	1 / (7-1)	0.1667

Closeness centrality



Nodes								Calculation	Normalized score
	1	2	3	4	5	6	7		
1	0	1	2	3	2	3	4	15 / (7-1)	0.4
2	1	0	1	2	1	2	3	10 / (7-1)	0.6
3	2	1	0	1	2	3	4	13/ (7-1)	0.4615
4	3	2	1	0	1	2	3	12/ (7-1)	0.5
5	2	1	2	1	0	1	2	9/ (7-1)	0.6667
6	3	2	3	2	1	0	1	12/ (7-1)	0.5
7	4	3	4	3	2	1	0	17/ (7-1)	0.3529

Betweenness centrality



Node	Score	Normalized score
1	0	0
2	6.5	0.4333
3	1	0.6677
4	1.5	0.1
5	9	0.6
6	5	0.3333
7	0	0

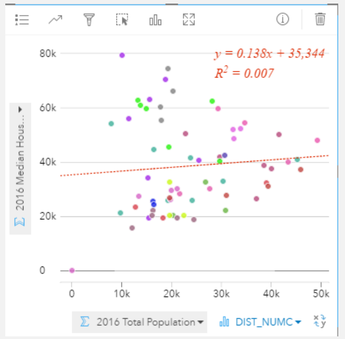
Node 2		
1,3	1 / 1	1
1,4	2 / 2	1
1,5	1 / 1	1
1,6	1 / 1	1
1,7	1 / 1	1
3,4	0 / 1	0
3,5	1 / 2	0.5
3,6	1 / 2	0.5
3,7	1 / 2	0.5
4,5	0 / 1	0
4,6	0 / 1	0
4,7	0 / 1	0
5,6	0 / 1	0
5,7	0 / 1	0
7,6	0 / 1	0
7,5	0 / 1	0
		6.5

Regression analysis

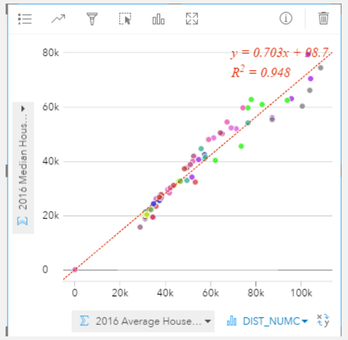
OLS regression is a statistical technique for the analysis and modeling of linear relationships.

SCATTERPLOT & CORRELATIONS

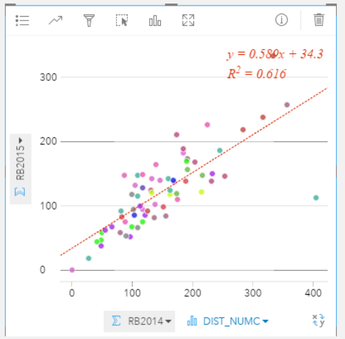
Strength Direction Type



No correlation



Good correlation



Positive, linear

OLS assumptions

DEPENDENT VARIABLE

Information being modeled or predicted

EXPLANATORY VARIABLES

Data that explain the dependent variable values

RESIDUALS

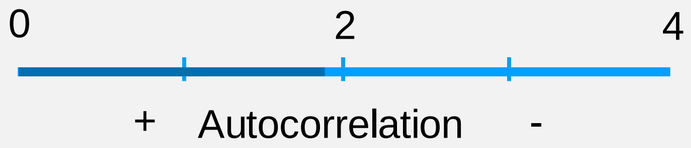
The unexplained in the dependent variable

- The data must be randomly sampled
- The expected value is a straight line function
- Must have negligible error in measurement
- The effects on the dependent variable are additive
- Should not show collinearity with one another
- The residuals have an expected sum of zero
- The residuals have homogeneous variance
- The residuals are normally distributed
- Adjacent residuals must not show autocorrelation

Model evaluation

MODEL STATISTICS

Explanatory variables
Coefficient
Standardized error
t-value
p-value
Standardized confidence intervals



y = 58,502 + 138popchangerate10 + 1,580f...
Prediction equation

R^2: 0.80456569
Measures model variablility

Adjusted R^2: 0.80401188
Predictor adjusted model variability

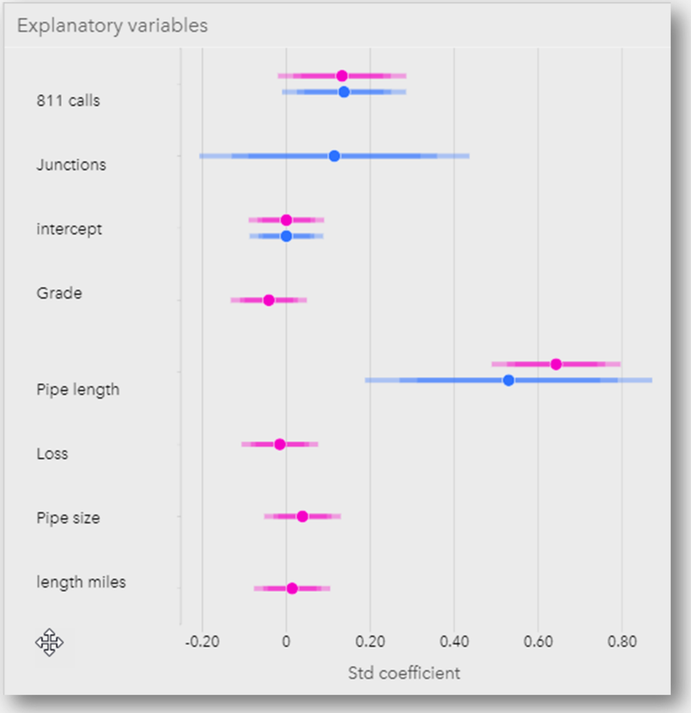
Durbin-Watson Test: 1.72817778
Describes model autocorrelation

p-Value: 0
Determines significance of the model

Residual standard error: 5,971.566 on 3176 df
Measures model fit

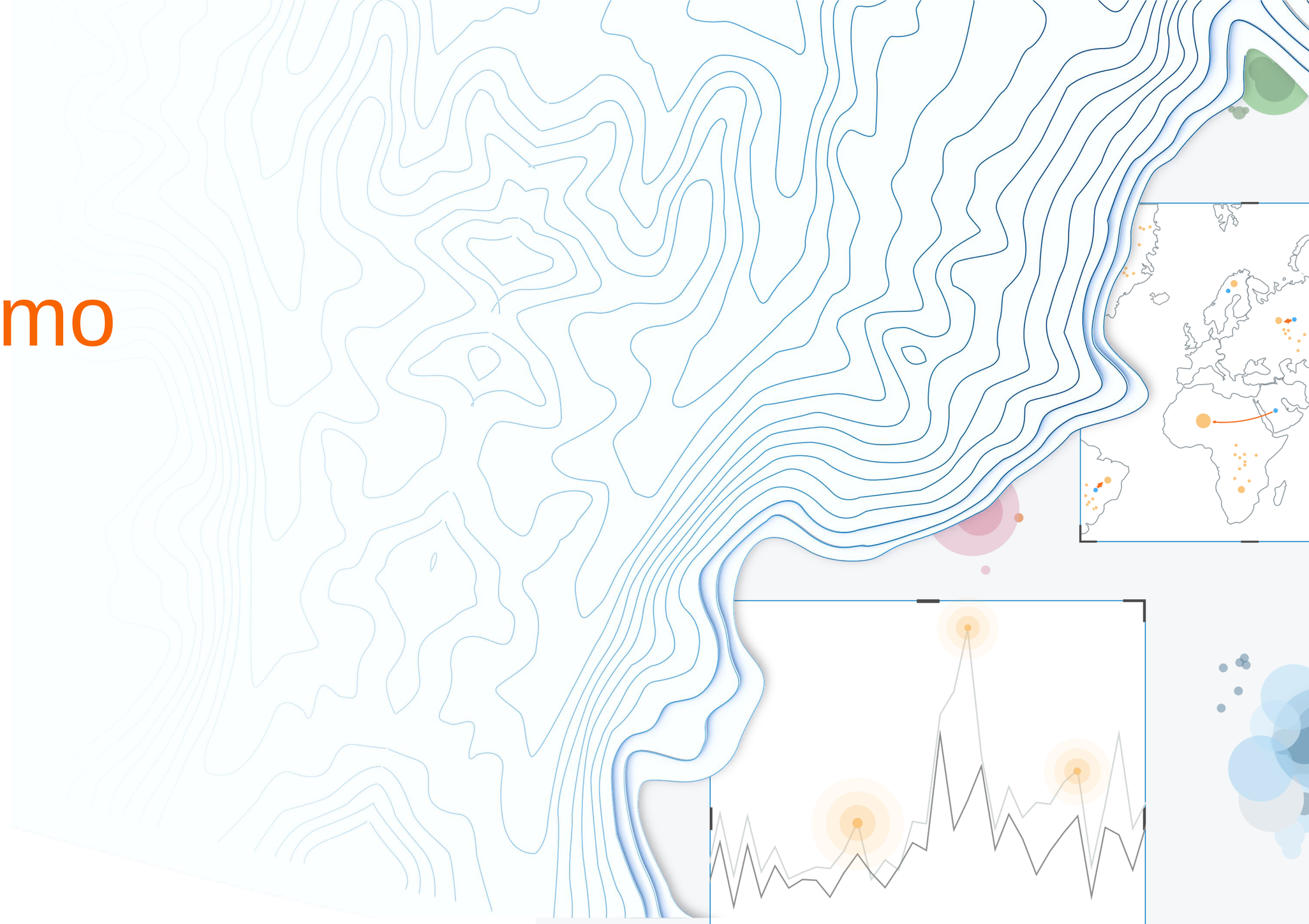
F statistic: 1,452.776 on 9 and 3176 df
Reference value for significance test

View confidence intervals



Demo

DEMO TITLE



Data science

Connect to your own Python and R kernels to extend analysis and visualization using both open source software platforms.

01

SETUP CONNECTION

Have complete control over your Python and/or R deployments

02

DATA ACCESS

Data can be passed from or to the Insights data pane

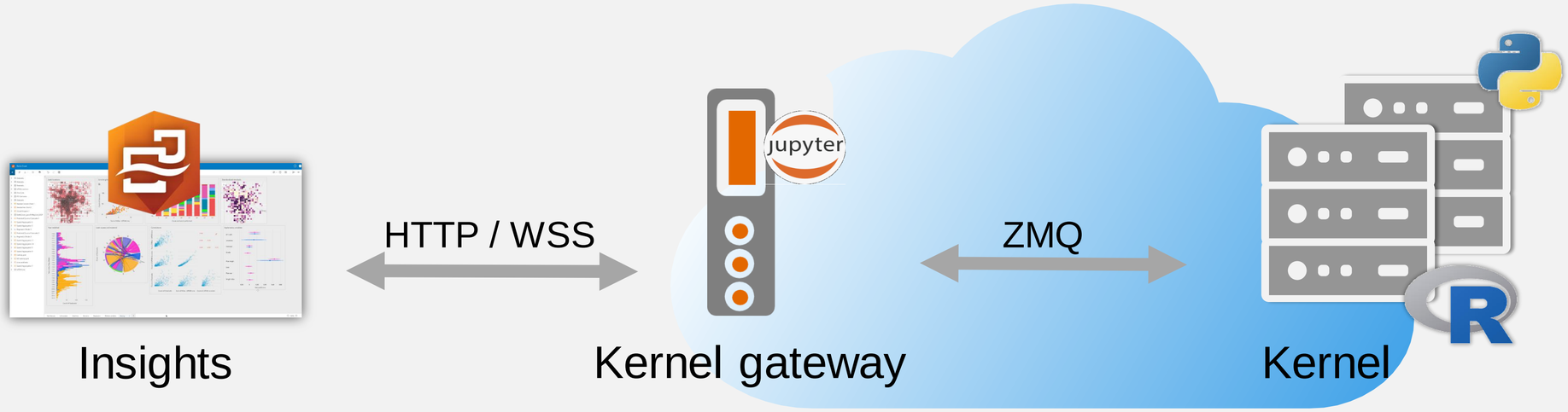
03

EXPAND VISUALS

Output visuals from Python or R can be added to Insights cards

Connecting to Python/R

Connect to your own Python and R kernels to extend analysis and visualization using both open source software platforms.



Extending analysis and visualizations

Extend your analysis using Python and/or R. Incorporate visualizations as cards. Manage your data.

Use the scripting editor to add scripts to models, save sessions and more…

TO COMPLETE

Sharing analysis

INTEGRATED WITH ANALYSIS

Share datasets

- workbooks
- pages (live/static)

Create packages

Use cross-filters

Change axis labels

Pop-out legends

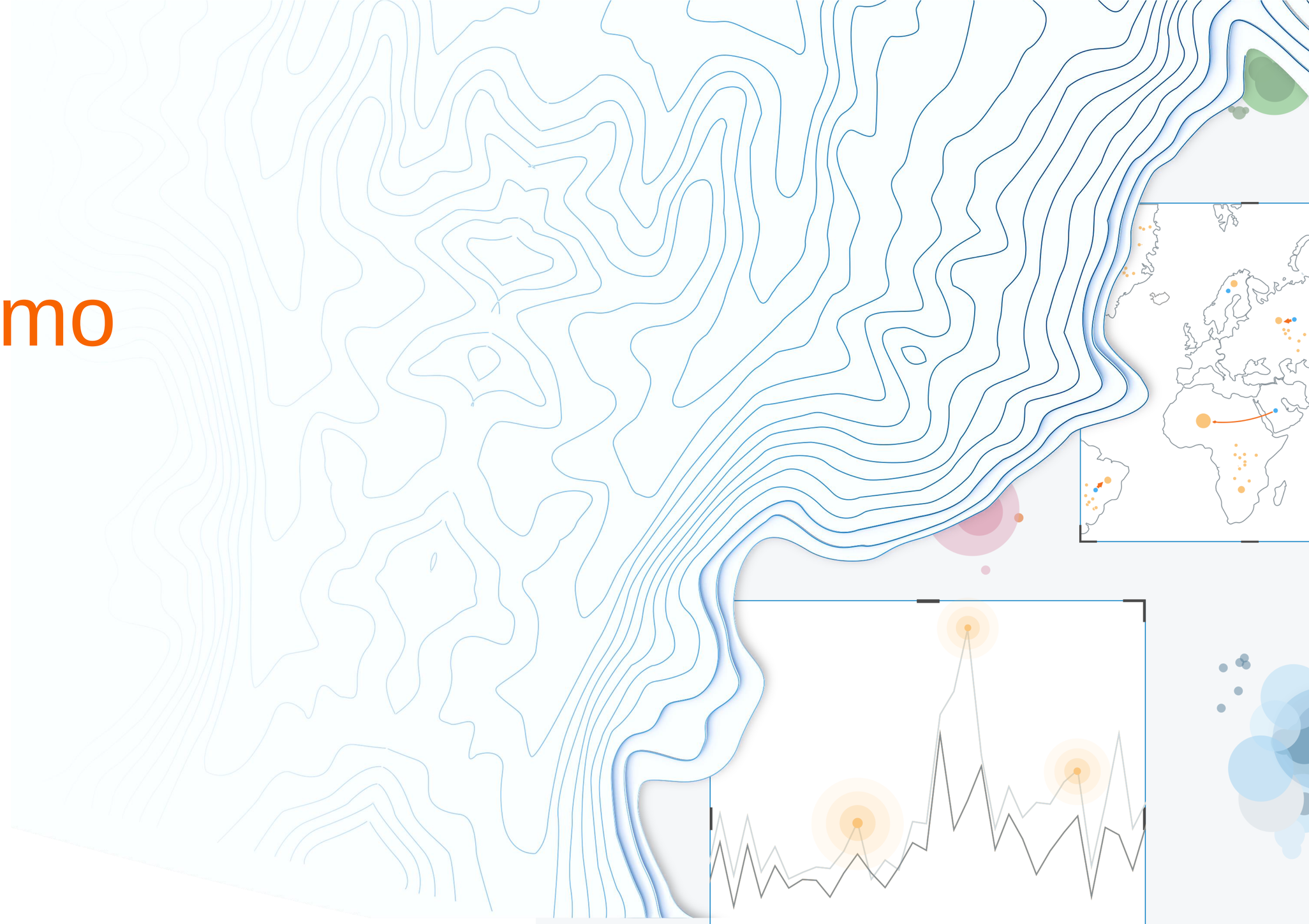
Foreground/background color

Add Filters

Include text/video/images

Demo

DEMO TITLE



Future plans

DATA

Auto update analysis
Packaging
SharePoint lists
OneDrive access
OS Authentication for SQL Server

ANALYSIS

Improvements to kernel density
Clustering
Labels on bar/column charts
Sync y-axis on combo charts