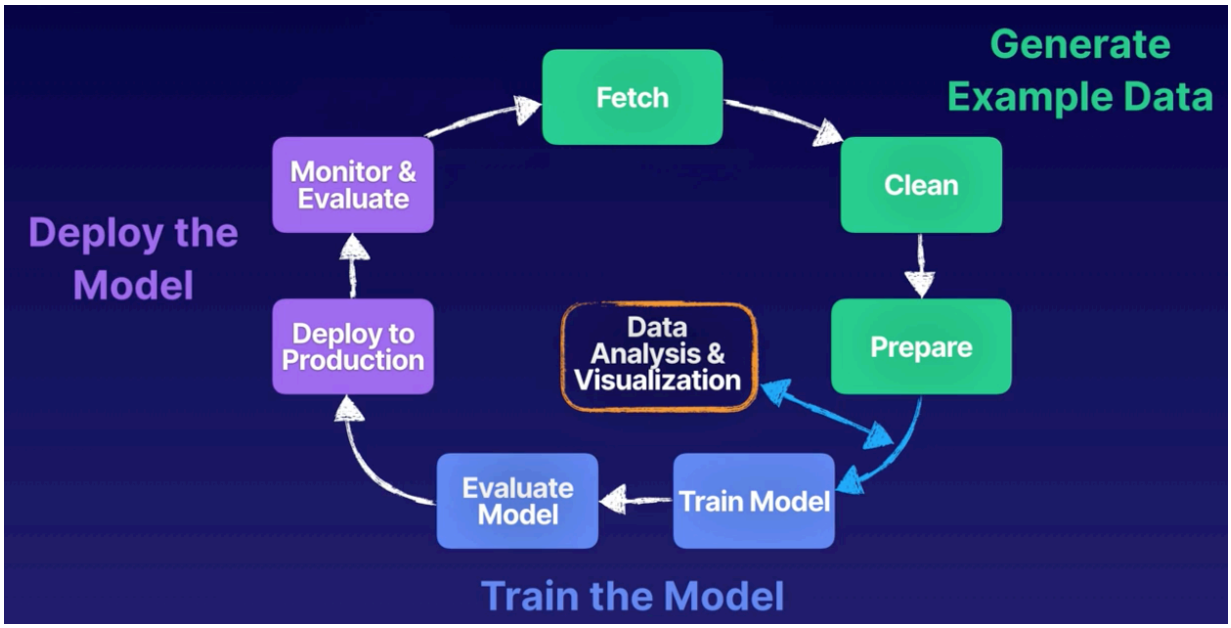
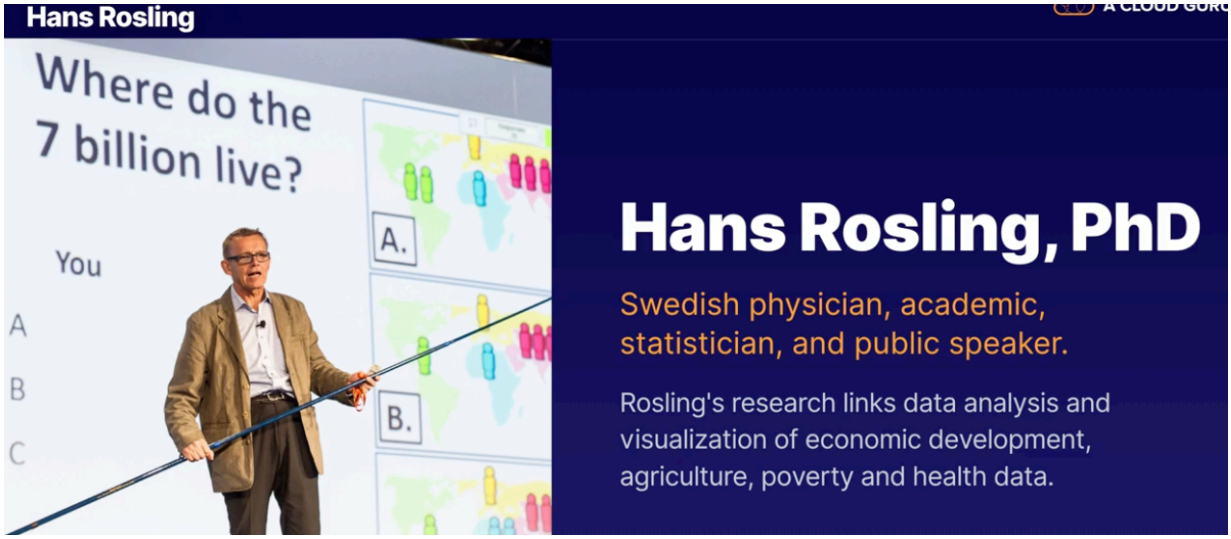
Cloud Guru - 4 - Data Analysis and Visualization

https://acloud.guru/course/aws-certified-machine-learning-specialty/learn/9ec6163d-ffe3-0975-5904-5d2b2d793493/c74480c5-9775-2d3d-4247-e3a5f3a2c21e/watch?backUrl=~2Fcourses&backUrl=~2Fcourses

This step is to make sure our data is ready for training our model.
This may help us identify the need for potentially more transformation or fetching more data



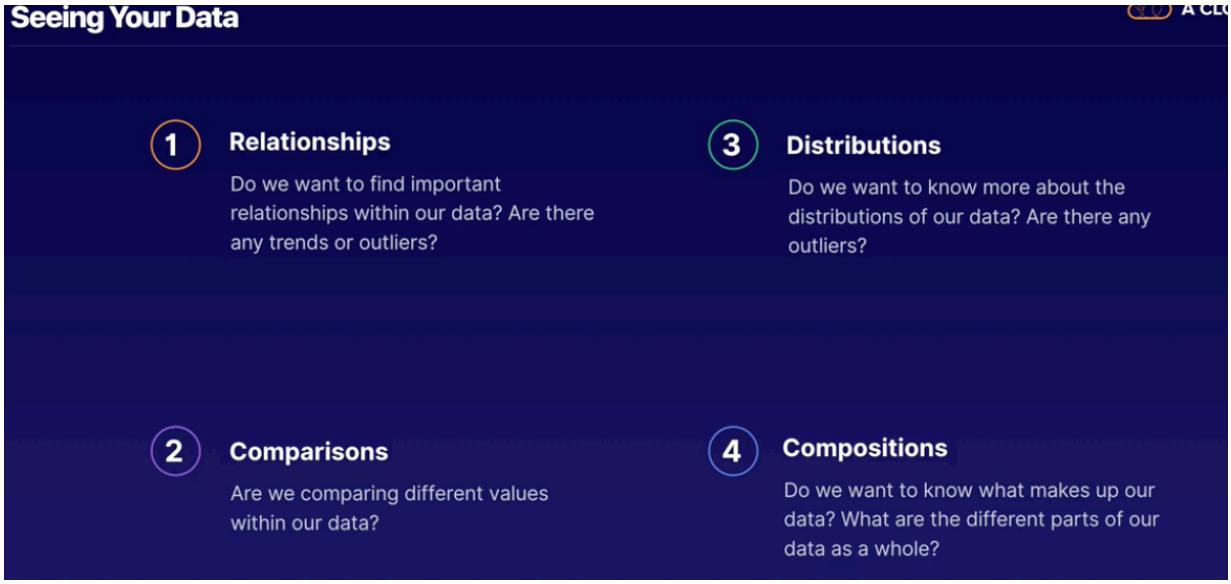


=> you can find out so much important information about the world with data

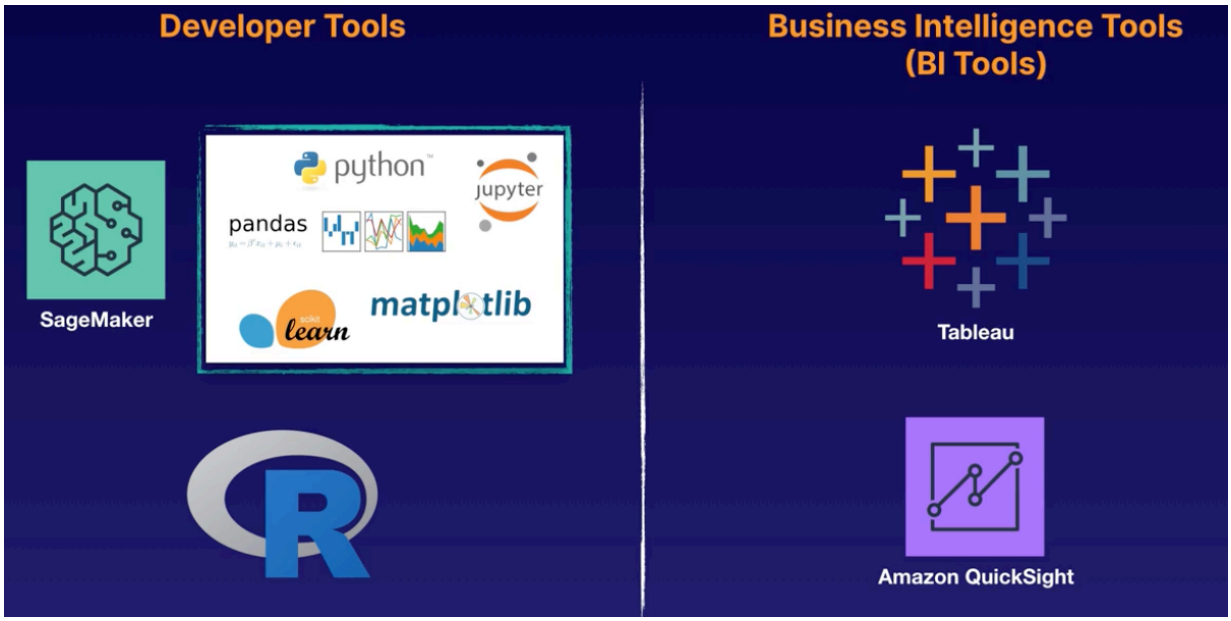
analysis and visualization

Him, his son and daughter created the below site to analyze world data with graphics:
GapMinder
https://www.gapminder.org/tools/

We may have questions that ask for best graph or visualization to help us analyze a specific type of data
4 types of data:

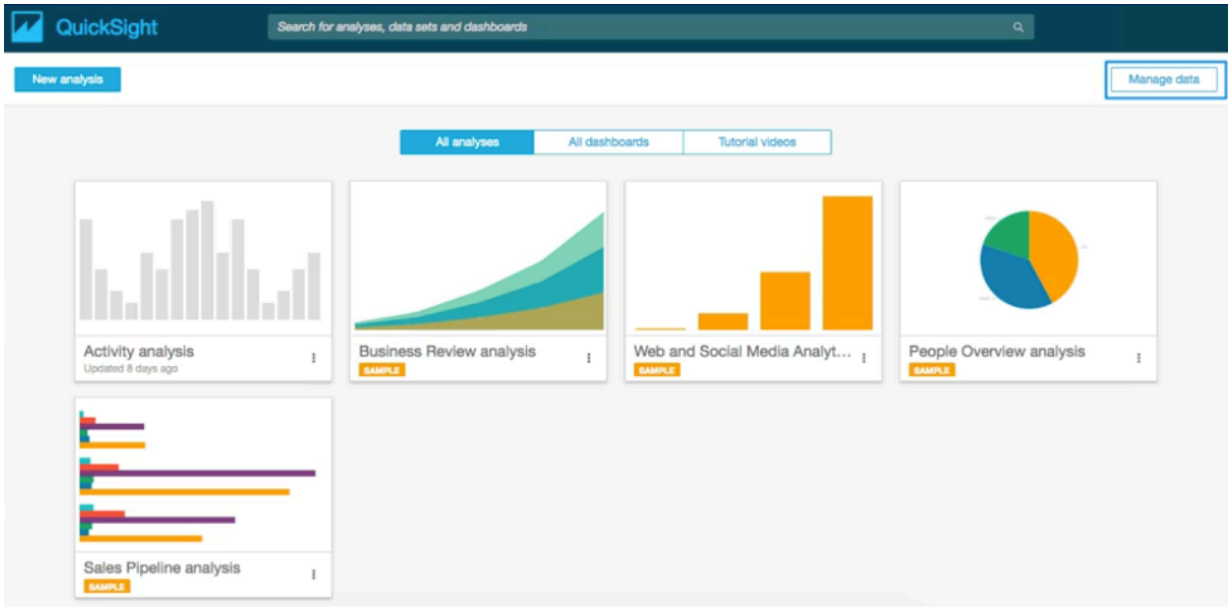


2 approaches:
- use Developer Tools
- use BI tools

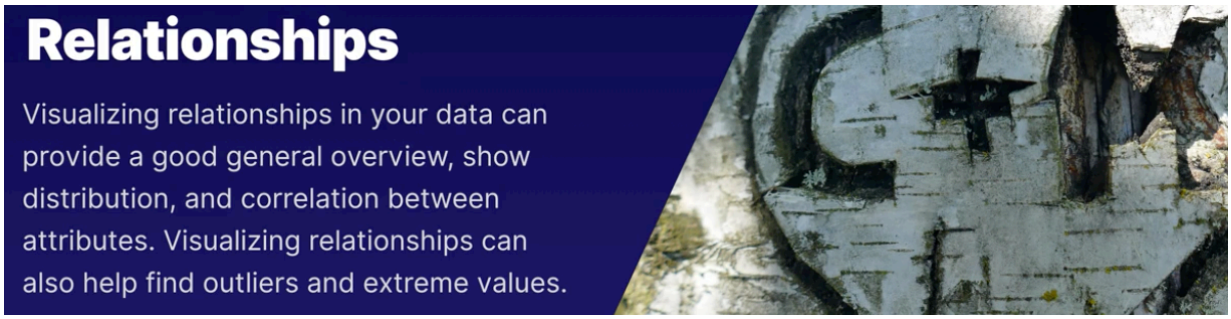


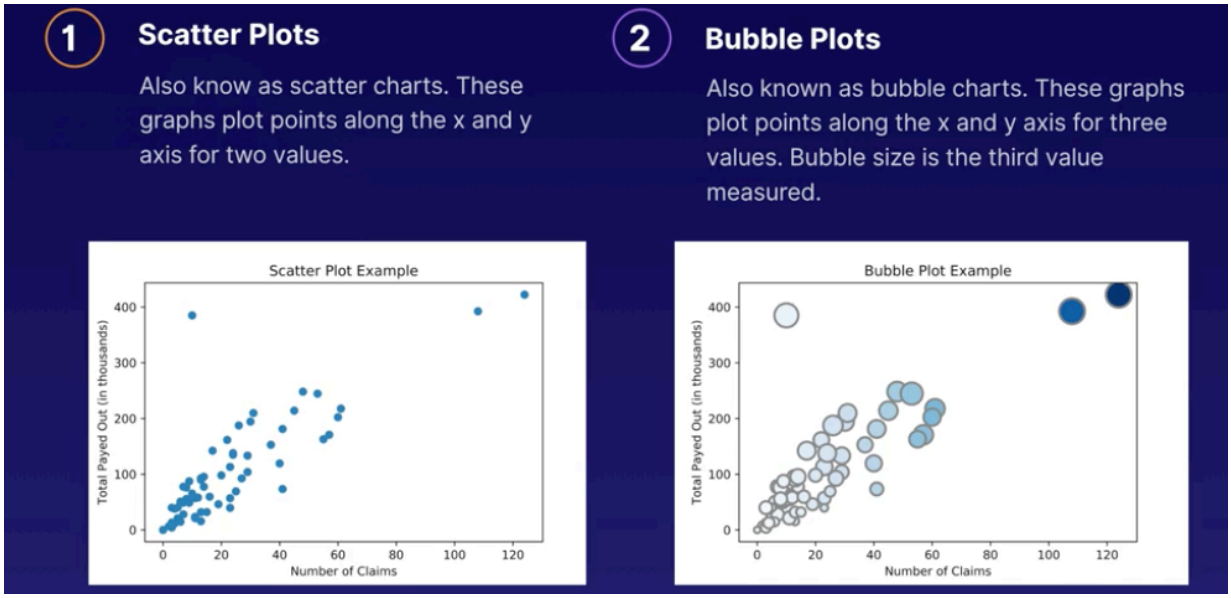
Amazon Quicksight

BI Tool that makes it easy to create visualization from our data, from the console



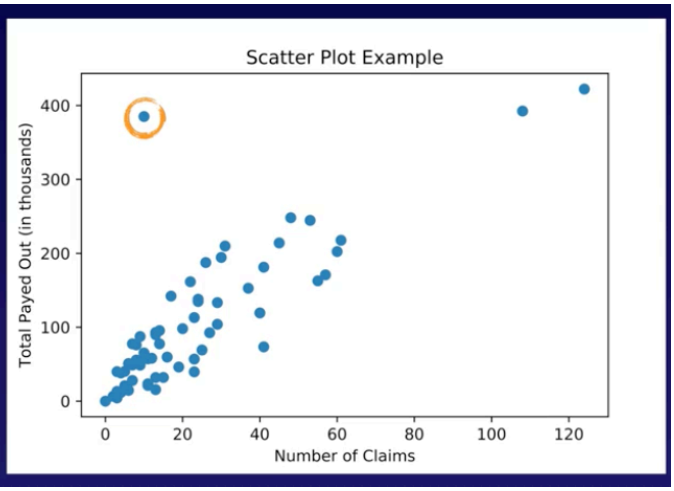
Relationships





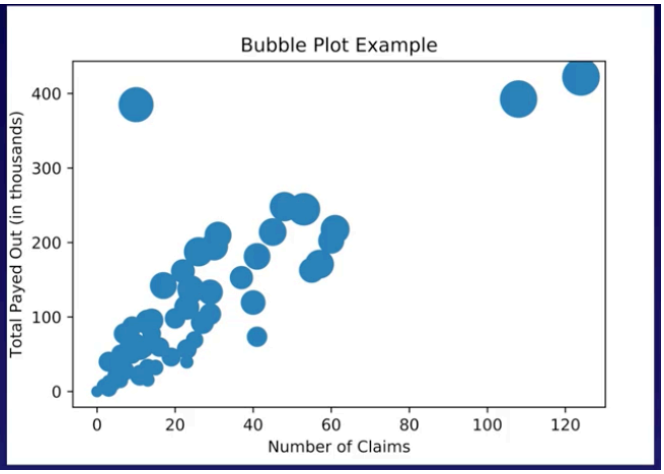
Example of scatter plot

num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...



And Bubble plot

num_of_claims	total_payment	days_processed
108	392.5	480
19	46.2	41
13	15.7	5
124	422.2	350
40	119.4	54
...	...	...

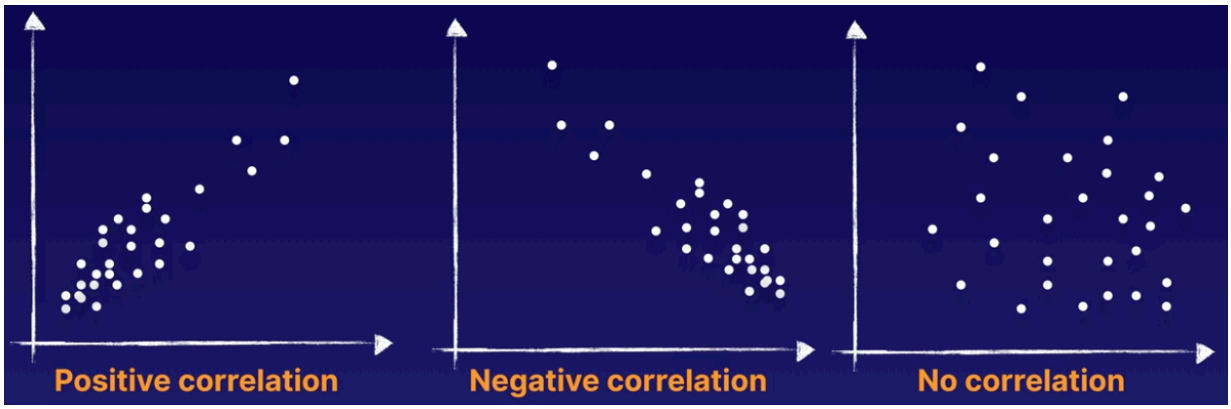


- **Scatter Plots**

  Example: Is there any relationship between the size of a home and the price?

- **Bubble Plots**

  Example: Are there any relationship between the size of a home, the age of the home, and the price?

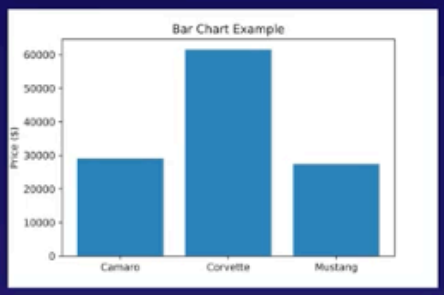


Comparisons

Use Bar charts or Line Charts

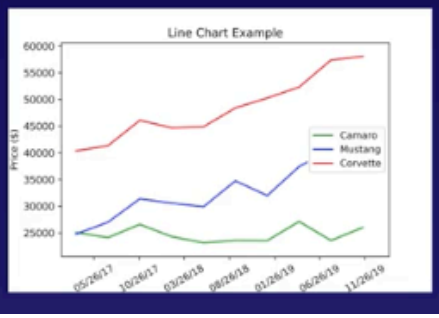
1 Bar Charts

Bar charts are graphs that use lines (bars) to mark single variable values. Provides a way to lookup and compare values.

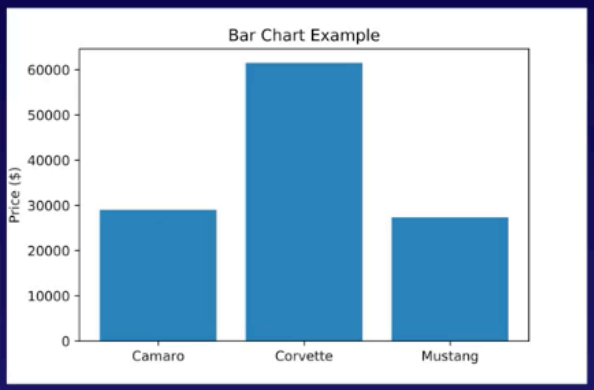


2 Line Charts

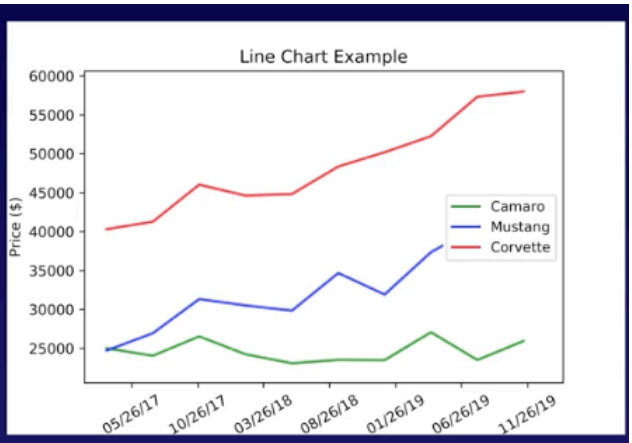
Line charts are graphs that use lines to show one or more variables changing over time.



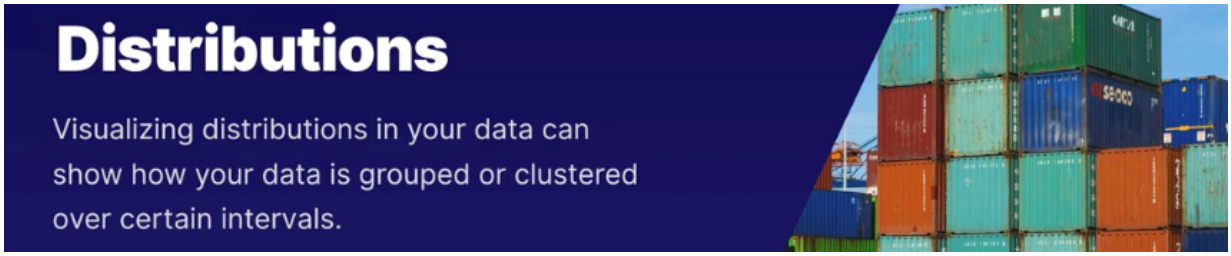
car	price
mustang	26395
corvette	59600
mustang	25900
camaro	25125
corvette	54988
camaro	23680
...	...

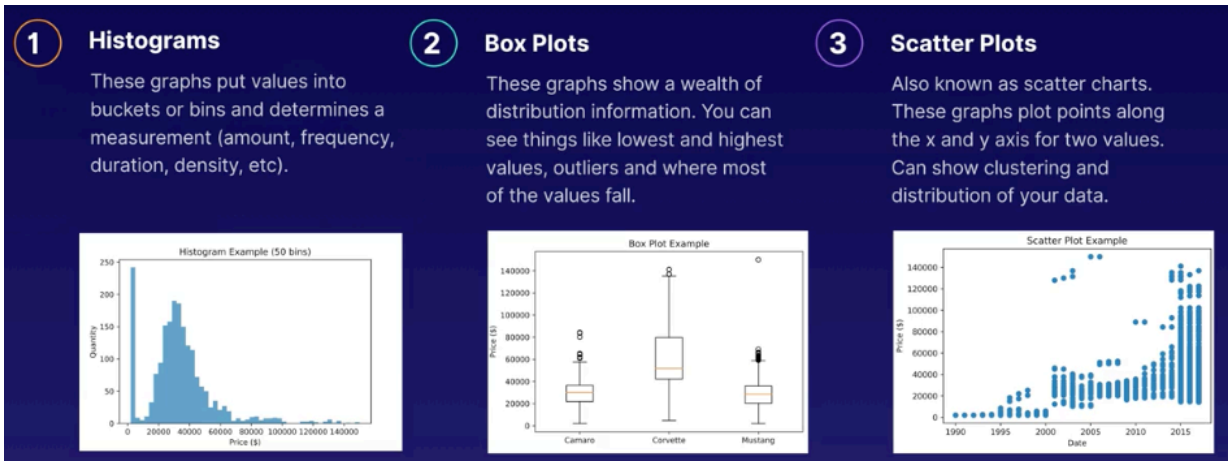


car	price	date
mustang	26395	02/2018
corvette	42500	03/2018
mustang	25900	01/2017
camaro	25125	10/2018
corvette	41700	04/2017
camaro	23680	01/2019
...	...	...

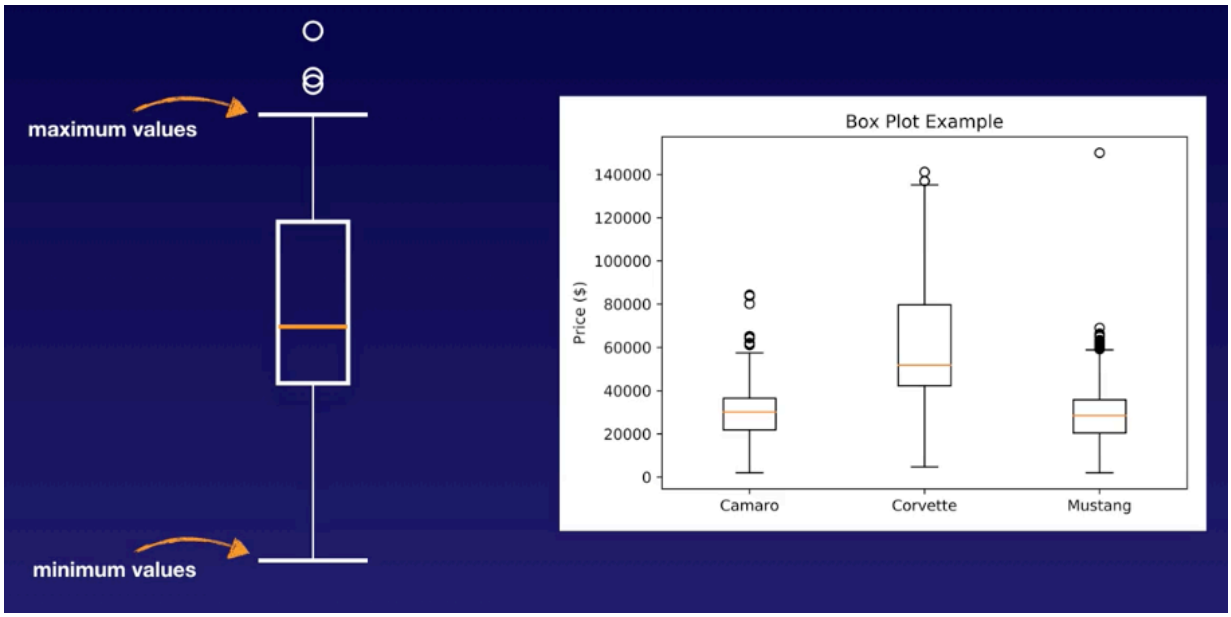


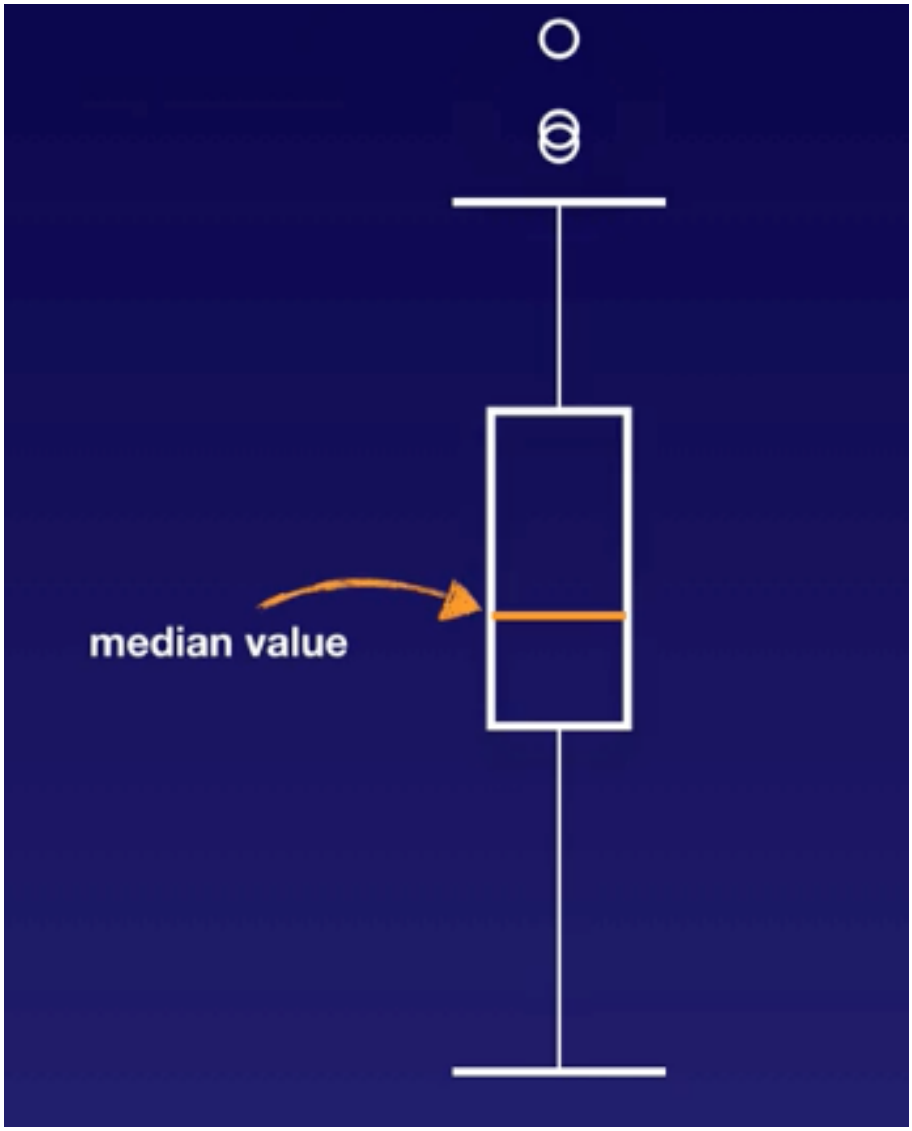
Distributions

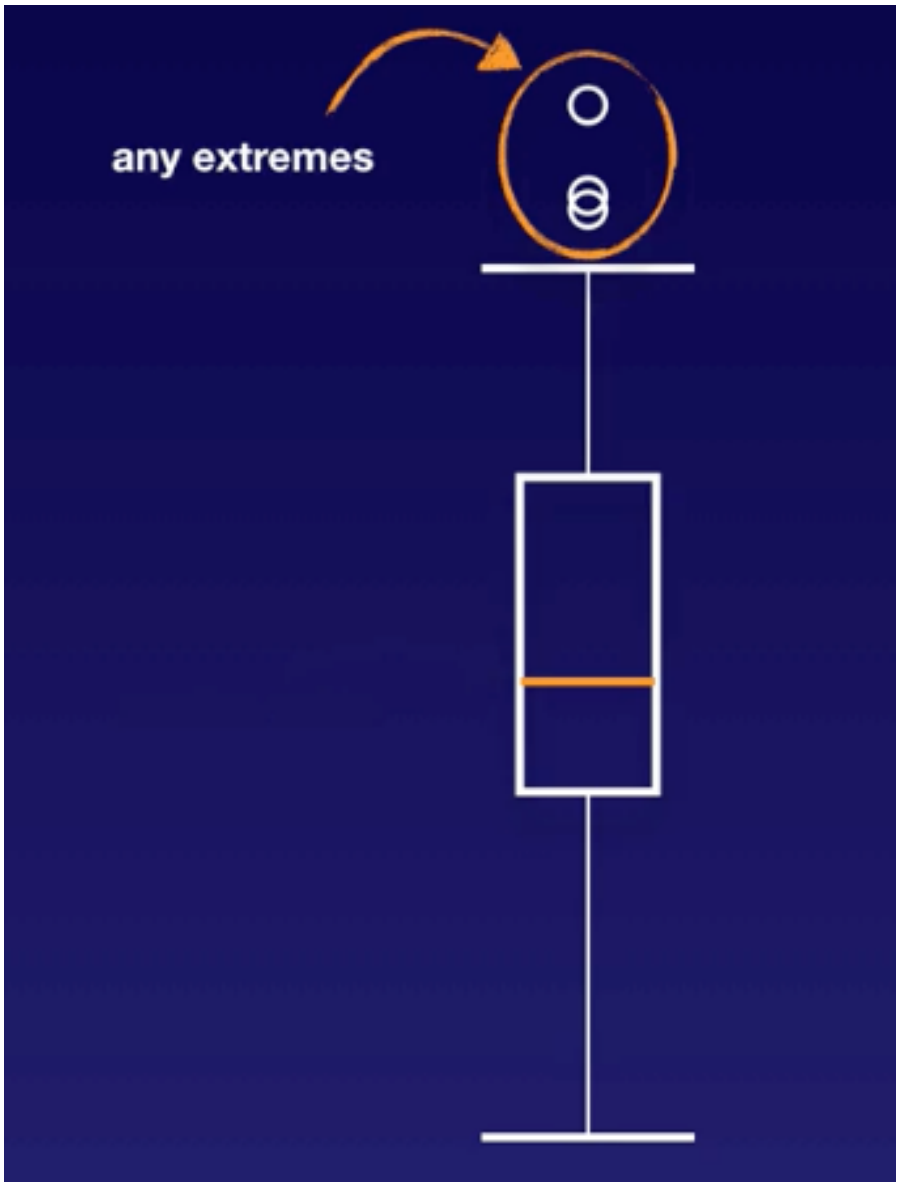


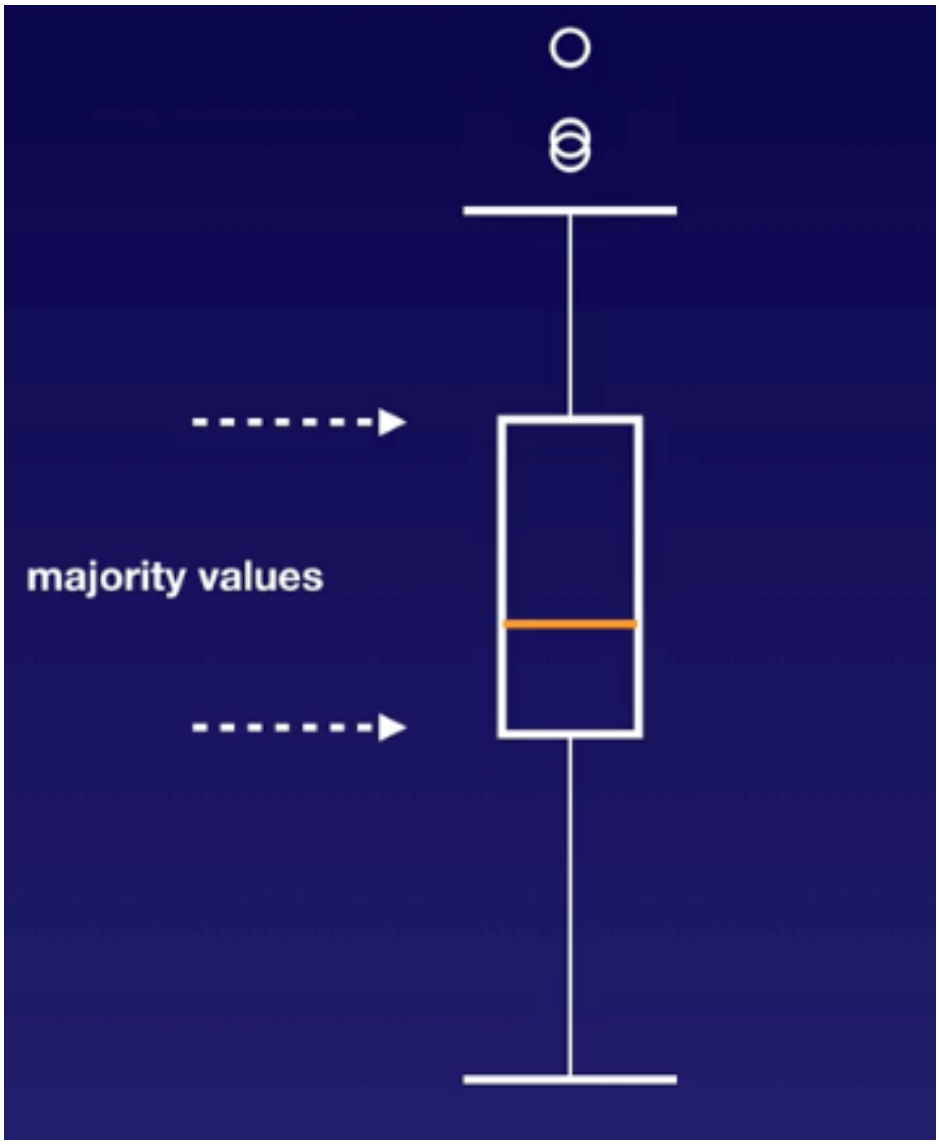


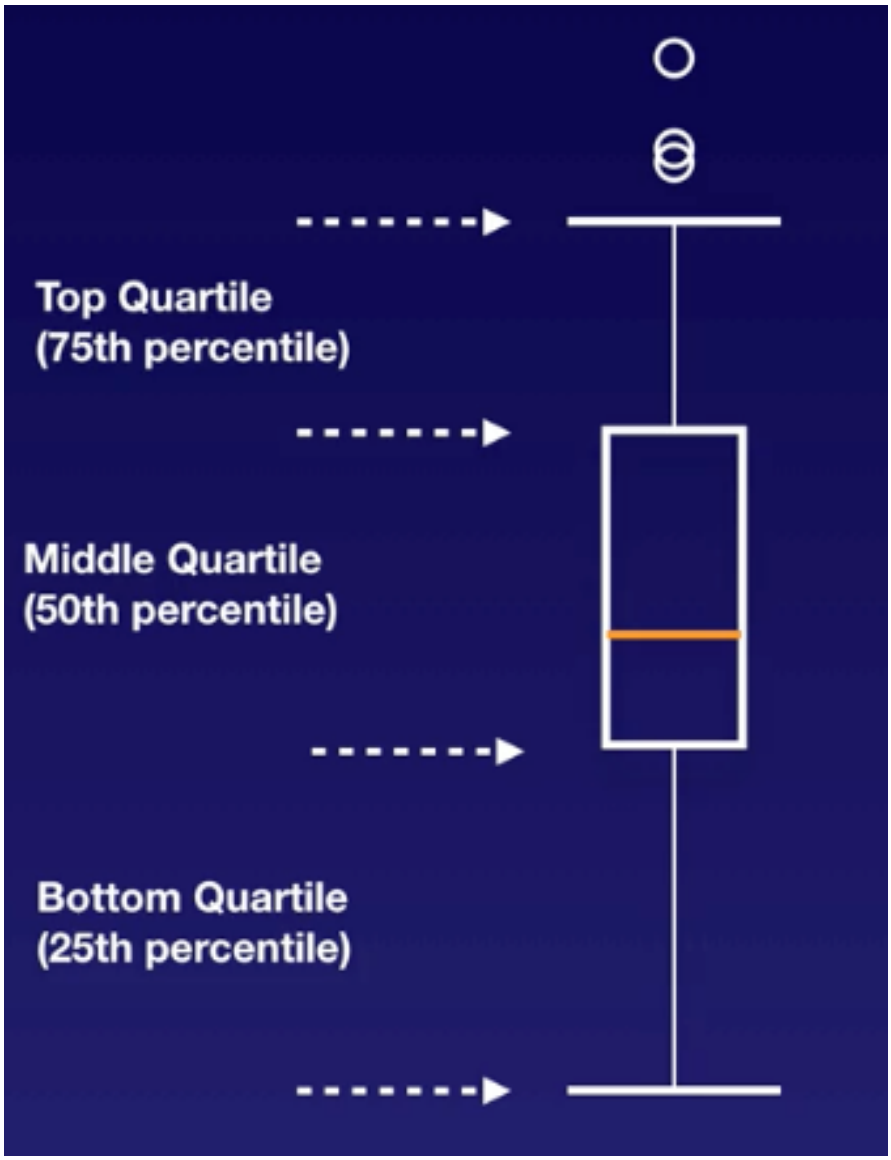
Box plots:
Show a lot of info into a single visualization
- min and max values
- median value
- extreme values
- majority of values (50% or more)
- 25/75 quartile


maximum values
minimum values
Box Plot Example
Price ($)
140000
120000
100000
80000
60000
40000
20000
0
Camaro
Corvette
Mustang


median value


any extremes


majority values



- **Histograms & Box Plots**

  Example: Showing the distribution of test scores for a given exam.

- **Scatter Charts**

  Example: Showing the return of investment (ROI) for the amount of money spent and the total time invested.

Data Compositions

Compositions

Visualizing composition of your data show the various elements and what your data is made of.

1 Pie Charts

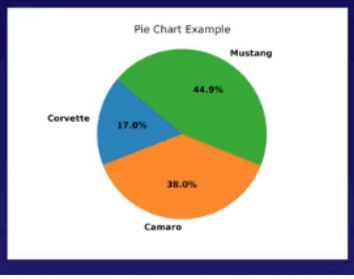
Pie charts show how various values compare as a whole, share of the total.

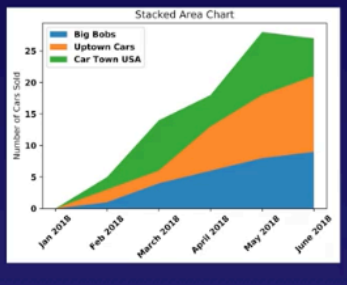
2 Stacked Area Charts

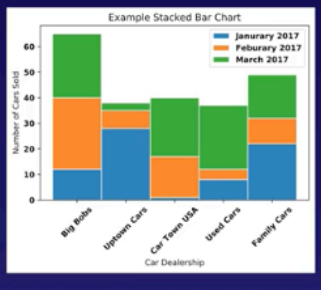
Stacked area charts shows the measurement of various items over longer periods of time.

3 Stacked Column Charts

Also known as stacked bar charts. These graph show quantity of various items over shorter periods of time.







- **Pie Charts**

  Example: Showing the sales figures for each region.

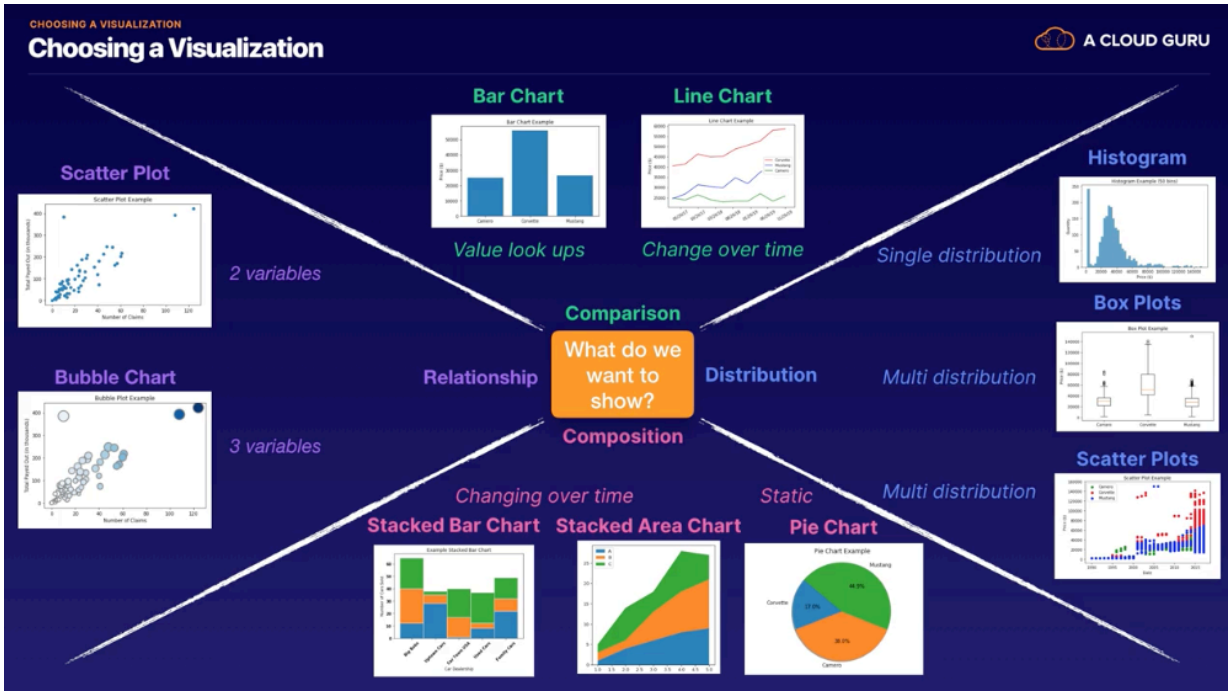
- **Stacked Area Charts**

  Example: Showing the number of products sold by different departments on a weekly basis.

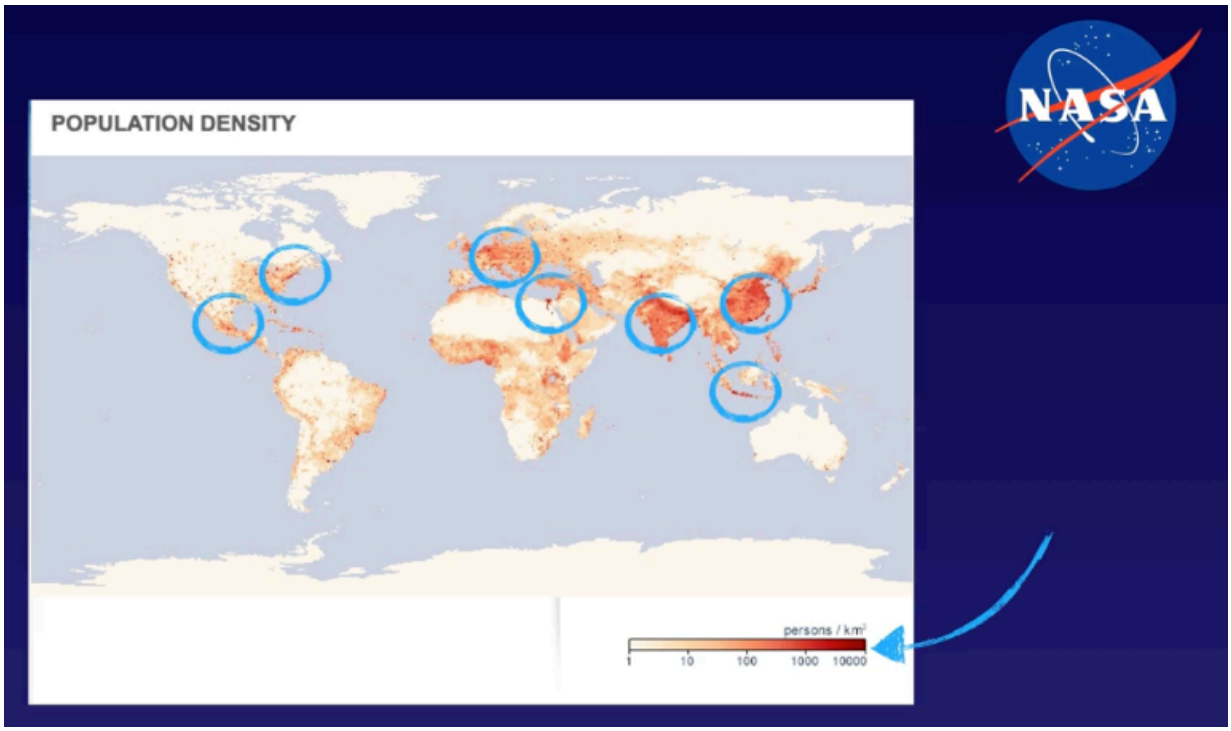
- **Stacked Bar Charts**

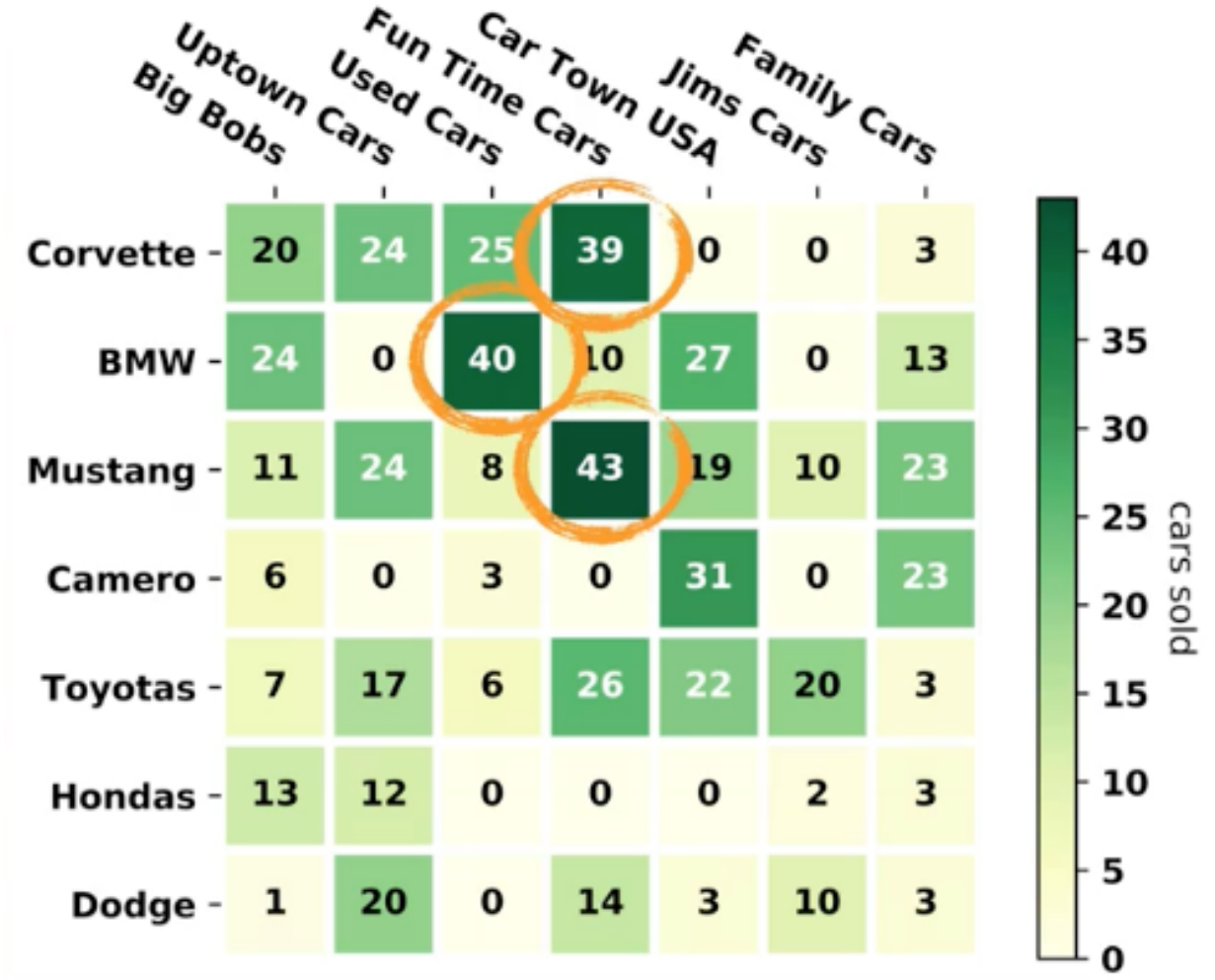
  Example: Showing the quarterly revenue totals for each region.

Choosing a visualization

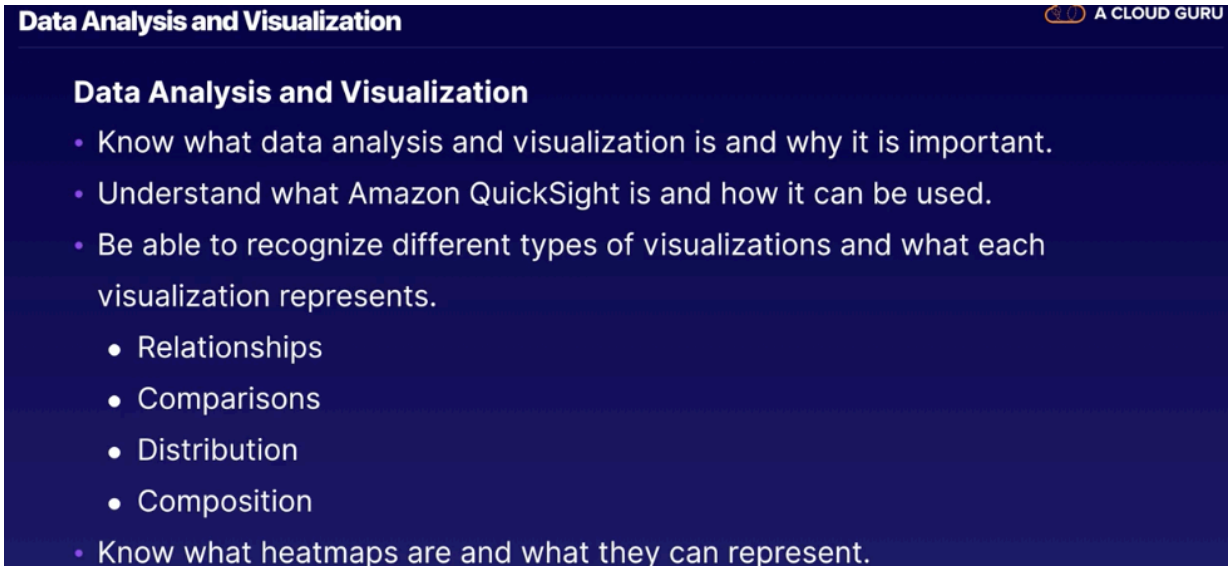


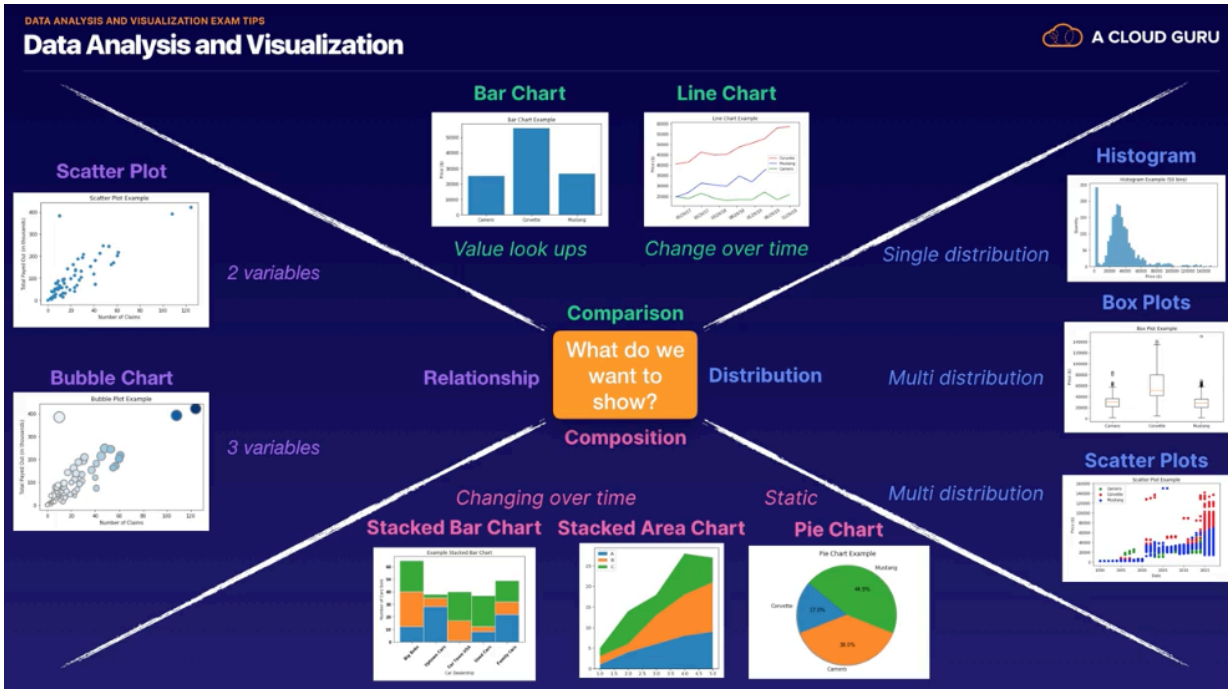
Another type: heat map


POPULATION DENSITY
NASA
persons / km²
1
10
100
1000
10000


Big Bobs
Uptown Cars
Used Cars
Fun Time Cars
Car Town USA
Jims Cars
Family Cars
Corvette 20 24 25 39 0 0 3
BMW 24 0 40 10 27 0 13
Mustang 11 24 8 43 19 10 23
Camero 6 0 3 0 31 0 23
Toyotas 7 17 6 26 22 20 3
Hondas 13 12 0 0 0 2 3
Dodge 1 20 0 14 3 10 3
cars sold
40
35
30
25
20
15
10
5
0

Exam tips

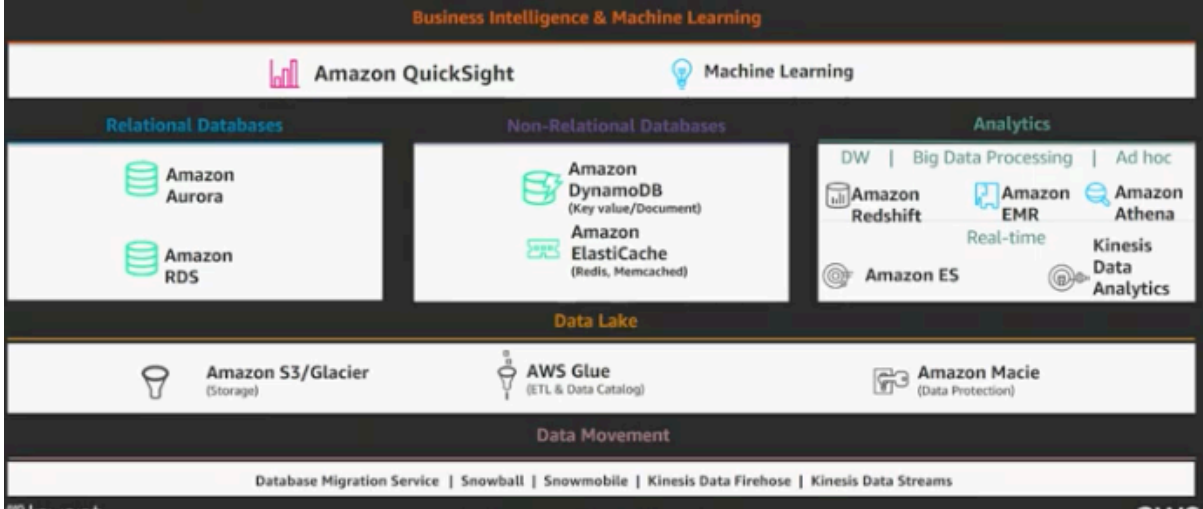




ML Insights AWS Insights:
https://www.youtube.com/watch?v=FhI1kVABMF0

One platform for data and analytics

Broadest and deepest portfolio of purpose-built services



One BI service for ALL of your users

Amazon QuickSight covers all of your users from casual data consumers, to dashboard creators, to power users and analysts that need self-serve analytics

Data Scientist (Author)

Give power users and analysts the freedom to do their own self-serve data discovery and analysis on governed data you control

Dashboard Creator (Author)

Create and publish rich, interactive dashboards to all of your users

End User (Reader)

With the new Reader Role, you can provide everyone in your organization secure, easy access to interactive dashboards and reports, on any device

Why Amazon QuickSight?

No servers to manage

Amazon QuickSight has no servers or software to manage, maintain, deploy, upgrade or migrate. We do the heavy lifting so you don't have to.

Native AWS integration

Amazon QuickSight securely integrates with your data sources and AWS services like Amazon Simple Storage Service (Amazon S3), Redshift, Amazon Athena, Amazon Aurora, Amazon Relational Database Service (Amazon RDS), AWS Identity and Access Management (IAM), AWS CloudTrail, Amazon Cloud Directory and more - providing you with everything you need to build an end-to-end BI solution.

Easily scale from 10 users to 10,000

QuickSight automatically scales with your usage and activity, with no need for additional infrastructure. QuickSight will grow with your organization's needs from a few users to tens of thousands of users.

Pay only for what you use

Provide read-only access to interactive dashboards and pay only when your users access them with Pay-per-Session pricing. With Amazon QuickSight there are no upfront costs, no annual commitments and no charges for inactive users.

Three powerful new features

Anomaly Detection

Uncover hidden insights by continuously analyzing across billions of data points.

ML Forecasting

Predict key business metrics with point and click simplicity.

Auto Narratives

Tell the story of your data in plain language narratives.

ML-powered anomaly detection

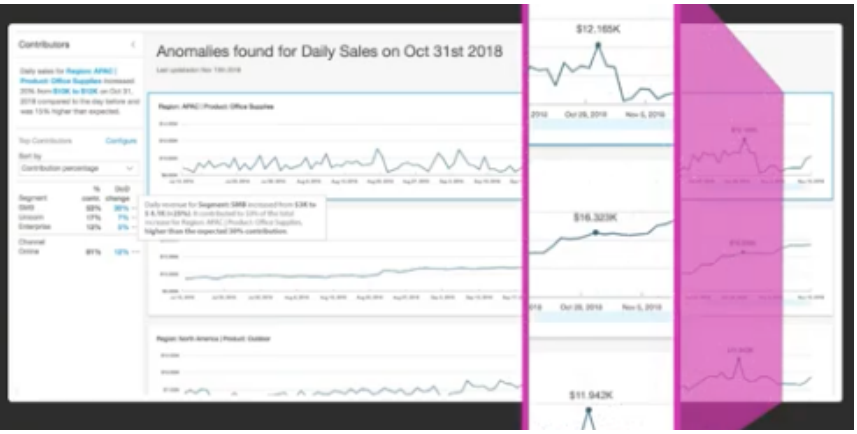
Challenges

Hidden in aggregates
Data blind spots
Slow to react to business incidents
Manual analysis doesn't scale
Not for BI

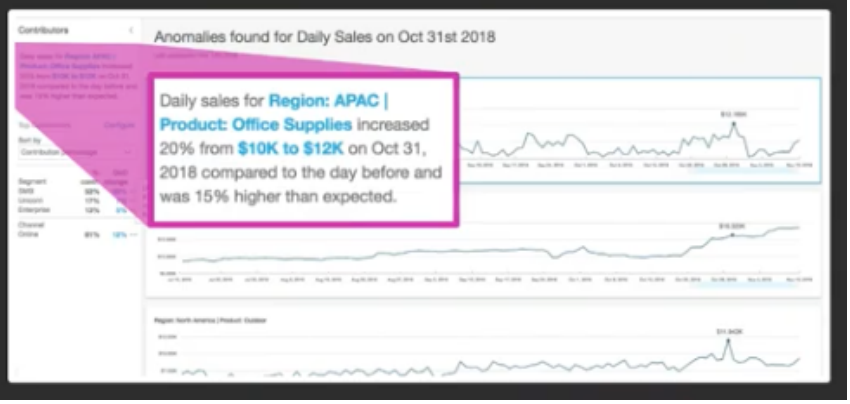
Our solution

Analyze millions of metrics
Continuously run in background
What happened AND why
Alerts delivered via email
No ML expertise needed

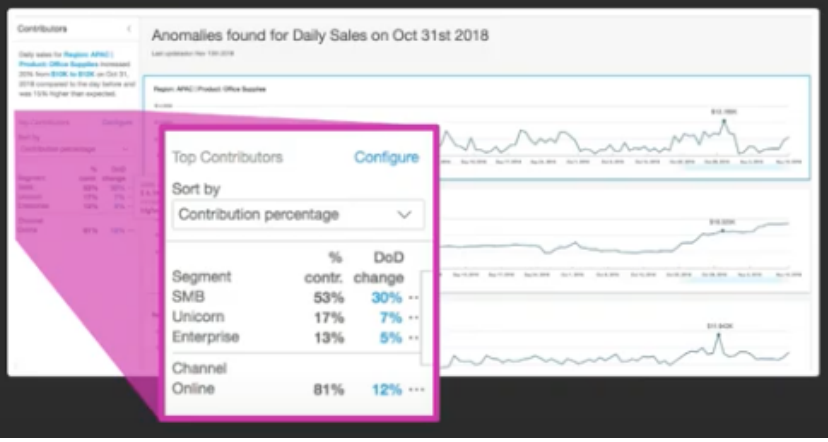
Discover all the hidden trends and anomalies



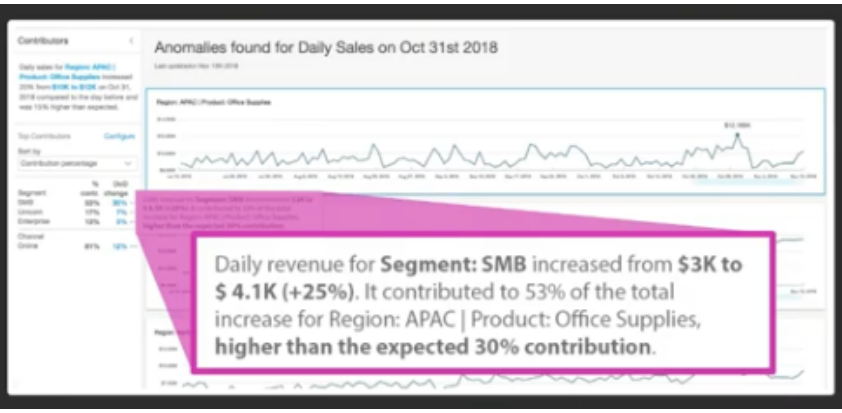
"Sales for Office Supplies in APAC was 15% above expected"



"SMB Segment was the top contributor."



"It's significant because SMB typically only accounts for 30% of sales."



ML-powered forecasting

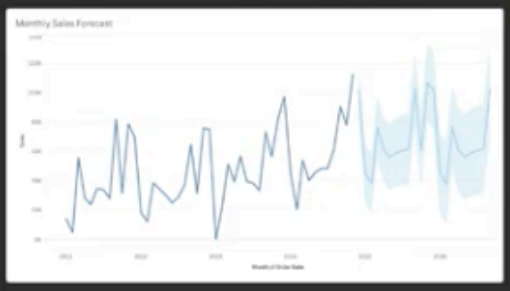
Challenges

Spreadsheet are complex to maintain
Doesn't learn, doesn't scale
How do you deal with outliers?
Hard to deal with changes
Simple things are hard to do
BI forecasting as too basic

Our solution

Point and click forecasting
Continuously learns from data
Automatically excludes outliers
Imputes missing values
Detects seasonality and trend
No ML or Excel needed

Traditional BI forecasting



Captures only seasonality

Missing upward trend

Confidence band influenced by bad data

QuickSight ML-powered forecasting

Captures seasonality and upward trends

Automatically excludes bad data

High confidence band

Auto Narrative

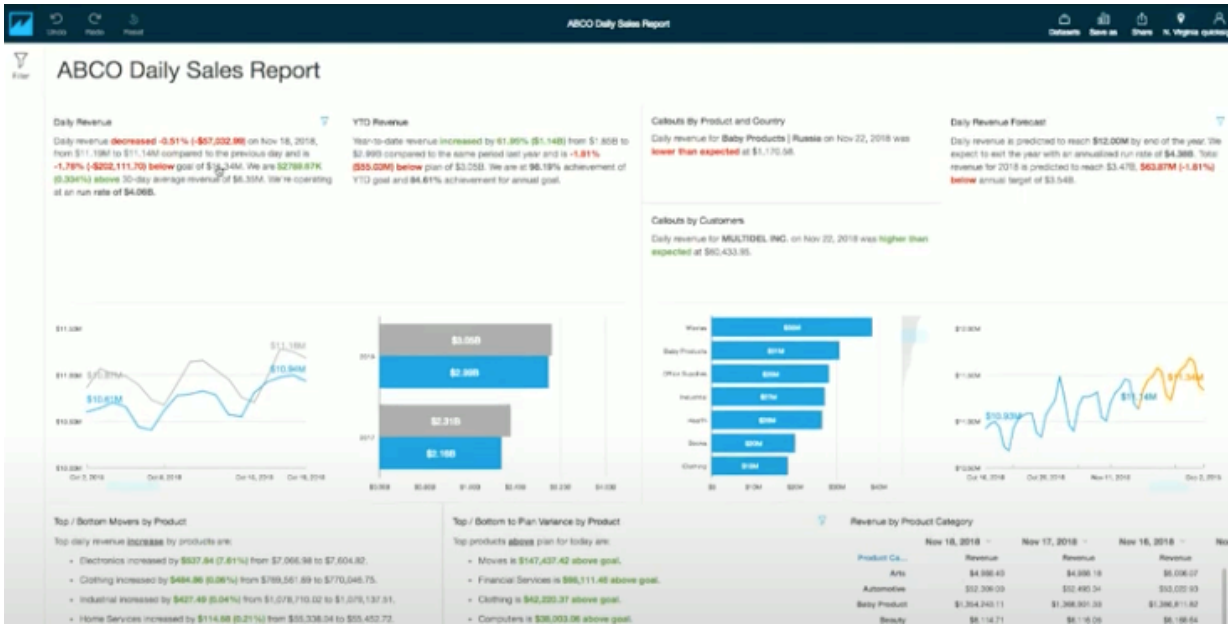
Challenges

What is the data trying to say?
Hours in front of dashboards
Sifting through charts and tables
Deciphering rows and columns of data
Drilling in and slicing and dicing
Time consuming, not scalable!

Our solution

Insights in plain language
Embedded in your dashboard
Fully customizable templates
Easy to get started
No coding necessary
Looking beyond visualizations


ABCO Daily Sales Report
Daily Revenue
Daily revenue decreased -0.51% (-$57,032.90) on Nov 18, 2018, from $11.19M to $11.14M compared to the previous day and is -1.78% (-$202,111.70) below goal of $11.34M. We are $2789.67K (0.334%) above 30-day average revenue of $8.35M. We're operating at an run rate of $4.06B.
Callouts By Product and Country
Daily Revenue Forecast
Top / Bottom Movers by Product
Top / Bottom to Plan Variance by Product
Revenue by Product Category


ABCO Daily Sales Report
Daily Revenue
YTD Revenue
Callouts By Product and Country
Callouts by Customers
Daily Revenue Forecast
Top / Bottom Movers by Product
Top / Bottom to Plan Variance by Product
Revenue by Product Category