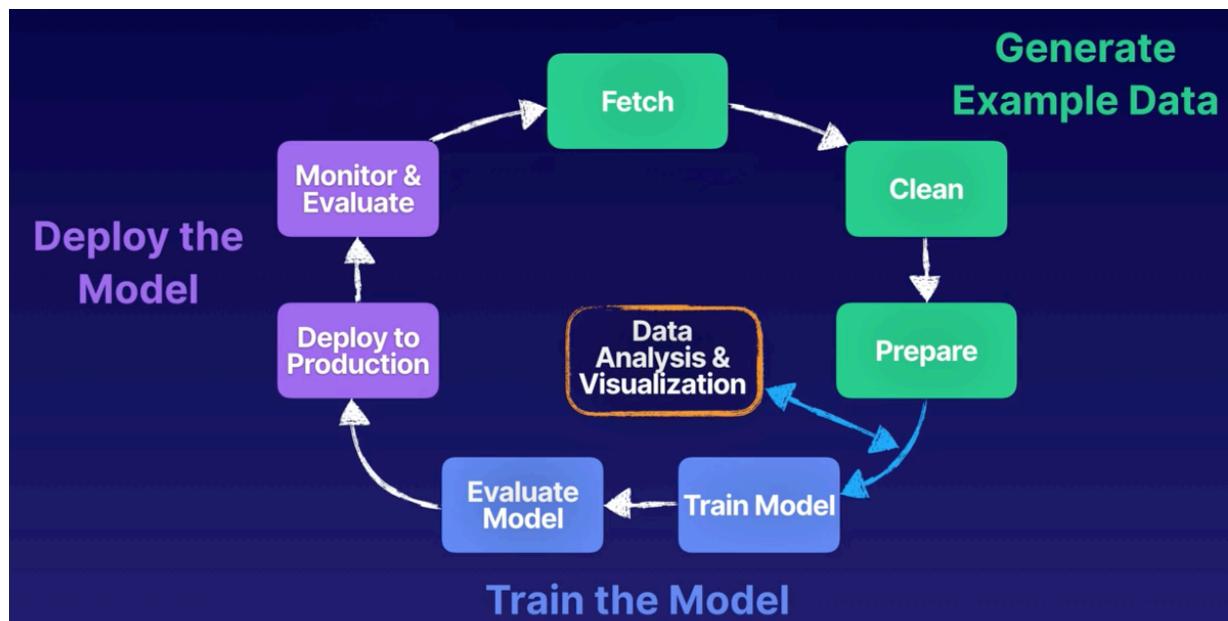


## Cloud Guru - 4 - Data Analysis and Visualization

<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



**Hans Rosling**

Where do the 7 billion live?

A. You  
B. A  
C. B.

**Hans Rosling, PhD**

Swedish physician, academic, statistician, and public speaker.

Rosling's research links data analysis and visualization of economic development, agriculture, poverty and health data.

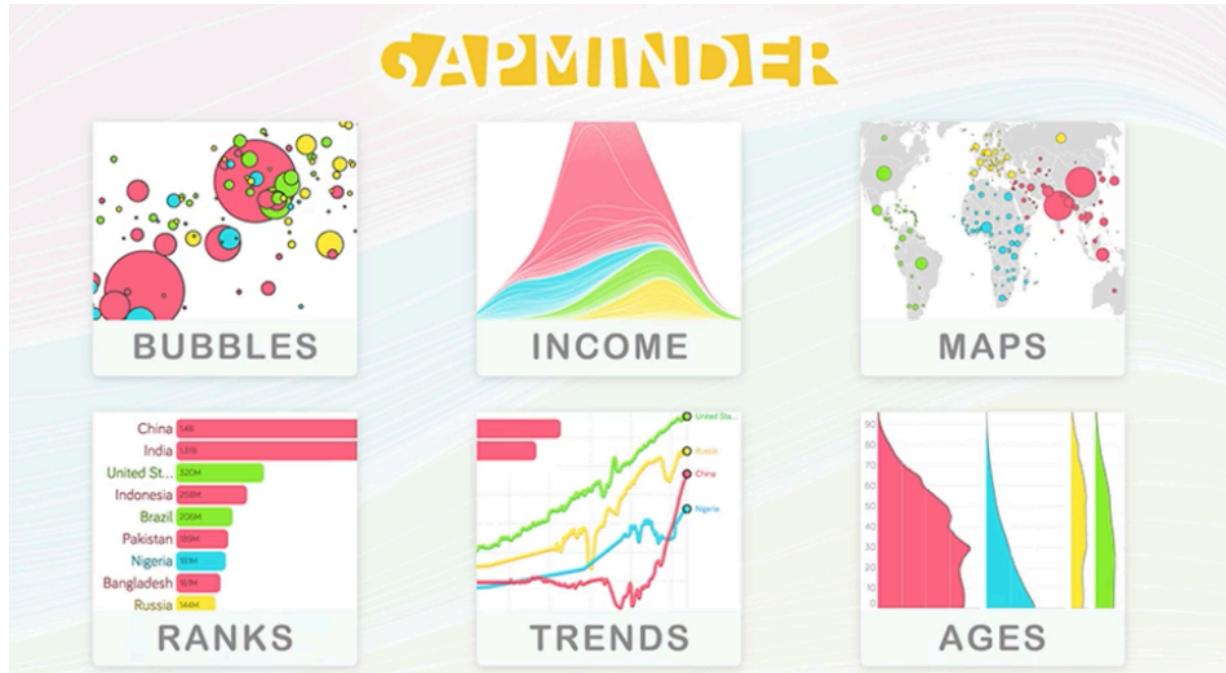
=> you can find out so much important information about the world with data

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:

### 1 Relationships

Do we want to find important relationships within our data? Are there any trends or outliers?

### 3 Distributions

Do we want to know more about the distributions of our data? Are there any outliers?

### 2 Comparisons

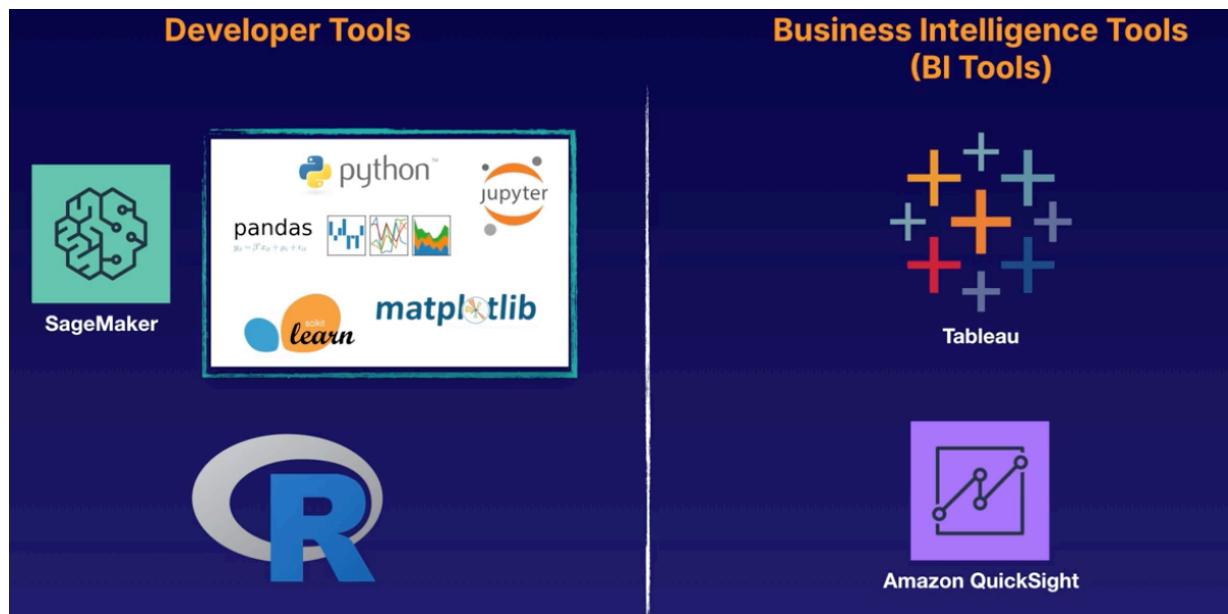
Are we comparing different values within our data?

### 4 Compositions

Do we want to know what makes up our data? What are the different parts of our data as a whole?

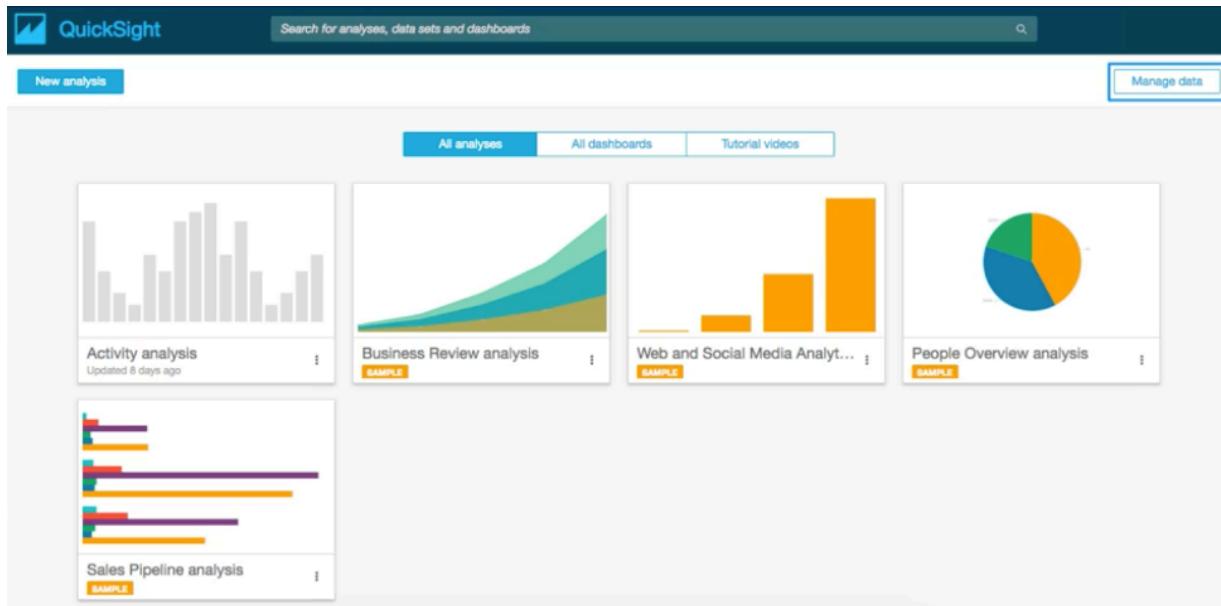
2 approaches:

- use Developer Tools
- use BI tools



### Amazon Quicksight

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



## Relationships

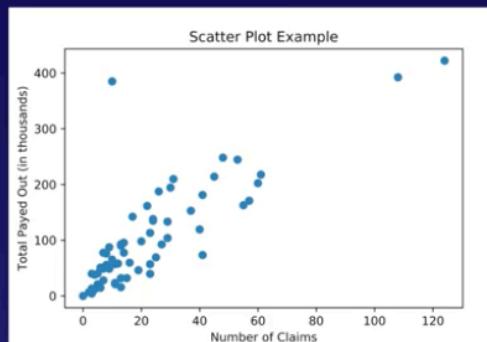
# Relationships

Visualizing relationships in your data can provide a good general overview, show distribution, and correlation between attributes. Visualizing relationships can also help find outliers and extreme values.



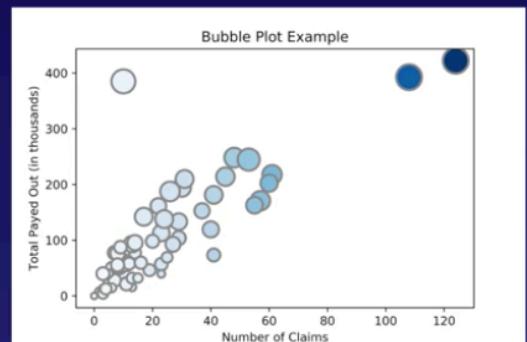
## 1 Scatter Plots

Also known as scatter charts. These graphs plot points along the x and y axis for two values.

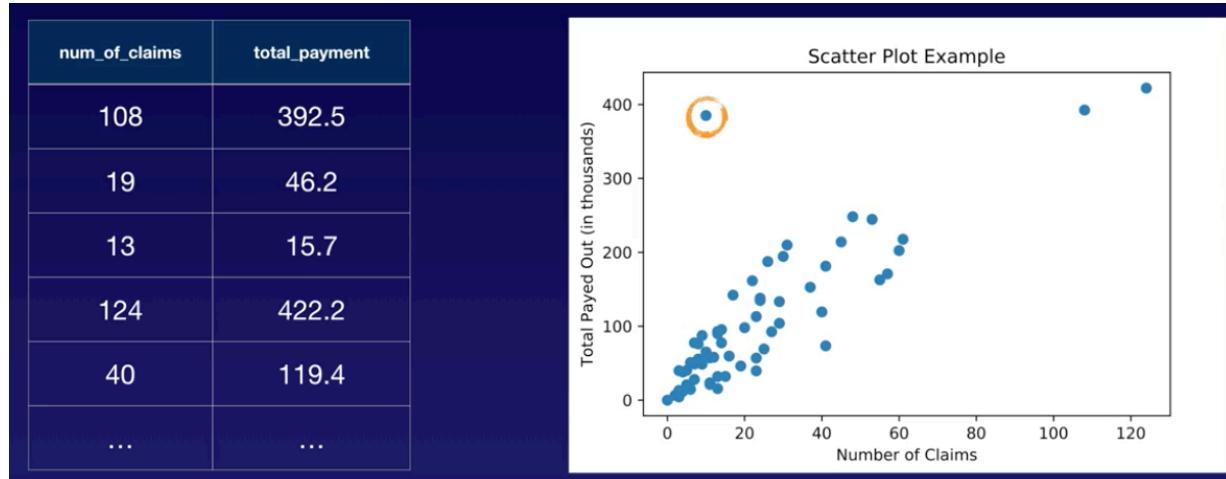


## 2 Bubble Plots

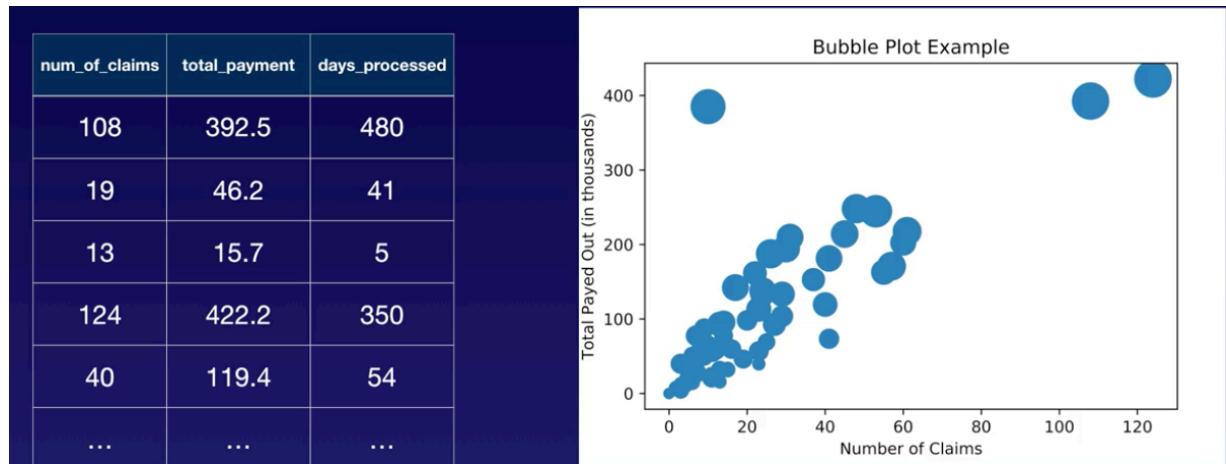
Also known as bubble charts. These graphs plot points along the x and y axis for three values. Bubble size is the third value measured.



Example of scatter plot



And Bubble plot

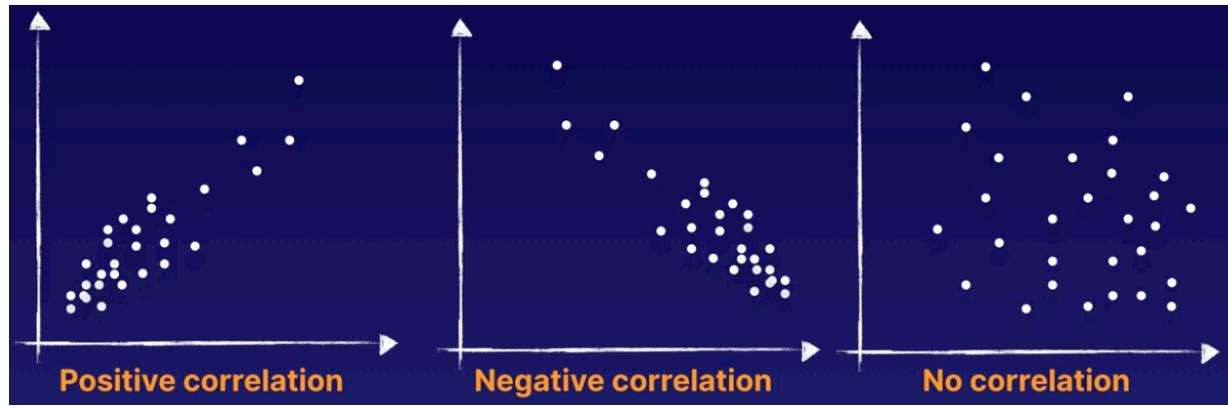


- **Scatter Plots**

**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?



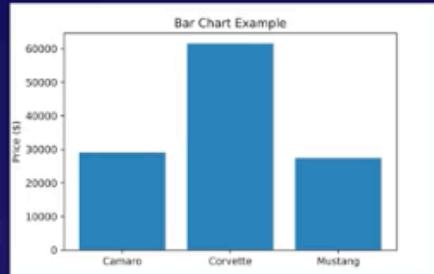
## Comparisons

Use Bar charts or Line Charts

1

### Bar Charts

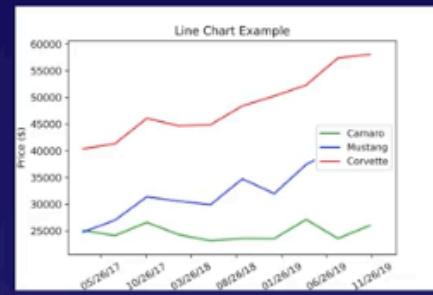
Bar charts are graphs that use lines (bars) to mark single variable values. Provides a way to lookup and compare values.



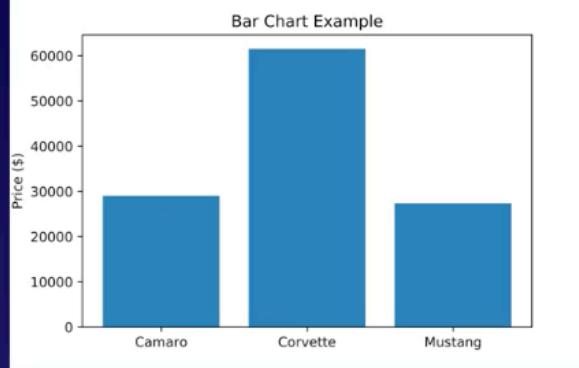
2

### Line Charts

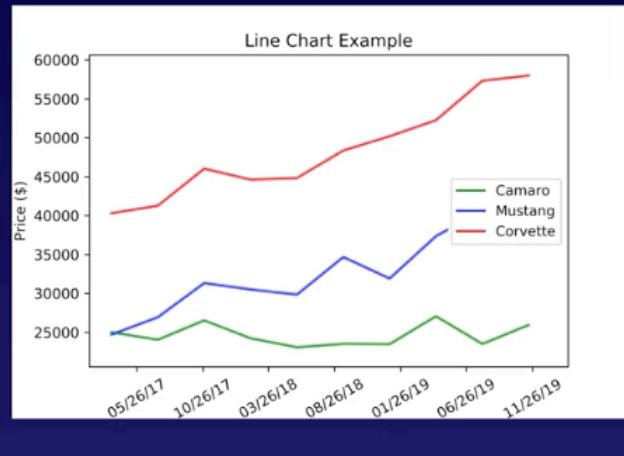
Line charts are graphs that use lines to show one or more variables changing over time.



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

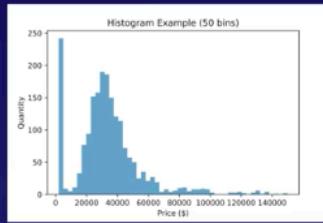
# Distributions

Visualizing distributions in your data can show how your data is grouped or clustered over certain intervals.



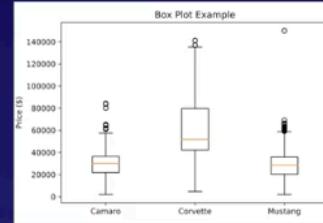
### 1 Histograms

These graphs put values into buckets or bins and determines a measurement (amount, frequency, duration, density, etc).



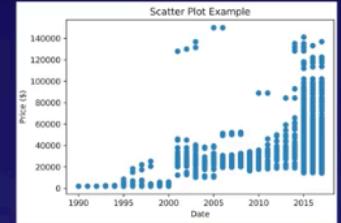
### 2 Box Plots

These graphs show a wealth of distribution information. You can see things like lowest and highest values, outliers and where most of the values fall.



### 3 Scatter Plots

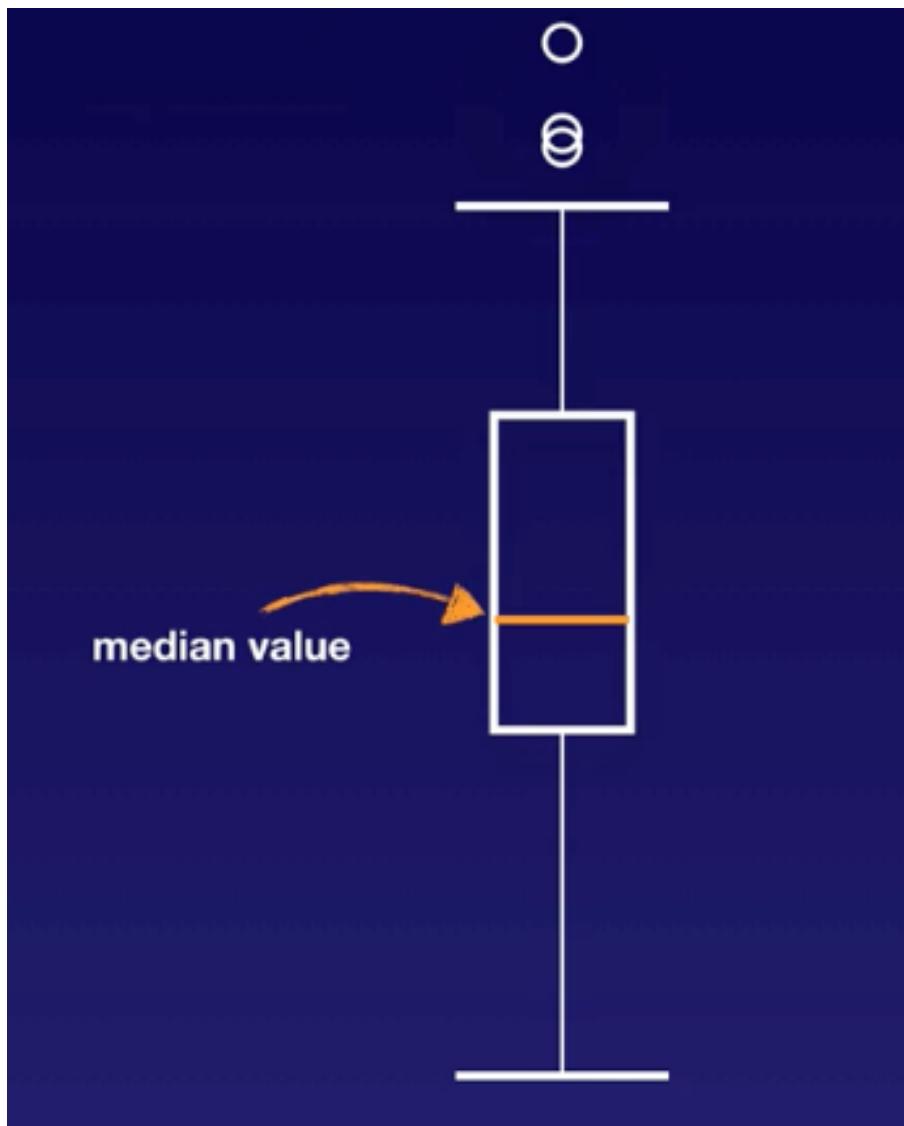
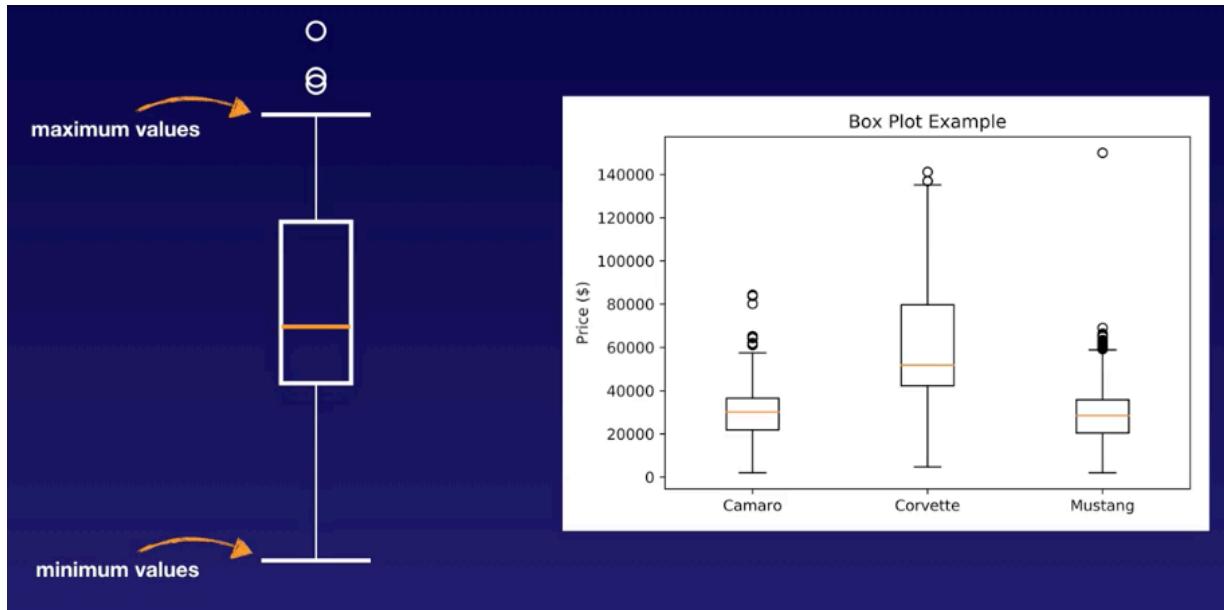
Also known as scatter charts. These graphs plot points along the x and y axis for two values. Can show clustering and distribution of your data.



Box plots:

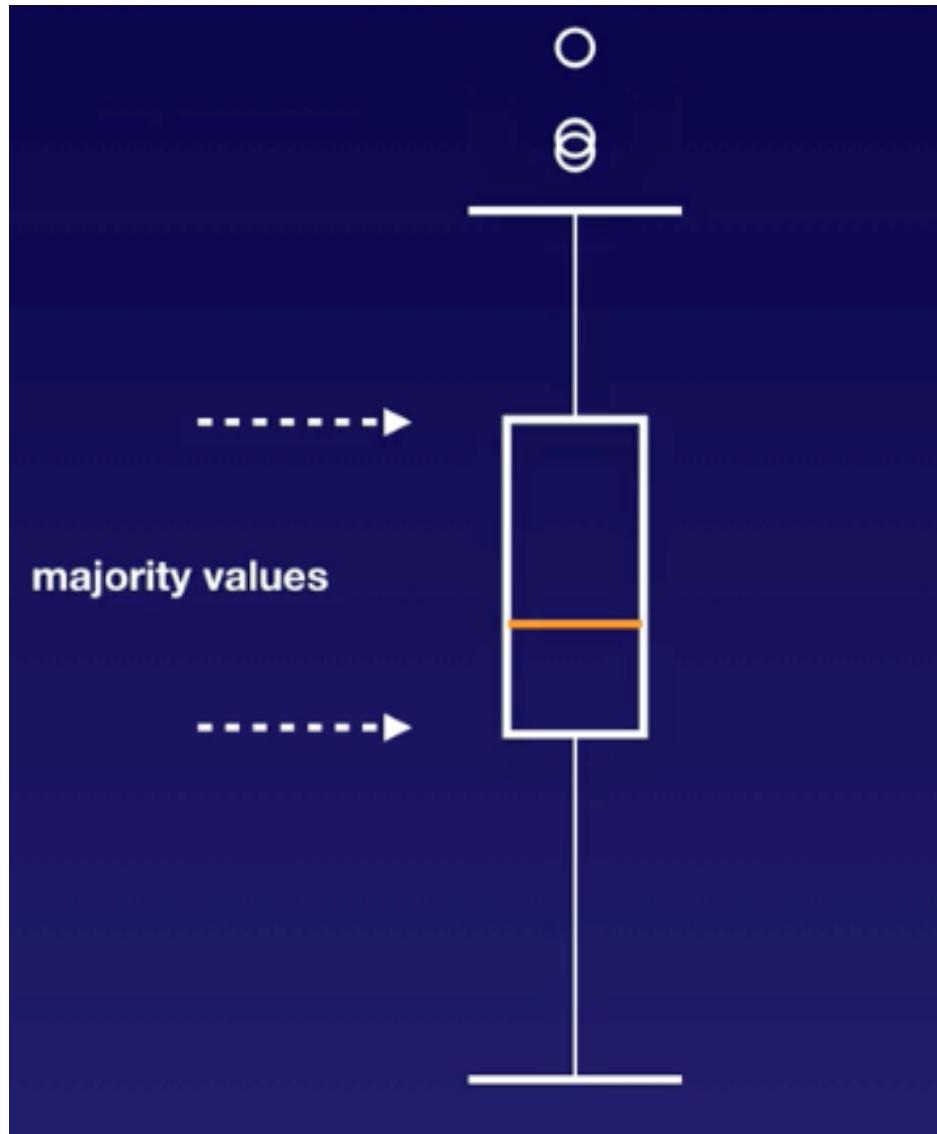
Show a lot of info into a single visualization

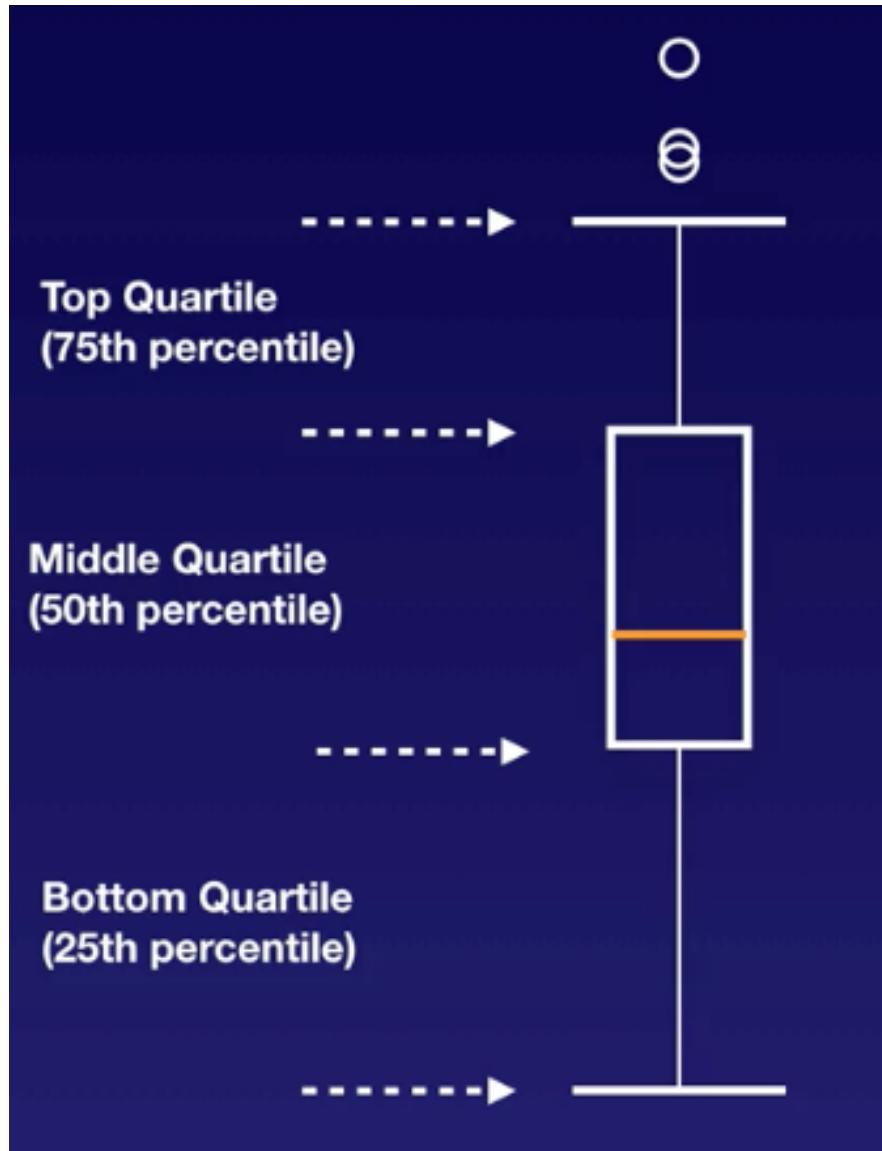
- min and max values
- median value
- extreme values
- majority of values (50% or more)
- 25/75 quartile



any extremes







- **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

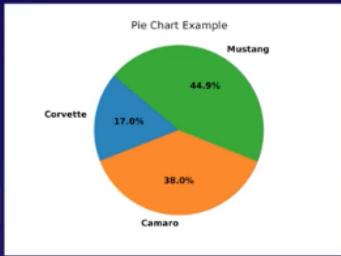
# Compositions

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



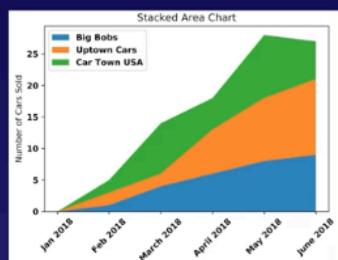
## 1 Pie Charts

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



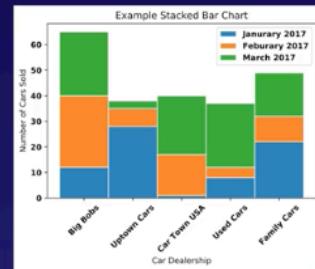
## 2 Stacked Area Charts

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.

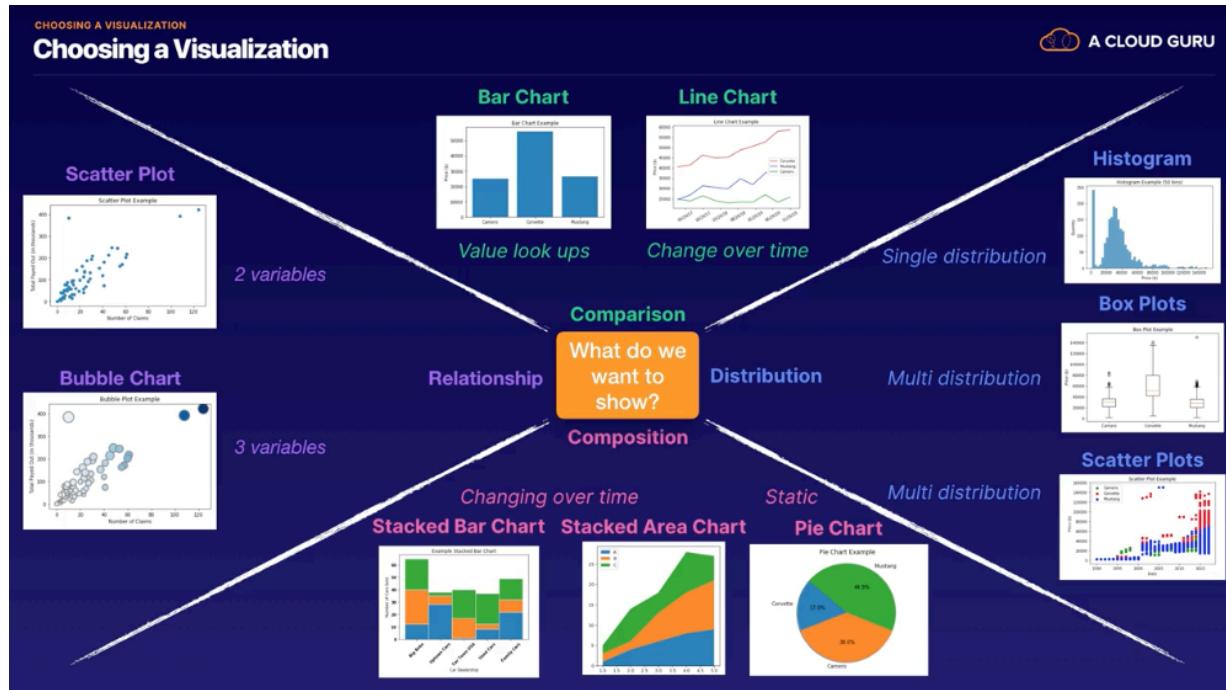
## • Stacked Bar Charts

**Example:** Showing the quarterly revenue totals for each region.

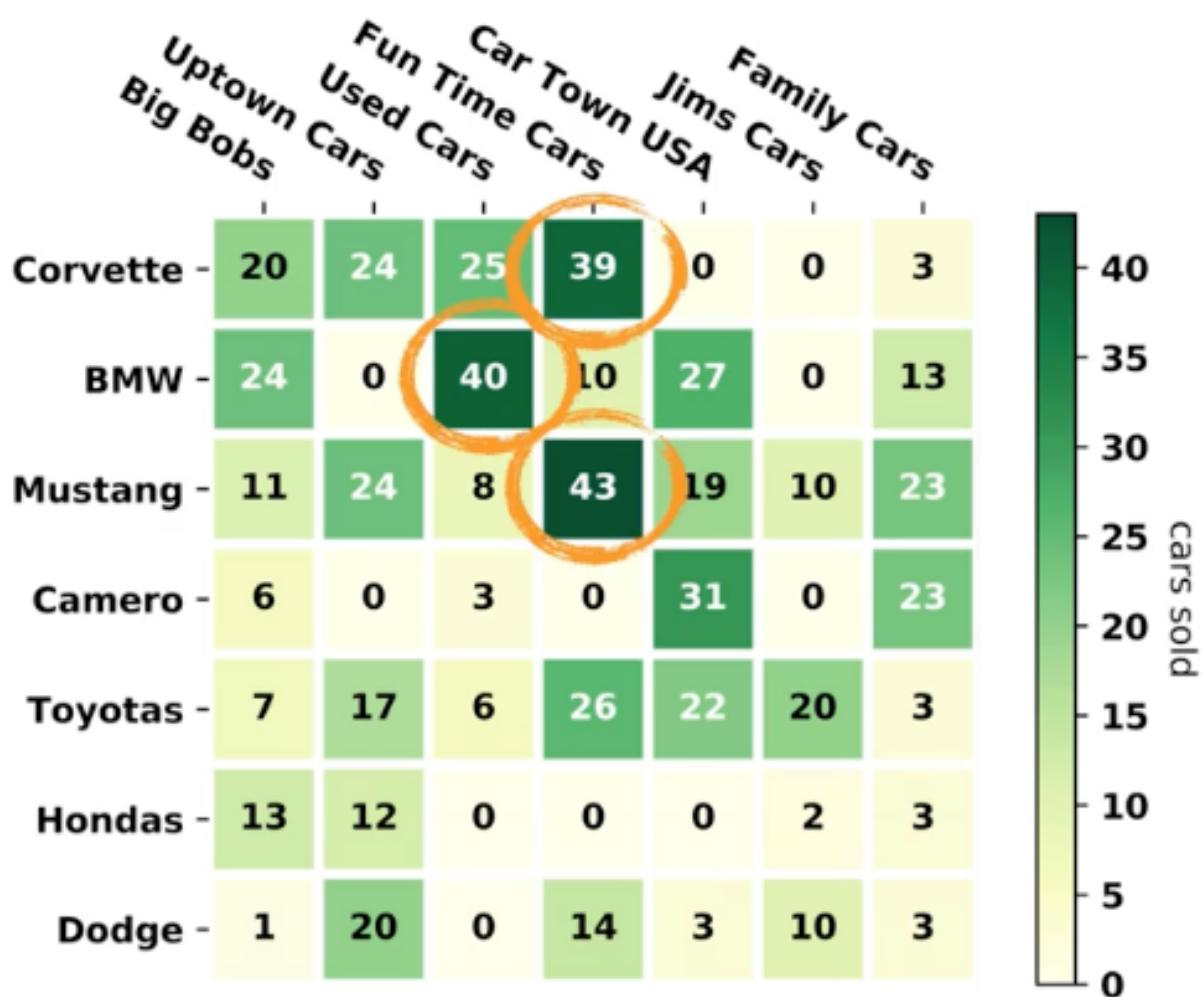
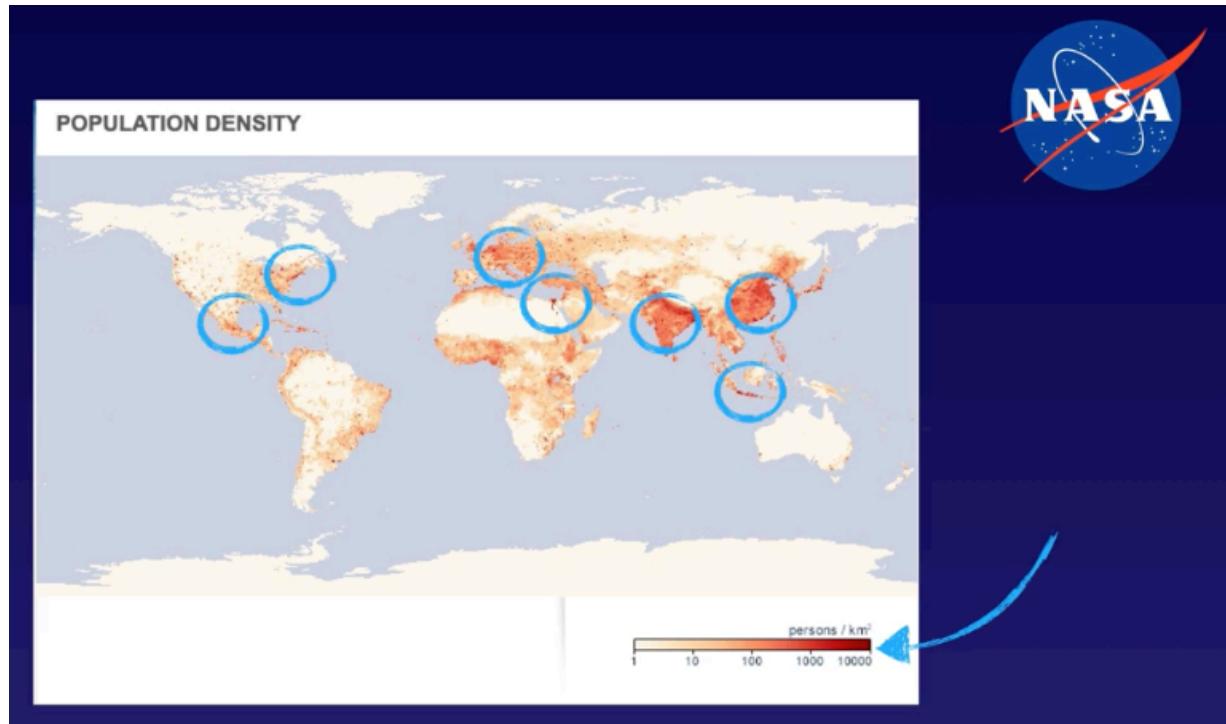
## Choosing a visualization

# Choosing a Visualization

Picking the right graph, chart, or visualization just depends on what you want to see.



Another type: heat map



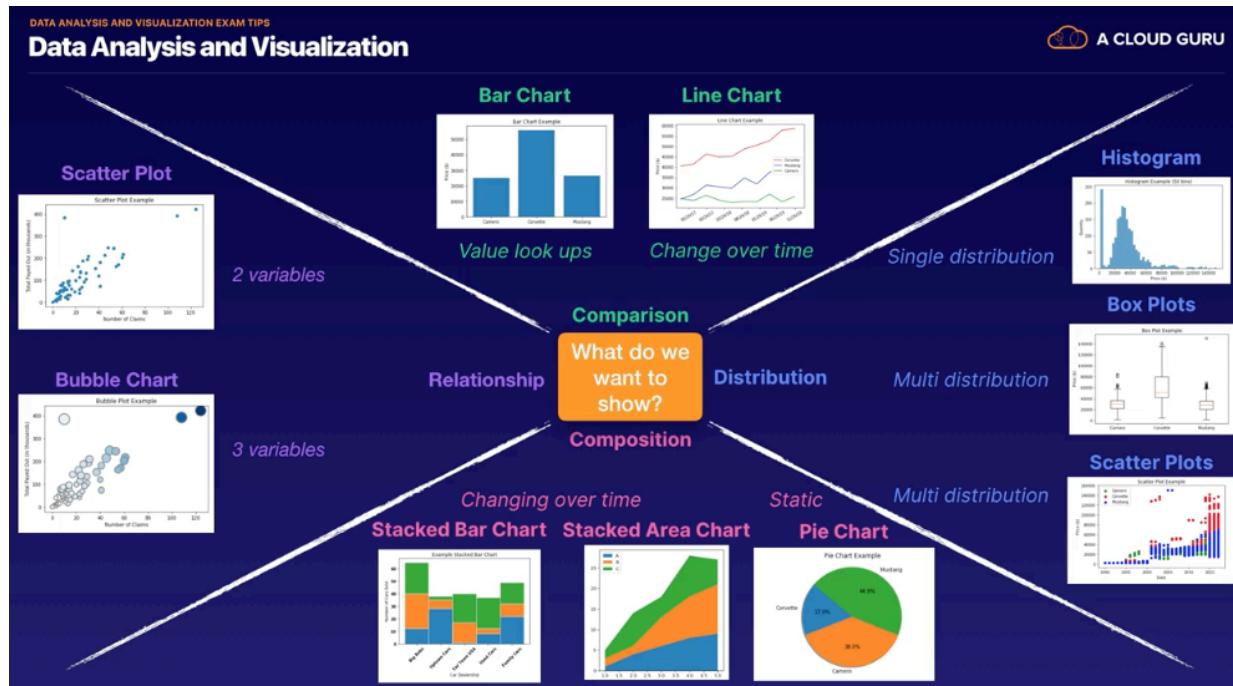
## Exam tips

### Data Analysis and Visualization



#### Data Analysis and Visualization

- Know what data analysis and visualization is and why it is important.
- Understand what Amazon QuickSight is and how it can be used.
- Be able to recognize different types of visualizations and what each visualization represents.
  - Relationships
  - Comparisons
  - Distribution
  - Composition
- Know what heatmaps are and what they can represent.



ML Insights AWS Insights:

<https://www.youtube.com/watch?v=Fhl1kVABMF0>

# One platform for data and analytics

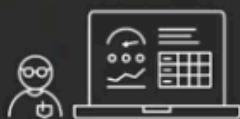
Broadest and deepest portfolio of purpose-built services

Business Intelligence & Machine Learning

The screenshot shows the AWS Data & Analytics landing page. At the top, there are links for Amazon QuickSight and Machine Learning. Below that, sections include Relational Databases (Amazon Aurora, Amazon RDS), Non-Relational Databases (Amazon DynamoDB, Amazon ElastiCache), Analytics (DW, Big Data Processing, Ad hoc, Amazon Redshift, Amazon EMR, Amazon Athena, Amazon ES, Kinesis Data Analytics), Data Lake (Amazon S3/Glacier, AWS Glue), Data Movement (Database Migration Service, Snowball, Snowmobile, Kinesis Data Firehose, Kinesis Data Streams), and Data Protection (Amazon Macie). The footer includes the re:Invent logo, a copyright notice (© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.), and the AWS logo.

## 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

re:Invent

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.



# 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.



## 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.



## 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.



## 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.

re:Invent

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.



# 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.

re:Invent

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.



# ML-powered anomaly detection

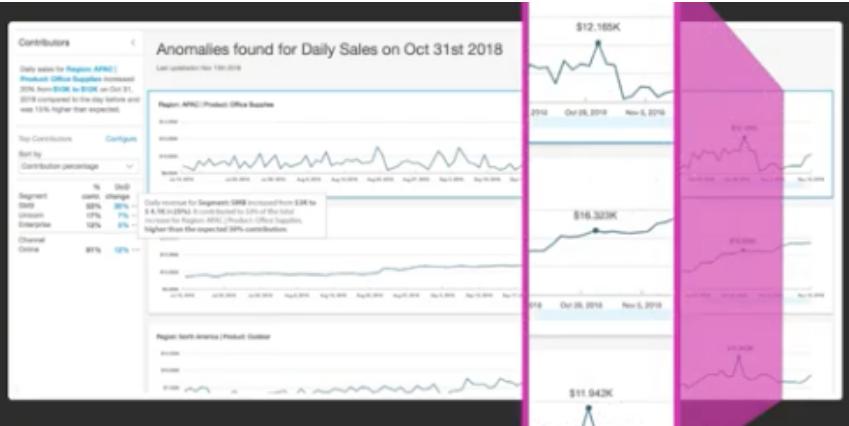
## Challenges

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

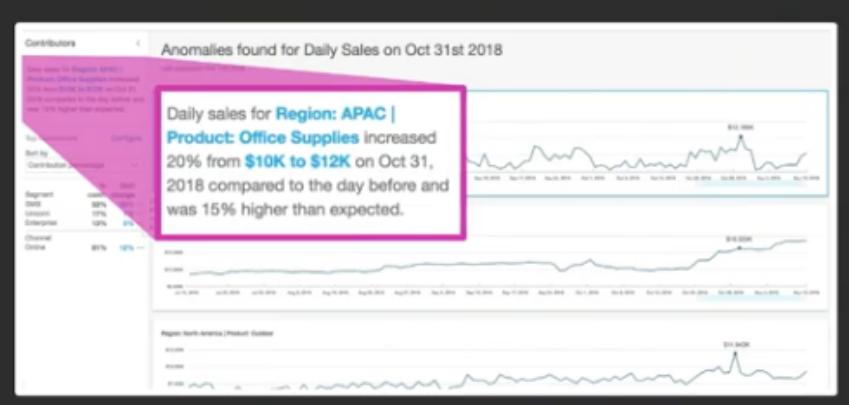
## Our solution

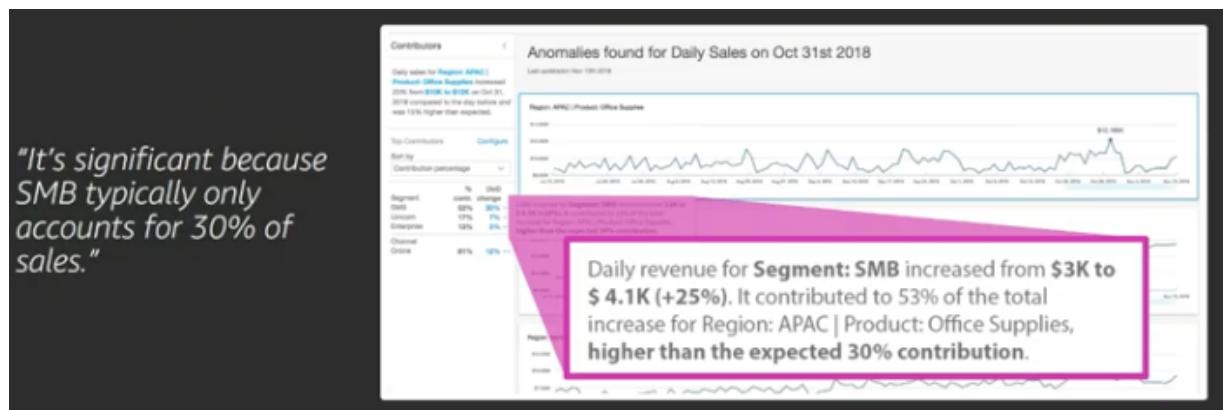
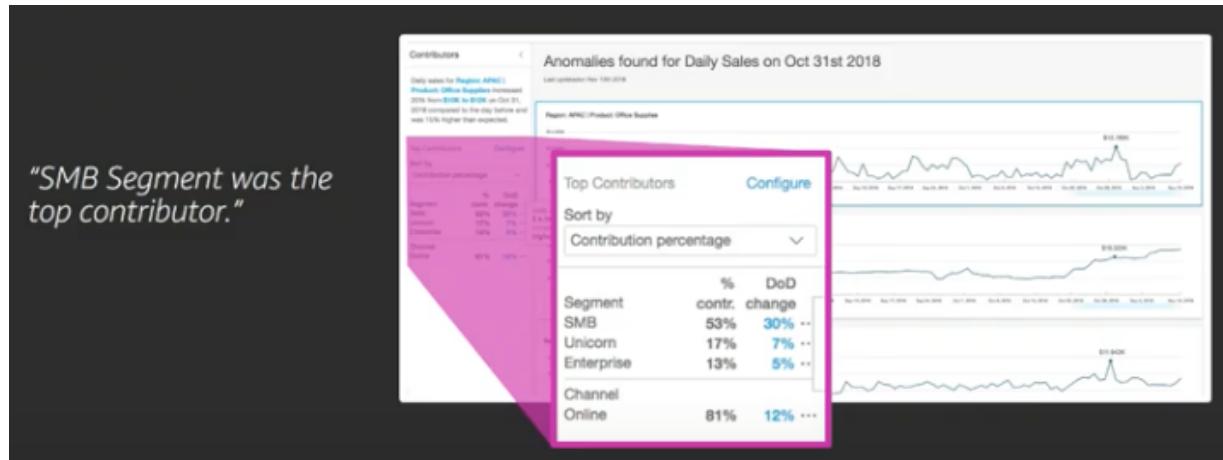
Analyze millions of metrics  
Continuously run in background  
What happened AND why  
Alerts delivered via email  
No ML expertise needed

Discover all the hidden trends and anomalies



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





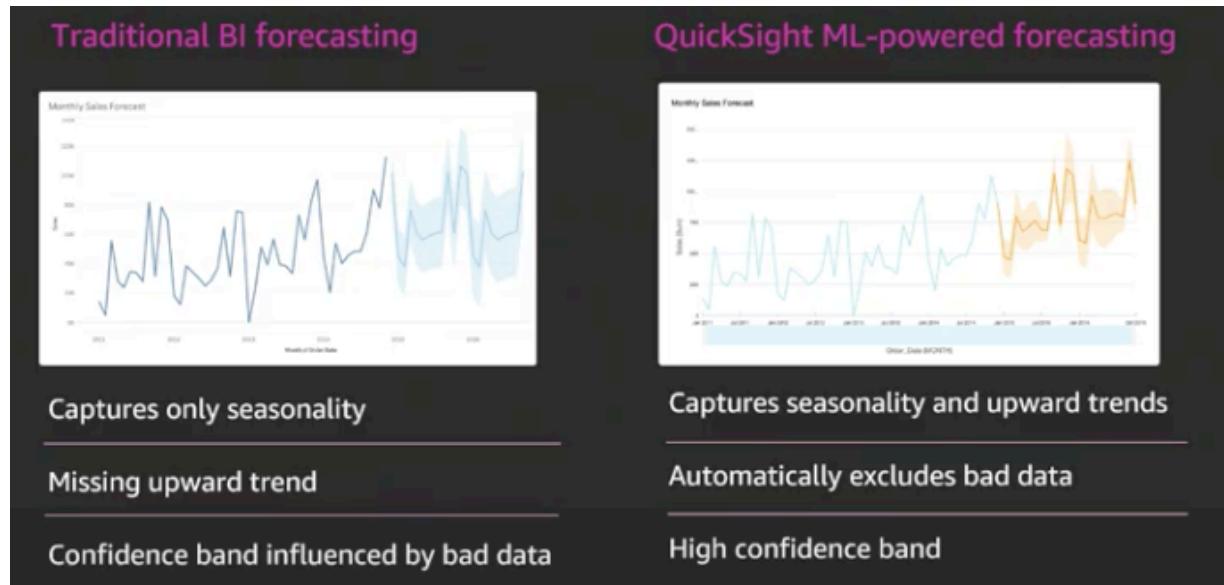
## ML-powered forecasting

### 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



## 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

**ABC Daily Sales Report**

**Daily Revenue**

Year-to-date revenue increased by 0.01% from \$107,002.90 to \$107,003.11. Total YTD revenue is \$107,003.11 compared to \$107,002.90 last year.

**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.

**YTD Revenue**

Year-to-date revenue increased by 0.01% from \$107,002.90 to \$107,003.11. Total YTD revenue is \$107,003.11 compared to \$107,002.90 last year.

**Callouts By Product and Country**

Daily revenue for Baby Products | Russia on Nov 22, 2018 was lower than expected at -\$35,000 (-1.81%).

**Customers**

Revenue for MULTIDEL, INC. on Nov 22, 2018 was higher than expected at \$80,433.95.

**Daily Revenue Forecast**

Daily revenue is predicted to reach \$10.00M by end of the year. We expect to exit the year with an annualized run rate of \$4.38B. Total revenue for 2018 is predicted to reach \$3.47B. **\$63,000M (-1.78%) below annual target of \$3.54B.**

**Top / Bottom Movers by Product**

Top daily revenue increase by products are:

- Electronics increased by \$887.04 (0.81%) from \$7,008.08 to \$7,004.82.
- Clothing increased by \$484.06 (0.08%) from \$769,581.00 to \$770,046.75.
- Industrial increased by \$427.49 (0.04%) from \$1,078,110.02 to \$1,079,137.51.
- Home Services increased by \$114.08 (0.21%) from \$55,338.04 to \$55,452.72.

**Top / Bottom-to-Plan Variance by Product**

Top products above plan for today are:

- Movies is \$147,437.42 above goal.
- Financial Services is \$86,111.46 above goal.
- Clothing is \$42,000.37 above goal.
- Computers is \$36,003.06 above goal.
- Outdoors is \$21,404.67 above goal.

**Revenue by Product Category**

Nov 18, 2018	Nov 17, 2018	
Product Cat... Arts	Revenue \$4,000.40	Revenue \$4,000.18
Automotive	\$62,300.00	\$62,400.34
Baby Product	\$1,384,240.11	\$1,386,061.33
Beauty	\$8,114.71	\$8,114.96
Books	\$1,350,700.60	\$1,351,260.71
Business	\$60,700.00	\$61,014.80
Clothing	\$770,046.75	\$780,001.80

**ABC 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.

**YTD Revenue**

Year-to-date revenue increased by 0.01% (\$1.14B) from \$107,002.90 to \$107,003.11. Total YTD revenue is \$107,003.11 compared to \$107,002.90 last year.

**Callouts by Product and Country**

Daily revenue for Baby Products | Russia on Nov 22, 2018 was lower than expected at -\$1.17M (-0.58%).

**Customers**

Revenue for MULTIDEL, INC. on Nov 22, 2018 was higher than expected at \$80,433.95.

**Daily Revenue Forecast**

Daily revenue is predicted to reach \$10.00M by end of the year. We expect to exit the year with an annualized run rate of \$4.38B. Total revenue for 2018 is predicted to reach \$3.47B. **\$63,000M (-1.78%) below annual target of \$3.54B.**

**Top / Bottom Movers by Product**

Top daily revenue increase by products are:

- Electronics increased by \$887.04 (0.81%) from \$7,008.08 to \$7,004.82.
- Clothing increased by \$484.06 (0.08%) from \$769,581.00 to \$770,046.75.
- Industrial increased by \$427.49 (0.04%) from \$1,078,110.02 to \$1,079,137.51.
- Home Services increased by \$114.08 (0.21%) from \$55,338.04 to \$55,452.72.

**Top / Bottom-to-Plan Variance by Product**

Top products above plan for today are:

- Movies is \$147,437.42 above goal.
- Financial Services is \$86,111.46 above goal.
- Clothing is \$42,000.37 above goal.
- Computers is \$36,003.06 above goal.

**Revenue by Product Category**

Nov 18, 2018	Nov 17, 2018	Nov 16, 2018	
Product Cat... Arts	Revenue \$4,000.40	Revenue \$4,000.18	Revenue \$4,000.18
Automotive	\$62,300.00	\$62,400.34	\$62,400.34
Baby Product	\$1,384,240.11	\$1,386,061.33	\$1,386,061.33
Beauty	\$8,114.71	\$8,114.96	\$8,114.96
Books	\$1,350,700.60	\$1,351,260.71	\$1,351,260.71
Business	\$60,700.00	\$61,014.80	\$61,014.80
Clothing	\$770,046.75	\$780,001.80	\$780,001.80

