Machine Learning Study

WEEK 2

Data Visualization Made Easy

Dr Tan Yan Bin

Data Engineer, Trainer

Yvaine YanBin Tan

Slides: bit.ly/3Ka6Bn9

TENSORFLOW & DEEP LEARNING MALAYSIA

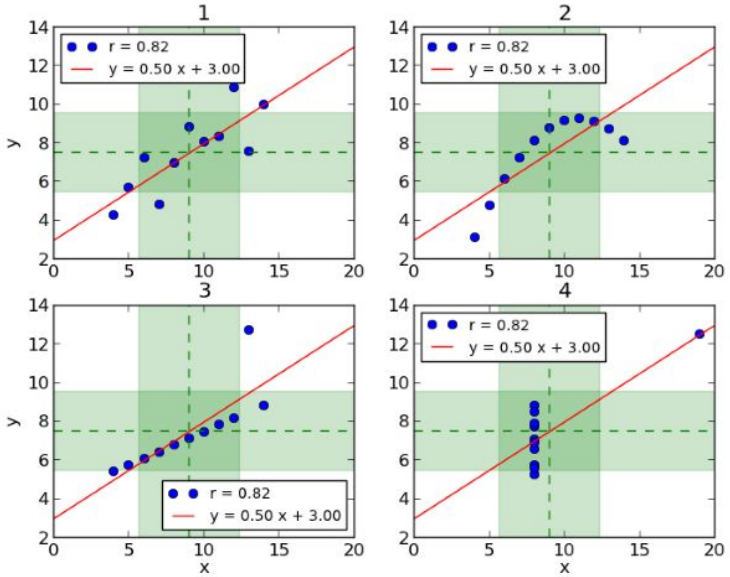
SUNWAY TECH CLUB

SUNWAY UNIVERSITY DATAXSIGHT Research Cluster

Why visualization?

Find stories from data

Get initial insights from available data

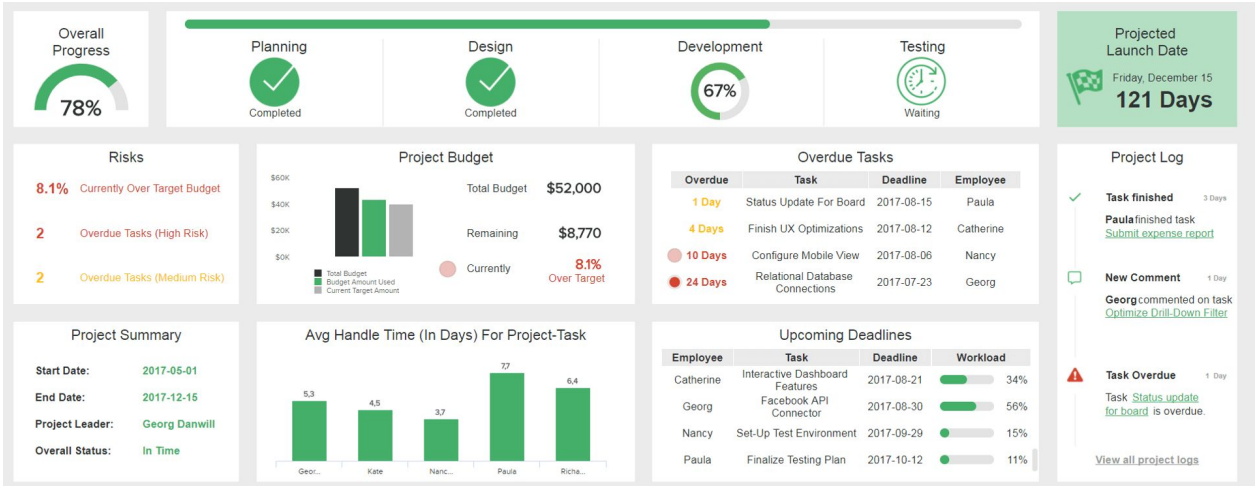


img source

Study

Tell stories from data

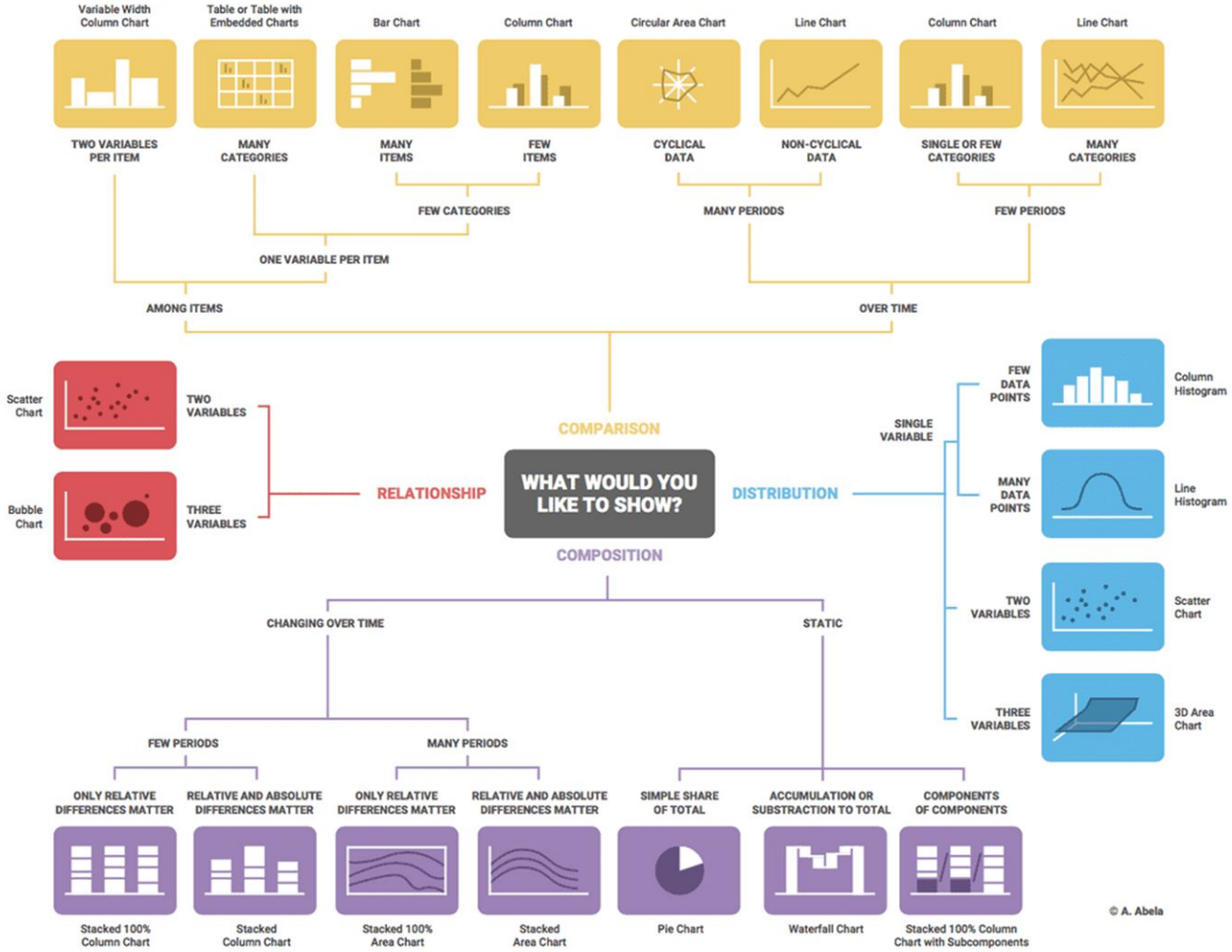
Select intended perceptions to convey our messages



Study

[img source]

Chart Types


Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart
TWO VARIABLES PER ITEM
MANY CATEGORIES
MANY ITEMS
FEW ITEMS
CYCLICAL DATA
NON-CYCLICAL DATA
SINGLE OR FEW CATEGORIES
MANY CATEGORIES
FEW CATEGORIES
MANY PERIODS
FEW PERIODS
ONE VARIABLE PER ITEM
AMONG ITEMS
OVER TIME
Scatter Chart
TWO VARIABLES
Bubble Chart
THREE VARIABLES
RELATIONSHIP
COMPARISON
WHAT WOULD YOU LIKE TO SHOW?
DISTRIBUTION
COMPOSITION
SINGLE VARIABLE
FEW DATA POINTS
Column Histogram
MANY DATA POINTS
Line Histogram
TWO VARIABLES
Scatter Chart
THREE VARIABLES
3D Area Chart
CHANGING OVER TIME
STATIC
FEW PERIODS
MANY PERIODS
ONLY RELATIVE DIFFERENCES MATTER
RELATIVE AND ABSOLUTE DIFFERENCES MATTER
ONLY RELATIVE DIFFERENCES MATTER
RELATIVE AND ABSOLUTE DIFFERENCES MATTER
SIMPLE SHARE OF TOTAL
ACCUMULATION OR SUBSTRACTION TO TOTAL
COMPONENTS OF COMPONENTS
Stacked 100% Column Chart
Stacked Column Chart
Stacked 100% Area Chart
Stacked Area Chart
Pie Chart
Waterfall Chart
Stacked 100% Column Chart with Subcomponents
© A. Abela

img source

If you wish to

highlight / draw attention

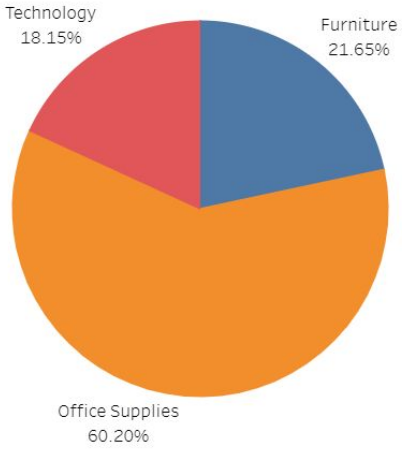
	Order Date											
Year of Ord..	January	February	March	April	May	June	July	August	Septem..	October	Novem..	Decemb..
2019	2,539	862	693	3,489	3,196	5,000	-841	5,765	8,594	3,469	9,363	9,555
2020	-3,190	2,814	9,931	4,187	4,677	3,336	3,289	5,372	8,220	2,818	12,475	8,093
2021	2,916	5,005	3,708	2,992	8,677	4,926	4,560	2,317	9,370	16,257	4,011	17,926
2022	7,232	1,614	15,013	958	6,300	8,247	7,007	9,488	11,051	10,671	9,692	8,656

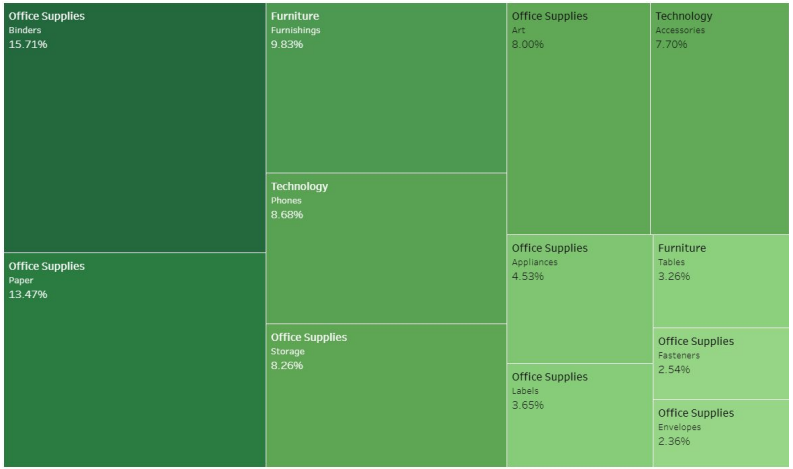
	Order Date											
Year	January	February	March	April	May	June	July	August	Septemb..	October	November	December
2019	2,539	862	693	3,489	3,196	5,000	-841	5,765	8,594	3,469	9,363	9,555
2020	-3,190	2,814	9,931	4,187	4,677	3,336	3,289	5,372	8,220	2,818	12,475	8,093
2021	2,916	5,005	3,708	2,992	8,677	4,926	4,560	2,317	9,370	16,257	4,011	17,926
2022	7,232	1,614	15,013	958	6,300	8,247	7,007	9,488	11,051	10,671	9,692	8,656

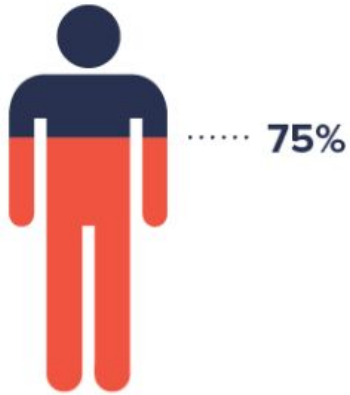
Study

If you wish to

show composition of a whole



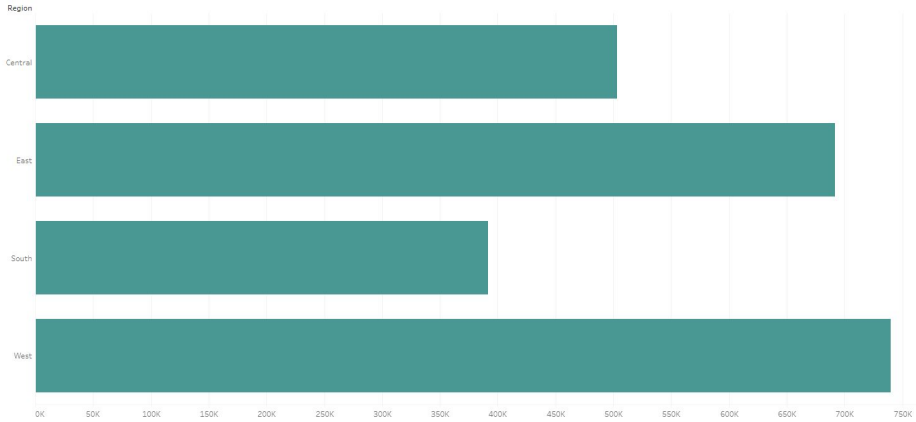


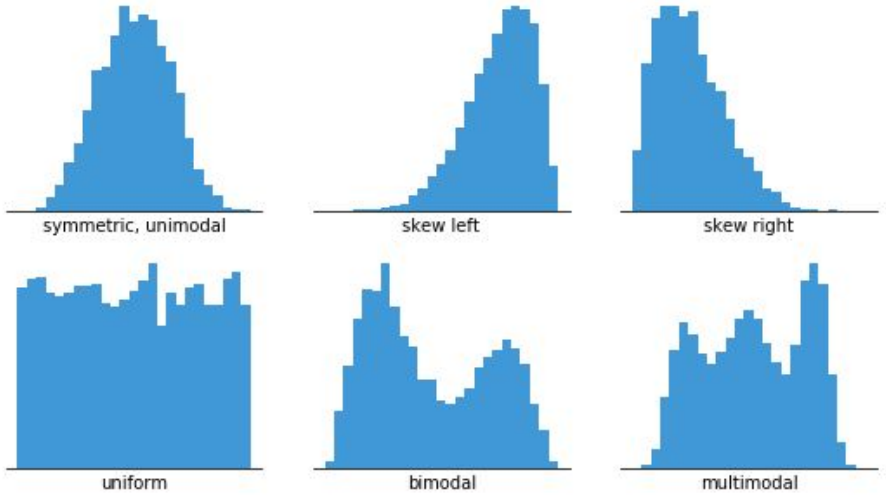


Study

If you wish to

show distribution

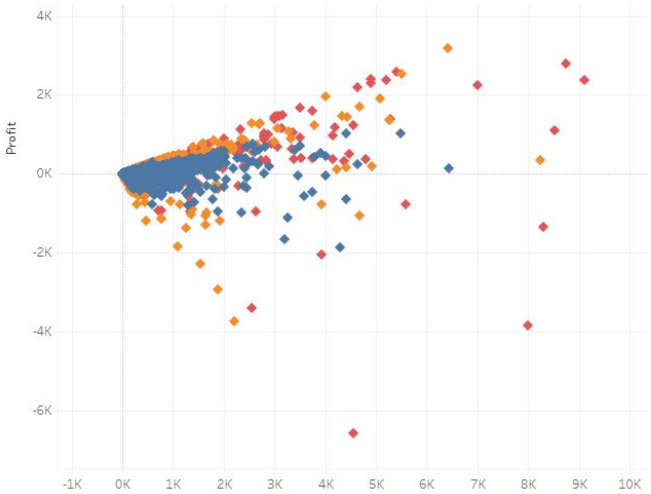


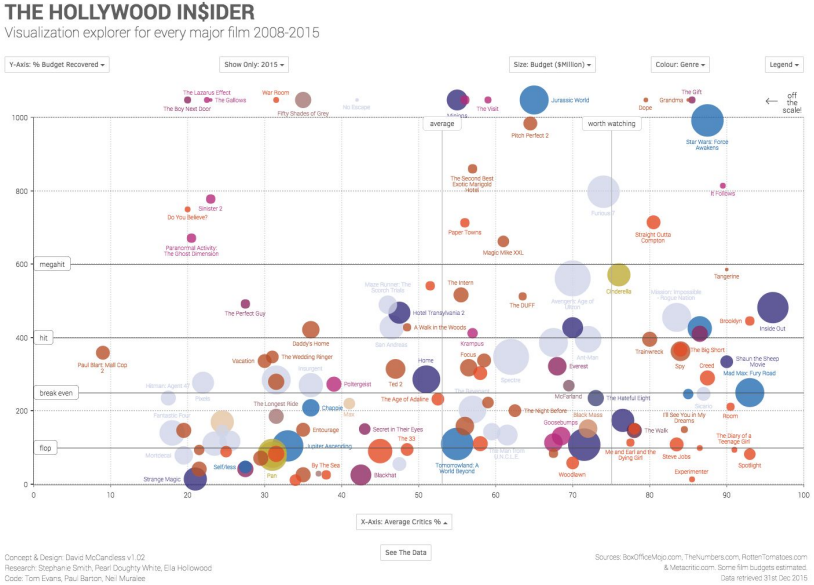


Study

If you wish to

show relationship

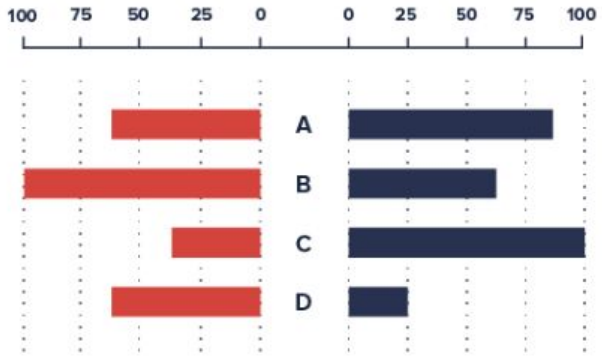


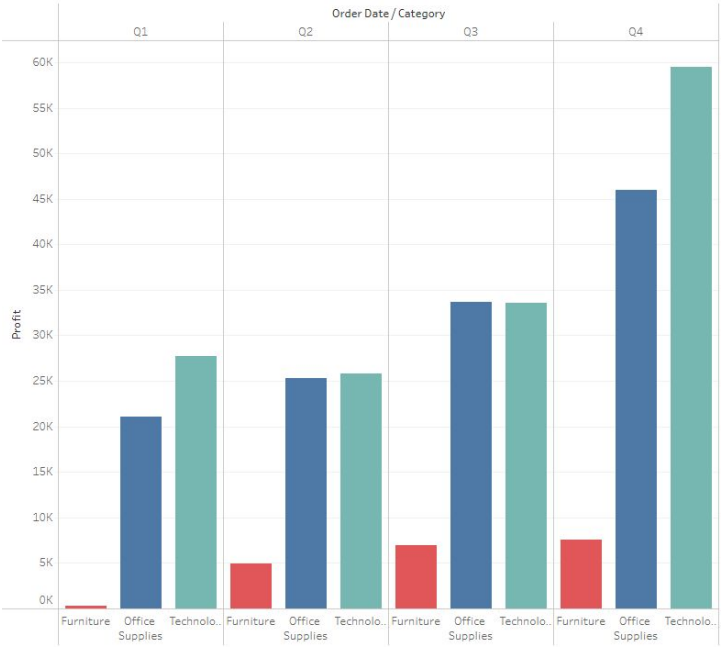


Study

If you wish to

compare among categories





Study

If you wish to

see how things change over time

Study

More Ideas:

https://datavizproject.com/

Tips & Tricks

Data Types

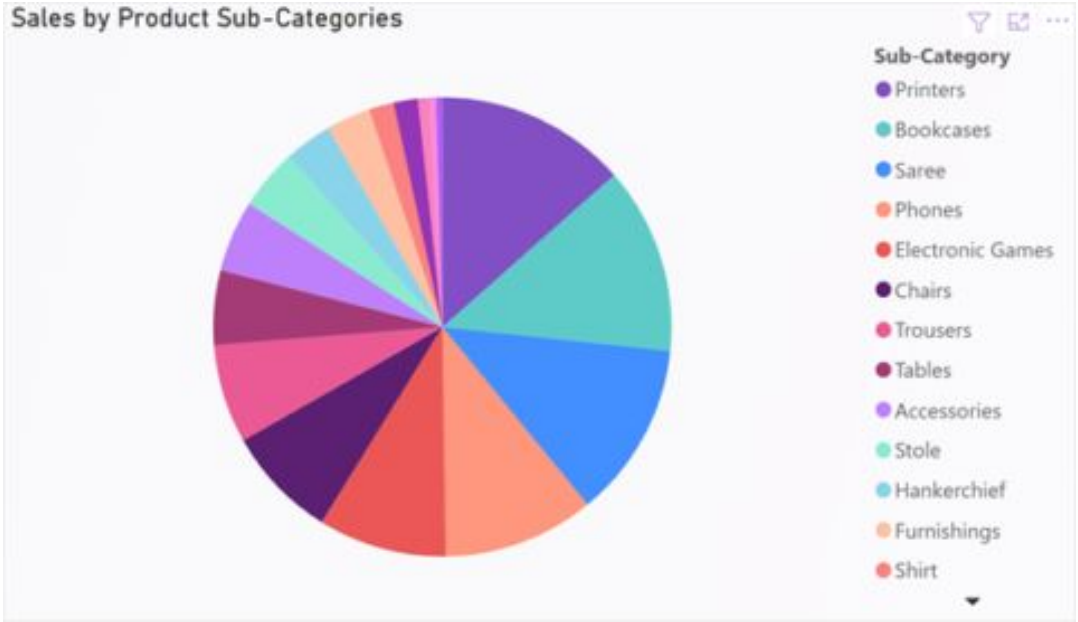
Discrete
Countable

Nominal
Categorical

Continuous
Measurable

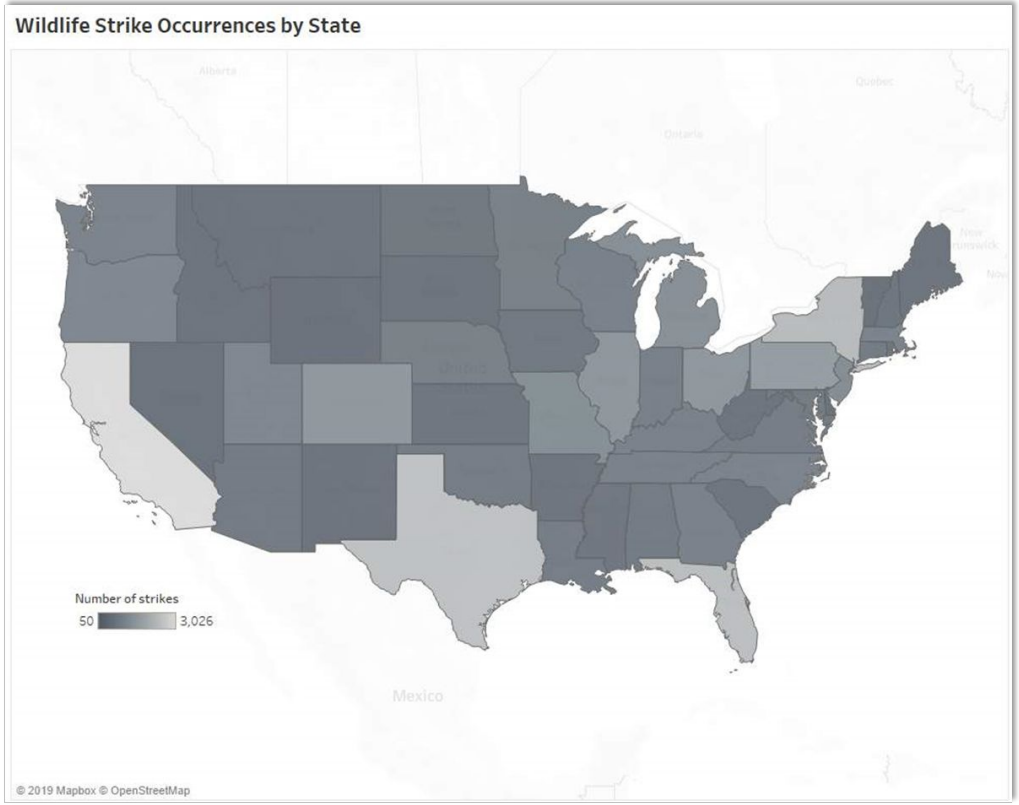
Ordinal
Categorical with order

Comment on this chart



Study

Anything weird with this chart?



Study

Common Data Viz Issues

- Inappropriate choice of chart type
- Too much information
- Poor choice of visuals (color, patterns, lines, ...)
- Inconsistent visuals and scales
- Inaccurate quantitative representation

Lastly ...

- Be curious. Think. Spot unusual patterns.
- Focus on our audience.
- Ensure clarity, precision, efficiency

Hands-on!

with Python Seaborn

https://www.kaggle.com/learn/data-visualization

https://colab.research.google.com/drive/1zRWfPmRCNoptTubWuK3x75paO8djY4Nv#scrollTo=U05NONlZKfn7