# Exploratory Data Analysis (EDA)

PART ONE

# Learning Objectives

- Introduction to Exploratory Data Analysis (EDA)
- Basic use of various visualization for different EDAs
- What to do with statistics and use them in visualization
- Other techniques used in EDA. (Spatialisation)

# **Exploratory Data Analysis**

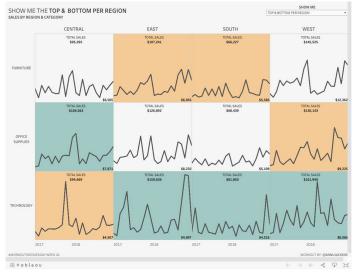
- EDA processes
  - Testing assumptions
  - Selecting models
  - Validating regression model
  - Selecting estimator
  - Identifying relationships
  - Detecting outlier
  - etc.





# **Exploratory Data Analysis (cont.)**

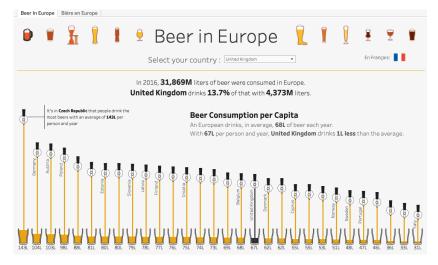
- Less focus on visual/artistic design
- Assists the viewers to pay attention to the data.



https://public.tableau.com/profile/ann.jackson#!/vizhome/WorkoutWednesd ayWeek41-TopBottomHighlights/WorkoutWednesdayWeek41-TopBottomHighlights

# **Explanatory Data Analysis**

- Main message to convey exists
- Typically used in presentation, hackathon, etc.



https://public.tableau.com/profile/guillevin#!/vizhome/BeerInEurope/BeerInEurope

# Types of EDA

# Exploratory Data Analysis

#### Utilize data's statistical attributes

- Temporal comparison
- Attributes comparison
- Ranking comparison
- Composition analysis
- Distributions analysis
- Variance analysis
- Correlation analysis
- Geographic analysis

# Temporal analysis/comparison

Display how indices change over a period of time

- Visualise
- Trend
- Pattern

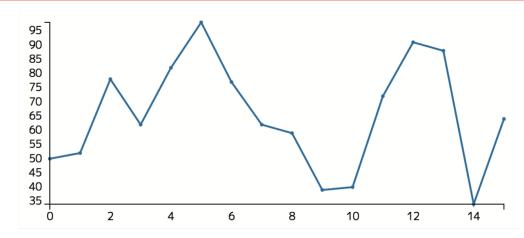
# Basic use of various visualisation for Temporal analysis/comparison

Basic	<b>☆</b>	Line graph
Emphasize the amount		Area graph
Change of composition	$\stackrel{\wedge}{\sim}$	Stacked Area graph
Change of % composition	<b>\}</b>	% Stacked Area graph
A pair comparison	X	Slope chart
Compare multiple trends/changes	<b>₩</b>	Spark line

Icons from Data Visualization, Toshikuni and Watanabe (2019)

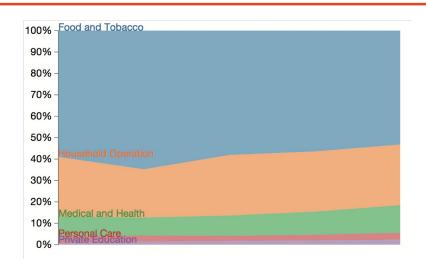
# **Line Graph/Chart**





Connected points

# **Area Graph**



Area

# Theme River (Area Graph)

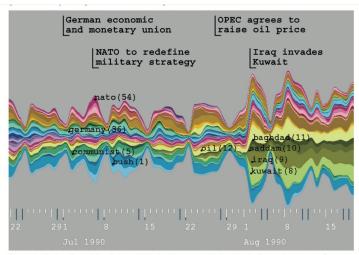
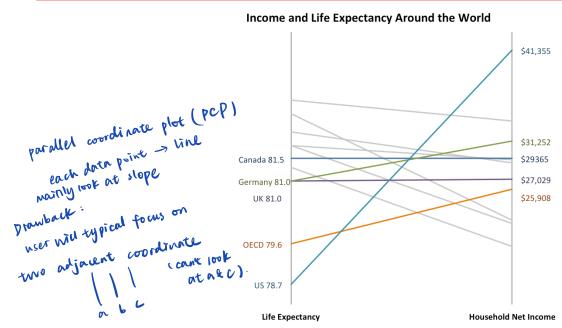


Figure 3: AP data from July - August 1990. A wide current in the river indicates heavy use of a topic, while changes in color distribution correlate to changes in themes.

\* [Visualizing Theme Changes over Time]

(http://www.ifs.tuwien.ac.at/~silvia/wien/vu-infovis/references/havre-ieeeinfovis00.pdf)

# **Slope Chart**



https://policyviz.com/2015/09/09/excel-slope-chart-with-two-metrics/

# **Spark Line**

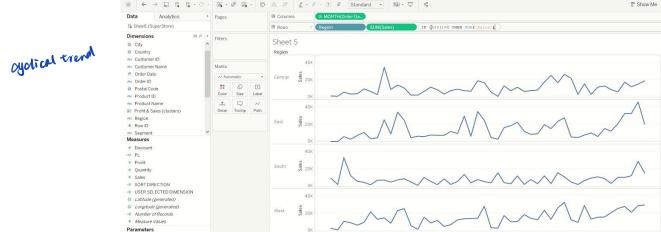
(i) Tableau - Book1

ALC CHOOSE SORT TYPE

AND SELECTED DIMENSION

# Number of States

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help



November 2013

May 2014

November 2014

Month of Order Date

May 2015

November 2015

o x

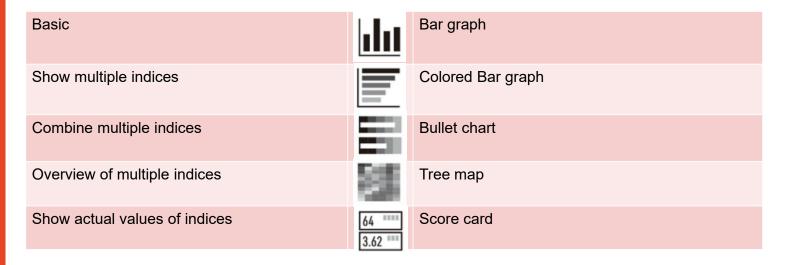
Sheet 1 Percent of Records with State = ... Sheet 3 Sheet 4 Sheet 5 🖳 🖽 👊

# **Attribute analysis/comparison**

Compare indices for each attribute

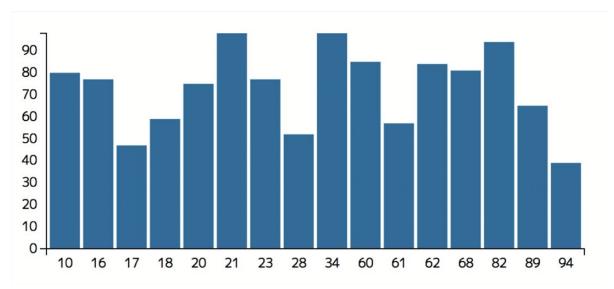
- Visualize/compare
  - Large/small
  - Good/bad

# Basic use of various visualisation for Attribute analysis/comparison



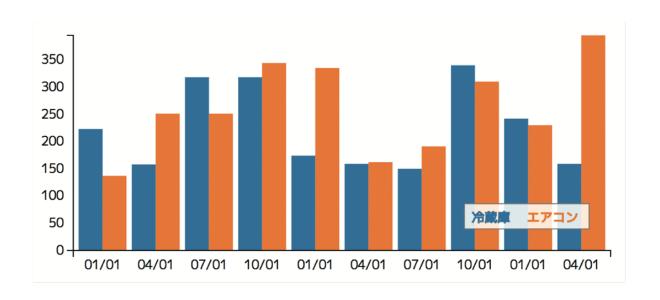
Icons from Data Visualization, Toshikuni and Watanabe (2019)

# **Bar Graph**



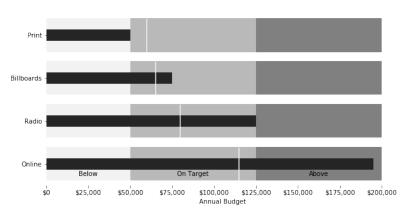
Line

### **Bar Chart**

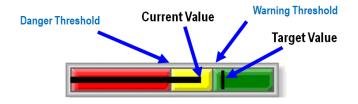


### **Bullet Chart**

#### Marketing Channel Budget Performance



https://pbpython.com/bullet-graph.html



 $\label{lem:https://glean.info/how-to-select-the-best-type-of-chart-to-visualize-your-data/$ 

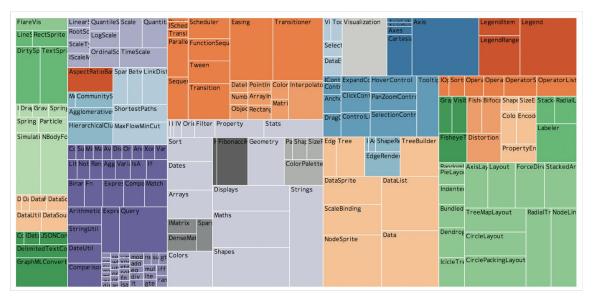
### **Scorecard**

INANCIAL DBJECTIVES	TARGET	CURRENT	TO TARGET	LEARNING & GI OBJECTIVES	ROWTH		
Increase Total Revenue	€5.458K	€6.197K	(+13.5%)	Open New Sales Regions	4	4	(O)
Increase Net Profit	€1.208K	€1.173K	(-2.9%)	Reduce Operating Costs	€30K	€26K	(-13%)
Increase Profit Margin	25.5%	23%	(-9.8%)	Develop New Proper Year	ducts 2	3	(+1)
Increase Profit per Customer	€100	€97	(-3%)	Standardise Staff Training Hours (me	onthly) 8 HRS	5 HRS	(-3 HRS)
CUSTOMER OBJECTIVES	TARGET	CURRENT	TO TARGET	INTERNAL OBJECTIVES	TARGET	CURRENT	TO TARGET
Increase Active Customers	5300	4.993	(-5.8%)	Reduce Average Absenteeism	6 DAYS	3 DAYS	(-3 DAYS)
Increase Customer Signups	876	819	(-6.5%)	Maximise Overall Labour Effectivene	ess 87%	76%	(-12.6%)
Sustain Customer Retention	151	128	(-14.6%)	Internally Fill Vancancies	30	29	(-1)
Improve Customer	85%	73%	(-14.1%)	Decrease Average Overtime (per Yea		30	(+O)

https://www.datapine.com/blog/kpi-scorecard-examples-templates-to-track-performance/

# **Tree Map**

Line/Area: size/location/nested



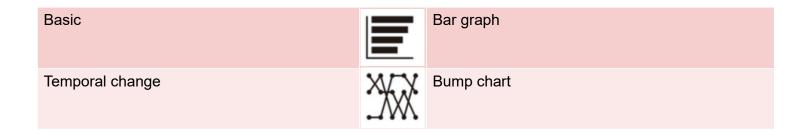
% [d3.js garally] (http://bl.ocks.org/mbostock/4063582)

# **Ranking Analysis**

Compare ranking of attributes based on index

- Visualize/compare
- Order

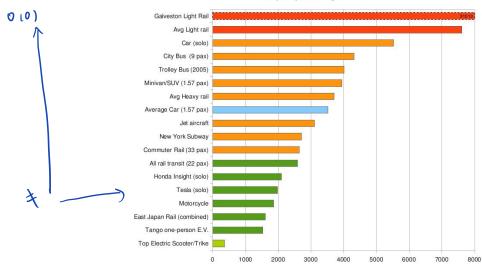
# Basic use of various visualisation for Ranking analysis



## **Bar Graph**

#### **USA Transportation Energy Use**



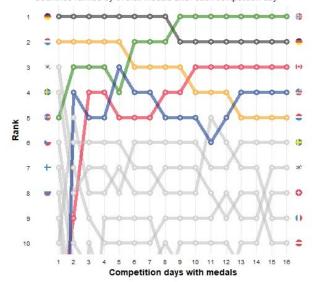


https://ideas.4brad.com/green-u-s-transit-whopping-myth

## **Bump Chart**

#### PyeongChang 2018 Olympic Winter Games

Countries ranked by overall medals after each competition day



# **Composition analysis**

Compare composition of attributes

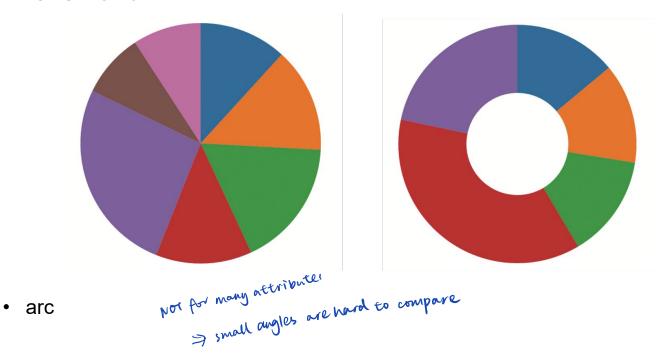
- Visualise /compare
- Order

# Basic use of various visualisation for Composition analysis

Small number of attributes		Pi chart
Many attributes		Tree map
Compare composition + amount	iİii	Stacked bar chart
Compare composition	İiii	% Stacked bar chart

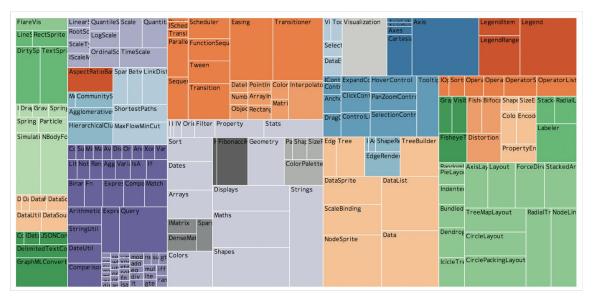
Icons from Data Visualization, Toshikuni and Watanabe (2019)

### **Pie Chart**



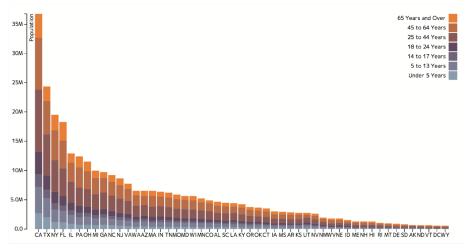
# **Tree Map**

Line/Area: size/location/nested



% [d3.js garally] (http://bl.ocks.org/mbostock/4063582)

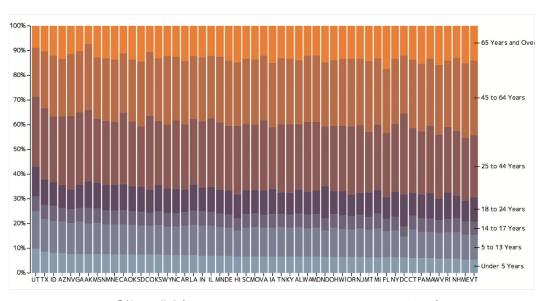
### **Cumulative Bar Chart**



% [d3.garally] (https://github.com/mbostock/d3/wiki/Gallery)

line with ratio (cumulative)

### % Cumulative Bar Chart



% [d3.garally] (https://github.com/mbostock/d3/wiki/Gallery)



# SYDNEY