

# **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

---

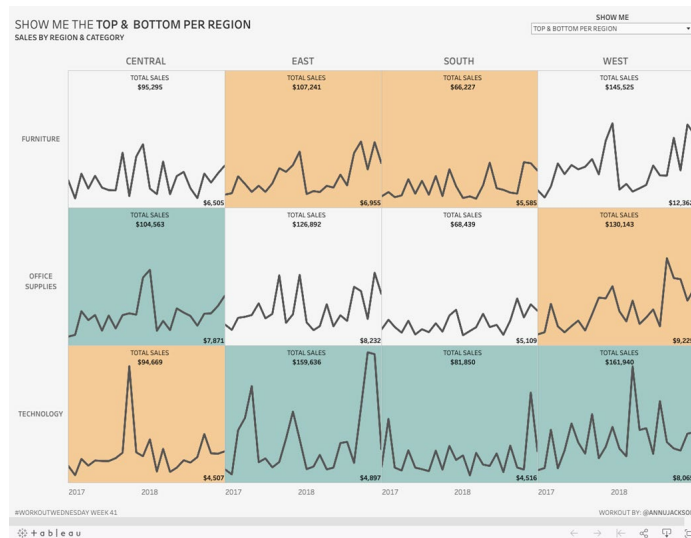
- EDA processes
  - Testing assumptions
  - Selecting models
  - Validating regression model
  - Selecting estimator
  - Identifying relationships
  - Detecting outlier
  - etc.
- Driven by the statistical features of datasets.

NEED  
hypothesis

# Exploratory Data Analysis (cont.)

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

*NOT distract viewer*



<https://public.tableau.com/profile/ann.jackson#!/vizhome/WorkoutWednesdayWeek41-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

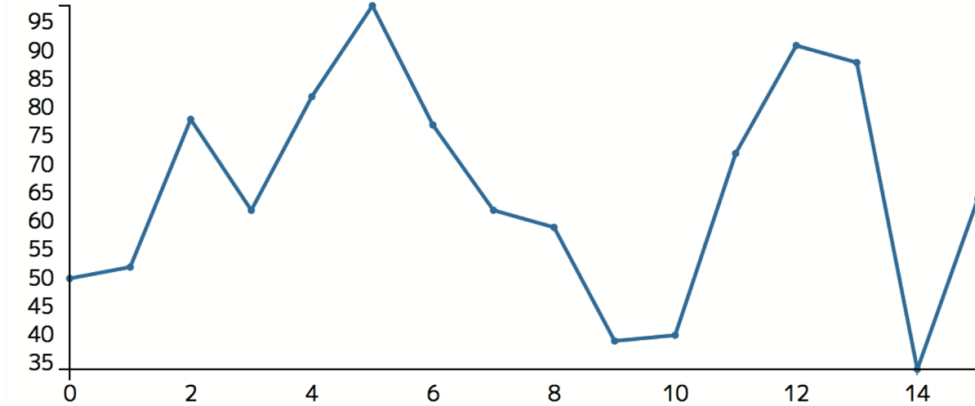
---

Basic		Line graph
Emphasize the amount		Area graph
Change of composition		Stacked Area graph
Change of % composition		% Stacked Area graph
A pair comparison		Slope chart
Compare multiple trends/changes		Spark line

# Line Graph/Chart

---

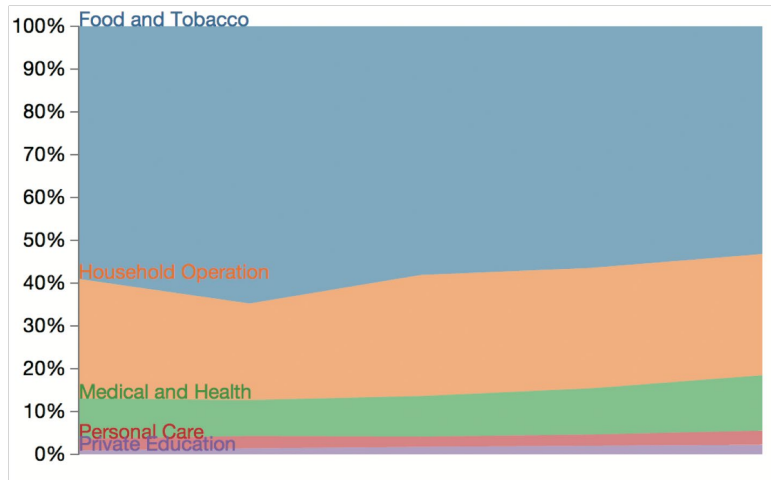
Peak  
cyclical trend  
slope



- Connected points

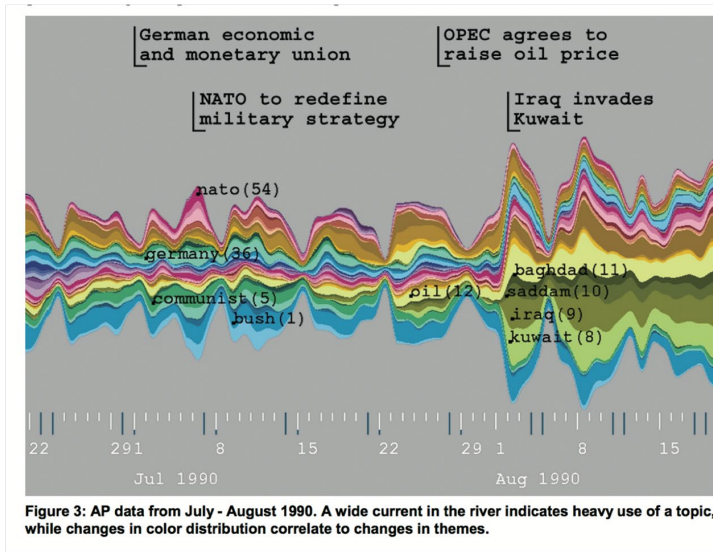
# Area Graph

---



- Area

# Theme River (Area Graph)

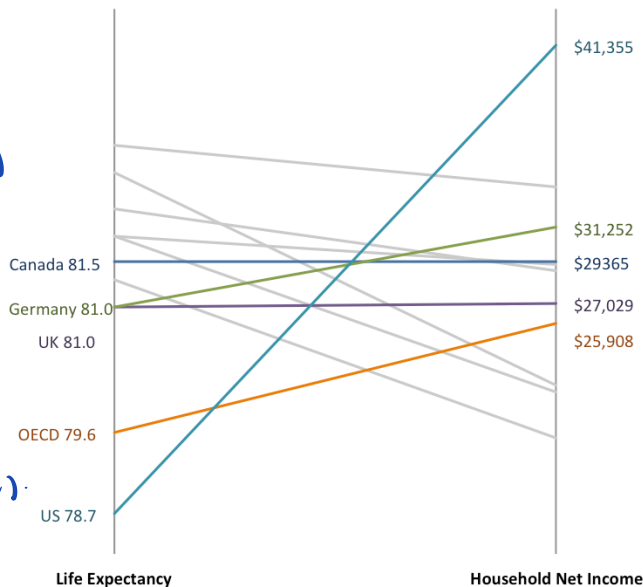


※ [Visualizing Theme Changes over Time]

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

# Slope Chart

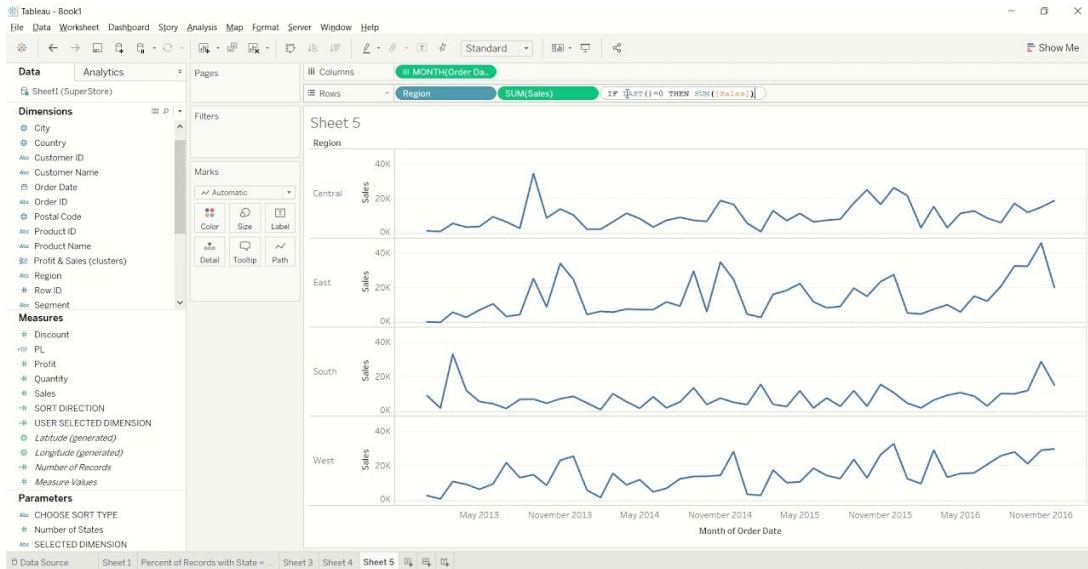
Income and Life Expectancy Around the World



parallel coordinate plot (PCP)  
each data point → line  
mainly look at slope  
Drawback:  
user will typically focus on  
two adjacent coordinate  
| | |  
a b c  
(can't look at a&c).

# Spark Line

*cyclical trend*








<https://i.ytimg.com/vi/Cg6UIVn0v8A/maxresdefault.jpg>

# Attribute analysis/comparison

Compare indices for each attribute

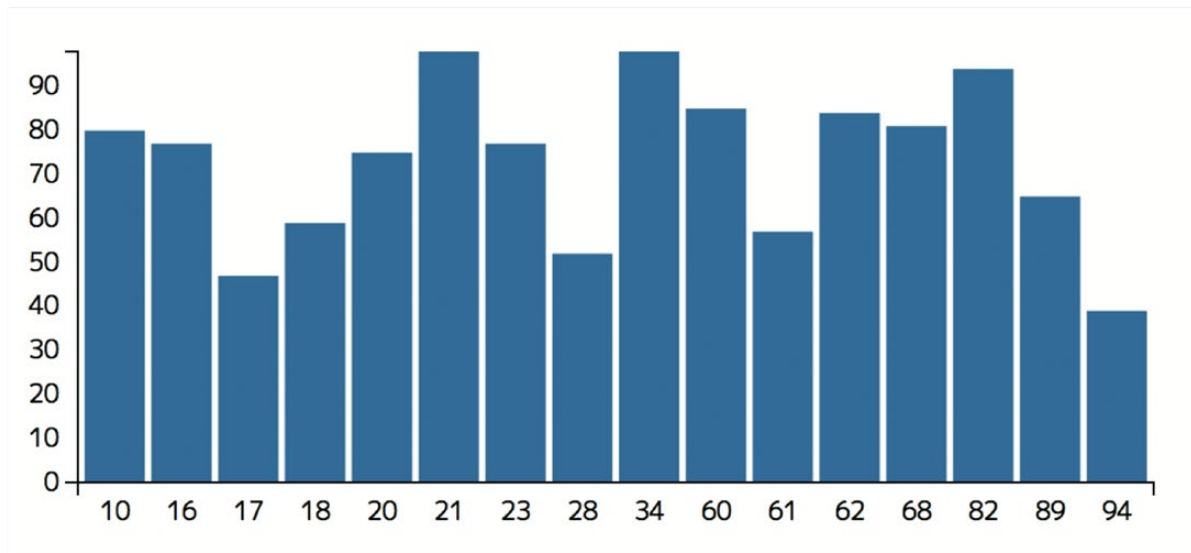
- Visualize/compare
  - Large/small
  - Good/bad

# Basic use of various visualisation for Attribute analysis/comparison

Basic		Bar graph
Show multiple indices		Colored Bar graph
Combine multiple indices		Bullet chart
Overview of multiple indices		Tree map
Show actual values of indices		Score card

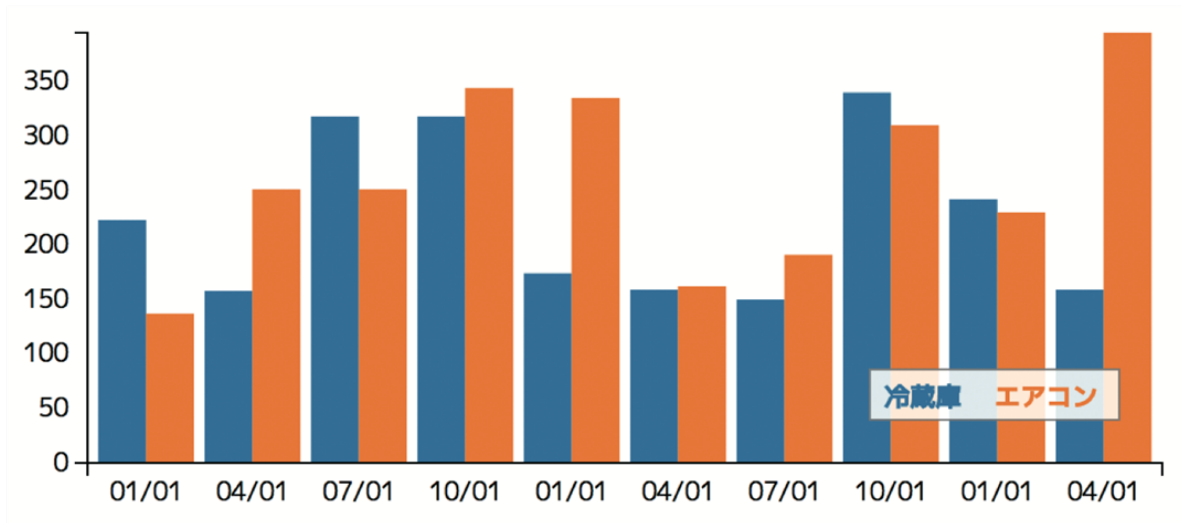


# Bar Graph

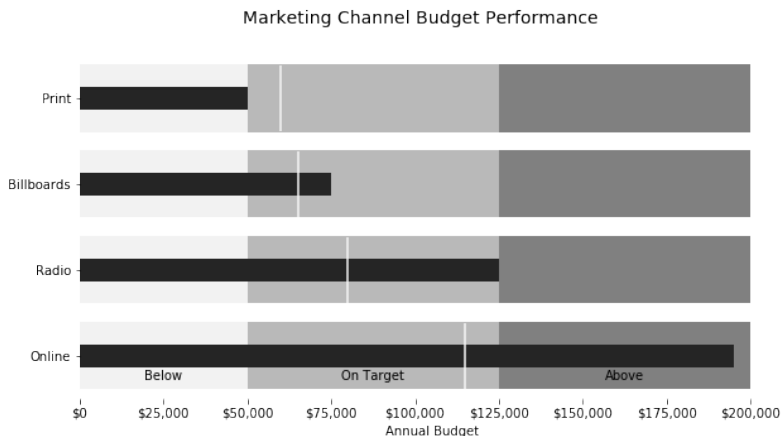


- Line

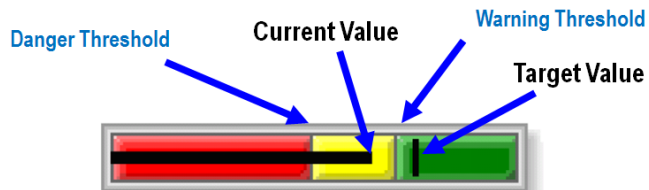
# Bar Chart



# Bullet Chart

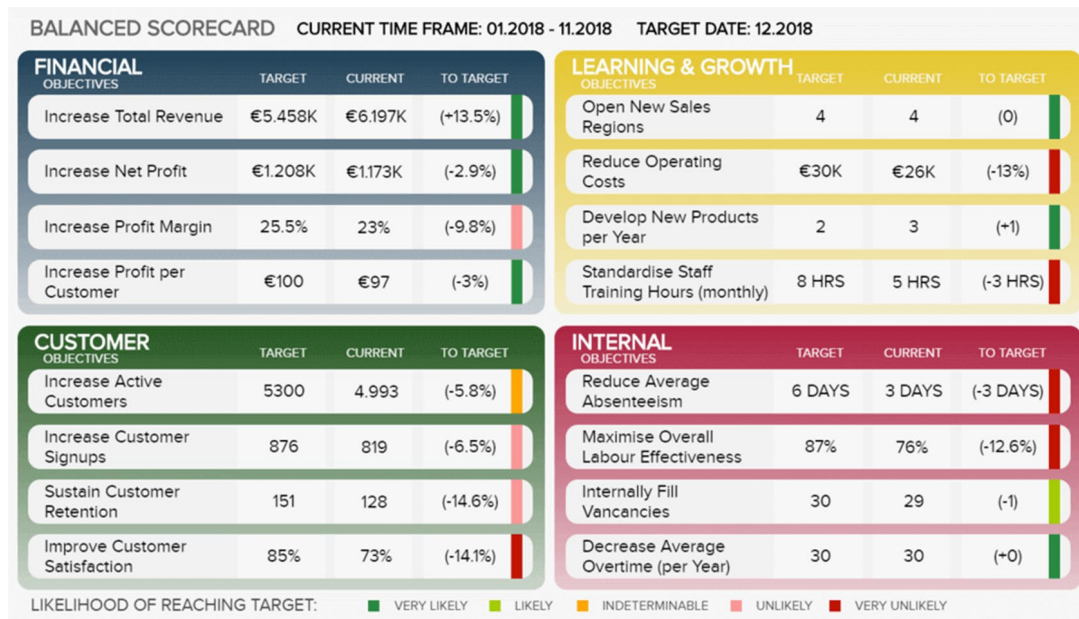


<https://pbpython.com/bullet-graph.html>



<https://glean.info/how-to-select-the-best-type-of-chart-to-visualize-your-data/>

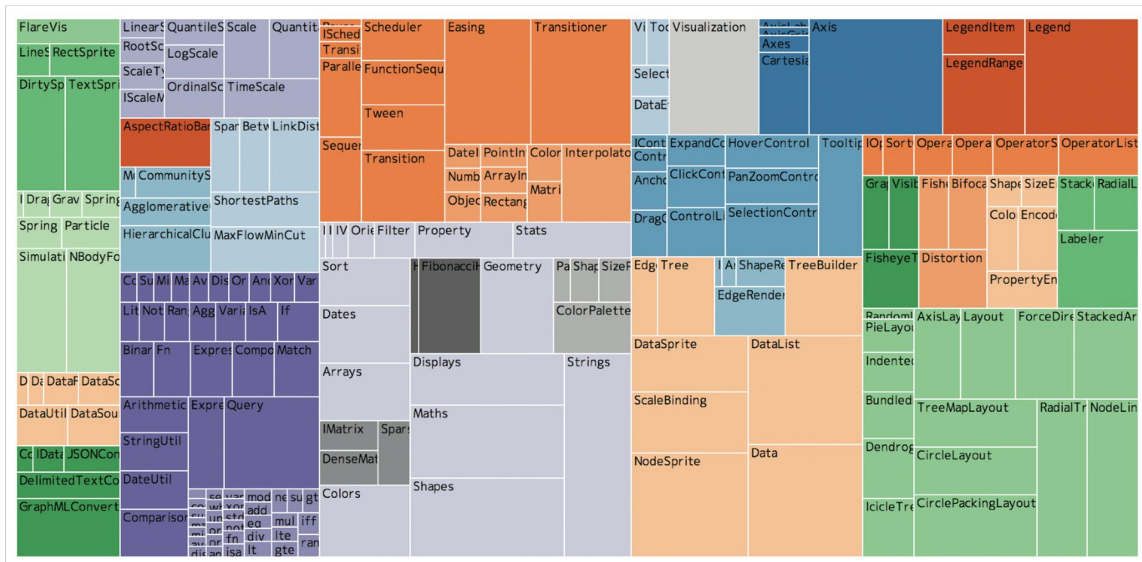
# Scorecard



<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

---

Basic		Bar graph
Temporal change		Bump chart

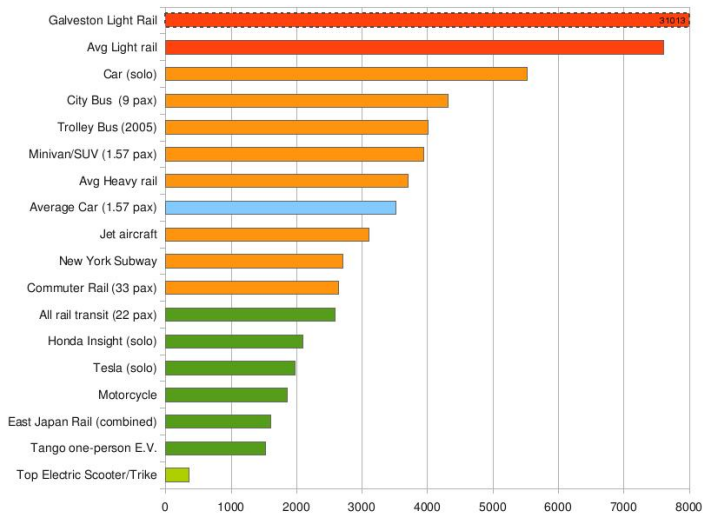
# Bar Graph

0 (0)



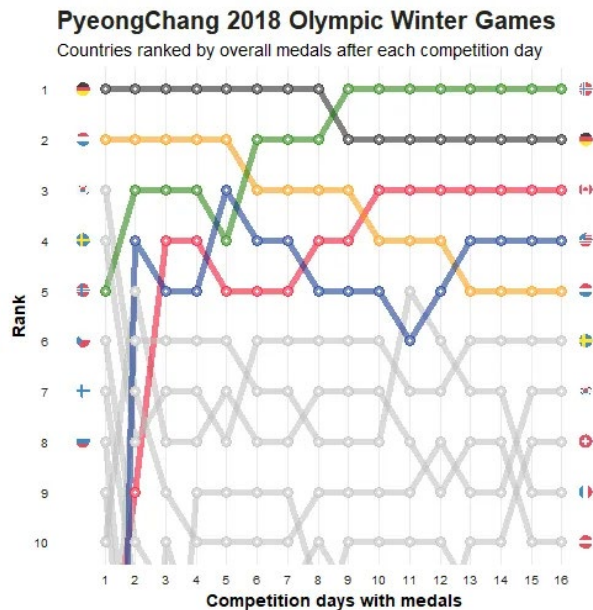
## USA Transportation Energy Use

BTUs per passenger-mile





# Bump Chart






# Composition analysis

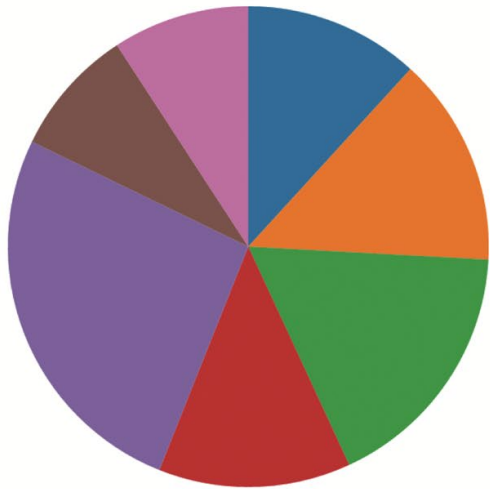
Compare composition of attributes

- Visualise /compare
- Order

# Basic use of various visualisation for Composition analysis

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

# Pie Chart

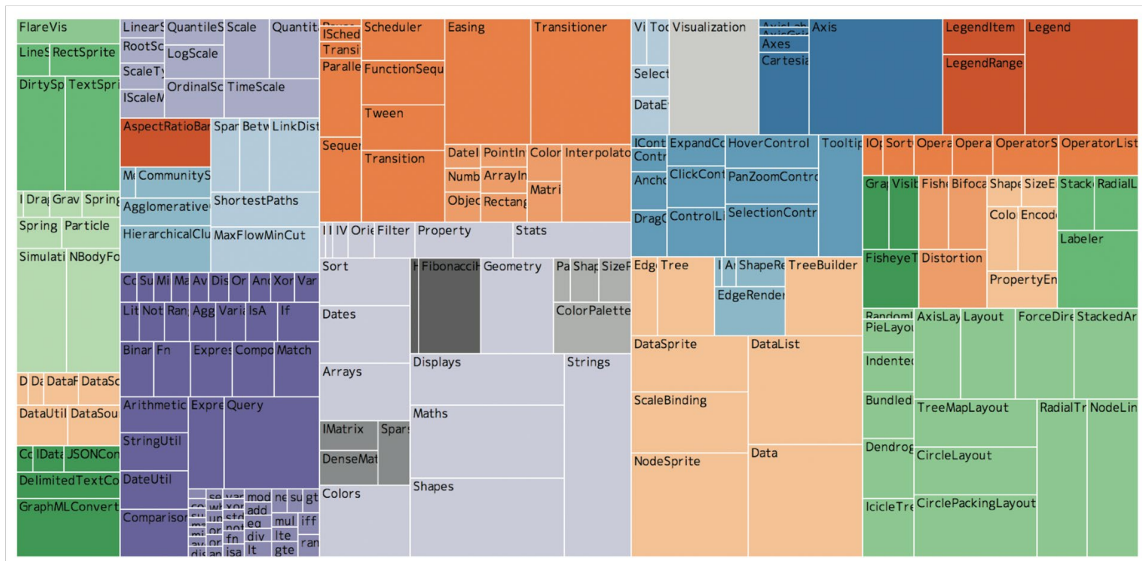


- arc

*Not for many attributes  
⇒ small angles are hard to compare*

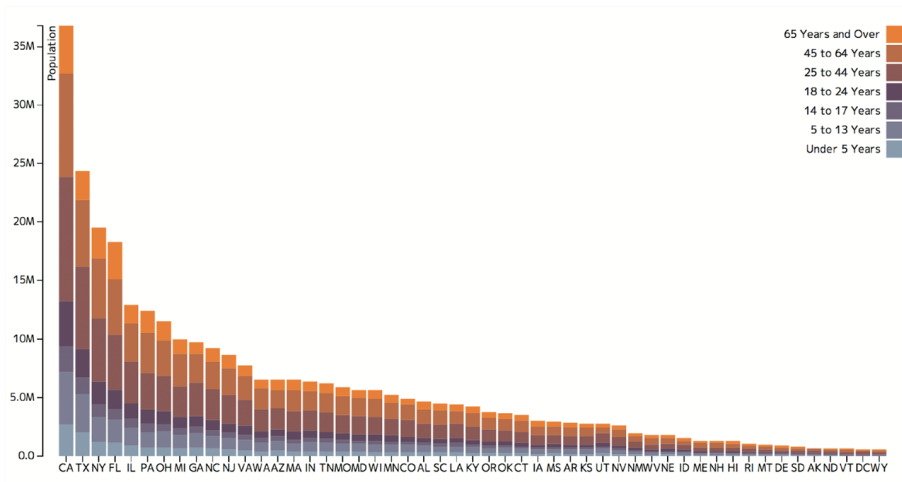
# Tree Map

- Line/Area : size/location/nested



※ 『d3.js garally』 (<http://bl.ocks.org/mbostock/4063582>)

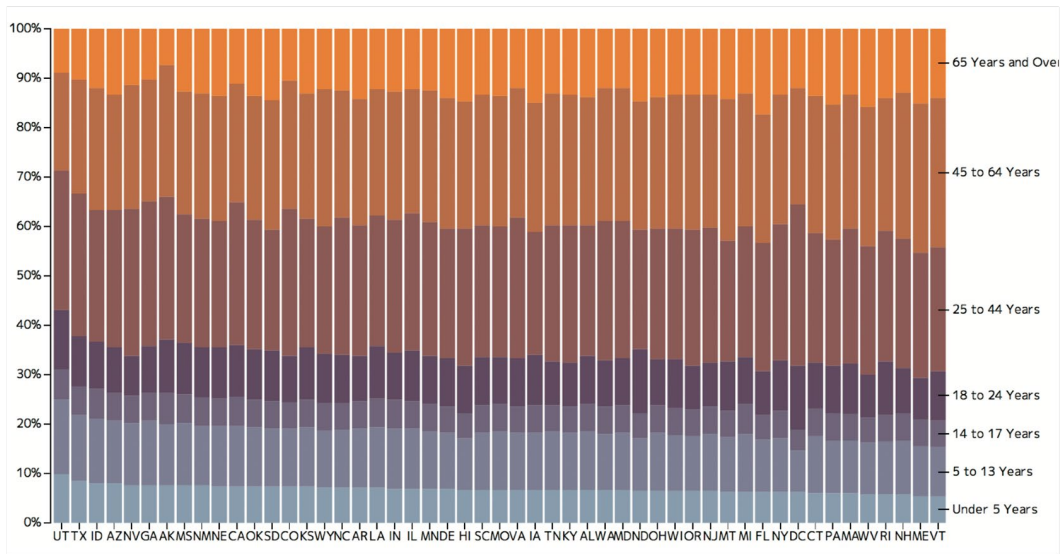
# Cumulative Bar Chart



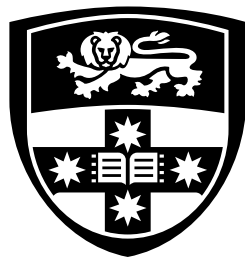
※ 『d3.garally』 (<https://github.com/mbostock/d3/wiki/Gallery>)

- line with ratio (cumulative)

# % Cumulative Bar Chart



※ 『d3.garally』 (<https://github.com/mbostock/d3/wiki/Gallery>)



THE UNIVERSITY OF  
**SYDNEY**