

The Role of Data Visualization in Big Data Analytics

- Chong Ho (Alex) Yu, Ph.D., D. Phil.
- Professor of Behavioral and Applied Science
 - Azusa Pacific University
- Paper Presented at 2020 IM Data Conference



Potential problems in Big Data Analytics/Data Science

- Over-rely on or pre-maturely hand over our judgment to AI and machine learning (Black Box) without examining (exploring) the data pattern carefully.
- Over-plotting: two many observations
- Curse of dimensionality: too many variables







Data mining and Exploratory Data Analysis (EDA)



- Data mining is not entirely new; it is an mining.
- techniques.
- Pattern-seeking
- data emerges (Like a detective).

extension of EDA; EDA is the precursor of data

• A philosophy about how data analysis should be carried out, rather than being a fixed set of

• Skepticism (detective spirit): John Tukey (not Turkey) suggested explore the data in as many ways as possible until a plausible story of the

EDA and data mining

- Same:
 - Data mining is an extension of EDA: it inherits the exploratory spirit; don't start with a preconceived hypothesis or theory.
 - Both heavily rely on data visualization.
- Difference:
 - DM: Use machine learning and resampling
 - DM: More robust
 - DM: Deal with much bigger sample size





Elements of EDA: 4Rs

- Velleman & Hoaglin (1981):
 - Residual analysis
 - Re-expression (data transformation)
 - Resistant
 - Display (revelation, data visualization)







- One of the great inventions of graphical techniques by John Tukey is the boxplot.
 - It is resistant against extreme cases (use the median)
 - It can easily spot outliers.
 - It can check distributional assumption using a quick 5point summary.







Cronbac	ch's a	
		a8642 0 .2 .4 .6 .8
Entire set	0.833	3
Excluded		
Col	α	8642 0 .2 .4 .6 .8
Q1	0.8160	
Q2	0.8181	
R3	0.8074	
R4	0.8102	
R5	0.8285	
Q6	0.8191	
Q7	0.8225	
Q8	0.8238	
R9	0.8171	
R10	0.8124	
R11	0.8059	
R12	0.8228	
Q13	0.8554	
Q14	0.8480	

- data.
- These are 14,819 responses to the **Consideration of Future** Consequences Scale (CFCS).
- When the *n* is huge, even random numbers might appear to be good.
- To demonstrate the problem, Q13 and Q14 are randomly generated.
- The Cronbach Alpha looks good! No bad items!

• Boxplot can be used in mediansmoothing when there are too any





Data visualization can distinguish patterns from noise.

- I want to know whether Q1 is strongly correlated with the total scale (Q2-Q14).
- Item-total correlation
- If the response pattern of Q1 is correlated with the total, then they are internally consistent.
- Too many data points!





• Change the display option to **Boxplot**.

• Now the median of the total by each category of Q1 is shown.

• There is a positive relationship between Q1 and all other items.





 When I use median – smoothing to check the association between Q13 and total without Q13., it was found that there is no relationship.

• The number can fool your mind, but the graph cannot fool your eyes.



What is data visualization?

- The process of exploring and displaying data in a manner that builds a visual analogy in the service of researcher insight and learning.
- A common ground shared by EDA and data mining: The data visualizer should explore the data in as many ways as possible until a plausible story of the data emerges.
- For the purpose of exploration the tool should be interactive and dynamic.





Static vs. dynamic

• Static

- What you see is what you get. The graph is frozen. After the graph is made, you cannot manipulate the graph (changing the background color or the line width is not considered "data manipulation" because it cannot reveal any insight about the data)

Dynamic

- The data table and different graphic panels are linked. Changing one would change all others.
- You can manipulate the graph to explore the data through different perspectives.





Dynamic graph

- Allow you to ask what-if questions:
 - What if I remove outliers?
 - Auto-refresh the model
- Combine multiple dimensions to see a holistic picture: Switch variables for comparison
- Divide and conquer: Localize the subgroups
- Example: COVID19 by US States





Dynamic graph

- Example: COVID19 by US States
- I am not endorsing or opposing any political party; I just show the data as is.
- The Cook Partisan Voting Index (PVI) is a measurement of how strongly a United States congressional district or states leans toward the Democratic or Republican Party compared to the national average.

PVIRescaled PVILeanR+1919Lean towards RepublicanR+1919Lean towards RepublicanR+1717Lean towards RepublicanR+1515Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1313Lean towards RepublicanR+1411Lean towards RepublicanR+1111Lean towards RepublicanR+1211Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+414Lean towards RepublicanR+314Lean towards RepublicanR+314Lean towards RepublicanR+414Lean towards Republican <th></th> <th></th> <th></th>			
R+1919Lean towards RepublicanR+1919Lean towards RepublicanR+1717Lean towards RepublicanR+1515Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1313Lean towards RepublicanR+1111Lean towards RepublicanR+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+1-1Lean towards RepublicanR+214Lean towards RepublicanR+33Lean towards RepublicanR+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards Democratic <th>PVI</th> <th>Rescaled PVI</th> <th>Lean</th>	PVI	Rescaled PVI	Lean
R+1919Lean towards RepublicanR+1717Lean towards RepublicanR+1515Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1313Lean towards RepublicanR+1111Lean towards RepublicanR+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+133Lean towards RepublicanR+99Lean towards RepublicanR+143Lean towards RepublicanR+155Lean towards RepublicanR+143Lean towards RepublicanR+155Lean towards RepublicanR+133Lean towards RepublicanR+22Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanC+1-1Lean towards Democratic <td>R+19</td> <td>19</td> <td>Lean towards Republican</td>	R+19	19	Lean towards Republican
R+1717Lean towards RepublicanR+1515Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1313Lean towards RepublicanR+1111Lean towards RepublicanR+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+131Lean towards RepublicanR+21Lean towards RepublicanR+131Lean towards RepublicanR+22Lean towards RepublicanR+131Lean towards RepublicanR+21Lean towards Republican <td>R+19</td> <td>19</td> <td>Lean towards Republican</td>	R+19	19	Lean towards Republican
R+1515Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1313Lean towards RepublicanR+1111Lean towards RepublicanR+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards Democratic </td <td>R+17</td> <td>17</td> <td>Lean towards Republican</td>	R+17	17	Lean towards Republican
R+1515Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1313Lean towards RepublicanR+1111Lean towards RepublicanR+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanCven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1<	R+15	15	Lean towards Republican
R+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1313Lean towards RepublicanR+1111Lean towards RepublicanR+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+1-1Lean towards RepublicanChen0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1 <t< td=""><td>R+15</td><td>15</td><td>Lean towards Republican</td></t<>	R+15	15	Lean towards Republican
R+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1313Lean towards RepublicanR+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanChen0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+14	14	Lean towards Republican
R+1414Lean towards RepublicanR+1414Lean towards RepublicanR+1313Lean towards RepublicanR+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+14	14	Lean towards Republican
R+1414Lean towards RepublicanR+1313Lean towards RepublicanR+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+14	14	Lean towards Republican
R+1313Lean towards RepublicanR+1111Lean towards RepublicanR+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+14	14	Lean towards Republican
R+1111Lean towards RepublicanR+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+55Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+13	13	Lean towards Republican
R+1111Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+55Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+22Lean towards RepublicanR+11Lean towards RepublicanR+22Lean towards RepublicanR+11Lean towards RepublicanR+21Lean towards RepublicanEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+11	11	Lean towards Republican
R+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+11	11	Lean towards Republican
R+99Lean towards RepublicanR+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+9	9	Lean towards Republican
R+99Lean towards RepublicanR+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+9	9	Lean towards Republican
R+99Lean towards RepublicanR+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+9	9	Lean towards Republican
R+88Lean towards RepublicanR+88Lean towards RepublicanR+55Lean towards RepublicanR+53Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+32Lean towards RepublicanR+22Lean towards RepublicanEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+9	9	Lean towards Republican
R+88Lean towards RepublicanR+55Lean towards RepublicanR+56Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+8	8	Lean towards Republican
R+55Lean towards RepublicanR+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+8	8	Lean towards Republican
R+55Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+5	5	Lean towards Republican
R+33Lean towards RepublicanR+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanEven0EvenEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+5	5	Lean towards Republican
R+33Lean towards RepublicanR+33Lean towards RepublicanR+22Lean towards RepublicanEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+3	3	Lean towards Republican
R+33Lean towards RepublicanR+22Lean towards RepublicanEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+3	3	Lean towards Republican
R+22Lean towards RepublicanEven0EvenEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+3	3	Lean towards Republican
Even0EvenEven0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	R+2	2	Lean towards Republican
Even0EvenEven0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	Even	0	Even
Even0EvenD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	Even	0	Even
D+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	Even	0	Even
D+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	D+1	-1	Lean towards Democratic
D+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	D+1	-1	Lean towards Democratic
D+1-1Lean towards DemocraticD+1-1Lean towards DemocraticD+3-3Lean towards Democratic	D+1	-1	Lean towards Democratic
D+1 -1 Lean towards Democratic D+3 -3 Lean towards Democratic	D+1	-1	Lean towards Democratic
D+3 -3 Lean towards Democratic	D+1	-1	Lean towards Democratic
	D+3	-3	Lean towards Democratic





Switch variables



Multiple variables

- Using "size" or adding more dimensions into the map is confusing.
- Difficult to see multiple dimensions concurrently
- Curse of dimensionality





Detect association

- More tests? → more confirmed cases?
- A flat regression line
- No relationship nationwide









File Edit Tables Rows Cols D	OE Analyze Graph Six Sigma Tools T	ools Add-Ins View	Window Help		
: 🖼 🔁 🖉 🔚 🖇 🖷 🗳 (S) () - : 📰 11 🚧 - : 🔽 🎗 🍄 🐵 🖓) & P Q + //	ਗ਼ਫ਼ਲ਼੦ੵ		
⊿ 💌 Local Data Filter	🖉 로 Graph Builder				
Clear Favorites 🕶	Undo Start Over Done		~// 08	┃ 📈 🖬 🖿 🚔 📲 🐂 🐂 📕 🛯 🖉 🖉	
27 matching rows	25 Columns			Tot Cases / 1M pop vs Tests / 1M por	
Inverse	Enter column name	◄ م			
💌 Lean (3) 🛛 🗡	🔥 USA State	~		Group X	
Even 3	ul PVI	600	000		
Lean towards Democrat 20	A Rescaled PVI			•	
Lean towards Republic27	Lean				
	Governor Party				
AND	Senate Party			•	
	House Balance	500	000 -		
	2020 Pop.				
	State_link_data				
	rank	~			
	⊿ Points Points				
	Summary Statistic None	× 0.40			
	Error Interval Auto	× d 400	000-	•	
	Interval Style Error Bar	× 1		:	
	Jitter Auto	< Ises/	-	• • •	
	Jitter Limit	ີ ເ			
	Variables	Tc			
	⊿ 💌 Line Of Fit	300	• 000		
	Fit Polynomial	¥			
	Degree Linear	*		• •	
	Confidence 🔽 Fit			•	
	Prediction			•	
	Statistics RMSE	200	000 -		
	R ²			•	
	Equation				
	Variables		-		
	Vanabes V			•	
		Ma	p 200000	300000 400000	500000
		Shaj	pe	Tests/ 1M pop	
		Where	e(Lean = Lean to	owards Republican)	
				L	





















Two or more filters





Two or more filters





Less is more: Binning can be helpful!

 Noise reduction 800 • 2018 PISA data 700 WLE: Weighted Likelihood 600 PV Science **Estimates** 500 400 • PV: Plausible values (test score) 300 Too many data! Overplotting! It 200 obscures us from seeing the 100 relationship between science test performance and teacher-directed instruction.



nnovative Methods with Data Science & Al

Select the variable and classify the data into different bins

—			
File Edit Tables Rows Cols	DOE Analyze Graph Six Sigma Tools Too		
New Columns	🌐 🗄 🖶 🚩 ≽ 🦻 🖕		
Column Selection			
Reorder Columns			
Column Info	haracter 1		
	instruction (WLE) Binned 2		
Standardize Attributes	3		
Preselect Role	WLE) Binned 4		
Formula	Positive affect (WLE) Binned 5		
Label/Uplabel	6		
	/		
Scroll Lock/Unlock	8		
Hide/Unhide	9		
Exclude/Unexclude	11		
Use for Marker	12		
	13		
Recode	14		
Columns Viewer	15		
Utilities	Compress Selected Columns		
Clear Cell Colors	Text to Columns		
Group Columns	Make Indicator Columns		
Ungroup Columns	Combine Columns		
Delete Columns	Make Binning Formula		



Innovative Methods with Data Science & Al

Select the variable and classify the data into different bins



Innovative Methods with Data Science & Al

Binning and median smoothing: Non-linear pattern emerges!





Violin plot and boxplot

 There are three displays in one: raw data, violin plot showing the distribution, and the boxplot showing the 5-point summary.



Innovative Methods with Data Science & Al

Revisit overplotting

- <u>2015 PISA data</u>
- *n* = 54,978
- A big cloud: Overplotting



Overplotting

- One way to "see through" the cloud is using the heat map.
- Limitation:
 - Density is shown by colors only.
 - It hides the raw data.



Nonparametric Bivariate Density

- The density of data points is represented by both colors and contour lines.
- Its interpretation is straight– forward.



nnovative Methods with Data Science & Al

Contour plot

- How about using the contour lines without showing the raw data?
- It could be confusing and misleading.
- The appearance of a one-dimensional histogram is tied to its binwidth or bandwidth.
- By the same token, the bandheight of the isolines determines how a contour plot appears.
- The two contour plots are based on the same data and they use different bandheights!



Innovative Methods with Data Science & Al

Can Parallel coordinate overcome the curse of dimensionality?

- By joining individual observations in parallel coordinate plot, you can look at multiple dimension.
- By you are facing the problem of over-plotting!



Curse of dimensionality

- Using colors, shape to represent different dimensions.
- Can be confusing.





Extreme examples





imda Innovative Methods with Data Science & Al

Conclusion

- Exploratory data analysis utilizing data visualization should work hand in hand with AI and machine learning.
- Data exploration requires dynamic, not static graphic.
- Overplotting can be overcome by using boxplot (median smoothing), heatmap, non-parametric contour...etc.
- Curse of dimensionality can be overcome by switching variables, local filters...etc.
- Too much data reduction is problematic (e.g. contour)
- No data reduction is also problematic (e.g. parallel coordinate plot)



For more info

 Google search for
"Oxford," "Exploratory data analysis" or/and my name.

Oxford Bibliographies

Your Best Research Starts Here

Browse by Subject

How to Subscribe

Free⁻





Introduction

Exploratory data analysis (EDA) is a strategy of data analysis that emphasizes maintaining an open mind to alternative possibilities. EDA is a philosophy or an attitude about how data analysis should be carried out, rather than being a fixed set of techniques. It is difficult to obtain a clear-cut answer from "messy" human phenomena, and thus the exploratory character of EDA is very suitable to psychological research. This research tradition was founded by John Tukey, who often relates EDA to detective work. In EDA, the role of the researcher is to explore the data in as many ways as possible until a plausible "story" emerges. A detective does not collect just any information. Instead, he or she collects clues related to the central question of the case. By the same token, EDA is not

Trial	S	Sign

Sign in

Exploratory Data Analysis

Chong Ho Yu

LAST MODIFIED: 29 NOVEMBER 2017 DOI: 10.1093/OBO/9780199828340-0200



For more info



Chong Ho Yu

Dancing with the Data: The Art and Science of Data Visualization

Dancing with the Data: The Art and Science of Data Visualization Paperback – June 19, 2014

by Chong Ho Yu (Author)

> See all formats and editions

Paperback \$88.00 <prime</pre>

\$00.00 Prime

1 Used from \$86.78 11 New from \$75.67





Contact Info

- Chong Ho (Alex) Yu
- <u>chonghoyu@gmail.com</u>
- cyu@apu.edu
- www.creative-wisdom.com/pub/pub.html





