Data Visualization（I）

崔雪彬
南京大学数字经济与管理学院
2024年10月8日

Data Visualization

- 拿到一个数据，用数据可视化展示什么?
- 什么类型的数据用什么样的图形展示?
- 怎么用软件来实现要展示的图形?

Descriptive Statistics

- Variables
- Definitions
- Number of observations
- Mean
- Standard deviation
- Min
- Max

Pan I.D.	Expend $	Income	HH Size	IPT	Quantity	Brand 1	Brand 2	Brand 3	Brand 4	Feature 1	Feature 2	Feature 3	Feature 4	Price 1
1	40.9	9	2	5	2	0	0	0	1	0	0	0	0	0.108
1	16.81	9	2	5	2	0	1	0	0	0	0	0	0	0.108
1	4.06	9	2	1	2	0	1	0	0	0	0	0	0	0.108
1	34.46	9	2	4	2	0	1	0	0	0	0	0	0	0.108
1	8.39	9	2	7	2	0	1	0	0	0	0	0	0	0.125
1	60.99	9	2	3	2	0	1	0	0	0	0	0	0	0.108
1	20.79	9	2	4	1	0	1	0	0	0	0	0	0	0.103
1	33.95	9	2	8	2	0	0	0	1	0	0	0	0	0.108
2	3.69	7	2	5	1	1	0	0	0	0	0	0	0	0.108
2	50.93	7	2	2	1	1	0	0	0	0	0	0	0	0.108
2	18.35	7	2	2	1	1	0	0	0	0	0	0	0	0.108
2	31.91	7	2	2	1	1	0	0	0	0	0	0	0	0.108

Descriptive Statistics

Table 2. Variable Definitions and Descriptive Statistics

Variable	Operationalization	Mean	Std. dev.	Min	Max
Frequency of purchases	Number of purchase transactions in the time period	0.81	1.58	0	101
Quantity of purchases	Number of items bought in the time period	1.50	3.47	0	542
Value of purchases	Monetary value of purchases in the time period ($)	43.94	117.61	0	8,993.81
Frequency of returns	Number of return transactions in the time period	0.11	0.48	0	51
Quantity of returns	Number of items returned in the time period	0.15	0.73	0	120
Value of returns	Monetary value of returns in the time period ($)	3.56	29.73	0	5,150.63
Net monetary value (NMV)	(Monetary value of purchases − Monetary value of returns) ($)	40.37	107.22	−1,937	8,993.81
App Adopters (TREAT)	Dummy indicating whether the shopper adopted the app (= 1) or not (= 0)	0.47	0.50	0	1
Time Period (POST)	Dummy indicating whether the period is before (= 0) or after (= 1) app launch	0.50	0.50	0	1
Recency	Number of days since the shopper's last purchase	193.60	293.71	0.08	1,493.23
Age	Age of the shopper in years at the start of the data period	31.98	10.78	11.00	115
Gender	Gender of the shopper (female = 2, male = 1, unknown = 0)	0.67	0.63	0.00	2.00
Distance to nearest store	Distance in miles between the geographical centers of the shopper's and the nearest store's zip codes	3.90	7.21	0.00	574.48
Number of stores	Number of the focal retailer's stores in the shopper's zip code	0.56	0.72	0.00	4
Loyalty program level	Dummy indicating whether the shopper is enrolled in the basic (= 0) or professional (= 1) membership category on app introduction date	0.16	0.37	0.00	1
Area population	Population of the shopper's zip code based on 2010 U.S. census	31,437	19,090	6.00	113,916
Competitor stores	Number of competing stores in the shopper's zip code	0.29	0.52	0.00	4
No. of unique products	Number of unique stock-keeping units that the shopper buys	1.21	2.71	0.00	313
No. of unique categories	Number of unique categories that the shopper buys	0.84	1.53	0.00	29
Pct. of top 100 products	Spending on the top 100 products/Total spending	0.21	0.34	0.00	1
Pct. of top 500 products	Spending on the top 500 products/Total spending	0.47	0.42	0.00	1
Cell towers	Number of cell towers in the shopper's zip code	6.99	6.93	0.00	52
Precipitation	Average precipitation level in the shopper's zip code as reported by the NOAA in millimeters in June 2014	102.82	74.81	0.00	498.5
Temperature	Average air temperature of the shopper's zip code in Celsius as measured by the NOAA in June 2014	23.53	3.57	0.96	36.38
Download speed	Percentage of the population in the shopper's county with download speeds less than 6,000 KBps as reported by the FCC in June 2014	0.01	0.02	0.00	0.81
Wireless access	Percentage of the population in the shopper's county with access to three or more wireless providers as reported by the FCC in June 2014	0.12	0.02	0.00	1

Notes. The statistics for the outcome variables (e.g., frequency, quantity, and value of purchases and returns) are averaged over the 36-month data period. NOAA, National Oceanic and Atmospheric Administration.

Source: Narang, U., V. Shankar. 2019. Mobile app introduction and online and offline purchases and product returns. Marketing Science 38(5) 756–772.

Descriptive Statistics

Table 3. Model-Free Evidence: Mean Statistics

Variable	Treated pre period	Treated post period	Control pre period	Control post period	Matched controls pre period	Matched controls post period
Frequency of purchases	1.211	1.368	0.407	0.363	1.197	0.963
Quantity of purchases	2.256	2.524	0.570	0.665	2.214	1.737
Value of purchases	63.60	75.84	21.56	20.93	63.10	52.09
Frequency of returns	0.177	0.196	0.047	0.041	0.169	0.129
Quantity of returns	0.238	0.258	0.062	0.053	0.226	0.167
Value of returns	5.369	6.406	1.584	1.452	5.257	4.146
Net monetary value of purchases	58.236	69.438	19.975	19.481	57.85	47.94
Frequency of purchases—online	0.019	0.033	0.006	0.007	0.020	0.017
Quantity purchased—online	0.028	0.050	0.009	0.010	0.028	0.025
Value of purchases—online	1.41	1.997	0.495	0.425	1.571	1.085
Frequency of purchases—stores	1.192	1.335	0.400	0.356	1.177	0.946
Quantity purchased—stores	2.228	2.474	0.741	0.655	2.186	1.713
Value of purchases—stores	62.20	73.85	21.06	20.51	61.53	51.01

Notes. The pre period (post period) statistics are monthly averages over the 18-month period before (after) app launch across shoppers; online purchases in the post period for the treated include purchases on the mobile site after clicking checkout in app. Matched controls are nonadopters similar to adopters based on nearest neighbor matching using preperiod covariates.

Source: Narang, U., V. Shankar. 2019. Mobile app introduction and online and offline purchases and product returns. Marketing Science 38(5) 756–772.

Descriptive Statistics

Table 1. Descriptive Statistics and Correlations.

	Mean	SD	Min	Max	1	2	3	4	5	6
1. BCEmotion	.103	.140	.000	1.000	1.000					
2. ViewerEmotion	.256	.356	.000	1.000	.081**	1.000				
3. AmountTips	.169	3.098	.000	396	.026**	.023**	1.000			
4. NumLikes	.007	.022	.000	.662	.018**	.093**	.012**	1.000		
5. LenCmt	.013	.017	.000	1.116	.103**	.299**	.023**	.141**	1.000	
6. NumViewers	4.797	5.010	1.000	119	.088**	.212**	.053**	.303**	.406**	1.000

**$p < .01$.
Notes: N(obs) = 105,453 (only from minute 3 to minute 120).

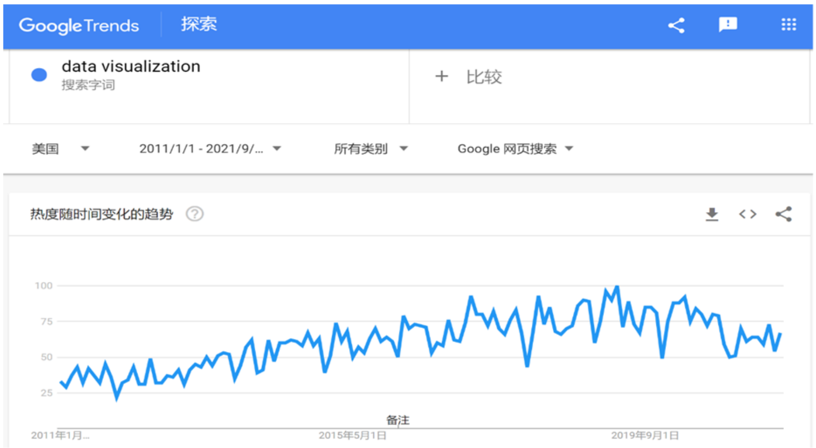
Source: Lin, Yan, Dai Yao, and Xingyu Chen (2021), "Happiness Begets Money: Emotion and Engagement in Live Streaming," Journal of Marketing Research, 58(3), 417-438.

What is Data Visualization?

- **Data visualization** gives us a clear idea of what the information means by giving it visual context through ***maps or graphs***. This makes the data more natural for the human mind to comprehend and therefore makes it easier to ***identify trends, patterns, and outliers*** within large data sets.

Baidu Index

Google Trend



Why Data Visualization?

"Human brains process visuals 60,000 times faster than they do text."

- Data visualization is important in data science because it helps us **make data 'speak' and provide all the hidden details it covers**. It also helps perform the exploratory analysis quickly, giving a massive boost to data science projects and effective decision-making.

When Data Visualization?

- **Preprocessing**
  - Supports data cleaning
  - Supports variable transformation and selection
- **Data exploration**
  - Explore the relationship between two variables
  - Present preliminary/model free results
- **Results presentation**

Examples: Distribution

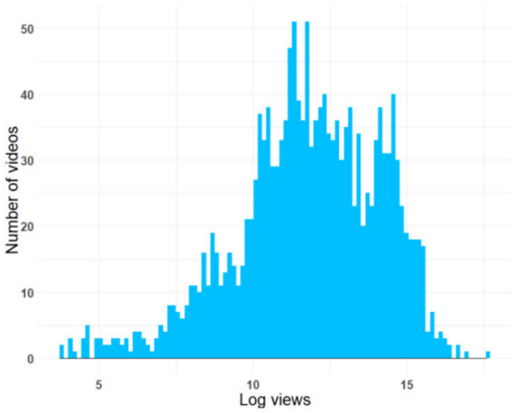


Figure 1 – Distribution of log view count

Source: Prashant Rajaram , Puneet Manchanda (2020), Video Influencers: Unboxing the Mystique.

Examples: Distribution

- (a) engagement = (# comments/# views), (b) popularity = (# likes/# views), and (c) likeability = (# likes/# dislikes)

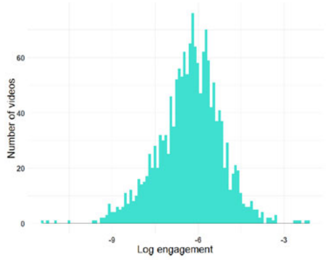


Figure 2a – Distribution of Log Engagement

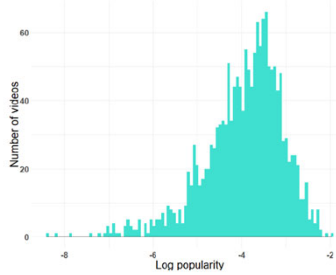


Figure 2b – Distribution of Log Popularity

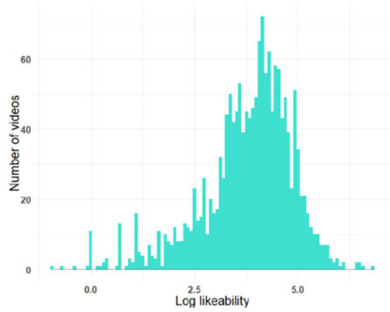


Figure 2c – Distribution of Log Likeability

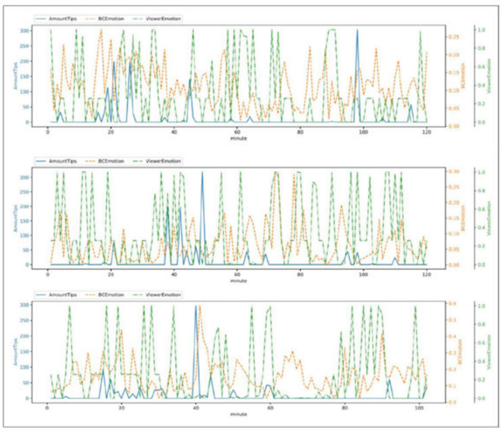
Source: Prashant Rajaram , Puneet Manchanda (2020), Video Influencers: Unboxing the Mystique.

Examples: Time Series

- Three live streams

Broadcaster emotions

Amount of Tips



Viewer emotions

Figure 2. Dynamics of variables in sample live streams.

Source: Lin, Yan, Dai Yao, and Xingyu Chen (2021), "Happiness Begets Money: Emotion and Engagement in Live Streaming," *Journal of Marketing Research*, 58(3), 417-438.

Examples: Time Series

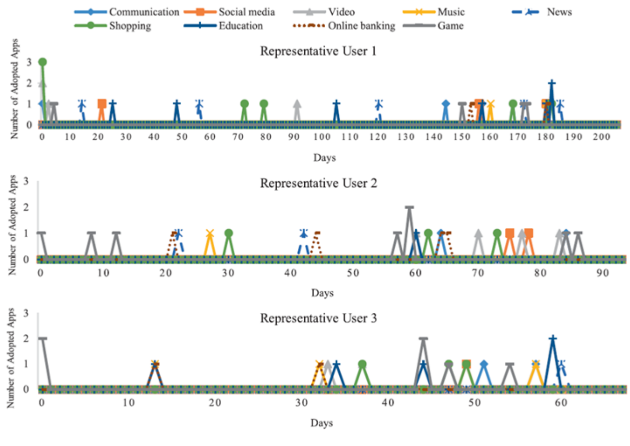
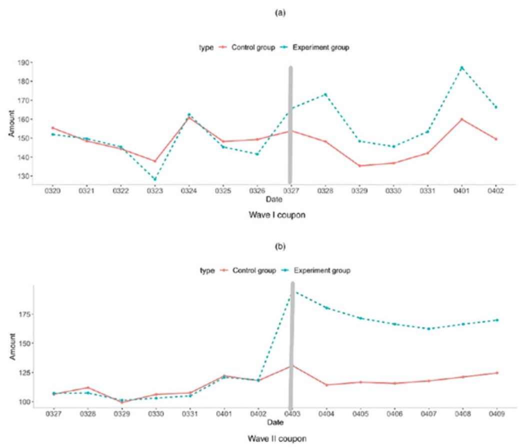


Fig. 3. Timelines of Representative Users' App Adoption.

Source: Xiaochi Sun, Cui, Xuebin*, Yacheng Sun (2023). Understanding the sequential interdependence of mobile app adoption within and across categories, *International Journal of Research in Marketing*, 40(3), 659-678

Examples: Model Free Evidence

Figure 2. (Color online) Time Trend of Spending



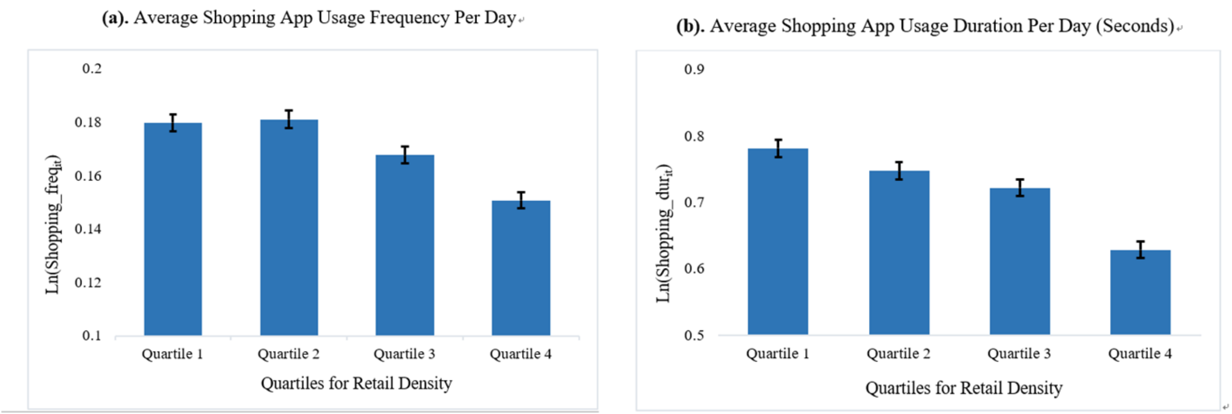
Note. The solid line represents spending by the control group, and the dotted line represents spending by the treatment group.

Treatment group

Matched control group

Qiao Liu, Qiaowei Shen, Zhenghua Li, Shu Chen (2021) "Stimulating Consumption at Low Budget: Evidence from a Large Scale Policy Experiment Amid the COVID-19 Pandemic" *Management Science*, 67(12), 7291-7307.

Examples: Model Free Evidence



Cui, Xuebin, Ting Zhu, Yubo Chen (2022). Where you live matters: The impact of local retail density on mobile shopping, conditional accepted *at Journal of Retailing.*

Examples: Model Free Evidence

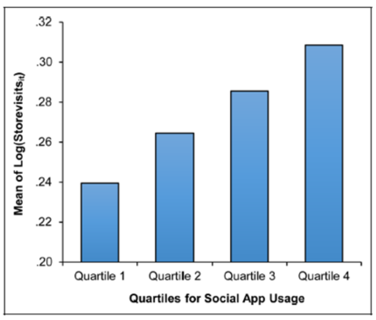


Figure 1. Relationship between social app usage and offline store visit frequency.
Notes: Quartiles 1–4 refer to the four quartiles of social app usage duration in all consumer samples (7,354). Quartile 1 represents the quartile with the lowest social app usage, and Quartile 4 represents the quartile with the highest social app usage. The natural log of the offline store visit frequency was taken after adding 1.

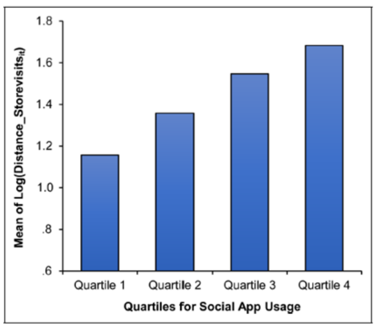


Figure 2. Relationship between social app usage and distance traveled to stores.
Notes: Quartiles 1–4 refer to the four quartiles of social app usage duration in all consumer samples (7,354). Quartile 1 represents the quartile with the lowest social app usage, and Quartile 4 represents the quartile with the highest social app usage. The natural log of the average daily distance traveled to offline retail stores was taken after adding 1.

Source: Cui, Xuebin, Yacheng Sun, Yubo Chen, Banggang Wu (2022). The impact of mobile social app usage on offline shopping store visits, *Journal of Interactive Marketing*, 57(3), 457-471.

Examples: Group Comparison

Who are the Early Mobile Adopters?

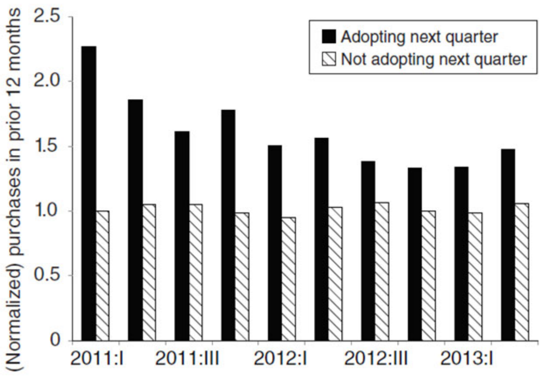


FIGURE 2. MOBILE ADOPTERS VERSUS NONADOPTERS

Source: Einav, L., Levin, J., Popov, I., & Sundaresan, N. (2014). Growth, adoption, and use of mobile E-commerce. *American Economic Review*, 104(5), 489-494.

Examples: Relationship

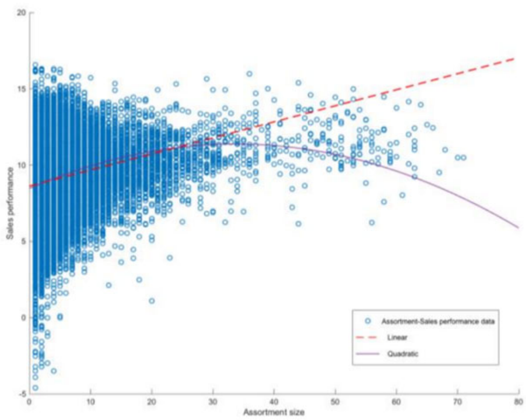
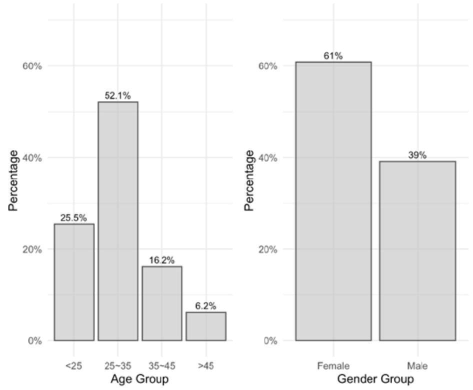


Figure 5. Actual Pattern Between Sales and Assortment size

Source: He, Y., Guo, X. and Chen, G. (2019). Assortment Size and Performance of Online Sellers: An Inverted U-Shaped Relationship, *Journal of the Association for Information Systems* 20(10): 1503–1530.

Examples: Group Comparison

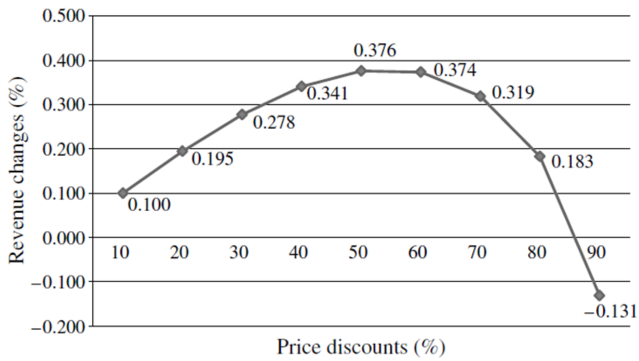
Figure A1 Adoption Rate across Age and Gender Observed in the Treatment Group



Source: Sun et al.(2021) , The Effect of Voice AI on Consumer Purchase and Search Behavior

Examples: Optimization

Figure 6 Optimal Price Discount to Maximize App Sales Revenues



Ghose and Han: Estimating Demand for Mobile Applications in the New Economy, *Management Science* 60(6), pp. 1470–1488.

Examples: Prediction

Figure 4 Predicted Share of Purchases Tracks Actual Share Closely for the Calibration Sample over the Calibration Period

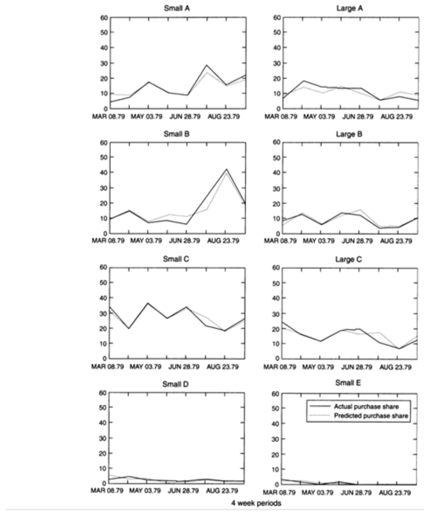
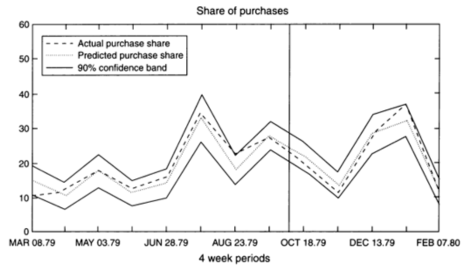


Figure 5 Tracking of Small A Purchase Share in a Holdout Sample of 100 Customers Shows Actual Share Falling Almost Entirely Within the Confidence Band



Notes. The vertical line separates the calibration time period from the forecast period. During the forecast period loyalty variables are constructed from simulated purchases.

Source: Guadagni, P. M., & Little, J. D. (1983). A logit model of brand choice calibrated on scanner data. *Marketing science*, 2(3), 203-238.

Examples: Word Cloud

Figure 2 Word Cloud by Decade

Note. Words that were used more frequently in a particular decade are shown in larger type.

- Based on the paper keywords

Source: Carl F. Mela, Jason Roos, Yiting Deng, (2013) Invited Paper—A Keyword History of Marketing Science. *Marketing Science* 32(1):8-18.

Word Cloud: Keywords

- Based on the paper keywords
- 2019-
- Journal of Consumer research
- Journal of Marketing
- Journal of Marketing Research
- Marketing Science
- Management Science (Marketing department)

Conclusion

- We need data visualization because the human brain is not well equipped to devour so much raw, unorganized information and turn it into something usable and understandable.
- We need graphs and charts to communicate data findings so that we can identify patterns and trends to gain insight and make better decisions faster.
- Data visualization could help preprocessing, data exploration and results presentation.