Python for Data Science

Data Visualization

Dr. Ilkay Altintas

Role of Visualization

- Define data visualization
- Describe the value of data visualization for data science

"The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill in the next decades… Because now we really do have essentially free and ubiquitous data." (emphasis mine)

- Hal Varian*, Google's Chief Economist

* Interview with James Manyika. Hal Varian on how the Web challenges managers. McKinsey&Company. Oct. 2008.

Defining Visualization

"The use of computer-supported, interactive, visual representations of abstract data to amplify cognition." [Card et al., 1999]

"The representation and presentation of data to facilitate understanding." [Kirk, 2016]

- Card, S. and Mackinlay, J. and Shneiderman, B., Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999.
- Kirk, A. Data Visualisation: A handbook for Data Driven Design. SAGE publications, 2016.

Defining Visualization

"The use of computer-supported, interactive, visual representations of abstract data to amplify cognition." [Card et al., 1999]

"The representation and presentation of data to facilitate understanding." [Kirk, 2016]

- Card, S. and Mackinlay, J. and Shneiderman, B., Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999.
- Kirk, A. Data Visualisation: A handbook for Data Driven Design. SAGE publications, 2016.

Defining Visualization

"The use of computer-supported, interactive, visual representations of abstract data to amplify cognition." [Card et al., 1999]

"The representation and presentation of data to facilitate understanding." [Kirk, 2016]

- Card, S. and Mackinlay, J. and Shneiderman, B., Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999.
- Kirk, A. Data Visualisation: A handbook for Data Driven Design. SAGE publications, 2016.

Example 1 – raw data

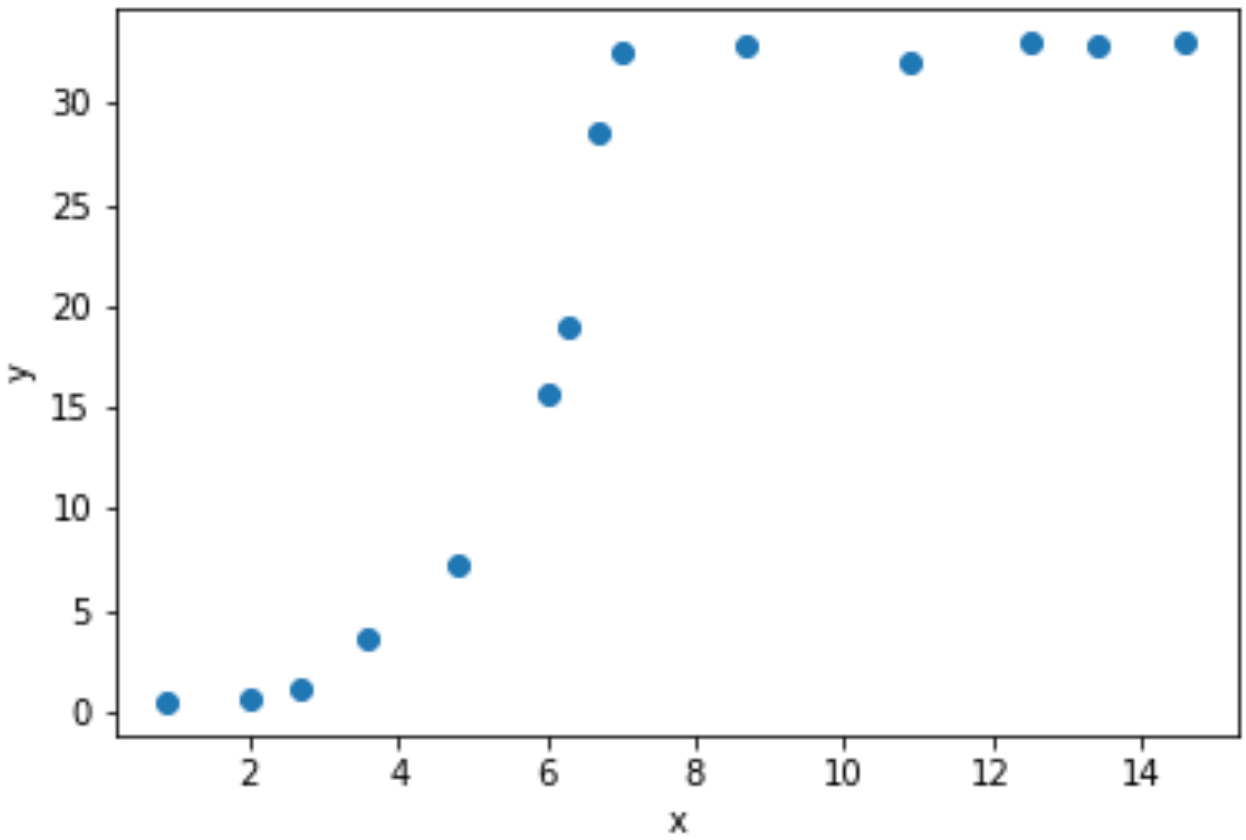
X	Y
0.9	0.5
2.7	1.1
6.7	28.6
10.9	32.8
6.0	15.7
6.3	19
7 .0	32.6

X	Y
8.7	32.3
4.8	7.3
12.5	33.1
13.4	32.9
2.0	0.75
3.6	3.6
14.6	33

	X	Y
Median	6.5	23.8
Mean	7.2	19.5
STD.DEV	4.2	13.6

Correlation = 0.88

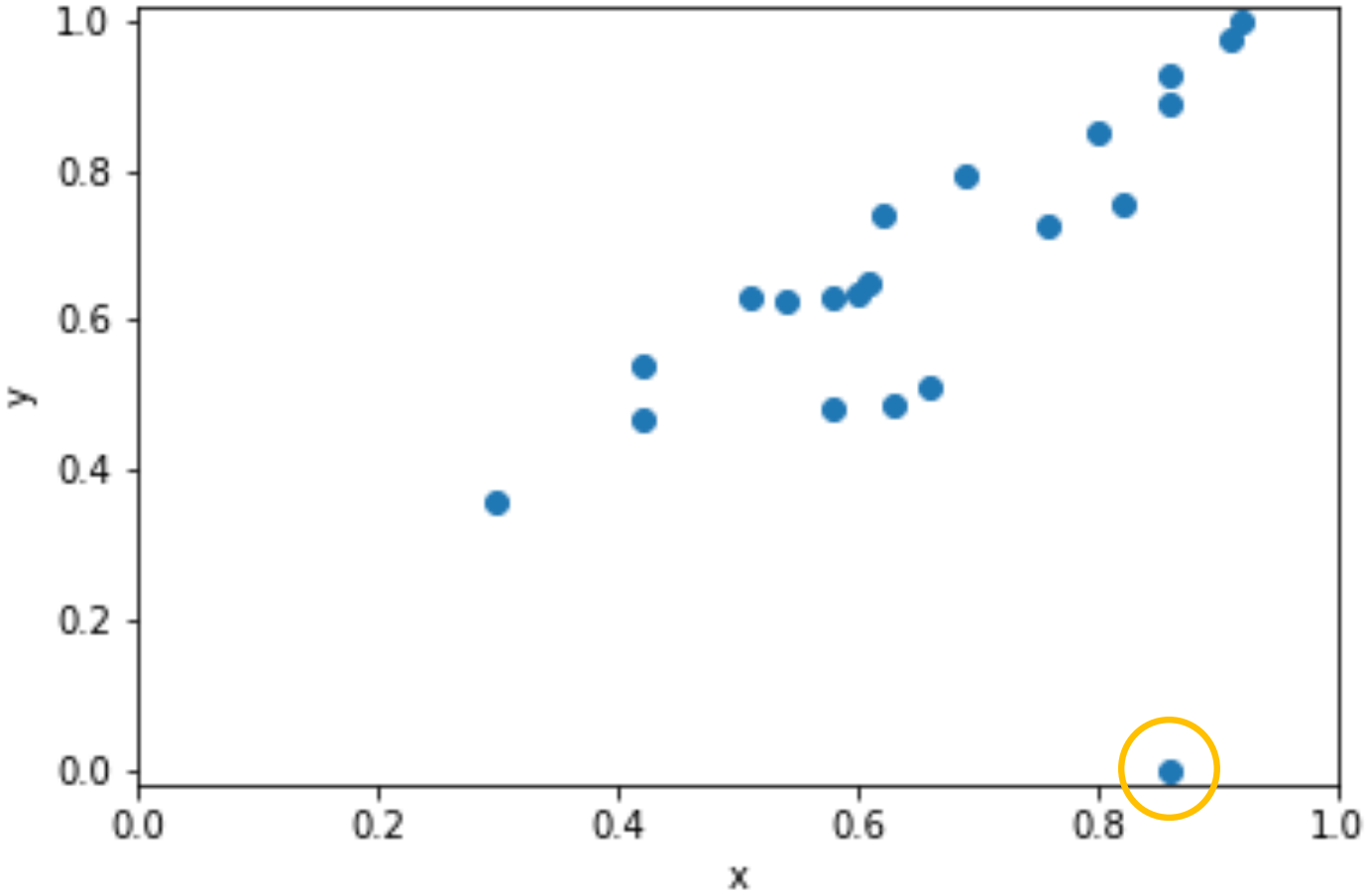
Example 1 – Visualized



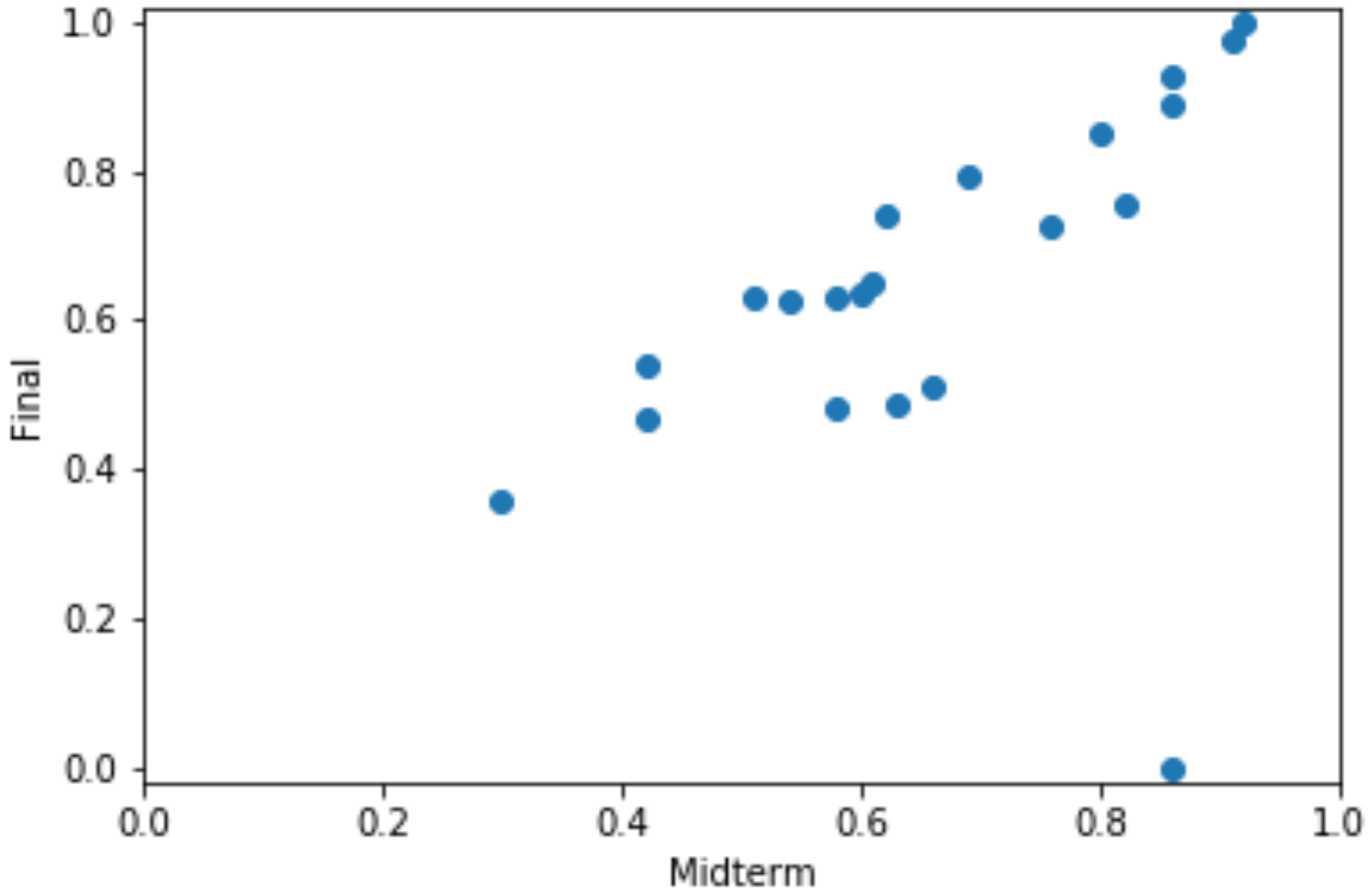
Example 2 – raw data

X	Y	X	Y
0.42	0.46750	0.66	0.51250
0.54	0.62500	0.63	0.48750
0.42	0.53750	0.92	1.00000
0.86	0.92750	0.86	0.88750
0.60	0.63750	0.91	0.97500
0.51	0.63125	0.82	0.75625
0.30	0.35625	0.86	0.00000
0.61	0.65000	0.80	0.85000
0.58	0.63125	0.69	0.79375
0.76	0.72500	0.62	0.74000
0.58	0.48125		

Example 2 – Visualized



Example 2 – Visualized



Types of Visualizations

- Understand the different ways data visualization is used in data science

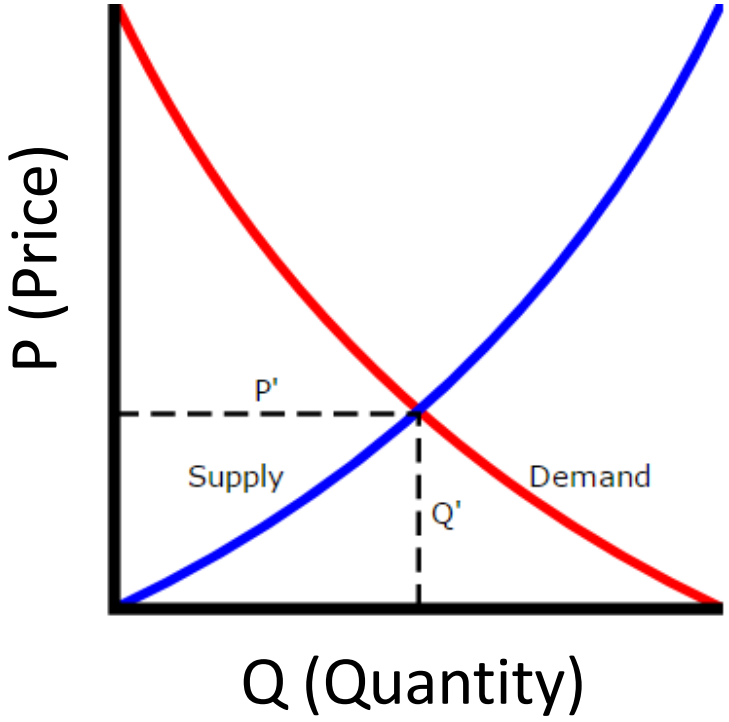
Two key categories*

Conceptual or data-driven

Declarative or exploratory

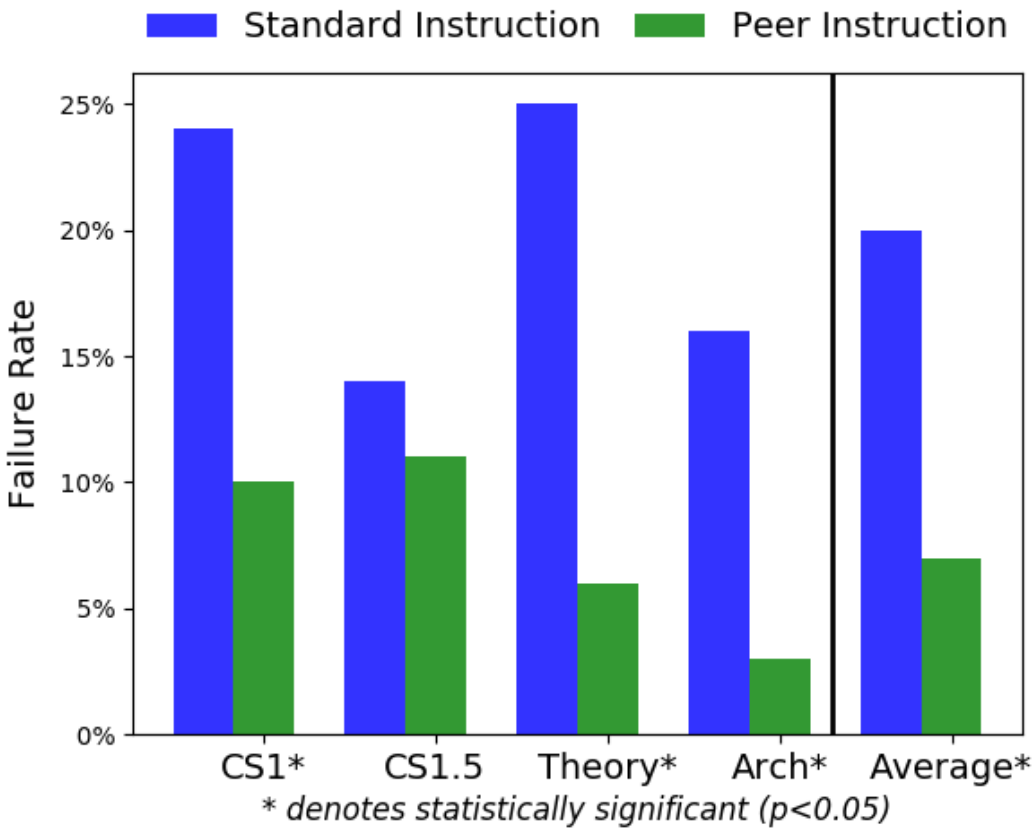
*Scott Berinato. Visualizations that really work. Harvard Business Review. June 2016.

Conceptual: Declarative



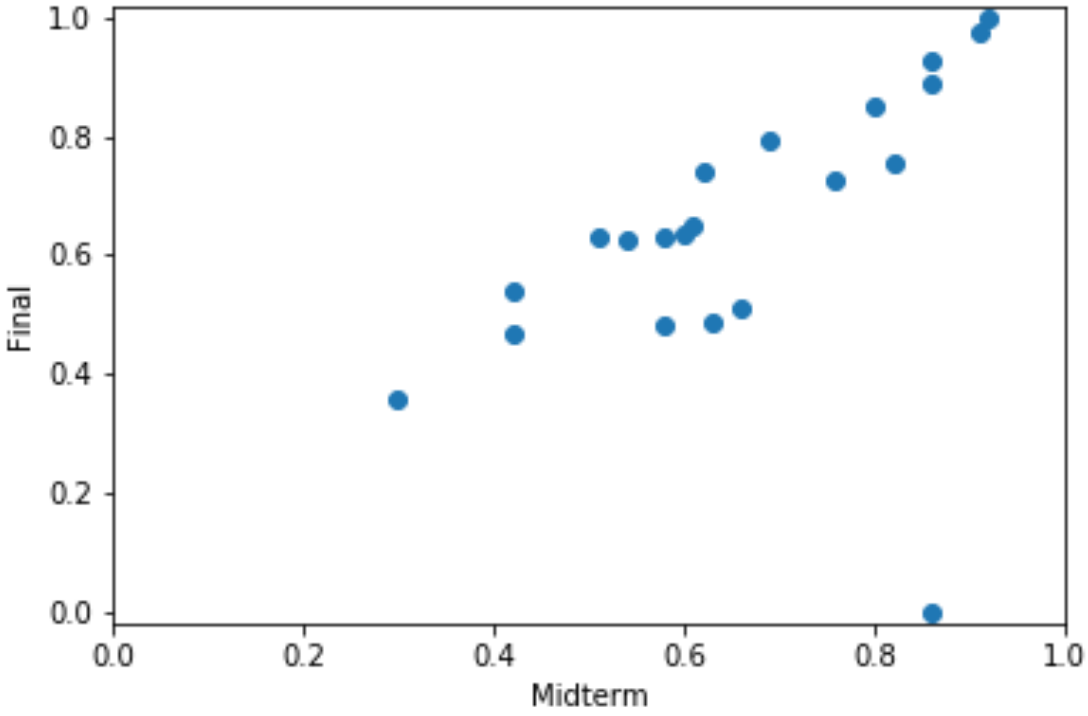
https://commons.wikimedia.org/wiki/File:Supply_and_demand_curves.svg

Data-Driven: Declarative



Porter, L., Bailey Lee, C. and Simon, B. Halving fail rates using peer instruction: a study of four computer science courses. In *Proceeding of the 44th ACM technical symposium on Computer science education*. March, 2013.

Data-Driven: Explorative



Visualization: Key design principles

- Recognize qualities of good data visualizations

Principles of Good Design

"Good data visualization is:

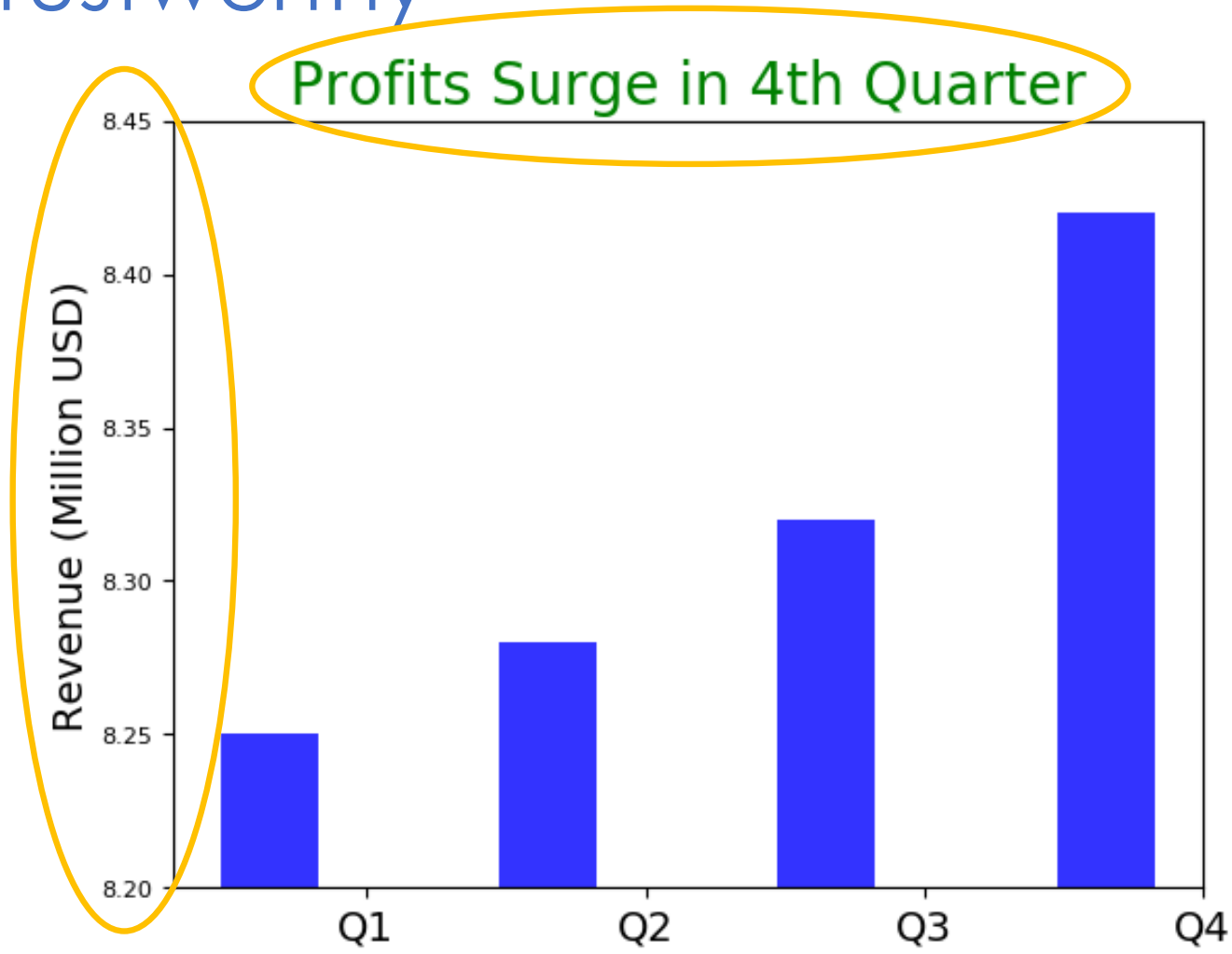
1. Trustworthy
2. Accessible
3. Elegant"

- Andy Kirk*

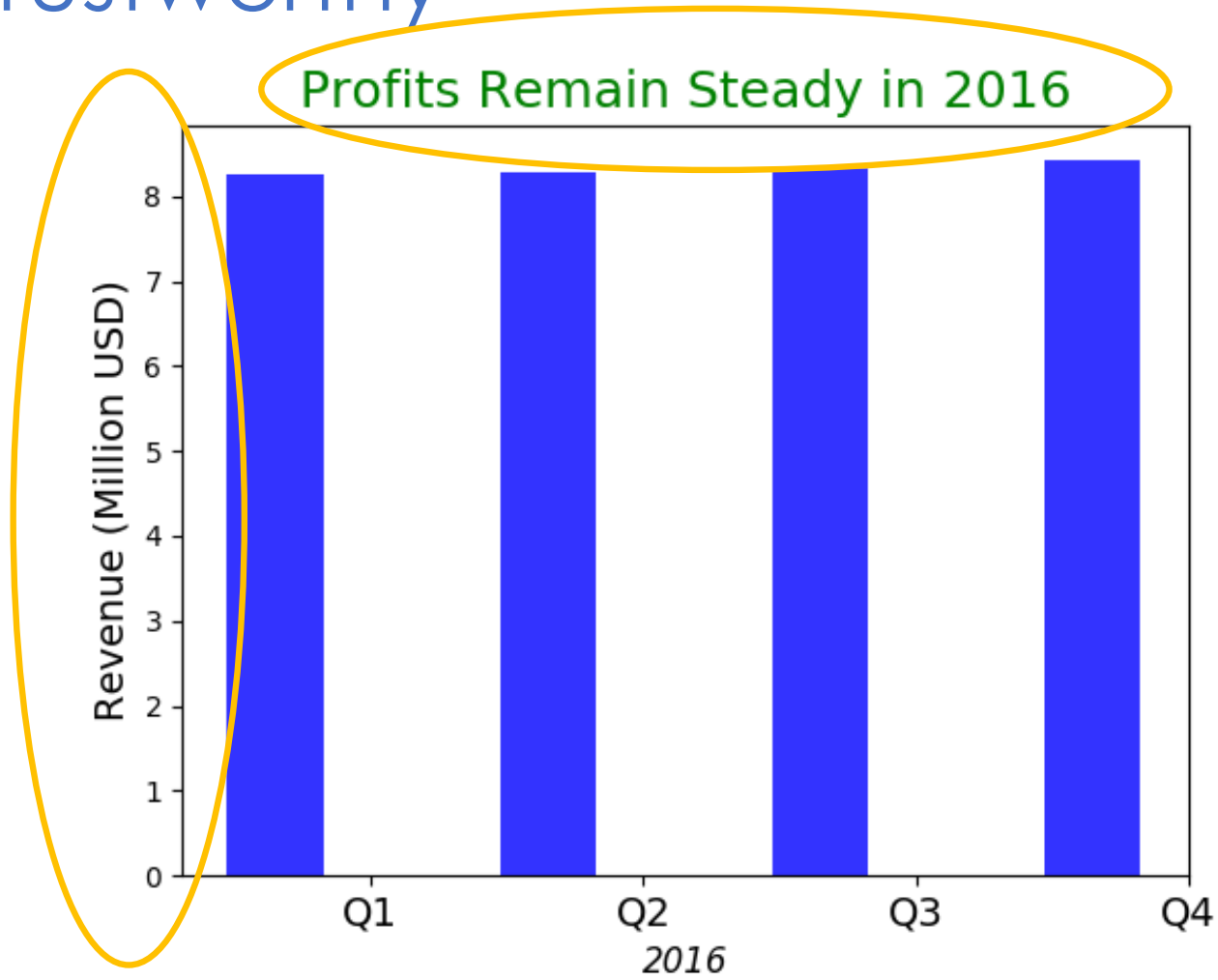
* Kirk, A. Data Visualisation: A handbook for Data Driven Design. SAGE publications, 2016.

Trustworthy

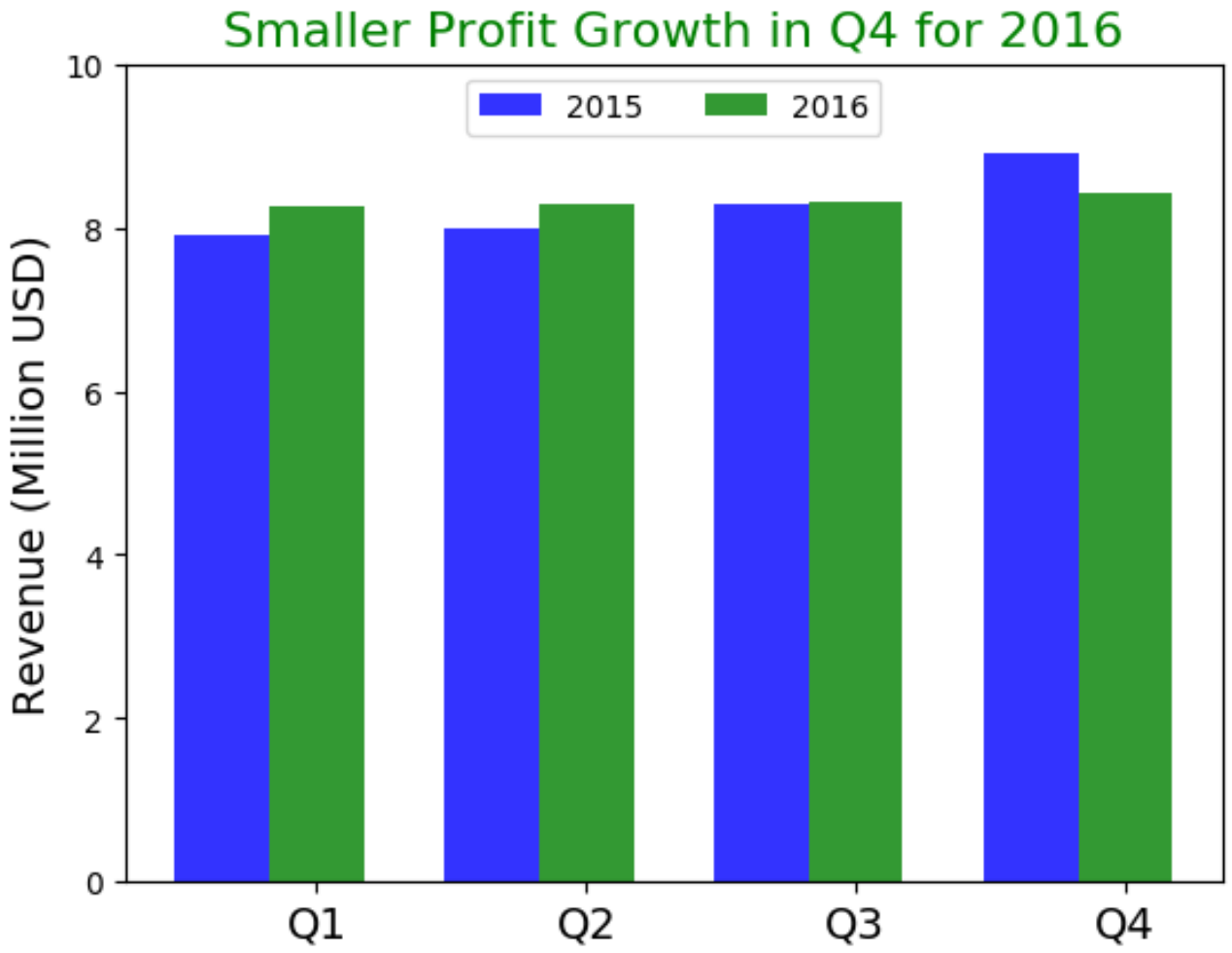
Trustworthy



Trustworthy



Trustworthy

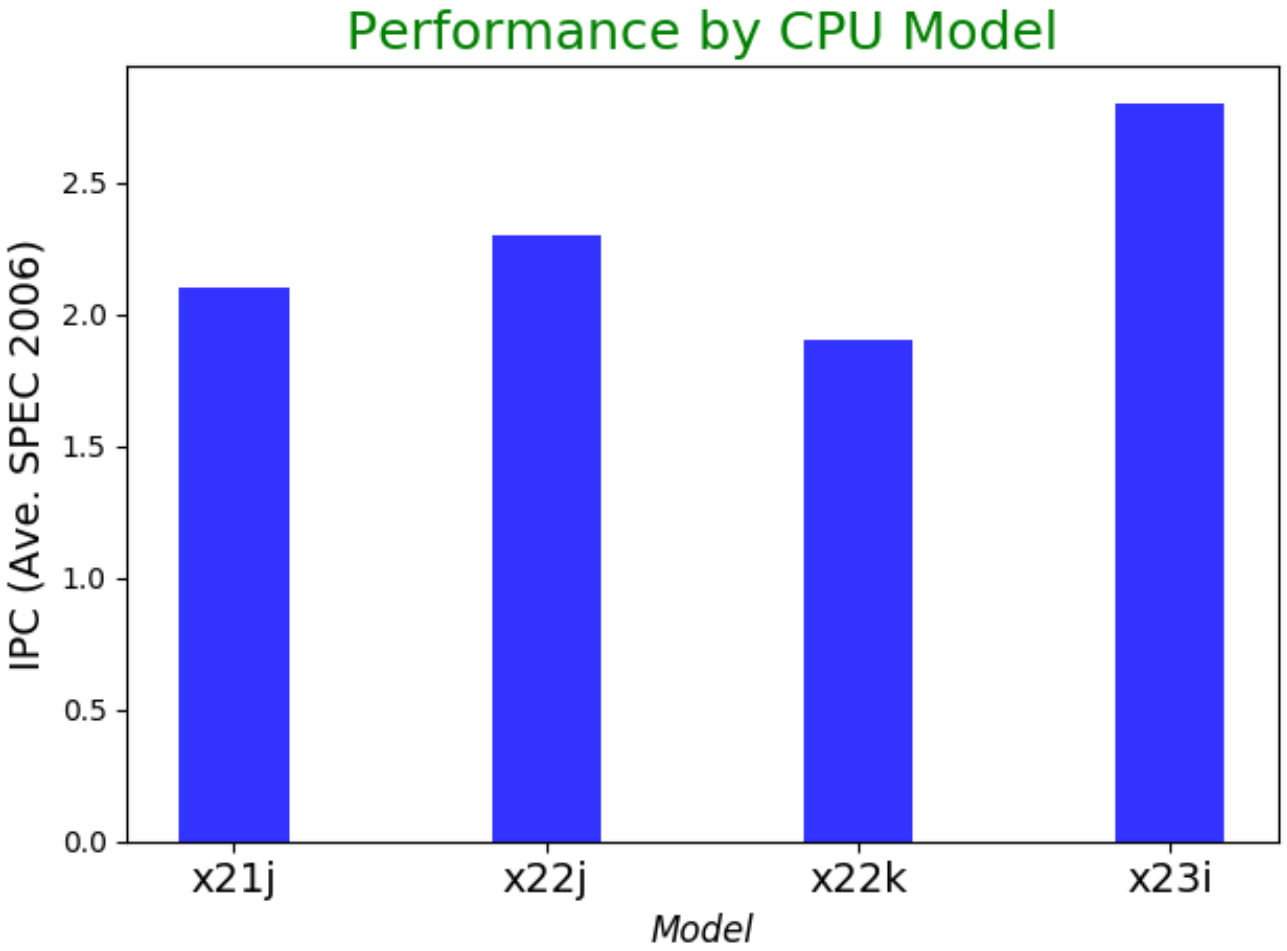


Trustworthy

- Trust is hard to earn, easy to lose.
- Honesty and integrity should be everywhere in the data science process

Accessible

Accessible

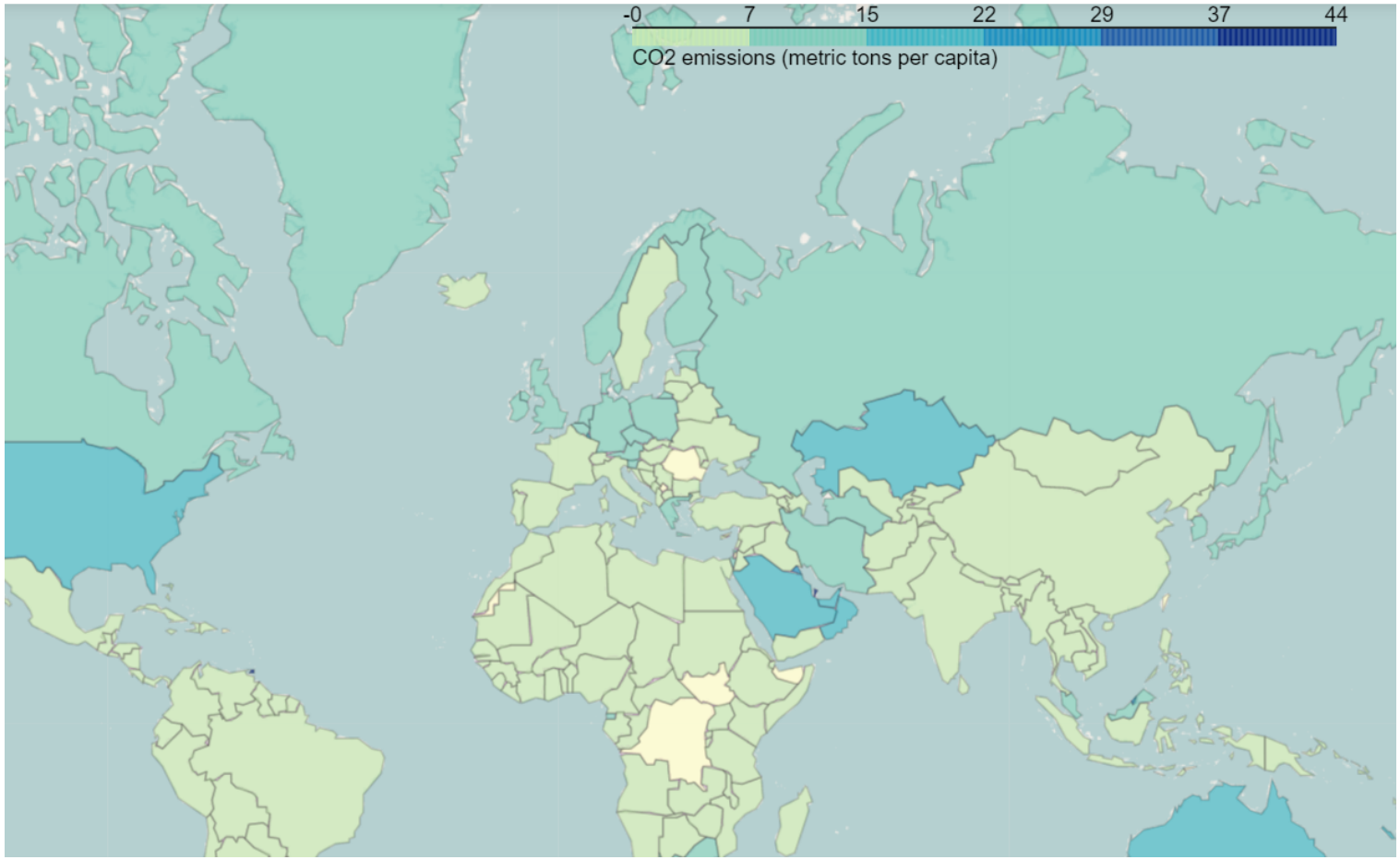

Performance by CPU Model
IPC (Ave. SPEC 2006)
2.5
2.0
1.5
1.0
0.5
0.0
x21j
x22j
x22k
x23i
Model

Accessible

- Know your audience
- Understand the purpose of the visualization

Elegant

Elegant



Elegant

- Focus on the relevant
- Be stylish if possible
- Think about decorations

Principles of Good Design

"Good data visualization is:

1. Trustworthy
2. Accessible
3. Elegant"

- Andy Kirk*

* Kirk, A. Data Visualisation: A handbook for Data Driven Design. SAGE publications, 2016.