

STA313 – Data Visualization | Lecture 1

What is visualization and why do we need it?

Prof. Fanny Chevalier

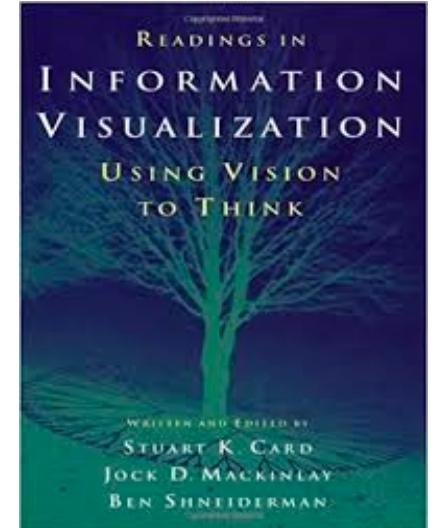
Acknowledgements:

The materials in this course is heavily based on the book and lecture materials from Tamara Munzner, as well as teaching materials and resources from John Stasko, Sheelagh Carpendale, Jeffrey Heer, Alex Lex, Miriah Meyer, Benjamin Bach, Charles Perin, Enrico Bertini, Evan Peck, Arvind Satyanarayan, Alberto Cairo, and Petra Isenberg.

Defining visualization

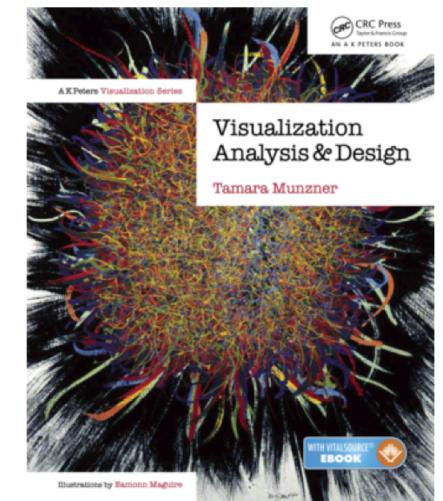
« *The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.* »

– Card et al. 1999



« *Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.* »

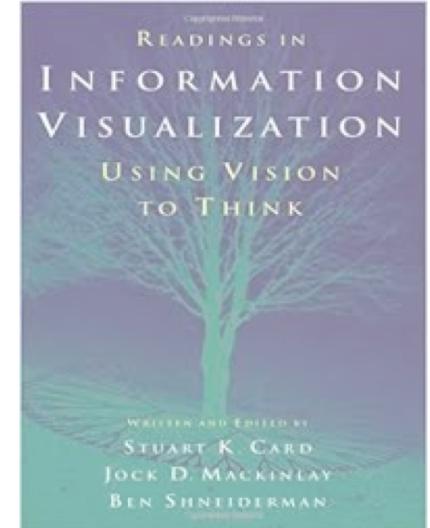
– Munzner 2014



Defining visualization

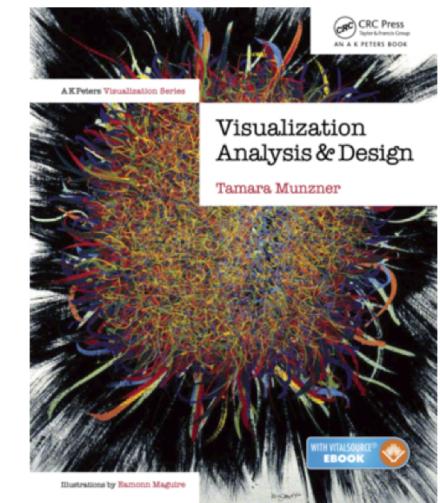
« *The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.* »

– Card et al. 1999



« *Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.* »

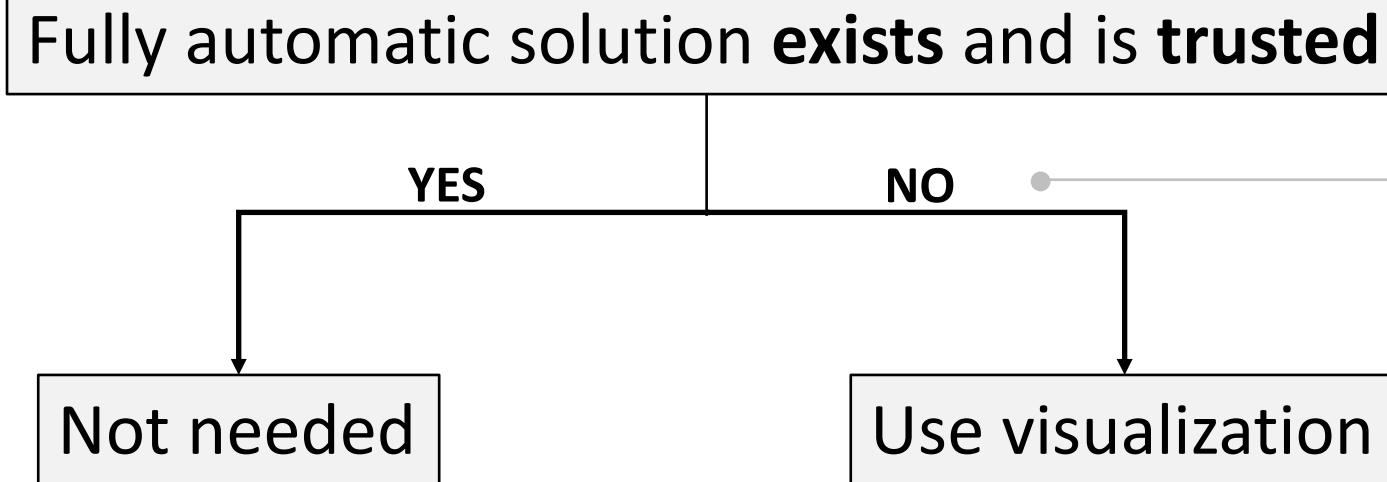
– Munzner 2014



Why have a human in the loop?

« Computer-based visualization systems provide **visual representations of datasets** designed to help **people** carry out tasks more effectively. »

– Munzner 2014

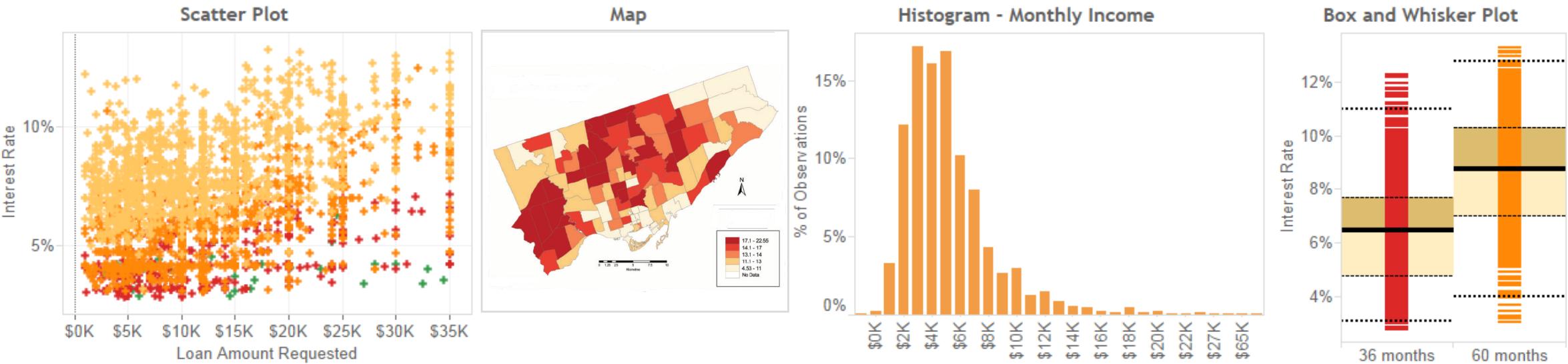


Many analysis problems are ill-specified

Why have a human in the loop?

« Computer-based visualization systems provide **visual representations of datasets** designed to help **people** carry out tasks more effectively. »

– Munzner 2014

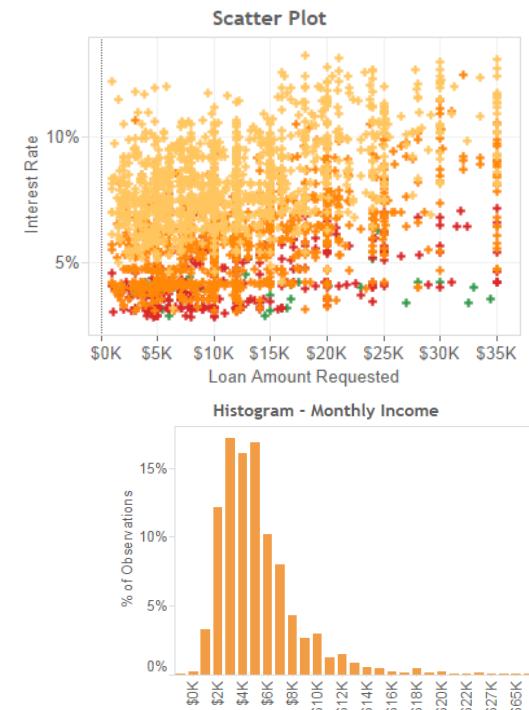


Why have a human in the loop?

« Computer-based visualization systems provide **visual representations of datasets** designed to help **people** carry out tasks more effectively. »

– Munzner 2014

Exploration / Analysis



Exploratory data analysis

assess requirements before developing a model

refine & determine parameters of automatic solution

verify, build trust in automatic solution

Why have a human in the loop?

« Computer-based visualization systems provide **visual representations of datasets** designed to help **people** carry out tasks more effectively. »

– Munzner 2014

Exploration / Analysis



Presentation



Communicate data insights
to an audience

Why rely on vision?

« Computer-based visualization systems provide **visual representations of datasets** designed to help people carry out tasks more effectively. »

– Munzner 2014

$$\mathcal{L} = \phi E, t$$

$$\nabla \cdot E = 0$$

$$= -\frac{1}{c} \frac{\partial H}{\partial t}$$

$$\nabla \times H = \vec{e}_t$$

$$V = \underline{H} \underline{\Psi}$$

$$f(\omega) = \int_{-\infty}^{\infty} f(x) e^{-j2\pi x \omega}$$

$$P\left(\frac{\partial V}{\partial t} + V\right)$$

$$\frac{1}{2} G^2 S^2 \frac{\partial^2 V}{\partial S^2} + r S^2$$

$$c(0, q_i, m_i) =$$

$$\frac{M_i}{2} \left(m_i \left(1 - \frac{D_i}{\sigma} \right) \right)$$

External representation:
replace cognition with perception

Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.1	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.1	4.0	5.39	19.0	12.5
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

requires a lot of cognitive resources to analyse in this form

Anscombe's Quartet

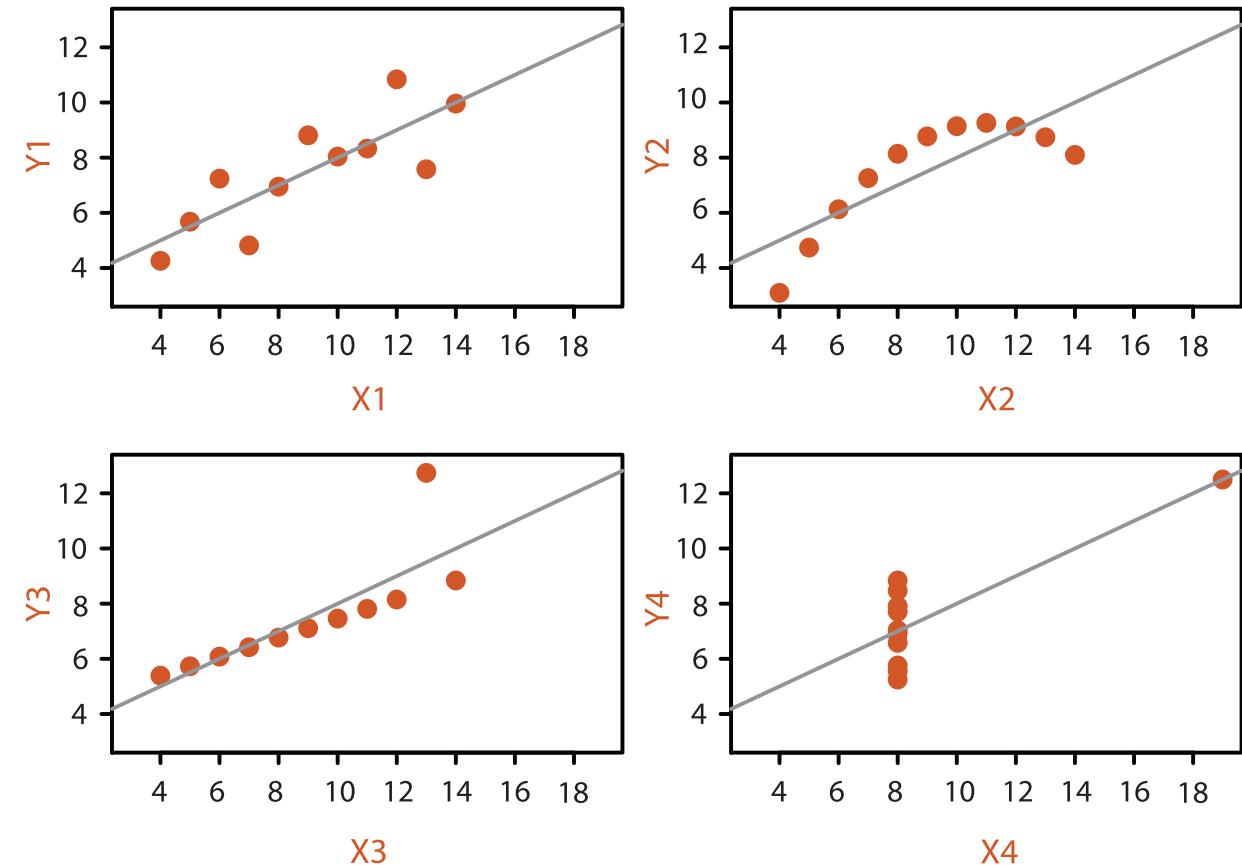
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.1	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.1	4.0	5.39	19.0	12.5
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Mean of x	9
Sample variance of x	11
Mean of y	7.50
Sample variance of y	4.12
Correlation between x and y	0.816
Linear regression line	$y = 3.00 + 0.500x$
Pearson's coefficient	r = 0.82

F.J. Anscombe, "Graphs in Statistical Analysis,"
American Statistician, vol. 27 (Feb 1973), pp. 17-21.

Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.1	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.1	4.0	5.39	19.0	12.5
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



F.J. Anscombe, "Graphs in Statistical Analysis,"
American Statistician, vol. 27 (Feb 1973), pp. 17-21.

Anscombe's Quartet

Raw data in tabular form

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.1	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.1	4.0	5.39	19.0	12.5
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

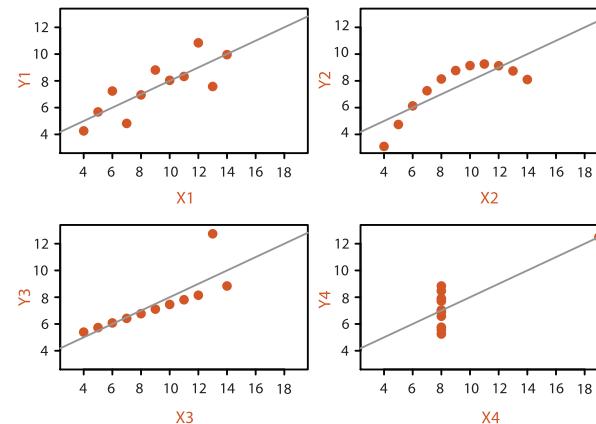
Information loss,
details matter

requires a lot of
cognitive resources to
analyse in this form

Statistical summary

Mean of x	9
Sample variance of x	11
Mean of y	7.50
Sample variance of y	4.12
Correlation between x and y	0.816
Linear regression line	$y = 3.00 + 0.500x$
Pearson's coefficient	r = 0.82

Visualization



Challenges

data

- quantity (e.g. large and streaming data)
- quality of data is often low
- dealing with uncertainty in the data

Challenges

human perception and cognition

- understanding and supporting how humans perceive and reason about data
- create representations that are fair to the data
- create (interactive) visualizations that are meaningful, clear, effective, and efficient

Lectures Summary

L1: What is visualiation and why do we need it?

L2: What? Data Abstraction & Why? Task Abstraction

L3: How? Marks and Channels

L4: More on Perception and Cognition

L5: Interaction

L6: Storytelling and Communication

L7: Spatial Structures: Maps

L8: Multidimensional Data and Matrices

L9: Hierarchical Structures: Trees, Graphs and Networks

L10: Advanced Topics in Visualization

Next up:

What?
Data Abstraction

Why?
Task Abstraction

Acknowledgements:

The materials in this course is heavily based on the book and lecture materials from Tamara Munzner, as well as teaching materials and resources from John Stasko, Sheelagh Carpendale, Jeffrey Heer, Alex Lex, Miriah Meyer, Benjamin Bach, Charles Perin, Enrico Bertini, Evan Peck, Arvind Satyanarayan, Alberto Cairo, and Petra Isenberg.