CP321 Data Visualization

- Introduction

Jiashu (Jessie) Zhao

About this Course

- 1:30-2:20pm, MWF, SC N1001
- Office Hour: Wednesday 2:20-3:20pm, N2076G
- If you are contacting my or the TA by email, please include **[CP321]** in the email subject line.
- Evaluation

Assessment	Weighting
Assignment	20%
Project	20%
Test 1	30%
Test 2	30%
Total	100%

- Project: You can work in a group of no more than 3.

- No mandatory textbooks
- Reference books:
  - Data Visualization with Python: Create an impact with meaningful data insights using interactive and engaging visuals. By Mario Dobler and Tim Großmann. (ISBN-13: 978-1789956467)
  - Interactive Visualization: Insight through Inquiry. By Bill Ferster and Ben Shneiderman. (ISBN-13: 978-0262018159)
  - Fundamentals of Data Visualization. By Claus O. Wilke. (ISBN-13: 978-1492031086)

Introduction to Data visualization

- What is data visualization?
- Why data visualization is valuable?
- History
- Models for Information Visualization

What is data visualization?

- Convert data in a systematic and logical way into the visual elements that make up the final graphic.

- Disciplinary boundaries of computer science, design, statistics, psychology, cognition, neuroscience, and the basic sciences.

Why data visualization is valuable?

- Useful in organizing and making sense out of data

- Without visualization, what can we do to understand the data?
  - Look at raw data
  - Look at statistical numbers, such as mean, variance, correlation, etc.

- Example: Anscombe ’ s Quartet

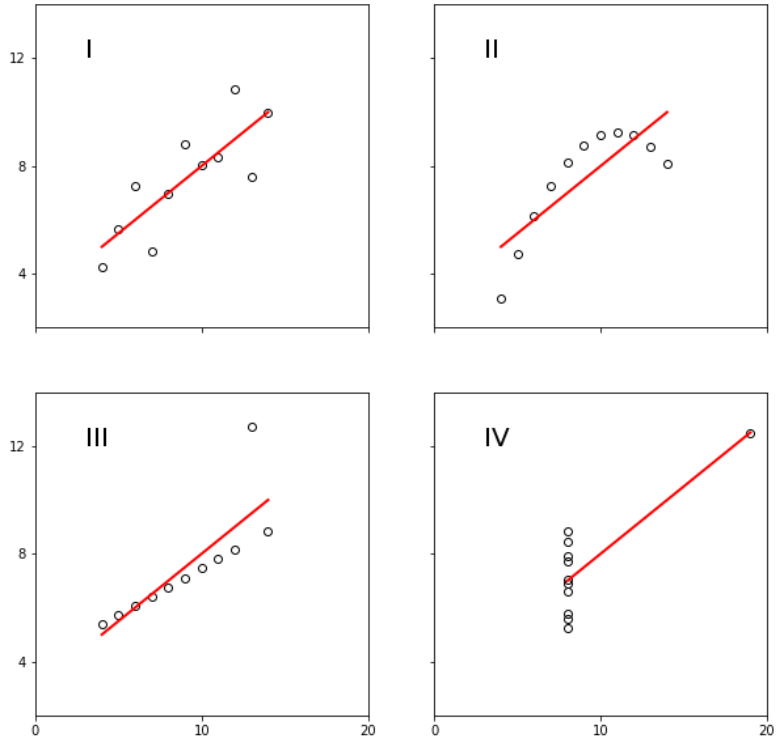
In [1]:
```
x1 = [10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5]
y1 = [8.04, 6.95, 7.58, 8.81, 8.33, 9.96, 7.24, 4.26, 10.84, 4.82, 5.68]
x2 = [10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5]
y2 = [9.14, 8.14, 8.74, 8.77, 9.26, 8.10, 6.13, 3.10, 9.13, 7.26, 4.74]
x3 = [10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5]
y3 = [7.46, 6.77, 12.74, 7.11, 7.81, 8.84, 6.08, 5.39, 8.15, 6.42, 5.73]
x4 = [8, 8, 8, 8, 8, 8, 8, 19, 8, 8, 8]
y4 = [6.58, 5.76, 7.71, 8.84, 8.47, 7.04, 5.25, 12.50, 5.56, 7.91, 6.89]
```

In [2]:
```
import numpy as np

pairs = (x1, y1), (x2, y2), (x3, y3), (x4, y4)
for ind, (x, y) in enumerate(pairs):
    print('Data', ind+1, 'Statistics: ',
          'mean_x=%1.2f, std_x=%1.2f, mean_y=%1.2f, std_y=%1.2f, corr_xy=%1.2f' %
          (np.mean(x), np.std(x),np.mean(y), np.std(y), np.corrcoef(x, y)[0][1]))
```

```
Data 1 Statistics:  mean_x=9.00, std_x=3.16, mean_y=7.50, std_y=1.94, corr_xy=0.82
Data 2 Statistics:  mean_x=9.00, std_x=3.16, mean_y=7.50, std_y=1.94, corr_xy=0.82
Data 3 Statistics:  mean_x=9.00, std_x=3.16, mean_y=7.50, std_y=1.94, corr_xy=0.82
Data 4 Statistics:  mean_x=9.00, std_x=3.16, mean_y=7.50, std_y=1.94, corr_xy=0.82
```

- What can we see by visualizing these four data sets?



In [3]:

```
import matplotlib.pyplot as plt

plt.rcParams['figure.figsize'] = [10, 10]

def fit(x):
    return 3 + 0.5 * x

fig, axs = plt.subplots(2, 2, sharex=True, sharey=True)
axs[0, 0].set(xlim=(0, 20), ylim=(2, 14))
axs[0, 0].set(xticks=(0, 10, 20), yticks=(4, 8, 12))

xfit = np.array([np.min(x1), np.max(x1)])
axs[0, 0].plot(x1, y1, 'ko', xfit, fit(xfit), 'r-', lw=2, fillstyle='none')
axs[0, 1].plot(x2, y2, 'ko', xfit, fit(xfit), 'r-', lw=2, fillstyle='none')
axs[1, 0].plot(x3, y3, 'ko', xfit, fit(xfit), 'r-', lw=2, fillstyle='none')
xfit = np.array([np.min(x4), np.max(x4)])
axs[1, 1].plot(x4, y4, 'ko', xfit, fit(xfit), 'r-', lw=2, fillstyle='none')

for ax, label in zip(axs.flat, ['I', 'II', 'III', 'IV']):
    ax.label_outer()
    ax.text(3, 12, label, fontsize=20)

plt.show()
```

Data Visualization with Python

- Jupyter Notebook
- Install

  https://jupyter.org/install
- Run
  - Command: jupyter notebook
  - Application opens in web browser: http://localhost:8888
  - Tutorial: https://www.datacamp.com/community/tutorials/tutorial-jupyter-notebook##InstallJupyter

Understanding of Data with Visualization

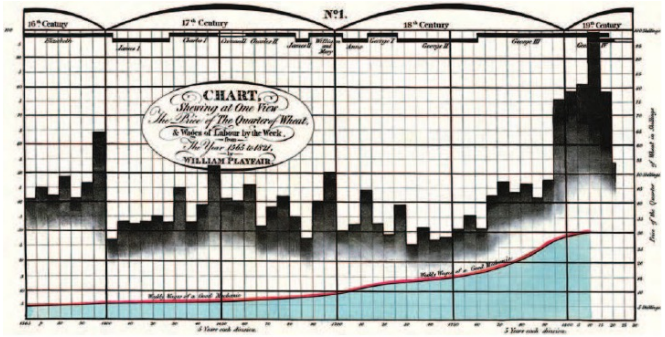
- Increase the number of resources available to the viewer while remaining cognizant of the limits of human working memory
- Reduce the need for time-consuming and tedious searching
- Enhance pattern recognition to detect meaningful trends and conditions
- Use the properties of the human perception system to effectively communicate meaning
- Allow the user to directly interact with the information to construct his or her own understandings

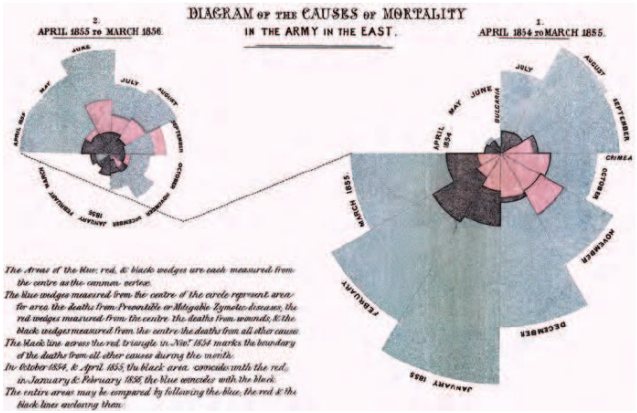
History

- People have been using graphics to represent knowledge for a long time

A Babylonian world map (6th century BC)

- Starting from 18th century, massive tables of numbers are collected.
  - William Playfair (1759 – 1823) devised a number of now-familiar graphical devices such as pie, bar, and line charts
  - Florence Nightingale expand on the pie charts to create rose diagrams (now known as coxcombs)

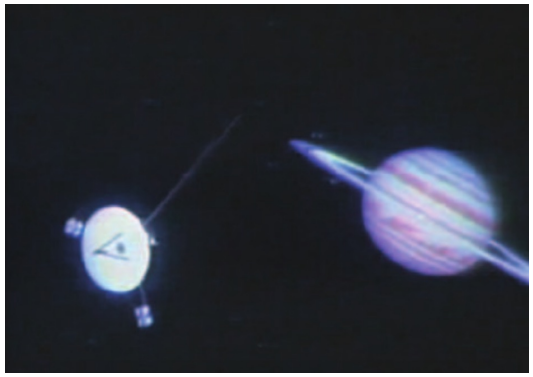




- In 1815, William Smith created the first published geological map of Britain.

A map that changed the world.

- The impact of computers
  - complex graphics based on large sets of data to create data-rich graphical visualizations
  - realistic models of complex scientific phenomena in three dimensions
  - interact with the graphics



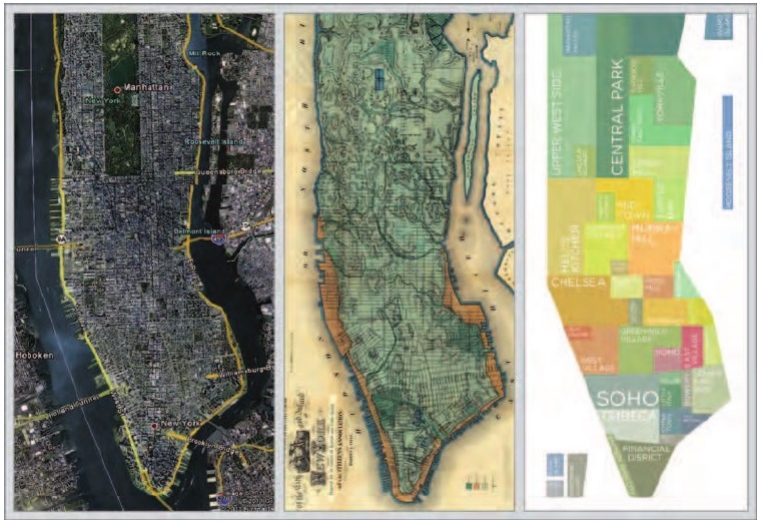
- The Impact of the Internet
  - Free accessibility have made data more widely available for instantaneous download
  - Internet companies have added simple API to their popular online services such as maps

Models for Information Visualization

- *Theoretical models* that try to offer *insight* as to the fundamental issues at stake.
- *Descriptive models* that create *taxonomies* of the elements that characterize their fundamental attributes.
- *Prescriptive models* that provide a framework to scaffold scholars through the *process* of creating and evaluating information visualizations.

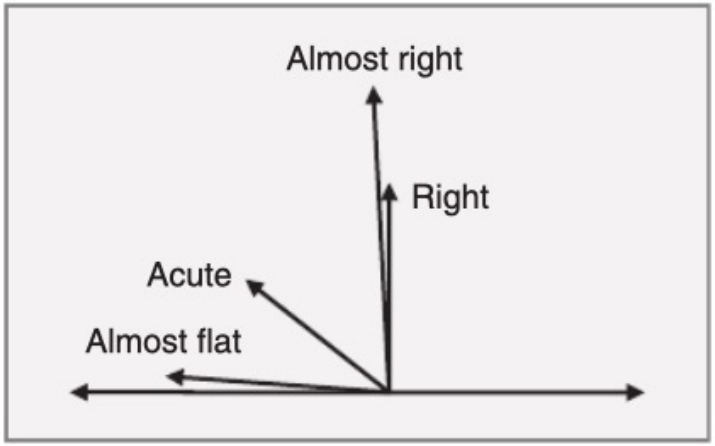
Theoretical Models

- Abstraction

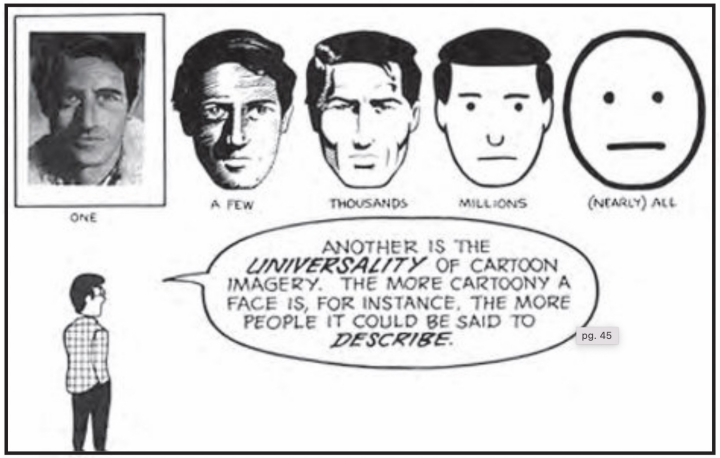


Degrees of abstraction in lower Manhattan

- Continuity
  - Although most measurements taken in the world are continuous in nature, our perceptions tend to divide them into discrete parts.
  - How we make sense of the world and are important to keep in mind when we design visualizations of continuous phenomena



Mileposts along the continuum



Part of the figures included are from the following sources:

- https://www.slideshare.net/jmborda86/the-split-attention-principle-in-multimedia-learning
- https://www.researchgate.net/figure/Modality-Effect_fig1_269112838
- https://3starlearningexperiences.wordpress.com/tag/modality-principle/
- http://ggregor.weebly.com/cognitive-load.html
- https://slideplayer.com/slide/5720365/
- https://sites.google.com/site/cognitivetheorymmlearning/contiguity-principle
- https://www.slideshare.net/Baynard/personalization-principle-by-group-a
- https://www.youtube.com/watch?v=BcWSUnXz8kw