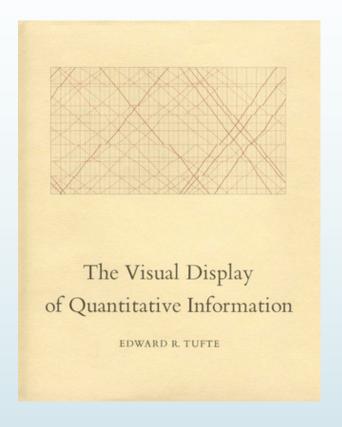
# Data Visualization

William Steimel B1778102

#### Contents

- Sources
- Introduction
- What is Data Visualization?
- Data Visualization by Charles Joseph Minard
- Tokyo Metro Case Study (Visualization of Passenger Flows)
  - Visualization of Passenger Flows Abstract
  - Tokyo Metro Data
  - Tokyo Metro Data Visualization
  - Data Visualization- Heatmap View
  - Data Visualization- RouteMap View
  - Tokyo Metro Accident Visualization
  - Impact
  - Conclusion

#### Sources



#### Visualization of Passenger Flows on Metro

Masahiko Itoh\* The University of Tokyo

Daisaku Yokoyama<sup>†</sup>
The University of Tokyo
Satoshi Kawamura
Tokyo Metro Co., Ltd.
The University of Tokyo

Masashi Toyoda<sup>‡</sup>
The University of Tokyo
Masaru Kitsuregawa<sup>§</sup>
National Institute of Informatics,
The University of Tokyo

Yoshimitsu Tomita Tokyo Metro Co., Ltd.



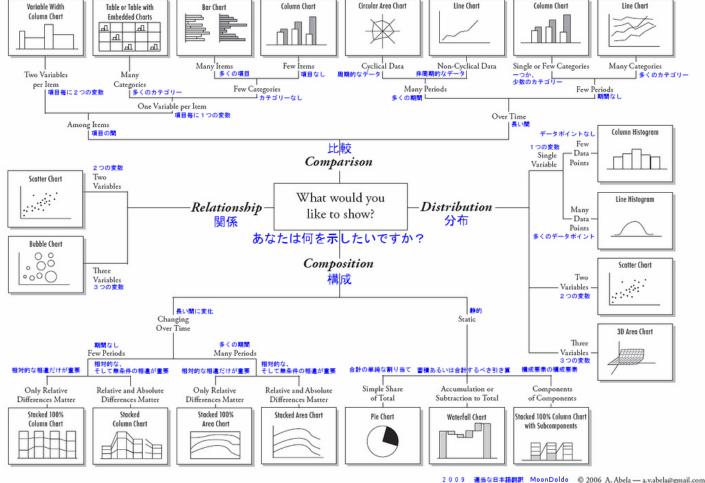
### Introduction

- Today I will discuss a few topics:
  - What is Data Visualization?
  - An example of a famous data visualization from Edward Tuftes Book that is considered one of the best in history.
  - Review of a research paper related to data visualization by Tokyo Metro titled Visualization of Passenger Flows

#### What is Data Visualization?

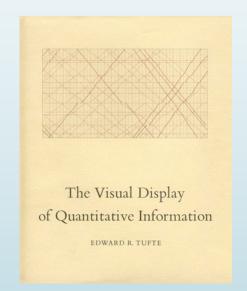
- "Data visualization is the process of presenting information through some visible means. Its goal is to make it easier to see patterns, and determine meaning, from complex data sets."
  - "There are two goals when presenting data: convey your story and establish credibility." -Edward Tufte
- There are many ways to visualize data depending on the problem.
  - The infographic on the right is useful when deciding which graphs to use.
- In Machine Learning- Data Visualization is typically for
  - Data Exploration
  - Visualization of Algorithm Behavior (Decision Boundary)
  - Visualization after Dimensionality Reduction
  - Metric Visualization

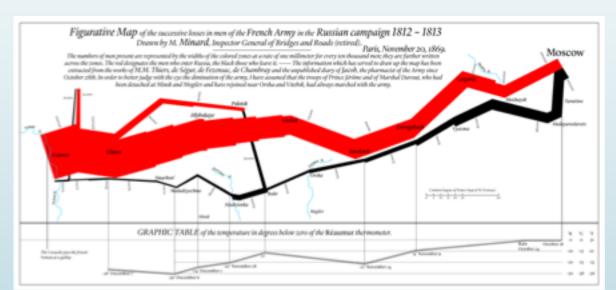
#### Chart Suggestions—A Thought-Starter



# Data Visualization by Charles Joseph Minard

- One Famous visualization from Edward Tuft's Book
- Drawn by M. Minard Paris, November 20, 1869.
  - One Visualization tells us many different things about Napolean's retreat from Russia in 1812.
  - Geographic Location, Total Soldiers, Temperature, Time/Date, Colors- Red: Men who enter Russia, Black: Men who retreat from Russia









# Tokyo Metro Case Study



- Visualization of Passenger Flows on Metro
  - Masahiko Itoh The University of Tokyo
  - Daisaku Yokoyama The University of Tokyo
  - Satoshi Kawamura Tokyo Metro Co., Ltd., The University of Tokyo
  - Masashi Toyoda The University of Tokyo
  - Masaru Kitsuregawa National Institute of Informatics, The University of Tokyo
  - Yoshimitsu Tomita Tokyo Metro Co., Ltd.

### Visualization of Passenger Flows Abstract

- **Business Case-** "Data Visualization is considered one of the most important techniques for examining influences of various kinds of phenomenon such as natural disasters, public gatherings, or accidents on changes of behavior in Metro System Passengers."
  - Understanding this behavior through data visualization is very useful to improving operational efficiency for managers of rail systems
- This paper visualizes transportation flows on Tokyo Metro based on from smart card data (PASMO) from March 2011 to April 2013

# Tokyo Metro Data

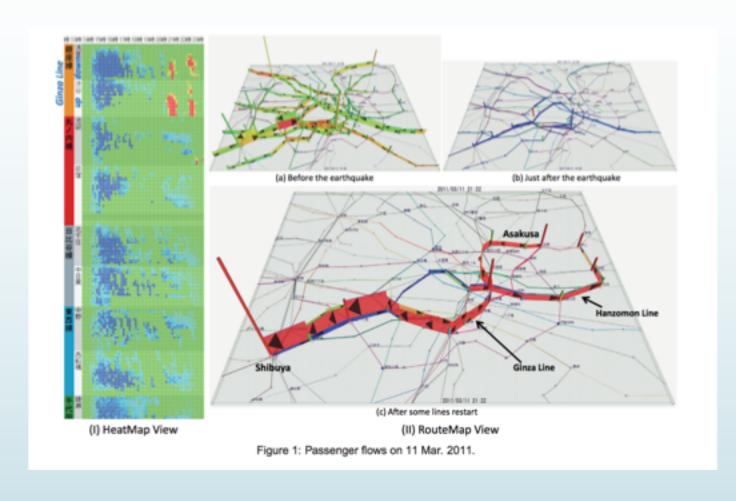
#### Methodology:

- Smart Card data (PASMO)
  - Each record contains origin, destination, and exit time (Touch in and Touch out data)
  - Goal Calculate total passengers on each part of the line every 10 minutes and calculate standard deviation to determine whether it is less or more crowded than usual
- Smart Card Challenge Does not contain data related to line transfers
  - Tokyo Metro Accounts for this through taking the most probable or likely route based on their origin and destination
- Log Data Includes almost all areas of Tokyo
  - 28 lines
  - 540 stations
  - 1 billion trips



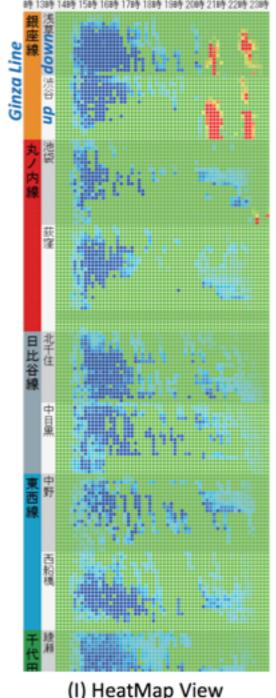
# Tokyo Metro Data Visualization

- These
   visualizations are
   two methods
   developed by
   Tokyo Metro for
   temporal analysis
   of passenger
   flows
- Details Passenger Flows on 11 Mar. 2011 (Tohoku Earthquake)



# Data Visualization Heatmap View

- X & Y Axis
  - X-axis Time (Blocks divided by every 10 minutes)
  - Y-axis Train Lines
- Color represents standard deviation in flow (d)
  - Red Higher d
  - Blue Lower d
- Up and Down Directions on lines are separated
  - Ginza Sen Up Towards Shibuya
  - Ginza Sen Down Towards Asakusa



(I) HeatMap View

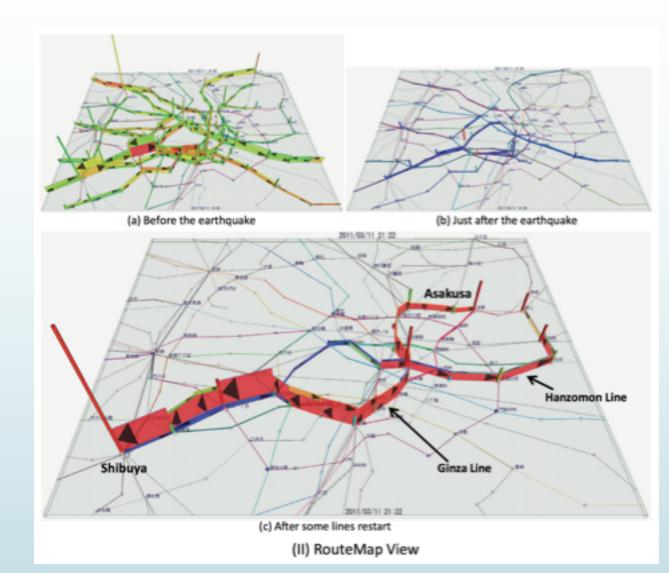
### Data Visualization – RouteMap View

#### Total Passengers Riding

- Number of passengers Height of 3D bands 2 Directions
- Color represents d (Standard Deviation of Passengers)
  - Red High, Blue Low

#### Passengers Exiting Station

- Bar on each station represents passengers exiting at each station
- Colors of bars also indicate d for (Standard Deviation of Passengers)
  - Red High, Blue Low



# Tokyo Metro Accident Visualization

- The next data visualization shows the before and after effect of an accident at JR Ueno station on multiple Tokyo Metro lines
  - Fukutoshin
  - Marunouchi
  - Nanboku

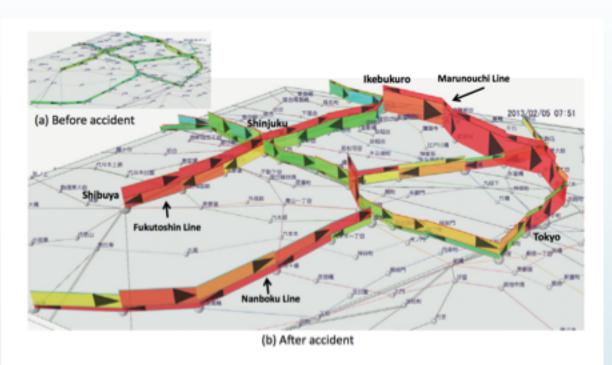


Figure 2: Effect of the accident at JR Ueno station on 5 Feb. 2013. (Only Fukutoshin, Marunouchi, and Nanboku Lines are selected.)

### Impact

- These Data Visualizations through the use of their data help Tokyo Metro understand the actions of their passengers in a variety of different conditions including:
  - Disasters
  - Accidents
  - Public Gatherings (Concerts, Shows, Etc.)
  - Everyday Use
- Improved Service Delivery
- Tokyo Metro plans to utilize this data in the future for automatic detection of future anomalous events and passenger flow simulation.
  - (Machine Learning)

#### Conclusion

- Data Visualization is extremely important for understanding trends and patterns when working in Data Science
- Tokyo Metro has gained enormous value through the use of Data Visualization to tell the story of how passengers are using their lines
  - This type of value can be gained when working with any business problem
- The visualizations created by Tokyo Metro are valuable as they illustrate multiple dimensions of information in one graph similar to M. Minard's
- Data Visualization is useful for finding patterns that are not easily seen in actual data
- Effective Data Visualization makes data easier to consume for audiences from a wide variety of backgrounds