Introduction to Data Visualization
: History, Concept, Methods

SONG SOOKYOUNG
2014.02.12.WED

HCI KOREA 2014
Data Visualization

Interest over time

Regional interest

<table>
<thead>
<tr>
<th>Region</th>
<th>Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Korea</td>
<td>100</td>
</tr>
<tr>
<td>India</td>
<td>97</td>
</tr>
<tr>
<td>United States</td>
<td>86</td>
</tr>
<tr>
<td>Canada</td>
<td>60</td>
</tr>
<tr>
<td>China</td>
<td>38</td>
</tr>
<tr>
<td>Netherlands</td>
<td>38</td>
</tr>
<tr>
<td>Australia</td>
<td>36</td>
</tr>
</tbody>
</table>
Infographic

Interest over time

Regional interest

- Singapore: 100
- South Korea: 94
- South Africa: 84
- Philippines: 73
- United States: 69
- New Zealand: 67
- Netherlands: 65
Contents

1. History

2. Concept

3. Methods (Process and Tools)
PART I.
HISTORY
# Short History of Data Visualization

<table>
<thead>
<tr>
<th>18C</th>
<th>19C</th>
<th>20C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joseph Priestley</td>
<td>John Snow</td>
<td>Jacques Bertin</td>
</tr>
<tr>
<td>William Playfair</td>
<td>Charles J. Minard</td>
<td>John Tukey</td>
</tr>
<tr>
<td>F. Nightingale</td>
<td>Edward Tufte</td>
<td>Leland Wilkinson</td>
</tr>
</tbody>
</table>
Joseph Priestley
영국의 신학자, 자연철학자, 화학자, 교육자(1733-1804)

A New Chart of History
William Playfair
스코틀랜드의 공학자, 정치경제학자 (1759-1823)

1786

The Commercial and Political Atlas
Inventing Bar Chart, Line Chart, and Pie Chart
Bar Chart
Line Chart

Exports and Imports to and from Denmark & Norway from 1700 to 1780

The Bottom line is divided into Years, the Right hand line into £10,000 each.
Pie Chart
John Snow
영국의 내과 의사 (1813-1858)

1854

Broad Street Cholera Outbreak
Dot Map
Charles J. Minard
프랑스의 도시공학자 (1781-1870)

1869

Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813

Napoleon’s Russian Campaign of 1812
Carte Figurative des progrès successifs de l'Armée Française dans la Campagne de Russie 1812-1813.

Dessiné par M. Minard, Ingenieur Général des Arts et Chanoine de Reims.

Paris, le 20 Novembre 1869.

Les nombres d'hommes pris enfin constituent par le longeur des lignes colorées à raison d'un millimètre pour vingt mille hommes ; ils sont de plus laissés en traces des gares. Le tracé tient compte des hommes qui retournent en Russie, le soir, ceux qui en sortent le matin, des convoyages qui se joignent à travers la carte, enfin qu'ils se sont faits dans les couverts à M. M. Chresto, de Soissons, de Hervieux, de Chambray et le journal militaire de Paris, pendant la campagne de l'Armée depuis le 28 Octobre.

L'un des moyens de juger de l'état de l'armée, c'est de supposer que les corps du Sénat, du Roi et du Maréchal Davoust, qui avaient été détachés au Mossk, à Minsk, et aux régions de Verkhi et Witerb, avaient toujours marché avec l'armée.

TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les erreurs de France (ou de M. de France)
Florence Nightingale
영국의 통계학자, 간호사 (1820-1910)

Mortality of the Civil Population in England

1873

Representing the Relative Mortality of the Army at Home and of the English Male Population at corresponding Ages.
Rose Diagram

The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent areas for the deaths from Preventable or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.
The black line across the red triangle in Nov 1854 marks the boundary of the deaths from all other causes during the month.
In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.
The entire areas may be compared by following the blue, the red & the black lines enclosing them.
Jacques Bertin
프랑스의 지도제작자 (1918-2010)

1967

Sémiologie Graphique.
Les diagrammes, les réseaux, les cartes.

Semiology of Graphics
: Diagrams, Networks, Maps
(1983)
Diagrams

Planar Dimensions
- Size
- Value
- Texture
- Color
- Orientation
- Shape
Networks

Composition
Linear
Circular
Grid
Organized Grid
Cubic
Maps
John W. Tukey
미국의 통계학자 (1915-2000)

1977

Exploratory Data Analysis
EDA Exploratory Data Analysis

Graphical Techniques
- Box Plot
- Histogram
- Run Chart
- Pareto Chart
- Scatter Plot
- Stem-and-Leaf Plot
- Parallel Coordinates
- Odds ratio
- Multidimensional Scaling
- Principal Component Analysis
- Multilinear PCA
Box Plot
Edward R. Tufte
미국의 통계학자 (1942~)

The Visual Display of Quantitative Information

1983

Visual Explanations: Images and Quantities, Evidence and Narrative

1997
Small Multiple
## Sparkline

<table>
<thead>
<tr>
<th></th>
<th>1999.1.1</th>
<th>65 months</th>
<th>2004.4.28</th>
<th>low</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>1.1608</td>
<td></td>
<td>1.1907</td>
<td>.8252</td>
<td>1.2858</td>
</tr>
<tr>
<td>¥</td>
<td>121.32</td>
<td></td>
<td>130.17</td>
<td>89.30</td>
<td>140.31</td>
</tr>
<tr>
<td>£</td>
<td>0.7111</td>
<td></td>
<td>0.6665</td>
<td>.5711</td>
<td>0.7235</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2003.4.28</th>
<th>12 months</th>
<th>2004.4.28</th>
<th>low</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>1.1025</td>
<td></td>
<td>1.1907</td>
<td>1.0783</td>
<td>1.2858</td>
</tr>
<tr>
<td>¥</td>
<td>132.54</td>
<td></td>
<td>130.17</td>
<td>124.80</td>
<td>140.31</td>
</tr>
<tr>
<td>£</td>
<td>0.6914</td>
<td></td>
<td>0.6665</td>
<td>0.6556</td>
<td>0.7235</td>
</tr>
</tbody>
</table>
Leland Wilkinson
미국의 통계학자, 컴퓨터 과학자

The Grammar of Graphics
Graphical Pipeline

Graphical Pipeline
Variables
Algebra
Scales
Statistics
Geometry
Coordinates
Aesthetics
Renderer

(a) Variables
   Variable set

(b) Algebra
   Variable set

(c) Scales
   Variable set

(d) Statistics
   Statistics graph

(e) Geometry
   Geometric graph

(f) Coordinates
   Coordinate graph

(g) Aesthetics
   Graphic

Renderer

Display

Manipulate data and combine into a set of variables

Apply scales for each variable

Calculate statistics

Create geometric object for mark

Apply coordinate system transformation

Map dimensions of graph to position and styles of marks
ggplot2 in R

Wickham(2009)

```r
p = qplot(Concentration, Percent.of.control, 
  data=screening_data, 
  geom=c("point", "smooth"), colour=Response.type) + 
  scale_x_log10() + 
  facet_grid(Compound ~ Cell.line) + 
  coord_cartesian(ylim=c(-10, 110)) 
print(p)
```
PART II.
CONCEPT
4 Concepts

- Visual Analytics
- Data Visualization
- Information Design
- Infographic
Data Visualization

the use of computer-supported, interactive, visual representations of abstract data to amplify cognition
Scientific Visualization

Scientific Visualization (SciVis) applies visualization to scientific data, typically physical data (the human body, the earth, molecules, or other) to enable scientists to perceive certain phenomena in the data
Information Visualization

Information Visualization (InfoVis) focuses on visualizing non-physical, abstract data such as financial data, business information, document collections, and abstract conceptions.
GeoVisualization

Geo-spatial data is special since it describes objects or phenomena that are related to a specific location in the real world.

Wind Map
Visual Analytics

the science of analytical reasoning facilitated by interactive visual interfaces
(Thomas and Cook, 2005)
the practice of presenting information in a way that fosters efficient and effective understanding of it
Travel Times on Commuter Rail

The map shows the travel times, in minutes, from Manhattan to stations in the region’s commuter rail system during the evening rush. Each alternating ring shows how much farther you can travel in an additional 15 minutes. Inbound times may differ.

**Express Train Effect**

Lines that appear to double back on themselves indicate when stations farther from Manhattan have shorter travel times because they are served by express trains.

**SPEED OF THE TRAINS**

Color-coding indicates how fast, on average, the train travels to reach that station:

- 15-25 m.p.h.
- 26-35 m.p.h.
- 36-45 m.p.h.
- 45+ m.p.h.

Travel times are based on trains departing Penn Station (New Jersey Transit, L.I.R.R. or Grand Central Terminal (Metro-North)) between 4 and 7 p.m. on weekdays. An average of the three fastest trains was used to capture the effect of express trains. Local trains may take significantly longer.

Sources: L.I.R.R., N.J. Transit, Metro-North; National Transit Database

<table>
<thead>
<tr>
<th></th>
<th>N.J. Transit</th>
<th>L.I.R.R.</th>
<th>Metro-North</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Annual Riders</strong></td>
<td>74 million</td>
<td>82.1 million</td>
<td>76.9 million</td>
</tr>
<tr>
<td><strong>Average Weekly Trips</strong></td>
<td>260,000</td>
<td>262,000</td>
<td>206,000</td>
</tr>
<tr>
<td><strong>Average Fare Revenue</strong></td>
<td>$361 million</td>
<td>$487 million</td>
<td>$465 million</td>
</tr>
<tr>
<td><strong>Average Fare Per Mile</strong></td>
<td>16 cents</td>
<td>20.7 cents</td>
<td>20.3 cents</td>
</tr>
<tr>
<td><strong>Farebox Ratio</strong></td>
<td>43.2 percent</td>
<td>46.4 percent</td>
<td>61.4 percent</td>
</tr>
<tr>
<td><strong>Stations</strong></td>
<td>162</td>
<td>124</td>
<td>109</td>
</tr>
<tr>
<td><strong>Average Fleet Age</strong></td>
<td>13 years</td>
<td>9 years</td>
<td>17 years</td>
</tr>
<tr>
<td><strong>On-Time Performance</strong></td>
<td>95%</td>
<td>93%</td>
<td>98%</td>
</tr>
</tbody>
</table>

*Percent of riders that cover operating costs. †Within six minutes.
Infographic

the graphic visual representations of information, data or knowledge intended to present complex information quickly and clearly
Data visualization is a popular new way of sharing research. Here is a look at some of the visual devices, informational elements, and general trends found in the modern day infographic.

**Design**

**Chart Style**
Percentage of infographics with the following charts:

- Pie Chart: 22%
- Pictorial Chart: 24%
- Line Chart: 24%
- Bar Chart: 32%

**Font**
- Sans Serif: 85%
- Condensed Sans Serif: 15%

**Key Info**
Percentage of infographics with key:
- 0: 33%

Average number of symbols per key: 5.1

**Base Color**
- Orange: 13%
- Green: 18%
- Blue: 29%
- Red: 18%
- Yellow: 10%
- Black: 25%

**Navigational Iconography**
Frequency of arrows & connecting lines in infographics:
- Arrows: 13%
- Lines: 38%
- Both: 13%

**Content**

**Countries Featured**

- United States: 88%
- China: 22%
- United Kingdom: 12%
- Australia: 12%
- Canada: 10%
- India: 10%
- France: 10%
- Mexico: 8%

**Theme**
Relative popularity of different infographic themes:
- Technology, Media, Culture: 15%
- Education: 15%
- Health: 12%
- Politics: 10%
- Transportation: 10%
- Environment: 8%

**Sections**
Average number of sources per infographic: 2.29

**Credited Sources**
Average number of words per infographic: 4,36

**Title**
“Richest and Poorest American Neigh”

---

**Concept & Design** Ivan Cash  
**Source** 49 infographics collected at random from www.good.is/infographics
PART III.
PROCESS
Data Visualization Process

Information Visualization Reference Model by S.K. Card
### Data types

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-dimensional</td>
<td>Linear data types are texts, program source code, or lists which are organized in a sequential manner.</td>
</tr>
<tr>
<td>2-dimensional</td>
<td>Mostly geographical data, such as maps, floorplans, or abstract 2D text layouts.</td>
</tr>
<tr>
<td>3-dimensional</td>
<td>Real-world objects like buildings, the human body, or molecules.</td>
</tr>
<tr>
<td>Temporal</td>
<td>Time-related items with a start and finish time, possibly overlapping.</td>
</tr>
<tr>
<td>Multi-dimensional</td>
<td>Data sets with many dimensions.</td>
</tr>
<tr>
<td>Tree (hierarchies)</td>
<td>Hierarchies or tree structures are defined by each item having a link to one parent item (except for the root).</td>
</tr>
<tr>
<td>Network</td>
<td>Items can be linked to an arbitrary number of other items.</td>
</tr>
</tbody>
</table>

Data types by Shneiderman (1996)
Univariate data

*Data Visualization* by Jarvinen (2013)
Bivariate data

*Data Visualization* by Jarvinen (2013)

scatterplot

bubbleplot
Multivariate data

*Data Visualization* by Jarvinen (2013)

- Parallel coordinates
- Scatterplot matrix
- Correlation matrix
- Correlation networks
- PCA (Principal Component Analysis)
Visual Mapping

Ranking of perceptual tasks by Cleveland and McGill (1984)
View Transformations

Classification of Visual Data Exploration Techniques

1. one-dimensional
2. two-dimensional
3. multi-dimensional
4. text/web
5. hierarchies/graphs
6. algorithm/software

Data to be Visualized

Visualization Technique

Interaction Technique

Standard 2D/3D Display
Geometrically-transformed Display
Iconic Display
Dense Pixel Display
Stacked Display

Standard Projection Filtering Zoom Distortion Link&Brush

Classification of Visual Data Exploration Techniques by Keim
Mistakes in Visualization
Mistakes in Visualization

The NSW Health system is...

RECRUITING MORE NURSES*

* Nursing headcount figures at June includes non casual staff and 3rd schedule

NSW Ministry of Health March 2013
Mistakes in Visualization

76% of the extreme poor live in rural areas
Mistakes in Visualization
Mistakes in Visualization

16 Userless Infographics
PART IV.
TOOLS
Many Eyes

An experiment brought to you by IBM Research and the IBM Cognos software group

Many Eyes (IBM)
Tableau Public
d3.js

D3.js is a JavaScript library for manipulating documents based on data. D3 helps you bring data to life using HTML, SVG and CSS. D3’s emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.

Download the latest version (3.4.1) here:
- d3.v3.zip

Or, to link directly to the latest release, copy this snippet:
R Studio
“In brief, the grammar tells us that a statistical graphic is a mapping from data to aesthetic attributes (color, shape, size) of geometric objects (points, lines, bars). The plot may also contain statistical transformations of the data and is drawn on a specific coordinate system”
download R and R Studio
installing packages

```
> install.packages("ggplot2")
Installing package into '/Users/song/Library/R/3.0/library'

trying URL 'http://cran.rstudio.com/bin/macosx/contrib/3.0/ggplot2_0.9.3.1.tgz'
Content type 'application/x-gzip' length 2650041 bytes (2.5 Mb)
downloaded 2.5 Mb

The downloaded binary packages are in
 /var/folders/t5/s8ml1mxlv6sv09nbtkmnxr0000gn/T/\RtmpEFw508/downloaded_packages
```
> install.packages("ggplot2")
Installing package into './Users/song/Library/R/3.0/library'
(as 'lib' is unspecified)
trying URL 'http://cran.rstudio.com/bin/macosx/contrib/3.0/ggplot2_0.9.3.1.tgz'
Content type 'application/x-gzip' length 2650041 bytes (2.5 Mb)
opened URL

downloaded 2.5 Mb

The downloaded binary packages are in
/var/folders/5s/6ml1m1v6x9nbttrnxr000gn/T/RtmpEFw508/downloaded_pack

> library(ggplot2)
>
install.packages("ggplot2")

The downloaded binary packages are in
/var/folders/t5/8ml1mxl1v6x6s09hbttkmzfl8000gq

library(ggplot2)

str(mpg)

$ manufacturer: Factor w/ 15 levels "audi","chevrolet",...: 1 1 1 1 1 1 1 1 1 1 ...
$ model: Factor w/ 38 levels "4runner 4wd",...: 2 2 2 2 2 2 3 3 3 3 ...
$ displ: num 1 1 1 1 1 1 1 1 1 1 ...
$ cyl: int 4 4 4 4 6 6 4 4 4 4 ...
$ trans: Factor w/ 10 levels "auto(aw)","auto(l3)",...: 4 9 10 1 4 9 1 9 4 1 ...
$ drv: Factor w/ 3 levels "4","F","r": 2 2 2 2 2 2 1 1 1 1 ...
$ cty: int 18 18 18 18 18 18 18 18 18 20 ...
$ hwy: int 29 29 31 30 26 26 27 26 25 28 ...
$ fl: Factor w/ 5 levels "c","d","e","p",...: 4 4 4 4 4 4 4 4 4 4 ...
$ class: Factor w/ 7 levels "2seater","compact",...: 2 2 2 2 2 2 2 2 2 2 ...

Factor label information for annotation
drawing ggplot

```r
> install.packages("ggplot2")
Installing package into '/Users/song/Library/R/3.0/library'
(as 'lib' is unspecified)
trying URL 'http://cran.rstudio.com/bin/macosx/contrib/3.0/ggplot2_0.9.3.1.tar.gz'
Content type 'application/x-gzip' length 2650041 bytes (2.5 Mb)
opened URL

downloaded 2.5 Mb

The downloaded binary package is in:
/var/folders/tn/kxmx000gnd89kgnhnt3h7d480000gn

> library(ggplot2)
> str(mpg)
'data.frame': 234 obs. of 11 variables:
$ manufacturer: Factor w/ 15 levels "audi","chevrolet",...: 1 1 1 1 1 1 1 1 1 1...
$ model       : Factor w/ 38 levels "runner 4wd",...: 1 2 3 4 5 6 7 8 9 10...
$ displ       : num 1.8 2.22 2.8 2.8 3.1 3.8 1.82 ...
$ year        : int 1999 1999 2008 1999 1999 2006 ...
$ cty         : int 15 15 21 20 16 18 18 16 20 ...
$ hwy         : int 19 19 25 26 27 26 25 28 28 ...
$ fl          : Factor w/ 5 levels "c","d","e","p": 4 4 4 4 4 4 4 4 4 4...
$ class       : Factor w/ 7 levels "2seater","compact",...: 1 2 2 2 2 2 2 2 2 2...

> qplot(displ, hwy, data=mpg)
```
install.packages("ggplot2")
Installing package into '/Users/song/Library/R/3.0/library'
(as 'lib' is unspecified)
trying URL 'http://cran.rstudio.com/bin/macosx/contrib/3.0/ggplot2_0.9.3.1.tgz'
Content type 'application/x-gzip' length 2655041 bytes (2.5 Mb)
opened URL
downloaded 2.5 Mb

The downloaded binary packages are in
/var/folders/t5/s8mllmvlv6xs09nbttkmnxr0000gn/T//RtmpEFw508/downloaded_packages
> library(ggplot2)
> str(mpg)
'data.frame': 234 obs. of 11 variables:
$ manufacturer: Factor w/ 15 levels "audi","chevrolet",..: 1 1 1 1 1 1 1 1 1 ....
$ model : Factor w/ 38 levels "4runner 4wd",..: 2 2 2 2 2 2 2 3 3 3 ....
$ displ : num 1.8 1.8 2 2 2 2.8 2.8 3.1 1.8 1.8 2 ....
$ cyl : int 4 4 4 4 6 6 6 4 4 4 ....
$ trans : Factor w/ 10 levels "auto(av)","auto(3l)",..: 4 9 10 1 4 9 1 9 4 1 ....
$ drv : Factor w/ 3 levels "4","f","r": 2 2 2 2 2 2 2 1 1 1 ....
$ cty : int 18 21 20 21 16 18 18 16 20 ....
$ hwy : int 29 29 31 30 26 26 26 25 25 28 ....
$ fl : Factor w/ 5 levels "e","d","p",..: 4 4 4 4 4 4 4 4 4 4 ....
$ class : Factor w/ 7 levels "2seater","compact",..: 2 2 2 2 2 2 2 2 2 2 ....
> ggplot(mpg, aes(displ, hwy))
modifying aesthetics

```r
> install.packages("ggplot2")
Installing package into '/Users/song/Library/R/3.0/library'
(as 'lib' is unspecified)
trying URL 'http://cran.rstudio.com/bin/macosx/contrib/3.0/ggplot2_0.9.3.1.tgz'
Content type 'application/x-gzip' length 2650041 bytes (2.5 Mb)
opened URL

The downloaded binary packages are in
/var/folders/5s/8nllmx1v6x09nbtknxv00wmc/TmpDIR92007/1/
> library(ggplot2)
> str(mpg)
'data.frame': 234 obs. of 11 variables:
$ manufacturer: Factor w/ 15 levels "audi","chevrolet",..: 1 1 1 1 1 1 1 1 1 1 ...
$ model: Factor w/ 38 levels "4runner 4wd",..: 2 2 2 2 2 2 3 3 3 ...
$ displ: num 1.8 1.8 2 2 2 2.8 2.8 3.1 3.1 3.1 ...
$ cyl: int 4 4 4 4 4 4 4 4 4 4 ...
$ trans: Factor w/ 10 levels "auto(m6)","auto(c4)",..: 4 9 10 1 4 9 1 9 4 1 0 ...
$ drv: Factor w/ 3 levels "4","f","p": 2 2 2 2 2 2 2 2 2 2 ...
$ cty: int 18 21 20 21 16 18 18 16 20 ...
$ hwy: int 29 32 31 30 26 25 27 26 25 28 ...
$ fl: Factor w/ 5 levels "c","d","e","p",..: 4 4 4 4 4 4 4 4 4 4 ...
$ class: Factor w/ 7 levels "compact","suv",..: 2 2 2 2 2 2 2 2 2 2 ...
> qplot(displ, hwy, data=mpg)
> qplot(displ, hwy, data=mpg, color=drv)
```
install.packages("ggplot2")
Installing package into "/Users/song/Library/R/3.0/library"
(as 'lib' is unspecified)
trying URL 'http://cran.rstudio.com/bin/macosx/contrib/3.0/ggplot2_0.9.3.1.tar.gz'
Content type 'application/x-gzip' length 2650041 bytes (2.5 Mb)
opened URL

downloaded 2.5 Mb

The downloaded binary packages are in
/var/folders/t5/s8m11nx1v6xs09nbttkmxr000gn/1//RtmpEfW508/downloaded_packages
> library(ggplot2)
> str(mpg)
'data.frame': 234 obs. of 11 variables:
$ manufacturer: Factor w/ 15 levels "audi","chevrolet",..: 1 1 1 1 1 1 1 1 1 1 ...
$ model : Factor w/ 38 levels "4runner 4wd",..: 2 2 2 2 2 2 2 2 3 3 ...
$ displ : num 1.8 1.8 2 2 2 2 2 2 3.1 3.1 ...
$ cty : int 18 21 20 21 16 18 18 18 18 20 ...
$ hwy : int 29 29 31 30 26 26 26 26 26 28 ...
$ fl : Factor w/ 5 levels "c","d","e","p",..: 4 4 4 4 4 4 4 4 4 4 ...
$ class : Factor w/ 7 levels "2seater","compact",..: 2 2 2 2 2 2 2 2 2 2 ...
> qplot(displ, hwy, data=mpg)
> qplot(displ, hwy, data=mpg, color=drv)
>
adding a geom

```r
> install.packages("ggplot2")
Installing package into '/Users/song/Library/R/3.0/library'
(as 'lib' is unspecified)
trying URL 'http://cran.rstudio.com/bin/macosx/contrib/3.0/ggplot2_0.9.3.1.tgz'
Content type 'application/x-gzip' length 2650041 bytes (2.5 Mb)
opened URL

downloaded 2.5 Mb

The downloaded binary packages are in
   /var/folders/t5/s8m1mmlx1v6x09nbttkx0000gn/T//RtmpEfW503/downloaded_packages
> library(ggplot2)
> str(mpg)
'data.frame': 234 obs. of 11 variables:
$ manufacturer: Factor w/ 15 levels "audi","chevrolet",...: 1 2 3 4 5 6 7 8 9 10 ... 31
$ model      : Factor w/ 38 levels "ar4runner 4wd",...: 1 2 3 4 5 6 7 8 9 10 ... 31
$ displacement: num 1.8 2 2.6 2.8 3 3.1 3.1 3.1 3.2 3.2 ... 3.9 4.0 4.0
$ cylinders  : int 4 4 4 4 6 6 6 4 4 4 ... 4 4 4 4 4 4 4 4 4 4
$ trans       : Factor w/ 10 levels "auto(id15)",...: 1 2 3 4 5 6 7 8 9 10
0 ...
$ drv         : Factor w/ 3 levels "4","f","r": 2 2 2 1 1 1 1 ... 1
$ cty         : int 18 21 20 11 16 18 18 16 20 ...
$ hwy         : int 29 29 31 30 26 26 26 26 25 28 ...
$ fl          : Factor w/ 5 levels "c","d","e","f",...: 4 4 4 4 4 4 4 4 4 4 ...
$ class       : Factor w/ 7 levels "2seat","compact",...: 2 2 2 2 2 2 2 2 2 2 ...

> ggplot(displ, hwy, data=mpg, color=dv)
> ggplot(displ, hwy, data=mpg, geom=("point","smooth"))
```
```r
install.packages("ggplot2")
Installing package into '/Users/song/Library/R/3.0/library'
(as 'lib' is unspecified)
trying URL 'http://cran.rstudio.com/bin/macosx/contrib/3.0/ggplot2_0.9.3.1.tgz'
Content type 'application/x-gzip' length 2650461 bytes (2.5 Mb)
opened URL

The downloaded binary packages are in
/var/folders/t5/s8mlllmxl1v6xss9nbttknnx000gn/T/RtmpEfw508/downloaded_packages
> library(ggplot2)
> str(mpg)
'data.frame': 234 obs. of 11 variables:
$ manufacturer: Factor w/ 15 levels "audi","chevrolet",...: 1 1 1 1 1 1 1 1 1 ...
$ model : Factor w/ 38 levels "amurder 4wd",...: 2 2 2 2 2 2 3 3 3 ...
$ displ : num 1.8 1.8 2 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
$ cyl   : int 4 4 4 4 6 6 4 4 4 ...
$ trans : Factor w/ 10 levels "auto(8v)","auto(13)",...: 4 9 10 1 4 9 1 9 4 1 ...
$ drv   : Factor w/ 3 levels "4","f","r": 2 2 2 2 2 2 1 1 1 ...
$ cty   : int 18 21 21 21 16 18 18 16 16 20 ...
$ hwy   : int 29 29 31 30 26 26 26 26 26 25 28 ...
$ fl    : Factor w/ 5 levels "c","d","e","p",...: 4 4 4 4 4 4 4 4 4 ...
$ class : Factor w/ 7 levels "2seater","compact",...: 2 2 2 2 2 2 2 ...
> qplot(displ, hwy, data=mpg)
> qplot(displ, hwy, data=mpg, color=dv)
> qplot(displ, hwy, data=mpg, geom="point", "smooth")
gem_smooth: method="auto" and size of largest group is <1000, so using loess. Use
'method = x' to change the smoothing method.
```
PART V.
TANGIBLE DATA
Data Anatomy [Civic] A data-driven display Ryoji Ikeda explore the interior of an automobile
Data Necklace: Wear your tweets
Physical data sculpture
World’s data pulsates through one-of-a-kind chandelier
감사합니다.

visualvoice@gmail.com