Data Visualization - Part I

Business Analytics - *Fall 2016*

NYU | TANDON SCHOOL OF ENGINEERING

What is data visualization?

- Representation of data in a pictorial or graphical format.
- A general way of talking about anything that converts data sources into a visual representation:
  - charts, graphs, maps, sometimes even just tables
- Combination of many disciplines
  - statistics, perception, graphic design, cognitive psychology, information design, communications, and data mining
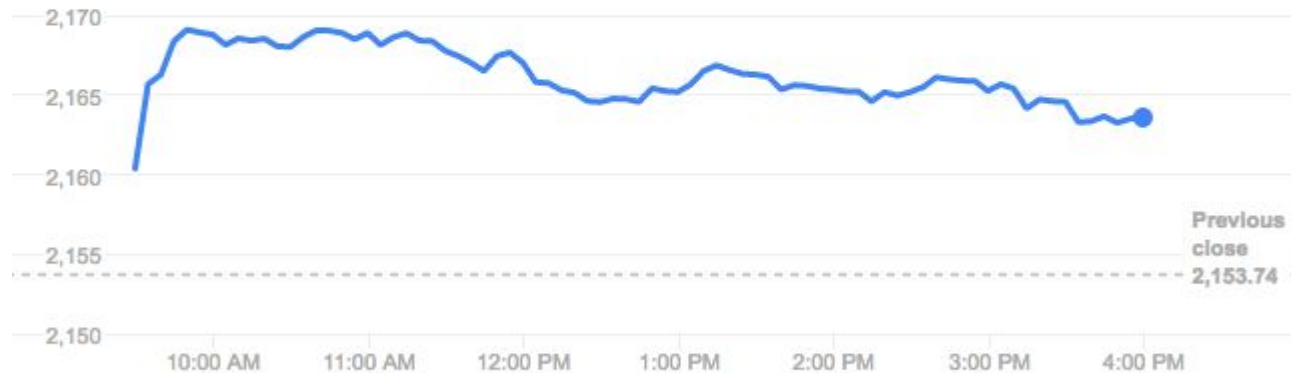
The Most Spoken Languages of the World

foursquare

PLANES, TRAINS, AND AUTOMOBILES **TRANSPORTATION CHECK-INS LAST YEAR**

PLANES
Based on same-day check-ins at two different airports

TRAINS
Based on same-day check-ins at two different train stations

AUTOMOBILES
Based on check-ins on highways and roads
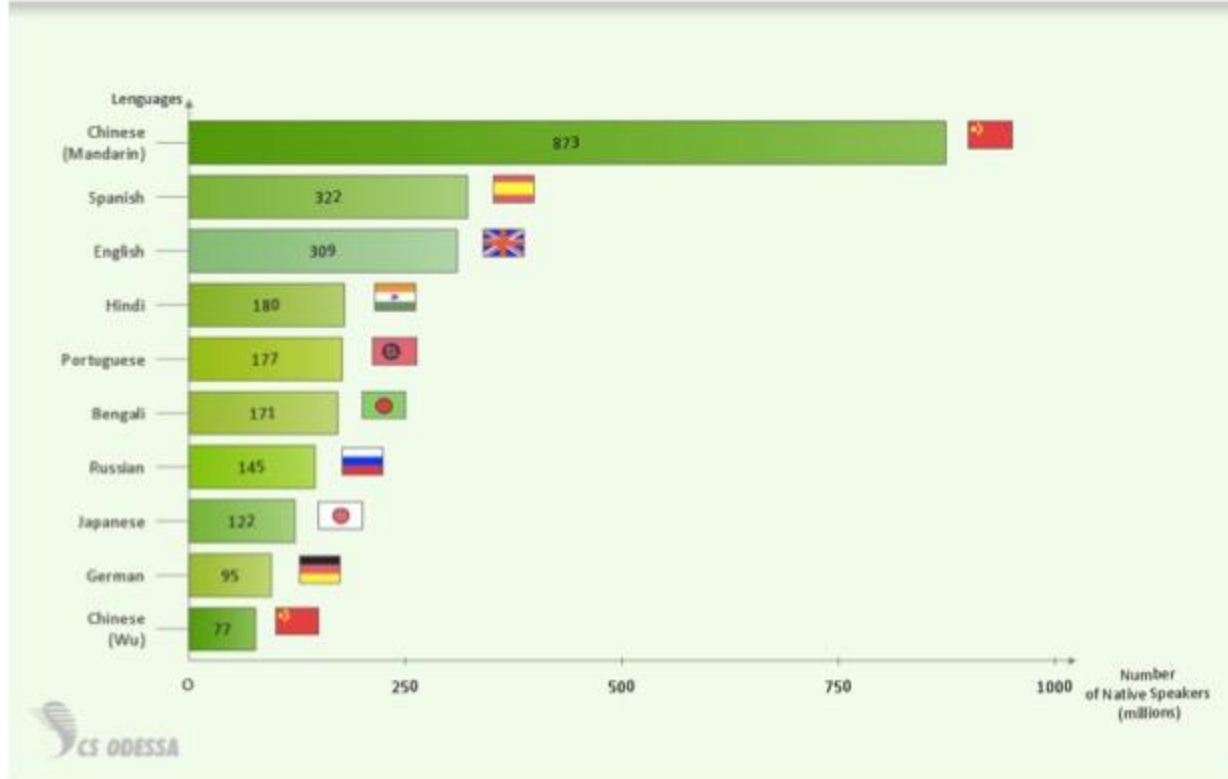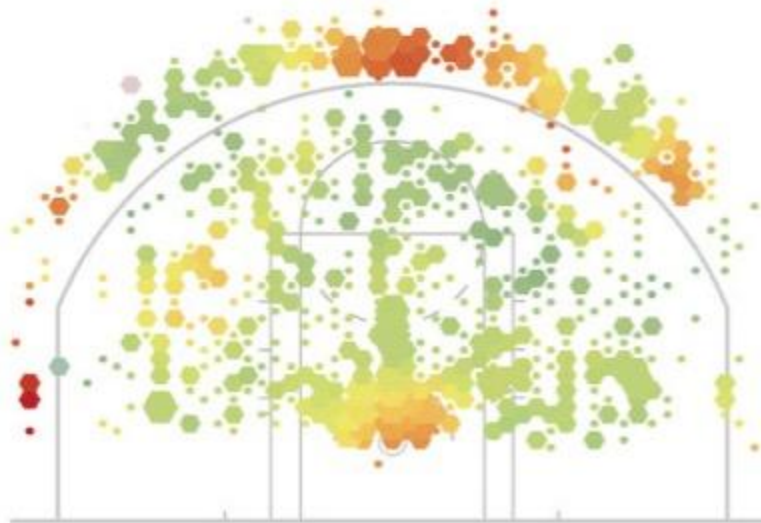
Number of attempts
Low ○ ◯ ⬡ High

Points per region
Low ▬▬▬▬▬ High

The Thunder are effective from almost any area on the court and shoot many more 3-point shots than the league average. Kevin Durant and James Harden are potent from the top of the arc.

Kevin Durant

VIEW: PHOTO | GRAPH

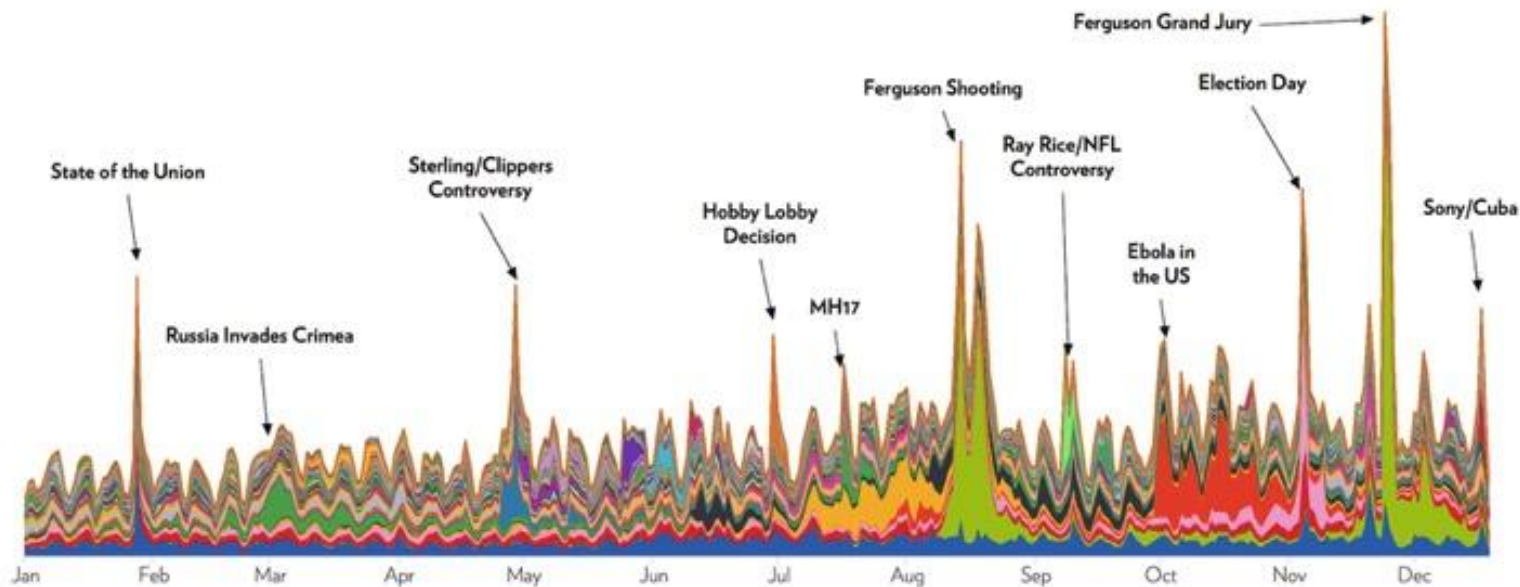TOTAL SHOTS **1,296** | POINTS PER SHOT **1.09** | F.G. PERCENT **49.6%**

THE YEAR IN NEWS from ECHELON INSIGHTS

What America talked about in 2014, as viewed through 184.5 million Twitter mentions.

Ferguson Grand Jury

Ferguson Shooting

Election Day

State of the Union

Sterling/Clippers Controversy

Ray Rice/NFL Controversy

Hobby Lobby Decision

Sony/Cuba

Russia Invades Crimea

MH17

Ebola in the US

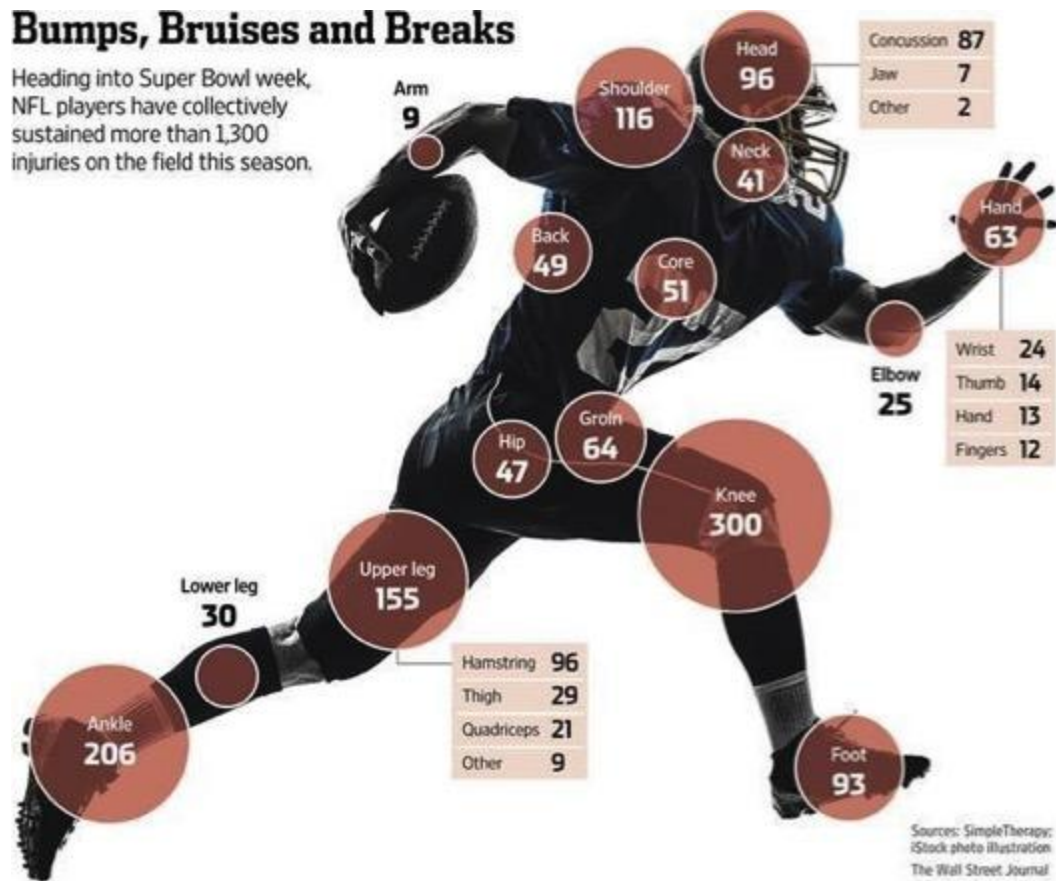Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

NYU TANDON SCHOOL OF ENGINEERING

Bumps, Bruises and Breaks

Heading into Super Bowl week, NFL players have collectively sustained more than 1,300 injuries on the field this season.

Arm 9

Shoulder 116

Head 96

Concussion	87
Jaw	7
Other	2

Neck 41

Hand 63

Wrist	24
Thumb	14
Hand	13
Fingers	12

Back 49

Core 51

Elbow 25

Hip 47

Groin 64

Knee 300

Upper leg 155

Hamstring	96
Thigh	29
Quadriceps	21
Other	9

Lower leg 30

Ankle 206

Foot 93

Sources: SimpleTherapy; iStock photo illustration
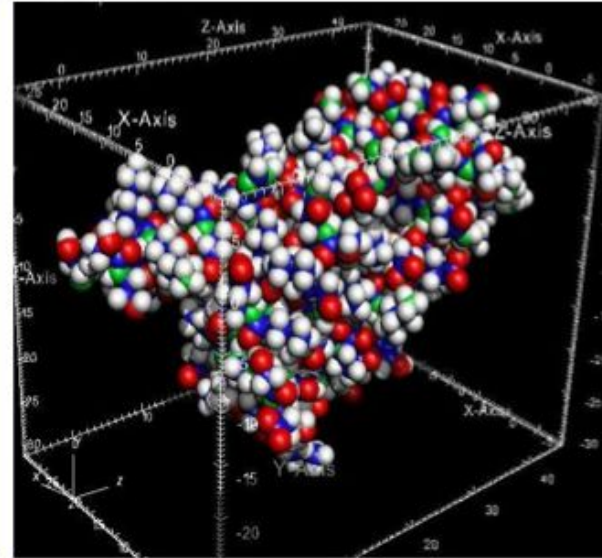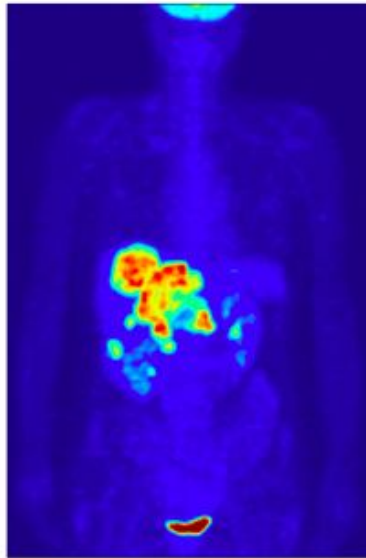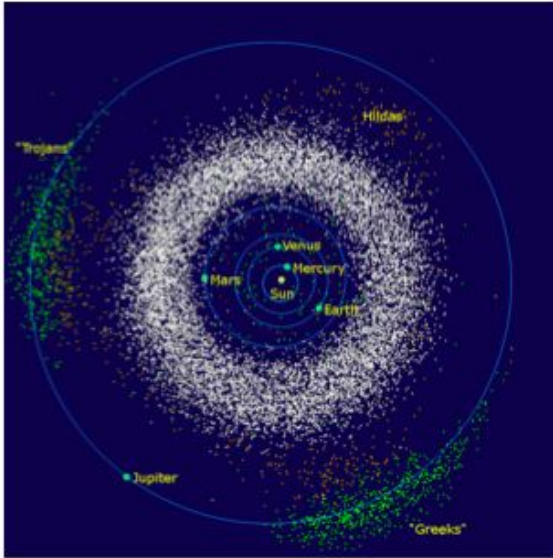The Wall Street Journal

NYU | TANDON SCHOOL OF ENGINEERING

Types of Data Visualization

- Scientific visualization

Types of Data Visualization

- Information visualization
  - Covers statistical charts and graphs as well as other visual/spatial metaphors that can be used to represent data sets that don't have inherent spatial components.
  - Relies more heavily on processing abstract data into a more concrete form that can be more effectively perceived by an observer

Why data visualization?

- Exploring and analyzing
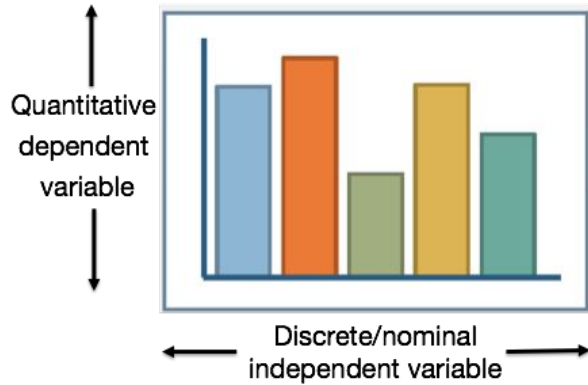
- Presenting and communicating

Types of Exploratory Data Analysis

- Non-graphical methods involve calculation of summary statistics.

- Graphical methods use charts and visual displays to summarize the data.

- Univariate methods look at one variable at a time.

- Multivariate methods look at two or more variables at a time to explore relationships.

- It is almost always a good idea to perform univariate EDA on each component of a multivariate EDA before performing the multivariate EDA.
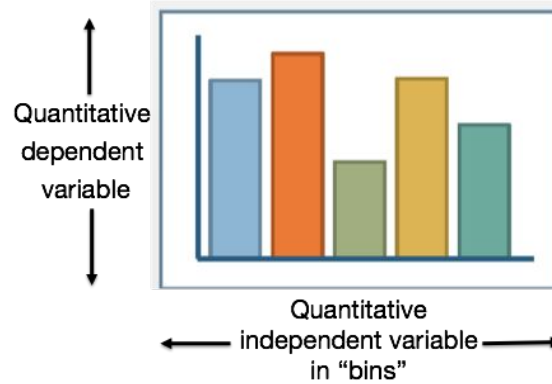
Univariate Graphical EDA

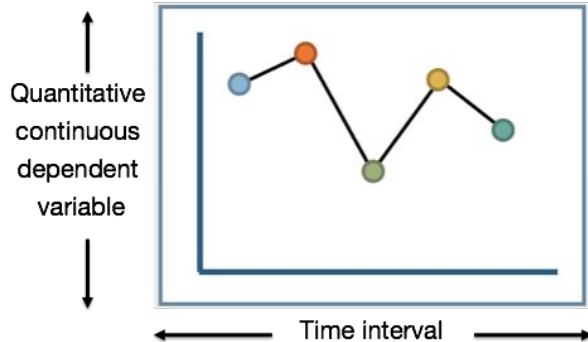- Exploring the distribution of the sample graphically

Bar Chart — Quantitative dependent variable (y-axis); Discrete/nominal independent variable (x-axis)

Histogram — Quantitative dependent variable (y-axis); Quantitative independent variable in "bins" (x-axis)

Box Plot — Max, Upper Quartile, Median, Lower Quartile, Min

Time Series — Quantitative continuous dependent variable (y-axis); Time interval (x-axis)

Scatter Plot — Quantitative dependent variable (y-axis); Quantitative independent variable (x-axis)
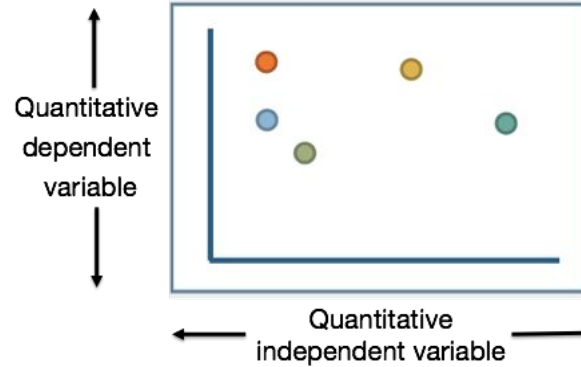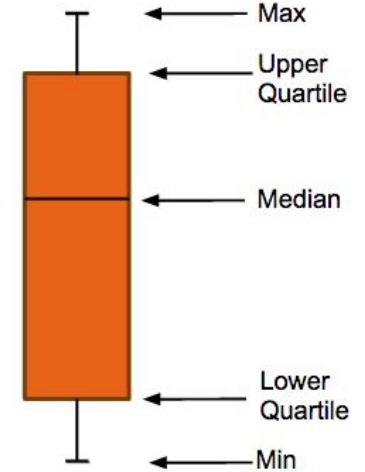
Data Types
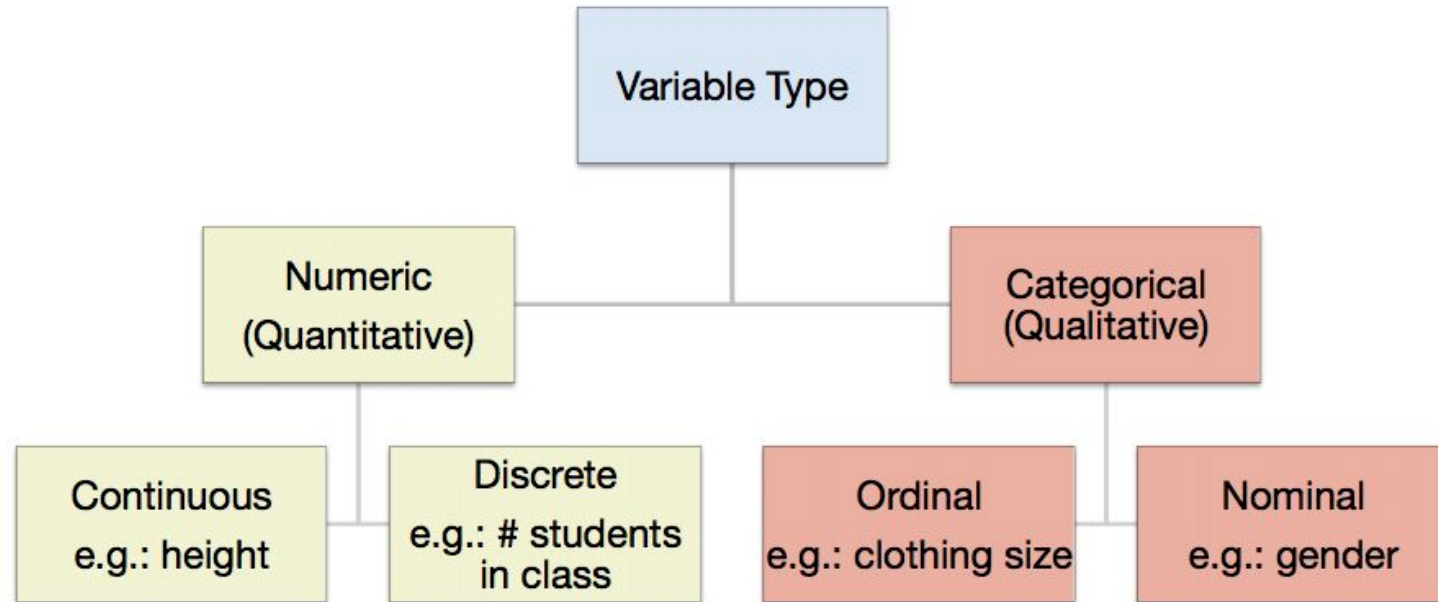
Customer ID	Date	Item ID	Item name	Order ID	Row ID	State	Gross profit	Number of Records	Sale amount	Sales target
18565	1/1/12	1	Calistoga Cheddar	106	1	CA	$8	1	$18	$22
22873	1/1/12	1	Calistoga Cheddar	145	1	CA	$10	1	$24	$22
38254	1/1/12	1	Calistoga Cheddar	179	1	CA	$10	1	$24	$22
46726	1/1/12	1	Calistoga Cheddar	207	1	CT	$8	1	$18	$22
41695	1/1/12	1	Calistoga Cheddar	188	1	DC	$21	1	$48	$67
18348	1/1/12	1	Calistoga Cheddar	102	1	FL	$10	1	$24	$22
4985	1/1/12	1	Calistoga Cheddar	29	1	IA	$10	1	$24	$22
16630	1/1/12	1	Calistoga Cheddar	92	1	IL	$10	1	$24	$22
45494	1/1/12	1	Calistoga Cheddar	202	1	LA	$18	1	$43	$44
46847	1/1/12	1	Calistoga Cheddar	208	1	MI	$50	1	$117	$156
41567	1/1/12	1	Calistoga Cheddar	187	1	MO	$7	1	$16	$22
42714	1/1/12	1	Calistoga Cheddar	191	1	NC	$7	1	$17	$22
19349	1/1/12	1	Calistoga Cheddar	112	1	NM	$10	1	$24	$22
4989	1/1/12	1	Calistoga Cheddar	30	1	NY	$10	1	$24	$22
9136	1/1/12	1	Calistoga Cheddar	50	1	NY	$10	1	$24	$22
42191	1/1/12	1	Calistoga Cheddar	190	1	NY	$22	1	$51	$67
34140	1/1/12	1	Calistoga Cheddar	169	1	OK	$8	1	$18	$22
43793	1/1/12	1	Calistoga Cheddar	195	1	SC	$10	1	$24	$22
29554	1/1/12	1	Calistoga Cheddar	159	1	SC	$14	1	$34	$44
46599	1/1/12	1	Calistoga Cheddar	206	1	VA	$36	1	$85	$89
40867	1/1/12	1	Calistoga Cheddar	184	1	VA	$41	1	$97	$133
210	1/1/12	2	Sonoma Stilton	3	1	AL	$9	1	$32	$48
19460	1/1/12	2	Sonoma Stilton	115	1	CA	$8	1	$28	$48
9828	1/1/12	2	Sonoma Stilton	53	1	CA	$9	1	$32	$48
19866	1/1/12	2	Sonoma Stilton	121	1	NH	$8	1	$28	$48
20503	1/1/12	2	Sonoma Stilton	127	1	TN	$9	1	$32	$48
12501	1/1/12	3	Reyes Raspberry	67	1	CA	$3	1	$8	$8
18703	1/1/12	3	Reyes Raspberry	108	1	MD	$4	1	$12	$8
3793	1/1/12	3	Reyes Raspberry	23	1	PA	$3	1	$8	$8
43770	1/1/12	4	Napa Cracker	194	1	CA	$13	1	$43	$77
33445	1/1/12	4	Napa Cracker	163	1	MN	$18	1	$61	$141

Multivariate Graphical EDA

In-Class Exercises

http://bit.ly/BA101116