Working with Data

Objectives

After completing this section of the course you should understand:

- The different Data Types
- The Data Acquisition & Examination process
- How to transform your Data
- Data Exploration: understanding your data

Data Types

Quantitative
Discrete or continuous data and typically any variables that are **countable**

Categorical
Nominal data offering a **way of distinguishing and labelling** values.

Ordered
Similar to categorical data but with an order within the data

Quantitative

Qualitative

Data Types' Influences

- The **type of exploratory data analysis** you can undertake
- The editorial thinking you establish
- The **specific chart types** you might use
- The **colour choices** and **layout decisions** around composition

Understanding your data

Understanding your data

- Mean vs. Median
- Actual vs. Rank Index vs. Percentile
- Change vs. Actual

Mean vs. Median

Hockey Player Points Scored

6	7	13	17	22	22	24	25	27	27
28	35	35	50	517					

Mean

- Add the values and divide by the number of items in the list
- Great for **consistent values**
- Real center

Median

- Take the number in the middle of the list
- Better when there are **outliers**

Mean
57

Median
25

Actual vs. Rank index vs. Percentile

1, 2, 3 %

GDP data

$15,684,800,000,000

Rank: 1
Percentile: 100th

$3,092,000,000

Rank: 161
Percentile: 20th

Change vs. Actual

##

GDP data

$59,000,000,000

$43,000,000,000

Where should I invest?

GDP Growth: -10%

GDP Growth: +8.5%

Two more concepts

Sample Size & Methodology

» How many entries?
» How data was collected?
» How was the data corrected?

Correlation vs. causation

» Does data move in the same direction?
» Is there a real causality?

Quality and reliability

Impact

» Footnotes
» Legends
» Labels

Converting your data

Common Data Adjustments

» Calculating indexes and ratios

» Calculating percentiles

» Aggregating

» Regrouping

» Converting from Excel/CSV to JSON/XML/MySQL

Calculating indexes and ratios

» Comparing data side-by-side

» Getting the data on a similar scale

» **Example:** Minimum Wage

Calculating percentiles

» Comparing data as part of a whole

» A portion of the entire dataset

» **Example**: GDP

Aggregating

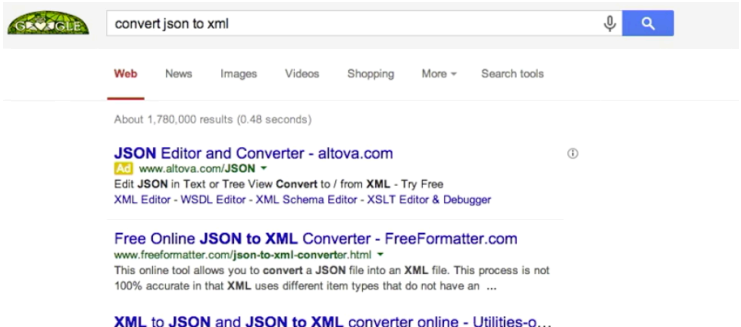
» Pivot the table

» Look at raw data from different granularity level

» **Example**: Excel Pivot Table

Converting from Excel/CSV to JSON/XML/MySQL

- Converting from one form to other
- Move to tabular data
- Move to Web standards



products.json:


```
[{
  "_id": {
    "$oid": "5968dd23fc13ae04d9000001"
  },
  "product_name": "sildenafil citrate",
  "supplier": "Wisozk Inc",
  "quantity": 261,
  "unit_cost": "$10.47"
}, {
  "_id": {
    "$oid": "5968dd23fc13ae04d9000002"
  },
  "product_name": "Mountain Juniperus ashei",
  "supplier": "Keebler-Hilpert",
  "quantity": 292,
  "unit_cost": "$8.74"
}, {
  "_id": {
```

THANKS!

Any questions?

You can find me at

- @manuGenard
- genard@fr.ibm.com