WEEK 2

Getting started with data

IST719 Information Visualization

AGENDA

Week 2

- Assignments for next week
- Data

Due Before Next Class

- READ
  - DP Chapter 1
  - DS Chapter 3 ← read first
  - VT Chapters 3 and 4
- SUBMIT
  - Download and install R: http://www.r-project.org/
  - Complete the 6 chart exercises in VT Chapter 4 ***** **R** portions ONLY *****
  - Upload to BB as a single, multi-page PDF
  - Filename: Ex2_LASTNAME.pdf

NEXT WEEK

Tips for R exercises

- Use both R console and document window.
  - Write your script in a new document window, then open from the console view
- Double check the year of the most recent data in the .CSV files provided on the VT website.
  - Make sure your labels reflect the data in the downloaded file, not just what is shown in the book.
- However, match the layout of the charts as exactly as possible.
  - Goal is to have them look **identical** to the images in the book
  - Resize window BEFORE saving as PDF

READINGS

Telling stories with data

- Patterns
  - Exploratory: Finding questions
  - Changes over time
  - In aggregate
- Relationships
  - In statistics, correlations and causation
  - Multiple variables
  - Between visualizations
- Encodings
  - Circle, bars, colors
  - Labels, legends and keys
  - Context

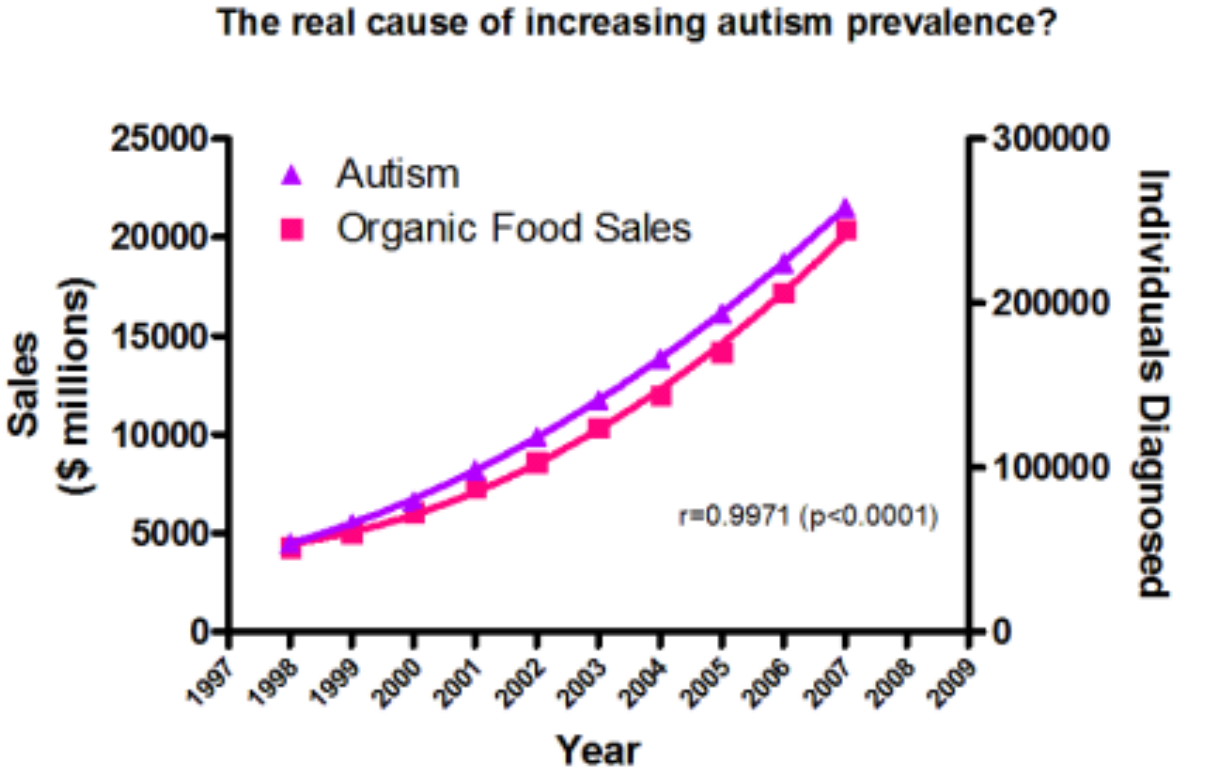
READINGS

Handling data

- Where do data come from?
  - Internet
  - Governments
  - Universities
- How do you get data?
  - Download
  - Scrape
  - Collect
- Using data
  - Identify sources
  - Verify the quality and accuracy of your data
  - Format data
  - First step: explore!

READINGS

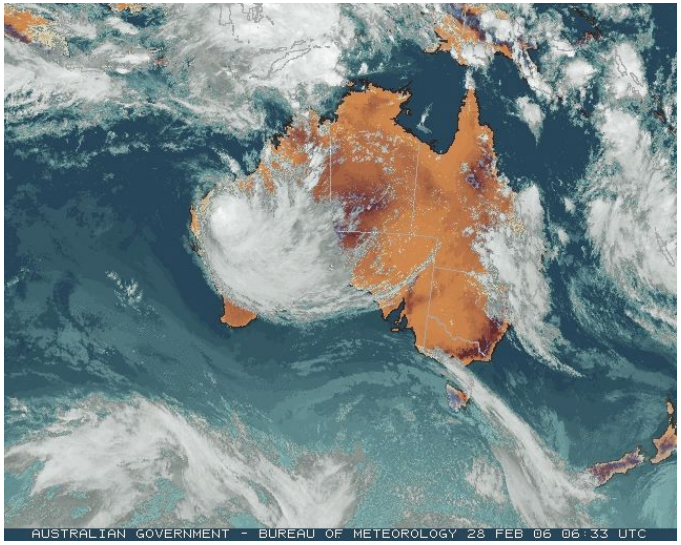
What's the story?

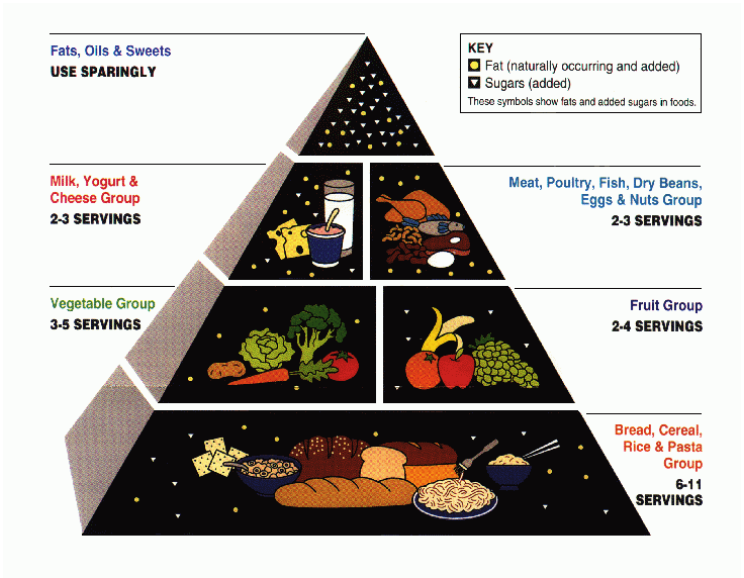


Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act

http://grist.org/list/this-graph-will-make-you-think-twice-about-eating-organic/

What are you interested in?





What is data?

- A plural of datum
- Data → Information → Knowledge → Wisdom

Data and Questions

- What question can your data answer?
- Test out an idea (hypothesis)
- Look for exceptional cases (both good and bad)
- Reduce uncertainty to maximize the chance of a favorable outcome

Data – group work

- Boolean / binary
- Categorical
- Ordinal / Rank
  - Integer, real, float
- Count
  - Integer
- Numeric / Continuous
  - Integer, real, float
- String / character data

Data in R – compound types

- Vectors
- Lists
- Data frame

Data – complex data types

- Date/time
- Documents
- Audio, video, photo
- Relational (network)
- Coordinate
  - (x,y), (x,y,z), (lat,long), (r,θ)

R – in class

For quiz

- Data management functions
  - list.files, read.csv, paste, read.table, load, save, rm, ls
- Data interrogation functions
  - colnames, fix, length, dim, str, summary, table, class, attrbitues
- Accessing data with [] and $
- Plots: barchart, boxplot, density

R – in class

- Review
  - R script, comments, pdfs
- Working with files
  - csv, tsv, Rdata, rda
- Working with data
  - Types
- Data interrogation
  - colnames, dim, str, subsetting, summary, plot

InformationIsBeautiful.net

David McCandless

[The Beauty of Data Visualization](#) (21:27)