

# Types of Quantitative Information

Continuous, real-valued

Integer, ordinal

Categorical

Temporal

Spatial, geographical

1

## Data Information

We're concerned with a body of available data that we'd like to gain insights from.

We want to be informed by the data.

The "information explosion" so widely discussed is about studying data to obtain information.

2

# Data



You marked this message as Junk Mail.

**From:** Geek Meister Leads <members@actnow-profits.net>  
**Subject:** Last chance! Free leads for your business - only 24 hrs.  
**Date:** September 12, 2009 7:21:09 PM CDT  
**To:** Dianne Cook



Mail thinks this message is Junk Mail.

(Instructions be

It doesn't matte  
Internet or trad  
business - you n  
buying from you  
business will be

**From:** Short Term Loans [Credit Dept] <promotions@frankdata.biz>  
**Subject:** Get Up To £400 In Less Than 24 Hours  
**Date:** September 14, 2009 3:06:59 AM CDT  
**To:** Dianne Cook  
**Reply-To:** Short Term Loans [Credit Dept] <promotions@frankdata.biz>



You marked this message as Junk Mail.

**From:** kate.parker@cantv.net  
**Subject:** Dear One In Love  
**Date:** September 14, 2009 5:23:24 AM CDT  
**To:** kare.parker@cantv.net  
**Reply-To:** kate.parker@cantv.net

Dear One In Love

3

# Data

We typically think of data as a table of numbers and text. Here's data collected to study spam email:

ID	Day of Week	Time of Day	Size (kb)	Domain	% Capitals	Sucker	Porn	Spam
1	Mon	7	3	se	14.0	yes	no	yes
2	Sun	8	3	com	12.8	yes	no	yes
3	Thu	7	158	edu	13.0	no	no	no
4	Sun	12	12	uk	1.3	no	yes	yes
5	Fri	12	5	com	13.0	no	no	no
6	Fri	11	2	com	10.0	no	yes	yes

4

# Types of information (variables)

- Categorical: Sucker, porn, spam, day, time, domain
- Continuous / Quantitative: Size, %capitals, ID
- Temporal: Day, Time
- Spatial: Domain

5

## Data



Here's data collected to study airline traffic out of Des Moines:

DayMonthYrTime	Year	Month	DayofMonth	DayOfWeek	DepTime	UniqueCarrier	FlightNum	ActualElapsedTime	Origin	Dest	Distance
1987-10-12 08:05:00	1987	10	12	1	805	TW	666	62	DSM	STL	260
1987-10-18 14:35:00	1987	10	18	7	1435	HP	31	166	DSM	PHX	1149
1987-10-26 11:45:00	1987	10	26	1	1145	TW	594	60	DSM	STL	260
1987-10-28 06:25:00	1987	10	28	3	625	HP	18	32	DSM	OMA	117
1987-10-01 12:45:00	1987	10	1	4	1245	HP	259	36	DSM	CID	104
1987-10-17 22:00:00	1987	10	17	6	2200	HP	763	45	DSM	OMA	117
1987-10-04 07:20:00	1987	10	4	7	720	UA	455	98	DSM	DEN	589
1987-10-03 16:10:00	1987	10	3	6	1610	TW	593	56	DSM	STL	260
1987-10-13 09:31:00	1987	10	13	2	931	UA	694	59	DSM	ORD	299
1987-10-05 06:45:00	1987	10	5	1	645	UA	402	55	DSM	ORD	299

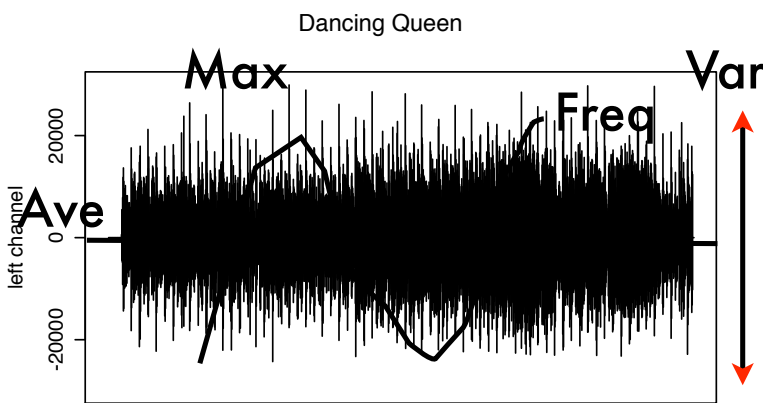
6

# Types of information (variables)

- Categorical: UniqueCarrier, FlightNum,
- Continuous / Quantitative: ActualElapsed Time, Distance
- Temporal: DayMonthYrTime, Year, Month, DayofMonth, DayofWeek, DepTime
- Spatial: Origin, Dest

7

## Data



Here's data collected to study music structure:

	Artist	Type	LVar	LAve	LMax	LFener	LFreq
I Feel Fine	Beatles	Rock	52497242.7	-5.731459	29911	110.6633	110.39972
The Memory of Trees	Enya	New wave	1135493.1	-10.618340	9994	102.1613	155.06430
V2	Vivaldi	Classical	771491.8	21.656094	6936	104.3674	843.83240
Yellow Submarine	Beatles	Rock	39908667.5	-6.261692	29061	109.4744	91.49508
Ticket to Ride	Beatles	Rock	68104547.0	-6.144911	30415	111.5996	107.30853
M5	Mozart	Classical	1568925.6	-6.199379	15839	99.4758	288.44819
Mamma Mia	Abba	Rock	7557437.3	-90.471062	28898	101.6165	48.76513
Love Me Do	Beatles	Rock	28806811.2	-5.745219	24159	109.9581	126.50757
I Want to Hold Your Hand	Beatles	Rock	61257693.6	-6.034068	28502	111.8181	294.67263
Rock Hard Times	Eels	Rock	54651552.8	1.984851	32759	108.8750	312.37864

8

# Types of information (variables)

- Categorical: Artist, Type
- Continuous / Quantitative: LVar, LAve, LMax, LFEner, LFreq
- Temporal:
- Spatial:

9

## Data

Here's data  
collected on wages:

**Subject  
demographics**

	id	ged	black	hispanic	hgc
	1	0	1	1	8
36	1	0	0	0	9
53	0	0	0	1	7
122	0	0	0	0	12
...					

**Repeated  
measurements**

	id	lnw	exper
	1.491	0.015	
	1.433	0.715	
	1.469	1.734	
	1.749	2.773	
	1.931	3.927	
	1.709	4.946	
	2.086	5.965	
	2.129	6.984	
36	1.982	0.315	
36	1.798	0.983	
36	2.256	2.040	
36	2.573	3.021	
36	1.819	4.021	
36	2.928	5.521	
36	2.443	6.733	
36	2.825	7.906	
36	2.303	8.848	
36	2.329	9.598	
...			

10

# Types of information (variables)

- Categorical: id (but there are a lot of them), race, ged, hgc
- Continuous/Quantitative:lnw
- Temporal:exper
- Spatial:

11

## Scales

### Data comes in different forms, resolution and scales:

- A table of information might be accompanied by a set of events, eg Napoleon's March.
- Geographical: some data might be recorded at a latitude and longitude, but other information might be reported along political boundaries
- Temporal: Date format is not easy to handle in most packages, needs to be converted to separate variables: day, month, year, order
- Longitudinal measurements such as health records of people, are made at repeated time points, but the descriptive information about the patient (male, female) doesn't change over time.

12

# Dimensionality

Each column of the data table is commonly considered to be a new dimension.

One column of data = univariate

Two columns of data = bivariate

Three columns = trivariate

More than three = multivariate

13

# Descriptive Terminology

- Words to describe the one variable (univariate) distribution of values: symmetric, skewed, outliers, clusters, unimodal, bimodal, multimodal.
- Describing relationships (associations, correlations) between variables: linear, non-linear, positive, negative, weak, moderate, strong.

14