Gesture recognition techniques

Definitions

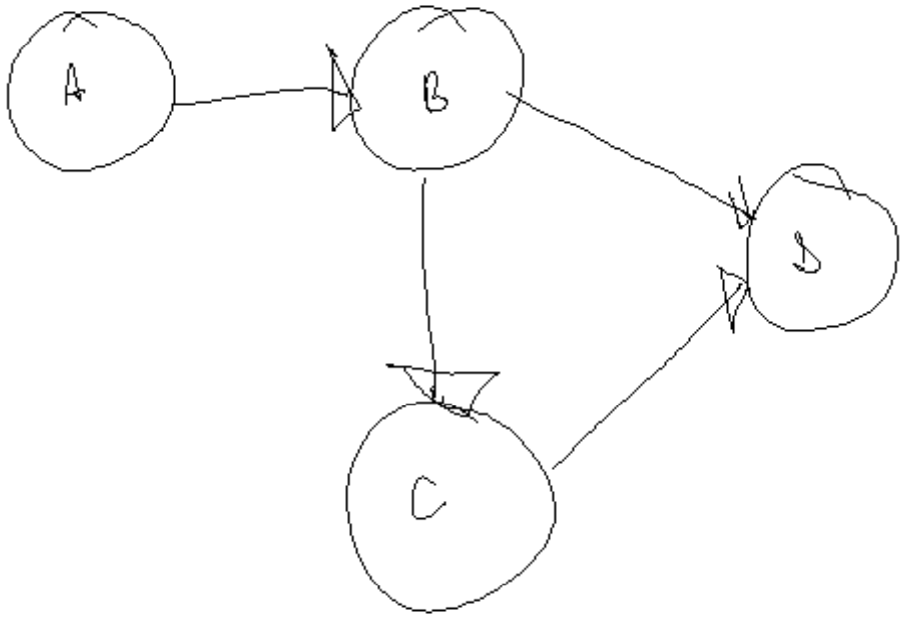
- Gesture – some type of body movement
  - a hand movement
  - Head movement, lips, eyes
- Depending on the capture this could be
  - Digital ink
  - Accelerometer data
  - Actual body movement detected by vision analysis (ie what the vision group do
- With digital ink
  - Stroke – time series of x,y points may include pressure and pen tilt data
  - Sometime people use the term ‘gesture’ to mean an editing stroke – delete, cut, copy etc

More definitions

- Bounding box, the smallest enclosing rectangle
- Distance is measured in himetric units = .01mm

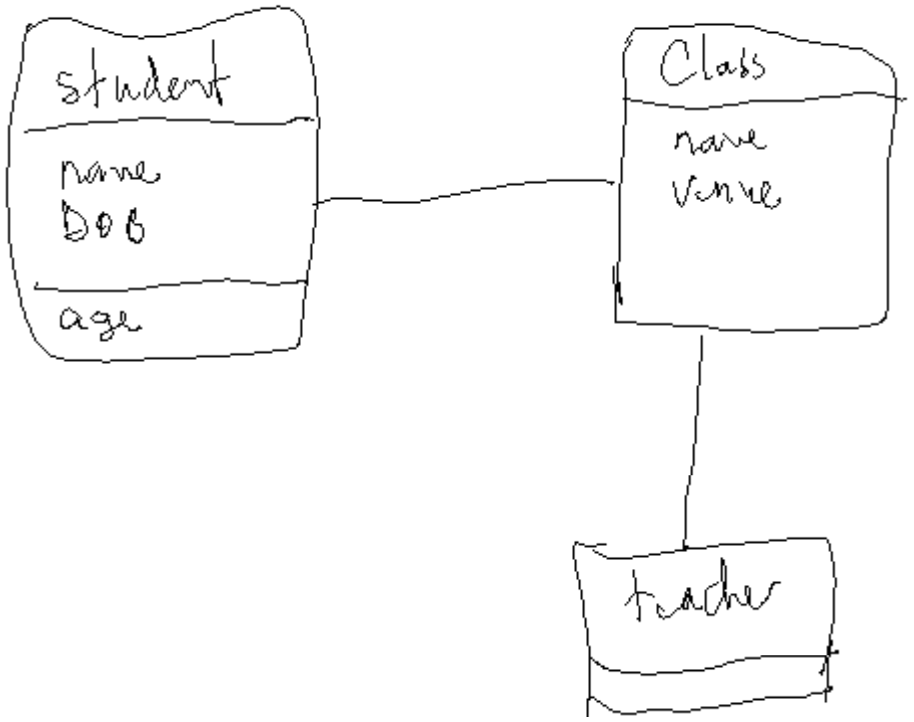
Dissecting a diagram

- Components
  - Nodes
    - Contain label
  - Arch/edge
    - Line and arrow
- Semantic meaning
  - Actions
  - Connections
  - Directed flow



What are the components here?

- What is the semantic meaning?



Where to start?

- Step 1 is dividing writing and drawing because there is a fundamental semantic are different
  - At least for languages that use an alphabet – languages such as Chinese there is some relationship
- The Microsoft OS (tablet and vista) has a pretty good writing recognizer
- It has a ‘divider’ that will separate writing from drawing
  - Trouble is it classifies nearly everything as writing (about 80%)
- How do you tell writing from drawing?

How do you find a better way?

- Look for effective ink features
  - We found 47!
    - Pressure
    - Time
    - Intersections
    - Curvature
    - OS values
  - We have another ~20 on the list we have thought of since

Time Features

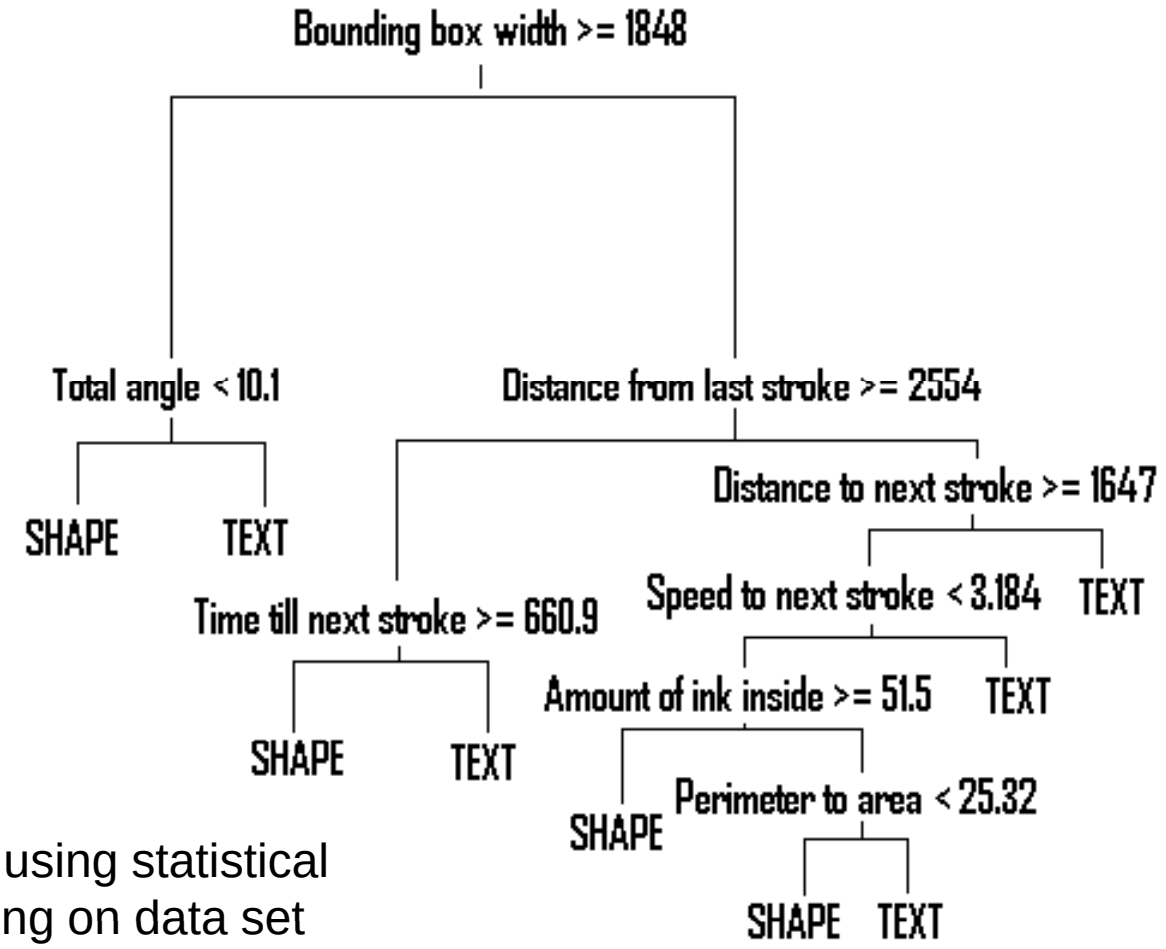
Feature	Description	Origin
Total duration	Total duration of the stroke from pen up to pen down.	(Rubine 1991)
Maximum speed	Maximum speed when drawing the stroke.	Adapted from (Rubine 1991)
Minimum speed	Minimum speed when drawing the stroke.	
Average Speed	Mean average speed when drawing the stroke.	
Time from last stroke	The time between the current stroke and the previous stroke in the sketch. Not applicable to the first stroke of a diagram.	New
Time till next stroke	The time between the current stroke and the next stroke in the sketch. Not applicable to the last stroke of a diagram.	
Speed from last stroke	Speed (distance/time) between the current stroke and the previous stroke in the sketch. Not applicable to the first stroke of a diagram.	
Speed to next stroke	Speed (distance/time) between the current stroke and the next stroke in the sketch. Not applicable to the last stroke of a diagram.	
# Speed minima	The number of extreme minima in the speed values for the stroke, this excludes the minima that occur at the beginning and end of the stroke for pen up/down events.	Adapted from (Sezgin, Stahovich et al. 2001)

From: Patel, Rachael. 2007, Exploring better techniques for diagram recognition, MSc Thesis, UoA

Data mining techniques

- R – statistical programming, used to partition the data
- Weka – a range of different techniques all in the one package.

Classification Tree



Created using statistical partitioning on data set from about 30 people

This is an improvement

- Increased % correct and decreased % incorrect
- About 80% correct
- How can we do better?

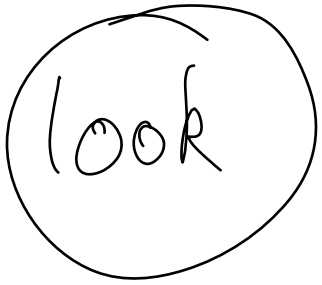
Features and Algorithms

- There are two parts to this divider
- The features – those most discriminating from the 47 examined
- The algorithm – a tree is a blunt instrument – more fuzziness is needed and also some idea of certainty
  - Probably the future down the tree the less confidence there is in the result

Context

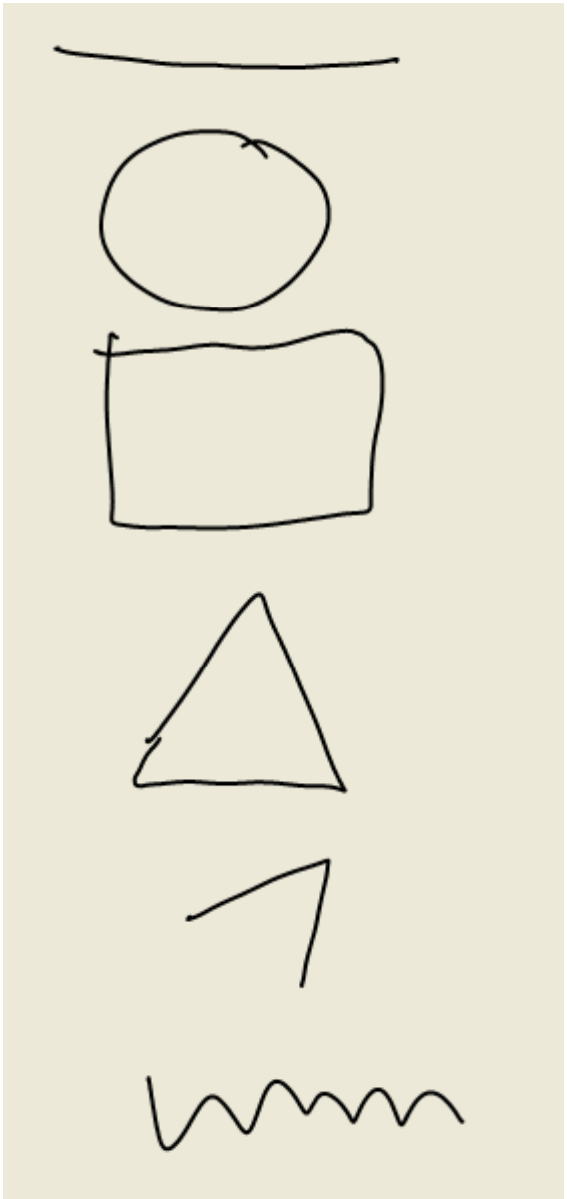
- The spatial relationship between strokes

- Semantic Rules – could this possibly be a character/drawing - containment



Shape recognition

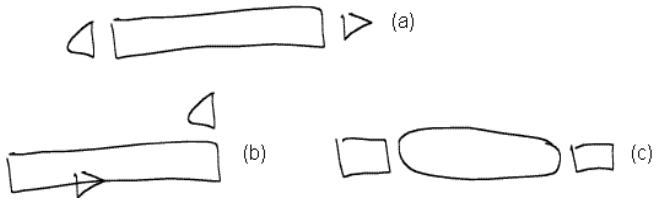
- There is quite a small set of basic shapes
- Most common approach Rubine's Algorithm[1]
  - 13 features – defined by heuristics
  - Hidden Markoff model (statistical pattern matching)
- This looks at single stroke
  - It isn't too difficult to join strokes



[1] Rubine, D. *Specifying gestures by example*. in *Proceedings of Siggraph '91*. 1991: ACM

Putting the bits together

- How would we classify these?
- Types
- Spatial relationships
  - Containment
  - Adjacency
- Connectivity



Inkkit component recognizer

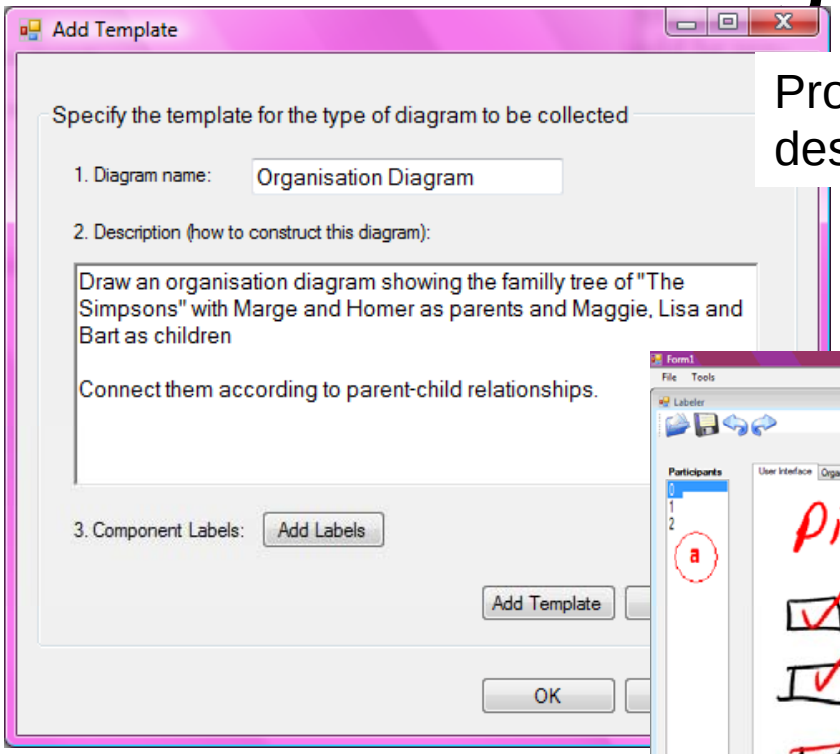
- Analyses relationships of shapes in library
- Does similar analysis on diagram
- Builds a huge tree of all possible relationships, each with a probability
- Parses the tree looking for most probably component
- Assigns that, removes it from the tree and searches again.

Other techniques

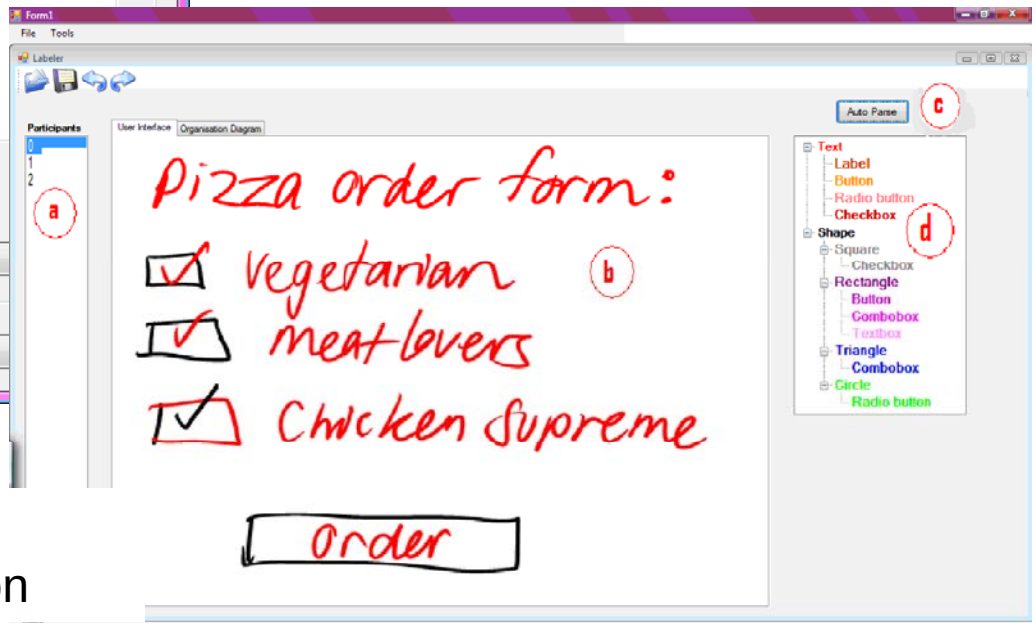
- Fuzzy logic
- User written rules
- Semantic networks
- Bayesian networks

- All of these rely on calculated ink features
  - How people arrived at the particular features they have chosen is rarely stated

Data Repositories for Recognition

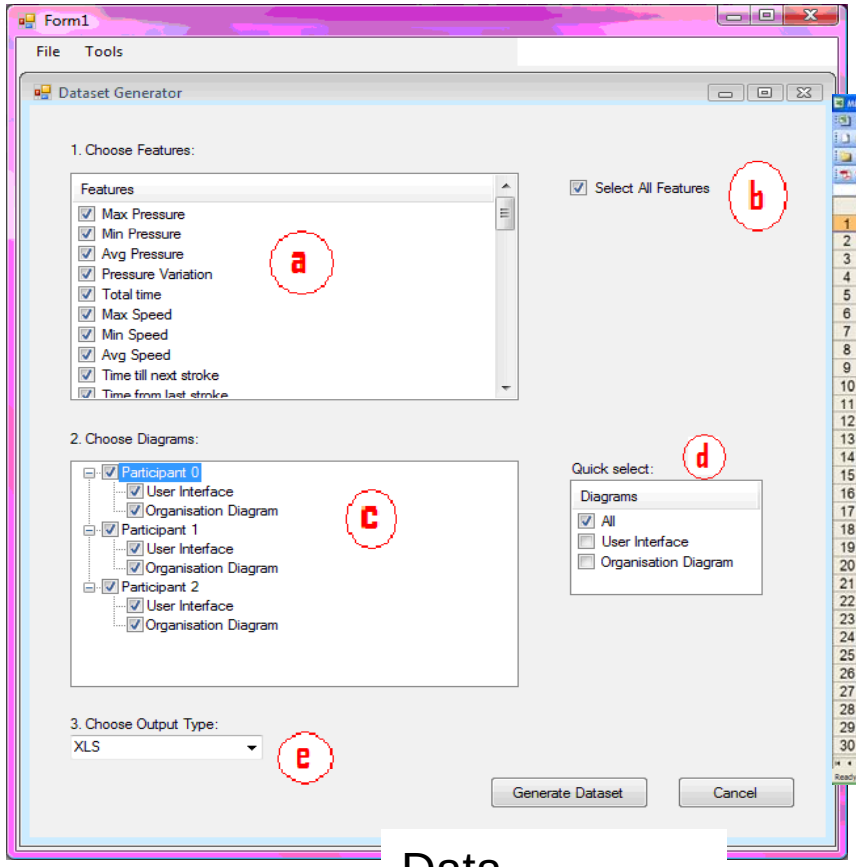


Problem description

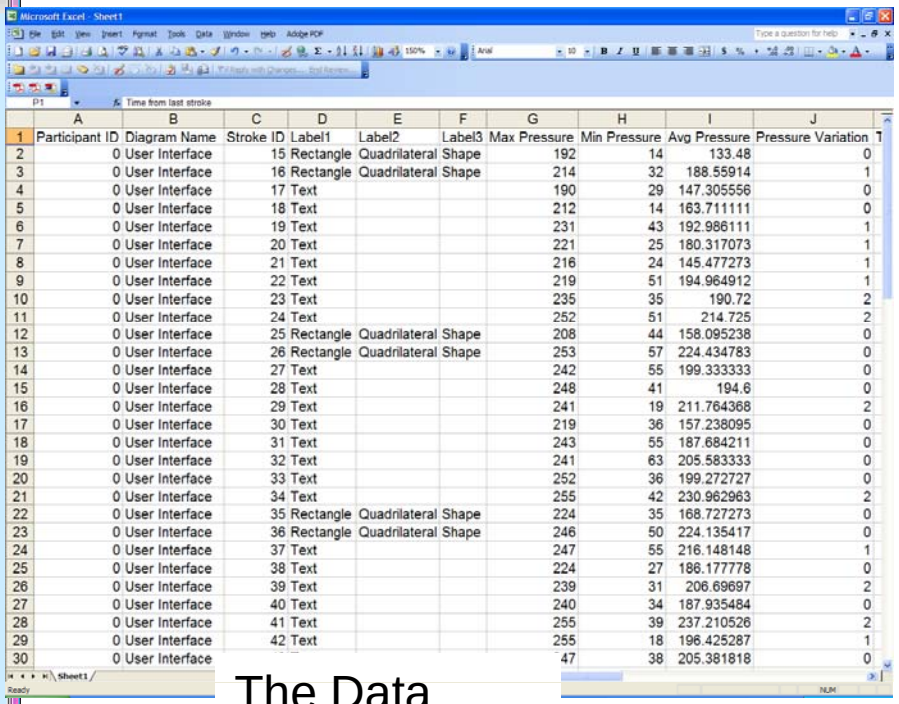


Data Collection

Dataset generation



Data generation



	A	B	C	D	E	F	G	H	I	J
1	Participant ID	Diagram Name	Stroke ID	Label1	Label2	Label3	Max Pressure	Min Pressure	Avg Pressure	Pressure Variation
2	0	User Interface	15	Rectangle	Quadrilateral	Shape	192	14	133.48	0
3	0	User Interface	16	Rectangle	Quadrilateral	Shape	214	32	188.55914	1
4	0	User Interface	17	Text			190	29	147.305556	0
5	0	User Interface	18	Text			212	14	163.711111	0
6	0	User Interface	19	Text			231	43	192.986111	1
7	0	User Interface	20	Text			221	25	180.317073	1
8	0	User Interface	21	Text			216	24	145.477273	1
9	0	User Interface	22	Text			219	51	194.964912	1
10	0	User Interface	23	Text			235	35	190.72	2
11	0	User Interface	24	Text			252	51	214.725	2
12	0	User Interface	25	Rectangle	Quadrilateral	Shape	208	44	158.095238	0
13	0	User Interface	26	Rectangle	Quadrilateral	Shape	253	57	224.434783	0
14	0	User Interface	27	Text			242	55	199.333333	0
15	0	User Interface	28	Text			248	41	194.6	0
16	0	User Interface	29	Text			241	19	211.764368	2
17	0	User Interface	30	Text			219	36	157.238095	0
18	0	User Interface	31	Text			243	55	187.684211	0
19	0	User Interface	32	Text			241	63	205.583333	0
20	0	User Interface	33	Text			252	36	199.272727	0
21	0	User Interface	34	Text			255	42	230.962963	2
22	0	User Interface	35	Rectangle	Quadrilateral	Shape	224	35	168.727273	0
23	0	User Interface	36	Rectangle	Quadrilateral	Shape	246	50	224.135417	0
24	0	User Interface	37	Text			247	55	216.148148	1
25	0	User Interface	38	Text			224	27	186.177778	0
26	0	User Interface	39	Text			239	31	206.69697	2
27	0	User Interface	40	Text			240	34	187.935484	0
28	0	User Interface	41	Text			255	39	237.210526	2
29	0	User Interface	42	Text			255	18	196.425287	1
30	0	User Interface					47	38	205.381818	0

The Data