

Disentangling a complex variable

Rob Kabacoff, Ph.D.

March 9, 2017

Untangling a complex variable

We are occasionally faced with a data set variable that contains multiple pieces of information. Consider the variable **charges** in the following dataset **test_data**.

id	charges
1	Murder,Kidnapping,Burglary
2	burglary: assault and battery
3	murder
4	

Multiple pieces of information are stored in this variable. In this case these pieces are delineated by a comma or colon. As is, the variable is not useful.

The **untangle** function can be used to create a set of dummy codes from this variable that will be more useful. This results in a new dataset with dummy codes:

```
test_data <- untangle(data=test_data, x="charges", pattern="[:,]", verbose=TRUE)
```

```
## Warning: package 'stringr' was built under R version 3.3.2
```

```
## [Unique Codes]
```

```
## murder
```

```
## burglary
```

```
## kidnapping
```

```
## assault and battery
```

Table 2: New test_data

id	charges	murder	burglary	kidnapping	assault_and_battery
1	Murder,Kidnapping,Burglary	1	1	1	0
2	burglary: assault and battery	0	1	0	1
3	murder	1	0	0	0
4		0	0	0	0