

Data Coding Report

How to use this tool

Coding data is an cyclical process; multiple attempts and refinements at coding data are required in order to be accurate. This tool should be used as a beginning cycle of the data coding process.

Coding data is a highly subjective process; this tool is less subjective, so take what it says with a grain of salt.

Plan to use this tool to help you get started or get unstuck during the coding process. It will give you some suggestions on what to look out for and how to proceed.

The First Cycle

Often, the first cycle of coding data involves going through each response and abstracting them individually. For example, there may be a long response complaining about the bathrooms in their dorm. This could be coded as “Housing” or “Bathrooms.” Identifying key words such as “dorm” and “bathroom” can be helpful in this first cycle. Below are the top ten keywords identified by this tool.

```
##      topKeywords$keyword[1:10]
## 1                people
## 2             activities
## 3                 area
## 4                 arts
## 5                chapel
## 6                 easy
## 7             education
## 8             facility
## 9                 hill
## 10                job
```

The Second Cycle

The second cycle involves grouping together responses based on their similarities. For example, “Bathrooms” and “Kitchens” could be grouped together as “Facilities.” Looking for the relationship between keywords help with this. The relationships between keywords are shown below.

```
##      keywords principalComponents directions
## 1    diversity                1    positive
## 2     people                  2    negative
## 3   cultural                  1    positive
## 4 opportunity                1    positive
## 5      city                   5    positive
## 6      dpac                   5    positive
## 7  proximity                  5    positive
## 8      close                  5    negative
```

## 9	restaurant	1	positive
## 10	family	5	negative
## 11	downtown	3	positive
## 12	art	1	positive
## 13	lot	5	positive
## 14	diverse	2	negative
## 15	weather	5	negative
## 16	food	3	negative
## 17	life	4	negative
## 18	nice	2	negative
## 19	access	3	positive
## 20	culture	4	negative
## 21	duke	3	positive
## 22	events	3	positive
## 23	good	2	negative
## 24	live	3	negative
## 25	medical	4	negative
## 26	community	2	negative
## 27	convenience	4	negative
## 28	entertainment	2	positive
## 29	hospital	3	positive
## 30	neighborhood	3	negative
## 31	variety	1	positive
## 32	work	5	positive
## 33	activities	5	negative
## 34	area	1	negative
## 35	arts	4	negative
## 36	chapel	3	negative
## 37	easy	2	positive
## 38	education	1	positive
## 39	facility	4	negative
## 40	hill	3	negative
## 41	job	5	negative
## 42	living	2	negative
## 43	parks	5	positive
## 44	place	1	negative
## 45	public	2	positive
## 46	quiet	3	negative
## 47	rtp	3	negative
## 48	school	1	positive
## 49	shopping	4	negative
## 50	thing	5	positive

So you can get a better idea of how the responses might be grouped, here is a sampling of the groups generated by this tool.

##	responses
## 39	PROXIMITY TO JOBS
## 6	DUKE UNIVERSITY & MED CENTER; DPAC
## 13	DPAC, RESTAURANTS, DIVERSITY
## 38	GREENWAYS FOR BICYCLING; DPAC; REVITALIZED DOWNTOWN; DBAP
## 43	DPAC
## 61	FAMILY IS NEARBY
## 5	OPPORTUNITIES; CLOSE TO RTP & JOBS

```

## 9                                CULTURE; FAMILY-LIKE
## 16                ACCESS TO ARTS-WEATHER-IMPROVEMENTS IN DOWNTOWN
## 2                DOWNTOWN; RESTAURANTS; EVENTS; JOB OPPORTUNITIES
## 56                                BASEBALL TEAM
## 162                PROGRESSIVE ATTITUDE & CULTURAL OPPORTUNITY
## 103                OPENNESS OF MAYOR & COUNCIL TO CITIZENS
## 30                                SCHOOLS
## 195                CONVENIENCE TO OTHER AREAS & WEATHER
## 62                                HOSPITAL FACILITIES
## 7  CONVENIENCE TO ALL FACILITIES--SHOPPING, MEDICAL, EDUC, ETC
## 21                                EXCELLENT MEDICAL CARE
## 51                                GOOD MEDICAL SERVICE
## 1                                MEDICAL FACILITIES-CULTURE
## 52                TAXES-NEIGHBORHOOD-CONVENIENT TO I-40-RTP-CHAPEL HILL
## 22                                CHEAPER THAN RALEIGH & CHAPEL HILL
## 11                PROXIMITY TO PARKS-NEIGHBORHOOD LIFE-FOOD OPTIONS
## 71                                PROXIMITY TO CHAPEL HILL
## 63                NOT AS CONGESTED AS CHAPEL HILL OR RALEIGH
##  clusters
## 39                1
## 6                1
## 13                1
## 38                1
## 43                1
## 61                2
## 5                2
## 9                2
## 16                2
## 2                2
## 56                3
## 162                3
## 103                3
## 30                3
## 195                3
## 62                4
## 7                4
## 21                4
## 51                4
## 1                4
## 52                5
## 22                5
## 11                5
## 71                5
## 63                5

```

Coding Sentiment

The final step in the coding process is to breakdown the groups by their sentiment. This step is optional, but in certain circumstances it is not enough to know that a response is about “Housing.” The response may indicate that they like or dislike housing, which are very different. This tool categorizes responses as having positive, negative, or neutral sentiment. A list of the most positive and negative responses are listed below. However, coding sentiment should be much easier and more accurate for a human than a computer, so trust yourself.

```

##                                vectorOfResponses
## 191                            QUALITY FOOD & ENTERTAINMENT
## 29                             EXCELLENT MEDICAL CARE
## 23                             FREEDOM; CONVENIENT
## 293                           CULTURAL & EDUCATIONAL OPPORTUNITIES
## 17                             ARTS, CULTURAL & ENTERTAINMENT OPPORTUNITIES
## 219                            LACK OF SNOW
## 94                             HOSPITALS
## 234                           NOISE AND JUST LIKE CITY OF MEDICINE
## 108 NO PLACE LIKE HOME; ESPECIALLY IF IT'S ALL YOU KNOW
## 151                            CLOSE TO HOSPITALS
## 147                            A BIG CITY, BUT NOT TOO BIG
##      sentiment$ave_sentiment
## 191                            1.2701706
## 29                             1.1547005
## 23                             1.0960155
## 293                           1.0392305
## 17                             0.9000000
## 219                            -0.4330127
## 94                             -0.2500000
## 234                           -0.1889822
## 108                            -0.1581139
## 151                            -0.1443376
## 147                            -0.1181139

```

Additional Tips

- Each of the above steps may take multiple cycles themselves.
- Since coding data is so subjective, be mindful of your own biases and how they might arise when coding.

To help with this, run your coding by another intern to get a fresh pair of eyes. Another intern can check your biases and help contribute to an additional cycle of coding.

For more detailed information about coding qualitative data, see [Coding qualitative data: a synthesis guiding the novice](#)